IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 25, 2020, accepted June 12, 2020, date of publication June 17, 2020, date of current version July 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003034

Real-Time Visual Localization of the Picking Points for a Ridge-Planting Strawberry Harvesting Robot

YANG YU, KAILIANG ZHANG, HUI LIU, LI YANG, AND DONGXING ZHANG
College of Engineering, China Agricultural University, Beijing 100083, China

Corresponding author: Kailiang Zhang (zhang_kailiang@cau.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61375189, and in part by the National Key Research and Development Plan of the Thirteenth Five-Year, China, under Grant 2017YFD0700503.

ABSTRACT At present, the primary technical deterrent to the use of strawberry harvesting robots is the low harvest rate, and there is a need to improve the accuracy and real-time performance of the localization algorithms to detect the picking point on the strawberry stem. The pose estimation of the fruit target (the direction of the fruit axis) can improve the accuracy of the localization algorithm. This study proposes a novel harvesting robot for the ridge-planted strawberries as well as a fruit pose estimator called rotated YOLO (R-YOLO), which significantly improves the localization precision of the picking points. First, the lightweight network Mobilenet-V1 was used to replace the convolution neural network as the backbone network for feature extraction. The simplified network structure substantially increased the operating speed. Second, the rotation angle parameter α was used to label the training set and set the anchors; the rotation of the bounding boxes of the target fruits was predicted using logistic regression with the rotated anchors. The test results of a set of 100 strawberry images showed that the proposed model's average recognition rate to be 94.43% and the recall rate to be 93.46%. Eighteen frames per second (FPS) were processed on the embedded controller of the robot, demonstrating good real-time performance. Compared with several other target detection methods used for the fruit harvesting robots, the proposed model exhibited better performance in terms of real-time detection and localization accuracy of the picking points. Field test results showed that the harvesting success rate reached 84.35% in modified situations. The results of this study provide technical support for improving the target detection of the embedded controller of harvesting robots.

INDEX TERMS Ridge-planting, harvesting robot, R-YOLO, fruit detection, rotated bounding box.

I. INTRODUCTION

As one of the most widely grown berries in the world, strawberry can be cultivated in the outdoors or controlled environments such as greenhouses and polytunnels [1]. In general, the cultivation modes in the greenhouses include table-top, bench-type, elevated-substrate, and ridge-planting (Fig.1a, b, c, d). Although the cultivation modes of Fig. a, b, c are more advanced in terms of the stereoscopic space utilization, as well as fruit yield and quality, ridge-planting is still widely implemented in China because of its lower initial costs and easy implementation. The ridge-planting cultivation area accounts for more than 90% of the total cultivation area [2]. However, strawberry cultivations have several drawbacks. Harvesting is the most time-consuming and labor-intensive step in strawberry production. Harvesting labor costs account for more than 75% of the total production costs, and this proportion continues to increase annually [3]. Labor shortages further restrict the economic benefits and development of the strawberry industry. Also, the low ridges and narrow roads between the ridges make ridge-planting more difficult, time-consuming, and labor-intensive than other cultivation modes. Therefore, research on harvesting robots can make great impact on reducing manual labor, improving harvesting efficiency, and reducing production costs. The market demand is especially strong for a harvesting robot used in ridge planting.

No reliable and cost-effective business system has been established in recent years, although scholars from Japan [5]–[7], China [10], [11], Norway [1], [4], [9] and Iran [8] have conducted extensive research on strawberry harvesting robots [12]. Some start-ups have also launched their

The associate editor coordinating the review of this manuscript and approving it for publication was Aysegul Ucar.

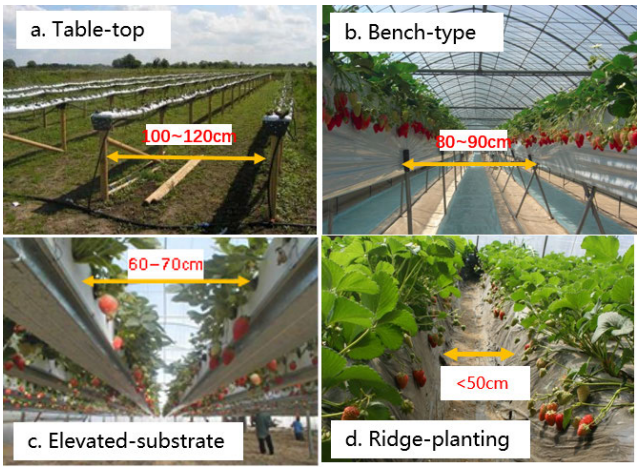


FIGURE 1. Strawberry cultivation modes.

robot designs. In 2013, Shibuya Seiki of Japan exhibited a strawberry harvesting robot that produced a >70% success rate; Harvest CROO of the United States designed a rotating device with multiple grippers that can pick strawberries on the ground. However, most of the harvesting robots have been designed for elevated cultivation. The sizes of the Cartesian-type and industrial robot arms are not suitable for the low and narrow ridge-planting environment of a greenhouse. As shown in Fig.1d, the width of the road between the two ridges is less than 50cm, and the height of the ridge is less than 40cm. The strawberry fruits cling to the side of the ridge and grow downward.

AGROBOT of Spain has developed an automatic strawberry harvesting machine (SW6010) that can be used in a ridge-planting mode. The machine has 24 independent manipulator arms, and the whole process (including searching for fruit and cutting fruit stems) takes only 4 seconds. However, no detailed information about the physical and performance parameters of this machine has yet been published. Therefore, research on strawberry harvesting robots in a ridge-planting mode is lacking.

Another major challenge in automated harvesting is that strawberries are easily damaged and bruised, so the end-effector (hand claw) of the harvesting robot cannot touch the surface of the fruits during harvesting [13]. Successful harvesting can only be achieved by cutting or burning the fruit stem. Therefore, the design of the contactless end-effector and the precise localization of the picking point on the fruit stem is essential for successful harvesting. Xiong *et al.* [1], [4] designed a cable-driven gripper for contactless harvesting of elevated strawberries that delivers them directly into a market pun-net, thereby eliminating the need for repacking. This machine produced a state-of-the-art harvesting performance. However, it could not be applied in the ridge-planting mode. The structured light RGB-D camera (Intel R200) used by Xiong to detect the depth of the fruit targets could not be adapted to the narrow and low passages as the camera's effective detection distance cannot be less than 80cm. Therefore, the implementation of effective target detection in a narrow and low environment of the ridge-planting mode is urgently needed. In general, the prediction of picking points requires the identification of the fruit targets, followed by the prediction of the location of the picking point based on differences in the size, shape, color, and texture between the fruit and the background. Furthermore, due to the complex operating environment of the harvesting robot and the various physical features of the fruit targets, fruit detection is susceptible to various factors, such as the intensity of the natural light, overlapping fruits, and the blocking of the fruit by stems and leaves. Low-precision target detection complicates the accurate prediction and localization of the picking points. Therefore, the rapid and accurate detection (identification and localization) of the picking points represents the core problem that has to be addressed in strawberry harvesting robots [14].

Target detection of fruits is similar in many ways to other target detecting applications, such as autonomous driving and face recognition. Therefore, the classical target detection models (R-CNN [15], SSP-Net [16], Faster R-CNN [16], YOLO [17], and SSD [18]) can also be applied to fruit detection. Of these models, R-CNN, SSP-Net and Faster R-CNN have a two-stage structure, which is slower than the one-stage methods [20], like YOLO and SSD. It is well known that the YOLO-V3 model is the preferred target detection algorithms in the engineering community due to its explicit structure and good real-time performance. However, the original YOLO model is not suitable for fruit target detection because of two reasons. First, unlike the COCO or VOC data set, which contains 80 or 20 classes, the fruit target has less than 10 classes. Therefore, the structure of the YOLO model needs to be modified and simplified to be suitable for fruit target detection. In addition, the simplification of the YOLO model will improve real-time performance. The second reason is that the target bounding box predicted from the original YOLO model is horizontal and contains many non-target pixels [21]; therefore, the pose of the fruit target is not well-defined, causing localizing errors for determining the picking point. Lei *et al*. [20] and Liu *et al.* [21] improved Faster RCNN and YOLO-V3, respectively, by using target rotation information for feature extraction to predict the rotation bounding box. The methods detected the orientation of the target in Remote Sensing Images, achieving a good balance between performance and efficiency. However, while the feature extraction networks (VGGNET and DarkNet53) of Liu and Lei run fast on the servers equipped with GPU accelerated computing cards (NVIDIA GTX 1080Ti / 4 TitanX), they cannot meet the real-time requirements of target detection in the robot embedded control terminal.

In this paper, a strawberry pose estimator called R-YOLO is proposed. It can be transplanted into the embedded control device of the robot to address the problems discussed above and can determine the picking-point position in real-time. This estimator not only identifies the strawberry targets but also generates the rotated bounding box containing the pose information of the fruit target. The slope of the fruit axis can be calculated using the rotated bounding box.

Subsequently, the picking point's position on the fruit stem can be located based on the direction of the fruit axis. This localization method substantially improves the harvesting success rate. Below is a summary of the proposal:

1. Designing a novel end-effector that is assembled on the servo control system of a strawberry harvesting robot suitable for the narrow ridge-planting mode. Unlike the others, the proposed end-effector is equipped with a pair of opposite fiber sensors between the two fingertips. Therefore, when approaching the fruit target, it does not need to measure the depth distance in real time, which simplifies the robot structure and speeds up the control.
2. Proposing R-YOLO, a target detector suitable for strawberry fruit in the narrow spaces of the ridge-planting mode. The rotated bounding box of the strawberry target is achieved by adding a rotation angle a to the anchors, thereby significantly improving the localization accuracy of the picking point. The proposed R-YOLO adopts a lightweight network for feature extraction, which demonstrated good real-time performance on the embedded control device of the robot.
3. The proposed method can be used easily and quickly to identify picking points of other fruits and vegetables. Also, it requires a small number of image samples for model retraining and minor modifications of the mechanical structure size.

The rest of the paper is arranged as follows: Section II reviews the literature of fruit target detection. In sections III, IV and V, we introduce the mechanical structure of the designed harvesting robot and the proposed object detection method for training and testing. The experimental results and discussions are provided in section VI. Section VII presents the conclusions of this work.

II. RELATED WORK

Fruit target detection is an important prerequisite for automatic harvesting. Various factors in the natural environment, such as the intensity of the light, overlapping fruits, and the occlusion of the fruit by stems and leaves, have resulted in many challenges. In recent years, numerous studies were conducted on fruit target detection. Commonly used methods include digital image processing, machine learning, and deep learning.

A. COMBINATION OF IMAGE PROCESSING AND MACHINE LEARNING

The combination of digital image processing and machine learning algorithms is the current mainstream approach for fruit target detection. In general, the first step includes image preprocessing operations, such as threshold segmentation, edge detection, and region growing in different color spaces to extract various features, such as the color, size, shape, and texture of the fruit target [22]–[26]. Subsequently, k-means clustering [27], the k-nearest neighbor method [28], support vector machine (SVM) [29], and artificial neural networks [30], and other machine learning algorithms have been used for target detection. Ouyang *et al.* [31] performed several processing operations on strawberry images to identify diseases, including median filtering to remove noise, the Otsu algorithm for image segmentation, and mean-shift clustering and morphological operations to obtain the most discriminative shape features. Wei *et al.* [32] extracted a new color feature in the OHTA color space, which was used to automatically calculate the segmentation threshold of fruit images using the improved Otsu algorithm; a recognition accuracy of more than 95% was achieved. An elevated strawberry harvesting robot designed by Qingchun *et al.* [11] used hue and saturation features to identify ripe fruits in the hue, saturation, and value (HSV) color space, and a binocular vision unit was used to determine the picking points. Benalia *et al.* [33] developed an automatic system to improve the quality control and sorting of dried figs (*Ficus carica*) based on computer vision. The browning index of each fruit and features extracted from the CIE XYZ, CIELab, and HunterLab color spaces were used as the input of a principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA); excellent results were obtained. Borges *et al.* [34] proposed a clustering method to detect and classify the severity of bacterial spot in tomatoes. The premise was to preprocess the images and extract the color features using the CIELab color space. In general, the above studies require expert knowledge to extract the features of the fruits. The target detection often suffered from low robustness and was greatly affected by the differences in the images and environmental factors. It is challenging to develop a method that can detect heterogeneous strawberry fruits and is not affected by multiple fruits, overlapping fruits, and the occlusion by stems and leaves. In addition, machine learning algorithms generally require a large number of samples, and methods that combine image processing and machine learning models are complex and have poor real-time performance.

B. FRUIT DETECTION BASED ON DEEP LEARNING

In recent years, object detection models based on deep learning and capable of good image representation and autonomous learning, such as R-CNN, faster R-CNN, YOLO, and SSD, have been widely used in research on fruit target detection [36]–[42]. Bargoti and Underwood [43] proposed an image processing framework for fruit detection and counting using images of orchards. A general image segmentation method was adopted, including two feature learning algorithms, *i.e.*, multi-scale multi-layered perceptrons (ms-MLP) and CNN; good fruit detection performance was obtained. Zhou *et al.* [44] optimized an 8-layer network based on VGGNet for feature extraction of the stem, flower, and fruit of tomatoes. Fu *et al.* [45] used the LeNet CNN to identify multiple clusters of kiwifruits; the method provided better performance in terms of speed and accuracy than other traditional methods. Inkyu *et al.* [46] applied the Faster R-CNN model to multi-band images

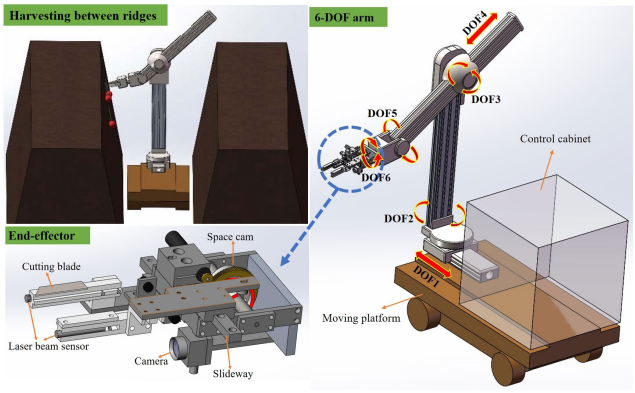


FIGURE 2. **The overall structure of the designed harvesting robot.**

(RGB and near-infrared) and used migration learning for sweet pepper detection. After retraining, this model was able to identify several other fruits, such as melons, apples, avocados, oranges, and strawberries. Tian *et al.* [47] improved the YOLO-V3 model to achieve real-time detection of apples in different natural environments and adopted the DenseNet network to process low-resolution feature layers. The experimental results showed that the improved model had better detection performance than the original YOLO-V3 model and faster R-CNN. However, the complexity of the model structure resulted in long running times on the embedded control devices and poor real-time performance. Moreover, the above-mentioned object detection algorithms can only roughly calculate the location of the fruit target because of the horizontal bounding box. The contour or pose information of the fruit target cannot be extracted, and the spatial relationship between the fruit and the picking point on the stem cannot be determined. In most cases, the target fruit can be located, but the picking point on the stem cannot. Unlike apple, citrus, and other fruits with a hard-outer cortex, harvesting of strawberries can only be achieved by cutting or burning the stem to avoid damage to the skin of the fruit. Therefore, the precise localization of the picking point is the ultimate goal of strawberry detection. None of the above methods meet the detection requirements for strawberry harvesting.

III. OVERVIEW OF THE ROBOT MECHANICAL STRUCTURE

Focusing on the low and narrow work environment of the ridge-planting cultivation mode, this paper designed a novel strawberry harvesting robot. As shown in Fig.2, the robot's hardware mechanism was independently designed, and assembled by the team to include three main modules: ''hand-eye / laser sensor'' end-effector, 6-degrees-of-freedom (6DOF) arm, and the moving chassis.

The end-effector was composed mainly of a space cam spring mechanism and two mechanical fingers. The maximum opening width of the fingertips was 45mm, and the diameter of the fruit stem that could be cut was within 3mm. The hand-eye visual system used a USB camera with a 640×480 resolution to capture video images in real-time for target detection. The head of the fingers was equipped with a pair of laser beam sensors, which emitted the signal on one side and received on the other. When the fruit stem entered the spaces between the two fingertips, it blocked the laser beam, triggering a ''fingertip closing'' signal, which was sent to the control module to perform an immediate closing action and cut the stem. Moreover, the end-effector could approach the fruit target at a fast and uniform speed (approximately 20cm/s) without detecting for the depth of the target fruit.

The 6DOF robotic arm was independently designed by the research team. Its overall height in the initial state was 80 cm, and it consisted of two rods, two moving joints, and four rotating joints. Each joint was equipped with a DC deceleration servo motor. When the end-effector needed to move to a specified position, the robot kinematics inverse solution was used to calculate the movement required for each joint, and then the pulse signal of each joint motor was outputted in sequence.

The moving chassis on four wheels was 50cm long and 30cm wide, which met the narrow environment requirements of the ridge-planting mode. In addition to the fixed 6-DOF arm, the upper space of the chassis was also equipped with a control cabinet. The control cabinet contained the main controller (Raspberry Pi 3B+), embedded target detection module (Jetson TX2), power transformer module, motor drivers, relay module and wiring terminals. All the circuits of motors and sensors were encapsulated in the control cabinet via the aviation plugs.

IV. R-YOLO DETECTOR

The proposed strawberry target pose detector (rotated YOLO (R-Yolo)) is an improvement of the YOLO-V3 model. The YOLO model, which is an object detection model, has the advantages of fast running speed and a simple model structure. The target detection task consists of classifying the target objects in the image and the generation of a bounding box of each target. YOLO reframes object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. First, fixed size ($n \times n$) feature maps are extracted from the input image through the feature extraction network. The input image is divided into $n \times n$ grid cells. Each grid cell predicts 3 bounding boxes with different sizes. If the center point of an object in the ground truth is located in a grid cell, then the grid cell will be used to predict the object. Compared with other classic target detection models such as R-CNN, fast R-CNN, and faster R-CNN, YOLO eliminates the complexity caused by a large number of sliding windows (anchor) and regional proposals generated in the region proposal network (RPN), which greatly improves the real-time performance of target detection.

R-YOLO uses the lightweight CNN MobileNet-V1 [48] as the backbone network for feature extraction to increase the running speed of YOLO on the embedded controller of the robot. MobileNet, which has excellent real-time performance when used in embedded devices, was proposed

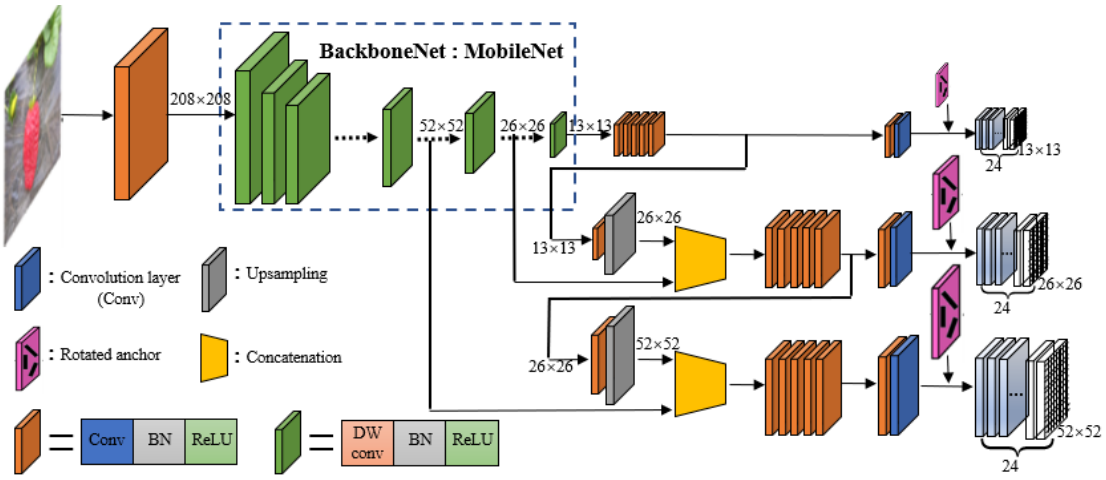


FIGURE 3. Overview of the R-YOLO model.

by Google in 2017. It significantly reduces the number of parameters in the convolutional networks by using depthwise separable convolution. In addition, the target bounding boxes predicted by conventional target detection models, such as fast R-CNN, faster R-CNN, and YOLO are horizontal and cannot describe the pose information of the target. However, the pose information of strawberry fruit is very important for automatic harvesting and improves the localization accuracy of the picking point for the end-effector. The target pose information minimizes the opening and closing range of the end-effector, thereby avoiding the harvesting of neighboring fruits or causing damage to the fruit. Therefore, the proposed R-YOLO not only needs to have high accuracy and good real-time performance for fruit detection but also needs to predict the rotation of the bounding box; therefore, the rotation angle of the fruit axis (pose information) has to be calculated. The framework of R-YOLO is shown in Fig.3.

A. THE MobileNet-V1 BACKBONE NETWORK

R-YOLO uses the lightweight network Mobilenet-V1 as the backbone network for feature extraction to improve the running speed in the feature extraction stage and reduce the memory requirements for a better real-time performance of the embedded controller. Mobilenet-V1 decomposes the standard convolution kernel into a 3 × 3 depthwise convolution and a 1 × 1 pointwise convolution. The depthwise separable convolution (DW Conv) reduces redundant expressions of the convolution kernel, which significantly decreases the number of convolutions and parameters and improves the real-time performance of the embedded controller. The backbone network structure of Mobilenet-V1 is shown in Fig.4.

Fig.4 shows the detailed structure of the backbone network of R-YOLO and the DW Conv. The DW Conv includes two independent modules: the depthwise convolution and the 1×1 pointwise convolution. Batch normalization is performed on the output, and a nonlinear activation unit (ReLU) is added. In Fig.4, s represents the step size of the depthwise convolution and k represents the number of the 1 × 1 pointwise convolution.

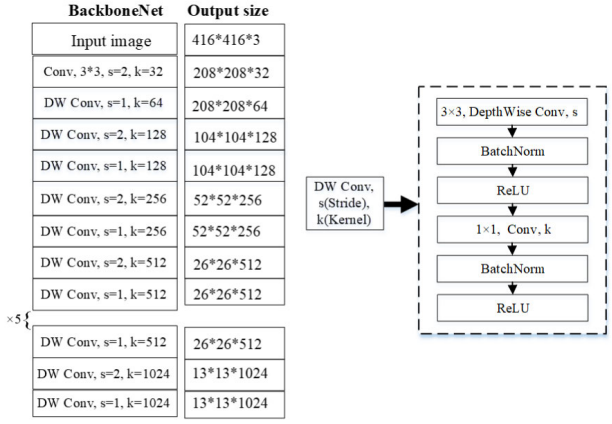


FIGURE 4. The backbone network of Mobilenet-V1.

B. THE PREDICTION OF ROTATED BOUNDING BOX

The feature map outputs from the backbone network divided the input image into n×n grid cells. Each grid cell contained the probability of 3 bounding boxes and 2 categories (ripe or unripe). The prediction of each bounding box consisted of 5 parameters (*x, y, w, h, confidence*). (*x, y*) is the center coordinates of the box, which was normalized to a range of 0 to 1; the size of the box, (*w, h*), was also normalized to [0, 1] relative to the size of the image.

1) THE IMAGE ANNOTATION OF THE ROTATED BOUNDING BOX

Different from the horizontal bounding box in the traditional labeled image, the rotated bounding box excludes most of the background and represents the smallest bounding box of the fruit target. The rotation parameter α was added during labeling the training image set to generate a rotated bounding box. Fig.5 shows the labeled strawberry image; the red dashed line is the traditional horizontal bounding box, and the yellow solid line box is the rotated bounding box with a rotation parameter α. The rotated bounding box is represented by five parameters: (*x, y, w, h,* α), where (*x, y*) represents

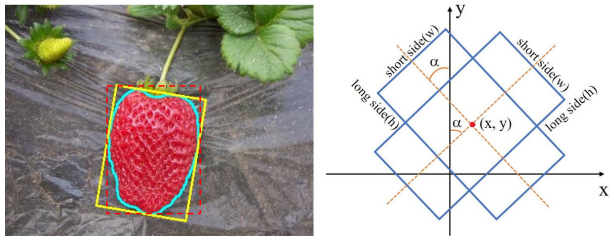


FIGURE 5. Image annotation and the rotation range of α.

the coordinate of the center point of the bounding box, w and h respectively represent the long and short sides of the bounding box, and α represents the angle between the y-axis and the long side of the bounding box. The rotation range of α is shown in Fig.5 and is $\alpha \in [-90°, 90°]$. If α is negative, the fruit axis is rotated to the left by α. Otherwise, the axis is rotated to the right. When α is 0, the fruit axis is vertical. The robot determines the pose of the fruit target according to the values of α and controls the rotation direction and angle of the end-effector joint to reduce the localization error of the picking point.

2) GENERATION OF THE ROTATED ANCHOR

The backbone network output three features maps with different sizes. Each grid cell on the feature map had 3 anchors with fixed sizes, which were used to predict the target bounding box using logistic regression. The rotation angle α was also added to the traditional horizontal anchor to match the ground truth in the annotated image. The distances between the camera that captured the strawberry images and the target fruits were variable, resulting in different sizes of the target bounding boxes labeled in the training set. K-means clustering was used on the sizes of the fruit bounding boxes in the training set, and R-YOLO used nine anchors with different lengths and widths. The smallest feature maps (13×13) with the largest receptive field were assigned the three largest anchors ((96×113), (154×181), and (286×319)), which were suitable for detecting larger fruits. Medium-sized feature maps (26×26) with medium receptive fields and medium-sized anchors ((31×58), (43×62), and (55×107)) were used to medium-sized fruits. The remaining three smallest anchors ((11×15), (17×29), and (32×34)) were assigned to the largest feature maps (52×52) and were suitable for detecting smaller targets. Each anchor had six rotation angles: $\{-\pi/3, -\pi/6, 0, \pi/6, \pi/3, \pi/2\}$, which meant that each grid cell in the feature maps had 18 anchors (3×6).

C. LOSS FUNCTION

The training loss of R-YOLO included two parts: classification loss and bounding box prediction loss. The prediction of the target bounding box was performed using logistic regression, which output six parameters: the coordinates of the center point (x, y), the box size (w, h), *confidence*, and the rotation angle α. In this study, different loss functions were formulated according to the calculation characteristics of the parameters. Each loss part was added to obtain the total training loss and perform end-to-end loss function training. The total training loss L_{total} is expressed as follows:

$$\begin{aligned} L_{total} &= L_{class} + L_{bbox} \\ &= L_{class} + (L_{x\_y} + L_{w\_h} + L_{conf} + L_{\alpha}) \end{aligned} \tag{1}$$

where L_{α} is the smooth L1 loss function [49]:

$$\begin{aligned} L_{\alpha} &= \sum_{i=0}^{S^2} \sum_{j=0}^{B} smooth_{L1}\left(v_i - v_i^*\right) \\ smooth_{L1}(x) &= \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise, \end{cases} \\ v &= \alpha - \alpha_{anchor} + k \cdot \pi, \quad k \in Z,\ v \in (-\frac{1}{2}\pi, \frac{1}{2}\pi] \end{aligned} \tag{2}$$

where S^2 represents the size of the extracted feature map ($S = 13/26/52$), B represents the total number of anchors in the feature map, v and v^* respectively represent the predicted and the ground-truth bounding boxes, α represents the predicted rotation angle, and α_{anchor} represents the rotation angle of the anchor with the largest intersection over union (IoU) value compared to the fruit target in the image. The designs of the loss function other than L_{α} were the same as that of YOLO-V3. L_{w_h} used the mean squared error (MSE) loss function:

$$L_{w\_h} = \lambda \cdot \sum_{i=0}^{S^2} \sum_{j=0}^{B} \left(\sqrt{w_i} - \sqrt{w_i^*}\right)^2 + \left(\sqrt{h_i} - \sqrt{h_i^*}\right)^2 \tag{3}$$

where w and w^* represent the widths of the predicted and ground-truth bounding boxes. h and h^* represent the lengths of both. The rest of the loss functions all used the binary cross-entropy:

$$\begin{aligned} &L_{Binary\_crossentropy}\left[s, s^*\right] \\ &\quad = \sum_{i=0}^{S^2} \sum_{j=0}^{B} s_i \log s_i^* + (1 - s_i) \log\left(1 - s_i^*\right), \\ &\quad s \in \{(x, y), confidence, class\} \end{aligned} \tag{4}$$

$$L_{x\_y} = \lambda \cdot L_{Binary\_crossentropy}\left[(x, y), (x^*, y^*)\right] \tag{5}$$

L_{class} and L_{conf} were calculated in the same way as L_{x_y}. The hyperparameter λ was used to balance the training losses of the classification and bounding box, and s and s^* represent the predicted and ground-truth bounding boxes, respectively.

D. THE LOCALIZATION OF THE PICKING POINT

R-YOLO predicted the rotated bounding boxes of the fruit targets in the view-field of the camera in real-time. A coordinate system was established by using the first pixel of the original image as the origin and the image length and width as the x- and y-axes, respectively. In this coordinate system, the straight line passing through the center point A (x, y) and the rotation angle α was generated (as shown in Fig.6); the line was the long axis of the fruit target.

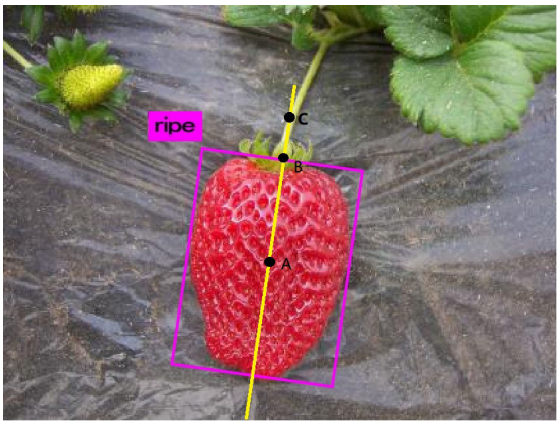


FIGURE 6. The localization of the fruit and picking point: A. barycenter, B. vertex, and C. picking point.

The robot hand-eye visual system calculated the axis direction (inclination slope) of the target fruit in real-time. The localization process of the picking point worked as follows:

1) The hand-eye visual system searched the target fruit and adjusted the initial state of the end-effector. Once the target detection module found the target fruit, the main controller adjusted the initial state of the end-effector to face the ridge wall and aimed at the target fruit.
2) The end-effector approached the target fruit and the wrist joint rotated in real time. The robotic arm control system received the slope value of the fruit axis from the visual system in real time, and dynamically adjusted the rotation angle of the wrist joint to keep the angle of the cutter consistent with the target fruit.
3) For visual detection of the picking point, statistical methods were used to measure the physical size of a large number of strawberry samples in the natural environment. The results showed that the best picking point for a strawberry is generally 13-20 mm above the calyx (top of the fruit) along the fruit stem. Once the laser beam sensor of the end-effector was triggered, the end-effector stopped moving towards the fruit. In a previous camera calibration study, the distance at which the end-effector cut the stem was 13-20mm, which corresponded to 20 pixels in the image. Therefore, at the moment when the cutting action was completed, the localization of the picking point (Fig.6C) was approximately 20 pixels away from the vertex of the contour (Fig.6B) along the axis of the fruit. The existing measurement error was caused by the distortion of the visual system when shooting at close range, but it was within the range of successful harvesting.

V. THE IMPLEMENTATION PLATFORM OF R-YOLO

After several operations, such as filtering, labeling, and image processing, the captured strawberry images were divided into a training set, validation set, and test set. The training set and validation set were used for model training and parameter

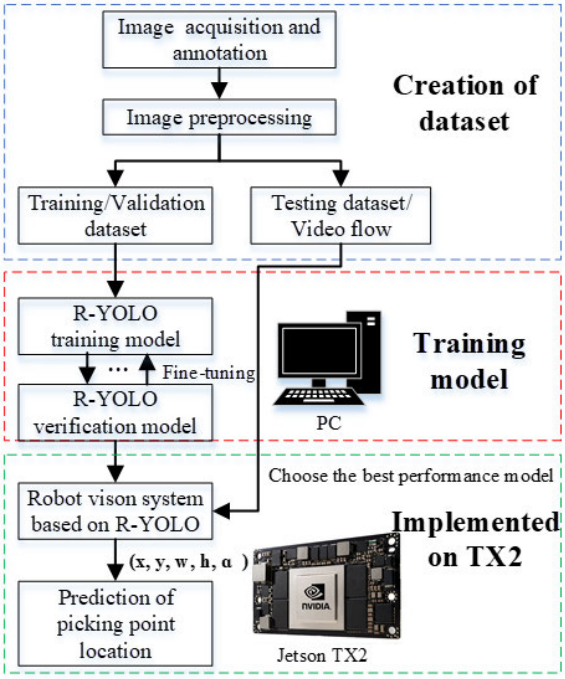


FIGURE 7. The flowchart of R-YOLO training and inference.

tuning of R-YOLO, and the test set was used to evaluate the performance of the trained model. The trained R-YOLO was implemented on the embedded control module for inference. The predicted rotated bounding box output from R-YOLO contained five parameters (x, y, w, h, α), which were used to calculate the coordinate of the picking point on the fruit stalk. The flowchart of R-YOLO training and inference is shown in Fig.7.

A. TRAINING THE R-YOLO MODEL

The proposed R-YOLO model was an improvement on the Darknet version of YOLO-V3. An Intel CPU (R) with a core (TM) of i7-8700k and 16 GB memory and an NVIDIA 1080 GPU for accelerated computing were used for training the model. In this experiment, 1900 out of 2000 strawberry images were selected for training (80% for the training set and 20% for the validation set). The remaining 100 images were used to evaluate the performance of the trained model. Four data enhancements were adopted for these images: the image brightness and contrast were enhanced by 1.5 times, respectively, and reduced to 50% of the original image. Therefore, the actual number of testing images was expanded to 400. Since the increase and decrease of the brightness and contrast do not modify the pixel coordinates in the originally- labeled images, no additional manual labeling was required. Before model training, migration learning was used for pretraining of the model based on the COCO dataset to address the problem of insufficient samples in the training set. The pre-training model extracted the general characteristics from the image set and provided good training performance, even for a relatively small dataset.

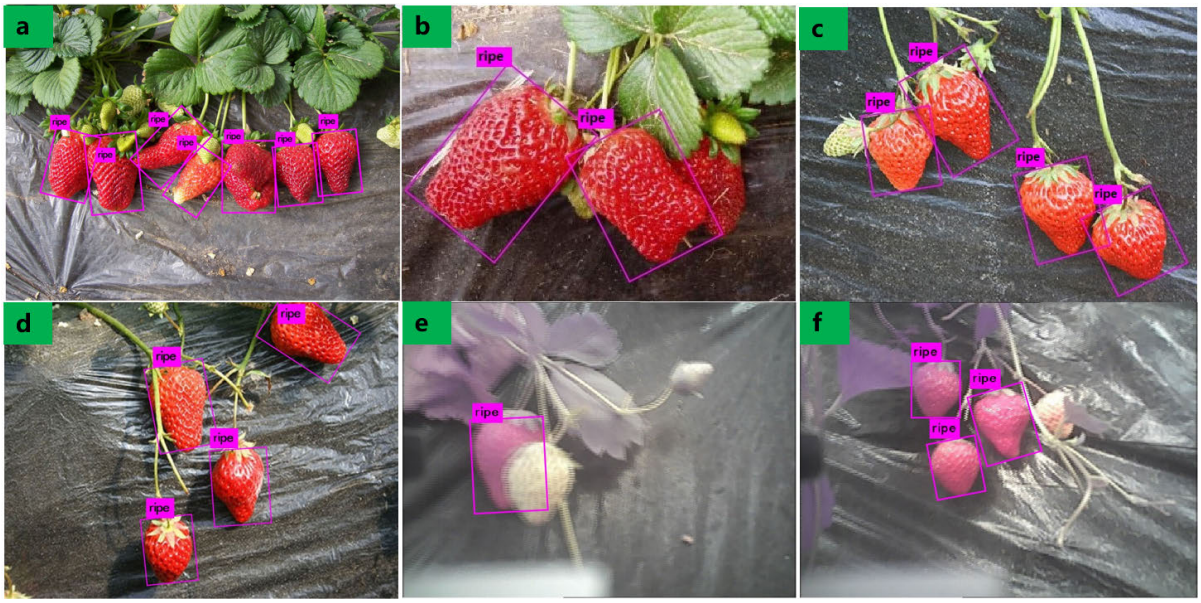


FIGURE 8. Results of strawberry detection: a. adherent fruit, b. overlapping fruit, c. separated fruit, d. occluded fruit, e and f. fruits under insufficient illumination.

B. THE EMBEDDED PLATFORM FOR THE TRAINED R-YOLO MODEL

The trained R-YOLO model was implemented on the embedded control platform NVIDIA Jetson TX2 for inference. The TX2 equipped with an NVIDIA Pascal ™GPU with 256 NVIDIA CUDA cores provides superior speed and energy efficiency for using embedded AI computing devices. Moreover, it should be emphasized that the module size of the TX2 is only 50mm×87mm, which meets the space size requirements of the robot control platform.

VI. EXPERIMENT RESULTS AND DISCUSSION

A. RESULTS AND EVALUATION OF STRAWBERRY DETECTION

The test set included 100 strawberry images (573 mature fruits and 315 immature fruits). In the experiment, the overlap coefficient (*OC*) [45] was used to evaluate the target detection accuracy. The *OC* was the ratio of the overlap between the detected target and the ground truth. The *OC* is calculated as follows:

$$OC = \frac{A_T \cap A_D}{A_T \cup A_D} \tag{6}$$

where A_r and, A_D respectively, represent the ground-truth bounding box and the detected bounding box. The successful harvest was performed when most of the area (≥90%) of the fruit was identified. Therefore, if the *OC* was 0.9 or above, the target detection result was considered correct. The detection performance of R-YOLO is shown in Fig.8.

As shown in Fig.8, R-YOLO not only showed good target detection performance for multiple separated fruits, overlap, and occlusion (Fig.8a, b, c, and d), but also for images with low light and interference (Fig.8e and f). The confusion matrix of the detection results for 100 image samples is listed in Table 1.

TABLE 1. Confusion matrix of R-YOLO detection results.

Ground Truth	Predicted Class		
	Ripe Fruits	Unripe Fruits	Background
Ripe Fruits	549	3	21
Unripe Fruits	12	287	16
Background	4	23	/

TABLE 2. Precision and recall rate of R-YOLO.

Evaluation parameter	Ripe fruit	Unripe fruit	Overall
Precision rate/%	97.17	91.69	94.43
Recall rate/%	95.81	91.11	93.46

The precision (*P*) and recall (*R*) rates were used to evaluate the target detection performance of R-YOLO:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \tag{7}$$

where *TP* is the number of cases that are correctly labeled as positive. *FP* is the number of cases that are incorrectly labeled as positive. *FN* is the number of cases that are positive but were labeled as negative [50]. The *P* and *R* results are shown in Table 2.

As shown in Tables 1 and 2, the results of the 100 test images showed that the overall *P* and *R* rates were 94.43% and 93.46%, respectively. The main reasons for the errors in fruit detection were as follows. The size of the unripe fruit samples was only 53% of that of the ripe fruit samples; thus, the feature extraction of unripe fruits did not provide reliable results. However, since the harvesting robot only picked ripe fruits, the omission of unripe fruits and the incorrect detection of other objects as unripe fruits in the background did not affect the harvesting performance. In addition, some

TABLE 3. Comparison of R-YOLO and other methods of strawberry detection.

Algorithm	Recognition method	Fruit features	Recognition performance	Speed (image size)
Wei *et al.* (2014) [32]	machine vision(color)	Occlusion, Overlap, Proximity, Separation	accuracy rate: >95%	/
Wang *et al.* (2015) [51]	machine vision (color+shape)	Proximity, Separation	accuracy rate: 100%	0.633-0.886s (640×480)
Inkyu *et al.* (2016) [46]	Faster R-CNN	Occlusion, Overlap, Proximity, Separation	F1:0.838	/
Bargoti *et al.* (2017) [43]	ms-MLP+CNN	Occlusion, Overlap, Proximity, Separation	F1:0.858	/
Fu *et al.* (2018) [45]	LeNet	Occlusion, Overlap, Proximity, Separation	accuracy rate: 89.29%	0.27s (600×400)
Yu *et al.* (2019) [2]	Mask R-CNN	Occlusion, Overlap, Proximity, Separation	accuracy rate: 95.78%	0.125s (640×480)
Method proposed in this paper	R-YOLO	Occlusion, Overlap, Proximity, Separation	accuracy rate: 94.43%	0.056s (640×480)

of the ripe and unripe fruits were misidentified. The main reason was that the ripeness of these fruits is difficult to determine, even by humans. The model classification results were affected by human errors when labeling the training images. Furthermore, some image features could not be detected because of illumination, occlusion, or the camera angle, resulting in misidentification.

R-YOLO had lower recognition accuracy and recall than the original YOLO-V3. The primary reason was that the simplified lightweight network MobileNet used as the backbone network of R-YOLO reduced the weight parameters of the residual convolution network in the original YOLO by several times. Although the speed of the model was improved, the accuracy of feature extraction was adversely affected. However, it was found that the detection accuracy of R-YOLO was only 1.3% lower than that of the original YOLO-V3. Moreover, the simplified network structure of R-YOLO greatly improved the model running speed, which was 3.6 times faster than that of the original YOLO-V3. R-YOLO implemented on the TX2 processed 18 frames per second (FPS), demonstrating excellent real-time performance.

B. COMPARISON WITH OTHER DETECTION ALGORITHMS

We tested several fruit detection methods proposed in previous studies to compare and verify the advantages and disadvantages of R-YOLO. The results of the performance comparison are shown in Table 3.

As shown in Table 3, the proposed fruit detection method has not only high accuracy but also good real-time performance. Although the precisions of the algorithms proposed by Wei *et al.* and Wang *et al.* were higher than that of the R-YOLO algorithm, the two studies used machine vision algorithms that are not very robust and may not be stable in a changing environment. In addition, these studies identified the fruits by extracting a single feature or few features and did not express the spatial relationships between multi-level features; therefore, the studies resulted in poor recognition performance for multiple fruits, overlap, and occlusion. Inkyu *et al.*, Bargoti *et al.*, and Fu *et al.* also used deep learning models, resulting in high accuracy and good robustness. The recognition precision of Yu *et al.* was 1.35% higher than that of R-YOLO. However, the real-time performance of the above methods was not as good as that of R-YOLO and is not be applicable to embedded control terminals. Moreover, R-YOLO also generated rotated bounding boxes for fruit targets, which increased the localization precision of the picking points.

In Fig.9, the three images in the first row show several strawberries in close proximity. The image processing methods based on machine vision mistakenly identified multiple fruits in close proximity as a single target (Fig.9a) and were not able to separate the fruits. In addition, the strawberry marked by the black circle in Fig.9d was occluded by the stalk. The machine vision method misidentified this target as two separate fruits. Although the use of CNN models such as faster R-CNN, mask R-CNN, and YOLO avoided the problems of the machine vision algorithms (Fig.9b, e), these traditional target detection models could only generate horizontal target bounding boxes and could not determine the pose information of the fruits. The proposed R-YOLO model not only provided good detection performance in complex environments but also generated the rotated bounding boxes of the fruit targets, thereby improving the localization precision of the picking points (Fig.9c, f).

We designed several different image acquisition schemes for the images used as training sets to prevent over-fitting of the model. First, we obtained the strawberry images from the

FIGURE 9. Comparison of the proposed method and other object detection methods: a. Detection result of machine vision, b. Detection result of CNN, c. Detection result of R-YOLO, d. Detection result of machine vision, e. Detection result of CNN, f. Detection result of R-YOLO.

same planting bases and took images of the same varieties at different times. Second, images of different strawberry varieties were captured in the same period. Third, we downloaded a large number of strawberry images from the internet. These images were used to fine-tune the pre-trained model. The proposed model provided good detection results using a small number of labeled images. Since strawberry images were obtained in different environments, the model was able to learn the features of various strawberry fruits, and over-fitting of the model was avoided. It was found that the different nature environments and different strawberry varieties had little effect on the detection results, indicating that we used a sufficiently large number of training samples with varying environmental conditions.

C. EVALUATION OF THE DETECTION OF THE PICKING POINT

The rotated bounding box of the fruit target, which was the output of R-YOLO, was used to find the fruit axis. The coordinate of the intersection between the fruit axis and the short side at the top of the bounding box was calculated; this represented the vertex of the bounding box. The picking point was approximately 20 pixels away from this vertex. The prediction results of the picking points from the 573 ripe strawberries showed that the average error of the proposed localization method was $\pm$ 2 mm. The maximum error was approximately 4 mm, which mainly occurred in the location of picking points of some malformed or flattened fruits. A comparison of the methods for detecting the picking point showed that the proposed method using the rotated bounding box (Fig.10b) resulted in an error that was 50% less than that using the horizontal bounding box (Fig.10a). The R-YOLO

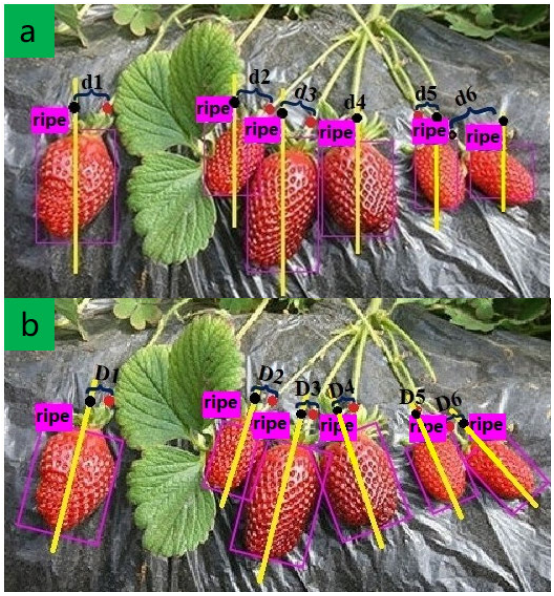


FIGURE 10. Comparison of localization methods for strawberry picking points: a. horizontal bounding box; b. rotated bounding box.

method resulted in relatively large errors for strawberries with an asymmetrical shape, such as deformities.

Each group of black and red dots in Fig.10 represents the predicted picking point and the ground truth, respectively. D_i and d_i (i = 1,2,3…6) represent the distance between the predicted picking point and the ground truth. D_i is generally smaller than d_i, indicating that the errors in predicting the picking point are lower for the rotated bounding boxes than that the horizontal bounding boxes.

To evaluate the impact of different picking point localization methods on the harvesting success rate,

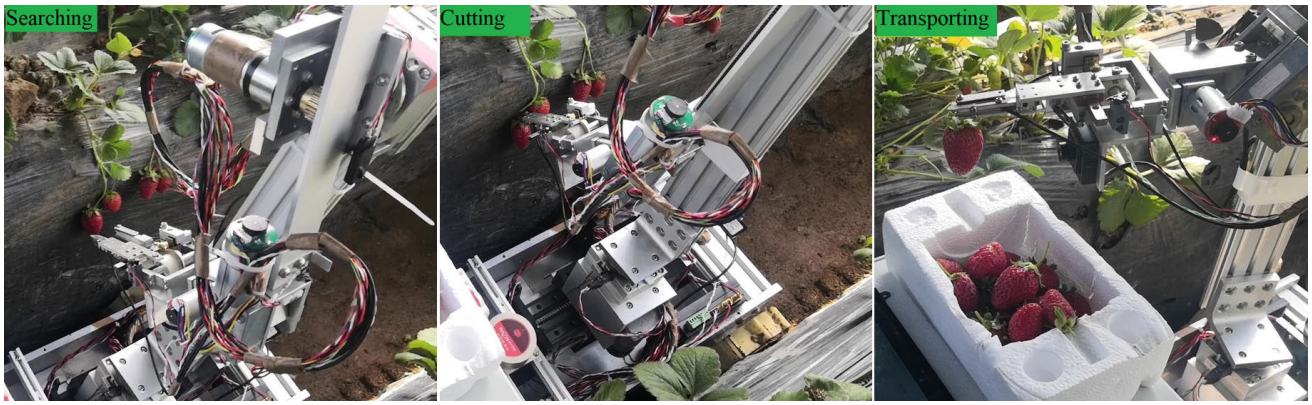


FIGURE 11. Action sequence of robotic harvesting operation in the field.

TABLE 4. Harvesting success rate with two localization methods.

Test number	Number of detected fruits	Number of harvested successfully	
		Original YOLO	R-YOLO
1	21	15	18
2	23	16	19
3	17	12	14
4	20	14	17
5	17	13	15
6	19	14	16
7	21	16	18
8	20	14	16
9	18	13	15
10	22	17	19
Accuracy		72.74%	84.35%

the proposed R-YOLO, and the original YOLO were both deployed for a comparison of the harvesting performance. Harvesting tests were conducted on a modified field that contained some isolated fruits and some multi-fruit adhesions. Also, some leaves were artificially removed to guarantee that the fruit stems were exposed and the occlusion area of the fruits did not exceed 50%. The results showed that the proposed method was indeed not effective at picking the fruits from the completely covered stems. For fruits with severe occlusion, the confidence scores predicted by R-YOLO were <60. At times the fruit was unrecognizable, and the robot gave up harvesting these fruits. Agronomy dictates that strawberry cultivation requires regular trimming of the branches and leaves at the top of the plant, which not only helps the fruit absorb nutrients and sunlight but also avoids fruit occlusion.

In this paper, a total of 10 groups of field tests were set up. In each test, the robot adopted two harvesting target detection models within the same scenario, and the success rate was used to evaluate the harvesting performance. While working, the hand-eye visual structure on the end-effector always faced the ridge surface, constantly searching for strawberry targets in the camera's view. After finding a target fruit, the robotic arm drove the end-effector to approach the fruit, cut the stem at a suitable place, transport it to the basket, and move to the next fruit. The action sequence of robotic harvesting operation in the field is shown in Fig.11. A harvesting process can only be defined as successful when the stem is cut off and the surface of the fruit is not damaged. The number of fruits successfully detected and harvested with the two target detection models in each test is recorded in Table 4.

It can be seen from Table 4 that within the same scenario, the harvesting success rate of R-YOLO is 84.35%, which is higher than that of the original YOLO model (72.74%). R-YOLO can be especially helpful in the harvesting of the strawberries that grow non-vertically downward and have an inclination angle of $\geq 45°$. The localization of the picking point predicted by the original horizontal bounding box is generally directly above the bounding box. However, the rotation bounding box generated by R-YOLO calculates the direction of the fruit axis, and then finds the picking point along the fruit axis, thus improving the positioning accuracy of the picking point.

VII. CONCLUSION

In this study, we designed a novel harvesting robot for the ridge-planting strawberry, and proposed the R-YOLO model for detecting the pose of strawberries in automatic harvesting. The proposed model achieved excellent robustness and real-time performance for the detection of fruits growing in various natural environments under varying light intensities and for multiple overlapping fruits. Moreover, the model could predict the rotation of the bounding box of the fruit target, which greatly improved the localization precision of the picking point. The following conclusions were drawn:

1) R-YOLO had good real-time performance while ensuring high accuracy of target detection. The overall P and R rates of 100 strawberry images were 94.43% and 93.46%. R-YOLO provided better detection performance and was more robust than traditional target detection methods based on machine vision for detecting strawberry fruits under different light intensities

and for multiple overlapping fruits. However, for those fruits whose area covered by leaves or other obstacles exceeds 50%, the confidence scores predicted by R-YOLO are lower than 60 or even unrecognizable, and the robot will give up harvesting these fruits. The target detection of the occlusion fruit is the focus of future research. Compared with other target detection models such as the Faster R-CNN, the accuracy of R-YOLO was not the highest, but the difference was relatively small. However, R-YOLO had the fastest running time and was 3.6 times faster than the original YOLO-V3. R-YOLO processed 18 images per second when implemented on TX2, meeting the real-time requirements of embedded controllers.

2) The rotated bounding box of the target generated by R-YOLO significantly improved the localization precision of the picking point. The localization results of 573 ripe fruits from 100 testing images showed that the average error was ± 2 mm, which met the error requirements of the end-effector of the harvesting robot. Meanwhile, it can be seen from the results of 10 sets of field harvesting tests that the robot implementing R-YOLO improved the localization accuracy with a harvesting success rate of 84.35% in modified situations. The primary reasons for the localization error were the curved stems of several strawberries and some malformed fruits that did not grow vertically. In a future study, we will increase the number of strawberry samples, optimize the model structure, and improve the performance for identifying the picking points. And we will also focus on determining priorities for fruit harvesting and optimal harvesting strategies. Moreover, the structure of the end-effector also needs to be upgraded. A temporary storage box or a conveyor belt should be added to provide continuous harvesting of multiple fruits.

REFERENCES

[1] Y. Xiong, Y. Ge, L. Grimstad, and P. J. From, "An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation," *J. Field Robot.*, vol. 37, no. 2, pp. 202–224, Mar. 2020.

[2] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit detection for strawberry harvesting robot in non-structural environment based on mask-RCNN," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104846.

[3] J. Linnan *et al.*, "A new type of facility strawberry stereoscopic cultivation mode," *J. China Agricult. Univ.*, vol. 24, no. 2, pp. 61–68, 2019.

[4] Y. Xiong, C. Peng, L. Grimstad, P. J. From, and V. Isler, "Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper," *Comput. Electron. Agricult.*, vol. 157, pp. 392–402, Feb. 2019.

[5] S. Hayashi, S. Yamamoto, S. Saito, Y. Ochiai, J. Kamata, M. Kurita, and K. Yamamoto, "Field operation of a movable strawberry-harvesting robot using a travel platform," *Jpn. Agricult. Res. Quart.*, vol. 48, no. 3, pp. 307–316, 2014.

[6] S. Yamamoto, S. Hayashi, S. Saito, Y. Ochiai, T. Yamashita, and S. Sugano, "Development of robotic strawberry harvester to approach target fruit from hanging bench side," *IFAC Proc. Volumes*, vol. 43, no. 26, pp. 95–100, 2010.

[7] Y. Cui, Y. Gejima, T. Kobayashi, K. Hiyoshi, and M. Nagata, "Study on Cartesian-type strawberry-harvesting robot," *Sensor Lett.*, vol. 11, no. 6, pp. 1223–1228, Jun. 2013.

[8] K. A. Vakilian, M. Jafari, and P. Zarafshan, "Dynamics modelling and control of a strawberry harvesting robot," in *Proc. 3rd RSI Int. Conf. Robot. Mechatronics (ICROM)*, Tehran, Iran, Oct. 2015, pp. 600–605.

[9] Y. Ge, Y. Xiong, G. L. Tenorio, and P. J. From, "Fruit localization and environment perception for strawberry harvesting robots," *IEEE Access*, vol. 7, pp. 147642–147652, 2019.

[10] Z. KaiLiang, L. Yang, L. Wang, L. Zhang, and T. Zhang, "Design and experiment of elevated substrate culture strawberry picking robot," *Trans. Chin. Soc. Agricult. Mach.*, vol. 43, no. 9, pp. 165–172, 2012.

[11] F. Qingchun, Z. Wengang, Q. Quan, J. Kai, and G. Rui, "Study on strawberry robotic harvesting system," in *Proc. IEEE Eng. CSAE*, May 2012, pp. 320–324.

[12] A. Silwal, J. R. Davidson, M. Karkee, C. Mo, Q. Zhang, and K. Lewis, "Design, integration, and field evaluation of a robotic apple harvester," *J. Field Robot.*, vol. 34, no. 6, pp. 1140–1159, Sep. 2017.

[13] F. Dimeas, D. V. Sako, V. C. Moulianitis, and N. A. Aspragathos, "Design and fuzzy control of a robotic gripper for efficient strawberry harvesting," *Robotica*, vol. 33, no. 5, pp. 1085–1098, Jun. 2015.

[14] C. W. Bac, E. J. van Henten, J. Hemming, and Y. Edan, "Harvesting robots for high-value crops: State-of-the-art review and challenges ahead," *J. Field Robot.*, vol. 31, no. 6, pp. 888–911, Nov. 2014.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2014.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016.

[20] J. Lei *et al.*, "Orientation adaptive YOLOv3 for object detection in remote sensing images," in *Proc. PRCV*, vol. 11857, 2019, pp. 586–597.

[21] L. Liu, Z. Pan, and B. Lei, "Learning a rotation invariant detector with rotatable bounding box," 2017, *arXiv:1711.09405*. [Online]. Available: https://arxiv.org/abs/1711.09405

[22] Y. Zhao, L. Gong, Y. Huang, and C. Liu, "A review of key techniques of vision-based control for harvesting robot," *Comput. Electron. Agricult.*, vol. 127, pp. 311–323, Sep. 2016.

[23] G.-Q. Jiang and C.-J. Zhao, "Apple recognition based on machine vision," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 2012, pp. 1148–1151.

[24] M. Rizon, N. A. N. Yusri, M. F. A. Kadir, A. R. B. Mamat, A. Z. A. Aziz, and K. Nanaa, "Determination of mango fruit from binary image using randomized Hough transform," in *Proc. 8th Int. Conf. Mach. Vis. (ICMV)*, Dec. 2015, vol. 9875, no. 3.

[25] J. Lu and N. Sang, *Detecting Citrus Fruits and Occlusion Recovery Under Natural Illumination Conditions*. Amsterdam, The Netherlands: Elsevier, 2015.

[26] A. Arefi, A. M. Motlagh, K. Mollazade, and R. F. Teimourlou, "Recognition and localization of ripen tomato based on machine vision," *Austral. J. Crop Sci.*, vol. 5, no. 10, pp. 1144–1149, 2011.

[27] J. P. Wachs, H. I. Stern, T. Burks, and V. Alchanatis, "Low and high-level visual feature-based apple detection from multi-modal images," *Precis. Agricult.*, vol. 11, no. 6, pp. 717–735, Dec. 2010.

[28] R. Linker, O. Cohen, and A. Naor, "Determination of the number of green apples in RGB images recorded in orchards," *Comput. Electron. Agricult.*, vol. 81, pp. 45–57, Feb. 2012.

[29] C. S. Nandi, B. Tudu, and C. Koley, "A machine vision-based maturity prediction system for sorting of harvested mangoes," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 7, pp. 1722–1730, Jul. 2014.

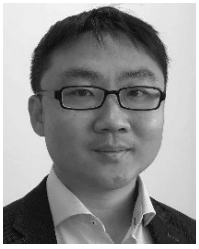
[30] A. Arefi and A. M. Motlagh, "Development of an expert system based on wavelet transform and artificial neural networks for the ripe tomato harvesting robot," *Austral. J. Crop Sci.*, vol. 7, no. 5, pp. 699–705, 2013.

[31] C. Ouyang, D. Li, J. Wang, S. Wang, and Y. Han, "The research of the strawberry disease identification based on image processing and pattern recognition," in *Proc. Int. Conf. Comput. Comput. Technol. Agricult.* Berlin, Germany: Springer, 2012.

[32] X. Wei, K. Jia, J. Lan, Y. Li, Y. Zeng, and C. Wang, "Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot," *Optik*, vol. 125, no. 19, pp. 5684–5689, Oct. 2014.

[33] S. Benalia, S. Cubero, J. M. Prats-Montalbán, B. Bernardi, G. Zimbalatti, and J. Blasco, "Computer vision for automatic quality inspection of dried figs (Ficus carica L.) in real-time," *Comput. Electron. Agricult.*, vol. 120, pp. 17–25, Jan. 2016.
[34] D. L. Borges, S. T. C. D. M. Guedes, A. R. Nascimento, and P. Melo-Pinto, "Detecting and grading severity of bacterial spot caused by Xanthomonas spp. in tomato (Solanum lycopersicon) fields using visible spectrum images," *Comput. Electron. Agricult.*, vol. 125, pp. 149–159, Jul. 2016.
[35] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2016, pp. 1440–1448.
[36] P. A. Dias, A. Tabb, and H. Medeiros, "Apple flower detection using deep convolutional networks," *Comput. Ind.*, vol. 99, pp. 17–28, Aug. 2018.
[37] L. Zhang, G. Gui, A. M. Khattak, M. Wang, W. Gao, and J. Jia, "Multi-task cascaded convolutional networks based intelligent fruit detection for designing automated robot," *IEEE Access*, vol. 7, pp. 56028–56038, 2019.
[38] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.
[39] G. Zeng, "Fruit and vegetables classification system using image saliency and convolutional neural network," in *Proc. IEEE 3rd Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Oct. 2017, pp. 613–617.
[40] T. Nishi, S. Kurogi, and K. Matsuo, "Grading fruits and vegetables using RGB-D images and convolutional neural network," in *Proc. IEEE Symp. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–6.
[41] Z. M. Khaing, Y. Naung, and P. H. Htut, "Development of control system for fruit classification based on convolutional neural network," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, Jan. 2018, pp. 1805–1807.
[42] L. Zhang, J. Jia, G. Gui, X. Hao, W. Gao, and M. Wang, "Deep learning based improved classification system for designing tomato harvesting robot," *IEEE Access*, vol. 6, pp. 67940–67950, 2018.
[43] S. Bargoti and J. P. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards," *J. Field Robot.*, vol. 34, no. 6, pp. 1039–1060, Sep. 2017.
[44] Y. Zhou, T. Xu, W. Zheng, and H. Deng, "Classification and recognition approaches of tomato main organs based on DCNN," *Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 15, pp. 219–226, 2017.
[45] L. Fu, Y. Feng, T. Elkamil, Z. Liu, R. Li, and Y. Cui, "Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks," *Trans. Chin. Soc. Agricult. Eng.*, vol. 34, no. 2, pp. 205–211, 2018.
[46] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "DeepFruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, no. 8, p. 1222, Aug. 2016.
[47] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Comput. Electron. Agricult.*, vol. 157, pp. 417–426, Feb. 2019.
[48] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861
[49] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3111–3122, Nov. 2018.
[50] Q. Yang, D. Xiao, and S. Lin, "Feeding behavior recognition for group-housed pigs with the faster R-CNN," *Comput. Electron. Agricult.*, vol. 155, pp. 453–460, Dec. 2018.
[51] L. Wang, L. Zhang, and Y. Duan, "Fruit localization for strawberry harvesting robot based on visual servoing," *Trans. Chin. Soc. Agricult. Eng.*, vol. 31, no. 22, pp. 25–31, 2015.

YANG YU is currently pursuing the Ph.D. degree with the College of Engineering, China Agricultural University, Beijing, China. His research interests include smart robotics, artificial intelligence technologies, object classification, target detection, semantic/instance segmentation, and image/video processing for the robot vision systems-based on deep learning and machine vision technologies.



KAILIANG ZHANG received the Ph.D. degree from China Agricultural University, in 2009. From 2009 to 2011, he was a Postdoctoral Fellow in industry and professional service robots with the Institute of Automation, Chinese Academy of Sciences. From 2014 to 2015, he held a postdoctoral position with the University of Ontario Institute of Technology, where he researched on high-performance parallel robotics, as well as rehabilitation robots. Since 2006, he has been engaged in key technologies of agricultural robots and involved in serials developing research on harvesting and grafting robots supported by the National Natural Science Foundation of China (NSFC). He currently works as an Associate Professor with China Agricultural University. His interests include intelligent agricultural equipment and robotics.

HUI LIU received the B.S. degree from China Agricultural University, Beijing, China, in 2019, where he is currently pursuing the master's degree with the College of Engineering. His research interests include mechanical arm behavior control and end-effector design for strawberry harvesting robots. He received the First Place in the 2018 China Agricultural Robot Competition for designing the orchard harvesting and storage robot.

LI YANG received the Ph.D. degree from China Agricultural University, in 2005. From 2012 to 2013, she was a Visiting Scholar with Iowa State University. She currently works as a Professor with China Agricultural University and a part-time expert in sowing and field management of the national edible bean industry technology system. Her research interests include intelligent agricultural equipment and agricultural robotics, including the development of high-efficiency precision seeding, pesticide application, harvesting, and other mechanized equipment. She has received the Second Prize of the National Science and Technology Progress Award, in 2019.

DONGXING ZHANG received the B.S. and S.M. degrees from China Agricultural University, in 1982 and 1991, respectively.

He is a Professor with China Agricultural University and a part-time Post Expert of the modern corn industry technology system, Ministry of Agriculture and Rural Affairs, China. He has studied in Italy, Israel, New Zealand. He has chaired and completed more than 20 scientific research projects, including the National Ninth Five-Year Research Project, the National Eleventh Five-Year, Twelfth Five-Year Support Plan Project, the 863 Project, the International Cooperation Projects, and the Horizontal Scientific Research Projects. His research achievements, such as Maize Precision Seeding Technology and Equipment and Vibration Subsoil Technology and Equipment designed by his team have been appraised as the international advanced level. He has published 12 national invention patents, five utility model patents, and four books. He has also published more than 100 academic papers (including more than 40 full-text SCI/EI searches) in leading domestic journals of agricultural machinery and agricultural engineering journals and international conferences. His current research interest includes full mechanization of corn production. He was a member and a Translator of the Conservation Farming Investigation Team of the Sino-US aScience and Technology Exchange Program. He has received the Second Prize of the National Science and Technology Progress Award, in 2019, the Second Prize of the Ministry of Agriculture's Scientific and Technological Progress Award, the Second Prize of the Beijing's Education and Teaching Achievements, and the Outstanding Innovation Team Award of the Ministry of Agriculture.