Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 8, No. 1, 2024

A Lightweight Object Detection Network for UAV Aerial Images

Lin Tang [1, 2, *], Shunyong Zhou [1, 2], Xinjie Wang [1, 2]

[1] School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin Sichuan, 644000, China
[2] Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yibin Sichuan, 644000, China
[*] **Corresponding author**: Lin Tang (Email: 690398280@qq.com)

Abstract: In order to solve the problems of poor detection algorithms, high network model complexity, and difficult deployment of algorithms in the field of aerial image target detection. In this paper, based on YOLOv7-tiny algorithm, a lightweight target detection network for UAV aerial images is designed. Partial convolutional PConv is introduced into the network, and the feature extraction block ELAN is improved, which reduces the computational volume of convolution and the number of model parameters in the feature extraction process, thus solving the problem of model lightweight. The feature fusion part of the network is optimal to improve the feature extraction ability of the network for small targets. At the same time, the large target detection layer in the original network is replaced with the small target detection layer in the aerial images, and the attention mechanism is embedded in the backbone network, which solves the problem of imperfect detection algorithms in aerial images. The loss function of the network is improved so that the prediction frames generated by the detection network and the truth frames match each other in the regression process, thus improving the training process of the network. The experimental results on the publicly available dataset VisDrone2019 dataset show that compared with the YOLOv7-tiny algorithm, the detection accuracy of the proposed model is improved by 0.7%, the recall R is improved by 2.2%, the F1 value is improved by 1.6%, the average detection accuracy mean is improved by 2.3%, and the number of parameters is reduced by 52.1%. Moreover, the image detection speed FPS reaches 66/f.s-1, which meets the real-time requirements of the aerial image detection model detection, and provides a research idea for the field of UAV aerial image detection.

Keywords: UAV; YOLOv7-tiny; Loss Function; Lightweight Model.

1. Introduction

In recent years, with the development of UAV technology becoming more and more mature, the aerial photography system composed of UAV equipped with camera has been widely used in both military and civil fields because of its advantages of small size, convenient carrying and simple operation, and also plays a very important role. In the military field, UAV is often used as a reconnaissance tool. By acquiring aerial images over the target area, the relevant target information of a specific area can be obtained. In the civil field, UAV is often used as an efficient image acquisition system, which is applied to large area and long distance inspection and search and rescue [1], smart city traffic inspection [2], agricultural planting and other fields [3-5].

With the rich application scenarios of UAV, the target detection field of UAV aerial images has attracted many researchers to study the target detection in UAV aerial images. And with the rapid development of computer vision technology and deep learning, the target detection algorithm based on deep learning has defeated the traditional target detection algorithm in many target detection fields by virtue of its excellent target detection performance, and become the mainstream target detection technology at present. However, in the actual use of UAV aerial photography, due to the impact of UAV flying attitude, flight height, shooting Angle and other factors, the target size in the images obtained by UAV aerial photography is small, the target distribution is relatively dense, and there is a blocking phenomenon between the targets, which makes the current mainstream detection algorithm when detecting the targets in the aerial images. The key information of the target can not be extracted well, so that the algorithm is prone to false detection and missed detection in the process of aerial image detection, which makes the target difficult to be accurately observed in the aerial image. At present, the mainstream detection algorithm model is large, and the requirements for the deployment of hardware platform are high. However, the hardware platform of the actual application of UAV is limited in computing resources, which cannot well meet the deployment of the current mainstream target detection algorithm, which brings great challenges to the target detection of UAV aerial photography in the practical application.

Above all, to design a at the same time meet the requirements of uav aerial image target detection accuracy and easy hardware platform test model of deployment requirements, for unmanned aerial vehicle (uav) in the practical application in real life is of great significance. In this paper, based on the lightweight model of YOLOV7 network model YOLOv7-Tiny, relevant research is carried out to solve the problems that the number of detection model parameters is large and the model calculation is complex and difficult to deploy. The contributions of this paper are as follows:

(1)The ELAN module in the backbone network of the original network model is improved, and partial convolution PCONV is introduced to replace the basic convolution CONV in the original ELAN module, so as to reduce the redundancy and internal access in the model calculation, and reduce the number of parameters and calculation of the model.

(2)On the basis of the Head part of the original detection model, a layer of upper sampling is added to enrich the high-level semantic information of features, thereby increasing the expression ability of features in the detection network. In the detection head part of the model, the detection head for large

targets in the original network is removed, and the detection head for smaller targets in aerial images is added;

(3)The SE attention mechanism is introduced into the backbone network to enable the detection model to pay attention to more target feature information during the feature extraction process.

(4)The loss function of the network is improved, so that the improved model has better performance in convergence speed and detection effect;

(5)Through the corresponding ablation experiments and the comparison experiments with the current mainstream single-stage detection model, the superiority and effectiveness of the model described in this paper are proved.

The corresponding structure of this paper is arranged as follows: Section 2 describes the current object detection algorithms and the related work in this field. Section 3 is the detailed introduction and elaboration of the improved method proposed in this paper. Section 4 is devoted to the implementation details of the relevant tests conducted in this paper, as well as the analysis and discussion of the corresponding experimental results. Section 5 provides a summary of the paper and an outlook for future work.

2. Related Work

Now target detection algorithm based on depth of learning has been widely used in various areas, according to the principle of the detection algorithm of network structure and detection can be divided into three categories: the target detection algorithm based on the transformer, single phase target detection algorithm and double stage target detection algorithm [6]. Transformer-based object detection algorithms include DETR [7], DINO[8], Pix2seq[9], PVT v2[10], etc. Among them, DETR algorithm eliminates many requirements for hand-designed components, such as the constraint of no anchor prior box, and processing steps such as non-maximum suppression. It simplifies the process of object detection, but the algorithm also has some shortcomings, such as slow convergence speed in the training process and poor detection effect for small objects. Single-stage object detection algorithms include RetinaNet[11], SSD[12] series, and YOLO[13-18] series. Single-stage object detection algorithms do not need a proposal box in the detection process, and directly regression the target category probability and position coordinates, so the detection speed of this kind of algorithm is faster. However, the detection accuracy of the algorithm is relatively low. In contrast, the two-stage object detection algorithm first generates the region, and then classifies the samples through the convolutional neural network. Common two-stage object detection methods include R-CNN[19], SPP-Net[20], Fast R-CNN[21], Faster R-CNN[22] and R-FCN[23], etc.Double stage target detection algorithm unique architecture makes algorithm has high detection accuracy, and for small target detection effect is relatively better, but the detection speed is slow, cannot satisfy the needs of real-time detection.

At present, in order to meet the detection speed requirements of UAV aerial images in practical use, many researchers choose the YOLO series of single-stage detection algorithm as the research object to study the target detection in UAV aerial images [24].For example, Liu et al. [25] proposed the YOLOv3_ReSAM algorithm based on the YOLOv3-Tiny backbone network to solve the problem of difficult target detection caused by small target size and fuzzy appearance in long-distance aerial photography, and used the pyramid structure of image features to achieve multi-level feature fusion prediction. At the same time, in order to solve the information loss caused by the network convolution process, a spatial attention mechanism based on residual structure is proposed, and the idea of reinforcement learning is introduced to guide boundary regression. Tan[26] et al, aiming at the problem of difficult target detection in UAV aerial images, proposed a YOLOv4_Drone target detection model based on YOLOv4 algorithm. In this model, void convolution was used to resample feature images, and spatial attention mechanism ULSAM was used to derive feature maps to realize multi-scale representation of feature maps. Non-maximal Suppression (Soft-NMS) was introduced to reduce the false detection caused by target occlusion. Liu[27] et al. designed a GBS-YOLOv5 algorithm suitable for UAV detection to solve the problem of low accuracy of UAV detection of small targets. This algorithm is based on the YOLOv5 detection model and uses efficient spatio-temporal interaction modules to replace the residual network structure in the original network. The spatial pyramid convolution module was added to enable the network to extract more small target feature information. In order to better interact with high-order spatial semantic information, recursive gated convolution was introduced into the feature fusion module. Aiming at the problem of limited onboard computing resources of Uavs, Li et al. [28] based on YOLOv5 algorithm, studied the detection accuracy of aerial images and the computational cost of the algorithm, reduced the computational cost by replacing the convolution operation with linear transformation, and introduced a transformer to enhance the feature extraction of images, and proposed a GGT-YOLO algorithm. Aiming at the problem of large number of targets and high proportion of small targets in aerial images, Zhao et al. [29] carried out relevant research based on YOLOv7 network, and proposed a multi-scale UAV aerial image target detection model MS-YOLOv7 by using multiple detection heads and CBAM attention mechanism. And achieved good results and verify the applicability of the proposed algorithm; Zeng[30] et al. improved the YOLOv7 algorithm by removing the second down-sampling layer and the deep detection head, introducing the DpSPPF module and adding the NWD measurement method to the indicators of positive and negative sample allocation, and proposed a real-time small target detection algorithm YOLOv7-UAV.Aiming at the current challenges in the field of UAV target detection, Li[31]et al. proposed an aerial image detection model with excellent performance based on YOLOv8 algorithm. The model introduced the idea of Bi-PAN-FPN to improve the neck part of the original network, so that the network could consider and reuse multi-scale features. In the backbone network part, the GhostblockV2 structure was used to replace the C2f module in the original network to achieve the purpose of reducing the number of model parameters, and the WiseIoU loss was used as the boundary regression loss to improve the overall detection performance of the model.

Inspired by the above articles and related research results, this paper selects the YOLOv7-Tiny detection model suitable for edge computing platform in YOLOv7 algorithm to improve the lightweight and detection accuracy based on the problems of poor aerial image target detection effect and difficult detection model deployment in the field of UAV aerial photography.

3. Improved Lightweight Model

3.1. YOLOv7-tiny Network Architecture

The YOLOv7 target detection algorithm is a newer generation of YOLO series network model proposed by Wang [18] et al. based on the YOLOv5 algorithm, which further improves and optimises the algorithm's detection accuracy and speed. Like the YOLOv5 algorithm, the YOLOv7 algorithm is proposed in different versions for different usage scenarios and hardware support platforms. Among them, Yolov7-tiny is the network model suitable for edge computing platforms in the YOLOv7 series of algorithms. Its network structure consists of three parts: Input, Backbone, and Head, as shown in Figure 1. The backbone network consists of CBL, ELAN and MP modules, which are used to complete the feature extraction of the image. The Head part consists of SPPCSPC, Upsample, ELAN and CBL modules, which fuses the high-level features with the low-level features to realise the fusion of high-resolution information and high-level semantic information. And the corresponding detection heads are designed for large, medium and small detection targets in the image.

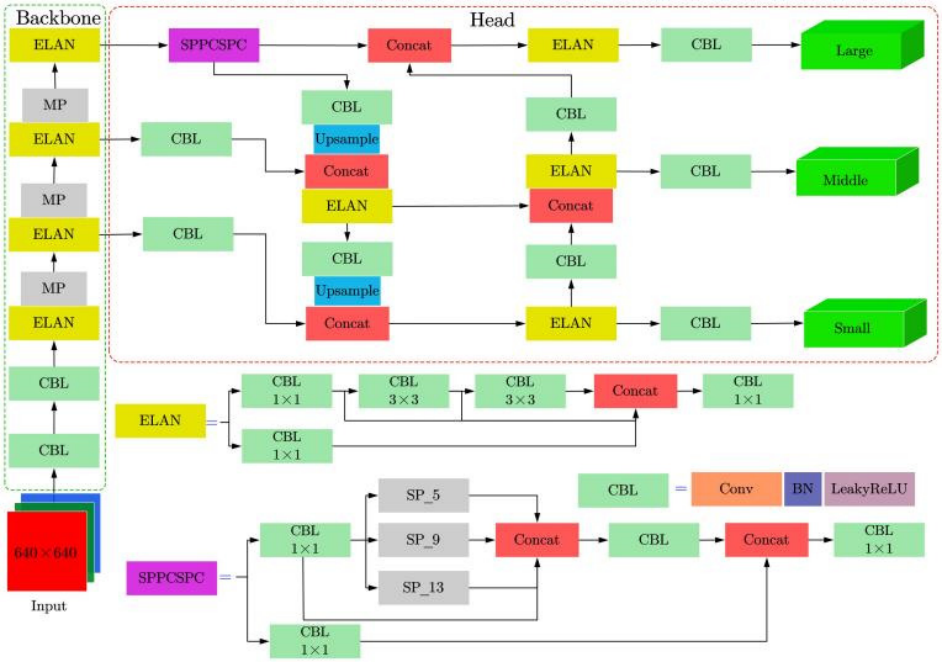


Figure 1. YOLOv7-Tiny network structure

3.2. Improvement of the ELAN Module

In order to reduce the number of floating-point operations and the number of memory accesses of the model in the detection process, CHEN [32] et al. proposed a new partial convolution PConv. By reducing redundant calculations and memory accesses at the same time, the convolution can extract spatial features more effectively. Based on partial convolutional PConv, a FasterNet neural network with faster running speed is further proposed to verify the detection effect of the new convolutional PConv in the network model. The reason why PConv can make the network have efficient and fast operation effect is that it only needs to apply filters on a part of the input channels to extract spatial features, and the rest channels remain unchanged. Figure 2 shows the different convolutional designs.

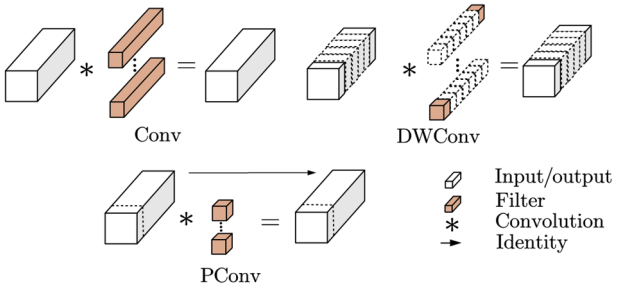


Figure 2. Convolution comparison diagram

Compared with the regular Conv, the partial conv has lower FLOPs. Relative to the depth of separable convolution DWConv part convolution PConv can in reducing network model and quantity and on the basis of the amount of calculation and to reduce memory access and reduce the cost of model checking effect; The improved network model can reduce the requirement of computing power of hardware equipment. In this paper, the ELAN module in the original network is improved by combining the partial convolution PConv.

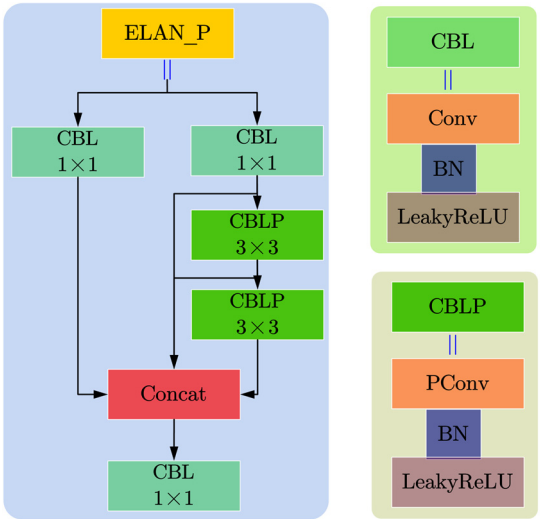


Figure 3. The improved ELAN_P module

The ordinary convolution Conv with the convolution kernel size of 3×3 in the original ELAN module is replaced by the partial convolution PConv with faster calculation speed, and the new ELAN_P module is obtained, as shown in Figure 3, and some ELAN modules in the original detection network are replaced. The detection model reduces the amount of calculation and access to memory in the feature extraction process of the backbone network, and realizes the improvement of lightweight model.

3.3. Network Structure Improvement

In the feature fusion part of the Head part of the original YOLOv7-tiny network, only the feature map is upsampled twice, which has a good detection effect for detecting large and medium-sized targets in the image. However, for small targets in UAV aerial images, such feature fusion methods cannot fully extract the feature information and corresponding semantic information of small targets in aerial images, so that the detection model is prone to false detection and missed detection when detecting small targets in aerial images. In order to make full use of the feature information of small targets in aerial images and improve the detection effect of the detection network for small targets in aerial images, an upsampling layer is added to the Head part of the original network to extract more target feature information. On this basis, a new detection head XSmall suitable for small targets in UAV aerial images is redesigned. and to test the model keep lightweight, decrease the number of arguments, to remove the original network for large target detection head, retained in the image is used for detection of medium and small target detection. The structure of the Head part after improvement are shown in Figure 4 below.

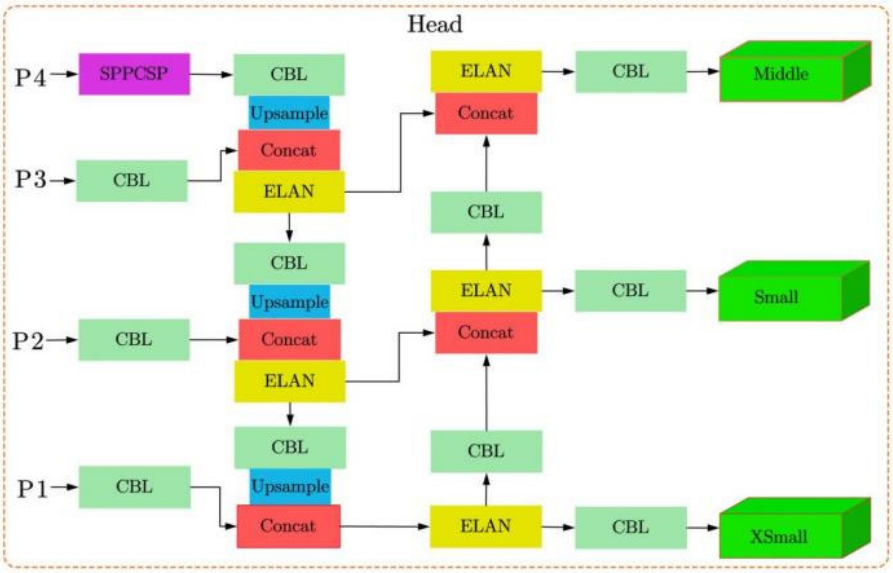


Figure 4. Improved Head section structure

3.4. SE Attention Mechanism

SE (Squeeze-Excitation) attention mechanism was proposed by Hu [33] et al. This attention mechanism uses a weight matrix to give different weights to different positions of the detection image from the perspective of the channel domain, so as to make the detection network pay attention to more target feature information, suppress complex background features, and improve the robustness of the network. and it does not add additional computing to the network. Figure 5 shows the structure of the SE attention mechanism.

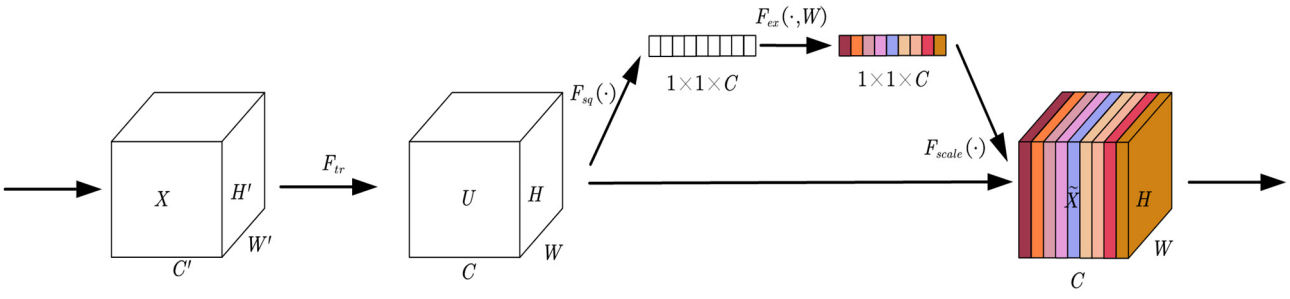


Figure 5. SE structure diagram

In the Figure 5, X is the original input data, H' is the height of the original input space, W' is the width of the original input space, C' is the number of original input channels, U is the feature map after the convolution operation, H is the space height after the convolution operation, W is the space width after the convolution operation, C is the number of channels after the convolution operation, and $\tilde{X}$ is the feature map with the final feature enhancement.

As shown in the figure, the SE attention mechanism works as follows: In the first step, the original feature map X of size $H'\times W'\times C'$ is input, and the feature map U of $H\times W\times C$ is generated after F_{tr} operation; In the second step, the squeeze operation is performed on $F_{sq}(\cdot)$ in the corresponding graph, and the feature map U is globally averaged pooled to generate a global feature of $1\times1\times C$. The corresponding formula is as

follows:

$$z = F_{sq}(U) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}U(i,j) \quad (1)$$

Where z is the global feature after pooling, i and j are the feature map size variables. In the third step, excitation operation is carried out, corresponding to $F_{ex}(\cdot,W)$ in the figure, to obtain the correlation of feature channels, retain the channels with large feature information, and suppress the channels with small feature information. So as to reduce the amount of calculation and improve the expression ability of feature information, the corresponding formula is as follows:

$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \sigma(W_2\delta(W_1 z)) \quad (2)$$

Where s is the weight normalization vector, W is the weight variable, W1 and W2 are the fully connected layer weights, σ and δ represent the Relu activation function and Sigmoid activation function, respectively. In the fourth step, the weight normalization vector s and feature map U are assigned weights, corresponding to $F_{scale}(\cdot)$ in the figure, and the feature map $\widetilde{X}$ that finally completes the SE attention weight assignment is obtained. The corresponding formula is as follows:

$$\widetilde{X} = F_{scale}(U, s) = sU \quad (3)$$

Through the structure diagram of SE attention mechanism and the corresponding relevant formula, it is easy to see that the size of the feature map is not changed by the assignment operation between the fourth step and the third step of the process realized by the attention mechanism. This makes the SE attention mechanism both high performance and efficiency, and also makes it widely used in the field of object detection.

3.5. Loss Function Improvement

The loss function adopted by the original YOLOv7-tiny network includes confidence loss, classification loss and regression loss of bounding boxes. The function used for the bounding box regression loss is the CIOU loss function, which was proposed by Zheng[34] et al. This function takes into account the overlap area between the prediction box and the truth box, the distance between the center points, and the aspect ratio, and is widely used in the current mainstream object detection algorithms. However, in the field of aerial image detection, the target to be detected accounts for a small number of pixels in the aerial image, which leads to the fact that the number of high-quality prediction boxes generated by the detection network is much less than that of low-quality prediction boxes in the regression process of the prediction box, and when the aspect ratio of the prediction box and the truth box is the same, the aspect ratio penalty term in the loss function will not play a role. This affects the training of the network.

In order to make the network better solve the matching problem between the prediction box and the truth box in the training, this paper chooses the SIOU loss function as the bounding box regression loss function of the detection network. The SIOU loss function [35] redefines the penalty metric based on the CIOU loss function and considers the vector Angle between the prediction box and the truth box, as shown in Figure 6. The loss function consists of four parts, namely Angle cost (Λ), Distance cost (Δ), Shape cost (Ω), and IOU cost (IOU). The relevant definitions are as follows:

$$L_{SIOU} = 1 - IOU + \frac{\Delta + \Omega}{2} \quad (4)$$

$$\Lambda = 1 - 2\times\sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \quad (5)$$

$$x = \frac{c_h}{\sigma} = \sin(\alpha), \sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \quad (6)$$

$$c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y}) \quad (7)$$

$$\Delta = \sum_{t=x,y}(1 - e^{-\gamma\rho t}) \quad (8)$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2, \rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2, \gamma = 2 - \Lambda \quad (9)$$

$$\Omega = \sum_{t=w,h}\left(1 - e^{-\omega_t}\right)^{\theta} \quad (10)$$

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (11)$$

$$IOU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (12)$$

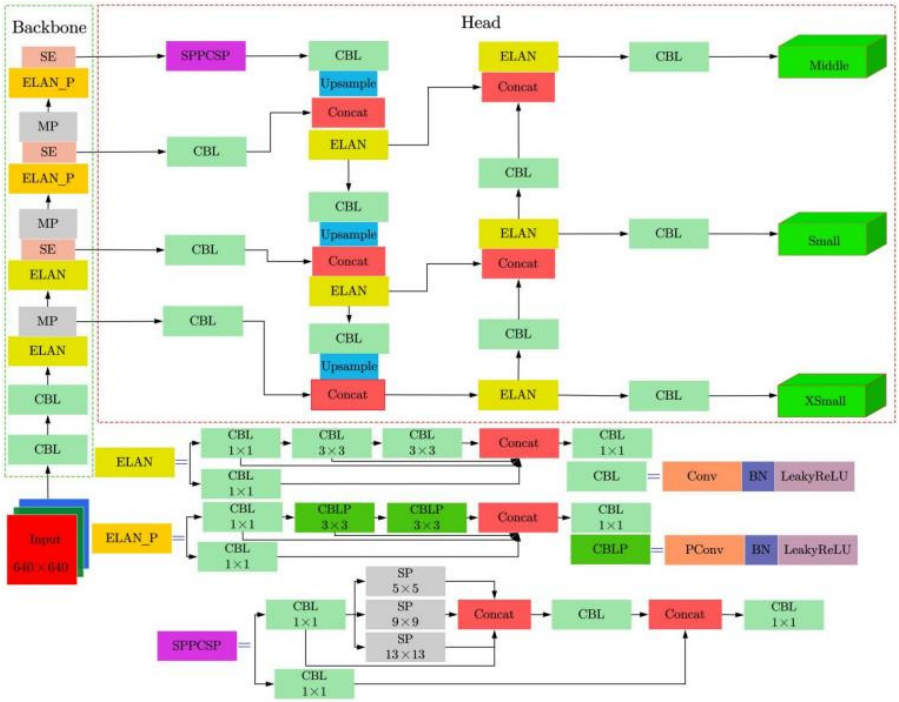


Figure 6. The overall structure diagram of this paper

Where b^{gt} and b represent the center point coordinates of the real box and the predicted box, σ represents the distance between the center points of the real box and the predicted box, C_h represents the distance between the center points of the real box and the predicted box in the y-direction, w^{gt} and h^{gt} represent the width and height of the real box, w and h represent the width and height of the predicted box, B and B_{gt} represent the predicted box and the true value box.

The above are all the improvements made in this paper based on the YOLOv7-tiny network, and a lightweight UAV aerial image detection network is obtained. The overall detection network structure is shown in Figure 6. Compared with the original YOLOv7-tiny network, the main network part, the neck network part, the detection head part and the loss function part are improved, which improves the detection effect of the algorithm for targets in aerial images.

4. Experiment and Result Analysis

The manuscript should include a conclusion. In this section, summarize what was described in your paper. Future directions may also be included in this section. Authors are strongly encouraged not to reference multiple figures or tables in the conclusion; these should be referenced in the body of the paper.

4.1. Data Set and Parameter Settings

The dataset used in this experiment is VisDrone2019 aerial image dataset. This data set is collected by the AISKYEYE team from Tianjin University's laboratory in 14 cities across China, separated by thousands of kilometers, through various drone cameras. Its benchmark dataset includes 288 video clips consisting of 261908 frames and 10209 still images. In this paper, 8629 labeled still images from 10209 still images are used to form the experimental data set, and the targets in this data set contain 10 categories: pedestrians, people, cars, vans, buses, trucks, motorcycles, bicycles, ahab tricycles and tricycles. among them, 6471 images are allocated to training data set, 548 images are allocated to validation data set and 1610 images are allocated to test data set. All experiments in this paper are carried out on the same platform and the same data set. The experiment system is Ubuntu18.04.6, the CPU is Intel Xeon(R)E-2276M, the running memory is 64G, the GPU is RTX5000 with 16G memory, and the deep learning framework is Pytorch2.0.1. CUDA version 11.4 and python version 3.9. In the training process of the network, SGD optimizer is adopted, the training batch size is 4, the initial learning rate setting is 0.01, the preheating initial momentum is 0.8, and the training is 200 rounds.

4.2. Data Set and Parameter Settings

In order to evaluate the improved network model, the evaluation indicators selected in this paper are: recall (R), precision (P), mean average precision (mAP), F1_Score(F1), the speed at which the Model detects images (FPS), the number of parameters used by the model (Parameters), the number of computations performed by the model (GFLOPs), and the resulting model weight file size (Model size) are calculated as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{13}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{14}$$

$$F1 = \frac{2PR}{P + R} \tag{15}$$

$$AP = \int_0^1 P(R)dR \tag{16}$$

$$mAP = \sum_{i=1}^{N} \frac{AP(i)}{N} \tag{17}$$

Where TP, FP and FN represent positive samples correctly identified as positive samples, negative samples identified as positive samples, and positive samples identified as negative samples, respectively.

4.3. Ablation Experiment

The lightweight UAV aerial image target detection network proposed in this paper is improved on the YOLOv7-tiny detection network. In order to verify the detection effect brought by the improvement points proposed in this paper, the experiment conducted by the unimproved basic network is defined as A, the experiment conducted by the improved ELAN module is defined as B, the experiment conducted by the improved neck network and detection head is defined as C, the experiment conducted by the embedded SE attention mechanism in the backbone network is defined as D, the experiment conducted by the improved loss function is defined as E, and the experiment containing all improvements performed is defined as F for ablation experiments and result analysis. The results obtained in the ablation experiments are shown in Table 1:

Table 1. Performance results of the ablation experiment algorithm

Method	P(%)	R(%)	F1(%)	mAP(%)
A	45.3	39.3	42	35.8
B	47.1	37.4	41.6	35.5
C	48.0	40.0	43.3	38.0
D	47.2	38.8	42.5	36.4
E	45.9	38.7	42	36.1
F	46.0	41.5	43.6	38.1

Table 2. Ablation experiment algorithm complexity results

Method	Parameters	GFLOPs	Model(MB)	FPS
A	6031950	13.1	12.3	94
B	4813390	11.5	9.9	90
C	4063150	13.8	8.5	73
D	6075854	13.1	12.4	88
E	6031950	13.1	12.3	93
F	2888494	12.3	6.2	66

4.4. Comparative Experiment

Table 3. Ablation experiment algorithm complexity results

Method	P (%)	R (%)	F1(%)	mAP (%)
CIOU	45.3	39.3	42	35.8
EIOU	46.0	38.6	41.9	35.7
Focal EIOU	45.7	39.1	42.1	36.0
GIOU	45.9	38.8	42	36.0
WIOU	47.3	37.3	41.7	35.7
SIOU	45.9	38.7	42.1	36.1

In this paper, in order to verify the effectiveness of the improved loss function in the UAV aerial image detection, a comparative experiment is conducted with the current mainstream bounding box regression loss function, including

EIOU, Focal EIOU [36], GIOU [37], SIOU, and WIOU [38]. The experimental results are shown in Table 3.

From the analysis of the results of the comparison experiments of different bounding box regression loss functions in Table 3, it can be seen that the detection performance of the algorithm brought by the improvement of the SIOU bounding box regression loss function is the best among the comparison experiments of the different bounding box regression loss functions, which indicates that the improved method of replacing the original network CIOU bounding box regression loss function with the SIOU bounding box regression loss function in this chapter is effective for the design of the target detection algorithm of the aerial image of the lightweight UAV.

In order to verify the superiority of the lightweight UAV aerial target detection algorithm proposed in this paper, in addition to the comparison with the basic algorithm YOLOv7-tiny, it also carries out comparison experiments with the mainstream lightweight target detection algorithms in the current YOLO series of algorithms YOLOv3-tiny, YOLOv5n, and YOLOv8n, and the comparison experimental results obtained by the used are shown in Table 4:

Table 4. Performance results of the ablation experiment algorithm

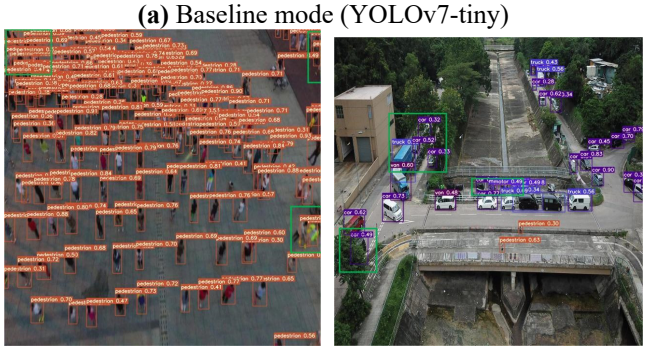
Method	P(%)	R(%)	mAP(%)	Parameters
YOLOv3-tiny	28.2	23.3	16.9	8687482
YOLOv5n	38.4	28.5	27.3	1772695
YOLOv7-tiny	45.3	39.3	35.8	6031950
YOLOv8n	46.9	34.2	35.1	3007598
Ours	46.0	41.5	38.1	2888494

From the comparison experimental results of different algorithms in Table 4, it can be seen that the lightweight UAV aerial image target detection algorithm proposed in this chapter is second only to the YOLOv5n algorithm, which has the smallest model among the mainstream lightweight target detection algorithms, in terms of model size, and the algorithm's precision rate P, recall rate R, F1 value, and average detection accuracy mean mAP are respectively 7.6%, 13%, 10.9% and 10.8%; compared to the latest lightweight target detection algorithm YOLOv8n, although the precision rate P of the algorithm is 0.9% lower, the algorithm's recall rate R, F1 value, and mean average detection accuracy mAP are 7.6%, 7.3%, 4%, and 3% higher, respectively. So comprehensively, the lightweight UAV aerial image target detection algorithm proposed in this chapter has better algorithmic superiority and can be better deployed and applied to edge computing platforms with better detection performance compared to the current mainstream lightweight target detection algorithms.

4.5. Visual Result Analysis

In order to more intuitively reflect the advantages of the proposed algorithm in the target detection effect of UAV aerial images, aerial images in different environments are selected from the VisDrone2019 data set as test images, and the lightweight detection network proposed in this paper and the basic network YOLOv7-tiny are used to test respectively, and the test results are shown in Figure 7.

(a) Baseline mode (YOLOv7-tiny)



(b) Algorithm of this paper

Figure 7. Comparison of detection results

From Figure 7, it is easy to see that the improved algorithm can effectively solve the problem of target missed detection due to the dense distribution of targets and the phenomenon of target occlusion in aerial images, and effectively improve the detection performance of the algorithm.

5. Conclusion

In this paper, based on the YOLOv7-tiny algorithm, a lightweight target detection algorithm for UAV aerial image detection is proposed by using four improvement methods: using partial convolutional PConv to improve the backbone network of the base algorithm, optimizing the target detection layer of the network, adding an attention mechanism, and improving the algorithm bounding box regression loss function. The results of ablation experiments and comparison experiments conducted on the public dataset VisDrone2019 show that the improved algorithm in this paper can satisfy the algorithm's demand for real-time detection of UAV aerial images and the edge computing platform's demand for algorithm lightweighting at the same time with good detection performance, which provides certain reference value for the application of UAV aerial photography technology. However, the algorithm in this paper has further room for improvement in the target detection accuracy of UAV aerial images, which needs to be further studied.

Acknowledgments

This work was supported by the Innovation Fund of Postgraduate, Sichuan University of Science and Engineering Y2022167 and Y2022155.

References

[1] Llanes L A C, Ulbis C R H, Garcia R G. Remote Controlled Unmanned Water Vehicle with Human Detection and GPS Using Yolov4 for Flood Search Operations[C]//2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2023, 1: 373-379.

[2] Ahmed I, Jeon G, Chehri A, et al. Adapting Gaussian YOLOv3 with transfer learning for overhead view human detection in smart cities and societies[J]. Sustainable Cities and Society, 2021, 70: 102908.

[3] Lin Y, Chen T, Liu S, et al. Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning[J]. Computers and Electronics in Agriculture, 2022, 197: 106938.

[4] Wang X, Yang W, Lv Q, et al. Field rice panicle detection and counting based on deep learning[J]. Frontiers in Plant Science, 2022, 13: 966495.

[5] Xu X, Wang L, Shu M, et al. Detection and Counting of Maize Leaves Based on Two-Stage Deep Learning with UAV-Based RGB Image[J]. Remote Sensing, 2022, 14(21): 5388.

[6] Zaidi S S A, Ansari M S, Aslam A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514.

[7] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.

[8] Zhang H, Li F, Liu S, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[J]. arXiv preprint arXiv:2203.03605, 2022.

[9] Chen T, Saxena S, Li L, et al. A unified sequence interface for vision tasks[J]. Advances in Neural Information Processing Systems, 2022, 35: 31333-31346.

[10] Wang W, Xie E, Li X, et al. Pvt v2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415-424.

[11] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[12] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

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

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

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

[16] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[17] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.

[18] 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. arXiv 2022[J]. arXiv preprint arXiv: 2207. 02696, 2022.

[19] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[20] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[21] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[22] 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.

[23] Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. Advances in neural information processing systems, 2016, 29.

[24] Srivastava S, Narayan S, Mittal S. A survey of deep learning techniques for vehicle detection from UAV images[J]. Journal of Systems Architecture, 2021, 117: 102152.

[25] Liu B, Luo H, Wang H, et al. YOLOv3_ReSAM: A small-target detection method[J]. Electronics, 2022, 11(10): 1635.

[26] Tan L, Lv X, Lian X, et al. YOLOv4_Drone: UAV image target detection based on an improved YOLOv4 algorithm[J]. Computers & Electrical Engineering, 2021, 93: 107261.

[27] Liu H, Duan X, Lou H, et al. Improved GBS-YOLOv5 algorithm based on YOLOv5 applied to UAV intelligent traffic[J]. Scientific Reports, 2023, 13(1): 9577.

[28] Li Y, Yuan H, Wang Y, et al. GGT-YOLO: a novel object detection algorithm for drone-based maritime cruising[J]. Drones, 2022, 6(11): 335.

[29] Zhao L L, Zhu M L. MS-YOLOv7: YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography[J]. Drones, 2023, 7(3): 188.

[30] Zeng Y, Zhang T, He W, et al. YOLOv7-UAV: An Unmanned Aerial Vehicle Image Object Detection Algorithm Based on Improved YOLOv7[J]. Electronics, 2023, 12(14): 3141.

[31] Li Y, Fan Q, Huang H, et al. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition[J]. Drones, 2023, 7(5): 304.

[32] CHEN J, KAO S, HE H, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 12021-12031.

[33] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[34] 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.

[35] Gevorgyan Z. SIoU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv:2205.12740, 2022.

[36] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression. arXiv 2021[J]. arXiv preprint arXiv:2101.08158.

[37] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 658-666.

[38] 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.