



Research article

Joint attention mechanism for the design of anti-bird collision accident detection system

Xuanfeng Li and Jiajia Yu*

Department of Electronic and Computer Engineering, Southeast University Chengxian College, Nanjing 210088, China

* **Correspondence:** Email: yujiajia8@126.com; Tel: +18021654925.

Abstract: Among various aviation accidents, bird collision is one of the most common accidents for civil passenger aircraft in recent years. With the significant breakthrough of deep convolutional neural networks in the field of target detection, this paper proposes a target detection method to prevent bird collision accidents. The algorithm in this paper integrates different attention mechanisms on the YOLOv5s network to solve the problems of small target detection miss, false detection and insufficient feature extraction capability. The trend-aware loss (TAL) and trend factor (W_i) are used to solve the drift of the prediction frame. After comprehensive ablation experiments, the improved algorithm shows significant improvement on the detection accuracy and speed. Results indicate that mean average precision (mAP) value reaches 99.8%, which is 6.3 percentage points higher than the original algorithm.

Keywords: bird collision accident; flight safety; YOLOv5s improvement; SE module; CBAM module; TAL; W_i

1. Introduction

With the development of national information technology, aircraft has become a key methods of transportation. In recent years, with the occurrence of low-altitude passenger aircraft accidents and low-altitude traffic management has become more and more strict, and the prevention of risks in civil passenger aircraft in advance has gradually become a new research hot spot for scholars. Among various aviation accidents, bird collision [1] is one of the most dangerous threats to civil airliners. In scenarios such as airports, the use of ultrasound to repel birds prior to incident [2] is the basis for

preventing bird collisions. Using infrared surveillance to capture multiple airfield scenes, using surveillance video to obtain the flight paths of birds and aircraft in real time to detect and identify them has practical research significance and scenario application value.

The traditional algorithms focus on the motion information of birds, detection methods, and object tracking. The literature [3,4] proposes a skeleton-based FBD method from the aspect of motion information in order to overcome the versatility of birds by describing the motion information of birds through a set of key poses. Based on the geometric topological relationship between key parts of the bird body, a set of key poses is described by extracting skeleton features, combining the flying bird skeleton features with the extracted keyword set, and the final detection results are verified using the consistency of the key frame poses variation set and the sequence image classification results. In terms of detection method, to control the cost, dedicated bird detection using 94 GHz millimeter wave radar is proposed in literature [5] during aircraft takeoff and landing, which can be scanned without gimbal or phased array components, but cannot be detected in real time. From the aspect of object tracking, a novel filtering method for fast and effective multi-scale and fast-connected speckle extraction is proposed in literature [6] for fast and accurate segmentation of moving objects in video sequences to handle various scene change sources. An intelligent video surveillance system is developed to test the performance of the algorithm by analyzing the properties of object motion in image pixels and time frames and combining two constraint levels to accomplish this for moving target localization. Therefore, in preventing bird collision accidents and precise control of birds and aircraft, research on more efficient, accurate and fast intelligent detection and identification methods for flight element information has become a key research direction at this time.

To address the detection of arbitrary directional targets and fine-grained recognition of aircraft types, a cascade framework based on convolutional neural networks for arbitrary directional and multi-type aircraft detection in remote sensing images is proposed in literature [7]. A fine-grained recognition sub-network with integrated learning and Fisher discriminant regularization is used to identify aircraft types in images for more accurate recognition. An edge-based hatch recognition and tracking method is proposed in literature [8,9] for identifying different hatches with similar shapes in order to solve the difficulties encountered when different replicas are covered. By means of simple geometrically constrained image contours, a new compound cover descriptor composed of edge features and position description vectors is used to identify those different compound covers with similar shapes. To solve the problem of high miss detection rate and false alarm rate when complex and dense targets, a Faster R-CNN based multi-angle feature driven and majority voting strategy is proposed in literature [10,11]. The multi-angle transform module is used to transform the input image to achieve multi-angle feature extraction of targets in the image.

There are many problems with the existing detection results, such as missed detection, false detection and insufficient feature extraction capability due to birds being too small relative to aircraft, drift and delay of the prediction frame due to the excessive flight speed of birds and aircraft. This paper proposes the following solutions based on YOLOv5s: introduces the attention mechanism SE and CBAM modules to solve the problem of missed detection of small targets such as birds; introduces a new loss function TAL and W_i to solve the prediction frame drift and delay problems. Comprehensive ablation experiments reveal that the improved YOLOv5s_SE&CBAM_TAL algorithm has significantly improved detection precision and detection speed.

2. The improved network structure of this paper

In this paper, we improve each component of YOLOv5s: introduce different attention mechanisms [12,13] in backbone and head to solve the problems of small target detection miss, false detection and insufficient feature extraction capability. The channel attention mechanism SE module [14] is introduced in the backbone for posterior improvements to form YOLOv5s_SEA. The hybrid domain attention mechanism CBAM module [15] is added in the head for channel domain and then spatial domain improvements to form YOLOv5s_CBAMA. The head output is changed to decoupled head approach and a new loss function, TAL, is introduced [16] to form YOLOv5s_SE&CBAM_TAL. The decoupled head approach increases the complexity of the operation, but the precision is improved and the convergence of the network is accelerated. The improved Intersection over Union (IoU) loss function is used to train the reg branch and the Binary Cross Entropy (BCE) loss function to train the cls branch. The data set is put into the improved network for training, and its structure is shown in Figure 1.

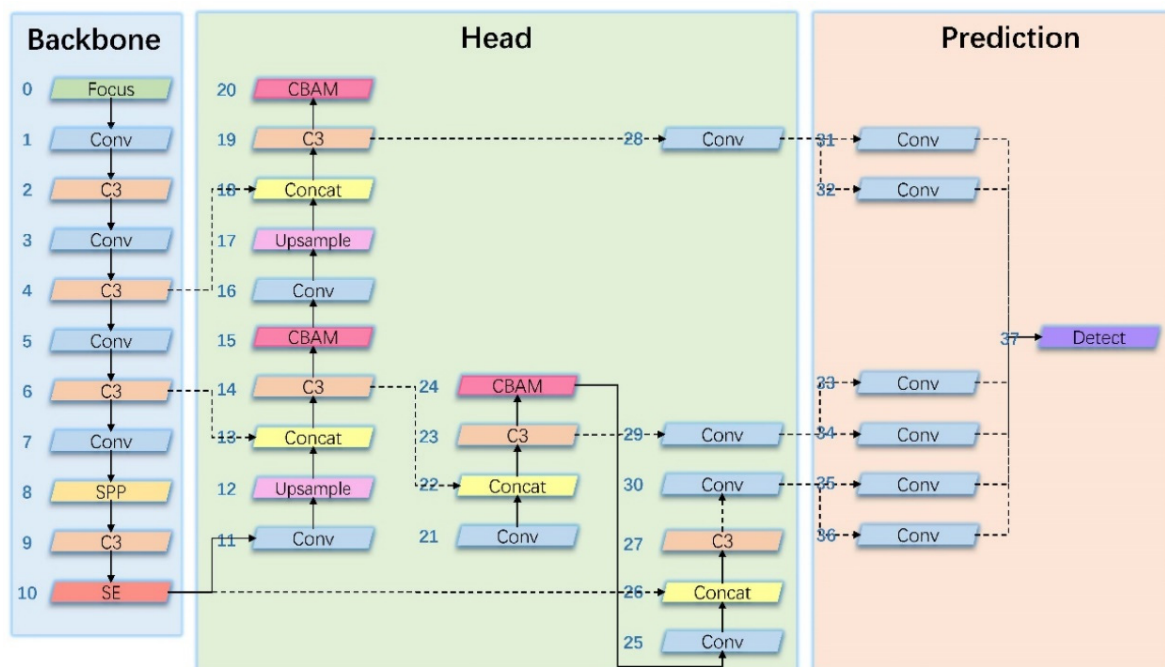


Figure 1. Improved YOLOv5s network structure.

3. Experimental results and analysis

3.1. Improvements based on backbone

This paper introduces the channel attention mechanism Squeeze-and-Excitation (SE) networks [17], aiming at autonomous learning to establish the mutuality among channels, and employs a dynamic weighting approach to rescale the channel weights. As over increases of the depth and breadth of the network can bring problems such as gradient disappearance and over-fitting. Based on the design principle of combination construction between similar modules, this paper embeds the SE module into

the backbone of YOLOv5s to produce a variety of combinations with rear, front, external rear, external front. The new network model YOLOv5s_SE is generated, as shown in Figure 2, as YOLOv5s_SEA, YOLOv5s_SEB, YOLOv5s_SEC, and YOLOv5s_SED, respectively.

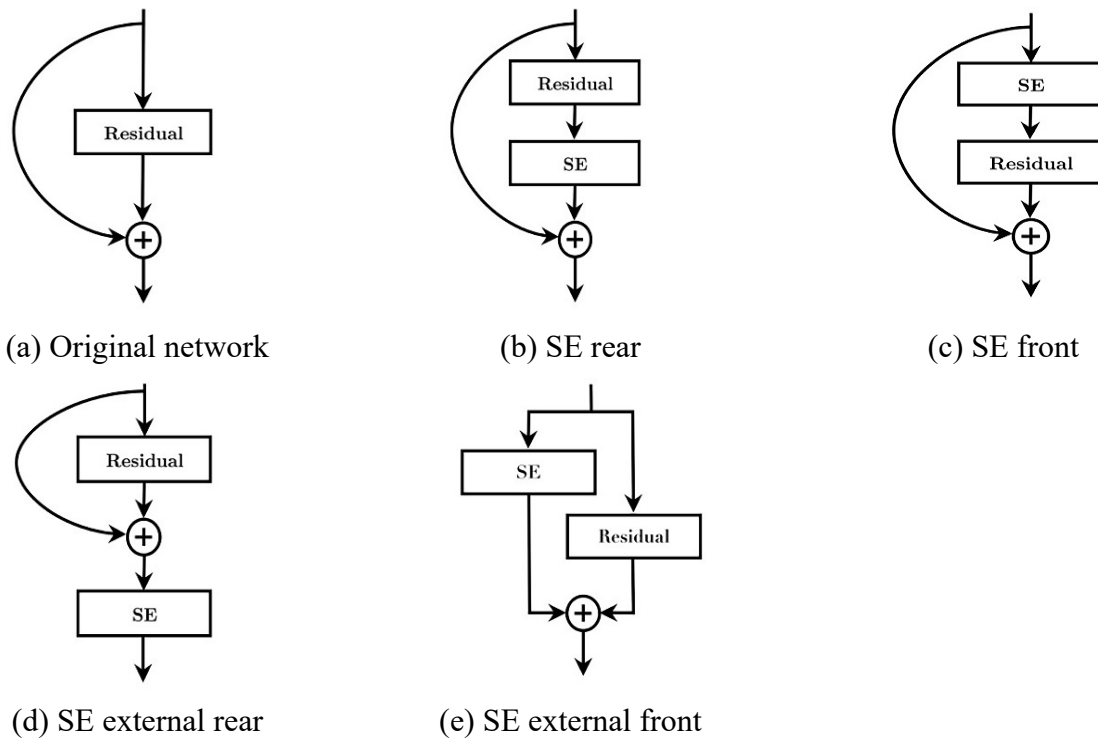


Figure 2. Various combinations of SE module and network.

Different combinations of SE modules are combined with untreated YOLOv5s for comprehensive ablation experiments. In this paper, three performance indexes: precision, recall and mAP are applied to the experiment; and the specific calculation formula can be expressed as below.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (3)$$

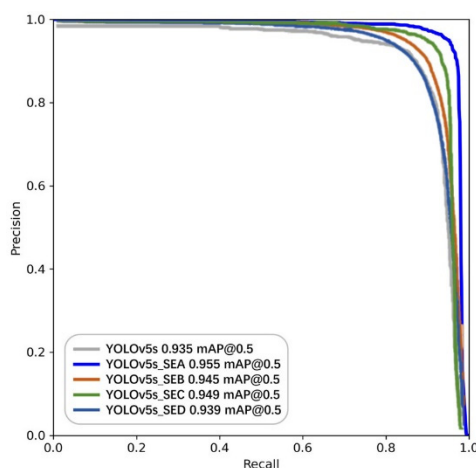
TP is the number of positive samples detected correctly. FP is the number of negative samples detected as positive. FN is the number of backgrounds incorrectly detected as positive, and N is the total number of categories.

A series of detection index data are obtained after 300 epochs of training for both training and testing phases, results as shown in Table 1. Based on the data and the structural analysis, the SE posterior is applied. The experimental results indicate that adding the SE module to the last layer of the backbone of YOLOv5 works best.

Table 1. Performance metrics data for the SE module and four different models.

Network model	Precision	Recall	mAP
YOLOv5s	84.1%	97.9%	0.935
YOLOv5s_SEA	85.2%	98.5%	0.955
YOLOv5s_SEB	84.5%	98.2%	0.945
YOLOv5s_SEC	84.8%	98.4%	0.949
YOLOv5s_SED	84.3%	97.9%	0.939

Based on the comprehensive ablation experimental results, YOLOv5s_SEB and YOLOv5s_SED can be excluded since the change of precision or recall rate are unobvious, while the other two models all have a large improvement. The most representative index mAP changes are analyzed again. As shown in Figure 3, the trend of the change of mAP obtained by training 300 epochs for different combinations of network models. The dark blue line represents YOLOv5s_SEA, whose mAP finally improved to 0.955, a 2-percent increase.

**Figure 3.** Variation trend of mAP between SE module and four different models.

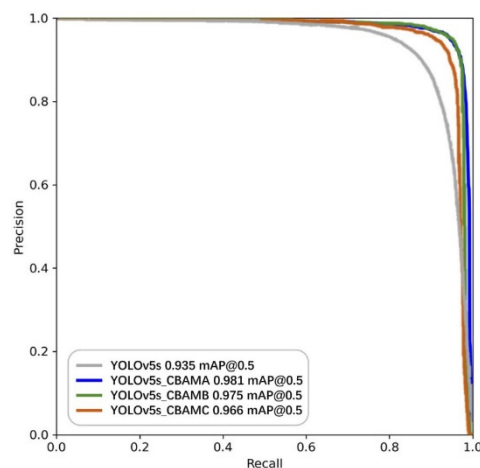
3.2. Improvements based on head

The introduction of the SE module brings only a small improvement in network performance, leading to the introduction of the hybrid domain attention mechanism Convolutional Block Attention Module (CBAM) [18]. CBAM is the attention module that integrates both channel and spatial dimensions in two different dimensions. This paper generates a new network model YOLOv5s_CBAM by embedding these two modules into the head of YOLOv5s in parallel or sequentially. They are: YOLOv5s_CBAMA, channel domain and then spatial domain; YOLOv5s_CBAMB, spatial domain and then channel domain; and YOLOv5s_CBAMB, channel domain and spatial domain in parallel. We conducted comprehensive ablation experiments with the above three networks and the unmodified YOLOv5s, as shown in Table 2. Based on the data and structural analysis, this paper adopts YOLOv5s_CBAMA as it saves parameters and computational power to some extent, and it is easy to apply to the new network architecture.

Table 2. Combination comparison of CAM and SAM in CBAM.

Network model	Precision	Recall	mAP
YOLOv5s	84.1%	97.9%	0.935
YOLOv5s_CBAMA	87.2%	98.6%	0.981
YOLOv5s_CBAMB	86.9%	98.4%	0.975
YOLOv5s_CBAMC	84.8%	98.2%	0.966

Based on the comprehensive ablation experimental results YOLOv5s_CBAMC can be excluded since the change of precision or recall rate are unobvious, while the other two models have a large improvement. The most representative metric mAP changes are analyzed again. As shown in Figure 4, the trend of the change of mAP obtained by training 300 epochs for different combinations of network models. The dark blue line represents the YOLOv5s_CBAMA and its mAP finally improves to 0.981, a 4.6-percent increase.

**Figure 4.** Combination of CAM and SAM in CBAM to compare mAP variation trends.

3.3. Integrating SE and CBAM modules

The improved YOLOv5s algorithm is introduced in different attention mechanisms here. The optimal SE module YOLOv5s_SEA is introduced in backbone; the optimal CBAM module YOLOv5s_CBAMA is introduced in head. Combined ablation experiments with the previous two sets of experiments are shown in Table 3.

Table 3. Analysis of performance index results of improved algorithm.

Network model	Precision	Recall	mAP
YOLOv5s	84.1%	97.9%	0.935
YOLOv5s_SE	85.2%	98.5%	0.955
YOLOv5s_CBAM	87.2%	98.6%	0.981
YOLOv5s_SE & CBAM	90.5%	98.9%	0.995

The next trend of mAP obtained by training 300 epochs of different improved algorithm network models is shown in Figure 5. The dark blue line represents YOLOv5s_SE&CBAM, and its mAP finally improves to 0.995, a 6-percent increase.

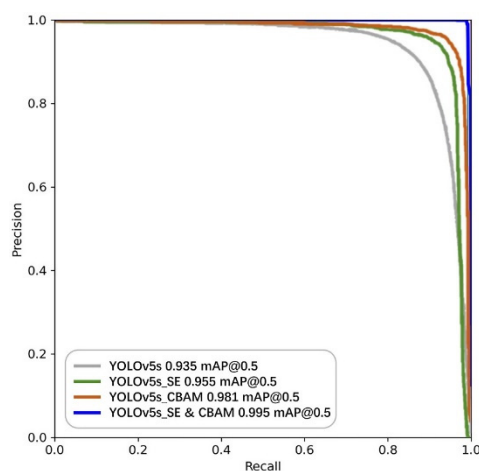


Figure 5. Trend of mAP for different improved algorithms.

3.4. Improvement based on prediction

The loss function can affect the detection performance of the network by influencing the learning of the network parameters. Due to the flexibility of bird movement, the requirements for the delay of the network model are extremely high. By the time of the detection of the target in the current frame completes, the next frame has already changed, thus bird collisions are not effectively prevented. Since the stream sensing is the result of the current frame, the calibration is always matched and evaluated by the next frame, and the performance gap makes the inconsistency between the current processing frame and the next matching frame. As shown in Figure 6(a), the green box indicates the actual object, while the red box indicates the predicted object, and the red arrow indicates the drift of the predicted frame due to the processing time delay. The improved schematic is shown in Figure 6(b). In order to solve the drift of the prediction frame, a TAL and a W_i are proposed in this paper, considering the delay and accuracy to measure the movement speed quantitatively.



(a) before improvement



(b) after improvement

Figure 6. Before-and-after visualization results.

Based on YOLOv5s_SE&CBAM, a decoupled head approach is changed at the head output in this paper. Although it increases the complexity of the operation, the precision is improved and the convergence speed of the network is accelerated. The improvement uses the IoU loss function to train the reg branch and the BCE loss function to train the cls branch. YOLOv5s is used in this article as the Baseline. Using the GT boxes of the previous frame, the current frame and the next frame (F_{t-1} , F_t , G_{t+1}), a triplet is constructed for training. Two adjacent frames (F_{t-1} , F_t) are taken as input to train the model to predict the GT boxes of the next frame. The GT boxes of G_{t+1} are supervised by the real GT boxes of F_t frames. Based on the input and supervised triples, this paper reconstructs the training data set into the form of $(F_{t-1}, F_t, G_{t+1})_{t=1}^{n_t}$, as shown in Figure 7.

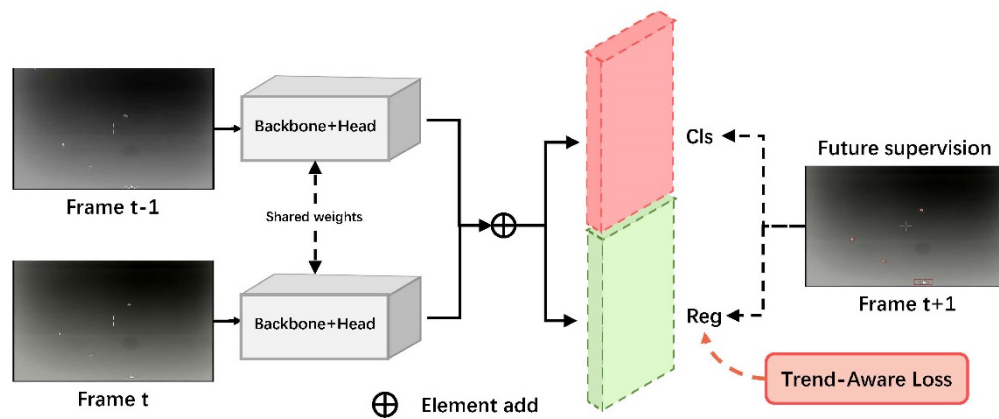


Figure 7. Training process.

The matching IoU of the detected object between two frames is obtained by calculating the IoU matrix between two GT boxes, and then performing the maximization operation on the dimensionality. A small value of this matching IoU means that the object moves fast, and vice versa. If a new object appears in the frame, there is no matching frame with it, at this time a threshold τ is set to deal with this situation, and the specific calculation formula can be expressed as below.

$$mIoU_i = \max_j (IoU(box_i^{t+1}, box_j^t)) \quad (4)$$

$$w_i = \begin{cases} \frac{1}{mIoU_i}, & mIoU_i \geq \tau \\ \frac{1}{v}, & mIoU_i < \tau \end{cases} \quad (5)$$

\max_j is the maximum operation value t between boxes in F_t , and v is the constant weight of the new object.

Trend-aware loss parameters are mainly set by two parameters τ and v . Then the parameter selection is crucial. In order to better evaluate the parameters good or bad. An accuracy streaming Average Precision (sAP) is proposed here to evaluate the accuracy by simultaneously evaluating the time delay and the accuracy of the detection. In order to determine an optimal set of τ and v suitable for the bird collision prevention phenomenon for this paper, several different sets of τ and v are selected for experiments in this paper. Where τ is denoted as a threshold to monitor the new object, and v is

denoted to control the degree of attention to the new object. In this paper, ν is set to be greater than 1.0, and the grid search data is performed for these two hyperparameters, and the results are shown in Table 4. Considered together, the optimal values of $\tau = 0.3$ and $\nu = 1.4$ for the two parameters are chosen to ensure a high sAP value and achieve the best performance.

Table 4. Parameter settings for TAL.

τ	N	sAP
0.2	1.3	33.5
0.2	1.4	33.7
0.2	1.5	33.7
0.3	1.3	33.7
0.3	1.4	34.1
0.3	1.5	33.8
0.4	1.3	33.2
0.4	1.4	33.6
0.4	1.5	33.3

In this paper, the task of processing delayed streams is focused. Under this task, TAL is proposed in this paper to alleviate the processing lag problem in stream perception. This paper uses a large number of approximation calculations based on deep reinforcement learning to obtain a better detection equilibrium. Compared with Baseline, YOLOv5s_SE&CBAM_TAL improves the mAP by 4.5% and achieves robust prediction at different birdie speeds. Next, the different improved algorithm network models are subjected to a comprehensive ablation experiment to train 300 epochs to obtain the trend of mAP, as shown in Figure 8. The dark blue line represents YOLOv5s_SE&CBAM_TAL, whose mAP is finally improved to 0.998, a 6.3-percent increase.

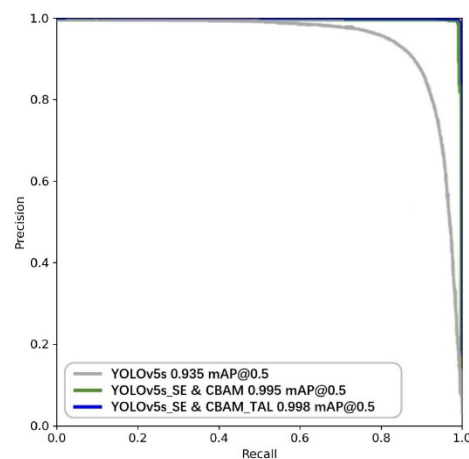


Figure 8. The changing trend of mAP with different improved algorithms.

3.5. Experimental data set and experimental environment deployment

There are two main sources of anti-bird collision detection datasets: on the one hand, the airfield

scene is captured by infrared surveillance shooting, and then the video is sliced and processed using python code [19]; on the other hand, valuable data are obtained in the Internet by python web crawlers [20]. Data cleaning is performed on the acquired large number of images in batch to remove the invalid images with low resolution and no detection target among them. The algorithm in this paper is to prevent the occurrence of bird collisions by identifying two major categories from representative small birds and airplanes, and adding difficult samples to improve the accuracy of detection. Labeling annotation software is used to annotate the cleaned data, and there are more than 5000 images in the annotated dataset.

As deep learning research progresses, the YOLOv5s algorithm tends to generate a large number of parameters during training and inference, thus requiring a computer with powerful computing power. Therefore, in this paper, the GeForce RTX 3070 Lite Hash Rate graphics card is selected for environment construction, and the YOLOv5s algorithm is deployed on Ubuntu 20.04 operating system. By configuring CUDA and CUDNN environments, we not only realize the parallel computing capability of GPU for data, but also speed up the training speed and improve the model accuracy. The system hardware and software configurations are shown in Table 5.

Table 5. System hardware and software configuration.

Operating System	Ubuntu 20.04.4 LTS
CPU	11th Gen Intel(R) Core (TM) i7-11700 @ 2.50 GHz × 16
Video Cards	GeForce RTX 3070 Lite Hash Rate
Memory	15.5 GiB
Graphics	NVIDIA Corporation
CUDA	11.4

This paper implements the training and testing of neural networks by porting the trained pt files on YOLOv5 to the Jetson Nano platform, a Linux system that includes advantages such as small size, powerful performance, and support for a range of popular AI algorithms. It has better application prospects in target detection. Therefore, Jetson Nano provides powerful support to enable real-time detection of flying birds, and the technical features of the embedded system in this paper are shown in Table 6. The trained pt weight file of YOLOv5s is only about 15 MB, and this greatly reduces the storage and processing capacity of the model. The platform can effectively detect small flying bird targets in different complex low-altitude traffic scenarios, and reduce the problem of insufficient real-time detection effect brought by delay. It has practical research significance and scenario application value.

Table 6. Technical characteristics of embedded systems.

Device	Technical features	
NVIDIA Jetson Nano	Processor	ARM Cortex A57, 1.42 GHz
	GPU	128 CUDA cores, 472 GFLOPS
	Memory	4 GB 64-bit LPDDR4 @ 25, 6 GB/s
	Power supply	20 Watt, 5V/4 A
	Size	100 × 80 × 29 mm

3.6. Visualization experiment analysis

The results are visualized in this paper and shown in Figure 9. For the baseline detector, the predicted bounding box encounters a severe lag. The faster a small bird moves, the greater the change in prediction. For small 5×5 objects like sparrows, the overlap between the prediction frame and the GT boxes becomes small or even absent. In contrast, the method in this paper mitigates the mismatch between the prediction frame and the moving object and fits the results accurately.

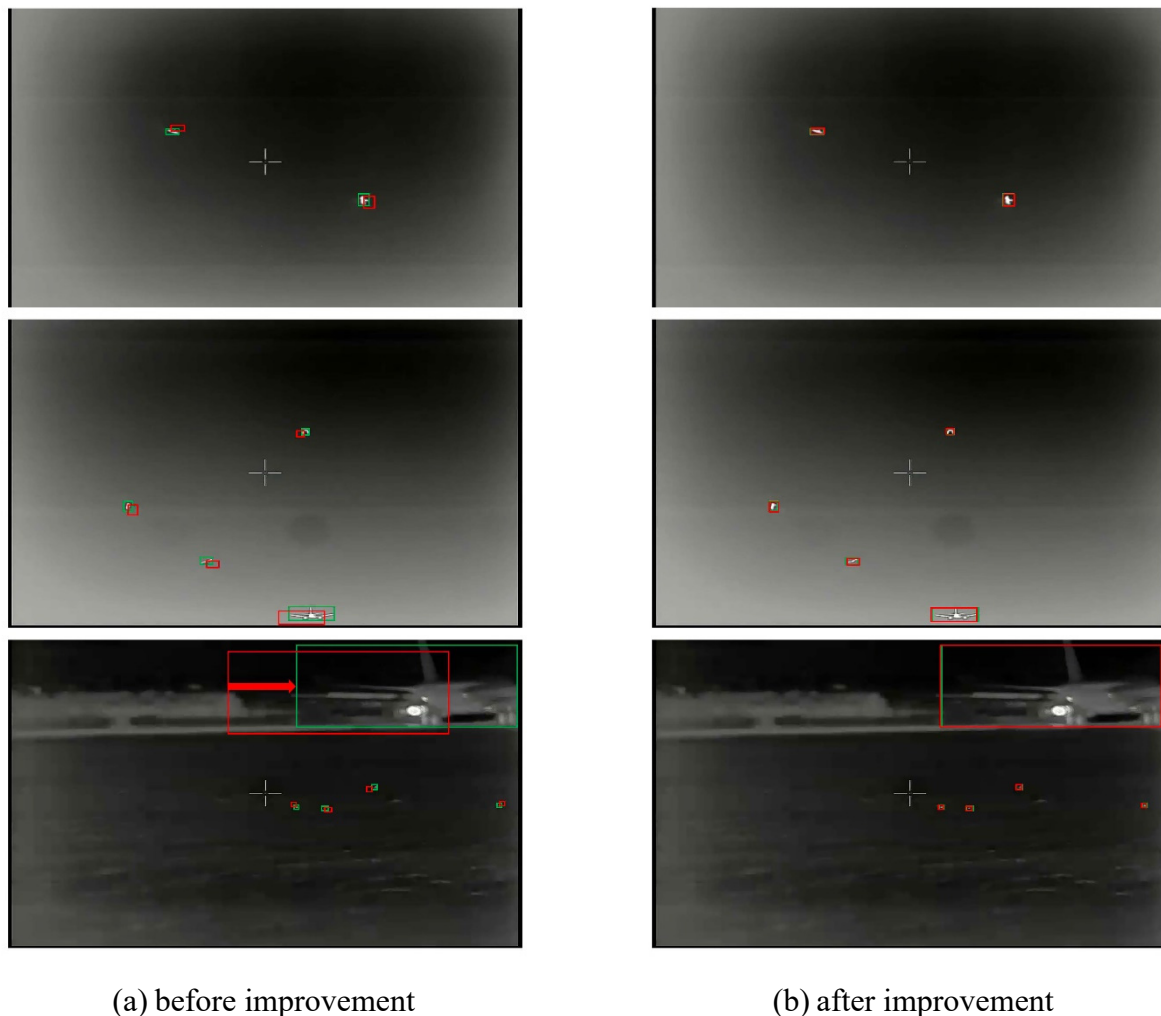


Figure 9. Detection results of the improved algorithm.

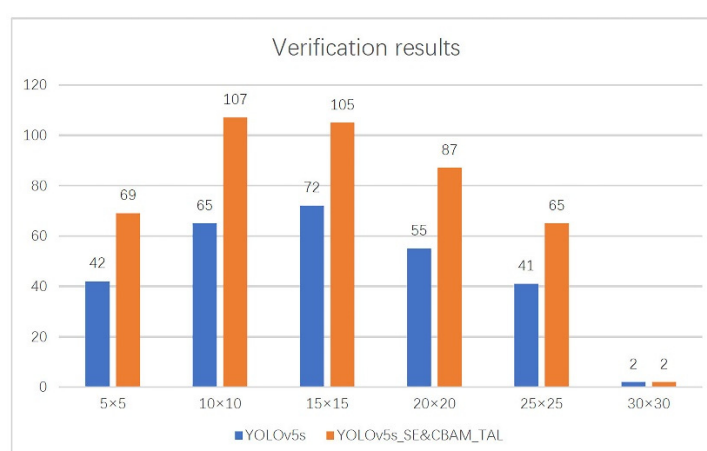
3.7. Experimental validation of YOLOv5s_SE&CBAM_TAL

In order to verify the performance of the algorithm here, this paper finds the difficult samples as the images of small sparrows and eagles as the test set for testing. Large number of valuable samples and difficult samples are selected here, and the image size is 960×576 , with the number of samples totaling 500. The sample sizes are: 5×5 for 75; 10×10 for 124; 15×15 for 123; 20×20 for 97; 25×25 for 79; 30×30 for 2. The samples are very small compared to the image size, and the results are shown in Table 7.

Table 7. Aggregate experimental results.

Sample sizes	5×5	10×10	15×15	20×20	25×25	30×30
Number of samples	75	124	123	97	79	2
YOLOv5s	42	65	72	55	41	2
YOLOv5s_CBAMA	55	95	102	70	58	2
YOLOv5s_SE & CBAM	65	102	100	92	64	2
YOLOv5s_SE&CBAM_TAL	69	107	105	87	65	2

In summary, among the 500 detected samples, the original network YOLOv5 detected only 277, and the improved YOLOv5s_SE & CBAM_TAL detected 435, which is 158 more than the original network, as shown in Figure 10.

**Figure 10.** Experimental verification results.

3.8. Comparison of several classical networks

In order to verify the advantages of the YOLOv5s_SE&CBAM_TAL model over other networks in this paper, commonly used models are selected for comparative performance analysis, and the training results are shown in Table 8.

Table 8. Comparison of the effects of different models.

Network model	Model volume /MB	FPS	mAP
Faster-RCNN	335	21.5	75.88%
SSD	315	57.5	79.72%
YOLOv3	256	34.6	76.38%
YOLOv4 [21]	253	36.8	86.15%
YOLOv5s	17	46.3	93.50%
YOLOv5s_SE&CBAM_TAL	15	65.5	99.80%

The result above shows that the mAP of the algorithm proposed in this paper improves by 6.3% over the original YOLOv5s, outperforming comparing to other network models in the same situation.

The improved algorithm, with a trained pt weight file of only about 15 MB and an FPS of 65.5, greatly reduces the storage and processing capacity of the model, thus allowing for real-time and fast detection of traffic elements on road traffic.

4. Conclusions

In scenarios such as airports, advance use of ultrasound to drive away birds is the basis for preventing bird collisions in presence. Real-time acquisition of flight paths of birds and aircraft using infrared surveillance photography, detection and identification are performed. A target detection method for bird collision prevention is proposed here. Different attention mechanisms SE and CBAM are integrated in the algorithm of this paper on YOLOv5s network to solve the problems of small target miss detection, false detection and insufficient feature extraction capability. TAL and W_i are used to solve the drift of the prediction frame. After comprehensive ablation experiments, the improved YOLOv5s_SE&CBAM_TAL algorithm has significantly improved the detection accuracy and speed, as the mAP value reaches 99.8% and a 6.3-percent increase compared with the original algorithm. Finally, the trained weights are deployed on the embedded system Jetson Nano platform. The platform is able to effectively detect small flying bird targets in different complex low-altitude traffic scenarios and reduce the problem of insufficient real-time detection effect due to delay. The platform has practical research significance and scenario application value.

Acknowledgments

The authors would like to thank the reviewers for their comments and suggestions, which helped to improve the manuscript. This work was supported by General Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (2022SJYB0712) and Research Development Fund for Young Teachers of Chengxian College of Southeast University (z0037).

Conflict of interest

The authors declare no conflict of interest.

References

1. D. Aleksandra, I. Cavka, V. Cokorilo, Bird strikes on an aircraft and bird strike prevention, *Tehnika*, **2** (2014), 291–298. <https://doi.org/10.5937/tehnika1402291D>
2. A. T. Marques, H. Batalha, J. Bernardino, Bird displacement by wind turbines: assessing current knowledge and recommendations for future studies, *Birds*, **2** (2021), 460–475. <https://doi.org/10.3390/birds2040034>
3. T. Wu, X. Luo, Q. Xu, A new skeleton based flying bird detection method for low-altitude air traffic management, *Chin. J. Aeronaut.*, **31** (2018), 2149–2164. <https://doi.org/10.1016/j.cja.2018.01.018>
4. X. Zhang, X. Wu, X. Zhou, X. Wang, Y. Zhang, Automatic detection and tracking of maneuverable birds in videos, in *2008 International Conference on Computational Intelligence and Security*, (2008), 185–189. <https://doi.org/10.1109/CIS.2008.46>

5. K. A. Klein, R. Mino, M. J. Hovan, P. Antonik, G. Genello, MMW radar for dedicated bird detection at airports and airfields, in *First European Radar Conference*, (2004), 157–160.
6. T. Yang, S. Z. Li, O. Pan, J. Li, Real-time and accurate segmentation of moving objects in dynamic scene, in *Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*, (2004), 136–143. <https://doi.org/10.1145/1026799.1026822>
7. Q. Hu, R. Li, Y. Xu, C. Pan, C. Niu, W. Liu, Toward aircraft detection and fine-grained recognition from remote sensing images, *J. Appl. Remote Sens.*, **16** (2022), 024516. <https://doi.org/10.1117/1.JRS.16.024516>
8. X. Yang, X. Fan, J. Wang, X. Yin, S. Qiu, Edge-based cover recognition and tracking method for an AR-aided aircraft inspection system, *Int. J. Adv. Manuf. Technol.*, **111** (2020), 3505–3518. <https://doi.org/10.1007/s00170-020-06301-x>
9. J. Yi, P. Wu, B. Liu, Q. Huang, H. Qu, D. Metaxas, Oriented object detection in aerial images with box boundary-aware vectors, in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (2021), 2149–2158. <https://doi.org/10.1109/WACV48630.2021.00220>
10. J. Feng, D. Ming, B. Zeng, J. Yu, Y. Qing, T. Du, et al., Aircraft detection in high spatial resolution remote sensing images combining multi-angle features driven and majority voting CNN, *Remote Sens.*, **13** (2021), 2207–2224. <https://doi.org/10.3390/RS13112207>
11. C. Wang, A. Bochkovskiy, H. M. Liao, Scaled-yolov4: Scaling cross stage partial network, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 13024–13033. <https://doi.org/10.1109/CVPR46437.2021.01283>
12. Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 13708–13717. <https://doi.org/10.1109/CVPR46437.2021.01350>
13. T. ting, L. Guo, H. Gao, T. Chen, Y. Yu, C. Li, A new time–space attention mechanism driven multi-feature fusion method for tool wear monitoring, *Int. J. Adv. Manuf. Technol.*, **120** (2022), 5633–5648. <https://doi.org/10.1007/S00170-022-09032-3>
14. H. Gong, L. Chen, H. Pan, S. Li, Y. Guo, L. Fu, et al., Sika deer facial recognition model based on SE-ResNet, *Comput. Mater. Con.*, **72** (2022), 6015–6027. <https://doi.org/10.32604/CMC.2022.027160>
15. H. Fu, G. Song, Y. Wang, Improved YOLOv4 marine target detection combined with CBAM, *Symmetry*, **13** (2021), 623–637. <https://doi.org/10.3390/sym13040623>
16. Z. Li, X. Huang, Z. Zhang, L. Liu, F. Wang, S. Li, Synthesis of magnetic resonance images from computed tomography data using convolutional neural network with contextual loss function, *Quant. Imag. Med. Surg.*, **12** (2022), 3151–3169. <https://doi.org/10.21037/qims-21-846>
17. D. Zhu, Overlapping boundary based multimedia slice transcoding method and its system for medical video and traffic video, *Multimed. Tools Appl.*, **75** (2016), 14233–14246. <https://doi.org/10.1007/s11042-015-3235-8>
18. Z. Tang, J. Li., Y. Zhou, Clothing information collection based on theme web crawler, *Int. J. Adv. Networking Appl.*, **10** (2019), 3919–3924. <https://doi.org/10.35444/IJANA.2019.10043>
19. J. Huang, D. Shao, H. Liu, Y. Xiang, L. Ma, S. Yi, et al., A lightweight segmentation method based on residual U-Net for MR images, *J. Intell. Fuzzy Syst.*, **42** (2022), 5085–5095. <https://doi.org/10.3233/JIFS-211424>

20. C. Wen, M. Hong, X. Yang, J. Jia, Pulmonary nodule detection based on convolutional block attention module, in *2019 Chinese Control Conference (CCC)*, (2019), 8583–8587. <https://doi.org/10.23919/ChiCC.2019.8865792>
21. H. Alqaysi, I. Fedorov, F. Z. Qureshi, M. O'Nils, A temporal boosted YOLO-Based model for birds detection around wind farms, *J. Imaging*, **7** (2021), 227–240. <https://doi.org/10.3390/jimaging7110227>



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)