

Received July 10, 2020, accepted August 7, 2020, date of publication August 11, 2020, date of current version August 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015759

# Hybrid Deblur Net: Deep Non-Uniform Deblurring With Event Camera

LIMENG ZHANG<sup>1</sup>, HONGGUANG ZHANG, JIHUA CHEN,  
AND LEI WANG, (Member, IEEE)

College of Computer, National University of Defense Technology, Changsha 410073, China

Corresponding author: Lei Wang (leiwang@nudt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB2202603, in part by Hegaoji of China under Grant 2017ZX01028103002, and in part by the National Natural Science Foundation of China under Grant 61802427 and Grant 61832018.

**ABSTRACT** Despite CNN-based deblur models have shown their superiority when solving motion blurs, restoring a photorealistic image from severe motion blur remains an ill-posed problem due to the loss of temporal information and textures. Event cameras such as Dynamic and Active-pixel Vision Sensor (DAVIS) can simultaneously produce gray-scale Active Pixel Sensor (APS) frames and events, which can capture fast motions as events of very high temporal resolution, *i. e.*,  $1\mu s$ , can provide extra information for blurry APS frames. Due to the natural noise and sparsity of events, we employ a recurrent encoder-decoder architecture to generate dense recurrent event representations, which encode the overall historical information. We concatenate the original blurry image with the event representation as our hybrid input, from which the network learns to restore the sharp output. We conduct extensive experiments on GoPro dataset and a real event blurry dataset captured by DAVIS240C. Our experimental results on both synthetic and real images demonstrate state-of-the-art performance for  $1280 \times 720$  images at 30 fps.

**INDEX TERMS** Event-based vision, high speed, image deblurring, real-time.

## I. INTRODUCTION

Motion blur is one kind of image degradation due to the long exposure time of a conventional camera. Object movement and camera shake during exposure contribute to complex blur kernels in captured pictures. Traditional deblurring models try to estimate blur kernels via a variety of priors or regularizations, and most of these approaches require intensive parameter-tuning and expensive computation.

Recent deep learning methods have shown their superiority in motion deblurring tasks. Early methods follow the idea of traditional methods, which leverage regularization priors, or substitute some operators with learned models [18], [19], [22]. Recent methods try to design end-to-end networks to learn the connections between blurry images and the corresponding sharp images without estimating the blur kernels [12], [20], [25], [26].

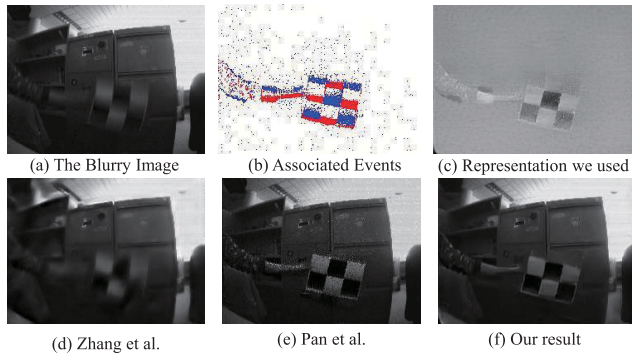
Non-uniform motion deblurring remains a highly ill-posed problem due to the loss of some important information, such as time information and image textures are destroyed. Although previous methods have made significant progress

in advancing the deblurring performance, they may fail in challenging cases, *e. g.*, images with severe motion blur and high dynamic range. As shown in Fig. 1(d), the state-of-the-art deep deblurring model [25] cannot restore the challenging motion blur information in such conditions.

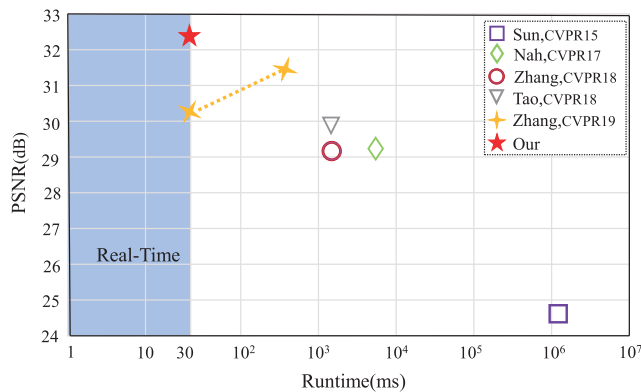
Numbers of approaches, *e. g.*, the use of the coarse-to-fine scheme and increasing model depth with finer-scale levels, have been proposed to address this problem, however, their benefits are marginal. Instead of merely relying on the blurry images, our work proposes to leverage the complete record of temporal information obtained by an event camera to solve the above issue.

Reference [3] are bio-inspired vision sensors that work differently from a traditional camera. Instead of accumulating light intensity during the fixed exposure time, the event camera records the changes of intensity asynchronously in microseconds. The output of event camera is a stream of events shaped in a four-dimensional array  $(x,y,t,p)$  that encodes time, location, and polarity of brightness changes at a very high dynamic range (140dB) respectively. As shown in Fig. 1(a), though the image is severely blurred, the corresponding event information Fig. 1(b) is abundant, and the recurrent event representation Fig. 1(c) used in this work

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan<sup>1</sup>.



**FIGURE 1.** Motivation of our model. A Severed motion-blurred image (a) can hardly be restored with the state-of-the-art deep learning model [25] (d) with an only blurry image. (e) [14] formulates the relationship between events (b) and blurry image (a) via a Event-based Double Integral (EDI) model. (f) refers to our proposed hybrid deblurring network, which learns to combine recurrent event representation (c) with a blurry image to restore a photorealistic image.



**FIGURE 2.** The PSNR vs. runtime of state-of-the-art learning-based motion deblurring models on the GoPro dataset [12]. The blue region indicates real-time inference. Our models achieve the best performance with the PSNR of 32.25dB for 1280 × 720 images.

include clear outlines, which are beneficial to restore blurry images.

In this work, we propose a two-phase hybrid deblurring network to capture motion blurs photorealistic. To deal with the noise and sparsity of event data, we employ a recurrent encoder-decoder architecture at phase 1 to generate recurrent event representations, then we concatenate the blurry image with its recurrent event representation (output of phase 1) as the input of phase 2 to restore the blurry images.

Inspired by [25], we use a simple multi-patch hierarchical encoder-decoder architecture at phase 2 to learn the residual between the abundant concatenated input and the target sharp image. We generate the simulated event data with ESIM [15], and evaluate the performance of our model on GoPro dataset [12]. Both qualitative and quantitative results show state-of-the-art performance *w. r. t.* Peak Signal to Noise Ratio (PSNR) as depicted in Fig. 2. We further evaluate our model on a real event camera dataset [14], which is captured by DAVIS240C [3]. Our qualitative results show our hybrid deblurring model can effectively restore a photorealistic sharp image in challenging conditions. To the best of

our knowledge, our proposed model is the first event-based deep learning deblurring model.

Our contributions in this paper are listed as follows.

- We formulate the motion deblurring as a residual learning task and propose to leverage recurrent event representations as latent image-like complementary.
- We propose a novel two-phase hybrid deblur net, in which phase 1 uses a recurrent encoder-decoder model to convert sparse events into detailed event representation, while phase 2 uses a multi-patch model to deblur in a fine-to-coarse manner. The deblur net concatenates a blurry image with the output of phase 1 to restore the image from non-uniform motion blurs.
- Our proposed hybrid architecture achieves state-of-the-art results on both synthetic and real blurry datasets and can deblur 1280 × 720 images at 30fps.

## II. RELATED WORK

### A. CONVENTIONAL IMAGE DEBLURRING

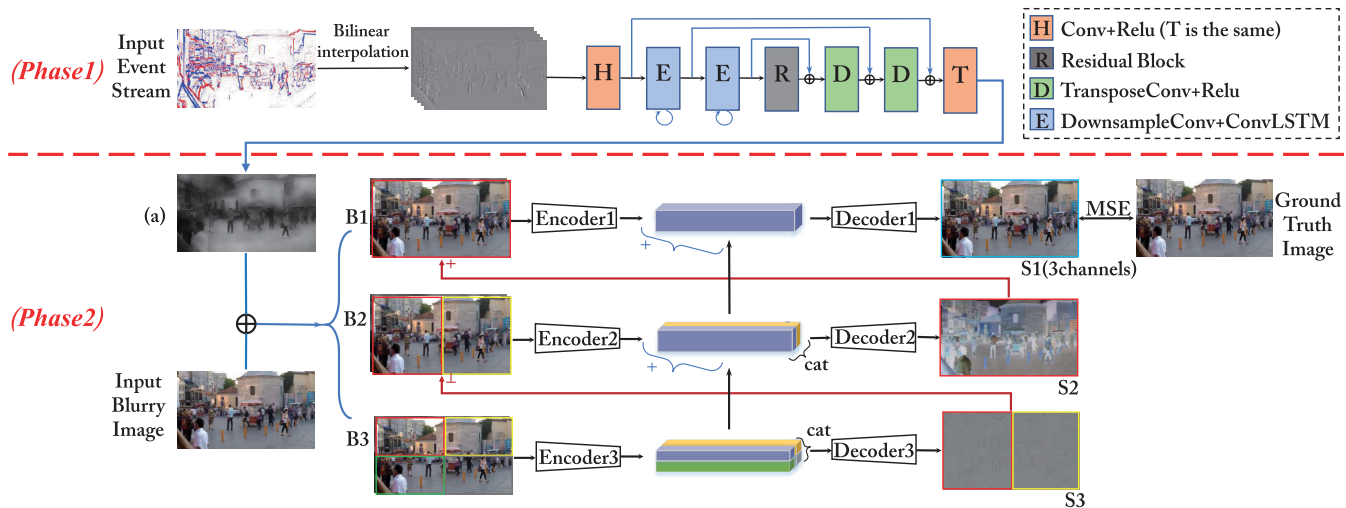
Early research into image deblurring usually presents many priors and assumptions. A lot of works [5], [13], [24] fail to remove non-uniform motion blur to estimate the global blur kernel. Recently, [14] formulation a deblurring method to an optimization problem by solving a single variable non-convex problem called the Event-based Double Integral (EDI) model. It performs well to solve the motion blur problem under low light and complex dynamic conditions. However, the natural noise of the event camera introduces accumulated noise and the loss of details, which makes the restored image non-photorealistic.

### B. LEARNING-BASED IMAGE DEBLURRING

Reference [19] proposes a convolution neural network to estimate locally blur kernel, then used the conventional energy-based optimization to estimate the latent sharp image. Reference [6] uses a fully convolutional neural network to estimate optical flow from a single blurry image, then restore the blurry image from the estimated optical flow. Reference [12] proposes a multi-scale CNN to restore sharp images in an end-to-end manner without estimating the blur kernel. Reference [20] proposes a coarse-to-fine SRN-Deblurnet to restore the blurry image on different levels. Reference [7] proposes to take consecutive multiple frames as input, restoring the middle sharp image. Reference [25] proposes a deep hierarchical multi-patch network via a fine-to-coarse hierarchical representation, exploiting the deblurring cues at different scales. It is the first real-time deep motion deblurring model for 720p images at 30fps.

### C. EVENT BASED INTENSITY IMAGE RECONSTRUCTION

Event cameras such as DAVIS [3] and Dynamic Vision Sensor (DVS) [10] recording intensity changes at a microsecond level do not suffer from motion blur. Reference [1] proposes to estimate optical flow and intensity images simultaneously by minimizing energy. Reference [11] restores intensity images through manifold regularization.



**FIGURE 3.** Our proposed Hybrid Deblurring Network(HDN) has two phases. Phase 1 uses a recurrent encoder-decoder network to generate a detailed event representation(a). Then we concatenate the representation with a corresponding blurry image to pass them into the 3-level multi-patch hierarchical network. Symbol + is a residual-like summation.

DAVIS camera uses a share camera sensor that can simultaneously output events and intensity images (APS). Due to the noise and loss of details reconstructed with only events, [17] proposes an asynchronous event-driven complementary filter to combine the APS frame with events. However, if the APS frame suffers from motion blur, the complementary filter can reconstruct the intensity image events only. Reference [29] directly integrates events to an APS frame and resets the integration to avoid accumulated noise. But this method lost the historical information of events. Reference [16] proposes a full convolutional recurrent Unet-like architecture to reconstruct intensity images from events only.

In this work, we propose a two-phase hybrid deblurring network to restore a sharp image by concatenating a blurry image with its recurrent event representation as the input for a simple multi-patch hierarchical deblurring model. Compared with conventional deblurring methods, we leverage the event information to help solve the ill-posed deblurring problem. Compared with event only image reconstruction, by using a detailed intensity image, our restored images are noise-less, dense, and photorealistic.

### III. HYBRID DEBLURRING MODEL

#### A. FORMULATION

We denote the blurry image input as  $B$ , our objective is to restore a sharp image from the blurry image with its corresponding events.  $E(t)$  refers to a set of events. Inspired by [14],

$$L(t) = L(t - T) + cE(t) \tag{1}$$

Equation (1) shows that the adjacent image can be achieved by integrating the events between exposure time, where  $t$  refers to the current timestamp,  $T$  refers to the exposure time and  $L(t)$  refers to the latent sharp image of current timestamp. Because the previous latent image is impossible to get, we decide to exploit the complete information of

the event, with a recurrent representation of the event, we can approximately get the similarly sharp image as  $L(t - T)$ .

In general, we formulate the deblurring task as a residual learning model and exploit the complete information encoded in the event stream. Due to the noise and sparsity of events, we adopt a recurrent event representation to be latent image-like input as  $L(t - T)$  in Equation (1).

#### B. OVERVIEW

The pipeline of our hybrid model is depicted in Fig. 3. Given blurry images with corresponding events, our network can output a sequence of deblurred images. The network contains two phases. Phase 1 is a preprocessing of event data. Inspired by [16], we use a fully convolutional encoder-decoder architecture, composed of 2 recurrent layers  $E$ , followed by 1 residual block  $R$  and 2 decoder layers  $D$ , with skip connections between symmetric layers. The encoders compose a stride 2 convolution followed by a convLSTM [23]. Decoder blocks use a transpose convolution. We use ReLU activations and batch normalization after each layer, the last prediction layer uses a sigmoid activation. We will discuss the differences between several event representations in Section 3.3 in detail.

Phase 2 concatenates the output of phase 1 and the blurry images as input, based on a multi-patch encoder-decoder architecture, which has shown superior performance in the deblurring task (refer to Section 3.4 for details.). We use the same architecture of encoder and decoder as [25] as shown in Fig.4. The numbers on each layer are their parameters. From top to bottom are the input channels, output channels, kernel size, and stride. Padding is 1 for all convolution layers.

#### C. EVENT REPRESENTATION (PHASE 1)

To process the event stream, we need to convert the event stream into an image-like representation. A natural choice is to directly integrate the event on a 2D plane.

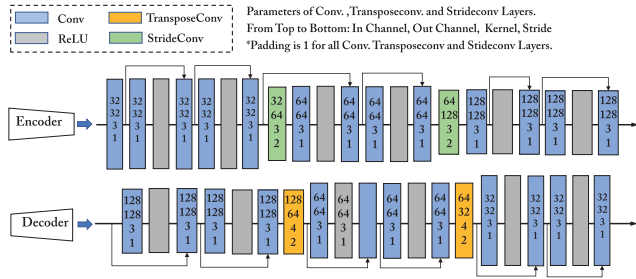


FIGURE 4. The architecture and layer configurations of our encoder and decoder.

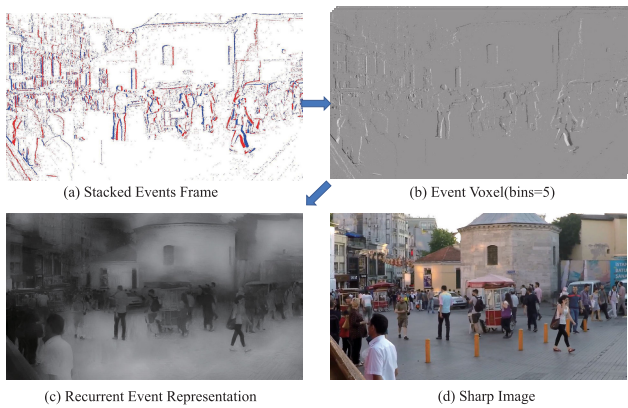


FIGURE 5. (a) A stacked frame that compresses events over exposure time. (b) A kind of event representation of voxel [28]. (c) Recurrent event representation of our phase 1 output. (d) Related sharp image.

Reference [28] proposes to encode the events in a spatial-temporal voxel grad. The events over a period of time will be discretized into  $B$  temporal bins, so the input of the network is a  $H \times W \times B$  image-like event tensor as shown in Fig. 5, where  $H$  and  $W$  are the sensor height and width.

The event tensor is usually using bilinear interpolation where each event  $(x, y, t, p)$  contributes its polarity to its two closest temporal bins according to:

$$E(x, y, t_n) = \sum_i p_i \max(0, 1 - |t_n - t_i^*|) \quad (2)$$

$$t_i^* = \frac{(t_i - t_{min})}{(t_{max} - t_{min})} (B - 1) \quad (3)$$

where  $n$  is the index of the temporal normalized index,  $p$  is the polarity, and  $t^*$  is the normalized timestamp of the  $i^{th}$  event. Following [16], we set to  $B$  to 5.

As depicted in Fig.5, for directly integration(a), spatial-temporal voxel(b), and others frames like representation [4], [21], [27], they may suffer from noise or lack details because of sparsity with little events. We need a representation that is similar to the latent sharp image with rich information.

In this work, we propose a two-phase hybrid deblurring network to restore a sharp image by putting a concatenated blurry image with a recurrent event representation into a simple multi-patch hierarchical deblurring model. Compared with the conventional deblurring method, we join the event

information which encodes complete temporal information to help solve the ill-posed deblurring problem. Compared with event only intensity image reconstruction, by using a detailed intensity image, our restored images are noise-less, dense, and photorealistic.

#### D. RESIDUAL LEARNING MODEL (PHASE 2)

Reference [25] presents a hierarchical multi-patch network inspired by Spatial Pyramid Matching [9]. The model makes the lower level focus on local information to produce residual information for the coarser level, and shows state-of-the-art performance. Different from [25], which stacks several (1-2-4-8) models to improve accuracy, We only adopt the simplest (1-2-4) model and achieve significant improvement by combining events. The notation (1-2-4) indicates the numbers of image patches from the coarsest to the finest level. i.e., a vertical split at the second level,  $2 \times 2 = 4$  splits at the third level.

Each level of our residual learning model consists of an encoder-decoder pair. The input of each level  $i$  is denoted as  $B_i$ , which is the sum of the concatenated input and the output of a lower-level  $S_{i-1}$ . Then the input  $B_i$  is into multiple non-overlapping patches as labeled by different colors in Fig. 3. The output of both encoder and decoder from a lower level will be added to the upper level so that the top-level can gather all information inferred in finer levels. Since two phases of our network can run parallelly, it takes about 30ms when processing 720p images in both two phases, which can satisfy real-time applications.

### IV. EXPERIMENT

#### A. DATASETS

##### 1) SIMULATED EVENT SEQUENCE

Phase 1 requires training data in the form of event sequences with corresponding ground-truth image sequences. We use the E2VID [16] dataset, which consists of 1000 sequences per 2 seconds.

##### 2) SYNTHETIC DATASET

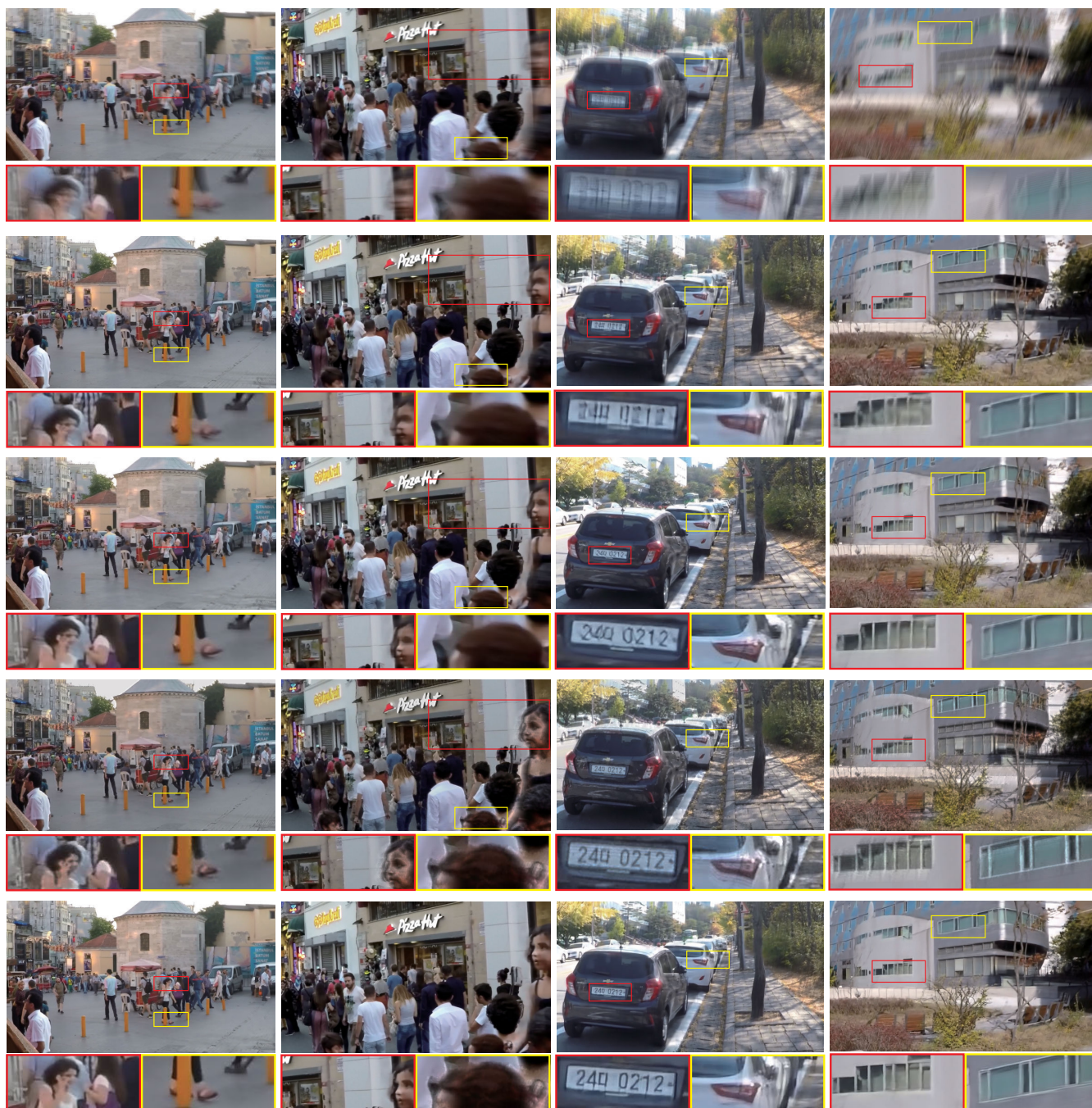
In order to quantitatively compare our experiment results, we use the popular GoPro blurry dataset [12], which consists of 3214 pairs of blurred and sharp images captured at  $720 \times 1280$  resolution. To get the simulated event data, we employ the ground-truth images to generate simulated event data based on ESIM [15].

##### 3) REAL DATASET

We evaluate our method on the real blurry event dataset [14], captured by DAVIS [3] under different conditions, such as indoor, outdoor, lowlighting conditions, and different motion patterns such as camera shake, object motion.

#### B. TRAINING DETAILS

We use a phase-to-phase training strategy, training phase 1 alone with the simulated event sequence. Then we fixed phase 1 to train phase 2 model. For the first stage, we split the simulated event sequences into 950 training sequences



**FIGURE 6.** Visual comparisons on GoPro dataset. In the top-down order, we show input blurry images, results of Tao et al. [20], Zhang et al. [25], Pan et al. [14], and Our results(best view).

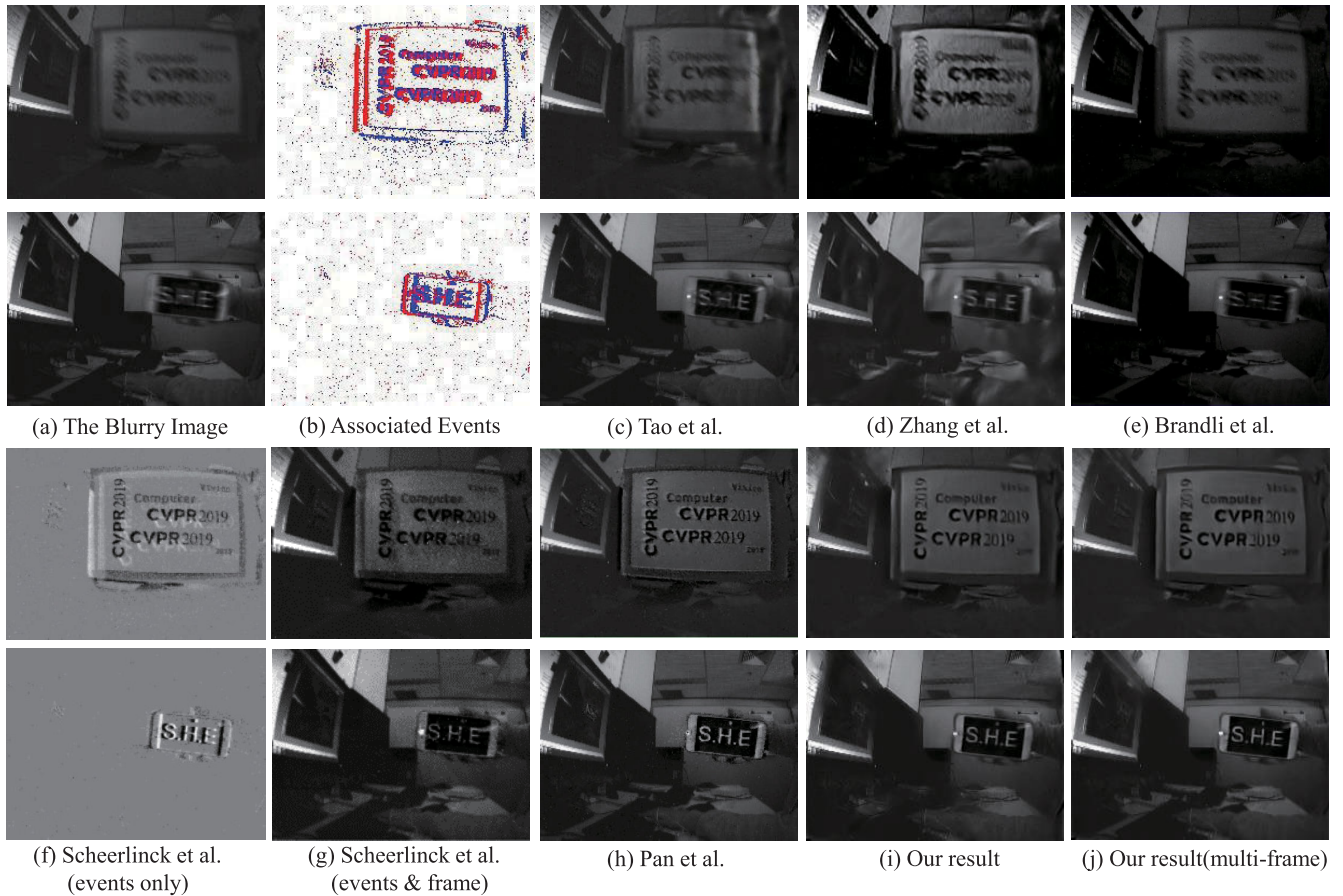
and 50 validation sequences follow the experience of [16], [30]. We augment the training data using random 2D rotations in the range of  $\pm 20$  degrees, horizontal and vertical flips, and random cropping with a crop size of  $128 \times 128$ . For the second phase, we take the synthetic dataset as our training dataset, with 22 sequences for training and the remaining 11 sequences for testing.

All our experiments are implemented in Pytorch and evaluated on a single NVIDIA Tesla P100 GPU. During the training of phase 2, we randomly crop images to  $256 \times 256$  pixel size and forward the cropped images to the

inputs of each level. The batch size is set to 6 during training. We use the Adam [8] as our optimizer set to 0.0001 and decay rate to 0.1 for 500 epoch, totally train our models for 2000 epochs. We normalize image to the range of [0, 1] and subtract 0.5. Besides, we use the mean-square error (MSE) loss at the output of level 1. Models' performance is measured by PSNR and Structural Similarity (SSIM).

### C. EXPERIMENT RESULTS

We compare our proposed network with state-of-the-art deblurring methods including conventional deblurring



**FIGURE 7.** Visual comparisons on real blurry event dataset [14] in low lighting and complex dynamic conditions. (a) Input blurry images. (b) Corresponding events information. (c) Deblurring results of [20]. (d) Deblurring result of [25]. (e) Reconstruction result of [29] with events and intensity frames. (f)Reconstruction result of [17]from only events. (g) Complementary filter reconstruction results of [17]. (h) Deblurring result of [14]. (i) Our deblurring result. (j) Our result with connecting consecutive 5 blurry images.

**TABLE 1.** Quantitative comparisons on the GoPro dataset 720p [12]. PSNR in dB, Runtime in ms, Model Size in MB. All models are tested under the same blurry condition and we use the pre-training models provided by these methods to conducting extensive testing. We train our model three times to get the average performance.

Models	PSNR	SSIM	Runtime	Model Size
Sun et al. [19]	25.30	0.8511	12000	54.1
Gong et al. [6]	26.05	0.8632	9600	39.2
Jin et al. [7]	26.98	0.8922	1100	57.3
Tao et al. [20]	30.26	0.9342	1600	33.6
Nah et al. [12]	29.08	0.9135	4300	303.6
Pan et al. [14]	29.06	0.9430	3220	<b>0.51</b>
Zhang(1-2-4) [25]	30.21	0.9023	<b>17</b>	21.7
Stack2-VMPHN [25]	31.5	<b>0.9483</b>	552	86.8
Ours	<b>32.25</b>	0.9285	31	57.8

methods [14], learning-based deblurring methods [7], [12], [20], [25] and event-based image reconstruction method [17], [29].

Evaluation on synthetic dataset is shown in Fig. 6 indicates that our deblurred images have the best view which is sharp and photorealistic. Qualitative comparisons in Tab.1 shows our proposed model achieves a significant performance promotion in terms of PSNR. We test these models in the same experimental environment several times.

The evaluation of the real dataset is shown in Fig. 7. The state-of-the-art deep deblur models (c)(d) fail to restore

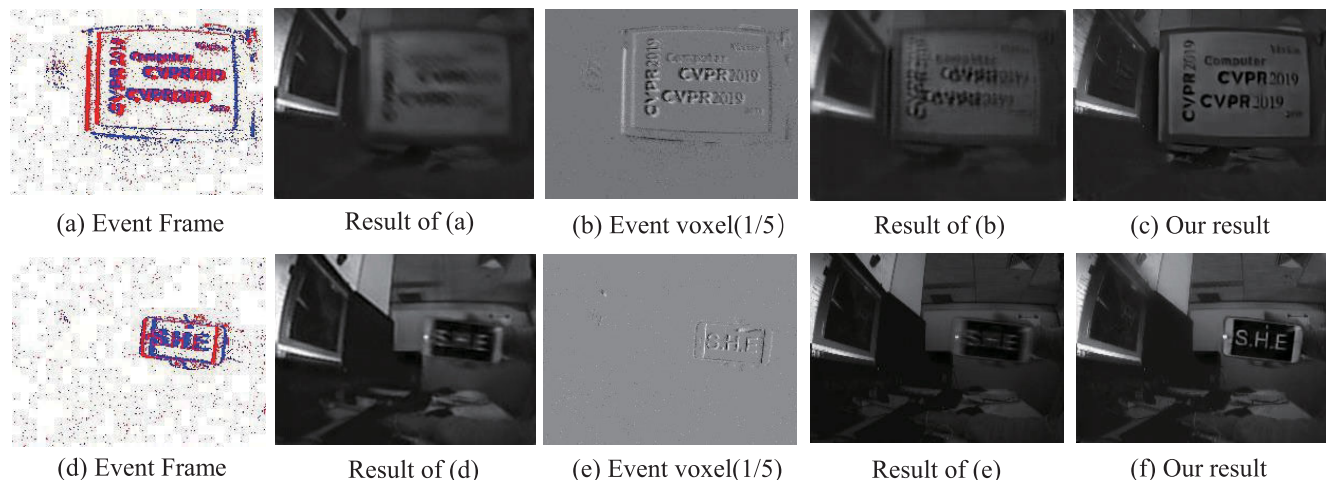
the sharp images due to the lack of time information caused by severed motion. (f) Reconstruction with only events lost backgrounds and has artifacts. (e)(g) Reconstruction methods use events and intensity frames still suffer from artifacts and noise. Reference [14] first proposed an EDI deblur model that uses both events and intensity frames. (h) A clear image is restored by EDI but it suffers from noise and is non-photorealistic. Our method first uses a deep deblur model that uses the recurrent representation as supplementary information and successfully restores a sharp image (i). As there are some artifacts in a small area of background, we then concatenated 5 consecutive blurry images with event recurrent representation to get more static information (j).

## V. ABLATION STUDIES

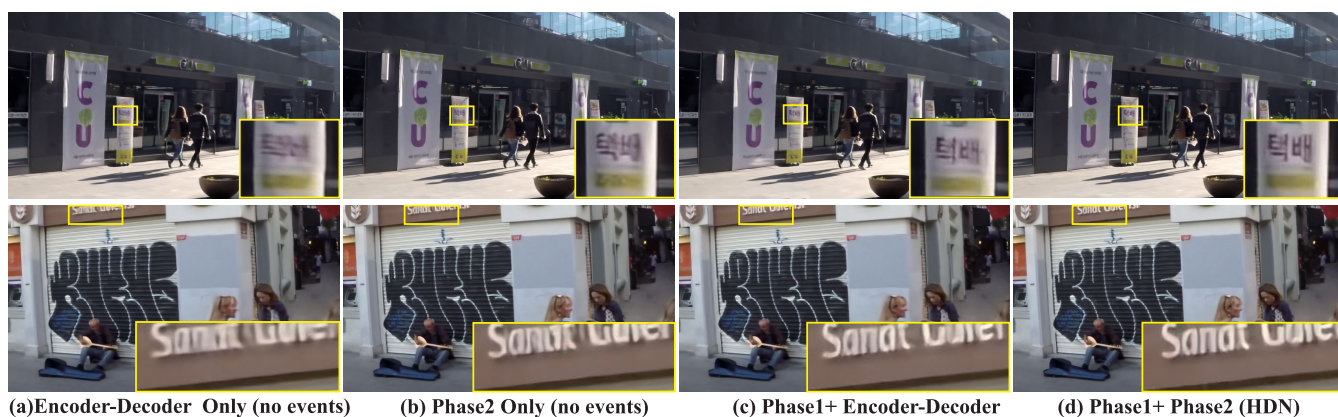
We now present ablation studies to discuss the contribution of each phase in our two-phase model.

### A. PHASE 1

We perform an ablation study to prove that our used recurrent event representation can provide more details and information than directly accumulating events or using the event voxel. (a)(d) We concatenate directly integrate event frames with a blurry image as input. (b)(e) We concatenate the commonly used event voxel [27]. As shown in Fig. 8, the deblur



**FIGURE 8.** Ablation study on event representations and results. (a,d) Direct integration. (b,d) one of five voxel. (c,f) Our results with recurrent event representation.



**FIGURE 9.** Visual comparisons on GoPro dataset. In the top-down order, we show input blurry images, results of Tao et al. [20], Zhang et al. [25], Pan et al. [14], and Our results(best view).

**TABLE 2.** Quantitative comparisons on the GoPro dataset 720p [12] of ablation study. PSNR in dB, Runtime in ms, Model Size in MB. We train each model three times and calculate the average performance.

Model Name	PSNR	SSIM	Runtime	Model Size
HDN	$32.25 \pm 0.01$	$0.9285 \pm 0.0001$	31	57.8
Baseline-A	$27.97 \pm 0.03$	$0.8488 \pm 0.0005$	7	6.9
Baseline-B	$30.21 \pm 0.03$	$0.9023 \pm 0.0005$	17	21.7
Baseline-C	$30.59 \pm 0.01$	$0.9071 \pm 0.0003$	31	47.6

results of those baselines are inferior to ours (c)(f). The result without any event information can refer to Fig 7.(d) and the result without the use of the blurry image can refer to Fig.7(f).

## B. PHASE 2

To measure the impact of our multi-patch hierarchical deblur model in phase 2. We propose three baseline networks and train these models under the same experimental conditions three times. Tab.2 shows the performance of these models on the GoPro dataset. Baseline-A uses a simple UNet-like encoder-decoder model which is commonly used in image restoring. Baseline-B is our phase 2 model without event information. Baseline-C uses the phase 1 of our model and

the encoder-decoder model. The visual comparison of these models on GoPro dataset is shown in Fig.9.

As shown in Fig.9 (c) and (d), our multi-patch hierarchical deblur model in phase 2 increasing performance a lot compared with the traditional encoder-decoder model. Without the event representation, the performance of Baseline-B still exceeds Baseline-A a lot. Besides, from the comparison results of Fig.9 (a)(c) and (b)(d), the use of event representation can significantly improve the performance of deblurring.

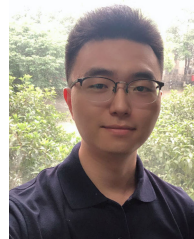
## VI. CONCLUSION

In this paper, we propose a two-phase hybrid model to restore blurry images with an event camera. We formulate the deblurring task as residual learning in our pipeline. Experiment results prove that use the recurrent event representation of our phase 1 output can provide a latent image-like supplementary information which is beneficial to deblur task. Phase 2 leverages a multi-patch hierarchical architecture to effectively fuse cues of blurs in local regions by levels. The experiments on both synthetic and real datasets show that our model

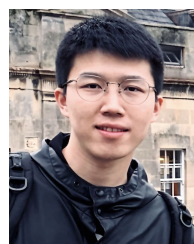
can restore sharp images from non-uniform blurry images. Due to the low resolution of event camera, we consider using event-camera to assist a high-resolution RGB camera to deblur in the future.

## REFERENCES

- [1] P. Bardow, A. J. Davison, and S. Leutenegger, "Simultaneous optical flow and intensity estimation from an event camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 884–892.
- [2] S. Barua, Y. Miyatani, and A. Veeraraghavan, "Direct face detection and video reconstruction from event cameras," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Lake Placid, NY, USA, Mar. 2016, pp. 1–9.
- [3] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "A 240×180 130 dB 3  $\mu$ s latency global shutter spatiotemporal vision sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, Oct. 2014.
- [4] Y. Chen, W. Chen, X. Cao, and Q. Hua, "Intensity-image reconstruction for event cameras using convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 1973–1977.
- [5] S. Cho and S. Lee, "Fast motion deblurring," in *Proc. ACM SIGGRAPH Asia Papers SIGGRAPH Asia*, Yokohama, Japan, 2009, pp. 1–8.
- [6] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. Van Den Hengel, and Q. Shi, "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3806–3815.
- [7] M. Jin, G. Meishvili, and P. Favaro, "Learning to extract a video sequence from a single motion-blurred image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 6334–6342.
- [8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015.
- [9] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jun. 2006, pp. 2169–2178.
- [10] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15  $\mu$ s latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008.
- [11] C. Reinbacher, G. Graber, and T. Pock, "Real-time intensity-image reconstruction for event cameras using manifold regularisation," in *Proc. Brit. Mach. Vis. Conf.*, York, U.K., 2016, pp. 1381–1393.
- [12] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3883–3891.
- [13] L. Pan, R. Hartley, M. Liu, and Y. Dai, "Phase-only image based kernel estimation for single image blind deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 6027–6036.
- [14] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, "Bringing a blurry frame alive at high frame-rate with an event camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 6813–6822.
- [15] H. Rebecq, D. Gehrig, and D. Scaramuzza, "ESIM: An open event camera simulator," in *Proc. Conf. Robot. Learn. (CoRL)*, Zurich, Switzerland, 2018, pp. 969–982.
- [16] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 3852–3861.
- [17] C. Scheerlinck, N. Barnes, and R. Mahony, "Continuous-time intensity estimation using event cameras," in *Proc. Asian Conf. Comput. Vis.*, in Lecture Notes in Computer Science, Perth, WA, Australia, 2018, pp. 308–324.
- [18] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Scholkopf, "Learning to deblur," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1439–1451, Jul. 2016.
- [19] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 769–777.
- [20] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 8174–8182.
- [21] L. Wang, I. S. M. Mostafavi, Y.-S. Ho, and K.-J. Yoon, "Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 10073–10082.
- [22] L. Xiao, J. Wang, W. Heidrich, and M. Hirsch, "Learning high-order filters for efficient blind deconvolution of document photographs," in *Proc. ECCV in Lecture Notes in Computer Science*, Amsterdam, The Netherlands, 2016, pp. 734–749.
- [23] S. Xingjian, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst. (ECCV)*, Montreal, QC, Canada, 2015, pp. 802–810.
- [24] L. Xu and J. Jia, "L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," *Proc. ECCV*, in Lecture Notes in Computer Science, Heraklion, Greece, 2010, pp. 157–170.
- [25] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 5971–5979.
- [26] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. H. Lau, and M.-H. Yang, "Dynamic scene deblurring using spatially variant recurrent neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2521–2529.
- [27] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 989–997.
- [28] Z. A. Zihao, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based optical flow using motion compensation," in *Proc. ECCV*, in Lecture Notes in Computer Science, Munich, Germany, 2018, pp. 711–714.
- [29] C. Brandli, L. Müller, and T. Delbruck, "Real-time, high-speed video decompression using a frame- and event-based DAVIS sensor," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Melbourne, VIC, Australia, Jun. 2014, pp. 686–689.
- [30] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Dec. 31, 2019, doi: 10.1109/TPAMI.2019.2963386.



**LIMENG ZHANG** received the B.S. degree from the National University of Defense Technology, in 2017, where he is currently pursuing the master's degree with the College of Computer Science and Technology. His research interests include event-based vision and image deblurring.



**HONGGUANG ZHANG** received the B.Sc. degree in electrical engineering and automation from Shanghai Jiao Tong University, Shanghai, China, in 2014, the M.Sc. degree in electronic science and technology from the National University of Defense Technology, Changsha, China, in 2016, and the Ph.D. degree in computer vision and machine learning from The Australian National University and Data61/CSIRO, Canberra, ACT, Australia, in 2020. His research interests include

fine-grained image classification, zero-shot learning, few-shot learning, and deep learning methods.





**JIHUA CHEN** received the B.S. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, and the M.S. degree in computer application from the National University of Defense Technology, Changsha, China. In 1986, he joined the School of Computer, National University of Defense Technology, where he was responsible for CAD and IC physical design and currently a Professor. His current research interest includes IC design and test.



**LEI WANG** (Member, IEEE) received the B.S. and Ph.D. degrees from the National University of Defense Technology, in 2000 and 2006, respectively. She is currently an Associate Professor with the College of Computer Science and Technology, National University of Defense Technology. Her current research interests include computer architecture, asynchronous circuit, artificial intelligence, and neuromorphic computation.

...