

MoBGS: Motion Deblurring Dynamic 3D Gaussian Splatting for Blurry Monocular Video

Minh-Quan Viet Bui^{1*}, Jongmin Park^{1*},
Juan Luis Gonzalez², Jaeho Moon¹, Jihyong Oh^{3†}, Munchurl Kim^{1†}

¹Korea Advanced Institute of Science and Technology (KAIST)

²Flawless AI

³Department of Imaging Science, GSAIM, Chung-Ang University

Abstract

We present MoBGS, a novel motion deblurring 3D Gaussian Splatting (3DGS) framework capable of reconstructing sharp and high-quality novel spatio-temporal views from blurry monocular videos in an end-to-end manner. Existing dynamic novel view synthesis (NVS) methods are highly sensitive to motion blur in casually captured videos, resulting in significant degradation of rendering quality. While recent approaches address motion-blurred inputs for NVS, they primarily focus on static scene reconstruction and lack dedicated motion modeling for dynamic objects. To overcome these limitations, our MoBGS introduces a novel Blur-adaptive Latent Camera Estimation (BLCE) method using a proposed Blur-adaptive Neural Ordinary Differential Equation (ODE) solver for effective latent camera trajectory estimation, improving global camera motion deblurring. In addition, we propose a Latent Camera-induced Exposure Estimation (LCEE) method to ensure consistent deblurring of both a global camera and local object motions. Extensive experiments on the Stereo Blur dataset and real-world blurry videos show that our MoBGS significantly outperforms the very recent methods, achieving state-of-the-art performance for dynamic NVS under motion blur.

Project Page — <https://kaist-viclab.github.io/mobgs-site>

Code — <https://github.com/KAIST-VICLab/MoBGS>

1 Introduction

Novel View Synthesis (NVS) has shown significant advancements in recent years, with applications spanning Virtual Reality (VR), Augmented Reality (AR), and film production. While dynamic NVS methods (Pumarola et al. 2021; Li et al. 2021, 2022; Wang et al. 2022; Weng et al. 2022; Oswald, Stühmer, and Cremers 2014; Collet et al. 2015; Broxton et al. 2020; Gao et al. 2021; Tretschk et al. 2021; Li et al. 2023b; Attal et al. 2023; Park et al. 2023; Shao et al. 2023; Fridovich-Keil et al. 2023; Yang et al. 2024; Wu et al. 2024a; Bae et al. 2024; Wang et al. 2024; Lei et al. 2024; Park et al. 2025; Huang et al. 2024) have made substantial progress in reconstructing realistic scenes

from monocular videos and multi-camera setups, their performance remains highly dependent on the quality of the given 2D observations. In particular, motion blur, which frequently occurs in casually captured videos due to fast-moving objects (Pan et al. 2016; Zhang et al. 2020) or camera shake (Bahat, Efrat, and Irani 2017; Zhang et al. 2018a), poses a major challenge. Since NVS techniques rely on the precise reconstruction of scene geometry and appearance, the loss of sharp details caused by blur can severely degrade the temporal consistency and rendering fidelity.

To reconstruct sharp 3D scenes from blurry 2D observations, several deblurring NVS methods (Ma et al. 2022; Wang et al. 2023; Lee et al. 2023a,b; Zhao, Wang, and Liu 2024; Lee et al. 2024a,b,c; Peng et al. 2025; Sun et al. 2024; Bui et al. 2025; Wu et al. 2024c) have been developed. These methods estimate camera trajectories during exposure, typically represented as latent camera poses, and generate a sequence of latent sharp images that are averaged to synthesize the blurred images, following the physical blur formation process (Zhao, Wang, and Liu 2024). Most of these methods focus only on static scene deblurring, without considering that object motion also occurs during the same exposure time along with camera movement, making them less effective for dynamic scenes with complex object motions.

A few recent methods (Sun et al. 2024; Bui et al. 2025; Wu et al. 2024c) have been explored for dynamic deblurring NVS from blurry monocular videos. While these dynamic deblurring NVS methods represent significant progress, they still face critical limitations in modeling both global camera and local object motion blur. First, the prior methods (Sun et al. 2024; Bui et al. 2025; Wu et al. 2024c) for latent camera pose estimation lack guidance from the input blur degree. Each frame’s blur degree offers a valuable prior for accurately predicting the set of latent camera poses during exposure, as more severe blurriness typically corresponds to longer and more complex motion paths (Chen and Clark 2025; Fang et al. 2025). The second limitation is inaccurate modeling of local motion blur, which arises from motion integrated over time. Since object-induced blur accumulates over this interval, accurately estimating this interval is crucial for capturing its complex and spatially varying behavior (Weng, Zhang, and Xiong 2023; Shang et al. 2023). However, recent methods do not explicitly incorporate this

*Co-first authors (equal contribution).

†Co-corresponding authors.

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

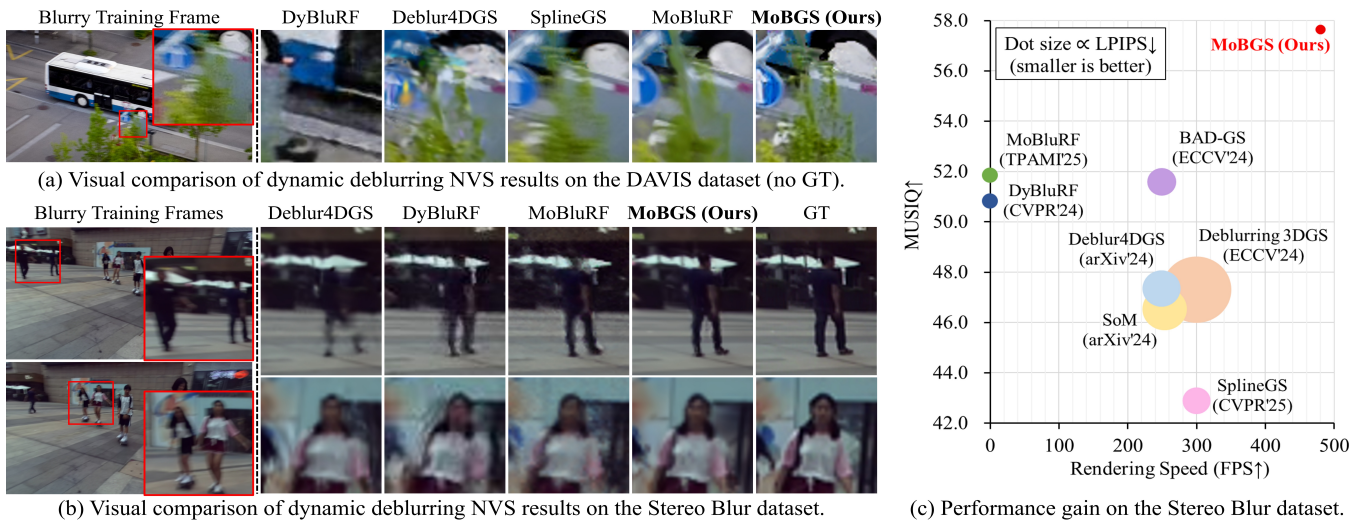


Figure 1: Our MoBGS achieves state-of-the-art dynamic deblurring novel view synthesis on (a) real-world casually captured blurry monocular videos (Pont-Tuset et al. 2018) and (b) synthesized blurry monocular videos (Sun et al. 2024), delivering (c) high perceptual quality (best LPIPS and MUSIQ scores) and fast rendering (~ 500 FPS). Each region is cropped and enlarged.

relationship when estimating the temporal span of motion integration or when modeling local motion blur.

To overcome these limitations and enable high-quality, sharp novel view synthesis from blurry monocular videos, we propose MoBGS, a novel Motion deBlurring method for NVS, based on dynamic 3D Gaussian Splatting (Kerbl et al. 2023). MoBGS first accurately predicts latent camera poses for each blurry frame by leveraging the input frame’s blur intensity as direct guidance using our Blur-adaptive Latent Camera Estimation (BLCE) method. Then, based on these estimated latent camera poses, MoBGS estimates a latent exposure time for motion blur via our Latent Camera-induced Exposure Estimation (LCEE) method. The estimated latent exposure duration is then used to model motion blur from moving objects while ensuring the temporal consistency of global camera motion and local object motion blur. Extensive quantitative and qualitative evaluations on the Stereo Blur dataset (Sun et al. 2024) and real-world videos (Pont-Tuset et al. 2018) show that our MoBGS substantially outperforms recent state-of-the-art (SOTA) methods in dynamic deblurring NVS. Our main contributions are as follows:

- We introduce a novel MoBGS framework for reconstructing high-quality spatio-temporal novel views from blurry monocular videos.
- A *Blur-adaptive Latent Camera Estimation* (BLCE) method is proposed to accurately estimate latent camera trajectories, considering the blurriness of video frames via our novel Blur-adaptive Neural ODE.
- A *Latent Camera-induced Exposure Estimation* (LCEE) method is proposed to estimate latent exposure time, ensuring consistent camera motion and object motion blur.
- Our MoBGS framework is extensively evaluated on the Stereo Blur dataset (Sun et al. 2024) and real-world blurry video sequences (Pont-Tuset et al. 2018), demonstrating *significant* improvements over SOTA methods.

2 Related Work

Dynamic Novel View Synthesis. Building on the success of Neural Radiance Field (NeRF) (Mildenhall et al. 2020) for static scenes, several methods have been developed for dynamic NVS. These include mapping observations to a shared canonical space (Park et al. 2021a,b; Song et al. 2023; Liu et al. 2023), estimating 3D scene flow (Li et al. 2021; Gao et al. 2021; Li et al. 2023b; Du et al. 2021), and discretizing space-time volumes using grid-based architectures (Cao and Johnson 2023; Fridovich-Keil et al. 2023; Shao et al. 2023; Attal et al. 2023). More recently, 3DGS-based methods (Kerbl et al. 2023) have been proposed for efficient, high-fidelity dynamic NVS. These approaches model temporal deformation of Gaussians using per-frame attribute offsets (Yang et al. 2024), grid-based encodings (Wu et al. 2024b), learned embeddings (Bae et al. 2024), rigid transformations (Wang et al. 2024), motion decomposition (Kwak et al. 2025), or spline-based trajectories (Park et al. 2025). While being effective under clean inputs, these methods fundamentally assume that each training frame is an instantaneous snapshot. As a result, they misinterpret motion blur as a feature of the scene. This leads to blur being baked into the 3D representation, resulting in blurry or ringing artifacts when trained on casually captured monocular videos.

Deblurring Novel View Synthesis. In deblurring NVS, early works like DeblurNeRF (Ma et al. 2022) and BAD-NeRF (Wang et al. 2023) focused on modeling blur caused by camera motion, either through predicted blur kernels or estimated camera trajectories. These concepts were later adopted into more efficient 3DGS representation methods like Deblurring 3DGS (Lee et al. 2024a) and BAD-GS (Zhao, Wang, and Liu 2024). SMURF (Lee et al. 2024c) was the first to use a Neural ODE (Chen et al. 2018) to predict continuous latent camera trajectories. However, their per-pixel blur kernel estimation introduces inefficiency due

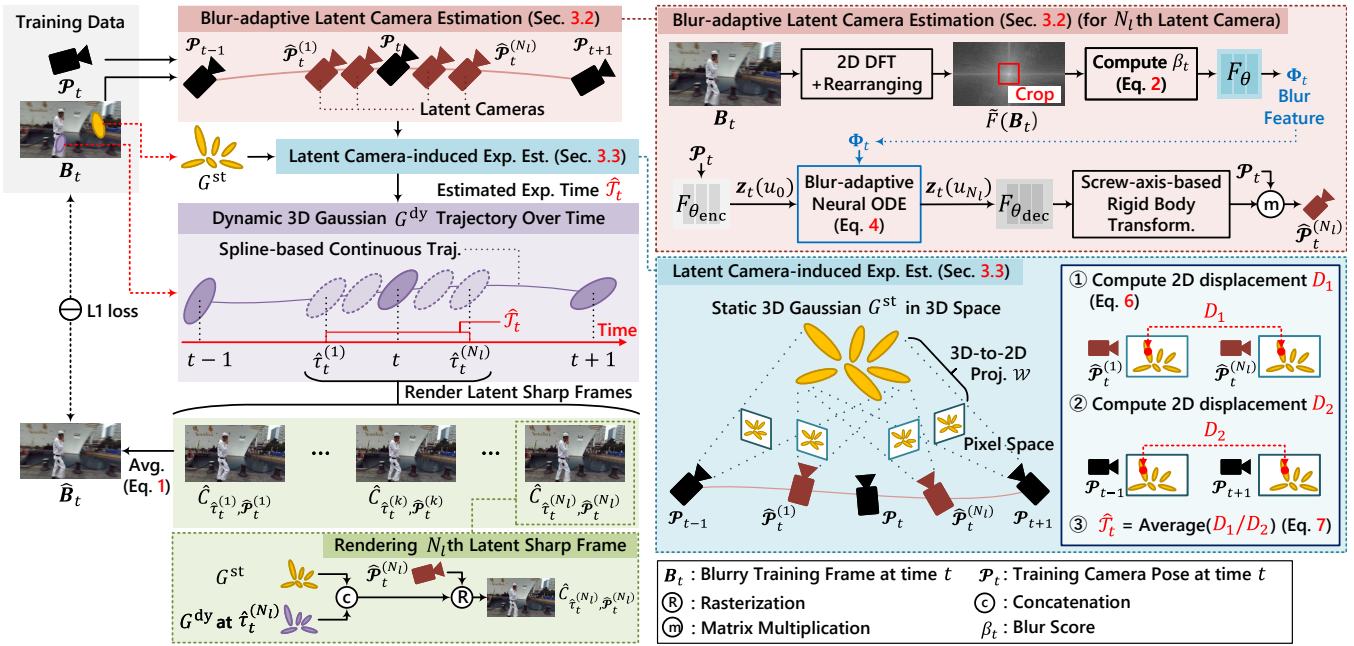


Figure 2: **Overview of MoBGS.** MoBGS accurately models scene blurriness by jointly considering global camera and local object motion over the same exposure time. It first estimates latent camera poses for each blurry frame using the Blur-adaptive Latent Camera Estimation (BLCE) method. Then, leveraging these poses, it estimates the corresponding exposure time via the Latent Camera-induced Exposure Estimation (LCEE) method, ensuring a consistent blur modeling of local moving objects.

to repeated computation. CRiM-GS (Lee et al. 2024b) improves efficiency by predicting global camera motion for all pixels but lacks the necessary prior information for accurate latent camera pose estimation. Importantly, none of these methods are designed to handle dynamic reconstruction.

The primary challenge in dynamic deblurring NVS has shifted to complex scenes where both camera and object motions contribute to the blur. Several methods have been proposed to tackle this problem (Sun et al. 2024; Wu et al. 2024c; Bui et al. 2025), marking significant progress in the field. DyBluRF (Sun et al. 2024) applies camera trajectory estimation from BAD-NeRF (Wang et al. 2023) on top of a dynamic NeRF, assuming fixed exposure time for object motion. MoBluRF (Bui et al. 2025) decomposes the 2D blur kernel into global camera and local object motion blur, enhancing sharp dynamic NeRF learning, although it omits exposure time estimation. Deblur4DGS (Wu et al. 2024c) addresses dynamic deblurring NVS using 3DGS, incorporating regularization techniques and modeling object motion blur with a learnable exposure time. Despite these advancements, the prior methods still encounter difficulties in accurately modeling both global camera and local object blur components. First, their camera motion estimation lacks guidance from the blurred input frames, limiting adaptability. Second, their approaches to exposure time estimation overlook the simultaneous occurrence of global camera motion and local object motion blur. As a result, they struggle to robustly estimate the exposure time for modeling dynamic object motion blur.

3 Proposed Method: MoBGS

3.1 Overview of MoBGS

Fig. 2 illustrates our MoBGS framework. Given a set of N_f blurry monocular video frames $\{B_t\}_{t=1}^{N_f}$ with corresponding camera poses $\{\mathcal{P}_t\}_{t=1}^{N_f}$, where each $\mathcal{P}_t \in \mathbb{R}^{4 \times 4}$, and a shared camera intrinsic matrix $K \in \mathbb{R}^{3 \times 3}$, MoBGS aims to synthesize sharp novel views by accurately modeling scene blurriness caused by both global camera motion and local object motion. MoBGS is built on SplineGS (Park et al. 2025), which represents each scene with static $\{G_i^{st}\}_{i=1}^{n^{st}}$ and dynamic 3D Gaussians $\{G_i^{dy}\}_{i=1}^{n^{dy}}$, where the latter are deformed via splines to ensure smooth and continuous motion. More details are provided in *Supplementary*.

Our blur modeling builds on two key insights: first, blur intensity serves as a valuable prior for predicting latent camera poses corresponding to global camera motion blur; second, both static backgrounds and moving objects undergo blurring during the same temporal interval in each frame. Based on these insights, MoBGS systematically decouples global camera motion and local object motion blur components. To handle camera motion blur, the Blur-adaptive Latent Camera Estimation (BLCE) method (Sec. 3.2) estimates a set of N_l latent camera poses $\{\hat{\mathcal{P}}_t^{(k)}\}_{k=1}^{N_l}$ for each blurry frame B_t by incorporating blur intensity as an additional input. These poses are then used by the Latent Camera-induced Exposure Estimation (LCEE) method (Sec. 3.3) to compute the frame-wise *latent* exposure time $\hat{\tau}_t$, which represents a motion-integration interval that characterizes how

much blur is accumulated within each frame. To model spatially-varying blur from local object motion, we sample latent timestamps $\{\hat{\tau}_t^{(k)}\}_{k=1}^{N_l}$ within the estimated latent exposure interval $\hat{\mathcal{T}}_t$ and use them to predict corresponding object motion, enabling fine-grained motion blur modeling. Finally, we render each latent sharp frame at a sampled timestamp $\hat{\tau}_t^{(k)}$ and its corresponding latent camera pose $\hat{\mathcal{P}}_t^{(k)}$, and denote it as $\hat{\mathcal{C}}_{\hat{\tau}_t^{(k)}, \hat{\mathcal{P}}_t^{(k)}}$. Following previous works (Wang et al. 2023; Lee et al. 2023a; Sun et al. 2024; Zhao, Wang, and Liu 2024; Luthra et al. 2024; Wu et al. 2024c), we formulate the final blurry frame by approximating continuous exposure with a discrete set of timestamps and averaging the corresponding latent sharp frames as:

$$\hat{\mathbf{B}}_t = \frac{1}{N_l} \sum_{k=1}^{N_l} \hat{\mathcal{C}}_{\hat{\tau}_t^{(k)}, \hat{\mathcal{P}}_t^{(k)}}, \quad \hat{\tau}_t^{(k)} = t + \hat{\mathcal{T}}_t \frac{k - \lceil \frac{N_l}{2} \rceil}{N_l}, \quad (1)$$

where $\lceil \cdot \rceil$ denotes the ceiling (round-up) operation. We refer readers to the *Supplementary* for a detailed explanation of the blur modeling process underlying Eq. (1).

3.2 Blur-adaptive Latent Camera Estimation

To address the blurriness caused by global camera motion, we propose the BLCE method which estimates latent camera poses guided by the blur intensity of each frame. To compute the blur intensity, we leverage the observation that blurry frames contain a higher proportion of low-frequency components due to the motion blur (Liu, Li, and Jia 2008; Shi, Xu, and Jia 2015, 2014; Yan and Shao 2016). Based on this, we define the blur score as the ratio of the magnitudes within a low-frequency region to the total magnitudes in the frequency spectrum of the frame. Specifically, we apply a 2D Discrete Fourier Transform (DFT) to \mathbf{B}_t , and rearrange the resulting DFT coefficients to center the low-frequency components, producing the shifted DFT $\tilde{\mathcal{F}}(\mathbf{B}_t)$. The blur score β_t is then given by:

$$\beta_t = \frac{\sum_{\xi \in \Lambda} M_t(\xi)}{\sum_{\xi} M_t(\xi)}, \quad \text{where } M_t = |\tilde{\mathcal{F}}(\mathbf{B}_t)|, \quad (2)$$

where ξ denotes a 2D frequency index and Λ represents the center-cropped square region with a side length of s , which covers the low-frequency components in $\tilde{\mathcal{F}}(\mathbf{B}_t)$. The blur feature Φ_t is extracted from β_t using a shallow MLP F_θ as:

$$\Phi_t = F_\theta(\phi(\beta_t)), \quad (3)$$

where $\phi(\cdot)$ denotes the positional encoding (Mildenhall et al. 2020). Leveraging this blur feature Φ_t , we propose a novel Blur-adaptive Neural ODE solver, built upon Neural ODE (Chen et al. 2018), which predicts latent camera poses $\{\hat{\mathcal{P}}_t^{(k)}\}_{k=1}^{N_l}$ along a smooth, temporally ordered trajectory. We encode the initial latent feature $\mathbf{z}_t(u_0)$ by feeding the camera pose \mathcal{P}_t into a shallow MLP $F_{\theta_{\text{enc}}}$, such that $\mathbf{z}_t(u_0) = F_{\theta_{\text{enc}}}(\mathcal{P}_t)$. A sequence of N_l latent vectors $\{\mathbf{z}_t(u_k)\}_{k=1}^{N_l}$ is then computed from $\mathbf{z}_t(u_0)$ as:

$$\mathbf{z}_t(u_k) = \mathbf{z}_t(u_0) + \int_{u_0}^{u_k} f(\mathbf{z}_t(u), u, \Phi_t; \psi) du, \quad (4)$$

where f is a shallow neural network parameterized by ψ , representing the derivative of the latent feature $d\mathbf{z}_t(u)/du$ and u denotes a temporal coordinate within the exposure duration. Unlike the existing methods (Lee et al. 2024c,b), we inject the blur feature Φ_t to f , enabling it to be guided by the blur intensity of each frame. $\mathbf{z}_t(u_k)$ is used to predict a screw axis $(\omega_t^{(k)}; \mathbf{v}_t^{(k)}) \in \mathbb{R}^6$ via a decoder $F_{\theta_{\text{dec}}}$ as:

$$(\omega_t^{(k)}; \mathbf{v}_t^{(k)}) = F_{\theta_{\text{dec}}}(\mathbf{z}_t(u_k)). \quad (5)$$

Finally, we predict the latent camera pose $\hat{\mathcal{P}}_t^{(k)}$ by applying a residual transformation $\Psi(\omega_t^{(k)}, \mathbf{v}_t^{(k)})$ to the initial camera pose \mathcal{P}_t , resulting in $\hat{\mathcal{P}}_t^{(k)} = \mathcal{P}_t \Psi(\omega_t^{(k)}, \mathbf{v}_t^{(k)})$, where $\Psi(\cdot)$ denotes the screw-axis-based rigid-body transformation introduced in Nerfies (Park et al. 2021a). By injecting blur intensity guidance, the BLCE method facilitates more stable learning of latent camera poses, leading to improved dynamic deblurring NVS performance across the entire frame, as demonstrated in Table 1, Table 2, and Table 3.

3.3 Latent Camera-induced Exposure Estimation

We propose the LCEE method to estimate the latent exposure time $\hat{\mathcal{T}}_t$ using prior knowledge from the latent camera trajectories predicted by BLCE (Sec. 3.2). To avoid confusion, $\hat{\mathcal{T}}_t$ is not the physical exposure time; rather, it represents the learned temporal span over which motion blur accumulates, allowing the model to adapt to varying blur levels. Given this interpretation of $\hat{\mathcal{T}}_t$, our key insight is that global camera motion and local object motion blur share the same exposure interval. Motivated by this, we interpret $\hat{\mathcal{T}}_t$ as being proportional to the temporal interval between the first and last latent camera poses, $\hat{\mathcal{P}}_t^{(1)}$ and $\hat{\mathcal{P}}_t^{(N_l)}$. As illustrated in the bottom-right blue box of Fig. 2, we estimate the latent exposure time $\hat{\mathcal{T}}_t$ by comparing the magnitude of camera motion across two intervals. Based on the assumption that the consecutive training camera poses \mathcal{P}_{t-1} and \mathcal{P}_t are evenly separated by a single timestep, we compute $\hat{\mathcal{T}}_t$ as the ratio between the camera movement from $\hat{\mathcal{P}}_t^{(1)}$ to $\hat{\mathcal{P}}_t^{(N_l)}$ and that from \mathcal{P}_{t-1} to \mathcal{P}_{t+1} . Since camera motion induces 2D displacements of static scene points on the image plane, we quantify the relative camera movement by measuring these projected displacements. Let $\mathcal{W}(\mathcal{P}, \mathbf{x})$ be a projection function that maps a 3D position \mathbf{x} to a 2D pixel coordinate under a given camera pose \mathcal{P} . Given the estimated latent camera poses $\{\hat{\mathcal{P}}_t^{(k)}\}_{k=1}^{N_l}$ corresponding to each blurry frame \mathbf{B}_t , we select $\hat{\mathcal{P}}_t^{(1)}$ and $\hat{\mathcal{P}}_t^{(N_l)}$. As shown in Fig. 2, we compute the 2D displacement of each static 3D Gaussian mean μ_i^{st} between $\hat{\mathcal{P}}_t^{(1)}$ and $\hat{\mathcal{P}}_t^{(N_l)}$ as:

$$D(\hat{\mathcal{P}}_t^{(1)}, \hat{\mathcal{P}}_t^{(N_l)}, \mu_i^{\text{st}}) = \|\mathcal{W}(\hat{\mathcal{P}}_t^{(1)}, \mu_i^{\text{st}}) - \mathcal{W}(\hat{\mathcal{P}}_t^{(N_l)}, \mu_i^{\text{st}})\|_2. \quad (6)$$

This projected displacement captures the shift of the static 3D Gaussian on the image plane caused by global camera motion between $\hat{\mathcal{P}}_t^{(1)}$ and $\hat{\mathcal{P}}_t^{(N_l)}$ during the exposure, representing the movement of the camera pair $\hat{\mathcal{P}}_t^{(1)}$ and $\hat{\mathcal{P}}_t^{(N_l)}$.

	Methods	Full image		Dynamic region		FPS	Train Time (hr)
		LPIPS↓ / MUSIQ↑ / tOF↓ / PSNR↑	LPIPS↓ / tOF↓ / PSNR↑	LPIPS↓ / tOF↓ / PSNR↑	LPIPS↓ / tOF↓ / PSNR↑		
Dynamic NVS	D3DGS (CVPR'24) (Yang et al. 2024)	0.446 / 34.33 / 6.972 / 15.91	0.542 / 8.215 / 13.81	57	0.30		
	4DGS (CVPR'24) (Wu et al. 2024b)	0.270 / 39.74 / 5.400 / 18.53	0.367 / 7.574 / 14.95	193	0.25		
	SoM (arXiv'24) (Wang et al. 2024)	0.219 / 46.52 / 2.395 / 22.92	0.272 / 2.876 / 17.18	254	1.2		
	SplineGS (CVPR'25) (Park et al. 2025)	0.141 / 42.88 / 1.409 / 26.41	0.168 / 1.417 / <u>22.31</u>	300	0.25		
Cascade	Restormer (CVPR'22) (Zamir et al. 2022) + SoM	0.167 / 53.06 / 2.101 / 22.82	0.238 / 2.646 / 17.09	254	1.2		
	GShiftNet (CVPR'23) (Li et al. 2023a) + SoM	0.158 / 54.76 / 2.070 / 22.62	0.253 / 2.604 / 16.64	254	1.2		
	Restormer (CVPR'22) (Zamir et al. 2022) + SplineGS	0.081 / 54.01 / 0.911 / <u>26.83</u>	0.118 / 1.249 / 22.26	<u>336</u>	0.25		
	GShiftNet (CVPR'23) (Li et al. 2023a) + SplineGS	<u>0.074</u> / <u>55.29</u> / <u>0.748</u> / 26.54	<u>0.108</u> / <u>1.174</u> / 22.06	329	0.25		
Deblurring NVS	Deblurring 3DGS (ECCV'24) (Lee et al. 2023b)	0.347 / 47.30 / 3.342 / 15.76	0.466 / 6.559 / 12.63	300	0.18		
	BAD-GS (ECCV'24) (Zhao, Wang, and Liu 2024)	0.146 / 51.58 / 1.551 / 21.43	0.330 / 4.207 / 15.18	250	<u>0.2</u>		
	DyBluRF (CVPR'24) (Sun et al. 2024)	0.079 / 50.82 / 0.889 / 25.62	0.158 / 1.367 / 19.41	0.2	51		
	Deblur4DGS (arXiv'24) (Wu et al. 2024c)	0.191 / 47.33 / 1.880 / 23.37	0.299 / 3.248 / 17.53	250	2.4		
	MoBluRF (IEEE TPAMI) (Bui et al. 2025)	0.078 / 51.84 / 0.816 / 25.69	0.155 / 1.456 / 20.63	0.1	50		
	MoBGS (Ours)	0.050 / 57.64 / 0.507 / 28.80	0.096 / 1.093 / 23.41	480	1.5		

Table 1: Dynamic deblurring novel view synthesis evaluation on the Stereo Blur dataset. **Bold** and underline denote the best and second best performances, respectively. Per-scene results are provided in the *Supplementary*.

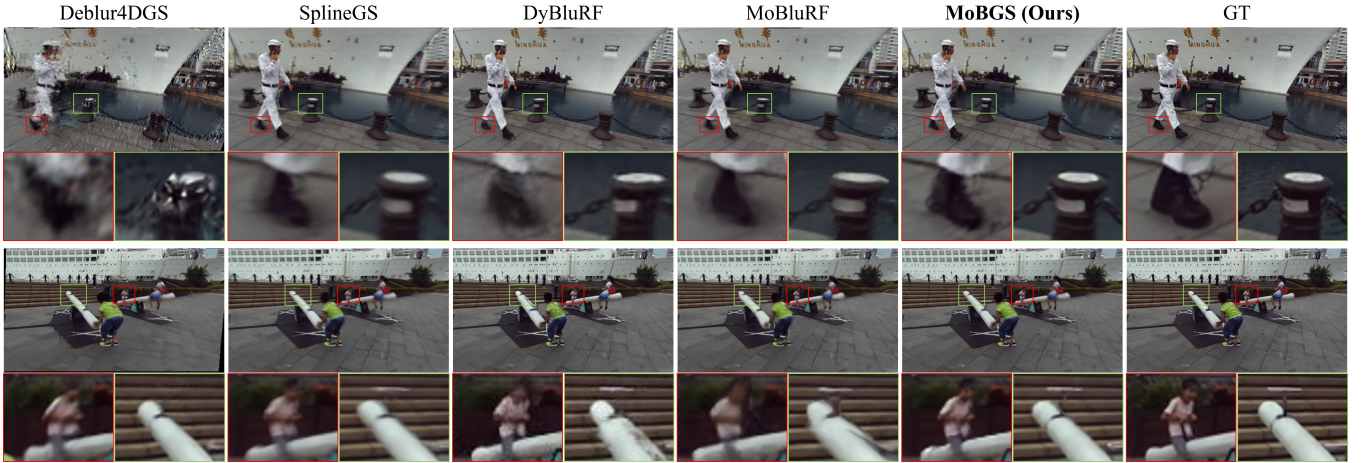


Figure 3: Visual comparisons for dynamic deblurring novel view synthesis on the Stereo Blur dataset.

Similarly, as shown in Fig. 2, we compute the 2D displacement between the two projections of the mean μ_i^{st} for the *training* (given) camera poses \mathcal{P}_{t-1} and \mathcal{P}_{t+1} , representing the displacement over the interval from $t-1$ to $t+1$, denoted as $D(\mathcal{P}_{t-1}, \mathcal{P}_{t+1}, \mu_i^{\text{st}})$. To estimate the per-frame latent exposure time $\hat{\mathcal{T}}_t$, we compute the average ratio between the two 2D displacements as:

$$\hat{\mathcal{T}}_t = \frac{2}{n^{\text{st}}} \sum_{i=1}^{n^{\text{st}}} \frac{D(\hat{\mathcal{P}}_t^{(1)}, \hat{\mathcal{P}}_t^{(N_i)}, \mu_i^{\text{st}}) + \epsilon}{D(\mathcal{P}_{t-1}, \mathcal{P}_{t+1}, \mu_i^{\text{st}}) + \epsilon}, \quad (7)$$

where n^{st} is the number of static 3D Gaussians and ϵ is a smoothing term to avoid numerical instabilities. As shown in Eq. (1), the estimated latent exposure time $\hat{\mathcal{T}}_t$ is used to compute the latent timestamps $\hat{\tau}_t^{(k)}$. Then, each latent sharp frame $\hat{\mathcal{C}}_{\hat{\tau}_t^{(k)}, \hat{\mathcal{P}}_t^{(k)}}$ at each latent timestamp $\hat{\tau}_t^{(k)}$ is independently rendered using the static $\{G_i^{\text{st}}\}_{i=1}^{n^{\text{st}}}$ and dynamic

3D Gaussians $\{G_i^{\text{dy}}\}_{i=1}^{n^{\text{dy}}}$. Finally, the rendered blurry frame $\hat{\mathcal{B}}_t$ is obtained by averaging these latent sharp frames. The LCEE effectively handles local object motion blur by leveraging the temporal alignment between global camera motion blur and local object motion blur. We evaluate LCEE's effectiveness by comparing dynamic-region deblurring performance in Table 1 and examining dynamic-region renderings under different exposure estimations in Table 4 and Fig. 6.

3.4 Optimization

We optimize our MoBGS using a photometric loss \mathcal{L}_{rgb} and a depth loss $\mathcal{L}_{\text{depth}}$ defined as:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rgb}} \mathcal{L}_{\text{rgb}} + \lambda_{\text{depth}} \mathcal{L}_{\text{depth}}, \quad (8)$$

where \mathcal{L}_{rgb} is an L1 loss between the given blurry input frame \mathcal{B}_t and our rendered blurry frame $\hat{\mathcal{B}}_t$. $\mathcal{L}_{\text{depth}}$ is an L1 loss between the rendered depth from rasterization and the ground-truth depth.

4 Experiments

Implementation Details. We build our framework upon the widely adopted open-source 3D Gaussian Splatting (3DGS) codebase (Kerbl et al. 2023). For depth estimation and 2D tracking, we utilize the pre-trained models UniDepth (Picinelli et al. 2024) and BootsTAP (Doersch et al. 2024), respectively. Our model is trained over 10K iterations, with the LCEE applied after 2K iterations. For each blurry input frame, our BLCE module estimates $N_l = 9$ latent images. The blur score β_t for the blur feature \mathcal{B}_t is computed using $s = 20$. The loss coefficients are empirically found and set as follows: $\lambda_{\text{rgb}} = 1.0$, $\lambda_{\text{depth}} = 0.2$. We use $N_c = 12$ control points for the spline-based motion modeling. Our MoBGS is trained and evaluated on a single NVIDIA RTX 3090Ti.

Datasets. To assess the performance of our MoBGS and other SOTA methods, we utilize two datasets: the Stereo Blur dataset (Sun et al. 2024) and the DAVIS dataset (Pont-Tuset et al. 2018). The Stereo Blur dataset consists of stereo videos from six scenes with significant motion blur, each featuring a blurry left-view video and a corresponding sharp right-view video for evaluation. We follow the same training and testing splits as DyBluRF (Sun et al. 2024). The DAVIS dataset represents real-world scenarios with diverse scenes containing rapid object motions, resulting in natural motion blur. As ground-truth (GT) for NVS is not available, we adhere to the protocol used in (Park et al. 2025) by visualizing fixed-view, varying-time renderings to assess NVS quality.

Metrics. We evaluate the rendering quality of each method using LPIPS (Zhang et al. 2018b), MUSIQ (Ke et al. 2021), and Peak Signal-to-Noise Ratio (PSNR). These metrics are computed across both the full image and dynamic regions. We further assess the temporal consistency across consecutive rendered frames using tOF (Chu et al. 2020) score. Notably, increased blur can sometimes artificially elevate PSNR due to reduced sensitivity to small pixel misalignments, while perceptual metrics like LPIPS and MUSIQ better reflect human visual perception (Zhang et al. 2018c; Barron et al. 2022; Tucker and Snavely 2020). Therefore, we emphasize the importance of perceptual metrics, particularly LPIPS and MUSIQ, in evaluating deblurring performance.

4.1 Comparison with State-of-the-Art Methods

Dynamic Deblurring NVS. The quantitative results for the joint evaluation of deblurring and dynamic NVS on the Stereo Blur dataset (Sun et al. 2024) are presented in Table 1. As demonstrated, recent dynamic NVS methods (Yang et al. 2024; Wu et al. 2024c; Bae et al. 2024; Wang et al. 2024; Park et al. 2025) suffer from poor perceptual quality and low temporal consistency due to motion blur, as evidenced by higher (worse) LPIPS and tOF, alongside lower (worse) MUSIQ scores. This degradation stems from the lack of a dedicated deblurring module, which limits their abilities to effectively reconstruct sharp and temporally coherent novel views. Notably, our MoBGS achieves **2.7× higher perceptual scores** compared to the SOTA method, SplineGS.

We then compare our results with cascade-based approaches which apply 2D deblurring networks as a preprocessing step. While the cascade approaches offer improve-

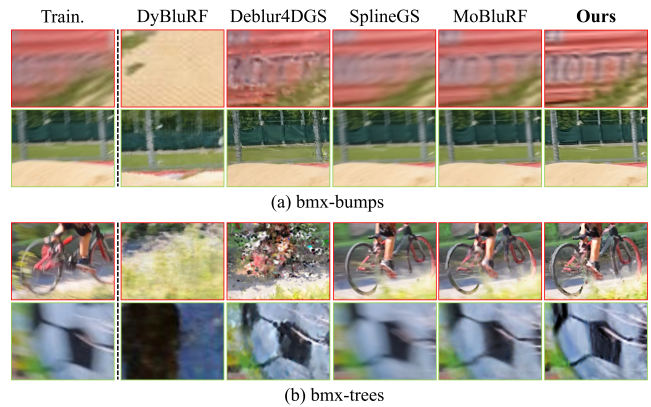


Figure 4: Visual comparisons for dynamic deblurring novel view synthesis on the DAVIS dataset. ‘Train.’ refers to the training frame. The results of DyBluRF (Sun et al. 2024) exhibit significant misalignment, as its latent camera pose optimization is overfitted to the training poses. Full image results are provided in the *Supplementary*.

ments over existing dynamic NVS methods, the combination of the two best methods (GShiftNet (Li et al. 2023a) + SplineGS (Park et al. 2025)) still lags behind our MoBGS, particularly in structural accuracy, as reflected by the **1.3-2dB PSNR gaps** in both the full and dynamic regions.

Table 1 also shows that the static deblurring NVS methods (Lee et al. 2024a; Zhao, Wang, and Liu 2024) fail to handle scene dynamics effectively. Notably, our MoBGS outperforms the latest state-of-the-art (SOTA) dynamic deblurring NVS method, MoBluRF (Bui et al. 2025), *by very large margins in all metrics*, while achieving **4,800× faster** rendering speeds and **34× faster** training time. Figs. 1, 3 and 4 illustrate the qualitative results on both the Stereo Blur dataset (Sun et al. 2024) and the real-world videos from the DAVIS dataset (Pont-Tuset et al. 2018). Our visualizations show that our MoBGS yields *significant* improvements in NVS quality, consistent with our quantitative results. More visualizations and demo videos are in the *Supplementary*.

Deblurring Effect. Table 2 shows the deblurring effect on blurry training views. As expected, the dynamic NVS methods (SoM and SplineGS) tend to overfit to the blurriness of input frames, leading to low visual quality, as evidenced by poor LPIPS and MUSIQ scores. Consistent with our findings in Table 1 for the joint deblurring and dynamic NVS evaluation, our MoBGS outperforms cascade-based and dynamic deblurring NVS methods *in all evaluation metrics*.

4.2 Ablation Study

Blur-adaptive Latent Camera Estimation (BLCE). As shown in Table 3, removing the deblurring mechanism (i.e., ‘No Deblurring module’) results in a significant degradation of performance across all metrics, underscoring the negative impact of blurry training frames. These results highlight the effectiveness of our joint 3D reconstruction and deblurring approach for dynamic NVS. Additionally, we compare different latent camera estimation methods by replacing our BLCE with commonly used approaches from (Sun et al.

Methods	LPIPS↓ / MUSIQ↑ / tOF↓ / PSNR↑
SoM	0.207 / 46.55 / 2.268 / 24.53
SplineGS	0.126 / 42.60 / 1.269 / 29.22
Restormer + SoM	0.151 / 52.60 / 1.905 / 24.70
GShiftNet + SoM	0.139 / 54.20 / 1.868 / 24.45
Restormer + SplineGS	0.064 / 53.66 / 0.787 / <u>30.31</u>
GShiftNet + SplineGS	<u>0.058</u> / 55.24 / 0.641 / 30.31
Deblurring 3DGS	0.220 / 50.88 / 2.262 / 20.47
BAD-GS	0.119 / 53.39 / 1.527 / 25.27
DyBluRF	0.064 / 52.16 / 0.740 / 29.15
Deblur4DGS	0.183 / 48.94 / 1.980 / 23.70
MoBluRF	0.073 / 49.99 / 0.742 / 29.27
MoBGS (Ours)	0.040 / 57.34 / 0.454 / 30.96

Table 2: Deblurring effect on *full images* on the Stereo Blur dataset. Per-scene results are in the *Supplementary*.

Methods	LPIPS↓	MUSIQ↑	tOF↓
No Deblurring module (SplineGS)	0.141	42.88	1.409
Deblurring w/ Spline Interp. (BAD-GS-like)	0.058	55.38	0.570
Deblurring w/ Neural ODE (CRiM-GS-like)	<u>0.057</u>	<u>55.49</u>	<u>0.536</u>
Deblurring w/ BLCE (Ours)	0.050	56.93	0.507

Table 3: Ablation on BLCE. We evaluate dynamic deblurring NVS results of *full image* on the Stereo Blur dataset.

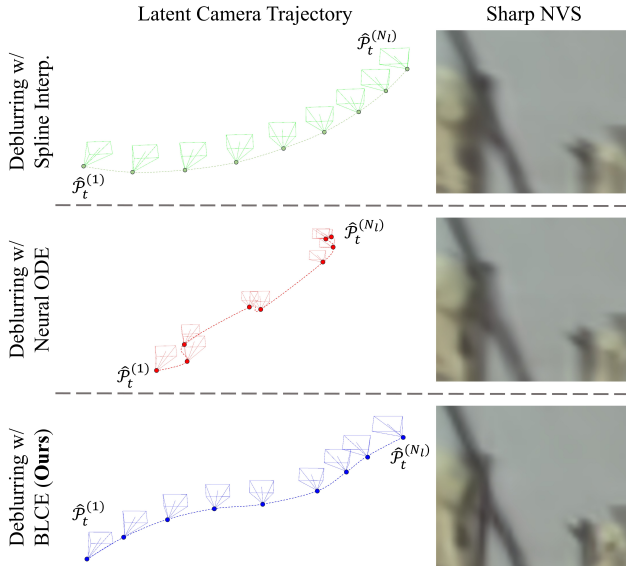


Figure 5: Visual comparisons for BLCE ablation study.

2024; Zhao, Wang, and Liu 2024), which use spline interpolation (denoted as ‘Deblurring w/ Spline Interp.’), and from (Lee et al. 2024c,b), which use Neural ODE (denoted as ‘Deblurring w/ Neural ODE’). Table 3 demonstrates that incorporating blur intensity guidance results in improvements across all perceptual metrics of the rendered images.

Methods	Dynamic deblurring NVS	Deblurring effect
	LPIPS↓ / tOF↓ / PSNR↑	LPIPS↓ / tOF↓ / PSNR↑
Fixed $\hat{\mathcal{T}}_t = 0.0$	0.120 / <u>1.237</u> / 23.20	0.112 / <u>1.055</u> / 26.39
Fixed $\hat{\mathcal{T}}_t = 0.5$	<u>0.117</u> / 1.276 / 23.12	<u>0.109</u> / 1.115 / 25.60
Fixed $\hat{\mathcal{T}}_t = 0.9$	0.157 / 2.014 / 19.41	0.150 / 1.863 / 20.84
Learnable $\hat{\mathcal{T}}_t$	0.128 / 1.261 / <u>23.24</u>	0.120 / 1.061 / <u>26.45</u>
LCEE (Ours)	0.096 / 1.093 / 23.41	0.085 / 0.896 / 26.54

Table 4: Ablation on LCEE. We evaluate both dynamic deblurring NVS and deblurring effect results of *dynamic region* on the Stereo Blur dataset.

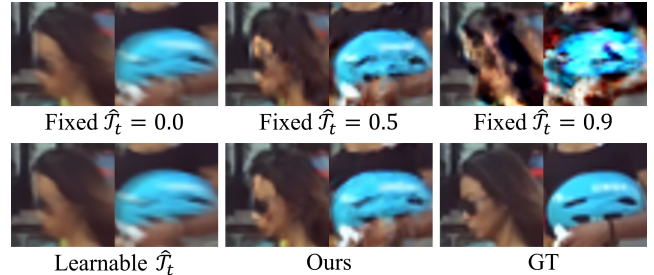


Figure 6: Visual comparisons for LCEE ablation study.

We present the latent camera trajectories and visual comparisons for the ablated models in Fig. 5. As shown, ‘Deblurring w/ Spline Interp.’ generates a smooth but oversimplified camera trajectory, which fails to capture fine details, leading to blurry reconstruction. On the other hand, ‘Deblurring w/ Neural ODE’ produces jagged and inconsistent camera poses. In contrast, our BLCE with the novel Blur-adaptive Neural ODE solver produces a smooth camera trajectory while ensuring sharp reconstruction, which is consistent with our superior quantitative results in Table 3.

Latent Camera-induced Exposure Estimation (LCEE). Unlike prior works (Bui et al. 2025; Sun et al. 2024; Wu et al. 2024c), our LCEE leverages the camera motion prior to estimate the latent exposure time, ensuring robust deblurring of dynamic regions. As shown in Fig. 6, the fixed latent exposure time setting (e.g., $\hat{\mathcal{T}}_t = 0.9$) introduces artifacts in dynamic regions, whereas setting $\hat{\mathcal{T}}_t = 0.0$ fails to account for any motion of dynamic objects, leading to blurry dynamic objects. The adverse effects of the above $\hat{\mathcal{T}}_t$ settings are also reflected in a worse perceptual score (LPIPS) and temporal inconsistencies (tOF) of *dynamic regions* in Table 4. While the fixed latent exposure time $\hat{\mathcal{T}}_t = 0.5$ mitigates these artifacts, it remains suboptimal for handling varying levels of motion blur, requiring manual fine-tuning across different scenarios. Alternatively, treating $\hat{\mathcal{T}}_t$ as a learnable parameter without any constraints, as in (Wu et al. 2024c), results in inconsistent deblurring between static and dynamic regions. This approach degrades the deblurring effect on moving objects and the temporal consistency, as reflected in the metrics of the ‘Learnable $\hat{\mathcal{T}}_t$ ’ variant in Table 4, and is evidenced by the noisy artifacts in Fig. 6.

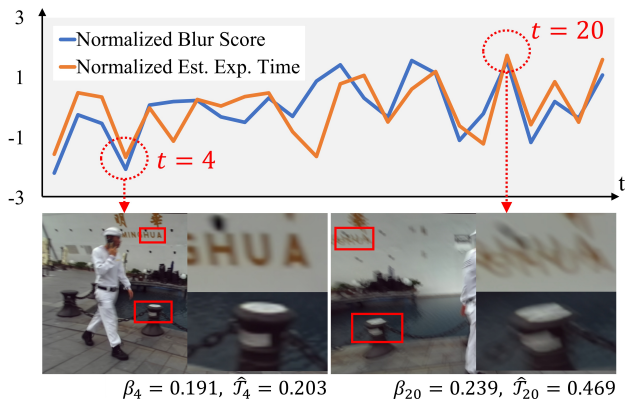


Figure 7: Correlation between our estimated \hat{T}_t (Eq. (7)) and our blur score β_t (Eq. (2)). For a clearer comparison of the trend, we normalize β_t and \hat{T}_t separately using their respective means and standard deviations for the graph.

We further evaluate the robustness of our estimated \hat{T}_t by plotting its normalized values alongside the normalization of blur score β_t in Fig. 7. By analyzing the correlation between these values and the corresponding blurry inputs, we observe that \hat{T}_t effectively aligns with the degree of blurriness present in the input frames. This correlation demonstrates that our latent exposure time estimation is both reliable and adaptive to varying levels of motion blur, leading to robust deblurring performance across a wide range of scenarios.

5 Conclusion

We introduce MoBGS, a novel dynamic deblurring NVS method for blurry monocular videos. We propose the Blur-adaptive Latent Camera Estimation (BLCE) and Latent Camera-induced Exposure Estimation (LCEE) methods to accurately model scene blurriness. Extensive experiments on both real-world and synthetic blurry video datasets demonstrate that MoBGS significantly outperforms recent SOTA methods in dynamic deblurring NVS.

Acknowledgments

This work was supported by Institute of Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korean Government [Ministry of Science and ICT (Information and Communications Technology)] (Project Number: RS-2022-00144444, Project Title: Deep Learning Based Visual Representational Learning and Rendering of Static and Dynamic Scenes, 100%).

References

Attal, B.; Huang, J.-B.; Richardt, C.; Zollhoefer, M.; Kopf, J.; O’Toole, M.; and Kim, C. 2023. HyperReel: High-fidelity 6-DoF video with ray-conditioned sampling. In *CVPR*, 16610–16620.

Bae, J.; Kim, S.; Yun, Y.; Lee, H.; Bang, G.; and Uh, Y. 2024. Per-gaussian embedding-based deformation for deformable

3d gaussian splatting. In *European Conference on Computer Vision*, 321–335. Springer.

Bahat, Y.; Efrat, N.; and Irani, M. 2017. Non-uniform blind deblurring by reblurring. In *ICCV*, 3286–3294.

Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 5470–5479.

Broxton, M.; Flynn, J.; Overbeck, R.; Erickson, D.; Hedman, P.; Duvall, M.; Dourgarian, J.; Busch, J.; Whalen, M.; and Debevec, P. 2020. Immersive light field video with a layered mesh representation. *ACM Transactions on Graphics (TOG)*, 39(4): 86–1.

Bui, M.-Q. V.; Park, J.; Oh, J.; and Kim, M. 2025. MoBluRF: Motion Deblurring Neural Radiance Fields for Blurry Monocular Video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Cao, A.; and Johnson, J. 2023. Hexplane: A fast representation for dynamic scenes. In *CVPR*, 130–141.

Chen, J.; and Clark, R. 2025. Image as an imu: Estimating camera motion from a single motion-blurred image. *arXiv preprint arXiv:2503.17358*.

Chen, R. T.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural ordinary differential equations. *Advances in neural information processing systems*, 31.

Chu, M.; Xie, Y.; Mayer, J.; Leal-Taixé, L.; and Thurey, N. 2020. Learning temporal coherence via self-supervision for GAN-based video generation. *ACM Transactions on Graphics (TOG)*, 39(4): 75–1.

Collet, A.; Chuang, M.; Sweeney, P.; Gillett, D.; Evseev, D.; Calabrese, D.; Hoppe, H.; Kirk, A.; and Sullivan, S. 2015. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics (ToG)*, 34(4): 1–13.

Doersch, C.; Luc, P.; Yang, Y.; Gokay, D.; Koppula, S.; Gupta, A.; Heyward, J.; Rocco, I.; Goroshin, R.; Carreira, J.; et al. 2024. Bootstap: Bootstrapped training for tracking-any-point. In *Proceedings of the Asian Conference on Computer Vision*, 3257–3274.

Du, Y.; Zhang, Y.; Yu, H.-X.; Tenenbaum, J. B.; and Wu, J. 2021. Neural radiance flow for 4d view synthesis and video processing. In *ICCV*, 14304–14314. IEEE Computer Society.

Fang, Z.; Wu, F.; Huang, T.; Dong, L.; Dong, W.; Li, X.; and Shi, G. 2025. Parameterized Blur Kernel Prior Learning for Local Motion Deblurring. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 23006–23015.

Fridovich-Keil, S.; Meanti, G.; Warburg, F. R.; Recht, B.; and Kanazawa, A. 2023. K-planes: Explicit radiance fields in space, time, and appearance. In *CVPR*, 12479–12488.

Gao, C.; Saraf, A.; Kopf, J.; and Huang, J.-B. 2021. Dynamic View Synthesis from Dynamic Monocular Video. In *ICCV*.

Huang, Y.-H.; Sun, Y.-T.; Yang, Z.; Lyu, X.; Cao, Y.-P.; and Qi, X. 2024. SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes. *CVPR*.

- Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; and Yang, F. 2021. Musiq: Multi-scale image quality transformer. In *ICCV*, 5148–5157.
- Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4): 139–1.
- Kwak, S.; Kim, J.; Jeong, J. Y.; Cheong, W.-S.; Oh, J.; and Kim, M. 2025. MoDec-GS: Global-to-Local Motion Decomposition and Temporal Interval Adjustment for Compact Dynamic 3D Gaussian Splatting. In *CVPR*.
- Lee, B.; Lee, H.; Sun, X.; Ali, U.; and Park, E. 2024a. Deblurring 3d gaussian splatting. In *European Conference on Computer Vision*, 127–143. Springer.
- Lee, D.; Lee, M.; Shin, C.; and Lee, S. 2023a. DP-NeRF: Deblurred Neural Radiance Field With Physical Scene Priors. In *CVPR*, 12386–12396.
- Lee, D.; Oh, J.; Rim, J.; Cho, S.; and Lee, K. M. 2023b. ExBluRF: Efficient Radiance Fields for Extreme Motion Blurred Images. In *ICCV*, 17639–17648.
- Lee, J.; Kim, D.; Lee, D.; Cho, S.; Lee, M.; and Lee, S. 2024b. CRiM-GS: Continuous Rigid Motion-Aware Gaussian Splatting from Motion-Blurred Images. arXiv:2407.03923.
- Lee, J.; Lee, D.; Lee, M.; Kim, D.; and Lee, S. 2024c. Smurf: Continuous dynamics for motion-deblurring radiance fields. *arXiv preprint arXiv:2403.07547*.
- Lei, J.; Weng, Y.; Harley, A.; Guibas, L.; and Daniilidis, K. 2024. MoSca: Dynamic Gaussian Fusion from Casual Videos via 4D Motion Scaffolds. *arXiv*.
- Li, D.; Shi, X.; Zhang, Y.; Cheung, K. C.; See, S.; Wang, X.; Qin, H.; and Li, H. 2023a. A Simple Baseline for Video Restoration With Grouped Spatial-Temporal Shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9822–9832.
- Li, T.; Slavcheva, M.; Zollhoefer, M.; Green, S.; Lassner, C.; Kim, C.; Schmidt, T.; Lovegrove, S.; Goesele, M.; Newcombe, R.; and Lv, Z. 2022. Neural 3D Video Synthesis From Multi-View Video. In *CVPR*, 5521–5531.
- Li, Z.; Niklaus, S.; Snavely, N.; and Wang, O. 2021. Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes. In *CVPR*.
- Li, Z.; Wang, Q.; Cole, F.; Tucker, R.; and Snavely, N. 2023b. Dynibar: Neural dynamic image-based rendering. In *CVPR*, 4273–4284.
- Liu, R.; Li, Z.; and Jia, J. 2008. Image partial blur detection and classification. In *2008 IEEE conference on computer vision and pattern recognition*, 1–8. IEEE.
- Liu, Y.; Gao, C.; Meuleman, A.; Tseng, H.; Saraf, A.; Kim, C.; Chuang, Y.; Kopf, J.; and Huang, J. 2023. Robust Dynamic Radiance Fields. In *CVPR*, 13–23.
- Luthra, A.; Gantha, S. S.; Song, X.; Yu, H.; Lin, Z.; and Peng, L. 2024. Deblur-NSFF: Neural Scene Flow Fields for Blurry Dynamic Scenes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3658–3667.
- Ma, L.; Li, X.; Liao, J.; Zhang, Q.; Wang, X.; Wang, J.; and Sander, P. V. 2022. Deblur-nerf: Neural radiance fields from blurry images. In *CVPR*, 12861–12870.
- Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.
- Oswald, M. R.; Stühmer, J.; and Cremers, D. 2014. Generalized connectivity constraints for spatio-temporal 3d reconstruction. In *ECCV*, 32–46. Springer.
- Pan, J.; Sun, D.; Pfister, H.; and Yang, M.-H. 2016. Blind image deblurring using dark channel prior. In *CVPR*, 1628–1636.
- Park, J.; Bui, M.-Q. V.; Bello, J. L. G.; Moon, J.; Oh, J.; and Kim, M. 2025. SplineGS: Robust Motion-Adaptive Spline for Real-Time Dynamic 3D Gaussians from Monocular Video. In *CVPR*.
- Park, K.; Sinha, U.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Seitz, S. M.; and Martin-Brualla, R. 2021a. Nerfies: Deformable neural radiance fields. In *ICCV*, 5865–5874.
- Park, K.; Sinha, U.; Hedman, P.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Martin-Brualla, R.; and Seitz, S. M. 2021b. HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields. *ACM Trans. Graph.*, 40(6).
- Park, S.; Son, M.; Jang, S.; Ahn, Y. C.; Kim, J.-Y.; and Kang, N. 2023. Temporal Interpolation Is All You Need for Dynamic Neural Radiance Fields. In *CVPR*, 4212–4221.
- Peng, C.; Tang, Y.; Zhou, Y.; Wang, N.; Liu, X.; Li, D.; and Chellappa, R. 2025. BAGS: Blur agnostic Gaussian splatting through multi-scale kernel modeling. In *European Conference on Computer Vision*, 293–310. Springer.
- Piccinelli, L.; Yang, Y.-H.; Sakaridis, C.; Segu, M.; Li, S.; Van Gool, L.; and Yu, F. 2024. UniDepth: Universal Monocular Metric Depth Estimation. In *CVPR*.
- Pont-Tuset, J.; Perazzi, F.; Caelles, S.; Arbeláez, P.; Sorkine-Hornung, A.; and Gool, L. V. 2018. The 2017 DAVIS Challenge on Video Object Segmentation. *arXiv*.
- Pumarola, A.; Corona, E.; Pons-Moll, G.; and Moreno-Noguer, F. 2021. D-nerf: Neural radiance fields for dynamic scenes. In *CVPR*, 10318–10327.
- Shang, W.; Ren, D.; Yang, Y.; Zhang, H.; Ma, K.; and Zuo, W. 2023. Joint video multi-frame interpolation and deblurring under unknown exposure time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13935–13944.
- Shao, R.; Zheng, Z.; Tu, H.; Liu, B.; Zhang, H.; and Liu, Y. 2023. Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering. In *CVPR*, 16632–16642.
- Shi, J.; Xu, L.; and Jia, J. 2014. Discriminative blur detection features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2965–2972.
- Shi, J.; Xu, L.; and Jia, J. 2015. Just noticeable defocus blur detection and estimation. In *Proceedings of the IEEE*

- Conference on Computer Vision and Pattern Recognition*, 657–665.
- Song, L.; Chen, A.; Li, Z.; Chen, Z.; Chen, L.; Yuan, J.; Xu, Y.; and Geiger, A. 2023. Nerfplayer: A streamable dynamic scene representation with decomposed neural radiance fields. *IEEE Transactions on Visualization and Computer Graphics*, 29(5): 2732–2742.
- Sun, H.; Li, X.; Shen, L.; Ye, X.; Xian, K.; and Cao, Z. 2024. DyBluRF: Dynamic Neural Radiance Fields from Blurry Monocular Video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7517–7527.
- Tretschk, E.; Tewari, A.; Golyanik, V.; Zollhöfer, M.; Lassner, C.; and Theobalt, C. 2021. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *ICCV*, 12959–12970.
- Tucker, R.; and Snavely, N. 2020. Single-View View Synthesis with Multiplane Images. In *CVPR*.
- Wang, L.; Zhang, J.; Liu, X.; Zhao, F.; Zhang, Y.; Zhang, Y.; Wu, M.; Yu, J.; and Xu, L. 2022. Fourier PlenOctrees for Dynamic Radiance Field Rendering in Real-time. In *CVPR*, 13524–13534.
- Wang, P.; Zhao, L.; Ma, R.; and Liu, P. 2023. BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields. In *CVPR*, 4170–4179.
- Wang, Q.; Ye, V.; Gao, H.; Austin, J.; Li, Z.; and Kanazawa, A. 2024. Shape of motion: 4d reconstruction from a single video. *arXiv preprint arXiv:2407.13764*.
- Weng, C.-Y.; Curless, B.; Srinivasan, P. P.; Barron, J. T.; and Kemelmacher-Shlizerman, I. 2022. Humannerf: Free-viewpoint rendering of moving people from monocular video. In *CVPR*, 16210–16220.
- Weng, W.; Zhang, Y.; and Xiong, Z. 2023. Event-based blurry frame interpolation under blind exposure. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1588–1598.
- Wu, G.; Yi, T.; Fang, J.; Xie, L.; Zhang, X.; Wei, W.; Liu, W.; Tian, Q.; and Wang, X. 2024a. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. In *CVPR*.
- Wu, G.; Yi, T.; Fang, J.; Xie, L.; Zhang, X.; Wei, W.; Liu, W.; Tian, Q.; and Wang, X. 2024b. 4d gaussian splatting for real-time dynamic scene rendering. In *CVPR*.
- Wu, R.; Zhang, Z.; Chen, M.; Fan, X.; Yan, Z.; and Zuo, W. 2024c. Deblur4DGS: 4D Gaussian Splatting from Blurry Monocular Video. *arXiv preprint arXiv:2412.06424*.
- Yan, R.; and Shao, L. 2016. Blind image blur estimation via deep learning. *IEEE Transactions on Image Processing*, 25(4): 1910–1921.
- Yang, Z.; Gao, X.; Zhou, W.; Jiao, S.; Zhang, Y.; and Jin, X. 2024. Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction. In *CVPR*.
- Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.
- Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Liu, W.; and Li, H. 2018a. Adversarial spatio-temporal learning for video deblurring. *IEEE Transactions on Image Processing*, 28(1): 291–301.
- Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Stenger, B.; Liu, W.; and Li, H. 2020. Deblurring by realistic blurring. In *CVPR*, 2737–2746.
- Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018b. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 586–595.
- Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018c. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.
- Zhao, L.; Wang, P.; and Liu, P. 2024. Bad-gaussians: Bundle adjusted deblur gaussian splatting. In *European Conference on Computer Vision*, 233–250. Springer.