Diffusion-based Event Generation for High-Quality Image Deblurring

Xinan Xie[1] Qing Zhang[1,2*] Wei-Shi Zheng[1,2]
[1]School of Computer Science and Engineering, Sun Yat-sen University, China
[2]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
xiexn9@mail2.sysu.edu.cn zhangq93@mail.sysu.edu.cn wszheng@ieee.org

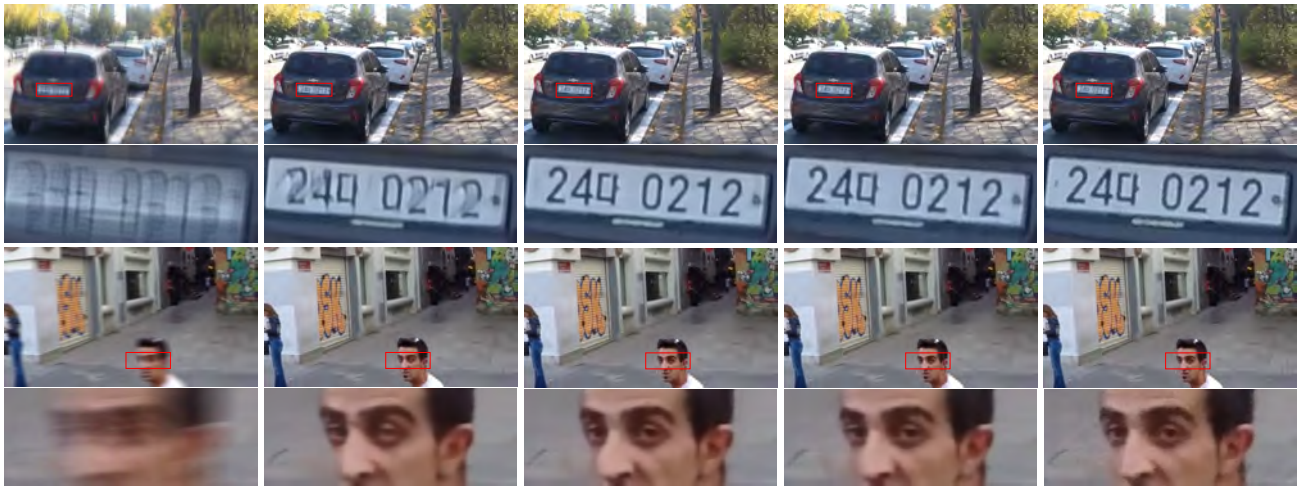


Figure 1. Comparison with previous methods on image blurring. As shown, by leveraging event guidance generated by a diffusion model, our method allows to produce high-quality deblurring results for images with complex motion blur, which are comparable to results of EFNet [43], a recent event-based deblurring method using GT event as input, and clearly better than results produced by HI-Diff [5]), a state-of-the-art image-only deblurring method.

Abstract

While event-based deblurring have demonstrated impressive results, they are impractical for consumer photos captured by cell phones and digital cameras that are not equipped with the event sensor. To address this problem, we in this paper propose a novel deblurring framework called Event Generation Deblurring (EGDeblurring), which allows to effectively deblur an image by generating event guidance describing the motion information using a diffusion model. Specifically, we design a motion prior generation diffusion model and a feature extractor to produce prior information beneficial for deblurring, rather than generating the raw event representation. In order to achieve effective fusion of motion prior information with blurry images and produce high-quality results, we develop a regression deblurring network embedded with a dual-attention channel fusion block. Experiments on multiple datasets demonstrate that our method outperforms state-of-the-art image deblurring methods. Our code is available at https://github.com/XinanXie/EGDeblurring.

1. Introduction

Image deblurring is an important problem in computer vision, aiming to recover a sharp image from the original blurred input. Traditional methods primarily address the problem by formulating an optimization that incorporates various hand-crafted priors and assumptions [2, 3, 7, 11, 21, 24, 32, 39, 54]. In recent years, learning-based methods [4, 8, 31, 42, 45, 47, 56] have made significant progress. However, they are often less effective for images with complex motion blur, as shown in Figure 1.

To enhance the robustness to complex blurry images, event-based deblurring methods [17, 27, 33, 40, 53, 60] have been developed, where the motion information cap-

*Corresponding author.

tured by event camera is utilized to achieve high-quality image deblurring. Although these methods demonstrate impressive results, they require events as additional inference input, which cannot be recorded by most consumer cameras, thereby having relatively limited practicability. Diffusion models [15] have recently gained widespread attention due to its powerful generation capability. This motivates us to explore the possibility of using diffusion models to generate event guidance for blurred images.

In this paper, we present EGDeblurring, a method that allows high-quality image deblurring by utilizing motion information from event guidance generated by a diffusion model. EGDeblurring consists of a motion prior generation diffusion model (MPG-Diff) and a regression deblurring network (RDNet), where MPG-Diff is a diffusion model used for generating event recording the motion information of a blurry image, while RDNet aims to recover a sharp image from the blurry image and the associated events. Note, we design MPG-Diff to generate event-based motion prior features instead of raw events. The reason behind is that raw event usually appears in terms of sparse representation, while diffusion models excel at generating dense representation with rich details. To ensure that MPG-diff generates reliable event guidance, we instead choose to predict motion prior feature rather than raw event. To effectively utilize the motion information recorded by events, we embed dual-attention channel fusion block (DACFB) at multiple levels of the RDNet, which enables aligning and fusing motion information features with image features. To enable better performance, a multi-stage training strategy is further introduced. During inference, our method receives only blurry images as input as it can generate event guidance by itself.

In summary, we make the following main contributions:

- We present a method that allows high-quality image deblurring by generating event guidance describing the motion information using a diffusion model.
- We develop a diffusion model for generating event guidance and a network to produce sharp image from a blurry image and its predicted event. Besides, we introduce an effective multi-stage training strategy.
- We compare our method with various state-of-the-art image deblurring methods on benchmark datasets. Results show that our method significantly outperforms existing image-only methods, and can produce comparable results to event-based methods with ground truth event as inference input.

2. Related Work

Image-only methods. Early image deblurring methods typically work by estimating the blur kernel using various natural image priors and assumptions [2, 3, 11, 21, 22, 25, 54]. However, these methods are mostly less effective for images with complex motion blur, as their leveraged priors and assumptions may not always hold in real-world scenarios. In the past decade, numerous deep learning based methods are developed for image deblurring [1, 4, 8, 16, 31, 38, 42, 45, 47, 56, 58] from different perspectives including multi-scale [31, 45], coarse-to-fine [8], and multi-stage [4]. More recently, transformer [20, 46, 57] and frequency analysis [10, 20] have been applied to image deblurring, which demonstrates promising results. However, due to lack of reliable modeling of the motion information, existing learning-based methods basically struggle to recover sharp images with clear and well-preserved details, especially when facing with blurry images incurred by complex object and camera motion.

Event-based methods. As event carries the motion information beneficial to image deblurring, there are various event-based methods [6, 14, 17, 19, 27, 33, 34, 40, 53, 60, 61] in the literature. Pan *et al.* [33] proposed an event-based double integral (EDI) model that integrates blurred images and events for deblurring. Shang *et al.* [40] developed a event fusion module that can be incorporated into existing image and video deblurring networks. RED-Net [53] proposes to combine optical flow estimation and for self-supervised deblurring. EFNet [43] designs an event-image cross-modal attention module to fuse event and image features. Although event-based methods achieve superior deblurring performance by leveraging motion information, to our knowledge, all of them require ground truth event as addition inference input, which limits their practicability.

Diffusion-based methods. Recently, diffusion models have demonstrated impressive performance in various low-level vision tasks including image/video deblurring [13, 26, 37, 51]. DvSR [51] presents a framework for stochastic blind image deblurring with a focus on perceptual quality using a conditional diffusion model. Ren *et al.* [36] proposed to improve the performance by leveraging multi-scale structure guidance. DiffIR [52] achieves efficient diffusion-based image restoration. HI-Diff [5] applies diffusion model in a highly compacted latent space to generate the prior feature for the deblurring process. There is also attempt towards integrating event information into diffusion models. Diffevent [50] formulates event-based image deblurring as an image generation problem by designing diffusion priors for both the image and residuals.

3. Method

Figure 2 presents the overview of the proposed EGDeblurring. As shown, during inference, our method only receives single blurry image as input. The EGDeblurring encompasses two key modules: the RDNet and the MPG-Diff. The RDNet embed the DACFB is designed to restore a blurred image into the corresponding sharp one, guided

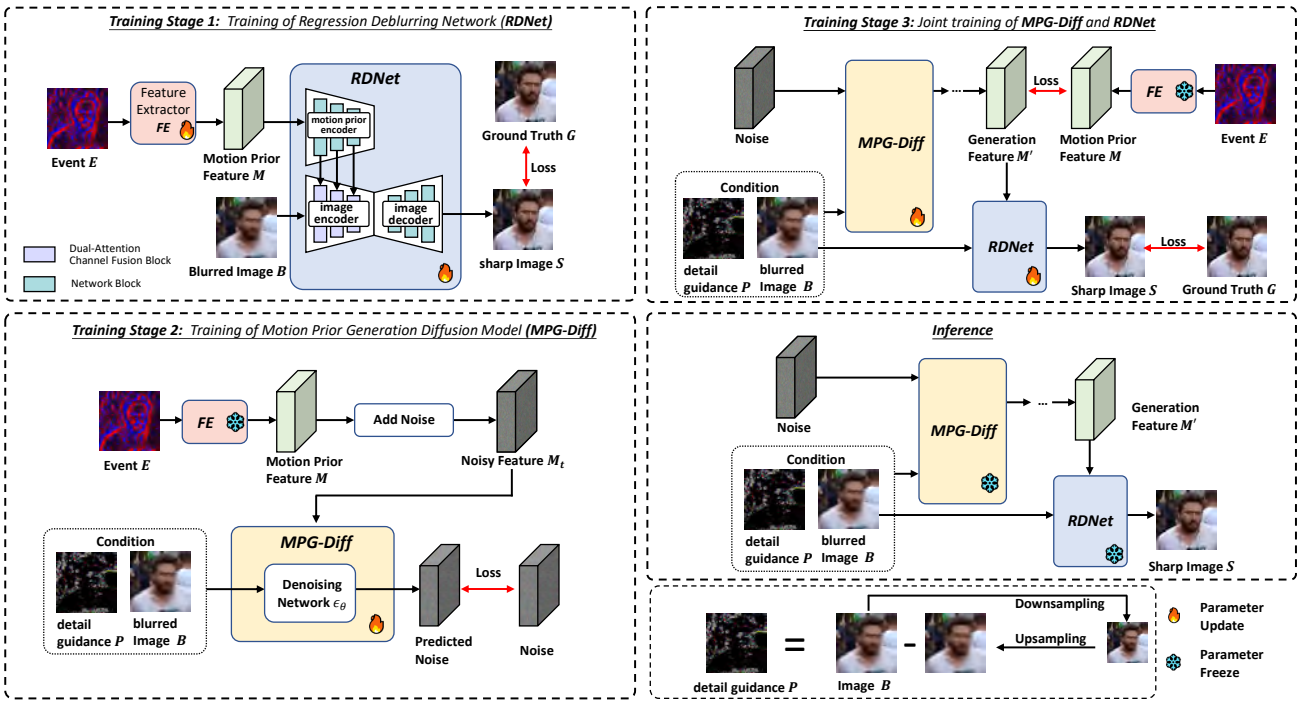


Figure 2. Overview of EGDeblurring. The EGDeblurring incorporates events during the training process, while it does not require event input during the inference stage, which is consistent with the image-only approach. The EGDeblurring are trained with a multi-stage training strategy. In Stage 1,the FE transforms the event representation E into the motion prior feature representation M which guides the RDNet to restore the sharp image S from the blurred image B. In Stage 2, the MPG-Diff is trained to generate motion prior feature M' conditioned on the blurred image B and the detail guidance P. In Stage 3, the MPG-Diff and the RDNet are jointly trained to achieve optimal deblurring performance. During inference, given a blurry image B, the MPG-Diff generates the corresponding motion prior feature M', guiding the RDNet to output the final sharp result S.

by motion prior features from events or the MPG-Diff. The MPG-Diff accepts the blurred image and the corresponding detail guidance as conditions, generating motion prior features. We adopt a multi-stage training strategy. In training stage 1 and training stage 2, the RDNet and MPG-Diff are trained respectively. Subsequently, in training stage 3, both are jointly trained to achieve coordination between the modules. During the inference process, given a blurred image, the MPG-Diff generates the corresponding motion prior features. The RDNet then utilizes these prior features to achieve high-quality image deblurring, eliminating the need for event assistance. Therefore, during inference, the EGDeblurring method is consistent with the image-only approach.

3.1. Regression Deblurring Network

To avoid undesired artifacts and achieve better results, we design the RDNet to perform image deblurring. The RDNet consists of a motion prior encoder, an image encoder, and an image decoder. Meanwhile, in order to match the generation of the subsequent diffusion model, we employ a Feature Extractor (FE) composed of a deep neural network to convert the event representation $E \in \mathbb{R}^{C_e \times H \times W}$ into a feature representation $M \in \mathbb{R}^{C_m \times H \times W}$. The motion prior encoder encodes M from the FE in training stage 1 (or M' from the MPG-Diff in training stage 3 and inference) into multi-scale motion prior features $F_m^k \in \mathbb{R}^{C' \times H' \times W'}$, where $k = 0, 1, 2$ and $C' = 2^k C_b, H' = H/2^k, W' = W/2^k$. C_b denotes the base number of channels. The blurred image $B \in \mathbb{R}^{3 \times H \times W}$ is encoded by the image encoder into image features F_i^k. F_i^k and F_m^k at the same scale are fused through the DACFB embedded in the image encoder. The image features are processed through the image decoder to produce the sharp image $S \in \mathbb{R}^{3 \times H \times W}$, which serves as the final output of the framework.

Details of DACFB. We propose a DACFB as the modal information regulator for the network. The utilization of a multi-branch structure enables the network to learn multiple sets of cross-modal values to modulate the image features. As shown in Figure 3, layer normalization and 1×1 convolutional layers process F_m^k and F_i^k to obtain $K \in \mathbb{R}^{H'W' \times C'}$ and $Q \in \mathbb{R}^{H'W' \times C'}$ respectively. The cross-channel attention map A is computed from K and Q. To ensure suf-

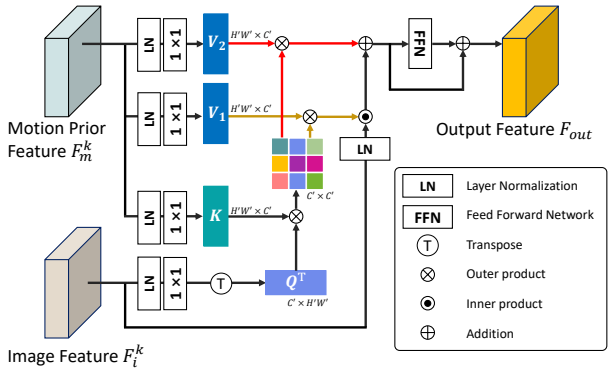


Figure 3. Network architecture of DACFB. The DACFB is embedded within the image encoder of the RDNet, aiming to fuse the motion feature F_m^k and the image feature F_i^k.

ficient modulation of features, we extract V_1 and V_2 from F_m^k, where $V_1, V_2 \in \mathbb{R}^{H'W' \times C'}$. The vectors V_1 and V_2 are respectively taken to compute the outer products with A, resulting in A_1 and A_2. Then the outputs of the attention operation are separately multiplied and added to F_i^k to obtain the output feature F'_{out}. F'_{out} is passed to the feed forward network and outputs the final result F_{out}. It can be represented by the following formula:

$$\begin{aligned} A &= softmax(\frac{Q^T K}{\sqrt{d}}), \\ A_1 &= V_1 \times A, \\ A_2 &= V_2 \times A, \\ F'_{out} &= LN(F_i^k) \cdot A_1 + A_2, \\ F_{out} &= FFN(F'_{out}) + F'_{out}, \end{aligned} \tag{1}$$

where $LN(\cdot)$ represents layer normalization and $FFN(\cdot)$ represents the feed forward network.

3.2. Motion Prior Generation Diffusion Model

We employ the MPG-Diff to generate the motion prior feature M' as a substitute for the motion information features M from events. We employ a conditional denoising diffusion probabilistic model [9, 15] to guide the generation of the corresponding motion prior information. We perform the forward diffusion process and the reverse denoising process on the motion prior features M instead of the events E.

In the forward diffusion process, M transformed from E by the FE is considered as the starting point of the forward Markov process (i.e., $M = M_0$). The diffusion model learns a Markov chain by progressively corrupting the target feature M_0 with Gaussian noise at each time step t, thereby transforming it into a Gaussian distribution (referred to as $\mathcal{N}$). The forward diffusion process is formulated as:

$$q(M_t|M_{t-1}) = \mathcal{N}(M_t; \sqrt{1-\beta_t}M_{t-1}, \beta_t \mathbf{I}), \tag{2}$$

where M_t is the feature at time step t, β_t is the predefined variance, and $\mathbf{I}$ represents the identity matrix for ensuring isotropic of the noise. After reparameterization [15], we obtain M_t from M_0 in a single step by:

$$M_t = \sqrt{\bar{\alpha}_t}M_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{3}$$

where $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$ is the cumulative product of forward noise weight parameters $\alpha_i = 1 - \beta_i$, while $\epsilon \sim \mathcal{N}(0, I)$ denotes random Gaussian noise.

In the reverse denoising process, Gaussian noise is iteratively denoised into the motion prior feature M'. We adopt the DDIM [41] sampling strategy. To synthesize the corresponding motion prior features, we utilize the blurred image B and detail guidance P as conditions. The reverse denoising process is defined as follows:

$$\begin{aligned} M_{t-1} &= \sqrt{\bar{\alpha}_{t-1}}(\frac{M_t - \sqrt{1-\bar{\alpha}_t} \cdot \mathbf{e}_t}{\sqrt{\bar{\alpha}_t}}) + \sqrt{1-\bar{\alpha}_{t-1}} \cdot \mathbf{e}_t, \\ \mathbf{e}_t &= \epsilon_\theta(M_t, B, P, t), \end{aligned} \tag{4}$$

where $\mathbf{e}_t$ is the noise estimated by the neural network ϵ_θ.

Detail guidance. To better constrain the diffusion model to output highly relevant information, we explore extracting details from the blurred image. We observed that when the image is downsampled, further detail information is lost. Thus, the difference between the downsampled image and the original image can serve as a guidance for the image details. We propose $P = B - \mathcal{T}(B)$ as the detail guidance, where $\mathcal{T}$ represents the operation of downsampling the image and then upsampling it back to its original size.

3.3. Multi-Stage Training Strategy

We adopt a three-stage joint training strategy to achieve synchronization between the modules.

Stage 1: training of RDNet. During the first stage of training, our goal is to introduce events and train a high-performance deblurring network. The training process at this stage is similar to that of event-based deblurring. We concatenate the FE and the RDNet in series and perform joint training. The event representation E is transformed into the feature representation M through the FE. Next, the RDNet accepts the M and the blurred image B and outputs the sharp image S. The overall process of training stage 1 can be expressed as follows:

$$\begin{aligned} M &= FE(E), \\ S &= RDNet(B, M). \end{aligned} \tag{5}$$

At this stage, we aim for the network to achieve optimal deblurring performance. We employ L_1 loss to constrain the

network output and jointly optimize the FE and the RDNet. The loss function is defined as follows:

$$L_{stage1} = ||S - G||_1, \tag{6}$$

where G represents the corresponding ground truth.

Stage 2: Training of MPG-Diff. We train the MPG-Diff using the basic training strategy of diffusion models. Given the event E, we first use the FE trained in the first stage to synthesize the M. Subsequently, M_t is generated using Equation 3. In the reverse denoising process, the blurred image B and detail guidance P serve as conditions, and the denoising network predicts the noise $\mathbf{e}_t$. We train the denoising network ϵ_θ using the following loss function:

$$L_{stage2} = E_{M_t,t,B,P,\epsilon\sim\mathcal{N}(0,I)}||\mathbf{e}_t - \epsilon||_2^2. \tag{7}$$

Note that the parameters of FE are frozen in this stage.

Stage 3: Joint training of MPG-Diff and RDNet. Since in training stage 2, we only trained the MPG-Diff without training the FE and the RDNet, the output M' from the MPG-Diff exhibits slight differences compared to the output M generated by the FE. As a result, directly using M' to guide the RDNet will not achieve optimal deblurring performance. To address this, we perform joint training of the MPG-Diff and the RDNet. Specifically, the MPG-Diff performs a complete iterative denoising process to obtain the predicted motion prior M'. Then we input both M' and the blurred image B into the RDNet to generate the deblurred result S. By constraining the final result, we jointly optimize the MPG-Diff and the RDNet. The loss function is defined as follows:

$$L_{stage3} = ||S - G||_1 + \lambda \cdot ||M' - M||_1, \tag{8}$$

where λ is the weight factor and M is the output of the FE. Since the denoising process is iterated multiple times, to reduce memory usage, we only record the gradients of the self-attention modules in the network ϵ_θ and update the weights of that specific part of the network.

3.4. Inference

Given a blurred image B, the MPG-Diff generates the corresponding motion prior features M'. The RDNet takes B and M' as inputs and outputs the sharp image S as the final result. The overall process is represented by the following equation:

$$\begin{aligned} M' &= Diff(B), \\ S &= RDNet(B, M'), \end{aligned} \tag{9}$$

where $Diff(\cdot)$ represents the complete iterative denoising process of the MPG-Diff using the network ϵ_θ.

Image-only method	SSIM↑	PSNR↑	Param.(M)
DeepDeblur [31]	0.914	29.08	11.7
DeblurGAN-v2 [23]	0.934	29.55	60.9
SRN [45]	0.934	30.26	6.8
DBGAN [59]	0.942	31.10	11.6
DMPHN [58]	0.940	31.20	21.7
Ren et al. [36]	0.954	32.11	52.0
MIMO-UNet+ [8]	0.957	32.45	16.1
HINet [4]	0.959	32.91	88.7
Restormer [57]	0.961	32.92	26.1
Stripformer [46]	0.962	33.08	20.0
HI-Diff [5]	0.964	33.33	28.49
MLWNet [12]	0.968	33.83	95.1
LoFormer [30]	0.969	34.09	49.0
MISC Filter [28]	0.969	34.10	16.0
FFTformer [20]	0.968	34.21	16.6
SegFFTformer [18]	0.970	34.38	-
AdaRevD [29]	0.972	34.60	220.8
Ours	0.973	34.62	80.3
D^{2}Net [40]	0.940	31.60	32.63
eSL-Net [49]	0.954	33.52	0.19
MADANET+ [55]	0.964	33.84	16.9
Vitoria et al. [48]	0.944	34.33	5.7
EFNet [43]	0.972	35.46	8.47
Ours *Stage 1*	0.974	35.50	20.7

Table 1. Quantitative comparison with state-of-the-art deblurring methods on synthetic GoPro dataset. Red background represents the best performance. Orange background represents the second best. yellow box represents the third best. Ours *Stage 1* represents the framework trained in Stage 1 with events. The underline denotes event-based method.

3.5. Implementation Details

In the RDNet, we set $C_b = 64$ and concatenate two DACFB modules at each scale of the image encoder. In the MPG-Diff, we use 1000 diffusion steps and a nosie schedule β_t linearly increasing from 1e-6 to 1e-2 for training, and 5 steps for inference. During training, we set the batch size to 4 and use the ADAM optimizer with an initial learning rate of 1×10^{-4}. We implement the algorithm in PyTorch and run it on the NVIDIA RTX 4090 GPU.

4. Experiments

4.1. Datasets and Metrics

GoPro Dataset. We evaluate the effectiveness of the method on the GoPro dataset [31], which is widely used in both image-only and event-based deblurring research. We use ESIM [35] to construct synthetic event data and create an image-like representation from the events [43].

REBlur Dataset. To reflect the performance on real-world

(a) Input (0.638, 21.31) (b) HINet[4] (0.730, 23.74) (c) FFTformer[20] (0.882, 29.37) (d) AdaRevD[29] (0.842, 27.53)

(e) HI-Diff[5] (0.836, 27.02) (f) Restormer [57] (0.793, 25.39) (g) Ours (0.889, 30.91) (h) Ground Truth

Figure 4. Qualitative results on the GoPro dataset. The values in parentheses represent (SSIM, PSNR).

(a) Input (0.833, 25.90) (b) HINet[4] (0.923, 32.51) (c) FFTformer[20] (0.934, 33.64) (d) AdaRevD[29] (0.935, 33.57)

(e) HI-Diff[5] (0.929, 32.91) (f) Restormer [57] (0.924, 32.52) (g) Ours (0.944, 35.24) (h) Ground Truth

Figure 5. Qualitative results on the GoPro dataset.

event datasets, we validate our approach on the REBlur dataset [43], which contains 1,469 group of blurred-sharp image pairs with the associated events captured by event cameras.

Metrics. We adopt two common metrics to quantitatively evaluate the performance of our method, including SSIM and PSNR, which are used to measure the similarity between the predicted results and the corresponding ground truth sharp images.

4.2. Comparison on Synthetic Datasets

Quantitative results on GoPro datasets are presented in Table 1, while visual results are shown in Figures 1, 4 and 5. We compare the model trained in stage 1 with event-based methods, while comparing the final EGDeblurring with image-only methods, as it only inputs the blurred image during inference. As shown in Table 1, the performance of our method in Stage 1 is better than that of other event-based methods, showing that our RDNet enables effective event-based image deblurring. In quantitative results, EGDeblurring outperforms all image-only deblurring methods. Even though our method has a relatively modest improvement compared to high-performance methods such as AdaRevD [29], we have achieved performance at the

Image-only method	SSIM↑	PSNR↑
SRN [45]	0.961	35.10
FFTformer [20]	0.960	35.44
Restormer [57]	0.959	35.50
HINet [4]	0.965	35.58
Stripformer [46]	0.960	35.66
HI-Diff [5]	0.963	35.91
Ours	0.969	36.21
EDI [33]	0.964	36.52
EFNet [43]	0.975	38.12
REFID [44]	0.975	38.34
Ours *Stage 1*	0.977	38.51

Table 2. Quantitative comparison with state-of-the-art methods for deblurring task on real world dataset: REBlur.

method	SSIM↑	PSNR↑
w/ raw event	0.921	32.33
w/o MPG-Diff	0.924	32.47
w/o DACFB	0.967	33.91
w/o the detail guidance	0.963	33.84
Ours	0.973	34.62

Table 3. Ablation studies on the GoPro dataset.

same level while the scale of our model is much smaller than theirs. This is attributed to the combination of the diffusion model and the regression network, which demonstrates advantages in distortion-based quantitative results. As shown in the visual results in Figure 4 and 5, under complex motion blur, other methods struggle to fully remove the blur due to the absence of motion information guidance. In contrast, our approach leveraging generated event guidance allows to recover high-quality sharp image. This indicates that utilizing the diffusion model to generate motion prior information can clearly enhance the overall deblurring performance and the robustness to deal with images with complex motion blur.

4.3. Comparison on Real-world Dataset

Quantitative results are presented in Table 2, while visual results are shown in Figure 6. In the quantitative results, our models from Stage 1 and Stage 3 perform the best among event-based methods and image-only methods, respectively. Our method achieves superior quantitative results, with a 0.3 dB improvement in PSNR compared to the second best method. The visual results in Figure 6 demonstrate that other methods leave residual blur along the edges of moving objects and exhibit noticeable ghosting artifacts, whereas our method restores sharp edge details. This validates the advantage of using the diffusion model to incorporate motion information into our approach.

4.4. Ablation Study

The results of the ablation studies are presented in Table 3 and Figure 7, where we investigate the contribution of each our design to our method.

Effect of MPG-Diff. MPG-Diff is responsible for generating motion prior information. To validate the effectiveness of the diffusion model, we replace it with a convolutional neural network and train it using the same training strategy. Compared to the complete method, the framework without the MPG-Diff shows a decrease in PSNR by 1.64 dB. This indicates that the motion information prior generated by the convolutional neural network is insufficient to provide adequate guidance for the deblurring task. In contrast, the introduction of the diffusion model enhances the robustness of the generated information, leading to improved overall performance of the method.

Metric	1 tensor	2 tensors (**ours**)	3 tensors	4 tensors	5 tensors
SSIM↑	0.969	**0.973**	0.971	0.973	0.972
PSNR↑	34.10	**34.62**	34.61	34.61	34.59

Table 4. Ablation study on different numbers of value tensors in DACFB.

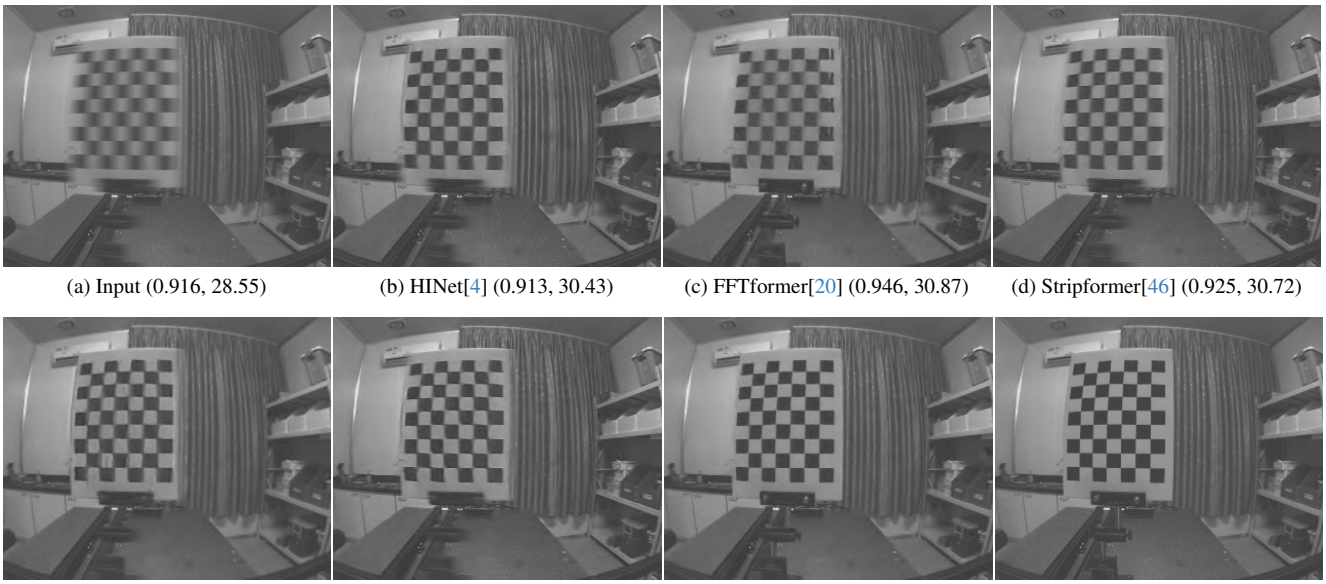
Metric	1× dimension	2× dimension	4× dimension	Ours with two-value
SSIM↑	0.969	0.970	0.971	**0.973**
PSNR↑	34.10	34.12	34.23	**34.62**

Table 5. Ablation study on feature dimensions of the vanilla cross-modal attention block.

Effect of motion prior feature generation. FE is designed to convert the raw event representation into the feature representation, which is the generation target of the diffusion model. Dense features represent motion information similar to event representations while ensuring that the diffusion model can generate corresponding prior features robustly. To validate that dense features is superior to directly generating event representation, we remove the FE and train the diffusion model to directly generate events (as shown in "w/ raw event" in Table 3 and 7). Both quantitative and qualitative results indicate a substantial decline in the framework's deblurring performance. This indicates that introducing dense feature representation is more beneficial to the generation of the diffusion model and the performance of the overall framework.

Effect of detail guidance. The detail guidance P, along with the blurred image, constrain the generation process of the diffusion model. We remove the detail guidance and only use the blurred image to constrain the output of the diffusion model. The deblurring results are negatively impacted, showing a 0.78 dB decrease in PSNR compared to the complete method. This indicates that the detail guidance constrain the output of the diffusion model, leading to the generation of more effective guiding features for deblurring.

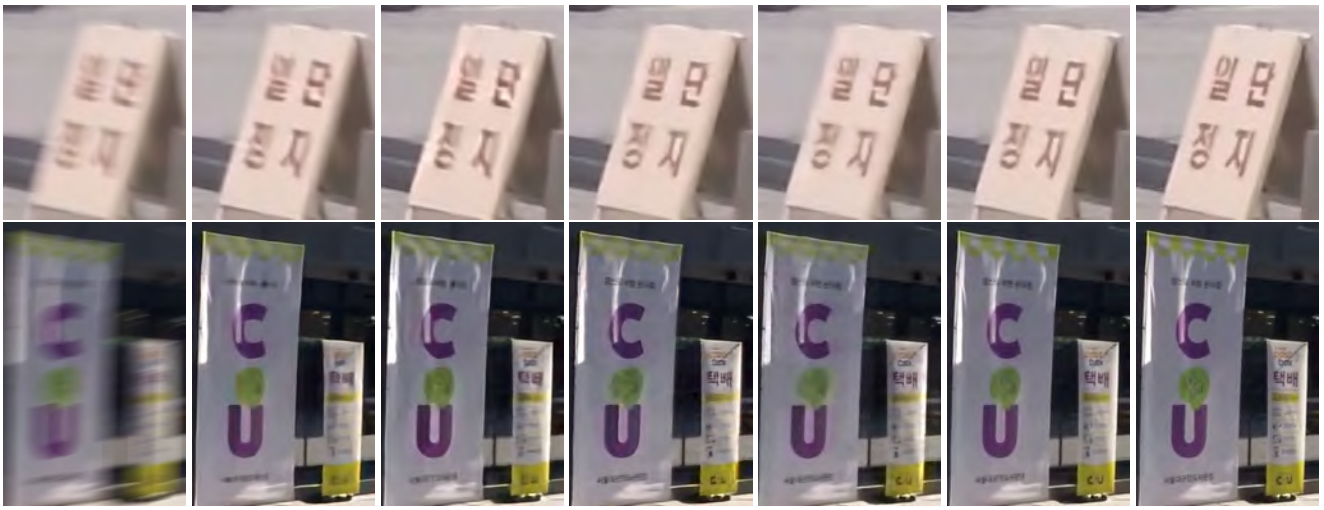
Effect of DACFB. The DACFB embedded in the RDNet fuses motion information features and image features through a cross-modal attention mechanism. We replace the DACFB in the RDNet with simple feature concatenation and convolutional layers, and train the model according to the original strategy. Quantitative results indicate a reduction in deblurring performance, with a decrease of 0.71 dB in PSNR. This demonstrates that, compared to simple concatenation, the DACFB effectively aligns and fuses the two different modalities. We verify the effectiveness of our scheme by adjusting the number of value tensors in the attention mechanism. As shown in Table 4, two value tensors we employ in our DACFB design produces best results, while either decrease or increase in the number of the tensors will lead to performance drop. Note, we show in



(a) Input (0.916, 28.55) (b) HINet[4] (0.913, 30.43) (c) FFTformer[20] (0.946, 30.87) (d) Stripformer[46] (0.925, 30.72)

(e) HI-Diff[5] (0.913, 30.43) (f) Restormer [57] (0.926, 30.96) (g) Ours (0.968, 34.78) (h) Ground Truth

Figure 6. Qualitative results on the REBlur dataset.



Input w/o MPG-Diff w/ raw event w/o DACFB w/o detail guidance Ours Ground Truth

Figure 7. Visual results of ablation experiments.

the Table 5 that our two-value design cannot be replaced by simple feature dimension increase with single-value motion prior feature (see 2nd, 3rd, and 4th columns).

5. Conclusion

We have presented EGDeblurring, a method that allows to achieve high-quality image deblurring by generating event guidance carrying the desired motion prior information using a diffusion model. Specifically, we design a motion prior generation diffusion model to generate event guidance from a single blurry image, and a regression deblurring network that can effectively utilize the motion prior information in the event guidance for image deblurring. Unlike existing event-based methods, our method can test with a single blurry image, without using ground truth event as input. Extensive experiments on both synthetic and real-world blur datasets demonstrate that our proposed method outperforms current state-of-the-art approaches.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (62471499), the Guangdong Basic and Applied Basic Research Foundation (2023A1515030002).

References

[1] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *ECCV*, 2020. 2

[2] Yuval Bahat, Netalee Efrat, and Michal Irani. Non-uniform blind deblurring by reblurring. In *ICCV*, 2017. 1, 2

[3] Leah Bar, Benjamin Berkels, Martin Rumpf, and Guillermo Sapiro. A variational framework for simultaneous motion estimation and restoration of motion-blurred video. In *ICCV*, 2007. 1, 2

[4] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *CVPR*, 2021. 1, 2, 5, 6, 8

[5] Zheng Chen, Yulun Zhang, Liu Ding, Xia Bin, Jinjin Gu, Linghe Kong, and Xin Yuan. Hierarchical integration diffusion model for realistic image deblurring. *NeurIPS*, 2024. 1, 2, 5, 6, 8

[6] Hoonhee Cho, Yuhwan Jeong, Taewoo Kim, and Kuk-Jin Yoon. Non-coaxial event-guided motion deblurring with spatial alignment. In *ICCV*, 2023. 2

[7] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *SIGGRAPH*. 2009. 1

[8] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *ICCV*, 2021. 1, 2, 5

[9] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *NeurIPS*, 2021. 4

[10] Jiangxin Dong, Jinshan Pan, Zhongbao Yang, and Jinhui Tang. Multi-scale residual low-pass filter network for image deblurring. In *ICCV*, 2023. 2

[11] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. Removing camera shake from a single photograph. In *SIGGRAPH*. 2006. 1, 2

[12] Xin Gao, Tianheng Qiu, Xinyu Zhang, Hanlin Bai, Kang Liu, Xuan Huang, Hu Wei, Guoying Zhang, and Huaping Liu. Efficient multi-scale network with learnable discrete wavelet transform for blind motion deblurring. In *CVPR*, 2024. 5

[13] Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In *CVPR*, 2023. 2

[14] Chen Haoyu, Teng Minggui, Shi Boxin, Wang YIzhou, and Huang Tiejun. Learning to deblur and generate high frame rate video with an event camera. *arXiv preprint arXiv:2003.00847*, 2020. 2

[15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 2020. 2, 4

[16] Seo-Won Ji, Jeongmin Lee, Seung-Wook Kim, Jun-Pyo Hong, Seung-Jin Baek, Seung-Won Jung, and Sung-Jea Ko. Xydeblur: divide and conquer for single image deblurring. In *CVPR*, 2022. 2

[17] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning event-based motion deblurring. In *CVPR*, 2020. 1, 2

[18] Insoo Kim, Jae Seok Choi, Geonseok Seo, Kinam Kwon, Jinwoo Shin, and Hyong-Euk Lee. Real-world efficient blind motion deblurring via blur pixel discretization. In *CVPR*, 2024. 5

[19] Taewoo Kim, Jeongmin Lee, Lin Wang, and Kuk-Jin Yoon. Event-guided deblurring of unknown exposure time videos. In *ECCV*, 2022. 2

[20] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, 2023. 2, 5, 6, 8

[21] Jan Kotera, Filip Šroubek, and Peyman Milanfar. Blind deconvolution using alternating maximum a posteriori estimation with heavy-tailed priors. In *CAIP*, 2013. 1, 2

[22] Dilip Krishnan, Terence Tay, and Rob Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, 2011. 2

[23] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*, 2019. 5

[24] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009. 1

[25] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *CVPR*, 2011. 2

[26] Xin Li, Yulin Ren, Xin Jin, Cuiling Lan, Xingrui Wang, Wenjun Zeng, Xinchao Wang, and Zhibo Chen. Diffusion models for image restoration and enhancement–a comprehensive survey. *CVPR*, 2023. 2

[27] Songnan Lin, Jiawei Zhang, Jinshan Pan, Zhe Jiang, Dongqing Zou, Yongtian Wang, Jing Chen, and Jimmy Ren. Learning event-driven video deblurring and interpolation. In *ECCV*, 2020. 1, 2

[28] Chengxu Liu, Xuan Wang, Xiangyu Xu, Ruhao Tian, Shuai Li, Xueming Qian, and Ming-Hsuan Yang. Motion-adaptive separable collaborative filters for blind motion deblurring. In *CVPR*, 2024. 5

[29] Xintian Mao, Qingli Li, and Yan Wang. Adarevd: Adaptive patch exiting reversible decoder pushes the limit of image deblurring. In *CVPR*, 2024. 5, 6

[30] Xintian Mao, Jiansheng Wang, Xingran Xie, Qingli Li, and Yan Wang. Loformer: Local frequency transformer for image deblurring. In *ACM MM*, 2024. 5

[31] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 1, 2, 5

[32] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Deblurring images via dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10):2315–2328, 2017. 1

[33] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *CVPR*, 2019. 1, 2, 6

[34] Liyuan Pan, Richard Hartley, Cedric Scheerlinck, Miaomiao Liu, Xin Yu, and Yuchao Dai. High frame rate video reconstruction based on an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2519–2533, 2020. 2

[35] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *CoRL*, 2018. 5

[36] Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *ICCV*, 2023. 2, 5

[37] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022. 2

[38] Christian J Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Learning to deblur. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(7):1439–1451, 2015. 2

[39] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. *ACM Trans. Graphics*, 27(3):1–10, 2008. 1

[40] Wei Shang, Dongwei Ren, Dongqing Zou, Jimmy S Ren, Ping Luo, and Wangmeng Zuo. Bringing events into video deblurring with non-consecutively blurry frames. In *ICCV*, 2021. 1, 2, 5

[41] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *ICLR*, 2021. 4

[42] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *CVPR*, 2020. 1, 2

[43] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based fusion for motion deblurring with cross-modal attention. In *ECCV*, 2022. 1, 2, 5, 6

[44] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. In *CVPR*, 2023. 6

[45] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, 2018. 1, 2, 5, 6

[46] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *ECCV*, 2022. 2, 5, 6, 8

[47] Fu-Jen Tsai, Yan-Tsung Peng, Chung-Chi Tsai, Yen-Yu Lin, and Chia-Wen Lin. Banet: A blur-aware attention network for dynamic scene deblurring. *IEEE Transactions on Image Processing*, 31:6789–6799, 2022. 1, 2

[48] Patricia Vitoria, Stamatios Georgoulis, Stepan Tulyakov, Alfredo Bochicchio, Julius Erbach, and Yuanyou Li. Event-based image deblurring with dynamic motion awareness. In *ECCV*, 2022. 5

[49] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. Event enhanced high-quality image recovery. In *ECCV*, 2020. 5

[50] Pei Wang, Jiumei He, Qingsen Yan, Yu Zhu, Jinqiu Sun, and Yanning Zhang. Diffevent: Event residual diffusion for image deblurring. In *ICASSP*. IEEE, 2024. 2

[51] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *CVPR*, 2022. 2

[52] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *ICCV*, 2023. 2

[53] Fang Xu, Lei Yu, Bishan Wang, Wen Yang, Gui-Song Xia, Xu Jia, Zhendong Qiao, and Jianzhuang Liu. Motion deblurring with real events. In *ICCV*, 2021. 1, 2

[54] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *CVPR*, 2013. 1, 2

[55] Dan Yang and Mehmet Yamac. Motion aware double attention network for dynamic scene deblurring. In *CVPR*, 2022. 5

[56] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 1, 2

[57] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 2, 5, 6, 8

[58] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *CVPR*, 2019. 2, 5

[59] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *CVPR*, 2020. 5

[60] Xiang Zhang and Lei Yu. Unifying motion deblurring and frame interpolation with events. In *CVPR*, 2022. 1, 2

[61] Xiang Zhang, Lei Yu, Wen Yang, Jianzhuang Liu, and Gui-Song Xia. Generalizing event-based motion deblurring in real-world scenarios. In *ICCV*, 2023. 2