Monochromatic Event Guided Image Deblurring with Event-triggering-aware Decomposition

Minggui Teng[1,2†] Boyu Li[1,2] Yixin Yang[1,2] Chu Zhou[3] Yan Chen[4] Jimmy S. Ren[4,5] Boxin Shi[1,2*]

[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University

[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[3] National Institute of Informatics [4] SenseTime Research [5] Hong Kong Metropolitan University

{minggui_teng, liboyu, yangyixin93, shiboxin}@pku.edu.cn

zhou_chu@hotmail.com {yanchenace, jimmy.sj.ren}@gmail.com

1. Ablation Studies

To verify the effectiveness of each component in our method, we conduct several ablation studies. The results are presented in Tab. 1. We evaluate the contribution of the position encoding by removing it (W/o PE). The impact of the edge map guidance is assessed by omitting this component (W/o EM). Additionally, the necessity of bi-directional event fusion is demonstrated by excluding this module (W/o BD). The impact of ImgColor-Net is shown by not enhancing the color terms (W/o CC). Finally, to validate our two-stage training strategy, we examine an end-to-end training approach (End2end). The results show that our complete model achieves the best performance.

Table 1. Quantitative results of ablation studies. The best performances are highlighted in **bold**.

	W/o PE	W/o EM	W/o BD	W/o CC	Add	End2end	Ours
PSNR	30.90	29.00	32.05	32.08	28.40	31.72	**33.08**
SSIM	0.9188	0.9248	0.9229	0.9317	0.8759	0.9289	**0.9377**

2. Additional Proof

In this section, we provide additional proof of bidirectional EDI model. As illustrated in [13], we can reverse event streams to bridge $\mathbf{S}_N$ with other frame $\mathbf{S}_i$ as:

$$\mathbf{S}_N = \mathbf{S}_i / \mathbf{E}(t_i, t_N) = \mathbf{S}_i \mathbf{E}'(t_i, t_N) \tag{1}$$

And then, we can bridge $\mathbf{B}_{\mathbb{g}}$ with $\mathbf{S}_N$, *i.e.*,

$$\begin{aligned} \mathbf{B}_{\mathbb{g}} &= \frac{\mathbf{S}_N}{T} \sum_{i=1}^{N} \mathbf{E}'(t_{i-1}, t_N) \\ &= \frac{\mathbf{S}_N}{T} \sum_{i=1}^{N} (\mathbf{E}'(t_{i-1}, t_i) + \mathbf{E}'(t_i, t_{i+1}) + \cdots + \mathbf{E}'(t_{N-1}, t_N)) \\ &= \frac{\mathbf{S}_N}{T} \sum_{i=1}^{N} i\mathbf{E}'(t_{i-1}, t_i). \end{aligned} \tag{2}$$

† This work is done during Minggui's internship at SenseTime.

* Corresponding author

Integrating Equation (1) and Equation (2), we can derive any arbitrary sharp frame at timestamp t_k:

$$\mathbf{S}' = \mathbf{B}_{\mathrm{g}} \frac{\sum_{i=k}^{N} \mathbf{E}'(t_{i-1}, t_i)}{\sum_{i=1}^{N} i\mathbf{E}'(t_{i-1}, t_i)} = \mathbf{B}_{\mathrm{g}} \cdot \mathcal{E}_{\mathtt{post}}. \tag{3}$$

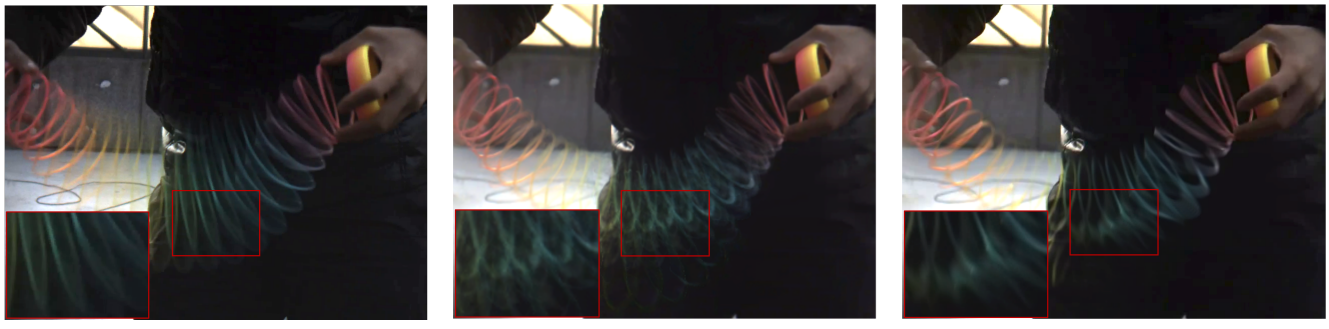
3. EvRGB-Deblur Dataset

Our hybrid camera system, shown on the left in Figure 1, is composed by three cameras: an event camera (Prophesee EVK4 HD) and two RGB cameras (Hikvision MV-CA050-12UC), paired with three beam splitters. For spatial calibration, we use a checkerboard to address homography and radial distortion across the three views. Temporal synchronization is achieved using an Arduino Uno Rev3 micro-controller board as the signal generator, which sends trigger signals to all cameras. The interface of signal generator and the GPIO ports of cameras are connected via cables. This hybrid system captures 10 groups of images from a variety of scenes featuring both ego-motion and object-motion, and the example scenes are shown on the right in Figure 1.

To better illustrate the distinction between synthetic and real domains of blurry images, we conduct an experiment that synthesizes blurry images from interpolated sharp videos. These synthesized images are then compared with our real-captured blurry image. As demonstrated in Figure 2, our real-captured image avoids unnatural blurs and artifacts in cases of large motion. Additionally, we display quantitative comparative results on our EvRGB-Deblur dataset in Tab. 2.

In Tab. 3, we compare our EvRGB-Deblur dataset with other publicly available event deblurring datasets [2, 4–6, 10, 11, 16]. Notably, Blur-DVS comprises two sub-datasets, which we refer to as Blur-DVS-`s` and Blur-DVS-`f`. For real-captured datasets (FEVD [6], EVRB [5], EventAid-B [2], and our EvRGB-Deblur dataset), we present typical scenarios of the each dataset in Figure 3. As shown in the examples, existing real-captured datasets do not specifically include colorful objects with diverse motion, which are essential for validating shape distortion and color bleeding issues. In the EventAid-B dataset, the

Figure 1. **Left**: Our hybrid camera system, which consists of three cameras: an event camera, and two RGB cameras with short and long exposure time, respectively. **Right**: Some scenes in the EvRGB-Deblur dataset.



(a) Real-captured **(b)** Synthetic-ev **(c)** Synthetic-img

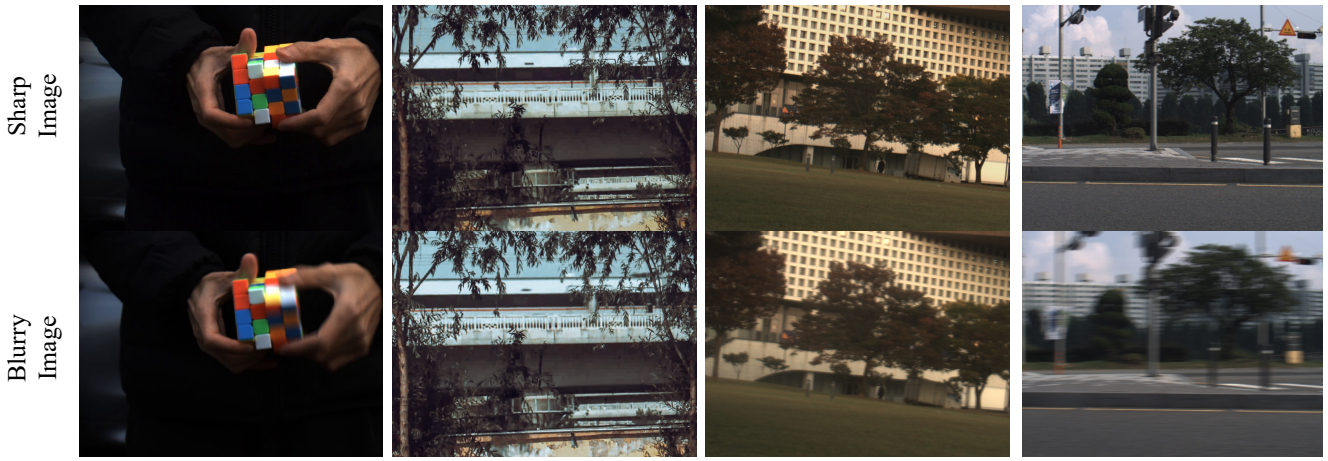
Figure 2. An example of blurry image synthesis. (a) Real-captured blurry image. (b) Synthetic blurry image interpolated by Timelens [14]. (c) Synthetic blurry image interpolated by RIFE [3]. Synthetic images exhibit unnatural blurs and artifacts in cases of large motion.

Table 2. Quantitative comparisons on the our EvRGB-Deblur dataset. The "Event" row specifies whether the methods require events as input (Yes [✓] or No [✗]).

Metric	NAFNet	Restormer	FFTformer	EDI	EFNet	NEST	REFID	MAENet	Ours
Event	✗	✗	✗	✓	✓	✓	✓	✓	✓
PSNR	27.09	27.06	27.12	24.24	27.25	20.03	27.23	24.79	**28.79**
SSIM	0.861	0.860	0.867	0.825	0.859	0.328	0.865	0.794	**0.876**

Table 3. Comparison of our collected EvRGB-Deblur dataset with other event deblurring datasets. The Blur and Sharp columns specify the image source, indicating whether it is synthetic (syn.), real-captured (real), or not applicable (n/a). The Resolution column indicates the resolution of the event sensor. The Image column represents whether the blurry images are grayscale or RGB.

Dataset	Blur	Sharp	Resolution	Image
Blur-DVS-s [4]	syn.	real	260×346	Gray
Blur-DVS-f [4]	real	n/a	260×346	Gray
REBlur [10]	syn.	real	260×346	Gray
HighREV [11]	syn.	real	1224×1632	RGB
MS-RBD [16]	real	n/a	192×288	Gray
EventAid-B [2]	real	real	620×835	RGB
FEVD [6]	real	real	768×1024	RGB
EVRB [5]	real	real	640×960	RGB
EvRGB-Deblur	**real**	**real**	624×840	RGB



(a) EvRGB-Deblur (b) EventAid-B (c) EVRB (d) FEVD

Figure 3. Examples from real-captured event-based deblurring datasets. (a) Our EvRGB-Deblur dataset captures object motion with colorful objects. (b) EventAid-B [2] has minimal motion. (c) EVRB [5] and (d) FEVD [6] primarily feature ego-motion blur.

captured motion is minimal, while the blur in the EVRB and FEVD datasets primarily results from ego-motion. In contrast, our dataset incorporates both ego-motion and object motion with colorful objects, enabling a comprehensive evaluation of deblurring performance in mitigating shape distortion and color bleeding artifacts.

4. Network Parameters

We list multiply–accumulate operations and network parameters of comparative image-based and event-based methods in the Tab. 4.

5. More Results on Real Dataset

In this section, we provide more qualitative comparisons among our method, NAFNet [1], Restormer [15], FFTformer [7], EDI [9], EFNet [10], NEST [13], REFID [11], and MAENet [12] on EvRGB-Deblur dataset, shown in Figure 4 and Figure 5. These comparisons are shown in Figure 4 and Figure 5. Across both sets of results, our method outperforms the other state-of-the-art methods. For example, in the first set of results in Figure 4 (first and second rows), our method recovers sharper

Table 4. The quantity of multiply-accumulate operations (MACs) and the count of network parameters (#Param) across image-based and event-baseds methods.

	NAFNet	Restormer	FFTformer	EFNet	NEST	REFID	Ours
MACs (G)	587	1128	1110	1018	4167	15570	13909
#Param (M)	67.8	26.1	14.9	8.5	19.7	15.8	30.5

details of the cartoon texture on the box. Additionally, in the first set of results in Figure 5, our deblurred result accurately preserves the edges of the slinky.

6. More Results on Synthetic Dataset

In this section, we provide more qualitative comparisons among our method, NAFNet [1], Restormer [15], FFTformer [7], EDI [9], EFNet [10], NEST [13], REFID [11], and MAENet [12] on REDS dataset [8], shown in Figure 6 and Figure 7. These comparisons are showcased in Figure 6 and Figure 7. Our method demonstrates superior performance compared to other state-of-the-art methods. In the second set of results in Figure 6 (third and fourth rows), our method successfully restores sharp and accurate edges of the text on the billboard, with minimal shape distortion. Furthermore, in the second set of results in Figure 7, our deblurred result shows significantly reduced color bleeding in the headscarf recovery.

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proc. of European Conference on Computer Vision*, 2022. 3, 4, 5, 6, 7, 8

[2] Peiqi Duan, Boyu Li, Yixin Yang, Hanyue Lou, Minggui Teng, Yi Ma, and Boxin Shi. EventAid: Benchmarking event-aided image/video enhancement algorithms with real-captured hybrid dataset. *arXiv preprint arXiv:2312.08220*, 2023. 2, 3

[3] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *Proc. of European Conference on Computer Vision*, 2022. 2

[4] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning event-based motion deblurring. In *Proc. of Computer Vision and Pattern Recognition*, pages 3320–3329, 2020. 2, 3

[5] Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. CMTA: Cross-modal temporal alignment for event-guided video deblurring. In *Proc. of European Conference on Computer Vision*, 2024. 2, 3

[6] Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. Frequency-aware event-based video deblurring for real-world motion blur. In *Proc. of Computer Vision and Pattern Recognition*, 2024. 2, 3

[7] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proc. of Computer Vision and Pattern Recognition*, 2023. 3, 4, 5, 6, 7, 8

[8] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study. In *Proc. of Computer Vision and Pattern Recognition Workshops*, 2019. 4, 7, 8

[9] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 3, 4, 5, 6, 7, 8

[10] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based fusion for motion deblurring with cross-modal attention. In *Proc. of European Conference on Computer Vision*, 2022. 2, 3, 4, 5, 6, 7, 8

[11] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. In *Proc. of Computer Vision and Pattern Recognition*, 2023. 2, 3, 4, 5, 6, 7, 8

[12] Zhijing Sun, Xueyang Fu, Longzhuo Huang, Aiping Liu, and Zheng-Jun Zha. Motion aware event representation-driven image deblurring. In *Proc. of European Conference on Computer Vision*, 2024. 3, 4, 5, 6, 7, 8

[13] Minggui Teng, Chu Zhou, Hanyue Lou, and Boxin Shi. NEST: Neural event stack for event-based image enhancement. In *Proc. of European Conference on Computer Vision*, 2022. 1, 3, 4, 5, 6, 7, 8

[14] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time lens: Event-based video frame interpolation. In *Proc. of Computer Vision and Pattern Recognition*, 2021. 2

[15] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. of Computer Vision and Pattern Recognition*, 2022. 3, 4, 5, 6, 7, 8

[16] Xiang Zhang, Lei Yu, Wen Yang, Jianzhuang Liu, and Gui-Song Xia. Generalizing event-based motion deblurring in real-world scenarios. In *Proc. of Computer Vision and Pattern Recognition*, 2023. 2, 3

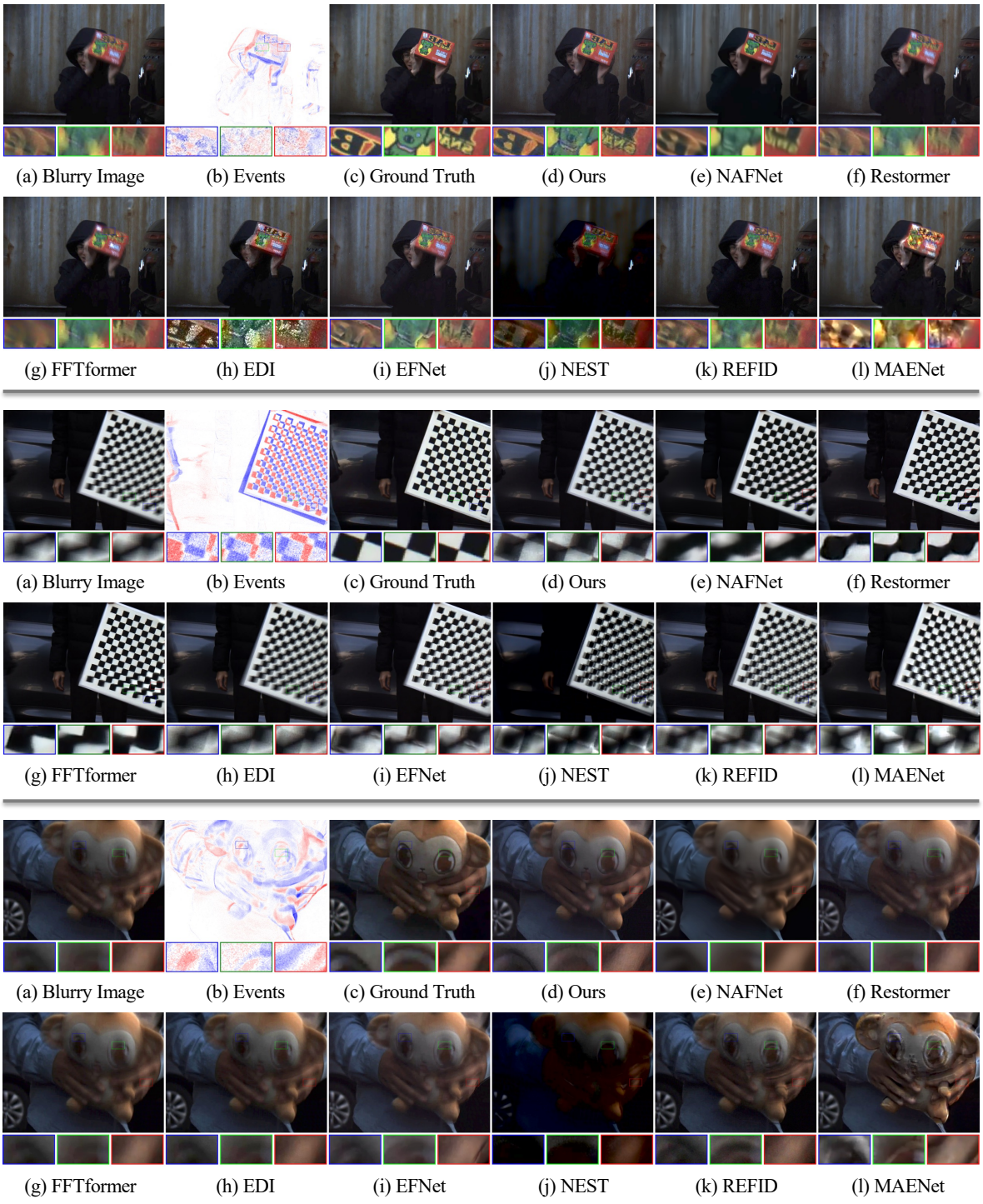


Figure 4. Visual quality comparison on real-captured EvRGB-Deblur dataset (Part I). (a) Blurry image. (b) Events. (c) Ground truth. (d)∼(l) Deblurred results of ours, NAFNet [1], Restormer [15], FFTformer [7], EDI [9], EFNet [10], NEST [13], REFID [11] and MAENet [12]. In the first set of results (first and second rows), our method recovers sharper details of the cartoon texture on the box.

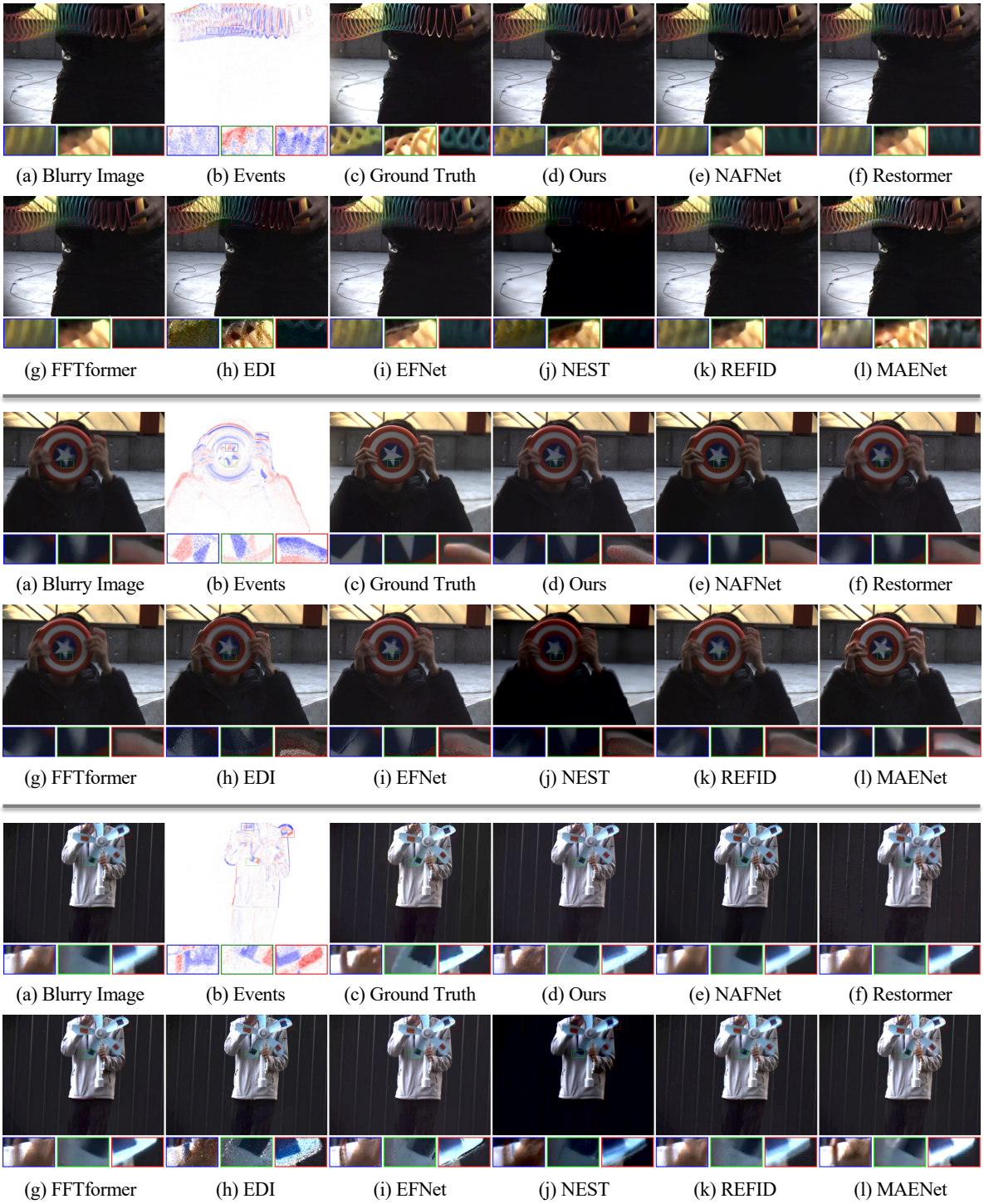


Figure 5. Visual quality comparison on real-captured EvRGB-Deblur dataset (Part II). (a) Blurry image. (b) Events. (c) Ground truth. (d)∼(l) Deblurred results of ours, NAFNet [1], Restormer [15], FFTformer [7], EDI [9], EFNet [10], NEST [13], REFID [11] and MAENet [12].. In the first set of results (first and second rows), our deblurred result accurately preserves the edges of the slinky.

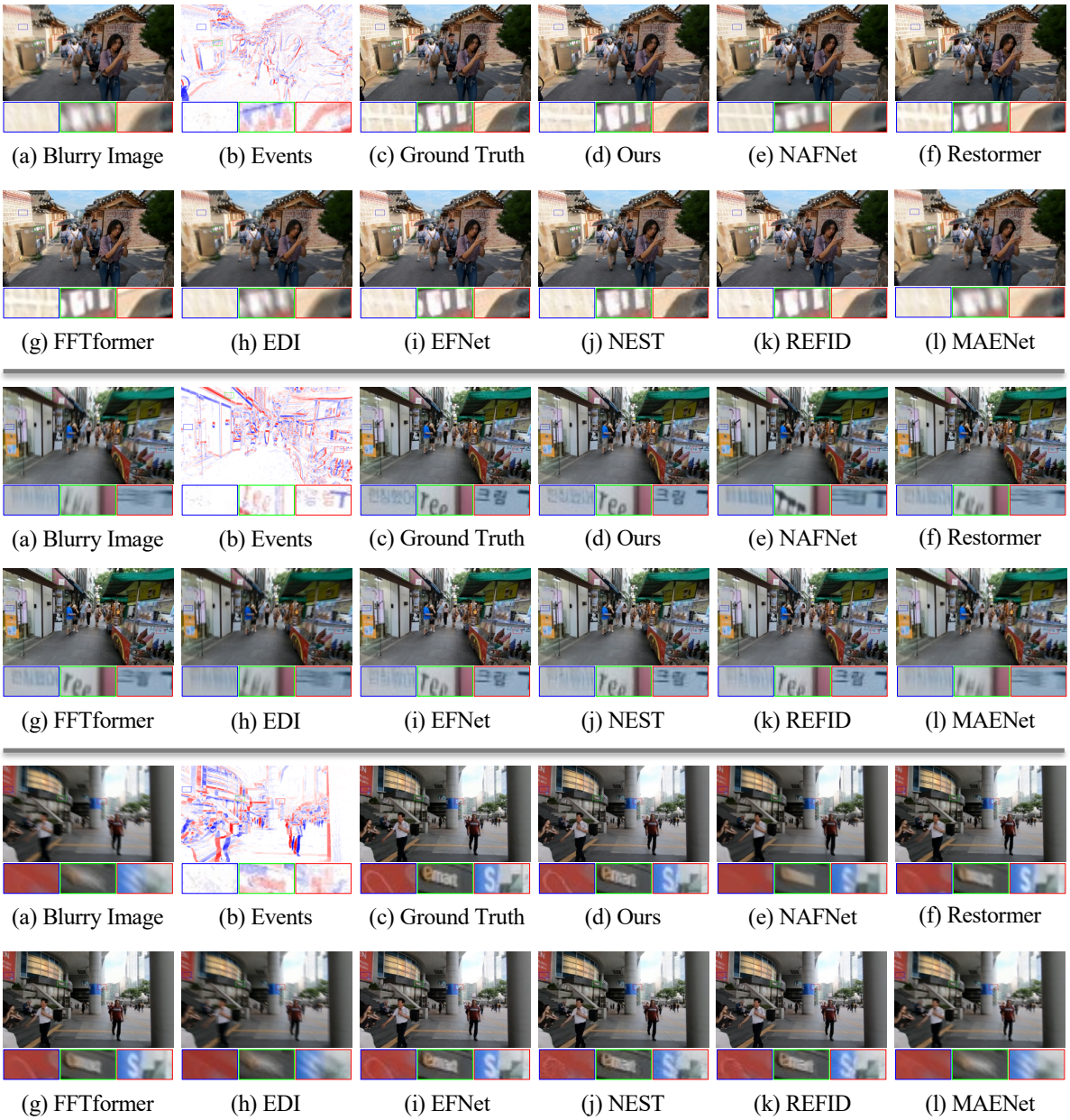


Figure 6. Visual quality comparison on REDS dataset [8] (Part I). (a) Blurry image. (b) Events. (c) Ground truth. (d)∼(l) Deblurred results of ours, NAFNet [1], Restormer [15], FFTformer [7], EDI [9], EFNet [10], NEST [13], REFID [11] and MAENet [12].. In the second set of results (third and fourth rows), our method successfully restores sharp and accurate edges of the text on the billboard, with minimal shape distortion.

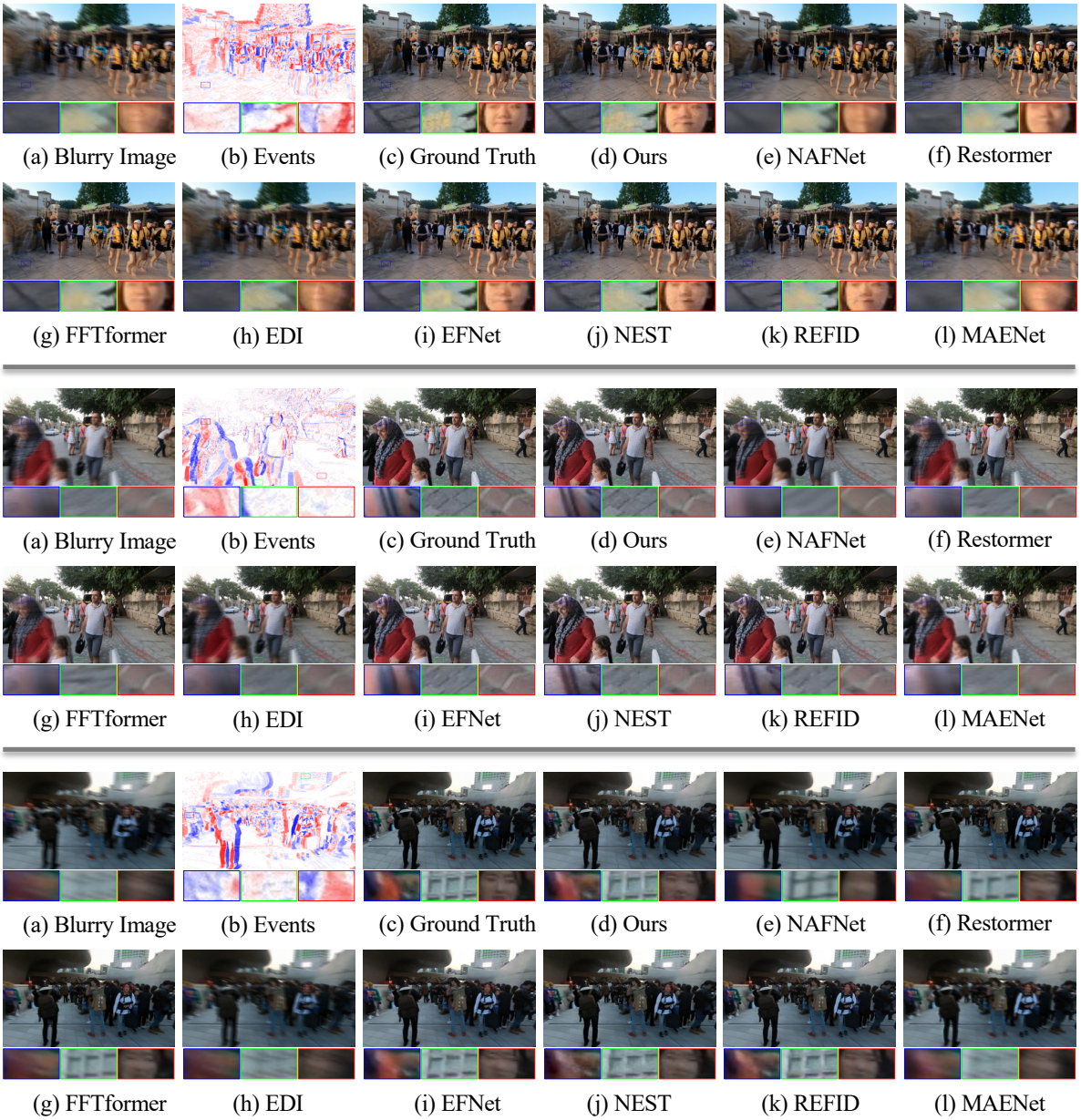


Figure 7. Visual quality comparison on REDS dataset [8] (Part II). (a) Blurry image. (b) Events. (c) Ground truth. (d)∼(l) Deblurred results of ours, NAFNet [1], Restormer [15], FFTformer [7], EDI [9], EFNet [10], NEST [13], REFID [11] and MAENet [12].. In the second set of results (third and fourth rows), our deblurred result shows significantly reduced color bleeding in the headscarf recovery.