A marigold corolla detection model based on the improved YOLOv7 lightweight

Yixuan Fan
Xinjiang University
Gulbahar Tohti
gulbahart@163.com
Xinjiang University
Mamtimin Geni
Xinjiang University
Guohui Zhang
Xinjiang University
Jiayu Yang
Xinjiang University

Research Article

Keywords: Deep learning, Object detection, Marigold, YOLOv7, Lightweight

Posted Date: November 29th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3654224/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Signal, Image and Video Processing on March 26th, 2024. See the published version at https://doi.org/10.1007/s11760-024-03107-2.

A marigold corolla detection model based on the improved YOLOv7 lightweight

Yixuan Fan[1], Gulbahar Tohti[1,2*], Mamtimin Geni[1], Guohui Zhang[1,], Jiayu Yang[1]

[1]College of Mechanical Engineering, Xinjiang University, Urumqi, 830017, Xinjiang, China.

[2] Xi'an Jiaotong University, Xi'an, 710049, Shaanxi, China.

*Corresponding author(s). E-mail(s): gulbahart@163.com;
Contributing authors: fanyixuan999@163.com; mgheni@263.net; 17863526497@163.com; 271741324@qq.com;

Abstract

Marigold, as an important traditional Chinese medicine with the functions of clearing heat and detoxification, protecting the liver, beautifying the face as well as promoting the health of the eyes, etc. Its demand is increasing day by day, and mechanized harvesting has become an inevitable trend in the industrialization of marigold harvesting. Therefore, this study tries to collect a new dataset in the south of Xinjiang, China, which is important for the production of marigold focusing on improving the YOLOv7 model by lightweighting, and proposing a set of detection models suitable for marigold. By deleting the redundant object detection layer of the YOLOv7 model, replacing the ordinary convolution of the backbone network with the DSConv convolution, replacing the SPPCSPC module with Simplified SPPF, and finally pruning and fine-tuning the model, it seeks to solve the trouble of the difficulty in deploying the mobile devices of the marigold picking robot and the inability to realize the high real-time detection. The experimental result shows that the accuracy and average accuracy mAP0.5 of the improved YOLOv7 model in marigold detection reach 93.9% and 97.7%, which are both improved compared with the original YOLOv7 model; the GFLOPs is only 2.3, which is 2.2% of the original model; and the parameter amount is 15.04M, which is 41.2% of the original model; The FPS is 166.7, which is 26.7% higher than the original model. It shows excellent accuracy and speed in detecting marigolds, providing strong technical support for marigold harvesting.

Keywords: Deep learning, Object detection, Marigold, YOLOv7, Lightweight

1 Introduction

Marigold is a persistent flowering plant with strong survivability, which can be planted in most parts of China, especially in Shache and Hotan areas of Xinjiang, China, where the marigold planting area reaches 13,400 hectares, which is the world's largest marigold planting base. Marigold has great medical value, with certain anti-inflammatory and heat-clearing effects and antibacterial detoxification and pain relief; the lutein contained in marigold can inhibit the degeneration of eyesight in the elderly and inhibit many diseases triggered by the aging of the organ-ism. It is widely used in health food, medicine, cosmetics and other products. [1]At present, marigold picking is mainly done by human labor, with a

per capita picking capacity of only 4kg/h, which makes the picking efficiency very inefficient.[2] In addition, marigold contains color hormone, which may lead to allergies of the pickers, therefore, mechanized harvesting has become an inevitable trend in the industrialization of marigold harvesting. The identification and positioning of marigold corolla become the key technology of mechanized harvesting.

In recent years, with the rapid development of deep learning and the continuous open source of algorithms such as Faster R-CNN[3] and YOLO series[4][5][6], the application of deep learning in the agricultural field has become more and more common. It is widely used on agricultural products such as citrus fruit[7], cherries[8], sugarcane[9], small wheat spikes[10], tomato[11], cotton[12], etc. Qi et al.[13] proposed a highly fused and lightweight deep learning architecture for tea chrysanthemum detection based on YOLO (TC-YOLO), using CSPDenseNet and CSPResNeXt as the backbone network, combining the recursive feature pyramid (RFP) multiscale fusion reflow structure and the Atrous Spatial Pyramid Pool (ASPP) module with cavity convolution, which can achieve an average accuracy of 92.49% on their own tea chrysanthemum dataset. Liu et al.[14] collected pear flowers from various environments, and used YOLOv5 deep learning framework for training, to complete the identification of pear flowers in different light densities and different complex environments, to improve the efficiency of pollination. .Wu et al. [15]proposed a real-time apple blossom detection method based on channel pruning YOLO v4 deep learning algorithm, constructed the YOLO v4 model under the framework of CSPDarknet53 and fine-tuned the model pruning, which simplified the detection model and had good detection performance. In order to accurately detect cucumber leaf diseases and insect pests, Saman M. Omer et al.[16] proposed an improved cucumber leaf disease and pest detection model based on the original YOLOv5l model, using the bottleneck CSP module instead of C3 as the backbone and neck network part, and combining the Convolutional Block Attention Module (CBAM) into the improved and original YOLOv5l model, and the overall performance of the improved model was better than that of the original YOLOv5 model. Qi et al.[17] in order to effectively ensure the quality and yield of crops for pest For real-time target detection of melon leaf diseases. Xu et al.[18] proposed a YOLO v5 s +Shuffle+pruned model and deployed it on Jeston Nano to verify the real-time performance of the model.Wang et al.[19] proposed a YOLOv7 model add the CBAM attention mechanism module in the front and back position of the enhanced Feature Pyramid Network (FPN) respectively, can better recognize the famous tea tender leaves.Yang et al.[20] proposed an improved algorithm based on YOLOv 7 . By introducing the MobileOne module into the YOLOv 7 backbone network, improving the SPPCSPS module and adding an auxiliary detection head. The improved model can significantly improve the recognition rate of fruit targets in high-density fruits.

The above research has made a great breakthrough in the field of agriculture, but there are very few ways to identify and pick marigolds. In view of the problems of slow detection speed, large number of parameters and large amount of calculation in the identification of marigold in the existing model, which is not conducive to the deployment of field manipulator operation, this paper proposes a method to lightweight marigold detection model.

2 YOLOv7

YOLOv7 [21]is an object detection and classification algorithm proposed in 2022, which employs a single network model, thus avoiding the cumbersome preprocessing steps in traditional two-stage target detection algorithms.YOLOv7 integrates a model reparameterization, a label assignment strategy, the ELAN efficient network architecture[22], and the auxiliary detection head training, effectively balancing detection speed and detection accuracy.YOLOv7 is mainly composed of 3 parts: Input, Backbone and Head.

The Input side uses adaptive image deflation and hybrid data enhancement techniques, and adopts the adaptive anchor frame calculation method as traditional YOLO, which solves the problem of varying sizes of input images and provides suitable pre-selected frames.

Backbone is responsible for feature extraction of the input image and final output of three feature layers of different sizes. It consists of three parts: CBS (Conv+BatchNorm+SiLU), ELAN and MPConv. Firstly, the feature of input image is

extracted by the CBS module , and three feature layers of different sizes are obtained. Secondly, These feature layers are subjected to feature fusion through the ELAN module to obtain enhanced feature layers. Thirdly, these enhanced feature layers are further extracted and fused through the MPConv module. Finally, output three enhanced and effective feature layers. These feature layers can be used in subsequent classifiers and regressors for target detection.

The Head part mainly consists of the SPPCSPC module, ELAN-W module, RepConv module, etc. The SPPCSPC consists of multiple convolution kernels of different sizes and SPP modules, which convolves and pools the input feature maps at different spatial locations in order to extract and enhance the feature information.The ELAN-W module is an extension of the ELAN module used in Backbone, ELAN-W is designed to provide efficient gradient propagation paths to improve the performance of the model.The RepConv module introduces reparameterized volumes and is more flexible, which will effectively reduce the computational complexity and the number of parameters of the model and improve the speed of operation.

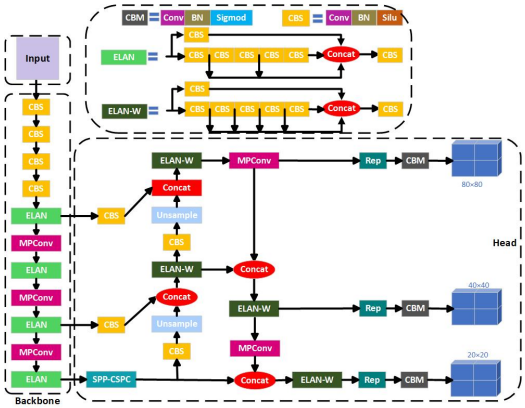

Fig. 1 Architecture of YOLOv7

3 Lightweight marigold detection model

Based on the improvement of the YOLOv7 model, the lightweight marigold detection model firstly replaces the ordinary convolution module with Distributed Shifting Convolution (DSConv)[23] in Backbone. Secondly, deleted the redundant layer of the Head part that is unfavorable to marigold detection. Thirdly, the SPPCSPC module is replaced with Simplified SPPF[24]. Finally the improved model is pruned and retrained at one time to complete the model improvement.

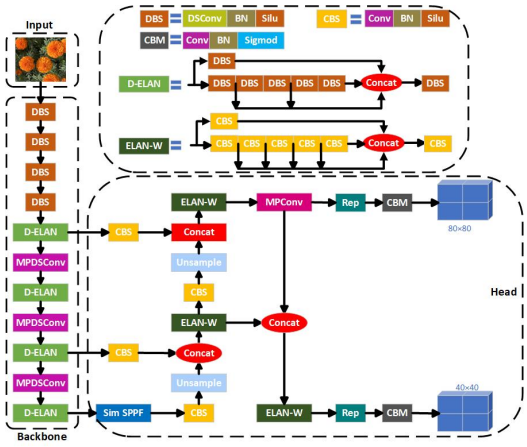

Fig. 2 Lightweight marigold detection mode

3.1 The backbone network is lightweight

In order to reduce the size of the model and the complexity of the calculation, the model can adapt to the resource-constrained edge devices, the Distribution Shifting Convolution is used to replace the ordinary convolution in the backbone network, so as to achieve the effect of model lightweight.

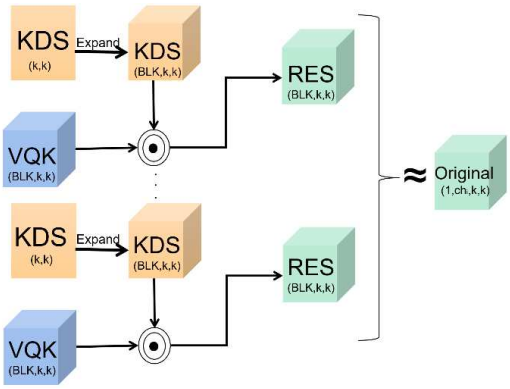

Fig. 3 The overall idea of DSConv

DSConv [23](Distribution Shifting Convolution) is an alternative convolution algorithm that can replace various convolutional neural networks for accelerated inference and training in a "plug-and-play" manner. DSConv replaces other common convolutions can achieve lower memory usage and faster computation. DSConv decomposes the traditional convolutional kernel into Variable Quantized Kernel (VQK) and Distribution Shifts, in which the Variable Quantized Kernel(VQK) only retains the variable long integer value and is the same size as the initial convolution kernel. VQK acts as a priori in the structure to capture the features extracted by a specific module. The Distribution Shifts is to move the distribution of VQK, and the output tensor matches the initial weight tensor through the Kernel Distribution Shifter (KDS) and the Channel Distribution Shifter (CDS) . The distribution shifts refers to the scaling and offset operations on the features. The structure diagram is shown in Figure 3, where the magnitude of the original convolutional tensor is (ch_o ,ch_i , k , k). ch_o is the number of channels in the next layer, ch_i is the channels in the current layer, k is the width and height of the kernel, and BLK is the block-size hyperparameter.

3.2 The head network is lightweight

3.2.1 The head structure is lightweight

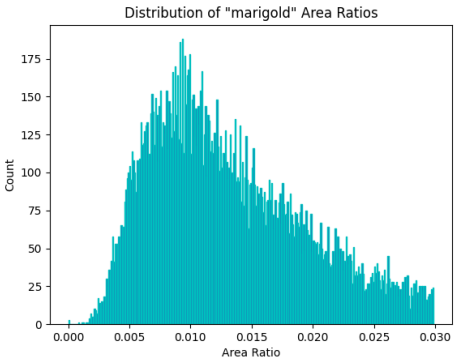

Fig. 4 Distribution of marigold area ratio

After the features of original image for the YOLOv7 model is extracted by Backbone, the feature maps of 20×20, 40×40, 80×80 dimensions will be obtained.In the YOLOv7 model, the smaller feature map size, is responsible for detecting larger targets. Specifically, if the feature map is more small, it means the more vague the details considered. Therefore, this is not good for the detection of small and medium-sized targets such as marigolds. According to Figure 4, the size of marigolds is relatively single, and there is no need for 20×20 detection layers for detection. The 20×20 detection layer can be discarded to reduce the number of parameters and computational complexity of the model, and improve the performance of the model detection algorithm.

3.2.2 Spatial pyramid pooling improvements

The Simplified SPPF module is used to replace the SPPCSPC module of the original network, which increases the receptive field and further reduces the number of parameters and calculations.

Simplified spatial pyramid pooling layer (SPPF) is a simplified version of the spatial pyramid pooling layer (SPPF). Spatial Pyramid Pooling Layer (SPPF) is a pooling layer for convolutional neural networks, which can pool the input feature map at different spatial scales to capture richer spatial information. However, the implementation of SPPF is complex and requires multiple convolution and pooling operations.

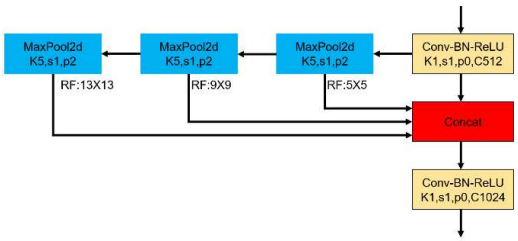

Fig. 5 Structure of simplified SPPF

Simplified SPPF simplifies SPPF and achieves similar functionality with a convolution and a pooling operation. Specifically, Simplified SPPF first performs a convolution operation on the input feature map, and then divides the convoluted feature map into two parts. A portion of the group performs 3 maximal pooling operations to obtain a low-resolution feature map. The other part is directly linked to after 3 maximum pooling operations. The obtained feature map contains the spatial information of the original feature map,

and has fewer parameters and faster operation speed.

3.3 Model pruning

Model pruning can be divided into training, pruning, and fine-tuning. Training refers to training a large model with the goal of optimal network performance and pruning it as a benchmark model. Pruning is the process of cutting a large model according to a specific pruning method to obtain a streamlined network structure. Fine-tuning refers to making adjustments to a cropped model to regain lost performance.

In this paper, a one-time pruning and retraining pruning method was adopted. [25] The pruning rate of the convolutional layer and the BN layer is set to 0.8 for pruning. In this method, the weight value is used to judge the importance of the filter, and the L1 norm is used as the evaluation index of the filter. By sorting the results, the filters with low contribution to the network in the first layer are cut out, which reduces the computational cost of CNN. Compared to other metrics, this method does not produce a sparse join pattern. It is possible to prune convolution kernels that are less useful to trained models without a loss of accuracy as much as possible.

4 Experimental data sets and evaluation indicators

4.1 Dataset production

The marigold dataset was collected from Yache County, Xinjiang Uygur Autonomous Region (geographical coordinates 37°27' 39°15'N, 76°10' 77°40'E). The collection of data is from July 16 to July 18, 2023. The shooting time is early morning, noon, and evening in order to obtain data at different times and under different lighting conditions.The dataset consists of pictures which were taken by a mobile phone as a device whose shooting angle is vertical downward overhead. The shooting distance is 20-50cm, and the images under different conditions are collected, including forward light, backlight, close distance, long distance and other situations. A total of 3320 marigold images were collected, and 3032 images were selected for dataset. These These filtered images are divided into two parts, they are respectively training set (2672 images) and test set (359 images) at a 9:1 ratio.

4.2 Dataset labeling

In this paper, Labelimg annotation software is used to manually label the marigold dataset, and the labeling box is for selecting the minimum external rectangle of marigold corolla. The annotation information file generated by labeling is an xml file, which stores the information such as the file name of the marigold image, the position information of the four corners of the rectangular box in the annotation area, and the annotation type. Since the annotation file required for YOLO training is a txt file, python is used to convert xml into a txt file to obtain a dataset that YOLO can run.

4.3 Experimental environment and parameter configuration

The experimental configuration is as follows: Win10 operating system. Graphics card model is Nvidia GeForce RTX 4060ti 8 GB. Processor model is 12th Gen Intel(R) Core (TM) i5-12490F CPU. The deep learning framework uses PyTorch1.8. The programming platform is PyCharm. All comparison algorithms run in the same environment.

4.4 Experimental evaluation indicators and experimental analysis

Combined with the actual situation of this experiment, P, R, mAP@0.5 and mAP@0.5:0.95, parameter quantity, GFLOPs and Frames Per Second (FPS) were selected as the evaluation criteria.

P is the precision, R is the recall, mAP is the mean average accuracy, mAP@0.5 Indicates an mAP at the intersection greater than the Union (IoU) threshold of 0.5. mAP@0.5:0.95 Represents the average accuracy over different loU threshold ranges (from 0.5 to 0.95 in 0.05 steps). mAP refers to the average AP of the model across all categories. FLOPs (Floating Point Operations). FPS(Frames Per Second) .

The AP equation (Eq.1) and mAP equation (Eq.2) are as follows:

$$AP = \int_0^1 P(R)dR \tag{1}$$

Table 1 Comparative experiments of different detection layers

	Model	FLOPs (G)	Parameters (M)	map@0.5 (%)	map@0.5:.95 (%)	FPS (img/s)
(1)	Yolov7	103.2	36.48	97.3	75.6	131.6
(2)	Yolov7-p6	106.6	53.55	96.8	74	128.2
(3)	Yolov7-p2	116.8	37.02	96.4	73.9	117.7
(4)	Yolov7-p34	94.6	25.97	97.8	76.1	149.3

$$mAP = \frac{\sum_{i=0}^{N} AP_i}{n} \tag{2}$$

5 Experimental results and analysis

5.1 Detect the influence of layers on the model

In order to verify that the 20×20 target detection layer proposed in this study is not significantly helpful for detecting marigolds, and the effectiveness of other target detection layers for marigold recognition, a series of verification experiments were carried out. The specific results are shown in Table 1. In Table 1, experiment (1) is the original model, (2) is the model with P6 (10×10) detection layer, (3) is the model with (160×160) detection layer, and (4) is the model with the 20×20 object detection layer removed. It can be seen from the table that the map@0.5 of model (2) and model (3) were 0.968 and 0.964, respectively, which were 0.5 percentage points and 0.9 percentage points lower than that of the original model.GFLOPs increased by 3.4 and 13.6, respectively, and FPS was also reduced by 3.4 and 13.9 compared to the original model. It can be seen from this that the addition of a small target detection layer and a large target detection layer does not improve the detection of marigolds.

The map@0.5 of model (4) can reach 97.8%, which is 0.5 percentage points higher than the original model. The map@0.5:.95 can reach 76.1, which is 0.5 percentage points higher than the original model. The GFLOPs are reduced by 8.3%, the number of parameters is reduced by 28.8%, and the FPS is increased by 17.7 and 13.5% compared with the original model.

5.2 The effect of different convolutions on the model

In order to analyze the performance of different convolution modules, the ordinary convolutions of model (4) Backbone determined in Section 5.1 were replaced with GhostConv, PConv, and DSConv.Table 2 shows the performance of the above stem convolution after replacing it.

Table 2 shows that the replacement of Backbone's normal convolution with GhostConv, PConv, and DSConv, map@0.5 are 97.7%, 97.8%, and 97.7%, respectively, with GFLOPs of 63.2, 74.3, and 29.0, respectively, and FPS of 120.5, 142.9, and 142.9, respectively. The purpose of replacing the backbone network convolution in this paper is to reduce the amount of computation and parameters without reducing other performance as much as possible, so as to achieve the effect of lightweight. Although DSConv is slightly lower than PConv in map@0.5 and map@0.5:.95, the GFLOPs of DSConv convolution are only 29, which is much lower than that of GhostConv and PConv, and the FPS is higher than that of GhostConv. Taking these factors into account, DSConv has better overall performance.

5.3 Ablation experimentsl

In order to further verify the optimization effect of the improved YOLOv7 in the marigold recognition and detection model, the improved YOLOv7 algorithm was compared with the initial algorithm step by step. The ablation experiment consists of 4 steps of improved comparison, the first step is to remove the redundant object detection layer, the second step is to replace the normal convolution in the backbone with the distributed shift convolution DSConv, the third step is to replace the SPPCSPC module in the neck with SimSPPF, and the fourth step is to prune the CNN model

Table 2 Trunk convolution substitution comparison experiments

Model	FLOPs (G)	Parameters (M)	map@0.5 (%)	map@0.5:.95 (%)	FPS (img/s)
GhostConv	63.2	19.42	97.7	0.753	120.5
PConv	74.3	21.46	97.8	0.762	142.9
DSConv	29.0	25.98	97.7	0.76	142.9

Table 3 Ablation experiments of lightweight model

	Model	FLOPs (G)	Parameters (M)	map@0.5 (%)	map@0.5:.95 (%)	FPS (img/s)
(1)	YOLOv7	103.2	36.48	97.3	75.6	131.6
(2)	YOLOv7-p34	94.6	25.97	97.8	76.1	149.3
(3)	YOLOv7-p34-DSConv	29.0	25.98	97.7	76	142.9
(4)	YOLOv7-p34-DSConv-SimSPPF	24.2	19.95	97.6	76	153.9
(5)	YOLOv7-p34-DSConv-SimSPPF-L2	2.3	15.04	97.6	75.9	153.9
(6)	YOLOv7-p34-DSConv-SimSPPF-L1	2.3	15.04	97.7	76.3	166.7

using the L1 norm for the model in step 3 and compare it with the model using the L2 norm.

According to the ablation experiments in Table 3, the replacement of the SPPCSPC module with SimSPPF can improve the detection speed and effectively reduce the number of GFLOPs and parameters by 16.6% and 23.2%, respectively, after replacing the SPPCSPC module with SimSPPF (3). FPS increased by 7.6%. Models (5) and (6) are the models obtained by pruning and fine-tuning using the L2 norm and L1 norm, respectively, and have advantages in map@0.5, map@0.5:.95 and FPS compared with model (5).

Based on the results of ablation experiments, a lightweight marigold target detection model based on the YOLOv7 model was carried out by using an improved method of L1 norm pruning in P3P4 detection layer + DSConv + SimSPPF. Compared with the YOLOv7 model, the improved YOLOv7 model has improved performance in every way. Compared with the YOLOv7 model, its map@0.5 is increased by 0.4 percentage points, map@0.5:.95 is increased by 0.7 percentage points, GFLOPs are reduced by 97.8%, parameters are reduced by 58.8%, FPS is increased by 35.1, and the growth rate is 26.7%.

Table 4 Comparison of the performance of different detection models

Model	map@0.5 (%)	P (%)	R (%)	FLOPs (G)
Yolov3	96.3	92.2	90.1	282.2
Yolov5	96.4	91.8	89.2	7.1
Yolov6	96.4	90.4	89.2	11.8
Yolov7	97.3	93.5	92.1	103.2
Yolov8	96.9	91	90.9	8.1
Ours	97.7	93.9	91.7	2.3

5.4 Comparison of the performance of different models

In order to compare the improved object detection model with the current mainstream YOLO series object detection model, the performance of different models was analyzed, and the superiority of the improved model in this study was explored. According to the comparative experiments in Table 4, The improved YOLO v7 detection model has the best precision, recall, and map@0.5 indicators , which are 93.9%, 91.7% and 97.7%, respectively.Compared with Yolov3, Yolov5, Yolov6, Yolov7 and Yolov8, the mean average accuracy of map@0.5 increased by 1.4, 1.3, 1.3, 0.4 and 0.8 percentage points, and the GFLOPs decreased by 279.9, 4.8, 9.5, 100.9 and 5.8, respectively.

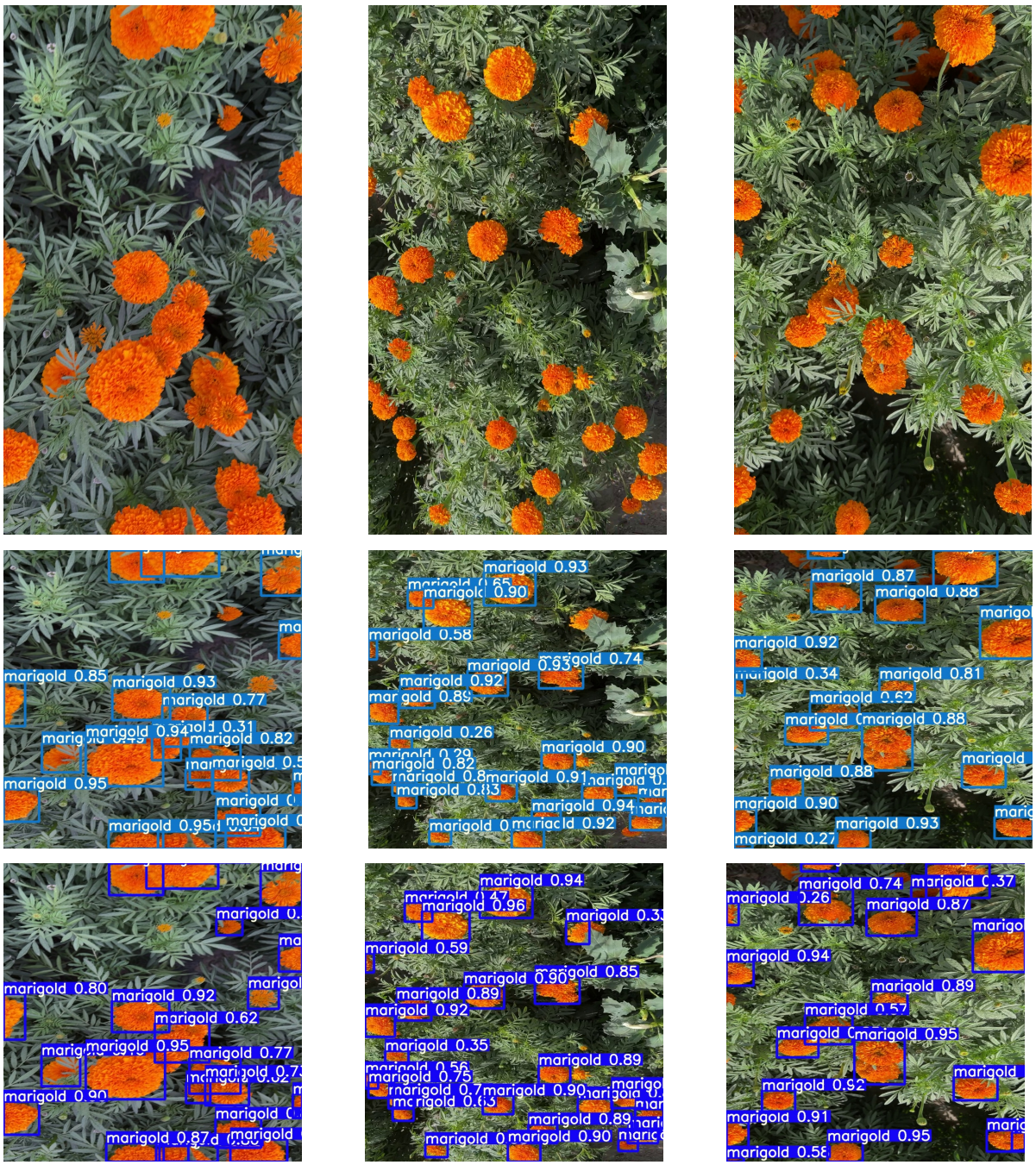

Fig. 6 Detection results of different models. The first row is the raw data, the second row is the image after YOLOv7 detection, and the third row is the image after Ours-inspection

5.5 The detection effect of YOLOv7 has been improved

Figure 6 shows the detection effect of the improved YOLOv7 model, the left figure shows the detection effect of the YOLOv7 model, and the right figure shows the model after the improvement of YOLOv7. Set the detection confidence threshold to 0.25 and the IOU threshold to 0.45. As can be seen from Figure 6, the improved YOLOv7 detection model will have a better detection effect on marigolds at the edge of the image, and can detect

objects that cannot be detected by YOLOv7 at the edge of the image, and the detection accuracy is higher.

6 Conclusion

In this paper, a dataset of marigold in different lighting environments is constructed for the recognition of marigold picking robots, and a lightweight marigold recognition method combining YOLOv7 and channel pruning algorithm is proposed. Firstly, the redundant detection layer of YOLOv7 is deleted to reduce the amount of computation and parameters. Secondly, the ordinary convolution in the Backbone is replaced by the Distribution Shifting Convolution to reduce the computational burden of the algorithm. Thirdly, the Simplified SPPF module is used to replace the SPPCSPC module of the original network, which reduces the amount of calculation and parameters of the model, and improves the detection speed of the model.Finally, through pruning and fine-tuning, the accuracy and speed of model detection were improved. The precision and mAP0.5 of the improved YOLOv7 model in marigold detection reached 93.9% and 97.7%, which were 0.4 percentage points higher than that of the original YOLOv7 model. The computational amount was only 2.3GFLOPs, which was 2.2% of the original model. The parameter amount was 15.04M, which was 41.2% of the original model. And The FPS is 166.7, which is 26.7% higher than the original model. The improved lightweight model has strong real-time detection ability and detection accuracy, which provides a reference for the real-time detection and high-precision demand detection of marigolds, and provides a reference for the design of marigold picking robots in the next step.

Declarations

Conflict of interest The authors declare that they have no competing interests.
Ethical Approval Not applicable.
Funding The National Natural Science Foundation of China (Project No. 12162031), State Key Laboratory for Manufacturing Systems Engineering of Xi'an Jiaotong University (Project No. sklms2022022).
Availability of data and materials Not applicable.

References

1. Li N, Wang P, Wu Z, et al (2010) Research status and development prospects on marigold. Northern Horticulture (10):228–231(in Chinese).

2. Zhengdong Q, Guodang L, Liangliang L, et al (2020) Present situation and suggestions on mechanical picking technology of marigold. Journal of Real-Time Image Processing 10(09):26–30

3. Ren S, He K, Girshick R, et al (2017) Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6):1137–1149. https://doi.org/{10.1109/TPAMI.2016.2577031}

4. Redmon J, Divvala S, Girshick R, et al (2016) You only look once: Unified, real-time object detection. IEEE Computer Society, Los Alamitos, CA, USA, pp 779–788, https://doi.org/{10.1109/CVPR.2016.91}

5. Redmon J, Farhadi A (2017) Yolo9000: Better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 6517–6525, https://doi.org/{10.1109/CVPR.2017.690}

6. Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv e-prints

7. Li C, Ma W, Liu F, et al (2023) Recognition of citrus fruit and planning the robotic picking sequence in orchards. Signal, Image and Video Processing 17(8):4425–4434

8. Gai R, Chen N, Yuan H (2021) A detection algorithm for cherry fruits based on the improved yolo-v4 model. Neural Computing and Applications (5)

9. Yu K, Tang G, Chen W, et al (2023) Mobilenet-yolo v5s: An improved lightweight method for real-time detection of sugarcane stem nodes in complex natural environments. IEEE Access 11:104070–104083. https://doi.org/{10.1109/ACCESS.2023.3317951}

10. Liu L, Li P (2023) An improved yolov5-based algorithm for small wheat spikes detection. Signal, Image and Video Processing 17(8):4485–4493

11. Chen J, Wang Z, Wu J, et al (2021) An improved yolov3 based on dual path network for cherry tomatoes detection. Journal of Food Process Engineering

12. Susa JAB, Nombrefia WC, Abustan AS, et al (2022) Deep learning technique detection for cotton and leaf classification using the yolo algorithm. In: 2022 International Conference on Smart Information Systems and Technologies (SIST), pp 1–6, https://doi.org/{10.1109/SIST54437.2022.9945757}

13. Qi C, Gao J, Pearson S, et al (2022) Tea chrysanthemum detection under unstructured environments using the tc-yolo model. Expert Systems with Applications 193:116473. https://doi.org/{https://doi.org/10.1016/j.eswa.2021.116473}

14. Liu Q, Wang S, He X, et al (2023) Pear flower recognition based on yolo v5s target detection model in complex orchard scenes pp 5961–5970

15. Wu D, Lv S, Jiang M, et al (2020) Using channel pruning-based yolo v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Computers and Electronics in Agriculture 178:105742. https://doi.org/{https://doi.org/10.1016/j.compag.2020.105742}

16. Omer SM, Ghafoor KZ, Askar SK (2023) Lightweight improved yolov5 model for cucumber leaf disease and pest detection based on deep learning. Signal, Image and Video Processing pp 1–14

17. Qi F, Wang Y, Tang Z, et al (2023) Real-time and effective detection of agricultural pest using an improved yolov5 network. Journal of Real-Time Image Processing 20(2):33

18. Xu Y, Chen Q, Kong S, et al (2022) Real-time object detection method of melon leaf diseases under complex background in greenhouse. Journal of Real-Time Image Processing 19(5):985–995

19. Wang Y, Xiao M, Wang S, et al (2023) Detection of famous tea buds based on improved yolov7 network. Agriculture 13(6):1190

20. Yang H, Liu Y, Wang S, et al (2023) Improved apple fruit target recognition method based on yolov7 model. Agriculture 13(7):1278

21. Wang CY, Bochkovskiy A, Liao HYM (2023) Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 7464–7475

22. Wang CY, Liao HYM, Yeh IH (2022) Designing network design strategies through gradient path analysis. arXiv preprint arXiv:221104800

23. Nascimento MGd, Fawcett R, Prisacariu VA (2019) Dsconv: Efficient convolution operator. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 5148–5157

24. Li C, Li L, Jiang H, et al (2022) Yolov6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:220902976

25. Li H, Kadav A, Durdanovic I, et al (2016) Pruning filters for efficient convnets