TaylorMoDe-GS: Taylor-Driven Gaussian Splatting Motion Model for Multi-View Dynamic Scene Deblurring

Xiaofeng Quan[1,3], **Junzhe Wan**[1,3], **Chao Cai**[1,3*]
Yifan Zuo[1,3], **Xiaoshui Huang**[2], **Yuming Fang**[1,3]
[1]Jiangxi University of Finance and Economics
[2]Shanghai Jiao Tong University
[3]Jiangxi Provincial Key Laboratory of Multimedia Intelligent Processing
ncccai@hotmail.com

Abstract

While 3D Gaussian Splatting (3DGS) has excelled in dynamic scene reconstruction, it struggles with multi-view object motion blur, where view-dependent non-uniform blur violates fundamental multi-view geometric constraints. Existing methods fail to balance complex motion fitting with physical consistency across different views. To address the challenge, we propose TaylorMoDe-GS, the first 3DGS framework tailored for multi-view dynamic object deblurring. Specifically, we shift the modeling paradigm from displacement fitting to velocity driven modeling. This is achieved by analytically deriving instantaneous 3D velocities via Taylor series. To map 3D physical motion onto the 2D image plane, we propose a velocity splatting technique. Building upon this, we introduce a neural Peano remainder network to compensate for high frequency nonlinear dynamics, effectively resolving the conflict between physical priors and fitting flexibility. Combined with a learnable physical blur synthesis mechanism, our framework ensures rigorous spatiotemporal and view consistency. Extensive experimental results validate that our method achieves notable advancements in restoring high-fidelity dynamic scenes and physically consistent motion fields, effectively addressing the challenges of multi-view dynamic object deblurring. The code is released at https://github.com/Chiffin-0816/TaylorMoDe-GS.

1 Introduction

Despite the success of 3DGS in dynamic reconstruction [Wu *et al.*, 2024], recovering multi-view dynamic scenes from motion-blurred observations remains a significant bottleneck [Kumar and Rajagopalan, 2025; Hu *et al.*, 2025a]. Unlike static deblurring which assumes global camera motion [Zhao *et al.*, 2024; Chen and Liu, 2024], dynamic blur arises from the independent 3D trajectories of local objects integrated over the exposure time [Lu *et al.*, 2025]. Due to motion parallax, these 3D motions project into anisotropic, view-dependent 2D blur kernels across different camera poses [Li *et al.*, 2022]. This lack of instantaneous consistency violates fundamental multi-view geometric constraints [Matsuki *et al.*, 2024; Keetha *et al.*, 2024], leading models to erroneously compensate for local motion through geometric deformation, which results in severe topological distortions and artifacts [Feng *et al.*, 2025].

Existing blurred scene reconstruction models exhibit significant inadequacies when addressing multi-view dynamic object motion blur [Yin *et al.*, 2025; Hu *et al.*, 2025b]. First, multi-view static deblurring methods generally assume that blur stems from global camera motion such as camera jitter or defocus [Zhao *et al.*, 2024; Lee *et al.*, 2024]. Their modeling logic relies on global camera pose transformations and lacks the perception of independent 3D trajectories for local objects within the scene. When applied to dynamic scenes, these methods fail to handle asymmetric local blur as the global geometric constraints are violated. Second, single-view dynamic deblurring methods [Gao *et al.*, 2021; Zhang *et al.*, 2025; Bui *et al.*, 2025] attempt to fit motion trajectories through temporal deformation fields. However, these are essentially highly under-determined black-box fittings. Without explicit cross space physical mapping, the learned blur kernels fail to converge into a unified motion vector field in 3D space, leading to severe ill-posed deformations during multi-view optimization rather than recovering true physical motion laws.

Recent dynamic reconstruction advancements primarily follow two paradigms: enhancing explicit representations like 4D-GS [Wu *et al.*, 2024] and Deformable 3DGS [Yang *et al.*, 2024b], or employing scene decomposition as seen in Swift4D [Wu *et al.*, 2025] and DeGauss [Wang *et al.*, 2025] to separate dynamic and static components. While these excel in inter-frame modeling, deblurring requires capturing microscopic motion within the exposure. Current frameworks, such as our baseline Learnable Infinite Taylor Gaussian [Hu *et al.*, 2025a], focus on "positional evolution" by using Taylor expansions to approximate Gaussian offsets. Nonetheless, such displacement-fitting tends to neglect the underlying physical constraints of intra-exposure motion.

In this work, we contend that modeling positional evolution alone is insufficient for the deblurring task. Blur [Zhong *et al.*, 2023; Zhong *et al.*, 2022] is the physical result of velocity integrated over time; displacement is merely the cumulative effect of this process. Therefore, we advocate for a paradigm shift from "displacement fitting" to "instantaneous velocity derivative modeling". As the first-order derivative of displacement,

*Corresponding author.

velocity more fundamentally carries the laws of dynamics and provides a physically meaningful integrand for blur kernel synthesis [Tran *et al.*, 2021]. Through this paradigm shift, we naturally incorporate differential priors and ensure that multi-view observations strictly converge to a unified 3D motion vector field [Lin *et al.*, 2024].

Based on the insights, we propose TaylorMoDe-GS, a framework that deeply couples physical analytical modeling with neural flexibility [Xie *et al.*, 2024; Li *et al.*, 2025]. First, we derive the instantaneous velocity by analytically differentiating the Taylor motion model, ensuring strict alignment with geometric positions. To bridge the representation gap between 3D physical motion and 2D pixel observations, we introduce velocity splatting. This technique utilizes a customized differentiable rasterization module to explicitly project and accumulate 3D velocities into a continuous 2D world velocity map. To address truncation errors in finite-order Taylor expansions, we introduce the Peano remainder network. Taking the splatted physical velocity map as a strong prior, this network employs a residual learning strategy to capture high frequency, non-linear velocity components. This hybrid "physical principle + neural residual" formulation overcomes the accuracy bottlenecks of traditional Taylor modeling in complex motions while avoiding the rigidity of partial differential equation-based constraints [Chen *et al.*, 2023].

Finally, we construct a learnable physical blur synthesis mechanism to close the optimization loop. By establishing a rigorous 2D-3D-2D mapping, back projecting pixels using hybrid depth priors, evolving trajectories via the learned velocity field, and re-projecting into image space, we explicitly synthesize physical blur kernels. This process naturally preserves motion parallax, ensuring that multi-view blurry observations provide strong constraints for a unified 3D motion field. Furthermore, we implement a gradient decoupling strategy to isolate static background gradients, ensuring the model fits observations by refining motion parameters rather than distorting static geometry. Our contributions are summarized as follows:

- We propose TaylorMoDe-GS, the first 3DGS framework tailored for multi-view dynamic object deblurring, enabling high-fidelity reconstruction from blurry video inputs.
- We propose Taylor-Driven velocity model, utilize taylor series to explicitly model 3D velocity and introduce velocity splatting to map spatial dynamics to the 2D plane, coupled with peano remainder compensation to maintain multi-view geometric consistency.
- We develop a learnable physical blur synthesis mechanism, accurately decouples motion blur through explicit physical projection, ensuring the structural stability of the reconstructed 3D scene.

2 Related Work

Dynamic Scene Reconstruction. In recent years, 3D Gaussian Splatting (3DGS) [Kerbl *et al.*, 2023] has emerged as a dominant paradigm for dynamic scene reconstruction due to its explicit representation and exceptional rendering efficiency. 4DGS [Wu *et al.*, 2024] decouples spatiotemporal features through 4D neural voxels, while Deformable 3DGS [Yang *et al.*, 2024b] proposes learning deformation fields in canonical space to predict positional offsets. Furthermore, SC-GS [Huang *et al.*, 2024] and Spacetime Gaussians [Li *et al.*, 2024] respectively introduce structure-consistent constraints and 4D temporal-spatial representations to improve dynamic modeling. To enhance reconstruction quality in complex environments, Swift4D [Wu *et al.*, 2025] and DeGauss [Wang *et al.*, 2025] respectively introduce efficient dynamic-static decomposition strategies, utilizing masking mechanisms to filter out dynamic distractors. Furthermore, Infinite Taylor Gaussian [Hu *et al.*, 2025a] employs Taylor series to capture the continuous evolution of scene attributes. Despite the superior performance of these methods in sharp image reconstruction, they generally assume ideal sharp observations and lack physical mechanisms to account for temporal integration effects. When addressing multi-view motion blur [Xu *et al.*, 2025], existing methods struggle to balance complex motion fitting with physical consistency, often erroneously compensating for local motion by distorting geometry, which leads to severe topological distortions.

Deblurring-based Scene Reconstruction. Existing deblurring approaches exhibit significant inadequacies when addressing multi-view dynamic scenarios. On one hand, static methods like BAD-Gaussians [Zhao *et al.*, 2024] and Deblur-GS [Chen and Liu, 2024] rely on a global camera motion assumption [Nah *et al.*, 2017], failing to perceive independent 3D trajectories of local objects. This leads to failure when handling asymmetric local blur where global geometric constraints are violated. On the other hand, monocular dynamic methods such as BARD-GS [Lu *et al.*, 2025] attempt to fit motion via deformation fields, but suffer from the inherent under-determination of monocular views and the "spectral bias" of Multi-Layer Perceptron. Without explicit 3D-2D physical mapping, the learned blur kernels often fail to converge to a unified 3D motion field, causing severe ill-posed deformations during multi-view optimization. In contrast, our TaylorMoDe-GS bridges this gap through velocity-driven modeling.

Physics-Driven Modeling. To mitigate the limitations of heuristic deformation, recent research has begun injecting physical laws into Gaussian dynamics. For instance, Phys-Gaussian [Xie *et al.*, 2024] and FreeGave [Li *et al.*, 2025] enhance 4D reconstruction realism by introducing velocity fields and newtonian constraints. Building upon these physics aware representations, our work shifts the paradigm from simple displacement fitting to instantaneous velocity derivative modeling. By utilizing explicit physical projection ,such as velocity splatting, we bridge the gap between 3D motion dynamics and 2D blurry observations, ensuring that the recovered motion fields and scene geometry remain strictly consistent across multiple views.

3 Methodology

We propose TaylorMoDe-GS, a physics-aware dynamic deblurring framework based on 3D Gaussian Splatting, specifically designed to recover sharp, high-fidelity scenes from multi-view videos with motion blur. Distinct from traditional implicit deformation fields or linear motion assumptions, we

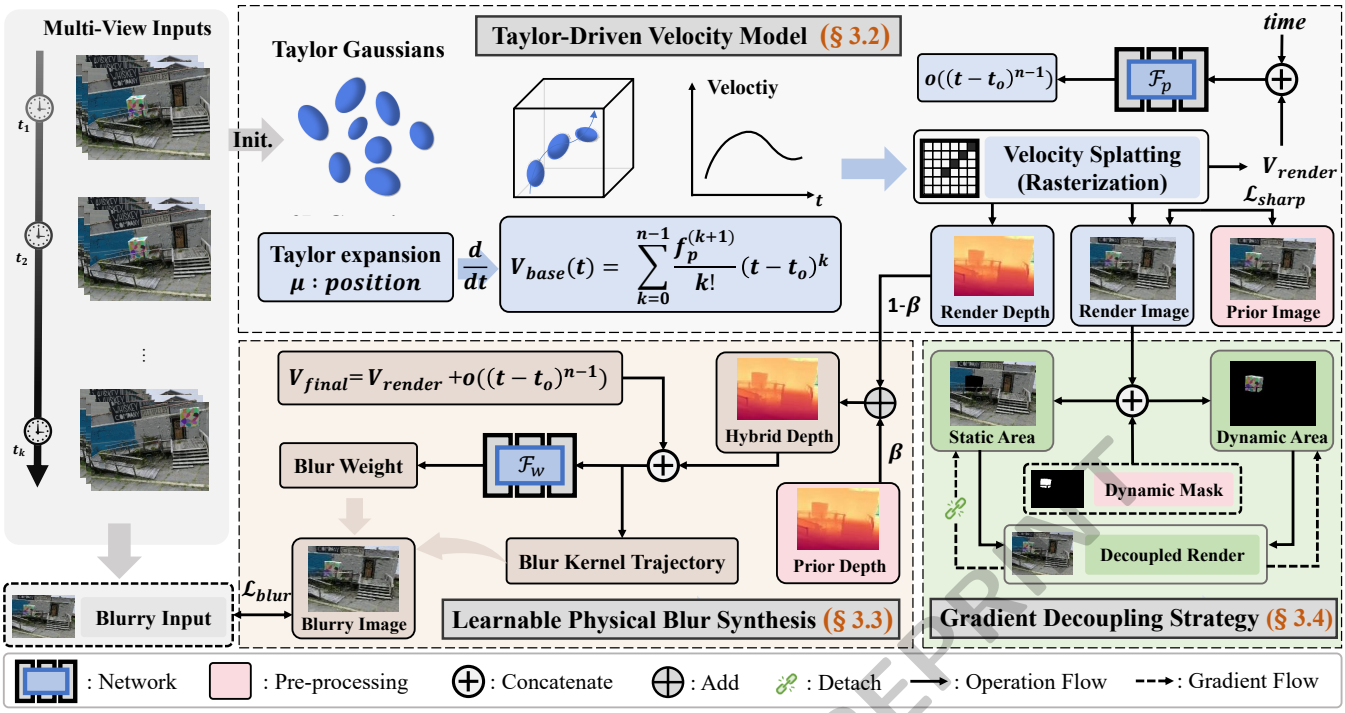

Figure 1: Overview of our TaylorMoDe-GS framework. Top: We model continuous 3D motion via Taylor expansion and introduce a Peano Remainder Network ($\mathcal{F}_p$) to compensate for high-order non-linear residuals, thereby obtaining the precise final velocity field V_{final}. Bottom-Left: In the physical blur synthesis module, we construct Hybrid Geometric Priors by fusing monocular priors with rendered depth to generate precise blur kernel trajectories, integrated with temporal weights predicted by $\mathcal{F}_w$. Bottom-Right: The Gradient Decoupling Strategy uses a mask to split rendering into static and dynamic branches. By blocking gradients in the static region, we prevent the motion blur synthesis process from contaminating the static background color.

employ Taylor expansions to explicitly model the physical blurring process. This approach bridges the gap between the continuous motion of Gaussian primitives and the resulting motion blur during exposure. In Sec. 3.1, we introduce the foundational theory. Subsequently, Sec. 3.2 details our motion model based on Taylor derivatives and the Peano remainder. Sec. 3.3 describes the learnable physical blur synthesis mechanism that simulates the exposure integration process. Finally, in Sec. 3.4, we present our optimization strategies and loss functions, incorporating a gradient decoupling strategy and hybrid priors to solve the ambiguities between geometry and motion.

3.1 Preliminaries

3D Gaussian Splatting (3DGS). 3DGS [Kerbl *et al.*, 2023] parameterizes a scene as a collection of 3D Gaussians $\mathcal{G} = \{G_1, \ldots, G_N\}$. Each primitive is parameterized by its center $\boldsymbol{\mu}_i \in \mathbb{R}^3$, covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3\times3}$, opacity $\alpha_i \in [0, 1]$, and spherical harmonics (SH) coefficients $\boldsymbol{c}_i$. To render an image, 3D Gaussians are projected onto the 2D image plane. The pixel color $\boldsymbol{C}$ is then synthesized via differentiable α-blending of N sorted primitives:

$$\boldsymbol{C} = \sum_{i\in\mathcal{N}} \boldsymbol{c}_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \tag{1}$$

where $\mathcal{N}$ denotes the set of Gaussians overlapping the target pixel.

Dynamic 3D Gaussians. We extend the static representation to the temporal domain by introducing a time variable t. A dynamic Gaussian primitive $\mathcal{G}(t)$ is formulated as a time-dependent tuple: $\mathcal{G}(t) = \{\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t), \boldsymbol{c}(t), \alpha(t)\}$. In this work, we focus on explicitly parameterizing the position trajectory $\boldsymbol{\mu}(t)$ to model the continuous motion required for blur synthesis during the exposure time τ.

Infinite Taylor Gaussian Model. Inspired by the Infinite Taylor Gaussian (ITG) framework [Hu *et al.*, 2025a], we parameterize the trajectory of the Gaussian position $\boldsymbol{\mu}(t)$ via a Taylor series expanded at a reference time t_0, rather than learning discrete parameters per timestamp. The trajectory is defined as:

$$\boldsymbol{\mu}(t) = \sum_{k=0}^{n} \frac{\boldsymbol{f}_p^{(k)}}{k!}(t-t_0)^k + R_n(t), \tag{2}$$

where $\boldsymbol{f}_p^{(k)} \in \mathbb{R}^3$ denotes the learnable k-th order Taylor coefficient, which represents the k-th derivative of the position function. $R_n(t)$ represents the neural Peano remainder, designed to compensate for nonlinearities that low-order polynomials fail to capture. This formulation enables the analytical velocity derivation in Sec. 3.2.

3.2 Taylor-Driven Velocity Model

Taylor Series Velocity Model. Departing from traditional methods that learn discrete displacement vectors, we model the short-term trajectory of each Gaussian as a continuous smooth function. By analytically differentiating the Taylor expansion defined in Eq. (2), we obtain the instantaneous base velocity $V_{base}(t)$ consistent with the positional change:

$$V_{base}(t) = \frac{d}{dt}\boldsymbol{\mu}(t) = \sum_{k=0}^{n-1} \frac{\boldsymbol{f}_p^{(k+1)}}{k!}(t-t_0)^k. \tag{3}$$

This explicit derivation ensures that the velocity field aligns strictly with the physical motion trajectory.

Velocity Splatting. To achieve precise blur synthesis based on 3D motion, it is necessary to accurately accumulate the motion of visible surfaces onto the image plane. Inspired by the accumulation mechanism in depth rendering, we propose a velocity splatting strategy based on the physical rendering logic of 3DGS. We develop a customized differentiable rasterization module to explicitly project and accumulate the 3D base velocity V_{base} onto the image plane. Utilizing the same α-blending mechanism used for color to account for perceptual visibility, we compute a dense rendered world velocity map V_{render}:

$$V_{render}(\boldsymbol{u}) = \sum_{i\in\mathcal{N}} T_i \alpha_i V_{base,i} \tag{4}$$

where $\boldsymbol{u}$ denotes pixel coordinates, α_i is opacity, and $T_i = \prod_{j=1}^{i-1}(1-\alpha_j)$ represents transmittance. This operation transforms the discrete, unstructured 3D Gaussian representation into a continuous 2D field V_{render}, representing the pixel-wise physical velocity.

Peano Remainder Compensation Network. Since the base velocity V_{base} derived from low order Taylor expansions suffers from truncation errors, it often misses complex nonlinear dynamics. To recover these residuals (mathematically corresponds to the Peano remainder), we introduce a Peano remainder network. Using the rendered velocity V_{render} as a physical prior, we feed it along with a time embedding $\gamma(t)$ into a lightweight MLP to predict the velocity residual R_{peano}:

$$R_{peano} = \mathcal{F}_p(V_{render}, \gamma(t)) \tag{5}$$

The final velocity is given by $V_{final} = V_{render} + R_{peano}$. By modeling only the residual, the network compensates for nonlinearities without needing to learn the entire motion field.

3.3 Learnable Physical Blur Synthesis Mechanism

To simulate the physical exposure process, we synthesize motion blur by integrating radiance over the exposure time τ. This synthesis relies on three key components: incorporating hybrid geometric priors, deriving the blur kernel trajectory, and predicting adaptive weights via a blur weight network.

Hybrid Geometric Priors. Precise 3D localization is a prerequisite for accurate physical blur synthesis. However, relying solely on the rendered depth D_{render} is unstable, particularly in early training. Depth errors result in incorrect 3D back-projection, distorting the blur kernel trajectory. To mitigate this geometric instability, we utilize hybrid geometric priors [Chung *et al.*, 2024]. We construct a stable depth map D_{in} by blending the rendered depth with a prior D_{prior} derived from a pre-trained monocular estimator:

$$D_{in}(\boldsymbol{u}) = \beta \cdot D_{prior}(\boldsymbol{u}) + (1-\beta)\cdot D_{render}(\boldsymbol{u}) \tag{6}$$

where β is a balancing coefficient. This hybrid depth D_{in} provides a robust geometric basis for trajectory estimation, ensuring stability despite noise in the reconstructed geometry [Guédon and Lepetit, 2024].

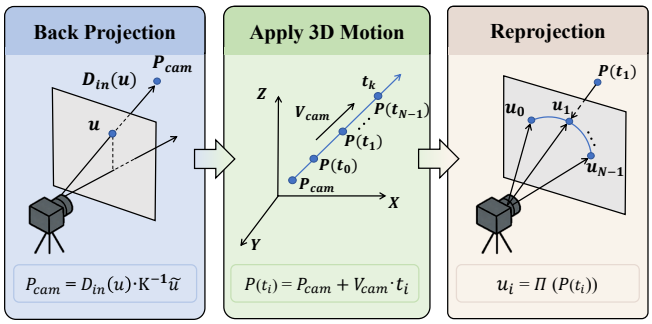

Figure 2: Physical formation of the blur kernel trajectory. The discrete kernel locations $\{\boldsymbol{u}_i\}$ are determined by re-projecting a 3D motion path, estimated from depth prior D_{in} and velocity $\boldsymbol{v}_{cam}$, back onto the image plane.

Blur Kernel Trajectory. We utilize the Track-Anything Model (TAM) to generate a binary mask M_{dyn} isolating dynamic objects. The dynamic pixel set is defined as $\Omega_{dyn} = \{\boldsymbol{u} \mid M_{dyn}(\boldsymbol{u}) = 1\}$. Each dynamic pixel $\boldsymbol{u} \in \Omega_{dyn}$ corresponds to a 3D point traversing a motion trajectory during the exposure time τ. We treat τ as a learnable variable rather than a fixed hyperparameter, allowing the integration window to adapt to scene dynamics.

To simulate shutter integration, we discretize the exposure interval $[-\tau/2, \tau/2]$ into N uniformly distributed timestamps. Let $i \in \{0, \ldots, N-1\}$ be the time index, the time offset t_i is defined as:

$$t_i = \left(\frac{i}{N-1} - \frac{1}{2}\right)\cdot \tau. \tag{7}$$

As shown in Figure 2, we back-project pixel $\boldsymbol{u}$ into camera space using the hybrid depth D_{in} to obtain the initial position P_{cam}:

$$P_{cam} = D_{in}(\boldsymbol{u})\cdot \mathrm{K}^{-1}\tilde{\boldsymbol{u}}, \tag{8}$$

where K is the camera intrinsic matrix, and $\tilde{\boldsymbol{u}}$ denotes homogeneous coordinates. The velocity V_{final} is transformed into the camera coordinate system via the rotation matrix R_{view}:

$$\boldsymbol{v}_{cam} = \mathrm{R}_{view} V_{final}. \tag{9}$$

For each timestamp t_i, projecting the instantaneous 3D position onto the image plane yields the sample coordinate $\boldsymbol{u}_i$:

$$\boldsymbol{u}_i = \Pi(P_{cam} + \boldsymbol{v}_{cam}\cdot t_i), \tag{10}$$

where $\Pi(\cdot)$ denotes the perspective projection function. Deriving the 2D trajectory from 3D geometry allows this formulation to preserve motion parallax.

Blur Weight Network. While the trajectory defines the sampling path, the contribution of each sample to the final pixel color varies. This variation arises from two factors: (1) Motion Speed: Slower motion increases pixel dwell time (higher weight), whereas faster motion reduces intensity (lower weight). (2) Perspective Scale: Depth modulates projected motion, as distant objects show smaller apparent motion than nearby objects, concentrating the blur intensity.

To capture these dynamics, we employ a lightweight MLP that maps the velocity in camera space $\boldsymbol{v}_{cam}(\boldsymbol{u})$ and hybrid depth $D_{in}(\boldsymbol{u})$ to a temporal weight vector $\boldsymbol{w}(\boldsymbol{u}) \in \mathbb{R}^N$:

$$\boldsymbol{w}(\boldsymbol{u}) = \mathcal{F}_w(D_{in}(\boldsymbol{u}), \boldsymbol{v}_{cam}(\boldsymbol{u})), \quad \text{s.t.} \sum_{i=0}^{N-1} w_i = 1, \quad (11)$$

where $\mathcal{F}_w$ denotes the weight prediction network, and w_i represents the i-th element of $\boldsymbol{w}$. The blurred color $\hat{C}(\boldsymbol{u})$ is then synthesized by accumulating colors along the trajectory:

$$\hat{C}(\boldsymbol{u}) = \sum_{i=0}^{N-1} w_i(\boldsymbol{u}) \cdot C(\boldsymbol{u}_i), \quad (12)$$

where C denotes the rendered sharp image.

3.4 Optimization and Loss Functions

Our optimization strategy aims to decouple the ambiguity between geometry and motion by imposing separate constraints on the static background and dynamic foreground.

Decomposed Hybrid Supervision. We enforce consistency across both sharp and blurry domains to jointly optimize static structure and motion dynamics. To restore sharp textures, we construct a hybrid target I_{target} by fusing the input background with the sharp foreground derived from a pre-trained deblurring model:

$$I_{target} = M_{dyn} \odot I_{pseudo} + (1 - M_{dyn}) \odot I_{input}. \quad (13)$$

We constrain the rendered sharp image C to match this target using a combination of L1 and D-SSIM losses:

$$\begin{aligned} \mathcal{L}_{sharp} = (1 - \lambda_{ssim})\|C - I_{target}\|_1 \\ + \lambda_{ssim}(1 - \text{SSIM}(C, I_{target})). \end{aligned} \quad (14)$$

For motion blur synthesis, we minimize the reconstruction error constrained to the dynamic region:

$$\mathcal{L}_{blur} = \|M_{dyn} \odot (\hat{C} - I_{input})\|_1. \quad (15)$$

To prevent the optimizer from distorting static geometry to fit this loss, we implement a **gradient decoupling strategy**. We explicitly detach the static background component during blur synthesis, ensuring that the model fits the observation by refining motion parameters rather than distorting the static background.

Physical Regularization. We impose three constraints to maintain physical consistency. First, to prevent geometric collapse, we align the rendered depth D_{render} with the prior D_{prior} by minimizing the L1 difference in the logarithmic domain:

$$\mathcal{L}_{depth} = \|\log(D_{render} + \epsilon) - \log(D_{prior} + \epsilon)\|_1, \quad (16)$$

where ϵ is a small constant for numerical stability. Second, for rigidly moving objects, we constrain the velocity field $\boldsymbol{v}_{cam}$ using the object center displacement $\boldsymbol{v}_{guide}$:

$$\mathcal{L}_{vel} = \|M_{dyn} \odot (\boldsymbol{v}_{cam} - \boldsymbol{v}_{guide})\|_1. \quad (17)$$

Finally, to enforce temporal continuity, we regularize the weight network by penalizing abrupt changes between adjacent timestamps:

$$\mathcal{L}_{smooth} = \|M_{dyn} \odot (w_i - w_{i-1})\|_2^2. \quad (18)$$

Total Objective. The final objective function is a weighted sum of the reconstruction and regularization terms:

$$\begin{aligned} \mathcal{L}_{total} = \mathcal{L}_{sharp} + \lambda_{blur}\mathcal{L}_{blur} \\ + \lambda_{depth}\mathcal{L}_{depth} + \lambda_{vel}\mathcal{L}_{vel} + \lambda_{smooth}\mathcal{L}_{smooth}. \end{aligned} \quad (19)$$

4 Experiments

This section details the experimental evaluation of TaylorMoDe-GS. We compare our method with existing approaches using both quantitative metrics and qualitative visualizations. Additionally, we conduct ablation studies to verify the effectiveness of our core components.

4.1 Experimental Settings

Training Details. Our framework is implemented in PyTorch, with all experiments conducted on a single NVIDIA RTX A5000 GPU (24GB). Before training, we preprocess the data to acquire necessary priors: (1) We use Depth-Anything V2 [Yang *et al.*, 2024a] to extract monocular depth maps for geometric guidance; (2) We use NAFNet [Chen *et al.*, 2022] to generate pseudo-sharp images for hybrid supervision; (3) We use Track-Anything Model (TAM) [Yang *et al.*, 2023] to segment dynamic object masks M_{dyn}. To ensure geometric consistency, we follow the standard protocol using COLMAP [Schönberger and Frahm, 2016; Schönberger *et al.*, 2016] to reconstruct sparse point clouds and camera poses.

For the introduced learnable modules, including the Peano remainder network ($\mathcal{F}_p$) and the blur weight network ($\mathcal{F}_w$), we use separate Adam optimizers with a learning rate of 1×10^{-4} and momentum parameters $(0.9, 0.999)$. The learnable exposure time τ is also jointly optimized. For hyperparameters, we discretize the exposure integration into $N = 7$ timestamps and set the weight for the monocular depth prior to 0.3. The loss weights are set to: $\lambda_{blur} = 0.5$, $\lambda_{smooth} = 0.01$, $\lambda_{depth} = 0.02$, and $\lambda_{vel} = 5 \times 10^{-5}$.

Datasets. We evaluate our method on the multi-view dynamic blur dataset provided by DynaMoDe-NeRF. This dataset contains 4 synthetic scenes (Cube, Lychee, Monkey, Sphere) rendered by Blender at 1024×768 resolution. Motion blur is simulated by temporally integrating high-frame-rate sharp videos. To ensure a unified experimental protocol across different baselines, we utilize a subset of 50 frames for training and evaluation. For each synchronized time step, we hold out one view as the test view and use the remaining views for training.

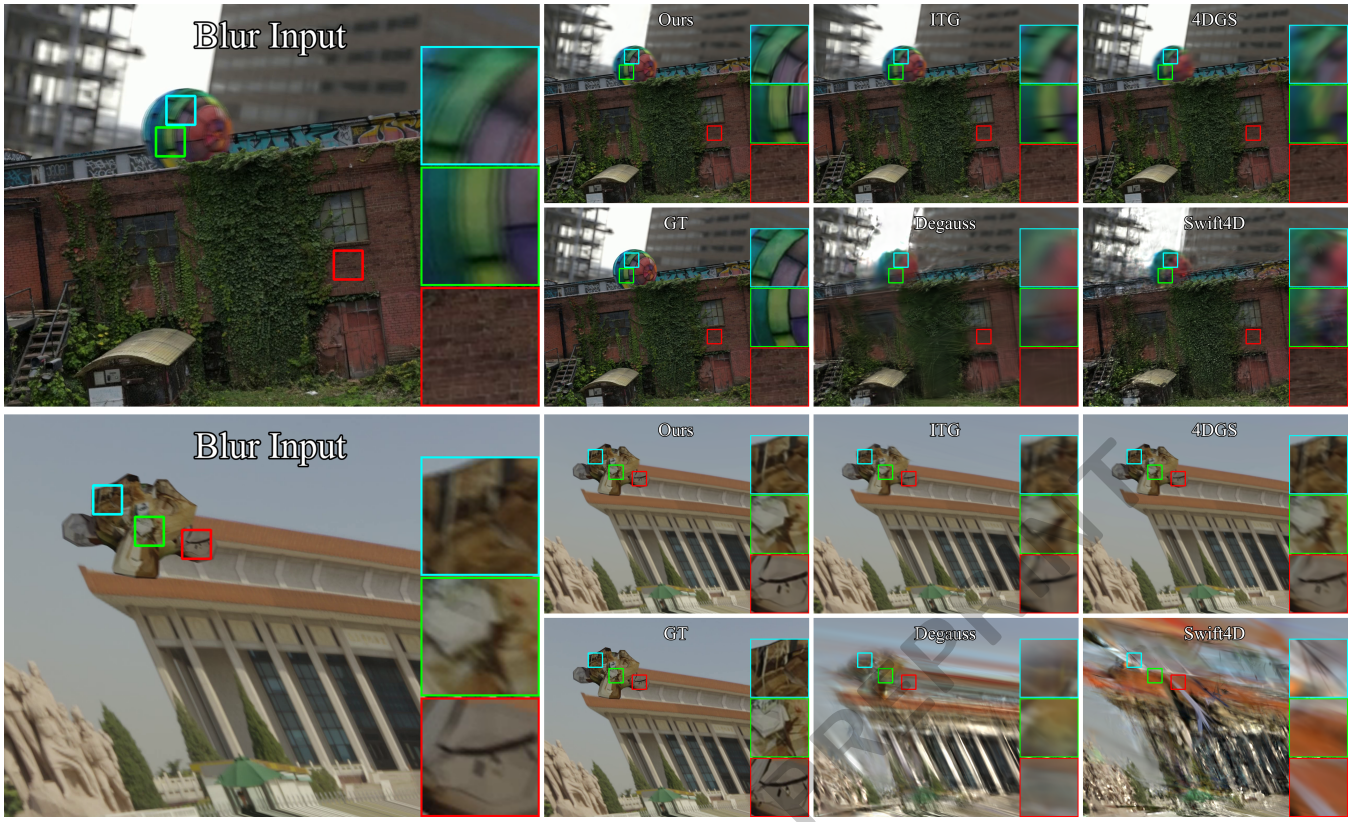

Figure 3: Qualitative comparisons on Monkey and Sphere. Swift4D and DeGauss fail with severe geometric distortion, while 4DGS and ITG suffer from residual ghosting on dynamic objects. In contrast, our method recovers sharp fine-grained details, such as facial geometry and surface textures.

Baselines. Since TaylorMoDe-GS pioneers 3DGS-based multi-view deblurring, we compare it against state-of-the-art dynamic reconstruction methods, including 4DGS and Infinite Taylor Gaussian (ITG). Additionally, we include Swift4D and DeGauss to evaluate performance against recent static-dynamic decomposition approaches.

Metrics. We use PSNR [Keleş *et al.*, 2021], SSIM [Wang *et al.*, 2004], and LPIPS [Zhang *et al.*, 2018] with AlexNet features to evaluate reconstruction and deblurring performance. Since motion blur primarily affects dynamic regions, we also report metrics computed within the dynamic mask Ω_{dyn} to assess the recovery of moving objects.

Figure 4: Visual ablation comparison of the physical motion model components. We compare the full model against settings without Taylor expansion and without the Peano remainder.

4.2 Results and Comparisons

Quantitative Evaluation. Table 1 presents the quantitative results across four scenes. Swift4D and DeGauss fail to reconstruct blurred inputs, yielding notably lower scores across all metrics. While 4DGS and ITG achieve reasonable quality by fitting temporal deformations, they remain limited by motion blur ambiguity. In contrast, TaylorMoDe-GS outperforms all baselines. In terms of average performance, our method improves Dynamic PSNR by 1.9 dB compared to the ITG baseline. Furthermore, our method achieves the lowest LPIPS scores across all scenes and regions. This reduction in perceptual error confirms that our physical modeling restores superior perceptual quality globally, especially in dynamic regions.

Qualitative Evaluation. Figure 3 compares the reconstruction results on the *Monkey* and *Sphere* scenes. Swift4D and DeGauss fail to model rapid motion, causing severe distortion across both background and foreground regions. While 4DGS and ITG maintain background stability, they produce visible residual blur and ghosting artifacts on moving objects. In contrast, TaylorMoDe-GS recovers sharp geometric details. As shown in the zoomed-in regions, our method restores fine structures, such as the facial geometry of the monkey and the surface texture of the sphere, which are lost in the baselines.

Method	Monkey						Sphere					
	Overall			Dynamic			Overall			Dynamic		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Swift4D[Wu *et al.*, 2025]	14.81	0.541	0.495	15.40	0.624	0.433	20.09	0.424	0.256	16.60	0.513	0.399
DeGauss[Wang *et al.*, 2025]	14.93	0.614	0.362	18.73	0.644	0.388	19.56	0.479	0.363	16.03	0.569	0.420
4DGS[Wu *et al.*, 2024]	39.40	0.977	0.028	30.15	0.868	0.210	28.75	0.841	0.135	21.48	0.653	0.285
ITG[Hu *et al.*, 2025a]	38.43	0.974	0.031	29.73	0.858	0.221	29.13	0.925	0.080	19.92	0.620	0.233
TaylorMoDe-GS (Ours)	**40.88**	**0.982**	**0.016**	**33.79**	**0.928**	**0.088**	**31.76**	**0.940**	**0.052**	**22.57**	**0.666**	**0.112**

Method	Cube						Lychee					
	Overall			Dynamic			Overall			Dynamic		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Swift4D[Wu *et al.*, 2025]	14.23	0.310	0.516	14.62	0.571	0.405	22.54	0.721	0.249	14.99	0.599	0.439
DeGauss[Wang *et al.*, 2025]	14.22	0.352	0.531	16.31	0.589	0.476	21.22	0.715	0.311	14.13	0.575	0.414
4DGS[Wu *et al.*, 2024]	30.75	0.888	0.127	**32.57**	**0.932**	0.083	**35.15**	**0.951**	0.059	26.22	0.824	0.247
ITG[Hu *et al.*, 2025a]	31.66	0.927	0.044	30.88	0.905	0.088	35.08	0.947	0.057	26.51	0.841	0.225
TaylorMoDe-GS (Ours)	**32.63**	**0.943**	**0.032**	31.65	0.920	**0.027**	35.08	0.948	**0.050**	**26.63**	**0.848**	**0.159**

Table 1: Quantitative comparison on the DynaMode-NeRF dataset. The **best** and second best results are bold and underlined, respectively.

Method	Overall			Dynamic		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
w/o Taylor Expansion	34.25	0.939	0.058	28.45	0.823	0.131
w/o Peano Remainder	34.14	0.948	0.047	27.85	0.838	0.185
w/o Gradient Decoupling	32.84	0.937	0.057	28.25	0.840	0.100
w/o Re-blurring Loss	31.97	0.916	0.088	26.97	0.828	0.153
w/o Hybrid Pseudo-GT	34.07	0.947	0.049	27.80	0.838	0.183
w/o Blur Weight Network	33.20	0.943	0.053	27.52	0.824	0.152
Ours Full	**35.09**	**0.953**	**0.038**	**28.66**	**0.841**	**0.096**

Table 2: Ablation study on core modeling components evaluated on the DynaMode-NeRF dataset. The **best** and second best results are bold and underlined, respectively.

Method	Overall			Dynamic		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
w/o Smoothness Reg.	33.59	0.946	0.047	28.43	0.838	0.102
w/o Velocity Reg.	32.88	0.939	0.057	28.28	0.833	0.143
w/o All Regs.	33.78	0.946	0.045	27.58	0.824	0.115
Ours Full	**35.09**	**0.953**	**0.038**	**28.66**	**0.841**	**0.096**

Table 3: Ablation study on regularization terms evaluated on the DynaMode-NeRF dataset. The **best** and second best results are bold and underlined, respectively.

4.3 Ablation Studies

We conduct ablation studies to evaluate the contribution of individual modules. The analysis is organized into two parts: core physical designs and auxiliary regularization terms.

Effectiveness of Physical Motion Model. We evaluate the physical motion model consisting of the Taylor expansion and the Peano remainder. Fig. 4 illustrates the necessity of these components. w/o Taylor Expansion replaces physics-based derivative modeling with direct velocity prediction using an MLP, which results in physically inconsistent motion estimates and introduces geometric distortions and rendering artifacts. w/o Peano Remainder limits the model to Taylor terms, which cannot capture high-order residuals or truncation errors, leaving residual blur and ghosting. Quantitative results in Table 2 indicate that black-box prediction and simplified Taylor modeling do not fully recover complex dynamic trajectories.

Analysis of Other Core Designs. Table 2 evaluates other core designs. w/o Gradient Decoupling allows dynamic error gradients to affect static parameter optimization, which causes background geometry to distort to fit dynamic residuals. w/o Re-blurring Loss removes the explicit physical constraint on the blur formation process and decreases reconstruction consistency. w/o Hybrid Pseudo-GT eliminates sharp image guidance, preventing the model from converging to a sharp state in dynamic regions. w/o adaptive weights replaces the learnable weighting network with uniform weights, ignoring photometric contribution variations from parallax and motion divergence and leading to integration bias.

Analysis of Regularization Terms. Table 3 assesses the impact of auxiliary regularization. Specifically, the rigid velocity constraint narrows the solution space, aligning 3D motion fields with blur trajectories. Temporal smoothness enforces trajectory continuity across frames. The performance drop observed upon removing these terms underscores their necessity for stabilizing the joint optimization process.

5 Conclusions

This paper presents TaylorMoDe-GS, the first 3DGS-based framework for multi-view dynamic object deblurring. By transitioning from displacement fitting to instantaneous velocity-derivative modeling, we overcome the physical limitations of existing models in handling intra-exposure integration effects. Our Taylor derived velocity field, coupled with neural peano remainder compensation, successfully balances rigid physical priors with high-frequency non-linear dynamics. Furthermore, the proposed velocity splatting and 2D-3D-2D physical mapping ensure rigorous cross-view geometric and motion consistency. Our velocity-driven method outperforms displacement-based approaches by analytically modeling physical motion from multiple blurred viewpoints. This multi-view motion representation effectively resolves the inherent ambiguity in motion-blurred observations, allowing reconstruction of sharp dynamic scenes and consistent 3D motion fields from blurry multi-view inputs.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62132006, U24A20220, and 62271237, the Natural Science Foundation of Jiangxi Province under Grants 20252BAC240227, 20252BAC230003, and 20242BAB26014, and the Postgraduate Innovation Special Fund Program of Jiangxi Province under Grant YC2024-B179.

References

[Bui *et al.*, 2025] Minh-Quan Viet Bui, Jongmin Park, Jihyong Oh, and Munchurl Kim. MoBlurF: Motion deblurring neural radiance fields for blurry monocular video. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2025.

[Chen and Liu, 2024] Wenbo Chen and Ligang Liu. DeblurGS: 3D Gaussian splatting from camera motion blurred images. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–15, 2024.

[Chen *et al.*, 2022] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 17–33. Springer, 2022.

[Chen *et al.*, 2023] Zhang Chen, Zhong Li, Liangchen Song, Lele Chen, Jingyi Yu, Junsong Yuan, and Yi Xu. NeuRBF: A neural fields representation with adaptive radial basis functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4182–4194, 2023.

[Chung *et al.*, 2024] Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-regularized optimization for 3D Gaussian splatting in few-shot images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 811–820, 2024.

[Feng *et al.*, 2025] Shengjie Feng, Xiaoqun Wu, and Jian Cao. A survey of multi-view stereo 3D reconstruction algorithms based on deep learning. *Digital Signal Processing*, 165:105291, 2025.

[Gao *et al.*, 2021] 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 (ICCV)*, pages 5712–5721, 2021.

[Guédon and Lepetit, 2024] Antoine Guédon and Vincent Lepetit. SuGaR: Surface-aligned Gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5354–5363, 2024.

[Hu *et al.*, 2025a] Bingbing Hu, Yanyan Li, Rui Xie, Bo Xu, Haoye Dong, Junfeng Yao, and Gim Hee Lee. Learnable infinite Taylor Gaussian for dynamic view rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26844–26854, 2025.

[Hu *et al.*, 2025b] Yuxi Hu, Jun Zhang, Zhe Zhang, Rafael Weilharter, Yuchen Rao, Kuangyi Chen, Runze Yuan, and Friedrich Fraundorfer. ICG-MVSNet: Learning intra-view and cross-view relationships for guidance in multi-view stereo. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2025.

[Huang *et al.*, 2024] 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 (CVPR)*, pages 4220–4230, 2024.

[Keetha *et al.*, 2024] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. SplaTAM: Splat track & map 3D Gaussians for dense RGB-D SLAM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21357–21366, 2024.

[Keleş *et al.*, 2021] Onur Keleş, M. Akın Yılmaz, A. Murat Tekalp, Cansu Korkmaz, and Zafer Doğan. On the computation of PSNR for a set of images or video. In *2021 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2021.

[Kerbl *et al.*, 2023] 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):139:1–139:14, 2023.

[Kumar and Rajagopalan, 2025] Ashish Kumar and Ambasamudram Narayanan Rajagopalan. Dynamode-nerf: Motion-aware deblurring neural radiance field for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21728–21738, June 2025.

[Lee *et al.*, 2024] Byeonghyeon Lee, Howoong Lee, Xiangyu Sun, Usman Ali, and Eunbyung Park. Deblurring 3D Gaussian Splatting. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 127–143. Springer, 2024.

[Li *et al.*, 2022] Tianye Li, Mira Slavcheva, Michael Zollhöfer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, Richard Newcombe, and Zhaoyang Lv. Neural 3D video synthesis from multi-view video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5521–5531, June 2022.

[Li *et al.*, 2024] Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime Gaussian feature splatting for real-time dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8508–8520, 2024.

[Li *et al.*, 2025] Jinxi Li, Ziyang Song, Siyuan Zhou, and Bo Yang. FreeGave: 3D physics learning from dynamic videos by Gaussian velocity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12433–12443, 2025.

[Lin *et al.*, 2024] Youtian Lin, Zuozhuo Dai, Siyu Zhu, and Yao Yao. Gaussian-Flow: 4D reconstruction with dynamic 3D Gaussian particle. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21136–21145, 2024.

[Lu *et al.*, 2025] Yiren Lu, Yunlai Zhou, Disheng Liu, Tuo Liang, and Yu Yin. BARD-GS: Blur-aware reconstruction of dynamic scenes via Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16532–16542, 2025.

[Matsuki *et al.*, 2024] Hidenobu Matsuki, Riku Murai, Paul H. J. Kelly, and Andrew J. Davison. Gaussian splatting SLAM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18039–18048, 2024.

[Nah *et al.*, 2017] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 257–265, 2017.

[Schönberger and Frahm, 2016] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-Motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016.

[Schönberger *et al.*, 2016] Johannes L. Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for unstructured Multi-View Stereo. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 501–518. Springer, 2016.

[Tran *et al.*, 2021] Phong Tran, Anh Tuan Tran, Quynh Phung, and Minh Hoai. Explore image deblurring via encoded blur kernel space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11956–11965, 2021.

[Wang *et al.*, 2004] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[Wang *et al.*, 2025] Rui Wang, Quentin Lohmeyer, Mirko Meboldt, and Siyu Tang. DeGauss: Dynamic-static decomposition with Gaussian splatting for distractor-free 3D reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6294–6303, 2025.

[Wu *et al.*, 2024] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4D Gaussian Splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20310–20320, 2024.

[Wu *et al.*, 2025] Jiahao Wu, Rui Peng, Zhiyan Wang, Lu Xiao, Luyang Tang, Jinbo Yan, Kaiqiang Xiong, and Ronggang Wang. Swift4D: Adaptive divide-and-conquer Gaussian splatting for compact and efficient reconstruction of dynamic scene. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

[Xie *et al.*, 2024] Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. PhysGaussian: Physics-integrated 3D Gaussians for generative dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4389–4398, 2024.

[Xu *et al.*, 2025] Zhankuo Xu, Chaoran Feng, Yingtao Li, Jianbin Zhao, Jiashu Yang, Wangbo Yu, Li Yuan, and Yonghong Tian. Breaking the vicious cycle: Coherent 3D Gaussian splatting from sparse and motion-blurred views. *arXiv preprint arXiv:2512.10369*, 2025.

[Yang *et al.*, 2023] Jinyu Yang, Mingqi Gao, Zhe Li, Shang Gao, Fangjing Wang, and Feng Zheng. Track Anything: Segment Anything meets videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[Yang *et al.*, 2024a] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth Anything V2. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:21875–21911, 2024.

[Yang *et al.*, 2024b] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20331–20341, 2024.

[Yin *et al.*, 2025] Zhenyu Yin, Xiaohui Wang, Feiqing Zhang, Xiaoqiang Shi, and Dan Feng. MVD-NeRF: Multi-view deblurring neural radiance fields from defocused images. *Image and Vision Computing*, 153:105876, 2025.

[Zhang *et al.*, 2018] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.

[Zhang *et al.*, 2025] Xuankai Zhang, Junjin Xiao, and Qing Zhang. Dynamic Gaussian splatting from defocused and motion-blurred monocular videos. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

[Zhao *et al.*, 2024] Lingzhe Zhao, Peng Wang, and Peidong Liu. BAD-Gaussians: Bundle adjusted deblur Gaussian Splatting. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 233–250. Springer, 2024.

[Zhong *et al.*, 2022] Zhihang Zhong, Xiao Sun, Zhirong Wu, Yinqiang Zheng, Stephen Lin, and Imari Sato. Animation from blur: Multi-modal blur decomposition with motion guidance. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 599–615. Springer, 2022.

[Zhong *et al.*, 2023] Zhihang Zhong, Mingdeng Cao, Xiang Ji, Yinqiang Zheng, and Imari Sato. Blur interpolation transformer for real-world motion from blur. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5713–5723, 2023.