Supplemental Material:
Joint Filtering of Intensity Images and Neuromorphic Events for High-Resolution Noise-Robust Imaging

Zihao W. Wang[†#§] Peiqi Duan[‡#] Oliver Cossairt[†] Aggelos Katsaggelos[†] Tiejun Huang[‡] Boxin Shi[‡*]
[†]Northwestern University [‡] Peking University
Project page: https://sites.google.com/view/guided-event-filtering

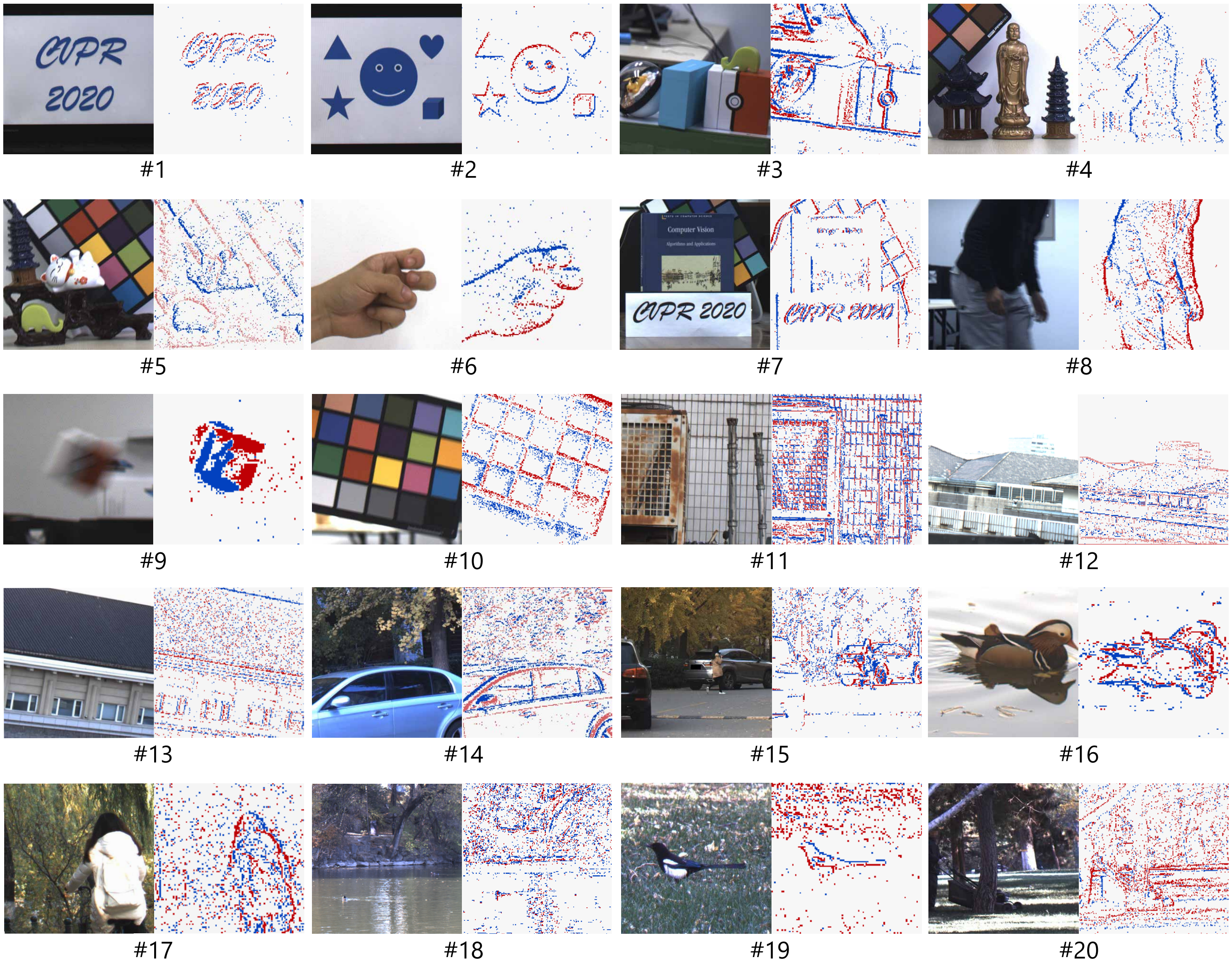


Figure 1: Our proposed RGB-DAVIS dataset. Shown images are screenshots of RGB videos (left) and event videos (right).

[#] Equal contribution. [*] Corresponding author.
[§] Part of this work was finished while visiting Peking University.

Table 1: Details of our RGB-DAVIS dataset

clip	# of images	# of events	indoor/ outdoor	camera motion	scene motion	description
#1	250	2.6M	indoor		✓	text displayed and animated on the monitor
#2	200	8.9M	indoor		✓	simple shapes displayed and animated on the monitor
#3	200	7.3M	indoor	✓		static objects with camera motion
#4	200	3.5M	indoor	✓		static objects with camera motion
#5	200	1.7M	indoor	✓		static objects with camera motion
#6	200	10.5M	indoor		✓	hand gestures
#7	150	2.4M	indoor	✓		textbook with background
#8	400	23.8M	indoor		✓	human body motion
#9	400	20.8M	indoor		✓	abruptly throwing an object
#10	400	21.8M	indoor		✓	color chart with hand-held motion
#11	200	3.8M	outdoor	✓		wall with grid structure
#12	190	3.6M	outdoor	✓		building with window
#13	200	4.3M	outdoor	✓		building
#14	400	9.3M	outdoor	✓	✓	car moving
#15	400	8.4M	outdoor	✓	✓	street with cars
#16	400	27.7M	outdoor	✓	✓	bird in a lake
#17	400	23.1M	outdoor	✓	✓	pedestrians walking on street
#18	400	22.6M	outdoor	✓		static objects with structured background
#19	400	20.7M	outdoor	✓	✓	bird on grass
#20	150	23.7M	outdoor	✓	✓	a weeding worker in a park

1. RGB-DAVIS dataset

We use our RGB-DAVIS camera system to collect various sequences of RGB-event video clips. As shown in Fig. 1, in total, there are 20 video clips. Both indoor and outdoor scenarios are captured. The scenes widely range from simple shapes to complex structures. All the clips involve camera motion and/or scene motion. The details are summarized in Table 1.

1.1. Geometric calibration

Here, we describe our calibration procedure between a 2448 × 2048 resolution machine vision camera (Point Grey Chameleon3) with a F/1.4 lens and a 180 × 190 resolution event camera (Davis240b) with a F/1.4 lens. The two cameras are physically co-located and they share a common viewing direction through a beam splitter (Thorlabs CCM1-BS013) with 50% splitting. In the calibration phase, the two cameras are kept stationary. Our geometric calibration is designed to geometrically transform event pixels to the machine vision camera image within the shared view. For this purpose, we display a *blinking* checkerboard pattern on a 13.9” 60Hz monitor. The monitor is placed around 2m away from the RGB-DAVIS imaging system to ensure both cameras can have full view of the checkerboard pattern. To form an event image, events are accumulated in a short time window (no longer than the blinking period of the pattern). On the machine vision camera, a 50fps video can be captured so as to select one frame with a stable checkerboard pattern. The keypoints can be extracted from the corners of the checkerboard images. Our 2D-based modeling includes a homography transformation estimated based on the central keypoints (inside the green box) and an anti-distortion transformation estimated based on all the keypoints. The distortion modeling is only for the event camera as the smart phone camera distortion has already been calibrated.

The homography is an affine transformation defined as a 3 × 3 matrix. Mathematically, it connects two sets of coordinates:

$$\mathbf{x}_i^I = \mathbf{H}\mathbf{x}_i^e, \tag{1}$$

where $\mathbf{x}_i^I = [x_i^I, y_i^I, 1]^T$ and $\mathbf{x}_i^e = [x_i^e, y_i^e, 1]^T$ are the homogeneous coordinates in the intensity image and the event image, respectively. $\mathbf{H}$ can be solved using a minimal of 4 points. In our implementation, we use more than 4 points (8 points) to avoid linear degeneration. The homography is estimated only applying to the central green area.

We use a radially symmetric model for distortion modeling:

$$\begin{aligned} x^u - x_c^e &= (x^e - x_c^e)(1 + k_x r^2) \\ y^u - y_c^e &= (y^e - y_c^e)(1 + k_y r^2), \end{aligned} \tag{2}$$

where x_c^e and y_c^e denote the distortion center. x^u and y_u are the distortion corrected coordinates. $r^2 = (x^e)^2 + (y^e)^2$. k_x and k_y are the distortion coordinates to be fitted.

The two transformations are enclosed in a single RANSAC loop to find the optimal solution that leads to the overall error. The error is computed on the intensity domain. Some unavoidable factors, such as slight deformation caused by camera movement and fluctuations in camera frame rate, still affect the alignment of two camera views. In dataset clips, video frames are fine-tuned by bicubic to match 8× resolution of the event camera.

2. Comparison between CM and JCM

In addition to the summarized results presented in Fig. 3 of the main manuscript, we present additional visual comparisons in Fig. 2 and Fig. 3. We capture image patches and use preset flow vectors to simulate events within a 50ms time window. In this simulation, the bipolar thresholds are set as $\epsilon_p = 0.2$ and $\epsilon_n = -0.2$, and standard deviation of the threshold noise is $\sigma_e = 0.03$. The polarity of events is not used during the computation of warped histogram [1]. The motion compensated results are shown in the subfigures (b) and (c) of Fig. 2 and Fig. 3. From the comparisons between subfigures (d) and (e) we can observe that the contrast maps of JCM have faster fall-off departing from the peaks than CM. The peak values also showcase more accurate flow estimation results for JCM. In terms of speed and efficiency, the JCM almost shares the same computational cost as CM, while achieves more accurate flow.

3. Additional results of GEF

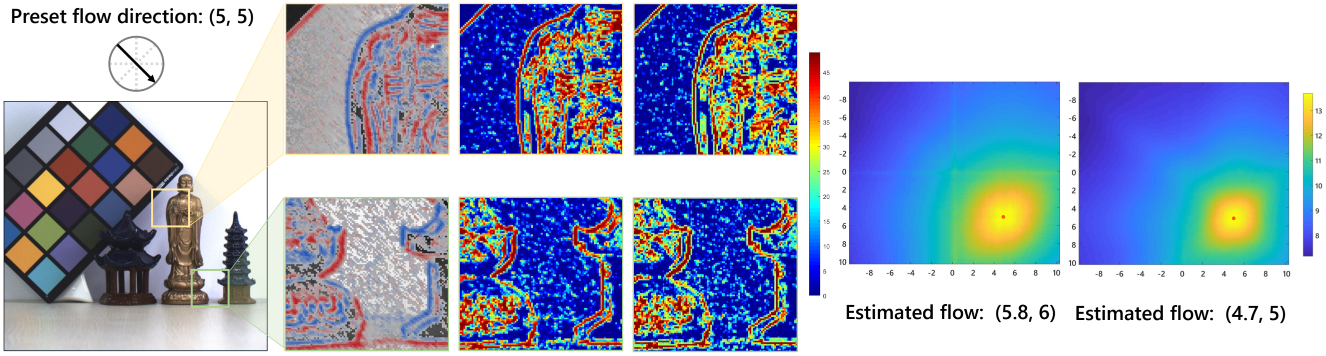
In this section, we present validation results on publicly available datasets, *e.g.*, DAVIS [6] and Color Event Dataset (CED) [7]. We also provide detailed explanation of our guided event super resolution processing steps and results.

3.1. Results on DAVIS dataset

The DAVIS [6] is a dataset containing intensity images and events at the same resolution (240 × 180). Here, we present additional denoise results for GEF with datas from the DAVIS dataset [6]. Two examples of events denoising are shown in Fig. 4 and Fig. 5. We compare the denoising perforemance between Liu *et al.* [5], EV-gait [9] and the proposed GEF, besides, the results of three candidate fliters of the GEF are also compared. Finally, we choose the MS-JF [8] as the optimal filter for the GEF.

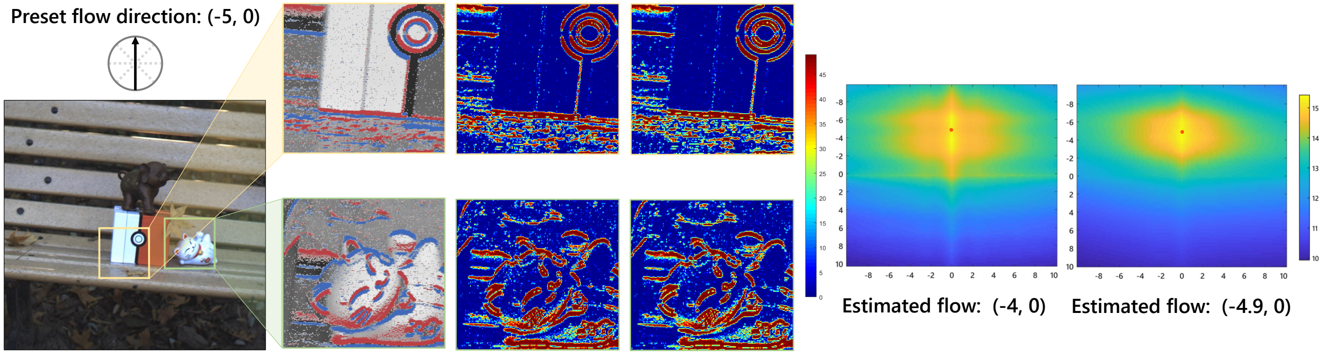
3.2. Results on Color Event Dataset (CED)

GEF can be applied to color image and events. We experiment on the recently released color event dataset [7]. The captured raw image is first demosaicked using the standard gradient-corrected linear interpolation to obtain RGB image. The flow is computed based on the grayscale image (converted from RGB) and all the captured events, which means the same flow is used across all three channels. The guided filtering process is run separately for each channel.



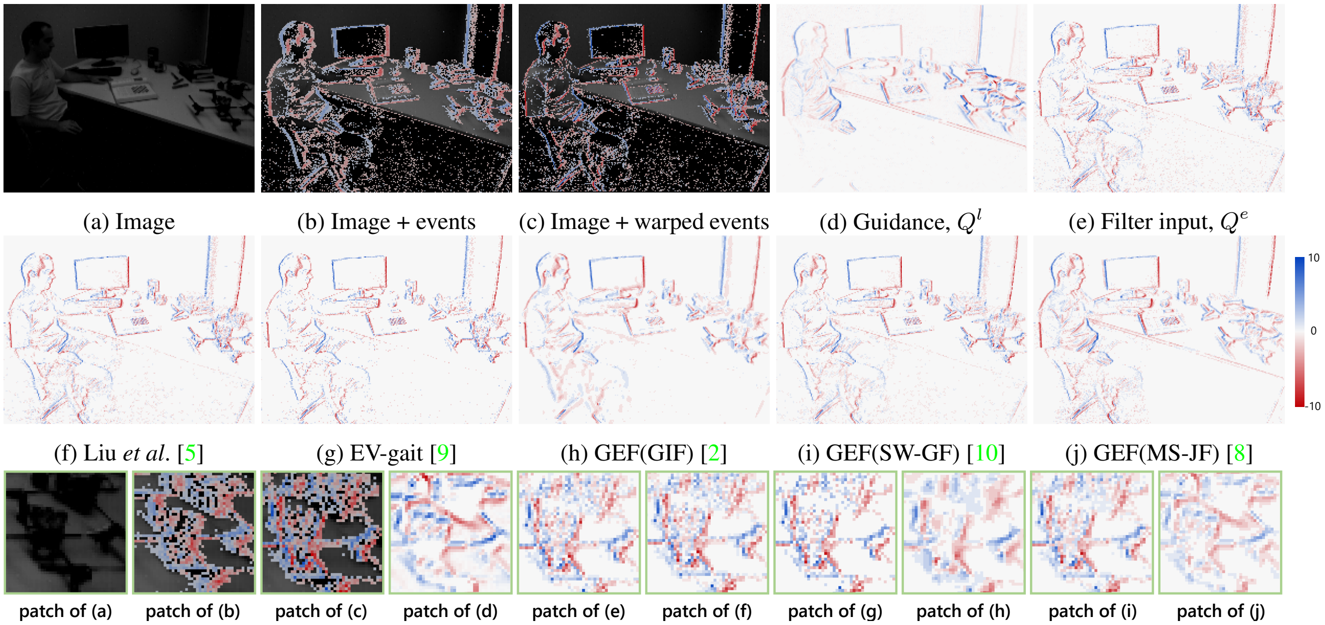
(a) image patches and events (b) CM (c) JCM (d) CM contrast map (e) JCM contrast map

Figure 2: First example of flow estimation. (b) and (c) denote the warped event image of the contrast maximization situation of CM and JCM. The red dot of (d) and (e) denote the point of the preset flow.



(a) image patches and events (b) CM (c) JCM (d) search map for CM (e) search map for JCM

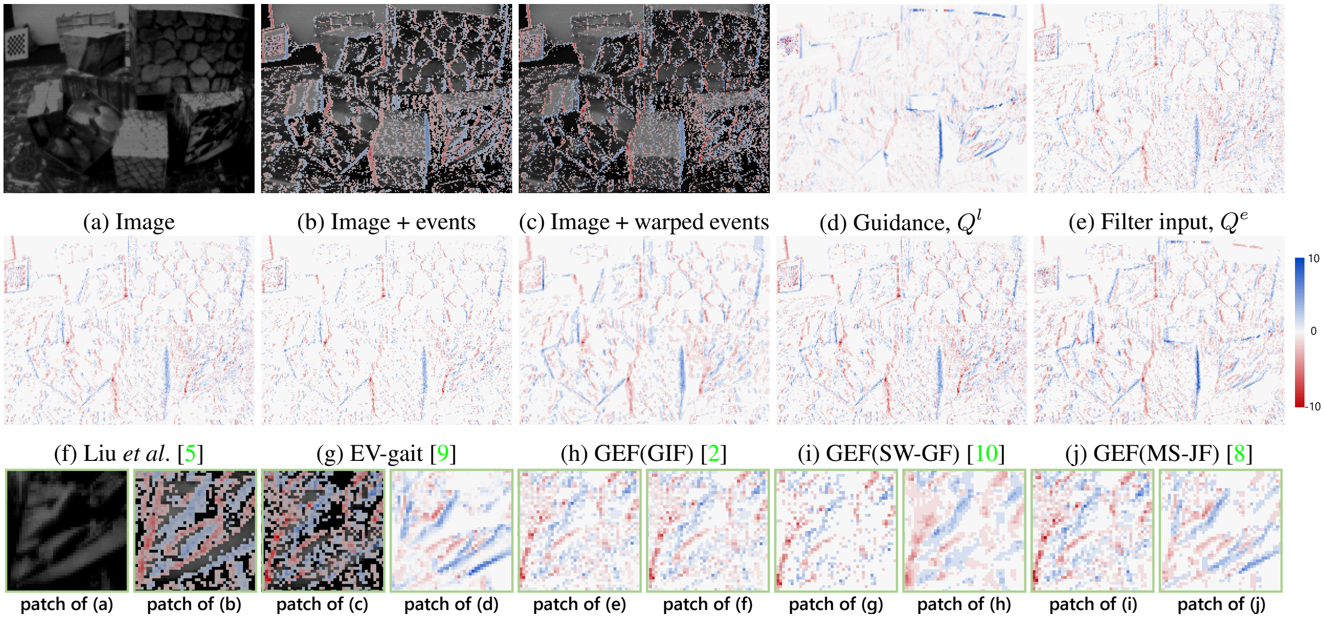
Figure 3: Second example of flow estimation. (b) and (c) denote the warped event image of the contrast maximization situation of CM and JCM. The red dot of (d) and (e) denote the point of the preset flow.



(a) Image (b) Image + events (c) Image + warped events (d) Guidance, Q^l (e) Filter input, Q^e

(f) Liu *et al*. [5] (g) EV-gait [9] (h) GEF(GIF) [2] (i) GEF(SW-GF) [10] (j) GEF(MS-JF) [8]

Figure 4: First denoise example of GEF on DAVIS [6] dataset. (b) An image overlaid with events (no warping); (c) An image overlaid with warped events (warped by JCM); (d) Q^l as filter guidance; (e) Q^e as filter input; (f-g) denoise output using Liu *et al*. [5] and EV-gait [9]; (h-j) GEF output using (h) GIF [2], (i) side-window guided filtering (SW-GF) [10], and (j) mutual-structure joint filtering (MS-JF) [8].



(a) Image (b) Image + events (c) Image + warped events (d) Guidance, Q^l (e) Filter input, Q^e

(f) Liu *et al.* [5] (g) EV-gait [9] (h) GEF(GIF) [2] (i) GEF(SW-GF) [10] (j) GEF(MS-JF) [8]

patch of (a) patch of (b) patch of (c) patch of (d) patch of (e) patch of (f) patch of (g) patch of (h) patch of (i) patch of (j)

Figure 5: Second denoise example of GEF on DAVIS [6] dataset. (b) An image overlaid with events (no warping); (c) An image overlaid with warped events (warped by JCM); (d) Q^l as filter guidance; (e) Q^e as filter input; (f-g) denoise output using Liu *et al.* [5] and EV-gait [9]; (h-j) GEF output using (h) GIF [2], (i) side-window guided filtering (SW-GF) [10], and (j) mutual-structure joint filtering (MS-JF) [8].

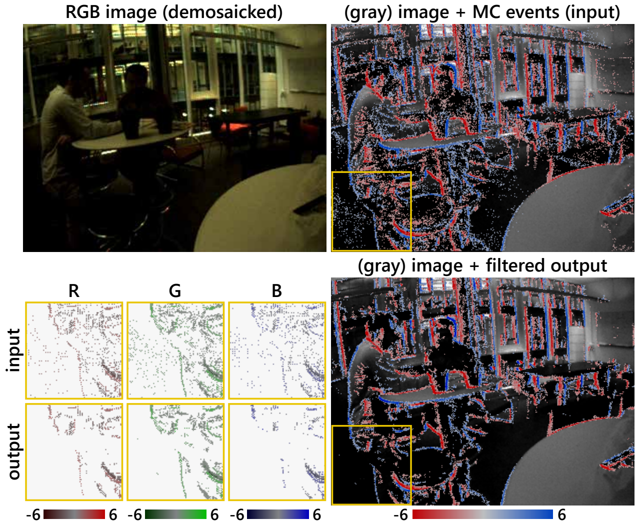

RGB image (demosaicked)
(gray) image + MC events (input)
(gray) image + filtered output
R
G
B
input
output
-6
6
-6
6
-6
6
-6
6

Figure 6: GEF applied on color event dataset [7].

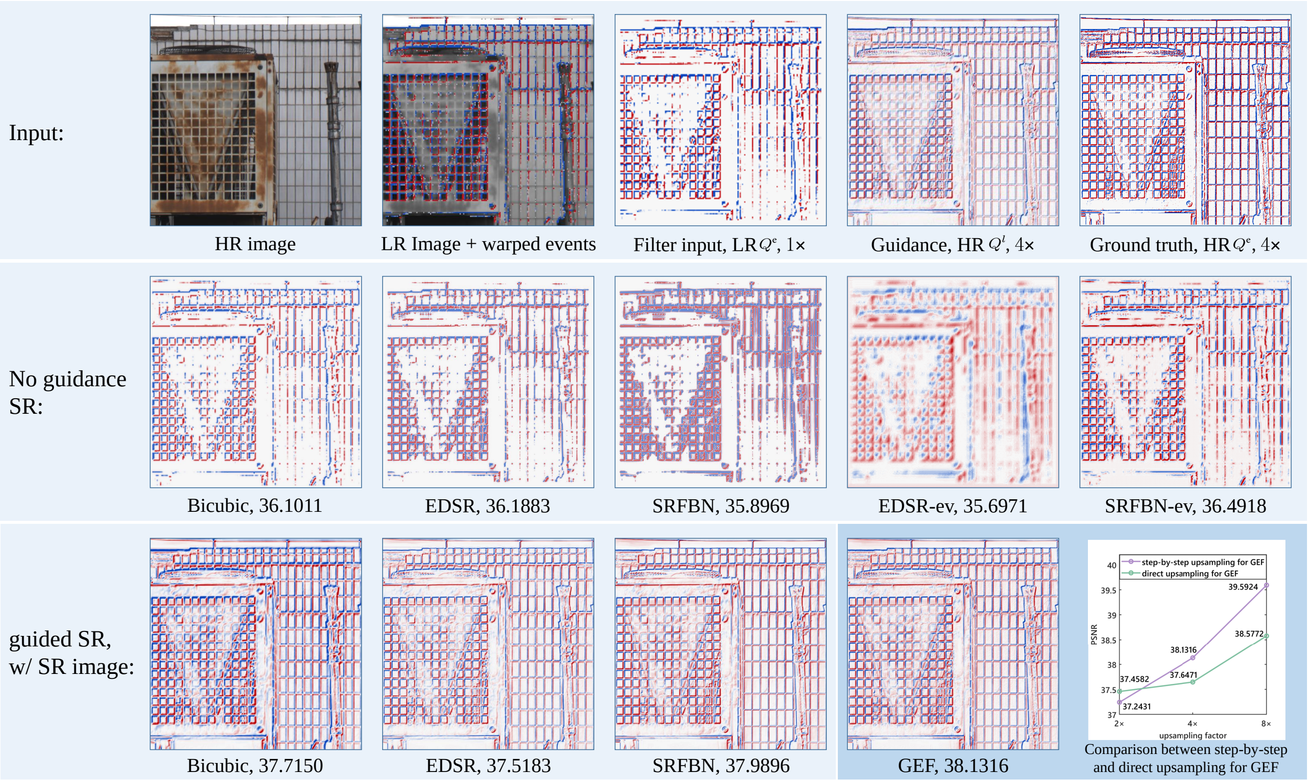
The results are shown in Fig. 6. From the results we can see the effectiveness of GEF denoising across all three channels, but with different performance for each. The green channel preserves more details than the red and blue channels (zoomed patches in Fig. 6).

4. Results for event super resolution

In this section, we describe our event super resolution (SR) processing steps. Figure 7 and Fig. 8 present comparison results for 4× SR and 8× SR, respectively. In both figures, *No guidance SR* includes bicubic upsampling, state-of-the-art SR algorithms EDSR [4] and SRFBN [3], both w/o and w/ re-training. The re-trained models are denoted as EDSR-ev and SRFBN-ev, respectively. We prepare 700 HR-LR event image pairs simulated from nature image dataset [3] to re-train the models.

The *guided SR, w/ SR image* corresponds to the approaches with intensity images as guidance. The same joint filtering algorithm is applied between the super-resolved intensity image and the event image. PSNR values indicate that the guided SR results are better than no guidance SR results. In the guided SR category, HR image as guidance provides the highest performance, both quantitatively and qualitatively.

For guided SR with HR image, we compare two strategies. (1) Step-by-step upsampling for GEF. In order to obtain the Q^o (filtered output image) at 2× scale, the filter output image at 1× scale is first bicubically upsampled, and then serves as Q^e at 2× to perform guided filtering. For 4× and 8×, this procedure is iteratively applied. (2) directly upsampling for GEF. For 2×, 4× and 8× upsampling, the Q^e at 1× is first bicubically upsampled to corresponding scales and then filter with Q^l computed at the same scales. The comparison results are presented at the bottom right corner of Fig. 7 and Fig. 8. As can be seen that Strategy (1) results in higher PSNR values for all 4× and 8× cases. In the joint filtering process, with an i7-8700K CPU, the average runtime is about 0.2s for 2× upsampling a 180×190 frame and 8s for 8×.

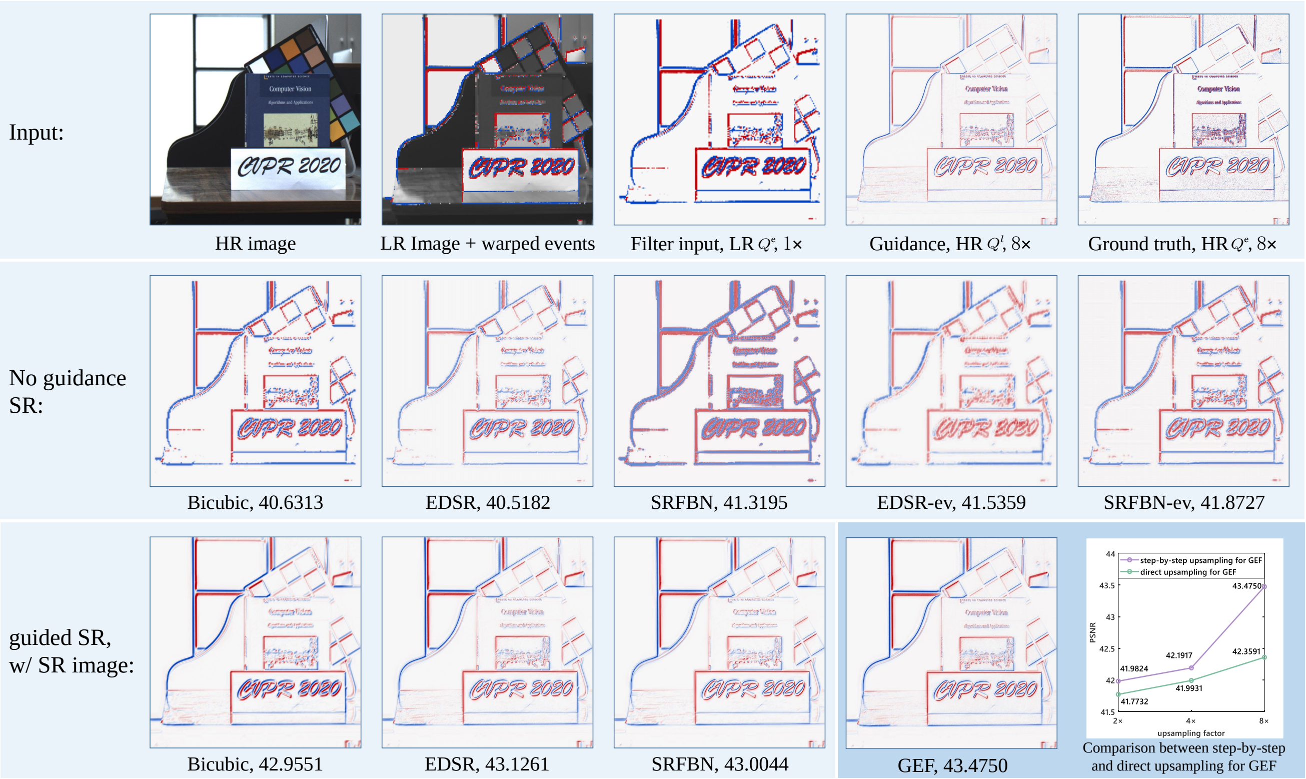


HR image
LR Image + warped events
Filter input, LR Q^e, 1×
Guidance, HR Q^l, 4×
Ground truth, HR Q^e, 4×

Bicubic, 36.1011
EDSR, 36.1883
SRFBN, 35.8969
EDSR-ev, 35.6971
SRFBN-ev, 36.4918

Bicubic, 37.7150
EDSR, 37.5183
SRFBN, 37.9896
GEF, 38.1316
Comparison between step-by-step and direct upsampling for GEF

Figure 7: First example of super-resolution results for $4\times$ upsampling. The numbers indicate the PSNR values.



HR image
LR Image + warped events
Filter input, LR Q^e, 1×
Guidance, HR Q^l, 8×
Ground truth, HR Q^e, 8×

Bicubic, 40.6313
EDSR, 40.5182
SRFBN, 41.3195
EDSR-ev, 41.5359
SRFBN-ev, 41.8727

Bicubic, 42.9551
EDSR, 43.1261
SRFBN, 43.0044
GEF, 43.4750
Comparison between step-by-step and direct upsampling for GEF

Figure 8: Second example of super-resolution results for $8\times$ upsampling. The numbers indicate the PSNR values.

References

[1] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3867–3876, 2018. 3

[2] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2012. 4, 5

[3] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3867–3876, 2019. 5

[4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, pages 136–144, 2017. 5

[5] Hongjie Liu, Christian Brandli, Chenghan Li, Shih-Chii Liu, and Tobi Delbruck. Design of a spatiotemporal correlation filter for event-based sensors. In *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 722–725, 2015. 3, 4, 5

[6] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International Journal of Robotics Research*, 36(2):142–149, 2017. 3, 4, 5

[7] Cedric Scheerlinck, Henri Rebecq, Timo Stoffregen, Nick Barnes, Robert Mahony, and Davide Scaramuzza. Ced: Color event camera dataset. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 0–0, 2019. 3, 5

[8] Xiaoyong Shen, Chao Zhou, Li Xu, and Jiaya Jia. Mutual-structure for joint filtering. In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. 3, 4, 5

[9] Yanxiang Wang, Bowen Du, Yiran Shen, Kai Wu, Guangrong Zhao, Jianguo Sun, and Hongkai Wen. Ev-gait: Event-based robust gait recognition using dynamic vision sensors. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 3, 4, 5

[10] Hui Yin, Yuanhao Gong, and Guoping Qiu. Side window guided filtering. *Signal Processing*, 165:315 – 330, 2019. 4, 5