Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AMFD-YOLO: An Adaptive Multi-scale Feature Modulation Network for Small Object Detection in Airport Surface Surveillance

Tian Luan
Civil Aviation Flight University of China
Fan Yang
Civil Aviation Flight University of China
Huanxia Wei
huanxia.wei@manchester.ac.uk
University of Manchester
Weijun Pan
Civil Aviation Flight University of China

Article

Keywords: Airport Surface Surveillance, YOLO11, Small Target Detection, Multi-scale Occlusion-aware Attention, Feature Distribution Modulation

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11306278/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

AMFD-YOLO: An Adaptive Multi-scale Feature Modulation Network for Small Object Detection in Airport Surface Surveillance

Tian Luan[1,2], Fan Yang[3*], Huanxia Wei[4*], Weijun Pan[3]

[1]Luoyang Beijiao Airport, Civil Aviation Flight University of China, Luoyang, 471099, China.
[2]National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu, 610065, China.
[3]Civil Aviation Flight Technology and Flight Safety Research Base, Civil Aviation Flight University of China, Guanghan, 618307, China.
[4]Department of Mechanical and Aerospace Engineering, The University of Manchester, Manchester, M13 9PL, UK.

*Corresponding author(s). E-mail(s): fy@cafuc.edu.cn; huanxia.wei@manchester.ac.uk;
Contributing authors: luantian@cafuc.edu.cn; wjpan@cafuc.edu.cn;

Abstract

Accurate detection of airport surface targets, including aircraft, tow tractors, catering trucks, refueling trucks, baggage carts, and ground staff, is essential for intelligent surface surveillance and automated ground operation management. However, detection remains challenging because of complex and time-varying backgrounds, large target-scale variation, a high proportion of small objects, and frequent mutual occlusion. To address these issues, we propose AMFD-YOLO based on the YOLO11 framework. First, a Multiscale Occlusion-aware Attention Module (MOAM) is embedded in the detection head to enhance responses to occluded targets while suppressing background interference. Second, a C3k2_FDM module incorporating the Feature Distribution Modulation (FDM) activation function is introduced to improve feature extraction under complex lighting and weather conditions. In addition, a P2 detection layer enhances the perception of tiny targets. Experiments on the Airport Surface Surveillance (ASS) dataset show that AMFD-YOLO improves precision, recall, mAP0.5, and mAP0.5-0.95 by 7.5%, 23.2%, 23.6%, and 12.5%, respectively, over baseline YOLO11. Ablation studies further verify the contribution of each component.

These results demonstrate the effectiveness of AMFD-YOLO for airport surface target detection, particularly under adverse environmental conditions and visually crowded apron and taxiway scenes, and its potential for intelligent airport surveillance.

Keywords: Airport Surface Surveillance, YOLO11, Small Target Detection, Multi-scale Occlusion-aware Attention, Feature Distribution Modulation

1 Introduction

With the rapid development of civil aviation systems, airport operation safety and efficiency have become increasingly important in modern air traffic management. Airport surface surveillance plays a critical role in ensuring safe and efficient operations by continuously monitoring aircraft, vehicles, and personnel. Traditional surveillance technologies, including multilateration (MLAT), automatic dependent surveillance–broadcast (ADS-B), and radar, can provide reliable target perception under all-weather conditions, but offer limited fine-grained semantic information about ground objects and operational activities [1]. Therefore, computer vision has become an important complementary modality for comprehensive airport surface perception.

Deep learning-based vision methods have substantially advanced intelligent perception in airport environments, including object detection, tracking, and anomaly recognition [2]. However, practical airport surveillance remains challenging due to adverse weather, nighttime conditions, uneven illumination, and complex ground environments, which can significantly degrade visual detection performance [3]. Moreover, aircraft, vehicles, and personnel often occupy only a small number of pixels in wide-area surveillance images, resulting in severe feature loss during network downsampling. Large scale variations, occlusion, and background clutter further increase the difficulty of accurate target detection. The limited availability of high-quality annotated airport surveillance datasets also restricts the development of robust vision-based perception methods [4].

Small-object detection has therefore attracted considerable attention, with existing studies mainly focusing on feature enhancement and multi-scale representation to compensate for the limited visual information of small targets [5]. Transformer-based and state-space-based architectures have further been explored to improve contextual modeling and feature representation for small objects in aerial imagery [6, 7]. In airport-specific applications, improved YOLO-based methods have also been developed for apron surveillance and ground-service-related object detection, demonstrating the effectiveness of task-oriented and lightweight detection frameworks [8–10]. FOD detection methods have additionally been investigated to enhance airport runway safety under complex environmental conditions [11]. Nevertheless, existing approaches still face challenges in simultaneously preserving discriminative features of small targets, capturing contextual information, and maintaining computational efficiency for practical deployment.

Recent studies have shown that modules such as Poly-Kernel Inception (PKI) and Context Anchor Attention (CAA) can enhance feature representation through multi-kernel feature extraction and contextual modeling [12]. However, existing approaches may still struggle to preserve discriminative features of small and partially occluded targets while suppressing background interference in complex airport scenes. In particular, repeated feature down-sampling can lead to the loss of fine-grained information required for extremely small targets, while occlusion and complex backgrounds further weaken target representation. Therefore, improving occlusion-aware feature representation, enhancing foreground-background discrimination, and preserving high-resolution features for extremely small objects remain important challenges in airport surface object detection.

Based on the above analysis, this paper proposes AMFD-YOLO, a small-object detection method for airport surface surveillance based on an improved YOLO11 framework. The main contributions of this paper are as follows:

(1) A Multi-scale Occlusion-aware Attention Module (MOAM) is introduced to enhance feature representation of occluded targets and improve robustness in complex airport scenes.

(2) The FDM module is embedded into the C3k2 structure to form a C3k2_FDM block for enhanced foreground feature extraction and background suppression, and an additional P2 detection head is introduced to improve extremely small object detection in high-resolution airport surveillance images.

Experimental results on the ASS dataset demonstrate that AMFD-YOLO achieves superior performance compared with other YOLO series detection methods in detecting aircraft, vehicles, and personnel under complex airport surface conditions.

The remainder of this paper is organized as follows: Section 2 introduces related work; Section 3 describes the proposed network architecture; Section 4 presents experimental settings and results; Section 5 discusses limitations and future work; Section 6 concludes the paper.

2 Related Work

2.1 Object Detection and YOLO Series Development

Object detection, as a core task in computer vision, aims to classify and localize specific objects in images. With the rapid development of deep learning, convolutional neural network-based methods have become the dominant paradigm, replacing traditional hand-crafted feature extraction approaches.

Object detection methods can be broadly divided into two categories: two-stage and one-stage detectors. Two-stage methods such as Fast R-CNN [13], Faster R-CNN [14], and Mask R-CNN [15] generally achieve high accuracy but suffer from high computational cost and limited real-time performance. In contrast, one-stage methods such as SSD [16], RT-DETR [17], and the YOLO series directly predict object classes and bounding boxes in an end-to-end manner, achieving a better trade-off between accuracy and speed, making them more suitable for real-time and resource-constrained scenarios.

Since the YOLO series of algorithms was launched in 2016, continuous technical iterations have been carried out. YOLOv1 uses the Darknet19 architecture to directly predict bounding boxes and class probabilities in one network, greatly improving the detection speed [18]. YOLOv2 expands the system resolution, supports the detection of more than 9,000 object categories, and improves the model's applicability and accuracy [19]. YOLOv3 introduces a multi-scale prediction mechanism and three sizes of detection kernels, enhancing the perception ability of different scale targets [20]. YOLOv4 adopts CSP connections and Mosaic data augmentation, improving the learning efficiency while reducing the computational load [21]. YOLOv5 optimizes the backbone, neck, and head structures, further improving the overall detection performance [22]. YOLOv6 introduces a bidirectional connection module and an anchor-assisted training strategy, achieving a good balance between detection speed and accuracy [23]. The Efficient Layer Aggregation Network (E-ELAN) proposed by YOLOv7 enhances the network's feature learning ability without disrupting the original architecture [24]. YOLOv8 replaces C3 with the C2f module with better gradient flow and combines a decoupled head with a task-aligned matching strategy to optimize the feature extraction and fusion efficiency [25]; YOLOv9 improves the model's generalization ability in diverse scenarios through programmable gradient information technology and a generalized efficient layer aggregation network [26]. YOLOv10 proposes a new real-time object detection method, addressing the post-processing and model architecture defects in previous YOLO versions [27]. While maintaining real-time detection capabilities, YOLO11 strengthens the extraction and expression of multi-scale features, providing a more solid foundation for detection in complex scenarios [28]. YOLOv12 introduces a regional attention mechanism and a residual efficient layer aggregation network, achieving higher detection accuracy while maintaining real-time performance [29]. YOLOv13 proposes the HyperACE and FullPAD mechanisms, enhancing multi-scale feature fusion and the ability to adapt to complex scenarios [30]. YOLOv26 is designed specifically for edge and low-power devices, simplifying the design architecture to achieve a lighter and faster deployment [31].

2.2 Research Status of Airport Scene Object Detection

Airport surface surveillance plays a critical role in ensuring aviation safety. Due to the elevated camera installation and wide monitoring coverage, most targets such as aircraft, vehicles, and personnel appear as small objects, resulting in limited feature information and increased detection difficulty compared with general object detection tasks [1, 32]. Moreover, the complex airport environment, including large-scale background variations and diverse target appearances, further challenges the effective representation of small objects. Apart from aircraft, ground vehicles, and on-site staff, runways and aprons may also contain intrinsically tiny foreign-object debris (FOD), further increasing the difficulty of fine-grained visual perception in airport surveillance [33]. Therefore, improving small-object feature representation while maintaining detection accuracy and computational efficiency has become an important research focus in airport scene object detection. To address these challenges, various improved methods have been proposed [12, 34, 35]. These approaches mainly focus on enhancing fine-grained and multi-scale feature representation, introducing attention

or Transformer-based mechanisms, and improving detection efficiency for practical airport surveillance scenarios.

Based on the ASS dataset, Zhou et al. constructed ASSD-YOLO by introducing attention mechanisms and Transformer encoders into YOLOv7, improving detection accuracy in complex scenes [36]. They further proposed LAS-YOLO, which adopts lightweight modules and replaces heavy pyramid structures with SPPF to reduce computational cost while maintaining performance [37]. To enhance robustness under adverse conditions, ADH-YOLO introduces a decoupled detection head and coordinate attention to improve performance in haze environments [9]. For real-time requirements, AD-YOLO integrates Swin Transformer and channel-spatial attention mechanisms to improve feature representation efficiency [7]. For airport apron scenarios, He et al. proposed DT-YOLO by integrating Transformer attention and deformable convolution to enhance multi-scale and deformable object representation [38]. Xu et al. introduced SPD-Conv into YOLOv5 to improve feature extraction for moving ground service targets [10]. In aerial surveillance scenarios, Li et al. combined multi-scale feature fusion and Wasserstein-based loss functions to improve small-object detection accuracy while maintaining efficiency [39]. For runway detection in remote sensing images, Wan et al. proposed CR-Mask R-CNN with rotated region proposals and attention mechanisms to reduce false and missed detections in complex backgrounds [40].

3 Airport Scene Recognition Algorithm Based on YOLO11

3.1 YOLO11 Network Architecture

The YOLO11 model adopts the C3k2 module, the Spatial Pyramid Fast Pooling (SPPF) module, and the Cross-Stage Partial Spatial Attention (C2PSA) mechanism. With a lightweight design and strong adaptability, YOLO11 can be deployed on edge devices and cloud platforms for real-time detection in airport scenarios.

Although YOLO11 performs well in general tasks, it has clear limitations in airport surveillance. The airport environment is complex and dynamic. Adverse weather such as haze, rain, snow, and low visibility degrades visible-light imaging, weakening feature extraction for aircraft, vehicles, and ground personnel, and reducing overall robustness for all-weather monitoring. In addition, airport targets vary widely in scale, from large aircraft to small vehicles and ground staff. The model tends to lose spatial details of small objects during deep feature extraction and lacks a dynamic multi-scale adaptation mechanism, leading to missed small targets and localization errors for large ones. Moreover, in cases of similar textures and occlusion on the apron, its discriminative ability is weakened, making it difficult to distinguish normal operations from abnormal intrusions, which fails to meet high-security and high-precision requirements.

To address these issues, this paper proposes AMFD-YOLO. A multi-scale attention mechanism (MOAM) is introduced to enhance occluded target features and reduce missed detections. Meanwhile, the normalization layer is replaced with a dynamic Tanh-based activation (FDM), which is integrated into the C3k2 module to form the C3k2_FDM block. This design enables adaptive feature scaling and improves the

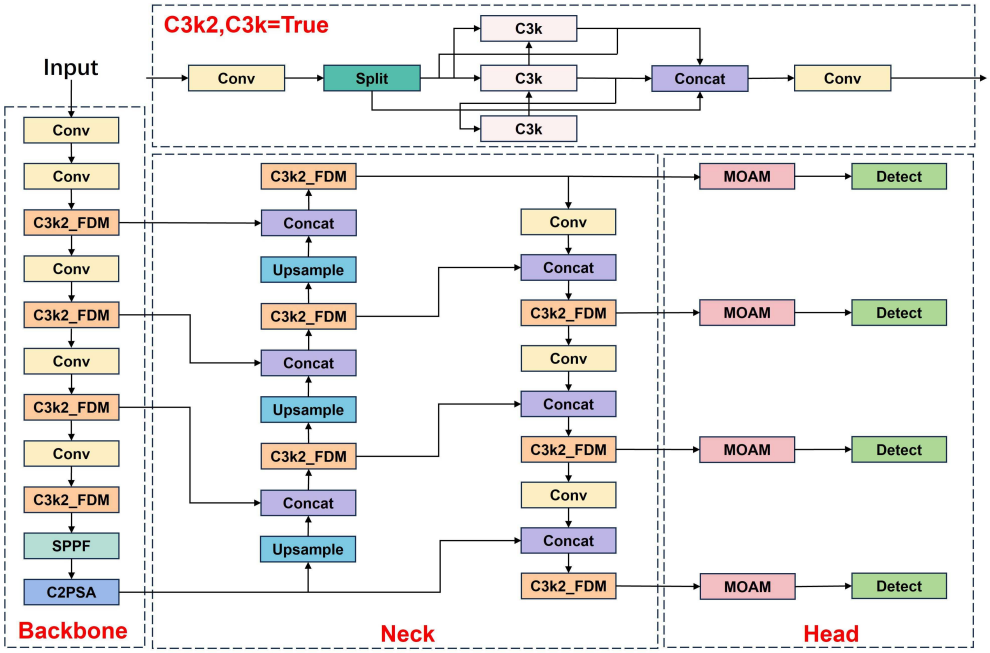


Fig. 1 AMFD-YOLO Network Architecture

model's sensitivity to complex environmental variations. In addition, a P2 detection head is added to enhance small-object perception. The overall improved architecture is illustrated in Figure 1.

3.2 MOAM

To address target occlusion in airport surveillance scenarios, we propose a Multi-scale Occlusion-aware Attention Module (MOAM), as shown in Figure 2. The module enhances feature representations of occluded targets in the network, thereby improving detection performance under severe occlusion conditions. It alleviates feature loss, localization errors, and local feature aliasing caused by object overlap, while strengthening salient foreground features and suppressing background interference. Consequently, it improves feature compensation for occluded objects and enhances localization stability.

Given an input feature map:

$$\mathbf{X} \in \mathbb{R}^{B\times C\times H\times W} \tag{1}$$

we construct multiple scale-specific transformation branches. For each scale p_k, feature embedding is implemented via a convolution with stride equal to the patch size:

$$\mathbf{E}^{(p_k)} = \mathrm{BN}\Big(\phi(\mathrm{Conv}_{p_k\times p_k, s=p_k}(\mathbf{X}))\Big) \tag{2}$$

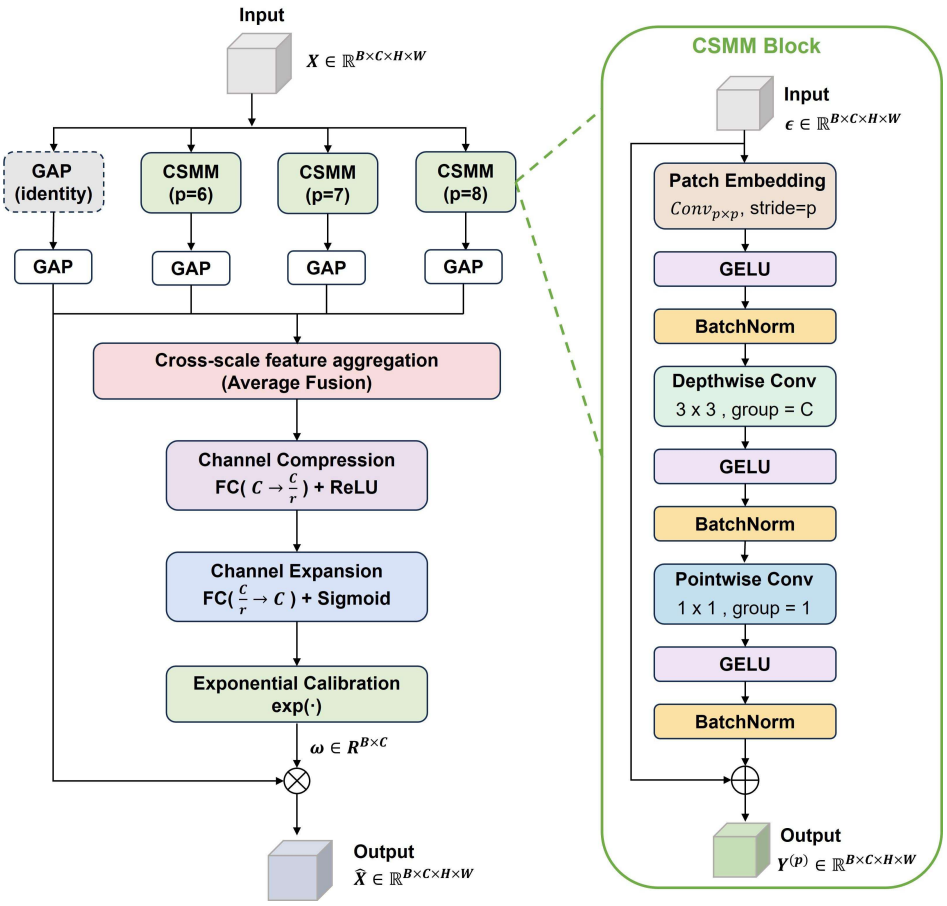


Fig. 2 Architecture of the MOAM Module

where $\phi(\cdot)$ denotes the GELU (or SiLU) activation function and BN denotes batch normalization.

Each scale branch further performs feature refinement through stacked depthwise separable convolution blocks with residual connections. Specifically, given the input feature representation $\mathbf{U}_{t-1}^{(p_k)}$ of the t-th block, depthwise convolution is first employed to enhance local spatial information, followed by pointwise convolution for channel interaction:

$$\widetilde{\mathbf{U}}_{t}^{(p_k)} = \mathbf{U}_{t-1}^{(p_k)} + \mathrm{BN}\Big(\phi(\mathrm{DWConv}_{3\times 3}(\mathbf{U}_{t-1}^{(p_k)}))\Big), \tag{3}$$

$$\mathbf{U}_{t}^{(p_k)} = \mathrm{BN}\Big(\phi(\mathrm{PWConv}_{1\times 1}(\widetilde{\mathbf{U}}_{t}^{(p_k)}))\Big). \tag{4}$$

After multiple stacked blocks, each branch produces:

$$\mathbf{Y}^{(p_k)} = F^{(p_k)}(\mathbf{X}) \tag{5}$$

To integrate multi-scale information, global average pooling is applied:

$$\mathbf{z}^{(p_k)} = \mathrm{GAP}(\mathbf{Y}^{(p_k)}), \quad \mathbf{z}^{(x)} = \mathrm{GAP}(\mathbf{X}) \tag{6}$$

The overall channel descriptor is obtained by averaging the original and multi-scale feature statistics:

$$\bar{\mathbf{z}} = \frac{1}{K+1}\left(\mathbf{z}^{(x)} + \sum_{k=1}^{K} \mathbf{z}^{(p_k)}\right). \tag{7}$$

Based on the aggregated descriptor, a lightweight channel attention mechanism is introduced to generate adaptive channel weights:

$$\mathbf{a} = \sigma\Big(\mathbf{W}_2\delta(\mathbf{W}_1\bar{\mathbf{z}})\Big), \quad \mathbf{a} \in (0,1)^C, \tag{8}$$

where $\delta(\cdot)$ denotes the nonlinear activation function and $\sigma(\cdot)$ represents the sigmoid function. To further strengthen the response of informative channels under complex occlusion conditions, an exponential calibration strategy is adopted:

$$\mathbf{w} = \exp(\mathbf{a}). \tag{9}$$

Finally, the output feature map is obtained via channel-wise modulation:

$$\hat{\mathbf{X}} = \mathbf{X} \odot \mathbf{w} \tag{10}$$

which can also be written in compact form as:

$$\hat{\mathbf{X}} = \mathbf{X} \odot \exp\Big(\sigma(\mathbf{W}_2\delta(\mathbf{W}_1 \cdot \mathrm{GAP}(F(\mathbf{X}))))\Big) \tag{11}$$

Overall, MOAM improves feature robustness by combining multi-scale spatial modeling and exponential channel reweighting, which is particularly effective in occluded and crowded scenes.

3.3 C3k2_FDM+P2 Detection Layer

The proposed C3k2_FDM module is built upon the original C3k2 architecture in YOLO11, which integrates multi-scale convolutional kernels and channel-splitting strategies to achieve efficient feature extraction under computational constraints.

To improve robustness in complex airport surveillance scenarios, we redesign the feature modulation mechanism within C3k2 by introducing Feature Distribution Modulation (FDM), while preserving the original topological structure of the network.

Let the input feature map be:

$$\mathbf{x} \in \mathbb{R}^{C\times H\times W}, \tag{12}$$

In standard YOLO architectures, feature modulation is implicitly performed via normalization layers and activation functions. However, this coupling leads to limited flexibility under severe distribution shifts in dense and cluttered scenes.

To address this limitation, we introduce an explicit modulation function, Feature Distribution Modulation (FDM), defined as:

$$\mathrm{FDM}(x) = \gamma \odot \tanh(\alpha x) + \beta, \tag{13}$$

where α is a learnable scalar controlling the saturation strength of the nonlinear mapping, and $\gamma, \beta \in \mathbb{R}^C$ are channel-wise affine parameters.

The identity branch in the C3k2_FDM module is defined as a direct shortcut path that preserves part of the input features without transformation:

$$\text{Identity branch: } \mathbf{z}_1 = \mathbf{x}, \tag{14}$$

which enhances feature reuse and stabilizes gradient propagation.

The transformation branch processes the remaining features through a sequence of Bottleneck or C3k blocks:

$$\mathbf{z}_{k+1} = \mathcal{F}(\mathbf{z}_k), \quad k = 1, \ldots, n, \tag{15}$$

where $\mathcal{F}(\cdot)$ denotes either Bottleneck or C3k transformation.

The modulated feature map is first projected via a 1×1 convolution:

$$\mathbf{z} = \text{Conv}_{1\times1}(\text{FDM}(\mathbf{x})) \in \mathbb{R}^{2C' \times H \times W}, \tag{16}$$

and then split into identity and transformation branches:

$$\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2]. \tag{17}$$

Finally, features are fused via concatenation followed by a 1×1 convolution:

$$\mathbf{y} = \text{Conv}_{1\times1}(\text{Concat}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{n+1})). \tag{18}$$

The overall C3k2_FDM structure is illustrated in Fig. 3.

This unified visualization avoids redundant depiction of the C3k2 architecture by integrating the C3k2 topology, the transformation branches, and the proposed FDM-based feature modulation mechanism into a single framework.

Finally, to further enhance small-object sensitivity in airport surveillance, a P2 detection head is introduced into the original YOLO11 detection pyramid, forming a four-scale detection structure $\{P2, P3, P4, P5\}$. The P2 layer preserves high-resolution spatial details and significantly improves the detection of small and densely distributed objects.

The original design of YOLO11 uses an 8-fold downsampled feature map, which has a relatively large receptive field for tiny targets and is prone to losing detailed information during feature extraction. To alleviate the resulting missed detection problem, this paper adds a P2 detection head for tiny targets on the basis of the original three-scale detection heads of P3, P4, and P5, forming a four-scale detection architecture. The P2 layer is located in the shallow layer of the network and has a high resolution. It retains rich fine-grained features and spatial location information, and can effectively capture the contour and texture details of small targets. After adding the detection head, the model's detection ability for targets of different scales is enhanced, and the overall detection accuracy and robustness are further improved.

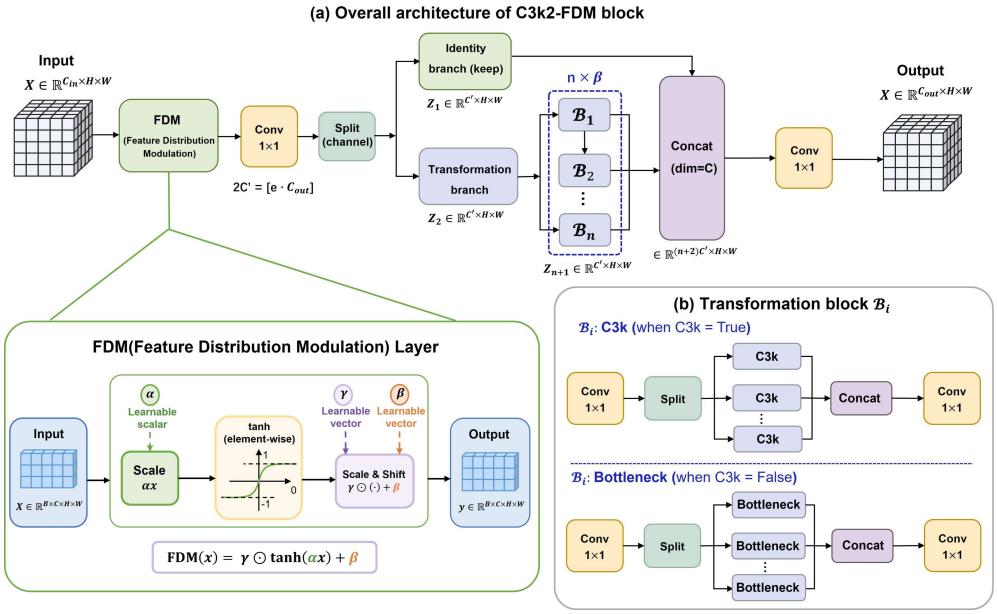


Fig. 3 Schematic Diagram of C3k2_FDM

4 Experiments and Result Analysis

4.1 Datasets

The ASS (Airport Surface Surveillance) dataset is an open dataset constructed by Zhou et al. for the research on object detection in airport surface surveillance scenarios [36]. This dataset is collected based on a real airport surface surveillance system and can better reflect the characteristics of actual airport surface surveillance application scenarios compared to other publicly available aircraft datasets. The ASS dataset contains two subsets: ASS1 (Airport Surface Surveillance Dataset) and ASS2 (Airport Panoramic Surveillance Dataset), corresponding to different surveillance perspectives and scenario characteristics. In this paper's experiment, the ASS1 dataset is mainly used. This dataset is constructed by collecting typical surveillance images from a real airport surface surveillance system, covering surveillance images under various environmental conditions such as sunny days, late nights, and rainy days. It contains a total of 2000 typical surveillance images, annotated with three common targets: airplanes, personnel, and vehicles, with a total of 8371 target bounding boxes. The ASS dataset is manually annotated using the Labeling graphical annotation tool and randomly divided into a training set and a test set in an 8:2 ratio.

4.2 Experimental Environment

In this paper, PyTorch 2.0.1 is used as the deep learning framework, and the Python version is 3.10.12. The experimental environment is based on the Ubuntu 20.04 operating system and the CUDA 11.8 computing platform. In terms of hardware, the model is trained using an NVIDIA GeForce RTX 4090 graphics card (with 25.2 GB of video memory). The training process is iterated 200 times, the batch size is set to 16, and the input image resolution is uniformly set to 640×640. The optimization algorithm uses Stochastic Gradient Descent (SGD), and the initial learning rate is set to 0.01. To ensure the fairness and reproducibility of the experimental results, all comparison methods and ablation experiments are executed under the above unified environment configuration and parameter conditions. By strictly controlling the experimental variables, external factors can be excluded from interfering with the experimental results, making the performance differences solely attributable to the algorithm design itself. This approach enhances the credibility of the experimental results and provides a reliable basis for horizontal comparison between subsequent methods.

4.3 Evaluation Metrics

The experiment uses multi-dimensional metrics to systematically evaluate the improved model. In terms of detection accuracy, Precision (P), Recall (R), and Mean Average Precision (mAP, including mAP0.5 and mAP0.5-0.95) are selected as the evaluation criteria. Precision is defined as

$$P = \frac{TP}{TP + FP} \tag{19}$$

reflecting how many of the positive examples predicted by the model are real, and the higher its value means fewer false positives;

Recall is defined as

$$R = \frac{TP}{TP + FN} \tag{20}$$

Recall represents the proportion of real targets that are correctly detected, and the higher its value means fewer missed detections;

The average precision measures the average precision performance of a single-class detector at different recall rate levels by calculating the area under the P-R curve with recall rate as the horizontal axis and precision as the vertical axis,

$$AP = \int_0^1 P\left(R\right) dR \tag{21}$$

while mAP takes the arithmetic mean of the APs of all classes and is a comprehensive indicator to measure the overall accuracy of a multi-class detection model.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{22}$$

The degree of model lightweight is measured by the number of parameters (nParam) and the number of floating-point operations (FLOPs). The lower the FLOPs, the smaller the computational burden.

4.4 Overall Model Comparative Analysis

In the airport scene detection task, the performance comparison between the improved YOLO11 model and the traditional YOLO11 model is clearly presented through three curve charts, and the number of training rounds is 200.

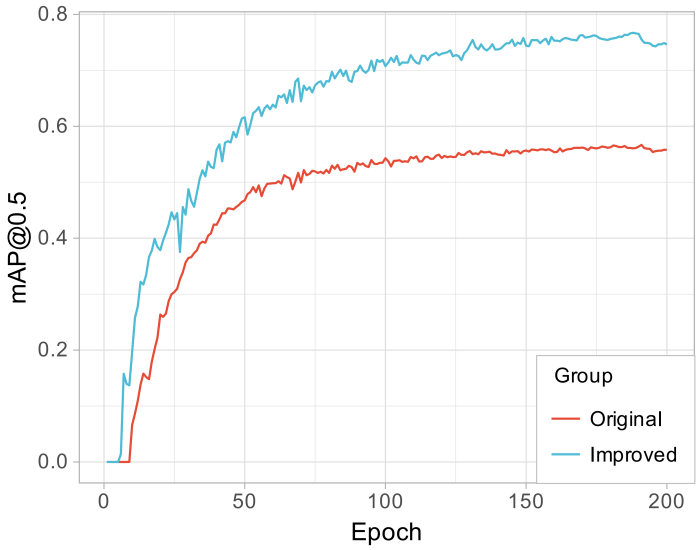


Fig. 4 Total loss

Figure 4 compares the total loss curves of the original YOLO11 model and the improved YOLO11: At the beginning of training, the loss values of both decrease rapidly, reflecting the efficient learning characteristics of the model. In the first 25 rounds, the loss of the improved model decreases faster, and the parameter adaptation is more rapid. However, the loss value of the improved model is always slightly higher than that of the original model in the middle and late stages of training. On the one hand, the improved model has a higher complexity, resulting in a slightly higher absolute value of the loss; on the other hand, the results of subsequent ablation experiments and comparative experiments show that the improved model has higher accuracy and greatly improved generalization ability in actual detection tasks. This reflects the effectiveness of the improvement strategy in balancing fitting and generalization.

Figure 5 compares the mAP@0.5 index of the original YOLO11 model and the improved model: In the early stage of training, the improved model learns faster. In

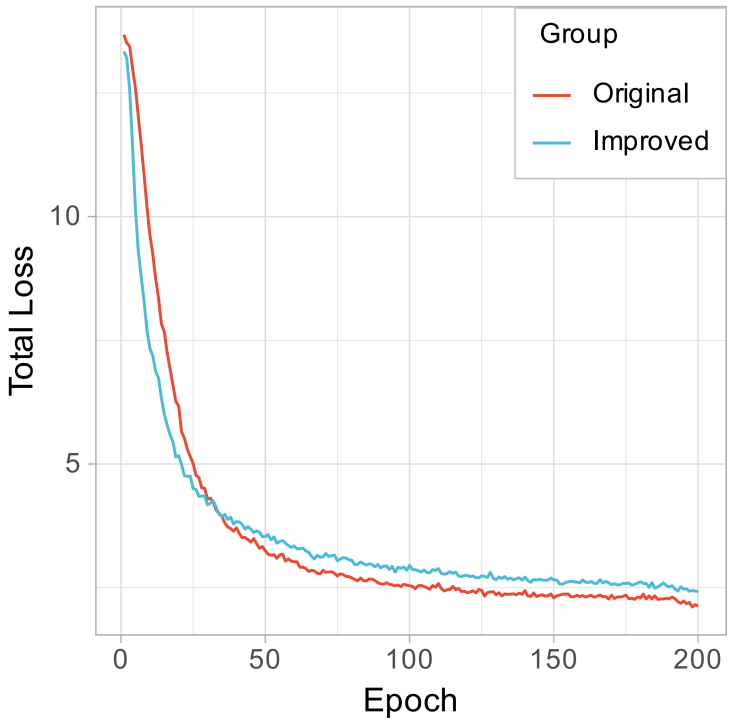


Fig. 5 mAP@0.5

the first 100 rounds, the mAP@0.5 value quickly rises to about 0.71, significantly higher than 0.53 of the original model, indicating that it can learn effective features faster; during the 100 - 200 rounds of training, the upward trend of the curve slows down, but the improved model always maintains a higher growth rate and reaches a higher index, verifying the continuous effectiveness of the improvement strategy; after about 200 rounds of stable training, the mAP@0.5 of the improved model stabilizes at about 0.76, a 20% increase compared to 0.56 of the traditional model. The improved model has achieved significant improvements in feature extraction and classification accuracy. Especially when IoU = 0.5, the matching degree between the detection box and the true bounding box has been significantly improved, achieving an efficient balance between detection rate and accuracy, and significantly enhancing the recognition ability of complex airport scenes.

Figure 6 shows that the improved YOLO11 model achieves higher mAP@0.5-0.95 values than the original model throughout training. It exhibits faster performance improvement in the early training stage and finally reaches a stable mAP@0.5-0.95 of about 0.56, which is 13% higher than the baseline result of 0.43. The improved model

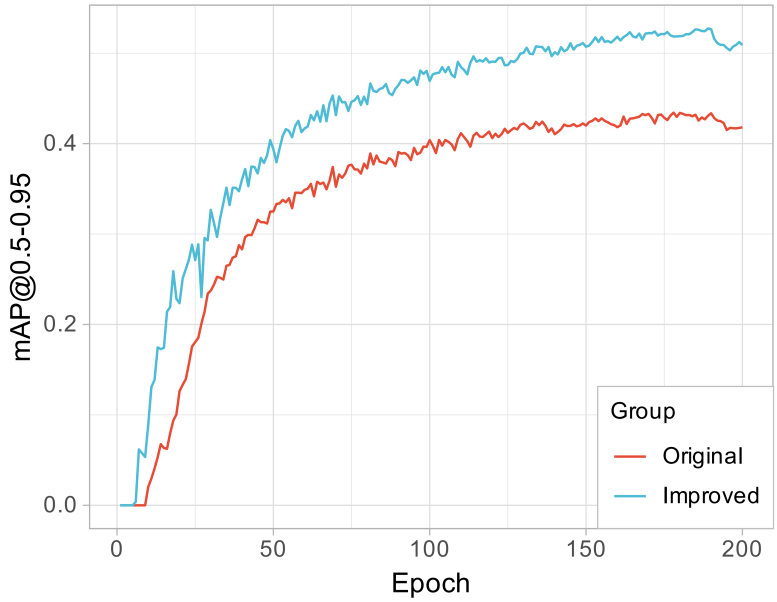


Fig. 6 mAP@0.5-0.95

still maintains obvious advantages at high IoU thresholds, verifying its strong ability to locate target boundaries accurately for airport-scene positioning tasks.

Overall, in the airport scene object detection, the improved YOLO11 has achieved significant improvements in both mAP@0.5 and mAP@0.5-0.95. Analyzing in combination with the total loss curve, it verifies the advantages of the model in recognition accuracy, training generalization ability and fine positioning, fully demonstrating the effectiveness and application value of the proposed improvement, and providing strong support for the efficient service of the airport scene recognition system.

4.5 Ablation Experiment

To verify the effectiveness of each improved module for the system, a set of ablation experiments was designed in this paper. All experiments were conducted under the same parameter configuration to ensure the comparability of the results. The specific experimental settings are as follows:

- Baseline model: YOLO11 was used as the baseline model, and its original performance on the target dataset was recorded.
- The second group of experiments: The MOAM attention mechanism was added to the detection head part of YOLO11.
- The third group of experiments: FDM was introduced into the C3k2 module and all C3k2 modules in the network were replaced with C3k2_FDM.

Table 1 Ablation experiment results

Model	Category	Precision	Recall	mAP@0.5	mAP@0.5-0.95
Baseline	all	0.774	0.528	0.563	0.435
	airplane	0.974	0.952	0.974	0.882
	person	0.486	0.0048	0.0111	0.00236
	truck	0.861	0.628	0.704	0.42
YOLO11 + MOAM	all	0.678	0.609	0.621	0.471
	airplane	0.962	0.97	0.979	0.896
	person	0.323	0.119	0.107	0.0325
	truck	0.748	0.738	0.776	0.484
YOLO11 + MOAM + FDM	all	0.849	0.76	0.799	0.56
	airplane	0.965	0.97	0.981	0.907
	person	0.761	0.475	0.55	0.225
	truck	0.82	0.836	0.865	0.548

Through these ablation studies, this paper analyzed the impact of each improvement on the model performance item by item, thus verifying the effectiveness of the proposed method in improving the detection accuracy and robustness. The experimental results are shown in Table 1.

The results of the ablation experiments show that each improvement measure has brought significant performance gains in the airport scene surveillance task. The overall precision rate of the original YOLO11 baseline model is 77.4%, the recall rate is 52.8%, and mAP@0.5 and mAP@0.5-0.95 are 56.3% and 43.5% respectively. The person class is almost impossible to be effectively detected, with a recall rate of only 0.48% and mAP@0.5 as low as 1.11%. In the second group of experiments (A), the MOAM attention mechanism was added to the baseline model. The experimental results show that the overall recall rate of the model is increased to 60.9%, mAP@0.5 is increased to 62.1%, and mAP@0.5-0.95 is increased to 47.1%; especially the detection performance of the person class is significantly improved, the recall rate is increased to 11.9%, mAP@0.5 is increased to 10.7%, and at the same time, the detection effects of the airplane and truck classes are also slightly enhanced. In the third group of experiments (A+B), on the basis of A, the C3k2 module was further replaced with the C3k2_FDM module, and the model performance was more comprehensively improved: the overall precision rate reached 84.9%, the recall rate reached 76%, mAP@0.5 reached 79.9%, and mAP@0.5-0.95 reached 56%. Among them, the detection effect of the person class is improved most significantly, the precision rate is increased to 76.1%, the recall rate is increased to 47.5%, and mAP@0.5 is increased to 55%. At the same time, the recall rate of the truck class is greatly increased to 83.6%, mAP@0.5 jumps to 86.5%, and mAP@0.5-0.95 is increased to 54.8%, showing stronger robustness to vehicle targets; all indicators of the airplane class reach the best, with P, R, mAP@0.5, and mAP@0.5-0.95 stabilizing at 96.5%, 97.0%, 98.1%, and 90.7% respectively. The above results fully prove the effectiveness of the combined improvement of MOAM and FDM.

In summary, the improved YOLO11 shows a significant performance leap in airport surface surveillance tasks. Compared with the baseline model, the overall precision, recall, mAP@0.5, and mAP@0.5-0.95 are increased by 7.5, 23.2, 23.6, and 12.5 percentage points respectively, reflecting the comprehensive optimization of the model in detection accuracy and robustness. In terms of sub-targets, the performance gain of the person class is the most prominent, with the recall rate jumping from 0.48% to 47.5%, and mAP@0.5 increasing from 1.11% to 55%. The detection robustness of the airplane and truck categories is also significantly enhanced. Among them, the mAP@0.5-0.95 of the airplane increases from 88.2% to 90.7%, and the mAP@0.5 of the truck increases from 70.4% to 86.5%. The above indicators fully demonstrate that the synergistic effect of the MOAM and C3k2_FDM mechanisms can effectively address the core challenges such as missed detection of small targets, target occlusion, and complex background interference in surface surveillance tasks, and have good practical application value.

4.6 Comparative Experiments

To further verify the effectiveness of AMFD-YOLO in airport surface surveillance tasks, this paper selects multiple YOLO series models for comparative experiments, and the experimental results are shown in Table 2.

From the perspective of overall performance, the precision of the improved model reaches 84.9%, significantly higher than 72.1% of YOLOv9t and 68.8% of YOLO26. In terms of recall, the improved model is increased to 76%, far ahead of YOLOv5 (54.3%) and YOLO12 (53.9%), achieving the coordinated optimization of precision and recall. In the comprehensive index mAP@0.5, the improved model reaches 79.9%, while YOLOv5, YOLOv6, and YOLO26 are only 56.6%, 57.1%, and 59.3% respectively. Under the more stringent mAP@0.5-0.95 standard, the improved model is also significantly better than YOLOv5 (44.1%) and YOLOv8 (45.4%) with 56%, fully verifying the comprehensive ability of the improved model in airport surface target detection tasks.

In the airport scene surveillance task, the target scale of the person category is the smallest and the detection difficulty is the highest. Most comparison models perform poorly on such targets, while the improved YOLO11 shows obvious advantages. Its precision rate reaches 76.1%, far exceeding YOLOv9t (36.1%) and YOLO26 (43.4%). The recall rate jumps to 47.5%, while the recall rates of YOLOv5 and YOLOv12 for this category are both 0%, and YOLOv10n is only 0.168%. In terms of accuracy, the mAP@0.5 of the person category of the improved model reaches 55%, while those of YOLOv6, YOLOv8, and YOLOv9t are only 1.71%, 5.64%, and 1.15% respectively. The mAP@0.5-0.95 of the improved model reaches 22.5%, while those of YOLOv5, YOLOv12, and YOLO26 are only 0%, 0.164%, and 2.54% respectively. The above data fully demonstrate that the improved model effectively overcomes the problem of missed detection of personnel targets and can more reliably capture the position information of the scene personnel. In addition to personnel targets, the detection robustness of the improved YOLO11 for two common airport targets, airplane and truck, is also significantly enhanced. The mAP@0.5-0.95 of the airplane category reaches 90.7%, better than YOLOv10n (87.7%) and YOLO26 (87.9%); the mAP@0.5 of the truck category reaches 86.5%, significantly higher than YOLOv5 (71.9%), YOLOv10n (69.3%),

Table 2 Comparative experiment results

Model	Category	Precision	Recall	mAP@0.5	mAP@0.5:0.95
YOLOv5	all	0.944	0.543	0.566	0.441
	airplane	0.982	0.963	0.980	0.889
	person	1.000	0.000	0.000	0.000
	truck	0.850	0.665	0.719	0.434
YOLOv6	all	0.949	0.544	0.571	0.451
	airplane	0.983	0.966	0.982	0.902
	person	1.000	0.00647	0.0171	0.00521
	truck	0.865	0.660	0.715	0.446
YOLOv8	all	0.753	0.583	0.599	0.454
	airplane	0.973	0.963	0.982	0.892
	person	0.432	0.05700	0.0564	0.0143
	truck	0.855	0.729	0.758	0.457
YOLOv9t	all	0.721	0.557	0.570	0.448
	airplane	0.971	0.966	0.973	0.892
	person	0.361	0.00671	0.0115	0.00320
	truck	0.830	0.697	0.724	0.450
YOLOv10n	all	0.896	0.526	0.559	0.435
	airplane	0.930	0.949	0.973	0.877
	person	0.978	0.00168	0.0114	0.00407
	truck	0.781	0.628	0.693	0.425
YOLO11	all	0.774	0.528	0.563	0.435
	airplane	0.974	0.952	0.974	0.882
	person	0.486	0.00480	0.0111	0.00236
	truck	0.861	0.628	0.704	0.420
YOLO12	all	0.947	0.539	0.566	0.441
	airplane	0.975	0.970	0.982	0.893
	person	1.000	0.00000	0.00656	0.00164
	truck	0.868	0.645	0.709	0.428
YOLO26	all	0.688	0.558	0.593	0.435
	airplane	0.926	0.945	0.973	0.879
	person	0.434	0.07550	0.0924	0.02540
	truck	0.705	0.653	0.712	0.399
AMFD-YOLO	all	0.849	0.760	0.799	0.560
	airplane	0.965	0.970	0.981	0.907
	person	0.761	0.475	0.550	0.225
	truck	0.820	0.836	0.865	0.548

and YOLO26 (71.2%), and its mAP@0.5-0.95 reaches 54.8%, also leading YOLOv5 (43.4%), YOLOv8 (45.7%), and YOLOv10n (42.5%). This shows that the improved model further enhances the anti-interference ability and overall robustness in complex scenarios while ensuring high recognition accuracy.

4.7 Multi-Model Scenario Application Comparison



Fig. 7 Comparison Results: (a) Original Image, (b) YOLOv5, (c) YOLOv6, (d) YOLOv8, and (e) YOLOv9

To visually demonstrate the effectiveness of the improved YOLO model in airport scene object detection, five groups of typical scenario images were selected: apron refueling scenario, nighttime apron scenario, runway takeoff scenario, taxiway guidance scenario, and rainy-day taxiway scenario. The algorithm models include YOLOv5, YOLOv6, YOLOv8, YOLOv9, YOLOv10, YOLO11, YOLOv12, YOLO26, and the improved YOLO11. The results are shown in Figures 10 and 11. The detection results of each algorithm are output in the form of bounding boxes with the target category and confidence. Through analysis, it can be found that in the first four groups of typical scenarios, all models effectively detected aircraft and vehicle targets, but only the improved YOLO11 model accurately detected tiny human targets. In the rainy-day taxiway scenario, YOLOv8 detected one person, while the improved model achieved accurate detection of two people. In the apron refueling scenario, the YOLOv6, v9, v10, 11 models had varying degrees of false detection and over-detection. In the nighttime scenario, the YOLOv9 and YOLO26 models also over-detected a vehicle, and YOLOv10 and YOLOv12 misdetected the airbridge as a truck in the taxiway guidance scenario.



Fig. 8 Comparison Results: (a) YOLOv10, (b) YOLO11, (c) YOLOv12, (d) YOLO26, and (e) Improved YOLO11

The above comparative analysis shows that through a series of optimizations, the improved YOLO11 model effectively enhances the object detection ability of the model under various complex scenario conditions. In contrast, other models such as YOLOv9, YOLOv10, and YOLO11 still have problems of over-detection, missed detection, and false detection in complex scenarios. In addition, the improved model also shows obvious advantages in terms of confidence, bounding box accuracy, and adaptability to complex environments, thus demonstrating its superior reliability and effectiveness in practical applications.

5 Discussion

In this paper, aiming at the problem of small object detection in airport scene surveillance, through the introduction of the MOAM multi-scale attention mechanism, FDM dynamic activation module, and P2 detection layer, the collaborative optimization of feature extraction and target localization is achieved. The P2 layer retains the shallow detail features, compensating for the information loss of small objects in the downsampling process; the MOAM module and C3k2_FDM enhance the model's adaptability to occluded targets and complex environments from the perspectives of spatial attention and dynamic non-linear transformation, respectively.

Although the improved model performs well on the ASS1 dataset, there are still two deficiencies:

First, the limitations at the data level. Although the ASS1 dataset covers various weather and lighting conditions, due to the limitations of the airport scenario, the image acquisition angles and heights are relatively fixed, there are insufficient samples

of extreme weather, and the actual operating conditions of the airport are not fully covered. Subsequently, we plan to cooperate with airport operation units to collect more diverse surface surveillance data and build an evaluation set covering a more comprehensive range of environmental conditions.

Second, the challenges at the engineering deployment level. Current research focuses on optimizing the detection performance of static images and does not fully consider the real-time requirements and computational resource constraints in actual surveillance systems. The P2 detection layer increases the computational load while improving the accuracy, and its inference efficiency on edge devices needs to be further verified. In addition, the detection results are not deeply integrated with the airport surface surveillance system, lacking support for functions such as continuous target tracking, trajectory prediction, and abnormal behavior recognition. Future work will further optimize the network structure in combination with lightweight technologies to adapt to airborne or edge deployment environments, and at the same time introduce a multi-target tracking module to build an intelligent surface surveillance solution with real-time positioning and continuous monitoring capabilities.

6 Conclusion

This paper focuses on the small target detection problem in airport surface surveillance and designs AMFD-YOLO, a target detection algorithm improved based on YOLO11. Through the analysis of the characteristics of airport surface images, it is found that there are challenges such as dense small targets, complex and changeable environments, and easy occlusion of targets in this scenario. To address the above challenges, this paper optimizes YOLO11 from three aspects: introducing the MOAM multi-scale attention mechanism to enhance the feature expression of occluded targets; using the FDM dynamic activation function to construct the C3k2_FDM module to achieve feature adaptive modulation and improve the adaptability of the model under different lighting and weather conditions; adding a P2 small target detection layer to retain the detailed information of the shallow layer and enhance the perception ability of tiny targets.

The experiments were carried out on the publicly available ASS1 dataset, in which the proportion of small targets reached 39.42%. The improved model achieved 84.9%, 76%, 79.9% and 56% in the four indicators of precision, recall, mAP@0.5 and mAP@0.5-0.95, respectively, which were 7.5, 23.2, 23.6 and 12.5 percentage points higher than the baseline YOLO11 model. The ablation experiments showed that the introduction of the C3k2_FDM module and the P2 detection layer contributed significantly to the improvement of the model performance, and their synergistic effect led to an order-of-magnitude improvement in the detection accuracy of the personnel category. Compared with mainstream algorithms such as YOLOv5, YOLOv8, YOLOv9t, YOLOv10n, YOLO11, YOLO12 and YOLO26, the method in this paper showed the best overall detection accuracy. Especially for small target categories such as personnel, the mAP@0.5 reached 55.0%, while the comparison methods did not exceed 10%, fully verifying the superiority of the improved algorithm in the small target detection task.

To sum up, AMFD-YOLO in this paper can effectively address the problem of small target detection in airport surface surveillance, maintain good detection performance under complex environmental conditions, and provide an effective technical solution for high-precision real-time surveillance of airport surfaces. In the future, the surface surveillance data will be improved, and research will be carried out in the directions of model lightweight and multi-object tracking, etc., to further enhance the practicality and robustness of the algorithm and promote the intelligent development of airport surface surveillance systems.

Supplementary information. No supplementary material is available for this article.

Acknowledgements. The authors thank the contributors of the ASS dataset for providing the publicly available benchmark dataset.

Declarations

Funding

This research was funded by the Henan Provincial Key Science and Technology Research Project, Grant No. 262102240002, the Regional Collaborative Innovation Special Project of the Xinjiang Uygur Autonomous Region (Science and Technology Assistance to Xinjiang Program), Grant No. 2026E2014, and the Major Science and Technology Project of Sichuan Province (Challenge-Based Bidding Initiative), Grant No. 2024ZDZX0046.

Conflict of interest

The authors declare no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The ASS1 dataset used in this study is publicly available at https://github.com/rookie257/rookie257.github.io.

Materials availability

Not applicable.

Code availability

The source code used in this study is available from the corresponding author upon reasonable request.

Author contribution

Conceptualization, T.L. and W.P.; methodology, F.Y. and H.W.; software, F.Y.; validation, H.W., T.L. and W.P.; formal analysis, T.L. and F.Y.; writing—original draft preparation, F.Y.; writing—review and editing, T.L. and H.W.; supervision, W.P. All authors have read and agreed to the published version of the manuscript.

References

[1] Fu, D., Wang, X., Wu, F., Han, S., Liang, B., Li, W., Yang, K.: Multi-task learning for airport surface surveillance: A review. Expert Systems **43**(4) (2026) https://doi.org/10.1111/exsy.70227

[2] Talbi, A., Abdellaoui, M., Pons-Prats, J., Kuljanin, J.: An overview on computer vision analysis in the airport applications. IEEE access (2025)

[3] Thai, P., Alam, S., Lilith, N.: Airside surveillance by computer vision in low-visibility and low-fidelity environment. Journal of Air Transportation **32**(4), 210–217 (2024) https://doi.org/10.2514/1.d0410

[4] Munyer, T., Brinkman, D., Zhong, X., Huang, C., Konstantzos, I.: Foreign object debris detection for airport pavement images based on self-supervised localization and vision transformer. CSCI, 1388–1394 (2022) https://doi.org/10.1109/csci58124.2022.00249

[5] Feng, Q., Xu, X., Wang, Z.: Deep learning-based small object detection: A survey. Mathematical Biosciences and Engineering **20**(4), 6551–6590 (2023)

[6] Verma, T., Singh, J., Bhartari, Y., Jarwal, R., Singh, S., Singh, S.: Soar: Advancements in small body object detection for aerial imagery using state space models and programmable gradients. arXiv preprint arXiv:2405.01699 (2024)

[7] Zhou, W., Cai, C., Li, C., Xu, H., Shi, H.: Ad-yolo: A real-time yolo network with swin transformer and attention mechanism for airport scene detection. IEEE Transactions on Instrumentation and Measurement, 73 (2024)

[8] Xu, Y., Liu, Y., Shi, K., Wang, X., Li, Y., Chen, J.: An airport apron ground service surveillance algorithm based on improved yolo network. Electronic Research Archive **32**(5), 3569 (2024)

[9] Zhou, W., Cai, C., Srigrarom, S., Wang, P., Cui, Z., Li, C.: Adh-yolo: a small object detection based on improved yolov8 for airport scene images in hazy weather: W. zhou et al. The Journal of Supercomputing **81**(3), 505 (2025)

[10] Xu, Y., Liu, Y., Shi, K., Wang, X., Li, Y., Chen, J.: An airport apron ground service surveillance algorithm based on improved yolo network. Electronic Research Archive **32**(5), 3569–3587 (2024)

[11] Mo, Y., Wang, L., Hong, W., Chu, C., Li, P., Xia, H.: Small-scale foreign object debris detection using deep learning and dual light modes. Applied Sciencesbi **14**(5), 2162 (2024)

[12] Yan, F., Guo, J., Wang, H.: Multiscale contextual fusion for robust airport surveillance. Mathematics **13**(20), 3350–3350 (2025) https://doi.org/10.3390/math13203350

[13] Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)

[14] Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence **39**(6), 1137–1149 (2016)

[15] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)

[16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European Conference on Computer Vision, pp. 21–37 (2016). Springer

[17] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J.: Detrs beat yolos on real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16965–16974 (2024)

[18] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)

[19] Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)

[20] Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)

[21] Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection (2020)

[22] Jocher, G.: Ultralytics YOLOv5. https://doi.org/10.5281/zenodo.3908559 . https://github.com/ultralytics/yolov5

[23] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., et al.: Yolov6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976 (2022)

[24] Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M.: 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 (2023)

[25] Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics

[26] Wang, C.-Y., Yeh, I.-H., Liao, H.-Y.M.: Yolov9: Learning what you want to learn using programmable gradient information. In: Computer Vision – ECCV 2024, pp. 1–21 (2024). Springer Nature Switzerland

[27] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G.: Yolov10: Real-time end-to-end object detection. Advances in neural information processing systems **37**, 107984–108011 (2024)

[28] Jocher, G., Qiu, J.: Ultralytics YOLO11. https://github.com/ultralytics/ultralytics

[29] Tian, Y., Ye, Q., Doermann, D.: Yolov12: Attention-centric real-time object detectors. In: Advances in Neural Information Processing Systems, vol. 38, pp. 78433–78457 (2025)

[30] Lei, M., Li, S., Wu, Y., Hu, H., Zhou, Y., Zheng, X., Ding, G., Du, S., Wu, Z., Gao, Y.: Yolov13: Real-time object detection with hypergraph-enhanced adaptive visual perception. arXiv preprint arXiv:2506.17733 (2025)

[31] Jocher, G., Qiu, J.: Ultralytics YOLO26. https://github.com/ultralytics/ultralytics

[32] Zhu, X., Liang, B., Fu, W. Fan Li: Airport small object detection based on feature enhancement. IET image processing **16**(11), 2863–2874 (2022)

[33] He, D., Sui, Y., Zhu, P., Yan, S., Shu, Y., Deng, L.: Mobile detecting methods for small objects on airport surfaces. 2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 506–511 (2023) https://doi.org/10.1109/iccasit58768.2023.10351756

[34] Ding, M., Zhou, W., Xu, Y., Xu, Y.: Two-stage framework for specialty vehicles detection and classification: Toward intelligent visual surveillance of airport surface. IEEE Trans. Aerosp. Electron. Syst. **60**(2), 1912–1923 (2023) https://doi.org/10.1109/taes.2023.3342797

[35] Zhao, Y., Kong, Z., Hu, Y., Wen, Z., Wang, S.: Airport-yolo: A fixed-camera solution for airport object detection. IET Image Processing (2026) https://doi.org/10.1049/ipr2.70387

[36] Zhou, W., Cai, C., Zheng, L., Li, C., Zeng, D.: Assd-yolo: a small object detection method based on improved yolov7 for airport surface surveillance. Multimedia Tools and Applications **83**(18), 55527–55548 (2024)

[37] Zhou, W., Cai, C., Wu, K., Li, C., Gao, B.: Las-yolo: a lightweight detection method based on yolov7 for small objects in airport surveillance. Journal of Supercomputing **80**(15) (2024)

[38] He, Z., He, Y., Lv, Y.: Dt-yolo: An improved object detection algorithm for key components of aircraft and staff in airport scenes based on yolov5. Sensors **25**(6), 1705 (2025)

[39] Li, Y., Fang, Y., Zhou, S., Zhang, Y., Ribeiro, N.A.: Robust small-object detection in aerial surveillance via integrated multi-scale probabilistic framework. Mathematics **13**(14), 2303 (2025)

[40] Lu, Y.: Cr-mask rcnn: An improved mask rcnn method for airport runway detection and segmentation in remote sensing images. Sensors **25** (2025)