Computers and Electronics in Agriculture 211 (2023) 108036

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

NDMFCS: An automatic fruit counting system in modern apple orchard using abatement of abnormal fruit detection

Zhenchao Wu [a], Xiaoming Sun [a], Hanhui Jiang [a], Wulan Mao [a,f], Rui Li [a,d], Nikita Andriyanov [e], Vladimir Soloviev [e], Longsheng Fu [a,b,c,d,*]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[c] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China
[d] Northwest A&F University Shenzhen Research Institute, Shenzhen, Guangdong 518000, China
[e] Department of Data Analysis and Machine Learning, Financial University under the Government of the Russian Federation, 125167 Moscow, Russia
[f] Institute of Agricultural Mechanization, Xinjiang Academy of Agricultural Sciences, Urumqi 830000, China

ARTICLE INFO

Keywords:
Apple orchard
Abatement of abnormal fruit detection
Object detection
Fruit counting
ID assignment

ABSTRACT

Automatic fruit counting is an important task for growers to estimate yield and manage orchards. Although many deep-learning-based fruit detection algorithms have been developed to improve performance of automatic fruit counting systems, abnormal fruit detection has often been caused by these algorithms detecting non-target fruits that have similar growth characteristics to target fruits. For abnormal fruit detection, detected fruits in the back row of the tree were defined as DFBRT, while detected fruits on the ground were defined as DFG. Both of them would result in a higher number of fruits counting than the ground truth. This study proposes an automatic fruit counting system called NDMFCS (Normal Detection Matched Fruit Counting System) to solve this problem for improving fruit counting accuracy in modern apple orchard. NDMFCS consists of three sub-systems, i.e. object detection based on You Only Look Once Version 4-tiny (YOLOv4-tiny), abatement of abnormal fruit detection based on threshold, and fruit counting based on trunk tracking and identity document (ID) assignment. YOLOv4-tiny was selected to implement detection of fruits and trunks, whose output is confidence and pixel coordinates of detected object. The DFBRT and DFG were abated by thresholds to improve detection performance of fruit. This meant that detected fruits were removed when their distance from camera is further than a distance threshold or the confidence of fruit detection is less than a confidence threshold. Finally, fruit counting was implemented by trunk tracking and ID assignment, where each fruit was assigned a unique tracking ID. Results on 10 sets of original videos indicated that average fruit detection precision was improved from 89.1% to 93.3% after abatement of abnormal fruit detection. Also, Multiple Object Tracking Accuracy and Multiple Object Tracking Precision were improved on average by 4.2% and 3.3%, respectively, while average ID Switch Rate was decreased on average by 1.1%. And average fruit counting accuracy was improved to 95.0% by 4.2%. Coefficient of determination (R^2) was 0.97, which indicated the number of fruits counted by NDMFCS was near to the ground truth. These results demonstrate that the abatement of abnormal fruit detection can improve performance of apple counting, which has the potential to provide a technical support for estimating fruit yield in modern apple orchards.

1. Introduction

Automatic counting of fruit is an important task for growers to estimate yield and manage orchards. For modern farms such as apple orchards, fruit numbers could help managers to determine some key decisions, which include the number of pickers, the number of storage bins and marketing strategies (He et al., 2022a; Villacrés et al., 2023; Wu et al., 2022; Zhang et al., 2022). In the absence of an automatic fruit counting system, fruits in orchards are usually counted manually by growers, which is tedious and time-consuming (Bhattarai and Karkee, 2022; Häni et al., 2020; Jiang and Li, 2020; Rong et al., 2023; Shen et al., 2023). Therefore, there is an increasing demand for automatic fruit

* Corresponding author at: College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China.
E-mail address: fulsh@nwafu.edu.cn (L. Fu).

https://doi.org/10.1016/j.compag.2023.108036
Received 1 May 2023; Received in revised form 29 June 2023; Accepted 30 June 2023
Available online 6 July 2023

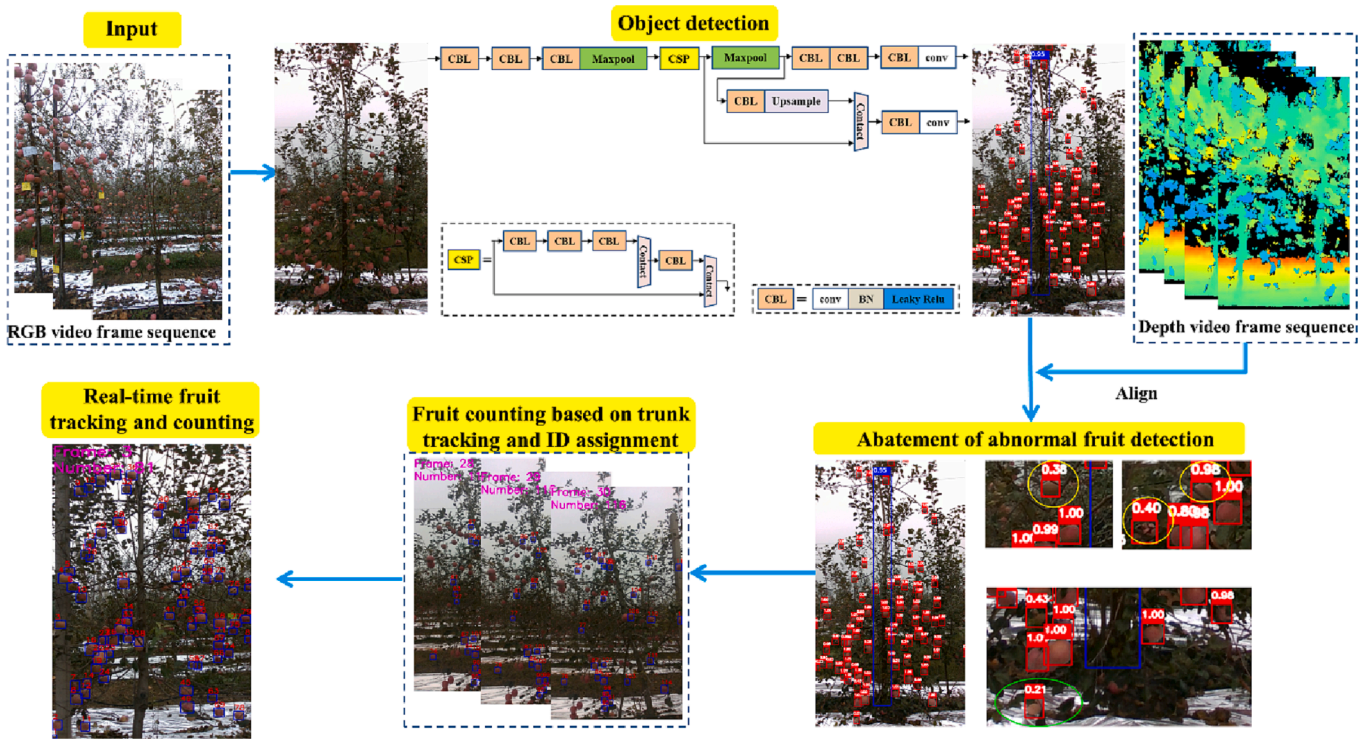


Fig. 1. Overall structure of NDMFCS.

counting systems, which enable accurate and reliable fruit counting in modern orchards before harvesting.

Typically, automatic fruit counting systems in modern apple orchards have mostly been developed for videos of tree-row and mainly included fruit detection and tracking. In modern apple orchards, tree canopies are pruned to create desired narrow canopy architectures with dwarf trees, thus creating tree-row with fruiting wall structure (Bhusal et al., 2022; Majeed et al., 2020), which are not often adequately covered by a single image to achieve counts of all fruit. Therefore, videos of tree-row, where the same fruit is detected and tracked to avoid repeated counts for fruit counting, have been recorded. Osman et al. (2021) achieved a counting accuracy of more than 90.6% by improving DeepSORT tracking network to track all detected fruits. He et al. (2022b) developed a robust fruit counting system by combining object detection with You Only Look Once (YOLO) Version 3, Kalman filter, and cascade matching, which reached a counting accuracy of 93.8%. Zhang et al. (2022) proposed an OrangeYOLO-based fruit counting system to overcome double-counting of the same fruit in videos, which reached a Mean Absolute Error of 0.081. Although many fruit detection or tracking algorithms have been proposed to improve performance of automatic fruit counting systems, high-precision fruit detection algorithms usually receive more attention from researchers, as it is a prerequisite for good tracking performance.

Many deep-learning-based fruit detection algorithms have been developed to improve the performance of automatic fruit counting systems. Due to complex backgrounds, variability of light in modern orchards, and several other factors lead to challenges in detecting fruit for automatic fruit counting systems (Abeyrathna et al., 2023; Vasconez et al., 2020; Wang et al., 2022; Wu et al., 2023; Zheng et al., 2022). YOLO, as one of the top deep-learning-based object detection networks, has been employed to address these challenges to develop real-time object detection models with superior performance (Biffi et al., 2021; Wan et al., 2022). Many variant YOLO networks have been developed based on size, color, cluster density, and other growth characteristics of different fruits, such as YOLO-Banana (Fu et al., 2022), YOLO-Grape (Li et al., 2021), MangoYOLO (Koirala et al., 2019), and OrangeYOLO (Zhang et al., 2022). Although these variant YOLO networks have improved detection accuracy for specific fruit varieties, it is still difficult to distinguish fruits in the back row of trees from those in the front row of trees based on growth characteristics.

Fruits in the back row of trees may be detected to result in a larger number of fruit than the ground truth, which is difficult to be resolved by improving fruit detection algorithms. Trees in modern apple orchards are planted in rows, whereas fruits in the back row of trees have always been detected by appearing in gaps of the front row of trees and need to be abated. Osman et al. (2021) developed a lightweight thresholding mechanism to correct detections returned by YOLO, which set height and width thresholds of rectangular boxes to determine whether detected fruit belongs to the back row of trees. However, due to front and rear extension of fruit tree branches and size difference of fruit, the size of the rectangular box cannot represent whether detected fruit belongs to the back row of trees or not (Fu et al., 2020). Therefore, it is more reasonable to determine whether detected fruit belongs to the back row of trees based on the distance between the fruit and camera rather than the size of the rectangular box.

In this study, a video-based automatic fruit counting system is proposed, called NDMFCS (Normal Detection Matched Fruit Counting System), which combines abatement of abnormal fruit detection with threshold. Remainder of this paper is structured as follows. Section 2 provides data sources and research methods employed to develop NDMFCS. Section 3 reports experimental results, followed by discussions of results and a comparison of performance with and without abatement of abnormal fruit detection. Conclusions are provided in Section 4, where limitations for practical applications of NDMFCS and future works are finally presented.

2. Materials and methods

This study develops an automatic fruit counting system for yield estimation in modern apple orchards based on videos. The overall structure of NDMFCS, as shown in Fig. 1, takes RGB (Red, Green, and Blue) as input of YOLOv4-tiny, where fruits and trunks in RGB video

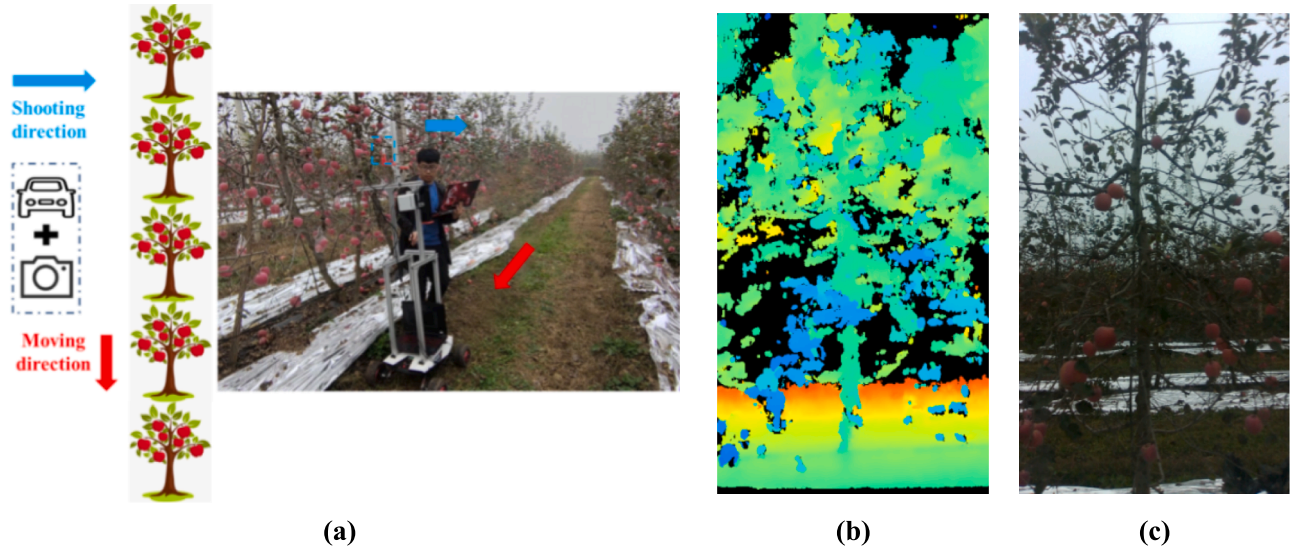


Fig. 2. Data acquisition in the experimental orchard. **(a)** Original images and videos were obtained by controlling the remote-controlled vehicle; Camera was about 1.4 m above the ground and had a distance of 2.0 m to tree-row; Blue arrow refers to shooting direction of the camera, while red arrow refers to moving direction of the remote-controlled vehicle; **(b)** A depth video frame, which will be aligned to **(c)** An RGB video frame. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

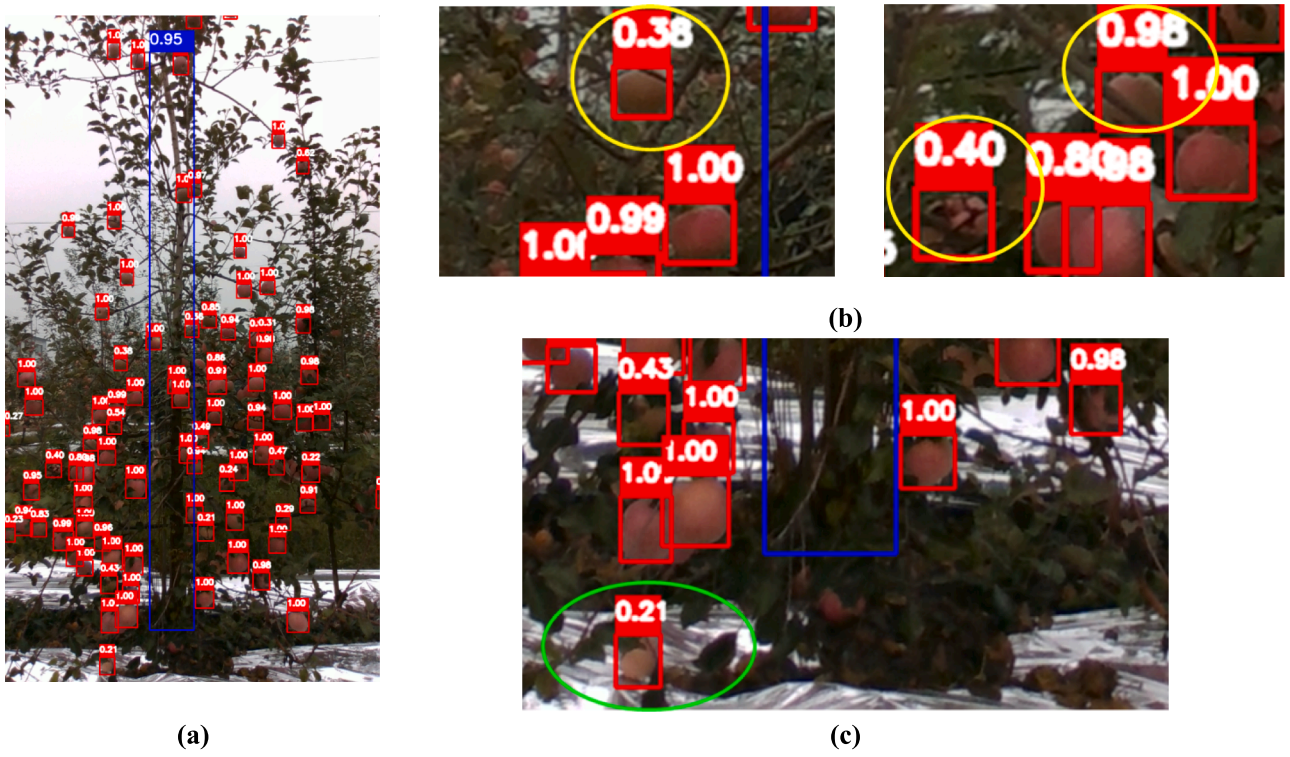


Fig. 3. Examples of abnormal fruit detection in an image of testing dataset, where the detected fruit and trunk are marked with red and blue rectangles, respectively. Numbers above rectangles represent confidence threshold of detection. **(a)** Example of fruit detection; **(b)** The DFBRT is marked with yellow circle; **(c)** The DFG is marked with green circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were detected. For abnormal fruit detection, detected fruits in the back row of the tree were referred to as DFBRT, while detected fruits on the ground were referred to as DFG. Then, the depth map was aligned to the RGB image to abate the DFBRT based on distance threshold. The DFG was abated based on the confidence of fruit detection. Finally, fruit counting was implemented based on trunk tracking and identity document (ID) assignment, which assigned a tracking ID to each fruit and correlated the same fruit in consecutive video frames.

2.1. Dataset preparation

2.1.1. Data acquisition and annotation

Datasets (i.e. original images and original videos) for experiment were collected at an orchard densely planted with dwarf crowns in Famen Town, Baoji City, Shaanxi Province, China (34°29′41″ N latitude, 107°51′40″ E longitude). Apple variety of this orchard is 'Fuji' with inter-plant of 1.5 m and inter-row spacing of 4.0 m. Field imaging platform

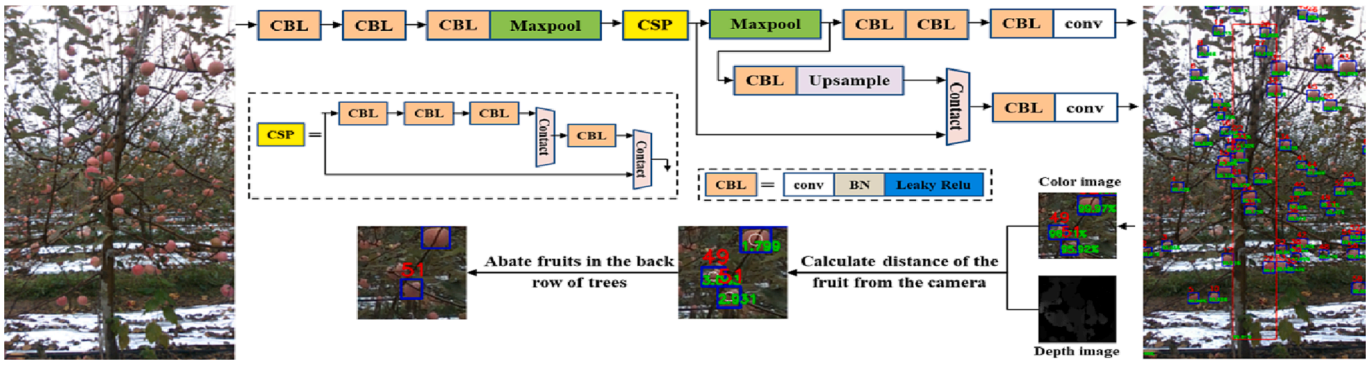


Fig. 4. The abatement process of the DFBRT, where detected fruits and trunks are marked with blue and red rectangles respectively; Green number in color image is confidence threshold of fruit detection, which is distance of the fruit from the camera after calculating distance of the fruit from the camera; Red number represents tracking ID of fruit. Note: conv is convolution; BN is stand for Batch Normalization; Leaky ReLU is the activation function; Maxpool indicts the maximum pool. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

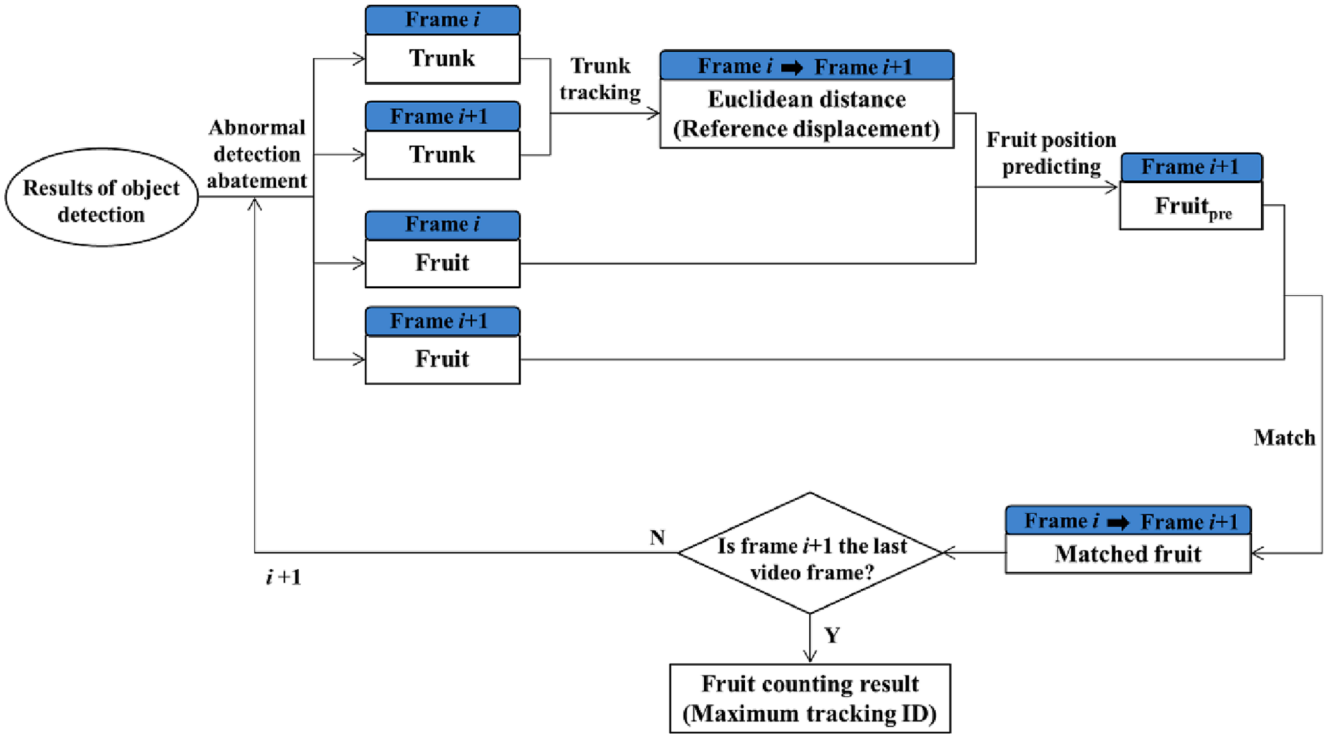


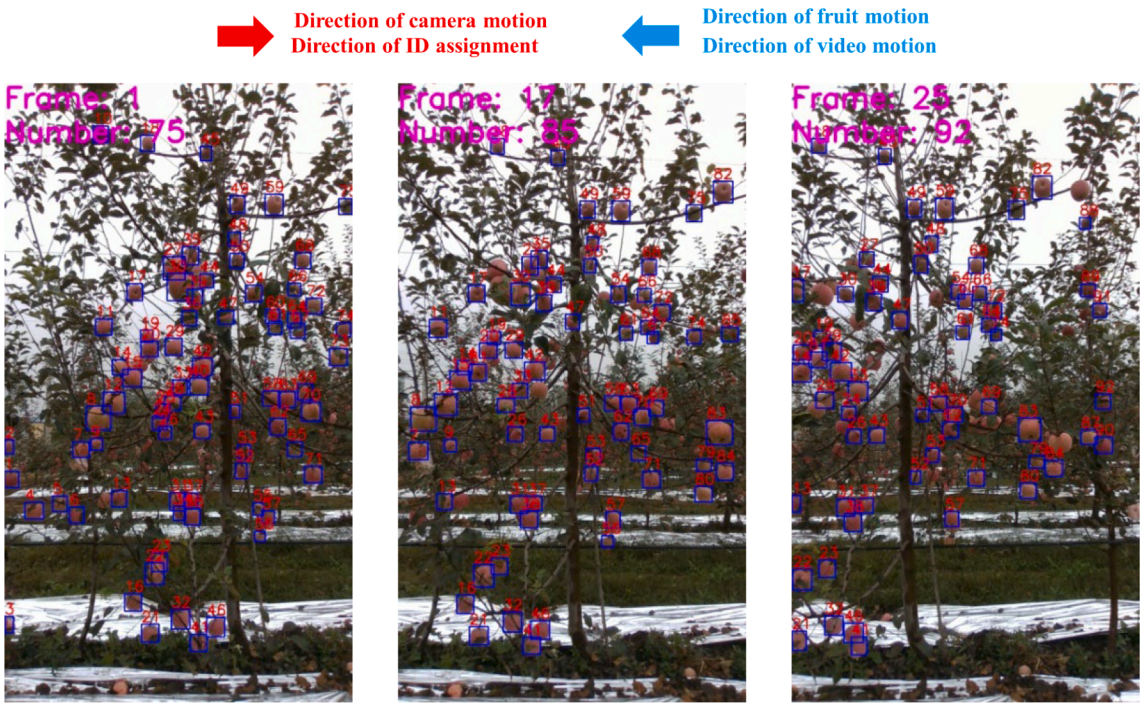
Fig. 5. Fruit counting flow based on trunk tracking and ID assignment, whose input is results of object detection after abnormal detection abatement, while its output is fruit counting result; Frame i or frame $i + 1$ in the blue box refers to the RGB video frame, which starts with the first RGB video frame. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

consisted of aluminum frames equipped with an Intel RealSense D435 camera, which was installed on a remote-controlled vehicle which moved at a speed of around1.0 m/s along the tree-row, as shown in Fig. 2**(a)**. A total of 1600 original images and 10 sets of original videos (i.e. V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}) both with a resolution of 720 × 1280 pixels were acquired in two consequent harvesting seasons of 2020 and 2021. Each set of original videos contained an RGB video and a corresponding depth video, where each depth frame was aligned to the RGB frame, as shown in Fig. 2**(b)** and Fig. 2**(c)**.

2.1.2. *Dataset building*

An original dataset was constructed using all original images for training and testing object detection model. The total of 1600 original images were randomly divided into a training dataset (80% of the images, 1280 images) and a testing dataset (20% of the images, 320 images). For each original image, fruits and trunks were manually labeled using rectangular boxes with different class labels, which was the same as Gao et al. (2022). A total of 93,050 samples were labeled, of which the number of 'fruit' and 'trunk' samples were 90,798 and 2252, respectively. Performance of NDMFCS was tested with the 10 sets of original videos.

Data augmentation was implemented to enrich the training dataset to help improve the overall learning and performance of object detection model. Data augmentations, i.e. motion blur transformation, contrast transformation, image mirroring in horizontal axis brightness transformation, Gaussian blur transformation, and sharpness transformation, were applied. The training dataset was augmented from the 1280 to 11,520 images and employed as an augmented training dataset to train object detection model.



(a) 1st frame **(b)** 17th frame **(c)** 25th frame

Fig. 6. Examples of fruit counting in different frames. Red arrow at the top represents directions of camera motion and ID assignment, while blue arrow represents directions of fruit and video motion; Numbers above blue rectangles refer to tracking IDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Detection results on the test dataset.

Object	*P*/%	*R*/%	*AP*/%	*mAP*/%	Speed/(ms/image)
Fruit	86.2	94.4	94.2	96.4	16
Trunk	99.1	98.7	98.6		

2.2. Object detection based on YOLOv4-tiny

YOLOv4-tiny was selected to implement detection of fruits and trunks, whose input is continuous RGB video frames, while its output is confidence and pixel coordinates of the detected object. YOLO light network has been widely used in object detection because of its good performance (Mirhaji et al., 2021; Tan et al., 2022). Gao et al. (2021) achieved an average accuracy of 94.5% and a detection speed of 18 ms for a dataset of 'Fuji' apple variety, leading to the selection of YOLOv4-tiny for its high detection accuracy and real-time speed.

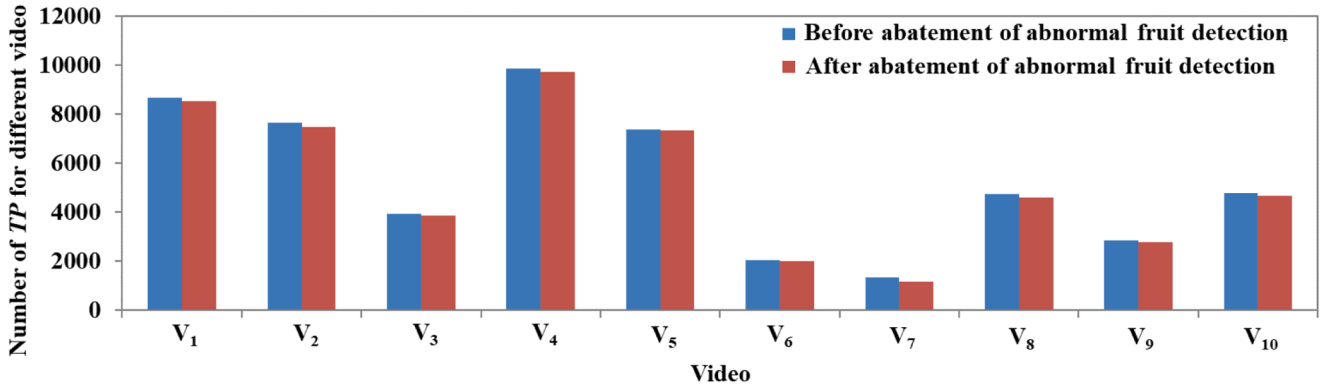
2.3. Abatement of abnormal fruit detection based on threshold

Abatement of abnormal fruit detection was implemented to improve detection performance and provide more accurate fruit number for counting. Abnormal fruit detection (i.e. the DFBRT or the DFG) was shown in Fig. 3. Detected in the back row of trees and on the ground would result in a higher number of fruit counting than the ground truth, which were abated by thresholds.

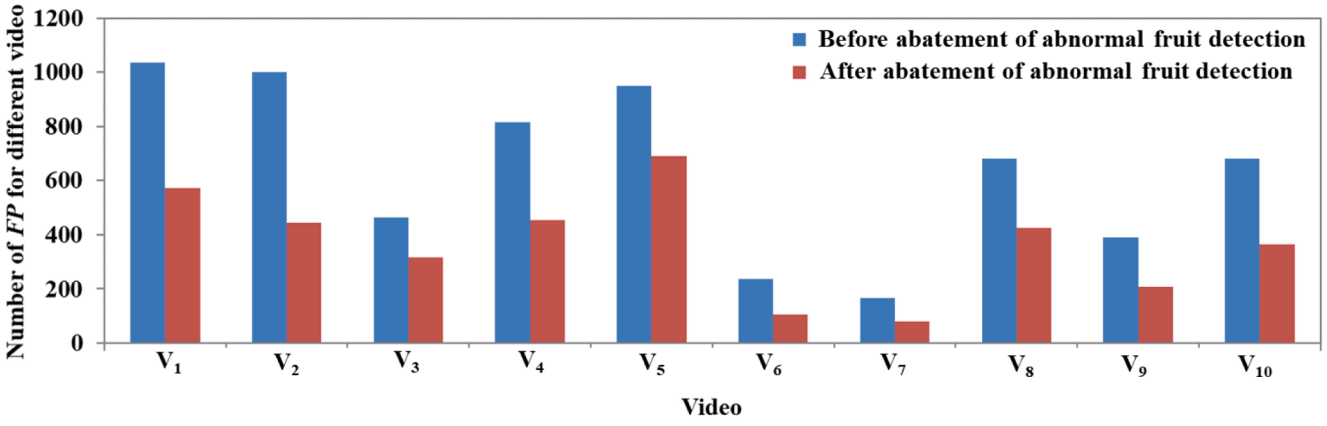
Abnormal fruit detection was abated by thresholds, which meant that detected fruits will be removed when their distance from the camera is further than a distance threshold or the confidence of fruit detection is less than a confidence threshold. The distance of fruits in the back row of trees from camera is much further than that of fruits in the front row of trees from the camera. Hence, abatement of the DFBRT was implemented by setting the distance threshold, as shown in Fig. 4. First, fruits in the RGB video frame were detected based on object detection model. Then, the depth video frame was aligned to the RGB video frame to extract the distance from pixels of the RGB video to the camera. Finally, since the distance between the camera and trunk and the maximum length of the branches along the shooting direction are about 2.0 m and 1.0 m, respectively, detected fruits with a distance between them and the camera of more than 3.0 m would be removed. Since few pixels in the depth video frame may be lost, the average of the distances of all pixels in the circle rather than the distances of individual pixels was represented as the distance from the fruit to the camera. This circle was draw with the center pixel of the fruit in the RGB video frame as the center and half the width of the detection rectangle of the fruit as radius. The average of the distances of all pixels in this circle represented the distance from the fruit to the camera. Besides, due to detected fruits on the ground generally have low confidence of fruit detection, confidence threshold of fruit detection was set to abate detected fruits on the ground, which would be removed when the confidence of fruit detection is less than 0.4 (Gao et al., 2021).

2.4. Fruit counting based on trunk tracking and ID assignment

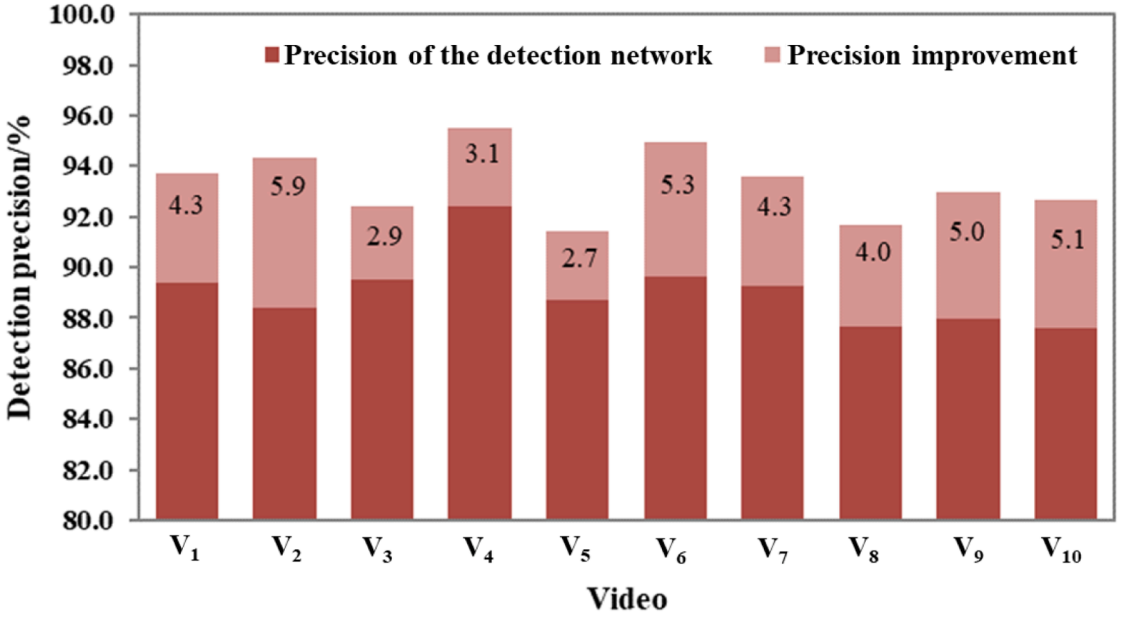
Fruit counting was implemented based on trunk tracking and ID assignment, which assigned a tracking ID (No. 1, No. 2, No. 3, No. 4, No. 5, etc.) to each fruit and correlated the same fruit in consecutive RGB video frames. As shown in Fig. 5, due to the relatively static position between trunks and fruits in modern apple orchard, displacement of the same trunk between consecutive RGB video frames was employed as a reference displacement to predict fruit position. In our previous work (Gao et al., 2022), although a correlation filter-based on trunk tracking method has been developed to improve the speed of fruit counting, it is only applicable to modern apple orchards with vertical fruiting-wall architecture. In modern apple orchards with vertical fruiting-wall



(a)

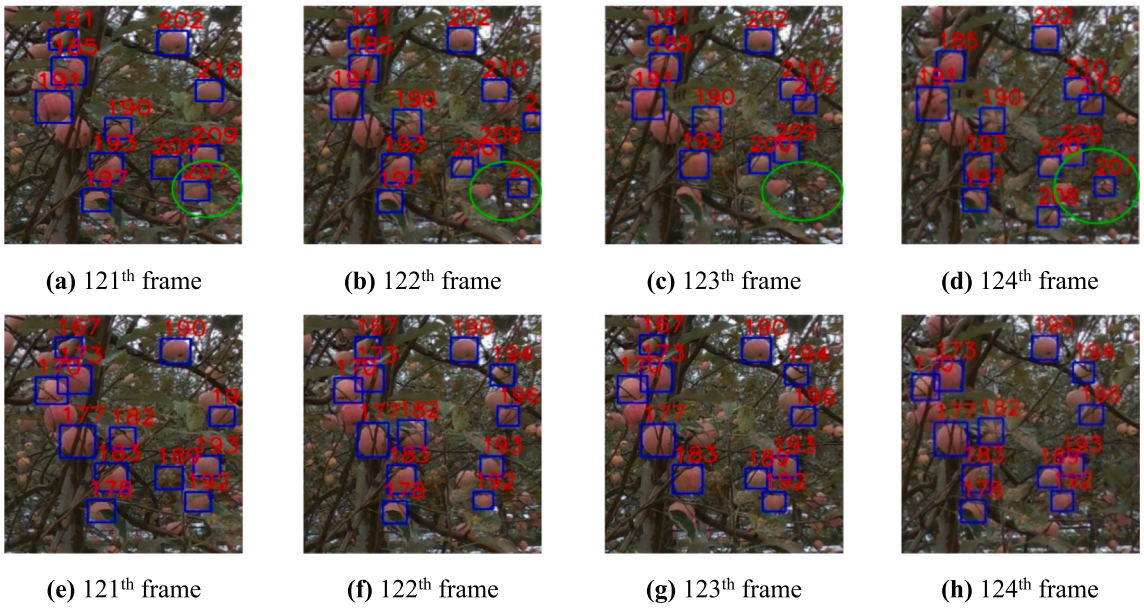


(b)



(c)

Fig. 7. Results of fruit detection before and after abatement of abnormal fruit detection in original videos. **(a)** and **(b)** are *TP* and *FP* numbers of fruits detected, respectively; **(c)** Improvement of fruit detection precision after abatement of abnormal fruit detection; The number on the column is an improved value of detection precision (%).



(a) 121th frame **(b)** 122th frame **(c)** 123th frame **(d)** 124th frame

(e) 121th frame **(f)** 122th frame **(g)** 123th frame **(h)** 124th frame

Fig. 8. Example of fruit tracking before and after abatement of abnormal fruit detection for four consecutive RGB video frames starting from the 121th frame in V_2; **(a), (b), (c),** and **(d)** are results of fruit tracking before abatement of abnormal fruit detection, while **(e), (f), (g),** and **(h)** are results of fruit tracking after abatement of abnormal fruit detection (The green circle highlights fruit where tracking error occurred). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

architecture, trunks of fruit trees are straight and have obvious features and could be accurately tracked based on the correlation filter. However, in other densely planted orchards with dwarf crowns, trunks often struggle to stay straight due to a large number of fruits and lack of enough wire restraints, making it difficult for the correlation filter to predict the area of the trunk. Hence, the same trunk in consecutive RGB video frames (frame i and frame $i+1$) was tracked using a minimum Euclidean distance (*Ed*) in two-dimensional space, which is calculated in Eq. (1) and regarded as the reference displacement. In NDMFCS, tracked trunk in first RGB video frame was first trunk detected in the direction of video motion and was subsequently transformed into another detected trunk when currently tracked trunk disappears from the RGB video frame. When transforming tracked trunk, average reference displacement of previous five RGB video frames will be as reference displacement. Fruit prediction in frame $i+1$ ($\mathbf{Fruit_{pre}}$), which was implemented based on the reference displacement and detected fruit in frame i, was matched using the minimum *Ed* to detected fruit in frame $i+1$ for fruit tracking. And tracked fruit was considered as the same fruit in consecutive RGB video frames and assigned a tracking ID. As shown in Fig. 6, the direction of ID assignment is the same as direction of camera motion and opposite to direction of fruit motion. The maximum tracking ID in the last RGB video frame will be fruit counting result.

$$Ed = \sqrt{(x^{i+1} - x^{i})^2 + (y^{i+1} - y^{i})^2} \tag{1}$$

where x^i and y^i refer to the horizontal and vertical coordinates of the center difference of detection rectangle for RGB video frame i, respectively; The range of i is from 1 to n_v, which is determined by the number of RGB video frames in each set of original videos; The n_v of V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10} were 135, 134, 105, 201, 171, 84, 45, 115, 52, 101 respectively.

2.5. Evaluation indicators

The performance of object detection model was evaluated using detection speed, average precision (*AP*), and mean average precision (*mAP*). Precision (*P*) and recall (*R*) of object detection model were calculated in Eq. (4) and Eq. (5). The *AP*, an indicator that reflects the performance of object detection model (Gao et al., 2020), was defined in Eq. (6) by *P* and *R*. The *mAP*, the mean of *AP* values of fruit and trunk detection, was calculated in Eq. (7).

Three more indicators, i.e. ID Switch Rate (*IDSR*), Multiple Object Tracking Accuracy (*MOTA*), and Multiple Object Tracking Precision (*MOTP*) were proposed to evaluate fruit tracking performance of NDMFCS, as shown in Eqs. (8) - (10). Additionally, counting accuracy (P_c) was defined in Eq. (11) to assess fruit counting performance of NDMFCS (Gao et al., 2022).

$$P = TP/(TP + FP) \tag{4}$$

$$R = TP/(TP + FN) \tag{5}$$

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{6}$$

$$mAP = \frac{1}{2}(AP_{\text{fruit}} + AP_{\text{trunk}}) \tag{7}$$

$$IDSR = \frac{Q_{\text{sh}}}{S} \tag{8}$$

$$MOTA = \frac{M_{\text{mat}}}{S} \tag{9}$$

$$MOTP = \frac{M_{\text{mat}}}{T_{\text{mat}}} \tag{10}$$

$$P_{\text{c}} = \left(1 - \frac{|S - G_{\text{manu}}|}{G_{\text{manu}}}\right) \times 100\% \tag{11}$$

where *TP*, *FP*, and *FN* mean true positive, false positive, and false negative, respectively. Q_{sh} and S refer to the number of fruits with ID switching and the total number of tracked fruits in the video, respectively; AP_{fruit} and AP_{trunk} refer to average precision of the fruit and trunk detection, respectively; M_{mat} and T_{mat} refer to the number of correctly tracked fruits and the total number of matched fruits in the video, respectively; G_{manu} refer to the ground truth in the video, which is the average of visual manual counts by three different operators.

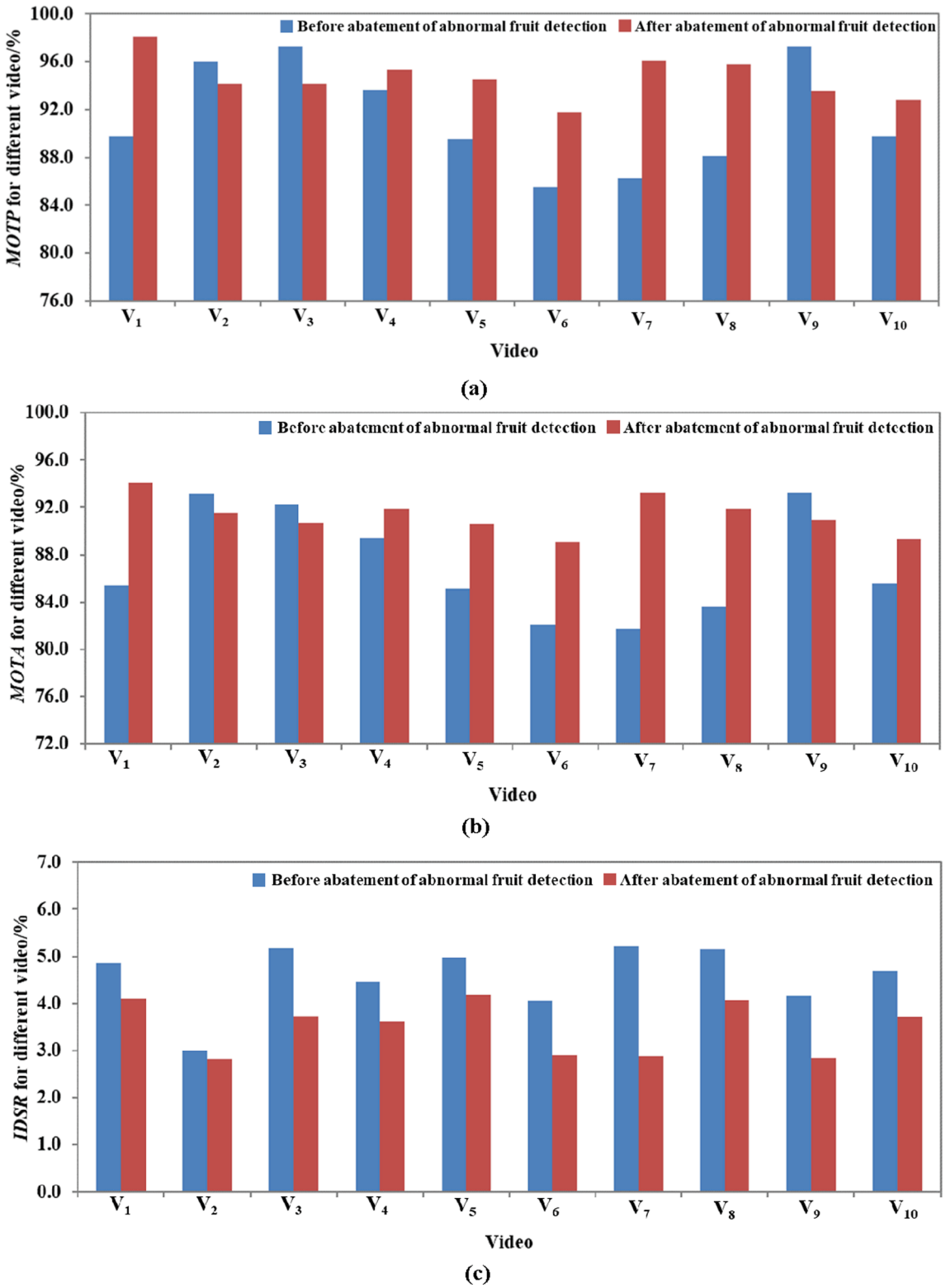


Fig. 9. Results of fruit tracking before and after abatement of abnormal fruit detection in original videos. **(a), (b),** and **(c)** are *MOTP*, *MOTA*, and *IDSR*, respectively.

2.6. Computing hardware and related settings

Network training and testing were implemented on a desktop computer with following specifications: Intel® Core™ i5-6400 CPU @ 2.70 GHz, NVIDIA GTX 1080 8 GB GPU, 16 GB of RAM, 64-bit Windows 10, CUDA 9.0, cuDNN 7.1.3, OpenCV 3.1.0, Python 3.6, CMake-3.16. The input size was set to 416 × 416. The learning rate, batch size and epoch were set to 0.001, 64 and 500, respectively. Stochastic gradient descent was used to iteratively optimize network parameters. And transfer learning technique was applied to train the network involved training on the COCO dataset (Lin et al., 2014).

For in-field application, a laptop computer was implemented with following specifications: Intel® Core™ i7-8565U CPU @ 1.80 GHz, 8 GB of RAM, NVIDIA GeForce MX250 2 GB GPU, 64-bit Windows 10, CUDA

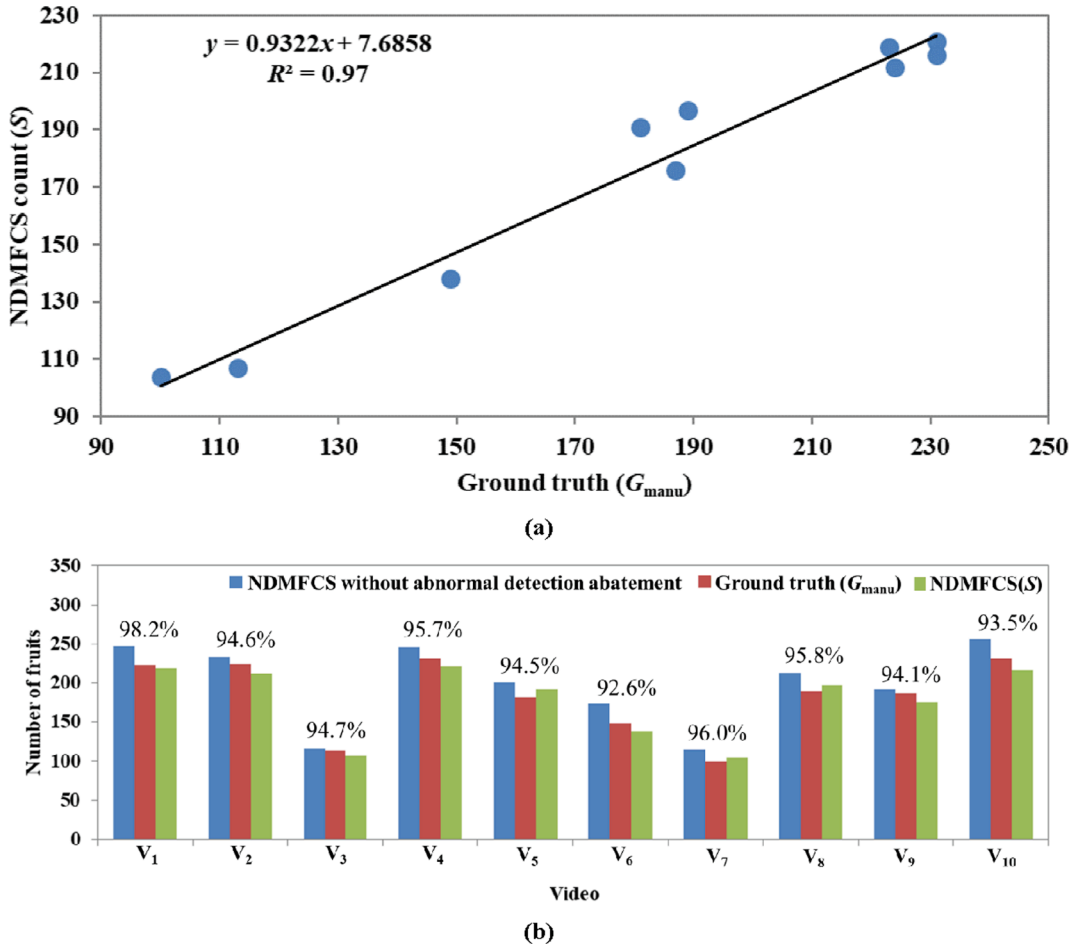


Fig. 10. Difference between the number of fruits obtained by manual and NDMFCS. **(a)** A linear regression between the number of fruits counted by NDMFCS and the number of fruits counted by manual vision; Each dot in the chart represents the maximum number of tracking IDs acquired in the last frame of a video found by NDMFCS in *x*-axis (*S*), and the summation of the visual manual counts found by the inspectors in the *y*-axis (G_{manu}); **(b)** Results of fruit counting for original videos by NDMFCS with or without abatement of abnormal fruit detection, which are compared with the ground truth; The P_c of NDMFCS with abatement of abnormal fruit detection is shown on the column.

10.0, cuDNN 7.6.5, OpenCV 4.4.0.

3. Results and discussions

3.1. Performance of object detection after abatement of abnormal fruit detection

The performance of object detection model for detecting both fruits and trunks in modern apple orchards was good, and its improvement was prevented by abnormal fruit detection. The performance of object detection model was tested with 320 images at the confidence threshold of 0.25, which contained 17,724 'fruit' and 451 'trunk' objects of interest. Object detection model took only 16 ms on average to process an image of a resolution of 720 × 1280 pixels, indicating that object detection model quickly detects the 'fruit' and 'trunk'. For the 'fruit', object detection model successfully detected 16,723 *TP*, with *FP* of 2,684 and *FN* of 1,001 (AP_{fruit} of 94.2%), as shown in Table 1. For the 'trunk', object detection model successfully detected 445 *TP*, with *FP* of 4 and *FN* of 6 (AP_{trunk} of 98.6%). Although these high *AP* made the *mAP* reach a high value of 96.4%, the large *FP* (abnormal fruit detection) number of fruits led to precision of object detection model for the 'fruit' was below 90% (86.2%), which prevents accurate fruit tracking and counting. Therefore, it is necessary to investigate performance changes of object detection model after abatement of abnormal fruit detection.

The performance of object detection model for detecting fruits was improved after abatement of abnormal fruit detection. As shown in Fig. 7, numbers of *TP* and *FP* after abatement of abnormal fruit detection were decreased on average by 3.0% and 44.1%, respectively, which resulted in the average fruit detection precision of object detection model for original videos was improved from 89.1% to 93.3%. Variant YOLO networks have been developed based on the growth characteristics of different fruits, which achieve good performance when the difference between the target to be detected and the background is obvious. However, they often fail to reach good detection performance when their background contains fruits in the back row of trees that are similar to fruits in the front row of trees (Bao et al., 2023; Cong et al., 2023; Zeng et al., 2023). Instead, it seems more sensible to abate fruits in the back row of trees based on the distance threshold (Fu et al., 2020; Jiang and Li, 2020), which provides a more reliable detection results for fruit tracking and counting.

3.2. Performance of fruit tracking and counting

Fruit tracking performance of NDMFCS was degraded due to the DFBRT and was improved after abatement of abnormal fruit detection. Compared to fruits in the front row of trees, fruits in the back row of trees are further away from the camera, which results in differences of motion parallax and causes tracking errors (Liu et al., 2019; Mizushina et al., 2020). As shown in Fig. 8, the displacement of detected fruits in the front row of trees between consecutive RGB video frames is larger

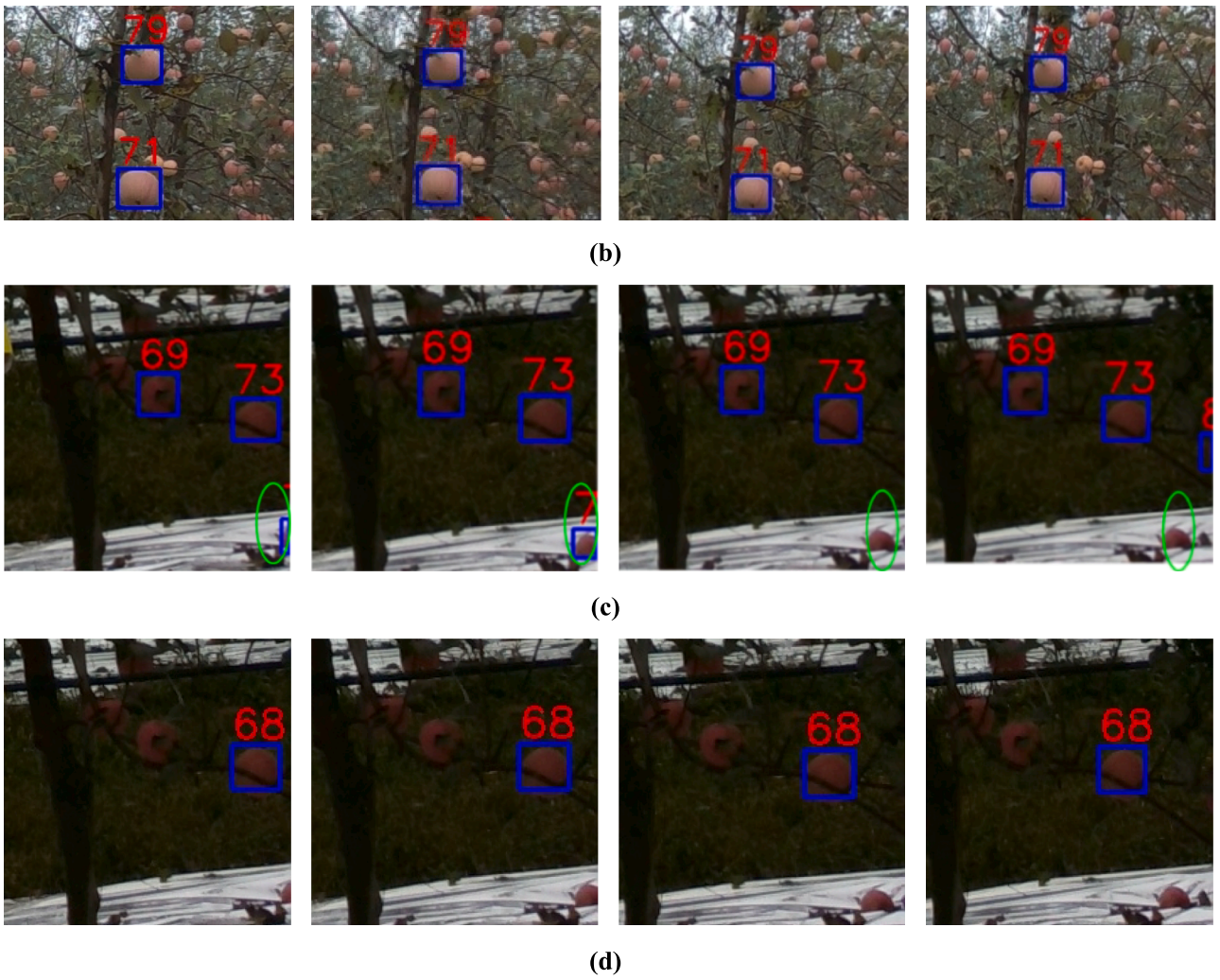


Fig. 11. Example of fruit counting before and after abatement of abnormal fruit detection for original videos. **(a)** Results of fruit counting before abatement of abnormal fruit detection for four consecutive RGB video frames starting from the 29th frame in V_2, which includes a fruit counting error of NDMFCS caused by the DFBRT (The green circle highlights fruit where counting error occurred); **(b)** Results of fruit counting after abatement of abnormal fruit detection for four consecutive RGB video frames starting from the 29th frame in V_2; **(c)** Results of fruit counting before abatement of abnormal fruit detection for four consecutive RGB video frames starting from the 48th frame in V_3, which includes a fruit counting error of NDMFCS caused by detected fruits on the ground (The green circle highlights fruit where counting error occurred); **(d)** Results of fruit counting after abatement of abnormal fruit detection for four consecutive RGB video frames starting from the 48th frame in V_3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

than that of the DFBRT, which causes tracking error of No. 207 fruit and resolved by abating abnormal fruit detection. As shown in Fig. 9, *MOTA* and *MOTP* for original videos were improved on average by 4.2% and 3.3% respectively after abatement of abnormal fruit detection, which indicated a good ability to maintain the normal trajectory of the fruit during tracking. Average *IDSR* for original videos was decreased from 4.6% to 3.5% after abatement of abnormal fruit detection, which means most fruits keep their tracking ID unchanged from appearance to disappearance.

Fruit counting performance of NDMFCS achieved an average accuracy of 95.0% in original videos after abatement of abnormal fruit detection, which was degraded due to the DFBRT and DFG. A linear regression between the number of fruits counted by NDMFCS (S) and the number of fruits counted by manual vision (G_{manu}) is illustrated, as shown in Fig. 10**(a)**. Coefficient of determination (R^2) of 0.97 was obtained, which indicates the number of fruits counted by NDMFCS (S) is near to the ground truth. Due to abnormal fruit detection, the number of fruits in original videos obtained by NDMFCS without abatement of abnormal fruit detection was on average nearly 20 more than the ground truth, as shown in Fig. 10**(b)**. Moreover, fruit counting performance of NDMFCS without abatement of abnormal fruit detection was not robust, with the maximum P_c of 97.3% (V_3) and the minimum P_c of only 83.9% (V_6). After abatement of abnormal fruit detection, NDMFCS achieved a P_c of 92.6% or higher for each of original videos, with the maximum P_c of 98.2% (V_1). And NDMFCS implemented on CPU at 3 ~ 5 frames per second (fps). All these results show that NDMFCS has achieved superior fruit counting performance, which would be degrade due to the DFBRT and DFG. As shown in Fig. 11, the DFBRT and the DFG were counted to result in a larger number of fruits than the ground truth. Detected fruits on the ground were abated by confidence threshold of fruit detection because their backgrounds are different from fruits in the front row of trees. The DFBRT, on the other hand, had a similar background to fruits in the front row of trees and therefore abated based on the distance threshold (Fu et al., 2020; Liu et al., 2020). Häni et al. (2020) also investigated abatement of the DFBRT based on the 3D reconstruction, and although it achieved good performance, it was computationally intensive and has the potential to be widely used only after the cost of high-performance computing hardware decreases in the future. Of course, it may also be promising research to transmit videos of orchards based on high-speed internet and cloud computing to abate the DFBRT

based on 3D reconstruction. The bounding box would not enclose the fruit perfectly due to the obscuration of leaves, branches and other fruits, which may cause the calculation of distance of fruits from camera to be inaccurate. However, obtaining more accurate fruit profiles by using instance segmentation rather than object detection seems to be a promising method in the future, with the potential to mitigate errors in the calculation of distance from the fruit to camera due to the occlusion of leaves, branches, and other fruits.

4. Conclusions

In this study, an automatic fruit counting system in modern apple orchard NDMFCS implemented video-based fruit counting, which achieved good performance using abatement of abnormal fruit detection. The performance of object detection model was improved by abating abnormal fruit detection, which includes the DFBRT and DFG. The DFBRT degraded fruit tracking performance of NDMFCS, while the DFBRT and DFG degraded fruit counting performance of NDMFCS. Promising results for fruit tracking and counting were obtained, illustrating the superiority of abatement of abnormal fruit detection in video-based counting applications. However, NDMFCS was developed based on videos of orchards acquired by the Intel RealSense D435 camera, which is difficult to be widely used because it requires computing equipment to achieve fruit counting. Some easy-carry smartphones have been equipped with a high-resolution RGB camera and system on chip and employed for classification and detection tasks with deep learning. Hence, a new automatic fruit counting system based on videos of orchards acquired by a smartphone could be developed in the future to help managers' decision-making.

CRediT authorship contribution statement

Zhenchao Wu: Data curation, Investigation, Writing – original draft. **Xiaoming Sun:** Methodology, Writing - review & editing. **Hanhui Jiang:** Software, Writing - review & editing. **Wulan Mao:** Investigation, Methodology, Writing – review & editing. **Rui Li:** Methodology, Writing – review & editing. **Nikita Andriyanov:** Methodology. **Vladimir Soloviev:** Investigation, Methodology, Writing – review & editing. **Longsheng Fu:** Conceptualization, Data curation, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32171897); Youth Science and Technology Nova Program in Shaanxi Province of China (2021KJXX-94); Science and Technology Promotion Program of Northwest A&F University (TGZX2021-29); National Foreign Expert Project, Ministry of Science and Technology, China (DL2022172003L, QN2022172006L).

References

Abeyrathna, R.M.R.D., Nakaguchi, V.M., Minn, A., Ahamed, T., 2023. Recognition and Counting of Apples in a Dynamic State Using a 3D Camera and Deep Learning Algorithms for Robotic Harvesting Systems. Sensors 23, 3810. https://doi.org/10.3390/s23083810.

Bao, W., Zhu, Z., Hu, G., Zhou, X., Zhang, D., Yang, X., 2023. UAV remote sensing detection of tea leaf blight based on DDMA-YOLO. Comput. Electron. Agric. 205, 107637 https://doi.org/10.1016/j.compag.2023.107637.

Bhattarai, U., Karkee, M., 2022. A weakly-supervised approach for flower/fruit counting in apple orchards. Comput. Ind. 138, 103635 https://doi.org/10.1016/j.compind.2022.103635.

Bhusal, S., Bhattarai, U., Karkee, M., 2022. Trellis wire detection for obstacle avoidance in apple orchards. IFAC-PapersOnLine 55, 72–77. https://doi.org/10.1016/j.ifacol.2022.11.117.

Biffi, L.J., Mitishita, E., Liesenberg, V., Dos Santos, A.A., Gonçalves, D.N., Estrabis, N.V., Silva, J. de A., Osco, L.P., Ramos, A.P.M., Centeno, J.A.S., Schimalski, M.B., Rufato, L., Neto, S.L.R., Junior, J.M., Gonçalves, W.N., 2021. Article atss deep learning-based approach to detect apple fruits. Remote Sens. 13, 54. https://doi.org/10.3390/rs13010054.

Cong, P., Feng, H., Lv, K., Zhou, J., Li, S., 2023. MYOLO: a lightweight fresh shiitake mushroom detection model based on YOLOv3. Agriculture 13, 392. https://doi.org/10.3390/agriculture13020392.

Fu, L., Majeed, Y., Zhang, X., Karkee, M., Zhang, Q., 2020. Faster R-CNN–based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting. Biosyst. Eng. 197, 245–256. https://doi.org/10.1016/j.biosystemseng.2020.07.007.

Fu, L., Yang, Z., Wu, F., Zou, X., Lin, J., Cao, Y., Duan, J., 2022. YOLO-Banana: a lightweight neural network for rapid detection of banana bunches and stalks in the natural environment. Agronomy 12, 391. https://doi.org/10.3390/agronomy12020391.

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M., Zhang, Q., 2020. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. Comput. Electron. Agric. 176, 105634 https://doi.org/10.1016/j.compag.2020.105634.

Gao, F., Wu, Z., Suo, R., Zhou, Z., Li, R., Fu, L., Zhang, Z., 2021. Apple detection and counting using real-time video based on deep learning and object tracking. Trans. Chinese Soc. Agric. Eng. 37, 217–224.

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., Li, R., Fu, L., Zhang, Q., 2022. A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. Comput. Electron. Agric. 197, 107000 https://doi.org/10.1016/j.compag.2022.107000.

Häni, N., Roy, P., Isler, V., 2020. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. J. F. Robot. 37, 263–282. https://doi.org/10.1002/rob.21902.

He, L., Fang, W., Zhao, G., Wu, Z., Fu, L., Li, R., Majeed, Y., Dhupia, J., 2022a. Fruit yield prediction and estimation in orchards: a state-of-the-art comprehensive review for both direct and indirect methods. Comput. Electron. Agric. 195, 106812 https://doi.org/10.1016/j.compag.2022.106812.

He, L., Wu, F., Du, X., Zhang, G., 2022b. Cascade-SORT: a robust fruit counting approach using multiple features cascade matching. Comput. Electron. Agric. 200, 107223 https://doi.org/10.1016/j.compag.2022.107223.

Jiang, Y., Li, C., 2020. Convolutional neural networks for image-based high-throughput plant phenotyping: a review. Plant Phenomics 2020, 4152816. https://doi.org/10.34133/2020/4152816.

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. Precis. Agric. 20, 1107–1135. https://doi.org/10.1007/s11119-019-09642-0.

Li, H., Li, C., Li, G., Chen, L., 2021. A real-time table grape detection method based on improved YOLOv4-tiny network in complex background. Biosyst. Eng. 212, 347–359. https://doi.org/10.1016/j.biosystemseng.2021.11.011.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft COCO: common objects in context, in: European Conference on Computer Vision. Springer, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R., Cui, Y., 2020. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. IEEE Access 8, 2327–2336. https://doi.org/10.1109/ACCESS.2019.2962513.

Liu, L., Zhang, T., Leighton, B., Zhao, L., Huang, S., Dissanayake, G., 2019. Robust global structure from motion pipeline with parallax on manifold bundle adjustment and initialization. IEEE Robot. Autom. Lett. 4, 2164–2171. https://doi.org/10.1109/LRA.2019.2900756.

Majeed, Y., Zhang, J., Zhang, X., Fu, L., Karkee, M., Zhang, Q., Whiting, M.D., 2020. Deep learning based segmentation for automated training of apple trees on trellis wires. Comput. Electron. Agric. 170, 105277 https://doi.org/10.1016/j.compag.2020.105277.

Mirhaji, H., Soleymani, M., Asakereh, A., Abdanan Mehdizadeh, S., 2021. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. Comput. Electron. Agric. 191, 106533 https://doi.org/10.1016/j.compag.2021.106533.

Mizushina, H., Kanayama, I., Masuda, Y., Suyama, S., 2020. Importance of visual information at change in motion direction on depth perception from monocular motion parallax. IEEE Trans. Ind. Appl. 56, 5637–5644. https://doi.org/10.1109/TIA.2020.3000135.

Osman, Y., Dennis, R., Elgazzar, K., 2021. Yield estimation and visualization solution for precision agriculture. Sensors 21, 6657. https://doi.org/10.3390/s21196657.

Rong, J., Zhou, H., Zhang, F., Yuan, T., Wang, P., 2023. Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion. Comput. Electron. Agric. 207, 107741 https://doi.org/10.1016/j.compag.2023.107741.

Shen, L., Su, J., He, R., Song, L., Huang, R., Fang, Y., Song, Y., Su, B., 2023. Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLOv5s. Comput. Electron. Agric. 206, 107662 https://doi.org/10.1016/j.compag.2023.107662.

Tan, C., Li, C., He, D., Song, H., 2022. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric. 193, 106683 https://doi.org/10.1016/j.compag.2021.106683.

Vasconez, J.P., Delpiano, J., Vougioukas, S., Auat Cheein, F., 2020. Comparison of convolutional neural networks in fruit detection and counting: a comprehensive evaluation. Comput. Electron. Agric. 173, 105348 https://doi.org/10.1016/j.compag.2020.105348.

Villacrés, J., Viscaino, M., Delpiano, J., Vougioukas, S., Auat Cheein, F., 2023. Apple orchard production estimation using deep learning strategies: a comparison of tracking-by-detection algorithms. Comput. Electron. Agric. 204, 107513 https://doi.org/10.1016/j.compag.2022.107513.

Wan, H., Fan, Z., Yu, X., Kang, M., Wang, P., Zeng, X., 2022. A real-time branch detection and reconstruction mechanism for harvesting robot via convolutional neural network and image segmentation. Comput. Electron. Agric. 192, 106609 https://doi.org/10.1016/j.compag.2021.106609.

Wang, X., Kang, H., Zhou, H., Au, W., Chen, C., 2022. Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. Comput. Electron. Agric. 193, 106716 https://doi.org/10.1016/j.compag.2022.106716.

Wu, Z., Li, G., Yang, R., Fu, L., Li, R., Wang, S., 2022. Coefficient of restitution of kiwifruit without external interference. J. Food Eng. 327, 111060 https://doi.org/10.1016/j.jfoodeng.2022.111060.

Wu, F., Yang, Z., Mo, X., Wu, Z., Tang, W., Duan, J., Zou, X., 2023. Detection and counting of banana bunches by integrating deep learning and classic image-processing algorithms. Comput. Electron. Agric. 209, 107827 https://doi.org/10.1016/j.compag.2023.107827.

Zeng, T., Li, S., Song, Q., Zhong, F., Wei, X., 2023. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment. Comput. Electron. Agric. 205, 107625 https://doi.org/10.1016/j.compag.2023.107625.

Zhang, W., Wang, J., Liu, Y., Chen, K., Li, H., Duan, Y., Wu, W., Shi, Y., Guo, W., 2022. Deep-learning-based in-field citrus fruit detection and tracking. Hortic. Res. 9, uhac003. https://doi.org/10.1093/hr/uhac003.

Zheng, Z., Xiong, J., Wang, X., Li, Z., Huang, Q., Chen, H., Han, Y., 2022. An efficient online citrus counting system for large-scale unstructured orchards based on the unmanned aerial vehicle. J. F. Robot. 40, 552–573. https://doi.org/10.1002/rob.22147.