

Supplemental Material:

Color4E: Event Demosaicing for Full-color Event Guided Image Deblurring

6 NETWORK DETAILS

In the EDM and IDM of our proposed Color4E network, cross attention modules integrate features from different modalities. The cross attention module we employ operates in a channel attention manner, meaning that the number of elements in the attention score matrix during intermediate computations is $c \times c$, where c is the dimensionality of the features. This allows the model to extract and fuse information along the feature dimension, better handling information from two modalities, and enhancing the model's applicability to inputs of arbitrary sizes. The experiment results in this paper verify that our model can handle input images of different sizes without modification, *e.g.*, while the model is trained with the C4E dataset at a resolution of 256×256 , the 1280×720 test images of GoPro [4] can still be effectively deblurred.

Formally, the input data undergoes the following operations in our cross attention module:

$$Q_X = XW^Q, \quad K_Y = YW^K, \quad V_Y = YW^V, \quad (9)$$

where X and Y are features after layer normalization. Depending on the modality of interest, in the EDM, the cross attention module takes events as queries and images as keys and values, and thus X is event feature and Y is image feature; while in the IDM, it is the reverse, with images as queries and events as keys and values. Q , K and V are the transformed queries, keys, and values, respectively. Then, attention scores are calculated, normalized by the softmax function, and finally multiplied with the values to obtain the fused result along the feature dimension:

$$\text{Attention}(Q_X, K_Y, V_Y) = \text{Softmax}\left(\frac{Q_X^T K_Y}{\sqrt{d_k}}\right) V_Y. \quad (10)$$

The attention results obtained are then residual-connected with the features of interest to obtain the corresponding attention features:

$$F_X = X + \text{Attention}(Q_X, K_Y, V_Y), \quad (11)$$

where F_X denotes the intermediate feature. Finally, this feature is fed into an MLP with residual connections for final feature integration, resulting in the output:

$$X' = F_X + \text{MLP}(F_X). \quad (12)$$

For the sake of training stability, we adopt the pre-layer normalization setting as mentioned above and incorporate modifications to the attention layer as proposed by [2, 3].

7 ANALYSIS OF C4E DATASET

We implement a display-filter-camera system that enables the synchronous recording of mosaic and full-color event to collect the high-resolution color event dataset C4E suitable for network training and evaluation, which can avoid the real-simulated gap of events. To verify the effectiveness of data collected by the display-filter-camera system in mitigating the real-simulated gap, we train the

proposed Color4E network with mosaic event data simulated from the event simulator DVS-Voltmeter [1] and mosaic event data captured from display-filter-camera system (*i.e.*, C4E dataset) respectively, then use the trained model to perform image deblurring on the real-captured dataset. Both models are trained with the same operational scheme and number of training epochs.

The experimental results are shown in Fig. 9, where input blurry images and mosaic events are outputted from a DAVIS346-color [5] camera. The comparison shows that the model trained with C4E can reconstruct sharper textures and recover tiny textures, such as the window in the first row and the traffic sign in the third row. This indicates that the event data collected based on the display-filter-camera system is more consistent with the feature distribution of real color mosaic events, so that the trained model can be generalized to the real-world scene.

8 ADDITIONAL RESULTS

We show additional examples of image deblurring results on the reprocessed GoPro dataset [4] and our collected C4E dataset in this section. Color4E is compared with recent event-based image deblurring methods eSL-Net [9], Red-Net [10], NEST [8], EF-Net [6] and REFID [7] on our reprocessed GoPro dataset (Fig. 10, Fig. 11, and Fig. 12) and our collected C4E dataset (Fig. 13 and Fig. 14). We also compare the performance of methods trained with different types of events, *i.e.*, mosaic color events and mono events, verifying the corresponding comparative result shown in Table 2.

REFERENCES

- [1] Songnan Lin, Ye Ma, Zhenhua Guo, and Bihan Wen. 2022. DVS-Voltmeter: Stochastic Process-Based Event Simulator for Dynamic Vision Sensors. In *Proc. of European Conference on Computer Vision*.
- [2] Liyuan Liu, Xiaodong Liu, Jianfeng Gao, Weizhu Chen, and Jiawei Han. 2020. Understanding the Difficulty of Training Transformers. In *Proc. of the Conference on Empirical Methods in Natural Language Processing*.
- [3] Xiaodong Liu, Kevin Duh, Liyuan Liu, and Jianfeng Gao. 2020. Very Deep Transformers for Neural Machine Translation. In *arXiv preprint arXiv:2008.07772*.
- [4] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. 2017. Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring. In *Proc. of Computer Vision and Pattern Recognition*.
- [5] Cedric Scheerlinck, Henri Rebecq, Timo Stoffregen, Nick Barnes, Robert Mahony, and Davide Scaramuzza. 2019. CED: Color event camera dataset. In *Proc. of Computer Vision and Pattern Recognition Workshops*.
- [6] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. 2022. Event-Based Fusion for Motion Deblurring with Cross-modal Attention. In *Proc. of European Conference on Computer Vision*.
- [7] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhong Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. 2023. Event-Based Frame Interpolation with Ad-hoc Deblurring. In *Proc. of Computer Vision and Pattern Recognition*.
- [8] Mingui Teng, Chu Zhou, Hanyue Lou, and Boxin Shi. 2022. NEST: Neural event stack for event-based image enhancement. In *Proc. of European Conference on Computer Vision*.
- [9] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. 2020. Event Enhanced High-Quality Image Recovery. In *Proc. of European Conference on Computer Vision*.
- [10] Fang Xu, Lei Yu, Bishan Wang, Wen Yang, Gui-Song Xia, Xu Jia, Zhendong Qiao, and Jianzhuang Liu. 2021. Motion Deblurring With Real Events. In *Proc. of International Conference on Computer Vision*.

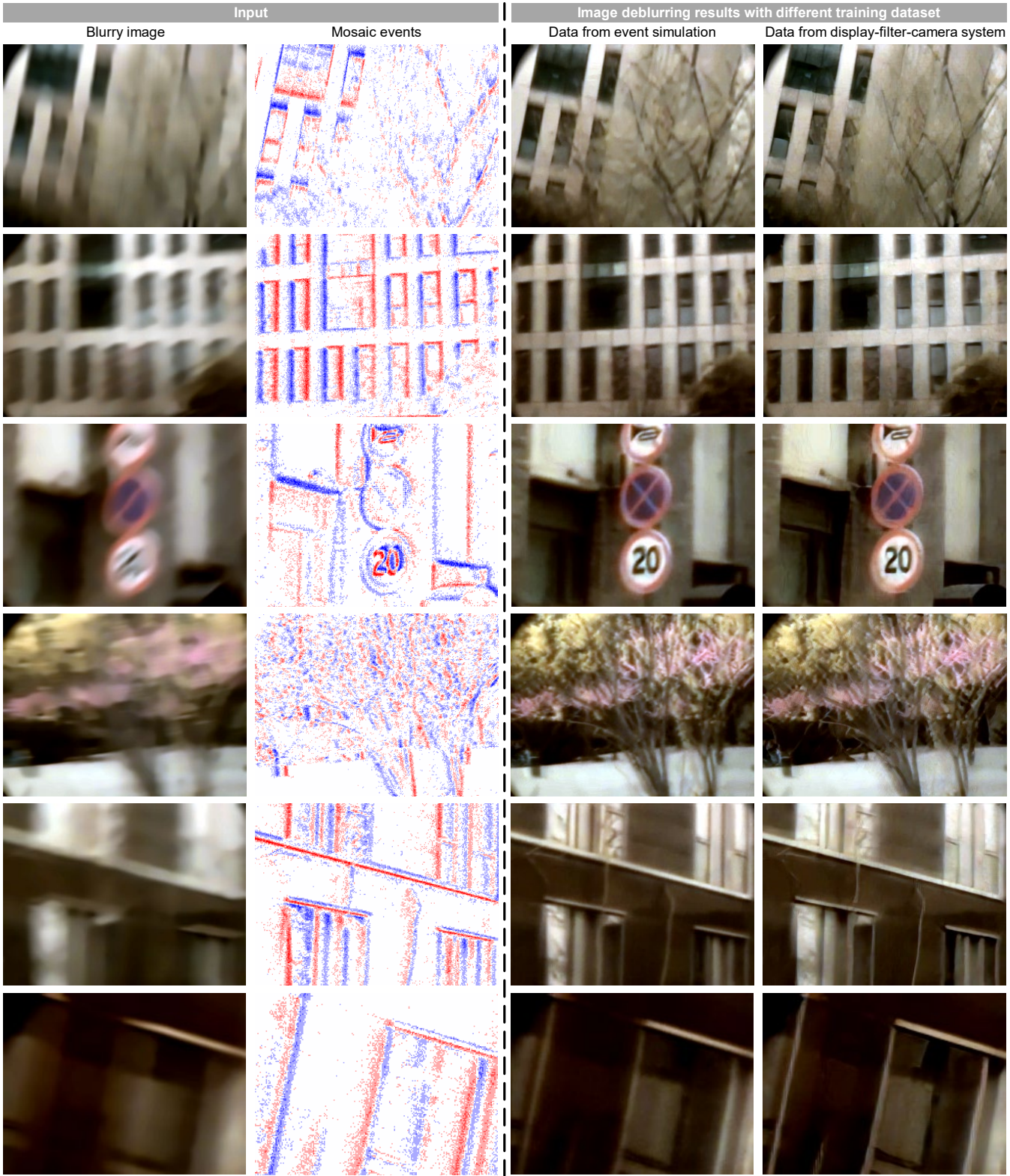


Figure 9: Image deblurring results reconstructed by our method for the real-captured dataset, our method is respectively trained with DVS-Voltmeter [1] simulation generated events and our display-filter-camera system captured events.

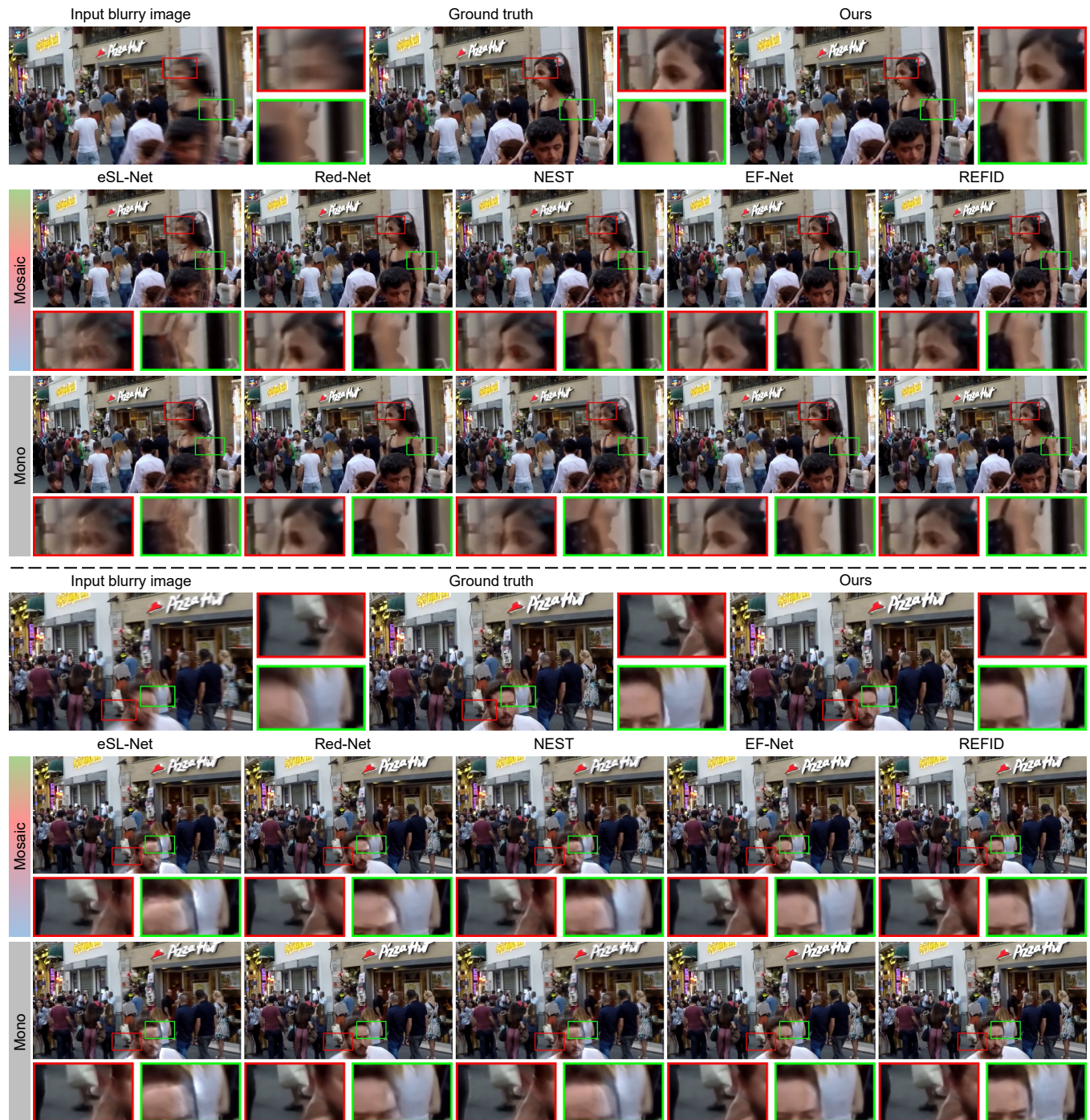


Figure 10: Image deblurring results on GoPro dataset. The labels “Mosaic” and “Mono” denote the networks are trained with different types of events.

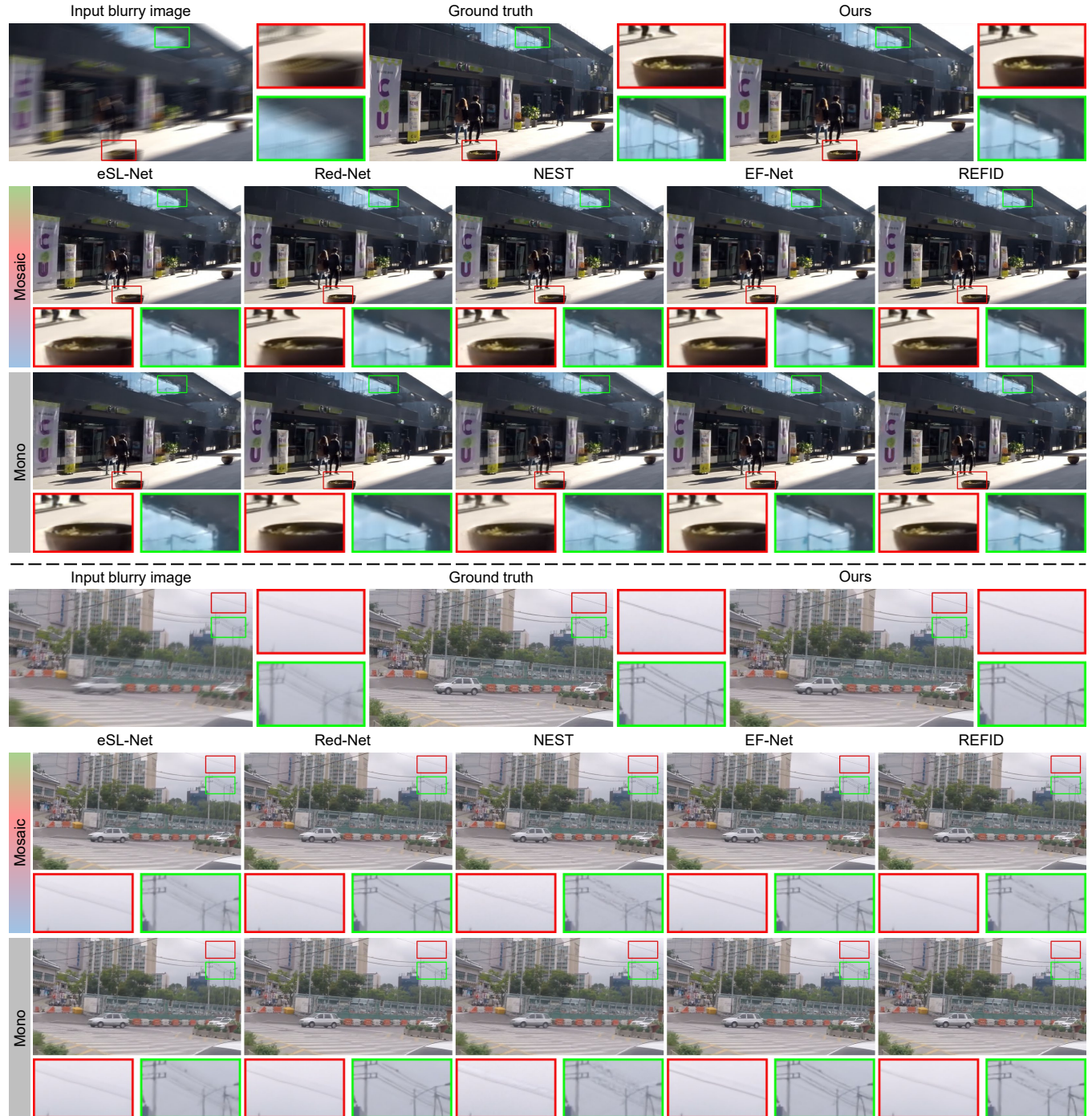


Figure 11: Image deblurring results on GoPro dataset. The labels “Mosaic” and “Mono” denote the networks are trained with different types of events.



Figure 12: Image deblurring results on GoPro dataset. The labels “Mosaic” and “Mono” denote the networks are trained with different types of events.



Figure 13: Image deblurring results on our C4E dataset. The labels “Mosaic” and “Mono” denote the networks are trained with different types of events.

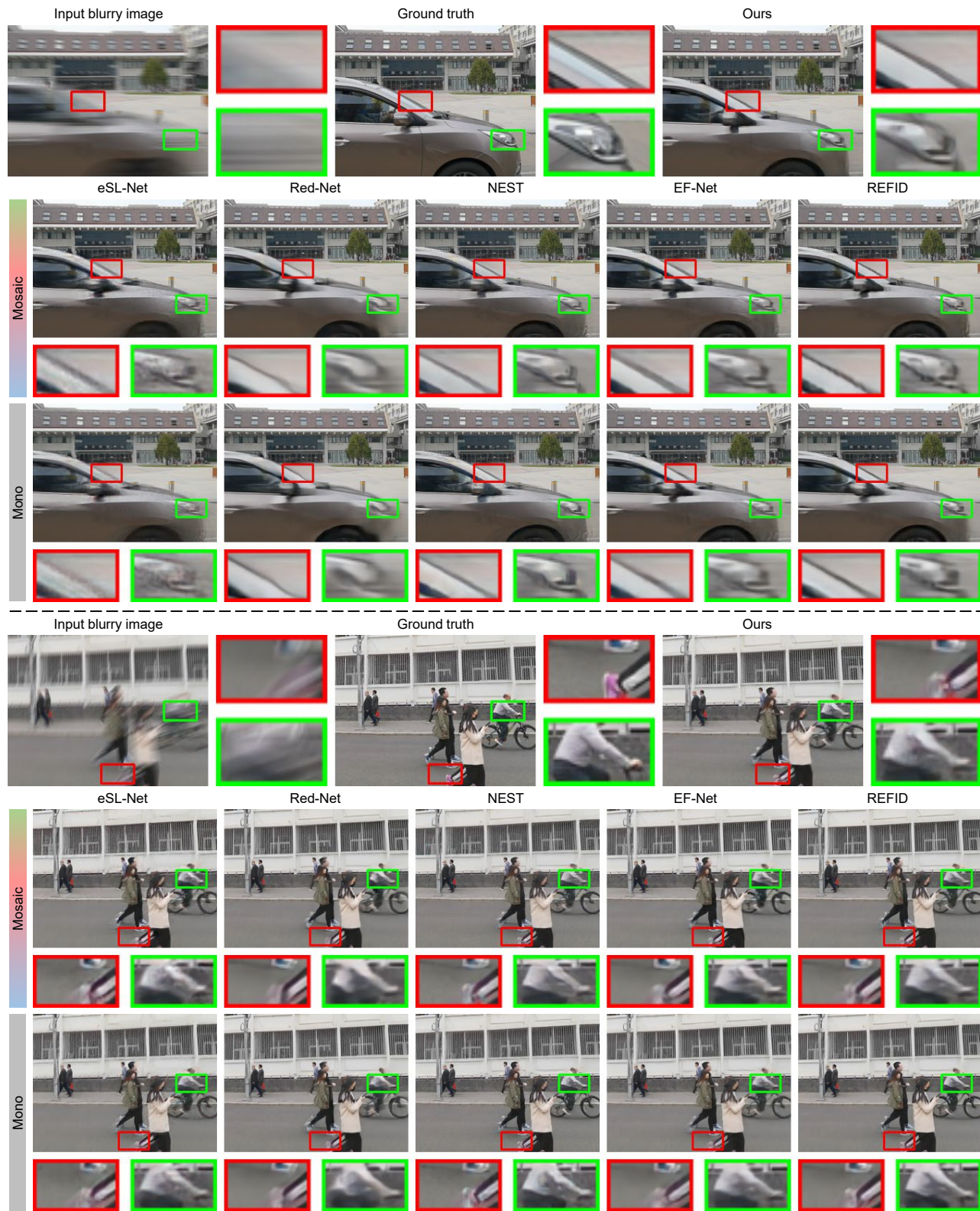


Figure 14: Image deblurring results on our C4E dataset. The labels “Mosaic” and “Mono” denote the networks are trained with different types of events.