I.J. Intelligent Systems and Applications, **2025, 2, 1-13**
Published Online on April 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2025.02.01

An Optimized YOLOv8-based Method for Airport Bird Detection: Incorporating ECA Attention, MBC3 Module, and SF-PAN for Improved Accuracy and Speed

Zia Ur Rehman*
Yangzhou University/ College of Information Engineering (College of Artificial Intelligence), Jiangsu, China
E-mail: ziaurrahman4h@gmail.com
ORCID iD: https://orcid.org/0009-0008-0638-7214
*Corresponding author

Abdul Ghafar
Yangzhou University/ College of Information Engineering (College of Artificial Intelligence), Jiangsu, China
E-mail: ayanghafar783@gmail.com
ORCID iD: https://orcid.org/0009-0001-5366-9133

Ahmad Syed
Yanshan University/ School of Electrical Engineering, Qinhuangdao, China
E-mail: asd@ysu.edu.cn
ORCID iD: https://orcid.org/ 0000-0002-1651-7288

Abu Tayab
Yanshan University/ school of Mechanical engineering, Qinhuangdao, China
E-mail: tayababu2591@gmail.com
ORCID iD: https://orcid.org/ 0009-0004-0613-8817

Md. Golam Rabbi
Yangzhou University/ College of Information Engineering (College of Artificial Intelligence), Jiangsu, China
E-mail: rabbi66@qq.com
ORCID iD: https://orcid.org/0009-0002-0934-9471

Received: 23 October 2024; Revised: 07 December 2024; Accepted: 05 February 2025; Published: 08 April 2025

Abstract: This research investigation utilizes deep learning object detection algorithms to achieve accurate recognition of birds near airports, thereby addressing the limitations of manual bird detection at airports, including low accuracy slow speed, and the high cost of radar detection. The ultimate goal is to ensure the safe operation of civil aviation. The following are the primary enhancements: First, an ECA (Efficient Channel Attention) attention mechanism was added to the Neck to enhance the network's emphasis on important characteristics. This resulted in a notable improvement in accuracy while only changing a few parameters. Second, by adding branches with various receptive fields, the MBC3 (Muti Branch C3) module was created to improve the expressiveness of the model. Thirdly, the model's right width and depth parameters will be chosen by investigating the effects of various network widths and depths on model performance. Fourth, to solve the problem of feature loss in recognizing tiny bird targets, the SF-PAN (Shallow Feature - Path Aggregation Network) structure was proposed. The model was evaluated using metrics such as mAP@50, FPS, precision, recall, and computational complexity on a test set derived from the dataset. Results show that the enhanced YOLOv8 achieves a mAP@50 of 83.1% and a speed of 31 FPS, a 2.5% improvement in accuracy and a 7 FPS increase over the baseline YOLOv8, while reducing parameters and weight size by approximately 48%. Comparative experiments further validate the model’s superiority over existing algorithms in terms of accuracy and resource efficiency. This upgraded YOLOv8 provides a novel, real-time solution for precise bird detection in challenging airport environments, ensuring safer civil aviation operations.

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

Index Terms: Airport Bird Detection, Bird Strike Prevention, Attention Mechanism, Multi-branch Convolution, Feature Fusion.

1. Introduction

Bird Strike usually refers to a collision between an aircraft and nearby birds during takeoff, flight, or landing, resulting in a flight safety accident [1]. The 2009 Hudson River emergency landing in New York was one of the typical civil aviation bird strike accidents. Between 1990 and 2022, the Federal Aviation Administration's Wildlife Impact Database recorded a total of 276846 bird strike event reports, and the database also showed an increasing trend in the number of bird strike events year by year. According to relevant statistics, 90% of bird strike incidents occur in airports and their surrounding airspace, and 76% occur below 60 meters. Therefore, airports and their surrounding airspace are key to preventing bird strike accidents.

Before driving away birds, the first step is accurately detecting bird targets in the airport airspace. There are two common methods: manual observation and radar detection. For small and medium-sized airports, the cost of radar detection is relatively high, and manual observation has disadvantages such as poor real-time performance and low accuracy. Deep learning object detection is a branch of computer vision that refers to locating and recognizing the category of an object in an image [2]. Two types of object detection algorithms are based on the detection stage: the two-stage and the single-stage methods. The two-stage method has a long training time, poor real-time performance, and poor accuracy in detecting small targets. The single-stage method is an end-to-end detection algorithm that maintains high detection speed without sacrificing accuracy. Representative algorithms include SSD (Single Shot MultiBox Detector) and YOLO (You Only Look Once) series [3-8]. The use of image-based bird target detection technology in small and medium-sized airports has the advantages of low cost, high accuracy, and fast speed. Directly applying object detection algorithms to bird detection tasks may result in low detection accuracy and slow speed. Therefore, it is necessary to make appropriate improvements to the algorithm to make it more suitable for bird detection tasks.

The main problems in small object detection tasks are as follows: firstly, the target occupies fewer pixels, so the detailed information contained in the target is not rich enough, which makes it difficult for the network to extract features, and when the image is downsampled multiple times, it is easy to cause the loss of target information. Secondly, there is no attention preference in the original YOLOv8 network, and the network cannot effectively allocate computing resources, ultimately resulting in a decrease in network detection performance. Thirdly, feature fusion methods can improve the accuracy of network object detection, while YOLOv8 network has poor feature fusion capability. In the improvement of object detection algorithms, Xia Tao et al. embedded a gate control channel attention mechanism in the feature extraction network to promote the relationship between feature channels [9]. The DenseNet (Densey Connected Convolutional Networks) model combined with a self-attention mechanism to better extract interference signal features, thereby avoiding complex convolution operations thus avoiding complex convolution operations [10]. In the Path Aggregation Network (PANet) applied structure in the SSD algorithm, fully utilizing the relationships between various feature layers to obtain richer semantic information of small targets, improve the network's feature fusion ability, and ultimately enhance the performance of small target detection [11]. The authors increased the coverage level of the FPN structure to obtain more complete small target information and improve the multi-scale detection performance of the model [12]. Bird strikes are a major concern for aviation safety, as collisions between birds and aircraft can cause significant damage, delays, and even accidents. Detecting birds near airports is challenging due to their unpredictable movement and environmental factors like weather and lighting, which complicate traditional detection methods such as radar or visual monitoring. Bird strikes result in thousands of incidents annually, leading to high costs for airlines and posing safety risks to passengers and crew. These issues highlight the urgent need for improved bird detection systems to reduce both the financial and safety impacts on the aviation industry. This research aims to address this problem by developing more effective bird detection technologies for airports.

This study selected the YOLOv8s model with moderate parameter quantity, high detection speed, and accuracy for bird detection research. Firstly, Efficient Channel Attention (ECA) [13] is introduced into Neck to make the model pay more attention to valuable channels, increase the weights of these channels, and suppress the weights of other channels, thereby improving the accuracy of object detection. Secondly, to address the issue of insufficient expressive power and poor small object detection performance in object detection models, a Multi Branch C3 (MBC3) module was designed, which integrates multiple parallel convolutional branch structures with different receptive fields to enhance the model's expressive power and improve small object detection performance. The goal of network lightweight was further achieved by adjusting the depth and width of the network. Finally, a Shallow Feature Path Aggregation Network (SF-PAN) was proposed to refuse the feature maps output by Backbone into the downsampling process of the Neck structure, improving detection performance. This algorithm can assist in detecting bird targets in the low-altitude airspace of airports, making up for the shortcomings of manual observation and radar detection, and ensuring the safe operation of aircraft.

Main Contributions of the Paper:

- Introduced Efficient Channel Attention (ECA) into YOLOv8s model's Neck to improve detection accuracy for bird targets in low-altitude airport airspace.

- Designed the Multi-Branch C3 (MBC3) module to integrate multiple convolutional branches with varying receptive fields, enhancing small object detection, especially for challenging targets like birds.
- Achieved a more efficient YOLOv8s model by adjusting its depth and width, optimizing detection speed and accuracy for real-time airport applications.
- Proposed a Shallow Feature Path Aggregation Network (SF-PAN) to enhance feature fusion during downsampling, improving detection performance for bird targets.
- Developed a cost-effective bird detection solution for low-altitude airport airspace, enhancing aviation safety by reducing reliance on manual observation and radar systems.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of existing bird detection algorithms and deep learning-based object detection methods. Section 3 outlines the methodology, detailing the enhancements made to the YOLOv8 model, including the integration of Efficient Channel Attention, the Multi-Branch C3 module, and the Shallow Feature Path Aggregation Network. Section 4 describes the experimental setup and evaluation criteria, followed by the presentation of bird detection experiment results conducted in low-altitude airport airspace. Section 5 discusses the findings, comparing the performance of the proposed algorithm with existing methods. Finally, Section 6 concludes the paper and identifies potential areas for future research.

2. Related Work

Several areas, such as traffic sign identification, facial recognition, and bird recognition, use image object detection [14]. Currently, medium- to large-sized target datasets have shown encouraging performance for deep learning-based object detection systems. Because there are fewer pixels in the image due to the tiny size of the bird target, the YOLOv8s has a harder time extracting features from the little target, which lowers detection performance. Two-stage algorithms have huge model sizes, complicated training, and sluggish frames per second (FPS) yet provide good accuracy in object identification applications. The YOLO algorithm, a single-stage deep learning-based object identification technique, was created to overcome these problems. The YOLO approach has minimal accuracy gain but much faster detection. In addition, the YOLO algorithm requires little memory during training, few parameters, and little time to execute. While the YOLO series algorithms are very good at identifying huge things, they are not as good at identifying little birds in airports. The main causes of this are the following: creating previous boxes, defining loss functions, matching negative and positive samples, issues in setting prior boxes, and unequal distribution of tiny object samples in training data [15].

Several techniques have been used to enhance YOLOv8s's tiny bird feature extraction capabilities. As an example, reference [16] presented the AF-FPN, a revised feature pyramid approach. This approach enhances the feature pyramid's representation capabilities by integrating the adaptive attention module (AAM) and feature enhancement module (FEM) to reduce information loss during feature map construction. A convolutional channel attention block was added by Reference [17] after each feature fusion to enhance mAP@50 detection for small flaws. A convolutional block attention module (CBAM) was added to the YOLOv7 model by Reference [18]. With this approach, feature responses across channels are adjusted by dynamically learning weights and modelling the interactions across the channels. This instructs the model to give rich information-rich features priority, hence improving the detection accuracy of tiny objects. A dense block was incorporated into the YOLOv2 network by Reference [19]. This addition improved the network's ability to extract features from small objects, which increased the detection precision.

According to reference [20], small object characteristics and spatial details tend to decrease over time or shrink over the network inference process. By combining the semantic information from higher-level feature maps with the fine-grained information from shallower-level feature maps, multiscale feature methods—also known as feature fusion—can help offset this problem by minimizing information loss and enhancing network accuracy. In order to accomplish feature fusion, Reference [21] presented the feature pyramid network (FPN), which consists of a top-down network, a bottom-up network, and lateral connections between them. By incorporating a second bottom-up network and lateral connections to form PANet, reference [22] built on the FPN and increased network accuracy and feature fusion efficiency even more.

The augmented feature fusion architecture (PB-FPN) described in Reference [23] was created from BiFPN and PANet and significantly improved the model's capacity to recognize tiny targets. Apart than adding methods for attention and using feature fusion, one may improve the detection accuracy by adjusting the model's loss function. In order to calculate bounding box regression loss, Reference [24] added a Focal DIoU loss to YOLOv3's loss function. This improved the algorithm's accuracy and convergence time.

CNNs have long dominated the field of computer vision because of their potent capacity to extract local characteristics. Nevertheless, vision transformers, or ViTs, have become a viable substitute in recent years. In some visual tasks, they have shown performance on par with or even better than CNNs. The authors of reference [25] successfully reduced the workload of radiologists by classifying brain cancers with a 98.7% classification accuracy using a ViT-based deep neural network. Using three datasets of traffic signs, Reference [25] examined the performance of seven CNNs and five ViTs; the findings showed that vision transformers are not competitively superior. Adaptive vision transformers (AdaViT) were described in Reference [26]. Their goal is to maximize the inference efficiency of vision transformers while limiting the accuracy reduction. AdaViT learns to decide to use rules for self-attention heads, patches, and transformer blocks independently for each input. Using two branches to analyse large- and small-scale image patches, reference [27] introduced the application of feature fusion algorithms to cross-attention multi-scale vision transformers (CrossViT) and developed a low-weight token fusion block based on cross-attention. The model's accuracy compares to or surpasses that of CNNs. The disadvantages of many ViT techniques include longer training times, more computation and model parameters, greater hardware requirements, and increased GPU memory use for model training. For instance, in order to keep the GPU memory needed for training from going above the GPU's memory limit, it was essential to decrease the number of channels in each layer throughout the real-time detection transformer (RT-DETR) training process [28].

The network's detection accuracy is increased by the aforementioned enhancement techniques, but they also have certain disadvantages. For example, the object detection network's parameters and computation are increased by some introduced advancements. Furthermore, it's possible that they aren't making the most of the greater detail information that the feature extraction network's little, low features carry. Our goal in this study is to address these issues as much as we can in order to improve the object detection network's overall performance.

3. Improvement Method

3.1. Improving the Model Structure of YOLOv8s

YOLOv8s consists of three parts: Backbone, Neck, and Head. The main function of Backbone is to extract image features; Neck uses the feature pyramid network PANet to process targets of different sizes, mainly fusing feature maps from multiple levels; Head is used to perform the final object detection, using feature maps extracted from the backbone network and Neck to predict the categories of bounding boxes and objects. The Head consists of three detection layers at different scales, each responsible for predicting targets of three different sizes. The detection at these three scales improves the accuracy of detecting small targets. The improved YOLOv8s object detection model is shown in Figure 1.

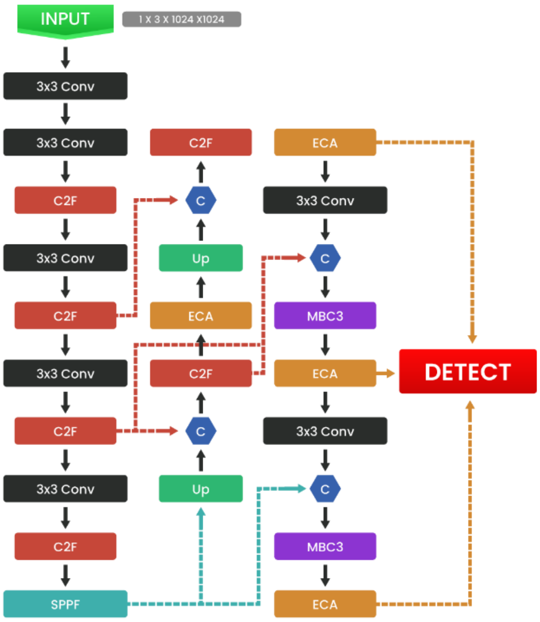


Fig.1. Model structure of improved YOLOv8s

This article mainly improves the Neck feature fusion part of YOLOv8s, including adding a 4-layer attention mechanism, using the proposed MBC3 module and SF-PAN structure. Compared with the original model, the improved YOLOv8s model has higher accuracy mAP@50 higher detection speed FPS, while reducing the number of parameters, weight size, and computational complexity of the model.

3.2. ECA Attention Mechanism

The channel attention mechanism can adaptively determine the objects that need attention and assign different weights to different channels [14]. In 2018 the authors proposed a Squeeze and Excitation Network (SENet) [29], where the SE module establishes correlations between different convolutional feature channels, making resource allocation more reasonable with only a few additional parameters. In 2018, the authors [16] proposed an ultra-lightweight channel attention module ECA module, which only contains k parameters (k<=9). The author found that avoiding dimensionality reduction and having appropriate channel interactions are crucial in SENet, so one-dimensional convolution was used to achieve local cross-channel interactions. Considering that the attention mechanism can make the network pay more attention to certain features of the input image, assign higher weights to that feature, and assign lower weights to other "unimportant" features, thus more efficiently processing the extracted features and achieving the goal of improving object detection performance. This study adds the efficient channel attention mechanism ECA to the Neck part of the YOLOv8s network so that the model can process the features extracted from the Backbone more efficiently. The structure of the ECA module is shown in Figure 2.

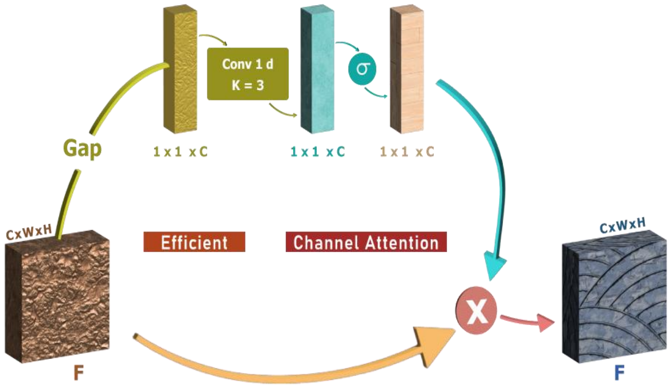


Fig.2. Module structure of ECA

In the ECA module, the input feature map is first subjected to global average pooling GAP, which changes the size from C × W × H (C represents the number of feature map channels, W represents the width of the feature map, and H represents the height of the feature map) to 1 × 1 × C. After converting to a one-dimensional vector, perform a one-dimensional convolution operation on the vector with a kernel size of K=3 to establish cross channel interaction between local channels. Subsequently, the output one-dimensional vector is activated through the sigmoid function. Finally, multiply the final output one-dimensional vector with the original feature map to obtain the channel-weighted feature map.

This study added an ECA module after each C2f module of Neck in YOLOv8s. Adding the ECA module improves the accuracy and speed of object detection in the network by adding a small number of parameters.

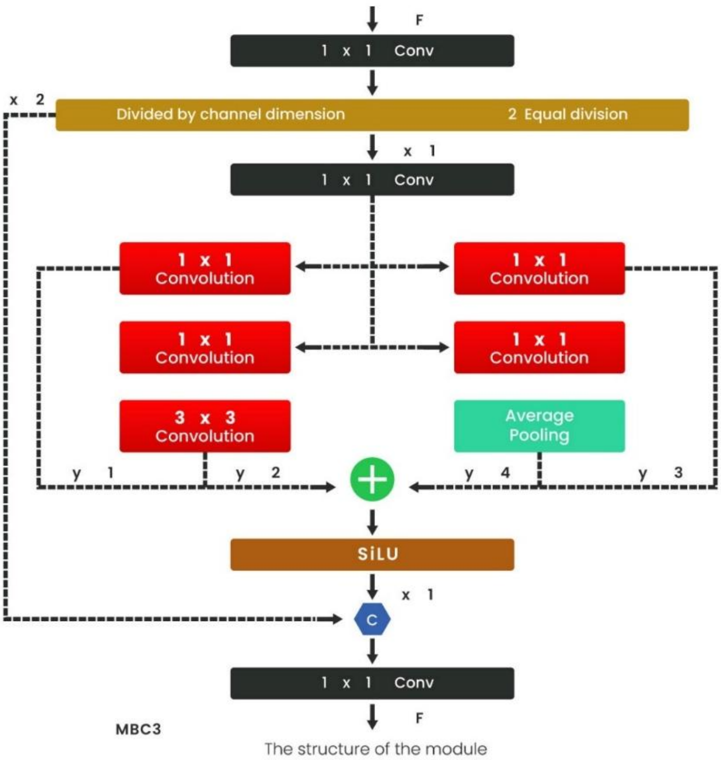


Fig.3. Module structure of MBC3

3.3. MBC3 Module

Among numerous neural network architectures, deeper network depth can significantly improve network detection performance, but there are fewer parallel branch structures in the network. Different convolutional branches have different receptive fields and feature extraction capabilities, which can extract different and more complex abstract information about bird targets. This study considers the addition of parallel multi-convolution branches and adds the feature maps extracted from the multi-branch structure to obtain an output feature map with richer feature information. The proposed module is named the MBC3 module. This module contains four parallel convolutional branch structures with different receptive fields, such as 1 × 1 convolution, 3 × 3 convolution, and 1 × 1 convolution + pooling layer. By applying the MBC3 module, the number of background predictions made by the network as flying birds has been reduced, and the accuracy of the network's prediction of small flying bird targets has been increased. The structure of the MBC3 module is shown in Figure 3.

In the MBC3 module, first, the input feature map F is processed through 1 1x1 convolutional layer (including 1x1 convolution operation + batch normalization) Batch Normalization (BN) + SiLU activation function, and then divide the output feature map into two parts, x1 and x2. Secondly, the x1 feature map is passed through a 1x1 convolutional layer, and its output feature map is parallelized through (1) 1x1 convolution operation; (2) 1x1 convolution operation, and 3x3 convolution operation; (3) 3 × 3 convolution operation; (4) 1x1 convolution operation and average pooling operation. Batch standardizes the output features of the four branches mentioned above, then adds them together and activates them using the SiLU activation function to output the feature map x1 '. Finally, concatenate the x1 'and x2 feature maps along the channel dimension, and use a 1x1 convolutional layer to obtain the output feature map F'.

The calculation formulas for the MBC3 module are shown in (1) ~ (3):

$$x1, x2 = \text{split}\left(f_{\text{Conv}}^{k=1}(F)\right) \tag{1}$$

$$\begin{aligned} x1' &= \text{SiLU}\left(f_{\text{conv}}^{k=1}\left(f_{\text{Conv}}^{k=1}(x1)\right) + f_{\text{conv}}^{k=3}\left(f_{\text{conv}}^{k=1}\left(f_{\text{Conv}}^{k=1}(x1)\right)\right)\right. \\ &\quad \left. + f_{\text{conv}}^{k=3}\left(f_{\text{Conv}}^{k=1}(x1)\right) + f_{\text{AvgPool}}\left(f_{\text{conv}}^{k=1}\left(f_{\text{Conv}}^{k=1}(x1)\right)\right)\right) \end{aligned} \tag{2}$$

$$F' = f_{\text{Conv}}^{k=1}\left(\text{Cat}\,(x2, x1')\right) \tag{3}$$

In the formula, () Split represents dividing the output feature map into 2 equal parts; $f_{\text{conv}}^{k=1}$ () represents a 1 × 1 monolayer; The convolution operation with $f_{\text{conv}}^{k=1}$() represents a 3x3 loop operation when k=3; SiLU() represents the number of SiLU cycles; Cat() represents concatenating feature maps according to channel dimensions.

Table 1. The influence of the position of MBC3 on model performance

method	mAP@50	FPS
case1	82.0%	28
case2	82.2%	30

3.4. SF-PAN Structure

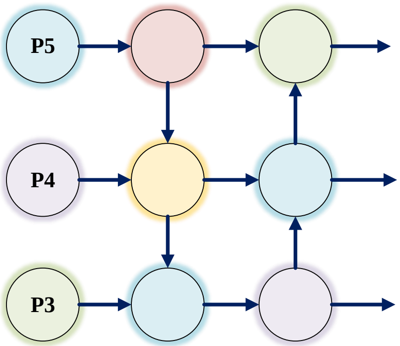
After undergoing multiple convolution operations, the network may result in the final output feature map losing small target information. Considering the Backbone area in the shallow position of the YOLOv8s network, which includes more bird orders the detailed information of the target, reusing this information will improve network performance Raising is helpful. To compensate for the network's loss during upsampling and downsampling processes the missing target information further increases the information of small bird targets in the image. This article proposes the SF-PAN structure, which is used in YOLOv8s during the second downsampling process in Neck, the data from Backbone was reused Shallow features. Different from a weighted bidirectional feature pyramid network the Bidirectional Feature Pyramid Network (BiFPN) is, the entire structure ultimately retains the network nodes with input 1, ensuring network connectivity. The depth of the network remains unchanged. The SF-PAN structure is shown in Figure 4 (b). In addition to the cross-sampling layer links mentioned above, SF-PAN also includes a weighted features Integration aspect. While integrating shallow and deep features, introduced learnable weights to learn different input features. P4 layer the output formula is shown in (4) - (5).

The SF-PAN structure reuses shallow information from the network, integrating shallow and deep feature maps not only enables information to bottom-up and top-down flow in the network but can also be more efficient and convenient lateral flow and cross sampling layer flow to help the network to be effective capture multi-scale features to improve network processing of different sizes and complexities the ability of the object. The introduction of SF-PAN structure increases the number of parameters by a small amount in terms of computational complexity, and the detection accuracy of the model has been increased.

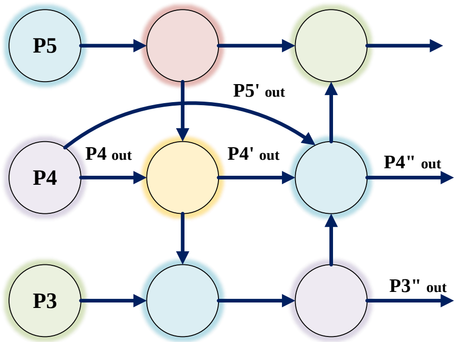
$$P4'_{\text{Out}} = \frac{\omega_1 \times P4_{\text{Out}} + \omega_2 \times P5'_{\text{Out}}}{\omega_1 + \omega_2 + \varepsilon} \tag{4}$$

$$P4''_{\text{Out}} = \frac{\omega_3 \times P4'_{\text{Out}} + \omega_4 \times P5'_{\text{Out}} + \omega_5 \times P3''_{\text{out}}}{\omega_3 + \omega_4 + \omega_5 + \varepsilon} \tag{5}$$

In the formula, ω represents the learned parameter; ε represents a number that is infinitely close to 0, preventing the formula from being divided by 0.



(a) Neck structure of YOLOv8s



(b) SF-PAN structure

Fig.4. The Neck structure of YOLOv8s and the SF-PAN structure proposed in this paper

4. Experiment and Results Analysis

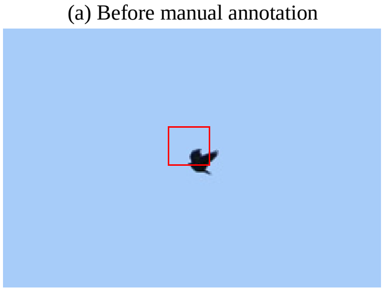
4.1. Dataset Production

The dataset of this experiment was trained, validated, and tested using partial AirBirds data published in reference [17]. The images in the AirBirds dataset are captured over four seasons of a year, including various bird species, different lighting conditions, and 13 weather conditions. Because the average annotation size of instances in this dataset is less than 10 pixels, this experiment selected 5000 images from AirBirds with the largest proportion of pixels occupied by the largest bird instance to form the dataset. In this way, these 5000 images include both larger bird instances and numerous smaller bird instances.

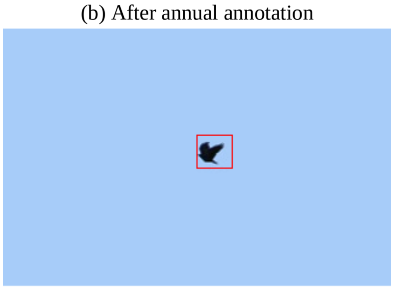
Data cleaning of 5000 datasets: (1) Manually re-label the annotation boxes with incorrect annotations (as shown in Figure 5 (a)), as shown in Figure 5 (b); (2) Expanding the dataset and introducing large instance bird images, the network went from being unable to detect large instances of birds well at the beginning (as shown in Figure 5 (c)) to being able to detect large instances of birds well (as shown in Figure 5 (d)).

(a) Before manual annotation

(b) After annual annotation



(c) Before expanding the dataset



(d) After expanding the dataset

Fig.5. Problems in the original dataset and modifications made to the dataset

The final dataset consists of 5354 images. Divide the dataset according to the ratio of the training set: validation set: and test set=8:1:1. The distribution of real box positions in the dataset and the size distribution.

4.2. Experimental Environment and Hyperparameter Configuration

The software and hardware environment configuration for this experiment is shown in Table 2.

Table 2. Software and hardware environment configuration

Name	Edition
CPU	Intel(R) Xeon(R) Silver 4110 CPU@2.10GHz
GPU	NVIDIA GeForce RTX 2080Ti 11GB
Memory	128GB
Operating System	Windows 10 Professional 22H2
Python	3.8.17
Pytorch	1.8.1
CUDA	10.2

The experimental hyperparameter configuration is: 300 training epochs, 8 batches, 8 workers, image size 1024 × 1024, initial learning rate 0.01, minimum learning rate 0.0001, optimizer SGD, momentum 0.937, and cosine annealing method is used to reduce the learning rate.

4.3. Evaluation Indicators

This experiment selects accuracy P (referring to the proportion of correctly predicted positive samples in the images recognized by the model), recall R (referring to the proportion of correctly recognized images in the test set), and average precision mean mAP@50 The evaluation indicators for this experiment include the average accuracy AP of all categories when the IoU threshold is 0.5, detection speed FPS (referring to the maximum number of images processed by the model within 1 second), parameter size Params, and weight size. The calculation formula for evaluation indicators is as follows as shown in (6) - (9):

$$P = \frac{TP}{TP+FP} \tag{6}$$

$$R = \frac{TP}{TP+FN} \tag{7}$$

$$AP = \int_0^1 P(\mathrm{t})dt \tag{8}$$

$$mAP = \frac{\sum_{n=1}^{N} AP_n}{N} \tag{9}$$

In the formula, TP represents the correct number of birds predicted by the model; FP is the model that predicts the background as the number of birds; FN is the number of birds predicted by the model as background; AP is the average accuracy, and P (t) represents the accuracy when the intersection to union ratio threshold is set to t; N represents the number of target categories in object detection. In this experiment, there was only one category, birds, with N=1; AP_n represents the nth category and the average accuracy of each category.

4.4. Improving the Visualization of Training and Detection Results of YOLOv8 Model

The training process of the improved YOLOv8s model mAP@50 curve showing in the figure 6. In the first 50 epochs, as shown in Figure 6(a), the model’s mAP@50 curve exhibited significant oscillations for both training and validation datasets. However, after 100 epochs, as seen in Figure 6(d), the curve gradually stabilized, reflecting improved consistency in performance. The early-stage oscillations were effectively reduced, and from around epoch 100 onward, the model showed more steady progress, with a smoother rise in mAP@50 for both training and validation sets. By focusing on the training up to 200 epochs, Figure 6(e) illustrates continued improvement in mAP@50, with the validation curve trailing the training curve, but the gap narrowing as training progressed. This suggests the model was stabilizing and optimizing its performance without signs of underfitting or overfitting.

As the training continued up to 300 epochs, shown in Figure 6(f), the mAP@50 of the validation set consistently exceeded the original YOLOv8s performance, indicating better generalization and no significant overfitting. The gap between the training and validation mAP@50 curves remained manageable, confirming the model was not overfitting, as no drastic divergence was observed. Additionally, the improved model was validated by running forward inference on five test images, which included birds of varying sizes, some of them flying. The improved YOLOv8s successfully detected all five instances of flying birds, showcasing its enhanced detection capabilities. These results, consistent with the mAP@50 trends observed during training, validate the improvements made to the model, confirming its ability to generalize effectively without significant underfitting or overfitting.

4.5. Ablation Experiment

This study not only considers the impact of introducing ECA attention mechanism and MBC3 module on model performance, but also considers the impact of adjusting network depth and width (AWD) on model performance based on the above improvements, as shown in Table 3. The α_w represents the width scaling factor, and α_d represents the depth scaling factor. The larger the α_w, the more feature maps the network outputs at each layer, and the larger the α_d, the deeper the network layers. Case4 mAP@50 is the highest.

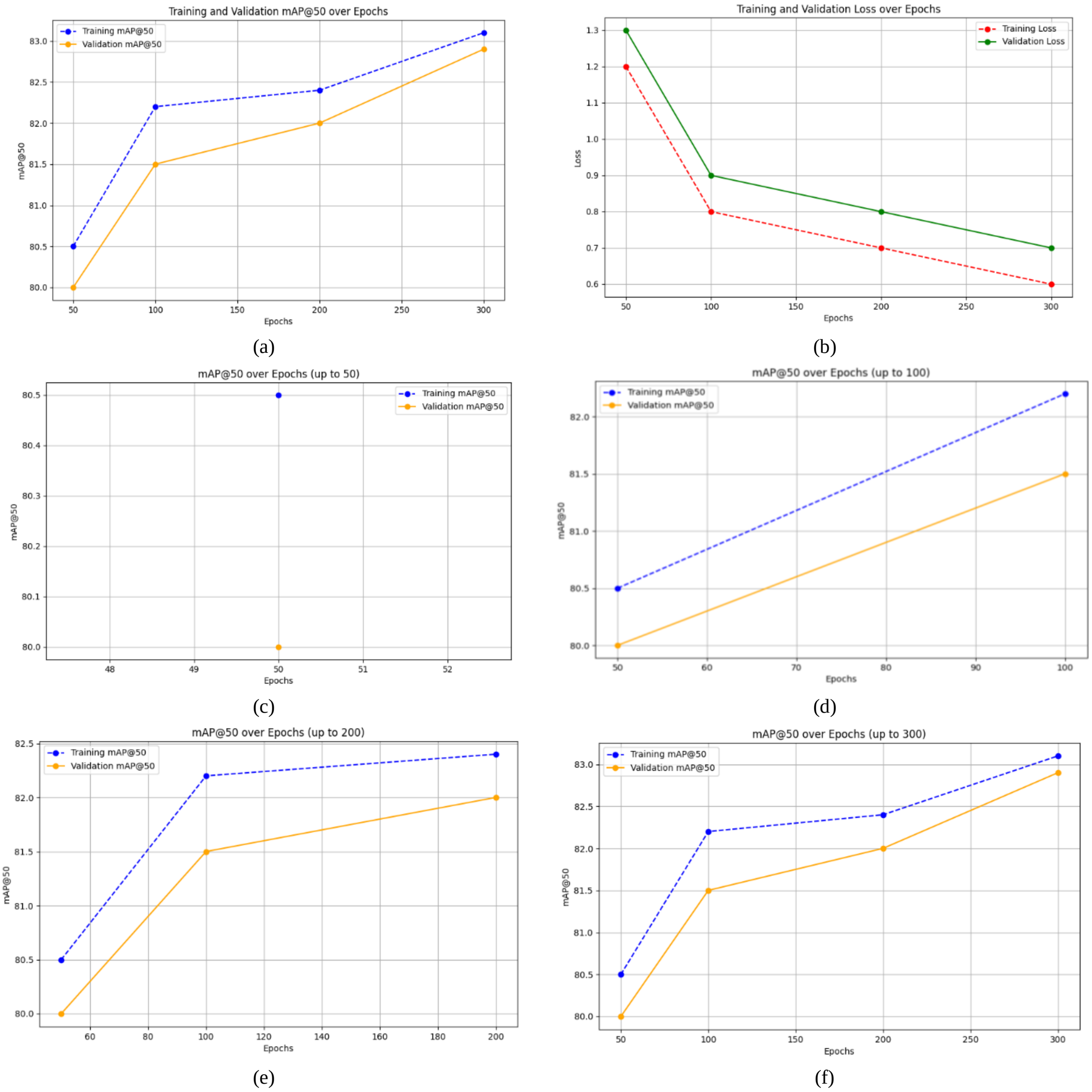


Fig.6. Visualization of training and detection results of improved YOLOv8s model (a) Training and Validation mAP@50 over Epochs, (b) Training and Validation Loss over Epochs, (c) mAP@50 over Epochs (up to 50), (d) mAP@50 over Epochs (up to 100), (e) mAP@50 over Epochs (up to 200),(f) mAP@50 over Epochs (up to 300)

Table 3. The impact of different network widths and depths on model performance

Method	α_w	α_d	mAP@50	FPS
case2	0.5	0.33	82.2%	30
case3	0.325	0.33	81.2%	27
case4	0.325	0.67	82.4%	31

The results of the ablation experiment are shown in table 4. From Table 4, it can be seen that the four improvements introduced significantly improved the model's performance mAP@50 FPS reduces the parameter count, computational complexity, and weight size of the model.

Table 4. Ablation experimental results

Method	mAP@50	FPS	P	R	Params	Weight Size/MB	GFLOPs
YOLOv8s	80.5%	24	78.9%	77.5%	11.1M	22.6	28.4
YOLOv8s + ECA	81.6%	28	80.9%	76.7%	11.1M	22.6	28.6
YOLOv8s + ECA + MBC3	82.2%	30	83.3%	75.7%	9.8M	21.1	26.9
YOLOv8s + ECA + MBC3 + AWD	82.4%	31	80.4%	78.4%	5.8M	12.9	17.4
YOLOv8s + ECA + MBC3 + AWD + SF-PAN	**83.1%**	**31**	80.3%	77.3%	**5.8M**	**12.9**	**17.5**

Compared to the original YOLOv8s model, the improved YOLOv8s mAP@50 the most significant improvement was achieved with a 2.5% increase, FPS increased by 7 frames per second, parameter count decreased by 5.8M from 11.1M, weight size decreased by 9.7MB, and the GFLOPs of the improved YOLOv8s were only 61.6% of the original model, indicating that the improved YOLOv8s consumes less computing resources compared to the original model as shown in figure 7.

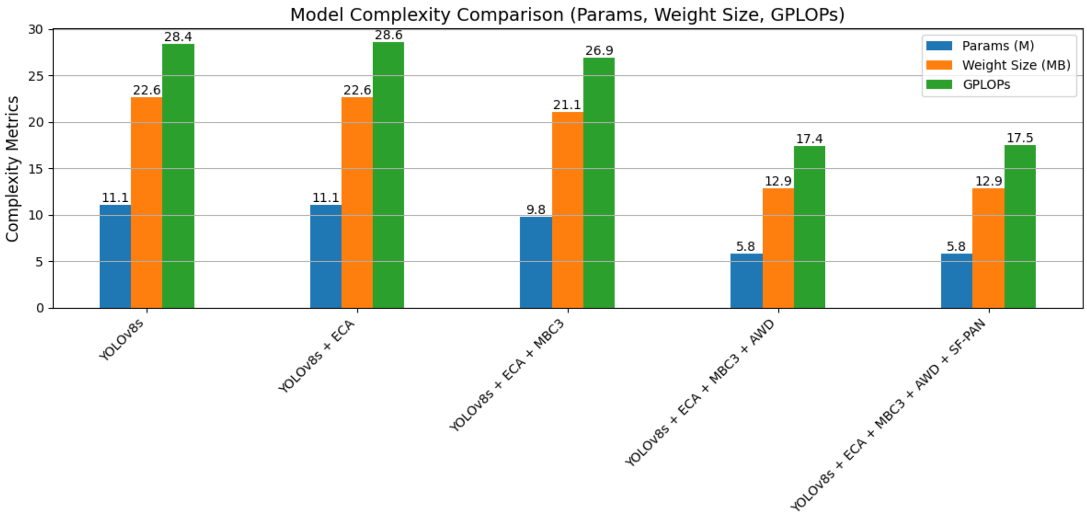


Fig.7. Ablation experimental results

4.6. *Comparative Experiment*

In this section, we present the results of the comparative experiments conducted between the improved YOLOv8s model and several well-established object detection models. The key evaluation metrics used are mAP@25, mAP@50, and mAP@75, which represent the mean average precision [30] at different Intersection over Union (IoU) thresholds, as well as FPS (Frames Per Second), parameter count, and weight size.

Table 5. Comparative experimental results

Method	mAP@25	mAP@50	mAP@75	FPS	Params	Weight size/MB
YOLOv8s	90.5%	80.5%	28.7%	24	11.1M	22.6
YOLOv3 [5]	92.6%	59.5%	10.8%	15	61.5M	235.0
YOLOv3-tiny [5]	51.8%	48.8%	17.6%	24	12.1M	24.4
YOLOv4-tiny [6]	89.3%	49.6%	10.2%	67	5.90M	22.4
YOLOv6s [7]	90.2%	78.7%	26.0%	24	16.3M	33.0
SSD [32]	88.6%	55.4%	1.70%	53	3.50M	14.3
CenterNet [33]	92.7%	68.2%	5.29%	28	32.7M	124.0
RT-DETR [34]	86.1%	66.1%	26.9%	29	9.70M	19.8
Faster RCNN [35]	7.76%	3.56%	0.63%	14	28.3M	108.0
Ours	**92.8%**	**82.9%**	**30.0%**	31	5.80M	**12.9**

The improved YOLOv8s model demonstrates strong performance, particularly with a mAP@50 value of 82.9%, outperforming all other models except YOLOv3 in mAP@25. With a mAP@25 of 92.8% and mAP@75 of 30.0%, the model achieves a good balance between accuracy and precision across various IoU thresholds. This performance makes it particularly robust for bird detection in low-altitude airport airspace. While YOLOv4-tiny achieves a higher FPS of 67, it compromises on accuracy, especially at higher mAP thresholds, whereas YOLOv8s achieves 31 FPS while maintaining higher accuracy. The YOLOv8s model is also smaller and more efficient, with only 5.8M parameters and a weight size of 12.9MB, making it ideal for deployment in systems with limited computational resources, such as embedded systems at airports [31]. It outperforms models like YOLOv3 and YOLOv6s, which have higher parameters and larger weight sizes but do not match the detection accuracy and speed of YOLOv8s.

The improved YOLOv8s model shows strong potential for real-time bird detection systems, offering a superior balance between accuracy and detection speed. Its efficiency in handling small, fast-moving targets, such as birds, makes it highly suitable for airport applications, outperforming other object detection models like Faster R-CNN and RT-DETR. However, there are areas for future improvement, such as enhancing robustness in diverse operational environments affected by lighting conditions or weather. Future research could explore domain adaptation techniques to address these challenges. Additionally, integrating YOLOv8s with other sensor types, such as radar or thermal cameras, could improve its detection capabilities, especially in low-visibility scenarios, extending its potential beyond visual detection.

5. Conclusions and Future Research Directions

In order to achieve more accurate and rapid detection of airport bird targets and provide a more effective decision-making basis for airport bird repelling, this paper proposes an airport bird detection algorithm based on YOLOv8 mAP@50 Reaching 82.9% and FPS of 31 frames per second, it meets the requirements of real-time object detection tasks. Compared to other algorithms, this algorithm has the highest detection accuracy. This algorithm can perform bird detection on image data transmitted from video surveillance devices deployed near airports in a distributed manner. Subsequently, the detection results of the algorithm can be transmitted to the airport bird repelling system to achieve accurate perception of birds in the airspace near the airport, and then corresponding repelling operations can be carried out on dangerous targets to ensure the safety of aircraft operation. Due to the small size and fast movement speed of bird targets, the detection effect is not ideal. So, continuing to improve the accuracy of bird detection and precise tracking of detected bird targets is the key direction of future research.

However, there are certain limitations to this study. The algorithm's performance may be affected by environmental factors such as lighting conditions, weather, and the complexity of the bird's flight path, which were not fully explored in this research. Additionally, the current system is limited to detecting birds in specific flight paths and may struggle with tracking birds in highly dynamic environments.

Future research should focus on enhancing the algorithm’s robustness under varying environmental conditions and improving its ability to track birds in complex and unpredictable flight paths. Moreover, the integration of additional sensor types (e.g., radar, thermal imaging) could further improve detection reliability, particularly in low-visibility situations. Expanding the system for use in other wildlife detection contexts, such as for monitoring other types of wildlife near airports or in other transportation settings, could also increase its applicability and impact.

References

[1] Qiao, Liang, and Liang Zhang. "Airport bird strike risk assessment and research." In *IOP Conference Series: Materials Science and Engineering*, vol. 612, no. 5, p. 052071. IOP Publishing, 2019.

[2] Kaur, R., & Singh, S. (2023). A comprehensive review of object detection with deep learning. Digital Signal Processing, 132, 103812.

[3] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[4] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[5] REDMON J, FARHADI A. Yolov3: An incremental improvement[C]//Computer vision and pattern recognition. Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.

[6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[EB/OB]. (2020-04-23)[2024-04-23].https://doi.org/10.48550/arXiv.2004.10934.

[7] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[EB/OB]. (2022-09-07)[2024-04-23].https://doi.org/10.48550/arXiv.2209.02976

[8] 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[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.

[9] Ma, H., Li, Y., Gao, Y., Zhang, Z., Abuadbba, A., Fu, A. & Abbott, D. TransCAB: Transferable clean-annotation backdoor to object detection with natural trigger in real-world. In 2023 42nd International Symposium on Reliable Distributed Systems (SRDS) (2023, September) (pp. 82-92). IEEE.

[10] Zhang, W., Sun, Y., Xue, K., & Yao, A. Research on modulation recognition algorithm based on channel and spatial self-attention mechanism. IEEE Access (2023).

[11] Tan, K., Ding, S., Wu, S., Tian, K., & Ren, J. A small object detection network based on multiple feature enhancement and feature fusion. Scientific Programming, 2023(1), 5500078.

[12] Wang, J., Yu, J., & He, Z. (2022). DECA: a novel multi-scale efficient channel attention module for object detection in real-life fire images. Applied Intelligence, 1-14.

[13] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.

[14] Lim, Xin Roy, Chin Poo Lee, Kian Ming Lim, and Thian Song Ong. "Enhanced traffic sign recognition with ensemble learning." Journal of Sensor and Actuator Networks 12, no. 2 (2023): 33.
[15] K. Tong, Y. Wu, and F. Zhou, ''Recent advances in small object detection based on deep learning: A review,'' Image Vis. Com put. vol. 97, May 2020, Art. no. 103910, doi: 10.1016/j.imavis.2020. 103910.
[16] J. Wang, Y. Chen, Z. Dong, and M. Gao, ''Improved YOLOv5 network for real-time multi-scale traffic sign detection,'' Neural Comput. Appl., vol. 35, no. 10, pp. 7853–7865, Apr. 2023.
[17] Li, Ning, Wenliang Zhang, Zhaoxu Liu, Kaifeng Liu, Junjie Wang, and Fan Zhang. "PDSE-YOLOv8: a lightweight detection method for internal defects in asphalt roads." Signal, Image and Video Processing (2024): 1-12.
[18] K. Jiang, T. Xie, R. Yan, X. Wen, D. Li, H. Jiang, N. Jiang, L. Feng, X. Duan, and J. Wang, ''An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation,'' Agriculture, vol. 12, no. 10, p. 1659, Oct. 2022, doi: 10.3390/agriculture12101659.
[19] Z. Huang, J. Wang, X. Fu, T. Yu, Y. Guo, and R. Wang, ''DC SPP-YOLO: Dense connection and spatial pyramid pooling based Yolo for object detection,'' Inf. Sci., vol. 522, pp. 241–258, Jun. 2020, doi: 10.1016/j.ins.2020.02.067.
[20] Cheng, Gong, Xiang Yuan, Xiwen Yao, Kebing Yan, Qinghua Zeng, Xingxing Xie, and Junwei Han. "Towards large-scale small object detection: Survey and benchmarks." IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).
[21] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, ''Feature pyramid networks for object detection,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.
[22] Tian, Zhi, Chunhua Shen, and Hao Chen. "Conditional convolutions for instance segmentation." In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16, pp. 282-298. Springer International Publishing, 2020.
[23] H. Liu, F. Sun, J. Gu, and L. Deng, ''SF-YOLOv5: A lightweight small object detection algorithm based on improved feature fusion mode,'' Sensors, vol. 22, no. 15, p. 5817, Aug. 2022, doi: 10.3390/ s22155817.
[24] L. Kong, J. Wang, and P. Zhao, ''YOLO-G: A lightweight network model for improving the performance of military targets detection,'' IEEE Access, vol. 10, pp. 55546–55564, 2022, doi: 10.1109/ACCESS.2022. 3177628.
[25] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, ''Classifi cation of brain tumor from magnetic resonance imaging using vision transformers ensembling,'' Current Oncol., vol. 29, no. 10, pp. 7498–7511, 2022.
[26] L. Meng, H. Li, B.-C. Chen, S. Lan, Z. Wu, Y.-G. Jiang, and S.-N. Lim, ''AdaViT: Adaptive vision transformers for efficient image recognition,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog nit. (CVPR), New Orleans, LA, USA, Jun. 2022, pp. 12299–12308, doi: 10.1109/CVPR52688.2022.01199.
[27] Guo, Jianyuan, Kai Han, Han Wu, Yehui Tang, Xinghao Chen, Yunhe Wang, and Chang Xu. "Cmt: Convolutional neural networks meet vision transformers." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 12175-12185. 2022.
[28] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, ''DETRs beat YOLOs on real-time object detection,'' 2023, arXiv:2304.08069.
[29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
[30] Wang, T., Su, J., Xu, C., & Zhang, Y. (2022). An intelligent method for detecting surface defects in aluminium profiles based on the improved YOLOv5 algorithm. Electronics, 11(15), 2304.
[31] SUN H, WANG Y, CAI X, et al. AirBirds: A Large-scale Challenging Dataset for Bird Strike Prevention in Real-world Airports[C]//Proceedings of the Asian Conference on Computer Vision. 2022: 2440-2456.
[32] LIU W, ANGUELOY D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
[33] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6569-6578.
[34] ZHAO Y, LV W, XU S, et al. Detrs beat yolos on real-time object detection[EB/OB]. (2023-04-17) [2024-04-23]. https://doi.org/10.48550/arXiv.2304.08069.
[35] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

Authors' Profiles

Zia Ur Rehman is a Master's candidate at Yangzhou University, China, specializing in artificial intelligence, software engineering, and machine learning. He holds a Master's degree in Software Engineering from Yangzhou University, Jiangsu, China, and a Bachelor's degree from Abdul Wali Khan University Mardan, Pakistan. His research focuses on intelligent robotics systems, AI-driven instruments and equipment, and big data technologies. He has expertise in deep learning algorithms, real-time computer vision systems, group sparsity techniques, and optimization methods. His work involves cutting-edge research in the areas of object detection and machine learning model optimization. His email is ziaurrahman4h@gmail.com

Abdul Ghafar received a master's degree in computer science and technology from Jiangsu University of Science and technology, China, in 2021. He is currently pursuing a PhD degree in Yangzhou University, china. His current research interests are machine learning and deep learning with a focus on crop disease classification using advanced computer vision techniques. His email is ayanghafar783@gmail.com

Ahmad Syed received his PhD degree from JNTU Hyderabad University, in 2018. He is currently working as Postdoc Researcher at Yanshan University, dept. of Electrical Engineering, China, since 2024. Dr. Ahmad Syed has extensive experience in renewable energy and power electronics, with a particular emphasis on the design and application of grid-connected PV inverters. He previously served as an Assistant Professor at the CBIT campus, Osmania University. He has authored or co-authored over 20 technical papers. He holds 1 patent. He is currently working on DC-DC converters for H2, EV applications. His contributions to research include numerous publications in prestigious international journals and book chapters, such as those by IET and Taylor & Francis. He has also actively participated in international conferences organized by IEEE, IET, Springer, and Elsevier.

Abu Tayab from Chittagong, Bangladesh, is a researcher focused on deep learning applications in vehicle-following technologies, enhancing accuracy, stability, object detection, and energy efficiency in Intelligent Connected Vehicles (ICVs). With expertise in MATLAB, machine learning, and IoT, he aligns his work with Industry 4.0 goals. Abu holds a bachelor's from Chongqing University of Science and Technology and is pursuing a master's in mechanical engineering at Yanshan University, China. His email is tayababu2591@gmail.com

Md. Golam Rabbi received a bachelor's degree in computer science and technology from Hebei University of Science and Technology, China, in 2020. He is currently pursuing a Master's degree at Yangzhou University, china. His current research interests are machine learning and Control systems with a focus on unmanned aerial vehicles and drones. His email is rabbi66@qq.com

How to cite this paper: Zia Ur Rehman, Abdul Ghafar, Ahmad Syed, Abu Tayab, Md. Golam Rabbi, "An Optimized YOLOv8-based Method for Airport Bird Detection: Incorporating ECA Attention, MBC3 Module, and SF-PAN for Improved Accuracy and Speed", International Journal of Intelligent Systems and Applications(IJISA), Vol.17, No.2, pp.1-13, 2025. DOI:10.5815/ijisa.2025.02.01