Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12720

Research Article Open Access

A System of Image Recognition-Based Railway Foreign Object Intrusion Monitoring Design

Beiyuan WANG, Lingqi WANG*, Chuanya GU

Shanghai Bearing Technology Research Institute Co., Ltd., Shanghai, 201801, China

****Corresponding Author:*** *Lingqi WANG, E-mail: 327611126@qq.com*

Abstract

The monitoring system designed in this paper is on account of YOLOv5 (You Only Look Once) to monitor foreign objects on railway tracks and can broadcast the monitoring information to the locomotive in real time. First, the general structure of the system is determined through demand analysis and feasibility analysis, the foreign object intrusion recognition algorithm is designed, and the data set required for foreign object intrusion recognition is made. Secondly, according to the functional demands, the system selects a suitable neural web, and the programming is reasonable. At last, the system is simulated to validate its functionality (identification and classification of track intrusion and determination of a safe operating zone).

Keywords: *Railway; Deep learning; YOLOv5; Image intelligent recognition; Obstacle detection*

1 Introduction

Object detection in YOLO is seen as a regression concern, which determines the position of a bounding box and works out the probability of it attributing to a particular category. Use a single neural network to predict bounding box and class probabilities directly from the complete image in one evaluation [1].

In the official code of YOLOv5 used in this project, the target detection networks given are respectively YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x models. YOLOv5 code network file is yaml format, network structure is very concise, using Leaky ReLU and Sigmoid as YOLOv5 activation function [2].

YOLOv5 passes each batch of training data through the data loader and simultaneously enhances the training data. The data loader performs three kinds of data enhancement: scaling, color space adjustment, and Mosaic enhancement (Mosaic data enhancement does effectively solve the most troublesome "small object problem" in model training, that is, small objects are not as accurately detected as large objects).

The authors of YOLOv5 provide us with two optimization functions, Adam, and SGD, and both have preset training hyperparameters to match (Default: SGD). The authors of YOLOv5 recommend that Adam is a more appropriate choice if you need to train smaller custom datasets, although Adam generally has a lower learning rate than SGD. But if you train large data sets, SGD works better for YOLOv5 than Adam. In fact, there has been no unified conclusion on which SGD or Adam is better in the academic community, depending on the actual project situation. If the optimization function is trained on a small custom data set, ADAM is a more suitable choice, but if the training of a large data set, the SGD optimization function is more frequently used.

2 Feasibility Analysis

2.1 Technical feasibility

The YOLOv5 network model is the underlying structure of the back end of the system. This system wants to achieve efficient and high-speed recognition, in that the YOLOv5m network model [3] is adopted for data training. The data set was processed and annotated using the LabelImg tool to obtain the normalized data needed for YOLOv5 training. To sum up, method used in this paper is simple and clear, and can be completed in a short time. Besides, the front- and back-end interactions are hierarchical, and technically feasible.

2.2 Operational feasibility

The system uses a simple framework, reasonable training mode and design language to ensure the order and efficiency in the development process. Users can quickly complete their desired results through a clear interface UI. Therefore, the system is feasible in operation.

Received on October 19, 2023; Accepted on November 27, 2023

2.3 Economic feasibility

The development cost is not high, since YOLOv5 is an open-source project on the Internet, it can be downloaded for free use. The back end uses the CPU for training. Users only need to access the system and upload pictures while let the system do the rest. In general, the development cost of this project is small, so it's economically feasible.

3 Design of the System

3.1 Foreign object recognition algorithm design

Based on YOLOv5 algorithm model, this paper realizes the monitoring function of railway foreign objects intrusion, therefore, by obtaining the best training model (. best-py) after training the data set through the YOLOv5 algorithm model is the key to this function, and then proceed to the next step of foreign objects category judgment. The program background can call the YOLOv5 algorithm model, and the PyQt5 module in Python is used to realize railway foreign object monitoring interface [4].

After the training of YOLOv5 algorithm model is completed, the training file will be output to the user. User can set the format of general input, usually used .jpg format. The system will monitor the input information by invoking the training file. After monitoring, the interface will display the category size and influence range of foreign objects invaded by railway.

3.2 General structure of system

With the continuous progress of artificial intelligent machine learning, the Chinese rail transit industry can cut the cost of monitoring the railway. Due to the gradual emergence of image recognition assisted artificial monitoring, the vital research is to solve the security and dependability of the train undertaking.

The biggest threat the train faces in the process is the impact of foreign objects on the railway, and the impact is mainly caused by human, animals, or stones etc. on the railway, hindering the normal train speed conditions. To avoid similar accidents, the railway safety department will set up warning signs or protective measures such as guardrail accordingly, but it's impossible to strictly inspect or clean every area all the time where the train travels, and because the standard rate of the train is generally above 150km/h, the operation of deceleration or emergency braking cannot be carried out in time when the railway foreign objects in the line of sight is encountered. Therefore, it's crucial to design a railway foreign object intrusion monitoring system reminding train drivers to implement different train speeds when encountering railway foreign objects at different distances.

The monitoring process's interface of the entire railway foreign object intrusion system is carried out according to the overall system flowchart as indicated in Figure 1.

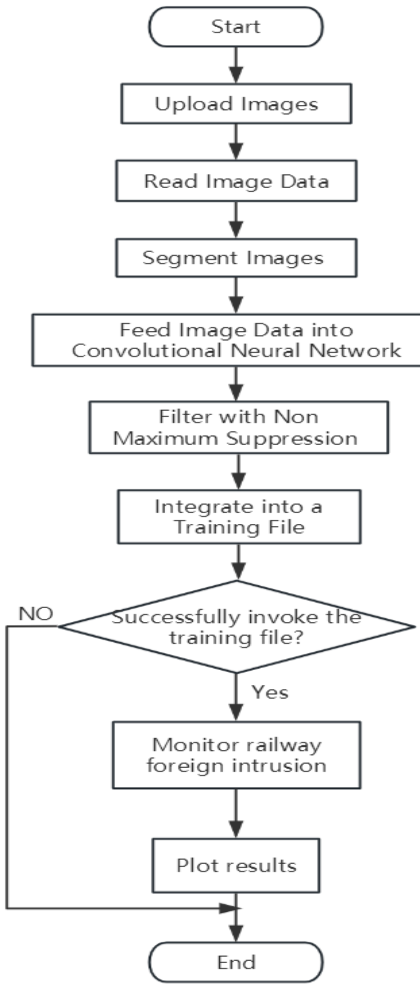

Figure 1 Flowchart of the system's overall structure

The specific process of this system is to upload the pictures that need to be monitored first and wait for the system to automatically read the images. The next step is to determine whether the system can successfully invoke the training file, which is trained previously. Therefore, a good recognition effect relates to the quality of the previously trained files. If the training file can be successfully called, the system will carry out the railway foreign object monitoring module, besides, it will return the foreign object position and affected region to the system after the checking is fulfilled. The system will classify different affected areas of the foreign object to different traveling speeds of the train according to the existing safe area range and transmit the data to the cab to give the driver the driving speed indication. Finish a series of commands to end the program of the system, exit the operation; If you cannot successfully call the training file, then directly end the program of the system, and exit the operation.

Based on the in-depth study of domestic and foreign railway foreign object intrusion monitoring technology, this subject intends to adopt YOLOv5 algorithm to conduct real-time monitoring of railway foreign objects intrusion penetration. Due to the constraints of the test environment, the experiment cannot be carried out while the train is running, so this subject will only analyze and discuss the theoretical test in the following chapters.

The system preparation work designed in this project: download the YOLOv5 project to the local, configure Python3.9, and the environment operating system is Windows 10; IDE: PyCharm; Python version: anaconda Python3.9. This system uses the CPU of this device for training. If large-scale training, image and video recognition and other operations are carried out, the computer should be equipped with GPU [5].

3.3 Creation of data set

Since, there are no training samples for railway foreign object intrusion monitoring in the existing known data sets, this topic intends to adopt the method of manual labeling to build a group of suitable data sets to ensure that the algorithm model can obtain good training results. With Baidu and other search engines as the main data set source, the web crawler developed by Python was used to collect and download related pictures. According to the actual railway operation, five types of foreign bodies are mainly monitored, including rock, person, animal, car, and rabbit. Integrate all image data and unify into.jpg format. A total of 324 data sets were used to train the algorithm model. After consulting the literature on image recognition, we can get a conclusion that the optimal data set required for computer image training is at least 300 pictures. Therefore, during the test, we plan to train more than 300 pictures first, and then continue to consider increasing the number of sample data sets if the effect is not as expected.

After sorting, the existing data sets need to be labeled to improve the performance of the algorithm model. LabelImg was used to manually mark 324 pictures in the foreign objects' image data set on the railway and label them with categories. The connection of LabelImg tag is indicated in Figure 2.

After manual annotation is completed, there are mainly two folders for storing related images and annotation information. One folder is called 'images and the other folder is called 'label'. The label folder stores the label information (in .txt format), that is, the label file for target detection. The label information corresponds to the name of the image data set file one by one.

Figure 2 The interface of LabelImg

Open the designated mark file, as shown in Figure 3 normalized data of railway foreign object categories and prior box coordinates. Each line in the .txt file represents a target, the x and y coordinates of the center point post-normalization, in addition to the breadth w and tallness h of the objective box. Besides, under the label folder, there is also a .txt file called 'classes', which contains information about the class names of all the label boxes. After the annotation of 324 images in the training data set, the pre-processing of the data set is completed.

```
0 0.382422 0.787500 0.757031 0.338889
1 0.417969 0.793750 0.031250 0.068056
```

Figure 3 Normalized data of railway foreign object categories and prior frame coordinates

3.4 Regional division function

Currently, two main methods of lane line detection exist domestically and internationally: one is a model-based approach, the other is a feature-based technique. The model-based detection method is to assign the lane to a sophisticated arithmetical model and fit the road line based on the model. The principle is to match the appropriate curve model for the lane lines according to the geometric features of the lane lines on the structured roadway, and then use the least square manner, Hough Transform to be the right size for lane principles [6]. The widespread mathematical representations are linear, allegorical and spline models. This method has strong anti-interference ability to noise, but it also has drawbacks, that is, one lane model cannot adapt to multiple road scenes at the same time.

Aspect-based detection method is to obtain the aspect information of lane lines from the pavement area through edge detection or threshold segmentation according to the feature information of lane lines themselves. This method has high requirements on the edge features of the lane lines and can obtain better results when the edge features are obvious, but it's very thoughtful to sound and has impoverished robustness.

This project adopts a Transmission Hove based road segment recognition algorithm, that is, a model-based recognition method to complete the division of this area. Figures 4 and 5 are the original diagram and the effect diagram after adding the algorithm respectively.

Figure 4 Original diagram

Figure 5 Effect diagram after adding the algorithm

First, the picture is changed into HLS hue space. And then using the edge detection and threshold methods to detect the lane lines [7], the edge detection results are combined with the color detection results, and the L-channel is used to suppress the non-white information. According to the lane line information detected above, histogram method and sliding window method [8] are used to determine the exact position of the track line. The safe region algorithm can realize 4 vertices in the return region, that is, the two regions in Figure 3.5 are separated into the crimson region and the yellow region. The algorithm can return a total of 8 vertices, and then judge which area in the crimson and yellow area the foreign object belongs to according to the center coordinates of the object position recognized by YOLO, to realize the train deceleration or braking operation.

3.5 Foreign object recognition function

In the railway foreign object identification module, the user uploads the images that need to be detected, and the system stipulates that the images can only be uploaded in .jpg format. If the uploaded format doesn't belong to the specified format, it won't be uploaded. If the user isn't satisfied with the uploaded picture, or if there is an error in the uploaded picture, the picture can be cleared and then uploaded again. If the user confirms that the uploaded picture is correct, the system can detect the current picture, and the successful detection result is returned to the system. The widespread procedure is indicated in Figure 6 Flowchart of railway foreign object recognition module.

The specific process of this system for foreign object recognition function is that the background monitoring personnel first upload the image that needs to be monitored, and if the image is successfully uploaded, wait for the system to automatically read the image; If the upload fails, upload the image again. The next step is to determine the user's next response, which can be divided into two steps, either to start recognizing the image uploaded by the user, or to clearly exit the image. If the user performs the identification undertaking, the system affects regardless of the successful identification. When the identification is successful, the monitoring result will be returned to the display without ending the program. While the discovery is not triumphant, the program of the system is ended directly, and the operation is withdrawn.

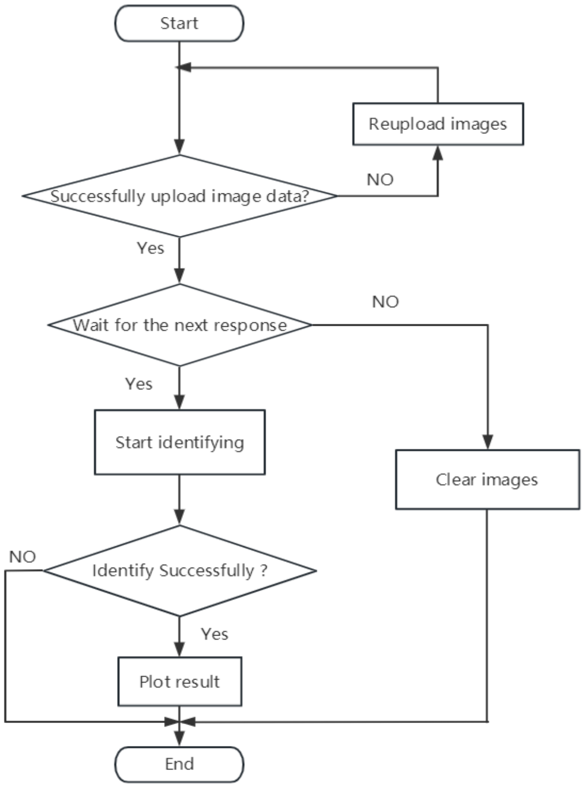

Figure 6 Railway foreign object identification module flowchart

Figure 7 shows the test recognition types and accurate results. It can be seen from the figure that the system has a high accuracy rate for the recognition effect of rail and a relatively low accuracy rate for the recognition effect of person, but the category judgment of both is correct. In view of the accuracy problem, in terms of subsequent training and data set collection, emphasis should be placed on the training and testing of 5 types of foreign bodies required for this subject (human, animal, stone, rabbit, car).

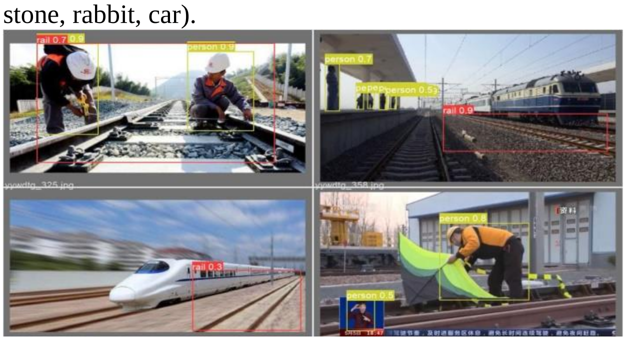

Figure 7 Recognition type and accuracy (partial)

4 Test and Analysis

4.1 Model training

YOLO algorithm is mainly composed of two parts, one is classification, the other is detection. The first method uses the image network for pre-training, while the second one uses precise settings on the detection data. YOLOv5 pre-training is based on COCO data set, its official pre-training weight is based on FP16, and the training weight is based on FP32, which can double the storage space. In common object detection algorithms, the image is uniformly reduced to a standard size (such as 416×416,608×608, commonly used by the YOLO algorithm) and transmitted to the detection network. However, YOLOv5 has been improved: In the actual use of the project, because various images, the size of each side is different, so after magnification, the size of the black edge on both sides is also discrepant, but more information will be added, thus affecting the speed of reasoning. In this way, on the original drawing, the adaptive addition of the smallest black edge can increase the reasoning speed by 37%.

When YOLOv5 code is trained, first modify the training configuration. To create a .yaml file, it's necessary to create a mask_data.yaml file in the data folder primarily, then to modify the address and category parameters of this file. The category name is changed to the data set label name used in this project (rail, animal, person, rock, car, rabbit), that is, the arrangement folder of the data set is modified, as indicated in Figure 8, and training can be carried out after modification.

```
# train and val data as 1) directory: path/images/, 2) file: path/images.txt, or 3)
train: D:/360Downloads/lunwen/ImgSpider/YOLO_Rail/score/images/train
val: D:/360Downloads/lunwen/ImgSpider/YOLO_Rail/score/images/val

# number of classes
nc: 6

# class names
names: ['rail','animal','person','rock','car','rabbit']
```

Figure 8 Modify the configuration file of the data set

4.2 Test result

Before model training, the labeling of sample data sets was divided into two days, and continuous labeling was not done, which affected the recognition of some classes in some labeling information, resulting in a certain deviation in the final picture identification of alien thing types. The confusion matrix of the training model is indicated in Figure 9.

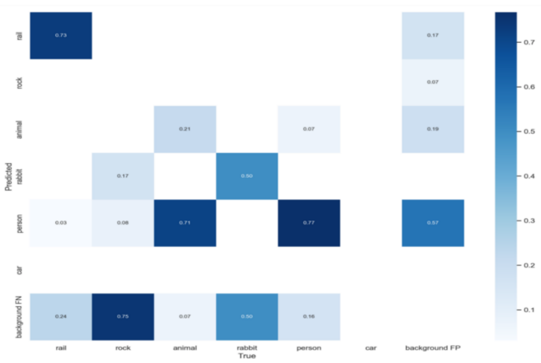

Figure 9 The confusion matrix of the training model

The explanation is as follows: the x-coordinate is the correct classification, and the y-coordinate is the classification predicted by the model.

The target detection model usually outputs many detection boxes, and we measure the detection effect of the model by counting and calculating whether each detection box can detect various proportions of the target. Therefore, we will divide the detection box into the following 4 cases, and the basis for judgment is mainly through calculating the intersection ratio IoU. The IoU intersection ratio function can be used to evaluate object detection algorithms [9]. By calculating the ratio of intersection and union of two bounding boxes, it can be used to determine whether the object detection algorithm is running well. The two compelled container are compelled box and ground reality reference standard respectively. Therefore, the IoU is designed to evaluate whether the object positioning algorithm is precise, that is, the IoU measures the relative size of the overlap between two compelled boxes. A way to measure positioning accuracy: As a general convention for computer detection, if the IoU is above 0.5, the prediction is acceptable; If the IoU equals to 1, the predicted and actual boxes overlap perfectly. The manual threshold for an IoU is 0.5. If you want to be strict, you can set a higher threshold, and the higher the threshold is, the better the accuracy is.

The confusion matrix, also known as the error matrix, is a level arrangement used to act for exactness evaluations and is represented by a matrix of n lines and n columns. The specific assessment lists of mistake matrix contain overall accuracy, cartographic accuracy, user accuracy, etc. These exactness lists can show the exactness of image categorization from different aspects.

According to the confusion matrix of the training model, it isn't difficult to draw a conclusion that the system has a total of 6 categories, of which 3 categories have recognition rates above 0.5, and 3 categories have recognition rates below 0.5, namely rock (1), animal (3) and car (6). The identification speeds of these three classes are relatively small. In the subsequent iterations, emphasis will be given to strengthening training of these categories.

In the P diagram of exactness speed, the level coordinate is confidence, and the upright coordinate is accuracy. In general, the accuracy P-graph raises with the increase of trust. After the training of this model, the average accuracy rate of all categories is around 0.901, and most bends indicate a slow rising tendency with the raise of confidence, among which rail has a better accuracy rate. It can be inferred that rail is better than other categories in terms of data set annotation, or it may be because rail has more pictures. Enables the program to recognize specific data more effectively, as shown in Figure 10.

The loss layer is the end point of the CNN. Blob two binary large objects can be entered. Blob contains the predicted result and the actual mark. After carrying out undertakings on the given two inputs, the Loss Function of the existing network, that is, L(θ) is produced, with θ standing for the vector space of the weight in the network. The intent of machine learning is to look for the best L(θ) that is available. Therefore, the weight of θ is the lowest, which can be accomplished by a series of optimization algorithms (such as SGD, etc.). The loss function is obtained through the calculation of forward transfer and is used as the beginning of backward transfer.

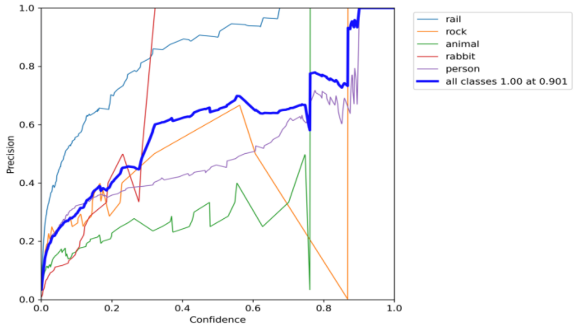

Figure 10 P-Graph

The YOLO series is based on objective scores, class probability scores, and border regression scores to calculate losses. YOLOv5 uses GLOU loss as the boundary box loss, YOLOv5 uses binary cross entropy and partial logarithm loss function to calculate the class probability and target score loss, and we can also use the fl_gamma parameter to activate the focus loss to calculate the loss function.

When the training network is normal, the loss value of the training set decreases and the val_loss value of the test set decreases, which is the most ideal situation. Box_loss refers to the loss function for predicting the position of the box, cls_loss refers to the class loss function, and obj_loss refers to the loss function for whether there are objects in the grid. As can be seen from Figure 11, the box_loss on the extreme left suggests that the training network is functioning correctly. The obj_loss and cls_loss images in the center demonstrate that the loss value of the training set is fluctuating, while the loss value of the test set stabilizes and increases. When the loss value of the test set is stable, the network overfitting can be solved by the following two methods: ① When the data set is not problem, the dropout layer can be added to the "middle depth" of the network or the network depth can be reduced successively; ②When there is a problem with the data set, all the data sets can be shuffled and redistributed. When the loss value of the training set and the loss value of the test set rise at the same time, this is mainly due to the structural design of the network, the setting of instruction hyperparameters, and the definition of the data set.

When the value of the loss function is small, it indicates that the forecasted value is nearer to the true value, producing a better effect. Box_loss is the loss function for the location of the prediction box, and cls_loss is the loss function for the class, and obj_loss refers to the loss function of whether there are objects in the grid. It can be seen from Figure 11 that the position and category of the prediction box in this data set are within a good measurement range, while the training effect of .obj is not good. As can be seen from the graph, the fluctuation is quite drastic.

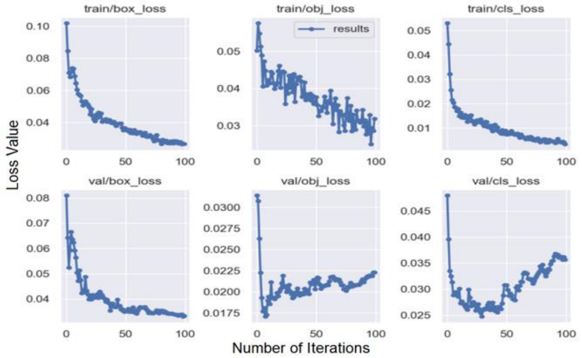

Figure 11 Loss function result curve

5 Conclusion

Based on the demand analysis of the overall system, the feasibility of railway foreign object intrusion monitoring system based on YOLOv5 was determined in this subject. The overall framework and process were designed through the improvement means of computer program engineering, and the program was grown. The railway foreign object intrusion monitoring system has an easy-to-understand interface, which facilitates users to expedite the monitoring process and accomplish the aim of high rate and high effectiveness.

The model recognition rate in the first training is less than 30%. After analysis, the selection of data sets has problems, and the overall data set features are not obvious, which has a great obstacle to training. The model recognition rate of the second training is about 50%. After analysis, it is found that the number of data sets is not enough, and there are fewer types of railway foreign objects, which leads to unsatisfactory overall training effect. After knowing the reason, the whole data set was adjusted, and training file and other parameters have been set. In short, the training model with the recognition rate of 73% was obtained, which was in line with the expected level.

References

[1]Ping H, Ming C, Kexin C, et al.. A combined real-time intelligent fire detection and forecasting approach through cameras based on computer vision method[J]. Process Safety and Environmental Protection,2022,164.

[2]GE Jing. Analysis of the development trend of computer image processing technology [J]. Information Technology Computer Products and circulation.2019(5):45-47.

[3]WANG Quandong, Yang Yue, Luo Yiping, et al.. Review on detection methods of foreign bodies in railway

encroachment [J]. Journal of Railway Science and Engineering,2019,16(12):3152-3159.
[4]Zheng Yuanpan, Li Guangyang, Li Ye. Application of Deep learning in image recognition [J]. Computer Engineering and Applications, 2019,55(12):20-36.
[5]Chen Chao, Qi Feng. Review on the development of Convolutional neural networks and their applications in computer vision [J]. Computer Science, 2019,46(3):63-73.
[6]Chen Rongbao, Zhao Dan, Wang Qianlong. A speed measurement method for vehicles moving ahead based on Image processing [J]. Sensors and Microsystems,2018,37(04):17-19+23.
[7]Xiang Jun, Zhang Jie, Pan Ru Ru et al.. Contour Extraction of Printed Fabric Pattern with Smooth Texture [J]. Journal of Textile Science,2017,38(11): 162-167.
[8]Girshick R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
[9]Joseph Redmonl. You Only Look Once: Unified,Real-Time Object Detection.[J]. CoRR,2015,abs/1506.02640.