

中图法分类号: TP18; P391 文献标识码: A 文章编号: 1006-8961(2026)04-1018-11

论文引用格式: Fan Y, Lu T Y and Wang Y Q. 2026. Robust scene deblurring and reconstruction with event cameras and 3DGS. Journal of Image and Graphics, 31(4):1018-1028(樊越, 陆韬宇, 王逸群. 2026. 融合事件相机与3DGS的鲁棒场景去模糊重建. 中国图象图形学报, 31(4):1018-1028)[DOI:10.11834/jig.250290]

## 融合事件相机与3DGS的鲁棒场景去模糊重建

樊越, 陆韬宇, 王逸群\*

重庆大学计算机学院, 重庆 400044

**摘要:** 目的 针对现有基于神经渲染的去模糊场景重建方法在图像清晰度、新视角泛化能力以及计算效率等方面的局限性, 提出一种融合事件相机数据与3D高斯泼溅(3D Gaussian splatting, 3DGS)框架的新方法。方法 一方面利用事件相机高时间分辨率优势, 捕捉动态场景运动轨迹与亮度变化, 弥补传统帧图像在动态场景下的信息缺失, 为去模糊重建提供更丰富的时空约束; 另一方面通过引入高斯形状属性变换网络, 预测高斯体缩放、旋转等变换属性, 动态调整其空间分布, 显著缓解了原视角过拟合问题, 同时避免新视角下因几何不匹配导致的模糊现象。结果 实验结果表明, 在合成数据集上, 本文方法在PSNR(peak signal-to-noise ratio)、SSIM(structural similarity index)和LPIPS(learned perceptual image patch similarity)指标上优于对比方法, 在真实数据集上, BRISQUE(blind/referenceless image spatial quality evaluator)值显著降低。与基于神经辐射场的方法相比, 训练时间从48 h缩减至不足1 h, 且实现了140 帧/s的实时渲染速度。结论 实验结果验证了提出的方法在复杂场景下的去模糊能力和改善新视角渲染效果方面的有效性。

**关键词:** 事件相机; 3D高斯泼溅(3DGS); 去模糊; 三维重建; 新视角合成

## Robust scene deblurring and reconstruction with event cameras and 3DGS

Fan Yue, Lu Taoyu, Wang Yiqun\*

School of Computer Science, Chongqing University, Chongqing 400044, China

**Abstract: Objective** In the tasks of 3D scene reconstruction and novel view synthesis, neural radiance fields (NeRF) stands as a landmark advancement in this field. It significantly drives the development and evolution of neural rendering-based 3D scene reconstruction methods. Moreover, 3D Gaussian splatting (3DGS) has emerged as a simple yet computationally efficient approach, thereby gaining recognition for its fast training and rendering capabilities. However, these methods typically operate only under ideal input conditions and often struggle to handle motion blur—a critical issue, given that motion blur in input images severely degrades rendering quality. Some existing methods attempt to address this issue by employing trainable parameters or models during reconstruction to predict the trajectory of camera motion blur. Nevertheless, this predictive approach fails to reflect the real trajectory of camera movement accurately, thereby limiting the accuracy of 3D scene reconstruction. Other methods integrate deblurring models with the 3DGS framework to accelerate model training and rendering speeds. Regrettably, these methods suffer from quality degradation issues, such as artifacts when rendering novel views, because of the risk of model overfitting during training. This study proposes a novel method

收稿日期: 2025-07-02; 修回日期: 2025-08-30; 预印本日期: 2025-09-06

\*通信作者: 王逸群 yiqun.wang@cqu.edu.cn

基金项目: 国家自然科学基金项目(52238003, 62202076); 重庆市自然科学基金项目(CSTB2023NSCQ-BSX0004, CSTB2022NSCQ-MSX0924)

**Supported by:** National Natural Science Foundation of China (52238003, 62202076); Chongqing Municipal Natural Science Foundation (CSTB2023NSCQ-BSX0004, CSTB2022NSCQ-MSX0924)

that fuses event camera data with the 3DGS framework to overcome the limitations of the aforementioned approaches.

**Method** An event camera triggers event responses only when the pixel brightness changes beyond a preset threshold, thereby enabling it to capture the instantaneous motion trajectories of objects and subtle brightness fluctuations in dynamic scenes accurately. The proposed method fully leverages the unique advantage of event cameras—microsecond-level high temporal resolution. For the sequential grayscale images output by the model, the method designs an event bin calculation module: first, it divides the grayscale image sequence at fixed time intervals and defines the time window corresponding to two consecutive frames as an event bin unit. Subsequently, it counts the number of events triggered by brightness changes for each pixel within this unit, thereby generating event bin data predicted by the model. The method further introduces real event bins as supervision signals to ensure the effectiveness of event information supervision. Real event bins are constructed using synchronously collected raw event camera data, and the pixel-level L2 error between the predicted event bins and the real event bins is calculated. Then, this error is incorporated into the model's loss function. This supervision mechanism forces the model to learn the laws of the continuous blurring process at high temporal resolution, thereby avoiding traditional methods' problem of losing details of dynamic processes because of reliance on discrete frame images. The information loss of traditional frame images in dynamic scenes can be compensated by combining the event bin calculation module with the 3DGS framework, thereby providing rich spatiotemporal constraints for deblurring reconstruction. Second, the method innovatively introduces a Gaussian shape attribute transformation network, which establishes a prediction and adjustment mechanism for attributes such as scaling and rotation of each 3D Gaussian primitive in the 3DGS framework. During the model training phase, this transformation mechanism achieves dynamic optimization of Gaussian primitives through iterative adjustments: after each training epoch, the parameters of all Gaussian primitives are updated based on the transformation attributes predicted by the network, gradually correcting the deviation between the initial Gaussian primitives and the real scene structure. This approach effectively prevents the Gaussian model from overfitting to the training views. During the rendering process, these transformation attributes can dynamically adjust the spatial distribution of 3D Gaussian primitives in real time, thereby allowing the Gaussian primitives to conform to the motion states and geometric structures of dynamic objects in the scene. As a result, the constraints of the original view data on the model are broken, and the degradation of rendering quality caused by geometric mismatches is avoided when switching to novel views.

**Result** Experimental results demonstrate the superiority of the proposed method. On synthetic datasets, it outperforms the existing methods in terms of peak signal-to-noise ratio, structural similarity index, and learned perceptual image patch similarity. This finding indicates a significant reduction in image noise and blur, and the rendered results exhibit structural information (such as edges and textures) that is highly consistent with real scenes. On real-world datasets, the method achieves a substantial reduction in the blind/referenceless image spatial quality evaluator value, thereby fully verifying the method's deblurring capability in real and complex scenarios. Compared with NeRF-based methods, the proposed method reduces the training time from 48 h to less than 1 h (representing an approximately 50-fold improvement in training efficiency) and achieves a real-time rendering speed of 140 frames/s. **Conclusion** Comprehensive experimental results confirm that the proposed method addresses the shortcomings of existing neural rendering-based deblurring reconstruction techniques in novel view generalization ability. It also achieves dual improvements in image clarity and computational efficiency through innovations in data fusion and network structure. This method provides a practical technical solution for high-fidelity and real-time reconstruction of dynamic scenes.

**Key words:** event camera; 3D Gaussian splatting(3DGS); deblurring; 3D reconstruction; novel view synthesis

## 0 引言

3D场景重建是计算机视觉的基本功能之一。随着神经辐射场(neural radiance field, NeRF)(Mildenhall等,2022)的出现,近年来,3D场景重建

的研究获得了更多关注。基于NeRF的方法通过隐式神经网络表示3D场景,取得了令人印象深刻的新颖视角合成效果。尽管出现了诸如Barron等人(2021)、Pumarola等人(2021)等多种方法来增强基于NeRF的方法的适用性,但是计算效率的问题仍旧突出。Müller等人(2022)在加速网络训练和图像

渲染速度方面进行了探索,但瓶颈问题依旧存在。遵循这一研究兴趣, Kerbl 等人(2023)提出了 3D 高斯泼溅(3D Gaussian splatting, 3DGS),它避免了对光线的密集采样,而是使用高斯球来表示 3D 空间场景信息,从而允许更快的训练和渲染。3DGS 通过可微光栅化技术,将 3D 高斯模型直接投影到图像平面,实现了实时渲染的同时保持高质量的视觉效果。

在实际应用中,常常会遇到由物体运动和相机抖动导致的图像模糊现象。一些方法(谢斌等, 2025; 邱云飞等, 2025; 陈加保等, 2023)仅聚焦于二维图像层面的去模糊处理,并未将研究重点延伸至 3D 场景重建任务。

Ma 等人(2022)提出了首个能够从模糊输入中恢复清晰 NeRF 的方法,它采用了一种分析合成的方式,使用一个可变形稀疏核(deformable sparse kernel, DSK)模块,通过模拟模糊过程重建模糊视图。Lee 等人(2023)则提出使用两个物理先验约束从模糊图像中恢复 3D 场景。随后,多个方法尝试解决 3D 高斯散射框架下的去模糊问题。Lee 等人(2024)提出使用一个小型的多层感知机(multilayer perceptron, MLP),来学习 3D 高斯模型不同属性的变化,通过调整每个 3D 高斯模型的协方差和均值,实现了对场景模糊程度的建模。Zhao 等人(2024)则采用样条函数描述相机曝光时间内的轨迹,在对高斯模型本身进行优化的同时,对曝光时间内的相机轨迹也进行联合优化。

然而,这些方法在重建过程中普遍采用可训练参数或模型预测相机运动模糊的轨迹,这种预测方式难以准确反映相机运动的真实轨迹,导致 3D 场景重建精度受限。

事件相机,也称为动态视觉传感器(dynamic vision sensor, DVS)(Lichtsteiner 等, 2008),异步捕捉像素亮度变化,借鉴了生物视觉系统。这种独特的记录框架有效地解决了帧间信息丢失的问题,这是帧式 RGB 相机常见的问题。事件相机具有多个优点,包括无运动模糊、高动态范围、低功耗和减少延迟。由于这些优势,它们在各种任务中表现出色,如光流估计(Akolkar 等, 2022)、深度估计(Takatani 等, 2021)以及特征检测和跟踪(Zhang 等, 2022)。

最近, Hwang 等人(2023)和 Rudnev 等人(2023)成功地仅使用事件数据训练了 NeRF 模型。然而,

这些方法经历了明显的伪影和色差,并且在神经表示学习的姿态估计中表现出有限的泛化能力。同时, Qi 等人(2023)通过利用模糊 RGB 图像和相应的事件数据成功训练了更清晰的 NeRF。Deguchi 等人(2024)第一个提出了将事件数据纳入高斯散射框架的方法。通过事件数据精确捕捉相机曝光时间内的运动轨迹,该方法显著提升了场景重建质量和图像去模糊效果。然而,该方法在训练过程中存在模型过拟合的风险,这会导致新视角渲染时出现伪影等质量下降问题。

针对上述方法的局限性,本文在现有研究的基础上提出了一种新的基于事件相机和 3DGS 的场景重建与去模糊方法。该方法的核心在于将模糊的 RGB 图像与事件数据有效结合,并引入一个形状属性变换网络预测场景中每个 3D 高斯的变换属性。这一设计在保持训练速度的基础上,实现了输入图像的高效去模糊处理,还进一步增强了模型生成高质量新视角图像的能力。

本文研究的主要贡献包括: 1)提出了一种基于事件相机和 3DGS 的场景重建与去模糊方法,有效解决了运动模糊重建的难题; 2)引入了形状属性变换网络  $F_\theta$  预测 3D 高斯的变换属性,显著改善了模型的新视角渲染能力,解决了过拟合问题; 3)在合成数据集和真实数据集上的实验表明,该方法在图像去模糊和新视角合成任务中均优于对比的基线方法。

## 1 方法

本文方法旨在解决 3D 场景重建中的运动模糊问题。通过结合事件相机数据与 3D 高斯散射(3DGS)框架,提出了一种高效的场景重建与去模糊方法。如图 1 所示,方法框架主要由 3D 高斯建模、事件数据处理、训练数据准备以及 3D 重建与去模糊流程 4 个关键部分构成。

### 1.1 3D 高斯建模

为了表示一个体积场并对其进行渲染,本研究采用了 3DGS 渲染中的方法,本方法提出了可微光栅化技术,利用 3D 高斯作为灵活高效的表示,高斯系数经过优化,可以准确地表示场景,结合了基于神经网络的优化和显式结构化数据存储的优点。这种混合方法旨在通过更快的训练和实时性能实现高质

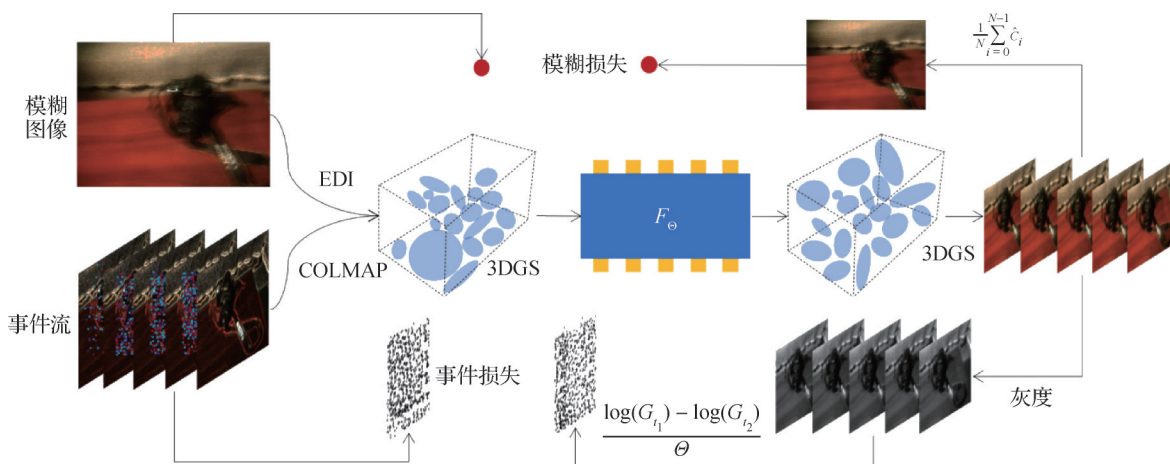


图1 模型框架

Fig. 1 Model framework

量渲染,特别是对于复杂的场景和高分辨率输出。高斯函数由世界空间中定义的全3D协方差矩阵 $\Sigma$ 定义,该矩阵以点(均值) $\mu$ 为中心(Zwicker等, 2001),具体为

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

式中, $\mathbf{x}$ 是3D高斯在空间中的位置坐标。然而,为了进行渲染,需要将3D高斯模型投影到2D空间,给定一个视图变换 $\mathbf{W}$ ,三维空间中的协方差矩阵 $\Sigma$ ,相机坐标系中的二维协方差矩阵 $\Sigma'$ 为

$$\Sigma' = \mathbf{J} \mathbf{W} \Sigma \mathbf{W}^T \mathbf{J}^T \quad (2)$$

式中, $\mathbf{J}$ 是通过投影转换矩阵的仿射近似变换, $\mathbf{W}$ 是视角变换矩阵。

3D高斯模型的协方差矩阵 $\Sigma$ 类似于描述一个椭球体的形状。给定一个缩放矩阵 $\mathbf{S}$ 和旋转矩阵 $\mathbf{R}$ ,计算对应的 $\Sigma$ ,具体为

$$\Sigma = \mathbf{R} \mathbf{S} \mathbf{S}^T \mathbf{R}^T \quad (3)$$

为了能够独立地对 $\mathbf{R}$ 和 $\mathbf{S}$ 进行优化,将 $\mathbf{R}$ 和 $\mathbf{S}$ 分别通过一个表示缩放的3D向量 $\mathbf{s}$ 和一个表示旋转的四元数 $\mathbf{r}$ 进行建模。 $\mathbf{s}$ 和 $\mathbf{r}$ 这两个参数可以很容易地分别转换为缩放矩阵 $\mathbf{S}$ 和旋转矩阵 $\mathbf{R}$ 并进行组合,同时要确保对 $\mathbf{r}$ 进行归一化以得到一个有效的单位四元数。

对于图像中每一个像素的RGB值,通过对覆盖在像素上的 $N$ 个有序投影的2D高斯进行累加来计算,具体为

$$\mathbf{C} = \sum_{i \in N} T_i \mathbf{c}_i \alpha_i, \quad T_i = \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (4)$$

式中, $\mathbf{c}_i$ 是2D高斯在当前像素位置的颜色, $T_i$ 是2D

高斯在当前像素位置的透射率。 $\alpha_i \in [0, 1]$ ,由 $1 - \exp(-\sigma_i \delta_i)$ 定义, $\sigma_i, \delta_i$ 分别表示2D高斯在当前像素位置的密度和距离其他2D高斯的间隔。

在上述3D高斯函数建模的基础上,为了能够更好地对新视角进行渲染,该研究提出形状属性变换网络,该网络使用一个参数量较少的小型MLP预测3D高斯的变换属性 $\Delta \mathbf{r}, \Delta \mathbf{s}$ ,具体为

$$(\Delta \mathbf{r}_i, \Delta \mathbf{s}_i) = F_\theta(\gamma(\mathbf{x}_i), \mathbf{s}_i, \mathbf{r}_i, \gamma(\mathbf{v})) \quad (5)$$

式中, $F_\theta$ 表示MLP, $\mathbf{x}_i, \mathbf{s}_i, \mathbf{r}_i$ 分别表示第 $i$ 个3D高斯的位置、缩放向量和四元数, $\mathbf{v}$ 表示视角观察方向。 $\gamma$ 表示位置编码,具体定义为

$$\gamma(\mathbf{x}) = (\sin(2^0 \pi \mathbf{x}), \cos(2^0 \pi \mathbf{x}), \dots, \sin(2^{L-1} \pi \mathbf{x}), \cos(2^{L-1} \pi \mathbf{x})) \quad (6)$$

式中, $L$ 表示编码的频数, $\gamma(\mathbf{x})$ 和 $\gamma(\mathbf{v})$ 分别表示对3D高斯的位置和观察方向进行位置编码。为了优化的稳定性,每个缩放因子 $(\Delta \mathbf{r}, \Delta \mathbf{s})$ 都乘以 $\lambda_{rs}$ ,再加上 $1 - \lambda_{rs}$ ,并将它们的最小值限制为1.0。将缩放因子 $(\Delta \mathbf{r}_i, \Delta \mathbf{s}_i)$ 分别和 $(\mathbf{r}_i, \mathbf{s}_i)$ 相乘,得到变换后的属性,具体为

$$\begin{cases} \hat{\mathbf{r}}_i = \mathbf{r}_i \cdot \min(1.0, \lambda_{rs} \Delta \mathbf{r}_i + (1 - \lambda_{rs})) \\ \hat{\mathbf{s}}_i = \mathbf{s}_i \cdot \min(1.0, \lambda_{rs} \Delta \mathbf{s}_i + (1 - \lambda_{rs})) \end{cases} \quad (7)$$

式中,函数 $\min(\cdot, \cdot)$ 返回较小的值。利用这些变换后的属性,构建变换后的3D高斯模型: $G(\mathbf{x}_i, \hat{\mathbf{r}}_i, \hat{\mathbf{s}}_i)$ 。在训练过程中,不仅对3D高斯进行优化,也对网络 $F_\theta$ 进行优化。

## 1.2 事件数据处理

事件相机检索到像素 $(x, y)$ 的亮度变化在时间 $t^e$ 超过阈值 $\Theta$ 时,会异步记录事件 $\mathbf{e}(x, y, t^e, d^e)$ 。事

件相机不会报告实际的亮度强度值 $L(x, y, t^e)$ , 而是报告亮度强度的变化方向 $d^e$ , 其定义为

$$d^e(x, y, t^e) = \begin{cases} -1 & \log(L(x, y, t^e)) - \log(L(x, y, t^e - \Delta t)) < -\Theta_{\text{neg}} \\ +1 & \log(L(x, y, t^e)) - \log(L(x, y, t^e - \Delta t)) > +\Theta_{\text{pos}} \end{cases} \quad (8)$$

式中,  $\Theta_{\text{neg}} = 0.3$ ,  $\Theta_{\text{pos}} = 0.2$ 。

给定一幅曝光时间从 $t_0$ 到 $t_{N-1}$ 的模糊图像以及相应的事件数据 $\{e_i\}_{t_0 < t'_i \leq t_{N-1}}$ , 设置 $N$ 个时间戳 $\{t_k\}_{k=0}^{N-1}$ , 将事件流平均分成 $N-1$ 个事件箱 $\{B_k\}_{k=1}^{N-1}$ , 以便更准确地估计曝光时间期间像素亮度的强度变化, 具体为

$$B_k = \{e_i(x_i, y_i, t'_i, d'_i)\}_{t_{k-1} < t'_i \leq t_k} \quad (9)$$

### 1.3 训练数据准备

对于合成数据集, 可以直接使用Blender中提供的真实相机位姿, 每一幅模糊的图像, 都通过相机位姿连续变化下的5幅清晰图像合成。而对于真实环境下拍摄到的模糊图像, 则需要使用COLMAP (structure-from-motion and multi-view stereo) (Schönberger和Frahm, 2016)估计相机位姿。然而, 当输入图像变得模糊时, COLMAP的位姿估计和点云重建过程就会失败。因此, 对于实际拍摄到的模糊数据, 该研究参考EDI(event-based deinterlacing)方法(Pan等, 2019), 对模糊图像进行初步去模糊处理, 然后将处理结果输入到COLMAP中, 以获取模糊过程中的相机位姿。

EDI模型利用事件数据将单幅模糊图像转换为多个按时间排列下的清晰图像。将其公式简化为离散版本。给定一幅模糊图像 $C_{\text{blur}}$ 以及相应的事件区间 $\{B_k\}_{k=1}^{N-1}$ , 假设在 $t_0$ 时刻的清晰图像为 $C_0$ , 根据式(10), 在划分每个事件区间的 $t_k$ 时刻的清晰图像 $C_k$ 可以表示为

$$C_k = C_0 e^{\theta \sum_{i=1}^k B_i}, k \geq 1 \quad (10)$$

根据一般的图像形成模型, 假设模糊图像是由多幅清晰图像进行加权得到的。由于已经将曝光时间均匀地划分为 $N$ 个部分, 因此模糊图像可以直接看做是这些图像的平均值, 具体为

$$C_{\text{blur}} = \frac{1}{N} \sum_{k=0}^{N-1} C_k \quad (11)$$

$$C_{\text{blur}} = \frac{C_0}{N} \left( 1 + e^{\theta \sum_{i=1}^1 B_i} + \dots + e^{\theta \sum_{i=1}^{N-1} B_i} \right) \quad (12)$$

式中, 在 $t_0$ 时刻的清晰图像为 $C_0$ , 可以表示为

$$C_0 = \frac{NC_{\text{blur}}}{1 + e^{\theta \sum_{i=1}^1 B_i} + \dots + e^{\theta \sum_{i=1}^{N-1} B_i}} \quad (13)$$

根据式(11), 一个事件箱中, 除了 $t_0$ 时刻的清晰图像外, 剩余事件区间 $\{t_k\}_{k=1}^{N-1}$ 时刻的清晰图像 $\{C_k\}_{k=1}^{N-1}$ 可以表示为

$$C_k = \frac{NC_{\text{blur}} e^{\theta \sum_{i=1}^k B_i}}{1 + e^{\theta \sum_{i=1}^1 B_i} + \dots + e^{\theta \sum_{i=1}^{N-1} B_i}} \quad (14)$$

将得到的 $N$ 个清晰图像 $\{C_k\}_{k=0}^{N-1}$ 输入到COLMAP中, 得到对应的 $N$ 个相机位姿 $\{P_k\}_{k=0}^{N-1}$ , 具体为

$$\{P_k\}_{k=0}^{N-1} = \text{COLMAP}(\{C_k\}_{k=0}^{N-1}) \quad (15)$$

这种基于事件数据的相机位姿估计, 增加了点云初始化的鲁棒性, 使得3D高斯和MLP能够在后续的优化过程中更好地收敛。

### 1.4 3D重建与去模糊流程

所提方法的大致流程如图1所示。在E2GS(event enhanced Gaussian splatting)的基础上, 该方法额外引入了形状属性变换网络 $F_\theta$ 预测场景中每个3D高斯的变换属性, 从而增强模型的渲染效果。输入是一幅模糊图像及该视角对应的事件流, 对于合成数据, 直接使用Blender中提供的图像的相机位姿; 对于真实数据, 使用上述方法获取相机曝光时间下的位姿序列。然后将这些位姿输入到基于3DGS框架的模型中, 同时使用 $F_\theta$ 预测每个3D高斯的缩放因子。经过整个模型的渲染, 得到每个相机位姿下的像素 $(x, y)$ 处的颜色, 并计算模型预测的模糊图像 $\hat{C}_{\text{blur}}$ 和预测的事件箱 $\hat{B}_{(t_1, t_2)}$ 。最后, 根据真实输入的模糊图像 $C_{\text{blur}}$ 和事件箱 $B_{(t_1, t_2)}$ , 计算模糊损失 $\mathcal{L}_{\text{blur}}$ 和事件损失 $\mathcal{L}_{\text{event}}$ , 以此监督整个模型的训练过程。

为了从模糊的图像中渲染三维场景, 使用两种类型的损失函数: 图像渲染的模糊损失和事件损失。对于每个模糊图像, 假设是由相机曝光时间下运动轨迹中的 $N$ 个相机位姿 $\{P_i\}_{i=0}^{N-1}$ 渲染得到的 $N$ 个清晰图像 $\{\hat{C}_i\}_{i=0}^{N-1}$ 加权合成。由于将这些时间戳设置为该曝光时间下的等分, 所以可通过取 $N$ 幅渲染图像的均值模拟模糊图像, 具体为

$$\hat{C}_{\text{blur}} = \frac{1}{N} \sum_{i=0}^{N-1} \hat{C}_i \quad (16)$$

损失函数是通过比较 $\hat{C}_{\text{blur}}$ 和 $C_{\text{blur}}$ 的L1损失(L1

loss)与D-SSIM损失(difference structural similarity loss)。使用权重参数 $w_{\{D-SSIM\}}$ 平衡这两个损失之间的权重,具体为

$$\mathcal{L}_{\text{blur}} = (1 - w_{D-SSIM})\mathcal{L}_1 + w_{D-SSIM}\mathcal{L}_{D-SSIM} \quad (17)$$

式中, $w_{\{D-SSIM\}}$ 设置为0.2,L1损失为

$$\mathcal{L}_1 = \frac{\sum_k |\hat{C}_{\text{blur}} - C_{\text{blur}}|}{M} \quad (18)$$

式中, $M$ 表示图像中像素点的数量,D-SSIM损失为

$$\mathcal{L}_{D-SSIM} = 1 - \frac{(2\mu_1\mu_2 + c_1)(2\sigma_{12} + c_2)}{(\mu_1^2 + \mu_2^2 + c_1)(\sigma_1^2 + \sigma_2^2 + c_2)} \quad (19)$$

式中, $\mu_1$ 和 $\mu_2$ 分别表示图像 $\hat{C}_{\text{blur}}$ 和 $C_{\text{blur}}$ 的均值, $\sigma_1^2$ 和 $\sigma_2^2$ 分别是图像 $\hat{C}_{\text{blur}}$ 和 $C_{\text{blur}}$ 的方差。 $\sigma_{12}$ 是图像 $\hat{C}_{\text{blur}}$ 和 $C_{\text{blur}}$ 的协方差。 $c_1 = (k_1L)^2$ 和 $c_2 = (k_2L)^2$ 是用于维持稳定性的常数, $L$ 是图像像素值的动态范围,对于合成数据集和真实数据集,都为255。 $k_1$ 和 $k_2$ 是设置为较小值的常数,在该模型框架中, $k_1 = 0.01$ , $k_2 = 0.03$ 。SSIM是一种衡量两幅图像相似度的指标,它考虑了图像的亮度、对比度和结构信息,相较于传统的基于像素差异(如均方误差(mean squared error, MSE))的方法,更符合人类视觉系统对图像质量的感知。

图像渲染的模糊损失只是使用平均 $N$ 幅图像的方法模拟一帧模糊图像,没有考虑形成模糊图像过程中,相机在持续曝光下运动的时间过程。为了利用事件信息监督高时间分辨率的连续模糊过程,同时采用了事件损失。给定 $N$ 个相机位姿下的渲染图像,首先从 $\{\hat{C}_{t_i}\}_{i=0}^{N-1}$ 中随机选择两帧 $\{C_{t_1}, C_{t_2}\}(t_1 < t_2)$ ,并将其转换为灰度强度图像 $G_{t_1}$ 和 $G_{t_2}$ 。在对数域中取两个灰度图像的强度值之差,并除以每个像素 $(x, y)$ 的事件阈值 $\Theta$ ,以估计两帧之间的事件箱 $\hat{B}_{(t_1, t_2)}$ ,具体为

$$\hat{B}_{(t_1, t_2)}(x, y) = \begin{cases} \left\lfloor \frac{\log(G_{t_1}) - \log(G_{t_2})}{\Theta} \right\rfloor & G_{t_1}(x, y) < G_{t_2}(x, y) \\ \left\lceil \frac{\log(G_{t_1}) - \log(G_{t_2})}{\Theta} \right\rceil & G_{t_1}(x, y) \geq G_{t_2}(x, y) \end{cases} \quad (20)$$

事件箱为每个像素存储了由于亮度变化而触发的事件的数量,使用均方误差计算估计的事件箱 $\hat{B}_{(t_1, t_2)}$ 与真实事件箱 $B_{(t_1, t_2)}$ 之间的误差。具体为

$$\mathcal{L}_{\text{event}} = \sum \left\| \hat{B}_{(t_1, t_2)}(x, y) - B_{(t_1, t_2)}(x, y) \right\|_2^2 \quad (21)$$

最终,通过使用一个权重参数 $w_{\text{event}}$ 将两个损失函数 $\mathcal{L}_{\text{blur}}$ 和 $\mathcal{L}_{\text{event}}$ 结合起来,从而得到总损失函数,具体为

$$\mathcal{L} = \mathcal{L}_{\text{blur}} + w_{\text{event}}\mathcal{L}_{\text{event}} \quad (22)$$

式中, $w_{\text{event}} = 0.005$ 。在每一轮的训练迭代中,依据损失函数关于模型参数的梯度信息,沿着梯度的反方向对3DGS模型和网络 $F_\theta$ 的参数进行更新,使得模型在后续的预测中能够更接近真实的场景。通过不断地优化模型和网络的参数,使得整个模型能够逐渐学习到输入数据中的特征和规律,从而实现真实场景的准确3D建模。在实际应用阶段,向整个训练好的模型输入一个特定的相机位姿信息,模型能够对场景中的物体进行光线追踪、辐射度计算等一系列复杂的渲染操作,最终输出一幅清晰的渲染图像。

## 2 实验结果

本文分别使用合成数据集和真实数据集训练该模型。

对于合成场景数据,在神经辐射场(NeRF)中的6个合成场景:椅子(chair)、榕树(ficus)、热狗(hotdog)、乐高(LEGO)、材质(materials)和麦克风(mic)的基础上进行了扩展。利用Blender中的Camera Shakify插件模拟相机抖动。对于每个图像视角,渲染出在相机抖动过程中拍摄到的17幅清晰图像,并记录它们相应的相机位姿。随后,将这17幅图像输入到事件模拟工具V2E(video to events)(Hu等,2021)中,以模拟相机抖动过程中产生的事件数据。此外,为了得到模拟的模糊图像,首先使用逆图像信号处理(inverse isp)将这17幅图像转换到原始域,然后将它们叠加。接着,再通过图像信号处理(isp)得到最终的模糊图像。每个场景都有100个视角的模糊图像以及相应的事件数据。

对于真实场景数据,使用DAVIS346(Brandli等,2014)彩色事件相机采集真实数据。这款相机能够同时采集时间域和空间域相对应的事件数据和RGB图像。相机的分辨率为 $346 \times 260$ 像素,RGB图像的曝光时间设置为100 ms。通过手持相机进行拍摄的方式,采集了5个具有挑战性的场景:字母(letter)、乐高(LEGO)、相机(camera)、植物(plant)和玩具(toys),这些场景在低光照环境(5~100 lux)下

包含丰富的颜色变化和纹理细节。每个场景在不同视角下有 30 幅具有不同程度模糊的图像以及相应的事件数据。

为了评估利用事件数据解决图像去模糊任务和新视角合成任务的有效性,该研究将该模型分别与 E2NeRF (event enhanced neural radiance fields) (Qi 等, 2023)、3DGS (Kerbl 等, 2023)、E2GS (Deguchi 等, 2024) 和 DeblurGS (deblurring 3D Gaussian splatting) (Lee 等, 2024) 进行了比较。E2NeRF 是一个利用基于神经辐射场 (NeRF) 的方式解决图像去模糊和新视角合成任务的方法。3DGS 方法利用 3D 高斯进行建模, 训练过程中高斯系数经过不断优化, 同时结合了基于神经网络的优化和显式结构化数据存储的优点, 可以准确地表示场景。E2GS 通过将事件相机数据和 3DGS 框架相结合, 从而实现场景重建和运动模糊的去除。DeblurGS 在 3DGS 框架的基础上, 通过改变高斯大小、增加高斯数量, 来模拟动态模糊的形成过程。

为了定量评估渲染图像的质量, 该研究采用了 3 种广泛认可的指标评估合成数据集的图像质量: 峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性指数度量 (structural similarity index measure, SSIM) 以及感知图像块相似性学习指标 (learned perceptual image patch similarity, LPIPS) (Zhang 等, 2018)。PSNR 常用于衡量图像的重建质量, 它通过比较原始图像和重建图像之间的差异进行计算, 值越大, 说明效果越好。SSIM 考虑了图像的亮度、对比度和结构信息, 更符合人类视觉系统对图像质量的感知, 值越大, 说明效果越好。LPIPS (Zhang 等, 2018) 是一种基于深度学习模型的感知相似性度量。它通过在预训练的深度学习神经网络 (如 VGG (Visual Geometry Group)) 中提取图像特征, 然后计算这些特征之间的距离来衡量图像的相似性, 值越小, 说明效果越好。由于真实世界的数据中不存在真实的清晰图像, 该研究使用盲/无参考图像空间质量评估器 (blind/referenceless image spatial quality evaluator, BRISQUE) (Mittal 等, 2012)。该指标基于亮度分布, 无需任何参考就能评估图像的自然度, 值越小, 说明效果越好。

### 2.1 合成数据集

图 2 和表 1 分别展示了不同模型在合成数据集上的渲染对比结果。图 2 中的参考图像为未经模糊

处理的清晰图像, 而其他方法的渲染结果均基于模糊图像输入进行 3D 场景重建。具体而言, 这些方法首先利用模糊图像完成场景重建, 随后通过输入固定相机位姿生成渲染图像。渲染结果与清晰参考图像的接近程度直接反映了模型从模糊图像重建 3D 场景的能力。表 1 中的数据则是通过计算 100 个视角的渲染结果并取平均值得到的。实验结果表明, 当输入图像为模糊图像时, 该方法能够更准确地重建场景, 并生成更为清晰的渲染图像。与 3DGS 和 DeblurGS 相比, 该方法在所有合成数据集场景中均表现出显著优势。此外, 与 E2NeRF 和 E2GS 相比, 该方法在大多数数据集上能够达到更优或相当的性能水平, 进一步验证了其在场景重建和图像去模糊任务中的有效性和鲁棒性。

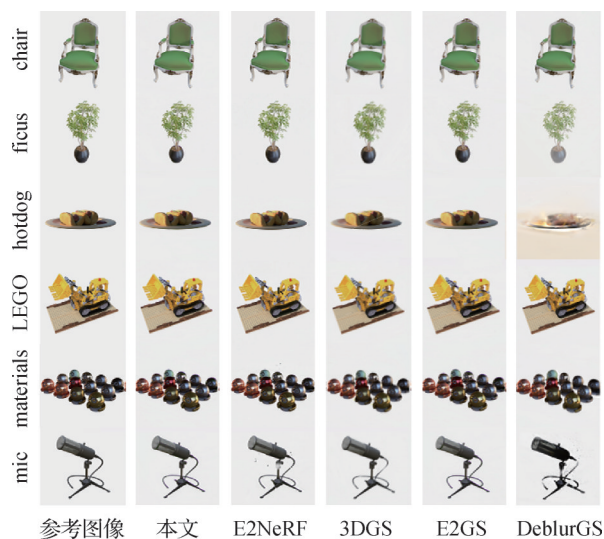


图 2 合成数据集渲染对比结果

Fig. 2 Comparison results of synthetic dataset rendering

### 2.2 真实数据集

图 3 和表 2 分别展示了不同模型在真实场景数据集上的渲染对比结果。由于真实场景数据中缺乏真实的清晰图像作为参考, 图 3 中最左侧的对比图直接采用了输入到模型中的模糊图像。表 2 中使用的评价指标为 BRISQUE (盲图像空间质量评估), 其值越小, 表明渲染图像的质量越高。表 2 中的数据是通过各模型对 17 个训练视角的渲染结果取均值后得到的。实验结果表明, 与 3DGS 和 DeblurGS 相比, 该方法在所有真实数据集场景中均表现出显著优势, 能够在重建高质量 3D 场景的基础上, 生成更为清晰的渲染图像。与 E2NeRF 和 E2GS 相比, 该方法

表1 合成数据集渲染对比结果  
Table 1 Rendering results on synthetic dataset

方法	chair			ficus			hotdog		
	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS
E2NeRF	0.96	31.54	<b>0.05</b>	0.91	25.51	0.12	<b>0.96</b>	<b>33.63</b>	<b>0.07</b>
3DGS	0.90	24.32	0.12	0.86	22.82	0.14	0.90	27.18	0.14
E2GS	<b>0.97</b>	31.27	<b>0.05</b>	<b>0.92</b>	25.64	0.11	<b>0.96</b>	32.71	<b>0.07</b>
DeblurGS	0.93	27.78	0.07	0.83	18.45	0.16	0.57	12.42	0.41
本文	<b>0.97</b>	<b>31.90</b>	<b>0.05</b>	<b>0.92</b>	<b>26.33</b>	<b>0.10</b>	<b>0.96</b>	31.67	<b>0.07</b>

方法	LEGO			materials			mic		
	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS
E2NeRF	0.91	28.11	0.12	<b>0.93</b>	27.81	<b>0.08</b>	<b>0.93</b>	27.08	<b>0.08</b>
3DGS	0.78	22.02	0.18	0.81	20.10	0.20	0.87	19.76	0.17
E2GS	<b>0.93</b>	28.49	<b>0.09</b>	<b>0.93</b>	<b>28.65</b>	<b>0.08</b>	<b>0.93</b>	27.68	<b>0.08</b>
DeblurGS	0.89	26.98	<b>0.09</b>	0.78	17.44	0.20	0.82	16.30	0.38
本文	<b>0.93</b>	<b>28.84</b>	<b>0.09</b>	<b>0.93</b>	27.86	<b>0.08</b>	<b>0.93</b>	<b>27.71</b>	<b>0.08</b>

注:加粗字体表示各列最优结果。

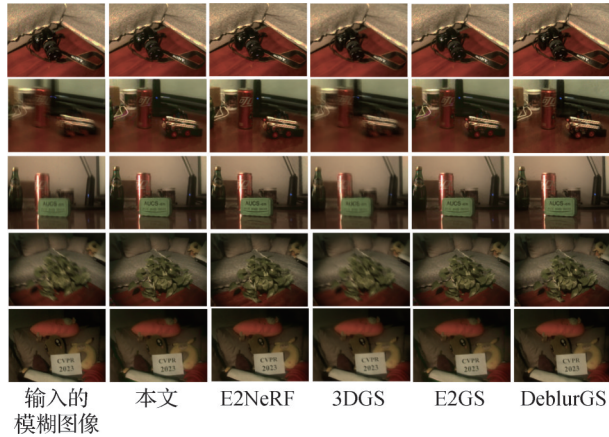


图3 真实数据集渲染对比结果

Fig. 3 Comparison results of rendering of real datasets

表2 真实数据集渲染对比结果  
Table 2 BRISQUE comparison results real datasets rendering

方法	camera	LEGO	letter	plant	toys
E2NeRF	28.76	35.12	43.27	33.54	44.72
3DGS	25.04	38.67	41.53	36.95	45.69
E2GS	21.43	<b>34.95</b>	37.83	31.79	<b>38.74</b>
DeblurGS	22.64	36.05	39.29	33.08	42.81
本文	<b>20.17</b>	35.81	<b>37.61</b>	<b>31.47</b>	39.28

注:加粗字体表示各列最优结果。

在大多数场景中均取得了更优的效果,仅在LEGO和toys两个场景中略微落后。这一结果进一步验证了该方法在真实场景数据集上的鲁棒性和有效性,展现了其在3D场景重建和图像去模糊任务中的潜力。

### 2.3 新视角渲染消融分析

基于3DGS框架的E2GS方法存在一个显著问题:模型容易对训练视角过度拟合,导致在新视角下的渲染效果显著下降。具体而言,模型在训练位姿下能够生成高质量的渲染结果,但在面对新视角时,渲染质量会出现明显退化。为了解决这一问题,该研究引入了网络 $F_\theta$ ,该网络通过预测3D高斯的变换属性(包括旋转变化 $\Delta r$ 和尺度变化 $\Delta s$ ),有效避免了3D高斯在空间中为拟合场景而过度调整自身形状,从而缓解了过拟合现象。图4和表3分别展示了该



图4 新视角渲染对比结果

Fig. 4 New perspective rendering comparison results

方法和E2GS对新视角的渲染效果对比。从图4中可以观察到,E2GS方法在新视角渲染时容易出现雾霭或颜色丢失等问题,尤其是在 chair、LEGO、material 和 mic 等场景中表现尤为明显。而表3的定量分

析进一步表明,引入 $F_\theta$ 后,模型在新视角下的渲染质量得到了显著提升。这一结果验证了 $F_\theta$ 在增强模型泛化能力和改善新视角渲染效果方面的有效性。

表3 新视角渲染对比结果  
Table 3 Novel-view rendering results

方法	chair			ficus			hotdog		
	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS
E2GS	0.72	16.69	0.37	0.89	24.27	0.21	0.95	30.62	0.08
本文	<b>0.97</b>	<b>30.87</b>	<b>0.05</b>	<b>0.92</b>	<b>25.23</b>	<b>0.10</b>	<b>0.96</b>	<b>31.49</b>	<b>0.07</b>
方法	LEGO			materials			mic		
	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS
E2GS	0.73	17.82	0.32	0.78	18.49	0.31	0.81	19.65	0.29
本文	<b>0.93</b>	<b>28.76</b>	<b>0.09</b>	<b>0.93</b>	<b>26.81</b>	<b>0.08</b>	<b>0.93</b>	<b>26.64</b>	<b>0.08</b>

注:加粗字体表示各列最优结果。

## 2.4 训练时间与渲染速度

此实验在 RTX4090 上对数据集进行训练和渲染。得益于3DGS框架对图像快速光栅化的优势,该方法大幅缩短了训练时间和渲染时间。如表4所示,相比于E2NeRF方法,平均训练时间从48 h缩减至30 min,且实现了140 帧/s的实时渲染速度。

表4 训练时间与渲染速度对比

Table 4 Comparison training time and rendering speed

方法	训练时间	渲染速度/(帧/s)
E2NeRF	48 h	0.04
DeblurGS	45 min	100
E2GS	<b>30 min</b>	<b>140</b>
本文	<b>30 min</b>	<b>140</b>

注:加粗字体表示各列最优结果。

## 2.5 基于观察视角建模的消融分析

该研究主要优势与贡献在于,通过引入网络 $F_\theta$ 预测3D高斯的变换属性,有效避免了基于3DGS框架的过拟合问题,加强模型对新视角的渲染能力。在新视角渲染任务中,观察视角对于网络的建模起到了至关重要的作用,为了进一步讨论该参数对于网络 $F_\theta$ 与整个模型的影响程度,该研究展示了不引入观察方向的实验对比结果。

当网络 $F_\theta$ 的输入不考虑观察方向 $\boldsymbol{v}$ 时,对3D高

斯的变换属性 $\Delta\boldsymbol{r}$ ,  $\Delta\boldsymbol{s}$ 预测表达式可以重写为

$$(\Delta\boldsymbol{r}_i, \Delta\boldsymbol{s}_i) = F_\theta(\boldsymbol{\gamma}(\boldsymbol{x}_i, \boldsymbol{s}_i, \boldsymbol{r}_i)) \quad (24)$$

此时,对于新视角的预测对比结果如图5和表5所示。通过图5可以观察到,当对网络 $F_\theta$ 的建模不再考虑观察视角的影响时,在所有的预测场景中,均出现了不同程度的色彩缺失等问题。对于每个场景,都进行了100个不同新视角的渲染,表5记录了所有渲染结果指标的均值,其中 $F_{\theta w/ov}$ 表示采用不考虑观察方向 $\boldsymbol{v}$ 的 $F_\theta$ 网络。这些定量结果进一步表明,建模过程中观察方向 $\boldsymbol{v}$ 的忽略,导致了新视角渲染质量的全面下降。这一实验结果验证了该方法建模的有效性,说明了基于观察视角的建模方法能有效提升模型新视角的渲染能力。

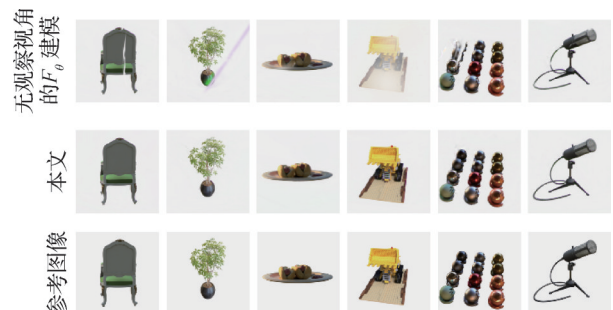


图5 有无观察视角建模对模型新视角渲染的影响  
Fig. 5 The impact of modeling with or without observation perspective on the rendering of the model from a new perspective

表5 有无观察视角建模对模型新视角渲染的影响

Table 5 Impact of viewpoint modeling on novel view rendering

方法	chair			ficus			hotdog		
	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS
无观察视角( $F_{\theta w/ov}$ )	0.95	28.73	0.06	0.90	24.35	0.12	<b>0.96</b>	31.01	0.08
有观察视角(本文)	<b>0.97</b>	<b>30.87</b>	<b>0.05</b>	<b>0.92</b>	<b>25.23</b>	<b>0.10</b>	<b>0.96</b>	<b>31.49</b>	<b>0.07</b>

方法	LEGO			materials			mic		
	SSIM	PSNR/dB	LPIPS	SSIM	PSNR/dB	LPIPS	SSIM	PSNR	LPIPS
无观察视角( $F_{\theta w/ov}$ )	0.91	26.32	0.11	0.91	24.09	0.13	0.92	23.76	0.11
有观察视角(本文)	<b>0.93</b>	<b>28.76</b>	<b>0.09</b>	<b>0.93</b>	<b>26.81</b>	<b>0.08</b>	<b>0.93</b>	<b>26.64</b>	<b>0.08</b>

注:加粗字体表示各列最优结果。

## 2.6 权重参数 $w_{event}$ 的消融分析

在损失函数的计算中,本研究设置了  $w_{event} = 0.005$ ,为了进一步讨论  $w_{event}$  的选值对于实验结果的影响,分别计算了  $w_{event} = 0.1$ 、 $w_{event} = 0.01$ 、 $w_{event} = 0.001$  和  $w_{event} = 0.0001$  时,模型在训练视角和新视角下的渲染结果。

实验表明,当  $w_{event} = 0.005$  时,模型在大部分场景下都得到了最优的渲染效果。并且在实验过程中发现,随着  $w_{event}$  的增大,模型的整体训练时间也随之增加。

## 3 结论

本文提出了一种基于事件引导和三维高斯泼溅的去模糊重建框架。通过融合事件相机数据并提出形状属性变换网络,实现从模糊图像中鲁棒地重建出清晰的3DGS场景。该方法不仅有效去除了RGB图像中的运动模糊,并实现了高质量的新视角实时合成。基于合成数据集与真实世界数据集的实验表明,该方法在渲染精度(如PSNR/SSIM指标显著提升)与训练效率方面均优于对比方法,其训练速度是基于NeRF的方法(如E2NeRF)的48倍(本文方法单场景训练时间 < 1 h,而E2NeRF需48 h)。框架保持了3DGS的实时渲染能力( $\geq 100$  帧/s),并通过高斯属性调制有效缓解了新视角生成中的过拟合问题。未来的研究方向包括将框架扩展至处理快速移动物体的动态场景(如体育赛事或低光环境),并围绕3DGS框架继续开展焦散模糊去除方向的研究工作。

## 参考文献(References)

- Akolkar H, Ieng S H and Benosman R. 2022. Real-time high speed motion prediction using fast aperture-robust event-driven visual flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 361-372 [DOI: 10.1109/TPAMI.2020.3010468]
- Barron J T, Mildenhall B, Tancik M, Hedman P, Martin-Brualla R and Srinivasan P P. 2021. Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 5855-5864 [DOI: 10.1109/ICCV48922.2021.00580]
- Brandli C, Berner R, Yang M H, Liu S C and Delbruck T. 2014. A 240 × 180 130 dB 3  $\mu$ s latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10): 2333-2341 [DOI: 10.1109/JSSC.2014.2342715]
- Chen J B, Xiong B S, Kuang F and Zhang Z Z. 2023. Motion deblurring based on deep feature fusion attention and double-scale. *Journal of Image and Graphics*, 28(12): 3731-3743 (陈加保, 熊邦书, 况发, 章照中. 2023. 深度特征融合注意力与双尺度的运动去模糊. *中国图象图形学报*, 28(12): 3731-3743) [DOI: 10.11834/jig.220931]
- Deguchi H, Masuda M, Nakabayashi T and Saito H. 2024. E2GS: event enhanced Gaussian splatting//*Proceedings of 2024 IEEE International Conference on Image Processing (ICIP)*. Abu Dhabi, United Arab Emirates: IEEE: 3456-3460 [DOI: 10.1109/ICIP.2024.10380860]
- Hu Y H, Liu S C and Delbruck T. 2021. v2e: from video frames to realistic DVS events//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Nashville, USA: IEEE: 1312-1321 [DOI: 10.1109/CVPRW53098.2021.00144]
- Hwang I, Kim J and Kim Y M. 2023. Ev-NeRF: event based neural radiance field//*Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 837-847 [DOI: 10.1109/WACV56688.2023.00090]

- Kerbl B, Kopanas G, Leimkuehler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 42(4): #139 [DOI: 10.1145/3592433]
- Lee B, Lee H, Sun X Y, Ali U and Park E. 2024. Deblurring 3D Gaussian splatting//*Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 127-143 [DOI: 10.1007/978-3-031-73636-0\_8]
- Lee D, Kim J, Park J and Kweon I S. 2023. DP-NeRF: deblurred neural radiance field with physical scene priors//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, Canada: IEEE/CVF: 18124-18133 [DOI: 10.1109/CVPR.2023.01762]
- Lichtsteiner P, Posch C and Delbruck T. 2008. A  $128 \times 128$  120 dB 15  $\mu$ s latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2): 566-576 [DOI: 10.1109/JSSC.2007.914337]
- Ma L, Li X Y, Liao J, Zhang Q, Wang X, Wang J, et al. 2022. Deblur-NeRF: neural radiance fields from blurry images//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 12861-12870 [DOI: 10.1109/CVPR52688.2022.01252]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2022. NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99-106 [DOI: 10.1145/3503250]
- Mittal A, Moorthy A K and Bovik A C. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12): 4695-4708 [DOI: 10.1109/TIP.2012.2214050]
- Müller T, Evans A, Schied C and Keller A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG)*, 41(4): #102 [DOI: 10.1145/3528223.3530127]
- Pan L Y, Scheerlinck C, Yu X, Hartley R, Liu M M and Dai Y C. 2019. Bringing a blurry frame alive at high frame-rate with an event camera//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 6820-6829 [DOI: 10.1109/CVPR.2019.00698]
- Pumarola A, Corona E, Pons-Moll G and Moreno-Noguer F. 2021. D-NeRF: neural radiance fields for dynamic scenes//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 10318-10327 [DOI: 10.1109/CVPR46437.2021.01018]
- Qi Y S, Zhu L, Zhang Y and Li J. 2023. E2NeRF: event enhanced neural radiance fields from blurry images//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 837-847 [DOI: 10.1109/ICCV51070.2023.01219]
- Qiu Y F, Liu Z Y and Wang M H. 2025. Image deblurring network combining Mamba and snake-like convolution. *Journal of Image and Graphics*, 30(10): 3187-3198 (邱云飞, 刘则延, 王茂华. 2025. 融合 Mamba 与蛇形卷积的图像去模糊网络. *中国图象图形学报*, 30(10): 3187-3198) [DOI: 10.11834/jig.240618]
- Rudnev V, Elgharib M, Theobalt C and Golyanik V. 2023. EventNeRF: neural radiance fields from a single colour event camera//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 4992-5002 [DOI: 10.1109/CVPR52729.2023.00483]
- Schönberger J L and Frahm J M. 2016. Structure-from-motion revisited//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 4104-4113 [DOI: 10.1109/CVPR.2016.445]
- Takatani T, Ito Y, Ebisu A, Zheng Y Q and Aoto T. 2021. Event-based bispectral photometry using temporally modulated illumination//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 15638-15647 [DOI: 10.1109/CVPR46437.2021.01538]
- Xie B, Li Y X, Shao X and Dai B Q. 2025. Multiscale image deblurring based on prompt learning and gated feedforward networks. *Journal of Image and Graphics*, 30(3): 755-768 (谢斌, 黎彦先, 邵祥, 戴邦强. 2025. 提示学习与门控前馈网络的多尺度图像去模糊. *中国图象图形学报*, 30(3): 755-768) [DOI: 10.11834/jig.240315]
- Zhang J Q, Dong B, Zhang H W, Ding J C, Heide F, Yin B C, et al. 2022. Spiking transformers for event-based single object tracking//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 8801-8810 [DOI: 10.1109/CVPR52688.2022.00860]
- Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]
- Zhao L Z, Wang P and Liu P D. 2024. Bad-Gaussians: bundle adjusted deblur Gaussian splatting//*Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 233-250 [DOI: 10.1007/978-3-031-72698-9\_14]
- Zwicker M, Pfister H, van Baar J and Gross M. 2001. EWA volume splatting//*Proceedings of 2001 Conference on Visualization*. San Diego, USA: IEEE: 29-36 [DOI: 10.1109/VISUAL.2001.964490]

## 作者简介

樊越,男,硕士研究生,主要研究方向为计算机视觉。

E-mail: linyao153580@gmail.com

王逸群,通信作者,男,副教授,主要研究方向为人工智能、计算机图形学、模式识别、表示学习和计算机视觉。

E-mail: yiqun.wang@cqu.edu.cn

陆韬宇,男,本科生,主要研究方向为计算机视觉。

E-mail: lutaoyu@stu.cqu.edu.cn