

RESEARCH

Open Access



# Mapping of cotton bolls and branches with high-granularity through point cloud segmentation

Lizhi Jiang<sup>1,2</sup>, Javier Rodriguez-Sanchez<sup>2,3</sup>, John L. Snider<sup>4</sup>, Peng W. Chee<sup>4</sup>, Longsheng Fu<sup>1\*</sup> and Changying Li<sup>2\*</sup>

## Abstract

High resolution three-dimensional (3D) point clouds enable the mapping of cotton boll spatial distribution, aiding breeders in better understanding the correlation between boll positions on branches and overall yield and fiber quality. This study developed a segmentation workflow for point clouds of 18 cotton genotypes to map the spatial distribution of bolls on the plants. The data processing workflow includes two independent approaches to map the vertical and horizontal distribution of cotton bolls. The vertical distribution was mapped by segmenting bolls using PointNet++ and identifying individual instances through Euclidean clustering. For horizontal distribution, TreeQSM segmented the plant into the main stem and individual branches. PointNet++ and Euclidean clustering were then used to achieve cotton boll instance segmentation. The horizontal distribution was determined by calculating the Euclidean distance of each cotton boll relative to the main stem. Additionally, branch types were classified using point cloud meshing completion and the Dijkstra shortest path algorithm. The results highlight that the accuracy and mean intersection over union (mIoU) of the 2-class segmentation based on PointNet++ reached 0.954 and 0.896 on the whole plant dataset, and 0.968 and 0.897 on the branch dataset, respectively. The coefficient of determination ( $R^2$ ) for the boll counting was 0.99 with a root mean squared error (RMSE) of 5.4. For the first time, this study accomplished high-granularity spatial mapping of cotton bolls and branches, but directly predicting fiber quality from 3D point clouds remains a challenge. This method provides a promising tool for 3D cotton plant mapping of different genotypes, which potentially could accelerate plant physiological studies and breeding programs.

**Keywords** Plant phenotyping, PointNet++, TreeQSM, Cotton boll distribution mapping

\*Correspondence:

Longsheng Fu  
fulsh@nwafu.edu.cn  
Changying Li  
cli2@ufl.edu

<sup>1</sup>College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>2</sup>Bio-Sensing, Automation, and Intelligence Laboratory, Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup>Department of Crop and Soil Sciences, University of Georgia, Tifton, GA 31793, USA

<sup>4</sup>Department of Crop and Soil Science, Institute of Plant Breeding, Genetics, and Genomics, University of Georgia-Tifton Campus, University of Georgia, Tifton, GA 31793, USA



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Cotton (*Gossypium hirsutum* L.) is the most widely grown natural fiber crop in the world [1]. Cotton fiber yield is the product of boll density (boll number per plant  $\times$  plant density), boll mass, and lint percent [2, 3]. The spatial distribution of bolls is closely related to yield and quality [4, 5]. It influences the growth conditions of the bolls, which directly impact key quality attributes such as fiber length, strength, uniformity, and maturity [6, 7]. Plant architecture, including the spatial distribution of fruits, has a significant impact on yield and fiber quality [8, 9]. For instance, monopodial branches do not appear to influence yield, whereas variations in the number of sympodial branches can result in notable yield differences between cultivars. Among fiber traits, Micronaire is particularly sensitive to fruiting nodes and their adaxial positions on a branch. Bolls located on upper internodes and farther from the main stem tend to exhibit lower Micronaire values. Additionally, bolls on branches situated toward the mid-section of the plant (typically nodes 9 through 14) generally have the highest fiber length and strength compared to those higher in the canopy (internodes 16 through 22). Although the total number of bolls per unit land area is an important driver of yield, boll weight and fiber quality can vary significantly at different main stem branching nodes and positions along a fruiting branch [10]. Thus, the contribution of each location on the plant to final lint yield and economic returns can vary as well. While the information obtained from phenotyping boll distribution is invaluable for breeders and physiologists alike, manual assessments of fruit distribution are impractical in breeding programs. Thus, the development of high-throughput methods to phenotype cotton boll distribution would be a major advancement.

Three-dimensional (3D) imaging technology preserves the structural characteristics of plants and allows for richer and more accurate information on plant architectural phenotypes. Light Detection and Ranging (LiDAR) is one of the technologies used for acquiring 3D plant models and can generate point clouds containing millions, or even tens of millions, of data points. These point clouds represent the structural features of the plant surface, including 3D coordinates ( $x$ ,  $y$ ,  $z$ ) and other information such as color and intensity. As it works with its own light source, LiDAR sensors greatly reduce the influence of environmental illumination and improve the quality of the collected data. Compared to two-dimensional (2D) images that are easily affected by lighting and lack depth information causing occlusion, 3D point clouds overcome this issue to a large extent and have been intensively investigated in plant phenotyping [11]. Point clouds have been applied to extract wheat canopy height ( $R^2=0.99$ , RMSE=0.017 m) and canopy cover ( $R^2=0.92$ ) accurately [12]. A 2D LiDAR mounted on a moving vehicle was

used to measure the maximum canopy height, projected canopy area, and plant volume of cotton plants, with  $R^2$  values of 0.97, 0.97, and 0.98, respectively [13]. Plant organ-level phenotyping using point clouds has also received increasing attention. Not only the angle [14], area [15], and length [16] of leaves, but also the number [17–20] and volume [21] of fruits have been studied. In particular, some studies have investigated machine learning methods to segment cotton main stem and individual leaves, enabling the estimation of main stem length, as well as the number, length, width, and area of the leaves [22–24]. However, the success of these phenotypic trait extractions depends on the point cloud instance segmentation, while the accuracy of segmentation is currently a bottleneck.

Traditional machine learning methods have been employed for point cloud segmentation to extract plant phenotypic traits. For example, geometric features were directly built within the point cloud to segment trellises, support poles, and tree trunks, enabling the successful generation of canopy density and depth maps [25]. A traditional machine learning algorithm support vector machine (SVM) was used to classify pomegranate tree point clouds into fruits, branches, and leaves [17]. An accuracy of 0.75 for fruit and 0.99 for non-fruit was achieved, reaching a fruit counting accuracy of 78%. Alternatively, TreeQSM is another non-deep learning instance segmentation method [26]. It is a 3D modeling approach specifically designed to create geometric and structural representations of trees from point clouds. By leveraging the tree's geometric features, this method effectively segments the entire point cloud into the main stem and branches. TreeQSM has already been used in phenotyping of apple trees and maize [27, 28]. Cotton plants also exhibit a tree-like structure, making this method applicable to the segmentation of cotton plants, and its segmentation performance is worth further exploration.

Deep learning techniques can automatically learn features from the data, which has a significant advantage over the hand-crafted features used in traditional machine learning methods. For example, one 3D deep learning approach PointNet [29] has been employed in point cloud segmentation tasks to achieve automatic segmentation of maize plant organs [30]. This approach enabled the extraction of various phenotypic traits, including stem height, leaf length, and leaf inclination. PointCNN [31] has also been applied to segment maize stems and leaves for extracting 3D phenotypic traits of plants from field-based terrestrial LiDAR data [32]. To obtain precise phenotypic traits of plant organs, it is essential to first accomplish instance segmentation of individual organs. Recently, several customized point cloud segmentation models have emerged. For

example, PlantNet was designed and used for semantic and instance segmentation of tobacco, tomato, and sorghum from point clouds [33]. An end-to-end model plant segmentation transformer (PST) was proposed to segment rapeseed at the pod stage [34], which proved that the point cloud segmentation method based on deep learning has great potential in analyzing dense plant point clouds with complex morphological features. A study explores the spatial distribution of apple fruit using deep learning techniques. Specifically, the spatial distribution of apples was detected and located using a methodology that combined Mask R-CNN for 2D fruit detection, structure-from-motion (SfM) for generating 3D point clouds, and a support vector machine to remove false positives [35]. The canopy volume, flowers, fruit, and yield maps of almond trees can be effectively used to map the distribution of flowers and fruits, as well as estimate and predict the yield of individual trees [36]. The development of deep learning-based point cloud segmentation techniques has improved the extraction accuracy of plant phenotypic traits and has provided data support for breeders to develop new crop varieties with high yields and high quality [37, 38].

In the field of high throughput phenotyping for cotton, there have been some studies to obtain phenotypic traits from point clouds. Some researchers import point clouds into CloudCompare and identify cotton bolls by color information, but color features alone may not always give a complete representation of the cotton bolls [39]. Some of our past work has developed a data processing pipeline for 3D boll mapping within a plot [40]. Additionally, the pipeline also provided estimations for the number, volume, and location of the bolls. Both of the above studies identified and localized cotton bolls, but the primary difference between them in obtaining cotton boll point clouds is that the first study only used color information to identify the point clouds, while the second study employed a machine learning classification algorithm, Support Vector Machines (SVM), utilizing both color and shape features to detect cotton bolls. In another study, individual cotton plant phenotypes were measured with  $R^2$  values and absolute percentage errors of 0.94 and 4.3% for main stem length and 0.7 and 5.1% for node counting, respectively [41]. However, the design of this method assumed that the highest point of the plant was on the main stem. It becomes challenging to identify the main stem when the highest point appears on the branches, thus limiting the applicability of this method. More recently, we investigated another deep learning model, PVCNN [42], to segment a single cotton plant into bolls, the main stem, and branches [43]. Seven architectural traits were extracted from the segmented plant parts, with  $R^2$  values all greater than 0.8, and the average absolute percentage errors less than 10%. This method

primarily focuses on the measurement of architectural traits of cotton plants and does not delve deeply into the spatial distribution of cotton bolls, especially their distribution on individual branches. In cotton breeding programs, it would be beneficial to measure fine-grained phenotypic traits such as the number and distribution of bolls not only on a plant but also on single branches. Based on the research, breeders can understand the relationship between cotton boll distribution and environmental stress (such as drought, pests and diseases), and select genotypes that can maintain good cotton boll distribution under adverse conditions.

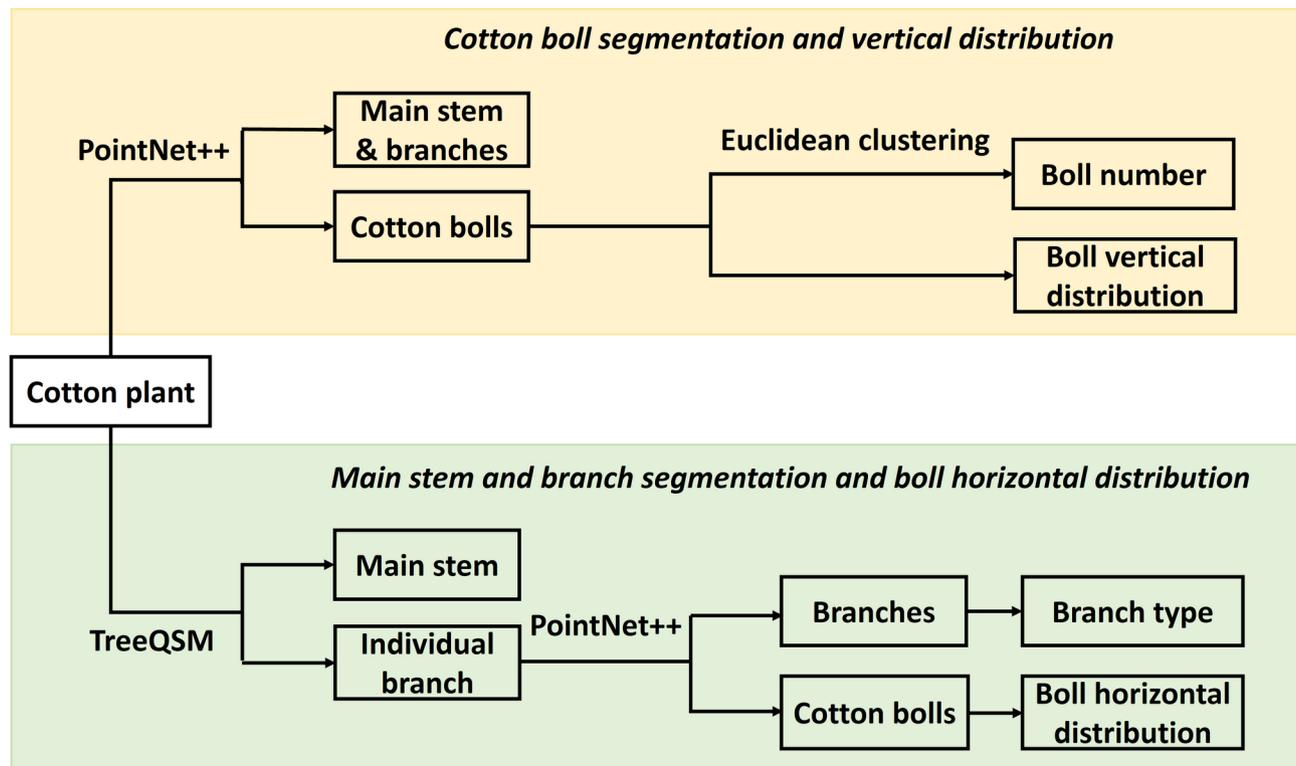
To fill the gap, the overall goal of this study was to develop a 3D point cloud data processing pipeline that enables the spatial mapping of cotton boll distribution both at the overall plant level and individual branch level for different genotypes. Specific objectives of this study were to: (1) develop a semantic segmentation method with PointNet++ to segment cotton bolls from individual plants and branches, (2) investigate TreeQSM-based instance segmentation method to segment cotton plants into main stems and individual branches, (3) determine branch types by combining point cloud meshing completion and the Dijkstra algorithm, and (4) map the spatial distribution of cotton bolls, including the distribution of bolls along the height of the plant and their position on each branch relative to the main stem.

## Methods

The overall workflow of data processing illustrates that there are two independent approaches (as two branches) to achieve the mapping of bolls along the vertical direction for a whole plant and a fine-grained mapping of bolls for each branch (Fig. 1). Breeders can select different branches based on their needs. For the first branch (upper part of Fig. 1), it utilizes PointNet++ to segment the entire plant into cotton bolls and non-boll cotton. With just a single segmentation step, it enables cotton boll counting and vertical distribution mapping. For the second branch (lower part of Fig. 1), it first employs TreeQSM to segment the plant into individual branches and the main stem. Then, PointNet++ is used to segment the cotton bolls on each branch. This process determines which cotton bolls belong to the same branch, eliminating the need for additional algorithms to map each boll back to its corresponding branch. Ultimately, this pipeline provides the horizontal distribution of cotton bolls on branches and identifies branch types. More detailed descriptions of the two approaches are introduced in the following sections.

## Data collection

In this experiment, individual cotton plants were studied. The datasets used were collected from the years in 2018,



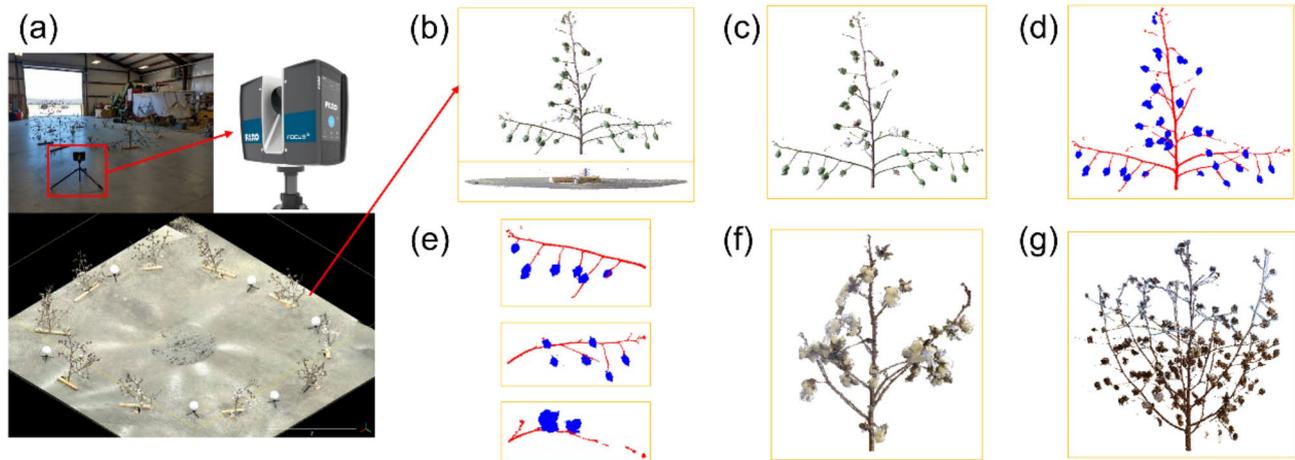
**Fig. 1** Cotton boll number and spatial distribution mapping workflow

2020, and 2021. The data collection considered two factors: the planting layout of the cotton plants for that year and the quality of the collected data. If the cotton plants were spaced far apart with no contact between adjacent plants, data were collected directly outdoors. For densely planted cotton plants, they were cut and brought indoors for data collection, which reduced environmental interference and resulted in higher-quality point cloud data. Dataset 1 was obtained in 2021 from a breeding field located at the Iron Horse Farm (latitude:37.730 N, longitude:83.303 W) in Greene County, Georgia, USA. A total of 34 individual plants were extracted from different field plots by severing them at the soil surface and were moved to an indoor setting. At this stage, the cotton bolls were green and not yet open, but the interior of the fruit is gradually maturing, forming fibers. The plants were held in place using brackets without overlaps between them. A high-resolution 3D LiDAR scanner (FARO Focus S70, FARO Technologies, USA) mounted on a tripod at a height of about 0.6 m was used to scan the cotton plants from multiple locations over a 360° horizontal plane (Fig. 2a). Color data was collected by the built-in color camera. Scans from multiple angles were registered as a single point cloud via FARO SCENE software (version 2019.2) (FARO Technologies, Florida, US). Dataset 2 was collected outdoors in December 2018, and 11 cotton plant samples were used [41], and the bolls were mature and mostly open. Dataset 3 was collected indoors in

December 2020 and February 2021, using 20 cotton plant samples [43], and the bolls in this batch were mature and fully open. A total of 65 plant samples were used in the experiment.

#### Point cloud data processing

The raw point cloud was preprocessed to isolate individual cotton plants. Initially, the individual plants were cropped from the plot level point cloud. Next, non-plant points such as ground and brackets were manually removed in CloudCompare. During the data collection process, it is common for outliers to appear due to factors like equipment accuracy and environmental conditions. A statistical outlier removal (SOR) filter was employed to eliminate these noisy data points. During the denoising process, the average distance of each point to its 6 nearest neighbors as well as the standard deviation of the distances were estimated, and those points whose distance exceeded one standard deviation were removed. After applying the preprocessing steps, individual cotton plant point clouds were obtained (Fig. 2c). All point cloud data underwent the same preprocessing operations, regardless of whether they were collected indoors or outdoors. Table 1 provides additional details about these data, with cotton boll points accounting for the majority of the entire plant, with an average proportion of approximately 67.95%.



**Fig. 2** Data acquisition and pre-processing workflow. (a) Experimental setup and raw point cloud collected by a terrestrial LiDAR. (b) Ground removal. (c) Denoised and down-sampled to 100,000 points. (d) Annotation of plant point cloud data: blue represents cotton bolls and red represents the main stem and branches. (e) Branches were extracted from a plant with 100,000 points to make a branch dataset. (f) and (g) are data collected from other years

**Table 1** Description of the cotton plant dataset. The non-boll proportion would be comprised of branches and main stems

	Date	Number of Samples	Boll proportion (%)	Non-boll proportion (%)
Dataset 1	November, 2021	34	67.36	32.64
Dataset 2	December, 2018	11	73.84	26.16
Dataset 3	December, 2020 & February, 2021	20	62.64	37.36

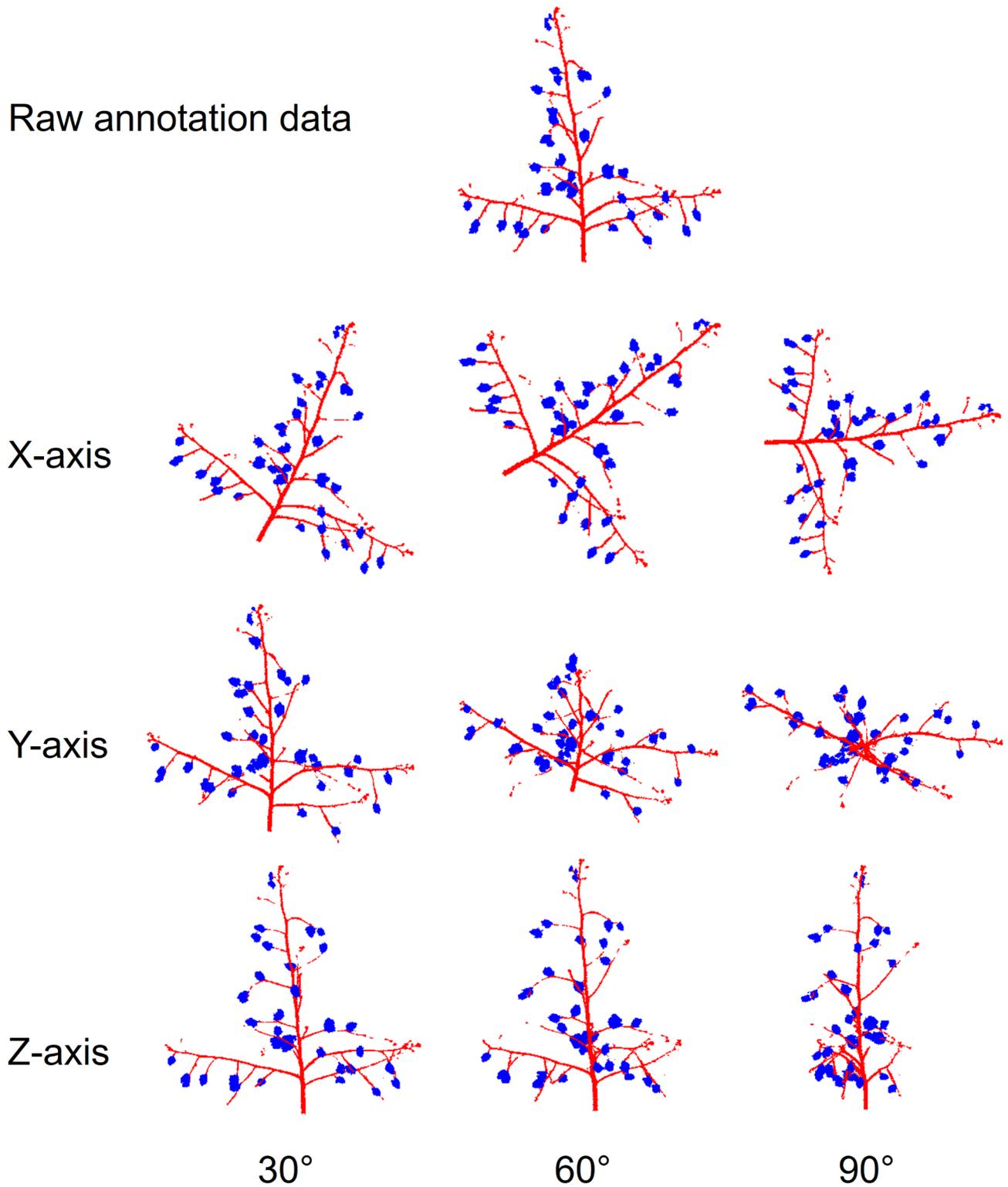
**Table 2** Overview of the number of different genotypes used in training and testing

Genotypes	Number of training samples	Number of testing samples
UA488	2	2
DP1646	5	2
ST5020	3	2
TO246BC3MDN	5	2
DESS6	3	4
DG3615	3	2
MDN0101	2	2
DP341	1	2
T0018MDN	2	1
Tamcot Sphinx	4	1
Acala Maxxa	2	3
T1046cBC1.GH212	1	1
T101MDN	-	1
Exotic T0281cMDN.GH200	2	-
Exotic T0151DN.GH180	1	-
Exotic T0347MDN.GH186	2	-
Exotic T0368BC3MDN.GH196	1	-
Elite Acala Maxxa 2011 (4201)	1	-

The pre-processed individual cotton plants were annotated to generate a ShapeNet format dataset [44]. The primary focus of this research was on cotton bolls, aiming to count and map them. Previous research by Saeed et al. [45] demonstrated that PointNet++ [46] achieved the highest segmentation accuracy for cotton bolls but struggled with differentiating between branches and main stems. As cotton bolls are spherical and both branches and main stems are cylindrical, distinguishing between bolls and branches was relatively easier compared to differentiating between branches and main stems. Therefore, this study focuses on using PointNet++ to classify cotton plants into two categories. To enhance the segmentation accuracy, branches and main stems were annotated as one class. The labeled data is shown in Fig. 2d, the blue point cloud represents cotton bolls ( $x, y, z, R, G, B, 0$ ), and the red point cloud represents non-bolls ( $x, y, z, R, G, B, 1$ ). Each point in the dataset contains coordinates, colors, and class labels. In our study, for mature cotton plants, the leaves were removed by spraying a defoliant, which is a common practice in commercial cotton harvesting; for immature plants, we manually removed the leaves before data collection to focus more on studying the distribution of cotton bolls.

The annotated dataset consists of 65 plants of 18 genotypes (Table 2), with data augmentation by rotating the point cloud around the three axes with different angles [47, 48]. Data augmentation is an effective technique to enhance the model training performance when data are limited [49, 50]. These samples were divided into training ( $n=40$ ) and testing ( $n=25$ ) sets. Subsequently, each sample was individually rotated around the  $x$ -,  $y$ -, and  $z$ -axes by angles of  $30^\circ$ ,  $60^\circ$ , and  $90^\circ$ , resulting in the generation of a total of 400 training samples (Fig. 3). To save computational resources, the labeled cotton plants were

## Raw annotation data



**Fig. 3** Data augmentation of annotated point cloud data. The annotated point cloud of a single plant was rotated 30, 60, and 90 degrees around the x-, y-, and z-axes, respectively

downsampled to approximately 100,000 points using spatial downsampling in CloudCompare. In addition, as the models trained on the whole plants are unsuitable for testing on individual branches, the branch dataset was created by selecting cotton plants with 100,000 points and cutting branches from cotton plants in contact with the main stem (Fig. 2e). There were 733 samples (390 for training and 343 for testing) in the branch dataset. Data pre-processing and labeling were performed using the open-source software CloudCompare v2.12. alpha. This tool was used to annotate the data, a crucial step for training PointNet++. To ensure accuracy, the labeled data was thoroughly reviewed three times to prevent annotation errors.

**Boll number and vertical distribution on a whole plant**

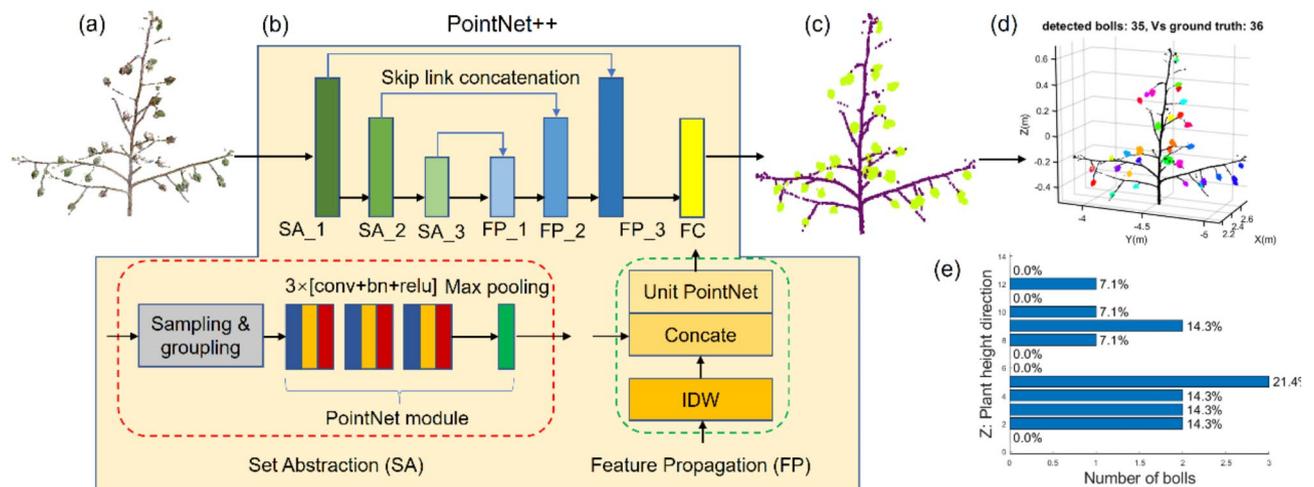
To obtain the number of cotton bolls and their distribution along the vertical axis for whole plants, a series of steps were followed. First, PointNet++ was trained to segment cotton bolls from individual plant point clouds. Then, Euclidean clustering was used to segment each cotton boll instance for counting. Finally, the centroid of each boll was determined by calculating the mean of each cluster, which was used to map the distribution of bolls at different plant heights (Z-axis).

**Semantic segmentation based on PointNet++**

Cotton bolls were segmented from plants by PointNet++. This model is a deep learning network that directly performs semantic segmentation on point clouds. One of the key advantages of deep learning is its ability to automatically extract features. Compared to manually designed features, deep features offer stronger representational power and greater robustness. The structure of

PointNet++ is shown in Fig. 4b. It completes the down-sampling and up-sampling of the point cloud by the two modules of set abstraction (SA) and feature propagation (FP), and gradually captures the feature information in the point cloud. The SA module consists of sampling, grouping, and PointNet, which performs sub-sampling, region proposal, and point feature extraction on point cloud. Stacking multiple SA layers (SA\_1, SA\_2, SA\_3) allows for capturing features at different scales. Feature Propagation (FP) is used to propagate high-level features to lower levels, thereby generating dense point cloud features. The purpose of stacking multiple FP layers (FP\_1, FP\_2, FP\_3) is to progressively recover and refine feature information, enabling the model to produce more accurate and detailed point clouds. Each FP layer gradually propagates high-level features to lower levels and uses the inverse distance weighting (IDW) method for interpolation, achieving step-by-step feature refinement.

PointNet++ was trained using the PyTorch framework. The labeled data was fed into the network with the batch size of 2. In the training process of PointNet++, an epoch refers to a complete pass of the entire training dataset through the network. During each epoch, the model processes all input data, makes predictions, and updates its weights based on the loss function. Training involves iteratively adjusting the model parameters across multiple epochs to minimize the loss and improve accuracy in tasks such as cotton boll segmentation. The initial learning rate was set to 0.001, and the Adam optimizer was used to adjust the learning rate over time. Point cloud normalization was applied for center alignment and scale unification, eliminating position and size differences. The model was trained for a total of 251 epochs. The input point features of PointNet++ have six dimensions,



**Fig. 4** Workflow of mapping cotton bolls along plant height. (a) Input of the cotton plant point cloud. (b) PointNet++ structure. Inverse distance weighting (IDW) is used for interpolating features from sampled points to the original points. Concat is then used to concatenate the interpolated features with low-level features. (c) Semantic segmentation results of PointNet++. Green represents cotton bolls and purple represents non-cotton bolls. (d) Euclidean clustering of cotton bolls. Each cluster was randomly assigned a color. (e) Distribution of cotton bolls at different heights

including coordinates ( $x, y, z$ ) and color (R, G, B). For a single plant, the number of input points is set to 100,000, while for branch data, the input point cloud is set to 10,000. The output after training is a point cloud with semantic information, where each point is assigned a semantic label. The PointNet++ model was trained on the HiperGator high-performance computing cluster with 8 AMD EPYC ROME CPU cores, one NVIDIA DGX A100 GPU node (80 GB). The operating system is Linux, and the software libraries include python 3.8, pytorch 1.12.1, and CUDA 11.4. The code, data, and training weights for cotton boll and branch mapping are available at [https://github.com/UGA-BSAIL/cotton\\_organ\\_mapping.git](https://github.com/UGA-BSAIL/cotton_organ_mapping.git).

#### ***Boll number counting based on Euclidean clustering***

The segmented bolls were counted by Euclidean clustering. Cotton plants were divided into cotton bolls and non-cotton bolls (branches and main stems) by PointNet++, and each predicted plant had 100,000 points. Euclidean clustering was performed on the points segmented as cotton bolls, with a threshold set empirically of 0.8 cm for the Euclidean distance in this study. Points with distances smaller than the threshold were grouped into one cluster (representing a cotton boll), while points with distances greater than the threshold were assigned to different clusters. This threshold was determined based on the spatial distribution characteristics of the cotton boll point cloud. In our dataset, the distances between points within most single cotton boll point clouds are typically less than 0.8 cm. If the threshold is too large, adjacent but independent bolls may be mistakenly grouped into the same cluster. Conversely, if the threshold is too small, a single boll may be incorrectly divided into multiple clusters. Through extensive trials with different thresholds, 0.8 cm was found to provide the best balance, aligning well with the natural size and point density distribution of the bolls, and effectively minimizing segmentation errors. During the counting process, each cluster with fewer than 100 points was regarded as noise and removed.

#### ***Vertical distribution of cotton bolls***

The vertical distribution reveals the position of each cotton boll along the height direction of the plant. To determine the distribution of cotton bolls at different heights, the entire plant was sliced vertically into sections with a thickness of 10 cm. Because the plants are mainly concentrated between 1.0 and 2.0 m, this resolution can effectively capture the spatial variation of the cotton bolls. The number of centroids representing cotton bolls in each slice was then counted, enabling the creation of a boll height distribution map with a resolution of 10 cm. This map provides valuable information about the concentration of bolls at various heights within the plant.

Additionally, the cotton plant was divided equally into three parts along the  $z$ -axis. This division allows for a rough estimation of the number of bolls in the upper, middle, and bottom parts of the plant. This estimation serves as a helpful tool in less demanding scenarios, such as gaining insights into the growth pattern of cotton plants. A few similar studies have been conducted using different data processing methods [39, 40].

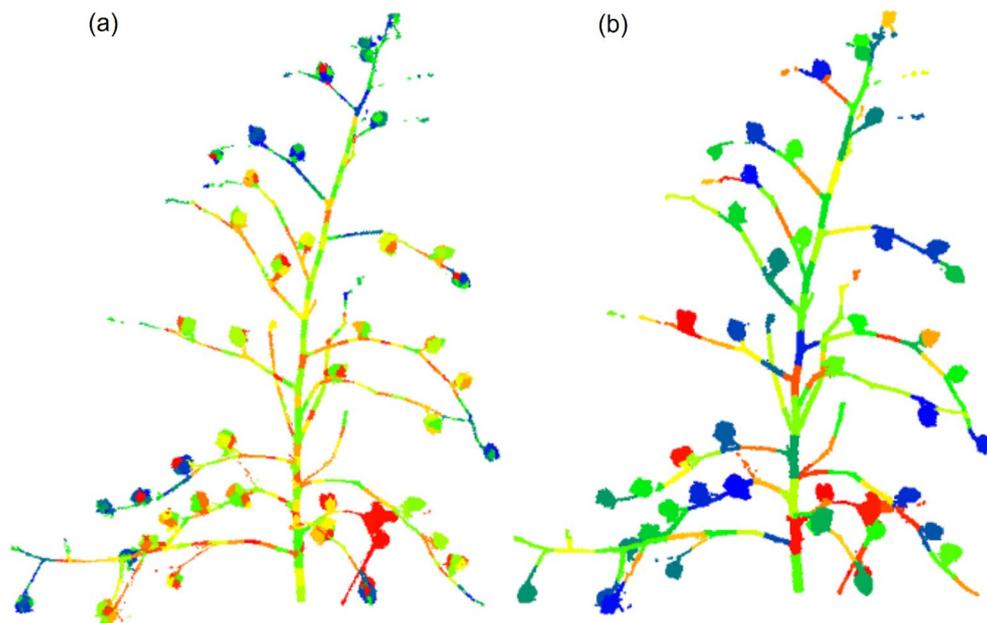
#### ***Horizontal distribution on individual branches***

To achieve boll mapping along individual branches, three main steps were followed. The first step was to segment the plant into the main stem and individual branches using TreeQSM. In the second step, PointNet++ was used to segment the cotton bolls on each branch, followed by the implementation of cotton boll instance segmentation through Euclidean clustering. Finally, the Euclidean distance from each cotton boll centroid to the main stem was calculated, and the distances were ranked to determine the distribution of cotton bolls relative to the main stem.

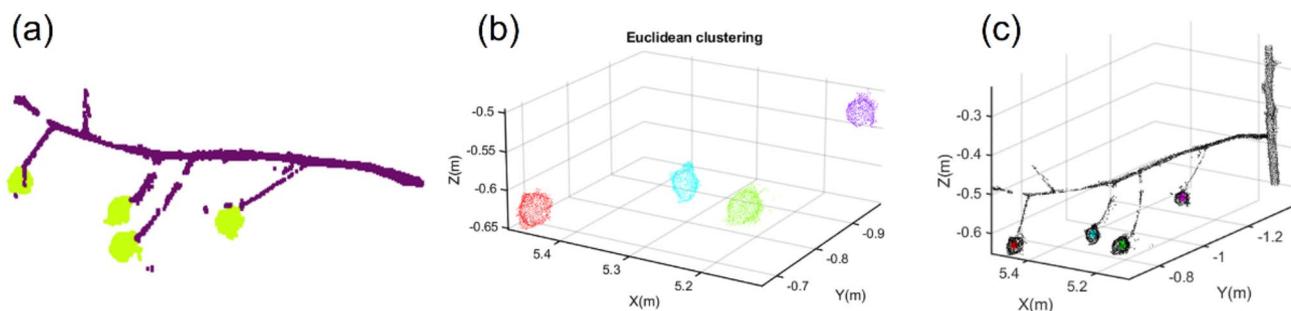
#### ***Instance segmentation of main stem and first-level branches based on TreeQSM***

Cotton plants were segmented into main stem and first-level branches (Individual branches in direct contact with the main stem) by TreeQSM. TreeQSM uses cover sets (small point cloud sets or small patches) to segment the point cloud along the main stem starting from its base in two phases. First, a large constant size cover set with radius Patch Diameter 1 (PD1) was applied to the entire plant. This step allows for quick and rough identification of the structural features of plants. Second, a finer patch of variable-sized from Patch Diameter 2 (min) (PD2Min) to Patch Diameter 2 (max) (PD2Max) determined the final branch topology. PD2Min plays a key role in TreeQSM tuning, as it defines the smallest features that will be segmented. Hence, PD2Min must be adapted to the smallest features of cotton plants. In the segmentation of cotton plants, a smaller patch diameter (such as PD2Min) captures more local details, making it suitable for segmenting smaller structural features, such as cotton bolls or small branches. In contrast, a larger patch diameter helps capture more global information, making it more suitable for coarse structural segmentation, such as the main stem and larger branches.

Cover sets are the smallest “units” that segment the cotton plant point cloud into main stem and individual branches. As shown in Fig. 5, cover sets with different diameters were shown. The smaller the cover set (Fig. 5a), the more detailed information can be captured. But it also creates more disconnected structures and consumes more time. The larger the cover set (Fig. 5b), the more global information can be obtained, and the



**Fig. 5** Comparison of cotton plant cover sets with different diameter sizes. The minimum diameters (PatchDiam) of the cover sets are 5 cm (a) and 10 cm (b). Each cover set was assigned a different color



**Fig. 6** Boll positioning on a single branch. (a) PointNet++ semantic segmentation. (b) Euclidean clustering of cotton bolls. (c) Boll centroid. The colored points on the bolls are the centroids

corresponding time can be saved. The patch diameter determines the size of the cover set. It needs to be set according to the specific structural features of the cotton plant, such as the diameter of the main stem and the size of the branches. PD2Min should be adjusted to match the smallest features of the cotton plant to ensure that the segmented structures are neither too coarse nor incomplete. By conducting multiple experiments and adjusting different patch diameters, a balance can be found that enables precise segmentation of the various parts of the cotton plant.

The hyperparameters were influenced by point cloud density, plant size, and structural complexity. Following the recommended settings of TreeQSM, the radius was determined using the point cloud from the bottom 10% of the plant's main stem. Based on this radius, PD1 was set to approximately one-third, PD2Min to one-sixth, and PD2Max to two-fifths of the radius. Given the

minimal variation in the main stem diameter of the cotton plants in this study, PD1, PD2Min, and PD2Max were manually set to [0.2, 0.15, 0.3] cm, respectively, and these parameters were consistently applied to all plants. The segmented point cloud was saved for further quantitative evaluation.

#### **Boll positioning on an individual branch**

To obtain the cotton boll class on individual branches, PointNet++ was used again for semantic segmentation (Fig. 6a). The boll categories were clustered by Euclidean clustering (distance threshold was fixed to 0.8 cm), and each small cluster obtained was considered a boll (Fig. 6b). Then, the mean of each cluster was calculated to determine the centroid representing the position of each cotton boll (Fig. 6c).

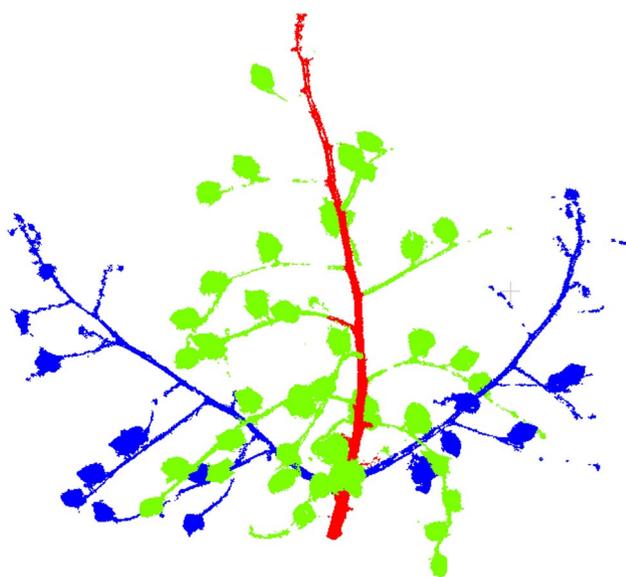
### Distribution of bolls relative to the main stem

After determining the positions of each cotton boll, the Euclidean distance from each cotton boll on the branches to the main stem was calculated. Then, the distances were sorted to get the distribution of bolls relative to the main stem. In the Euclidean clustering process, there will be noise that was mistaken for bolls, and there will be very small bolls. In the experiment, clusters with fewer than 100 points were reclassified as non-cotton bolls. Finally, the distribution of bolls relative to the main stem was mapped.

### Branch type identification

Cotton branches can be classified as vegetative branches and fruiting branches [27]. As can be seen in Fig. 7, cotton bolls are usually attached directly to the fruiting branches via a short stalk, while on vegetative branches, cotton bolls grow on secondary branches (branches that grow from a primary branch). These two types of branches can be distinguished based on the Euclidean distance from the cotton bolls to the primary branch (branches that grow directly from the main stem).

The shortest path method was used to obtain primary branches. In some cases, there will be gaps in the branches, which makes it relatively difficult to obtain the path, so the point cloud completion needs to be done first. A point cloud meshing method based on concave hull was used to fill the gaps. The point cloud was meshed to generate triangular patches that tightly surrounded all points, and the single branch formed a connected whole (Fig. 8b). Each triangle patch was then uniformly down-sampled such that the final single branch contained 2000



**Fig. 7** Vegetative branches and fruiting branches on a cotton plant. Blue points represent vegetative branches, green points represent fruiting branches, and red points represent the main stem

points. As shown in Fig. 8c, all gaps in the point cloud were filled. Although some extra points (non-branch points) are introduced, it does not affect subsequent processing. Next, the two endpoints of the path were determined by calculating the nearest and farthest points of the branch from the main stem of the cotton plant. Finally, the Dijkstra algorithm [51] was used to obtain the points of the shortest path between the two endpoints (red points in Fig. 8d). Those points were used to represent primary branches.

The branch type was determined according to the distance from the cotton bolls to the primary branch. This distance was calculated for each boll (represented by the cluster's centroid) to the main branch (represented by the shortest path points). If any of the calculated distances were larger than the set threshold (10 cm), the branch was considered a vegetative branch. Otherwise, it was considered a fruiting branch. The threshold was manually determined based on the studied plants. As shown in Fig. 7, cotton bolls on fruiting branches are directly connected to the main branch through a short peduncle. In contrast, vegetative branches typically grow secondary branches from the main branch, and the cotton bolls on vegetative branches are located on these secondary branches, resulting in a greater distance from the main branch. By comparing multiple threshold values, we found that 10 cm was the most appropriate threshold.

### Evaluation metrics

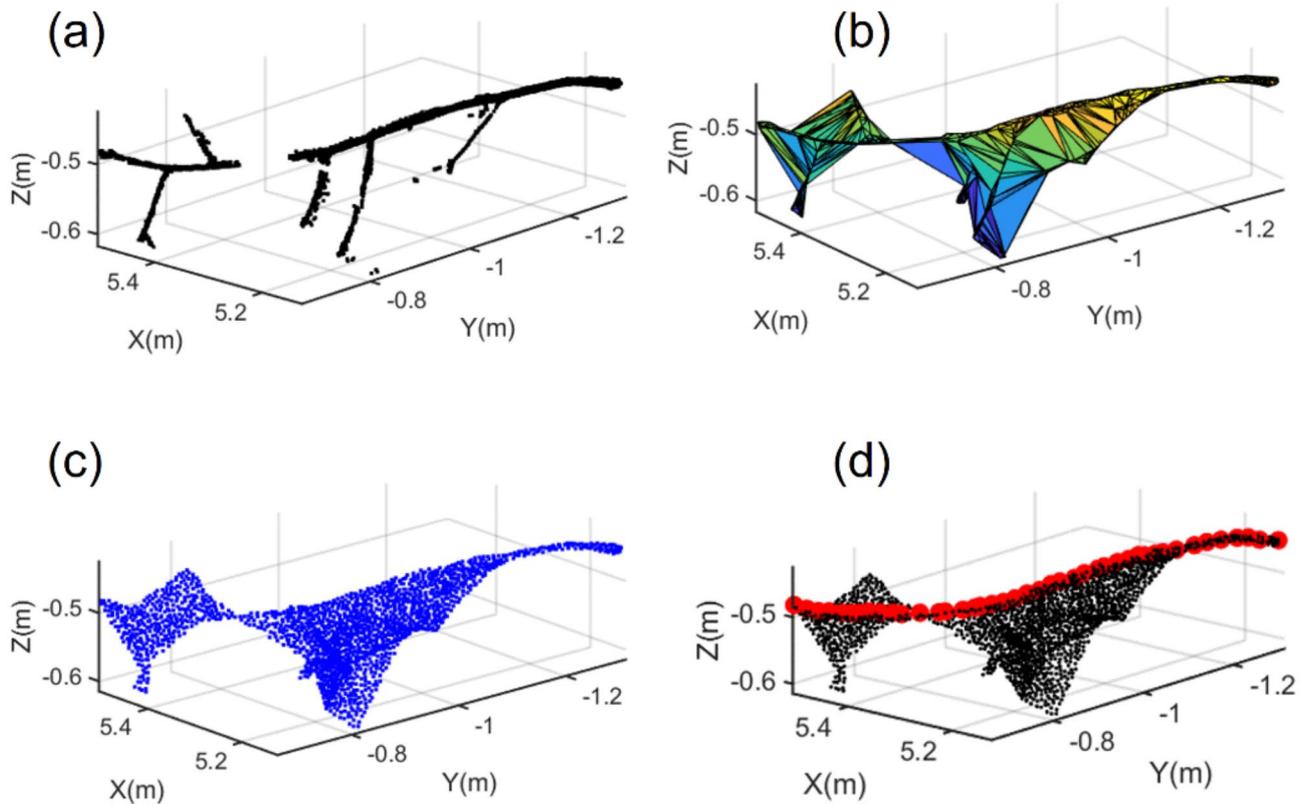
Accuracy, intersection over union (IoU), precision (P) recall (R) and F1-score (F1) were used to evaluate the performance of the point cloud segmentation. Accuracy is the ratio of correctly predicted data points to the total number of data points. IoU indicates the similarity of the predicted region of a category to the ground truth region. P and R denote the proportion of correctly predicted points to the total predicted points and total ground truth points, respectively. F1 is a comprehensive indicator calculated as the harmonic mean of P and R. For all the above indicators, the larger the value, the better the segmentation result. The five indicators are defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (2)$$

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$



**Fig. 8** Primary branch point acquisition based on point cloud meshing. (a) Point cloud with multiple gaps. (b) Point cloud meshing processing. (c) Mesh sampling into point cloud. (d) Shortest path points (The red points represent the primary branches defined in this study)

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \quad (5)$$

Where TP, TN, FP, and FN represent the number of true positive, true negative, false positive, and false negative points of a certain class, respectively.

The evaluation of traits was conducted using root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Here,  $N_i$  represents the predicted value of the  $i$ -th sample,  $m_i$  represents the true value of the  $i$ -th sample, and  $n$  is the total number of samples.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (N_i - m_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |N_i - m_i| \quad (7)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{N_i - m_i}{m_i} \right| \quad (8)$$

**Table 3** Segmentation performance of PointNet++ on the individual plant and branches

	Accuracy	mIoU	Boll IoU	Branch IoU
Whole plant dataset	0.954	0.896	0.932	0.860
Branch dataset	0.968	0.897	0.929	0.865

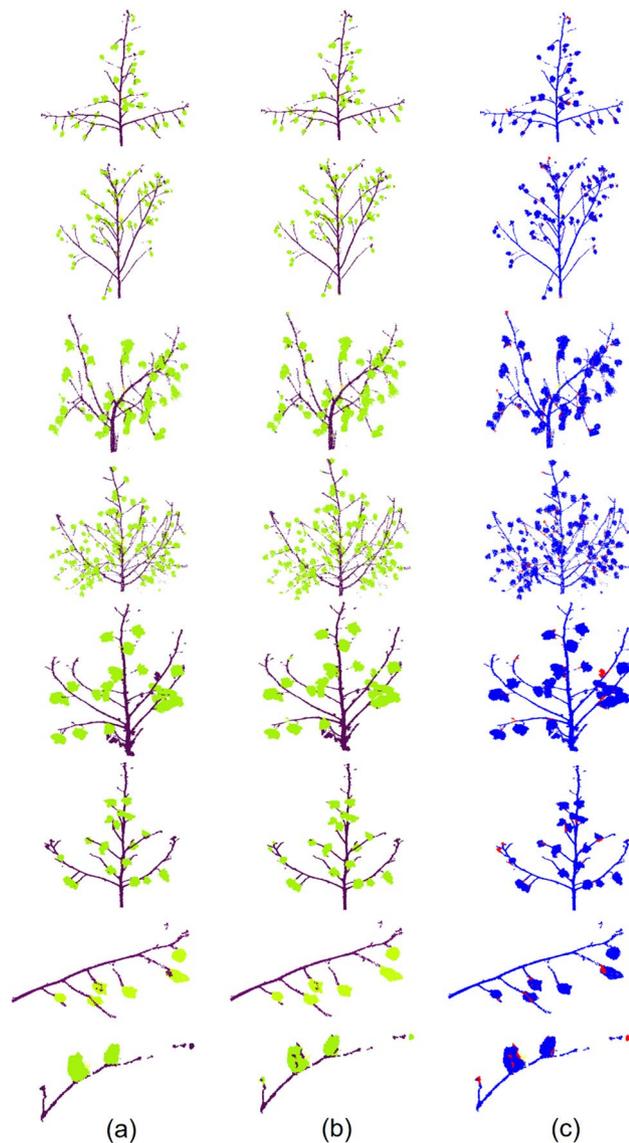
## Results

### Performance of PointNet++ on semantic segmentation

PointNet++ achieved excellent performance in semantic segmentation on both the cotton plant (mIoU=0.896) and branch (mIoU=0.897) datasets, successfully segmenting cotton bolls from the branches. Table 3 presents the good performance of PointNet++ in terms of both accuracy and mIoU on two distinct datasets. In terms of categorization, the model efficiently partitions the point cloud into cotton bolls and non-bolls, with particular effectiveness in segmenting cotton bolls, which are of primary interest in this study. It's worth noting that the intersection over union (IoU) for cotton bolls exceeds 0.9, indicating a higher level of segmentation accuracy compared to non-cotton bolls. The model size of PointNet++ was 20 MB, with 1.48 million parameters, demonstrating that despite the model's relatively small size, it was capable of segmenting dense cotton plant point clouds. On the NVIDIA DGX A100 GPU, the training

time for a single cotton plant dataset was 7.1 h, and the testing time for 25 samples was 81s (with 100,000 sampling points). The training time for the branch dataset was 6.3 h, and the testing time for 343 samples was 190s (with 5,000 sampling points).

Figure 9 shows the test results achieved by PointNet++. The model demonstrates its ability to effectively segment plants and branches of varying shapes, with only a few points predicted incorrectly (highlighted as red points in Fig. 9c). The main errors are concentrated in the areas where cotton bolls and branches are in contact, mainly due to their high similarity, which makes it challenging



**Fig. 9** Visualization of semantic segmentation using PointNet++ on whole plants and branches. (a) Ground truth of the annotated data. (b) Prediction from semantic segmentation by PointNet++. (c) Difference between the ground truth and the prediction (red points represent errors while blue points represent correctly predicted points). The first six rows are individual plant data, and the last two rows are branch data

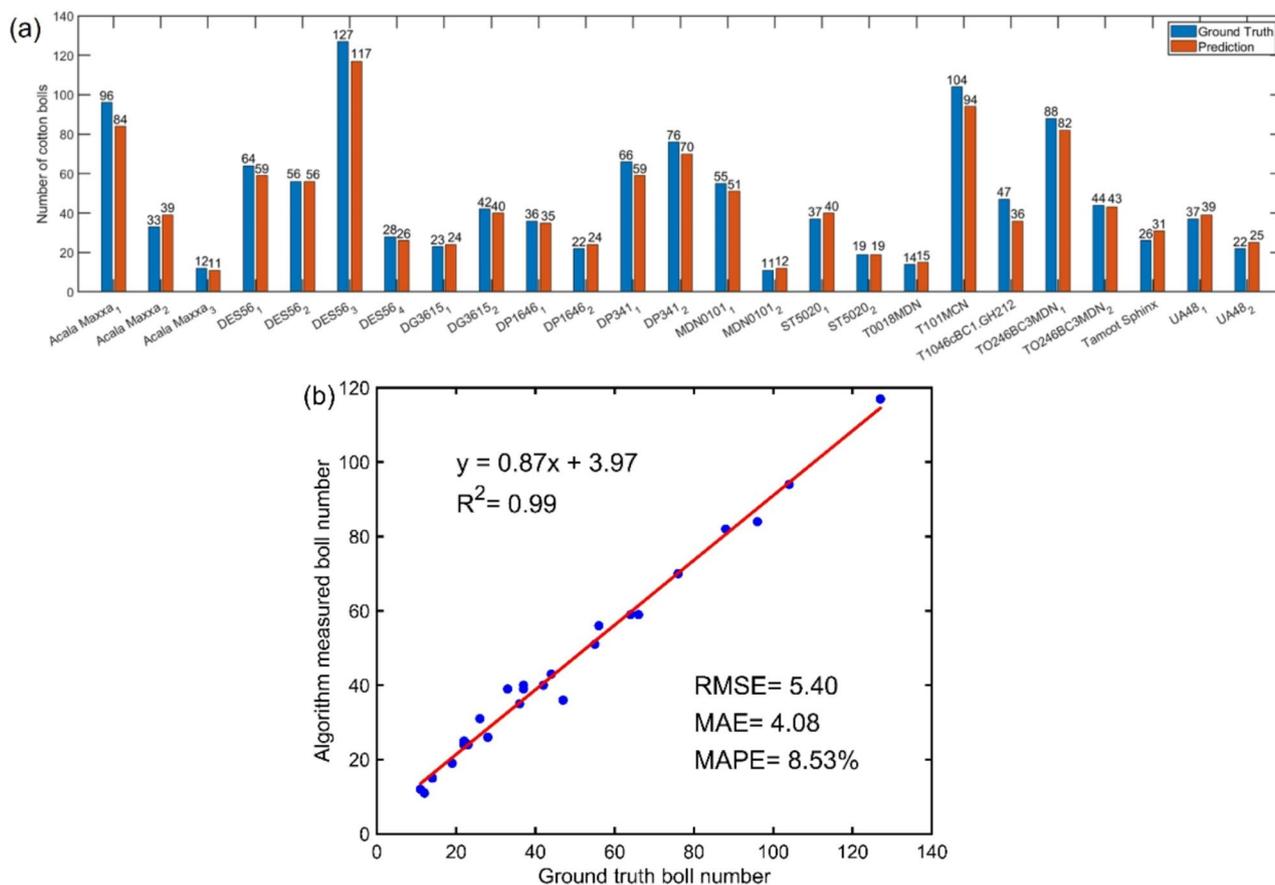
for the model to accurately distinguish the boundaries. Furthermore, there is a tendency for small spherical point cloud regions to be erroneously segmented as cotton bolls. However, by applying a point threshold (e.g., 100 points), these erroneous segmentations can be further eliminated. The PointNet++ model has shown good performance in accurately depicting cotton bolls.

#### Boll number counting for the single cotton plant

The boll counting results obtained from the proposed method were highly accurate and demonstrated a strong correlation with manual counting results (Fig. 10). Twenty-five representative samples were carefully selected for the counting evaluation, considering their diverse shapes, varying structural complexities, and different numbers of bolls per plant. The evaluation revealed an overall boll counting accuracy of over 90% (MAPE=8.53%) with an  $R^2$  value of 0.99. However, it is worth noting that in 14 out of the 25 samples illustrated in Fig. 10a, the algorithm predictions underestimated the actual values. This underestimation can be attributed to the clustering process, where the algorithm erroneously groups different cotton bolls into a single cluster due to their close proximity. Additionally, 9 samples were overestimated. This is because some non-boll points are predicted to be bolls during the PointNet++ semantic segmentation process, causing these isolated points to be aggregated into boll instances. This evidence that selecting an appropriate clustering threshold is also crucial for accurate counting.

Overall, despite these challenges, the counting results obtained in this study are considered acceptable. While the method we developed does not yet achieve the 100% accuracy of manual counting, the high correlation of 0.99 between the predicted and actual boll counts demonstrates the reliability of our method and has already met the requirements of our breeding program (counting errors are required to be less than 20%). Although 14 out of the 25 samples underestimated boll counts due to clustering of closely positioned bolls and 9 samples overestimated due to misclassified points, these deviations did not significantly affect the overall ranking or selection of high-yielding plants. More importantly, our predicted boll counts provide a practical and efficient means to estimate the yield of individual cotton plants (as shown by the red bar in Fig. 10a), which is highly beneficial for selecting high-yielding plants in cotton breeding.

Different plant structures influence the accuracy of cotton boll counting. When the bolls on a plant are sparse, the larger gaps between them result in predicted counts that closely match the ground truth. However, when the bolls are densely packed, they tend to touch each other, which can cause multiple bolls to be grouped into a single cluster, leading to counting errors. Once multiple



**Fig. 10** Cotton boll counting results from whole plants. **(a)** Comparison of predicted and ground truth cotton boll numbers from 25 plants and 13 genotypes. **(b)** Linear regression of cotton boll numbers between predicted from our algorithm and the ground truth

bolls merge into a cluster, the cluster’s volume increases significantly. Optimizing boll counting accuracy based on volume is a potential direction for future improvements.

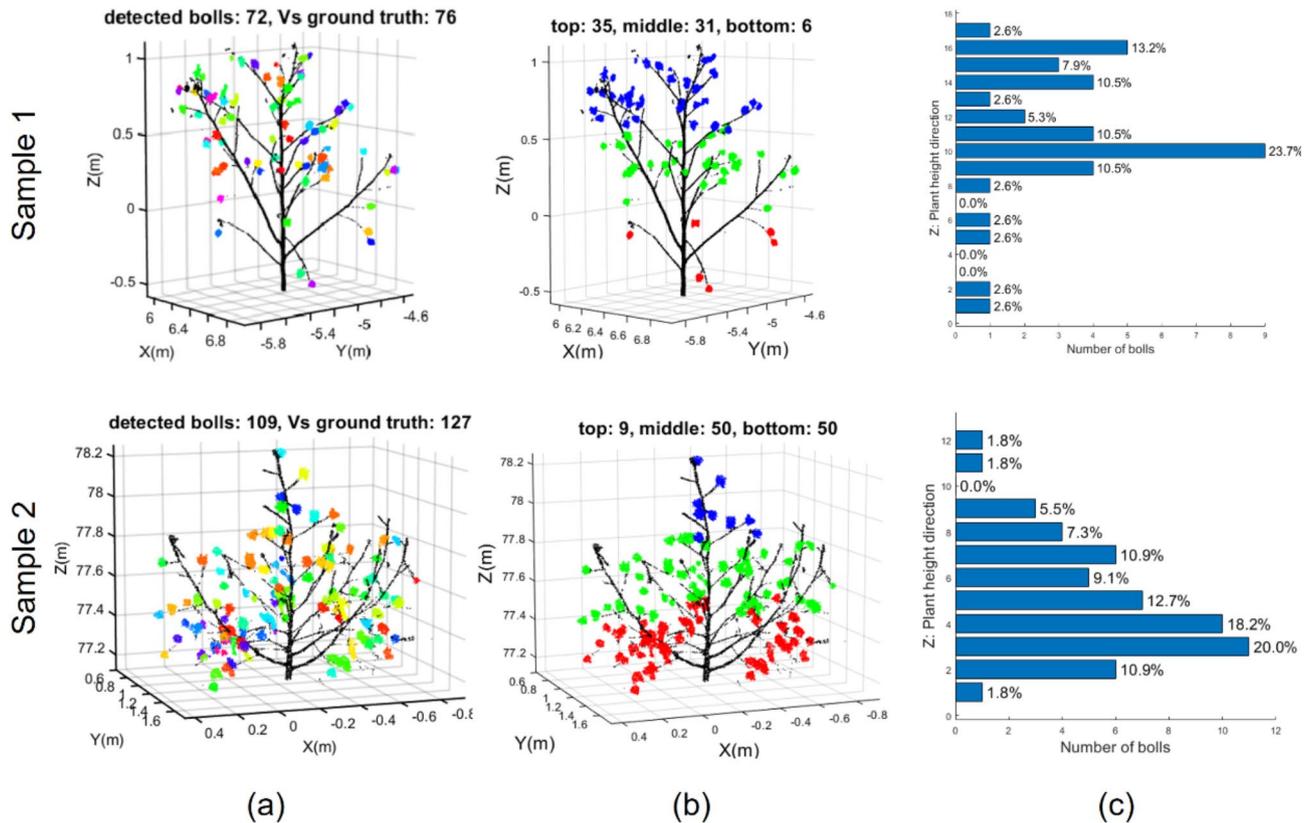
**Mapping vertical distribution of cotton bolls on a single plant**

The spatial position and distribution of cotton bolls at different heights can be easily observed and understood by data visualization (Fig. 11). In Fig. 11a, each cluster was assigned a color to represent a predicted boll, and the location of each cotton boll was plotted across the entire plant. Figure 11b provides a rough estimation of cotton boll numbers for the equally divided upper, middle, and lower parts. It allows us to observe the yield distribution in these three parts for plants with different structures. During the growth of cotton bolls, the lower bolls mature first but receive limited sunlight due to their position. The upper bolls, which mature last, are more affected by water and nutrient availability, often resulting in relatively lower quality. In contrast, the middle bolls benefit from adequate sunlight and nutrient supply, making them more likely to produce high-quality cotton bolls. We can roughly estimate the yield distribution by

dividing the cotton bolls into three Sects. [52, 53]. Figure 11c is a highly detailed yield distribution map that counts the number of cotton bolls at every 10 cm height interval. The proportions of cotton bolls at different heights are also displayed in the figure. Since boll density is one of the most important determinants of yield [2], this comprehensive depiction offers insights into the overall yield distribution across the entire plant.

**Evaluation of TreeQSM on main stem and branch instance segmentation**

The TreeQSM algorithm showed promising results in effectively separating plants into main stems and individual branches. Table 4 shows that the method is particularly effective for segmenting the main stem, while the segmentation performance for individual branches is slightly less effective. A plant has only one main stem, which reduces the sources of segmentation errors. The main stem is usually thicker and straighter than branches, while branches are sometimes densely intertwined, increasing the complexity of branch segmentation. Overall, TreeQSM is suitable for various cotton plant structures, including upright and slightly curved



**Fig. 11** Illustrate of the vertical distribution of cotton bolls from two samples. (a) Instance segmentation of cotton bolls based on Euclidean clustering. Each cotton boll is randomly assigned a color in HSV color space. (b) Number of bolls in the top, middle and bottom parts. Blue points represent the top bolls, green points represent the middle bolls, and red points represent the bottom bolls. (c) Number of bolls at different heights with 10 cm increments

**Table 4** Evaluation of TreeQSM segmentation results

	P	R	F1	Accuracy
Main stem	0.945	0.825	0.875	0.952
Branch	0.797	0.884	0.807	0.796
Average	0.871	0.855	0.841	0.874

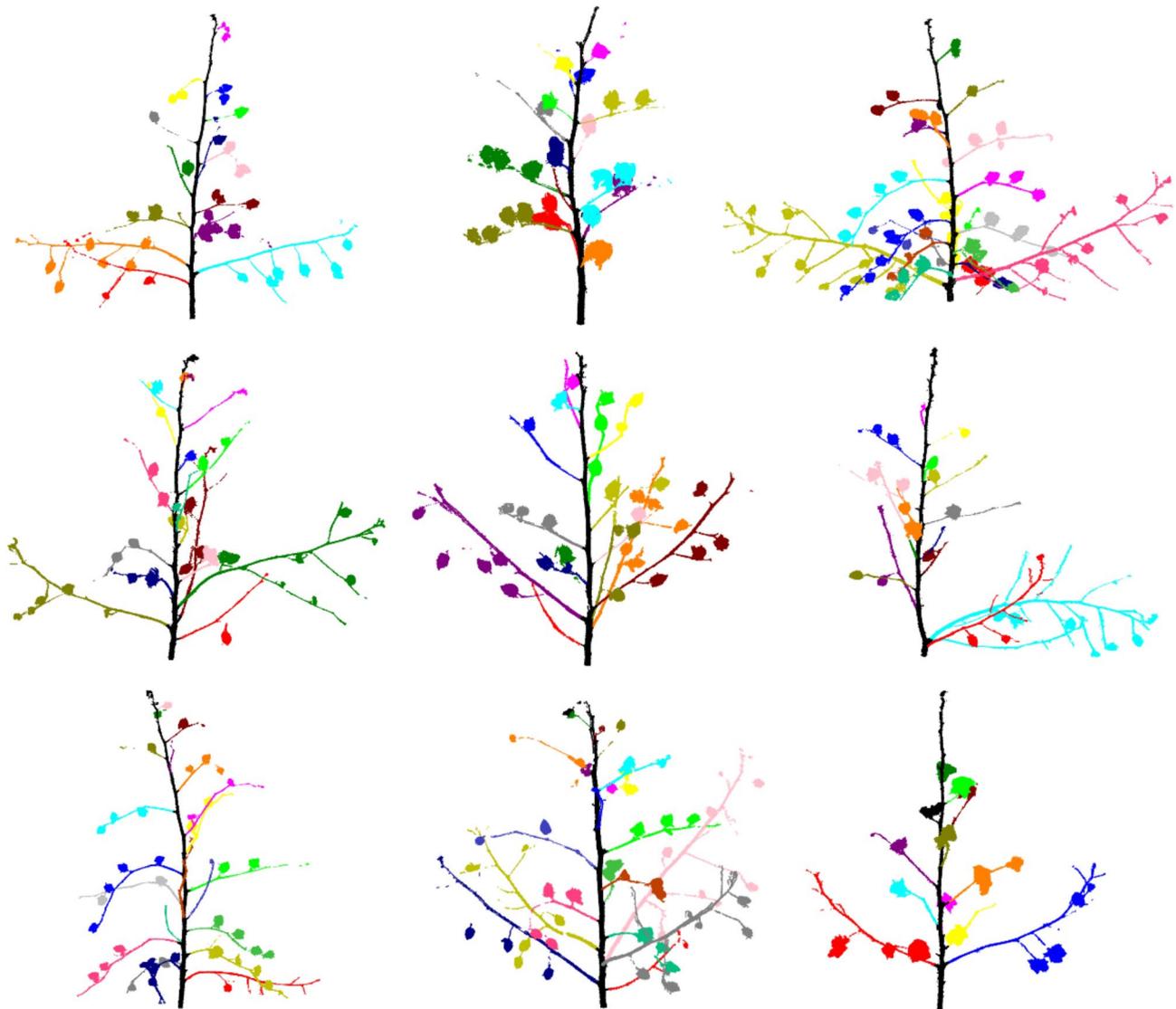
main stems and sparse or dense but non-overlapping branches (Fig. 12). This flexibility makes the algorithm highly versatile, capable of addressing the needs of different plant architectures. This method is suitable not only for large, robust trees but also for plants with fine main stems and branches, such as cotton. The segmentation of the main stem and branches of cotton plants achieved an average F1 of 0.841 and an Accuracy of 0.874, effectively meeting the needs of subsequent research.

TreeQSM segmentation results overall achieved consistently high accuracy across different genotypes but also vary to certain extent in the test set (Fig. 13). When individual branches do not touch—whether they are sparse (e.g., genotype ST5020) or relatively dense (e.g., genotype Acala Maxxa)—TreeQSM can accurately segment the entire plant into individual branches and the main stem. However, in some genotypes, branches are densely packed and touch each other, leading to suboptimal

segmentation results. In the T101MCN genotype, some cotton bolls are tightly attached to the main stem, causing boll point clouds to be included in the segmented main stem. TreeQSM can handle slight curvature in the main stem, but when the main stem is severely curved, the segmentation results are less accurate (e.g., genotype T1046cBC1.GH212). The above results are based on the currently collected data. Since the structure of cotton plants is influenced not only by genotype but also by the growing environment, which in turn affects the segmentation results, further analysis with additional plant samples is required in the future.

**Mapping of the boll distribution relative to the main stem**

The bolls on each branch were mapped relative to the main stem according to their Euclidean distance from it. To evaluate boll distribution, this experiment treated bolls of the same position from the main stem as one class. In the assessment results shown in Table 5, Level\_1 (represented by the red bolls in Fig. 14) exhibits the highest accuracy in cotton boll distribution, while the accuracy visibly decreases as the distance of the bolls from the main stem increases. The boll distribution map presents the count of bolls at each position from the main stem,



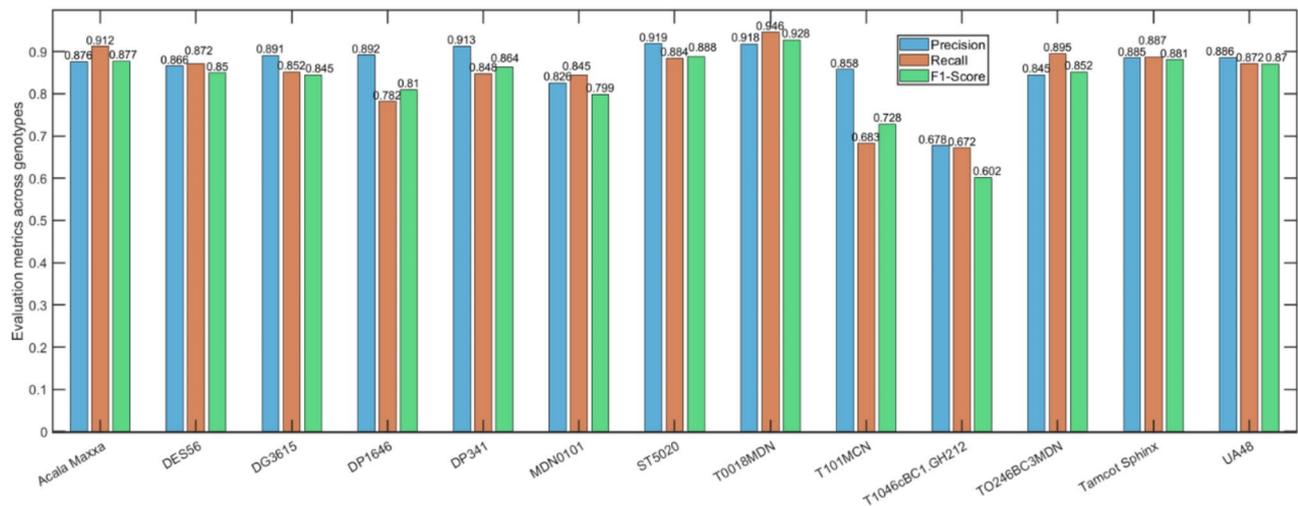
**Fig. 12** Illustration of branch and main stem instance segmentation based on TreeQSM for nine plant samples. Black points represent the main stem, while the remaining branches connected to the main stem are assigned different colors

and it consistently shows that cotton plants with different structures exhibit the highest number of bolls at the first level. Since the first and second levels make up the majority of the bolls, we only evaluated the first two levels.

The accuracy of the cotton boll distribution map is the result of multiple factors. First and foremost, it is related to TreeQSM, as only by accurately segmenting individual branches can each cotton boll be mapped to its corresponding branch. Another factor is PointNet++, whose primary responsibility is to separate bolls from individual branches. Next is the clustering operation to obtain individual boll instances to calculate the distance from each boll to the main stem. All the aforementioned factors affect the quality of the cotton boll distribution map.

### Branch type classification

By leveraging the instance segmentation results of main stems and branches from the TreeQSM, individual branches of each plant were further classified into vegetative branches and fruiting branches. Figure 15 presents a visualization of branch classification for different samples, where red, green, and black correspond to vegetative branches, fruiting branches, and main stems, respectively. Based on the evaluation results, the method demonstrates robustness ( $R^2 = 0.83$ ) when comparing predicted values of the vegetative branches with their corresponding real values (Fig. 16). It can be seen from the displayed samples that the number of vegetative branches is less than that of fruiting branches, and the vegetative branches grow only at the lower part of the main stem.



**Fig. 13** Comparison of TreeQSM segmentation results across different genotypes

**Table 5** Evaluation of the distribution of bolls relative to the main stem. Level\_1 includes the cotton bolls on each branch that are closest to the main stem (the red Bolls in Fig. 14); level\_2 includes the cotton bolls on each branch that are the second closest to the main stem (the green bolls in Fig. 14)

	P	R	F1
Level_1	0.762	0.635	0.688
Level_2	0.622	0.461	0.510
Average	0.692	0.548	0.599

## Discussion

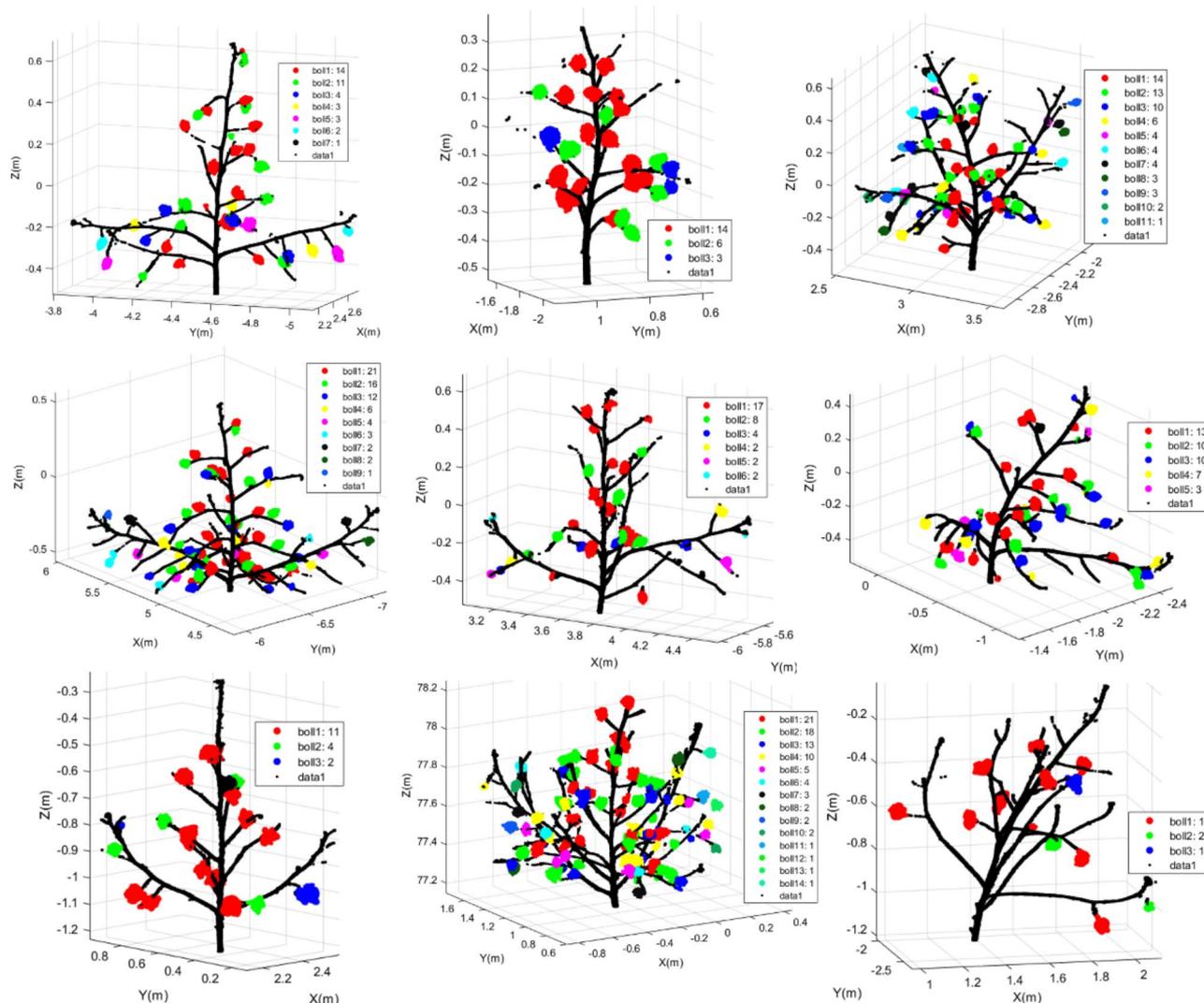
### Research problem and practical significance

This research presents a methodology to map cotton bolls, main stems, and branches with a high spatial resolution. The method provides a convenient tool for breeders and physiologists to identify genotypes with desirable traits. The first aspect of the method involves estimating the boll number per plant. The method used in this study achieved a strong correlation with the ground truth ( $R^2=0.99$ ) and a low mean absolute percentage error (MAPE = 8.53%). Compared to the segmentation of maize ears [54], our segmentation of cotton bolls achieved a mIoU that was 5% higher, indicating that PointNet++ was better suited for cotton boll segmentation. In the organ-level trait assessment of maize plants [30], the estimated values for leaf length, plant height, and stem length had R values exceeding 0.9, showing a higher correlation with the actual values compared to our cotton boll counting correlation. This is a particularly notable achievement since boll density (bolls per plant  $\times$  plant density) is a major driver of environment-induced yield variability, and has been shown to be correlated with genetic yield improvement as well [55]. For example, boll number is the dominant driver of nitrogen and water-induced yield variation in cotton [2, 41, 56], and biotic stressors such as pathogenic nematodes and viruses can substantially

limit boll numbers and yield [5, 10]. Campbell et al. [57] showed that yield improvement in South Carolina Breeding programs was partially associated with an increase in boll production.

In the second aspect of the method, the spatial distribution of cotton bolls on plants was mapped, which allows for the visualization of the vertical (along plant height) and horizontal (relative to the main stem) distributions of cotton bolls. Regarding the vertical distribution of bolls in space, for instance, phenotypic traits (such as boll numbers and retention rates, contributions to yield, and fiber quality) peak at main stem nodes near the middle of the canopy [58]. From the perspective of horizontal distribution along a fruiting branch, fruiting sites closest to the main stem have higher retention rates and the bolls are larger and of better fiber quality than those fruit nodes that are farther away from the main stem [5, 59]. Fruit distribution patterns can also affect the suitability of cotton cultivars for specific environments [60]. For example, in short-season environments, cultivars with more compact fruit distribution patterns (greater fruit set at lower nodes and the majority of bolls distributed over a smaller range of nodes) would be favored. In contrast, cultivars with fruit distributed over a wider range of nodes would be favored in long-season environments to maximize yield potential.

Our methods not only offer a comprehensive approach to cotton boll analysis, providing estimates of boll numbers and insightful visualizations of their spatial distribution, but also could be readily applied to stress physiology research in cotton. For example, biotic and abiotic stressors have been shown to affect boll distribution patterns [5, 61], but the application of plant mapping methods broadly to stress physiology in cotton has been limited due to the time and labor constraints associated with manual mapping. Similarly, genotypic differences in

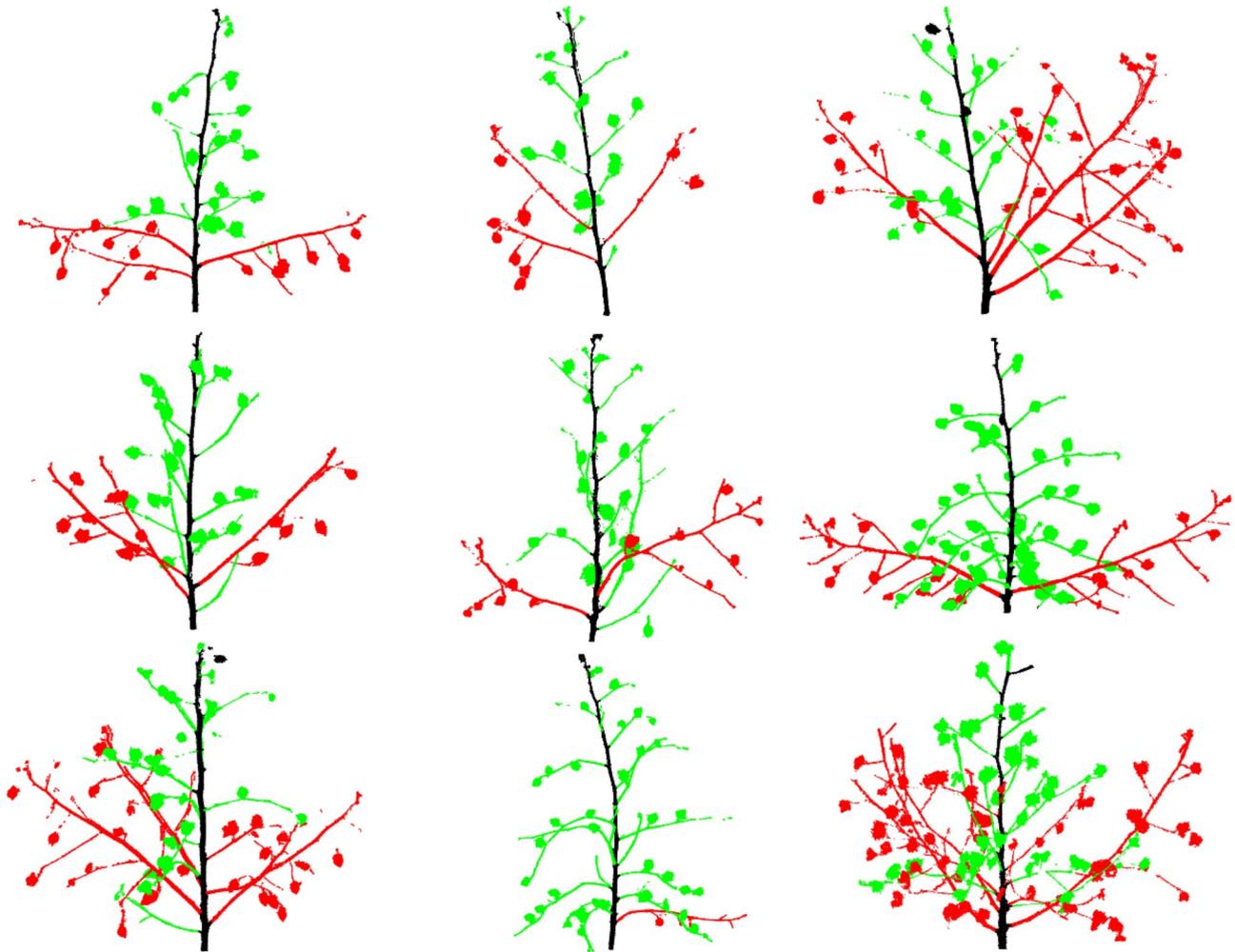


**Fig. 14** Mapping the distribution of cotton bolls relative to the main stem for nine plant samples. Among them, boll 1 represents the boll closest to the main stem on each branch (shown in red); boll 2 represents the next closest boll (shown in green); boll 3 represents the third boll closest to the main stem (shown in blue). Each level was assigned a different color

fruit distribution patterns affect a cultivar’s maturity and yield potential in a given environment [60]. However, the application of plant mapping to cultivar improvement efforts has been limited due to the time constraints associated with traditional hands-on approaches. The methods presented here will significantly advance efforts to alleviate this bottleneck.

From a cotton breeding perspective, the ability to successfully map plant organ and architectural traits could open a new paradigm in yield improvement. While cotton breeders have made steady genetic gains in lint yield, it came at the expense of smaller seed size because selective breeding for new cultivars has relied heavily on driving higher lint percent [62, 63] also called gin turnout, which is the measurement of the weight ratio of lint to seedcotton. The lint percent of modern cultivars has been

increasing in the last three decades, is now approaching 45%, meaning that close to half of the seedcotton by weight is composed of lint fiber. Consequently, seed size being inversely related to lint percent, has become smaller over the same time period and is now negatively impacting germination and seedling vigor [64, 65]. In fact, seedling vigor and stand establishment rank among the top concerns for US cotton producers. The long-term solution to the challenges of lint yield improvement would require targeting other yield component traits that do not contribute to smaller seed size such as increasing the number of cotton bolls per plant by extending the reproductive time period to produce higher numbers of fruiting internodes. However, these plant architecture traits are seldom collected in cotton breeding programs due to added cost and labor requirements. An accurate



**Fig. 15** Visualization of qualitative results of branch type classification from nine examples. The red points are the vegetative branches (VB), the green points are the fruiting branches (FB), and the black points are the main stem

measurement on the spatial and temporal distribution of cotton bolls offers unprecedented opportunity for breeders to select for new improved plant architectural traits.

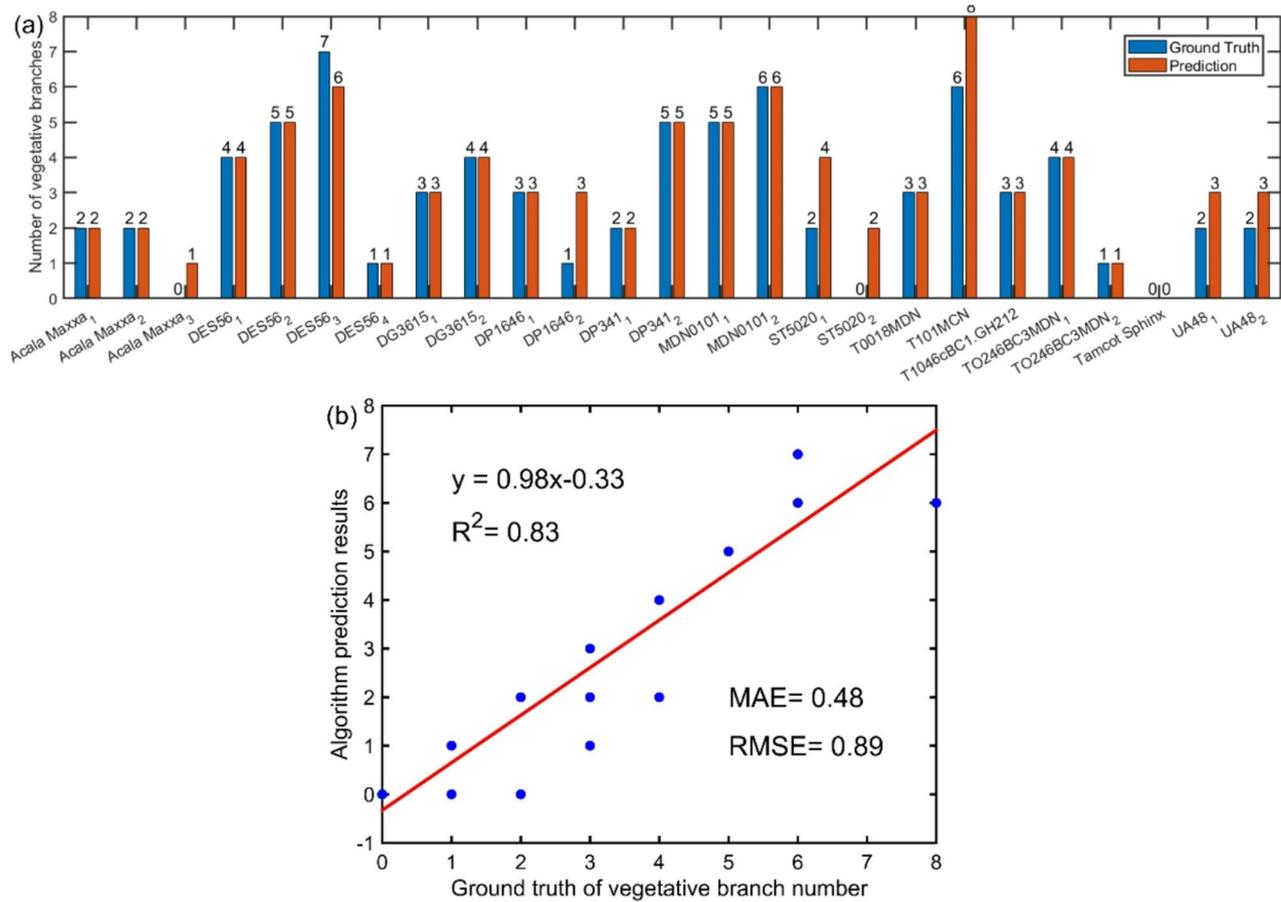
#### Error analysis

The accuracy of cotton boll counting is related to the segmentation results of PointNet++ and Euclidean clustering (Fig. 17). After PointNet++ segments the cotton plants, a small portion of branches may be incorrectly segmented as cotton bolls, as shown in the red boxes in Fig. 17b. Subsequent clustering operations will inevitably group these mis-segmented points into cotton bolls. Although clusters with fewer than 100 points were removed, this does not guarantee that all mis-segmented points are eliminated (Fig. 17a). The threshold for Euclidean clustering is an empirically chosen value that relates to the density of the point cloud and the distribution characteristics of the cotton bolls. Since the point clouds used in this study are dense and the cotton bolls are relatively compact, we selected a relatively small threshold. For the selection of

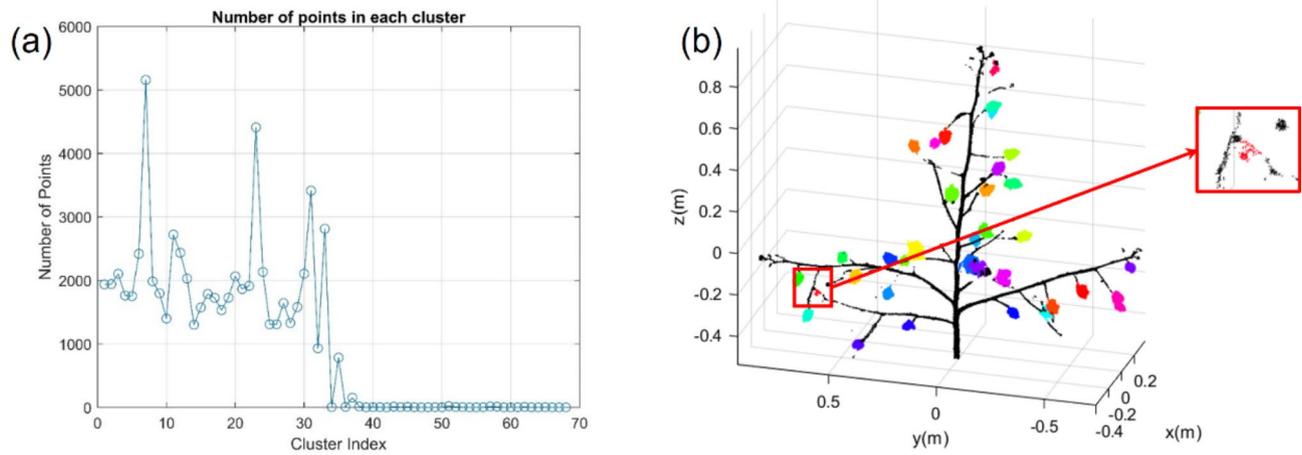
the clustering threshold, this experiment tested three clustering distance thresholds: 0.6 cm, 0.8 cm, and 1.0 cm, to determine the optimal value (Fig. 18). The results indicate that a 0.8 cm threshold yields the most accurate counts. A larger threshold tends to merge multiple bolls into a single cluster, while a smaller threshold would split a single boll into multiple clusters.

Segmentation errors are related to cotton boll yield estimation. When under-segmentation occurs, multiple cotton bolls are merged into a single cluster, which often leads to underestimation of yield. In contrast, over-segmentation may cause a single boll to be mistakenly identified as multiple bolls or non-boll point clouds to be misclassified as bolls, resulting in overestimation of yield. Additionally, the size of the cotton bolls also influences yield estimation.

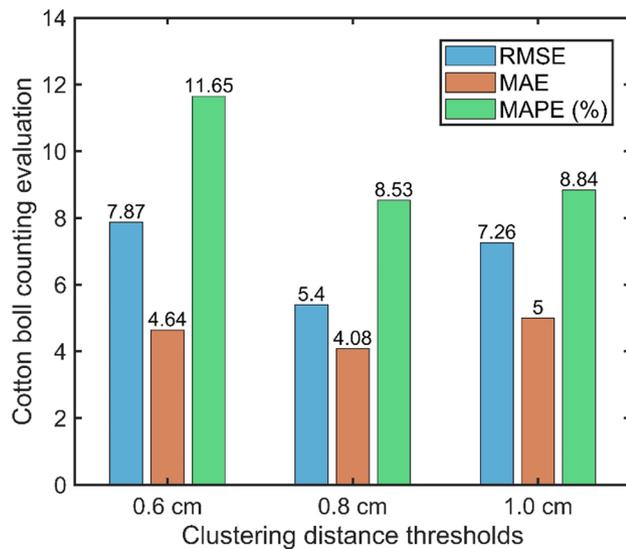
The segmentation performance of TreeQSM directly impacts the distribution of cotton bolls on each branch and the classification of branch types. The main issue with TreeQSM on cotton plants is that a small portion of



**Fig. 16** Quantitative results of branch type classification from 25 samples. (a) Comparison of vegetative branch counting results between the ground truth and predictions of different samples. (b) Linear regression of vegetative branch numbers between the predicted number and ground truth



**Fig. 17** Error analysis of cotton boll counting. (a) The number of points in each cluster after Euclidean clustering. (b) Cotton boll instance segmentation results, with red boxes indicating non-cotton bolls predicted as cotton bolls



**Fig. 18** Effect of different clustering thresholds on cotton boll counting. Three distance thresholds—0.6 cm, 0.8 cm, and 1.0 cm—were compared

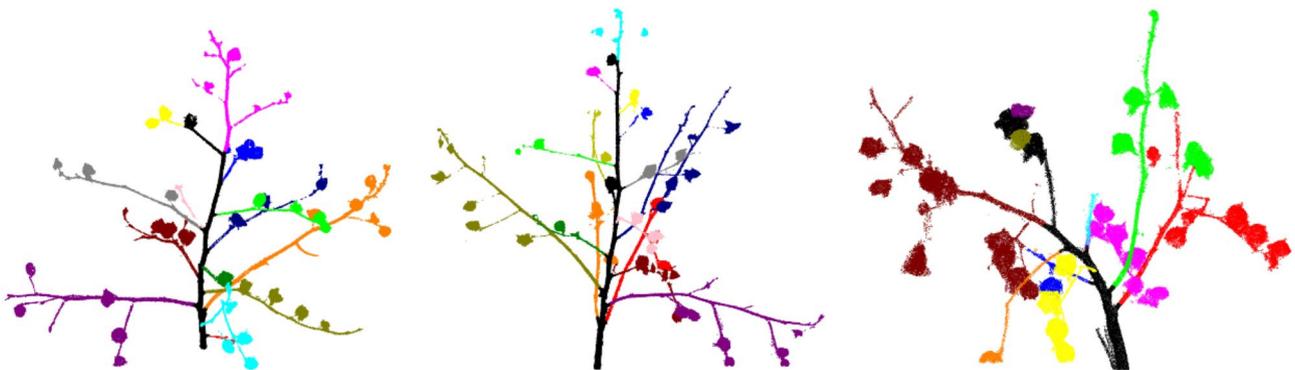
the main stem cannot be fully segmented. In the first two plants shown in Fig. 19, only a small part at the top of the main stem is segmented incorrectly. In contrast, the plant on the far right suffers from significant segmentation issues due to the main stem being excessively curved, resulting in most of the main stem and its branches not being segmented.

#### Advantages and limitations

Compared to previous imaging-based methods for plant phenotyping, our approach offers several advantages. Firstly, it utilizes 3D point cloud data, which helps reduce the occlusion problem encountered when working with 2D images [66, 67]. Point clouds have an additional depth dimension compared to 2D images, which enables them to not only achieve object counting like in 2D images but also provide more 3D traits such as volume, spatial shape, and more. Additionally, the identification of cotton bolls is implemented using the deep learning model

PointNet++, which automatically learns point cloud features, eliminating the need for manual feature engineering common in traditional machine learning approaches. In certain segmentation tasks involving sorghum, grapes, and wheat, extracting specific features like eigenvalues of the local covariance matrix, fast point feature histogram, and principal curvature for different plants requires significant manual effort. However, this manual feature extraction approach falls short in terms of the generalization achieved by deep learning methods [68–70]. In the end, the experiment successfully mapped the distribution of cotton bolls within each branch, which is more refined than the previous analysis of the overall distribution of cotton bolls in the plot [39, 40]. The resulting boll distribution map provides a higher level of detail with organ-level characteristic features.

While the methodology proposed in this study has accomplished the main objectives, there are areas that warrant further refinement. One limitation arises during data processing when specific thresholds need to be set according to cotton plant characteristics, thereby somewhat hindering the pipeline's general applicability. These thresholds include the distance parameters during boll clustering, the size used for generating the cover set in TreeQSM, and the distance threshold employed when classifying branch types. In this research, while the PointNet++ semantic segmentation demonstrated strong performance on the dataset and accurately differentiated cotton bolls from branches, the TreeQSM instance segmentation requires additional enhancements. Current challenges include inaccurate segmentation when dealing with smaller branch ends. Moreover, the clustering process needs refinement since bolls tend to cluster together when they are in contact. Another major limitation of this study is that the method is not an end-to-end model for obtaining results, which reduces computational efficiency. We need to store the segmented point clouds and then perform post-processing steps to achieve cotton boll counting, spatial distribution analysis, and other tasks.



**Fig. 19** Three examples of TreeQSM mis-segmentation. Black points represent the main stem, while each of the remaining branches is assigned a different color

This somewhat limits the overall flexibility of the method. These issues constitute the primary focus of our upcoming research. During the data preprocessing stage, this study manually removed the ground to isolate individual plants, a method that proves inefficient for large-scale datasets. The plan is to develop an automated ground removal algorithm that leverages deep learning techniques to automatically identify and separate the ground from the plant parts, thereby enhancing the efficiency of data processing.

This study has a few other limitations regarding both data collection and the research subject. First, data collection was conducted in an indoor environment, which may limit the model's applicability in other settings. While lighting conditions and background are relatively controlled indoors, outdoor environments pose challenges due to variations in lighting, wind speed, and complex backgrounds, which can interfere with data collection and result in missing plant point clouds or excessive noise.

Second, the LiDAR device (FARO Focus S70) we used was expensive, costing over \$20,000, which made its practical application in plant phenotyping challenging. At the time, some low-cost RGB-D cameras had made contributions to the field of plant phenotyping [71], but the quality of the point clouds they generated was lower compared to those captured by LiDAR. Additionally, emerging high-quality 3D reconstruction methods, such as Neural Radiance Fields (NeRF), have become promising research directions for our future work [72].

Third, the current research focuses on processing point cloud data from individual cotton plants, which limits its applicability to scenarios involving multiple or densely packed plants. Studies of individual plants provide guidance for large-scale breeding programs. In cotton breeding, early-generation screening in the F3 and sometimes F4 generations is typically performed through single-plant selection. Cotton breeders routinely make thousands of observations on individual plants to identify candidates for advancement to progeny rows, representing a major bottleneck in the breeding pipeline. Furthermore, the entire selection process is subjective, relying on personal judgment rather than data-driven approaches. The method we propose enables the accurate capture of detailed architectural and phenotypic traits, which can then be used to objectively cull plants with undesirable traits, such as those possessing a low number of fruiting internodes or failing to retain fruits in the second or third positions on the adaxial branch. While the method is more suited to small-scale breeding programs, addressing the complexities of large-scale breeding requires new approaches capable of processing plot-level data. Large-scale plant analysis, particularly under challenges such as mutual occlusion and intertwined branches, is a key

focus of our future research. Studying individual plants serves as an important foundation, providing a controlled environment to accurately capture architectural and phenotypic traits. This precision is essential for developing and validating effective methods for 3D segmentation, trait measurement, and spatial analysis. By evaluating the accuracy of point cloud segmentation on individual plants, we establish a basis for extending these techniques to broader applications and advancing research in cotton plant phenotyping.

Finally, while the spatial distribution and size of cotton bolls inferred from 3D point cloud data provide valuable insights into yield estimation, this approach has inherent limitations in directly assessing fiber quality. Point cloud data cannot directly provide information about the internal properties of the material, whereas fiber quality attributes—such as length, strength, maturity, and uniformity—require more in-depth microstructural and physical analyses. Estimating boll mass and lint percentage is crucial for breeding, and we plan to measure these parameters in the future.

This approach has its promises as well as limitations. To extend its applicability to high-throughput phenotyping (HTP) to be integrated into cotton breeding programs, several key improvements are needed in the future. First, using our laboratory's self-designed ground robot system, MARS [73], for automated data collection will improve the workflow's scalability for large-scale breeding programs. Second, automating data preprocessing (such as ground removal and plant segmentation) using deep learning techniques will significantly increase processing efficiency and reduce manual intervention. Developing an integrated, end-to-end model based on deep learning will simplify the point cloud segmentation process. The ultimate goal is to transition from individual plant analysis to plot-level breeding applications, enabling large-scale selection of superior varieties.

## Conclusions

This study introduces a novel approach, integrating PointNet++ and TreeQSM, to map cotton bolls not only on the whole plant, but also on individual branches, a feat not previously accomplished. This innovative methodology provides a valuable tool for 3D plant mapping, potentially providing an alternative path to improve cotton yield to complement traditional breeding methods without sacrificing seed size and seeding vigor. A potential limitation arises in the necessity for configuring specific algorithm parameters in accordance with the unique structure of each cotton plant with substantial structural differences. Future work will also involve the exploration of an end-to-end 3D deep learning model with a larger sample size.

### Acknowledgements

This work is very grateful to Javier Rodriguez-Sanchez for data acquisition.

### Author contributions

L.J., C.L. and J.R. conceived the idea and designed the experiments; C.L. and J.S. contributed to the preparation of materials and equipment; J.R. conducted the data collection. L.J. analyzed the data; L.J. interpreted results and wrote the manuscript draft; C.L., J.R., J.S., P.C. and L.F. edited the manuscript. All authors read and approved the final manuscript.

### Funding

This work was partially funded by the United States Department of Agriculture National Institute of Food and Agriculture grant (Award No. 2023-67021-40646), the National Science Foundation Growing Convergence Research (Award No. 1934481), Cotton Incorporated, and the Georgia Cotton Commission.

### Data availability

The code, data, and training weights for cotton boll and branch mapping are available at [https://github.com/UGA-BSAIL/cotton\\_organ\\_mapping.git](https://github.com/UGA-BSAIL/cotton_organ_mapping.git).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 26 July 2024 / Accepted: 8 April 2025

Published online: 20 May 2025

### References

- Zhao H, Chen Y, Liu J, Wang Z, Li F, Ge X. Recent advances and future perspectives in early-maturing cotton research. *New Phytol*. 2023;237:1100–14. <https://doi.org/10.1111/nph.18611>.
- Hu W, Snider JL, Wang H, Zhou Z, Chastain DR, Whitaker J, et al. Water-induced variation in yield and quality can be explained by altered yield component contributions in field-grown cotton. *Field Crops Res*. 2018;224:139–47. <https://doi.org/10.1016/j.fcr.2018.05.013>.
- Virk G, Snider JL, Bourland F. Genotypic and environmental contributions to Lint yield, yield components, and fiber quality in upland cotton from Arkansas variety trials over a 19-year period. *Crop Sci*. 2023. <https://doi.org/10.1002/csc2.20948>.
- Zhang J, Han Y, Li X, Wang G, Wang Z, et al. Inhibition of apical dominance affects boll spatial distribution, yield and fiber quality of field-grown cotton. *Ind Crops Prod*. 2021;173:114098. <https://doi.org/10.1016/j.indcrop.2021.114098>.
- Snider JL, Davis RF, Earl HJ, Timper P. Water availability and root-knot nematode management alter seedcotton yield through similar effects on fruit distribution patterns. *Field Crops Res*. 2019;233:88–95. <https://doi.org/10.1016/j.fcr.2019.01.008>.
- Schaefer C, Nichols B, Collins G, Whitaker J, Bednarz C, Main C, et al. Cotton maturity determination through vertical mapping. *Crop Sci*. 2017;57:62–70. <https://doi.org/10.2135/cropsci2016.03.0168>.
- Nie J, Yuan Y, Qin D, Liu Y, Wang S, Li J, et al. Spatial distribution of Bolls affects yield formation in different genotypes of Bt cotton varieties. *J Integr Agric*. 2019;18:2492–504. [https://doi.org/10.1016/S2095-3119\(19\)62617-1](https://doi.org/10.1016/S2095-3119(19)62617-1).
- Davidonis GH, Johnson AS, Landivar JA, Fernandez CJ. Cotton fiber quality is related to boll location and planting date. *Agron J*. 2004;96:42–7. <https://doi.org/10.2134/agronj2004.4200>.
- Sequeira RA, Cochran M, El-Zik KM, Stone ND, Makela ME. Inclusion of plant structure and fiber quality into a distributed delay cotton model to improve management and optimize profit. *Ecol Model*. 1994;71:161–86. [https://doi.org/10.1016/0304-3800\(94\)90080-9](https://doi.org/10.1016/0304-3800(94)90080-9).
- Parkash V, Sharma DB, Snider J, Bag S, Roberts P, Tabassum A, et al. Effect of cotton leafroll Dwarf virus on physiological processes and yield of individual cotton plants. *Front Plant Sci*. 2021;12:734386. <https://doi.org/10.3389/fpls.2021.734386>.
- Li D, Shi G, Kong W, Wang S, Chen Y. A leaf segmentation and phenotypic feature extraction framework for multiview stereo plant point clouds. *IEEE J Sel Top Appl Earth Observations Remote Sens*. 2020;13:2321–36. <https://doi.org/10.1109/JSTARS.2020.2989918>.
- Jimenez-Berni JA, Deery DM, Rozas-Larraondo P, Condon ATG, Rebetzke GJ, James RA, et al. High throughput determination of plant height, ground cover, and above-ground biomass in wheat with lidar. *Front Plant Sci*. 2018;9:1–18. <https://doi.org/10.3389/fpls.2018.00237>.
- Sun S, Li C, Paterson AH, Jiang Y, Xu R, Robertson JS, et al. In-field high throughput phenotyping and cotton plant growth analysis using lidar. *Front Plant Sci*. 2018;9:1–17. <https://doi.org/10.3389/fpls.2018.00016>.
- Vicari MB, Pisek J, Disney M. New estimates of leaf angle distribution from terrestrial lidar: comparison with measured and modelled estimates from nine broadleaf tree species. *Agric For Meteorol*. 2019;264:322–33. <https://doi.org/10.1016/j.agrformet.2018.10.021>.
- Xiang L, Bao Y, Tang L, Ortiz D, Salas-Fernandez MG. Automated morphological traits extraction for sorghum plants via 3D point cloud data analysis. *Comput Electron Agric*. 2019;162:951–61. <https://doi.org/10.1016/j.compag.2019.05.043>.
- Miao T, Zhu C, Xu T, Yang T, Li N, Zhou Y, et al. Automatic stem-leaf segmentation of maize shoots using three-dimensional point cloud. *Comput Electron Agric*. 2021;187:106310. <https://doi.org/10.1016/j.compag.2021.106310>.
- Yu T, Hu C, Xie Y, Liu J, Li P. Mature pomegranate fruit detection and location combining improved F-PointNet with 3D point cloud clustering in orchard. *Comput Electron Agric*. 2022;200:107233. <https://doi.org/10.1016/j.compag.2022.107233>.
- Gené-Mola J, Gregorio E, Auat Cheein F, Guevara J, Llorens J, Sanz-Cortierra R, et al. Fruit detection, yield prediction and canopy geometric characterization using lidar with forced air flow. *Comput Electron Agric*. 2020;168:105121. <https://doi.org/10.1016/j.compag.2019.105121>.
- Jiang L, Li C, Sun J, Chee P, Fu L. Estimation of cotton boll number and main stem length based on 3D Gaussian splatting written for presentation at the 2024 ASABE annual international meeting sponsored by ASABE. 2024;1–11. <https://doi.org/10.13031/aim.202400898>.
- Jiang L, Li C, Fu L. 3D deep Learning-based segmentation to reveal the spatial distribution of cotton Bolls. 2022 ASABE annual international meeting. American Society of Agricultural and Biological Engineers; 2022. p. 1. <https://doi.org/10.13031/aim.202200361>.
- Ni X, Li C, Jiang H, Takeda F. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. *Hortic Res*. 2020;7. <https://doi.org/10.1038/s41438-020-0323-3>.
- Shen J, Wu T, Zhao J, Wu Z, Huang Y, Gao P, et al. Organ segmentation and phenotypic trait extraction of cotton seedling point clouds based on a 3D lightweight network. *Agronomy*. 2024;14:1–19. <https://doi.org/10.3390/agronomy14051083>.
- Hao H, Wu S, Li YK, Wen W, Fan J, Zhang Y, et al. Automatic acquisition, analysis and wilting measurement of cotton 3D phenotype based on point cloud. *Biosyst Eng*. 2024;239:173–89. <https://doi.org/10.1016/j.biosystemseng.2024.02.010>.
- Yan J, Tan F, Li C, Jin S, Zhang C, Gao P, et al. Stem–Leaf segmentation and phenotypic trait extraction of individual plant using a precise and efficient point cloud segmentation network. *Comput Electron Agric*. 2024;220:108839. <https://doi.org/10.1016/j.compag.2024.108839>.
- Zeng L, Feng J, He L. Semantic segmentation of sparse 3D point cloud based on geometrical features for trellis-structured Apple orchard. *Biosyst Eng*. 2020;196:46–55. <https://doi.org/10.1016/j.biosystemseng.2020.05.015>.
- Raunonen P, Kaasalainen M, Markku Å, Kaasalainen S, Kaartinen H, Vastaranta M, et al. Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sens*. 2013;5:491–520. <https://doi.org/10.3390/rs5020491>.
- Zhang C, Jiang Y, Xu B, Li X, Zhu Y, Lei L, et al. Apple tree branch information extraction from terrestrial laser scanning and backpack-LiDAR. *Remote Sens*. 2020;12:1–17. <https://doi.org/10.3390/rs1213592>.
- Xu B, Wan X, Yang H, Feng H, Fu Y, Cen H, et al. TIPS: A three-dimensional phenotypic measurement system for individual maize Tassel based on TreeQSM. *Comput Electron Agric*. 2023;212. <https://doi.org/10.1016/j.compag.2023.108150>.
- Charles RQ, Su H, Kaichun M, Guibas LJ, PointNet. Deep Learning on Point Sets for 3D Classification and Segmentation. 2017 IEEE Conference on

- Computer Vision and Pattern Recognition (CVPR). 2017:77–85. <https://doi.org/10.1109/CVPR.2017.16>
30. Li Y, Wen W, Miao T, Wu S, Yu Z, Wang X, et al. Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning. *Comput Electron Agric.* 2022;193:106702. <https://doi.org/10.1016/j.compag.2022.106702>.
  31. Li Y, Bu R, Di X. PointCNN. Convolution On X-Transformed Points. in: *NeurIPS, Proceedings of the Advances in Neural Information Processing Systems.* 2018;Montréal:820–30. <https://doi.org/10.48550/arXiv.1801.07791>
  32. Ao Z, Wu F, Hu S, Sun Y, Su Y, Guo Q, et al. Automatic segmentation of stem and leaf components and individual maize plants in field terrestrial lidar data using convolutional neural networks. *Crop J.* 2021. <https://doi.org/10.1016/j.cj.2021.10.010>.
  33. Li D, Shi G, Li J, Chen Y, Zhang S, Xiang S, et al. PlantNet: A dual-function point cloud segmentation network for multiple plant species. *ISPRS J Photogrammetry Remote Sens.* 2022;184:243–63. <https://doi.org/10.1016/j.isprsjprs.2022.01.007>.
  34. Du R, Ma Z, Xie P, He Y, Cen H. PST: plant segmentation transformer for 3D point clouds of rapeseed plants at the podding stage. *ISPRS J Photogrammetry Remote Sens.* 2023;195:380–92. <https://doi.org/10.1016/j.isprsjprs.2022.11.022>.
  35. Gené-Mola J, Sanz-Cortiella R, Rosell-Polo JR, Morros JR, Ruiz-Hidalgo J, Vilaplana V, et al. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Comput Electron Agric.* 2020;169:105165. <https://doi.org/10.1016/j.compag.2019.105165>.
  36. Underwood JP, Hung C, Whelan B, Sukkarieh S. Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. *Comput Electron Agric.* 2016;130:83–96. <https://doi.org/10.1016/j.compag.2016.09.014>.
  37. Adhikari J, Petti D, Ployaram W, Li C, Paterson AH. Characterizing Season-Long floral trajectories in cotton with Low-Altitude remote sensing and deep learning. *Plants People Planet.* 2025. <https://doi.org/10.1002/PPP3.10644>.
  38. Walter JDC, Edwards J, McDonald G, Kuchel H. Estimating biomass and canopy height with lidar for field crop breeding. *Front Plant Sci.* 2019;10:1145. <https://doi.org/10.3389/fpls.2019.01145>.
  39. Dube N, Bryant B, Sari-Saraf H, Ritchie GL. Cotton boll distribution and yield Estimation using three-dimensional point cloud data. *Agron J.* 2020;112:4976–89. <https://doi.org/10.1002/agj2.20412>.
  40. Sun S, Li C, Chee PW, Paterson AH, Jiang Y, Xu R, et al. Three-dimensional photogrammetric mapping of cotton Bolls in situ based on point cloud segmentation and clustering. *ISPRS J Photogrammetry Remote Sens.* 2020;160:195–207. <https://doi.org/10.1016/j.isprsjprs.2019.12.011>.
  41. Sun S, Li C, Chee PW, Paterson AH, Meng C, Zhang J, et al. High resolution 3D terrestrial lidar for cotton plant main stalk and node detection. *Comput Electron Agric.* 2021;187:106276. <https://doi.org/10.1016/j.compag.2021.106276>.
  42. Liu Z, Tang H, Lin Y, Han S. Point-voxel Cnn for efficient 3d deep learning. *Adv Neural Inf Process Syst.* 2019;32. <https://doi.org/10.48550/arXiv.1907.03739>.
  43. Saeed F, Sun S, Sanchez JR, Snider J, Liu T, Li C. Cotton plant part 3D segmentation and architectural trait extraction using point voxel convolutional neural networks. *Plant Methods.* 2023;19:1–23. <https://doi.org/10.1186/s13007-023-00996-1>.
  44. Chang AX, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z et al. ShapeNet: An Information-Rich 3D Model Repository. 2015 <http://arxiv.org/abs/1512.03012>
  45. Saeed F, Li C. Plant organ segmentation from point clouds using Point-Voxel CNN. American Society of Agricultural and Biological Engineers Annual International Meeting, ASABE 2021. American Society of Agricultural and Biological Engineers. 2021:1316–26. <https://doi.org/10.13031/aim.202100428>
  46. Qi C, Yi L, Su H, Guibas L, PointNet++. Deep Hierarchical Feature Learning on NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017;2017-Decem:5105–5114. <https://doi.org/10.48550/arXiv.1706.02413>
  47. Zhao C, Yang J, Xiong X, Zhu A, Cao Z, Li X. Rotation invariant point cloud analysis: where local geometry Meets global topology. *Pattern Recogn.* 2022;127:108626. <https://doi.org/10.1016/j.patcog.2022.108626>.
  48. Wu X, Lao Y, Jiang L, Liu X, Zhao H. Point transformer V2: grouped vector attention and Partition-based pooling. *Adv Neural Inf Process Syst.* 2022;35:1–13. <https://doi.org/10.48550/arXiv.2210.05666>.
  49. Yang X, Miao T, Tian X, Wang D, Zhao J, Lin L, et al. Maize stem-leaf segmentation framework based on deformable point clouds. *ISPRS J Photogrammetry Remote Sens.* 2024;211:49–66. <https://doi.org/10.1016/j.isprsjprs.2024.03.025>.
  50. Xin B, Sun J, Bartholomeus H, Kootstra G. 3D data-augmentation methods for semantic segmentation of tomato plant parts. *Front Plant Sci.* 2023;14:1–17. <https://doi.org/10.3389/fpls.2023.1045545>.
  51. Wang H, Yu Y, Yuan Q. Application of Dijkstra algorithm in robot path-planning. 2011 Second International Conference on Mechanic Automation and Control Engineering. 2011:1067–9. <https://doi.org/10.1109/MACE.2011.5987118>
  52. Bednarz CW, Nichols RL, Brown SM. Plant density modifications of cotton within-boll yield components. *Crop Sci.* 2006;46:2076–80. <https://doi.org/10.2135/cropsci2005.12.0493>.
  53. Zhai M, Wei X, Pan Z, Xu Q, Qin D, Li J, et al. Optimizing plant density and canopy structure to improve light use efficiency and cotton productivity: two years of field evidence from two locations. *Ind Crops Prod.* 2024;222. <https://doi.org/10.1016/j.indcrop.2024.119946>.
  54. Yang R, He Y, Lu X, Zhao Y, Li Y, Yang Y, et al. 3D-based precise evaluation pipeline for maize ear rot using multi-view stereo reconstruction and point cloud semantic segmentation. *Comput Electron Agric.* 2024;216:108512. <https://doi.org/10.1016/j.compag.2023.108512>.
  55. Li X, Lei Y, Han Y, Wang Z, Wang G, Feng L, et al. The relative impacts of plant density and weather on cotton yield variability. *Field Crops Res.* 2021;270:108202. <https://doi.org/10.1016/j.fcr.2021.108202>.
  56. Ermanis A, Gobbo S, Snider JL, Cohen Y, Liakos V, Lacerda L, et al. Defining physiological contributions to yield loss in response to irrigation in cotton. *J Agron Crop Sci.* 2021;207:186–96. <https://doi.org/10.1111/jac.12453>.
  57. Campbell BT, Chee PW, Lubbers E, Bowman DT, Meredith WR, Johnson J, et al. Genetic improvement of the pee Dee cotton germplasm collection following Seventy years of plant breeding. *Crop Sci.* 2011;51:955–68. <https://doi.org/10.2135/cropsci2010.09.0545>.
  58. NIE J, jun YUANY, chao QIND, lin LIUY, hui WANGS, lei LIJ, pu, et al. Spatial distribution of Bolls affects yield formation in different genotypes of Bt cotton varieties. *J Integr Agric.* 2019;18:2492–504. [https://doi.org/10.1016/S2095-3119\(19\)62617-1](https://doi.org/10.1016/S2095-3119(19)62617-1).
  59. Bednarz CW, Nichols RL, Brown SM. Within-boll yield components of high yielding cotton cultivars. *Crop Sci.* 2007;47:2108–12. <https://doi.org/10.2135/cropsci2006.12.0827>.
  60. Constable GA, Bange MP. The yield potential of cotton (*Gossypium hirsutum* L.). *Field Crops Res.* 2015;182:98–106. <https://doi.org/10.1016/j.fcr.2015.07.017>.
  61. Snowden C, Ritchie G, Cave J, Keeling W, Rajan N. Multiple irrigation levels affect boll distribution, yield, and fiber micronaire in cotton. *Agron J.* 2013;105:1536–44. <https://doi.org/10.2134/agronj2013.0084>.
  62. Conaty WC, Constable GA. Factors responsible for yield improvement in new *Gossypium hirsutum* L. cotton cultivars. *Field Crops Res.* 2020;250:107780. <https://doi.org/10.1016/j.fcr.2020.107780>.
  63. Singh J, Gamble AV, Brown S, Campbell BT, Jenkins J, Koebernick J, et al. 65 Years of cotton Lint yield progress in the USA: Uncovering key influential yield components. *Field Crops Res.* 2023;302:109058. <https://doi.org/10.1016/j.fcr.2023.109058>.
  64. Snider JL, Collins GD, Whitaker J, Chapman KD, Horn P. The impact of seed size and chemical composition on seedling vigor, yield, and fiber quality of cotton in five production environments. *Field Crops Res.* 2016;193:186–95. <https://doi.org/10.1016/j.fcr.2016.05.002>.
  65. Bourland FM. Functional characterization of seed and seedling vigor in cotton. *J Cotton Sci.* 2019;23:168–76. <https://doi.org/10.56454/fxqj8103>.
  66. Sun S, Li C, Paterson AH, Chee PW, Robertson JS. Image processing algorithms for infield single cotton boll counting and yield prediction. *Comput Electron Agric.* 2019;166:104976. <https://doi.org/10.1016/j.compag.2019.104976>.
  67. Liu Q, Zhang Y, Yang G. Small unopened cotton boll counting by detection with MRF-YOLO in the wild. *Comput Electron Agric.* 2023;204:107576. <https://doi.org/10.1016/j.compag.2022.107576>.
  68. Wahabzada M, Paulus S, Kersting K, Mahlein AK. Automated interpretation of 3D laserscanned point clouds for plant organ segmentation. *BMC Bioinform.* 2015;16:1–11. <https://doi.org/10.1186/s12859-015-0665-2>.
  69. Sodhi P, Vijayarangan S, Wettergreen D. In-field segmentation and identification of plant structures using 3D imaging. *IEEE Int Conf Intell Robots Syst.* 2017;2017-Sept:5180–7. <https://doi.org/10.1109/IROS.2017.8206407>.
  70. Ziamtsov I, Navlakha S. Machine learning approaches to improve three basic plant phenotyping tasks using three-dimensional point clouds. *Plant Physiol.* 2019;181:1425–40. <https://doi.org/10.1104/pp.19.00524>.

71. Jiang L, Li C, Fu L. Apple tree architectural trait phenotyping with organ-level instance segmentation from point cloud. *Comput Electron Agric.* 2025;229:109708. <https://doi.org/10.1016/j.compag.2024.109708>.
72. Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R, Nerf. Representing scenes as neural radiance fields for view synthesis. *Commun ACM.* 2021;65:99–106. <https://doi.org/10.1145/350325>.
73. Xu R, Li C. A modular agricultural robotic system (MARS) for precision farming: concept and implementation. *J Field Robot.* 2022;39:387–409. <https://doi.org/10.1002/rob.22056>.

### **Publisher's note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.