Supplemental Material: NEST: Neural Event Stack for Event-based Image Enhancement

Minggui Teng[1], Chu Zhou[2], Hanyue Lou[1], and Boxin Shi[1,3,4]✉

[1] NERCVT, School of Computer Science, Peking University
[2] KLMP (MOE), School of Artificial Intelligence, Peking University
[3] Institute for Artificial Intelligence, Peking University
[4] Beijing Academy of Artificial Intelligence

6 Data Preparation

Our training and testing datasets are adopted from Wang *et al.* [12]. As their datasets only contain the gray-scale images, we regenerate RGB blurry images and LR images from the original REDS dataset [8] as Wang *et al.* [12] suggested, which is shown in Figure 7. We first down-sample image resolution from 720×1280 to 180×320, and add some Gaussian noise to obtain the LR images. To avoid unnatural artifacts in the blurry images, we increase the frame rate of LR images by a video frame interpolation method [9], and then we generate blurry images by averaging 17 continuous LR images. The events are simulated by an event simulator ESIM [11]. It is worth mentioning that we conduct deblurring and super-resolution operations on luminance ($\mathbf{Y}$) channel of input images, and then concatenate chrominance ($\mathbf{U}$ and $\mathbf{V}$) channels to restore color images.

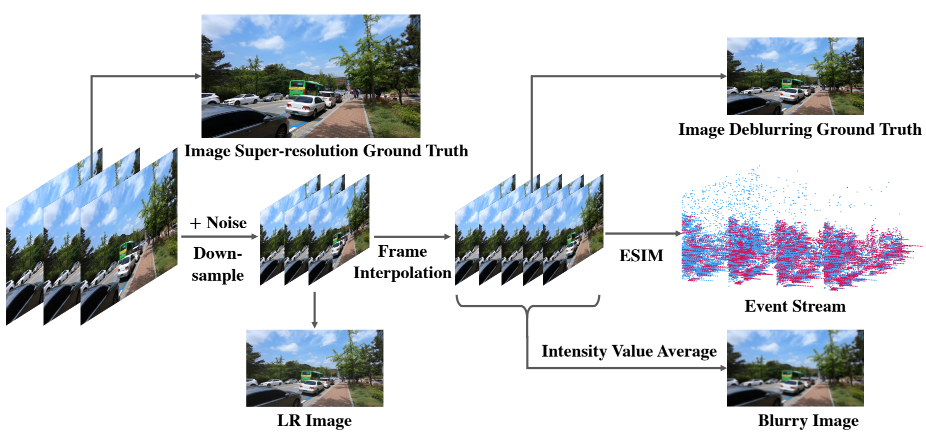


Fig. 7. The overview of our data preparation pipeline.

✉ Corresponding author: `shiboxin@pku.edu.cn`
Project page: `https://github.com/ChipsAhoyM/NEST`

7 Comparison with EDI Model

In this section, we analyze the key difference of our method from the Event-based Double Integral (EDI) model [10] in detail. The EDI model [10] treats the event threshold as a constant, which brings the error when calculating the event integral. As shown by Equation (5) in Pan *et al.* [10] , the EDI model is:

$$\mathbf{B} = \left(\frac{\mathbf{L}(f)}{T} \int_{f-\frac{T}{2}}^{f+\frac{T}{2}} \exp\left(\mathbf{E}(t)\right) \mathrm{d}t \right), \tag{14}$$

where $\mathbf{B}$ denotes the blurry image, $\mathbf{L}(f)$ denotes the latent image in time f, and $\mathbf{E}(t)$ denotes the sum of events between time f and t, *i.e.*,

$$\mathbf{E}(t) = c \int_{f}^{t} e(s)\mathrm{d}s. \tag{15}$$

The exposure time of blurry image is $[f - \frac{T}{2}, f + \frac{T}{2}]$. In Equation (15), $e(s)$ is the delta function, with unit integral, at time s. As the green line shown in Figure 8, when calculating the forward procedure ($f \leq t$) with different thresholds for positive and negative events, Equation (15) can process it correctly.

However, based on the integral operation, when $f > t$, $\mathbf{E}(t)$ should be calculated by changing the polarity of events, *i.e.*,

$$\mathbf{E}(t) = -c \int_{t}^{f} e(s)\mathrm{d}s. \tag{16}$$

Since the event positive and negative thresholds are unequal, when $f > t$, calculating the $\mathbf{E}(t)$ by Equation (16) leads to the incorrect value **4.5**, causing inconsistency between the ground truth **3.0**, as shown in the orange line of Figure 8. Therefore, we propose the bidirectional event summation to handle this issue, shown in Equation (3) of the main paper.

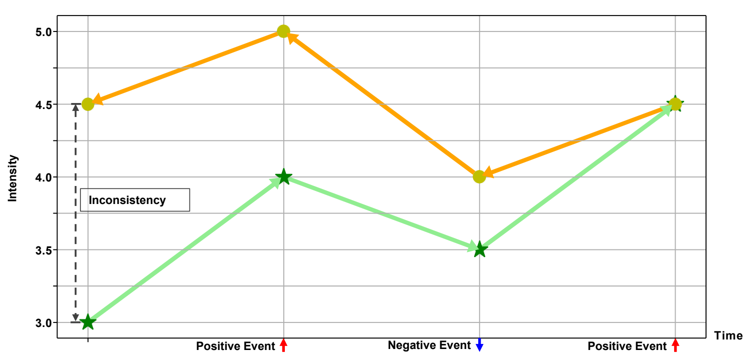


Fig. 8. A toy example shows inconsistency caused by simply reversing the event polarity. The forward calculation illustrates by the green line while the backward by the orange line.

8 Comparison between NESTs training from SR/Deblurring

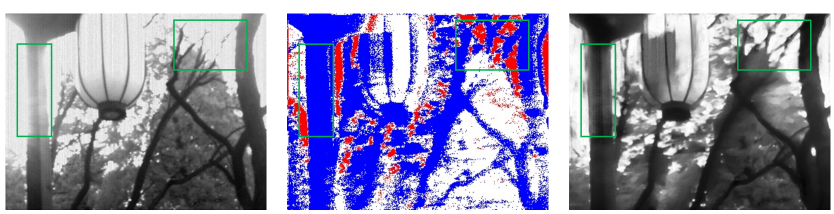
As Equation (9) and Equation (10) show, the image deblurring and SR both need the bi-directional summation of events. The difference between NESTs from different tasks is small, and it only affects the quality of restored images slightly. To verify this, we conduct an experiment with different NESTs generating strategies. We try encoding events once, and then apply them to image deblurring and SR sequentially (named "A"); we also try restoring the sharp image and HR images with corresponding NESTs separately (named "B"). As the Table 4 shows, NESTs trained from different tasks perform similarly.

Table 4. Quantitative evaluation between different NESTs encoding strategy.

Experiment	PSNR ↑	SSIM ↑	LPIPS ↓
A	26.32	0.6899	0.3869
B	26.51	0.7032	0.3570

9 Additional Application: High Dynamic Range

Since event cameras also demonstrate the high dynamic range property (130 dB for DAVIS240), the event data also benefits the over-exposed or under-exposed region recovery with more details. As shown in Figure 9, our method can fuse the over-exposed details hidden in the events to the intensity image. Due to the lack of HDR paired images in our training dataset, our method is not specialized to effectively handle the HDR issue from a single LDR image with corresponding events. Actually, we do not recover the HDR image, since the data range is still limited in 0 to 255. However, we demonstrate the possibility of restoring an HDR image from an LDR image using NEST representation in Figure 9, extending NEST with a well-designed HDR dataset.



(a) LDR image (b) Events (c) Results

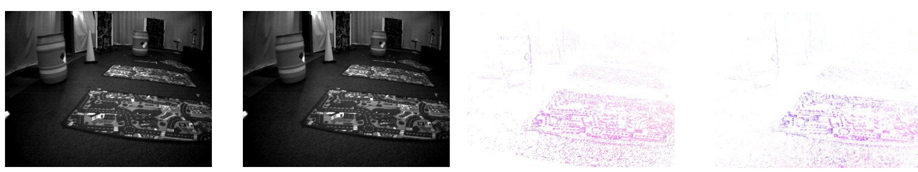
Fig. 9. An example of applying the NEST to HDR image recovery. (a) LDR image with over-exposed region. (b) The corresponding events. (c) The results of our method. As highlighted in green boxes, our method can recover some details in the over-exposed region.

10 Additional Application: Optical Flow Estimation

Since NESTs also contain global semantic information, we also try to apply NESTs to the optical flow estimation task (an example is shown in Figure 10). We use the EV-FlowNet [16] as the baseline model and replace its input with our NEST representation. We adopt the MVSEC dataset [16], train our model on the "outdoor day1" and "outdoor day2" sequences, and test on the "indoor flying1", "indoor flying2" and "indoor flying3" sequences. The quantitative comparison is shown in Table 5. As the results shown, our NEST representation can improve the performance of EV-FlowNet [16].

Table 5. Quantitative comparisons for optical flow estimation application on the MVSEC dataset.

	indoor flying1		indoor flying2		indoor flying3	
dt = 1 frame	AEE ↓	% Outlier ↓	AEE ↓	% Outlier ↓	AEE ↓	% Outlier ↓
EV-FlowNet [16]	1.03	2.2	1.72	15.1	1.53	11.9
Ours	**0.94**	**0.8**	**1.53**	**10.9**	**1.31**	**8.1**



(a) Previous image (b) Next image (c) Masked ground truth (d) Our results

Fig. 10. An example of applying the NEST to optical flow estimation. (a) The previous image. (b) The next image. (c) Masked ground truth. Since the events are sparse, we only visualize the optical flow where events have triggered. (c) The results of our method.

11 Comparison with state-of-the-art image-based method

To better illustrate the effectiveness of our NEST-based image enhancement method, we compare image-based deblurring/SR methods MPRNet [14] and BasicVSR++ [2]. Since input images are LR and noisy (generally with a different noise model from RGB cameras) limited by DAVIS346 cameras, MPRNet [14] and BasicVSR++ [2] cannot work well. Qualitative comparisons for deblur/SR applications on real data are shown below. Besides, we test MPRNet and BasicVSR++ on our synthetic testing data and show average quantitative metrics below the name of each method.[1]

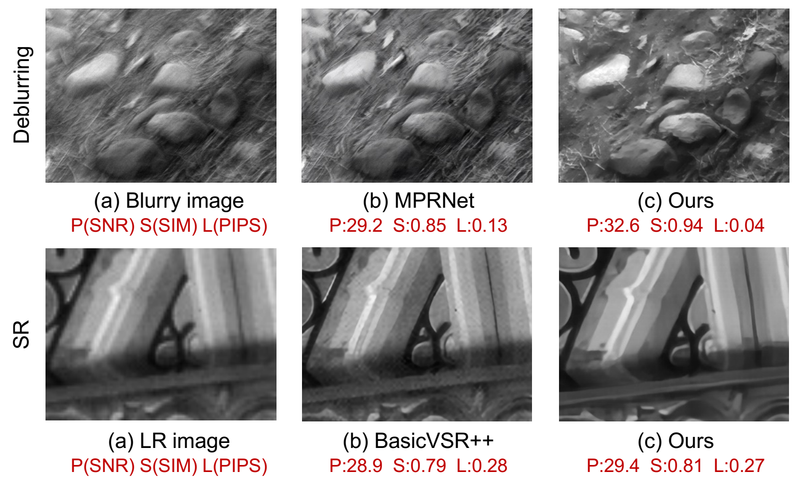


Fig. 11. Qualitative comparisons for deblurring/SR application on real data with state-of-the-art image-based methods. (a) Blurry image (upper)/ LR image (lower). (b) Results of MPRNet [14] (upper)/ BasicVSR++ [2] (lower). (c) Results of ours.

[1] Note these quantitative results are reported on the pre-trained model, and the results may get improved by retraining on images from event cameras.

12 Network Architecture in Detail

In this section, we present the architecture details of our proposed NEST-guided D-Net (shown in Figure 12) and S-Net (shown in Figure 13).

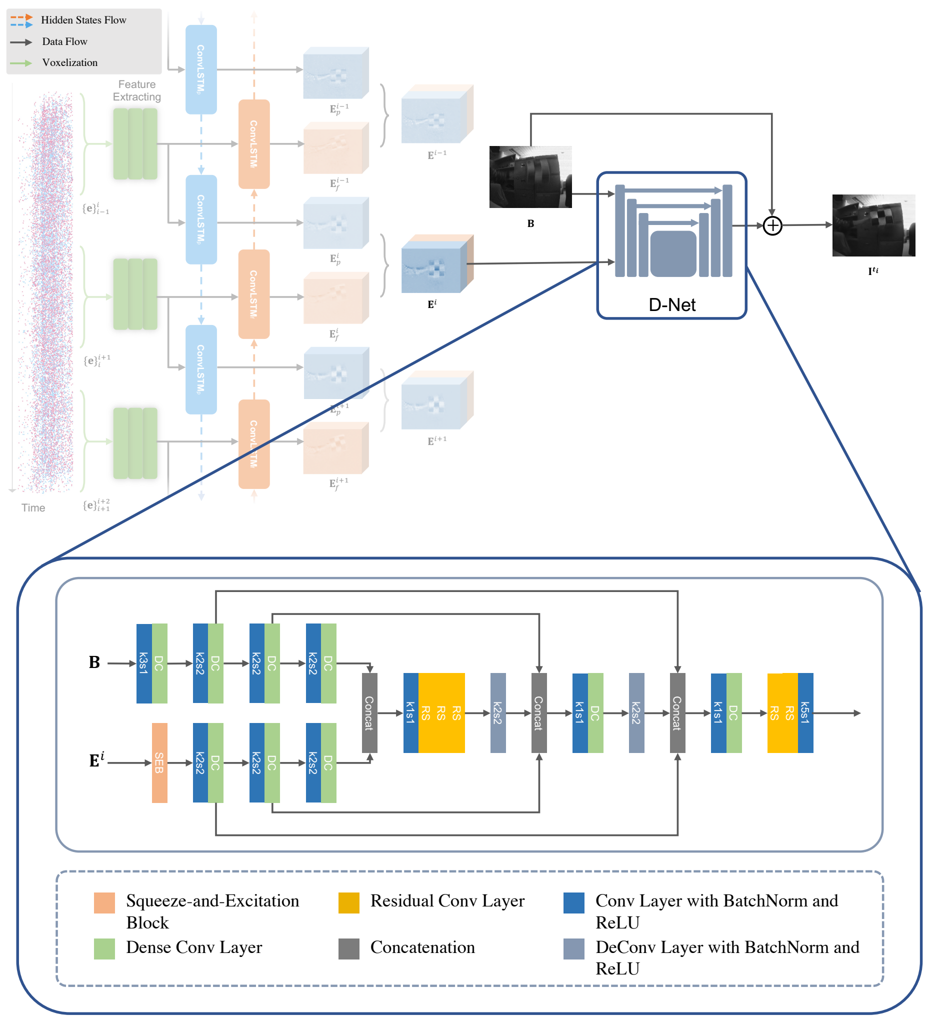


Fig. 12. The architecture of our NEST-guided D-Net in detail.

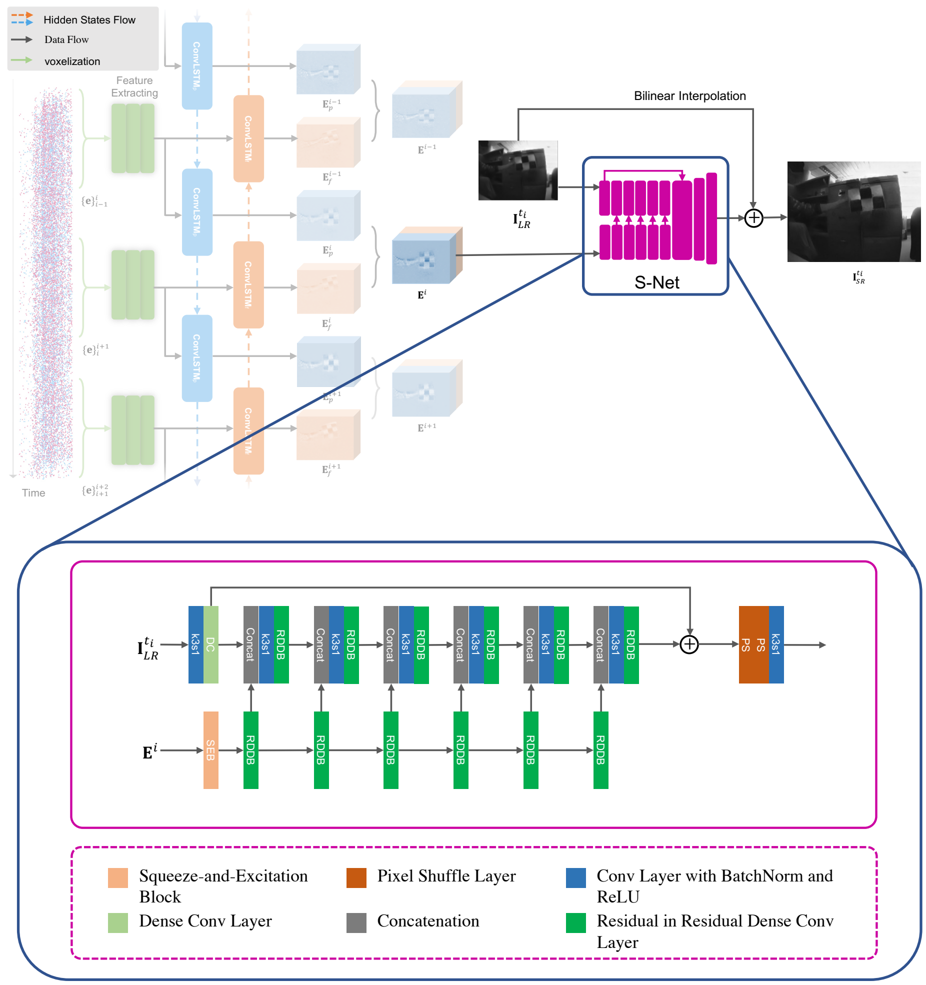


Fig. 13. The architecture of our NEST-guided S-Net in detail.

13 Comparison between eSL-Net and NEST+eSL

In this section, we provide qualitative comparisons between eSL-Net [12] and NEST+eSL. Comparison on deblurring application is shown in Figure 14 (upper). Comparison on super-resolution application is shown in Figure 14 (lower).

Fig. 14. Qualitative comparisons for deblurring (upper) and super-resolution (lower) application on synthetic data. (a) Blurry image / LR image. (b) Result of eSL-Net [12]. (c) Result of NEST+eSL.

14 More Results on Event-based Image Deblurring Application

In this section, we provide more qualitative comparisons among our method, ESTRNN [15], EDI [10], LEDVDI [6], eSL-Net [12], EvStack [13], EST$^+$ [3] and MatrixLSTM$^+$ [1]. Comparisons on synthetic data are shown in Figure 15 and Figure 16. Comparisons on real data are shown in Figure 17 and Figure 18.

15 More Results on Event-based Image Super-resolution Application

In this section, we provide more qualitative comparisons among our method, SPSR [7], RBPN [5], EvIntSR [4], eSL-Net [12], EvStack [13], EST$^+$ [3] and MatrixLSTM$^+$ [1]. Comparisons on synthetic data are shown in Figure 19 and Figure 20. Comparisons on real data are shown in Figure 21 and Figure 22.

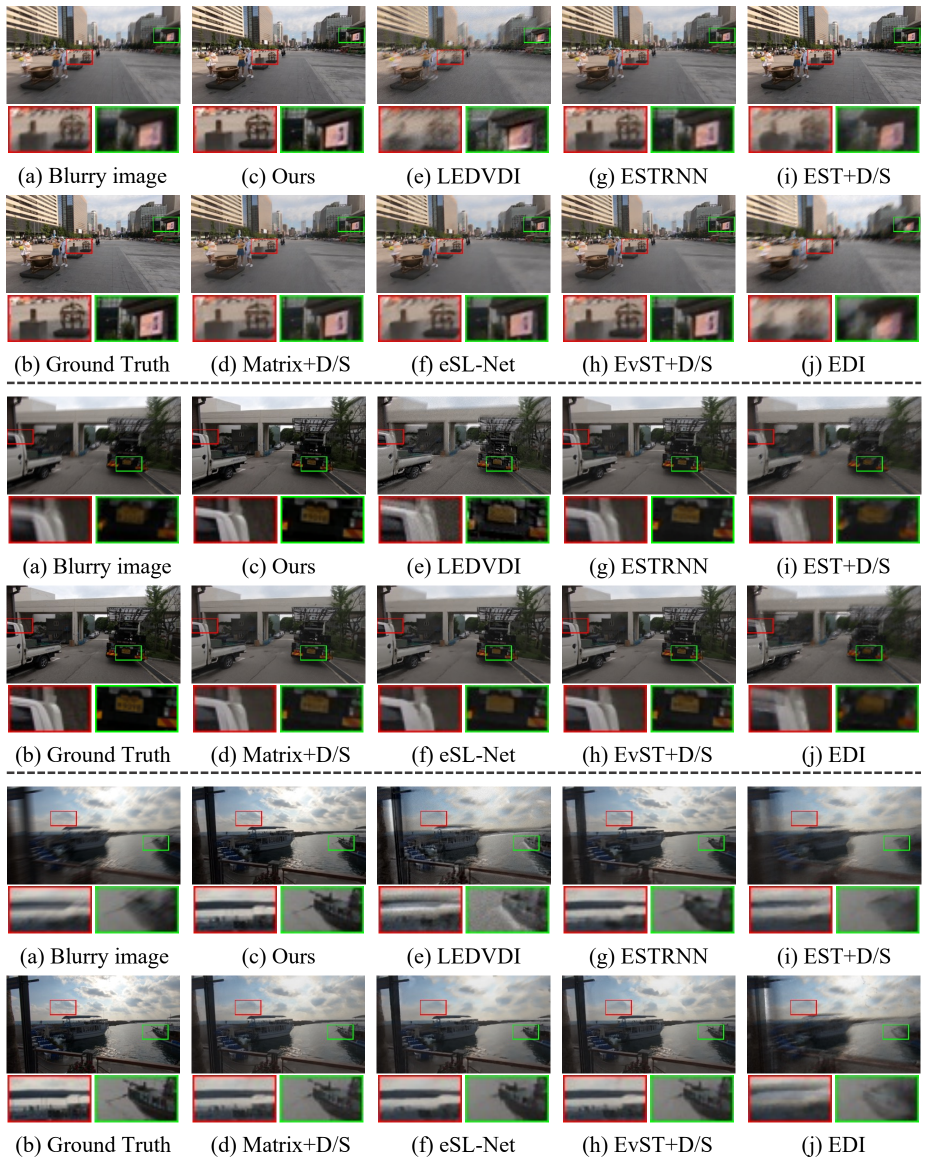


Fig. 15. Qualitative comparisons for deblurring application on synthetic data (part 1). (a) Blurry image. (b) Ground truth. (c)∼(j) Deblurring results of ours,Matrix+D/S [1], LEDVDI [6], eSL-Net [12], ESTRNN [15], EvST+D/S [13], EST+D/S [3], and EDI [10]. Close-up views are provided below each image.

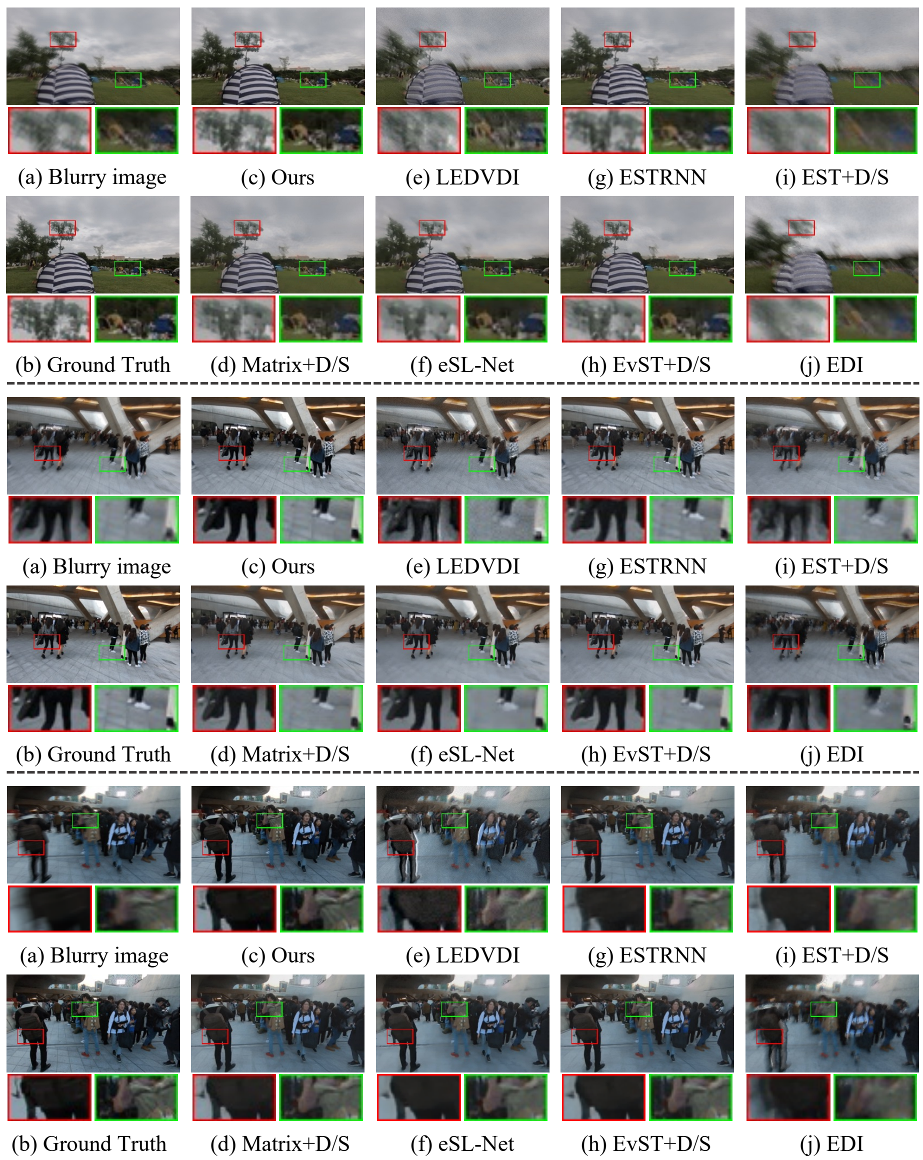


Fig. 16. Qualitative comparisons for deblurring application on synthetic data (part 2). (a) Blurry image. (b) Ground truth. (c)∼(j) Deblurring results of ours,Matrix+D/S [1], LEDVDI [6], eSL-Net [12], ESTRNN [15], EvST+D/S [13], EST+D/S [3], and EDI [10]. Close-up views are provided below each image.

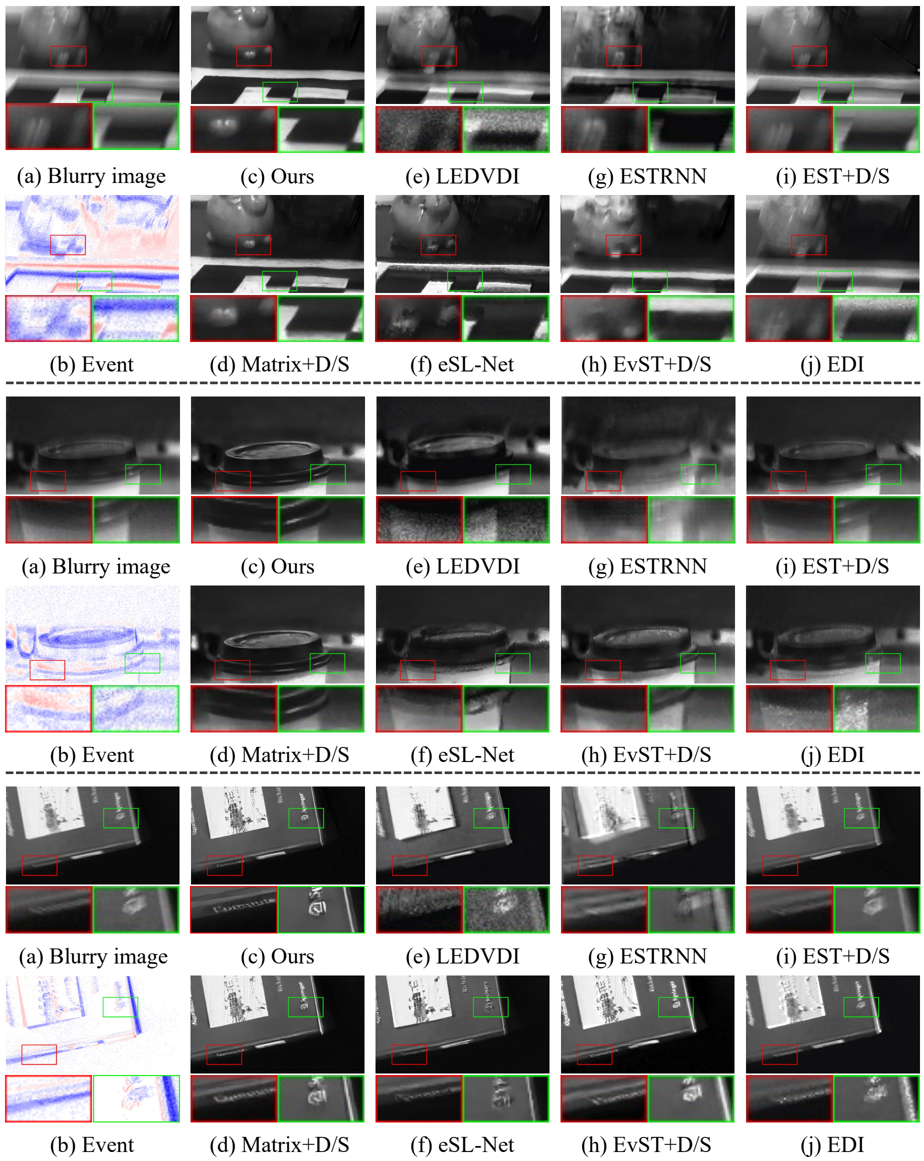


Fig. 17. Qualitative comparisons for deblurring application on real data (part 1). (a) Blurry image. (b) Event. (c)∼(j) Deblurring results of ours,Matrix+D/S [1], LEDVDI [6], eSL-Net [12], ESTRNN [15], EvST+D/S [13], EST+D/S [3], and EDI [10]. Close-up views are provided below each image.

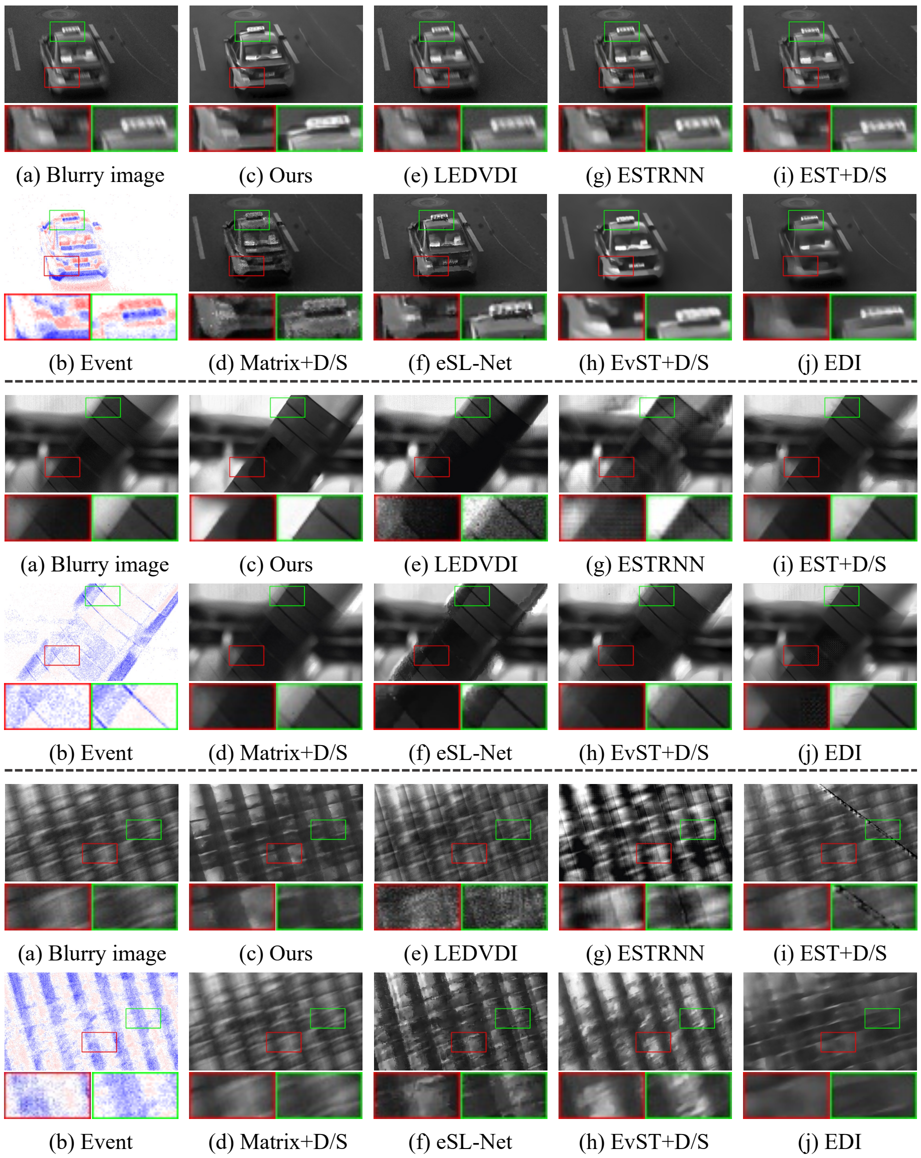


Fig. 18. Qualitative comparisons for deblurring application on real data (part 2). (a) Blurry image. (b) Event. (c)∼(j) Deblurring results of ours,Matrix+D/S [1], LEDVDI [6], eSL-Net [12], ESTRNN [15], EvST+D/S [13], EST+D/S [3], and EDI [10]. Close-up views are provided below each image.

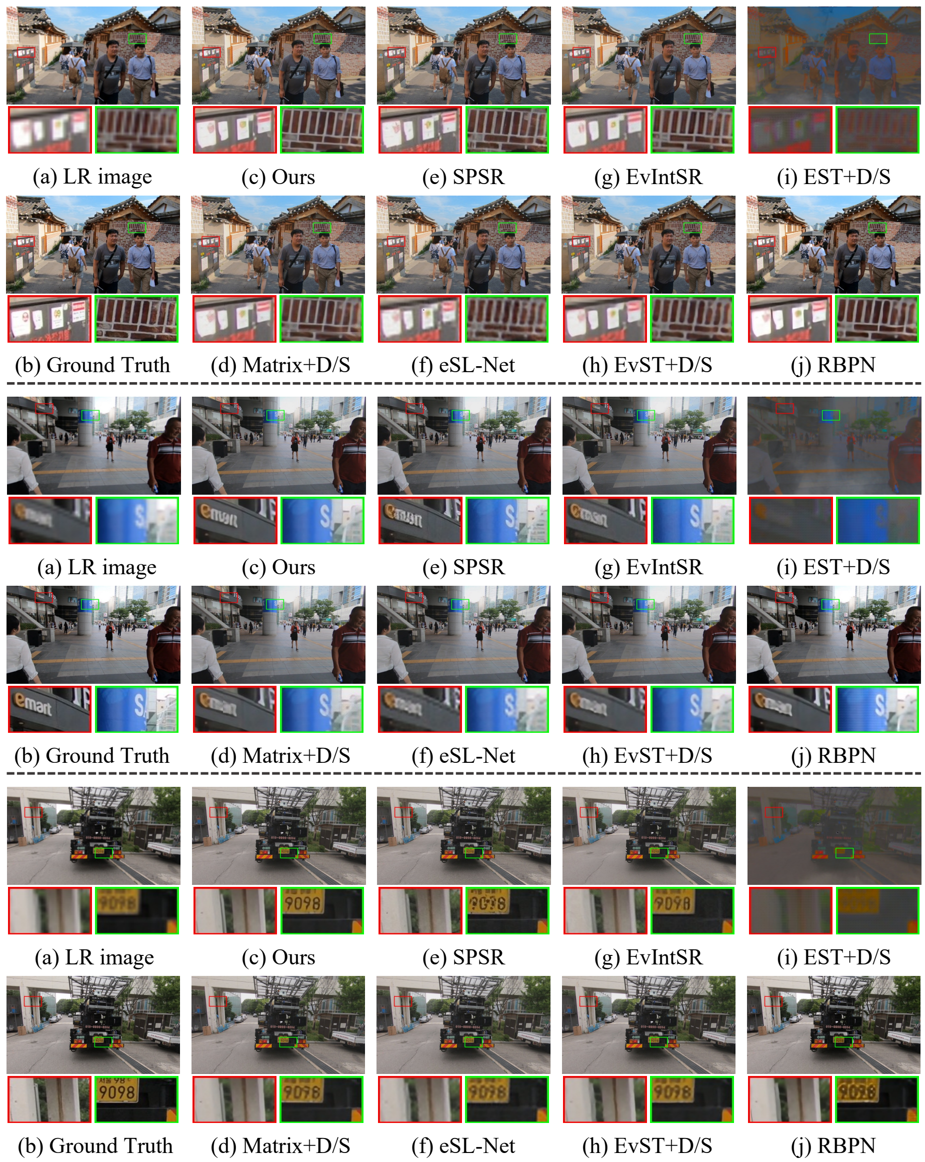


Fig. 19. Qualitative comparisons for super-resolution application on synthetic data (part 1). (a) LR image. (b) Ground truth. (c)∼(j) Super-resolved 4× results of ours, Matrix+D/S [1], SPSR [7], NEST+eSL [12], EvIntSR [4], EvST+D/S [13], EST+D/S [3], and RBPN [5]. Close-up views are provided below each image.

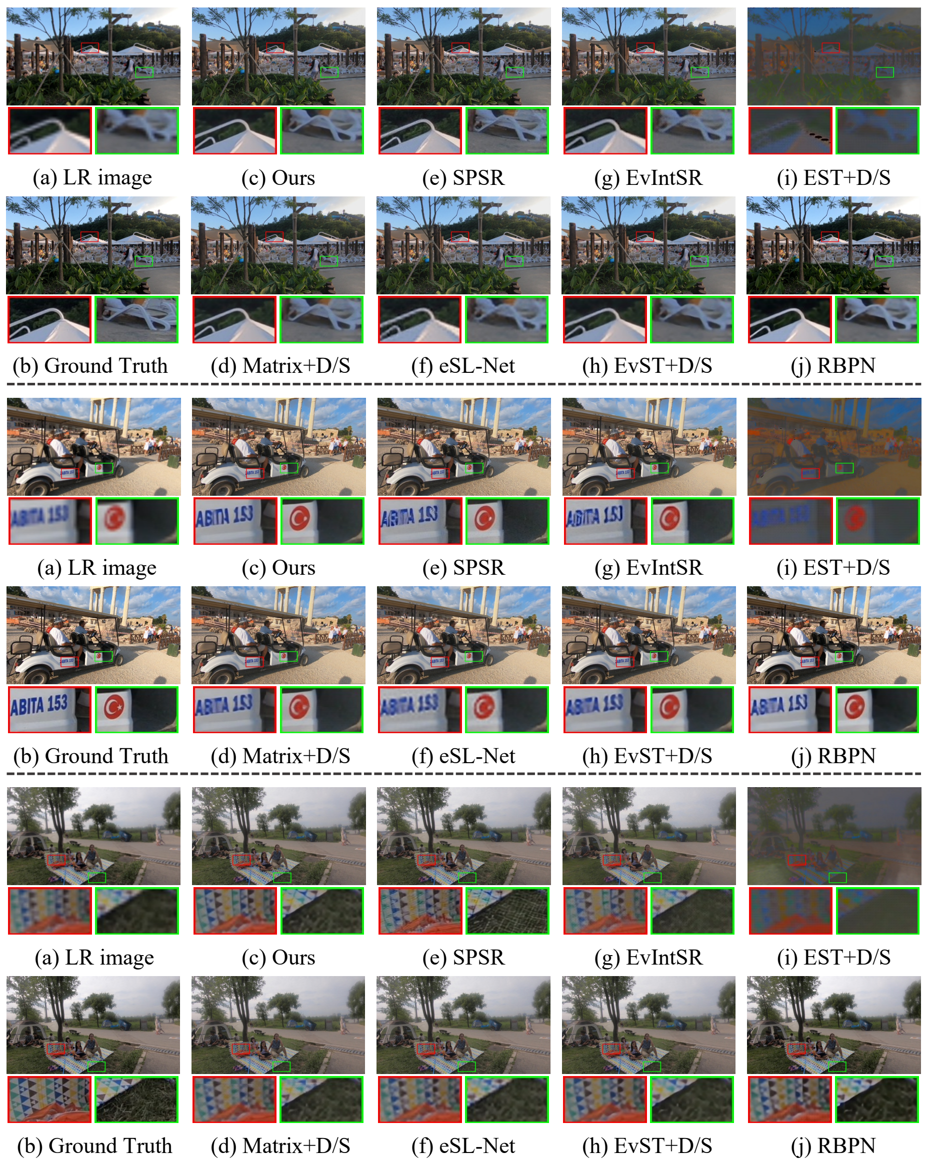


Fig. 20. Qualitative comparisons for super-resolution application on synthetic data (part 2). (a) LR image. (b) Ground truth. (c)∼(j) Super-resolved 4× results of ours, Matrix+D/S [1], SPSR [7], NEST+eSL [12], EvIntSR [4], EvST+D/S [13], EST+D/S [3], and RBPN [5]. Close-up views are provided below each image.

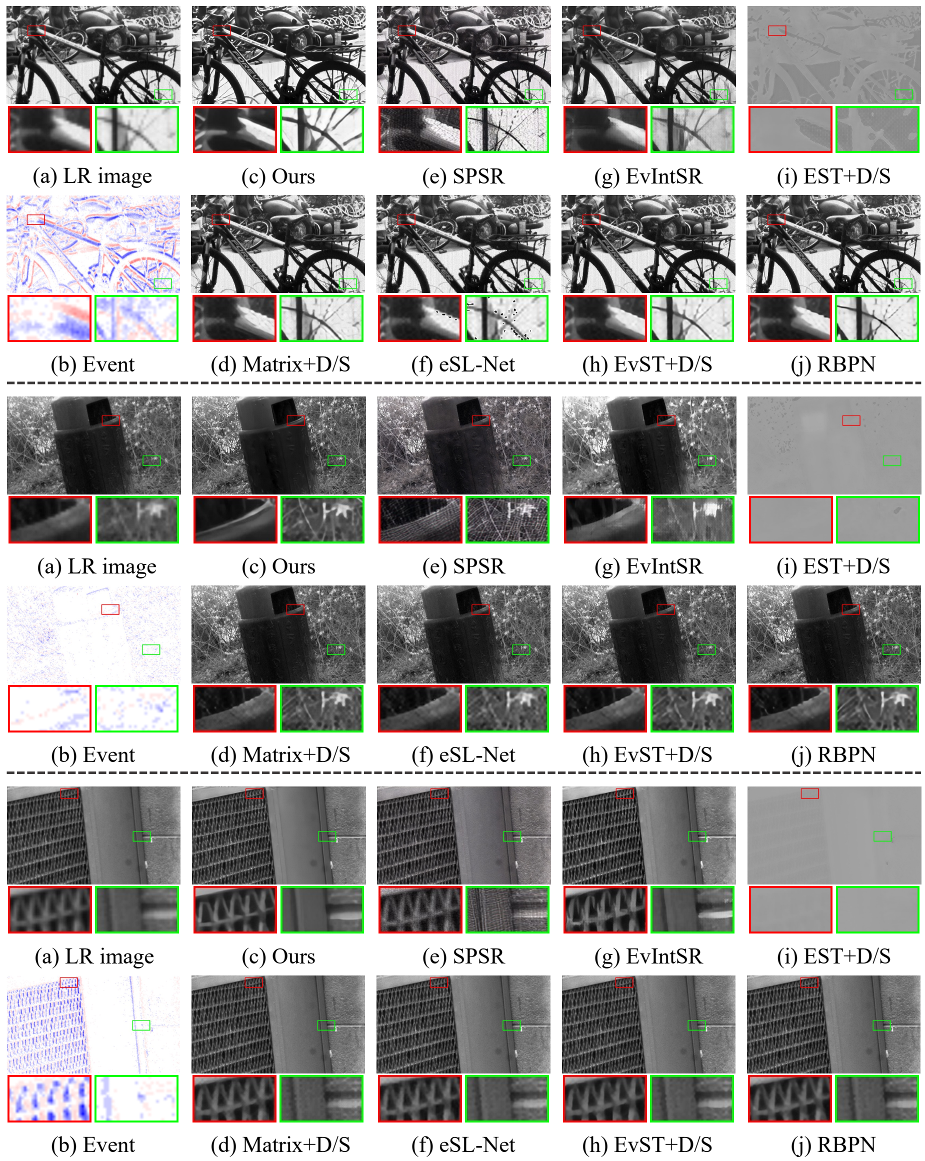


Fig. 21. Qualitative comparisons for super-resolution application on real data (part 1). (a) LR image. (b) Event. (c)$\sim$(j) Super-resolved 4$\times$ results of ours, Matrix+D/S [1], SPSR [7], NEST+eSL [12], EvIntSR [4], EvST+D/S [13], EST+D/S [3], and RBPN [5]. Close-up views are provided below each image.

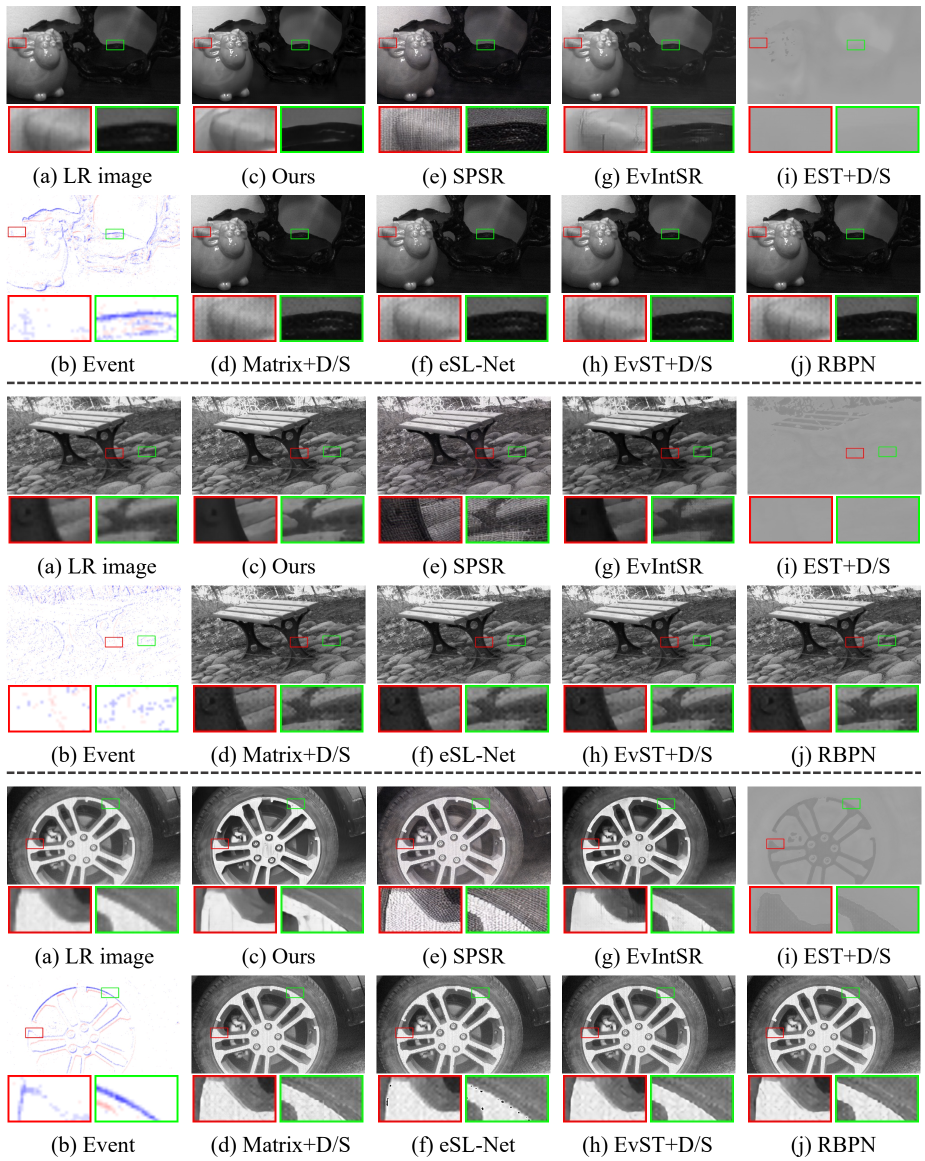


Fig. 22. Qualitative comparisons for super-resolution application on real data (part 2). (a) LR image. (b) Event. (c)$\sim$(j) Super-resolved $4\times$ results of ours, Matrix+D/S [1], SPSR [7], NEST+eSL [12], EvIntSR [4], EvST+D/S [13], EST+D/S [3], and RBPN [5]. Close-up views are provided below each image.

References

1. Cannici, M., Ciccone, M., Romanoni, A., Matteucci, M.: A differentiable recurrent surface for asynchronous event-based data. In: Proc. of European Conference on Computer Vision. pp. 136–152 (2020)
2. Chan, K.C., Zhou, S., Xu, X., Loy, C.C.: Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In: Proc. of Computer Vision and Pattern Recognition. pp. 5972–5981 (2022)
3. Gehrig, D., Loquercio, A., Derpanis, K.G., Scaramuzza, D.: End-to-end learning of representations for asynchronous event-based data. In: Proc. of International Conference on Computer Vision. pp. 5633–5643 (2019)
4. Han, J., Yang, Y., Zhou, C., Xu, C., Shi, B.: EvIntSR-Net: Event guided multiple latent frames reconstruction and super-resolution. In: Proc. of International Conference on Computer Vision. pp. 4882–4891 (2021)
5. Haris, M., Shakhnarovich, G., Ukita, N.: Recurrent back-projection network for video super-resolution. In: Proc. of Computer Vision and Pattern Recognition. pp. 3897–3906 (2019)
6. Lin, S., Zhang, J., Pan, J., Jiang, Z., Zou, D., Wang, Y., Chen, J., Ren, J.: Learning event-driven video deblurring and interpolation. In: Proc. of European Conference on Computer Vision (2020)
7. Ma, C., Rao, Y., Cheng, Y., Chen, C., Lu, J., Zhou, J.: Structure-preserving super resolution with gradient guidance. In: Proc. of Computer Vision and Pattern Recognition (2020)
8. Nah, S., Baik, S., Hong, S., Moon, G., Son, S., Timofte, R., Mu Lee, K.: Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In: Proc. of Computer Vision and Pattern Recognition Workshops. pp. 1996–2005 (2019)
9. Niklaus, S., Mai, L., Liu, F.: Video frame interpolation via adaptive separable convolution. In: Proc. of International Conference on Computer Vision. pp. 261–270 (2017)
10. Pan, L., Scheerlinck, C., Yu, X., Hartley, R., Liu, M., Dai, Y.: Bringing a blurry frame alive at high frame-rate with an event camera. In: Proc. of Computer Vision and Pattern Recognition. pp. 6820–6829 (2019)
11. Rebecq, H., Gehrig, D., Scaramuzza, D.: ESIM: an open event camera simulator. In: Conference on Robot Learning. pp. 969–982 (2018)
12. Wang, B., He, J., Yu, L., Xia, G.S., Yang, W.: Event enhanced high-quality image recovery. In: Proc. of European Conference on Computer Vision (2020)
13. Wang, L., Ho, Y.S., Yoon, K.J., et al.: Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In: Proc. of Computer Vision and Pattern Recognition. pp. 10081–10090 (2019)
14. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proc. of Computer Vision and Pattern Recognition. pp. 14821–14831 (2021)
15. Zhong, Z., Gao, Y., Zheng, Y., Zheng, B.: Efficient spatio-temporal recurrent neural network for video deblurring. In: Proc. of European Conference on Computer Vision. pp. 191–207 (2020)
16. Zihao Zhu, A., Yuan, L., Chaney, K., Daniilidis, K.: Unsupervised event-based optical flow using motion compensation. In: Proc. of European Conference on Computer Vision Workshops (2018)