Latency Correction for Event-guided Deblurring and Frame Interpolation

Yixin Yang[1,2,†] Jinxiu Liang[1,2] Bohan Yu[1,2] Yan Chen[3] Jimmy S. Ren[3] Boxin Shi[1,2,*]

[1] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University

[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University [3] SenseTime Research

{yangyixin93,cssherryliang,ybh1998,shiboxin}@pku.edu.cn yanchenace@gmail.com rensijie@sensetime.com

Abstract

Event cameras, with their high temporal resolution, dynamic range, and low power consumption, are particularly good at time-sensitive applications like deblurring and frame interpolation. However, their performance is hindered by latency variability, especially under low-light conditions and with fast-moving objects. This paper addresses the challenge of latency in event cameras — the temporal discrepancy between the actual occurrence of changes in the corresponding timestamp assigned by the sensor. Focusing on event-guided deblurring and frame interpolation tasks, we propose a latency correction method based on a parameterized latency model. To enable data-driven learning, we develop an event-based temporal fidelity to describe the sharpness of latent images reconstructed from events and the corresponding blurry images, and reformulate the event-based double integral model differentiable to latency. The proposed method is validated using synthetic and real-world datasets, demonstrating the benefits of latency correction for deblurring and interpolation across different lighting conditions.

1. Introduction

Emerging as a groundbreaking visual sensing paradigm, event cameras [1, 12, 13, 20, 22] provide a novel perspective to capturing and interpreting dynamic scenes. They have advantages of high temporal resolution (the discrete measurement resolution of the event detection time, in the order of microseconds [8]), high dynamic range (*e.g.*, 120 dB for DAVIS346), and low power consumption. They asynchronously record changes in logarithmic pixel intensity only when they exceed a predefined threshold. Their unique characteristics have led to impressive performance gains in various time-sensitive applications, including deblurring [5, 19, 24, 26, 32] and frame interpolation [19, 25, 30].

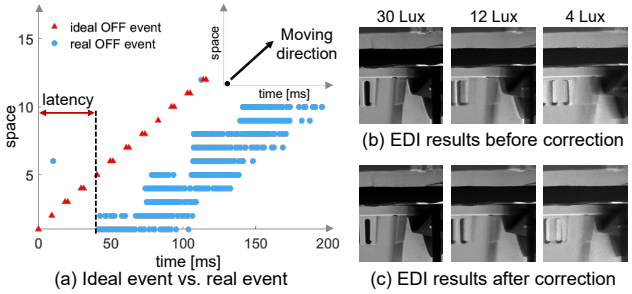


Figure 1. Illustration of latency in an event camera, and its influence on an event-guided deblurring algorithm. (a) The aerial view of ideal *vs.* real events triggered by a black point moving upwards on a white background, where the real events are captured in low-light conditions (below 0.1 Lux). (b) Deblurring results using the Event-based Double Integral (EDI) [19] method, guided by events captured under different illumination conditions. (c) Improved deblurring results from EDI [19] using event with latency corrected by the proposed method.

The ideal event model posits that event triggering is instantaneous upon the detection of intensity changes. However, real-world event timestamps deviate from this, exhibiting a temporal discrepancy between the actual moment of intensity change and the corresponding event timestamp. As claimed by the inventors of the event-based sensors, in well-lit environments above 1k Lux, event cameras can achieve latency as low as 3 μs [22]. However, in dimmer conditions for a few Lux, the latency can exceed 1 ms [22]. The challenge in modeling this temporal discrepancy arises from the sensor pixels' non-linear and variable response to illumination, which is further complicated by manufacturing imperfection, inherent electronic noise, and the unpredictable nature of real-world scenes [13, 15, 16, 21]. Therefore, the profile of the temporal discrepancy is spatial-varying, inconsistent, and noisy. All these complexities make algorithms for handling them scarce.

Without specifying the nature of the temporal discrepancy, we consider the collective temporal discrepancies in the timing of the camera's response to real-world changes, a phenomenon we term "*latency*", where the aforementioned sensor-level latency is one of the dominant components.

†This work is done during Yixin's internship at SenseTime.
*Corresponding author: shiboxin@pku.edu.cn

Despite the complexity of the latency formation model, latency presents a strong relation with illuminance. As illustrated in Figure 1 (a), compared with the ideal events generated instantaneously when the intensity changes, the real events suffer from latency in illuminance below 0.1 Lux. The latency presents a significant challenge for time-sensitive applications [19, 24, 25, 30], which assume ideal event triggering and struggle to restore sharp edges due to the severe latency, as demonstrated in Figure 1 (b) *vs.* (c).

In this paper, we focus on latency correction for event-guided imaging deblurring [19, 24, 28] and interpolation [19, 25, 31] in low-light and fast-moving scenes, where event cameras should play an important role. Based on the analysis and modeling of latency, we propose to predict latency in event cameras by conceptualizing latency as a variable function—potentially a constant for a specific scene or a polynomial function of the illuminance level, given a fixed event camera type—we endeavor to elucidate the relationship between latency and illuminance.

Specifically, we propose a latency correction method for captured events and blurry images in an event-image hybrid sensor (*e.g.*, DAVIS346). Inspired by the latency properties related to illuminance observed through controlled experiments, we characterize the latency-intensity relationship using a parameterized curve. To estimate the optimal latency, we introduce a novel objective to assess the sharpness profile of a latent image reconstructed from the latency-corrected events and the blurry image, which we term event-based temporal fidelity. Building upon them, we collect data under different scene illuminance and advocate for a self-supervised approach to latency correction. The EDI model [18, 19] is reformulated to be differentiable for latency to enable gradient-based optimization, therefore enhancing the performance of deblurring and interpolation.

The efficacy of the proposed method is validated using both synthetic and real data, demonstrating improvements in image quality for deblurring and interpolation under challenging lighting conditions where the latency in events becomes severe. The key contributions of our work include:

- the first data-driven latency correction method grounded in a differentiable reformulation of the event integral with respect to latency;
- the event-based temporal fidelity for measuring the sharpness profile of a latent image reconstructed by latency-corrected events and the corresponding blurry image; and
- a parameterized latency-intensity curve for modeling the relationship of latency in events and the intensity of the corresponding blurry image.

The proposed method provides an approach to understanding and mitigating the effects of latency, paving the way for more accurate and reliable event-based imaging in dynamic environments with challenging lighting conditions.

2. Related works

We investigate the influence of latency on imaging algorithms in event-image hybrid systems, particularly focusing on deblurring and interpolation tasks due to their sensitivity to timing in events.

Analysis of latency in event cameras. The triggering of events involves three subprocesses: photon reception, event trigger, and event readout. Notably, latency primarily arises during photon reception and event readout due to the electronic and physical design of these cameras [8]. Latency characteristics vary among different event camera models, ranging from minimal latency of 15 μs [13] to 4 μs [20] under 1k Lux conditions. Despite improvements in camera design, latency persists at nearly 1 ms under 1 Lux [1], posing challenges in low-light scenarios. Existing studies predominantly focus on sensor design aspects of event latency [1, 12, 13, 22] with a limited number examining latency under varying illuminance levels [6, 21]. In low-light conditions, latency induced by photoreceptor bandwidth becomes predominant [16], which is proportional to photocurrent (or illuminance) at low illuminance level, increasing monotonically with the illuminance level [2, 13, 15, 16, 21]. Consequently, the latency decreases as illuminance increases [7, 13, 21]. The photoreceptor circuit's low-pass characteristic [15, 16], has led to its modeling as a first- or second-order low-pass filter in various event simulators [7, 10]. While this provides a simplified relationship between latency and photocurrent, accurately determining photocurrent is often invalid for arbitrary imaging setups due to the complex imaging processes involving shuttering, apertures, *etc*. In order to save costs, most event cameras timestamp each event as they are being read out from the pixel array [8], which introduces readout latency that is significantly influenced by sensitivity and contrast [2, 8, 10, 15, 21]. The complexity of these factors contributing to latency in event cameras explains the scarcity of algorithms specifically designed to address this challenge for event-image hybrid systems.

Event guided imaging algorithms. Event cameras record changes during the exposure time of blurry images, which makes the problem of deblurring and interpolation more well-posed. Pan *et al*. [18, 19] introduced the EDI model for high-frame-rate video reconstruction from blurry videos using event data, applicable in both deblurring and interpolation. For deblurring, Jiang *et al*. [9] combined convolutional recurrent networks with directional event filtering, Shang *et al*. [23] utilized the nearest sharp frames and events for deblurring, Xu *et al*. [28] improved model generalization through self-supervised learning, Teng *et al*. [26] focused on event representations for image enhancement, and Sun *et al*. [24] developed an event-image fusion module. For joint deblurring and interpolation, Lin *et al*. [14] employed neu-

ral networks and dynamic filtering, while Wang *et al*. [27] and Zhang *et al*. [30, 31] integrated multiple techniques, including event-enhanced sparse learning and self-supervised frameworks. Sun *et al*. [25] proposed a bidirectional recurrent network for fusion tasks.

For these algorithms assuming ideal event triggering, the reconstruction of the latent image between or within exposure time heavily depends on the precision of timestamps in accompanying events. Latency presence can significantly impair algorithm performance. To the best of our knowledge, these has been no approach to address latency for event-guided deblurring and interpolation.

3. Proposed Method

In this paper, we aim to tackle the latency in events for event-guided deblurring and frame interpolation. Section 3.1 delves into the deviation of events with latency from the ideal event triggering mechanism and the event-based double integral model with latency. Based on this model, we propose a learning-based latency correction method for event-guided deblurring and frame interpolation. Section 3.2 introduces the objective for optimization, the event-based temporal fidelity. To allow gradient-based optimization, Section 3.3 reformulates event-based double integral to be differentiable respective to latency, where the latency is further parameterized into a polynomial of intensity on a per-pixel basis. Training details are shown in Section 3.4.

3.1. Analysis on the latency in events

Latency in event triggering. In an event camera, the photons hitting the photodetector of a pixel are transducted into an electrical signal and then compared against a predefined threshold to determine whether the change in intensity is substantial enough to trigger an "event". If this threshold is surpassed, the event is recorded with a high-resolution timestamp. Mathematically, the pixels of an event camera independently respond to variations in the continuous brightness signal $I_{\boldsymbol{p}}(t)$ at pixel position $\boldsymbol{p} = (p_x, p_y)^\top$ and time t. An event $e = (\boldsymbol{p}, t, \sigma)$ is initiated at pixel $\boldsymbol{p}$ and time t with polarity $\sigma \in \{+1, -1\}$ indicating `ON` or `OFF` events when the logarithmic change of $I_{\boldsymbol{p}}$ exceeds a dispatched threshold θ since the last event triggered at the pixel $\boldsymbol{p}$ and time t_{ref}:

$$\|\log I_{\boldsymbol{p}}(t) - \log I_{\boldsymbol{p}}(t_{\text{ref}})\| \geq \theta. \tag{1}$$

In tasks involving image reconstruction where per-pixel intensity is needed, the logarithmic intensity changes for continuous time neighboring recorded timestamps are often approximated as an equality involving the threshold value θ and polarity σ with the quantization errors ignored.

The ideal model posits that event triggering is instantaneous upon the detection of intensity changes. However, real-world event timestamps deviate from this ideal, exhibiting a temporal *latency* l between the actual change in intensity and the event's triggering. The presence of latency exacerbates the approximation error of the intensity changes (*e.g*., Eq. (2) in the following), particularly affecting time-sensitive tasks like deblurring and frame interpolation. In environments with controlled lighting and motion, empirical methods have been employed to study and characterize latency in event cameras, such as using a second-order low-pass filter [13, 20] with parameters that are influenced by illumination conditions [6, 7, 13]. In the unpredictable flux of real-world settings, latency becomes markedly more unpredictable and defies precise modeling. The complexity arises from the non-linear and non-uniform factors. For instance, pixel responses across the sensor array exhibit variability due to manufacturing differences, and the light received by each pixel fluctuates rapidly with movement in the scene or changes in illumination conditions. Moreover, the existence of noise exacerbates the difficulty.

Event-based double integral model with latency. An event-image hybrid sensor (*e.g*., DAVIS346) is able to capture pixel-wise aligned events and images, synergistically harnessing the strengths of both modalities. This integration involves vibrant color representation characteristic of frame-based sensors, coupled with the high temporal resolution and dynamic range offered by event cameras. The mathematical foundation for the interplay between events and images in such a system is encapsulated by the EDI model, which is originally proposed for applications in deblurring and frame interpolation [18, 19].

The proportional change for images at time τ and time t can be measured by the integral of the N events $\mathbb{E} = \{e_k\}_{k=1}^{N}$ triggered from τ to t as:

$$I_{\boldsymbol{p}}(t) = I_{\boldsymbol{p}}(\tau) \exp\left(\theta \int_{\tau}^{t} E_{\boldsymbol{p}}(s) ds\right), \tag{2}$$

where $E_{\boldsymbol{p}}(s)$ is a function of continuous time:

$$E_{\boldsymbol{p}}(s) = \sigma_k \delta_{t_k}(s), \quad s \in (t_{k-1}, t_k], \tag{3}$$

whenever there is an event $e_k = (\boldsymbol{p}_k, t_k, \sigma_k)$ triggered and $\delta_{t_k}(s)$ is an impulse function at time t_k with unit integral. A blurry image $B_{\boldsymbol{p}}(\tau)$ with exposure duration $t \in [\tau - \Delta\tau, \tau + \Delta\tau]$ can be represented as the average of the latent image $I_{\boldsymbol{p}}(t)$ over the exposure duration [18], which can be formulated as:

$$B_{\boldsymbol{p}}(\tau) = \frac{1}{2\Delta\tau} I_{\boldsymbol{p}}(\tau) R_{\boldsymbol{p}}(\tau), \tag{4}$$

where $R_{\boldsymbol{p}}(\tau)$ is defined as:

$$R_{\boldsymbol{p}}(\tau) = \int_{\tau-\Delta\tau}^{\tau+\Delta\tau} \exp\left(\theta \int_{\tau}^{t} E_{\boldsymbol{p}}(s) ds\right) dt. \tag{5}$$

It indicates that the latent image can be reconstructed by:

$$I_{\boldsymbol{p}}(\tau) = 2\Delta\tau \frac{B_{\boldsymbol{p}}(\tau)}{R_{\boldsymbol{p}}(\tau)}. \tag{6}$$

We introduce a *spatially-varying latency* $l_{\boldsymbol{p}}(\tau)$ during exposure time $[\tau-\Delta\tau, \tau+\Delta\tau]$ into the EDI map $R_{\boldsymbol{p}}(\tau)$ in Eq. (5):

$$\widehat{R}_{\boldsymbol{p}}(\tau) = \int_{\tau-\Delta\tau}^{\tau+\Delta\tau} \exp\left(\theta \int_{\tau-l_{\boldsymbol{p}}(\tau)}^{t-l_{\boldsymbol{p}}(\tau)} E_{\boldsymbol{p}}(s)ds\right) dt. \quad (7)$$

In practice, the event as well as the exposure time for each latent image are discrete. Suppose the blurry image B compromises N_I latent images, and N_j events at pixel $\boldsymbol{p}$ are triggered during exposure time $[\tau - l_{\boldsymbol{p}}(\tau), t - l_{\boldsymbol{p}}(\tau)]$, the discrete formulation of Eq. (7) is:

$$\widehat{R}_{\boldsymbol{p}}(\tau) = \sum_{j=1}^{N_I} \exp\left(\theta \sum_{k=1}^{N_j} \sigma_k\right). \quad (8)$$

The latent image reconstructed from blurry image B and event image E can then be obtained by:

$$\widehat{I}_{\boldsymbol{p}}(\tau) = N_I \frac{B_{\boldsymbol{p}}(\tau)}{\widehat{R}_{\boldsymbol{p}}(\tau)}. \quad (9)$$

$\boldsymbol{p}$ and τ will be omitted in the following for simplicity.

3.2. Event-based temporal fidelity

We consider events $\mathbb{E} = \{e_k\}_{k=1}^{N}$ captured in the exposure duration of the blurry image B. To achieve the goal of latency correction, a no-reference-based objective should be defined since the ground truth is hard to collect. To this end, we propose event-based temporal fidelity (ETF) $\mathcal{L}_{\text{ETF}}$ to quantify the sharpness profile of latent images I reconstructed from events with latency $l_{\boldsymbol{p}}$ corrected and their corresponding blurry image B, as shown in Figure 2 (a).

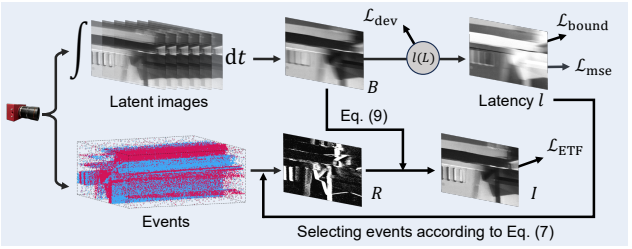
Intuitively, $\mathcal{L}_{\text{ETF}}$ should have smaller values as the estimated latency approaches its optimal values. Sharp images have more high-frequency contents compared with blurry images, which can be assessed by measuring the magnitude of its derivatives [3]. As shown in Figure 1 (d), the latent image results of EDI obtained from events with large latency appear with less sharpness and thicker edges. This motivates us to optimize the latency $l_{\boldsymbol{p}}$ by enforcing regularizations of the sharpness profile of the latent image I reconstructed from the EDI model in Eq. (9). Specifically, we define the average edge thickness of image gradients $\nabla I_{\boldsymbol{p}} = (\nabla_x I_{\boldsymbol{p}}, \nabla_y I_{\boldsymbol{p}})$ as the number of pixels with values greater than a parameter λ,

$$s_\lambda(I) = \sum_{\boldsymbol{p}} H(\|\nabla I_{\boldsymbol{p}}\|_2 - \lambda), \quad (10)$$

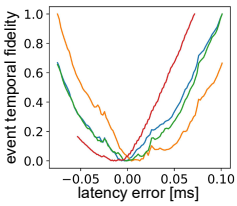
where $H(\cdot)$ is the Heaviside function.

Integrating the right-side of the equation over λ with a decreasing function ρ, we can obtain an λ-irrelevant representation of $s(I)$:

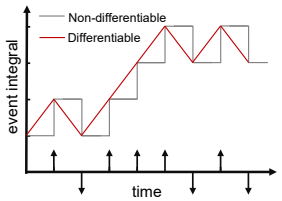
$$s(I) = \int_0^\infty \sum_{\boldsymbol{p}} \rho(\lambda) H\left(\|\nabla I_{\boldsymbol{p}}\|_2 - \lambda\right) \mathrm{d}\lambda. \quad (11)$$



(a) The proposed latency correction method



(b) $\mathcal{L}_{\text{ETF}}$ vs. constant l



(c) Differentiable event integral

Figure 2. Illustration of the proposed latency correction method. (a) The pipeline of the proposed method, which takes synchronized blurry image B and events as input, and output the latency l by polynomial function $l(L)$. (b) The value of the proposed event temporal fidelity (ETF) $\mathcal{L}_{\text{ETF}}$ decreases as the estimated latency approaches ground-truth latency, which validates the efficacy of ETF as an objective. (c) Latency-differentiable reformulation of event integration, enabling gradient-based optimization.

We adopt the function $\rho(\lambda) = e^{-\lambda}$ in our experiments[1]. To pay more attention to the smoothness area, we multiply the image gradients $\nabla I_{\boldsymbol{p}}$ with an amplifier ϵ.

It is noted that, for the robustness to scenes with possible thick edges on the image, ϵ can not be set too small, which brings challenges for detecting high-frequency details. To compensate for this, we define another term about the contrast of the latent image I. Motivated by the fact that sharper images tend to have large contrast , we evaluate the contrast as the average local variation with local window $\mathcal{N}(\boldsymbol{p})$ centered at pixel $\boldsymbol{p}$ [17]:

$$h(I) = \sum_{\boldsymbol{p}} \frac{1}{|\mathcal{N}_{\boldsymbol{p}}|} \sum_{\boldsymbol{q}\in\mathcal{N}_{\boldsymbol{p}}} (I_{\boldsymbol{q}} - \mu_I(\mathcal{N}_{\boldsymbol{p}}))^2, \quad (12)$$

where $\mu_I(\mathcal{N}_{\boldsymbol{p}})$ is the mean of the pixel value inside the window $\mathcal{N}_{\boldsymbol{p}}$. The reason we use a local variation rather than a global one is that the blurry effects caused by motion are not always globally consistent. Maximizing the local variation can emphasize the local sharpness and reduce the blur for each local window. In practice, we adopt local standard deviation $\sqrt{h(I)}$ in our experiments.

For consecutive frames in a video, we can assume that their overall brightness level are similar, and thus the latency values are similar. Therefore, we define an indicator of temporal consistency by assuming Gaussian-distributed latency across consecutive frames. Given the latency of the last adjacent frame is l', the indicator defined on the latency

[1]Please see the supplementary material for more details about $s(I)$.

l of the current frame is formulated as:

$$g(l) = \frac{1}{2\pi} \exp\left(-\frac{(l-l')^2}{2}\right). \tag{13}$$

Overall, $\mathcal{L}_{\text{ETF}}$ to determine optimal latency is defined as:

$$\mathcal{L}_{\text{ETF}}(I, l) = \frac{s(\epsilon I)}{\sqrt{h(I)g(l)}}. \tag{14}$$

The efficacy of the proposed $\mathcal{L}_{\text{ETF}}$ is validated in Figure 2 (b), which shows that its value decreases as the latency estimation error decreases.

3.3. Learning-based latency correction

Based on the objective defined by $\mathcal{L}_{\text{ETF}}$, we introduce a self-supervised approach founded on a parametrized form of latency, and a reformulated event-based double integral model that is differentiable with respect to latency.

Latency-illuminance curve. The physical factors that have influences on the latency of events cannot be easily measured for the arbitrary situation. Fortunately, for deblurring and frame interpolation algorithms using event-image hybrid sensors such as [18, 19], the intensity of the image can indirectly indicate the illuminance levels in certain regions. Inspired by the previous findings about the relationship between latency and illuminance levels [7, 16, 21], we establish a latency-intensity relationship using a parameterized curve, which is validated in controlled environments (see illustrations in Figure 3 and discussions in Section 4.1 for more details, which shows that the latency increases as the illuminance decreases). To elucidate this relationship, we model the latency-illuminance curve as a K-th order polynomial ψ of the illuminance L at pixel $\boldsymbol{p}$:

$$\psi(L) = \sum_{i=0}^{K} a_i \cdot f(L)^i, \tag{15}$$

where $\{a_i\}_{i=0}^{K}$ are the parameters, $f(\cdot)$ is a fixed function[2] to map the illuminance L into monotone decrement for consideration of easier optimization. The curve in Eq. (15) can effetively be a guidance for regularizing the estimation of spatially varying latency. In experiments, the illuminance L is unavailable. We adopt the blurry image B to approximate illuminance L to estimate the spatially-varying latency, resulting in $l_{\boldsymbol{p}} = \psi(\frac{B_{\boldsymbol{p}}}{2\Delta\tau})$.

Latency-differentiable event integral. It is noted that the latency $l_{\boldsymbol{p}}$ derived from Eq. (9) is an index-like parameter, which has no derivatives. To allow gradient-based optimization, we establish a differentiable relationship between latency in events and specific indicators as shown in Figure 2 (c). By reformulating the event integral to be latency-differentiable, we derive a gradient chain linking $\{a_i\}_{i=0}^{K}$ to $\mathcal{L}_{\text{ETF}}$ via the EDI model with latency.

[2]Please see the supplementary material for more details about $f(\cdot)$.

In contrast to Eq. (3) where events $e_k = (\boldsymbol{p}_k, t_k, \sigma_k)$ are represented by an impulse function, we employ a piece-wise constant function:

$$E_{\boldsymbol{p}}^{(k)}(s) = \frac{\sigma_k}{t_k - t_{k-1}}, \quad s \in (t_{k-1}, t_k], \tag{16}$$

where the changes σ_k is equally split into the several duration. Then integration of events in duration $(t_{k-1}, t_k]$ becomes:

$$\int E_{\boldsymbol{p}}^{(k)}(s)ds = \sigma_k, \quad s \in (t_{k-1}, t_k]. \tag{17}$$

For simplicity, we denote $E_{\boldsymbol{p}}^{(k)}(s)$ as $E_{\boldsymbol{p}}^{(k)}$. Consider the N events $\{e_k\}_{k=1}^{N}$ triggered at pixel $\boldsymbol{p}$ during exposure time $[\tau - l_{\boldsymbol{p}}, t - l_{\boldsymbol{p}}]$ where $t_0 < \tau - l_{\boldsymbol{p}} < t_1$ and $t_{N-1} < t - l_{\boldsymbol{p}} < t_n$, we can rewrite the inner integral part in Eq. (7) into:

$$\begin{aligned}
&\int_{\tau - l_{\boldsymbol{p}}}^{t - l_{\boldsymbol{p}}} E_{\boldsymbol{p}}(s)ds \\
=&\int_{\tau - l_{\boldsymbol{p}}}^{t_1} E_{\boldsymbol{p}}^{(1)}ds + \sum_{k=2}^{N-1} \int_{t_{k-1}}^{t_k} E_{\boldsymbol{p}}^{(k)}ds + \int_{t_{N-1}}^{t - l_{\boldsymbol{p}}} E_{\boldsymbol{p}}^{(N)}ds \\
=&(t_1 - l_{\boldsymbol{p}} - \tau)\sigma_1 + \sum_{k=2}^{N-1} \sigma_k + (t - l_{\boldsymbol{p}} - t_{N-1})\sigma_n,
\end{aligned} \tag{18}$$

which is differentiable in the neighborhood of $l_{\boldsymbol{p}}$ with fixed N. It is noted that N will change when $\tau - l_{\boldsymbol{p}} < t_0$ or $t - l_{\boldsymbol{p}} \geq t_N$, in which discontinuous changes on event triggering time will occur. While this might lead to a local minimum solution in optimization for a specific sample, it can be alleviated when optimizing over different samples.

Spatially-varying latency correction. With the differentiable form of event integral for parameters $\{a_i\}_{i=0}^{n}$, we can now perform gradient descent in the dataset by defining a loss function $\mathcal{L}_{\text{total}}$. Firstly, $\mathcal{L}_{\text{ETF}}$ defined in Eq. (14) is used as a term of the loss function. Heuristically, most of the latency $l_{\boldsymbol{p}}$ should lie in the range from 0 to the exposure time of the last blurry image, which can be optimized by the reformulated EDI. To ensure the latency does not exceed the previous exposure time so that the derivative can be calculated, a clipping operation is applied on $l_{\boldsymbol{p}}$. A boundary loss is also used to limit the range of latency:

$$\mathcal{L}_{\text{bound}} = \sum_{l_{\boldsymbol{p}} > 0} l_{\boldsymbol{p}} + \sum_{l_{\boldsymbol{p}} < \tau - \tau' - \Delta\tau} -l_{\boldsymbol{p}}. \tag{19}$$

To deal with $l_{\boldsymbol{p}}$ exceeds the range of $[0, \tau - \tau' - \Delta\tau]$ and disables derivative computation due to gradients with too large magnitudes, we additionally optimize an intermediate *constant latency* l_{const} shared across the sensor array during the exposure time of the blurry image. It has a relatively small solution space and can be solved efficiently by searching. l_{const} provides an effective estimation of the overall latency (whose effectiveness is validated in the ex-

periment results on synthetic data in Section 4.2). Thus, it is being adopted in the term $\mathcal{L}_{\text{mse}}$ to minimize the distance between the current estimated latency and l_{const}. l' in Eq. (13) is replaced by l_{const} to regularize using the frame's constant latency, rather than adjacent frames'.

From the latency-illuminance curve estimated previously, we can observe that the curve should be monotonic. Thus we enforce that the function $l_{\boldsymbol{p}}$ is monotonically decreasing as the intensity increases. To apply the monotonicity constraint, we introduce the following constraint on the derivative of a function $l_{\boldsymbol{p}}$:

$$\mathcal{L}_{\text{dev}} = \sum \max(f'(\frac{B}{2\Delta\tau}) \sum_{i=0}^{K} i \cdot a_i \cdot f(\frac{B}{2\Delta\tau})^{i-1}, 0), \quad (20)$$

where B contains different values on each pixel to ensure monotonically decreasing on all the data samples.

Overall, the loss function to be minimized is defined as:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{ETF}} + \beta\mathcal{L}_{\text{bound}} + \eta\mathcal{L}_{\text{mse}} + \zeta\mathcal{L}_{\text{dev}}, \quad (21)$$

where $\alpha, \beta, \eta, \zeta$ are hyper-parameters for balancing different terms. Empirically, we set $\alpha = 1, \beta = 1, \eta = 20, \zeta = 0.1$ in our experiments.

3.4. Training details

Dataset preparation. For latency correction training, we utilized a DAVIS346 event camera [1] to obtain blurry images and corresponding events. Our real-scene dataset comprises 9 videos, where 100 frames and their neighboring events from each video are selected for training.

Other implementation details. To simplify the implementation, we stack events into 500 bins to maintain high temporal resolution. Our model is implemented on the Pytorch framework and runs on an NVIDIA GeForce RTX 3090 GPU. We use ADAM [11] with default parameter setting for optimization. The batch size is set as 1. The learning rate was fixed to 10^{-5} in the earlier 2 epochs and reduced to 0 in the last 3 epochs by the linear decay strategy. Constant initialization to zero for the non-latency assumption is used for initialization.

4. Experiments

4.1. Latency estimation in controlled environments

Experiments in controlled settings measure actual latency and illuminance. Our setup, shown in Figure 3 (a), includes a light source, an Arduino board, a power source, a photometer, and an event camera (DVS346Mono) with its lens removed to prevent light attenuation. The Arduino board was programmed to switch off the light at a known time t_{ideal}, and the subsequent `OFF` events were assumed to be triggered by this change. The latency of the i-th triggered event was calculated as the difference between t_{ideal} and the time $t^{(i)}_{\text{actual}}$. Illuminance is modulated by adjusting light voltage and measured alongside events by a photometer (with the lowest detectable value being 0.1 Lux). This process was repeated to mitigate noise effects (about 80 times in our experiments). The latency level of a given illuminance was calculated by the event triggering rate as shown in Fig. 3 (b), which is approximated by the median latency when the first 1% `OFF` events among all the samples are triggered for each process. By varying the illuminance, we empirically find that the latency increases as the illuminance decreases, as shown in Fig. 3 (c). We use the proposed parameterized curve to approximate the relationship between latency and illuminance, whose results are shown in Fig. 3 (d). It demonstrates an accurate curve fitting with $K = 5$. Due to the difficulty in illuminance measuring, the estimated parameter cannot be directly applied in deblurring and interpolation for arbitrary scenes.

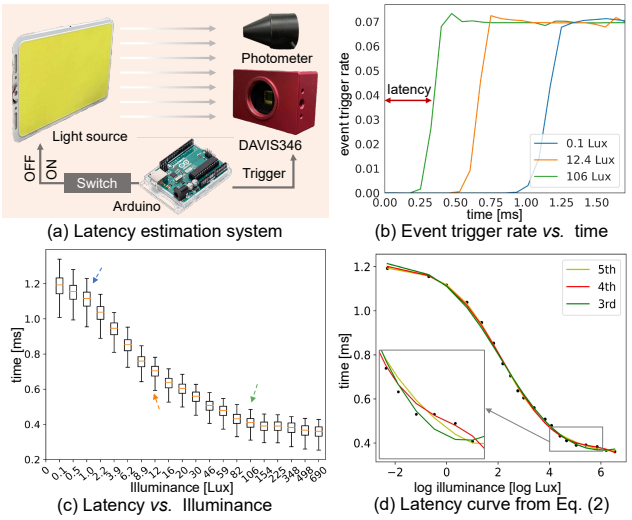


Figure 3. Estimation of latency-illuminance curve. (a) A system designed to investigate the relationship between illuminance levels and latency, approximated using an K-th order polynomial. (b) The comparison of event triggering rate under three varied lighting conditions. It is approximated by the first 1% `OFF` events in our experiments. (c) Estimated latency under controlled lighting, highlighting increased latency with diminishing light. (d) Results of the polynomial approximation.

4.2. Evaluation on synthetic data

Compared methods. As the proposed method serves as a preprocessing step for latency correction of events, we evaluate the proposed methods based on two downstream tasks, that is, deblurring, and interpolation. Specifically, we test our method for deblurring algorithms including EDI [19], EFNet [24], and GEM [31], and for interpolation algorithms including EDI [19] and REFID [25]. We directly use the released pretrain model for comparison.

Dataset. To qualitatively evaluate downstream tasks, we

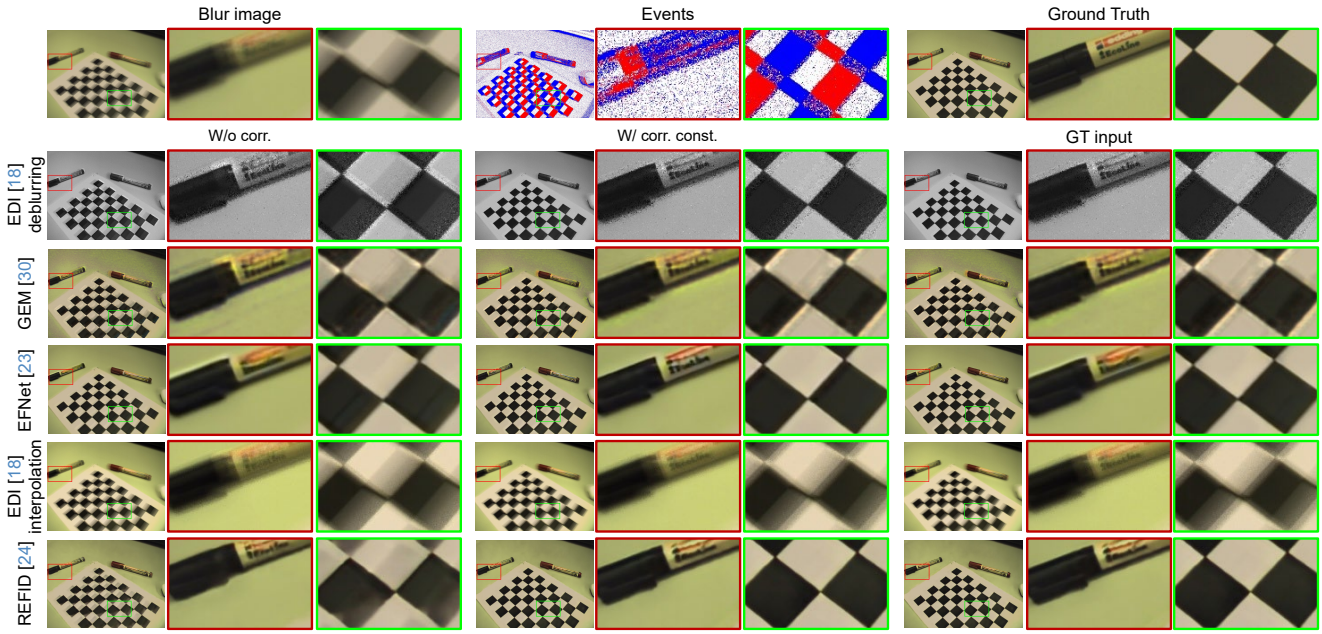


Figure 4. Qualitative comparisons on synthetic data. We correct the latency of events by the constant latency correction ("W/ corr. const.") method and compared it with the ground truth events (GT input). All the methods are tested with the same input.

Table 1. Quantitative evaluation on synthetic data. ↑ (↓) means higher (lower) is better. The "Diff" shows the percentage improvement from "W/o corr." to "W/ corr. const." compared with "GT input".

		PSNR↑	SSIM↑	LPIPS↓
EDI [19] Deblurring	GT input	28.489	0.839	0.489
	W/o corr.	26.001	0.810	0.500
	W/ corr. const.	**27.471**	**0.828**	**0.488**
	Diff. (%)	+5.15	+2.15	-2.45
EFNet [24]	GT input	33.761	0.972	0.105
	W/o corr.	33.330	0.971	0.107
	W/ corr. const.	**33.778**	0.971	**0.106**
	Diff. (%)	+0.13	0	-0.95
GEM [31]	GT input	30.611	0.963	0.183
	W/o corr.	27.714	0.934	0.198
	W/ corr. const.	**30.372**	**0.958**	**0.185**
	Diff. (%)	+8.68	+2.49	-7.10
EDI [19] Interpolation	GT input	22.293	0.840	0.330
	W/o corr.	22.125	0.834	0.331
	W/ corr. const.	**22.265**	**0.838**	0.331
	Diff. (%)	+0.63	+0.47	0
REFID [25]	GT input	34.522	0.974	0.109
	W/o corr.	28.244	0.927	0.142
	W/ corr. const.	**34.217**	**0.970**	**0.111**
	Diff. (%)	+17.30	+4.41	-28.44

synthesize events with constant latency l_τ for each pair of events and its corresponding blurry image. In particular, we utilize the event provided in the HighREV dataset [25] as ideal events, and then we add a constant latency to the event-image pairs for different event sequences. We employ the training data in the HighREV dataset which contains 1615 frames from 19 videos for testing to provide more convincing results. The spatially-varying latency correction is highly related to the properties of event cameras and depends on different scene illuminances, which can not be faithfully synthetic by event simulators [7, 10]. Therefore, we only provide results of the constant latency estimated by the proposed methods for quantitative comparisons.

Metrics. To show the corrected performance, all the tested methods are evaluated with four kinds of input, events without latency (terms "GT input"), events with latency which is not corrected by our method (terms "W/o corr."), the events corrected by constant latency l_{const} (terms "W/ corr. const."). We adopt three commonly used metrics to measure the performance of deblurring and interpolation, peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and the perceptual error with learned perceptual image patch similarity (LPIPS) [29], respectively.

Results. The quantitative evaluation results are shown in Table 1. We add the "Diff." row to show the improvement of the proposed method. It shows that the proposed method can improve the performance of deblurring and interpolation by event latency correction. For deep learning methods like EFNet [24], they implicitly considered latency effects in training data generation, which makes them more robust to latency in events. However, their performance can also be improved with the latency correction. The qualitative evaluation results are shown in Figure 4, which show that the proposed method can explicitly reduce the artifacts brought by latency, and it performs more closely to the ground truth events (GT input). Specifically, the latency correction can

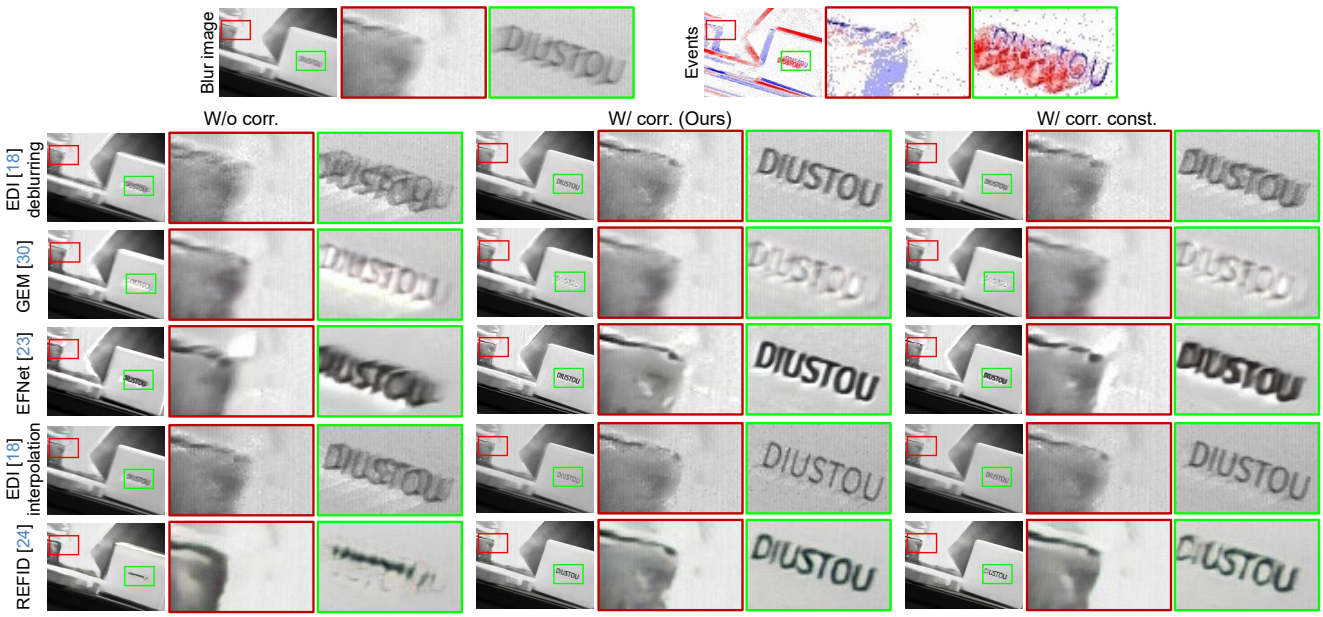


Figure 5. The qualitative comparison on real data. We correct the latency of events by constant latency correction ("W/ corr. Const.") and spatially-varying latency correction ("W/ corr."). All the methods are tested with the same input.

reduce the light strips brought by previous edges, in which the events are triggered with latency. The non-learning interpolation algorithm EDI [19] is more strict with the time accuracy and the noise in events, which makes it still blurry even with the GT input.

4.3. Evaluation on real data

In order to reveal the performance of the proposed method, we evaluate our method on real data captured by DAVIS346Mono. Because the ground truth sharp images are unavailable, we can only qualitatively evaluate the proposed method on real data. The results on real data[2] are shown in the left two columns of Figure 5. It shows that both the proposed latency correction methods can reduce the artifacts and blurry effects. Processed by the proposed latency correction methods, the number of inaccuracy events was explicitly reduced to help the deblurring and interpolation algorithms to produce natural and sharp edges.

4.4. Ablation study

We compare the performance of spatially-varying latency correction with constant latency correction on real data as shown in the right two columns in Figure 5. It shows that events corrected by the spatially-varying method have more natural views than constant latency correction, which can also provide sharper results[3]. Benefiting from the data captured under different illumination conditions, the spatially varying latency correction can also provide more accurate latency estimation for specific scenes. Along with the different illumination conditions, it can also address the different latency properties for different pixels.

[3]More comparisons are provided in the supplementary material.

5. Conclusion

In this paper, we present the first method for latency correction in improving event-guided deblurring and interpolation tasks. Our approach introduces event-based temporal fidelity to evaluate the sharpness of latent images reconstructed from latency-corrected events and their corresponding blurry images. This objective guides the optimization of parameters in a parametrized latency model through gradient-based optimization, facilitated by a reformulated, latency-differentiable, event-based double integral model. The efficacy of our methods is demonstrated using both synthetic and real-world data sets.

Limitations. Although with potential for broad applicability across different event cameras, the present inplementation of the proposed method is specifically tailored for DAVIS346Mono cameras due to that it is the only event camera we have with on-chip synchronization and our current method requires perfect temporal synchronization between the image and events. Moreover, due to challenges in measuring illuminance and curve variability, the practicality of estimated parameters of the latency-illuminance curve varies across sensors. The curves for more sensors, while significant, has not been explored in this paper.

Acknowledgement

This work was supported by National Science and Technology Major Project (Grant No. 2021ZD0109803), Beijing Natural Science Foundation (Grant No. L233024), National Natural Science Foundation of China (Grant No. 62302019, 62088102, 62136001), and SenseTime Collaborative Research Grant.

References

[1] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240 × 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 2014. 1, 2, 6

[2] Tobi Delbruck and Carver A Mead. Adaptive photoreceptor with wide dynamic range. In *Proc. of IEEE International Symposium on Circuits and Systems-ISCAS*, pages 339–342, 1994. 2

[3] Guillermo Gallego, Mathias Gehrig, and Davide Scaramuzza. Focus Is All You Need: Loss Functions for Event-Based Vision. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 4

[4] Michael D. Grossberg and Shree K. Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1272–1282, 2004. 1

[5] Chen Haoyu, Teng Minggui, Shi Boxin, Wang YIzhou, and Huang Tiejun. Learning to Deblur and Generate High frame rate video with an event camera. *arXiv preprint arXiv:2003.00847*, 2020. 1

[6] Ondřej Holešovský, Radoslav Škoviera, Václav Hlaváč, and Roman Vítek. Experimental Comparison between Event and Global Shutter Cameras. *Sensors*, 21(4), 2021. 2, 3

[7] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic DVS events. In *Proc. of Computer Vision and Pattern Recognition Workshops*, 2021. 2, 3, 5, 7

[8] inivation. Understanding the Performance of Neuromorphic Event-based Vision Sensors. https://inivation.com/wp-content/uploads/2020/05/White-Paper-May-2020.pdf, 2020. 1, 2

[9] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning Event-Based Motion Deblurring. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[10] Damien Joubert, Alexandre Marcireau, Nic Ralph, Andrew Jolley, André van Schaik, and Gregory Cohen. Event camera simulator improvements via characterized parameters. *Frontiers in Neuroscience*, 15:702765, 2021. 2, 7

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[12] Juan Antonio Leñero-Bardallo, Teresa Serrano-Gotarredona, and Bernabé Linares-Barranco. A 3.6 μs Latency Asynchronous Frame-Free Event-Driven Dynamic-Vision-Sensor. *IEEE Journal of Solid-State Circuits*, 46(6):1443–1455, 2011. 1, 2

[13] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128× 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008. 1, 2, 3

[14] Songnan Lin, Jiawei Zhang, Jinshan Pan, Zhe Jiang, Dongqing Zou, Yongtian Wang, Jing Chen, and Jimmy Ren. Learning event-driven video deblurring and interpolation. In *Proc. of European Conference on Computer Vision*, 2020. 2

[15] Brian McReynolds, Rui Graca, and Tobi Delbruck. Experimental methods to predict dynamic vision sensor event camera performance. *Optical Engineering*, 61(7):074103–074103, 2022. 1, 2

[16] Brian J Mcreynolds, Rui Graca, Daniel O'Keefe, Rachel Oliver, Richard Balthazor, Nathaniel George, and Matthew McHarg. Modeling and decoding event-based sensor lightning response. In *Proc. of Unconventional Imaging, Sensing, and Adaptive Optics 2023*, 2023. 1, 2, 5

[17] Patrenahalli M. Narendra and Robert C. Fitch. Real-time Adaptive contrast enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):655–661, 1981. 4

[18] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2, 3, 5

[19] Liyuan Pan, Richard Hartley, Cedric Scheerlinck, Miaomiao Liu, Xin Yu, and Yuchao Dai. High frame rate video reconstruction based on an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 1, 2, 3, 5, 6, 7, 8

[20] Christoph Posch, Daniel Matolin, and Rainer Wohlgenannt. A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS. *IEEE Journal of Solid-State Circuits*, 46(1):259–275, 2011. 1, 2, 3

[21] Jonah P Sengupta. Demystifying Event-Based Camera Latency: Sensor Speed Dependence on Pixel Biasing, Light, and Spatial Activity. In *Proc. of SPIE Conference 13045: Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXXV*, 2023. 1, 2, 5

[22] Teresa Serrano-Gotarredona and Bernabé Linares-Barranco. A 128 × 128 1.5% Contrast Sensitivity 0.9% FPN 3 μs Latency 4 mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers. *IEEE Journal of Solid-State Circuits*, 48(3):827–838, 2013. 1, 2

[23] Wei Shang, Dongwei Ren, Dongqing Zou, Jimmy S Ren, Ping Luo, and Wangmeng Zuo. Bringing events into video deblurring with non-consecutively blurry frames. In *Proc. of International Conference on Computer Vision*, 2021. 2

[24] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based Fusion for Motion Deblurring with Cross-modal Attention. In *Proc. of European Conference on Computer Vision*, 2022. 1, 2, 6, 7, 5

[25] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based Frame Interpolation with Ad-Hoc Deblurring. In *Proc. of Computer Vision and Pattern Recognition*, 2023. 1, 2, 3, 6, 7, 5

[26] Minggui Teng, Chu Zhou, Hanyue Lou, and Boxin Shi. NEST: Neural event stack for event-based image enhancement. In *Proc. of European Conference on Computer Vision*, 2022. 1, 2

[27] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. Event enhanced high-quality image recovery. In *Proc. of European Conference on Computer Vision*, 2020. 3

[28] Fang Xu, Lei Yu, Bishan Wang, Wen Yang, Gui-Song Xia, Xu Jia, Zhendong Qiao, and Jianzhuang Liu. Motion Deblurring with Real Events. In *Proc. of International Conference on Computer Vision*, pages 2583–2592, 2021. 2

[29] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. of Computer Vision and Pattern Recognition*, 2018. 7

[30] Xiang Zhang and Lei Yu. Unifying motion deblurring and frame interpolation with events. In *Proc. of Computer Vision and Pattern Recognition*, 2022. 1, 2, 3

[31] Xiang Zhang, Lei Yu, Wen Yang, Jianzhuang Liu, and Gui-Song Xia. Generalizing Event-Based Motion Deblurring in Real-World Scenarios. In *Proc. of International Conference on Computer Vision*, 2023. 2, 3, 6, 7, 5

[32] Chu Zhou, Minggui Teng, Jin Han, Jinxiu Liang, Chao Xu, Gang Cao, and Boxin Shi. Deblurring Low-Light Images with Events. *International Journal of Computer Vision*, 131 (5):1284–1298, 2023. 1