UAV-BASED HIGH-THROUGHPUT PHENOTYPING TO SEGMENT INDIVIDUAL APPLE TREE ROW BASED ON GEOMETRICAL FEATURES OF POLES AND COLORED POINT CLOUD

Collection
Research

Wulan Mao[1,2], Bryan Murengami[1], Hanhui Jiang[1], Rui Li[1], Long He[3], Longsheng Fu[1,4,*]

[1] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi, China.
[2] Institute of Agricultural Mechanization, Xinjiang Academy of Agricultural Sciences, Urumqi, Xinjiang, China.
[3] Department of Agricultural and Biological Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.
[4] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi, China.
[*] Correspondence: fulsh@nwafu.edu.cn

HIGHLIGHTS

- Integrating RGB values and 3D coordinates of fruit trees provided phenotype-related data.
- The use of point clouds projection in tree rows segmentation has been confirmed.
- Using geometric features increased the accuracy of removing poles in an orchard.
- Segmenting fruit trees after removing poles on a single tree row was more accurate than traditional segmentation method.
- UAV equipped with LiDAR and a camera is promising for orchard high-throughput phenotyping.

ABSTRACT. *High-throughput phenotyping (HTP) of fruit trees is important for providing crop geometrical information to evaluate their high yield genotypes. Unmanned aerial vehicle (UAV) is suitable for HTP by obtaining remote sensing data of large modern apple orchards, where each tree row needs to be segmented before segmenting a single tree. This study aims to develop a method for segmenting each row without noise (ERWON) of apple trees based on integrating RGB values and three-dimensional coordinates by UAV. A robust, real-time, RGB-colored, and LiDAR-inertial-visual tightly-coupled state estimation network was used to form a dense map of the orchard, which provided datasets of colored point clouds. Supporting poles were removed from the point clouds based on the consistent number of half upper parts and lower parts. Random sampling and an effective local feature aggregator were trained to segment ERWON after pole segmentation. Results showed that a precision of 0.971, a recall of 0.984, and an intersection-over-union of 0.817 for ERWON segmentation were achieved. This method proposed a potential solution for addressing the challenge of accurately and efficiently segmenting ERWON in large orchards. It is expected to be helpful for obtaining general parameters, such as geometric, morphological, and textural characteristics, as well as more specific parameters relevant to a particular phenotyping task.*

Keywords. *Apple trees, Detection, Point cloud, RGB-colored, Segmentation.*

High throughput phenotyping (HTP) of fruit trees plays a key role in providing geometrical features to identify important genes and evaluate high yield genotypes. HTP of fruit trees includes some key indicators, such as tree height, crown length, and crown width (Herzig et al., 2021; Fullana-Pericàs et al., 2022). It is a representation of the relative effects of genetic and environmental variables on production, as well as their interactions (Li and Cheng, 2022). Accurate HTP is crucial to manage fruit trees in orchards for improving fruit quality (Yang et al., 2017). HTP parameters provide additional insights into tree health and vigor, such as assessing canopy density, identifying potential stress or disease symptoms, and monitoring the overall condition of the trees (Chivasa et al., 2021).

Unmanned aerial vehicles (UAVs) are instrumental in acquiring HTP data for fruit trees, enabling the collection of remote sensing data across extensive orchard areas. UAVs can rapidly gather pertinent information when equipped with various sensors, such as multispectral data for distinguishing between vegetation types and mapping crop yields; visible light imagery for evaluating fruit ripeness, canopy coverage, and the extent of pest and disease damage; as well as Light Detection and Ranging (LiDAR) data for detailed mapping

Submitted for review on 15 November 2023 as manuscript number ITSC 15895; approved for publication as a Research Article and as part of the Artificial Intelligence Applied to Agricultural and Food Systems Collection by Associate Editor Dr. Jianfeng Zhou and Community Editor Dr. Yiannis Ampatzidis of the Information Technology, Sensors, & Control Systems Community of ASABE on 20 May 2024.

Citation: Mao, W., Murengami, B. G., Jiang, H., Li, R., He, L., & Fu, L. (2024). UAV based high throughput phenotyping to segment individual apple tree row based on geometrical features of poles and colored point cloud. *J. ASABE, 67*(5), 1231-1240. https://doi.org/10.13031/ja.15895

  ISSN 2769-3295 https://doi.org/10.13031/ja.15895

of fruit orchards (Lyu et al., 2020; Delavarpour et al., 2023). Additionally, UAVs can autonomously navigate predetermined flight paths, altitudes, and overlaps after takeoff. UAVs are noted for their high spatial and temporal resolution and cost-efficiency. Yet, battery dependency limits their flight time. They can operate independently of weather conditions and cloud cover, providing real-time data, flexible flight operations, and adept navigation over complex farmland terrain (Bai et al., 2022). Consequently, despite the constraint of limited flight duration, UAVs are a preferred choice for swiftly collecting comprehensive remote sensing data at appropriate flight altitudes, significantly aiding the efficient acquisition of HTP data for fruit trees.

Accurate segmentation of a single tree in an orchard is a major work of HTP (Jang et al., 2020; Xie and Yang, 2020). To segment a single fruit tree, the first step is the segmentation of individual tree rows from remote sensing data of the whole orchard. In this study, remote sensing data of the whole orchard, including several tree rows, were acquired by UAV from a top view, which affected accurate segmentation of a single tree. Segmentation of one tree row first can reduce interference from other rows. It thus helps to more accurately separate a single tree from one tree row rather than the whole orchard. Therefore, a pre-work of single tree segmentation is to segment its row without other tree rows. A single tree row can be segmented by red-green-blue (RGB) images or a point cloud.

RGB images have been widely used to segment various objects from agricultural environments. Sun et al. (2022) applied the SOLOv2 model to segment grafting positions of apple trees from RGB images, which got an average precision of 0.811. Suo et al. (2022) applied a segmentation method of different apple seedling parts in images based on Blend Mask, which achieved an average precision of 85.6% on scion. Song et al. (2021) developed a method of discrete wire pixels reconstruction on progressive probabilistic Hough transform to help segment wires from images, which reached an intersection-over-union (IoU) of 0.424. Nevertheless, most of these approaches only relied on RGB values without considering three-dimensional (3D) coordinates during object segmentation.

Point cloud has been employed by considerable studies for object segmentation in orchards. Trees in the orchard were segmented by 3D coordinates obtained from LiDAR sensors with an average root mean square error of 68 mm (Liu and Zhu, 2016). Kang and Wang (2023) proposed a 3D segmentation network to segment apples based on coordinate information from point clouds and achieved a mean IoU of 0.862. Some other studies focused on canopy or tree trunk segmentation based on the coordinate information and provided similar results ($R^2 \geq 0.85$) of crown height or crown width (Kang et al., 2018; Ghanbari Parmehr and Amati, 2021; Hanssen et al., 2021; Sultan Mahmud et al., 2021). These studies demonstrated that segmentation of objects in orchards can be achieved based on the 3D coordinates of a point cloud. However, point clouds without RGB images led to inaccurate object boundaries, which affected the accuracy of object segmentation.

Integration of RGB values and 3D coordinates improves accuracy of segmenting single tree rows, which is affected by background noise. This noise includes any portion of the point cloud data that is unrelated to the intended target. Although the integration of RGB values and 3D coordinates can improve the accuracy of segmenting single tree rows (Zhang et al., 2022), issues such as over-segmentation and under-segmentation persist due to the interference of background noise including poles, wires, and ground (Yu et al., 2022). Therefore, it is critical to understand the accurate segmentation of a tree row without interference from this background noise.

The primary goal of this study was to develop a method for segmenting each row without noise (ERWON) of apple trees in large orchards by integrating RGB values with 3D coordinates obtained by UAV-based LiDAR sensors. Utilizing the geometrical features of poles, the method facilitates the automatic removal of background noise from the RGB-colored point cloud. The structure of this paper is as follows: Materials and Methods illustrates the whole workflow of the proposed extraction and classification method for segmenting the background noise and each row of apple trees in an orchard. Experimental verification of the proposed method with corresponding discussions and analyses is described in the Results and Discussions section. Conclusions of this study and future work are summarized in the Conclusions section.

Materials and Methods

Data Acquisition

A sensing system was attached to the UAV to acquire RGB values and 3D coordinates of the whole orchard. The sensing system mainly included Avia LiDAR (Livox, Shenzhen, Guangdong, China), MV-CA050-20UC RGB camera (HikRobot, Hangzhou, Zhejiang, China), Next Unit of Computing (NUC) (Intel, Santa Clara, California, USA), clamping device, and supporting frame, as shown in figure 1. LiDAR had an accuracy of 0.02 m at a range of 20 m with a field of view of 70.4° × 77.2°. The number of LiDAR returns was set from one to three. Three returns were used in this study to get more point clouds in the orchard. The RGB images had a resolution of 2592 × 2048 pixels, which were captured by a camera with a field of view of 75.6° × 70.2°. Data transmission was conducted by using the next unit of computing with a four-core I7-8550U CPU and 16 GB of RAM.

The UAV completed the orchard data acquisition task in autonomous mode. A speed of 2 m/s was chosen to complement the LiDAR sensor's frame rate of 10 fps and the

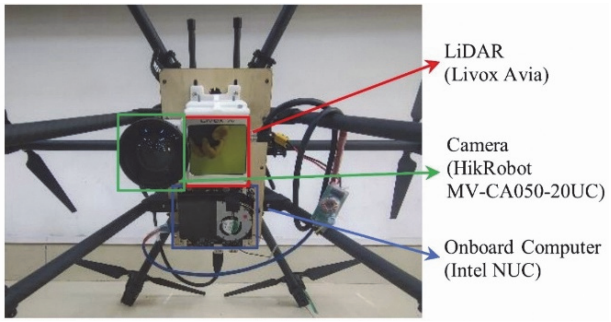

Figure 1. Sensing system with different sensors of a LiDAR and RGB camera on a UAV.

camera's frame rate of 60 fps, which allows for capturing high-quality data. A flight altitude of 15 m was selected to ensure the UAV safely passed the tallest trees in the orchard while remaining sufficient distance to capture high-resolution data. Using Mission Planner V1.3.37 (3D Robotics Inc., USA), the flight route was designed with a 30% side overlap rate to guarantee comprehensive data capture. During the data acquisition, the UAV was flown over apple tree rows at an almost constant speed. A total of 9,428 frames of point cloud and one corresponding RGB video of apple trees were acquired by the sensing system in June 2022. There were about 10,000 to 14,000 points in a frame of point clouds.

The synchronization of RGB images and LiDAR point clouds was achieved by aligning timestamps with the NUC system time, which was integrated into the drivers of both devices. This modification of the device drivers to use system time ensured precise alignment of starting times, eliminating potential data misalignment or inconsistency. Frames of point cloud and RGB video were put into a bag file for simultaneous localization and mapping (SLAM) to obtain a dense map of the whole orchard.

All data was acquired in a commercial apple orchard with a vertically supporting system near Yangling, Shaanxi, China. A supporting system in an orchard refers to an infrastructure or equipment used to facilitate the growth and health of fruit trees, such as trellises made up with poles and wires, which is shown in figure 2. Different varieties of apple trees were planted in the test orchard, such as 'YanFu No. 6,' 'YingWei,' and 'MiCui.' The height of these trees ranged from 0.6 to 2.5 m with an inter-tree spacing of 1.5 m and inter-row spacing of 3.3 m. Concrete poles were placed between every seven or eight trees to support them.

SLAM Network for Orchard Dense Mapping

After acquiring data in orchard, a SLAM network named Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state estimation and mapping package (R^{3}live) (Lin and Zhang, 2022) was applied to form a dense map of orchard. R^{3}live comprises two core subsystems: LiDAR-Inertial Odometry (LIO) and Visual-Inertial Odometry (VIO). The LIO subsystem is responsible for constructing dense global maps that capture the geometric structure through 3D positions, utilizing data from inertial and LiDAR sensors. Concurrently, the VIO subsystem enhances these maps with texture by incorporating RGB values associated with the 3D points, derived from visual-inertial sensor information. The integration within R^{3}live processes both the RGB values and the 3D coordinates to create a comprehensive and detailed representation of the environment. RGB-colored point clouds of orchard were thus gained for further apple tree segmentation.

Figure 2. Apple trees and supporting systems (including poles and wires) in orchard. X axis was along tree rows, Y axis was perpendicular to tree rows, and Z axis represented a direction of poles.

Algorithm of Segmenting Single Tree Row

RGB-colored point clouds of all trees were generated by R^{3}live, which needs automatic segmentation of each row with noise (ERWN) for further segmentation ERWON. The supporting system was a parallel plane along with tree rows, which was parallel to the XZ plane. A difference between tree rows and ground is the height in the Z axis. Therefore, Z axis values were used to segment ERWN. All RGB-colored point clouds were projected on the YZ plane for obtaining a projection perpendicular to rows. The projection was then differentiated along the Y axis to get mean points along the Z axis of different parts. These mean points formed a set with several local maximum and minimum points. These maximum and minimum points were confirmed by comparing each point's Z axis value with its two nearest points. The dividing line between tree rows was determined by the Y value of the minimum points. Therefore, each row of the apple tree was separated from RGB-colored point clouds of the whole orchard. The algorithm can be described with the following four steps.

1. Compute the projection of all RGB-colored point clouds (x, y, z) in the orchard onto the YZ plane, resulting in the set of points P.

$$P = \left\{ (y,z)_{projected} \mid (x,y,z)_{original} \right\} \tag{1}$$

2. Compute the mean Z axis coordinate M for each point in P, resulting in the function f.

$$M = \left\{ \left(y, f(y) = \frac{\sum_{(y_i,\, z_i) \in P, y_i = y} z_i}{count\left\{ (y_i,\, z_i) \in P, y_i = y \right\}} \right) \mid y in P \right\} \tag{2}$$

3. Identify the maximum and minimum points $M_{\max}$ and $M_{\min}$, of M by comparing each point's Z axis value with its two nearest points.

$$M_{max} = \left\{ (y,z)\, in\, M \mid z > f(y_{i-1})\, and\ z > f(y_{i+1}) \right\} \tag{3}$$

$$M_{min} = \left\{ (y,z)\, in\, M \mid z < f(y_{i-1})\, and\ z < f(y_{i+1}) \right\} \tag{4}$$

4. The final step is computing the Y-coordinate d of the dividing line between tree rows.

$$d = min\left\{ y \mid (y,z)\, in\, M_{min} \right\} \tag{5}$$

Methods of Removing Supporting Poles

To achieve precise HTP of apple trees, an initial and essential step involves accurately segmenting ERWON within ERWN. This necessitates the removal of poles from the background, a task attainable by leveraging their geometric characteristics. The upper half-parts depicted point clouds with a Z-value greater than half the height in each X-Y position, and the rest were lower-half parts. Poles had a rectangular-like shape in contrast to apple trees with their irregular geometrical features, which was a base for segmenting poles. Apple trees exhibited a generally spindle-shaped appearance with a higher density of canopy points than trunk points. Therefore, greater differences were found between the upper and lower-half points in apple trees than those in poles, as shown in figure 3.

A two-step method was thus proposed to remove poles from each apple tree row. First, positions of apple trees and poles in the X-Y plane were first detected (Step 1). And then the number of upper and lower parts was calculated to segment poles (Step 2). The X-Y plane was traversed using a 2D filter with a size of 0.5 m × 0.5 m to find the X-Y positions of apple trees and poles in Step 1. The point clouds with a height over 0.5 m and more than 100 points were kept to segment point clouds of ground and wire. Point clouds only included apple trees and poles after Step 1. Then the poles were targeted to be removed for a better segmentation of ERWON in Step 2. Apple trees have spindle-shaped geometry, while poles have rectangular geometry. As a result, the disparity in the number of point clouds between the upper and lower parts of apple trees was greater than that of poles. The number of point clouds between upper and lower parts of objects was counted as a part of poles if there was a difference less than 5%.

RandLA-Net for ERWON Segmentation

This study applied the random sampling local feature aggregation network (RandLA-Net), a semantic segmentation network, to segment ERWON from ERWN based on a point cloud with integrated RGB values and 3D coordinates (Chen et al., 2021). The methodology employed RandLA-Net's strategy of random point sampling from RGB-colored point clouds for efficient data processing, complemented by the training of the model on a combination of RGB values and 3D coordinates to effectively capture the semantic information of the scene. This approach demonstrated its effectiveness in several public datasets, including Semantic 3D Point Cloud Dataset (Semantic3D), Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS), and Semantic Understanding of Scenes through a 3D Mapping Dataset from Karlsruhe Institute of Technology in Germany and the Toyota Technological Institute at Chicago (Semantic KITTI) (Hu et al., 2021). This network utilized point sampling to improve segmentation efficiency in terms of memory and computation. A local feature aggregation module was further proposed to capture and preserve geometric features.

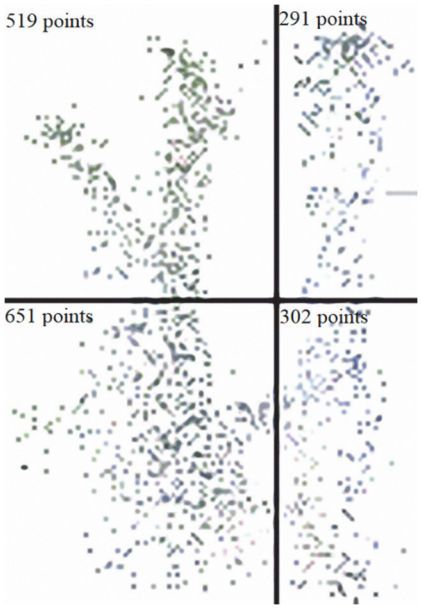

Figure 3. Different points of upper half and lower half of apple tree (left) and pole (right).

Dataset Building

The dataset was constructed using RGB-colored point clouds to train and test network performance. A TXT annotation file was created by open-source software, Cloud Compare, which contained coordinates of the point cloud (x, y, z), color components (R, G, B), and labels. ERWON and poles with noise were labeled with different labels in the TXT file, as shown in figure 4. A total of 18,470,000 points were labeled from the dense map. Annotated data was divided into a training dataset (80% of point clouds) and a testing dataset (20% of point clouds) to train and test the RandLA-Net model, respectively.

Data augmentation was applied to enlarge the point cloud training dataset, which helped to improve segmentation accuracy of minority data. To achieve a better segmentation of ERWON, this study considered most kinds of interference that may occur when segmenting objects, such as adding shelters, rotation, and down-sampling (Seidel et al., 2021). Although many studies have proved that adding shelters can improve network performance, this study did not choose adding shelters as an augmentation method because poles in orchards generated by adding shelters do not conform to the actual situation from a top view (Vargas Rivero et al., 2021; Xu et al., 2022). Rotation in the X-Y plane and region-based down-sampling on the dataset were thus implemented for data augmentation. The initial phase of augmentation through rotation involved centering the point cloud. This was accomplished by subtracting the mean of the point cloud from every individual point within the cloud. Subsequently, the rotation matrix was computed utilizing the following formula:

$$R=\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & -\cos\theta\end{bmatrix} \tag{6}$$

where θ is rotation angle, the point cloud underwent rotation by multiplying each point with the rotation matrix. Following this rotation, the points were translated back to their original center point, ensuring a coherent transformation process. Augmentation through down-sampling was executed

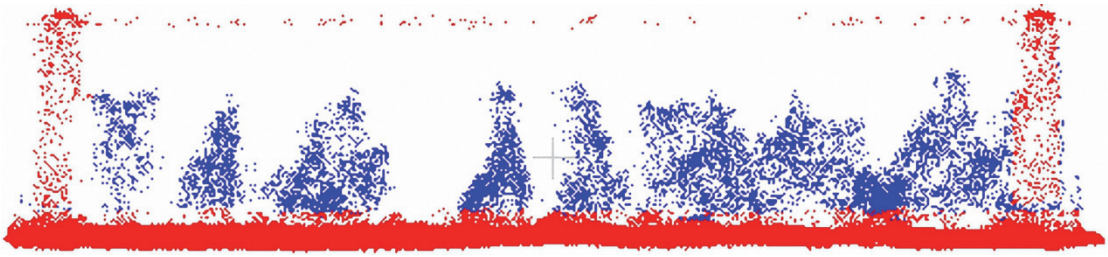

Figure 4. Different labels for ERWON (blue dots) and noise (red dots) in Cloud Compare.

using a region-based methodology. This approach partitioned the point cloud data into voxel grids, with each region uniformly down-sampled. Subsequently, these down-sampled regions were amalgamated to reconstruct new point cloud data. After this, the original 131,000 points of poles were augmented to 1,048,000 points.

Networks and Training Hyperparameters

RandLA-Net is an efficient and lightweight neural network architecture for semantic segmentation of large-scale 3D point clouds. It uses random sampling to select a subset of points from the input point cloud, and then uses a local feature aggregation module to extract both local and global features from the selected points. The extracted features are then used to predict the semantic label for each point in the point cloud. RandLA-Net was employed to train the segmentation model of ERWON. The training of the RandLA-Net network was conducted on a computer with an AMD Ryzen 7 5800X 8-Core Processor (3.80 GHz) CPU, an NVidia GeForce GTX 3080 Ti 12 GB GPU (10,240 CUDA cores), and 64 GB of memory, running on the Ubuntu 18.04 system. Software included CUDA 11.0, cuDNN 8.2.1, Tensorflow 2.6.0, and Python 3.8. Batch size and initial learning rate were set to 4 and 0.001, respectively. A total of 500,000 iterations were set to observe the training process.

Evaluation Indicators

Precision, recall, and IoU were calculated to evaluate the performance of segmenting ERWON and removing pole methods (Olson and Delen, 2008) by equations 7, 8, and 9, respectively.

$$\text{Precision} = TP / (TP + FP) \quad (7)$$

$$\text{Recall} = TP / (TP + FN) \quad (8)$$

$$\text{IoU} = TP / (TP + FP + \text{FN}) \quad (9)$$

where

TP = The number of correctly predicted labeled objects in the semantic segmentation

FP = The number of unlabeled objects with incorrect predictions in the semantic segmentation

FN = The number of missed objects in the semantic segmentation.

Results and Discussion

R^{3}live Performance in Orchard

A dense map generated by R^{3}live provided a dataset of this study. The dense map included RGB-colored point clouds of all objects in orchard, as shown in figure 5. The area covered by black point clouds indicates the location of tree rows since each row of trees was covered with a black plastic mulch. The point cloud density of black plastic mulch was higher than that of apple trees from a top view, which caused a black appearance of tree rows. Point clouds of ground were displayed in remaining parts. The dense map provided rich and detailed information about the orchard, including the location and size of the apple trees, the presence of black plastic mulch, and the topography of the ground. This information was essential for the model to accurately identify and segment the different objects in the orchard. In addition, the RGB-colored point clouds provided additional information about the objects in the orchard, such as their color and texture. This information can be helpful for the model to distinguish between different types of objects, such as trees, ground, and black plastic mulch.

The performance of R^{3}live was not quantified separately for apple trees, poles, or weeds because this was not within the scope of this study. However, the performance of R^{3}live directly influences the segmentation accuracy of poles and ERWON, and therefore, we can infer its overall performance. Nevertheless, we can conclude from the clear images provided that combining RGB values and 3D coordinates as input to R^{3}live increases the density of the generated map, facilitating the interpretation of orchard conditions.

Performance of Segmenting ERWN

The segmentation of ERWN provided robustness to ERWON segmentation. In figure 6, a visual representation of point cloud data depicting tree rows and the ground is presented in the YZ plane. The figure depicts six tree rows of similar heights, aligned parallel to the Z axis. Adjacent to the Y axis, the ground appears uneven, potentially indicating the presence of weeds amidst the apple tree rows. Notably, the tree rows have higher elevations than the ground along the Z-axis. The ground varies in height along the Z axis and forms a straight, sloping line along the Y axis. This illustration underscores the effectiveness of the proposed segmentation method for accurately delineating ERWN.

Figure 7 shows the local minimal points of the set, which were found by comparing Z-values. These points are shown as blue dots in the figure. The Y-values of the local minimal points were used to automatically segment the ERWN, which is a challenging task due to the variability in tree size and shape. The results of the segmentation are shown in figure 8. As can be seen from the figure, six apple tree rows are successfully segmented by removing ground and weeds in between rows. Unlike methods that mainly identify and

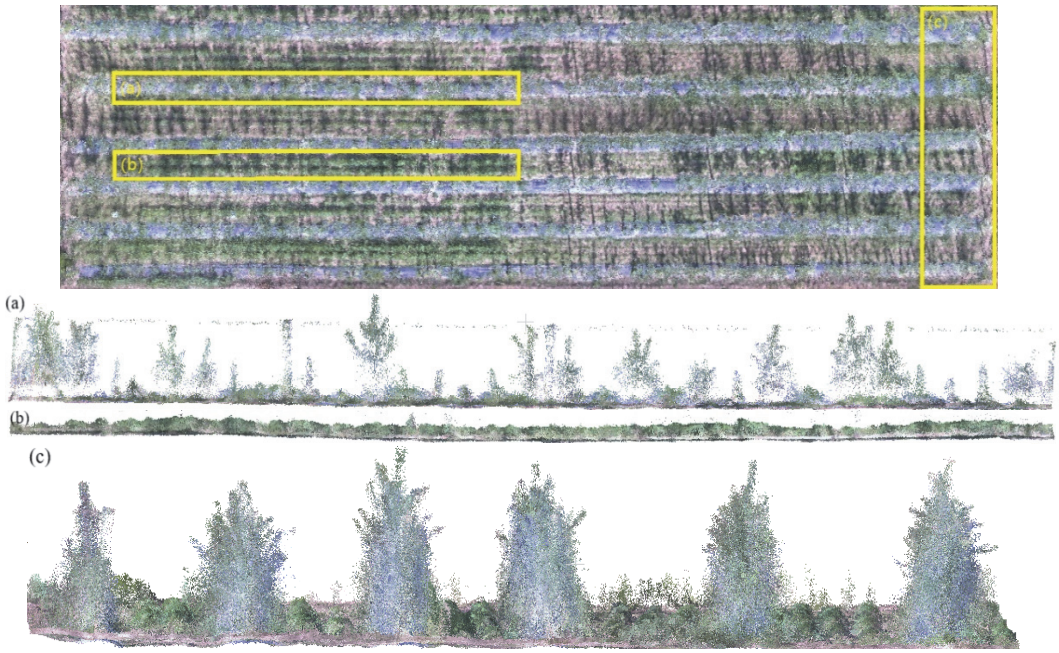

Figure 5. Dense map generated by R³live. (a) Trees and poles, (b) weeds, and (c) a side view of tree rows.

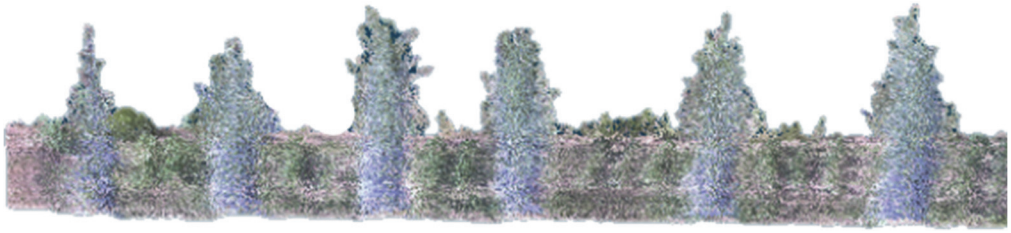

Figure 6. Point cloud data in the YZ plane, showing tree rows and the ground.

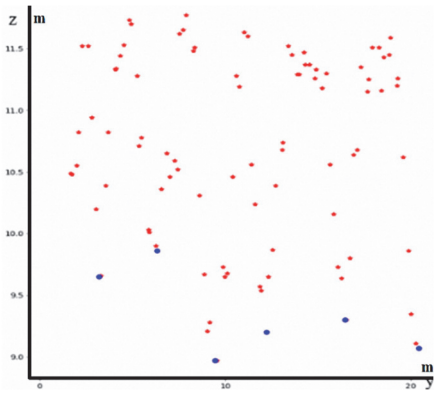

Figure 7. A set of differentiated mean points (red pots) with several local minimum points (blue pots).

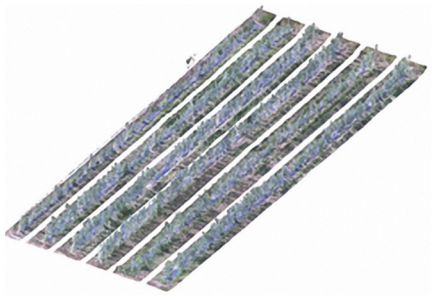

Figure 8. Segmentation of single apple tree row in orchard.

segment features by selecting specific 3D coordinates in a single image (Zine-El-Abidine et al., 2021; Straub et al., 2022), our LiDAR-based data processing approach can accurately segment a detailed map into individual rows of apple trees. This unique feature enables the rapid and efficient acquisition of high-throughput processing data.

While the performance of segmenting ERWN from the dense map was not quantified, it was a necessary step for obtaining accurate ERWON segmentation, which is the main goal in this work. The accuracy of ERWON segmentation is intrinsically linked to the initial segmentation ERWN. The effectiveness of the ERWN segmentation is contingent upon the linear arrangement of apple trees. A notable limitation of this projection method arises when confronting non-linear planting patterns, such as curved rows, where the method's applicability is constrained. Additionally, figure 8 clearly shows six segmented ERWN, which suggests that the segmentation of ERWN was successful.

Performance of Removing Poles

Different numbers of point clouds in the upper and lower parts of apple trees and poles were key to segmenting poles. As depicted in figure 9, the performance results indicated the precision, recall, and IoU values of 0.896, 0.912, and 0.817, respectively. In a top view from the UAV, some point clouds of apple trees' upper regions (crown) and lower regions (trunk) were obscured by the crown. Comparatively, there

Figure 9. Segmentation of poles (red dots) and other information (blue dots) by the proposed method.

was a higher abundance of point clouds in the upper regions of apple trees than in the lower regions. However, point clouds associated with poles remained unaffected by the apple tree's crown, resulting in general consistency between the upper and lower regions. Consequently, the method proposed in this study effectively facilitated pole segmentation and exhibited the potential for segmenting various types of fruit orchards with poles.

In this study, the pole removal method exhibited a reduced impact on other objects when compared to alternative approaches. Zeng et al. (2020) achieved a precision of 0.685 in pole segmentation, as detailed in table 1. However, their method sometimes misclassified tree trunks as poles due to their similar characteristics, leading to a blend of labels. In contrast, the proposed method began with a pre-segmentation step for poles, which attained a higher precision of 0.896. This approach significantly lowered the incidence of false positives relative to Zeng et al. (2020), thereby enhancing the precision. Such results affirm the suitability of the proposed method for pole segmentation within an orchard, optimizing the ERWON segmentation process. Nevertheless, the method has its own constraints; it is not applicable when the poles deviate from a regular rectangular shape, underscoring the need for a standardized pole structure to ensure the method's effectiveness.

Table 1. Results of different methods for segmenting poles.

Method	Object	Precision	Recall	IoU
Zeng et al. (2020)	Poles	0.685	0.981	/
Proposed method	Poles	0.896	0.912	0.817

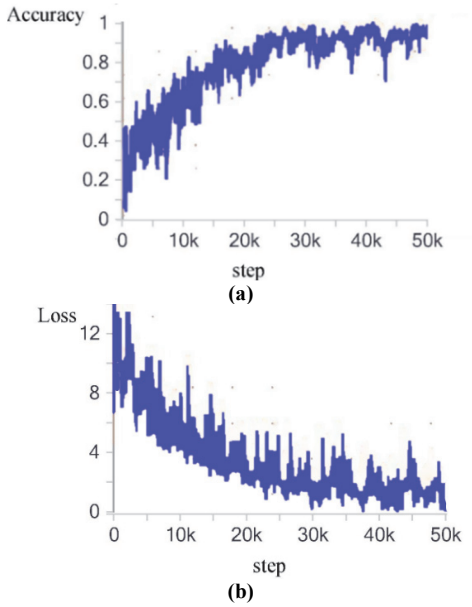

Figure 10. Training accuracy (a) and loss curve (b) of RandLA-Net network.

Training Assessment of Trained RandLA-Net Model

Loss was defined as a difference between a model's label and its predictions. Training accuracy and loss value of the trained RandLA-Net model are shown in figures 10a and 10b, respectively. The training accuracy of semantic segmentation was 0.952 when the number of epochs reached 50. The training accuracy kept improving until it ultimately stabilized at around 93%, with an increase in training iterations. A segmentation accuracy of over 90% can be applied in real applications (Wang et al., 2021). Hence, the developed segmentation method can effectively segment ERWON, which is a critical step in obtaining HTP data for orchard fruits. The loss curve exhibited a decreasing trend, demonstrating the model's learning progress over iterations. Specifically, the loss curve of the RandLA-Net model converged within 50,000 iterations. A pivotal observation was made: the loss value experienced stabilization around 42,000 iterations, reaching its lowest point at 0.05. This stabilization is indicative of the model's robustness, suggesting that it had successfully learned the intricate features of the apple trees without overfitting.

ERWON Segmentation of Trained RandLA-Net Model

Although there were some reports of applying RandLA-Net on several datasets, it showed an IoU of 0.680 on trees due to an error segmentation with cylinder like objects (Zeng et al., 2020). Therefore, RandLA-Net was applied to segmenting objects in an orchard without removing poles for comparing the performance of the method proposed in this research. The proposed method used RandLA-Net to segment RGB-colored point clouds after removing poles, as shown in figure 11.

Several performance indices, including precision, recall, and IoU, were employed to evaluate the segmentation performance of the two methods. RandLA-Net was trained under the same hyperparameters as the proposed method to ensure a fair comparison. As shown in table 2, RandLA-Net achieved a precision of 0.837, a recall of 0.876, and an IoU of 0.680 on ERWON segmentation. This relatively low accuracy can be attributed to the misclassification of certain point clouds belonging to apple trees and poles. These misclassified point clouds were typically cylinder-like objects with low heights, indicating a weakness of RandLA-Net in handling such objects. In contrast, the proposed method achieved a significantly higher precision of 0.971, a recall of 0.984, and an IoU of 0.817 on ERWON segmentation. This improvement can be attributed to the pre-processing step of removing poles before applying the RandLA-Net network. By eliminating these potentially confusing objects, the proposed method effectively mitigates the segmentation errors observed in RandLA-Net. To further enhance the segmentation performance, particularly for cylinder-like objects, several strategies can be considered, for instance augmenting the training dataset, including more examples of cylinder-like objects, such that during training the model can learn the

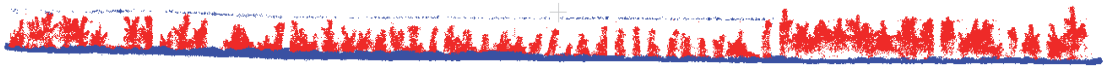

Figure 11. Segmentation of ERWON (red dots) and noise (blue dots) in orchard.

Table 2. Results of different methods for segmenting ERWON.

Method	Object	Precision	Recall	IoU
RandLA-Net	ERWON	0.837	0.876	0.680
Proposed method	ERWON	0.971	0.984	0.817

feature and the weights will be able to classify correctly during testing. Another potential improvement would be to utilize features such as texture or shape information, in addition to RGB-colored, to provide more discriminative cues for accurate segmentation.

The current study demonstrates the effectiveness of combining RGB values and 3D coordinates to enhance the acquisition of HTP data in orchards. Our results indicate a notable improvement in segmentation accuracy when these two data sources are combined, suggesting a synergistic relationship between spectral and spatial information. However, our findings also highlight the limitations of relying solely on RGB values and 3D coordinates. The presence of background elements and poles can significantly hinder the accurate segmentation of ERWON. To address this issue, we proposed a pre-processing step involving pole removal, which effectively eliminated these interfering objects and led to a substantial improvement in segmentation accuracy.

The pre-processing procedure for ERWON segmentation unfolds in four distinct steps. In Step 1, the R^{3}live was utilized to integrate RGB values with 3D coordinates captured by UAV, creating a densely mapped representation of the orchard. It is important to note the possibility of point cloud data loss due to systematic errors in LiDAR and camera data acquisition, as well as potential inaccuracies in the R^{3}live-generated colored point cloud. In Step 2, leveraging the agronomic traits of apple trees, the colored point cloud is projected onto a plane perpendicular to the tree rows, enhancing the precision of ERWN segmentation. In Step 3, the unique geometric characteristics of the supporting poles and trees are used to differentiate and remove the poles effectively. This process maintains the integrity of the point cloud data, contributing to the high precision and recall rates observed for ERWON segmentation. Nevertheless, the segmentation's success is contingent upon the regularity of orchard planting patterns and the geometric consistency of the poles. Finally, in Step 4, random sampling and a sophisticated local feature aggregation, refined through training with RandLA-Net, are employed to segment ERWON post-pole removal. Despite its improved accuracy over traditional methods, it is acknowledged that some data may still be overlooked due to the intrinsic limitations of the RandLA-Net.

Compared to methods presented in the Introduction, which have relied solely on RGB imagery or only on LiDAR point clouds for orchard segmentation, our approach yields higher precision and IoU scores. Our method demonstrated superior performance in segmenting tree rows over those that have used either data type in isolation. Our pole removal method capitalizes on the distinctive geometric features of poles and trees. Poles typically exhibit a rectangular-like shape due to their construction materials and fabrication processes, while apple trees display a more irregular and organic shape. By leveraging these geometric differences, we effectively identified and removed poles from the dense map, achieving an accuracy of 89.6%, surpassing the 68.5% achieved by Zeng et al. (2020). The high accuracy rates attained for both pole removal and ERWON segmentation demonstrated the effectiveness of our proposed method. However, it is important to note that our method for segmenting ERWON is based on orchard planting patterns and is therefore limited by the alignment of apple trees.

In the study, the integration of RGB values and 3D coordinates not only enhanced the texture features learned by RandLA-Net but also established guidance to facilitate the future extraction of information such as the number of apples, their size, color, and maturity stage (Zine-El-Abidine et al., 2021). This information is crucial for yield estimation, quality assessment, and optimization of harvest timing. The accuracy of all phenotyping parameters derived from segmented trees is inherently tied to the precision of the segmentation process. An accurate segmentation algorithm ensures that the extracted parameters faithfully represent the true characteristics of individual trees, enabling reliable phenotyping assessments.

Overall, the development of a robust single-tree segmentation method is a crucial step in advancing orchard phenotyping capabilities. The ability to accurately segment individual trees enables the measurement of a wide range of HTP parameters, providing valuable insights into tree growth, health, productivity, and overall management strategies.

CONCLUSIONS

The proposed method minimized the effect of background noise with accurate segmentation of apple tree rows, poles, and ground, which is essential in obtaining the HTP data for orchard fruits. UAVs with cameras and LiDAR can effectively collect RGB values and 3D coordinates in an orchard. The information provided phenotype-related data based on integrating the RGB-colored point clouds. Apple trees were segmented accurately using the integrated data after removing background noise. Therefore, a fruit tree grower could use the method to know the locations of each tree row and apple tree. The potential applications of this method extend to mechanized pruning, precision spraying, and various other automated agricultural operations, underscoring its promise as a valuable tool for the industry.

However, it is important to note that the current method does not encompass instance segmentation for individual apple trees. Therefore, future research endeavors may extend this methodology to include instance segmentation based on parameters such as distances and angles relative to the tree center point, enabling the acquisition of single-tree point

clouds. This evolution represents an avenue for further refinement and enhancement of the proposed approach in the context of precision agriculture.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32171897), Science and Technology Program of Yulin City, China (2023-CXY-183), and the National Foreign Expert Project, Ministry of Science and Technology, China (DL2022172003L, QN2022172006L).

References

Bai, D., Li, D., Zhao, C., Wang, Z., Shao, M., Guo, B.,... Jin, X. (2022). Estimation of soybean yield parameters under lodging conditions using RGB information from unmanned aerial vehicles. *Front. Plant Sci., 13*. https://doi.org/10.3389/fpls.2022.1012293

Chen, Y., Xiong, Y., Zhang, B., Zhou, J., & Zhang, Q. (2021). 3D point cloud semantic segmentation toward large-scale unstructured agricultural scene classification. *Comput. Electron. Agric., 190*, 106445. https://doi.org/10.1016/j.compag.2021.106445

Chivasa, W., Mutanga, O., & Burgueño, J. (2021). UAV-based high-throughput phenotyping to increase prediction and selection accuracy in maize varieties under artificial MSV inoculation. *Comput. Electron. Agric., 184*, 106128. https://doi.org/10.1016/j.compag.2021.106128

Delavarpour, N., Koparan, C., Zhang, Y., Steele, D. D., Betitame, K., Bajwa, S. G., & Sun, X. (2023). A review of the current unmanned aerial vehicle sprayer applications in precision agriculture. *J. ASABE, 66*(3), 703-721. https://doi.org/10.13031/ja.15128

Fullana-Pericàs, M., Conesa, M. À., Gago, J., Ribas-Carbó, M., & Galmés, J. (2022). High-throughput phenotyping of a large tomato collection under water deficit: Combining UAVs' remote sensing with conventional leaf-level physiologic and agronomic measurements. *Agric. Water Manag., 260*, 107283. https://doi.org/10.1016/j.agwat.2021.107283

Ghanbari Parmehr, E., & Amati, M. (2021). Individual tree canopy parameters estimation using UAV-based photogrammetric and LiDAR point clouds in an urban park. *Remote Sens., 13*(11), 2062. https://doi.org/10.3390/rs13112062

Hanssen, F., Barton, D. N., Venter, Z. S., Nowell, M. S., & Cimburova, Z. (2021). Utilizing LiDAR data to map tree canopy for urban ecosystem extent and condition accounts in Oslo. *Ecol. Indic., 130*, 108007. https://doi.org/10.1016/j.ecolind.2021.108007

Herzig, P., Borrmann, P., Knauer, U., Klück, H.-C., Kilias, D., Seiffert, U.,... Maurer, A. (2021). Evaluation of RGB and multispectral unmanned aerial vehicle (UAV) imagery for high-throughput phenotyping and yield prediction in barley breeding. *Remote Sens., 13*(14), 2670. https://doi.org/10.3390/rs13142670

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z.,... Markham, A. (2022). Learning semantic segmentation of large-scale point clouds with random sampling. *IEEE Trans. Pattern Anal. Mach. Intell., 44*(11), 8338-8354. https://doi.org/10.1109/TPAMI.2021.3083288

Jang, G., Kim, J., Yu, J.-K., Kim, H.-J., Kim, Y., Kim, D.-W.,... Chung, Y. S. (2020). Review: Cost-effective unmanned aerial vehicle (UAV) platform for field plant breeding application. *Remote Sens., 12*(6), 998. https://doi.org/10.3390/rs12060998

Kang, H., & Wang, X. (2023). Semantic segmentation of fruits on multi-sensor fused data in natural orchards. *Comput. Electron. Agric., 204*, 107569. https://doi.org/10.1016/j.compag.2022.107569

Kang, Z., Yang, J., Zhong, R., Wu, Y., Shi, Z., & Lindenbergh, R. (2018). Voxel-based extraction and classification of 3-D pole-like objects from mobile LiDAR point cloud data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 11*(11), 4287-4298. https://doi.org/10.1109/JSTARS.2018.2869801

Li, J., & Cheng, X. (2022). Supervoxel-based extraction and classification of pole-like objects from MLS point cloud data. *Opt. Laser Technol., 146*, 107562. https://doi.org/10.1016/j.optlastec.2021.107562

Lin, J., & Zhang, F. (2022). R^3LIVE: A robust, real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state estimation and mapping package. *Proc. 2022 Int. Conf. on Robotics and Automation (ICRA)* (pp. 10672-10678). IEEE. https://doi.org/10.1109/ICRA46639.2022.9811935

Liu, H., & Zhu, H. (2016). Evaluation of a laser scanning sensor in detection of complex-shaped targets for variable-rate sprayer development. *Trans. ASABE, 59*(5), 1181-1192. https://doi.org/10.13031/trans.59.11760

Lyu, B., Smith, S. D., Xue, Y., Rainey, K. M., & Cherkauer, K. (2020). An efficient pipeline for crop image extraction and vegetation index derivation using unmanned aerial systems. *Trans. ASABE, 63*(4), 1133-1146. https://doi.org/10.13031/trans.13661

Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques.* Springer. https://doi.org/10.1007/978-3-540-76917-0

Seidel, D., Annighöfer, P., Thielman, A., Seifert, Q. E., Thauer, J.-H., Glatthorn, J.,... Ammer, C. (2021). Predicting tree species from 3D laser scanning point clouds using deep learning. *Front. Plant Sci., 12*. https://doi.org/10.3389/fpls.2021.635440

Song, Z., Zhou, Z., Wang, W., Gao, F., Fu, L., Li, R., & Cui, Y. (2021). Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting. *Comput. Electron. Agric., 181*, 105933. https://doi.org/10.1016/j.compag.2020.105933

Straub, J., Reiser, D., Lüling, N., Stana, A., & Griepentrog, H. W. (2022). Approach for graph-based individual branch modelling of meadow orchard trees with 3D point clouds. *Precis. Agric., 23*(6), 1967-1982. https://doi.org/10.1007/s11119-022-09964-6

Sultan Mahmud, M., Zahid, A., He, L., Choi, D., Krawczyk, G., & Zhu, H. (2021). LiDAR-sensed tree canopy correction in uneven terrain conditions using a sensor fusion approach for precision sprayers. *Comput. Electron. Agric., 191*, 106565. https://doi.org/10.1016/j.compag.2021.106565

Sun, X., Fang, W., Gao, C., Fu, L., Majeed, Y., Liu, X.,... Li, R. (2022). Remote estimation of grafted apple tree trunk diameter in modern orchard with RGB and point cloud based on SOLOv2. *Comput. Electron. Agric., 199*, 107209. https://doi.org/10.1016/j.compag.2022.107209

Suo, R., Fu, L., He, L., Li, G., Majeed, Y., Liu, X.,... Li, R. (2022). A novel labeling strategy to improve apple seedling segmentation using BlendMask for online grading. *Comput. Electron. Agric., 201*, 107333. https://doi.org/10.1016/j.compag.2022.107333

Vargas Rivero, J. R., Gerbich, T., Buschardt, B., & Chen, J. (2021). Data augmentation of automotive LIDAR point clouds under adverse weather situations. *Sensors, 21*(13), 4503. https://doi.org/10.3390/s21134503

Wang, K., Zhou, J., Zhang, W., & Zhang, B. (2021). Mobile LiDAR scanning system combined with canopy morphology extracting methods for tree crown parameters evaluation in orchards. *Sensors, 21*(2), 339. https://doi.org/10.3390/s21020339

Xie, C., & Yang, C. (2020). A review on plant high-throughput phenotyping traits using UAV-based sensors. *Comput. Electron.*

Agric., 178, 105731. https://doi.org/10.1016/j.compag.2020.105731
Xu, S., Zhou, X., Ye, W., & Ye, Q. (2022). Classification of 3-D point clouds by a new augmentation convolutional neural network. *IEEE Geosci. Remote. Sens. Lett., 19*, 1-5. https://doi.org/10.1109/LGRS.2022.3141073
Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H.,... Yang, H. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci., 8*. https://doi.org/10.3389/fpls.2017.01111
Yu, Q., Yang, H., Gao, Y., Ma, X., Chen, G., & Wang, X. (2022). LFPNet: Lightweight network on real point sets for fruit classification and segmentation. *Comput. Electron. Agric., 194*, 106691. https://doi.org/10.1016/j.compag.2022.106691
Zeng, L., Feng, J., & He, L. (2020). Semantic segmentation of sparse 3D point cloud based on geometrical features for trellis-structured apple orchard. *Biosyst. Eng., 196*, 46-55. https://doi.org/10.1016/j.biosystemseng.2020.05.015
Zhang, B., Wang, R., Zhang, H., Yin, C., Xia, Y., Fu, M., & Fu, W. (2022). Dragon fruit detection in natural orchard environment by integrating lightweight network and attention mechanism. *Front. Plant Sci., 13*. https://doi.org/10.3389/fpls.2022.1040923
Zine-El-Abidine, M., Dutagaci, H., Galopin, G., & Rousseau, D. (2021). Assigning apples to individual trees in dense orchards using 3D colour point clouds. *Biosyst. Eng., 209*, 30-52. https://doi.org/10.1016/j.biosystemseng.2021.06.015