Monochromatic Event Guided Image Deblurring with Event-triggering-aware Decomposition

Minggui Teng[1,2†] Boyu Li[1,2] Yixin Yang[1,2] Chu Zhou[3] Yan Chen[4] Jimmy S. Ren[4,5] Boxin Shi[1,2*]

[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University
[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
[3] National Institute of Informatics [4] SenseTime Research [5] Hong Kong Metropolitan University

{minggui_teng, liboyu, yangyixin93, shiboxin}@pku.edu.cn
zhou_chu@hotmail.com {yanchenace, jimmy.sj.ren}@gmail.com

Abstract

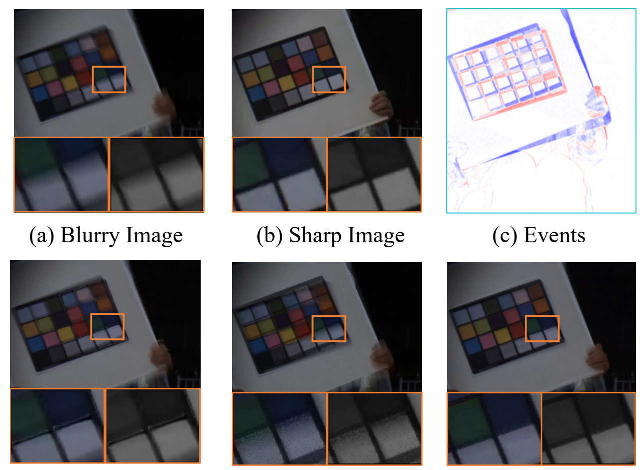
Event cameras have been applied to restore sharp images from blurry ones with high-temporal-resolution events. However, these event cameras typically record in monochromatic luminance space, leading to significant color information loss. Addressing this challenge, this paper introduces EvTraDe-Net, a novel framework for color image deblurring. It uniquely decomposes color images into luminance terms, aligned with event streams from the event-triggering process, and color terms that capture variations across color channels. EvTraDe-Net employs the event fusion for luminance deblurring and an edge map guidance for color compensation, and mutually utilizes the luminance deblurring to optimize the image decomposition in a task-oriented manner. It effectively reduces shape distortion and color bleeding, as validated on both synthetic data and our new EvRGB-Deblur dataset.

1. Introduction

Conventional cameras require exposure time to open the shutter for capturing images. However, object motion or camera shake during this time can lead to image blur, particularly in low-light environments or high-speed scenarios. During image formation, the blurry region can be perceived as sharp regions convolved with a blur kernel. Unfortunately, obtaining or estimating the blur kernel from blurry images is challenging. Moreover, in most scenarios with moving objects, the blur kernel is spatially variant, further complicating image deblurring.

Event vision sensors [14] detect scene radiance changes at each pixel. When the radiance change surpasses a preset threshold, the event pixel triggers a signal with coordinates, timestamp, and polarity. This unique characteristic allows for recording scene information with high temporal resolution. Leveraging the precise motion information encoded in event streams, blurry and sharp images can be linked through the Event-based Double Integral (EDI) model [20]. Building on the EDI model, learning-based strategies [10, 15, 24, 26, 30] have been designed in the feature domain for complementing event and image data, reducing their data distribution gap, to obtain clearer deblurred results.

However, the EDI model [20] only correlates grayscale images with monochromatic events and previous methods [23, 26] overlook the distinction in color formation between image and event acquisition processes. It results



(a) Blurry Image (b) Sharp Image (c) Events
(d) EFNet (e) EDI (f) Ours

Figure 1. Our goal is to deblur a blurry color image (a) for a sharp image (b), guided by monochromatic events (c). But the chromatic gap between images and events leads to shape distortion (d) and color bleeding (e), evident in results from EFNet [23] and EDI [20]. We introduce an event-triggering-aware image decomposition method to ease these issues (f).

† This work is done during Minggui's internship at SenseTime.
* Corresponding author

in shape distortion (Figure 1(d)) and color bleeding (Figure 1(e)) in deblurred images. To address these issues, the first step is to carefully consider the "alignment" between events and color images in the monochromatic space. That is, deblurring should be performed with monochromatic images aligned with events, while colors are handled separately. Recognizing that images hold color information absent in event signals, we are prompted to explore: *Can a color image be decomposed into a component aligned with monochromatic events triggering for improving the color image deblurring?*

Conventional color and luminance separation methods (*e.g.*, separation with `Lab` transformation) do not incorporate events signals. In this paper, we elaborately design an event guided objective function for image decomposition, consisting of a gradient constraint and an event-intensity constraint. The gradient constraint ensures that the decomposed color component remains smooth. Meanwhile, the event-intensity constraint, optimized during luminance deblurring, accounts for the pixel-wise correlation between monochromatic event signals and blurry images. By aligning with monochromatic events through this objective function, we decompose color images into two components: the `luminance` term, which is pixel-wise related to the event triggering process, and the `color` term, comprising low-frequency information that does not trigger event signals. After obtaining these image components, we introduce an event fusion module for deblurring the luminance term. To compensate for the blurry color term, we utilize edge map guidance computed from the deblurred luminance term. By recombining these deblurred components, we finally produce sharp images from our proposed **Ev**ent-**Tr**iggering-**a**ware image **De**composition and Deblurring Network (**EvTraDe-Net**). Our result exhibits improved shape distortion and less color bleeding (Figure 1(f)) by narrowing the gap between color images and event signals. Our solution makes contribution to event guided color image deblurring by:

- proposing an event-triggering-aware color image decomposition strategy, separating event-related luminance term from color term composed by:
  - + an event fusion module exploiting event temporal information to prevent shape distortion;
  - + a decomposition module optimized in a task-oriented manner during training phase via mutually improving the image deblurring;
  - + an edge map guided color compensation network to alleviate color bleeding;
- introducing the EvRGB-Deblur dataset, featuring color-diverse scenes essential for evaluating color image deblurring, with paired blurry, sharp images, and corresponding event streams.

Quantitative and qualitative results on both synthetic and our newly collected EvRGB-Deblur dataset demonstrate that our method outperforms state-of-the-art methods in recovering sharp images with more faithful color appearance.

2. Related Work

Single image deblurring. Single image deblurring techniques focus on deriving deblurred results from individual images. Traditional approaches to image deblurring either process image features [4, 27] or utilize parametric prior models to estimate motion kernels [17, 22], followed by deconvolution. With the advent of deep learning, methods employing neural networks have become dominant. Notably, a method for predicting deconvolution filter coefficients that are applied directly to input images was developed [1], and a deep CNN using a multi-scale framework for end-to-end learning was introduced [18]. To address the increasing complexity of these systems, the Nonlinear Activation Free Network [3] and the Restoration Transformer model, which processes features along the channel dimension [31], were proposed. Furthermore, the exploration of frequency domain properties with the Transformer-based FFTformer method was conducted [13]. Recently, diffusion models have been introduced for image deblurring [5] with multi-modal priors. However, due to the limited preserved information encoded inside the blurry image, these methods may underperform in complex real-world scenes.

Event-based image deblurring. Event-based image deblurring methods exploit the physical relationship between blurry images and event data to restore sharp images. The Event-based Double Integral (EDI) model [20] formulated an iterative method to recover deblurred frames from blurry frames and corresponding events. Building on EDI, CNNs were employed to refine frames and enhance details [2]. More recently, the Recurrent Event-based Frame Interpolation with ad-hoc Deblurring network was introduced for general event-based frame interpolation and deblurring [24]. Methods to extract motion cues from event streams for low-light image reconstruction were proposed in [33], while novel event representations designed to improve performance led to the development of the Event Fusion Network [23] and the Neural Event Stack [26]. Additionally, an exposure-guided event representation using a self-supervised network for generalizing motion deblurring was introduced in [32]. To handle cross-resolution gap, the Scale-Aware Spatio-temporal Network was proposed to implicitly aggregate both spatial and temporal correspondence features of images and events in [30]. However, these methods, based on the EDI model, frequently encounter issues such as shape distortion and color bleeding. This is primarily due to the domain gap between event and RGB cameras, which should be carefully considered when deblurring color images using monochromatic events.

3. Proposed Method

Monochromatic event formulation. Unlike conventional cameras, which capture images over an exposure period, event vision sensors detect changes in luminance radiance space without a color filter array. When the radiance change from t_i to t_j at a pixel exceeds a preset threshold, the event camera triggers a signal $e = (x, y, p, t)$, *i.e.*,

$$||\log \mathbf{S}_i(x,y) - \log \mathbf{S}_j(x,y)|| > c, \tag{1}$$

where (x, y) denotes the coordinates of the triggered pixel at latent images $\mathbf{S}_i$ and $\mathbf{S}_j$, and c represents the threshold of the event camera. The polarity p indicates whether the pixel intensity is increasing ($+1$) or decreasing (-1), and t is the timestamp of the event trigger. Due to the absence of exposure time in event cameras, they achieve high temporal resolution, making them well-suited for capturing motion cues essential for addressing motion blur.

EDI model. As event signals encode high-temporal information, they directly link the latent sharp images by capturing pixel value displacement, *i.e.*,

$$\mathbf{S}_j = \mathbf{S}_i \int_k \exp(c \cdot p(k)) \mathrm{d}k, \;\; \text{for } t(k) \in (t_i, t_j), \tag{2}$$

where a sequence of events $\mathbf{e} = \{x, y, p, t\}_k$ occurring between t_i and t_j can bridge two images through event integration using their corresponding thresholds. Integrating the latent sharp image again allows the events to link blurry and sharp images via a double integral [20], known as Event-based Double Integral (EDI) model, *i.e.*,

$$\mathbf{B} = \frac{\mathbf{S}_1}{T} \int_t^{t+T} \int_k \exp(c \cdot p(k)) \mathrm{d}k \; \mathrm{d}t, \tag{3}$$

where $\mathbf{B}$ and $\mathbf{S}_1$ are the blurry image with exposure time T and the sharp image at timestamp t_1, respectively.

3.1. Event-triggering-aware Image Decomposition

A blurry color image can be conceptualized as the average of multiple N images across all three channels, *i.e.*,

$$\mathbf{B}(c) = \frac{1}{N} \sum\nolimits_{k=1}^{N} \mathbf{S}_k(c), \; c \in \{R, G, B\}. \tag{4}$$

Since event cameras record radiance changes without color information, Equation (3) can be applied only to monochromatic images with event signals. Although color event cameras exist (*e.g.*, DAVIS 346C [21]), they suffer from lower resolution and mosaic issues. Directly combining color images with monochromatic events leads to image distortion (Figure 1(d)) and color bleeding (Figure 1(e)), as the image formation differ between event and frame-based cameras.

Blurry image decomposition. The previous kernel-based method [28] has attempted to decompose images into two components and use texture component to estimate a more accurate blur kernel. This inspires us to decompose the color image into two components for event-related information decoupling: one related to the luminance domain, triggered by event signals, and the other containing color information, observable in the blurry image, *i.e.*,

$$\mathbf{B}_{\mathrm{g}}, \mathbf{B}_{\mathrm{c}} \Leftarrow \mathcal{D}(\mathbf{B}), \tag{5}$$

where $\mathcal{D}$ denotes the image decomposition operator. Here, $\mathbf{B}_{\mathrm{g}}$ represents the pixel value related to event signal triggering, and $\mathbf{B}_{\mathrm{c}}$ embodies the color information. We refer to $\mathbf{B}_{\mathrm{g}}$ as `luminance` term, and to $\mathbf{B}_{\mathrm{c}}$ as `color` term.

Objective function. Decomposing a single image is an ill-posed problem. Traditional image conversion methods (such as `YUV` and `Lab`), based on the human vision system, convert RGB images with the linear transformation. As events record the pixel value changes in the log space, combining the luminance channel from conventional methods with events leads to color bleeding and shape distortion.

To reduce the chromatic gap between blurry color images and monochromatic events, our goal is to find a decomposition method that is aligned with the event-triggering process. Given that event signals record motion information in the luminance space, and that smooth regions are less likely to trigger events, we require $\mathbf{B}_{\mathrm{g}}$ to have a single channel, extracting blurry luminance information aligned with the event triggering process across RGB channels. Meanwhile, we set $\mathbf{B}_{\mathrm{c}}$ to have three channels, preserving smooth color contents, which contains non-overlapping information with events. And objective function is:

$$\underset{\mathbf{B}_{\mathrm{g}}, \mathbf{B}_{\mathrm{c}}}{\operatorname{argmin}} ||\mathcal{H}(\mathbf{B}_{\mathrm{g}}, \mathbf{B}_{\mathrm{c}}) - \mathbf{B}|| + ||\nabla \mathbf{B}_{\mathrm{c}}|| + \mathcal{R}(\mathbf{B}_{\mathrm{g}}, \mathbf{e}), \tag{6}$$

where $\mathcal{H}$ represents the inverse operator of decomposition operator $\mathcal{D}$, and $\mathbf{B}_{\mathrm{c}}$ captures smooth color information with a smaller gradient value. $\mathcal{R}$ denotes the regularization function, which is implemented as evaluating the deblurring performance of $\mathbf{B}_{\mathrm{g}}$ with monochromatic event $\mathbf{e}$, *i.e.*,

$$\mathcal{R}(\mathbf{B}_{\mathrm{g}}, \mathbf{e}) = ||\mathtt{Deblur}(\mathbf{B}_{\mathrm{g}}, \mathbf{e}) - \mathbf{S}_{\mathrm{g}}||, \tag{7}$$

in which $\mathbf{S}_{\mathrm{g}}$ is the decomposed luminance term of the sharp image $\mathbf{S}$, and `Deblur` is deblurring operation. Guided by Equation (6), our decomposition method leverages the event-triggering information and aligns images with events in monochromatic domain, we refer to it as the *event-triggering-aware image decomposition*.

A straightforward decomposition approach initially converts the image to the monochromatic space, followed by computing color term with $\mathbf{B}/\mathbf{B}_{\mathrm{g}}$. However, this handcrafted decomposition method does not account for the

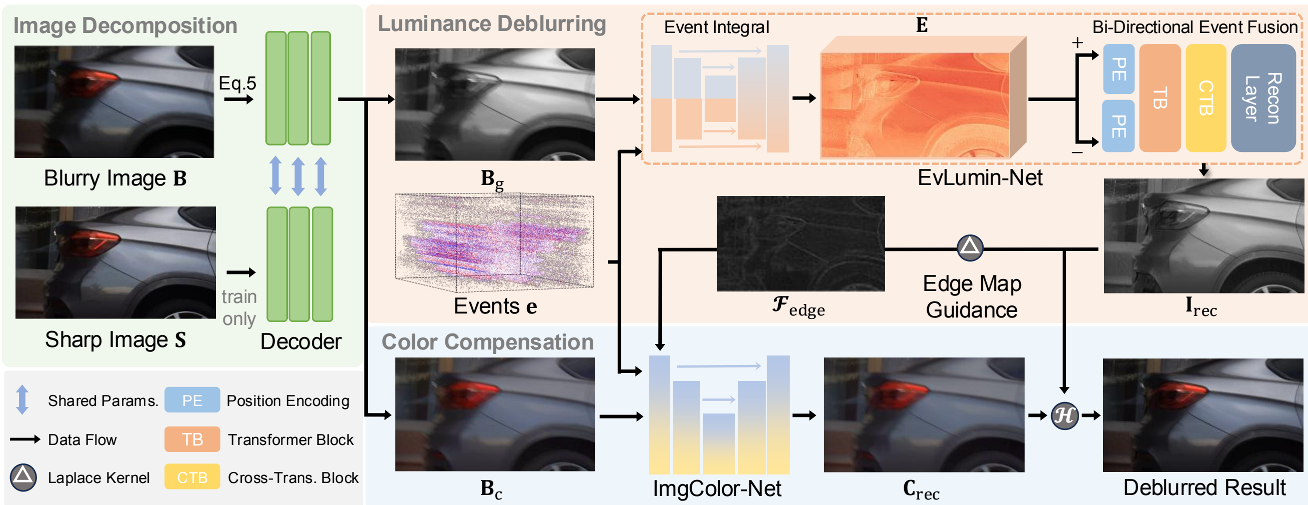


Figure 2. The pipeline of EvTraDe-Net. We first decompose the blurry image $\mathbf{B}$ with event-triggering-aware strategy by decoder network f_{d}, resulting in a luminance term $\mathbf{B}_{\mathrm{g}}$ and a color term $\mathbf{B}_{\mathrm{c}}$. The sharp image $\mathbf{S}$ is similarly decomposed, utilizing shared parameters. This step is performed exclusively for subsequent supervision and to optimize the decoder in a task-oriented manner during the training stage. And then EvLumin-Net estimates the event summation $\mathbf{E}$ by fusing the events $\mathbf{e}$ with the blurry luminance term using an event integral module. This is followed by a bi-directional event fusion module that merges the event summation to restore the sharp luminance term $\mathbf{I}_{\mathrm{rec}}$. After generating an edge map $\mathcal{F}_{\mathrm{edge}}$ from the deblurred luminance term using a Laplacian kernel, we employ ImgColor-Net to adjust the distribution of the color term, guided by edge map and events. Finally, the deblurred luminance and color terms are combined as $\mathcal{H}(\mathbf{I}_{\mathrm{rec}}, \mathbf{C}_{\mathrm{rec}})$ to produce the sharp image.

pixel-wise correlation between the event signals and the blurry image, as depicted in Equation (3). Consequently, it cannot optimally utilize the regularization function outlined in Equation (7). Therefore, our decomposition strategy also needs to ensure that the decomposed luminance term is pixel-wise aligned with the events in the monochromatic space. We propose a multi-layer convolutional decoder to fit the decomposition operation $\mathcal{D}$ in Equation (5) and introduce two-stage strategy to optimize both gradient and event-intensity constraints, shown in Figure 2.

Gradient constraint. Instead of directly adopting the gradient constraint of the color term as depicted in Equation (6), we apply the constraint to various combinations of the luminance and color terms from sharp and blurry images. To optimize the first term of the objective function Equation (6), we recombine the decomposed terms to minimize the reconstruction error. For recombination operation $\mathcal{H}$ referenced in Equation (6), we adopt a simple multiplication of the luminance and color terms. Given that the color term contains smooth information with less gradient detail, exchanging the color term impacts the merged image minimally. Thus, we utilize a cross-combination strategy to meet the second term constraint of the objective function Equation (6), as detailed in Sec. 3.4.

Event-intensity constraint. For the regularization term Equation (7), we adopt the *task-oriented methodology* [29]. We combine the training of image decomposition and luminance deblurring modules, which enhances the performance of decomposing images into event-triggering-aware luminance and color terms mutually in a task-oriented manner.

3.2. Task-oriented Luminance Deblurring

Our objective function Equation (6) encourages that image decomposition and the performance of luminance deblurring are mutually reinforcing. Thus, a more effective luminance deblurring method is also essential for improving the overall decomposition performance. To address this issue, we propose the bi-directional EDI model by incorporating forward and backward event temporal information. As Equation (3) interprets the blurry image as an integration of latent sharp images, by combining Equation (4) and Equation (5), we replace the integral operation with summation and reformulate Equation (3) as:

$$\begin{aligned}\mathbf{B}_{\mathrm{g}} &= \frac{\mathbf{S}_1}{N}\sum_{i=1}^{N}\sum_{e_k\in(t_1,t_i)}\exp(c\cdot p_k)\\ &= \frac{\mathbf{S}_1}{N}\sum_{i=1}^{N}(N-i+1)\sum_{e_k\in(t_{i-1},t_i)}\exp(c\cdot p_k).\end{aligned} \tag{8}$$

Integrating Equation (2) and Equation (8), we can derive any arbitrary sharp frame at time t_k as:

$$\mathbf{S} = \mathbf{B}_{\mathrm{g}}\frac{\sum_{i=1}^{k}\mathbf{E}(t_{i-1},t_i)}{\sum_{i=1}^{N}(N-i+1)\mathbf{E}(t_{i-1},t_i)} = \mathbf{B}_{\mathrm{g}}\cdot\mathcal{E}_{\mathrm{pre}}, \tag{9}$$

where $\mathbf{E}(t_{i-1}, t_i)$ represents the summation of events triggered by luminance changes between t_{i-1} and t_i, and $\mathbf{S}$ is the sharp image located at timestamp t_k.

Since event streams record continuous temporal information from t_1 to t_N, we can also bridge the relationship between the blurry image $\mathbf{B}$ and the sharp image $\mathbf{S}$ using inverse event streams. By substituting $\mathbf{S}_1$ with $\mathbf{S}_N$ in Equation (8), the sharp image can derived as follows:

$$\mathbf{S}' = \mathbf{B}_{\mathrm{g}} \frac{\sum_{i=k}^{N} \mathbf{E}'(t_{i-1}, t_i)}{\sum_{i=1}^{N} i \cdot \mathbf{E}'(t_{i-1}, t_i)} = \mathbf{B}_{\mathrm{g}} \cdot \mathcal{E}_{\mathtt{post}}, \tag{10}$$

where $\mathbf{E}'(t_{i-1}, t_i) = 1/\mathbf{E}(t_{i-1}, t_i)$, and $\mathbf{S}'$ represents the dual solution obtained by reversing the event streams. As both Equation (9) and Equation (10) can restore sharp latent images, we merge them when applying the EDI model bidirectionally to produce a more reliable estimation:

$$\mathbf{S} = \mathbf{B}_{\mathrm{g}} \cdot f_{\mathrm{w}}(\mathcal{E}_{\mathtt{pre}}, \mathcal{E}_{\mathtt{post}}), \tag{11}$$

where f_{w} denotes a weighted combination operator. In Equation (11), we reformulate the EDI model by complementarily utilizing bi-directional event information. Due to introducing more motion information encoded in the event streams via the bi-directional path, bi-directional EDI model is expected to restore sharper image in the luminance space.

EvLumin-Net. As described above, accurate event integration is a key step. To this end, we propose the EvLumin-Net tailored for luminance deblurring shown in Figure 2, which consists of two modules: event integral module for event integral summation and bi-directional event fusion module for summation combination. Given that the triggering threshold is spatial and temporal variant [6], we combine $\mathbf{B}_{\mathrm{g}}$ and event $\mathbf{e}$ to obtain the event summation $\mathbf{E}$:

$$\mathbf{E} = \mathcal{M}_{\mathtt{int}}(\mathbf{B}_{\mathrm{g}}, \mathbf{e}), \tag{12}$$

where $\mathcal{M}_{\mathtt{int}}$ is the implicit function of event integral module. To better explore the relationship between the blurry image and event streams, we employ a U-Net backbone to fuse features from both modalities by using a dual-encoder and single-decoder strategy, to predict event summations.

As indicated in Equation (9) and Equation (10), different event summations carry varying weights in the merging process. To better formulate this process, we propose a bi-directional event fusion module to combine the initial event summations with a dual-way position encoding. Then, we apply a Transformer-based block [31] to obtain the final restored sharp image, *i.e.*,

$$\mathbf{I}_{\mathtt{rec}} = \mathcal{M}_{\mathtt{bi}}(\mathbf{B}_{\mathrm{g}}, \mathbf{E} + \mathtt{PE}_1, \mathbf{E} + \mathtt{PE}_2), \tag{13}$$

A detailed proof can be found in supplementary.

where $\mathtt{PE}_1, \mathtt{PE}_2 \in \mathbb{R}^N$ represent two position encoding vectors with learned parameters, $\mathcal{M}_{\mathtt{bi}}$ represents the implicit function of bi-directional event fusion module, and $\mathbf{I}_{\mathtt{rec}}$ is the deblurred luminance term. Note that we assign the same position encoding value to each channel of the event summation. Since the original Transformer block only computes self-attention, to better utilize the luminance cues in the blurry image, we introduce a cross-Transformer block with cross-attention mechanism, where the key vector is derived from the blurry image feature maps. After extracting and fusing the feature maps from the event summations and blurry image, we use multiple convolution layers to reconstruct the final sharp image.

Overall, during the luminance deblurring stage shown in Figure 2, EvLumin-Net calculates the event summation using appropriate thresholds and fuses features from the blurry image with event integral module. After obtaining the initial event summations over different time intervals, we elaborately design a bi-directional event fusion module to merge the image features and event summation, finally restoring the deblurred luminance image.

3.3. Color Compensation

Since event signals are recorded in monochromatic space, directly combining them with color terms is not feasible. Thus, we decompose blurry images first and deblur luminance terms, processed with pixel-wise aligned event signals. Since deblurred luminance term delineates the boundaries of regions within the images, we propose an edge map-guided color compensation strategy, to compensate non-overlapping color information.

ImgColor-Net. Guided by edge map computed from the deblurred luminance term, we employ another U-Net backbone, named ImgColor-Net, to reconstruct color terms. Furthermore, increased event triggering at the same pixel suggests greater color displacement. Therefore, we first fuse the events and edge map and then use the resulting features to compensate chromatic information in the color term, *i.e.*,

$$\mathbf{C}_{\mathtt{rec}} = f_{\mathrm{c}}(\mathbf{B}_{\mathrm{c}}, \mathcal{F}_{\mathtt{edge}}, \mathbf{e}), \tag{14}$$

where $\mathcal{F}_{\mathtt{edge}}$ is the edge map computed from the deblurred luminance term $\mathbf{I}_{\mathtt{rec}}$, f_{c} is the implicit function of ImgColor-Net and $\mathbf{C}_{\mathtt{rec}}$ is the deblurred color term.

As depicted in Figure 2, we extract the edge map from the deblurred luminance term $\mathbf{I}_{\mathtt{rec}}$ using a Laplacian kernel. Before fusing the color term features, we combine the edge map and events using a convolution layer and an SE block [8]. These are fed into a U-Net to reconstruct the deblurred color term, which adjust the blurry color term $\mathbf{B}_{\mathrm{c}}$ to the sharp one $\mathbf{S}_{\mathrm{c}}$ in feature domain, guided by events and edge maps. After reconstructing the color term, we can ob-

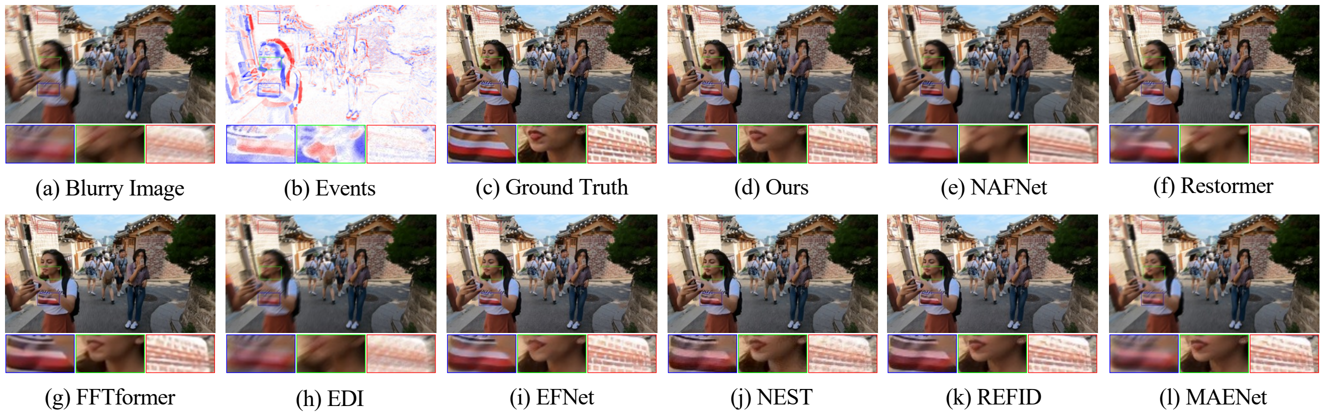


Figure 3. Visual quality comparison on REDS dataset [19]. (a) Blurry image. (b) Events. (c) Ground truth. (d)∼(l) Deblurred results of ours, NAFNet [3], Restormer [31], FFTformer [13], EDI [20], EFNet [23], NEST [26], REFID [24], and MAENet [25].

tain the sharp image by merging the deblurred color term with the luminance term.

3.4. Implementation Details

Loss function. The loss function of EvTraDe-Net, consists of two components: decomposition loss and reconstruction loss. The decomposition loss $\mathcal{L}_d$ is defined as:

$$\mathcal{L}_d = \sum ||X_\mathrm{c} \cdot \mathbf{S}_\mathrm{g} - \mathbf{S}|| + \sum ||X_\mathrm{c} \cdot \mathbf{B}_\mathrm{g} - \mathbf{B}||, \quad (15)$$

where $\mathbf{B}$ and $\mathbf{S}$ represent the blurry and sharp images, respectively, X_c is the color term of blurry image $\mathbf{B}$ or sharp image $\mathbf{S}$. The reconstruction loss $\mathcal{L}_r$ is defined as:

$$\mathcal{L}_r = \beta_1 \mathcal{L}_2(\mathbf{S}_\mathtt{rec}, \mathbf{S}) + \beta_2 \mathcal{L}_\mathtt{perc}(\mathbf{S}_\mathtt{rec}, \mathbf{S}), \quad (16)$$

where $\mathbf{S}_\mathtt{rec}$ is the deblurred image, $\mathcal{L}_2$ and $\mathcal{L}_\mathtt{perc}$ represent the mean squared error (MSE) loss and perceptual loss based on a pretrained VGG-16 network, and β_1 and β_2 are hyper-parameters, empirically set to 10 and 0.5 in our experiments, respectively. The total loss is optimized jointly as:

$$\mathcal{L} = \alpha_1 \mathcal{L}_d + \alpha_2 \mathcal{L}_r, \quad (17)$$

with α_1 and α_2 set to 1 and 0.1, respectively.

Training details. We adopt a two-stage training strategy for EvTraDe-Net. In the first stage, we train the decoder network and EvLumin-Net using both the decomposition and reconstruction losses. For computing the reconstruction loss, we directly combine the deblurred luminance term $\mathbf{I}_\mathtt{rec}$ with the decomposed color term $\mathbf{B}_\mathrm{c}$. By jointly optimizing image decomposition and luminance deblurring, decoder network is trained in a task-oriented manner.

In the second stage, we freeze parameters of decoder network and EvLumin-Net, and train ImgColor-Net with the reconstruction loss. We implement our method using the PyTorch platform on an NVIDIA RTX 3090 GPU. For each stage, we train our network for 100 epochs with the AdamW optimizer and a cosine learning rate scheduler. The initial learning rates are set to 1×10^{-3} and 1×10^{-4} for the first and second stages, respectively.

4. Experiment

Synthetic Dataset. We perform both quantitative and qualitative experiments on the REDS dataset [19] and the GoPro dataset [18]. For event data simulation, we first interpolate the frame rate by a factor of 16 using RIFE [9]. Event streams are then simulated from high-frame-rate videos using the DVS-Voltmeter [16], while blurry images are generated by averaging sharp images over a sequence.

Our approach is evaluated against three image-based deblurring methods: NAFNet [3], Restormer [31], and FFTformer [13], as well as five event-based methods: EDI [20], NEST [26], EFNet [23], and REFID [24], and MAENet [25]. To measure the quality of deblurred images, we utilize commonly adopted metrics: PSNR and SSIM. Quantitative comparisons are presented in Tab. 1, and qualitative comparisons are shown in Figure 3. Our method demonstrates superior performance compared to other state-of-the-art approaches on both metrics, effectively restoring more texture details encoded within the event streams and producing deblurred images with reduced color bleeding and fewer artifacts. Image-based methods (NAFNet [3] and Restormer [31]) fail to recover high-frequency details. And event-based methods (REFID [24], MAENet [25]) suffer from color bleeding, as highlighted in the blue box in Figure 3, since they did not properly solve the domain gap between the color images and monochromatic events.

EvRGB-Deblur Dataset. We evaluate our method on a real-world dataset [7] to assess its performance. The quantitative results are provided in Tab. 1, while the qualitative

Table 1. Quantitative comparisons on the three benchmark datasets [7, 18, 19]. The "Event" column specifies whether methods require events as input (Yes [✓] or No [✗]). The best performances are highlighted in **bold**.

Method	Event	REDS [19]		GoPro [18]		EventAid-B [7]	
		PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
NAFNet [3]	✗	26.13	0.795	33.69	0.967	26.00	0.694
Restormer [31]	✗	29.27	0.874	32.92	0.961	26.06	0.680
FFTformer [13]	✗	29.19	0.881	34.21	0.969	25.09	0.660
EDI [20]	✓	25.16	0.800	29.06	0.943	26.63	0.656
NEST [26]	✓	32.39	0.919	34.59	0.967	19.89	0.576
EFNet [23]	✓	32.46	0.930	35.46	0.972	26.72	0.681
REFID [24]	✓	32.54	0.930	35.91	0.973	26.19	0.765
MAENet [25]	✓	31.91	0.917	36.07	0.976	26.01	0.744
Ours	✓	**33.08**	**0.938**	**36.67**	**0.980**	**27.28**	**0.800**

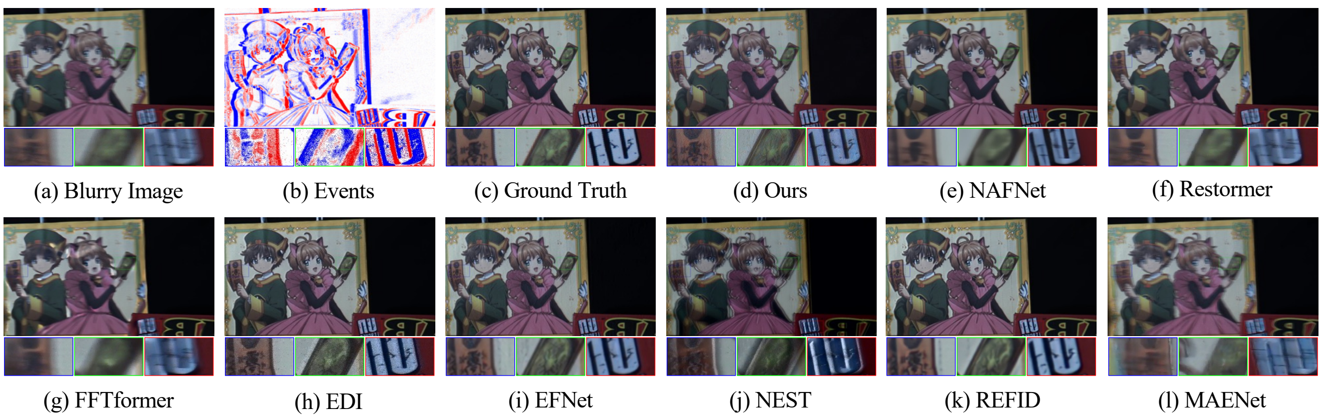


Figure 4. Visual quality comparison on EventAid-B dataset [7]. (a) Blurry image. (b) Events. (c) Ground truth. (d)∼(l) Deblurred results of ours, NAFNet [3], Restormer [31], FFTformer [13], EDI [20], EFNet [23], NEST [26], REFID [24], and MAENet [25].

results are shown in Figure 4. These results highlight the effectiveness of our method in restoring sharp images from blurry ones in real-world scenarios, demonstrating its practicality and robustness.

However, existing datasets [7, 11, 12] do not contain a sufficient number of color-diverse scenes, which are essential for evaluating the performance of color image deblurring methods. Thus, we introduce the EvRGB-Deblur dataset, which utilizes a hybrid camera system to collect a set of paired blurry and sharp images, along with corresponding event streams. And we capture data under ten different lighting conditions, incorporating a variety of colorful objects (*e.g.*, cube and colorchecker). Each scene consists of 100 paired data points, resulting in a total of 1000 image-event pairs.

To obtain the blurry images, we set the exposure time as 50ms, while for sharp images, the exposure time is set to 5ms. The captured blurry images have a frame rate of 24 FPS, aligning with standard datasets, such as the GoPro dataset, which has a frame rate of 30 FPS. We achieve temporal synchronization using a signal generator, and ensure spatial alignment through camera calibration. This novel dataset empowers a more comprehensive evaluation of the performance of event-based color image deblurring methods. Detailed dataset comparison can be found in the supplementary material.

Qualitative comparisons are in Figure 5 and quantitative comparisons in Figure 6. As demonstrated in the radar plots within Figure 6, our method consistently outperforms other state-of-the-art techniques on both metrics. Notably, our method exhibits significant improvements on PSNR in the case of group #6, which involves a rotating fan and colorful blocks. We further illustrate an example from group #6 in Figure 5. As the results shown, MAENet deblurred image exhibits color bleeding (blue box), and NEST shows shape distortion (red box). However, our method manages to preserve sharper edges and exhibits less shape distortion.

Decomposition Validation. To validate the improvements offered by our proposed decomposition method, we conduct an experiment involving the deblurring of the luminance term using the EDI method (denoted as EDI*),

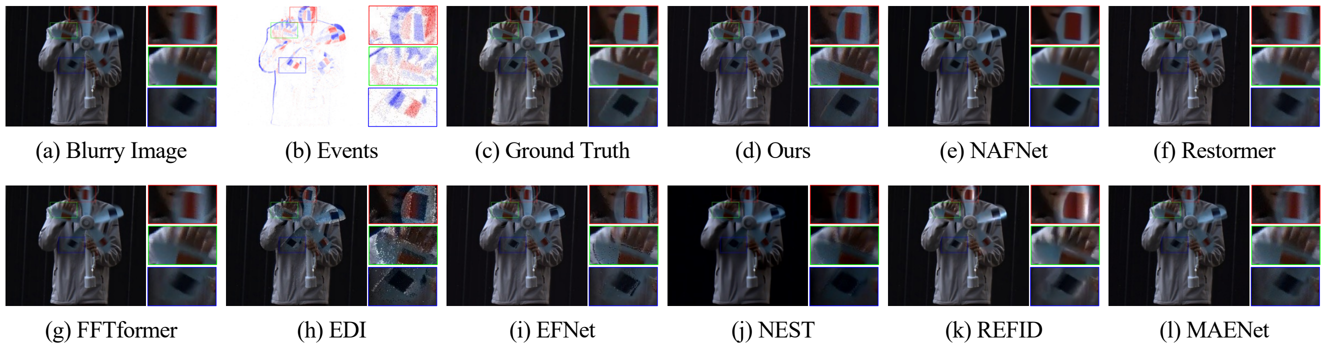


Figure 5. Visual quality comparison on real-captured EvRGB-Deblur dataset. (a) Blurry image. (b) Events. (c) Ground truth. (d)∼(l) Deblurred results of ours, NAFNet [3], Restormer [31], FFTformer [13], EDI [20], EFNet [23], NEST [26], REFID [24], and MAENet [25].

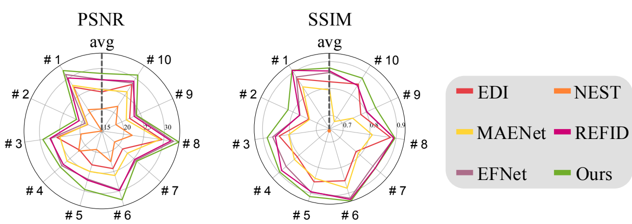


Figure 6. The quantitative results on EvRGB-Deblur dataset. In the radar plots, performance improves with distance from the center. The 11 axes in each radar plot correspond to 10 groups and average results across all groups.

Table 2. Quantitative results of decomposition validation.

	EDI	EDI*	Lab	YUV	Ours
PSNR	25.16	31.64	30.67	30.39	**33.08**
SSIM	0.800	0.911	0.923	0.897	**0.938**

to verify that the luminance term is pixel-wise correlated with the event signals as depicted in Equation (3). Furthermore, we input the separable `L` / `Y` and `ab` / `UV` channels into EvLumin-Net and ImgColor-Net, respectively (denoted as `Lab` and `YUV`), to compare the conventional color transformation method with our decomposition approach. Additionally, we explore the effectiveness of using the multiplication to implement combination operation $\mathcal{H}$ in Equation (6) (denoted as Add). The results, presented in Tab. 2, demonstrate the effectiveness of our proposed event-triggering-aware strategy. Specifically, we show the visual comparison between EDI* and vanilla EDI method [20] in Figure 7. As the results shown, EDI* can restore sharper edges, since inputting our decomposed luminance term significantly enhances its performance.

5. Conclusion

In this paper, we introduce a novel approach to color image deblurring with monochromatic event signals. Our event-

(a) EDI (b) EDI*

Figure 7. Visual comparison between the original EDI method [20] (a) and EDI* (b).

triggering-aware decomposition method allows for a more effective deblurring process. By employing the EvLumin-Net and ImgColor-Net for the decomposed luminance and color terms, our method achieves significant improvements in reducing shape distortion and color bleeding. Besides, the proposed EvRGB-Deblur dataset offers a more comprehensive tool for evaluating deblurring methods.

Limitation. Despite the notable achievements of our method, several limitations remain. In terms of color restoration, accurately recovering the saturation of the deblurred image is challenging, as it is difficult to infer from both the event streams and blurry images. Moreover, the integration of mosaic RGB events with color images, especially in contexts where color fidelity is crucial, remains a topic for further exploration.

Acknowledgment

This work was supported by National Natural Science Foundation of China (Grant No. 62088102, 62136001), Beijing Natural Science Foundation (Grant No. L233024), and Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Grant No. Z241100003524012).

References

[1] Ayan Chakrabarti. A neural approach to blind motion deblurring. In *Proc. of European Conference on Computer Vision*, 2016. 2

[2] Haoyu Chen, Minggui Teng, Boxin Shi, Yizhou Wang, and Tiejun Huang. A residual learning approach to deblur and generate high frame rate video with an event camera. *IEEE Transactions on Multimedia*, 25:5826–5839, 2023. 2

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proc. of European Conference on Computer Vision*, 2022. 2, 6, 7, 8

[4] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *Proc. of ACM SIGGRAPH Asia*, 2009. 2

[5] Hyungjin Chung, Jong Chul Ye, Peyman Milanfar, and Mauricio Delbracio. Prompt-tuning latent diffusion models for inverse problems. In *Proc. of International Conference on Machine Learning*, 2024. 2

[6] Tobi Delbruck, Yuhuang Hu, and Zhe He. v2e: From video frames to realistic DVS event camera streams. *arXiv preprint arXiv:2006.07722*, 2020. 5

[7] Peiqi Duan, Boyu Li, Yixin Yang, Hanyue Lou, Minggui Teng, Yi Ma, and Boxin Shi. EventAid: Benchmarking event-aided image/video enhancement algorithms with real-captured hybrid dataset. *arXiv preprint arXiv:2312.08220*, 2023. 6, 7

[8] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proc. of Computer Vision and Pattern Recognition*, 2018. 5

[9] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *Proc. of European Conference on Computer Vision*, 2022. 6

[10] Dachun Kai, Jiayao Lu, Yueyi Zhang, and Xiaoyan Sun. EvTexture: Event-driven texture enhancement for video super-resolution. In *Proc. of International Conference on Machine Learning*, 2024. 1

[11] Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. CMTA: Cross-modal temporal alignment for event-guided video deblurring. In *Proc. of European Conference on Computer Vision*, 2024. 7

[12] Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. Frequency-aware event-based video deblurring for real-world motion blur. In *Proc. of Computer Vision and Pattern Recognition*, 2024. 7

[13] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proc. of Computer Vision and Pattern Recognition*, 2023. 2, 6, 7, 8

[14] 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

[15] 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. 1

[16] Songnan Lin, Ye Ma, Zhenhua Guo, and Bihan Wen. DVS-Voltmeter: Stochastic process-based event simulator for dynamic vision sensors. In *Proc. of European Conference on Computer Vision*, 2022. 6

[17] Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. In *Proc. of European Conference on Computer Vision*, 2014. 2

[18] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proc. of Computer Vision and Pattern Recognition*, pages 3883–3891, 2017. 2, 6, 7

[19] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study. In *Proc. of Computer Vision and Pattern Recognition Workshops*, 2019. 6, 7

[20] 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. 1, 2, 3, 6, 7, 8

[21] Cedric Scheerlinck, Henri Rebecq, Timo Stoffregen, Nick Barnes, Robert Mahony, and Davide Scaramuzza. CED: Color event camera dataset. In *Proc. of Computer Vision and Pattern Recognition Workshops*, 2019. 3

[22] Libin Sun, Sunghyun Cho, Jue Wang, and James Hays. Edge-based blur kernel estimation using patch priors. In *Proc. of International Conference on Computational Photography*, 2013. 2

[23] 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, 8

[24] 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, 6, 7, 8

[25] Zhijing Sun, Xueyang Fu, Longzhuo Huang, Aiping Liu, and Zheng-Jun Zha. Motion aware event representation-driven image deblurring. In *Proc. of European Conference on Computer Vision*, 2024. 6, 7, 8

[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, 6, 7, 8

[27] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *Proc. of European Conference on Computer Vision*, 2010. 2

[28] Yuquan Xu, Xiyuan Hu, Lu Wang, and Silong Peng. Single image blind deblurring with image decomposition. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012. 3

[29] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *International Journal of Computer Vision*, 127:1106–1125, 2019. 4

[30] Wei Yu, Jianing Li, Shengping Zhang, and Xiangyang Ji. Learning scale-aware spatio-temporal implicit representation for event-based motion deblurring. In *Proc. of International Conference on Machine Learning*, 2024. 1, 2

[31] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. of Computer Vision and Pattern Recognition*, 2022. 2, 5, 6, 7, 8

[32] Xiang Zhang, Lei Yu, Wen Yang, Jianzhuang Liu, and Gui-Song Xia. Generalizing event-based motion deblurring in real-world scenarios. In *Proc. of Computer Vision and Pattern Recognition*, 2023. 2

[33] 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. 2