Boosting Event Stream Super-Resolution with A Recurrent Neural Network §Supplementary Material§

Wenming Weng, Yueyi Zhang(✉), and Zhiwei Xiong

University of Science and Technology of China, Hefei, China
wmweng@mail.ustc.edu.cn, {zhyuey, zwxiong}@ustc.edu.cn

1 Additional Visual Results

We provide additional visual results of event SR and video reconstruction among bicubic, SRFBN [4], EventZoom [1], EventZoom-cr and our method. For large factor SR results of SRFBN and EventZoom, we run 2×-model multiple times due to high cost of training a single large factor SR model, which is also adopted in [1]. The 16× SR results of EventZoom-cr are omitted because we cannot train a 16×-model of EventZoom-cr.

On real-world dataset. We present more real-world results on ENFS-real [1] in Fig. 1. The size of real LR events is 56×31, which is severely degraded. For all methods, we use the pre-trained model on synthetic dataset (ENFS-syn) to super-resolve real LR events. It can be clearly observed that we can achieve satisfactory results with fine details and sharp edges, demonstrating strong ability of generalization and robustness when applied for real sensors.

On synthetic dataset. We present more synthetic results on ENFS-syn and RGB-DAVIS-syn in Fig. 2 and Fig. 3. The sizes of LR events on ENFS-syn and RGB-DAVIS-syn are 80×45 and 190×180, respectively.

Video reconstruction. We present more visual results for event-based video reconstruction on ENFS-syn in Fig. 4. We show LR/ground-truth events and corresponding frames for better comparison. We also convert ground-truth events to video, named "GT events recons", which is upper bound. Compared with baselines, our method is capable of producing relatively good reconstruction results. To the best of our knowledge, we benchmark the methods of event-to-frame with large factor (8×) for the first time.

2 Video Demonstration for Event SR and Video Reconstruction

We also provide a video demonstration for event SR and video reconstruction. We strongly recommend referring to it for intuitive comparisons of bicubic, SRFBN [4], EventZoom [1], EventZoom-cr and our method.

Table 1. The comparisons of parameters and inference time among EventZoom [1], SRFBN [4] and ours. Best in bold.

Methods	#Param	inference time (ms, LR size: 45x60)			
		2×	4×	8×	16×
EventZoom	11.5M	178.5	192.5	227.2	326.9
SRFBN	2.1M	179.6	215.3	289.4	428.6
RecEvSR (Ours)	**1.8M**	**173.8**	**184.6**	**191.1**	**207.8**

3 Model Performance Analysis

We add the comparisons of parameters and inference time in Tab. 1. The experiments are conducted using an NVIDIA GTX 3090 GPU. As can be seen, our model RecEvSR with 1.8M achieves the smallest parameters, while EventZoom [1] with 11.5M is five times larger. In terms of inference time, our RecEvSR is more efficient than other two methods, which is more portable for deployment in the mobile devices.

4 Implementation Details

4.1 Training Settings

For training SRFBN [4], EventZoom [1], EventZoom-cr and our RecEvSR, we use ADAM as the optimizer, and set initial learning rate as 10^{-3}, which is decayed by a factor of 0.95 every 4000 iterations. We train all methods for 100000 iterations and set batch size as 2. We also apply the early-stop strategy for all methods. The horizontal, vertical and polarity flips are performed with probability of 0.5 for data augmentation. We implement our network using Pytorch 1.8 and run all experiments on an NVIDIA GTX 3090 GPU.

4.2 Details of Our Synthetic Datasets

We generate two synthetic datasets for conducting experiments. For ENFS-syn, we first down-sample the original frames from NFS [3] using bicubic interpolation to form multi-scale frames. We then convert the generated multi-scale frames into events using event simulator [2], producing multi-scale events. For event simulator, we adopt the default initial settings as in [2], except that the positive contrast thresholds ($\mathrm{C_p}$) and negative contrast thresholds ($\mathrm{C_n}$) are restricted to the limitation: $\mathrm{C_p} = \mathrm{C_n} \times x, x \in \mathcal{N}(\mu = 1.0, \sigma = 0.1)$. The same synthesizing process is employed for producing RGB-DAVIS-syn.


LR
Frame
2x
4x
8x
16x
N/A
bicubic
SRFBN
EventZoom
EventZoom-cr
Ours
LR
Frame
2x
4x
8x
16x
N/A
bicubic
SRFBN
EventZoom
EventZoom-cr
Ours
LR
Frame
2x
4x
8x
16x
N/A
bicubic
SRFBN
EventZoom
EventZoom-cr
Ours

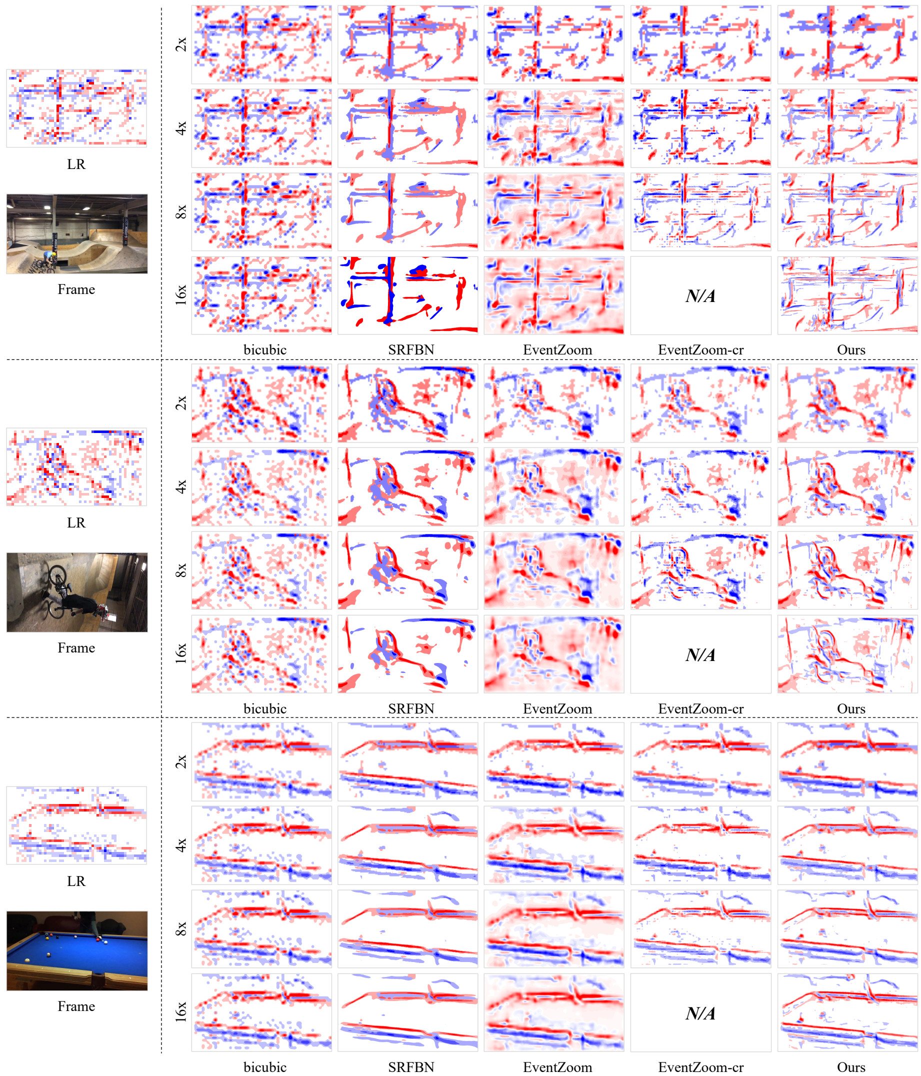


Fig. 1. Additional real-world visual results on ENFS-real [1] among bicubic, SRFBN [4], EventZoom [1], EventZoom-cr and ours for event SR.

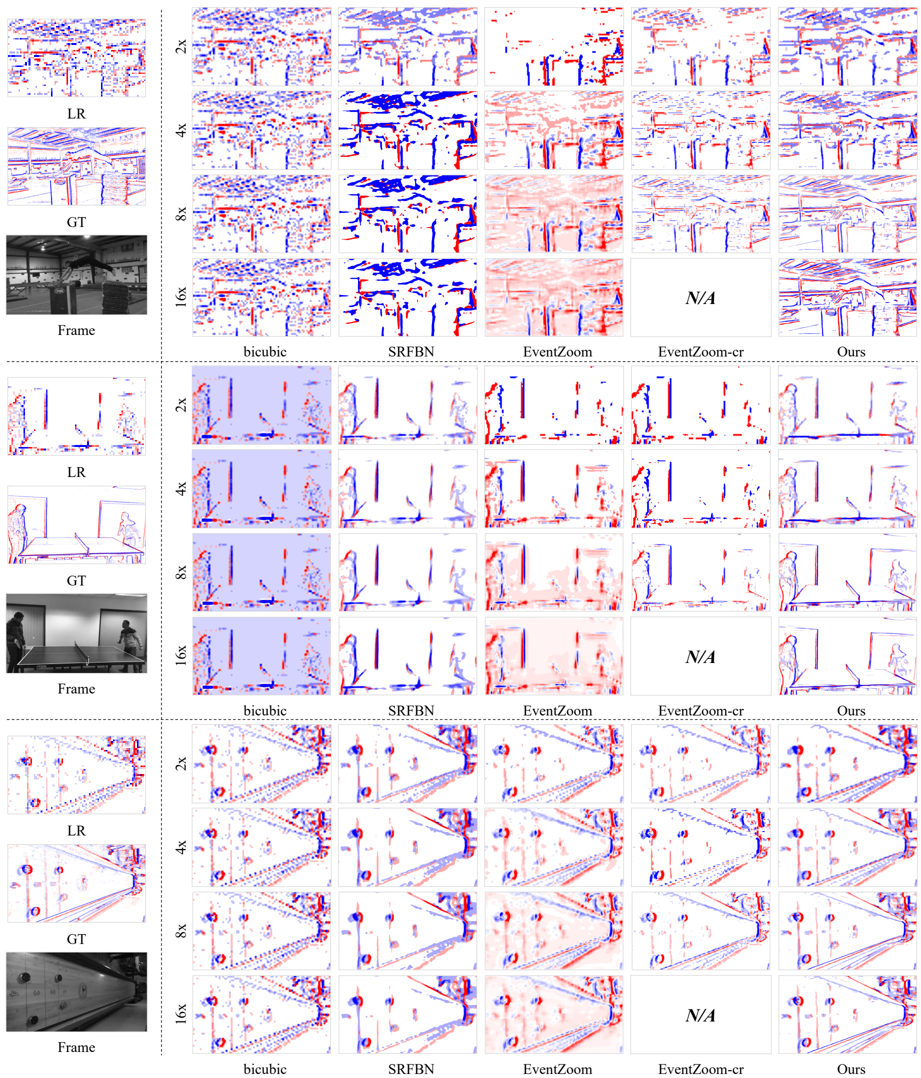


Fig. 2. Additional visual results on ENFS-syn among bicubic, SRFBN [4], EventZoom [1], EventZoom-cr and ours for event SR.

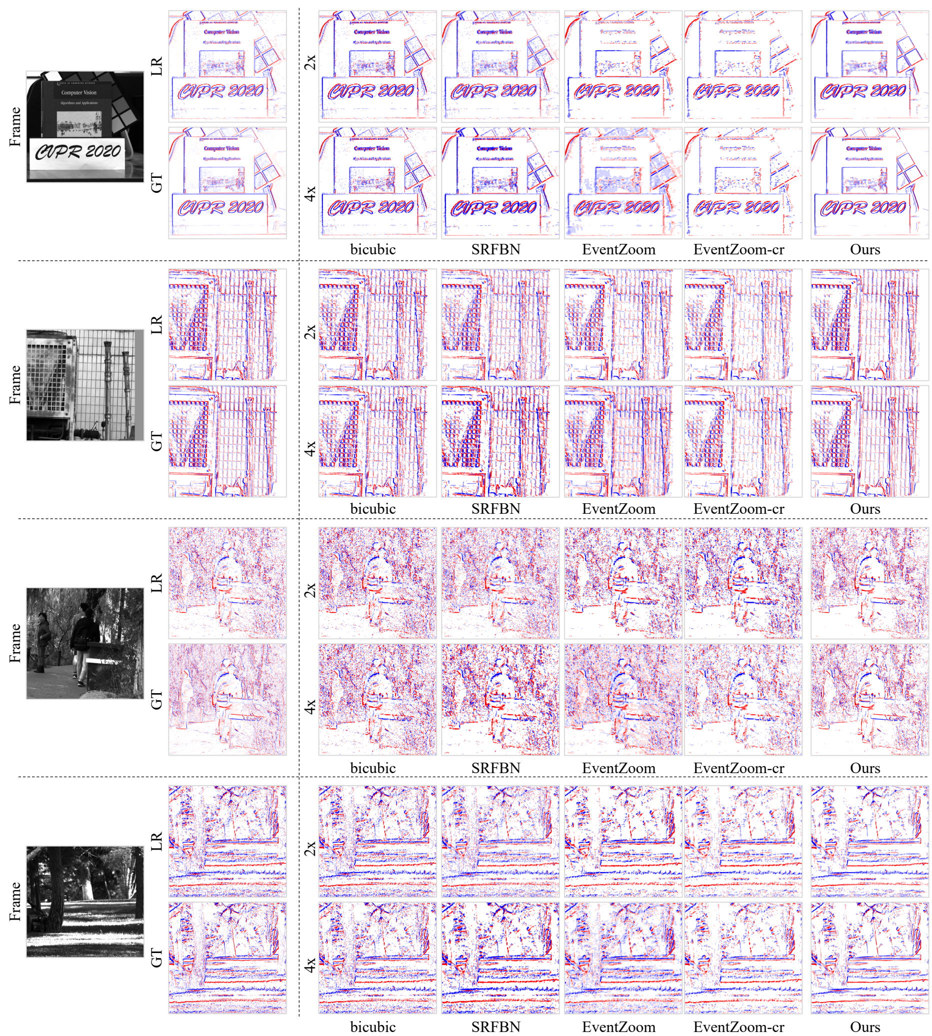
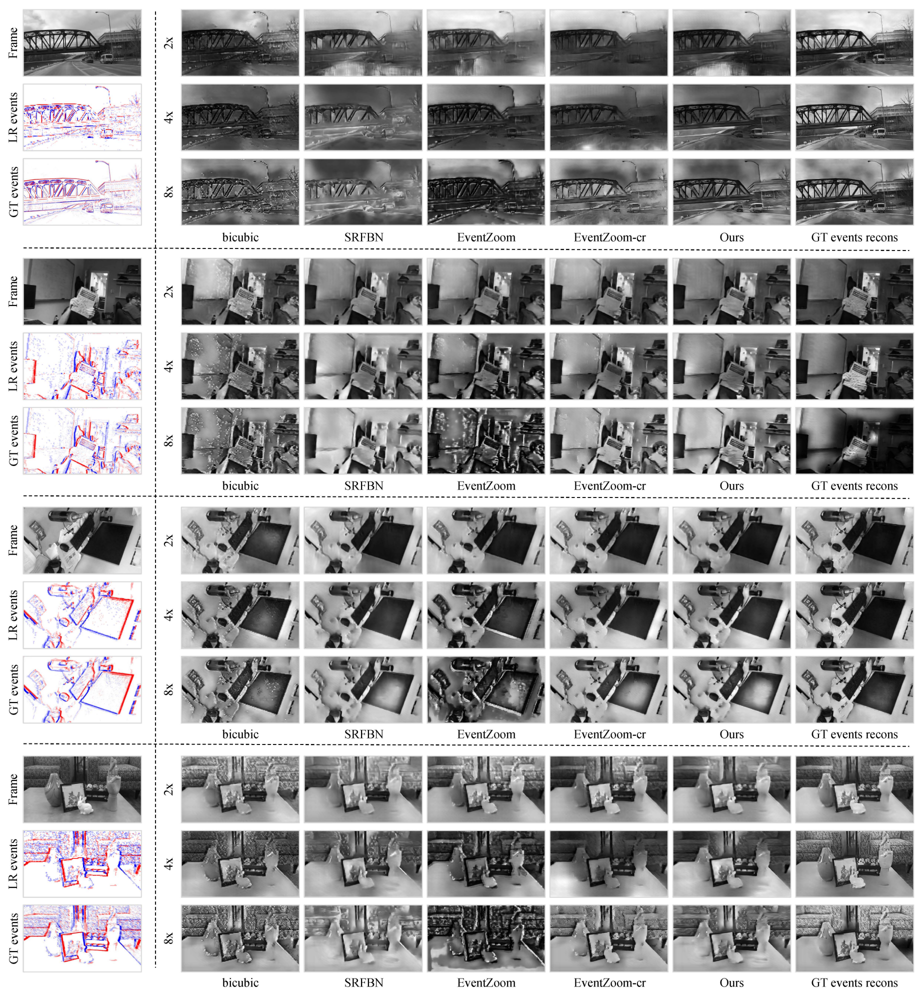


Fig. 3. Additional visual results on RGB-DAVIS-syn among bicubic, SRFBN [4], EventZoom [1], EventZoom-cr and ours for event SR.


Frame
LR events
GT events
2x
4x
8x
bicubic
SRFBN
EventZoom
EventZoom-cr
Ours
GT events recons

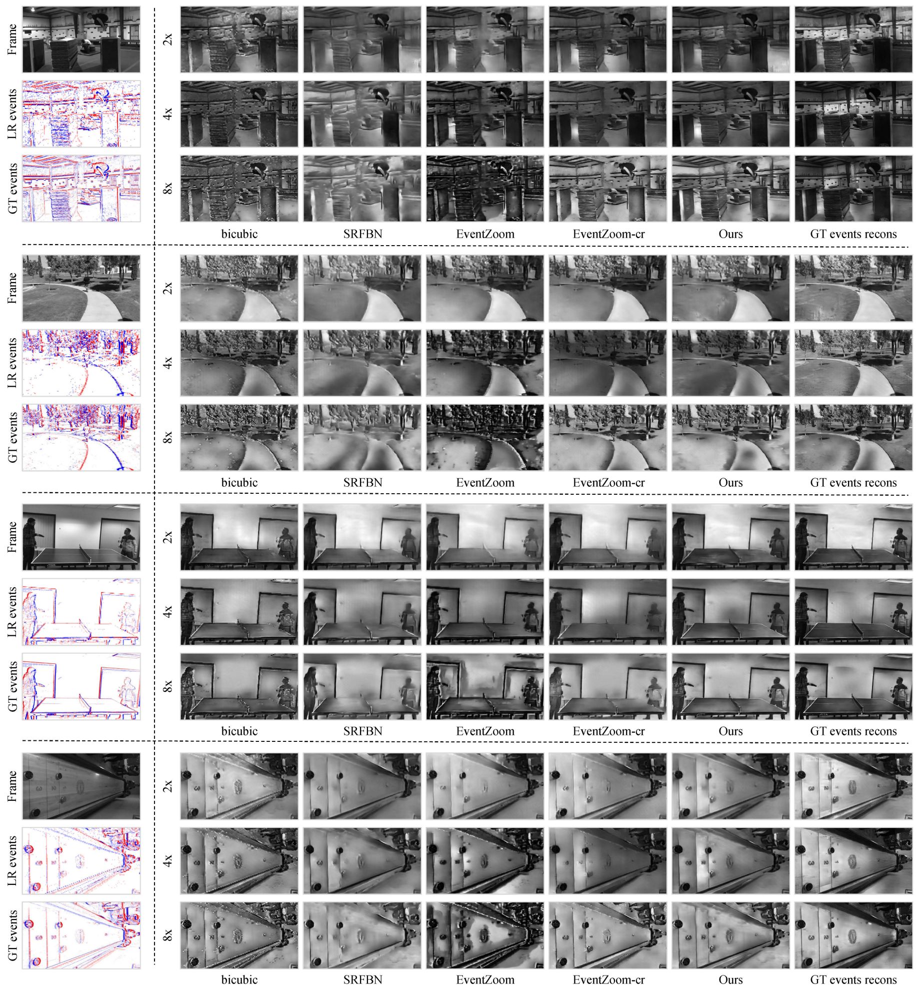


Fig. 4. Additional visual results on ENFS-syn among bicubic, SRFBN [4], EventZoom [1], EventZoom-cr and ours for video reconstruction.

References

1. Duan, P., Wang, Z.W., Zhou, X., Ma, Y., Shi, B.: Eventzoom: Learning to denoise and super resolve neuromorphic events. In: CVPR (2021) 1, 2, 4, 5, 6, 8
2. Gehrig, D., Gehrig, M., Hidalgo-Carrió, J., Scaramuzza, D.: Video to events: Recycling video datasets for event cameras. In: CVPR (2020) 2
3. Kiani Galoogahi, H., Fagg, A., Huang, C., Ramanan, D., Lucey, S.: Need for speed: A benchmark for higher frame rate object tracking. In: ICCV (2017) 2
4. Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., Wu, W.: Feedback network for image super-resolution. In: CVPR (2019) 1, 2, 4, 5, 6, 8