Supplementary Material: EvIntSR-Net: Event Guided Multiple Latent Frames Reconstruction and Super-resolution

Jin Han[1] Yixin Yang[2] Chu Zhou[1] Chao Xu[1] Boxin Shi[2,3,4]
[1]Key Lab of Machine Perception (MOE), Dept. of Machine Intelligence, Peking University
[2]NELVT, Dept. of Computer Science and Technology, Peking University
[3]Institute for Artificial Intelligence, Peking University
[4]Beijing Academy of Artificial Intelligence

6. Additional Results of EvIntSR-Net

6.1. Results on Image Super-resolution

In addition to Fig. 4 and Fig. 5 of the main paper, we provide more comparisons on synthetic and real data between the proposed EvIntSR-Net and other state-of-the-art methods, including E2SRI [1], eSL-Net [6], EV [5]+SISR [3], and APS+MISR [2]. Fig. 7 $\sim$ Fig. 10 show $2\times$ and $4\times$ SR results on synthetic data. Fig. 11 compares our SR results with those from eSL-Net [6] on real data.

6.2. High-frame-rate Video Generation

High-frame-rate (HFR) videos with super-resolved frames are shown in the supplementary video[1]. We first reconstruct multiple latent frames, then put each latent frame on the central position, which is viewed as the target frame to super-resolve. We interpolate continuous frames with frame-rate 8 times higher than the original APS frames (*e.g.*, 240 FPS videos from 30 FPS videos). In supplementary video, we compare our HFR videos with those generated from eSL-Net [6] on both simulated data and real-captured data. Results show that our reconstruction videos look smoother and reserve more details than eSL-Net [6].

7. Ablation Study on Loss Functions

We ablate different loss functions from the complete model and evaluate them quantitatively in Table 3. The comparing results show that using the combination of L2 loss and perceptual loss helps the network to perform better in reconstructing SR images.

[1] https://youtu.be/3Uc1MMiYiO4

Table 3: Ablation study on loss functions. "Perc. loss" means perceptual loss in this table.

	PSNR↑	SSIM↑	LPIPS↓
L2 loss only	23.03	0.767	0.170
Perc. loss only	22.65	0.740	0.132
L1 loss + perc. loss	22.35	0.764	0.140
L2 loss + perc. loss (Ours)	**23.12**	**0.776**	**0.130**

References

[1] Jonghyun Choi, Kuk-Jin Yoon, et al. Learning to super resolve intensity images from events. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 1

[2] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Recurrent back-projection network for video super-resolution. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 1

[3] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 1

[4] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International Journal of Robotics Research*, 2017. 6

[5] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. Events-to-video: Bringing modern computer vision to event cameras. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 1

[6] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. Event enhanced high-quality image recovery. In *Proc. of European Conference on Computer Vision*, 2020. 1, 6

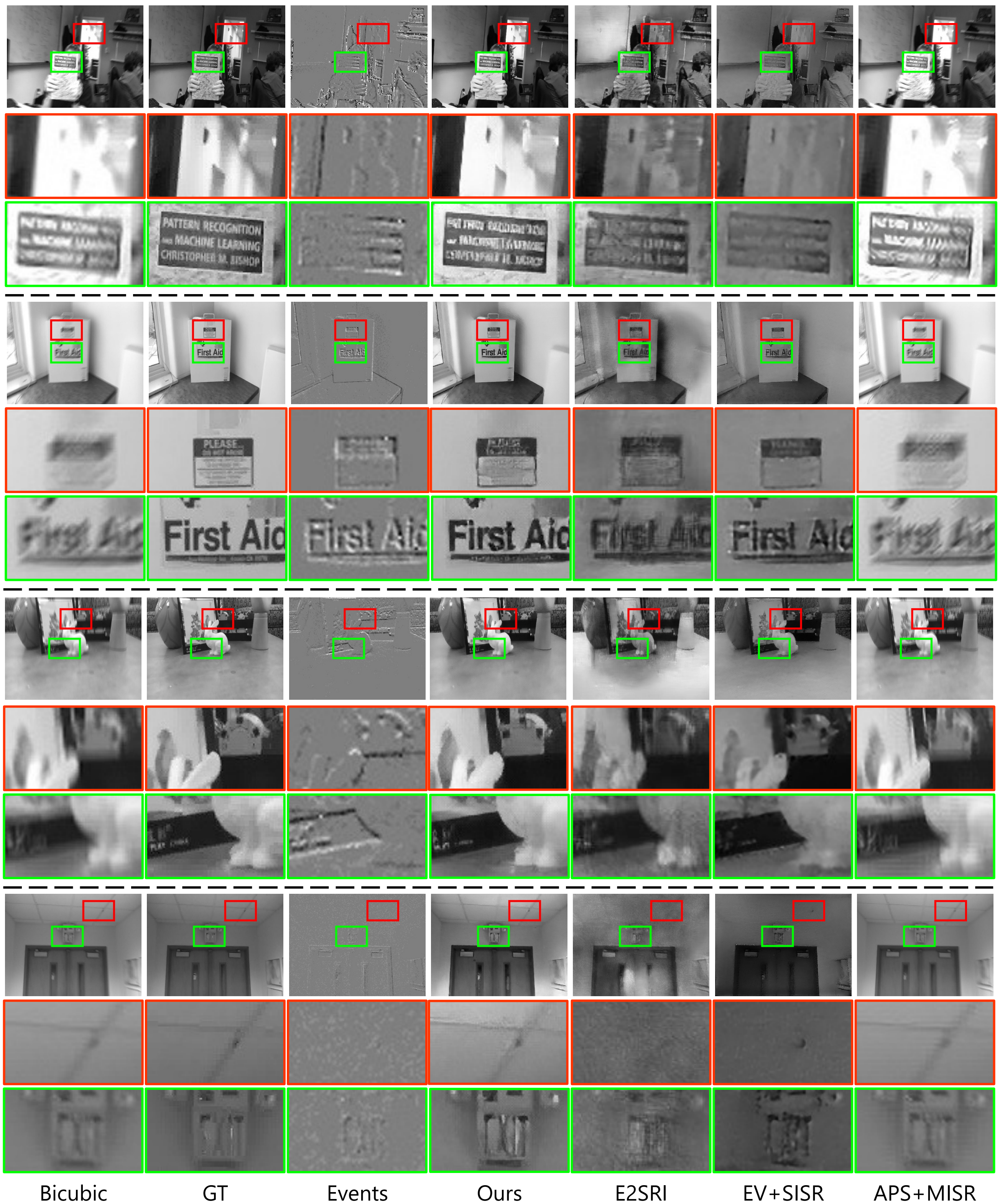

Figure 7: Visual quality comparison of 2× SR on synthetic dataset between EvIntSR-Net and other state-of-the-art super-resolution methods, including both event-based approaches and image-based methods. The APS frames (first column) and event stacks (third column) are upsampled with bicubic interpolation to the corresponding scale for reference.

Figure 8: Visual quality comparison of $2\times$ SR on synthetic dataset between EvIntSR-Net and other state-of-the-art super-resolution methods.

Figure 9: Visual quality comparison of $4\times$ SR on synthetic dataset between EvIntSR-Net and other state-of-the-art super-resolution methods.

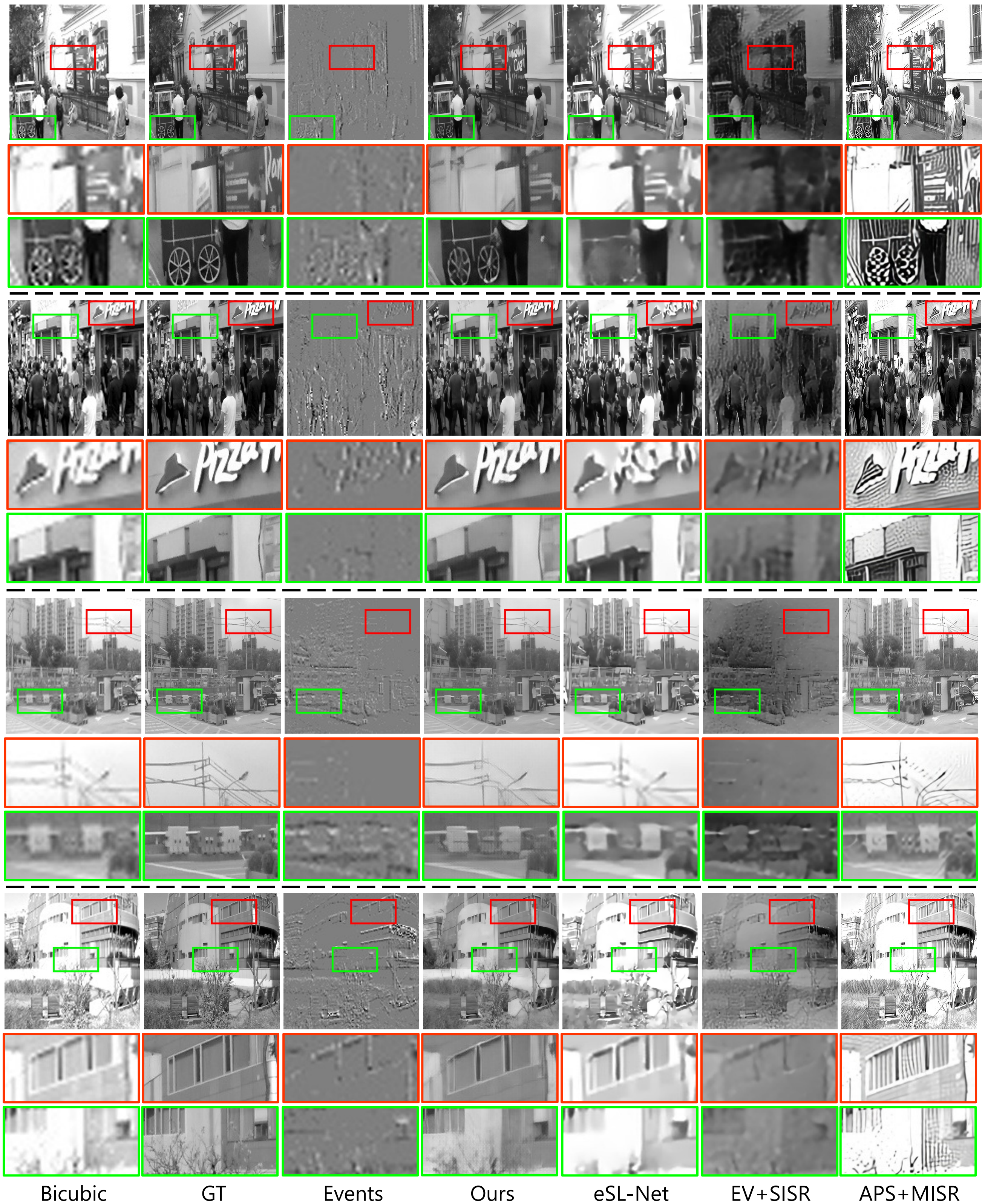

Figure 10: Visual quality comparison of $4\times$ SR on synthetic dataset between EvIntSR-Net and other state-of-the-art super-resolution methods.

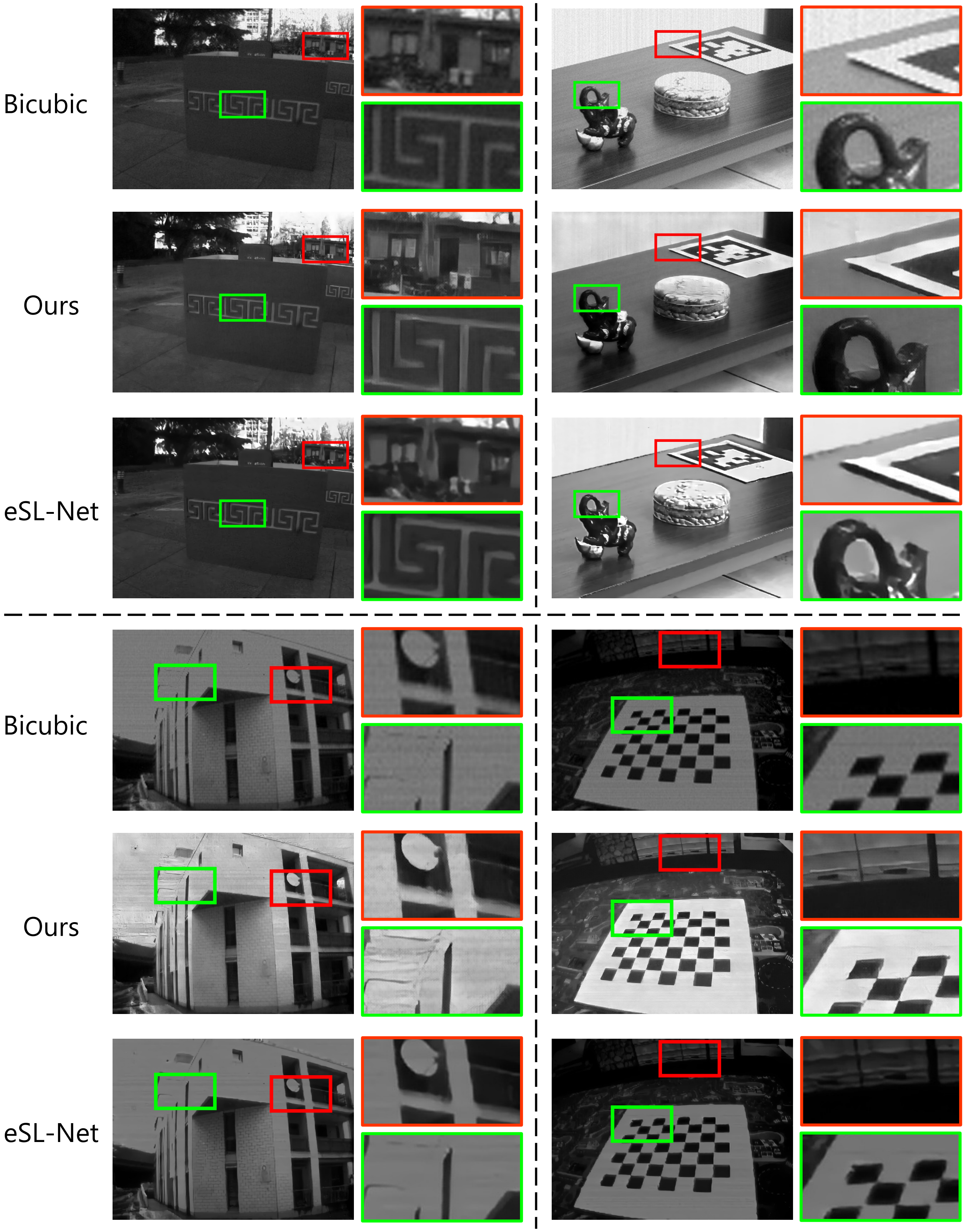

Figure 11: $4\times$ SR visual quality comparison between EvIntSR-Net and eSL-Net [6] on real samples from DAVIS346 captured by us (top 2 cases) and public dataset [4] (bottom 2 cases).