A Light-weight Deep Neural Network for Vehicle Detection in Complex Tunnel Environments

ZHENG Lie[1], REN Dandan[1,2]

(1. *Hubei University of Technology*, *Wuhan 430000*; 2. *Hubei University of Technology*, *Wuhan 430000*)

Abstract: With the rapid development of social economy, transportation has become faster and more efficient. As an important part of goods transportation, the safe maintenance of tunnel highways has become particularly important. The maintenance of tunnel roads has become more difficult due to problems such as sealing, narrowness and lack of light. Currently, target detection methods are advantageous in detecting tunnel vehicles in a timely manner through monitoring. Therefore, in order to prevent vehicle misdetection and missed detection in this complex environment, we propose aYOLOv5-Vehicle model based on the YOLOv5 network. This model is improved in three ways. Firstly, The backbone network of YOLOv5 is replaced by the lightweight MobileNetV3 network to extract features, which reduces the number of model parameters; Next, all convolutions in the neck module are improved to the depth-wise separable convolutions to further reduce the number of model parameters and computation, and improve the detection speed of the model; Finally, to ensure the accuracy of the model, the CBAM attention mechanism is introduced to improve the detection accuracy and precision of the model. Experiments results demonstrate that the YOLOv5-Vehicle model can improve the accuracy.

Keywords: CBAM, Depth-wise Separable Convolution, MobileNetV3, Vehicle Detection, YOLOV5

1 Introduction

With the rapid development of the global economy, the number of transportation vehicles is also increasing, which brings a lot of convenience to people and speeds up the transportation of goods. However, this has also aggravated the problems of road traffic congestion, frequent traffic accidents, serious energy wastage and deterioration of environmental quality. Therefore, these issues have put forward higher requirements for tunnel highwayssafety, andalso driven the construction and wide adoption of various high-level intelligent highways. Especially in mountainous areas, tunnel highways are an important part of road traffic, and their safety is directly related topeople's lives[1].

The driving environment of a tunnel highways is more complex than that of a regular highway because of the narrow, enclosed space and numerous facilities, which makes its operation and management particularly important. In tunnelhighways, the driving speed is high, the number of vehicles is large, the lighting is poor, the noise is loud, and the air quality is poor, which results in a high accident rate and difficult rescue efforts, especially when a traffic accident occurs on a vehicle carrying flammable and explosive materials, the traffic problem will be more serious.

Therefore, it is essential to establish an effective and comprehensive real-time monitoring system for tunnel safety to prevent and reduce the occurrence of accidents in the tunneland to minimize their destructive impact[2].

Currently, object detection technologiesare the key to vehicle detection based on traffic surveillance

videos. traditional methods use the sliding window approach to extract candidate frames and then extract information of each frame, which is input into a classifier for recognition[3]. These algorithms include Haar+Adaboost[4], Hog+SVM[5], and DPM[6]. However, traditional target detection algorithms generate a lot of redundant candidate frames during the sliding window extraction process, resulting in slower detection speeds and lower efficiency[7].

With the development of deep learning and GPU, object detection models based on deep learning are becoming more and more diverse. Nowadays, target detection methods canbe classified into two main types: Two-Stage methods and One-Stage methods[8]. In the Two-Stage methods, the target candidate regions are generated, and then the generated candidate regionsare classified and calibrated for position to obtain the final detection result. Such algorithms include fast R-CNN, R-CNN, and faster-CNN[9~11]. In the One-Stage methods, object classification and position prediction are performed directly in the network without generating a preselection box. This converts the entire target detection process into a regression problem, so that the class and location information of the target object is obtained by processing the input image once. The network structure is simple and reduces a lot of redundant computation. The more common algorithms are SSD[12-13], YOLO[14], YOLOv2[15], YOLOv3[16], YOLOv4[17], YOLOv5.

Therefore, the single-stage approach has faster detection speed and higher accuracy than the two-stage approach. Shu Liu, Lu Qi, et al.[18] proposed the R-CNN detection model, which applied deep learning to target detection for the first time. Cai et al. [19] introduced cascaded classifiers and proposed the Cascade R-CNN, which optimized the noise interference problem in the detection frame and effectively improveddetection accuracy by adjusting the overlap rate (Intersection over Union, lOU) threshold. Chen[20] choose to classify objects of different sizes. To detect medium objects, they deepened the network's depth to extract more semantic features.To detect small objects, they introduceddeconvolution and region mapping to obtain a higherresolution feature map, which greatly improved the detection accuracy of small and medium targets. Reference [21] used the k-means algorithm to cluster the dataset and learned from the dense network idea to improve the YOLOv3 network for detecting aircraft targets in remote sensing images, which greatly improved the detection accuracy. Yuchun Chu, Hang Gong, et al.[22] proposed a knowledge distillation algorithm based on yolov4 for target detection, which can improve the accuracy of the model and reduce the parameters, but there is still room for improvement in detection speed and efficiency. In summary, deeplearning methods have high application value in vehicle target detection.

A large number of video surveillance devices have been deployed and installed in actual expressway tunnels, which contain a wealth of vehicle information. Based on this, more and more researchers use video data as the entry point to study various vehicle behavior states, gradually improve management efficiency and alleviate the contradiction between massive surveillance cameras and limited surveillance capabilities. Reference [23] introduced a convolutional neural network model to identify vehicles based on traditional image processing methods, which overcomes the false detection caused by light interference to some extent but cannot fundamentally improve the vehicle detection problem. Shixu Shi et al.[24] used a hybrid differenttechnique to extractvehicle targets in the video and used a particle filtering algorithm to track the moving vehicle to realize the parking detection behavior of the vehicle. Du et al.[25] improved YOLOv3 for highway vehicle target detection, but this algorithm is greatly influenced by light and fast target movements and is not suitable for tunnel highways.

Considering the constraints of tunnel highways, such as low light conditions, the possibility of missed and false detections, this paper proposes a network model, YOLOv5-Vehicle, which is based on an improved version of YOLOv5. This model reduces the number of parameters in the original model, while improving its accuracy and detection speed.

The main strategies used in this paper are as follows:

(1) Modifications to the backbone of YOLOv5. The backbone network of YOLOv5 is replaced with a lightweight MobileNetV3 network to extract features and reduce the number of model parameters.

(2) Replacement the convolutional network. All traditional convolutions in the neck module are replaced with depth-wise separable convolutions, which further reduce the number of model parameters and computation, thus improve the detection speed of the model.

(3) Modifications to the attention mechanism of YOLOv5. The CBAM attention mechanism is introduced to ensure the accuracy of the model and to improve its detection precision.

Experiments results demonstrate that the YOLOv5-Vehicle model can effectively improve the accuracy and the detection speed of vehicle detection in the tunnel environments.

2 Introduction of Yolov5 Detection Network

On June 25, 2020, YOLOv5 proposed by Ultralytics LLC, which is an improved version of YOLOv4. This algorithm contains the advantages of many deep learning algorithm frameworks. It has high accuracy, fast detection, and better performance on open source data[27].

Therefore, this paper uses YOLOv5 as the detection model. Its network model is composed of four parts: Input, Backbone, Neck and Prediction[28]. The network structure of YOLOv5 is shown in Fig.1.

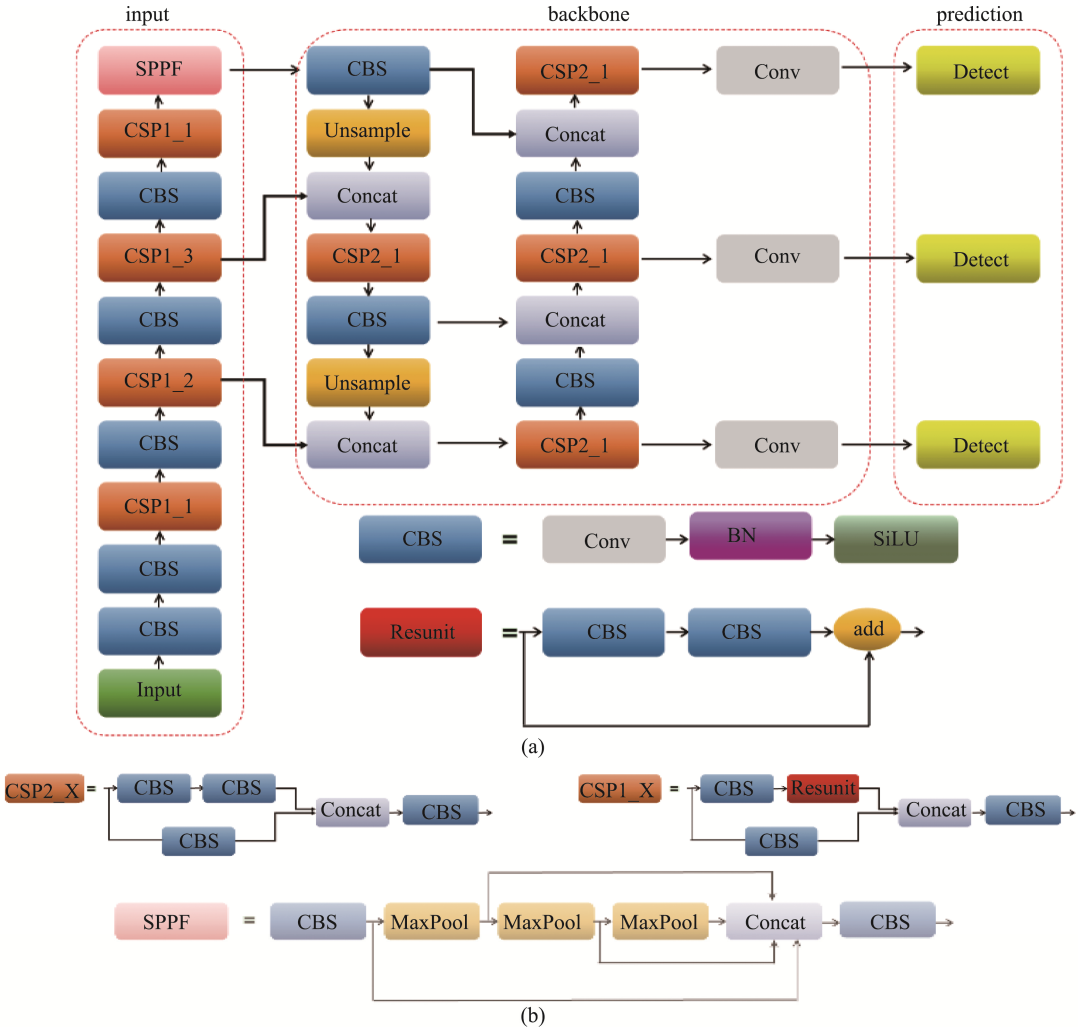

Fig.1 Structure Diagram of YOLOv5. (a) Structure Diagram of YOLOv5. (b) Some Combinations in Structural Diagrams

First, the inputsection consists of an image size of 640×640. There is Mosaic data augmentation, which involves splicing images using methods such as random cropping, scaling and placement, and it can enrich our dataset. In addition, adaptive anchor box calculation and adaptive image scaling are included. The former involves constructing anchor boxes based on the offset of the real border position relative to the preset boundary during training. This process involves outlining targetsat potential position and adjusting them to predefined boundaries. The latter involves adaptively adding minimal black borders to the original image, thus effectively increasing the inference speed. Secondly,the backbone network includes the Focus,CSP and spatial pyramid pooling SPPF [30]. In the Focus section takes the original 640×640×3 image as input and passes it through the Focus structure. A slice operation produces a 320×320×12 feature map, which is then convolved with 32 convolution kernels to create a 320×320×12 feature map. Meanwhile, CSP1_X enhances the gradient values by adding a residual structure during backpropagation, effectively preventing the gradient from disappearing due to the network being too deep. Third, the neck network combines the feature pyramid FPN [29] and PAN [30]. FPN transfers and fuses high-level feature information top-down by upsampling to obtain a feature map for prediction. In contrast, PAN transfers strongly localized features from the bottom up. The combination of these methods aggregates the parameters of different detection layers from various backbone layers to enhance the feature fusion ability of the network. The CSP2_X structure is employed in the neck module, mainly composed of a series of convolutional layers, which strengthen the network's ability to integrate features and retain more feature information. Finally, the Prediction part includesGIOU_Loss and a weighted NMS (Non-Maximum Suppression) to screen multiple target anchor boxes to improve detection accuracy.

3 Improvement of Yolov5

3.1 Improvement of Network Structure

MobileNetV3 was published in 2019[31]. It combines the depth-wise separable convolution of MobileNetV1, the inverted residual structure of MobileNetV2, linear bottleneck and SE modules. It is characterized by fewer parameters and fast speed, which can greatly reduce the demand for computational power. TheSE attention mechanism adopted by MobileNetV3 obtains the importance of each channel by averaging pooling and fully connection layers, and then suppressing or enhancing the channels for different tasks. Inaddition, it also uses the Hard-Sigmoid activation function at the second connection layer, which can improve the operation speed to some extent. Our article performs feature extraction by replacing thebackbone network of YOLOv5 with the MobileNetV3 network. Its network structure is shown in fig.2.

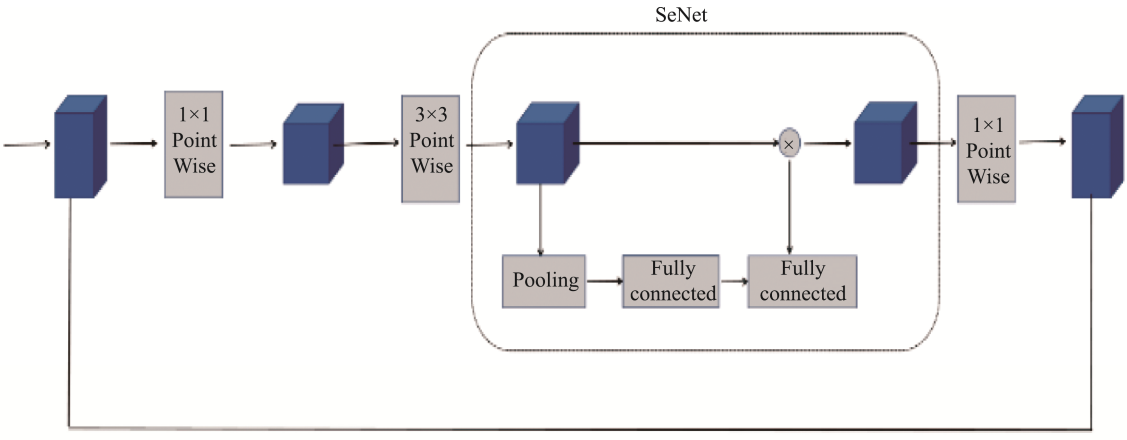

Fig.2 Structure Diagram of MobileNetV3

3.2 Introduce of Depth Wise-separable Convolution

The traditional convolution principle in YOLOv5 is to multiply the input feature maps of each channel with the corresponding convolution kernel and accumulate them, and finally output the feature maps. The traditional convolution structure is shown in fig.3.

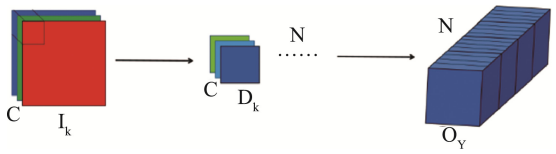

Fig.3 The Traditional Convolution Decomposition Process

I_x and O_y in the figure represent the size of the input image and the output image respectively, D_k represents the size of the convolution kernel, C and N are the number of input and output channels. The definition of traditional convolution is shown in equation (1):

$$Q_2=D^2{}_k\times C\times N\times Q_y \tag{1}$$

In 2017, Sandler et al. proposed the concept of depth-wise separable convolution as a lightweight method for embedded devices[32]. The well-known Xception and MobileNet models use depth-separable convolutions to replace traditional convolutions to reduce model parameters and improve computing speed. Therefore, it greatly reduces parameters and computation of the model and effectively speeds up its detection. Depth-wise separable convolution is a plug-and-play module that is also widely used in convolutional neural network models. It is easyto deploy and can meet the needs of lightweight parameters and computation.

It consists of two small operations: depth-wise convolution and pointwise convolution. Depth-wise separable convolution separates the partial convolution in traditional convolution into a D×D depth convolution and a 1×1 point-by-point convolution. Fig.4 shows the structure of depth- wise separable convolution.

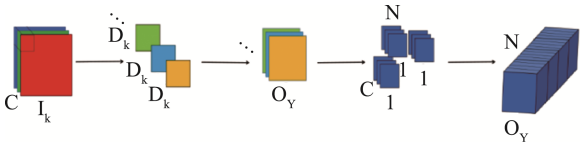

Fig.4 The Depth Wise-separable Convolution Decomposition Process.

I_x and O_y in the fig.5 represent the size of the input image and the output image respectively, D_k represents the size of the convolution kernel, C and Nare the number of input and output channels. the definition of traditional convolution is (2), (3):

$$\frac{Q_1}{Q_2}=\frac{D_k^2\times C*\times+C\times N\times O_y^2}{D_k^2\times C\times N\times O_y^2}=\frac{1}{N}+\frac{1}{D_k^2} \tag{2}$$

$$Q_1=D_{\mathrm{k}}^2\times C\times O_{\mathrm{y}}^2 \tag{3}$$

Comparing the computation of depth-wise separable convolution and traditional convolution, we can observe that the model parameters and computation can be reduced to 1/D^2 of conventional convolution when the model uses depth-separable convolution. Obviously, this can make the detection of the model speed is significantly improved.

3.3 Introduce of CBAM Attention

The CBAM (Convolutional Block Attention Module) attention mechanism is a lightweight attention module proposed in 2018. It can focus on channels and spaces[33]. Reference [34] added the CBAM module to Resnet and MobileNet for comparison and conducted experiments on the application of the two attention modules. They observed that the attention mechanism pays more attention to the target object. Given any intermediate feature map in the convolutional neural network, CBAM injects the attention map along two independent dimensions of the channel and space of the feature map, and then multiplies the attention by the input feature map to perform adaptive feature refinement. Since the CBAM attention mechanism is an end-to-end generic module, it can also be integrated into CNNs and trained together with basic CNNs.

In this paper, the CBAM attention mechanism is

used to replace the SeNet module in the network model to strengthen the focus on the detection target, which reduces the problem of detection accuracy degradation due to the complex environment. The structure of CBAM attention mechanism is shown in Fig.5.

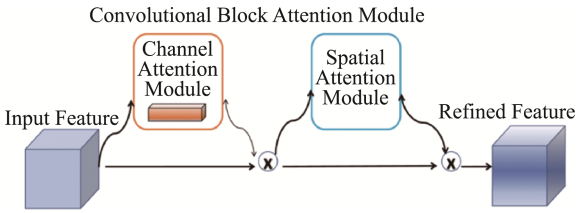

Fig.5 Structure Diagram of CBAM

First, the input feature mapsare subjected to global maximum pooling and global average pooling through the channel attention mechanism, and then the two output feature maps are added using a shared neural network. Finally, the resulting feature mapsare activated by the sigmoid function to obtain the final channel attention feature maps. The channel attention mechanism is defined in equation (4) and its structure is shown in Fig.6.

$$M_c(F) = \delta\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) = \delta(W_1(\mathrm{W}_0(F^c_{avg})) + W_1(\mathrm{W}_0(F^c_{max}))) \quad (4)$$

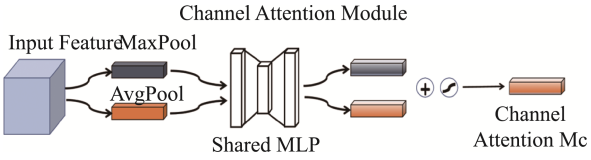

Fig.6 The Channel Attention Module.

We perform element-wise multiplication on the obtained channel attention feature maps to obtain the input feature maps required by the spatial attention, perform global maximum pooling and global average pooling on the channel dimension, and output the feature maps based on the channel to make connections. Next, convolution is performed to reduce its dimension to 1 channel. Finally, the spatial attention feature map is generated through sigmoid. The definition of the spatial attention is shown in formula (5) and the structure of the spatial attention is shown in Fig.7.

$$M_s(F) = \delta\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right) = \delta(f^{7\times7}(F^s_{avg}; F^s_{max})) \quad (5)$$

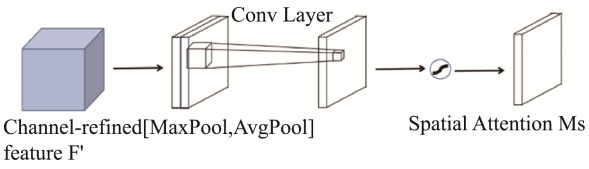

Fig.7 The Spatial Attention Module.

3.4 The Yolov5_Vechicle Model Structure

This article is mainly based on the improved model of YOLOv5. We have made three key improvements: (1) We replaced the backbone network of YOLOv5 with the MobileNetV3 network, which can greatly reduce the parameters of the model. (2) We replaced the traditional convolution of the neck module with a depth-wise separable convolution, which has fewer parameters, this further reduces the parameters of the model and improves the detection speed of the model. (3) To avoid excessive loss of accuracy in the above steps, we introduce a lightweight CBAM attention mechanism module, which improves model's detection precision through two dimensions: channel attention mechanism and spatial attention mechanism.

Experiments have demonstrated that, with the above modifications, we have achieved both high precision of the model and greatly reduced the number of mode parameters. In addition, the detection speed of the model has been significantly improved, making it more suitable for real-time monitoring of embedded devices. The YOLOv5 Vehicle structure is shown in Fig.8.

4 Experiments

4.1 Experimental Environment

In this paper, the deep learning platform is built in Pytorch. The detailed parameters of the experimental environment are shown in Table 1 and Table 2.

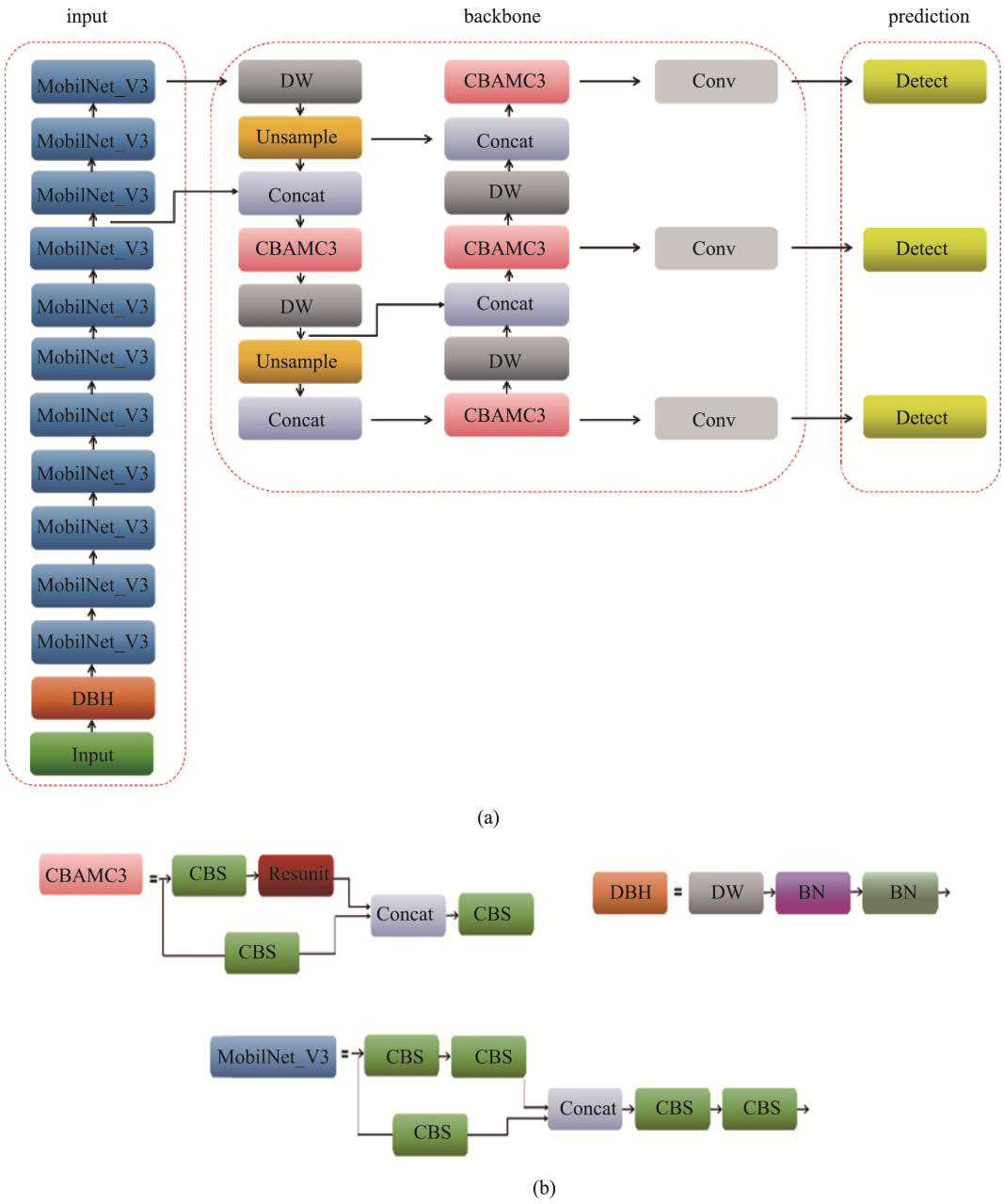

Fig.8 Structure Diagram of YOLOv5_Vehicle.(a) Structure Diagram of YOLOv5_Vehicle. (b) Some Combinations in Structural Diagrams

Table 1 Experimental Environment

Deep Learning Framework	PyTorch1.8.1
CPU	Inter(R) Xeon(R) Platinum 8255C
GPU	RTX 2080Ti 11GB
Operating System	Ubunyu 18.04

Table 2 Experimental Paramenters

Parameter	values
Batch	16
Width ×Height	640× 640
Decay	0.0005
IOU_Thresh	0.2
Epochs	300
Momentum	0.937

4.2 Evaluation Indicator

In the field of object detection, recall, precision, AP and mAP(mean Average Precision) are commonly used to evaluate the performance of object detection algorithm. The confusion matrix is shown in Fig.9.

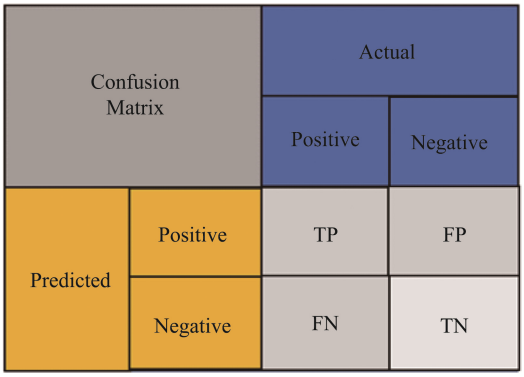

Fig.9 The Confusion Matrix

where TP indicates the number of samples predicted as positive by the algorithm, FN is the number of samples predicted as negative, i.e., the number of missed detections, and FP is the number of samples predicted as positive, i.e., the number of false detections. TN indicates the number of samples identified by the model as negative when their true category is negative. However, it is difficult to maintain a high recall and accuracy rate at the same time during model evaluation. The formula is as follows:

$$R = \frac{TP}{TP+FN} * 100\% \quad (6)$$

$$P = \frac{TP}{TP+FP} * 100\% \quad (7)$$

Therefore, a parameter is needed to integrate these two parameters. The AP and mAP are used to measure the algorithm performance of the detection network. It is applicable to both single-label and multi-label image classification and computation. Their equations can be written as follows:

$$AP = \frac{TP+TN}{TP+TN+FP} \quad (8)$$

$$mAP = \frac{\sum AP}{N} \quad (9)$$

4.3 Data Pre-processing

Due to the lack of effective public datasets in the field of vehicle target detection in tunnel environments, the datasets used in this paper are videos from certain tunnel highway in Hubei, China.

4.4 Construction of Dataset

All the images are processed by blurring, adding noise, flipping, brightness transformation, contrast enhancement, etc.

As shown in Table 4, when the backbone network of the model is replaced by MobileNetV3, the precision rate decreases by 2% and the mAP decreases by 2.9%,while the parameter is 44.9% and the weight is 57.6% of the original model. When the depth-wise separable convolution(DW) is introduced, the precision rate decreases by 0.2% and the mAP decreased by 1.3%, however, the number of parameters is 65.3% and the weight is 62.6% of the original model. Thus, it can be observed that the introduction of MobileNetV3 and depthwise separable convolution can effectively reduce the size of the model, making it more conducive to deployment in mobile devices.

The expansion of the dataset makes the environment more complex, thus increasing the difficulty of inspection. Our experimental datasetincludes a total of 1856 images, which were expanded to 3343 images. The labeling was done using the makesense_ai tool, where vehicles larger than 6 meters are defined as Vehicle_L and vehiclessmaller than 6 meters are defined as Vehicle_S, where the training set is 3008 images, and the validation set is 335 images.

Table 3 The Number of Images

Vehicle	Train	Test	Total
Vehicle_L	903	2105	3008
Vehicle_S	101	234	335

5 Experimental Results and Analysis

5.1 Introducing Mobilenetv3 and Depth Wise Separable Convolution

To simplify the model parameters, this paper introduces the MobileNetV3 module to replace the

backbone network in YOLOv5and replaces all traditional convolutions in the neck module with depth-wise separable convolutions.The experimental results are shown in Table 4.

5.2 Introducing CBAM Attention

The tunnel vehicle data were experimentally verified on different attention detection algorithms and the results are shown in Table 5.

It can be observed that the precision rate and mAP of our CBAM attention mechanismare the highest compared to other attention mechanism, at 96.8% and 89.1% respectively.Compared with the YOLOv5 model, the precision, recall rate and mAP value of YOLOv5_CBAM have increased by 5.6%, 1.8%, and 1.1% respectively.

5.3 Ablation Experiments

Based on the original YOLOv5 algorithm, we added different improved methods to design ablation experiments to verify the improvement effect of each improved method Table 6. Ablation Experiment Results on the original algorithm.

The experimental results are shown in Table 6.

The symbol"√" indicates the introduction of modification methods, A, B, C, D, E and F indicate the algorithm models of introducing different modules introduced on the basis of YOLOv5 model, and YOLOv5_Vehicle indicates the improved algorithm model in this paper. From the experimental results in

Table 4 Improved Experimental Results

Model	P%	R%	mAP%	Parameters	Model size/MB
YOLOv5s	0.912	0.833	0.880	1761871	13.9
YOLOv5_mobil	0.892	0.842	0.851	792725	8.0
YOLOv5_DW	0.910	0.828	0.867	1151343	8.7

Table 5 Results of Different Attention Mechanisms

Model	P%	R%	mAP%	Parameters	Model Size/MB
YOLOv5s	0.912	0.833	0.880	1761871	13.9
YOLOv5_CA	0.927	0.885	0.871	1768511	14.2
YOLOv5_CBAM	0.968	0.851	0.891	1775063	14.3
YOLOv5_ECA	0.913	0.838	0.874	1606812	14.2
YOLOv5_SE	0.936	0.841	0.860	1610663	14.4

Table 6 Ablation Experiment Results

Model	CBAM	DW	MobilnetV3	P%	R%	mAP%	Parameters	Model Size/MB
YOLOv5s				0.912	0.833	0.880	1761871	13.9
A	√			0.968	0.851	0.881	1775063	14.3
B		√		0.91	0.828	0.871	1151343	8.7
C			√	0.90	0.842	0.868	792725	8.0
D	√	√		0.930	0.855	0.871	1165043	9.2
E	√		√	0.936	0.835	0.851	796317	8.6
F		√	√	0.901	0.841	0.838	744181	8.2
YOLOv5_Vehicle	√	√	√	0.930	0.854	0.894	747773	8.4

Table 6, compared with the originalYOLOv5, the number of parameters andweights of the model can be effectively reduced by introducing MobileNetV3 and depth-wise separable convolution. When MobileNetV3 is introduced, the parameter amount and weight size are 44.9% and 57.6% of the original model, respectively. When depth-wise separable convolution is introduced, the parameter amount and weight size are 65.3% and 62.6% of the original model, respectively. When both are introduced simultaneously, the model parameters and weights are 42.2% and 58.9% of the original model, respectively. The precision rate and mAP are 90.1% and 83.8%, respectively, which are 1.1% and 4.2% lower than the original model, but the size of the model is substantially reduced. Then, a lightweight CBAM attention mechanism, named YOLOv5_Vehicle, is introduced based on the F model. Its parameters and weights are 42.4% and 60.4% of the original model, respectively. The precision and mAP are 93.0% and 89.4% respectively. Compared with the original model, the precision rate and mAP increased by 1.8% and 1.4%. It can be seen that the model proposed in this paper not only greatly reduces the computational volume and parameters and shrinks the size of the model, but also ensures a high accuracy rate and mAP.

The positioning loss box_loss, classification loss cls_loss, and confidence lossobj_loss curves of the model in this paper are shown in Fig.10:

It can be observed that the box_loss and obj_loss graphs both level off around 300 epochs and almost converge after 300 epochs. It can be observed from the cls_loss graph that the curve gradually flattens out until the 50th epochs and converges iteratively after 50 epochs. In order to observe detection differences between the improved model YOLOv5_Vehicle and YOLOv5 more intuitively, two different situations with smaller and denser target vehicles are selected for detection comparison. The results are presented in Fig.12, where the first row shows the original images, the second row shows the YOLOv5 model detection images, and the third row shows the detection images of the improved model YOLOv5_vehicle proposed in this paper.

The results show that when the detection target is small, the improved model YOLOv5_Vehicle in this paper has better detection accuracy compared with YOLOv5; when the detection target is denser, YOLOv5 has missed detection, while the YOLOv5_Vehicle model proposed in this paper has no missed detection, and the detection accuracy is relatively high. Therefore, the improved model in this paper maintains high accuracy while greatly reducing the parameters and computation, and has higher degree of recognition for small targets.

5.4 Comparisons with Mainstream Model in Object Images Detection

Comparing the accuracy and model size of the YOLOv5_Vehicle model proposed in this paper with other mainstream models further proves its superiority and feasibility. The experimental results are shown in Table 7.

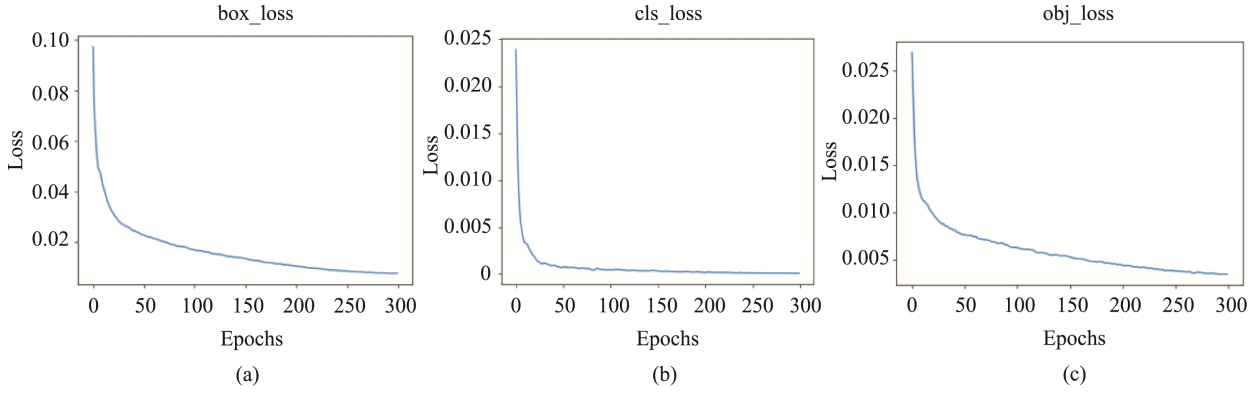

Fig.10 The Loss Curves. (a) Box_loss (b) Cls_loss (c) Obj_loss

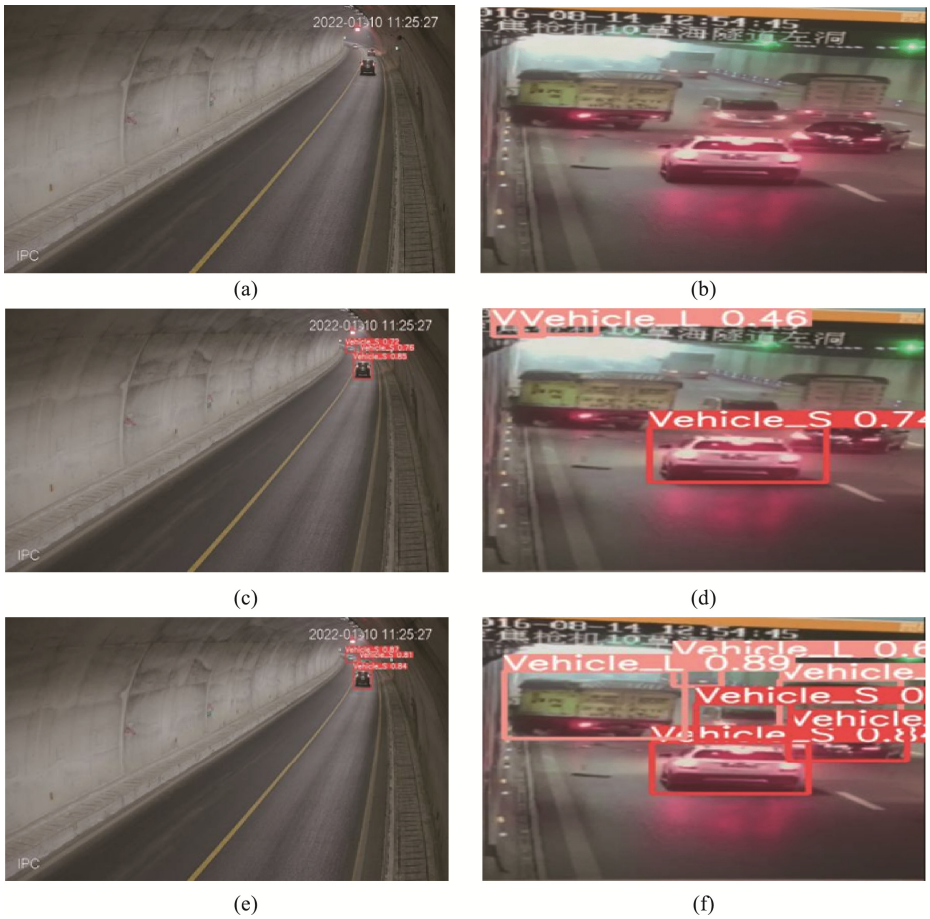

Fig.11 The Detection Examples of YOLOv5 and YOLOv5-Vehicle. (a), (b) Original Images; (c), (d) YOLOv5s Detection Images; (e), (f) YOLOv5-Vehicle Detection Images.

Table 7 The Detection Performances under Different Models

Model	P%	R%	mAP%	Model Size/MB
YOLOv4	0.881	0.836	**0.842**	173.1
YOLOv3	0.871	0.826	0.833	203.5
SSD	0.861	0.786	0.782	191.1
Faster R-CNN	0.882	0.848	0.850	230.6
YOLOv5_Vehicle	0.930	0.854	0.894	8.4

According to the experimental results, compared with the YOLOv4 model, when the model size of YOLOv5_Vehicle is greatly reduced, it precision and mAP are increased by 4.9% and 5.2%, respectively, but the model size is only 4.89% of the YOLOv4 model; Compared with the YOLOv3 model, the precision and mAP of YOLOv5_Vehicle have increased by 5.9% and 6.1%, respectively, the model size is 4.13% of the YOLOv4 algorithm. Compared with the SSD, the precision and mAP of the YOLOv5_Vehicle have increased by 6.9% and 11.2%, and the model size is 4.4% of the SSD model. Meanwhile compared with the Faster R-CNN, the precision and mAP of YOLOv5_Vehicle have increased by 4.8% and 4.4%. Therefore, the improved model in this paper has better performance while reducing the number of parameters and model size. Compared with the previous model our model has an advantage in detecting large trucks with an accuracy of 0.82 and no false detections.

6 Conclusion

1) Replacing the backbone network of the YOLOv5 model with MobileNetV3 reduces the number of parameters in the model by 44.9% and the model size

by 57.6%, which greatly reduces the computational effort and facilitates the deployment of the model.

2) Using depth-wise separable convolution instead of the ordinary convolution of the YOLOv5 model reduces the model parameters by 62.6% and the model size by 65.3%, which greatly improves the running speed of the model and makes it more suitable for quick testing.

3) Replacing the attention mechanism in the YOLOv5model with the CBAM attention mechanism further improves the precision rate and mAP of the model, with precision rate of 93.0% and mAP of 92.4%, thereby reducing the missed detection and false detection of the model.

4) The above three modules are integrated into the YOLOv5 model to obtain the final model, which not only has fewer parameters and faster detection speed, but also ensures higher precision rate and mAP.

This indicates that our proposed method can effectively detect vehicles on tunnel highways, which has a certain promotion effect on promoting further research in the field of vehicle detection and tracking.

References

[1] Zhang, S.R. Ma, Z.L. Shi, Q.(2017). Distribution Characteristics and Preventive Measures of Traffic Accidents in Expressway Tunnel Group. Engineering traffic, pp. 63-66.

[2] Jiang, R.X.Peng, Y.P.(2021).Improved YOLOv4 small target detection algorithm with embedded scSE module. Journal of Graphics, pp: 546-555.

[3] Zhu, H.G.(2021). An efficient lane line detection method based on computer vision, Journal of Physics: Conference Series, 1802(3): 032006.

[4] Tian, D.X. Zhang, C. Duan, X.T. Zhou, J.S. Sheng, Z.G.(2017).The Cooperative Vehicle Infrastructure System Based on Machine Vision, Association for Computing Machinery, pp. 85-89.

[5] K. D. Tomasz, T. Kryjak, "FPGA Implementation of multi-scale face detection using HOG features and SVM classifier, " *Image Processing & Communications,* 2016, 21(3): 27-44.

[6] Felzenszwalb, P. F. Girshick, R. B.McAllester, D. Ramanan, D.(2010). Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1627-1645.

[7] Li, N.N. Wang, X.N. Fu, Zheng, Y.He, F.X.(2022). A traffic police object detection method based on optimized YOLO model, Journal of Graphics.43(2): 296-305.

[8] Hu, J. Wang, Z.B. Chang, M.J. Xie, L.H. Xu, W.-C. Chen, Nan.(2022). PSG-Yolov5: A Paradigm for Traffic Sign Detection and Recognition Algorithm Based on Deep Learning, Symmetry, 14, 2262.

[9] Girshick, R. Donahue, B. Darrell, J. T. Malik, J.(2013). Rich feature hierarchies for accurate object detection and semantic segmentation, arXiv: 1311.2524.

[10] Girshick, R. B.(2015). Fast R-CNN, In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp. 1440-1448.

[11] Ren, S.Q. He, K.M. Girshick, R. B. Sun, J.(2015). Faster R-CNN: Towards real-time object detection with region proposal networks, arXiv:1506.01497v3.

[12] Liu, W. Anguelov, D. Erhan, D. Szegedy, C. Fu, S. Berg, C.Y.(2016). SSD: Single shot MultiBox detector, arXiv:1512.02325v5.

[13] Fu, C.Y. Liu, W. Ranga, A. Tyagi, A.(2017). SSD: Deconvolutional single shot detector, arXiv: 1701.06659.

[14] Li, Z.X. Zhou, F.Q.(2017). FSSD: Feature fusion single shot MultiBox detector, arXiv:1712.00960.

[15] Redmon, J. Divvala, S. Girshick, R.B. Farhadi, A.(2015). You only look once: Unified, real-time object detection, arXiv:1506.02640v5.

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

[17] Bochkovskiy, A. Wang, C.Y. Yhuan, H. Liao, M.(2004).Yolov4: Optimal speed and accuracy of object detection, arXiv: 2004.10934v1.

[18] Liu, S. Qi, L.Qin, H.F.Shi, J.P.Jia, J.Y.(2018).Path aggregation network for instance segmentation, arXiv:1803.01534v4.

[19] Cai, Z.W. Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection, In Proceedings of the IEEE conference on computer vision and pattern recognition, 6154-6162.

[20] Chen, H.J. Wang, Q.Q. Yang, G.W. Han, J.L. Yin, C.J. Chen, J. Wang, Y.Z.(2019). SSD Object Detection Algorithm with Multi-Scale Convolution Feature Fusion,

13(6): 1049-1061.

[21] Redmon, Farhadi, A.(2018). Yolov3: An incremental improvement, arXiv: 1804.02767.

[22] Chu, Y.C. Hang, G. Wang, X.F. Liu, P.S.(2022). Study on Knowledge Distillation of Target Detection Algorithm Based on YOLOv4, Computer Science, 49(6A): 337-344.

[23] Yang, Z.L. Ding, J. Liu, J.F.(2021). A new tunnel parking detection method combined with convolutional neural network, Journal of Chongqing University, 4(6): 49-59.

[24] Shi, X.S.(2008). Vehicle breaking detection on express way based on particle filter algorithm, Computer Engineering and Applications. 44(34)：239-242.

[25] Du, J.H. He, N.(2020). Real-time road vehicles detection based on improved YOLOv3, Computer Engineering and Applications, 56(11): 26-32.

[26] Zhao, L.L. Wang, X.Y. Zhang, Y. Zhang, M.Y.(2022). Vehicle target detection based on YOLOv5s fusion Senet, Journal of Graphics, 43(5): 776-782.

[27] Tan, X.L. Bie, X.B. Lu, G.L. Tan, X.H.(2021). Real-time detection for mask-wearing of personnel based on YOLOv5 network model, Laser Journal, 42(2):147-150.

[28] Liu, Z.G. Gao, Y. Du, Q.Q. Chen, M. Lv, W.-Q. (2022). YOLO-Extract: Improved YOLOv5 for Aircraft Object Detection in Remote Sensing Images, IEEE Access, vol. 11, pp. 1742-1751.

[29] Wu, D.H. Lv, S.-C. Jiang, M. Song, H.B.(2021). Using channel pruning based YOLOv4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments, Computers and Electronics in Agriculture, vol. 178, no. 5, pp. 174–178.

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

[31] Howard, A. L. Sandler, M. Chu, G. Chen, L.C. Chen, B. Tan, M.X. Wang, W.J. Zhu, W.-K. Pang, R.M. Vasudevan, V.J.Le, Q.V.Adam, H.(2019). Searching for MobileNetV3, arXiv :1905.02244v5.

[32] Sandler, M. Howard, A. Zhu, M.L. Zhmoginov, A. Chen, L.C.(2019). MobileNetV2: Inverted residuals and linear bottlenecks, arXiv:1801.04381v4.

[33] Woo, S. Park, J. C. Lee, J. Y. Kweon, I. S.(2018). CBAM: convolutional block attention module, arXiv:1807. 06521v2.

[34] Li, J.N. Zhang, J.Z. Zhang, X.Y. Wang, S.(2022). Lightweight Helmet Wearing Detection Algorithm of Improved YOLOv5, Computer Engineering and Applications. 58(9): 201-207.

Author Biographies

REN Dandan received her B.Sc. degree in school of mathematics from China University of Mining and Technology, China in 2021. She is now a M.Sc. candidate in school of science, Hubei University of Technology, China. Her main research interests include computer vision, object detection, urban traffic intelligent system.

E-mail: 102112270@hbut.edu.cn

ZHENG Lie received his M.Sc. degree in school of mathematics from Hubei University, China from 1992, then as a visiting scholar in department of mathematics and mechanics from University of Warsaw, Poland. He concurrently serves as a master tutor and vice president of the Faculty of Science, he is also the director of Hubei Computational Mathematics Society. His main research interests include applied mathematics and computer technology.

E-mail: Teacherzhenglie@hotmail.com