Bio-Inspired Robust Spiking Object Detection for Flapping-Wing UAVs

En Lin
School of Software Technology
Zhejiang University
Hangzhou, China
linen@zju.edu.cn

Huaning Li
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
lhnkevin1014@zju.edu.cn

Guanghua Song
School of Aeronautics and Astronautics
Zhejiang University
Hangzhou, China
ghsong@zju.edu.cn

Rui Yan
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
ryan@zjut.edu.cn

Huajin Tang*
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
htang@zju.edu.cn

***Abstract*—Resource-constrained visual perception is crucial for flapping-wing UAVs, where reliable object detection is required for autonomous flight. Spiking Neural Networks (SNNs) provide an attractive solution due to their event-driven computation and sparse activation. However, in practical flapping-wing flight, periodic wingbeat-induced vibration and rapid ego-motion inevitably introduce severe motion blur and vibration-induced jitter, which suppress high-frequency details and significantly degrade the feature representations of existing SNN-based object detectors. To address this challenge, we propose an Eagle-Eye Perception Interaction Module (E-PIM) for fully spiking object detection. E-PIM introduces a dual-branch interaction strategy to enhance blur-sensitive local cues while aggregating wider contextual information, thereby improving the robustness of spiking feature maps under motion degradation. We evaluate the proposed method on VisDrone, a motion-blurred VisDrone benchmark, and a real-world flapping-wing UAV dataset collected during flight with wingbeat-induced vibration and ego-motion blur. Results show consistent improvements under motion degradation while preserving spike-driven inference with minimal additional runtime overhead.**

***Index Terms*—Spiking neural networks, object detection, motion blur, bio-inspired vision, flapping-wing UAVs**

I. Introduction

Bird-inspired flapping-wing UAVs are receiving increasing attention for their high maneuverability and potential in cluttered or confined environments [1]. However, their limited payload and onboard resources make reliable visual perception difficult, and deploying modern object detectors on such platforms remains challenging under strict computation and memory constraints [2], [3]. This challenge is further aggravated in real flight, where periodic wingbeat-induced vibration and rapid ego-motion introduce motion blur and image jitter, degrading visual features that are critical for drone-view detection of small and densely distributed targets.

Spiking Neural Networks (SNNs) have been explored for resource-constrained onboard perception due to their event-driven computation and sparse activations [4]–[6]. Recent studies have extended SNNs from image classification to dense prediction tasks such as object detection [7]–[9]. In particular, a fully spiking detector based on integer-valued training and spike-driven inference has achieved strong detection performance while preserving spiking-domain processing [10]. Building on this baseline, we focus on improving robustness under motion degradation induced by flapping-wing flight.

Although existing SNN detectors have achieved encouraging progress, their robustness under real flapping-wing flight conditions remains underexplored. In practice, periodic wingbeat-induced vibration and rapid ego-motion introduce motion blur and image jitter, which suppress high-frequency details and local contrast that are critical for small-object aerial detection, and may further lead to unstable spiking feature responses. To address this issue, we abstract two complementary mechanisms from biological vision: one branch enhances blur-sensitive local details, while the other aggregates wider contextual cues to stabilize degraded feature representations. Based on this design, we propose the Eagle-Eye Perception Interaction Module (E-PIM) and insert it into the backbone of a fully spiking detector to improve robustness with minimal changes to the original detection pipeline.

The main contributions of this work are summarized as follows:

- We propose E-PIM, a dual-branch interaction module for fully spiking object detection under motion degradation, which enhances local detail cues while aggregating wider contextual information to improve the robustness of spiking feature representations.
- We integrate E-PIM into a fully spiking detector as a lightweight plug-in module, preserving spike-driven inference with little additional runtime overhead.
- We validate the proposed method on both public bench-

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62236007 and 32441113). *Corresponding author.

marks and real-world flapping-wing UAV data, demonstrating consistent detection improvements under motion blur.

II. Related Work

A. SNN-based Object Detection

Spiking Neural Networks (SNNs) have been explored for resource-constrained visual perception due to their event-driven computation and sparse activations. Compared with image classification, extending SNNs to object detection is more challenging because detection requires denser feature representations and more precise localization [4]–[6]. Recent studies have adapted modern detection frameworks to the spiking domain and explored training strategies and architectural designs for object detection [7]–[9], [11], [12]. However, under aerial scenarios with frequent motion degradation and viewpoint changes, SNN detectors may still suffer from unstable feature representations [13], [14]. In this work, we build upon a recent high-performance fully spiking detection baseline [10] and focus on improving robustness under motion blur and jitter.

B. Onboard Perception for Flapping-Wing UAVs

Flapping-wing UAVs introduce additional challenges for onboard perception due to their periodic wingbeat-induced vibration and oscillatory motion patterns, which often cause motion blur and image jitter and reduce the reliability of vision-based detection [15], [16]. Existing efforts mainly address these issues through system-level stabilization or robustness-oriented perception designs [16]–[18]. However, robust object detection under motion degradation in flapping-wing flight remains insufficiently explored [1], [3], [19].

C. Robust Aerial Detection under Motion Blur

Motion blur is a common degradation in aerial imagery caused by rapid ego-motion, platform vibration, and exposure limitations [20]. Such blur suppresses high-frequency details and weakens local contrast, which can be harmful for drone-view object detection where many targets are small and densely distributed [13]. Prior studies investigate robustness evaluation under motion blur and show that detectors trained on clean images may suffer noticeable performance drops when tested on blurred inputs [21]. In this work, we focus on improving robustness and generalization under motion blur without relying on blur-specific training data.

III. Methods

A. Architecture Overview

Our method is built upon the fully spiking one-stage detector SpikeYOLO [10]. SpikeYOLO follows the Backbone–Neck–Head design of YOLOv8 [22] with spike-compatible building blocks. We use three model scales (n/s/m) with depth-width settings $(0.33, 0.375)$, $(0.33, 0.5)$, and $(0.5, 0.625)$, where the number of NSM blocks and channel width are scaled by d and w, respectively.

On top of this baseline, we introduce the Eagle-Eye Perception Interaction Module (E-PIM) to enhance spiking feature representations under motion degradation. As shown in Fig. 1, we keep the overall SpikeYOLO pipeline unchanged and only insert E-PIM into backbone stages for robust feature enhancement. All remaining components (e.g., SNNConv and SNNBlock) follow the baseline design for fair comparison.

B. Spiking Temporal Modeling

We follow the integer-valued training and spike-driven inference scheme of SpikeYOLO [10]. Specifically, the I-LIF neuron emits integer activations during training to reduce the representation loss caused by binary quantization, while these integer activations are decomposed into binary spike sequences during inference to preserve spike-driven computation. The integer activation is given by

$$S[t] = \mathrm{Clip}\big(\mathrm{round}(U[t]),\, 0,\, D\big), \tag{1}$$

where D is the maximum integer output of I-LIF. For frame-based aerial detection, the same image is fed at each timestep, and the feature at a given stage is denoted by $\mathbf{U} \in \mathbb{R}^{T\times C\times H\times W}$. Based on this spatiotemporal representation, E-PIM enhances degraded features while preserving the original temporal dynamics.

C. E-PIM: Eagle-Eye Perception Interaction Module

Motivation and bio-inspired abstraction. Motion blur and jitter suppress high-frequency details and local contrast, leading to unstable spiking feature responses in aerial small-object detection. Inspired by biological vision, we design a dual-pathway interaction module in which one branch enhances local details and the other aggregates wider contextual cues. Given an input feature tensor $\mathbf{X} \in \mathbb{R}^{T\times C\times H\times W}$, the module outputs an enhanced tensor $\mathbf{Y}$ of the same shape.

(1) Expansion with spiking activation. We first apply a 1×1 projection followed by normalization and spiking activation:

$$\mathbf{S} = \mathrm{SN}\Big(\mathrm{BN}\big(\mathrm{Conv}_{1\times1}(\mathbf{X})\big)\Big), \tag{2}$$

where $\mathrm{SN}(\cdot)$ denotes the spiking neuron layer (I-LIF in Sec. 3.2). This step increases feature capacity before interaction and keeps subsequent operations spike-driven, maintaining compatibility with fully spiking inference.

(2) Dual-pathway perception interaction. To capture complementary cues under motion degradation, we employ two depthwise pathways with different receptive fields: a 3×3 pathway emphasizing local detail cues that are most vulnerable to blur, and a 7×7 pathway aggregating wider contextual information to stabilize features under jitter and contrast degradation. The two pathways are fused by additive interaction:

$$\mathbf{F} = \mathrm{BN}\big(\mathrm{DWConv}_{3\times3}(\mathbf{S})\big) \oplus \mathrm{BN}\big(\mathrm{DWConv}_{7\times7}(\mathbf{S})\big). \tag{3}$$

We adopt depthwise convolutions to keep the interaction lightweight while preserving channel-wise spiking patterns.

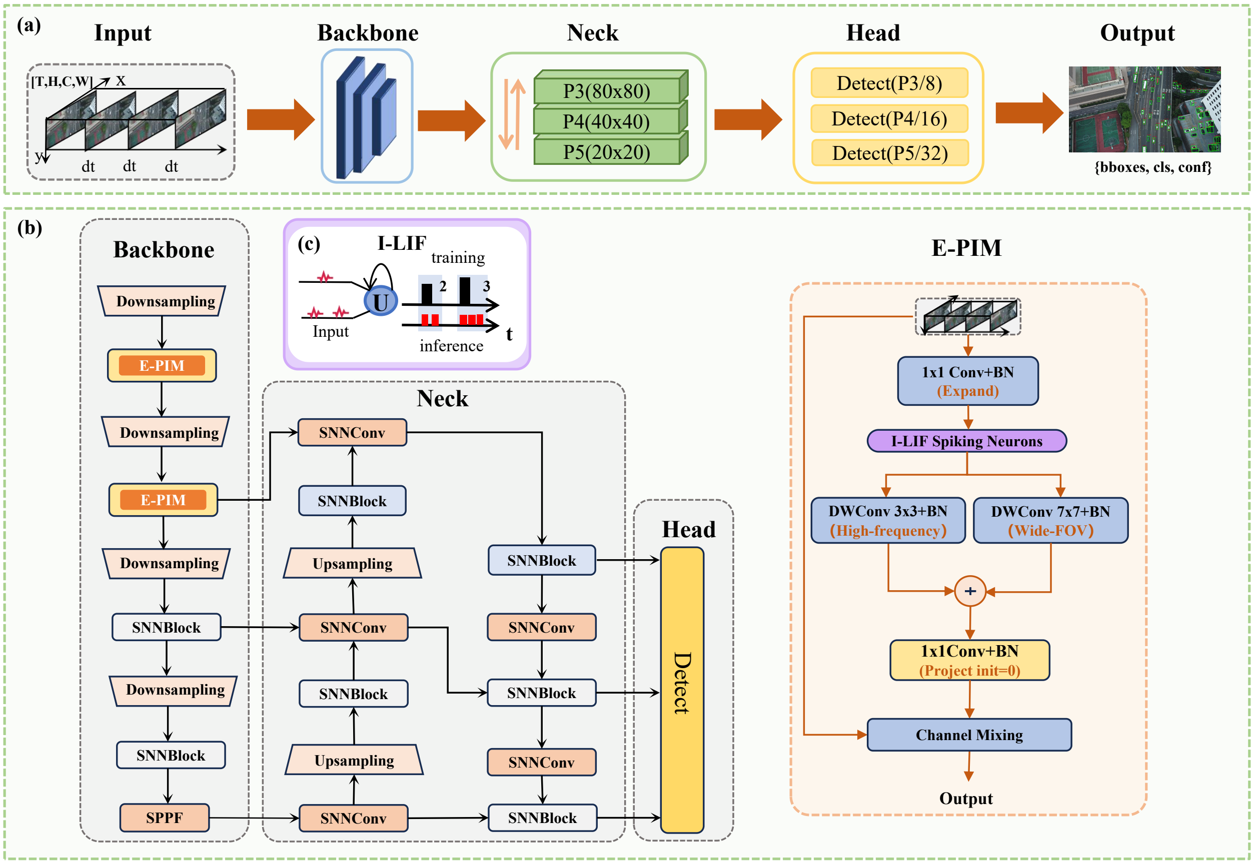

Fig. 1. Overview of the proposed fully spiking detector. (a) Overall detection pipeline. (b) Structure of E-PIM and its insertion into the backbone. (c) Conversion from integer activations during training to binary spike sequences during inference.

(3) Residual projection with zero initialization. After interaction, we project features back and inject them through a residual pathway:

$$\mathbf{Y} = \mathrm{CM}\Big(\mathbf{X} \oplus \mathrm{Proj}(\mathbf{F})\Big), \tag{4}$$

where $\mathrm{Proj}(\cdot)$ is implemented as a 1×1 projection. We initialize $\mathrm{Proj}(\cdot)$ with zeros ("project init=0"), so that the module starts from an approximate identity mapping. This stabilizes plug-in optimization in deep spiking backbones and avoids disrupting spike propagation at the early training stage. $\mathrm{CM}(\cdot)$ denotes a spiking-compatible channel mixing operator (following the baseline implementation) to enhance inter-channel interactions.

IV. Experiments

A. Experimental Settings

We evaluate the proposed E-PIM on aerial object detection benchmarks and real-world flapping-wing UAV data, first reporting detection performance on VisDrone under the standard evaluation protocol. To further assess robustness to motion degradation, we construct *VisDrone-Blur*, a motion-blurred benchmark from the VisDrone validation set. Unless otherwise stated, all models are trained on the original VisDrone training set and directly evaluated on VisDrone-Blur without additional fine-tuning. Moreover, we conduct real-world experiments on a self-collected flapping-wing UAV dataset to validate practical effectiveness under wingbeat-induced vibration and ego-motion blur.

1) Datasets: **VisDrone.** We evaluate our method on the VisDrone aerial object detection benchmark [13]. VisDrone contains diverse drone-view scenes with small and densely distributed targets, complex backgrounds, and large viewpoint variations. All models are trained and evaluated following the official protocol for fair comparison.

VisDrone-Blur. To evaluate robustness under motion blur, we construct VisDrone-Blur by applying synthetic motion blur to the VisDrone validation images. The blur kernels are generated with random directions and magnitudes to simulate motion-induced degradations caused by rapid ego-motion and wingbeat vibrations in flapping-wing UAV flight. VisDrone-Blur is used for testing only, while the training set remains unchanged. Note that VisDrone-Blur focuses on motion blur only, while the real-world flapping-wing UAV dataset contains both blur and vibration-induced jitter.

Real-world Flapping-wing UAV Dataset. To further validate real-world performance, we collect an onboard RGB aerial dataset in a playground scene. The dataset consists of two non-overlapping flight videos and contains 165 annotated

TABLE I
COMPARISON RESULTS ON VISDRONE AND VISDRONE-BLUR UNDER DIFFERENT MODEL SCALES.

0026598*/ Arch.	Model	$T\times D$	Params (M)	VisDrone (Clean)		VisDrone-Blur	
				AP_{50}	AP	AP_{50}	AP
ANN	YOLOv8-N [22]	1	3.1	31.8	18.3	29.5	16.1
	YOLOv8-S [22]	1	11.1	38.1	22.5	34.5	19.3
	YOLOv8-M [22]	1	25.8	41.6	25.1	37.2	21.1
SNN	EMS-YOLO [9]	4	17.3	15.1	6.35	14.0	5.65
	SpikeYOLO-N [10]	1×4	13.2	34.8	19.9	32.2	17.6
	SpikeYOLO-N [10]	4×1	13.2	30.5	17.1	28.3	15.0
	SpikeYOLO-S [10]	1×4	23.1	37.1	21.5	34.3	18.9
	SpikeYOLO-M [10]	1×4	45.7	39.9	23.3	37.0	20.4
	Ours (E-PIM)-N	1×4	13.2	36.1	20.8	33.2	18.2
	Ours (E-PIM)-N	4×1	13.2	29.3	16.5	27.1	14.6
	Ours (E-PIM)-S	1×4	23.1	38.4	22.4	35.1	19.4
	Ours (E-PIM)-M	1×4	47.8	41.6	24.6	38.0	21.1

images with a single category (*pedestrian*). The captured data exhibits realistic degradations such as vibration-induced jitter and motion blur. We apply a hybrid data cleaning strategy that combines automatic filtering and manual inspection to remove low-quality frames. This dataset is used *only for evaluation* and is not involved in training or fine-tuning. As shown in Fig. 2, the *input* denotes the original captured image.

2) Compared Methods: **ANN reference models.** For reference, we include YOLOv8 [22] as a representative ANN-based one-stage detector and report results under three model scales (N/S/M) on both VisDrone and VisDrone-Blur.

SNN baselines. For spiking object detection, we include EMS-YOLO [9] as an existing SNN detector. We further adopt SpikeYOLO [10] as the main baseline fully spiking one-stage detector. SpikeYOLO is designed for aerial detection with spike-compatible architectural components and training strategies, providing a strong baseline under the fully spiking inference setting. To examine scalability across model capacities, we evaluate three scales following the baseline configurations: SpikeYOLO-N, SpikeYOLO-S, and SpikeYOLO-M.

Proposed method. We integrate the proposed E-PIM into the SpikeYOLO backbone and evaluate the resulting models under the same three scales. Unless otherwise specified, all components other than E-PIM (including spiking blocks, training strategies, and detection head) are kept identical to the baseline to ensure fair comparison and isolate the contribution of E-PIM.

Implementation details. All models are implemented in PyTorch and trained/tested on 6 NVIDIA RTX 3090 GPUs. For fair comparison, we follow the training and inference settings of SpikeYOLO [10]. We only insert E-PIM into the baseline backbone, with all other components unchanged.

Spiking settings. The spiking network runs for T timesteps. For I-LIF neurons, the maximum integer activation value is set to D, resulting in an effective temporal budget of $T\times D$. We use 1×4 as the main experiment default and additionally report 4×1 results to analyze different T/D allocations. Other

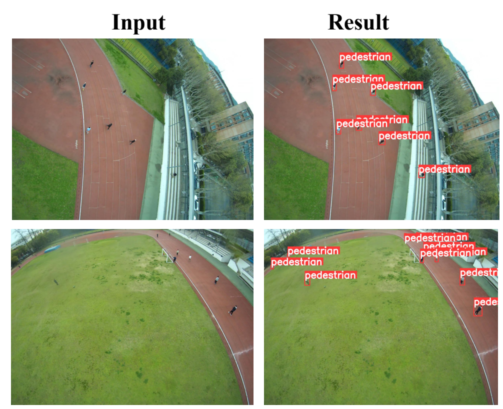

Fig. 2. Qualitative results on the real-world flapping-wing UAV dataset. Left: input images. Right: detection results.

hyperparameters follow the baseline configuration.

3) Evaluation Metrics: **Detection metrics.** We evaluate detection performance using standard object detection metrics, including AP_{50} (mAP at $IoU = 0.5$) and AP (averaged over IoU thresholds from 0.5 to 0.95 with a step of 0.05). Unless otherwise stated, all results are computed under the same evaluation protocol for fair comparison.

B. Main Results on VisDrone

On VisDrone (Clean), we compare ANN detectors (YOLOv8), an existing SNN detector (EMS-YOLO), the SpikeYOLO baseline, and our E-PIM enhanced models, as summarized in Table I. Overall, E-PIM brings consistent improvements over SpikeYOLO across different model scales under the same $T\times D$ setting. Specifically, AP_{50} increases from 34.8/37.1/39.9 to 36.1/38.4/41.6 for N/S/M, corresponding to gains of +1.3/+1.3/+1.7. Similarly, AP improves from 19.9/21.5/23.3 to 20.8/22.4/24.6, yielding gains of +0.9/+0.9/+1.3. These results indicate that E-PIM enhances spiking feature representations and improves aerial detection performance without modifying the overall detection pipeline.

Notably, these improvements are obtained without changing the overall detector pipeline or introducing blur-specific training data. Meanwhile, E-PIM remains computationally lightweight, with a small parameter overhead mainly introduced by the additional projection layers in E-PIM. These results indicate that E-PIM serves as an effective plug-in module for enhancing spiking feature representations on standard aerial detection benchmarks.

C. Robustness Evaluation on VisDrone-Blur

To evaluate robustness under motion blur, we report results on VisDrone-Blur in Table I. Compared with the SpikeYOLO baseline, E-PIM consistently improves detection performance under motion degradation. Specifically, AP_{50} increases from 32.2/34.3/37.0 to 33.2/35.1/38.0 for N/S/M, achieving gains of +1.0/+0.8/+1.0. AP also improves from 17.6/18.9/20.4 to 18.2/19.4/21.1, corresponding to gains of +0.6/+0.5/+0.7. These results suggest that E-PIM enhances the stability of spiking feature representations when motion blur weakens high-frequency details and local contrast, leading to more reliable detection in degraded aerial imagery. Notably, the improvements are consistent across model scales, indicating that the proposed module generalizes well under different capacity settings.

To provide qualitative evidence of the robustness improvement, we visualize the response heatmaps and detection results of the baseline and our method on both VisDrone (clean) and VisDrone-Blur (blur), as shown in Fig. 3. Compared with the baseline, E-PIM produces more concentrated and target-related responses under motion blur, reducing background interference and improving detection stability. These observations are consistent with the quantitative results in Table I.

D. Real-world flapping-wing UAV Experiments

To evaluate practical deployment performance, we conduct experiments on a self-collected real-world flapping-wing UAV dataset. We compare our method with representative detectors, including YOLOv8-N, EMS-YOLO-N, and SpikeYOLO-N, where all models adopt the N scale. The quantitative results are summarized in Table II. Our method achieves the best AP_{50} and comparable AP, indicating consistent performance gains on real-world data. We note that the overall AP ($AP_{50:95}$) is relatively low for all methods, which is mainly attributed to annotation-related localization noise in bounding boxes, making the metric more sensitive to small misalignments. Fig. 2 further provides qualitative examples, where the *input* denotes the original captured images and the *result* shows the corresponding detection outputs.

E. Efficiency Analysis

Besides detection accuracy, runtime efficiency is also important for onboard perception on resource-constrained flapping-wing UAVs. We therefore further compare inference speed under the same evaluation setting, as reported in Table III. All FPS results are measured on an NVIDIA RTX 3090 with batch size 1. As shown in Table III, YOLOv8-S achieves higher throughput on GPU (47.3 FPS), while fully spiking detectors exhibit lower FPS (e.g., 13.63 FPS for SpikeYOLO-S). Under the same model scale, integrating E-PIM keeps the parameter count unchanged (23.1M) and introduces only negligible runtime overhead (13.69 FPS vs. 13.63 FPS). These results indicate that E-PIM improves robustness while maintaining comparable GPU inference speed. We attribute the lower FPS mainly to the timestep-wise state updates required in fully spiking inference and the limited support of standard GPU implementations for sparse event-driven computation.

TABLE II
RESULTS ON THE REAL-WORLD FLAPPING-WING UAV DATASET.

Method	AP_{50} (%)	AP (%)
YOLOv8 [22]	47.3	14.7
EMS-YOLO [9]	39.8	9.76
SpikeYOLO [10]	43.0	13.0
Ours (E-PIM)	49.1	14.7

TABLE III
EFFICIENCY COMPARISON IN TERMS OF RUNTIME SPEED (FPS).

Method	Arch.	Params (M)	FPS
YOLOv8-S [22]	ANN	11.1	47.3
EMS-YOLO [9]	SNN	17.3	16.8
SpikeYOLO-S [10]	SNN	23.1	13.63
Ours (E-PIM)-S	SNN	23.1	13.69

F. Ablation Study

We conduct ablation studies on **SpikeYOLO-N (N-scale)** and report results on VisDrone (Clean), as shown in Table IV. Compared with the SpikeYOLO-N baseline, integrating the full E-PIM improves AP_{50} from 34.8 to 36.1 and AP from 19.9 to 20.8, demonstrating the effectiveness of the proposed module. When keeping only the high-frequency branch, the performance still increases to 35.2 AP_{50} and 20.3 AP, indicating that enhancing local detail cues benefits aerial detection. In contrast, using only the wide-FOV branch leads to a significant drop (27.1 AP_{50} and 14.6 AP), suggesting that relying solely on large-context aggregation may weaken discriminative local features, which is harmful for small and dense targets in aerial imagery. However, when fused with the high-frequency branch, the wide-FOV pathway provides complementary contextual stabilization, resulting in the best overall performance. Overall, the dual-branch interaction in E-PIM provides a better balance between detail preservation and contextual stabilization, resulting in consistent performance gains.

V. CONCLUSION

In this paper, we propose E-PIM, an Eagle-Eye Perception Interaction Module for robust fully spiking object detection

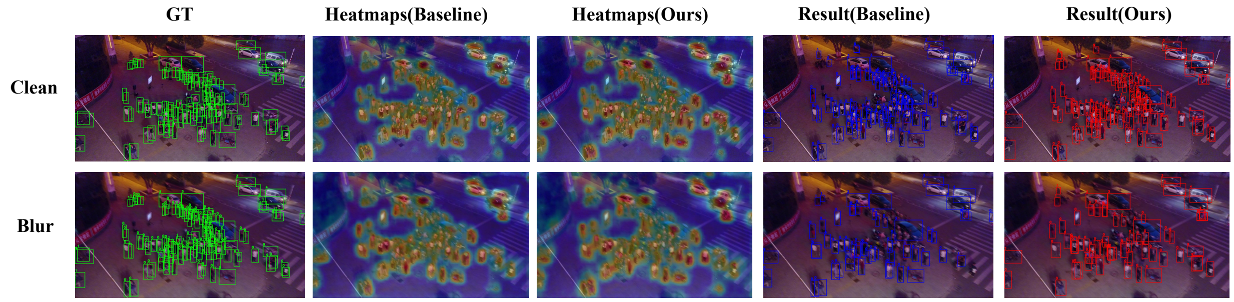

Fig. 3. Qualitative comparison of baseline and E-PIM on VisDrone (clean) and VisDrone-Blur (blur), showing enhanced blur robustness.

TABLE IV
ABLATION STUDY OF E-PIM ON VISDRONE (CLEAN).

Variant	AP_{50} (%)	AP (%)
Baseline (SpikeYOLO)	34.8	19.9
+ E-PIM (full)	36.1	20.8
E-PIM (high-frequency only)	35.2	20.3
E-PIM (wide-FOV only)	27.1	14.6

in flapping-wing UAV scenarios under motion degradation. By integrating a dual-branch perception interaction mechanism into a SpikeYOLO-style fully spiking detector, E-PIM enhances discriminative feature representations under motion blur and vibration-induced jitter. Extensive experiments on VisDrone, our constructed VisDrone-Blur benchmark, and a real-world flapping-wing UAV dataset demonstrate consistent improvements in detection performance and robustness. Moreover, our method improves robustness while introducing only limited additional runtime overhead, indicating its potential for resource-constrained onboard deployment. Future work will extend the evaluation to more challenging real-world degradations and explore practical onboard deployment.

REFERENCES

[1] J. W. Gerdes, S. K. Gupta, and S. A. Wilkerson, “A review of bird-inspired flapping wing miniature air vehicle designs,” in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 44106, 2010, pp. 57–67.

[2] G. de Croon, “Flapping wing drones show off their skills,” *Science Robotics*, vol. 5, no. 44, p. eabd0233, 2020.

[3] R. Zufferey, J. Tormo-Barbero, M. M. Guzman, F. J. Maldonado, E. Sanchez-Laulhe, P. Grau, M. Pérez, J. Á. Acosta, and A. Ollero, “Design of the high-payload flapping wing robot e-flap,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3097–3104, 2021.

[4] S. Ghosh-Dastidar and H. Adeli, “Spiking neural networks,” *International journal of neural systems*, vol. 19, no. 04, pp. 295–308, 2009.

[5] K. Yamazaki, V.-K. Vo-Ho, D. Bulsara, and N. Le, “Spiking neural networks and their applications: A review,” *Brain sciences*, vol. 12, no. 7, p. 863, 2022.

[6] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, “Deep learning in spiking neural networks,” *Neural networks*, vol. 111, pp. 47–63, 2019.

[7] J. Qu, Z. Gao, T. Zhang, Y. Lu, H. Tang, and H. Qiao, “Spiking neural network for ultralow-latency and high-accurate object detection,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 3, pp. 4934–4946, 2024.

[8] S. Kim, S. Park, B. Na, and S. Yoon, “Spiking-yolo: spiking neural network for energy-efficient object detection,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11 270–11 277.

[9] Q. Su, Y. Chou, Y. Hu, J. Li, S. Mei, Z. Zhang, and G. Li, “Deep directly-trained spiking neural networks for object detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 6555–6565.

[10] X. Luo, M. Yao, Y. Chou, B. Xu, and G. Li, “Integer-valued training and spike-driven inference spiking neural network for high-performance and energy-efficient object detection,” in *European Conference on Computer Vision*. Springer, 2024, pp. 253–272.

[11] Y. Ma, B. Guo, and X. Zhang, “Multi-layer feature cascade fusion spiking neural network for object detection,” *International Journal of Neural Systems*, vol. 35, no. 11, p. 2550063, 2025.

[12] L. Yu, H. Chen, Z. Wang, S. Zhan, J. Shao, Q. Liu, and S. Xu, “Spikingvit: A multiscale spiking vision transformer model for event-based object detection,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 17, no. 1, pp. 130–146, 2024.

[13] D. Du, P. Zhu, L. Wen, X. Bian, H. Lin, Q. Hu, T. Peng, J. Zheng, X. Wang, Y. Zhang *et al.*, “Visdrone-det2019: The vision meets drone object detection in image challenge results,” in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 0–0.

[14] H. Zhang, N. Chen, M. Li, and W. An, “Spiking swin transformer for uav object detection based on event cameras,” in *2024 12th International Conference on Information Systems and Computing Technology (ISCTech)*. IEEE, 2024, pp. 1–6.

[15] D. Xue, B. Song, W. Song, W. Yang, W. Xu, and T. Wu, “Computational simulation and free flight validation of body vibration of flapping-wing mav in forward flight,” *Aerospace Science and Technology*, vol. 95, p. 105491, 2019.

[16] Z. Wang, S. Wang, Y. Hu, Y. Zhou, N. Li, and X. Zhang, “Multimodal fusion image stabilization algorithm for bio-inspired flapping-wing aircraft,” *Biomimetics*, vol. 10, no. 7, p. 448, 2025.

[17] Q. Fu, X. Wang, Y. Zou, and W. He, “A miniature video stabilization system for flapping-wing aerial vehicles,” *Guidance, Navigation and Control*, vol. 2, no. 01, p. 2250001, 2022.

[18] E. Pan, X. Liang, and W. Xu, “Development of vision stabilizing system for a large-scale flapping-wing robotic bird,” *IEEE Sensors Journal*, vol. 20, no. 14, pp. 8017–8028, 2020.

[19] S. Rafee Nekoo, R. Rashad, C. De Wagter, S. B. Fuller, G. d. Croon, S. Stramigioli, and A. Ollero, “A review on flapping-wing robots: Recent progress and challenges,” *The International Journal of Robotics Research*, p. 02783649251343638, 2025.

[20] S. Dai and Y. Wu, “Motion from blur,” in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–8.

[21] L. Lelégard, M. Brédif, B. Vallet, and D. Boldo, “Motion blur detection in aerial images shot with channel-dependent exposure time,” in *Symposium on Photogrammetric Computer Vision and Image Analysis (PCV)*, vol. 38, 2010, pp. 180–185.

[22] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics yolov8,” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics