PHYSICALLY PLAUSIBLE AND TEMPORALLY COHERENT MOTION IN 4D GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Synthesizing novel views of dynamic scenes is challenging in computer vision. Without explicit constraints coupling Gaussian distributions to motion, 3DGS often yields non-physical renderings, a limitation amplified in 4DGS. The temporal dimension increases complexity and leads to geometric degradation. Although higher point cloud density can improve accuracy, it notably increases computational cost. To achieve physically consistent and temporally coherent rendering of dynamic scenes, we introduce PhysMotion4DGS, a physics-aware 4D Gaussian Splatting framework. It introduces 4D physical realism consistency regularization, which constrains the motion velocity of Gaussian distributions using cues from both 2D image space and 3D physical space. Furthermore, PhysMotion4DGS introduces temporal consistency regularization to ensure smooth, continuous motion over time, preventing unnatural dynamics. To capture global structure and fine details, we adopt a progressive multi-scale training strategy that optimizes Gaussian distributions in stages, enabling high-quality rendering without increasing point cloud size. Experimental results demonstrate that PhysMotion4DGS outperforms leading methods in quality and efficiency, improving PSNR by 5.58 dB, reducing LPIPS by 80.00%, and decreasing training time by 23.08%.

INTRODUCTION

Neural rendering methods have advanced 3D rendering using neural networks to implicitly model volumetric scene properties, such as density and color, enabling high-quality novel view synthesis Mildenhall et al. (2020); Pumarola et al. (2021); Park et al. (2021b); Li et al. (2022b); Wu et al. (2020); Cheng et al. (2023); Park et al. (2021a); Tretschk et al. (2021). However, the reliance on dense sampling and complex network inference results in high computational costs, limiting its suitability for real-time applications. To improve efficiency, 3D Gaussian Splatting (3DGS) Kerbl et al. (2023) explicitly represents scenes with Gaussian primitives and leverages GPU-accelerated methods. Despite this progress, dynamic scene rendering methods are still hard to maintain the temporal continuity crucial to realistic motion modeling. These limitations are amplified in real-time applications, e.g., immersive VR/AR, where stereoscopic views, high refresh rates, and tight motion-to-photon budgets demand temporally stable rendering under rapid viewpoint changes. Image-level constraints often lead to temporal flicker, cross-view drift, and appearance inconsistencies that degrade presence.

Existing dynamic scene rendering methods lack effective motion modeling. Real-time 4DGS models time independently, failing to capture the evolution of the 3D structure; this per-time-step treatment disrupts temporal continuity. Deformation-based methods constrain Gaussian distributions only at separate time steps, without modeling the underlying physical dynamics. Effective modeling faces three recurrent issues. First, the motion representation of the same object often lacks temporal consistency, especially in regions affected by occlusion or motion blur. Second, object deformations frequently appear physically implausible, resulting in stretching, collapsing, and jittering artifacts. Finally, the heavy reliance on 2D constraints causes entanglement between depth and motion. This induces spatial drift, such as geometric and occlusion errors, and temporal inconsistencies, including flicker and unstable motion. One feasible way is to add more Gaussian primitives for motion modeling, which increases representational capacity but introduces two challenges. First, enforcing spatiotemporal consistency typically requires additional constraints whose cost grows with the number of primitives, reducing optimization efficiency and scalability. Second, the model should

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.

remain robust to noisy or sparse observations and to dynamic scene topology, such as occlusions, disocclusions, and merge events.

To construct physically accurate motion representations, PhysMotion4DGS jointly models the dynamic evolution of Gaussian distributions from the dual perspectives of motion fields and physical dynamics. This ensures that predicted motion is not only visually plausible but also physically interpretable and consistent. By leveraging optical flow as a dense proxy for observed motion, PhysMotion4DGS accurately captures object movement in complex scenes, enabling realistic rendering of physical interactions such as object deformations and environmental contact. Furthermore, we explicitly model physical velocity as a dynamic parameter within the 4D Gaussian representation, and constrain its temporal evolution through physically grounded regularization inspired by classical mechanics. This prevents abrupt velocity changes that often lead to unrealistic deformations in dynamic scenes. Our proposed 4D physical realism consistency regularization further improves the robustness of PhysMotion4DGS by suppressing irregular and unstable Gaussian motion. It also preserves topological consistency between the predicted motion fields and the underlying geometric structures throughout dynamic evolution. Our approach substantially enhances the temporal coherence and physical plausibility of reconstructed motion in complex dynamic scenes.

To resolve scene degradation caused by abrupt motion or sudden object appearances in dynamic environments, PhysMotion4DGS introduces image-space supervision by projecting 3D velocity vectors into the same coordinate space as optical flow. Using the pinhole camera model, the velocities of 3D Gaussian primitives are projected onto the image plane, yielding 2D displacements. Pixel-level constraints are then applied to these projections, and by minimizing the discrepancy between the predicted 2D displacements and the observed optical flow, PhysMotion4DGS aligns 3D motion predictions with observed image-space motion, ensuring visual consistency between physical dynamics and rendered appearance. To further prevent error accumulation from relying solely on 2D motion cues, PhysMotion4DGS lifts the 2D optical flow into 3D space by combining it with depth information, estimating accurate 3D velocities. These reconstructed 3D velocities are then compared against the predicted velocities of the Gaussian primitives, ensuring physical consistency in the 3D space. This two-way projection mechanism, aligning both 2D and 3D motion, enables PhysMotion4DGS to maintain faithful correspondence between observed image dynamics and physically grounded motion in the scene.

To address temporal inconsistencies in dynamic scene reconstruction, we extend the 4D physical realism consistency regularization to operate over continuous unit time intervals. Prior dynamic neural rendering methods focus on static appearance accuracy. The lack of temporal coherence results in abrupt transitions and physically implausible motion, especially under complex deformations, occlusions, and non-rigid interactions. Such artifacts degrade both perceptual continuity and physical plausibility in the rendered output. Our temporal regularization addresses this by explicitly constraining the evolution of 4D Gaussian primitives, enforcing smoothness in motion attributes such as position and velocity across time. Formulated continuously, the constraint ensures gradual, coherent dynamics, allowing Gaussians representing the same underlying structure to maintain consistent motion trajectories. Importantly, this formulation is model-agnostic regarding frame rate or time step resolution, enabling robust performance across variable temporal sampling conditions. By anchoring the predicted motion to temporally coherent trajectories, the regularization addresses high-frequency fluctuations and discontinuities that commonly emerge in unconstrained settings. As a result, rendered sequences exhibit improved temporal fidelity, with smooth object motion and enhanced visual stability.

To further improve physical realism and robustness in dynamic scene rendering, PhysMotion4DGS employs a progressive multi-scale training strategy consisting of two stages. In the coarse stage, PhysMotion4DGS focuses on learning global motion consistency by incorporating supervision from 4D physically-aware consistency, enabling it to capture stable, large-scale motion patterns and build resilience to noise during early training. This forms a strong foundation for fine-grained rendering while reducing the risk of overfitting to noisy and incomplete data. In the fine stage, the training shifts to refining local motion details and enhancing physical accuracy. 4D physics motion consistency regularization aligns predicted 3D velocities with those derived from depth-enhanced optical flow, ensuring that each Gaussian motion is image-consistent and physically plausible. PhysMotion4DGS progressively optimizes motion, from global structure to local detail, to produce physically grounded, temporally coherent renderings of complex, time-varying scenes without increasing the number of Gaussian primitives. Its hierarchical training integrates image-based supervision with

physics-informed constraints, improving performance across diverse motion regimes and yielding stable rendering with realistic dynamics. In summary, our contributions are as follows:

- The introduction of 4D physical realism consistency regularization effectively addresses physical inconsistencies in dynamic scenes, ensuring that motion renderings remain physically interpretable and consistent over time.
- A 4D physically-aware consistency regularization, which is the first approach to explicitly intergrate the velocity of Gaussian distributions with observed optical flow, enforcing rendered motion to conform to real-world physical dynamics.
- A 4D physics motion consistency regularization, which matches velocities derived from optical flow and depth with the model's predicted velocities, ensuring that motion remains dynamically consistent with real-world physical behavior.
- An optimization framework for 4D Gaussian representations that progressively refines spatiotemporal consistency and physical realism in dynamic scenes, without increasing point cloud complexity.

RELATED WORK

We first review classical rendering, then neural methods, focusing on NeRF and 3D Gaussian Splatting (3DGS), which have advanced rendering fidelity, yet accurately modeling motion in dynamic scenes remains challenging.

Neural Rendering Traditional rendering methods Broxton et al. (2020); Collet et al. (2015); Li et al. (2018); Guo et al. (2015) based on geometry, materials, and lighting are effective for static scenes but struggle with dynamic environments, complex interactions, and high-frequency detail Thies et al. (2019); Waechter et al. (2014); Kutulakos & Seitz (2000); Penner & Zhang (2017); Mildenhall et al. (2019); Srinivasan et al. (2019); Zhou et al. (2018). Neural rendering has emerged as a powerful alternative, with neural radiance field (NeRF) achieving high-fidelity view synthesis by encoding volumetric scenes via multi-layer perceptrons (MLPs) Mildenhall et al. (2020); Hu et al. (2022); Verbin et al. (2022); Kopanas et al. (2022); Bemana et al. (2022); Yan et al. (2023); 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); Guo et al. (2023). However, its dense sampling and MLP inference result in high computational costs Fang et al. (2022); Gao et al. (2021); Yi et al. (2023); Liu et al. (2023); Li et al. (2022b); Zhang et al. (2022); Xu et al. (2022a); Abou-Chakra et al. (2022); Mildenhall et al. (2021); Barron et al. (2021); Zhang et al. (2020); Pumarola et al. (2021); Park et al. (2021a;b), limiting real-time applicability. To reduce NeRFs' computational overhead, various acceleration strategies have been proposed to speed up rendering Chen et al. (2022); Song et al. (2023); Attal et al. (2023); Li et al. (2022a); Wang et al. (2023b); Müller et al. (2022); Fridovich-Keil et al. (2022). 3D Gaussian Splatting (3DGS) Kerbl et al. (2023) provides a more efficient solution by representing scenes with Gaussian primitives and leveraging rasterization for fast rendering. While 3DGS achieves real-time performance, it lacks temporal coherence in dynamic scenes, relying on frame-wise updates that lead to motion artifacts and inconsistent representations. To address this, we introduce a physics-aware motion framework that effectively enforces consistency in image and 3D physical space to enable high-fidelity, physically plausible, real-time dynamic scene rendering.

Dynamic Point-Based Rendering NeRF represents a scene as a continuous volumetric field encoded by a neural network to generate high-quality, continuous novel view synthesis and is capable of modeling complex materials and lighting variations Park et al. (2021a); Tretschk et al. (2021); Fang et al. (2022); Wang et al. (2023c); Gao et al. (2022); Peng et al. (2023); Xu et al. (2022b); Lin et al. (2023); Wang et al. (2023b). Different from traditional methods Li et al. (2012); Collet et al. (2015); Kanade et al. (1997); Zitnick et al. (2004); Du et al. (2021), by augmenting points with learned features and rendering them via CNNs, NeRFs encode not just geometry, but also rich appearance attributes such as color, texture, and transparency, enabling high-fidelity reconstruction in complex scenes. This fusion of volumetric rendering and neural architectures allows fine-grained, data-driven control over point attributes. 3D Gaussian Splatting (3DGS) extends this paradigm by representing scenes as Gaussian point clouds, naturally supporting soft blending, surface detail, and

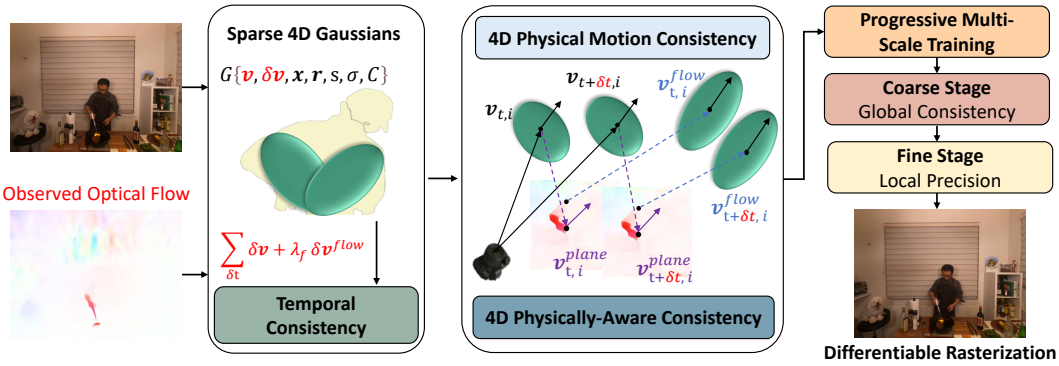


Figure 1: **Framework Overview.** First, 4D physics-aware consistency aligns projected 3D velocity with observed optical flow. Then, 4D physics-based motion consistency matches predicted per-primitive velocities to flow-/depth-derived estimates. Third, temporal consistency enforces smooth evolution over time. Furthermore, a progressive, multi-scale training strategy optimizes motion from global structure to local detail, yielding physically plausible, temporally stable renderings.

complex lighting, while avoiding aliasing and sparsity issues common in traditional methods. However, in dynamic scenarios, 3DGS struggles with weak temporal correspondences, limiting its ability to capture consistent motion under fast or non-linear dynamics. Existing 3DGS extensions Li et al. (2023); Huang et al. (2024); Yang et al. (2023); Wu et al. (2023); Yang et al. (2024); Duan et al. (2024) often result in temporal artifacts and unrealistic motion transitions. PhysMotion4DGS addresses these limits by adding physics-based constraints and higher-dimensional Gaussian features to enforce spatiotemporal consistency in motion and appearance. A progressive multi-stage training scheme optimizes Gaussians from global to local scales, yielding temporally coherent renderings with improved physical realism.

METHOD

PHYSICAL MOTION 4D GAUSSIAN SPLATTING

As illustrated in Figure 1, PhysMotion4DGS imposes physically grounded constraints on object motion to maintain spatiotemporal consistency. Dynamic objects are represented as Gaussian primitives whose state includes a physically interpretable 3D velocity, enabling per-primitive motion prediction over time. A 4D physics-aware consistency regularization projects each Gaussian primitive's 3D motion through a pinhole camera to generate predicted 2D optical flow aligned with observed flow to improve image-space motion accuracy. In parallel, a 4D motion-consistency term fuses depth with optical flow to lift 2D flow into 3D and estimate per-primitive velocities, enforcing agreement between predicted and observed 3D motion and improving physical fidelity. Temporal-consistency regularization further enforces smooth velocity and depth evolution between adjacent frames, suppressing jitter and discontinuities. To efficiently capture dynamics across scales, we adopt a hierarchical training scheme: a coarse stage establishes global motion alignment via 4D flow consistency, followed by a refinement stage that sharpens local motion and structure via 4D velocity consistency. Integrating these multi-level, physics-aware constraints yields a robust, scalable framework that captures complex object motion while addressing temporal artifacts and geometric drift.

OPTIMIZATION SCHEME

Projecting Velocity into Image Space To further improve the accuracy of the final rendering, we incorporate observed optical flow to constrain the motion of Gaussians within the PhysMotion4DGS framework. Specifically, we project the 3D motion of each Gaussian onto the image plane using a pinhole camera model and then design a regularization term to enforce consistency between this projected motion and the observed optical flow. Consider a Gaussian primitive with: 3D position $\mathbf{x}_{t,i} = \left(X_{t,i},\, Y_{t,i},\, Z_{t,i}\right)$ and 3D velocity $\mathbf{v}_{t,i} = \left(v_{x,t,i},\, v_{y,t,i},\, v_{z,t,i}\right)$. Assume a camera with focal length f and the unit time interval δt. For small displacements, the 3D motion can be linearly

approximated and projected onto the image plane as 2D optical flow:

$$\delta \mathbf{p}_{t,i} = \Big(f \, \frac{v_{x,t,i} \, Z_{t,i} - X_{t,i} \, v_{z,t,i}}{Z_{t,i}^2}, \quad f \, \frac{v_{y,t,i} \, Z_{t,i} - Y_{t,i} \, v_{z,t,i}}{Z_{t,i}^2} \Big) \, \delta t. \tag{1}$$

This equation captures the perspective projection effect, mapping the velocity into a 2D displacement in the image coordinate system.

4D Physically-Aware Consistency Let $\mathbf{f}_t \in \mathbb{R}^2$: the observed optical flow vector from time t to $t + \delta t$. To ensure that the projected motion aligns with the observed flow, we define the following regularization term:

$$\mathcal{L}_{\text{4DPhysically-Aware}} = \sum_{t,i} \Big\| \delta \mathbf{p}_{t,i} - \mathbf{f}_{t,i} \Big\|^2. \tag{2}$$

where $\delta \mathbf{p}_{t,i}$ represents the predicted 2D displacement from the projection, while $\mathbf{f}_{t,i}$ is the observed optical flow. By minimizing the difference between these two, PhysMotion4DGS is encouraged to align its predicted velocity with the actual scene motion, leading to physically interpretable dynamics. Our regularization ensures not only accurate rendering but also coherent and realistic motion over time.

Elevating Optical Flow to Higher Dimensions We utilize camera intrinsics and depth information to lift optical flow into a displacement. This displacement is then compared with predicted 3D velocity. By elevating optical flow from the 2D image plane to 3D space using depth, we constrain motion predictions effectively, reducing error accumulation that typically occurs when working solely in the image domain. The flow-guided 3D velocity is used as a supervisory signal, ensuring that the predicted velocity for each Gaussian primitive remains physically consistent. Given $\mathbf{x}_t \in \mathbb{R}^3$: the 3D position of a Gaussian primitive at time t, d_t: the depth at the corresponding time, $K \in \mathbb{R}^{3\times3}$: the camera intrinsic matrix, which includes the focal lengths (f_x, f_y) and the principal point (c_x, c_y). We define the projection function $\Pi(\cdot)$ that maps a 3D Gaussian primitive $\mathbf{x} = (x, y, z)$ onto the image plane as:

$$\Pi(\mathbf{x}) \;=\; \begin{bmatrix} f_x \cdot \frac{x}{z} + c_x \\ f_y \cdot \frac{y}{z} + c_y \end{bmatrix}. \tag{3}$$

Using the depth value $d_{t+\delta t}$, we define an inverse projection function Π^{-1} that back-projects a pixel coordinate (with its depth) to a 3D Gaussian primitive. The 3D displacement inferred from the optical flow is computed as:

$$\mathbf{x}_{t+\delta t}^{\text{flow}} \;=\; \Pi^{-1}\Big(\Pi(\mathbf{x}_t) \;+\; \mathbf{f}_t, \; d_{t+\delta t} \Big). \tag{4}$$

where $\Pi(\mathbf{x}_t)$ gives the projection of the Gaussian primitive at time t onto the image plane, and adding the optical flow $\mathbf{f}_t$ provides its expected location in time $t + \delta t$. Finally, we compute the flow-based 3D velocity by:

$$\mathbf{v}_t^{\text{flow}} \;=\; \frac{\mathbf{x}_{t+\delta t}^{\text{flow}} - \mathbf{x}_t}{\delta t}. \tag{5}$$

4D Physics Motion Consistency To ensure that predicted 3D velocity aligns with the flow-guided 3D velocity inferred from depth and optical flow, we enforce a consistency loss:

$$\mathcal{L}_{\text{4DPhysicsMotion}} \;=\; \frac{1}{N} \sum_{i=1}^{N} \Big\| \mathbf{v}_t^{(i)} - \mathbf{v}_{t,\text{flow}}^{(i)} \Big\|^2. \tag{6}$$

where N is the number of Gaussian primitives, $\mathbf{v}_t^{(i)}$ is predicted velocity for the i-th Gaussian primitive, and $\mathbf{v}_{t,\text{flow}}^{(i)}$ is the flow-guided 3D velocity. This loss term enforces that the predicted motion is consistent with the physically plausible motion derived from the combined depth and optical flow, improving motion estimation accuracy.

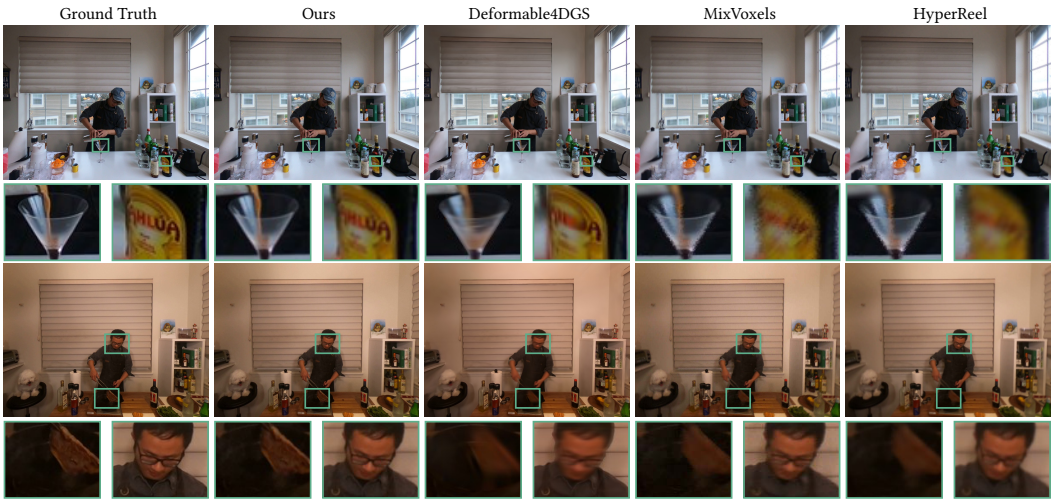


Figure 2: **Qualitative Comparison on Plenoptic Video Dataset**. PhysMotion4DGS renders sharp and coherent motion boundaries, demonstrating superior temporal consistency and physical realism across diverse motion scenarios.

Temporal Consistency To incorporate the temporal dimension, we further extend the loss over continuous time intervals rather than computing it at a single time instance. The loss can be written as:

$$\mathcal{L}_{\text{Temporal}} = \sum_{i=1}^{N} \left(\left\| \mathbf{v}_{t+\delta t}^{(i)} - \mathbf{v}_{t}^{(i)} \right\|^2 + \lambda_{\text{f}} \left\| \mathbf{v}_{t+\delta t,\text{flow}}^{(i)} - \mathbf{v}_{t,\text{flow}}^{(i)} \right\|^2 \right). \quad (7)$$

To ensure real-time rendering, our approach normalizes across unit time intervals without explicitly depending on the separate time steps. This design makes the approach scalable and consistent across varying temporal resolutions. Enforcing temporal regularization encourages PhysMotion4DGS to produce temporally coherent motion predictions that agree with the observed optical flow at all time instances. Our temporal extension is critical for dynamic scenes where the consistency of motion from time to time is important.

4D Physical Realism Consistency Leveraging the temporal dimension, we develop a unified framework for both 3D static and 4D dynamic rendering by combining 4D physical realism consistency regularization with 4D physically-aware consistency regularization. The final 4D physically-aware motion regularization is expressed as:

$$\mathcal{L}_{\text{4DPhysics}} = \lambda_1 \mathcal{L}_{\text{4DPhysicsMotion}} + \lambda_2 \mathcal{L}_{\text{4DPhysically-Aware}} + \lambda_3 \mathcal{L}_{\text{Temporal}}. \quad (8)$$

where λ_o is the hyperparameter. Our extended temporal regularization strategy not only enforces that the predicted velocities align with the depth-guided and flow-guided velocities but also encourages smooth motion over time, which is critical for realistic and physically plausible dynamic scene rendering.

Progressive Multi-Scale Training In our approach, we adopt a progressive multi-scale training method to achieve both optimized rendering performance and improved training efficiency. During the coarse stage, the primary goal is to establish stable spatial consistency across the entire scene. At this stage, we emphasize the use of 4D physically-aware consistency regularization, which ensures PhysMotion4DGS to learn global motion patterns in a stable manner. The regularization term encourages PhysMotion4DGS to predict motion that aligns with observed optical flow over time, ensuring that the broader motion structure of the scene is captured accurately. This global motion consistency is important in the early stages of training, where noise, especially in depth estimation, can substantially impact the quality of the predictions. By focusing on coarse-scale motion consistency, PhysMotion4DGS becomes less sensitive to errors caused by noisy or incomplete depth information, which is common in the early phases of training. This step lays a solid foundation for

future stages, enabling PhysMotion4DGS to learn the primary motion characteristics of the scene without getting distracted by fine details that may not be as reliable.

The fine stage takes the previously established motion patterns and refines them, focusing on scene detail and texture accuracy. At this stage, we incorporate 4D physics motion consistency regularization, which further improves PhysMotion4DGS's ability to predict detailed local motion. The velocity consistency term ensures that the motion at a finer scale is physically accurate, aligning the predicted 3D velocities with the motion inferred from optical flow and depth data. During this phase, we calculate two distinct loss terms at different spatial scales. The coarse scale loss focuses on ensuring overall consistency of the scene, leveraging 4D physically-aware consistency to maintain a stable global motion structure. In contrast, the fine scale loss focuses on improving local physical accuracy, using 4D physics motion consistency to refine the finer details of the scene. This approach allows PhysMotion4DGS to capture subtle movements and rapid local variations that are important for achieving realistic motion in complex dynamic environments.

The combination of these two training stages and loss terms creates a multi-scale framework that effectively balances the need for capturing broad, global motion patterns while simultaneously refining the fine details and textures of the scene. By separating the training process into coarse and fine scales, PhysMotion4DGS is able to learn and optimize the different levels of motion complexity in a structured way. This progressive training method is especially effective in handling scenes with rapidly changing motion or complex structures, where different parts of the scene may require different levels of detail. The coarse stage stabilizes global motion, while the fine stage enables PhysMotion4DGS can capture intricate, localized variations. The multi-scale approach enhances PhysMotion4DGS's robustness, ensuring that it performs well even in challenging dynamic environments with diverse motion patterns.

EXPERIMENTS

Datasets We conduct a thorough evaluation of PhysMotion4DGS on two widely recognized dynamic scene reconstruction benchmarks. **Plenoptic Video Dataset** Li et al. (2022b) employs a synchronized array of 15–20 GoPro cameras to capture a variety of challenging phenomena, including rapidly flickering flames, complex shadow interplay, partially translucent materials, and glossy, specular surfaces, at a high resolution of 1352×1014. Each sequence comprises 300–400 frames recorded at 60 fps, with 17–20 distinct camera views used for training and one held-out view reserved for quantitative testing. We extract SfM points from the first frame, cap them at 100 K, and, with our 4D Gaussian splatting and lightweight point cloud, converge in 14 K iterations. **D-NeRF Dataset Pumarola et al. (2021)** provides monocular, background-free video resized to 800×800, with 50–200 training and 20 test frames. We train for 20K iterations, upsample ×2, prune every 8K, and stop growth at 15K.

Table 1: **Quantitative Comparison on Plenoptic Video Dataset**. We compare our approach with both NeRF-based and Gaussian-based methods. Our approach outperforms the baselines in PSNR. *: 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
STG Li et al. (2023)	30.43	0.94	0.16	62 min	27.51
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	34.00	0.95	0.05	60 min	45.00

Efficiency We pre-allocate device buffers for visibility lists, per-tile work queues, and intermediate gradients using a caching allocator to avoid per-iteration allocations, and we apply selective checkpointing to retain only gradients that materially affect convergence. Parallelism and I/O are organized to keep the device saturated. We overlap communication with computation on dedicated CUDA streams. Data ingestion employs pinned host memory, double buffering, and asynchronous H2D copies; image decoding and augmentation are pipelined with GPU execution to eliminate stalls. We discard Gaussian primitives outside the frustum and those whose projected ellipse footprint falls below a pixel-dependent threshold. We also adapt the order of spherical harmonics for distant or sub-pixel primitives. In inference, we tighten footprint culling to maximize FPS, while training uses more conservative settings to avoid bias.

RESULTS

EVALUATION ON PLENOPTIC VIDEO DATASET

Table 1 presents a quantitative comparison between PhysMotion4DGS and recent NeRF-based and Gaussian-based approaches on the Plenoptic Video dataset. PhysMotion4DGS achieves a PSNR of 34.00, exceeding all baselines by a large margin, outperforming the best existing method, MixVoxels, by over 3.1. It also attains the lowest LPIPS score, indicating superior perceptual quality and fine detail preservation, and maintains a strong SSIM of 0.95, reflecting high structural consistency over time. While individual baselines such as K-Planes report competitive LPIPS or SSIM scores, they fall short in either reconstruction accuracy or efficiency, highlighting the balanced performance of our approach. In terms of training time, PhysMotion4DGS converges in just 60 minutes, faster than all NeRF-based counterparts, thanks to our progressive multi-scale training strategy that accelerates optimization without compromising quality. Critically, PhysMotion4DGS achieves real-time rendering at 45.00 FPS, which is nearly three times faster than MixVoxels and substantially ahead of all other baselines. This combination of speed and fidelity positions PhysMotion4DGS as a practical solution for time-sensitive applications such as interactive dynamic scene exploration and live relighting. Furthermore, as shown in Figure 2, the motion boundaries around dynamic regions, such as human hands, moving objects, and flowing liquid, are sharp and well-aligned, reflecting the ability to preserve structural consistency even under complex deformations.

EVALUATION ON D-NERF DATASET

Table 2 reports a quantitative comparison of our method against leading NeRF-based and Gaussian-based baselines on the D-NeRF dataset. Our approach achieves a PSNR of 37.00, outperforming Deformable4DGS by 4.01 and exceeding TiNeuVox by 4.13. We also report a SSIM of 0.99 and LPIPS of 0.01, the latter being the lowest among all methods, indicating superior perceptual quality and sharp detail preservation, even under complex motion and deformation. These results confirm a clear advantage in reconstruction accuracy, demonstrating that our physically-aware motion regularization and progressive optimization strategy can more effectively capture fine geometric and temporal details in dynamic scenes. In terms of efficiency, our method converges in just 10 minutes, making it the fastest to train among all evaluated models. Additionally, it achieves real-time rendering at 120.00 FPS, substantially outperforming NeRF-based baselines, which remain below 2 FPS. This performance reflects both the scalability and practicality of our framework in real-time scenarios. Experimental results confirm the effectiveness of our PhysMotion4DGS in balancing accuracy, perceptual quality, and speed for high quality dynamic scene reconstruction.

Table 2: **Quantitative Comparison on D-NeRF Dataset**. Our approach surpasses both NeRF-based and Gaussian-based baselines in PSNR.

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
Deformable3DGS Yang et al. (2023)	39.31	0.99	0.01	26 min	85.45
Deformable4DGS Wu et al. (2023)	32.99	0.97	0.05	13 min	104.00
Ours	37.00	0.99	0.01	10 min	120.00

ABLATION STUDIES

4D Physically-Aware Consistency. We first examine the impact of incorporating only the 4D physically-aware consistency regularization (Table 3 (*b*)). Compared to the base model, this configuration yields consistent PSNR improvements across all scenes in the D-NeRF dataset. This module projects the predicted 3D Gaussian motion onto the image plane and enforces alignment with observed optical flow, effectively constraining the model to produce motion patterns that are consistent with 2D visual cues. The observed performance gains validate the effectiveness of image-space motion supervision in guiding early-stage learning and improving the fidelity of appearance changes associated with motion. When 4D physically-aware consistency is removed (Figure 3), we observe spatial misalignment and motion drift, especially in regions involving complex depth interactions or occlusions. This suggests that 2D image-space constraints derived from optical flow are essential for anchoring projected motion and reinforcing alignment between rendered motion and observed trajectories in the input.

Table 3: **Ablation Study with Quantitative Comparison on the D-NeRF Dataset.** We evaluate the base model and the effect of three components, PA (4D physically-aware consistency), PM (4D physics-motion consistency), and T (temporal consistency).

ID	Ablation Items			Overall Metrics (Avg.)			Sequence Results (PSNR ↑)		
	PA	PM	T	PSNR↑	SSIM↑	LPIPS↓	Jumping Jacks	Stand Up	Trex
a				31.97	0.91	0.19	34.85	36.38	32.60
b	✓			33.11	0.94	0.15	35.12	36.71	33.05
c		✓		35.30	0.96	0.10	35.60	37.49	33.80
d			✓	34.12	0.93	0.13	35.35	37.02	33.45
e	✓	✓	✓	**37.00**	**0.99**	**0.01**	**35.92**	**38.21**	**34.32**

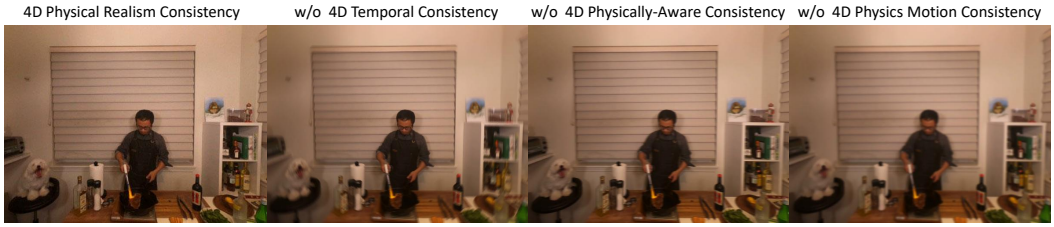


Figure 3: **Visual Comparison under Ablations.** Rendered results from our physically consistent 4D framework and from variants obtained by removing individual components.

4D Physics Motion Consistency. When enabling only the 4D physics motion consistency regularization (Table 3 (*c*)), the model exhibits the most substantial performance improvement among the single-component ablations. The magnitude of the improvement demonstrates that physical motion supervision is a critical factor in resolving ambiguities inherent in monocular dynamic scene reconstruction. Unlike image-space constraints, which operate on projected motion, this component leverages geometry-aware information to regulate the motion field in volumetric space, leading to accurate and coherent dynamics, especially in scenes with complex deformations, occlusions, or nonrigid motion. When 4D physics motion consistency is excluded, we can observe that the most severe degradation occurs (Figure 3). In this case, we observe severe deformation and geometric instability in moving regions, indicating that physically grounded 3D velocity supervision plays a critical role in preserving motion realism.

Temporal Consistency. Activating only the temporal consistency regularization (Table 3 (*d*)) leads to notable improvements over the base model. Unlike the other two regularization components, which focus on spatial alignment, temporal consistency operates along the temporal dimension, acting as a structural prior to suppress abrupt or unrealistic motion fluctuations. The effectiveness of this constraint is evident in scenes exhibiting subtle temporal deformation, where minor motion inconsistencies in separate time steps can accumulate into visually disruptive artifacts. Figure 3 shows that removing the temporal consistency results in perceptible temporal flickering and motion instability, especially around dynamic regions such as the subject's hands and moving objects. This highlights the importance of enforcing smooth velocity transitions over time to maintain temporal coherence in dynamic scenes.

CONCLUSION

We introduce an innovative physics-aware motion rendering approach that enhances physical consistency and temporal coherence in dynamic scene motion prediction. PhysMotion4DGS establishes a deep connection between 3D space and physical space. This not only ensures the visual accuracy of Gaussian distribution dynamics but also guarantees their physical plausibility. Using a progressive multi-scale training strategy, PhysMotion4DGS optimizes both global motion consistency and local detail accuracy, substantially improving the reconstruction of complex motion scenes. Experimental results show that PhysMotion4DGS outperforms existing methods in dynamic scene modeling, offering stable, coherent, and physically consistent reconstructions without increasing point cloud size. Future work will extend PhysMotion4DGS to interactive applications, such as head-tracked VR, targeting stable, temporally coherent rendering of dynamic scenes.

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.

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.

Wei Cheng, Ruixiang Chen, Siming Fan, Wanqi Yin, Keyu Chen, Zhongang Cai, Jingbo Wang, Yang Gao, Zhengming Yu, Zhengyu Lin, et al. Dna-rendering: A diverse neural actor repository for high-fidelity human-centric rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19982–19993, 2023.

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.

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.

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.

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.

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. IEEE, 2022.

Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4220–4230, 2024.

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.

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.

Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime gaussian feature splatting for real-time dynamic view synthesis. *arXiv preprint arXiv:2312.16812*, 2023.

Zhong Li, Minye Wu, Wangyiteng Zhou, and Jingyi Yu. 4d human body correspondences from panoramic depth maps. In *Proceedings of the IEEE 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.

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.

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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 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.

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.

Minye Wu, Yuehao Wang, Qiang Hu, and Jingyi Yu. Multi-view neural human rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1682–1691, 2020.

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.