Taming Dynamic Clutter: Variance-Driven Adaptive Gain Control for Bio-inspired Small Target Detection

Jiaxiang Li[1], Shaobing Gao[2⋆], Qinbing Fu[4], Meiyi Li[1], Tiansheng Lu[1], and Minjie Tan[3]

[1] School of Mathematics/College of Computer Science, Sichuan University
[2] School of Artificial Intelligence/College of Computer Science, Sichuan University
[3] College of Computer Science and Artificial Intelligence, Southwest Minzu University
[4] School of Mathematics and Information Science, Guangzhou University
jiaxiang589@gmail.com; gaoshaobing@scu.edu.cn; qifu@gzhu.edu.cn

Abstract. Bio-inspired Small Target Motion Detection (STMD) models often suffer "false alarm collapse" due to non-stationary environmental interference, failing to adapt to dynamic background statistics. We propose a bio-inspired architecture integrating macro-motion decoupling with micro-scale Adaptive Gain Control (AGC). Macroscopically, it leverages lobula plate tangential cells (LPTC)-based population coding to eliminate ego-motion-induced baseline drifts in real-time with minimal overhead. Microscopically, we introduce an adaptive shunting inhibition operator inspired by Natural Scene Statistics (NSS). By quantifying local temporal variance, this mechanism maps non-stationary clutter into a gain control factor to suppress heavy-tailed noise from high-frequency flickers. This versatile plug-and-play module integrates seamlessly into existing STMD frameworks. Extensive experiments on synthetic and real-world datasets demonstrate that our model significantly reduces false alarms while maintaining high sensitivity for dim targets, achieving state-of-the-art precision and robustness in complex dynamic environments. Our code is publicly available at AGC.

Keywords: Bio-inspired vision · Small target motion detection · Non-stationary environmental interference · Shunting inhibition · Natural scene statistics

1 Introduction

The real-time and robust detection of extremely small moving targets (typically subtending 1–3 degrees of the visual field [30, 41]) in complex dynamic environments remains a formidable challenge in computer vision [23, 44]. Due to the severe lack of target features and the pervasive sensor noise [41], this task is exceptionally difficult, yet it is crucial for strategic applications such as low-altitude UAV defense, infrared search and track, and autonomous navigation [7, 24, 29].

⋆ Corresponding author.

To tackle this challenge, bio-inspired STMD models [41], derived from insect visual systems, have emerged as a promising paradigm owing to their high efficiency and low-power potential. By simulating neural cascade computations, a series of STMD architectures–ranging from the seminal ESTMD [41] to the recent Feedback STMD [39, 40] have demonstrated remarkable feature extraction capabilities under stationary background assumptions.

Despite their theoretical elegance, the practical deployment of existing STMD models is fundamentally bottlenecked by a two-fold environmental challenge. First, because their core mechanisms rely heavily on basic temporal luminance contrast [42], traditional feed-forward architectures are notoriously fragile against non-stationary local flickers (e.g., rustling leaves or water glints). These localized stochastic oscillations violate the smoothness assumptions of NSS [11, 33], causing models to frequently misclassify them as genuine moving targets. Second, this vulnerability is drastically compounded under camera ego-motion. When global optical flow couples with local dynamic textures, conventional spatial-temporal filters fail to decouple the target's smooth physical translation from chaotic background disturbances. Consequently, even with basic motion compensation, these models lack adaptive regulation and inevitably suffer from a catastrophic surge of background false positives in complex real-world scenarios [26].

To reconcile the discrepancy between biological efficiency and artificial fragility in cluttered environments, this paper proposes a hierarchical spatio-temporal visual enhancement system. Our framework orchestrates a synergistic cascade of macro-scale global kinematic decoupling and micro-scale adaptive neuromodulation. This architecture replicates the dual-stream processing found in insect vision, where the LPTC integrate wide-field motion for ego-motion estimation, while the STMDs maintain high-fidelity selectivity for micro-transients. At the macro-scale, we leverage the population coding principles of the LPTC network [39] to perform feedforward compensation, effectively stabilizing the visual input into a quasi-stationary reference frame. At the micro-scale, we transcend traditional static filters by embedding the canonical mechanisms of sensory adaptation [9, 14, 22, 31] and shunting inhibition [6, 12, 16, 33] into the detection pathway. By mathematically abstracting these processes as steady-state solutions to nonlinear neural dynamics, our system quantifies local spatio-temporal uncertainty to pinpoint high-entropy flicker sources. This triggers a rapid, variance-driven "Gain Collapse" that suppresses heavy-tailed noise while strictly preserving legitimate target signals. Consequently, this design confers a robust, context-aware homeostatic regulation capability on the artificial visual system, as illustrated in the system architecture (Fig. 1). The primary contributions of this paper are summarized as follows:

- **Macro-Statistical Stabilization via LPTC:** We leverage a hierarchical LPTC population coding mechanism [39] to compensate for camera ego-motion, providing a stable foundation for micro-motion extraction.
- **Sealing the Gaps of Local Uncertainty:** We propose an adaptive AGC operator that eliminates non-stationary local disturbances–a critical vulnerability that LPTC-based stabilization [39] alone cannot resolve.

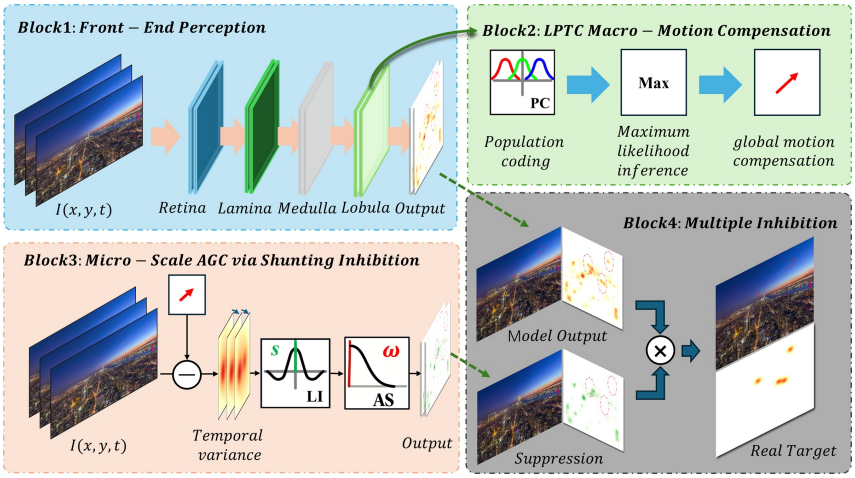


Fig. 1: The system is organized into four synergistic blocks: Block 1: The front-end visual cascade, simulating the insect Retina-Lamina-Medulla-Lobula pathway for spatio-temporal feature extraction. Block 2: Macro-motion compensation, which leverages LPTC-based population coding and Maximum Likelihood Estimation (MLE) to decouple global camera ego-motion. Block 3: Micro-scale AGC module, which quantifies local temporal variance and maps statistical uncertainty into an inhibitory gain factor ω via a Lateral Inhibition (LI) operator and Adaptive Suppression (AS) mapping. Block 4: Multiple Inhibition stage, where the raw model output is modulated by the suppression mask through multiplicative gating, effectively sealing false-alarm vulnerabilities induced by non-stationary flickers while preserving genuine moving targets.

– **Synergistic "Plug-and-Play" Architecture:** Designed as an independent adaptive gate, our module is purely plug-and-play and can be seamlessly integrated into mainstream Bio-inspired STMD architectures.

2 Related Work

2.1 Small Target Detection Paradigms

Two primary computational paradigms exist for extracting extremely small targets. Deep learning (DL) models [7, 29], leveraging feature pyramids [24], attention mechanisms, and context modulation [23, 44], achieve high accuracy on controlled benchmarks. However, their practical edge deployment is hindered by three major bottlenecks: the prohibitive cost of pixel-level annotations, excessive convolutional overhead, and the reliance on computationally heavy optical flow networks [35, 36] for ego-motion compensation. To circumvent these bottlenecks, bio-inspired STMDs [30] offer a lightweight alternative. Foundational architectures–ranging from the seminal ESTMD [41] to variants with directional selectivity [38] and temporal feedback [40] excel under stationary background

assumptions but are highly vulnerable to global optical flow. While the recent STFeedbackSTMD [39] effectively mitigates this by compensating for global camera ego-motion, its inherent reliance on relative temporal contrast leaves a critical structural vulnerability: the inability to adapt to non-stationary local high-frequency clutter. Consequently, existing models still frequently fail when exposed to the complex real-world dual impact of global moving backgrounds and localized dynamic disturbances [3].

2.2 Global Dynamics Compensation for Moving Backgrounds

In dynamic observation platforms, global background displacement induces massive false alarms. Traditional methods for motion compensation generally fall into two categories: classical iterative algorithms, such as Lucas-Kanade [27] or dense feature alignment, which suffer from local optima in low-contrast, textureless scenarios; and modern DL-based optical flow models (e.g., RAFT [36], StreamFlow [35], and DPFlow [28]). While these deep learning approaches offer enhanced precision, they are typically plagued by prohibitive computational overhead and high latency when deployed on resource-constrained UAV platforms [1]. Conversely, insects exhibit remarkable agility by utilizing LPTC to perceive global optical flow efficiently [4, 17, 20]. Recent neuro-computational studies reveal that LPTC arrays can encode complex global motion through low-dimensional population coding rather than pixel-wise optimization [5, 18]. By integrating this bio-inspired MLE architecture [39], our framework bypasses the heavy burden of iterative pixel-level alignment. This enables ultra-fast, feed-forward global motion compensation, establishing a quasi-stationary reference frame at a minimal computational cost–a critical prerequisite for our downstream micro-feature extraction module [3].

2.3 Non-stationary Clutter Suppression and Adaptive Gain Control

Even with ideal background alignment, pervasive micro-scale clutter (e.g., foliage, water glints) induces catastrophic false alarms in traditional STMDs [2]. This stems from their reliance on static normalization, which fails to adapt to non-stationary local statistics. While recent approaches like context-aware fusion [43] and attention-based filtering [23] mitigate such interference, they often impose prohibitive computational costs. Neuro-biological evidence suggests that neurons maintain homeostasis through sensory adaptation [9, 22] and shunting inhibition [6, 33], dynamically suppressing high-variance noise that violates NSS [32]. Drawing on these principles and top-down modulation [15], we introduce a temporal-variance-based AGC mechanism into the STMD domain. By mapping non-stationary local statistics into a gating factor, our plug-and-play module suppresses high-frequency flickers while preserving the lightweight, real-time efficiency of bio-inspired architectures.

3 Method

3.1 Macro-motion Compensation and Residual Generation

When dealing with global optical flow shifts induced by a moving camera, traditional dense optical flow methods incur prohibitive computational costs. To establish a reliable reference frame for subsequent target detection, our first step is to compute a globally aligned difference residual signal, $E(x, y, t)$, which strips away the low-frequency macro-displacement caused by camera translation:

$$E(x, y, t) = I(x, y, t) - I(x - \phi, y - \psi, t - \Delta t) \tag{1}$$

where I denotes the input image sequence, and (ϕ, ψ) represent the cumulative macro-displacements of the background in the 2D physical plane.

To efficiently estimate macro-dynamic parameters (ϕ, ψ) without pixel-level iterative optimization, we adopt the LPTC-based ego-motion compensation framework from [39]. This biological module decodes global kinematic cues–specifically, motion direction $\tilde{\theta}$ and velocity $v(\epsilon)$—via MLE on a population of directionally-tuned neurons. By integrating these parameters over time, we derive the accumulated background displacements (ϕ, ψ) between reference time s and current time t:

$$\phi(t, s) = \int_s^t v(\epsilon) \cos\tilde{\theta} d\epsilon, \quad \psi(t, s) = \int_s^t v(\epsilon) \sin\tilde{\theta} d\epsilon \tag{2}$$

where (ϕ, ψ) map historical background coordinates into the current quasi-stationary reference frame, enabling ultra-fast, feedforward motion compensation.

3.2 Micro-scale AGC via Shunting Inhibition

Ideally, the macro-compensated residual $E(x, y, t)$ (Eq. 1) should contain only low-energy Gaussian noise. However, the LPTC module, acting as a macroscopic integrator, fails to address ubiquitous micro-scale, non-stationary clutter. These high-frequency flickers manifest as intense, heavy-tailed distributions [33], which trigger catastrophic false alarm collapses in existing STMDs [26, 42].

To suppress this residual noise, we draw inspiration from biological *sensory adaptation* [9, 22], which endows neurons with homeostatic regulation to dynamically adjust gain based on local statistics. Specifically, we model the neuron's adaptation state $\sigma^2(t)$ for variance sensing using the *Leaky Integrator* dynamics [8, 9]:

$$\tau \frac{d\sigma^2(t)}{dt} = -\sigma^2(t) + E^2(t) \tag{3}$$

where τ governs the memory decay (forgetting rate), and $E^2(t)$ denotes the instantaneous input energy of the residual.

To implement this continuous neural process in our discrete detection framework, we apply the Euler method to discretize Eq. (3) with a temporal step Δt:

$$\frac{\sigma^2(t) - \sigma^2(t - \Delta t)}{\Delta t} = \frac{1}{\tau}\left(-\sigma^2(t - \Delta t) + E^2(t)\right) \tag{4}$$

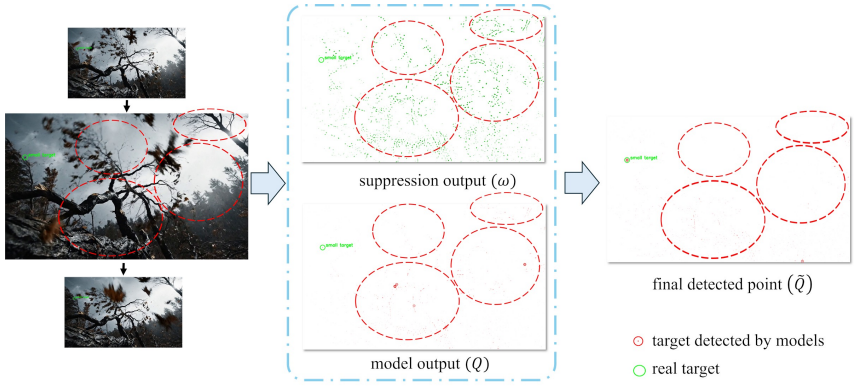


Fig. 2: An input video sequence featuring extreme non-stationary clutter caused by violently swaying branches (highlighted by red dashed circles) alongside a genuine small target (green circle). Middle: The raw baseline output (Q) suffers from massive false alarms in these high-variance regions. Concurrently, the AGC module generates a suppression mask (ω) by precisely capturing the local temporal uncertainty. Right: Through multiplicative gating ($\tilde{Q} = \omega \cdot Q$), the final detection map completely eliminates the motion-induced artifacts within the red boundaries while strictly preserving the signal of the legitimate target. In the detection maps, red solid circles denote the objects detected by the model.

Rearranging the terms yields the discrete-time update rule:

$$\sigma^2(t) = \left(1 - \frac{\Delta t}{\tau}\right) \sigma^2(t - \Delta t) + \frac{\Delta t}{\tau} E^2(t) \tag{5}$$

Starting from $\sigma^2(x, y, 0) = 0$, Eq. (5) is mathematically equivalent to the Exponential Moving Average (EMA) with temporal forgetting characteristics:

$$\sigma^2(x, y, t) = (1 - \gamma)\sigma^2(x, y, t - \Delta t) + \gamma E^2(x, y, t) \tag{6}$$

where the adaptation parameter γ corresponds precisely to the biophysical ratio $\Delta t/\tau$. Subsequently, a spatial receptive field W_s integrates this fluctuation information over the image domain Ω, yielding the clutter metric $C(x, y, t)$:

$$C(x, y, t) = \iint_{\Omega} \sigma^2(u, v, t) \cdot W_s(x - u, y - v)\, du dv \tag{7}$$

The kernel W_s is formulated as a center-surround antagonistic profile to distinguish target-like transients from broad clutter:

$$W_s = A \cdot [G_{s_2} - e \cdot G_{s_3} - \rho]^+ + B \cdot [G_{s_2} - e \cdot G_{s_3} - \rho]^- \tag{8}$$

where G_{s_2} and G_{s_3} are 2D Gaussian kernels. To optimize the response, s_2 is tuned to the expected target scale, while s_3 defines the inhibitory context. Regulatory weights A and B, along with inhibitory strength e and threshold ρ, further refine

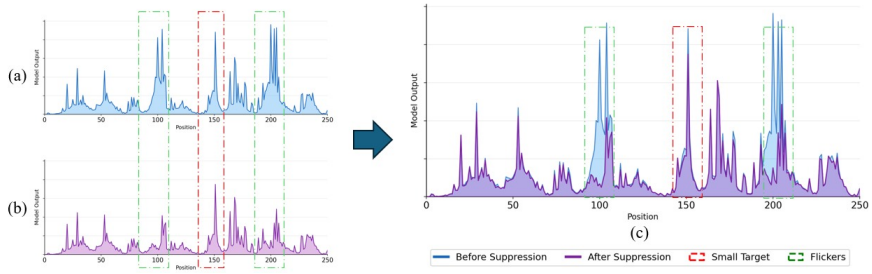


Fig. 3: (a) Raw model output (blue line) where both the genuine small target (red dashed box) and non-stationary flickers (green dash-dotted boxes) exhibit indistinguishable high-amplitude peaks, posing a severe false-alarm risk for traditional STMD operators. (b) Suppressed output (purple line) after adaptive modulation. (c) Comparative overlay highlighting selective suppression. By dynamically perceiving the high temporal variance inherent in flickering disturbances, the system triggers a rapid "Gain Collapse" in these regions. Consequently, non-stationary flickers are significantly dampened, while the detection strength of the legitimate small target is precisely preserved, effectively widening the margin between the true target and background clutters.

the suppression of heavy-tailed noise. $^+$ and $^-$ denote the positive and negative parts, respectively.

Grounded in *Efficient Coding Theory* [13, 22], which suggests neurons maximize information efficiency by suppressing high-variance redundant noise, we utilize C as a metric for local spatio-temporal uncertainty. In neural computational terms, C corresponds to the magnitude of *inhibitory conductance* [6]. Specifically, the membrane potential $\tilde{Q}$ of a neuron driven by an excitatory feedforward input $Q(t)$ and an inhibitory conductance $g_{inh}(t)$ is governed by the shunting inhibition dynamical system [16]:

$$\tau_m \frac{d\tilde{Q}}{dt} = Q(t) - \big(1 + g_{inh}(t)\big)\tilde{Q} \tag{9}$$

Assuming the fast-spiking dynamics reach a rapid steady-state ($\frac{d\tilde{Q}}{dt} = 0$), the response converges to $\tilde{Q} = Q/(1+g_{inh})$, the canonical divisive normalization [6].

Inspired by this mechanism, we design an adaptive gain inhibition factor $\omega(x, y, t)$ as a functional abstraction of the shunting process. By mapping the inhibitory conductance to a nonlinear exponential proxy of the uncertainty metric C, i.e., $g_{inh} = \exp(C^2/\lambda^2) - 1$, the modulation emerges as a dynamic gating factor:

$$\omega(x, y, t) = \frac{1}{1 + g_{inh}} = \exp\left(-\frac{C(x, y, t)^2}{\lambda^2}\right) \tag{10}$$

where the constant λ determines the system's tolerance threshold for heavy-tailed noise. This formulation effectively identifies high-frequency oscillations inexplicable by the LPTC, signaling clutter zones where Eq. (10) triggers a rapid "Gain Collapse" to normalize neural responses while strictly preserving the transients of genuine moving targets.

Acting as a versatile, late-stage plugin, this dynamic gain factor ω is then directly applied to the raw output $Q(x, y, t)$ of any baseline feedforward STMD operator via multiplicative gating, yielding the final modulated output $\tilde{Q}(x, y, t)$:

$$\tilde{Q}(x, y, t) = \omega(x, y, t) \cdot Q(x, y, t) \tag{11}$$

When encountering a genuine small moving target, its displacement is relatively smooth and brief, resulting in an extremely low local uncertainty C. Consequently, the gain $\omega \approx 1$, and the system strictly maintains its utmost detection sensitivity for dim targets. Conversely, once a region exhibits high-frequency flickering, the surge in C triggers a rapid exponential "Gain Collapse" via Eq. (10), driving $\omega \to 0$. As visually demonstrated in Fig. 2, when compared to the raw output of the state-of-the-art STFeedbackSTMD [39], our suppression mechanism empowers the system to correctly identify the genuine small target within a highly cluttered dynamic background, effectively eradicating the vast majority of false alarms.

Beyond these two extremes, the exponential mapping in Eq. (10) is itself Lipschitz-continuous, smoothly bounding the unbounded clutter metric C^2 into a gain within $(0, 1]$. Acting as a continuous soft-threshold rather than a hard cut-off, it avoids the abrupt, discontinuous transitions of binary gating, thereby ensuring a graceful and stable degradation of the response as local uncertainty grows. This property prevents legitimate yet faint targets near the suppression boundary from being erroneously discarded, further enhancing the robustness of the mechanism.

This mechanism effectively simulates the adaptivity of biological visual systems in complex dynamic environments. By selectively dampening non-stationary false alarms without altering the native motion-selective structural properties of the network, it provides a robust filter against severe clutter. From a signal-level perspective, Fig. 3 illustrates the selective suppression capability of our proposed AGC mechanism.

4 Experiments

4.1 Experimental Setup

Details regarding the dataset specifications, the construction of the synthetic flicker scenarios, the definitions of quantitative evaluation metrics, and comprehensive implementation configurations (including hardware deployment details and baseline parameter settings) are provided in the Supplementary Material (SM). To assess the performance of the proposed AGC-augmented STMD framework, we evaluate the system across three domains—Vision Egg [34], RIST [37], and the Infrared Detection Dataset [19]—using AUC, Precision, $F1$-score, Detection Rate, and FA Suppression Ratio (FASR) as primary evaluation criteria.

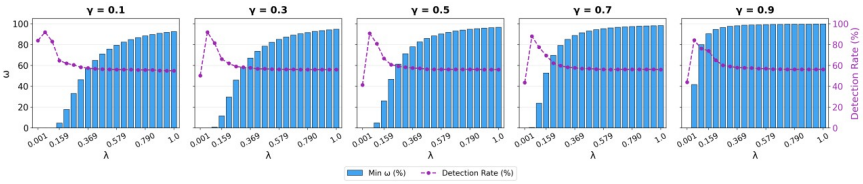


Fig. 4: Parameter sensitivity analysis of the temporal update rate γ and the tolerance threshold λ. Each sub-panel illustrates the variations of the minimum suppression gain ω (blue bars, left y-axis) and the Detection Rate (purple dashed line, right y-axis) as a function of λ. The results demonstrate that the optimal balance between noise suppression and target preservation is achieved at small λ values across various γ configurations.

4.2 Parameter Sensitivity Analysis

Sensitivity analysis via grid search on the temporal update rate γ and threshold λ (Fig. 4) reveals that λ regulates "Gain Collapse" intensity–where excessively small values risk over-suppressing dim targets, while large values fail to reject clutter–and a small γ maintains a stable temporal context for distinguishing smooth target motion from stochastic oscillations. We achieve peak performance at $\lambda = 0.05$ and $\gamma = 0.1$, which provides sufficient suppression depth for robust generalization across diverse dynamic environments without sacrificing sensitivity to ultra-weak targets. While currently fixed as a heuristic, these parameters offer theoretical potential for future kinematic-temporal coupling to enable dynamic velocity-adaptive decay.

4.3 Impact of LPTC Displacement Compensation

Experimental results demonstrate that LPTC-based macro-motion compensation is essential for the stability of our AGC module. Without this alignment, global background ego-motion induces significant ghosting artifacts in the temporal residual (E), which the AGC mechanism erroneously identifies as local statistical uncertainty (C). As visualized in Fig. 5, this misidentification triggers an undesired "Gain Collapse" in texture-rich regions such as leaf edges and sky-background transitions, leading to severe desensitization and the loss of legitimate target signals. Conversely, by incorporating LPTC-based feedforward compensation, we effectively eliminate motion-induced residuals and restore a quasi-stationary background. This ensures that the AGC operator can precisely distinguish between genuine, non-stationary flickers and simple background translations, thereby maintaining high detection sensitivity in complex dynamic environments.

To further probe the influence of LPTC compensation, we conduct a controlled robustness study in which we deliberately inject a bias into the estimated ego-motion, emulating imperfect macro-motion alignment, and examine whether the AGC module remains effective. As shown in Fig. 6(a), detection accuracy degrades gradually and monotonically as the injected bias grows,

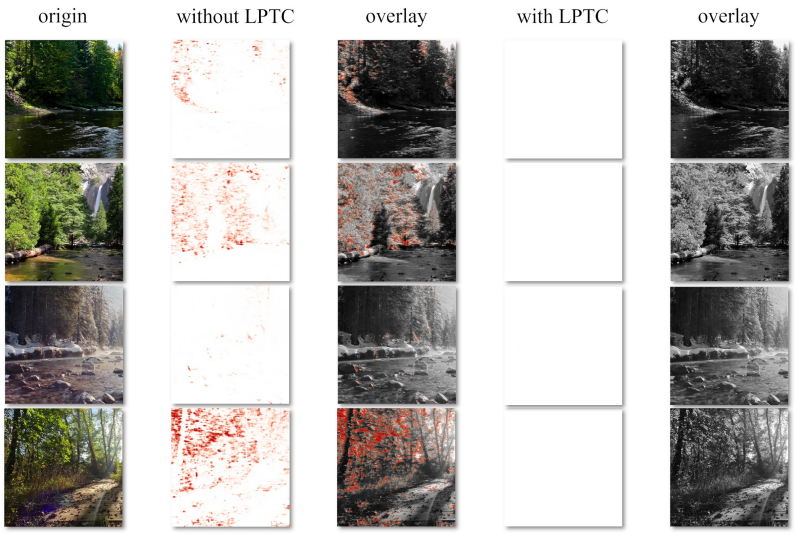
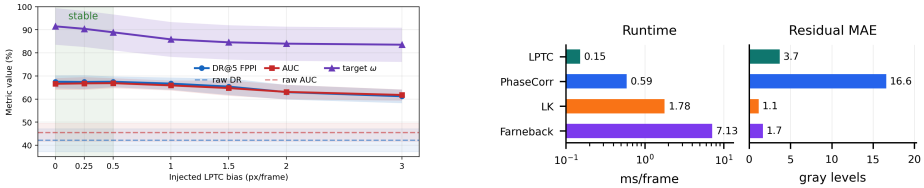


Fig. 5: Visualization of the suppression gain ω with and without LPTC compensation. Red regions in the overlay columns indicate areas where "Gain Collapse" ($\omega \to 0$) is triggered. Without LPTC, background ego-motion induces spurious residuals that are misidentified as clutter. LPTC alignment effectively eliminates these motion-induced artifacts, restoring detection sensitivity across diverse natural scenes.

rather than collapsing abruptly, and consistently surpasses the raw STFeedback-STMD [39] baseline across the tested range. This graceful degradation indicates that the AGC operator does not rely on perfectly accurate ego-motion estimation; instead, it tolerates the bounded residual error inherent in the LPTC population-coding framework, since moderate misalignment perturbs the temporal residual E only mildly and is therefore absorbed by the variance-based gating rather than being misinterpreted as genuine clutter. Beyond this robustness, we benchmark the computational efficiency of LPTC against representative classical motion-compensation methods–Phase Correlation [21], Lucas-Kanade [27], and Farneback optical flow [10]. As reported in Fig. 6(b), LPTC attains the lowest per-frame latency of 0.15 ms–roughly 4× to 47× faster than these classical alternatives–while retaining a highly competitive Residual MAE. Together, these results confirm that LPTC-based compensation supplies the quasi-stationary reference frame required by our AGC operator at a favorable robustness–efficiency trade-off, making the full pipeline well suited to real-time, resource-constrained deployment.

4.4 Comparison on Synthetic and Real-world Datasets

Baseline Performance in Flicker-free Environments To ensure our mechanism does not penalize legitimate targets, we first evaluated it on flicker-free



(a) (b)

Fig. 6: Robustness and efficiency of LPTC-based macro-motion compensation, evaluated on a controlled translation sequence (true motion 0.25 px/frame). (a) Detection performance and suppression gain ω under artificially injected LPTC compensation bias; solid lines denote our full pipeline, dashed lines the raw STFeedbackSTMD [39] baseline without AGC. (b) Module-level benchmark against classical motion-compensation methods (Phase Correlation [21], Lucas-Kanade [27], Farneback [10]), comparing per-frame runtime (left, ms/frame, log scale) and Residual MAE (right, gray levels).

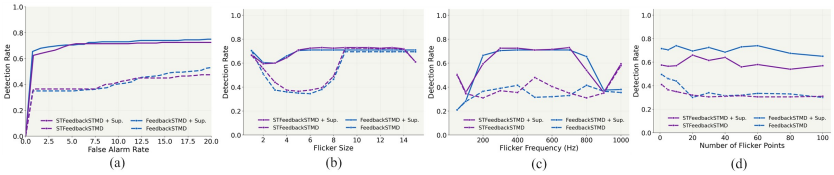


Fig. 7: (a) ROC curves, (b) flicker size, (c) flicker frequencies, (d) flicker quantities.

synthetic sequences (details in SM). In these clean scenarios, our module maintains signal integrity and yields marginal improvements in detection rates at low false alarm levels, demonstrating its compatibility and reliability as a versatile plug-and-play enhancement.

Robustness Under Controlled Flickering Scenarios To further demonstrate the superiority of the proposed adaptive mechanism in highly cluttered environments, we introduced three critical flicker-related parameters: quantity, frequency, and spatial size, using the Vision Egg framework [34]. By default, these parameters were set to 25, 250 Hz, and 5×5 pixels, respectively. To distinguish flickers from legitimate targets, all flicker points were motion-synchronized with the background. We evaluated the impact of each parameter by varying it individually while keeping the others constant.

As shown in Fig. 7, the integration of our adaptive suppression module yields substantial and consistent performance gains over the original STMD baselines. Our approach maintains a stable and significant detection margin across varying flicker frequencies and quantities (Fig. 7a, c, d). However, a performance boundary is observed regarding the spatial scale of disturbances (Fig. 7b): as flicker size exceeds 9×9 pixels, the baselines' inherent size-selectivity begins to recover, narrowing the performance gap. At extreme sizes (near 15×15 pixels), a minor

Table 1: Performance Comparison on Synthetic Data.

Method	Variant	AUC	DR	Prec	F1	FASR	Spd
ESTMD [41]	Orig.	0.172	0.010	0.008	0.009	–	4
	Ours ↓	0.202	0.000	0.000	0.000	27.8	4
DSTMD [38]	Orig.	0.473	0.380	0.085	0.139	–	17
	Ours ↑	0.571	0.390	0.213	0.276	64.9	17
FeedbackSTMD [40]	Orig.	0.606	0.610	0.112	0.190	–	3
	Ours ↑	0.706	0.710	0.807	0.755	96.5	4
STFeedbackSTMD [39]	Orig.	0.858	0.840	0.257	0.393	–	10
	Ours ↑	0.910	0.910	1.000	0.953	100.0	11

crossover occurs, indicating that while our local-variance operator is highly effective for micro-scale clutter, extending its robustness to large-scale non-stationary disturbances remains a promising direction for future optimization.

Table 1 evaluates our AGC plugin across four STMD baselines. We find that the plugin's efficacy correlates with the baseline's ability to generate distinct feature responses. For rudimentary models like ESTMD [41], the mechanism yields marginal results; we attribute this to the inherently low output feature contrast, where our variance-based operator may inadvertently penalize faint targets alongside noise. In contrast, the advantages become prominent as baseline complexity increases. For sophisticated architectures like FeedbackSTMD [40] and STFeedbackSTMD [39], the integration of AGC delivers substantial improvements across all key metrics, achieving near-total suppression of flickering-induced false alarms. These findings highlight the high synergy between our AGC and spatio-temporal feedback loops: by effectively mapping local temporal variance to a gain control factor, our plugin enables precise localization of non-stationary clutter via dynamic "Gain Collapse". With negligible overhead, this versatile module achieves a superior performance-efficiency trade-off, serving as an effective, universal enhancement for bio-inspired vision systems.

Comparison in Controlled Real-World Scenarios Since the original RIST dataset focuses primarily on natural backgrounds, it lacks the high-frequency point flickers–such as neon lights or aviation obstruction lights–commonly encountered in dynamic surveillance. To rigorously verify the robustness of our adaptive suppression operator ω, we introduce a controlled experimental design by injecting simulated flicker points into these real-world video sequences. As evidenced in Table 2, integrating our AGC plugin delivers consistent performance gains across all STMD baselines, particularly in precision and $F1$-score. By quantifying local temporal uncertainty to trigger precise "Gain Collapse" on non-stationary flickers, the AGC operator effectively mitigates the false-alarm vulnerabilities inherent in relative temporal contrast models. These results confirm the plugin's efficacy as an architecture-agnostic enhancement, significantly

Table 2: Performance Comparison on RIST.

Method	Variant	AUC	DR	Prec	F1	FASR	Spd
ESTMD [41]	Orig.	0.566	0.270	0.140	0.184	–	6
	Ours ↑	0.824	0.340	0.352	0.346	62.3	7
DSTMD [38]	Orig.	0.346	0.260	0.045	0.076	–	26
	Ours ↑	0.360	0.205	0.181	0.192	83.3	26
FeedbackSTMD [40]	Orig.	0.408	0.380	0.039	0.070	–	7
	Ours ↑	0.776	0.540	0.543	0.541	95.2	8
STFeedbackSTMD [39]	Orig.	0.209	0.190	0.017	0.032	–	19
	Ours ↑	0.387	0.290	0.228	0.256	90.9	19

Table 3: Performance Comparison on Infrared Data.

Method	Variant	AUC	DR	Prec	F1	FASR	Spd
ESTMD [41]	Orig.	0.391	0.210	0.100	0.136	–	4
	Ours ↑	0.425	0.280	0.145	0.191	12.2	4
DSTMD [38]	Orig.	0.700	0.420	0.139	0.209	–	18
	Ours ↑	0.707	0.420	0.313	0.359	64.6	18
FeedbackSTMD [40]	Orig.	0.734	0.520	0.204	0.293	–	4
	Ours ↑	0.760	0.710	0.464	0.561	59.6	4
STFeedbackSTMD [39]	Orig.	0.724	0.584	0.248	0.348	–	11
	Ours ↑	0.760	0.713	0.735	0.724	85.5	11
DNANet [23]	–	0.086	0.290	0.029	0.053	–	44
SCTransNet [44]	–	0.215	0.265	0.072	0.113	–	34
HCFNet [43]	–	0.394	0.796	0.200	0.320	–	68

stabilizing bio-inspired vision pipelines against deceptive background noise in complex, real-world environments.

Performance on Infrared Datasets Performance on flicker-injected infrared sequences, characterized by low signal-to-noise ratio and complex thermal clutter, further validates our AGC plugin's generalization capabilities. Table 3 shows consistent performance gains across all STMD baselines when integrated with our mechanism. Notably, while deep learning models [23, 43, 44] struggle with these non-stationary disturbances, our approach maintains superior detection precision by selectively suppressing flicker-induced false alarms. This transformative improvement in $F1$-score and precision, coupled with the high FASR values, confirms our AGC module's robust suppression capability. Given its negligible computational overhead, this bio-inspired enhancement provides a hardware-efficient, plug-and-play solution for small target detection in diverse and challenging infrared environments.

Table 4: Quantitative Comparison of Different Models

Model	Precision	Recall	F1	AUC	AP	Speed(ms)
DNANet [23]	0.1320	0.4800	0.2071	0.3958	0.2838	46.7
SCTransNet [44]	0.1572	0.5000	0.2392	0.4117	0.3769	41.6
HCFNet [43]	0.1580	0.2750	0.2007	0.2410	0.1753	126.9
ResUNet_DTUM [25]	0.0191	0.0550	0.0283	0.0553	0.0190	37.1
Ours	0.7908	0.9450	0.8610	0.9518	0.9410	12.4

A Stress Test for Robustness To evaluate generalization against out-of-distribution dynamic clutter, we designed a stress test using synthetic sequences injected with high-frequency flickering. As shown in Table 4, state-of-the-art DL models [23, 25, 43, 44] experience severe performance degradation due to their reliance on static training distributions, which renders them vulnerable to unseen temporal noise. In stark contrast, augmenting the STFeedbackSTMD [39] baseline with our adaptive suppression mechanism achieves superior precision and recall. This advantage stems from our bio-inspired prior: by modeling temporal uncertainty as a deterministic physical constraint, our model effectively suppresses unpredictable flickers to precisely localize genuine small targets (Fig. 8). These results demonstrate that biological priors are indispensable for robust detection against dynamic interference, where data-driven DL models falter. Finally, our approach maintains a significantly lower computational footprint, underscoring its efficiency for real-time, resource-constrained edge deployment.

Complexity Analysis: We analyze the per-frame complexity in terms of the image resolution $H \times W$. The pipeline comprises three sequential feedforward stages: the front-end visual cascade performs spatial filtering at $O(HWK^2)$, where K is the receptive-field radius; the LPTC compensation, using population coding over L directionally-tuned neuron groups, costs $O(HWL)$ with L a small constant; and the AGC module combines element-wise EMA variance sensing, exponential gain mapping, and multiplicative gating ($O(HW)$) with a single fixed-size center-surround convolution ($O(HWk^2)$, with a constant kernel of side k). Since K, L, and k are all constants independent of the resolution, the total per-frame time complexity is $O(HW)$ up to a constant per-pixel factor, with a spatial complexity of $O(HW)$ for buffering intermediate states. Unlike methods relying on dense optical flow or iterative optimization, whose heavy convergence requirements incur prohibitive cost, our feedforward architecture scales linearly with the pixel count, incurring only constant per-pixel overhead. Together with its reliance on a few 2D buffers and basic arithmetic, this makes the pipeline inherently amenable to parallel acceleration on FPGA or neuromorphic hardware, and well suited to real-time, resource-constrained edge deployment.

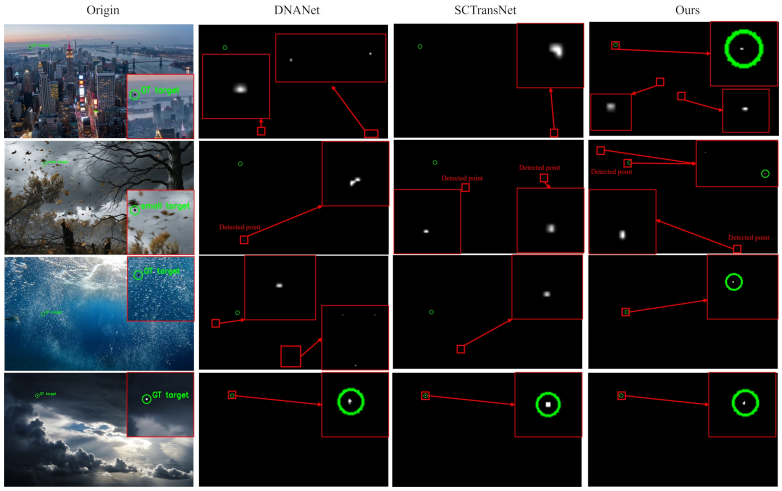


Fig. 8: Qualitative comparison in natural scenes with injected flickers. Green circles mark ground truth. DL baselines are misled by non-stationary clutter, whereas our framework suppresses distractors and preserves the true targets.

5 Conclusion

To bridge theoretical bio-inspired models and complex real-world dynamics, we present a hierarchical spatio-temporal adaptive system. By decoupling macro-scale ego-motion via LPTC population coding and suppressing micro-scale flickers through AGC, our model resolves the long-standing "false alarm collapse" in small target detection. Extensive evaluations show that this architecture-agnostic plugin significantly boosts the precision of SOTA models with negligible added latency, offering a robust, hardware-friendly paradigm for bio-inspired vision. Despite its effectiveness, the framework faces integration challenges with standard DL pipelines: global competition mechanisms such as Softmax or NMS often cause "signal quenching", where strong distractors suppress weak targets. This "winner-take-all" logic is fundamentally incompatible with the localized, parallel autonomy of our biological model. Future work will embed our localized adaptive operators into DL backbones to alleviate these bottlenecks, toward a robust hybrid sensing paradigm, and implement the full processing cascade on neuromorphic hardware for ultra-low-power, real-time performance on resource-constrained UAV platforms.

Acknowledgments. This work was supported in part by Sichuan Science and Technology Program under Grant 2026NSFSC0438, in part by the National Natural Science Foundation of China under Grant 62076170, 62376063, 12571558.

References

1. Alfarano, A., Maiano, L., Papa, L., Amerini, I.: Estimating optical flow: A comprehensive review of the state of the art. Computer Vision and Image Understanding **249**, 104160 (2024). https://doi.org/10.1016/j.cviu.2024.104160
2. Bagheri, Z., Wiederman, S.D., Cazzolato, B., Grainger, S., O'Carroll, D.C.: Robustness and real-time performance of an insect inspired target tracking algorithm under natural conditions. In: 2015 IEEE Symposium Series on Computational Intelligence. pp. 97–102 (2015). https://doi.org/10.1109/SSCI.2015.24
3. Bagheri, Z.M., Wiederman, S.D., Cazzolato, B.S., Grainger, S., O'Carroll, D.C.: Performance of an insect-inspired target tracker in natural conditions. Bioinspiration & Biomimetics **12**(2), 025006 (2017). https://doi.org/10.1088/1748-3190/aa5b48
4. Borst, A., Egelhaaf, M.: Principles of visual motion detection. Trends in Neurosciences **12**(8), 297–306 (1989)
5. Borst, A., Weber, F.: Neural action fields for optic flow based navigation: A simulation study of the fly lobula plate network. PLoS ONE **6**(1), e16303 (2011)
6. Carandini, M., Heeger, D.J.: Normalization as a canonical neural computation. Nature Reviews Neuroscience **13**(1), 51–62 (2012)
7. Cheng, G., Yuan, X., Yao, X., Yan, K., Zeng, Q., Xie, X., Han, J.: Towards large-scale small object detection: Survey and benchmarks. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(11), 13467–13488 (2023). https://doi.org/10.1109/TPAMI.2023.3290594
8. Dayan, P., Abbott, L.F.: Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, Cambridge, MA (2001)
9. Fairhall, A.L., Lewen, G.D., Bialek, W., de Ruyter van Steveninck, R.R.: Efficiency and ambiguity in an adaptive neural code. Nature **412**(6849), 787–792 (2001)
10. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Bigun, J., Gustavsson, T. (eds.) Image Analysis. pp. 363–370. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
11. Field, D.J.: Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A **4**(12), 2379–2394 (Dec 1987). https://doi.org/10.1364/JOSAA.4.002379
12. Fu, Q., Qin, Z.: Emergence of robust looming selectivity via coordinated inhibitory neural computations. Applied Soft Computing **187**, 114337 (2026). https://doi.org/10.1016/j.asoc.2025.114337
13. Gao, S., Li, Y.: Primary visual cortex contributes to color constancy by predicting rather than discounting the illuminant. International Journal of Neural Systems p. 2650034 (2026)
14. Gao, S., Liu, X.: Explaining orientation adaptation in v1 by updating the state of a spatial model. Frontiers in Computational Neuroscience **15**, 759254 (2022)
15. Gilbert, C.D., Li, W.: Top-down influences on visual processing. Nature Reviews Neuroscience **14**(5), 350–363 (2013)
16. Grossberg, S.: Nonlinear neural networks: Principles, mechanisms, and architectures. Neural Networks **1**(1), 17–61 (1988). https://doi.org/10.1016/0893-6080(88)90021-4
17. Hausen, K.: Functional characterization and anatomical identification of motion sensitive neurons in the lobula plate of the blowfly calliphora erythrocephala. Zeitschrift für Naturforschung C **31**(9-10), 629–634 (1976). https://doi.org/10.1515/znc-1976-9-1001

18. Henning, M., Ramos-Traslosheros, G., et al.: Populations of local direction-selective cells encode global motion patterns generated by self-motion. Science Advances **8**(3), eabi7112 (2022)
19. Hui, B., Song, Z., Fan, H., et al.: A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background. Sci. Data Bank **5**, 12 (2020)
20. Joesch, M., Schnell, B., Raghu, S.V., Reiff, D.F., Borst, A.: On and off pathways in drosophila motion vision. Nature **468**(7321), 300–304 (2010)
21. Kuglin, C., Hines, D.: The phase correlation image alignment method. inproceedings of the ieee international conference on cyberneticsand society (1975)
22. Laughlin, S.: A simple coding procedure enhances a neuron's information capacity. Zeitschrift für Naturforschung C **36**(9–10), 910–912 (1981)
23. Li, B., Xiao, C., Wang, L., Wang, Y., Lin, Z., Li, M., An, W., Guo, Y.: Dense nested attention network for infrared small target detection. IEEE Transactions on Image Processing **32**, 1745–1758 (2023). https://doi.org/10.1109/TIP.2022.3199107
24. Li, K., Wan, G., Cheng, G., Meng, L., Han, J.: Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing **159**, 296–307 (2020). https://doi.org/10.1016/j.isprsjprs.2019.11.023
25. Li, R., An, W., Xiao, C., Li, B., Wang, Y., Li, M., Guo, Y.: Direction-coded temporal u-shape module for multiframe infrared small target detection. IEEE Transactions on Neural Networks and Learning Systems **36**(1), 555–568 (2025). https://doi.org/10.1109/TNNLS.2023.3331004
26. Ling, J., Wang, H., Xu, M., Chen, H., Li, H., Peng, J.: Mathematical study of neural feedback roles in small target motion detection. Frontiers in Neurorobotics **16**, 984430 (2022). https://doi.org/10.3389/fnbot.2022.984430
27. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proc. 7th Int. Joint Conf. Artificial Intelligence (IJCAI). pp. 674–679 (1981)
28. Morimitsu, H., Zhu, X., Cesar-Jr., R.M., Ji, X., Yin, X.C.: DPFlow: Adaptive optical flow estimation with a dual-pyramid framework. In: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025)
29. Nikouei, M., Baroutian, B., Nabavi, S., Taraghi, F., Aghaei, A., Sajedi, A., Moghaddam, M.E.: Small object detection: A comprehensive survey on challenges, techniques and real-world applications. Intelligent Systems with Applications **27**, 200561 (2025). https://doi.org/10.1016/j.iswa.2025.200561
30. Nordström, K., Barnett, P.D., O'Carroll, D.C.: Insect detection of small targets moving in visual clutter. PLoS Biology **4**(3), e54 (2006)
31. Qiu, Y., Klindt, D.A., Szatko, K.P., Gonschorek, D., Hoefling, L., Schubert, T., Busse, L., Bethge, M., Euler, T.: Efficient coding of natural scenes improves neural system identification. PLoS Computational Biology **19**(4), e1011037 (2023). https://doi.org/10.1371/journal.pcbi.1011037
32. Ruderman, D.L.: The statistics of natural images. Network: Computation in Neural Systems **5**(4), 517 (nov 1994). https://doi.org/10.1088/0954-898X/5/4/006
33. Schwartz, O., Simoncelli, E.P.: Natural signal statistics and sensory gain control. Nature Neuroscience **4**(8), 819–825 (2001)
34. Straw, A.D.: Vision Egg: An open-source library for realtime visual stimulus generation. Frontiers in Neuroinformatics **2**, 1–10 (2008)
35. Sun, S., Liu, J., Li, H., Liu, G., Li, T.H., Gao, W.: Streamflow: Streamlined multiframe optical flow estimation for video sequences. In: Globerson, A., Mackey, L.,

Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (eds.) Advances in Neural Information Processing Systems. vol. 37, pp. 9205–9228. Curran Associates, Inc. (2024). https://doi.org/10.52202/079017-0292
36. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. pp. 402–419. Springer (2020)
37. Wang, H.: RIST data set. https://sites.google.com/view/hongxinwang-personalsite/download (2020), accessed: Feb. 7, 2026
38. Wang, H., Peng, J., Yue, S.: A directionally selective small target motion detecting visual neural network in cluttered backgrounds. IEEE Transactions on Cybernetics **50**(4), 1541–1555 (2020)
39. Wang, H., Zhong, Z., et al.: Bio-inspired small target motion detection with spatio-temporal feedback in natural scenes. IEEE Transactions on Image Processing **33**, 451–465 (2024)
40. Wang, H., Wang, H., Zhao, J., Hu, C., Peng, J., Yue, S.: A time-delay feedback neural network for discriminating small, fast-moving targets in complex dynamic environments. IEEE Transactions on Neural Networks and Learning Systems **34**(1), 316–330 (2023). https://doi.org/10.1109/TNNLS.2021.3094205
41. Wiederman, S.D., Shoemaker, P.A., O'carroll, D.C.: A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS one **3**(7), e2784 (2008)
42. Xu, M., Wang, H., Chen, H., Li, H., Peng, J.: A fractional-order visual neural model for small target motion detection. Neurocomputing **550**, 126459 (2023). https://doi.org/10.1016/j.neucom.2023.126459
43. Xu, S., Zheng, S., Xu, W., Xu, R., Wang, C., Zhang, J., Teng, X., Li, A., Guo, L.: Hcf-net: Hierarchical context fusion network for infrared small object detection. In: 2024 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6 (2024). https://doi.org/10.1109/ICME57554.2024.10687431
44. Yuan, S., Qin, H., Yan, X., Akhtar, N., Mian, A.: Sctransnet: Spatial-channel cross transformer network for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–15 (2024). https://doi.org/10.1109/TGRS.2024.3383649