ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

Survey and Application of Target Detection Algorithms Based on Deep Learning

Tianyu Li[1, a], Hao Wu[1, b], Yanling Mao[1, c], Dong Li[1, d], Lei Chen[1, e]

[1]Artificial Intelligence Key Laboratory of Sichuan Province, Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong, 643000 China.

[a]litianyu207@163.com, [b]wuhao801212@163.com, [c]1061143634@qq.com, [d]1017130100@qq.com, [e]2795725046@qq.com

Abstract

This paper reviews the target detection algorithm based on deep learning and its application in the power job site. Firstly, the development of the target detection algorithm and the research directions in recent years are sorted out; then the single-stage and two-stage detection algorithms are separately reviewed. To elaborate, describe its advantages and disadvantages in detail, and on this basis, analyze in detail the improved network architecture based on two methods, and at the same time analyze the application of target detection technology in the power job site; finally, the current stage of deep learning target detection The shortcomings of the algorithm and its development direction are discussed.

Keywords

Deep learning, Target detection, One- stage, Two-stage, Power job site.

1. INTRODUCTION

Target detection technology is a very important research direction in the field of computer vision. This research direction combines the two topics of target positioning and target recognition, that is, determining the category and location of the target in a picture or a video. At present, target detection technology has achieved great success in the fields of face recognition, automatic driving, video surveillance, etc.

Traditional target detection algorithms generally use the sliding window method for detection, which is generally divided into three stages: first, use the sliding window to generate some candidate regions on the image, then extract the feature of the candidate region, and finally use the pre-trained classifier to perform classification. The main representative algorithms are: Viola-Jones algorithm [2], HOG + SVM algorithm [3]-[4], DPM algorithm [5]-[6]. The Viola-Jones algorithm is mainly used for face detection. It uses Haar features to describe the commonality of faces and build integral images. Then it uses Adaboost to train the classifier to build a cascade classifier, and finally uses non-maximum suppression. The HOG + SVM algorithm is generally used for pedestrian detection, mainly by extracting the HOG features of positive and negative samples and training it in the SVM classifier. The DPM algorithm uses improved HOG features, SVM classifier and sliding window detection ideas, for the multi-view problem of the target, a multi-component (Component) strategy is adopted, and for the deformation problem of the target itself, a graph-based structure (Pictorial Structure) part model strategy. However, the traditional target detection algorithm based on sliding window selection strategy is not targeted, so the time complexity is high, redundant windows will be generated, and the manually designed features are not very robust.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

In the 2012 ImageNet Challenge, the deep learning model AlexNet won the championship, igniting the wave of deep learning research. At present, the target detection technology is mainly based on deep learning research. According to the detection process, it is mainly divided into a target detection algorithm (two stage) based on regional suggestions and a target detection algorithm (one stage) based on target regression, The difference between the two is that the two-stage target detection algorithm needs to first generate a pre-selection box that may contain the object to be detected, and then perform object detection; while the one stage algorithm treats the task as a regression problem and directly extracts features from the network to predict the object category and location. The representative two stage algorithms are R-CNN, fast-RCNN, faster-RCNN, etc.; the representative algorithms of one stage are mainly YOLO, SSD, retinanet, etc.

2. TARGET DETECTION ALGORITHM BASED ON REGION SUGGESTION

The basic steps of the two-stage detection algorithm based on the area recommendation target detection algorithm are: first, the candidate area is generated by the candidate area generation algorithm Slective Search, then the candidate area is sent to the CNN for feature extraction, and finally the extracted information is processed Classification and regression.

2.1. Introduction to R-CNN, SPP-Net, Fast R-CNN, Faster R-CNN Algorithms

Literature [7] proposed the first detection network R-CNN based on deep learning. This algorithm increased the mAP value from 35.1% to 53.7%, and also greatly improved the speed. The author combined Slective Search and CNN to form R-CNN. The algorithm first generates candidate regions on the input image, and then uses a convolutional network to extract features for each candidate region, and then sends the features to the SVM classifier to determine the category sum. The linear regressor finely corrects the predicted frame position. However, the network has three major shortcomings. First, all candidate regions will be sent to the convolution network for feature extraction and classification, then the amount of calculation is too large, resulting in low detection speed; second, R-CNN's calculation of each All of the data is stored, which will consume a lot of disk space; third, due to the existence of the fully connected layer, the input image size will be limited to a fixed size.

In 2015, He Kaiming et al [8] proposed SPP-Net (Spatial Pyramid Pooling), which made great improvements to the problems of R-CNN. First, SPP-Net adds a pyramid pooling layer after the last convolutional layer in the network, which solves the problem that the input image must be enlarged or reduced to the same size. Second, SPP-Net only extracts the features from the original image once, and then maps the candidate boxes generated by Slective Search to the feature map, which is about 100 times faster. However, SPP-Net also does not solve the problem of storing the calculation results at each step, and the candidate frame generation and feature extraction are not in the same network.

Fast R-CNN [9] is improved on the basis of R-CNN, and its detection speed and detection accuracy are further improved than R-CNN and SPP-Net. The algorithm first extracts the features of the original image, then maps the candidate regions to the feature map to find the region of interest, and obtains fixed-scale features after ROI pooling, and finally performs classification and regression. Compared with the previous detection network, Fast R-CNN has three main improvements: (1) simplify the pyramid pooling layer, downsample the feature maps of different scales to a fixed scale, and obtain the ROI Pooling layer; (2) A multi-task loss function is proposed, and the border regression is added to the convolutional network for training, and the loss of classification and regression is obtained; (3) A large amount of disk space is not required to store the calculation result of each step.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

In the same year, Faster R-CNN [10] integrated the four steps of target detection into the same network, and truly realized end-to-end calculation. The biggest advantage is that the selective search of the candidate region generation algorithm is replaced by a candidate region generation network (RPN), which solves the redundant calculation of the original candidate region generation algorithm and improves the detection speed. The candidate area generation network generates 9 anchors for each pixel position in the image according to the 9 sizes of area and ratio, and uses IoU in the classification layer to determine the foreground and background of each area, and is used in the window regression layer Softmax corrects the anchor points of the bounding box to obtain accurate candidate regions.

2.2. Improved Network based on Faster R-CNN

The improved network based on Faster R-CNN is mainly improved in four aspects: more complete ROI classification, better feature network, more accurate RPN network and larger mini-Batch.

Literature [11] proposed R-FCN on the basis of fully convolutional neural network to solve the problem that Faster R-CNN cannot realize the weight sharing of candidate region extraction network and feature extraction network. First, the fully connected layer in the network is removed and replaced with a convolutional layer, almost all network layer parameter sharing is achieved, which greatly improves the calculation speed of the network. Second, in order to solve the situation where the image distortion will change in the target detection task, a position-sensitive ROI Pooling layer is used in the network, and a "position-sensitive map" is obtained as an output to solve the problem.

Reference [12] added a branch for segmentation task based on Faster R-CNN, and obtained Mask R-CNN that can be used not only for target detection but also for instance segmentation. There are two innovations: (1) add a third branch under the two branches of the original classification and regression, and output the Mask of each ROI for segmentation; (2) improve the ROI Pooling of Faster R-CNN and propose ROI Align method is used to eliminate the original quantization error. Since the accurate pixel position is required for segmentation tasks, ROI Align uses the "bilinear difference" algorithm, which uses four true pixel values around the virtual point of the original image to jointly determine the pixel value of the point. The disadvantage of this network is that the detection speed is slower than Faster R-CNN, about 5f / s.

Reference [13] mainly improves the feature network of Faster R-CNN, and proposes a multi-scale hyper feature (Hyper Feature), which is a special space obtained by integrating multi-layer feature maps. The core of the algorithm is the layer-jumping extraction feature, which can not only obtain high-level semantics, but also obtain low-level high-resolution position information, which improves the detection effect of small targets, and the network has more advantages at high IOU, and the detection accuracy can be achieved 73%.

For the improvement of the feature extraction network, the FPN network proposed in [14] introduces a feature pyramid network to solve the problem of detecting multiple scales. The network has designed a top-down structure and horizontal connections, and based on this, it combines low-level features with high resolution and high-level features with rich semantic information for prediction. Therefore, the detection performance of this network for small targets is outstanding.

Reference [15] uses FPN as the detection framework network, and proposes a large mini-batch detection model MegDet, which can use a large mini-batch training network, and can use multi-GPU joint training to reduce the training time to 4 hour. For the shortcomings of large mini-batch training, the author proposes a learning rate selection strategy and a cross-GPU

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

batch normalization (BN) method. The two strategies can be used together to overcome the shortcomings of large mini-batch training and reduce training time Get higher accuracy

3. TARGET DETECTION ALGORITHM BASED ON TARGET REGRESSION

The detection algorithm based on target regression directly extracts the features of the input image, and calculates the class probability of the target object and the position of the target through the convolutional neural network.

3.1. YOLO, SSD, CornerNet Algorithm Introduction

Literature [16] proposed an end-to-end target detection algorithm using a regression strategy-YOLO. This algorithm does not need to generate a suggestion box, and uses the entire image as the input of the network to directly return the target position and category to the output layer . The YOLO algorithm uses the GoogleNet framework to divide an image into grids, each grid is only responsible for detecting the target whose center point falls within the grid; each grid needs to predict B bounding boxes, The output of is the bounding box coordinates (x, y, w, h) and the confidence Confidence. The confidence calculation formula is defined as:

$$Confidence = \Pr(Object) \times IOU_{pred}^{truth}$$

When the detection target center does not fall in the grid, confidence = 0, and if it is in the grid, it takes 1; IOU is the maximum intersection ratio between the prediction frame and the real frame. Assuming that C categories need to be predicted at this time, the final output tensor of the network is $S \times S \times (5 \times B + C)$. The advantage of YOLO algorithm is that the detection speed is fast, reaching 45fps, which can meet real-time detection; the disadvantage is that the detection accuracy is low, and when there are several object center points in a grid at the same time, it will cause a deviation in the detection position.

Reference [17] combines the advantages of YOLO and Faster R-CNN to propose an SSD algorithm. This algorithm has the same detection accuracy as Faster R-CNN, and the detection speed is faster than YOLO. The core of the SSD has three points (1) extracting feature maps of different scales for detection, large-scale feature maps are used to detect small objects, and small-scale feature maps are used to detect large objects; (2) borrow the anchor mechanism, each Feature maps can be generated through the Prior box layer. (3) The SSD uses a priori frames of different scales and aspect ratios, which can reduce the training difficulty to a certain extent.

In view of the shortcomings of YOLO, literature [18] proposed YOLOv2 on the basis of YOLO, and mainly made the following improvements: (1) Introduce batch standardization in all convolutional layers of YOLO, which can play a regularizing effect to a certain extent and reduce the over-fitting of the model; (2) Introduced High Resolution Classifier to increase the resolution of the training network to 48x48, and finally the mAP in the article is increased by about 4%; (3) Perform k-means clustering on the border of the training set to automatically obtain the prior information of the anchor; (4) A new full-convolution feature extraction network Darknet-19 is used, and anchor boxes are introduced to predict the bounding box; (5) A new joint training algorithm is proposed. The algorithm trains a classifier on the detection and classification data set at the same time, uses the detection data to train the object position, and the classification data set increases the amount of classification. Compared with the previous version, YOLOv2 has greatly improved the detection accuracy and detection speed.

Reference [19] was modified on the basis of YOLOv2, and got the YOLOv3 network. First, Darknet-53 was used as the basic network, and some residual modules were added, and shortcut links were set up between some layers. In addition, YOLOv3 uses a multi-scale prediction method, using k-means clustering to obtain a priori frames, a total of 9 anchors and

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

3 scales are selected, and each scale corresponds to an average of 3 anchors. Finally, YOLOv3 modifies the loss function. The softmax function is not suitable for multi-label classification, so logistic regression is used instead of softmax as the classifier, and the accuracy will not decrease after being replaced. The model of YOLOv3 is more complicated, so its detection speed is lower than that of YOLOv2, but its detection accuracy is greatly increased, especially after using multi-scale feature maps, the detection of small targets has also been greatly improved.

Reference [20] draws on the design of human pose estimation and proposes a target detection algorithm based on key points. The algorithm uses a fully convolutional neural network to detect a pair of key points in the upper left and lower right corners of the target position to obtain a prediction frame. This can solve the problem of too many useless prediction frames generated by the previous algorithm; And the idea of associative embedding is introduced to judge whether the pair of key points come from the same target prediction frame. If the pair of key points come from the same prediction frame, the distance between their embedding will be small. The article also proposes a new pooling layer corner pooling, which can help the network to more accurately locate the corners of the prediction frame. The algorithm obtained 42.1% AP on MS COCO and the detection speed was 1.4fps.

3.2. Improved Target Detection Network Based on SSD Algorithm

In the current improved network based on SSD, good detection performance is a major advantage of the network framework. To solve the problem that SSD is not robust enough to detect small targets, literature [21] proposed an improved detection network DSSD. The algorithm replaces VGG with the deep residual network Resnet as the backbone network. Using deep networks to extract features can give shallower feature maps better expression capabilities; and introduces a deconvolution layer, which uses deconvolution to form high-level semantics. The fusion of information and low-level semantic information improves the progress of detection. Literature [22] proposed a network RSSD for efficient feature fusion. The author uses pooling and deconvolution to perform feature fusion. On the one hand, the classification network is used to increase the connection of feature maps between different layers and reduce the appearance of duplicate frames; On the one hand, it increases the number of feature maps of different layers, so that it can detect small-sized objects. On the VOC2007 dataset, RSSD takes 48x48 pictures as input, mAP can reach 80.8%, and the detection speed is 16.6fps.

Since both DSSD and RSSD use a more complex network as the backbone network, this will reduce the detection speed, so the literature [23] draws on the idea of FPN and proposes the FSSD algorithm. FSSD takes VGG-16 as the backbone network and proposes a feature fusion network, which is to adjust some features in the network to the same size to obtain a pixel layer, and then use this layer as the base layer to generate a feature pyramid. The feature fusion network can fuse high-level semantic information and low-level semantic information to regenerate the feature pyramid. Compared with the previous two improved algorithms, the detection progress and speed have been improved.

Aiming at the pre-training problem of target detection network, the literature [24] uses DenseNet as the main frame and proposes a detection algorithm DSOD without pre-training. Since the pre-trained models in the past may not be well transferred to the detection model framework; and the pre-trained model architecture is relatively fixed, and it is difficult to modify it according to the specific detection task later. Therefore, this paper proposes an algorithm that can start training data from 0, without pre-training the model in advance, and the final detection effect can also be comparable to the pre-training model.

The general single-stage detection algorithm will produce a lot of useless anchors, resulting in algorithms such as YOLO and SSD always lagging behind the two-stage detection algorithm in the detection accuracy. Literature [25] is an algorithm with important meaning under the SSD algorithm. This algorithm has two main characteristics: (1) a new target detection algorithm,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

retinanet, is proposed. Firstly, RESNET is used as the backbone network to obtain higher dimension feature expression, then FPN network is used to integrate the extracted features, improve the richness of feature information, and finally classify and regress. (2) A new loss function focal loss is proposed, which adds a weight coefficient before the original cross entropy function, which can make a small amount of data have more influence and reduce the influence of large amount of data. The algorithm not only has a fast detection speed, but also greatly improves the detection accuracy.

4. DETECTION OF VIOLATIONS OF POWER JOB SITE BASED ON DEEP LEARNING

At present, with the rise of deep learning, it is no longer only rely on artificial detection, but rely on artificial intelligence for detection. Firstly, we use neural network to detect pedestrians, and stipulate some inaccessible areas on the operation site. If the block diagram obtained by pedestrian detection enters the area, it will be regarded as violation; then we use the area obtained by pedestrian detection as the area of interest, as the input of safety helmet and work clothes to detect the wearing of safety equipment, so the violation detection on the power operation site is divided into work Personnel testing and wear testing.

4.1. Staff Detection based on Deep Learning

The detection of electric field workers is based on an important application of pedestrian detection in target detection tasks. There are complex posture changes, uneven illumination and local occlusion in life scenes. Traditional machine learning cannot achieve ideals With the effect of deepening, the use of deep learning for staff detection can obtain more accurate detection results. Compared with general object detection, pedestrian detection focuses on real-time applications (autonomous driving and video surveillance scenarios), so it has higher requirements for speed and accuracy.

Literature [26] proposed an efficient pedestrian detection network. In the SSD-based detection framework, a progressive positioning fitting module was introduced, and multiple positioning modules were trained using the ever-increasing IOU threshold. The specific steps are to use a series of detection anchor boxes applied to each step. At each step, use the regression anchor box to optimize the classifier. As the anchor box is gradually refined, more positive samples are obtained, and later use of higher anchor boxes Thresholds are used for training to produce more accurate positioning. Experiments prove that multi-step prediction is the key to improve the detection and positioning accuracy. The selection of appropriate positive and negative samples plays a very important role in the training process of the detector. Reference [27] proposes a new perspective of advanced semantic feature detection to solve the problem that the traditional target detection algorithm based on sliding window or a priori frame method requires complicated configuration, and converts the coordinates and scales of pedestrian center points into advanced semantics Feature information. The algorithm simplifies pedestrian detection to a task of calculating the heat map of the pedestrian center through a convolution network and directly predicting the coordinates and width and height of the center point. Experiments show that this method can accurately detect the position of pedestrians.

Since the occlusion situation is very common in pedestrian detection tasks, in real scenes, the occlusion situations encountered by pedestrian detection tasks mainly fall into two categories: one is mutual occlusion between the detected pedestrian individuals, introducing a lot of interference information, which will cause False detection; second is the occlusion of the detection object by some other objects, which will lose some information of the target and cause missed detection. Reference [28] proposes to solve the problem of misdetection by setting the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

loss function (Repnlsion Loss) for the situation where dense pedestrians cause mutual interference. This function calculates the loss functions of the prediction box and the real target box, and the adjacent real target box and the adjacent different target prediction boxes respectively, so that the distance between the prediction box and the real target box is reduced, and the The distance increases, which greatly reduces the false detection rate. Reference [29] takes both occlusion situations into consideration and proposes a new detection network OR-CNN. The network follows the Faster R-CNN framework. To deal with the problem of occlusion, two strategies are proposed to solve the two problems of dense pedestrian occlusion and occlusion of unrelated objects. First, a loss function is designed in the RPN network to make multiple The candidate frame matched to the real target is as close as possible; then in the second stage, the human body is divided into five parts according to the a priori information to extract features separately, and then these local features are combined with the global feature weighted summation to obtain the fused feature; finally Then perform classification and regression.

4.2. Hard Hat and Workwear Inspection Based on Deep Learning

For the power job site, it is a major trend to apply deep learning to the power job site for detection. According to the principle of the target detection network, the detection of hard hat workwear is also mainly divided into two types of algorithms. One is to first extract the region suggestion frame, and then to classify and locate the target; the other is to directly regress the input image. Get the difference and position of the target.

In order to meet the needs of the detection tasks of safety helmets and work clothes in electric work sites, literature [30] proposed an improved YOLOv2 safety helmet detection algorithm. This algorithm first introduced a dense network to fuse shallow low semantic information and deep high semantic information to The detection accuracy of small targets; then introduce the lightweight structure of MobileNet to compress the detection network, which increases the practicality of the network; and finally start training from 0 on the data set made by yourself. The detection accuracy of the algorithm is 87.42%, but the detection speed is 148fp / s. With the introduction of the YOLOv3 detection network, the literature [31] proposed an improved network based on YOLOv3 to detect safety helmets. The algorithm uses a pyramid structure to fuse feature maps at different levels to obtain three prediction maps, and uses these three maps for prediction and positioning; but because this article is pre-trained on the COCO data set, it is adapted to the helmet detection For the task, the k-means algorithm is used to cluster the prior frame of the self-made training set, so that the detection accuracy and speed of the improved network are improved compared to the original network.

Compared with the single-stage detection in the previous two documents, the literature [32] uses a two-stage detection algorithm to propose an improved helmet detection algorithm based on Faster R-CNN. Based on the original Faster R-CNN, the algorithm uses multi-scale training and increases the number of anchor points to improve the network's robustness to targets of different sizes, and introduces an online difficult sample mining strategy to prevent the imbalance of positive and negative samples; Finally, the multi-component combination method is adopted to detect the detected target to propose the false detection target. The accuracy of this algorithm is greatly improved compared with the previous two documents, but due to the characteristics of two-stage detection, the detection speed is reduced. Reference [33] used an SSD network to detect wearable devices, replaced the backbone network VGG-16 in the original algorithm with the MobileNet network, and produced its own security device data set based on the VOC2007 data set standard. The algorithm uses its own data set to train the model, recognizes eight kinds of security equipment including hard hats and work clothes, and judges whether the worker is wearing security according to the highest confidence result of the worker's multiple photos from different angles device. The algorithm in this paper can well

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

detect the staff's equipment wearing situation, also meet the real-time requirements, and has good stability and anti-interference.

5. SUMMARY AND OUTLOOK

This article mainly summarizes the target detection based on deep learning and its application in the power industry. It summarizes the single-stage target detection and the two-stage target detection framework, and analyzes the improvement network of the two types of methods in detail. The advantages and disadvantages of these two methods are discussed, and finally the application of target detection based on power industry tasks is briefly explained. The advantage of the single-stage detection network is its high speed, but it inevitably leads to its low detection accuracy. However, due to the continuous introduction of improved algorithms in recent years, the single-stage detection network has greatly improved in the direction of detection progress. In the same way, the two-stage detection network due to the existence of the regional recommendation network, the early calculation process is complicated, which makes the detection speed slower but the accuracy is high. The application scenarios of target detection involve multiple aspects, This article mainly reviews the application of power operation scenarios. Hard hat and staff detection are two major hot spots in the power industry, most of them rely on traditional machine learning algorithms for detection, Due to the high speed and accuracy of deep learning, they have now replaced the traditional detection algorithms and become the mainstream. However, there are still many defects in the adaptability of deep learning to the real environment of this task, and the actual detection task based on deep learning still needs a long development time.

At present, there are still many aspects of the target detection algorithm based on deep learning that have not been properly resolved. The main research directions and problems are as follows:

The detection sensitivity of the general detection network to small targets is not enough, large-scale occlusion, and detection targets are similar to the background and difficult to detect. Further research on these issues is needed.

The detection effect for some small numbers of samples is not ideal. At present, most training methods are pre-training on a large number of data sets, and then fine-tuning the model, which leads to the model relying heavily on pre-training, often moving to real tasks Adaptability is not strong.

The detection speed and accuracy cannot be possessed at the same time. The single-stage detection will discard the detection accuracy, otherwise the two-stage detection will discard the detection speed. At this stage, an efficient detection algorithm with interval speed and accuracy should be designed, which is particularly important for the target detection direction.

ACKNOWLEDGEMENTS

Sichuan Science and Technology Department Project (2017JY0338, 2019YJ0477, 2018GZDZX0043, 2018JY0386); Artificial Intelligence Sichuan Key Laboratory Project (201 9RYY01); Sichuan Institute of Technology Talent Introduction Project (2017RCL53); Enterprise Informatization and Internet of Things Measurement and Control Technology Laboratory project (2018WZY0 1); Sichuan Institute of Technology (specialist) workstation project of Sichuan Institute of Technology (2018YSGZZ04); Science and Technology Project Funding of State Grid Corporation of China (contract number: 521997180016).

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

REFERENCES

[1] Yuan Luo, Boyu Wang, Xu chen. A Summary of Research on Target Detection Technology Based on Deep Learning[J]. Semiconductor Optoelectronics, 2020, 41(01): 1-10.

[2] Ríha Kamil, Mašek Jan, Burget Radim, Beneš Radek, Závodná Eva. Novel method for localization of common carotid artery transverse section in ultrasound images using modified Viola-Jones detector. [J]. Ultrasound in medicine & biology,2013,39(10).

[3] Hou Beiping, Zhu Wen. Fast Human Detection Using Motion Detection and Histogram of Oriented Gradients. 2011, 6(8):1597-1604.

[4] M. Fatih Talu, Mehmet Gül, Nuh Alpaslan, Birgül Yiğitcan. Calculation of melatonin and resveratrol effects on steatosis hepatis using soft computing methods[J]. Computer Methods and Programs in Biomedicine,2013,111(2).

[5] Xiong Cong, Wang Wenwu. Research on pedestrian detection technology based on DPM model [J]. Electronic Design Engineering,2014,22(23):172-173.

[6] Zeng Jiexian, Cheng Xiao. Pedestrian detection in traffic scenarios combined with single and double pedestrian DPM models [J]. Acta Electronica Sinica, 2016,44(11):2668-2675.

[7] Girshick R, Donahue J, Darrell and T, etal. Rich feature hierarchies for object detection and semantic segmentation[C]//2014 IEEE Conf. on Computer Vision and Pattern Recognition, 2014;1-21.

[8] He K, Zhang X, Ren S, etal. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Trans. onPattern Analysis & MachineIntelligence, 2014, 37(9): 1904-1916.

[9] Girshick R. Fast R-CNN[J]. Computer Science, 2015(4): 169-178.

[10] Ren S, He K, Girshick R, etal. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Trans. On Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[11] Dai J, Li Y, He K, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks[J].2016.

[12] He K, Gkioxari G, Dollar P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):1-1.

[13] Kong T, Yao A, Chen Y, et al. HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection[J].2016.

[14] LIN T Y, DOLLAR 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.

[15] Peng C, Xiao T, Li Z, et al. MegDet: A Large Mini-Batch Object Detector[J].2017.

[16] Redmon J, Divvala S, Girshick R, etal. You only look once: unified, real-time object detection[J]. arXive-prints,2015(6):2640-2650.

[17] Liu W, Anguelov D, Erhan D, etal. SSD: Single shot multibox detector[J]. ECCV2016: Computer Vision, 2016: 21-37.

[18] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//IEEE 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.

[19] Redmon J, Farhadi A. YOLOv3: An incremental improvement[C]//2018 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2018: 2767-2773.

[20] LIU W, ANGUELOV D, ERHAN D, etal. Ssd: Single shot multibox detector[C]// European conference on computer vision. Springer, Cham, 2016: 21-37.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0010

[21] Fu C, Liu W, Ranga A, et al. DSSD: Deconvolutional Single Shot Detector.[J]. arXiv: Computer Vision and Pattern Recognition, 2017.

[22] JEONG J, PARK H, KAWK N. Enhancement of SSD by concatenating feature maps for object detection[J]. arXiv preprint arXiv:1705.09587, 2017.

[23] LI Z, ZHOU F. FSSD: feature fusion single shot multibox detector[J]. arXiv preprint arXiv:1712.00960, 2017.

[24] Shen Z, Liu Z, Li J, et al. DSOD: learning deeply supervised object detectors from scratch [C]/ /IEEE International Conference on Computer Vision, 2017: 1937-1945.

[25] Lin T, Goyal P, Girshick R B, et al. Focal Loss for Dense Object Detection[J]. international conference on computer vision, 2017: 2999-3007.

[26] Wei Liu, Shengcai Liao, et al. Learning Efficient Single-stage Pedestrian Detectors by Asymptotic Localization Fitting[C]//Proceedings of the European Conference on computor Vision (ECCV), 2018.

[27] Wei Liu, Shengcai Liao, Weiqiang Ren, et al. High-level Semantic Feature Detection: A New Perspective for Pedestrian Detection[J]. 2019IEEE Conf.on Computer Vision and Pattern Recognition, 2018: 2167-2173.

[28] Wang Xinlong, Xiao Tete, Jiang Yuning, et al. Repulsion loss:detecting pedestrians in a crowd[C]//2018IEEE/CVF Conf. On Computer Vision and Pattern Recognition, 2018, 1(4):7774-7783.

[29] Shifeng Zhang, Longyin Wen, et al. Occlusion-aware R-CNN: Detecting Pedestrians in a Crowd[C]//Proceedings of the European Conference on computor Vision(ECCV), 2018: 637-653.

[30] Ming Fang, Tengteng Sun, Zhen Shao. Fast helmet wearing detection based on improved YOLOv2[J]. Optical Precision Engineering, 2019, 27(05): 1196-1205.

[31] Hui Shi, XianQiao Chen, Ying Yang. Improve YOLO v3's helmet wearing detection method[J]. Computer Engineering and Applications, 2019, 55(11): 213-220.

[32] Shoukun Xu, Yaru Wang, et al. Research on Hard Hat Wear Detection Based on Improved FasterRCNN[J]. Computer Engineering and Applications: 1-6[2020-04-14].

[33] Chuntang Zhang, Licong Guan. SSD-MobileNet-based miner security wearable device detection system[J]. Industrial and mining automation, 2019, 45(06): 96-100.