ITC 4/52

Information Technology and Control

Vol. 52 / No. 4 / 2023

pp. 878-886

DOI 10.5755/j01.itc.52.4.34183

Pepper Target Recognition and Detection Based on Improved YOLO v4

Received 2023/05/22 | Accepted after revision 2023/07/06

HOW TO CITE: Tan, Z., Chen, B., Sun, L., Xu, H., Zhang, K., Chen, F. (2023). Pepper Target Recognition and Detection Based on Improved YOLO v4. *Information Technology and Control*, 52(4), 878-886. https://doi.org/10.5755/j01.itc.52.4.34183

Pepper Target Recognition and Detection Based on Improved YOLO v4

Zhiyuan Tan, Bin Chen, Liying Sun

Department of Information Engineering, Hefei Vocational College of Science and Technology, Hefei, 231201, China

Huimin Xu, Kun Zhang

School of Mechanical Engineering, Zhejiang Sci-Tech University, HangZhou, 310018, China

Feng Chen

Institute of Mechanical Engineering, Anhui Science and Technology University, Chuzhou, 233100, China

Corresponding author: chenf@ahstu.edu.cn

In order to improve visual recognition accuracy of pepper and provide reliable technical support for agricultural production, an improved YOLOv4 algorithm for pepper target recognition and detection was proposed in this paper. By adding Mosaic data enhancement and CBAM (Conventional block attention module) attention mechanism to the primitive character extraction network, the method enhanced the learning ability of the target detection algorithm, made the network effectively suppress the interference features, and increased the attention to effective features. To improve the accuracy of identification. The improved network model was trained, verified and tested on the self-made data set. The results showed that the proposed algorithm could effectively improve the accuracy of pepper recognition under natural light, and finally improved the mean Average Precision (mAP) of the existing YOLOv4 algorithm from 88.95% to 98.36%.

KEYWORDS: Improved YOLOv4, Data augmentation, CBAM attention mechanism.

1. Introduction

With the acceleration of aging year by year, it is extremely urgent to propose a picking robot that can replace human labor, so as to alleviate the substantial increase in labor costs. Great progress has been made in the research of picking robot abroad, but it is still in the primary stage in China. In the field of fruit and vegetable picking, researches mainly focus on the picking and recognition of spherical fruits, while researches on the picking and recognition of irregular fruits and vegetables such as ripe peppers in natural environment are few. Accurate crop target identification and detection technology is one of the key technologies to realize agricultural automation. As one of the important cash crops, the improvement of production and quality of pepper is of great significance for ensuring domestic market supply and promoting agricultural modernization. Therefore, to carry out pepper target identification and detection research, to improve the precision and intelligent level of pepper planting, promote our agricultural modernization has important significance.

Some researchers in China have identified and detected pepper by various means, among which Ding [5] team from Tianjin University of Technology adopted the optimized convolution neural network method with deep learning mechanism to improve the recognition rate. First of all, the image is collected and binarized, the neural network modeling is carried out by Matlab, and the advantages of self-learning are used for training and testing. The accuracy of recognition rate of chili image by convolutional neural network was verified by simulation. Compared with the traditional BP neural network, this phenomenon shows that it has good generalization ability and robustness, but the recall rate of this method is low, and the recognition accuracy needs to be improved. By deepening the network depth, Yang [20] and his team from Wuhan University of Technology designed a four-layer network structure to recognize millet pepper image, which can effectively realize the recognition of millet pepper and has high recognition accuracy, but poor recognition efficiency. Li [9] from Agricultural University of Hebei collected images of dried chili, designed the background, lighting and display of the images, used line analysis method and minimum boundary rectangle method to segment images of dried chili, and adopted naive Bayes method to grade dried chili, effectively improving the accuracy of classification of dried chili. However, the complexity of this method is high, and the efficiency of this method is poor.

In view of the problems of low recognition accuracy and poor recognition efficiency of existing methods, high precision identification of pepper can only be achieved under ideal industrial environment. This paper studies a pepper target recognition and detection method based on improved YOLO v4 to solve the problem of low recognition success rate. By adding Mosaic data enhancement and CBAM (Conventional block attention module) attention mechanism to the primitive character extraction network, the method enhanced the learning ability of the target detection algorithm, made the network effectively suppress the interference features, and increased the attention to effective features. In order to improve the accuracy of identification, to complete the accurate identification of pepper target.

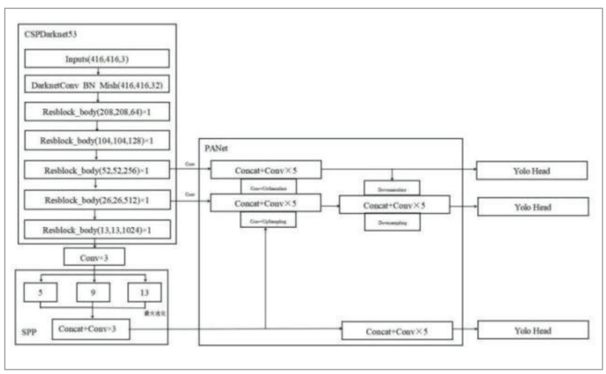
2. Yolov4 Network

YOLO network has developed rapidly in recent years, which is one of the representative target detection networks. Its main working principle is to take the detection task as a regression problem for rapid detection, which has faster detection speed than regional candidate networks and can complete end-to-end prediction [11].

YOLOv4, proposed in 2020, is an improved version of YOLOV3. It is based on the original YOLO target detection framework, and tries the advanced optimization methods in the field of deep learning in recent years, and carries out some improvements in data augmentation, backbone network, network training, activation function, loss function and other aspects [3]. YOLOv4 develops a simple and efficient model, which realizes the perfect combination of detection speed and detection accuracy. It is one of the efficient and powerful models in the existing target detection algorithms. At the same time, YOLO series belongs to an open source Python language code, which is convenient to conduct improvement on this basis [8], As shown in Figure 1.

Figure 1

YOLOv4 Network Architecture



The improvement points of its Backbone feature extraction network Backbone are as follows:

1 The main trunk feature extraction network is changed from DarkNet53 to CSPDarkNet53.
2 Activation Function: Mish activation Function is adopted.

In YOLOV3, the structure of Darknet 53 is composed of a series of residual network structures. Darknet 53 has a resblock_body module, which consists of a single down-sampling and a stack of multiple residual structures.

YOLOV4 modifies the activation function of Darknet-Conv2D from ReLU to Mish function, and transforms the convolution block from DarknetConv2D_BN_Leaky to Mish.

The Mish function has the following formula:

$$\mathrm{Mish} = x \cdot \tanh(\ln(1 + e^{x})), \tag{1}$$

where, x is the number of features and h is the convolution parameter.

3. Improved Yolov4 Network

The traditional Yolov4 network may not be able to meet the demand for accurate target recognition and detection of pepper due to the differences between pepper and other objects in form, color, size and other characteristics. Therefore, the Yolov4 network needs to be improved to improve its accuracy and efficiency in pepper target recognition and detection.

3.1. Mosaic Data Augmentation

Traditional data augmentation methods include rotation, clipping, flipping, deformation and scaling, noise addition, color disturbance, etc. In the improved YOLOv4 algorithm, Mosaic data enhancement was added after the traditional data enhancement to achieve the purpose of expanding data [23], so as to improve the Yolov4 network's ability to recognize and detect pepper targets. Mosaic is the first new data augmentation technology introduced in YOLOv4, which enables the model to learn how to recognize objects smaller than normal. Mosaic data augmentation draws lessons from CutMix data augmentation method to a certain extent. As shown in Figure 2, after a series of operations such as flipping, rotating, cropping and scaling the image, the number of images is changed to 4, and the training images are combined into one according to a certain proportion [2]. After 4-in-1, large samples or medium samples in the data set will become small samples with high probability, so Mosaic data augmentation not only expands the data set, but also increases the number of small samples, which has advantages for data sets with a small number of small samples [1, 6, 15, 22].

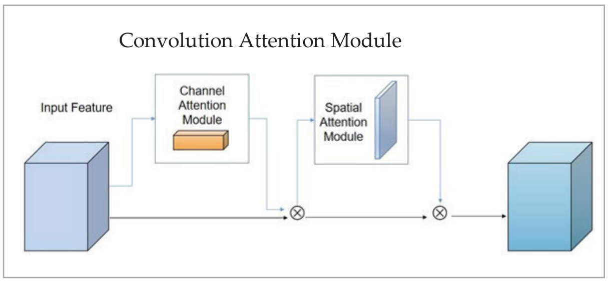
Figure 2
Schematic diagram of Mosaic data after enhancement processing

3.2. CBAM Attenti on Mechanism

Attention mechanism is a way to realize network adaptive attention. In real life, when observing things, human beings generally focus on the obvious characteristic parts of things quickly, while ignoring some irrelevant parts that affect the judgment results, so as to obtain information quickly and accurately and make accurate judgments. Attention mechanism should be added to neural algorithm because of this phenomenon, and the performance of network model will be greatly improved by adding only a few algorithms [10, 13, 21].

CBAM is a combination of channel attention mechanism and spatial attention mechanism, while SENet only focuses on channel attention mechanism, so CBAM can achieve better effects than SENet, and its adaptability is strong. It can be applied to a variety of different neural network structures, and the implementation is very simple. The ability to understand and express input features can be effectively improved, thus improving the performance of Yolov4 network [9]. The schematic diagram of its implementation is as follows. CBAM will process the channel and spatial attention mechanism for the input feature layer, respectively. This is a simple and effective attention module for feedforward convolution neural network. Given an intermediate feature graph, the attention module of the feedforward convolutional neural network will successively infer the attention graph along two independent dimensions, and then multiply the attention graph by the input feature graph for adaptive feature modification. Because CBAM is a lightweight general-purpose module, it can be seamlessly integrated into any CNN architecture and can be trained end-to-end with basic CNN [12, 14, 16-17, 19], as shown in Figure 3.

Figure 3
Convolution Attention Module



The specific implementation of channel attention mechanism and spatial attention mechanism is shown in Figure 4.

Figure 4
Channel Attention Module

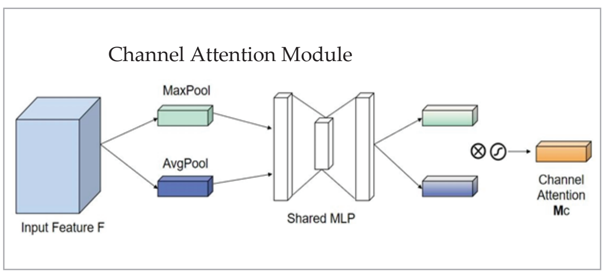


Figure 4 shows the channel attention mechanism. The implementation of the channel attention mechanism can be divided into two parts. First, the input single feature layer is pooled by global average and global maximum. After that, the results of average pooling and maximum pooling are processed by using the shared full connection layer, and then the two processed results are added, and then a sigmoid is taken to obtain the weight of each channel in the input feature layer. After obtaining this weight, multiply this weight by the original input feature layer [4]. The expression to generate the channel concern feature is shown as follows:

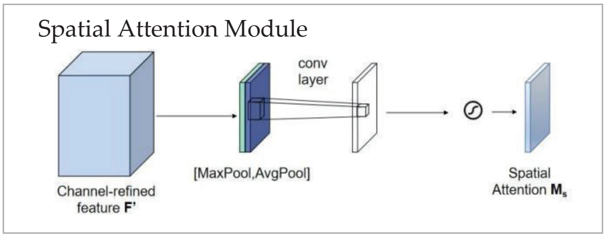
$$\begin{aligned} M_c(F) &= \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MPL}(\mathrm{MaxPool}(F))) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)\right) \end{aligned} \tag{2}$$

In the formula, σ represents Sigmoid function, AvgPool, MaxPool represents global average pooling and global maximum pooling, respectively, F represents specific stream input characteristic mapping, MPL represents multi-layer perceptron, W0 and W1 represent weights of neurons at the first layer and neurons at the second layer, respectively, and F^C_{avg}, F^C_{max} represents global average pooling and global maximum pooling specific stream input characteristic mapping, respectively.

The specific flow is that the input feature map F (H × W × C) is pooled by global maximum pooling and global average pooling based on width and height, respectively, to obtain two feature maps of 1 × 1 × C, and then they are sent to a two-layer neural network MLP, where the number of neurons in the first layer is C/r (r is the reduction rate), the activation function is Relu, and the number of neurons in the second layer is C. Then, the MLP output features are added based on element-wise operation, and then the sigmoid activation operation is performed to generate the final channel attention feature that is, $M_C(F)$. Finally, $M_C(F)$ and the input feature graph F are multiplied by element-wise operation to generate the input features needed by the spatial attention module.

Figure 5 shows the working principle of spatial attention mechanism. First, the maximum value and average value will be taken on the channels of the input feature points in the feature layer. After that, the two results are stacked, and the number of channels is adjusted by convolution with one channel number at a time, and then sigmoid is taken, at which time, the weight value of each feature point in the input feature layer (between 0 and 1) is obtained. A process of multiplying this weight by the original input feature layer after obtaining this weight [18].

Figure 5
Spatial Attention Module



The expression for the spanning space attention feature is as follows:

$$\begin{aligned} M_s(F) &= \sigma\left(f^{7\times7}\left(\left[AvgPool(F); MaxPool(F)\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right), \end{aligned} \quad (3)$$

where $f^{7\times7}$ represents convolution operation with filter size of 7×7, F^s_{avg}, F^s_{max} represent the mapping of specific stream input characteristics of spatial attention module global average pooling and global maximum pooling, respectively.

The specific flow is to take the feature map F'output by the Channel attention module as the input feature map of this module. Firstly, a channel-based global maximum pooling and global average pooling are done to get two H × W × 1 feature maps, and then these two feature maps are spliced as channels. Then, after a 7 × 7 convolution operation, the dimension is reduced to one channel, that is, H × W × 1. Then, the spatial attention feature ($M_s(F)$) is generated by sigmoid. Finally, the feature and the input feature of the module are multiplied to get the final feature.

4. Experiment and Analysis

4.1. Design the Steps and Flow of the Experiment

In the experiment, the data set was first made, then build the pytorch network framework in Python environment, build YOLOv4 model in pytorch, improve the model, and finally implement training, and analyze and evaluate the training data.

4.2. Experimental Environment

All experiments are run on the same workstation, and the specific configuration is shown in Table 1.

Table 1
Experimental environment

Name	Model/version
Python	3.7.6
CPU	Intel (R) i9-9900K CPU
GPU	NVIDIA GeForce RTX 2080
Network framework	Pytorch1.2
Industrial camera	HIKVISION MV-CA050-10GC 500
Industrial lens	HIKROBOT MVL-MF0828M-8MP

4.3. Experimental Parameters

By evaluating the performance of the workstation experimental equipment and the experimental environment, the experimental parameters are set as follows: the number of training periods (epoch) of images is 100, the first 50 times are freezing stages, the number of images sent to the network in each batch (Batch_size) is 32, and the learning rate (Learning_rate, Lr) is 0.001; The last 50 times are the thawing stage, the number of images sent to the network in each batch is 16, and the learning rate is 0.0001. In order to speed up the data reading, num_works is set to 4, and the input image resolution (Input_shape) is 416 × 416.

4.4. Preparation of Material Data Set

The production process of material data set is as follows: (1) A total of 500 images of pepper in natural light environment are collected by Hikvision industrial camera, and different pepper plants are photographed from different angles during the collection process; (2) Use PhotoScape to unify the name and size of all images; (3) labelImg image annotation tool is used to annotate images, and different materials are manually labeled to generate XML label files; Finally, the mature pepper label name Red pepper is formed, and the training set and verification set are generated according to the ratio of 9: 1.

4.5. Evaluation Indicators

Commonly used evaluation indicators in YOLOv4-tiny are as follows: Precision, Recall and F1 values are the harmonic mean values of Precision and Recall, which are equivalent to the comprehensive evaluation indexes of Precision and Recall values, AP mean values (mAP) of all categories in the multi-classification detection model and frames per second (FPS). Among them, the larger the mAP is, the higher the recognition accuracy of the model is. The above indexes have corresponding calculation formulas, and the specific formulas are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (5)$$

$$F_1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \quad (6)$$

$$\text{Total number of samples} = \text{TP} + \text{FP} + \text{TN} + \text{FN} \quad (7)$$

In the Formula (4)-(7), TP is True Positive, FP is False Positive, TN is True Negative, FN is False Negative, AP is the area under P-R curve, and P-R curve is Precision-Recall curve [9].

4.6. Pepper Recognition Experiment

In the material recognition experiment, firstly, the original YOLOv4 model recognition algorithm is used to train the data set, and the training weight is YOLOv4_weights_voc. pth, which is an open source weight file on the Internet. After the training, the weight file is generated and applied to image prediction and real-time detection. Because the shape of distant pepper is not obvious, it can be seen from Figure 6 that the fuzzy recognition rate of distant pepper is lower.

As shown in Figure 7, Epoch-Loss diagram shows the data saved after three trainings. During the training period, a weight file is generated every poch, and the graph of train loss and val loss decreasing with the increase of poch is drawn. It can be seen from the figure that with the addition of the two mechanisms, the loss descending speed and the final convergence position of the improved YOLOv4 are obviously improved during training, so it is concluded that the introduction of the two mechanisms can improve the fitting ability of the network.

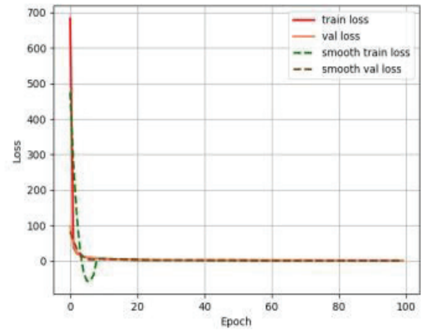
Figure 6

YOLOv4 image recognition effect

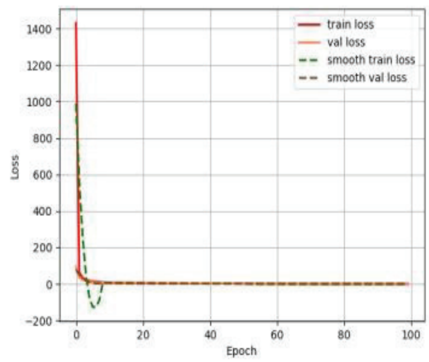
(a) Image for detection (b) Image for YOLOv4 detection

Figure 7

Epoch-Loss Diagram



(a) YOLOv4



(b) YOLOv4 + Mosaic + CBAM

After running the map_out program, the related files of running results appear. Through the statistics of the original YOLOv4-tiny model and the data obtained after adding Mosaic data augmentation and ECA attention mechanism, respectively, the statistics are shown in Tables 2-3.

Table 2

Values of four related parameters of original YOLOv4-tiny

Parameter name	Precision	Recall	F_1	AP
Parameter value	85.00%	75.00%	0.87	88.93%

Table 3
YOLOv4 + Mosaic data augmentation + CBAM attention mechanism related parameters of four materials

Parameter name	Precision	Recall	F_1	AP
Parameter value	100.00%	100.00%	0.94	98.36

From the statistical data of the above three tables, it can be seen that with the enhancement of Mosaic data and the introduction of CBAM attention mechanism, the learning intensity of the model has increased, and more attention has been paid to the key details of the image. The recognition accuracy of the improved YOLOv4 in distant fuzzy peppers has steadily improved.

$$mAP = \frac{\sum_{k=1}^{c} AP_k}{C} \quad (8)$$

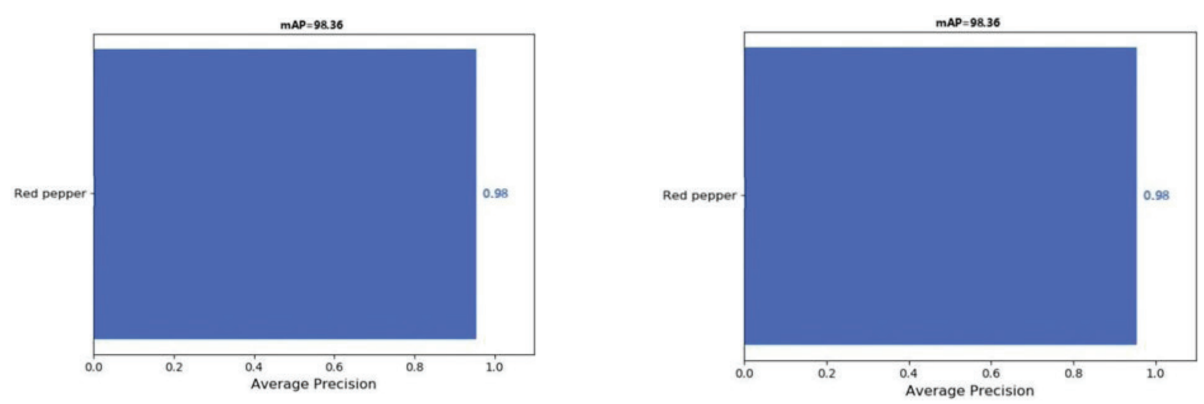
From the statistical data in Table 4 and Figure 8, it can be seen that with the enhancement of Mosaic data and the introduction of CBAM attention mechanism, the recognition accuracy of improved YOLOv4-tiny in brass and red copper is steadily improved, the comprehensive Precision value is increased by 4.66%, Recall is slightly decreased after improvement, F1 is increased by 0.01 as a whole, and AP value is increased by 3.83% after improvement. The mAP value increased by 3.84%.

Finally, as shown in Figure 9, the image recognition results in the data set show that the improved YOLOv4 has a great improvement in pepper recognition accuracy compared with the original YOLOv4, and its recognition accuracy is higher, which has a strong comprehensive recognition ability for common peppers under natural light in multi-spatial angle situations.

Table 4
Statistical table of test data

Evaluation index	Precision	Recall	F_1	AP	mAP	FPS
Before improvement	85.00%	75.00%	0.87	88.93%	88.95%	24
After improvement	100.00%	100.00%	0.94	98.36%	98.36%	30

Figure 8
Four types of materials and comprehensive mAP values



(a) Before improvement (b) After improvement

Figure 9

Improved YOLOv4 image recognition effect

(a) Image for detection (b) Image for YOLOv4 detection

5. Summary

This paper introduces an automatic identification and detection method of pepper under natural light based on improved YOLOv4. In this method, Mosaic data enhancement algorithm and CBAM attention mechanism are added to the main trunk feature extraction network of the network model. By expanding the data set and paying attention to useful features and ignoring useless features, it makes up for the problem of less pictures in the data set and makes the learning intensity of the network higher. Meanwhile, it also meets the requirements of grasping important features, and effectively improves the utilization rate of model features.

By analyzing and evaluating the experimental results with various evaluation indexes, the improved YOLOv4 has improved the recognition accuracy of pepper, and the final mAP value has increased from 88.95% to 98.36%, thus ensuring the recognition speed and accuracy on the premise of real-time. However, the network has weak ability to recognize overlapping peppers, and the overlapping peppers in the lens will make the recognition rate drop sharply. In the future, it is necessary to improve the weak recognition rate of stacked materials.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Conflicting Interests

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

Funding

This article was supported by the Key Project of Natural Science Foundation of Anhui Province (2022AH052978).

References

1. Cheng, L. Research and Implementation of Remote Sensing Image Target Detection Technology Based on Deep Learning. University of Electronic Science and Technology of China, 2022. https://doi.org/10.27005/d.cnki.gdzku.20005.1000100001006.
2. Dong, L. J., Zeng, Z. G., Yi, S. Q., Wen, Z. Q., Meng, C. Remote Sensing Image Target Detection Based on YOLOv5. Journal of Hunan University of Technology, 2022, 36(03), 44-50.
3. Huang, H. C. Study on the Maturity and Damage Identification of Fresh Pepper Based on Hyperspectral Technology. Guizhou University, 2022. https://doi.org/10.27047/d.cnki.ggudu.2022.002030.
4. Li, H. Fruit Detection and Attitude Estimation Based on RGB-D Camera. Jiangnan University, 2019.https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C475K-Om_zrgu4lQARvep2SAkOsSuGHvNoCRcTRpJSuX-uqfL8LFmdGUbDJz-QvdWYqhwen6gGyvcO8n3YVp-CxpZz0&uniplatform=NZKPT&src=copy
5. Li, L., Ding, W. K. Pepper Recognition Based on Convolutional Neural Network. Journal of Tianjin University of Technology, 2017, 33(03), 12-15.
6. Li, L., Liu, Y., Wang, W., Geng, H., Li, L.H. A Mask Wearing Detection Algorithm Based on Multi-Scale Channel Attention Mechanism. Computer Engineering and Design, 2023, 44 (02), 598-604. https://doi.org/10.16208/j.issn1000-7024.2023.02.038.
7. Li, S. Classification of Dried Peppers Based on Machine Vision. Hebei Agricultural University, 2019. https://doi.org/10.27109/D.CNKI.GHBNU.2019.2019.000377.

8. Li, S. Research on Key Technologies of Pepper Autonomous Mobile Picking Platform Based on ROS and YOLOv5. 2022, 000452. https://doi.org/10.27441/d.cnki.gyzdu.2022.000452.
9. Li, X., Pan, J. D., Xie, F. P., Zeng, J. P., Li, Q., Huang, X. J., Liu, D.W., Wang, X.S. Fast and Accurate Green Pepper Detection in Complex Backgrounds via an Improved Yolov4-Tiny Model. Computers and Electronics in Agriculture, 2021, 191. https://doi.org/10.1016/j.compag.2021.106503
10. Liu, X. Z., Yu, J., Kurihara, T., Li, K., Niu, Z., Zhan, S. Learning an Optical Filter for Green Pepper Automatic Picking in Agriculture. Computers and Electronics in Agriculture, 2021, 191. https://doi.org/10.1016/j.compag.2021.106521
11. Liu, Z. J. Improved Face Mask Detection Algorithm Based on YOLOX. Nanjing University of Posts and Telecommunications, 2022. https://doi.org/ 10.27251/d.cnki.gnjdc.2022.000194.
12. Nan, Y. L., Zhang, H. C., Zeng, Y., Zheng, J. Q., Ge, Y. F. Faster and Accurate Green Pepper Detection Using NSGA-II-Based Pruned YOLOv5l in the Field Environment. Computers and Electronics in Agriculture, 2023, 205. https://doi.org/10.1016/j.compag.2022.107563
13. Ning, Z. T., Luo, L. F., Ding, X. M., Dong, Z. Q., Yang, B. F., Cai, J. H., Chen, W. L., Lu, Q. H. Recognition of Sweet Peppers and Planning the Robotic Picking Sequence in High-Density Orchards. Computers and Electronics in Agriculture, 2022, 196. https://doi.org/10.1016/j.compag.2022.106878
14. Shang, K. G. Research on Mechanical Structure Design and Control System of Tea Picking Robot. Changchun University of Science and Technology, 2019. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C475KOm_zrgu4lQARvep2SAkEcTGK3Qt5VuzQzk0e7M1z5JjaqXN-d3dE4VGgs1HshPRoRG4B4REjFG5-35wsq9q&uniplatform=NZKPT&src=copy
15. Wang, A. Y., Wang, Y. Y. Modulation Recognition Algorithm of Multi-Attention Mechanism Network. Computer Engineering and Design, 2023, 44 (02), 328-334. https://doi.org/10.16208/j.ISSN 1000-7024.2023.02.002.
16. Wang, S. J. Study on Growth Dynamics and Water Yield Effect of Drip Irrigation under Film in Oasis. Gansu Agricultural University, 2017. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C475KOm_zrgu4lQARvep2SAk-6BvX81hrs37AaEFpExs0GuPQlEfuyw060KlKiK3uho8gJmwkAv53lEQeE5sKNIj&uniplatform=NZKPT&src=copy
17. Wei, T. Y., Liu, T. H., Zhang, S. W., Li, S., Miao, H., Liu, S. X. Identification and Location Method of Pepper Picking Robot Based on Improved YOLOv5s. Journal of Yangzhou University (Natural Science), 2023, 26(01), 61-69. DOI: 10.19411/J.1007-824.
18. Wu, H., Dong, Z. Fruit and Vegetable Recognition Based on Microcomputer and Neural Network. Journal of Ningxia University (Natural Science Edition), 2021, 42(01), 39-44
19. Wu, Y. H., Huo, W. T., Zhang, X. Q., Niu, W. Q., Liang, F., Luo, J. W., Liao, Q. Research on Pepper Picking Robot Based on Convolutional Neural Network. Industrial Control Computer, 2023, 36(01), 30-31.
20. Yang, C. Study on Defect Identification and Classification of Xiaomi Pepper Based on Deep Learning. Wuhan University of Light Industry, 2019. DOI: 10.27776/d.cnki.gwhgy.2019.000213.
21. Yang, F., Ding, Z. T., Xing, M. M., Ding, B. Overview of the Improvement of Object Detection Algorithm in Deep Learning. Computer Engineering and Application, 2023, P1-17. http://kns.cnki.net/kcms/detail/11.2127.tp.20230214.1459.040.html
22. Zhong, Y.W., Che, W.G. Ship Target Detection Based on Improved YOLOv5. Journal of Shaanxi University of Technology (Natural Science Edition), 2023, 39(01), 42-50.
23. Zhu, W.D., He, Y.S., Chen, J., Ren, W.M., Sun, Y.P. Algorithm for Marine Biological Detection Based on Improved YOLOv5. Computer and Digital Engineering, 2022, 50(08), 1631-1636.