Denoising Event Data with NN

Tianshu Kuai, Yan Ma, Yihan Ni
University of Toronto

Motivation

Event Camera

- Detect pixel-wise binary brightness changes in the scene and output asynchronous sequences of "events"
- Event Stream: $E = \{e_k\}_{k=1}^{N} = \{(x_k, y_k, t_k, p_k)\}_{k=1}^{N}$
- Advantages: High dynamic range, microsecond-level temporal resolution, and no motion blur
- Applications: well-suited for high-speed use cases such as driving scenarios

Event Data Denoising

- Problem: Methods for denoising RGB images cannot be directly used on noisy event data due to the difference in data representations
- Our goal: To explore deep neural networks' capabilities in terms of denoising event data, and evaluate the performance of denoising on classification task

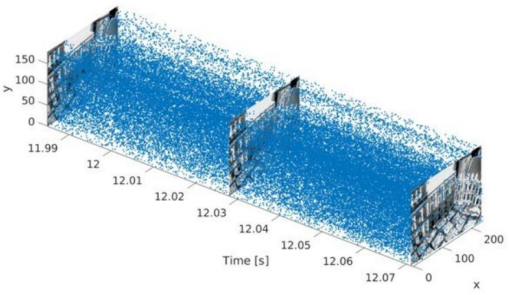

Figure 1: 3D visualization of an event stream.

Related Work

Event Voxel Representation[1]

- Divide event streams into portions with equal temporal length and accumulate them into spatio-temporal voxels
- Events of opposite polarities are handled separately
- The value of each voxel represents the number of occurred events

Event Denoising – EventZoom[2]

- Built upon the 3D U-Net backbone
- Raw events are voxelized to 3D tensors as the input to the denoising network

Datasets

- DVS128 Gesture Dataset[3] captures 11 different types of human gestures movements
- N-Caltech101 Dataset[4] contains event data of 101 different classes of static objects

References

[1] Gehrig, Daniel, et al. "End-to-end learning of representations for asynchronous event-based data." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
[2] Duan, Peiqi, et al. "EventZoom: Learning to denoise and super resolve neuromorphic events." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[3] Amir, Arnon, et al. "A low power, fully event-based gesture recognition system." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[4] Orchard, Garrick, et al. "Converting static image datasets to spiking neuromorphic datasets using saccades." Frontiers in neuroscience. 2015.

Proposed Denoising Pipeline

- Raw event stream is converted to event voxels for both polarities
- 3D U-Net processes and outputs clean voxels of the same dimension as the input voxels for downstream classification
- In each down-sampling block, the spatial data dimension is halved, and the number of channels doubles. The process is reversed during up-sampling blocks

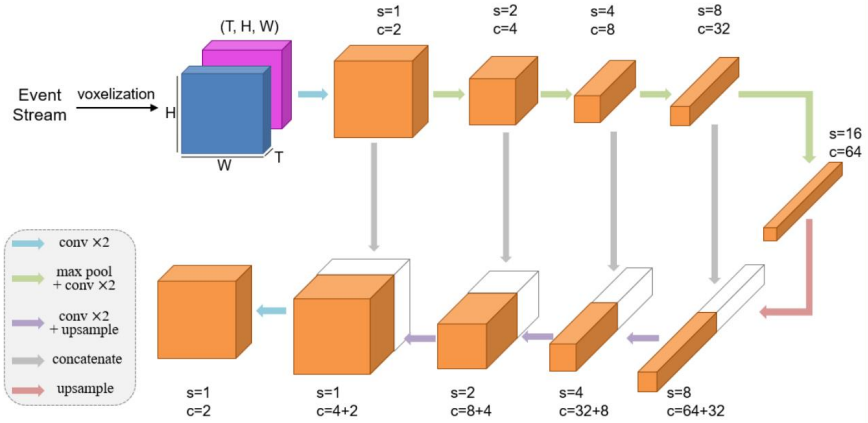

Figure 2: Our event denoising pipeline based on a 3D U-Net backbone.

Experimental Results

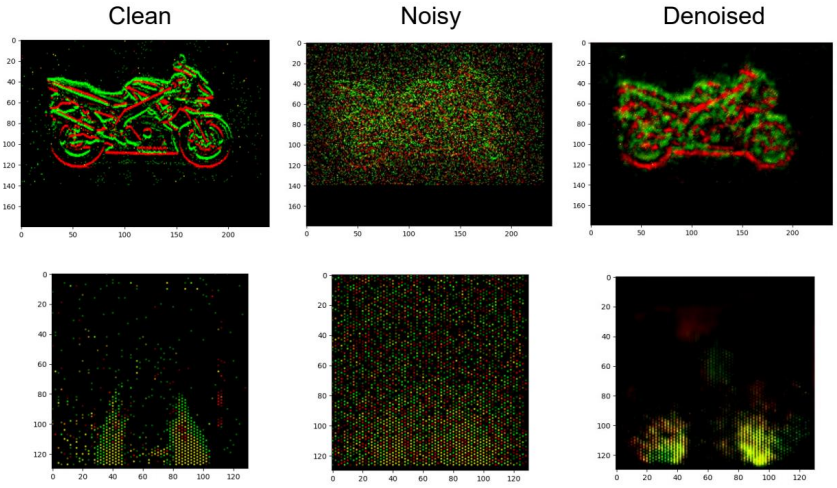

Figure 3: Denoising results shown as a single time channel event image from N-Caltech101 (top) and DVS128 Gesture (bottom). Red pixels represent positive events, green pixels represent negative events, and yellow pixels indicate that events of both polarities are present.

Classification accuracy (%) on noisy test sets from both datasets

	Method	DVS128 Gesture[3]	N-Caltech 101[4]
(1)*	ResNet-34 based classifier trained using clean data	17.32	5.89
(2)	Ours + (1) without additional training	**87.24**	**66.89**
(3)	ResNet-34 based classifier trained using noisy data	89.45	71.63
(4)	Ours + ResNet-34 based classifier trained using noisy data	**91.02**	**77.24**

*(1) Classification accuracies on clean test sets from DVS128 Gesture and N-Caltech 101 are 93.36% and 85.01%, respectively.