Noname manuscript No.
(will be inserted by the editor)

Deblurring Low-Light Images with Events

Chu Zhou · Minggui Teng · Jin Han · Jinxiu Liang · Chao Xu · Gang Cao · Boxin Shi

Received: date / Accepted: date

Abstract Modern image-based deblurring methods usually show degenerate performance in low-light conditions since the images often contain most of the poorly visible dark regions and a few saturated bright regions, making the amount of effective features that can be extracted for deblurring limited. In contrast, event cameras can trigger events with a very high dynamic range and low latency, which hardly suffer from saturation and naturally encode dense temporal information about motion. However, in low-light conditions existing event-based deblurring methods would become less robust since the events triggered in dark regions are often severely contaminated by noise, leading to inaccurate reconstruction of the corresponding intensity values. Besides, since they directly adopt the event-based double integral model to perform pixel-wise reconstruction, they can only handle low-resolution grayscale active pixel sensor images provided by the DAVIS camera, which cannot meet the requirement of daily photography. In this paper, to apply events to deblurring low-light images robustly, we propose a unified two-stage framework along with a motion-aware neural network tailored to it, reconstructing the sharp image under the guidance of high-fidelity motion clues extracted from events. Besides, we build an RGB-DAVIS hybrid camera system to demonstrate that our method has the ability to deblur high-resolution RGB images due to the natural advantages of our two-stage framework. Experimental results show our method achieves state-of-the-art performance on both synthetic and real-world images.

Keywords Low-light · Image deblurring · Event camera · Deep learning · Hybrid camera system

✉ Boxin Shi
E-mail: shiboxin@pku.edu.cn

Chu Zhou · Chao Xu
Key Laboratory of Machine Perception (MOE), School of Intelligence Science and Technology, Peking University, China

Minggui Teng · Jinxiu Liang · Boxin Shi
National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, China

Jin Han
Graduate School of Information Science and Technology, The University of Tokyo, Japan

Gang Cao · Boxin Shi
Beijing Academy of Artificial Intelligence, China

1 Introduction

Images captured in low-light conditions[1] are prone to motion blur caused by camera shakes since the sensor requires a longer exposure time to receive adequate light. Due to the broad existence of light sources in such conditions, bright regions with saturated pixels are often captured. Together with the inevitable noise in dark regions, the amount of effective features that can be extracted for deblurring is often limited, which degenerates the performance of modern image-based methods (Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) for handling low-light images.

Unlike conventional frame-based cameras that can only capture image frames, event cameras (*e.g.*, DAVIS camera (Brandli et al, 2014)) not only capture grayscale active pixel sensor (APS) images, but also detect per-pixel brightness changes in an asynchronous manner by triggering events whenever the logarithmic change of latent irradiance exceeds a preset threshold. These events have a very high dynamic range (HDR) so they hardly suffer from saturation. Besides,

[1] In this paper, we use the term "low-light images" to refer to the images containing a majority of pixels in dark regions (poorly visible with low contrast) and a few bright (usually saturated) pixels (*e.g.*, night scenes containing light sources (Hu et al, 2018b)), instead of the images containing only dark regions (*e.g.*, images with short-exposure (Chen et al, 2018a; Zhang et al, 2020b)).

events are triggered with very low latency, so that they naturally encode dense temporal information about motion, which is particularly useful in motion deblurring applications. Current event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021) aim to deblur the APS images under the guidance of events that have the same spatial resolution as APS images (the output of DAVIS camera). These methods directly model the mapping from events to latent irradiance in a pixel-wise manner based on the event-based double integral (EDI) model (Pan et al, 2019), which demonstrate higher performance and better generalization ability than image-based ones. However, when taking photos in low-light conditions, both the contrast and signal-to-noise ratio (SNR) are much lower in dark regions (Mitrokhin et al, 2018) than in bright regions, so that in dark regions the "good events" (triggered by brightness changes) are less observed while the "bad events" (triggered by noise) become dominated. In such a situation, the pixels in dark regions cannot be reconstructed robustly by these event-based deblurring methods due to their sensitivity to the lack of "good events" as well as the abundance of "bad events", this is because for a certain pixel, the EDI model relies on events triggered at its position to reconstruct the corresponding intensity value. Furthermore, the APS images are grayscale and often have low spatial resolution (typically 346×260 pixels for the DAVIS346 camera), which cannot meet the requirement of daily photography, leading to limited application scenarios. So, it is of great interest to develop a new event-based deblurring method for handling low-light images robustly, with the ability to deblur high-resolution RGB images.

An intuitive strategy could be discarding the events triggered in dark regions (which are often noisy and mainly contain "bad events") and focusing on filtering the events triggered in bright regions (which are often clean and mainly contain "good events") for image deblurring. To apply this strategy, our preliminary work (Zhou et al, 2021a) introduces a two-stage deblurring pipeline instead of performing pixel-wise reconstruction based on the EDI model. It first selects an image patch containing a light streak (blurred light source caused by camera shakes) and utilizes the light streak to filter the clean local events to estimate the spatially-uniform 2D blur kernel, then performs non-blind deconvolution with the estimated blur kernel. Besides, the two-stage pipeline is ready to be applied to deblurring high-resolution RGB images captured by an RGB-DAVIS hybrid camera system, by estimating the blur kernel from the APS image and corresponding events (captured by the DAVIS camera) in the first stage and deconvolving the high-resolution RGB image (captured by the RGB camera) using the estimated blur kernel in the second stage. Despite that Zhou et al (2021a) for the first time demonstrate the ability to deblur high-resolution RGB images with events, they adopt the assumption that the blur is caused by in-plane camera shakes, which cannot deal with spatially-variant blur. To solve this problem, a new approach for extracting global motion clues is required to replace the process that estimates a 2D blur kernel from a local patch reflecting the local motion only.

In this paper, we propose to extend our preliminary work (Zhou et al, 2021a) to a *unified* two-stage framework to apply events to deblurring low-light images with spatially-variant blur. The key observation is that strong edges in low-light images (they could be but not necessarily be caused by light streaks) could generally trigger events with higher SNR and they encode spatial information about motion (Joshi et al, 2008; Cho and Lee, 2009; Fu et al, 2022). We therefore design the first stage to extract global motion clues by utilizing the edge map of the APS image to filter the clean events; then, in the second stage the estimated global motion clues are used for image deblurring. Tailored to such a framework, we further propose a motion-aware neural network to perform the deblurring process: First, it extracts features from the edge map of the APS image and corresponding events jointly to obtain both spatial and temporal information about motion, and encodes the information into high-fidelity motion clues by explicitly using bi-directional optical flows to supervise this process in the latent space; then, it adopts a denoising module to perform blind noise suppression in the image domain to avoid ringing artifacts, and reconstructs the sharp image under the guidance of motion clues. Besides, due to the natural advantages of our two-stage framework that does not directly adopt the EDI model to perform pixel-wise reconstruction, by building an RGB-DAVIS hybrid camera system, the proposed method demonstrates the ability to deblur high-resolution RGB images with events, without the spatially-uniform blur assumption. To summarize, this paper makes contributions by demonstrating:

- a unified two-stage framework to apply events to deblurring low-light images;
- a motion-aware neural network tailored to such a framework to perform the deblurring process;
- ability to deblur high-resolution RGB images with events, without the spatially-uniform blur assumption.

Compared with our preliminary work (Zhou et al, 2021a), the main improvement is replacing its deblurring pipeline (based on blur kernel estimation) with a unified framework (based on motion clue extraction) and redesigning the network architecture along with the loss functions to adapt spatially-variant blur. Besides, we improve the synthetic dataset generation pipeline by adopting the model proposed by Whyte et al (2012) and using the RealBlur dataset (Rim et al, 2020) as the data source to generate images with spatially-variant blur for evaluation. We also recapture the real data containing spatially-variant blur to show that our new framework has a good generalization ability on real low-light images and events.

2 Related works

2.1 Image deblurring methods

Generally, image deblurring methods could be divided into two categories: image-based methods, which aims to directly reconstruct the sharp image from a single blurry image, and event-based methods, which use events to guide the image deblurring process. We will have an overview on them respectively in the following.

Image-based deblurring. Image-based deblurring is a highly ill-posed problem due to the complexity of natural image structures and the diversity of blur patterns. Some works treated this problem as a maximum *a posteriori* (MAP) estimation problem and proposed several handcrafted image priors (*e.g.*, total variation regularization (Chan and Wong, 1998), heavy-tailed gradient distributions (Fergus et al, 2006), local smoothness prior (Shan et al, 2008), normalized sparsity prior (Krishnan et al, 2011), and ℓ_0-regularized prior (Xu et al, 2013; Pan et al, 2016a)) to relieve its ill-posedness. To reduce the ill-posedness and improve the robustness, several methods tried to exploit the latent priors lying in the image itself, such as strong edges (Joshi et al, 2008; Cho and Lee, 2009; Fu et al, 2022), patch recurrences (Michaeli and Irani, 2014), blurry image outliers (Dong et al, 2017), channel statistics (Pan et al, 2016b; Yan et al, 2017), noise pattern (Zhong et al, 2013), and light streaks (Hu et al, 2018b; Chen et al, 2021a). Although these prior-based methods are successful in restoring plausible sharp contents in a large variety of scenes, their applicability is still limited since they are based on time-consuming numerical optimization, which cannot meet the requirement of real-time deblurring. Recently, deep neural networks have been adopted to handle this problem. They usually run much faster than prior-based methods. These learning-based methods could be divided into two categories: direct methods and indirect methods. Direct methods, which try to deblur in an end-to-end manner using convolutional neural networks (CNN) (Nah et al, 2017; Zhang et al, 2018a; Tao et al, 2018; Gao et al, 2019; Zhang et al, 2019; Suin et al, 2020; Cho et al, 2021; Chi et al, 2021) or generative adversarial networks (GAN) (Kupyn et al, 2018, 2019; Zhang et al, 2020a), often demonstrate visually impressive results and are easy to deploy. Indirect methods, which use blur-related physical quantities (*e.g.*, blur kernels (Kaufman and Fattal, 2020; Ren et al, 2020; Dong et al, 2021; Tran et al, 2021; Chen et al, 2021b) or their Fourier coefficients (Chakrabarti, 2016), patch-wise motion vectors (Sun et al, 2015), and dense motion flows (Chen et al, 2018b; Gong et al, 2017; Yuan et al, 2020)) to implicitly or explicitly supervise the extraction of image features for guiding the deblurring process, usually show better generalization ability and suffer less from overfitting.

Event-based deblurring. Event cameras (Lichtsteiner et al, 2008; Brandli et al, 2014) are neuromorphic sensors that can asynchronously detect per-pixel brightness changes and trigger events whenever the logarithmic change of latent irradiance exceeds a preset threshold. They have many attractive properties that frame-based cameras do not possess: high temporal resolution, very high dynamic range, low power consumption, and high pixel bandwidth, which could naturally benefit the image deblurring task. Recent event cameras (*e.g.*, DAVIS camera (Brandli et al, 2014)) are able to capture grayscale APS images along with events, since they contain a global shutter APS in addition to the dynamic vision sensor that shares the same photosensor array. This unique advantage makes it possible to use events to guide the deblurring process of APS images. Pan et al (2019) proposed the EDI model that clarifies the relationship among the blurry image, events, and latent irradiance. Based on the EDI model, several new methods have been proposed to solve the event-based deblurring problem. Jiang et al (2020) used a convolutional recurrent neural network that integrates visual and temporal knowledge of both global and local scales to recover image details. Lin et al (2020) proposed an end-to-end trainable neural network to generate high-speed videos and used dynamic filtering to handle the events triggered by the spatially-varying threshold. Chen et al (2020) proposed a residual model suitable for learning image deblurring and high frame rate video generation with events. Pan et al (2020) proposed a method based on numerical optimization to estimate the optical flow from a single image and corresponding events for deblurring. Xu et al (2021) exploited the blurry consistency and photometric consistency to enable self-supervision on the event-based deblurring network with real-world data. Shang et al (2021) developed a principled framework and a flexible event fusion module for tackling video deblurring with the help of events.

2.2 Low-light processing methods

Since our method is designed for low-light conditions, low-light processing methods are also closely related with it. Low-light processing methods focus mainly on handling the noise in events or images, and we will briefly overview the methods about event denoising and image denoising respectively in the following.

Event denoising. Existing event denoising methods mainly focus on background activity noise produced by temporal noise and junction leakage currents. Liu et al (2015) designed a correlation filter chip for removing background uncorrelated noise events. Barrios-Avilés et al (2018) proposed a bioinspired filtering algorithm which reduces the generated data from event-based sensors without loss of relevant information and performs denoising. Khodamoradi and Kastner

(2018) introduced a hardware friendly spatiotemporal correlation filter with $O(N)$ memory complexity for reducing noise in neuromorphic vision sensors. Baldwin et al (2020) presented a novel method named "event probability mask" for labeling event data and proposed a CNN for event denoising. Based on the assumption that events are triggered by edge motion and therefore shall follow the same spatiotemporal motion projection within a local window if valid (Gallego et al, 2018; Stoffregen et al, 2019), Wang et al (2019) proposed to filter events by their motion association likelihood. By further making use of the motion compensation between the image and event signals, Wang et al (2020b) proposed to use guided image filtering techniques to obtain events with low noise. Recently, Duan et al (2021) proposed a deep neural framework to achieve noise-free event restoration.

Image denoising. Maharjan et al (2019) proposed a residual learning based deep neural network for end-to-end low-light image denoising to improve the image quality with low computational cost. Gu et al (2019) designed a top-down self-guided network to better exploit image multi-scale information, achieving excellent denoising performance for low-light images. Wei et al (2020) formulated a noise model to synthesize realistic noisy images that can match the quality of real data under extreme low-light conditions. Moseley et al (2021) extended existing learning-based low-light image denoising approaches by combining a physical noise model with real noise samples and scene selection based on 3D ray tracing to generate training data; and by conditioning their model on the camera's environmental metadata at the time of image capture. Besides, existing low-light image enhancement methods also involve low-light noise reduction in addition to visibility enhancement, which could be found in the surveys of Li et al (2021); Liu et al (2021).

3 Framework

In this section, we will first introduce the background and motivation of our unified framework in Section 3.1, then detail its design in Section 3.2.

3.1 Background and motivation

For a typical event camera (*e.g.*, the DAVIS346 camera used in this paper), each event can be described as $(\mathbf{u}, t, \sigma)$, meaning that the event is triggered at pixel position $\mathbf{u} = (u_x, u_y)^\top$ and time t when the logarithmic change of latent irradiance $\mathbf{I}(t) = I(\mathbf{u}, t)$ exceeds a preset spatially-varying threshold $\mathbf{c} = c(\mathbf{u})$. Here σ is the polarity given by:

$$\sigma = \begin{cases} 1 & \text{if } \log(I(\mathbf{u},t)) - \log(I(\mathbf{u},t-\delta t)) \geq c(\mathbf{u}) \\ -1 & \text{if } \log(I(\mathbf{u},t)) - \log(I(\mathbf{u},t-\delta t)) \leq -c(\mathbf{u}) \end{cases}, \tag{1}$$

where δt denotes the time interval since the last event occurred at the same pixel position. In addition to events, a grayscale APS image that has the same spatial resolution as events could also be recorded. Since camera shakes would occur during the exposure period, the captured image tends to be blurry. Denoting the blurry image as $\mathbf{B} \triangleq B(\mathbf{u})$, its formation could be described as the integral of latent irradiance $\mathbf{I}(t)$ during the exposure period $[0, T]$:

$$\mathbf{B} = \frac{1}{T}\int_0^T \mathbf{I}(t)\mathrm{d}t + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{\epsilon} = \epsilon(\mathbf{u})$ represents the image domain noise generated with the in-camera imaging pipeline.

Combining Equation (1) and Equation (2) and denoting $\mathbf{I}(0)$ (the latent irradiance at time 0) as the sharp image $\mathbf{S}$, the relationship among the blurry image, events, and sharp image can be described using the EDI model (Pan et al, 2019):

$$\mathbf{B} = \frac{\mathbf{S}}{T}\int_0^T \exp(\mathbf{c}\cdot\mathbf{E}(t))\mathrm{d}t + \boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{E}(t) = E(\mathbf{u}, t)$ stands for the sum of events triggered by brightness changes instead of noise between time 0 and t. However, in low-light conditions, images usually contain a majority of pixels in dark regions where events are mainly triggered by noise, and the performance of current event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021) could be degenerated in those regions since they directly adopt the EDI model to perform pixel-wise reconstruction in an end-to-end manner. To appropriately apply events to deblurring low-light images, we could divide the end-to-end pixel-wise reconstruction process into two stages: extracting motion clues from the blurry image and corresponding events, and deblurring the blurry image using the motion clues instead of events. Therefore, high-fidelity motion clues could be extracted in the first stage by discarding the events triggered in dark regions and focusing on filtering the events triggered in bright regions, and the sharp image could be restored robustly in the second stage under the guidance of motion clues. In such a setting, deblurring high-resolution RGB images also becomes possible since the second stage does not require events so that it is not limited to processing the APS image which has the same spatial resolution as events.

3.2 Framework design

Our goal is to reconstruct the sharp image $\mathbf{S}$ from the blurry image $\mathbf{B}$ and corresponding events $\mathbf{e} = \{e_i\}_{i=0}^{T}$ (all events triggered during the exposure period $[0, T]$) in low-light conditions in a two-stage manner. As shown in Figure 1 (left), our

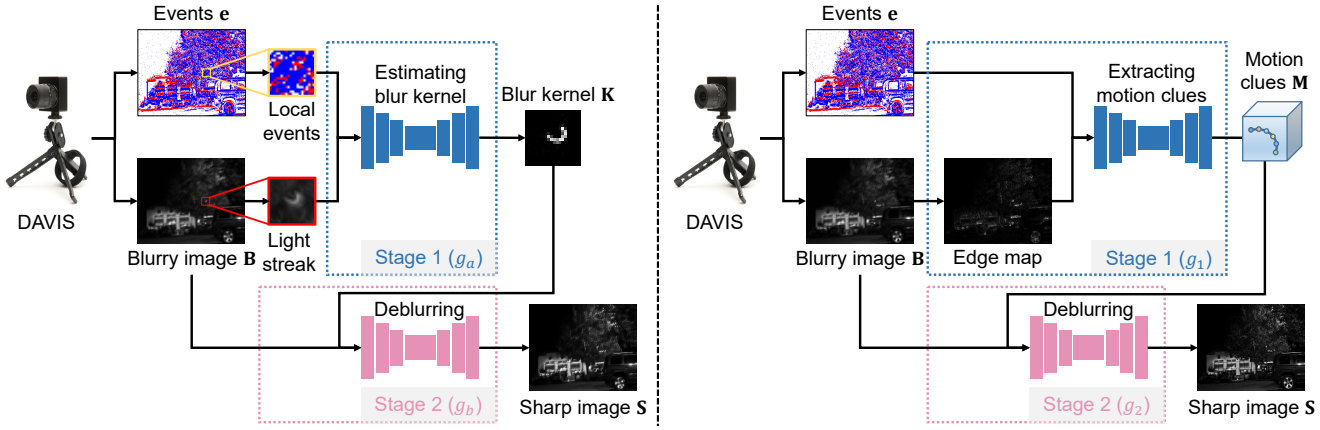

Fig. 1 Left: The deblurring pipeline adopted by our preliminary work (Zhou et al, 2021a), which first selects an image patch containing a light streak and utilizes the light streak to filter the clean local events to estimate the spatially-uniform 2D blur kernel, then performs non-blind deconvolution with the estimated blur kernel. Right: The proposed unified two-stage deblurring framework, which first utilizes the edge map of the APS image to filter the clean events in the whole image to extract motion clues, then use the extracted motion clues to guide the deblurring process. We use color pair (red, blue) to represent the event polarity $(1, -1)$ throughout this paper.

preliminary work (Zhou et al, 2021a) first selects an image patch containing a light streak and utilizes the light streak to filter the clean local events to estimate the spatially-uniform 2D blur kernel, then performs non-blind deconvolution with the estimated blur kernel. Denoting the two stages as g_a and g_b respectively, the whole deblurring pipeline adopted by Zhou et al (2021a) could be described as

$$\mathbf{K} = g_a(p(\mathbf{B}), p(\mathbf{e})) \quad \text{and} \quad \mathbf{S} = g_b(\mathbf{B}, \mathbf{K}), \tag{4}$$

where $\mathbf{K}$ is the estimated blur kernel and p denotes the operation on an extracted patch containing a light streak. However, the first stage requires to detect a light streak manually or using the search-based algorithm proposed by Hu et al (2018b), which is inconvenient. Besides, since the estimated 2D blur kernel can only reflect local motion, the second stage is only able to handle spatially-uniform blur.

To apply events to deblurring low-light images with spatially-variant blur, we propose to extend our preliminary work (Zhou et al, 2021a) to a unified two-stage framework. We notice that in low-light images, strong edges can also encode spatial information about motion (Joshi et al, 2008; Cho and Lee, 2009; Fu et al, 2022) like light streaks due to strong local contrast. These edges are suitable for filtering the clean events when dealing with spatially-variant blur, because they are usually caused by light streaks observed in the whole image, and the SNR of events triggered by them is relatively high. Besides, they can be detected easily by convolving the image with a Laplace kernel, which is less time-consuming than detecting an image patch containing a light streak. As shown in Figure 1 (right), instead of using a light streak to filter the clean local events in a patch (Zhou et al, 2021a), we choose to use the edge map of $\mathbf{B}$ to filter the clean events in the whole image in the first stage. In addition, we choose to use high-dimensional image features (named motion clues $\mathbf{M}$) instead of a 2D blur kernel to guide the deblurring process in the second stage, since high-dimensional image features extracted by neural networks are good at encoding complicated motion information and can also be explicitly supervised in the latent space using blur-related physical quantities (*e.g.*, bi-directional optical flows (Yuan et al, 2020)). Denoting the two stages as g_1 and g_2 respectively, the proposed unified two-stage deblurring framework could be described as

$$\mathbf{M} = g_1(\mathbf{B}, \mathbf{e}) \quad \text{and} \quad \mathbf{S} = g_2(\mathbf{B}, \mathbf{M}). \tag{5}$$

Such a two-stage framework can be applied to deblurring high-resolution RGB images without the spatially-uniform blur assumption adopted in our preliminary work (Zhou et al, 2021a), by building an RGB-DAVIS hybrid camera system that simultaneously captures a high-resolution RGB image $\mathbf{B}_{\text{RGB}}$ along with the APS image $\mathbf{B}$ and corresponding events $\mathbf{e}$, and simply substituting $\mathbf{B}$ with $\mathbf{B}_{\text{RGB}}$ in the second stage.

4 Network

Tailored to the proposed two-stage framework, we design a neural network to perform the deblurring process in a motion-aware manner, as shown in Figure 2. Following Equation (5), it consists of two stages for extracting motion clues and reconstructing the sharp image. We will detail them in Section 4.1 and Section 4.2 respectively.

4.1 Extracting motion clues

The first stage aims to extract motion clues $\mathbf{M}$ from the blurry image $\mathbf{B}$ and corresponding events $\mathbf{e}$. As shown in the first stage of Figure 2, we first adopt a Laplace kernel to convolve the blurry image $\mathbf{B}$ to acquire its edge map

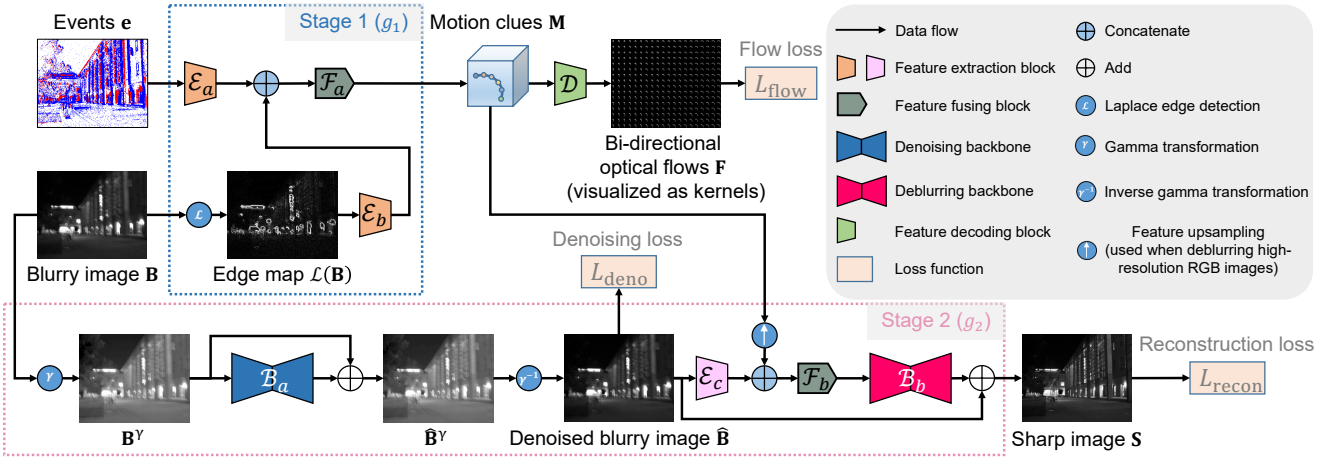

Fig. 2 We design a motion-aware neural network tailored to our unified two-stage framework (Figure 1 (right)). In the first stage, it extracts features from the edge map of the APS image and corresponding events jointly to obtain both spatial and temporal information about motion, and encodes the information into high-fidelity motion clues by explicitly using bi-directional optical flows to supervise this process in the latent space. In the second stage, it adopts a denoising module to perform blind noise suppression in the image domain to avoid ringing artifacts, and reconstructs the sharp image under the guidance of motion clues.

$\mathcal{L}(\mathbf{B})$ (where $\mathcal{L}$ denotes the operation of convoving with a Laplace kernel), and use two feature extraction blocks ($\mathcal{E}_a$ and $\mathcal{E}_b$) to extract features from $\mathbf{e}$ (stacked into a 13-channel spatiotemporal voxel grid (Zhu et al, 2019)) and $\mathcal{L}(\mathbf{B})$ respectively for obtaining both spatial and temporal information about motion. Then, we adopt a feature fusing block $\mathcal{F}_a$ to fuse the extracted features for encoding the spatial and temporal motion information into high-fidelity motion clues $\mathbf{M}$ (a 32-channel feature map which has the same spatial resolution as $\mathbf{B}$), by implicitly filtering the clean events using the edge map in the latent space. The complete process of this stage could be described as

$$\mathbf{M} = \mathcal{F}_a(\texttt{concat}(\mathcal{E}_a(\mathbf{e}), \mathcal{E}_b(\mathcal{L}(\mathbf{B})))). \tag{6}$$

Since $\mathbf{M}$ should encode spatially-variant motion information during the whole exposure period, blur-related physical quantities such as bi-directional optical flows are suitable for supervising this stage. Inspired by Yuan et al (2020), by adopting a feature decoding block $\mathcal{D}$ to decode $\mathbf{M}$, we explicitly let the output of $\mathcal{D}$ to be bi-directional optical flows $\mathbf{F}$:

$$\mathbf{F} = \mathcal{D}(\mathbf{M}). \tag{7}$$

In such a way, $\mathbf{M}$ can be effectively supervised in the latent space, providing guidance to the reconstruction of the sharp image in the second stage.

Layer details. Since both the edge map $\mathcal{L}(\mathbf{B})$ and events $\mathbf{e}$ are sparse, noisy, and non-uniformly distributed signals, extracting their features requires large receptive fields and long-range spatial dependencies. Therefore, we design the feature extraction blocks $\mathcal{E}_a$ and $\mathcal{E}_b$ to include a 1×1 convolution layer, a non-local block (Wang et al, 2018), and a dense block (Huang et al, 2017). Note that we add an instance normalization layer (Ulyanov et al, 2016) and a `ReLU` activation function after the 1×1 convolution layer for bringing sufficient non-linearity. Besides, to reduce the usage of GPU memory and save inferring time, we split the output feature map of the convolution layer into a grid of non-overlapping patches before the non-local block (Li et al, 2018). As for the feature fusing block $\mathcal{F}_a$, since it aims to implicitly filter the clean events using the edge map in the latent space, we design it to include a 1×1 convolution layer for efficient usage of the shallow features and a squeeze-and-excitation block (Hu et al, 2018a) for adaptively recalibrating channel-wise feature responses by modeling interdependencies between the feature channels of the edge map $\mathcal{L}(\mathbf{B})$ and events $\mathbf{e}$. The feature decoding block $\mathcal{D}$ is designed to be a bottleneck block (He et al, 2016) with a 1×1 convolution layer for decoding $\mathbf{M}$ into bi-directional optical flows $\mathbf{F}$.

4.2 Reconstructing the sharp image

The second stage aims to reconstruct the sharp image $\mathbf{S}$ from the blurry image $\mathbf{B}$ under the guidance of motion clues $\mathbf{M}$. Since directly using $\mathbf{M}$ to perform deconvolution-like operations on $\mathbf{B}$ would bring about ringing artifacts due to noise, we propose to adopt a denoising backbone $\mathcal{B}_a$ to perform blind noise suppression in the image domain first. Despite the noise usually resides in dark regions which makes it difficult to distinguish, we observe that by applying a gamma transformation ($\gamma < 1$), the contrast between noise and original signal could be magnified. Therefore, as shown in the second stage of Figure 2, we first apply a gamma transformation (we choose $\gamma = \frac{1}{2.2}$) to convert $\mathbf{B}$ to $\mathbf{B}^\gamma$, and use $\mathcal{B}_a$ to learn the residual between $\mathbf{B}^\gamma$ and $\hat{\mathbf{B}}^\gamma$ (where $\hat{\mathbf{B}}$ denotes the denoised

blurry image):

$$\hat{\mathbf{B}}^{\gamma} = \mathcal{B}_a(\mathbf{B}^{\gamma}) + \mathbf{B}^{\gamma}, \tag{8}$$

and then apply a inverse gamma transformation to convert $\hat{\mathbf{B}}^{\gamma}$ back to $\hat{\mathbf{B}}$. As the denoised blurry image $\hat{\mathbf{B}}$ becomes available, we use another feature extraction block $\mathcal{E}_c$ to extract its features, adopt another feature fusing block $\mathcal{F}_b$ to fuse the extracted features and the motion clues $\mathbf{M}$, and then send the fused features into a deblurring backbone $\mathcal{B}_b$ to learn the residual between $\hat{\mathbf{B}}$ and $\mathbf{S}$. This process can be written as

$$\mathbf{S} = \mathcal{B}_b(\mathcal{F}_b(\texttt{concat}(\mathcal{E}_c(\hat{\mathbf{B}}), \uparrow(\mathbf{M})))) + \hat{\mathbf{B}}, \tag{9}$$

where $\uparrow$ denotes the feature upsampling block. Note that the feature upsampling block $\uparrow$ is only used when deblurring high-resolution RGB images. As for supervision, instead of supervising in the latent space like the first stage, we propose to use the noise-free blurry image and sharp image to explicitly supervise the outputs of denoising backbone $\mathcal{B}_a$ and deblurring backbone $\mathcal{B}_b$ respectively.

Layer details. The denoising backbone $\mathcal{B}_a$ is designed to be a modified autoencoder architecture (Hinton and Salakhutdinov, 2006), since it can incorporate multi-scale information for enriching detail contents, which is proved to be effective in denoising (Gu et al, 2019; Moseley et al, 2021). Inside $\mathcal{B}_a$, we embed 3 residual bottleneck blocks (He et al, 2016) in the coarsest layer for more fine-grained contextual information. The feature extraction block $\mathcal{E}_c$ consists of only a 7×7 convolution layer (also followed by an instance normalization layer (Ulyanov et al, 2016) and a `ReLU` activation function) since it only extract features directly from the denoised blurry image $\hat{\mathbf{B}}$. The feature fusing block $\mathcal{F}_b$ is designed to be the same as $\mathcal{F}_a$ in the first stage. As for the deblurring backbone $\mathcal{B}_b$, we design it to be an attention U-Net architecture (Oktay et al, 2018), since it shows excellent localization and context generalization ability in other works that also performs the guided reconstruction operation like ours (Han et al, 2020; Zhou et al, 2020, 2021b).

5 Data preparation and implementation details

In this section, we will first detail our synthetic dataset generation pipeline in Section 5.1, then show our loss function and training strategy in Section 5.2 and Section 5.3 respectively.

5.1 Synthetic dataset generation pipeline

It is difficult to get pairwise blurry and sharp low-light images with corresponding events triggered during the exposure period. Besides, obtaining the ground truth noise-free blurry image and bi-directional optical flows for supervision is not feasible. So, we propose to generate a synthetic dataset for training our network. Since existing event-based deblurring benchmarks (Wang et al, 2020a; Xu et al, 2021) do not contain images captured in low-light conditions, we cannot use them to generate our dataset. However, we find that the RealBlur dataset (Rim et al, 2020) contains abundant night scenes with noise-free high-resolution RGB images, making it the desired data source for generating our dataset. In short, for each scene, our synthetic dataset generation pipeline could be described as the following steps:

(1) Resize the source image to 960×760 pixels to serve as the ground truth noise-free high-resolution RGB image $\mathbf{S}_{\text{RGB}}$;
(2) convert $\mathbf{S}_{\text{RGB}}$ to grayscale, and resize it to 320×256 pixels to serve as the ground truth noise-free APS image $\mathbf{S}$;
(3) randomly adjust the dynamic range of $\mathbf{S}_{\text{RGB}}$ and $\mathbf{S}$ to make some pixels saturated;
(4) randomly generate a base camera motion trajectory using the algorithm proposed in (Boracchi and Foi, 2012), and use the spatially-variant blur model proposed by Whyte et al (2012) to make it pixel-wise;
(5) obtain the bi-directional optical flows $\mathbf{F}$ from the pixel-wise trajectory using the method proposed by Hyun Kim and Mu Lee (2015);
(6) move $\mathbf{S}_{\text{RGB}}$ and $\mathbf{S}$ along the pixel-wise trajectory to get multiple (25 in our experiments) latent frames during the exposure period, and use V2E (Delbruck et al, 2020) (applying the "noisy mode") to generate corresponding events $\mathbf{e}$ from the latent frames of $\mathbf{S}$;
(7) average the latent frames to get the noise-free blurry images $\hat{\mathbf{B}}_{\text{RGB}}$ and $\hat{\mathbf{B}}$, and adopt the low-light noise model proposed by Lv et al (2021) to acquire the blurry images $\mathbf{B}_{\text{RGB}}$ and $\mathbf{B}$.

For evaluation, we choose the images provided by Hu et al (2018b) as source images, and also adopt the above dataset generation pipeline to generate the test dataset. Note that for each scene, we randomly generate 20 different camera motion trajectories and further perform data augmentation (*e.g.*, flipping and rotating) to avoid overfitting, so that our training (test) dataset contains 4680 (220) different images in total.

5.2 Loss function

The total loss function of our network L consists of three terms: flow loss L_{flow}, denoising loss L_{deno}, and reconstruction loss L_{recon}, which is defined as

$$\begin{aligned} &L(\mathbf{F}, \hat{\mathbf{B}}, \mathbf{S}, \mathbf{F}_{\text{gt}}, \hat{\mathbf{B}}_{\text{gt}}, \mathbf{S}_{\text{gt}}) \\ &= L_{\text{flow}}(\mathbf{F}, \mathbf{F}_{\text{gt}}) + L_{\text{deno}}(\hat{\mathbf{B}}, \hat{\mathbf{B}}_{\text{gt}}) + L_{\text{recon}}(\mathbf{S}, \mathbf{S}_{\text{gt}}), \end{aligned} \tag{10}$$

where the subscript gt labels the ground truth throughout this paper. We will detail each of them in the following.

Flow loss. The flow loss term L_{flow} aims to supervise the first stage (the extraction of motion clues) in the latent space, which could be written as

$$L_{\text{flow}}(\mathbf{F}, \mathbf{F}_{\text{gt}}) = \beta_{\text{flow}_a} L_1(\mathbf{F}, \mathbf{F}_{\text{gt}}) + \beta_{\text{flow}_b} L_{\text{tv}}(\mathbf{F}), \tag{11}$$

where L_1 denotes the ℓ_1 loss, L_{tv} is the total variation loss to enforce smoothness, β_{flow_a} and β_{flow_b} are empirically set to be 0.1 and 0.01 respectively.

Denoising loss. The denoising loss term L_{deno} aims to supervise the denoising backbone for obtaining noise-free blurry images in the second stage, which could be written as

$$L_{\text{deno}}(\hat{\mathbf{B}}, \hat{\mathbf{B}}_{\text{gt}}) = \beta_{\text{deno}_a} L_1(\hat{\mathbf{B}}, \hat{\mathbf{B}}_{\text{gt}}) + \beta_{\text{deno}_b} L_2(\hat{\mathbf{B}}, \hat{\mathbf{B}}_{\text{gt}}), \tag{12}$$

where L_2 denotes the ℓ_2 loss, β_{deno_a} and β_{deno_b} are empirically set to be 10.0 and 10.0 respectively.

Reconstruction loss. The reconstruction loss term L_{recon} aims to supervise the deblurring backbone for reconstructing high-quality sharp images in the second stage, which could be written as

$$L_{\text{recon}}(\mathbf{S}, \mathbf{S}_{\text{gt}}) = \beta_{\text{recon}_a} L_2(\mathbf{S}, \mathbf{S}_{\text{gt}}) + \beta_{\text{recon}_b} L_{\text{perc}}(\mathbf{S}, \mathbf{S}_{\text{gt}}), \tag{13}$$

where L_{perc} denotes the perceptual loss, β_{recon_a} and β_{recon_b} are empirically set to be 100.0 and 0.1 respectively. The perceptual loss L_{perc} is defined as

$$L_{\text{perc}}(\mathbf{S}, \mathbf{S}_{\text{gt}}) = L_2(\phi_h(\mathbf{S}), \phi_h(\mathbf{S}_{\text{gt}})), \tag{14}$$

where ϕ_h denotes the feature map from h-th layer of VGG-19 network (Simonyan and Zisserman, 2014) pretrained on ImageNet (Russakovsky et al, 2015), and here we use activations from $VGG_{3,3}$ convolutional layer.

5.3 Training strategy

We implement our method using PyTorch on a PC with an Intel Core i7-8700K CPU and an NVIDIA 2080Ti GPU. The network is trained for 100 epochs with a batch size of 8. For optimization, we use Adam optimizer (Kingma and Ba, 2014) with $\beta_1 = 0.5$, $\beta_2 = 0.999$. The learning rate is set to be 0.001 during the training process without change. The network parameters are initialized with Xavier initialization (Glorot and Bengio, 2010).

6 Experiments

In this section, we make comparisons with other methods and conduct ablation study on the synthetic dataset in Section 6.1 and Section 6.2 respectively, and use an event camera (DAVIS346) and our RGB-DAVIS hybrid camera system to capture real-world images for further evaluation in Section 6.3 and Section 6.4 respectively.

6.1 Evaluation on synthetic data

We compare our method with four state-of-the-art event-based deblurring methods including (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021) and our preliminary work (Zhou et al, 2021a), and four state-of-the-art image-based deblurring methods including (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021). Visual quality comparisons are shown in Figure 3[2]. From the results we can see that the result generated by our method resembles the ground truth more closely. Despite the method proposed by Pan et al (2019) could relieve motion blur, it generates over-smooth result since it is based on numerical optimization which cannot make full use of semantic information. The method proposed by Lin et al (2020) fails to deblur since it requires information of multiple adjacent frames, which is unavailable in our settings (see Section 3.2). The method proposed by Xu et al (2021) darkens the whole scene since it is not good at handling the dark regions. Our preliminary work (Zhou et al, 2021a) does not perform well since it cannot deal with spatially-variant blur. The image-based deblurring methods (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) cannot handle the regions with strong local contrast (*e.g.*, edges) robustly since they cannot obtain temporal information about motion from events; in addition, they often suffer from ringing artifacts due to noise.

To evaluate the deblurring results quantitatively, we adopt four frequently-used image quality metrics including PSNR (peak signal-to-noise ratio), SSIM (structural similarity), MS-SSIM (multi-scale SSIM), and LPIPS (learned perceptual image patch similarity (Zhang et al, 2018b), higher (lower) means more different (similar) to ground truth, which is different from other metrics). Results are shown in Table 1 (also labeled in the top right of corresponding examples in Figure 3). Our model consistently outperforms the compared methods on all metrics.

6.2 Ablation study

To verify the validity of each design choice, we conduct a series of ablation studies and show comparisons in Table 2. We first verify the significance of events that encode temporal information about motion by comparing with a model that performs deblurring without using events (W/o events). From the result we can see that the performance degenerates severely, since the problem turns into an image-based deblurring problem which is too ill-posed. Then, we demonstrate the necessity of taking the edges of the blurry image as input by removing them (W/o edges). We find that this model does not perform as well as our complete model since the edges can filter clean events to avoid the artifacts caused

[2] Additional results can be found in the supplementary material.

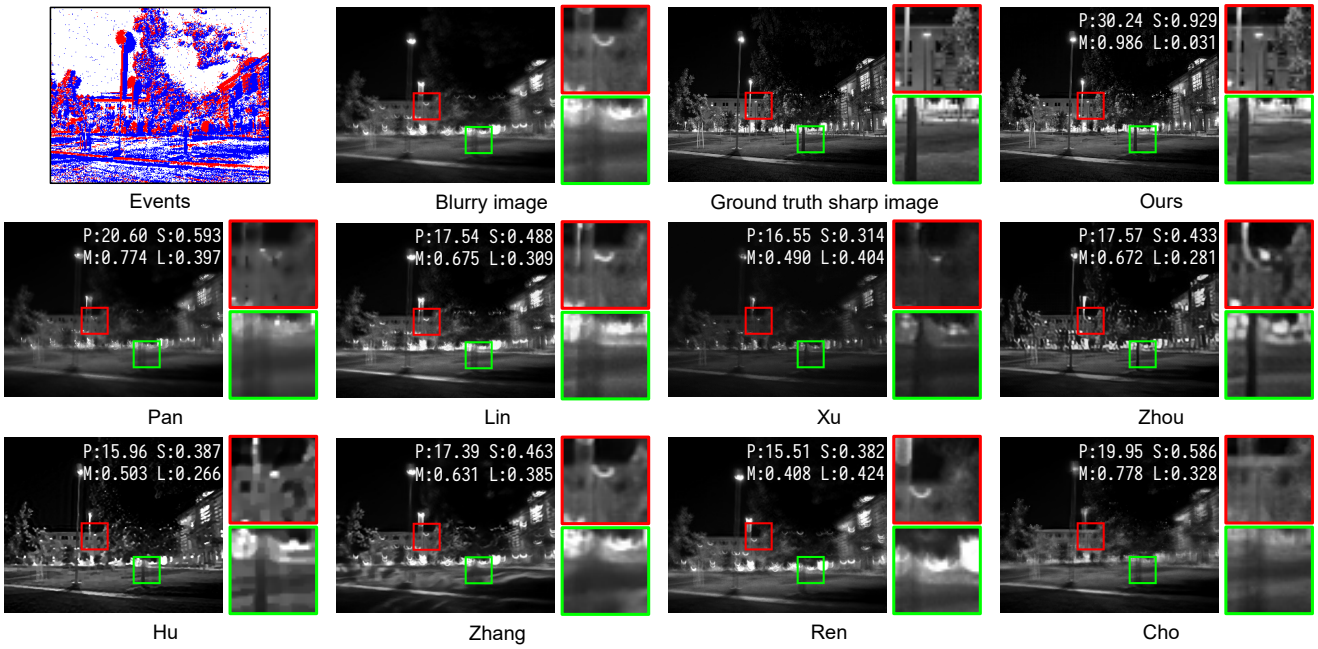

Fig. 3 Qualitative comparisons on synthetic data among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021a), and four image-based deblurring methods (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021). Quantitative results evaluated using PSNR (P), SSIM (S), MS-SSIM (M), and LPIPS (L) are labeled in the top right of each image.

Table 1 Qualitative comparisons on synthetic data among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021a), and four image-based deblurring methods (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021). ↑ (↓) means the higher (lower) the metrics the better the results throughout this paper. **Bold** font indicates the best performance throughout this paper.

	(Pan et al, 2019)	(Lin et al, 2020)	(Xu et al, 2021)	(Zhou et al, 2021a)	(Hu et al, 2018b)	(Zhang et al, 2019)	(Ren et al, 2020)	(Cho et al, 2021)	Ours
PSNR↑	21.92	19.18	16.63	18.13	18.13	18.47	17.10	19.47	**31.61**
SSIM↑	0.699	0.604	0.497	0.541	0.543	0.557	0.475	0.604	**0.912**
MS-SSIM↑	0.821	0.736	0.602	0.696	0.669	0.668	0.515	0.721	**0.982**
LPIPS↓	0.329	0.307	0.353	0.275	0.270	0.382	0.376	0.323	**0.051**

Table 2 Quantitative evaluation results of ablation study.

	PSNR↑	SSIM↑	MS-SSIM↑	LPIPS↓
W/o events	23.11	0.724	0.875	0.155
W/o edges	31.18	0.892	0.976	0.056
W/o flows	31.11	0.905	0.977	0.058
W/o denoising	30.99	0.897	0.975	0.061
W/o residual	27.82	0.776	0.946	0.058
Our complete model	**31.61**	**0.912**	**0.982**	**0.051**

by noisy events. Furthermore, we validate the effectiveness of using the bi-directional optical flows for supervision by removing the flow loss (W/o flows), and show the importance of adopting the denoising module to perform blind noise suppression in the image domain by removing it (W/o denoising). Finally, we validate the residual learning strategy adopted in the second stage (learn the residual between $\hat{\mathbf{B}}$ and $\mathbf{S}$, see Equation (9) for details) by comparing with a model directly learning $\mathbf{S}$ (W/o residual). These results show that our complete model achieves the optimal performance with the proposed specific designs.

6.3 Evaluation on real data captured by an event camera

To show that our method has a good generalization ability on real low-light images and events, we capture several images from various scenes along with corresponding events using an event camera (DAVIS346). As shown in Fig. 4[3], our method generalizes well with excellent performance. Taking the green box as an example, the shadow on the ground cannot be restored robustly by event-based deblurring methods that directly adopt the EDI model to perform pixel-wise reconstruction (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021), because the EDI model is vulnerable to noisy events in such a dark region; the bicycle wheel restored by our preliminary work (Zhou et al, 2021a) is still blurry due to its spatially-uniform blur assumption; the image-based deblurring methods (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) suffer from ringing artifacts severely, leading to overlapped fringes.

[3] Additional results can be found in the supplementary material.

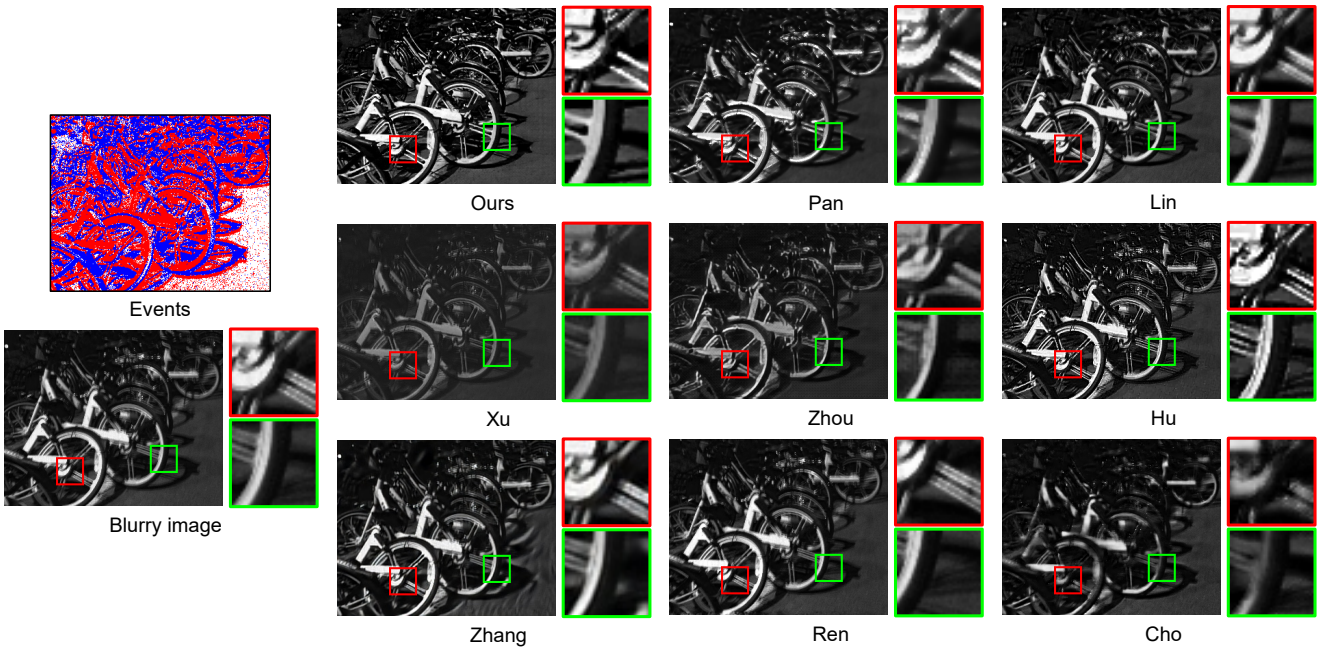

Fig. 4 Qualitative comparisons on real data captured by an event camera (DAVIS346) among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021a), and four image-based deblurring methods (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021).

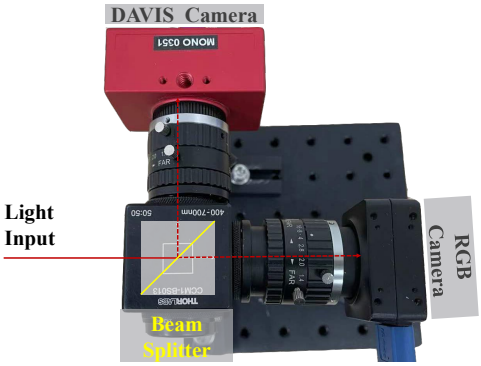

Fig. 5 Our RGB-DAVIS hybrid camera system, which consists of an RGB camera (PointGrey Chameleon3) and an event camera (DAVIS346). A beam splitter is placed in front of them to make their fields of view aligned.

6.4 Evaluation on real data captured by a hybrid camera

To demonstrate that out method has the ability to deblur high-resolution RGB images with events, we build an RGB-DAVIS hybrid camera system consisting of an RGB camera (PointGrey Chameleon3) and an event camera (DAVIS346) with the same F/1.4 lens to capture high-resolution RGB images and low-resolution APS images along with corresponding events, as shown in Fig. 5. To ensure the motion trajectories of the two sensors are approximately the same, we use a beam splitter in front of them to make their fields of view aligned (Han et al, 2020; Wang et al, 2020b). Note that in our experiments we resize and crop the central part of the RGB and APS images to 960×768 and 320×256 pixels respectively.

To deblur an RGB image, we need an APS image captured simultaneously with it, along with corresponding events triggered during the exposure period. However, it is non-trivial to achieve precise temporal synchronization unless we can configure a synchronized clock to trigger two cameras simultaneously at the chip level, which is beyond the scope of this paper. Therefore, we propose an alternative strategy to achieve approximated temporal synchronization. First, to alleviate the software delay, we write a script to simultaneously trigger the capturing programs of two cameras. Then, to relieve the negative impact caused by hardware delay, we periodically capture a scene and select the "best" matched image pair between RGB and APS images. In such a periodical capturing process, we first set the exposure time of the RGB images to the same value as the APS images and set the RGB camera to burst mode, then capture a sequence of RGB images and APS images and select an image pair with the closest appearance by scaling them to the same size and seeking a pair with maximum MS-SSIM value. After selecting the APS image, the events triggered during the exposure period can be extracted since APS images and events are well synchronized in the event camera.

Visual quality comparisons are shown in Fig. 6[4]. Note that we only compare our method with our preliminary work (Zhou et al, 2021a) and image-based deblurring methods (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al,

[4] Additional results can be found in the supplementary material.

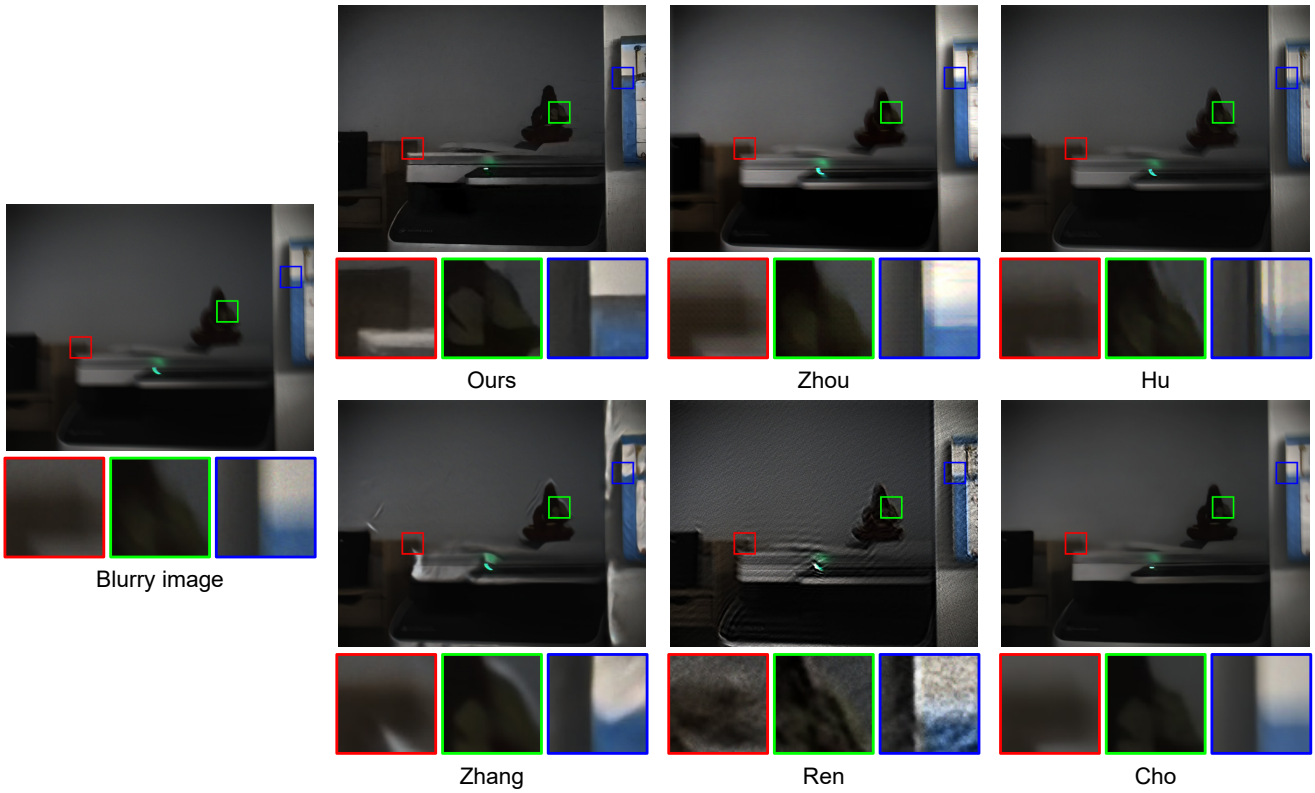

Fig. 6 Qualitative comparisons on real data captured by our RGB-DAVIS hybrid camera system among our method, our preliminary work (Zhou et al, 2021a), and four image-based deblurring methods (Hu et al, 2018b; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021).

2021) since event-based methods that directly adopt the EDI model can only handle low-resolution APS images (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021). This proof-of-concept experiment shows a great potential of applying events to deblurring images satisfying modern camera specifications and daily photography.

7 Conclusion and discussion

We propose a unified two-stage framework to apply events to deblurring low-light images. Instead of performing pixel-wise reconstruction based on the EDI model, it extracts high-fidelity motion clues by utilizing the edge map of the APS image to filter the clean events in the first stage, and reconstructs the sharp image under the guidance of motion clues in the second stage. Tailored to such a framework, we further design a motion-aware neural network to perform the deblurring process, and demonstrate the ability to deblur high-resolution RGB images with events, without the spatially-uniform blur assumption. Experimental results show our method achieves state-of-the-art performance over image-based and event-based solutions on both synthetic and real-world images.

Flexibility of the proposed framework. Due to the complexity of the proposed framework, errors could be accumulated in the deblurring process. For example, in addition to strong edges, weak edges containing noisy textures could also be detected by the Laplace kernel. Therefore, once the first stage fails to process the less reliable features extracted from those weak edges, the filtered events would still be noisy so that the obtained motion clues would become inaccurate, leading to artifacts in the reconstructed sharp images. This problem could be mitigated by adopting modern learning-based edge detection algorithms Yu et al (2017); Liu et al (2019) to estimate the edge map in a noise-resistant manner. In our future work, we plan to add a plug-and-play edge detection network module to replace the Laplace edge detection operation in our framework to further increase the robustness and flexibility.

Limitations. Since our method is designed for deblurring a single image with corresponding events triggered during the exposure period, we cannot reconstruct high-frame-rate video like other event-based deblurring methods which directly adopt the EDI model (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021). In addition, our RGB-DAVIS hybrid camera system cannot achieve precise temporal synchronization, and this could be solved by locating the two sensors in the same chip with different resolutions (pixel sizes) in the future.

Acknowledgement

This work is supported by National Key R&D Program of China (2020AAA0105200) and National Natural Science Foundation of China under Grant No. 62136001, 62088102, 62276007, and 61876007.

References

Baldwin R, Almatrafi M, Asari V, Hirakawa K (2020) Event probability mask (EPM) and event denoising convolutional neural network (EDnCNN) for neuromorphic cameras. In: Proc. of Computer Vision and Pattern Recognition, pp 1701–1710

Barrios-Avilés J, Rosado-Muñoz A, Medus LD, Bataller-Mompeán M, Guerrero-Martínez JF (2018) Less data same information for event-based sensors: A bioinspired filtering and data reduction algorithm. Sensors 18(12):4122

Boracchi G, Foi A (2012) Modeling the performance of image restoration from motion blur. IEEE Transactions on Image Processing 21(8):3502–3517

Brandli C, Berner R, Yang M, Liu SC, Delbruck T (2014) A 240 $\times$ 180 130 dB 3 μs latency global shutter spatiotemporal vision sensor. IEEE Journal of Solid-State Circuits 49(10):2333–2341

Chakrabarti A (2016) A neural approach to blind motion deblurring. In: Proc. of European Conference on Computer Vision, pp 221–235

Chan TF, Wong CK (1998) Total variation blind deconvolution. IEEE Transactions on Image Processing 7(3):370–375

Chen C, Chen Q, Xu J, Koltun V (2018a) Learning to see in the dark. In: Proc. of Computer Vision and Pattern Recognition, pp 3291–3300

Chen H, Gu J, Gallo O, Liu MY, Veeraraghavan A, Kautz J (2018b) Reblur2Deblur: Deblurring videos via self-supervised learning. In: Proc. of International Conference on Computational Photography, pp 1–9

Chen H, Teng M, Shi B, Wang Y, Huang T (2020) Learning to deblur and generate high frame rate video with an event camera. arXiv preprint arXiv:200300847

Chen L, Zhang J, Lin S, Fang F, Ren JS (2021a) Blind deblurring for saturated images. In: Proc. of Computer Vision and Pattern Recognition, pp 6308–6316

Chen L, Zhang J, Pan J, Lin S, Fang F, Ren JS (2021b) Learning a non-blind deblurring network for night blurry images. In: Proc. of Computer Vision and Pattern Recognition, pp 10542–10550

Chi Z, Wang Y, Yu Y, Tang J (2021) Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning. In: Proc. of Computer Vision and Pattern Recognition, pp 9137–9146

Cho S, Lee S (2009) Fast motion deblurring. In: Proc. of ACM SIGGRAPH Asia, pp 1–8

Cho SJ, Ji SW, Hong JP, Jung SW, Ko SJ (2021) Rethinking coarse-to-fine approach in single image deblurring. In: Proc. of International Conference on Computer Vision, pp 4641–4650

Delbruck T, Hu Y, He Z (2020) V2E: From video frames to realistic DVS event camera streams. arXiv preprint arXiv:200607722

Dong J, Pan J, Su Z, Yang MH (2017) Blind image deblurring with outlier handling. In: Proc. of International Conference on Computer Vision, pp 2478–2486

Dong J, Roth S, Schiele B (2021) Learning spatially-variant MAP models for non-blind image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 4886–4895

Duan P, Wang ZW, Zhou X, Ma Y, Shi B (2021) EventZoom: Learning to denoise and super resolve neuromorphic events. In: Proc. of Computer Vision and Pattern Recognition, pp 12824–12833

Fergus R, Singh B, Hertzmann A, Roweis ST, Freeman WT (2006) Removing camera shake from a single photograph. In: Proc. of ACM SIGGRAPH, pp 787–794

Fu Z, Zheng Y, Ma T, Ye H, Yang J, He L (2022) Edge-aware deep image deblurring. Neurocomputing

Gallego G, Rebecq H, Scaramuzza D (2018) A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In: Proc. of Computer Vision and Pattern Recognition, pp 3867–3876

Gao H, Tao X, Shen X, Jia J (2019) Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: Proc. of Computer Vision and Pattern Recognition, pp 3848–3856

Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proc. of International Conference on Artificial Intelligence and Statistics, pp 249–256

Gong D, Yang J, Liu L, Zhang Y, Reid I, Shen C, Van Den Hengel A, Shi Q (2017) From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In: Proc. of Computer Vision and Pattern Recognition, pp 2319–2328

Gu S, Li Y, Gool LV, Timofte R (2019) Self-guided network for fast image denoising. In: Proc. of International Conference on Computer Vision, pp 2511–2520

Han J, Zhou C, Duan P, Tang Y, Xu C, Xu C, Huang T, Shi B (2020) Neuromorphic camera guided high dynamic range imaging. In: Proc. of Computer Vision and Pattern Recognition, pp 1730–1739

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proc. of Computer Vision and Pattern Recognition, pp 770–778

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507

Hu J, Shen L, Sun G (2018a) Squeeze-and-excitation networks. In: Proc. of Computer Vision and Pattern Recognition, pp 7132–7141

Hu Z, Cho S, Wang J, Yang MH (2018b) Deblurring low-light images with light streaks. IEEE Transactions on Pattern Analysis and Machine Intelligence 40(10):2329–2341

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proc. of Computer Vision and Pattern Recognition

Hyun Kim T, Mu Lee K (2015) Generalized video deblurring for dynamic scenes. In: Proc. of Computer Vision and Pattern Recognition, pp 5426–5434

Jiang Z, Zhang Y, Zou D, Ren J, Lv J, Liu Y (2020) Learning event-based motion deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 3320–3329

Joshi N, Szeliski R, Kriegman DJ (2008) PSF estimation using sharp edge prediction. In: Proc. of Computer Vision and Pattern Recognition, pp 1–8

Kaufman A, Fattal R (2020) Deblurring using analysis-synthesis networks pair. In: Proc. of Computer Vision and Pattern Recognition, pp 5811–5820

Khodamoradi A, Kastner R (2018) $O(N)$-Space spatiotemporal filter for reducing noise in neuromorphic vision sensors. IEEE Transactions on Emerging Topics in Computing 9(1):15–23

Kingma DP, Ba J (2014) ADAM: A method for stochastic optimization. arXiv preprint arXiv:14126980

Krishnan D, Tay T, Fergus R (2011) Blind deconvolution using a normalized sparsity measure. In: Proc. of Computer Vision and Pattern Recognition, pp 233–240

Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J (2018) DeblurGAN: Blind motion deblurring using conditional adversarial networks. In: Proc. of Computer Vision and Pattern Recognition, pp 8183–8192

Kupyn O, Martyniuk T, Wu J, Wang Z (2019) DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In: Proc. of International Conference on Computer Vision, pp 8878–8887

Li C, Guo C, Han LH, Jiang J, Cheng MM, Gu J, Loy CC (2021) Low-light image and video enhancement using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence pp 1–1

Li G, He X, Zhang W, Chang H, Dong L, Lin L (2018) Non-locally enhanced encoder-decoder network for single image de-raining. In: Proc. of ACM MM, pp 1056–1064

Lichtsteiner P, Posch C, Delbruck T (2008) A 128 $\times$ 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. IEEE Journal of Solid-State Circuits 43(2):566–576

Lin S, Zhang J, Pan J, Jiang Z, Zou D, Wang Y, Chen J, Ren J (2020) Learning event-driven video deblurring and interpolation. In: Proc. of European Conference on Computer Vision

Liu H, Brandli C, Li C, Liu SC, Delbruck T (2015) Design of a spatiotemporal correlation filter for event-based sensors. In: International Symposium on Circuits and Systems, pp 722–725

Liu J, Xu D, Yang W, Fan M, Huang H (2021) Benchmarking low-light image enhancement and beyond. International Journal of Computer Vision 129(4):1153–1184

Liu Y, Cheng MM, Hu X, Bian JW, Zhang L, Bai X, Tang J (2019) Richer convolutional features for edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 41(08):1939–1946

Lv F, Li Y, Lu F (2021) Attention guided low-light image enhancement with a large scale low-light simulation dataset. International Journal of Computer Vision 129(7):2175–2193

Maharjan P, Li L, Li Z, Xu N, Ma C, Li Y (2019) Improving extreme low-light image denoising via residual learning. In: Proc. of International Conference on Multimedia and Expo

Michaeli T, Irani M (2014) Blind deblurring using internal patch recurrence. In: Proc. of European Conference on Computer Vision, pp 783–798

Mitrokhin A, Fermüller C, Parameshwara C, Aloimonos Y (2018) Event-based moving object detection and tracking. In: Proc. of International Conference on Intelligent Robots and Systems, pp 1–9

Moseley B, Bickel V, López-Francos IG, Rana L (2021) Extreme low-light environment-driven image denoising over permanently shadowed lunar regions with a physical noise model. In: Proc. of Computer Vision and Pattern Recognition, pp 6317–6327

Nah S, Hyun Kim T, Mu Lee K (2017) Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 3883–3891

Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018) Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999

Pan J, Hu Z, Su Z, Yang MH (2016a) L_0-regularized intensity and gradient prior for deblurring text images and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(2):342–355

Pan J, Sun D, Pfister H, Yang MH (2016b) Blind image deblurring using dark channel prior. In: Proc. of Computer Vision and Pattern Recognition, pp 1628–1636

Pan L, Scheerlinck C, Yu X, Hartley R, Liu M, Dai Y (2019) Bringing a blurry frame alive at high frame-rate with an event camera. In: Proc. of Computer Vision and Pattern Recognition, pp 6820–6829

Pan L, Liu M, Hartley R (2020) Single image optical flow estimation with an event camera. In: Proc. of Computer Vision and Pattern Recognition, pp 1669–1678

Ren D, Zhang K, Wang Q, Hu Q, Zuo W (2020) Neural blind deconvolution using deep priors. In: Proc. of Computer Vision and Pattern Recognition, pp 3341–3350

Rim J, Lee H, Won J, Cho S (2020) Real-world blur dataset for learning and benchmarking deblurring algorithms. In: Proc. of European Conference on Computer Vision

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. International Journal of Computer Vision 115(3):211–252

Shan Q, Jia J, Agarwala A (2008) High-quality motion deblurring from a single image. ACM Transactions on Graphics (Proc of ACM SIGGRAPH) 27(3):1–10

Shang W, Ren D, Zou D, Ren JS, Luo P, Zuo W (2021) Bringing events into video deblurring with non-consecutively blurry frames. In: Proc. of International Conference on Computer Vision, pp 4531–4540

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556

Stoffregen T, Gallego G, Drummond T, Kleeman L, Scaramuzza D (2019) Event-based motion segmentation by motion compensation. In: Proc. of Computer Vision and Pattern Recognition, pp 7244–7253

Suin M, Purohit K, Rajagopalan A (2020) Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 3606–3615

Sun J, Cao W, Xu Z, Ponce J (2015) Learning a convolutional neural network for non-uniform motion blur removal. In: Proc. of Computer Vision and Pattern Recognition, pp 769–777

Tao X, Gao H, Shen X, Wang J, Jia J (2018) Scale-recurrent network for deep image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 8174–8182

Tran P, Tran AT, Phung Q, Hoai M (2021) Explore image deblurring via encoded blur kernel space. In: Proc. of Computer Vision and Pattern Recognition, pp 11956–11965

Ulyanov D, Vedaldi A, Lempitsky V (2016) Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:160708022

Wang B, He J, Yu L, Xia GS, Yang W (2020a) Event enhanced high-quality image recovery. In: Proc. of European Conference on Computer Vision, pp 155–171

Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: Proc. of Computer Vision and Pattern Recognition, pp 7794–7803

Wang Y, Du B, Shen Y, Wu K, Zhao G, Sun J, Wen H (2019) EV-Gait: Event-based robust gait recognition using dynamic vision sensors. In: Proc. of Computer Vision and Pattern Recognition, pp 6358–6367

Wang Z, Duan P, Cossairt O, Katsaggelos A, Huang T, Shi B (2020b) Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In: Proc. of Computer Vision and Pattern Recognition, pp 1609–1619

Wei K, Fu Y, Yang J, Huang H (2020) A physics-based noise formation model for extreme low-light raw denoising. In: Proc. of Computer Vision and Pattern Recognition, pp 2758–2767

Whyte O, Sivic J, Zisserman A, Ponce J (2012) Non-uniform deblurring for shaken images. International Journal of Computer Vision 98(2):168–186

Xu F, Yu L, Wang B, Yang W, Xia GS, Jia X, Qiao Z, Liu J (2021) Motion deblurring with real events. In: Proc. of International Conference on Computer Vision, pp 2583–2592

Xu L, Zheng S, Jia J (2013) Unnatural L_0 sparse representation for natural image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 1107–1114

Yan Y, Ren W, Guo Y, Wang R, Cao X (2017) Image deblurring via extreme channels prior. In: Proc. of Computer Vision and Pattern Recognition, pp 4003–4011

Yu Z, Feng C, Liu MY, Ramalingam S (2017) CASENet: Deep category-aware semantic edge detection. In: Proc. of Computer Vision and Pattern Recognition, pp 5964–5973

Yuan Y, Su W, Ma D (2020) Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training. In: Proc. of Computer Vision and Pattern Recognition, pp 3555–3564

Zhang H, Dai Y, Li H, Koniusz P (2019) Deep stacked hierarchical multi-patch network for image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 5978–5986

Zhang J, Pan J, Ren J, Song Y, Bao L, Lau RW, Yang MH (2018a) Dynamic scene deblurring using spatially variant recurrent neural networks. In: Proc. of Computer Vision and Pattern Recognition, pp 2521–2529

Zhang K, Luo W, Zhong Y, Ma L, Stenger B, Liu W, Li H (2020a) Deblurring by realistic blurring. In: Proc. of Computer Vision and

Pattern Recognition, pp 2737–2746
Zhang R, Isola P, Efros AA, Shechtman E, Wang O (2018b) The unreasonable effectiveness of deep features as a perceptual metric. In: Proc. of Computer Vision and Pattern Recognition
Zhang S, Zhang Y, Jiang Z, Zou D, Ren J, Zhou B (2020b) Learning to see in the dark with events. In: Proc. of European Conference on Computer Vision, pp 666–682
Zhong L, Cho S, Metaxas D, Paris S, Wang J (2013) Handling noise in single image deblurring using directional filters. In: Proc. of Computer Vision and Pattern Recognition, pp 612–619
Zhou C, Zhao H, Han J, Xu C, Xu C, Huang T, Shi B (2020) UnModNet: Learning to unwrap a modulo image for high dynamic range imaging. In: Proc. of Advances in Neural Information Processing Systems
Zhou C, Teng M, Han J, Xu C, Shi B (2021a) DeLiEve-Net: Deblurring low-light images with light streaks and local events. In: Proc. of International Conference on Computer Vision Workshops, pp 1155–1164
Zhou C, Teng M, Han Y, Xu C, Shi B (2021b) Learning to dehaze with polarization. In: Proc. of Advances in Neural Information Processing Systems
Zhu AZ, Yuan L, Chaney K, Daniilidis K (2019) Unsupervised event-based learning of optical flow, depth, and egomotion. In: Proc. of Computer Vision and Pattern Recognition, pp 989–997

Noname manuscript No.
(will be inserted by the editor)

Supplementary Material: Deblurring Low-Light Images with Events

Chu Zhou · Minggui Teng · Jin Han · Jinxiu Liang · Chao Xu · Gang Cao · Boxin Shi

Received: date / Accepted: date

8 Additional results on synthetic data

In this section, we provide additional qualitative comparisons on synthetic data among four state-of-the-art event-based deblurring methods including (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021) and our preliminary work (Zhou et al, 2021), and four state-of-the-art image-based deblurring methods including (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021), as shown in Fig. 7, Fig. 8, and Fig. 9, corresponding to Footnote 2 in Section 6.1 of the paper.

9 Additional results on real data captured by an event camera

In this section, we provide additional qualitative comparisons on real data captured by an event camera (DAVIS346) among four state-of-the-art event-based deblurring methods including (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021) and our preliminary work (Zhou et al, 2021), and four state-of-the-art image-based deblurring methods including (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021), as shown in Fig. 10, Fig. 11, and Fig. 12, corresponding to Footnote 3 in Section 6.3 of the paper.

10 Additional results on real data captured by a hybrid camera

In this section, we provide additional qualitative comparisons on real data captured by our RGB-DAVIS hybrid camera system among our preliminary work (Zhou et al, 2021) and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021), as shown in Fig. 13, corresponding to Footnote 4 in Section 6.4 of the paper.

✉ Boxin Shi
E-mail: shiboxin@pku.edu.cn

Chu Zhou · Chao Xu
Key Laboratory of Machine Perception (MOE), School of Intelligence Science and Technology, Peking University, China

Minggui Teng · Jinxiu Liang · Boxin Shi
National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, China

Jin Han
Graduate School of Information Science and Technology, The University of Tokyo, Japan

Gang Cao · Boxin Shi
Beijing Academy of Artificial Intelligence, China

References

Cho SJ, Ji SW, Hong JP, Jung SW, Ko SJ (2021) Rethinking coarse-to-fine approach in single image deblurring. In: Proc. of International Conference on Computer Vision, pp 4641–4650

Hu Z, Cho S, Wang J, Yang MH (2018) Deblurring low-light images with light streaks. IEEE Transactions on Pattern Analysis and Machine Intelligence 40(10):2329–2341

Lin S, Zhang J, Pan J, Jiang Z, Zou D, Wang Y, Chen J, Ren J (2020) Learning event-driven video deblurring and interpolation. In: Proc. of European Conference on Computer Vision

Pan L, Scheerlinck C, Yu X, Hartley R, Liu M, Dai Y (2019) Bringing a blurry frame alive at high frame-rate with an event camera. In: Proc. of Computer Vision and Pattern Recognition, pp 6820–6829

Ren D, Zhang K, Wang Q, Hu Q, Zuo W (2020) Neural blind deconvolution using deep priors. In: Proc. of Computer Vision and Pattern Recognition, pp 3341–3350

Xu F, Yu L, Wang B, Yang W, Xia GS, Jia X, Qiao Z, Liu J (2021) Motion deblurring with real events. In: Proc. of International Conference on Computer Vision, pp 2583–2592

Zhang H, Dai Y, Li H, Koniusz P (2019) Deep stacked hierarchical multi-patch network for image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 5978–5986

Zhou C, Teng M, Han J, Xu C, Shi B (2021) DeLiEve-Net: Deblurring low-light images with light streaks and local events. In: Proc. of International Conference on Computer Vision Workshops, pp 1155–1164

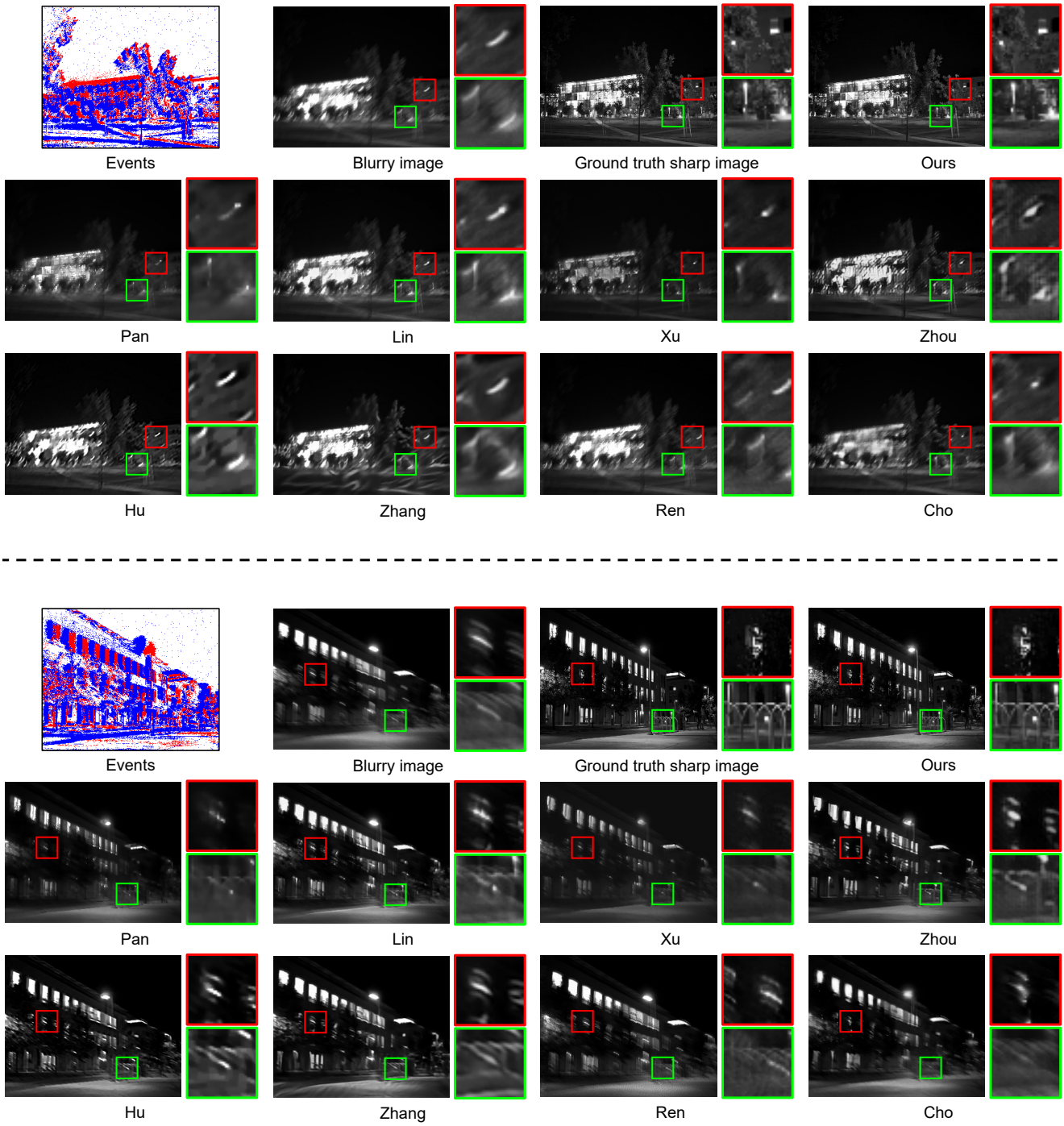

Fig. 7 Additional qualitative comparisons on synthetic data among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021), and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) (part 1).

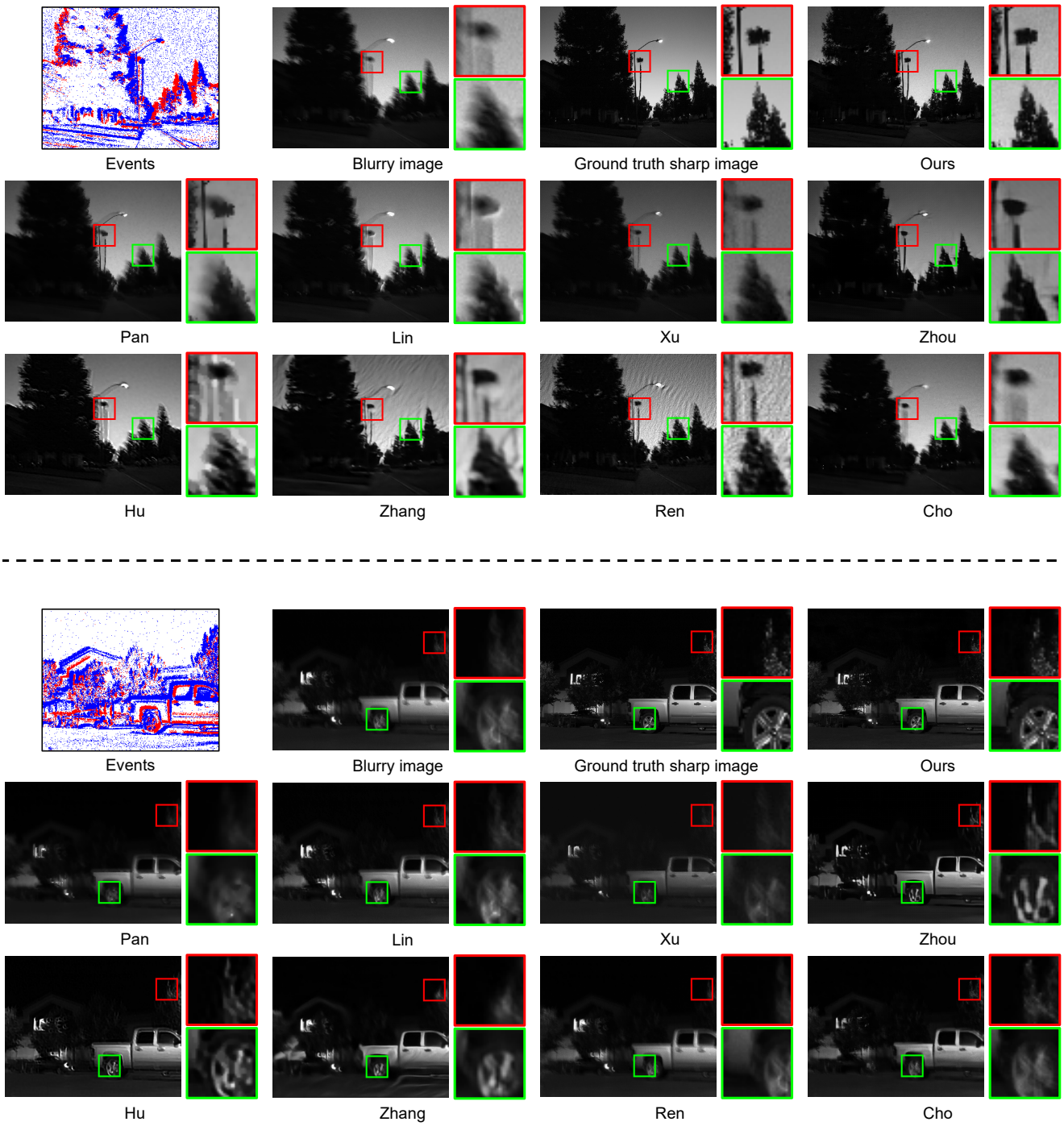

Fig. 8 Additional qualitative comparisons on synthetic data among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021), and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) (part 2).

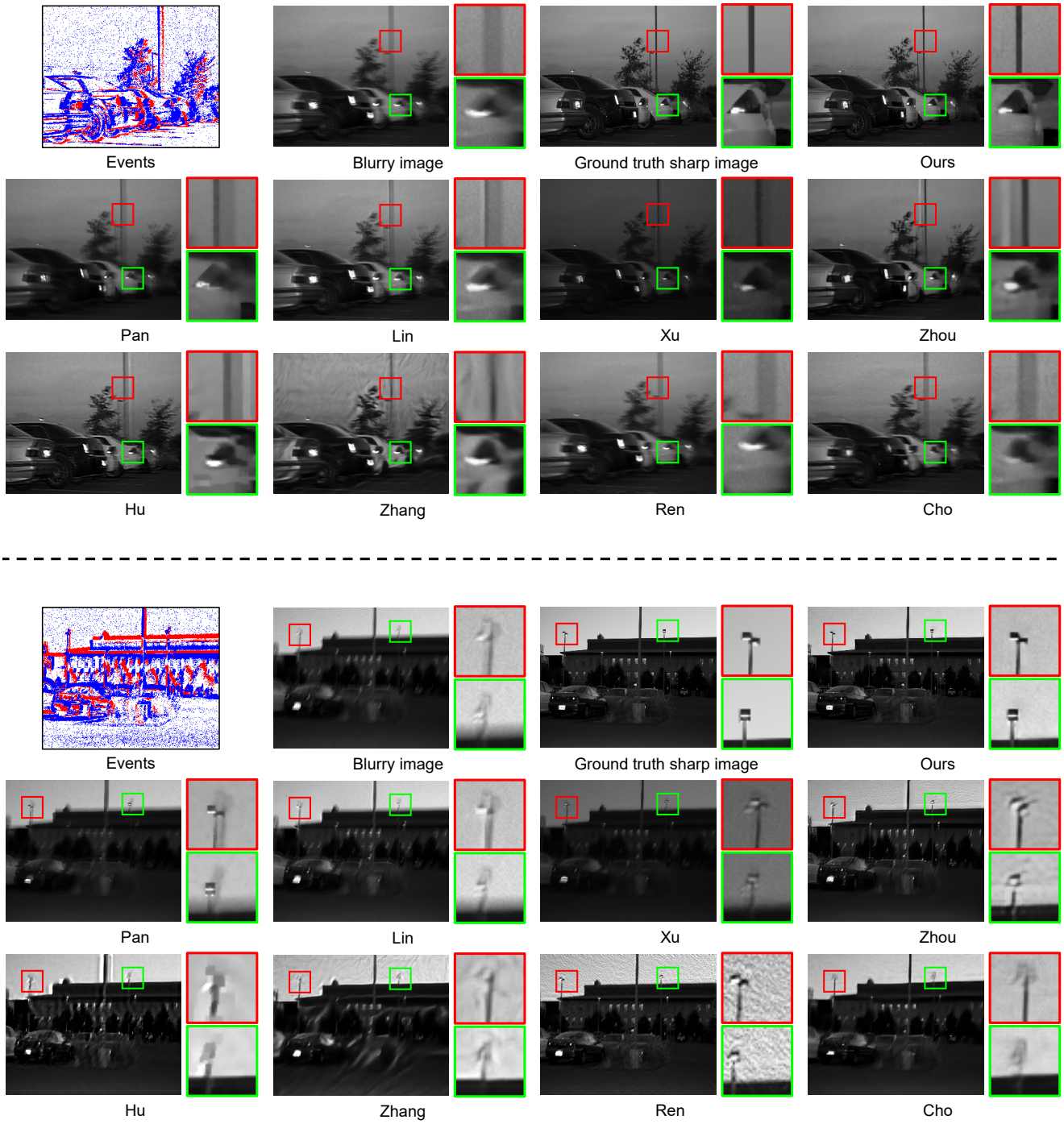

Fig. 9 Additional qualitative comparisons on synthetic data among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021), and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) (part 3).

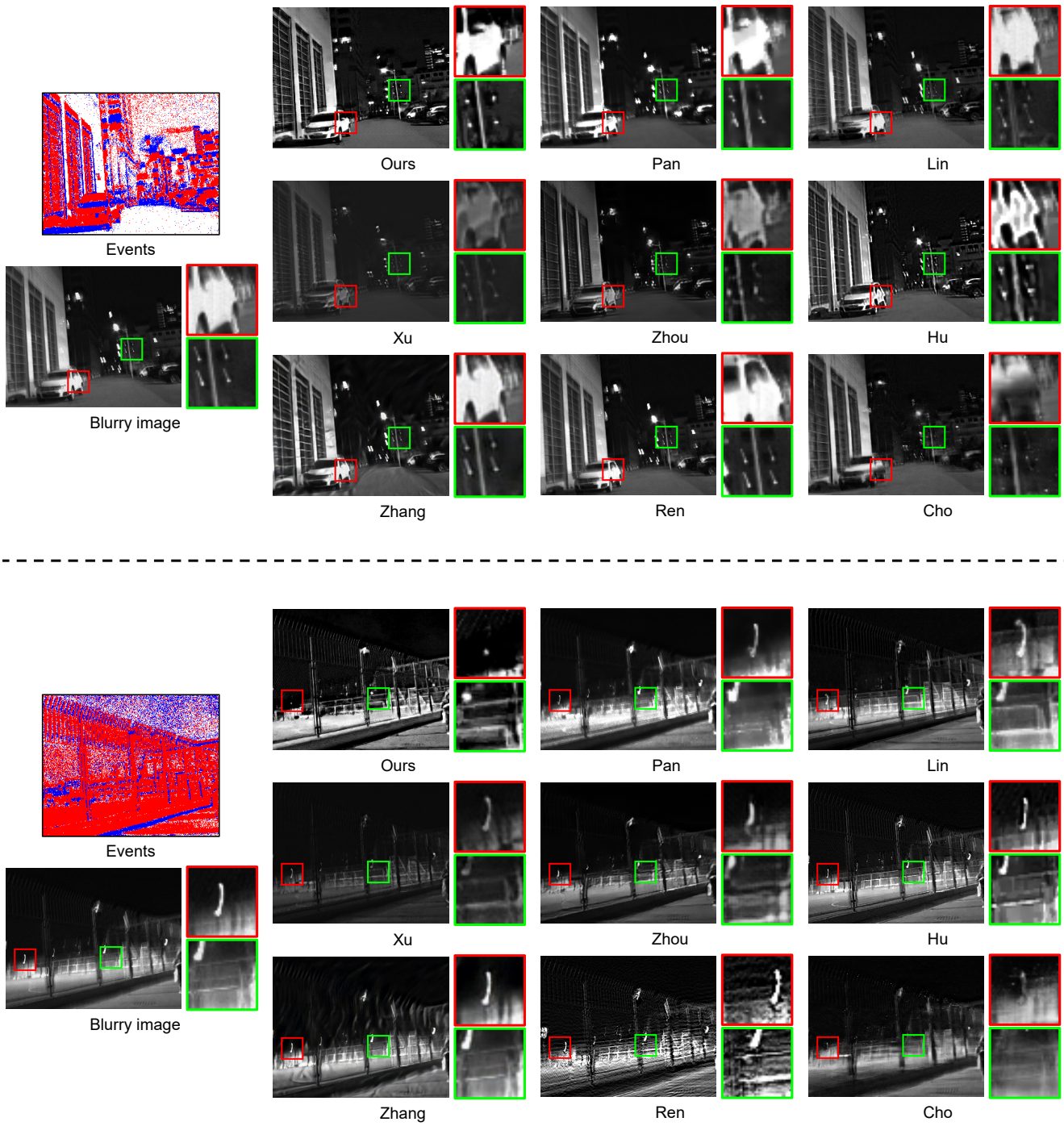

Fig. 10 Additional qualitative comparisons on real data captured by an event camera among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021), and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) (part 1).

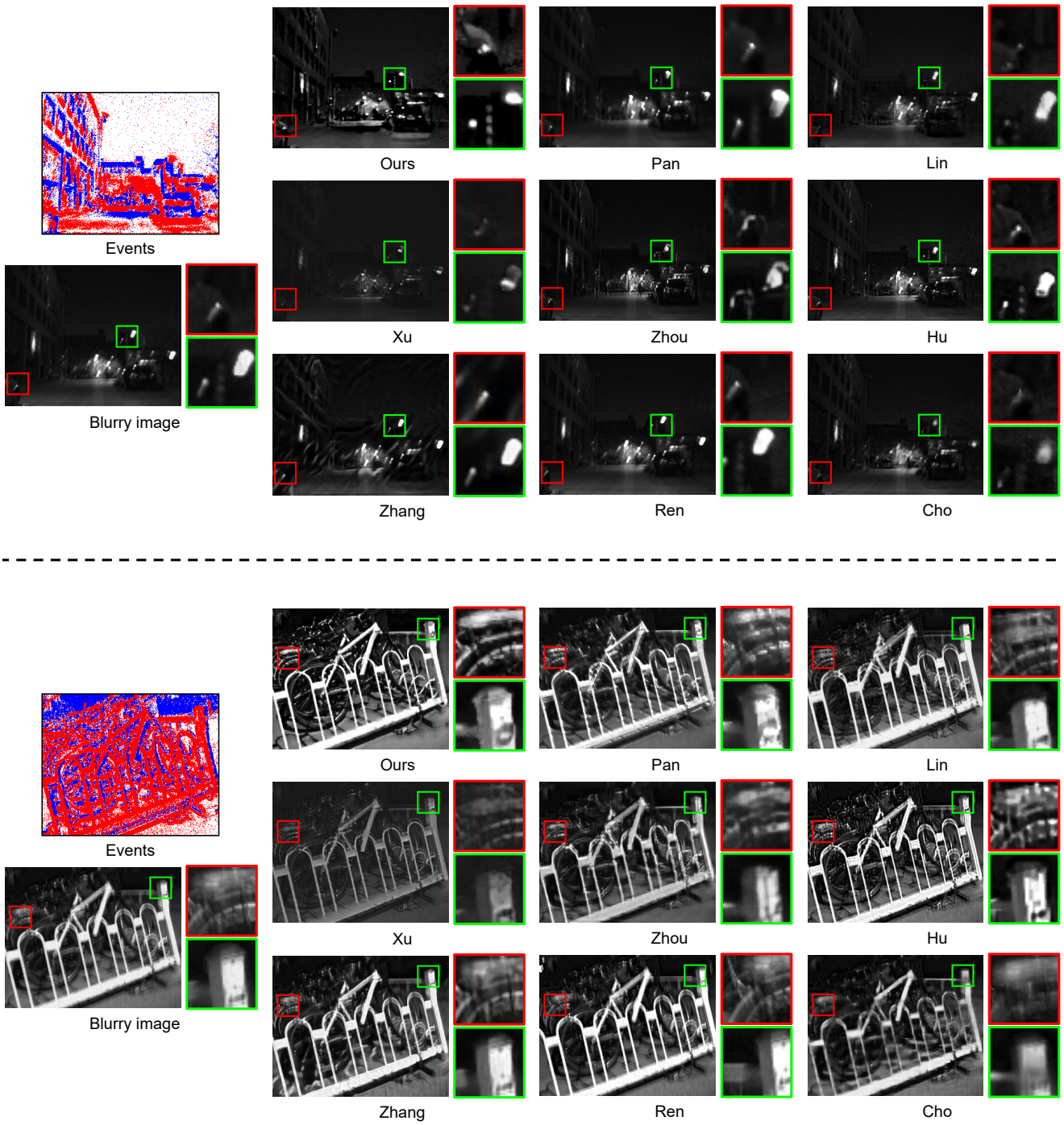

Fig. 11 Additional qualitative comparisons on real data captured by an event camera among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021), and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) (part 2).

Fig. 12 Additional qualitative comparisons on real data captured by an event camera among our method, four event-based deblurring methods (Pan et al, 2019; Lin et al, 2020; Xu et al, 2021; Zhou et al, 2021), and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021) (part 3).

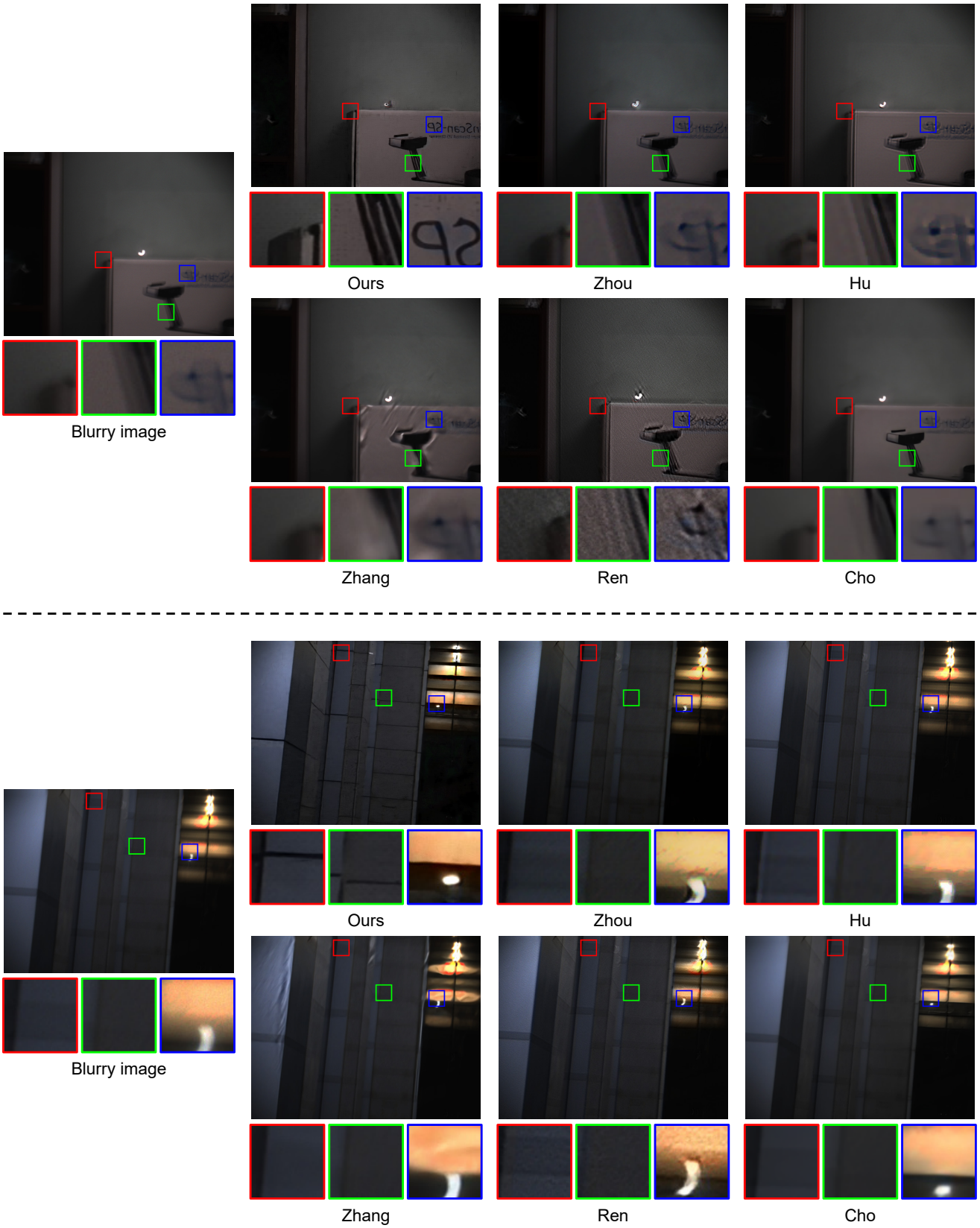

Fig. 13 Additional qualitative comparisons on real data captured by our RGB-DAVIS hybrid camera system among our method, our preliminary work (Zhou et al, 2021), and four image-based deblurring methods (Hu et al, 2018; Zhang et al, 2019; Ren et al, 2020; Cho et al, 2021).