GeMS: Efficient Gaussian Splatting for Extreme Motion Blur

Anonymous authors
**Paper under double-blind review**

Abstract

We introduce *GeMS*, a framework for 3D Gaussian Splatting designed to handle severely motion-blurred images. State-of-the-art deblurring method for extreme motion blur, such as ExBluRF, as well as Gaussian Splatting-based approaches like Deblur-GS, typically assume access to corresponding sharp images for camera pose estimation and point cloud generation, which is an unrealistic assumption. Additionally, methods relying on COLMAP initialization, such as BAD-Gaussians, fail due to the lack of reliable feature correspondences in cases of severe motion blur. To address these challenges, we propose *GeMS*, a 3D Gaussian Splatting framework that reconstructs scenes directly from extremely motion-blurred images. GeMS integrates: (1) *VGGSfM*, a deep learning-based SfM pipeline which estimates camera poses and generates point clouds directly from severely motion-blurred images; (2) *MCMC*-based Gaussian Splatting, which enables robust scene initialization by treating Gaussians as samples from an underlying probability distribution, eliminating heuristic densification and pruning strategies; and (3) Joint optimization of camera motion trajectory and Gaussian parameters which ensures stable and accurate reconstruction. While this pipeline produces reasonable reconstructions, extreme motion blur can still introduce inaccuracies, especially when all input views are severely blurred. To address this, we propose *GeMS-E*, which integrates a progressive refinement step when event data is available. Specifically, we perform (4) *Event-based Double Integral (EDI)* deblurring, which first restores deblurred images from motion-blurred inputs. These deblurred images are then fed into the GeMS framework, leading to improved pose estimation, point cloud generation, and hence overall reconstruction quality. Both GeMS & GeMS-E achieve state-of-the-art performance on synthetic as well as real-world datasets, demonstrating their effectiveness in handling extreme motion blur. To the best of our knowledge, we are the first to effectively address this problem in extreme blur scenarios within a 3D Gaussian Splatting framework, without requiring sharp images for SfM (pose and point cloud) initialization.

1 Introduction

Motion blur poses a significant challenge in 3D scene reconstruction and novel view synthesis, particularly in real-world scenarios involving high-speed camera motion or low-light conditions. Recent advancements in Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS) have greatly improved the quality and efficiency of 3D reconstruction. However, these methods rely on sharp images and accurate camera poses, making them vulnerable to severe motion blur. Traditional Structure-from-Motion (SfM) pipelines, such as COLMAP Schonberger & Frahm (2016), fail under heavy blur due to unreliable feature correspondences, leading to degraded pose estimation and poor reconstruction quality.

NeRF-based deblurring methods, including Deblur-NeRF Ma et al. (2022), and voxel-based radiance field approaches like ExBluRF Lee et al. (2023), attempt to model the motion blur process and recover sharp scene representations. However, they assume precise camera poses, which are difficult to obtain in real-world cases. BAD-NeRF Wang et al. (2023) mitigates this issue by jointly optimizing camera trajectories and radiance fields, but its implicit volumetric representation results in slow training and rendering speeds, limiting its practical usability.
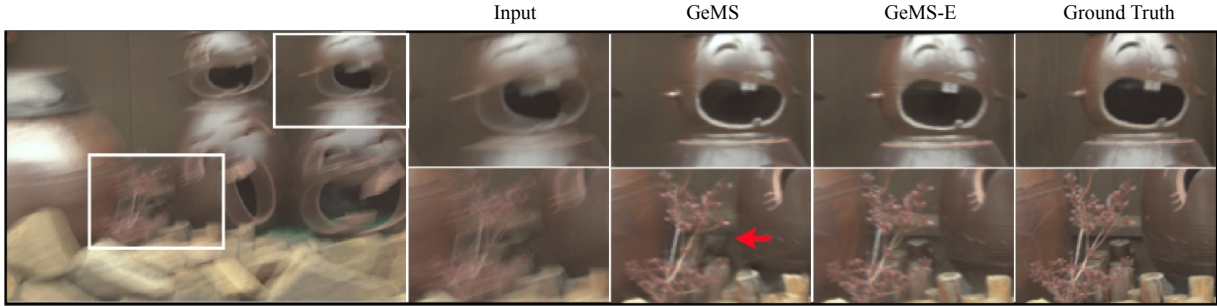
Figure 1: Overview: GeMS reconstructs sharp 3D scenes directly from extremely motion-blurred images by optimizing Gaussian Splatting without relying on COLMAP or sharp image supervision. When event data is available, GeMS-E leverages event streams to further enhance reconstruction quality, especially when all input views are severely blurred.

Recently, explicit representations like 3D Gaussian Splatting have gained traction due to their superior efficiency and rendering speed. BAD-Gaussians Zhao et al. (2024) and Deblur-GS Chen & Liu (2024) extend 3DGS to motion-blurred images by optimizing camera motion and Gaussian parameters simultaneously. However, these approaches still rely on COLMAP for SfM initialization, making them ineffective in extreme motion blur scenarios where traditional feature-based pose estimation fails.

Recently, several SfM pipelines have been proposed that demonstrate improved robustness and accuracy compared to COLMAP. However, existing methods like GLOMAP Pan et al. (2024) and HLOC Sarlin et al. (2019) still rely on pairwise feature correspondences, making them vulnerable to extreme motion blur. Pixel-Perfect SfM Lindenberger et al. (2021) improves geometric accuracy by optimizing feature metric error using dense deep features, but it remains dependent on keypoint refinement. VGGSfM Wang et al. (2024) addresses these limitations by replacing keypoint matching with deep 2D point tracking, eliminating the reliance on pairwise correspondences. By jointly optimizing camera poses and 3D points through a differentiable bundle adjustment layer, it achieves robust performance across diverse datasets. While not explicitly designed for extreme motion blur, its deep point tracking and end-to-end optimization make it well-suited for our COLMAP-free 3DGS initialization in severely blurred scenarios.

Building on this foundation, we propose *GeMS*, a robust *COLMAP-free 3D Gaussian Splatting framework* for reconstructing 3D scenes directly from motion-blurred images. Unlike prior methods, GeMS avoids the limitations of COLMAP by leveraging *VGGSfM* Wang et al. (2024) for camera pose estimation and point cloud generation. Although VGGSfM provides a reliable initialization, the resulting point clouds tend to be coarse and sparse under severe motion blur. To address this limitation, we incorporate the *3DGS-MCMC* method proposed by Kheradmand et al. (2025) into our pipeline. This probabilistic formulation models Gaussians as samples from a distribution and employs MCMC-based sampling to adaptively initialize and densify the scene, eliminating the need for heuristic densification and improving early training stability. However, 3DGS-MCMC only accounts for inaccuracies in the point cloud and does not correct errors in camera poses, errors which under extreme blur, can substantially degrade reconstruction quality. To mitigate this, we perform joint optimization Zhao et al. (2025) of camera poses and Gaussian parameters during training. This joint refinement is essential for correcting the coarse initialization from VGGSfM and leads to significantly improved geometric consistency and photometric accuracy. These components work in synergy to make GeMS robust against the challenges of severe motion blur, enabling accurate and consistent deblurring and novel view synthesis across diverse scenes and blur conditions.

While GeMS significantly enhances reconstruction quality, extreme motion blur can still lead to pose inaccuracies and scene inconsistencies, especially when all input views are severely blurred. When event data is available, we address this limitation by introducing GeMS-E, which incorporates event-based deblurring into our framework. Event cameras, which capture intensity changes with high temporal resolution, are well-suited for deblurring fast-motion scenes. Several NeRF-based event-driven methods, such as E2NeRF Qi et al. (2023) and EBAD-NeRF Qi et al. (2024), have explored event-based 3D reconstruction. However, they suffer from the inefficiencies of NeRF's implicit representation. More recently, E2GS Deguchi et al. (2024)
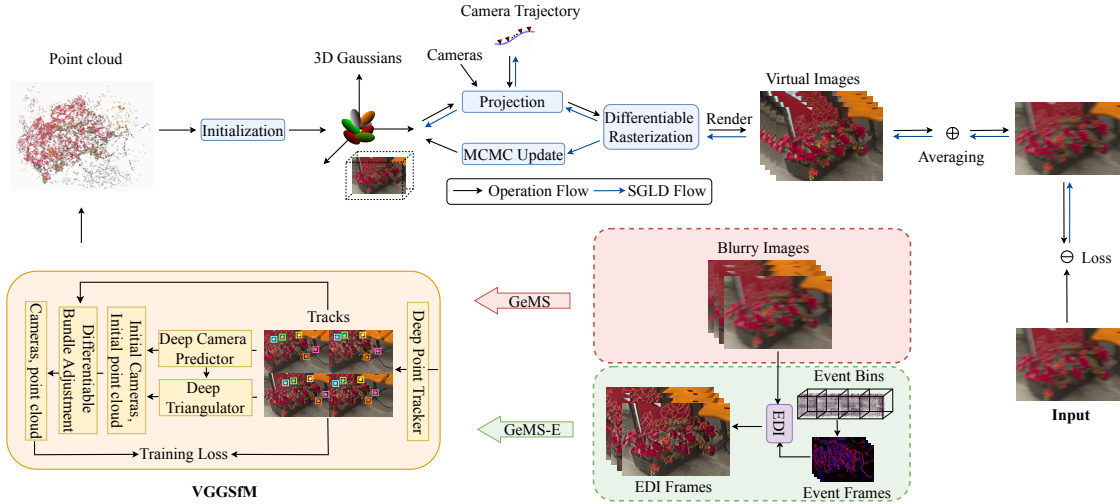
Figure 2: Our Method: *GeMS* addresses extreme motion blur in 3D Gaussian Splatting framework. GeMS directly optimizes Gaussians on blurred images without requiring COLMAP or sharp supervision, leveraging VGGSfM for robust SfM initialization, 3DGS-MCMC with joint optimization for effective refinement of camera poses and Gaussian parameters. When event data is available, GeMS-E enhances this by incorporating the EDI model to recover deblurred images, which are then fed into the GeMS pipeline for improved reconstructions.

has attempted to integrate event streams into 3D Gaussian Splatting, but it fails to deblur images effectively, introducing severe color artifacts. A detailed discussion of related works is provided in appendix A. In contrast, our approach seamlessly integrates event-based deblurring within GeMS. We leverage the Event-based Double Integral (EDI) model to reconstruct sharp images from event streams. These deblurred images are then processed through our COLMAP-free GeMS pipeline, ensuring robust and high-fidelity reconstructions even in extreme motion scenarios. By combining event-driven deblurring with an efficient Gaussian Splatting framework, GeMS-E surpasses existing methods in both reconstruction quality and computational efficiency, when event data is available. Our GeMs/GeMS-E results are illustrated in Figure 1.

Our key contributions are as follows:

- We propose GeMS, an efficient COLMAP-free 3D Gaussian Splatting framework that enhances robustness in extreme motion-blurred scenarios by leveraging VGGSfM for camera pose estimation and point cloud generation without sharp image supervision. The framework further incorporates MCMC-based Gaussian initialization and optimization, along with joint optimization of camera poses and Gaussian parameters, to effectively restore sharp scene representations.

- When event data is available, GeMS-E integrates event-based deblurring using the EDI model, enabling sharp image reconstruction and novel sharp view synthesis even in extreme motion blur conditions, particularly in cases where all input views are severely blurred. To further support research in this area, we introduce a complementary synthetic event dataset *EveGeMS* specifically curated for extreme blur scenarios.

- Through extensive evaluations on both synthetic and real-world datasets, we demonstrate superior performance in deblurring and novel view synthesis, achieving state-of-the-art reconstruction quality while significantly improving computational and memory efficiency.

To the best of our knowledge, this is the first work to effectively address the challenge of extreme motion blur within a 3D Gaussian Splatting framework, without relying on sharp images for pose and point cloud initialization.

2 Method

GeMS is a COLMAP-free 3D Gaussian Splatting framework for deblurring and sharp novel view synthesis under extreme motion blur. It leverages VGGSfM Wang et al. (2024) for camera pose estimation and point cloud generation, followed by MCMC-based Gaussian initialization and optimization Kheradmand et al. (2025). A joint optimization of camera trajectories and Gaussian parameters further refines the reconstruction, ensuring robustness in highly blurred scenarios. GeMS-E (Event-Assisted Deblurring): When event data is available, we first recover sharp images using the EDI model Pan et al. (2019). These deblurred images are then processed through our GeMS framework to facilitate both motion deblurring and sharp novel view synthesis, even in the presence of extreme motion blur across all input images. Our overall framework is illustrated in Figure 2. Background of 3D Gaussian Splatting is presented in appendix B.1.

2.1 GeMS (Direct Deblurring)

COLMAP Schonberger & Frahm (2016) is the de facto choice for structure-from-motion (SfM) initialization in NeRF and 3D Gaussian Splatting (3DGS) pipelines. However, it requires sharp, high-quality images for reliable feature matching, rendering it ineffective under severe motion blur. Consequently, state-of-the-art deblurring methods that rely on COLMAP become ineffective in such challenging scenarios. In contrast, our proposed GeMS framework eliminates dependence on COLMAP by employing a robust and efficient Gaussian Splatting pipeline that integrates VGGSfM with 3DGS-MCMC based optimization. We then jointly optimize Gaussian parameters and camera motion trajectories to refine pose inaccuracies. This systematic approach enables accurate and sharp 3D scene reconstruction directly from severely motion-blurred images.

2.1.1 Robust Initialization and Optimization with VGGSfM and 3DGS-MCMC

Accurate initialization is critical for 3D Gaussian Splatting (3DGS), particularly in scenarios with severe motion blur where traditional methods like COLMAP fail due to unreliable feature correspondences. This limitation affects several SOTA deblurring pipelines such as ExBluRF Lee et al. (2023) and BAD-Gaussians Zhao et al. (2024) which depend on COLMAP initialization. To address this, we integrate two complementary techniques: (1) **VGGSfM**, a fully differentiable deep-learning-based Structure-from-Motion (SfM) pipeline for robust camera pose estimation and point cloud generation. (2) **MCMC based Gaussian Splatting with joint optimization**, which treats Gaussians as probabilistic samples, eliminating heuristic densification and pruning strategies for more adaptive initialization together optimizing for camera poses.

VGGSfM for SfM initialization: Unlike traditional SfM pipelines that rely on incremental image registration and non-differentiable optimization, VGGSfM estimates all camera poses in an end-to-end manner. It builds on recent advances in deep 2D point tracking to extract reliable, pixel-accurate tracks without requiring explicit feature matching between image pairs. Instead of gradual camera registration, VGGSfM employs a Transformer-based model for global pose estimation, improving robustness in motion-blurred conditions. Furthermore, it utilizes a differentiable bundle adjustment layer based on the Theseus solver, seamlessly integrating into learning-based frameworks. By incorporating VGGSfM, we eliminate the need for sharp images in initialization, significantly enhancing stability and accuracy in pose estimation.

3DGS-MCMC for joint optimization: Following 3DGS-MCMC Kheradmand et al. (2025), we reinterpret 3D Gaussian splatting as a probabilistic sampling process using Markov Chain Monte Carlo (MCMC) techniques. Instead of treating Gaussians as fixed entities requiring heuristic adjustments, MCMC considers them as samples drawn from an underlying probability distribution of the scene. The updates to Gaussian parameters are formulated using Stochastic Gradient Langevin Dynamics (SGLD), which naturally aligns with MCMC sampling. This approach allows Gaussians to relocate dynamically, removing the need for explicit cloning while maintaining statistical consistency. To further address pose inaccuracies introduced by severe motion blur, we also incorporate joint optimization of Gaussian parameters and camera poses within the 3DGS-MCMC framework. This enables consistent refinement of both scene geometry and motion, improving overall reconstruction accuracy under challenging severely motion blurred conditions.

Synergistic integration of VGGSfM and 3DGS-MCMC joint optimization: By combining VG-GSfM and 3DGS-MCMC joint optimization, our framework achieves robust initialization and optimization without requiring sharp images. This integration significantly improves the stability, accuracy, and adaptability of 3DGS, making it well-suited for severe motion blurred scenarios.

2.1.2 Camera Motion Trajectory Modeling in 3DGS-MCMC for Pose optimization

To model camera motion during exposure, we parameterize the pose in $SE(3)$ and approximate its trajectory in the Lie algebra of $SE(3)$. While BAD-Gaussians Zhao et al. (2024) employ *linear* and *cubic spline* interpolation for trajectory estimation, these methods prove inadequate for handling severe motion blur. Hence our method adopts *Bézier curve interpolation* inspired from Chen & Liu (2024), which provides a smoother and more accurate representation. Given a Bézier curve of degree M, the camera motion is represented using $M+1$ control points T_j ($j = 0, ..., M$). The interpolated camera pose at time t is given as:

$$T_t = \prod_{j=0}^{M} \exp\left(\binom{M}{j}(1-u)^{M-j}u^j \cdot \log(T_j)\right) \tag{1}$$

where $u = t/\tau \in [0, 1]$ and τ is the exposure time. This formulation ensures smooth motion trajectory estimation while remaining differentiable, enabling joint optimization for accurate deblurring.

2.1.3 Loss Functions

Given a set of K motion-blurred images, the goal is to jointly estimate the camera motion trajectory for each image and the learnable parameters of 3DGS, $\boldsymbol{\theta}$ (i.e., mean position $\boldsymbol{\mu}$, 3D covariance $\boldsymbol{\Sigma}$, opacity $\mathbf{o}$, and color $\mathbf{c}$). For this joint estimation framework, we draw inspiration from BAD-Gaussians Zhao et al. (2024). This estimation is achieved by minimizing the following loss function, which combines an $\mathcal{L}_1$ loss with a D-SSIM term. The loss is computed between $\mathbf{B}_k(\mathbf{u})$, the kth synthesized blurry image generated via 3DGS and its corresponding real captured counterpart, $\mathbf{B}_k^{gt}(\mathbf{u})$. The blur image formation model and joint optimization method are presented in appendix B.2 and appendix C.2 respectively.

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{\text{D-SSIM}} \tag{2}$$

2.2 GeMS-E (Event-Assisted Deblurring)

When event data is available, we further refine our pipeline, particularly in scenarios where all input views are severely blurred and optimization struggles to converge. In such extreme motion blur cases, event cameras offer a compelling alternative, as they asynchronously capture intensity changes with high temporal resolution, making them inherently robust to motion blur. To harness this advantage, we introduce an event-guided scene reconstruction framework that generates sharp images from event streams and feeds them into our proposed GeMS framework.

2.2.1 Event Generation

Unlike frame-based cameras that record pixel brightness at a fixed frame rate, event cameras asynchronously generate an event $\mathbf{e}(x, y, \tau, p)$ when the change in brightness of pixel (x, y) in the logarithmic domain exceeds a threshold Θ at time τ.

$$p_{x,y,\tau} = \begin{cases} -1, & \log(\mathcal{I}_{x,y,\tau}) - \log(\mathcal{I}_{x,y,\tau-\Delta\tau}) < -\Theta \\ +1, & \log(\mathcal{I}_{x,y,\tau}) - \log(\mathcal{I}_{x,y,\tau-\Delta\tau}) > \Theta \end{cases} \tag{3}$$

where p denotes the direction of the brightness change, and $\mathcal{I}_{(x,y,\tau)}$ represents the brightness value of pixel (x, y) at time τ. Since events are generated asynchronously, they are typically grouped into b event bins, divided equally over time, to facilitate processing. Given a blurred image with exposure time from t_{start} to

t_{end} and the associated event data $\{\mathbf{e}_i\}_{t_{\text{start}} < \tau_i \leq t_{\text{end}}}$, we can generate $\{B_k\}_{k=1}^b$ as follows:

$$B_k = \{\mathbf{e}_i(x_i, y_i, \tau_i, p_i)\}_{t_{k-1} < \tau_i \leq t_k} \tag{4}$$

where $t_k = t_{\text{start}} + \frac{k}{b} t_{\text{exp}}$ is the time division point between bins and $t_{\text{exp}} = t_{\text{end}} - t_{\text{start}}$ represents the exposure time. The EDI deblurring method is detailed in appendix C.1.
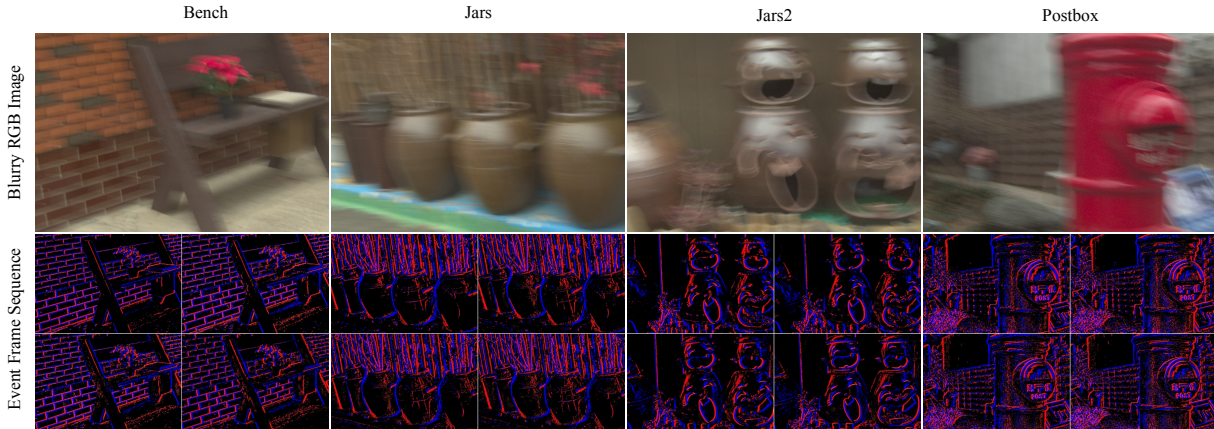
Figure 3: Our Complementary Event Dataset (EveGeMS): We present a synthetic event dataset designed for scenarios involving extreme motion blur. While the RGB frames suffer from severe blur, each is accompanied by a sequence of event frames that preserve fine structural and motion cues. This complementary information facilitates robust and accurate deblurring and novel view synthesis especially where all input views are severely blurred.

3 Experiments

3.1 Experimental Setup

Datasets: For evaluation, we use the synthetic dataset provided by ExBluRF Lee et al. (2023), which includes eight diverse outdoor scenes captured with challenging camera motion. Each scene contains 20 to 40 blurry views for training and 4 to 6 test views for evaluating novel view synthesis. Each blurry image in ExBluRF is paired with sequences of sharp images, which we utilize to create our complementary synthetic event dataset, *EveGeMS*, as shown in Figure 3. Specifically, the sharp frames recorded during the camera motion are processed using ESIM Rebecq et al. (2018) to generate the corresponding event stream, similar to the synthetic datasets in E2NeRF. Additionally, we use a real-world dataset from E2NeRF Qi et al. (2023), captured with the DAVIS346 color event camera. This dataset consists of five challenging scenes (i.e. letter, lego, camera, plant, and toys) with complex textures and varied motion, where RGB frames were captured with a 100ms exposure time, resulting in motion blur and complex camera trajectories. By incorporating event streams into the ExBluRF dataset, we contribute an extreme blurry and event pair dataset. We will release our complementary synthetic event dataset publicly upon acceptance to support future research in event-based deblurring and view synthesis under severe motion blur conditions. The complete event frame sequences of our EveGeMS dataset for all scenes are provided in the supplementary material.

Baseline Methods and Evaluation Metrics: Our baselines include: state-of-the-art deep learning-based single-image motion deblurring methods (MPRNet Zamir et al. (2021), Restormer Zamir et al. (2022)); event-based motion deblurring methods (EDI Pan et al. (2019), E2NeRF Qi et al. (2023), EBADNeRF Qi et al. (2024)); and motion deblurring method designed for extreme motion blur (ExBluRF* Lee et al. (2023)) and 3D Gaussian Splatting-based deblurring (BAD-Gaussians* Zhao et al. (2024)). Note that both ExBluRF* and BAD-Gaussians* rely on pose and point cloud initialization from sharp images, which is an unrealistic assumption. In real-world scenarios, obtaining pose and point cloud initializations directly from extremely blurred images is not possible for NeRF and 3DGS based deblurring methods that rely on COLMAP for initialization. Therefore, we exclude direct comparisons with ExBluRF* and BAD-Gaussians* on real dataset.

However, we consider event-based methods as EDI deblurred images from events enable COLMAP initialization even under severe motion blur. We provide details of the implementation details in (appendix D.1).

3.2 Quantitative Results

Reconstruction Quality: Our method achieves superior reconstruction quality, outperforming existing approaches in PSNR, SSIM, and LPIPS. Notably, GeMS provides an average **1 dB** improvement in PSNR over competing methods, as detailed in Table 1. A key limitation of competing methods like ExBluRF* and BAD-Gaussians* is their reliance on sharp image-based pose and point cloud initialization, which is not feasible in real-world scenarios with extreme motion blur. Hence, for a fair comparison with the state-of-the-art (SOTA) event-based method, E2NeRF, our approach GeMS-E achieves an average **2.5 dB** improvement in PSNR, as shown in Table 1, highlighting its effectiveness in leveraging motion-blurred inputs with event information. Even without event data, our method GeMS surpasses all competing event-based methods across all metrics, highlighting its robustness in extreme motion blur scenarios. The integration of event-based deblurring, VGGSfM based SfM initialization, 3DGS-MCMC joint optimization enables our GeMS-E to achieve high-fidelity reconstructions with significantly lower computational cost.

Table 1: Quantitative comparisons for novel sharp view synthesis (deblurring + novel view synthesis) on the Synthetic Dataset. Note that ExBluRF* and BAD-Gaussians* rely on SfM initialization from sharp images, as COLMAP fails on extremely blurred images. While their metric values are shown here for reference, they are excluded from the main rankings for fairness. The best and second-best results are highlighted in green and orange , respectively. For completeness, rankings that include ExBluRF* and BAD-Gaussians* are provided in appendix D.2.

| Scene | Metric | Single Image Deblurring | | Event-Based Deblurring | | | Sharp Image Supervision(*) | | Ours | |
		MPRNet	Restormer	EDI	E2NeRF	EBAD-NeRF	ExBluRF*	BAD-Gaussians*	GeMS	GeMS-E
Bench	PSNR↑	25.35	26.39	24.78	25.41	28.15	31.93	32.54	29.86	33.55
	SSIM↑	0.678	0.720	0.642	0.708	0.822	0.877	0.901	0.841	0.924
	LPIPS↓	0.425	0.356	0.366	0.438	0.172	0.111	0.046	0.118	0.063
Camellia	PSNR↑	24.84	25.14	20.65	28.07	24.33	28.02	28.83	28.56	29.47
	SSIM↑	0.669	0.690	0.560	0.721	0.743	0.715	0.815	0.821	0.873
	LPIPS↓	0.395	0.351	0.391	0.329	0.192	0.313	0.099	0.129	0.108
Dragon	PSNR↑	29.96	28.37	28.93	30.89	33.99	33.45	36.98	32.43	37.01
	SSIM↑	0.731	0.704	0.648	0.697	0.864	0.828	0.930	0.818	0.925
	LPIPS↓	0.454	0.465	0.380	0.433	0.202	0.180	0.045	0.171	0.069
Jars	PSNR↑	25.36	25.57	25.15	29.85	28.89	30.85	31.52	31.42	32.35
	SSIM↑	0.680	0.687	0.637	0.775	0.838	0.840	0.867	0.879	0.898
	LPIPS↓	0.406	0.371	0.385	0.334	0.198	0.156	0.078	0.127	0.108
Jars2	PSNR↑	24.33	26.43	24.09	27.71	27.39	30.89	28.94	28.14	28.79
	SSIM↑	0.745	0.814	0.678	0.770	0.863	0.860	0.851	0.873	0.906
	LPIPS↓	0.358	0.275	0.359	0.383	0.171	0.113	0.114	0.173	0.133
Postbox	PSNR↑	25.89	26.52	23.50	30.66	26.82	31.40	26.40	27.74	31.33
	SSIM↑	0.736	0.753	0.658	0.813	0.826	0.864	0.757	0.788	0.906
	LPIPS↓	0.318	0.286	0.334	0.262	0.151	0.095	0.123	0.15	0.070
Stone Lantern	PSNR↑	24.97	26.68	24.14	30.47	26.29	28.24	28.29	28.29	29.43
	SSIM↑	0.785	0.831	0.697	0.836	0.802	0.765	0.843	0.849	0.894
	LPIPS↓	0.342	0.280	0.346	0.324	0.264	0.236	0.143	0.195	0.152
Sunflowers	PSNR↑	28.86	29.55	27.16	31.74	30.98	34.46	34.06	29.47	33.69
	SSIM↑	0.837	0.847	0.775	0.850	0.903	0.920	0.942	0.854	0.938
	LPIPS↓	0.242	0.206	0.287	0.310	0.117	0.093	0.065	0.163	0.077
Average	PSNR↑	26.19	26.83	24.80	29.35	28.36	31.15	30.95	29.49	31.95
	SSIM↑	0.733	0.756	0.662	0.771	0.833	0.834	0.863	0.840	0.908
	LPIPS↓	0.368	0.324	0.356	0.352	0.183	0.162	0.089	0.153	0.097

Training Time and GPU Memory Consumption: We evaluate the efficiency of our approach against state-of-the-art event-based methods, EBAD-NeRF and E2NeRF, in terms of training time and GPU memory usage on real dataset. As shown in Table 3, our method significantly reduces training time, completing optimization in approximately **7 minutes** per scene, whereas EBAD-NeRF and E2NeRF require over 6

hours and 14 hours, respectively. Moreover, Table 4 highlights the GPU memory consumption across different scenes, where our approach requires only ∼**1.55 GiB** on an average, compared to the excessive demands of EBAD-NeRF (∼14.49 GiB) and E2NeRF (∼15.16 GiB). These results confirm the scalability and hardware efficiency of our method, making it practical for real-world applications.

3.3 Qualitative Results
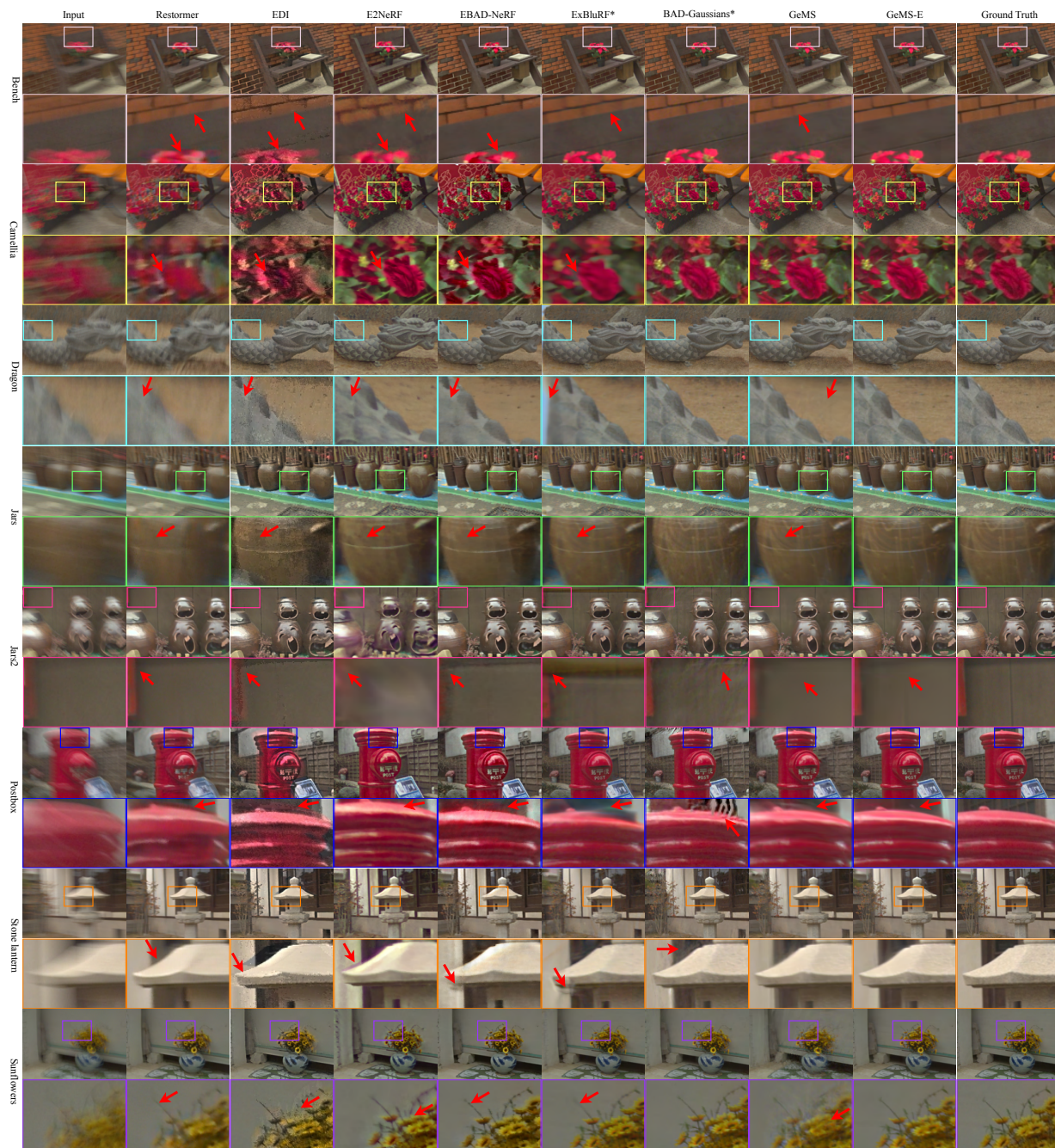
Figure 4: Results on the Synthetic Dataset: Our method effectively removes severe motion blur, reconstructing sharp results with high fidelity. Compared to existing approaches, it better preserves fine details and structural consistency while reducing color artifacts, demonstrating robustness under extreme blur conditions. Note that ExBluRF* and BAD-Gaussians* rely on pose and point cloud initializations from sharp images.
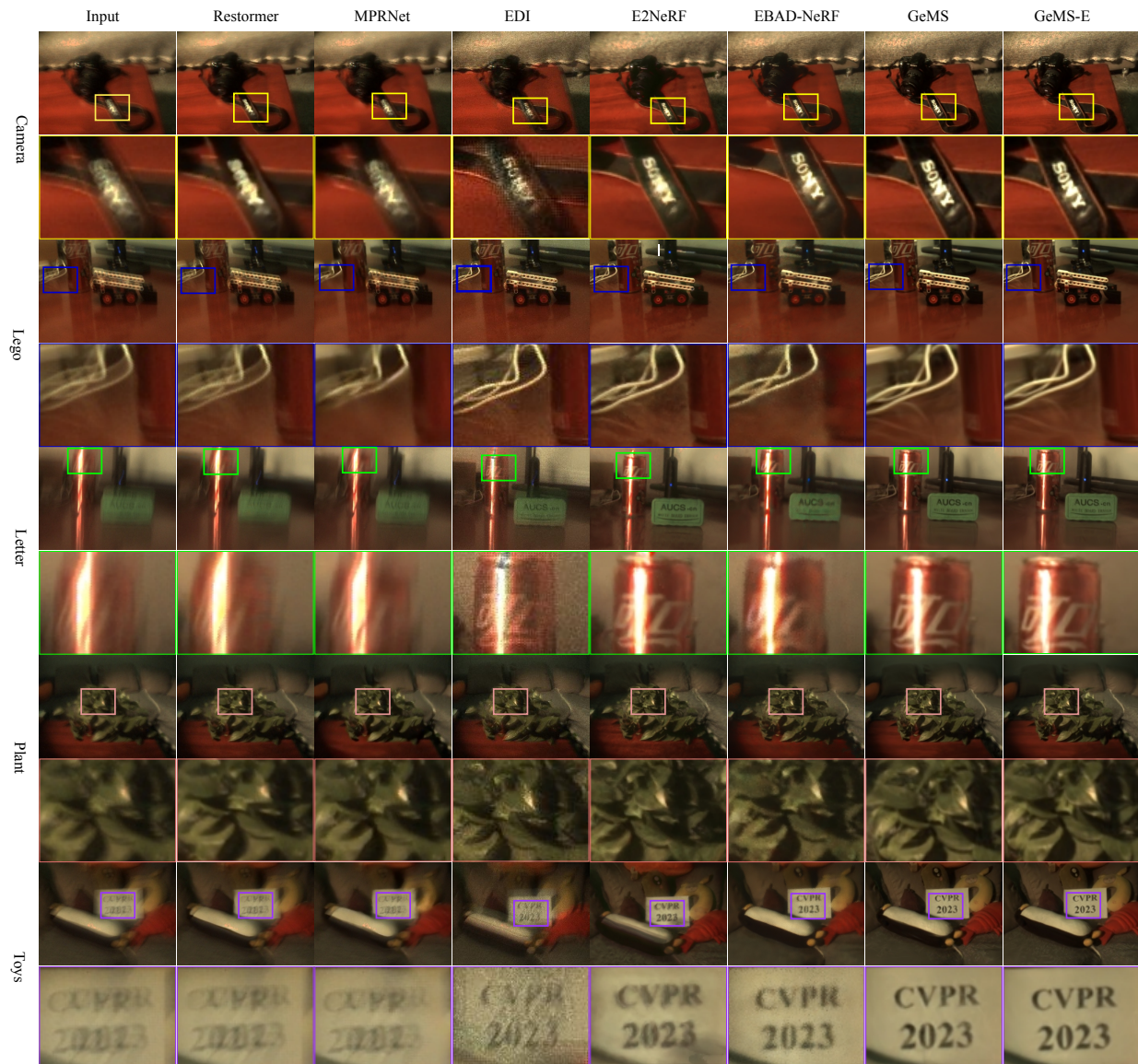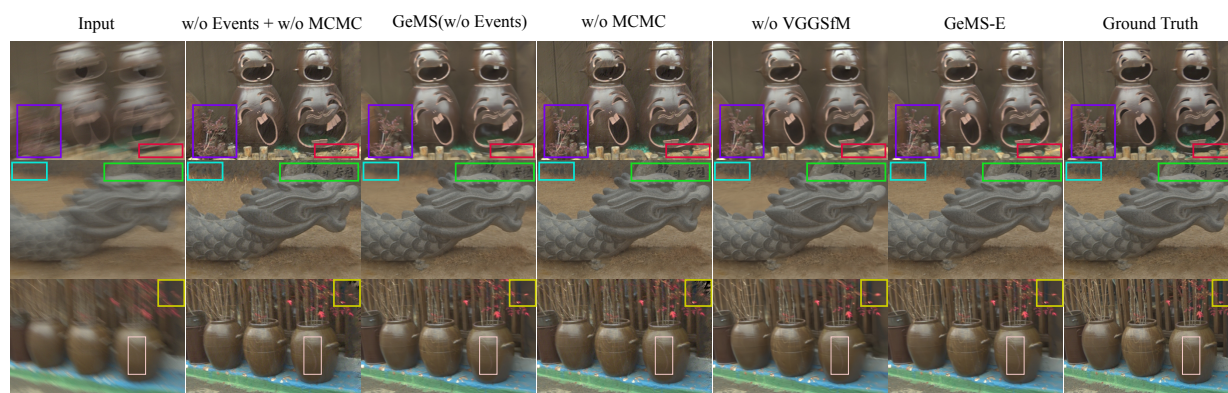
Figure 5: Results on the Real Dataset: Our method reconstructs sharp and high-quality images from severely motion-blurred real-world inputs. In contrast, existing methods struggle with artifacts, noise, loss of fine details, and text degradation. Our framework effectively restores textures and structural consistency, as evident in the insets.

We present qualitative comparisons on both synthetic and real datasets in Figure 4 and Figure 5, respectively. Our method consistently outperforms existing approaches, producing sharper reconstructions with fewer artifacts and improved texture details. In synthetic scenes such as *Stone Lantern* and *Jars*, competing methods struggle to recover fine structures, often leading to over-smoothed or distorted reconstructions, as indicated by the red arrows in Figure 4. In contrast, our approach faithfully restores object details and edges. In particular, in extremely blurred scenes like *Sunflowers* where all the input views are severely blurred, leveraging event-based deblurring further enhances reconstruction quality by recovering sharp details. Similarly, in real-world datasets as shown in Figure 5, our method achieves superior clarity, particularly in challenging regions with high-frequency textures, such as text on the *CVPR 2023* poster or specular highlights on metallic surfaces. Unlike prior event-based methods, such as E2NeRF and EBAD-NeRF, which introduce color distortions or fail to fully remove motion blur, our approach effectively preserves accurate color distributions while restoring sharp details. Moreover, our method demonstrates improved robustness

9

in handling fine-grained textures, such as plant leaves and intricate object boundaries, where others exhibit blurring or ghosting artifacts. This advantage is evident in diverse scenes, reinforcing the effectiveness of our framework. Additional qualitative results of our methods, GeMS and GeMS-E, are presented in appendix E. The supplementary material provides novel sharp free-viewpoint renderings for both synthetic and real-world scenes.

3.4 Ablations

Component-wise Analysis. To quantify the contribution of each module in GeMS / GeMS-E, we perform ablation experiments on the synthetic dataset. The results are presented in Table 2 for the quantitative evaluation of novel sharp view synthesis (deblurring + novel view synthesis). Details of the abbreviations used for the ablation methods and their corresponding deblurring results are provided in appendix D.3. Additionally, Figure 6 and Figure 7 provide qualitative comparisons of deblurring on synthetic and real datasets, respectively, further highlighting the effectiveness of our approach with inclusion of each module.

Figure 6: Ablations on different modules of our framework on the Synthetic Dataset: Our method (GeMS) achieves strong deblurring with MCMC helping to reduce artifacts and enhance reconstruction quality. With events (GeMS-E), the results become even sharper and more detailed, demonstrating the added benefit of event information.
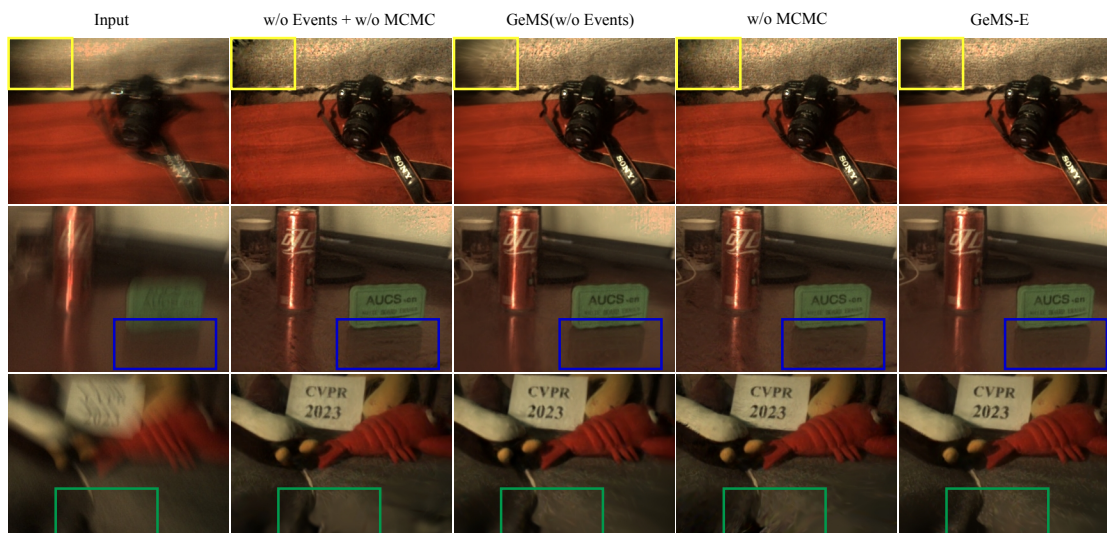
Figure 7: Ablation study on different modules of our framework on the Real Dataset: GeMS produces high-quality reconstructions, with MCMC effectively suppressing artifacts. Incorporating event data in GeMS-E further refines the results, demonstrating the effectiveness of our approach across different blur settings.

Table 2: Ablation study for novel sharp view synthesis (deblurring + novel view synthesis) on the synthetic dataset. We evaluate the impact of different components, VGGSfM, MCMC and EDI on novel view synthesis performance. The inclusion of each module leads to considerable improvements in PSNR, SSIM, and LPIPS, with their combination yielding the best overall results, demonstrating the synergistic benefits of our approach. Abbreviations used for the ablation methods are outlined in appendix D.3.

Method	Metric	Bench	Camellia	Dragon	Jars	Jars2	Postbox	Stone L.	Sunflowers	Average
w/o MCMC + w/o EDI + w/ VGGSfM	PSNR↑	30.06	27.80	33.19	30.89	28.51	25.86	27.22	28.64	29.02
	SSIM↑	0.832	0.772	0.831	0.842	0.838	0.688	0.821	0.827	0.806
	LPIPS↓	0.097	0.118	0.081	0.097	0.102	0.177	0.169	0.226	0.133
w/ MCMC + w/ EDI + w/ HLOC	PSNR↑	30.32	27.07	31.01	28.62	27.85	22.94	24.73	31.76	28.04
	SSIM↑	0.838	0.717	0.773	0.752	0.816	0.586	0.706	0.898	0.761
	LPIPS↓	0.101	0.154	0.118	0.146	0.104	0.218	0.251	0.090	0.148
GeMS (w/ MCMC + w/o EDI + w/ VGGSfM)	PSNR↑	29.86	28.56	32.43	31.42	28.14	27.74	28.29	29.47	29.49
	SSIM↑	0.841	0.821	0.818	0.879	0.873	0.788	0.849	0.854	0.840
	LPIPS↓	0.118	0.129	0.171	0.127	0.173	0.150	0.195	0.163	0.153
w/o MCMC + w/ EDI + w/ VGGSfM	PSNR↑	32.63	28.70	36.41	31.66	28.60	28.52	27.97	33.97	31.06
	SSIM↑	0.901	0.811	0.933	0.865	0.839	0.803	0.855	0.942	0.869
	LPIPS↓	0.042	0.101	0.039	0.084	0.104	0.085	0.145	0.069	0.083
w/ MCMC + w/ EDI + w/o VGGSfM	PSNR↑	32.33	30.05	35.03	31.44	28.93	30.67	28.29	33.51	31.28
	SSIM↑	0.914	0.872	0.866	0.876	0.886	0.887	0.852	0.933	0.886
	LPIPS↓	0.081	0.111	0.203	0.147	0.161	0.098	0.224	0.086	0.139
GeMS-E (w/ MCMC + w/ EDI + w/ VGGSfM)	PSNR↑	33.55	29.47	37.01	32.35	28.79	31.33	29.43	33.69	31.95
	SSIM↑	0.924	0.873	0.925	0.898	0.906	0.906	0.894	0.938	0.908
	LPIPS↓	0.063	0.108	0.069	0.108	0.133	0.070	0.152	0.077	0.097

Our results demonstrate the effectiveness of VGGSfM for robust camera pose estimation and accurate initial point cloud generation, particularly in challenging motion-blurred scenarios where traditional methods like COLMAP fail. In GeMS, the combination of VGGSfM and 3DGS-MCMC based joint optimization leads to sharp and precise reconstructions, even in extreme blur conditions, offering a significant advantage over COLMAP-dependent methods. Furthermore, compared to HLOC, VGGSfM achieves an average PSNR improvement of **1.45 dB** for novel sharp view synthesis, as shown in Table 2, demonstrating its superior performance in motion-blurred environments. Hence, VGGSfM emerges as the most suitable choice for severely motion-blurred scenarios. Additionally, event-based deblurring (EDI) in **GeMS-E** further enhances sharpness, particularly in severe motion blur cases. As shown in 2, **GeMS-E** achieves the best overall performance, while **GeMS** provides a strong alternative when event data is unavailable.

Table 3: Training Time Comparison on the Real Dataset (hh:mm:ss): Comparison of training times for different methods on real datasets, highlighting the efficiency of our approach.

	Camera↓	Lego↓	Letter↓	Plant↓	Toys↓	Avg↓
EBAD-NeRF	06:44:00	06:44:00	06:44:00	06:44:00	06:44:00	06:44:00
E2NeRF	14:58:00	14:58:00	14:58:00	14:58:00	14:58:00	14:58:00
Ours	00:07:14	00:07:23	00:07:18	00:06:58	00:07:11	00:07:13

Table 4: GPU Memory Usage on the Real Dataset (MiB): Evaluation of GPU memory consumption across different methods, demonstrating the significant reduction in GPU memory usage.

	Camera↓	Lego↓	Letter↓	Plant↓	Toys↓	Avg↓
EBAD-NeRF	14836	14836	14836	14836	14836	14836
E2NeRF	15532	15532	15532	15532	15532	15532
Ours	1571	1897	1730	1341	1382	1584

Number of Virtual Cameras & Trajectory Representations: We conducted experiments to analyze the impact of the number of interpolated virtual camera poses within the duration of exposure, denoted as n, and the effectiveness of different motion trajectory representations. We varied n across $\{5, 10, 15, 20\}$, with the corresponding rendering results summarized in Figure 8. Our findings indicate that increasing n up to a certain threshold effectively mitigates severe motion blur, beyond which performance begins to decline; hence, we adopt $n = 15$. Additionally, we conducted ablations using linear interpolation, cubic B-splines, and Bézier curves for trajectory representations, with results also summarized in Figure 8. Our findings demonstrate that Bézier curves outperform other methods, as they better capture the complex, non-uniform motion present in heavily blurred scenes. Bézier curves consistently achieve superior performance across all metrics in severe motion blur scenarios, providing smoother and more accurate trajectory representations, making them the optimal choice for our framework.
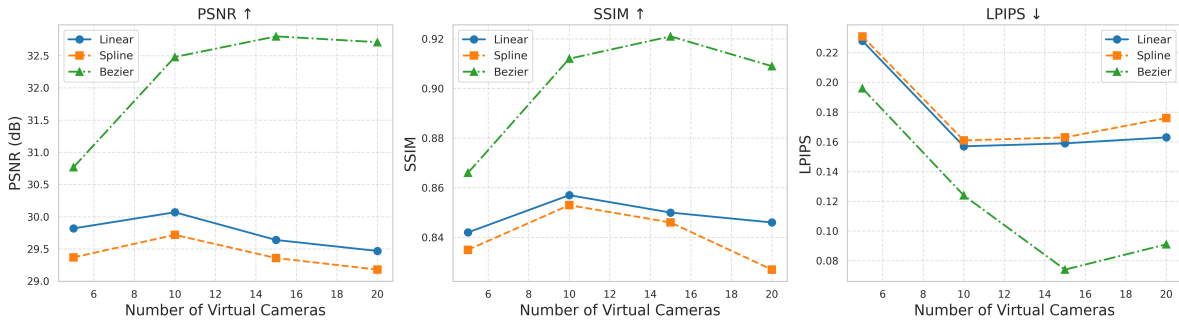
Figure 8: Impact of trajectory representations and the number of virtual cameras: Comparison of Linear, Spline, and Bezier trajectory representations across different virtual camera counts. Bezier interpolation consistently performs better than Linear and Spline representations, demonstrating its advantage in generating high-quality sharp novel views.

4 Conclusion

We introduced GeMS, an efficient 3D Gaussian Splatting framework designed to reconstruct scenes directly from severely motion-blurred images without relying on COLMAP or sharp image priors. Our approach integrates VGGSfM for robust pose estimation and point cloud generation directly from blurred images. MCMC-based Gaussian Splatting helps suppress artifacts and improve stability, while the joint optimization of camera motion trajectories and Gaussian parameters further refines poses, leading to a more accurate scene representation, even under extreme motion blur. When event data is available, we propose GeMS-E, which integrates event-based deblurring using the EDI model to enhance reconstruction quality. This refinement ensures sharper and more reliable results, particularly in scenarios where all input views are severely blurred. To advance future research, we make available our EveGeMS, a complementary synthetic event dataset tailored for scenarios involving extreme motion blur. Experimental results on both synthetic and real-world datasets demonstrate that GeMS / GeMS-E outperforms state-of-the-art methods in accuracy and efficiency, delivering superior reconstruction quality while significantly reducing training time and memory consumption. This work empowers NeRF and 3DGS pipelines to adopt robust deep SfM methods, such as VGGSfM, for reliable initialization in challenging scenarios like extreme motion blur, moving beyond the conventional reliance on COLMAP. Future work will explore deblurring with an extremely sparse set of images, ultimately pushing the framework's limits to handle the most challenging scenario: reconstructing sharp scene from a single motion-blurred image under extreme motion blur. Additionally, we aim to extend GeMS / GeMS-E to dynamic scenes.

References

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *CVPR*, 2021.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 2022.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *ECCV*, 2022.

Wenbo Chen and Ligang Liu. Deblur-gs: 3d gaussian splatting from camera motion blurred images. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–15, 2024.

Hiroyuki Deguchi, Mana Masuda, Takuya Nakabayashi, and Hideo Saito. E2gs: Event enhanced gaussian splatting. In *2024 IEEE International Conference on Image Processing (ICIP)*, pp. 1676–1682. IEEE, 2024.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *CVPR*, 2022.

Yifan Jiang, Peter Hedman, Ben Mildenhall, Dejia Xu, Jonathan T Barron, Zhangyang Wang, and Tianfan Xue. Alignerf: High-fidelity neural radiance fields via alignment-aware training. *CVPR*, 2023.

Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2):517–547, 2021.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *TOG*, 2023.

Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3d gaussian splatting as markov chain monte carlo. *Advances in Neural Information Processing Systems*, 37:80965–80986, 2025.

Dongwoo Lee, Jeongtaek Oh, Jaesung Rim, Sunghyun Cho, and Kyoung Mu Lee. Exblurf: Efficient radiance fields for extreme motion blurred images. In *ICCV*, 2023.

David B Lindell, Julien NP Martel, and Gordon Wetzstein. Autoint: Automatic integration for fast neural volume rendering. In *CVPR*, 2021.

Philipp Lindenberger, Paul-Edouard Sarlin, Viktor Larsson, and Marc Pollefeys. Pixel-perfect structure-from-motion with featuremetric refinement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5987–5997, 2021.

Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. Deblur-NeRF: Neural Radiance Fields from Blurry Images. In *CVPR*, 2022. URL `https://limacv.github.io/deblurnerf/`.

N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, pp. 99–108, 1995. doi: 10.1109/2945.468400.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, August 2020.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *TOG*, 2022.

Linfei Pan, Dániel Baráth, Marc Pollefeys, and Johannes L Schönberger. Global structure-from-motion revisited. In *European Conference on Computer Vision*, pp. 58–77. Springer, 2024.

Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a Blurry Frame Alive at High Frame-Rate With an Event Camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6820–6829, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An Imperative Style, High-performance Deep Learning Library. *Advances in neural information processing systems*, 32, 2019. URL `https://pytorch.org/`.

Yunshan Qi, Lin Zhu, Yu Zhang, and Jia Li. E2NeRF: Event Enhanced Neural Radiance Fields from Blurry Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13254–13264, 2023.

Yunshan Qi, Lin Zhu, Yifan Zhao, Nan Bao, and Jia Li. Deblurring neural radiance fields with event-driven bundle adjustment. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 9262–9270, 2024.

Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. ESIM: An Open Event Camera Simulator. In *Proceedings of The 2nd Conference on Robot Learning*, pp. 969–982. PMLR, October 2018.

Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *ICCV*, 2021.

Jeremy Reizenstein, Roman Shapovalov, Philipp Henzler, Luca Sbordone, Patrick Labatut, and David Novotny. Common objects in 3d: Large-scale learning and evaluation of real-life 3d category reconstruction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10901–10911, 2021.

Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019.

Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020.

Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion Revisited. In *CVPR*, 2016. URL `https://github.com/colmap/colmap`.

Thomas Schops, Johannes L Schonberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3260–3269, 2017.

Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. *CVPR*, 2022.

Jianyuan Wang, Nikita Karaev, Christian Rupprecht, and David Novotny. Vggsfm: Visual geometry grounded deep structure from motion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 21686–21697, 2024.

Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields. In *CVPR*, 2023. URL `https://wangpeng000.github.io/BAD-NeRF/`.

Liwen Wu, Jae Yong Lee, Anand Bhattad, Yuxiong Wang, and David Forsyth. Diver: Real-time and accurate neural radiance fields with deterministic integration for volume rendering, 2022.

Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, et al. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025.

Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. In *ICCV*, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage Progressive Image Restoration. In *CVPR*, 2021. URL `https://github.com/swz30/MPRNet`.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022.

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

Lingzhe Zhao, Peng Wang, and Peidong Liu. BAD-Gaussians: Bundle Adjusted Deblur Gaussian Splatting. In *ECCV*. Springer, 2024.

Lingzhe Zhao, Peng Wang, and Peidong Liu. Bad-gaussians: Bundle adjusted deblur gaussian splatting. In *European Conference on Computer Vision*, pp. 233–250. Springer, 2025.

Appendix

In the appendix, we provide additional information as outlined below:

- appendix A presents an overview of related works.

- appendix B details the background of 3D Gaussian Splatting and blur image formation model.

- appendix C elaborates on the EDI model and the joint optimization process of 3DGS.

- appendix D presents implementation details, additional quantitative results and ablations from experimental section.

- appendix E presents additional qualitative results of our methods, GeMS and GeMS-E, on both synthetic and real datasets across two different views.

A Related Work

A.1 SfM for NeRF/3DGS Initialization

Accurate camera pose estimation and sparse 3D reconstruction are essential for initializing NeRF and 3D Gaussian Splatting (3DGS). COLMAP Schonberger & Frahm (2016) is the most widely used incremental SfM pipeline, but its reliance on SIFT-based feature matching makes it vulnerable to motion blur and low-texture failures. GLOMAP Pan et al. (2024) improves scalability with a global SfM approach, while HLOC Sarlin et al. (2019) enhances wide-baseline localization using SuperGlue Sarlin et al. (2020). However, both methods remain dependent on pairwise feature correspondences, limiting their effectiveness under extreme motion blur. To address this, Pixel-Perfect SfM Lindenberger et al. (2021) refines both keypoint locations and camera poses by optimizing a featuremetric error using dense deep features. This improves the geometric accuracy of SfM pipelines, making it more robust to detection noise and appearance variations. Further advancing SfM, VGGSfM Wang et al. (2024) replaces traditional keypoint matching with deep 2D point tracking, eliminating the need for pairwise matching. It jointly optimizes camera poses and 3D points via a fully differentiable bundle adjustment layer, achieving state-of-the-art performance on CO3D Reizenstein et al. (2021), IMC Phototourism Jin et al. (2021), and ETH3D Schops et al. (2017). Given its superior robustness in extreme motion blur scenarios, VGGSfM is the preferred SfM method for initializing 3DGS in our deblurring pipeline.

A.2 Novel View Synthesis

NeRF Mildenhall et al. (2020) has garnered significant attention in 3D vision due to its remarkable ability to generate photo-realistic novel views. At its core, NeRF employs a neural implicit representation optimized via differentiable volume rendering. Numerous works have sought to enhance its rendering quality Barron et al. (2021; 2022); Jiang et al. (2023); Wu et al. (2022); Zhang et al. (2020), while others have focused on accelerating both training and rendering Lindell et al. (2021); Reiser et al. (2021); Yu et al. (2021); Fridovich-Keil et al. (2022); Müller et al. (2022); Chen et al. (2022); Sun et al. (2022), leading to significant improvements in efficiency.

Recently, *3D Gaussian Splatting* (3DGS) Kerbl et al. (2023) has emerged as an efficient alternative to radiance field models, excelling in both fine-grained scene reconstruction and real-time rendering. By replacing NeRF's computationally expensive ray marching Max (1995) with a deterministic rasterization technique, 3DGS preserves visual fidelity while enabling rapid rendering. However, 3DGS requires *accurate camera poses and point cloud initialization* for effective reconstruction, which can be challenging in motion-blurred scenarios. To further enhance stability and accuracy, recent works have explored *MCMC-based Gaussian Splatting* Kheradmand et al. (2025), treating Gaussians as probabilistic samples for more robust initialization and optimization. Inspired by this, our method integrates *MCMC-based Gaussian Splatting*, improving reconstruction quality in challenging motion-blurred scenarios.

A.3 Radiance Field Deblurring

NeRF-based approaches have explored various techniques to reconstruct sharp scene representations from motion-blurred images. Deblur-NeRF Ma et al. (2022) and ExBluRF Lee et al. (2023) model the blur process during scene optimization but assume fixed camera poses, which can be inaccurate under severe motion blur. To address this, BAD-NeRF Wang et al. (2023) jointly optimizes camera motion and radiance fields, allowing pose refinement. However, its reliance on implicit MLP-based representations leads to slow optimization and rendering times. Explicit representations, such as 3D Gaussian Splatting, have recently emerged as efficient alternatives for real-time rendering. BAD-Gaussians Zhao et al. (2024) extends 3DGS by jointly optimizing Gaussians and camera trajectories, improving both efficiency and reconstruction quality. Similarly, Deblur-GS Chen & Liu (2024) refines Gaussian parameters to achieve sharper reconstructions but still relies on COLMAP for initialization, making it ineffective in extreme motion blur scenarios.

Event-based methods have also been introduced to tackle motion blur in 3D reconstruction. E2NeRF Qi et al. (2023) and EBAD-NeRF Qi et al. (2024) integrate event streams into NeRF-based frameworks, leveraging high temporal resolution for deblurring. However, they inherit the inefficiencies of NeRF's implicit representations. More recently, E2GS Deguchi et al. (2024) has attempted to incorporate event data into Gaussian Splatting, but it fails to effectively remove motion blur, introducing severe color artifacts. To overcome these challenges, we propose a novel COLMAP-free Gaussian Splatting framework that directly processes motion-blurred images while incorporating event-based deblurring in extreme cases where event data is available. By leveraging VGGSfM for robust pose estimation and MCMC-based Gaussian Splatting for adaptive initialization, our approach jointly optimizes camera poses and Gaussian parameters, overcoming the limitations of previous methods and achieving high-quality, real-time 3D reconstruction under severe motion blur.

B Background

B.1 3D Gaussian Scene Representation

In the 3DGS framework Kerbl et al. (2023), a scene is modeled as a collection of 3D Gaussian distributions, each described by its mean position $\boldsymbol{\mu} \in \mathbb{R}^3$, a 3D covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{3 \times 3}$, opacity $\mathbf{o} \in \mathbb{R}$, and color $\mathbf{c} \in \mathbb{R}^3$. The distribution of each Gaussian is represented by the following formula:

$$\mathbf{G}(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{5}$$

The covariance matrix $\boldsymbol{\Sigma}$ is parameterized by a scale matrix $\mathbf{S} \in \mathbb{R}^3$ and a rotation matrix $\mathbf{R} \in \mathbb{R}^{3 \times 3}$. The rotation matrix $\mathbf{R}$ is represented using a quaternion $\mathbf{q} \in \mathbb{R}^4$, while the decomposition

$$\boldsymbol{\Sigma} = \mathbf{R} \mathbf{S} \mathbf{S}^T \mathbf{R}^T \tag{6}$$

ensures that $\boldsymbol{\Sigma}$ remains positive definite.

For rendering, 3D Gaussians are projected into 2D space from the camera pose $\mathbf{T}_c = \{\mathbf{R}_c \in \mathbb{R}^{3 \times 3}, \mathbf{t}_c \in \mathbb{R}^3\}$ using the following equations:

$$\boldsymbol{\Sigma}' = \mathbf{J} \mathbf{R}_c \boldsymbol{\Sigma} \mathbf{R}_c^T \mathbf{J}^T \tag{7}$$

where $\boldsymbol{\Sigma}' \in \mathbb{R}^{2 \times 2}$ is the 2D covariance matrix and $\mathbf{J} \in \mathbb{R}^{2 \times 3}$ is the Jacobian matrix for the projection.

To compute the rendered pixel colors, the 2D Gaussians are rasterized based on their depth values:

$$\mathbf{C} = \sum_{i=1}^{N} \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1}(1 - \alpha_j) \tag{8}$$

where $\mathbf{c}_i$ is the color of the i-th Gaussian, and α_i is the corresponding alpha value:

$$\alpha_i = \mathbf{o}_i \cdot \exp(-\sigma_i), \quad \sigma_i = \frac{1}{2} \Delta_i^T \boldsymbol{\Sigma}'^{-1} \Delta_i \tag{9}$$

Here, $\Delta_i \in \mathbb{R}^2$ represents the distance between the pixel center and the center of the 2D Gaussian. The pixel color $\mathbf{C}$ is differentiable with respect to both the Gaussian parameters $\mathbf{G}$ and the camera pose $\mathbf{T}_c$.

B.2   Physical Motion Blur Image Formation Model

The process of image formation in a digital camera involves the accumulation of photons during the exposure period, which are subsequently converted into electrical signals. Mathematically, this can be expressed as an integration over a sequence of virtual latent sharp images:

$$\mathbf{B}(\mathbf{u}) = \phi \int_0^\tau \mathbf{C}_t(\mathbf{u}) \mathrm{dt} \tag{10}$$

where $\mathbf{B}(\mathbf{u}) \in \mathbb{R}^{\mathrm{H} \times \mathrm{W} \times 3}$ represents the captured motion-blurred image, with $\mathbf{u} \in \mathbb{R}^2$ denoting the pixel location in an image of height H and width W. Here, ϕ is a normalization factor, τ is the exposure time, and $\mathbf{C}_t(\mathbf{u}) \in \mathbb{R}^{\mathrm{H} \times \mathrm{W} \times 3}$ corresponds to the latent sharp image at a given timestamp $t \in [0, \tau]$. The motion-blurred image $\mathbf{B}(\mathbf{u})$ arises due to camera movement during the exposure and is effectively the average of all latent sharp images $\mathbf{C}_t(\mathbf{u})$ over time. In practice, this integral is approximated using a finite number n of discrete samples, leading to the following discrete formulation:

$$\mathbf{B}(\mathbf{u}) \approx \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{C}_i(\mathbf{u}) \tag{11}$$

The extent of motion blur in an image is influenced by the camera's movement during exposure. A rapidly moving camera within a given exposure time creates severe motion blur. Note that $\mathbf{B}(\mathbf{u})$ remains differentiable with respect to each virtual sharp image $\mathbf{C}_i(\mathbf{u})$, which is a key property for optimization in motion deblurring tasks.

C   Method

C.1   EDI (Event Double Integral)

The event-based double integral model (EDI) Pan et al. (2019) reconstructs multiple sharp images from a single motion-blurred image using event data. Given a blurred image I_{blur} and a sequence of event bins $\{B_k\}_{k=1}^b$, the simplified EDI formulation adopted by E2NeRF Qi et al. (2023) models event accumulation to estimate sharp images over time.

We define the accumulated event count up to bin k as:

$$S_k = \sum_{i=1}^k B_i \tag{12}$$

The event-compensated image intensity is iteratively computed as:

$$E_k = E_{k-1} + e^{\Theta S_k}, \quad E_0 = 1, \quad k = 1, 2, \ldots, b \tag{13}$$

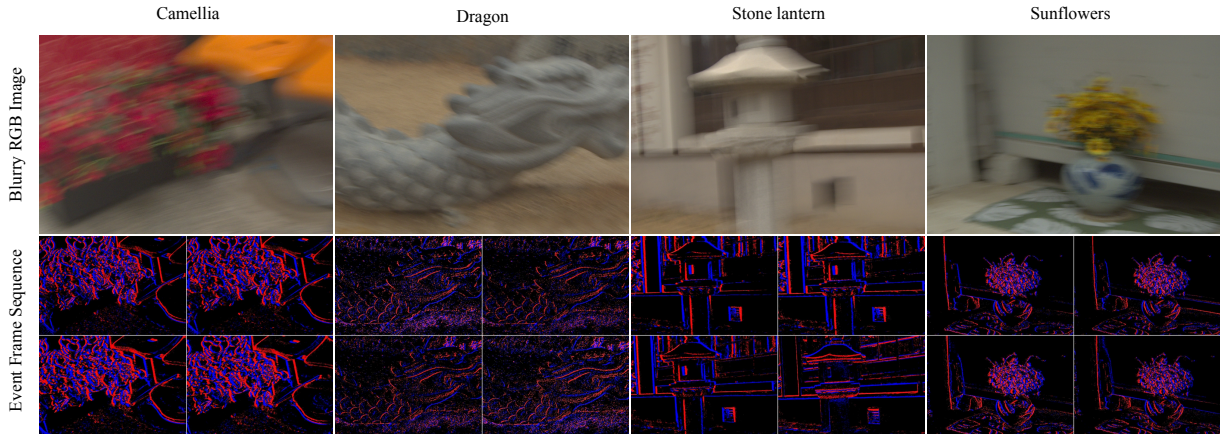where Θ is the event threshold controlling the influence of accumulated events.

The initial reference sharp image I_0 is estimated as:

$$I_0 = \frac{(b+1) I_{\mathrm{blur}}}{E_b} \tag{14}$$

where E_b corresponds to the final accumulated intensity estimate.

Each deblurred image I_k is obtained by propagating the initial image using event accumulation:

$$I_k = I_0 e^{\Theta S_k}, \quad k = 1, 2, \ldots, b \tag{15}$$

Figure 9: Our Complementary Event Dataset (EveGeMS): We introduce a synthetic event-based dataset tailored for cases with intense motion blur. Despite the heavy degradation in the RGB frames, each is paired with a stream of event frames that capture precise structural and motion details. This complementary signal enables reliable deblurring and novel view reconstruction, even when all input views are significantly blurred.

Among the reconstructed deblurred images $\{I_k\}_{k=0}^{b}$, we use the middle frame as a reference for the corresponding blurry image. These reference deblurred images are then fed into our GeMS framework. Using event data for initialization, *GeMS-E* enhances robustness against extreme motion blur, enabling reliable 3D reconstruction even when all input views are severely blurred.

C.2 Joint Optimization

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{\text{D-SSIM}} \tag{16}$$

To optimize both the learnable Gaussian parameters $\boldsymbol{\theta}$ and the camera poses $\mathbf{T}$ (represented using a Bézier curve of degree M with $M + 1$ control points T_j, $j = 0, ..., M$), the required Jacobians are computed to ensure proper gradient flow. As shown in Zhao et al. (2025), the gradient of the loss with respect to $\boldsymbol{\theta}$ is given by:

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}} = \sum_{k=0}^{K-1} \frac{\partial\mathcal{L}}{\partial\mathbf{B}_k} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{\partial\mathbf{B}_k}{\partial\mathbf{C}_i} \frac{\partial\mathbf{C}_i}{\partial\boldsymbol{\theta}} \tag{17}$$

while the gradient with respect to the camera pose is:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{T}} = \sum_{k=0}^{K-1} \frac{\partial\mathcal{L}}{\partial\mathbf{B}_k} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{\partial\mathbf{B}_k}{\partial\mathbf{C}_i} \frac{\partial\mathbf{C}_i}{\partial\boldsymbol{\theta}} \frac{\partial\boldsymbol{\theta}}{\partial\mathbf{T}} \tag{18}$$

For clarity, the explicit dependence on $\mathbf{u}$ in $\mathbf{B}_k(\mathbf{u})$ and $\mathbf{C}_i(\mathbf{u})$ is omitted. The camera poses are parameterized using their corresponding Lie algebra representations in $SE(3)$, each expressed as a 6D vector.

D Experiments

D.1 Implementation Details:

Our method is implemented in PyTorch Paszke et al. (2019) within the 3DGS-MCMC Kheradmand et al. (2025) framework using the *gsplat* Ye et al. (2025) pipeline. We optimize both Gaussians and camera poses in $SE(3)$ Bézier space using the Adam optimizer. For Bézier, we use 9 control points. The learning rate for Gaussians follows the original 3DGS Kerbl et al. (2023), while for camera poses, it is set to 1×10^{-3}. We set the number of virtual camera poses (n in Eq. 11) to 15, ensuring a balance between performance and efficiency. We use 13 event bins for event-based deblurring (EDI). All experiments are conducted on an NVIDIA RTX 4090 GPU with a data factor of 2, using 7k iterations for all experiments and comparisons.

D.2 Quantitative Results

Table 5: Quantitative comparisons for novel sharp view synthesis on the Synthetic Dataset: Note that ExBluRF* and BAD-Gaussians* rely on pose and point cloud initializations from sharp images, as COLMAP fails to work on extremely motion-blurred images. Best and second-best results are highlighted in green and orange, respectively. Here ExBluRF* and BAD-Gaussians* are also included in rankings.

Scene	Metric	Single Image Deblurring		Event-Based Deblurring			Sharp Image Supervision		Ours	
		MPRNet	Restormer	EDI	E2NeRF	EBAD-NeRF	ExBluRF*	BAD-Gaussians*	GeMS	GeMS-E
Bench	PSNR↑	25.35	26.39	24.78	25.41	28.15	31.93	32.54	29.86	33.55
	SSIM↑	0.678	0.720	0.642	0.708	0.822	0.877	0.901	0.841	0.924
	LPIPS↓	0.425	0.356	0.366	0.438	0.172	0.111	0.046	0.118	0.063
Camellia	PSNR↑	24.84	25.14	20.65	28.07	24.33	28.02	28.83	28.56	29.47
	SSIM↑	0.669	0.690	0.560	0.721	0.743	0.715	0.815	0.821	0.873
	LPIPS↓	0.395	0.351	0.391	0.329	0.192	0.313	0.099	0.129	0.108
Dragon	PSNR↑	29.96	28.37	28.93	30.89	33.99	33.45	36.98	32.43	37.01
	SSIM↑	0.731	0.704	0.648	0.697	0.864	0.828	0.930	0.818	0.925
	LPIPS↓	0.454	0.465	0.380	0.433	0.202	0.180	0.045	0.171	0.069
Jars	PSNR↑	25.36	25.57	25.15	29.85	28.89	30.85	31.52	31.42	32.35
	SSIM↑	0.680	0.687	0.637	0.775	0.838	0.840	0.867	0.879	0.898
	LPIPS↓	0.406	0.371	0.385	0.334	0.198	0.156	0.078	0.127	0.108
Jars2	PSNR↑	24.33	26.43	24.09	27.71	27.39	30.89	28.94	28.14	28.79
	SSIM↑	0.745	0.814	0.678	0.770	0.863	0.860	0.851	0.873	0.906
	LPIPS↓	0.358	0.275	0.359	0.383	0.171	0.113	0.114	0.173	0.133
Postbox	PSNR↑	25.89	26.52	23.50	30.66	26.82	31.40	26.40	27.74	31.33
	SSIM↑	0.736	0.753	0.658	0.813	0.826	0.864	0.757	0.788	0.906
	LPIPS↓	0.318	0.286	0.334	0.262	0.151	0.095	0.123	0.15	0.070
Stone Lantern	PSNR↑	24.97	26.68	24.14	30.47	26.29	28.24	28.29	28.29	29.43
	SSIM↑	0.785	0.831	0.697	0.836	0.802	0.765	0.843	0.849	0.894
	LPIPS↓	0.342	0.280	0.346	0.324	0.264	0.236	0.143	0.195	0.152
Sunflowers	PSNR↑	28.86	29.55	27.16	31.74	30.98	34.46	34.06	29.47	33.69
	SSIM↑	0.837	0.847	0.775	0.850	0.903	0.920	0.942	0.854	0.938
	LPIPS↓	0.242	0.206	0.287	0.310	0.117	0.093	0.065	0.163	0.077
Average	PSNR↑	26.19	26.83	24.80	29.35	28.36	31.15	30.95	29.49	31.95
	SSIM↑	0.733	0.756	0.662	0.771	0.833	0.834	0.863	0.840	0.908
	LPIPS↓	0.368	0.324	0.356	0.352	0.183	0.162	0.089	0.153	0.097

D.3 Ablations

Component-wise Analysis.

- **w/o MCMC + w/o EDI + w/ VGGSfM**: Baseline method without MCMC sampling and event-driven deblurring, relying solely on VGGSfM for Structure-from-Motion (SfM).

- **w/ MCMC + w/o EDI + w/ HLOC**: Incorporates MCMC-based optimization but excludes EDI deblurring, using HLOC for SfM instead of VGGSfM.

- **GeMS (w/ MCMC + w/o EDI + w/ VGGSfM)**: Our core approach, integrating MCMC-based optimization while using VGGSfM.

- **w/o MCMC + w/ EDI + w/ VGGSfM**: Introduces EDI-based event-driven deblurring while omitting MCMC optimization.

- **w/ MCMC + w/ EDI + w/o VGGSfM**: Replaces VGGSfM with COLMAP while incorporating both MCMC-based optimization and EDI deblurring.

- **GeMS-E (w/ MCMC + w/ EDI + w/ VGGSfM)**: Our full pipeline which incorporates MCMC-based optimization, event-driven deblurring (EDI), and VGGSfM.

Table 6: Ablation study for deblurring on the Synthetic Dataset. We assess the contribution of VGGSfM, MCMC, and EDI to deblurring performance. Each component progressively enhances PSNR, SSIM, and LPIPS, with their combined effect leading to the most accurate and visually consistent reconstructions.

Method	Metric	Bench	Camellia	Dragon	Jars	Jars2	Postbox	Stone L.	Sunflowers	Average
w/o MCMC + w/o EDI + w/ VGGSfM	PSNR↑	30.78	28.54	34.21	31.57	28.73	26.63	27.98	28.39	29.60
	SSIM↑	0.853	0.801	0.857	0.858	0.841	0.694	0.846	0.823	0.822
	LPIPS↓	0.080	0.101	0.083	0.079	0.096	0.176	0.137	0.234	0.123
w/ MCMC + w/o EDI + w/ HLOC	PSNR↑	30.65	28.32	32.32	28.82	28.50	25.76	26.34	32.72	29.18
	SSIM↑	0.838	0.774	0.807	0.755	0.831	0.664	0.770	0.909	0.794
	LPIPS↓	0.085	0.123	0.110	0.136	0.097	0.186	0.186	0.073	0.124
GeMS (w/ MCMC + w/o EDI + w/ VGGSfM)	PSNR↑	30.43	29.45	33.09	32.22	28.74	27.48	28.63	30.60	30.08
	SSIM↑	0.855	0.848	0.837	0.897	0.882	0.760	0.864	0.873	0.852
	LPIPS↓	0.098	0.110	0.150	0.093	0.129	0.164	0.148	0.146	0.130
w/o MCMC + w/ EDI + w/ VGGSfM	PSNR↑	33.06	29.45	37.69	32.30	29.39	29.87	29.66	34.57	32.00
	SSIM↑	0.910	0.834	0.938	0.878	0.852	0.823	0.891	0.947	0.884
	LPIPS↓	0.034	0.086	0.039	0.067	0.091	0.075	0.099	0.061	0.069
w/ MCMC + w/ EDI + w/o VGGSfM	PSNR↑	32.93	30.68	35.56	32.14	29.64	31.39	29.34	34.31	32.00
	SSIM↑	0.924	0.890	0.879	0.894	0.898	0.895	0.872	0.941	0.899
	LPIPS↓	0.059	0.092	0.173	0.115	0.118	0.088	0.182	0.075	0.113
GeMS-E (w/ MCMC + w/ EDI + w/ VGGSfM)	PSNR↑	34.23	30.12	37.48	33.30	30.01	32.22	30.38	34.63	32.80
	SSIM↑	0.933	0.890	0.932	0.917	0.921	0.913	0.916	0.947	0.921
	LPIPS↓	0.045	0.092	0.060	0.076	0.090	0.061	0.107	0.062	0.074

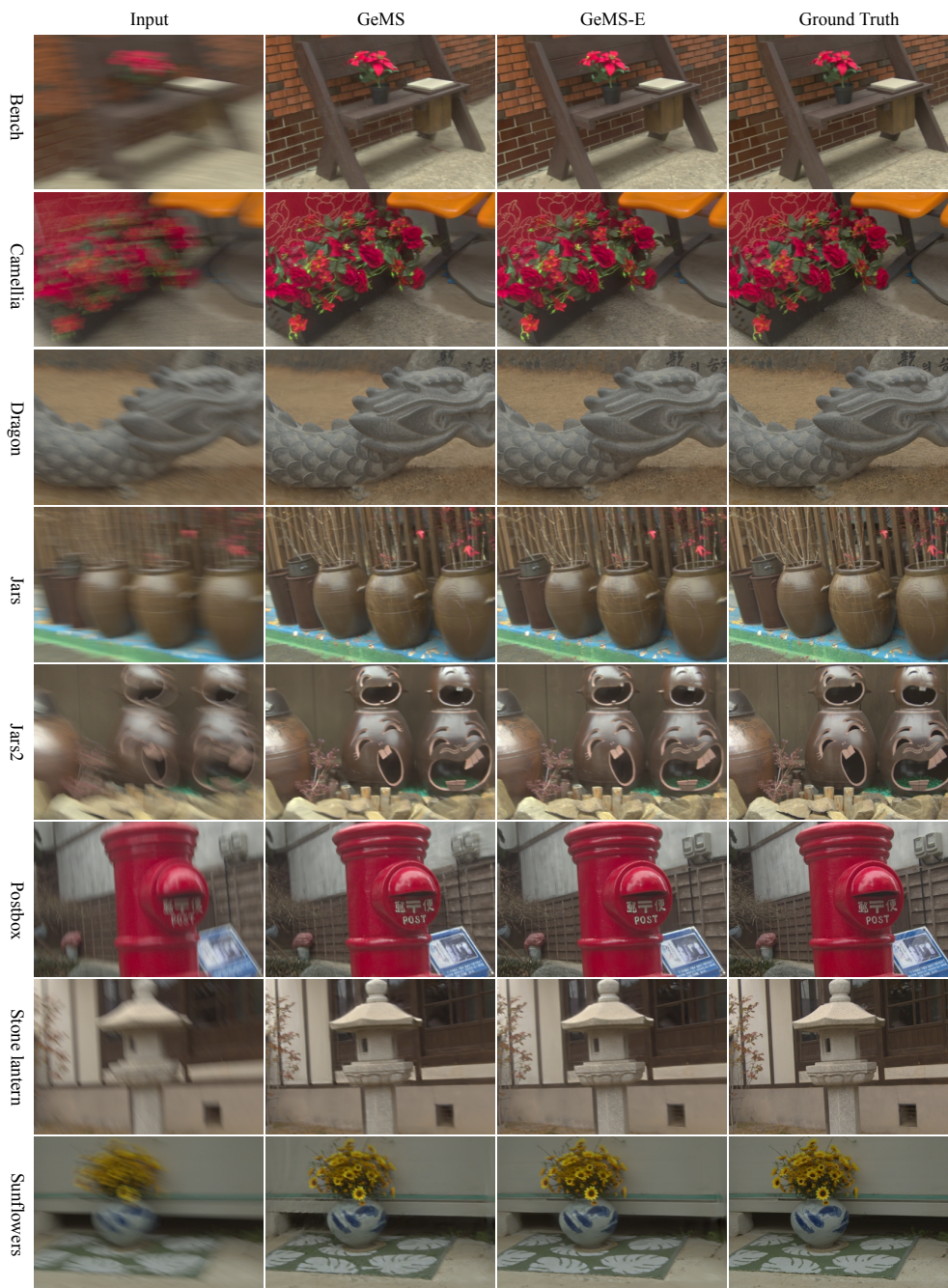# E   Qualitative Results

E.1   Synthetic Dataset(View-1)

Figure 10: Our Results on Synthetic Dataset
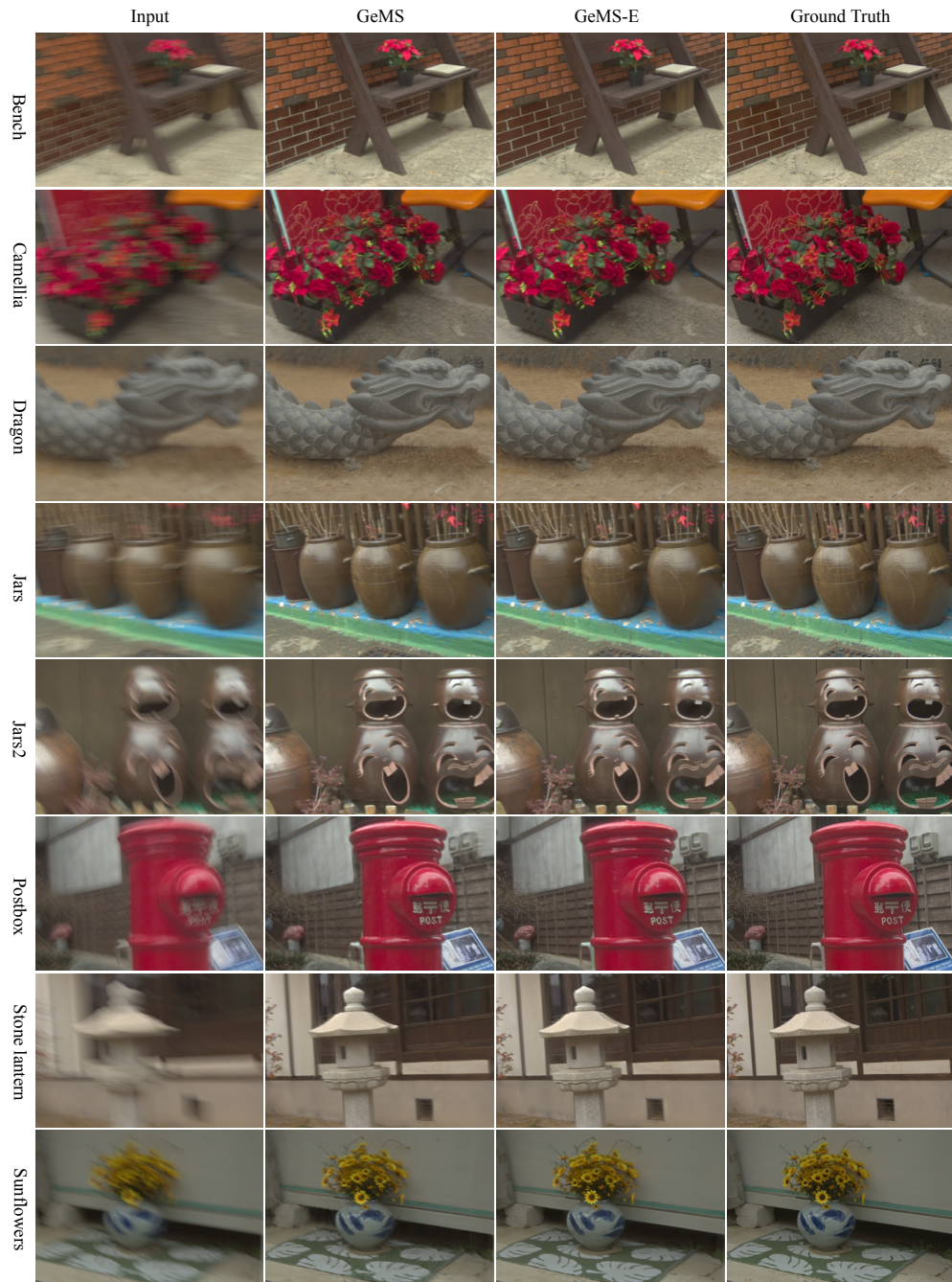
E.2 Synthetic Dataset(View-2)

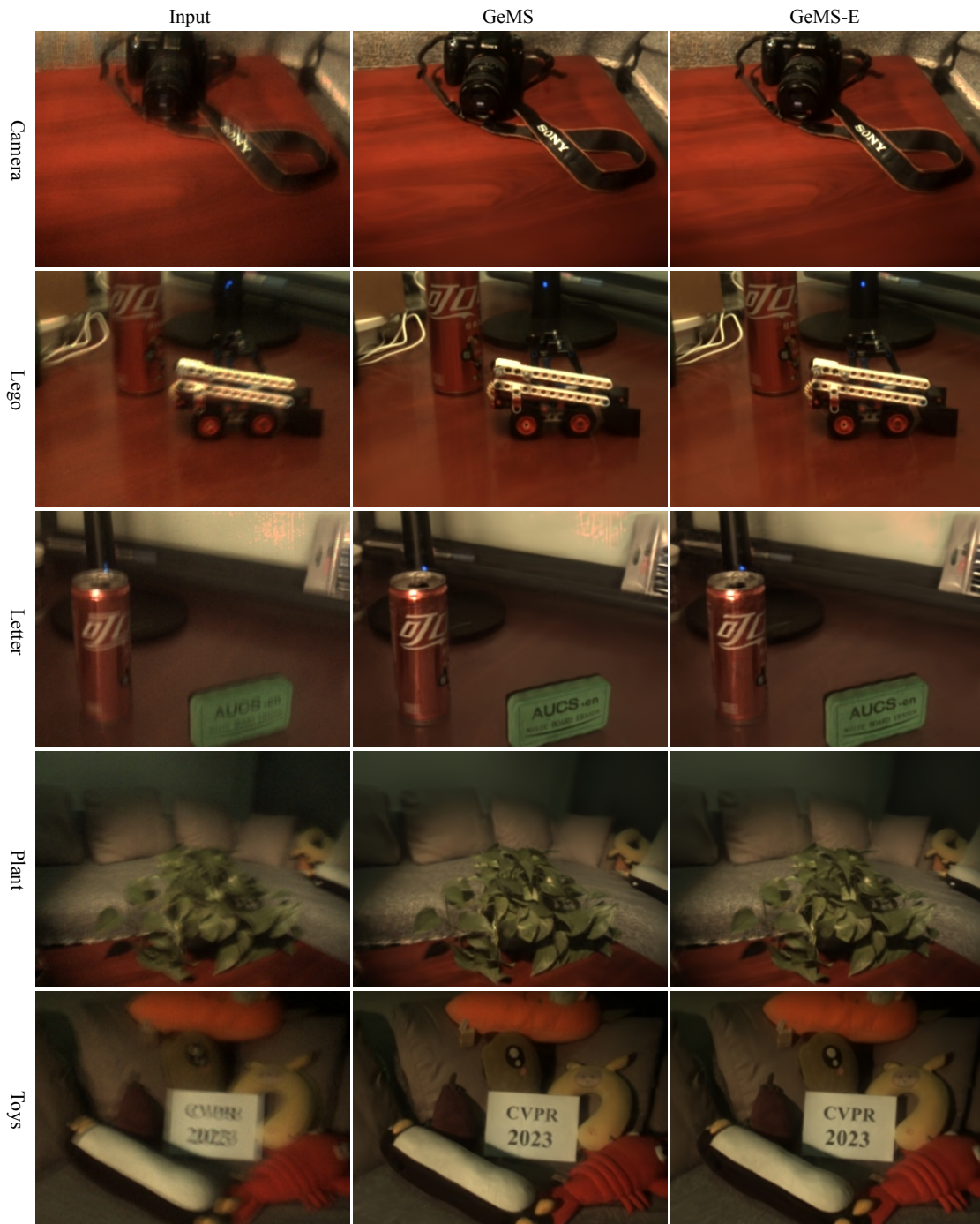Figure 11: Our Results on Synthetic Dataset

E.3 Real Dataset(View-1)

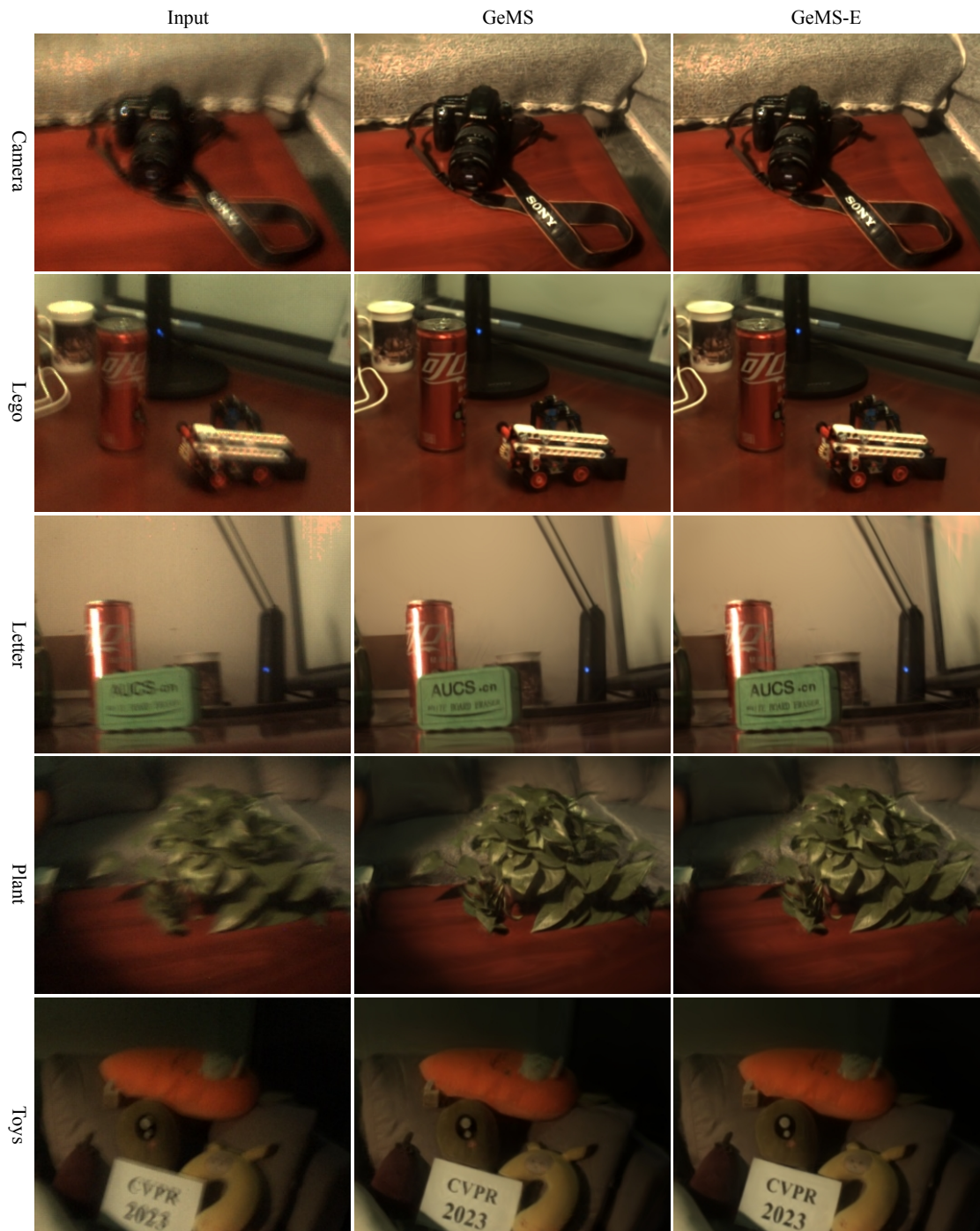
Figure 12: Our Results on Real Dataset

E.4   Real Dataset(View-2)

Figure 13: Our Results on Real Dataset