Supplemental Material: Learning to See in the Dark with Events

Song Zhang*[1,3], Yu Zhang*†[3,4], Zhe Jiang*[3], Dongqing Zou[3], Jimmy Ren[3], and Bin Zhou†[1,2]

[1] State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China
[2] Peng Cheng Laboratory, Shenzhen, China
[3] SenseTime Research, Beijing, China [4] Tsinghua University, Beijing, China
{zhangs,zhoubin}@buaa.edu.cn
{zhangyu1,jiangzhe,zoudongqing,rensijie}@sensetime.com

1 Introduction

In this supplementary material, we provide:

1) The detailed layer/parameter configurations of the proposed network architecture (Sect. 2);
2) Illustration of the captured DVS-Dark dataset (Sect. 3);
3) Additional visual comparisons with state-of-the-art models (Sect. 4).

Note that for 3), abbrevations of different approaches are: Pix2Pix [2, 3], CycleGAN [8], I2I [5], SDA [1], CIE [7], RIRM [4] and E2V [6].

References

1. Hong, W., Wang, Z., Yang, M., Yuan, J.: Conditional generative adversarial network for structured domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1335–1344 (2018)
2. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
3. L. Wang, S. M. Mostafavi, Y.S.H., Yoon, K.J.: Event-based High Dynamic Range Image and Very High Frame Rate Video Generation using Conditional Generative Adversarial Networks (2019)
4. 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)
5. Murez, Z., Kolouri, S., Kriegman, D.J., Ramamoorthi, R., Kim, K.: Image to image translation for domain adaptation. arXiv preprint arXiv:1712.00479 (2017)

*Equal contribution.
†Corresponding authors: Bin Zhou (zhoubin@buaa.edu.cn) and Yu Zhang (zhangyulb@gmail.com).

6. Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D.: Events-to-video: Bringing modern computer vision to event cameras. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
7. Scheerlinck, C., Barnes, N., Mahony, R.: Continuous-time intensity estimation using event cameras. In: Asian Confence on Compututer Vision (ACCV) (2018)
8. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on (2017)

2 Architecture Details

As described in Sect. 3 of the paper, the proposed architecture contains 5 main modules: the shared and private encoders E_c and E_p, the discriminaor D, the reconstruction decoder R and the detail enchancing branch T_e, whose layer configurations are listed as follows. Note that the structure of the residual block is referred to Table 6.

Table 1: Layer configurations of shared encoder E_c.

E_c	Input	Kernel size	Stride	In channels	Out channels	Output
Conv1	Events	7	1	8	32	Conv1
Norm1	Conv1	Instance Norm + ReLU				Norm1
Conv2	Norm1	3	2	32	64	Conv2
Norm2	Conv2	Instance Norm + ReLU				Norm2
Conv3	Norm2	3	2	64	128	Conv3
Norm3	Conv3	Instance Norm + ReLU				Norm3
Res1	Norm3	Residual block × 9				X_c^f/X_{LE}^f

Table 2: Layer configurations of private encoder E_p.

E_p	Input	Kernel size	Stride	In channels	Out channels	Output
Conv1	$concatenate(Events, noise)$	7	1	9	32	Conv1
Norm1	Conv1	Instance Norm + ReLU				Norm1
Conv2	Norm1	3	2	32	64	Conv2
Norm2	Conv2	Instance Norm + ReLU				Norm2
Conv3	Norm2	3	2	64	128	Conv3
Norm3	Conv3	Instance Norm + ReLU				Norm3
Res1	Norm3	Residual block × 9				X_p^f

Table 3: Layer configurations of discriminator D.

D	Input	Kernel size	Stride	In channels	Out channels	Output
Conv1	X_{DE}^f/X_{LE}^f	5	2	128	32	Conv1
Norm1	Conv1	Leaky ReLU				Norm1
Conv2	Norm1	5	2	32	64	Conv2
Norm2	Conv2	Instance Norm + Leaky ReLU				Norm2
Conv3	Norm2	5	2	64	128	Conv3
Norm3	Conv3	Instance Norm + Leaky ReLU				Norm3
Conv4	Norm3	5	1	128	256	Conv4
Norm4	Conv4	Instance Norm + Leaky ReLU				Norm4
Conv5	Norm4	5	1	256	1	R/F

Table 4: Layer configurations of the detail enchancing branch T_e.

T_e	Input	Kernel size	Stride	In channels	Out channels	Output
Conv1	$concatenate(X_{DE}^f, noise)$	3	1	160	128	Conv1
Norm1	Conv1	Instance Norm + ReLU				Norm1
Res1	Norm1	Residual block × 9				Δy

Table 5: Layer configurations of reconstruction decoder R.

R	Input	Kernel size	Stride	In channels	Out channels	Output
Res1	$X_{DE}^f + \Delta y/X_{DE}^f$	Residual block × 9				Res1
Deconv1	Res1	3	2	128	64	Deconv1
Norm1	Deconv1	Instance Norm + ReLU				Norm1
Deconv2	Norm1	3	2	64	32	Deconv2
Norm2	Deconv2	Instance Norm + ReLU				Norm2
Conv1	Norm2	7	1	32	1	Conv1
Sigmoid	Conv1	sigmoid				$\hat{y}/\bar{y}$

Table 6: Layer configurations of a residual block.

Residual block	Input	Kernel size	Stride	In channels	Out channels	Ouput
Conv1	*	3	1	128	128	Conv1
Norm1	Conv1	Instance Norm + ReLU				Norm1
Conv2	Norm1	3	1	128	128	Conv2
Norm2	Conv2	Instance Norm				Norm2

3 Illustration of the DVS-Dark Dataset

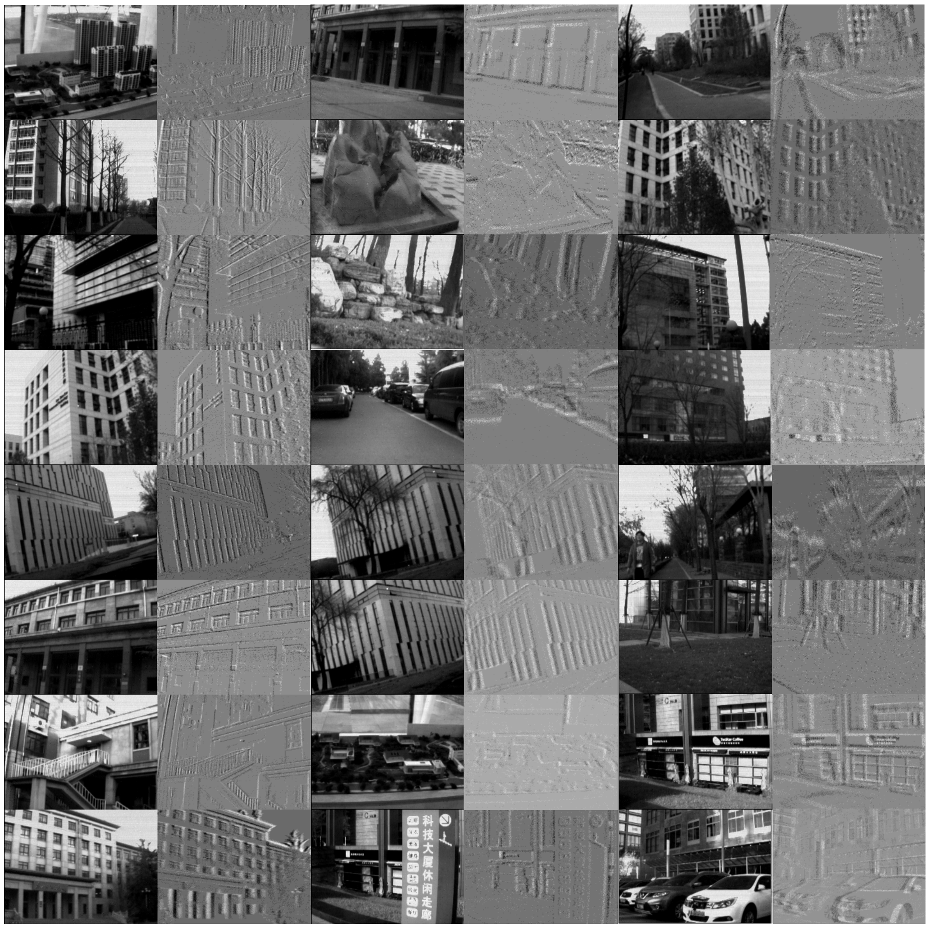

Fig. 1: Representative intensity images and the corresponding event frames of the daylight split of DVS-Dark dataset.

Fig. 2: Representative intensity images and the corresponding event frames of the low-light split of DVS-Dark dataset.

4 More Visual comparisons with State-of-the-Art Models

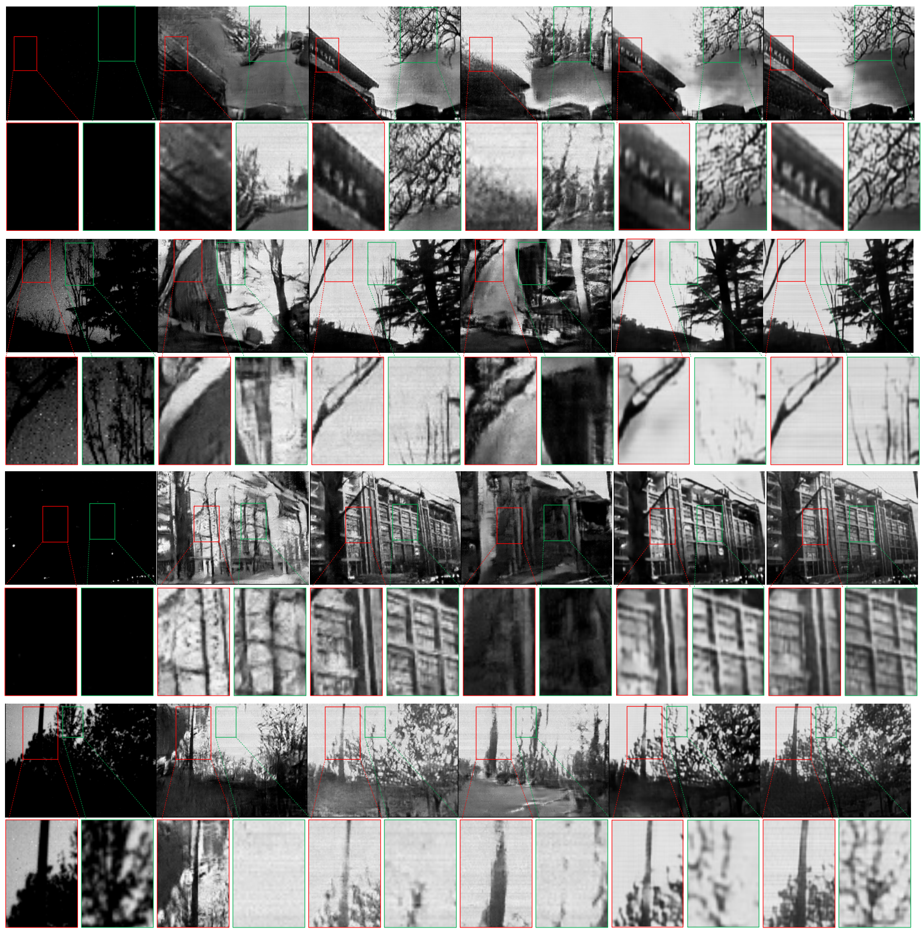

Fig. 3: Additional visual comparisons with domain-adaptation approaches Pix2Pix, CycleGAN, I2I and SDA, tested on the real low-light events.

Fig. 4: Additional visual comparisons with domain-adaptation approaches Pix2Pix, CycleGAN, I2I and SDA, tested on the real low-light events.

Fig. 5: Additional visual comparisons with domain-adaptation approaches Pix2Pix, CycleGAN, I2I and SDA, tested on the real low-light events.

Fig. 6: Additional visual comparisons with domain-adaptation approaches Pix2Pix, CycleGAN, I2I and SDA, tested on the real low-light events.

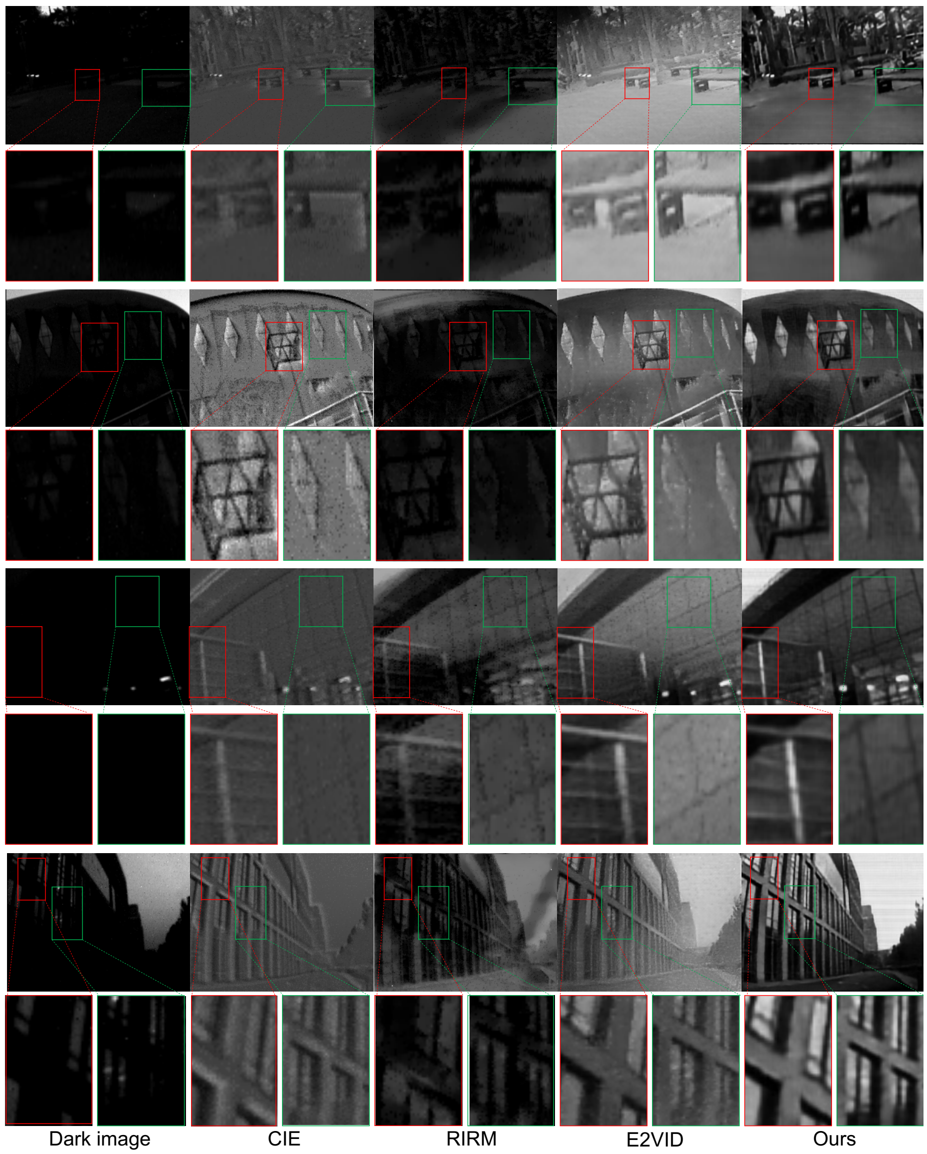

Fig. 7: Additional visual comparisons with event based intensity reconstruction approaches CIE, RIRM, E2VID, tested on the real low-light events.

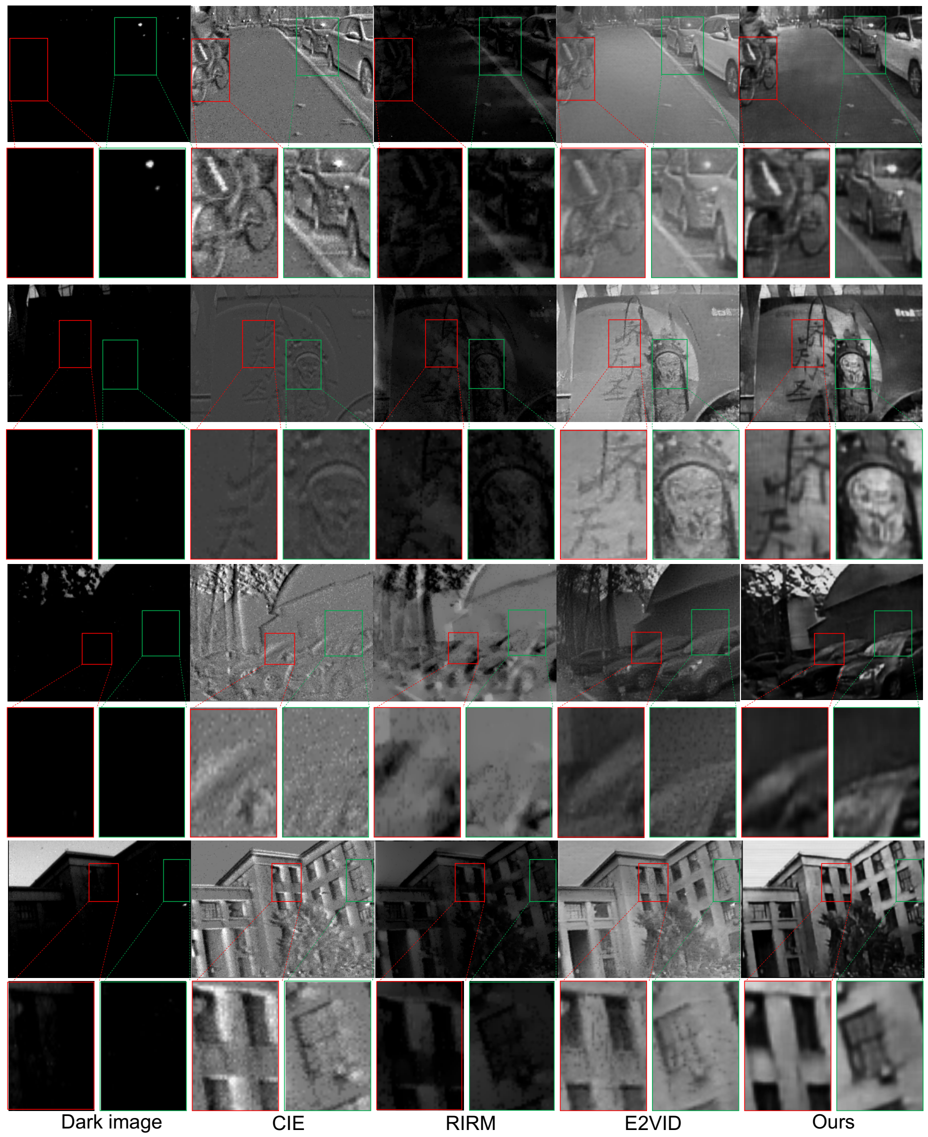

Fig. 8: Additional visual comparisons with event based intensity reconstruction approaches CIE, RIRM, E2VID, tested on the real low-light events.

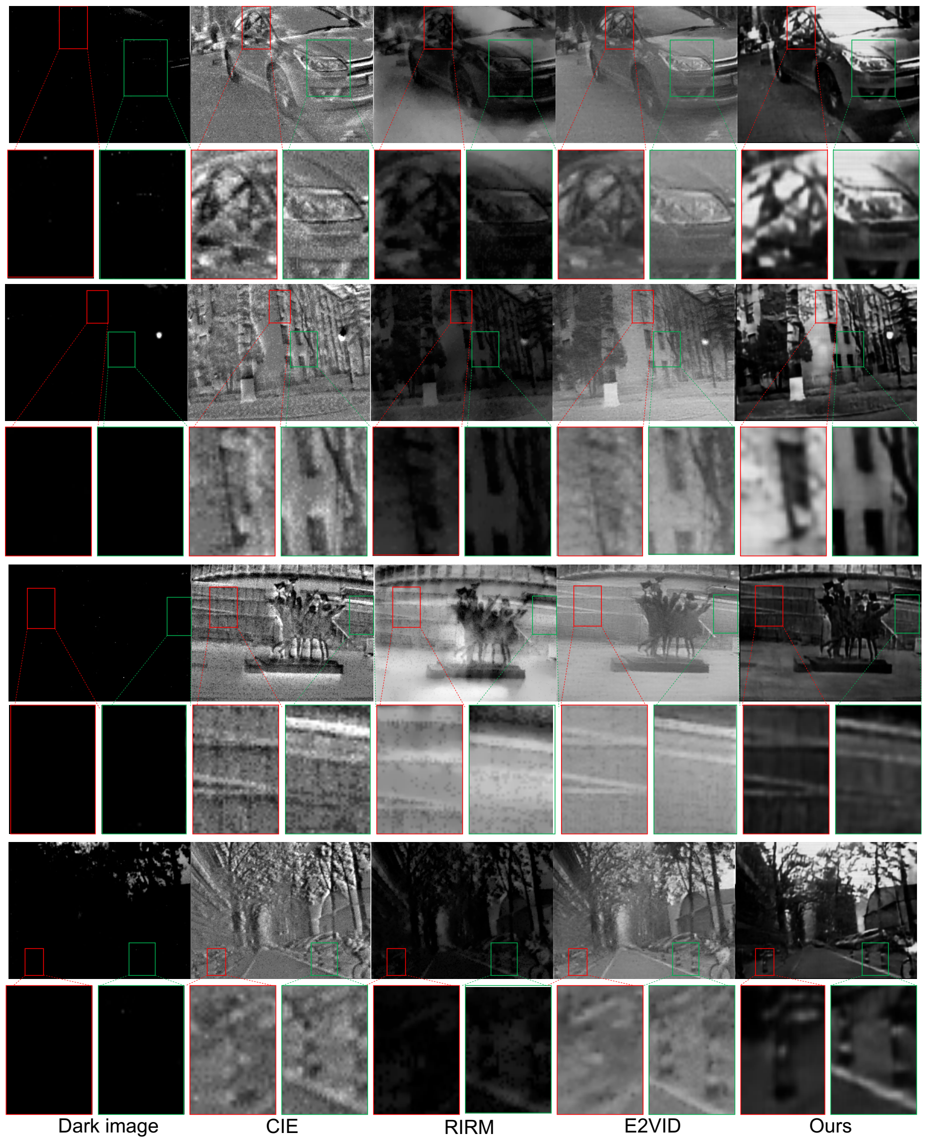

Fig. 9: Additional visual comparisons with event based intensity reconstruction approaches CIE, RIRM, E2VID, tested on the real low-light events.

Fig. 10: Additional visual comparisons with event based intensity reconstruction approaches CIE, RIRM, E2VID, tested on the real low-light events.