Supplemental Material:
EvUnroll: Neuromorphic Events based Rolling Shutter Image Correction

Xinyu Zhou[1#] Peiqi Duan[1#] Yi Ma[1] Boxin Shi[1,2,3] ✉

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

[2]Institute for Artificial Intelligence, Peking University [3]Beijing Academy of Artificial Intelligence

7. Dataset

7.1. Gev-RS dataset

We compare our collected Gev-RS dataset with other publicly available RS correction datasets [1,2,6] in Table 4. As can be seen, our dataset has the advantage of high-frame-rate, which is helpful for synthesizing data closer to real RS images and event streams. We also consider the diversity of the samples by including building, road and park scenes (Fig. 10), with both ego-motion and object-motion.

Table 4: Comparison of our collected Gev-RS dataset with other RS correction datasets.

Dataset	Camera properties	Captured Resolution	Frame rate
Hedborg *et al.* [1]	Cell phone	1280×720	30
Fastec-RS [2]	High-speed camera	640×480	2400
BS-RSCD [6]	Hybird system	1280×920	15
Gev-RS	High-speed camera	1280×720	5700

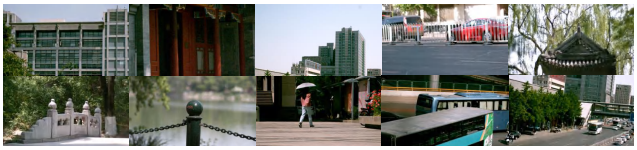


Figure 10: Examples of scenes in the Gev-RS dataset.

7.2. Real-captured data

To test EvUnroll on real-captured data, we build a hybrid camera system (Fig. 11) consisting of an RS camera and an event camera. For geometric calibration, we use a checkerboard to deal with homography and radial distortion between two views by the correction method used in JCD [6] and GEF [5]. For temporal synchronization, we shoot a high-precision stopwatch when capturing data, and synchronize the two signals by aligning the timestamp in the RS frame sequence and the event stream in post-processing.

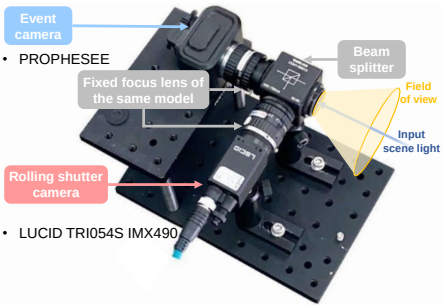


Figure 11: Our RS-Event hybrid imaging system.

8. GS2RS velocity connection

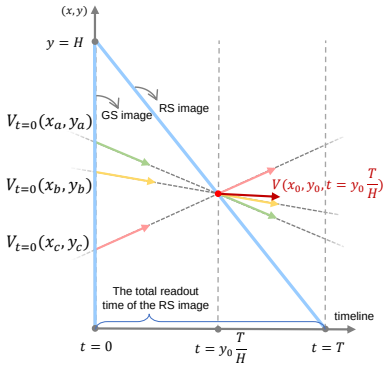


Figure 12: Illustration of GS2RS velocity connection.

We show the GS2RS velocity connection proposed in Sec. 3.3 of the main paper. As Fig. 12 shows, under the optical flow assumption, the velocity of a random pixel of RS image (*i.e.*, $V(x_0, y_0, t = y_0 \frac{T}{H})$) can be expressed as a vector mean of the set $\{V_{t=0}(x_a, y_a), V_{t=0}(x_b, y_b), V_{t=0}(x_c, y_c)\}$, which collects elements whose velocity direction passes through the point $(x_0, y_0, t = y_0 \frac{T}{H})$. Therefore, after we estimate $V_{t=0}$ by

# Contributed equally to this work as first authors

✉ Corresponding author: shiboxin@pku.edu.cn

Project page: https://github.com/zxyemo/EvUnroll

the GS2GS flow-Net, the velocity vector of each RS image pixel can be calculated by a mean process.

9. Qualitative ablation results

We provide qualitative ablation results in Fig. 13. Our image deblurring module outputs sharp images, which improves the performance of both the flow-based connection module and synthesis-based connection module. In flow case, there are black edges of corrected images due to the lack of prior information to complete occlusion regions. In syn. case, the synthesis-based connection module is better at reconstructing occluded regions, but causes artifacts in the RS distorted areas when event information is noisy. The refine module successfully avoids the major drawbacks of the flow-based correction result $I_{t=t_s}^{\text{GS\_F}}$ and the synthesis-based correction result $I_{t=t_s}^{\text{GS\_S}}$.

10. More ablation studies

10.1. Necessity of two pathways

Due to the inability to recover out-of-view or occluded regions, the result of the flow module suffers from black edge artifacts as shown in Fig. 13, resulting in a lower PSNR gain than the synthesis (Syn.) module as indicated in Table 3 in the main paper. Additionally, the flow module (68.4G FLOPs) has lower complexity than the synthesis module (120.6G FLOPs), which further quantifies the efficacy of each module.

10.2. Ablation on the refine module

Cases 1-4 in Table 3 are ablated without the refine module, and the comparison verify its effectiveness. Intuitively, we directly averaging the output of two modules, *i.e.* $I_{t=t_s}^{\text{GS\_F}}$ and $I_{t=t_s}^{\text{GS\_S}}$, and obtain a numerical degradation of 1.72 dB for PSNR on the Gev-RS dataset.

10.3. Ablation on the loss function

We remove each loss term from the overall loss function L, and report the impact in Table 5. Note that removing each loss term will cause a degradation in corrected images, our overall loss function achieve the best result.

Table 5: Ablation study on training loss.

loss case	L_c	$L_c + L_{tv}$	$L_c + L_p$
PSNR	29.01	29.20	30.01
SSIM	0.889	0.894	0.904

11. Additional results on Gev-RS dataset

More qualitative comparisons on the Gev-RS dataset are presented in Fig. 14. We also compare our method with the event-based image reconstruction method E2VID [3] in Fig. 15. E2VID [3] reconstructs monochrome images only from an event stream. We input the simulated event stream paired with Gev-RS data into the E2VID network, and output intensity images under a default parameter setting. Compared with our results, the reconstructed images of E2VID contain more artifacts caused by event noise and show few details in still objects (*e.g.*, the street light in the third example) due to that there are no events being triggered. The performance on generating consecutive GS frames is compared in the video accompanied with this supplemental material.

12. Comparison with Time Lens [4]

The event-based video interpolation method Time Lens [4] adopts a similar two-branch network design, but our "warping+synthesis" design is essentially from that of Time Lens. For concept, Time Lens uses two frames and events to achieve inter-frame interpolation, while EvUnroll uses one RS blurry image and events to achieve intra-frame deblurring and RS correction. For implementation, we explicitly calculate the row-wise optical (undistorted) flow between RS and GS, employ the attention of time-offset to enhance the perception of row-wise features, and introduce the Bi-LSTM to bidirectionally correlate features of adjacent time. For comparison, we slightly modify the Time Lens to adapt the RS task and perform the same experiment in Table 3 after retraining. EvUnroll achieves better PSNR (30.14 *vs.* 28.33) and SSIM (0.912 *vs.* 0.862) scores than Time Lens.

13. Impact of unsynchronization

As indicted in Sec. 5, when shooting high-speed motion scenes, our current prototype of a simple hybrid camera system cannot ensure ensure the microsecond-level synchronization of the RS image with the event stream. Here, we evaluate the performance of EvUnroll with an unsynchronized setting. As shown in Table 6, the performance of EvUnroll drops below the state-of-the-art methods [2, 6] if time shift reaches 5ms.

Table 6: Correlation of performance with time shift.

TS (ms)	-5	-4	-3	-2	-1	0	1	2	3	4	5
PSNR	23.60	24.91	26.02	27.18	28.68	30.14	29.30	27.36	26.18	25.27	23.55

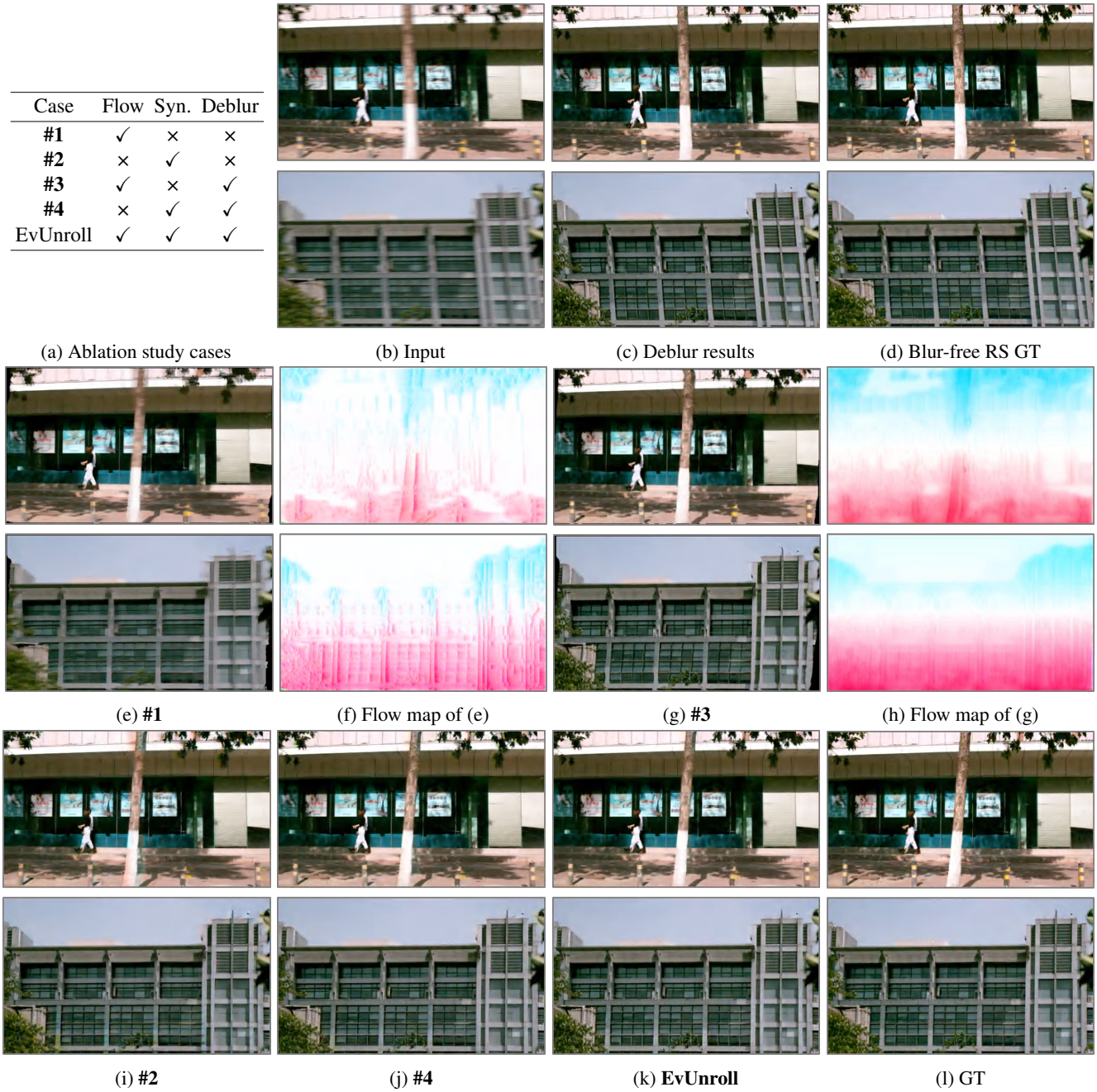
References

[1] Johan Hedborg, Per-Erik Forssén, Michael Felsberg, and Erik Ringaby. Rolling shutter bundle adjustment. In *Proc. of Computer Vision and Pattern Recognition*, 2012. 1

[2] Peidong Liu, Zhaopeng Cui, Viktor Larsson, and Marc Pollefeys. Deep shutter unrolling network. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 1, 2, 4

[3] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with

Case	Flow	Syn.	Deblur
#1	✓	×	×
#2	×	✓	×
#3	✓	×	✓
#4	×	✓	✓
EvUnroll	✓	✓	✓



(a) Ablation study cases (b) Input (c) Deblur results (d) Blur-free RS GT

(e) **#1** (f) Flow map of (e) (g) **#3** (h) Flow map of (g)

(i) **#2** (j) **#4** (k) **EvUnroll** (l) GT

Figure 13: Two qualitative comparison results on different cases of the ablation study. (a) Different module combinations proposed in Table 3 of the main manuscript.

an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2, 4

[4] 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

[5] Zihao Winston Wang, Peiqi Duan, Oliver Cossairt, Aggelos Katsaggelos, Tiejun Huang, and Boxin Shi. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 1

[6] Zhihang Zhong, Yinqiang Zheng, and Imari Sato. Towards rolling shutter correction and deblurring in dynamic scenes. In *Proc. of Computer Vision and Pattern Recognition*, 2021. 1, 2, 4

(a) RS Frame (b) JCD [6] (c) DSUN [2] (d) Ours (e) GS Frame

Figure 14: Qualitative comparisons on Gev-RS dataset. Even rows: absolute difference between the RS frame and correction results with the corresponding GS frame. (b)-(d) Correction results of different methods.

Figure 15: Qualitative comparisons with E2VID [3].