SpikeTrack: High-performance and Energy-efficient Event-Based Object Tracking with Spiking Neural Network

Supplementary Material

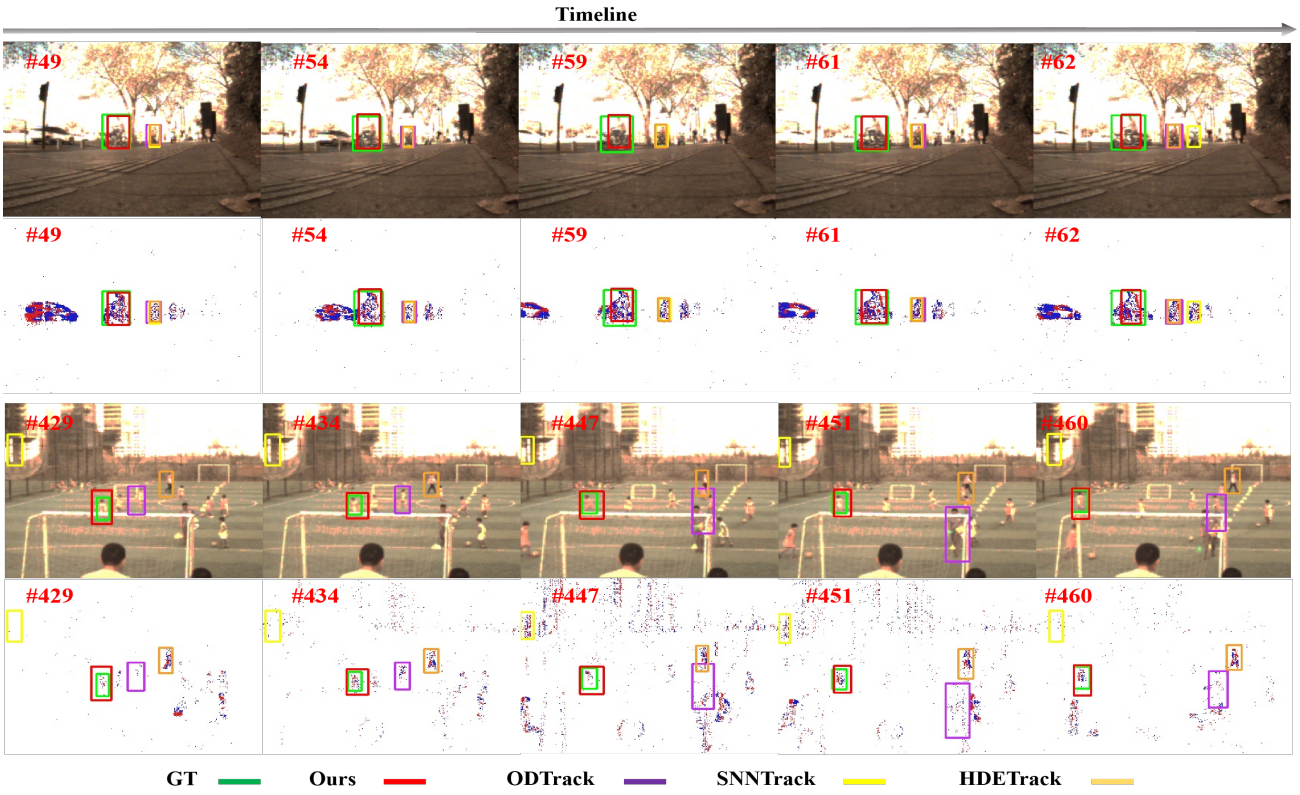

Figure 1. Visual comparison of our SpikeTrack against other SOTA trackers under fast motion condition on VisEvent.

1. Overview

In this supplementary material, we first provide additional qualitative comparisons between our proposed SpikeTrack and several state-of-the-art tracking methods. We then present more detailed experimental settings, followed by further details on the MSST training paradigm and energy estimation. Finally, we discuss the broader societal impacts of SpikeTrack.

2. More Visual Comparisons

Figure 1 presents qualitative comparisons between SpikeTrack and three state-of-the-art trackers—ODTrack [8], SNNTrack [7], and HDETrack [5]—under fast motion scenarios in the VisEvent dataset. Under such conditions, SpikeTrack consistently maintains accurate bounding boxes on target objects, even with large inter-frame displacements. In contrast, conventional trackers either fail to keep up with the target or experience significant drift. Notably, SpikeTrack's multi-frame search training strategy and DI-LIF neuron design contribute to stable localization throughout the sequence. These results highlight the effectiveness of SpikeTrack in dynamic and complex environments.

3. More Experimental Settings

We implement our SpikeTrack using the PyTorch-based SpikingJelly [1] framework. For training the SNN, we adopt the surrogate gradient method, specifically following the STBP [6] algorithm. The STBP algorithm operates across both the spatial domain layer-by-layer and the temporally correlated time domain, enabling the training of high-performance SNNs without requiring additional complex techniques.

4. More MSST training paradigm details

MSST introduces training overhead by using longer sequences with backpropagation through time (BPTT) for temporal learning, but only supervises the final frame to reduce computational cost. During inference, it runs sequentially online without additional latency. This design aligns with the inherent temporal dynamics of SNNs through persistent membrane potentials across frames. As shown in Table 1, SpikeTrack achieves real-time GPU inference at 35 FPS with significantly lower energy consumption than ANNs, and is expected to gain further efficiency on neuromorphic hardware, making it well-suited for power-constrained event-based tracking.

Method	Type	FPS ↑	Latency ↓
TransT	ANN	45	22.2 ms
SNNTrack	ANN + SNN	38	26.3 ms
SpikeTrack (ours)	SNN	35	28.6 ms

Table 1. Inference speed comparison on GPU.

5. More Energy estimation details

The mean firing rate S_{frs} is calculated as a layer-wise weighted average over the entire test set. Specifically, for each SNN layer, the total spike count across all time steps is recorded. This count is divided by the number of neurons in the layer and the total number of time steps to obtain the average firing rate. The resulting value is then used for energy estimation based on the SynOps model. We provide a comparative analysis of energy estimation in Table 2. Compared to I-LIF, DI-LIF suppresses noise spikes and employs highly sparse spike patterns to encode features efficiently. This sparse coding strategy maintains competitive performance while significantly improving energy efficiency, which is crucial for low-power neuromorphic computing.

Setting	Firing rates (%)	Power (mJ)	SR (%)
w/ I-LIF	12.5	9.17	38.1
w/ DI-LIF	10.8	7.92	38.9

Table 2. Comparison of firing rates, power consumption, and tracking performance.

6. Broader Impacts of SpikeTrack

Event-based cameras [2, 4] and spiking neural networks (SNNs) [3] offer ultra-low latency and high energy efficiency. SpikeTrack exemplifies these advantages by demonstrating that a purely spike-driven tracker can achieve performance comparable to leading frame-based trackers while consuming only a fraction of the energy. We outline three principal societal benefits: (i) SpikeTrack's all-integer training and inference enable deployment on hardware with minimal digital logic, reducing silicon area, memory footprint, and battery consumption. This is particularly suitable for edge devices such as agricultural drones, wearable assistants, and IoT sensors, especially in energy- and bandwidth-constrained environments. (ii) Our proposed Multi-Search-sequence-and-Single-Template (MSST) training paradigm leverages the event stream's fine temporal resolution, offering a blueprint for robust motion understanding in autonomous driving, industrial robotics, and assistive navigation where fast reaction is critical. (iii) By maintaining sparse representations through event streams, SpikeTrack minimizes data transmission, thereby lowering bandwidth usage and long-term privacy risks compared to frame-based approaches.

References

[1] Wei Fang, Yanqi Chen, Jianhao Ding, Zhaofei Yu, Timothée Masquelier, Ding Chen, Liwei Huang, Huihui Zhou, Guoqi Li, and Yonghong Tian. Spikingjelly: An open-source machine learning infrastructure platform for spike-based intelligence. *Science Advances*, 9(40):eadi1480, 2023. 1

[2] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 2

[3] Wulfram Gerstner and Werner M Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002. 2

[4] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128 × 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008. 2

[5] Xiao Wang, Shiao Wang, Chuanming Tang, Lin Zhu, Bo Jiang, Yonghong Tian, and Jin Tang. Event stream-based visual object tracking: A high-resolution benchmark dataset and a novel baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19248–19257, 2024. 1

[6] Yujie Wu, Lei Deng, Guoqi Li, Jun Zhu, and Luping Shi. Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers in neuroscience*, 12:331, 2018. 1

[7] Jiqing Zhang, Malu Zhang, Yuanchen Wang, Qianhui Liu, Baocai Yin, Haizhou Li, and Xin Yang. Spiking neural networks with adaptive membrane time constant for event-based tracking. *IEEE Transactions on Image Processing*, 2025. 1

[8] Yaozong Zheng, Bineng Zhong, Qihua Liang, Zhiyi Mo, Shengping Zhang, and Xianxian Li. Odtrack: Online dense temporal token learning for visual tracking. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7588–7596, 2024. 1