https://doi.org/10.31449/inf.v49i18.7918 Informatica **48** (2025) 191-204 **191**

Enhanced YOLOv5s-Based Apple Detection for Harvesting Robots Using a Multi-Rotated Box Algorithm

Wen Jian Feng[1], Yi Lin Zhao[1], Yongsheng Xie[2*], Linbing Wei[2]

[1]Guangxi Eco-engineering Vocational & Technical College Liuzhou, Guangxi 545004, China

[2]School of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin, 546199, China

E-mail: fwjian@163.com, silvia-0023@163.com, xieyongsheng@gxstnu.edu.com, 023505@163.com

[*]Corresponding author

Keywords: apple harvesting, robot, object detection, image processing techniques, navigation techniques

Recieved: December 27, 2024

The apple picking robot makes use of a number of technologies, one of which is the apple target identification algorithm. When it comes to automated apple picking, the robots' optical systems are crucial. Generally speaking, it finds ripe apples by taking photographs of its environment, processing them, and then analyzing the findings. The inability of traditional vision algorithms to process complex backdrops hinders the efficiency of harvesting robots. The continuous development and refining of the CNN have led to a substantial improvement in its efficacy in target identification during the last several years. The current crop of apple recognition algorithms struggles to tell the difference between partially obscured apples and ones entirely concealed by tree branches. Direct use of the algorithm endangers the harvesting robot's mechanical arm, apples, as well as gripping end-effector. In response to this real-world issue, we provide a lightweight apple targets identification approach for picking robots based on enhanced YOLOv5s. This method can automatically identify which apples in an apple tree picture are graspable and which ones are not. This method is able to circumvent the impact of light transformation, in contrast to the conventional segmentation approach. When there is a lot of resemblance between the fruit and the backdrop, though, it becomes more challenging to get strong recognition results. With a recall rate of 98%, a detection speed of 47 f/s, and a mAP (mean Average Precision) of apple detection of 98.13%, the findings demonstrate that the YOLO v5 network has perfect properties. The YOLO v5 is able to simultaneously fulfill the accuracy and speed criteria of apple identification, in contrast to more conventional network models like Faster R-CNN and YOLO v4. The experiment culminates with the employment of the apple-harvesting robot that the researcher developed themselves. Results demonstrate that the robot has a harvesting success rate of 99.2% in 9.5 seconds. Because of these improvements in accuracy and speed, the suggested apple detecting approach is preferable. Innovative concepts for intelligent agriculture and apple-harvesting robotics may emerge from it.

Povzetek: Prispevek predstavlja izboljšan model YOLOv5s z zasukanim okvirjem za kvalitetno in hitro zaznavo jabolk v realnem času v neurejenih sadovnjakih za uporabo na robotskih obiralcih.

1 Introduction

An ever-increasing quantity of undeveloped rural property has contributed to China's "rural labor shortage" in recent years, as the country's youth have flocked to urban centers in search of employment. A growing need for agricultural robots was caused by China's rapidly aging population and the decline of the country's agricultural workforce. As farming gear and automation technologies have advanced at a fast pace, so too have agricultural robots [1]. Because of the dramatic drop in agricultural manpower, robotic harvesting is now a must for the growth of the apple sector. It is still very difficult to use robots to pick apples in unstructured orchards [2]. The modern apple industry has progressed thanks to the standard apple orchard approach, which promotes the growing of apple trees with a spindle form. While picking fruit from trees, the robotic arm might snag on branches or other impediments if an adequate obstacle avoidance route isn't in place, which could ruin the fruit, harm the tree, or both [3]. Optimal scheduling, selection allowing, increased operation efficiency, and decreased labor expenses may all be achieved with robotic harvesting. Because of these features, robot harvesters help farmers make the most of their harvests [4]. Horticulture and farming are two of the cornerstones of every economy. The atomization of agriculture has been aided by a number of new technologies that have emerged as a consequence of recent technological developments. Apple picking is a common agricultural task in Italy that is mostly reliant on human labor but is amenable to automation because to the prevalence of apple orchards [5]. The most popular alternative to the inefficient and costly human apple harvester is a robot that picks the apples. Recent studies on apple-picking robots have shown promising results in the lab, but their poor apple-positioning performance makes them impractical for use in real orchards. There is currently no widely-used, precise technique for placing apples for an apple-picking robot. Some orchards were able to achieve satisfactory results using positioning approaches that used detection-based deep learning [6]. The process of citrus fruit picking requires a lot of time and effort. Labor expenditures are increasing due to the world's aging population. Hence, both the corporate and academic sectors have shown significant interest in the citrus-harvesting robot [7]. The drastic reduction in agricultural manpower has made robot harvesting an absolute need in the apple business.

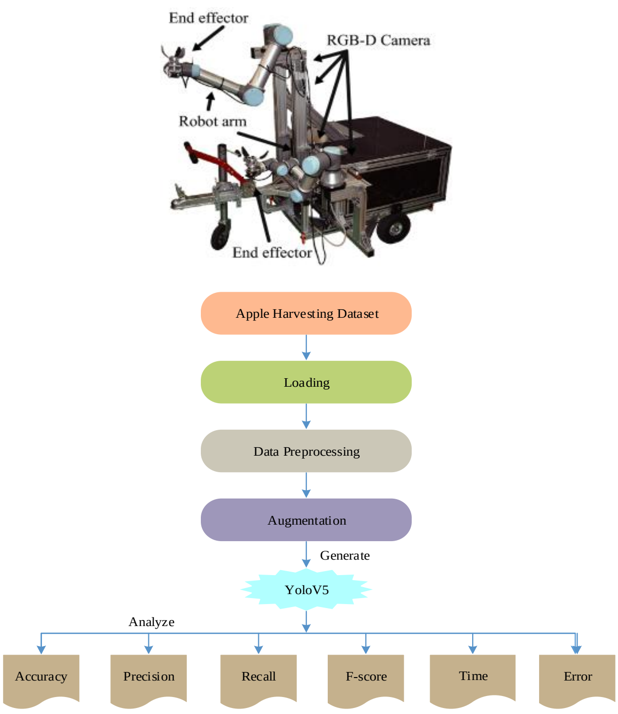


Figure1: General model for apple harvesting robot

A lot of people are interested in harvesting robots because of the potential benefits of using multiple robotic arms to increase industrial applications and operational efficiency. Despite significant advancements in the field, multi-arm harvesting robots have not yet seen widespread use in orchard production due to issues with operating efficiency and fruit positioning accuracy [8]. The angle at which the end controller holds the apple and the method it uses to spin apart the apples have a significant influence on the picking outcome of the apple harvesting robot [9]. Research on automating and intelligent systems via the integration of machine vision and image processing has recently attracted a lot of attention. Intelligent management improves the effectiveness of planting, harvesting, as well as picking fruits and vegetables while simultaneously reducing the intensity and amount of work required. Before feature extraction can take place, the dataset must be checked for errors, low-quality pictures and labels removed, image sizes adjusted, color spaces transformed, denoising applied, and normalization applied [10]. Feature extraction will improve along with the dataset's size and variety. An enhanced apple picture segmentation method built on the Deeplabv3 framework is presented here as AppleDNet. Its goal is to help the apple harvesting robot choose apples appropriately by differentiating them from their complex natural environment backdrops. You may get better and quicker segmentation results using the well-known Deeplabv3 approach [11]. It uses Atrous convolution, Depthwise separated convolution, and transfer learning. Due to the worsening labor shortage that has persisted since the rapid outbreak, fruit picking has become more problematic. Robotic harvesting has focused on the autonomous collection of fruits, such as apples. When it comes to apple orchards, however, current robots aren't very useful because to their inefficient robotic grippers [12]. The collective demand for agricultural harvests necessitates the use of more efficient equipment and the guarantee of a more affordable energy supply. Robots that collect fruit are becoming more common as AI makes it easier and faster to select fruits, which means more orange fruit exports. However, an efficient and inexpensive energy source is required to guarantee the fruit harvesting robot's successful operation [13]. Due to their stringent demand for well-manicured canopies to work consistently, specific apple harvesting robots have not yet found widespread implementation [14]. Picking fruit by hand is a tedious and time-consuming process. Despite their high prices and limited efficiency, robots can automate fruit harvesting to a large extent, drastically cutting down on manpower needs [15]. Robots can now accurately perceive depth, analyze scenes, and identify objects in real time thanks to AI-powered visual systems. From healthcare gadgets and driverless cars to drones and factory robots, these skills are crucial.

2 Related works

A new apple-harvesting robot was introduced and its field assessment detailed in [16]. The primary parts of the robot include a vision component that was specifically created for it, a manipulator with four degrees of freedom, a soft end effector that is upgraded and based on vacuum technology, and a dropping/catching element that can accept and transfer fruits that have been collected. For effective, automated apple harvesting in difficult orchard settings, software algorithms are created to allow hardware components to work together synergistically. In order to accomplish accurate localization and strong apple recognition, a new perception method is created by combining algorithms for processing and analyzing images with modified triangulation. New planning and control systems lead the robot to its intended destinations. Field tests were conducted in two apple orchards that had various tree layouts with leaf conditions to assess the robotic system's effectiveness. The robot had a harvest success rate of 82.4% even while working with new, well-pruned trees in the orchard. Despite the dense foliage and closely packed branches of an older orchard, the robot managed a 65.2% success rate.

Using the standard genetic algorithm as a foundation, the researchers of [17] provide a new encoding technique and an improved double-encoding GA. Although the crossover link employs that new encoding method, the mutation link keeps using the route node sequencing encoding methodology. It is possible to put the selection process after the mutation and speed up convergence by doing merging sorting while elitist selecting the parent, crossover, as well as mutation populations prior to selection. We construct a three-dimensional route for the apple-harvesting robot using the enhanced GA; along the way, we consist of an adaptive adjustment mechanism. The results of the experimental simulations demonstrate that the mobile robot's enhanced GA-based three-dimensional route planning has the ability to reduce the number of pathways and loops while effectively meeting the robot's operating needs.

By studying the current apple-picking behavior, the authors of [18] create a novel design for the flexible three-fingered end-effector. We observe an alternate method of selecting apples by encircling the end-effector and lowering it first. Additionally, there are two different ways to hold an apple with three fingers: one is horizontal, such as a ruler, and the other is vertical, such as your fingers. Researching the effects of different apple plucking methods led to the development of a simulation model including branches, stems, and apples. The ideal angle of circular movement among the force needed to make it happen may be determined by conducting simulations that take into consideration all the potential limitations, such as the angle that separates the apple stem as well as the vertical direction, the root impulse, the apple's rate of motion, and so on. At last, we experimented with several apple-picking methods using the bendable three-fingered end-effector. Using the circular-pull-down movements separation approach, the trial results showed that an angle between 15° and 20° was optimal for apple selection.

Describes a way to pick apples with both hands at the same time while concentrating on solving problems with eyesight, posture, and dual-arm stability [19]. The article begins by outlining the robot's software and hardware systems, their integration, procedures, and control structure. It then moves on to discuss the systems themselves. In addition, the research employs a multi-task network framework to identify and locate fruits by combining a dual-vision perception method with a frustum-based fruit localization strategy. Lastly, a multi-arm work planning approach using evolutionary algorithms is used to optimize the goal harvesting frequency for each limb, leading to greater teamwork. An orchard was the site of many field experiments designed to determine the robot system's overall efficacy. The total success rate of the robot system in field testing was 76.77%, indicating good performance. On average, it took 7.29 seconds to pluck fruit and only dealt 5.56% damage.

The authors of [20] propose a YOLOv5-RACF algorithm that can detect apples and determine their diameter as a solution to the problem of automated apple harvesting in orchards. This kind of software might automate apple sorting by directing the robot's navigation, robotic arm posture, locomotion framework, Lidar mapping, and gripping actions via the operating system. This trial took place in a genuine orchard. Our results for apple recognition using the algorithm model were 91.02% accuracy (mAP@0.5:0.95) and 99.88% accuracy (the mAP@0.5:0.95). Using a measurement accuracy of 1-3 mm, finding the apple's diameter would take a mere 0.13 seconds. The robot typically requires nine seconds to get an apple and put it back where it came from. In real-world agricultural settings, these results demonstrate the system's dependability and effectiveness.

If you're selecting apples from the spindle-shaped trees that are common in modern apple orchards, this study suggests a three-dimensional path-planning approach to avoid these problems utilizing full-field fruit avoidance. Using the free spindle, raised spindle, and thin spindle as a starting point, a three-dimensional spatial representation of fruit tree branches was built. These tree topologies are often seen in apple trees. Creating a setting representation obstacle map based on grids was the subsequent stage for the apple tree models. The next stage was to enhance the original pheromones by dispersing them in a non-uniform manner using the first ant colony approach. By incorporating a biomimetic optimization strategy into the antenna system of the beetle and decreasing the need for pheromone updates, they improved both the stability and the speed of route finding. We used a cubic B-spline curve to smooth out the robotic arm's anticipated path, making its harvesting technique more successful and reducing the likelihood of unexpected halt or twists. An improved ant colony optimization algorithm was used to simulate three types of spindle-shaped apple trees for the purpose of three-dimensional route planning in order to avoid obstacles. Based on the findings, free-spindle-shaped trees had a 96% success percentage, high-spindle-shaped trees 86%, and slender-spindle-shaped trees 92% when it came to obstacle avoidance route planning.

In [22], a robotic grasping device was introduced that integrates tactile detecting, deep learning, with soft robots.

Limit the amount of mechanical harm to fruits that you can. Robots equipped with fin-ray fingers, integrated tactile sensor arrays, and special perception algorithms can detect and deal with branch interference throughout harvesting. A strong strategy for limiting interference and a test validation success rate of 83.3-87.0% have been shown in relation to understanding status identification. The proposed grasping technique has potential for usage in a variety of robotic grasping tasks where the handling of undesirable foreign items is required.

Agricultural combine harvesting machinery often pauses between rows of apple trees while working in single master-slave navigation setting, as described in [23]. Depending on the distances estimated from ground-based GPS stations, the transportation as well as pickup robots might switch between various navigation modes. A cloth-simulated filter with a random sample agreement mechanism were used to produce the inter-row waypoints. While the turn waypoints for the master were determined using a kinematic model, those for the slave were manually picked using GNSS data. At last, they accomplished master-slave navigation by ground head master-slave command routing by relentlessly pursuing these waypoints using a pure pursuit algorithm. The results of the testing demonstrate that the robot satisfies the requirements for robot orchard communication, as it can communicate with an information loss rate of fewer than 1.2% despite being near 50 meters of an orchard row. The master-slave robot satisfies the requirements of cooperative orchard harvesting by completing sequential pauses made possible by its follow navigation abilities.

In [24], a new approach to apple identification in orchards in Kashmir was shown. It comprised YOLOv8's deep learning algorithm, an Apple Harvester Robot, and a RealSense camera. The proposed method aims to enhance the precision and effectiveness of apple harvesting by considering the unique challenges posed by different lighting conditions, vegetation, and apple varieties, shapes, and colors. The state-of-the-art object identification method known as YOLOv8 is used to locate apples in their natural habitat. To ensure that the YOLOv8 algorithm can accurately recognize apples in their entire splendor regardless of lighting or weather, they trained it using a vast collection of annotated photographs shot in Kashmir orchards. They used bounding boxes to not only identify apples, but also to locate their precise position, so the robot's manipulator could treat the fruit with care. Their proposed approach has, on average, yielded 91.2% success. They achieved a recall of 98% and an accuracy of 93.9%. Every classifier in the dataset was correct about one instance's prediction.

By outlining a plan for improving the design of configurations with two manipulators, the researchers of [25] met the R&D demands of a robot that could efficiently pick apples. The "high spindle," a typical Chinese tree form, was used to demonstrate their argument. From the characteristics of the geographical placement of fruits under a typical dwarf and close-planted canopy, a three-degree range of motion, two types of vertically synchronous processes, and a Cartesian coordinate dual-manipulator were constructed. The plucking robotic arm can adapt its movement to the height of the tree thanks to its two-stage telescopic components, that may be powered by gas or electricity. A model for optimizing the critical configuration parameters is developed through integrating a quantitative account of the operational setup to the dual-manipulator configuration parameters. We provide a technique for comprehensive assessment of the CRITIC-TOPSIS integrated dual-manipulator system. There should be 1119.3 mm and 39.4° in the upper telescopic portion of the dual-manipulator, and 898.7 mm and 26° and 755.3 mm in the lower telescopic section, respectively. In addition, you need to find the exact distance between the tree trunk's center and its attached base of the highest frame.

Table 1: Comparison of the proposed and existing methods

Reference	Method	Result	Remarks
[16]	Automated apple harvesting two apple orchards	65.2% success rate	The primary parts of the robot include a vision component that was specifically created for it, a manipulator with four degrees of freedom, a soft end effector that is upgraded and based on vacuum technology, and a dropping/catching element that can accept and transfer fruits that have been collected.
[17]	Standard genetic algorithm improved double-encoding GA	75% accuracy	We construct a three-dimensional route for the apple-harvesting robot using the enhanced GA; along the way, we consist of an adaptive adjustment mechanism.

[18]	We construct a three-dimensional route for the apple-harvesting robot	enhanced GA; along the way, we consist of an adaptive adjustment mechanism; 78% accuracy	The ideal angle of circular movement among the force needed to make it happen may be determined by conducting simulations that take into consideration all the potential limitations, such as the angle that separates the apple stem as well as the vertical direction, the root impulse, the apple's rate of motion, and so on.
[19]	Dual-Vision Perception Method	The total success rate of the robot system in field testing was 76.77%, indicating good performance. On average, it took 7.29 seconds to pluck fruit and only dealt 5.56% damage.	An orchard was the site of many field experiments designed to determine the robot system's overall efficacy.
[20]	YOLOv5-RACF algorithm	Our results for apple recognition using the algorithm model were 91.02% accuracy (mAP@0.5:0.95) and 99.88% accuracy (the mAP@0.5:0.95). Using a measurement accuracy of 1-3 mm, finding the apple's diameter would take a mere 0.13 seconds. The robot typically requires nine seconds to get an apple and put it back where it came from. In real-world agricultural settings, these results demonstrate the system's dependability and effectiveness.	This kind of software might automate apple sorting by directing the robot's navigation, robotic arm posture, locomotion framework, Lidar mapping, and gripping actions via the operating system.
[21]	We used a cubic B-spline curve to smooth out the robotic arm's anticipated path, making its harvesting technique more successful and reducing the likelihood of unexpected halt or twists.	Based on the findings, free-spindle-shaped trees had a 96% success percentage, high-spindle-shaped trees 86%, and slender-spindle-shaped trees 92% when it came to obstacle avoidance route planning.	The next stage was to enhance the original pheromones by dispersing them in a non-uniform manner using the first ant colony approach. By incorporating a biomimetic optimization strategy into the antenna system of the beetle and decreasing the need for pheromone updates, they improved both the stability and the speed of route finding
[22]	Robotic Grasping Device	A strong strategy for limiting interference and a test validation success rate of 83.3-87.0% have been	Robots equipped with fin-ray fingers, integrated tactile sensor arrays, and special perception algorithms can detect and deal

		shown in relation to understanding status identification. The proposed grasping technique has potential for usage in a variety of robotic grasping tasks where the handling of undesirable foreign items is required.	with branch interference throughout harvesting.
[23]	Agricultural combine harvesting machinery	The results of the testing demonstrate that the robot satisfies the requirements for robot orchard communication, as it can communicate with an information loss rate of fewer than 1.2% despite being near 50 meters of an orchard row. The master-slave robot satisfies the requirements of cooperative orchard harvesting by completing sequential pauses made possible by its follow navigation abilities.	A cloth-simulated filter with a random sample agreement mechanism were used to produce the inter-row waypoints. While the turn waypoints for the master were determined using a kinematic model, those for the slave were manually picked using GNSS data
[24]	Apple identification in orchards in Kashmir	Their proposed approach has, on average, yielded 91.2% success. They achieved a recall of 98% and an accuracy of 93.9%. Every classifier in the dataset was correct about one instance's prediction.	The state-of-the-art object identification method known as YOLOv8 is used to locate apples in their natural habitat.
[25]	CRITIC- TOPSIS integrated dual-manipulator system	There should be 1119.3 mm and 39.4° in the upper telescopic portion of the dual-manipulator, and 898.7 mm and 26° and 755.3 mm in the lower telescopic section, respectively. In addition, you need to find the exact distance between the tree trunk's center and its attached base of the highest frame.	From the characteristics of the geographical placement of fruits under a typical dwarf and close-planted canopy, a three-degree range of motion, two types of vertically synchronous processes, and a Cartesian coordinate dual-manipulator were constructed.

3 Research method

In order for robots to gather apples autonomously, real-time recognition of apples in their natural environments is essential. This technology is also crucial for predicting orchard productivity and managing fines. The detection performance of YOLO v5 is superior than that of the existing popular methods. Nevertheless, the robot's picking efficiency will be diminished by YOLO v5's intricate network architecture. To solve the problem of automated apple harvesting in orchards, we provide an Improved Multi-Rotated Box Algorithm for apple identification and diameter calculation. An apple-harvesting robot that utilizes many technologies is suggested in this research. Radar, inertial measurement units (IMUs), and vision sensors are among the many sensors used by the robot. When the robot's sensors share data, it may reach a higher degree of automation, allowing it to harvest crops without human intervention. The robot's arm position, navigation, Lidar mapping, and locomotion structure are all controlled by this software via the robot operating system (ROS); it then uses these to achieve autonomous apple harvesting and placing. An authentic orchard setting served as the testing ground. In terms of speed and accuracy, Des-YOLO v5 outperforms more conventional network models like YOLO v5 or Faster R-CNN when it comes to apple identification. One apple's diameter could be accurately measured within a margin of error of 1-3 mm in about 0.12 seconds. Picking up an apple and getting back into its starting position takes the robot, on average, nine seconds. The system's efficiency and dependability in actual agricultural contexts are shown by these outcomes.

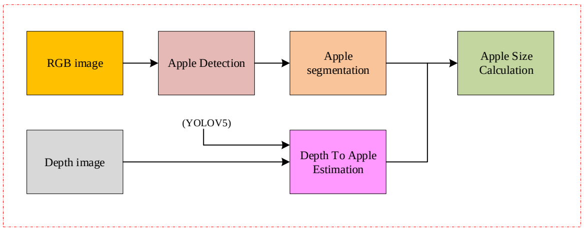


Figure 2: Flow model for apple harvesting

3.1 Dataset description

The data used in this research came from two places: the internet and the actual circumstances in the orchard. We chose 2044 high-quality photographs of apples taken in a variety of settings from varied perspectives and in a range of lighting conditions from various web sources. We used the JPG format to save these pictures. In July 2024, more data was gathered from the orchard base at Jilin Farming Institute in Changchun, Jilin Province, China, to enhance the dataset and fulfill the requirements of accurate identification in intricate agricultural situations. An iPhone 13 smartphone was used for data gathering, and it produced JPG photos having a resolution of 4096 × 3072 pixels. In a complicated orchard setting, 1026 photographs of orchard apples were captured from a variety of angles, sizes, and degrees of occlusion and backdrops to guarantee the images' unpredictability and diversity.

Table 1: The apple dataset in all its detail

Label	Original Image Data	Validation	Training	Augmented Image Data
goodapple	3122	994	8834	6945

3.2 Data augmentation

To further enhance the dataset and make the algorithm more robust, we divided it in a 9:1 ratio. This helped to boost recognition accuracy. Some of the enhancements are shown in Figure 3. Improvements to the image's brightness, histogram equalization for better recognition in varying lighting, vertical and horizontal image inversion for wider viewing angles, and the use of Gaussian and salt-and-pepper noise to simulate camera shake blur are all part of the process. The dataset now includes 9934 apple photos after augmentation. We decided to partition the dataset before augmenting it in order to lower the frequency of overfitting. Furthermore, in order to lessen the resemblance to the initial dataset, every augmentation makes use of two distinct data augmentation techniques concurrently. Furthermore, we tagged the original photographs by hand and gave them the moniker goodapple. Table 1 displays the specifics of the program. To train a YOLO neural network, additional practice sets are often required. To improve the network model's generalizability, it may be helpful to apply more training sets. This will help the neural network understand the apple picture attributes to a sufficient degree. On the other hand, gathering a plethora of training resources is a genuine challenge when it comes to limited material collecting capabilities. It is also challenging to fully extract the form properties of the fruit since apples have a unique growing posture and a severe overlap phenomenon. Hence, prior to YOLO training, the apple picture must be preprocessed. The initial information set is processed using Matlab in this research to obtain the impact of data improvement.

- To create additional training sets, the picture is either vertically or horizontally rotated at a predetermined angle, and its aspect ratio is altered.

- Image processing methods such as median filtering, histogram equalization, and saturation/hue adjustment improve the data.

- The Mosaic data enhancement approach is used to randomly trim four photos and combine them into one image for the purpose of training the model. This improves the model's capacity to generalize.

3.3 Apple localization

Before you can set up your D435i depth camera on a ROS system, you must install the packages and drivers. Using the depth camera to focus on the apple aim is the next step. Next, we'll put the trained YOLOv5 algorithm to work identifying apples. Pixel values ("min, y min") are shown in the bottom-left corner of the bounding box, whereas xmax and ymax are shown in the bottom-right corner. The center pixel's position (px, py) is retrieved when the apple focus has been identified. These may only be obtained by those who have subscribed to the subject on the camera's built-in features. Two sets of properties inherent to the camera are the focal length (fx, fy) with the image center coordinates (ppx, ppy). We can determine the two-dimensional positions of the apple (x, y) using the camera's orienting mechanism. When the camera's technology obtains the pixel location at the center of the apple through subscription (z), it combines these two-dimensional coordinates with the apple's coordinates to produce its three-dimensional geographical position. Using the centers of the pixels and the following calculations, we can find the apple's three-dimensional location:

$$\begin{aligned} px &= (xmin + xmax)/2 \\ py &= (ymin + ymax)/2 \\ x &= (px - ppx)/fx \times z \\ y &= (py - ppy)/fy \times z \end{aligned} \tag{1}$$

The final set of coordinates is obtained by averaging the three-dimensional geographic position of the apple over 10 frames, starting with the current frame. By reducing the effect of noise in each frame and smoothing out short-term measuring changes, this approach improves localization accuracy and makes the results more reliable. For computations, the following equations are utilized:

$$\begin{aligned} x &= \frac{1}{10}\sum_{1=t-9}^{t} x_1 \\ \bar{y} &= \frac{1}{10}\sum_{1-t-9}^{t} y_1 \\ \bar{z} &= \frac{1}{10}\sum_{1-t-9}^{t} z_t \end{aligned} \tag{2}$$

where x ¯, ¯y, and ¯z Averaging the apples organizes—which are saved as x_i, y_i, and z_i in each frame—results in, where x_i is the index of the current frames.

By modifying YOLOv5, we were able to pinpoint Apple targets. The newest network in the YOLO series, YOLOv5, offers better detection accuracy than its predecessors (YOLOv4) and a smaller model that saves processing resources, creating it perfect for mobile positioning.

3.4 Improved multi-rotated box algorithm

In order to find the ships' targets, we obtained the convolutional feature maps and then placed the rotated previous boxes on each feature map point. Figure 2 shows the blue-colored rotated preceding boxes. We may anticipate the target's class name and position data in different orientations using the rotating preceding boxes. Before calculating the loss during network training, we get the positive sample information by matching the previous boxes with the ground-truth bounding boxes.

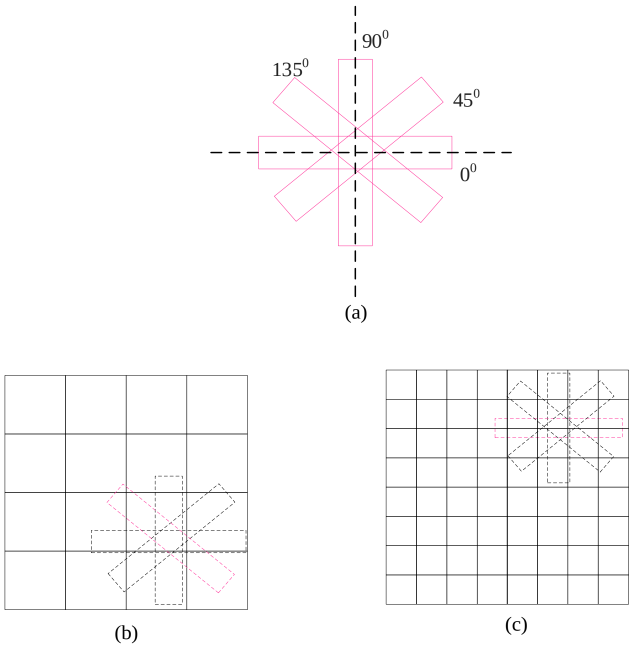


Figure 3

We input 600 × 600-pixel pictures into the network after dividing the large-scale image. The previous box's dimensions are {32, 64,128,256} {32,64,128,256}, and its aspect ratio is 5:1, yielding {71,142,284,568} {71,142,284,568} for the long side h. The box's step size is 16. The multi-scale targets may be more accurately detected with the help of multi-scale previous boxes. When the ship's size is tiny, it is difficult to determine the head or tail direction, therefore the rotation angle is set as {0, 45,90,135} {0,45,90,135}. Figure 3 displays the earlier boxes that have been rotated by one scale. These settings are selected according to the network architecture and the features of the ships that are targets in the dataset.

Rotated box matching

When selecting positive samples for training, implementing nonmaximum suppression in evaluation, and determining whether it is the correct detection in evaluation, rotated box matching is an essential component of target detection. Here we'll pretend that two boxes, labeled box _1 and box _2, have been rotated. To capture the box, we use a matrix that contains only zeros outside of the box region. By adding up the matrix components, we can calculate the box's size. The entries for the two matrix are I_1 and I_2. Box x_2's area and box _1's area are both documented as Area_1 with Area 2, respectively. We get the overlap zone by multiplying I_1 and I_2 together. Matrix I_3 is the output, and Area _n is the total of the components that make up the overlap area. So, we can determine the degree to which the two boxes match by

$$\text{matching degree} = \frac{\text{Area}_{\cap}}{\text{Area}_{U}} = \frac{\text{Area}_{\cap}}{\text{Area}_1 + \text{Area}_2 - \text{Area}_n} \quad (3)$$

It will take a lot of computation to match the rotating boxes because of the calculation among each preceding box. Therefore, we need to focus our selecting efforts first. After deciding on a range of values for t_c, we measure the distance among the two centers d_c=‖(x_1^c,y_1^c)-(x_2^c,y_2^c)‖, where (x_1^c,y_1^c) and (x_2^c,y_2^c) are the center point coordinates of the two boxes. If d_c<t_c, For the purpose of determining the matching degree, we choose a preceding box whose angle is closest to that of the ground-truth box. The calculating amount is drastically cut by using these procedures.

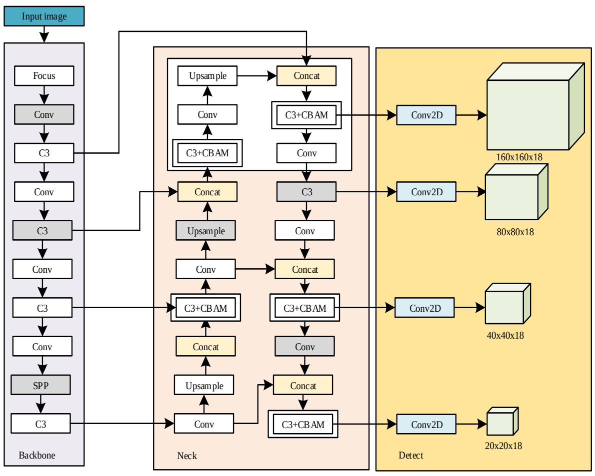


Figure 4: YoloV5 modified model

3.5 Improvement of fusion feature layer

The target detection network's identification performance may be greatly enhanced by merging feature maps of varying sizes. A more discriminating feature than the input features is what feature fusion is all about. Combining characteristics obtained from pictures creates this feature. In addition to greater resolution, the lower-level feature map includes more exact location data and detailed item descriptions. The convolutional layer did not extract enough features, which leads to a low-level feature map with poor semantics and extraneous noise. Although the high-level feature map contains a wealth of semantic information, the feature map precision is poor, and the capacity to perceive picture details is rather inadequate. Improving the model's detection performance relies on the successful integration of low-level and high-level information.

3.6 Apple size calculation method

In order to plan for harvesting labor, fruit box needs, and storage space, growers rely on precise estimates of fruit size and quantity. Classification of quality according to various sizes is also made possible by it. Finding out how big apples are will help autonomous harvesting robots adjust the opening and closing angles of their grippers in future research, which in turn reduces mechanical damage when picking apples. Importantly, segmenting the apple targets is the first step before getting the apple sizes. A lot of people have put a lot of time and energy into studying target segmentation. An Enhanced Multi-Rotated Box Algorithm was used in this study. In order to plan for harvesting labor, fruit box needs, and storage space, growers rely on precise estimates of fruit size and quantity. Classification of quality according to various sizes is also made possible by it. Finding out how big apples are will help autonomous harvesting robots adjust the opening and closing angles of their grippers in future research, which in turn reduces mechanical damage when picking apples. Importantly, segmenting the apple targets is the first step before getting the apple sizes.

Here is the representation of the circle's equation:

$$(x-a)^2 + (y-b)^2 = r^2 \quad (4)$$

where (a, b) is the point where the two lines meet, besides r is its radius.

A minimum of three points are required to establish a circle. So, consider these (x_1, y_1), (x_2, y_2), and (x_3, y_3) The structure that follows of linear equations may be used to reach the solution:

$$AX = B$$
$$A = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix}, B = \begin{bmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ x_3^2 + y_3^2 \end{bmatrix} \quad (5)$$

The answer X, obtained by solving the aforementioned system of linear formulas, is given by:

$$X = \begin{bmatrix} D_x \\ D_y \\ C \end{bmatrix} \quad (6)$$

After that, we may express the center (a, b) besides radius r as follows:

$$a = \frac{D_x}{2}, b = \frac{D_y}{2}, r = \sqrt{C + a^2 + b^2} \quad (7)$$

For each point (x_i, y_i), The following is how we get its distance from the fitted circle's center (a, b):

$$d_1 = \sqrt{(x_1 - a)^2 + (y_1 - b)^2} \tag{8}$$

If |d_i-r|<ϵ, at this stage, we believe the point to be a viable fitting point. We employ a bespoke nonlinear function for exact fitting after repeating the procedure 1000 times. Points with activation levels higher than 0.5 are retained by this function, which allocates distinct weights to fitted points and outliers. Let me give you the formula:

$$Y = \frac{1}{1+e^{a(d_1-r)}} \tag{9}$$

where d_i thus, where d is the distance from every point to the center of the circle, r is the radius of the corresponding circle, while an is the steepness of the function.

3.7 Model training and tuning

The procedure for training and fine-tuning the model is described in this section. Partitioning the training dataset is the first topic covered, followed by the platforms and models used for training. After these prerequisites were met, we set up a performance baseline and fine-tuned it from there. When a machine learning algorithm is trained using data, this procedure is called model training. Conversely, hyperparameter tuning is choosing the combination of hyperparameters that yields the greatest effect on the validation set. Importantly, hyperparameters are not learnt from the data but rather defined before training begins. Critical to the operation and output of a machine learning model are these configuration options, called hyperparameters. Model training is concerned with getting the model to perform well on the training information, while hyperparameter tuning is all about getting the hyperparameters just right so the model can perform its best on new, unknown data.

3.8 The harvesting robot's navigation method integration

As seen in Figure 8, two primary communication channels are available in ROS. A publisher disseminates messages to certain topics on a regular basis according to the message transmission model. The subject is the medium via which the publisher along with subscribers convey messages to one another. In order to get the messages that the publisher publishes, users must subscribe to the subject. Thanks to this technique, nodes may communicate with one another in an asynchronous fashion, meaning that publishers and subscribers can both send and receive messages at any time. This study primarily uses it for collecting data from sensors and providing status updates, tasks that do not need quick answers. Asynchronous communication, however, does not ensure that subscribers will get all messages. A customer may also submit a service request as another way to communicate. After the service gets the request, it sends it on to the server. The client is notified of the result of the request by the server, which processes it and delivers a response. The client sends the request and then waits for the server to respond in order to ensure the service request was properly executed; this is an example of synchronous communication. The overall testing tests of the harvesting robot make use of this synchronous communication approach to coordinate the operations of numerous components. This system makes sure that everything runs well, which allows autonomous navigation, identification while localization, and harvesting to be done efficiently. For instance, when a client asks for directions to a certain place, the server plans the route and then gives the navigation result. The server performs the client's request and provides the result when the image processing component is asked to find and identify apples. Just like a robotic arm, when a client asks it to harvest, the server organizes the arm's joint motions to accomplish the task and gives a response showing whether or not it was successful.

4 Evaluation metrics

In this work, we employ a number of metrics for assessing models: GFLOPS, model variable count, mean average precision, recall, as well as precision. Two measures of a test's efficacy compare the ratio of expected positive testers to the proportion of real positive samples; these are recall (R) and precision (P). The mAP is just the sum of the areas under the recall PR and accuracy PR curves, expressed simply. The total amount of parameters and the computing cost of the model are also provided. The second unit is gigaflops, which stands for giga computations performed in a second. The research assessed the efficacy of the trained apple objects recognition algorithm using objective evaluation metrics including recall (2), precision (1), mAP (3), and F1 score (4). Here are the equations for the calculation:

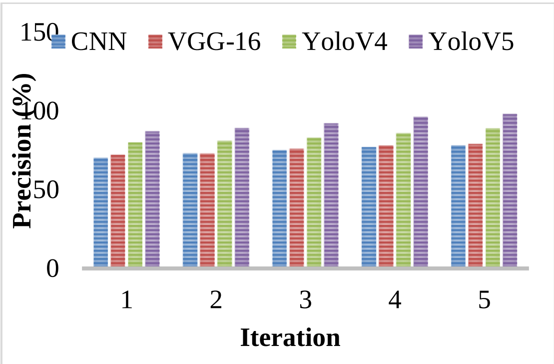
$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ mAP &= \frac{1}{C}\sum_{K-1}^{N} P(k)\Delta R(k) \\ F1 &= \frac{2}{\frac{1}{P_{\text{rectstom}}}+\frac{1}{R_{\text{coall}}}} \end{aligned} \tag{10}$$

where TP is the number of successful identifications of two apple types; FP indicates the number of false positives for apple targets in the background; An integer representation of the total amount of apple objects that have not been recognized is FN. With "C" representing the entire assortment of apple consumption categories, N representing the total amount of IOU thresholds, K representing the IOU threshold itself, P(k) representing the recall, and R(k) representing the accuracy.

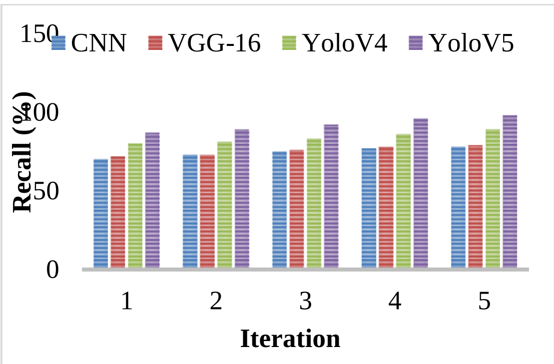
Table 2: Examination of the outcomes of apple diameter measurements

Object	Variance/ mm^2	Standard Deviation /mm	Actual Measurement Value/mm	Camera Measurement Value/mm
Apple1	8.933321	4.126392	86.66	92.792459
Apple2	4.255878	2.367900	79.18	75.823300
Apple3	1.324617	4.113411	78.44	79.812499
Apple4	7.268401	3.262963	81.53	69.320226

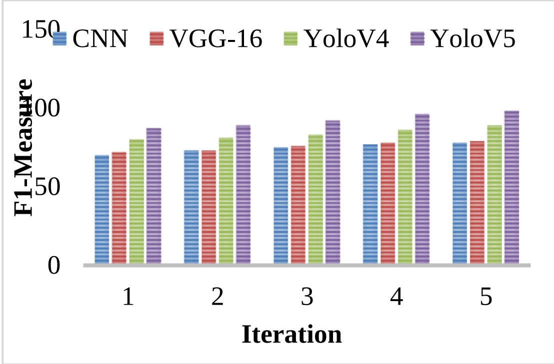
Table 2 uses standard deviation as well as variance to compare the uniformity and precision of diameter measurements made using actual and camera methods for different apple varieties grown in orchards. The variance shows how the camera's measurement results are distributed relative to the actual measurement values, while the standard deviation shows how much the actual measurement values vary from the camera's measuring values. With inaccuracies ranging from 1.13 mm to 3.14 mm, measurements taken by camera tend to be more accurate than those taken by hand.



(a) Precision



(b) Recall



(c) F1-Measure

Figure 5: Comparative analysis on various measures

4.1 Results discussion

Depending on the specifics of their project, users can select the model that works best for them in terms of creation and usage size. Our primary objective in developing and implementing the recognition approach for this research was to achieve real-time fruit targets detection. This capability will subsequently be included into our apple picking robot. A detection framework based on updated YOLOv5s is a good choice for inclusion into integrated components of choosing robot vision systems because to the properties of the YOLOv5s system, such as its quick detection speed and low simulation size. Consequently, the expenses of implementing the recognition model will decrease.

Among the various benefits of the offered apple detection technique are: The first thing it can do is recognize apple objectives in images, whether they are graspable or not. The second advantage is its high detection efficiency; the picking robot can identify apples in real time with ease thanks to the enhanced YOLOv5s model. Finally, hardware devices might make advantage of the proposed detecting designs because of how little and light they are. Given that bigger models need more setup and processing power, this is relevant to both the algorithm's broad applicability and the visual system cost of the selecting robot.

5 Conclusion

Our research focused on an apple-harvesting robot that used the ROS framework in conjunction with image processing and deep learning. This apple-harvesting robot can navigate itself around orchards, identify apples in real time, and measure their diameter with pinpoint accuracy. Automated apple harvesting is now within reach, thanks to this robotic technology, which is a huge step forward in agricultural automation that will help with manpower shortages and boost production efficiency. Having said that, the experiment still has a few holes and might need some further investigation. In complicated orchard settings, for example, navigation precision is lacking.

Obstacles to autonomous apple picking by agricultural robots include unsteady light intensity, branches and leaves getting in the way, and occlusion of the fruit by other objects. So, improving the apple-harvesting robot's ability to navigate and recognize targets in cluttered environments may be the focus of future research into making the robot more robust and adaptable to more complex agricultural settings. Our investigation centred on a robotic apple harvester that integrated deep learning, image processing, and the ROS platform. This apple-harvesting robot can navigate itself around orchards, identify apples in real time, and measure their diameter with pinpoint accuracy. Automated apple harvesting is now within reach, thanks to this robotic technology, which is a huge step forward in agricultural automation that will help with manpower shortages and boost production efficiency. Having said that, the experiment still has a few holes and might need some further investigation. In complicated orchard settings, for example, navigation precision is lacking. Obstacles to autonomous apple picking by agricultural robots include unsteady light intensity, branches and leaves getting in the way, and occlusion of the fruit by other objects. In order to create the apple-harvesting robot more resilient and suitable for more complicated agricultural settings, future research might concentrate on developing better navigation algorithms and increasing the detection of apple targets in obstructed situations. In future work, such as processing times under various conditions and the challenges posed by the physica constraints of the robots.

Funding

Talent Specialization Program of Guangxi Eco-engineering Vocational & Technical College（No.GXSTKYZX202307006）; This work was supported by the Youth scientific research and innovation team of Guangxi Normal University of science and Technology (No.GXKS2020QNTD02) Research on the Extraction Algorithm of Field Navigation Line in Visual Navigation of Mountain agricultural robot (No.2020KY23023) , Laibin City Scientific Research and Technology Development Plan Project “Research and application demonstration and promotion of sugarcane breeding based on artificial intelligence”(No.20211216121424482) and Guangxi Science and Technology Plan Project "Key Technologies and Industrial Cluster Applications for Manufacturing Human Machine Intelligent Interactive Touch Terminals" (No.Guike AA21077018).

References

[1] Kang, H., & Chen, C. (2020). Fast implementation of real-time fruit detection in apple orchards using deep learning. Comput. Electron. Agric., 168. https://doi.org/10.1016/j.compag.2019.105108. https://doi.org/10.1016/j.compag.2019.105108

[2] Lu, S., Chen, W., Zhang, X., & Karkee, M. (2022). Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. Comput. Electron. Agric., 193, 106696. https://doi.org/10.1016/j.compag.2022.106696

[3] Wu, Z., Sun, X., Jiang, H., Mao, W., Li, R., Andriyanov, N., Soloviev, V., & Fu, L. (2023). NDMFCS: An automatic fruit counting system in modern apple orchard using abatement of abnormal fruit detection. Comput. Electron. Agric., 211, 108036. http://dx.doi.org/10.2139/ssrn.4442720

[4] Journal, I. (2023). Apple Fruit Disease Detection for Hydroponic Plants Using Machine Learning and Image Processing. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. DOI:10.1109/ICOSEC51865.2021.9591903

[5] Lee, J., Gadsden, S.A., Biglarbegian, M., Cline, J.A., & Martínez, M. (2022). Smart Agriculture: A Fruit Flower Cluster Detection Strategy in Apple Orchards Using Machine Vision and Learning. Applied Sciences. https://doi.org/10.3390/app122211420

[6] Li, T.F., Fang, W., Zhao, G., Gao, F., Wu, Z., Li, R., Fu, L., & Dhupia, J.S. (2021). An improved binocular localization method for apple based on fruit detection using deep learning. Information Processing in Agriculture.DOI: 10.1016/j.inpa.2021.12.003

[7] Marandi, S., & Chakravorty, P.C. (2021). Apple Fruit Disease Detection Using Image Processing. DOI: 10.17148/IJARCCE.2022.11645

[8] Subha, V., & Kasturi, K. (2024). Structural Invariant Feature Segmentation Based Apple Fruit Disease Detection Using Deep Spectral Generative Adversarial Networks. SN Computer Science, 5, 1-11. https://doi.org/10.1007/s42979-024-02994-z

[9] Johanson, R., Wilms, C., Johannsen, O., & Frintrop, S. (2023). S3AD: Semi-supervised Small Apple Detection in Orchard Environments. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 7061-7070. https://doi.org/10.48550/arXiv.2311.05029

[10] Ulla, S., Shariff, M.G., & Basavanna, G. (2020). Fruit Categorization and Disease Detection Using Ml Raspberry Pi Based Fruit Categorization and Quality Maintenance with Disease Detection Using Ai and Machine Learning. International Journal of Scientific & Technology Research, 9, 219-227. https://doi.org/10.3390/ai4040041

[11] Wang, X., Kang, H., Zhou, H.X., Au, W., Wang, M.Y., & Chen, C. (2023). Development and evaluation of a robust soft robotic gripper for apple harvesting. Comput. Electron. Agric., 204, 107552. https://doi.org/10.1016/j.compag.2022.107552

[12] Au, W., Zhou, H., Liu, T., Kok, E., Wang, X., Wang, M.Y., & Chen, C. (2023). The Monash Apple Retrieving System: A review on system intelligence and apple harvesting performance. Comput. Electron. Agric., 213, 108164. https://doi.org/10.1016/j.compag.2023.108164

[13] Pupin, D.S., & Chilikin, A. (2023). Development of control system for robotic apple harvesting device. E3S Web of Conferences. https://doi.org/10.1051/e3sconf/202345810022

[14] Pupin, D.S., & Chilikin, A. (2023). Development of control system for robotic apple harvesting device. E3S Web of Conferences. https://doi.org/10.1051/e3sconf/202345810022

[15] Chen, K., Li, T., Yan, T., Xie, F., Feng, Q., Zhu, Q., & Zhao, C. (2022). A Soft Gripper Design for Apple Harvesting with Force Feedback and Fruit Slip Detection. Agriculture. https://doi.org/10.3390/agriculture12111802

[16] Zhang, K., Lammers, K., Chu, P., Li, Z., & Lu, R. (2023). An automated apple harvesting robot - From system design to field evaluation. J. Field Robotics, 41, 2384-2400. DOI:10.1002/rob.22268

[17] Yan, Z., & Sun, M. (2023). THREE-DIMENSIONAL PATH PLANNING OF APPLE HARVESTING ROBOT BASED ON IMPROVED GENETIC ALGORITHM. INMATEH Agricultural Engineering. https://doi.org/10.35633/inmateh-71-40

[18] Ji, W., He, G., Xu, B., Zhang, H., & Yu, X. (2024). A New Picking Pattern of a Flexible Three-Fingered End-Effector for Apple Harvesting Robot. Agriculture. https://doi.org/10.3390/agriculture14010102

[19] Huang, W., Miao, Z., Wu, T., Guo, Z., Han, W., & Li, T. (2024). Design of and Experiment with a Dual-Arm Apple Harvesting Robot System. Horticulturae. https://doi.org/10.3390/horticulturae10121268

[20] Zhu, F., Zhang, W., Wang, S., Jiang, B., Feng, X., & Zhao, Q. (2024). Apple-Harvesting Robot Based on the YOLOv5-RACF Model. Biomimetics, 9. https://doi.org/10.3390/biomimetics9080495

[21] Yan, B., Quan, J., & Yan, W. (2024). Three-Dimensional Obstacle Avoidance Harvesting Path Planning Method for Apple-Harvesting Robot Based on Improved Ant Colony Algorithm. Agriculture.; https://doi.org/10.3390/agriculture14081336

[22] Zhou, H.X., Kang, H., Wang, X., Au, W., Wang, M.Y., & Chen, C. (2023). Branch Interference Sensing and Handling by Tactile Enabled Robotic Apple Harvesting. Agronomy. https://doi.org/10.3390/agronomy13020503

[23] Wenju, M., Heng, L., Wei, H., Fuzeng, Y., & Zhijie, L. (2022). Development of a Combined Orchard Harvesting Robot Navigation System. Remote Sensing. https://doi.org/10.3390/rs14030675

[24] Das, P., Chakraborty, A., Sankar, R., Singh, O.K., Ray, H., & Ghosh, A. (2023). An Apple Detection Approach Based on Deep Learning Algorithm in Kashmir Orchards for Apple Harvester Robot. 2023 3rd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), 1-6. DOI:10.1109/ICEFEET59656.2023.10452237

[25] Xiong, Z., Feng, Q., Li, T., Xie, F., Liu, C., Liu, L., Guo, X., & Zhao, C. (2022). Dual-Manipulator Optimal Design for Apple Robotic Harvesting. Agronomy. DOI:10.3390/agronomy12123128