



Study on Feedback and Correction of Tomato Picking Localization Information

Shuangyou Wang^{1,2}, Guohua Gao^{1*}, Ciyin Shuai¹

¹ Faculty of Materials and Manufacturing, Beijing University of Technology, Beijing 100124, China

² School of Software, Handan University, Handan 056005, China

Corresponding Author Email: gaoguohua@bjut.edu.cn

<https://doi.org/10.18280/ts.400107>

Received: 15 October 2022

Accepted: 2 January 2023

Keywords:

tomato detection, ShuffleNetV2, visual feedback, localization correction

ABSTRACT

In the process of picking tomatoes, due to the mechanical error caused by the mechanical arm, the tomato positions cannot be detected accurately, and the information feedback of the positioning is not available, affecting the picking efficiency. Therefore, this article proposed visual feedback information and correction, and designed an improved yolov5s model lightweight detection method, whose backbone network was replaced with lightweight ShuffleNetV2. In addition, the Bidirectional Feature Pyramid Network (BiFPN) was added to obtain richer feature information. Experimental results showed that the improved model achieved 97.4 percent mAP, 97.5 percent accuracy and 1.89 MB model size, with inference time of 4.8 ms per image. This detection method quickly calculated the Euclidean distance between the reference point and the target tomato. The target tomato, with the Euclidean distance less than 58.12 mm, was picked successfully, while the one, with the Euclidean distance greater than 58.12 mm, was not picked. Then the error needs to be calculated and fed back to the robot for picking again. The whole process realized information feedback and correction and improved the picking efficiency with less feedback time.

1. INTRODUCTION

China is the country with the largest cultivation area and the largest total production of tomato in the world. At present, tomato picking mainly depends on manual picking, which has high labor intensity, high cost and low degree of automation. Therefore, the research and development of tomato picking robot is of great significance to picking operation. However, the difficulty of the current problem lies in how to improve the recognition and localization precision of tomato picking robot [1].

There are two kinds of localization errors of picking robot, one is the localization error of tomato recognition and detection, and the other is the localization error of mechanical movement. They determine the accuracy and efficiency of the picking robot. In practical application, the phenomenon that the tomato can not be picked successfully often appears, which is mainly caused by the localization error of computer vision and robot arm movement. The reason for the localization error in computer vision is that the algorithm of recognition and detection can not adapt to the needs of the natural growing environment in the actual field. The localization error caused by the movement of the manipulator comes from the hardware equipment such as the manipulator and the motor.

Previous studies have studied tomato detection and tomato localization [2, 3]. This paper mainly focuses on the error caused by mechanical motion. The current picking robot does not give feedback on the picking results after picking the target. It does not form a closed-loop information. Little research has been done on providing corrective information to robots in the event of picking failure. In this paper, the computer vision technology was used to detect the feedback of the target tomato after picking and to provide the robot with corrective

information for the target that failed to pick. After field testing, we find that if we can find a connection between a relative fixed point and the target point, in which the problem can be solved well. Therefore, we put forward an idea to add a reference point in the site environment to analyze and judge the picking results. The main principle is to determine whether the tomato is successfully picked by comparing the spatial position information between the target tomato and the reference point after picking with the spatial position information of the picking template. At the same time, the error of the failed target is calculated to provide the robot with correction information. In order to improve the efficiency of the picking robot, firstly, it is necessary to detect the tomatoes and reference points with high precision, and then reduce the feedback time. Therefore, as long as 3D information of target tomato and reference point can be detected and located quickly, then the purpose of this study can be achieved.

2. RELATED RESEARCH

In recent years, scholars at home and abroad have carried out research on the localization technology of picking robot, and achieved certain results. Kondo et al. [4] developed a tomato picking robot, and mainly studied the customization of the end effector according to the characteristics of structured tomatoes planted in greenhouse environment. The end effector developed by the study is not only aimed at a single tomato, but directly at picking the whole string of tomatoes. Experiments show that the success rate of picking tomato string is only 50%, and it takes 15 seconds to pick the whole tomato string. Mehta and Burks [5] of the University of Florida in the United States has studied citrus picking robots. They

proposed a visual recognition scheme with two cameras, one installed in the center of the end effector position and the other installed in a fixed position. Through the picking experiment in the laboratory, the picking success rate reached 95%. Experimental measurement shows that the picking precision of this institute is about 15mm. Hayashi et al. [6] developed strawberry picking robot, which mainly includes machine vision system, mobile platform system, picking device system and so on. Through the experiment, the pick success rate can reach 54.99%, and the average pick success time is 8.6 seconds. Feng et al. [7, 8] of the National Engineering Research Center of Intelligent Equipment for Agriculture (NERCIEA) developed a picking robot for hanging-line cultivation of tomatoes, which adopts the rail-type mobile lifting platform, is equipped with a 4-DOF articulated mechanical arm, and has the end effector structure of sucking and pulling sleeve, air bag clamping and screwing separation. It's equipped with line laser vision system, and the fruit recognition and localization are realized by CCD camera and laser vertical scanning respectively. The results show that the picking time of single tomato fruit is about 24s, and the picking success rate is 83.9% under strong light and 79.4% under weak light, respectively. As to the tomato picking robot designed by Wang et al. [9], the picking claw consists of fruit adsorption, tightening and rotation; the vacuum generator can absorb tomatoes when they move to the target; through laser ranging, the fruits are covered by telescopic cylinders to complete picking. Experiments show that it takes 4 seconds to locate tomato fruit, 12 seconds to move the mechanical arm, 8 seconds to pick fruit and 12 seconds to reset the mechanical arm, and the success rate of robot picking reaches 83.9%.

Ling et al. [10] designed a dual-arm tomato picking robot. Two-DOF mechanical arms are symmetrically distributed, the target fruit is grasped by a vacuum cup, and then the fruit is separated by cutting. In its vision system, the sliding window method is used to extract haar-like features in each sub-window. The AdaBoost classifier is also used to detect tomatoes. In 171 target tomatoes, 60 sample images were used to test, the recognition success rate was 96.5%, and the average detection speed was 85ms per image. Eighty ripe tomatoes were randomly selected for picking experiment, and 70 target tomatoes were picked by the picking robot, with a success rate of 87.5% and an average picking rate of 29 seconds/tomato. Yaguchi et al. [11] designed a rotary claw end effector. After the end effector is positioned on the target fruit, the three claws first approach to clamp the tomato, drive the tomato to rotate relative to the stem, and separate it from the stem to realize picking. In the recognition process, color features are extracted by using hue, saturation and intensity space, then Euclidean distance is selected to cluster point clouds, and finally tomatoes are recognized by spherical fitting, which takes 200ms. The picking experiment shows that the picking efficiency is 23S/piece, and the picking success rate is 60%. Williams et al. [12] showed a robot with multiple mechanical arms to pick kiwifruit. The average cycle of each fruit is 5.5 seconds, and the robot picker can successfully pick 51.0% kiwifruit in the orchard. Jia et al. [13] proposed an apple recognition method based on pulse coupled neural network and genetic Elman neural network (GA-Elman) to improve the efficiency of picking apples. Xiong et al. [14] developed a machine vision system for strawberry localization, which is implemented on strawberry picking robot and tested in greenhouse strawberry production. Their test results show that the picking robot with optimized localization method can

achieve 74.1% picking rate under structured conditions. Miao et al. [15] proposed an algorithm for estimating the maturity of truss tomato and a synthesis method for stalk localization based on the experimental errors of each method. Both indoor and field tests were carried out using robot pickers. The results show that the proposed algorithm has high precision under different illumination conditions, and the average deviation is 2 mm. It can guide the robot to pick truss tomatoes effectively, and the average running time is 9 seconds/cluster. Rong et al. [16] proposed yolov5m model to recognize tomato in the greenhouse, and adopted the optimal sorting algorithm and the nearest neighbor localization algorithm to design directional grasping tomato. The Experimental results showed that the recognition precision of tomato is 97.3%, and the average harvest time of single fruit is 14.6s.

To sum up, a lot of work has been done at home and abroad in the aspect of fruit and vegetable picking and harvesting robots, and many achievements have been made.

In the research process, researches are mostly aimed at tomato recognition and localization of picking robot, and some researches are on end executive structure localization. However, there are few studies related to inaccurate fruit localization caused by mechanical motion errors. Therefore, this article proposes to study the localization error caused by robot movement, and analyze and compensate the position error of grasping fruit by machine vision, so as to achieve successful picking of fruits.

3. MATERIALS AND METHODS

3.1 Image acquisition

The research site is tomato greenhouse of China International Intelligent Agriculture Demonstration Base. In the experiment, tomato images are collected by mobile phone and ZED camera from multiple angles, and the imaging distance is 300mm-1200mm. The image resolution is 1280*720. The tomato images collected are shown in Figure 1.



Figure 1. Tomato images collected

3.2 Data enhancement mode

In the tomato planting environment of intelligent agricultural greenhouse, different light intensities and angles bring different image features, and the number of datasets will affect the learning ability and generalization ability of deep learning neural network training model. This requires that enough datasets be used to train the model, which can represent the image data of different environments and

different perspectives in greenhouse. The data needed in this article are tomato data and the relative position mark data of mechanical arm, which are basically to-be-picked tomato data and mark data. In order to achieve the general capability of the depth network model, image flipping, brightness balance, image rotation and image scaling are used to enhance the collected images. Among them, image flipping and rotation can improve the detection ability and stability of the network model, and brightness balance can avoid the influence of performance deviation of the network model due to sensor differences and ambient illumination changes [17, 18]. Finally, a total of 1000 images of tomato data samples are obtained, including 800 training sets, 100 verification sets and 100 test sets.

3.3 Method

3.3.1 Principle of YOLO

In YOLO algorithm, the object detection is directly regarded as the regression of position coordinates and confidence score. Therefore, YOLO algorithm can predict the categories and positions of multiple objects in real time once. Different from traditional target detection algorithms such as selecting sliding window method and Faster R-CNN algorithm to extract candidate regions, YOLO directly inputs the whole image into the network model for training and detection. This idea greatly improves the training and detection speed of network model.

In 2016, YOLO network was proposed by Redmon et al. [19]. Based on YOLO, yolov2 (Redmon and Farhadi [20]),

Yolov3 (Redmon and Farhadi, [21]) and Yolov4 (Bochkovskiy et al. [22]) were proposed. As a new excellent target detection technology, YOLO network has been widely recommended by scholars. It only needs a neural network to detect objects. YOLO can read the whole image once, and can identify the local information of the image, which greatly reduces the error detection rate of the background.

2020 saw the release of yolov5, which was well reflected in precision and speed. yolov5 model is divided into four versions: yolov5l, yolov5m, yolov5x and yolov5s according to the parameters depth multiple and width multiple. Among them, yolov5s model has the fastest detection speed and the smallest model parameters. Its network structure is shown in Figure 2.

The localization error feedback of tomatoes is mainly realized by tomato visual recognition technology, so the model is required to have high real-time and lightweight performance. This article studies the improved design based on yolov5s network structure, and the main improvements are as follows.

3.3.2 Backbone network improvement

When yolov5 algorithm is used to detect small targets, its detection effect is not good. There are many parameters in the training network model, and the memory space consumed by the model is large. Under the requirement of high real-time detection, the reasoning speed is not fast enough. Therefore, this article replaces the backbone network of yolov5s model with a lightweight ShuffleNetV2 [23], which reduces the training model parameters and makes the model lighter.

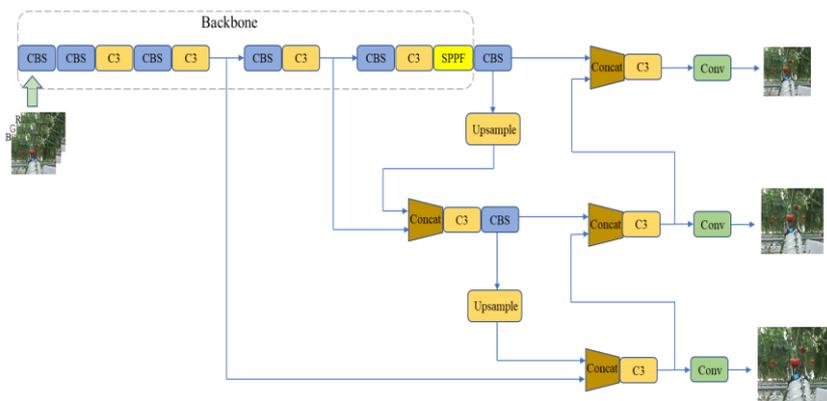


Figure 2. Yolov5 network structure diagram

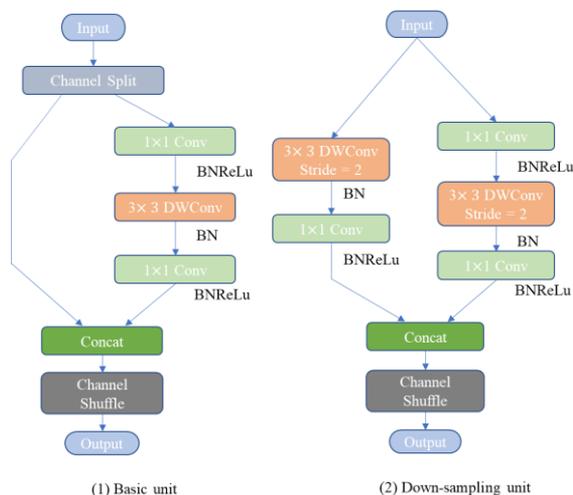


Figure 3. ShuffleNetV2 structure

The ShuffleNetV2 architecture consists mainly of a base unit and a down-sampling unit, as shown in Figure 3. The basic unit divides the number of input feature channels into two groups, the left branch is not processed, and the right branch is subject to convolution operation and batch normalization, which fuses the output features of the left and right branches and shuffles the channels [24], thus strengthening the fusion of sub-channel graph information. The down-sampling unit does not adopt channel separation operation, directly increases the number of network channels and the width of the network, and further strengthens the ability to extract network features [25].

3.3.3 Add BiFPN (Bi-directional feature pyramid network)

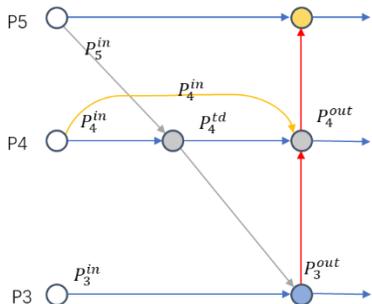


Figure 4. BiFPN structure

The original path aggregation network (PANet) was substituted with the BiFPN [26] to introduce bidirectional weighted fusion. In most networks, different image features are simply superimposed or added up, without any differentiation. With different resolutions, different features contribute variedly to the feature map outputted after feature fusion. For example, the shallow network has a high resolution and relatively clear position information, while the deep network has a wide perceptive field and many high-dimensional semantics. The BiFPN reduces the loss of characteristic information, and realizes multiscale and cross scale optimization by learning the importance of different

input features through weighted feature fusion, applying two-way (bottom-up and top-down) features repeatedly, and adding horizontal connections between input and output features on the same scale. The improved yolov5 adopts the BiFPN to achieve the weighted fusion of features from the third to the fifth layer. Figure 4 shows the structure of the BiFPN.

The calculation of Level 4 can be expressed as:

$$P_4^{td} = Conv\left(\frac{\omega_1 \cdot P_4^{in} + \omega_2 \cdot Resize(P_5^{in})}{\omega_1 + \omega_2 + \epsilon}\right) \quad (1)$$

$$P_4^{out} = Conv\left(\frac{\omega'_1 \cdot P_4^{in} + \omega'_2 \cdot P_4^{td} + \omega'_3 \cdot Resize(P_3^{out})}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon}\right) \quad (2)$$

where, *Resize* is usually down sampling or up sampling; *w* is the parameter learned to differentiate the importance of different features during feature fusion.

4. EXPERIMENTAL AND ANALYSIS

4.1 Experimental environment and parameter setting

The experimental equipment is Intel (R) Core (TM) i7-9700 CPU processor, WIN10 64-bit operating system, and the graphics card is NVIDIA GeForce GTX1660. The number of training data iterations is 300, batchsize is 6, and Works is 4.

4.2 Lightweight analysis of model

As can be seen from Table 1, it shows a comparison with other backbone models. The improved model has a high degree of lightweight, with parameters reduced by about 8 times to 0.8M, model size reduced by more than 7 times to 1.89MB and inference time reduced by more than 3 times to 4.8ms. From the experimental results, the overall detection performance is guaranteed, which is convenient for mobile device deployment and real-time requirements.

Table 1. Lightweight comparison of models

Model	Params (M)	FLOPs (G)	mAP@.5	P	R	Size (MB)	Inference time (ms)
yolov5	7.0	15.8	0.988	0.981	0.965	13.6	18.6
mobilenetV2	2.9	7.0	0.979	0.975	0.965	5.92	12.4
mobilenetV3	3.5	6.3	0.981	0.977	0.959	7.06	7.6
ShuffleNetV2+bifpn(ours)	0.8	1.9	0.974	0.975	0.963	1.89	4.8

4.3 Target detection and analysis

The yolov5s model and the improved yolov5s model are trained and tested under the same dataset. Their PR curves are shown in Figure 5 and Figure 6. It can be seen from the figures that the improved yolov5(ShuffleNetV2+bifpn) training model achieves precision 0.975 and mAP 0.974. Although the accuracy is reduced, but still maintain the high accuracy. It can fully meet the requirements of detection targets in the feedback stage.

The comparison results of mAP@0.5 are shown in Figure 7. The model training starts from scratch. After 300 iterations, the curve tends to the highest value. The original model has

slight oscillation, and the improved model is relatively smooth, without fitting. On the whole, the improved model yolov5(ShuffleNetV2+bifpn) is stable and reliable.

For the two models, target detection is carried out in different target scenes, as shown in Figure 8 and Figure 9. The yolov5_improved represents backbone (ShuffleNetV2+bifpn) network. The detection precision of tomato class and bzu class in all model is close to the same mAP@0.5 value in the range of mechanical arm picking. In the middle view area of the image, the yolov5 improved confidence of the bzu class is a little higher than the original, and the small target detection is more accurate, which provides a good basis for error analysis and correction in the process of mechanical arm picking.

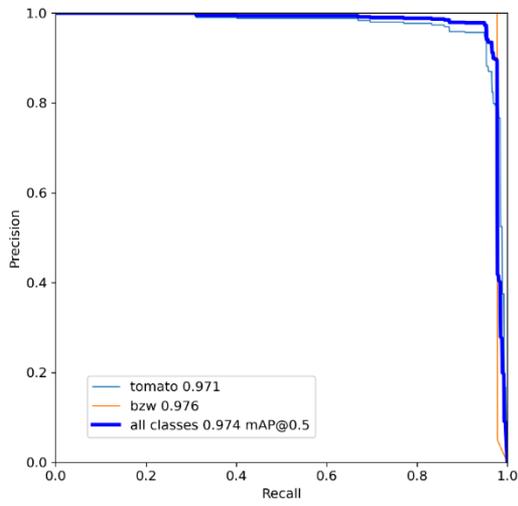


Figure 5. PR curves of yolov5(ShuffleNetV2+bifpn)

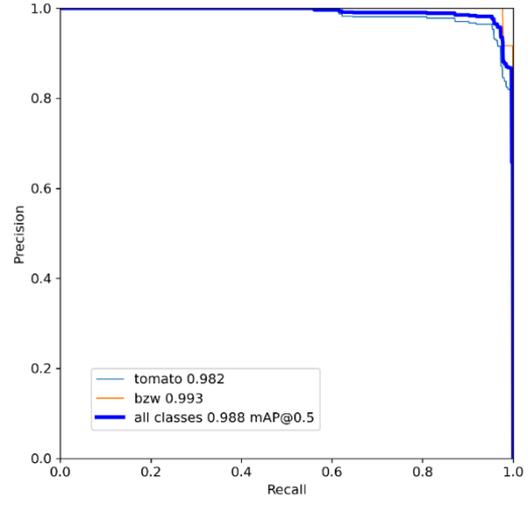


Figure 6. PR curves of yolov5

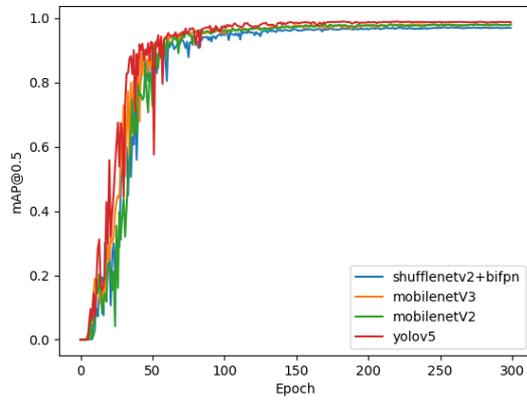


Figure 7. The mAP@0.5 curves



Figure 8. Scene I target detection



Figure 9. Scene II target detection

5. FEEDBACK OF TOMATO PICKING INFORMATION

5.1 Feedback principle

In the process of tomato picking and localization, there are two main errors, one is the error caused by tomato recognition and detection, and the other is the mechanical error caused by mechanical arm operation. The former can be solved by a more accurate algorithm, which has been studied before. This article mainly focuses on the latter and carries out experimental analysis. Mechanical errors are often transmitted to the controller by detecting the coordinate information of tomatoes, and the controller controls the movement of the mechanical arm according to the transmitted tomato localization information; then it is impossible to determine whether it reaches the accurate position, and most of them have no feedback of the information. This stage mainly determines whether the target tomato has been picked successfully by the

robot arm. That is, whether it has reached the position of the specified target tomato. If the target tomato has been picked successfully, then go on the next tomato. If failed, it needs to correct information which will be sent to picking robot and pick again.

Figure 10 shows the process of picking tomatoes by mechanical arm, and the images are taken and collected by binocular stereo camera. In the previous articles, the 3D information of tomatoes was obtained by binocular stereo vision calculation. In this article, binocular vision is used to move the range information of tomatoes. There are two types of images, one is tomato, and the other is b/w; the b/w tag is on the end paw of the mechanical arm. Its main function is to determine whether the mechanical arm accurately reaches the position of the tomato by comparing the tomato to be picked with its 3D coordinates, and to realize the whole closed loop by feeding back to the controller. Figure 11 shows a side view of the picking process.

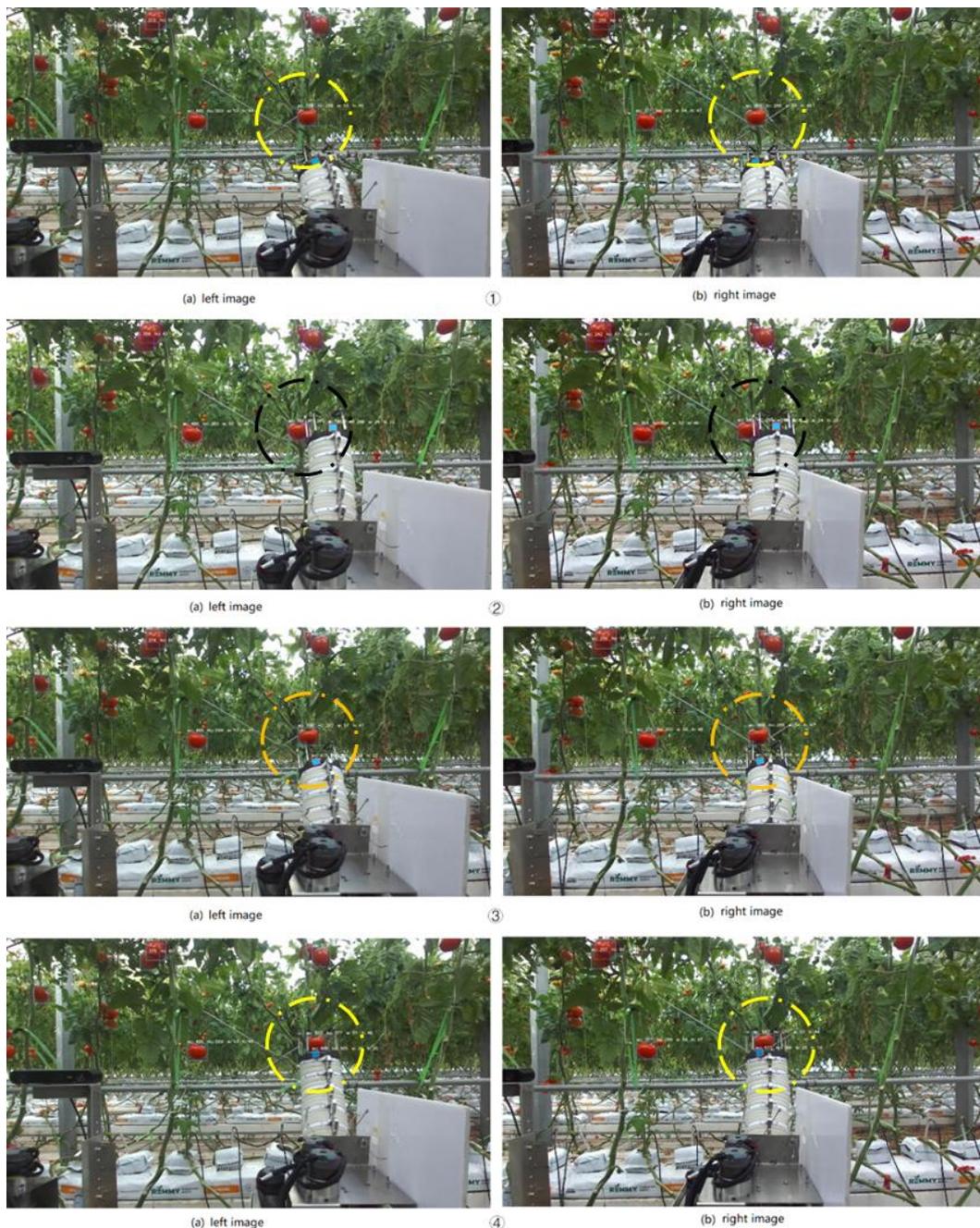


Figure 10. Picking and localization process



Figure 11. Side view of tomato picking

5.2 Detection and localization in Feedback stage

According to the previous study [3], it combined with the detection of reference points and target tomatoes in the feedback stage, their 3D coordinates can be calculated. The main representative renderings are shown in Figure 12. It only displays the 3D coordinates of the reference point and the picking point.

In order to observe and analysis of information between reference points and target tomatoes in the feedback stage, a full range of tomatoes was picked in the different scenes, it was up to a total of 90 picking sample data. These data are displayed in 3D space, as shown in Figure 13. Among them, they were picked successfully by 50 times, and the 50 sample data was shown in line Figure 14.

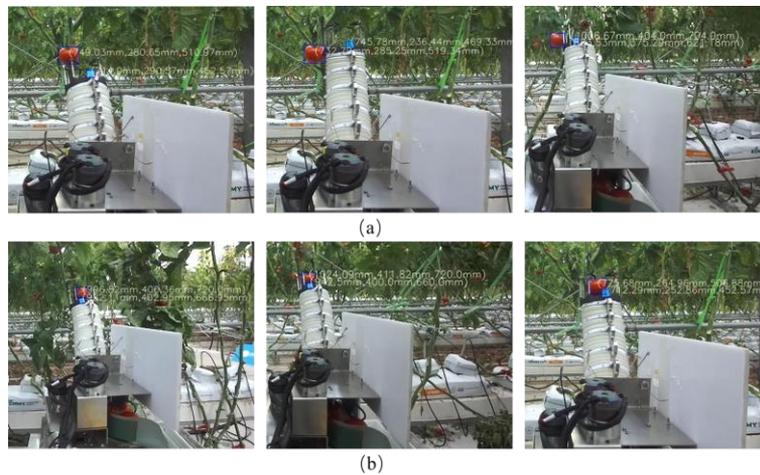


Figure 12. Reference point and target point 3D coordinate: (a) The 3D coordinates of tomato picked unsuccessfully; (b) The 3D coordinates of tomato picked successfully

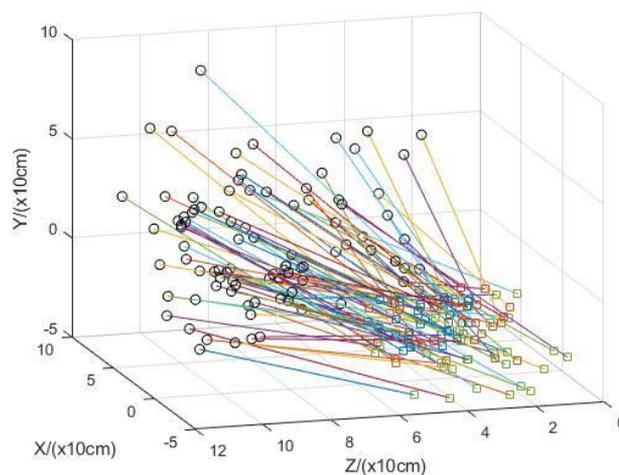


Figure 13. The spatial distribution map of reference points and target points for picking success and failure in the feedback stage (Square head represents reference point, round head represents tomatoes, they are connected by solid lines)

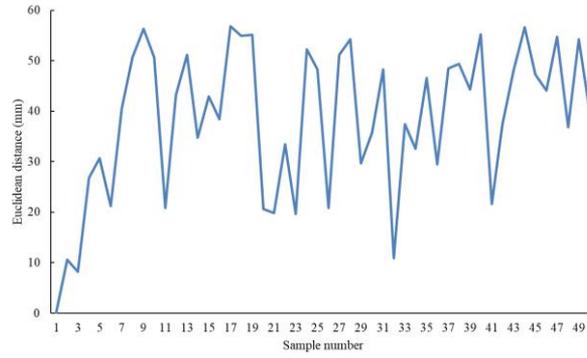


Figure14. Euclidean distance between reference point and target tomato picked successfully

Figure 13 and Figure 14 show that the distance is in range of 8.14 mm to 58.12 mm between tomatoes picked successfully and the reference point. So, when the distance between the picking point and the reference point is less than 58.12 mm, it is considered that the tomato is picked successfully. The calculation formula is expressed as (3) and (4).

$$L_i = \rho_i(C_i, T_i) \quad (3)$$

$$f(L_i) = \begin{cases} 1, & L_i \leq 58.12 \text{ mm} \\ 0, & L_i > 58.12 \text{ mm} \end{cases} \quad (4)$$

T_i represents the 3D coordinates of the i th tomato, C_i represents the 3D coordinates of reference point of the i th corresponding tomato, ρ_i represents 3D euclidean distance, the $f(\cdot)$ represents the state of the result, with 1 indicating success and 0 indicating failure.

5.3 Localization information feedback and correction

Figure 15 shows the left and right images under the picking

tomatoes, and the information detected by tomato and mark recognition. The specific data are shown in Table 2. Through these information and binocular hardware parameters, according to the previous research, 3D coordinates can be solved as tomato (x_1, y_1, z_1) and bzu (x_2, y_2, z_2).

From the data in Table 2, the 3D difference between the tomato to be picked and the mark can be obtained, as shown by the following formula:

$$\begin{cases} \Delta x = x_1 - x_2 \\ \Delta y = y_1 - y_2 \\ \Delta z = z_1 - z_2 \end{cases} \quad (5)$$

The distance between the tomato to be picked and the mark can be obtained from formula (5) as follows:

$$L = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \quad (6)$$

Since both the mark and the picking claw are fixed, the center of the claw (tomato) and the center of the reference point are fixed, which is L in Formula (6). It can be obtained from Table 2.

Table 2. Recognition and detection information of tomatoes and marks

Category	Left image coordinates (pixel)	Right image coordinates (pixel)	Right image width (w) (pixel)	Right image height (h) (pixel)	3D coordinates (mm)
Tomato	(810,277)	(683,277)	53	44	166.30, -74.65, 498.90
Bzu	(809,305)	(673,306)	20	19	154.41, -45.00, 465.88

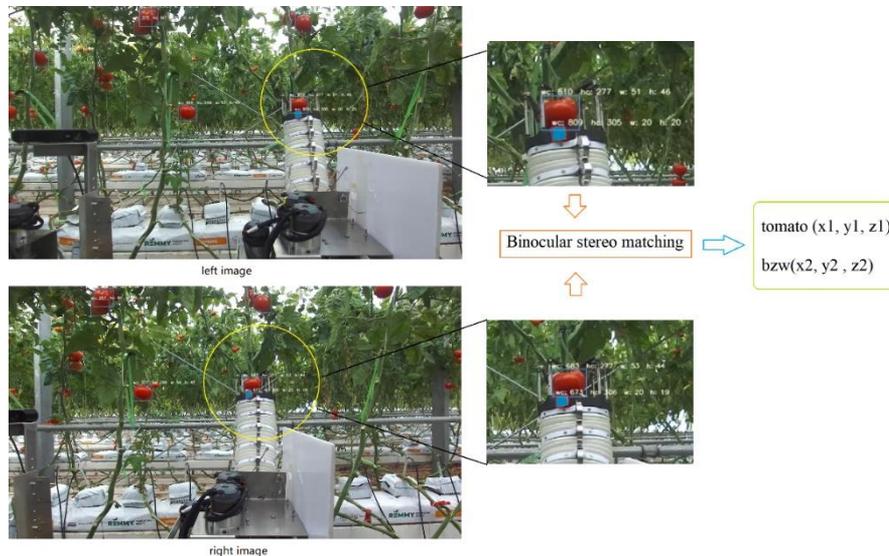


Figure 15. Binocular stereo 3D calculation

Therefore, in the picking process, after the mechanical arm moves, it is possible to calculate the ratio of the 3D distance L' to L between the tomato to be picked and the reference point. In this design, the fault tolerance range of the mechanical arm is 58.12 mm, that is to say, when $L \leq 58.12$ mm, the mechanical arm accurately reaches the position of the tomato to be picked and can perform the picking action.

When $L'(x', y', z')$ is not within the range that can be picked, it is necessary to correct the position information. Correction information as formula (7).

$$\Delta L = L' - 58.12 \text{ mm} \quad (7)$$

Then, the feedback information ΔL is sent to the robot, it completed visual feedback and correction.

6. CONCLUSION

The improved yolov5s detection method proposed in this article replaces the lightweight ShuffleNetV2 as the backbone network, and also add BIFPN. The actual experimental results show that the performance of model parameters is reduced by 8 times, the size of model is reduced by 7 times, and the inference time of each image is reduced by 3 times, which meets the requirements of lightweight deployment and real-time performance. Especially in the visual perspective of front view, tomato and bzu classes can be correctly identified, which provides timely feedback and correction for the localization information of the mechanical arm and improves the efficiency. In the complex environment of greenhouse, in order to make tomato localization detection information more accurate and faster, it is necessary to optimize the algorithm detection precision and detection speed and various parameters of the model. At the same time, it is also suggested to strengthen the precision of mechanical motion.

ACKNOWLEDGMENT

The work is supported by S&T Program of Hebei (Grant No.: 21327212D) and funded by Science and Technology Project of Hebei Education Department (Grant No.: ZC2022095).

REFERENCES

- [1] Liu, C.L., Lin, H.Z., Li, Y.M., Gong, L., Miao, Z.H. (2020). Analysis on status and development trend of intelligent control technology for agricultural equipment. *Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery*, 51(1): 1-18.
- [2] Gao, G.H., Wang, S.Y., Shuai, C.Y., Zhang, Z.H., Zhang, S., Feng, Y.B. (2022). Recognition and detection of greenhouse tomatoes in complex environment. *Traitement du Signal*, 39(1): 291-298. <https://doi.org/10.18280/ts.390130>
- [3] Gao, G., Wang, S., Shuai, C. (2023). Optimization of greenhouse tomato localization in overlapping areas. *Alexandria Engineering Journal*, 66: 107-121. <https://doi.org/10.1016/j.aej.2022.11.036>
- [4] Kondo, N., Yata, K., Iida, M., Shiigi, T., Monta, M., Kurita, M., Omori, H. (2010). Development of an end-effector for a tomato cluster harvesting robot. *Engineering in Agriculture, Environment and Food*, 3(1): 20-24. [https://doi.org/10.1016/S1881-8366\(10\)80007-2](https://doi.org/10.1016/S1881-8366(10)80007-2)
- [5] Mehta, S.S., Burks, T.F. (2014). Vision-based control of robotic manipulator for citrus harvesting. *Computers and Electronics in Agriculture*, 102: 146-158. <https://doi.org/10.1016/j.compag.2014.01.003>
- [6] Hayashi, S., Yamamoto, S., Tsubota, S., Ochiai, Y., Kobayashi, K., Kamata, J., Peter, R. (2014). Automation technologies for strawberry harvesting and packing operations in Japan I. *Journal of Berry Research*, 4(1): 19-27. <https://doi.org/10.3233/JBR-140065>
- [7] Feng, Q., Wang, X., Wang, G., Li, Z. (2015). Design and test of tomatoes harvesting robot. In 2015 IEEE International Conference on Information and Automation, Lijiang, China, pp. 949-952. <https://doi.org/10.1109/ICInfA.2015.7279423>
- [8] Feng, Q., Wang, X., Wu, P., Wang, G. (2016). Design and test of tomatoes harvesting robot. *Journal of Agricultural Mechanization Research*, 38(4): 94-98.
- [9] Wang, L.L., Zhao, B., Fan, J.W., Hu, X.A., Wei, S., Li, Y.S., Zhou, Q.B., Wei, C.F. (2017). Development of a tomato harvesting used in greenhouse. *International Journal of Agricultural and Biological Engineering*, 10(4): 140-149. <https://doi.org/10.25165/j.ijabe.20171004.3204>
- [10] Ling, X., Zhao, Y., Gong, L., Liu, C., Wang, T. (2019). Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision. *Robotics and Autonomous Systems*, 114: 134-143. <https://doi.org/10.1016/j.robot.2019.01.019>
- [11] Yaguchi, H., Nagahama, K., Hasegawa, T., Inaba, M. (2016). Development of an autonomous tomato harvesting robot with rotational plucking gripper. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea (South), pp. 652-657. <https://doi.org/10.1109/IROS.2016.7759122>
- [12] Williams, H.A., Jones, M.H., Nejati, M., Seabright, M.J., Bell, J., Penhall, N.D., MacDonald, B.A. (2019). Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. *Biosystems Engineering*, 181: 140-156. <https://doi.org/10.1016/j.biosystemseng.2019.03.007>
- [13] Jia, W., Mou, S., Wang, J., Liu, X., Zheng, Y., Lian, J., Zhao, D. (2020). Fruit recognition based on pulse coupled neural network and genetic Elman algorithm application in apple harvesting robot. *International Journal of Advanced Robotic Systems*, 17(1): 1-14. <https://doi.org/10.1177/1729881419897473>
- [14] Xiong, Y., Ge, Y., From, P.J. (2020). An obstacle separation method for robotic picking of fruits in clusters. *Computers and Electronics in Agriculture*, 175: 105397. <https://doi.org/10.1016/j.compag.2020.105397>
- [15] Miao, Z., Yu, X., Li, N., Zhang, Z., He, C., Li, Z., Sun, T. (2023). Efficient tomato harvesting robot based on image processing and deep learning. *Precision Agriculture*, 24(1): 254-287. <https://doi.org/10.1007/s11119-022-09944-w>
- [16] Rong, J., Wang, P., Wang, T., Hu, L., Yuan, T. (2022). Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots. *Computers and Electronics in Agriculture*, 202: 107430. <https://doi.org/10.1016/j.compag.2022.107430>

- [17] Ma, J., Li, Y., Chen, Y., Du, K., Zheng, F., Zhang, L., Sun, Z. (2019). Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *European Journal of Agronomy*, 103: 117-129. <https://doi.org/10.1016/j.eja.2018.12.004>
- [18] Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, 157: 417-426. <https://doi.org/10.1016/j.compag.2019.01.012>
- [19] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [20] Redmon, J., Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263-7271. <https://doi.org/10.1109/CVPR.2017.690>
- [21] Redmon, J., Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. <https://doi.org/10.48550/arXiv.1804.02767>
- [22] Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. <https://doi.org/10.48550/arXiv.2004.10934>
- [23] Ma, N., Zhang, X., Zheng, H.T., Sun, J. (2018). Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 116-131.
- [24] Zhang, X., Zhou, X., Lin, M., Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848-6856.
- [25] Zhang, X., Zhou, Y.C., Liu, Z.Y., Li, X.Z. (2022). Identification and application of apple leaf diseases based on improved ShuffleNet V2 Model. *Journal of Shenyang Agricultural University*, 53(1): 110-118.
- [26] Tan, M., Pang, R., Le, Q.V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10781-10790. <https://doi.org/10.48550/arXiv.1911.09070>