Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 14, No. 2, 2024

Research on Visual Recognition of Service Robots Based on YOLO Algorithm

Bo Yuan, Jiajun Su, Changlong Dai, Jingjing Zha, Jialin Sun, Xinyi Hou

School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin,300222, China;

Abstract: In modern society, based on the aging of population and the rapid development of intelligent technology, the demand for intelligent service robots is increasing, and visual recognition is one of the key technologies of service robots. The rapid development of deep learning technology provides new solutions for visual recognition. Existing target recognition algorithms have the problem of complex model and slow inference speed. According to the above problems, this paper proposes a visual recognition algorithm based on YOLO series, which is used for the visual recognition design scheme of service robot. It can ensure the detection accuracy, reduce the speed of the model complexity, speed up the reasoning, and determine whether to meet the actual needs. Through randomly selected pictures, the results show that the design scheme can effectively improve the detection accuracy and real-time of the service robot on targets, and enhance its autonomous service ability in various complex environments.

Keywords: Service robot; YOLO; Visual recognition.

1. Introduction

In recent years, the service robot industry has developed rapidly, the market demand is increasingly strong, and the output has also increased substantially. In 2022, China's service robot market will reach 51.6 billion yuan. In the first half of 2023, the industry will maintain a stable growth trend, with the output reaching 3.53 million sets. It is expected that the output of Chinese service robots will reach 7.06 million sets in 2023, and the output will further increase to 7.718 million sets in 2024. Intelligent service robots are being widely used in housekeeping, medical care, elderly care, logistics, catering, hotel, fire protection and other fields. For example, the medical transport robot can automatically deliver medicine to the ward, and corridor disinfection can also be done by the corresponding robots, in some shopping malls, the cooperative robot can massage customers, in industrial production, the cooperative robot can participate in various work. With the continuous progress of technology and the continuous growth of market demand, service robots are expected to be applied and developed in more fields, bringing more convenience to people's life and work. Visual recognition is the basic function of service robots. As a common visual recognition algorithm, YOLO (You Only Look Once) algorithm has been widely used and developed in the service robot industry. Some studies are devoted to apply the YOLO algorithm in various scenarios of service robots to improve the perception and recognition ability of the robot. Fruit identification in the complex environment, for example, by improving YOLOv3 network, replace backbone and adjust the number of anchor frame, improve the identification accuracy of machine vision, can all-weather in different light environment to occlusion, adhesion and bagging a variety of cases such as fruit identification positioning, it is of great significance for picking robot in the field of agriculture. In addition, some service robot enterprises have developed YOLO algorithm-based image annotation system and complex background human behavior recognition classification algorithm to enhance the recognition ability of the robot to human behavior, and apply them to home service robots, so as to realize remote care, security patrol and other functions. To sum up, the intelligent recognition algorithm represented by YOLO algorithm has been widely used in the field of visual recognition of service robots, including improving the ability to identify small targets, reducing information loss, and enhancing real-time performance and accuracy. At the same time, it is also trying to expand its application fields, such as home furnishing, logistics, medical care, agriculture, etc. To provide support for the intelligent development of service robots. However, the algorithm may still face some challenges in practical application, such as adaptability to complex environments, limitations in computing resources, and customization requirements for different types of service robots, which are also issues that need to be further addressed in future research.

2. Fundamental model

The YOLO series algorithm is a typical algorithm in the single-stage visual recognition algorithm. YOLOv1 As the origin of this series of algorithms, it is also the first algorithm of single-stage visual recognition. Different from the sliding window and candidate region technology of two-stage visual recognition, YOLOv1 takes the whole image as input to train and detect the area based on the image globally, so it can better distinguish the foreground and background. Several versions of the YOLO visual recognition algorithm have been introduced, each of which has been improved and optimized from the previous version. YOLOv2 Mechanisms such as batch normalization (Batch Normalization) and anchor box (Anchor Boxes) are introduced to further improve the detection accuracy and speed. YOLOv3 The deeper network structure (Darknet-53) and multi-scale prediction enable the algorithm to detect targets of different sizes. Based on the basis of YOLOv3, the algorithm is updated with some new ideas in the field of CNN.

YOLOv5, the fifth version of the YOLO series, is a visual recognition algorithm based on deep learning and convolutional neural network, which has the advantages of fast reasoning, high detection accuracy and real-time detection. The author of YOLOv5 publicly released version

1.0 of YOLOv5 on May 27,2020. Based on the previous generations of YOLOv1 visual recognition to YOLOv4 visual recognition, the YOLOv5 visual recognition algorithm model has been updated on many levels. YOLOv5 improves the CSPNet and SPP modules. In terms of training, YOLOv5 uses GIOU as the loss function of the algorithm, which also improves the SPP module in the backbone network.

YOLOv5 The algorithm model mainly consists of 4 parts, including input end Input, backbone network Backbone, Neck network Neck and head network Head. First, the YOLOv5 model delivers the image to the backbone network through the input end. The input end usually accepts the image of 640x640x3 size, and then extracts the features of the input image through the backbone network. Next, the neck network integrates multi-scale features through the features extracted by the backbone network, and transmits these features to the head network. Finally, the head network detects the position and category of the target through the fusion features, and finally output the prediction results.

YOLOv5 In the initial version of the visual recognition algorithm [1], the author provided four basic pre-trained models of different sizes, including YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x (as shown in Figure 1).

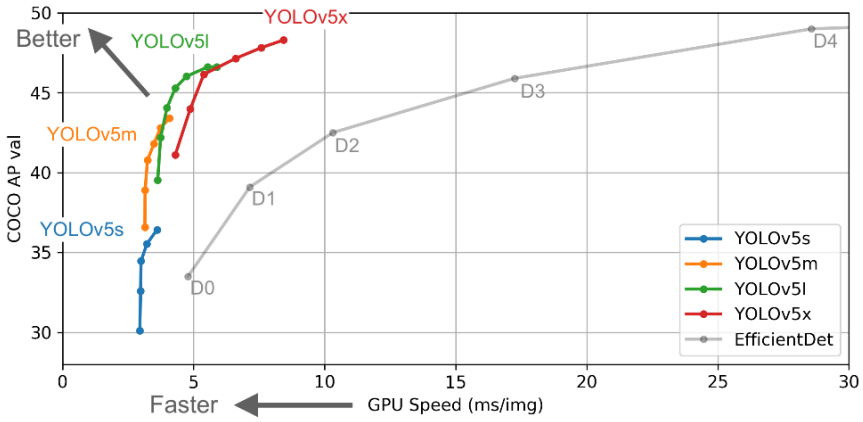

Figure 1. YOLOv5 algorithm model Fig

The average accuracy, computational speed and model size of these models vary, so it is necessary to choose a model suitable for the detection of industrial production line parts. YOLOv5s is the smallest model in the YOLOv5 algorithm series [2], it has the least amount of calculation and parameters, so its detection speed is the fastest among the four models, fast detection speed, also let YOLOv5s model can be used in high real-time occasions, because YOLOv5s fewer parameters, also lead to YOLOv5s model detection accuracy will be relatively low. On the basis of YOLOv5s model, YOLOv5m model increases model depth and model width, so that YOLOv5m model achieves a certain balance between detection accuracy and detection speed, with both certain detection accuracy and certain detection speed, which also leads to YOLOv5m model cannot reflect its characteristics in specific scenarios [3]. The parameters and computation of the YOLOv5l model are larger than those of the previous two models. More parameters and computation make the detection accuracy of the YOLOv5l model more accurate, thus losing a certain detection speed. This model is suitable for scenarios where the accuracy requirements are high and the speed requirements are not particularly strict. YOLOv5x Model is the largest model in the YOLOv5 algorithm, compared with the other three models, YOLOv5x parameters and calculation minimum, huge parameters and calculation of the most accurate, but also make it detection speed is slow, computing resource consumption is very big, YOLOv5l model is suitable for high accuracy requirements, detection speed requirement is not high [4]. Considering the complex application scenario of visual recognition algorithm and the high requirements on the model size and computational amount of the algorithm, this paper chooses to use the YOLOv5s algorithm model for experiments.

As shown in Figure 2, for the network structure diagram of the YOLOv5s algorithm model, the picture is first input and sliced [5]. Then, convolution and feature fusion are performed through the convolution module and C3 module, and then enter the neck network through the spatial pyramid pooling layer SPPF module, and finally the results are output through the head network.

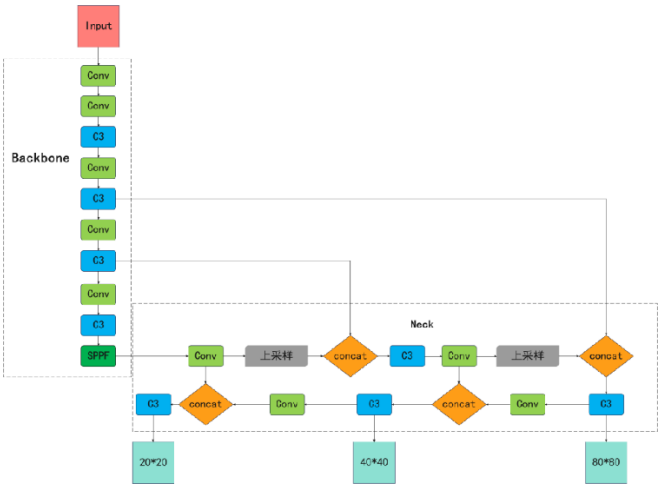

Figure 2. YOLOv5s network structure diagram

3. Experimental environment and the datasets

Visual recognition algorithm model detection speed and detection accuracy more than related to visual recognition algorithm model and parameters[6], and the visual recognition algorithm experiment environment also have relationship, different experiment environment will more or less affect the visual recognition algorithm experiment results, in order to ensure the accuracy of the experimental results, this paper in the same experiment environment, the experimental environment as shown in Table 1.

Table 1. Table of the experimental environment configuration

The parameter name	Parameter setting
operating system	Window11 The Home version is 64-bit
CPU processor	Intel(R)Core(TM)i7-9750H CPU
Number of CPU cores	12
GPU	RTX 3080 Ti
Memory	12GB
Python	3.6
Pytorch	2.3.0
CUDA	12.3

When using YOLO algorithm to train your robot on object recognition detection algorithm model, it is very important to make a data set about robot detection objects[7]. The production quality of the data set is also related to the various indicators of the trained network model algorithm. If the calibration of the data set is very well completed when making the data set, then the accuracy of the trained algorithm model will be higher. This paper selected some common objects, including airplanes, bicycles, birds, and boats and other common things, according to these common things, use cameras to take pictures of some common things in life, take and collect pictures, of planes, bicycles, birds, and boats, 1645 pictures,The 1,645 pictures were classified into the data set to be calibrated, including 321 plane pictures, 373 bicycle pictures, 465 bird pictures and 486 boat pictures, as shown in Table 2 .

Table 2. Identification quantity table of intelligent service robots

Part category	The number of pictures
airplane	321
bicycle	373
bird	465
boat	486

Install the Anaconda software on the computer. Use Anaconda software to configure the image annotation environment[8]. Install the LabelImg image annotation tool. Set the output format in the annotation tool to a format recognizable by the YOLO algorithm. Start annotating objects. In the LabelImg annotation tool, set the four items of airplane, bicycle, bird, and boat to the English labels aeroplane, bicycle, bird, and boat respectively. Annotate a total of 1645 pictures for intelligent service robot recognition. Obtain 1645 coordinate files with calibration results. Divide the annotated pictures into training set and test set in a ratio of 8:2. Separate the pictures in the data set and the calibrated documents and make them correspond one by one. Put them into the folders of training set and test set respectively. Compress the data set to complete the dataset production.

4. Experimental results

After completing the dataset production, train the robot detection algorithm [9]. Make the completed dataset into a compressed file and upload it to the experimental environment. Then use the code unzip.mydata to decompress the compressed dataset into the yolov5-7.0 folder. Modify the path of the training set in the VOC.yaml program to mydata/images/train, and modify the path of the test set to mydata/images/val. Modify the label names under the Classes program to the English names of the objects, such as aeroplane, bicycle, bird, and boat. The label names under the Classes program must be consistent with the order when the LabelImg annotation software is calibrating so that there will be no errors during model training. The network model used in this experiment is the YOLOv5s network model [10]. In the train.py program, modify the weight file weights of the model algorithm to yolov5s.yaml and modify the number of iterative training epochs to 300 rounds to ensure that the trained model can converge.

5. Evaluating indicator

When the training of YOLOv5 visual recognition algorithm is completed, it is usually necessary to analyze some model evaluation indicators to judge the quality of the evaluation model after training. The analysis of the evaluation indicators plays a crucial role in the optimization and use of the model. For the trained industrial production line part detection algorithm model, the following evaluation indexes are used to analyze the results of the model.

(1)Precision

The ratio of the number of the correct positive samples to the total number of the visual recognition is used to evaluate the accuracy of the algorithm model for the sample prediction. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP}$$

Where TP is the true example, representing the number of identified as detected objects and predicted as detected objects; FP is the number of identified as non-visually recognized objects and predicted as detected objects.

(2) Recall

It represents the ratio of the correct number of predicted positive samples to the actual number of positive samples. The recall rate can reflect the missed detection situation of the algorithm, and the high recall rate indicates that the model can cover more true positive samples. The calculation formula is as follows:

$$R = \frac{TP}{TP + FN}$$

Where FN is a false negative case, representing the number of objects identified as visually recognized object species and predicted as non-visually recognized object species.

(3) Mean average precision

The mean value of average accuracy mAP is more volatile than P and R, and it can better reflect the global performance. The larger the calculated average accuracy value, the better the training effect of the model. The mAP calculation formula is as follows:

$$mAP = \frac{\sum_{i=1}^{q} AP(i)}{q}$$

(4) Number of model parameters

The number of parameters in a model is usually composed of parameters in the model convolution layer, pooling layer and full connection layer, etc. The sum of the number of these parameters constitutes the size of the algorithm model. Algorithm model of the number of size will affect the model training speed and robustness, algorithm model of the smaller the number, the smaller the requirement of the model, the algorithm running speed, download and deployment will be more convenient, can achieve better lightweight effect, but the model of the number size will affect the effect of the model detection, as the number of parameters reduced, the

accuracy of the model may also fall.

(5) FPS

The FPS index is a measure of the speed of the model detection, indicating how many frames the model can process or how many images it detects per second. When the FPS value is higher, the more images the model can process in unit time, that is, the faster the model detection speed, the faster the algorithm model speed, the better the real-time performance of the model detection.

(6) Floating-point operations on the volume of FLOPs

The amount of floating point operations of the model mainly refers to the number of floating point operations needed in the actual operation process. It is an indirect criterion used to measure the complexity of the algorithm and the speed of the model. By analyzing the size of the floating point operation amount, we can understand the calculation amount that the model needs to carry out when processing the data, so as to evaluate the detection speed of the model.

6. Experimental results and the analysis

After training the YOLOv5 algorithm model with the completed data set, the completed data results will be obtained in the exp folder. These data intuitively show the classification accuracy and model size of the trained model. The following data will be analyzed and evaluated below.

Figure 3 shows the confusion matrix of the service robot identifying object types after the model training. The confusion matrix is a specific two-dimensional matrix. Each row in the matrix represents the real category of the calibration label, and each column in the matrix represents the category predicted by the algorithm model. The left to the bottom right diagonal in the matrix model correctly detected accuracy TP, called true positive, the higher the diagonal value, represents the accuracy of the higher, the upper right triangle area in the matrix belongs to the model of samples, also known as false positive FP, means the model classified the background or other categories into the current category, the lower left triangle area in the matrix belongs to the number of missed samples, also known as false negative FN, said the model failed to correctly detect the true samples, resulting in the target failed to be identified or misclassified into other categories. It can be seen from the figure that the correct probability of the model predicts the service robot to identify the object type. The accuracy of the bicycle is 0.95, which is the highest among the four objects. The lowest accuracy of aircraft is 0.85, the accuracy of bird is 0.92, and the accuracy of ship is 0.87. From the results of the training analysis, also illustrates the production data set, the bicycle data set of the best, calibration results is the most effective, at the end of the model training, bicycle accuracy and average accuracy will reach high value, and the aircraft data set is not very good, lead to low accuracy after model training.

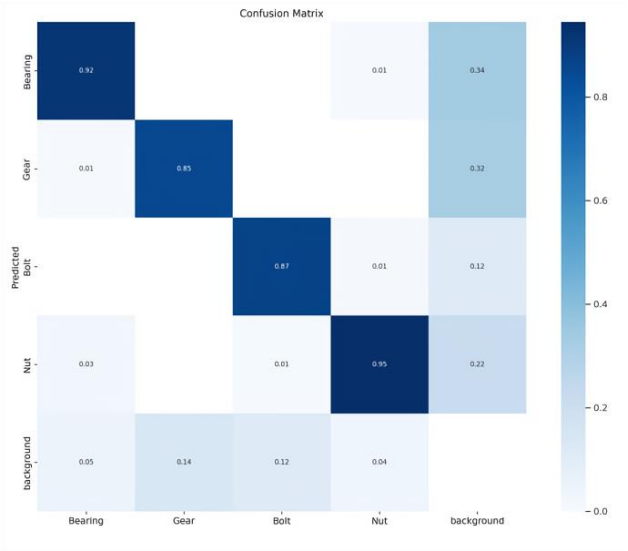

Figure 3. Confusion matrix

7. Conclusion

Under the modern social trend, the application field of service robot is more and more extensive, which is also the key technology to solve the aging of the population. With the continuous development of science and technology, service robots are becoming more and more intelligent, and deep learning visual recognition technology has also become the core of robot intelligent manufacturing. In this paper, the visual recognition algorithm based on YOLO is used for the visual recognition of the service robot, and while ensuring the accuracy of the visual recognition, the number of parameters of the model is reduced to realize the lightweight of the model algorithm. The principle and advantages of YOLO algorithm are elaborated, including its efficient visual recognition ability and real-time characteristics. The YOLO algorithm is combined with the perception system of the service robot to realize the accurate identification and positioning of the objects in the environment. A large amount of experimental data verify the effectiveness of YOLO algorithm in service robots. Research shows that the YOLO algorithm can significantly improve the perception accuracy and response speed of the robot, so that it can perform tasks more accurately in complex environments.

Although YOLO algorithm is widely used in service robots due to its powerful functions, there are still some problems and shortcomings: the detection accuracy of small targets and blocked targets is not high, and customized for different service scenarios, such as families, hospitals, shopping malls, so that it can better adapt to the characteristics and requirements of specific scenarios and identify specific items or behaviors.

Acknowledgements

College Students' Innovation and Entrepreneurship Training Program. Project number: 202310066108.

References

[1] Wang Xu, Wu Yanxia, Zhang Xue, et al. Review of rotational object detection studies under computer vision [J]. Computer Science, 2023,50 (08): 79-92.

[2] Redmon J, Divvala SGirshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]. Computer Vision & Pattern Recognition. IEEE, 2016, 779–788.

[3] Guan Jiacheng, Ren Hongwei, Zhou Songjia. Lightweight target detection based on YOLOv5 [J]. Application of computer systems, 2023,32 (09): 132-142.

[4] Li Hang. Research and Application of Image Classification Based on Convolutional Neural Network [D]. Hangzhou: Hangzhou Dianzi University, 2023.

[5] Wang Yongsheng, Ji Siyu. Summary of deep learning-based object detection algorithms [J]. Computer and Digital Engineering, 2023,51 (06): 1231-1237.

[6] Tian Manjun, Li, Kong Shihan, et al. Visual detection algorithm based on YOLOv4 (English) [J]. Information and Electronic Engineering Frontier, 2022,23 (08): 1217-1229.

[7] Chen Yimin, Li Wanyi, Weng Hanrui, et al. Review of the two-stage target detection algorithms based on deep learning [J]. Information and Computer (theoretical edition), 2023,35 (14): 112-114.

[8] Zhou Keyu, Li Jun. Progress in deep learning-based object detection research [J]. SCM and Embedded System Application, 2023,23 (07): 38-40.

[9] Qi Xulei. Single-stage object detection algorithm based on deep learning [D]. Hangzhou: Zhejiang Sci-Tech University, 2023.

[10] Zhao Dongdong, Xie Dunhan, Chen Peng, et al. Lightweight YOLOv5 sonar image object detection algorithm based on ZYNQ [J]. Opto-Engineering, 2024,51 (01): 60-72.