Computers and Electronics in Agriculture 199 (2022) 107209

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Remote estimation of grafted apple tree trunk diameter in modern orchard with RGB and point cloud based on SOLOv2

Xiaoming Sun [a], Wentai Fang [a], Changqing Gao [a], Longsheng Fu [a,b,c,*], Yaqoob Majeed [d], Xiaojuan Liu [a], Fangfang Gao [a], Ruizhe Yang [a], Rui Li [a]

[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] Faculty of Agricultural Engineering and Technology, University of Agriculture, Faisalabad 38000, Pakistan

ARTICLE INFO

Keywords:
Apple tree phenotyping
Deep learning
Grafted apple tree
Grafting position
Trunk diameter

ABSTRACT

Apple tree phenotyping can reflect individual development of single apple tree, which mainly involves tree height, crown width, and diameter of apple tree trunk (DATT). This study aimed to estimate diameter of grafted apple tree trunk, whose target position of diameter estimation is about 10 cm above grafting position. An estimated DATT approach of combining red–greenblue-depth (RGB-D) sensor with SOLOv2 was proposed. Firstly, Kinect V2 was employed to obtain original RGB images and point clouds of the grafted apple trees simultaneously. There were 120 and 60 RGB images and corresponding point clouds randomly collected from two modern apple orchards. Secondly, SOLOv2 deep learning model was selected and trained to instance segment grafting position from RGB image for determining it automatically. Then, corresponding exact position of the grafting position in point cloud was mapped by coordinate transformation of its pixel coordinates, which was obtained by trained SOLOv2 model. Finally, DATT was estimated by calculating the difference between maximum and minimum Y coordinates of points selected by distance thresholds in X, Y, and Z directions near the target position, which were 0.10 m, 0.035 m, and 0.20 m, respectively. Results showed that average precision and average recall of the trained SOLOv2 model for instant segmenting the grafting position were 0.811 and 0.830, respectively. Mean absolute error, mean absolute percentage error, and root mean square error of the proposed method were 3.01 mm, 5.86%, and 3.79 mm, respectively. It illustrates that the proposed method can estimate DATT and thus contribute to automatic apple tree phenotyping.

1. Introduction

Apple tree phenotyping can monitor growth tendency of individual fruit tree in apple orchard, which facilitates farmers manage the orchard precisely. It is external manifestation of growth and development of individual apple tree under combined action of genes and environment (Dungey et al., 2018). Tree phenotyping mainly includes key indicators such as tree height, crown width, trunk diameter, and fruit properties (Iizuka et al., 2018; Swayze et al., 2021; Zhang et al., 2021; Gao et al., 2022; Wu et al., 2022). These indicators assist to fully reflect overall situation of entire orchard. Therefore, it is significant to obtain accurate phenotypic information of apple tree.

Diameter of apple tree trunk (DATT) is one of the important indicators of apple tree phenotyping. DATT can reflect growth situation of single apple tree in orchard (Hao et al., 2022). Estimating diameter of tree trunk needs to accurately locate measurement height (Montoya et al., 2021; Trochta et al., 2017). For natural growing tree that is not grafted, the height of measuring diameter at breast height (DBH) is usually about 1.3 m above the ground (Moe et al., 2020; Wang et al., 2019; Wu et al., 2019). While for grafted apple tree, target position of estimating DATT is about 10 cm above the grafting position (Foster et al., 2015). DATT is significantly important parameter to calculate (by converting it to cross-sectional area and then dividing it by yield of a single apple tree) the production efficiency of an apple tree, which reflects its ability of bearing apples (Coupel-Ledru et al., 2019; Mokroš et al., 2018; Sosna, 2014). As a consequence, it is important to accurately estimate DATT.

Fruit growers usually measure DATT manually with diameter tapes

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

https://doi.org/10.1016/j.compag.2022.107209
Received 21 April 2022; Received in revised form 25 June 2022; Accepted 8 July 2022
Available online 14 July 2022
0168-1699/© 2022 Elsevier B.V. All rights reserved.

or calipers, which is time-consuming and labor-intensive. To estimate DATT, two investigators need to work together and maintain a squatting position for a long time, which makes it extremely difficult and tedious to complete apple tree surveys within limited timeframe (Krisanski et al., 2020; Liu et al., 2021; Luoma et al., 2017). Consequently, realizing remote and accurate estimation of DATT has become critically important problem needs to be solved in apple tree phenotyping.

Some researchers had estimated diameter of tree trunk by using tools including digital camera, laser rangefinder, or mobile phone, which was influenced by investigators subjectivity. Song et al. (2021) developed a handheld device with a digital camera and a laser rangefinder for DBH measurement of a single tree, which had a measurement bias of −1.78 mm and needed to keep parallel to target tree trunk. Yuan et al. (2021) designed an electronic device using a high-precision laser ranging sensor and a dip sensor to measure DBH after manually determining exact measurement location, which had a estimation bias of 1.6 mm compared with calipers. Fan et al. (2018) utilized a mobile phone with red–greenblue-depth (RGB-D) and simultaneous localization and mapping (SLAM) to measure DBH and hence decided whether to remeasure, which had a large measurement bias of 330 mm. Shuib et al. (2018) installed customized girth measurement equipment based on radio frequency identification/wireless sensor network technology on tree trunk to measure girth of rubber tree, which reported an average error value of 1 mm. Based on above studies, investigator's subjectivity affects diameter estimation of tree trunk using the customized equipment.

Other researchers proposed automatic methods for measuring diameter of tree trunk without additional subjectivity of investigators by employing different three-dimensional sensors, such as light detection and ranging (LiDAR) or RGB-D sensor. Corte et al. (2020) mounted LiDAR on unmanned aerial vehicle to estimate DBH in a forest plantation with approximately 17 ha, which obtained root mean square error (RMSE) of 34.6 mm. McGlade et al. (2020) employed Kinect V2 as RGB-D sensor to estimate DBH of urban tree, which obtained the RMSE of 35.3 mm after removing trees with irregular or non-circular stems. RGB-D sensor was helpful for estimating DATT without additional subjectivity of investigators.

RGB-D sensor was employed to estimated DATT in this study. Estimating DATT directly from point cloud collected by Lidar was difficult because features of grafting position in point cloud are not as obvious as those in RGB image. Although point cloud acquired by LiDAR is better than RGB-D sensor (Bruggisser et al., 2020; Guo et al., 2021; Hyyppä et al., 2020; Pires et al., 2022; Sun et al., 2021), low-cost RGB-D sensor is conducive to automatically identify grafting position of the grafted apple tree for estimating DATT, which can acquire RGB image and point cloud (Fu et al., 2020a, 2020b; Zhang et al., 2018; Zhang et al., 2020). Low-cost RGB-D sensor had unique advantage than LiDAR, which was able to acquire RGB image to locate the grafting position for estimating DATT.

Identifying the grafting position and setting suitable distance thresholds in different directions were two critical steps in estimating DATT by RGB-D sensor. Identifying the grafting position is a crucial step to achieve remote and automatic estimation of DATT. In this study, pixel mask of the grafting position was acquired through instance segmentation. SOLOv2 has better instance segmentation results than other classic instance segmentation networks, such as Mask R-CNN (Wang et al., 2020b). The more accurate targets edge information segmented by SOLOv2, the better DATT estimation results. Therefore, SOLOv2 was selected as an instance segmentation network for identifying the grafting position and gaining corresponding pixel mask based on two considerations as follows. Firstly, only target apple tree in center of image is segmented by SOLOv2 when several apple trees appear in RGB image because SOLOv2 can segment targets by locations (Wang et al., 2020a). Other apple trees in RGB image will not be identified so that they don't interfere with estimating DATT. Secondly, precise edge information of the grafting position is required to accurately estimate DATT. Establishing appropriate distance thresholds in different directions is another key step during estimating DATT. The results of estimating DATT will be affected due to sparsity of point cloud collected by RGB-D sensor (Gollob et al., 2021; Neuville et al., 2021). Stacking points within a certain distance threshold near the target position in the X-axis direction can compensate for the sparsity of point cloud collected by RGB-D sensor. Different distance thresholds setting in X direction result in different thicknesses of the stacked point clouds, thus affecting results of estimating DATT (Koreň et al., 2020). Distance thresholds in Y and Z directions are set to determine points belonging to grafted apple tree trunk, which were near the target position in point cloud. Distance thresholds setting in Y and Z directions depend on width range of the grafted apple tree trunk. Setting appropriate thresholds in different directions is beneficial to obtain more accurate results of estimating DATT.

In this study, RGB image and point cloud obtained by RGB-D sensor were used to realize remote and automatic DATT estimation by identifying the grafting position from RGB image and estimating diameter from corresponding point cloud. Instance segmentation of the grafting position helped to locate its position in the point cloud and thus facilitated to obtain the grafted apple tree phenotyping remotely and automatically. The structure of this paper was as follows: Section 2 illustrated the whole workflow of the proposed method for estimating DATT, recognition of the grafting position, method for determining the target position, details of estimating DATT, and performance evaluation of method proposed in this study. Section 3 described performance of the trained SOLOv2 model, appropriate distance thresholds setting, and DATT estimation evaluation. Conclusions of this study were summarized in section 4.

2. Materials and methods

2.1. Analysis overview

An automatic DATT estimation method combining instance segmentation from RGB image with point cloud was proposed in this study. Its workflow is shown in Fig. 1. Firstly, original RGB image and corresponding point cloud based on depth map of the grafted apple tree were acquired by RGB-D sensor. For RGB-D sensor, original RGB image and depth map were directly collected by RGB camera and depth sensor, respectively. Then, depth map was aligned to the original RGB image by employing Kinect Software Development Kit (SDK) build-in function, which was provided by Kinect V2 manufacturer Microsoft. After that, point cloud was then generated from the aligned depth map. Z coordinate of points in point cloud was directly obtained from aligned depth map. X and Y coordinates of points in point cloud were thus calculating by corresponding Z coordinate, pixel coordinates of this points, and internal paraments of Kinect V2. Therefore, points in point cloud can be indexed by pixel coordinates of original RGB image. The point cloud was filtered based on statistical outlier filter to remove abnormal noise points near tree trunk. Then grafting position of the grafted apple tree was annotated for making datasets that was used to train SOLOv2 model for segmenting the grafting position. Secondly, the trained SOLOv2 model was utilized to visualize grafting position mask of the grafted apple tree and determine its pixel coordinates. To index the point cloud by pixel coordinates of the original RGB image, X, Y, and Z coordinates of points in point cloud were respectively and sequentially written into three corresponding coordinates matrices as matrix elements. Then, the grafting position and target position of estimating DATT were determined. Finally, the points that were picked by distance thresholds in X, Y, and Z directions near the target position were stacked in the X-axis direction to estimate DATT.

2.2. Data acquisition

Data acquisition system, mainly composed of Kinect V2 sensor, tripod, and computer to acquire RGB image and corresponding point cloud of the grafted apple tree. As RGB-D sensor, Kinect V2 was fixed on

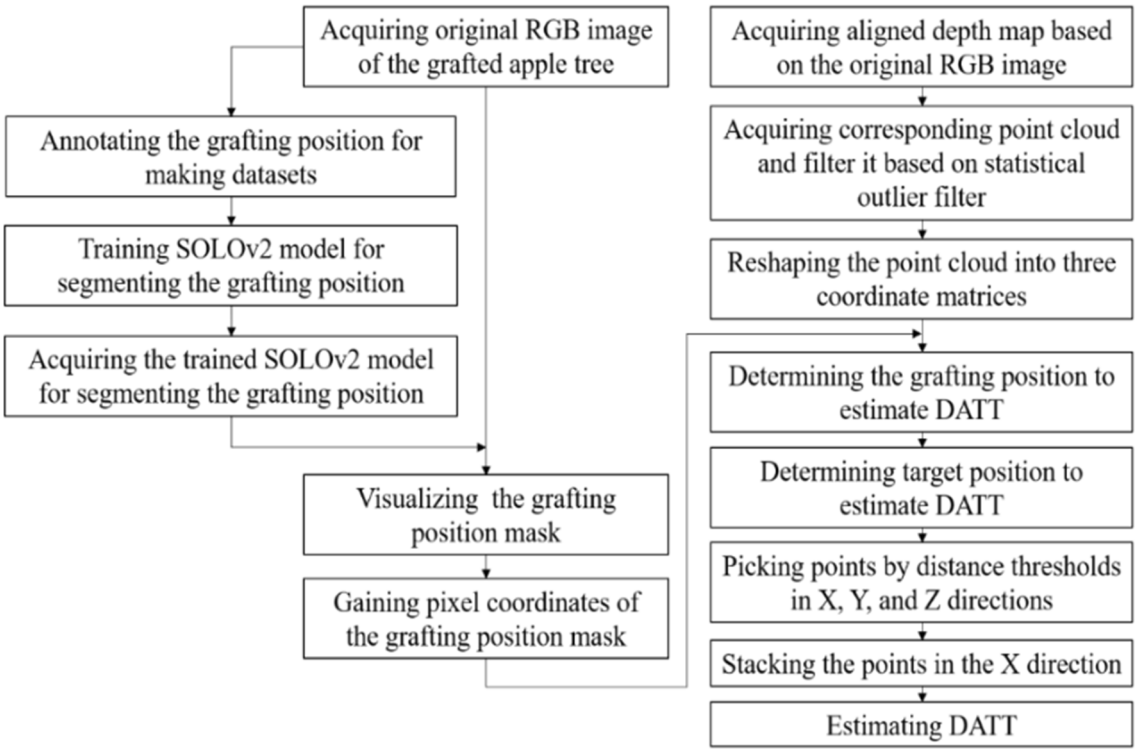


Fig. 1. Workflow of estimating DATT of the grafted apple tree.

a tripod at the height of 1.2 m and was about 2.0 m far away from the target apple tree to acquire original RGB image with a resolution of 1,920 × 1,080 pixels, depth map aligned to RGB image, and point cloud. Depth values were collected by depth sensor of Kinect V2, which was Z coordinates of points in point cloud. X and Y coordinates of points in point cloud were calculated by corresponding Z coordinate, internal parameters of Kinect V2, and pixel coordinates of points, which were calculated by Eqs. (1) and (2), respectively. After Kinect V2 acquired data, the point cloud was filtered based on statistical outlier filter. Schematic diagram of the data acquisition system is shown in Fig. 2. Original depth map with 512 × 424 pixels resolution was registered to original RGB image with resolution of 1,920 × 1,080. The point cloud generated by aligned depth map was thus correspondingly expanded from 512 × 424 to 1,920 × 1,080 points. During collecting data under the natural light, the whole target apple tree was always in center of the original RGB image.

$$x = \frac{(u - c_x)z}{f_x} \tag{1}$$

$$y = \frac{(v - c_y)z}{f_y} \tag{2}$$

Where u and v are pixel coordinates of point. z is depth value of point collected by depth sensor of Kinect V2. c_x, c_y, f_x, and f_y are internal parameters of Kinect V2.

The data was collected from two modern apple orchards in Qianxian and Fufeng, Shaanxi Province, China. Both modern apple orchards adopted dwarf stock dense planting cultivation mode with 1.0 m × 3.5 m spacing among apple trees. Variety of apples grown in Fufeng modern apple orchard was 'Gala' while it was 'Qincui' in Qianxian modern apple orchard. In total, 120 and 60 pairs of original RGB images and corresponding point clouds of the grafted apple trees were randomly collected

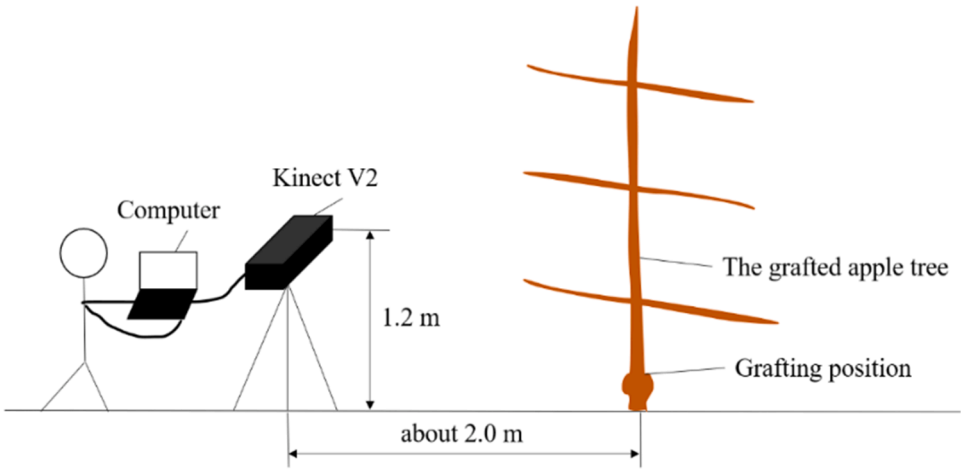


Fig. 2. Schematic diagram of data acquisition system.

in two apple orchards, respectively. The diameters of 24 and 30 grafted apple trees were manually measured with a caliper in Qianxian and Fufeng respectively, which ranged from 31 mm to 68 mm. Manual measurement results were used as ground truth to subsequently compare with the proposed method of estimating DATT. The target position of estimating DATT was 10 cm higher than the grafting position of the grafted apple tree, as shown in Fig. 3.

2.3. Grafting position recognition

2.3.1. Dataset preparation

Grafting position of the grafted apple tree was annotated for making a suitable dataset as an input for SOLOv2. Only target apple tree in the center of original RGB image was annotated. RGB image was manually labeled as ground truth with two classes: grafting position and trunk, as shown in Fig. 4. To make the trained SOLOv2 model more suitable for different lighting conditions, 180 pairs of original RGB images and corresponding annotation files were augmented to 900 pairs by sequentially setting both brightness and contrast to 0.9 and 1.1, respectively. Training set and testing set were randomly divided into 4 to 1 ratio (80 % for training set and 20 % for testing set) so that 720 pairs were randomly selected for training set while the remaining 180 pairs were used as the testing set.

2.3.2. SOLOv2 model training

Instance segmentation of the grafting position from RGB image with the trained SOLOv2 model was basis for obtaining pixel coordinates of its mask. Framework 'solov2-r101-dcn-fpn-8gpu-3x' was built on Linux system, which employed GeForce GTX 3080Ti as its GPU. Learning rate and epochs were set to 0.002 and 16, respectively. After SOLOv2 model training completed, the original RGB images of the grafted apple trees that had the manual measurement results of DATT were visualized through the trained SOLOv2 model to obtain instance segmentation mask of the grafting position and thus gain its pixel coordinates.

2.4. Determining target position of estimating DATT

Point cloud was transformed to three 1,920-by-1,080 coordinate matrices for indexing the grafting position by pixel coordinates of its mask. Elements in three coordinate matrices were X, Y, and Z coordinates of points in point cloud, respectively. The point cloud contained three-dimensional information of object, including X, Y, and Z coordinate values. Unit of these coordinate values in point cloud was in meter. The X-axis direction represented height direction of the grafted apple tree. The Y-axis direction represented width direction of the grafted apple tree while the Z-axis direction represented depth direction of the grafted apple tree. In process of collecting data by Kinect V2, coordinate values of points were successively written into point cloud with PCD format in sequence by column of the aligned depth map. To index exact grafting position by pixel coordinates of its mask, X, Y, and Z coordinates of points in point cloud were written into three coordinate matrices with a resolution of 1,080 × 1,920 pixels as matrix elements, sequentially and respectively.

The grafting position was indexed by pixel coordinate transformation of the highest middle point (HMP) on its mask. HMP was a special point on pixel mask of the grafting position, which was the highest point on vertical axis of its geometric center. Pixel abscissa of HMP was pixel ordinate of the highest point of the grafting position mask. Its pixel ordinate belonged to the pixel abscissa of geometric center of the grafting position mask. Progress of mapping HMP from RGB image to three coordinate matrices is shown as Fig. 5, which are transformed by point cloud. After gaining pixel coordinates of HMP, corresponding point in point cloud was thus matched by coordinate transformation formulas, which are shown in Eq. (3) and Eq. (4). Abscissa and ordinate pixel coordinates of HMP were employed to index corresponding X, Y, and Z coordinates in three coordinate metrices, which were transformed by point cloud.

$$x_p = 1079 - x_i \tag{3}$$

$$y_p = 1919 - y_i \tag{4}$$

Where (x_i, y_i) is pixel coordinate of one point in original RGB image and (x_p, y_p) is coordinate of corresponding point in three coordinate matrices.

Coordinates of points in point cloud were possibly null values. A square was thus employed to calculate average coordinate values of points in it except for null values as exact grafting position. The grafting position was exactly determined by it. Its side length can be automatically changed, which was odd to ensure that HMP was its center. Its size automatically adjusted from 1 × 1 until average coordinate values of it were all not empty in X, Y, and Z directions. The average values in X, Y, and Z directions were separately regarded as the height, width, and



Fig. 3. Grafting position of a grafted apple tree in Fufeng modern apple orchard (a) and schematic diagram of target position of estimating DATT (b).

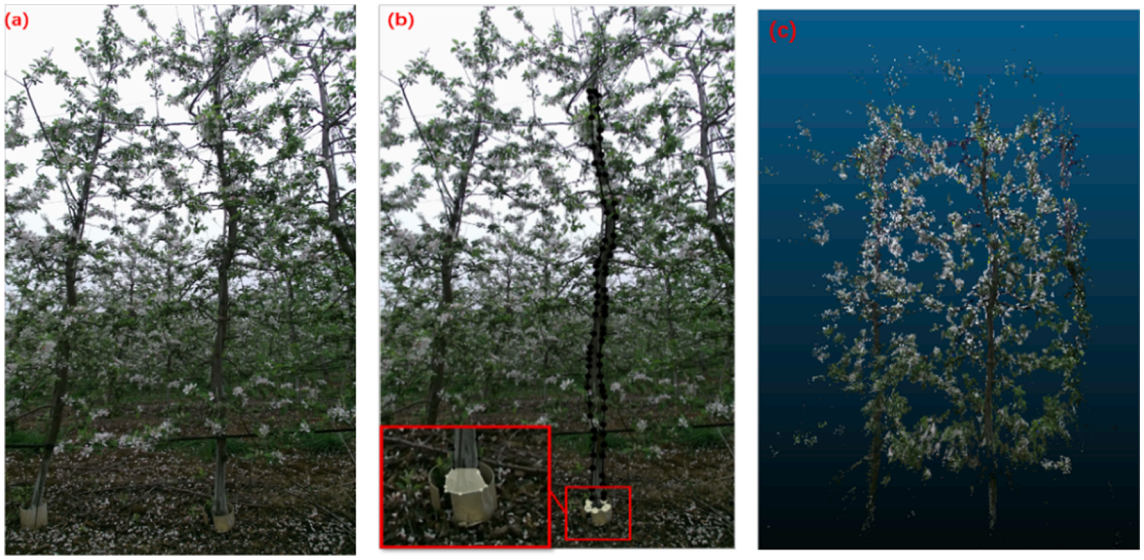


Fig. 4. An example of manual labelling result of RGB image. (a) Original image collected by Kinect V2; (b) Ground truth of grafting position labelled by white points and ground truth of trunk labelled by black points; (c) Corresponding point cloud of this RGB image.

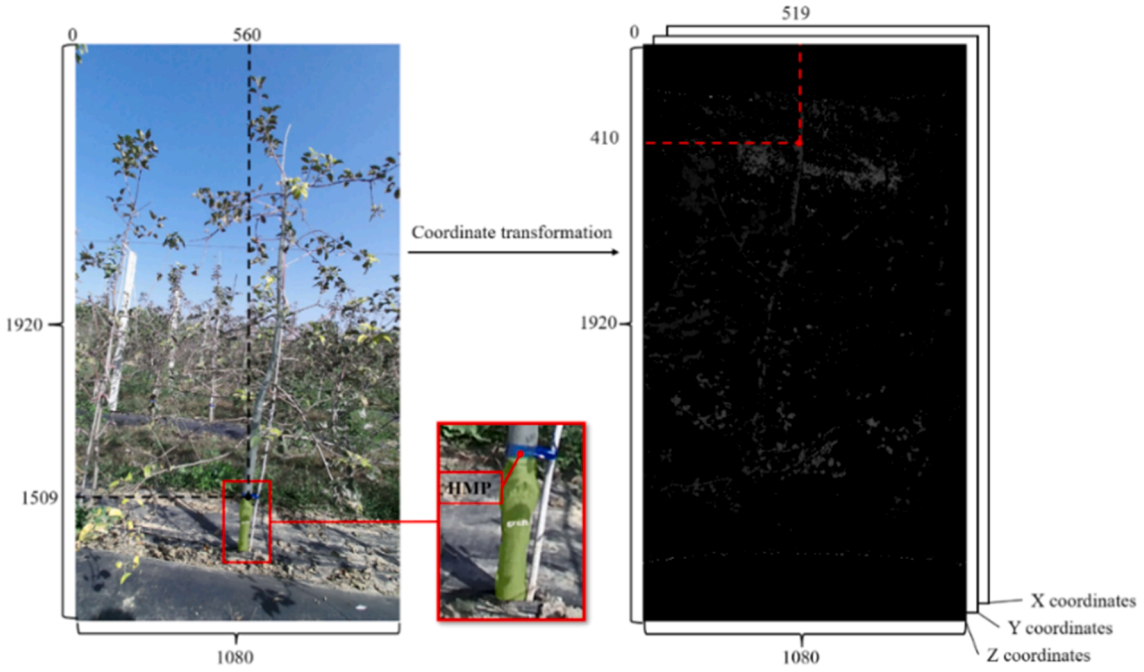


Fig. 5. An example of mapping HMP from RGB image to three coordinate matrices transformed by point cloud.

depth of the grafting position.

After the grafting position was located, the target position of estimating DATT was determined. The height of the target position was determined by adding 0.10 m to the height of the grafting position. The width and depth of the grafting position were separately considered as the width and depth of the target position. The target position of estimating DATT was located from the height, width, and depth directions. It was located and further utilized for subsequent DATT estimation.

2.5. DATT estimation

Distance thresholds in X, Y, and Z directions were set to select points near the target position, which were picked to estimate DATT, as shown in Fig. 6. Since density of the point cloud collected by Kinect V2 was relatively sparse, picked points were stacked in the X-axis direction to make up for defect of the sparsity of the point cloud. In this paper, distance thresholds in X, Y, and Z directions were 0.10 m, 0.035 m, and 0.20 m, respectively. All points conforming to Eq. (5) were used to estimate DATT.

$$\begin{aligned} x_{height} - x_{threshold} < x < x_{height} + x_{threshold}, \\ y_{width} - y_{threshold} < y < y_{width} + y_{threshold} \\ z_{depth} - z_{threshold} < z < z_{depth} + z_{threshold}, \end{aligned} \tag{5}$$

Where (x, y, z) is coordinate of one point to estimate DATT. The x_{height}, y_{width}, and z_{depth} are X, Y, and Z coordinate values of the target position, respectively. The $x_{threshold}$, $y_{threshold}$, and $z_{threshold}$ are distance thresholds in X, Y, and Z directions, respectively.

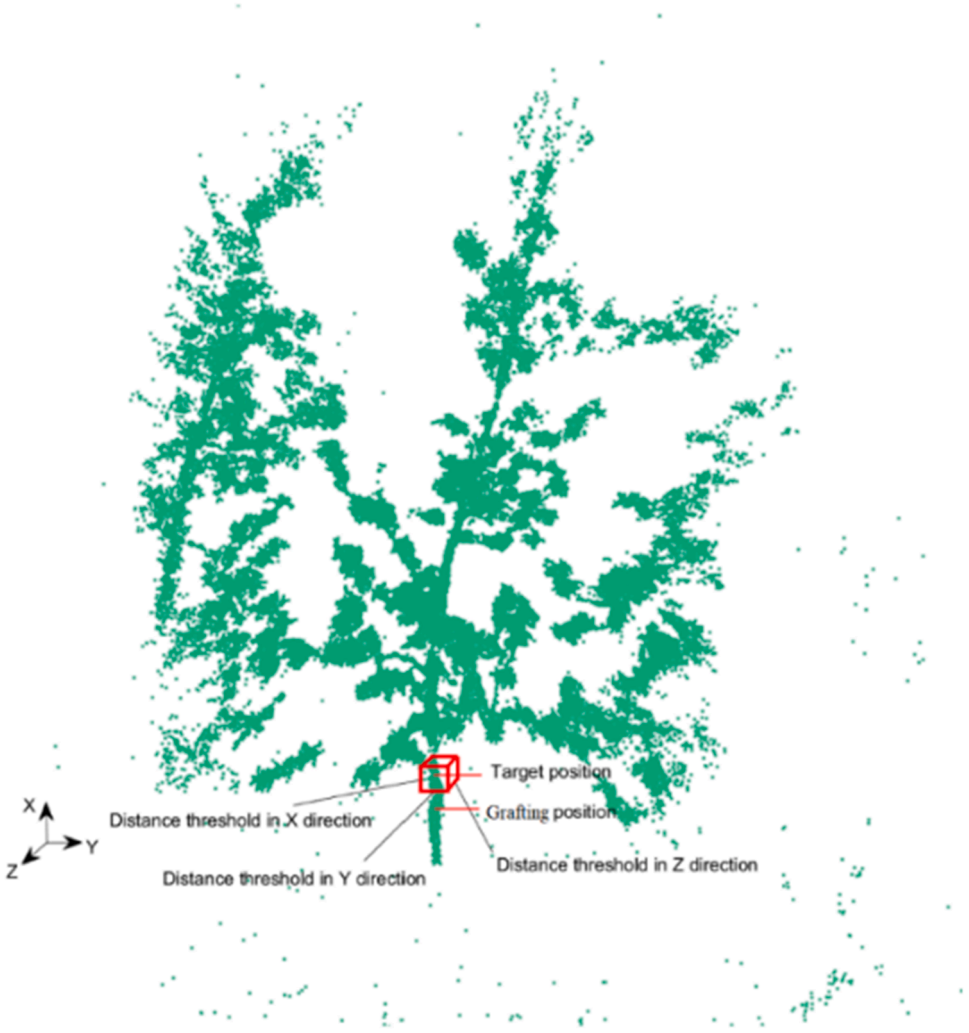


Fig. 6. Points in the red cuboid were picked by distance thresholds in X, Y, and Z directions for estimating DATT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

DATT was estimated by the stacked points that were selected by distance thresholds in X, Y, and Z directions near the target position. All points that were picked by Eq. (3) were stacked in the X-axis direction, as shown in Fig. 7. According to values of Y coordinate from small to large, all stacked points were arranged in sequence. DATT was estimated by calculating difference between the maximum and the minimum coordinate values of these stacked points in the Y-axis direction. Therefore, diameter of the grafted apple tree was automatically estimated by the

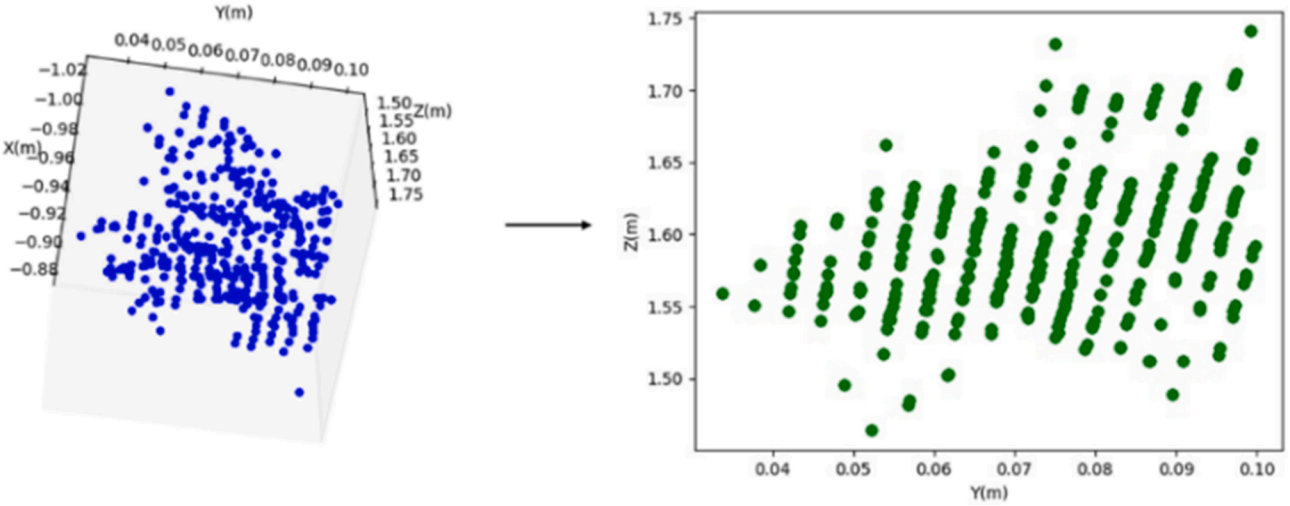


Fig. 7. Points picked for estimating DATT were stacked in the X-axis direction.

proposed method. Pseudo-code of automatically estimating DATT are shown in Fig. 8.

2.6. Performance evaluation

Performance of the trained SOLOv2 model for segmenting the grafting position was evaluated by average precision (AP), AP_{50}, AP_{75}, AP_S, average recall (AR), and AR_S. Precision (P) and recall (R) are defined in Eqs. (6) and (7), respectively. Intersection over Union (IoU) is calculated by Eq. (8).

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$IoU = \frac{D \cap G}{D \bigcup G} \tag{8}$$

Where *TP* (True Positives) represents number of correctly predicted labeled objects, *FP* (False Positives) represents number of unlabeled objects with incorrect predictions, and *FN* (False Negatives) represents number of missed objects. *D* is area of detection result of object and *G* is corresponding area of ground truth.

AP, AP_{50} and AP_{75} are calculated by Eq. (9), Eq. (10), and Eq. (11), respectively. AR is recall value given one detection per image. AP_S and AR_S indicate AP and AR when small objects with pixel area less than 32 × 32, respectively.

$$AP = \frac{1}{10}\sum_{i=0}^{9}\left(\int_0^1 P_i(R_i)dR_i\right) \tag{9}$$

$$AP_{50} = \int_0^1 P_i(R_i)dR_i, i = 0 \tag{10}$$

$$AP_{75} = \int_0^1 P_i(R_i)dR_i, i = 5 \tag{11}$$

Where i is the i^{th} *IoU* when *IoU* is from 0.50 to 0.95 and corresponding step size is 0.05. P_i and R_i are precision and recall with the i^{th} *IoU*.

Performance of the proposed method for estimating DATT in this study was evaluated by mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) compared with manual measurement method. The results of manual measurement with a caliper were considered as the ground truth. The MAE, MAPE and

Algorithm 1 Algorithm of estimating DATT

Input: Abscissa and ordinate coordinates of HMP: x_i, y_i; Coordinates of points in point cloud: x, y, z;
Output: Diameter of the grafting position: D;

$x_i \leftarrow (1079 - x_i)$, $y_i \leftarrow (1919 - y_i)$; // Locate square by coordinate transform
$kernel \leftarrow 1, judge \leftarrow$ True; // Determine side length of square
while $(judge)$ **do**
    Find coordinates of points in square. Its side length is kernel and central point is HMP;
    $flag \leftarrow$ True;
    for $(i \leftarrow 0$ to $(kernel - 1))$ **do**
        if (There are coordinates of points in square arent NAN value) **then**
            $flag \leftarrow$ False;
            break;
        end if
    end for
    if $(flag)$ **then**
        $kernel \leftarrow (kernel + 2)$;
        continue;
    else
        $judge \leftarrow$ False;
    end if
end while
Calculate averages of x, y, z coordinates of points in square, respectively;
$height \leftarrow average_x$, $width \leftarrow average_y$, $depth \leftarrow average_z$;
$height \leftarrow height + 0.1$; // Target position is 0.1 m higher than grafting position
$threshold_x \leftarrow 0.1$, $threshold_y \leftarrow 0.035$, $threshold_z \leftarrow 0.2$; // Distance thresholds
for $(i \leftarrow 0$ to $(1{,}920 * 1{,}080 - 1))$ **do**
    if $(height - threshold_x) < x < (height + threshold_x))$ **then**
        if $((width - threshold_y) < y < (width + threshold_y))$ **then**
            if $((depth - threshold_z) < z < (depth + threshold_z))$ **then**
                Find coordinates of points that are within range of distance thresholds;
            end if
        end if
    end if
end for
Sort y coordinates of selected points from small to large;
$D \leftarrow (max_y$ - $min_y)$;
return D

Fig. 8. Pseudo-code of automatically estimating DATT.

RMSE are calculated by Eq. (12), Eq. (13), and Eq. (14), respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(e_i - g_i)^2 \tag{12}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{e_i - g_i}{g_i}\right)^2 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - g_i)^2} \tag{14}$$

Where n is number of samples to estimate DATT, e_i is estimation value of the i^{th} sample, and g_i is ground truth of the i^{th} sample.

3. Results and discussions

3.1. Performances of trained SOLOv2 model

3.1.1. Training assessment of SOLOv2 model

Loss value was an indicator reflecting whether training process of SOLOv2 model was effective. Loss curves of the trained SOLOv2 model are shown in Fig. 9. Loss of the trained SOLOv2 model was equal to 'loss_ins' plus 'loss_cate', which represented loss of instance mask segmentation and loss of instance category prediction, respectively. The loss decreased as iterations increased. After the iterations reached 11,520, the loss curve converged and stabilized. The lowest loss of the trained SOLOv2 model reached 0.0988, while the lowest 'loss_ins' and 'loss_cate' were 0.0782 and 0.0206, respectively. The low and stable loss showed that the trained SOLOv2 model learned features of the grafting position of the grafted apple tree, which had potential to accomplish automatic DATT estimation.

3.1.2. Instance segmentation results of trained SOLOv2 model

Instance segmentation results of the trained SOLOv2 model provided a basis for subsequent DATT estimation. Examples of instance segmentation results of the trained SOLOv2 model and corresponding exact location of the grafting position in point cloud are shown in Fig. 10. Instance segmentation results that were validated on the testing set by the trained SOLOv2 model are as shown in Table 1. The AP and AR of the trained SOLOv2 model were 0.811 and 0.830, respectively. For small object with area less than 32 × 32 pixels, such as the grafting position, AP_S and AR_S were 0.829 and 0.832, respectively. The AP of the trained SOLOv2 model were higher than the AP of Mask R-CNN model for identifying and segmenting minerals to estimate surface modal mineralogy (AP = 0.504 vs AP = 0.443) (Koh et al., 2021). Similar to Wang et al. (2020), the results illustrated that SOLOv2 had better instance segmentation than Mask R-CNN. AP_{50} and AP_{75} of the trained SOLOv2 model were 0.993 and 0.990, respectively. The AP_{50} and AP_{75} were both over 0.6, which demonstrated that the trained SOLOv2 model can acquire accurate instance segmentation results of the grafting position for estimating DATT (Lumnitz et al., 2021).

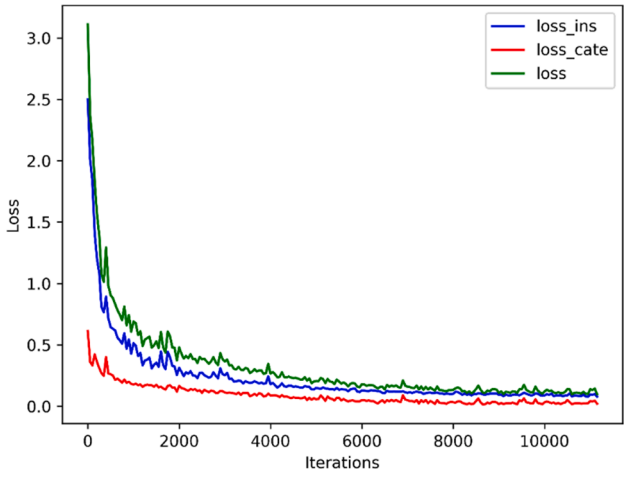


Fig. 9. Loss curve of the trained SOLOv2 model.

3.2. Determining distance threshold in X direction

Appropriate distance thresholds in different directions were set to reduce error of DATT estimation results, especially distance threshold in X direction. Distance threshold in Y direction can be set based on experience about the maximum diameter for the grafted apple trees of different ages and varieties. Since the maximum diameter of the collected grafted apple trees was 68 mm, distance threshold in Y direction was set to 0.035 m, which was half of an integer that was greater than the largest diameter of them. As there are not many interference points from other objects in depth direction, distance threshold in Z direction was automatically set to 0.20 m, which should be as large as possible. Distance threshold in X direction influenced the thickness of the stacked points, which would have greater impact on the results of estimation DATT than distance thresholds in Y and Z directions. Distance threshold in X direction was automatically set to 0.10 m by comparing five different distance thresholds in X direction, which were set to 0.01 m, 0.05 m, 0.10 m, 0.15 m, and 0.20 m, respectively.

Determining appropriate distance threshold in X, Y, and Z directions was significant to estimate DATT. To explore impact of different distance thresholds in X direction on results of estimating DATT, five different distance thresholds in X direction were set to 0.01 m, 0.05 m, 0.10 m, 0.15 m, and 0.20 m, respectively. Meanwhile, distance thresholds in Y and Z directions remained at 0.035 m and 0.20 m, respectively. When different distance thresholds in X direction were set, the method proposed in this study was applied to estimate DATT. Fig. 11 shows the DATT estimation results when distance thresholds in X direction are different compared with manual measurement using a caliper.

Five distance thresholds in X direction were set to estimate DATT with the method proposed in this study. The MAEs and MAPEs of the five distance thresholds in X direction are shown in Table 2. When distance threshold in X direction was less than 0.10 m, the MAE and MAPE decreased as distance threshold in X direction increased due to the sparsity of the point cloud collected by Kinect V2. However, when distance threshold in X direction was more than 0.10 m, the MAE and MAPE increased as distance threshold in X direction increased because the stacked points contained too many points that didn't belong to the target position. When distance threshold in X direction was 0.10 m, the result was optimal among five distance thresholds in X direction. The MAE and MAPE of the best result were 3.01 mm and 5.86%, respectively. In consequence, the distance threshold in X, Y, and Z directions for estimating DATT were set to 0.10 m, 0.035 m, and 0.2 m, respectively.

3.3. DATT estimation evaluation

In this paper, the proposed method of estimating DATT showed potential for automatic estimation of diameter of the grafted apple tree trunk. Results of estimating DATT of all grafted apple trees are as shown in Fig. 12, which are collected in two modern apple orchards. It showed that a strong relationship was established between the proposed method and manual method for estimating DATT ($R^2 = 0.8333$). For the grafted apple trees collected in Qianxian, the MAE, MAPE, and RMSE were 3.11 mm, 6.59%, and 3.53 mm, respectively. As for the grafted apple trees gathered in Fufeng, the MAE, MAPE, and RMSE were 2.93 mm, 5.28%, and 3.98 mm, respectively. The MAE, MAPE, and RMSE of all data were 3.01 mm, 5.86%, and 3.79 mm, respectively. Similarly to Tomaštík et al. (2017), the point cloud of tree was collected by RGB-D sensor. However, Tomaštík et al. (2017) estimated the diameter of tree trunk by calculating the Euclidean distance between two points on left and right

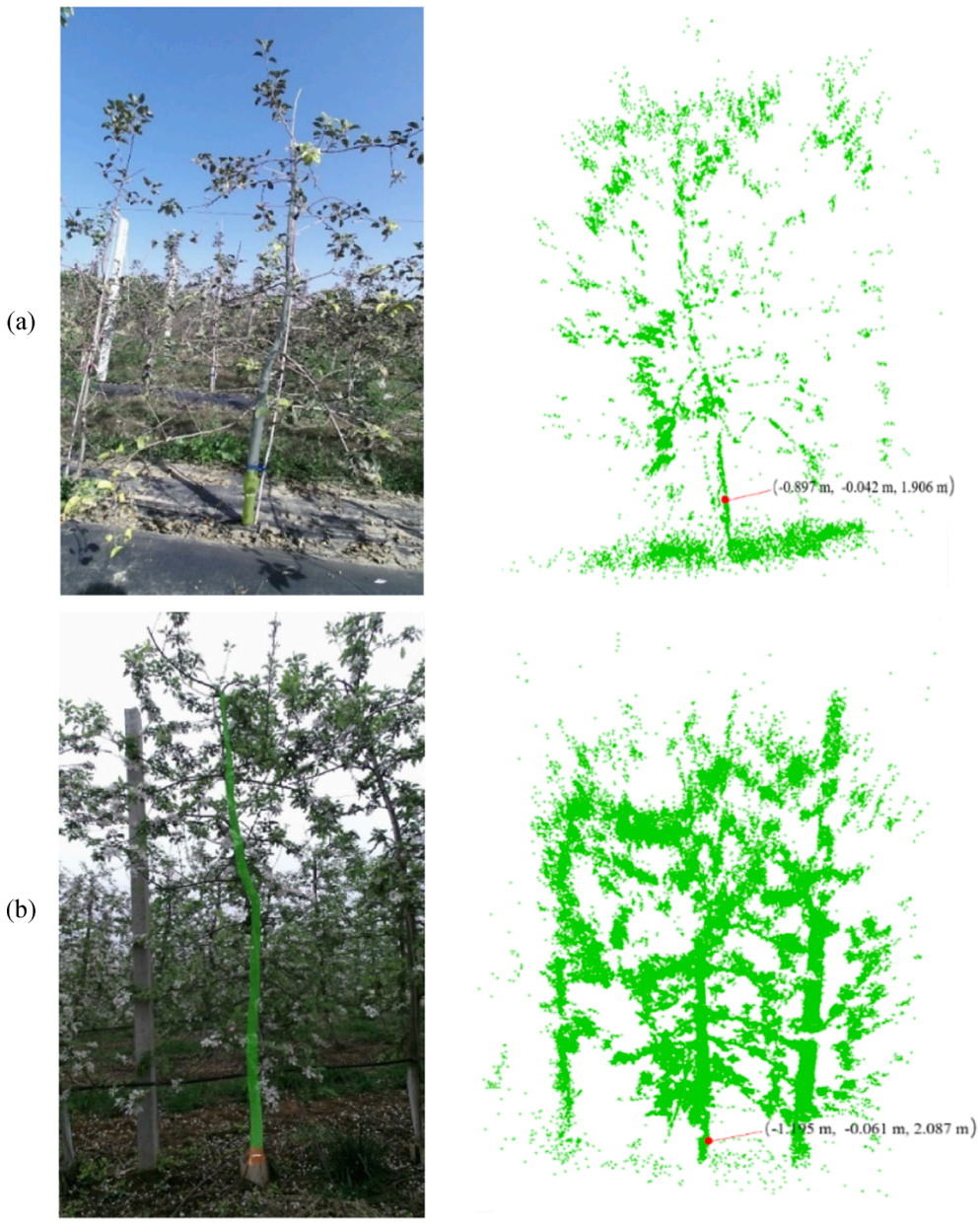


Fig. 10. Examples of instance segmentation results of the trained SOLOv2 model and corresponding exact location of the grafting position in point cloud. (a) and (b) were collected in Qianxian and Fufeng, respectively.

Table 1
Instance segmentation results of the trained SOLOv2 model.

Model	Instance segmentation results					
Solov2_r101_dcn_fpn_8gpu_3x	AP	AP_{50}	AP_{75}	AP_S	AR	AR_S
	0.811	0.993	0.990	0.829	0.830	0.832

borders of tree trunk at the measurement location, which gained the RMSEs between 16.1 mm and 21.0 mm. Hence, a conclusion that the proposed method of estimating DATT was conducive to improve accuracy of DATT estimation and realize automatic DATT estimation can be drawn in this study.

4. Conclusions

In this paper, a method of DATT estimation was proposed, which combined RGB-D sensor with trained SOLOv2 model. There were 120 and 60 pairs of original RGB image and corresponding point cloud collected by Kinect V2 from two modern apple orchards. The trained SOLOv2 model was applied to segment the grafting position from RGB image and then gain pixel mask of it. This was because of its height that was above the ground did not fix. The target position of estimating DATT was determined by the pixel mask of the grafting position and the corresponding point cloud. DATT was estimated by stacking points in the X-axis direction, which were picked by three distance thresholds in X, Y, and Z directions near the target position. Distance thresholds in Y and Z directions were 0.035 m and 0.20 m, which were determined by the

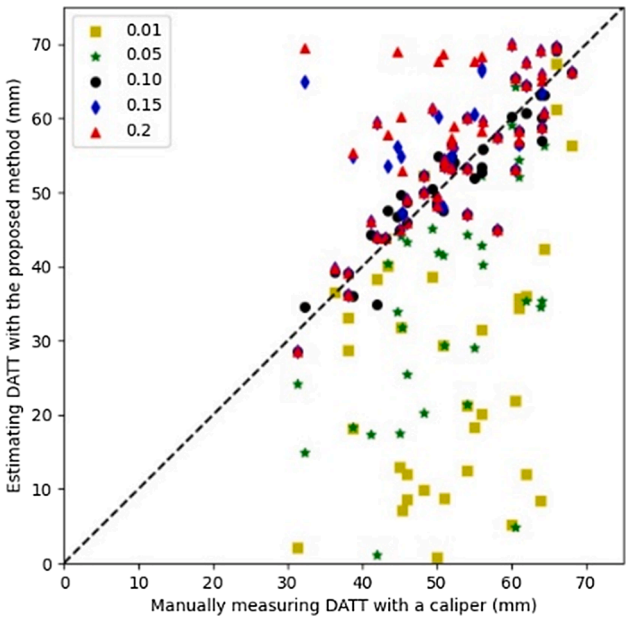


Fig. 11. Results of estimating DATT when different distance thresholds in X direction were set.

Table 2
The MPEs and MAPEs of estimating DATT when different distance thresholds in X direction were set.

Distance threshold in X direction (m)	MAE (mm)	MAPE
0.01	24.53	47.28%
0.05	11.80	23.59%
0.10	3.01	5.86%
0.15	5.37	11.39%
0.20	6.53	13.89%

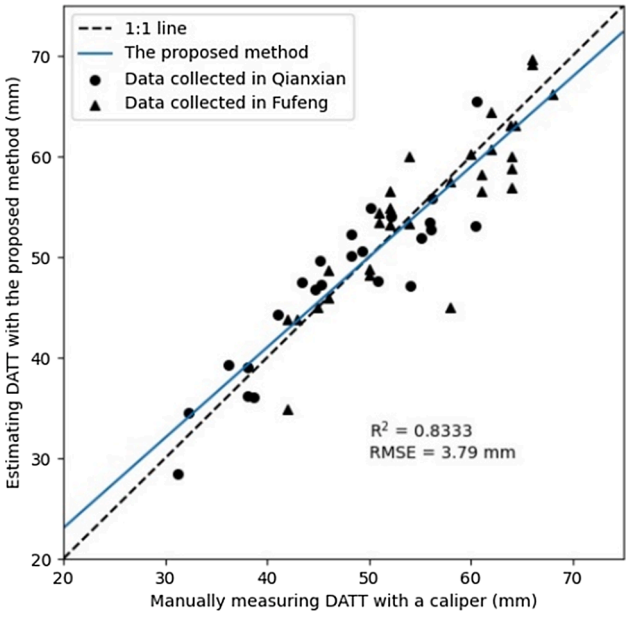


Fig. 12. Results of estimating DATT compared with manual measurement.

width range of the grafted apple tree trunk. Different distance thresholds in X directions that determined the thicknesses of the picked points affected the results of estimating DATT. The optimal distance threshold in X direction was 0.10 m. The AP and AR of the trained SOLOv2 model for segmenting the grafting position were 0.811 and 0.830, respectively. The MAE, MAPE, and RMSE of the proposed method were 3.01 mm, 5.86%, and 3.79 mm, respectively. The method proposed in this study has potential to remotely and automatically estimate DATT. Deep learning will contribute to automatically locate the target position to estimate DATT and thus further promote to precisely estimate the grafted apple tree phenotyping.

CRediT authorship contribution statement

Xiaoming Sun: Data curation, Investigation, Writing – original draft. **Wentai Fang:** Methodology, Writing – review & editing. **Changqing Gao:** Methodology, Writing – review & editing. **Longsheng Fu:** Conceptualization, Data curation, Methodology, Supervision, Writing – review & editing. **Yaqoob Majeed:** Writing – review & editing. **Xiaojuan Liu:** Methodology, Writing – review & editing. **Fangfang Gao:** Investigation, Methodology, Writing – review & editing. **Ruizhe Yang:** Methodology, Writing – review & editing. **Rui Li:** Methodology, 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

The data that has been used is confidential.

Acknowledgements

This work was partially 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); China Postdoctoral Science Foundation funded project (2021 M692656); Recruitment Program of High-End Foreign Experts of the State Administration of Foreign Experts Affairs, Ministry of Science and Technology, China (G20200027075).

References

Bruggisser, M., Hollaus, M., Otepka, J., Pfeifer, N., 2020. Influence of ULS acquisition characteristics on tree stem parameter estimation. ISPRS J. Photogramm. Remote Sens. 168, 28–40. https://doi.org/10.1016/j.isprsjprs.2020.08.002.

Corte, A.P.D., Rex, F.E., de Almeida, D.R.A., Sanquetta, C.R., Silva, C.A., Moura, M.M., Wilkinson, B., Zambrano, A.M.A., da Cunha Neto, E.M., Veras, H.F.P., de Moraes, A., Klauberg, C., Mohan, M., Cardil, A., Broadbent, E.N., 2020. Measuring individual tree diameter and height using gatoreye high-density UAV-lidar in an integrated crop-livestock-forest system. Remote Sens. 12, 863. https://doi.org/10.3390/rs12050863.

Coupel-Ledru, A., Pallas, B., Delalande, M., Boudon, F., Carrié, E., Martinez, S., Regnard, J.L., Costes, E., 2019. Multi-scale high-throughput phenotyping of apple architectural and functional traits in orchard reveals genotypic variability under contrasted watering regimes. Hortic. Res. 6, 52. https://doi.org/10.1038/s41438-019-0137-3.

Dungey, H.S., Dash, J.P., Pont, D., Clinton, P.W., Watt, M.S., Telfer, E.J., 2018. Phenotyping whole forests will help to track genetic performance. Trends Plant Sci. 23 (10), 854–864. https://doi.org/10.1016/j.tplants.2018.08.005.

Fan, Y., Feng, Z., Mannan, A., Khan, T.U., Shen, C., Saeed, S., 2018. Estimating tree position, diameter at breast height, and tree height in real-time using a mobile phone with RGB-D SLAM. Remote Sens. 10 (11), 1845. https://doi.org/10.3390/rs10111845.

Foster, T.M., Celton, J.M., Chagne, D., Stuart Tustin, D., Gardiner, S.E., 2015. Two quantitative trait loci, Dw1 and Dw2, are primarily responsible for rootstock-induced dwarfing in apple. Hortic. Res. 2, 15001. https://doi.org/10.1038/hortres.2015.1.

Fu, L., Gao, F., Wu, J., Li, R., Karkee, M., Zhang, Q., 2020a. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. Comput. Electron. Agric. 177, 105687 https://doi.org/10.1016/j.compag.2020.105687.

Fu, L., Majeed, Y., Zhang, X., Karjee, M., Zhang, Q., 2020b. 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.

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.

Gollob, C., Ritter, T., Kraßnitzer, R., Tockner, A., Nothdurft, A., 2021. Measurement of forest inventory parameters with apple ipad pro and integrated lidar technology. Remote Sens. 13, 3129. https://doi.org/10.3390/rs13163129.

Guo, W., Carroll, M.E., Singh, A., Swetnam, T.L., Merchant, N., Sarkar, S., Singh, A.K., Ganapathysubramaian, B., 2021. UAS-based plant phenotyping for research and breeding applications. Plant Phenomic. 2021, 9840192. 10.34133/2021/9840192.

Hao, Y., Widagdo, F.R.A., Liu, X., Quan, Y., Dong, L., Li, F., 2022. Estimation and calibration of stem diameter distribution using UAV laser scanning data: A case study for larch (Larix olgensis) forests in Northeast China. Remote Sens. Environ. 268, 112769 https://doi.org/10.1016/j.rse.2021.112769.

Hyyppä, E., Hyyppä, J., Hakala, T., Kukko, A., Wulder, M.A., White, J.C., Pyörälä, J., Yu, X., Wang, Y., Virtanen, J.P., Pohjavirta, O., Liang, X., Holopainen, M., Kaartinen, H., 2020. Under-canopy UAV laser scanning for accurate forest field measurements. ISPRS J. Photogramm. Remote Sens. 164, 41–60. https://doi.org/10.1016/j.isprsjprs.2020.03.021.

Iizuka, K., Yonehara, T., Itoh, M., Kosugi, Y., 2018. Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a Japanese cypress (Chamaecyparis obtusa) forest. Remote Sens. 10, 13. https://doi.org/10.3390/rs10010013.

Koh, E.J.Y., Amini, E., McLachlan, G.J., Beaton, N., 2021. Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy. Minerals Eng. 173, 107230 https://doi.org/10.1016/j.mineng.2021.107230.

Koreň, M., Hunčaga, M., Chudá, J., Mokroš, M., Surový, P., 2020. The influence of cross-section thickness on diameter at breast height estimation from point cloud. ISPRS Int. J. Geo-Inf. 9, 495. https://doi.org/10.3390/ijgi9090495.

Krisanski, S., Taskhiri, M.S., Turner, P., 2020. Enhancing methods for under-canopy unmanned aircraft system based photogrammetry in complex forests for tree diameter measurement. Remote Sens. 12, 1652. https://doi.org/10.3390/rs12101652.

Liu, L., Zhang, A., Xiao, S., Hu, S., He, N., Pang, H., Zhang, X., Yang, S., 2021. Single tree segmentation and diameter at breast height estimation with mobile LiDAR. IEEE Access 9, 24314–24325. https://doi.org/10.1109/ACCESS.2021.3056877.

Lumnitz, S., Devisscher, T., Mayaud, J.R., Radic, V., Coops, N.C., Griess, V.C., 2021. Mapping trees along urban street networks with deep learning and street-level imagery. ISPRS J. Photogramm. Remote Sens. 175, 144–157. https://doi.org/10.1016/j.isprsjprs.2021.01.016.

Luoma, V., Saarinen, N., Wulder, M.A., White, J.C., Vastaranta, M., Holopainen, M., Hyyppä, J., 2017. Assessing precision in conventional field measurements of individual tree attributes. Forests 8 (2), 38. https://doi.org/10.3390/f8020038.

McGlade, J., Wallace, L., Hally, B., White, A., Reinke, K., Jones, S., 2020. An early exploration of the use of the Microsoft Azure Kinect for estimation of urban tree Diameter at Breast Height. Remote Sens. Lett. 11 (11), 963–972. https://doi.org/10.1080/2150704X.2020.1802528.

Moe, K.T., Owari, T., Furuya, N., Hiroshima, T., 2020. Comparing individual tree height information derived from field surveys, LiDAR and UAV-DAP for high-value timber species in Northern Japan. Forests 11 (2), 223. https://doi.org/10.3390/f11020223.

Mokroš, M., Výbošt'ok, J., Tomaštík, J., Grznárová, A., Valent, P., Slavík, M., Merganič, J., 2018. High precision individual tree diameter and perimeter estimation from close-range photogrammetry. Forests 9 (11), 696. 10.3390/f9110696.

Montoya, O., Icasio-Hernández, O., Salas, J., 2021. TreeTool: A tool for detecting trees and estimating their DBH using forest point clouds. SoftwareX 16, 100889. https://doi.org/10.1016/j.softx.2021.100889.

Neuville, R., Bates, J.S., Jonard, F., 2021. Estimating forest structure from UAV-mounted LiDAR point cloud using machine learning. Remote Sens. 13, 352. https://doi.org/10.3390/rs13030352.

Pires, R.P., Olofsson, K., Persson, H.J., Lindberg, E., Holmgren, J., 2022. Individual tree detection and estimation of stem attributes with mobile laser scanning along boreal forest roads. ISPRS J. Photogramm. Remote Sens. 187, 211–224. https://doi.org/10.1016/j.isprsjprs.2022.03.004.

Shuib, N.H., Ahmad, N., Akmal Amran, M.K., Rifdi Rizuan, M.I., Rizal Wong, F.R., 2018. Girth measuring device and system based on RFID/WSN technology in rubber plantation: A preliminary data. 2018 2nd Int. Conf. Smart Sensors Appl. 2, 158–162. https://doi.org/10.1109/ICSSA.2018.8535692.

Song, C., Yang, B., Zhang, L., Wu, D., 2021. A handheld device for measuring the diameter at breast height of individual trees using laser ranging and deep-learning based image recognition. Plant Methods. 17 (1), 67. https://doi.org/10.1186/s13007-021-00748-z.

Sosna, I., 2014. Estimation of productive value of Czech origin scab-resistant apple cultivars on different rootstocks. J. Hortic. Res. 22 (2), 115–121. https://doi.org/10.2478/johr-2014-0028.

Sun, S., Li, C., Chee, P.W., Paterson, A.H., Meng, C., Zhang, J., Ma, P., Robertson, J.S., Adhikari, J., 2021. High resolution 3D terrestrial LiDAR for cotton plant main stalk and node detection. Comput. Electron. Agric. 187, 106276 https://doi.org/10.1016/j.compag.2021.106276.

Swayze, N.C., Tinkham, W.T., Vogeler, J.C., Hudak, A.T., 2021. Influence of flight parameters on UAS-based monitoring of tree height, diameter, and density. Remote Sens. Environ. 263, 112540 https://doi.org/10.1016/j.rse.2021.112540.

Tomaštík, J., Saloň, Š., Tunák, D., Chudý, F., Kardoš, M., 2017. Tango in forests – An initial experience of the use of the new Google technology in connection with forest inventory tasks. Comput. Electron. Agric. 141, 109–117. https://doi.org/10.1016/j.compag.2017.07.015.

Trochta, J., Kruček, M., Vrška, T., Kraâl, K., 2017. 3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR. PLoS ONE 12 (5), e0176871. https://doi.org/10.1371/journal.pone.0176871.

Wang, P., Li, R., Bu, G., Zhao, R., 2019. Automated low-cost terrestrial laser scanner for measuring diameters at breast height and heights of plantation trees. PLoS ONE 14 (1), e0209888. https://doi.org/10.1371/journal.pone.0209888.

Wang, X., Kong, T., Shen, C., Jiang, Y., Li, L., 2020a. SOLO: Segmenting Objects by Locations. Eur. Conf. Computer Vision 12363 LNCS, 649–665. 10.1007/978-3-030-58523-5_38.

Wang, X., Zhang, R., Kong, T., Li, L., Shen, C., 2020b. SOLOv2: Dynamic and fast instance segmentation. Adv. Neural Inf. Process. Syst. 34, 1–17. https://arxiv.org/abs/2003.10152v3.

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, X., Zhou, S., Xu, A., Chen, B., 2019. Passive measurement method of tree diameter at breast height using a smartphone. Comput. Electron. Agric. 163, 104875 https://doi.org/10.1016/j.compag.2019.104875.

Yuan, F., Fang, L., Sun, L., Zheng, S., Zheng, X., 2021. Development of a portable measuring device for diameter. Austrain J. For. Sci. 138, 25–50. https://doi.org/10.1016/j.measurement.2018.03.008.

Zhang, C., Serra, S., Quirós-Vargas, J., Sangjan, W., Musacchi, S., Sankaran, S., 2021. Non-invasive sensing techniques to phenotype multiple apple tree architectures. Inf. Process. Agric. 1–12 https://doi.org/10.1016/j.inpa.2021.02.001.

Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X., Gao, Z., 2018. Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN). Comput. Electron. Agric. 155, 386–393. https://doi.org/10.1016/j.compag.2018.10.029.

Zhang, Z., Igathinathane, C., Li, J., Cen, H., Lu, Y., Flores, P., 2020. Technology progress in mechanical harvest of fresh market apples. Comput. Electron. Agric. 175, 105606 https://doi.org/10.1016/j.compag.2020.105606.