# Event-based Blurry Frame Interpolation under Blind Exposure

Wenming Weng Yueyi Zhang* Zhiwei Xiong
University of Science and Technology of China, Hefei, China
wmweng@mail.ustc.edu.cn, {zhyuey, zwxiong}@ustc.edu.cn
https://github.com/WarranWeng/EBFI-BE

## Abstract

*Restoring sharp high frame-rate videos from low frame-rate blurry videos is a challenging problem. Existing blurry frame interpolation methods assume a predefined and known exposure time, which suffer from severe performance drop when applied to videos captured in the wild. In this paper, we study the problem of blurry frame interpolation under **blind exposure** with the assistance of an event camera. The high temporal resolution of the event camera is beneficial to obtain the exposure prior that is lost during the imaging process. Besides, sharp frames can be restored using event streams and blurry frames relying on the mutual constraint among them. Therefore, we first propose an exposure estimation strategy guided by event streams to estimate the lost exposure prior, transforming the blind exposure problem well-posed. Second, we propose to model the mutual constraint with a temporal-exposure control strategy through iterative residual learning. Our blurry frame interpolation method achieves a distinct performance boost over existing methods on both synthetic and self-collected real-world datasets under blind exposure.*

## 1. Introduction

Blurry frame interpolation (BFI) [23, 46, 73] aims to restore sharp high frame-rate videos from blurry low frame-rate videos, which is highly desirable for a wide range of applications, such as novel view interpolation synthesis [7], frame rate conversion [32], slow motion [21] and inter-frame video compression [64]. Compared with the cascade scheme, *i.e.*, combining frame deblurring [20, 25, 33, 39, 40, 51, 53, 59] with frame interpolation [2, 21, 31, 32, 35, 37, 38, 68], the joint method [46, 73] is more effective, which overcomes the error accumulation problem and ambiguity of the temporal scope [46].

Despite remarkable improvement, prior works [23, 46] assume that the exposure time is predefined or known as the shutter period, which we call **non-blind exposure** as shown in the left part of Fig. 1 (a). However, the complicated video capturing in the wild gives rise to variable and unknown exposure time, which we call **blind exposure**. The right part of Fig. 1 (a) presents that the shutter period is the summation of exposure time and data readout time. For better imaging, the exposure time is variable to fit the changing light conditions in the real imaging environments, especially when the auto-exposure function turns on [73].

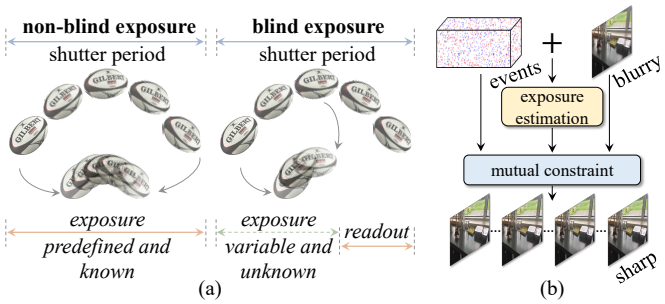



Figure 1. (a) Non-blind exposure setting and blind exposure setting. (b) Illustrative flow of our solution to blurry frame interpolation under blind exposure.

**Challenges and Motivation.** In this paper, we focus on the problem of BFI under blind exposure, which is practical yet has barely been investigated explicitly. The first challenge of this problem can be attributed to the intrinsic imaging mechanism of frame-based cameras. As can be seen from Fig. 1 (a), the accumulation operation of frame-based cameras inevitably results in the loss of motion information during the exposure time. Particularly, the variable exposure time further leads to the temporal jittering, which degrades the performance of the video enhancement algorithms with constant/predefined exposure time assumption. Another challenge is that there misses a decent model as the effective guidance. Conventional frame-based methods for BFI are vulnerable to the artifacts introduced by flow-based warping or straight forward prediction.

To overcome the above challenges, we attempt to provide a new perspective, by resorting to a novel sensor.

*Corresponding author. This work was supported in part by the National Natural Science Foundation of China under Grants 62131003 and 62021001.

Event cameras [8, 69], also known as neuromorphic sensors, are bio-inspired visual sensors that output events by detecting spatiotemporal brightness changes. We demonstrate that event cameras are suitable for BFI under blind exposure from two points as shown in Fig. 1 (b). First, the high temporal resolution property of event cameras is able to compensate for the exposure prior that is lost during the imaging process of frame-based cameras. To this end, we propose an exposure estimation strategy guided by event streams to estimate the lost exposure prior. In such a way, the blind exposure problem can be made well-posed, which eases the difficulty of video restoration. Second, the event stream is a natural constraint between blurry frames and sharp frames, providing a physical model for video restoration. To effectively model this physical correlation, we propose a temporal-exposure control strategy that takes timestamps and exposure priors as inputs for interpolation through iterative residual learning. By exploiting the proposed strategies of exposure estimation and time-exposure control, we are able to perform arbitrary-time interpolation from blurry frames under blind exposure. To evaluate our method in real-world scenarios, we collect a real blurry video dataset using a DAVIS-346 color event camera, which includes multiple exposure assumptions. Experimental results on synthetic and self-collected real datasets demonstrate our superior performance over existing methods.

The contributions of this paper can be summarized as follows: **1**) we provide a decent solution for BFI under blind exposure by using event cameras, for the first time; **2**) we propose an exposure estimation strategy guided by event streams, which makes the blind exposure problem well-posed; **3**) we propose a temporal-exposure control strategy, which enables BFI at an arbitrary timestamp under blind exposure; **4**) we achieve superior performance over existing state-of-the-art methods on both synthetic and self-collected real-world datasets.

## 2. Related Work

**Non-blind Exposure Based Video Restoration.** Temporal information (*e.g.*, frame rate, exposure time) and spatial visual effects (*e.g.*, blurry or sharp) measure the overall quality of video data. For acquiring high quality and high frame-rate video, early works exploited hand-crafted information such as optical flow [19, 44] and dark channel priors [40] for interpolation and motion deblurring. Recently, deep learning has dominated the video restoration, presenting flow-based techniques [26, 29, 36, 68], kernel-based techniques [35, 37, 38] and phased-based techniques [31, 32] for interpolating, and recurrent or spatio-temporal filter adaptive architecture [20, 34, 53, 74, 75] for deblurring. More recently, event cameras have shown potential for video restoration. These works [13, 18, 22, 27, 45, 50, 52, 54–56, 62, 63, 66, 67, 70, 76] demonstrated impressive results by designing an effective aggregation module of event streams and frames. However, previous methods usually ignore the exposure time setting using the predefined and known exposure time instead, which leads to a severe performance drop when applied to real-world scenarios.

**Blind Exposure Based Video Restoration.** In general, blind exposure is a real-world setting that describes the complex light conditions in the wild. Intuitively, blind exposure challenges video restoration in threefold: 1) sequence-to-sequence (or scene-to-scene) exposure variance, which means that different video sequences or scenes have diverse exposure time for adapting to the light-specific photography; 2) frame-to-frame exposure variance, which is a more general case especially when the auto-exposure function of cameras turns on; 3) for most scenes, it is difficult to obtain the accurate exposure time due to the existence of data readout time and the open of auto-exposure function, which means that the unknown exposure is a common case in real-world capturing. To the best of our knowledge, there are only a few works on this problem. Zhang *et al*. [73] proposed a state estimation network to estimate the start and end key-state frames of a blurry frame. Kim *et al*. [52] attempted to correlate frames and event streams for selecting the events within the exposure time. Despite the corresponding merits, these methods suffer from respective compromises: 1) difficulty in estimating the accurate key-state frames solely from frames due to exposure ambiguity that will be discussed in Sec. 4.1, 2) loss of a decent physical model as the effective guidance. In this work, we realize an event-based unified model in an effective and efficient way, leveraging the high temporal resolution property and mutual constraint provided by event cameras.

## 3. BFI under Non-blind Exposure

Starting with an easy case, we first present the solution of non-blind exposure problem, where the accurate exposure time is provided. We specifically study the event-based BFI. As shown in Fig. 1, the shutter period $\mathcal{T}$ consists of exposure time $\mathcal{T}_{ex} = [t_s, t_e]$ ($t_s, t_e$ denote the start and end timestamp of exposure time) and data readout time $\mathcal{T}_{re}$. According to the EDI model [41], given blurry frame $B$ averaged within the exposure time and corresponding event stream $\mathcal{E}_{[t_s,t_e]}$ triggered during the exposure time, we can obtain the sharp latent frame $L(f)$ at timestamp $f \in \mathcal{T}_{ex}$

$$L(f) = \frac{B}{E(f, \mathcal{T}_{ex})}, \tag{1}$$

where $E(\cdot)$ is the event double integral

$$E(f, \mathcal{T}_{ex}) = \frac{1}{t_e - t_s} \int_{t_s}^{t_e} \exp\left(c \int_f^t e(s) ds\right) dt, \tag{2}$$

in which $e(t) \triangleq p \cdot \delta(t - \tau)$ represents the continuous representation of events with the Dirac function $\delta(\cdot)$, and $c$ denotes the contrast threshold. Obviously, we need to compute the event double integral for interpolation. However,

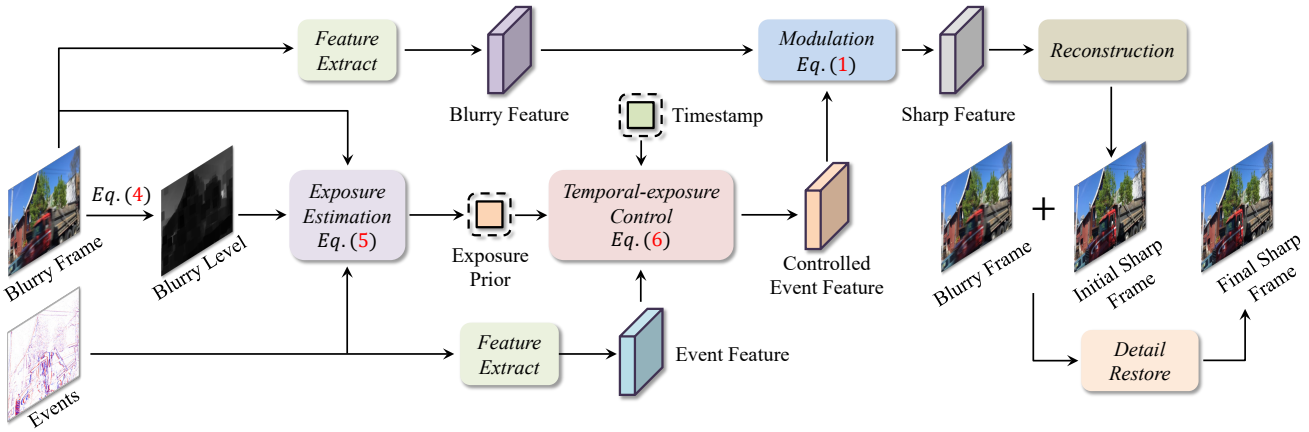



Figure 2. The proposed pipeline for BFI under blind exposure. The blurry frame and events are first used to estimate the lost exposure prior, which is then combined with timestamp for temporal-exposure aware controlling. Eventually, the sharp frame is reconstructed by event-based modulation followed by a reconstruction decoder and a detail restoration network. Note that if we lower the step of timestamp, we can obtain **extreme interpolation videos** (up to 640 FPS from 5 FPS), which is exhibited in the supplement video.

Eq. (2) only holds when timestamp $f$ is within the exposure time. Sometimes the latent sharp frames within the data readout time are also desirable for the fluent motion. To overcome this drawback, EVDI [72] is proposed and revises EDI to achieve arbitrary-time interpolation. Specifically, the event double integral $E(f, \mathcal{T}_{ex})$ at arbitrary timestamp $f$ is reformulated as

$$E(f, \mathcal{T}_{ex}) \approx w_1 \, \mathrm{LDI}\left(\mathcal{P}\left(\mathcal{E}_{[f,t_s]}\right)\right) + w_2 \, \mathrm{LDI}\left(\mathcal{P}\left(\mathcal{E}_{[f,t_e]}\right)\right), \tag{3}$$

where $\mathrm{LDI}(\cdot)$ is a learnable network approximating $E(0, [0, t_e - t_s])$, $w_1 = (f - t_s)/(t_e - t_s), w_2 = (t_e - f)/(t_e - t_s)$ are weights, and $\mathcal{P}(\cdot)$ defines the preprocessing event operator for time shift/flip and polarity reversal [72]. Based on Eq. (1) and Eq. (3), EVDI [72] demonstrates that the sharp frame at an arbitrary timestamp can be restored from a blurry frame and corresponding event streams.

So far, it is possible to interpolate sharp frames from blurry frames under non-blind exposure using Eq. (1) and Eq. (3). However, a key factor to guarantee the excellent performance of EVDI is that the exposure time $\mathcal{T}_{ex}$ must be known exactly. Otherwise, the poor results will be generated, which can be validated in Sec. 5. In short, EVDI can not address the blind exposure problem.

# 4. BFI under Blind Exposure

Recalling the solution to the easy case of non-blind exposure, it suggests that accurate exposure time has a dominant influence on BFI. However, for the blind exposure problem, it is difficult to obtain the accurate exposure time. To make this problem well-posed, a possible solution for remedy is to estimate the lost exposure prior, which is the main idea of this paper. For blind exposure based BFI, we show that event cameras play an important role in estimating the lost exposure prior and modeling the physical correlation of Eq. (3). Fig. 2 illustrates the overall pipeline, which consists of three main components: 1) exposure estimation for estimating the lost exposure prior, 2) temporal-exposure control that takes timestamp and exposure prior as inputs for temporal-exposure aware controlling, and 3) sharp frame reconstruction by event-based modulation followed by a reconstruction decoder and a detail restoration network.

## 4.1. Exposure Estimation Guided by Event Streams

Considering the unavailability of exposure information in the blind exposure problem, we choose to explicitly estimate it. Analogously, blind SR [3,11,47,49,57,71] faces the challenge of unknown degradation kernels in the real-world scenarios. To tackle this, given the input low resolution image, the kernel prior can be approached by explicit estimation [11], internal learning [3,47] or implicit estimation [57] via contrastive learning [4, 6, 12, 16]. Similar to kernel estimation, for estimating the exposure prior, we choose to estimate exposure duty instead of estimating the absolute exposure time. The exposure prior equals $\mathcal{T}_{ex}/(\mathcal{T}_{ex} + \mathcal{T}_{re})$ and ranges from 0 to 1. In practice, it is difficult to compute the exposure prior from the blurry frames directly. Because the blurry effect of an image is determined by the motion speed and duration of exposure time, which we call **exposure ambiguity**. It typically happens that blur due to fast motion within short time equals to blur due to slow motion within long time. This reveals that a blurry frame alone is not enough for estimating exposure prior due to the loss of motion prior. As a remedy, event cameras are capable of recording motion prior in microseconds and faster motion triggers a large number of events, which helps clear exposure ambiguity. In other words, the fact that high (or low) count number of events suggests fast (or slow) motion regularizes the exposure estimation.

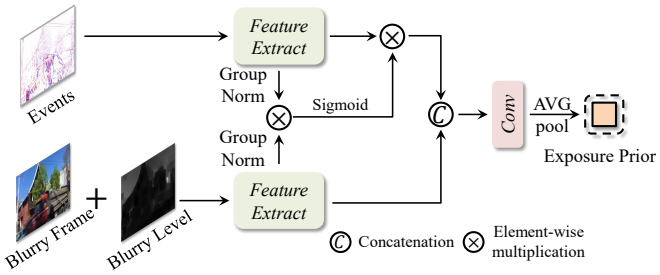



Figure 3. Illustration of exposure estimation component.

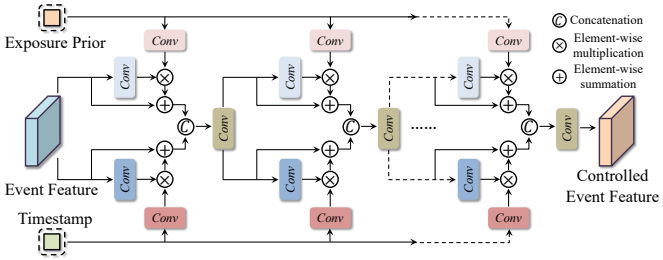



Figure 4. Illustration of temporal-exposure control component.

By exploiting the frame and corresponding events, we are able to bypass the exposure ambiguity for accurate exposure estimation. However, for some challenging scenarios, such as non-uniform motions, the blurry frame will present an unbalanced blurry effect, which will degrade the exposure estimation. In addition, the color information is trivial for the exposure estimation to some extent. In order to further improve the estimation process, we opt to provide an additional blurry prior called the *blurry level* using dark channel [17, 40] or Laplacian [1], which can be considered as a measurement of the degree of blurriness in a frame. The blurry level can be described as

$$BlurryLevel = T(Frame), \tag{4}$$

where $T(\cdot)$ denotes the dark channel function [17] or the Laplacian function [1].

Fig. 3 illustrates the proposed exposure estimation module. We first concatenate the blurry frame and the blurry level, which are then fed into a frame feature extractor to output a frame feature. The event stream is accumulated as an event count map ($ECM$) [10] with dimensions of $B \times 2 \times H \times W$, where $B$ means the number of time bins that is fixed as 16 for all the experiments, and 2 denotes positive and negative polarities. The $ECM$ is then fed to another feature extractor, generating an event feature. To alleviate the modality diversity of frames and events, the two features are first regularized by group normalization [65], which are then multiplied together and activated by the Sigmoid function to produce the event mask for further rectifying the event feature. Finally, we use a convolutional block followed by average pooling to compress the concatenated frame and event features to acquire the exposure prior. We formulate the exposure estimation as

$$EP = P([Frame, BlurryLevel], ECM), \tag{5}$$

where $P(\cdot)$ denotes the whole operation of exposure estimation in Fig. 3, $[\cdot]$ and $EP$ are the concatenation operator and the exposure prior, respectively.

### 4.2. Temporal-exposure Control

Through exposure estimation, we are able to estimate the lost exposure prior, which can transfer blind exposure well-posed. Now, the remaining question is how to interpolate the sharp frames from blurry frames given arbitrary timestamp and the estimated exposure prior. Fortunately, the Eq. (1) and Eq. (3) provide a physical model for this purpose. However, it is inevitably disturbed by the complicated noise distribution of real-world events due to the event preprocessing operation and weighted summation, even though the learnable network $LDI(\cdot)$ is adopted. In contrast, we propose a temporal-exposure control framework, which is an end-to-end fully learnable network without any handcrafted operation.

For simplicity, we reformulate the event double integral in Eq. (3) as

$$E(t, \mathcal{T}_{ex}) = F(t, EP, EventFeat), \tag{6}$$

where $t$ is the normalized timestamp ranging from 0 to 1, $EP$ is the exposure prior, and $EventFeat$ denotes the event feature. Particularly, for acquiring $EventFeat$, we convert the event stream to an $ECM$ with dimensions of $B \times 2 \times H \times W$, which is further embedded in the event feature by a feature extractor. As can be clearly seen from Eq. (6), the function $F(\cdot)$ is temporal-exposure aware, which is an important property for the BFI under blind exposure.

In specific, we attempt to implement Eq. (6) by regarding timestamp and exposure prior as two controlling coefficients. Recently, a few works [14, 15, 48, 58, 60] have presented learnable controlling frameworks for image restoration and image synthesis. They demonstrated that the unbalanced learning in a multiple degradation problem can be modulated in an interactive manner. Following this way, the restoration results become controllable. Inspired by these works, we embrace the success of the control framework and design our temporal-exposure control strategy, which is illustrated in Fig. 4. Generally, timestamps and exposure priors are fundamentally different. In contrast to the single path framework [15] that combines the multiple controlling coefficients as a whole block, we specifically design a framework consisting of dual paths, each of which is responsible for controlling one coefficient. The learning process is unfolded in multiple steps, which enables an iterative residual fashion. At the end of each step, the controlled results from the dual paths are fused using a simple convolutional block. In such a way, we eventually acquire a controlled event feature that represents the event double integral in Eq. (6).

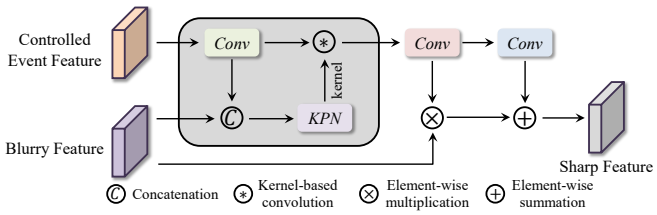



Figure 5. Illustration of modulation component.

### 4.3. Sharp Frame Reconstruction

With the event double integral controlled by timestamp and the estimated exposure prior, we are able to reconstruct the sharp frame using the event-based modulation in Eq. (1). In particular, we make some modifications to the physical model in Eq. (1). To avoid divide-by-zero, we use the multiplication between event and frame features instead of the division in Eq. (1). We embrace the kernel prediction network (KPN) [75] to learn a per-pixel kernel for alleviating the spatially variant noise disturbances. More precisely, as shown in Fig. 5, the blurry frame feature and controlled event double integral are jointly utilized to predict a learnable kernel. After rectification by the dynamic network, we proceed to deblur the blurry frame feature by multiplication with the event double integral in Eq. (6), to acquire the sharp frame feature added with a learnable bias.

In order to acquire the sharp frame, we choose a two-step reconstruction process as shown in Fig. 2. The initial sharp frame $I_{init}$ is generated by a reconstruction decoder. Then, we adopt a U-Net based detail restoration network [5] to produce the final sharp frame $I_{final}$. The underlined reason for our two-step process is that, for some non-event regions that mean a slight blurriness, the modulation in Eq. (1) will jumble the sharp feature, which leads to texture loss. To complement this, the aforementioned detail restoration network is exploited.

### 4.4. Training Implementation

**Loss Functions.** Two loss functions are used to reduce the difference between the estimated sharp frame $I$ and the ground-truth frame $I_{gt}$. The first is the Laplacian pyramid loss [29, 35, 36], which is defined as

$$\mathcal{L}_{\text{lap}}\left(I, I_{gt}\right)=\sum_{s=1}^{S} 2^{s-1}\left\|L^{s}\left(I\right)-L^{s}\left(I_{gt}\right)\right\|_{1}, \quad (7)$$

where $L^s\left(I\right)$ represents the $s^{th}$ level of the Laplacian pyramid of an image $I$, and $S$ is set as 5 in our implementation. We also adopt the census loss $\mathcal{L}_{cen}$ [24, 30] to alleviate the over-tighten constraint of the $\mathcal{L}_1$ loss, which can be formulated as

$$\mathcal{L}_{cen}(I, I_{gt}) = Dis(Cen(I), Cen(I_{gt})), \quad (8)$$

where $Dis(\cdot)$ and $Cen(\cdot)$ denote the soft Hamming distance and census transform [43], respectively. Then the final loss can be computed as $\mathcal{L} = \mathcal{L}_{lap} + \mathcal{L}_{cen}$.

**Stage-wise Training.** We adopt a stage-wise training strategy for optimizing our network. In specific, we first train the exposure estimation module by using the ground-truth exposure prior. We then train the remaining network (which means dropping exposure estimation module) by using the ground-truth exposure prior as the controlling coefficient. Eventually, we finetune the whole network. To optimize the exposure estimation module, we use the $\mathcal{L}_2$ loss to measure the difference between the estimated and ground-truth exposure priors. For constraining the initial and final sharp frames, we use loss $\mathcal{L}_{net} = \lambda\mathcal{L}(I_{init}, I_{gt}) + \mu\mathcal{L}(I_{final}, I_{gt})$, where $\lambda$, $\mu$ are the weighting factors.

## 5. Experiments

**Dataset Preparation.** We comprehensively evaluate our method on both synthetic and real-world datasets, in both quantitative and qualitative ways. We make our synthetic datasets using GoPro [33] and RealSharp-DAVIS [52]. For GoPro, the events are synthesized from the high frame-rate video using an event simulator [9]. In contrast, RealSharp-DAVIS provides real events captured by a DAVIS-346 color event camera. To mimic the real-world video frame acquisition in Fig. 1, we average consecutive frames within the exposure time to generate a blurry frame and discard several frames to simulate the data readout time. We use the pattern "$m$-$n$" for the exposure assumption, meaning that the frame number of exposure time and that of data readout time are $m$ and $n$, respectively. Experimentally, we define that the exposure time of the training dataset ranges from 9 to 15 with a shutter period of 16, termed TRAIN-MulEX. For the testing datasets, we define two types: TEST-I, whose exposure time ranges from 9 to 15 with a shutter period of 16 and is consistent with TRAIN-MulEX; and TEST-II, whose exposure time ranges from 7 to 11 with a shutter period of 12 and is inconsistent with TRAIN-MulEX. We follow the official data split for training and testing. Additionally, we also collect a real blurry video dataset called RealBlur-DAVIS, which contains real blurry frames and events captured by a DAVIS-346 color event camera. We set the exposure time as $\{30, 60, 90, 120, 150, 180\}$ ms, with a shutter period of 200 ms for generating low frame-rate blurry videos.

**Training Details and Evaluation Metrics.** We implement our method using PyTorch [42]. The Adam solver is used for optimization. To pretrain our exposure estimation module and the remaining network, we set the learning rate as $10^{-4}$, which is decayed by a factor of 0.5 every 200K iterations. For joint finetuning, we set the learning rate as $10^{-5}$. Before 10K iterations, the weighting factors are set as $\lambda = 0.1$ and $\mu = 1$, which are then changed to $\lambda = 1$ and $\mu = 0.1$. The batch size is set as 8. For quantitative evaluation, we adopt distortion metrics PSNR (dB) and SSIM [61] to compute the difference of all interpolated sharp frames (*i.e.*, $m + n$) and ground-truth latent sharp frames.

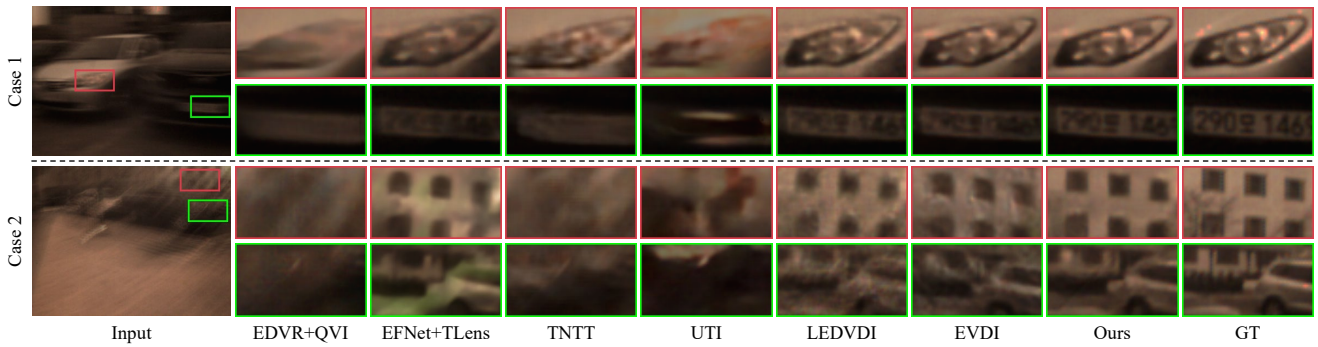



Figure 6. Visual comparisons of interpolated frames on semi-synthetic dataset RealSharp-DAVIS [52]. The events are captured by a DAVIS-346 color event camera, and the blurry frame is generated by averaging the consecutive sharp frames. *Case 1 is with train-test consistency and case 2 is with train-test inconsistency*. Zoom in for a better visual experience. **More results are in the supplement video**.

Table 1. Quantitative comparisons with train-test consistency. All interpolated frames (*i.e.*, $m$ + $n$) are used for evaluation. The four exposure assumptions are choosen from the training datasets. PSNR (the higher, the better) and SSIM (the higher, the better) are adopted for the evaluation of reconstruction accuracy. The best results are shown in bold font.

| Methods | Inputs | GoPro [33] | | | | | | | | RealSharp-DAVIS [52] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 9-7 | | 11-5 | | 13-3 | | 15-1 | | 9-7 | | 11-5 | | 13-3 | | 15-1 | |
| | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| EDVR [59]+SloMo [21] | RGB | 19.73 | 0.683 | 20.01 | 0.693 | 20.24 | 0.702 | 20.53 | 0.713 | 20.69 | 0.777 | 21.02 | 0.787 | 21.32 | 0.796 | 21.62 | 0.805 |
| EDVR [59]+QVI [68] | RGB | 19.80 | 0.683 | 20.10 | 0.694 | 20.35 | 0.703 | 20.65 | 0.714 | 20.94 | 0.786 | 21.28 | 0.797 | 21.56 | 0.805 | 21.81 | 0.813 |
| EFNet [27]+TLens [55] | RGB+Events | 22.80 | 0.820 | 22.84 | 0.821 | 22.87 | 0.822 | 22.84 | 0.822 | 23.66 | 0.879 | 23.77 | 0.880 | 23.86 | 0.881 | 23.88 | 0.881 |
| UEVD [52]+TLens [55] | RGB+Events | 22.78 | 0.818 | 22.85 | 0.820 | 22.91 | 0.822 | 22.92 | 0.822 | 23.65 | 0.879 | 23.77 | 0.881 | 23.85 | 0.882 | 23.83 | 0.882 |
| TNTT [23] | RGB | 19.44 | 0.661 | 19.90 | 0.679 | 20.30 | 0.695 | 20.69 | 0.710 | 20.76 | 0.780 | 21.21 | 0.790 | 21.59 | 0.798 | 21.97 | 0.808 |
| BIN [46] | RGB | 16.42 | 0.562 | 16.73 | 0.571 | 17.05 | 0.580 | 17.60 | 0.589 | 17.77 | 0.702 | 18.12 | 0.712 | 18.45 | 0.722 | 18.80 | 0.732 |
| UTI [73] | RGB | 19.61 | 0.660 | 19.63 | 0.660 | 19.60 | 0.659 | 19.57 | 0.659 | 20.05 | 0.756 | 20.14 | 0.760 | 20.20 | 0.763 | 20.25 | 0.767 |
| EDI [41] | Mono+Events | 19.29 | 0.690 | 19.28 | 0.695 | 19.22 | 0.697 | 19.13 | 0.694 | 19.13 | 0.722 | 19.17 | 0.729 | 19.17 | 0.734 | 19.11 | 0.736 |
| LEDVDI [28] | RGB+Events | 27.60 | 0.903 | 27.82 | 0.907 | 27.83 | 0.907 | 27.57 | 0.904 | 30.12 | 0.949 | 30.82 | 0.953 | 31.04 | 0.954 | 30.41 | 0.953 |
| EVDI [72] | RGB+Events | 25.87 | 0.871 | 25.86 | 0.872 | 25.62 | 0.869 | 25.21 | 0.862 | 27.51 | 0.912 | 27.37 | 0.912 | 26.99 | 0.909 | 26.49 | 0.906 |
| Ours | RGB+Events | **28.46** | **0.918** | **28.55** | **0.920** | **28.52** | **0.920** | **28.41** | **0.919** | **31.89** | **0.965** | **32.05** | **0.966** | **32.06** | **0.966** | **31.65** | **0.965** |

**Comparison Baselines.** For comparison, we employ two types of baselines. The first is the cascade solution, which stacks a deblurring model and a sharp frame interpolation model. For frame-based methods, we choose a representative deblurring model EDVR [59] with two interpolation models SloMo [21] and QVI [68]. For event-based methods, we choose two state-of-the-art deblurring models EFNet [27] and UEVD [52] with an interpolation model TimeLens (TLens) [55]. The other baselines are the joint solutions, which aim to directly interpolate sharp frames from blurry frames. The frame-based methods include TNTT [23], BIN [46] and UTI [73]. The event-based methods include EDI [41], LEDVDI [28] and EVDI [72].

**Results with Train-test Consistency.** Instead of specifying the fixed and known exposure time, we assume that the exposure time is variable and unknown. In specific, the training and testing datasets contain more than one exposure time assumption. For fair comparisons, we retrain EFNet, UEVD, TLens, LEDVDI and EVDI on TRAIN-MulEX. For frame-based methods, we directly use the pretrained models provided by the projects. We first demonstrate the results on TEST-I, which show train-test consistency and are presented in Tab. 1. According to these results, some important conclusions can be made.

First, event streams are able to compensate for the lost motion information, boosting the video restoration by a large margin. Tab. 1 suggests that for the cascade models, event-based methods consistently outperform frame-based methods, achieving a near 3 dB gain in terms of PSNR on both GoPro and RealSharp-DAVIS. For the joint models, event-based methods reach the second best performance except EDI. Note that EDI is an optimization method, which shows the limited ability of restoration.

Second, the joint models show more promising results than the cascade models. This is supported by the fact that event-based joint models significantly surpass the event-based cascade models, yielding over a 5 dB average PSNR boost on both GoPro and RealSharp-DAVIS. Generally, cascade models are sub-optimal due to the error accumulation of two separated deblurring and interpolation models, resulting in inferior performance. For frame-based methods, the joint models and cascade models show similar results, which indicates that the blind exposure setting degenerates the effectiveness of these methods. Obviously, our method consistently outperform existing methods whether the cascade or joint models, on both GoPro and RealSharp-DAVIS. The visual results in the Case 1 of Fig. 6 can also support the above conclusions.

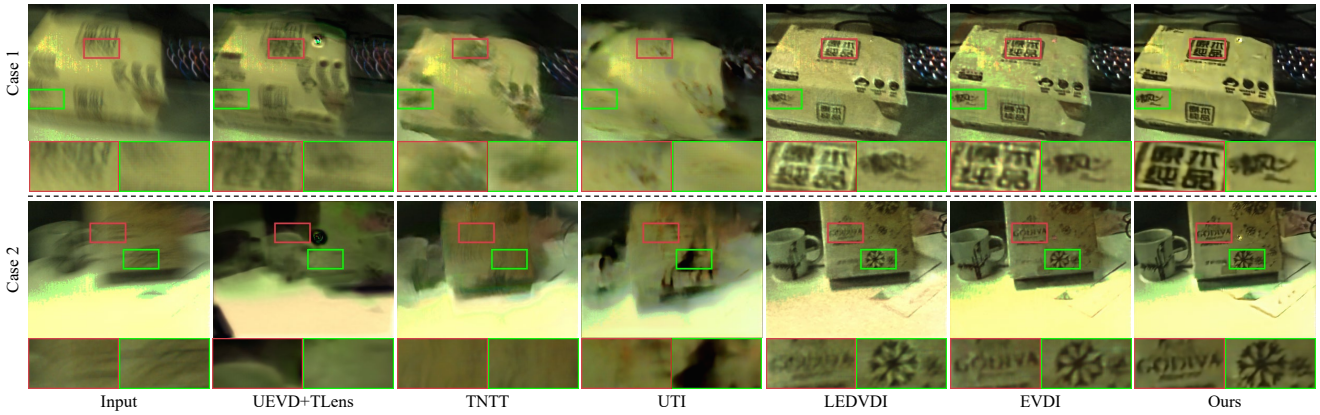



Figure 7. Visual comparisons of interpolated frames on the self-collected real-world blurry video dataset RealBlur-DAVIS. In particular, the blurry frame and events are captured by a DAVIS-346 color event camera. We show two cases here, which are *more challenging and are all train-test inconsistency*. Please zoom in for a better visual experience. **More results are in the supplement video**.

Table 2. Quantitative comparisons with train-test inconsistency. Compared with Tab. 1, **the exposure assumptions here are different from the training datasets**. Complexity analysis is also reported in terms of FLOPs (G) and runtime (ms) that are measured using an RTX 3090 GPU. The best results are shown in bold font. The results on GoPro [33] are in the supplement.

| Methods | RealSharp-DAVIS [52] | | | | | | FLOPs(G) /Time(ms) |
|---|---|---|---|---|---|---|---|
| | 7-5 | | 9-3 | | 11-1 | | |
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | |
| EDVR [59]+SloMo [21] | 21.74 | 0.799 | 22.18 | 0.812 | 22.60 | 0.824 | 119.19/144 |
| EDVR [59]+QVI [68] | 21.84 | 0.805 | 22.27 | 0.819 | 22.64 | 0.830 | 134.09/208 |
| EFNet [27]+TLens [55] | 24.48 | 0.888 | 24.61 | 0.889 | 24.66 | 0.890 | 117.18/54 |
| UEVD [52]+TLens [55] | 24.45 | 0.888 | 24.59 | 0.890 | 24.63 | 0.890 | 216.20/96 |
| TNTT [23] | 21.63 | 0.791 | 22.22 | 0.804 | 22.78 | 0.816 | 88.75/34 |
| BIN [46] | 18.46 | 0.713 | 18.92 | 0.726 | 19.41 | 0.738 | 467.00/204 |
| UTI [73] | 20.53 | 0.777 | 20.62 | 0.779 | 20.67 | 0.781 | 231.66/199 |
| EDI [41] | 19.85 | 0.739 | 19.92 | 0.749 | 19.88 | 0.754 | -/- |
| LEDVDI [28] | 25.48 | 0.862 | 25.97 | 0.870 | 26.10 | 0.875 | 200.25/44 |
| EVDI [72] | 28.51 | 0.925 | 28.23 | 0.923 | 27.59 | 0.919 | **29.04**/66 |
| Ours | **31.27** | **0.958** | **31.34** | **0.959** | **31.28** | **0.959** | 39.10/**29** |

**Results with Train-test Inconsistency.** We then present the results on TEST-II in Tab. 2 using the models trained on TRAIN-MulEX, which are more challenging due to the severe exposure inconsistency of the training and testing datasets. In particular, we focus more on the results of LEDVDI, EVDI and ours, which are the top three most competent methods. Recalling the results of train-test consistency in Tab. 1, LEDVDI presents better results than EVDI, yielding a near 2 dB PSNR gain on GoPro and a near 3 dB PSNR gain on RealSharp-DAVIS. However, from the results of train-test inconsistency as shown in Tab. 2, we can observe a reversed conclusion that EVDI performs favorably than LEDVDI especially for the case of "7-5" that shows a large train-test gap. This can be attributed to the fact that LEDVDI is specifically designed based on the assumption of a known and fixed exposure time. When transferred to the blind exposure, the problem formulation of LEDVDI can no longer hold, leading to severe degradation in reconstruction quality. While EVDI builds a physical model that supports diverse light conditions, which guarantees better results compared with LEDVDI. In contrast, thanks to the exposure estimation, our method is able to adaptively estimate the lost exposure prior, thus maintaining superior performance, which is also validated by the Case 2 of Fig. 6. Being a blind exposure based deblurring method, UEVD [52] lacks a decent physical model as an effective guidance, leading to under-performant results. Additionally, our method is efficient with fewer FLOPs and runtime compared with other competitors as shown in Tab. 2.

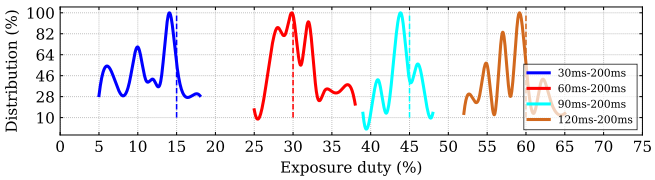
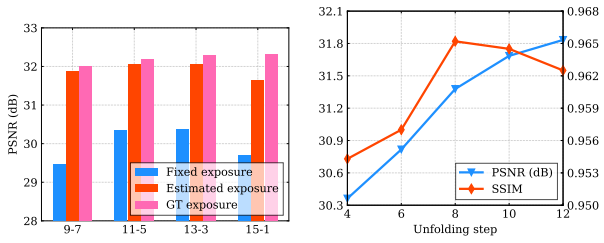



Figure 8. Exposure estimation results on the self-collected RealBlur-DAVIS dataset using the model trained on synthetic data. The color solid and color dash denote the estimated and ground-truth exposure duty, respectively. The pattern "$A$ms-$B$ms" in the legend denotes that the exposure time is $A$ ms with a shutter period of $B$ ms. The exposure duty can be calculated as $A/B \times 100\%$ .

**Results on Real Blurry Video.** We make real-world evaluations on the self-collected dataset RealBlur-DAVIS that contains 20 scenes with diverse light conditions and motion speeds. Due to the difficulty of capturing ground-truth sharp frames, we only present the qualitative results. As shown in Fig. 7, our method achieves the best visual results in comparison with the competitors. Specifically, more sharper edges and details are reconstructed by our method. In contrast, frame-based methods fail to reconstruct a satisfactory sharp frame due to the variable and unknown exposure time in the blind exposure problem. Event-based methods such as LEDVDI and EVDI, present remarkable improvements over frame-based methods. However, undesired visually blurry effects also exist, which are introduced by the inaccurate problem definition.




(a) (b)

Figure 9. Investigation results of exposure estimation (a) and unfolding step (b) on RealSharp-DAVIS [52].

**Exposure Estimation Results on Real Blurry Video.** Furthermore, we conduct exposure estimation experiments on the self-collected real blurry video dataset RealBlur-DAVIS. In specific, for each blurry frame, we estimate the corresponding exposure duty using our exposure estimation module. Subsequently, we count the distribution of the estimated exposure duty. The ground-truth exposure duty is also illustrated for better comparison. As shown in Fig. 8, *the estimated exposure duty peak is near the ground-truth exposure duty without any temporal prior provided*. Such results validate the adaptation ability of our method, even though our model is trained solely on the synthetic datasets.

## 6. Methodology Analysis

**Effectiveness of Exposure Estimation.** We propose to estimate the exposure prior to make the blind exposure problem well-posed. To validate the effectiveness of our exposure estimation, we drop this module and manually fix the value of exposure prior, which is then provided to the temporal-exposure control. As shown in Fig. 9 (a), when disabling the exposure estimation, the performance drops significantly for all the exposure assumptions. We also present the results using the ground-truth exposure prior, which is the upper bound. These results concretely demonstrate the effectiveness of the proposed exposure estimation.

**Investigation on Unfolding Step.** As mentioned in Sec. 4.2, in order to learn the temporal-exposure aware property of event double integral, we unfold our temporal-exposure control module in multiple steps. For investigating the influence of the unfolding step, we conduct some experiments. We illustrate the results in Fig. 9 (b). As can be observed, the PSNR intensifies as the unfolding step increases. However, SSIM shows a weak drop when the unfolding step exceeds 8. Moreover, the larger unfolding steps lead to the larger complexity. Therefore, we choose 8 as our final unfolding step.

**Ablation on Inputs of Exposure Estimation.** In order to estimate a more accurate exposure prior, we make an elaborate design by utilizing the mutual correlation of a blurry frame, corresponding events and the blurry level. To validate the necessity of our design, we conduct an ablation study on these inputs. As shown in Tab. 3, given events as a complement, we can achieve a remarkable performance boost and a more stable training in comparison to the model with only RGB. This indicates that the exposure ambiguity can be well regularized by the events, benefiting the estimation process. Furthermore, the performance can be boosted when the blurry level is provided. We experimentally choose the Laplacian function [1] for calculating the blurry level due to its clear performance advantage.

Table 3. Ablation results of each component on RealSharp-DAVIS [52]. Abbreviation for some components: "Ev." (events), "Lap." (Laplacian), "DCH." (dark channel), "mod." (modulation) and "rest." (restoration). The best results are shown in bold font.

| exposure estimation | | | | dual | revised | detail | stage-wise | metrics | |
|---|---|---|---|---|---|---|---|---|---|
| RGB | Ev. | Lap. | DCH. | path | mod. | rest. | training | PSNR | SSIM |
| ✓ | | | | | | | | 28.420 | 0.931 |
| ✓ | ✓ | | | | | | | 29.529 | 0.943 |
| ✓ | ✓ | | ✓ | | | | | 29.944 | 0.948 |
| ✓ | ✓ | ✓ | | | | | | 30.185 | 0.952 |
| ✓ | ✓ | ✓ | | ✓ | | | | 30.675 | 0.958 |
| ✓ | ✓ | ✓ | | ✓ | ✓ | | | 30.759 | 0.962 |
| ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | 31.378 | 0.965 |
| ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | **31.913** | **0.966** |

**Ablation on Model Architecture.** In addition to the exposure estimation, our model is constructed with the proposed temporal-exposure control, revised modulation and detail restoration. To investigate the influence of these designs, we make additional ablations. For ablating temporal-exposure control, we replace our dual path design with the single path [15]. We ablate the KPN based modulation in Fig. 5 using a simple convolutional network or directly drop the detail restoration network for the remaining ablations. Tab. 3 reflects that all these designs further elevate the performance, revealing the potential to facilitate the BFI.

**Effectiveness of Stage-wise Training.** For training our network, we devise a stage-wise training strategy as demonstrated in Sec. 4.4. Instead of using the warm-up pretraining followed by the joint finetuning, we directly adopt an end-to-end training manner for comparison. As shown in Tab. 3, our stage-wise training brings a PSNR gain of about 0.5 dB, which validates its effectiveness.

## 7. Conclusion

In this paper, we reveal a new challenge of blurry frame interpolation under blind exposure. To conquer this challenge, we provide a new perspective, by using an event camera. We implement an event-based unified model in an effective and efficient way, by proposing the exposure estimation strategy guided by event streams and the temporal-exposure control strategy achieved with iterative residual learning. As validated by comprehensive experiments, we demonstrate the superior performance in dealing with blind exposure based blurry frame interpolation in practice.

## References


[1] Raghav Bansal, Gaurav Raj, and Tanupriya Choudhury. Blur image detection using laplacian operator and open-cv. In *SMART*, 2016. 4, 8

[2] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. Depth-aware video frame interpolation. In *CVPR*, 2019. 1

[3] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *NeurIPS*, 2019. 3

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020. 3

[5] Duolikun Danier, Fan Zhang, and David Bull. St-mfnet: A spatio-temporal multi-flow network for frame interpolation. In *CVPR*, 2022. 5

[6] Alexey Dosovitskiy, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with convolutional neural networks. In *NeurIPS*, 2014. 3

[7] John Flynn, Ivan Neulander, James Philbin, and Noah Snavely. Deepstereo: Learning to predict new views from the world's imagery. In *CVPR*, 2016. 1

[8] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020. 2

[9] Daniel Gehrig, Mathias Gehrig, Javier Hidalgo-Carrio, and Davide Scaramuzza. Video to events: Recycling video datasets for event cameras. In *CVPR*, 2020. 5

[10] Daniel Gehrig, Antonio Loquercio, Konstantinos G Derpanis, and Davide Scaramuzza. End-to-end learning of representations for asynchronous event-based data. In *ICCV*, 2019. 4

[11] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, 2019. 3

[12] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *CVPR*, 2006. 3

[13] Jin Han, Yixin Yang, Chu Zhou, Chao Xu, and Boxin Shi. Evintsr-net: Event guided multiple latent frames reconstruction and super-resolution. In *ICCV*, 2021. 2

[14] Jingwen He, Chao Dong, and Yu Qiao. Modulating image restoration with continual levels via adaptive feature modification layers. In *CVPR*, 2019. 4

[15] Jingwen He, Chao Dong, and Yu Qiao. Interactive multi-dimension modulation with dynamic controllable residual learning for image restoration. In *ECCV*. Springer, 2020. 4, 8

[16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020. 3

[17] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12), 2010. 4

[18] Weihua He, Kaichao You, Zhendong Qiao, Xu Jia, Ziyang Zhang, Wenhui Wang, Huchuan Lu, Yaoyuan Wang, and Jianxing Liao. Timereplayer: Unlocking the potential of event cameras for video interpolation. In *CVPR*, 2022. 2

[19] Tae Hyun Kim and Kyoung Mu Lee. Generalized video deblurring for dynamic scenes. In *CVPR*, 2015. 2

[20] Tae Hyun Kim, Kyoung Mu Lee, Bernhard Scholkopf, and Michael Hirsch. Online video deblurring via dynamic temporal blending network. In *ICCV*, 2017. 1, 2

[21] Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik Learned-Miller, and Jan Kautz. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In *CVPR*, 2018. 1, 6, 7

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

[23] Meiguang Jin, Zhe Hu, and Paolo Favaro. Learning to extract flawless slow motion from blurry videos. In *CVPR*, 2019. 1, 6, 7

[24] Lingtong Kong, Boyuan Jiang, Donghao Luo, Wenqing Chu, Xiaoming Huang, Ying Tai, Chengjie Wang, and Jie Yang. Ifrnet: Intermediate feature refine network for efficient frame interpolation. In *CVPR*, 2022. 5

[25] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *CVPR*, 2018. 1

[26] Hyeongmin Lee, Taeoh Kim, Tae-young Chung, Daehyun Pak, Yuseok Ban, and Sangyoun Lee. Adacof: Adaptive collaboration of flows for video frame interpolation. In *CVPR*, 2020. 2

[27] Sun Lei, Sakaridis Christos, Liang Jingyun, Jiang Qi, Yang Kailun, Sun Peng, Ye Yaozu, Wang Kaiwei, and Gool Luc Van. Event-based fusion for motion deblurring with cross-modal attention. In *ECCV*. Springer, 2022. 2, 6, 7

[28] 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*. Springer, 2020. 6, 7

[29] Yihao Liu, Liangbin Xie, Li Siyao, Wenxiu Sun, Yu Qiao, and Chao Dong. Enhanced quadratic video interpolation. In *ECCV*. Springer, 2020. 2, 5

[30] Simon Meister, Junhwa Hur, and Stefan Roth. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In *AAAI*, 2018. 5

[31] Simone Meyer, Abdelaziz Djelouah, Brian McWilliams, Alexander Sorkine-Hornung, Markus Gross, and Christopher Schroers. Phasenet for video frame interpolation. In *CVPR*, 2018. 1, 2

[32] Simone Meyer, Oliver Wang, Henning Zimmer, Max Grosse, and Alexander Sorkine-Hornung. Phase-based frame interpolation for video. In *CVPR*, 2015. 1, 2

[33] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 1, 5, 6, 7

[34] Seungjun Nah, Sanghyun Son, and Kyoung Mu Lee. Recurrent neural networks with intra-frame iterations for video deblurring. In *CVPR*, 2019. 2

[35] Simon Niklaus and Feng Liu. Context-aware synthesis for video frame interpolation. In *CVPR*, 2018. 1, 2, 5

[36] Simon Niklaus and Feng Liu. Softmax splatting for video frame interpolation. In *CVPR*, 2020. 2, 5

[37] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive convolution. In *CVPR*, 2017. 1, 2

[38] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In *ICCV*, 2017. 1, 2

[39] Thekke Madam Nimisha, Akash Kumar Singh, and Ambasamudram N Rajagopalan. Blur-invariant deep learning for blind-deblurring. In *ICCV*, pages 4752–4760, 2017. 1

[40] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *CVPR*, 2016. 1, 2, 4

[41] 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. 2, 6, 7

[42] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NeurIPSW*, 2017. 5

[43] Zabih Ramin and Woodfill John. Non-parametric local transforms for computing visual correspondence. In *ECCV*. Springer, 1994. 5

[44] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *CVPR*, 2015. 2

[45] 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. 2

[46] Wang Shen, Wenbo Bao, Guangtao Zhai, Li Chen, Xiongkuo Min, and Zhiyong Gao. Blurry video frame interpolation. In *CVPR*, 2020. 1, 6, 7

[47] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *CVPR*, 2018. 3

[48] Alon Shoshan, Roey Mechrez, and Lihi Zelnik-Manor. Dynamic-net: Tuning the objective without re-training for synthesis tasks. In *ICCV*, 2019. 4

[49] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR*, 2020. 3

[50] Wu Song, You Kaichao, He Weihua, Yang Chen, Tian Yang, Wang Yaoyuan, Zhang Ziyang, and Liao Jianxing. Video interpolation by event-driven anisotropic adjustment of optical flow. In *ECCV*. Springer, 2022. 2

[51] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *CVPR*, 2017. 1

[52] Kim Taewoo, Lee Jeongmin, Wang Lin, and Yoon Kuk-Jin. Event-guided deblurring of unknown exposure time videos. In *ECCV*. Springer, 2022. 2, 5, 6, 7, 8

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

[54] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li, and Davide Scaramuzza. Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion. In *CVPR*, 2022. 2

[55] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time lens: Event-based video frame interpolation. In *CVPR*, 2021. 2, 6, 7

[56] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. Event enhanced high-quality image recovery. In *ECCV*. Springer, 2020. 2

[57] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *CVPR*, 2021. 3

[58] Wei Wang, Ruiming Guo, Yapeng Tian, and Wenming Yang. Cfsnet: Toward a controllable feature space for image restoration. In *ICCV*, 2019. 4

[59] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *CVPRW*, 2019. 1, 6, 7

[60] Xintao Wang, Ke Yu, Chao Dong, Xiaoou Tang, and Chen Change Loy. Deep network interpolation for continuous imagery effect transition. In *CVPR*, 2019. 4

[61] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[62] Wenming Weng, Yueyi Zhang, and Zhiwei Xiong. Event-based video reconstruction using transformer. In *ICCV*, 2021. 2

[63] Wenming Weng, Yueyi Zhang, and Zhiwei Xiong. Boosting event stream super-resolution with a recurrent neural network. In *ECCV*. Springer, 2022. 2

[64] Chao-Yuan Wu, Nayan Singhal, and Philipp Krahenbuhl. Video compression through image interpolation. In *ECCV*, 2018. 1

[65] Yuxin Wu and Kaiming He. Group normalization. In *ECCV*, 2018. 4

[66] Zeyu Xiao, Wenming Weng, Yueyi Zhang, and Zhiwei Xiong. Eva2: Event-assisted video frame interpolation via cross-modal alignment and aggregation. *IEEE Transactions on Computational Imaging*, 8:1145–1158, 2022. 2

[67] 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. 2

[68] Xiangyu Xu, Li Siyao, Wenxiu Sun, Qian Yin, and Ming-Hsuan Yang. Quadratic video interpolation. In *NeurIPS*, 2019. 1, 2, 6, 7

[69] Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding. *IEEE Journal of Solid-State Circuits*, 50(9):2149–2160, 2015. 2

[70] Zhiyang Yu, Yu Zhang, Deyuan Liu, Dongqing Zou, Xijun Chen, Yebin Liu, and Jimmy S Ren. Training weakly supervised video frame interpolation with events. In *ICCV*, 2021. 2

[71] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *CVPR*, 2018. 3

[72] Xiang Zhang and Lei Yu. Unifying motion deblurring and frame interpolation with events. In *CVPR*, 2022. 3, 6, 7

[73] Youjian Zhang, Chaoyue Wang, and Dacheng Tao. Video frame interpolation without temporal priors. In *NeurIPS*, 2020. 1, 2, 6, 7

[74] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *ECCV*. Springer, 2020. 2

[75] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Haozhe Xie, Wangmeng Zuo, and Jimmy Ren. Spatio-temporal filter adaptive network for video deblurring. In *ICCV*, 2019. 2, 5

[76] Yunhao Zou, Yinqiang Zheng, Tsuyoshi Takatani, and Ying Fu. Learning to reconstruct high speed and high dynamic range videos from events. In *CVPR*, 2021. 2