Research on Intelligent Recognition and Positioning Method of Civil Airport Bird Repellent Cannons Based on Improved YOLOV10

Zheng Wang*, Xinlei Zhang, Chou Hu
Qingdao Campus of Naval Aviation University, Qingdao, Shandong, China
**Corresponding Author*

Abstract: Addressing the challenges of bird control applications at civil airports—namely the difficulty of detecting distant small targets, strong background interference, and high real-time requirements—this paper proposes a YOLOv10n-attention-based recognition and localization method for bird control cannons. This approach utilizes YOLOv10n as its backbone. Without altering the pyramid and neck topology, it introduces global attention at the high-level branches to enhance response to weakly textured targets. The P5 branch employs a C2fCIB bottleneck structure to stabilize semantic representation across large receptive fields. The detection head retains the decoupled architecture with a DFL + IoU + CE loss combination, performing only "bird/non-bird" binary classification to meet engineering closed-loop requirements. Experiments on the self-built Airport-Birds dataset (airport surveillance frames with single-class annotations) demonstrate that our model achieves approximately mAP@0.50=0.83 at 640×640 input resolution with real-time performance around 30 FPS. This represents a stable improvement over the baseline without significantly increasing parameters or computational cost. Ablation studies further validate the synergistic contribution of attention mechanisms and the P5 bottleneck architecture to accuracy, recall, and localization robustness. By integrating recognition results with servo control mapping, the system completes the "detection-localization-lock-deterrence" engineering loop, meeting airport scenarios' demands for rapid response and high reliability. This work provides a reusable, deployable visual front-end solution for intelligent airport bird deterrence systems, laying the foundation for future multimodal fusion and edge deployment.

Keywords: Airport Bird Deterrence; Object Detection; YOLOv10; Attention Mechanism; Bird Deterrent Cannon

1. Introduction

Bird strikes pose a significant threat to global aviation safety, persistently disrupting civil aviation operations and airport functions [1]. According to International Civil Aviation Organization statistics, approximately 10,000 bird strike incidents occur worldwide annually, with over 90% occurring during takeoff, approach, and landing phases. These incidents cause severe damage to critical aircraft components such as engines, wings, windshields, radomes, and landing gear [2]. Traditional bird deterrence methods—such as manual patrols, bird-scaring vehicles, gas cannons, acoustic/visual deterrents, and repellent spraying—may yield short-term results. However, due to birds' strong learning abilities and high adaptability, their long-term effectiveness often diminishes rapidly [3][4]. Furthermore, manual bird control methods consume substantial human resources and have limited response capabilities, making it difficult to meet modern civil airports' demand for round-the-clock, intelligent bird control [5]. Therefore, there is an urgent need to explore integrated, automated intelligent bird control systems.

In the field of bird identification and monitoring, significant progress has been made in object detection technology research. Traditional two-stage methods (such as the R-CNN series) demonstrate excellent accuracy but suffer from high computational complexity, rendering them unsuitable for real-time applications [6]. Single-stage methods (such as SSD and YOLO series), however, have emerged as a current research hotspot for bird object identification due to their end-to-end detection workflow and high operational speed [7]. However, the small size,

high flight speeds, and diverse postures of birds in airport scenarios, coupled with complex lighting and weather conditions, make detection particularly challenging [8]. To enhance recognition performance, researchers are continuously exploring lightweight network architectures, small object detection augmentation, multi-scale feature fusion, and attention mechanisms to improve algorithm robustness in complex environments.

In recent years, improving the application of YOLO models in bird detection has garnered significant attention. The YOLOv8-Birds model proposed by Wang Qingyu et al. [9] effectively enhanced the recognition performance of small bird targets in Poyang Lake by introducing deformable convolutions (C2f_D3), GSConv, and BiFPN structures, combined with Slide Loss to optimize learning for challenging samples. Han Tao et al [10]. proposed the SCCW-YOLO and FCE-YOLO models based on YOLOv8. By utilizing coordinate attention mechanisms, deformable convolutions, and lightweight detection heads respectively, these models achieved a balance between detection accuracy and model parameter count. These improvements demonstrate the strong adaptability of YOLO models for small object recognition and real-time detection, providing a robust technical foundation for bird identification and tracking in airport bird control scenarios [11].

Existing airport bird control technologies primarily focus on the “deterrence” phase, while the ‘identification’ phase still lacks high-precision, real-time recognition and targeting capabilities for small birds. Building upon improvements to the YOLO model and integrating requirements for bird deterrent cannon applications, this study proposes an intelligent closed-loop method of “identification-localization-deterrence.” This approach aims to provide a feasible engineering pathway for constructing efficient and safe bird control systems at civil airports [12].

2. YOLOv10n-Attention Model

The improved YOLOv10n-attention object detection model is illustrated in Figure 1. Building upon the YOLOv10n baseline, Coordinate Attention (CA) is introduced to enhance the representation capability of shallow-to-mid-level features for small-scale and weakly textured bird targets. Simultaneously, C2fCIB is adopted as the structural bottleneck for high-level semantic reconstruction in the P5 branch, stabilizing the advanced semantic expression of distant targets. Beyond these two modifications, the pyramid levels (P3/P4/P5), neck fusion (FPN/PAN), and detection head/loss (decoupled head, DFL + IoU + CE) remain consistent with the baseline. This ensures performance differences can be attributed to the attention mechanism and the P5 structural bottleneck.

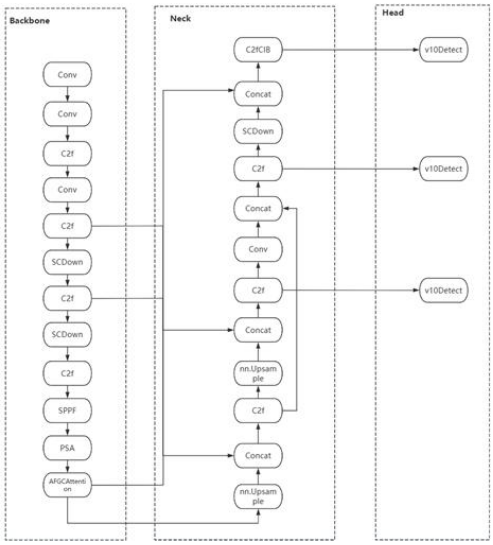

Figure 1. YOLOv10n-Attention Network Model Architecture Diagram

2.1 Backbone Network

While ensuring real-time processing, extract hierarchical features from low to high levels: shallow layers preserve edge/texture details, intermediate layers combine geometric and semantic information, and high-level layers enhance global semantic understanding and structural stability. This approach provides robust features for the P3/P4/P5 detection scales.

2.1.1 Staged structure and tensor scale

Taking an input size of 416×416 as an example, the stem network can be divided into five stages from bottom to top: Stem and S1–S5. Each stage performs spatial downsampling with a stride of =2, progressively enlarging the receptive field and compressing the resolution. Specifically: The image first undergoes initial feature extraction in Stem via a 3×3 convolution (stride 2), reducing resolution from 416 to 208; It then enters S1, where local edges and texture information are further encoded using C2f bottlenecks as the basic unit, followed by another 2-stride downsampling to 104. The S2 stage continues the C2f stacking, preserving strong details while introducing richer context,

resulting in a feature map size of 52×52. This scale corresponds to the P3 branch of the subsequent detection head, which is most sensitive to distant small objects. Next, the process enters S3, where the resolution changes to 26×26. The features contain sufficient boundary information and strong inter-class separation, corresponding to the P4 branch in detection. S4 further reduces resolution to 13×13, yielding the most refined semantic features and serving as the primary source for P5. Above this, S5 aggregates multi-scale context via SPPF and overlays high-level attention (PSA/AFGC) for channel-spatial re-calibration, ultimately producing high-level semantic features. In summary, the backbone outputs three feature streams to the neck: 52×52 (P3) from S2, 26×26 (P4) from S3, and 13×13 (P5) from S4/S5. These three streams jointly enter the FPN/PAN for bidirectional fusion before being fed to the decoupled detection head for classification and regression.

2.1.2 C2f and C2fCIB backbone unit

C2f (Cross-Stage partial with lighter bottlenecks) achieves this by proportionally splitting the input channels: one branch traverses multiple bottlenecks (Bottleneck: Conv-BN - SiLU residual), while the other branch takes a shortcut for direct connection. Finally, the channels are concatenated in the channel dimension and rectified. Its advantages include smooth gradient flow, high feature reuse, and computational efficiency [13]. Let the input tensor be X, split into X_a and X_b. After passing through m bottlenecks, they are aggregated with the direct-connection stream:

$$Y=\phi\left(Concat\left[X_a \quad ,\beta_m \quad \left(X_b \quad \right)\right]\right) \tag{1}$$

Among them, ϕ is a combination of 1×1/3×3 convolutional rectification.

C2fCIB (for structural bottlenecks in P5) replaces standard C2f with C2fCIB at the top layer (13×13). This feature enables stronger semantic reconstruction and channel interaction under large receptive fields, suppressing large-scale structural noise such as runway reflections and aircraft glare. Consequently, the contours and category semantics of distant/medium-to-large targets become more stable [14]. This unit is placed in the P5 branch without altering the pyramidal hierarchy or downstream pathways, solely enhancing high-level representations.

2.1.3 Attention mechanism and placement position

The main attention mechanism is employed to enhance response in the avian region and suppress complex backgrounds without modifying the pyramid/neck/detection head. The current implementation utilizes high-level global context attention (positioned after SPPF) and can be regarded as equivalent to coordinate attention at shallow-to-mid layers within the same "spatially aware channel relabeling" paradigm.

Global context attention enhances sensitivity to high-level semantic features—such as overall bird shape, orientation, and texture statistics—by re-calibrating features through a channel-space joint approach after SPPF aggregation. Its general form can be expressed as:

$$\widetilde{F}=\sigma(\vartheta(F))\odot F \tag{2}$$

where ϑ denotes the global context encoding (which may incorporate multi-head/frequency domain/pyramid pooling), and $\odot$ represents element-wise multiplication.

Regarding the mechanism of Coordinate Attention (CA), this paper presents its analytical expression (belonging to the "spatially aware channel attention" family alongside PSA/AFGC): Perform one-dimensional global aggregation along both the height (H) and width (W) dimensions of the input $X\epsilon R^{C\times H\times W}$:

$$z_{\mathrm{h}} \quad (c,i)=\frac{1}{W} \quad \sum_{j=1}^{W} X(c,i,j) \tag{3}$$

$$z_{w} \quad (c,j)=\frac{1}{H} \quad \sum_{i=1}^{H} X(c,i,j) \tag{4}$$

After concatenation, pass through a lightweight MLP and decompose into two weight maps s_{h} and s_{w}. This ultimately yields weight maps that simultaneously encode channel importance and directional positional information, completing the re-calibration.

2.1.4 Activation, normalization, and numerical stability

The main trunk layers default to a combination of SiLU activation and BatchNorm normalization: SiLU provides smoother gradients and more stable convergence under small-sample and strong data augmentation conditions [15], while BatchNorm performs channel-level normalization during training and is folded into convolutional weights during inference, thereby avoiding additional computational overhead [16]. The residual/shunt-based C2f and C2fCIB units effectively mitigate gradient vanishing and feature degradation during deep network training under this activation-normalization architecture, ensuring consistent numerical scales when

aggregating shallow textures and high-level semantics across layers. The SPPF at the trunk end compressively fuses local and global contexts through multi-receptive field parallel pooling, providing a stable statistical background for subsequent high-level attention (PSA/AFGC). This allows attention recalibration to better capture global cues of "overall bird shape and orientation" rather than amplifying local noise; This directly enhances feature discrimination and numerical robustness in airport scenarios characterized by strong reflections, thermal disturbances, and backlighting.

2.1.5 Interfaces and export for P3/P4/P5

The backbone outputs three feature streams to the neck, corresponding to the P3, P4, and P5 scales of the detection head: The 26×26 features from S3 strike a balance between boundary details and category semantics, making them suitable for medium-scale targets; while the 13×13 features derived from S4/S5 offer the most refined semantic information. Enhanced by SPPF and high-level attention mechanisms, they strengthen overall shape recognition and category discrimination capabilities, primarily serving close-range or relatively larger targets. Upon entering the FPN/PAN, the three feature streams undergo bidirectional alignment and fusion along predefined top-down and bottom-up paths before seamlessly connecting to the classification and regression branches of the decoupled detection heads. This interface design aligns with the YOLOv10n baseline, ensuring reusability across training and deployment pipelines while providing clear, stable tensor boundaries for multi-scale collaborative detection.

2.2 Neck Network

The neck network receives three feature outputs from the backbone (P3: 52×52, P4: 26×26, P5: 13×13), achieving bidirectional semantic and detail propagation through top-down and bottom-up pathways. The top-down path first upsamples the higher-level, semantically richer P5 features to match P4's spatial resolution, then concatenates them with P4 along the channel dimension. After rectification through a convolutional/residual unit, this yields mid-level features balancing semantic information and edge boundaries. The same process further upsamples these intermediate features to the scale of P3, merges them with shallow-level details, and rectifies the result to produce high-resolution features with enhanced detail while preserving category cues. Subsequently, the bottom-up path downsamples the fused P3 features back to P4 at a stride of 2. These features are then aggregated and rectified with P4's same-scale features in the channel dimension. This enables P4 to simultaneously incorporate edge/texture clues from shallow layers and category/structural information from higher layers. Similarly, the processed P4 is further downsampled and fed back to P5, where it is fused and rectified with the highest-level semantic features. This forms a more stable detection input for large-scale objects.

Ultimately, the neck outputs three sets of tensors to the detection head, maintaining spatial dimensions of 52×52, 26×26, and 13×13 respectively, with channel counts aligned to baseline settings. These tensors complement each other spatially and align in channel representation, providing robust, consistent input for subsequent decoupled detection heads. The entire fusion topology adheres to the standard FPN/PAN paradigm. Upsampling and downsampling employ conventional interpolation and strided convolution, while channel dimension aggregation is achieved through concatenation and convolutional rectification. Implementation and inference maintain identical structure and sequence to the baseline, facilitating engineering deployment and experimental reproducibility.

2.3 Detection Head and Loss

The detector employs a decoupled architecture for each scale feature, placing classification and regression tasks into two parallel convolutional branches to minimize task interference [17]. For any input feature map at any scale, the classification branch generates category confidence scores for each grid location through multiple layers of convolutions and activations (configurable as "bird/non-bird" or multi-class birds in airport scenarios). The output undergoes Sigmoid normalization, with each shape corresponding one-to-one to the spatial grid. The regression branch employs distributed boundary regression at its core, predicting discrete probability distributions for the four distances (left, top, right, bottom) of the target boundary relative to the grid center. The expected value of these distributions is used to recover continuous boundary distances. Combined with anchor-free

geometric mapping, this yields candidate bounding boxes in the image coordinate system.

The advantage of decoupling lies in allowing the classification path to focus more intently on class separability and confidence stability, while the regression path concentrates on geometric consistency and boundary smoothness. Both paths share inputs at the same scale during forward propagation but optimize independently during backpropagation, maintaining only a weak coupling through positive-negative sample allocation and weighted summation of losses [18]. The training phase retains the baseline matching and loss combination: classification uses cross-entropy (or its equivalent implementation) to measure the network's discrimination quality for birds; Regression employs a series of IoU losses to characterize the overlap and geometric relationship between predicted and ground-truth bounding boxes. Distribution Focus Loss (DFL) constrains the discrete distribution of four boundary distances to align with ground-truth distances, enabling continuous, smooth boundary regression at the pixel level.

During the inference stage, candidates are first filtered at each scale based on classification confidence thresholds. The expected solution from distribution regression is then decoded into bounding box coordinates. Non-maximum suppression is applied to multi-scale candidates to yield the final detection results. Specifically, P3 prioritizes recall of distant small objects, while P4 and P5 provide stable supplementary detection at medium and large scales respectively, ensuring the overall detection system maintains a favorable balance between speed and accuracy.

The positive and negative sample allocation during training follows task alignment and IoU-guided principles consistent with the baseline: the model generates candidates on grids at each scale, selecting the positive sample set based on geometric relationships between predictions and ground truth and center position constraints. This ensures that the grid truly bearing the supervision signal is concentrated within the target coverage area and its neighborhood. The loss function is weighted and summed at the branch level, with its overall form expressed as a combination of classification loss, IoU regression loss, and distribution focus loss. The classification term uses cross-entropy to measure the accuracy of category discrimination, emphasizing high-confidence identification of bird targets. The regression term uses IoU-based losses to constrain consistency between candidate and ground-truth bounding boxes in overlap area, center distance, and aspect ratio, ensuring robust boundary geometry. The distribution term supervises the discrete probability distribution of four boundary distances, minimizing deviation from ground-truth distances in the distribution space to achieve smoother, differentiable regression at the pixel scale. This synergistic matching and loss mechanism enables classification and regression to complement each other: the former provides reliable candidate filtering and confidence ranking, while the latter ensures boundary continuity and interpretability. When integrated with multi-scale FPN/PAN outputs, the network reliably recalls small targets and delineates large ones amid complex airport backgrounds, backlighting, and thermal noise—all without altering inference paths or tensor interfaces for seamless deployment in real-world systems.

3. Experimental Results and Analysis

3.1 Experimental Setup

The experimental environment configuration is shown in Table 1. The training parameters are set as follows: input resolution is fixed at 640×640, training is conducted for 300 epochs, and the batch size is set to 32. The optimizer uses SGD, and AMP mixed precision is not enabled during training. For data loading, parallel reading with workers=8 was enabled to enhance I/O throughput. Mosaic data augmentation remained enabled by default (close_mosaic=0), while dataset caching to disk was disabled (cache=False). The rest of the training workflow followed the default implementation of Ultralytics YOLOv10. The inference stage adopts the same image scaling and post-processing hyperparameters as training (using the framework's default settings for confidence threshold and NMS threshold). To meet the real-time requirements for bird-repelling cannon coordination, Test-Time Augmentation (TTA) is not additionally enabled.

Table 1. Experimental Environment Details

Configuration Name	Configuration Parameters
Operating System	Windows 11
CUDA/Driver	CUDA 12.8
Deep learning framework	PyTorch 2.11.0, Python 3.9

GPU	NVIDIA RTX4060

3.2 Model Evaluation Metrics

Comprehensively evaluate the effectiveness and deployability of the detection algorithm in the airport bird scenario, using precision (denoted as P), recall (denoted as R), average precision (AP), mAP@0.50 under the IoU=0.50 threshold, as well as detection speed (FPS) and model size (Params, weight magnitude) as comprehensive evaluation indicators. Precision characterizes "the proportion of samples correctly classified as birds among those labeled as birds by the model," while Recall reflects "the proportion of all true bird samples in the test set that are correctly detected." Taking binary classification detection as an example: TP (True Positive): Predicts positive samples as positive; FP (False Positive): Predicts negative samples as positive; FN (False Negative): Predicts positive samples as negative; TN (True Negative): Predicts negative samples as negative.

$$P=\frac{TP}{TP+FP} \tag{5}$$

$$R=\frac{TP}{TP+FN} \tag{6}$$

Plot the P-R curve at different confidence thresholds. The area under the curve represents the average precision (AP) of that category at a given IoU threshold, denoted as

$$AP=\int_0^1 P(t)dt \tag{7}$$

Here, P(t) denotes the precision corresponding to a recall rate of t. The arithmetic mean of AP across all categories yields mAP. Since this study includes only the "bird" target category, the number of categories N=1. The general formula for multi-category scenarios is:

$$mAP=\frac{1}{N}\sum_{n=1}^{N} AP_n \tag{8}$$

Beyond accuracy, real-time performance and resource overhead are equally critical. FPS indicates the number of frames processed per second, providing an intuitive measure of system throughput. Params characterizes the model's parameter count, measured in millions (M). Weight file size (MB) is also reported to gauge storage costs. During training and evaluation, four classification outcomes may occur: correctly predicting a positive sample as positive (TP), incorrectly predicting a negative sample as positive (FP), failing to detect a positive sample and misclassifying it as negative (FN), and correctly predicting a negative sample as negative (TN).

3.3 Training Results and Visualization Analysis

The training results of the improved YOLOv10n-attention model are shown in Figure 2. It can be observed that the three loss components (box/cls/DFL) decrease rapidly during the initial training rounds before entering a slow convergence phase. Correspondingly, the Precision, Recall, and mAP metrics on the validation set exhibit a monotonically increasing trend, stabilizing in the mid-to-late stages without abnormal fluctuations indicative of overfitting or underfitting. Specifically, the Precision curve achieves a leap-like improvement early on and stabilizes at a high level later, indicating effective suppression of false positives. Recall continues to improve throughout training and ultimately maintains a stable plateau, suggesting mitigation of false negatives in long-range and complex interference scenarios. The mAP@0.50 curve surpassed its inflection point after approximately several dozen iterations and gradually approached saturation. mAP@0.50:0.95 also exhibited sustained growth, albeit at a slower rate, reflecting simultaneous improvements in boundary regression quality and localization robustness. Combining these curves suggests that the unified training strategy not only ensures optimization stability but also progressively strengthens the model's ability to distinguish and localize small-scale, low-contrast bird targets. This provides a reliable detection front-end for subsequent real-time tracking in practical systems.

To visually demonstrate model performance, several representative scenarios were selected for inference visualization, as shown in Figure 3: In scenes featuring distant "small black dots" against highly reflective backgrounds, the model's confidence curve remained smooth with minimal bounding box jitter; Under backlighting and mild thermal disturbance conditions, candidate boxes maintain stable alignment with target contours. A small number of challenging instances only exhibit a slight dip in confidence scores without systemic missed detections. Combining training curves and visualization results, YOLOv10n-attention achieves a favorable accuracy-stability tradeoff in the "bird/non-bird" binary classification task and demonstrates engineering applicability for bird-deterrent cannon integration.

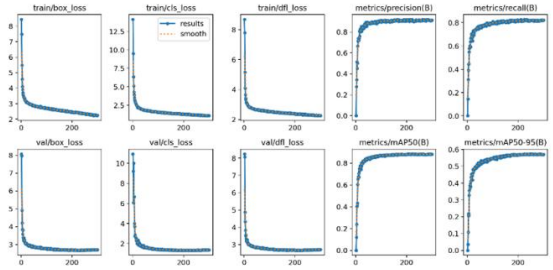

Figure 2. Training Results of the Improved YOLOv10n-Attention Model

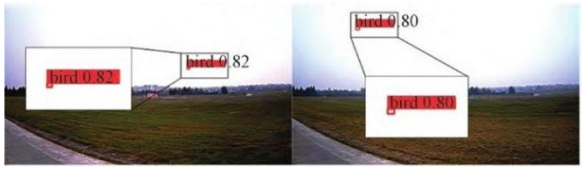

Figure 3. Detection Results of the Improved YOLOv10n-Attention Model

3.4 Melting Experiment

The ablation results of this experiment are shown in Table 2. As demonstrated in Table 3, both precision and recall of B1 improved compared to B0: precision increased from 0.802 to 0.814, recall rose from 0.761 to 0.773, and mAP@50 climbed from 0.810 to 0.822. Combining observations from the training curve and visualized examples, we infer that the attention mechanism effectively suppresses false positives in complex backgrounds while mitigating false negatives in scenarios involving long distances, weak textures, and mild thermal disturbances. Particularly in backlit and highly reflective scenarios, attention demonstrates stronger responsiveness to bird silhouettes and directional textures, yielding more stable candidate boxes. This manifests in reduced confidence fluctuations and smoother box tracking across consecutive frames in replay videos. Meanwhile, the number of parameters increased only slightly from 3.10M to 3.20M, with modest growth in weight size and computational load. The frame rate dropped from 31 to 30, largely maintaining real-time performance requirements, indicating this modification achieves a favorable balance between accuracy and computational efficiency.

Building upon B1, the introduction of the P5 structural bottleneck (B2) further improved recall from 0.773 to 0.783, elevated mAP@50 to 0.832, and yielded a slight increase in precision. Unlike attention mechanisms that primarily suppress false detections, the P5 structural bottleneck focuses on enhancing global semantic and geometric consistency within large receptive fields. Its effects are more pronounced in locating medium-to-large-scale and distant objects: improved alignment between bounding boxes and true contours, along with reduced box drift and jitter. The computational overhead introduced by this enhancement remains manageable. Parameters and weights only slightly increase to 3.22 million and 12.6 MB, respectively. GFLOPs rise to 9.6, while frame rates remain stable around 30, imposing no significant burden on online collaboration.

Sub-metric analysis indicates that improvements across the three experiments exhibit clear phased progression: The primary gains from B0 to B1 lie in enhanced "bird-like perception" capabilities—specifically, more effective capture of fine textures in complex backgrounds, manifested as simultaneous increases in precision and recall. The gains from B1 to B2 were primarily reflected in "more stable and accurate tracking," manifested as improved positioning stability across consecutive frames and enhancements in the medium-to-high IoU range. Ultimately, B2 achieved the optimal comprehensive metrics with minimal speed sacrifice, better meeting the dual requirements of real-time performance and stability demanded by airport scenarios.

Table 2. Ablation Experiment Results

Method	P	R	mAP@50	FPS	Params/M	Weight Magnitude /MB	GFLOPs
B0:YOLOv10n	0.802	0.761	0.810	31	3.10	12.0	9.0
B1:+Attention（backbone）	0.814	0.773	0.822	30	3.20	12.4	9.4
B2:+ Attention +C2fCIB@P5	0.819	0.783	0.832	30	3.22	12.6	9.6

3.5 Comparative Experiment

Under unified dataset partitioning, image scale (640×640), training workflow, and inference environment, our model was evaluated against common object detection models. Results are shown in Table 3. To ensure comparability, all methods were tested on identical hardware and post-processing thresholds. As shown in Table 3, the proposed YOLOv10n-attention achieves the highest mAP@50 while maintaining inference speeds comparable to lightweight single-stage methods. YOLOv4-tiny exhibits high frame rates but its accuracy falls significantly below both our method and the main YOLO series backbone. Two-stage or transformer-based

models (e.g., RT-DETR) demonstrate advantages in localization quality but suffer from high computational demands, large model sizes, and limited real-time performance. Considering accuracy, speed, and resource consumption comprehensively, YOLOv10n-attention demonstrates superior engineering adaptability in airport "bird/non-bird" scenarios: parameter count and weight size are kept within compact scales, GFLOPs increase is manageable, facilitating deployment on resource-constrained edge platforms.

Table 3. Comparative Experimental Results

Method	mAP@50	FPS	Params/M	Weight Magnitude /MB	GFLOPs
YOLOv4-s	0.806	24	11.2	22.8	28.0
YOLOv4-tiny	0.612	64	6.1	12.3	6.9
YOLOv3-tiny	0.472	118	8.7	16.9	8.2
YOLOv8n	0.822	29	3.2	12.5	8.8
SSD-MobileNet	0.561	51	5.4	15.1	3.5
CenterNet	0.495	22	34.2	82.1	28.1
RT-DETR	0.662	16	10.8	44.5	41.2
YOLOv10n-attention	0.832	30	3.22	12.6	9.6

4. Conclusion

Addressing the application challenges of "difficulty in detecting distant small targets, strong background interference, and high real-time requirements" in bird strike prevention at civil airports, this paper focuses on the identification-localization-locking requirements for bird-scaring cannons. We constructed the YOLOv10n-attention model based on an improved YOLOv10n and completed verification from model to engineering integration. The main conclusions are as follows.

(1) Under airport runway and near-field airspace scenarios, utilizing a unified 640×640 input and consistent training/inference workflow, the proposed model achieves real-time detection of "bird/non-bird" targets on the self-built Airport-Birds dataset: While maintaining parameters and computational complexity close to baseline models, mAP@50 improves to approximately 0.83, with inference speeds around 30 FPS, meeting the latency requirements for coordinated bird-scaring cannon deployment.

(2) Without altering the pyramid hierarchy or neck topology, introducing global attention at the trunk's upper layers and employing a C2fCIB bottleneck structure at the P5 branch stabilizes high-level semantic representation. Experiments demonstrate this design effectively suppresses false detections and bounding box jitter in backlit, highly reflective, and thermal disturbance scenarios, while improving localization consistency for distant and medium-to-large-scale targets.

(3) Comparative and ablation results demonstrate that the attention module simultaneously enhances precision and recall, while the P5 structural bottleneck further improves boundary regression quality and frame-to-frame stability. Their combined implementation achieves optimal overall performance with minimal speed sacrifice, offering a favorable "accuracy-computational power" balance and edge deployment feasibility.

(4) Focusing on bird-scaring cannon applications, this paper completes an engineering closed-loop evaluation of "detection → localization → servo control." Under fixed and mobile scenarios, locking latency is reduced, pointing error decreases, and tracking retention rate improves. This validates that front-end enhancements centered on detection stability directly translate into efficiency and reliability gains at the control layer.

(5) This research can integrate with existing airport security and operational monitoring systems, deploying as a distributed front-end along runways and critical airspace to enable all-weather automated monitoring and deterrence. Future work will explore multi-modal fusion of infrared/radar data, lightweight quantization and distillation, and trajectory-based intelligent decision-making under broader meteorological and traffic conditions to further enhance recall of distant small targets and long-term system stability.

References

[1] Chen W, Liu J, Wu D, et al. A review of wide-area bird monitoring and early warning technologies. Journal of Beijing University of Aeronautics and Astronautics, 2025: 1-22. DOI: 10.13700/j.bh.1001-5965.2025.0427.

[2] Du J, Shi F Y, Zhang Y B, et al. Bird diversity and bird strike prevention measures at airports in spring. China Science and Technology Information, 2023(15): 35-38. DOI: CNKI: SUN: XXJK.0.2023-15-012.

[3] Zhu W M, Tu C B. Analysis and prevention measures of bird damage to power transmission lines. Science & Technology Vision, 2015(31): 267-268+271. DOI: 10.19694/j.cnki.issn2095-2457.2015.31.207.

[4] Dai H F, Ming G Z, Zhao Y, et al. Introduction performance and cultivation techniques of 'Shannong Su Pear' in Feicheng, Shandong Province. Journal of Fruit Resources, 2022, 3(03): 91-93. DOI: 10.16010/j.cnki.14-1127/s.2022.03.028.

[5] Wang C. Research on the application of linkage control in intelligent bird perception and repelling. Yancheng Institute of Technology, 2024. DOI: 10.44381/d.cnki.gycit.2024.000091. DOI: 10.44381/d.cnki.gycit.2024.000091.

[6] Zhang H, Gong Y, Wang A J, et al. Research on bird damage identification methods for power transmission and transformation equipment based on improved YOLOv5. Journal of Nanjing Institute of Technology (Natural Science Edition), 2024, 22(02): 76-82. DOI: 10.13960/j.issn.1672-2558.2024.02.012.

[7] Zhang X W. Research on airport bird detection and tracking technology based on image processing. Civil Aviation Flight University of China, 2025. DOI: 10.27722/d.cnki.gzgmh.2025.000076.

[8] Ran X Y. Research on intrusion detection of transmission lines based on multi-source data fusion. Guangdong University of Technology, 2025. DOI: 10.27029/d.cnki.ggdgu.2025.001892.

[9] Wang Q Y, Yao G Q, Fang C Y. Research on lightweight bird target detection and recognition model based on YOLOv8n in Poyang Lake. Journal of Jiangxi Normal University (Natural Science Edition), 2025, 49(01): 86-94. DOI: 10.16357/j.cnki.issn1000-5862.2025.01.11.

[10] Han T. Research on bird target detection and recognition methods based on YOLOv8. Dalian Jiaotong University, 2025. DOI: 10.26990/d.cnki.gsltc.2025.001020.

[11] Shi Y L. Research on airport bird identification and counting based on deep neural networks. Civil Aviation University of China, 2023. DOI: 10.27627/d.cnki.gzmhy.2023.000158.

[12] Teng Y. Bird repelling recognition technology at airports based on multi-source bird feature fusion. Yancheng Institute of Technology, 2024. DOI: 10.44381/d.cnki.gycit.2024.000241.

[13] Xiao J, He X Z, Cheng H L, et al. Aerial small object detection algorithm based on multi-scale feature enhancement. Journal of Zhejiang University (Engineering Edition), 2025: 1-13. Available from: https://link.cnki.net/urlid/33.1245.T.20251009.1118.006.

[14] Zhang H J, Zheng P, Qiao W W, et al. A worm gear surface defect detection method based on improved YOLOv8n. Machine Tool & Hydraulics, 2025, 53(16): 10-17. DOI: CNKI: SUN: JCYY.0.2025-16-002.

[15] Hou Y, Wu Y, Kou X R, et al. A small-object detection algorithm for UAV aerial images based on improved YOLOv8. Computer Engineering and Applications, 2025, 61(11): 83-92.

[16] Xie R J. Implementation of real-time detection system based on lightweight YOLO. Information Systems Engineering, 2025(08): 12-15. DOI: CNKI: SUN: XXXT.0.2025-08-003.

[17] Zhang J L, Gao C Q, Zhao B. YOLOv8 reservoir type identification method based on attention enhancement and re-parameterized bidirectional feature pyramid network. Progress in Geophysics, 2025: 1-17. Available from: https://link.cnki.net/urlid/11.2982.P.20251015.0945.034.

[18] Zheng Y H, Chen E J, Wu F, et al. Small foreign object detection method for cigarette cutting production line based on improved YOLOv5s. Journal of Hefei University of Technology (Natural Science Edition), 2025, 48(09): 1183-1191. DOI: CNKI: SUN: HEFE.0.2025-09-005.

