

BAD-Gaussians: Bundle Adjusted Deblur Gaussian Splatting

Lingzhe Zhao^{1*}, Peng Wang^{1,2*}, Peidong Liu¹

¹Westlake University ²Zhejiang University

{zhaolingzhe, wangpeng, liupeidong}@westlake.edu.cn

<https://lingzhezhao.github.io/BAD-Gaussians/>

Abstract. While neural rendering has demonstrated impressive capabilities in 3D scene reconstruction and novel view synthesis, it heavily relies on high-quality sharp images and accurate camera poses. Numerous approaches have been proposed to train Neural Radiance Fields (NeRF) with motion-blurred images, commonly encountered in real-world scenarios such as low-light or long-exposure conditions. However, the implicit representation of NeRF struggles to accurately recover intricate details from severely motion-blurred images and cannot achieve real-time rendering. In contrast, recent advancements in 3D Gaussian Splatting achieve high-quality 3D scene reconstruction and real-time rendering by explicitly optimizing point clouds into 3D Gaussians. In this paper, we introduce a novel approach, named BAD-Gaussians (Bundle Adjusted Deblur Gaussian Splatting), which leverages explicit Gaussian representation and handles severe motion-blurred images with inaccurate camera poses to achieve high-quality scene reconstruction. Our method models the physical image formation process of motion-blurred images and jointly learns the parameters of Gaussians while recovering camera motion trajectories during exposure time. In our experiments, we demonstrate that BAD-Gaussians not only achieves superior rendering quality compared to previous state-of-the-art deblur neural rendering methods on both synthetic and real datasets but also enables real-time rendering capabilities.

Keywords: 3D Gaussian Splatting · Deblurring · Bundle Adjustment · Differentiable Rendering

1 Introduction

Acquiring accurate 3D scene representations from 2D images has long been a challenging problem in computer vision. Serving as a fundamental component in various applications such as virtual/augmented reality and robotics navigation, substantial efforts have been dedicated to addressing this challenge over the last few decades. Among those pioneering works, Neural Radiance Fields (NeRF) [29], parameterized by Multi-layer Perceptrons (MLP), stands out for

* Equal contribution.

its utilization of differentiable volume rendering technique [22, 27] and has garnered significant attention due to its capability to recover high-quality 3D scene representation from 2D images.

Numerous works have focused on enhancing the performance of NeRF, particularly in terms of training [6, 30, 37] and rendering efficiency [13, 52]. A recent advancement, 3D Gaussian Splatting (3D-GS) [15], extends the implicit neural rendering [29] to explicit point clouds. By projecting these optimized point clouds (Gaussians) onto the image plane, 3D-GS [15] achieves real-time rendering while enhancing the efficiency of NeRF in both training and rendering, and also improves rendering quality. However, both NeRF-based [1, 29, 30] methods and 3D-GS [15] heavily rely on well-captured sharp images and accurately pre-computed camera poses, typically obtained from COLMAP [38]. Motion-blurred images, a common form of image degradation, often encountered in low-light or long-exposure conditions, can notably impair the performance of both NeRF and 3D-GS. The challenges posed by motion-blurred images to NeRF and 3D-GS can be attributed to three primary factors: (a) NeRF and 3D-GS rely on high-quality sharp images for supervision. However, motion-blurred images violate this assumption and exhibit notably inaccurate corresponding geometry between multi-view frames, thus presenting significant difficulties in achieving accurate 3D scene representation for both NeRF and 3D-GS; (b) Accurate camera poses are essential for training NeRF and 3D-GS. However, recovering accurate poses from multi-view motion-blurred images using COLMAP [38] is challenging. (c) 3D-GS necessitates sparse cloud points from COLMAP as the initialization of Gaussians. The mismatched features between multi-view blurred images and the inaccuracies in pose calibration further exacerbate the issue, leading to COLMAP producing fewer cloud points. This introduces an additional initialization issue for 3D-GS. Therefore, these factors result in a notable drop in performance for 3D-GS when dealing with motion-blurred images.

Implicit neural representations, i.e. NeRF [29], have been employed to reconstruct sharp 3D scenes from motion-blurred images [20, 26, 46]. For example, Deblur-NeRF [26] introduces a deformable sparse kernel that alters a canonical kernel at spatial locations to simulate the blurring process. DP-NeRF [20] integrates physical priors derived from the motion-blurred image acquisition process into Deblur-NeRF [26] to construct a clean NeRF representation. In contrast, BAD-NeRF [46] models the physical process of capturing motion-blurred images and jointly optimizes NeRF while recovering the camera trajectory within the exposure time. However, these implicit deblur rendering methods encounter significant challenges in achieving real-time rendering and producing high-quality outputs with intricate details. Additionally, the implicit representation introduces extra difficulties in optimizing the neural parameters and camera poses as mentioned in [12].

In order to address these challenges, we propose **Bundle Adjusted Deblur Gaussian Splatting**, the first motion deblur framework based on 3D-GS, which we refer to as **BAD-Gaussians**. We incorporate the physical process of motion blur into the training of 3D-GS, employing a spline function to characterize the

trajectory within the camera’s exposure time. In the training of BAD-Gaussians, the camera trajectory within exposure time is optimized using gradients derived from the Gaussians of the scene, while jointly optimizing the Gaussians themselves. Specifically, the trajectory of each motion-blurred image is represented by the initial and final poses at the beginning and end of the exposure time, respectively. By assuming the exposure time is typically short, we can interpolate between the initial and final poses to obtain every camera pose along the trajectory. From this trajectory, we generate a sequence of virtual sharp images by projecting the scene’s Gaussians onto the image plane. These virtual sharp images are then averaged to synthesize the blurred images, following the physical blur process. Finally, the Gaussians along the trajectory are optimized by minimizing the photometric error between the synthesized blurred images and the input blurred images through differentiable Gaussian rasterization.

We evaluate BAD-Gaussians using both synthetic and real datasets. The experimental results demonstrate that BAD-Gaussians outperforms prior state-of-the-art implicit neural rendering methods by explicitly incorporating the image formation process of motion-blurred images into the training of 3D-GS, achieving better rendering performance in terms of real-time rendering speed and superior rendering quality. In summary, our **contributions** can be outlined as follows:

- We introduce a photometric bundle adjustment formulation specifically designed for motion-blurred images, achieving the first real-time rendering performance from motion-blurred images within the framework of 3D Gaussian Splatting;
- We demonstrate how this formulation enables the acquisition of high-quality 3D scene representation from a set of motion-blurred images;
- Our approach successfully deblurs severe motion-blurred images, synthesizes higher-quality novel view images, and achieves real-time rendering, surpassing previous state-of-the-art implicit deblurring rendering methods.

2 Related Work

2.1 Neural Radiance Fields

NeRF [29], employing implicit MLP, exhibits remarkable performance in synthesizing high-quality novel view images and accurately representing 3D scenes. Numerous extension works have been proposed to enhance NeRF’s performance, including improvements in training [6, 30, 37, 42] and rendering [13, 19, 36, 48, 52] efficiency, as well as anti-alias capabilities [1–3]. Additionally, various methods aim to bolster NeRF’s robustness against imperfect inputs, such as inaccurate camera poses [14, 24, 34, 47], few-shot images [9, 16, 32], and low-quality images [20, 23, 26, 28, 46]. In the following section, we will primarily focus on reviewing methods closely related to our work.

Fast Neural Rendering. Numerous approaches inspired by NeRF [29] have sought to enhance its rendering efficiency by employing advanced data structures to reconstruct radiance fields, thereby minimizing the computational cost

associated with implicit MLPs used in NeRF. These approaches are primarily categorized into grid-based [5, 6, 11, 37] and hash-based [30] methods. Despite these efforts, achieving real-time rendering for unbounded and complete scenes, as well as high-resolution images, remains challenging. In contrast to methods based on volume rendering and implicit representations, which may hinder fast rendering, recent advancements like 3D-GS [15] achieve real-time high-quality rendering through pure explicit point scene representation and the differentiable Gaussian rasterization. Nevertheless, these 3D scene representations heavily rely on accurately posed high-quality images.

NeRF for Camera Optimization. The pioneering work, BARF [24], was the first to propose simultaneous optimization of camera parameters alongside NeRF. They employed a coarse-to-fine bundle adjustment strategy to enhance camera pose recovery. Concurrently, SC-NeRF [14] introduced a method to learn various camera models, encompassing both extrinsic and intrinsic parameters, along with scene representation. In the latest development, CamP [34] introduced a preconditioner to mitigate correlations between camera parameters, aiming for improved joint optimization of camera and NeRF parameters. In contrast to the aforementioned works [14, 24, 34], which focus solely on optimizing camera poses for sharp images, our approach goes further by recovering the trajectory of each blurred image within the exposure time.

NeRF for Deblurring. Several scene deblurring methods based on NeRF have been proposed, including Deblur-NeRF [26] and DP-NeRF [20], which reconstruct sharp scene representations from sets of motion-blurred images. However, these methods fix inaccurate camera poses recovered from blurred images during training, leading to degradation in reconstruction performance. BAD-NeRF [46] jointly learns camera motion trajectories within exposure time and radiance fields, following the physical blur process. Despite these advancements, existing deblurring neural rendering strategies struggle to achieve real-time rendering and reconstruct intricate scene details due to the implicit MLP structure, necessitating significant improvements in rendering efficiency and quality. Moreover, optimizing the 3D scenes jointly with camera poses faces additional challenges due to NeRF’s implicit representations. To address these issues, we propose achieving deblurring capability within the framework of Gaussian Splatting [15].

2.2 Image Deblurring

Two primary categories typically classify existing techniques for addressing the motion deblurring problem: the first involves formulating the issue as an optimization task, wherein gradient descent is employed during inference to jointly refine the blur kernel and the latent sharp image [7, 10, 17, 21, 33, 40, 49]. Another pipeline phrases the deblurring task as an end-to-end learning, particularly leveraging deep learning methodologies. With the support of substantial datasets and deep neural networks, superior results have been achieved for both single image deblurring [18, 31, 43] and video deblurring [41]. However, these 2D deblurring methods cannot exploit the 3D scene geometry between multi-view images, thus failing to ensure the view consistency of the scene from different viewpoints.

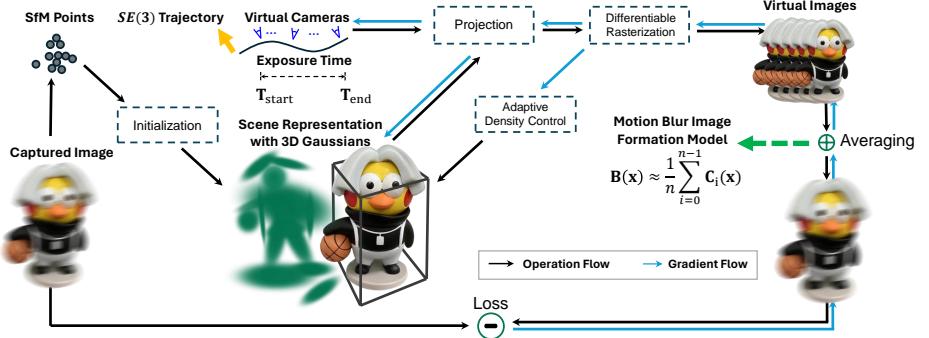


Fig. 1: The pipeline of BAD-Gaussians. Our approach utilizes Gaussian representations to depict sharp 3D scenes derived from a series of motion-blurred images, along with their inaccurate poses and sparse point clouds from COLMAP, serving as the initialization for the Gaussians. Employing forward projection and differentiable Gaussian rasterization, we jointly optimize the Gaussians in the scene and the camera trajectory within exposure time, by backpropagating gradients from Gaussians to camera poses. Following the physical process of motion blur, we model motion-blurred images by averaging the virtual sharp images captured during the exposure time. These virtual camera poses are represented and interpolated using a continuous spline within the $\text{SE}(3)$ space. The joint optimization of Gaussians and camera trajectories is achieved by minimizing the photometric loss between synthesized and actual blurry images.

In contrast, our approach focuses on leveraging the geometry within multi-view blurry images to reconstruct a high-quality 3D scene.

3 Method

BAD-Gaussians aims to recover a sharp 3D scene representation by jointly learning the camera motion trajectories and Gaussians parameters, given a sequence of motion-blurred images, along with their inaccurate poses and sparse point clouds estimated from COLMAP [38], as shown in Fig. 1. This is achieved by minimizing the photometric error between the input blurred images and the synthesized blurred images generated based on the physical motion blur image formation model. We will deliver each content in the following sections.

3.1 Preliminary: 3D Gaussian Splatting

Following 3D-GS [15], the scene is represented by a series of 3D Gaussians. Each Gaussian, denoted as \mathbf{G} , is parameterized by its mean position $\mu \in \mathbb{R}^3$, 3D covariance $\Sigma \in \mathbb{R}^{3 \times 3}$, opacity $\mathbf{o} \in \mathbb{R}$ and color $\mathbf{c} \in \mathbb{R}^3$. The distribution of each scaled Gaussian is defined as:

$$\mathbf{G}(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)}. \quad (1)$$

To ensure that the 3D covariance Σ remains positive semi-definite, which is physically meaningful, and to reduce the optimization difficulty, 3D-GS represents Σ using a scale $\mathbf{S} \in \mathbb{R}^3$ and rotation matrix $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ stored by a quaternion $\mathbf{q} \in \mathbb{R}^4$:

$$\Sigma = \mathbf{RSS}^T \mathbf{R}^T. \quad (2)$$

In order to enable differentiable Gaussian rasterization, 3D-GS projects 3D Gaussians to 2D from a given camera pose $\mathbf{T}_c = \{\mathbf{R}_c \in \mathbb{R}^{3 \times 3}, \mathbf{t}_c \in \mathbb{R}^3\}$ for rasterizing and rendering using the following equation, as described in [54]:

$$\Sigma' = \mathbf{JR}_c \Sigma \mathbf{R}_c^T \mathbf{J}^T, \quad (3)$$

where $\Sigma' \in \mathbb{R}^{2 \times 2}$ is the 2D covariance matrix, $\mathbf{J} \in \mathbb{R}^{2 \times 3}$ is the Jacobian of the affine approximation of the projective transformation. Afterward, each pixel color is rendered by rasterizing these N sorted 2D Gaussians based on their depths, following the formulation:

$$\mathbf{C} = \sum_i^N \mathbf{c}_i \alpha_i \prod_j^{i-1} (1 - \alpha_j), \quad (4)$$

where \mathbf{c}_i is the learnable color of each Gaussian, and α_i is the alpha value computed by evaluating a 2D covariance Σ' multiplied with the learned Gaussian opacity \mathbf{o} :

$$\alpha_i = \mathbf{o}_i \cdot \exp(-\sigma_i), \quad \sigma_i = \frac{1}{2} \Delta_i^T \Sigma'^{-1} \Delta_i, \quad (5)$$

where $\Delta \in \mathbb{R}^2$ is the offset between the pixel center and the 2D Gaussian center.

The derivations presented above demonstrate that the rendered pixel color, denoted as \mathbf{C} in Eq. (4), is a function that is differentiable with respect to all of the learnable Gaussians \mathbf{G} , and the camera poses \mathbf{T}_c . This facilitates our bundle adjustment formulation, accommodating a set of motion-blurred images and inaccurate camera poses within the framework of 3D-GS.

3.2 Physical Motion Blur Image Formation Model

The physical process of image formation in a digital camera encompasses the gathering of photons during the exposure period, followed by their conversion into measurable electric charges. Mathematically representing this phenomenon involves integrating across a sequence of simulated virtual latent sharp images, as follows:

$$\mathbf{B}(\mathbf{u}) = \phi \int_0^\tau \mathbf{C}_t(\mathbf{u}) dt, \quad (6)$$

where $\mathbf{B}(\mathbf{u}) \in \mathbb{R}^{H \times W \times 3}$ denotes the real captured motion-blurred image, $\mathbf{u} \in \mathbb{R}^2$ represents the pixel location in the image with height H and width W , ϕ serves as a normalization factor, τ is the camera exposure time, $\mathbf{C}_t(\mathbf{u}) \in \mathbb{R}^{H \times W \times 3}$ is the virtual latent sharp image captured at timestamp $t \in [0, \tau]$ within the

exposure time. The blurred image $\mathbf{B}(\mathbf{u})$, resulting from camera motion during the exposure time, is calculated by averaging all the virtual images $\mathbf{C}_t(\mathbf{u})$ across different timestamps t . This discrete approximation of this model is dependent on the number n of discrete samples, as denoted:

$$\mathbf{B}(\mathbf{u}) \approx \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{C}_i(\mathbf{u}). \quad (7)$$

The level of motion blur within an image is contingent upon the movement of the camera during the exposure time. For instance, a swiftly moving camera results in minimal relative motion, particularly with shorter exposure times, while a slowly moving camera yields motion-blurred images, especially in low-light scenarios with prolonged exposure times. Additionally, it can be deduced that $\mathbf{B}(\mathbf{u})$ demonstrates differentiability concerning each virtual sharp image $\mathbf{C}_i(\mathbf{u})$.

3.3 Camera Motion Trajectory Modeling in 3D-GS

Based on Eq. (7), a straightforward approach to address motion-blurred images involves determining each virtual sharp image \mathbf{C}_i , which serves as the dependent variable of the motion-blur image \mathbf{B} . Given that a sharp image \mathbf{C}_i can be rendered from a specified camera pose \mathbf{T}_i within the framework of 3D-GS (i.e. \mathbf{G}_θ) [15], establishing a one-to-one correspondence between poses and virtual sharp images is feasible. Consequently, we formulate the corresponding poses of each latent sharp image within the exposure time τ by employing a camera motion trajectory represented through linear interpolation between two camera poses, one at the beginning of the exposure $\mathbf{T}_{\text{start}} \in \mathbf{SE}(3)$ and the other at the end $\mathbf{T}_{\text{end}} \in \mathbf{SE}(3)$. The virtual camera pose at time $t \in [0, \tau]$ can thus be expressed as follows:

$$\mathbf{T}_t = \mathbf{T}_{\text{start}} \cdot \exp\left(\frac{t}{\tau} \cdot \log(\mathbf{T}_{\text{start}}^{-1} \cdot \mathbf{T}_{\text{end}})\right), \quad (8)$$

where τ represents the exposure time. $\frac{t}{\tau}$ can be further discretized and derived as $\frac{i}{n-1}$ for the i^{th} sampled virtual sharp image (i.e. C_i) in Eq. (7) (i.e. with pose denoted as \mathbf{T}_i), when there are n sampled images in total. It follows that \mathbf{T}_i is differentiable with respect to both $\mathbf{T}_{\text{start}}$ and \mathbf{T}_{end} . We refer to prior works [25, 46] for a comprehensive understanding of the interpolation and derivations of the related Jacobian. [25]. The objective of BAD-Gaussians is to estimate both $\mathbf{T}_{\text{start}}$ and \mathbf{T}_{end} for each frame, along with the learnable parameters of Gaussians \mathbf{G}_θ .

3.4 Loss Function

From a collection of K motion-blurred images, we can proceed to estimate both the learnable parameters $\boldsymbol{\theta}$ (i.e. mean position $\boldsymbol{\mu}$, 3D covariance $\boldsymbol{\Sigma}$, opacity \mathbf{o}

and color \mathbf{c}) of 3D-GS and the camera motion trajectories (i.e. $\mathbf{T}_{\text{start}}$ and \mathbf{T}_{end}) for each image. This estimation is accomplished by minimizing the following loss function, which includes an \mathcal{L}_1 loss and a D-SSIM term between $\mathbf{B}_k(\mathbf{u})$, the k^{th} blurry image synthesized from 3D-GS using the aforementioned image formation model (i.e. Eq. (7)), and $\mathbf{B}_k^{gt}(\mathbf{u})$, the corresponding real captured blurry image:

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{\text{D-SSIM}}. \quad (9)$$

To optimize the learnable Gaussians parameter $\boldsymbol{\theta}$, and camera pose \mathbf{T} (i.e. $\mathbf{T}_{\text{start}}$ and \mathbf{T}_{end} in our model) for each image, it is necessary to derive the corresponding Jacobians to complete the gradient flow:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = \sum_{k=0}^{K-1} \frac{\partial \mathcal{L}}{\partial \mathbf{B}_k} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{\partial \mathbf{B}_k}{\partial \mathbf{C}_i} \frac{\partial \mathbf{C}_i}{\partial \boldsymbol{\theta}}, \quad (10)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{T}} = \sum_{k=0}^{K-1} \frac{\partial \mathcal{L}}{\partial \mathbf{B}_k} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{\partial \mathbf{B}_k}{\partial \mathbf{C}_i} \frac{\partial \mathbf{C}_i}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{T}}, \quad (11)$$

where $\mathbf{B}_k(\mathbf{u})$ and $\mathbf{C}_i(\mathbf{u})$ are denoted as \mathbf{B}_k and \mathbf{C}_i for simplification, and we parameterize both $\mathbf{T}_{\text{start}}$ and \mathbf{T}_{end} with their corresponding Lie algebras of $\mathbf{SE}(3)$, which can be represented by a 6D vector respectively. For further details on the Jacobian of Gaussians to camera pose, $\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{T}}$, please refer to our supplemental materials.

4 Experiments

4.1 Experimental Settings

Implementation Details. We implemented our method using PyTorch [35] within the 3D-GS [15] framework. Both the optimization of Gaussians and camera poses are performed using the Adam optimizer. The learning rate for Gaussians remains identical to the original 3D-GS [15], while for camera poses (i.e., $\mathbf{T}_{\text{start}}$ and \mathbf{T}_{end} in Equation (8)), it is exponentially decreased from 1×10^{-3} to 1×10^{-5} . The number of virtual camera poses (i.e., n in Equation (7)) is set to 10, considering the trade-off between performance and efficiency. We initialize camera poses and Gaussians using estimations obtained from COLMAP [38]. All experiments are conducted on an NVIDIA RTX 4090 GPU.

Benchmark Datasets. We evaluate the performance of our method using both synthetic and real datasets provided by Deblur-NeRF [26] and BAD-NeRF [46]. Deblur-NeRF [46] contains five scenes, where the blurred images are generated via Blender [8] by averaging sharp virtual images captured during the exposure time, under the assumption of consistent velocity camera motion. For our evaluation, we utilize the dataset from BAD-NeRF [46], which expands upon the number of virtual sharp images to create more realistic motion-blurred images while keeping other settings consistent with [26]. Additionally, Deblur-NeRF [26]

Table 1: Ablation studies on the number of virtual camera poses n . The results indicate that performance reaches a saturation point with increasing number n .

n	Cozy2room			Tanabata		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3	31.63	.9389	.0810	25.43	.8215	.2287
4	32.99	.9552	.0558	27.08	.8704	.1764
5	33.68	.9595	.0458	28.39	.9011	.1400
6	33.86	.9627	.0368	29.57	.9214	.1101
8	34.48	.9654	.0315	31.15	.9440	.0696
10	34.68	.9521	.0258	32.12	.9481	.0464
15	34.78	.9679	.0257	33.58	.9652	.0198
20	34.63	.9646	.0307	33.87	.9668	.0151

Table 2: Ablation studies on the effect of trajectory representations. We denote Deblur-NeRF-S and Deblur-NeRF-R as the synthetic and real data from Deblur-NeRF, respectively. The results demonstrate that cubic interpolation improves performance in scenes with complex camera trajectories (i.e. *MBA-VO* and *Deblur-NeRF-R*).

	Deblur-NeRF-S			MBA-VO			Deblur-NeRF-R		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Linear Interpolation	33.92	.9467	.0422	30.96	.8868	.1491	24.77	.7778	.1291
Cubic B-spline	33.24	.9395	.0473	31.06	.8883	.1454	26.11	.8157	.1067

Table 3: Quantitative comparisons for deblurring on the synthetic dataset of Deblur-NeRF [26], referred to as DB-NeRF in the table due to space constraints. Notably, DB-NeRF* and DP-NeRF* are trained with ground-truth poses, while the others are trained with poses estimated by COLMAP [38]. The experiments highlight the superior performance of our method over previous approaches. Additionally, the results demonstrate the sensitivity of Deblur-NeRF and DP-NeRF to pose accuracy. Regarding rendering efficiency, our method achieves over 200 FPS, whereas Deblur-NeRF, DP-NeRF, and BAD-NeRF fall below 1 FPS. Our method takes about 30 minutes to train, while other methods take more than 10 hours. Each color shading indicates the best and second-best result, respectively.

	Cozyroom			Factory			Pool			Tanabata			Trolley		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [29]	26.13	.7886	.2484	22.02	.5581	.4550	29.90	.7901	.2595	20.57	.5653	.4584	21.71	.6413	.3814
3D-GS [15]	25.86	.7908	.2270	21.73	.5623	.4503	29.37	.7816	.2561	20.51	.5773	.4278	21.65	.6587	.3614
MPR [53]	29.90	.8862	.0915	25.07	.6994	.2409	33.28	.8938	.1290	22.60	.7203	.2507	26.24	.8356	.1762
SRN [43]	29.47	.8759	.0950	26.54	.7604	.2404	32.94	.8847	.1045	23.19	.7274	.2438	25.36	.8119	.1618
DB-NeRF [26]	29.53	.8786	.0879	25.85	.7651	.2340	31.07	.8477	.1397	23.20	.7243	.2471	25.68	.8111	.1683
DB-NeRF*	30.26	.8933	.0791	26.40	.7991	.2191	32.30	.8755	.1345	24.56	.7749	.2166	26.24	.8254	.1671
DP-NeRF [20]	29.02	.8761	.0773	25.42	.7559	.2139	30.48	.8402	.1127	23.93	.7540	.2038	26.17	.8269	.1359
DP-NeRF*	30.77	.9020	.0584	27.69	.8328	.1847	33.22	.8922	.0954	25.27	.7973	.1779	26.99	.8413	.1312
BAD-NeRF [46]	32.11	.9137	.0514	32.18	.9105	.1189	32.22	.8680	.0908	25.80	.8047	.1563	29.68	.8954	.0752
Ours	34.68	.9521	.0258	31.88	.9270	.0952	36.95	.9434	.0225	32.12	.9481	.0464	33.97	.9628	.0209

captured a real motion-blurred dataset by intentionally shaking a handheld camera during exposure.

To enhance the evaluation of our method on severely motion-blurred images, we incorporate a dataset tailored for motion blur-aware visual odometry

Table 4: Quantitative comparisons for **novel view synthesis** on the synthetic dataset of Deblur-NeRF [26], referred to as DB-NeRF in the table due to space constraints. The results demonstrate Our methods outperform previous state-of-the-art approaches, delivering the best performance across the board.

	Cozyroom			Factory			Pool			Tanabata			Trolley		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [29]	25.45	.7734	.2514	22.51	.5672	.4199	29.23	.7711	.2756	20.74	.5774	.4341	21.09	.6157	.3924
3D-GS [15]	25.24	.7789	.2084	21.34	.5532	.4371	28.61	.7562	.2610	20.37	.5755	.3958	20.86	.6167	.3701
MPR [53]+ [15]	29.00	.8685	.0894	23.38	.6848	.2374	29.04	.8170	.1761	22.53	.7349	.2162	26.27	.8352	.1694
SRN [43]+ [15]	28.64	.8662	.0903	25.45	.7740	.2045	29.97	.8207	.1650	23.09	.7456	.2141	24.89	.8103	.1666
DB-NeRF [26]	29.09	.8718	.0937	25.19	.7359	.2315	30.97	.8401	.1600	23.71	.7483	.2369	24.46	.7758	.2215
DB-NeRF*	29.88	.8901	.0747	26.06	.8023	.2106	30.94	.8399	.1694	24.82	.7861	.2045	25.78	.8122	.1797
DP-NeRF [20]	29.56	.8831	.0683	27.91	.8317	.1833	30.98	.8454	.1265	24.73	.7867	.1865	25.62	.8123	.1606
DP-NeRF*	30.16	.8958	.0608	28.43	.8420	.1771	31.78	.8623	.1306	25.37	.8046	.1701	26.44	.8307	.1424
BAD-NeRF [46]	31.07	.9027	.0551	31.71	.9038	.1204	30.95	.8389	.1107	25.42	.8011	.1566	28.60	.8806	.0843
Ours	32.35	.9278	.0340	28.18	.9164	.0999	33.30	.8964	.0551	31.00	.9427	.0478	31.44	.9443	.0345

Table 5: Quantitative comparisons for **deblurring** on the synthetic dataset of MBA-VO [26]. The results demonstrate that our method achieves the best performance even when the camera undergoes acceleration

	ArchViz-low			ArchViz-high			Average		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [29]	26.26	.7887	.3775	23.29	.7098	.4910	24.78	.7493	.4343
3D-GS [15]	26.54	.8113	.3385	23.43	.7243	.4614	24.99	.7678	.4000
MPR [53]	29.60	.8757	.2103	25.04	.7711	.3576	27.32	.8234	.2840
SRN [43]	30.15	.8814	.1703	27.07	.8190	.2796	28.61	.8502	.2250
DB-NeRF [26]	28.38	.8484	.1792	25.71	.7762	.3220	27.05	.8123	.2506
DB-NeRF*	29.65	.8744	.1764	26.44	.8010	.3172	28.05	.8377	.2468
DP-NeRF [20]	28.24	.8490	.1615	25.74	.7839	.2913	26.99	.8165	.2264
DP-NeRF*	28.57	.8481	.1511	26.80	.8081	.2736	27.69	.8281	.2124
BAD-NeRF [46]	30.51	.8749	.1654	27.54	.8109	.2679	29.03	.8429	.2167
Ours	32.28	.9167	.1134	29.64	.8568	.1847	30.96	.8868	.1491

benchmarking (i.e. MBA-VO [25]). Different from the assumption of constant velocity in the Deblur-NeRF dataset [26], the blur images synthesized from MBA-VO [25] are based on real camera motion trajectories obtained from the ETH3D dataset [39], which do not exhibit constant velocity and include accelerations, thus presenting notable challenges.

Baselines and Evaluation Metrics. We conduct a comparative **deblurring** analysis between our method and state-of-the-art learning-based single image deblurring algorithms including SRN [43] and MPR [53]. Additionally, we include evaluations against closely related approaches such as Deblur-NeRF [26], DP-NeRF [20], and BAD-NeRF [46]. To evaluate deblurring performance, we render high-quality images corresponding to the midpoint (i.e. $\mathbf{C}_{\frac{T}{2}}$ in Eq. (6)) of exposure time of each training image from the optimized Gaussians.

To facilitate **novel view synthesis** using SRN [43] and MPR [53], we use the deblurred images obtained from pre-trained models as inputs for training 3D-GS, as these single deblurring methods are not optimized for synthesizing novel

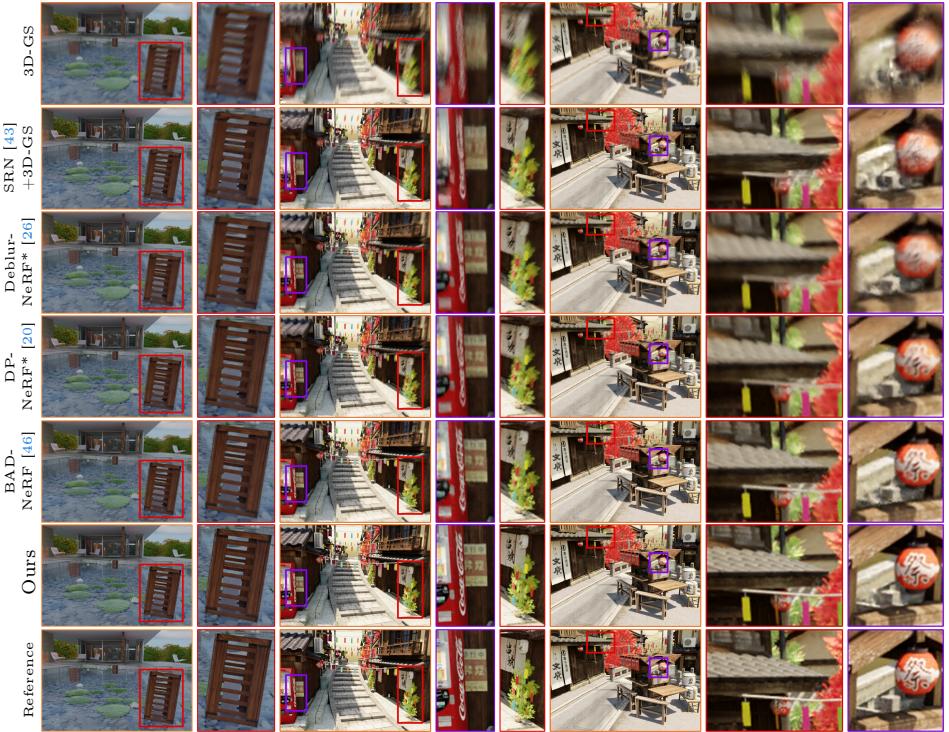


Fig. 2: Qualitative novel view synthesis results of different methods with synthetic datasets. Despite being trained with ground truth poses (*), BAD-Gaussians outperforms Deblur-NeRF* and DP-NeRF* in recovering high-quality scenes from motion-blurred images with inaccurate camera poses, showcasing its superior performance.

images. During the training stage, all poses are estimated using COLMAP [38]. Additionally, we train Deblur-NeRF [26] and DP-NeRF [20] with ground-truth camera poses from Blender to assess the impact of inaccurate pose estimation on these two methods. The quality of the rendered sharp image is evaluated using PSNR, SSIM, and LPIPS metrics.

For **pose accuracy** evaluation, we compare the Absolute Trajectory Error (ATE) metric, against classical structure-from-motion framework COLMAP [38] and BAD-NeRF [46].

4.2 Ablation Study

Virtual Camera Poses. We experimented to investigate the effect of the number of interpolated virtual camera poses within the exposure time, as described by n in Eq. (7). Two of the five synthetic scenes provided by Deblur-NeRF [26] were selected, representing sequences with minor and severe motion blur (i.e. *Cozy2room* and *Tanabata*), respectively. We varied the number, n , from 3 to

Table 6: Quantitative comparisons for **novel view synthesis** on the real captured dataset of Deblur-NeRF [26].

	Ball			Basket			Buick			Coffee			Decoration		
	PSNR↑	SSIM↑	LPIPS↓												
NeRF [29]	24.08	.6237	.3992	23.72	.7086	.3223	21.59	.6325	.3502	26.48	.8064	.2896	22.39	.6609	.3633
3D-GS [15]	23.72	.6321	.3210	23.96	.7466	.2600	21.53	.6630	.2743	27.44	.6630	.2208	22.29	.6826	.2868
DB-NeRF [26]	27.15	.7641	2112	27.35	.8367	.1347	24.93	.7791	.1545	30.72	.8949	.1070	24.15	.7730	.1700
DP-NeRF [20]	26.86	.7522	.2066	27.71	.8434	.1301	25.52	.7847	.1433	31.35	.9013	.0987	24.42	.7820	.1611
BAD-NeRF [46]	26.71	.7480	.2120	26.40	.8024	.1268	23.06	.7104	.2099	29.08	.8401	.1941	22.09	.6067	.3079
Ours	28.11	.8041	.2058	27.41	.8632	.0987	24.22	.8064	.1161	32.17	.9280	.0659	25.53	.8442	.0937
	Girl			Heron			Parterre			Puppet			Stair		
	PSNR↑	SSIM↑	LPIPS↓												
NeRF [29]	20.07	.7075	.3196	20.50	.5217	.4129	23.14	.6201	.4046	22.09	.6093	.3389	22.87	.4561	.4868
3D-GS [15]	19.97	.7276	.2613	20.28	.5254	.3109	22.98	.6326	.2967	22.38	.6463	.2645	22.68	.4709	.3911
DB-NeRF [26]	22.27	.7976	1687	22.63	.6874	.2099	25.82	.7597	.2161	25.24	.7510	.1577	25.39	.6296	.2102
DP-NeRF [20]	23.43	.8148	.1459	22.79	.7010	.1891	25.90	.7658	.1893	25.56	.7560	.1469	25.53	.6326	.1778
BAD-NeRF [46]	19.72	.6194	.3378	21.81	.6249	.2340	24.86	.7066	.2131	24.14	.7073	.1833	25.64	.6370	.1768
Ours	21.28	.8152	.1040	24.52	.7657	.1187	25.94	.8133	.0983	25.25	.7991	.0948	26.63	.7177	.0685

20 in our study, and the rendering metric results are presented in Table 1. Our experimental results demonstrate that increasing the number of interpolated virtual camera poses helps in addressing severe motion blur. However, marginal improvements are observed for minor blur instances. Based on our experiments, we choose $n = 10$ interpolated virtual images to strike a balance between rendering performance and training efficiency (larger n means more computational resources).

Trajectory Representations. To assess the impact of different trajectory representations, we conduct two experiments: one involves optimizing $\mathbf{T}_{\text{start}}$ and \mathbf{T}_{end} to depict a linear trajectory, while the other utilizes a higher-order spline (i.e., cubic B-spline) that jointly optimizes four control knots \mathbf{T}_1 , \mathbf{T}_2 , \mathbf{T}_3 , and \mathbf{T}_4 to capture more complex camera motions. We refer [45] for more details about cubic B-spline. The average quantitative results on the synthetic datasets of *Deblur-NeRF-S* [26] (Cozy2room, Factory, Pool, Tanabata and Trolley) and *MBA-VO* [25] (ArchViz-low and ArchViz-high), along with real datasets, *Deblur-NeRF-S*, which includes 10 real captured scenes from Deblur-NeRF, are presented in Table 2. In synthetic scenes, linear interpolation demonstrates comparable performance to cubic B-spline interpolation. However, in real captured scenes, cubic B-spline interpolation outperforms linear interpolation, particularly due to the longer exposure time during image capture. The effectiveness of cubic B-spline interpolation in real scenes can be attributed to its ability to better model the nuances of camera motion over longer time intervals. Conversely, linear interpolation is sufficient to accurately represent camera motion trajectories within shorter time intervals, as observed in synthetic scenes. Combining the training efficiency and rendering quality, we employ linear interpolation in synthetic datasets and cubic B-spline in real data.

4.3 Results

Results on Synthetic Data. We evaluate our approach against baseline methods using synthetic datasets obtained from both Deblur-NeRF [26] and MBA-

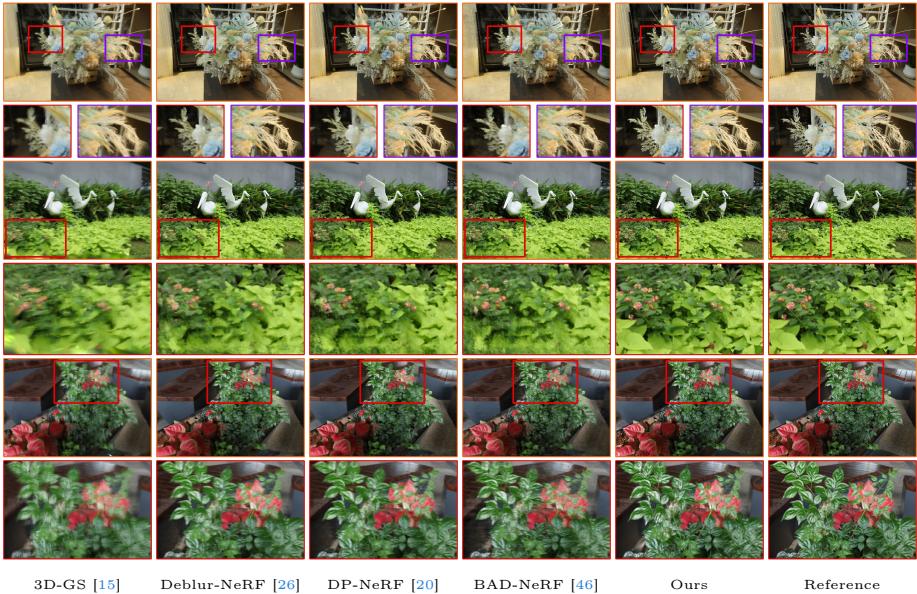


Fig. 3: Qualitative novel view synthesis results of different methods with the real datasets. The experimental results demonstrate that our method achieves superior performance over prior methods on the real dataset as well. In contrast, BAD-NeRF yields poorer results when applied to real data and exhibits satisfactory performance only within synthetic datasets.

VO [25]. The quantitative evaluation results on **deblurring** and **novel view synthesis** with the Deblur-NeRF dataset are presented in Table 3 and Table 4 respectively. All metrics (i.e. PSNR, SSIM, LPIPS) demonstrate substantial improvements over prior state-of-the-art methods (i.e. 3.6 and 1.7 dB higher than the second best method on average in Table 3 and Table 4, respectively).

The results demonstrate that both NeRF [29] and 3D-GS [15] are suffering from motion-blurred images, which motivates the necessity of our method. Single-stage methods such as MPR [53] and SRN [43] fall short of matching the performance of our approach due to their limited utilization of geometric information across multi-view images. Additionally, our method surpasses Deblur-NeRF [26] and DP-NeRF [20], partly because they fail to optimize the inaccurate camera poses estimated from COLMAP during training. We additionally trained Deblur-NeRF and DP-NeRF with ground truth poses (denoted as *), and the results demonstrate improved performance, reflecting the sensitivity to the pose accuracies of both Deblur-NeRF and DP-NeRF. A notable result from both Table 3 and Table 4 is that our method performs worse than BAD-NeRF on the Factory sequence. We find that it is caused by the inferior capability to represent the sky by Gaussian splatting compared to NeRF. Nevertheless, our method still performs better than all prior methods with a large margin on average.

Table 7: Pose estimation performance of BAD-Gaussians on various blur sequences. The results are in the absolute trajectory error metric in centimeters (ATE/cm). The COLMAP-sharp / COLMAP-blur represents the result of COLMAP with sharp/blurry images respectively.

	Cozy2room	Factory	Pool	Tanabata	Trolley	ArchViz-low	ArchViz-high
COLMAP-sharp [38]	.133±.079	.101±.052	.172±.047	.277±.103	.173±.079	.148±.078	.115±.066
COLMAP-blur [38]	.287±.161	1.69±1.24	.633±.236	1.59±1.06	.912±.559	.536±.309	.766±.391
BAD-NeRF [46]	.316±.101	.600±.208	.095±.046	.134±.543	.122±.070	1.24±.635	2.18±1.54
Ours	.165±.057	.395±.426	.264±.098	1.20±.621	.134±.070	.520±.591	1.41±1.11

Qualitative results are presented in Fig. 2. It demonstrates that our method effectively attains high-quality scene representations with intricate details, being trained from a series of blurred images. In Fig. 2, it is evident that Deblur-NeRF [26] and DP-NeRF [20] encounter difficulties in modeling regions with significant depth variations, attributable to their method of synthesizing blurred images through convolution with a blur kernel applied to rendered images. The physically based method BAD-NeRF [46] exhibits superior performance overall, yet it still presents some deficiencies, particularly noticeable around areas with significant color and depth variation. It demonstrates the effectiveness of the explicit Gaussian splatting representation over the implicit neural representation.

We further assess the performance of our methods using a dataset characterized by severe motion blur and camera movements with varying velocities, sourced from MBA-VO [25]. The results presented in Table 5 demonstrate that our method achieves the best performance even when the camera undergoes acceleration.

Results on Real Data. The performance of **novel view synthesis** on real data sourced from Deblur-NeRF [26] are also evaluated. The quantitative results, as presented in Table 6, demonstrate the superior performance of our method compared to other approaches. Furthermore, the qualitative results illustrated in Fig. 3 vividly demonstrate our method’s ability to deliver intricate details.

Pose Estimation. Due to the unknown metric scale, we align the estimated trajectories against their corresponding ground truth for the computation of the absolute trajectory error metric. The experimental results presented in Table 7 demonstrate the effectiveness of our approach in recovering precise camera poses.

5 Conclusion

We presented the first pipeline to learn Gaussian splattings from a set of motion-blurred images with inaccurate camera poses. Our pipeline can jointly optimize the 3D scene representation and camera motion trajectories. Extensive experimental evaluations demonstrate that our method can deliver high-quality novel view images, and achieve real-time rendering compared to prior state-of-the-art works.

BAD-Gaussians: Bundle Adjusted Deblur Gaussian Splatting

Supplementary Materials

A Introduction

In this supplementary material, we present the derivation of the analytical jacobian of the Gaussians w.r.t the camera poses and some additional qualitative and quantitative evaluation of our BAD-Gaussians.

B Jacobian of the Gaussians w.r.t the Camera Poses

In

$$\frac{\partial \mathcal{L}}{\partial \mathbf{T}_i} := \underbrace{\sum_{k=0}^{K-1} \frac{\partial \mathcal{L}}{\partial \mathbf{B}_k} \cdot \underbrace{\frac{1}{n} \sum_{i=0}^{n-1} \frac{\partial \mathbf{B}_k}{\partial \mathbf{C}_i} \frac{\partial \mathbf{C}_i}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{T}_i}}_{\text{auto-diff}}}_{\text{auto-diff}}, \quad (1)$$

we have

$$\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{T}_i} = \left[\frac{\partial \boldsymbol{\mu}'}{\partial \mathbf{T}_i} \quad \cancel{\frac{\partial \boldsymbol{\Sigma}'}{\partial \mathbf{T}_i}} \quad \cancel{\frac{\partial \mathbf{c}}{\partial \mathbf{T}_i}} \quad \cancel{\frac{\partial \mathbf{o}}{\partial \mathbf{T}_i}} \right], \quad (2)$$

where color \mathbf{c} and opacity \mathbf{o} of the Gaussian are independent with the virtual camera pose $\mathbf{T}_i = [\mathbf{R} \; \mathbf{t}] \in \mathbf{SE}(3)$ of the i^{th} virtual sharp image \mathbf{C}_i . Also, following GS-SLAM [50], we ignore $\frac{\partial \boldsymbol{\Sigma}'}{\partial \mathbf{T}_i}$ for efficiency.

As for $\frac{\partial \boldsymbol{\mu}'}{\partial \mathbf{T}_i}$, since the motion-blurred image synthesis is implemented in PyTorch [35], the first part of the gradient in Eq. (1) can be computed by the auto-diff module of PyTorch. The remaining part of the gradient in Eq. (1) can be computed as follows:

$$\frac{\partial \mathbf{C}_i}{\partial \boldsymbol{\mu}'} \frac{\partial \boldsymbol{\mu}'}{\partial \mathbf{T}_i} = \frac{\partial \mathbf{C}_i}{\partial \boldsymbol{\mu}'} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\mu}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\mu}_c} \frac{\partial \boldsymbol{\mu}_c}{\partial \mathbf{T}_i}, \quad (3)$$

where $\boldsymbol{\mu}_c = \mathbf{R}\boldsymbol{\mu} + \mathbf{t}$ represents $\boldsymbol{\mu}$ transformed into the camera's coordinate space.

The first term $\frac{\partial \mathbf{C}_i}{\partial \boldsymbol{\mu}'} \frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\mu}} = \frac{\partial \mathbf{C}_i}{\partial \boldsymbol{\mu}}$ is the Jacobian w.r.t. the mean position of the Gaussians, which is already computed in the CUDA backend of the differentiable projection and rasterization. The second term can be simplified as follows:

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\mu}_c} = \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{R}\boldsymbol{\mu} + \mathbf{t}} = \mathbf{R}^\top; \quad (4)$$

And the last term:

$$\frac{\partial \boldsymbol{\mu}_c}{\partial \mathbf{T}_i} = \mathbf{I}_3 \otimes [\boldsymbol{\mu}^\top \; 1] \in \mathbb{R}^{3 \times 12}, \quad (5)$$

where \otimes is the Kronecker operator and \mathbf{I}_3 is a 3×3 identity matrix [4] [51].

Finally, note that the virtual camera pose \mathbf{T}_i is interpolated from the control knots of the $\mathbf{SE}(3)$ continuous trajectory, e.g. $\mathbf{T}_{\text{start}}, \mathbf{T}_{\text{end}} \in \mathbf{SE}(3)$ for linear interpolation, and $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \mathbf{T}_4 \in \mathbf{SE}(3)$ for cubic B-spline. We use PyPose [44] to implement the interpolations, thus the corresponding Jacobian of \mathbf{T}_i w.r.t. the pose adjustments (the actual parameters being optimized), e.g. $\boldsymbol{\varepsilon}_{\text{start}}, \boldsymbol{\varepsilon}_{\text{end}} \in \mathfrak{se}(3)$ for linear interpolation and $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4 \in \mathfrak{se}(3)$ for cubic B-spline, are handled by auto-diff of PyTorch [35].

C On Ablation Studies

C.1 Full Table of Ablations on Trajectory Representations

The full results of our ablation study on trajectory representations are presented in Table A. The results demonstrate that linear interpolation adequately represents the camera motion trajectory for synthetic datasets, such as *MBA-VO* and *Deblur-NeRF-Synthetic*. However, cubic B-spline outperforms linear interpolation in real data scenarios (i.e. *Deblur-NeRF-Real*), attributed to the extended exposure time.

C.2 Details of Ablations on the Number of Virtual Cameras

In our ablation study on the number of virtual cameras n , for a fair comparison, we make the number of the densified Gaussians roughly the same by adjusting the threshold of the gradient in densification with n . This is based on the following derivation: In Eq. (1), during the synthesis of motion-blurred image, the gradient of every Gaussian is scaled by $\frac{1}{n}$. Therefore, if we change n to n' , the densification threshold should be multiplied by $\frac{n}{n'}$, in order to match the scaled gradients.

D Additional Qualitative Evaluation

We provide further qualitative experimental results on both the synthetic and real datasets, showcased in Fig. A and Fig. B respectively. These results demonstrate the superior performance of our method over previous state-of-the-art approaches.

D.1 Visualization of Pose Estimation Results

In this section, we present the visualization results in terms of camera pose estimation. The experiments are conducted on the synthetic dataset of Deblur-NeRF [26]. We present the comparison result of BAD-Gaussians against COLMAP [38] and BAD-NeRF [46] in Fig. C. It demonstrates that our method recovers motion trajectories more accurately.

Table A: Ablation studies on the effect of trajectory representations. The results demonstrate that cubic interpolation improves performance in scenes with complex camera trajectories (i.e. *MBA-VO* and *Deblur-NeRF-Real*).

<i>Dataset</i>	<i>Sequence</i>	Linear Interpolation			Cubic B-spline		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
<i>Deblur-NeRF</i> <i>Synthetic</i> [26]	Cozy2room	34.68	.9521	.0258	33.74	.9446	.0346
	Factory	31.88	.9270	.0952	31.87	.9324	.0897
	Pool	36.95	.9434	.0225	34.77	.9107	.0440
	Tanabata	32.12	.9481	.0464	32.09	.9477	.0464
	Trolley	33.97	.9628	.0209	33.73	.9619	.0220
<i>MBA-VO</i> [25]	ArchViz-low	32.28	.9167	.1134	32.43	.9165	.1118
	ArchViz-high	29.64	.8568	.1847	29.68	.8601	.1789
<i>Deblur-NeRF</i> <i>Real</i> [26]	Ball	23.10	.6423	.2778	28.11	.8041	.2078
	Basket	27.03	.8564	.0998	27.41	.8632	.0987
	Buick	23.44	.7939	.1118	24.22	.8064	.1161
	Coffee	30.52	.9017	.0913	32.17	.9280	.0659
	Decoration	24.99	.8305	.0992	25.53	.8442	.0937
	Girl	21.23	.8028	.1449	21.28	.8152	.1040
	Heron	21.70	.6929	.1495	24.52	.7657	.1187
	Parterre	25.03	.7817	.1173	25.94	.8133	.0983
	Puppet	24.78	.7811	.1021	25.25	.7991	.0948
	Stair	25.87	.6944	.0975	26.63	.7177	.0685

D.2 Video of Novel View Synthesis Results

To showcase the effectiveness of our approach, we provide supplementary videos illustrating the capability of BAD-Gaussians to recover high-quality latent sharp video from blurry images. The videos contain results on both synthetic and real scenes from Deblur-NeRF [26]. In the provided videos, on the left are our rendered novel view images and on the right are the input blurry images.

Notably, in the provided videos, due to the fast training speed and low GPU memory requirements of our method, we are able to train real scenes at the native resolution 2400×1600 to achieve maximum reconstruction quality in about 1.5 hours, compared to the resolution of 600×400 that we used in all experiments above in this paper for a fair comparison.

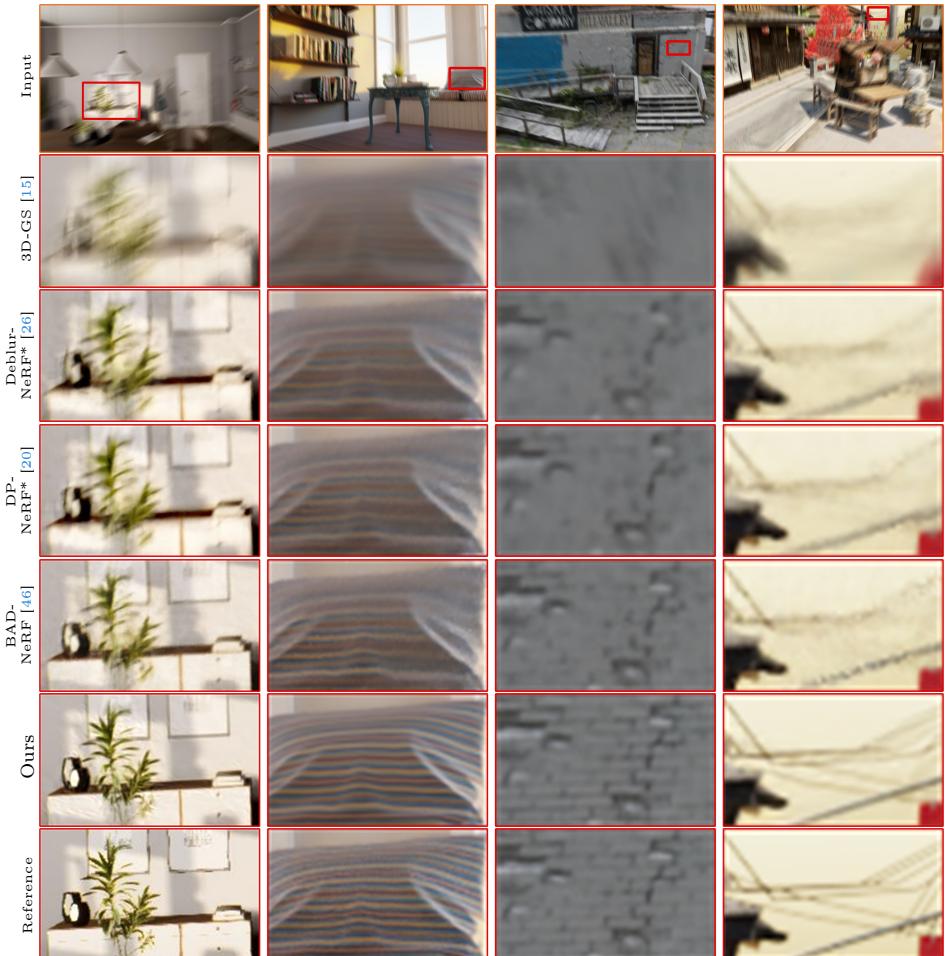


Fig. A: Qualitative deblurring results of different methods with synthetic datasets from MBA-VO [25] and Deblur-NeRF [26]. The scenes, from left to right, encompass *ArchViz-high*, *Cozy2room*, *Factory*, and *Trolley*. Despite being trained with ground truth poses (*), BAD-Gaussians outperforms Deblur-NeRF* and DP-NeRF* in recovering high-quality scenes from motion-blurred images with inaccurate camera poses, showcasing its superior performance.

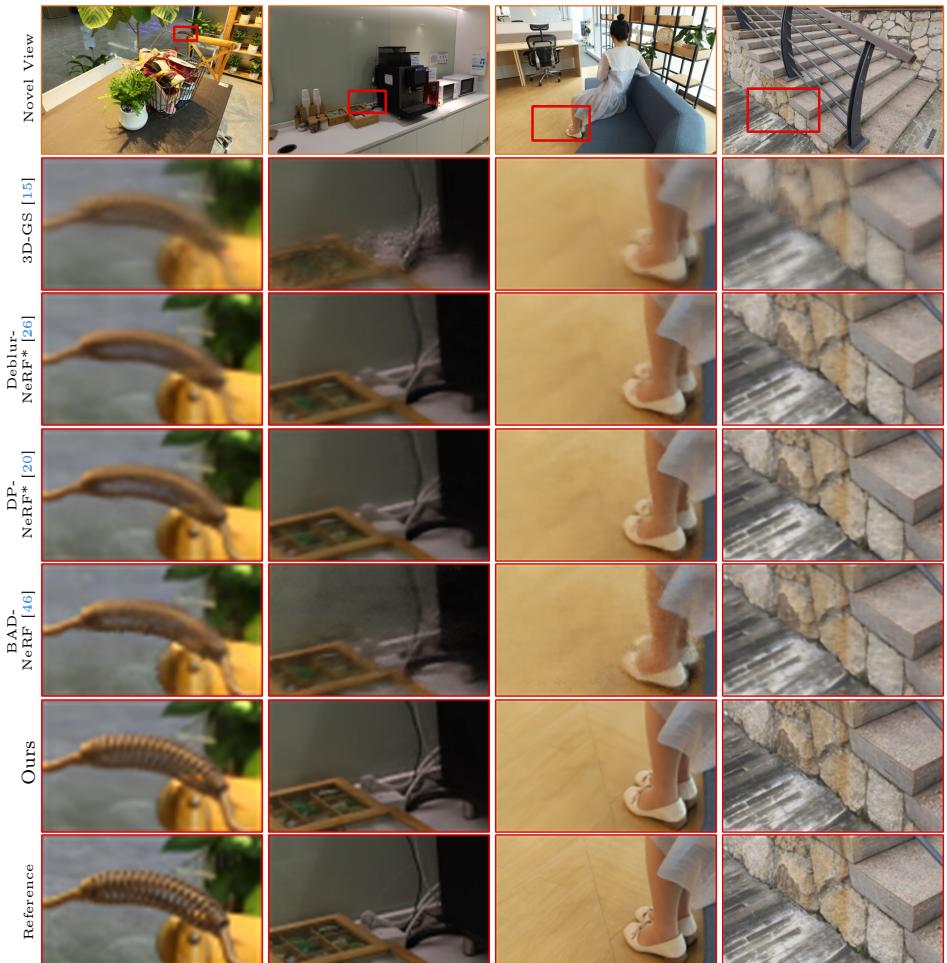


Fig. B: Qualitative novel view synthesis results of different methods with the real datasets from Deblur-NeRF [26]. The scenes, from left to right, encompass *Basket*, *Coffee*, *Girl*, and *Stair*. The experimental results demonstrate that our method achieves superior performance over prior methods on the real dataset as well.

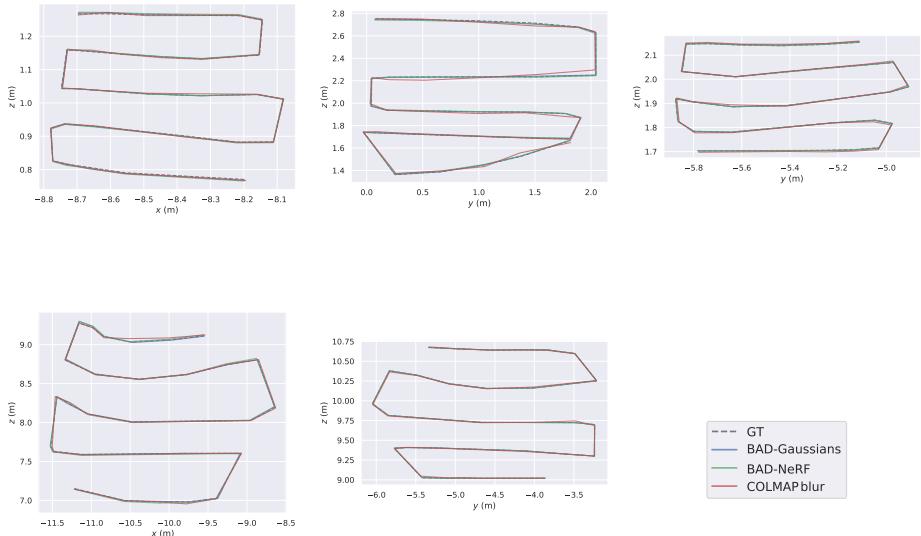


Fig. C: Qualitative Comparisons of estimated camera poses on Deblur-NeRF dataset. These are results on *Cozy2room*, *Factory*, *Pool*, *Tanabata* and *Trolley* sequences respectively. The results demonstrate that our method recovers motion trajectories more accurately compared with both COLMAP [38] and BAD-NeRF [46].

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In: International Conference on Computer Vision (ICCV) (2021), <https://jonbarron.info/mipnerf/> 2, 3
2. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://jonbarron.info/mipnerf360/> 3
3. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. In: International Conference on Computer Vision (ICCV) (2023), <https://jonbarron.info/zipnerf/> 3
4. Blanco, J.L.: A tutorial on SE(3) transformation parameterizations and on-manifold optimization. Tech. Rep. 012010, University of Malaga (2010), http://ingmec.ual.es/~jlblanco/papers/jlblanco2010geometry3D_techrep.pdf 16
5. Cao, A., Johnson, J.: HexPlane: A Fast Representation for Dynamic Scenes. In: Computer Vision and Pattern Recognition (CVPR) (2023), <https://caoang327.github.io/HexPlane/> 4
6. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: TensoRF: Tensorial Radiance Fields. In: European Conference on Computer Vision (ECCV) (2022), <https://apchenstu.github.io/TensoRF/> 2, 3, 4
7. Cho, S., Lee, S.: Fast Motion Deblurring. ACM Transactions on Graphics (TOG) **28**(5) (2009) 4
8. Community, B.O.: Blender - a 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam (2018), <http://www.blender.org> 8
9. Deng, K., Liu, A., Zhu, J.Y., Ramanan, D.: Depth-supervised NeRF: Fewer Views and Faster Training for Free. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://www.cs.cmu.edu/~dsnerf/> 3
10. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. In: ACM Transactions on Graphics (TOG) (2006) 4
11. Fridovich-Keil, S., Meanti, G., Warburg, F.R., Recht, B., Kanazawa, A.: K-Planes: Explicit Radiance Fields in Space, Time, and Appearance. In: Computer Vision and Pattern Recognition (CVPR) (2023), <https://sarafridov.github.io/K-Planes/> 4
12. Fu, Y., Liu, S., Kulkarni, A., Kautz, J., Efros, A.A., Wang, X.: Colmap-Free 3D Gaussian Splatting. arXiv preprint arXiv:2312.07504 (2023), <https://oasisyang.github.io/colmap-free-3dgs/> 2
13. Garbin, S.J., Kowalski, M., Johnson, M., Shotton, J., Valentin, J.: FastNeRF: High-Fidelity Neural Rendering at 200FPS. In: International Conference on Computer Vision (ICCV) (2021), <https://microsoft.github.io/FastNeRF/> 2, 3
14. Jeong, Y., Ahn, S., Choy, C., Anandkumar, A., Cho, M., Park, J.: Self-Calibrating Neural Radiance Fields. In: International Conference on Computer Vision (ICCV) (2021), <https://postech-cvlab.github.io/SCNeRF/> 3, 4
15. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics (TOG) **42**(4) (July 2023), <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/> 2, 4, 5, 7, 8, 9, 10, 12, 13, 18, 19

16. Kim, M., Seo, S., Han, B.: InfoNeRF: Ray Entropy Minimization for Few-Shot Neural Volume Rendering. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://cv.snu.ac.kr/research/InfoNeRF/> 3
17. Krishnan, D., Fergus, R.: Fast image deconvolution using Hyper-Laplacian priors. In: Neural Information Processing Systems (NIPS) (2009) 4
18. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. In: International Conference on Computer Vision (ICCV) (2019), <https://github.com/VITA-Group/DeblurGANv2> 4
19. Lassner, C., Zollhofer, M.: Pulsar: Efficient sphere-based neural rendering. In: Computer Vision and Pattern Recognition (CVPR) (2021), <https://github.com/facebookresearch/pytorch3d/> 3
20. Lee, D., Lee, M., Shin, C., Lee, S.: DP-NeRF: Deblurred Neural Radiance Field With Physical Scene Priors. In: Computer Vision and Pattern Recognition (CVPR) (2023), <https://dogyoonlee.github.io/dpnerf/> 2, 3, 4, 9, 10, 11, 12, 13, 14, 18, 19
21. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding and evaluating blind deconvolution algorithm. In: Computer Vision and Pattern Recognition (CVPR) (2009) 4
22. Levoy, M.: Efficient Ray Tracing of Volume Data. ACM Transactions on Graphics (TOG) 9(3), 245–261 (1990) 2
23. Li, M., Wang, P., Zhao, L., Liao, B., Liu, P.: USB-NeRF: Unrolling Shutter Bundle Adjusted Neural Radiance Fields. In: International Conference on Learning Representations (ICLR) (2024), https://moyangli00.github.io/usb_nerf/ 3
24. Lin, Chen-Hsuan and Ma, Wei-Chiu and Torralba, Antonio and Lucey, Simon: BARF: Bundle-Adjusting Neural Radiance Fields. In: International Conference on Computer Vision (ICCV) (2021), <https://chenhsuanlin.bitbucket.io/bundle-adjusting-NeRF> 3, 4
25. Liu, P., Zuo, X., Larsson, V., Pollefeys, M.: MBA-VO: Motion Blur Aware Visual Odometry. In: International Conference on Computer Vision (ICCV) (2021), <https://github.com/ethliup/MBA-VO> 7, 10, 12, 13, 14, 17, 18
26. Ma, L., Li, X., Liao, J., Zhang, Q., Wang, X., Wang, J., Sander, P.V.: Deblur-NeRF: Neural Radiance Fields from Blurry Images. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://limacv.github.io/deblurnerf/> 2, 3, 4, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19
27. Max, N.: Optical Models for Direct Volume Rendering. ACM Transactions on Graphics (TOG) 1(2), 99–108 (1995) 2
28. Mildenhall, B., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., Barron, J.T.: NeRF in the dark: High dynamic range view synthesis from noisy raw images. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://bmild.github.io/rawnerf/> 3
29. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In: European Conference on Computer Vision (ECCV) (2020), <https://www.matthewtancik.com/nerf> 1, 2, 3, 9, 10, 12, 13
30. Müller, T., Evans, A., Schied, C., Keller, A.: Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. ACM Transactions on Graphics (TOG) 41(4), 102:1–102:15 (Jul 2022), <https://nvlabs.github.io/instant-nfp/> 2, 3, 4
31. Nah, S., Kim, T.H., Lee, K.M.: Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring. In: Computer Vision and Pattern Recognition (CVPR) (2017), https://github.com/SeungjunNah/DeepDeblur_release 4

32. Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S., Geiger, A., Radwan, N.: RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://m-niemeyer.github.io/regnerf/> 3
33. Park, H., Lee, K.M.: Joint Estimation of Camera Pose, Depth, Deblurring, and Super-Resolution from a Blurred Image Sequence. In: International Conference on Computer Vision (ICCV). vol. 2017-Octob (2017). <https://doi.org/10.1109/ICCV.2017.494> 4
34. Park, K., Henzler, P., Mildenhall, B., Barron, J.T., Martin-Brualla, R.: CamP: Camera Preconditioning for Neural Radiance Fields. ACM Transactions on Graphics (TOG) (2023), <https://camp-nerf.github.io/> 3, 4
35. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An Imperative Style, High-performance Deep Learning Library. Advances in neural information processing systems **32** (2019), <https://pytorch.org/> 8, 15, 16
36. Piala, M., Clark, R.: TermiNeRF: Ray Termination Prediction for Efficient Neural Rendering. In: International Conference on 3D Vision (3DV) (2021), <https://projects.mackopes.com/terminerf/> 3
37. Sara Fridovich-Keil and Alex Yu, Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance Fields without Neural Networks. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://alexyu.net/plenoxels/> 2, 3, 4
38. Schonberger, J.L., Frahm, J.M.: Structure-from-motion Revisited. In: Computer Vision and Pattern Recognition (CVPR) (2016), <https://github.com/colmap/colmap> 2, 5, 8, 9, 11, 14, 16, 20
39. Schops, T., Sattler, T., Pollefeys, M.: BAD SLAM: Bundle Adjusted Direct RGB-D SLAM. In: Computer Vision and Pattern Recognition (CVPR) (2019), <https://github.com/ETH3D/badslam> 10
40. Shan, Q., Jia, J., A., A.: High-quality Motion Deblurring from a Single Image. ACM Transactions on Graphics (TOG) **27**(3) (2008) 4
41. Su, S., Delbracio, M., Wang, J.: Deep Video Deblurring for Hand-held Cameras. In: Computer Vision and Pattern Recognition (CVPR) (2017), <https://github.com/shuochsu/DeepVideoDeblurring> 4
42. Sun, C., Sun, M., Chen, H.: Direct Voxel Grid Optimization: Super-fast Convergence for Radiance Fields Reconstruction. In: Computer Vision and Pattern Recognition (CVPR) (2022), <https://sunset1995.github.io/dvgo/> 3
43. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: Computer Vision and Pattern Recognition (CVPR) (2018), <https://github.com/jiangsutx/SRN-Deblur> 4, 9, 10, 11, 13
44. Wang, C., Gao, D., Xu, K., Geng, J., Hu, Y., Qiu, Y., Li, B., Yang, F., Moon, B., Pandey, A., Aryan, Xu, J., Wu, T., He, H., Huang, D., Ren, Z., Zhao, S., Fu, T., Reddy, P., Lin, X., Wang, W., Shi, J., Talak, R., Cao, K., Du, Y., Wang, H., Yu, H., Wang, S., Chen, S., Kashyap, A., Bandaru, R., Dantu, K., Wu, J., Xie, L., Carbone, L., Hutter, M., Scherer, S.: PyPose: A library for robot learning with physics-based optimization. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023), <https://github.com/pypose/pypose> 16
45. Wang, P., Zhao, L., Ma, R., Liu, P.: Supplementary Material for BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields https://openaccess.thecvf.com/content/CVPR2023/supplemental/Wang_BAD-NeRF_Bundle_Adjusted_CVPR_2023_supplemental.pdf 12

46. Wang, P., Zhao, L., Ma, R., Liu, P.: BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields. In: Computer Vision and Pattern Recognition (CVPR) (2023), <https://wangpeng000.github.io/BAD-NeRF/> 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 19, 20
47. Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V.A.: NeRF--: Neural radiance fields without known camera parameters. arXiv preprint arXiv:2102.07064 (2021), <https://nerfmm.active.vision/> 3
48. Wizadwongsa, S., Phongthawee, P., Yenpraphai, J., Suwananakorn, S.: NeX: Real-time View Synthesis with Neural Basis Expansion. In: Computer Vision and Pattern Recognition (CVPR) (2021), <https://nex-mpi.github.io/> 3
49. Xu, L., Jia, J.: Two-Phase Kernel Estimation for robust motion deblurring. In: European Conference on Computer Vision (ECCV) (2010) 4
50. Yan, C., Qu, D., Wang, D., Xu, D., Wang, Z., Zhao, B., Li, X.: GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting. arXiv preprint arXiv:2311.11700 (2023) 15
51. Ye, V., Kanazawa, A.: Mathematical supplement for the gsplat library. arXiv preprint arXiv:2312.02121 (2023), <https://github.com/nerfstudio-project/gsplat> 16
52. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: PlenOctrees for Real-time Rendering of Neural Radiance Fields. In: International Conference on Computer Vision (ICCV) (2021), <https://alexyu.net/plenoctrees/> 2, 3
53. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage Progressive Image Restoration. In: Computer Vision and Pattern Recognition (CVPR) (2021), <https://github.com/swz30/MPRNet> 9, 10, 13
54. Zwicker, M., Pfister, H., Van Baar, J., Gross, M.: EWA Volume Splatting. In: Proceedings Visualization, 2001. VIS'01. pp. 29–538. IEEE (2001), https://cgl.ethz.ch/research/past_projects/surfels/ewavolumesplattting/index.html 6