Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.063906

ARTICLE

DNEFNET: Denoising and Frequency Domain Feature Enhancement Event Fusion Network for Image Deblurring

Kangkang Zhao[1], **Yaojie Chen**[1,*] **and Jianbo Li**[2]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, 430081, China
[2]CSSC Huangpu Wenchong Shipbuilding Company Limited, Guangzhou, 510727, China

*Corresponding Author: Yaojie Chen. Email: chenyaojie@wust.edu.cn

Received: 28 January 2025; Accepted: 27 April 2025; Published: 09 June 2025

ABSTRACT: Traditional cameras inevitably suffer from motion blur when facing high-speed moving objects. Event cameras, as high temporal resolution bionic cameras, record intensity changes in an asynchronous manner, and their recorded high temporal resolution information can effectively solve the problem of time information loss in motion blur. Existing event-based deblurring methods still face challenges when facing high-speed moving objects. We conducted an in-depth study of the imaging principle of event cameras. We found that the event stream contains excessive noise. The valid information is sparse. Invalid event features hinder the expression of valid features due to the uncertainty of the global threshold. To address this problem, a denoising-based long and short-term memory module (DTM) is designed in this paper. The DTM suppressed the original event information by noise reduction process. Invalid features in the event stream and solves the problem of sparse valid information in the event stream, and it also combines with the long short-term memory module (LSTM), which further enhances the event feature information in the time scale. In addition, through the in-depth understanding of the unique characteristics of event features, it is found that the high-frequency information recorded by event features does not effectively guide the fusion feature deblurring process in the spatial-domain-based feature processing, and for this reason, we introduce the residual fast fourier transform module (RES-FFT) to further enhance the high-frequency characteristics of the fusion features by performing the feature extraction of the fusion features from the perspective of the frequency domain. Ultimately, our proposed event image fusion network based on event denoising and frequency domain feature enhancement (DNEFNET) achieved Peak Signal-to-Noise Ratio (PSNR)/Structural Similarity Index Measure (SSIM) scores of 35.55/0.972 on the GoPro dataset and 38.27/0.975 on the REBlur dataset, achieving the state of the art (SOTA) effect.

KEYWORDS: Image deblurring; event camera; denoising; frequency domain

1 Introduction

Motion blur commonly occurs in conventional cameras, particularly when the camera is shaking or objects are moving at high speed. This phenomenon primarily results from long exposure times, during which the camera is unable to effectively accumulate information about the light emitted or reflected by objects within a short period. Consequently, important textures and details in the image are lost. In high-speed motion scenarios, conventional cameras often struggle to capture sufficient dynamic information, leading to severe image blur and posing a significant challenge to most existing deblurring methods. Traditional approaches treat deblurring as an inverse problem and attempt to recover a sharp image via inverse convolution by estimating the blur kernel. However, these methods heavily depend on accurate blur kernel estimation, which is difficult to achieve in real-world scenarios with unknown or complex motion

patterns. In contrast, deep learning-based deblurring methods learn the mapping from blurred to sharp images through large-scale data training. Although such methods have significantly improved deblurring performance, they still exhibit notable limitations in handling fast-moving objects. This is mainly because conventional frame-based images [1] fail to capture motion information during exposure. Event cameras, by contrast, operate on an asynchronous event-driven mechanism that records changes in luminance with extremely high temporal resolution. They generate sparse event streams that accurately reflect the direction, speed, and edge information of object motion. This enables not only explicit capture of dynamic scene changes but also implicit provision of positional cues for blurred regions. Traditional methods are limited by their inability to accurately perceive motion trajectories, which restricts their deblurring performance. Therefore, leveraging the high temporal resolution and motion-awareness capabilities of event cameras for image deblurring has emerged as a promising and increasingly mainstream approach.

Fig. 1 presents a comparison of the architectural differences between our method and recent event-based deblurring approaches. These methods are described in more detail in the next related methods section. Sun et al. [2] proposed the EFNet network architecture and introduced the event-image feature fusion module (EICA) for the first time, which achieves significant state-of-the-art (SOTA) results in the field of image deblurring through the dual-channel extraction and fusion of event features and traditional RGB image features. Subsequently, Yang et al. [3] further improved EFNet and proposed the DLEFNet network architecture. Unlike Sun et al., Yang et al. borrowed the EICA module and applied deformable convolution and the LSTM network model to event data processing for the first time. This further enhanced the ability to extract temporal information from event features. By improving the modeling of temporal information in event features, DLEFNet achieves a more significant improvement in deblurring performance. The SOTA method DiffEvent [4] treats image deblurring as a generative problem by introducing diffusion priors and a sampling strategy that jointly estimates the sharp image and residual image. This approach effectively reduces residuals and more accurately restores details, significantly improving deblurring performance.

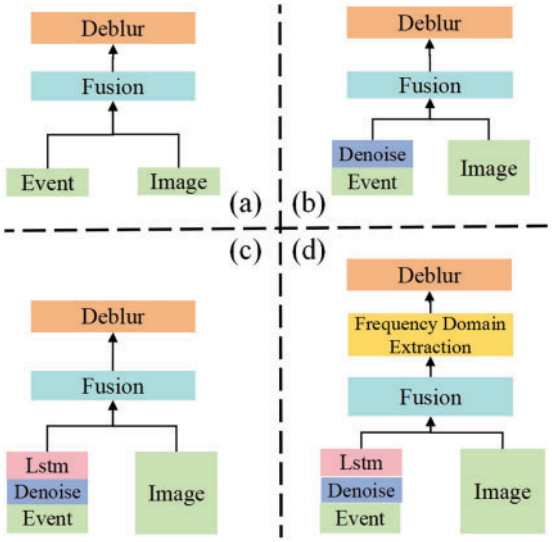

Figure 1: **(a)** Conventional model **(b)** Noise estimation model **(c)** Time-series modeling model **(d)** Our model

However, the above methods still exhibit significant shortcomings. In the traditional model shown in Fig. 1a, insufficient consideration of the random noise problem–due to the high sensitivity of the event camera and the uncertainty in setting the global threshold [5] leads to the event stream being interspersed

with a large amount of redundant random noise and invalid events. This results in sparse effective information and limited expressive capability, reducing the model's effectiveness in recovering details in dynamically ambiguous regions. The noise estimation-based model in Fig. 1b mitigates the interference of excessive noise in the event data by introducing a noise estimation mechanism. However, the denoising process inevitably removes some effective information, leading to a loss of temporal information. This limits the model's ability to capture the completeness and continuity of motion details, making it difficult to accurately restore complex dynamic changes. Furthermore, although the model based on temporal modeling in Fig. 1c improves the temporal representation of event features by introducing an LSTM-based module, it overlooks the inter-modal resolution differences and inconsistencies in feature representation during the fusion process [6] with RGB features. Consequently, the high-frequency details of the moving object structure, edges, and textures contained in the event features are not fully exploited, hindering the model's performance in high-frequency detail reconstruction and edge recovery.

To address the above issues, we thoroughly analyze the characteristics of the event stream. We found that due to the high sensitivity of event sensors and the high degree of integration within electronic devices, there is a substantial amount of redundant random noise in the event stream. This noise leads to sparse valid information and somewhat inhibits the expression of meaningful features. At the same time, the event stream contains rich high-frequency feature information, particularly high-frequency components related to detailed features such as the structure, texture, and edges of moving objects. However, traditional spatial-domain feature processing methods, which typically rely on convolutional operations, tend to focus on extracting low-frequency information, failing to effectively utilize these important high-frequency features. To overcome these challenges, we propose the DNEFNET network, which introduces the DTM and RES-FFT [7] modules to enhance deblurring performance. The DTM module suppresses invalid features in the event data, emphasizing valid features for more accurate representation and improved temporal expressiveness. The RES-FFT module shifts image feature extraction from the spatial domain to the frequency domain, thereby enhancing high-frequency details in the fused features and better preserving the edges and fine details captured by event data. The specific process is outlined in Algorithm 1. Experimental results on both synthetic and real datasets (e.g., GoPro and REBlur) demonstrate that our method outperforms existing state-of-the-art approaches in terms of deblurring performance and visual quality.

Algorithm 1: DNEFNET image processing

Require: Event data $E \in \mathbb{R}^{C_{ev} \times H \times W}$, Blurred image $B \in \mathbb{R}^{C \times H \times W}$

$Conv(E, B) \rightarrow E_1, B_1$
DTM$(E_1) \rightarrow e^*1, e^*2, e^*4$
for i = 1 to 3 **do**
    Downsample$(b_{\text{input}}, \text{scale} = 0.5) \rightarrow b^*$
    EICA$(e_i^*, b^*) \rightarrow F_{\text{fusion}}^i$
    $F_{\text{fusion}}^i \rightarrow b_{\text{input}}$
end for
Upsample$(b_{\text{input}}, \text{scale} = 4) \rightarrow b_{\text{input}}$
SAM$(b_{\text{input}}) \rightarrow F_{\text{SAM}}$
RES-FFT$(F_{\text{SAM}}) \rightarrow F_{\text{out}}$
Conv3x3$(F_{\text{out}}) \rightarrow I_{\text{sharp}}$
Return I_{sharp}

In summary, our main contributions are as follows.

1. We deeply analyze the imaging principle of event cameras and find that there is a large amount of redundant random noise in event data, which leads to sparse and limited expression of effective features, while high-frequency features are underutilized. To this end, we propose the DNEFNET network, which effectively solves the core problems of insufficient expression of event features and underutilization of high-frequency features by combining frequency-domain enhancement and denoising mechanisms.

2. We design a novel Denoising Long and Short-Term Memory (DTM) module that integrates denoising with time-series modeling. By suppressing redundant noise and emphasizing effective features, this module significantly enhances feature representation and improves the model's ability to express temporal features.

3. We introduce a RES-FFT-based residual block that extends feature processing from the spatial domain to the frequency domain. This block effectively isolates and enhances high-frequency information in the fused features, improving edge recovery and detail reconstruction, thereby providing a robust solution for event-based deblurring tasks.

4. Our deblurring network (DNEFNET) achieves PSNR and SSIM scores of 35.55/38.26 dB and 0.972/0.975 on the GoPro and REBlur datasets, respectively. These results are significantly better than existing deblurring methods, demonstrating state-of-the-art performance.

2 Related Work

2.1 Frame-Based Image Deblurring Method

Traditional deblurring methods typically treat the task as an inverse problem. In image processing, blurring is viewed as the result of convolving the original image with a blurring kernel. These methods construct a blurring model based on the known or estimated convolutional structure of the blurring kernel and apply inverse convolution to achieve deblurring. These methods can be broadly categorized into two approaches: blind deblurring and non-blind deblurring, depending on whether the blur kernel is assumed to be known. Early non-blind deblurring methods use classical image inverse convolution algorithms. For example, Ref. [8] proposes the use of Wiener filtering based on minimum mean square error to deblur images, while Ref. [9] introduces the Lucy-Richardson inverse convolution method for image clarification. However, these approaches often fail to fully exploit the a priori information inherent in natural images, leading to inaccurate recovery. In blind deblurring architectures, where the blurring kernel is unknown, the goal is to simultaneously recover the blurred image and estimate the blur kernel [10,11] propose solutions that involve adding various constraints for regularization and incorporating additional priors. While these non-deep learning methods perform well in many cases, they tend to produce unsatisfactory results in complex real-world scenarios, particularly under extreme blur conditions.

2.2 Image Deblurring Method Based on CNN Network

With the success of deep learning, many Convolutional Neural Network (CNN) deep models have been proposed, leading to significant progress in image deblurring. For example, Ronneberger et al. [12] propose a multiscale CNN network (U-NET) that recovers clear images from blurred ones by integrating multiscale feature information. Zhu et al. [13] propose utilizing Deformable Convolutional Networks (DCNs) to estimate blur patterns and regions. By specifically learning the relationship between blurred areas and clear images, their approach optimizes feature alignment, further enhancing deblurring performance. Cho et al. [14] introduce MIMO-UNET, which facilitates inter-scale information exchange and improves deblurring performance by adding input channels for different scale images, extracting surface information using the Shallow Convolutional Module (SCM), and applying the attention mechanism for selective feature

extraction. Tao et al. [15] propose a coarse-to-fine scale recurrent network to enhance the efficiency of multiscale image deblurring. Similarly, Zamir et al. [16] introduce inter-stage feature fusion [17] to further optimize deblurring performance. To reduce the computational cost associated with multiscale frameworks, Zhang et al. [18] propose a network architecture based on a multi-patch strategy (DMPHN), exploring different stacking methods and achieving good results in image deblurring. Although these methods demonstrate strong performance for mildly blurred images, they remain ineffective in cases of extreme blurring due to their inability to capture temporal features, such as the motion information of objects.

2.3 Event-Based Image Deblurring

In extreme environments, traditional image sensors often fail to effectively record object motion information due to exposure time limitations, resulting in poor performance of traditional deblurring methods in complex scenarios such as high-speed motion and rapid lighting changes. In contrast, event cameras, with their asynchronous imaging and microsecond time resolution, accurately capture dynamic changes in scenes. While image data excels in texture and spatial details, it can be complemented by event data, leading to increased attention on deblurring methods based on event-image fusion. Pan et al. [19] introduce the earliest event-based deblurring model, the Event-based Double Integration (EDI) model, which integrates event data for deblurring. However, due to sampling limitations of event cameras and the resulting noise, Pan et al. [20] further propose the MEDI model, which mitigates noise interference by obtaining smaller noise estimates through multi-image and event integration. Wang et al. [21] demonstrate that adding event data channels to image-based deblurring networks improves their performance, while Xu et al. [22] devise a semi-supervised framework to address data inconsistency issues by utilizing real-world events. Recent advancements include EFNET [2], which constructs a two-segment U-Net architecture: one segment processes event information, while the other focuses on image deblurring. This network introduces new event representations and fusion modules, achieving excellent results. Yang et al. [3] propose using an LSTM network with deformable convolution to further extract event information, enhancing temporal feature representation and achieving strong performance. Although these methods address event camera noise caused by global threshold uncertainty, they fail to deeply analyze the causes of noise generation. Redundant and invalid feature information in event data leads to sparse valid information, impacting effective feature representation. Moreover, most methods rely on spatial-domain feature processing, which struggles to fully utilize the structural motion and rich high-frequency details provided by event data. This limitation hampers the performance of existing event-based deblurring models in complex real-world scenarios. Effectively leveraging high-frequency information and temporal dynamic features in event data remains a critical challenge in current event-driven deblurring research.

3 Method

3.1 Representation of Events

Event cameras [23] are bio-inspired sensors that asynchronously record logarithmic changes in image intensity. Unlike conventional cameras, which capture full image frames at fixed intervals, event cameras generate an event whenever the intensity at a specific pixel exceeds a predefined threshold. Free from the limitations of fixed exposure times and restricted dynamic range, event cameras are capable of capturing high-speed motion with microsecond-level temporal resolution. Instead of producing discrete frames, they detect rapid intensity variations and output a continuous stream of asynchronous events that encode these changes. Assume a pixel location k and a manually defined exposure time T. Then, an event e can be represented as (x_k, y_k, t_k, p_k), where (x_k, y_k) denotes the spatial coordinates of the pixel, and t_k represents the timestamp of the event. The polarity $p_k \in \{+1, -1\}$ indicates the direction of the intensity change (increase

or decrease) at pixel k at time t_k. The generation of this polarity can be expressed as follows:

$$p_k = \begin{cases} +1, \text{ if } \log\left(\dfrac{J_t(x_k, y_k)}{J_{t-\Delta t}(x_k, y_k)}\right) > c \\ -1, \text{ if } \log\left(\dfrac{\mathcal{J}_t(x_k, y_k)}{J_{t-\Delta t}(x_k, y_k)}\right) < c. \end{cases} \tag{1}$$

In Eq. (1), c denotes the pixel brightness change threshold specified by the event camera, if the intensity change $\log\left(\frac{J_t(x_k, y_k)}{J_{t-\Delta t}(x_k, y_k)}\right)$ produced by pixel k during the instantaneous time Δt exceeds the threshold c, a corresponding event is generated and pixel point k is updated.

3.2 Overall Architecture

As shown in Fig. 2, DNEFNET consists of two main components: the feature extraction module and the deblurring module. The feature extraction module processes the input image and event data through two separate channels. In the event channel, the event stream contains a significant amount of random noise, leading to sparse and ineffective representation of valid information. To address this issue, we design the DTM module, which first suppresses invalid features through a denoising process, allowing the remaining valid information to become more concentrated. Although this denoising process may result in the loss of some temporal details, DTM compensates for this by performing temporal modeling on the denoised event features, thereby enhancing their temporal expressiveness. As a result, the DTM module not only improves the denoising effectiveness of the event stream but also strengthens its ability to represent temporal information.

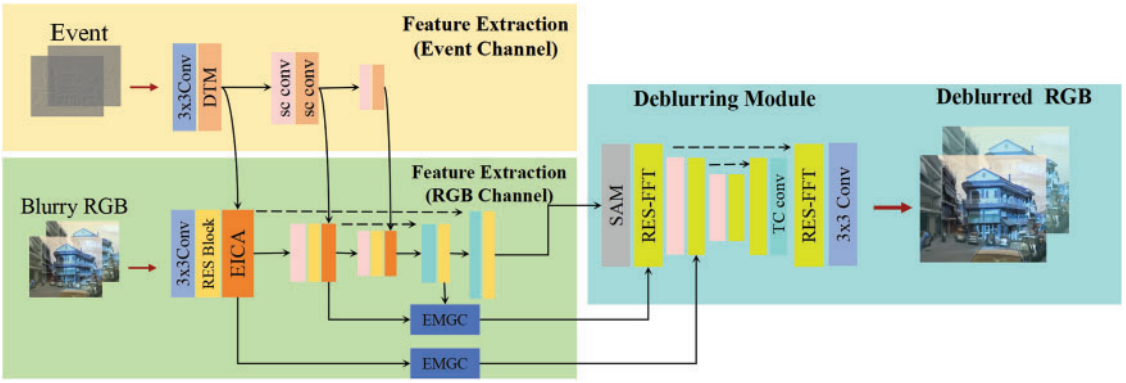

Figure 2: Overall network architecture

For the image channel, we perform multi-level fusion of the processed event features with image features at different scales, enhancing image sharpness and improving edge recovery. The second part of the network is the deblurring module. Traditional deblurring methods typically operate in the spatial domain, where high-frequency components of the fused features are often underutilized. To overcome this limitation, we introduce the RES-FFT residual block, replacing the conventional residual blocks in U-Net with RES-FFT blocks, thereby shifting feature processing from the spatial domain to the frequency domain. This enhances and separates the high-frequency components within the fused features, significantly improving the model's capability to reconstruct motion edges and texture details and leading to clearer and more detailed deblurred results.

DNEFNET fully leverages the high spatial resolution and rich texture detail of image data while simultaneously applying fine-grained denoising and temporal modeling to event data. This effectively solves the problem of sparse and ineffective feature representation in event streams. Moreover, through deep analysis of event characteristics, DNEFNET extracts and enhances high-frequency information in the fused features, enabling the model to more accurately capture motion edges and fine textures, thereby improving deblurring quality and detail preservation.

Thanks to these innovations, DNEFNET achieves superior deblurring performance on both the GoPro [24] and REBlur [2] datasets, especially under challenging scenarios involving complex scenes and extreme motion blur, demonstrating strong robustness and enhanced detail recovery.

3.3 DTM

Given a potentially clear image $L(f)$, according to the EDI model proposed in [19] we can derive its fuzzy image based on the event data e. The specific derivation process can be expressed as follows.

$$\begin{aligned}\mathbf{B} &= \frac{1}{T}\int_{f-T/2}^{f+T/2}\mathbf{L}(t)dt, \\ &= \frac{\mathbf{L}(f)}{T}\int_{f-T/2}^{f+T/2}\exp\left(c\int_{f}^{t}p_k ds\right)dt.\end{aligned} \tag{2}$$

In Eq. (2), B denotes the observed blurred image, p_k represents the polarity component of the event stream, T is the manually set exposure time, and f refers to the midpoint of the exposure interval. $L(f)$ denotes the latent sharp image corresponding to B. Due to the high sensitivity of event sensors and the high degree of electronic integration, electronic noise is inevitably introduced. This makes it challenging to apply a globally fixed threshold c uniformly across all pixels, resulting in significant uncertainty at the pixel level. Such threshold deviations cause the event stream to contain a large amount of redundant random noise, which lacks meaningful information. This noise degrades feature representation quality and contributes to the sparse nature of valid event data.

The EDI model proposed in [19] is developed under idealized conditions. While it effectively establishes a mathematical relationship between blurred images and latent sharp images, its performance deteriorates in real-world scenarios, where complex and unknown environmental factors limit its practical applicability.

Therefore, to mitigate the sparsity of valid information in the event stream and enhance the concentration of meaningful features, we propose a denoising-based Long and Short-Term Memory module (DTM), whose structure is illustrated in Fig. 3. The DTM module is designed to address two key challenges: suppressing redundant random noise and preserving temporal information in event features. It achieves this by first applying a denoising process to eliminate invalid features caused by noise, thereby enhancing the compactness and relevance of the feature representation. However, since denoising may inevitably remove some valid information, the module further incorporates a temporal modeling mechanism to recover and enhance the temporal expressiveness of the denoised features.

The DTM module thus consists of two main components: a denoising sub-module that filters out random noise and a temporal modeling sub-module that captures long- and short-term dependencies in the denoised feature stream.

In the denoising stage, we first normalize the input event data $e \in \mathbb{R}^{C\times H\times W}$ to ensure training stability and a reasonable data distribution. Specifically, we compute the mean μ and standard deviation σ of the input and normalize it to obtain a standardized distribution e_{norm} with zero mean and unit variance. This normalization provides a more stable basis for subsequent network processing. Then, e_{norm} is passed through

two convolutional layers: a 1×1 convolution to perform inter-channel linear mapping, followed by a 3×3 convolution to extract local edge features.

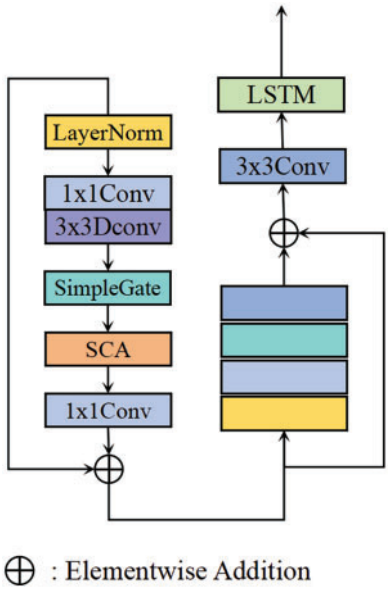

Figure 3: The detailed structure of DTM block

This sequence of operations yields the shallow feature representation F_{shallow}. The deep convolution process enhances salient local regions, effectively suppressing noise events, which typically exhibit weak or irregular patterns due to the threshold uncertainty c. This initial suppression of noise provides a crucial pre-processing step for the denoising pipeline. The full denoising process is summarized in Eq. (3), where μ denotes the mean and σ represents the standard deviation of the input e.

$$\begin{aligned} e_{norm} &= \frac{e-\mu}{\sigma}, \\ F_{shallow} &= Conv_2\left(Conv_1(e_{norm})\right). \end{aligned} \tag{3}$$

In traditional denoising methods, although nonlinear activation functions such as ReLU and GELU can effectively enhance shallow features and suppress noise, they often lead to increased inter-block complexity and computational overhead. To address this issue, inspired by NAF [25], we introduce a simplified activation structure called *SimpleGate*, as illustrated in Fig. 4a. Specifically, the input shallow feature map F_{shallow} is split into two equal parts, F_1 and F_2 along the channel dimension, which are then fused through element-wise multiplication to generate the output F_{SG}, as defined in Eq. (4). This lightweight gating mechanism enables selective enhancement and suppression of channel-wise features, effectively reducing the influence of redundant noise in event data. Moreover, SimpleGate achieves strong noise suppression capability with significantly lower computational cost, providing an efficient and practical solution for feature modulation in the denoising module.

$$\begin{aligned} \mathrm{F}_1, \mathrm{F}_2 &= \mathrm{split}(F_{Shallow}), \\ \mathrm{F}_{\mathrm{SG}} &= \mathrm{SimpleGate}(\mathrm{F}_1, \mathrm{F}_2) = \mathrm{F}_1 \odot \mathrm{F}_2. \end{aligned} \tag{4}$$

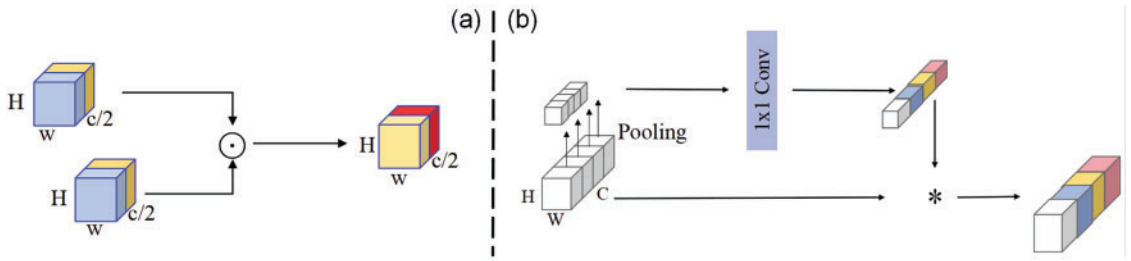

Figure 4: **(a)** SimpleGate implementation **(b)** SCA implementation $*$: Channel-wise multiplication

To further enhance the denoising capability in conjunction with SimpleGate, we introduce a spatially compressed channel attention mechanism (SCA), as depicted in Fig. 4b. As described in Eq. (5), SCA first applies global average pooling to the input feature map F_{SG} to compute the spatially averaged channel-wise vector V. This vector is then passed through a 1×1 convolution to generate the channel attention weights W_{CA}. These weights are subsequently multiplied with F_{SG} to produce the refined feature map F_{CA}. Since noisy events typically exhibit weak or insignificant activation across the channel dimension, the channel attention mechanism adaptively reweights the feature map, selectively enhancing the more informative channels while attenuating the impact of noisy or irrelevant ones. This process strengthens the representation of valid features and further suppresses the influence of invalid event-induced artifacts, thereby improving the effectiveness and robustness of the denoising operation.

$$\begin{aligned} \mathrm{V} &= \mathrm{GAP}(F_{SG}), \\ \mathrm{W}_{\mathrm{CA}} &= Conv_{1x1}(V), \\ F_{CA} &= W_{CA} \odot F_{SG}. \end{aligned} \tag{5}$$

However, some temporal information may be lost during the inevitable processing of valid features in the denoising module. To address this issue, we introduce an LSTM-based temporal enhancement module within DTM. This module takes the denoised event features as input and compensates for potential information loss by leveraging historical data, thereby ensuring the completeness and coherence of the output in the temporal domain. The LSTM network effectively restores the temporal continuity of the denoised features, preserving the integrity of valid information and mitigating the loss typically associated with single-frame denoising. Furthermore, by processing the event features frame by frame, the LSTM embeds temporal dependencies into the output, ensuring consistency across adjacent time steps and improving both the temporal resolution and feature representation capability.

$$\begin{aligned} C_t &= f_t \odot C_{t-1} + i_t \odot \tanh\left(\mathrm{Conv}\left(\left[F_{CA}, h_{t-1}\right]\right), W_c\right), \\ h_t &= o_t \odot \tanh\left(C_t\right). \end{aligned} \tag{6}$$

In Eq. (6), h_t and C_t represent the hidden state and memory unit at the current time step, respectively. h_{t-1} and C_{t-1} correspond to the hidden state and memory unit at the previous time step. o_t denotes the candidate value at the current time step. f_t represents the activation value of the forget gate, i_t denotes the input gate, and w_c refers to the convolution weight. By leveraging the unique memory unit of the LSTM and the combined action of the three gates, the model effectively extracts temporal sequences from the event stream data. This approach not only captures local spatial correlations but also preserves and enhances temporal information. During frame-by-frame processing of event features, the model integrates temporal

dependencies into the output features, ensuring temporal compensation and consistency in the denoised features, thus improving feature representation capability.

3.4 FFT Based U-Net Network Architecture

The traditional U-Net architecture performs well in image deblurring tasks; however, when fusing event data and image data, existing methods often struggle to fully exploit the high-frequency feature information of the event data. Since event data primarily consists of sparse edges and motion trajectories, which contain crucial information for accurately localizing object motion and blur regions, the extraction and utilization of high-frequency details are particularly important in feature fusion. However, conventional spatial-domain convolution operations have limited capacity to capture these high-frequency components, which can lead to the degradation of edges, structures, and texture details of moving objects, thus negatively impacting the deblurring performance. To address this issue, we design an enhanced U-Net architecture that incorporates RES-FFT residual blocks. Unlike traditional spatial-domain residual blocks, the RES-FFT block shifts the feature processing from the spatial domain to the frequency domain, isolating and enhancing high-frequency components, such as edges and motion trajectories, to more effectively recover detailed features. The structure of the RES-FFT residual block is illustrated in Fig. 5a. In Eq. (7), we transform the input features from the spatial domain to the frequency domain using a Fourier Transform (FFT) to isolate high-frequency components that contain essential information, such as edges and motion trajectories. The fused features processed by the feedforward network are denoted as F_{fusion}. In the frequency domain, these high-frequency components are convolved with 1 × 1 and 3 × 3 convolutions to further extract detailed features, enhancing sensitivity to ambiguous regions and motion information, thereby generating high-frequency features F_{high} for further processing. Meanwhile, the low-frequency information is retained in the spatial domain through two standard convolution operations. Subsequently, an inverse Fourier Transform (IFFT) is applied to map the enhanced high-frequency features back to the spatial domain, outputting the Fourier feature F_{fft}. This is then pixel-wise added to the original spatial features, which have undergone two convolution operations, ensuring effective complementarity between the spatial and frequency domains. This process strengthens the high-frequency information while preserving low-frequency and overall structural information. The final processed feature, F_{out}, is then output. It is important to note that the object motion information and high-frequency features of the blurred region recorded in the event data are crucial for the deblurring process. The RES-FFT module effectively extracts these features through precise frequency-domain operations, significantly enhancing the fusion feature representation and improving the model's ability to recover edge details and reconstruct high-frequency features.

$$
\begin{aligned}
&F_{\text{conv}} = \text{Conv}_{3\times3}\left(\text{Relu}\left(\text{Conv}_{3\times3}\left(\text{F}_{\text{fusion}}\right)\right)\right),\\
&\text{Y} = \text{FFT}\left(\text{F}_{\text{fusion}}\right),\\
&\text{F}_{\text{high}} = \text{Conv}_{1\times1}\left(\text{Relu}\left(\text{Conv}_{3\times3}(\text{Y})\right)\right),\\
&\text{F}_{\text{fft}} = \text{IFFT}\left(\text{F}_{\text{high}}\right),\\
&\text{F}_{\text{out}} = \text{F}_{\text{fft}} + \text{F}_{\text{high}}.
\end{aligned}
\tag{7}
$$

In addition, to enhance the deep feature extraction capability of the network, we adopt a stacked design in both the encoder and decoder, extending the number of residual blocks to four layers. This further deepens the feature learning capacity of the network. The specific architecture of the network is shown in Fig. 5b. This multi-module synergistic design, which combines event features and image features, enables our variant of the U-Net architecture to handle the deblurring task more efficiently in event and image fusion scenarios.

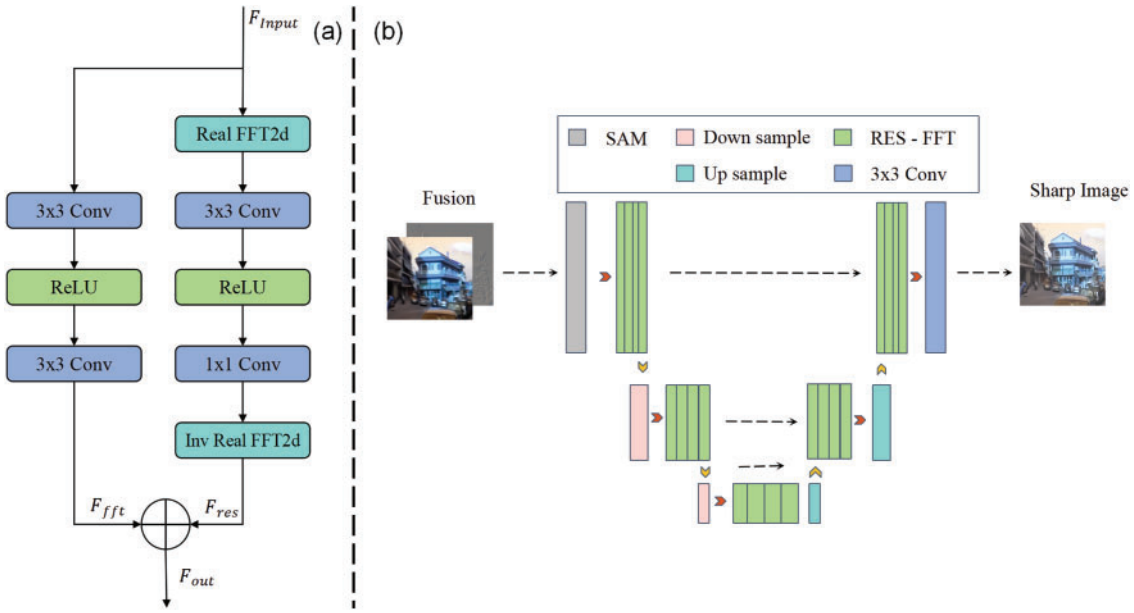

Figure 5: **(a)** Specific architecture of RES-FFT **(b)** The deblurring part of DNEFNET

4 Experiments

4.1 Dataset

GoPro: We used the GoPro dataset [24], which is widely used for motion deblurring tasks, for training and evaluation. The dataset contains 3214 pairs of blurred and clear image pairs with an image resolution of 1280 × 720. The blurred images are generated by averaging multiple frames of clear images taken at high speed. In order to adapt our two-channel model to event data, we generated event stream data corresponding to each pair of blurred images using the ESIM event simulator. The training set contains 2103 pairs of images, and the test set contains 1111 pairs of images for evaluating the performance of the model.

REBlur: We pre-trained the model on the GoPro dataset and fine-tuned it on the REBlur [2] real dataset to assess the model's generalization ability on real event data. The REBlur dataset is designed for event-image deblurring tasks and includes three motion modes and 12 linear and non-linear blurring scenarios, totaling 36 sequences and 1469 image-event data pairs. Of these, 486 pairs are used for training and 983 pairs for testing. Since the REBlur dataset contains blurred images with varying speeds and motion modes, we further evaluate the model's deblurring performance at different blurring levels to assess its robustness in extreme blurring and complex motion scenarios.

Table 1: Comparative results on motion deblurring on the GoPro dataset

Method	PSNR ↑	SSIM ↑
BANET [26]	32.54	0.957
MPRNet	32.66	0.959
MIMO-unet	32.68	0.959
HINET [27]	32.71	0.959
Restormer [28]	32.92	0.961
U-former [29]	33.06	0.967

(Continued)

Table 1 (continued)

Method	PSNR ↑	SSIM ↑
SSAMAN [30]	33.53	0.965
NAFNET	33.71	0.967
MADANET [31]	33.84	0.964
RED*	28.98	0.849
ERDNet* [32]	32.99	0.935
HINET*	33.69	0.961
EFNET*	35.46	0.972
DiffEvent*	35.55	0.972
Ours	**35.55**	**0.972**

Note: * denotes event-based approach, bolded meanings denote best results.

4.2 Experiment Parameters

Experiments were performed on a server with four 2080Ti GPUs, using data augmentation [33] and distributed training. We used the AdamW optimizer with an initial learning rate of $2e^{-4}$ and a minimum learning rate of $1e^{-7}$. 200,000 iterations were performed on the GoPro dataset in 43 hours. Subsequently, 1800 iterations were fine-tuned on the REBlur dataset, using a single GPU, with all other configurations remaining the same. In addition, SCER preprocessing was performed on the event data.

4.3 Comparisons with State-of-the-Art Methods

We compare our approach to state-of-the-art image-only and event-based deblurring methods on the GoPro and REBlur datasets. These include image-based methods such as Restormer, SSAMAN, NAFNET, HINET, MIMO-unet++, U-former, MADANET+, BANET, MPRNet, and SRN, as well as event-based deblurring methods like ERDNet, DiffEvent, EFNET, RED, and other recent advancements. Since most event-based deblurring methods do not have publicly available implementations, to ensure a fair comparison and fill this gap, we introduce an event channel in HINET, making it adaptable to event datasets. This modified version is labeled as HINET*.

GOPRO: We present the comparative results of our deblurring experiments on the GOPRO dataset in Table 1. Compared to existing event-based and image-based methods, our approach outperforms the others, with subjective metrics comparable to the best event-based deblurring method (DiffEvent). The main reason for this is that the GOPRO dataset is synthetic, lacking the complex, non-uniform blurring and varied motion patterns found in real-world scenes. The simple and consistent blurring in this dataset is more suited to DiffEvent's global fusion strategy, which excels at recovering image details. In contrast, our method focuses on the recovery of local details and high-frequency features, with an emphasis on enhancing effective features and recovering high-frequency details. However, due to the lower level of blurring and the single type of blur in the GOPRO dataset, noise and information sparsity in the event stream are less problematic, and the image's edge and high-frequency details are not significantly impacted. As a result, our method performs similarly to DiffEvent on this dataset. However, in more realistic and complex blurring environments, our method demonstrates superior performance and better detail recovery.

In Fig. 6, we visually present some qualitative results of our method on the GoPro dataset. It is evident that image-based methods suffer from the loss of sharp edge information during the deblurring process,

resulting in more blurred images. The main reason for this is that, as shown in [25–28], image-based deblurring methods, limited by RGB images, do not capture the object's motion trajectory during the exposure time, leading to the loss of some temporal information. As a result, these methods fail to achieve targeted deblurring. EFNET, on the other hand, incorporates event data and performs a good fusion of event and image information, resulting in a better deblurring effect compared to image-based methods. However, EFNET does not fully account for the high-frequency details captured in the event features or the impact of redundant noise, leading to a lack of some details and textures in the output. In contrast, DNEFNET addresses both the high-frequency details in the events and the effect of redundant noise, allowing it to better restore the structural and textural details of the blurred image compared to other methods.

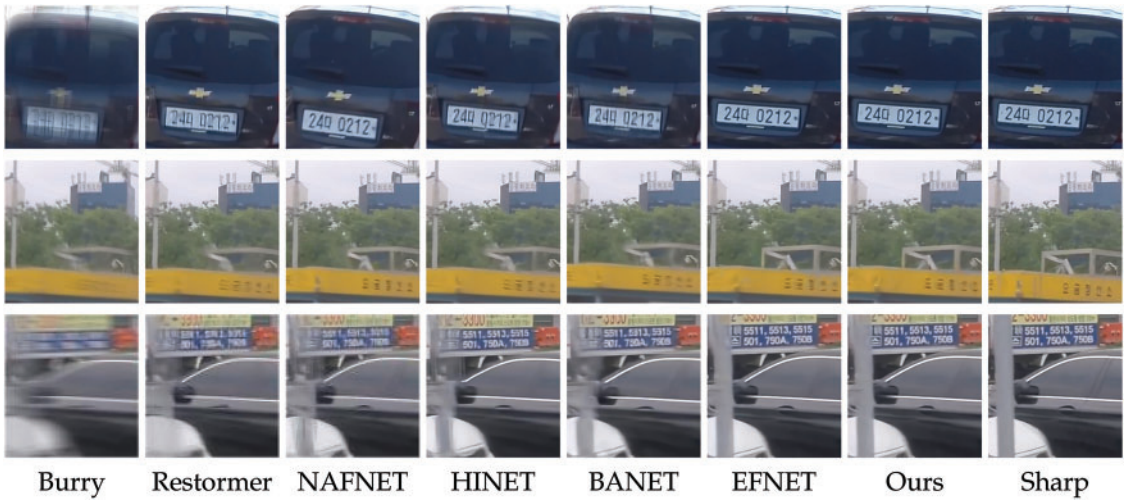

Figure 6: Visual comparison on GoPro

REBlur: In Table 2, we present the comparative results of deblurring experiments on the REBlur dataset, where our method achieves a new state-of-the-art (SOTA) performance, outperforming the current diffevent method in both PSNR and SSIM metrics. This improvement is attributed to our DTM module and RES-FFT-based residual block, specifically designed for event data, which enable more precise capture of localized motion trajectories and high-frequency features. In contrast, Diffevent adopts a global fusion strategy based on diffusion models and Transformers, treating event image deblurring as a generative task. However, this approach has limitations in local detail recovery, leading to ineffective utilization of local high-frequency information in non-uniform blurring and subtle motion features.

Additionally, in Fig. 7, we visually demonstrate the deblurring effects of different methods under various blurring conditions, where the blur intensity per frame is set based on pixel displacement, ranging from 0.5–5 pixels/frame and 5–20 pixels/frame. In the two sets of comparative experiments simulating low-speed and medium-speed motion, while all methods generally achieve deblurring, they struggle to recover fine details such as text, image edges, and other intricate structures. In contrast, in the 20–50 pixels/frame fast-motion simulation, our method produces deblurred images that are noticeably more consistent with the ground truth clear images. The superior visual quality compared to other methods further demonstrates that our approach can effectively restore local high-frequency details and achieve exceptional deblurring performance, even under extreme blurring conditions.

Table 2: Comparative results on motion deblurring on the REBlur dataset

Method	PSNR ↑	SSIM ↑
BANET	31.16	0.925
Restormer	34.82	0.955
SRN	35.10	0.961
HINET	35.58	0.965
NAFNet	36.00	0.964
HINET*	37.68	0.973
EFNET*	38.12	0.973
DiffEvent*	38.23	0.974
Ours	**38.27**	**0.975**

Note: * denotes event-based approach, bolded meanings denote best results.

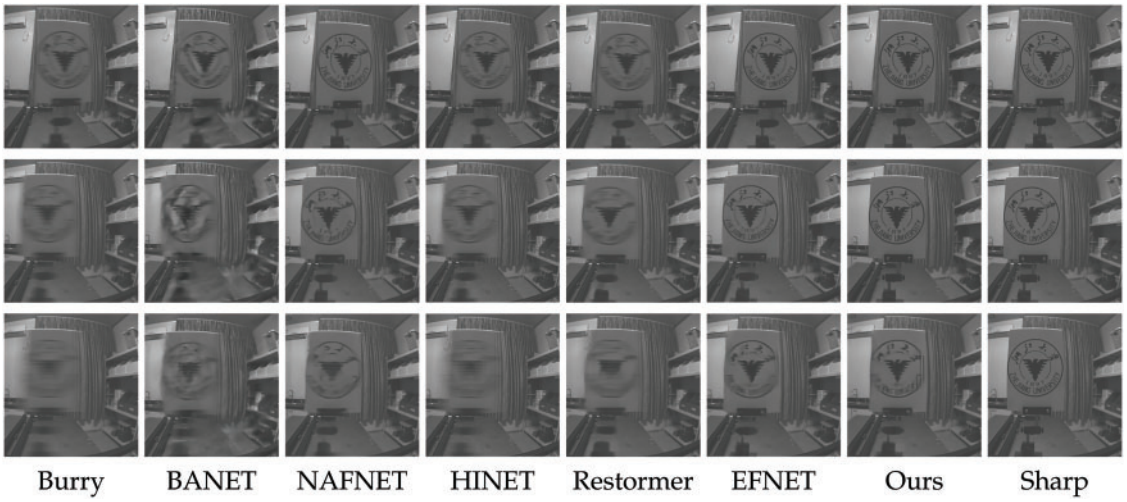

Figure 7: Visual comparison on REBlur

4.4 Hyperparametric Analysis

In Fig. 7, we visualize the deblurring effect of our method under different blurring levels. In Fig. 8, we further compare and analyze the quantitative deblurring performance of DNEFNET and EFNET under various blurring conditions. As shown in the line graph, DNEFNET and EFNET exhibit similar deblurring effects under low blurring conditions. However, as the blurring level increases, the PSNR of EFNET decreases significantly, while DNEFNET maintains high PSNR and SSIM, indicating that it is able to effectively recover clear images even in complex, non-uniform blurring environments. DNEFNET reduces blur residue and enhances image quality in extreme blurring situations by accurately capturing localized motion trajectories and enhancing high-frequency details. In contrast, EFNET only performs a simple fusion of event and image features, failing to fully utilize the temporal dynamics of the event data. This limitation makes it difficult to effectively remove blur in highly blurred situations, and the lack of a mechanism to recover high-frequency information ultimately leads to a loss of clarity and detail. Therefore, under the high blur condition of the REBlur dataset, DNEFNET demonstrates stronger deblurring capabilities.

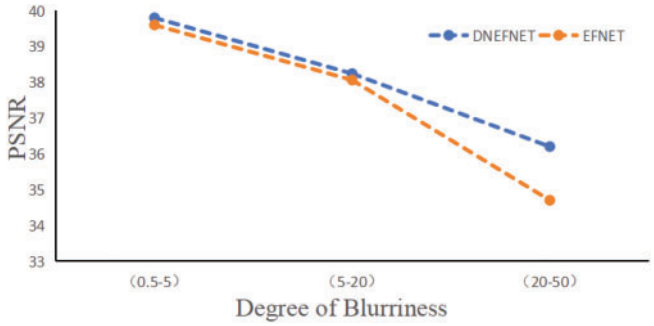

Figure 8: Indicator values for different levels of fuzziness under the REBlur dataset

4.5 Ablation Experiment

On the REBlur dataset, we conducted ablation experiments to investigate the contribution of different modules in the network. Table 3 and Fig. 9 show the results of our defuzzification experiments on the REBlur dataset. First, as shown in Table 3, DTM significantly improves the deblurring effect compared to the original baseline model, with a significant increase in the SSIM value despite a slight decrease in the PSNR value, indicating that the denoising improvement effectively enhances the structural consistency of the image, especially in terms of detail retention. In Fig. 9, we compare the DTM-processed event data with the original event data, and it can be clearly seen that the DTM-processed event information has clearer contours and restores more texture details. This shows that DTM effectively solves the problem of sparse and ineffective representation of valid information in the event stream and, at the same time, improves the expression ability of event features on the time scale. Through subjective performance and qualitative analysis, we verify the effectiveness of the DTM module. Second, the introduction of the RES-FFT residual block shifts the feature processing from the spatial domain to the frequency domain, thus further enhancing the high-frequency details in the fused features. This strategy effectively enhances the model's ability to recover motion edges and detailed textures and improves the reconstruction of high-frequency information. In terms of PSNR and SSIM metrics, the incorporation of RES-FFT improves them by 0.09 dB and 0.3%, respectively. This further demonstrates the advantage of the RES-FFT-based residual block over the traditional U-Net network architecture in processing event-image fusion data and improves the deblurring performance of the model.

Table 3: Quantitative study of different components in the method on the REBlur dataset

Method	PSNR ↑	SSIM ↑
Base	38.12	0.9730
Base+DTM	38.06	0.9753
Base+RES-FFT	38.21	0.9755
Ours	**38.27**	**0.9758**

Note: Bolded meanings denote best results.

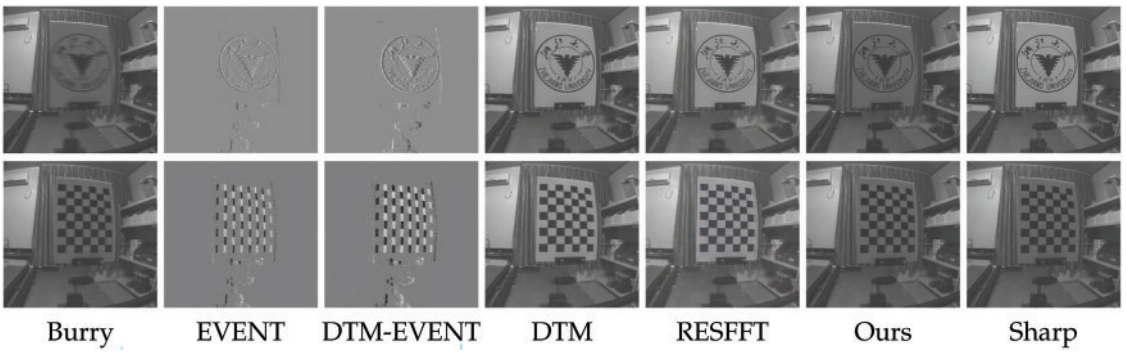

Figure 9: Visual comparison of ablation experiments under the REBlur dataset

5 Conclusion

In this study, we proposed a novel deblurring network architecture, DNEFNET, which successfully addressed the issues of effective information sparsity and limited utilization of high-frequency features in event-based deblurring tasks. To this end, first, we designed the DTM module, which effectively mitigated the problems of redundant noise and information sparsity in the event stream, significantly enhancing the representation of event features. Second, we introduced a residual block based on RES-FFT, which improved the model's ability to recover edge information and reconstruct high-frequency details, thereby compensating for the shortcomings of traditional spatial-domain methods in high-frequency information extraction. This provided an innovative solution for the event-based deblurring task.

Future research can further optimize the proposed method in several directions. Since event data captures the trajectory of a moving object during exposure, the high-frequency features it contains can be leveraged to estimate the object's motion direction and the blurred regions in the image. This feature is not only essential for deblurring tasks but also has wide applications in target tracking. By utilizing the motion trajectory information from event data, the motion trends of the target can be effectively predicted, thus improving the accuracy and robustness of target tracking. Therefore, an important direction for our future research will be to explore how to fully exploit the high-frequency information in event data to enhance the cooperative performance of target tracking and deblurring tasks.

Acknowledgement: The authors are grateful to all the editors and anonymous reviewers for their comments and suggestions.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Kangkang Zhao; draft manuscript preparation: Kangkang Zhao, Yaojie Chen; review: Jianbo Li. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Zhang X, Liu J, Zhang X, Lu Y. Self-supervised graph feature enhancement and scale attention for mechanical signal node-level representation and diagnosis. Adv Eng Inform. 2025;65:103197. doi:10.1016/j.aei.2025.103197.
2. Sun L, Sakaridis C, Liang J, Jiang Q, Yang K, Sun P, et al. Event-based fusion for motion deblurring with cross-modal attention. In: Proceedings of the 17th European Conference on Computer Vision (ECCV 2022); 2022 Oct 23–27; Tel Aviv, Israel: Springer; 2022. p. 412–28. doi:10.1007/978-3-031-19797-0_24.
3. Yang D, Yamac M. Deformable convolutions and LSTM-based flexible event frame fusion network for motion deblurring. arXiv:2306.00834. 2023.
4. Wang P, He J, Yan Q, Zhu Y, Sun J, Zhang Y. Diffevent: event residual diffusion for image deblurring. In: ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2024 May 19–24; Seoul, Republic of Korea: IEEE; 2024. p. 3450–4. doi:10.1109/ICASSP48485.2024.10446822.
5. Stoffregen T, Scheerlinck C, Scaramuzza D, Drummond T, Barnes N, Kleeman L, et al. Reducing the sim-to-real gap for event cameras. In: Computer Vision–ECCV 2020: 16th European Conference; 2020 Aug 23–28; Glasgow, UK: Springer; 2020. p. 534–49.
6. Zhang X, Liu J, Zhang X, Lu Y. Multiscale channel attention-driven graph dynamic fusion learning method for robust fault diagnosis. IEEE Trans Ind Inform. 2024;20(9):11002–13. doi:10.1109/tii.2024.3397401.
7. Mao X, Liu Y, Liu F, Li Q, Shen W, Wang Y. Intriguing findings of frequency selection for image deblurring. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI 2023); 2023 Feb 7–14; Washington, DC, USA: AAAI Press; 2023. p. 1905–13. doi:10.1609/aaai.v37i2.25281.
8. Banham MR, Katsaggelos AK. Digital image restoration. IEEE Signal Process Mag. 1997;14(2):24–41. doi:10.1109/79.581363.
9. Richardson WH. Bayesian-based iterative method of image restoration. J Opt Soc Am A. 1972;62(1):55–9. doi:10.1364/JOSA.62.000055.
10. Bahat Y, Efrat N, Irani M. Non-uniform blind deblurring by reblurring. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy: IEEE; 2017. p. 3286–94.
11. Cho S, Lee S. Fast motion deblurring. In: Proceedings of the ACM SIGGRAPH Asia 2009 Conference; 2009 Dec 1; Yokohama, Japan: ACM; 2009. p. 1–8. doi:10.1145/1661412.1618491.
12. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Proceedings, Part III; 2015 Oct 5–9; Munich, Germany: Springer; 2015. p. 234–41.
13. Zhu X, Hu H, Lin S, Dai J. Deformable convnets v2: more deformable, better results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019); 2019 Jun 16–20; Long Beach, CA, USA: IEEE; 2019. p. 9308–16. doi:10.1109/cvpr.2019.00953.
14. Cho SJ, Ji SW, Hong JP, Jung SW, Ko SJ. Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2021); 2021 Oct 11–17; Montreal, QC, Canada: IEEE; 2021. p. 4641–50.
15. Tao X, Gao H, Shen X, Wang J, Jia J. Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018); 2018 Jun 18–22; Salt Lake City, UT, USA: IEEE; 2018. p. 8174–82.
16. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH, et al. Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021); 2021 Jun 19–25; Virtual Event: IEEE; 2021. p. 14821–31.
17. Zhang X, Zhang X, Liu J, Wu B, Hu Y. Graph features dynamic fusion learning driven by multi-head attention for large rotating machinery fault diagnosis with multi-sensor data. Eng Appl Artif Intell. 2023;125(6):106601. doi:10.1016/j.engappai.2023.106601.
18. Zhang H, Dai Y, Li H, Koniusz P. Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019); 2019 Jun 16–20; Long Beach, CA, USA: IEEE; 2019. p. 5978–86.

19. Pan L, Scheerlinck C, Yu X, Hartley R, Liu M, Dai Y. Bringing a blurry frame alive at high frame-rate with an event camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019); 2019 Jun 16–20; Long Beach, CA, USA: IEEE; 2019. p. 6820–9.
20. Pan L, Hartley R, Scheerlinck C, Liu M, Yu X, Dai Y. High frame rate video reconstruction based on an event camera. IEEE Trans Pattern Anal Mach Intell. 2020;44(5):2519–33. doi:10.1109/TPAMI.2020.3036667.
21. Wang Z, Ren J, Zhang J, Luo P. Image deblurring aided by low-resolution events. Electronics. 2022;11(4):631. doi:10.3390/electronics11040631.
22. Xu F, Yu L, Wang B, Yang W, Xia G, Jia X, et al. Motion deblurring with real events. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2021); 2021 Oct 11–17; Montreal, QC, Canada: IEEE; 2021. p. 2583–92.
23. Kogler J, Sulzbachner C, Kubinger W. Bio-inspired stereo vision system with silicon retina imagers. In: Proceedings of the International Conference on Computer Vision Systems (ICVS 2009); 2009 Apr 1–3; Graz, Austria: Springer; 2009. p. 174–83.
24. Nah S, Kim T-H, Lee M. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017); 2017 Jul 21–26; Honolulu, HI, USA: IEEE; 2017. p. 3883–91.
25. Chen L, Chu X, Zhang X, Sun J. Simple baselines for image restoration. In: European Conference on Computer Vision (ECCV 2022); 2022 Aug 23–27; Tel Aviv, Israel: Springer; 2022. p. 17–33.
26. Tsai FJ, Peng YT, Tsai CC, Lin YY, Lin CW. Banet: a blur-aware attention network for dynamic scene deblurring. IEEE Trans Image Process. 2022;31:6789–99. doi:10.1109/TIP.2022.3216216.
27. Chen L, Lu X, Zhang J, Chu X, Chen C. Hinet: half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021); 2021 Jun 19–25; Virtual: IEEE; 2021. p. 182–92.
28. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH. Restormer: efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022); 2022 Jun 19–24; New Orleans, LA, USA: IEEE; 2022. p. 5728–39.
29. Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H. Uformer: a general U-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022); 2022 Jun 19–24; New Orleans, LA, USA: IEEE; 2022. p. 17683–93.
30. Zafar A, Aftab D, Qureshi R, Fan X, Chen P, Wu J, et al. Single stage adaptive multi-attention network for image restoration. IEEE Trans Image Process. 2024;33:2924–35. doi:10.1109/TIP.2024.3384838.
31. Yang D, Yamac M. Motion aware double attention network for dynamic scene deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022); 2022 Jun 19–24; New Orleans, LA, USA: IEEE; 2022. p. 1113–23.
32. Chen H, Teng M, Shi B, Wang Y, Huang T. Learning to deblur and generate high frame rate video with an event camera. arXiv:2003.00847. 2020.
33. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25(6):84–90. doi:10.1145/3065386.