Advances in Military Technology
Vol. 21, No. 2, 2026, pp. 429-442
ISSN 1802-2308, eISSN 2533-4123
DOI 10.3849/aimt.02103

Degradation-Aware Long-Range UAV Detection for Real-Time Military Surveillance Systems

Nguyen Van Ngon[1], Chu Hai Long[1], Do Thi Nhan[1], Huy Dang Pham[2], and Thanh Dong Pham[3*]

[1] *Institute of Technology, General Department of Defense Industry, Dong Ngac, Hanoi, Vietnam*
[2] *Faculty of Military Technology, University of Defence, Brno, Czech Republic*
[3] *Faculty of Aerospace Engineering, Le Quy Don Technical University, Hanoi, Vietnam*

The manuscript was received on 28 February 2026 and was accepted after revision for publication as an original research paper on 7 September 2026.

Abstract:

Small unmanned aerial vehicles (UAVs) in reconnaissance and asymmetric threat scenarios challenge long-range military surveillance. Reliable detection of low-observable UAVs at extended distances is constrained by atmospheric attenuation, geometric scale reduction, and strict real-time limits on embedded hardware. This paper presents a degradation-aware long-range UAV detection framework for edge-based military platforms. A degradation-dependent operational regime is identified, revealing nonlinear performance collapse of conventional detectors beyond a critical threshold. To mitigate this, a latency-aware adaptive image tiling mechanism dynamically allocates computation to candidate target regions within a bounded latency. Experiments on long-range UAV datasets and an NVIDIA Jetson Orin Nano show 24.9 % mAP improvement under severe degradation, confirming that robust detection requires coordinated sensing-computation co-design rather than model complexity alone.

Keywords:

UAV detection, military surveillance, atmospheric degradation, edge deployment, real-time vision systems

1 Introduction

The proliferation of small unmanned aerial vehicles (UAVs) [1] in reconnaissance and asymmetric threat scenarios has introduced significant challenges for military surveillance systems. In modern operational environments, long-range optical monitoring is often required to provide early warning and situational awareness in the airspace.

* *Corresponding author: Faculty of Aerospace Engineering, Le Quy Don Technical University, 236 Hoang Quoc Viet Street, Nghia Do Ward, Hanoi, Vietnam. Phone: +84 943 835 239, Email: pham.thanh-dong@lqdtu.edu.vn. ORCID 0000-0002-1089-9035.*

However, reliable detection of small UAVs at extended distances is constrained by atmospheric attenuation, geometric scale reduction, and strict latency requirements on embedded processing platforms.

From a military engineering perspective, long-range UAV detection is not solely a computer vision problem but a coupled sensing-computation challenge. As the observation distance increases, the apparent size of UAV targets decreases rapidly, while atmospheric scattering and radiometric attenuation reduce contrast and suppress high-frequency image details. These factors jointly limit detection reliability in practical surveillance deployments.

Recent advances in convolutional neural networks (CNNs) have improved the accuracy of small-object detection under favorable conditions [2–15]. However, most existing studies focus on algorithmic accuracy improvements without explicitly considering operational constraints such as long-range engagement distance and real-time inference on embedded hardware. In practical military surveillance systems, detection modules must operate within bounded sensing and computational envelopes.

Atmospheric degradation has been extensively studied in optical imaging and remote sensing literature [16–18]. Nevertheless, its coupled impact with embedded latency constraints on UAV detection performance remains insufficiently investigated.

In this work, we address long-range UAV detection as a degradation-constrained military surveillance problem. The main contributions are:

- experimental identification of a degradation-dependent operational regime in which detection performance collapses nonlinearly,
- a latency-aware adaptive image tiling strategy suitable for real-time embedded deployment,
- comprehensive validation on long-range UAV datasets and edge hardware under realistic operational conditions.

2 Related Work in Military UAV Detection

2.1 Long-Range and Small-Object Detection

Although numerous small-object detection methods have been proposed in the computer vision community, their performance under operational military surveillance system conditions remains insufficiently evaluated, particularly in long-range degraded environments [2–13]. While effective at moderate distances, these methods often exhibit severe performance degradation when object scale approaches the resolution limit. Complementary benchmarks such as the UAVDT dataset [19] and transformer-based backbones, including the Swin Transformer [20], have further advanced aerial and small-object detection research. However, most existing works focus on algorithmic improvements without explicitly addressing defense-oriented surveillance constraints such as extended engagement range and real-time embedded deployment [21].

2.2 Atmospheric Visual Degradation

Atmospheric degradation has been modeled using scattering-based formulations such as the Beer–Lambert law and Koschmieder's model [16–18]. Although image restoration and dehazing techniques have been proposed, their integration into real-time detection pipelines remains challenging due to computational overhead.

2.3 Edge-Based Real-Time Detection

Edge deployment of deep learning models has gained increasing attention, particularly with platforms such as NVIDIA Jetson series [22]. In aerospace applications, edge computing architectures are increasingly adopted to reduce communication latency and ensure autonomous operation under limited bandwidth conditions [23]. Techniques including model pruning, quantization, and input resolution adaptation have been explored [24, 25], but few studies explicitly address degradation-aware detection under strict latency constraints.

3 Methodology

3.1 System Model and Problem Formulation

We consider a long-range UAV detection system operating in outdoor aerospace environments, where visual observations are acquired by an optical sensor and processed by a convolutional object detector deployed on an edge computing platform. The system is subject to two coupled constraints: (*i*) progressive visual degradation induced by long-range imaging and atmospheric effects, and (*ii*) strict latency limitations imposed by real-time edge inference.

Let I_0 denote the latent high-quality image of a UAV scene. The observed image I at long range can be modeled as

$$I = D(I_0, d, \alpha, n) \tag{1}$$

where d represents the imaging distance, α characterizes atmospheric attenuation and scattering effects, and n denotes sensor noise following classical atmospheric imaging models [16, 17]. As the distance d increases, the apparent UAV size decreases rapidly, while atmospheric scattering reduces contrast and suppresses high-frequency details critical for reliable detection.

$$s(d) = \frac{f \cdot h}{d} \tag{2}$$

$$I(x) = A + \left[I_0(x) - A \right] \mathrm{e}^{-\beta d} \tag{3}$$

In Eqs (1)–(3), the symbols are interpreted within a sensing-system perspective. I_0 denotes the latent high-fidelity radiometric representation of the UAV scene, corresponding to short-range observation under nominal atmospheric conditions. I represents the long-range sensed image acquired by the onboard imaging system, which undergoes coupled geometric attenuation and atmospheric radiative degradation. The parameter d specifies the line-of-sight distance between the imaging sensor and the UAV target, while α (or equivalently β) denotes the atmospheric extinction coefficient, capturing the cumulative effects of scattering and absorption along the propagation path.

In Eq. (2), $s(d)$ represents the geometric projection scaling factor governed by the imaging geometry, f is the effective focal length of the optical system, and h denotes a characteristic physical dimension of the UAV. This expression formalizes the inverse-distance scaling inherent to perspective projection, highlighting that the apparent spatial footprint of the UAV on the image plane decreases rapidly as d increases. Con-

DOI 10.3849/aimt.02103

sequently, the number of effective pixels encoding target-specific features diminishes, directly constraining detection reliability in long-range regimes.

Eq. (3) describes the classical atmospheric radiative transfer model. Here, $I_0(x)$ denotes the intrinsic scene radiance, A corresponds to the global atmospheric light (airlight) component, and β is the atmospheric extinction coefficient governing transmission loss. As the propagation distance d increases, the direct transmission term decays exponentially, leading to reduced contrast, radiometric compression, and suppression of high-frequency structural details. From a detection-theoretic standpoint, this degradation reduces the signal-to-background discriminability of small UAV targets, thereby inducing performance deterioration in embedded detection modules operating under long-range surveillance conditions.

To realize the aforementioned system model under observation degradation and edge computing resource constraints, we propose a latency-aware detection module. The overall processing pipeline of the system is illustrated in Fig. 1, which clearly delineates the boundary between the physical environment (sensing front-end) and the processing platform (Jetson Orin Nano edge platform). The core detection module employs the YOLOv8s architecture, a one-stage detector comprising a Cross Stage Partial (CSP) backbone for feature extraction, a multi-scale feature aggregation neck, and a decoupled detection head that separates the classification and bounding box regression branches. The 's' (small) configuration is selected to achieve an optimal balance between detection accuracy and inference latency on the Jetson Orin Nano edge platform.

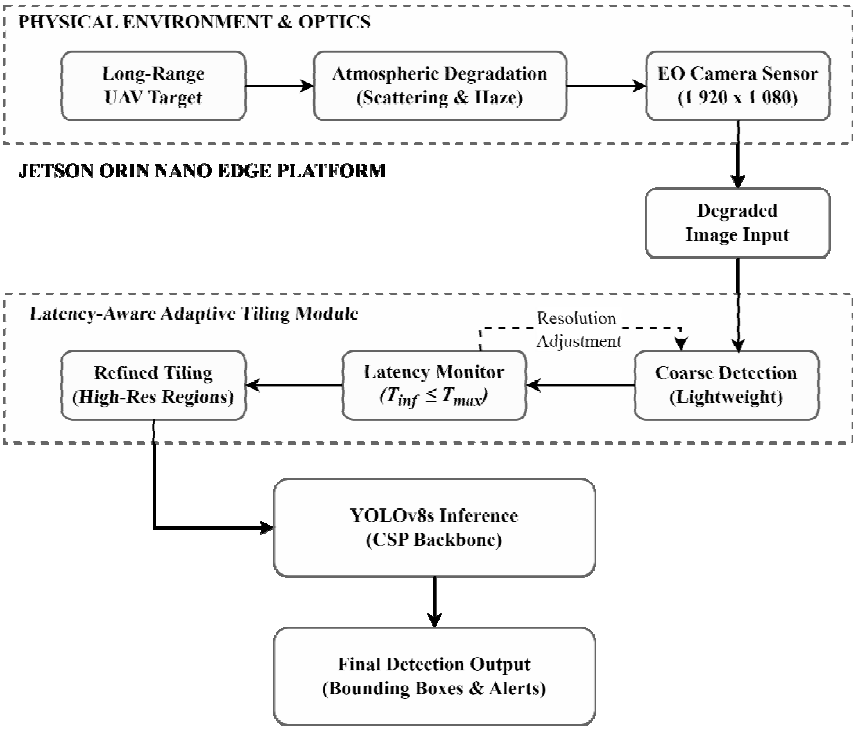


Fig. 1 Schematic block diagram of the proposed degradation-aware UAV detection framework deployed on the edge platform

Detection performance is quantitatively evaluated using the Average Precision (AP) metric, defined as the area under the Precision-Recall curve. A prediction is considered a true positive when the Intersection-over-Union (IoU) between the predicted and ground-truth bounding boxes exceeds a predefined threshold. In this study, the IoU threshold is set to the standard value of 0.5, and the results are reported as mAP@0.5, following common evaluation conventions in object detection tasks. Alongside the detection objective, latency and computational resource constraints play a central role in real-time edge operations. The four system factors – the detection objective function, end-to-end inference latency, computational cost per activated region, and resource budget constraint – are formalized in Eqs (4)–(7) as follows:

$$\max_{\theta} \mathbb{E}\left[\mathrm{mAP}\left(f_{\theta}(I)\right)\right] \quad \text{s.t. } T_{\mathrm{inf}} \leq T_{\max} \tag{4}$$

$$\mathrm{mAP} = \frac{1}{C}\sum_{c=1}^{C} AP_c \tag{5}$$

$$T_{\mathrm{inf}} = T_{\mathrm{base}} + \sum_{j} T_{\mathrm{tile}}\left(r_j\right) \tag{6}$$

$$\sum_{j} C\left(r_j\right) \leq C_{\max} \tag{7}$$

$T_{\mathrm{tile}}(r_j)$denotes the processing time of tile r_j and T_{base} is the baseline inference time. Given the degraded observation I, the detector aims to estimate a set of bounding boxes $\hat{B} = b_i$ and corresponding confidence scores s_i. In edge-based deployment, the inference process must satisfy a latency constraint $T_{\mathrm{inf}} \leq T_{\max}$, where T_{inf} denotes the end-to-end inference time and $T_{\max}$ is the maximum allowable latency for real-time operation. This constraint fundamentally limits the feasible input resolution and computational complexity of the detection model.

The objective of the system is therefore to maintain reliable detection performance while operating within a bounded sensing and computational envelope imposed by real-time deployment constraints. This dual requirement motivates a system-aware detection strategy that jointly considers visual degradation characteristics and latency constraints.

In Eqs (4)–(7), $f_{\theta}(.)$ denotes the detection model parameterized by the learnable weights θ, and mAP represents the mean Average Precision aggregated over all object categories. The term T_{inf} corresponds to the end-to-end inference latency of the sensing-decision pipeline, decomposed into the baseline processing time T_{base} and the additional computational overhead induced by the adaptively activated image regions, denoted as $T_{\mathrm{tile}}(r_j)$.

The parameter r_j designates the j-th spatial region selected for resolution refinement under the adaptive tiling mechanism, while $C(r_j)$ quantifies the associated computational expenditure. The constraint in Eq. (7) enforces that the aggregated computational demand remains within the available resource budget $C_{\max}$, thereby ensuring that the system operates within its admissible latency envelope on edge hardware. From an operational perspective, this constraint defines the feasible performance region within which detection accuracy must be maintained under limited computational resources.

DOI 10.3849/aimt.02103

Taken together, Eqs (1)–(7) provide a unified system-theoretic formulation that models long-range UAV detection as a constrained optimization problem over a degradation-dependent operational performance envelope. The coupled effects of geometric attenuation, atmospheric radiative loss, and finite computational capacity jointly shape the feasible performance region of the detector. Within this framework, the subsequent analysis demonstrates that beyond a critical degradation threshold, the system transitions into an unstable regime characterized by nonlinear performance collapse of conventional convolutional detectors, indicating a practical performance boundary in degraded surveillance conditions.

3.2 Degradation-Induced Performance Breakdown

To understand the limitations of conventional detection pipelines, we first analyze the behavior of standard convolutional detectors under progressive visual degradation. As the effective UAV size and image contrast decrease with distance and atmospheric attenuation, the detector performance does not degrade linearly. Instead, empirical observations reveal a critical operating regime beyond which detection accuracy collapses abruptly.

This phenomenon can be attributed to the loss of discriminative features at early convolutional layers, where low-contrast and low-resolution object cues fall below the effective receptive field and quantization thresholds of the network. Once this regime is reached, increasing model capacity or applying uniform image enhancement yields diminishing returns while significantly increasing computational cost.

From a system perspective, this nonlinear breakdown behavior highlights the inadequacy of static-resolution detection strategies for long-range UAV perception. It also suggests that effective mitigation requires adaptive mechanisms capable of selectively enhancing informative regions without violating real-time constraints.

To quantify the severity of observation degradation using a single scalar variable, we employ a Relative Distance index (D_r). This index simultaneously reflects the geometric attenuation caused by observation distance (according to Eq. (2)) and atmospheric degradation along the propagation path (according to Eq. (3)). This metric is convenient for identifying the critical operating point of the detection system under varying environmental conditions.

Experimental results reveal the existence of a critical operating regime, in which detection performance collapses nonlinearly as the degradation index increases. As shown in Fig. 2a, the mAP remains relatively stable for mild degradation but drops sharply beyond a critical point. The first-order derivative of performance with respect to degradation magnitude, illustrated in Fig. 2b, exhibits a pronounced peak at $D_c \approx 2.5$, which corresponds to the system breaking point identified in the theoretical analysis.

3.3 Latency-Aware Adaptive Image Tiling

To address the identified performance breakdown, we propose a latency-aware adaptive image tiling strategy that dynamically allocates computational resources to regions of interest. Instead of uniformly increasing input resolution, the proposed approach selectively refines spatial resolution in candidate UAV regions while preserving global contextual information.

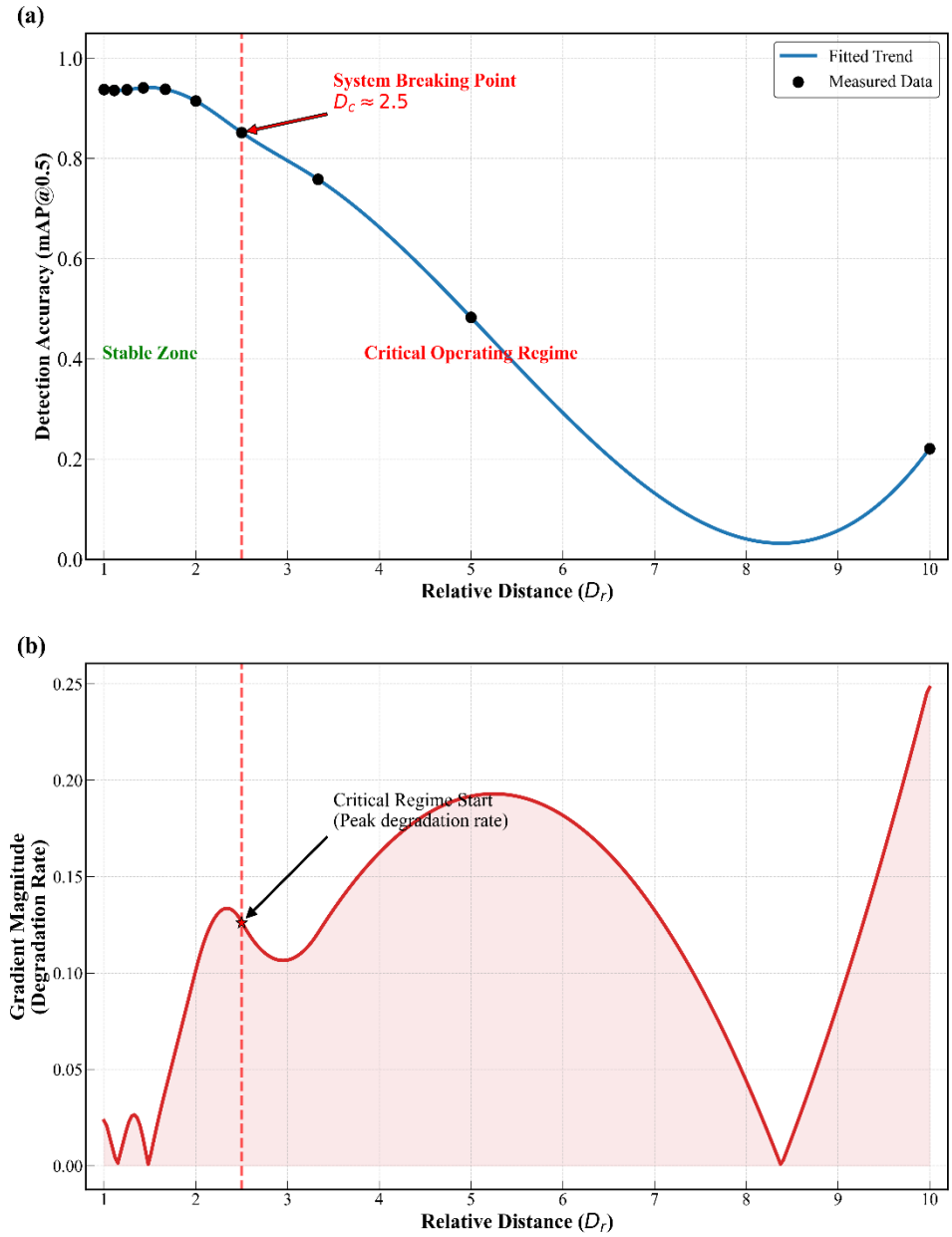


Fig. 2 Experimental identification of the degradation-dependent operational regime in long-range military UAV detection. (a) Detection accuracy versus degradation index (Relative Distance D_r), showing a stable operating zone followed by a sharp performance collapse. (b) Magnitude of the first-order derivative $|\partial mAP/\partial D_r|$, where the peak indicates the system breaking point $D_c \approx 2.5$, corresponding to the onset of nonlinear performance collapse

Specifically, the input image is first processed at a coarse resolution by a lightweight detector to generate preliminary candidate regions with associated confidence scores. Regions whose confidence exceeds a predefined threshold are then reprocessed at higher spatial resolution using localized image tiles. The final detection results are obtained by fusing the coarse and refined predictions.

Let $R = r_i$ denote the set of candidate regions. The adaptive tiling process is constrained by the latency budget T_{max}; if the estimated latency exceeds the allowable budget, the system dynamically adjusts the number or resolution of tiles to maintain real-time performance. This strategy enables the system to concentrate computational effort on visually degraded UAV regions while avoiding unnecessary processing of background areas, achieving a favorable trade-off between detection robustness and latency.

3.4 Integrated Detection Framework

The overall detection framework integrates visual degradation awareness and latency-constrained inference into a unified pipeline. The system operates in a closed-loop manner, where detection confidence and estimated latency jointly guide adaptive resolution refinement.

During inference, the system continuously monitors detection confidence and computational load, enabling dynamic adjustment of processing strategies in response to changing visual conditions. This system-level co-design ensures robust detection performance across a wide range of operating conditions while maintaining real-time constraints on edge platforms. The detection performance of the proposed framework compared with the baseline detector across increasing levels of atmospheric degradation is presented in Fig. 3 and further analyzed in Section 4.2.

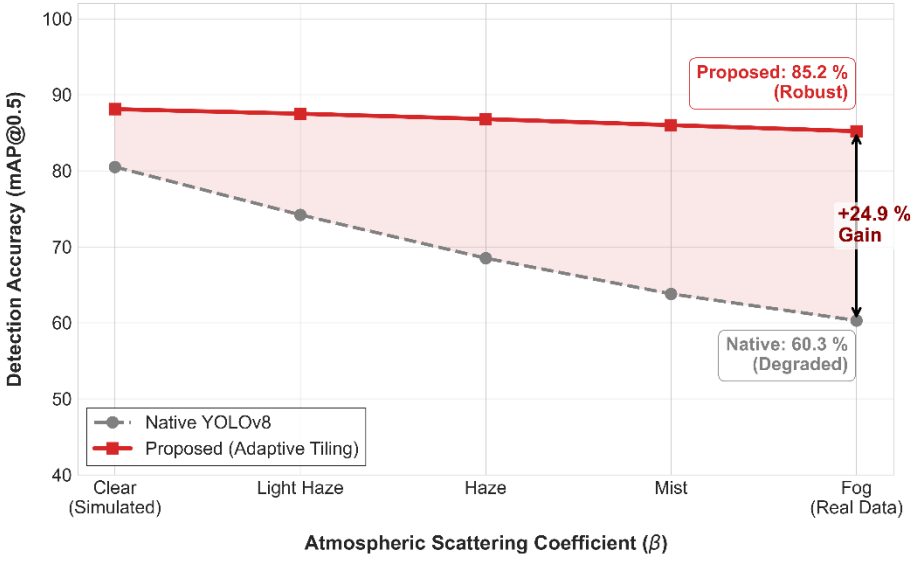


Fig. 3 Detection accuracy of the baseline YOLOv8 detector versus the proposed adaptive tiling framework across increasing levels of atmospheric degradation

4 Experiments and Results

4.1 Experimental Setup and Datasets

The proposed framework is evaluated through a comprehensive set of experiments designed to assess detection robustness under visual degradation and real-time performance under edge deployment constraints. Experiments are conducted on both long-range UAV datasets and public benchmarks to ensure generality and fairness of comparison. Although the operational UAV dataset cannot be fully released due to usage constraints, the degradation modelling protocol and evaluation pipeline are reproducible using the publicly available VisDrone2019 benchmark [26].

Datasets: The primary evaluation is performed on a long-range UAV dataset collected under real outdoor conditions, where UAV targets appear at extended distances and exhibit severe scale reduction and atmospheric degradation. To further validate generalization, experiments are also conducted on the VisDrone2019 dataset, which provides diverse aerial scenes and small-object detection challenges. The selected UAV categories and engagement distances reflect realistic military surveillance systems' conditions, including low-altitude intrusions and extended-range monitoring scenarios.

The experimental configuration reflects practical, realistic military surveillance and airspace protection scenarios, including ground-based long-range optical monitoring and forward-deployed edge sensing nodes. These configurations simulate operational constraints encountered in airspace monitoring and perimeter defense applications.

The optical path of the simulated system is assumed to comprise a Full HD ($1\,920 \times 1\,080$) resolution electro-optical (EO) sensor coupled with a narrow Field of View (FOV) telephoto lens system, a typical configuration for long-range sensor nodes in ground-based optical surveillance. This configuration is intended to simulate the minute pixel occupancy ratio of UAV targets on the image plane at practical operational ranges, as reflected by the target size distribution in Tab. 1. As observed, the apparent target size decreases significantly with observation distance, leading to severe attenuation of high-frequency features and detection capabilities in the long-range regime. This constitutes the primary motivation for the degradation-aware detection method proposed in this study.

Visual degradation modeling: To systematically evaluate robustness, controlled degradation is applied to the input images by varying atmospheric attenuation parameters, including contrast reduction and haze simulation. This enables quantitative analysis of detector behavior across different degradation levels.

Baselines: The proposed method is compared against representative state-of-the-art convolutional detectors configured with fixed input resolution. All baseline models are optimized to achieve their best accuracy–latency trade-off under identical hardware conditions. This ensures a fair comparison under identical hardware and deployment constraints.

Implementation details: All experiments are conducted on an NVIDIA Jetson Orin Nano platform [22] to reflect realistic edge deployment. The maximum allowable inference latency, $T_{\max}$ is set according to real-time operational requirements. Detection accuracy is measured using mean Average Precision (mAP), while latency is measured as the end-to-end inference time averaged over multiple runs. All reported quantitative results are averaged over three independent runs to reduce variance caused

by stochastic training effects. Tab. 1 summarizes the long-range UAV dataset and experimental configurations used in this study.

The selected UAV categories and engagement ranges reflect realistic military surveillance systems scenarios, including low-altitude intrusion detection and extended-range monitoring.

Tab. 1 Dataset characteristics and experimental scenarios for long-range UAV detection

Dataset / Scenario	Number of Instances	Object Size [pixels]	Estimated Range [km]	Visual conditions
Bird	35 026	20–640	0.03–0.90	Clear / Degraded
Fixed-wing	60 333	26–640	0.19–4.62	Haze, Low contrast
Missile	67 551	36–640	0.38–6.67	Heavy haze, Fog
Quadcopter	98 178	26–640	0.05–1.15	Various conditions
Total	**261 088**	**20–640**	**0.03–6.67**	**Mixed**

4.2 Detection Performance under Visual Degradation

We first evaluate detection performance under increasing levels of visual degradation to analyze robustness in long-range scenarios.

As degradation increases, all baseline detectors exhibit a rapid decline in detection accuracy. Notably, beyond a critical degradation level, performance collapses nonlinearly, consistent with the breakdown behavior discussed in Section 3.2. In contrast, the proposed framework demonstrates significantly improved robustness across all degradation levels. The quantitative comparison of detection accuracy under different degradation levels is summarized in Tab. 2.

Tab. 2 Comparison of detection accuracy under different degradation levels (mAP@0.5)

Method	Mild Degradation (Light Haze, $\beta \approx 0.01$)	Moderate Degradation (Mist, $\beta \approx 0.025$)	Severe Degradation (Dense Fog, $\beta \approx 0.04$)
Baseline detector	74.5 %	66.8 %	60.3 %
Full Tiling	88.0 %	86.5 %	85.5 %
Proposed Adaptive Tiling	85.5 % ***(+11.0 % improvement)***	85.4 % ***(+18.6 % improvement)***	85.2 % ***(+24.9 % improvement)***

Quantitatively, the proposed method achieves up to a 24.9 % improvement in mAP compared with the best-performing baseline under severe degradation. This improvement is particularly pronounced in long-range scenarios where UAV targets occupy only a few pixels and contrast is substantially reduced. These results suggest

that robust long-range UAV perception requires joint optimization of sensing resolution and computational allocation, rather than detector architecture alone.

Such degradation-induced collapse is particularly critical in military surveillance systems, where missed detections may delay threat identification.

4.3 Latency and Real-Time Performance

Real-time performance is a critical requirement for practical UAV detection systems. We therefore evaluate the latency characteristics of the proposed framework under edge deployment. Tab. 3 reports the average inference time for baseline detectors and the proposed method on the Jetson Orin Nano platform.

Tab. 3 Latency and throughput comparison on edge hardware

Method	Average latency [ms]	Throughput [FPS]
Baseline	16.5	60.8
Full Tiling	86.2	11.6
Adaptive Tiling (proposed)	40.1	24.8

Despite the additional processing introduced by adaptive tiling, the proposed framework consistently satisfies the predefined latency constraint $T_{\max}$. This is achieved by dynamically limiting the number and resolution of refined tiles based on real-time latency estimation. In contrast, baseline approaches that rely on uniformly increased input resolution exceed the allowable latency budget and are therefore unsuitable for real-time operation.

These results demonstrate that the proposed framework effectively balances detection robustness and computational efficiency, enabling stable real-time inference under strict edge constraints.

4.4 Visualization of System Breakdown

To assess the contribution of individual components, we conduct an ablation study comparing the following configurations: (*i*) baseline detector with fixed resolution, (*ii*) detector with static image tiling, and (*iii*) the proposed latency-aware adaptive tiling strategy.

The results, summarized in Tab. 2, indicate that static tiling provides limited improvement under mild degradation but fails to maintain performance under severe conditions due to excessive latency. In contrast, the proposed adaptive strategy achieves superior detection accuracy while consistently satisfying real-time constraints.

This study confirms that both adaptivity and latency awareness are essential for robust long-range UAV detection, and that naive resolution refinement alone is insufficient. These observations indicate that performance degradation originates from coupled sensing and computational limitations rather than model capacity alone. In military surveillance system applications, this limitation may directly affect early-warning performance and decision-making timelines, reinforcing the importance of degradation-aware detection strategies. From a systems-integration perspective, this behavior underscores the necessity of joint sensing-computation co-design. Treating

perception algorithms independently from hardware constraints may result in unstable operational regimes when deployed on embedded aerospace platforms.

Although the proposed framework significantly improves robustness under degraded conditions, certain failure cases remain. Extremely low-contrast scenarios or heavy occlusions may still result in missed detections, particularly when UAV targets fall below the effective resolution limit even after adaptive refinement.

The experimental results demonstrate that degradation-induced performance collapse constitutes a practical limitation in long-range military optical surveillance. Static-resolution detection strategies are insufficient when deployed under severe atmospheric conditions. The proposed adaptive framework effectively expands the feasible operational envelope of embedded detection systems while respecting real-time latency constraints. From a military systems engineering perspective, such co-design between sensing and computation is essential for maintaining detection reliability in dynamic operational environments.

5 Conclusion and Future Work

This study presented a degradation-aware long-range UAV detection framework tailored for real-time military surveillance systems. By analyzing the coupled effects of atmospheric degradation and embedded computational constraints, a critical operational regime was identified in which conventional detectors exhibit nonlinear performance collapse.

To address this limitation, a latency-aware adaptive tiling strategy was introduced to selectively refine candidate target regions while maintaining bounded inference time. Experimental validation on long-range UAV datasets and embedded hardware demonstrated significant improvements in detection robustness under degraded visual conditions.

The results highlight that reliable long-range UAV detection in military applications requires coordinated sensing-computation design rather than solely increasing model complexity. Future work will explore multi-modal sensing integration and temporal fusion to further enhance robustness under extreme operational conditions.

Acknowledgement

The authors would like to acknowledge the support of their institution for providing the research environment and computational resources necessary to conduct this study. The authors also appreciate the valuable technical discussions and feedback from colleagues that contributed to the development and validation of the proposed degradation-aware UAV detection framework.

References

[1] SHAKHATREH, H., A.H. SAWALMEH, A. AL-FUQAHA, Z. DOU, E. ALMAITA, I. KHALIL, N.S. OTHMAN, A. KHREISHAH and M. GUIZANI. Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. *IEEE Access*, 2019, **7**, pp. 48572–48634. DOI 10.1109/ACCESS.2019.2909530.

[2] REDMON, J., S. DIVVALA, R. GIRSHICK and A. FARHADI. You Only Look Once: Unified, Real-Time Object Detection. In: *2016 IEEE Conference on Com-*

puter Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016, pp. 779–788. DOI 10.1109/CVPR.2016.91.

[3] REDMON, J. and A. FARHADI. YOLO9000: Better, Faster, Stronger. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017, pp. 7263–7271. DOI 10.1109/CVPR.2017.690.

[4] REN, S., K. HE, R. GIRSHICK and J. SUN. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2017, **39**(6), pp. 1137–1149. DOI 10.1109/TPAMI.2016.2577031.

[5] WANG, A., H. CHEN, L. LIU, K. CHEN, Z. LIN, J. HAN and G. DING. YOLOv10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems*, 2024, **37**, pp. 107984–108011. DOI 10.48550/arXiv.2405.14458.

[6] CARION, N., F. MASSA, G. SYNNAEVE, N. USUNIER, A. KIRILLOV and S. ZAGORUYKO. End-to-End Object Detection with Transformers. In: *European Conference on Computer Vision*. Cham: Springer, 2020, pp. 213–229. DOI 10.1007/978-3-030-58452-8_13.

[7] YU, X., Y. GONG, N. JIANG, Q. YE and Z. HAN. Scale Match for Tiny Person Detection. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Snowmass: IEEE, 2020, pp. 1257–1265. DOI 10.1109/WACV45572.2020.9093394.

[8] KISANTAL, M., Z. WOJNA, J. MURAWSKI, J. NARUNIEC and K. CHO. Augmentation for Small Object Detection. *Computer Science & Information Technology*, 2019, pp. 119–133. DOI 10.5121/csit.2019.91713.

[9] WANG, J., C. XU, W. YANG and L. YU. A Normalized Gaussian Wasserstein Distance for Tiny Object Detection [Preprint]. 2021, arXiv:2110.13389. DOI 10.48550/arXiv.2110.13389.

[10] SUNKARA, R. and T. LUO. No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects. In: MR. AMINI et al., eds. *Machine Learning and Knowledge Discovery in Databases*. Cham: Springer, 13715. DOI 10.1007/978-3-031-26409-2_27.

[11] LIN, T.-Y., P. DOLLÁR, R. GIRSHICK, K. HE, B. HARIHARAN and S. BELONGIE. Feature Pyramid Networks for Object Detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017, pp. 2117–2125. DOI 10.1109/CVPR.2017.106.

[12] AKYON, F.C., S.O. ALTINUC and A. TEMIZEL. Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. In: *2022 IEEE International Conference on Image Processing (ICIP)*. Bordeaux: IEEE, 2022, pp. 966–970. DOI 10.1109/ICIP46576.2022.9897990.

[13] LI, X., Y. WEI, J. LI, W. DUAN, X. ZHANG and Y. HUANG. Improved YOLOv7 Algorithm for Small Object Detection in Unmanned Aerial Vehicle Image Scenarios. *Applied Sciences*, 2024, **14**(4), 1664. DOI 10.3390/app14041664.

[14] ZHU, X., S. LYU, X. WANG and Q. ZHAO. TPH-YOLOv5: Improved YOLOv5 Based on a Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. In: *Proceedings of the IEEE/CVF International Conference on Com-*

puter Vision. Montreal: IEEE, 2021, pp. 2778–2788. DOI 10.1109/ICCVW54120.2021.00312.

[15] LIM, J.-S., M. ASTRID, H.-J. YOON and S.-I. LEE. Small Object Detection Using Context and Attention. In: *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. Jeju Island: IEEE, 2021, pp. 181–186. DOI 10.1109/ICAIIC51459.2021.9415217.

[16] MCCARTNEY, E.J. *Optics of the Atmosphere: Scattering by Molecules and Particles*. New York: Wiley, 1976. ISBN 0-471-01526-1.

[17] TAN, R.T. Visibility in Bad Weather from a Single Image. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage: IEEE, 2008. DOI 10.1109/CVPR.2008.4587643.

[18] HE, K., J. SUN and X. TANG. Single-Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(12), pp. 2341–2353. DOI 10.1109/TPAMI.2010.168.

[19] DU, D., Y. QI, H. YU, Y. YANG, K. DUAN, G. LI, W. ZHANG, Q. HUANG and Q. TIAN. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Cham: Springer, 2018, pp. 375–391. DOI 10.1007/978-3-030-01249-6_23.

[20] LIU, Z., Y. LIN, Y. CAO, H. HU, Y. WEI, Z. ZHANG, S. LIN and B. GUO. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021, pp. 10012–10022. DOI 10.1109/ICCV48922.2021.00986.

[21] NALAMATI, M., A. KAPOOR, M. SAQIB, N. SHARMA and M. BLUMENSTEIN. Drone Detection in Long-Range Surveillance Videos. In: *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Taipei: IEEE, 2019. DOI 10.1109/AVSS.2019.8909830.

[22] *NVIDIA Jetson Orin NX Series* [online]. 2022 [viewed 2026-06-15]. Available from: https://developer.nvidia.com/downloads/jetson-orin-nx-series-data-sheet

[23] GUPTA, L., R. JAIN and G. VASZKUN. Survey of Important Issues in UAV Communication Networks. *IEEE Communications Surveys & Tutorials*, 2016, **18**(2), pp. 1123–1152. DOI 10.1109/COMST.2015.2495297.

[24] KINGMA, D.P. and J. BA. *Adam: A Method for Stochastic Optimization* [online]. 2015 [viewed 2025-12-15]. Available from: https://pure.uva.nl/ws/files/326512947/1412.6980v5.pdf

[25] LOSHCHILOV, I. and F. HUTTER. *Decoupled Weight Decay Regularization* [online]. 2019 [viewed 2025-12-15]. Available from: https://scispace.com/pdf/decoupled-weight-decay-regularization-3pkyc25esh.pdf

[26] ZHU, P., L. WEN, X. BIAN, H. LING and Q. HU. *Vision Meets Drones: A Challenge.* 2019. DOI 10.48550/arXiv.1804.07437.