ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

Application of YOLOv3-based Industrial Image Recognition

Wenjie Liu [1, a], Yuming Qi[1, b, *] and Zhongmin Wang[1, c]

[1]Institute of robotics and intelligent equipment, Tianjin University of Technology and Education, Tianjin 300222, China

[a]867911017@qq.com, [b]chigym@163.com, [c]873746812@qq.com

Abstract

This paper proposed an industrial image recognition system based on YOLOv3 in automatic production line. Complete the implementation of detection and run the identification program by using a Windows computer. The YOLOv3 is used for detection framework, and its network can directly predict the bounding box and classification probability in an image frame. Finally, the product identification results are output according to the images recognized by industrial cameras in automatic production line. Its recognition accuracy can reach 90% for magic cube.

Keywords

Industrial image recognition; YOLOv3; Object detection; Automation.

1. INTRODUCTION

Traditional garbage sorting is in the form of pipeline + manual sorting. Some electronic garbage and chemical garbage do great harm to human body. Vision-based intelligent manipulator system can automatically identify different kinds and sizes of garbage, identify reusable parts and automatically realize sorting. In traditional manipulator object grabbing, single object with fixed position and posture is grabbed by manual teaching. In the conventional visual servo system, the position of the camera and the manipulator is relatively fixed, the object is single and the position and posture are fixed. Because the information of working environment, object type, shape, size and posture can not be perceived independently, the object capture method of the traditional manipulator system has many uncertainties, which can only be made in a specific environment. Use. Many problems, such as the existence of multiple objects, the different types and sizes of objects, the change of position and pose of objects, the relative position of camera and manipulator is not fixed, make the traditional visual manipulator system unable to complete complex grasping tasks. In order to grasp autonomous objects in natural environment, researchers continue to improve visual-based manipulator object grasping methods. Literature [1] introduces the traditional method of robot arm object capture based on image binary image position detection. Literature [2]-[4] proposes a visual servo control scheme based on Jacobian matrix estimation. The above methods are all relatively fixed in a single object scene. In the stage of visual recognition, most methods still set features manually, but in the actual grasping task, the size, shape, external illumination intensity, angle change and sampling angle of the target object are uncertain. The robustness of object features extracted by traditional feature extraction methods is poor, and it can not adapt to new objects and changing environment.

In recent years, machine vision is more and more used in human transportation, logistics and security, which plays an important role in replacing the feature recognition of artificial repetitive machinery. With the development of in-depth learning, many researchers at home and abroad use CNN (convolutional neural network) for image recognition. Huang Yue [5]and

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

others use the algorithm to recognize the car icon, which improves the recognition accuracy and stability of the system. Wang Fujian [6] and others designed the vehicle information detection and recognition system, which designed different color recognition to make the classification more detailed. But there is a common problem in traditional algorithms, that is, the detection and recognition speed is slow, the amount of calculation is large and it is not suitable for miniaturization [7]. YOLO (you only look once) algorithm proposed by Redmond can improve the operation efficiency and speed up the detection while ensuring the recognition accuracy. Aiming at the grasping environment of multi-target, cluttered environment, unstable object pose, size and relative position of camera and manipulator, this paper presents an improved YOLOv3 algorithm to realize automatic detection of multi-object in cluttered environment, overcome the problem of high overlap rate of object frame detected by the original algorithm, and identify objects. Classification of body. The innovation of this method is as follows: Compared with the traditional artificial feature extraction method, it improves the characteristics of the target object of YOLOv3 algorithm, and has high generalization ability and stability by means of in-depth learning pre-training. Compared with hand-eye coordinated uncalibrated grabbing method for large-scale data sets [8], this method avoids using expensive equipment to collect data sets. This method is more economical, faster and conforms to the concept of AI production line.

2. TARGET DETECTION ALGORITHMS

The classical target detection algorithm (R-CNN)[9]firstly uses selective search method[10]to search for 1000~2000 candidate frames in the image, and then uses convolutional neural network (CNN) model to extract features. Each candidate box needs to be input into the model to extract features. There are a lot of overlaps in the range of thousands of candidate boxes. Repeated feature extraction generates a huge amount of computation, which makes target detection not real-time. Faster-RCNN [11]achieves fast target detection. This method improves the accuracy and speed of target detection. However, the generation and classification of candidate frames are too computational to achieve real-time detection targets.

Literature [12] proposes YOLO object detection method, which treats object detection task as a regression problem and inputs the whole picture into YOLO network. The advantage of this method is that it can detect objects quickly, avoid background errors effectively and learn generalization characteristics of objects, but its object detection accuracy is low. Dense small object detection results are poor. In order to make the target detection algorithm have the advantages of fast detection speed and high detection accuracy, YOLOv2 object detection algorithm is proposed in reference [13]. The YOLOv2 model is trained with VOC data set. The mean average precision is 76.8 and the detection speed is 67 FPS. VOC dataset trains Faster R-CNN model with mAP of 73.2 and detection speed of 7 FPS. The detection accuracy of YOLOv2 is better than Faster R-CNN, and the detection speed is faster than YOLO. Compared with YOLOv2, the backbone network of YOLOv3 ranges from darknet-19 of V2 to darknet-53, and there is also a small replaceable backbone network tiny_ darknet. As shown in Figure 1, there is no pooling layer and full connection layer in the whole YOLOv3 network structure [13].

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

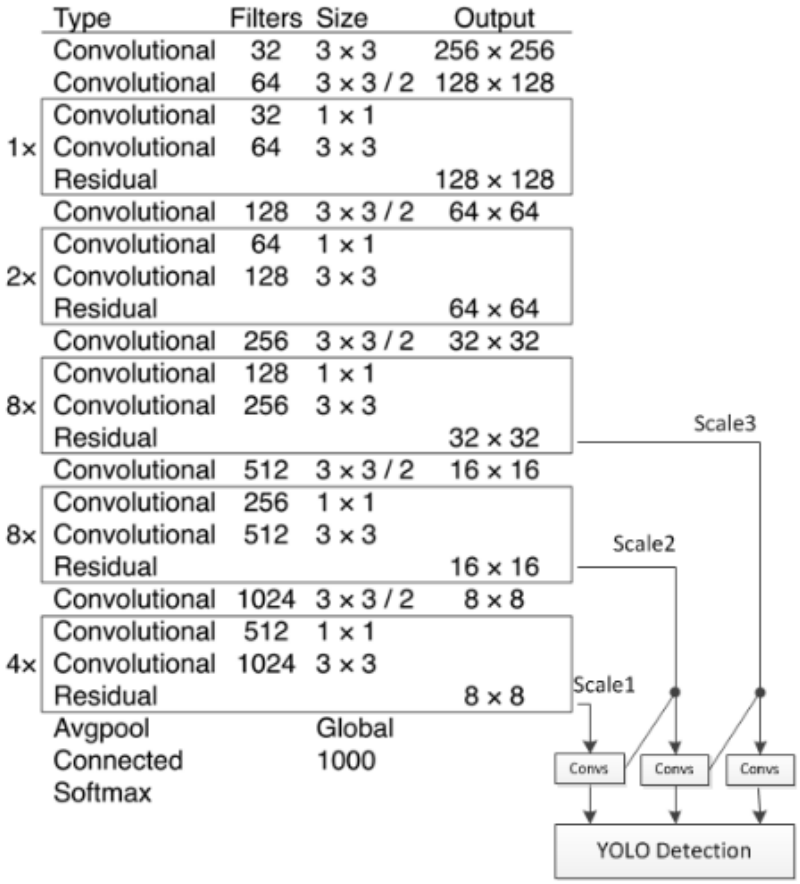

Figure 1. YOLOv3 network architecture

In the forward propagation process, the dimension transformation of tensor is realized by changing the step size of convolution core, such as stride= (2, 2), which is equivalent to reducing the image edge length by half (reducing the area to 1/4 of the original). Five times of reduction is required, and then the absolute (x, y, w, h, c) is calculated by formula (1).

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{(t_w)} \\ b_h = p_h e^{t_h} \\ \mathrm{c} = \Pr(objiect) * IOU(b, object) \end{cases} \tag{1}$$

For each bounding box network, four coordinate offsets tx, ty, tw, th are predicted. In feature map, a cell is offset relative to the upper left corner (cx, cy). The size of bounding box is pw, ph, the target center coordinates (bx, by), width bw, height bh and confidence probability C are calculated. As shown in Figure 2, the coordinates of the center point of the prediction grid and the size of the recognition frame are illustrated.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

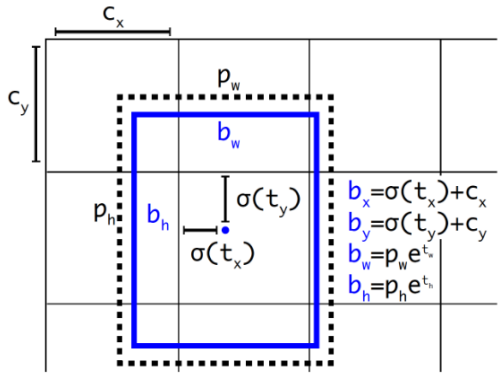

Figure 2. Predicts grid center coordinates and recognition frame size

3. IMAGE RECOGNITION SYSTEM BASED ON YOLOV3

In the automatic production process, a production line can only produce one product for a period of time, and the forming process of the product requires one or several steps. At this time, the production process of the same product is divided into several steps to operate independently according to the process requirements, but each step and process will cause a certain amount of deformation of the parts. Before each step, the product needs to check whether the product completes the previous step or not. In the process of transmission, due to the different speed of vibration, friction and transmission belt, the position and direction of the product may change in the transmission process, such as clockwise or counterclockwise rotation of an angle, left offset or right offset of a distance; therefore, a system that can quickly identify industrial images is needed. The depth image system that appears on the market is mainly implemented under Linux system, which requires high technical requirements for front-line operators and requires certain programming basis, such as Python and Linux operation instructions. In our country, the commonly used computer system is Windows, while Linux, MAC, Unix, Chrome OS and other systems are manufactured in industry. The usage rate is low. Most of the electrical components commonly used in the market, especially those connected with computers, are driven by Windows. If the existing electrical components need to be redesigned to drive on Linux system, the hidden production cost of the products will increase.

Therefore, the system designed is based on YOLOv3 algorithm, which is mounted on Win7. It can complete product identification in industrial production. In the automatic production environment, the accuracy and illumination are the difficulties of image recognition. The accuracy of industrial image recognition in conventional visual inspection is very low. In general, the target object can be distinguished from the background, and the central coordinates and boundaries of the target object can be given. When the target is detected, the feature of the target object in the target area is extracted and classified with the trained classifier, so as to complete the recognition and classification of industrial images.

Install Anaconda on win7 system and create virtual environment to install Tensorflow framework, keras and opencv. Industrial camera (GY200) is used for image acquisition. In the existing data model, training pictures are common things in life, such as birds, dogs, cattle, people, bicycles, mobile phones, cars, etc.[14]. Because of the rapid updating of industrial products, high similarity and wide range of products, there is no suitable data model for industrial products, so the corresponding training model needs to be made. VOC2007 data set structure model is used to produce data sets. The first step is to collect target images. When collecting images of products, it is necessary to take pictures of products in different postures, that is, data enhancement. The number of images collected from the same product is 30, and too many will increase training time in the process of training. The second step is to modify the name of the image. The name of the image is changed so that it can be easily found when

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

marking and modified when mistakes occur. The third step is to modify the image size to 416 * 416, and the required image size in the yolov3 algorithm is a multiple of 32. The fourth step uses LabelImg software to mark and generate XML files. In the XML files, the information of the picture will be generated, such as the position of the tag box relative to the upper left corner of the picture, the size of the tag box, and the coordinates of the center point of the tag box relative to the upper left corner of the picture. The fifth step is to train the model and finally get the weight. This step requires higher performance of the computer. At this time, other processors can be trained by linking to the network. The sixth step is to acquire the image of the detected object, acquire the real-time product image by the industrial camera (GY200) installed on the automatic production line, and detect the current state of the product through the generated weight file, and display it on the image. Fig. 3, flow chart of image recognition system based on yolov3.

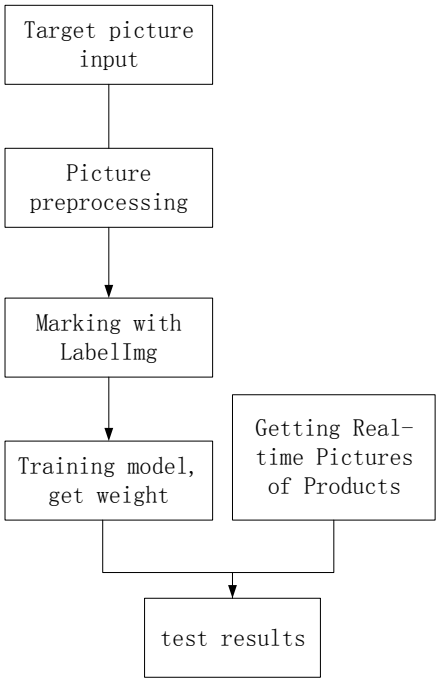

Figure 3. Flow chart of image recognition system based on yolov3

4. SYSTEM TEST AND EXPERIMENTAL RESULTS

After a series of preparatory work and framework deployment, in order to validate the industrial image recognition system based on YOLOv3 algorithm proposed in this paper, we use our own data set to test, and obtain a small number of magic cube pictures through the network. After renaming, modifying the size, training and testing as training pictures, we get the weight. Pictures are randomly extracted. As shown in Figure 4, the test part of the data set model has been marked, and the central coordinates and lengths of the target area have been saved.

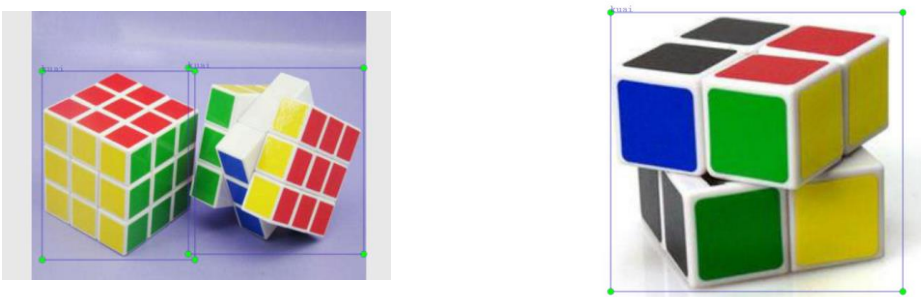

Figure 4. Magic Cube Marker Picture

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

On the automatic production line, in order to simplify the manual workload, after extracting video from industrial cameras, verify whether the magic cube is correctly recognized. Thirty tests were carried out. Test 1: The number of tests of Rubik Cube on white background was 10 times, Test 2: The number of tests of Rubik Cube on wooden desktop was 10 times, Test 3: The number of tests of Rubik Cube and bottle mixed on white background was 10 times. The test results are shown in Table 1.

Table 1. Three Scheme comparing

Numble	Number	Distinguish	Accuracy
1	10	9	90%
2	10	10	100%
3	10	8	80%

Testing 1: Under the white background, the accuracy of Rubik's cube detection is 90%, and the average verification time is 1.44s. As shown in Fig. 5-1 and Fig. 5-2, the object can be recognized when the detection surface of magic cube is color (that is, the color difference between the detection surface and the background is large); when the color of the detection surface is similar to the background color, the object can not be detected. The image recognition system based on YOLOv3 can recognize the magic cube when the magic cube rotates at a certain angle or changes its face.

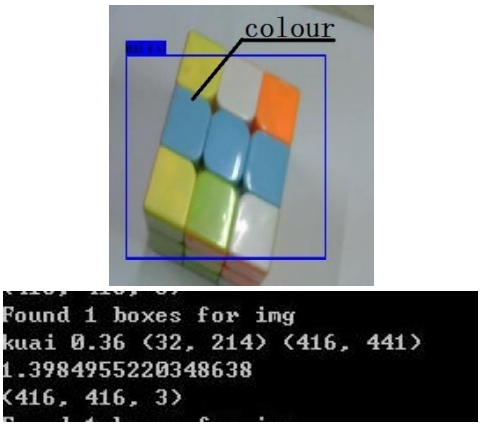

Figure 5-1. Color Detection Surface

Figure 5-2. White Detection Surface

Note:

found n boxes for img: n targets are identified and N recognition boxes appear；

Kuai m < a, b > < c, D >: It means that the target type is Kuai (only one is marked when training data set). m is the accuracy value of recognition, < a, b > is the nearest point coordinate between the recognition frame and the upper left corner of the picture, and < c, d > is the farthest point coordinate between the recognition frame and the upper left corner of the picture;

T: Recognition time;

<416, 416, 3>: Picture size 416*416, 3: RGB three-color channel.

Test 2: Under the wooden background, the accuracy of Rubik's cube detection is 100%, and the average detection time is 1.41s. As shown in Fig. 1-3, when the magic cube rotates at a certain angle, the recognition box can only recognize part of the object, and the recognition accuracy is 0.31; as shown in Figs. 6-1 and 6-2, the recognition accuracy of the object is 0.83.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

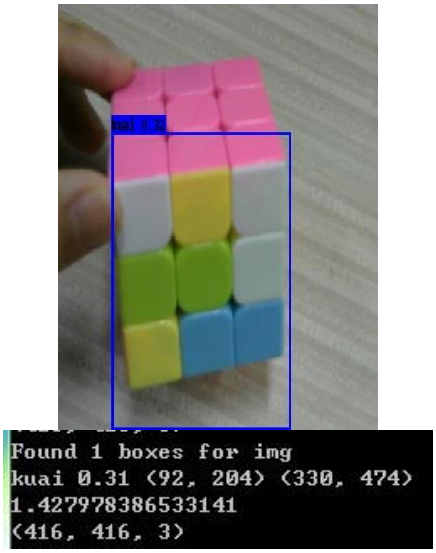

Figure 6-1. Rotation Angle Detection

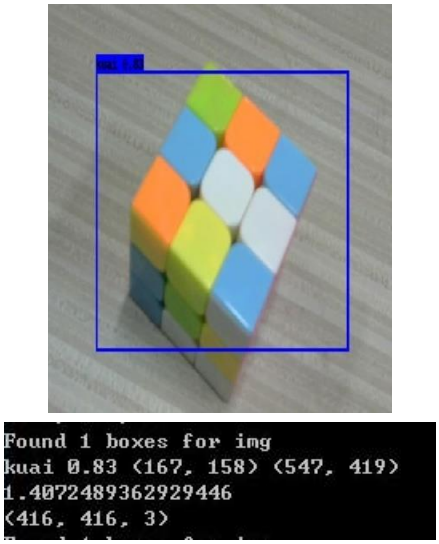

Figure 6-2. Color Surface Detection

Test 3: Under the white background, the detection accuracy of Rubik's Cube is 80%, and the average detection time is 1.44s. As shown in Figures 7-1 and 7-2, when the background object is similar to the target object's sign, there will be misrecognition. In Figure 7-2, the keyboard will be misidentified as the target object.

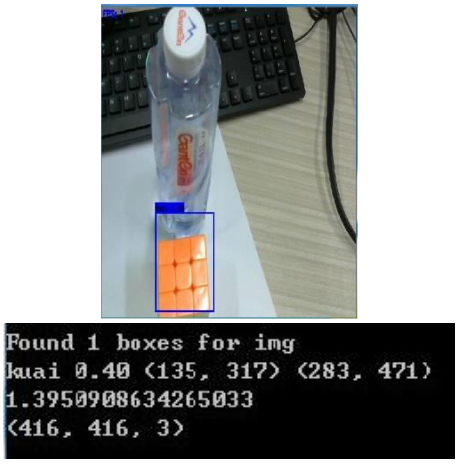

Figure 7-1. Mixed Detection with Bottles

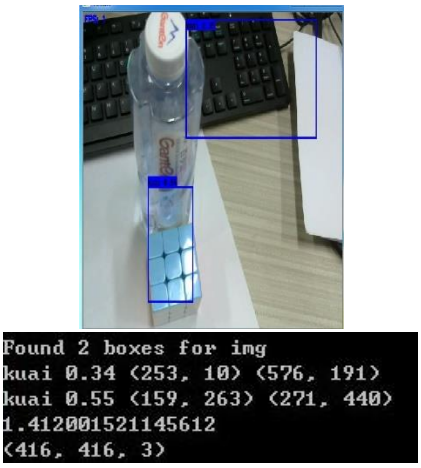

Figure 7-2. Error detection

To sum up: Test 1 and Test 2 show that the Rubik Cube has a white detection surface similar to the white background, and the recognition system can not detect the target body at this time; When Rubik Cube uses other color surfaces or mixed color surfaces to detect in the white background, the recognition system can quickly identify the target body; Rubik Cube in the wooden background; Rubik Cube in the wooden background. When detecting, the recognition system can quickly identify the target body. The visual recognition system based on YOLOv3 can recognize the object which is quite different from the background environment. For the object which is similar in color, there will be no recognition or recognition error. At the same time, in identifying magic cubes with different illumination in the same environment, because the material and color of magic cubes are relatively reflective, under strong illumination, the recognition system based on YOLOv3 can not detect magic cubes.

Test 1 compared with test 3, the background environment is white. When the magic cube and bottle are mixed together, the visual recognition system based on YOLOv3 can recognize the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

magic cube very well and can not recognize the bottle. In data set training, there is no training bottle feature, that is, there is no document in weight file that can verify bottle feature. However, in Figure 1-6, it can be seen that the keyboard in the industrial camera recognition box has been misidentified. Because the height of the collected image is too high, the gap between keyboard keys is small in the collected image of industrial camera, and the keyboard is misidentified as magic cube.

According to the analysis results of the above errors, the following improvements can be made.

(1) Optimizing the design of loss function to achieve the optimal solution in three aspects: predictive coordinates (x, y, w, h), confidence and classification. To improve the predictive target detection model, the height and width of YOLOv3 output border are reduced by k_w and k_h times respectively [15]. To reduce the boundary coincidence rate between objects. To avoid misidentification when multiple target weights are overlapped, and to improve its accuracy.

(2) In the automatic production line, the industrial camera uses constant light source to enhance the image exposure of the camera, avoiding the similarity between the detection object and the background environment, and adopts fixed height to optimize the focal length of the industrial camera, so as to improve the image recognition accuracy.

5. CONCLUSION

This paper uses YOLOv3 algorithm to realize industrial image recognition. Using tiny_yolo in YOLOv3 algorithm can greatly reduce the CPU occupation and improve the processing efficiency. The efficiency of identification is good, but the accuracy needs to be improved. In this paper, the YOLOv3 deep learning classifier based on Tensorflow framework is studied, and the target detection and recognition of magic cube pictures are carried out. The comprehensive detection and recognition rate is 90%. Windows computer can run the system in industrial production, and it can be used in a wide range of occasions. The industrial image recognition system based on YOLOv3 is applied to modern intelligent production line, which provides a theoretical basis for real-time identification of industrial products detection system.

It can be used in residential intelligent identification household monitoring probe. In the early stage of YOLOv3 algorithm, the feature extraction and partial weight allocation are improved, but the increase of feature dimension will result in the geometric increase of computation amount, which will lead to the decrease of overall recognition efficiency and increase the production cost of product automation. At the same time, image recognition network structure based on YOLOv3 realizes the detection of graph input-recognition-classification, simplifies the complex neural network of traditional workstation, and provides reliable theoretical verification for portable image recognition equipment with low power consumption.

REFERENCES

[1] Du Xuedan, Cai Yinghao, Lu Tao, Wang Shuo and Yan Zhe. A method of manipulator grasping based on deep learning [J].Robot, 2017,39(06): 820-828+837.

[2] Hosoda K , Asada M . Versatile visual servoing without knowledge of true Jacobian[C]// Intelligent Robots and Systems '94. 'Advanced Robotic Systems and the Real World', IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on. IEEE, 1994.

[3] Su J , Zhang Y , Luo Z . Online estimation of Image Jacobian Matrix for uncalibrated dynamic hand-eye coordination[J]. International Journal of Systems, Control and Communications, 2008, 1(1):31.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0006

[4] Horaud R , Dornaika F , Espiau B . Visually Guided Object Grasping[J]. IEEE Transactions on Robotics and Automation, 1998, 14(4):525-532.

[5] Huang Y , Wu R , Sun Y , et al. Vehicle Logo Recognition System Based on Convolutional Neural Networks With a Pretraining Strategy[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4):1951-1960.

[6] Wang Fujian, Zhang Jun, Lu Guoquan, et al. Vehicle Information Detection and Tracking System Based on YOLO [J]. 2018, v.31(07):92-94.

[7] Dlagnekov L , Belongie S . Recognizing cars[J]. Ecological Modelling, 2005, 113(13):71-81(11).

[8] Levine S , Pastor P , Krizhevsky A , et al. Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection[J]. International Journal of Robotics Research, 2016.

[9] Girshick R , Donahue J , Darrelland T , et al. Rich feature hierarchies for object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014.

[10] Uijlings J R R , K. E. A. van de Sande…. Selective Search for Object Recognition[J]. International Journal of Computer Vision, 2013, 104(2):154-171.

[11] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. 2015.

[12] Redmon J , Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016.

[13] Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[14] Cheng Xi, Wu Yunzhi, Zhang Youhua, et al. Image Recognition of Stored Grain Pests Based on Deep Convolution Neural Network [J]. China Agricultural Bulletin, 2018, v.34;No.472(01):160-164.

[15] Yu Yuqin, Wei Guoliang, Wang Yongxiong. Autonomous grasping method of uncalibrated 3D manipulator based on improved YOLOv2[J/OL]. Computer Applied Research.