RESEARCH ARTICLE

MFE-YOLO: A Multi-feature Fusion Algorithm for Airport Bird Detection

Hui Sun[1], **Yukun Wang**[1], **Jingjin Du**[1] **and Rui Wang**[1,*]

[1] School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China

Abstract

To address the issues of low accuracy in manual observation and slow detection by radar in airport bird detection, this paper designs a lightweight bird detection network named MFE-YOLO. This network is based on the YOLOv8 framework, with the main body part featuring an MF module replacing the original C2f module to enhance the network's feature extraction capability. An EMA mechanism is added to increase the focus on bird targets, and the Focal-Modulation module is introduced to reduce background interference. Additionally, a DCSlideLoss is designed during the supervised network training process to alleviate the imbalance of samples. Finally, the real-time detection performance is verified by combining the Byte Track algorithm, and generalization experiments are conducted on the public MS COCO dataset. The experimental results show that the MFE-YOLO algorithm has certain improvements in evaluation metrics such as Precision, Recall, and mean Average Precision, indicating that this algorithm has good detection performance and can achieve precise detection of birds in the low-altitude area of airports.

Keywords: airport bird detection, MFE-YOLO, multi-feature, attention mechanism, byte track algorithm.

1 Introduction

Bird strike incidents have always been a challenge for the civil aviation industry. With the vigorous development of the civil aviation industry and the sharp increase in passenger flow, the frequency of bird strike incidents worldwide has been on the rise, and the safety of the aircraft operating environment has received widespread attention [1, 2]. It is of great significance to effectively monitor the birds in the low-altitude areas of the airport to avoid the threat to civil aviation safety, especially the effective supervision of the aircraft runway and the surrounding low-altitude airspace is an urgent problem to be solved to ensure civil aviation safety, as quantitative risk assessment studies have confirmed the severity of bird strike threats at airports [3].

Traditional airport bird detection methods, such as manual observation and acoustic monitoring systems, suffer from several limitations. Manual observation is labor-intensive and lacks real-time responsiveness, while acoustic detection systems are highly susceptible to interference from surrounding environmental noise. To address these issues, a small radar-based detection system has been proposed to improve airport airspace

Academic Editor:
Bernard De Baets

Submitted: 28 March 2025
Accepted: 25 April 2025
Published: 21 May 2025

Vol. 2, **No.** 2, 2025.
10.62762/TIS.2025.323887

***Corresponding author:**
Rui Wang
ruiwang@cauc.edu.cn

Citation
Sun, H., Wang, Y., Du, J., & Wang, R. (2025). MFE-YOLO: A Multi-feature Fusion Algorithm for Airport Bird Detection. *ICCK Transactions on Intelligent Systematics*, 2(2), 85–94.

© 2024 ICCK (Institute of Central Computation and Knowledge)

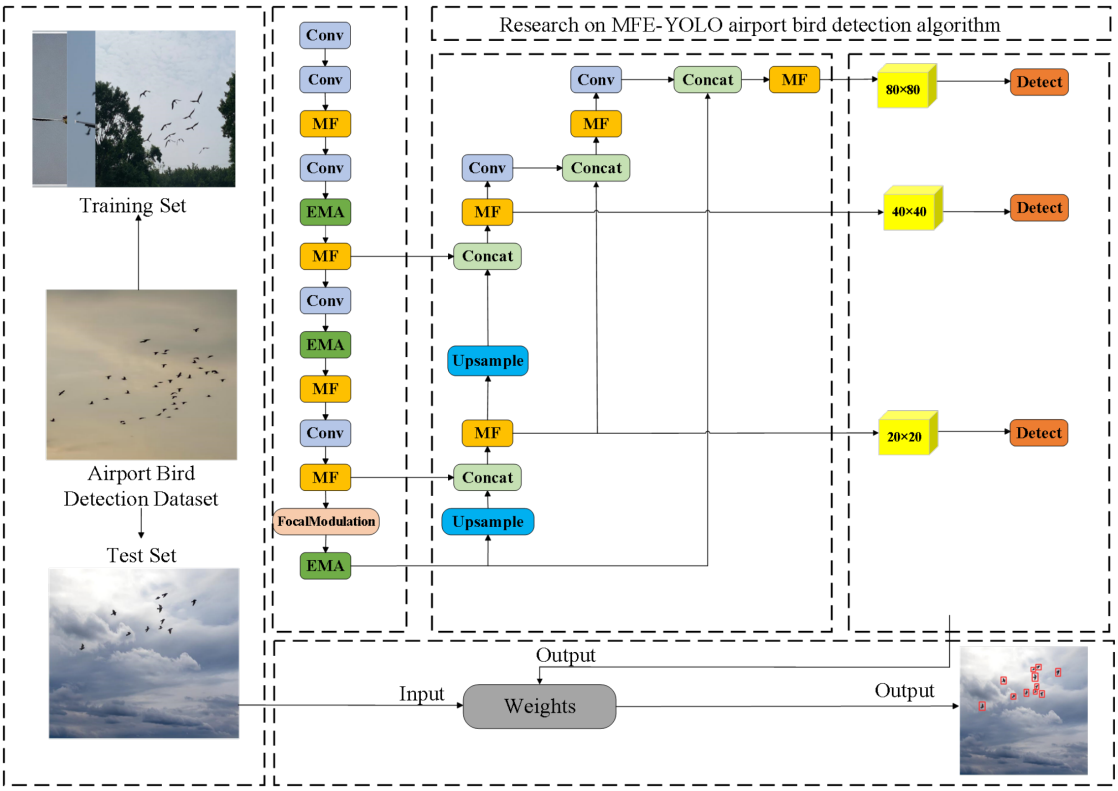


Figure 1. Structure of the MFE-YOLO network model.

safety through the effective detection of birds, drones, and wake vortices [4]. Reference [5] uses radar equipment for detection, but when the distance is relatively close, the resolution ability is limited and it is difficult to detect accurately. Reference [6] conducts bird detection through background subtraction and tracking corresponding points. It relies on background environment modeling and the detection is easily affected by light changes. Reference [7] applies deep learning to detect marine birds using high-resolution aerial images, but the detection process is computationally intensive and the method is not directly adaptable to the dynamic low-altitude airport environment. Reference [8] develops a vision-based path planning system for unmanned bird-repelling vehicles deployed at airports. While the system integrates visual perception to navigate and operate in airport environments, it is focused on vehicle autonomy rather than directly addressing the problem of accurately detecting and classifying bird targets in complex large-scale scenarios. Reference [9] proposes a clustered object detection method for aerial images that handles densely packed small targets, offering a useful reference for bird flock detection. However, its performance degrades significantly under varying illumination conditions in airport environments. Reference [10] proposes an airport bird detection algorithm based on a four-frame symmetric difference method combined with edge feature extraction. By integrating Canny edge detection, morphological processing, and connected domain analysis, the method improves the detection accuracy of low, slow, and small bird targets. However, the algorithm still relies on multi-frame differential operations and complex image-processing procedures, leading to relatively high computational complexity. Reference [11] proposes a cascaded sparse query method to accelerate small object detection in high-resolution images, which is applicable to detecting small bird targets. However, the method has limited adaptability to the wide range of scale variations exhibited by flying birds at different distances. To sum up, this study adopts a relatively advanced deep learning algorithm for airport bird detection, and it mainly faces the following difficulties: (1) the body size of birds varies greatly in scale in different images and is easily affected by background factors such as lighting. (2) the shape of the bird is similar to the aircraft, which is easy to cause misdetection. (3) the presence of flying birds to cover each other, the problem of leakage is serious. (4) the current detection of the public datasets shot at a relatively close distance, which can not be applied to the detection of airport birds.

To overcome the above difficulties, this paper proposes

a lightweight MFE-YOLO (Multi-Features-Efficient YOLO, MFE-YOLO) recognition network. The Multi-Features (MF) module is utilized to enhance the network's ability to extract different morphological characteristics of birds and reduce the computational complexity of model parameters. The Efficient Multi-Scale Attention (EMA) mechanism improves the degree of attention to bird targets rather than aircraft. Focal-Modulation reduces the interference of environmental factors such as illumination, and the adaptability to bird flock samples is improved through the DCSlideLoss (Decay-Slide Loss, DCSlideLoss). Finally, the detection ability of the algorithm for birds is verified through experiments and the real-time processing ability for birds is verified through Byte Track.

2 Overview of MFE-YOLO

The architecture of the MFE-YOLO network for airport bird detection based on deep learning is illustrated in Figure 1. First, this study introduces the MF module, which integrates dilated convolution, to replace the original C2f module in the network. This enhances the network's ability to extract multi-scale features of flying birds while effectively reducing the number of parameters and improving computational speed. Second, to mitigate the interference of environmental factors such as lighting, the original SPPF module is replaced with the Focal-Modulation module. Third, the EMA mechanism is incorporated to enhance the network's focus on bird targets. Fourth, the DCSlideLoss function is adopted during training to suppress gradients of extreme samples and increase the weight of regular samples. Finally, the Byte Track algorithm is employed to evaluate the real-time detection performance of the network, and generalization experiments are conducted using the MS COCO dataset. The following sections introduce the MF module, EMA mechanism, Focal-Modulation, DCSlideLoss, and Byte Track structure, followed by the experimental evaluation.

3 Methodology

3.1 MF Module

The target during bird flight is small and fewer features are obtained only by convolution stacking, in order to obtain more features for small targets, this paper proposes the MF module with reference to the CG Module structure of CG Network [12] and replaces the C2f module in the original structure with this module, This allows the network to generate a richer feature map while also reducing the number of parameters and the network's complexity of the original structure.

As seen in Figure 2, the MF block has three primary steps: Feature Learning, Connection Layer, and Feature Fusion. While the feature improvement phase seeks to refine joint features on channels, the feature learning step concentrates on learning spatially joint features.

As shown in the Feature Learning step in Figure 2, $G_{loc}(*)$, a local feature extractor, and $G_{sur}(*)$, a contextual feature extractor, are the primary components of the feature learning phase. They are used to learn the local features and the surrounding environmental context, respectively. The primary function of $G_{loc}(*)$, which is instantiated as a 3×3 conventional convolutional layer, is to learn the local features from the eight neighboring feature vectors, which correspond to the purple area in Figure 2(a). A 3×3 dilated convolution with a dilation rate of 2 is used to instantiate $G_{sur}(*)$. The broader sensory field easily learns contextual environment characteristics, corresponding to the yellow area in Figure 2(b).

This paper uses the Batch Normalization (BN) and Activation Function (SILU) as the connectivity layer between the Feature Learning and Feature Fusion phases, the contextual environment and local feature union of $G_{sur}(*)$ and $G_{loc}(*)$, and lastly, the output of the features to the feature fusion phase to perform the multilevel feature fusion in order to minimize the amount of computation.

As shown in the Feature Fusion step in Figure 2, the global context extractor, $G_{glo}(*)$, is a component of the Feature Fusion section that enhances the joint features by extracting global contextual information. The red-boxed area in Figure 2(c) represents the global average pooling layer and multilayer perceptron that are instantiated as $G_{glo}(*)$. The multilayer perceptron, which is composed of two fully-connected layers, is used to further extract the global contextual relationships after the global average pooling layer effectively aggregates the scene images. Lastly, the features are re-weighted for fusion using the extracted global contextual relationships.

3.2 EMA Mechanism

Due to the considerable variations in posture and shape exhibited by birds during flight, this paper integrates the EMA mechanism into the network architecture, as depicted in Figure 3. This integration enhances the network's feature extraction capability

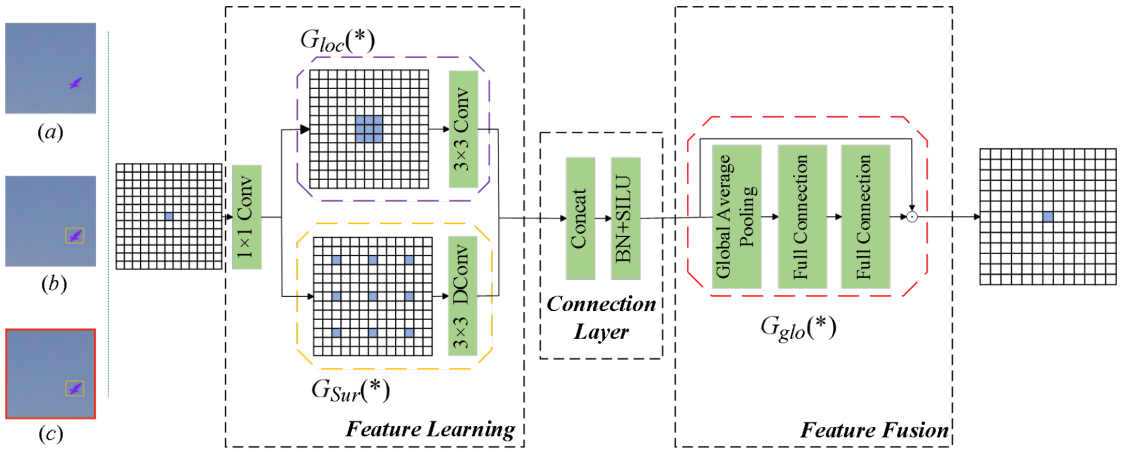


Figure 2. Structure of MF Module.

and improves its robustness to image transformations such as flipping and rotation, all while maintaining computational efficiency under limited resource constraints [13].

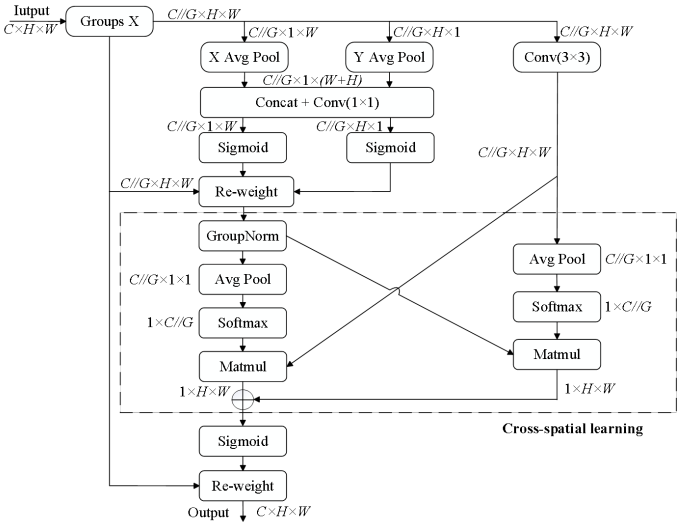


Figure 3. Structure of the EMA mechanism.

The EMA mechanism structure has three primary feature extraction routes, i.e., extracting attention weights for two 1×1 convolved paths and one 3×3 path convolution. The outputs of the two 1×1 branches and the 3×3 branch is introduced in the cross-space learning section, which models the cross-channel interaction of the information in both directions to achieve richer feature aggregation. The output of the smallest branch's channel features is then transformed into dimensional shapes that match those of the 3×3 branch, or $C//G$, and the global spatial information output from the 1×1 branch is encoded using 2D global average pooling. After reweighting the input feature map X using the Res-Net structure, the same dimensional feature vector is finally produced using the Sigmoid function, where C represents the number of input channels, H and W represent the input features' height and breadth in the spatial dimension, respectively, and $C//G$ represents the number of channels that each sub-feature occupies.

3.3 Focal-Modulation

This study introduces a Focal-Modulation module, grounded in focal modulation theory [14], to further improve the model's efficiency and accuracy in addressing multi-scale variations of flying birds while mitigating background interference.

Replacing the original SPPF module, the proposed module leverages the focal modulation mechanism to emphasize salient regions within the image, thereby capturing long-range dependencies and contextual information.

This enhancement significantly increases the model's suitability for detecting small and hard-to-identify bird targets, particularly in complex background environments.

Hierarchical context extraction from local to global scope at various granularity levels and gated aggregation of the extracted context features compressed into the modulator at various granularities comprise the two parts of this module's procedure, which is seen in Figure 4.

Given an input feature X, a context hierarchical representation of $L+1$ deep convolutional directions is obtained by projecting it into the new feature space of the linear layer, which is finally sent to the modulator. This part is passed through the linear layer to obtain spatial and level-aware weights, and then an element-by-element multiplication is performed to perform a weighted sum, finally obtaining a single

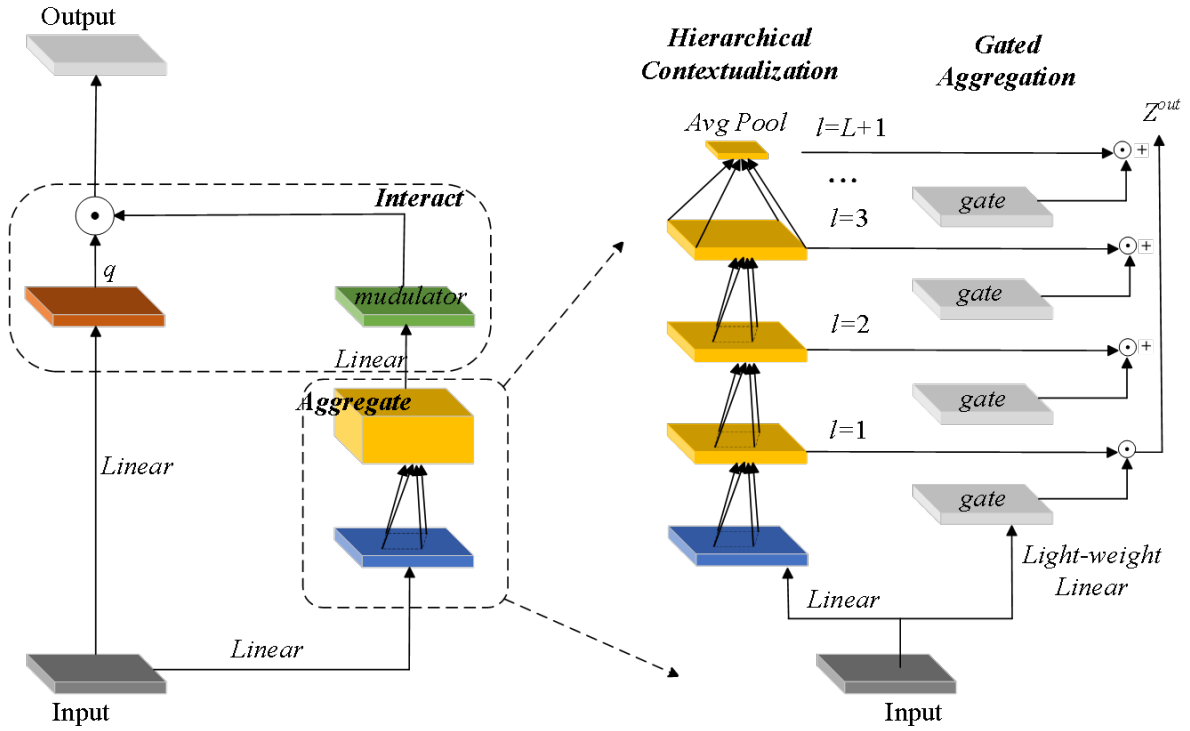


Figure 4. Structure of Focal-Modulation.

feature mapping of the same size as the input X.

3.4 DCSlideLoss

Due to the varying number of flying birds over the airport, high-density scenes contain significantly more complex local patterns and texture features compared to low-density or background regions. The gap between the regions is large, and the number of simple samples is significantly greater than the number of difficult samples, making it much easier to cause duplicate identification and omission detection. As a result, the joint loss function suggested in this study is the DCSlideLoss sample weighting function, which is shown below:

$$L = L_{DFL} + L_{CIoU} + L_{DCS} \tag{1}$$

where L_{DFL} is the distribution focus loss and L_{CIoU} is the bounding box loss, which is functionally defined as:

$$Loss_{DFL}(S_i, S_{i+1}) = (y - y_{i+1}) \log(S_i) + (y_i - y) \log(S_{i+1}) \tag{2}$$

$$L_{CIoU} = 1 - q + \left(\frac{\rho^2(b, b^{gt})}{c^2} + \frac{v^2}{(1-q)+v} \right) \tag{3}$$

where y is the label value, s is the target's label score, v is a scale parameter that measures the ratio between the anchor and target frames, b and b^{gt} are the anchor and target frames' centroids, respectively, ρ is the Euclidean distance between the two centroids, c is the diagonal distance of the smallest rectangle that covers both the anchor and target frames, p is the flyer's probability score, and q is the ratio of where the predicted and real frames intersect and concatenate (IoU).

L_{DCS} is the DCSlideLoss sample weighting function. The function is defined as follows:

$$L_{DCS} = \begin{cases} -\dfrac{v}{(1-q)+v} p \log(1-p) \\ [-q(q \log(p) + (1-q) \log(1-p))]\, e^{1-\mu} \\ [-q(q \log(p) + (1-q) \log(1-p))]\, e^{1-q} \end{cases} \tag{4}$$

This study presents the parameter μ as the threshold parameter of simple and difficult samples in the sample because the loss value is bigger near the threshold of difficult and simple samples. The DCSlideLoss sample weighting function uses an exponential function to weight the loss weight coefficients at the thresholds during the training process, while smoothing the loss weights of the samples near the transition thresholds, and the loss weight coefficients are shown in Figure 5.

When condition $0 \leq q \leq \mu - 0.1$ is met, indicating that the detection target is a simple sample, the weight coefficient is assigned a value of 1. Under condition $\mu - 0.1 < q < \mu$, where the target lies within the critical transition zone between simple and difficult samples,

to emphasize its intermediate complexity, the weight coefficient is adjusted to $e^{1-\mu}$. In the case of condition $\mu \leq q \leq 1$, where the target represents a difficult sample, the weight coefficient is set to e^{1-q} accordingly.

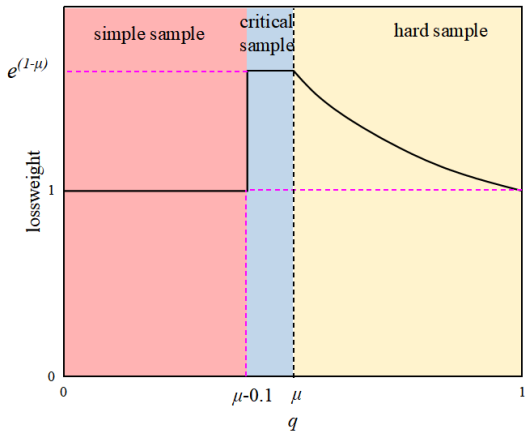


Figure 5. Sample weight coefficient function.

3.5 Joint Byte Track Structure

This paper uses the Byte Track algorithm in conjunction with the MFE-YOLO bird detection algorithm to validate the algorithm's detection ability in real-time scenarios because of the complex characteristics of flying birds, which include small targets, similar shapes, and a high overlap rate of flight processes. The block diagram of the matching strategy is displayed in Figure 6.

The exact procedure of the Byte Track algorithm is as follows: the detection targets of the high-confidence detection frame and the low-confidence detection frame are generated by first using the MFE-YOLO method for target identification in each picture frame. The Kalman filter predicts the motion trajectory of the detected target and generates the predicted trajectory frame. Next, the Hungarian algorithm is used to match the high-confidence target detection frame with the existing trajectory for the first association. For successfully matched trajectories, their Kalman filters are updated and put into the trajectory set, unmatched trajectory frames are put into the set T_{remin}, and unmatched detection frames are put into the set D_{remin}. Secondly by matching the low confidence prediction frame with the trajectory in T_{remin} for the second time. For successfully matched trajectories, the Kalman filter is updated and stored in the trajectory collection of the current frame, and those unsuccessful in both matches are deleted. The collection of trajectories is then produced for Kalman filtering to anticipate the new location of the trajectories. Lastly, all of the trajectories of the current frame are returned as the set of existing trajectories for the next frame of the picture.

4 Experiments

4.1 Airport Bird Detection Dataset

To thoroughly evaluate the performance of the proposed model, this study utilizes a dataset comprising images of flying birds near airports, which were collected both through field photography and online sources. A total of 5,186 images were selected and divided into training, testing, and validation sets at an approximate ratio of 7:2:1, comprising approximately 3,630 training images, 1,037 test images, and 519 validation images. Considering the visual similarity between birds and aircraft in real-world scenarios, 102 images containing both birds and airplanes were included in the dataset to reduce the risk of misclassification. Additionally, 72 images containing only airplanes were added as negative samples. To further assess the model's detection capability in dynamic scenes, five video clips of flying birds were also collected for training and evaluation, as illustrated in Figure 7.

4.2 Evaluation criteria

Following the standard evaluation protocols adopted in deep learning-based object detection surveys [16], this study utilizes the following evaluation metrics in the experiments: Precision (P), Recall (R), mean Average Precision (mAP), the number of Parameters (Param), and Giga Floating-point Operations (GFLOPs). These metrics are also widely adopted in detection and tracking studies covering person re-identification and multi-target scenarios [15]. For real-time detection performance, the Identification F1 score (IDF1), Multiple Object Tracking Accuracy (MOTA), and Multiple Object Tracking Precision (MOTP) metrics are employed, consistent with standard multi-object tracking evaluation practice. Since the tracked targets in this work are small flying birds, the evaluation additionally references benchmarks and protocols established for small object detection [17], which provide guidance on assessing detection quality under the scale and density conditions characteristic of bird targets.

4.3 Comparison with other methods

To further illustrate the robustness and excellent efficacy of the suggested approach in this article. In this paper, comparative experiments with SSD,

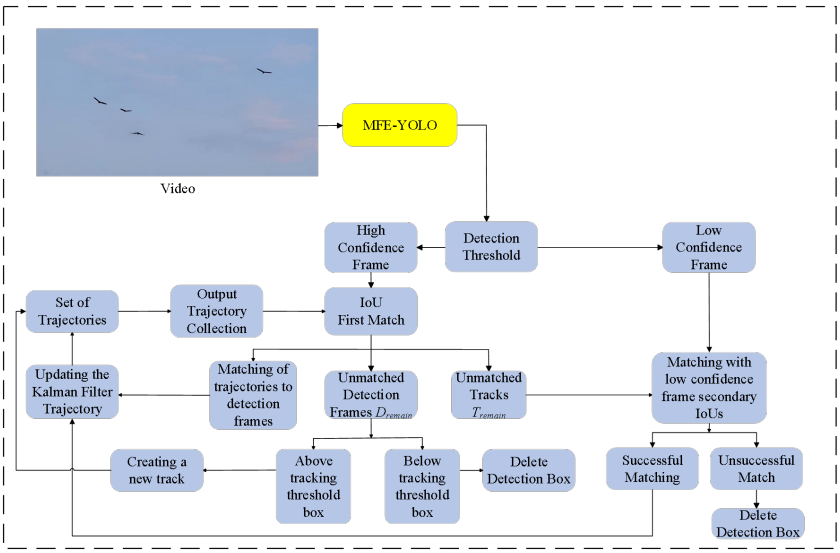


Figure 6. Joint structure of MFE-YOLO and Byte Track.



Figure 7. Airport Bird detection dataset (excerpt).

SMB-YOLOv5 [18], YOLOv5, YOLOv6, YOLOv7 [19], and YOLOv8 algorithms are conducted in the same configuration environment. Table 1 displays the outcomes of the experiment.

Table 1. Comparison experiment with other algorithms.

	P	R	mAP	Param (M)	GFLOPs
SSD	0.863	0.77	0.842	26	16.95
SMB-YOLOv5	0.814	0.706	0.750	12.3	-
YOLOv5	0.926	0.82	0.893	7.3	16.5
YOLOv6	0.887	0.72	0.872	4.63	11.34
YOLOv7	0.899	0.82	0.893	6	13
YOLOv8	0.926	0.82	0.896	3	8.1
Ours	0.949	0.84	0.91	2.2	6.1

It is fully demonstrated by the comparative verification of other methods that the MFE-YOLO detection technique, which is suggested in this study, is more suitable for detecting flying birds inside the airport and has a higher detection accuracy for flying birds above the airport. This work conducts comparison tests using the benchmark YOLOv8-Byte Track method, and ours-Byte Track algorithm to further validate the algorithm's real-time detection impact. The findings are shown in Table 2.

Experimental results demonstrate that the proposed method surpasses the benchmark network YOLOv8 combined with Byte Track in terms of detection performance. The algorithm effectively identifies flying birds in real-world airport scenarios, indicating that the MFE-YOLO framework exhibits strong stability and real-time capability in airport bird detection tasks.

Table 2. Comparison of real-time detection effect.

Methods	Dataset	IDF1	MOTA	MOTP
YOLOv8+Byte Track	001	19.7	11.5	28.02
	002	11.3	5.9	27.06
	003	57.8	69.1	75.67
	004	6.5	10.8	24.44
	005	51.1	71.6	84.69
Ours+Byte Track	001	63.2	82.6	49.06
	002	36.2	46	27.04
	003	85.4	88.2	75.67
	004	7.1	33.8	47.77
	005	51.7	71.6	90.12

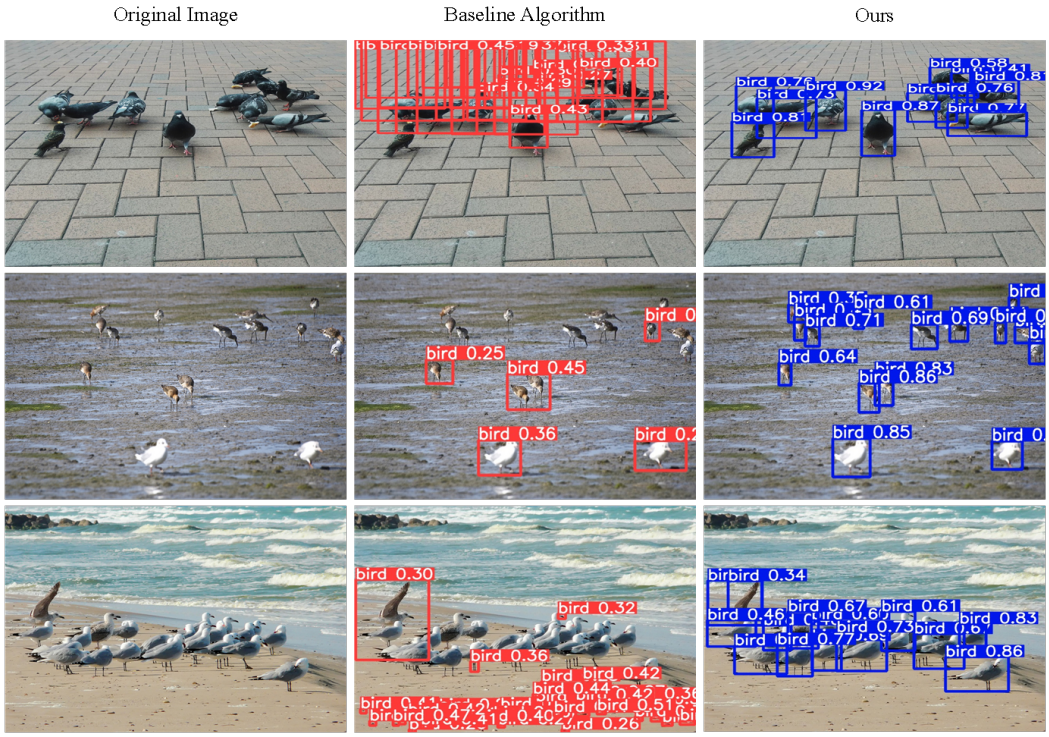


Figure 8. Visualization of the results of the baseline algorithm and the proposed method on the MS COCO dataset.

To further verify the generalization capability of the proposed method, localization and detection experiments were conducted on the publicly available MS COCO dataset. The proposed algorithm was compared with several state-of-the-art models, including YOLOv5, YOLOv7, YOLOv8, Faster R-CNN, and SSD. As illustrated in the Table 3, the proposed airport bird detection algorithm also demonstrates high localization and recognition performance for bird targets in natural environments.

Table 3. Comparative Experiments on the MS COCO Dataset.

	mAP	Param (M)	GFLOPs
YOLOv5	0.568	7.3	16.5
YOLOv7	0.528	6	13
YOLOv8	0.526	3	8.1
SSD	0.431	26	16.95
Faster R-CNN	0.592	28.3	-
Ours	0.596	2.2	6.1

Figure 8 intuitively presents visual comparisons among the original images, baseline algorithms, the proposed method, and the corresponding results on the MS COCO dataset.

4.4 Ablation Experiment

In this paper, ablation studies were systematically conducted to validate the contributions of each proposed module. As shown in Table 4, introducing the Multi-scale Fusion (MF) module alone reduces model parameters by 30% (from 3.0 M to 2.1 M) and computational cost by 28.4% (from 8.1 GFLOPs to 5.8 GFLOPs), but temporarily sacrifices detection performance (mAP drops 1.2%). The Dynamic Contextual Slide Loss (DCSlideLoss) independently improves mAP by 1.0% (0.896 → 0.905) without increasing model complexity.



Figure 9. Thermal map and localization of bird detection.

The combination of EMA attention and Focal-Modulation mechanisms increases parameters by 3.3% (3.0 M → 3.1 M) and GFLOPs by 3.7% (8.1 → 8.4), yet enhances precision by 1.3% (0.926 → 0.938). When fully integrating all components (MF + EMA + Focal-Modulation + DCSlideLoss), our final model achieves 2.2% higher precision (0.949

Table 4. Comparison of ablation experiment results of MFE-YOLO modules.

Improvements				P	R	mAP	Param/M	GFLOPs
MF	EMA	Focal-Modulation	DCSlideLoss					
-	-	-	-	0.926	0.82	0.896	3	8.1
✓	-	-	-	0.915	0.80	0.884	2.1	5.8
-	-	-	✓	0.912	0.82	0.905	3	8.1
-	✓	✓	-	0.935	0.80	0.892	3.1	8.4
✓	✓	✓	-	0.941	0.81	0.9	2.2	6.1
-	✓	✓	✓	0.938	0.82	0.887	3.1	8.4
✓	✓	✓	✓	0.949	0.84	0.91	2.2	6.1

vs 0.926), 2.4% better recall (0.84 vs 0.82), and 1.6% mAP improvement (0.910 vs 0.896) compared to the baseline YOLOv8, while maintaining 26.7% fewer parameters (2.2 M vs 3.0 M) and 24.7% lower computation (6.1 vs 8.1 GFLOPs). These results conclusively demonstrate the effectiveness of our co-design methodology for balancing efficiency and accuracy. Figure 9 shows the positioning map and heat map of the bird target predicted by the model. It vividly demonstrates that the MFE-YOLO algorithm can effectively detect bird targets in the airport scene and has good detection performance.

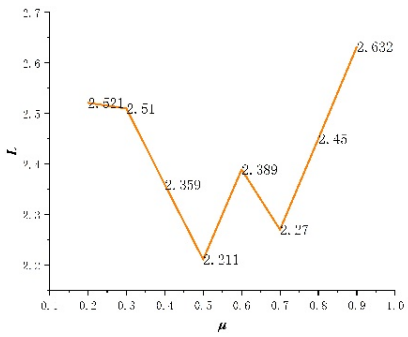


Figure 10. Line graph of experimental loss values L for μ values.

Furthermore, the DCSlideLoss mentioned in this paper is subjected to ablation experiments with weighted modifications using different values of μ. The experimental findings are shown in Figure 10. Loss value L relative to the parameter μ different values of the curve first decreases and then increases, when $\mu = 0.5$ when the loss of the minimum, so this paper takes $\mu = 0.5$.

5 Conclusion

This study addresses the safety risks of flying birds during airport aircraft takeoff and landing by proposing a more lightweight bird detection network (MFE-YOLO) for bird detection in low altitude airport locations. Experimental results demonstrate that the proposed network achieves improved stability, reduced model complexity, and higher accuracy in identifying flying birds at airports. The MF module is used to lighten the network, and the DCSlideLoss, the Focal-Modulation module, and the EMA mechanism are introduced to enhance the model's fusing capacity to the multi-scale aspects of the flying birds. Future work will focus on expanding the training dataset, optimizing the model for deployment on edge devices, and evaluating the proposed method in broader scenarios such as public safety monitoring and UAV-based obstacle avoidance.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the Key Research and Development Program of Tianjin, China under Grant 22YFZCSN00210.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Liu, Z., Wang, Z., Cao, Y., Cai, S., Chen, W., & Lu, K. (2025). Risk assessment of bird strike at airport based on improved CRITIC-cloud model. *Ecology and Environment*, *34*(1), 135. [CrossRef]

[2] Hu, Y., Xing, P., Yang, F., Feng, G., Yang, G., & Zhang, Z. (2020). A birdstrike risk assessment model and its application at Ordos Airport, China. *Scientific Reports*, *10*(1), 19627. [CrossRef]

[3] Yu, C., Zhao, F., Yin, Y., & Wu, Y. (2025). Entropy-weighted TOPSIS-based bird strike risk assessment for an airport. *International Journal of Computational Intelligence Systems*, *18*(1), 28. [CrossRef]

[4] Gong, J., Yan, J., Kong, D., & Li, D. (2024, September). Enhancing airport airspace safety: A small radar system for mitigating bird, drone, and wake vortex hazards. In *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC)* (pp. 1-10). IEEE. [CrossRef]

[5] Wu, H., Cheng, Z., Wei, K., Ma, J., & Li, X. (2022, October). Performance Evaluation of Bird Detection Radar in the Application of Airports. In *Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering* (pp. 756-760). [CrossRef]

[6] Shakeri, M., & Zhang, H. (2012, July). Real-time bird detection based on background subtraction. In *Proceedings of the 10th World Congress on Intelligent Control and Automation* (pp. 4507-4510). IEEE. [CrossRef]

[7] Boudaoud, L. B., Maussang, F., Garello, R., & Chevallier, A. (2019, June). Marine bird detection based on deep learning using high-resolution aerial images. In *OCEANS 2019-Marseille* (pp. 1-7). IEEE. [CrossRef]

[8] Wang, R., Li, J., Shi, Y., & Sun, H. (2022). Vision-based path planning algorithm of unmanned bird-repelling vehicles in airports. *Journal of Beijing University of Aeronautics and Astronautics*, *50*(5), 1446-1453. [CrossRef]

[9] Yang, F., Fan, H., Chu, P., Blasch, E., & Ling, H. (2019). Clustered object detection in aerial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8311-8320).

[10] Ma, Q., Cai, J., & Chen, Q. (2025, July). Design of Airport Bird Detection Algorithm. In *2025 4th International Conference on Electronics, Integrated Circuits and Communication Technology (EICCT)* (pp. 392-395). IEEE. [CrossRef]

[11] Yang, C., Huang, Z., & Wang, N. (2022). QueryDet: Cascaded sparse query for accelerating high-resolution small object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13668-13677). [CrossRef]

[12] Wu, T., Tang, S., Zhang, R., Cao, J., & Zhang, Y. (2020). CGNet: A light-weight context guided network for semantic segmentation. *IEEE Transactions on Image Processing*, *30*, 1169-1179. [CrossRef]

[13] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., & Huang, Z. (2023, June). Efficient multi-scale attention module with cross-spatial learning. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. [CrossRef]

[14] Yang, J., Li, C., Dai, X., & Gao, J. (2022). Focal modulation networks. *Advances in Neural Information Processing Systems*, *35*, 4203-4217.

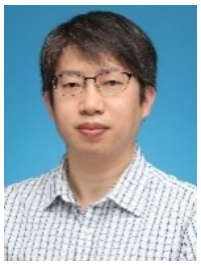
[15] Amirgaliyev, B., Mussabek, M., Rakhimzhanova, T., & Zhumadillayeva, A. (2025). A review of machine learning and deep learning methods for person detection, tracking and identification, and face recognition with applications. *Sensors*, *25*(5), 1410. [CrossRef]

[16] Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, *111*(3), 257-276. [CrossRef]

[17] Cheng, G., Yuan, X., Yao, X., Yan, K., Zeng, Q., Xie, X., & Han, J. (2023). Towards large-scale small object detection: Survey and benchmarks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(11), 13467-13488. [CrossRef]

[18] Liang, H., Zhang, X., Kong, J., Zhao, Z., & Ma, K. (2024). Smb-yolov5: A lightweight airport flying bird detection algorithm based on deep neural networks. *IEEE Access*, *12*, 84878-84892. [CrossRef]

[19] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7464-7475). [CrossRef]



Hui Sun, born in 1979, M.S., lecturer. His main research interests include wireless sensor networks, multi-agent system control, UAV swarm networking and coordination, and communications. (Email: h-sun@cauc.edu.cn)



Yukun Wang is a 2022 graduate student studying Control Engineering at the Civil Aviation University of China. His research focuses on airport bird defense. (Email: 1589179618@qq.com)



Jingjin Du is a 2024 graduate student studying Control Science and Engineering at the Civil Aviation University of China. Her research focuses on image processing. (E-mail: 15822168025@163.com)



Rui Wang, born in 1980, Ph.D, professor, master's advisor. Her main research interests include aerospace material management strategies, complex system modeling, nonlinear system modeling. (Email: ruiwang@cauc.edu.cn)