Foreign Body Detection in the Electrified Area of Urban Rail Trains Using Improved Yolov3 Algorithm

Chensong Wang
Wei Cui
Xingguang Li
Xinrou Liu

Foreign body invade the electric receiving area of urban rail train, interfere with the operation of electric equipment on the roof, and affect the normal operation of urban rail traffic. Aiming at the problems of the traditional non-contact foreign body detection in the electric area of urban rail train, such as slow detection speed and poor detection accuracy of small target foreign body, An improved YOLOV3 (You Only Look Once) network model based on PAN feature pyramid structure and adaptive spatial feature fusion is proposed. By improving the main body of the YOLOv3 network model, it can alleviate the problem that the network prediction size map is too large and the experience field is too small. The features of different levels of foreign objects are initially fused with PAN's feature pyramid to extract strong location information and strong semantic information of the foreign objects, then the method of adaptive spatial feature fusion was used to learn the spatial weights of the fusion of feature maps at various scales, obtaining more effective prediction feature maps at different scales after fusion and improving the detection ability of small targets. The improved k-means clustering algorithm is used to obtain the size of anchor and match it to the corresponding feature layer, which can mark the position of foreign body more accurately. Experimental results show that the detection accuracy of the improved YOLOV3 reaches 95.7%, which is 5.1% higher than the detection effect of the original network. It can accurately and quickly identify the different size of intrusive foreign body in the electric area of the roof of the urban rail train.

Keywords:urban rail train ; foreign body detection; YOLOv3; PAN; Adaptive spatial feature fusion ;dimension clustering
***Tob Regul Sci.*™ 2021;7(5):1059-1066**
DOI: doi.org/10.18001/TRS.7.5.23

In the previous studies based on the condition of the electrified area of the train roof, most of the attention was focused on the abnormal state detection of the pantograph,[1,2,3] but less attention was paid to the foreign body detection in the electrified area, while the foreign body in the electrified area of the train roof would become an unsafe factor, such as lost tools, dead birds, plastic bags. These foreign bodies will damage the pantograph equipment due to inertia impact during the train operation, and will also be wound to the cable and transmission line equipment, causing the equipment fail to work normally and affecting the traffic operation.

Due to the complexity of the vehicle roof equipment, it is difficult to apply the common foreign body detection method to the vehicle roof foreign body detection.[4] The earlier railway foreign body detection methods[5][6] based on infrared curtain, laser scanning and other methods cannot accurately judge the type of foreign body, and the calculated parameters do not have significant characteristics, which increases the difficulty of foreign body identification. With the development of deep learning in the field of target detection, some foreign object detection methods based on this have been proposed[7,,8,9], and have been used in the field of railway foreign object detection[10,11,12].However, this kind of detection method has some shortcomings in the detection of irregular and small targets, and the

Chensong Wang, College of Electronic Information Engineering, Changchun University of Science and Technology,Changchun, China. Wei Cui, College of Electronic Information Engineering, Changchun University of Science and Technology,Changchun, China, Xingguang Li, College of Electronic Information Engineering, Changchun University of Science and Technology,Changchun, China, Xinrou Liu, College of Electronic Information Engineering, Changchun University of Science and Technology,Changchun, China.*
Corresponding author: Wei Cui, College of Electronic Information Engineering, Changchun University of Science and Technology,Changchun, China*

detection accuracy needs to be improved.

In this paper, combined with the YOLOv3 network, an improved YOLOv3 network based on the PAN feature pyramid structure combined with adaptive spatial feature fusion is proposed to complete the task of detecting foreign objects in the power receiving area of urban rail trains. By increasing the depth of the network, the deeper characteristic information of the foreign body can be obtained. Then, using the PAN image pyramid structure was used to perform initial fusion of foreign body features at different levels, and adaptive spatial feature fusion was used to make full use of the semantic information of feature pyramid to obtain the best fusion method of feature images with different resolutions and feature images of different scales, which could be used to predict the location and category of foreign body. Finally, the data set was re-clustered to obtain a new anchor box suitable for improving the network and data set, marking the location of foreign body more accurately. By using the above method to optimize and improve the YOLOv3 network, a more efficient detection network can be obtained, which is suitable for the real-time detection of foreign body in the electric area of the roof on urban rail train.

YOLOV3 DETECTION NETWORK

At present, the representative deep learning networks such as RCNN,[13] Fast R-CNN,[14] Fast-er R-CNN,[15] SPP-Net.[16] this kind of detection algorithm adopts a method based on area detection, including various kind of candidate region generation and different feature layer processing process that cannot guarantee the real-time performance of the algorithm.[17] Redmon proposed the YOLO series of algorithms[18,19,20] that transforms target detection into a regression problem. It takes the entire image as the input of the network and only passes through a neural network to get the location of the bounding box and its category, which improves the real-time performance of target detection.

YOLOv3 uses Darknet-53 as the backbone network, which is mainly composed of convolutional layer and residual structure.[21] The convolutional layer consists of a series of depth separable convolution. After each convolutional layer, a batch normalization operation[22] and a LeakyReLU layer are added to prevent overfitting phenomenon. Residual structure completes the jump link between layers, which effectively solve the problem of vanishing gradient, and greatly improves the efficiency of deep network training. In the prediction network of the network, YOLOv3 uses a multi-scale fusion method similar to FPN[23] for prediction, which improves the accuracy of target detection.

The detection accuracy and speed of YOLOV3 on VOC2007 data set have achieved good results (Table 1). It is more competent for real-time detection tasks in complex environments. For the YOLOv3 network, multi-scale foreign body detection in a complex background is still a challenge. In the task of detecting foreign body in the power receiving area of urban rail trains, the detection background is complicate and contains many small-sized foreign body. It is difficult to obtain good results by directly using YOLOv3 network to detect foreign body, so it is necessary to improve the network to adapt to the foreign body detection task in the electric area of urban rail train.

Table 1 YOLOv3 Compare. other Network Performance Tables (VOC2007 data Set)

Model name	mAP(%)	Detection speed($f \cdot s^{-1}$)
Fast R-CNN	68.4	0.5
FasterR-CNN	70.4	5
SSD300	72.4	46
SSD512	74.9	19
YOLO	57.9	45
YOLOv2	73.4	67
YOLOv3	74.5	35

METHOD

The Darknet-53 network is used in the YOLOv3 network to extract features. Through five times consecutive down-sampling and twice up-sampling, three size feature maps of 52×52, 26×26, and 13×13 are obtained and used for location and category prediction. Aiming at the problems appearing in YOLOv3 algorithm, such as large size of feature map and small perception field, this paper adds 4 continuous residual units on the basis of DarkNet-53, then continues to perform a down-sampling to build deeper network structure, add the number and size of feature map to enhance the detection accuracy about the network, and extract richer deep features of foreign body. The input image size of the network is normalized to 512×512 to retain more image information and enlarge the network receptive field. After the input image is down-sampled six times, four feature maps of different sizes are output, which are 64×64, 32×32, 16×16, and 8×8, serveing as the yolo layer to be processed in the improved network. The backbone structure of the optimized network is shown in Figure 1.

Fig.1 Backbone network optimization

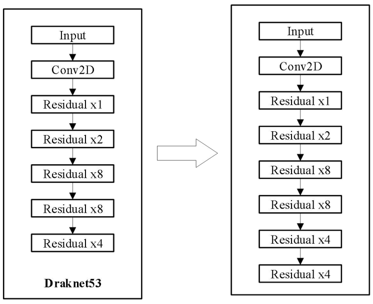
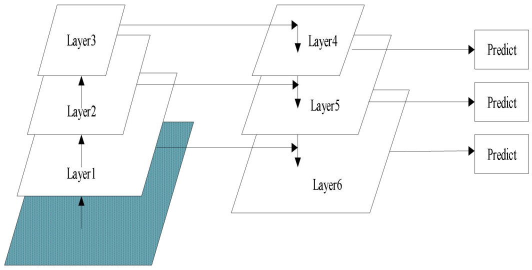


Fig.2 FPN network architecture



In deep convolutional networks, image pyramid is used to fuse the information among various scale features and forecast on multiple feature graphs to improve the detection performance of the network.

Feature Pyramid Networks (FPN) is shown in Figure 2. Multiple deep convolution is take on input image to obtain Layer layers of different sizes. After sampling Layer4, the feature map with the same size as Layer2 is obtained. After channel processing, tensor addition is carried out to obtain the feature Layer Layer5. FPN up-samples the higher-level feature map with stronger semantics from top-down, then connects the feature to the previous layer horizontally, Therefore, the high-level features have been enhanced, and each prediction layer incorporates features of different resolution and semantic intensity, which can complete the detection of objects with corresponding resolutions to ensure that each layer has a suitable resolution. And strong semantic features.

To increase location information about foreign body on the feature map, the feature pyramid network of PAN (Path Aggregation Network)[25] is used to promote the flow of image information. It uses accurate low-level positioning information to enhance the entire feature hierarchy, by shortening the information path between the low-level and top-level features, so that each feature map used for detection inherits the location information of the stronger foreign body.

When YOLOv3 uses the feature pyramid detection, the small targets of foreign body are more related to the shallow feature maps, and the large target foreign body are more related to the high-level feature maps. When a foreign object is assigned a positive value in the feature map of a certain level, the corresponding area in the feature map of other levels is regarded as the background. Therefore, if an image contains both large and small foreign body, the conflicts between features at different levels often occupy most of the feature pyramid. This inconsistency interferes with the gradient calculation during training and reduces the effectiveness of the feature pyramid. Adaptive Spatial Feature Fusion[23] is a new data-driven pyramid feature fusion method. It learns to filter different levels of feature conflict information in space to suppress the inconsistency of gradient back propagation, and improves the invariance of feature proportions, which alleviates the interference during gradient calculation and improves the effectiveness of the feature pyramid.

Adaptive spatial feature fusion can adaptively learn the spatial weights of feature map fusion at various scales. In a multi-scale training task, to obtain a feature map of a certain scale, it is necessary to compress the feature maps of other scales on the channel and adjust the size to the same level. And then find the optimal fusion of the target on each scale feature map. At each spatial location, through training, the contradictory information and insignificant features at that location are filtered out, making more representative features easier to discover and learn.

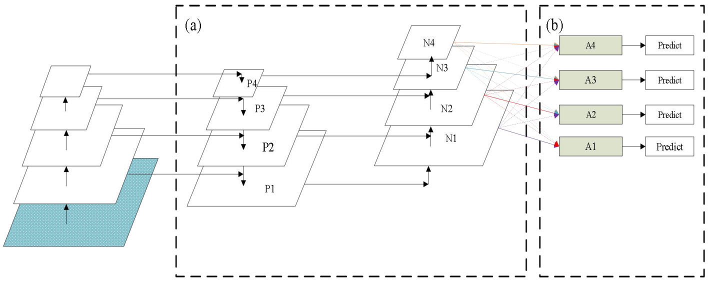
The model will find the most suitable spatial weight of the feature map fusion during the training process. The feature fusion formula is as follows:

$$y_{ij}^{l} = \alpha_{ij}^{l} \cdot N_{ij}^{1\to l} + \beta_{ij}^{l} \cdot N_{ij}^{2\to l} + \gamma_{ij}^{l} \cdot N_{ij}^{3\to l} + \xi_{ij}^{l} \cdot N_{ij}^{4\to l}$$

$$\alpha_{ij}^{l} + \beta_{ij}^{l} + \gamma_{ij}^{l} + \xi_{ij}^{l} = 1$$

Among them, y_{ij}^{l} represents the mapping of the vector (i, j) on the feature map y^{l}, $N_{ij}^{n\to l}$ represents the feature adjusted from the n-level feature to the l-level feature at the position (i, j) vector. $\alpha_{ij}^{l}, \beta_{ij}^{l}, \gamma_{ij}^{l}, \xi_{ij}^{l}$ are the four different levels of features learned through the network adaptively and are mapped to the spatial weights of layer l. After the softmax operation in the channel direction, they are normalized between [0, 1].

Fig 3 Improved YOLOv3 detection network (a) PAN characteristic pyramid structure (b) Adaptively Spatial Feature Fusion



After the addition of adaptive spatial feature fusion, the neck of the improved YOLOv3 network is shown in Figure 3. Among the above net-work structure features, the PAN structure aggregates the characteristic parameters of different detection layers from different main layers to enrich the characteristics of foreign body in each detection layer. The adaptive spatial feature fusion preserves useful information for combination during training and finds out the best fusion method of feature images at different scales, so that the fused feature images have more significant feature information of foreign body and better ability of foreign body detection at different scales.

Yolov3 used K-means algorithm to cluster the 9 groups of priori box dimensions on COCO dataset. However, in the actual detection task, the priori box dimension calculated on the COCO data set is not suitable for our foreign body detection scene. Since the structure of the network has been optimized, the number and size (pixels) of the acquired detection feature map are also changed, and the required priori box needs to be updated accordingly. Therefore, in order to be suitable for the optimized network and the produced data set, an improved K-means algorithm[26] was used to carry out cluster analysis on the foreign body data set, and 12 groups of prior box dimension centers were obtained. The clustering centers were used to carry out detection experiments on the foreign body data set. The specific steps are as follows:

(1): Select the cluster center cluster, the selection principle: the surrounding points of the cluster center point are relatively dense, and the distance between the cluster center points is relatively far. Define that if the number of sample data containing points in the neighborhood of avgDist of point a is greater than min-Points, then it is a dense point, and min-Points is the minimum sample number threshold.

(2): Calculate the IOU of the two boxes. The smaller the d is, the more similar box_j is to the cluster, and box_j can be classified as the cluster.

$$d(box_j, centriod_i) = 1 - IOU(box_j, centriod_i)$$

(3): Update each cluster to the mean value of the points of the current cluster.

(4): Repeat the above (1)~(3) process continuously until the cluster center changes very little, and the optimal cluster center is obtained.

The above K-means clustering algorithm was used to conduct cluster analysis on the self-made foreign body data set, and 12 groups of prior box dimension centers were obtained, which were evenly divided into four feature maps of different scales for foreign body detection experiments.

EXPERIMENT

For the task of detection foreign body on the roof of urban rail trains, there is no existing roof foreign body data set available. To better solve that problem and verify the applicability of improved network foreign body detection, select five common types of foreign body that are easy to invade the roof of the car are used as the basis for the construction of the data set. Thousands of images are obtained through shooting and online collection, and a set of industrial foreign body data sets are self-made through marking tools. The data set contains a total of 8,500 pictures, divide into 5 categories: wrenches, gloves, birds, plastic bags, and kites. respectively, some data samples are shown in Figure 4.

Fig.4 Sample data set

The main hardware configuration of this experimental platform are Intel i7-9700 processor, 16GB running memory, 8GB GTX1650 GPU. The software environment is CUDA10, Python3.6, Pytorch and various dependent libraries. Network training adopts the idea of transfer learning and uses the official weight of YOLOv3 as the initial weight to improve the training speed. Some training parameter Settings are shown in Table 1.

Table 2 Experimental parameters

Parameter names	Value	Parameter names	Value
Learning rate	0.001	Weight_decay	0.0005
Epoch	300	Max_batches	500200
Batch size	16	Ir_factor	0.1
Momentum	0.9	Nms	0.5

In the experiment, the same data set was simultaneously used to train Faster R-CNN, SSD[27], YOLOV3 networks for the same period. The trained model was used to detect the same images, and mAP and detection speed were used as evaluation indexes of detection effect.

Fig.5 Comparison of detection effects

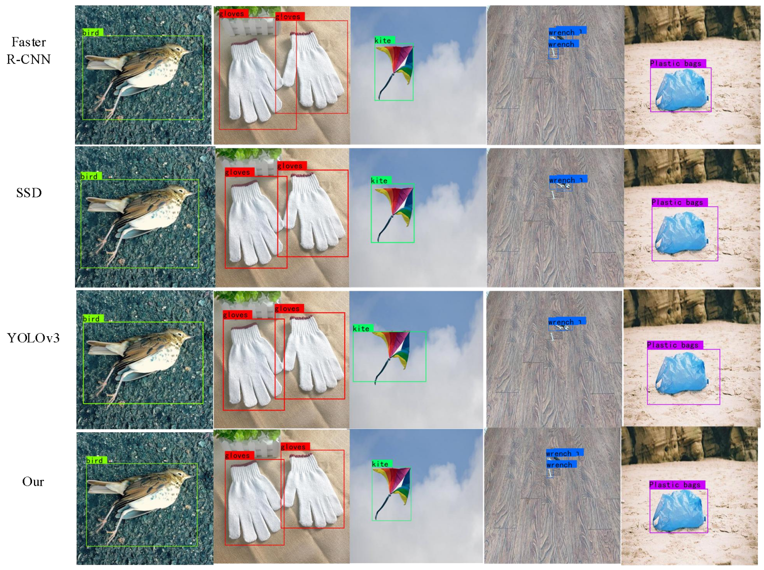


Table 3 Comparison of experimental results of self-made data set

Model name	mAP(%)	Detection speed($f{\cdot}s^{-1}$)
Faster R-CNN	93.2	0.6
SSD	88.7	22
YOLOv3	90.6	28
Our	95.7	23

Figure 6 visually shows the detection effects of various algorithms. It can be seen that Faster R-CNN, SSD, YOLOv3 and improved YOLOv3 algorithms can all correctly detect the target category. After the improved YOLOv3 algorithm uses the improved K-means clustering algorithm, the return position of the detection frame is more accurate. At the same time, in the detection scene where there are small targets, the SSD algorithm and the YOLOv3 algorithm have missed the detection of small targets, and the improved YOLOv3 algorithm performs better in the detection of multiple targets and small targets.

From the results in Table 3, it can be seen that the Faster R-CNN and SSD algorithms are inferior to the improved YOLOv3 algorithm in terms of mAP and recognition frame rate. The real-time performance of the YOLOv3 is better than the improved YOLOv3, but in terms of detection accuracy, the improved YOLOv3 algorithm works better. The improved YOLOv3 algorithm has promoted detection accuracy by 4.9% compared to the original YOLOv3 algorithm. Under the condition of maintaining high detection accuracy and real-time detection, the improved algorithm in this paper is more suitable for the detection of foreign body intruding on the roof of urban rail trains.

In order to show more intuitively and improve the performance of the network for small target detection, this paper sorts the 1700 images of the test data set according to the target size, and divides them into three categories: S, M, and L, which represent the three size targets of small, medium and large respectively. Including 1050 large-size targets, 2160 medium-size targets, and 521 small-size targets. The experiment compares the detection performance of different size of targets based on various detection network, and counts the number of accurate detections, the number of false detections and the number of missed detections in the detection of small-size targets by each detection network. The experimental results are as Fig.6.

Table 4 Comparison of small target test results

	Number of actual	Number of errors	Number of missed	Accuracy	Error rate	Misdetection rate
Faster R-CNN	419	26	76	80.4%	5.0%	14.5%
SSD	377	39	105	72.3%	7.5%	20.2%
YOLOv3	407	31	83	78.1%	5.9%	16.0%
Our	480	16	25	92.1%	3.1%	4.8%

It can be seen from the experiment that the improved YOLOv3 algorithm is better than other types of algorithms in the detection performance of the three types of targets, especially in the performance of small target detection. Compared with the original YOLOV3 algorithm, the error detection rate and omission rate of the improved YOLOV3 algorithm for small targets are reduced by 2.8% and 11.2%, respectively, which improving the detection accuracy of small targets. It also show that the improved YOLOv3 network uses a feature extraction network based on the PAN feature pyramid structure combined with adaptive spatial feature fusion, which combines high-level and low-level semantic features. It effectively solves the problem of misdetection and missed detection of small-scale targets, so that the entire network has better detection capabilities for small foreign body, stronger robustness, and more suitable for foreign object detection environments in the electrified area on the roof of urban rail trains.

Fig. 6 Comparison of three types of target detection performance

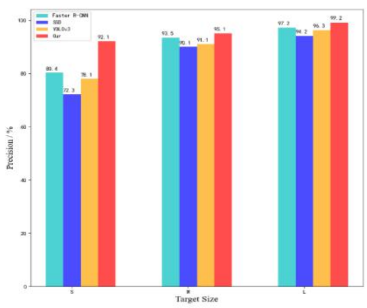
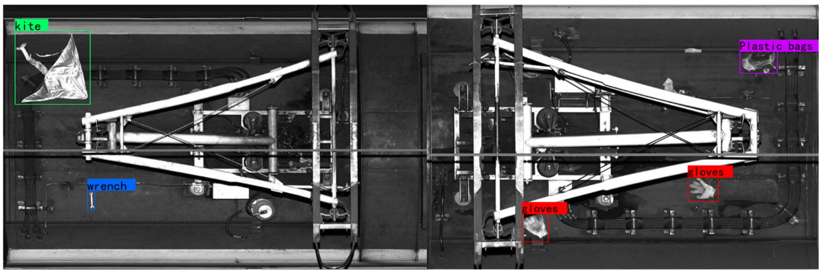


Figure 7 shows the improved YOLOv3 network for the detection results of foreign body in the power receiving area on the roof of urban rail trains. Improved YOLOv3 network can handle targets of different sizes well and accurately detect foreign body intruding in the pantograph area of urban rail trains.

Fig.7 Multidimensional foreign body detection and verification



CONCLUSION

The improved YOLOv3 network based on the PAN feature pyramid structure combined with adaptive spatial feature fusion detects invading foreign body of different sizes in the electrified area of urban rail train roofs, and enriches the high-level semantic features of foreign body by deepening the backbone network structure, and alleviating the large size of the original network prediction map ,the small experience field. This paper improve the prediction network to increase the accuracy of the network's detection of small-size foreign body, and use the improved K-means clustering algorithm to more accurately mark the location of foreign body. The experimental results show that the method can accurately and quickly identify the intrusion of foreign body of different sizes in the power receiving area on the roof of urban rail trains. Due to the limitation of the experimental conditions, this article did not do the foreign body detection experiment on the hanging bow net. At the same time, the algorithm has room for improvement. Combined with the deep network model pruning technology, the detection speed of the improved network can be further promoted.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

China Jilin Province Science and Technology Plan Development Project (20180201042G X)

REFERENCES

1. Lu S, Liu Z, Li D, et al. Automatic wear measurement of pantograph slider based on multi-view analysis[J]. IEEE Transactions on Industrial Informatics, 2020.
2. Gao S. Automatic Detection and Monitoring System of Pantograph–Catenary in China's High-Speed Railways[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1-12.
3. Xin T, Roberts C, Weston P, et al. Condition monitoring of railway pantographs to achieve faultdetection and fault diagnosis[J]. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2020, 234(3): 289-300.
4. Yang K, Li J, Lin P, et al. Shape-based foreign body recognition of train roof using invariant moments[J]. Optik, 2013, 124(21): 5181-5183.
5. OH S, Kim G, Lee H. A Monitoring System with Ubiquitous Sensors for Passenger Safety in Railway Platform[C] //Proceedings of IEEE 7th Internatonal Conferenceon Power Electronics. NewYork:IEEE, 2007: 289-294.
6. Guo B Q, Zhu L Q, Shi H M. Intrusion Detection Algorithm for Railway Clearance with Rapid DBSCAN Clustering [J].Journal of Instruments and Instruments, 2012,33(2): 241-247.
7. Liang H, Zuo C, Wei W. Detection and Evaluation Method of Transmission Line Defects Based on Deep Learning[J]. IEEE Access, 2020, 8: 38448-38458.
8. Rong D, Xie L, Ying Y. Computer vision detection of foreign body in walnuts using deep learning[J]. Computers and Electronics in Agriculture, 2019, 162: 1001-1010.
9. Cao X, Wang P, Meng C, et al. Region based CNN for foreign object debris detection on airfield pavement[J]. Sensors, 2018, 18(3): 737.
10. Xu Y, Tao H Q, Hu L L. Railway Foreign Body Intrusion Detection Based on Faster R-CNN Network Model[J]. Journal of the China Railway Society, 2020, 42(05): 91-98.
11. He D, Yao Z, Jiang Z, et al. Detection of foreign matter on high-speed train underbody based on deep learning[J]. IEEE Access, 2019, 7: 183838-183846.
12. Wang Y, Wang J G. Study on safety inspection of railway train operation based on deep learning algorithm[J]. China Safety Science Journal, 2018, 28(S2): 41-45.
13. Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. 580-587.
14. Girshick R, Fast R-CNN[J]. Proceedings of the IEEE Anternational Conference on Computer Vision, 2015.

1440-1448.
15. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015. 91-99.
16. He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Re-cognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9). 1904-1916.
17. He K , Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 770-778.
18. Redmon J, Divvala S, Girshick R, et al. You only look once unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Boston, America. 2015: 779-788.
19. Redmon J, Farhadi A. YOLO9000 : better, faster, stronger[C]//IEEE Conference on Computer Vision & Pattern Recognition, Hawaii, America. 2017: 6517-652.
20. Redmon J, Farhadi A. YOLOv3: an incremental improvement[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, America 2018.
21. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
22. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]//International Conference on Machine Learning. Guangzhou, China.2015: 448-456.
23. Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection[J]. arXiv preprint arXiv:1911.09516, 2019.
24. Lin T Y, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:2117-2125.
25. Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
26. Zhang S J, Zhao H C. Algorithm research of optimal cluster number and initial cluster center. Application Research of Computers, 2017, 34(06): 1617-1620.
27. Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multi Box Detector[C]//European Conference on Computer Vision. Springer, Cham, 2016.