Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 3, 2023

Object Detection for Aircraft Turnover Milestone based on Modified

Qirui Jiang *, Yuqi Liu

School of Software Engineering, Chengdu University of Information Technology, Chengdu, Sichuan, 610225, China
*** Corresponding author:** Qirui Jiang (Email: 1037030742@foxmail.com)

Abstract: In the target detection task of aircraft turnaround milestone in foggy scenario, there are some problems such as unstable location of prediction frame boundary, high error detection rate and poor detection effect of small target. A new target detection method BTM-YOLO (Broad-sighted upsample and three-dimensional attention multiple detection head YOLO) is proposed, which is based on YOLOv7 network. Add a small target detection head to improve the ability of small target detection; The up-sampling module OVRAFE is introduced to reduce the information loss in the up-sampling process. Replace CIoU with Median Wise IoU (MWIoU) to suppress the problem of poor sample swelling in data sets. The improved model makes up for the performance shortcomings of small target detection in foggy days, and the average detection accuracy on the real foggy day test set is 76.2%, which is 3.32% higher than that of the original model, basically meeting the task requirements.

Keywords: Object Detection; Haze Environment; Aircraft Turnaround Milestone; YOLOv7; Multiple Detection Head; Attention Mechanism.

1. Introduction

Video analysis is an important means to ensure the normal operation of all aspects of the airport, and the aircraft turnaround milestone is completed by real-time monitoring of the tarmac monitoring, which has been transformed from traditional manual to automatic. The biggest change is that the use of a breakthrough target detection algorithm based on deep neural network has been tried to replace manual work and has achieved certain results. The aircraft turnaround milestone is applied to different links of the whole flight process, and it is particularly important to guarantee multiple nodes such as water truck entry, loading bridge entry and follow me car entry during the period before take-off and after landing. Most of the existing target detection studies on airport targets mainly focus on ideal images, while the target detection studies in harsh environments such as haze weather are mostly seen in other common scenes such as road vehicles, pedestrians, traffic signs, etc. The research on ground target detection in civil airports under haze weather is still vacant.

This research is part of the optimization project of Video analysis support Node automatic acquisition technology of Civil Aviation Research Institute. The main purpose of the whole work process from flight landing to take-off is to make up for the shortcomings of the target recognition of this technology in the work process, improve the shortcomings of poor target recognition in fog-day scenarios, and lay the groundwork for further work. Keep the project moving forward.

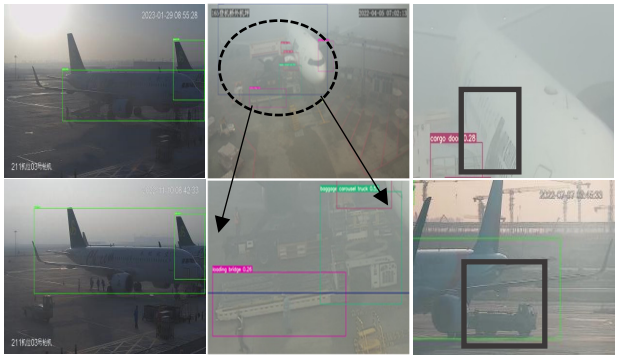
2. Background and Motivation

Since the introduction of deep learning, object detection algorithms can be divided into traditional and deep learning-based algorithms. The main features of traditional algorithms include artificially designed object features, sliding window mechanism to extract features, and traditional classifiers with cumbersome steps. The main feature of the detection algorithm based on deep learning is that it is based on the characteristics of deep network learning, the target selection is based on candidate box or direct regression, and relies on deep network classification.

According to the steps, the deep learning based target detection network can be divided into one-stage direct regression to the target class and two-stage pre-generated candidate box and then predicted. In recent years, the single-stage detection algorithm is represented by YOLO series. Classical methods include YOLOv3, YOLOX[1], YOLOv7 [2], etc., which directly carry out the detection task as a regression task. The two-stage detection method generates a certain number of candidate boxes and then conducts the detection, such as the classic RCNN, Fast R-CNN, Faster R-RCNN[3], etc. For example, Huang et al.[4], who adopted the single-stage detection method, combined the improved SSD algorithm with the feature pyramid fusion network to improve the multi-scale target detection performance of the detection algorithm. In the detection of personnel in the tarmac area, Wang et al.[5] combined Ghost-NET network with YOLOv3 algorithm and added SE attention mechanism to improve the algorithm's ability to extract important features. In addition, convolution layer and normalization layer are combined to reduce the number of parameters. Based on the YOLOv3 algorithm, Xia et al.[6] replaced the original convolutional module with the depth-separable convolutional module, and replaced the original intersection ratio loss function with the distance-based loss function DIoU to improve the detection speed and accuracy of the algorithm for targets in the airport. Based on the YOLO v5 algorithm, Yi et al.[7] reduced the depth of the feature pyramid and limited the maximum downsampling multiple to solve the problem of difficult identification of small targets. In addition, by adjusting the depth of feature transmission of the residual module, the repetitive superposition of background features was alleviated. Yin et al.[8] effectively transformed the algorithm's learning of physical texture and surface information of target categories into learning of contouring through style transfer of fog weather images, alleviating the influence of light and noise and improving the detection

performance of foggy targets. Based on YOLOv4, Liu et al.[9] replaced the backbone with ShuffleNet V2[10] network to reduce the number of parameters, and combined with DeblurGANv2[11] image enhancement algorithm to improve the detection accuracy and speed. Based on YOLOv3 network, Yang et al.[12] introduced Xception[13] module to improve the accuracy, and GridDehazeNet[14] network to de-fog the image, which enhanced the image contrast. DIoU function was introduced to replace the original IoU function, which improved the feature extraction capability and positioning accuracy, and the accuracy and location performance of the network are improved. Xie et al.[15] proposed DONet, a joint learning framework for image de-fogging and target detection, for target detection tasks in fog scenes. By jointly optimizing the de-fogging network and detection network, the network can more truly recover image details and simulate color features during the process of learning de-fogging, thus improving the accuracy of target detection. In addition, Lv et al.[16], who adopted the two-stage detection method, improved the father-RCNN algorithm, added SA attention mechanism in the feature extraction process, improved the feature extraction capability of the algorithm, and replaced the original loss function with DIoU function to locate the target more accurately and improve the performance of the algorithm.

The main problems encountered in the research process include: 1.Lack of applicable data sets, and the existing public data sets are not suitable for the research content of this paper; 2. The model loses image details and target information during the up-sampling process; 3. After high-multiple downsampling, the feature map information is missing, which hinders the subsequent network layer's perception of the target features; 4. The existing YOLOv7 model has poor detection effect on small-size targets in the research scene; 5. The YOLOv7 model has unstable positioning of target objects and high false detection rate. See Figure 1 for details.



a b c

Fig 1. The problem of identifying results from the underlying model

For the above problems, after practical exploration, the contribution is as follows:

1) In view of problem 1, self-built data set is targeted according to the task scenario of the study, frame slices are made for the surveillance video during the aircraft turnaround milestone process, and the data set is made, in which the real fog image is used as the fog test set.

2) For problem 2, OVRAFE upsample module is introduced to replace UPsample module in Head network. OVRAFE upsample module has a wider feature receptive field and retains more details of target objects in the process of upsampling, reducing information loss, easing the impact of noise, facilitating deep network perception and extracting feature information. The OVRAFE module improves the model performance at the cost of increasing the number of model parameters.

3) For problem 3, the Head layer is connected to the shallow layer of the Backbone network, so that the deep network can obtain part of the feature information of the shallow network feature map and strengthen the feature perception ability of the deeper network.

4) For problem 4, a small target detection head is added to the existing YOLOv7 network structure to enhance the small target detection capability of the model.

5) For problem 5, in the prediction part of the model, Mwise IoU loss function is introduced to replace the original CIoU loss function, which enhances the location ability of the boundary frame and improves the stability of the detection frame.

In figure a, the detection frame selection of the target aircraft is extremely unstable and fluctuates greatly. In Figure b, the picture below is a partial enlargement of the above picture, and the lifting platform vehicle is mistakenly detected as a baggage conveyor truck and the tail is mistakenly detected as a gallery bridge. The thick wire frame in figure c is the missing target, which is the cabin door and the water truck respectively.

3. BTM-YOLO Network

3.1. YOLOv7 Introduction

the target detection task as a regression problem. Compared with the previous version, the detection accuracy and speed have been improved. ELAN module is a new structure proposed by YOLO v7, which is used to replace the transition layer. Its main function is to solve the problem of excessive training difficulty caused by excessive number of transition layers and increasing of shortest gradient path when stacking blocks. The algorithm is mainly divided into two parts: Backbone network and Head network. The input image is extracted by Backbone backbone network, and after repeated subsampling, three different size feature maps are output respectively, and then transmitted to the subsequent Head network for feature map fusion and detection.

3.2. BTM-YOLO Network

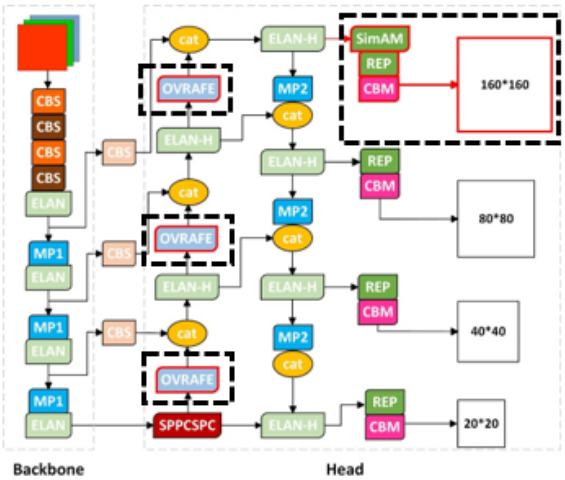


Fig 2. BTM-YOLO Network

On the basis of YOLO v7, a lightweight OVRAFE module is introduced to replace the original upsampling module, the

Head network is connected to the shallow part of the backbone network, and a new small target detection head NeHead is added. The specific structure is shown in Figure 2. The improved upsampling structure and the added small target detection head with 3D attention are shown in the dotted box.

3.3. Upsampling Module with Global View

In YOLOv7, the nearest neighbor up-sampling of feature maps are simple and efficient, but the disadvantage is that semantic information on feature maps cannot be utilized, the range of sensing areas is insufficient, and important target detail features will be lost after upsampling. Because there are quite a number of images distorted by fog deterioration in the data set, the sampling module needs to have a larger receptive field and stronger semantic information acquisition ability. In this regard, OVRAFE, an upsampling module with a broader field of view, is introduced, which is improved based on the CARAFE upsampling operator, and the receptive field with channel weights is introduced, which is improved from the previous local receptive field around the central point to the global receptive field, and the range of information acquisition is further improved. OVRAFE module contains three sub-modules, including channel weighting module, upsampled kernel prediction module and feature recombination module, as shown in Figure 3. The process is as follows:

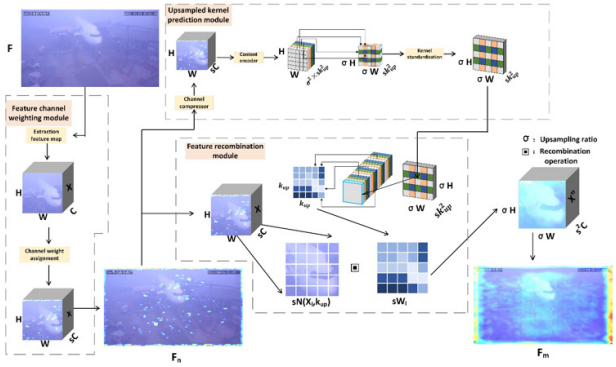


Fig 3. The working process of OVRAFE module

1) The channel weighting module assigns weights to all channels in the feature graph $\boldsymbol{F}$ and multiplicates with the feature graph $\boldsymbol{F}$ to obtain the weighted feature graph $\boldsymbol{F_n}$. The up-sampled kernel D is predicted by the up-sampled kernel prediction module. Finally, the feature graph $\boldsymbol{F}$ and the up-sampled kernel D are recombined to obtain the final output feature graph $\boldsymbol{F_m}$. All symbols are shown in Table 1. Assume that the input size of feature graph $\boldsymbol{F}$ is H×W×C. In the module of channel weighting, the feature graph $\boldsymbol{F}$ is globally averaged, and each channel corresponds to a vector. After integration, activation function ReLU and Sigmoid are used to process the channel weight matrix s, and the matrix is multiplied with the corresponding channel of feature graph $\boldsymbol{F}$ to obtain the feature graph $\boldsymbol{F_n}$:

$$\boldsymbol{F_n} = \boldsymbol{F} \times s \tag{1}$$

2) Assume that the up-sampling rate is σ in the up-sampling kernel prediction module, and the processing process of the input feature graph $\boldsymbol{F}$ is as follows:

(1) 1×1 convolution was used to compress the number of channels to H×W×C_m to reduce the amount of computation in subsequent steps;

(2) For content coding and up-sampling kernel prediction, assuming that the size of the up-sampling kernel is k_j×k_j, the size of the up-sampling kernel to be predicted is σH×σW×k_j^2 on the premise that the up-sampling kernel is used for all positions of the output feature map;

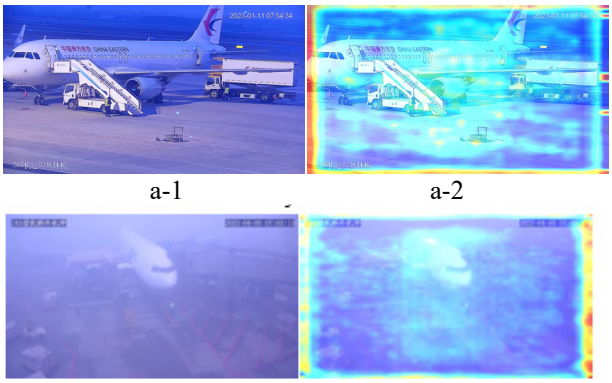
(3) A k_i×k_i convolution layer is used to predict the compressed feature graph in the first step. The number of input channels remains unchanged, and the number of output channels becomes σ^2×k_j^2. The channel dimension is expanded in spatial dimension, and the size of the up-sampled kernel is σH×σW×k_j^2, and the reconstructed kernel is denoted as W_l;

(4) SoftMax function is used to process the upsampled kernel obtained in the third step, the main purpose is to normalize, so that the weight sum of the convolution kernel is 1.

3) In the feature recombination module, for each part of the feature map output from the previous module, map them in situ according to the position of the input feature map, and then select this position as the center of the square region with the range of k_j×k_j. In the feature map $\boldsymbol{F_n}$, mark all such square fields as N($\boldsymbol{F_n}$, k_j), and do the dot product with the predicted upper sampling kernel of this point. The output feature graph $\boldsymbol{F_m}$ is obtained:

$$\begin{aligned} \boldsymbol{F_m} &= sN(\boldsymbol{F_n}, k_j) \times sW_l \\ \boldsymbol{F_m} &= \sigma H \times \sigma W \times s^2 C \end{aligned} \tag{2}$$

As shown in Figure 3, the feature map is input into the OVRAFE module, and after feature recombination operation, a new feature map is output. In the output feature map, the size and brightness of the shadow block at different positions represent the model's attention to this place.



a-1 a-2

b-1 b-2

Fig 4. The shallow feature map and deep feature thermal map

Table 1. Symbolic meaning

Symbol	meaning
σ	Upsampling ratio
H, W, C	Feature map size
D	Upsampled kernel
$\boldsymbol{F}$	Initial feature map
$\boldsymbol{F_n}$	Feature map with weights
$\boldsymbol{F_m}$	The final output feature map
s	Vector matrix with weights
k_j	Size of the upsampled kernel
W_l	The reassembled kernel
N($\boldsymbol{F_m}$,k_j)	All feature plots $\boldsymbol{F_m}$ of size k_j

As shown in Figure 4, subfigures a-1 and b-1 are the feature maps of the shallow network, and a-2 and b-2 respectively correspond to the feature heat maps of the deep network after

a-1 and b-1 have been sampled by multiple OVRAFE modules. The highlighted color blocks covered in the feature maps will attract more attention, so that the network has a bias to retain correct target feature information. Areas with fewer highlighted color blocks will reduce attention, suppressing the impact of incorrect feature information and unimportant background information. The formula symbols are shown in Table 1.

3.4. Optimized Head Network

In the foggy airport ground scene, the size of the target category is different, the image quality is degraded under the influence of fog, and the features of small targets are seriously lost. It is difficult for the model to extract the feature details of small targets in the foggy scene, and it is unable to effectively learn the correct features of small targets in the foggy day, which ultimately leads to the insufficient detection performance of the model for small targets in the foggy day. In order to optimize the performance of YOLO v7 model in detecting small targets in fog environment, the network structure is deepened on the basis of the original three size detection heads of 20×20, 40×40 and 80×80, and a target detection head NeHead with 3D attention mechanism is added, which outputs 160×160 large-size feature maps. In the deep network, the feature map is undersampled at a high multiple and the feature details of small targets are mixed with fog in the background, so the model cannot extract enough correct features. More detailed features of small targets are needed. Residual connections are made between the Head layer and the shallow layer of BackBone network to obtain features from the feature map of the shallow layer of the network. Improve the feature extraction ability of small targets in foggy days. The structure of the improved BTM-YOLO network is shown in Figure 2 above, and the structure of the detection head is shown in Figure 5 below:

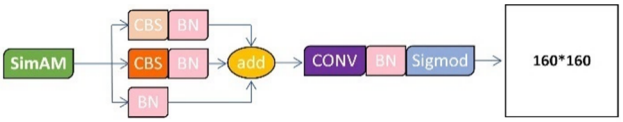


Fig 5. Detection head structure with three dimensional attention mechanism

On the basis of two-dimensional channel attention and spatial attention, SimAM attention mechanism[18] introduces the third dimension "energy" to obtain more information. In neuroscience, the firing patterns of information-rich neurons are different from those of nearby neurons, and upon activation, nearby neurons are inhibited, i.e., spatially inhibited. Then switch to the perspective of feature extraction, features with more information are different from other features in terms of energy magnitude, and enhance the energy value of features with more information, so that the correct features are easier to learn. According to the formula defined by the energy paradigm and assuming that each channel has M = H×W energy functions, in order to simplify the calculation, the mean value of all neurons $\hat{\mu}$ and the variance $\hat{\sigma}^2$ are used to replace the strain value in the formula, and the expression of the minimum energy value is obtained:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{3}$$

The target neuron has lower energy, is more different from the surrounding neuron, and is more important. When the feature is enhanced by the obtained energy value, its reciprocal is used. The expression is as follows:

$$X_n = \text{sigmoid}(\frac{1}{e_t^*}) \odot X \tag{4}$$

By using the three-dimensional attention mechanism, the target features are weighted and enhanced from the three dimensions of channel, space and dimensional energy to extract richer feature information. By optimizing the attention mechanism, this method can extract the feature information of small targets more effectively. Small targets often have small size and low contrast in the image, which is difficult to be accurately identified by traditional detection methods. In this method, energy is introduced on the two-dimensional basis of channel and space to optimize the attention mechanism, which can enhance the contrast and significance of small targets in the feature map, and improve the recognition accuracy of the detection head. The meanings of related symbols are shown in Table 2.

Table 2. Symbol meaning

Symbol	Meaning
t	The target neuron of the input feature
$M = H \times W$	The number of neurons in a channel
e_t^*	The minimum energy of a channel
$\hat{\mu}$	Input neuron mean
$\hat{\sigma}^2$	The variance of the input neuron
X	The original eigenvalue
X_n	The enhanced eigenvalue
$\odot$	For reinforcement operations

3.5. Optimize the Boundary Frame Loss Function

The boundary frame loss function of YOLOv7 is CIoU[19], which takes into account geometric factors such as distance, aspect ratio and Angle, effectively improves the regression accuracy of the prediction frame and speeds up the regression efficiency. However, the regression samples preset by CIoU are all high-quality samples, but it ignores that there are a considerable number of average quality samples and low quality samples in the samples. Therefore, combined with Wise IoU(WIoU) bounding frame loss function with dynamic non-monotony static focusing mechanism[20], a new Median Wise IoU(MWIoU) loss function is proposed according to the current situation of low-quality samples flooding in foggy scenes, and both high-quality and low-quality samples are kept under suppression. And the tolerance of general quality samples is improved. The function is divided into three parts, IoU part, distance loss and monotonic focusing coefficient. Figure 6 shows the schematic diagram of WIoU parameters. Dark prediction box D, center coordinates (x,y), light color box is the real box, center coordinates (x_{gt}, y_{gt}). According to the IoU definition, the improved distance loss is defined as follows:

$$R_{MWIoU} = \exp(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}) \tag{5}$$

During the calculation of the loss terms, Wg and Hg are

separated from the calculation graph to avoid the negative gradient that they may appear in, as mentioned above, impeding convergence.

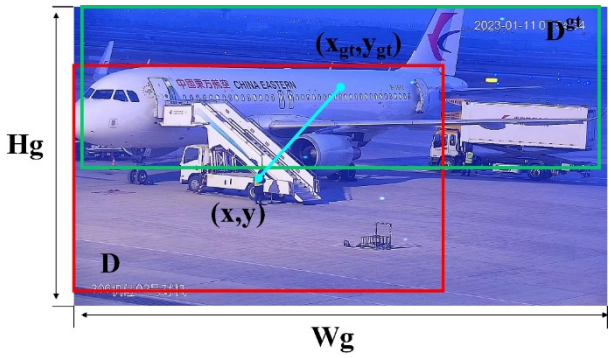


Fig 6. MWIoU Distance loss function parameter diagram

The non-monotonic focusing coefficient q is defined as follows:

$$q_M = \frac{\varepsilon}{\delta \alpha^{\varepsilon-\delta}} \tag{6}$$

$$\varepsilon = \frac{L_{IoU}^*}{M(L_{IoU})} \in [0,+\infty) \tag{7}$$

In the above formula, the parameter ε is introduced to describe the outlier, which is the ratio of L_{IoU}^* to $M(L_{IoU})$, the $L_{IoU}^*$is the loss contribution value of each training sample, and $M(L_{IoU})$ represents the median value of the loss contribution of all samples. The smaller the outlier, the higher the anchor frame quality and the greater the outlier. It means that the mass of anchor frame is lower, so both types of anchor frame will get a small gradient gain. Compared with the mean value, the median $M(L_{IoU})$ is also dynamic, and is not affected by individual limit samples, and the quality division criteria of the anchor frame still retains the advantage of dynamic change, so MWIoU can formulate appropriate strategies to complete gradient gain allocation according to the conditions faced at different times. The final definition of MWIoU is as follows:

$$L_{MWIoU} = q_M R_{WIoU} L_{IoU} \tag{8}$$

In the frame regression, the model will focus on the anchor frame of ordinary quality. The idea of this strategy is to enhance the whole, so the gradient contribution of low quality anchor frame and high quality anchor frame is suppressed at the same time. The model can be focused on the common quality of the anchor frame, thereby improving the model performance from an overall perspective.

4. Experimental Environment and Result Analysis

4.1. Experimental Environment

All experiments were performed in the following environments: Intel® Xeon® Processor E5-2690 v3 CPU, NVIDIA Quadro P5000 (16GB video memory) GPU, 64GB memory, and Windows10 operating system. Development environment PyCharm2022, Python3.9, Pytorch1.10.1.

4.2. Data Sets and Data Enhancement

The research data is derived from the ground surveillance video of multiple civil aviation airports, and the surveillance video is processed to form a data set. The real fog scene images were extracted as the real test set, a total of 375 images, and the remaining clear images were used as the training set. Two fog synthesis methods were used to enhance the image pre-processing of the training set in the way of uniform interval extraction. One was reverse fog synthesis based on the dark channel de-fogging process, and the other was fog synthesis based on the atmospheric light scattering model formula and the higher the concentration of the distance. After data enhancement, a total of 3144 images were obtained. The training set and test set instances are shown in Figure 7 and Table 3.

Fig 7. All categories

Table 3. Category and number of instances

Category	Number	
	Train Dataset	Test Dataset
plane	5751	716
cabin door	4656	105
cargo door	1750	175
baggage carousel truck	1099	174
stair truck	852	173
food truck	756	91
water service truck	512	103
follow me car	484	99

4.3. Evaluation Index

The experimental results used R (Recall), mean Average Precision mAP (mean Average Precision), parameter number (Params) and floating-point operations per second (FLOPs) as evaluation indexes.

4.4. Training Strategies

The training rounds of the model were set to 300, the picture size was set to 960×960, the batch size was set to 7, the optimizer was set to SGD, the initial learning rate was set to 1e-2, the final learning rate to 5e-4, the initial moment was set to 0.937, and image-weights was turned on.

4.5. Experiments on COCO2017 Dataset

In order to get close to the target categories in the airport, all the four categories of aircraft, cars, buses and trucks are extracted from COCO2017 data set, and 4,500 pieces are randomly selected, of which 4,000 pieces are used as the training set, and the remaining 500 pieces are used as the test set based on the synthesis of fog images based on the principle of atmospheric light scattering. The test results are shown in Table 4 The visualization results are shown in Figure 8. From the comparison between a-1, b-1, c-1 and a-2, b-2, c-2, it can be observed that the improved algorithm can still obtain more feature information and detect more small targets in the case of dense fog, thus improving the overall performance of the

model. It can also be seen from Table 4 that the performance of the improved detection algorithm is improved by 2.95%.

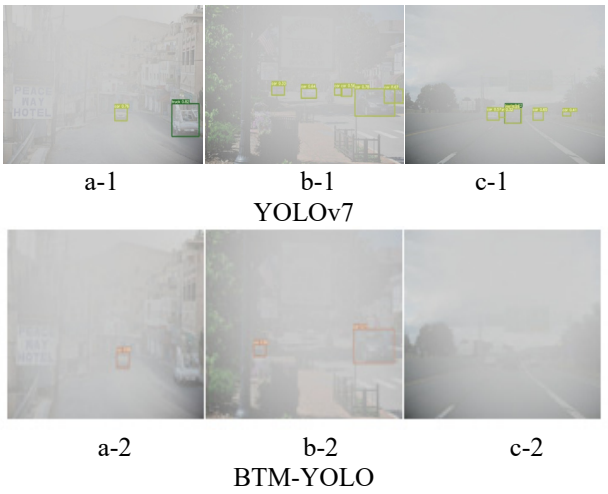


Fig 8. Detection results on the COCO dataset

Table 4. Detection results on the COCO dataset

Model	airplane	car	bus	truck	mAP0.5/%
	AP0.5/%				
YOLOv7	76.51	74.33	82.19	78.64	77.92
BTM-YOLO	76.37	83.26	84.01	79.83	80.87

4.6. Ablation Experiment

In order to verify whether all the improvements are effective for the detection task of flight support nodes in fog-day scenarios, all the improvements are verified, and the results are shown in Table 5. The visualization results are shown in Figure 8. After adding NeHead detection head, the small target detection capability of the model is enhanced, along with the overall detection performance, the number of parameters is increased slightly, and the average accuracy is increased by 1.74%. After the OVRAFE module is added, the deep network perceives more feature details, the detection head perceives more target feature information, and the recognition ability is enhanced, and the average accuracy is increased by 0.91%. After the loss function WIoU is adopted, the loss contribution weight of ordinary quality samples is increased in training, and low quality samples are suppressed. The average accuracy of the improved model BTM-YOLO is increased by 0.67% without increasing the number of model parameters.

Compared with the original model YOLO v7, the detection accuracy of the improved model BTM-YOLO is increased by 3.32% and the number of parameters is increased by 0.61M. The amount of calculation increased a lot, but the accuracy improved. From the comparison of test results before and after the model improvement in FIG. 9, it can be seen that in the first row of comparison, YOLO v7 in FIG. A-2 only selected the main part of the aircraft class, and the wing part was excluded, and there was difficulty in classifying food trucks.

The BTM-YOLO result of A-3 included the aircraft wing part. The position of the aircraft was accurately located, and small targets such as the passenger door and the cargo door at the arrow in the figure were identified; In the second row, YOLOv7 of Figure b-2 misidentifies the cargo trailer at the arrow as a food truck, while b-3's BTM-YOLO does not, and correctly identifies the small target cabin door; In the third row, c-2's YOLOv7 did not identify the target, while c-3's BTM-YOLO identified the passenger elevator and cargo door.

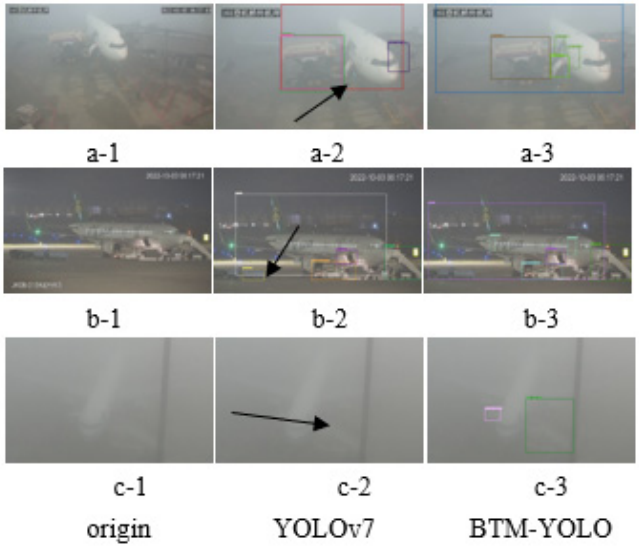


Fig 9. Comparison of detection results before and after algorithm improvement

Table 5. Ablation experiment

Model	Params/M	FLOPs/G	mAP0.5/%
Model 1: YOLOv7	37.01	105.1	72.88
Model 2:1+NeHead	37.42	116.8	74.62
Model 3: 2+ OVRAFE	37.62	120.1	75.53
Model 4: 3+ MWIoU	37.62	120.1	76.20

4.7. Performance Comparison Experiments of Various Algorithms

The algorithm in the research is based on the improvement of YOLO v7. In order to verify the performance of BTM-YOLO algorithm, the image size is 960×960, which is compared with other target detection algorithms. It includes two-stage target detection algorithm Cascade R-CNN and single-stage target detection algorithm YOLOv4, YOLOv5, YOLOv7 and YOLOv8. The Cascade R-CNN backbone network adopts ResNet50 FPN, while several algorithms of the YOLO series are based on Darknet53. Table 6 shows the detection result data, and Figure 10 shows the detection result of the visual test set. The thick wire frame identification of the target is not detected, and the misdetected target is marked by an arrow. Large targets such as aircraft and food trucks are detected stably. The Cascade R-CNN results of FIG.

a-1 and b-1 are good, with only one cabin door missing, while YOLOv4 of FIG. a-2 and b-2 and YOLOv5 of FIG. a-3 and b-3 are mostly missing. Small targets such as cabin door are not detected. YOLOv7 in FIG. a-4 and b-4 and YOLOv8 in FIG. a-5 and b-5 have better results and less missed detection, while the algorithm proposed in this paper significantly improves the detection performance of small targets. While maintaining the overall detection performance, the detection effect of small targets is significantly improved.

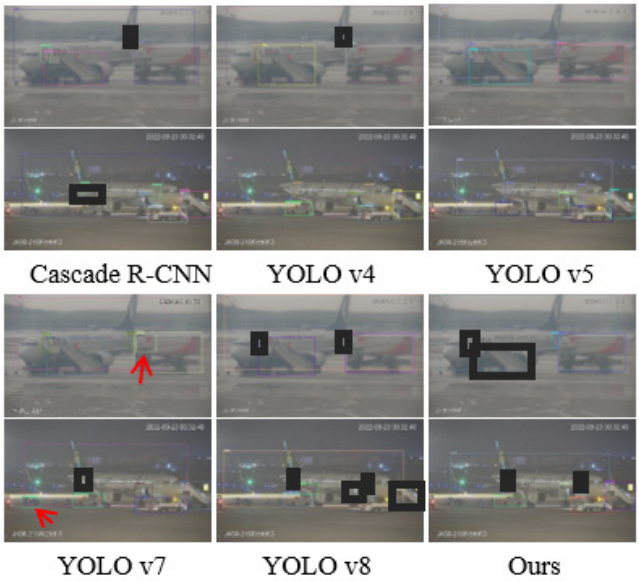


Fig 10. Comparison of detection results of multiple algorithms

5. Conclusion

OVREFA up-sampling module is introduced to expand the field of perception, retain more features in the sampling process, and alleviate the problems of image detail distortion and target feature information loss. The residual connection between the shallow layer and the deep layer of the backbone network is made to improve the ability of the network to feel the characteristic information stably. Adding a small target detection head improves the performance of small target recognition; The CIoU function is replaced by the MWIoU bounding loss function to suppress the low-quality samples whose number expands due to fog interference, and to increase the weight of high-quality and ordinary quality samples. From the test set results and COCO2017 data set test results, the improved algorithm has higher precision and better robustness than other algorithms, and performs better in the target detection task of flight support nodes in foggy days. However, the current algorithm still has shortcomings, such as the large number of parameters in the entire model, which will be targeted and lightweight for subsequent deployment.

References

[1] GE Z, LIU S , WANG F ,et al.YOLOX: Exceeding YOLO Series in 2021[J]. 2021.DOI:10.48550/arXiv.2107.08430.

[2] 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[J].arXiv e-prints, 2022.DOI: 10.48550/ arXiv. 2207. 02696.

[3] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[4] Huang G W, Li W, Zhang B H, et al. Improved SSD-Based Multi-scale Object Detection Algorithm in Airport Surface[J]. Computer Engineering and Applications, 2022,58(05):264-270.

[5] Wang Y, Yuan G W, Qu R, et al. Target Detection Method of Airport Apron Based on Improved YOLOv3[J].Journal of Zhengzhou University (Natural Science Edition), 2022,54 (05): 22-28.DOI:10.13705/j.issn.1671-6841.2021287.

[6] Xia Z H,Wei R X,Tu J,et al. Moving Target Detection Method for General Aviation Airport[J]. Science Technology and Engineering, 2022,22(29):13114-13119.

[7] Yi J H, Qu S J, Yao Z K,et al. Traffic sign recognition model in haze weather based on YOLOv5[J]. Journal of Computer Applications, ,2022,42(09):2876-2884.

[8] Yin X P, Zhong P, Xue W,et al. The Method of UAV Image Object Detection under Foggy Weather by Style Transfer[J]. Aero Weaponry, ,2021,28(03):22-30.

[9] Liu S G, Zhang L K,Du H D,et al. Improved Object Detection of YOLOv4 in Foggy Conditions[J/OL]. Journal of System Simulation: 1-10[2023-08-07]. DOI: 10. 16182/ j.issn 1004731 x.joss.22-0423.

[10] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]// Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.

[11] KUPYN O, MARTYNIUK T, WU J, et al. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better[C]// Proceedings of the IEEE/CVF international conference on computer vision. 2019: 8878-8887.

[12] Yang K Z, Yan X N,Sun J, et al. A DeRf-YOLOv3-X Object Detection Method for Rainy and Foggy Background[J]. Chinese Journal of Sensors and Actuators, 2022,35(09):1222-1229.

[13] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.

[14] LIU X, MA Y, SHI Z, et al. Griddehazenet: Attention-based multi-scale network for image dehazing[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 7314-7323.

[15] Xie Y H, Xie Y, Chen L,et al. Object Detection in Real-World Hazy Scene[J]. Journal of Computer-Aided Design & Computer Graphics,2021,33(05):733-745.

[16] Lv Z L, Chen L Y. A Novel Deep Neural Network Compression Model for Airport Object Detection[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2021,38 (04): 571-586.DOI:10.16356/j.1005-1120.2021.04.004.

[17] LOY C C , LIN D , WANG J ,et al. CARAFE: Content-Aware ReAssembly of FEatures.2019[2023-08-07].

[18] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]//International conference on machine learning. PMLR, 2021: 11863-11874.

[19] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]// Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.

[20] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J]. arXiv preprint arXiv:2301.10051, 2023.