Supplemental Material:
EventZoom: Learning to Denoise and Super Resolve Neuromorphic Events

Peiqi Duan† Zihao W. Wang‡ Xinyu Zhou† Yi Ma† Boxin Shi†§
†NELVT, Department of Computer Science and Technology, Peking University
‡Department of Computer Science and Engineering, Northwestern University
§Institute for Artificial Intelligence, Peking University

7. Architectural alternatives

We considered and compared several alternatives in designing EventZoom. The comparisons include:

- Losses: we compared with the perceptual loss [2] and the $L1$ loss.
- Partial 3D convolution [3] in replacement of regular 3D convolution.
- A spatially-adaptive feature aggregation mechanism, SPADE [4] in replacement of simple feature concatenation.

Table 1 records the minimal validation loss values and the average run times per channel of an event stack. The experimental results show that the partial convolution variant shares similar performance with EventZoom, perceptual loss and $L1$ loss variants improve the running speed but decline the reconstruction quality. Besides, the SPADE variant obtains both the worst reconstruction quality and run time values. In order to qualitatively compare the performance of different variants, we show three sets of comparison figures in Fig. 1 and Fig. 2. As can be seen that the perceptual loss and $L1$ loss, which are often used in image super resolution tasks, can not take on event super resolution task. The SPADE variant is failed to transfer the semantic information of images into the E2I module to reconstruct the event frame better.

In addition, the qualitative comparison results between w/ and w/o image features in E2I module are also presented in Fig. 1 and Fig. 2. The comparison between subfigure (g) and (h) shows that by incorporating an E2I module to leverage HR and LR image features, EventZoom can reconstruct sharper and clearer event frames.

Table 1: Ablation on Lossthe several architectural variants.

	MSE	runtime
Partial Convolution [3]	0.056	0.065s
Perceptual Loss [2]	0.064	0.061s
L1 Loss	0.071	0.058s
SPADE [4]	0.074	0.093s
EventZoom (MSE)	0.057	0.064s

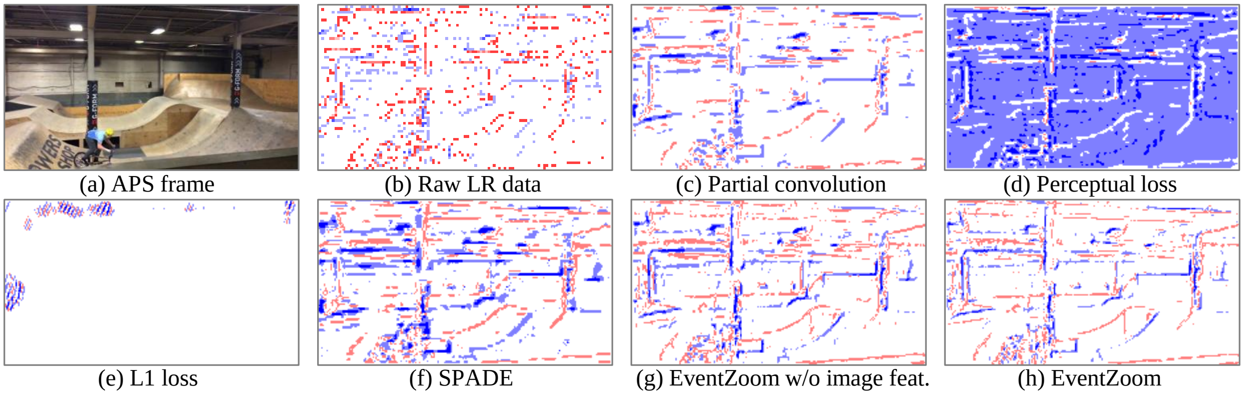
8. Additional results of EventZoom

8.1. The RPMD scores comparison with EDnCNN

The Relative Plausibility Measure of Denoising (RPMD) is an objective metric for benchmarking DVS denoising proposed in EDnCNN [1]. This method infers the log–likelihood probability of each event by combining the intensity from APS and the camera motion captured by IMU, and uses RPMD score to benchmark denoising performance by comparing denoising labels to the marked probabilities. We tested the middle 2% temporal window of the 14 out of 16 sequences except Scene-1 and Scene-16, because of the impact of severe bad pixels. The RPMD scores are shown in Table 2, smaller values indicate better performance. As can be seen that the EventZoom achieved competitive result by taking less time compared to EDnCNN. EventZoom performed poorly in scene-2, because this scene had highly textured scene contents and our model eliminated them as noise, which led to a large RPMD score.

8.2. Qualitative comparison with GEF

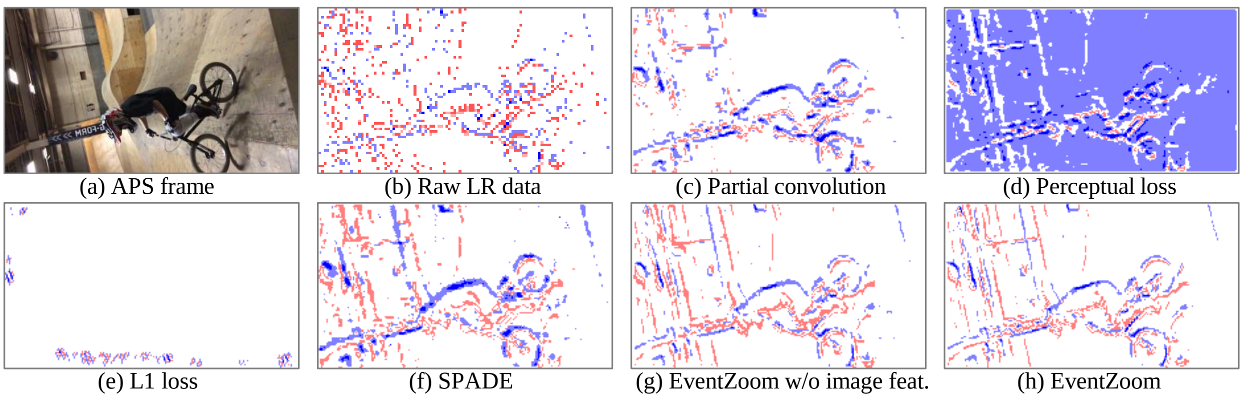
We compare EventZoom with GEF [5] for $2\times$ and $4\times$ event-to-event SR on RGB-DAVIS dataset [5], which contains 1520×1440 HR images and corresponding 190×180 LR event data. The qualitative results are shown in Fig. 3. We can find that GEF's performance relies heavily on the quality of HR images. The local regions where HR images are blurry or in lack of spatial features are suffered compromised quality.



(a) APS frame (b) Raw LR data (c) Partial convolution (d) Perceptual loss

(e) L1 loss (f) SPADE (g) EventZoom w/o image feat. (h) EventZoom

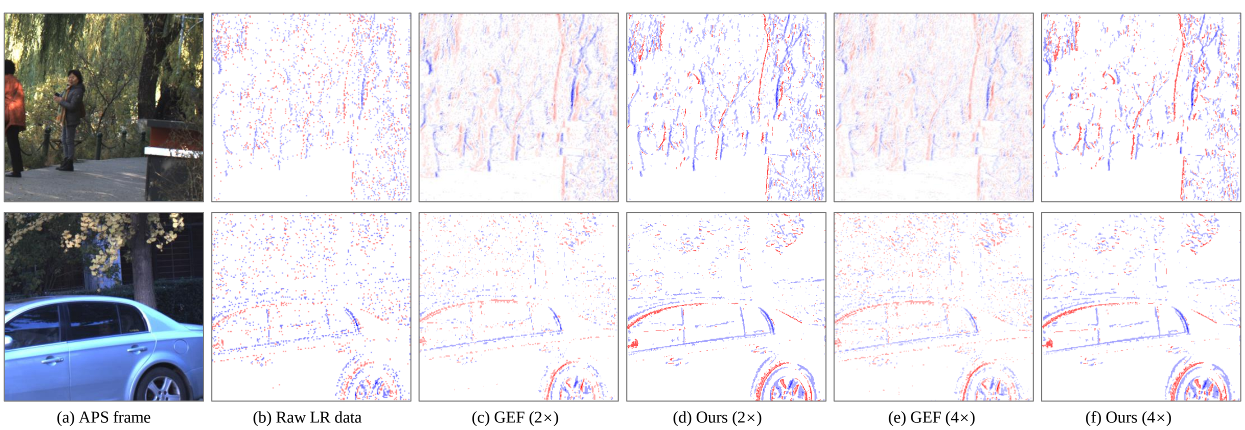
Figure 1: Qualitative comparison results on the several architectural variants. (First example)



(a) APS frame (b) Raw LR data (c) Partial convolution (d) Perceptual loss

(e) L1 loss (f) SPADE (g) EventZoom w/o image feat. (h) EventZoom

Figure 2: Qualitative comparison results on the several architectural variants. (Second example)



(a) APS frame (b) Raw LR data (c) GEF (2×) (d) Ours (2×) (e) GEF (4×) (f) Ours (4×)

Figure 3: Qualitative comparison results on RGB-DAVIS dataset [5] between GEF and EventZoom

Table 2: RPMD [1] scores comparison between EDnCNN [1] and EventZoom

	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	average
EDnCNN	20.12	22.53	7.34	5.29	15.40	13.06	16.13	28.59	5.43	34.35	16.97	2.25	6.41	30.14	16.00
Ours	35.71	17.81	9.56	7.41	12.78	11.39	16.52	30.24	5.31	40.33	18.98	2.63	8.44	27.90	17.50

References

[1] R Baldwin, Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. Event probability mask (epm) and event denoising convolutional neural network (edncnn) for neuromorphic cameras. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 1701–1710, 2020. 1, 2

[2] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016. 1

[3] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Eur. Conf. Comput. Vis.*, 2018. 1

[4] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. *CVPR*, 2019. 1

[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 *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2020. 1, 2