Context-aware Sparse Spatiotemporal Learning for Event-based Vision

Shenqi Wang[(1)] and Guangzhi Tang[(2)]

[(1)] TU Delft, [(2)] Maastricht University

Event cameras offer high temporal resolution, high dynamic range, and motion-blur robustness, yet most deep learning methods process event streams with dense activations, forfeiting the sparsity that makes event data attractive for resource-constrained robotic platforms. Neuromorphic processors can exploit activation sparsity for energy-efficient inference, but achieving high sparsity in practice typically requires careful manual tuning of sparsity-inducing loss terms.

We introduce Context-aware Sparse Spatiotemporal Learning (CSSL), a framework that replaces fixed activation thresholds with learned, input-dependent thresholds derived from the local feature distribution. A lightweight convolution branch produces a per-pixel sigmoid threshold that gates the main convolution output through a Heaviside step function, retaining only informative activations. This context-aware mechanism generalizes to standard convolution layers, residual blocks, and convolutional recurrent units (MGU, GRU, MinimalRNN), enabling spatiotemporal sparsity control across diverse architectures without explicit sparsity regularization.

Evaluated on event-based object detection (1 Mpx and Gen1 datasets) and optical flow estimation (MVSEC dataset), CSSL achieves strong performance with high efficiency. On 1 Mpx object detection, CSSL-SEED attains 46.4 mAP with only 2.8 GSOp, roughly 32% of the operations required by RVT-S and 7.4% of leading SNN-based methods. On optical flow, CSSL-EV-FlowNet reaches 2.38 average endpoint error at 10.6% neuron activation density, outperforming prior sparse recurrent baselines in both accuracy and sparsity. Per-layer analysis confirms that context-aware thresholding reduces activation density throughout convolutional layers more effectively than post-hoc L1 sparsity losses, while maintaining or improving task accuracy in a single training stage.

These results position CSSL as a practical route toward deploying event-based vision on neuromorphic hardware. Because high sparsity emerges naturally from context-aware gating, networks trained with CSSL are directly amenable to sparse, event-driven execution without additional sparsity tuning. Future work targets integration with neuromorphic processors and extension to further event-based tasks including motion prediction and autonomous navigation.

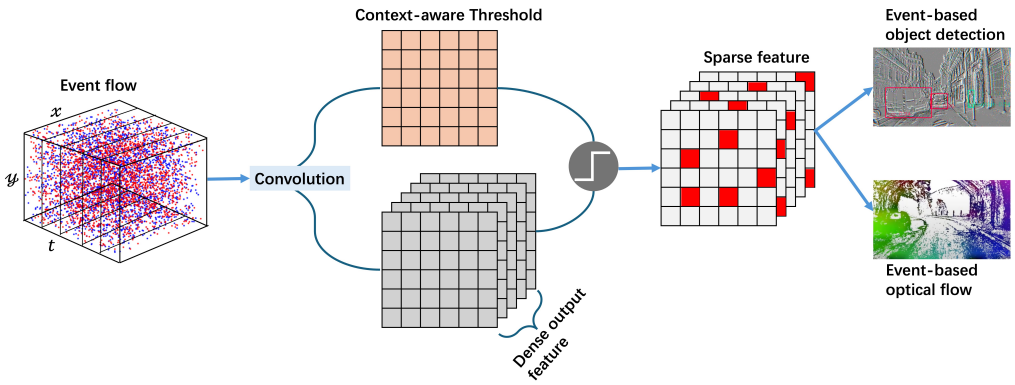

Figure 1: An overview of the proposed Context-aware Sparse Spatiotemporal Learning (CSSL) framework. CSSL introduces context-aware thresholding to dynamically regulate activations in convolutional modules, selectively filtering out redundant activations while preserving essential information. The framework is applied to event-based object detection and optical flow estimation.