Computer Communications 206 (2023) 124–132

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

High-precision real-time UAV target recognition based on improved YOLOv4

Yuxing Dong [a,*], Yujie Ma [b], Yan Li [a], Zhen Li [a]

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, China
[b] Shanghai Electro-Mechanical Engineering Institute, Shanghai 201109, China

ARTICLE INFO

Keywords:
UAV
Photoelectric detection system
Superresolution reconstruction
Object detection algorithm

ABSTRACT

In recent years, unmanned aerial vehicles (UAVs) have gained widespread use in both military and civilian fields with the advancement of aviation technology and improved communication capabilities. However, the phenomenon of unauthorized UAV flights, or "black flying", poses a serious threat to the safe flight of aircraft in airspace and public safety. To effectively interfere with and attack UAV targets, it is crucial to enhance the detection and identification of "low, slow and small" UAVs. This study focuses on achieving high-precision and lightweight detection and identification of four-rotor, six-rotor, and fixed-wing UAVs in low-altitude complex environments. By combining deep learning target detection with superresolution feature enhancement, a lightweight UAV detection model is designed and field-tested for verification. To address the challenge of detecting small UAV targets with limited information, the feature fusion network is enhanced based on the traditional YOLOv4 algorithm to improve the detection ability of small targets via small target enhancement and candidate box adjustment. The feasibility of the improved network is quantitatively and qualitatively analyzed. Channel pruning and layer pruning are then applied to the network, significantly reducing its depth and width and realizing a lightweight network. Finally, reasoning quantification is conducted on the embedded platform to enable end-side deployment of the target detection algorithm.

1. Introduction

In recent years, with the rapid development of aviation technology and the upgrading of communication technology, UAV products and functions have emerged in endlessness. From aerial photography, geological surveys, vegetation irrigation, and fire fighting operations to combat reconnaissance, UAVs are closer to life, but they also produce new threats [1,2].

There are two main categories of low-altitude UAV threats. (1) Passive threats due to communication failures. Because the UAV is out of control, it may pose a threat to the safety of flying objects in the nearby airspace and people and property on the ground. For example, in May 2017, due to the interference of nearby drones, the normal flight of many aircraft at Kunming Airport was seriously threatened, and the travel of tourists was disturbed. (2) UAV pilots use UAVs to maliciously invade military bases and secret-related sites, threatening the security of the country, important targets and the public. For example, the White House in the United States, the Japanese Prime Minister's official residence and other confidential places have been invaded by drones, hidden information has been stolen, and national security has been threatened. Moreover, armed drones are playing an increasingly important role in the battlefield. Therefore, to protect the personal and property safety of citizens, it is necessary to counter UAVs, and improving the detection and identification ability of "low, slow and small" UAVs is the key to detecting, interfering with and attacking UAV targets.

Due to the characteristics of low flight altitude, low speed flight or hovering, and small size, high-precision UAV detection is easily interfered with by birds, airborne objects, and complex environmental backgrounds. At present, the main detection methods of UAVs include radar, photoelectric, radio frequency and acoustic detection [3,4]. Radar realizes UAV detection by transmitting signals and receiving echoes. However, at low altitudes, radar receives more clutter, which is easily interfered with, and there is a blind field of view, which makes it difficult to detect small targets of UAVs. Rf detection is equally susceptible to clutter interference, while acoustic detection has a short detection range and high cost, making it unsuitable for UAV detection. By obtaining video in real time and observing the contour, size and other information of UAVs, photoelectric detection can effectively extract UAV features and has a fast detection speed, which can realize the detection and identification of close-range UAVs [5].

For small UAV targets, traditional computer vision methods have a poor ability to extract target features, are not easy to detect, and are prone to environmental interference, resulting in low target detection accuracy and poor application effects. The explosive development of deep learning in recent years has solved many difficulties and pain points of computer vision. In the field of UAV detection, using deep

* Corresponding author.
E-mail address: dongyuxing@126.com (Y. Dong).

https://doi.org/10.1016/j.comcom.2023.04.019
Received 22 December 2022; Received in revised form 14 April 2023; Accepted 18 April 2023
Available online 29 April 2023

learning methods to let the machine autonomously learn the contour, texture and other features of the UAV and perform effective feature extraction can greatly improve the detection accuracy of the UAV. At a long distance, UAV target imaging is small, the semantic information is insufficient, the target recognition effect is limited, and there will be certain misjudgment. Weak image enhancement through image super-resolution can effectively restore high-resolution detailed information and improve the accuracy of subsequent detection.

Anti-uav detection is developing toward high accuracy, intelligence and miniaturization. It is of great significance to study UAV detection algorithms and weak image enhancement algorithms based on deep learning for UAV detection and early warning.

The major contributions of this paper are summarized as follows.

(1) We investigate the improvement of the YOLOv4 object detection algorithm. According to the characteristics of UAV target detection, the UAV detection accuracy of the YOLOv4 network is improved by adding large-scale detection layers, improving anchors and enhancing small targets.

(2) The improved algorithm can provide more semantic information, such as contour and texture, and effectively improve the target detection effect.

(3) The mAP is improved to 86.7%, which realizes high-precision UAV small target detection.

The remainder of this paper is organized as follows. Section 2 discusses the related work in UAV Target Recognition. Section 3 presents the small UAV target detection algorithm based on the improved YOLOv4. Section 4 proves the YOLOv4 model improvement and optimization. In Section 5, we analyze the experimental results. Finally, Section 6 concludes our work.

2. Related work

Uav target detection is a typical moving target detection problem. Common detection methods include classical moving object detection based on the optical flow method, frame difference method and background modeling method [3], traditional object detection methods and object detection based on deep learning methods.

(1) Classical moving object detection method

Optical flow methods determine the relative motion of 3D objects by finding the change in imaging pixels in the time domain. The optical flow method can overcome the jitter of the lens itself and realize the detection of moving objects. Although the optical flow method has a good target extraction effect, its algorithm complexity is very large, which is not suitable for real-time detection, and it is greatly affected by illumination and sensitive to noise, which is not suitable for UAV detection.

The frame difference method detects the target by the difference of two or more frames, which easily causes the target hole, and the target extraction effect is poor. However, the frame difference method easily extracts the moving target. Yang Guodong et al. used the frame difference search method to detect UAVs. The pixel-level features of the image are integrated into the frame difference method, and the heuristic search of the selective search algorithm is used to segment the specific moving target to realize the detection of UAV small targets.

The background modeling method constructs the background model, differentiates the image from the background image, and obtains the moving target. It is greatly affected by illumination, and the deviation is large when the camera shakes. The detection effect is determined by background model modeling. Common modeling methods include Gaussian modeling and the Vibe algorithm. Wang Chuanyun et al. used Gaussian mixture background modeling combined with a compressed sensing domain for UAV target detection. First, the image is decomposed into blocks, and the Gaussian mixture background model is used to identify the candidate blocks that may contain the target in the compressed sensing domain. The background image and target image are separated by sparse matrix factorization to realize UAV small target detection, which improves the detection performance of the UAV.

(2) Traditional object detection methods

The traditional UAV target detection method uses the idea of a sliding window to traverse the whole image, extracts the features of the corresponding region, and uses the commonly used classifier to detect and classify the extracted features. In UAV detection, the selected UAV features and classifiers are very important. For Uavs, common features mainly include contour features, texture features, edge features, and other human-defined features. In practical applications, feature fusion is often used to fuse different features to improve the accuracy of object detection. Classifiers generally include SVM support vector machine, naive Bayes estimation and improved classifiers of related algorithms.

In the field of traditional target detection, much research has been carried out on the problems related to UAV detection. In the face of complex scenes, Lei L 2018 proposed an aerial target detection algorithm based on AM5728 with a dynamic complex background. The Lucas–Kanade (LK) optical flow method was used in the algorithm, and the median filter was used for the adaptive threshold segmentation method to predict the UAV under a dynamic complex background, which could realize the effective detection of UAVs with different scale changes and cloudy backgrounds. However, there is a problem that the detected target looks slightly larger than the actual target. . In the same year, Wang W et al. used multiscale decomposition filtering in space and multiscale difference processing in time to effectively detect UAV targets with complex backgrounds and different distances in scenes with unknown UAV sizes and speeds.

In selecting UAV features and classifiers, HOG features and SVM classifiers are commonly utilized. In 2019, Abu-Jamous M et al. employed a histogram of oriented gradients to extract UAV features and applied a support vector machine (SVM) to distinguish specific features of UAVs for detection and classification. The team combined image feature pyramids, non-local maximum suppression, geometric and illumination transformation of datasets, and other strategies to enhance UAV detection accuracy and achieved impressive results. [6]. On this basis, Xie X et al. proposed a detection and recognition algorithm based on block diagonal features in 2021. By reading the HOG features of the UAV and then using low-rank recovery technology for block diagonalization of the HOG features, the low-rank features of the UAV were constructed by introducing the block-diagonal sparse regularization term to increase the discrimination of the HOG features of the UAV. Then, SVM is used for classification to improve the accuracy of UAV identification [6,7]. In the realm of UAV detection, Xie J et al. proposed a novel spatiotemporal feature fusion method based on a data-driven SVM in 2021 to address the issue of low-contrast scenes in UAV detection. The method utilizes spatiotemporal contour features to describe the discontinuity of each pixel in the spatial and temporal domains, which are derived via the Black-hat filter and Ghostfree dark-focusing frame difference. The SVM classifier is trained using supervised spatiotemporal contours to automatically learn the spatiotemporal feature fusion mechanism, yielding excellent results in the detection of small, low-contrast UAVs. [8].

(3) Object detection method based on deep learning

Traditional target detection methods have difficulty extracting features when the scene is complex, and with the increase in extracted features, the calculation amount also increases greatly, which puts forward higher requirements for the real-time detection of UAVs. Since AlexNet was proposed in 2012, deep learning has been widely used in feature extraction and target classification [9]. A convolutional neural network (CNN) was used to replace manual extraction of features, and with the expansion and extension of the network, low-level contour, texture and other semantic information could be fully extracted [10,11]. The CNN network can effectively improve the detection accuracy of UAVs by replacing manual design and selecting UAV features.

The search algorithm was used to replace the traditional method of using a sliding window to extract regional features, and the detection

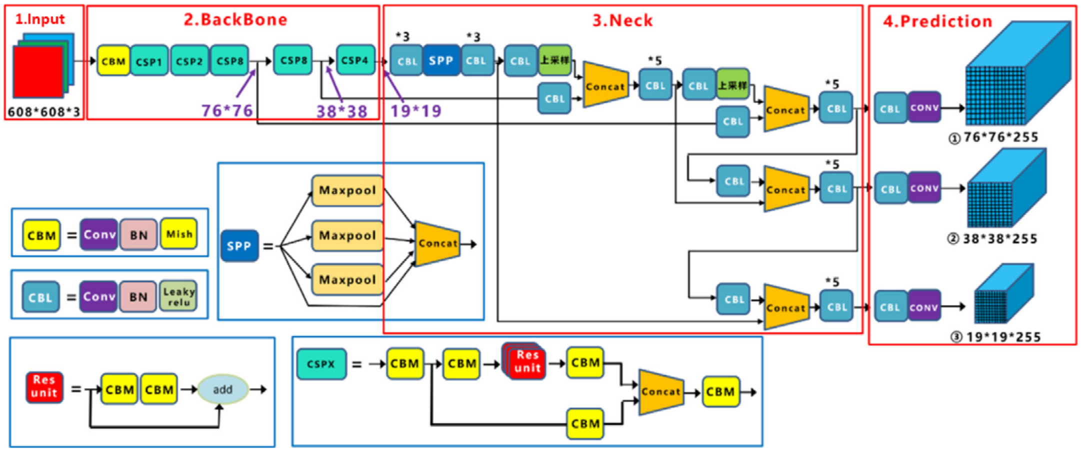


Fig. 1. Structure diagram of YOLOv4.

accuracy reached twice that of the traditional detection algorithm. In 2015, a simplified SPP-Net network [12] was introduced, softmax was used to replace SVM for target classification and localization, and Fast R-CNN [13] was proposed, which greatly improved the speed of target detection [12,13]. In 2016, on the basis of the original Fast R-CNN, the RPN network was used to generate candidate regions instead of the original region search algorithm, and Faster R-CNN was proposed to realize the CNN network implementation of the whole process of object detection, which further improved the detection speed [8,13]. According to the different target objects, Faster R-CNN also has many improvements and variations. In small target detection, Liu Jiahao et al. introduced the top-downtop feature pyramid fusion structure. Through the fusion of three different scale feature maps, the low-level detailed information and high-level semantic information were highly fused, the UAV features were enhanced, and the accuracy of small target detection was improved [14]. Cao et al. using the idea of two-stage detection, used bilinear interpolation to improve ROI pooling to solve the localization bias problem and used multiscale convolution feature fusion to improve the recognition effect [15,16].

The two-stage algorithm represented by the Faster R-CNN algorithm has greatly improved the detection accuracy, but the detection rate is still insufficient and cannot meet the requirements of real-time engineering. In 2015, Redmon et al. proposed the YOLO algorithm, which uses whole-process convolution for object discrimination and candidate box prediction [17]. After several years of development, the YOLO series has been widely used in various object detection problems by changing the network architecture, adding multiscale fusion, changing the activation function, optimizing the loss function and other aspects of improvement. Hu et al. used the last four scales of the YOLOv3 feature mAP instead of the original three scale feature map to predict the bounding box of the object, which could obtain more texture and contour information to detect small UAV objects. In the same UAV test sample, the MAP was improved by approximately 4.16% [18].

In 2019, Craye C utilized the U-Net architecture to identify small target flying object areas in large scenes and determined the target category in the area through the ResNet network, enabling identification of UAVs and birds. In addition to employing classical target detection networks, this approach offers an innovative solution to detecting small targets [14,19]. In 2020, Sun H et al. proposed the TIB-Net UAV lightweight detection network, which enhances the feature extraction ability of UAVs through the cyclic path structure and integrates the cyclic path into the EXTD detector. At the same time, the spatial attention module is integrated into the network backbone, which can filter the background noise and achieve good results [20]. In 2021, Ding L et al. deleted the low-resolution layer of the SSD network and enhanced the high-resolution network layer for the lack of texture and shape features of the infrared UAV. The adaptive pipeline filter (APF) based on temporal correlation and motion information was used to correct the results, which achieved higher accuracy and better robustness [21].

Although the UAV detection effect based on the deep learning method is good, the current UAV dataset is small, and the UAV target is small, which will cause the loss of detection accuracy to a certain extent. Image processing methods such as image superresolution reconstruction technology can be used to improve the detection accuracy.

3. Small UAV target detection algorithm based on improved YOLOv4

The YOLOv4 algorithm was proposed in 2020, and it performs well in the field of regular object detection. Taking the COCO dataset as an example, its AP is increased to 43.5%. Therefore, based on the YOLOv4 algorithm, combined with the characteristics of UAV small targets, this paper improves the algorithm to achieve effective UAV detection.

3.1. YOLOv4 object detection algorithm

Compared with YOLOv3, YOLOv4 is optimized in the backbone network, multiscale fusion, activation function, loss function, etc. The structure diagram of YOLOv4 is shown in Fig. 1.

3.2. Method process

(1) Backbone network: Based on DarkNet53 and referring to the idea of skip connection of the cross-stage local network (CSPNet), the Backbone network of CSPDarkNet53 is constructed. In Fig. 2, a parallel network is added to the original residual block for splicing and fusion, and skip connections are performed inside and outside the residual block, which enhances the feature extraction ability of the network and accelerates the training optimization of the network.

(2) Multiscale feature fusion: The SPP structure is used after the backbone network of CSPDarkNet53, and the SPP structure diagram is shown in Fig. 3.

Max pooling at four scales is used for feature processing to improve the receptive field size, and the context features are separated under the premise of ensuring the speed of the network. After SPP, the PANet network is used to achieve feature map fusion of different scales and sizes. Through repeated feature extraction, the feature extraction ability of the YOLO network for different sizes of objects is effectively enhanced. PANet is a top-down extraction after bottom-up feature extraction from FPN.

Max pooling at four scales is used for feature processing to improve the receptive field size, and the context features are separated under the premise of ensuring the speed of the network. After SPP, the

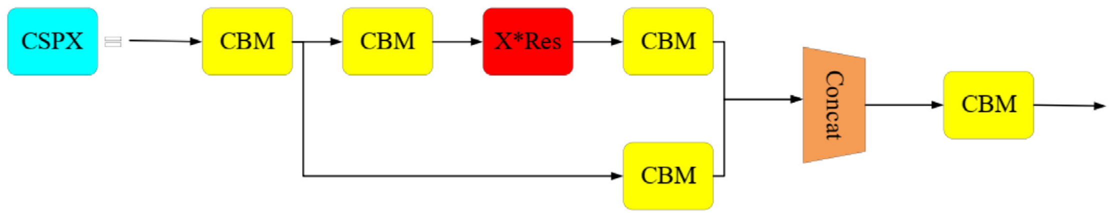


Fig. 2. CSP structure diagram.

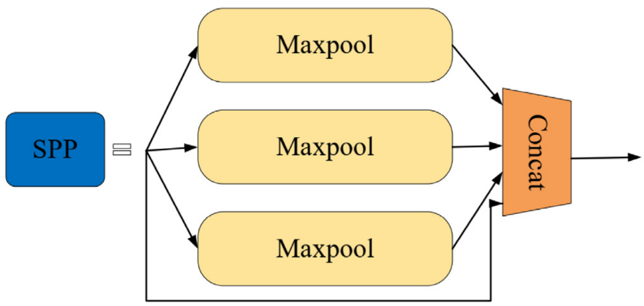


Fig. 3. SPP network structure.

PANet network is used to achieve feature map fusion of different scales and sizes. Through repeated feature extraction, the feature extraction ability of the YOLO network for different sizes of objects is effectively enhanced. PANet is a top-down extraction after bottom-up feature extraction from FPN, and its schematic diagram is shown in Fig. 4.

(3) Loss function: The loss function continues the composition of the YOLOv3 loss function but introduces CIoU instead of IoU in the position loss function. On the basis of considering the intersection and union ratio, the overlap area of the two boxes, the ratio of width to height, the distance of the center position and other factors are considered, so the CIoU expression is given by Formula (1):

$$R_{CIoU} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{1}$$

where c is the diagonal length of the intersection of the two boxes, and the expressions for v and α are as follows.

$$v = \frac{4}{\pi}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \tag{2}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{3}$$

Therefore, the expression for the position loss function is:

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj}[(1 - IoU) + \frac{d^2}{c^2} + \alpha v] \tag{4}$$

(4) Activation function: The Mish live function is introduced on the basis of LReLU, and the calculation expression is given by Eq. (5).

$$Mish(x) = x \times \tanh(\ln(1 + e^x)) \tag{5}$$

On the one hand, the Mish function is used to avoid gradient saturation, and on the other hand, it can solve the problem that ReLu does not activate negative numbers at all. Its schematic diagram is shown in Fig. 5.

4. YOLOv4 model improvement and optimization

For the same class of UAVs, the UAV target imaging is large at close range, while at long range, the imaging meets the criteria of small targets. Fig. 6 by counting the object size distribution in UAV data in common scenarios. There is a large difference in the size of the far-near target, but the size is mostly concentrated within 100 × 100. Although YOLOv4 has an AP of 43.5% on the COCO dataset and the network has

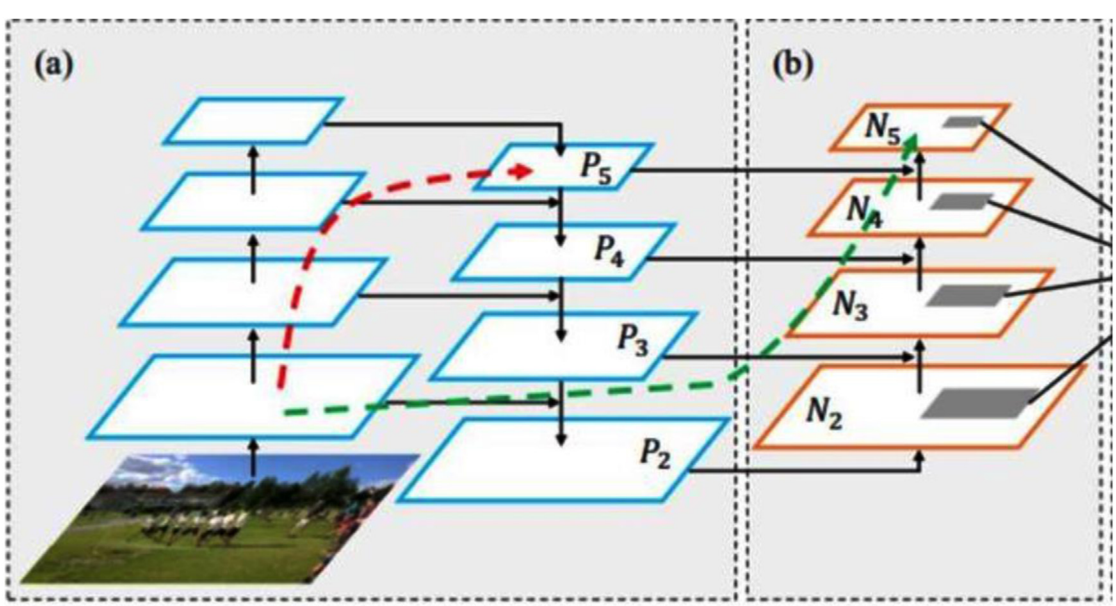


Fig. 4. Schematic diagram of the PANet.

Table 1
Comparison of output sizes of multiscale feature maps.

algorithm	Scale1	Scale2	Scale3	Scale4
YOLOv4	13 13 × 255 ×	26 26 × 255 ×	52 × 52 × 255	
Improved YOLOv4	13 × 13 × 18	26 * 26 * 18	52 × 52 by 18	104 × 104 × 18

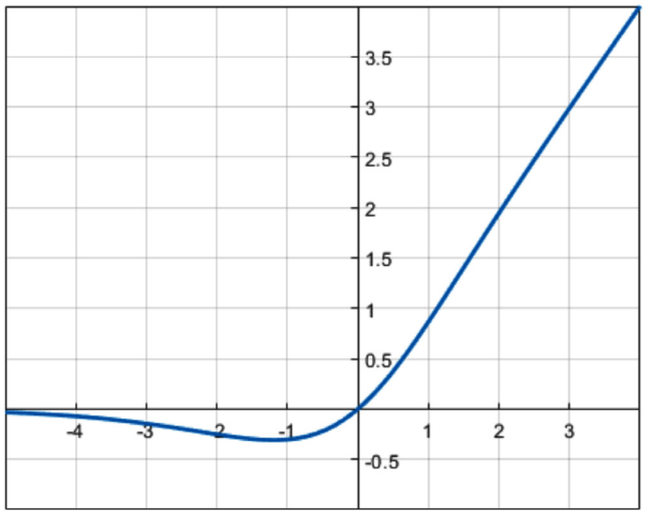


Fig. 5. Schematic diagram of the Mish activation function.

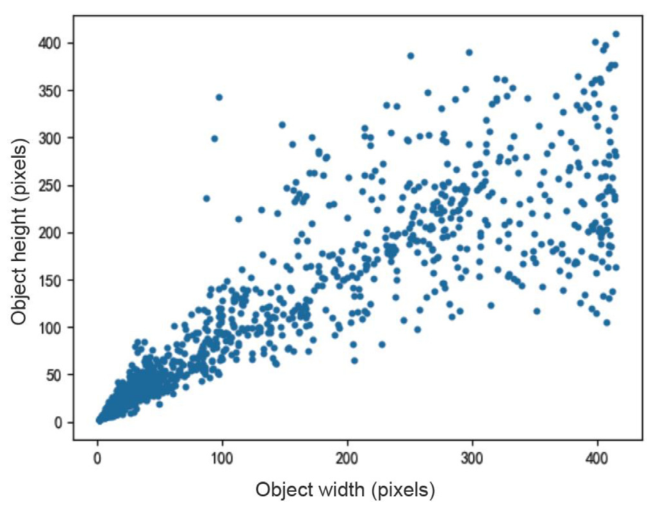


Fig. 6. Object size distribution diagram.

been optimized for small object detection, there are still some problems in "low, slow and small" UAV small object detection:

(1) The feature map of the deep network is used for classification and prediction, which contains few small object features, and the accuracy of small object detection is limited.

(2) The feature extraction layers are deep, and the UAV target features are few, so the UAV features are easily lost in the feature extraction process.

(3) The anchor used in the algorithm has weak generalization ability for small targets.

In response to the above problems, the YOLO model is improved from the aspects of network structure, small target enhancement and candidate box adjustment to improve the accuracy of UAV small target detection [22].

4.1. Improved YOLOv4 network structure

In Fig. 7, the YOLOv4 network uses CSPDarknet53 feature extraction, which is down sampling 32 times after 5 times. SPP and PAN are used to increase the receptive field, and the image feature pyramid is constructed between the feature maps. This improved algorithm focuses on enhancing the feature fusion of YOLOv4 by upsampling the shallow feature map and concatenating it with the shallower UAV feature image. This process increases the output size of the 104 × 104 small object detection scale, resulting in a four-scale detection output from the detection module. The shallow feature maps contain valuable low-level information such as target shape and texture, while the deep feature maps contain rich high-level semantic information. The improved branch effectively utilizes both low-level and high-level semantic information from the backbone, while simultaneously realizing small object scale detection through a new detection layer.

In the head part, the final prediction is carried out through 3 × 3 convolutional layers and 1 × 1 convolutional layers, and the dimension of the output vector is 3 × (k+5) at each scale, where k is the output category, and there is only one category of UAV in this paper, so k is 1. Each position in the final output feature map is associated with three different anchor boxes, and 5 represents the localization prediction and classification prediction of the prediction box. The size of the input image is 416 × 416 (see Table 1).

By adding a large-scale detection layer for feature enhancement for small target detection, the performance of UAV target detection in far and near ranges is effectively balanced.

4.2. Candidate tuning

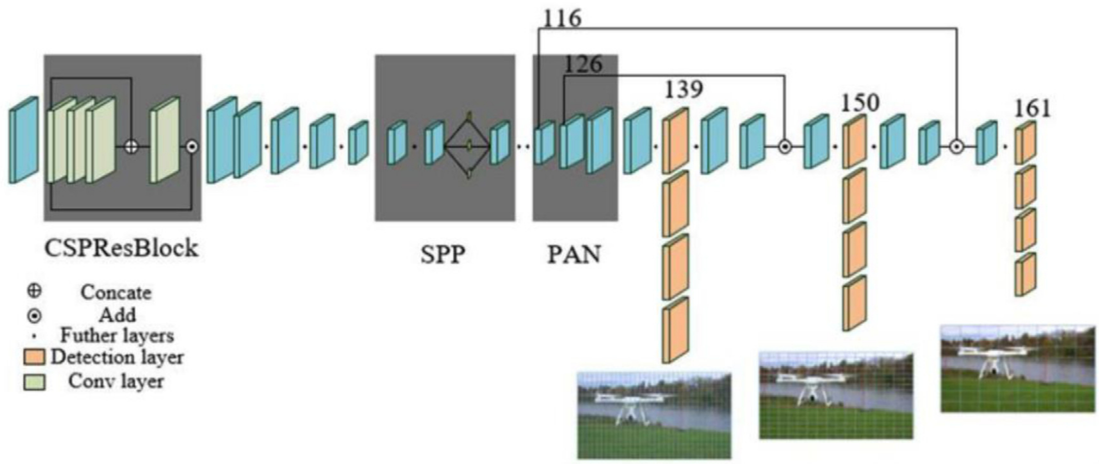
YOLOv4 uses K-means to cluster the true box sizes to obtain nine anchor boxes, and three anchor boxes are assigned to each scale. A higher value of K will improve the quality of the obtained preset anchor boxes for various objects and help the model converge during training. In the improved YOLOv4 algorithm, there are four scales of feature maps. To balance the performance of UAV target detection at near and far distances, three anchor boxes are still allocated to each scale feature map, and a total of 12 anchor boxes are set. However, K-means initializes different cluster centers and generates anchor boxes of different sizes, which will have different effects on the detection results. Therefore, taking UAVs as samples, the K-means++ algorithm is used to cluster the size of the labeling box.

The K-means++ algorithm mainly modifies the selection process for initial clustering centers. In this method, a real box in the training set (based on width and height) is randomly chosen as the first cluster center. The algorithm then calculates the Euclidean distances between the first cluster center and other real boxes in the training set. The box with a larger distance is more likely to be selected as the next cluster center, and this process continues until k-1 cluster centers are selected. The final clustering results are presented in Table 2. By making the anchor box of clustering more focused on small targets, K-means++ produces clustering results that are more consistent with real labels.

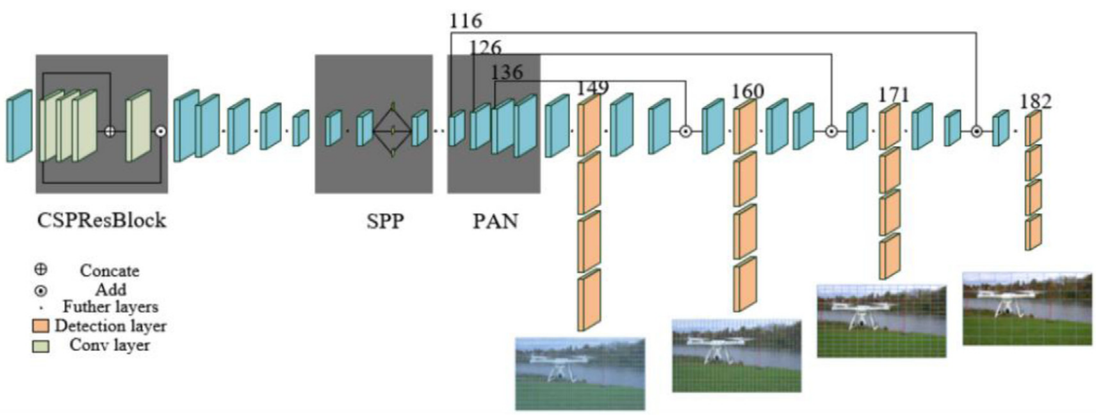
Small target augmentation.

(1) Small target Data Augmentation

Although YOLOv4 utilizes mosaic data augmentation, the random scaling of UAV targets can result in significant loss of target information and limited effectiveness. To address the challenges of small target datasets, this study implements traditional data augmentation techniques such as rotation, scaling, and cropping. Additionally, multiple UAVs can be copied into a single image to increase the number of



(a) Improve the former network structure



(b) The improved network structure

Fig. 7. Comparison diagram of network structure improvement.

Table 2
Anchor box clustering results under the same training set.

Algorithm	First layer	Second layer	Third layer	Fourth layer
YOLOv4	(282,242) (203,160) (110,95)	(73,58) (38,32) (21,16)	(15,24) (11,15) (8,11)	
Improved YOLOv4 (K-means)	(331,301) (292,238) (252,186)	(227,250) (195,159) (144,118)	(99,86) (74,57) (40,36)	(20,24) (12,17) (8,11)
Improved YOLOv4 (k-means ++)	(295,274) (261,196) (196,163)	(143,120) (103,87) (80,62)	(55,46) (40,30) (28,42)	(20,24) (12,17) (8,11)

targets, as shown in Fig. 8. This approach increases the number of matching anchor boxes and drives the model to place greater emphasis on small UAVs. As a result, the contribution of small UAVs to the loss function calculation is increased.

(2) Small target image enhancement

At a long distance, after optical magnification, the target still belongs to the range of small targets, the feature degradation is serious, the contour and texture features are not obvious, and the features of the target extracted by putting into the network are still very limited. Therefore, in this paper, the MFSRCNN method in Chapter 2 is used to reconstruct the low-resolution samples, generate high-resolution clear images, and restore the contour and texture features of the UAV to a certain extent. The clear pictures were extended to the training dataset to enhance the ability of the network to learn UAV features to improve the accuracy of UAV target detection.

5. Experimental results and analysis of algorithm improvement

The experimental environment in this section is the Windows 10 operating system, the hardware is an i7-7700 processor, and the GPU model is NVIDIA GeForce GTX 1080 Ti.

5.1. Training process

(1) Experiment to obtain the dataset.
Camera data: Sensor: 1/2.8 inch CMOS;
Focal length: 8.1 mm–310 mm (F1.8-F5.6).

Fig. 8. UAV data augmentation.

Fig. 9. Schematic diagram of the dataset.

The size of the acquired image is 1920 × 1080. (2) Download the UAV dataset from the Internet.

The UAV dataset includes the Drone Dataset (UAV) dataset and the Drone-data-2021 dataset. The dataset is shown in Fig. 9.

The initial UAV dataset contains 17,590 images, including the image features of various typical types of UAVs, such as quadrotor, six-rotor, and fixed-wing UAVs, under close- and long-distance test conditions. After small target data enhancement, 20,000 UAV data points are finally obtained to form the experimental dataset. We selected 80% of the data as the training set and the rest as the test set.

(2) Image annotation

The LabelImg tool was used to manually label the data, the UAV target was labeled as airplane, and the picture frame was saved as xml. The xml file holds the category and bounding box information of the annotated Uavs.

5.2. Training on data

(1) Hyperparameter settings

The size of the input image is uniformly transformed to 416 × 416. Because of the limited GPU memory, the batch size is set to 8, the learning rate is set to 0.002324, and the learning rate decays to a certain extent as the training epoch becomes larger.

Adam optimization strategy.

(2) Transfer learning

In the control experiment, the YOLOv4 network adopts the Fine-Tune method. First, the COCO dataset data are trained to obtain the YOLOv4 model corresponding to the large dataset, and then the UAV training set is trained to obtain the final model. Through transfer learning, training from 0 is avoided, and the training time is reduced.

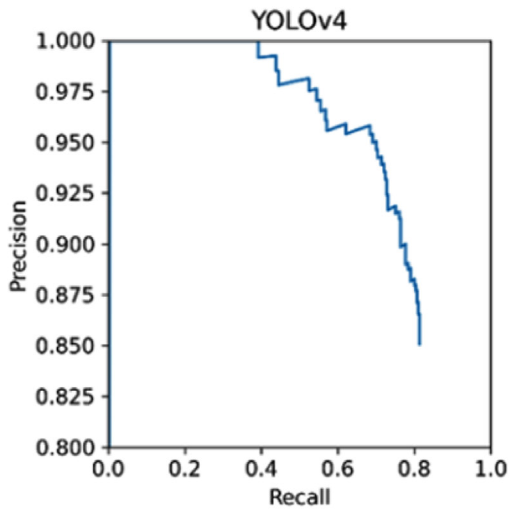
After training, both models achieve good fitting effects after more epochs, and the loss value decreases smoothly in the process of repeated training. The loss function value has an obvious decrease because the learning rate has an exponential decay, so the network learns more UAV information from the dataset and has a better convergence effect.

5.3. Analysis of experimental results

The model obtained by the algorithm before and after improvement was tested, the IoU threshold was set to 0.5, the precision rate, recall rate and F1 value were calculated, and Table 3 was obtained.

After the algorithm is improved, the accuracy and recall rate of UAV detection are improved to a certain extent. Results demonstrate that the improved model enhances detection accuracy and reduces the false detection of UAV targets. The improved model achieves an increase of 5.8% in F1 value, indicating a better detection performance compared to the original model.

The closer the PR curve is to the coordinate (1,1), the better the detection performance of the algorithm is. Fig. 10 illustrates that the improved algorithm yields an improved precision–recall (PR) curve for UAV targets compared to the PR curve generated by the original



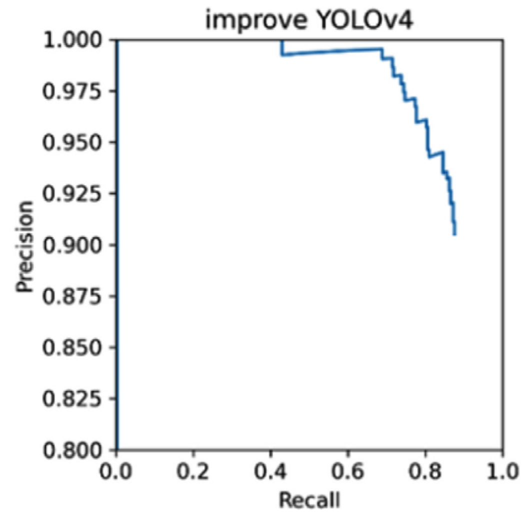


Fig. 10. Comparison of PR curves of different algorithms.

Table 3
Comparison of the algorithm before and after improvement.

Algorithm	Accuracy/%	Recall/%	F1 value/%
YOLOv4 algorithm	85.7	81.4	83.5
Improved YOLOv4 algorithm	91.2	87.5	89.3

Table 4
Comparison of model checking.

Algorithm	Faster R-CNN	SSD	YOLOv3	YOLOv4	Improve YOLOv4
mAP (%)	74.5	76.1	76.5	79.6	86.7
FPS	21.7	28.93	43.57	53.11	49.26

YOLOv4 algorithm. The PR curve of the improved algorithm entirely covers the curve of the original YOLOv4, suggesting that the improved algorithm exhibits stronger detection capabilities for UAVs.

Table 4 is obtained by reasoning on the UAV test data and setting the IoU threshold to 0.5.

The mAP of the original YOLOv4 model is 79.6%, which is higher than that of YOLOv3, SSD and other algorithms. Due to the characteristics of multiscale fusion, it has certain advantages in the recognition of UAV small targets.

The improved model achieves an mAP of 86.7%, representing an overall improvement of 7.1%. Furthermore, the improved algorithm significantly enhances the recognition effect of long-distance UAV small targets. On the NVIDIA GeForce GTX 1080 Ti platform, the original YOLOv4 model delivers an FPS of 53, while the improved algorithm's FPS is 49, indicating slightly slower detection speed. However, this minor speed sacrifice results in improved detection accuracy. Experimental results demonstrate that the improved YOLOv4 algorithm outperforms YOLOv3, SSD, and other algorithms in terms of mAP and FPS. The algorithm strikes a balance between FPS and mAP, exhibiting stronger detection capabilities for UAV small targets.

5.4. Comparative analysis of experiments

The feature map used for prediction in the YOLOv4 network only contains three scales. Due to the lack of rich semantic information such as details and contours of small objects during feature fusion, false detection occurs, and some UAV targets may be missed during detection. However, in the improved YOLOv4 algorithm, the new high-resolution feature map (receptive field 4×4) is used to introduce the scale prediction of UAV small targets, increase the detailed information of small targets, and increase the feature information of UAV detection. Therefore, the improved method can reduce missed detections with fewer false detections, the position of the bounding box is more accurate, and the size is closer to the actual size of the UAV.

6. Conclusions

Our work addresses the limitations of the YOLOv4 network for small UAV target detection by enhancing the accuracy of the network through the addition of large-scale detection layers, anchor improvements, and enhanced small target images. Experimental validation shows an overall increase of 7.1% in mean average precision (mAP) to 86.7%, with improved precision and recall rates. The improved algorithm is capable of providing more semantic information and focuses on contour and texture details to enhance object detection effectiveness.

7. Future work

While the photonic system has successfully passed the joint debugging test and acceptance, there are still areas for improvement in the system. These include

(a) The current unmanned aerial vehicle (UAV) detection system needs further improvement in terms of scale and background contrast. To increase the accuracy of detecting small targets, especially in non-sky backgrounds that may be affected by interference from buildings or mountains, a three-dimensional scan using a laser radar can be employed to provide distance information.

(b) The current tracking algorithm in the system is achieved through a tracking board, which leaves room for further research in UAV tracking algorithms to achieve more stable and sustainable tracking.

(c) By utilizing the parameters of the camera, information about the detected UAV, and the background scene, reinforcement learning can be employed to enhance the system's environmental awareness and intelligent response, ultimately improving automation and reducing the need for human intervention.

CRediT authorship contribution statement

Yuxing Dong: Conceptualization, Methodology, Software, Writing – review & editing. **Yujie Ma:** Data curation, Writing – original draft. **Yan Li:** Data curation, Writing – original draft. **Zhen Li:** Visualization, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

Data will be made available on request.

References

[1] J. Shi, P. Cong, L. Zhao, et al., A two-stage strategy for UAV-enabled wireless power transfer in unknown environments, IEEE Trans. Mob. Comput. (01) (2023) 1–15.
[2] Z. Shao, G. Cheng, J. Ma, et al., Real-time and accurate UAV pedestrian detection for social distancing monitoring in COVID-19 pandemic, IEEE Trans. Multimed. 24 (2021) 2069–2083.
[3] M.A. Mohammed, M.K. Abd Ghani, N. Arunkumar, et al., A real time computer aided object detection of nasopharyngeal carcinoma using genetic algorithm and artificial neural network based on haar feature fear, Future Gener. Comput. Syst. 89 (2018) 539–547.
[4] Z. Yu, Y. Shen, C. Shen, A real-time detection approach for bridge cracks based on YOLOv4-FPM, Autom. Constr. 122 (2021) 103514.
[5] Y. Cai, T. Luan, H. Gao, et al., YOLOv4-5D: An effective and efficient object detector for autonomous driving, IEEE Trans. Instrum. Meas. 70 (2021) 1–13.
[6] S. Cao, D. Zhao, X. Liu, et al., Real-time robust detector for underwater live crabs based on deep learning, Comput. Electron. Agric. 172 (2020) 105339.
[7] M. Feng, H. Lu, Y. Yu, Residual learning for salient object detection, IEEE Trans. Image Process. 29 (2020) 4696–4708.
[8] Xie Jiayang, Gao Chengxing, Wu Junfeng, et al., Small low-contrast target detection: Data-driven spatiotemporal feature fusion and implementation, IEEE Trans. Cybern. (2021) 1–12.
[9] J. Hao, S. Luo, L. Pan, Computer-aided intelligent design using deep multiobjective cooperative optimization algorithm, Future Gener. Comput. Syst. 124 (2021) 49–53.
[10] Y. Wu, L. Zhang, Z. Gu, et al., Edge-AI-driven framework with efficient mobile network design for facial expression recognition, ACM Trans. Embed. Comput. Syst. (2018).
[11] C. Chen, C. Wang, B. Liu, et al., Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles, IEEE Trans. Intell. Transp. Syst. (2023).
[12] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al., Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.
[13] R. Girshick, Fast R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.
[14] J. Liu, Z. Wang, Y. Wu, An improved faster R-CNN for UAV-based catenary support device inspection, Int. J. Software Eng. Knowl. Eng. (2020).
[15] Cao Changqing, Wang Bo, Zhang Wenrui, et al., An improved faster R-CNN for small object detection, IEEE Access 7 (2019) 106838–106846.
[16] Y. Zhang, F. Zhang, Y. Jin, et al., Local correlation ensemble with GCN based on attention features for cross-domain person re-ID, ACM Trans. Multimedia Comput., Commun. Appl. 19 (2) (2023) 1–22.
[17] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, et al., Selective search for object recognition, Int. J. Comput. Vis. 104 (2) (2013) 154–171.
[18] Y. Hu, X. Wu, G. Zheng, et al., Object detection of UAV for anti-UAV based on improved YOLO v3, in: 2019 Chinese Control Conference, CCC, Technical Committee on Control Theory, Chinese Association of Automation, 2019, pp. 8386–8390.
[19] C. Craye, S. Ardjoune, Spatio-temporal semantic segmentation for drone detection, in: 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS, IEEE, 2019, pp. 1–5.
[20] Sun Han, Yang Jian, Shen Jiaquan, et al., TIB-net: Drone detection network with tiny iterative backbone, IEEE Access 8 (2020) 130697–130707.
[21] Ding Lianghui, Xu Xin, Cao Yuan, et al., Detection and tracking of infrared small target by jointly using SSD and pipeline filter, Digit. Signal Process. 110 (2021) 102949.
[22] Xue Qiao, Zhu Youwen, Wang Jian, Joint distribution estimation and Navie Bayes classification under local differential privacy, IEEE Trans. Emerg. Top. Comput. 9 (4) (2021) 2053–2063.