# Event-based Motion Deblurring with Blur-aware Reconstruction Filter

Nuo Chen , Chushu Zhang , Wei An , Longgaung Wang , Miao Li ,Qiang Ling

***Abstract*—Event-based motion deblurring aims at reconstructing a sharp image from a single blurry image and its corresponding events triggered during the exposure time. Existing methods learn the spatial distribution of blur from blurred images, then treat events as temporal residuals and learn blurred temporal features from them, and finally restore clear images through spatio-temporal interaction of the two features. However, due to the high coupling of detailed features such as the texture and contour of the scene with blur features, it is difficult to directly learn effective blur spatial distribution from the original blurred image. In this paper, we provide a novel perspective, i.e., employing the blur indication provided by events, to instruct the network in spatially differentiated image reconstruction. Due to the consistency between event spatial distribution and image blur, event spatial indication can learn blur spatial features more simply and directly, and serve as a complement to temporal residual guidance to improve deblurring performance. Based on the above insight, we propose an event-based motion deblurring network consisting of a Multi-Scale Event-based Double Integral (MS-EDI) module designed from temporal residual guidance, and a Blur-Aware Filter Prediction (BAFP) module to conduct filter processing directed by spatial blur indication. The network, after incorporating spatial residual guidance, has significantly enhanced its generalization ability, surpassing the best-performing image-based and event-based methods on both synthetic, semi-synthetic, and real-world datasets. In addition, our method can be extended to blurry image super-resolution and achieves impressive performance. Our code is available at: https://github.com/ChenYichen9527/MBNet now.**

***Index Terms*—Motion deblurring, Event-based vision, Super-resolution**

## I. Introduction

Motion deblurring aims at restoring the sharp image from originally captured images with motions caused by objects moving or cameras shaking during the exposure time [1, 2]. Given the inherently ill-posed problem, many learning-based methods have been proposed to learn the deblurring mapping functions by specific network designs, including multi-stage structures [3, 4], multi-scale structures [5–8], recurrent mechanisms [9–14], and self-attention mechanisms [15–19]. Despite the continuous improvements in certain scenes, these methods often struggle to effectively handle real distorted images, owing to the complicated blur patterns with unknown scene dynamics [5, 20].

This work was partially supported in part by the Independent Innovation Science Fund Project of National University of Defense Technology (Nos. 22–ZZCX-042) (Corresponding author: Miao Li, Wei An)

Nuo chen, Chushu Zhang, Miao Li, Qiang Ling, Wei An are with the College of Electronic Science and Technology, National University of Defense Technology (NUDT), P. R. China. Emails: chennuo97, zhangchushu, lm8866, lingqiang16, anwei@nudt.edu.cn. Longguang Wang is with the Aviation University of Air Force, Changchun 130012, China. E-mail: wanglongguang15@nudt.edu.cn.

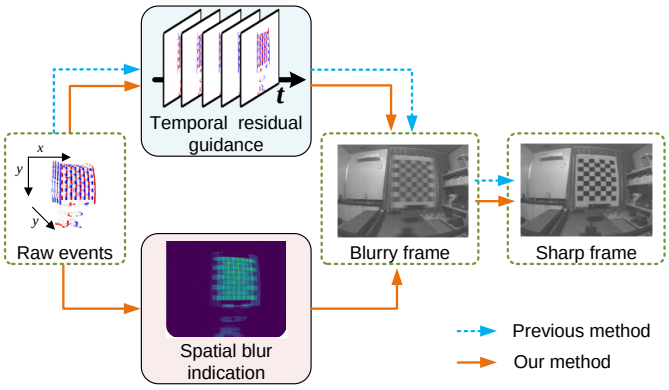



Fig. 1. Comparison between our method and the previous method. The previous method treats events as temporal residuals to guide deblurring, while our method also utilizes additional spatial blur indication of events to better reconstruct clear images.

Since deblurring methods are limited by unknown motion patterns during the exposure time, a new approach is brought up by resorting to a novel and promising sensor, the event camera. An event camera can record the precise luminance change in the scene with high response speed due to the asynchronous imaging design [21–23]. This variant record during the exposure time compensates for the elimination of texture details in a motion-distorted image and enhances the performance in image restoration, especially in a highly dynamic change environment. Recent event-based motion deblurring methods have made some attempts to bring the events into the motion deblurring task and achieved significant improvements [24]. Commonly, these methods follow the scheme as the "Temporal residual guidance" in Fig. 1. The blurry image is regarded an integrated frame from a frame sequence within the exposure time, which are bridged by event-formed residual frames. To this end, the temporal-spatial interaction built from events and images is explicitly constructed. Based on this interaction, existing methods make various network designs including event-based double integral (EDI) based networks [25, 26], flow-based networks [27], and attention-based networks [28, 29], which are validated to be effective.

However, existing deblurring methods require learning the spatial distribution of blur from the original blurred images to perform precise deblurring on different regions. This task is quite challenging because frame-based images not only encompass the spatial information of blur but also integrate intricate details such as scene structure and texture. The

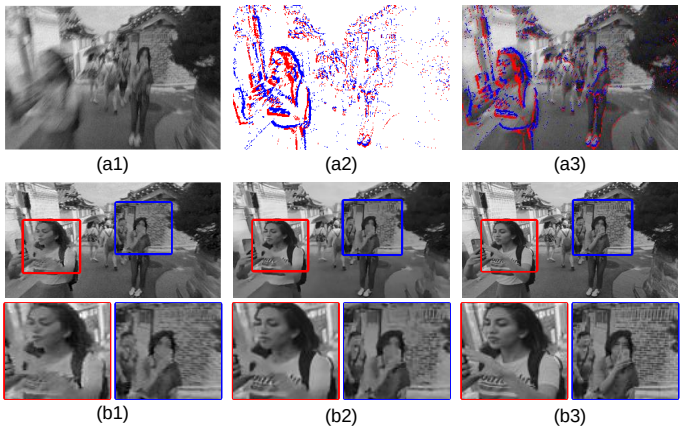



Fig. 2. Illustrative example of consistency between blur score map and actual blur regions. (a1) a blurry image; (a2) the event blur score map; (a3) the overlapping result of the first two images; (b1) to (b3) show the deblurring results guided only by event temporal residuals, the deblurring results guided by adding spatial blur, and the ground truth, respectively. In the areas with serious blur, the event blur score map has a higher response. And after adding spatial guidance to remove blurring, there are fewer halo artifacts generated and the better visualization performance is achieved.

entanglement of these details with blur information poses significant difficulties for the network to extract discriminative semantic features, hindering more refined deblurring. In this paper, we explore the application of event data in deblurring from a novel perspective, specifically, the network learns the actual blur distribution directly from the spatial distribution of events, as shown in Fig. 1. We accumulate events within the exposure time along the time dimension to obtain an event blur score map. Since events in each pixel indicate luminance change during the exposure time and it's spatial distribution reflects the degree of blurriness in different regions of a blurry image, so they can provide blur indication in a blur image. This blur indication can be explained simply as a principle that the more events generated, the larger luminance change in the area occurs, and the severer blur retains. In areas with severe blurring, the event blur score map has a higher response, that is, the event spatial blur score map is consistent with the actual blur area, as shown in Fig. 2. With the blur indication information, a spatially differentiated reconstruction filter can be predicted, mapping each pixel on the blurry image to the same pixel on the sharp image.

Based on this insight, we propose a novel framework called MBNet, which consists of two core modules, a multi-scale event-based double integral module (MS-EDI) and a Blur-Aware Filter Prediction (BAFP) module. These two modules respectively learn temporal residual features and spatial blur features from events and predict the spatio-temporal filter. In MS-EDI, the split events are fed into two weight-sharing feature encoders for multi-scale event residual feature output. The BAFP module predicts a spatially blur-aware filter from the event blur score map. These two modules are sequentially applied to the distorted image to achieve coarse-to-fine deblurring. Notably, since the output channel of the reconstruction filter is adjustable, our proposed network can be easily extended to blurry image super-resolution. Qualitative and quantitative experiments on the synthetic and real-world datasets show the outstanding performance of our proposed method compared to state-of-the-art methods and demonstrate the effectiveness of our network in motion deblurring.

To sum up, in this work, we propose a network for motion deblurring by using both the guidance of temporal residuals and the blur indication by events in the exposure time. The contributions of this work are three-fold:

1. We first argue that the motion information captured by event cameras can be regarded as a novel blur indication, which is beneficial in spatially differentiated image reconstruction for motion deblurring. Based on this, we develop a blur-aware reconstruction module to validate this concept, demonstrating its effectiveness.

2. We propose an event-based motion deblurring network called MBNet. This network integrates temporal residual guidance and spatial blur indication to fully utilize the rich information provided by event cameras.

3. Our method significantly outperforms existing state-of-the-art methods on the synthetic and real-world benchmarks for the motion deblurring task. In addition, we extend our method to the blurry image super-resolution task, showing that our design is generic and effective for various blurry image restoration tasks.

## II. Related Work

### A. Image-based Motion Deblurring

Image-based motion deblurring aims to solve the problem of spatial blurring from the original blurred image and restore the blurred image into a clear image with distinct texture contours. Early methods attempted to apply some priors to estimate the blur kernel, such as linear motion, Gaussian scale mixture, and sparsity. However, if the above assumptions are not fulfilled, these methods will degrade rapidly, leading to artifacts and reduced deblurring performance.

With the development of deep neural networks, various effective network designs have been proposed that enable the direct learning of neural networks from data for motion deblurring. Such as encoder-decoder based Unet framework[30, 31], multi-scale fusion[5–7, 32], multi-stage fusion[3, 4], recurrent refinement[9–13] and self-attention mechanism[15–18]. These methods have achieved significant performance improvements. However, these works fail in extreme scenarios since the considerable combinations of real-world textures and motions are severely missing in a blurred image.

### B. Event-based Motion Deblurring

Event cameras measure the changes in light intensity at each pixel and output events when the change exceeds a threshold. They can asynchronously output events at extremely high frequencies, thus providing missing information during the exposure time of blurred images caused by motion. Compensating for blurred images using events can easily resolve issues such as temporal ambiguity and texture erasure.

Event-based motion deblurring methods can be categorized into EDI-based methods, flow-based methods, and attention-based methods. EDI-based methods build the relation between latent sharp frames and blurry frames with double

integral feature of events [25, 26]. Early methods estimated unknown parameters in the EDI model through traditional optimization approaches. After that, lots of network-based methods motivated by the model are proposed, with specific designs including dynamic filter generation[33], sparse learning[34, 35], deformable convolutions[36], kernel prediction network[37], self-supervised learning framework[38, 39], multi-patch architecture[40], bi-directional recurrent architecture[41], efficient design[42] and spiking or polynomial representation[43, 44]. Events with high temporal resolution provide the residual luminance information for the spatial-temporal interaction construction. Flow-based methods[27, 45, 46] assume that events with high temporal resolution provide residual brightness information for spatiotemporal interaction construction, and then utilize the residual brightness of the front and back frames to compensate for the current frame through a warping operation. Attention-based methods[28, 29, 47, 48] regard the event stream as event frame sequences in a different modality, which can be leveraged for feature fusion of the current frame through attention-based modules.

Although these aforementioned event-based methods vary in the network designs, they are common in event information utilization: events with low latency provide rich temporal information, which is complementary to the current sharp spatial frame reconstruction through temporal-spatial interaction. While in this work, a novel perspective is brought up, that events not only provide temporal residual cues but can also serve as blur indicators for the sharp image reconstruction.

## III. Methods

We first present two formulations from temporal and spatial perspectives to bridge the sharp images and blurry images through events in Sec. III-A. In Sec. III-B, we introduced our designed network architecture, MBNet, which estimates temporal guidance and spatial guidance filters from events based on these two formulas. Finally, we present our training loss in Sec. III-C.

### A. Problem Formulation

Given a blurry image $\boldsymbol{B}$ and the events $\boldsymbol{e}(s)$ within the exposure time $\tau$, we aim to get the sharp image $\boldsymbol{I}$ by a reconstruction filter $\mathscr{F}$. This formulation can be constructed as:

$$\boldsymbol{I} = \mathscr{F}\left(\boldsymbol{B}, \boldsymbol{e}(s), \tau\right), \tag{1}$$

Event cameras trigger an event $\boldsymbol{e}(x, y, t, p)$ when the local logarithmic brightness variation $\log \Delta \boldsymbol{I}$ of each pixel reaches the threshold value $c$. The event response mechanism can be formulated as:

$$p = \begin{cases} +1, & \log \Delta \boldsymbol{I}(x, y) \geq c \\ -1, & \log \Delta \boldsymbol{I}(x, y) \leq -c \end{cases}, \tag{2}$$

where $p$ is polarity, indicating whether the brightness increases or decreases.

*1) **Temporal residual guidance deblurring**:* From the temporal perspective, the relation between a blurry frame $\boldsymbol{B}$ and a sharp frame $\boldsymbol{I}$ can be constructed as an integral process:

$$\boldsymbol{B}(x, y, f) = \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} \boldsymbol{I}(x, y, f + \Delta t) d\Delta t, \tag{3}$$

where the blurry image $\boldsymbol{B}(f)$ with timestamp $f$ is the integral of all the sharp images $\boldsymbol{I}$ within the exposure time $\tau$. Given that events indicate brightness variation as Eq. 2, the residual frame can be constructed as:

$$\begin{aligned} \log \Delta \boldsymbol{I}(x, y, t) &= \log \boldsymbol{I}(x, y, t + \Delta t) - \log \boldsymbol{I}(x, y, t) \\ &= c \int_{t}^{t+\Delta t} \boldsymbol{e}(s) ds, \end{aligned} \tag{4}$$

where $e(t) \triangleq p \cdot \delta(t)$ denotes the continuous representation of events with $\delta(\cdot)$ indicating the Dirac function. Thus, the sharp image can be constructed by the blurry image and the events within the exposure time $\tau$ as:

$$\boldsymbol{B}(x, y, f) = \frac{\boldsymbol{I}(x, y, f)}{\tau} \int_{-\tau/2}^{\tau/2} \exp\left(c \int_{f}^{f+t} \boldsymbol{e}\left(s\right) ds\right) dt. \tag{5}$$

To avoid zero division, we transform Eq. 5 from real domain to logarithmic domain:

$$\tilde{\boldsymbol{B}}(x, y, f) = \tilde{\boldsymbol{I}}(x, y, f) + \frac{\exp c}{\tau} \int_{-\tau/2}^{\tau/2} \int_{f}^{f+t} \boldsymbol{e}\left(s\right) ds dt, \tag{6}$$

where $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{I}}$ refer to the blurry and sharp frames in the logarithmic domain respectively. The Eq. 6 can be transformed as:

$$\begin{aligned} &\tilde{\boldsymbol{B}}(x, y, f) = \tilde{\boldsymbol{I}}(x, y, f) + \\ &\frac{\exp c}{\tau} \left( \int_{0}^{\tau/2} \int_{f}^{f+t} \boldsymbol{e}\left(s\right) ds dt - \int_{0}^{-\tau/2} \int_{f}^{f+t} \boldsymbol{e}\left(s\right) ds dt \right), \end{aligned} \tag{7}$$

where events are split into two parts through exposure timestamp $f$. When the events before the exposure timestamp $f$ are reversed by time, the two parts of events can be processed by the same function, which is presented as:

$$\begin{aligned} \tilde{\boldsymbol{B}}(x, y, f) = \tilde{\boldsymbol{I}}(x, y, f) &+ \Theta_{EDI}\left(\boldsymbol{e}\left(s\right), \tau/2\right) \\ &- \Theta_{EDI}\left(\boldsymbol{e}\left(s\right), -\tau/2\right), \end{aligned} \tag{8}$$

where $\Theta_{EDI}$ refers to the function of

$$\Theta_{EDI}(e(s), x) = \frac{\exp c}{\tau} \int_{0}^{x} \int_{f}^{f+t} \boldsymbol{e}\left(s\right) ds dt. \tag{9}$$

*2) **Spatial blur indication deblurring**:* From the spatial perspective, the relation between a blurry frame $\boldsymbol{B}$ and a sharp frame $\boldsymbol{I}$ can be constructed as:

$$\boldsymbol{B} = \boldsymbol{I} \otimes k + n. \tag{10}$$

where $n$ denotes some additive white Gaussian noise and $k$ denotes the blur kernel. Thus, the deblurring process can be described as an inverse convolution process, which can be regarded as a mapping function to map $\boldsymbol{B}$ and $\boldsymbol{I}$. Each pixel $(i, j)$ on the sharp image $\boldsymbol{I}$ is decided by the feature of the same pixel $(i, j)$ on the blurry image $\boldsymbol{B}$ and the weights of the

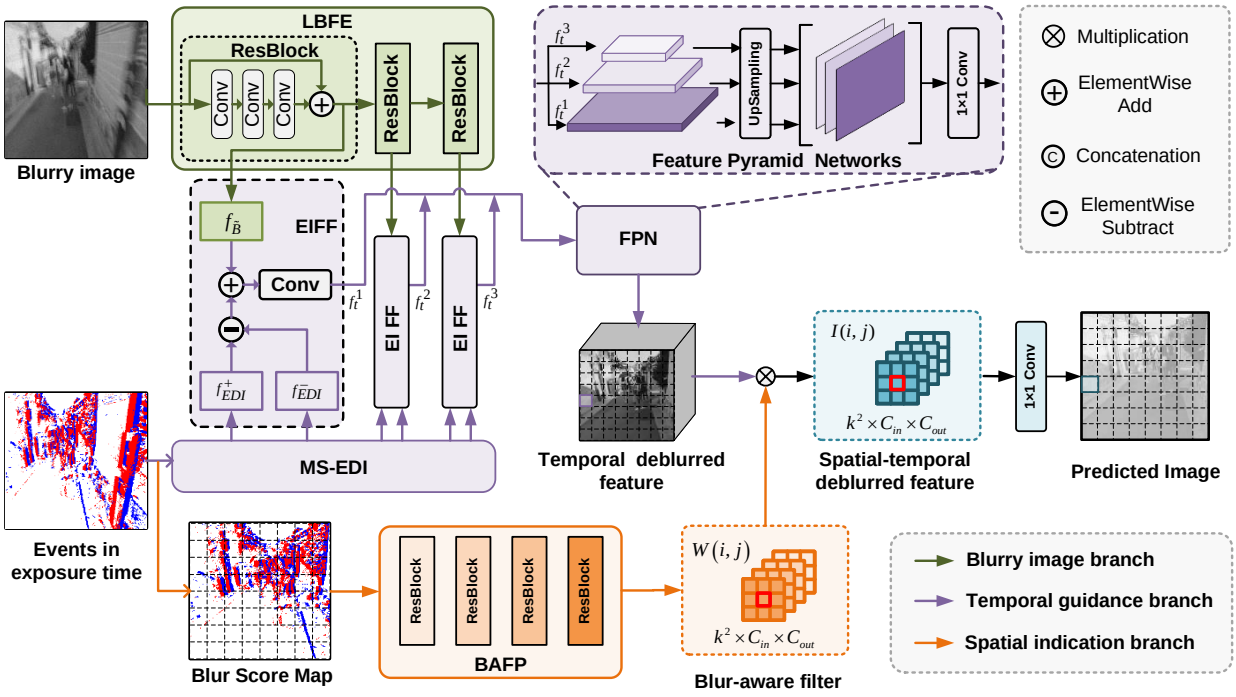



Fig. 3. The pipeline of the proposed MBNet for reconstructing blurry images with events. $f_{\tilde{B}}$ and $f_{EDI}^{\pm}$ are respectively the event temporal residual feature and the blurry image feature. Given a blurry image and its associated events in the exposure interval, MBNet first learns the blurred features in the logarithmic domain from the blurred image by Logarithmic Blurry Feature Encoder (LBFE). Then we use the Mutil-scale event-based double integral module (MS-EDI) to learn the temporal residual feature from spilt events, and utilize the Blur-Aware Filter Prediction (BAFP) module to learn the spatial blurred filter from the blur score map. Finally, the blurred image features are fused with the temporal residual features through the Event and Image Feature Fusion (EIFF) module and restored to a clear image after passing through the spatial blur filter.

corresponding filter. The deblurring process can be formulated as:

$$\boldsymbol{I}(i,j) = \Phi\left(f_B(i,j), W(i,j)\right), \tag{11}$$

where $\boldsymbol{I}(i,j)$ denotes the pixel value at $(i,j)$ on sharp image. $f_B(i,j)$ denotes the feature of pixel $(i,j)$ on the blurry image. $W(i,j)$ is the weight of the filter for pixel $(i,j)$. $\Phi(.)$ is the feature mapping function to calculate the pixel value.

Refer to the Eq. 2, as events represent the extremely slight value change at each pixel during exposure time, they indicate the precise blur and form the blur score map as:

$$M_{blur}^{h\times w}(i,j) = \int_{t\in\tau} \boldsymbol{e}(i,j,t)\,dt. \tag{12}$$

The filter can be learned from the blur score map. Thus, we update Eq. 11 as:

$$\boldsymbol{I}(i,j) = \Phi\left(f_B(i,j), W\left(\boldsymbol{e}(s),\tau\right)(i,j)\right). \tag{13}$$

In summary, the relations between a sharp frame and a blurry frame with events within the exposure time are presented in Eq. 8 and in Eq. 13, where events are treated as the residual frames and the blur score map respectively.

### B. Network Architecture

*1) **Overview**:* The overall architecture consists of three branches, as illustrated in Fig .3. The blurry image branch is responsible for extracting features of blurred images, the temporal guidance branch learns temporal residual guidance from events, and the spatial indication branch learns spatial blur indication from events. The process of deblurring involves the following two steps. In the first stage, the MS-EDI module eliminates extreme blur under the guidance of temporal residuals. In the second stage, the BAFP module achieves more refined deblurring by estimating the blur-aware kernel.

*2) **Temporal guidance branch**:* This branch is designed to use temporal residuals of events to guide image deblurring.

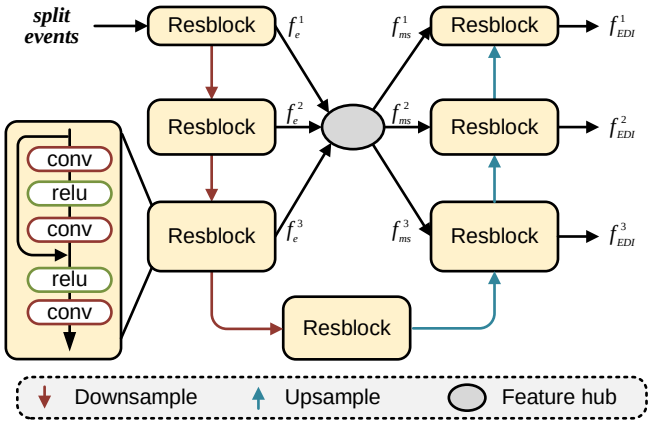



Fig. 4. Mutil-scale event-based double integral (MS-EDI) module. The feature hub is designed to increase the flexibility of feature flow. It takes three feature inputs of different scales, which are then added together as the output after being downsampled or upsampled to the same size.

Blurry images and sharp images can be bridged by event temporal residuals, as described in Eq. 8.

**Logarithmic Blurry Feature Encoder:** As shown in Fig. 3, the logarithmic blurring feature $\tilde{\boldsymbol{B}}$ (shown as "$f_{\tilde{B}}$" in Fig. 3 ) in Eq. 8 is learned by Logarithmic Blurry Feature Encoder (LBFE), which is designed as three stacked residual convolution blocks (Resblocks) to ensure the compactness of the network.

**Mutil-scale Event-based Double Integral Module:** In order to enable compatibility with convolutional neural network layers, the event sequences need to be transformed into tensor-like formats. We accumulate events by polarity into $N$ temporal bins. Then, a $2N \times H \times W$ tensor of events is formed. The mapping function $\Theta_{EDI}$ in Eq. 8 is learned by the Mutil-scale Event-based Double Integral Module (MS-EDI), which has a similar structure to Unet [49]. As shown in Fig. 4, in MS-EDI, the split events yield three features of different scales: $f_e^1$, $f_e^2$, and $f_e^3$, after passing through three resblocks. Within the feature hub, $f_e^2$, and $f_e^3$ are resized to the same scale of $f_e^1$. These three are then added together to form a multi-scale feature. This multi-scale feature is further resized into three distinct scales for output (i.e., $f_{ms}^1$, $f_{ms}^2$, and $f_{ms}^3$). Finally, these outputs are send to resblock and produce the final features:$f_{EDI}^1$, $f_{EDI}^2$, and $f_{EDI}^3$. The MS-EDI module enhances the flexibility of feature flow, ensuring thorough integration among features of various scales. This module extracts event double integral features at different scales, as shown by "$f_{EDI}^{\pm}$" in Fig. 3 .

**Event and Image Feature Fusion:** After the LBFE and MS-EDI modules, we obtain the blurry image features $F_B$ and event temporal residual features $f_{EDI}^{\pm}$. These two types of features are then fused in the Event and Image Feature Fusion module (EIFF). In this module, $f_{EDI}^{-}$ and $f_{EDI}^{+}$ are first subtracted and then added to $F_B$, and finally a convolution is performed to integrate all features.

In particular, we conduct the function Eq. 8 at different scales to excavate more constraint information between sharp and blurry images. All the deblurred features at different scales as shown "$f_t^1, f_t^2, f_t^3$" in Fig. 4 are upsampled hierarchically through the Feature Pyramid Networks (FPN), and fused by a 3×3 convolution.

*3) **Spatial indication branch**:* The spatial distribution of events reflects the degree of blurriness in different regions of a blurry image. Therefore, utilizing blur indication information can predict spatial derivative reconstruction filters, enabling the mapping of pixels from a blurred image to their correct positions in a sharp image. Blurry images and sharp images can be bridged by event spatial blur indication, as described in Eq. 13.

**Blur-Aware Filter Prediction:** The Blur-Aware Filter Prediction module (BAFP) is composed as four stacked residual convolution blocks, and it takes the blur score map $M$ described in Eq. 12 as input. This module is designed for spatial blur filter prediction as:

$$W^{h\times w} = BAFP\left(M_{blur}^{h\times w}(\boldsymbol{e}(s),\tau)\right), \tag{14}$$

for each pixel $(i,j)$, the reconstruction filter $W^{h\times w}$ at this pixel is denoted as $W(i,j)^{k\times k\times C_{in}\times C_{out}}$, where $C_{in}$ is the number of channels of the temporal deblurred feature, $C_{out}$ is the number of channels of the predicted sharp image and $k$ represents the size of the convolution kernel. The module is effective for the reconstruction filter prediction since spatial information is integrated into the feature encoder part.

In the spatial indication branch, spatial blurring filter filters the temporal deblurring features to obtain spatio-temporal deblurring features. Finally, these features are sent to a 1×1 convolution kernel, and the final clear image is output.

### C. *Training Loss*

To train the network, a combination of an $L_1$ image reconstruction loss and a frequency reconstruction loss function is used.

The reconstruction loss $L_{rec}$ is defined as follows:

$$L_{rec} = \|I - \tilde{I}\|_1, \tag{15}$$

where $I$ and $\tilde{I}$ refer to the ground truth sharp image and the estimated image respectively.

The frequency reconstruction measures the $L_1$ distance between the ground-truth and deblurred images in the frequency domain as follows:

$$L_{fr} = \|FFT(I_{\downarrow}) - FFT(\tilde{I})\|_1, \tag{16}$$

where $FFT$ denotes the fast Fourier transform that transfers image signals to the frequency domain.

The final loss function for training our network is determined as follows:

$$L_{total} = L_{rec} + \lambda * L_{fr}, \tag{17}$$

where $\lambda$ is a hyperparameter.

## IV. Experiments

### A. *Datasets.*

We evaluate the proposed method with three various datasets, including synthetic, semi-synthetic, and real-world ones.

**Synthetic dataset**. We adopt the GoPro dataset as a benchmarking dataset released by [34], which contains a sharp image set, a blurry image set, and a event stream set. The sharp image set contains 25,650 sharp frames with the resolution of 320×180 from the original GoPro dataset [53] with downsampling and grayscale conversion. Blurry images are synthesized by averaging 17 consecutive interpolated frames. Event streams are synthesized as well, using ESIM [54] with added noise to imitate real scenarios. The training set contains images and events from 240 video sequences and the testing set contains the rest from 30 video sequences.

**Semi-synthetic dataset**. The HQF dataset [55] contains real-world events and sharp ground-truth frames that are well-exposed to avoid motion blur. The blurry frames are synthesized using these clear frames. Same as [34, 35], we first increase the frame rate of sharp images by [56] and then average over 49 consecutive sharp image frames to synthesize the motion blur. In order to test the generalization ability of the methods, we take GoPro trained model and directly evaluate it on the test set of the HQF dataset.

TABLE I
QUANTITATIVE COMPARISON OF THE PROPOSED METHOD TO STATE-OF-THE-ART MOTION DEBLURRING METHODS ON THE SYNTHETIC DATASET, SEMI-SYNTHETIC DATASET, AND REAL-WORLD DATASET.

| Method | Publication | Inputs | | Synthetic dataset | | | Semi-synthetic dataset | | | Real-world dataset | | | Params. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Image | Events | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | |
| SRNN[6] | CVPR2018 | ✓ | ✗ | 25.32 | 0.6989 | 0.391 | 23.32 | 0.6875 | 0.401 | 35.10 | 0.9618 | 0.213 | 6.8M |
| HINet[3] | CVPRW2021 | ✓ | ✗ | 25.46 | 0.7051 | 0.321 | 23.51 | 0.6752 | 0.393 | 35.58 | 0.9651 | 0.191 | 10.6M |
| MIMOUNet[5] | ICCV2021 | ✓ | ✗ | 25.71 | 0.7064 | 0.311 | 23.08 | 0.6745 | 0.408 | 35.84 | 0.9672 | 0.182 | 16.1M |
| Restormer[15] | CVPR2022 | ✓ | ✗ | 26.11 | 0.7076 | 0.301 | 25.05 | 0.7190 | 0.399 | 36.21 | 0.9623 | 0.193 | 25.3M |
| E2VID[50] | CVPR2019 | ✗ | ✓ | 12.55 | 0.3569 | 0.664 | 8.86 | 0.4836 | 0.531 | 8.961 | 0.4942 | 0.564 | 41.8M |
| E2SRI[51] | CVPR2020 | ✗ | ✓ | 11.77 | 0.4350 | 0.610 | 10.98 | 0.5368 | 0.537 | 9.867 | 0.5681 | 0.536 | 40.9M |
| BHA[26] | CVPR2019 | ✓ | ✓ | 29.06 | 0.8653 | 0.211 | 24.64 | 0.7230 | 0.391 | 36.52 | 0.9642 | 0.135 | 12.3M |
| eSL-Net[34] | ECCV2020 | ✓ | ✓ | 30.23 | 0.8703 | 0.203 | 24.57 | 0.7370 | 0.371 | 36.28 | 0.9664 | 0.152 | **1.32M** |
| EFNett[28] | ECCV2022 | ✓ | ✓ | 26.41 | 0.7280 | 0.309 | 24.31 | 0.6986 | 0.375 | 38.12 | <u>0.9753</u> | 0.123 | 14.4M |
| eSL-Net++[35] | TPAMI2023 | ✓ | ✓ | 31.02 | 0.8849 | <u>0.181</u> | 24.22 | 0.7185 | <u>0.351</u> | 38.13 | 0.9693 | 0.126 | <u>1.41M</u> |
| REFID[52] | CVPR2023 | ✓ | ✓ | <u>31.56</u> | <u>0.8876</u> | 0.183 | <u>26.32</u> | <u>0.7561</u> | 0.356 | <u>38.16</u> | 0.9652 | <u>0.119</u> | 15.9M |
| DeblurSR[43] | AAAI2024 | ✓ | ✓ | 27.92 | 0.7960 | 0.252 | 25.85 | 0.7310 | 0.398 | 38.09 | 0.9671 | 0.123 | 9.5M |
| MBNet(Ours) | - | ✓ | ✓ | **33.15** | **0.9164** | **0.153** | **28.06** | **0.8185** | **0.273** | **38.21** | **0.9796** | **0.105** | 2.34M |

[1] The **bold** and the <u>underline</u> represent the best and second-best performance, respectively.


Fig. 5. Visual comparison with other methods on the synthetic dataset. Our method generates sharper and visually-faithful result.

**Real-world dataset**. We adopt the REBlur dataset [28] to validate the effectiveness of the proposed network in real-world scenarios with real blurry images. This dataset contains 1469 groups of blurry-sharp images pairs from 36 sequences, which are collected in a high-precision optical laboratory with stable illumination in a two-shot way. The training set contains 486 pairs and the testing set contains 983 pairs.

### *B. Implementation Details.*

We implement the proposed network using PyTorch on an NVIDIA Titan-RTX 3090 GPU. Our network is trained by Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, on 160×160 randomly cropped patches with the batches of size 4. The initial learning rate is $1 \times e^{-4}$, which decreases by a factor of 10 every 10 epochs. We train our proposed network for 50 epochs. The hyperparameter $\lambda$ of the loss function is set to 0.1. In the spatial indication branch, the number of input channels $C_{in}$ is set to 32, and the number of output channels $C_{out}$ is set to 3 for color images and 1 for grayscale images.

### *C. Comparisons with the State-of-the-Art*

In this section, we evaluate the proposed network and compare it with recent motion deblurring methods on the synthetic dataset, semi-synthetic dataset, and the real-world

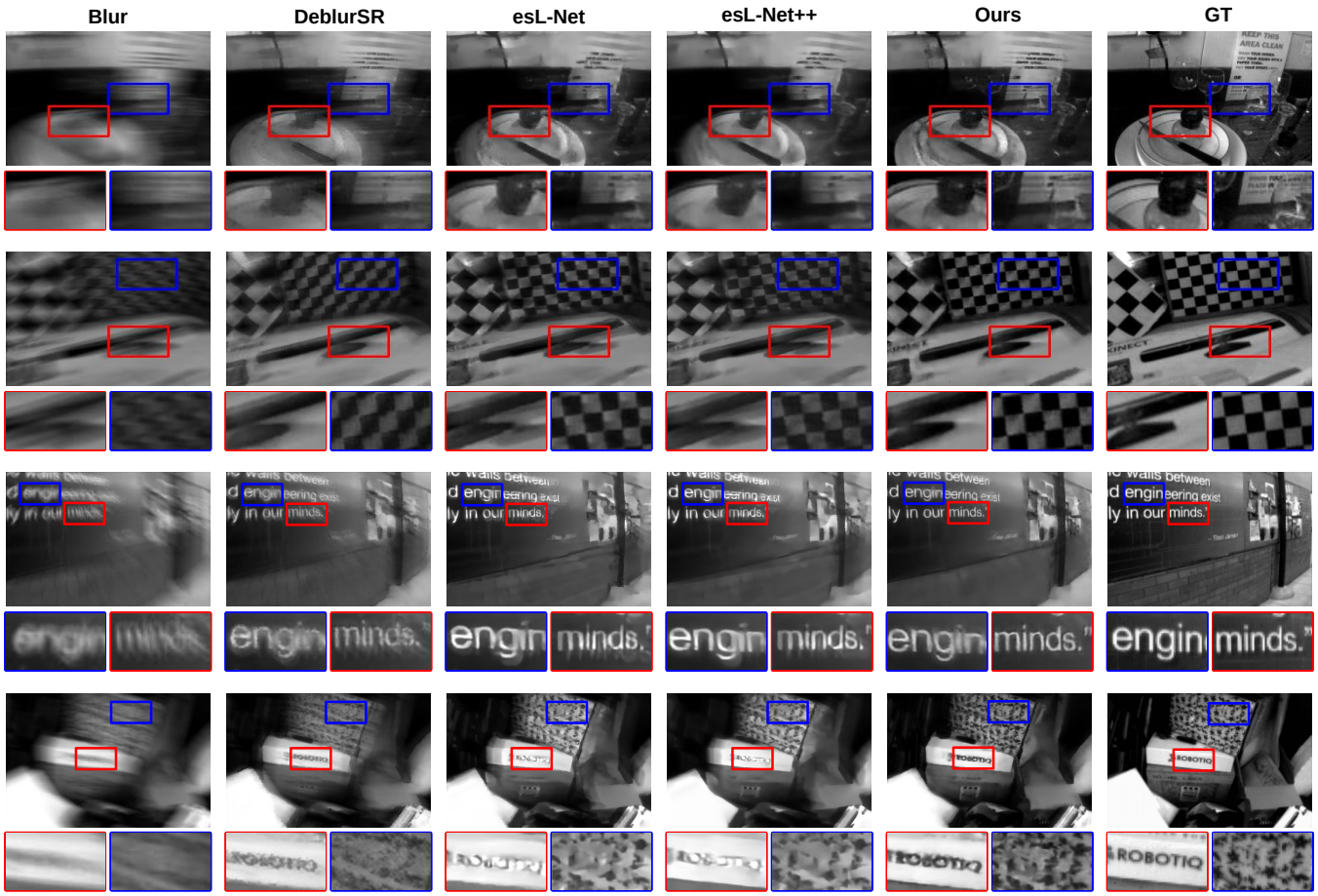



Fig. 6. Visual comparison with other methods on the semi-synthetic dataset. Our method effectively removes blur while preserving the fine image details.

dataset. The performance of our methods is quantitatively evaluated by PSNR, SSIM [57] and LPIPS [58] and qualitative results are presented for further analysis.

*1) Quantitative results:* We compare our approach with state-of-the-art deblurring methods, including event-only intensity restoration methods [3, 5, 6, 15], frame-only deblurring methods [50, 51] and image-event fusion based methods [26, 28, 34, 35, 43, 52]. The results are reported in Tab. I, compared with the second-best model(REFID [52]), our method achieves remarkable performance, with 1.59 dB improvement in PSNR, 2.88% improvement in SSIM and 2.8% reduction in LPIPS on synthetic dataset. When compared with all methods on the semi-synthetic dataset, our proposed method achieves more impressive performance, with 1.74 dB improvement in PSNR, 6.24% improvement in SSIM and 8.3% reduction in LPIPS compared to the second best method. Since we employed the pre-trained model for the evaluation and didn't fine-tune our model, this improvement demonstrates that the generalization ability of our model is better compared to other methods. This can be attributed to the reconstruction process in the BAFP module since the EDI feature varies in different event camera settings and different surroundings. To evaluate the motion deblurring performance of our method in real blur situations, we also compare our network with other methods on the real-world dataset. Due to the large gap between the distribution of real data and synthetic data, we use the model fine-tuned on the real dataset for comparison. As shown in Tab. I, our model surpasses the best existing method, with 0.05 dB improvement in PSNR, 0.4% improvement in SSIM and 1.4% reduction in LPIPS.

*2) Qualitative results:* We further evaluate the performances qualitatively on the synthetic dataset, the semi-synthetic dataset, and the real-world dataset. The results are visualized in Fig. 5 for the synthetic dataset, Fig. 6 for the semi-synthetic dataset and Fig. 7 for the real-world dataset. Our proposed network can reconstruct sharper details than other methods, while deblurSR [43] recovers unclear images with ghosting, and eSL-Net [34] eliminates the image details with the blur, producing an unfavorable over-smoothing effect. Our methods can reconstruct sharper edges while eSL-Net++ [35] predicts unclear edges and slight slur in the corner.

### D. Model Analyses

We conducted extensive ablation experiments to validate the effectiveness of the designed architecture and each component in our network. All the models were trained on the image patches of size 160×160, with the same optimizer setting, and

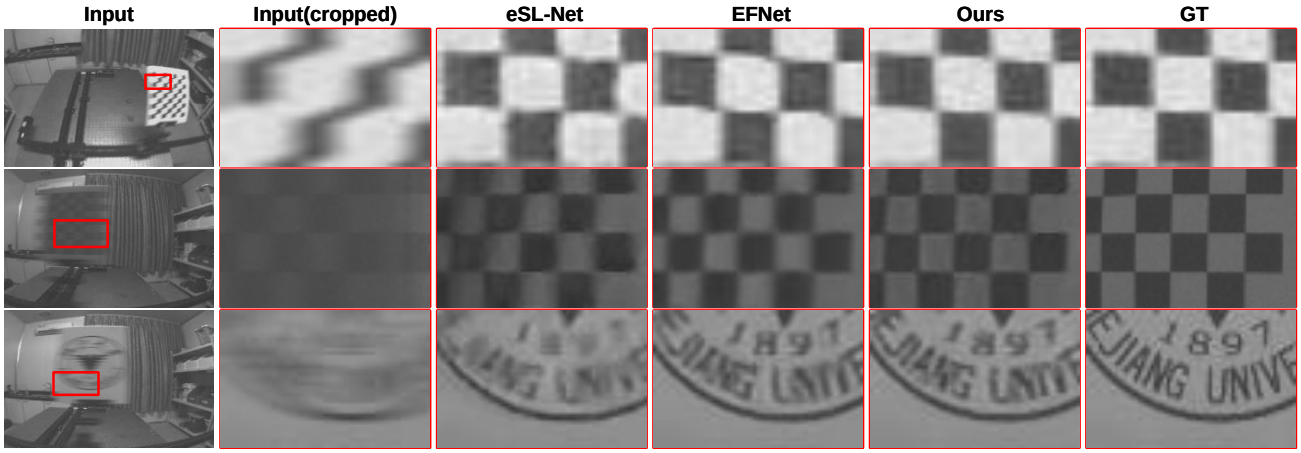



Fig. 7. The visual results of motion deblurring methods on the real-world dataset. The image reproduction quality of our method is more faithful to the ground truth than other methods.

TABLE II
ABLATION STUDY ON FRAMEWORK SCHEME AND LOSS FUNCTION.

| Arch. | Synthetic dataset | | | Semi-synthetic dataset | | | Real-world dataset | | | Params. |
|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | |
| MS-EDI only | 32.08 | 0.8856 | 0.264 | 25.52 | 0.7628 | 0.361 | 36.93 | 0.9433 | 0.166 | 1.70M |
| BAFP only | 32.26 | 0.8964 | 0.233 | 27.20 | 0.7838 | 0.295 | 37.52 | 0.9562 | 0.155 | 1.05M |
| MBNet* | 32.61 | 0.9011 | 0.201 | 27.56 | 0.7932 | 0.289 | 37.82 | 0.9613 | 0.145 | 2.34M |
| MBNet | 33.15 | 0.9164 | 0.153 | 28.06 | 0.8185 | 0.273 | 38.21 | 0.9796 | 0.105 | 2.34M |

[1] *MBNet** indicates that the input of the spatial blur guidance branch is replaced with the original blurred image from the event blur score map.

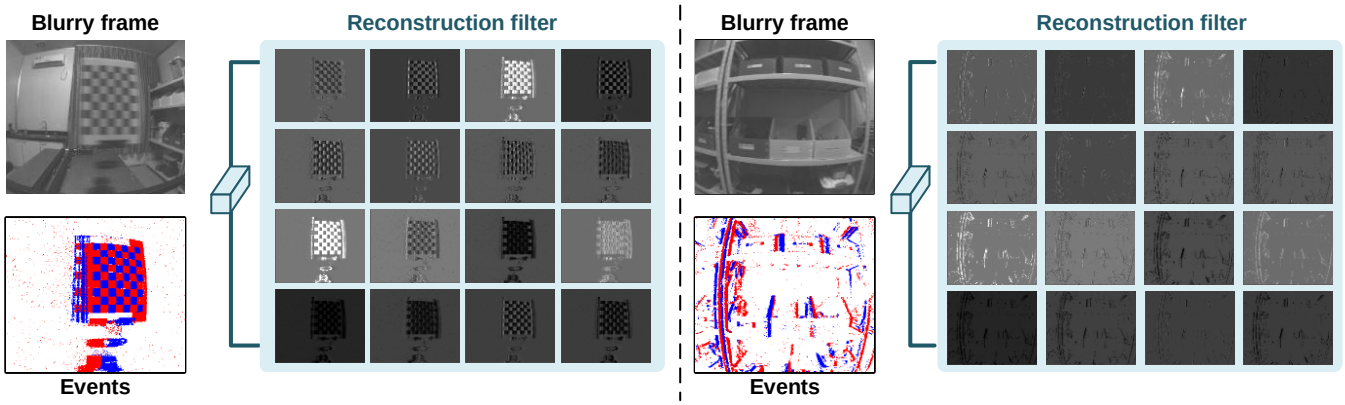



Fig. 8. The visual examples of the reconstruction filter. The 16 arranged feature maps represent the weights of the first 16 channel dimensions in the reconstruction filter of the whole image.

tested on the synthetic dataset. Next, we analyze the results of each experiment individually.

**MS-EDI vs. BAFP**. As mentioned in Sec. III-A, in the MS-EDI module, events are provided as residual frames to construct the temporal-spatial interaction, while in BAFP, events are provided as motion blur indicator for the reconstruction filter prediction. The ablation experiment in Tab. II shows that the proposed combined network surpasses both MS-EDI-only and BAFA-only methods, demonstrating the event information leverage from both perspectives can be complementary and boost the quality of the predicted sharp images. It is worth mentioning that the MS-EDI-only architecture achieves less impressive results compared to BAFP-only architecture in network generalization, i.e., the performance of the MS-EDI-only method trained on synthetic datasets decreased by 6.56 dB on the semi-synthetic dataset, while the BAFP only method decreased by 5.06 dB on the semi-synthetic dataset. This is because the learned event-based double integral feature can be various in different scenarios and camera settings. By incorporating our proposed spatial blur guidance, the generalization ability of the model can be effectively enhanced.

**Blurred image vs. Event blur score map**. To compare the effectiveness of learning blurred spatial features from the original image and from the event blur score map, we conduct an additional ablation experiment. Specifically, we replace the input of the event spatial indication branch from the event blur

TABLE III
ABLATION STUDY OF ARCHITECTURES WITH DIFFERENT COMPONENTS OF OUR METHOD.

| MS-EDI | | | BAFP | PSNR↑ | SSIM↑ | LPIPS↓ | Params. |
|---|---|---|---|---|---|---|---|
| *MS.* | *RE.* | *FB.* | *#ResBlock* | | | | |
| | | | 4 | 31.72 | 0.8887 | 0.253 | 1.93M |
| | ✓ | | 4 | 31.91 | 0.9029 | 0.201 | 1.93M |
| ✓ | ✓ | | 4 | 32.41 | 0.9102 | 0.182 | 2.34M |
| ✓ | ✓ | ✓ | 4 | 33.15 | 0.9164 | 0.153 | 2.34M |
| ✓ | ✓ | ✓ | 5 | 33.19 | 0.9169 | 0.146 | 3.07M |
| ✓ | ✓ | ✓ | 6 | 33.25 | 0.9173 | 0.139 | 3.72M |

[1] *MS.* refers to the multi-scale feature fusion strategy.
[2] *RE.* refers to reversed events in the temporal dimension as input.
[3] *FB.* refers to the feature hub design in the EDI feature encoder.

score map to the original blurred image. The results, as shown in Tab. II "MBNet*", indicate that the algorithm's performance has improved compared to MS-EDI only but is not as good as using event spatial blur indicators. This may be because image blur is not as intuitive as events. The relationship between events and blur is clear: blur occurs where there is fast movement, and areas with faster movement have greater pixel intensity changes, thus generating more events. Conversely, frame-based images contain not only blurred spatial information but also details such as scene structure and texture. Extracting discriminative deblurring features from these images can be challenging.

**Ablation studies of the component design**. We conduct the ablation studies in Tab. III to analyze the contribution of components of two modules. In the MS-EDI module, our multi-scale fusion strategy and the multi-scale EDI feature extraction design improve the performance, as well as events temporal reversion. In the BAFP module, the design of more stacked convolution blocks yields improvement but enlarges the network capacity, which indicates that a simple convolution network is sufficient for the reconstruction filter prediction.

The filter map is visualized in Fig. 8. The weight maps of the filter in the first 16 channel dimensions are successively arranged for better visualization. The weights of the filters are varied in different sharp and blurry areas, which demonstrates that the predicted filter can learn various reconstruction operators in different areas from the event-based blur indication. Besides, the filter map is textured in blurry areas and plain in sharp areas since the complex filter is generated for context reconstruction with severe motion blur in the image.

### E. Robustness to Event Noise

To verify the robustness of our proposed method, we conduct quantitative comparisons of deblurring under different events noise levels on the synthetic dataset. Regarding event noise, we add uniform noise with noise ratios ranging from 20% to 100% and evaluate the performance of REFID, eSL-Net++ and our MBNet. Quantitative results are shown in Figure 9. The quantitative results confirm that MBNet outperforms REFID and eSL-Net++ at different noise levels, verifying the better robustness of MBNet.

TABLE IV
COMPARISON OF BLURRY IMAGE SUPER-RESOLUTION METHODS ON THE SYNTHETIC DATASET AND SEMI-SYNTHETIC DATASET.

| Method | Mode | GoPro dataset | | HQF dataset | |
|---|---|---|---|---|---|
| | | PSNR↑ | SSIM ↑ | PSNR↑ | SSIM↑ |
| MIMOUNet+RCAN4× | Cascade | 22.71 | 0.4654 | 22.83 | 0.6656 |
| Restormer+RCAN4× | Cascade | 22.83 | 0.4653 | 24.23 | 0.6810 |
| EFNet+RCAN4× | Cascade | 22.78 | 0.4657 | 23.54 | 0.6657 |
| DeblurSR [43] | Unified | 24.22 | 0.6176 | <u>25.11</u> | <u>0.7343</u> |
| eSL-Net [34] | Unified | 25.57 | 0.6774 | 24.04 | 0.7287 |
| eSL-Net++ [35] | Unified | <u>25.93</u> | <u>0.6876</u> | 23.76 | 0.7203 |
| Ours | Unified | **26.73** | **0.7082** | **27.01** | **0.7787** |

[1] The **bold** and the <u>underline</u> represent the best and second-best performance, respectively.

### F. Extension to Blurry Image Super-Resolution

Our proposed network can be extended to blurry image super-resolution since the blur-aware reconstruction network can expand the pixels by adjusting the output channels of the reconstruction filter.

*1) Datasets for Super-Resolution:* We keep the same training dataset and testing datasets as [34, 35] for a fair comparison, including the GoPro dataset and HQF dataset. The GoPro dataset consists of SR image and LR image pairs and synthetic events. The original HQF dataset contains real sharp images and events and we generate the blurry images and SR images following the generation instruction in [34].

*2) Comparisons with the State-of-the-Art:* We compare our approach with state-of-the-art blurry image super-resolution methods including the end-to-end methods, i.e., DeblurSR [43], eSL-Net [34], and eSL-Net++ [35], and the two-stage Deblur-then-SR methods by cascading the motion deblurring with the image SR, i.e., MIMOunet [5], Restormer [15], and EFNet [28]. The task of the image SR is fulfilled by RCAN [59].

*Quantitative results:* We report the quantitative result in Tab. IV, our model performs favorably against other methods. Compared to the previous best-performing method, our method achieves performance improvement by a large margin on both datasets, i.e., 0.8dB/0.0206 on the GoPro dataset and 1.9dB/0.0444 on the HQF dataset. The results demonstrate that our model is effective in blurry image super-resolution task as well. Although our method is trained on a synthetic GoPro dataset, it still performs well on real data, indicating the generalization of our method from synthetic scenes to the real world.

*Qualitative results:* Visual results are illustrated in Fig. 10. Compared to cascade methods, e.g., MIMOUNet, end-to-end methods can better eliminate motion blur, which is consistent with quantitative evaluations. Our method can give precise high resolution (HR) reconstructions that exhibit the most similar appearances to the ground-truth HR images, primarily because we leverage events to establish the connection between clear and blurred images from two distinct perspectives.

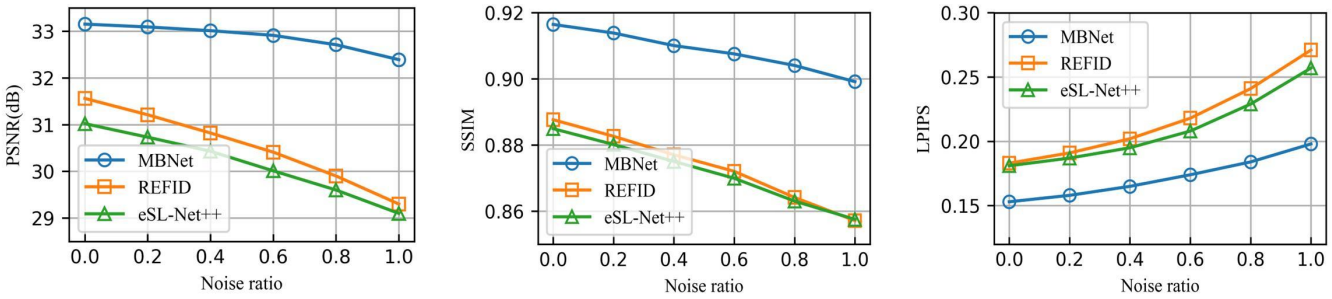



Fig. 9. Quantitative comparisons of different deblurring methods under different levels of event noise on the synthetic dataset.

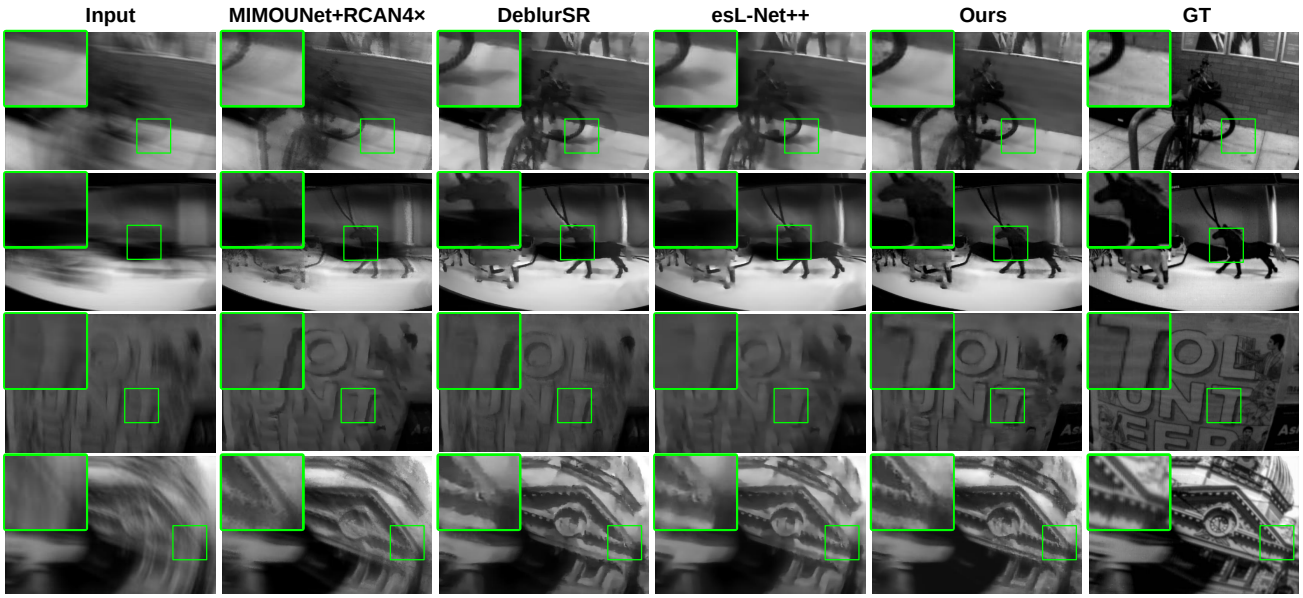



Fig. 10. The visual results of blurry image super-resolution methods on GoPro Dataset.

### G. Limitation and Future Work

In this paper, we present a novel event-based approach for image deblurring, which utilizes temporal residual guidance and spatial blur indications. Experiments on various datasets demonstrate the effectiveness of our method. However, our approach has the following two limitations.

One limitation is that our method is limited to utilizing event information within the exposure time and has not yet involved feature extraction and learning between multiple frames. This limitation may compromise the algorithm's performance when dealing with scenes featuring occlusions and complexities. In future research, we aim to further improve the deblurring algorithm's performance by introducing long-term temporal context features.

Another limitation is that the image resolution of event cameras is far lower than that of RGB cameras (346x260 or lower). This limitation seriously affects the application scope of event cameras. Although we can restore higher-quality images through super-resolution technology, in situations with strict resolution requirements, the restored images often appear distorted, and the effect is not ideal. Therefore, in our future research work, we plan to integrate high-resolution RGB cameras with event cameras to create a hybrid system. We hope that through this system, we can achieve higher-resolution deblurring tasks and capture more real-world scene data to further enrich and improve our research work.

## V. Conclusion

In this paper, we a novel perspective, i.e., employing the blur indication provided by events for spatially differentiated image reconstruction for motion deblurring. Based on this, we propose a network, MBNet, which integrates temporal residual guidance and spatial blur indication to fully utilize the rich information provided by event cameras. Moreover, our network can be easily extended to blurry image super-resolution. Experiments show that our proposed network outperforms state-of-the-art approaches in motion deblurring and the designed blur-aware reconstruction module is effective as well.

## References

[1] B. Sheng, P. Li, X. Fang, P. Tan, and E. Wu, "Depth-aware motion deblurring using loopy belief propagation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 955–969, 2019.

[2] Y. Mao, Z. Wan, Y. Dai, and X. Yu, "Deep idempotent network for efficient single image blind deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 1, pp. 172–185, 2022.

[3] L. Chen, X. Lu, J. Zhang, X. Chu, and C. Chen, "Hinet: Half instance normalization network for image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops(CVPRW)*, 2021, pp. 182–192.

[4] S. Zamir, A. Arora, S. Khan, M. Hayat, F. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14 816–14 826.

[5] S. Cho, S. Ji, J.-P. Hong, S. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *IEEE International Conference on Computer Vision (ICCV)*, 2021, pp. 4621–4630.

[6] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2018, pp. 8174–8182.

[7] J. Dong, J. Pan, Z. Yang, and J. Tang, "Multi-scale residual low-pass filter network for image deblurring," in *IEEE International Conference on Computer Vision (ICCV)*, 2023, pp. 12 311–12 320.

[8] Y. Bai, H. Jia, M. Jiang, X. Liu, X. Xie, and W. Gao, "Single-image blind deblurring using multi-scale latent structure prior," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2033–2045, 2019.

[9] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. Lau, and M.-H. Yang, "Dynamic scene deblurring using spatially variant recurrent neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2018, pp. 2521–2529.

[10] P. Wieschollека, M. Hirsch, B. Schölkopf, and H. Lensch, "Learning blind motion deblurring," in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 231–240.

[11] S. Nah, S. Son, and K. Lee, "Recurrent neural networks with intra-frame iterations for video deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 8094–8103.

[12] Z. Zhong, Y. Gao, Y. Zheng, and B. Zheng, "Efficient spatio-temporal recurrent neural network for video deblurring," in *European Conference on Computer Vision(ECCV)*, 2020, pp. 191–207.

[13] S. Zhou, J. Zhang, J. Pan, H. Xie, W. Zuo, and J. Ren, "Spatio-temporal filter adaptive network for video deblurring," in *IEEE International Conference on Computer Vision(ICCV)*, 2019.

[14] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3025–3036, 2020.

[15] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[16] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "Maxim: Multi-axis mlp for image processing," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5759–5770.

[17] Y. Li, Y. Fan, X. Xiang, D. Demandolx, R. Ranjan, R. Timofte, and L. Gool, "Efficient and explicit modelling of image hierarchies for image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023.

[18] F.-J. Tsai, Y.-T. Peng, C. Tsai, Y. Lin, and C. Lin, "Banet: A blur-aware attention network for dynamic scene deblurring," *IEEE Transactions on Image Processing*, vol. 31, pp. 6789–6799, 2022.

[19] X. Zhang, T. Wang, R. Jiang, L. Zhao, and Y. Xu, "Multi-attention convolutional neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1986–1997, 2021.

[20] K. Zhang, W. Ren, W. Luo, W. Lai, B. Stenger, M. Yang, and H. Li, "Deep image deblurring: A survey," *International Journal of Computer Vision*, vol. 130, no. 9, p. 2103–2130, 2022.

[21] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[22] Z. Chen, J. Wu, J. Hou, L. Li, W. Dong, and G. Shi, "Ecsnet: Spatio-temporal feature learning for event camera," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 2, pp. 701–712, 2022.

[23] M. Liu and T. Delbruck, "Edflow: Event driven optical flow camera with keypoint detection and adaptive block matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 5776–5789, 2022.

[24] Z. Liu, J. Wu, G. Shi, W. Yang, W. Dong, and Q. Zhao, "Motion-oriented hybrid spiking neural networks for event-based motion deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[25] L. Pan, R. Hartley, C. Scheerlinck, M. Liu, X. Yu, and Y. Dai, "High frame rate video reconstruction based on an event camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 5, pp. 2519–2533, 2022.

[26] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, "Bringing a blurry frame alive at high frame-rate with an event camera," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6813–6822.

[27] W. Shang, D. Ren, D. Zou, J. Ren, P. Luo, and W. Zuo, "Bringing events into video deblurring with non-consecutively blurry frames," in *IEEE International Conference on Computer Vision(ICCV)*, 2021, pp. 4531–4540.

[28] L. Sun, C. Sakaridis, J. Liang, Q. Jiang, K. Yang, P. Sun, Y. Ye, K. Wang, and L. Gool, "Event-based fusion for motion deblurring with cross-modal attention," in *European Conference on Computer Vision(ECCV)*, 2022, pp. 412–428.

[29] H. Cho, Y. Jeong, T. Kim, and K.-J. Yoon, "Non-coaxial event-guided motion deblurring with spatial alignment," in *IEEE International Conference on Computer Vision (ICCV)*, 2023, pp. 12 492–12 503.

[30] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, "Deep video deblurring for hand-held cameras," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 237–246.

[31] M. Jin, G. Meishvili, and P. Favaro, "Learning to extract a video sequence from a single motion-blurred image," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2018, pp. 6334–6342.

[32] S. Deng, W. Ren, Y. Yan, T. Wang, F. Song, and X. Cao, "Multi-scale separable network for ultra-high-definition video deblurring," in *IEEE International Conference on Computer Vision (ICCV)*, 2021, pp. 14 010–14 019.

[33] S. Lin, J. Zhang, J. Pan, Z. Jiang, D. Zou, Y. Wang, J. Chen, and J. Ren, "Learning event-driven video deblurring and interpolation," in *European Conference on Computer Vision (ECCV)*, 2020.

[34] B. Wang, J. He, L. Yu, G.-S. Xia, and W. Yang, "Event enhanced high-quality image recovery," in *European Conference on Computer Vision(ECCV)*. Springer, 2020, p. 155–171.

[35] L. Yu, B. Wang, X. Zhang, H. Zhang, W. Yang, J. Liu, and G. Xia, "Learning to super-resolve blurry images with events," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[36] P. Vitoria, S. Georgoulis, S. Tulyakov, A. Bochicchio, J. Erbach, and Y. Li, "Event-based image deblurring with dynamic motion awareness," in *European Conference on Computer Vision*. Springer, 2022, pp. 95–112.

[37] W. Weng, Y. Zhang, and Z. Xiong, "Event-based blurry frame interpolation under blind exposure," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023.

[38] X. Zhang and L. Yu, "Unifying motion deblurring and frame interpolation with events," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 17 744–17 753.

[39] X. Zhang, L. Yu, W. Yang, J. Liu, and G. Xia, "Generalizing event-based motion deblurring in real-world scenarios," in *IEEE International Conference on Computer Vision (ICCV)*, 2023.

[40] H. Zhang, L. Zhang, Y. Dai, H. Li, and P. Koniusz, "Event-guided multi-patch network with self-supervision for non-uniform motion deblurring," *International Journal of Computer Vision*, vol. 131, no. 2, pp. 453–470, 2023.

[41] L. Sun, C. Sakaridis, J. Liang, P. Sun, J. Cao, K. Zhang, Q. Jiang, K. Wang, and L. Van Gool, "Event-based frame interpolation with ad-hoc deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[42] S. Lin, Y. Zhang, D. Huang, B. Zhou, X. Luo, and J. Pan, "Fast event-based double integral for real-time robotics," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 796–803.

[43] C. Song, C. Bajaj, and Q. Huang, "Deblursr: Event-based motion deblurring under the spiking representation," in *AAAI Conference on Artificial Intelligence*, 2024.

[44] C. Song, Q. Huang, and C. Bajaj, "E-cir: Event-enhanced continuous intensity recovery," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 7793–7802.

[45] F. Xu, L. Yu, B. Wang, W. Yang, G.-S. Xia, X. Jia, Z. Qiao, and J. Liu, "Motion deblurring with real events," in *IEEE International Conference on Computer Vision(ICCV)*, 2021, pp. 2583–2592.

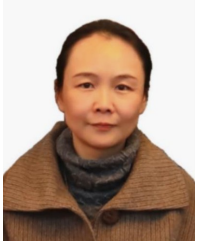
[46] W. Shang, D. Ren, Y. Yang, H. Zhang, K. Ma, and W. Zuo, "Joint video multi-frame interpolation and deblurring under unknown exposure time," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 13 935–13 944.

[47] T. Kim, J. Lee, L. Wang, and K.-J. Yoon, "Event-guided deblurring of unknown exposure time videos," in *European Conference on Computer Vision (ECCV)*, 2022, pp. 519–538.
[48] C. Zhang, X. Zhang, M. Lin, C. Li, C. He, W. Yang, G.-S. Xia, and L. Yu, "Crosszoom: Simultaneous motion deblurring and event super-resolving," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
[49] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
[50] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2019, pp. 3857–3866.
[51] S. M. Mostafavi I., J. Choi, and K.-J. Yoon, "Learning to super resolve intensity images from events," *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 2768–2786, 2020.
[52] L. Sun, C. Sakaridis, J. Liang, P. Sun, J. Cao, K. Zhang, Q. Jiang, K. Wang, and L. Van Gool, "Event-based frame interpolation with ad-hoc deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 043–18 052.
[53] S. Nah, S. Baik, S. Hong, G. Moon, S. Son, R. Timofte, and K. M. Lee, "Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1996–2005.
[54] H. Rebecq, D. Gehrig, and D. Scaramuzza, "Esim: an open event camera simulator," in *Conference on robot learning*. PMLR, 2018, pp. 969–982.
[55] T. Stoffregen, C. Scheerlinck, D. Scaramuzza, T. Drummond, N. Barnes, L. Kleeman, and R. Mahony, "Reducing the sim-to-real gap for event cameras," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*. Springer, 2020, pp. 534–549.
[56] L. Kong, B. Jiang, D. Luo, W. Chu, X. Huang, Y. Tai, C. Wang, and J. Yang, "Ifrnet: Intermediate feature refine network for efficient frame interpolation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 1959–1968.
[57] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
[58] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.
[59] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *European Conference on Computer Vision (ECCV)*, 2018.

**Nuo Chen** received the B.E. degree in Mechanical Design manufacture and Automation and M.S. degree in Mechanical and Electrical Engineering from the Central South University (CSU), China, in 2019 and 2022, respectively. He is currently working toward the PhD degree in Information and Communication Engineering from National University of Defense Technology (NUDT), Changsha, China. His research interests include image processing, neuromorphic cameras and computer vision.

**Longguang Wang** received the B.E. degree in Electrical Engineering from Shandong University (SDU), Jinan, China, in 2015, and the Ph.D. degree in Information and Communication Engineering from National University of Defense Technology (NUDT), Changsha, China, in 2022. His current research interests include low-level vision and 3D vision.

**Miao Li** received the M.E. and Ph.D. degrees from the National University of Defense Technology (NUDT) in 2012 and 2017, respectively. He is currently an Associate Professor with the College of Electronic Science and Technology, NUDT. His current research interests include infrared dim and small target detection.



**Wei An** received the Ph.D. degree from the National University of Defense Technology (NUDT), Changsha, China, in 1999. She was a Senior Visiting Scholar with the University of Southampton, Southampton, U.K., in 2016. She is currently a Professor with the College of Electronic Science and Technology, NUDT. She has authored or co-authored over 100 journal and conference publications. Her current research interests include signal processing and image processing.