Event-driven Motion Deblurring via Trajectory-based Kernel Reconstruction

Zhiwei Zhong[1], Peilin Chen[1], Wei Dong[2], Bo Li[2], Anmin Liu[2], and Shiqi Wang[1⋆]

[1] City University of Hong Kong, Hong Kong, China
[2] vivo BlueImage Lab, vivo Mobile Communication Co., Ltd., China
zhwzhong.cs@gmail.com, shiqwang@cityu.edu.hk

Abstract. Recovering a sharp image from a motion-blurred observation remains challenging due to the inherently ill-posed nature of blind deblurring, especially under real-world non-uniform blur caused by complex camera and scene motion. Event cameras, with their microsecond-level temporal resolution and inherent sensitivity to motion, provide rich motion cues that can serve as strong priors for resolving such ambiguities. However, existing event-guided deblurring methods either rely on simplified blur assumptions or treat event data merely as generic features, without fully exploiting their physical relationship with the image formation process. In this paper, we propose a complete event-driven non-uniform blind deblurring framework that explicitly models the physical formation of spatially varying blur kernels. Specifically, we first estimate dense pixel-wise motion trajectories from the event stream via a differentiable event-alignment objective. These trajectories are then used to construct per-pixel point spread functions (PSFs), which provide a physically grounded blur operator. Based on this operator, we formulate a joint optimization framework that enforces blur consistency, event consistency, and image regularization. Finally, we design an unrolled optimization network that alternates between data-consistency reconstruction and learned image priors. Experiments on both synthetic and real-world datasets demonstrate that the proposed method improves deblurring performance while maintaining low computational complexity.

Keywords: Image deblurring · Event camera · Physics-based Vision

1 Introduction

Motion blur occurs when the camera and the scene move during exposure [11,18]. It removes fine details and harms the performance of tasks such as detection, tracking, and SLAM [15, 38, 65]. As a result, robust motion deblurring remains a long-standing and fundamental problem in computer vision. Traditional image deblurring methods often rely on the assumption of spatially invariant blur, where a single blur kernel is shared across the entire image [41, 54, 58]. While

⋆ Corresponding author.

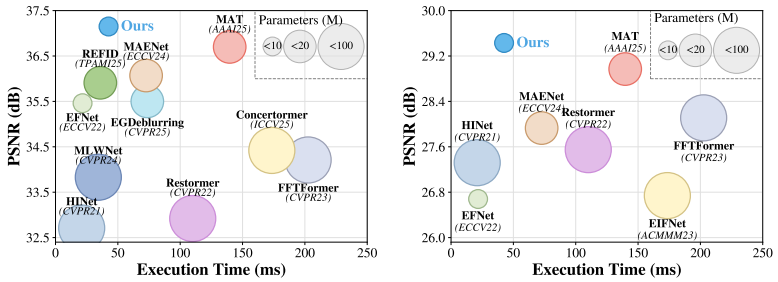
this assumption simplifies the problem formulation and optimization, it is rarely satisfied in real-world scenarios. In practice, motion blur is usually spatially varying due to complex camera motion, depth variations in the scene, and non-rigid object motion [16,24,28]. In such cases, the blur kernel differs across pixels, making blind deblurring significantly more ill-posed. This spatially varying nature of motion blur poses substantial challenges for classical single-image deblurring methods and often leads to suboptimal reconstruction quality [10, 32].

Recent deep learning approaches have improved non-uniform deblurring by learning strong image priors [13, 21, 42, 48]. However, most methods rely only on blurred RGB inputs and face a fundamental ambiguity: many combinations of sharp images and spatially varying blur kernels can lead to the same blurred observation. Without reliable motion cues, these methods tend to approximate blur effects rather than explicitly recover the underlying motion structure, which often leads to artifacts, especially under fast or spatially inconsistent motion.

Event cameras (or Dynamic Vision Sensors), representing a paradigm shift in visual sensing, offer a promising solution to this challenge. Unlike standard frame-based cameras that integrate intensity over a fixed exposure time, event cameras asynchronously record changes in log-intensity with microsecond-level temporal resolution [3, 9, 33, 59]. This unique mechanism allows them to capture rich, high-frequency motion information with high dynamic range and no motion blur. Since the event stream essentially encodes the continuous trajectory of motion relative to the scene texture, it provides strong, natural physical cues that are complementary to the accumulated intensity of the blurred frame. Theoretically, if the motion trajectory during the exposure can be accurately recovered from events, the ill-posed nature of deblurring can be largely resolved.

Despite this potential, effectively leveraging event data for non-uniform blind image deblurring remains an open research question. Existing event-guided deblurring methods can be broadly categorized into two groups. The first group adopts end-to-end data-driven approaches, where event streams are treated as supplementary feature modalities and fused with image features to directly learn the mapping from blurred to sharp images via deep neural networks [5, 8, 56, 67, 70]. While conceptually straightforward, these methods often ignore the underlying physical relationship between the event stream and the blur formation process, leading to suboptimal exploitation of motion cues and limited generalization capability. The second group attempts to incorporate physical constraints but often relies on simplified assumptions, such as uniform blur kernels or linear motion models [31, 53, 68]. Such assumptions rarely hold in dynamic real-world environments where motion is complex and spatially varying. Consequently, there is a lack of a unified framework that can handle complex non-uniform blur while explicitly respecting the physical formation model of the event-blur relationship.

To bridge this gap, we propose an event-driven non-uniform motion deblurring framework that explicitly models the physical formation of spatially varying blur kernels. Instead of treating events as latent features, we first estimate dense, pixel-wise motion trajectories from the event stream via a differentiable event-alignment objective. These trajectories are then utilized to construct precise



(a) GoPro dataset [30]. **(b)** HS-ERGB dataset [44].

Fig. 1: Model complexity and performance comparison on benchmark datasets. The figure compares different methods in terms of PSNR, model parameters, and execution time. For a fair comparison, the execution time of all methods is measured on a single NVIDIA RTX 5090 GPU, with the input image resolution fixed to 480×480.

per-pixel point spread functions (PSFs), providing a physically grounded blur operator. Based on this operator, we formulate the deblurring task as a joint optimization problem that enforces consistency among the blurred image, the event stream, and image regularity. To solve this efficiently, we adopt a deep unfolding strategy and design an unrolled optimization network that alternates between a data-consistency module and a prior-learning module. This architecture combines the interpretability of model-based optimization with the strong representation capability of deep learning. Our main contributions are as follows:

- We propose a novel event-driven framework for non-uniform motion deblurring that explicitly models the physical formation of spatially varying blur. Unlike prior works that either oversimplify blur assumptions or treat events as generic features, our method leverages the physical relationship between events and blur as the core modeling principle.
- We introduce a differentiable event-alignment objective to estimate dense, pixel-wise motion trajectories directly from the event stream. By leveraging these trajectories to construct precise per-pixel point spread functions (PSFs), we alleviate the ill-posed ambiguity of blind deblurring with reliable, high-frequency motion cues.
- We design a deep unfolding architecture that unrolls the joint optimization of the blurred image and event data. This approach integrates the interpretability of physical models with the representation power of deep learning to achieve superior deblurring performance. As shown in Fig. 1, our method achieves better performance with lower computational complexity.

2 Related Works

2.1 Frame-based Image and Video Deblurring

Frame-based image and video deblurring has been extensively studied over the past decades. Early methods mainly rely on variational optimization frameworks,

where blur is modeled as a convolution between a latent sharp image and a blur kernel, together with handcrafted image priors such as total variation or sparsity constraints [22, 25, 54]. However, these approaches often assume spatially invariant blur or simple parametric motion models, which limits their applicability in real-world scenarios. To address spatially varying blur, several works introduce locally adaptive kernels or segmentation-based strategies [16, 47, 55]. While these methods improve modeling flexibility, they significantly increase computational complexity and remain sensitive to noise and initialization due to the inherently ill-posed nature of blind deconvolution. Video-based approaches further exploit temporal redundancy and motion coherence across frames, leveraging optical flow or multi-frame alignment to better constrain the deblurring problem [6,34]. However, their performance heavily depends on accurate motion estimation, which itself becomes unreliable under severe motion blur.

Recent years have witnessed significant progress in image and video deblurring driven by deep learning. Convolutional neural network (CNN) based methods employ multi-scale, multi-branch, or divide-and-conquer architectures to directly regress sharp images from blurred inputs, achieving substantial improvements over traditional optimization-based approaches [17,30,50,66]. Meanwhile, lightweight and efficient baseline models demonstrate that carefully designed network structures can achieve state-of-the-art performance with significantly reduced computational cost [1]. To better handle spatially varying blur, recent works incorporate kernel-aware modeling and disparity cues into the learning framework [62]. In addition, Transformer-based architectures capture long-range spatial and temporal dependencies via self-attention, improving robustness to large-motion blur [27,63]. More recently, diffusion-based generative models have been explored to model the inherent uncertainty of deblurring and synthesize realistic blur patterns for data augmentation [46, 48]. Despite these advances, most frame-based methods lack explicit supervision of motion during exposure, limiting their ability to recover physically accurate non-uniform blur.

2.2 Event-based Motion Deblurring

Event cameras provide an alternative sensing modality that naturally avoids motion blur by asynchronously recording brightness changes with microsecond temporal resolution [12]. Their ability to capture fine-grained motion information has inspired growing interest in event-based motion deblurring. Early works focus on reconstructing sharp images or videos directly from events by integrating events over time or solving inverse problems with handcrafted priors [29,36].

More recent approaches leverage the complementary strengths of events and blurred intensity frames. A common strategy is to treat events as an additional input modality and fuse them with RGB images in an end-to-end deep network to predict sharp images [4,5,7,40,49,51,56,57,61,64,67]. While effective in practice, these methods typically regard events as generic spatiotemporal features and do not explicitly model the physical relationship between event generation and blur formation. As a result, motion information contained in events is not fully exploited, and generalization to unseen motion patterns remains limited.

Another line of research incorporates physical constraints by using events to estimate motion cues such as optical flow or camera trajectories, which are then employed to guide the deblurring process [31, 53]. To keep the problem tractable, some methods adopt simplified blur models, including assumptions of uniform blur kernels or linear motion during the exposure time [68]. However, such assumptions are rarely satisfied in real-world scenarios, including depth discontinuities, non-rigid object motion, and complex camera dynamics. As a result, performance often degrades under spatially varying blur.

In summary, although event cameras provide rich motion cues that are highly beneficial for deblurring, existing methods either underutilize this information in purely data-driven frameworks or rely on overly restrictive physical assumptions. This leaves a gap for a unified approach that can fully exploit event data while explicitly modeling the physical formation of spatially varying motion blur. Our work addresses this gap by estimating dense pixel-wise motion trajectories from events and constructing physically grounded per-pixel blur operators within a deep unfolding optimization framework.

3 Method

We propose an event-guided motion deblurring framework that explicitly exploits the event stream to model pixel-wise motion. In contrast to prior event-based approaches that treat events merely as auxiliary features, our method leverages the physical relationship between motion blur formation and event-triggered brightness changes to construct a principled, spatially varying blur model.

The key observation is that motion blur arises from the temporal integration of pixel-wise motion trajectories over the exposure interval. Since event cameras record brightness changes along true motion trajectories with high temporal resolution, these trajectories can be directly inferred from the event stream.

Motivated by this insight, we first estimate dense motion trajectories from events via a differentiable event-alignment objective. Based on the estimated trajectories, we construct per-pixel point spread functions (PSFs), yielding a spatially varying blur operator. Motion deblurring is then formulated as a joint optimization problem that enforces blur consistency, event consistency, and image regularization. An overview of the proposed framework is shown in Fig. 2.

3.1 Physical Modeling of Spatially Varying Blur

Motion blur is caused by continuous camera and scene motion during the exposure time. Let $\mathbf{x}$ denote a pixel location in the image plane. During exposure, this pixel follows a motion trajectory $\Phi_t(\mathbf{x})$ over the interval $t \in [0, T]$. As a result, the observed blurry image B is formed by integrating the latent sharp image S along this trajectory:

$$B(\mathbf{x}) = \frac{1}{T} \int_0^T S\big(\Phi_t(\mathbf{x})\big)\, dt + n(\mathbf{x}), \tag{1}$$

where $n(\mathbf{x})$ denotes sensor noise.

By discretizing the exposure time into M temporal samples $\{t_m\}_{m=1}^{M}$, Eq. (1) can be approximated as a spatially varying convolution:

$$B(\mathbf{x}) = \sum_{\mathbf{d}} k_{\mathbf{x}}(\mathbf{d})\, S(\mathbf{x}+\mathbf{d}) + n(\mathbf{x}), \tag{2}$$

where $\mathbf{d}$ denotes a spatial offset and $k_{\mathbf{x}}(\mathbf{d})$ is the pixel-wise point spread function:

$$k_{\mathbf{x}}(\mathbf{d}) = \frac{1}{M} \sum_{m=1}^{M} \delta\left(\mathbf{d} - (\Phi_{t_m}(\mathbf{x}) - \mathbf{x})\right), \tag{3}$$

with $\delta(\cdot)$ being the Dirac delta function. Once the motion trajectories are known, the spatially varying blur kernel field $K = \{k_{\mathbf{x}}\}$ can be analytically derived.

3.2 Event-driven Motion Estimation

Event cameras asynchronously trigger events in response to changes in log-intensity. An event $e_k = (\mathbf{x}_k, t_k, p_k)$ is generated when

$$\log I(\mathbf{x}_k, t_k) - \log I(\mathbf{x}_k, t_k^-) = p_k C, \tag{4}$$

where $p_k \in \{-1, +1\}$ denotes the polarity and C is the contrast threshold. Since events are generated along true motion trajectories, they provide reliable motion cues even under severe motion blur. Let $v(\mathbf{x}, t)$ denote the dense motion field. Given a reference time t_0, an event can be motion-compensated to

$$\tilde{\mathbf{x}}_k = \mathbf{x}_k - \int_{t_k}^{t_0} v(\mathbf{x}_k, \tau)\, d\tau. \tag{5}$$

If the motion field is accurately estimated, motion-compensated events corresponding to the same scene edge will become spatially aligned. Therefore, we formulate an event-alignment objective that encourages the formation of sharp and high-contrast structures after motion compensation:

$$\mathcal{L}_{evt}(v) = -\sum_{\mathbf{x}} \left\| \nabla \left(\sum_{k} p_k\, \delta(\mathbf{x} - \tilde{\mathbf{x}}_k) \right) \right\|_2^2. \tag{6}$$

To improve robustness in regions with sparse events and suppress noise, we further impose a spatial smoothness prior on the motion field:

$$\mathcal{R}_v(v) = \|\nabla v\|_1. \tag{7}$$

The final motion estimation problem is formulated as

$$\hat{v} = \arg\min_{v} \left(\mathcal{L}_{evt}(v) + \lambda_v \mathcal{R}_v(v) \right). \tag{8}$$

This optimization can be solved using gradient-based methods or approximated by a compact neural module. With the estimated motion field, pixel-wise trajectories $\Phi_{t_m}(\mathbf{x})$ are obtained and used to construct physically grounded PSFs via Eq. (3). In practice, temporal displacements are accumulated onto a discrete kernel grid using bilinear splatting, followed by normalization to preserve energy.

3.3 Joint Deblurring with Event Consistency

Given the event-derived spatially varying blur operator, our goal is to recover the latent sharp image S from the observed blurry image B. We formulate motion deblurring as the following optimization problem:

$$\min_{S} \ \mathcal{L}_{blur}(S) + \mu\, \mathcal{L}_{cons}(S, E) + \lambda_s \mathcal{R}_S(S), \tag{9}$$

where $\mathcal{L}_{blur}$ enforces physical blur consistency, $\mathcal{L}_{cons}$ incorporates event-guided structural constraints, and $\mathcal{R}_S$ denotes an image regularization term.

Blur consistency.

$$\mathcal{L}_{blur}(S) = \|B - \mathcal{K}(K)S\|_F^2. \tag{10}$$

This term ensures that the reconstructed image, when re-blurred using the generated kernels, faithfully reproduces the observed blurry measurement.

Event consistency. Beyond motion estimation, the event stream also encodes reliable structural information. To transfer such information into the deblurring process, we enforce consistency between the reconstructed image S and an event-derived representation E at the level of structural responses:

$$\mathcal{L}_{cons} = \|\mathcal{A}(S) - \mathcal{A}(E)\|_F^2, \tag{11}$$

where $\mathcal{A}(\cdot)$ denotes an anisotropic structure response operator. Formally, for each pixel location i, the operator is defined as

$$\mathcal{A}(X)_i = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \big(X_j - X_i\big), \tag{12}$$

where $\mathcal{N}(i)$ denotes a local neighborhood centered at pixel i. The operator aggregates local intensity differences to highlight edge-aligned structures while suppressing isolated noise.

3.4 ADMM-inspired Unrolled Optimization

To solve Eq. (9), we derive an ADMM-inspired unrolled optimization architecture. By introducing an auxiliary variable Z, the original problem is decomposed into simpler subproblems. The augmented Lagrangian is given by

$$\begin{aligned} \mathcal{L}_\rho(S, Z, U) = \|B - \mathcal{K}(K)S\|_F^2 + \mu\|\mathcal{A}(S) - \mathcal{A}(E)\|_F^2 \\ + \lambda_s \mathcal{R}_S(Z) + \frac{\rho}{2}\|S - Z + U\|_F^2, \end{aligned} \tag{13}$$

where U denotes the scaled dual variable and ρ is the penalty parameter. The optimization proceeds by alternatively updating S, Z, and U according to

$$\begin{aligned} S^{k+1} &= \arg\min_{S} \ \|B - \mathcal{K}(K)S\|_F^2 + \mu\|\mathcal{A}(S) - \mathcal{A}(E)\|_F^2 + \frac{\rho}{2}\|S - Z^k + U^k\|_F^2, \\ Z^{k+1} &= \arg\min_{Z} \ \lambda_s \mathcal{R}_S(Z) + \frac{\rho}{2}\|S^{k+1} - Z + U^k\|_F^2, \\ U^{k+1} &= U^k + S^{k+1} - Z^{k+1}. \end{aligned} \tag{14}$$

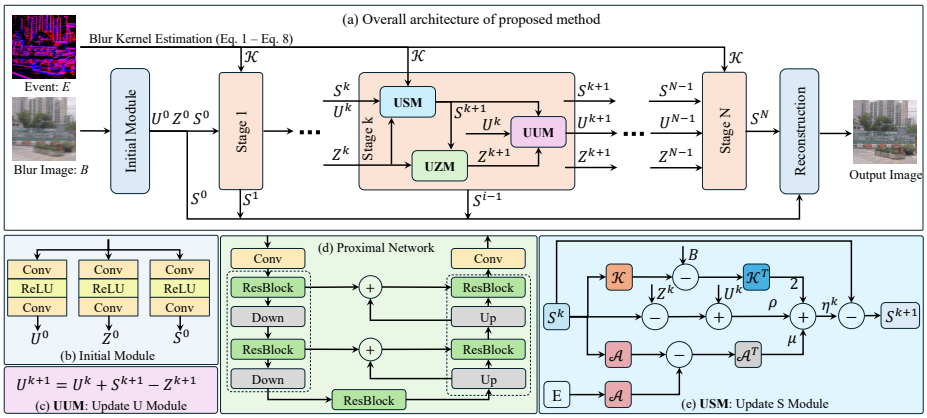


Fig. 2: Overview of the proposed event-guided deblurring framework. Given a blurred image B and the corresponding event data E, the event stream is first used to estimate dense motion trajectories and construct blur kernels $\mathcal{K}$. The deblurring problem is then solved using an ADMM-inspired unrolled network composed of multiple stages, each including an update S module (USM) for data consistency, an update Z module (UZM) for image priors, and an update U module (UUM) for dual variable update.

S-update. With Z and U fixed, the S-subproblem is solved via gradient descent:

$$S^{(k+1)} = S^{(k)} - \eta^{(k)} \nabla_S, \tag{15}$$

where the step size $\eta^{(k)}$ is a learnable parameter and the gradient is given by

$$\nabla_S = 2\mathcal{K}^\top(\mathcal{K}(K)S - B) + 2\mu\,\mathcal{A}^\top(\mathcal{A}(S) - \mathcal{A}(E)) + \rho\left(S - Z^{(k)} + U^{(k)}\right). \tag{16}$$

Z-update. The Z-subproblem takes the form of a proximal mapping associated with the image regularization term:

$$Z^{k+1} = \mathrm{prox}_{\frac{\lambda_S}{\rho}\mathcal{R}_S}\left(S^{k+1} + U^k\right). \tag{17}$$

Dual update. Finally, the scaled dual variable is updated by

$$U^{k+1} = U^k + S^{k+1} - Z^{k+1}. \tag{18}$$

3.5 Unrolled Network Architecture

Based on the proposed joint optimization formulation, we design an ADMM-inspired unrolled network for learnable motion deblurring. As shown in Fig. 2, the overall framework consists of two main components: a blur kernel estimation module and an unrolled optimization network. The kernel estimation module exploits the event stream to estimate dense motion trajectories and construct

spatially varying blur kernels in a physically grounded manner. Based on the estimated kernels, the unrolled network embeds the corresponding blur operator and solves the deblurring problem by unrolling the ADMM optimization for N stages. The unrolled network includes an initialization module followed by multiple iterative stages. The initialization module provides the initial estimates of the optimization variables, and each stage corresponds to one ADMM iteration composed of an update S module (USM), an update Z module (UZM), and an update U module (UUM). The details of each module are described below.

Initialization module. As shown in Fig. 2b, the initialization module takes the blurred image B as input and produces the initial estimates $\{S^0, Z^0, U^0\}$. It consists of two convolutional layers with a ReLU activation in between and is designed to provide a reasonable starting point for the subsequent optimization:

$$\{S^0, Z^0, U^0\} = \mathcal{F}_{\text{init}}(B), \tag{19}$$

where $\mathcal{F}_{\text{init}}(\cdot)$ denotes the initialization network.

Update S module (USM). This module updates the latent sharp image according to the ADMM framework (Eq. 15 and Fig. 2e). In this module, one gradient descent step is performed on the data fidelity and event consistency terms. Importantly, the spatially varying blur operator constructed from the event-derived kernels is explicitly embedded in this module, ensuring physical consistency between the reconstructed image and the observed blurry input.

Update Z module (UZM). The UZM corresponds to the proximal update in the ADMM framework, which is associated with the image regularization term. In traditional optimization-based deblurring methods, this proximal operator is usually defined by hand-crafted priors, such as total variation [37] or sparsity-based [14] regularization, leading to fixed and limited representations. Different from these approaches, we adopt a learning-based proximal operator to model the image prior. Specifically, the proximal mapping in Eq. 17 is parameterized by a neural network, which enables the model to learn more flexible and expressive priors directly from data. As illustrated in Fig. 2d, we implement the learned proximal operator using a lightweight encoder–decoder network. This design allows the model to effectively capture both local textures and global image structures while maintaining high computational efficiency, providing stronger regularization for the reconstruction process.

Update U module (UUM). The UUM updates the dual variable in a manner consistent with Eq. 18.

3.6 Network Training

The final output is obtained by a weighted aggregation of the intermediate reconstructions from all stages: $\hat{S} = \sum_{i=0}^{N} w_i \, S^i$, where w_i denotes the learnable weight of the i-th stage. We supervise the network using the Charbonnier loss between the final output $\hat{S}$ and the ground-truth image S^*:

$$\mathcal{L}_{\text{rec}} = \sum_{\mathbf{x}} \sqrt{\left(\hat{S}(\mathbf{x}) - S^*(\mathbf{x})\right)^2 + \epsilon^2}. \tag{20}$$

Table 1: Quantitative results (PSNR and SSIM) on the GoPro dataset.

Method	SRN [43] ICCV'18	HINet [2] CVPR'21	MSDINet [26] ECCV'22	Restormer [63] CVPR'22	FFTFormer [20] CVPR'23	MLWNet [13] CVPR'24	Concertormer [23] ICCV'25	EVSSM [21] CVPR'25
PSNR	30.26	32.71	33.28	32.92	34.21	33.83	34.42	34.51
SSIM	0.934	0.959	0.964	0.961	0.969	0.968	0.971	0.971
Method	EFNet [40] ECCV'22	MAENet [42] ECCV'24	REFID [39] TPAMI'25	CMTA-7 [19] ECCV'24	EGDeblurring [52] CVPR'25	MAT [56] AAAI'25	Zhu *et al* [70] ICCV'25	**Ours** 2026
PSNR	35.46	36.07	35.91	<u>36.78</u>	35.50	36.67	36.70	**37.15**
SSIM	0.972	0.976	0.973	<u>0.977</u>	0.974	**0.978**	<u>0.977</u>	**0.978**

By aggregating and supervising the intermediate results, this loss design effectively constrains intermediate reconstructions and stabilizes the training process.

4 Experiments and Analysis

4.1 Experimental Settings

Datasets. We evaluate our method on three public benchmarks that include synthetic, semi-synthetic, and real-world event-based deblurring data, following the common experimental protocol in recent event-guided deblurring works.
GoPro dataset. We use the GoPro dataset [30] as the main training benchmark. It contains blurry images generated by averaging consecutive sharp frames, together with corresponding ground-truth sharp images. The dataset includes 3,214 blurry–sharp image pairs, among which 2,103 pairs are used for training and 1,111 pairs for testing. For event data, we adopt the synthetic event stream released by [40], where events are generated using the ESIM simulator [35] with randomly sampled contrast thresholds. For each blurry image, the raw events within the exposure time are converted into multiple event frames using the same preprocessing strategy as the original dataset.
HS-ERGB dataset. HS-ERGB provides real event data captured together with high-frame-rate videos [44]. Following [42], we use the normal-blur version [69], where blurry images are synthesized by averaging interpolated sharp frames. Since the exposure time of each frame is known, event frames are generated by accumulating events within the corresponding exposure interval.
REBlur dataset. REBlur consists of real-world event streams paired with blurred images and high-quality sharp ground truth [40]. The dataset covers various motion types, including both linear and nonlinear motions. We follow the official split and use 486 image pairs for training and 983 pairs for testing.
Implementation Details. The proposed method is implemented in PyTorch and trained on four NVIDIA RTX 5090 GPUs. During training, we randomly crop patches of size 256×256 from the blurry images and their corresponding event representations as inputs. The batch size is set to 8. We adopt the AdamW optimizer [71] with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The initial learning rate is set to 2×10^{-4} and is scheduled using cosine annealing. Random horizontal and vertical

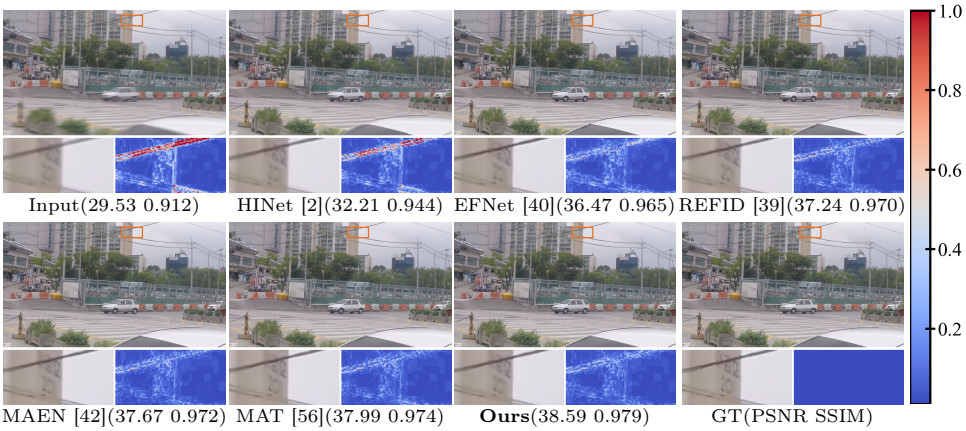


Fig. 3: Visual comparison of different methods on the GoPro dataset [30]. For better clarity, a local region is enlarged, along with its corresponding error map.

Table 2: Quantitative results (PSNR and SSIM) on the HS-ERGB dataset [44].

Method	HINet [2]	Restormer [63]	FFTFormer [20]	EFNet [40]	EIFNet [60]	MAENet [42]	MAT [56]	**Ours**
PSNR	27.32	27.55	28.11	26.68	26.74	27.93	<u>28.97</u>	**29.43**
SSIM	0.807	0.808	0.813	0.800	0.797	0.812	<u>0.816</u>	**0.818**

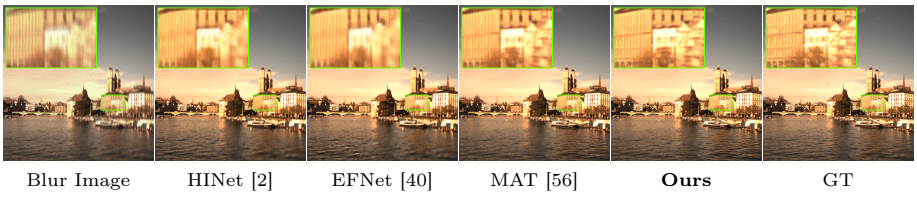


Fig. 4: Visual comparison of different methods on the HS-ERGB dataset [44]. Please enlarge the PDF for more details.

flipping are applied for data augmentation during training. Following [45], the event stream is converted into event frames by discretizing events into fixed temporal bins. For simplicity, the number of bins is set to be consistent with M (Eq. 3), where $M = 8$ by default. The predicted kernel size K is fixed to 5×5, and the unrolled network consists of $N = 6$ iterative stages. The influence of different values of M, N, and K is further investigated in the ablation study. For the HS-ERGB [44] and REBlur [40] datasets, existing methods follow different training strategies, where some are trained from scratch while others are fine-tuned from models pretrained on the GoPro dataset. We observe that even for the same method, using different pretrained models can lead to noticeably different performance. Therefore, to ensure a fair comparison, all competing methods are retrained from scratch on these two datasets in our experiments.

Table 3: Quantitative results (PSNR and SSIM) on the REBlur dataset [40].

Method	HINet [2]	Restormer [63]	EFNet [40]	EIFNet [60]	EGDeblurring [52]	MAENet [42]	MAT [56]	**Ours**
PSNR	35.03	35.22	36.33	36.56	36.52	36.78	<u>36.97</u>	**37.24**
SSIM	0.960	0.961	0.964	0.968	0.965	0.967	<u>0.969</u>	**0.971**

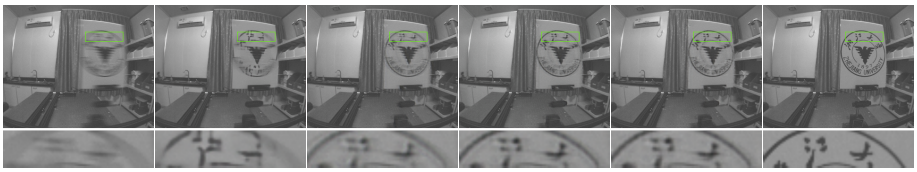


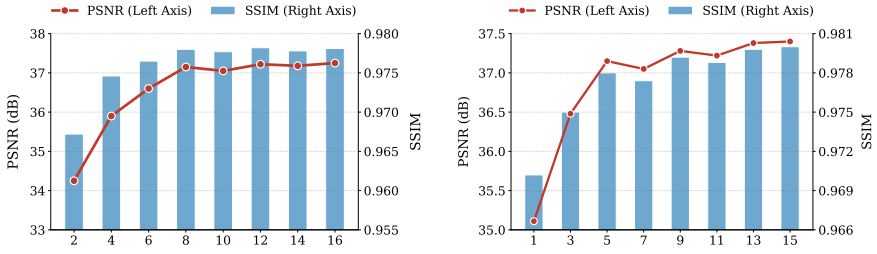
Fig. 5: Visual comparison of different methods on the REBlur dataset [40]. Please enlarge the PDF for more details.

4.2 Comparison with State-of-the-Art Methods

We compare our method with a wide range of state-of-the-art image deblurring approaches, including frame-based methods such as HINet, Restormer, and Concertormer [2,23,63], as well as event-guided methods including EFNet, MAENet, REFID, EGDeblurring, and MAT [39, 40, 42, 52, 56].

Quantitative comparison. Tables 1, 2, and 3 present the quantitative results on the GoPro [30], HS-ERGB [44], and REBlur [40] datasets, respectively. As shown in these tables, our method consistently achieves the best or highly competitive PSNR and SSIM scores across all three benchmarks. On the GoPro dataset, our method substantially outperforms all frame-based approaches and shows clear improvements over existing event-guided methods, highlighting the effectiveness of explicitly incorporating event-derived blur modeling into the deblurring process. Similar performance trends are observed on the HS-ERGB [44] and REBlur [40] datasets, which involve real event data and more complex motion patterns. These results demonstrate that our method generalizes well across synthetic, semi-synthetic, and real-world scenarios.

Qualitative comparison. In addition to quantitative evaluation, we present visual comparisons on the GoPro, HS-ERGB, and REBlur datasets in Figs. 3, 4, and 5, respectively. Compared with frame-based methods, which often leave residual blur or introduce artifacts in regions with fast motion, our method produces sharper structures and more faithful textures. Event-based baselines improve motion handling to some extent, but still suffer from incomplete blur removal or inconsistent details. By contrast, our method effectively suppresses motion blur while preserving fine details, as evidenced by the clearer edges and lower error responses in the highlighted regions.



(a) Number of temporal bins M. **(b)** Kernel size $K \times K$.

Fig. 6: Ablation studies on the number of temporal bins M (a), and kernel size K (b).

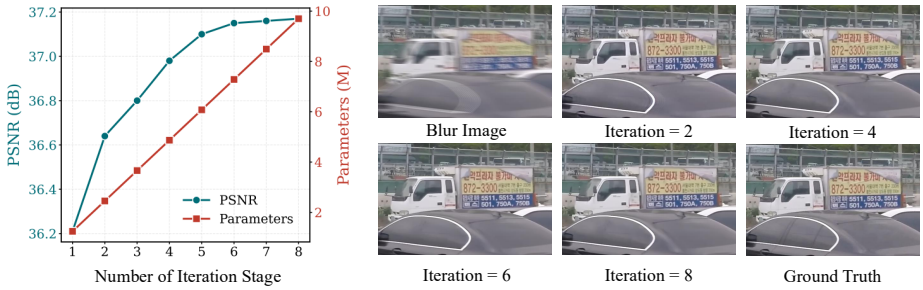


Fig. 7: Effect of the number of unrolled stages N. Left: PSNR and parameter growth with respect to iteration stages. Right: Visual comparisons at different iteration stages.

4.3 Ablation Studies

We conduct ablation studies on the GoPro dataset [30] to analyze the impact of key hyper-parameters and major components. Unless otherwise specified, we use the default setting $M = 8$, $K = 5 \times 5$, and $N = 6$.

Effect of temporal bins M. Fig. 6a analyzes the effect of the number of temporal bins M used to discretize the exposure interval. When M is small, motion information is sparsely sampled, resulting in inaccurate blur kernel estimation and degraded deblurring performance. As M increases, both PSNR and SSIM improve consistently, indicating more accurate motion integration. However, the performance gain gradually saturates for larger values of M, while the computational cost continues to increase. Therefore, we set $M = 8$ to achieve a balance between reconstruction accuracy and computational efficiency.

Effect of kernel size K. Fig. 6b shows the effect of the kernel size on deblurring performance. Small kernels (*e.g.*, 1×1 or 3×3) cannot sufficiently model large motion blur, resulting in inferior performance. Increasing the kernel size leads to a clear improvement up to 5×5. Beyond this point, the performance continues to improve slightly but with diminishing returns. Considering both reconstruction quality and efficiency, we adopt $K = 5 \times 5$ as the default kernel size.

Table 4: Ablation study. `Model1` replaces the spatially varying blur operator with a standard convolution, and `Model2` removes the event consistency term.

Method	Event-derived blur operator	Event consistency	PSNR	SSIM
`Model1`	×	✓	36.25	0.976
`Model2`	✓	×	36.60	0.977
Ours	✓	✓	**37.15**	**0.978**

Effect of the number of stages N. We further analyze the influence of the number of unrolled stages N in the proposed network. As shown in Fig. 7 (left), increasing N consistently improves PSNR, indicating that additional stages enable progressive refinement of the reconstruction. Meanwhile, the number of model parameters grows linearly with N, reflecting the increased model capacity. Notably, the performance gain gradually saturates after a certain number of stages. The qualitative results in Fig. 7 (right) further illustrate this behavior. With more iterations, motion blur is progressively removed and fine details become clearer, while later stages mainly provide subtle refinements. Considering the trade-off between performance and model complexity, we set $N = 6$ as the default number of stages in our experiments.

Replacing the spatially varying blur operator with convolution. In many existing deep unfolding network (DUN)-based deblurring methods, the blur operator is unknown and is therefore commonly approximated by a learnable convolution layer. While this strategy simplifies the optimization and avoids explicit kernel modeling, it ignores the spatially varying nature of real-world motion blur and lacks physical interpretability. To examine the effect of this approximation, we follow this common practice and replace the spatially varying blur operator $\mathcal{K}(K)S$ in the blur consistency term with a standard convolution operation. Specifically, the blur consistency term in Eq. 9 is reformulated as

$$\mathcal{L}_{blur}^{\text{conv}}(S) = \|B - (k * S)\|_F^2, \tag{21}$$

where k denotes a learnable convolution kernel and $*$ is the standard convolution operator. As shown in Table 4, this variant leads to a clear performance degradation compared to the full model. The results indicate that replacing the event-derived blur operator with a generic convolution is insufficient to capture complex, spatially varying motion blur, thereby highlighting the benefit of explicitly embedding event-based blur modeling into the unfolding framework.

Removing event consistency. We further study the role of event consistency by removing the corresponding term in Eq. 9, *i.e.*, setting $\mu = 0$. Without this constraint, the optimization relies solely on the blur consistency term and learned image priors. As shown in Table 4, this simplification leads to noticeable performance degradation. These results indicate that the event consistency term provides complementary supervision that cannot be fully compensated by blur consistency alone.

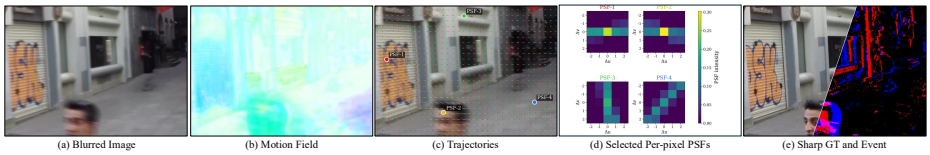


Fig. 8: Motion and point spread function (PSF) visualization. We show the estimated motion field, event-derived trajectories, and selected per-pixel PSFs, where different PSF orientations and spatial supports reflect spatially varying blur patterns.

Motion/PSF visualization. As shown in Fig. 8, we further visualize the estimated motion field, representative trajectories, and selected per-pixel PSFs to examine whether the learned kernels capture meaningful blur patterns. The trajectories vary across image regions, and the corresponding PSFs exhibit different orientations and spatial supports, which is consistent with the spatially varying motion in the blurred input.

5 Conclusion

In this paper, we presented a physically grounded event-driven framework for non-uniform motion deblurring. By leveraging the high temporal resolution of event streams, we explicitly estimate dense pixel-wise motion trajectories and construct spatially varying blur kernels that accurately model the blur formation process. Based on the resulting blur operator, we formulate a joint optimization problem and solve it using an ADMM-inspired unrolled network. The resulting deep unfolding architecture effectively combines the interpretability of model-based optimization with the representation power of deep learning, enabling stable training and efficient inference. Extensive experiments on both synthetic and real-world benchmarks demonstrate that the proposed method consistently outperforms existing frame-based and event-guided approaches while maintaining favorable computational efficiency. In future work, we plan to extend the framework to more challenging real-world scenarios, such as dynamic scenes with complex occlusions and illumination variations. Moreover, incorporating more expressive event representations and exploring lightweight architectures for real-time deployment are promising directions for further research.

Acknowledgements

This work is partially supported by the Research Grant Council (RGC) of Hong Kong General Research Fund (GRF) under Grant 11200323, the NSFC/RGC JRS Project N_CityU198/24, and RGC Research Fellow Scheme under Grant RFS2526-1S04.

References

1. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European conference on computer vision. pp. 17–33. Springer (2022)
2. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 182–192 (2021)
3. Chen, N., Li, B., Wang, Y., Ying, X., Wang, L., Zhang, C., Guo, Y., Li, M., An, W.: Motion and appearance decoupling representation for event cameras. IEEE Transactions on Image Processing **34**, 5964–5977 (2025)
4. Chen, N., Zhang, C., An, W., Wang, L., Li, M., Ling, Q.: Event-based motion deblurring with blur-aware reconstruction filter. IEEE Transactions on Circuits and Systems for Video Technology **35**(9), 8508–8519 (2025)
5. Cho, H., Jeong, Y., Kim, T., Yoon, K.J.: Non-coaxial event-guided motion deblurring with spatial alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12492–12503 (2023)
6. Cho, S., Wang, J., Lee, S.: Video deblurring for hand-held cameras using patch-based synthesis. ACM Transactions on Graphics (TOG) **31**(4), 1–9 (2012)
7. Dong, X., Yokoya, N.: Understanding dark scenes by contrasting multi-modal observations. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 840–850 (January 2024)
8. Dong, X., Yokoya, N., Wang, L., Uezato, T.: Learning mutual modulation for self-supervised cross-modal super-resolution. In: European Conference on Computer Vision. pp. 1–18. Springer (2022)
9. Duan, P., Li, B., Yang, Y., Lou, H., Teng, M., Zhou, X., Ma, Y., Shi, B.: Eventaid: Benchmarking event-aided image/video enhancement algorithms with real-captured hybrid dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(8), 6959–6973 (2025)
10. Feijoo, D., Garrido-Mellado, P., Conde, M.V., Rim, J., Garcia, A., Cho, S., Timofte, R.: Efficient real-world deblurring using single images: Aim 2025 challenge report. In: 2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). pp. 5772–5780 (2025)
11. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. ACM Trans. Graph. **25**(3), 787–794 (Jul 2006)
12. Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A.J., Conradt, J., Daniilidis, K., et al.: Event-based vision: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(1), 154–180 (2020)
13. Gao, X., Qiu, T., Zhang, X., Bai, H., Liu, K., Huang, X., Wei, H., Zhang, G., Liu, H.: Efficient multi-scale network with learnable discrete wavelet transform for blind motion deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2733–2742 (2024)
14. Ge, X., Liu, J., Hu, D., Tan, J.: An extended sparse model for blind image deblurring. Signal, Image and Video Processing **18**(2), 1863–1877 (2024)
15. Guo, Q., Feng, W., Gao, R., Liu, Y., Wang, S.: Exploring the effects of blur and deblurring to visual object tracking. IEEE Transactions on Image Processing **30**, 1812–1824 (2021)
16. Gupta, A., Joshi, N., Lawrence Zitnick, C., Cohen, M., Curless, B.: Single image deblurring using motion density functions. In: European conference on computer vision. pp. 171–184. Springer (2010)

17. Ji, S.W., Lee, J., Kim, S.W., Hong, J.P., Baek, S.J., Jung, S.W., Ko, S.J.: Xydeblur: Divide and conquer for single image deblurring. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17421–17430 (2022)
18. Jia, J.: Single image motion deblurring using transparency. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2007)
19. Kim, T., Cho, H., Yoon, K.J.: Cmta: Cross-modal temporal alignment for event-guided video deblurring. In: European Conference on Computer Vision. pp. 1–19. Springer (2024)
20. Kong, L., Dong, J., Ge, J., Li, M., Pan, J.: Efficient frequency domain-based transformers for high-quality image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5886–5895 (2023)
21. Kong, L., Dong, J., Tang, J., Yang, M.H., Pan, J.: Efficient visual state space model for image deblurring. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 12710–12719 (2025)
22. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: CVPR 2011. pp. 233–240. IEEE (2011)
23. Kuo, P.H., Pan, J., Chien, S.Y., Yang, M.H.: Efficient concertormer for image deblurring and beyond. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 14665–14675 (October 2025)
24. Laroche, C., Almansa, A., Tassano, M.: Deep model-based super-resolution with non-uniform blur. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1797–1808 (2023)
25. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding and evaluating blind deconvolution algorithms. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 1964–1971. IEEE (2009)
26. Li, D., Zhang, Y., Cheung, K.C., Wang, X., Qin, H., Li, H.: Learning degradation representations for image deblurring. In: European conference on computer vision. pp. 736–753. Springer (2022)
27. Liang, J., Fan, Y., Xiang, X., Ranjan, R., Ilg, E., Green, S., Cao, J., Zhang, K., Timofte, R., Gool, L.V.: Recurrent video restoration transformer with guided deformable attention. Advances in Neural Information Processing Systems **35**, 378–393 (2022)
28. Lu, X., Sun, Z., Ge, C., Peng, Y., Zhou, Z., Li, Z., Liao, Z., Li, D., Kang, Q., Fu, X., Zha, Z.J.: Efficient high fps non-uniform motion deblurring via progressive learning. In: 2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). pp. 5644–5653 (2025)
29. Munda, G., Reinbacher, C., Pock, T.: Real-time intensity-image reconstruction for event cameras using manifold regularisation. International Journal of Computer Vision **126**(12), 1381–1393 (2018)
30. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3883–3891 (2017)
31. Nakabayashi, T., Hasegawa, K., Matsugu, M., Saito, H.: Event-based blur kernel estimation for blind motion deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4120–4128 (2023)
32. Noki, M.M.H., Mahmud, S.M., Majumder, P.S., Radi, A.M.A., Ali, M.H., Khan, M.M.: Deblurring in the wild: A real-world dataset from smartphone high-speed videos. arXiv preprint arXiv:2506.19445 (2025)
33. Pan, L., Hartley, R., Scheerlinck, C., Liu, M., Yu, X., Dai, Y.: High frame rate video reconstruction based on an event camera. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(5), 2519–2533 (2020)

34. Rao, C., Li, G., Lan, Z., Sun, J., Luan, J., Xing, W., Zhao, L., Lin, H., Dong, J., Zhang, D.: Rethinking video deblurring with wavelet-aware dynamic transformer and diffusion model. In: European Conference on Computer Vision. pp. 421–437. Springer (2024)
35. Rebecq, H., Gehrig, D., Scaramuzza, D.: Esim: an open event camera simulator. In: Conference on robot learning. pp. 969–982. PMLR (2018)
36. Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D.: Events-to-video: Bringing modern computer vision to event cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3857–3866 (2019)
37. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena **60**(1-4), 259–268 (1992)
38. Sayed, M., Brostow, G.: Improved handling of motion blur in online object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1706–1716 (2021)
39. Sun, L., Gehrig, D., Sakaridis, C., Gehrig, M., Liang, J., Sun, P., Xu, Z., Wang, K., Van Gool, L., Scaramuzza, D.: A unified framework for event-based frame interpolation with ad-hoc deblurring in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(4), 2265–2279 (2025)
40. Sun, L., Sakaridis, C., Liang, J., Jiang, Q., Yang, K., Sun, P., Ye, Y., Wang, K., Gool, L.V.: Event-based fusion for motion deblurring with cross-modal attention. In: European conference on computer vision. pp. 412–428. Springer (2022)
41. Sun, L., Cho, S., Wang, J., Hays, J.: Good image priors for non-blind deconvolution: generic vs. specific. In: European conference on computer vision. pp. 231–246. Springer (2014)
42. Sun, Z., Fu, X., Huang, L., Liu, A., Zha, Z.J.: Motion aware event representation-driven image deblurring. In: Proceedings of the European conference on computer vision (ECCV). pp. 418–435. Springer (2024)
43. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8174–8182 (2018)
44. Tulyakov, S., Gehrig, D., Georgoulis, S., Erbach, J., Gehrig, M., Li, Y., Scaramuzza, D.: Time lens: Event-based video frame interpolation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16155–16164 (2021)
45. Wang, B., He, J., Yu, L., Xia, G.S., Yang, W.: Event enhanced high-quality image recovery. In: European Conference on Computer Vision. pp. 155–171. Springer (2020)
46. Whang, J., Delbracio, M., Talebi, H., Saharia, C., Dimakis, A.G., Milanfar, P.: Deblurring via stochastic refinement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16293–16303 (2022)
47. Whyte, O., Sivic, J., Zisserman, A., Ponce, J.: Non-uniform deblurring for shaken images. International journal of computer vision **98**(2), 168–186 (2012)
48. Wu, J.H., Tsai, F.J., Peng, Y.T., Tsai, C.C., Lin, C.W., Lin, Y.Y.: Id-blau: Image deblurring by implicit diffusion-based reblurring augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25847–25856 (2024)
49. Xiao, Z., Li, Z., Zhao, Y., Liu, Y., Zhang, Z., Jia, W.: Learning dual modality interactions for event-based motion deblurring. IEEE Transactions on Multimedia **28**, 3652–3666 (2026)

50. Xiao, Z., Wang, X.: Asymmetric dual-lens video deblurring. In: Belgrave, D., Zhang, C., Lin, H., Pascanu, R., Koniusz, P., Ghassemi, M., Chen, N. (eds.) Advances in Neural Information Processing Systems. vol. 38, pp. 42422–42437. Curran Associates, Inc. (2025)
51. Xiao, Z., Wang, X.: Event-based video super-resolution via state space models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12564–12574 (June 2025)
52. Xie, X., Zhang, Q., Zheng, W.S.: Diffusion-based event generation for high-quality image deblurring. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 2194–2203 (2025)
53. Xu, F., Yu, L., Wang, B., Yang, W., Xia, G.S., Jia, X., Qiao, Z., Liu, J.: Motion deblurring with real events. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2583–2592 (2021)
54. Xu, L., Jia, J.: Two-phase kernel estimation for robust motion deblurring. In: European conference on computer vision. pp. 157–170. Springer (2010)
55. Xu, L., Jia, J.: Depth-aware motion deblurring. In: 2012 IEEE International Conference on Computational Photography (ICCP). pp. 1–8. IEEE (2012)
56. Xu, S., Sun, Z., Zhong, M., Cao, C., Liu, Y., Fu, X., Chen, Y.: Motion-adaptive transformer for event-based image deblurring. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 8942–8950 (2025)
57. Xue, S., Ma, H., Chen, H., Yang, Z., Deng, Y.: Event-based motion deblurring using task-oriented 3d gaussian event representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 29547–29556 (June 2026)
58. Yan, Y., Ren, W., Guo, Y., Wang, R., Cao, X.: Image deblurring via extreme channels prior. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6978–6986 (2017)
59. Yan, Z., Jiao, J., Wang, Z., Lee, G.H.: Event-driven dynamic scene depth completion. In: Belgrave, D., Zhang, C., Lin, H., Pascanu, R., Koniusz, P., Ghassemi, M., Chen, N. (eds.) Advances in Neural Information Processing Systems. vol. 38, pp. 142971–142993. Curran Associates, Inc. (2025)
60. Yang, W., Wu, J., Li, L., Dong, W., Shi, G.: Event-based motion deblurring with modality-aware decomposition and recomposition. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 8327–8335 (2023)
61. Yang, W., Wu, J., Li, L., Dong, W., Shi, G.: Asymmetric hierarchical difference-aware interaction network for event-guided motion deblurring. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 9265–9273 (2025)
62. Yang, Y., Pan, L., Liu, L., Liu, M.: K3dn: Disparity-aware kernel estimation for dual-pixel defocus deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13263–13272 (2023)
63. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
64. Zhang, C., Zhang, X., Jiang, C., Xia, G.S., Yu, L.: Evdi++: Event-based video deblurring and interpolation via self-supervised learning. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–18 (2026)
65. Zhang, J., Li, J., Li, J., Sun, Y., Liu, X., Zheng, Z., Lu, G.: Mbrvo: A blur robust visual odometry based on motion blurred artifact prior. IEEE Robotics and Automation Letters **9**(10), 8418–8425 (2024)

66. Zhang, J., Pan, J., Ren, J., Song, Y., Bao, L., Lau, R.W., Yang, M.H.: Dynamic scene deblurring using spatially variant recurrent neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2521–2529 (2018)
67. Zhang, L., Zhang, H., Chen, J., Wang, L.: Hybrid deblur net: Deep non-uniform deblurring with event camera. IEEE Access **8**, 148075–148083 (2020)
68. Zhang, P., Liu, H., Ge, Z., Wang, C., Lam, E.Y.: Neuromorphic imaging with joint image deblurring and event denoising. IEEE Transactions on Image Processing **33**, 2318–2333 (2024)
69. Zhang, X., Yu, L., Yang, W., Liu, J., Xia, G.S.: Generalizing event-based motion deblurring in real-world scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10734–10744 (2023)
70. Zhu, Y., Chen, H., Deng, Y., You, W.: Separation for better integration: Disentangling edge and motion in event-based deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14732–14742 (2025)
71. Zhuang, Z., Liu, M., Cutkosky, A., Orabona, F.: Understanding adamw through proximal methods and scale-freeness. arXiv preprint arXiv:2202.00089 (2022)