electronics

MDPI

Article

Image Deblurring Aided by Low-Resolution Events

Zhouxia Wang [1,*], Jimmy Ren [2,3], Jiawei Zhang [4] and Ping Luo [1]

[1] Department of Computer Science, The University of Hong Kong, Hong Kong, China; pluo@cs.hku.hk
[2] SenseTime Research, Hong Kong, China; jimmy.sj.ren@gmail.com
[3] Qing Yuan Research Institute, Shanghai Jiao Tong University, Shanghai 200240, China
[4] SenseTime Research, Shenzhen 518067, China; zhjw1988@gmail.com
* Correspondence: wzhoux@connect.hku.hk

Abstract: Due to the limitation of event sensors, the spatial resolution of event data is relatively low compared to the spatial resolution of the conventional frame-based camera. However, low-spatial-resolution events recorded by event cameras are rich in temporal information which is helpful for image deblurring, while intensity images captured by frame cameras are in high resolution and have potential to promote the quality of events. Considering the complementarity between events and intensity images, an alternately performed model is proposed in this paper to deblur high-resolution images with the help of low-resolution events. This model is composed of two components: a DeblurNet and an EventSRNet. It first uses the DeblurNet to attain a preliminary sharp image aided by low-resolution events. Then, it enhances the quality of events with EventSRNet by extracting the structure information in the generated sharp image. Finally, the enhanced events are sent back into DeblurNet to attain a higher quality intensity image. Extensive evaluations on the synthetic GoPro dataset and real RGB-DAVIS dataset have shown the effectiveness of the proposed method.

Keywords: image deblurring; event camera; event super resolution; complementarity

check for updates

Citation: Wang, Z.; Ren, J.; Zhang, J.; Luo, P. Image Deblurring Aided by Low-Resolution Events. *Electronics* **2022**, *11*, 631. https://doi.org/10.3390/electronics11040631

Academic Editors: Leonardo Galteri, Claudio Ferrari and Stefanos Kollias

Received: 16 December 2021
Accepted: 15 February 2022
Published: 18 February 2022

1. Introduction

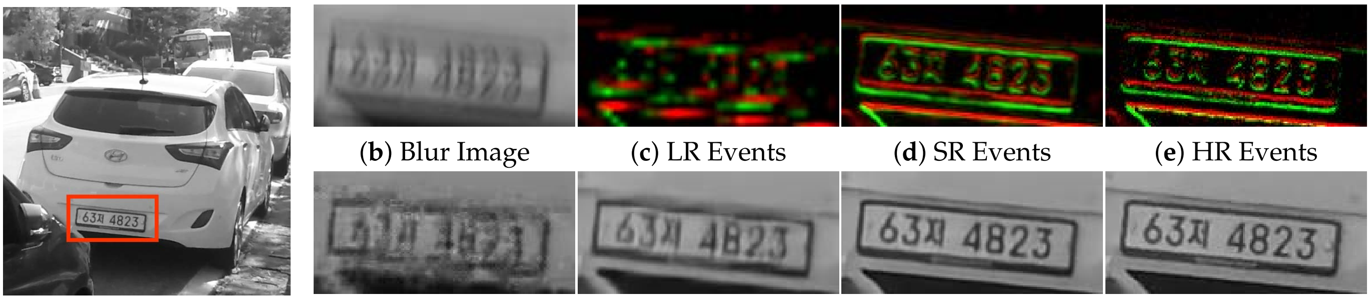
In contrast to the conventional frame-based camera which represents a dynamic scenario with a sequence of still images, an event-based vision sensor [1–4] tends to detect per-pixel brightness changes in microsecond resolution. Once the logarithm of the intensity changes exceeds a preset threshold c in a given pixel (x, y), an event will be triggered which can be formulated as

$$log(\mathbf{I_{xy}}(\mathbf{t})+b)-log(\mathbf{I_{xy}}(\mathbf{t}-\mathbf{\Delta t})+b)=p\cdot c \tag{1}$$

where $\mathbf{I_{xy}}(\mathbf{t})$ and $\mathbf{I_{xy}}(\mathbf{t}-\mathbf{\Delta t})$ denote the intensities at time t and $t-\Delta t$, respectively. Δt is the time since the last event happened in location (x, y). b is a small constant used for avoiding $\mathbf{I_{xy}}(\mathbf{t})$ to be zero. $p\in\{+1,-1\}$ is the polarity representing the direction (increase or decrease) of the intensity change. Finally, an event is represented as $e=(x,y,t,p)$. Since event data are discrete and own high temporal resolution, a high dynamic range, and a low motion blur, the event camera has shown its potential in several robotic and computer vision tasks, such as image reconstruction [5–12], image deblurring [13–15], object detection [16,17], and SLAM [18].

However, the spatial resolution of existing event cameras is still not larger than 1 megapixel [1–4,17], which is far from the 12 or 24 megapixel spatial resolution of existing frame-based cameras. Considering the high temporal events recorded by event cameras and the high-resolution images captured by conventional frame-based cameras, constructing a mixed imaging system with an event camera and a frame-based camera with different spatial resolutions may be a means of achieving high-quality images. A similar imaging system was suggested in [19], where an algorithm is provided for calibrating these two kinds of cameras.However, this work assumes that the intensity images captured by the frame-based camera are sharp, or the intensity image will not produce effectiveness in their method. Since motion in the world is very common and that high-speed frame-based cameras are expensive, it is difficult to acquire sharp intensity images in all scenarios. Therefore, our research is based on low-spatial-resolution events and high-spatial-resolution images with blurry intensity, in the aim of deblurring the image with the help of low-resolution events.

There are some representative works related to event-based image deblurring [14,15]. Pan et al. [15] restored a high frame-rate sharp video by modeling the relationship between the events and the blurry intensity images as a double integral, while Lin et al. [14] implemented this physical model with a neural network and attained high performance in video deblurring and interpolation. However, these methods aim to process the blurry intensity images and events which are in the same spatial resolution. We find experimentally that once the spatial resolution of events degrades, these methods' performance rapidly declines (as shown in Figure 1).



(**b**) Blur Image (**c**) LR Events (**d**) SR Events (**e**) HR Events

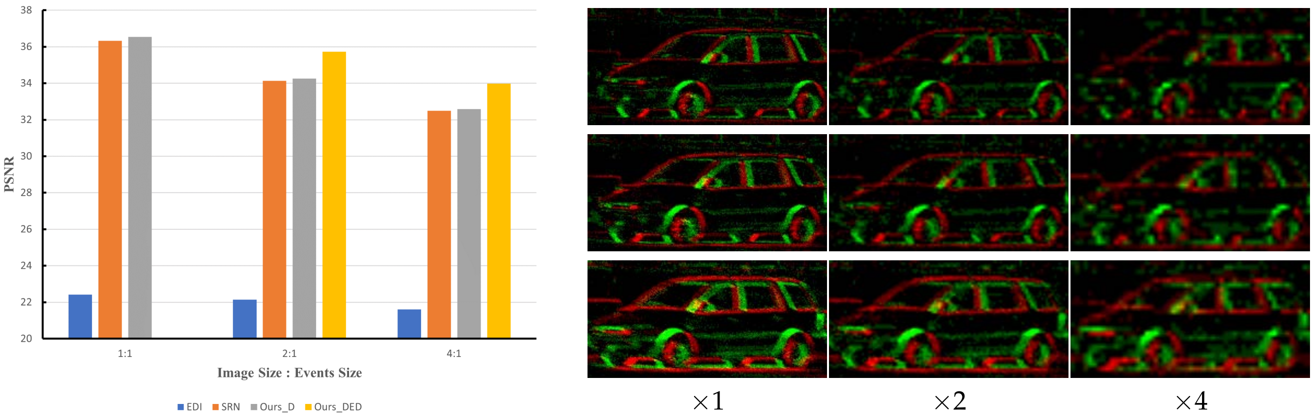
(**a**) Whole View (**f**) Pan et al., 2019 (**g**) Lin et al., 2020 (**h**) Ours_DED (**i**) GT

Figure 1. Results of image deblurring when facing low-resolution events. Methods proposed in [14,15] are presented in terms of their image deblurring potential and both of them aim to process events and intensity images of the same resolution. (**a**) Whole View; (**b**) Blur Image; (**c**) LR Events; (**d**) SR Events; (**e**) HR Events; (**f**) Pan et al., 2019; (**g**) Lin et al., 2020; (**h**) Ours_DED; (**i**) GT. When they confront the events (**c**) whose spatial resolution is 4 times lower than that of the blurry image, their performance becomes worse, as shown in (**f**,**g**). Facing these two kinds of degraded data, this paper tends to update the intensity image and events alternately, and finally attains a higher quality intensity image (**h**) and events (**d**).

Although both events and intensity images are degraded in our research, they still hold significant information. Specifically, on the one hand, the events are considered as changes between a sequence of latent images that lead to the blurring of the final image. Although they are in low spatial resolution, they still own high-resolution temporal information (as shown in Figure 2b), and the high-resolution temporal information is rather helpful for image deblurring. On the other hand, events are triggered by the local edges of the scene and it is possible to reconstruct high-resolution events with the assistance of high-resolution intensity images. Therefore, we propose an alternately performed model for a deblurring image aided by low-resolution events. This model mainly consists of two components: a physical model-based Deblur network (DeblurNet) which takes events and blurry image as inputs and aims to output a sharp high-resolution image; an EventSR network (EventSRNet) which takes low-resolution events and the high-resolution intensity images attained by DeblurNet as inputs and aims to derive the corresponding high-resolution events. Since the spatial resolution of events affects the performance of image deblurring (as shown in Figure 2a, the higher the spatial resolution of events, the better of image deblurring), we sought to attain an even better sharp image by another DeblurNet. At this time, DeblurNet takes the high-quality events generated by EventSRNet as input rather than the original low-resolution events.

The main contributions of this method are as follows.

- It aims to deblur images with the help of by low-resolution events, which is more practical;
- It proposes a model to explicitly restore the degraded intensity image and events by considering the complementarity between them, and an alternated strategy is applied to progressively promote the quality of the intensity image;
- Extensive evaluations of the synthetic GoPro [20] dataset and real RGB-DAVIS [19] dataset show the effectiveness of the proposed method.



(**a**)

(**b**)

Figure 2. (**a**) The image deblurring performance of EDI [15,21] and our method in a different setting, in terms of PSRN; (**b**) The event sequence in a different spatial resolution. The first column is the event sequence that was simulated with a video in GoPro [20] in the original spatial resolution while the event sequences in the next two columns are simulated with the same video that has been downsized to ×2 and ×4, respectively. This shows that events with low-resolution still hold as much rich temporal information as the high-resolution events, which is very helpful for image deblurring.

2. Related Works

2.1. Event Enhancement

Since events captured by neuromorphic vision sensors are brightness changes of scenes, they are highly sensitive to noise. In order to enhance the quality of events, some previous works [22–24] have added an additional pre-processing operation implemented by a spatiotemporal filter into sensors. Other works such as [25] preferred to fuse it into a neural network optimized with motion consistency. Although all these works have shown their advantages in event enhancement, they only focus on denoising and do not consider the situation that the events are in a low spatial resolution. In contrast to previous works, Wang et al. [19] proposed to obtain high-resolution and noise-robust events by joint filtering low-resolution events with high-resolution intensity images. However, the high-resolution intensity images required here are supposed to be very clear and sharp, or they will be disabled in the algorithm. In reality, the assumption is too strong to implement because of the existing low-frame rate cameras and moving scenes. Unlike these methods, our work is based on low-spatial-resolution events and high-spatial-resolution images with blurry intensity and aims to iteratively update events and intensity images with the help of one another.

2.2. Event-Based Image Enhancement

Image enhancement aims to improve the image quality by removing blur, noise, or increasing its resolution. Since event cameras have the advantages of a high dynamic range, high temporal resolution, and low motion blur, there is increasing interest in enhancing the quality of images by event data. The method proposed by Wang et al. [26] could directly reconstruct, restore, and super-resolve images from events by three GANs.

Mostafavi et al. [27] proposed to reconstruct high-resolution intensity images from events with the help of a flownet. The information used in both of these works are just events that are brightness changes, without a base brightness, and the reconstructed images tend to be miss-matched with the real world. Therefore, DAVIS, an event camera that can not only capture events at a high temporal rate but also attain a sequence of low-frame-rate-intensity images (denoted as APS), appears, and a lot of works started based on these events and APS. Pan et al. [15] proposed an event-based double integral model for obtaining a high-frame-rate video from events and a blurry intensity image and Lin et al. [14] implemented the physical model proposed by [15] as a neural network, which achieved a high performance in terms of video deblurring and interpolation. Moreover, Wang et al. [28] unified denoising, deblurring, and super-resolution in one model by an event-enhanced degeneration model and Zhang et al. [29] proposed a hybrid deblur net for image deblurring with learned event representation. These methods have shown their advantage of image enhancement. However, they are all adapted to the intensity images and the events that are in the same and low spatial resolution. They have ignored the advantages (high spatial resolution) of the frame-based camera and once the spatial resolution of events is lower than that of the intensity images, the performance of these methods rapidly declines. Our method in this work is based on the promising premise that events captured by event camera which are in low-resolution and images captured by the conventional frame-based camera which are blurry but have high spatial resolution can function alternately to provide a solution to enhance the blurry intensity image and low-resolution events.

3. Method

This paper aimed to attain a blur-less and high-resolution intensity image with a sequence of low-spatial-resolution events recorded by an event camera and a high-spatial-resolution blurry intensity image captured by a frame-based camera. In order to make full use of the complementarity between these two kinds of data, an alternately performed model consisting of a Deblur network (DeblurNet) and an EventSR network (EventSRNet) is proposed here. DeblurNet is used for deblurring with the help of events data that own temporal information and an EventSR network (EventSRNet) used for enhancing event data assisted by high spatial resolution intensity images with enriched structure information. DeblurNet first produces a preliminary sharp image with original low-resolution events. Then, EventSRNet enhances the quality of events with the generated preliminary sharp image which potentially owns the rich structure information needed by low-resolution events. Since high-quality events yield high-quality intensity images (as shown in Figure 2a), another DeblurNet inputted with the enhanced events generated by EventSRNet is applied for further attaining a better sharp intensity image.

This section aims to introduce the details of the proposed method, and at the beginning, we will introduce the stacking method used in this paper in Section 3.1. Then, the details of the architecture and its learning method will be introduced in Section 3.2 and Section 3.3, respectively. The final part (Section 3.4) is the training procedure.

3.1. Representation of Event Data

In order to adapt the discrete event data to the conventional convolutional neural network, we stack event data based on time which is similar to the SBT proposed in [12]. Specifically, the exposure time for collecting a blurry intensity image $\mathbf{I_b}$ is denoted as $\mathbf{t_b}$. Then, events triggered during $\mathbf{t_b}$ are separated into $\mathbf{n}$ parts. For the ith ($i = 1, 2, \ldots, \mathbf{n}$) part, negative and positive events in the time interval $[\frac{(i-1)\mathbf{t_b}}{\mathbf{n}}, \frac{i\mathbf{t_b}}{\mathbf{n}}]$ are accumulated into two channels $\mathbf{E}^i_-(x, y)$ and $\mathbf{E}^i_+(x, y)$ for each pixel (x, y), respectively. That is, while $e = (x, y, t, p)$:

$$\begin{aligned}\mathbf{E}_{-}^{i}(x,y) &= \sum_{t=\frac{(i-1)\mathbf{t_b}}{\mathbf{n}}}^{\frac{i\mathbf{t_b}}{\mathbf{n}}} p, \quad p=-1,\\ \mathbf{E}_{+}^{i}(x,y) &= \sum_{t=\frac{(i-1)\mathbf{t_b}}{\mathbf{n}}}^{\frac{i\mathbf{t_b}}{\mathbf{n}}} p, \quad p=1.\end{aligned} \tag{2}$$

In conclusion, events for $\mathbf{I_b}$ can be expressed as $\mathbf{E} = \left[\mathbf{E}_{-}^{1}, \mathbf{E}_{+}^{1}, \ldots, \mathbf{E}_{-}^{n}, \mathbf{E}_{+}^{n}\right]$, and $\mathbf{E} \in \mathbf{R}^{2\mathbf{n} \times H \times W}$, where H and W are the height and width of image $\mathbf{I_b}$.

3.2. Architecture

The overview of the architecture is shown in Figure 3. It contains two DeblurNet and one EventSRNet. The first DeblurNet aims to yield a preliminary sharp image with the low-resolution events. This preliminary sharp image will be sent into the EventSRNet to guide the enhancement of events with its rich structure information. Then, another DeblurNet takes the enhanced events generated by EventSRNet to further enhance the quality of the sharp image. Both DeblurNet and EventSRNet are implemented with neural networks. The following was detailed in the introduction.

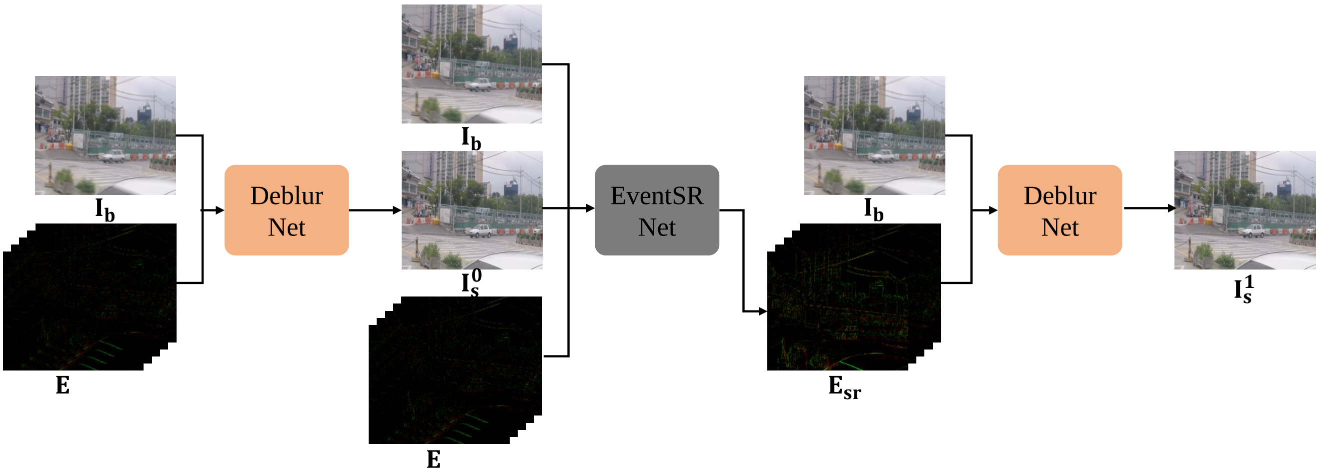


Figure 3. Framework of the proposed model. It contains two components: DeblurNet and EventSRNet. The whole reference process consists of three steps. The first step is to attain the preliminary sharp intensity images $\mathbf{I_s^0}$ from low-resolution events $\mathbf{E}$ and high-resolution blurry intensity image $\mathbf{I_b}$. The second step tends to promote the resolution of events $\mathbf{E}$ to $\mathbf{E_{sr}}$ with the help of $\mathbf{I_b}$ and $\mathbf{I_s^0}$. Finally, the super-resolved events $\mathbf{E_{sr}}$ and $\mathbf{I_b}$ are sent into another DeblurNet for achieving a better sharp intensity image.

3.2.1. DeblurNet

The method proposed by Lin et al. [14] has shown the efficiency of the event-based convolutional neural network implemented with a physical model for video deblurring. We tend to adopt this method for event-based image deblurring. Since the original model is designed for video deblurring and interpolation, by abandoning the video-related components, the IntegralNet implemented with an Unet [30] and a dynamic filter network [31] remained as our DeblurNet. Only a blurry intensity image $\mathbf{I_b}$ and its events $\mathbf{E}$ which occurred during $\mathbf{t_b}$ are required as inputs. Therefore, the DeblurNet can be formulated as

$$\mathbf{I_s^0} = \mathit{DeblurNet}(\mathbf{I_b}, \mathbf{E}). \tag{3}$$

3.2.2. EventSRNet

EventSRNet is implemented with a Unet with skip connections. For making full use of the complementarity between intensity images and events, EventSRNet takes the combination of a blurry intensity image $\mathbf{I_b}$, events $\mathbf{E}$, and the sharp image $\mathbf{I_s}$ generated by DeblurNet as input. Its goal is to generate a sequence of high spatial resolution events $\mathbf{E_{sr}}$. This network can be formulated as

$$\mathbf{E_{sr}} = EventSRNet([\mathbf{I_s}, \mathbf{I_b}, \mathbf{E}]). \tag{4}$$

More details of these two networks are provided in Appendix A.

3.2.3. Deblurring Refinement

For further achieving a high-quality intensity image, the super-resolved events and blurry intensity image will be sent back into the DeblurNet, and it can be formulated as

$$\mathbf{I_s^1} = DeblurNet(\mathbf{I_b}, \mathbf{E_{sr}}). \tag{5}$$

3.3. Learning

3.3.1. DeblurNet Loss

DeblurNet is constrained with MSE loss:

$$\mathcal{L}_d = \frac{1}{HW}(\alpha\|\mathbf{I_s^0} - \hat{\mathbf{I}}\|^2 + \beta\|\mathbf{I_s^1} - \hat{\mathbf{I}}\|^2) \tag{6}$$

where H and W are the height and width of high-resolution intensity images, respectively. $\hat{\mathbf{I}}$ is the groundtruth of intensity images. α and β are the loss weight of the first deblurring and the second deblurring.

3.3.2. EventSRNet Loss

Since the events are the records of brightness changes, they tend to be sparse, especially when they have been stacked. Most values in $\mathbf{E}$ are zeros. Standard regression loss is not appropriate for the learning of EventSRNet. For attaining a better convergence, we adopt an asymmetric L1 loss rather than the standard L1 loss here. The asymmetric L1 loss will pay more attention to the place where occurs events and pay less attention to the place that has no brightness change. It can be formulated as

$$\mathcal{L}_e = \frac{1}{HW}(\mathcal{L}_{e+} + \gamma\mathcal{L}_{e-}) \tag{7}$$

where:

$$\begin{aligned} \mathcal{L}_{e+} &= |\mathbf{E_{sr}}(x,y) - \mathbf{E_{hr}}(x,y)|, \quad && \mathbf{E_{hr}}(x,y) > 0 \\ \mathcal{L}_{e-} &= |\mathbf{E_{sr}}(x,y)|, && \mathbf{E_{hr}}(x,y) == 0 \end{aligned} \tag{8}$$

where (x,y) is the coordination of stacked events and $\mathbf{E_{hr}}$ is the high spatial resolution events. While $\gamma < 1$, the place where no event has happened will apply less effect on $\mathcal{L}_e$. In our experiments, $\gamma = 0.1$.

In conclusion, the whole loss function is:

$$\mathcal{L} = \mathcal{L}_d + \theta\mathcal{L}_e \tag{9}$$

3.4. Training Procedure

There are three steps for training:

- Training the first DeblurNet with low-resolution events upsampled by bilinear interpolation and blurry intensity images. Here, $\alpha = 1.0$, $\beta = 0$, and $\theta = 0$;

- Training EventSRNet with low-resolution events upsampled by bilinear interpolation, blurry intensity image, and the preliminary sharp intensity image generated by the first DeblurNet. Here, $\alpha = 0$, $\beta = 0$, and $\theta = 1.0$;
- Training the second DeblurNet with super-resolved events enhanced by EventSRNet and blurry intensity images. Here, $\alpha = 0.1$, $\beta = 1.0$, and $\theta = 1.0$.

3.5. Experimental Settings

The proposed method is implemented on the basis of PyTorch [32] and optimized with AdamW [33] whose momentum, momentum2, and weight decay are 0.9, 0.999, and 1×10^{-4}. The first DeblurNet and EventSRNet are trained with a learning rate of 1×10^{-4} for 200 epochs, and then the learning rate descends to 1×10^{-5} for continually fine-tuning for another 100 epochs. As for the second DeblurNet, since its parameters are initially with the first DeblurNet, its learning rate is set to 10^{-5} and it only needs to be trained for 80 epochs. In addition, our model is trained with the batch size and patch size of the training dataset as 16 and 256 × 256.

3.6. Datasets

We evaluate the proposed model with two datasets: GoPro [20] and RGB-DAVIS [19].

GoPro [20] is a widely used dataset for dynamic scene deblurring. It contains a total of 33 scenes that have been split into a training set (22 scenes) and a test set (11 scenes). As for every scene, it consists of a sequence of sharp images. To adapt it to our task which requires a sequence of low-resolution events, a sequence of high-resolution events, a blurry high-resolution intensity image, a sharp high-resolution intensity image, and a series of operations have been applied to this dataset. Firstly, for attaining finer events and blurring images, we increase the frame-rate of videos from 240 fps to 960 fps by an existing high-performance video frame interpolation method proposed by Niklaus et al. [34]. Secondly, $(m-1) \times 4 + 1$ (m is the number of sharp images before interpolation and $m = 11$ in this paper) sharp images are fused into the blurry intensity image with the methods mentioned in [20]. Thirdly, events between every two interpolated frames are simulated by an event simulator ESIM [35]. The events generated with the original intensity images are considered high-resolution events, while the low-resolution events are synthetic with the intensity images degraded via bilinear interpolation. As such, we finally 2040 and 1101 samples in the training set and test set, respectively.

RGB-DAVIS [19] is a real event dataset. It contains low-spatial-resolution events captured by event cameras and high-spatial-resolution intensity images attained by conventional frame-based camera. We used the same pipeline (including the video frame interpolation, sharp image fusion, and ESIM simulation) described above to acquire the additional high spatial resolution events and blurry intensity images. After simulating, we split this dataset into a training set and a test set based on scenes. The training set contains 918 samples while the test set owns 230 samples. More details are shown in Appendix C.

4. Results and Analysis

4.1. Comparison with State-of-the-Art Methods

To validate the effectiveness of our proposed method aiming to perform image deblurring aided by low-resolution events, we conducted experiments on the GoPro dataset and the RGB-DAVIS dataset and compared them with state-of-the-art works about image deblurring, including EDI [24] and SRN [21]. As for the metric, we used PSRN and SSIM [36] for the quantitative comparison of intensity images.

Note that EDI [24] only works for the events that have the same spatial resolution with the intensity image. Therefore, the low-resolution events will be upsampled into the spatial resolution of the intensity image by bilinear interpolation before sending them into this model. Since this work aimed to perform video deblurring and interpolation, for adapting the events to image deblurring, we choose the center frame as its deblurring output. As for SRN [21], it was originally designed for image deblurring without events and the synthetic blurry images here are a little bit different from those released in [20] (the synthetic blurry images here are harder than before for deblurring). For a fair comparison, we fine-tuned SRN with our synthetic data to obtain a comparable result. In addition, for demonstrating the basic ability of low-resolution events in the image deblurring task, we conducted an experiment that simultaneously sends the combination of upsampled events and blurry intensity images into SRN, and named it as SRN$^+$. Since SRN is a relative heavy model that aims to process three scale inputs, it is time-consuming. In order to save time, we adopted the physical model-based methods proposed by [14] for DeblurNet. It is approximately 4-fold smaller than SRN, and it can achieve a comparable or even better performance for event-based image deblurring. Here, we name the results generated in the first DeblurNet as Ours_D, while we name the results generated in the second DeblurNet as Ours_DED.

The results of these methods on the GoPro dataset and RGB-DAVIS dataset are shown in Tables 1 and 2, respectively. Figures 4 and 5 are the qualitative visualizations of some examples in their test set. Each dataset contains two event settings. ×2 means that the spatial resolution of the intensity image is two times that of its events and ×4 means that the spatial resolution of the intensity image is four times that of its events. The proposed method performs favorably against the existing methods on both settings. Since EDI is optimized with a traditional algorithm, it cannot borrow useful information from sharp intensity images or high-resolution events which are considered the supervisor in deep-learning-based methods, regardless of the fact that its performance in PSRN or SSIM is relatively low. The visualized results tend to be artifacts, especially when the resolution of events is four times lower than the intensity images. We then observe SRN and SRN$^+$. The performance of SRN$^+$ is greater than that of SRN by more than 1 dB in PSNR for both the GoPro dataset and RGB-DAVIS dataset, which proves the effectiveness of the low-resolution events for image deblurring. Ours_D, the results of our first DeblurNet, despite being implemented with a smaller model, achieves a comparable or even better performance than that of the deblurring image. Furthermore, by adding additional EventSRNet and DeblurNet, Ours_DED attains even further performance enhancement. The visualized results also appear more smooth and clear. We made a comparison between the different event scale settings. It is obvious that a smaller resolution gap between the intensity image and events means a higher image deblurring performance. This phenomenon reflects the possibility of the performance promotion of image deblurring while enhancing the quality of events. This also proves the effectiveness of our EventSRNet.

Table 1. Image deblurring results of GoPro in terms of average PSNR and SSIM. SRN and SRN$^+$ represent the results of image deblurring without and with low-resolution events, respectively, while Ours_D and Ours_DED represent the results of our method without and with EventSRNet, respectively. This shows that the proposed method outperforms the methods from previous works and image deblurring can benefit from low-spatial resolution events and event-enhancing operations.

Method	EDI [15]	SRN [21]	SRN$^+$	Ours_D	Ours_DED
			2×		
PSRN	22.14	30.99	34.13	34.26	35.73
SSIM	0.8029	0.9446	0.9651	0.9630	0.9742
			4×		
PSRN	21.61	30.99	32.49	32.58	33.92
SSIM	0.7997	0.9446	0.9568	0.9570	0.9651

4.2. Ablation Study

4.2.1. Effectiveness of EventSRNet

EventSRNet takes low-resolution events, blurry image, and sharp image as inputs. It not only tends to reconstruct high spatial resolution events from temporal information in low-spatial-resolution events, but also from the structure information in the blurry image and sharp image based on the assumption that local events are triggered with edges. Therefore, we conduct experiments with EventSRNet's inputs in three settings: only low-resolution events, low-resolution events and blurry image, and low-resolution events, blurry image, and sharp images. The metric used for evaluating the quality of events is MSE (mean squared error), which is better the smaller it is. The corresponding results are shown in Table 3 and Figure 6. We can see that both the blurry image and sharp image are useful for the enhancement of events. However, with only blurry images, the generated events also tend to be blurry, while the sharp image can guide the events to be more clear.

4.2.2. Effectiveness of Asymmetric L1 Loss

Since the data structure of events is sparse, we adopted asymmetric L1 loss for the optimization of EventSRNet. To evaluate the usefulness of this setting, we replace the asymmetric L1 loss with normal L1 loss. Table 3 and Figure 6 are the results of event enhancement learning with L1 loss (L1) and asymmetric L1 loss (AL1), in terms of MSE. With low-resolution events, blurry image, and sharp image as inputs, the enhanced events learning by L1 are clear but lose a lot of information that is not near the edges in the intensity image. Asymmetric L1 loss can learn this sparse information well.

Table 2. Image deblurring results of RGB_DAVIS in terms of average PSNR and SSIM. SRN and SRN$^+$ represent the results of image deblurring without and with low-resolution events, respectively, while Ours_D and Ours_DED represent the results of our method without and with EventSRNet, respectively. This shows that the proposed method outperforms the methods from previous works and image deblurring can benefit from the -spatial resolution events and event-enhancing operations.

Method	EDI [15]	SRN [21]	SRN$^+$	Ours_D	Ours_DED
			2×		
PSRN	20.16	26.96	27.60	27.73	27.98
SSIM	0.7636	0.8792	0.8901	0.8927	0.9033
			4×		
PSRN	19.79	26.51	27.27	27.33	27.65
SSIM	0.7582	0.8756	0.8851	0.8874	0.8907

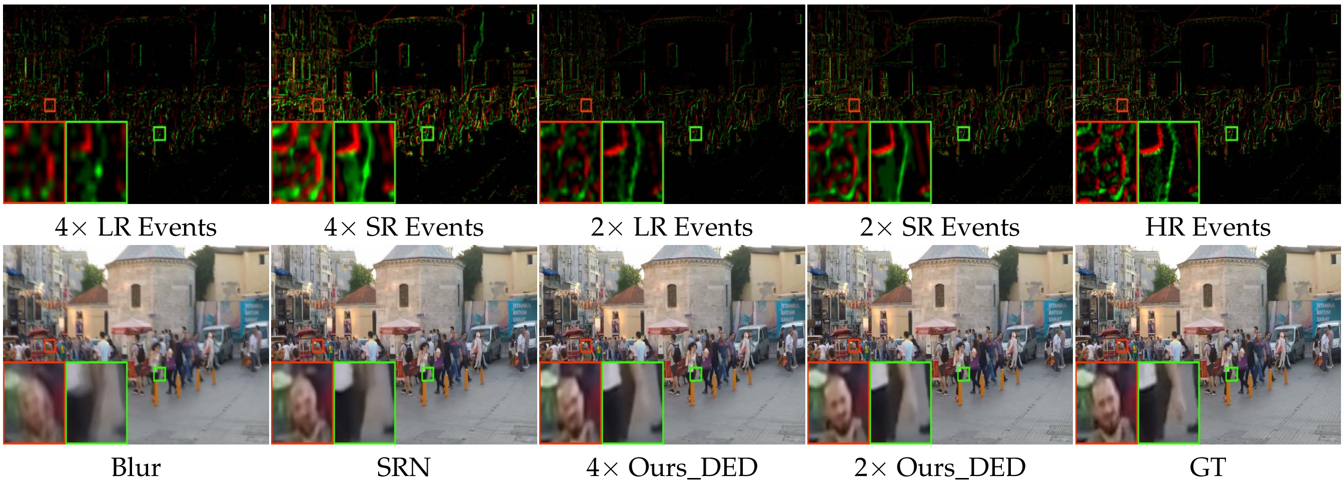


Figure 4. Visual comparisons of image deblurring without events with 2× and 4× events. Obviously, the results of image deblurring with events (2× Ours_DED and 4× Ours_DED) are better than the result without events (SRN [21]), and the 2× spatial resolution events are better than the 4× spatial resolution events.

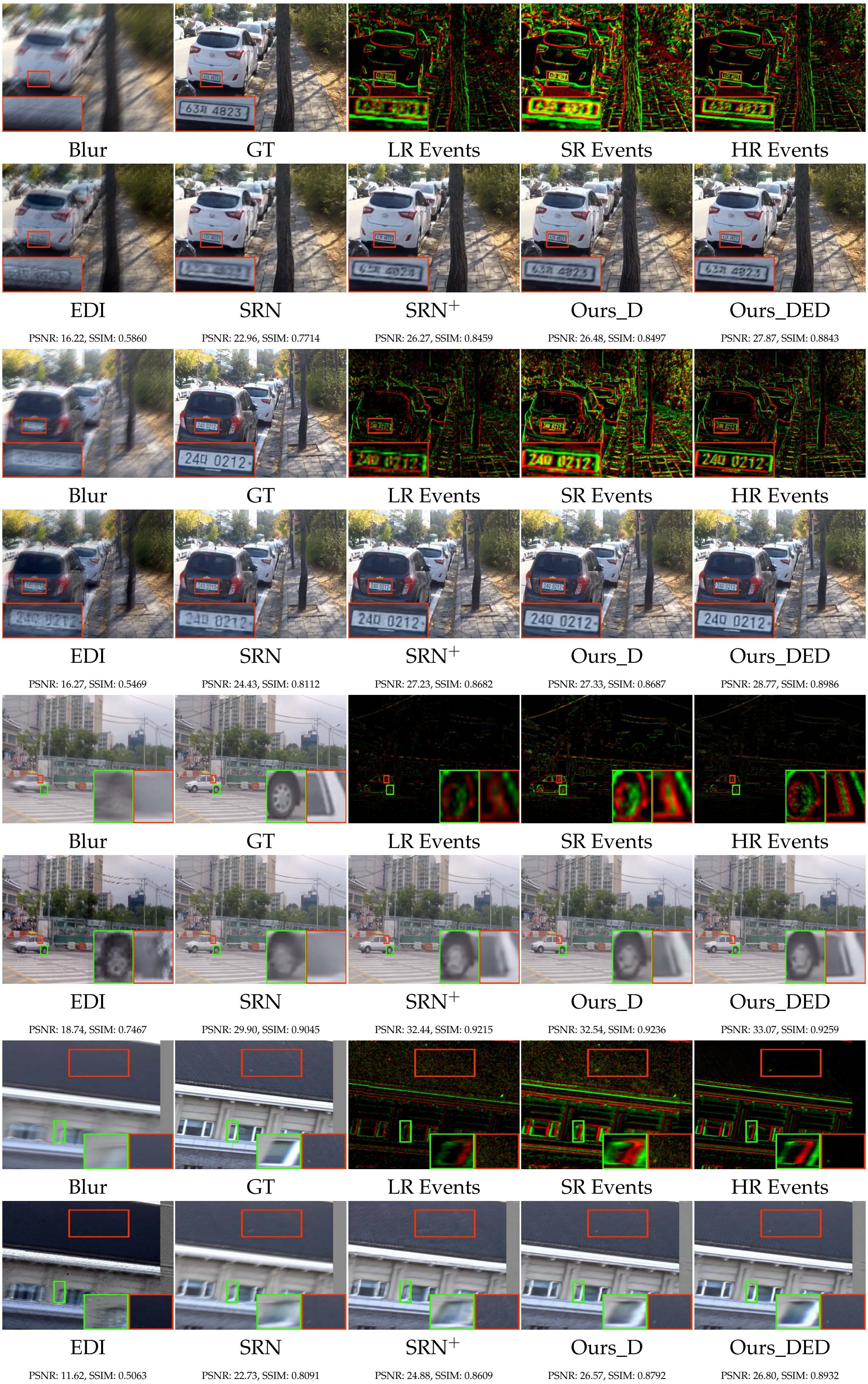

Blur
GT
LR Events
SR Events
HR Events
EDI
SRN
SRN$^+$
Ours_D
Ours_DED
PSNR: 16.22, SSIM: 0.5860
PSNR: 22.96, SSIM: 0.7714
PSNR: 26.27, SSIM: 0.8459
PSNR: 26.48, SSIM: 0.8497
PSNR: 27.87, SSIM: 0.8843
Blur
GT
LR Events
SR Events
HR Events
EDI
SRN
SRN$^+$
Ours_D
Ours_DED
PSNR: 16.27, SSIM: 0.5469
PSNR: 24.43, SSIM: 0.8112
PSNR: 27.23, SSIM: 0.8682
PSNR: 27.33, SSIM: 0.8687
PSNR: 28.77, SSIM: 0.8986
Blur
GT
LR Events
SR Events
HR Events
EDI
SRN
SRN$^+$
Ours_D
Ours_DED
PSNR: 18.74, SSIM: 0.7467
PSNR: 29.90, SSIM: 0.9045
PSNR: 32.44, SSIM: 0.9215
PSNR: 32.54, SSIM: 0.9236
PSNR: 33.07, SSIM: 0.9259
Blur
GT
LR Events
SR Events
HR Events
EDI
SRN
SRN$^+$
Ours_D
Ours_DED
PSNR: 11.62, SSIM: 0.5063
PSNR: 22.73, SSIM: 0.8091
PSNR: 24.88, SSIM: 0.8609
PSNR: 26.57, SSIM: 0.8792
PSNR: 26.80, SSIM: 0.8932

Figure 5. Visual comparisons of image deblurring on ×4 spatial resolution. Obviously, image

deburring takes benefits from the events, although they are in low spatial resolution (SRN and SRN^{+}). Furthermore, the enhancement of the event proposed in this paper can further improve the quality of image deblurring (Ours_D and Ours_DED).

Table 3. The results of EventSRNet with different input settings and learning losses in terms of MSE. These are tested on GoPro with the ×4 spatial resolution gap between intensity the image and events. The guides from the sharp images and blurry images and the asymmetric L1 loss benefit the enhancement of events.

Events	Blurry Image	Sharp Image	L1	AL1	MSE
√	-	-	-	√	0.01964
√	√	-	-	√	0.01065
√	√	√	-	√	0.00706
√	√	√	√	-	0.01478

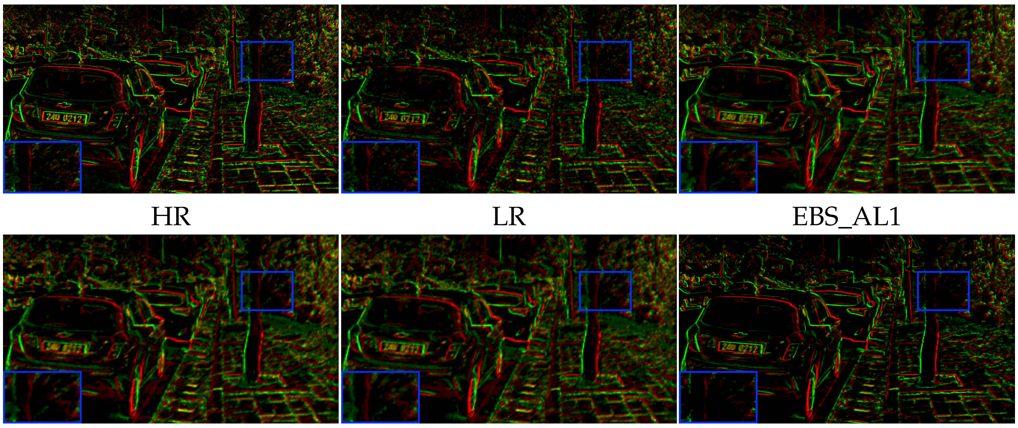


Figure 6. Visualization of EventSRNet learning with different settings. HR and LR are events captured under ×1 and ×4 spatial resolution. 'E', 'B', and 'S' here mean that the inputs of EventSRNet contain events, blurry images, and sharp image, respectively. 'AL1' and 'L1' mean learning EventSRNet with asymmetric L1 loss and normal L1 loss. The proposed method with all EBS and AL1 tends to obtain a clearer and more structured result. Without sharp images, BE_AL1 and E_AL1 appear blurry. Furthermore, the result attained with L1 loss tends to lose information.

5. Conclusions

In contrast to previous single image deblurring works or low-resolution-event-based image deblurring works, this work takes both the advantages and disadvantages of the conventional frame-based camera and event camera into consideration and proposes to deblur images with the help of low-resolution events. Although both events and intensity images are degraded, the low spatial resolution events are still rich in temporal information that is useful for image deblurring, and the intensity images have potential use in the reconstruction of events. Therefore, this paper proposes an alternately performed model for enhancing the quality of intensity images and events by exploiting the complementarity between them. By firstly deblurring the image with state-of-the-art image deblurring methods, we can obtain a reasonably sharp image for providing rich structure information for the enhancement of events. The temporal information in the processed events will further promote the quality of the image. Extensive experiments show the effectiveness of the proposed method.

This work can be extended in several directions. The enhanced events can be further used for many downstream tasks, such as object recognition, object detection, and segmentation. Additionally, it can also be used for the frame interpolation of high-resolution video. We will thus consider further improving the quality of the events to make a contribution to downstream tasks.

Author Contributions: Conceptualization, Z.W., J.R., J.Z. and P.L.; methodology, Z.W., J.R., J.Z. and P.L.; software, Z.W.; validation, Z.W. and J.Z.; formal analysis, Z.W., J.R., J.Z. and P.L.; investigation, Z.W., J.R., J.Z. and P.L.; resources, J.R. and P.L.; data curation, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, J.R., J.Z. and P.L.; visualization, Z.W.; supervision, P.L.; project administration, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Details of the Structures of DeblurNet and EventSRNet

Appendix A.1. DeblurNet

The framework of DeblurNet is shown in Figure A1. The methods proposed in [14,15] have proven that a sharp image can be achieved by combining a blurry image and the double integral of events. Therefore, similarly to [14], DeblurNet tends to extract the features of events $\mathbf{E}$ with a skipped Unet [30] (the network in the bottom of Figure A1) and these features work as the double integral of events which will be further combined with a blurry image $\mathbf{I_b}$ for achieving the sharp image $\mathbf{I_s}$. Since events are triggered by the spatially variant threshold, another branch (the network at the top of Figure A1) is designed for generating a dynamic filter (**DF**). **DF** contains different filters for each position in feature maps and by applying these filters to the feature of events, the final feature of events tends to be more robust to the variant threshold. In this paper, the filter size K of the dynamic filter is 5.

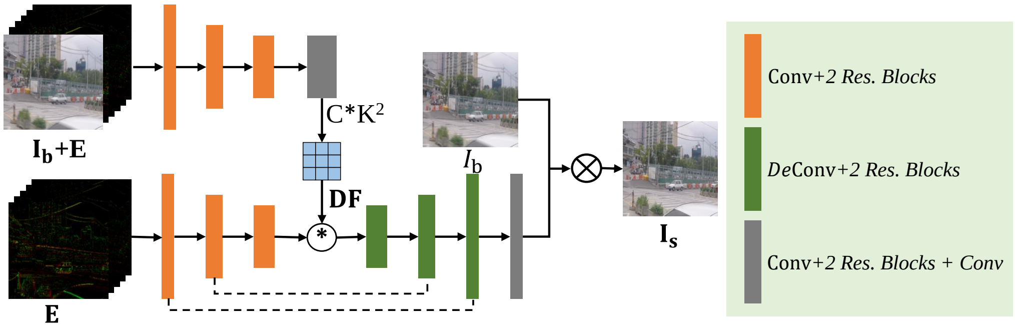


Figure A1. Framework of DeblurNet.

Appendix A.2. EventSRNet

Since Unet [30] showed its advantages on image super-resolution, we also adopted a lightweight Unet for events' super-resolution. As shown in Figure A2, it takes the combination of a sharp image $\mathbf{I_s}$, a blurry image $\mathbf{I_b}$, and low-resolution events $\mathbf{E}$ as input, and which yields as output a sequence of high-resolution events. The structure of EventSRNet is similar to the events' feature extractor in DeblurNet except for the branch about the dynamic filter. It contains two downsampling steps and two upsampling steps, which are implemented with convolution operation and deconvolution operation, respectively. The skips between layers are implemented with concatenation.
For performance improvement, the Unet [30] used in our EventSRNet can be replaced with other more powerful adapted networks, such as EDSR [37] and U-Net++ [38], though they may result in additional power consumption. Once we replace the Unet with U-Net++, the MSE of event ×4 super-resolution on GoPro drops from 0.00706 to 0.00627.

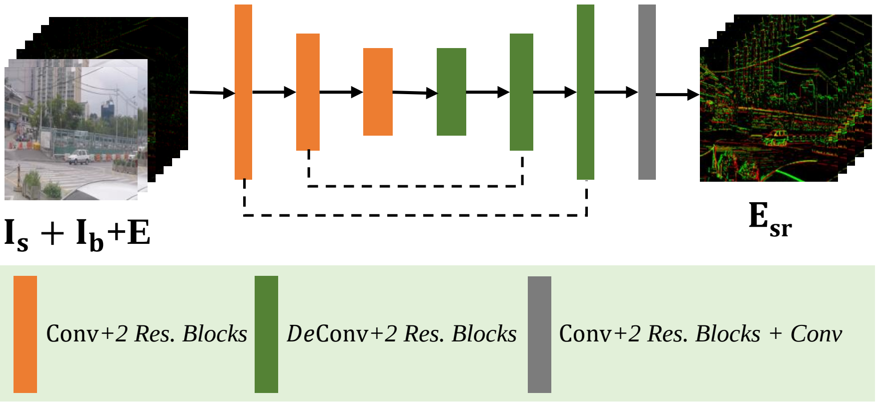


Figure A2. Framework of EventSRNet.

Appendix B. Details of RGB-DAVIS Dataset

RGB-DAVIS [19] is a real-world event dataset. It includes 10 indoor scenes (named Indoor1∼10) and 10 outdoor scenes (Outdoor1∼10)—a total of 20 scenes. Samples in each scene contain a sequence of low-spatial-resolution events (180 × 190) and a high-spatial-resolution sharp image (1440 × 1520). It was collected by Wang et al. [19] and originally used for denoising events and achieving the events' super-resolution with traditional algorithm. In order to adapt it to our CNN-based model, we needed to split it into a training set and test set, and synthetic high-resolution events and high-resolution blurry images required by our method.

In this paper, we randomly selected 3 indoor scenes and 2 outdoor scenes out of these 20 scenes for the test set, and the remaining 15 scenes are used for training. The selected scenes were Indoor2, Indoor4, Indoor7, Outdoor3, and Outdoor5.

Similarly to the GoPro [20] dataset, the synthetic procedure of RGB-DAVIS includes video frame interpolation, sharp image fusion, and ESIM simulation, which were described in the script. In contrast to GoPro, which needs to simulate low-resolution events, RGB-DAVIS has real-world low-resolution events and what it needs to do is simulate the corresponding high-resolution events and blurry images. The high-resolution blurry images are synthetic by fusing $(m-1) \times 4 + 1$ (m is the number of sharp images before interpolation) sharp frames downsampled from the original sharp images (the resolution of the original sharp image is 8 times that of the resolution of the low-resolution events). Since the time between two frames in RGB-DAVIS is longer than that in GoPro, m is set to 5 in RGB-DAVIS rather than 11.

Appendix C. Evaluation of Generalization

For evaluating the generalization of the proposed method, we tested our model trained with GoPro [20] on new data from Need for Speed [39] following the synthesized process described in Section 3.6. The result is shown in Figure A3. It can also attain a reasonable deblurring result.

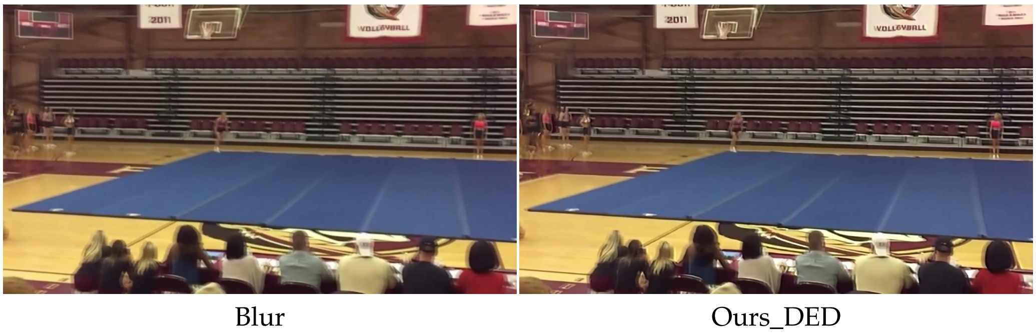


Figure A3. Results of deblurring on Need for Speed [39] with the model trained with GoPro [20].

References

1. Finateu, T.; Niwa, A.; Matolin, D.; Tsuchimoto, K.; Mascheroni, A.; Reynaud, E.; Mostafalu, P.; Brady, F.; Chotard, L.; LeGoff, F.; et al. 5.10 A 1280 × 720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86 µm Pixels, 1.066 GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline. *IEEE Conf. Proc.* **2020**, *2020*, 112–114.
2. Patrick, L.; Posch, C.; Delbruck, T. A 128 × 128 120 dB 15 µs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE J. Solid-State Circuits* **2008**, *43*, 566–576.
3. Posch, C.; Serrano-Gotarredona, T.; Linares-Barranco, B.; Delbruck, T. Retinomorphic event-based vision sensors: Bioinspired cameras with spiking output. *Proc. IEEE* **2014**, *102*, 1470–1484. [CrossRef]
4. Son, B.; Suh, Y.; Kim, S.; Jung, H.; Kim, J.S.; Shin, C.; Park, K.; Lee, K.; Park, J.; Woo, J.; et al. 4.1 A 640 × 480 dynamic vision sensor with a 9 µm pixel and 300 meps address-event representation. In Proceedings of the 2017 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 5–9 February 2017.
5. Brandli, C.; Muller, L.; Delbruck, T. Real-time, high-speed video decompression using a frame-and event-based DAVIS sensor. In Proceedings of the 2014 IEEE International Symposium on Circuits and Systems (ISCAS), Melbourne, Australia, 1–5 June 2014.
6. Bardow, P.; Davison, A.J.; Leutenegger, S. Simultaneous optical flow and intensity estimation from an event camera. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
7. Munda, G.; Reinbacher, C.; Pock, T. Real-time intensity-image reconstruction for event cameras using manifold regularisation. *Int. J. Comput. Vis.* **2018**, *126*, 1381–1393. [CrossRef]
8. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. Events-to-video: Bringing modern computer vision to event cameras. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
9. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. High speed and high dynamic range video with an event camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1964–1980. [CrossRef] [PubMed]
10. Scheerlinck, C.; Rebecq, H.; Gehrig, D.; Barnes, N.; Mahony, R.; Scaramuzza, D. Fast image reconstruction with an event camera. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 1–5 March 2020.
11. Scheerlinck, C.; Barnes, N.; Mahony, R. Continuous-time intensity estimation using event cameras. In Proceedings of the ACCV 2018, Perth, Australia, 2–6 December 2018.
12. Wang, L.; Mohammad Mostafavi, I.S.; Ho, Y.S.; Yoon, K.J. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
13. Jiang, Z.; Zhang, Y.; Zou, D.; Ren, J.; Lv, J.; Liu, Y. Learning event-based motion deblurring. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3320–3329.
14. Lin, S.; Zhang, J.; Pan, J.; Jiang, Z.; Zou, D.; Wang, Y.; Chen, J.; Ren, J. Learning Event-Driven Video Deblurring and Interpolation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Volume 3.
15. 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 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
16. Mitrokhin, A.; Fermüller, C.; Parameshwara, C.; Aloimonos, Y. Event-based moving object detection and tracking. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018.
17. Perot, E.; de Tournemire, P.; Nitti, D.; Masci, J.; Sironi, A. Learning to Detect Objects with a 1 Megapixel Event Camera. *arXiv* **2020**, arXiv:2009.13436.
18. Vidal, A.R.; Rebecq, H.; Horstschaefer, T.; Scaramuzza, D. Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. *Robot. Autom. Lett.* **2018**, *3*, 994–1001. [CrossRef]
19. Wang, Z.W.; Duan, P.; Cossairt, O.; Katsaggelos, A.; Huang, T.; Shi, B. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
20. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
21. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-recurrent network for deep image deblurring. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
22. Khodamoradi, A.; Kastner, R. O (N)-space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Trans. Emerg. Top. Comput.* **2018**, *9*, 15–23. [CrossRef]
23. Liu, H.; Brandli, C.; Li, C.; Liu, S.C.; Delbruck, T. Design of a spatiotemporal correlation filter for event-based sensors. In Proceedings of the 2015 IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, Portugal, 24–27 May 2015.
24. Padala, V.; Basu, A.; Orchard, G. A noise filtering algorithm for event-based asynchronous change detection image sensors on truenorth and its implementation on truenorth. *Front. Neurosci.* **2018**, *12*, 118. [CrossRef] [PubMed]

25. Wang, Y.; Du, B.; Shen, Y.; Wu, K.; Zhao, G.; Sun, J.; Wen, H. EV-gait: Event-based robust gait recognition using dynamic vision sensors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6358–6367.
26. Wang, L.; Kim, T.K.; Yoon, K.J. EventSR: From Asynchronous Events to Image Reconstruction, Restoration, and Super-Resolution via End-to-End Adversarial Learning. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),Seattle, WA, USA, 13–19 June 2020.
27. Mostafavi, M.; Choi, J.; Yoon, K.J. Learning to Super Resolve Intensity Images from Events. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
28. Wang, B.; He, J.; Yu, L.; Xia, G.S.; Yang, W. Event Enhanced High-Quality Image Recovery. *arXiv* **2020**, arXiv:2007.08336.
29. Zhang, L.; Zhang, H.; Chen, J.; Wang, L. Hybrid deblur net: Deep non-uniform deblurring with event camera. *IEEE Access* **2020**, *8*, 148075–148083. [CrossRef]
30. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015.
31. Jia, X.; De Brabandere, B.; Tuytelaars, T.; Van Gool, L. Dynamic filter networks. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016.
32. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in pytorch. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
33. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
34. Niklaus, S.; Mai, L.; Liu, F. Video frame interpolation via adaptive separable convolution. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
35. Rebecq, H.; Gehrig, D.; Scaramuzza, D. ESIM: An open event camera simulator. In Proceedings of the 2nd Conference on Robot Learning, Zürich, Switzerland, 29–31 October 2018.
36. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003.
37. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017.
38. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef] [PubMed]
39. Kiani Galoogahi, H.; Fagg, A.; Huang, C.; Ramanan, D.; Lucey, S. Need for speed: A benchmark for higher frame rate object tracking. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.