Data-driven optimization of YOLOv8n for drone search missions

Dinmukhammed Orkushpayev[1,*,†], Saya Sapakova[1,†], Zhansaya Bekaulova[1,†], Aknur Yemberdiyeva[1,†] and Aigul Agybayeva[2,†]

[1] International Information Technology University, Manas St. 34/1, Almaty, 050000, Kazakhstan
[2] Communal State Institution "Abai Village Secondary School", Almaty, Kazakhstan

Abstract

Real-time object detection on Unmanned Aerial Vehicles (UAVs) is critical for time-sensitive applications like search and rescue, yet its efficacy is severely constrained by on-board computational limitations and the challenging nature of aerial imagery. While lightweight architectures such as YOLOv8 are available, a systematic framework for their optimal configuration across diverse operational scenarios is largely absent. This study addresses this gap by conducting an empirical analysis of the YOLOv8n model to establish a hierarchical methodology for its performance optimization. We performed a comprehensive suite of ablation studies on Heridal dataset with contrasting native resolutions (416×416 and 4000×3000), systematically evaluating the impact of data augmentation strategies, input processing resolution, and post-processing parameters. Our results demonstrate that a specialized augmentation pipeline is a foundational prerequisite, improving mAP50 by 15-19 percentage points over standard approaches. We identified an "optimal operational window" for input resolution, with 640×640 pixels being ideal for low-resolution data and 768×768 pixels for high-resolution data, achieving a peak of mAP50 of 70.5%. Furthermore, we establish that the confidence threshold acts as a tactical regulator for the Precision-Recall trade-off, with a value of ~0.2 providing the best-balanced performance. This work provides a data-driven framework for configuring detectors on UAVs, enabling practitioners to maximize performance for specific mission profiles and enhancing the reliability of autonomous search systems.

Keywords

UAV, object detection, deep learning, computer vision, YOLO, optimization

1. Introduction

Unmanned Aerial Vehicles (UAVs) have become instrumental in time-critical applications such as search and rescue, where their ability to rapidly survey hazardous and inaccessible terrain offers a significant advantage over traditional methods. The operational success of these missions is fundamentally predicated on the performance of their on-board perception systems, which are tasked with real-time object detection from a challenging aerial perspective. However, deploying high-performance deep learning models onto UAVs introduces a severe challenge defined by a dual set of constraints: the intrinsic limitations of the platform including restricted computational power, memory, and energy budgets - and the extrinsic difficulties of the data, which is characterized by small object scales, extreme perspective shifts, and variable illumination. This conflict creates a critical tension between the demand for high-accuracy detection and the operational necessity for low-latency, real-time processing.

While a significant body of research focuses on developing novel network architectures, the performance of any autonomous system is ultimately bottlenecked by the efficacy of its foundational perception module. Lightweight, single-stage architectures like YOLO (You Only Look Once) are often selected for their favorable speed-accuracy trade-off, yet their out-of-the-box performance can

[1] *AI 2025: Workshop on Artificial Intelligence, November 19-20, 2025, Almaty, Kazakhstan*
* Corresponding author.
† These authors contributed equally.
s.sapakova@iitu.edu.kz (S. Sapakova); zh.bekaulova@iitu.edu.kz (Zh. Bekaulova); a.yemberdiyeva@iitu.edu.kz (A. Yemberdiyeva)
0000-0001-6541-6806 (S. Sapakova); 0009-0000-9339-9222 (Zh. Bekaulova); 0009-0005-5078-2412 (A. Yemberdiyeva)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-07-16

degrade substantially when faced with the diverse conditions of real-world deployment. The key challenge is that a single, fixed configuration is rarely optimal across different sensor types, environmental conditions, or mission priorities. This reveals a critical research gap: the lack of a systematic, empirical framework for configuring and optimizing an existing detector to maximize its effectiveness under specific operational constraints.

To address these limitations, we move beyond proposing new architecture and instead conduct a rigorous empirical analysis of the YOLOv8n model, an architecture explicitly designed for high-efficiency edge deployment. Our approach is to deconstruct the model's performance by systematically investigating the hierarchical impact of the key factors that govern its behavior from training to inference. We hypothesize that by isolating and quantifying the effects of data augmentation, input resolution, and post-processing parameters, we can establish a data-driven methodology for practical, mission-specific optimization. This investigation is performed on two datasets with contrasting native resolutions to ensure our findings are generalizable across a spectrum of sensor qualities.

The primary objective of this study is to establish a clear, hierarchical framework for optimizing a lightweight object detector for aerial search missions. Specifically, our research contributions are: to quantify the foundational impact of specialized data augmentation strategies on model generalization; to demonstrate the existence of an "optimal operational window" for input processing resolution and identify the performance collapse caused by the scale-shift phenomenon; and to establish a methodology for tactical adaptation of the detector's operational mode using post-processing parameters. We systematically evaluate these factors using a comprehensive suite of metrics, including mean Average Precision (mAP), Precision, Recall, F1-score, and inference latency, providing a holistic view of the accuracy-efficiency trade-off. This work provides a practical and reproducible framework for maximizing the operational effectiveness of object detection on UAVs.

2. Literature review

Vision-based perception systems are fundamental for the autonomy and operational effectiveness of UAVs across a spectrum of critical applications, including infrastructure inspection, precision agriculture, and disaster response [1]. In scenarios like search and rescue, the ability to perform real-time object detection from an aerial perspective is a core mission requirement, offering a significant advantage over traditional methods [2]. These applications may involve individual UAVs or coordinated multi-UAV systems [3], often referred to as swarms, where real-time detection and communication among team members enable collaborative mission execution [4]. Regardless of the deployment architecture, all UAV systems demand robust object detection from onboard cameras, requiring detectors to operate under strict computational constraints. Aerial systems encounter unique challenges rarely addressed in standard detection benchmarks, including significant variations in image resolution, lighting conditions, and sensor quality, which can degrade performance [5]. These practical constraints highlight the need for a systematic understanding of detection performance under varying deployment conditions.

The field of object detection has evolved from computationally intensive two-stage methods to highly efficient single-stage architectures. The YOLO (You Only Look Once) family of detectors, in particular, has become the de facto standard for real-time applications by reformulating detection as a single regression problem [6]. Subsequent iterations have progressively improved the trade-off between accuracy, inference speed, and model size through architectural innovations such as advanced backbones and efficient feature fusion necks. These design choices, along with the availability of scaled variants, make the YOLO framework particularly well-suited for deployment on embedded systems and UAVs where computational resources are a primary constraint. Architectural modifications and optimizations, such as those proposed for YOLOv8 [7], continue to push the performance envelope, especially for challenging tasks like small object detection in aerial views. Aerial imagery presents distinctive challenges for object detection systems, as summarized in table below (Table 1).

Table 1
Summary of Key Challenges in UAV-based Object Detection

Challenge	Core Problem Description	Primary Mitigation Approaches
Small Object Size	Insufficient pixel data for feature extraction	Increased input resolution; multi-scale feature fusion (e.g., FPNs)
Occlusion	Partial or complete object obstruction	Contextual inference models; simulated occlusion augmentation
Scale Variations	High variance in intra-class object scale	Multi-scale feature pyramids; scale jitter augmentation
Class Imbalance	Skewed class distribution in training data	Data resampling techniques; weighted loss functions (e.g., Focal Loss)
Perspective Changes	Severe appearance distortion from variable viewing angles	Viewpoint-diverse training; perspective transformation augmentation
Dense Object Groups	Clustered instances leading to merged detections	Advanced NMS algorithms; instance segmentation methods
Illumination Variability	Feature degradation from non-uniform lighting (shadows, glare)	Photometric normalization; brightness/contrast augmentation
Adverse Weather	Signal degradation from atmospheric obscurants (fog, rain)	Multi-modal sensor fusion (e.g., thermal); image restoration pre-processing

The significant variation in image resolution depending on flight altitude and camera specifications creates a complex performance landscape [8]. UAV systems may carry sensors ranging from legacy low-resolution cameras to modern high-resolution systems, a factor that becomes particularly important in coordinated operations where heterogeneous platforms might need to share detection information. Additionally, aerial platforms impose strict constraints on computational resources, with limited power budgets and onboard processing capacity.

While this work focuses on single-agent perception, it is important to situate it within the broader context of multi-UAV systems [9]. Such systems employ various coordination strategies - ranging from centralized command structures to fully decentralized [10] and distributed decision-making - to execute complex tasks efficiently [11]. These strategies rely on robust communication and, in some cases, advanced optimization approaches like learning-based methods to govern collective behavior [12]. However, the efficacy of any coordination algorithm is fundamentally bottlenecked by the quality and efficiency of the foundational perception capabilities of each individual UAV. Whether coordination employs reinforcement learning PPO [13], DQN [14], DDPG [15]or traditional planning algorithms [16], the ability of each agent to reliably and rapidly detect objects of interest directly impacts overall team performance [17]. This study addresses the foundational challenge of optimizing this individual UAV detection performance, which forms the basis for any coordinated multi-agent system.

Beyond model architecture, the practical robustness of a deployed detector is highly sensitive to its inference-time hyperparameters, primarily the confidence threshold and the Non-Maximum Suppression (NMS) IoU threshold. These parameters directly control the operational trade-off between missing true positives (low Recall) and generating false alarms (low Precision). This sensitivity becomes particularly critical in multi-UAV systems where consistent detection behavior across the fleet is essential for reliable data fusion and collaborative task execution. While some studies have explored robustness through techniques like ensemble learning [18], a systematic analysis of hyperparameter sensitivity, especially in conjunction with varying input resolutions, remains an understudied area. Understanding this sensitivity is critical for practitioners deploying detection systems in resource-constrained environments that demand both reliability and efficiency. The paper [19] discusses hardware and information support for automated electromagnetic monitoring of geodynamic objects, enabling continuous data collection and transmission for early diagnostics.

The existing literature has firmly established that YOLO-family detectors are highly effective for real-time object detection. However, what remains largely unaddressed is a systematic methodology for their practical optimization in resource-constrained aerial applications [20]. Specifically, a critical research gap exists in the quantitative understanding of how detection performance is jointly affected by training strategies, input resolution, and post-processing parameters. Practitioners deploying UAV systems currently lack data-driven guidelines for answering key questions: What is the optimal input resolution for a given sensor? How does this choice affect the accuracy-latency trade-off? How should post-processing be configured to align with specific mission goals? The challenges of curating and annotating diverse aerial datasets further compound this issue.

This study aims to fill this gap by conducting a rigorous and systematic empirical analysis of the YOLOv8n detector. Our primary contributions are: A quantitative demonstration of the foundational importance of specialized data augmentation for achieving robust baseline performance, a key strategy for overcoming limited data diversity. A comprehensive evaluation of detection performance across a wide spectrum of input resolutions, on two dataset variants simulating different sensor qualities, leading to the identification of an "optimal operational window" for resolution. A sensitivity analysis of key post-processing parameters, providing a clear methodology for tactical, mission-specific adaptation of the detector. Collectively, our findings provide a data-driven, hierarchical framework for optimizing detection systems, moving beyond anecdotal tuning to a principled approach for configuring detectors on UAV platforms, a task that requires robust information and hardware support to be successful in the field.

3. Materials and methods

3.1. Model description and evaluation metrics

To identify a suitable baseline model for in depth analysis, one that embodies an optimal trade-off between detection accuracy and computational efficiency for resource constrained UAV platforms, a preliminary comparative evaluation of three prominent architectures was performed. The candidates included three single-stage models from the YOLOv8 family: YOLOv8n, YOLOv8m, YOLOv8x.

The YOLOv8 model family was selected for particular focus as its variants ('n' for nano, 'm' for medium, and 'x' for extra-large) are explicitly designed to represent different points on the accuracy-efficiency spectrum. Architecturally, these are single-stage detectors that have been optimized for speed and performance. Their design incorporates a CSPDarknet-derived backbone for efficient feature extraction, a Path Aggregation Network (PAN) neck for robust multi-scale feature fusion critical for detecting objects of varying sizes in aerial views, and an anchor-free, decoupled head for streamlined bounding box regression and object classification as shown in the figure below (Figure 1).

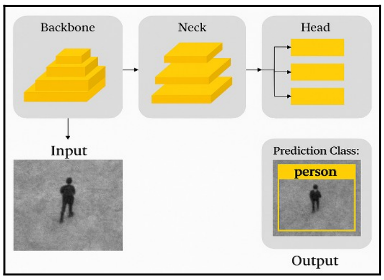

Figure 1: YOLO architecture design.

Our evaluation protocol employs a comprehensive suite of metrics: Recall to prioritize survivor detection, Precision to ensure operational efficiency, F1-Score to measure their balance, and mAP50/mAP50-95 for a robust assessment of overall detection and localization performance, all deliberately chosen to align with the critical and distinct priorities of Search and Rescue missions. To quantitatively assess the performance of the object detection models, we utilize the mean Average Precision (mAP) metric, specifically at an Intersection over Union (IoU) threshold of 0.5 (denoted as mAP50 or mAP@0.5) and averaged over IoU thresholds from 0.5 to 0.95 in steps of 0.05 (denoted as mAP50-95 or mAP@[.5:.95]). These metrics are standard benchmarks in object detection tasks.

The IoU is the fundamental measure for evaluating the localization accuracy of a predicted bounding box. It quantifies the overlap between the predicted bounding box (B_p) and the ground truth bounding box (B_{gt}):

$$IoU=\frac{Area\left(B_p \bigcap B_{gt}\right)}{Area\left(B_p \bigcup B_{gt}\right)} \tag{1}$$

Detections with $IoU \geq$ threshold are True Positives (TP); others may be False Positives (FP). Undetected ground truths are False Negatives (FN). From these, Precision (P) and Recall (R)

$$P=\frac{TP}{TP+FP} \tag{2}$$

$$R=\frac{TP}{TP+FN} \tag{3}$$

The mAP is the mean of the AP values calculated for each individual class (k) over a total of N classes:

$$mAP=\frac{1}{N}\sum_{k=1}^{N} AP_k \tag{4}$$

mAP50 evaluates general detection performance with a lenient localization requirement ($IoU \geq 0.50$). In contrast, mAP50-95, a more stringent metric, averages APs calculated at ten IoU thresholds (0.50 to 0.95), thus heavily penalizing imprecise localization and reflecting both classification and localization proficiency.

3.2. Dataset description

The model was trained on a composite dataset created by combining two sources: the SARD (2021) [21] dataset, which provides imagery specific to search and rescue scenarios, and the MIXED (2024) [22] dataset, which adds greater environmental and perspectival diversity. This combined training corpus, totaling 5981 images, was divided into subsets for training (4201 images), validation (1378 images), and testing (402 images).

To assess the model's generalization capabilities, we employed the HERIDAL (2022) benchmark for all final evaluations. This dataset is specifically designed for aerial search and rescue, consisting of 1546 test images that feature human subjects in a wide range of poses, scales, and perspectives. Crucially, HERIDAL incorporates many of the real-world challenges central to our study such as small object sizes, partial occlusions, and variable lighting making it an ideal testbed for assessing model robustness.

To conduct a more thorough analysis of how performance is affected by sensor quality, we utilized two distinct versions of this dataset. The first is a low-resolution (416x416) version, representative of standard UAV video streams or legacy hardware. The second is a high-resolution (4000x3000) version, simulating data from modern, high-fidelity sensors. This dual resolution evaluation strategy allows us to rigorously assess not only the model's baseline performance but also its scalability and robustness to drastically different data characteristics.

Our training methodology was designed to maximize model performance and ensure the reproducibility of our results, anchored by two core strategies: transfer learning and extensive data augmentation. To leverage knowledge from a large-scale visual domain and accelerate convergence on our specialized dataset, we employed a transfer learning strategy. All models were initialized with weights pre-trained on the COCO dataset. This approach allows the model to utilize powerful, generalized feature representations learned from millions of images, which significantly enhances its ability to generalize from our more limited training corpus.

3.3. Training configuration settings

All models were fine-tuned for a maximum of 150 epochs using an early stopping patience of 40. The training was performed on two NVIDIA T4 GPUs with a batch size of 32, an input resolution of 640x640, and was accelerated using Automatic Mixed Precision (AMP). We employed the AdamW optimizer with a cosine annealing learning rate schedule, configured with an initial learning rate of 0.0015, a weight decay of 0.0005, and a 3-epoch warmup period for stability. Key regularization techniques included label smoothing (factor=0.05). Furthermore, our protocol included a final training phase where mosaic augmentation was disabled for the last 10 epochs to allow the model to fine-tune on standard image compositions. An example of a training patch with the augmentation used is shown in figure below (Figure 2).

Figure 2: Sample training batch demonstrating the applied data augmentations.

Table 2
Hyperparameters of data augmentations

Parameter	Proposed Augmentation	Strong Augmentation	Standard Augmentation	No Augmentation
Color augmentations				
hsv_h (shade)	0.02	0.03	0.0	0.0
hsv_s (saturation)	0.4	0.9	0.0	0.0
hsv_v (brightness)	0.3	0.6	0.0	0.0
Geometric augmentations				
degrees	10.0	10.0	0.0	0.0
translate	0.2	0.2	0.0	0.0
scale	0.3	0.6	0.0	0.0
shear	2.0	2.0	0.0	0.0
perspective	0.0003	0.0005	0.0	0.0
flipud	0.3	0.1	0.0	0.0
fliplr	0.5	0.5	0.0	0.0
Combined augmentations				
mosaic	0.9	1.0	1.0	0.0
mixup	0.1	0.3	0.0	0.0
cutmix	-	0.2	0.0	0.0
erasing	-	0.4	0.0	0.0
copy_paste	0.3	-	-	-
auto_augment	-	randaugment	-	-

To evaluate robustness on aerial imagery, we compared four augmentation regimes: a balanced "Proposed Augmentation" (moderate geometric and color transforms plus Mosaic, MixUp, Copy-

Paste), a Strong setup (aggressive transforms including CutMix, Erasing, RandAugment), and Standard / No Augmentation baselines; this controlled comparison measures how augmentation intensity affects detection stability and overall performance. Detailed parameters are given in table below (Table 2).

4. Results and discussion

4.1. Comparative analysis and baseline model selection

This section presents the empirical evaluation of the selected object detection models and a systematic deconstruction of their performance through a series of ablation studies. The primary objective was to identify architecture offering an optimal balance between detection accuracy and computational efficiency for deployment on resource-constrained UAV platforms. All experiments were conducted on an NVIDIA T4 GPU; while absolute latency values are hardware-dependent, the relative performance differences provide a robust and generalizable basis for model comparison.

The initial phase of our investigation focused on selecting a baseline model that embodies the optimal trade-off between detection accuracy and computational efficiency for deployment on resource-constrained UAV platforms. This selection process was predicated on two key findings.

First, we established that domain-specific adaptation via fine-tuning is not merely beneficial but an absolute prerequisite for effective performance. The dramatic performance delta between baseline models pre-trained on the general-purpose COCO dataset and models fine-tuned on our specialized SARD+MIXED corpus is quantified in table below (Table 3).

Table 3
The Impact of Domain-Specific Fine-Tuning on Model Performance (mAP@0.5)

Detector	Baseline mAP@50 (%) (COCO Pre-trained)	Fine-Tuned mAP@50 (%) (SARD+MIXED)	Absolute Gain (p.p.)	Relative Increase (Fold)
YOLOv8n	2.8	62.9	+60.1	22.6x
YOLOv8m	8.6	69.9	+61.3	8.1x
YOLOv8x	6.8	68.9	+62.1	10.1x

As the results show, all models experienced a monumental improvement after fine-tuning. The YOLOv8n model, for instance, saw its mAP@0.5 score increase from a near-unusable 2.8% to a robust 62.9%, a relative increase of over 22-fold. This result unequivocally demonstrates that only fine-tuned models are viable for this application, rendering a direct comparison of the baseline models irrelevant for selecting a deployment candidate.

With the necessity of fine-tuning established, we then conducted a comparative analysis of the three fine-tuned YOLOv8 variants to quantify their respective trade-offs between accuracy and computational cost, with results summarized in table below (Table 4).

Table 4
Comparative Performance of Fine-Tuned YOLOv8 Variants on the HERIDAL Dataset (416x416)

Detector	mAP50 (%)	mAP50-95 (%)	Precision (%)	Recall (%)	F1-score (%)	Latency (ms)	Size (MB)
YOLOv8n	62.9	28.8	81.4	54.4	65.2	5.2	5.97
YOLOv8m	69.9	34.1	82.5	61.2	70.3	26.6	49.6
YOLOv8x	68.9	34.4	84.2	60.8	70.6	76.6	130

The data in Table 4 clearly illustrates the inherent compromise between model complexity and operational efficiency. While the YOLOv8m model achieves the highest mAP50 of 69.9%, this marginal accuracy gain comes at the cost of a 5-fold increase in latency and an 8-fold increase in model size compared to YOLOv8n. The YOLOv8n model, conversely, offers a dramatically superior performance-per-resource ratio. Its ability to achieve substantial accuracy (62.9% mAP50) at a very low latency (5.2 ms) and with a minimal memory footprint (5.97 MB) makes it the most suitable candidate for real-time object detection goals.

Therefore, based on this analysis, YOLOv8n was selected as the baseline model for all subsequent in-depth ablation studies. Its unparalleled balance of speed, efficiency, and robust accuracy after fine-tuning establishes it as the most pragmatically viable candidate for real-world UAV deployment.

To move beyond surface-level comparisons and develop a deep understanding of the model's behavior, we conducted a series of ablation studies. These experiments were designed to isolate and quantify the impact of three hierarchical factors: 1) the foundational data augmentation strategy, 2) the critical input processing resolution, and 3) the tactical post-processing parameters. By performing this analysis on Heridal dataset with contrasting native resolutions (Low-Res: 416x416, High-Res: 4000x3000), we aimed to derive generalizable principles for model optimization.

4.2. The foundational impact of data augmentation on generalization

The initial investigation aimed to establish a robust baseline model by assessing the impact of different data augmentation strategies during training. Four models were compared: one trained with our proposed augmentation pipeline tailored for aerial imagery (mine-aug), one with the library's standard pipeline (default-aug), one with no augmentation (no-aug), and one with only noise augmentation (noise-aug) as shown in table below (Table 5).

Table 5
The effect of augmentation strategy on the performance of YOLOv8n

Detector	Dataset Resolution	*mAP50 (%)*	*mAP50-95 (%)*	*Precision (%)*	*Recall (%)*	*F1-score (%)*
YOLOv8n-mine-aug	416×416	62.91	28.77	81.4	54.4	65.2
YOLOv8n-default-aug		47.66	17.70	62.8	43.1	51.1
YOLOv8n-no-aug		50.32	22.00	63.3	44.8	52.5
YOLOv8n-noise-aug		37.21	11.59	53.44	36.27	43.21
YOLOv8n-mine-aug	4000×3000	68.24	34.28	78.1	60.7	68.3
YOLOv8n-default-aug		49.26	17.75	59.1	47.6	52.7
YOLOv8n-no-aug		42.16	15.66	55.4	41.1	47.2
YOLOv8n-noise-aug		46.36	16.53	57.27	46.61	51.39

The results unequivocally demonstrate that a specialized, task-aware augmentation strategy is the single most critical factor for achieving high performance. On the high-resolution dataset, our proposed pipeline (mine-aug) surpasses the default strategy by a staggering 19 percentage points in mAP50. This substantial margin confirms that generic augmentation techniques are insufficient to prepare the model for the unique challenges of aerial imagery, such as extreme variations in object scale, viewing angles, and lighting conditions. The model trained without augmentation expectedly suffers from poor generalization. Consequently, the YOLOv8n-mine-aug model was established as the superior baseline for all subsequent experiments.

4.3. Input resolution is the dominant factor defining the performance window

The most critical parameter affecting both accuracy and speed is the input processing resolution. Figure below (Figure 3) illustrates the performance metrics as a function of input resolution for both low- and high-resolution datasets.

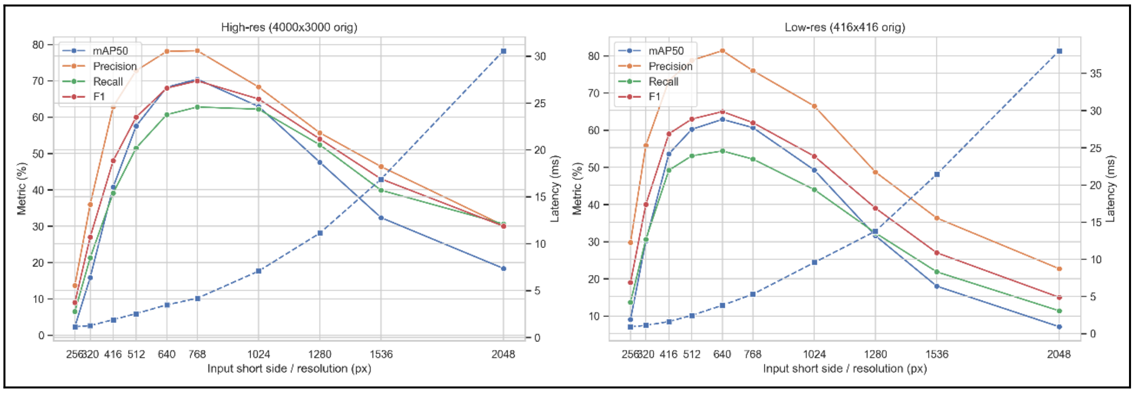

Figure 3: Detection metrics vs. input resolution for high- and low-resolution datasets.

As the plots unequivocally demonstrate, there is no universally optimal resolution. Instead, an "optimal operational window" exists, which is highly dependent on the source data's native resolution. For the high-resolution dataset (left panel), the peak shifts upward to 768x768 pixels, as a higher resolution is required to preserve details. Conversely, for the low-resolution dataset (right panel), performance peaks at 640x640 pixels, beyond which accuracy begins to decline due to upscaling artifacts. In both cases, extreme increases in resolution lead to a performance collapse, a phenomenon attributable to the scale shift effect, where object scales at inference deviate too far from those seen during training.

This fundamental trade-off between accuracy and latency is further visualized in figure below (Figure 4). Bubble size denotes F1-score, and color denotes the dataset.

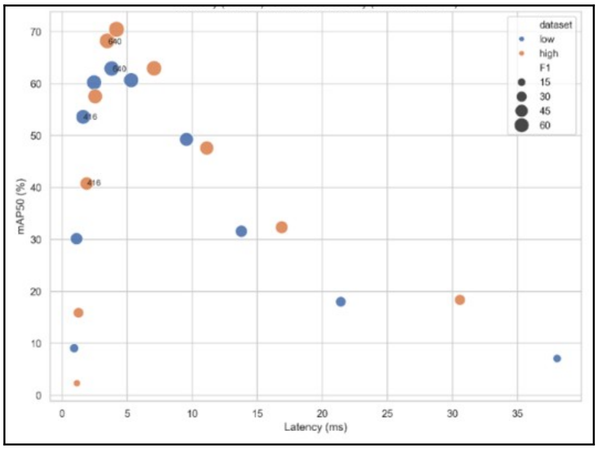

Figure 4: Trade-off between detection accuracy (mAP50) and inference latency.

The scatter plot clearly identifies the most efficient configurations. The points corresponding to 640px (low-res) and 768px (high-res) occupy the desirable top-left quadrant, signifying high accuracy at low latency. As resolution increases, the points move to the right (higher latency) and eventually down (lower accuracy), making them operationally inefficient.

4.4. Post-processing parameters tools for tactical adaptation and robustness confirmation

With optimal resolutions identified, we investigated the parameters that fine-tune the model's output at the inference stage. Figure below (Figure 5) shows how the confidence threshold acts as a regulator for the Precision-Recall trade-off. This illustrates the selection of operational thresholds for different mission priorities.

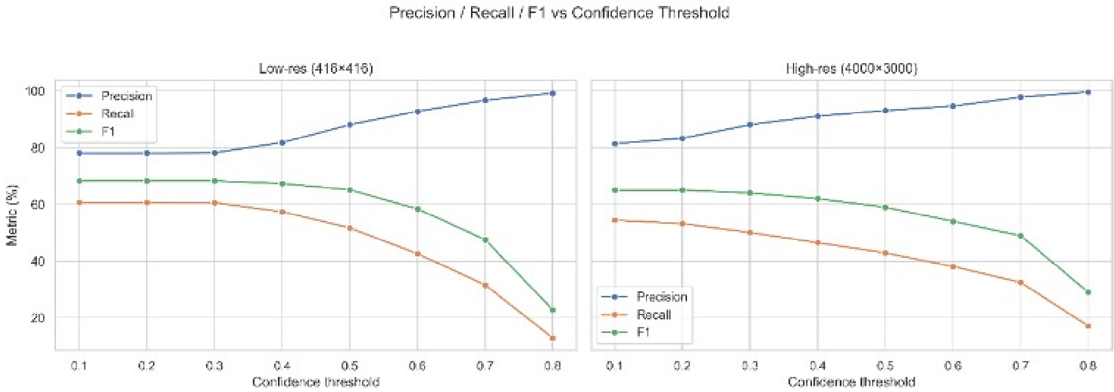

Figure 5: Precision, Recall, and F1-score as functions of the confidence threshold for both datasets.

The plots show that the confidence threshold (conf) is a powerful tool for tactical adaptation. For search-and-rescue missions where maximizing Recall is paramount, a lower threshold (e.g., 0.1-0.2) is optimal. For tasks like asset counting where Precision is critical, a higher threshold (e.g., >0.5) is preferable.

Finally, we assessed the model's sensitivity to the NMS IoU threshold, a parameter that can be difficult to tune as shown in figure below (Figure 6).

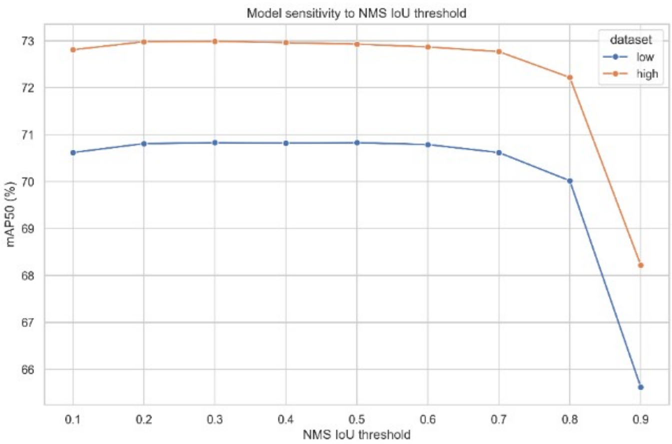

Figure 6: Sensitivity of detection performance (mAP50) to the NMS IoU threshold for low- and high-resolution datasets.

As Figure 6 illustrates, the model exhibits exceptional robustness to this parameter. Performance remains on a stable plateau across a vast range of IoU values (0.1 to 0.6) for both datasets. This is a highly desirable characteristic, as it indicates high-quality localization and eliminates the need for meticulous iou tuning during deployment.

4.5. Synthesis of findings and optimal configurations

To provide a holistic, at-a-glance comparison of all tested configurations and synthesize the findings from our multi-faceted analysis, the normalized performance metrics are visualized as a heatmap in figure below (Figure 7).

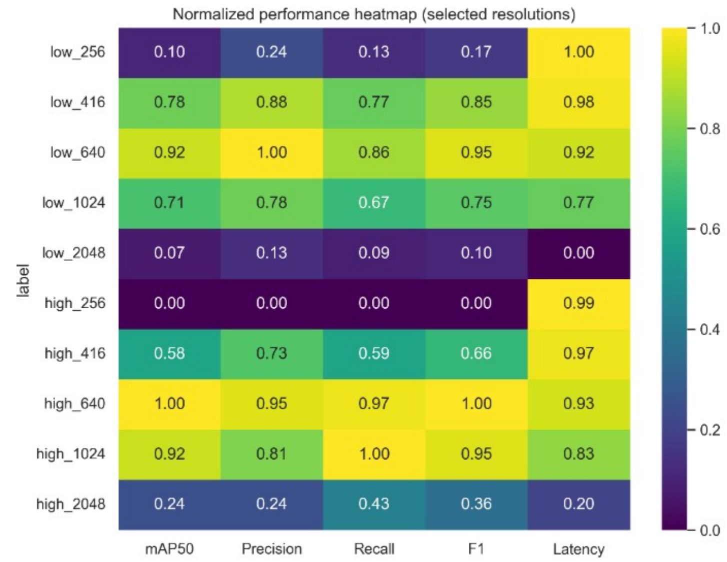

Figure 7: Normalized performance heatmap summarizing key detection metrics across representative resolutions.

In this heatmap, each metric column is independently normalized from 0 (worst performance in that column, dark purple) to 1 (best performance, bright yellow). This visualization allows for the direct comparison of each configuration's relative strengths and weaknesses. The heatmap provides a clear visual confirmation of the "optimal operational windows" identified in Figure 3. The rows corresponding to low640 and high640/high1024 exhibit the brightest overall profiles across the accuracy metrics (mAP50, Precision, Recall, F1). Conversely, configurations at the resolution extremes (low256, low2048, high2048) are dominated by dark colors, indicating suboptimal performance.

Crucially, the Latency column must be interpreted in reverse, where brighter colors signify inferior (higher) latency. The heatmap visually captures the core engineering trade-off: the configuration low640 presents a remarkably balanced profile with strong accuracy and good latency (bright accuracy cells, moderately bright latency cell), whereas high1024 achieves excellent Recall at the cost of significantly worse latency.

While the heatmap provides a powerful visual summary, a selection of key configurations with their absolute performance values is presented in table below (Table 6) to ground the analysis in concrete data.

Table 6
Summary of Representative Configurations and Performance Metrics

Dataset	Resolution	mAP50 (%)	Precision (%)	Recall (%)	F1-score (%)	Latency (ms)	Note
Low-Res	416x416	53.6	73.3	49.2	59.0	1.62	Native Resolution
	640x640	62.9	81.4	54.4	65.0	3.81	Optimal Performance
	1024x1024	49.3	66.5	44.0	53.0	9.56	Performance Degradation
High-Res	640x640	68.2	78.1	60.7	68.0	3.47	Training Resolution
	768x768	70.5	78.3	62.8	70.0	4.18	Optimal Performance
	1280x1280	47.6	55.7	52.4	54.0	11.12	Performance Degradation

Ablation studies show that the maximum effectiveness of the YOLOv8n detector is achieved through a hierarchical approach based on specialized data augmentation during training. Figure below (Figure 8) shows YOLOv8n model predictions on high-res Heridal dataset with optimized parameters: resoution=768, conf=0.2, iou=0.5.

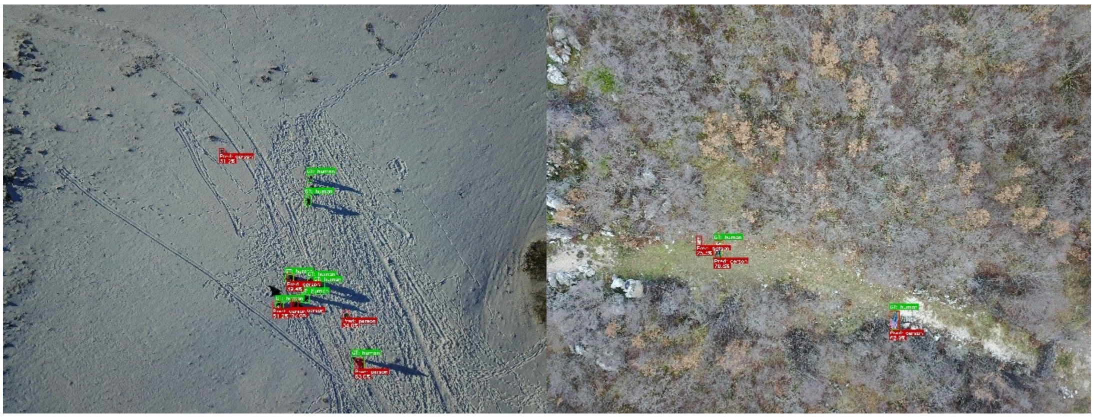

Figure 8: YOLOv8n model prediction results (red) with ground truth annotations (green).

For search and rescue operations, the optimal configuration is to use a processing resolution of 640x640 or 768x768 pixels, depending on the native resolution of the drone camera, in order to maximize the quality of the extracted features. At the inference stage, a confidence threshold in the range of 0.15-0.25 should be set, which ensures the highest possible detection completeness (Recall) while maintaining high overall accuracy. The model's stability to the IoU NMS parameter allows the use of a standard value of ~0.5, which greatly simplifies the practical deployment of the system in real conditions.

5. Conclusion

This study addressed the critical challenge of optimizing lightweight object detectors for real-time aerial search missions, focusing on the practical deployment constraints of UAV platforms. The primary objective was to move beyond architectural comparisons and conduct a rigorous empirical analysis to establish a systematic framework for maximizing the performance of a state-of-the-art model, YOLOv8n, through meticulous hyperparameter tuning.

Our investigation yielded a hierarchical framework for optimization, demonstrating that performance is governed by a sequence of interdependent factors. First, we empirically proved that a specialized, task-aware data augmentation strategy is the foundational prerequisite for achieving high performance, yielding a substantial 15-19 percentage point uplift in mAP50 compared to standard augmentation pipelines. Second, our analysis of input resolution revealed the existence of an "optimal operational window" that is critically dependent on the source data's native resolution. For low-resolution (416x416) data, peak performance (mAP50 of 62.9%) was achieved at 640x640, whereas for high-resolution (4000x3000) data, the optimum shifted to 768x768 (mAP50 of 70.5%). This finding quantitatively demonstrates the performance collapse caused by the scale shift phenomenon when processing resolution deviates too far from the model's training parameters. Finally, we established that post-processing parameters serve as tactical adaptation tools: the confidence threshold was identified as a flexible regulator for the Precision-Recall trade-off, with an optimal value for balanced performance found at ~0.2, while the NMS IoU threshold was shown to be non-critical due to the model's high localization accuracy.

The practical significance of these findings lies in providing a clear, data-driven methodology for practitioners to configure perception systems on UAVs for specific missions. For search and rescue operations, where maximizing Recall is paramount, our results recommend using the YOLOv8n model with a specialized augmentation pipeline, an input resolution of 640x640 or 768x768 (depending on camera fidelity), and a confidence threshold of ~0.2. This configuration ensures the highest probability of detecting targets while maintaining real-time latency of under 5 ms on a standard GPU. These guidelines can directly inform the development of more reliable and efficient systems for emergency services, environmental monitoring, and infrastructure inspection.

Several limitations of this study should be acknowledged. All experiments were conducted on a single GPU architecture (NVIDIA T4), and while the relative performance trends are expected to hold, absolute latency figures will vary on different edge computing hardware. The study was also confined to a single object class (persons), and the optimal hyperparameters may require recalibration for other object types. Furthermore, the analysis was performed on pre-recorded datasets and does not account for real-world complexities such as motion blur, communication delays, or dynamic environmental changes during a live mission.

Building upon these limitations, future research should proceed in several key directions. First, the proposed optimal configurations should be validated on a diverse range of embedded hardware platforms (e.g., Radxa, Rasspberry Pi) to create a more comprehensive performance profile. Second, the framework should be extended to multi-class detection scenarios to assess its generalizability. The third and most important direction is the integration of the optimized detector into a high-precision drone simulation environment. This will enable a quantitative assessment of how component-level improvements in perception lead to improved mission performance at the system level, such as detection time and area coverage, thereby bridging the gap between detector optimization and drone search research. Finally, studying the impact of model optimization techniques, such as quantization and pruning, on the identified optimal configurations will be a valuable next step toward further improving deployment efficiency.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Mishra B., Garg D., Narang P. & Mishra V. (2020). Drone-surveillance for search and rescue in natural disaster. Computer Communications. 156, 1–10. https://doi.org/10.1016/j.comcom.2020.03.012.

[2] Mittal P., Singh R. & Sharma A. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. Image and Vision Computing. 104, 104046. https://doi.org/10.1016/j.imavis.2020.104046.

[3] Azar A. T., Koubaa A., Ali Mohamed N., Ibrahim H. A., Ibrahim Z. F., Kazim M., ... Casalino G. (2021). Drone Deep Reinforcement Learning: A Review. Electronics. 10(9), 999. https://doi.org/10.3390/electronics10090999.

[4] Dong J., Ota K. & Dong M. (2021). UAV-Based Real-Time Survivor Detection System in Post-Disaster Search and Rescue Operations. IEEE Journal on Miniaturization for Air and Space Systems. 2(4), 209–219. https://doi.org/10.1109/JMASS.2021.3083659.

[5] Alvey B., Anderson D. T., Buck A., Deardorff M., Scott G. & Keller J. M. (2021). Simulated Photorealistic Deep Learning Framework and Workflows to Accelerate Computer Vision and Unmanned Aerial Vehicle Research. ICCVW 2021. 3882–3891. https://doi.org/10.1109/ICCVW54120.2021.00435.

[6] Redmon J., Divvala S., Girshick R. & Farhadi A. (2016). You Only Look Once: Unified, Real-Time Object Detection. CVPR 2016. 779–788. https://doi.org/10.1109/CVPR.2016.91.

[7] Li Y., Fan Q., Huang H., Han Z. & Gu Q. (2023). A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. Drones. 7(5), 304. https://doi.org/10.3390/drones7050304.

[8] Wang H. & Gao P. (2024). Survey of small object detection methods based on deep learning. ICIIBMS 2024. 9, 221–224. https://doi.org/10.1109/ICIIBMS62405.2024.10792837.

[9] Arnold R., Jablonski J., Abruzzo B. & Mezzacappa E. (2020). Heterogeneous UAV Multi-Role Swarming Behaviors for Search and Rescue. CogSIMA 2020. 122–128. https://doi.org/10.1109/CogSIMA49017.2020.9215994.

[10] Balázs B., Vicsek T., Somorjai G., Nepusz T. & Vásárhelyi G. (2024). Decentralized traffic management of autonomous drones. Swarm Intelligence. https://doi.org/10.1007/s11721-024-00241-y.

[11] Bu Y., Yan Y. & Yang Y. (2024). Advancement Challenges in UAV Swarm Formation Control: A Comprehensive Review. Drones. 8(7), 320. https://doi.org/10.3390/drones8070320.

[12] Jamshidpey A., Wahby M., Heinrich M. K., Allwright M., Zhu W. & Dorigo M. (2024, August 13). Centralization vs. decentralization in multi-robot coverage: Ground robots under UAV supervision. arXiv. https://doi.org/10.48550/arXiv.2408.06553.

[13] Zhu W., Fang W. & Su Y. (2024). Path Planning for Multi-UAV Based on Improved Proximal Policy Optimization Algorithm. YAC 2024. 1890–1894. https://doi.org/10.1109/YAC63405.2024.10598516.

[14] Xu C., Zhang K. & Song H. (2020). UAV Swarm Communication Aware Formation Control via Deep Q Network. IPCCC 2020. 1–2. https://doi.org/10.1109/IPCCC50635.2020.9391509.

[15] Wu R., Gu F. & Huang J. (2020). A multi-critic deep deterministic policy gradient UAV path planning. CIS 2020. 6–10. https://doi.org/10.1109/CIS52066.2020.00010.

[16] Wang X., Wang S., Liang X., Zhao D., Huang J., Xu X., Dai B. & Miao Q. (2024). Deep reinforcement learning: A survey. IEEE Transactions on Neural Networks and Learning Systems. 35(4), 5064–5078. https://doi.org/10.1109/TNNLS.2022.3207346.

[17] Xia Z., Du J., Wang J., Jiang C., Ren Y., Li G. & Han Z. (2022). Multi-Agent Reinforcement Learning Aided Intelligent UAV Swarm for Target Tracking. IEEE Transactions on Vehicular Technology. 71(1), 931–945. https://doi.org/10.1109/TVT.2021.3129504.

[18] Dousai N. M. K. & Lončarić S. (2022). Detecting humans in search and rescue operations based on ensemble learning. IEEE Access. 10, 26481–26492. https://doi.org/10.1109/ACCESS.2022.3156903.

[19] Bykov A., Kuzichkin O., Dorofeev N. & Koskin A. (2017). Information-hardware Support of Systems of the Automated Electromagnetic Monitoring of Geodynamic Objects. Procedia Computer Science. 103(C), 253–259. https://doi.org/10.1016/j.procs.2017.01.098.

[20] Kumar T., Brennan R., Mileo A. & Bendechache M. (2024). Image data augmentation approaches: A comprehensive survey and future directions. IEEE Access. 12, 187536–187571. https://doi.org/10.1109/ACCESS.2024.3470122.

[21] Sasa Sambolek & Ivasic-Kos M. (2021). SEARCH AND RESCUE IMAGE DATASET FOR PERSON DETECTION - SARD. IEEE Dataport. https://dx.doi.org/10.21227/ahxm-k331.

[22] FYP 2. (2024). Mixed Dataset - Aerial Person Dataset [Open Source Dataset]. Roboflow Universe. Roboflow. https://universe.roboflow.com/fyp-2-tmvit/mixed-dataset-aerial-person.