Apple recognition and picking sequence planning for harvesting robot in a complex environment

Wei Ji, Tong Zhang, Bo Xu, Guozhi He

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China

Abstract

In order to improve the efficiency of robots picking apples in challenging orchard environments, a method for precisely detecting apples and planning the picking sequence is proposed. Firstly, the EfficientFormer network serves as the foundation for YOLOV5, which uses the EF-YOLOV5s network to locate apples in difficult situations. Meanwhile, the soft non-maximum suppression algorithm is adopted to achieve accurate identification of overlapping apples. Secondly, the adjacently identified apples are automatically divided into different picking clusters by the improved density-based spatial clustering of applications with noise. Finally, the order of apple harvest is determined to guide the robot to complete the rapid picking, according to the weight of the Gauss distance weight combined with the significance level. In the experiment, the average precision of this method is 98.84%, which is 4.3% higher than that of YOLOV5s. Meanwhile, the average picking success rate and picking time are 94.8% and 2.86 seconds, respectively. Compared with sequential and random planning, the picking success rate of the proposed method is increased by 6.8% and 13.1%, respectively. The research proves that this method can accurately detect apples in complex environments and improve picking efficiency, which can provide technical support for harvesting robots.

Correspondence: Wei Ji, Jiangsu University, Xuefu Road 301 Zhenjiang 212013, China.
E-mail: jiwei@ujs.edu.cn

Key words: apple recognition; picking planning; EF-YOLOV5s; DBSCAN.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: none.

Availability of data and material: the data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: this work was supported in part by the National Natural Science Foundation of China (NSFC) [grant number 61973141].

Received: 7 July 2023.
Accepted: 25 September 2023.

Journal of Agricultural Engineering 2024; LV:1549
doi:10.4081/jae.2024.1549

Introduction

The planting area and output of apples in China account for more than 50% of the wors production, but picking apples is still mainly manual and expensive (Wang *et al.*, 2016). Therefore, the apple picking robot is the future development direction (Hu *et al.*, 2022; Ji *et al.*, 2021). However, the apple picking robot still has problems with low precision and low picking efficiency in complex environments (Wang *et al.*, 2017), so how to improve the picking efficiency of the apple-picking robot is the focus and difficulty of the research on the apple-picking robot (Bu *et al.*, 2022). At present, in order to solve the problem of the low efficiency of apple robot picking, researchers have carried out research on apple identification and picking planning (Zhang *et al.*, 2016; Tang *et al.*, 2023; Wu *et al.*, 2023).

For apple target recognition, Ji *et al*. (2022) described an apple detection approach based on Shufflenetv2-YOLOX, which could achieve 26.3 frames per second on the Jetson Nano. And the method boosted the precision and speed of detection. Jia *et al*. (2020) proposed a visual detection network based on Mask R-CNN, which took advantage of the detection of overlapping apples. The precision rate of the model is 97.31%, and the recall rate is 95.70%. Gao *et al*. (2020) proposed a multi-class apple identification method for dense fruit trees based on a fast regional convolutional neural network, which could effectively identify occluded apples with an average precision of 0.879.

In the research on apple-picking planning, Zhang *et al*. (2021) drove a three-degree-of-freedom apple-picking manipulator based on a nonlinear control plan, and the average time to plan an apple was 8.8 seconds. Zhao *et al*. (2011) succeeded in planning apples in 77% of their experiments. The complex calculation of algorithms resulted in an average picking time of 15 seconds. Yu *et al*. (2021) developed a new apple-picking robot, which had a precision rate of 82.5% and a success rate of 72%. The single apple-picking time was approximately 14.6 seconds, which was still much slower than manual apple-picking. The aforementioned research findings essentially show the average precision and speed of detection, but the anti-interference capacity and efficiency of robots need to be further enhanced in subsequent studies.

Despite the fact that the above study has significantly improved the precision of recognition, there are still problems. In challenging environments, high precision and speed of detection cannot coexist (Xu *et al.*, 2023). It is also hard to automatically identify the picking cluster and locate the best path in dense planting, and robotic picking planning is complicated and inefficient. The outcome is that the final apple-picking takes a long time and does not yield the intended results. In order to increase the precision of detection and reduce the picking time, this study offers a method to precisely detect apples and plan the picking sequence.

The following is a summary of the major contributions of this study: i) in order to improve the precision, a method for apple detection in the complex orchard is provided based on the EF-YOLOV5s network, which uses the EfficientFormer structure and soft non-maximum suppression (NMS) to improve the precision and effectiveness; ii) in order to improve the efficiency of picking, a picking sequence plan is suggested based on an enhanced density-based spatial clustering of applications with noise (DBSCAN) algorithm, which is logically designed by combining Gaussian weights of distance and significance level. The function of the improved DBSCAN algorithm is automatically dividing the nearby apples into several picking clusters.

Materials and Methods

Image acquisition

In this study, the Fengxian Apple Demonstration Base in Xuzhou City, Jiangsu Province, China, provided the apples used in the research. Fuji Apple, the primary apple variety in China, served as the research object. The distance between the camera and the apple is maintained at 0.3-2 m. The richness of the data set was ensured by the simultaneous capture of 1877 images in various natural settings, including 1234 images of unbagged apples, 316 images of apples at night, and 327 images of bagged apples, as shown in Figure 1. It is required to increase the number of images, such as apples at night and bagged apples, because deep learning has restrictions on the size of the data set. In order to extend these two types of images, the image enhancement approach is utilized (Ji *et al.*, 2023). The images are enhanced by using the Albalentations library in PyTorch, which contains mirror image flipping, scaling rotation, randomly rearranging the RGB channels of the input image, image exposure, composition enhancement, and other operations. Finally, 9358 images were acquired. By using LabelImg annotation, a training set and a verification set are created in the ratio of 9:1. A deep-field camera captured 50 apple images simultaneously to assess the effectiveness of the method.

Apple target detection based on EF-YOLOV5s network

This research suggests an EF-YOLOV5s network based on YOLOV5s to improve the precision and speed of detection in a real-time apple-picking robot, and its network topology is depicted in Figure 2. And replacing the C3 structure with the EfficientFormer structure can increase the speed and precision of detection (Li *et al.*, 2022). Additionally, the SoftNMS algorithm, which is an improvement of the original NMS method, increases the network's precision for overlapping apples (Bodla *et al.*, 2017).

Backbone design

The C3 structure, which has many parameters and a slow speed of detection, is used by YOLOV5's feature extraction network. This vast and sophisticated network is challenging to implement in some real-world application settings, such as mobile or embedded devices. Additionally, the transformer has the ability to gather global information, which can compensate for YOLOV5's inadequacies, since it is a convolutional neural network and cannot conduct global modeling (Han *et al.*, 2021). The newest lightweight Transformer network, EfficientFormer, can simultaneously increase speed and fix flaws in convolutional neural networks. EfficientFormer serves as the foundation of YOLOV5's feature extraction network in this study. The construction of EfficientFormer is depicted in Figure 3.

Two convolution layers are the first levels in EfficientFormer, which is then followed by a MetaBlock layer and a patch embedding layer. Both of these layers contain pooled and multi-head self-attention mixers in various arrangements that are all the same size.

Anchor frame algorithm improvement

YOLOV5s automatically adopts the NMS algorithm, primarily using Intersection over Union (IoU) to weed out eligible boxes. IoU is one of them and serves as a benchmark for gauging the precision of matching items in a given data set. The final box that satisfies the requirements is obtained by the NMS method after iteratively performing IoU operations with other boxes and filtering those with high IoU (Wu *et al.*, 2020).

The NMS algorithm forcibly discards consecutive identification frames with overlaps bigger than the overlapping threshold as a result of this filtering technique, which lowers the precision of detection of genuine objects in overlapped areas. Additionally, the NMS threshold value must be carefully chosen. It's necessary to use experience to debug threshold values.

The screening formula of the NMS identification box is as follows:

$$s_i = \begin{cases} s_i, & iou(\mathcal{M}, b_i) < N_t \\ 0, & iou(\mathcal{M}, b_i) \geq N_t \end{cases} \quad (1)$$

where s_i is the score of the recognition frame, b_i represents the candidate box, $\mathcal{M}$ is the maximum score, and N_t is the overlap threshold.

This study therefore opts to apply the soft-NMS method based on NMS. Instead of removing identification boxes forcibly like NMS does, the soft-NMS algorithm primarily maintains identification accuracy by lowering the score of overlapping boxes, which significantly improves the accuracy of the identification of overlapping apples.

The following formula is used in the soft-NMS identification box screening:

$$s_i = s_i e^{-\frac{iou(\mathcal{M}, b_i)^2}{\sigma}}, \ \forall b_i \notin D \quad (2)$$

where D is the last detection frame, and σ is the parameter of Gaussian penalty function. Non-maximum suppression starts with a list of detection boxes B with scores S. After selecting the detection with the maximum score $\mathcal{M}$, it removes it from the set B and appends it to the set of final detections D. It also removes any box that has an overlap greater than a threshold N_t with $\mathcal{M}$ in the set B. This process is repeated for the remaining boxes B. A major issue with non-maximum suppression is that it sets the score for neighboring detections to zero. Thus, we propose a single-line modification to the traditional greedy NMS algorithm in which we decrease the detection scores as an increasing function of overlap instead of setting the score to zero as in NMS. This soft-NMS algorithm is shown in *Supplementary Figure 1*.

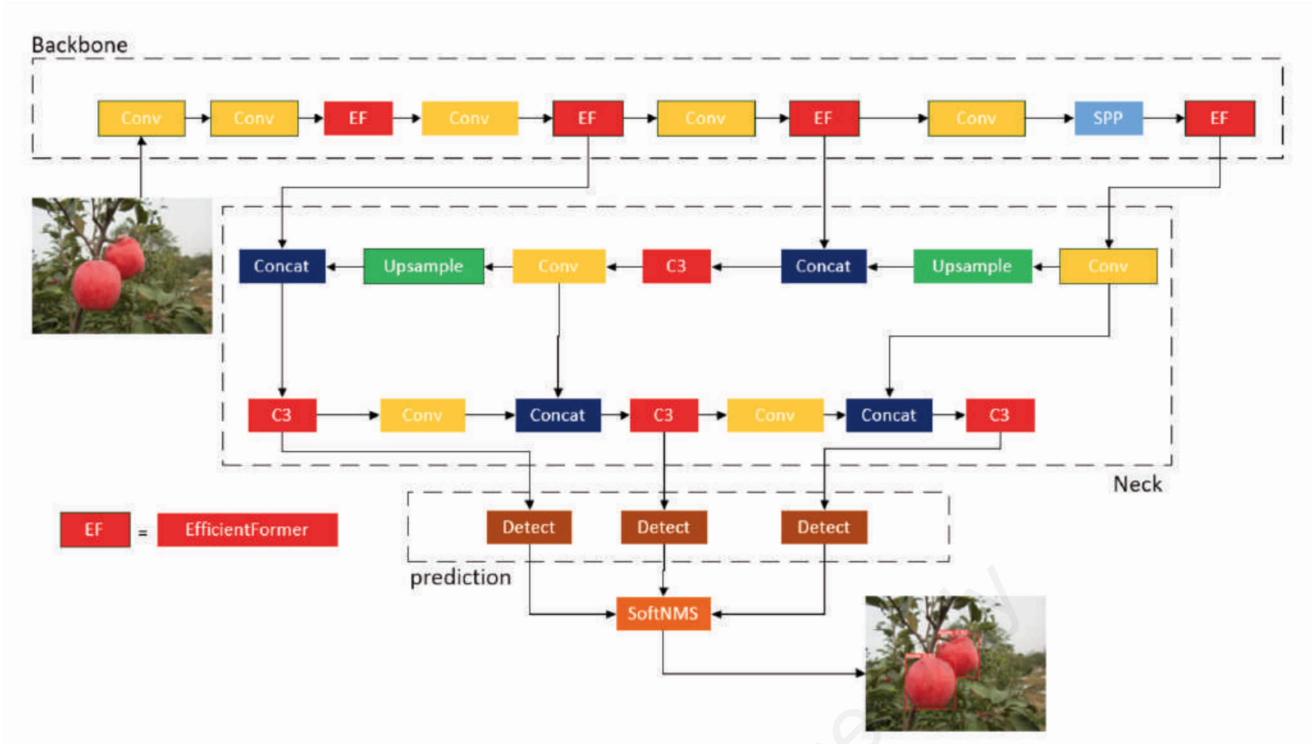

Figure 2. EF-YOLOV5s network structure schematic diagram.

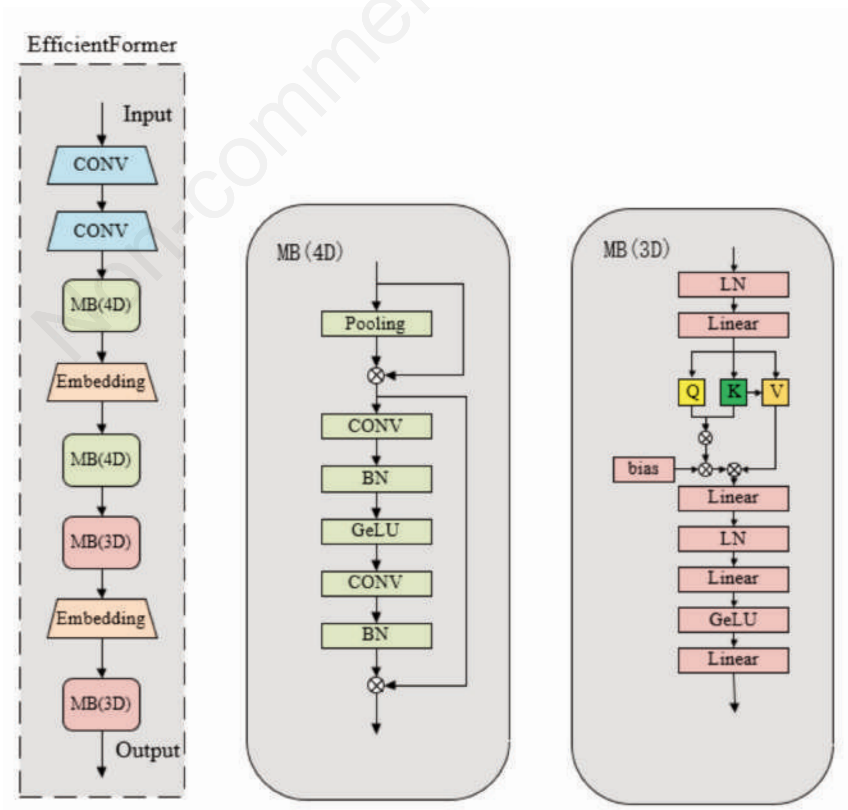

Figure 3. Structure of EfficientFormer.

Identify the location of apples

Planning the picking sequence begins with accurate apple positioning and detection. Every apple in a dense orchard may have a complicated background. The three-dimensional location coordinates L_i of apples are expected for the end effector's picking action after the robot has identified the apple. In this study, the center coordinates of the detection frame on the detection map are used as the position of the apple anchor point, and the three-dimensional coordinates of the apple anchor point in the camera coordinate system are calculated according to the depth information and camera parameters.

$$L_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = z\rho \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (3)$$

where z is the depth measured by the camera, and ρ is a 3×3 matrix containing the internal parameters of the camera. u and v are the horizontal and vertical coordinates of the apple center point.

According to Eq. (3), the spatial distance D_{ij} between two picking points can be obtained as follows:

$$D_{ij} = D_{L_i} - D_{L_j} = \sqrt{(x_i^2 - x_j^2) + (y_i^2 - y_j^2) + (z_i^2 - z_j^2)} \quad (4)$$

Design of apple picking sequencing based on the improved density-based spatial clustering of applications with noise algorithm

A crucial requirement for effective robot picking is the automatic sorting of recognized apples into picking clusters. As a result, the picking planning method presented in this research is based on the improved DBSCAN clustering algorithm. The high-density area is referred to as an "apple cluster" to distinguish it from the low-density area. Then, by utilizing depth information and the significance level of the Gaussian kernel function, the picking order of each apple fruit is planned.

Apple cluster classification based on the improved density-based spatial clustering of applications with noise clustering algorithm

DBSCAN is a density-based spatial clustering algorithm that can divide regions with sufficient density into clusters and find clusters with arbitrary shapes in noisy spatial databases. However, the DBSCAN clustering algorithm relies on manual judgment when deciding on the domain radius (Eps) and the minimum coverage points of core points (MinPts), making it difficult to apply automation to this process (Schubert *et al.*, 2017). The study therefore suggests a novel technique based on the K-distance function (Swanepoel *et al.*, 1999). The construction of the k-distance function and its integration with real objects yield the values of Eps and MinPts. The following definitions are given:

Eps: domain radius, that is, the maximum distance between two apples.

MinPts: the minimum number of domain points required for a given apple to become the core object in the domain.

Selection of Eps

To realize the function of automatically clustering the identified apples quickly and accurately, this study proposes to use the k-distance function to determine the Eps value in the DBSCAN clustering algorithm. The spatial distance D_{ij} is sorted by size using the k-distance function, and the value of Eps is based on where the first inflection point is located. The k-distance plot is shown in *Supplementary Figure 2*.

In this study, the characteristic number, dim, is used to derive the k value according to Eq. (5), where k equals 3.

$$k = 2 \times dim - 1 \quad (5)$$

Selection of MinPts

The choice of MinPts is guided by the principle that dim denotes the dimension of the data to be clustered. Setting MinPts to 1 would be illogical because each single apple fruit would constitute a separate cluster. The outcome was the same for the neighboring cluster at that time. MinPts must therefore select a number greater than or equal to 3. Two nearby clusters with high densities, though, might merge into one cluster if the value is sufficiently high. Given the analysis above, the value of MinPts in this study is:

$$MinPts = k + 1 \quad (6)$$

Process of improving the density-based spatial clustering of applications with noise algorithm

Algorithmic processes are optimized by combining deep learning reliability and coordinate information. The accuracy of robot clustering is improved, which lays a good foundation for picking sequence planning. The process of the improved DBSCAN algorithm is shown in Figure 4.

Picking sequence planning method based on Gaussian weights of distance and significance level

In order to ensure efficient picking in a complex environment, this study determines the following picking plan based on the following information: i) divide prepared apples with improved DBSCAN density clustering algorithms; ii) according to the weight of distance, all apples are picked one by one, starting from the nearest cluster to the farthest cluster; iii) for each cluster, the apple is picked according to its significance level, and the outermost apple is given priority to prevent collision with other apples.

To plan the picking sequence according to the above picking plan, it is necessary to sort the order of clusters first and then sort the apples in each cluster in turn. Therefore, this study uses the Gaussian kernel function to rank each cluster. Then, according to the significance level of each apple, the picking order of apples in the cluster is planned (Sun *et al.*, 2018).

The center of the coordinate system is where the depth camera of the apple picking robot is located. The Gaussian kernel function is used to assign weights of distance to each apple cluster in the detection area (Wang *et al.*, 2003):

$$k(x, x^*) = e^{-\frac{\|x - x^*\|^2}{2\sigma^2}} \quad (7)$$

where $k(x,x^*)$ is the Gaussian kernel function, x is the three-dimensional coordinate of the core point P of the apple cluster, and x^* is the center point. The closer the apple cluster is to the center point, the greater its weight for distance. To sum up, this research sorts the Gaussian weights of distance across apple clusters after first grouping the discovered apples, by using the DBSCAN. The apples with significance levels ranging from high to low were then selected one at a time, after the association between significance level and depth information had been established. Repeat the process until all of the apples have been selected.

Results and Discussion

Introduction of the experimental platform

The configuration of the industrial computer used in this study is shown in Table 1. The detection algorithm is written in Python on PyCharm. In model training, epoch is set to 150, batch size is set to 64, and input image size is set to 640× 640. In order to verify the performance of the identification algorithm proposed in this study, 50 test sets in complex situations were selected for identification and location experiments. Recall, precision, and *F1* score are used to evaluate the detection performance.

$$recall = \frac{TP}{TP + FN} \times 100\% \tag{8}$$

$$precise = \frac{TP}{TP + FP} \times 100\% \tag{9}$$

$$F1 = \frac{2 \times precision \times recall}{precision \times recall} \times 100\% \tag{10}$$

where *TP* is the number of positive samples predicted by positive samples, *FP* is the number of positive samples predicted by negative samples, and *FN* is the number of negative samples predicted by positive samples.

EF-YOLOV5s network performance verification

In order to verify the detection effect of the network on apples in complex environments, this study takes 50 complex orchard pictures as the test set, including 10 unbagged apples, 20 apples at night, and 20 bagged apples. Because night and bagging are the key in the visual research of picking robots at present, this study chooses night and bagging test images to account for a high proportion, which can better reflect the detection effect of the network on apples in complex environments. At the same time, this study chose to conduct ablation experiments to evaluate each step. As evaluation indices, average precision (AP), frame per second (FPS), and parameter quantity are chosen. The results of ablation experiments are shown in Table 2.

As can be seen from the data in Table 2, every step of improvement is an effective improvement, that effectively improves the

Table 1. Industrial computer configuration.

Computer configuration	Specific parameters
Operating system	Windows 10
CPU	Intel e5-2683
Random access memory	64GB
GPU	GTX1080ti

CPU, central processing unit; GPU, graphics processing unit.

Figure 4. Process of the improved density-based spatial clustering of applications with noise algorithm.

Table 2. Industrial computer configuration.

YOLOV5s	EfficientFormer	SoftNMS	AP	FPS	Param (M)
✔			94.36%	19	7.20
✔	✔		97.21%	33	10.22
✔	✔	✔	98.81%	28	10.78

AP, average precision; FPS, frame per second.

identification speed or precision of the network. The AP of EF-YOLOV5s is 98.81%. Although the parameters increased by 3.58 M, the detection speed increased by 32% to 28 FPS. EfficientFormer network structure effectively improves the detection effect of the network. And this module can effectively enhance the feature extraction ability of backbone.

In order to verify the effectiveness of the method, we took the feature extraction effect and precision of the improved model as the basis for judgment, and conducted strict comparative experiments. The experimental results were obtained as shown in *Supplementary Figure 3*. Compared with three groups of experiments in different environments, the improved model obviously has a larger and more accurate feature extraction area, which means that the improved model can bring better training effects and higher training efficiency.

The soft-NMS algorithm is used in this study to help boost the recognition success rate of obstructed apples, and the comparison of recognition accuracy between NMS and soft-NMS algorithms is shown in Figure 5. The unidentified apple fruit is represented among them by the black spherical box.

In parallel, we carried out meticulous comparative studies in three challenging circumstances as an extremely thorough research project on apple-picking robots, as illustrated in Figure 6. Apples in complex environments such as unbagging, bagging, and nighttime were selected as the test objects, and the comparative experiments of apple fruit detection were carried out by using YOLOX-tiny, YOLOV5s, and EF-YOLOV5s detection networks. Results for average precision are shown in Figure 7.

The average identification precision of different methods is shown in Figure 7. The EF-YOLOV5 network can detect most apples, but only a few small targets that are far away can't. Because the bagged apples are irregular in shape and the white bags will reflect light, the precision of detection is obviously not as good as that of the unpacked apples. At night, the precision is worse than in the daytime because of the uneven illumination of the light source and the reflection of the apple surface. EF-YOLOV5s clearly outperforms the YOLOV5s and YOLOX-tiny networks in terms of precision. Especially, EF-YOLOV5s has a better detection effect on apples, which are seriously blocked. Although, in actual recognition, small targets in the distance will be unrecognizable, this does not affect the work of the picking robot. When pick-up robots actually work, they don't need to consider small targets in the distance.

The comparison of precision between EF-YOLOV5s networks and other networks is shown in Figure 8. From the peak value of the curve, it can be seen that the precision of EF-YOLOV5s network is better than that of YOLOX-tiny and YOLOV5s. From the rising speed of the curve, it can be seen that EF-YOLOV5s is obviously faster than the other two networks, which shows the advantage of the EF-YOLOV5s network, which is that it is easy to train.

Although the parameters of the improved EF-YOLOV5s network have increased, it has brought about a great improvement in precision and FPS. The YOLOX-tiny network, as one of the most effective lightweight networks at present, has better performance than YOLOV5s. The improved network EF-YOLOV5s is better than YOLOX-tiny in precision, recall, and FPS of detection. Compared with other lightweight networks, EF-YOLOV5s can accurately and quickly detect apples in complex environments with higher precision and speed, making it more suitable for deployment in apple-picking robots. Table 3 demonstrates how this network differs significantly from other networks.

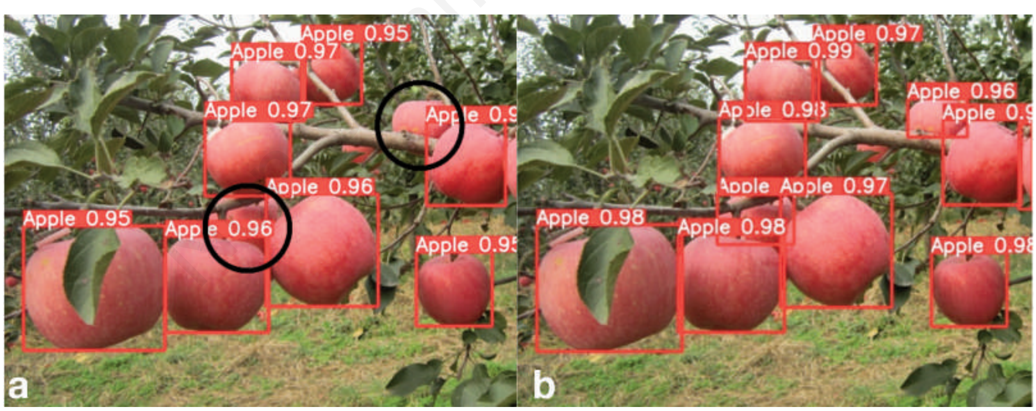

Figure 5. Comparison of the recognition of overlapping apples using non maximum suppression and soft non maximum suppression algorithm. **a**) Non maximum suppression; **b**) soft non maximum suppression.

Table 3. Comparison of performance between the EF-YOLOV5s network and other networks.

Model	AP	Precision	Recall	F1	Param（M）	FPS
YOLOV5s	94.36%	94.85%	92.32%	0.93	7.20	19
EF-YOLOV5s	98.81%	98.95%	96.45%	0.98	10.78	28
YOLOX-tiny	96.58%	96.89%	94.62%	0.94	5.03	21

AP, average precision; FPS, frame per second.

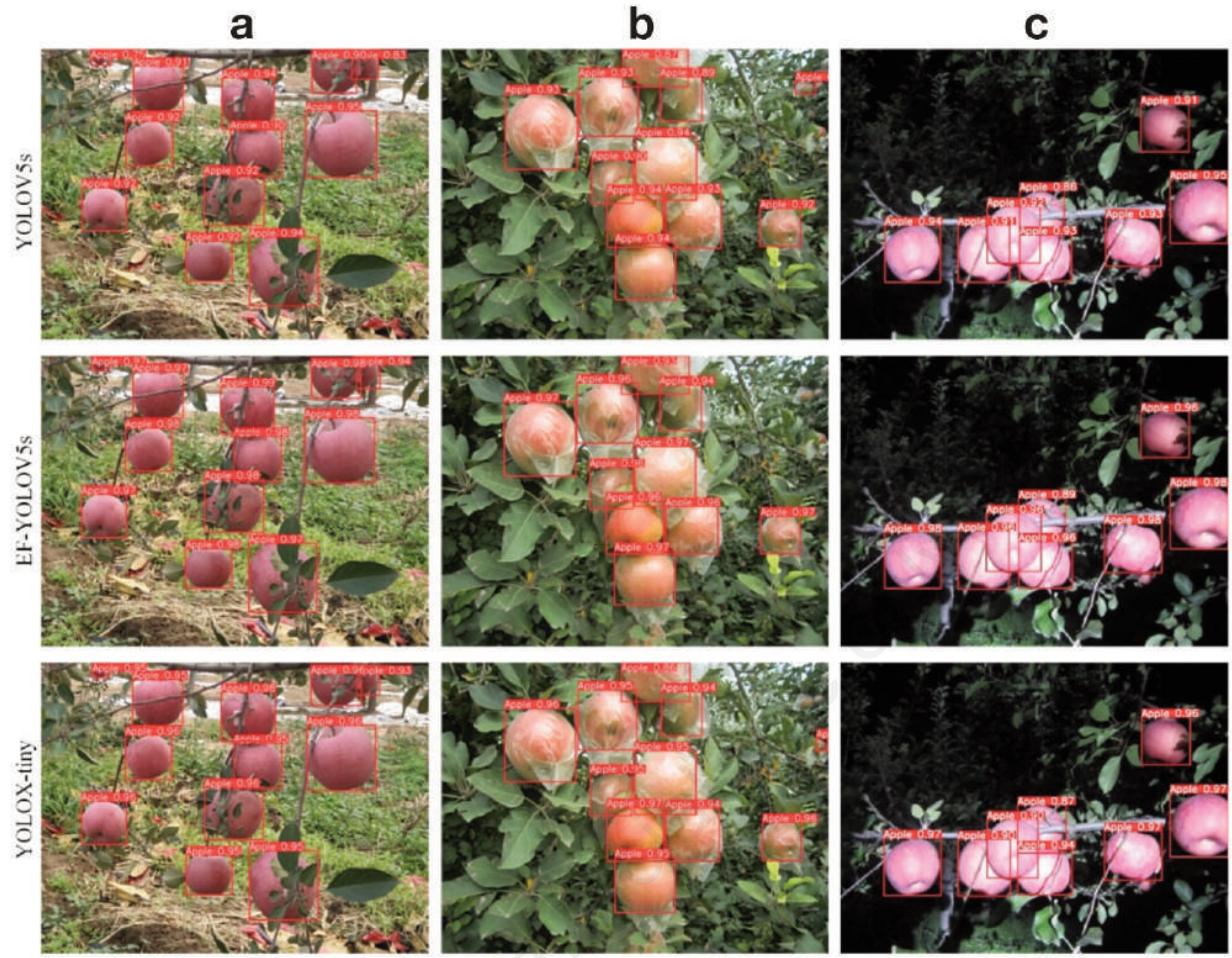

Figure 6. Comparative diagram of apple detection with different networks in different environments. **a**) Unbagged apples; **b**) bagged apples; B apples at night.

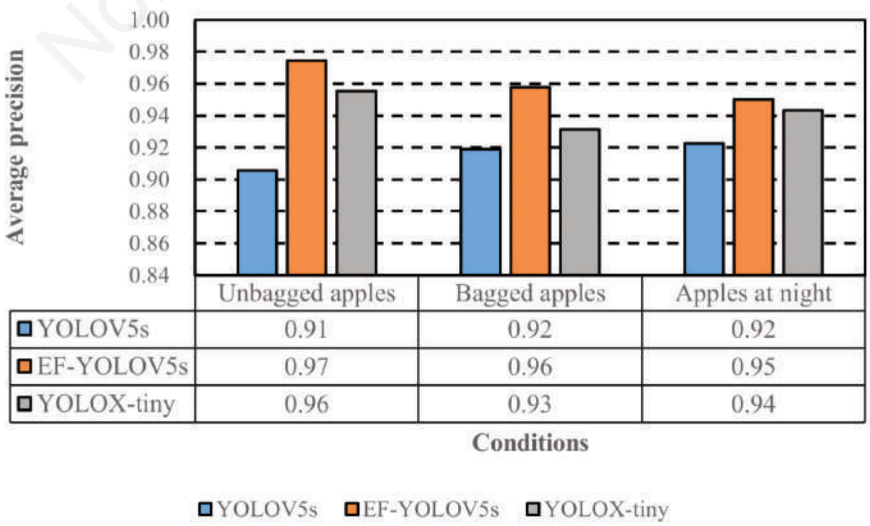

Figure 7. Comparison of average identification precision of different methods.

Simulation verification of apple picking sequence planning

According to the three-dimensional information of apple fruit recognition, a schematic diagram of the apple spatial network is established, as shown in Figure 9. Three methods, namely, picking sequence planning based on the improved DBSCAN algorithm (referred to as "proposed method" in the following table), sequence planning, and random planning, were used to carry out comparative experiments on apples in unbagged, night, and bagging environments (Karche *et al.*, 2022; Gangammanava *et al.*, 2021). Use Matlab to build 3D positions and make a schematic diagram of the picking sequence. Given the orchard's complex environment, the experiment does not include any other obstacles besides leaves and branches. Leaves are considered collision obstacles, while branches are non-collision obstacles. It is assumed that the robot arm moves at a speed of 100 mm/s according to the

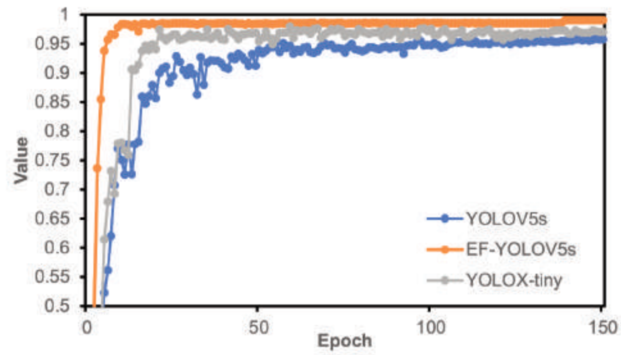

Figure 8. Comparison chart of the detectionprecision of different networks.

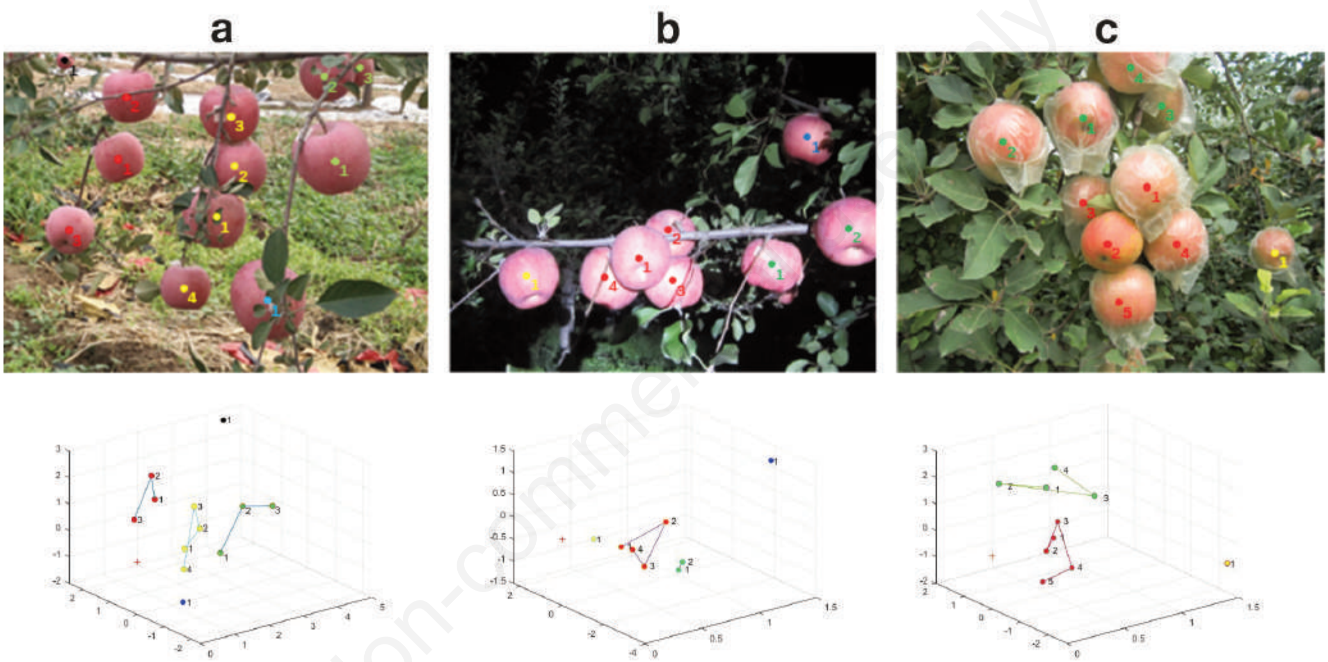

Figure 9. Schematic diagram of apple positioning and spatial network construction. a) unbagged apples; b) apples at night; c) bagged apples.

Table 4. Experimental comparison of different planning in complex environments.

Environment	Planning	Path length/mm	Total picking time/s	Single fruit picking time/s
Unbagged	Proposed	3262.35	42.43	3.86
	Sequential	3648.68	48.95	4.45
	Random	5624.95	57.49	5.23
Night	Proposed	4562.30	52.67	4.79
	Sequential	4896.26	60.45	5.50
	Random	5849.35	72.95	6.63
Bagged	Proposed	3195.46	40.86	3.71
	Sequential	3654.80	46.25	4.20
	Random	4815.49	50.47	4.59
Average	Proposed	3673.37	45.32	4.12
	Sequential	4066.58	51.89	4.72
	Random	5429.93	60.30	5.48

planned sequence to pick all the identified apples. The results of different picking sequence planning methods in different environments are shown in Table 4. In the simulation experiments of night and bagging, the success rate of picking in a night and bagging environment is lower than that in an unbagging environment due to the interference of night light and reflected light from white bags. Compared with sequential and random planning, the proposed method reduces the average picking time by 12.66 and 24.84 percentage points, respectively, and shortens the traversal path by 9.67 and 32.35 percentage points. These results show that, even for apples in a complex environment, the picking efficiency of the improved picking sequence planning method is obviously improved compared with the other two methods, and the picking time and path length are reduced.

Picking failure will occur in the experiment due to the branch blocking. According to the findings, the improved picking sequence planning may considerably enhance the picking efficiency of the apple-picking robot and limit the possibility of collision damage when the apple-picking robot performs multi-objective apple picking.

Apple picking experiment

In order to verify the effectiveness of the picking plan proposed in this study in actual apple picking, a four-degree-of-freedom single-arm apple-picking robot platform, as shown in Figure 10, is used for comparative experiments. Pick from different angles and positions to ensure the difference between each picking experiment. Limited by the growing and harvesting season and conditions of apples, the robot picks simulated apples in the laboratory. The experimental results obtained by using the improved DBSCAN clustering algorithm are shown in Figure 11. The experimental results of picking paths with different picking plans are shown in Figure 12, and the experimental results of picking in different positions for 10 times in the laboratory environment are shown in Figure 13.

In this study, the EF-YOLOV5s identification algorithm is used to provide the spatial location information for detecting apples, and Figure 11 shows the identified apple. Then, the improved DBSCAN clustering algorithm is used to automatically divide the apple into clusters with different densities, and then the picking planning of each apple fruit is calculated by combining the Gaussian weights of distance and significance level. By combining Gauss weights with the improved DBSCAN clustering algorithms, rational planning of the apple's picking sequence was achieved, significantly improving the picking efficiency. In a simulated environment, the proposed planning method will be experimentally compared with two other methods, as shown in Figure 12.

Figure 10. Apple picking robot.

Figure 11. Apple detection and clustering diagram in laboratory environment. a) Apple detection; b) clustering diagram.

Figure 12. Schematic diagram of picking trajectory under the proposed, sequential and random planning. a) Proposed planning; a) sequential planning; c) random planning.

In the laboratory, three distinct planning methods were used to carry out 10 groups of comparative experiments on picking sequence, and the result is shown in Figure 13.

During the actual experiment, sequential planning and random planning do not take into account the circumstances where the apple is covered up, which could lead to picking failure. However, the planning proposed in this study fully considers the occlusion. From the experimental results, the method proposed in this study has indeed improved the average picking success rate, reaching 94.8%. Compared with the other two methods, the method proposed in this study is superior to random planning in terms of average picking success rate and average picking time. Even though it occasionally necessitates a longer picking path in order to boost the picking success rate, the proposed method generally outperforms the other methods in terms of picking success rate and efficiency. To sum up, the experiments show that the apple identification network and picking planning method designed in this study play a significant role in improving the efficiency and success rate of agricultural apple-picking robots.

Conclusions

In order to solve the problem of low efficiency in apple picking in complex environments, this study proposes an apple-picking method combining target detection and picking sequence planning. Firstly, the EF-YOLOV5s target detection network is used to detect apples. Compared with the original YOLOV5s network, the improved network has obviously enhanced the efficiency of detection, with the precision of detection increasing by 4.3% and the speed increasing by 32%. Secondly, by introducing apple-picking sequence planning based on the improved DBSCAN clustering algorithm, this method effectively divides apples into clusters with different densities and improves the efficiency of the robot. Finally, the picking order of each apple is determined by combining the weights of Gaussian distance and significance level, which avoids the repeated identification and processing of information in the picking process and significantly improves the picking efficiency. Overall, compared with sequential planning and random planning, the success rate of the proposed planning has increased

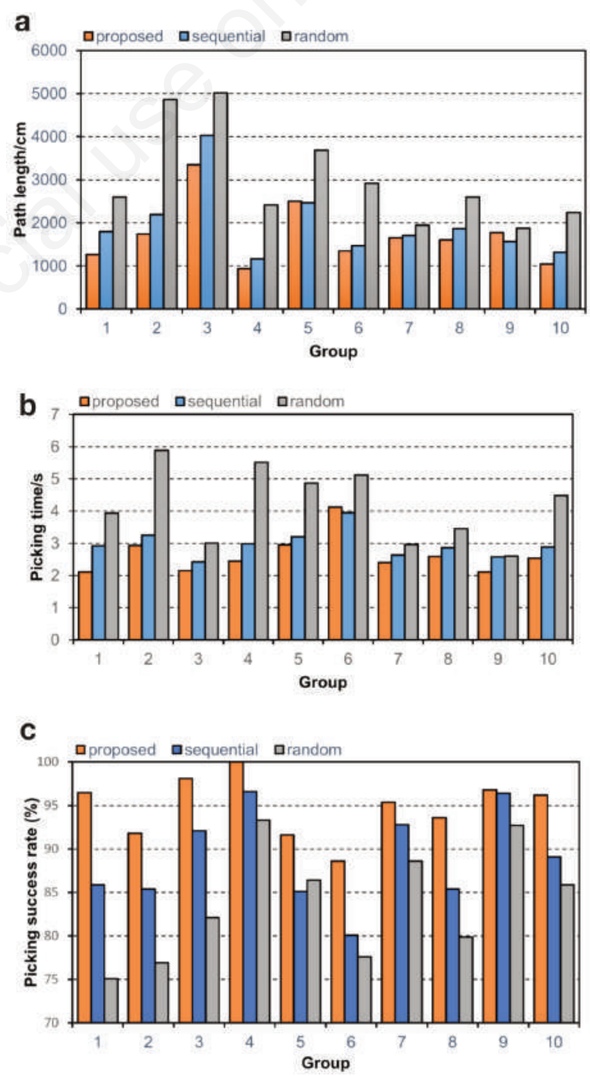

Figure 13. Comparative experiments on picking sequence. a) Average picking time of single apple; b) path length; c) picking success rate.

by 6.8% and 13.1%, respectively. In addition, the more targets that are detected, the greater the advantages of EF-YOLOV5s identification network. In order to accurately detect and increase harvest efficiency in the actual environment, this method can successfully address the issue of the significant quantity of information processing and reprocessing that robots must deal with.

References

Bodla N., Singh B., Chellappa R. 2017. Soft-NMS—improving object detection with one line of code. Proceedings of the IEEE International Conference on Computer Vision. pp. 5561-9.

Bu L., Chen C., Hu G. 2022. Design and evaluation of a robotic apple harvester using optimized picking patterns. Comput. Electron. Agric. 198:107092.

Gangammanavar H., Sen S. 2021. Stochastic dynamic linear programming: A sequential sampling algorithm for multistage stochastic linear programming. SIAM J. Optim. 31:2111-40.

Gao F., Fu L., Zhang X. 2020. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. Comput. Electron. Agric. 176:105634.

Han K., Xiao A., Wu E. 2021. Transformer in transformer. Advances in Neural Information Proces. Syst. 34:15908-19.

Hu G., Chen C., Chen J., Sun L. 2022. Simplified 4-DOF manipulator for rapid robotic apple harvesting. Comput. Electron. Agric. 199:107-77.

Ji W., Gao X., Xu B., Pan Y., Zhang Z., Zhao D. 2021. Apple target recognition method in complex environment based on improved YOLOv4. J. Food Process Eng. 44: e13866.

Ji W., Pan Y., Xu B., Wang J. 2022. A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. Agriculture. 12:856.

Ji W., Peng J., Xu B., Zhang T., 2023. Real-time detection of underwater river crab based on multi-scale pyramid fusion image enhancement and MobileCenterNet model. Comput. Electron. Agric. 204:107522.

Jia W., Tian Y., Luo R. 2020. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. Comput. Electron. Agric. 172:105380.

Karcher C.J. 2022. Logspace sequential quadratic programming for design optimization. AIAA J. 60:1471-81.

Li Y., Yuan G., Wen Y. 2022. EfficientFormer: Vision transformers at mobileNet speed. arXiv preprint arXiv. 2206:01191.

Schubert E., Sander J., Ester M. 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Trans. Database Syst. 42:1-21.

Swanepoel K.J. 1999. Cardinalities of k-distance sets in Minkowski spaces. Discrete Math. 197:759-67.

Sun T., Wang H.H., He D.J. 2018. Segmentation and picking sequence planning of clustered apples. Int. Agric. Eng. J. 27:309-17.

Tang Y., Zhou H., Wang H., Zhang Y. 2023a. Fruit detection and positioning technology for a Camellia oleifera C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. Expert Syst. Appl. 211:118573.

Tang Y., Qiu J., Zhang Y., Wu D., Cao Y., Zhao K., Zhu L. 2023b. Optimization strategies of fruit detection to overcome the challenge of unstructured background in field orchard environment: A review. Precis. Agric. 1-37.

Wu F., Yang Z., Mo X., Wu Z., Tang W., Duan J., Zou X., 2023. Detection and counting of banana bunches by integrating deep learning and classic image-processing algorithms. Comput. Electron. Agric. 209:107827.

Wu S., Li X., Wang X. 2020. IoU-aware single-stage object detector for accurate localization. Image Vis. Comput. 97:103911.

Wang W., Xu Z., Lu W. 2003. Determination of the spread parameter in the Gaussian kernel for classification and regression. Neurocomputing. 55:643-63.

Wang N., Joost W., Zhang F.S. 2016.Towards sustainable intensification of apple production in China-Yield gaps and nutrient use efficiency in apple farming systems. J. Integr. Agric. 15:716-25.

Wang D., Song H., He D. 2017. Research advance on vision system of apple picking robot. Trans. the Chin. Soc. Agric. Eng. 33:59-69.

Xu B., Cui X., Ji W., Yuan H., Wang J. 2023. Apple grading method design and implementation for automatic grader based on Improved YOLOV5. Agriculture. 13:124.

Yu X.J., Fan Z.M., Wang X.D. 2021. A lab-customized autonomous humanoid apple harvesting robot. Comput. Electr. Eng. 96:107459.

Zhang F. 2016. Design of apple picking robot based on machine vision and binocular distance measurement. Int. J. Simul. Syst. Sci. Technol. 17.

Zhang K., Lammers K., Chu P. 2021. System design and control of an apple harvesting robot. Mechatronics. 79:102644.

Zhao D.A., Lv J., Wei J. 2011. Design and control of an apple harvesting robot. Biosyst. Eng. 110:112-22.

Online supplementary material:

Figure S1. The pseudo-code in green is replaced with the one in red in soft non maximum suppression. We propose to revise the detection scores by scaling them as a linear or Gaussian function of overlap.

Figure S2. K-distance diagram.

Figure S3. Feature extraction heat map distribution in different environments. a) Original images; b) YOLOV5s; c) EF-YOLOV5s.