

# —Supplementary Material—

## Event Enhanced High-Quality Image Recovery

Bishan Wang\*, Jingwei He\*, Lei Yu†, Gui-Song Xia, and Wen Yang

Wuhan University, Wuhan, China  
{wangbs, jingwei.he, ly.wd, guisong.xia, yangwen}@whu.edu.cn

### 1 Overview

In this material, we first give the detailed architectures of eSL-Net and more training details. Then, we present ablation study to show the effectiveness of our network architecture. Finally we provide additional visual results to show the superiority of our eSL-Net. Dataset, code, and more results are available at: <https://github.com/ShinyWang33/eSL-Net>.

### 2 Architecture Details of Proposed eSL-Net

The detailed architectures of proposed eSL-Net are illustrated here.

**eSL-Net without super-resolution** The network architecture of eSL-Net without super-resolution is shown in Fig. 1, where  $\mathbf{D}_X$  is the low resolution dictionary. Note: “Conv(in\_channels, out\_channels, kh, kw)” where “in\_channels” is the number of input channels, “out\_channels” is the number of output channels, “kh” represents the height of kernels, and “kw” is the width of kernels. There are 187920 parameters in the eSL-Net without super-resolution.

**eSL-Net with super-resolution** The network architecture of eSL-Net with super-resolution is shown in Fig. 1, where  $\mathbf{D}_X$  is the high resolution dictionary. There are 1322281 parameters in the eSL-Net with super-resolution. We have also evaluated the proposed eSL-Net with input of size  $240 \times 180$  and the corresponding GFLOPs is 230.01, which is less than half of RCAN [6] (563 GFLOPs), a state-of-the-art image super-resolution network exploited to further super-resolve the results of the compared algorithms in Section 7 of our paper. Moreover, eSL-Net is only with about 10% parameters of RCAN.

### 3 More Training Details

We trained eSL-Net on the synthetic training dataset. For eSL-Net without super-resolution, the training dataset includes 22800 low-resolution clear images, 22800 low-resolution blurry images and 22800 event sequences. For eSL-Net with super-resolution, the training dataset includes 22800 high-resolution clear images, 22800 low-resolution blurry images and 22800 event sequences.

---

† Corresponding Author

\* Equal contribution

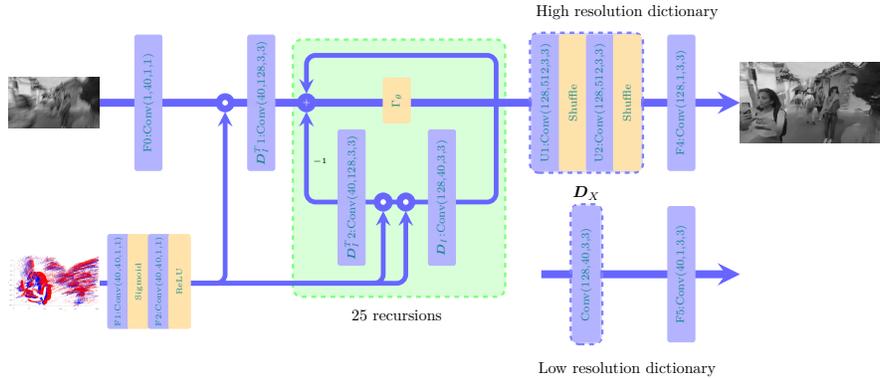


Fig. 1. Network architecture of eSL-Net

The network is optimized using Adam with batch size of 12, momentum parameter of 0.9 and weight decay of  $10^{-4}$ . The learning rate is initially set to 0.0008 and then decreased by a factor of 10 every 10 epochs. We train a total of 50 epochs as no further increase of PSNR on the synthetic testing dataset can be observed.

## 4 Ablation Study

To validate the effectiveness of the proposed architecture and each specific module, we conduct ablation study on three variants as follows.

**Single input architecture** We compare our network with a single input architecture without the encoder of  $\mathbf{E}$ . This can be achieved by concatenating events and input image  $\mathbf{Y}$ , then sending the  $1+2 \times k$  channel tensor to the single input network.

**Without sparse learning architecture** We investigate the effectiveness of the iteration module of sparse learning. In the iteration module we replace relu-conv block with conv-relu block that is common in CNN.

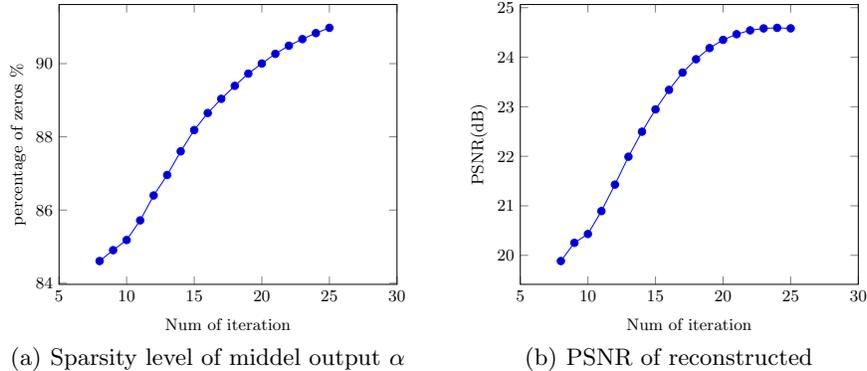
We calculated PSNR for above models on the synthetic testing dataset, which are shown in the table 1.

**Different iterations and middle output of sparse coefficients** the iteration block is the most important module composed by a series of iterations corresponding to Eq. (9), aiming at denoising by representing inputs with sparse coefficients  $\alpha$ . Thus as increasing iterations,  $\alpha$  becomes sparser and the reconstruction is more accurate as shown in the Fig. 2.

These results demonstrate our complete model achieves the optimal performance with these specifically designed strategies.

**Table 1.** Quantitative comparison of our eSL-Net and ablation study on the synthetic testing dataset.

Methods	Single input	Without sparse learning	eSL-Net
PSNR(dB)	25.13	6.5998	<b>25.41</b>

**Fig. 2.** Middle output and Reconstruction of eSL-Net for different number of iterations on the synthetic testing dataset.

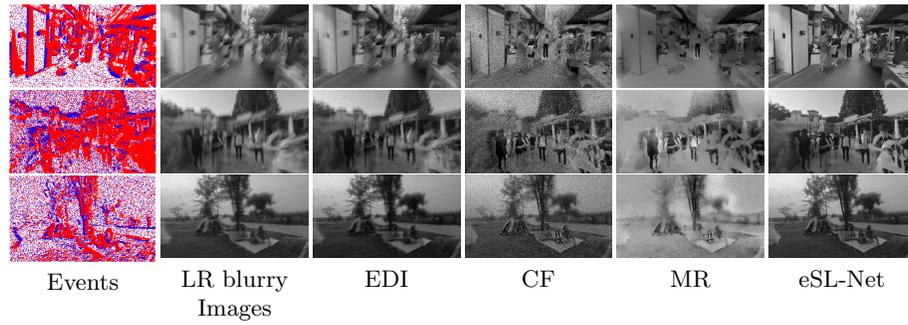
## 5 More Qualitative Results

### 5.1 Intensity Reconstruction without Super-Resolution

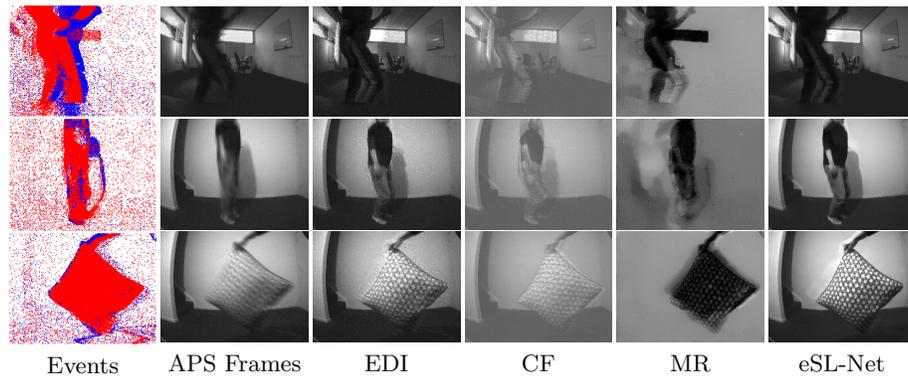
In Fig. 3 and Fig. 4, we provide additional intensity reconstruction results without super-resolution on the synthetic dataset and the real scenes to clearly show the superiority of our proposed eSL-Net. The event-based double integration (EDI) [4], complementary filter method (CF) [5] and manifold regularization method (MR) [3] are used for comparisons.

### 5.2 Intensity Reconstruction with Super-Resolution

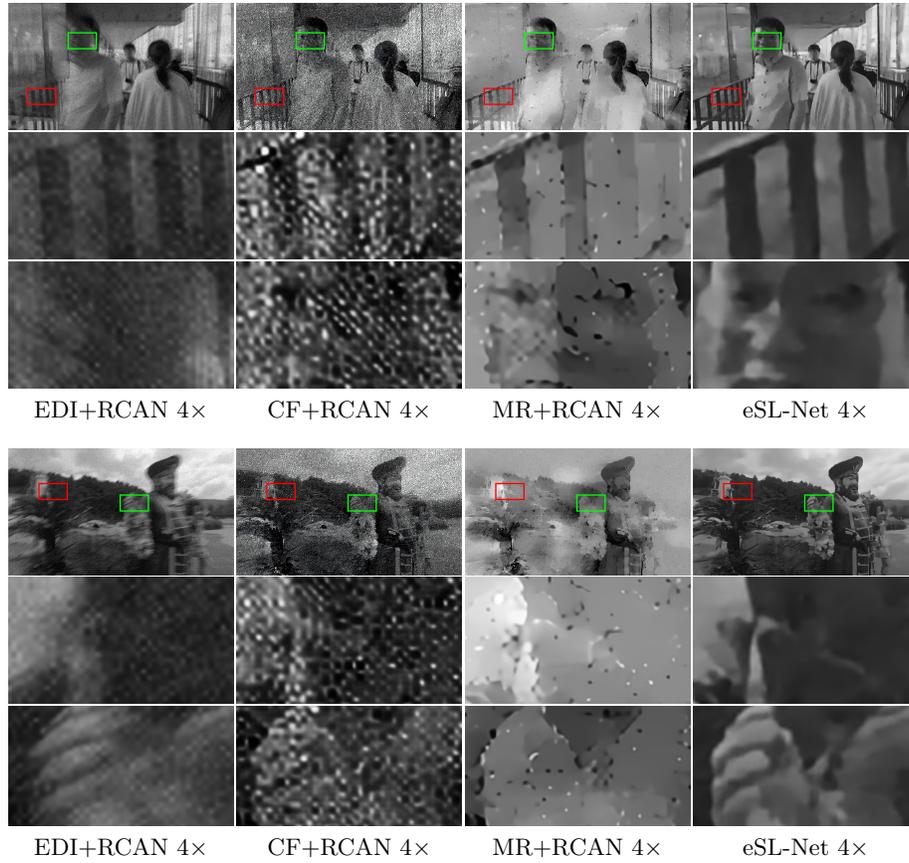
In Fig. 5 and Fig. 6, we provide additional intensity reconstruction results with super-resolution on the synthetic dataset and the real scenes to clearly show the superiority of our proposed eSL-Net. EDI [4], CF [5] and MR [3] armed with an excellent SR network RCAN [6] are used for comparisons.



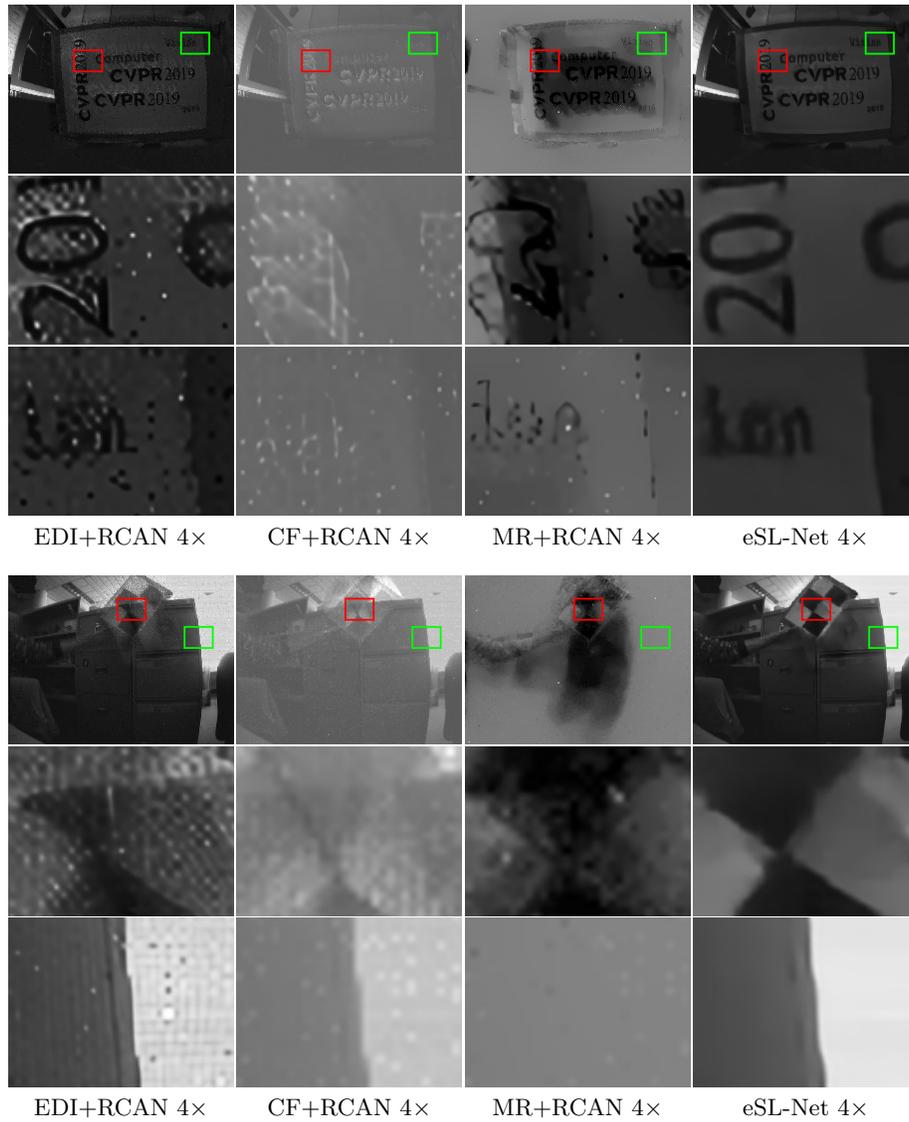
**Fig. 3.** Qualitative comparison of our outputs without SR to EDI, CF and MR on the synthetic testing dataset.



**Fig. 4.** Qualitative comparison of our outputs without SR to EDI, CF and MR on the real dataset [4].



**Fig. 5.** Qualitative comparison of our outputs to EDI, CF and MR with SR method on the synthetic testing dataset. The second and third lines are detailed images zoomed for better comparison.



**Fig. 6.** Qualitative comparison of our outputs to EDI, CF and MR with SR method on the real dataset [4]. The second and third lines are detailed images zoomed for better comparison.

## References

1. Brandli, C., Berner, R., Yang, M., Liu, S.C., Delbruck, T.: A  $240 \times 180$  130 db 3  $\mu$ s latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits* **49**(10), 2333–2341 (2014)
2. Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A., Conradt, J., Daniilidis, K., et al.: Event-based vision: A survey. *arXiv preprint arXiv:1904.08405* (2019)
3. Munda, G., Reinbacher, C., Pock, T.: Real-time intensity-image reconstruction for event cameras using manifold regularisation. *International Journal of Computer Vision* **126**(12), 1381–1393 (2018)
4. 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: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 6820–6829 (2019)
5. Scheerlinck, C., Barnes, N., Mahony, R.: Continuous-Time Intensity Estimation Using Event Cameras. In: Jawahar, C., Li, H., Mori, G., Schindler, K. (eds.) *Computer Vision – ACCV 2018*. vol. 11365, pp. 308–324. Springer International Publishing, Cham (2019)
6. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: *Proceedings of the European Conference on Computer Vision*. pp. 286–301 (2018)