

Deblur4DGS: 4D Gaussian Splatting from Blurry Monocular Video

Renlong Wu, Zhilu Zhang, Mingyang Chen, Xiaopeng Fan, Zifei Yan, Wangmeng Zuo
Harbin Institute of Technology

hirenlongwu@gmail.com, cszlzhang@outlook.com, youngmchan269@gmail.com,

fxp@hit.edu.cn, yanzifei@hit.edu.cn, wzmuzo@hit.edu.cn

<https://deblur4dgs.github.io/>

Abstract

Recent 4D reconstruction methods have yielded impressive results but rely on sharp videos as supervision. However, motion blur often occurs in videos due to camera shake and object movement, while existing methods render blurry results when using such videos for reconstructing 4D models. Although a few NeRF-based approaches attempted to address the problem, they struggled to produce high-quality results, due to the inaccuracy in estimating continuous dynamic representations within the exposure time. Encouraged by recent works in 3D motion trajectory modeling using 3D Gaussian Splatting (3DGS), we suggest taking 3DGS as the scene representation manner, and propose the first 4D Gaussian Splatting framework to reconstruct a high-quality 4D model from blurry monocular video, named Deblur4DGS. Specifically, we transform continuous dynamic representations estimation within an exposure time into the exposure time estimation. Moreover, we introduce exposure regularization to avoid trivial solutions, as well as multi-frame and multi-resolution consistency ones to alleviate artifacts. Furthermore, to better represent objects with large motion, we suggest blur-aware variable canonical Gaussians. Beyond novel-view synthesis, Deblur4DGS can be applied to improve blurry video from multiple perspectives, including deblurring, frame interpolation, and video stabilization. Extensive experiments on the above four tasks show that Deblur4DGS outperforms state-of-the-art 4D reconstruction methods. The codes are available at <https://github.com/ZcsrenlongZ/Deblur4DGS>.

1. Introduction

Substantial efforts have been made for 4D reconstruction, which has extensive applications in augmented reality and virtual reality. To model static scenes, Neural Radiance Field (NeRF) [52] and 3D Gaussian Splatting (3DGS) [27] propose implicit neural representation manner and explicit Gaussian ellipsoids one, respectively. To model dynamic

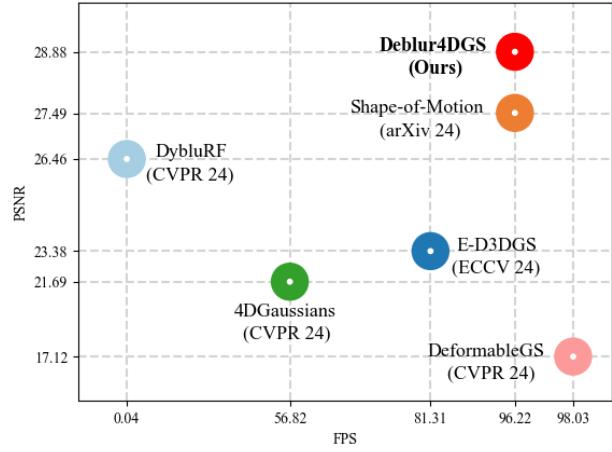


Figure 1. Performance and rendering speed comparisons for novel-view synthesis on the 720×1080 images. Deblur4DGS produces better results with real-time rendering speed.

objects, implicit neural fields [5, 48, 60, 84, 87, 90, 108] and explicit deformation [8, 10, 25, 30, 40, 44, 49, 79, 89] are suggested for motion representation. While achieving great progress, most methods rely on synchronized multi-view videos. They yield unsatisfactory results when applied to monocular video, where dynamic objects are only observed once at each timestamp. To alleviate the under-constrained nature of the problem, recent studies have introduced data-driven priors, such as depth maps [35, 88], optical flows [16, 17, 19, 35, 46, 76, 80, 108], tracks [36, 65, 70, 79], and generative models [8, 86, 91, 93] to improve 4D reconstruction.

Unfortunately, motion blur often arises in the videos due to camera shake and object movement, while the above methods heavily rely on sharp videos for supervision. When reconstructing the 4D scene from the blurry video, these methods can render blurry results. The first step to solving this problem is to deal with camera motion blur, which is relatively simple. Some NeRF-based [32, 33, 39, 50, 78] and 3DGS-based [7, 55, 58, 102] methods have suggested jointly optimizing 3D static scene representation and camera poses within the exposure time by calculating the reconstruction loss between the synthetic blurry images and

the input blurry frames. In contrast, the object motion blur is more challenging to address, as the solution has to estimate continuous and sharp dynamic representations within the exposure time to simulate blurry frames. As far as we know, only DybluRF [4, 72] attempted this with NeRF but struggled to produce high-quality and real-time rendering results, as shown in Fig. 1.

In this work, we propose taking 3DGS [27] as the scene representation manner to explore the problem, driven by two main motivations. First, the real-time rendering capability of 3DGS and its successful application in 4D reconstruction make this method highly promising. Second, the explicit representation property of 3DGS makes it easier to model 3D object motion than NeRF. In particular, recent advances [30, 65, 70, 79] in reconstructing 3D motion trajectories present an opportunity to simplify this continuous dynamic representations estimation problem. Based on this, the challenging problem can be transformed into simple exposure time estimation, without the extra motion trajectory modeling in DybluRF [4, 72]. With the estimated exposure time, continuous dynamic representations can be obtained by directly interpolating between representations at the nearest integer timestamps.

Specifically, we name it Deblur4DGS, which is the first Gaussian Splatting framework designed for 4D reconstruction from blurry monocular video. For the static scene, we present a camera motion predictor to learn camera poses at exposure start and end, which is jointly optimized with static Gaussians. For the dynamic objects, we optimize learnable exposure time parameters and dynamic Gaussians of the integer timestamps, simultaneously. Then continuous camera poses and dynamic Gaussians within exposure time can be obtained by interpolation, and they are used to render continuous sharp frames to calculate the reconstruction loss. Moreover, the exposure regularization term is introduced to avoid trivial solutions, while multi-frame and multi-resolution consistency regularization terms are used to alleviate artifacts. Furthermore, existing 4D reconstruction methods generally select Gaussians at a single timestamp as canonical Gaussians (*e.g.*, the ones at the first timestamp). However, it may produce results with missing details in scenes with large motion, especially when processing blurry videos with a low frame rate. To alleviate this issue, we suggest variable canonical Gaussians as time progresses based on the image blur level. Gaussians corresponding to the sharper frame are selected as the canonical ones for better blur removal, and each canonical Gaussian is only used for some nearby timestamps to reduce the difficulty of modeling large motion.

Blurry videos suffer from not only motion blur, but also low frame rates and scene shake generally. Beyond novel-view synthesis, the optimized Deblur4DGS can also be applied to address these problems, achieving deblurring,

frame interpolation, and video stabilization. We evaluate the 4D reconstruction quality of Deblur4DGS from all four perspectives. Extensive experiments demonstrate that Deblur4DGS outperforms state-of-the-art 4D reconstruction methods quantitatively and qualitatively while maintaining real-time rendering speed. Furthermore, Deblur4DGS has competitive capabilities in comparison with task-specific video processing models trained in a supervised manner.

The main contributions can be summarized as follows:

- We propose Deblur4DGS, the first 4D Gaussian Splatting framework specially designed to reconstruct a high-quality 4D model from blurry monocular video.
- We propose transforming dynamic representation estimation into exposure time estimation, where a series of regularizations are suggested to tackle the under-constrained optimization and blur-aware variable canonical Gaussians is present to better represent objects with large motion.
- Extensive experiments on novel-view synthesis, deblurring, frame interpolation, and video stabilization tasks show that Deblur4DGS significantly outperforms state-of-the-art 4D reconstruction methods.

2. Related Work

2.1. Image and Video Deblurring

Deep learning-based image [29, 37, 38, 43, 62, 73, 82, 92, 99, 101] and video [6, 23, 24, 41, 42, 54, 56, 66, 69, 83, 95, 96, 104–106] deblurring methods have been widely explored. Compared to image deblurring methods, video ones can leverage temporal clues between consecutive frames for more effective restoration. For example, DBLRNet [96] employs spatial-temporal 3D convolution for spatial and temporal features aggregation. DSTNet [56] develops a deep discriminative spatial and temporal network. BasicVSR++ [6] improves feature fusion with second-order feature propagation and flow-guided deformable alignment. BSSTNet [95] introduces a blur map to sufficiently utilize the entire video, achieving recent state-of-the-art. When reconstructing from a blurry video, pre-processing it with the 2D deblurring method is a straightforward manner. However, 2D deblurring methods cannot perceive 3D structures and maintain scene geometric consistency, leading to unsatisfactory scene reconstruction.

2.2. 3D and 4D Reconstruction

To reconstruct 3D models, NeRF [52] and 3DGS [27] introduce implicit neural representation manner and explicit Gaussian ellipsoids one respectively, where the latter generally achieves better results in both image quality and rendering speed. To reconstruct 4D models, most works [1, 5, 8, 10, 11, 13–15, 18, 20, 25, 30, 40, 44, 45, 48, 49, 51, 60, 67, 68, 74, 84, 87, 89, 90, 103] incorporate implicit neural fields and explicit deformation for

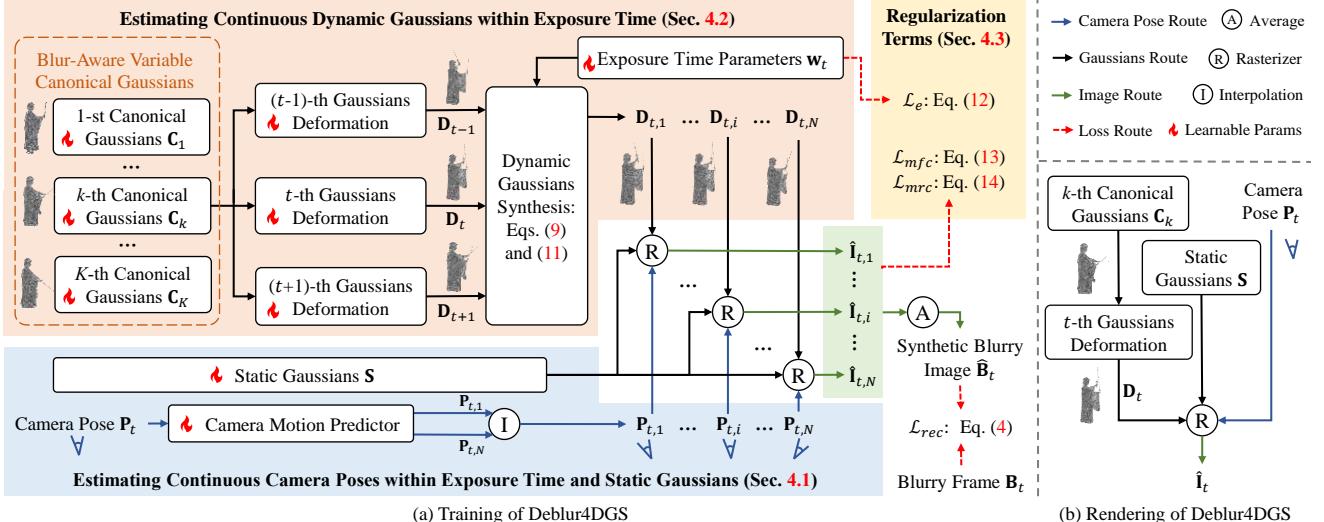


Figure 2. (a) Training of Deblur4DGS. When processing t -th frame, we first discretize its exposure time into N timestamps. Then, we estimate continuous camera poses $\{\mathbf{P}_{t,i}\}_{i=1}^N$ (see Sec. 4.1) and dynamic Gaussians $\{\mathbf{D}_{t,i}\}_{i=1}^N$ (see Sec. 4.2) within exposure time. Next, we render each latent sharp image $\mathbf{I}_{t,i}$ with the camera pose $\mathbf{P}_{t,i}$, dynamic Gaussians $\mathbf{D}_{t,i}$ and static Gaussians \mathbf{S} . Finally, $\{\hat{\mathbf{I}}_{t,i}\}_{i=1}^N$ are averaged to obtain the synthetic blurry image $\hat{\mathbf{B}}_t$, which is used to calculate the reconstruction loss \mathcal{L}_{rec} with the given blurry frame \mathbf{B}_t . To regularize the under-constrained optimization, we introduce exposure regularization \mathcal{L}_e , multi-frame consistency regularization \mathcal{L}_{mfc} and multi-resolution consistency regularization \mathcal{L}_{mrc} (see Sec. 4.3). (b) Rendering of Deblur4DGS. Deblur4DGS produces the sharp image with user-provided timestamp t and camera pose \mathbf{P}_t .

motion representation. For example, DeformableGS [90] achieves it with multi-layer perceptrons (MLPs). 4DGaussians [1] improves it with a multi-resolution HexPlane [5]. E-D3DGS [84] introduces per-Gaussian embeddings and temporal embeddings. Moreover, to better reconstruct from monocular video, some studies enhance 4D reconstruction with data-driven priors, such as depth maps [35, 88], optical flows [16, 17, 19, 35, 46, 76, 80, 108], tracks [36, 65, 70, 79], and generative models [8, 86, 91, 93, 98].

Note that these methods heavily rely on high-quality sharp videos for supervision and perform poorly when facing blurry inputs. To process camera motion in static areas, recent works [7, 31–34, 39, 50, 55, 58, 78, 102] suggest jointly optimizing the scene representation and recovering the camera poses within the exposure time. To process object motion blur in dynamic scenes, DybluRF [4, 72] incorporates object motion blur formation into dynamic model optimization but faces challenges in producing high-quality images and achieving real-time rendering. In this work, with 3DGS [27] as the scene representation manner, we develop Deblur4DGS to reconstruct a high-quality 4D model from a blurry video.

3. Preliminary

3.1. 4D Gaussian Splatting

A 3D Gaussian [27] is parameterized by $\{\mathbf{x}, \mathbf{r}, \mathbf{s}, \alpha, \mathbf{c}\}$, where \mathbf{x} characterizes the center position in the world space, rotation matrix \mathbf{r} and scale matrix \mathbf{s} define the shape, α is opacity, and spherical harmonics (SH) coefficients \mathbf{c} repre-

sent the view-dependent color.

In 4D Gaussian Splatting (4DGS), the static and dynamic regions are usually processed, respectively. Static regions can be represented by a set of 3D Gaussians, named \mathbf{S} . For the dynamic areas, 4DGS generally selects a timestamp (e.g., the first timestamp) and represents the objects by canonical dynamic Gaussians, i.e., \mathbf{C} . Then, \mathbf{C} is deformed to other timestamps by an implicit neural field [5, 48, 60, 84, 87, 90, 108] or explicit manner [8, 10, 25, 30, 40, 44, 49, 79, 89] for motion representation. Denote by \mathbf{D}_t the dynamic Gaussians at t -th timestamp, it can be written as,

$$\mathbf{D}_t = \mathcal{F}(\mathbf{C}, t; \Theta_{\mathcal{F}}). \quad (1)$$

\mathcal{F} is the deformation operation with parameters $\Theta_{\mathcal{F}}$. The Gaussians for t -th timestamp is the union of \mathbf{S} and \mathbf{D}_t .

Collectively, 4DGS models a scene with static Gaussians \mathbf{S} , canonical dynamic Gaussians \mathbf{C} , and a deformation operation \mathcal{F} . With the provided camera pose, the Gaussians at t -th timestamp \mathbf{D}_t can be projected into 2D spaces and rasterized to obtain the corresponding image.

3.2. Motion Blur Formation

Motion blur occurs due to camera shake and object movement, which can be simply regarded as the integration of a latent sharp image sequence [53], i.e.,

$$\mathbf{B}(u, v) = \phi \int_0^\tau \mathbf{I}_t(u, v) dt. \quad (2)$$

$\mathbf{B}(u, v) \in \mathbb{R}^{H \times W \times 3}$ is the blurry image and $\mathbf{I}_t(u, v)$ is the latent sharp one at t -th timestamp. (u, v) is pixel location, τ

is the camera exposure time, and ϕ is a normalization factor. To approximate the integral operation, recent works [72, 78, 102] divide the exposure time into N timestamps and regard the blurry image as the average of N sharp images, *i.e.*,

$$\mathbf{B}(u, v) \approx \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{I}_i(u, v). \quad (3)$$

In this work, we reconstruct a high-quality 4D model from a blurry video by incorporating the physical motion blur formation into model optimization.

4. Proposed Method

Let $\{\mathbf{B}_t\}_{t=1}^T$ and $\{\mathbf{M}_t\}_{t=1}^T$ denote a blurry video with T timestamps and the corresponding masks indicating dynamic areas, respectively. As shown in Fig. 2(a), when processing t -th frame, we first evenly divide its camera exposure time into N timestamps. Then, we estimate continuous camera poses $\{\mathbf{P}_{t,i}\}_{i=1}^N$ and dynamic Gaussians $\{\mathbf{D}_{t,i}\}_{i=1}^N$ to simulate camera shake and object movement. Next, we render each sharp image $\hat{\mathbf{I}}_{t,i}$ with the corresponding camera pose $\mathbf{P}_{t,i}$, dynamic Gaussians $\mathbf{D}_{t,i}$ and static Gaussians \mathbf{S} . After that, we average $\{\hat{\mathbf{I}}_{t,i}\}_{i=1}^N$ to obtain the synthetic blurry image $\hat{\mathbf{B}}_t$, which is used to calculate the reconstruction loss \mathcal{L}_{rec} with the given blurry frame \mathbf{B}_t , *i.e.*,

$$\mathcal{L}_{rec} = (1 - \beta)\mathcal{L}_1(\hat{\mathbf{B}}_t, \mathbf{B}_t) + \beta\mathcal{L}_{ssim}(\hat{\mathbf{B}}_t, \mathbf{B}_t). \quad (4)$$

\mathcal{L}_1 and \mathcal{L}_{ssim} are ℓ_1 loss and SSIM [81] loss, respectively. β is set to 0.2. The settings all follow 3DGS [27].

4.1. Continuous Camera Poses Estimation

To estimate continuous camera poses, recent methods [7, 55, 58, 102] directly optimize exposure start and end poses (*i.e.*, $\mathbf{P}_{t,1}$ and $\mathbf{P}_{t,N}$). Then, the linear interpolation is performed between $\mathbf{P}_{t,1}$ and $\mathbf{P}_{t,N}$ in the Lie-algebra of $\text{SE}(3)$ to obtain the camera pose at i -th intermediate timestamp (*i.e.*, $\mathbf{P}_{t,i}$). It can be written as,

$$\mathbf{P}_{t,i} = \mathbf{P}_{t,1} \odot \exp\left(\frac{i-1}{N-1} \odot \log\left(\frac{\mathbf{P}_{t,N}}{\mathbf{P}_{t,1}}\right)\right). \quad (5)$$

\exp and \log are exponential and logarithmic functions, respectively. \odot is a pixel-wise multiply operation.

We follow the manner but deploy a tiny MLP as the camera motion predictor (see details in *Suppl.*) for more stable optimization. We pre-train the camera motion predictor and static Gaussians \mathbf{S} with static reconstruction loss \mathcal{L}_{rec}^s , *i.e.*,

$$\mathcal{L}_{rec}^s = (1 - \beta)\mathcal{L}_1(\hat{\mathbf{B}}_t^s, \mathbf{B}_t^s) + \beta\mathcal{L}_{ssim}(\hat{\mathbf{B}}_t^s, \mathbf{B}_t^s). \quad (6)$$

$\hat{\mathbf{B}}_t^s = (1 - \mathbf{M}_t) \odot \hat{\mathbf{B}}_t$ and $\mathbf{B}_t^s = (1 - \mathbf{M}_t) \odot \mathbf{B}_t$ are the static areas of $\hat{\mathbf{B}}_t$ and \mathbf{B}_t , respectively.

4.2. Continuous Dynamic Gaussians Estimation

We first introduce blur-aware variable canonical Gaussians for better dynamic representation at integer timestamps.

Then, we describe Gaussian deformation manner. Finally, we detail how to take learnable exposure time parameters to obtain continuous dynamic Gaussians within exposure time.

Blur-Aware Variable Canonical Gaussians. Existing 4D reconstruction methods generally select a single canonical Gaussians \mathbf{C} across the entire video, which may produce results with missing details in scenes with large motion. To alleviate the issue, we suggest variable the canonical Gaussians as time progresses. In such case, the k -th canonical Gaussians \mathbf{C}_k is only used for some nearby timestamps thus reducing the difficulty of modeling large motion. One way to achieve this is to uniformly divide the video into L segments and select \mathbf{C}_k for k -th segment. Although it improves performance, selecting the one corresponding to the sharper frame is better for blur removal. In particular, we first uniformly divide the video into L segments and calculate the blur level b_t of dynamic areas for t -th frame following [2, 63]. It can be written as,

$$b_t = \sum_{(u,v) \in \mathbf{M}_t} (\Delta \mathbf{B}_t(u, v) - \overline{\Delta \mathbf{B}_t})^2, \quad (7)$$

where \mathbf{M}_t indicates dynamic areas. $\Delta \mathbf{B}_t$ is the image Laplacian and $\overline{\Delta \mathbf{B}_t}$ is its mean value. The larger b_t is, the sharper the frame is. To make the start and end frame of the segment as sharp as possible, we look for the sharp frame among their surrounding H frames and redefine them as the start and end of the current segment. After that, we select the Gaussians corresponding to the sharpest frame in each segment as its canonical ones. **Gaussian Deformation.** We follow Shape-of-Motion [79] to deform dynamic Gaussians, where a set of rigid transformation matrices are deployed. Let $\{\mathbf{x}_c, \mathbf{r}_c, \mathbf{s}, \mathbf{o}, \mathbf{c}\}$, $\{\mathbf{x}_t, \mathbf{r}_t, \mathbf{s}, \mathbf{o}, \mathbf{c}\}$, and $\{\mathbf{A}_t, \mathbf{E}_t\}$ denote a Gaussian in \mathbf{C}_k , the ones in \mathbf{D}_t , and the corresponding transformation matrix, respectively. It can be written as,

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_c + \mathbf{E}_t, \quad \mathbf{r}_t = \mathbf{A}_t \mathbf{r}_c. \quad (8)$$

Kindly refer to Shape-of-Motion [79] for details.

Interpolation with Exposure Time Parameters. To get continuous dynamic Gaussians $\{\mathbf{D}_{t,i}\}_{i=1}^N$, one straightforward way is to deploy a series of learnable Gaussian or deformation parameters, but it is unstable to optimize. With the explicit object motion representation in Eq. (8), $\mathbf{D}_{t,i}$ can be calculated by interpolating between the ones at the nearest integer timestamps, *i.e.*,

$$\begin{aligned} \mathbf{D}_{t,i} &= \mathbf{w}_{t,i} \odot \mathbf{D}_{t-1} + (1 - \mathbf{w}_{t,i}) \odot \mathbf{D}_t, \quad i \in [1, N/2], \\ \mathbf{D}_{t,i} &= (1 - \mathbf{w}_{t,i}) \odot \mathbf{D}_t + \mathbf{w}_{t,i} \odot \mathbf{D}_{t+1}, \quad i \in [N/2, N]. \end{aligned} \quad (9)$$

$\mathbf{w}_{t,i}$ is the normalized time interval between $\mathbf{D}_{t,i}$ and $\mathbf{D}_{t,N/2}$. Thus, the problem is transformed to estimate $\mathbf{w}_{t,i}$. In the implementation, we can estimate the one at exposure start and end (*i.e.*, $\mathbf{w}_{t,1}$ and $\mathbf{w}_{t,N}$) and then interpolate between them to get the i -th intermediate one $\mathbf{w}_{t,i}$, *i.e.*,

$$\mathbf{w}_{t,i} = (1 - \frac{i-1}{N-1}) \odot \mathbf{w}_{t,1} + \frac{i-1}{N-1} \odot \mathbf{w}_{t,N}. \quad (10)$$

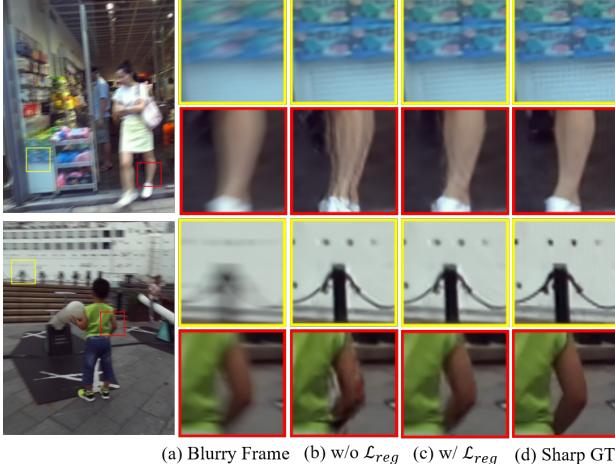


Figure 3. Illustration of artifacts in dynamic areas. After optimization with Eq. (4), static areas of $\hat{\mathbf{I}}_{t,i}$ are sharp while dynamic areas exist notable artifacts, as shown in the yellow and red boxes in (b), respectively. The proposed regularization terms \mathcal{L}_{reg} (see Eq. (15)) improve the visual effect, as shown in (c).

As the object motion within the exposure can be regarded as uniform, the absolute value of $\mathbf{w}_{t,1}$ and $\mathbf{w}_{t,N}$ are equal, which is half the exposure time \mathbf{w}_t . Thus, Eq. (10) can be re-written as,

$$\mathbf{w}_{t,i} = \left(1 - \frac{i-1}{N-1}\right) \odot \frac{\mathbf{w}_t}{2} + \frac{i-1}{N-1} \odot \left(-\frac{\mathbf{w}_t}{2}\right). \quad (11)$$

Thus, we set learnable parameters \mathbf{w}_t in Deblur4DGS for continuous dynamic Gaussians estimation within the exposure time. The canonical Gaussians, Gaussian deformation modules, and \mathbf{w}_t are jointly optimized. The reconstruction loss for dynamic areas is similar to Eq. (6).

4.3. Regularization Terms

After optimization with Eq. (4), static areas of $\hat{\mathbf{I}}_{t,i}$ are sharp while dynamic areas can with notable artifacts, as shown in the yellow and red boxes of Fig. 3(b) respectively. The reasons are below. (1) Note that multiple solutions exist for the model to fulfill Eq. (4). The most ideal one is that every $\hat{\mathbf{I}}_{t,i}$ is sharp, and the most trivial one is that every $\hat{\mathbf{I}}_{t,i}$ is as blurry as \mathbf{B}_t . (2) As static areas are consistent across the entire video, the model tends to learn the underlying sharp representation for inter-frame consistency. In other words, the inter-frame consistency implicitly regularizes model optimization. To further validate this, we conduct an experiment that removes the inter-frame consistency by reducing the number of frames to one. In such a case, the static areas are blurry after optimization with Eq. (4), which supports our confirmation. (3) Compared to static areas, the inter-frame consistency in dynamic ones is weaker due to object motion. It may provide insufficient regularization to guide sharp representation learning, thus leading to results with artifacts. To avoid this, we introduce some regularization terms, including exposure regularization \mathcal{L}_e , multi-frame

consistency regularization \mathcal{L}_{mfc} , and multi-resolution consistency regularization \mathcal{L}_{mrc} .

First, the continuous dynamic Gaussians $\{\mathbf{D}_{t,i}\}_{i=1}^N$ should not be the same. In other words, the value of exposure time parameters \mathbf{w}_t should not be too small. If \mathbf{w}_t is too small, $\mathbf{D}_{t,i}$ is nearly the same as \mathbf{D}_t , leading to trivial solutions. We constrain \mathbf{w}_t by \mathcal{L}_e , as,

$$\mathcal{L}_e = \max(0, \epsilon - \mathbf{w}_t). \quad (12)$$

\max is the maximum function and ϵ is a threshold.

Second, despite different motions, the content of multiple frames within exposure time should be similar. We utilize \mathcal{L}_{fc} to constrain consistency between neighbor frames, and that between each frame and the first frame, i.e.,

$$\begin{aligned} \mathcal{L}_{mfc} = & \frac{1}{N-1} \sum_{i=2}^N \left(\left\| \mathbf{M}_{t,i} \odot (\hat{\mathbf{I}}_{t,i-1 \rightarrow i} - \hat{\mathbf{I}}_{t,i}) \right\|_1 \right. \\ & \left. + \left\| \mathbf{M}_{t,1} \odot (\hat{\mathbf{I}}_{t,i \rightarrow 1} - \hat{\mathbf{I}}_{t,1}) \right\|_1 \right). \end{aligned} \quad (13)$$

$\hat{\mathbf{I}}_{t,i-1 \rightarrow i}$ and $\hat{\mathbf{I}}_{t,i \rightarrow 1}$ are obtained by aligning $\hat{\mathbf{I}}_{t,i-1}$ to $\hat{\mathbf{I}}_{t,i}$ and aligning $\hat{\mathbf{I}}_{t,i}$ to $\hat{\mathbf{I}}_{t,1}$ with a pre-trained optical flow network [71], respectively. $\mathbf{M}_{t,i}$ and $\mathbf{M}_{t,1}$ are masks that indicate dynamic areas in $\hat{\mathbf{I}}_{t,i}$ and $\hat{\mathbf{I}}_{t,1}$, respectively.

Third, the blur in the lower resolution is lower level and is easier to remove [28, 75], thus the artifacts are less in models trained with down-sampled blurry video. Taking advantage of this, we impose \mathcal{L}_{mrc} to assist the optimization of high-resolution models with results from low-resolution models. It can be written as,

$$\mathcal{L}_{mrc} = \|(\mathbf{M}_{t,i})_{\downarrow} \odot ((\hat{\mathbf{I}}_{t,i})_{\downarrow} - sg(\hat{\mathbf{I}}_{t,i}^l))\|_1. \quad (14)$$

$\hat{\mathbf{I}}_{t,i}^l$ is the rendered sharp image from the low-resolution model, which is pre-trained by taking the down-sampled video as supervision. $(\cdot)_{\downarrow}$ is an image down-sampling operation. sg is the stop-gradient operation.

Overall, the proposed regularization terms \mathcal{L}_{reg} can be denoted as,

$$\mathcal{L}_{reg} = \lambda_e \mathcal{L}_e + \lambda_{mfc} \mathcal{L}_{mfc} + \lambda_{mrc} \mathcal{L}_{mrc}. \quad (15)$$

λ_e , λ_{mfc} , and λ_{mrc} are set to 0.1, 2, 1, respectively. Besides, following Shape-of-Motion [79], we also use some other regularization terms \mathcal{L}_{oth} to help reconstruct 3D motion better, and the details can be seen in the *Suppl*.

4.4. Application to Multiple Tasks

The blurry videos suffer from not only motion blur, but also low frame rates and scene shake generally. Beyond novel-view synthesis, Deblur4DGS can adjust the camera poses and timestamps to address these problems, achieving video deblurring, frame interpolation, and video stabilization. First, when inputting camera poses of the blurry video, Deblur4DGS can render corresponding deblurring results. Second, when feeding the interpolated camera poses and timestamps, Deblur4DGS can produce frame-interpolated results. Third, Deblur4DGS can render a more stable video with the smoothed camera poses as inputs.

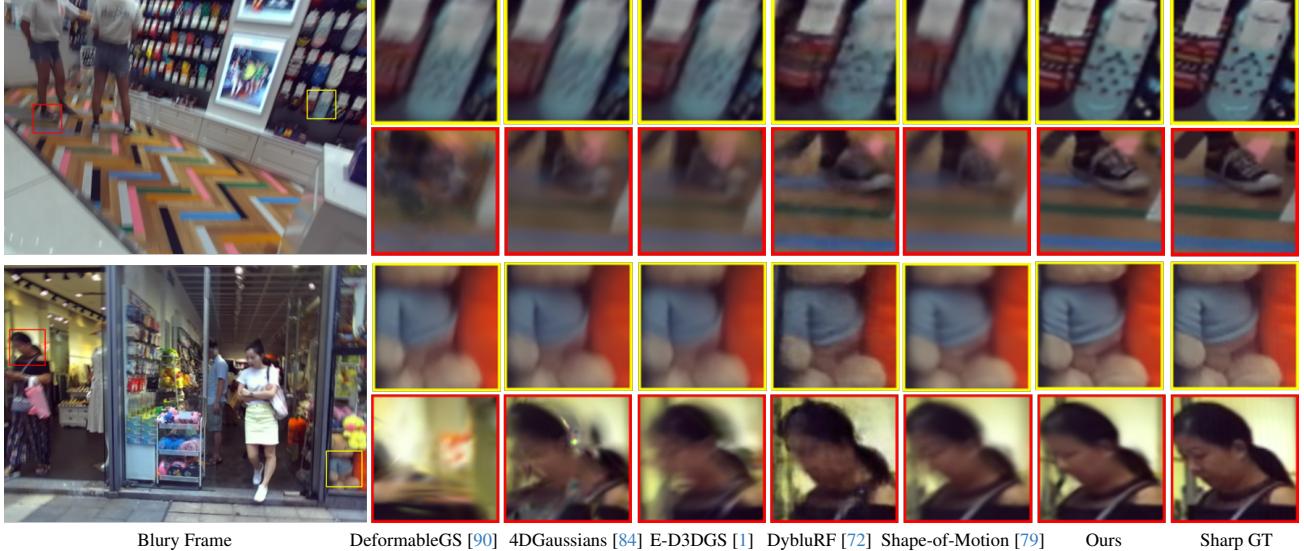


Figure 4. Visual comparisons of novel-view synthesis on the 720×1080 images. Our method produces more photo-realistic details in both static and dynamic areas, as marked with yellow and red boxes respectively.

5. Experiments

5.1. Experimental Settings

Datasets. We evaluate methods on 6 scenes with significant motion blur from Stereo Blur Dataset [107], where each one contains blurry stereo videos and the corresponding sharp ones. For each scene, we extract 24 frames from the original video and use COLMAP [64] to calibrate the camera extrinsic and intrinsic parameters, following DybluRF [72]. Note that DybluRF [72] conducts experiments on $\times 2.5$ down-sampled data. We employ both $\times 2.5$ down-sampled ones (*i.e.*, 288×512) and the original ones (*i.e.*, 720×1080).

Training Configurations. We use the blurry image sequences from the left camera for training. For stable optimization, we pre-train the camera motion predictor and static Gaussians S for 200 epochs. After that, we jointly optimize the camera motion predictor, S , canonical dynamic Gaussians $\{\mathbf{C}_k\}_{k=1}^K$, deformable operation \mathcal{F} and exposure time parameters $\{\mathbf{w}_t\}_{t=1}^T$ for 200 epochs. The learning rate for the camera motion predictor is set to 5×10^{-4} and decayed to 1×10^{-5} . The learning rate for $\{\mathbf{w}_t\}_{t=1}^T$ is set to 1×10^{-1} and decayed to 1×10^{-5} . The learning rate for S , $\{\mathbf{C}_k\}_{k=1}^K$ and \mathcal{F} are the same as Shape-of-Motion [79]. N is set to 11. L and H are set to 5 and 3 respectively. ϵ is set to 1.0. All experiments are conducted with PyTorch [57] on one Nvidia GeForce RTX A6000 GPU.

Evaluation Configurations. We evaluate methods on novel-view synthesis, deblurring, frame interpolation, and video stabilization tasks. To evaluate novel-view synthesis and deblurring performance, we use the sharp image sequences of the right and left cameras as ground truth respectively. Due to calibrated camera parameters are not absolutely accurate, the rendering results may be slightly

Table 1. Quantitative comparisons for **novel view synthesis**. \uparrow denotes the higher metric the better, and \downarrow denotes the lower one the better. The best is **boldfaced**, and the second is underlined.

Methods	PSNR \uparrow /SSIM \uparrow /LPIPS \downarrow	PSNR \uparrow /SSIM \uparrow /LPIPS \downarrow
	288×512	720×1080
DeformableGS [90]	17.82 / 0.722 / 0.292	17.12 / 0.751 / 0.363
4DGaussians [84]	21.94 / 0.833 / 0.180	21.69 / 0.852 / 0.238
E-D3DGS [1]	22.85 / 0.843 / 0.175	22.38 / 0.854 / 0.240
DyBluRF [72]	27.09 / 0.919 / <u>0.090</u>	26.46 / 0.913 / <u>0.142</u>
Shape-of-Motion [79]	<u>27.89</u> / <u>0.928</u> / 0.116	27.49 / 0.922 / 0.192
Deblur4DGS (Ours)	28.92 / 0.949 / 0.060	28.88 / 0.947 / 0.098

Table 2. Quantitative results that pre-process the blurry video with an image or video deblurring method before 4D reconstruction. ‘None’ denotes no image or video deblurring method being used.

Reconstruction Methods	Deblurring Methods	PSNR \uparrow / SSIM \uparrow / LPIPS \downarrow	PSNR \uparrow / SSIM \uparrow / LPIPS \downarrow
		288×512	720×1080
E-D3DGS [1]	None	22.85 / 0.843 / 0.175	22.38 / 0.854 / 0.240
	Restormer [92]	23.55 / 0.854 / 0.142	22.77 / <u>0.867</u> / 0.186
	DSTNet [56]	23.21 / <u>0.863</u> / <u>0.132</u>	<u>22.80</u> / 0.865 / <u>0.177</u>
	BSSTNet [95]	<u>23.28</u> / 0.857 / 0.128	23.08 / <u>0.876</u> / 0.163
Shape-of-Motion [79]	None	27.89 / 0.928 / 0.116	27.49 / 0.922 / 0.192
	Restormer [92]	28.39 / 0.941 / 0.080	28.25 / 0.940 / 0.117
	DSTNet [56]	28.07 / 0.938 / 0.072	28.22 / 0.941 / 0.104
	BSSTNet [95]	<u>28.34</u> / 0.940 / <u>0.079</u>	<u>28.24</u> / <u>0.943</u> / 0.100
Deblur4DGS (Ours)	None	28.92 / 0.949 / 0.060	28.88 / <u>0.947</u> / <u>0.098</u>

spatially misaligned with the ground truth. We adopt the aligned PSNR [22], SSIM [81] and LPIPS [97] as evaluation metrics, following similar works [3, 12, 85]. Specifically, we align rendering results to the ground truth by a pre-trained optical flow network (*i.e.*, PWC-Net [71]) and then calculate metrics between the aligned results and the ground truth. Simultaneously, we report results that calculate metrics between the original rendering images and the ground truth in the *Suppl*. As there is no ground truth for frame interpolation and video stabilization tasks, we evaluate method performance with recent no-reference metrics,

Table 3. Quantitative comparisons for **deblurring**. The first category is supervised image or video deblurring methods, and the second is 4D reconstruction-based ones. The best one in each category is **boldfaced**, and the second one is underlined.

Methods	PSNR↑/SSIM↑/LPIPS↓	PSNR↑/SSIM↑/LPIPS↓
	288 × 512	720 × 1080
Restormer [92]	<u>33.33 / 0.974 / 0.043</u>	<u>32.43 / 0.962 / 0.064</u>
DSTNet [56]	<u>33.13 / 0.971 / 0.039</u>	<u>32.42 / 0.962 / 0.059</u>
BSSTNet [95]	34.13 / 0.980 / 0.027	33.63 / 0.973 / 0.039
DeformableGS [90]	30.92 / 0.949 / 0.103	29.11 / 0.926 / 0.191
4DGaussians [84]	<u>31.25 / 0.949 / 0.097</u>	<u>29.95 / 0.932 / 0.169</u>
E-D3DGS [1]	<u>31.28 / 0.949 / 0.098</u>	<u>29.98 / 0.931 / 0.170</u>
DyBluRF [72]	29.87 / 0.947 / 0.054	29.23 / 0.943 / 0.105
Shape-of-Motion [79]	29.83 / 0.946 / 0.107	28.93 / 0.932 / 0.190
Deblur4DGS (Ours)	32.55 / 0.975 / 0.041	31.60 / 0.965 / 0.084

i.e., CLIPQA [77] and MUSIQ [26].

5.2. Comparison with State-of-the-Art Methods

We compare Deblur4DGS with 5 state-of-the-art 4D reconstruction methods (*i.e.*, DeformableGS [90], 4DGaussians [84], E-D3DGS [1], DyBluRF [72], and Shape-of-Motion [79]) on two resolution (*i.e.*, 228 × 512 and 720 × 1080) datasets. We input poses of right and left cameras to produce novel-view synthesis and deblurring results, respectively. We linearly interpolate the left camera poses and timestamps to generate ×16 frame interpolation results. We employ a Gaussian filter to smooth left camera poses for video stabilization, following [59]. The results for novel-view synthesis, deblurring, frame interpolation, and video stabilization are summarized in Tab. A, Tab. B, Tab. 4, and Tab. 5, respectively. Besides, rendering speed comparisons on 720 × 1080 images are reported in Fig. 1.

Novel-view synthesis Comparisons. As shown in Tab. A, among all competing methods, DyBluRF [72] gets the best LPIPS score, while Shape-of-Motion [79] achieves the best PSNR and SSIM scores. Benefiting the explicit 3D representation manner and elaborate design for processing motion blur, our Deblur4DGS obtains all the best scores on both datasets. In addition, Deblur4DGS removes blur more clearly and produces more photo-realistic details in both static and dynamic areas, as shown in Fig. 4. More visual comparisons can be seen in the *Suppl.*

In addition, to further demonstrate the effectiveness of Deblur4DGS, we first pre-process the blurry videos with state-of-the-art image [92] or video [56, 95] deblurring methods and then perform 4D reconstruction. The results are summarized in Tab. 2. Compared with reconstruction from blurry videos, the incorporation of deblurring models improves performance. This is because the deblurring models remove some blur, facilitating sharp scene reconstruction. However, as the image and video deblurring methods cannot perceive 3D structure and maintain scene geometric consistency, the scene reconstruction results are still unsatisfactory. In contrast, Deblur4DGS jointly reconstructs

Table 4. Quantitative comparisons for **frame interpolation**. The first category is supervised video frame interpolation methods, and the second is 4D reconstruction-based ones.

Methods	CLIPQA↑/MUSIQ↑	CLIPQA↑/MUSIQ↑
	288 × 512	720 × 1080
RIFE [21]	0.179 / <u>44.090</u>	0.147 / 32.637
EMAVFI [94]	<u>0.181</u> / 42.773	<u>0.170</u> / 32.564
VIDUE [66]	0.267 / 60.854	0.185 / 49.560
DeformableGS [90]	0.172 / 41.603	0.167 / 31.205
4DGaussians [84]	0.177 / 44.132	0.177 / 33.836
E-D3DGS [1]	<u>0.178</u> / 43.250	0.181 / 32.945
DyBluRF [72]	<u>0.166</u> / <u>51.513</u>	<u>0.132</u> / <u>36.768</u>
Shape-of-Motion [79]	0.176 / 43.020	0.202 / 33.817
Deblur4DGS (Ours)	0.200 / 54.043	0.213 / 42.139

Table 5. Quantitative comparisons for **video stabilization**. The first category is 2D video stabilization methods, and the second is 4D reconstruction-based ones.

Methods	CLIPQA↑/MUSIQ↑	CLIPQA↑/MUSIQ↑
	288 × 512	720 × 1080
MeshFlow [47]	<u>0.140</u> / <u>38.289</u>	<u>0.116</u> / 34.140
NNDVS [100]	0.186 / 45.613	0.137 / <u>33.269</u>
DeformableGS [90]	0.170 / 40.846	0.168 / 30.971
4DGaussians [84]	0.145 / 44.673	0.172 / 34.416
E-D3DGS [1]	<u>0.181</u> / 44.364	<u>0.182</u> / 33.828
DyBluRF [72]	0.164 / <u>53.115</u>	0.132 / <u>37.556</u>
Shape-of-Motion [79]	0.174 / 43.293	0.202 / 33.750
Deblur4DGS (Ours)	0.199 / 54.299	0.211 / 42.241

scene geometry and processes motion blur in 3D space, achieving better scene reconstruction results.

Deblurring Comparisons. Apart from 4D reconstruction-based methods, we additionally compare with some state-of-the-art image [92] and video [56, 95] deblurring ones. As shown in Tab. B, Deblur4DGS obtains better results than 4D reconstruction-based methods and comparable ones to the deblurring-specific (*i.e.*, Restormer [92], DSTNet [56], and BSSTNet [95]) methods. Compared with the former, Deblur4DGS better reconstructs the scene, thus performing better. Note that the latter ones are trained on large paired data in a fully supervised manner while Deblur4DGS is optimized only with the given blurry video in a self-supervised manner. Although the large data prior makes these methods perform better, Deblur4DGS is more convenient to use without paired data collection.

Frame Interpolation Comparisons. We compare Deblur4DGS with 4D reconstruction-based methods and some state-of-the-art supervised video frame interpolation ones (*i.e.*, RIFE [21], EMAVFI [94], and VIDUE [66]), as shown in Tab. 4. Deform4DGS generally performs satisfactorily. Note that VIDUE [66] is specially designed and trained with large paired data in a supervised manner for joint deblurring and frame Interpolation, thus achieving better results.

Video Stabilization Comparisons. The results are summarized in Tab. 5. Benefiting from the better geometry reconstruction, Deblur4DGS achieves pleasant scores compared with both 2D video stabilization methods (*i.e.*, Mesh-

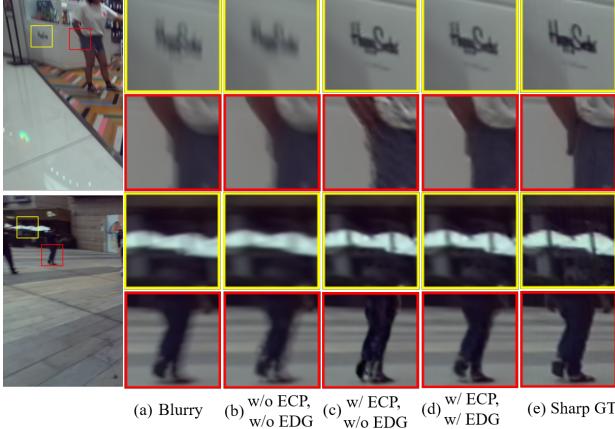


Figure 5. Effect about the estimation of continuous camera poses (ECP) and dynamic Gaussians (EDG). Static areas are marked with yellow boxes while dynamic ones are marked with red boxes. Without ECP and EDG, the results are almost as blurry as the input frame, as shown in (b). Only with ECP, static areas are sharp but may lead to visual artifacts in dynamic areas, as shown in (c). Further introducing EDG improves this, as shown in (d).

Flow [47] and NNDVS [100]) and 4D reconstruction-based ones, which further demonstrates its effectiveness.

6. Ablation Study

We first discuss the estimation of continuous camera poses (ECP) and dynamic Gaussians (EDG). Then, we validate the effect of regularization terms, as well as the blur-aware variable (BAV) canonical Gaussians.

6.1. Effect of ECP and EDG

ECP and EDG are introduced to process camera motion blur and object motion blur, respectively. Visual comparisons are shown in Fig. 5, and quantitative results are in the *Suppl.* First, without ECP and EDG, the results are almost as blurry as the input frame, as shown in Fig. 5(b). Second, only with ECP, the static areas are sharp but may lead to visual artifacts in dynamic areas, as shown in Fig. 5(c). It is because ECP cannot simulate the object movement. Third, we further introduce EDG to simulate that produces visually pleasant results in both areas, as shown in Fig. 5(d).

6.2. Effect of Regularization Terms

The effect of exposure regularization \mathcal{L}_e (see Eq. (12)), multi-frame consistency regularization \mathcal{L}_{mfc} (see Eq. (13)) and multi-resolution consistency regularization \mathcal{L}_{mrc} (see Eq. (14)) are shown in Tab. D. Without these regularization terms, the performance drops significantly and the visual results in Fig. 3 show that dynamic areas are with notable artifacts. By regularizing the object motion within the exposure time distinguished, \mathcal{L}_e alleviates the trivial solutions and improves performance. Besides, \mathcal{L}_{mfc} and \mathcal{L}_{mrc} ad-

Table 6. Ablation studies on regularization terms (see Eq. (15)).

\mathcal{L}_e	\mathcal{L}_{mfc}	\mathcal{L}_{mrc}	PSNR↑ / SSIM↑ / LPIPS↓
✗	✗	✗	28.69 / 0.947 / 0.063
✗	✓	✓	28.86 / 0.949 / 0.060
✓	✗	✗	28.73 / 0.947 / 0.060
✓	✓	✗	28.87 / 0.948 / 0.060
✓	✗	✓	28.86 / 0.948 / 0.060
✓	✓	✓	28.92 / 0.949 / 0.060

Table 7. Effect of blur-aware variable (BAV) canonical Gaussians. ‘None’ denotes selecting a single one across the entire video.

Methods	PSNR↑ / SSIM↑ / LPIPS↓
None	28.67 / 0.946 / 0.065
w/o Blur-Aware	28.85 / 0.949 / 0.062
Ours	28.92 / 0.949 / 0.060

ditionally regularize multi-frame and multi-resolution consistency respectively, helping to alleviate artifacts. Their combinations perform best.

6.3. Effect of BAV Canonical Gaussians.

The effect of BAV canonical Gaussians is summarized in Tab. E. First, selecting a single canonical Gaussians across the entire video (*i.e.*, None) leads to poor performance, due to the challenge of modeling large object motion. Second, selecting variable canonical Gaussians uniformly (*i.e.*, w/o Blur-Aware) alleviates the challenge, thus leading to performance improvement. We also experiment with an optical flow-based strategy [67] to select the canonical Gaussians, and it achieves a similar performance to the uniform selection. This may be due to the inaccurate estimation of optical flow from blurry images. Third, our blur-aware selection is better, as the canonical Gaussians corresponding to the sharper frame help blur removal.

7. Conclusions

In this work, we propose Deblur4DGS, the first 4D Gaussian Splatting framework to reconstruct a high-quality 4D model from blurry monocular video. In particular, with the explicit motion trajectory modeling based on 3D Gaussian Splatting, we propose to transform the challenging continuous dynamic representation estimation within an exposure time into the exposure time estimation, where a series of regularizations are suggested to tackle the under-constrained optimization. Besides, a blur-aware variable canonical Gaussians is present to represent objects with large motion better. Beyond novel-view synthesis, Deblur4DGS can improve blurry video quality from multiple perspectives, including deblurring, frame interpolation, and video stabilization. Extensive results show Deblur4DGS outperforms state-of-the-art 4D reconstruction methods.

Deblur4DGS: 4D Gaussian Splatting from Blurry Monocular Video

Supplementary Material

The content of the supplementary material involves:

- Structure of camera motion predictor in Sec. A.
- Other regularization terms in Sec. B.
- More result comparisons in Sec. C.
- Effect of ECP and EDG in Sec. D.
- Effect of regularization terms in Sec. E.
- Effect of BAV canonical Gaussians in Sec. F.

A. Structure of Camera Motion Predictor

The structure of camera motion predictor is provided in Fig. A. It first embeds the camera pose \mathbf{P}_t to a higher dimensional space, following [52]. Then, three FC blocks are stacked, each consisting of an FC layer followed by a ReLU operation. Finally, we deploy two heads to predict the camera pose at exposure start and end (*i.e.*, $\mathbf{P}_{t,1}$ and $\mathbf{P}_{t,N}$), respectively.

B. Other Regularization Terms

Following Shape-of-Motion [79], we use some other regularization terms \mathcal{L}_{oth} to help reconstruct 3D motion better, including mask regularization \mathcal{L}_{mask} , 2D tracks regularization \mathcal{L}_{track} and distance-preserving regularization \mathcal{L}_{rigid} .

Specifically, we render the masks within the exposure time $\{\hat{\mathbf{M}}_{t,i}\}_{i=1}^N$ to indicate dynamic areas. To supervise the training, we synthesize the mask $\hat{\mathbf{M}}_t^B$ for the synthetic blurry image $\hat{\mathbf{B}}_t$ as,

$$\hat{\mathbf{M}}_t^B(u, v) = \max\{\hat{\mathbf{M}}_{t,1}(u, v), \hat{\mathbf{M}}_{t,2}(u, v), \dots, \hat{\mathbf{M}}_{t,N}(u, v)\}. \quad (\text{A})$$

(u, v) is the pixel location. The mask regularization \mathcal{L}_{mask} can be written as,

$$\mathcal{L}_{mask} = \mathcal{L}_1(\hat{\mathbf{M}}_t^B, \mathbf{M}_t), \quad (\text{B})$$

where \mathbf{M}_t is the mask that obtained by applying SAM2 [61] to the ground truth blurry frame. Besides, we render the 2D tracks $\hat{\mathbf{U}}_{t \rightarrow t'}$ from a pair of randomly sampled query time t and target time t' . We supervise it by the lifted long-range 2D tracks $\mathbf{U}_{t \rightarrow t'}$ that are extracted from TAPIR [9], *i.e.*,

$$\mathcal{L}_{track} = \mathcal{L}_1(\hat{\mathbf{U}}_{t \rightarrow t'}, \mathbf{U}_{t \rightarrow t'}). \quad (\text{C})$$

Finally, we enforce a distance-preserving loss \mathcal{L}_{rigid} between randomly sampled dynamic Gaussians and their J -nearest neighbors. Let \mathbf{x}_t and $\mathbf{x}_{t'}$ denote the position of a

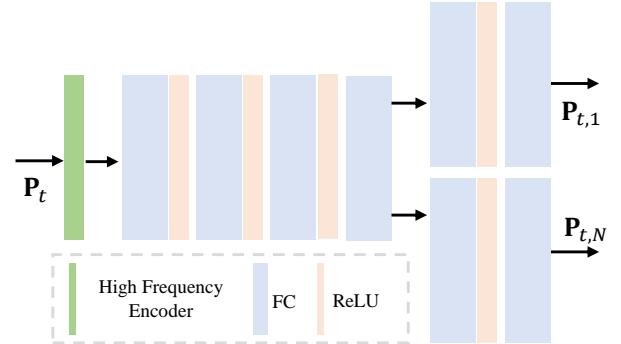


Figure A. Structure of camera motion predictor.

Gaussian at time t and t' , and $\mathcal{C}_J(\mathbf{x}_t)$ denote the set of J -nearest neighbors of \mathbf{x}_t . \mathcal{L}_{rigid} can be written as,

$$\mathcal{L}_{rigid} = \|\text{dist}(\hat{\mathbf{x}}_t, \mathcal{C}_J(\mathbf{x}_t)) - \text{dist}(\hat{\mathbf{x}}_{t'}, \mathcal{C}_J(\mathbf{x}_{t'}))\|_2^2. \quad (\text{D})$$

dist measures the Euclidean distance.

Overall, \mathcal{L}_{oth} can be written as,

$$\mathcal{L}_{oth} = \lambda_{mask} \mathcal{L}_{mask} + \lambda_{track} \mathcal{L}_{track} + \lambda_{rigid} \mathcal{L}_{rigid}. \quad (\text{E})$$

λ_{mask} , λ_{track} and λ_{rigid} are set to 1, 2, and 2, respectively.

C. More Result Comparisons

Due to calibrated camera parameters are not absolutely accurate, the rendering results may be slightly spatially misaligned with the ground truth. In the main text, we calculate the metrics (*i.e.*, PSNR [22], SSIM [81] and LIPIS [97]) between the aligned rendering results and the ground truth. Here, we provide the novel-view synthesis and deblurring comparisons that directly calculate metrics between the rendering results and the ground truth, as shown in Tab. A and Tab. B respectively. It shows that Deblur4DGS still outperforms state-of-the-art 4D reconstruction-based methods. More visual results for novel view synthesis in Fig. B show that our method produces more photo-realistic details. Moreover, visual results for deblurring in Fig. C show that compared with 4D reconstruction methods, Deblur4DGS produces sharper contents and fewer visual artifacts. Furthermore, the frame interpolation and video stabilization results, as well as some novel view synthesis videos are shown in the [site](#).

D. Effect of ECP and EDG

The visual results are provided in the main text. Here we provide the quantitative results in Tab. C. Without ECP and

EDG, the reconstruction results are unsatisfactory due to motion blur. ECP and EDG enable processing of the camera motion blur and object motion blur respectively, thus leading to performance improvement.

E. Effect of regularization terms

We validate the effect of exposure regularization \mathcal{L}_e , multi-frame consistency regularization \mathcal{L}_{mfc} and multi-resolution consistency regularization \mathcal{L}_{mrc} . In the main text, we provide the results on 288×512 data. Here, we provide that on 720×1080 data in Tab. D. The results are consistent for both data. \mathcal{L}_e regularizes the object motion within the exposure time distinguished, alleviating the trivial solutions and improving performance. \mathcal{L}_{mfc} and \mathcal{L}_{mrc} regularize multi-frame and multi-resolution consistency respectively, helping to alleviate artifacts. Their combinations perform best.

F. Effect of BAV canonical Gaussians

Tab. E additionally provides the quantitative results on 720×1080 data, which are consistent with the ones on 288×512 data. Compared with selecting a single one across the entire video (*i.e.*, None), selecting variable canonical Gaussians uniformly (*i.e.*, w/o Blur-Aware) alleviates the challenge of modeling large object motion, thus leading to performance improvement. Our method selects the canonical Gaussians corresponding to the sharper frame, which is beneficial to blur removal and performs better.

Table A. Quantitative comparisons for **novel view synthesis**. The metrics are directly calculated between rendering results and ground truth. The best is **boldfaced**, and the second is underlined.

Methods	PSNR↑/SSIM↑/LPIPS↓	PSNR↑/SSIM↑/LPIPS↓
	288×512	720×1080
DeformableGS [90]	16.32 / 0.644 / 0.356	16.06 / 0.711 / 0.424
4DGaussians [84]	20.58 / 0.795 / 0.186	20.31 / 0.821 / 0.258
E-D3DGS [1]	20.99 / 0.804 / 0.179	20.95 / 0.829 / 0.257
DyBluRF [72]	23.39 / 0.823 / <u>0.088</u>	22.59 / 0.815 / <u>0.168</u>
Shape-of-Motion [79]	<u>26.47</u> / <u>0.908</u> / 0.104	<u>25.76</u> / <u>0.896</u> / 0.195
Deblur4DGS (Ours)	26.69 / 0.918 / 0.052	25.93 / 0.904 / 0.102

Table B. Quantitative comparisons for **deblurring**. The metrics are directly calculated between rendering results and ground truth. The first category is supervised image or video deblurring methods, and the second is 4D reconstruction-based ones.

Methods	PSNR↑/SSIM↑/LPIPS↓	PSNR↑/SSIM↑/LPIPS↓
	288×512	720×1080
Restormer [92]	<u>36.74</u> / <u>0.987</u> / 0.019	35.97 / <u>0.981</u> / 0.028
DSTNet [56]	<u>36.49</u> / <u>0.986</u> / <u>0.015</u>	35.24 / 0.977 / <u>0.027</u>
BSSTNet [95]	36.88 / 0.988 / 0.013	35.87 / 0.983 / 0.021
DeformableGS [90]	30.97 / 0.950 / 0.096	28.95 / 0.924 / 0.189
4DGaussians [84]	31.30 / <u>0.951</u> / 0.090	<u>29.76</u> / 0.929 / 0.167
E-D3DGS [1]	<u>31.34</u> / <u>0.951</u> / 0.091	29.74 / 0.929 / 0.170
DyBluRF [72]	29.32 / 0.942 / <u>0.049</u>	28.67 / 0.936 / 0.101
Shape-of-Motion [79]	29.51 / 0.945 / 0.095	28.35 / 0.927 / 0.187
Deblur4DGS (Ours)	31.77 / 0.972 / 0.033	30.48 / 0.956 / 0.081

Table C. Effect about the estimation of continuous camera poses (ECP) and dynamic Gaussians (EDG).

ECP	EDG	PSNR↑/SSIM↑/LPIPS↓	PSNR↑/SSIM↑/LPIPS↓
		288×512	720×1080
✗	✗	28.00 / 0.929 / 0.119	27.70 / 0.923 / 0.191
✓	✗	28.86 / 0.948 / 0.062	28.68 / 0.945 / 0.100
✓	✓	28.92 / 0.949 / 0.060	28.88 / 0.947 / 0.098

Table D. Ablation studies on regularization terms.

\mathcal{L}_e	\mathcal{L}_{mfc}	\mathcal{L}_{mrc}	PSNR↑/SSIM↑/LPIPS↓	PSNR↑/SSIM↑/LPIPS↓
			288×512	720×1080
✗	✗	✗	28.69 / 0.947 / 0.063	28.70 / 0.946 / 0.099
✗	✓	✓	28.86 / 0.949 / 0.060	28.86 / 0.947 / 0.098
✓	✗	✗	28.73 / 0.947 / 0.060	28.74 / 0.945 / 0.099
✓	✓	✗	28.87 / 0.948 / 0.060	28.78 / 0.946 / 0.098
✓	✗	✓	28.86 / 0.948 / 0.060	28.79 / 0.946 / 0.099
✓	✓	✓	28.92 / 0.949 / 0.060	28.88 / 0.947 / 0.098

Table E. Effect of blur-aware variable (BAV) canonical Gaussians. ‘None’ denotes selecting a single one across the entire video.

Methods	PSNR↑/SSIM↑/LPIPS↓	PSNR↑/SSIM↑/LPIPS↓
	288×512	720×1080
None	28.67 / 0.946 / 0.065	28.66 / 0.945 / 0.104
w/o Blur-Aware	28.85 / 0.949 / 0.062	28.83 / 0.947 / 0.098
Ours	28.92 / 0.949 / 0.060	28.88 / 0.947 / 0.098

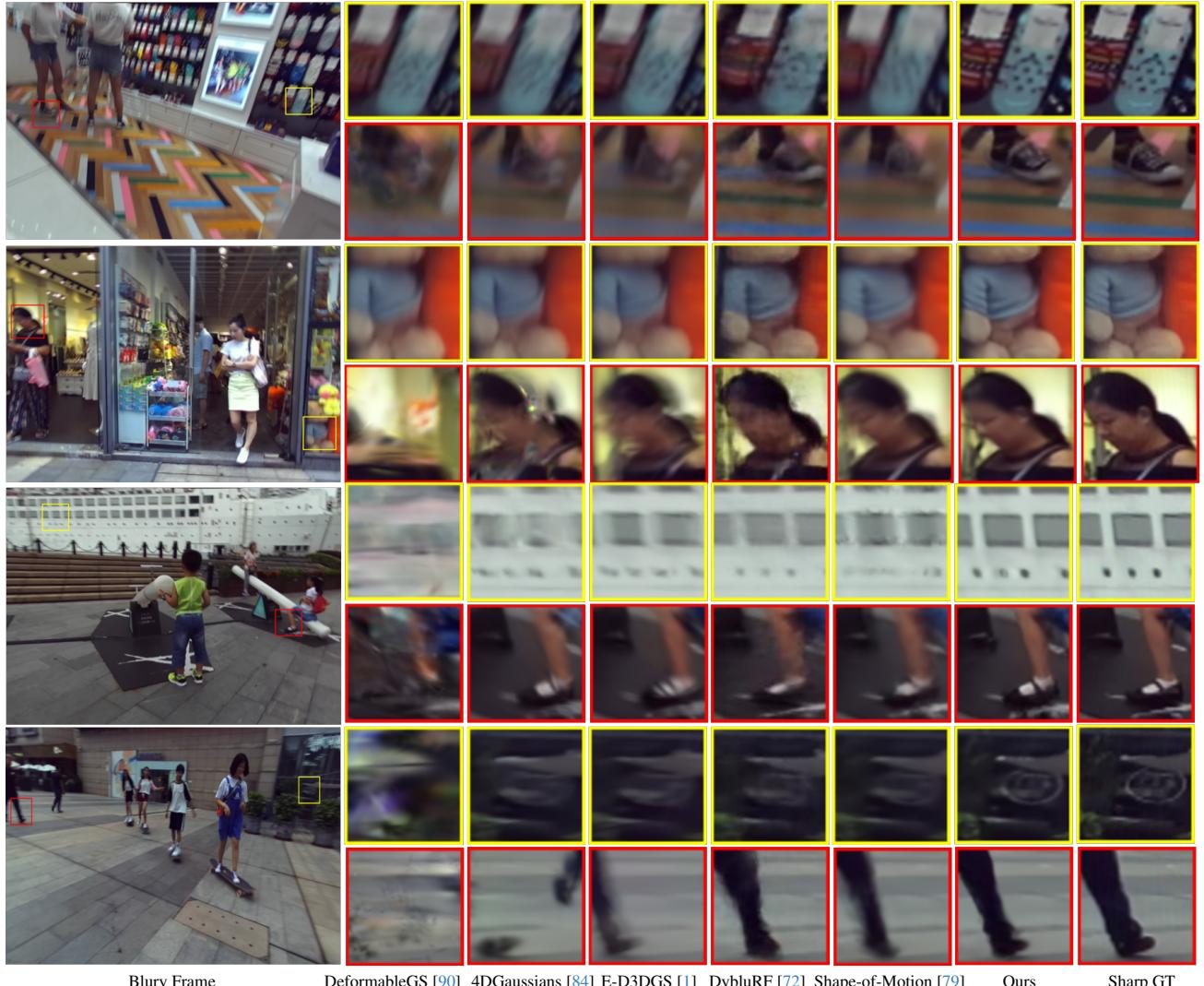


Figure B. Visual comparisons of novel-view synthesis on the 720×1080 images. Our method produces more photo-realistic details in both static and dynamic areas, as marked with yellow and red boxes respectively.

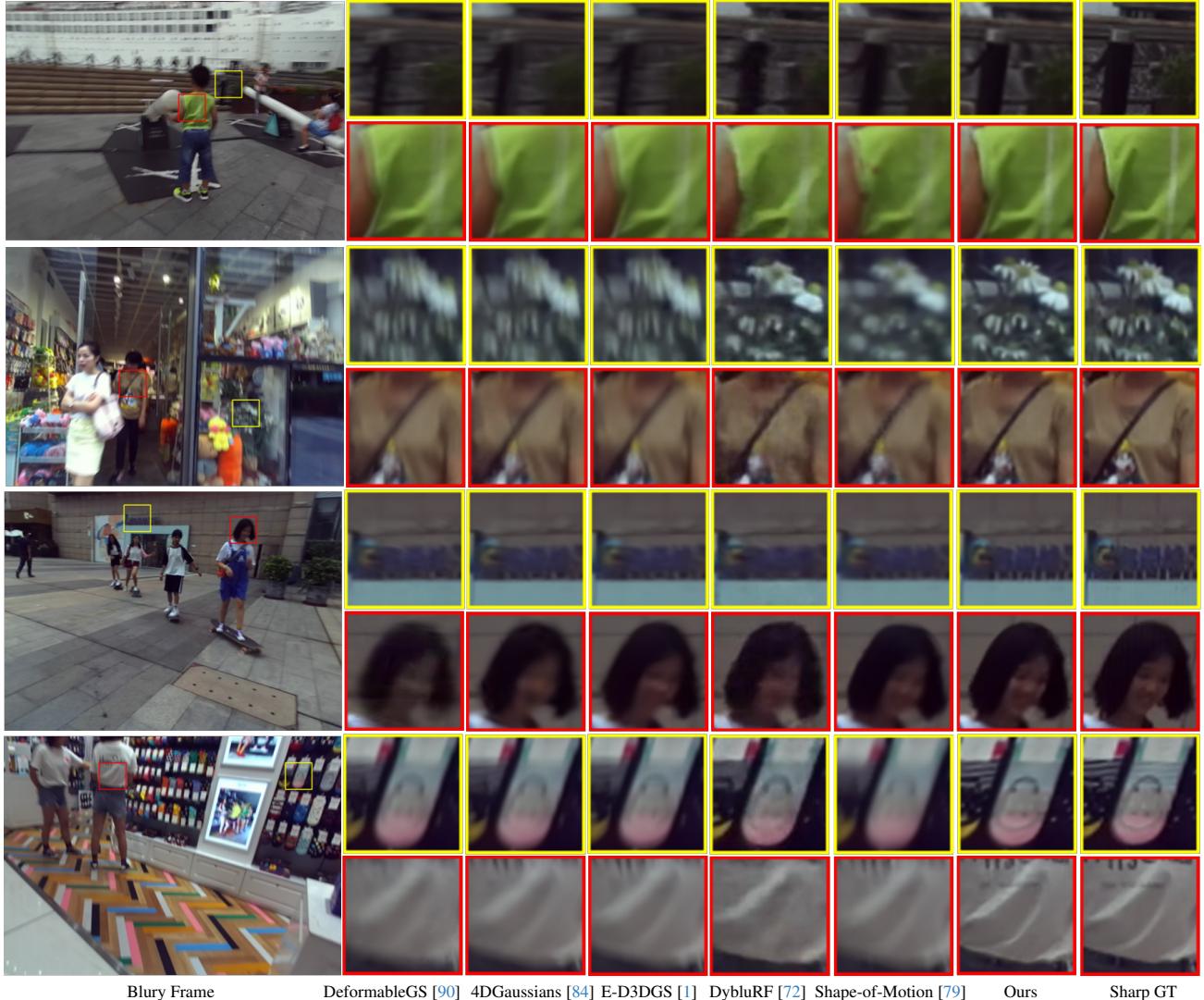


Figure C. Visual comparisons of deblurring on the 720×1080 images. Compared with 4D reconstruction-based methods, Deblur4DGS produces sharper contents and fewer artifacts in both static and dynamic areas, as marked with yellow and red boxes respectively.

References

- [1] Jeongmin Bae, Seoha Kim, Youngsik Yun, Hahyun Lee, Gun Bang, and Youngjung Uh. Per-gaussian embedding-based deformation for deformable 3d gaussian splatting. *arXiv preprint arXiv:2404.03613*, 2024. [2](#), [3](#), [6](#), [7](#), [10](#), [11](#), [12](#)
- [2] Raghav Bansal, Gaurav Raj, and Tanupriya Choudhury. Blur image detection using laplacian operator and open-cv. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, pages 63–67. IEEE, 2016. [4](#)
- [3] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *CVPR*, 2021. [6](#)
- [4] Minh-Quan Viet Bui, Jongmin Park, Jihyong Oh, and Munchurl Kim. Dyblurf: Dynamic deblurring neural radiance fields for blurry monocular video. *arXiv preprint arXiv:2312.13528*, 2023. [2](#), [3](#)
- [5] Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 130–141, 2023. [1](#), [2](#), [3](#)
- [6] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5972–5981, 2022. [2](#)
- [7] 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. [1](#), [3](#), [4](#)
- [8] Wen-Hsuan Chu, Lei Ke, and Katerina Fragkiadaki. Dreamscene4d: Dynamic multi-object scene generation from monocular videos. *arXiv preprint arXiv:2405.02280*, 2024. [1](#), [2](#), [3](#)
- [9] Carl Doersch, Yi Yang, Mel Vecerik, Dilara Gokay, Ankush Gupta, Yusuf Aytar, Joao Carreira, and Andrew Zisserman. Tapir: Tracking any point with per-frame initialization and temporal refinement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10061–10072, 2023. [9](#)
- [10] Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d-rotor gaussian splatting: towards efficient novel view synthesis for dynamic scenes. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. [1](#), [2](#), [3](#)
- [11] Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d gaussian splatting: Towards efficient novel view synthesis for dynamic scenes. *arXiv preprint arXiv:2402.03307*, 2024. [2](#)
- [12] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *CVPR*, 2022. [6](#)
- [13] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. [2](#)
- [14] Brandon Y Feng, Hadi Alzayer, Michael Rubinstein, William T Freeman, and Jia-Bin Huang. 3d motion magnification: Visualizing subtle motions from time-varying radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9837–9846, 2023.
- [15] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. [2](#)
- [16] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5712–5721, 2021. [1](#), [3](#)
- [17] Quankai Gao, Qiangeng Xu, Zhe Cao, Ben Mildenhall, Wenchao Ma, Le Chen, Danhang Tang, and Ulrich Neumann. Gaussianflow: Splatting gaussian dynamics for 4d content creation. *arXiv preprint arXiv:2403.12365*, 2024. [1](#), [3](#)
- [18] Xiang Guo, Jiadai Sun, Yuchao Dai, Guanying Chen, Xiaojing Ye, Xiao Tan, Errui Ding, Yumeng Zhang, and Jingdong Wang. Forward flow for novel view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16022–16033, 2023. [2](#)
- [19] Zhiyang Guo, Wengang Zhou, Li Li, Min Wang, and Houqiang Li. Motion-aware 3d gaussian splatting for efficient dynamic scene reconstruction. *arXiv preprint arXiv:2403.11447*, 2024. [1](#), [3](#)
- [20] Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4220–4230, 2024. [2](#)
- [21] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *European Conference on Computer Vision*, pages 624–642. Springer, 2022. [7](#)
- [22] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. [6](#), [9](#)
- [23] Tae Hyun Kim, Kyoung Mu Lee, Bernhard Scholkopf, and Michael Hirsch. Online video deblurring via dynamic temporal blending network. In *Proceedings of the IEEE international conference on computer vision*, pages 4038–4047, 2017. [2](#)
- [24] Bangrui Jiang, Zhihuai Xie, Zhen Xia, Songnan Li, and Shan Liu. Erdn: Equivalent receptive field deformable network for video deblurring. In *European Conference on Computer Vision*, pages 663–678. Springer, 2022. [2](#)
- [25] Kai Katsumata, Duc Minh Vo, and Hideki Nakayama. A compact dynamic 3d gaussian representation for real-time dynamic view synthesis. In *ECCV*, 2024. [1](#), [2](#), [3](#)
- [26] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer.

- In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 7
- [27] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1, 2, 3, 4
- [28] Kiyeon Kim, Seungyong Lee, and Sunghyun Cho. Mssnet: Multi-scale-stage network for single image deblurring. In *European conference on computer vision*, pages 524–539. Springer, 2022. 5
- [29] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023. 2
- [30] Agelos Kratimenos, Jiahui Lei, and Kostas Daniilidis. Dynmf: Neural motion factorization for real-time dynamic view synthesis with 3d gaussian splatting. *arXiv preprint arXiv:2312.00112*, 2023. 1, 2, 3
- [31] Byeonghyeon Lee, Howoong Lee, Xiangyu Sun, Usman Ali, and Eunbyung Park. Deblurring 3d gaussian splatting. *arXiv preprint arXiv:2401.00834*, 2024. 3
- [32] Dogyoon Lee, Minhyeok Lee, Chajin Shin, and Sangyoun Lee. Dp-nerf: Deblurred neural radiance field with physical scene priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12386–12396, 2023. 1
- [33] Dongwoo Lee, Jeongtaek Oh, Jaesung Rim, Sunghyun Cho, and Kyoung Mu Lee. Exblurf: Efficient radiance fields for extreme motion blurred images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17639–17648, 2023. 1
- [34] Junghe Lee, Donghyeong Kim, Dogyoon Lee, Suhwan Cho, and Sangyoun Lee. Crim-gs: Continuous rigid motion-aware gaussian splatting from motion blur images. *arXiv preprint arXiv:2407.03923*, 2024. 3
- [35] Yao-Chih Lee, Zhoutong Zhang, Kevin Blackburn-Matzen, Simon Niklaus, Jianming Zhang, Jia-Bin Huang, and Feng Liu. Fast view synthesis of casual videos. *arXiv preprint arXiv:2312.02135*, 2023. 1, 3
- [36] Jiahui Lei, Yijia Weng, Adam Harley, Leonidas Guibas, and Kostas Daniilidis. Mosca: Dynamic gaussian fusion from casual videos via 4d motion scaffolds. *arXiv preprint arXiv:2405.17421*, 2024. 1, 3
- [37] Dasong Li, Yi Zhang, Ka Chun Cheung, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Learning degradation representations for image deblurring. In *European conference on computer vision*, pages 736–753. Springer, 2022. 2
- [38] Ji Li, Weixi Wang, Yuesong Nan, and Hui Ji. Self-supervised blind motion deblurring with deep expectation maximization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13986–13996, 2023. 2
- [39] Moyang Li, Peng Wang, Lingzhe Zhao, Bangyan Liao, and Peidong Liu. Usb-nerf: Unrolling shutter bundle adjusted neural radiance fields. *arXiv preprint arXiv:2310.02687*, 2023. 1, 3
- [40] Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime gaussian feature splatting for real-time dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8508–8520, 2024. 1, 2, 3
- [41] Jingyun Liang, Yuchen Fan, Xiaoyu Xiang, Rakesh Ranjan, Eddy Ilg, Simon Green, Jiezhang Cao, Kai Zhang, Radu Timofte, and Luc V Gool. Recurrent video restoration transformer with guided deformable attention. *Advances in Neural Information Processing Systems*, 35:378–393, 2022. 2
- [42] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *IEEE Transactions on Image Processing*, 2024. 2
- [43] Jingbo Lin, Zhilu Zhang, Yuxiang Wei, Dongwei Ren, Dongsheng Jiang, Qi Tian, and Wangmeng Zuo. Improving image restoration through removing degradations in textual representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2866–2878, 2024. 2
- [44] Youtian Lin, Zuozhuo Dai, Siyu Zhu, and Yao Yao. Gaussian-flow: 4d reconstruction with dynamic 3d gaussian particle. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21136–21145, 2024. 1, 2, 3
- [45] Jia-Wei Liu, Yan-Pei Cao, Weijia Mao, Wenqiao Zhang, David Junhao Zhang, Jussi Keppo, Ying Shan, Xiaohu Qie, and Mike Zheng Shou. Devrf: Fast deformable voxel radiance fields for dynamic scenes. *Advances in Neural Information Processing Systems*, 35:36762–36775, 2022. 2
- [46] Qingming Liu, Yuan Liu, Jiepeng Wang, Xianqiang Lv, Peng Wang, Wenping Wang, and Junhui Hou. Modgs: Dynamic gaussian splatting from causally-captured monocular videos. *arXiv preprint arXiv:2406.00434*, 2024. 1, 3
- [47] Shuaicheng Liu, Ping Tan, Lu Yuan, Jian Sun, and Bing Zeng. Meshflow: Minimum latency online video stabilization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VI 14*, pages 800–815. Springer, 2016. 7, 8
- [48] Zhicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, Min Yang, Xiao Tang, Feng Zhu, and Yuchao Dai. 3d geometry-aware deformable gaussian splatting for dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8900–8910, 2024. 1, 2, 3
- [49] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. *arXiv preprint arXiv:2308.09713*, 2023. 1, 2, 3
- [50] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12861–12870, 2022. 1, 3
- [51] Marko Mihajlovic, Sergey Prokudin, Siyu Tang, Robert Maier, Federica Bogo, Tony Tung, and Edmond Boyer.

- Splatfields: Neural gaussian splats for sparse 3d and 4d reconstruction. *arXiv preprint arXiv:2409.11211*, 2024. 2
- [52] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 2, 9
- [53] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 3
- [54] Seungjun Nah, Sanghyun Son, and Kyoung Mu Lee. Recurrent neural networks with intra-frame iterations for video deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8102–8111, 2019. 2
- [55] Jeongtaek Oh, Jaeyoung Chung, Dongwoo Lee, and Kyoung Mu Lee. Deblurgs: Gaussian splatting for camera motion blur. *arXiv preprint arXiv:2404.11358*, 2024. 1, 3, 4
- [56] Jinshan Pan, Boming Xu, Jiangxin Dong, Jianjun Ge, and Jinhui Tang. Deep discriminative spatial and temporal network for efficient video deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023. 2, 6, 7, 10
- [57] 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. 6
- [58] Cheng Peng, Yutao Tang, Yifan Zhou, Nengyu Wang, Xijun Liu, Deming Li, and Rama Chellappa. Bags: Blur agnostic gaussian splatting through multi-scale kernel modeling. *arXiv preprint arXiv:2403.04926*, 2024. 1, 3, 4
- [59] Zhan Peng, Xinyi Ye, Weiyue Zhao, Tianqi Liu, Huiqiang Sun, Baopu Li, and Zhiguo Cao. 3d multi-frame fusion for video stabilization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7507–7516, 2024. 7
- [60] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021. 1, 2, 3
- [61] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. 9
- [62] Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10721–10733, 2023. 2
- [63] Xuanchi Ren, Zian Qian, and Qifeng Chen. Video deblurring by fitting to test data. *arXiv preprint arXiv:2012.05228*, 2020. 4
- [64] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 6
- [65] Jenny Seidenschwarz, Qunjie Zhou, Bardienus Duisterhof, Deva Ramanan, and Laura Leal-Taixé. Dynomo: Online point tracking by dynamic online monocular gaussian reconstruction. *arXiv preprint arXiv:2409.02104*, 2024. 1, 2, 3
- [66] Wei Shang, Dongwei Ren, Yi Yang, Hongzhi Zhang, Kede Ma, and Wangmeng Zuo. Joint video multi-frame interpolation and deblurring under unknown exposure time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13935–13944, 2023. 2, 7
- [67] Richard Shaw, Michal Nazarczuk, Jifei Song, Arthur Moreau, Sibi Catley-Chandar, Helisa Dhamo, and Eduardo Pérez-Pellitero. Swings: Sliding windows for dynamic 3d gaussian splatting. 2, 8
- [68] Nagabhushan Somraj, Kapil Choudhary, Sai Harsha Mupparaju, and Rajiv Soundararajan. Factorized motion fields for fast sparse input dynamic view synthesis. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. 2
- [69] Hyeongseok Son, Junyong Lee, Jonghyeop Lee, Sunghyun Cho, and Seungyong Lee. Recurrent video deblurring with blur-invariant motion estimation and pixel volumes. *ACM Transactions on Graphics (TOG)*, 40(5):1–18, 2021. 2
- [70] Colton Stearns, Adam Harley, Mikaela Uy, Florian Dubost, Federico Tombari, Gordon Wetzstein, and Leonidas Guibas. Dynamic gaussian marbles for novel view synthesis of casual monocular videos. *arXiv preprint arXiv:2406.18717*, 2024. 1, 2, 3
- [71] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8934–8943, 2018. 5, 6
- [72] Huiqiang Sun, Xingyi Li, Liao Shen, Xinyi Ye, Ke Xian, and Zhiguo Cao. Dyblurf: Dynamic neural radiance fields from blurry monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7517–7527, 2024. 2, 3, 4, 6, 7, 10, 11, 12
- [73] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 769–777, 2015. 2
- [74] Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, and Wei Xing. 3dgstream: On-the-fly training of 3d gaussians for efficient streaming of photo-realistic free-viewpoint videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20675–20685, 2024. 2

- [75] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018. 5
- [76] Fengrui Tian, Shaoyi Du, and Yueqi Duan. Monon-erf: Learning a generalizable dynamic radiance field from monocular videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17903–17913, 2023. 1, 3
- [77] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023. 7
- [78] Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. Bad-nerf: Bundle adjusted deblur neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4170–4179, 2023. 1, 3, 4
- [79] Qianqian Wang, Vickie Ye, Hang Gao, Jake Austin, Zhengqi Li, and Angjoo Kanazawa. Shape of motion: 4d reconstruction from a single video. *arXiv preprint arXiv:2407.13764*, 2024. 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12
- [80] Shizun Wang, Xingyi Yang, Qihong Shen, Zhenxiang Jiang, and Xinchao Wang. Gflow: Recovering 4d world from monocular video. *arXiv preprint arXiv:2405.18426*, 2024. 1, 3
- [81] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4, 6, 9
- [82] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 2
- [83] Patrick Wieschollek, Michael Hirsch, Bernhard Scholkopf, and Hendrik Lensch. Learning blind motion deblurring. In *Proceedings of the IEEE international conference on computer vision*, pages 231–240, 2017. 2
- [84] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20310–20320, 2024. 1, 2, 3, 6, 7, 10, 11, 12
- [85] Renlong Wu, Zhilu Zhang, Shuhao Zhang, Longfei Gou, Haobin Chen, Lei Zhang, Hao Chen, and Wangmeng Zuo. Self-supervised video desmoking for laparoscopic surgery. *arXiv preprint arXiv:2403.11192*, 2024. 6
- [86] Zijie Wu, Chaohui Yu, Yanqin Jiang, Chenjie Cao, Fan Wang, and Xiang Bai. Sc4d: Sparse-controlled video-to-4d generation and motion transfer. In *European Conference on Computer Vision*, pages 361–379. Springer, 2025. 1, 3
- [87] Zhiwen Yan, Chen Li, and Gim Hee Lee. Nerf-ds: Neural radiance fields for dynamic specular objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8285–8295, 2023. 1, 2, 3
- [88] Xiaodong Yang, Weixing Xie, Yihang Fu, Wentao Fan, and Xiao Dong. 4d gaussian splatting for high-fidelity dynamic reconstruction of single-view scenes. Available at SSRN 4895336. 1, 3
- [89] Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. *arXiv preprint arXiv:2310.10642*, 2023. 1, 2, 3
- [90] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20331–20341, 2024. 1, 2, 3, 6, 7, 10, 11, 12
- [91] Heng Yu, Chaoyang Wang, Peiye Zhuang, Willi Menapace, Aliaksandr Siarohin, Junli Cao, Laszlo A Jeni, Sergey Tulyakov, and Hsin-Ying Lee. 4real: Towards photorealistic 4d scene generation via video diffusion models. *arXiv preprint arXiv:2406.07472*, 2024. 1, 3
- [92] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 2, 6, 7, 10
- [93] Yifei Zeng, Yanqin Jiang, Siyu Zhu, Yuanxun Lu, Youtian Lin, Hao Zhu, Weiming Hu, Xun Cao, and Yao Yao. Stag4d: Spatial-temporal anchored generative 4d gaussians. In *European Conference on Computer Vision*, pages 163–179. Springer, 2025. 1, 3
- [94] Guozhen Zhang, Yuhuan Zhu, Haonan Wang, Youxin Chen, Gangshan Wu, and Limin Wang. Extracting motion and appearance via inter-frame attention for efficient video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5682–5692, 2023. 7
- [95] Huicong Zhang, Haozhe Xie, and Hongxun Yao. Blur-aware spatio-temporal sparse transformer for video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2673–2681, 2024. 2, 6, 7, 10
- [96] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Wei Liu, and Hongdong Li. Adversarial spatio-temporal learning for video deblurring. *IEEE Transactions on Image Processing*, 28(1):291–301, 2018. 2
- [97] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6, 9
- [98] Yabo Zhang, Yuxiang Wei, Dongsheng Jiang, Xiaopeng Zhang, Wangmeng Zuo, and Qi Tian. Controlvideo: Training-free controllable text-to-video generation. *arXiv preprint arXiv:2305.13077*, 2023. 3
- [99] Zhilu Zhang, RongJian Xu, Ming Liu, Zifei Yan, and Wangmeng Zuo. Self-supervised image restoration with blurry and noisy pairs. *Advances in Neural Information Processing Systems*, 35:29179–29191, 2022. 2

- [100] Zhuofan Zhang, Zhen Liu, Ping Tan, Bing Zeng, and Shuaicheng Liu. Minimum latency deep online video stabilization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23030–23039, 2023. 7, 8
- [101] Zhilu Zhang, Shuoqiao Zhang, Renlong Wu, Zifei Yan, and Wangmeng Zuo. Bracketing is all you need: Unifying image restoration and enhancement tasks with multi-exposure images. *arXiv preprint arXiv:2401.00766*, 2024. 2
- [102] Lingzhe Zhao, Peng Wang, and Peidong Liu. Bad-gaussians: Bundle adjusted deblur gaussian splatting. *arXiv preprint arXiv:2403.11831*, 2024. 1, 3, 4
- [103] Xiaopeng Zhao, Zhenlin An, Qingrui Pan, and Lei Yang. Nerf2: Neural radio-frequency radiance fields. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023. 2
- [104] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 191–207. Springer, 2020. 2
- [105] Zhihang Zhong, Mingdeng Cao, Xiang Ji, Yinqiang Zheng, and Imari Sato. Blur interpolation transformer for real-world motion from blur. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5713–5723, 2023.
- [106] Zhihang Zhong, Ye Gao, Yinqiang Zheng, Bo Zheng, and Imari Sato. Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network. *International Journal of Computer Vision*, 131(1):284–301, 2023. 2
- [107] Shangchen Zhou, Jiawei Zhang, Wangmeng Zuo, Haozhe Xie, Jinshan Pan, and Jimmy S Ren. Davanet: Stereo deblurring with view aggregation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10996–11005, 2019. 6
- [108] Ruijie Zhu, Yanzhe Liang, Hanzhi Chang, Jiacheng Deng, Jiahao Lu, Wenfei Yang, Tianzhu Zhang, and Yongdong Zhang. Motions: Exploring explicit motion guidance for deformable 3d gaussian splatting. *arXiv preprint arXiv:2410.07707*, 2024. 1, 3