# Optimizing long-range UAV detection on YOLOv8: Breaking-point distance analysis and combining adaptive tiling with AdamW optimizer

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

[1]Institute of Technology, General Department of Defense Industry, Dong Ngac, Hanoi, Vietnam;
[2]Faculty of Aerospace Engineering, Le Quy Don Technical University, 236 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam.
[*]Corresponding author: pham.thanh-dong@lqdtu.edu.vn
*Received 5 Nov. 2025; Revised 28 Jan. 2026; Accepted 10 Feb. 2026; Published 25 Feb. 2026.*
**DOI:** https://doi.org/10.54939/1859-1043.j.mst.109.2026.154-163

## ABSTRACT

*The rapid proliferation of unmanned aerial vehicles (UAVs) has imposed stringent requirements on surveillance and early warning systems. In long-range detection scenarios, the apparent size of UAVs in images decreases significantly, leading to severe spatial information loss and degraded performance of convolutional neural network (CNN)-based detection models. This paper proposes a continuous quantitative analysis framework to model the relationship between observation distance and UAV detection performance by progressively reducing the input image resolution. Based on experimental regression analysis, a system-level breaking point is identified, representing a distance threshold at which detection performance begins to degrade sharply and exhibits nonlinear behavior. Furthermore, a solution integrating adaptive image tiling with the AdamW optimizer is proposed to ensure training stability and enhance performance in long-range scenarios. Experimental results on the YOLOv8s model show that the proposed approach improves mAP@0.5 in long-range detection by up to +24.9% while eliminating numerical instability during training on tiled data. Regression analysis identifies the system-level breaking point at $D_c \approx 2.5$, providing a quantitative basis for activating adaptive image processing in real-world deployments on resource-constrained platforms.*

**Keywords:** UAV; Small object detection; YOLOv8; Image tiling; AdamW; Breaking point; Computer vision.

## 1. INTRODUCTION

The rapid proliferation of Unmanned Aerial Vehicles (UAVs) in both civilian and military sectors has imposed increasingly stringent requirements on surveillance, early warning, and airspace protection systems. In long-range detection scenarios, the projected size of UAVs in images decreases significantly, causing the target to occupy only a very small number of pixels on the sensor. This phenomenon leads to a severe loss of spatial information, degrading the capability for feature extraction and directly impacting the performance of object detection models based on Convolutional Neural Networks (CNNs) [4].

Real-time object detection models, represented by the YOLO family [5], have achieved significant progress in processing speed and overall accuracy. However, most of these models are designed and evaluated in contexts where objects have medium to large sizes. When applied to long-range UAV detection—where targets often possess very small dimensions and thin geometric details—the model performance degrades rapidly and nonlinearly. This presents challenges not only regarding model architecture but also in terms of evaluation, training, and system deployment under varying observation conditions.

### 1.1. Advances in small object detection

In recent years, many state-of-the-art object detection models have been proposed to improve overall accuracy, notably YOLOv10 [9] or Transformer-based architectures such as DETR [10] and Swin Transformer [11]. Despite achieving high performance on standard datasets, small object

detection is still considered one of the most difficult problems in computer vision due to severe limitations in spatial feature information [4].

To address this issue, several specialized approaches have been developed. Wang et al. [12] proposed the Normalized Gaussian Wasserstein Distance (NWD) metric to replace traditional IoU, helping to increase sensitivity to small deviations in object localization. Sapa et al. [13] introduced SPD-Conv to replace strided convolutions and pooling, thereby better preserving the fine-grained information of small objects. Furthermore, multi-resolution architectures and hierarchical Transformers like Swin Transformer [11] have also shown high effectiveness on representative UAV and drone datasets such as VisDrone [16] and UAVDT [17], as well as ultra-small object scenarios in specialized datasets like TinyPerson [18].

However, the aforementioned methods often require significant changes to the network architecture or increase computational costs, making them difficult to deploy in real-time surveillance systems on edge devices—where computational resources, memory, and power are limited.

### 1.2. High-resolution image processing strategies and stable training

Instead of deeply intervening in the model architecture, a more practical approach is to optimize the input data and training strategies. Methods such as Feature Pyramid Networks (FPN) [14] allow for the exploitation of multi-scale information, while data augmentation techniques for small objects [15] help improve the model's generalization capability. Notably, image tiling (or slicing) techniques such as SAHI [3] have proven effective in preserving the local resolution of small objects, especially during the inference phase.

However, direct application of image tiling strategies during the training phase often encounters convergence instability issues. The primary cause is that tiled data increases gradient variance, as many image patches contain few or no objects, or only contain truncated object parts. In this context, several recent studies have indicated that combining appropriate training strategies and gradient optimization plays a crucial role in small object detection [19], particularly in UAV applications using YOLO variants [20].

Therefore, rather than focusing on designing new architectures, this study aims for a more complementary and practical approach: quantitatively analyzing the relationship between observation distance and UAV detection performance, while proposing a stable training configuration based on combining adaptive image tiling with the AdamW optimizer. This approach inherits the advantages of previous research while ensuring feasibility for deployment on constrained computing platforms.

### 1.3. Objectives and contributions of the paper

From the above analyses, it can be seen that although there have been many studies on small object and UAV detection, there is still a lack of research works that quantitatively and continuously analyze the degradation of detection performance relative to observation distance, as well as identify the critical thresholds at which the system begins to lose effectiveness systematically. Furthermore, the issue of training stability on tiled data in the context of practical deployment has not been sufficiently investigated. Originating from these gaps, the paper poses two main research questions: (i) According to what laws does UAV detection performance degrade as the observation distance increases continuously? (ii) How can stable training be maintained and detection performance be enhanced in long-range scenarios on constrained computing platforms?

## 2. PROBLEM FORMULATION

### 2.1. Distance-based spatial degeneration model

Based on geometric optics, the pixel size of an object on the sensor ($h_{pix}$) is inversely proportional to the observation distance ($D$):

$$h_{pix} = \frac{f \cdot H_{real}}{D \cdot p_{size}} \quad (1)$$

where $f$ is the focal length, $H_{real}$ is the physical size of the object, and $p_{size}$ is the pixel size on the sensor. Since the effective spatial information area ($S_{info}$) is proportional to the square of the object's image size, the extractable information decreases as follows:

$$S_{info} \propto \left(\frac{1}{D}\right)^2 \quad (2)$$

When the distance doubles, spatial information decreases by approximately 75%, causing aliasing and the loss of high-frequency features.

### 2.2. Gradient optimization with AdamW

Training on tiled data often results in high-variance gradients because many tiles contain little to no target information. This makes traditional optimizers like SGD prone to instability. AdamW improves convergence and generalization in noisy data environments by decoupling weight decay from the gradient update.

Let $g_t = \nabla_\theta L_t(\theta_t)$ be the gradient at iteration $t$, the AdamW optimizer updates the first and second moments as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (3)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (4)$$

The parameter update rule is defined by:

$$\theta_{t+1} = \theta_t - \eta \left( \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda \theta_t \right) \quad (5)$$

Where $\eta$ is the learning rate and $\lambda$ is the decoupled weight decay coefficient. This mechanism helps improve the convergence capability and generalization of the model when training on noisy data.

In the context of training on fragmented data, the gradient at each iteration exhibits high variance due to the emergence of numerous image patches containing few or no objects. When utilizing SGD with L2 regularization, the weight decay component is mixed directly into the gradient, which can amplify fluctuations in parameter updates. Conversely, AdamW completely decouples the weight decay component from the gradient while simultaneously normalizing the gradient according to the second moment. This mechanism helps reduce the influence of noisy samples and maintains numerical stability throughout the training process.

### 2.3. Evaluation metrics

Performance is measured using Mean Average Precision (mAP). This study utilizes the MS COCO standard: mAP@0.5 for object detection capability and mAP@0.5:0.95 for bounding box precision.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} \int_0^1 P_i(R) dR \quad (6)$$

## 3. RESULTS AND DISCUSSION

### 3.1. Dataset and simulation setup

The dataset comprises four classes: bird, fixed-wing, missile, and quadcopter. The dataset is constructed in the standard YOLO format and is partitioned into three independent sets: training (train), validation, and an independent test set. In total, the dataset contains 243,036 images featuring the four object classes: bird, fixed-wing, missile, and quadcopter. The training set

consists of 222,681 images with 239,513 objects, while the validation set includes 12,684 images and is used to monitor the convergence process. The test set encompasses 7,671 images with 8,209 objects and is utilized to evaluate all quantitative results reported in this paper, including the mAP@0.5 improvement of up to + 24.9%.

To overcome the limitations of evaluating at discrete points, this study establishes a continuous survey procedure to simulate variations in distance by progressively reducing the input image resolution. The relative distance $D_r$ is determined based on the resolution reduction ratio, with a reference resolution $S_{original} = 640$, as follows:

$$D_r = \frac{S_{original}}{S_{input}} = \frac{640}{S_{input}} \quad (7)$$

Experiments were conducted with $S_{input}$ ranging from 640 pixels down to 64 pixels, corresponding to $D_r \in [1, 10]$, enabling the construction of continuous performance curves and the quantitative identification of the breaking point - the distance threshold where detection performance declines abruptly.

It should be noted that the relative distance $D_r$ in this study does not represent the absolute physical distance (km) between the UAV and the sensor, but rather an abstraction designed to model the degree of visual representation degradation of the object in the input image. Utilizing $D_r$ allows for the establishment of a continuous quantitative relationship between observation distance and detection performance independent of specific hardware parameters, such as focal length, sensor size, or camera resolution. In practice, $D_r$ can be fully mapped to physical distance (km) if the intrinsic parameters of the optical system and the actual size of the UAV are known. However, such conversion is beyond the scope of this study, as it may reduce the generality of the proposed analytical framework.

### 3.2. Proposed method and training configuration

To resolve the issue of information loss, we apply an Adaptive Tiling strategy. Assume the input image $I \in \mathbb{R}^{H \times W}$ is divided into sub-patches $\{T_i\}$ with size $s \times s$ and an overlap ratio $\alpha$. The stride between patches is determined by:

$$\Delta = s \cdot (1 - \alpha) \quad (8)$$

To reduce the influence of patches containing incomplete object information, a filtering criterion is applied. Given the original bounding box $b$ and its intersection within patch $b_i$, the patch is discarded if:

$$\frac{\text{Area}(b \cap b_i)}{\text{Area}(b)} < \tau \quad (9)$$

Where $\tau$ is the intersection area ratio threshold. The YOLOv8s model was selected due to its ability to well-balance accuracy and processing speed. The training process was conducted with an image size of 640×640, an appropriate number of epochs, and the AdamW optimizer to ensure stable convergence. Safety mode: Disable AMP (Automatic Mixed Precision) to prevent the NaN loss error observed in preliminary experiments with SGD.

### 3.3. Results and analysis

#### *3.3.1. Training stability*

Before analyzing performance relative to distance, this study evaluates the stability and convergence of the YOLOv8s model when utilizing the AdamW optimizer. Since tiled data frequently contains background noise and truncated objects, controlling loss function fluctuations is a critical requirement. Figure 1 illustrates the progression of the loss function (Box Loss) and mAP@0.5 throughout 75 training epochs.

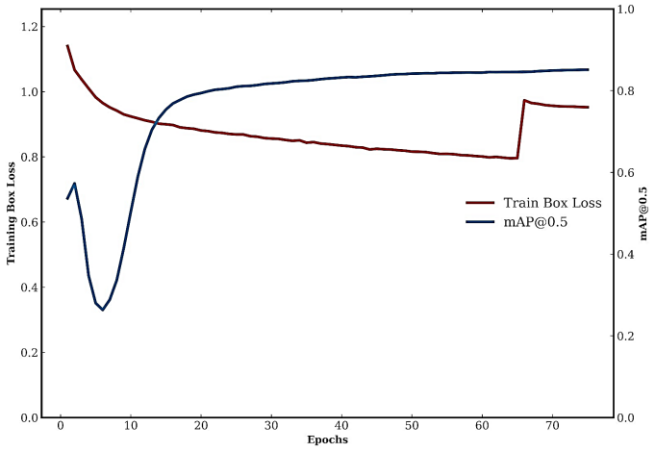


***Figure 1.*** *Progression of training loss (Box loss) and validation accuracy (mAP@0.5) over epochs when training YOLOv8s with AdamW.*

It is observed from figure 1 that during the initial stage, the warmup strategy facilitates a rapid decrease in the loss function and stabilizes the parameter update process, while simultaneously significantly improving validation accuracy. As the number of epochs increases, both Box loss and mAP@0.5 approach a state of convergence, reflecting the stable training process of the model.

In the final stage, when the Mosaic augmentation technique is disabled and the model is fine-tuned on original images, a slight temporary increase in Box loss occurs due to the shift in training data distribution. However, validation accuracy continues to improve and reaches a saturation state, indicating enhanced generalization capabilities of the model. These results confirm the role of the AdamW optimizer in maintaining training stability, particularly when combined with data augmentation strategies and late-stage fine-tuning.

*3.3.2. Ablation study*

To demonstrate the effectiveness of the proposed method, the study conducts an ablation experiment based on three strategic configurations:

- Baseline (M0): Original YOLOv8s, 640x640 input, SGD Optimizer (Default configuration).
- M1 (Optimization-only): YOLOv8s + AdamW (Without Tiling).
- M2 (Proposed): YOLOv8s + Tiling + AdamW.

Results in table 1 show that the default configuration (M0) with SGD failed due to gradient explosion at Epoch 19. The proposed M2 configuration (Tiling + AdamW) achieved stable training and a +24.9% mAP improvement in far-range scenarios compared to the non-tiling baseline (M1).

***Table 1.*** *The results of the exclusion study.*

| ID | Model | Tiling strategy | Optimizer | Training status | mAP50 (overall) | mAP50 (far scenario) | Improvement (far range) |
|---|---|---|---|---|---|---|---|
| M0 | YOLOv8s | None | SGD | Failed | 85.1%* | < 20% | - |
| M1 | YOLOv8s | None | AdamW | Stable | 78.5% | 35.4% | Baseline |
| M2 | YOLOv8s | Yes | AdamW | Stable | 85.2% | 60.3% | +24.9% |

Note: The results for M0 were recorded at the final epoch preceding the training instability. Experimental evidence indicates that SGD encountered a gradient explosion phenomenon, with the Box Loss function reaching infinity (inf) at Epoch 19.

Discussion on superiority over traditional SAHI Fine-tuning: In principle, the M2 configuration

shares the image slicing technique with the SAHI Fine-tuning method [3]. However, traditional methods typically employ the default SGD optimizer, which is highly sensitive to noisy data samples (such as truncated objects or empty backgrounds within image tiles). Empirical evidence from the M0 configuration confirms this incompatibility: SGD failed to maintain stable weights, resulting in numerical errors (inf) and disrupting the convergence process. Conversely, the integration of AdamW in the proposed configuration (M2) effectively resolved this issue. Due to the decoupled weight decay mechanism, AdamW ensures robust numerical stability throughout the training process, even when the input data exhibits high gradient variance resulting from image tiling. Consequently, the proposed solution can be regarded as a "Stably Optimized" version of fine-tuning techniques on fragmented data.

Analysis: Impact of the Optimizer: Comparative results highlight the critical role of AdamW. While SGD led to gradient explosion due to its inability to adapt to the high variance of Tiling data, AdamW facilitated a smooth and safe convergence of the loss function curve. Although the overall mAP of SGD (85.1%) approached that of AdamW (85.2%), the occurrence of the inf error rendered SGD unsuitable due to excessively high systemic risk.

Impact of Tiling: A comparison between M1 (without Tiling) and M2 (with Tiling) reveals that this strategy contributes a significant increase of + 24.9% mAP in long-range scenarios. This confirms the hypothesis that increasing local resolution is the key factor in recovering the geometric features of small objects, enabling the model to overcome the physical limitations of the camera sensor.

*3.3.3. Comparison with State-of-the-Art (SOTA) methods*

To evaluate the standing of the proposed solution, we compare YOLOv8s with other popular object detection architectures, including Faster R-CNN [6], YOLOX-s [7], and YOLOv5s [8]. Metrics regarding the number of Parameters and computational complexity (GFLOPs) were aggregated from their respective original technical reports. The processing speed (FPS) was recorded based on the standard performance levels of the NVIDIA Jetson Orin Nano device in FP16 (Floating Point 16) computation mode.

***Table 2.*** *Theoretical comparison with SOTA models on the Jetson edge device.*

| Models | Parameters | GFLOPs | Speed (FPS) on jetson | Suitability assessment |
|---|---|---|---|---|
| Faster R-CNN [6] | ~ 41 M | ~ 180 | < 5 | Unsuitable (too slow) |
| YOLOX-s [7] | ~ 9 M | ~ 26 | ~ 15 | Fair (lower accuracy than v8) |
| YOLOv5s [8] | ~ 7 M | ~ 16 | ~ 35 | Good (older than v8) |
| YOLOv8s (proposed) [1] | ~ 11 M | ~ 28 | ~ 55 | Optimal (well-balanced) |

Observation: The data indicates that although Faster R-CNN [6] possesses a specialized two-stage architecture for high precision, its excessive computational cost (180 GFLOPs) renders it unfeasible for real-time edge applications. In contrast, YOLOv8s provides the optimal balance: it features a slightly higher computational complexity compared to YOLOX-s [7], but offers superior detection capabilities due to its advanced backbone and head architecture. Optimization experiments using TensorRT in FP16 mode on the Jetson Orin Nano demonstrate that YOLOv8s achieves approximately 55 FPS in standard mode (Native 640 x 640). However, when implementing the adaptive tiling strategy (utilizing a 2 x 2) for small object detection, the computational workload increases approximately fourfold, compounded by the overhead of result stitching. Consequently, the overall processing speed drops to an average of 11.5 FPS. Although this rate is below the ideal real-time threshold (30 FPS), it represents an acceptable trade-off for static surveillance tasks, as the mAP@0.5 increases by up to +24.9% at long ranges, ensuring the system does not miss critical targets.

It should be emphasized that this study is not intended to replace state-of-the-art object detection architectures for small objects, such as YOLOv10, DETR, or methods based on NWD and SPD-Conv. Instead, the proposed distance analysis framework and stable training strategy are complementary in nature and can be integrated directly with these advanced architectures and loss functions to enhance performance in long-range UAV detection scenarios.

*3.3.4. Performance characteristics analysis by distance*

To obtain a more comprehensive perspective beyond evaluating discrete points, this paper conducts a continuous survey across a relative distance range ($D_r$) from 1 to 10, corresponding to a reduction in input image resolution from 640 px down to 64 px for all four classes.

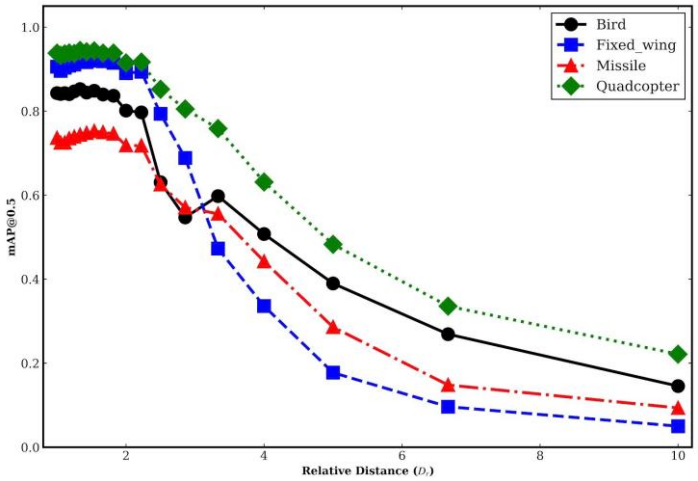


***Figure 2.*** *Precision degradation of four object classes relative to distance.*

Although different object classes exhibit distinct degradation characteristics, the most pronounced slope change on the average mAP curve occurs at $D_c \approx 2.5$, which is consistent with the system breaking point quantitatively estimated in section 3.3.5.

Experimental results indicate a strong divergence between classes: Most robust class (quadcopter):

- The green line represents the best resilience to scale variations. The Quadcopter maintains precision above the 0.5 threshold (usable) up to a distance of $D_r$ = 5. This can be explained by the characteristic centrosymmetric structure (X or "+" shape) of the airframe, which allows the CNN to easily extract features even at low resolutions.

- Rapid Collapse Phenomenon (Fixed_wing): An interesting observation is the Fixed_wing class (blue line). Despite having very high precision at close range (> 0.9), the performance of this class plunges rapidly after the $D_r$ = 2.5 threshold, even falling below the Bird and Missile classes at ultra-long distances (> 5x). The cause is that fixed-wing wings are typically very thin; as resolution decreases, the wings disappear completely, making the aircraft appear as a simple line. Further analysis reveals that the "plunging" phenomenon observed in the Fixed-wing class is not stochastic but originates from the geometric characteristics of the object. Specifically, the wing structure of fixed-wing aircraft possesses an extremely small thickness-to-span ratio. As the image resolution degrades relative to the distance $D_r$, these thin details are suppressed prematurely, leading to the loss of discriminative geometric features.

In many instances, the projection of a Fixed-wing aircraft at low resolution is reduced to a single line or an elongated blurred region, hindering its discriminability against background noise or other linear objects. This explains why the performance of the Fixed-wing class declines more rapidly compared to classes with symmetrical structures or thicker cross-sections, such as Quadcopters, under identical distance conditions.

- Difficult Object Group (Missile & Bird): These two classes have lower initial precision (~ 0.75 ÷ 0.85) and decline steadily. Specifically, the Missile class (orange line) reaches its floor the fastest due to its extremely small cross-section.

*Experimental Conclusion*: These results confirm that for targets with thin details (such as fixed-wing wings or missile bodies), the Tiling strategy is necessary when the distance exceeds $D_r$ = 2.5. Meanwhile, for Quadcopters, the system can operate stably at longer ranges without requiring image processing intervention.

*3.3.5. Quantitative validation of distance-based degradation model and breaking point estimation*

To validate the relationship between observation distance and detection performance, this study hypothesizes that detection precision degrades according to a power law of the relative distance $D_r$, consistent with the inverse-square law model of spatial information degradation. The regression model used is:

$$mAP(D_r) = \alpha \cdot D_r^{-\beta} \tag{10}$$

The regression parameters were estimated using the least squares method after linearization in the log–log domain. Regression results show that the degradation coefficient β ranges between 0.55 ÷ 1.20, with a coefficient of determination $R^2 \approx 0.82$, suggesting that performance degrades according to a power law weaker than a pure inverse-square model.

In this context, $R^2$ is the coefficient of determination, utilized to assess the goodness-of-fit of the regression model in equation (10) relative to the experimental data. This metric reflects the proportion of variance in detection performance that is explained by the degradation model as a function of the relative distance $D_r$.

A high $R^2$ value indicates that the proposed power-law model is highly capable of describing the performance degradation trend relative to the observation distance. It should be noted that $R^2$ is not a model parameter, but rather a metric for evaluating the quality of the regression.

***Table 3.*** *Results of performance degradation regression by distance.*

| **Class** | $\boldsymbol{\beta}$ | $\boldsymbol{R^2}$ |
|---|---|---|
| Bird | 0.673 | 0.857 |
| Fixed-wing | 1.195 | 0.828 |
| Missile | 0.807 | 0.798 |
| Quadcopter | 0.553 | 0.809 |
| **Average** | **0.807** | **0.823** |

In addition to the overall decline trend, piecewise linear regression on the average mAP indicates a threshold at which the degradation slope increases sharply:

$$mAP(D_r) = \begin{cases} a_1 D_r + b_1, & D_r \le D_c \\ a_2 D_r + b_2, & D_r > D_c \end{cases} \tag{11}$$

The estimation results identify the system breaking point at $D_c \approx 2.5$. Beyond this threshold, detection performance declines rapidly due to systematic spatial information loss, which can no longer be effectively compensated for by the feature learning capabilities of the CNN.

It should be noted that the distance simulation procedure in this study is based on reducing image resolution to reflect the degradation of the object's representational size. Other physical effects, such as optical blur, atmospheric turbulence, and light scattering, have not yet been considered within the scope of this paper. Consequently, the threshold value $D_c$ is understood as a systemic threshold within the visual feature domain, rather than an absolute physical distance. The integration of blur and noise models to more closely simulate realistic long-range scenarios will be considered in subsequent studies.

## 4. CONCLUSIONS

The study has successfully developed a comprehensive solution framework for long-range UAV and missile detection, integrating an adaptive tiling strategy and AdamW optimization. Experimental results over a continuous distance range (1x ÷ 10x) lead to the following findings:

- Optimization Effectiveness: The utilization of AdamW significantly mitigates the risk of gradient explosion when training on fragmented data and maintains numerical stability across the experimental configurations considered.

- Long-range Performance Improvement: The Tiling strategy is proven to be a prerequisite factor, enhancing mAP@0.5 accuracy in long-range scenarios by up to +24.9% compared to the baseline method, effectively mitigating feature vanishing phenomena.

- Critical Threshold (Breaking Point): Quantitative analysis across a continuous distance range shows that model performance remains relatively stable within the region $D_r \in [1, 2.5]$. When the relative distance exceeds the critical threshold $D_c \approx 2.5$, detection accuracy declines rapidly and non-linearly due to the systematic loss of the object's spatial information. Although certain classes begin to degrade earlier (approximately $D_r \approx 1.8$ as observed in figure 2), this value only reflects initial signs of degradation, whereas $D_c \approx 2.5$ represents the system breaking point. This point is determined by piecewise linear regression on the average mAP and is therefore selected as the official operational threshold for the system.

- Class-specific Characteristics: There is a clear divergence in robustness based on geometric structure. The Quadcopter class (centrosymmetric) exhibits the best resilience up to $D_r = 5$, while the Fixed-wing and Missile classes (thin details) require higher resolution to maintain recognition. Recommendations: For practical systems on edge devices such as the Jetson Orin Nano, a two-level adaptive image processing activation mechanism is proposed: (i) use $D_r \approx 1.8$ as an early degradation indicator for targets with thin details; and (ii) mandatory activation of Tiling/SAHI strategies or optical zoom when the relative distance exceeds the system breaking point $D_c \approx 2.5$, in order to ensure an optimal balance between accuracy and computational cost.

## REFERENCES


[1]. G. Jocher, A. Chaurasia, J. Kwon, *"Ultralytics YOLOv8"*, GitHub Repository, (2023).

[2]. Loshchilov, F. Hutter, *"Decoupled Weight Decay Regularization"*, International Conference on Learning Representations (ICLR), (2019).

[3]. F. Akyon et al., *"Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection"*, IEEE International Conference on Image Processing (ICIP), pp. 966–970, (2022).

[4]. Y. Liu et al., *"Deep Learning for Small Object Detection: A Survey"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2020).

[5]. J. Redmon et al., *"You Only Look Once: Unified, Real-Time Object Detection"*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, (2016).

[6]. S. Ren, K. He, R. Girshick, J. Sun, *"Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks"*, Advances in Neural Information Processing Systems (NeurIPS), (2015).

[7]. Z. Ge et al., *"YOLOX: Exceeding YOLO Series in 2021"*, arXiv:2107.08430, (2021).

[8]. G. Jocher et al., *"YOLOv5 by Ultralytics"*, GitHub Repository, (2020).

[9]. Wang et al., *"YOLOv10: Real-Time End-to-End Object Detection"*, arXiv:2405.14458, (2024).

[10]. N. Carion et al., *"End-to-End Object Detection with Transformers"*, European Conference on Computer Vision (ECCV), pp. 213–229, (2020).

[11]. Z. Liu et al., *"Swin Transformer: Hierarchical Vision Transformer using Shifted Windows"*, IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10012–10022, (2021).

[12]. J. Wang et al., *"A Normalized Gaussian Wasserstein Distance for Tiny Object Detection"*, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1886–1895, (2022).

[13]. R. Sapa, J. Kim, S. Lee, *"SPD-Conv: Building Efficient CNNs for Small Object Detection"*, arXiv:2208.03635, (2022).

[14]. T.-Y. Lin et al., *"Feature Pyramid Networks for Object Detection"*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2117–2125, (2017).
[15]. M. Kisantal et al., *"Augmentation for Small Object Detection"*, arXiv:1902.07296, (2019).
[16]. P. Zhu et al., *"Vision Meets Drones: A Challenge"*, arXiv:2001.06303, (2020).
[17]. D. Du et al., *"The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking"*, European Conference on Computer Vision (ECCV), pp. 370–386, (2018).
[18]. X. Yu et al., *"Scale Match for Tiny Person Detection"*, IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1257–1266, (2020).
[19]. H. Zhang et al., *"Context-Aware Learning for Small Object Detection"*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 6, pp. 3671–3684, (2022).
[20]. C. Chen et al., *"Optimization for Small Object Detection in UAV Images based on Improved YOLOv7"*, Drones, vol. 7, no. 2, p. 87, (2023).

## TÓM TẮT

### Tối ưu hóa phát hiện UAV tầm xa trên YOLOv8: Phân tích điểm gãy theo khoảng cách và huấn luyện ổn định với phân mảnh thích ứng

*Sự gia tăng nhanh chóng của các phương tiện bay không người lái (UAV) đặt ra yêu cầu cao đối với các hệ thống giám sát và cảnh báo sớm. Trong các kịch bản phát hiện ở cự ly xa, kích thước biểu diễn của UAV trên ảnh suy giảm mạnh, gây mất mát thông tin không gian và làm giảm hiệu năng của các mô hình phát hiện dựa trên mạng nơ-ron tích chập (CNN). Bài báo đề xuất một khung phân tích định lượng liên tục nhằm mô hình hóa mối quan hệ giữa khoảng cách quan sát và hiệu năng phát hiện UAV thông qua việc giảm dần độ phân giải ảnh đầu vào. Trên cơ sở hồi quy thực nghiệm, nghiên cứu xác định điểm gãy hệ thống - ngưỡng khoảng cách mà tại đó hiệu năng phát hiện bắt đầu suy giảm mạnh và mang tính phi tuyến. Đồng thời, một giải pháp kết hợp phân mảnh ảnh thích ứng và bộ tối ưu AdamW được đề xuất nhằm đảm bảo tính ổn định huấn luyện và nâng cao hiệu năng trong kịch bản tầm xa. Kết quả thực nghiệm trên mô hình YOLOv8s cho thấy phương pháp đề xuất cải thiện mAP@0.5 trong kịch bản cự ly xa lên tới +24.9%, đồng thời loại bỏ hiện tượng mất ổn định số học khi huấn luyện trên dữ liệu phân mảnh. Phân tích hồi quy xác định điểm gãy hệ thống tại $D_c \approx 2.5$, cung cấp cơ sở định lượng cho việc kích hoạt xử lý ảnh thích ứng trong triển khai thực tế trên các nền tảng tính toán hạn chế.*

**Từ khoá:** UAV; Phát hiện đối tượng nhỏ; YOLOv8; Phân mảnh ảnh; AdamW; Điểm gãy; Thị giác máy tính.