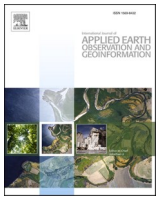
International Journal of Applied Earth Observations and Geoinformation 114 (2022) 103055

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



An automated, high-performance approach for detecting and characterizing broccoli based on UAV remote-sensing and transformers: A case study from Haining, China

Chengquan Zhou [a], Hongbao Ye [a,*], Dawei Sun [a], Jibo Yue [b], Guijun Yang [c], Jun Hu [d,e,*]

[a] Institute of Agricultural Equipment, Zhejiang Academy of Agricultural Sciences, Hangzhou 310000, China
[b] College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China
[c] Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture P. R. China, Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China
[d] Food Science Institute, Zhejiang Academy of Agricultural Sciences, Hangzhou 310000, China
[e] Key Laboratory of Postharvest Preservation and Processing of Vegetables (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Hangzhou 310000, China

ARTICLE INFO

Keywords:
Transformers
Multi-sensor
Broccoli detection and characterization
Canopy mapping
Volume estimation

ABSTRACT

Accurate canopy mapping and head-volume estimation of large areas of broccoli is an important prerequisite for precision farming since it provides important phenotypic traits associated with field management, environmental control, and yield prediction. Currently, the detection and characterization of broccoli mostly rely on ground surveys and human interpretation, which is often time- and labor-intensive. Recent developments based on unmanned aerial vehicle (UAV) remote sensing offer low cost, timely, and flexible data acquisition, thereby providing a potential alternative technique to enhance *in situ* field surveys. The combination of UAV data and deep learning has led to a series of breakthroughs in rapid and automated collection of simultaneous multisensor and multimodal plant phenotyping data. However, their application for monitoring broccoli remains problematic when faced with the significant spatial scale involved and the variety of vegetation species. To address this problem, we propose herein a fast and reliable semi-automatic workflow based on deep learning to process UAV RGB imagery and LiDAR point clouds and thereby remotely detect and characterize broccoli canopy and heads. First, we explore the use of TransUNet to differentiate canopy and non-canopy regions in RGB images at the individual-plant scale. The results demonstrate that TransUNet consistently achieves the highest accuracy (average returned Precision, Recall, F1 score, and IoU of 0.917, 0.864, 0.901, and 0.895, respectively) compared with three CNN-based and two shallow learning-based approaches. In addition, TransUNet performs best in terms of robustness against variations in training samples. Subsequently, to estimate the volume of broccoli heads, a point cloud transformer (PCT) network is developed for point cloud segmentation. Improving upon the results of three existing methods PointNet, PointNet++, and K-means that were applied to the same datasets, the best-performing PCT produced a precision of 0.914, an overall recall of 0.899, an overall F1 score of 0.901, and an overall IoU of 0.879. A regression analysis indicates that the PCT estimates had $R^2 = 0.875$, RMSE = 18.62, and rRMSE = 3.64 %, which is also superior to the results from other comparison approaches. Collectively, the wide application of such technology would facilitate applied research in plant phenotyping and precision agro-ecological applications and field management.

1. Introduction

Broccoli (*Brassica oleracea* L. var. *italica*), a popular vegetable that is widely cultivated across the world, is rich in glucosinolates, fibers, and vitamins (Li et al., 2019). The demand for broccoli is increasing and the production was over 27.5 million tons in 2020, with China as the leading producer, accounting for more than 39 % of the global production (https://www.fao.org). Sustainable management for large areas of

* Corresponding authors at: Institute of Agricultural Equipment, Zhejiang Academy of Agricultural Sciences, Hangzhou 310000, China (H. Ye). Food Science Institute, Zhejiang Academy of Agricultural Sciences, Hangzhou 310000, China (J.Hu).
E-mail addresses: yhb2008@zaas.ac.cn (H. Ye), hujun@zaas.ac.cn (J. Hu).

https://doi.org/10.1016/j.jag.2022.103055
Received 5 July 2022; Received in revised form 9 October 2022; Accepted 10 October 2022
Available online 19 October 2022

broccoli increasingly requires accurate information about the canopy and heads. Estimations of broccoli canopy and broccoli-head volume must be adaptable to the different varieties or genotypes, especially for field management and automated harvest. Traits derived from such information offer useful data for monitoring the growth stages and provide reference points for harvest strategies (Blok et al., 2021). To date, the most accurate approach to identifying such agronomic traits is still *in situ*, manual field surveying, which is extremely time- and labor-intensive, making it expensive (Kusumam et al., 2017). In addition, it is not operationally feasible for large-scale measurements. Therefore, a critical need exists for an efficient method to collect broccoli phenotypic data with high accuracy in a non-invasive manner (see Fig. 1).

Recently, various technological developments have created opportunities to replace routine manpower field surveys. A combination of two advances offers significant potential for accurate detection and characterization of broccoli: unmanned aerial vehicles (UAVs) and deep learning (DL) (Zhang et al., 2016). UAVs have a high potential for fine-grained observation and analysis because they can deliver ultra-high-resolution data and permit a flexible revisit time (Deng et al., 2018). In recent decades, UAVs have been extensively applied in many domains to enhance survey quality, such as urban remote sensing (Lobo Torres et al., 2020), wildlife population census (Rey et al., 2017), and land cover mapping (Mahdianpari et al., 2018). Furthermore, the UAVs are also advantageous for crop- and vegetation-related applications involving biomass estimation (Maimaitijiang et al., 2019), disease diagnosis (Kerkech et al., 2020), and canopy mapping (Alonzo et al., 2020).

Information on the location and size of the vegetation canopy is essential for farmers because it describes productivity and growth vigor (Yuan et al., 2019). Although they return a good performance for canopy mapping, sensors such as thermal infrared (Shirzadifar et al., 2020) and hyperspectral cameras (Nezami et al., 2020) are high-priced and computationally expensive. To circumvent this, recent studies involving canopy mapping have used RGB-based sensors because of their low cost, fine spatial detail, and high market availability (Schiefer et al., 2020; López-Jiménez et al., 2019; Franklin and Ahmed, 2018). For instance, Santos et al. (2020) applied an UAV structure-from-motion (SfM) approach to determine whether SfM can be used to estimate the height and crown diameter of coffee trees. Their research highlights the potential of UAV-SfM for determining the biophysical parameters of trees on a large scale. Yan et al. (2019) used a color mixture analysis (CMA) method to improve the accuracy and efficiency of mapping fractional vegetation cover. A comparison with three other estimation algorithms (e.g. FCLS, HAGFVC and LAB2) shows that the proposed CMA is more accurate and robust against variations in environmental conditions. In other work, Hassanein et al. (2019) introduced a low-cost approach to detect crop rows based on UAV RGB imagery. Their method has three main steps: color space conversion, section generation, and scan-line generation. After evaluation based on images acquired at various heights and on various dates, this method proved to be useful for field applications.

Nevertheless, using only RGB information, accurate structure information on vegetation cannot be obtained (Jin et al., 2021). As a promising active technology, LiDAR can penetrate canopies and detect understory plants by emitting laser pulses and recording return pulse time, thereby providing high-precision, three-dimensional (3D) information on plants. Many successful efforts have been put forth for estimating vegetation height (Fagua et al., 2019), coverage (Wu et al., 2019), leaf area index (Jin et al., 2020), above-ground biomass (Wiering et al., 2019), crown size and volume (Duncanson et al., 2015; Colaço

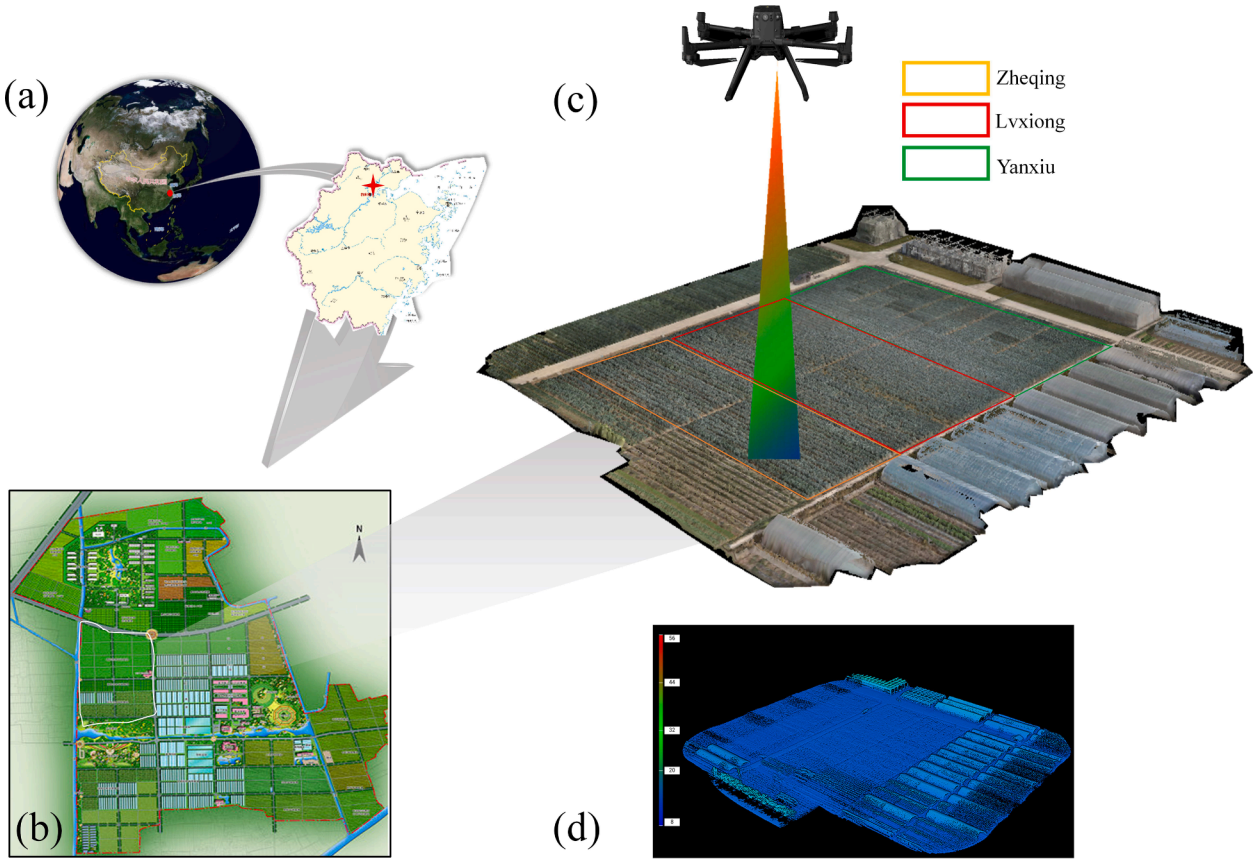


Fig. 1. Overall visualization of study area. (a) Location of study site, Haining County, Zhejiang, China. (b) Overview of Zhejiang Academy of Agricultural Sciences Yangdu Scientific Research Innovation Base. (c) Orthorectified image produced from the UAV flight with field plots of broccoli overlaid. The colored rectangles show the location of the three broccoli cultivars. (d) Digital surface model obtained from UAV.

et al., 2017), etc. using LiDAR sensors. For example, Brede et al., (2019) compared UAV laser scanning with terrestrial laser scanning for estimating tree volume with a TreeQSM method. Moorthy et al. (2019) presented a machine-learning-based leaf and wood classification method to classify 3D point clouds into woody and leafy components. The proposed method outperformed existing methods in most cases without a post-processing step. Hadas et al. (2019) described a workflow to determine the geometric parameters of apple trees in orchards. A robust method was proposed that combines an alpha-shape algorithm, principal components analysis, and detection of local minima on crown profiles. The experimental results reveal a strong correlation between LiDAR data analysis and field measurements.

Early efforts such as watershed (Natesan et al., 2019), support vector machine (Fassnacht et al., 2016), and region growing (Tao et al., 2015) focus on extracting low-level color or geometric features from pixels and point clouds. DL algorithms, especially convolutional neural networks (CNNs), autonomously extract deep, high-level features to better describe targets and uncover more complex and hierarchical relationships (Hoeser and Kuenzer, 2020). Unlike the traditional shallow-learning approaches, as an end-to-end inductive learning process, CNNs no longer rely on rule-based algorithms but can learn the most robust characteristics automatically, thereby reducing reliance on human expertise. Following the success of DL, many canopy mapping studies used CNNs for mapping complex scenes over large areas, such as U-Net, FCN, and DeepLab v3+ (Kattenborn et al., 2021). For point cloud segmentation, the Point-based approaches such as PointNet (Li et al., 2018) and PointNet++ (Qi et al., 2017) run directly on raw point clouds to learn 3D representation by using Multi-Layer Perceptrons, max-pooling, and rigid transformations. The combination of UAV data and CNNs may thus facilitate vegetation detection and delineation (Osco et al., 2020).

Nevertheless, the CNN-based approach adaptation for monitoring and characterizing broccoli via the UAV platform remains challenging (Reichstein et al., 2019). Initially, although UAVs and CNN-based methods have proven their feasibility in RGB and LiDAR data processing, they are rarely applied to broccoli-related tasks. Workers with a background in broccoli breeding and planting may lack the requisite high levels of domain expertise and practical experience in computer science. Conversely, the missing of related prior knowledge in horticulture and agronomy leading to the declines in model performance. Another notable limitation of methods based on CNNs is the need for an adequate number of training samples (Martins et al., 2021). The high labor and time costs for additional flights and field surveys make it difficult to train a deep CNN with massive network parameters. Currently, most standard DL models are pre-trained based on publicly accessible labeled datasets. Many such datasets contain multiclass targets (e.g., cats, dogs, houses, people, ships). CNNs perform well because the objects of different classes have significant distinguishable features, but broccoli, weeds, or other vegetation appear much more similar. In this case, even the most advanced transfer learning algorithm cannot take full advantage of the existing annotations without negative-transfer problems (Pandey and Jain, 2021). Finally, limited by the locality of convolution operations, CNNs tend to focus on local details, without capturing the global information. Therefore, these architectures usually perform poorly, especially for targets that show large variation in terms of texture, size, and shape. As the above analysis indicates, a need exists for a practical solution that can systematically enhance the feasibility of UAVs and DL for monitoring and characterizing broccoli. Recently, transformers have been applied to image vision tasks, performing better than popular CNNs due to their special decoder-encoder structure. As its core component, the self-attention mechanism generates a refined attention feature for its input feature based on the global context. Unlike prior CNN-based methods, all transformer operations are parallelizable and order-independent, which make it not only powerful for modeling global contexts but also demonstrates superior transferability for downstream tasks under large-scale pre-training.

Inspired by the superior performance of transformers in the field of natural language processing (NLP), the primary aim of this study is to integrate and validate methods for detecting and characterizing broccoli by using a lightweight UAV platform and Transformers. Our main contributions can be summarized as follows:

(1) We introduce an advanced transformer-based algorithm (TransUNet) to improve the accuracy and efficiency of broccoli canopy mapping across different cultivars, regions, and light conditions.
(2) We estimate broccoli-head volume by using a point cloud transformer (PCT) with reference data manually labeled in a dense point cloud and also provided by field surveys.
(3) We apply the results of the study in real applications and thereby demonstrate that the proposed integrated framework outperforms existing methods applied to the same datasets.

2. Materials and methods

2.1. Study site

This study was performed in a 0.75 ha area at the Zhejiang Academy of Agricultural Sciences, Yangdu Scientific Research Innovation Base, Haining County, Zhejiang Province, China (30°27′ N, 120°25′ E). The region has a subtropical monsoon climate, with an average rainfall of 1187 mm and an annual average temperature of 15.9 ℃. The mean elevation is 30 m above sea level with the hottest temperature in July and the coldest in January. Broccoli was grown in soil with a pH of 6.5 and an organic matter content of over 30 g/kg. After tillage, the test site design was broken up into three main plot replications for three broccoli cultivars: Zheqing, Lvxiong, and Yanxiu. We chose these three cultivars because they are the dominant broccoli types in the area. Next, each plot was split into 150 sub-plots (1.5 m × 6 m in size). All cultivars were transplanted on September 10, 2021 into 7-cm-deep rows 0.5 m apart and at a density of 2600 seeds per 667 m^2. Before cultivation, 1000 kg organic fertilizer, 20 kg ternary compound fertilizer, and 1 kg borax were applied per 667 m^2. We determined the nitrogen level based on the growth status and weather conditions. Herbicide was applied at 8.4 kg/ha during the experiment to control weeds.

2.2. Data acquisition

2.2.1. RGB and airborne laser scanning data

We used visible light and LiDAR sensors with a real-time kinematic global positioning system to explore the potential of UAV data to detect and estimate changes in broccoli canopy and head which have not been quantified previously due to the lack of a dataset. The UAV campaigns were conducted on October 31 and November 25, 2021, under clear skies and with low wind speeds from 10:30 am to 12:00 am local time. The flight mission was pre-programmed to fly at 30 m height and at a speed of 2.5 m/s to achieve an 80 % forward overlap rate and 80 % sidelap. The entire study area was covered by 22 long parallel flight lines and 1 short parallel flight line. The UAV was constantly oriented parallel to the flight line, so that the mission was completed in 20 min. A DJI Matrice M300 RTK (DJI Technology Co., Shenzhen, China) was used as a base for the platform, with mission planning done by using the DJI iPad Ground Station Pro application. The maximum horizontal flight speed is 23 m/s, and the flight duration varies between 25 and 55 min depending on the payload. A ZENMUSE *P*1 digital camera and a ZENMUSE L1 laser scanner were mounted underneath the UAV to a motorized gyroscopically stabilized gimbal to acquire RGB images and point cloud data. The ZENMUSE *P*1 is a 45-megapixel camera equipped with a 2.67-inch full-frame CMOS sensor. The focal length was 35 mm, and a shutter speed of 1/1000 s was used to ensure the best optimal exposure without being affected by motion. For each flight, the ISO was adjusted from 500 to 600 in view of the local brightness conditions. All pictures were recorded in RAW format and then converted into JPEG format without any

compression. According to the chosen flight height, a ground sample distance of 1.5 cm was maintained. Likewise, the ZENMUSE L1 is a survey-grade laser scanner with a field of view of 70.4° × 4.5° under linear repeat scan mode. The instrument has a maximum effective point cloud data rate of 240 000 points/s and a maximum range of 450 m. Prior to scanning, the L1 was preheated for several minutes. The sampling frequency was set to 240 kHz with a mean LiDAR point density of 1465 points/m^2 produced at a flight speed of 2.5 m/s. (Fig. 2).

2.2.2. Field-survey data

The field data included a variety investigation, and the broccoli heads were immediately measured after the UAV survey to provide ground-truth information. First, reference data of planting varieties were collected by visual inspection. The survey factors include the coordinates of the four corner points of each sample plot, broccoli species, and seeding density. Second, we used a measuring rod to measure the diameters and heights of the broccoli heads. Note that we measured the heights of the broccoli heads regardless of stems. To ensure the visibility of the broccoli heads from 30 m above, only broccoli heads with a diameter at least of 50 mm or more were recorded in the survey data. Before the field investigation, the study area, sampling plots, planting varieties, and number of seeds were predetermined by visual interpretation of UAV RGB data. For this study, a total of 28 512 broccoli heads were accurately measured, including 9750 Zheqing, 9346 Lvxiong, and 9416 Lvxiong, which corresponded to all broccoli heads visible on the UAV imagery.

2.3. Data preprocessing

A total of 727 RGB images with spatial dimension of 8192 × 5460 pixels were acquired and used in this study. The DJI Terra (DJI Technology Co., Shenzhen, China) software was used to orthorectify and mosaic the RGB images. This tool combines flight route planning with two- and three-dimensional reconstruction to produce unordered but overlapping remote images. Nine accurate and reliable ground control points, distributed evenly in the study site, were established in the study area to improve the vertical and horizontal accuracy, which not only facilitated the extraction of geometrical features between images but also helped compare multiple datasets captured throughout the season. We followed the standard of DJI Terra processing procedures: importing the raw images, arranging and selecting the appropriate images according to the overlap rate and definition, importing the corresponding POS data, manually aligning the ground targets to differentially corrected ground control points, and selecting the reconstruction type and clarity. The reconstruction clarity was determined by setting the module in DJI Terra to "high" to maximize the accuracy of sparse and dense 3D clouds, the digital surface model, and mosaics.

The raw lidar data were provided as LAS files containing the corresponding reflectance value for each point. The raw data were preprocessed following the guidelines of the DJI Terra software for point cloud data preprocessing. This procedure imports raw data, chooses the point cloud density, selects the coordinate and elevation system, and sets the effective distance of the point cloud. In addition, the point cloud accuracy was used to improve the consistency of the processing result.

2.4. TransUnet framework for broccoli canopy mapping

In brief, training a CNN model with a deeper architecture and more parameters may be challenging because of restrictions on the amount of training data and the computational cost. Recently, Transformers have emerged as a new deep learning paradigm due to their novel attention mechanism and superior performance in capturing global information (Zhang et al., 2022). In this section, we train a TransUNet framework for precise broccoli canopy mapping from UAV imagery for all experiments.

Our workflow for mapping the broccoli canopy consists of mixed manual and automated steps which can be divided into three main steps: (1) data organization, (2) data augmentation, and (3) architecture design and training.

2.4.1. Data organization

In this first step, we split the original mosaic images into a number of square 512 × 512-pixel sub-images to save running time and reduce the memory budget of our GPU. In this study, 3000 candidate patches were extracted from the acquired images using ArcGIS v.10.6.1 (ESRI, Redlands, CA, USA) software. In the labeling step, a LabelMe tool (MIT's Computer Science and Artificial Intelligence Laboratory, Massachusetts, USA) was applied to label the canopy images. Two expert workers hand-labeled each sub-image at the pixel level into two classes: canopy and background. Manual correction was done by another expert worker, which consisted in correcting possible misclassification and ensuring the reliability of the reference data. The polygon modular was used to label the outline of the broccoli canopy, and all annotated files were stored in JSON format.

2.4.2. Data augmentation

To improve the quantity of training data and generalization of the mapping method, a huge amount of data was required. The number of labeled sub-images in our study remains very low and is insufficient for

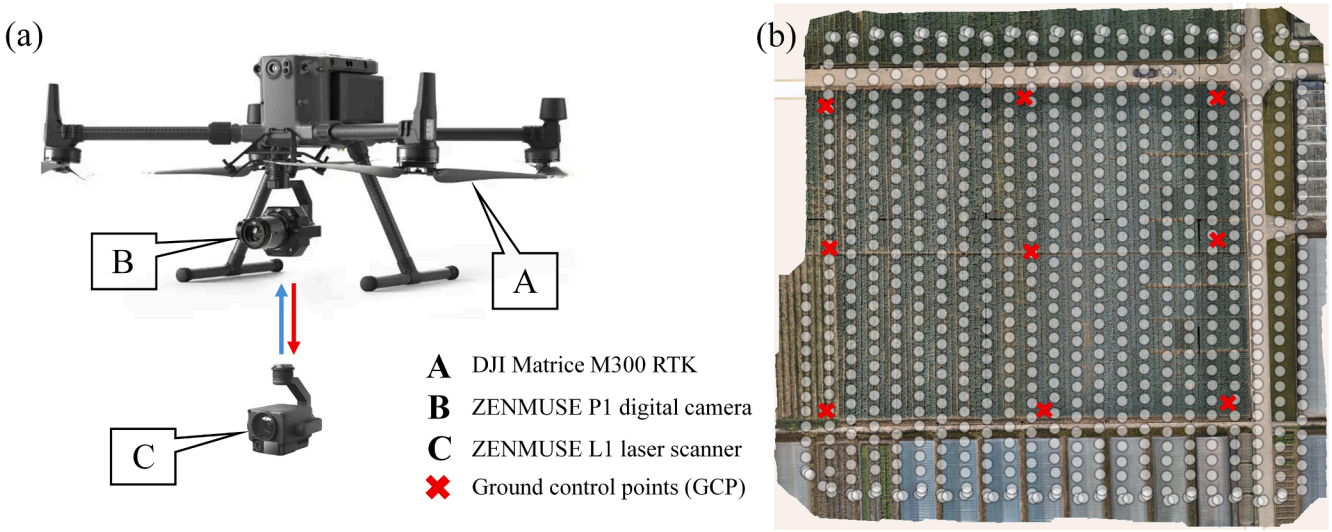


Fig. 2. (a) UAV systems and integrated sensors. (b) Location of flight path.

proper training of the CNN modules. To improve the overall learning procedure and performance, several data augmentation practices, including rotating (45°, 90°, 135°, and 180°), mirroring (about the horizontal and vertical axes), and modifying the brightness (coefficients between 0.8 and 1.2 with a step of 0.1 multiplied by the grey level), were implemented by using the "*imgaug*" Python package, which strives to increase the amount of labeled data through this transformation process. Eventually, 3000 training samples along with the corresponding ground-truth data were augmented to 9000 samples, with 90 % used for training and validation and the remaining 10 % held for testing. This approach made the training model much more robust against variations in light and viewpoint and other small variations in outdoor conditions or devices.

2.4.3. Architecture design and training

This subsection briefly introduces TransUNet, which is based on the open-source implementation by Chen et al. (2021) (https://github.com/Beckschen/TransUNet). TransUNet follows the basic design of U-Net, which uses the skip-connections strategy between layers to enhance local details. In contrast with traditional U-Net, TransUNet combines a transformer with the U-Net architecture and introduces a CNN-transformer hybrid encoder by incorporating several transformer layers during feature extraction. The advantage of TransUNet comes from combining a CNN and a transformer, which leverages both detailed high-resolution spatial information from CNN features and the global context encoded by Transformers. In the present study, the CNN-transformer hybrid encoder consists of a ResNet-50 and a Vision Transformer. Given an augmented sample, the ResNet-50 first down-samples three times to generate a feature map, then splits the feature map into N non-overlapping 1×1 patches for patch embedding. The transformer encoder consists of 12 blocks containing a multi-head self-attention and a multi-layer perceptron. Ultimately, a cascaded upsampler, which upsamples several times to decode the hidden feature for generating the final outputs, was used as decoder.

Fig. 3 shows the training architecture for broccoli canopy mapping based on TransUNet. Note that we did not change the architecture of the original model but adapted the input data source to achieve fast and accurate processing. The training platform included a Precision 3630 desktop (Dell Inc., Texas, USA) with Intel (R) Core (TM) i7-8700 K 3.7 GHz × 12 processor CPU, 32 GB of memory and a NVIDIA RTX 3080Ti GPU (10240 CUDA cores) with 12 GB of internal RAM under Microsoft Windows 10 Professional operating system. Other software tools including CUDA 9.0, CUDNN 7.1, and Python 3.7 were configured in the PyTorch 1.0.1 framework. For network training, the batch size was set to 24 and 20 000 training iterations were used. The optimizer, learning rate, momentum, weight decay, and loss function were set as Adam, 0.01, 0.9, 0.0001, and focal loss, respectively.

2.5 Vol. ume estimates

This section discusses the framework for estimating the broccoli-head volume from the laser data. The estimate of broccoli-head volume was preprocessed with the help of the LiDAR 360 software. LiDAR 360, developed by Green Valley Technology. Co., ltd. (Haidian, Beijing, China), was used to separate aboveground points from ground points. The UAV lidar data were processed using the toolbox for individual tree segmentation to (1) remove the noise point through a filtering program, (2) classify the lidar point cloud into ground points and non-ground points, (3) divide the *xy* plane into a 0.5 m × 0.5 m grid and generate a digital terrain model by interpolating the ground points, and (4) normalize the z coordinates of points by using the final digital terrain model. Inspired by the superior performance of transformer, a PCT network was chosen to detect broccoli heads at segment levels (https://github.com/MenghaoGuo/PCT).

Briefly, the PCT contains three important features: an input embedding module, an attention module, and a classification and segmentation network. The aim of input embedding was to map the point cloud from the *xyz* space to 128-dimensional space in two ways: point embedding and neighbor embedding. In the attention module, self-attention was combined with the offset-attention mechanism. Fig. 4 shows the overall framework of the PCT. A total of 6000 broccoli-head regions were first manually marked for training along with 2000 broccoli-head regions for evaluation. The Adam strategy was used to optimize the networks with the batch size, initial learning rates, and weight decay set to 32.20, 0.01, and 0.0001, respectively. The soft cross-entropy loss function used by Wang et al. (2019) was adopted. After segmenting the point cloud, the average broccoli-head volume was calculated by using the convex hull algorithm.

2.6. Accuracy assessment

The accuracy of automated canopy mapping and volume estimation for broccoli was first accessed at the pixel level. We define model performance in terms of the usefulness of the predictions for precise mapping of the canopy. Our evaluation metric consists of Precision (also known as user's accuracy), Recall (also known as producer's accuracy),

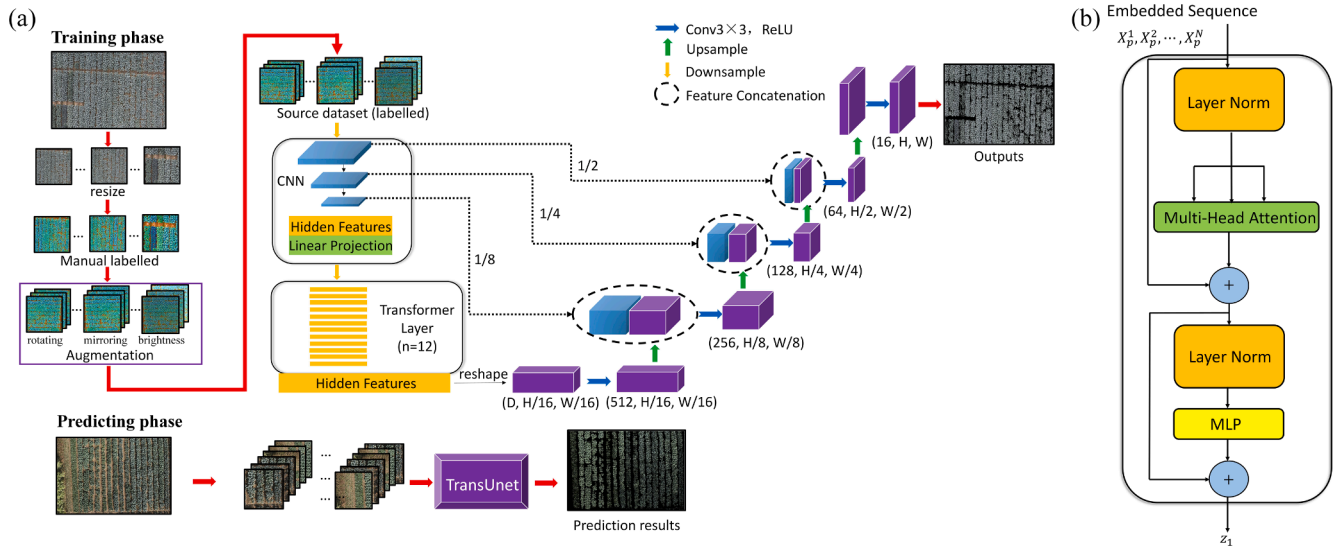


Fig. 3. (a) Overall framework of TransUNet. (b) Architecture of transformer layer.

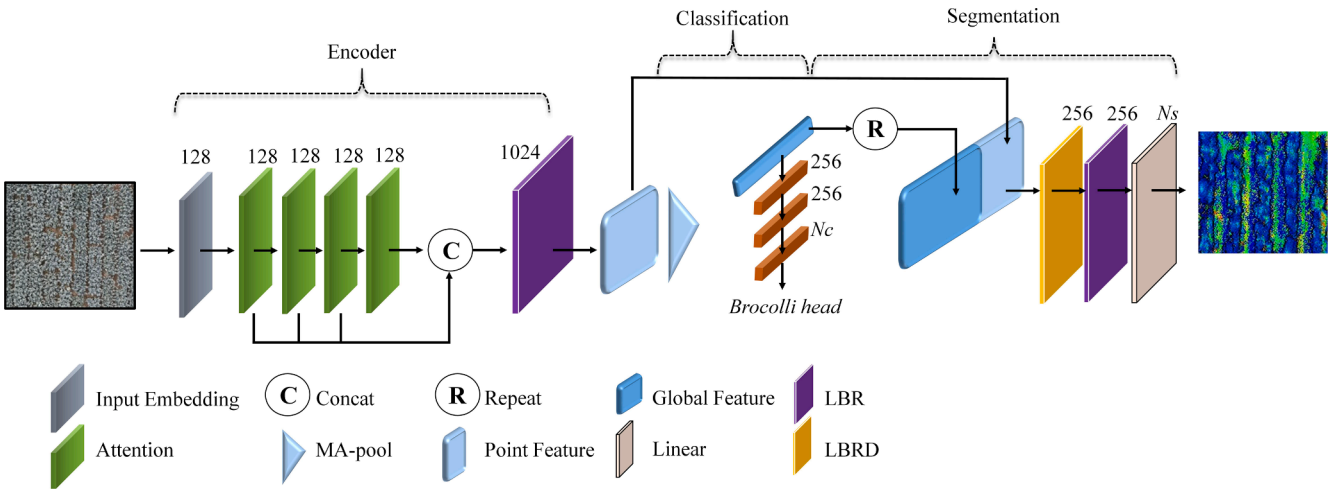


Fig. 4. Flowchart of point cloud transformer to estimate individual broccoli-head volume in this study. MA-pool: Max-Pool + Average-Pool; LBR: Linear + BatchNorm, +ReLU layers; LBRD: LBR + Dropout layer.

F1 score, and Intersection over Union (IoU), which are widely used in previous studies. Precision depicts the ratio of correctly classified pixels to the number of points of the mode. Recall describes the capability of the model to segment the broccoli canopy or head and is inversely related to omission error. The overall accuracy is quantified by the F score, which provides the harmonic mean of recall and precision. Three types of classified results are produced: true positives (TP, for correctly classified), false positives (FP, for erroneously classified), and false negatives (FN, for not classified) to calculate the above metrics. The ratio of the union and the intersection of the labeled region and the predicted polygons from TransUNet and PCT is called the IoU. Mathematically, the metrics are given by.

$$Precison = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - score = 2 \times \frac{Precison \times Recall}{Precison + Recall} \tag{3}$$

$$IoU = \frac{TP}{FP + TP + FN} \tag{4}$$

Next, the accuracy assessment at the plant level was obtained by comparing field measurements with PCT estimates of broccoli-head volume. The accuracy was evaluated by computing the determination coefficient (R^2), the root mean square error (RMSE), and the relative RMSE (rRMSE%). As usual, R^2, the RMSE, and the rRMSE are obtained by using.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - x_{i,ref})^2}{\sum_{i=1}^{N}(x_i - \bar{x}_i)^2} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - x_{i,ref})^2}{N}} \tag{6}$$

$$rRMSE = \frac{RMSE}{\bar{x}_{i,ref}} \tag{7}$$

where N is the number of the broccoli heads, x_i is the broccoli-head volume from the assessed datasets, $x_{i,ref}$ is the broccoli-head volume from the reference dataset, and $\bar{x}_{i,ref}$ is the average volume from the reference dataset.

3. Results

We conducted three groups of comparison experiments. First, the broccoli canopy mapping results were presented to examine the improvement of using the TransUnet architecture. Second, we evaluated the accuracy of point cloud segmentation on different approaches and highlight how the overall segmentation accuracy is improved by PCT. Third, we solely compared volume estimation results of our findings with ground-truth, following a brief analysis of accuracy assessment.

3.1. Accuracy analysis of broccoli canopy mapping

As stated above, three main plots with three broccoli cultivars were used to demonstrate the applicability of the proposed method. To validate the performance of TransUNet, it was compared with two traditional machine-learning methods (i.e., random forest and support vector machines) and three other standard CNN-based networks: U-Net, FCN, and SegNet. The mapping results were evaluated through visual inspection and quantitative assessment. As reported in numerous previous works, the chosen prevalent approaches were successful. Note that we used the same data preprocessing procedures for both machine-learning-based and CNN-based methods to make the comparisons with the same dataset.

As depicted in Fig. 5, TransUNet produced outstanding results for broccoli canopy mapping with an average Precision of 0.917, an average Recall of 0.864, an average F1 score of 0.901, and a mean IoU of 0.895, which far exceed those of U-Net (0.844, 0.832, 0.839, 0.832), FCN (0.868, 0.845, 0.855, 0.834), and SegNet (0.875, 0.862, 0.868, 0.841). Note that random forest performs the worst of all techniques (0.71, 0.691, 0.707, 0.658), followed by support vector machines (0.733, 0.708, 0.713, 0.707). The mapping accuracy of the three DL methods differs strongly from that of the two standard machine-learning-based methods. Although the classical machine learning-based methods can map the canopy, their inferior performance can be ascribed to the inherent weaknesses of the hand-crafted spectral and texture features in representing the broccoli canopy characteristics, which may explain why highly accurate canopy mapping needs manual or semi-automatic labeling. Moreover, with respect to different varieties, TransUNet performs best for Zheqing, with improvements of 7.1 %–9.4 % and 5.2 %–6.8 % with respect to the average accuracy for Lvxiong and Yanxiu, respectively. For the Lvxiong and Yanxiu species, TransUNet performed slightly worse, which may be due to (a) the measurement uncertainty in field, and/or (b) the relatively complex canopy structure of Lvxiong and Yanxiu. We conclude that TransUNet is competitive for broccoli canopy

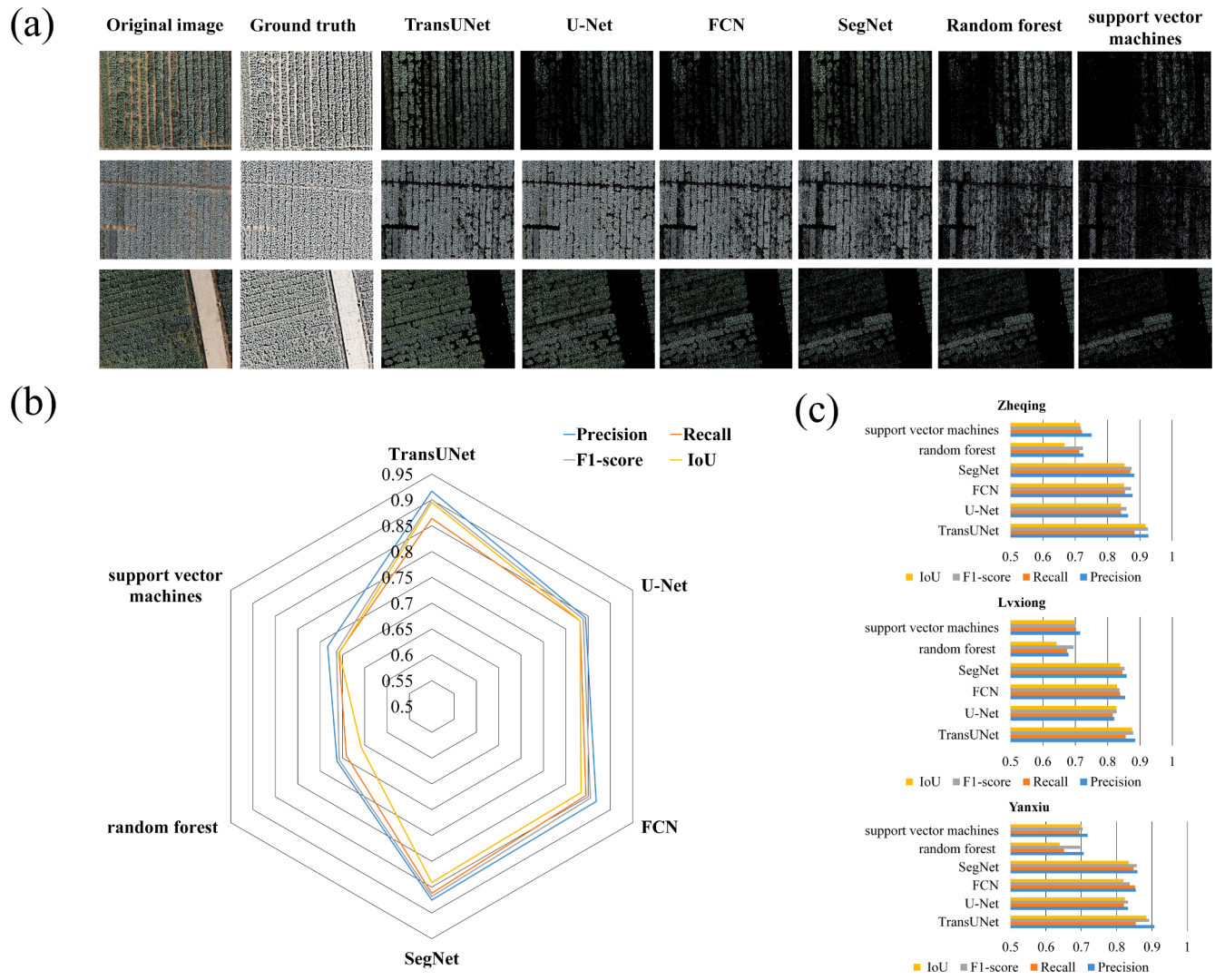


Fig. 5. (a) Example of broccoli canopy mapping results. The row indicates the test sample, and the columns indicate the model. (b) Accuracy of broccoli canopy mapping using different models. The value inside the circle of the radar plot is 50% and the value of the outer boundary is 100%.

mapping, which gives it a strong potential for accurate plant characterization in a real-field environment.

3.2. Accuracy assessment of UAV laser scanning point cloud segmentation algorithm

In the second assessment, we considered the accuracy of point cloud segmentation based on UAV laser scanning data using the process described in Section 2.5. For comparison, we also applied two recent point-cloud-based models, PointNet and PointNet++, and a classical K-means clustering algorithm. Note that we still use Precision, Recall, F1 score, and IoU as evaluation metrics to determine whether a laser point is classified correctly.

Fig. 6(a) shows the original point cloud, the manual labeling, the results of PCT, the results of PointNet, the results of PointNet++, and the results of K-means. Of the four comparison approaches, the metrics show that the PCT performs best (Precision = 0.914, Recall = 0.899, F1 score = 0.901, IoU = 0.879), followed by PointNet (Precision = 0.877, Recall = 0.853, F1 score = 0.864, IoU = 0.863), PointNet++ (Precision = 0.816, Recall = 0.798, F1 score = 0.809, IoU = 0.708), and K-means (Precision = 0.753, Recall = 0.697, F1 score = 0.734, IoU = 0.654).

Furthermore, we analyzed in detail the case of different species by using the quantitative results shown in Fig. 6(b). The two broccoli species Lvxiong and Yanxiu have similar high accuracies except for K-means, which correctly classifies only 75 % of the points. More broccoli-head points were misclassified or incorrectly identified for Zheqing, probably because the topological structure and surface texture of the Zheqing category is similar those of shrubs or weeds. Also, in this case, results for PCT are slightly worse, but the error is not large. Better accuracy was achieved by PCT for three plots with the four metrics ranging from 0.892 to 0.937, 0.876 to 0.907, 0.881 to 0.928, and 0.839 to 0.905, respectively. This result demonstrates that the accuracy of PCT does not change significantly between the different varieties.

3.3. Analysis of accuracy of broccoli-head-volume estimation

To evaluate the accuracy of the predictive model established for broccoli-head volume, the estimation results and the corresponding field survey were selected and compared in terms of R^2, RMSE, and rRMSE. In Table 1, we selected one model for each regression to demonstrate its expression. As shown in Table 1, the PCT-based approach produces the highest overall accuracy with R^2 = 0.875, followed by PointNet++ (0.813) and PointNet (0.738), with K-means producing the poorest performance with R^2 = 0.654. The results also show that the PCT-based approach decreases the rRMSE by 1.65 %, 7.23 %, and 11.98 % with respect to the PointNet, PointNet++, and K-means methods. The decent performance of the PCT-based approach for different genotypes may be attributed to the precise classification of the point cloud. The modest

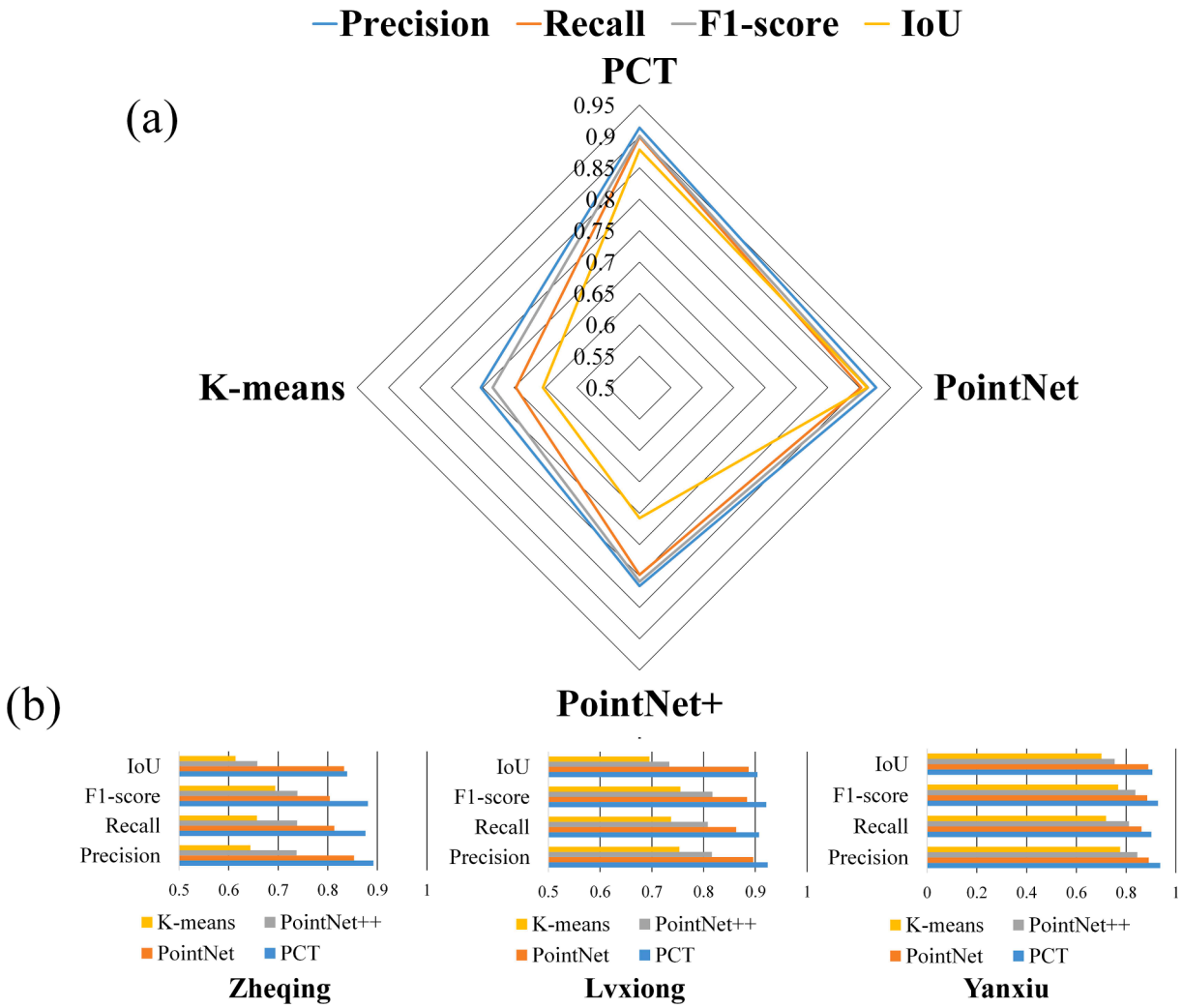


Fig. 6. Accuracy of UAV laser scanning point cloud segmentation using different models. (a) Overall accuracy of laser point segmentation with different methods. (b) Comparison of laser point segmentation accuracy for different varieties.

Table 1
Field reference vs predicted broccoli-head volume.

Methods	R^2	RMSE (cm^3)	rRMSE (%)
PCT	0.875	18.62	3.64
PointNet++	0.813	26.54	5.29
PointNet	0.738	50.75	10.87
K-means	0.654	78.94	15.62

performance obtained by PointNet and PointNet++ demonstrates their reliability and adaptability, even for uncommon applications. In contrast, although the K-means technique could solve the collinearity classification problems, it is unable to map non-linear and complex relationship which result in poor performance.

4. Discussion

The discussion focuses on three themes: (a) the practicality offered by TransUNet and PCT for real-world monitoring, (b) the contribution of these techniques in terms of efficiency and cost, and (c) detailed limitations and perspective.

4.1. The practical effect of canopy mapping based on TransUNet

The first contribution of this study is that it demonstrates a feasible alternative to simultaneously detecting a broccoli canopy in UAV-based RGB imagery. Previous results demonstrated that TransUNet accurately differentiates canopy from non-canopy and captures time-series changes in the canopy at both the plant and plot scales. For further validation purposes, we test the accuracy and robustness of TransUNet to show its potential for practical use.

4.1.1. Sensitivity in relation to environmental conditions

Atmospheric conditions, illumination intensity, and soil reflection change over time. To better understand the robustness of our canopy mapping model when the quality of the input images varies in terms of light intensity and soil reflection, a new *Dataset_test* was used as the test data, which consisted of field images taken under normal and intense conditions (i.e., strong shadowing). Fig. 7(a) shows typical images of the mapping outputs from the TransUNet and the average accuracies rates of this dataset, which reveal that variation of illumination intensity imposed by weather conditions does not significantly affect the mapping. However, a 5 %–10 % drop in accuracy appears when the soil is moist, which may indicate that changes in background affect the extraction of the canopy region, as expected in Section 2.2.1.

4.1.2. Influence of training-set size on accuracy of canopy mapping

The need for extensive training samples is a limitation of CNN models, although increasing the size of training data improves the accuracy. To test whether a large amount of data is crucial for success, we

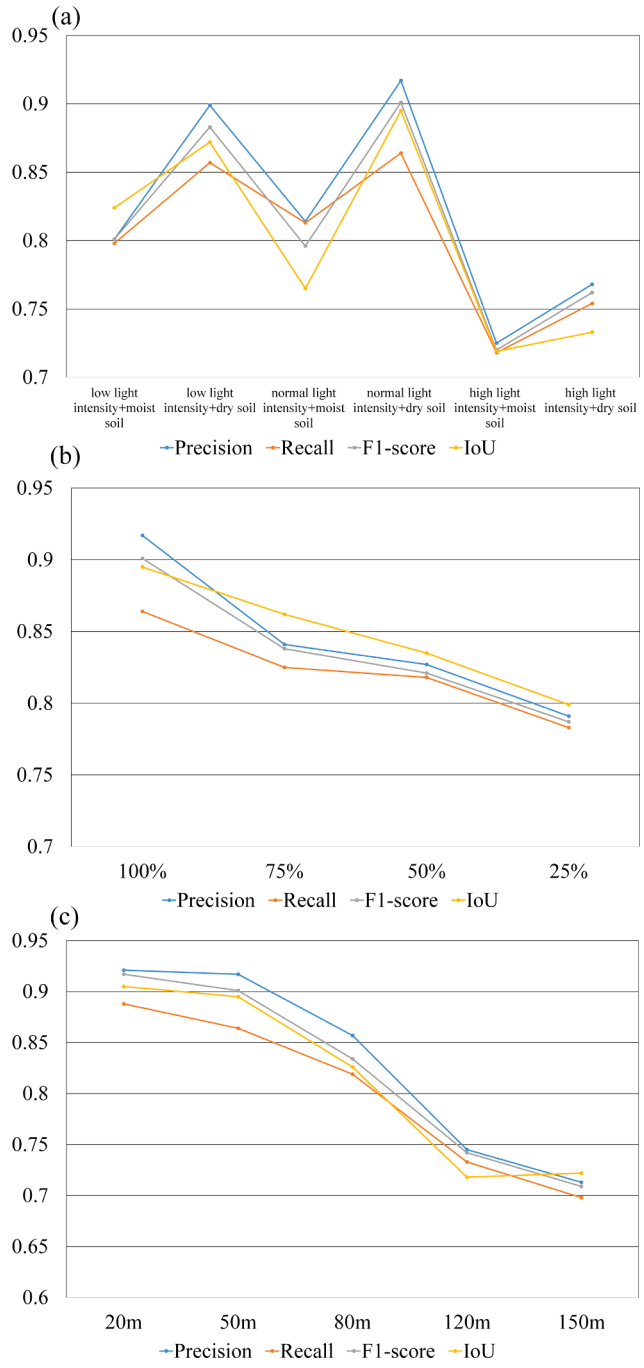


Fig. 7. (a) Accuracies of broccoli canopy mapping results under various environmental conditions. (b) Effect of sample size (randomly selected subsets representing 25%, 50%, 75% of the original training sample size) on overall accuracy. (c) Effect of spatial resolution of the proposed method on overall accuracy.

analyze how different training sizes affect the results by selecting at random different numbers of images (25 %, 50 %, and 75 %) from the original dataset as training samples. Fig. 7(b) shows the mapping accuracy of TransUNet for training sets of different sizes. The curves demonstrate that, as the number of the training samples is reduced to the previous quarter, the performance of the method decreases slightly. In particular, for the minimum training samples, the overall accuracy of the broccoli canopy mapping remains sufficiently precise for field monitoring, which indicates that TransUNet works well with a small number of samples. The experimental results show that the increased number of training images improves the accuracy by 8 %–15 % but has a limited impact on the results of TransUNet. Despite using only a quarter of the images to train the model in the extreme case, the proposed model still identifies most of the canopy pixels.

4.1.3. *Impact of spatial resolution on accuracy of canopy mapping*

Due to their typical low flight altitude, UAVs can play a key role in vegetation monitoring, given that airborne or satellite monitoring cannot provide imagery with a comparable ultrafine spatial resolution. A qualitative study was conducted to analyze how the spatial resolution affects the performance of the method [Fig. 7(c)]. The proposed method was verified over the three plots with training images acquired at 20–150 m altitude, corresponding respectively to 0.002–0.01 m spatial resolution. Analyzing the precision in mapping canopy under different spatial resolution shows that the best evaluation metrics are obtained for 20 m image-acquisition height, with Precision = 0.921, Recall = 0.888, F1 score = 0.917, and IoU = 0.905. Results show that the mapping accuracy of the model decreases linearly with spatial resolution, which is consistent with the results of Schiefer et al. (2020). Very poor estimates are obtained for 150 m image-acquisition height, with Precision = 0.713, Recall = 0.698, F1 score = 0.709, and IoU = 0.722. A likely reason for the decreasing accuracy with decreasing spatial resolution is the decreased abundance of spatial information. Note also that high resolution does not significantly translate into accurate CNN models, likely because the optimal spatial resolution is affected by numerous factors, including spatial distribution and target size. This indicates that a knowledge gap exists for determining the optimal resolution to balance the accuracy and efficiency.

4.2. *Generality and applicability of PCT for calculating broccoli-head volume*

Lidar, which simultaneously provides horizontal and vertical information, is promising for monitoring vegetation. However, the current performance of point segmentation using lidar data is insufficient for field-monitoring applications. An investigation of the factors affecting the estimation of broccoli-head volume shows that the accuracy is affected not only by the model chosen but also by the data quality (Lin and Habib, 2021). This section thus compares the performance of PCT with various data combinations.

4.2.1. *Effect of point density on estimation of broccoli-head volume*

We first characterized how LiDAR point density affects the accuracy of estimations of broccoli-head volume. As shown in Table 2, the overall accuracy generally increases with increasing point density of LiDAR data. Overall accuracy was about 0.75 at a point density of 500 points/m^2 and increased to about 0.9 with a point density above 1500 points/m^2. Note also that R^2, RMSE, and rRMSE are much lower at minimum density than at the highest point density. This observation is consistent with the work of Li et al. (2013) and is likely explained by the fact that lower LiDAR point densities may not adequately characterize structure in similar vegetations.

4.2.2. *Effect of training data size on volume estimation*

To test the sensitivity of the proposed PCT method to training-sample size, the approach was run with randomly selected training subsets representing 25 %, 50 %, 75 %, and 90 % of the original training samples. Table 3 shows that the success rate generally increases with increasing training data. Using 25 % and 100 % of the training data produces the lowest and highest accuracy, respectively. In particular, the PCT is almost unaffected by sample size, showing a small reduction in accuracy: about 0.85, 0.87, 0.88, and 0.9 when using 25 %, 50 %, 75 %, and 90 % of the original sample size, respectively. For all test samples, R^2 decreases slightly whereas RMSE and rRMSE remain almost unchanged. This result indicates that the proposed method has significant advantages for dealing with training data of various sizes.

Table 2
Volume-estimation accuracies measured as a function of point density.

Point density (N/m^2)	Precision	Recall	F1-score	IoU	R^2	RMSE (cm^3)	rRMSE (%)	Average running time (h)
500	0.752	0.699	0.734	0.757	0.758	43.63	9.67	7.54
700	0.794	0.785	0.793	0.801	0.779	45.55	9.33	8.15
1000	0.802	0.797	0.80	0.823	0.784	21.28	4.27	8.34
1200	0.859	0.846	0.851	0.862	0.823	20.01	3.95	8.56
1500	0.914	0.899	0.901	0.879	0.875	18.62	3.64	9.21

Table 3
Accuracy of estimation of broccoli-head volume as a function of the size of the training data.

Training data size	Precision	Recall	F1-score	IoU	R^2	RMSE (cm^3)	rRMSE (%)
25 %	0.857	0.831	0.845	0.837	0.835	23.69	4.35
50 %	0.871	0.864	0.869	0.841	0.841	22.07	4.23
75 %	0.882	0.862	0.873	0.858	0.853	20.56	4.01
90 %	0.901	0.874	0.885	0.867	0.866	19.24	3.75
100 %	0.914	0.899	0.901	0.879	0.875	18.62	3.64

4.2.3. Impact of clumping on broccoli-head-volume estimation

Clumping is a severe problem for vegetation delineation and one of the main causes preventing such studies from moving to the individual-plant level. To qualitatively analyze the effect of the plant density, we test the average results of our integrated volume estimation method for three groups of clumping: isolated (>90 % visible), boundary (50 %–90 % visible), and clumped (<50 % visible). The results show that the appropriate plant density is essential for the accurate estimation of the volume of broccoli heads using LiDAR data. As shown in Table 4, the isolated group produces the highest prediction accuracy, followed by the boundary group and then the clumped group. Collectively, these results also highlight that the PCT-based method is more adaptive than the other methods. The decent performance with the isolated group may be attributed to the clear boundaries of this group, whereas the clumped broccoli heads often strongly overlap.

4.3. Efficiency

Efficiency is an important consideration for processing UAV data, so we list the time complexity for the different methods for canopy mapping and for estimating the volume of broccoli heads. As shown in Table 5, all DL models required a longer time to learn features compared with traditional machine-learning methods, which is because they require significantly more time for scale selection and for calculating weights or parameters. We then compared different DL models, which showed that our TransUet and PCT produces the highest accuracy with time complexity improved by about 17.88 % and 26.86 %, respectively. Upon increasing the total number of labeled samples, the computational cost of TransUet and PCT do not increase much and tend to translate into immediate quality gains.

4.4. Limitations of the current study and potential future work

At present, more and more studies are focusing on agricultural remote sensing via UAVs and DL algorithms. Combined with RGB/LiDAR devices, the workflow provided herein has the potential to characterize vegetation and crops on an ultrafine spatial scale. Nevertheless, some limitations remain. First, like other DL methods, the proposed method relies on large and representative training samples, which require substantial manual effort to build. Other than the traditional augmentation methods, artificial intelligence techniques (e.g., generative adversarial networks) are recommended for future attempts to generate reliable samples from limited original datasets. Second, the proposed method uses some technical strategies to enhance robustness and accuracy, requiring additional computing resources. The running cycle may thus be longer, especially for larger study areas. In the future, parallel processing could be integrated to improve efficiency. Third, the broad applicability of the model for monitoring broccoli has not been assessed.

However, despite the encouraging results, further research is required. For example, to adapt to the various conditions, similar plants in different growth stages and more high-quality data should be acquired for training. More species from different types of broccolis should also be considered in future studies. Another improvement involves compressing the model size and improving the speed while reducing the overall complexity and computational cost. Additional developments from our recent work on unsupervised labeling and training should also

Table 5
Analysis of efficiency of different approaches. The bold font represents the running time and the weight size of the proposed methods.

Canopy mapping			Volume estimation		
Methods	Average running time	Weight Size	Methods	Average running time	Weight Size
TransUNet	**5.34 h**	**135.8 MB**	**PCT**	**9.21 h**	**91.7 MB**
U-Net	8.86 h	273.5 MB	PointNet++	12.78 h	163.5 MB
FCN	7.94 h	543.8 MB	PointNet	15.93 h	144.1 MB
SegNet	6.78 h	364.2 MB	K-means	7.26 h	87.5 MB
Random forest	4.53 h	164.1 MB	/	/	/
support vector machine	5.26 h	122.3 MB	/	/	/

Table 4
Effect of clumping on estimation of broccoli-head volume.

Clumpiness group	Precision	Recall	F1-score	IoU	R^2	RMSE (cm^3)	rRMSE (%)
Isolated	0.914	0.899	0.901	0.879	0.875	18.62	3.64
Boundary	0.853	0.836	0.849	0.819	0.801	36.24	7.05
Clumped	0.786	0.771	0.753	0.705	0.752	46.21	9.53

prove helpful to improve the efficiency of the proposed method. Finally, the field-survey data may also lack entries or have incorrect entries because the reference data are not without challenges. Human interpretation may scale better, but in some cases, it fails to pick up the characteristic parameters of plants due to the complexity of the scene. This is especially true when plant density is relatively high (e.g., due to serious overlapping). Thus, techniques of automated or semi-automated field observations should be explored.

5. Conclusions

Although accurate monitoring of broccoli canopy and heads is essential for understanding the growth status and managing the harvest, it remains challenging. The objective of this study was to develop an automated method for mapping and characterizing broccoli. The proposed method was derived from RGB imagery and LiDAR point cloud data acquired by using a UAV-based remote-sensing system. The workflow presented constitutes a feasible alternative to field surveys and should advance the state of precision agriculture. For canopy mapping, the proposed TransUnet provides highly accurate results, achieving a mean Precision = 0.917, Recall = 0.864, F1 score = 0.901, and IoU = 0.895, which are 4 %–30 % better than results from other methods evaluated with the same data. For estimating the volume of broccoli heads, the proposed PCT produces satisfactory results, with Precision = 0.914, Recall = 0.899, F1 score = 0.901, and IoU = 0.879. Comparative experiments show that PCT produces the more accurate results, with R^2 increased by 0.062–0.221, RMSE decreased by 7.92–60.32, and rRMSE decreased by 1.65 %–11.98 % with respect to other state-of-art approaches. Moreover, we demonstrate the robustness of the proposed method by performing the same task with an adjusted dataset. The results clearly reveal the advantages of the proposed method for monitoring wild vegetation.

CRediT authorship contribution statement

Chengquan Zhou: Data curation, Writing – original draft. **Hongbao Ye:** Visualization, Investigation. **Dawei Sun:** Software. **Jibo Yue:** Validation. **Guijun Yang:** Writing – review & editing. **Jun Hu:** Supervision.

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.

Acknowledgments

This study was partially funded by the National Natural Science Foundation of China (Grant No. 32000283, No. 31901662 and No. 31901722), and partially supported by the Beijing Natural Science Foundation (No. 6182011). The corresponding author thanks the DJI Technology Co., Ltd for providing guidance on UAV control technology.

References

Alonzo, M., Dial, R.J., Schulz, B.K., Andersen, H., Lewis-Clark, E., Cook, B.D., Morton, D. C., 2020. Mapping tall shrub biomass in alaska at landscape scale using structure-from-motion photogrammetry and lidar. Remote Sen. Environ. 245. 10.1016/j.rse.2020.111841.

Blok, P.M., Henten, E., Evert, F., Kootstra, G., 2021. Image-based size estimation of broccoli heads under varying degrees of occlusion. Biosyst. Eng. 208, 213–233. 10.1016/j.biosystemseng.2021.06.001.

Brede, B., Calders, K., Lau, A., Raumonen, P., Bartholomeus, H.M., Herold, M., Kooistra, L., 2019. Non-destructive tree volume estimation through quantitative structure modelling: Comparing UAV laser scanning with terrestrial LIDAR. Remote Sens. Environ. 233. 10.1016/j.rse.2019.111355.

Chen, J., Lu, Y., Yu, Q., Luo, X., Zhou, Y., 2021. Transunet: transformers make strong encoders for medical image segmentation. 10.48550/arXiv.2102.04306.

Colaço, A., Trevisan, R., Molin, J., Rosell-Polo, J., Escolà, A., 2017. Orange tree canopy volume estimation by manual and LiDAR-based methods. Adv. Anim. Biosci. 8, 477–480. 10.1017/S2040470017001133.

Deng, L., Mao, Z., Li, X., Hu, Z., Duan, F., Yan, Y., 2018. UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. ISPRS J. Photogramm. Remote Sens. 146, 124–136. https://doi.org/10.1016/j.isprsjprs.2018.09.008.

Duncanson, L.I., Dubayah, R.O., Cook, B.D., Rosette, J., Parker, G., 2015. The importance of spatial detail: Assessing the utility of individual crown information and scaling approaches for lidar-based biomass density estimation. Remote Sens. Environ. 168, 102–112. 10.1016/j.rse.2015.06.021.

Fagua, J.C., Jantz, P., Rodriguez-Buritica, S., Duncanson, L., Goetz, S.J., 2019. Integrating LiDAR, multispectral and SAR data to estimate and map canopy height in tropical forests. Remote Sens. 11, 2697–2716. 10.3390/rs11222697.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87. https://doi.org/10.1016/j.rse.2016.08. 013.

Franklin, S.E., Ahmed, O.S., 2018. Deciduous tree species classification using objectbased analysis and machine learning with unmanned aerial vehicle multispectral data. Int. J. Remote Sens. 39, 5236–5245. https://doi.org/10.1080/01431161.2017.1363442.

Hadas, E., Jozkow, G., Walicka, A., Borkowski, A., 2019. Apple orchard inventory with a LiDAR equipped unmanned aerial system. Int. J. Appl. Earth Obs. Geoinf. 82, 101911. 10.1016/j.jag.2019.101911.

Hassanein, M., Khedr, M., El-Sheimy, N., 2019. Crop row detection procedure using low-cost UAV imagery system. ISPRS Archives. 42 (2/W13), 349–356. https://doi.org/10.5194/isprs-archives-XLII-2-W13-349-2019.

Hoeser, T., Kuenzer, C., 2020. Object detection and image segmentation with deep learning on earth observation data: a review-Part I: Evolution and recent trends. Remote Sens. 12, 1667. https://doi.org/10.3390/rsl2101667.

Jin, S., Su, Y., Song, S., Xu, K., Hu, T., Yang, Q., Wu, F., Xu, G., Ma, Q., Guan, H., Pang, S., Li, Y., Guo, Q., 2020. Non-destructive estimation of field maize biomass using terrestrial lidar: An evaluation from plot level to individual leaf level. Plant methods. 16, 69–87. 10.1186/s13007-020-00613-5.

Jin, S., Sun, X., Wu, F., Su, Y., Guo, Q., 2021. Lidar sheds new light on plant phenomics for plant breeding and management: recent advances and future prospects. ISPRS J. Photogramm. Remote Sens. 171, 202–223. 10.1016/j.isprsjprs.2020.11.006.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (cnn) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. 10.1016/j.isprsjprs.2020.12.010.

Kerkech, M., Hafiane, A., Canals, R., 2020. Vine disease detection in uav multispectral images using optimized image registration and deep learning segmentation approach. Comput. Electron. Ag. 174. 10.1016/j.compag.2020.105446.

Kusumam, K., Krajník, T., Pearson, S., Duckett, T., Cielniak, G., 2017. 3d-vision based detection, localization, and sizing of broccoli heads in the field. J. Field Robot. 34, 1505–1518. 10.1002/rob.21726.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B., 2018. PointCNN: Convolution on X-transformed points. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, 31, 820–830.

Li, J., Hu, B., Noland, T.L., 2013. Classification of tree species based on structural features derived from high density lidar data. Agr. Forest Meteorol. 171–172, 104–114. 10.1016/j.agrformet.2012.11.012.

Li, Z., Mei, Y., Liu, Y., Fang, Z., Yang, L., Zhuang, M., Zhang, Y.Y., Lv, H.H., 2019. The evolution of genetic diversity of broccoli cultivars in china since 1980. Sci. Hortic. 250, 69–80. 10.1016/j.scienta.2019.02.034.

Lin, Y.C., Habib, A., 2021. Quality control and crop characterization framework for multi-temporal UAV LiDAR data over mechanized agricultural fields. Remote Sens. Environ. 256 https://doi.org/10.1016/j.rse.2021.112299.

Lobo Torres, D., Feitosa, R.Q., Nigri Happ, P., Elena Cúe La Rosa, L., Marcato Junior, J., Martins, J., Olã Bressan, P., Gonçalves, W.N., Liesenberg, V., 2020. Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on High Resolution UAV Optical Imagery. Sensors 20, 563. https://doi.org/10.3390/s20020563.

López-Jiménez, E., Vasquez-Gomez, J.I., Sanchez-Acevedo, M.A., Herrera-Lozada, J.C., Uriarte-Arcia, A.V., 2019. Columnar cactus recognition in aerial images using a deep learning approach. Ecol. Inform. 52, 131–138. https://doi.org/10.1016/j.ecoinf.2019.05.005.

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y., 2018. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sens. 10 https://doi.org/10.3390/rs10071119.

Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K.T., Maw, M.J., Shakoor, N., Mockler, T., Fritschi, F.B., 2019. Vegetation Index Weighted Canopy Volume Model (CVMVI) for soybean biomass estimation from Unmanned Aerial System-based RGB imagery. ISPRS J. Photogramm. Remote Sens. 151, 27–41. 10.1016/j.isprsjprs.2019.03.003.

Martins, V.S., Kaleita, A.L., Gelder, B.K., 2021. Digital mapping of structural conservation practices in the Midwest U.S. croplands: implementation and preliminary analysis. Sci. Total Environ. 772, 145191. https://doi.org/10.1016/j.scitotenv.2021.145191.

Moorthy, S., Calders, K., Vicari, M.B., Verbeeck, H., 2019. Improved supervised learning-based approach for leaf and wood classification from lidar point clouds of forests. IEEE Trans. Geosci. Remote Sens. 58, 3057–3070. 10.1109/TGRS.2019.2947198.

Natesan, S., Armenakis, C., Vepakomma, U., 2019. Resnet-based tree species classification using uav images. Int. Arch. Photogramm., Remote Sens. Spatial Inf.

Sci.- ISPRS Arch. 42 (2/WlS), 475–481. https://doi.org/10.5194/isprs-archives-XLII-2-W13-475-2019.

Nezami, S., Khoramshahi, E., Nevalainen, O., Pol¨onen,¨I., Honkavaara, E., 2020. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. Remote Sens. 12, 1–19. https://doi.org/10.20944/preprints202002.0334.v1.

Osco, L.P., de Arruda, M.d.S., Marcato Junior, J., da Silva, N.B., Ramos, A.P.M., Moryia, E.A.S., Imai, N.N., Pereira, D.R., Creste, J.E., Matsubara, E.T., Li, J., Gongalves, W. N., 2020. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. ISPRS J. Photogramm. Remote Sens. 160, 97–106. https://doi.org/10. 1016/j.isprsjprs.2019.12.010.

Pandey, A., Jain, K., 2021. An intelligent system for crop identification and classification from UAV images using conjugated dense convolutional neural network Comput. Electr. Agric, 106-543, https:// doi.org/ 10.1016/j.compag.2021.106543.

Qi, C., Yi, L., Su, H., Guibas, L., 2017. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Adv. Neural Informat. Process. Syst. 5100–5109. https://arxiv.org/pdf/1706.02413.pdf.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven earth system science. Nature. 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Rey, N., Volpi, M., Joost, S., Tuia, D., 2017. Detecting animals in African savanna with UAVs and the crowds. Remote Sens. Environ. 200, 341–351. https://doi.org/10.1016/j.rse.2017.08.026.

Santos, L.M., Ferraz, G., Barbosa, B., Diotto, A.V., Xavier, L.A., 2020. Biophysical parameters of coffee crop estimated by UAV RGB images. Precis. Agric. 21, 1227–1241. 10.1007/s11119-020-09716-4.

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., Schmidtlein, S., 2020. Mapping forest tree species in high resolution uav-based rgb-imagery by means of convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 170.205-215. 10.1016/j.isprsjprs.2020.10.015.

Shirzadifar, A., Bajwa, S., Nowatzki, J., Bazrafkan, A., 2020. Field identification of weed species and glyphosate-resistant weeds using high resolution imagery in early growing season. Biosyst. Eng. 200, 200–214. 10.1016/j.biosystemseng.2020.10.001.

Tao, S., Wu, F., Guo, Q., Wang, Y., Li, W., Xue, B., Hu, X., Li, P., Tian, D., Li, C., Yao, H., Li, Y., Xu, G., Fang, J., 2015. Segmenting tree crowns from terrestrial and mobile LiDAR data by exploring ecological theories. ISPRS J. Photogramm. Remote Sens. 110, 66–76. 10.1016/j.isprsjprs.2015.10.007.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. 38, 1–12. 10.1145/3326362.

Wiering, N.P., Ehlke, N.J., Sheaffer, C.C., 2019. Lidar and RGB image analysis to predict hairy vetch biomass in breeding nurseries. Plant Phenome J. 2, 1–8. 10.2135/tppj2019.02.0003.

Wu, X., Shen, X., Cao, L., Wang, G., Cao, F., 2019. Assessment of individual tree detection and canopy cover estimation using unmanned aerial vehicle based light detection and ranging (UAV-LiDAR) data in planted forests. Remote Sens. 11, 908–928. 10.3390/rs11080908.

Yan, G.J., Li, L., Coy, A., Mu, X.H., Chen, S.B., Xie, D.H., Zhang, W.H., Shen, Q.F., Zhou, H.M., 2019. Improving the estimation of fractional vegetation cover from UAV RGB imagery by colour unmixing. ISPRS J. Photogramm. Remote Sens. 158, 23–34. 10.1016/j.isprsjprs.2019.09.017.

Yuan, W., Wijewardane, N.K., Jenkins, S., Bai, G., Ge, Y., Graef, G.L., 2019. Early prediction of soybean traits through color and texture features of canopy RGB imagery. Sci. Rep. 9, 14089. https://doi.org/10.1038/s41598-019-50480-x.

Zhang, J., Liu, C., Wang, B., Chen, C., He, J., Zhou, Y., 2022. An infrared pedestrian detection method based on segmentation and domain adaptation learning. Comput. Electr. Eng. 99, 107781-. 10.1016/j.compeleceng.2022.107781.

Zhang, L.P., Zhang, L.F., Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci. Remote Sens. M. 4, 22–40. 10.1109/MGRS.2016.2540798.