

SPIF: A Spatio-Temporal Polarity Interaction Filter for Reliable Event Selection

Jiaxu He ¹, Zhan Sun ², Juncheng Li ¹, Hao Chen ¹, Junxiang Ma ¹ and Bo Zhou ^{1,*}

¹ School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China; m202473048@hust.edu.cn (J.H.); junchengli@hust.edu.cn (J.L.); m202573049@hust.edu.cn (H.C.); majunx@hust.edu.cn (J.M.)

² School of Integrated Circuits, Huazhong University of Science and Technology, Wuhan 430074, China; d202480649@hust.edu.cn

* Correspondence: zhoubo@hust.edu.cn

Highlights

What are the main findings?

- SPIF scores each event by combining neighborhood support in time, space, and polarity with the recent activity of the same pixel.
- The results show that polarity information helps separate target events from background responses, and the hot-state term improves the score by weakening evidence from repeatedly triggered pixels.

What are the implications of the main findings?

- Event selection can be handled with a single reliability score instead of separate hand-crafted rules for target preservation and noise suppression.
- SPIF provides a compact training-free filter for retaining sparse target events and reducing unreliable events in event streams.

Abstract

Event cameras provide high temporal resolution and sparse asynchronous output for small-target monitoring. However, distant weak targets often generate sparse and fragmented events that are easily obscured by responses from background structures and sensor background activity (BA) noise. This paper proposes a Spatio-Temporal Polarity Interaction Filter (SPIF), which assigns a reliability score to each incoming event. SPIF constructs weighted neighborhood support from temporal proximity, spatial distance, and polarity relationships, and uses the recent firing history of the center pixel to discount unreliable evidence caused by persistent activation. On the complete test split of the event-based unmanned aerial vehicle (EV-UAV) benchmark, SPIF achieves a macro-averaged F1 score (macro-F1) of 0.7732 and a target retention rate of 0.8896, exceeding the corresponding values of 0.7583 and 0.6740 obtained by the supervised EV-SpSegNet. Driving and the ED24 real-world denoising dataset further evaluate separation between valid events and BA noise. SPIF obtains the highest area under the receiver operating characteristic curve (AUC) on Driving at 3–10 Hz/pixel and under all evaluated ED24 settings. The C++ implementation processes 6.72 million events per second on the tested CPU.

Keywords: event camera; reliable event selection; spatio-temporal correlation; polarity-aware filtering; hot-state discount; background activity noise



Academic Editor: Zhenming Peng

Received: 26 June 2026

Revised: 27 July 2026

Accepted: 28 July 2026

Published: 3 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

A dynamic vision sensor (DVS), also known as an event camera, records pixel-level brightness changes as asynchronous event streams [1]. Compared with frame cameras, a DVS provides microsecond-level temporal resolution, high dynamic range, low latency, and sparse output. These properties reduce static redundancy while preserving transient responses caused by fast motion, supporting applications such as high-speed target tracking, motion estimation, and object detection [2–4]. They also make event cameras suitable for observing fast and weak targets, especially when conventional frame cameras are limited by exposure time and motion blur.

Distant small-target monitoring usually involves wide fields of view, weak responses, and complex backgrounds [5,6]. In these scenes, event cameras capture short-term motion responses without accumulating redundant static frames. Recent MVT and EV-UAV studies further show that event streams provide a high-temporal-resolution representation for weak-target monitoring [7,8]. In this paper, small target follows the EV-UAV setting and refers to a distant UAV whose average annotated size is approximately 6.8×5.4 pixels. The challenge is that an event stream is a mixture of target responses, background responses, and sensor-induced events rather than a direct target trajectory.

Two sources of interference are especially relevant here. The first is scene-induced background activity. A distant target may occupy only a few pixels and provide limited texture, shape, and local contrast, whereas building edges, vegetation textures, cloud variations, platform motion, and other background structures can also trigger events through relative motion or illumination changes. EV-UAV explicitly describes event sources as target events, background events, and noise events, supporting a distinction among these three sources in event-based UAV monitoring [8]. The second source is sensor-generated background activity (BA) noise, including photon shot noise, dark-current shot noise, leakage-current noise, and hot-pixel noise [9–11]. Yan et al. further showed that event noise is not always isolated: near-edge and recessive noise can have stronger spatio-temporal correlations with real events, making them harder to suppress [12]. Although scene-induced background events and BA noise differ in origin, both can form locally coherent event support. The practical goal is to keep sparse task-relevant responses while suppressing interfering events that mislead local decisions.

Lightweight event filters commonly base their decisions on local temporal correlation or event density. BAF, STCF, and density-based filters use temporal windows, neighborhood counts, or spatio-temporal density to distinguish valid events from BA noise [10,13–15]. Their low computational cost is attractive for event-by-event processing. However, these filters usually treat local support as a binary or weakly weighted cue. Such support becomes ambiguous when background structures produce genuine events near a target trajectory or when dense BA noise produces accidental neighborhood correlations.

Recent rule-based filters improve local evidence by incorporating polarity, spatial distance, event attributes, or statistical dependencies. PFD distinguishes neighboring events according to their polarity relationships, whereas EBF jointly weights temporal, spatial, and event-attribute information. Attribute-weighted time surfaces and probabilistic graph models estimate reliability from attribute differences or inter-event statistics [16–19]. These approaches replace equal neighborhood counting with differentiated evidence and improve event ranking under BA noise. However, their scores are determined mainly by historical events in the neighborhood. When the center pixel triggers persistently, its events may repeatedly receive neighborhood support. The recent activity of the source pixel is therefore a useful reliability cue, but neighborhood-only scores do not model it explicitly.

Motion-consistency and learning-based methods model the scene more explicitly. Contrast-maximization and motion-compensation methods identify inconsistent events by

estimating optical flow, camera motion, or a dominant motion model [20–22]. EDnCNN, AEDNet, and EDFormer learn complex spatio-temporal dependencies or event probabilities from training data [11,23,24]. Graph-based methods exploit the irregular structure of event streams: graph-neural-network-driven Transformers and DBRGNN classify events using learned graph representations and residual graph networks after density clustering, respectively [25,26]. Diffusion models and pretrained visual models have also been applied to event-based video interpolation and transferable event representations [27,28]. EV-SpSegNet uses sparse convolutions to segment target events from background and noise events on EV-UAV [8]. These methods exploit richer context, but they also require motion estimation, task-specific training, or network inference. This motivates a local, training-free reliability model that distinguishes reliable support from evidence produced by background structures or abnormal pixel activity.

The above discussion suggests that it is not sufficient to ask whether recent neighboring events exist. A filter also needs to assess whether this support is likely to come from a target or coherent structure, or from background structures and abnormal activity at the source pixel. We therefore propose the Spatio-Temporal Polarity Interaction Filter (SPIF), which extends local neighbor counting to a more discriminative continuous score while retaining local state updates and event-by-event processing.

The remainder of this paper is organized as follows. Section 2 presents SPIF, the datasets, and the evaluation protocols. Section 3 reports the quantitative, ablation, runtime, and qualitative results. Section 4 discusses the behavior and limitations of the reliability score, and Section 5 concludes the paper.

2. Materials and Methods

2.1. Problem Definition

Given a temporally ordered event stream $\mathcal{E} = \{e_i\}_{i=1}^N$, each event is represented as $e_i = (x_i, y_i, t_i, p_i)$. The integer coordinates (x_i, y_i) refer to the sensor array, with the origin at the upper-left corner, x_i increasing to the right, and y_i increasing downward. For a sensor of width W and height H , $x_i \in [0, W - 1]$ and $y_i \in [0, H - 1]$. The timestamp is t_i , and $p_i \in \{-1, +1\}$ denotes the polarity of the brightness change. SPIF assigns a continuous score S_i to each incoming event and retains it when $S_i \geq \theta$. The score ranks event reliability rather than representing a class probability. On EV-UAV, it is used to select target events; on Driving and ED24, it ranks valid events above BA noise. Both settings require sparse or fragmented useful responses to be preserved while unreliable local support is suppressed.

SPIF builds S_i from two reliability cues: neighborhood support and the recent firing history of the center pixel. The first indicates whether the current event belongs to a plausible local structure in time, space, and polarity. The second reduces the reliability of that support when the same pixel has fired repeatedly within a short interval. Their product preserves local structure as positive evidence while attenuating support associated with persistent triggering or local pseudo-support.

2.2. SPIF Reliability Score

Figure 1 shows the SPIF workflow. The method maintains three local state tables: the latest-timestamp table $T(x, y)$, the polarity table $P(x, y)$, and the hot-state table $H(x, y)$. For the current event e_i , SPIF first reads the states within the radius- r neighborhood $\mathcal{N}_r(c_i)$ centered at $c_i = (x_i, y_i)$. The neighboring timestamps and polarities are used to estimate local structural support. In parallel, the center timestamp and hot state indicate whether the current pixel has fired persistently over a short interval. The two results are combined into S_i before thresholding and state writeback.

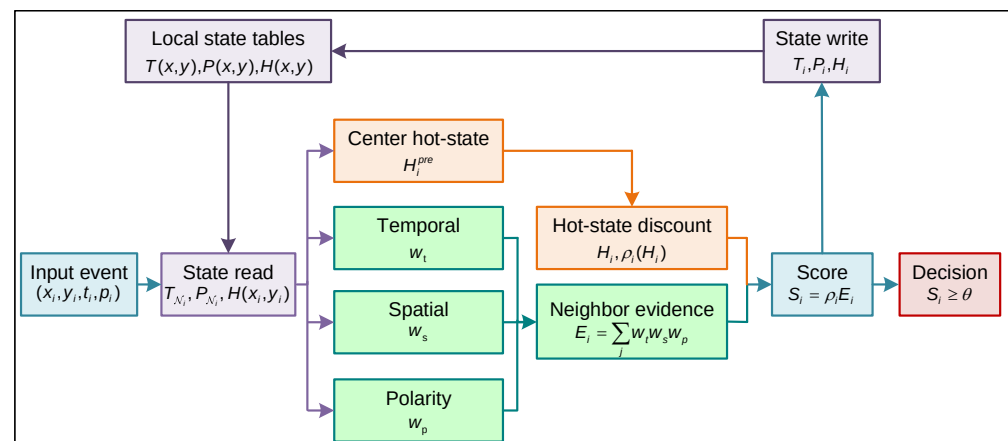


Figure 1. Flowchart of the proposed algorithm.

For a neighboring pixel j , the temporal interval between the current event and the latest historical event at that position is

$$\Delta t_{ij} = t_i - T(x_j, y_j). \quad (1)$$

The temporal kernel uses a squared decay within a finite time window:

$$w_t(i, j) = \begin{cases} \left(1 - \frac{\Delta t_{ij}}{\tau}\right)^2, & 0 < \Delta t_{ij} \leq \tau, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where τ controls the temporal range within which local support can be formed. Historical events with small Δt_{ij} receive high weights. As the interval increases, their contributions decay quadratically and reach zero continuously at $\Delta t_{ij} = \tau$. Events outside the temporal window are ignored. Compared with a linear decay over the same interval, the squared term suppresses older events more rapidly while retaining a continuous boundary at τ . The score is therefore concentrated on short-term structures close to the current event.

Spatial distance decay distinguishes near-neighbor support from distant-neighbor support:

$$w_s(i, j) = \exp\left(-\frac{\|c_i - c_j\|_2^2}{2\sigma^2}\right). \quad (3)$$

The Gaussian kernel provides a non-negative, isotropic decay from the current event. The radius r truncates the support region, whereas σ independently controls how quickly the contribution decreases within that region. This separation preserves a finite neighborhood while avoiding equal weights for all events inside it.

The polarity-aware weight is defined as

$$w_p(i, j) = \begin{cases} 1, & P(x_j, y_j) = p_i, \\ \alpha, & P(x_j, y_j) \neq p_i, \end{cases} \quad (4)$$

where α is the relative weight assigned to opposite-polarity neighborhood evidence. Same-polarity evidence dominates when $\alpha < 1$, whereas opposite-polarity evidence receives greater weight when $\alpha > 1$. Polarity is ignored when $\alpha = 1$. This design does not assume that same- and opposite-polarity neighbors contribute equally in all scenes. Instead, α is selected once for each dataset, allowing polarity to act as a reliability cue rather than a fixed rejection rule.

The weighted neighborhood support is then

$$E_i = \sum_{j \in \mathcal{N}_r(c_i)} w_t(i, j) w_s(i, j) w_p(i, j). \quad (5)$$

A larger E_i indicates stronger local support. This evidence alone is insufficient because the same local accumulation can arise from different sources. A moving target may generate temporally adjacent events along a sparse trajectory, whereas background edges, textured regions, or repeated sensor firings may provide nearby events with compatible timestamps and polarities. The center hot state determines whether recent activity at the current pixel should reduce confidence in the accumulated evidence.

The center hot state describes the repeated triggering strength of the current pixel. Let

$$\Delta t_i^0 = t_i - T(x_i, y_i), \quad (6)$$

and define the hot state before the current triggering as

$$H_i^{\text{pre}} = \max\left(0, H(x_i, y_i) - \frac{2\Delta t_i^0}{\tau}\right). \quad (7)$$

When the current pixel is triggered again, the hot state is updated as

$$H_i = H_i^{\text{pre}} + 1. \quad (8)$$

This update is an event-driven leaky accumulator. The $+1$ term records the current trigger, while the time-normalized leakage removes accumulated activity between successive triggers. More generally, replacing 2 with λ gives $H_n \approx H_{n-1} + 1 - \lambda\delta/\tau$ for an approximately constant inter-event interval δ before clipping. The state therefore accumulates when $\delta < \tau/\lambda$, so λ controls the repeated-firing timescale relative to the neighborhood window. The supplementary experiments on Driving and ED24 identify $\lambda = 2$ as the best common setting when the two mean AUC values are weighted equally; this value is used throughout the experiments.

The reliability discount generated from the center hot state is

$$\rho_i = \frac{H_i + 1}{2H_i + 1}, \quad (9)$$

and the final SPIF score is

$$S_i = \rho_i E_i. \quad (10)$$

When the center pixel triggers repeatedly within a short interval, H_i increases and ρ_i decreases. Strong neighborhood evidence from a persistently active pixel is therefore discounted. The mapping remains positive, monotonic, and bounded, so the firing history of the center pixel attenuates the accumulated evidence without imposing a separate binary hot-pixel decision.

Thus, S_i is a local ranking score rather than a normalized probability. Unlike the fixed neighbor-count decision in STCF, it weights individual neighbors and then adjusts their accumulated support using the firing history of the center pixel.

SPIF retains an event when $S_i \geq \theta$, where θ determines the operating point. After the decision, the current timestamp and polarity are written to $T(x_i, y_i)$ and $P(x_i, y_i)$, respectively. The updated H_i is written to the center hot-state table. The state tables record every arriving event, regardless of whether the event is retained. This keeps state evolution independent of previous threshold decisions and lets subsequent scores use the same

unfiltered event history. Each event is processed in four stages: state reading, reliability scoring, threshold-based decision, and state writeback.

2.3. Complexity Analysis

SPIF uses only local state tables and finite-neighborhood scanning. For each input event, the algorithm reads a neighborhood of radius r and performs one weighted accumulation, giving a per-event complexity of $O(r^2)$. The timestamp, polarity, and hot-state tables are stored per pixel, so the memory cost scales linearly with sensor resolution and does not increase with sequence length. Because the score depends only on the current event and historical local states, SPIF processes the stream in a single pass without buffering the complete sequence or estimating a global motion model.

Unlike methods based on optical-flow estimation, global optimization, or neural-network inference, SPIF requires neither training samples nor learned network parameters. The neighborhood scale determines its main computational cost. A larger r provides broader structural support but increases state accesses and weighted-accumulation operations. The hyperparameter analysis for each dataset selects a finite neighborhood that balances event-selection performance and computational cost.

2.4. Datasets, Baselines, and Evaluation Metrics

The experiments followed the two interference sources introduced above. EV-UAV was used as the main application case to test sparse target retention under background scene responses. Driving and ED24 provided supporting tests on BA noise, evaluating whether the same reliability score ranks valid events above noise under simulated and real low-light conditions, respectively.

EV-UAV is an event-based anti-UAV monitoring dataset acquired with a stationary DAVIS346 camera that observes distant UAV targets under diverse background and illumination conditions [8]. It provides approximate event-level target annotations by extending frame-level two-dimensional target boxes into spatio-temporal boxes. Events within these boxes are treated as target events; the remaining events are treated as background events when evaluating target selection.

Driving is a widely used event-denoising benchmark. It comprises a dynamic driving sequence generated with v2e and mixed with simulated BA noise at 1, 3, 5, 7, and 10 Hz/pixel [10,29]. These rates follow the established public benchmark protocol and provide five progressively denser noise conditions for evaluating how score ranking changes with BA noise intensity.

ED24 combines labeled sequences of valid events with pure BA noise acquired using a DAVIS346 under different illumination conditions. This design preserves the illumination-dependent statistics of real sensor noise [11]. The noise includes photon shot noise, dark-current shot noise, leakage-current noise, and hot-pixel noise. As illumination decreases, mid-frequency dark-current noise increases, and more high-frequency hot pixels become active. We used the Pedestrian06 and Bicycle02 scenes and report results at attenuator voltages of 1.8, 2.5, and 3.3 V. A higher voltage corresponds to lower illumination and stronger BA noise. At these voltages, the signal-to-noise ratios for Pedestrian06 are 5.18:1, 0.40:1, and 0.22:1, respectively. The corresponding ratios for Bicycle02 are 3.51:1, 0.32:1, and 0.16:1.

For EV-UAV, the target retention rate (TRR), non-target suppression rate (NSR), precision, F1 score, and target/background ratio gain (TBR gain) were used to evaluate separation between target events and labeled background events. Here, NSR refers to suppression of labeled non-target events, which may include genuine scene responses as well as sensor noise. Under this annotation protocol, TP, FP, TN, and FN denote retained

target events, retained non-target events, suppressed non-target events, and suppressed target events, respectively. The EV-UAV metrics are defined as follows:

$$\text{TRR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{NSR} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (11)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{F1} = \frac{2\text{Precision} \cdot \text{TRR}}{\text{Precision} + \text{TRR}}. \quad (12)$$

To give each sequence equal weight in the overall EV-UAV evaluation, the F1 score was first computed separately for each sequence and then averaged over the N test sequences:

$$\text{macro-F1} = \frac{1}{N} \sum_{n=1}^N \text{F1}_n. \quad (13)$$

The target/background ratio gain is defined as

$$\text{TBR gain} = \frac{\text{TP}/\text{FP}}{(\text{TP} + \text{FN})/(\text{FP} + \text{TN})}, \quad (14)$$

where TBR gain measures the multiplicative improvement in the ratio of target events to background events after filtering relative to the raw stream. TRR quantifies target retention, whereas NSR quantifies the suppression of labeled non-target events. Precision measures the proportion of target events among retained events, and F1 summarizes the balance between target retention and precision after filtering. Because TBR gain can increase when most events are removed, it was interpreted only together with TRR and macro-F1. For Driving and ED24, AUC was used as the primary metric. AUC measures the ability of filtering scores to rank valid and BA-noise events without relying on a single threshold [10].

For EV-UAV, the parameters and one common decision threshold were selected using all 24 sequences in the official validation split and then frozen for the complete official test split. No parameter or threshold was adjusted for an individual test sequence. Results for each sequence and a sensitivity analysis of sequences with extremely sparse target events are provided in the Supplementary Materials.

The baselines included BAF, STCF, TS, PFD, EBF, EDnCNN, EDFormer, and EV-SpSegNet. Methods with binary outputs generated operating points by scanning temporal windows or neighborhood-count settings, whereas methods that return continuous scores scanned their final decision thresholds. BAF and STCF used $\tau \in [2, 200]$ ms. PFD used $\tau \in [2, 200]$ ms, $[1, 32]$ ms, and $[16, 256]$ ms on EV-UAV, Driving, and ED24, respectively. TS, EBF, and SPIF initially scanned θ over $[0.05, 0.5]$, $[0, 8]$, and $[0, 8]$, respectively. On EV-UAV, if the best validation point lay at the upper boundary and improved macro-F1 over the adjacent candidate by more than 5×10^{-4} , the range was extended until the gain fell below this tolerance. Each training-free method then used its validation macro-F1 to select one parameter setting and one common operating point, which were frozen on the official test split. Driving and ED24 used threshold-independent AUC. EDnCNN and EDFormer used the pretrained models from their original studies. EV-SpSegNet was evaluated only on EV-UAV using the authors' pretrained model and official threshold of 0.9. We retained this published operating point in the main comparison and report a threshold sweep from 0.5 to 0.99 in the Supplementary Material. For Driving, the BAF, TS, EDnCNN, and EDFormer entries were taken from the EDFormer paper [11]; the remaining entries were produced by our implementations.

2.5. Hyperparameter Settings and Sweeps

The main SPIF hyperparameters are the neighborhood radius r , time window τ , distance-decay scale σ , and polarity weight α . EV-UAV used macro-F1 over the complete official validation split to select one structural setting and one decision threshold. Driving used the public 5 Hz/pixel condition for calibration and retained the selected structure at 1, 3, 5, 7, and 10 Hz/pixel. ED24 selected one structure using the mean AUC of Pedestrian06 and Bicycle02 at 2.5 V and retained it for both scenes at 1.8, 2.5, and 3.3 V. Thus, each benchmark used one common setting rather than a separate setting for every condition. During each sweep, one parameter was varied while the other three were fixed. Table 1 lists the final settings.

Table 1. SPIF parameter settings for the three datasets.

Dataset	r (pixel)	τ (ms)	σ (pixel)	α	θ
EV-UAV	5	128	2.0	22	8.0
Driving	2	32	1.75	0.05	—
ED24 Pedestrian06/Bicycle02	5	256	2.75	0.25	—

Figures 2–4 show the hyperparameter analyses for EV-UAV, Driving, and ED24. EV-UAV selected $r = 5$, $\tau = 128$ ms, $\sigma = 2$, $\alpha = 22$, and a common threshold of 8.0. The heatmap shows that preferred values differ among validation sequences; the final setting therefore maximizes macro-F1 over the full validation split rather than any single sequence. The large α and the ablation results indicate that opposite-polarity evidence is particularly informative for separating sparse target traces from background responses. Driving and ED24 selected smaller α values, for which same-polarity continuity provides the principal polarity cue under BA noise.

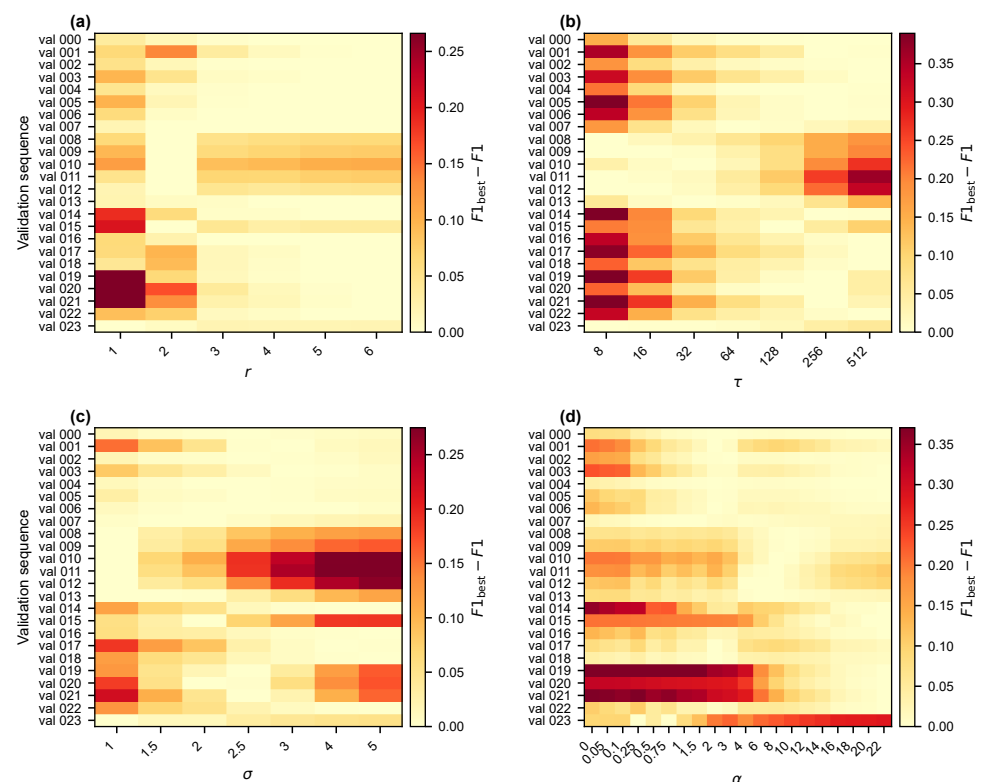


Figure 2. SPIF hyperparameter analysis on the EV-UAV validation split. Each cell shows the F1 loss relative to the best candidate for the corresponding sequence; lighter colors indicate values closer to the best value for that sequence.

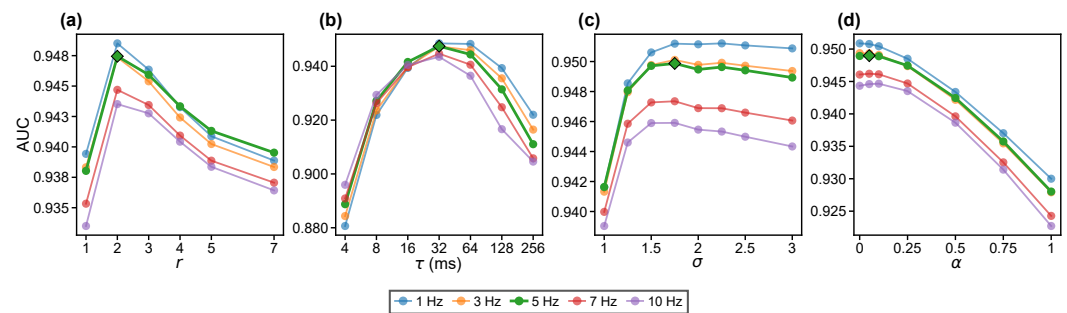


Figure 3. SPIF hyperparameter sweeps on Driving using AUC: (a) neighborhood radius r ; (b) time window τ ; (c) distance-decay scale σ ; and (d) polarity weight α . Green diamonds mark the selected parameter values.

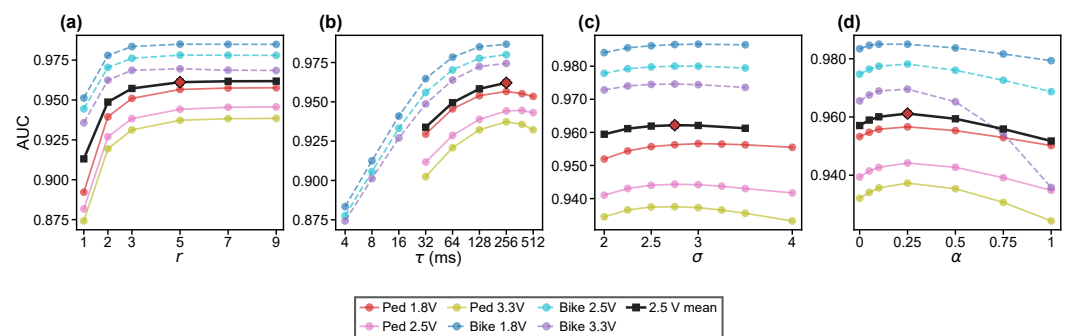


Figure 4. SPIF hyperparameter sweeps on the ED24 Pedestrian06 and Bicycle02 scenes using AUC: (a) neighborhood radius r ; (b) time window τ ; (c) distance-decay scale σ ; and (d) polarity weight α . Red diamonds mark the selected parameter values.

Driving selected $r = 2$ and $\tau = 32$ ms under the 5 Hz/pixel calibration condition. For ED24, increasing r from 5 to 9 raised the mean calibration AUC only from 0.96118 to 0.96183; therefore, $r = 5$ and $\tau = 256$ ms were retained to limit the neighborhood size. The parameter σ controls distance decay within the truncated support region, whereas α controls the relative contributions of the two polarity relationships.

3. Results

3.1. Quantitative Results

This section first reports target selection on EV-UAV, and then uses Driving and ED24 to analyze how SPIF ranks valid events above noise under simulated and real BA noise.

3.1.1. EV-UAV Target Selection

We first examined event selection under scene-induced background responses. The parameters and decision thresholds were frozen after validation, and the results were macro-averaged over the official test sequences. Target events were generally sparse and short-lived, whereas background events were genuine scene responses produced by scene structures, background dynamics, and illumination changes.

As shown in Table 2, SPIF achieved the highest macro-F1 (0.7732) and TRR (0.8896). Its macro-F1 exceeded those of EV-SpSegNet and the next-best training-free method, STCF, by 0.0149 and 0.0988, respectively. At its official threshold of 0.9, EV-SpSegNet achieved the highest NSR, precision, and TBR gain, but its TRR was 0.6740. This operating point therefore applies stronger suppression to improve the proportion of target events after filtering. SPIF retained 88.96% of target events while maintaining an NSR of 0.9268, giving a better overall balance between target retention and non-target suppression on the complete test split.

Table 2. Target selection results on the EV-UAV test split. Macro-F1 and its standard deviation (SD) across sequences are reported separately; the other metrics are macro-averaged. The upward arrow indicates that higher values are better. Best values are boldfaced, and second-best values are underlined. SD is descriptive and is not ranked.

Method	Macro-F1 ↑	SD	TRR ↑	NSR ↑	Precision ↑	TBR Gain ↑
BAF [13]	0.6411	0.3088	<u>0.8650</u>	0.9022	0.5942	27.27
STCF [10]	0.6744	0.3002	0.7639	0.9392	0.7275	239.43
PFD [16]	0.6405	0.2843	0.6921	<u>0.9422</u>	0.7253	231.57
TS [30]	0.6240	0.2621	0.7150	0.9391	0.6144	24.67
EBF [17]	0.6223	0.2857	0.6904	0.9353	0.7278	293.64
EV-SpSegNet [8]	<u>0.7583</u>	0.1936	0.6740	0.9977	0.9041	1414.03
SPIF	0.7732	0.3143	0.8896	0.9268	<u>0.8010</u>	<u>516.75</u>

3.1.2. Driving with Simulated BA Noise

Table 3 presents the AUC comparison for Driving under different simulated BA noise levels, providing a controlled analysis that complements the EV-UAV experiment. SPIF achieved the highest AUC at medium and high noise levels, although its AUC was slightly lower than that of EDFormer at 1 Hz. EV-SpSegNet was not included in the Driving and ED24 AUC tables because it is designed to identify target events in EV-UAV, whereas these two datasets evaluate separation between valid events and BA noise.

Table 3. AUC comparison on Driving under different noise levels. Best values are boldfaced, and second-best values are underlined.

Method	1 Hz	3 Hz	5 Hz	7 Hz	10 Hz
BAF	0.8479	0.8155	0.7930	0.7732	0.7479
STCF	0.9136	0.9075	0.8965	0.8927	0.8830
TS	0.9307	0.9260	0.9270	0.9241	0.9202
EBF	0.9484	0.9444	0.9408	<u>0.9387</u>	<u>0.9346</u>
PFD	0.9135	0.9079	0.8886	0.8849	0.8794
EDnCNN [23]	0.8873	0.8771	0.8748	0.8654	0.8574
EDFormer [11]	0.9541	<u>0.9472</u>	<u>0.9424</u>	0.9343	0.9264
SPIF	<u>0.9512</u>	0.9502	0.9500	0.9475	0.9463

As the simulated BA noise increased from 1 to 10 Hz/pixel, the AUC values of most methods decreased continuously. The AUC of SPIF decreased from 0.9512 to 0.9463, a reduction of 0.0049. In comparison, the AUC of EDFormer decreased from 0.9541 to 0.9264. EDFormer exceeded SPIF by 0.0029 at 1 Hz/pixel, but SPIF achieved the highest AUC at every noise level from 3 to 10 Hz/pixel. At 10 Hz/pixel, SPIF exceeded EBF and EDFormer by 0.0117 and 0.0199, respectively. Within the evaluated range, SPIF was less sensitive to increasing simulated BA noise, suggesting that the weighted reliability score remains discriminative when random background activity becomes denser.

3.1.3. ED24 with Real Low-Light BA Noise

Table 4 presents the AUC comparison for the ED24 Pedestrian06 and Bicycle02 scenes, further analyzing real low-light BA noise. ED24 complements Driving by using BA noise acquired from a real sensor under decreasing illumination. SPIF achieved the highest AUC at all three noise levels in both scenes, indicating that its reliability score remained effective under real low-light BA noise.

Table 4. AUC comparison on ED24 under real low-light BA noise. Best values are boldfaced, and second-best values are underlined.

Method	Pedestrian06			Bicycle02		
	1.8 V	2.5 V	3.3 V	1.8 V	2.5 V	3.3 V
BAF	0.9013	0.8417	0.8165	0.9489	0.9149	0.8960
STCF	0.9013	0.8598	0.8402	0.9501	0.9302	0.9169
TS	0.8619	0.8528	0.8465	0.9102	0.8977	0.9148
EBF	<u>0.9427</u>	<u>0.9187</u>	<u>0.9078</u>	<u>0.9789</u>	<u>0.9679</u>	<u>0.9604</u>
PFD	0.8030	0.7932	0.7856	0.8853	0.8759	0.8682
SPIF	0.9563	0.9443	0.9375	0.9865	0.9800	0.9746

As the attenuator voltage increased and illumination decreased, all methods were affected by stronger real BA noise. Nevertheless, SPIF achieved the highest AUC under all six settings. Relative to EBF, the closest method, the AUC margin of SPIF increased from 0.0136 to 0.0297 on Pedestrian06 and from 0.0077 to 0.0142 on Bicycle02. SPIF therefore remained effective not only under simulated BA noise but also under low-light conditions containing multiple real noise components, including shot, leakage-current, and hot-pixel effects.

3.2. Ablation Study

The ablation study analyzes how each score term contributes under the two interference sources. Four variants were used: No spatial removes spatial distance decay; No polarity assigns equal weights to the two polarity relationships; No hot removes the center hot-state discount; and Time only keeps polarity-blind temporal-kernel accumulation while removing spatial decay, polarity-dependent weighting, and the hot-state discount. For EV-UAV, each variant retained the fixed structural parameters and selected only one common threshold on the validation split before frozen test evaluation. EV-UAV is used for target selection, while Driving and ED24 are used to evaluate separation between valid events and BA noise.

Table 5 shows that the score terms make different contributions to EV-UAV target selection. Removing polarity-aware weighting and retaining only the temporal kernel reduced macro-F1 by 0.0748 and 0.0946, respectively. Polarity relationships and multidimensional neighborhood support are therefore the principal cues in this setting. Removing the hot-state discount reduced macro-F1 by 0.0189. The No spatial variant increased TRR slightly but reduced NSR, precision, and macro-F1. Spatial decay and the hot-state discount therefore improve the balance between target retention and non-target suppression rather than simply maximizing the number of retained target events.

Table 5. Macro-averaged ablation results on the EV-UAV test split. The upward arrow indicates that higher values are better, and the best value in each metric column is boldfaced.

Variant	Macro-F1 ↑	TRR ↑	NSR ↑	Precision ↑	ΔMacro-F1
SPIF	0.7732	0.8896	0.9268	0.8010	0
No spatial	0.7647	0.9063	0.8876	0.7782	−0.0085
No polarity	0.6984	0.8398	0.9255	0.7296	−0.0748
No hot	0.7543	0.8706	0.9313	0.7906	−0.0189
Time only	0.6786	0.8522	0.8799	0.7108	−0.0946

Tables 6 and 7 show that the complete SPIF score achieved the highest AUC under every BA-noise setting. On Driving, the largest reductions occurred for No pol. and Time only. On ED24, every ablated variant lost performance, and Time only degraded most under stronger noise. Spatial decay, polarity weighting, and the center hot state act on the

distance distribution, polarity composition, and repeated center triggering of neighborhood support, respectively, and provide complementary evidence in the complete score.

Table 6. Ablation results for SPIF on Driving (AUC). The best value in each row is boldfaced.

Noise Level	SPIF	No Spat.	No Pol.	No Hot	Time Only
1 Hz	0.9512	0.9497	0.9317	0.9496	0.9249
3 Hz	0.9502	0.9480	0.9300	0.9488	0.9225
5 Hz	0.9500	0.9475	0.9302	0.9489	0.9227
7 Hz	0.9475	0.9444	0.9267	0.9468	0.9192
10 Hz	0.9463	0.9426	0.9254	0.9460	0.9240

Table 7. Ablation results for SPIF on ED24 (AUC). The best value in each row is boldfaced.

Scene	Noise Level	SPIF	No Spat.	No Pol.	No Hot	Time Only
Pedestrian06	1.8 V	0.9563	0.9514	0.9507	0.9522	0.9393
	2.5 V	0.9443	0.9303	0.9357	0.9314	0.8970
	3.3 V	0.9375	0.9283	0.9267	0.9212	0.8958
Bicycle02	1.8 V	0.9865	0.9841	0.9834	0.9841	0.9761
	2.5 V	0.9800	0.9738	0.9747	0.9746	0.9615
	3.3 V	0.9746	0.9694	0.9666	0.9675	0.9515

3.3. Runtime Evaluation

We measured runtime on the Driving 5 Hz stream. The traditional filters and SPIF were implemented in C++17 and compiled with Microsoft Visual C++ 19.44 (Microsoft Corporation, Redmond, WA, USA) using /O2, and evaluated on an Intel Core Ultra 7 265 K CPU (Intel Corporation, Santa Clara, CA, USA). Their timing covered event scoring and state updates. EDFormer used PyTorch 2.2.0 with CUDA 11.8, and EDnCNN used MATLAB R2022b; both released pretrained pipelines were evaluated on an NVIDIA RTX 3090 GPU (NVIDIA Corporation, Santa Clara, CA, USA). Timing included model-specific preprocessing, data transfer, and inference. File I/O and metric computation were excluded, and each method was run ten times. Table 8 summarizes the results.

Table 8. Runtime, storage complexity, and neighborhood work on Driving 5 Hz. Runtime is reported as the mean $\pm$ standard deviation over ten runs.

Method	Platform	Mem	Neighbor Visits/Event	Time/Event (μ s)	Throughput (Meps)
BAF	CPU (C++)	WH	8	0.0188 $\pm$ 0.0009	53.21
STCF	CPU (C++)	WH	8	0.0305 $\pm$ 0.0010	32.78
TS	CPU (C++)	WH	9	0.0500 $\pm$ 0.0018	19.99
EBF	CPU (C++)	WH	24	0.1151 $\pm$ 0.0041	8.69
PFD	CPU (C++)	WH	16	0.1420 $\pm$ 0.0046	7.04
SPIF	CPU (C++)	WH	24	0.1489 $\pm$ 0.0056	6.72
EDFormer	GPU (Python)	$\sim$ 2.3 M	–	20.4534 $\pm$ 3.8153	0.0489
EDnCNN	GPU (MATLAB)	>48 M	–	68.4042 $\pm$ 0.2539	0.0146

Here, Mem follows the storage-complexity notation used by EBF [17]; W and H denote sensor width and height, M denotes 10^6 , and the values for the two neural networks are the published estimates reported in that study. SPIF sustained 6.72 Meps on the tested CPU. It was slower than the simpler traditional filters but required much less measured time per event than the released neural-network pipelines on the tested GPU. CPU and GPU timings are reported separately because the implementations and timing scopes differ.

3.4. Qualitative Results

Figure 5 shows the EV-UAV test_018 sequence. Target events account for 21.4% of the complete sequence, and the selected window contains a continuous target trajectory together with a compact cluster of background events near the image center.

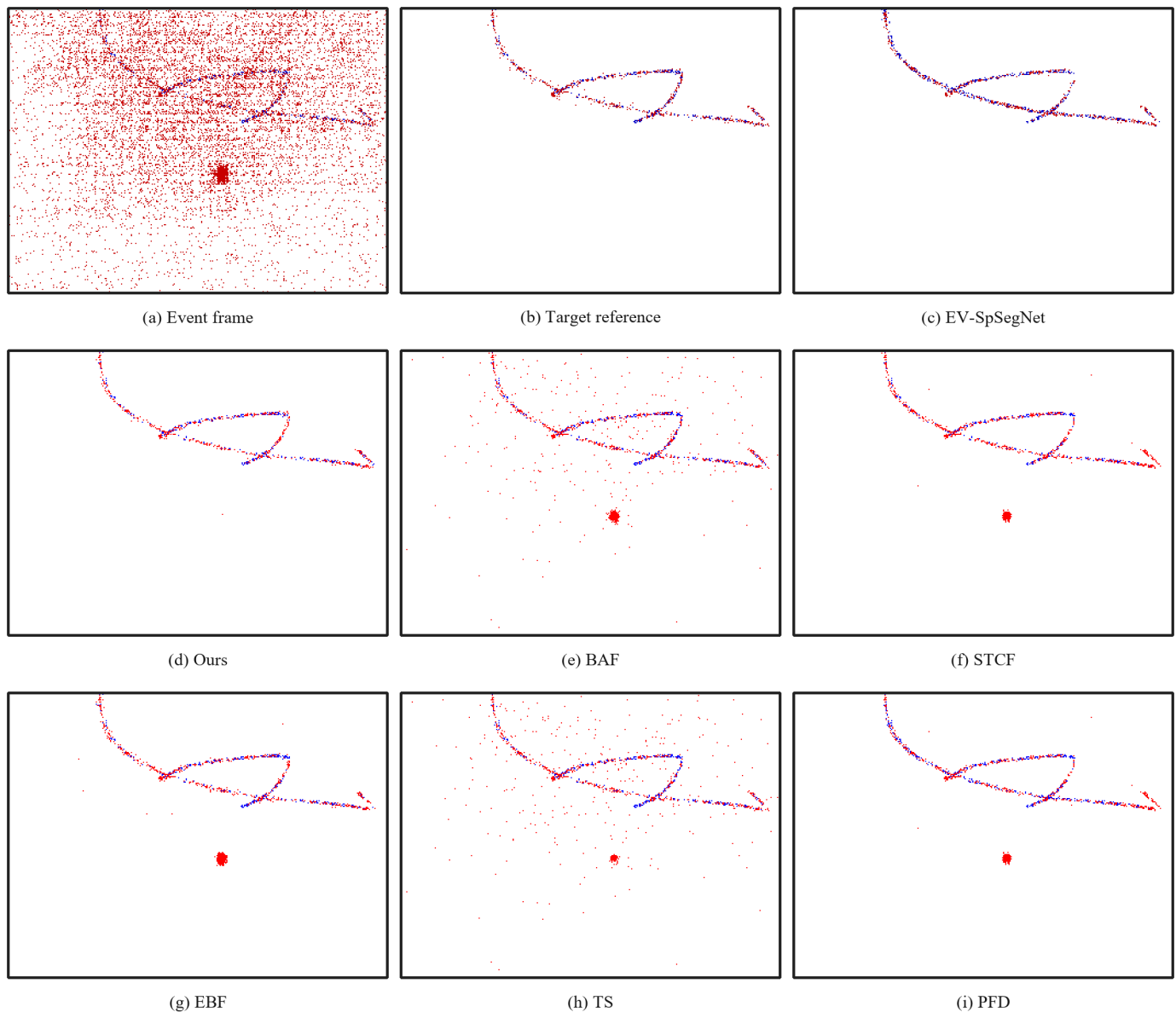


Figure 5. Qualitative comparison on the EV-UAV test_018 sequence. Red and blue denote positive- and negative-polarity events, respectively. EV-SpSegNet uses its official threshold of 0.9, and all results are generated from the same 3-s time window.

The 3-s accumulation is used only to make sparse structures visible. SPIF and EV-SpSegNet both substantially suppress the compact central cluster of background events, whereas the traditional filters retain visible residual responses in this region. SPIF also preserves a more complete target trajectory than EV-SpSegNet, consistent with its higher TRR in this window. Supplementary Figure S3 provides the sparser test_003 case, and Supplementary Figure S4 examines test_023, where extreme target sparsity leads to markedly different retention–suppression trade-offs.

Figure 6 shows the qualitative results for the Driving 5 Hz subset. SPIF suppressed locally dense BA noise while preserving relatively continuous event structures. EBF and

EDFormer also retained comparatively clear structures, whereas EDnCNN over-suppressed valid events in this example.

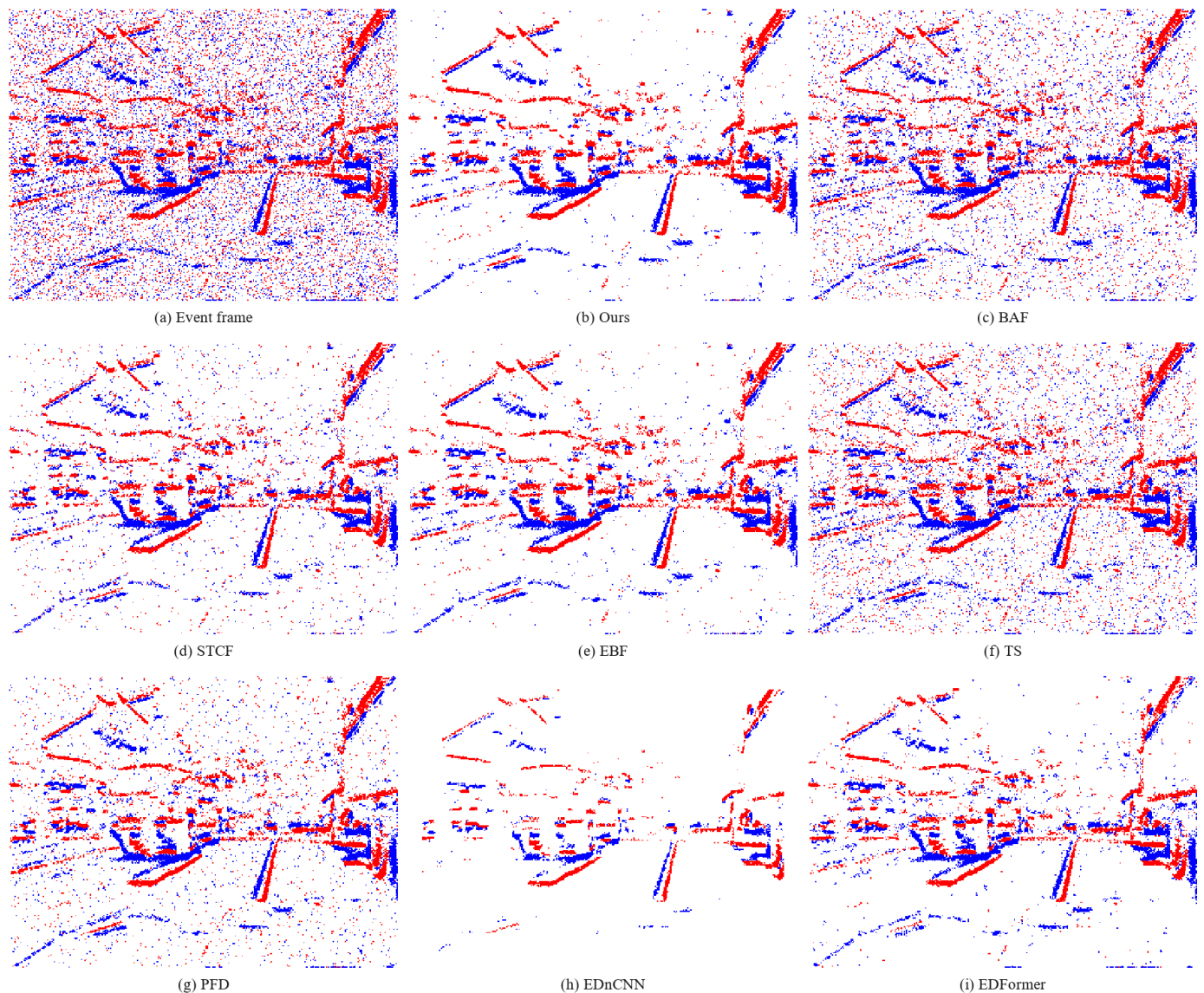


Figure 6. Qualitative comparison of multiple methods on the Driving 5 Hz subset, showing preservation of valid structures and suppression of BA noise. Red and blue denote positive- and negative-polarity events, respectively.

These qualitative results are consistent with the quantitative findings: SPIF reduces dispersed background responses in EV-UAV while preserving the main target trajectory, and it suppresses locally dense BA noise in Driving without removing the dominant event structure.

4. Discussion

Taken together, the three datasets show how SPIF balances event retention and interference suppression under different event-generation mechanisms. On EV-UAV, SPIF achieved the best macro-F1 by retaining more target events than EV-SpSegNet, whereas the supervised network favored stronger background suppression and higher precision. Driving and ED24 further showed that the same reliability score remained discriminative

as simulated or real BA noise increased. Supplementary Figure S1 visualizes the score distributions of individual events.

The ablations explain this behavior. EV-UAV selected a large α , and No polarity produced the second-largest macro-F1 loss after Time only, showing that opposite-polarity evidence is important for distinguishing sparse target traces from background responses. Driving and ED24 selected smaller α values, for which same-polarity continuity was the main polarity cue under BA noise. Removing the hot-state discount reduced performance on all three datasets, so recent activity at the center pixel contributes information that is not captured by neighborhood structure alone. The tests of the leakage coefficient and discount function in the Supplementary Material indicate that gradual accumulation and forgetting are more important than the exact form of the monotonic discount.

The extreme test_023 case in Supplementary Figure S4 reveals a limit of local reliability scoring. SPIF retains most target events but also leaves many background responses, whereas EV-SpSegNet strongly suppresses the background at the cost of removing most target events. Several traditional filters obtain higher F1 scores in this window by taking intermediate operating points. When target events are overwhelmed by background responses, local evidence alone is therefore insufficient to maintain both high retention and strong suppression.

The present experiments do not cover event pile-up during very fast motion, refractory effects, long-term bias drift, or adverse weather such as rain and fog. These factors may alter target density, polarity statistics, and background firing patterns, requiring a new assessment of the structural parameters and operating threshold. Runtime also depends on the compiler, processor, and implementation. Because SPIF was not coupled to a retrained detector or segmenter, the current results support event selection and denoising claims but do not establish an improvement in downstream recognition accuracy.

5. Conclusions

SPIF treats event filtering as a two-part reliability framework rather than a sequence of independent rules. Weighted neighborhood support emphasizes coherent local structures, while the center-pixel state discounts support associated with persistent firing. Both terms use local event history and are updated in event order, without task-specific training, motion estimation, or global optimization.

On EV-UAV, SPIF achieved the highest macro-F1 (0.7732) and TRR (0.8896), while EV-SpSegNet favored stronger suppression at the cost of lower target retention. Driving and ED24 showed that the same score also separates valid events from simulated and real BA noise, with the highest AUC at 3–10 Hz/pixel on Driving and under all six ED24 settings. The ablation results linked these gains to polarity weighting, spatial decay, and the center hot state, and the C++ implementation sustained 6.72 Meps on the tested CPU.

SPIF therefore provides a compact balance between retaining target events in monitoring data and retaining valid events in denoising benchmarks. Its local evidence cannot fully resolve coherent background structures whose support resembles that of a target. Future work will incorporate longer-range motion cues and evaluate the filter jointly with detection or tracking models under adverse weather and changes in sensor bias.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/rs18152551/s1>.

Author Contributions: Conceptualization, J.H. and B.Z.; methodology, J.H.; software, J.H.; validation, J.H.; investigation, J.H. and Z.S.; writing—original draft preparation, J.H.; writing—review and editing, B.Z., J.L., H.C. and J.M.; supervision, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets are publicly available from their original sources. The SPIF implementation is available at https://github.com/3137959601/SPIF_CPP (accessed on 26 June 2026). Section 3 reports the evaluation protocol and parameter settings.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

DVS	Dynamic Vision Sensor
SPIF	Spatio-Temporal Polarity Interaction Filter
BA	Background Activity
TRR	Target Retention Rate
NSR	Non-Target Suppression Rate
TBR	Target-to-Background Ratio
AUC	Area Under the ROC Curve

References

- Gallego, G.; Delbruck, T.; Orchard, G.; Bartolozzi, C.; Taba, B.; Censi, A.; Leutenegger, S.; Davison, A.; Conradt, J.; Daniilidis, K.; et al. Event-Based Vision: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 154–180. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ramesh, B.; Zhang, S.; Yang, H.; Ussa, A.; Ong, M.; Orchard, G.; Xiang, C. E-TLD: Event-Based Framework for Dynamic Object Tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 3996–4006. [\[CrossRef\]](#)
- Gehrig, M.; Scaramuzza, D. Recurrent Vision Transformers for Object Detection with Event Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 13884–13893.
- Gehrig, D.; Scaramuzza, D. Low-Latency Automotive Vision with Event Cameras. *Nature* **2024**, *629*, 1034–1040. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object Detection in Optical Remote Sensing Images: A Survey and a New Benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [\[CrossRef\]](#)
- Du, D.; Qi, Y.; Yu, H.; Yang, Y.; Duan, K.; Li, G.; Zhang, W.; Huang, Q.; Tian, Q. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 370–386.
- Jing, S.; Lv, H.; Zhao, Y.; Liu, H.; Sun, M. MVT: Multi-Vision Transformer for Event-Based Small Target Detection. *Remote Sens.* **2024**, *16*, 1641. [\[CrossRef\]](#)
- Chen, N.; Xiao, C.; Dai, Y.; He, S.; Li, M.; An, W. Event-Based Tiny Object Detection: A Benchmark Dataset and Baseline. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Honolulu, HI, USA, 19–23 October 2025; pp. 7209–7218.
- Berthelon, X.; Chenegros, G.; Finatou, T.; Ieng, S.H.; Benosman, R. Effects of Cooling on the SNR and Contrast Detection of a Low-Light Event-Based Camera. *IEEE Trans. Biomed. Circuits Syst.* **2018**, *12*, 1467–1474. [\[CrossRef\]](#) [\[PubMed\]](#)
- Guo, S.; Delbruck, T. Low Cost and Latency Event Camera Background Activity Denoising. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 785–795. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jiang, B.; Xiong, B.; Qu, B.; Salman Asif, M.; Zhou, Y.; Ma, Z. EDformer: Transformer-Based Event Denoising Across Varied Noise Levels. In *Computer Vision—ECCV 2024*; Springer: Berlin/Heidelberg, Germany, 2025; Volume 15085, pp. 200–216.
- Yan, C.; Wang, X.; Zhang, X.; Li, X. Adaptive Event Address Map Denoising for Event Cameras. *IEEE Sens. J.* **2022**, *22*, 3417–3429. [\[CrossRef\]](#)
- Delbruck, T. Frame-Free Dynamic Digital Vision. In Proceedings of the International Symposium on Secure-Life Electronics, Advanced Electronics for Quality Life and Society, Tokyo, Japan, 6–7 March 2008; pp. 21–26.
- Liu, H.; Moeys, D.P.; Das, G.; Neil, D.; Liu, S.C.; Delbruck, T. Design of a Spatiotemporal Correlation Filter for Event-Based Sensors. In Proceedings of the IEEE International Symposium on Circuits and Systems, Lisbon, Portugal, 24–27 May 2015; pp. 722–725.
- Zhang, P.; Ge, Z.; Song, L.; Lam, E.Y. Neuromorphic Imaging with Density-Based Spatiotemporal Denoising. *IEEE Trans. Comput. Imaging* **2023**, *9*, 530–541. [\[CrossRef\]](#)
- Shi, C.; Wei, B.; Wang, X.; Liu, H.; Zhang, Y.; Li, W.; Song, N.; Jin, J. Polarity-Focused Denoising for Event Cameras. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 4370–4383. [\[CrossRef\]](#)
- Guo, S.; Shi, C.; Wang, L.; Jin, J.; Lu, Y. EBF: An Event-Based Bilateral Filter for Effective Neuromorphic Vision Sensor Denoising. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 9889–9894. [\[CrossRef\]](#)

18. Yin, X.; Zhang, Y.; Chen, G.; Zhang, D.; Yin, J. Attribute-Weighted Time Surface-Based Denoising Method for Space Object Event Streams. *IEEE Sens. J.* **2025**, *25*, 17276–17285. [[CrossRef](#)]
19. Wu, J.; Ma, C.; Li, L.; Dong, W.; Shi, G. Probabilistic Undirected Graph Based Denoising Method for Dynamic Vision Sensor. *IEEE Trans. Multimed.* **2021**, *23*, 1148–1159. [[CrossRef](#)]
20. Gallego, G.; Rebecq, H.; Scaramuzza, D. A Unifying Contrast Maximization Framework for Event Cameras, with Applications to Motion, Depth, and Optical Flow Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3867–3876.
21. Shiba, S.; Klose, Y.; Aoki, Y.; Gallego, G. Secrets of Event-Based Optical Flow, Depth and Ego-Motion Estimation by Contrast Maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 7742–7759. [[CrossRef](#)] [[PubMed](#)]
22. Shiba, S.; Aoki, Y.; Gallego, G. Simultaneous Motion and Noise Estimation with Event Cameras. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Honolulu, HI, USA, 19–23 October 2025; pp. 6959–6969.
23. Baldwin, R.W.; Almatrafi, M.; Asari, V.; Hirakawa, K. Event Probability Mask and Event Denoising Convolutional Neural Network for Neuromorphic Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1701–1710.
24. Fang, H.; Wu, J.; Li, L.; Hou, J.; Dong, W.; Shi, G. AEDNet: Asynchronous Event Denoising with Spatial-Temporal Correlation among Irregular Data. In Proceedings of the ACM International Conference on Multimedia, Lisbon, Portugal, 10–14 October 2022; pp. 1427–1435.
25. Alkendi, Y.; Azzam, R.; Ayyad, A.; Javed, S.; Seneviratne, L.; Zweiri, Y. Neuromorphic Camera Denoising Using Graph Neural Network-Driven Transformers. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 4110–4124. [[CrossRef](#)] [[PubMed](#)]
26. Feng, W.; Wang, X.; Zhan, X.; Huang, H. Event Denoising for Dynamic Vision Sensor Using Residual Graph Neural Network with Density-Based Spatial Clustering. *Neurocomputing* **2025**, *636*, 130026. [[CrossRef](#)]
27. Chen, J.; Feng, B.Y.; Cai, H.; Wang, T.; Burner, L.; Yuan, D.; Fermuller, C.; Metzler, C.A.; Aloimonos, Y. Repurposing Pre-Trained Video Diffusion Models for Event-Based Video Interpolation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 11–15 June 2025; pp. 12456–12466.
28. Chen, Z.; Hou, J.; Zhu, Z.; Wu, J.; Shi, G. Scaling Dense Event-Stream Pretraining from Visual Foundation Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Denver, CO, USA, 3–7 June 2026; pp. 8011–8022.
29. Hu, Y.; Liu, S.C.; Delbruck, T. V2E: From Video Frames to Realistic DVS Events. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 1312–1321.
30. Lagorce, X.; Orchard, G.; Galluppi, F.; Shi, B.E.; Benosman, R.B. HOTS: A Hierarchy of Event-Based Time-Surfaces for Pattern Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1346–1359. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.