



SCAG: A stratified, clustered, and growing-based algorithm for soybean branch angle extraction and ideal plant architecture evaluation

Songyin Zhang¹, Yinmeng Song^{1,2}, Ran Ou^{1,2}, Yiqiang Liu^{1,3}, Shaochen Li¹, Xinlan Lu¹
Shan Xu¹, Yanjun Su⁴, Jiang Dong^{1,2,3}, Yanfeng Ding^{1,2,3}, Haifeng Xia⁵, Qinghua Guo⁶, Jin
Wu⁷, Jiaoping Zhang^{1,2}, Shichao Jin^{1,3} *

¹ Plant Phenomics Research Centre, Academy for Advanced Interdisciplinary Studies, Collaborative Innovation Centre for Modern Crop Production cosponsored by Province and Ministry, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China

² National Center for Soybean Improvement, Key Laboratory for Biology and Genetic Improvement of Soybean (General, Ministry of Agriculture), College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China

³ Sanya Research Institute of Nanjing Agriculture University, Sanya 572024, China

⁴ State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

⁵ School of Automation, Southeast University, Nanjing 210096, China

⁶ Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871, China

⁷ Division for Ecology and Biodiversity, School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, China

*** Corresponding authors**

jinshichao1993@gmail.com; jschaon@njau.edu.cn

Abstract: Three-dimensional (3D) phenotyping is important for studying plant structure and function. Light detection and ranging (LiDAR) has gained prominence in 3D plant phenotyping due to its ability to collect 3D point clouds. However, organ-level branch detection remains challenging due to small targets, sparse points, and low signal-to-noise ratios. In addition, extracting biologically relevant angle traits is difficult. In this study, we developed a stratified, clustered, and growing-based algorithm (SCAG) for soybean branch detection and branch angle calculation from LiDAR data, which is heuristic, open-source, and expandable. SCAG achieved high branch detection accuracy ($F\text{-score}=0.77$) and branch angle calculation accuracy ($r=0.84$) when evaluated on 152 diverse soybean varieties. Meanwhile, the SCAG outperformed two other classic algorithms, the support vector machine ($F\text{-score}=0.53$) and density-based methods ($F\text{-score}=0.55$). Moreover, after applying the SCAG to 405 soybean varieties over two consecutive years, we quantified various 3D traits, including canopy width, height, stem length, and average angle. After data filtering, we identified novel heritable and repeatable traits for evaluating soybean density tolerance potential, such as the average angle to height (AHR) and the ratio of average angle to stem length (ALR), which showed greater potential than the well-known ratio of canopy width to height (CHR) trait. Our work demonstrates significant advances in 3D phenotyping and plant architecture screening. The algorithm can be applied to other crops, such as maize and tomato. Our dataset, scripts, and software are public, which can further benefit the plant science community by enhancing plant architecture characterization and ideal variety selection.

Keywords: Soybean; LiDAR; branch angle calculation; phenotype; ideal plant type.

Introduction

Soybean is a major source of plant oil and protein used for human consumption. The burgeoning global population demands a significant increase in soybean production (Guo et al., 2022). Increasing soybean productivity is, therefore, a long-term breeding demand. Because the area of arable cropland is limited, increasing crop productivity per unit of land area is a widely recognized breeding target. Therefore, it is urgent to screen density-tolerant soybeans, discover new genes controlling plant architecture, and design/breed elite density-tolerant soybeans.

Plant architecture is a comprehensive trait that characterizes crop production potential and is widely recognized as the main direction of high-yield breeding (Gao et al., 2017). Plant architecture refers to the three-dimensional (3D) organization of a plant, including stem height, branching pattern, and the shape and position of leaves and reproductive organs (Huang, X et al., 2021). The success of the “Green Revolution” for rice and wheat inspired us to improve yield by designing a semidwarf plant architecture. Unlike grain crops that have panicles/heads above the canopy, soybean is a typical pod crop whose yield organ grows throughout the plant (Liu et al., 2020). With respect to the “Green Revolution” for soybean plants, one key aspect of plant architecture is branch formation, particularly the branch angle, which directly determines canopy coverage and further influences plant functions, especially light interception, photosynthetic efficiency, and yield (Takahashi et al., 2021; Clark et al., 2022). Smaller branch angles contribute to narrow architecture, enabling dense planting (Tian et al., 2019; Liu et al., 2020; You & Sun, 2022). More importantly, small (vertical) angles at the top of the plant and large (horizontal) angles at the bottom can maximize the light interception of an individual plant and optimize the light distribution and utilization of a crop population, forming a “smart canopy” (Mantilla-Perez et al., 2020) for high-yield production. Therefore, the branch angle is vitally important for evaluating soybean density tolerance and yield potential.

High-throughput branch phenotyping is one of the key challenges for 3D plant architecture studies. The traditional methods for branch angle measurement mainly rely on manual operations, such as the use of inclinometers (Gratani & Ghia, 2002), which are tedious and difficult to perform. Two-dimensional (2D) image-based techniques have

improved the efficiency and precision of branch angle acquisition (Abdel-Khalek et al., 2017). For example, image-based leaf angle extractors were developed to improve the efficiency of leaf angle extraction in maize and sorghum (Kenchanmane Raju et al., 2020). However, these methods are mainly focused on crops (e.g., maize and sorghum) with simple canopy structures (e.g., symmetrical distribution of leaves) and require specifically selected shooting angles. In contrast, soybean crops have more complex 3D plant architecture, thus making it difficult to determine an optimal shooting angle.

LiDAR is an active remote sensing technique used to directly acquire the 3D information of an object, shedding new light for 3D plant phenomics (Jin et al., 2021b; Tao et al., 2022). LiDAR is superior to other 3D reconstruction techniques, such as stereo imagers (Yidan et al., 2016) and depth cameras (Xiong et al., 2017), which are easily affected by environmental conditions (e.g., illumination) (Ma & Liu, 2018; Ghahremani et al., 2021; Ingale & J, 2021), object surface properties (e.g., continuity and flatness) (Nuijten et al., 2019), and camera calibration (Huang, K et al., 2021; Tsai et al., 2021; Zhang et al., 2022). LiDAR has been widely applied to phenotypic tasks due to its high accuracy and insensitivity to environmental light (Guo et al., 2018; Jin et al., 2020a; Jayakumari et al., 2021; Li, Q et al., 2022; Sun et al., 2022; Zang et al., 2023). However, LiDAR point cloud-based methods have been mainly developed for crops with simple structures, such as for calculating the leaf angles in maize and sorghum (Thapa et al., 2018; Jin et al., 2019; Su et al., 2019; Wu et al., 2019; Jin et al., 2021a). These methods can be categorized into two major types: geometric methods and machine learning-based methods. Geometric methods usually use the geometric distance between the samples in the point clouds (Jin et al., 2019) and the curvature calculation method (Su et al., 2019) for skeleton extraction and trait calculation. Notably, geometric methods mostly rely on handcrafted features and expert experiences. In contrast, machine learning-based methods automatically learn features based on given abundant data samples and their corresponding labels. The labeled key points of plants allow for machines to learn the rules for expected targets and automatically recognize plant phenotypic traits (Dutagaci et al., 2020). However, it is difficult to interpret machine learning-based methods and apply them to other scenarios (e.g., crop types). Therefore, geometric algorithms are usually selected by

the research community if effective features and rules can be designed and large-volume datasets are unavailable.

Branch detection algorithms are critical for extracting complex 3D branch angles and empowering further 3D phenomics applications. Skeleton extraction is the most frequently used method for branch angle detection based on point clouds (Zhang et al., 2020). The Laplacian contraction algorithm has been widely used in skeleton extraction by shrinking point clouds, such as in maize (Wu et al., 2019). In addition, several heuristic methods have been developed (Jin et al., 2019; Su et al., 2019). The MNVG (median normalized-vector growth) method was proposed for generating the skeleton by calculating the median points of a ball region from the bottom upward (Jin et al., 2019). Similarly, the curvature calculation method was developed for extracting the skeleton of maize by finding points with large curvatures (Su et al., 2019). Notably, skeleton-based methods are effective for large crops with simple canopy structures. However, these methods are not applicable to soybean plants due to various challenges. 1) *Complex 3D structure*: The branches of soybean plants are distributed throughout 3D space, thus making their structure more complex than that of crops with leaves almost symmetrically distributed along a main stem, such as maize and sorghum; 2) *Fewer curved branching organs*: Plant architecture is closely related to branching organs. Compared with the flexible leaves of maize and sorghum, the branches of soybean plants are wood-like with no obvious curvature. 3) *Smaller crop size*: The average height, stem width, and branch width of soybean are smaller than those of other crops, such as maize, thus leading to a smaller signal-to-noise ratio in soybean point cloud data under the same scanning specifications.

The main objectives of this study are presented below.

1) To develop a new soybean branch detection algorithm, i.e., the stratified, clustered, and growing-based (SCAG) algorithm, based on LiDAR data.

2) To evaluate the performance and robustness of the proposed algorithm through model comparison and parameter sensitivity analysis.

3) To verify the applicability and transferability of the proposed algorithm to other crop types and its value in evaluating the potential of soybean density tolerance based on plant architecture (e.g., branch angles).

The proposed SCAG algorithm could facilitate high-throughput soybean branch angle extraction and accelerate branch angle-based plant architecture screening, which is of great interest to the breeding community.

Materials and Methods

Study area and data collection

The study area was located at the Baima Experimental Station of Nanjing Agricultural University in Nanjing, Jiangsu Province, China (119°18'71"E, 31°62'00"N). A two-year germplasm screening experiment was repeatedly conducted in both 2021 and 2022. The length and width of the experimental field were approximately 200 m and 30 m, respectively. There were 810 plots in the field that included two replications of 405 soybean varieties (Fig. 1). Each plot had three plant rows with a row spacing of 0.7 m and a row length of 1 m. The distance between adjacent plants within a row was approximately 0.07 m. Two border rows were planted around the target plots for protection. These varieties have significant diversity and representativeness in terms of plant architecture and were selected from all over the world by the National Center for Soybean Improvement (<https://ncsi.njau.edu.cn/sysgk1/sysjj.htm>).

Although we aimed to assess soybean plant architecture in real production conditions, it was difficult to acquire high-quality individual point clouds in a changeable field environment with a dense canopy. Therefore, three plants of each variety per plot were randomly selected from the field and transferred to an indoor platform for scanning to ensure data accuracy during the maturity stage. The leaves of plants at the maturity stage were manually defoliated before scanning in 2021, and the leaves and pods were removed before scanning in 2022. The soybean individuals of different varieties were fixed in rows with a row spacing of approximately 0.7 m and a plant spacing of approximately 1 m on a platform with a length of 23 m and a width of 3.5 m (Fig. 1 b). The point clouds of these varieties are scanned by using a terrestrial LiDAR scanner (TLS, Faro Focus3D S70) at multiple stations. The scanning time of each station was approximately 4 minutes, and the point cloud resolution was approximately 2 mm. There were 10-14 stations in the platform area, which were distributed along the edge of the platform, and the distance between two

stations was approximately 3.7 m (Fig. 1 c). Finally, these multiple stations of TLS data were registered into a whole point cloud using SCENE software.

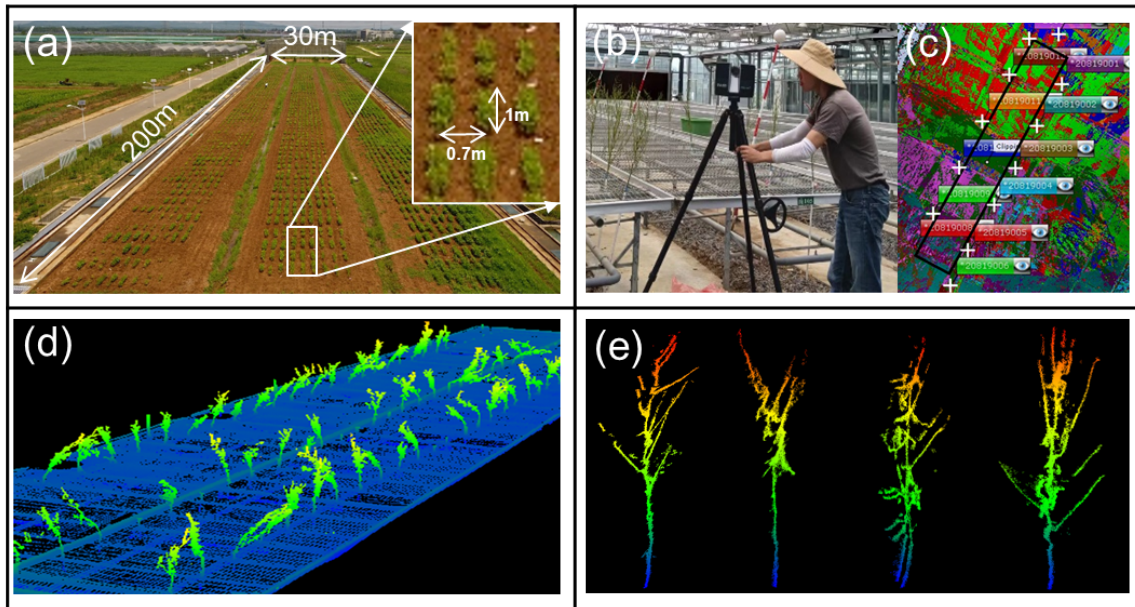


Fig. 1. The study area and process of data collection. (a) The experimental field was approximately 200 m by 30 m. Each plot consisted of three rows of the same soybean variety, with a row width of 1 m and a row spacing of 0.7 m. The field contained 810 plots, which included two replications of 405 soybean varieties. (b) Soybean plants were fixed on an indoor platform and scanned by using a TLS. (c) Point cloud data were registered using the SCENE software, where the black box is the platform and the white "+" represent the TLS locations. (d) An overview of the collected LiDAR data on October 22, 2021. (e) The point cloud data of several representative soybean varieties.

Data preprocessing and dataset construction

In both 2021 and 2022, the registered LiDAR data were manually segmented into 3218 individual point clouds of 405 varieties (each variety had 1-2 individual samples) and denoised using the statistical outlier removal (SOR) algorithm in *CloudCompare* software. A total of 152 individual samples were randomly selected to construct a well-labeled dataset (*Soybean3D*) for algorithm development (see details in Supplemental Materials S1). The *Soybean3D* dataset is diverse (Fig. S1) and can be further divided into three different complex groups based on the number of branch angles and average branch angle (Table S1).

Algorithm design of the SCAG

This work presents an SCAG framework for branch angle detection and calculation. The SCAG method includes three steps, i.e., branch detection, node location, and branch point optimization (Fig. 2). The branch detection step was designed to coarsely locate the branches from a series of point clouds of a slice height (H) from the bottom to the top. The node point location step aims to find node points (connection points between stems and branches) by using the downward region growth method based on the above-detected branch points. Branch point optimization is a postprocessing step for optimizing the above-detected branch locations that may be too close to the node points. The angle calculation uncertainty can be decreased based on the finely optimized branch points and node points.

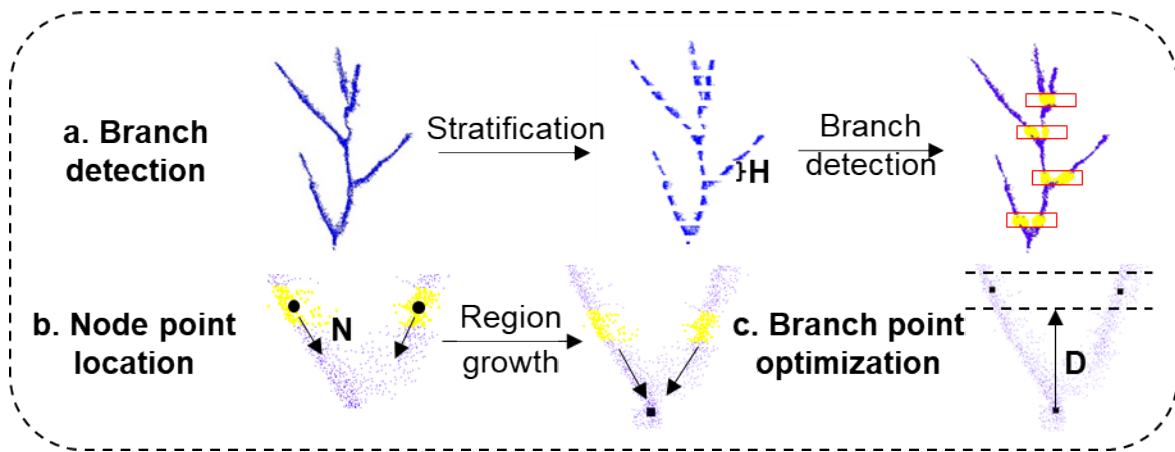


Fig. 2. An overview of the proposed method. (a) Branch detection: The soybean point cloud was sliced into multiple layers of the same height. The branch locations were detected based on the cluster relationships of adjacent layers. (b) Node point location: The node point was located by using the downward region growth method based on the branch points detected in the previous step. (c) Branch point optimization: Two branching points were relocated by slicing a new upward layer that was closer/similar to manual operations from the node point. The optimized branch points and node points were used to calculate branch angles.

Step 1: Branch detection

Branch detection aims to coarsely find the locations of the branches for branch angle calculation, which requires finding two branch points and one node point. The parameter slice height (H) was used to slice a point cloud into different layers from the bottom to the top because the branches were located at different heights. The point cloud of each layer was clustered by using density-based spatial clustering of applications with noise (DBSCAN) (Schubert et al., 2017). The branch locations were detected by analyzing the

number of clusters and the relationship between any two adjacent layers. The number of clusters in each layer and its next bottom layer mainly included three cases.

Case 1: The number of clusters in the upper slice layer (CNU) was greater than the number of clusters in the lower slice layer (CNL), i.e., $CNU > CNL$.

This was the simplest case, i.e., the number of clusters increased from the bottom to the top. This case meant that a new branch was formed, as presented in Fig. 3. b.1. Therefore, the clusters in the CNU were preserved for locating branches.

Case 2: The number of clusters in the upper slice equaled the number of clusters in the lower slice, i.e., $CNU = CNL$.

This case included two possible results: (I) the upper layer contained a newly split branch (Fig. 3. b.2.I), which should be retained; and (II) the upper layer did not contain a newly split branch but rather a continuation of the previous branch (Fig. 3. b.2.II), which should be ignored.

To determine if a layer contained a new branch, a distance-based rule was proposed by considering the distances (i.e., $D_{CNL} = [d_1, \dots, d_{CNL}]$) between each cluster in the lower layer and its nearest cluster in the upper layer. Each cluster was represented by a median point calculated from points belonging to the cluster. We used the median value instead of the mean value of the point cloud as the cluster center because our previous findings indicated that the median value is insensitive to noisy points (Jin et al., 2019). If one of the clusters in the lower layer did not have a neighboring cluster within the given distance threshold in the upper layer, then the upper layer contained new branches (Fig. 3. b.2.I); otherwise, if each cluster in the lower layer had neighboring clusters within the given distance threshold in the upper layer, the upper layer did not contain a new branch but a continuation of a lower branch (Fig. 3. b.2.II). This biological observation can be formulated into a distance-based rule, i.e., if the maximum value of the nearest distances (i.e., $\max(D_{CNL})$) is greater than twice the minimum value (i.e., $\min(D_{CNL})$), the upper layer should be retained for locating the new branch. Otherwise, the upper layer was considered a continuation of the bottom branch and was discarded. The “twice” threshold was selected according to repeated trials and errors.

Case 3: The number of clusters in the upper slice is smaller than the number of clusters in the lower slice, i.e., $CNU < CNL$.

This case also includes two possible results: (I) the upper layer contained a new branch (Fig. 3. b.3.I), which should be retained; (II) the upper layer did not contain a newly split branch but rather a continuation of several previous branches (Fig. 3. b.3.II), which should be ignored.

To determine if a layer contained a new branch, the proposed distance-based rule was used. First, the distances between each cluster in the lower layer and its nearest neighbor cluster in the upper layer were calculated, i.e., D_{CNL} . Afterward, the number of elements in D_{CNL} that satisfied the rule of being smaller than two times $Min(D_{CNL})$ was counted and named $Nmin$. When $Nmin$ was less than CNU , some clusters in the upper layer were unable to find their source branches in the lower layer. Here, the upper layer contained new branches (Fig. 3. b.3. I). In contrast, if $Nmin$ was equal to or greater than CNU , then all the clusters in the upper layer could find their source branch in the lower layer, which meant that the upper layer had no new branches (Fig. 3. b.3. II).

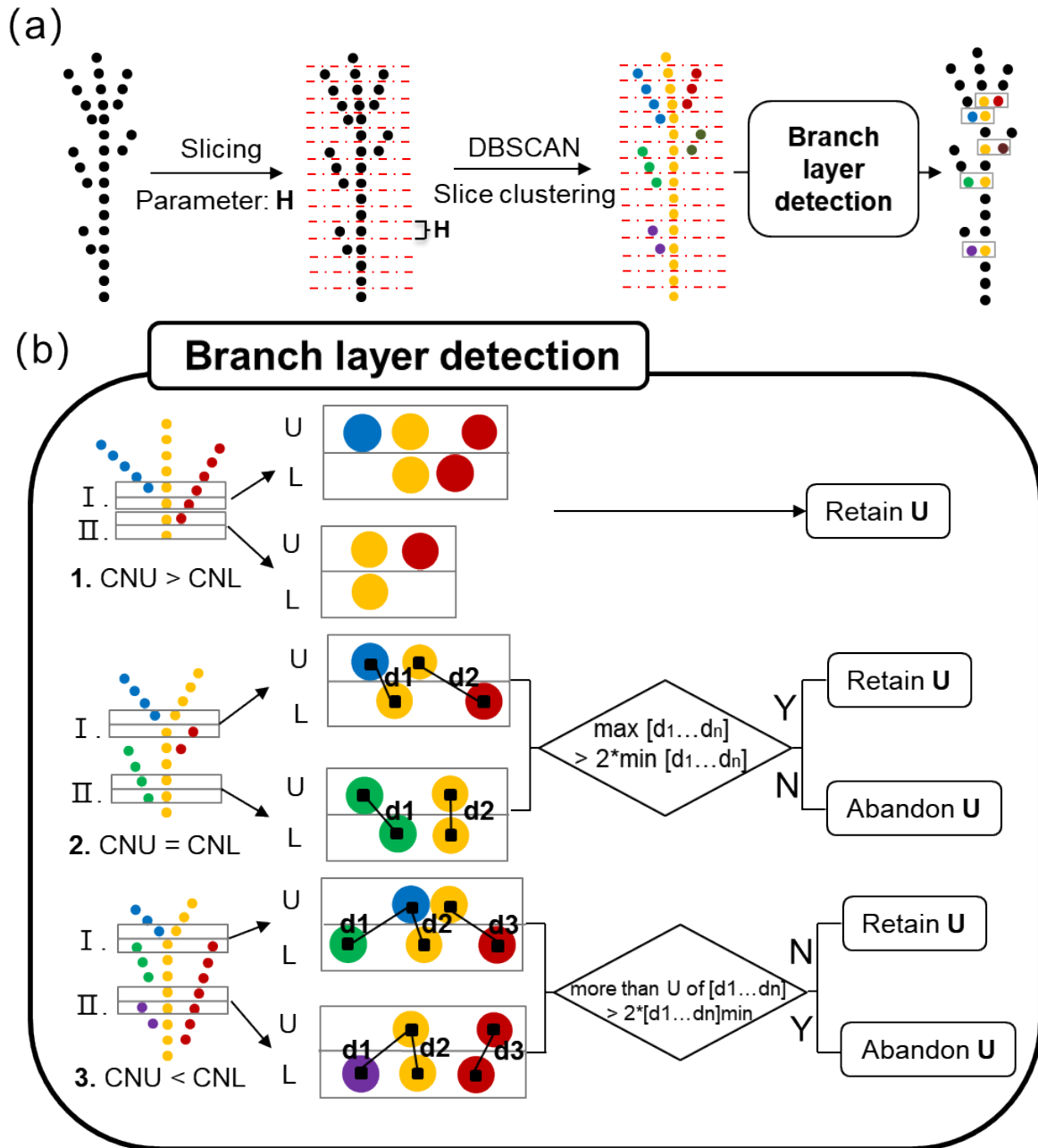


Fig. 3. Flowchart of slice clustering and branch detection. (a) Branch detection workflow. The parameter slice height (H) represents the height of the slice. (b) The logic for detecting the branch layers is based on the number of clusters in two adjacent layers. U and L represent the upper and lower layers, respectively. CNU and CNL represent the number of clusters in the upper and lower layers, respectively. d_n (e.g., d_1 , d_2 , and d_3) represents the nearest distance from a cluster in the lower layer to a cluster in the upper layer.

Step 2: Node point location

The node point location was designed to find the intersection point of the main stem and branches (Fig. 4). It was implemented by considering regional growth based on the detected branch clusters. As the clusters forming a branch angle were often the closest clusters, this study considered only two clusters belonging to a new branch by finding the two closest clusters in each layer, although there was a possibility of more than two clusters existing in the same layer. The median values of two detected clusters were calculated and used as the seed points in the three steps of the downward regional growth method. First, the points with z values larger than the two seed points were removed from the point cloud of the plant. Second, the two seed points (on the main stem and the branch) were used for searching for the N nearest neighbors, i.e., growth point number, with the *KD-Tree* (Bentley, 1975) searching method. The resulting N points on each side were extracted as a new cluster, which was then used to recalculate the new seed points. Third, the new seed points were used to repeat the aforementioned downward growth process, i.e., the first two steps. The growth ended when the number of points (overlapping points) obtained on both sides during the search was greater than N . Finally, the overlapping points were extracted, and their median was used as the node point.

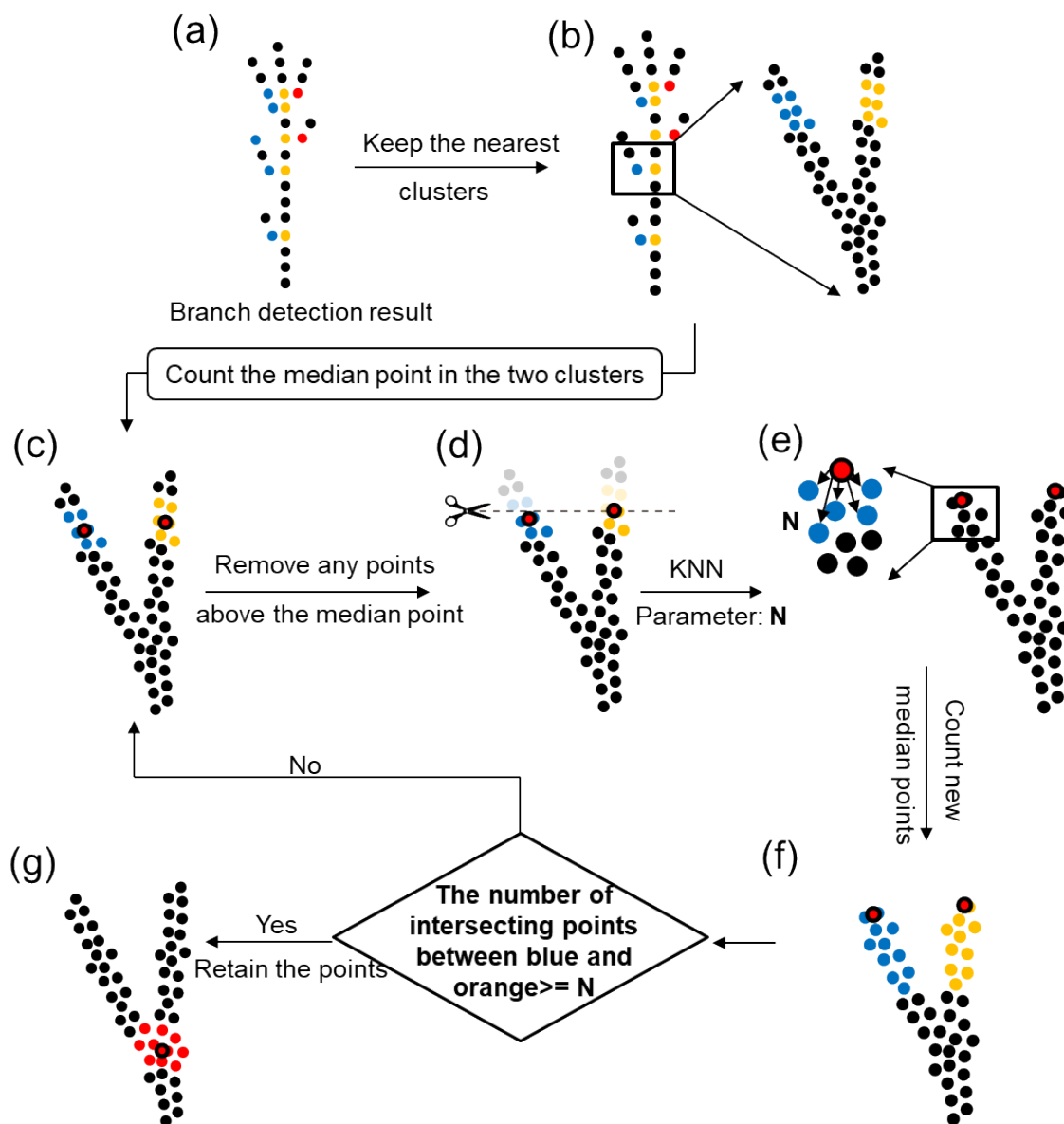


Fig. 4. Flowchart of the node location and downward regional growth. (a) The layers include new branches that are marked with colorful points after branch detection. (b) The two closest clusters were kept in each layer. (c) The median points of the two clusters were calculated. (d) Points greater than the median point were removed. (e) The KNN method was used to search for N nearest neighbor points with two median points as the seed points. (f) The two searched new clusters are presented in yellow and blue. If the number of points (overlapping points) obtained during the search on both sides was greater than N , the two clusters were used to extract new median points, which were then used for repeating steps (c) to (f). Otherwise, (g) the downward search ended, and the median of the overlapping points was extracted as the node point.

Step 3: Branch point optimization

Branch point optimization was designed to improve the angle calculation accuracy by finely adjusting the locations of branch points. As the branch points obtained by slicing and branch detecting were too close to the node point, the angle calculation had large uncertainties based on the measuring technique. The branch points can be optimized based on the distance between them and the newly founded node point. First, we selected a layer above the node point with a thickness of 1 cm and a slice depth (D) for simulating the manual measurement. Second, DBSCAN was used for dividing the points in the sliced layer into two groups and calculating the medians of each group as the optimized branch points. Third, two spatial vectors, i.e., \vec{a} and \vec{b} , were formed by connecting the node point and the optimized branch points, which were used for calculating the branch angle (Fig. 5d).

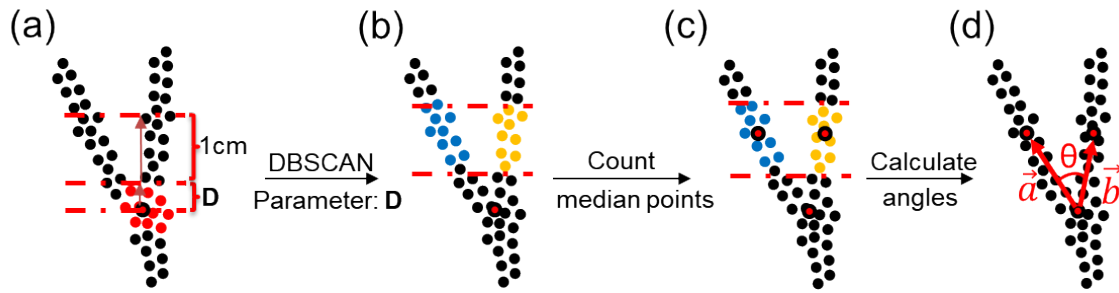


Fig. 5. Branch point optimization and branch angle calculation. The parameter slice depth (D) denotes the distance used for upward optimization based on the node point. (a) The points from a 1-cm deep layer were selected above the branch point with a height of D . (b) The selected points were clustered into two groups using DBSCAN. (c) The median points of two clusters were calculated, which represented the two branch points after optimization. (d) The branch angle was calculated based on two spatial vectors formed by the optimized branch points and the node point.

Accuracy assessment of the proposed SCAG

The accuracy assessment of the proposed SCAG algorithm mainly consisted of two parts, namely, branch detection accuracy assessment and angle calculation accuracy assessment. The branch detection accuracy was evaluated by analyzing the Euclidean distance between the SCAG-predicted node points and the manually labeled node points. If there was one or more labeled node points surrounding the predicted node point with a distance of less than 1 cm (approximately the average diameter of a soybean node area), the prediction was correct and was called a true-positive (TP) case; otherwise, the

prediction was treated as a false-positive (*FP*) case. If there was no predicted node point in the 1 cm neighborhood of a labeled node point, the node point was missed and called a false-negative (*FN*). Finally, the *recall* (*R*), *precision* (*P*), and *F-score* (*F*) (Goutte & Gaussier, 2005) were calculated. The values of all three metrics range from 0 to 1. A larger value indicates better accuracy. These metrics can be mathematically expressed as follows:

$$R = \frac{TP}{TP + FN} \quad (1)$$

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$F = 2 \times \frac{R \times P}{R + P} \quad (3)$$

The angle calculation accuracy was evaluated by using the coefficient of correlation (*r*) and root mean squared error (RMSE) between the SCAG predictions and manual measurements.

Parameter sensitivity analysis of SCAG

The proposed SCAG algorithm has three important parameters, namely, the slice height (*H*), growth point number (*N*), and slice depth (*D*). Theoretically, the branch detection accuracy may be influenced by parameters *H* and *N*, whereas parameter *D* may influence only the accuracy of the branch angle calculation.

The sensitivity analysis of parameters *H* and *N* was evaluated with the *F-score* under different parameter combinations using a grid-searching method in Python. The parameter *H* was set as 0.3 cm, 0.5 cm, 1 cm, 1.5 cm, 2 cm, or 2.5 cm relative to the height of the adjacent branching node. The parameter *N* was chosen from 10, 20, 30, 40, 60, 80, 100, 120, and 140 by considering the point density around the branching node. Similarly, the sensitivity analysis of parameter *D* was evaluated with the correlation coefficient *r* under different selections using an equally spaced searching method in Python. Parameter *D* was chosen from 1.5 cm, 2 cm, 2.5 cm, 3 cm, 3.5 cm, and 4 cm based on manual measurements.

Cross-comparison with other methods

The proposed SCAG method was compared with two other methods, including the heuristic method (i.e., a threshold method based on point density) and the machine learning-based method (i.e., support vector machine (Noble, 2006)). These three methods mainly differ in terms of the techniques used for detecting the branches, but they use the same approach for calculating the angles. The density-based method (DB) is a heuristic method that is based on the knowledge that the local point density of a branch node point is greater than that of other locations. In contrast, the support vector machine (SVM) method performs point classification (i.e., node point or not) in a supervised manner based on manually labeled targets and handcrafted features. The reasons for selecting SVM for comparison were based on its wide usage and low dataset requirements (Tsang et al., 2005). The implementation details of the two methods are provided in Supplemental Materials S2.

Results

The SCAG performed well in branch detection and branch angle calculation

The proposed SCAG method was evaluated qualitatively and quantitatively by using the *Soybean3D* dataset. Twelve representative samples were selected to visually demonstrate the branch detection performance due to limited space (Fig. 6). The samples included maximum and minimum values of the plant height, canopy width, CHR, average angle, number of branch angles, and nearest point distance (NPD). In each subplot, each detected branch consisted of three points of the same color, which represented two branch points and one node point. NPD had a minor effect on SCAG performance because branches were accurately detected in large and small NPD cases (Fig. 6a). Similarly, the proposed SCAG accurately detected the branch angles of samples with different plant height, canopy width, CHR, branch angle number, and average angle values. Some underdetections existed when the points were missing (Fig. 6c) or the branch angle was too large (Fig. 6e-f). Overall, the proposed SCAG method performed effectively for different soybean varieties.

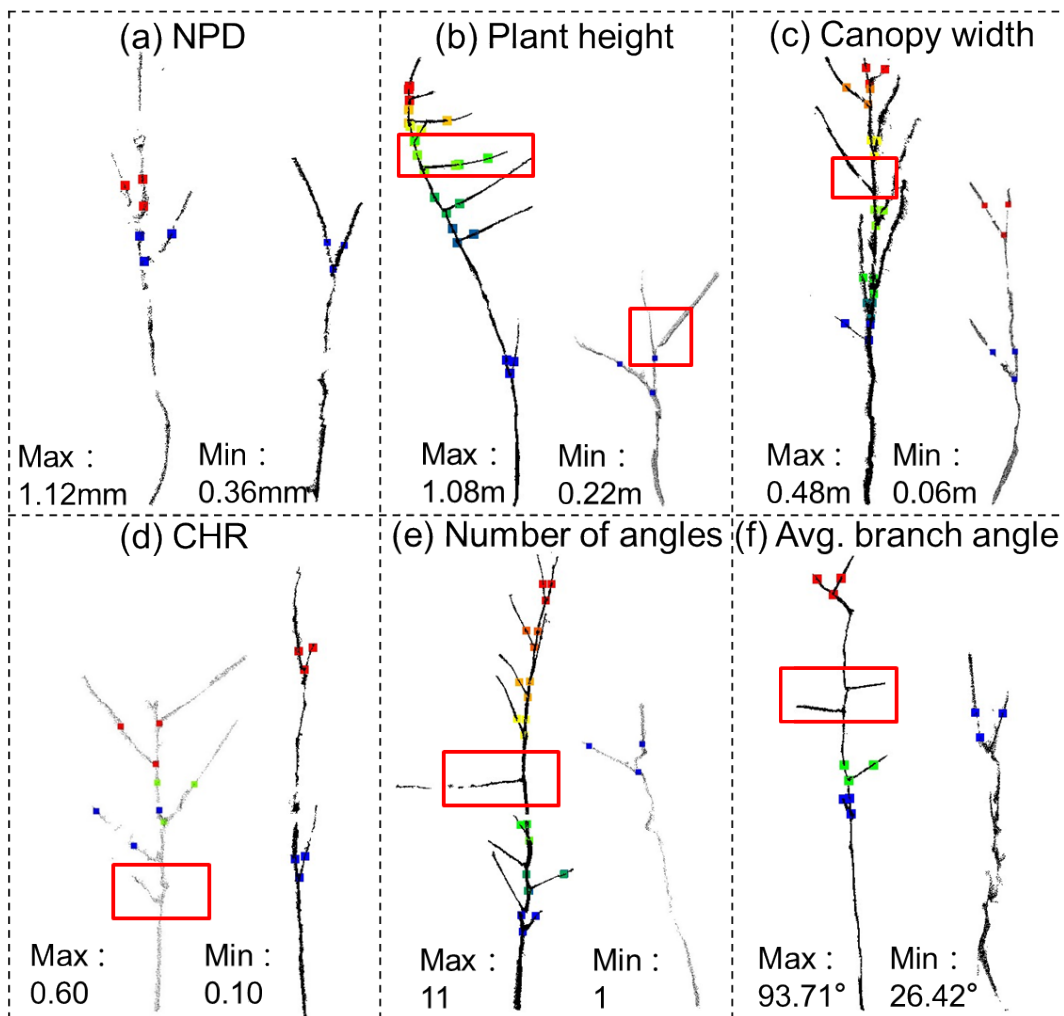


Fig. 6. Branch detection and node location results for 12 representative soybean varieties with high phenotypic diversity. The three points in different colors on each soybean crop represent one node point and two branch points of the branch detected by the SCAG method. The left and right subplots represent the point clouds of the plants with the largest and smallest (a) NPD, (b) plant height, (c) canopy width, (d) CHR, and (e) number of branch angles, respectively. (f) Average branch angles. The red boxes indicate branches not detected by the algorithm.

The quantitative results showed that the mean R , P , and F -score of branch detection were 0.81, 0.73, and 0.77, respectively (Table 1). Moreover, the F -scores of the simple, medium, and complex groups were 0.80, 0.77, and 0.66, respectively. Furthermore, the quantitative results of the branch angle calculations showed that the mean r of the proposed SCAG was 0.84. The r values for the simple, medium, and complex samples were 0.88, 0.85, and 0.83, respectively (Supplemental Materials S2), which showed a decreasing trend.

In addition, the proposed SCAG-predicted angles showed a slight overestimation for small angles and an underestimation for large angles in the different complex groups.

Table 1. The accuracy of branch detection and angle calculation for the proposed SCAG method based on the Soybean3D dataset.

Group	Sample number	Recall	Precision	F-score	r
Simple	50	0.75	0.86	0.80	0.88
Medium	51	0.77	0.78	0.77	0.85
Complex	51	0.71	0.62	0.66	0.83
All	152	0.81	0.73	0.77	0.84

The SCAG was superior to the other methods and was robust

For branch detection, the proposed SCAG method detected most of the angles ($R=0.81$), which was better than the SVM ($R=0.52$) and DB ($R=0.55$) methods (Table 2). Moreover, SCAG outperformed the other two methods in terms of the *precision* (0.73) and *F-score* (0.77). This showed that the proposed SCAG method detects more branches with higher *precision*. In the angle calculation, all three methods achieved high performance with *r* values greater than 0.80, and the differences among the three methods were relatively small.

Table 2. The best results of the three calculation methods for branch angle detection after parametric sensitivity analysis.

Method	Recall	Precision	F-score	r
SVM	0.52	0.55	0.53	0.84
DB	0.55	0.56	0.55	0.88
SCAG	0.81	0.73	0.77	0.84

Parameter sensitivity analyses were separately conducted for branch detection and angle calculation. The branch detection was insensitive to parameter N but influenced by parameter H following a clear pattern (Fig. 7). On the one hand, the optimal value of H , according to the best *F-score* (0.77; Fig. 7c), was between 0.5 cm and 1.0 cm, which was

around the branch diameter. On the other hand, the R value decreased with H (Fig. 7a), and the P value first increased and then decreased (Fig. 7b).

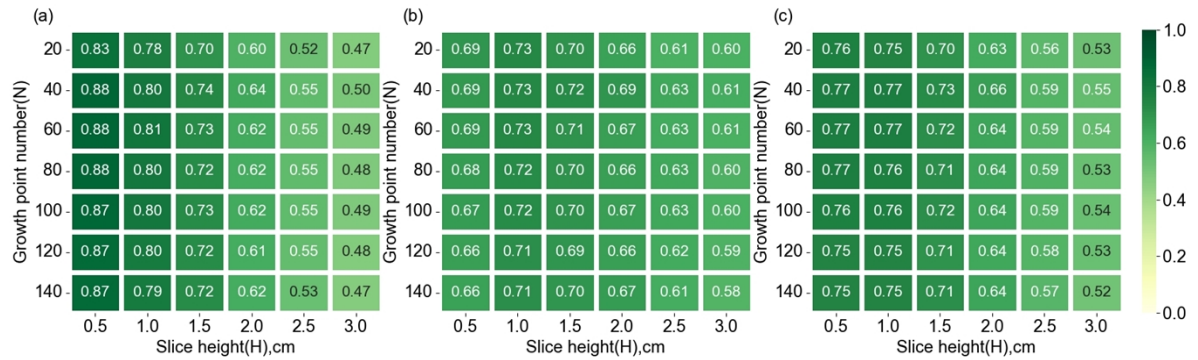


Fig. 7. Parameter sensitivity analysis of branch detection. (a), (b), and (c) *Recall*, *precision*, and *F-score* of branch detection under various parameter combinations of slice height (H) and growth point number (N).

In the angle calculation, D was the only parameter that affected the result. However, the angle calculation accuracy may be indirectly influenced by the branch detection process, which is influenced mainly by parameter H but is insensitive to parameter N according to Fig. 7. Therefore, we analyzed the influences of parameters D and H on the angle calculation. The results showed that H has only a subtle influence. The r was relatively higher when D was approximately 2.0~3.0 cm (Fig. 8), which was closer to the threshold height between branch points and node points when measured manually.

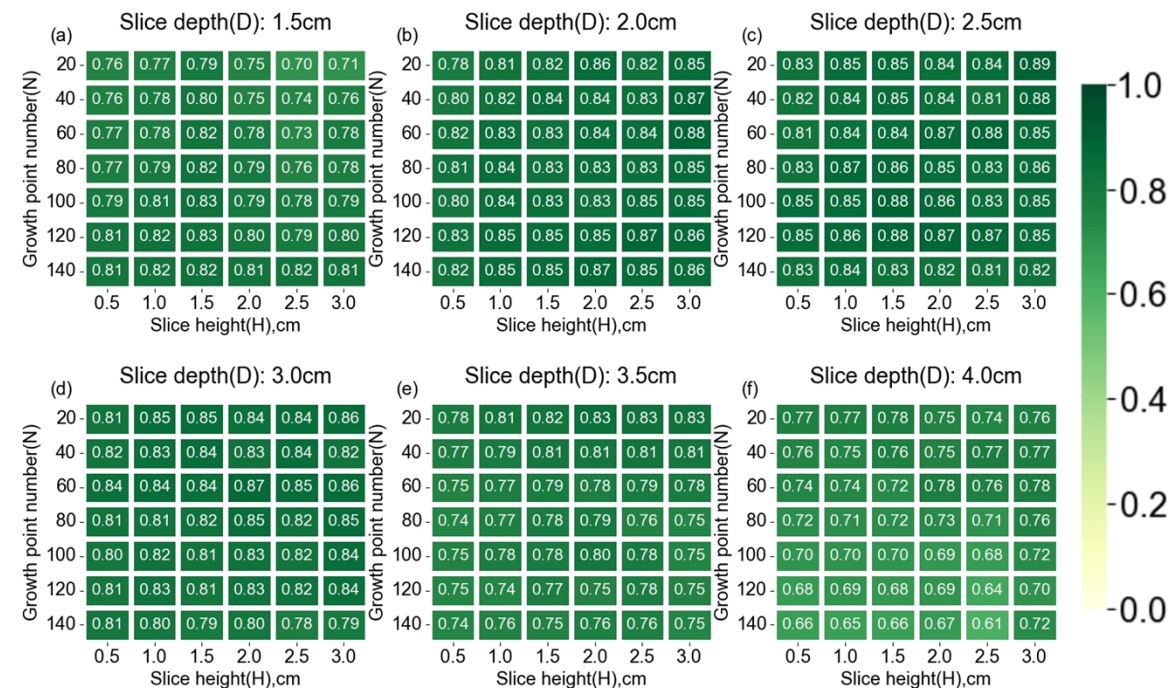


Fig. 8. Parameter sensitivity analysis of the branch angle calculation according to the angle calculation accuracy (r). The title of each subplot denotes the value of parameter D . The horizontal and vertical coordinates of each subplot represent the parameters H and N , respectively.

Discussions

Automatic, accurate, and robust branch detection and angle calculation

This study proposed an accurate method (SCAG) for branch detection inspired by the growth laws of plants. The proposed SCAG is accurate in most cases (Fig. 6; Table 1). Some underdetection cases were mainly caused by missing data due to overlapping (Fig. 6c), too short branches (Fig. 6d), or too large branch angles that do not obey the principle based on the upward growing angle (Fig. 6e-f). On the other hand, overdetection cases are rare unless excessive noise points exist.

In addition to high accuracy, the proposed SCAG method is not very sensitive to parameter N within the range of 20 to 140 or parameter D within the range of 1.5 cm to 4.0 cm, and it is affected by parameter H (i.e., slice height in branch detection). (Fig. 7). If H is too large, a slice layer may include multiple node points, resulting in a lower *recall* (Fig. 7a). If H is too small, a slice may have too few points, which may be considered noise during the clustering process, thus reducing *precision* (Fig. 7b). Theoretically, the proposed SCAG method should be robust because it detects branches based on the logic of branch growth based on clustered median points instead of considering every point. This robustness of the median points has been discussed in previous studies (Jin et al., 2019; Wang et al., 2023). In practice, the parameter H should be set to approximately two times the branch diameter according to the sensitivity analysis (Fig. 7). Parameter D is suggested to be consistent with manual measurements (Fig. 8). The parameter N is insensitive and can be set by simple pretesting (Fig. 7).

The proposed SCAG outperforms the SVM and DB methods (Table 2) because it is difficult to define a feature set for SVM and a robust density threshold for the DB method, given a sparse, unordered, and unstructured point cloud. In the DB method, the reason for the low angle detection rate is that the density of the point cloud is uneven. In the SVM method, the handcrafted features are also experienced and may not be optimal, thus

resulting in a low branch detection rate. Despite the rapid development of deep learning methods (Jin et al., 2018; Guo, Q et al., 2020; Jin et al., 2020b; Li, D et al., 2022; Xu & Li, 2022), they have high requirements for building large-volume, well-labeled, and rich-diverse datasets.

Extensibility analysis of the proposed SCAG algorithm

The transferability of an algorithm is crucial for ensuring its widespread application. The proposed SCAG algorithm was therefore evaluated with an open-source point cloud dataset, namely, Pheno4D (Schunck et al., 2021), to show its general applicability. The Pheno4D dataset contains point clouds of two different kinds of crops, i.e., maize and tomato. Maize is a monocotyledonous crop that has a relatively simple structure due to its asymmetrically distributed leaves along a single stem. In contrast, tomato is a dicotyledonous crop that has a much more complex structure than soybean, such as multiple branches around the main stem. This work selected six different maize and six different tomato point clouds to evaluate the transferability performance of the SCAG algorithm.

As expected, SCAG works well for maize (parameter setting: $H=20$; $N=100$; $D=30$) and tomato (parameter setting: $H=5$; $N=120$; $D=8$) varieties (qualitative results in Fig. S11). The accuracy of the leaf/branch angle calculations was high for maize ($r=0.95$) and tomato ($r=0.94$), thus demonstrating its effectiveness across multiple crop types (Fig.11). The predicted angle values show an overall overestimation in maize, which is mainly caused by the adhesion between the leaf sheath and stem of corn, resulting in the predicted node points being at higher positions than the manually measured node points (Fig. S11).

Therefore, SCAG can be applied not only to soybean for branch angle-related trait extraction but also to other crops, such as maize and tomato, thereby improving architecture-related breeding applications (van Esse, 2022; Wang et al., 2022). Through a better understanding of how plant architecture affects light penetration, air circulation, and disease susceptibility, breeders can design new varieties that maximize the yield, quality, efficiency (Coupel-Ledru et al., 2022), and tolerance potential of crops in various

environments (Guo, W et al., 2020), contributing to global food security and agricultural innovation in the future.

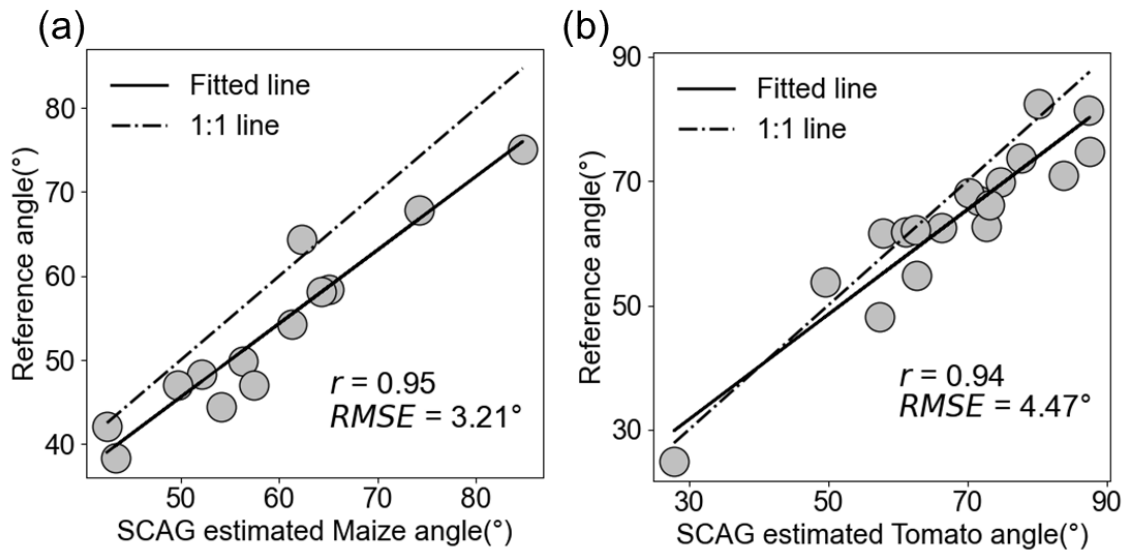


Fig. 9. The quantitative results of leaf/branch angle calculations obtained by using the proposed SCAG algorithm on six (a) maize and (b) tomato point clouds.

The application value of the SCAG algorithm in ideal soybean type selection

Selecting ideal soybean types is a key step for breeding applications. To construct new traits for ideal soybean type selection, we first extracted four basic three-dimensional phenotypes from all 405 varieties in both years, including the average angle (computed by SCAG), canopy width, height, and stem length (Supplemental Materials S5). In this study, we noted that there was an inconsistency of phenotypes between the two replications. This inconsistency may be caused by data scanning. One main reason was that the individual plant structure can be damaged when it is transported from the field to the indoors for scanning. The other possible reason is that the scanner has a point resolution of approximately 2 mm, which makes it challenging to characterize thin branches and small stems, especially under occlusions. Therefore, data quality control was conducted based on the basic traits before constructing the new traits. First, if a variety had no sample in any replication of two years, it was removed. Second, the consistency of traits between two replications in each year was evaluated based on stem length because it had high accuracy (Fig. S7) and was insensitive to plant posture (e.g., tilted or not). If the relative error (Park & Stefanski, 1998) of the stem length between two replications was greater than 30%, the variety was excluded. After data quality filtering, 96 varieties were retained for 2021, and 76 varieties were retained for 2022. Finally, 55 common varieties from both years were identified for further trait exploration.

Second, 11 composite indices were constructed to evaluate density tolerance based on basic traits, including the CHR (ratio of canopy width to height, a classic density tolerance index), CLR (ratio of canopy width to stem length), LHR (ratio of stem length to height), CAR (ratio of canopy width to average angle), AHR (ratio of average angle to height), ALR (ratio of average angle to stem length), ANR (ratio of average angle to angle number), ACHR (product of average angle and CHR), ACLR (product of average angle and CLR), ANRCHR (product of ANR and CHR) (Supplemental Materials S6). Smaller values for these composite indices meant that the plant looks more compact and should be useful for screening potentially density-tolerant varieties.

The single-year H^2 (Visscher et al., 2008; Zang et al., 2023) was calculated for the 96 varieties in 2021 and the 76 varieties in 2022, while the multiyear H^2 was computed for the

55 varieties across the two years. Moreover, two-way analysis of variance (ANOVA) revealed significant differences among varieties for different traits, suggesting the presence of distinct plant types among our materials (Table S2). To better reveal the ability of traits to describe the relative differences among varieties, we evaluated the consistency between the two years using the DTW (dynamic time warping) method based on the ranking results of traits for varieties (Muda et al., 2010). DTW offers a flexible alternative that does not require data to adhere to specific distribution assumptions and enables the repeatability assessment of trends despite overall differences in absolute (mean) values. A smaller DTW indicates better consistency between two sequences. All indices were normalized to the range of 0 to 1 using the min–max scaling method before calculating DTW. The DTW of two series of data may be influenced by their series order (the variety number in this study). However, the relative relationship between DTW and different traits was not influenced by the variety order (Table S3), making it suitable for comparing the repeatability of different traits across years. Overall, we expected to screen traits with higher H^2 and smaller DTW, which should be more robust in evaluating the attributes of density tolerance across years and may be more valuable in future genetic applications.

Heritability (H^2) and repeatability (DTW) analyses revealed that stem length and height had better H^2 values than canopy width and average angle (Fig. 10). Notably, the relatively high H^2 for stem length may be a result of data quality control. Moreover, different traits had similar repeatability between the two years, although the average angle had the best (smallest) DTW.

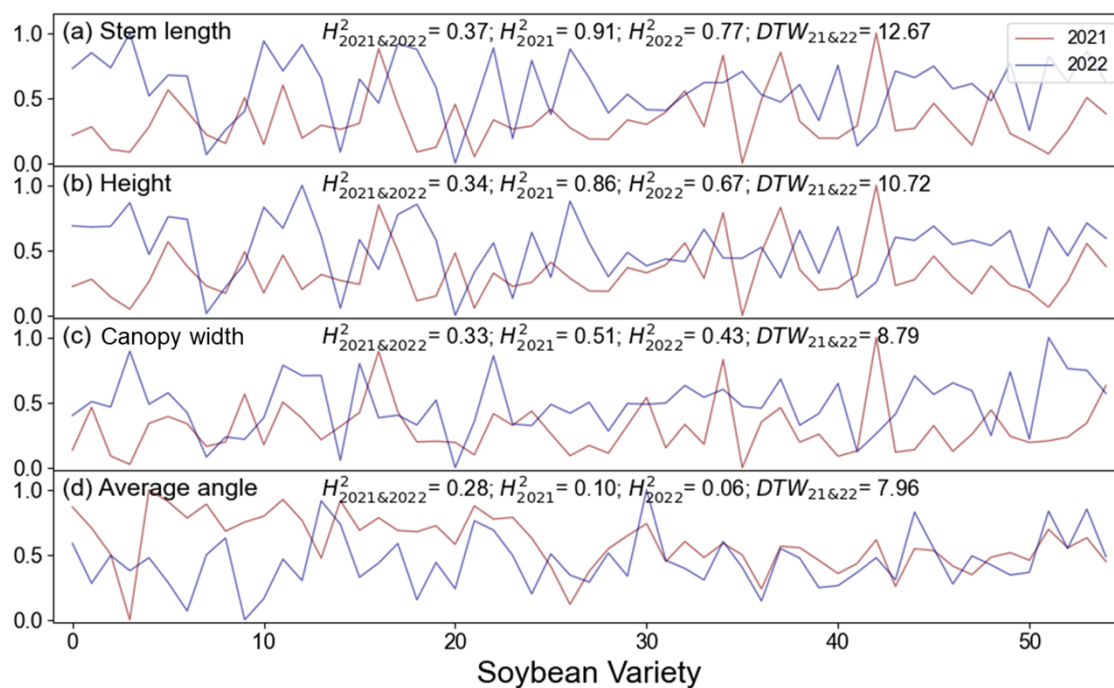


Fig. 10. The repeatability (DTW) of various density tolerance indices across two years was compared for 55 soybean types, and their heritability (H^2) was assessed. Among them, $H^2_{2021\&2022}$ exhibited broad-sense heritability between the two years. H^2_{2021} showed broad-sense heritability in 2021. H^2_{2022} showed broad-sense heritability in 2022. The vertical axis of each subplot represents the normalized value of each index, and the horizontal axis represents the different soybean varieties.

The analysis of the 11 composite indices showed that the AHR and ALR were clearly better than the CHR in H^2 , while the ANR and CAR were similar to the CHR. All of these angle-related indices had similar repeatability to the CHR (i.e., similar DTW). In contrast, we found that the stem length-related traits LHR and CLR had lower heritability and worse DTW than the CHR for the selection of density-tolerant varieties (Fig. S8). Furthermore, composite indices can be further combined to construct other new traits that are marginally better than the CHR, such as the ACHR and ACLR (Fig. S9), but simple traits are usually preferred among useful traits in practice.

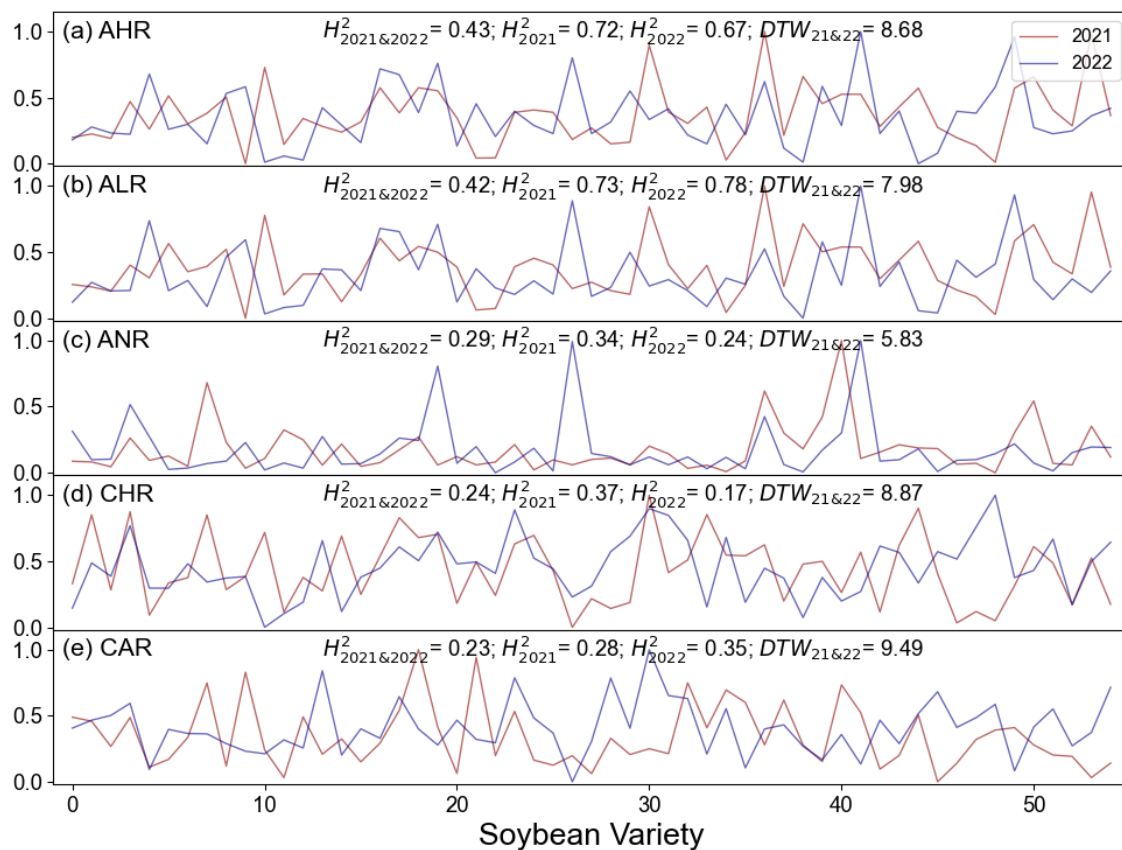


Fig. 11. Comparison of the repeatability (DTW) of different composite density tolerance indices for 55 soybean varieties in 2021 (red line) and 2022 (blue line) and assessment of their heritability (H^2). $H^2_{2021\&2022}$ showed broad-sense heritability between the two years. H^2_{2021} showed broad-sense heritability for 2021. H^2_{2022} showed broad-sense heritability for 2022.

The application value of the SCAG algorithm holds great potential for enhancing crop development and agricultural productivity. One significant application of SCAG is its ability to screen for potentially ideal plant architecture. Different crops have distinct branch angle preferences, with some requiring more upright architecture to optimize light interception and promote yield production (Percy et al., 2005). This study explored novel angle-related traits (e.g., AHR and ALR) that have better heritability (higher H^2) and repeatability (lower DTW) across years for evaluating density-tolerant varieties (Fig. 11). These novel traits hold potential value for selecting ideal plant types and screening germplasm resources (Ao et al., 2013; Liu et al., 2020; Wang & Wang, 2015; Moeinizade et al., 2020; Singh et al., 2020; Basu & Parida, 2023).

Contributions and implications

This work proposed a biologically inspired method for soybean branch detection and angle calculation based on point clouds. The proposed SCAG method incorporates the idea of logical judgment based on continuous changes in the number of branches and stratification, clustering, and growth technologies. The proposed SCAG achieves high branch angle calculation accuracy ($r=0.84$) and outperforms two other branch angle extraction algorithms. Moreover, the proposed SCAG is insensitive to hyperparameters and is suitable for other crops, such as maize and tomato. Moreover, the SCAG-predicted angles are valuable for selecting heritable and repeatable traits to identify density-tolerant varieties. Additionally, diverse datasets (*Soybean3D*) and source code, such as for deep learning-based branch detection, are publicly available for future studies.

Despite the aforementioned contributions, this work also reveals some implications to be considered in future studies. These implications primarily revolve around data collection and sample quality. First, point cloud data collection is time-consuming due to transportation and scanning. One way to improve the efficiency of 3D data acquisition is to use 3D reconstruction techniques from high-resolution images based on new techniques, such as neural radiance fields (Mildenhall et al., 2021), which is much less expensive but still challenging in the field. Second, the point cloud accuracy needs to accurately describe the branch thickness. It is worth further exploring how point cloud quality affects the performance of a geometry-based method. In addition, missing data problems are common due to factors such as leaf occlusion, which can be alleviated by introducing data reconstruction algorithms, such as 3D model-based methods (Pound et al., 2014; Kar et al., 2015) and 3D geometry-based methods (Furukawa & Hernández, 2015; Verbin & Zickler, 2020), during the data preprocessing steps. The underdetection of flattened/broken branches may require a combination of upward and downward detection results or the use of state-of-the-art deep learning methods (Huang et al., 2020) that may have a better ability to capture long-range/global relationships of branches. The deep learning methods may also be beneficial for detecting small branch targets due to their multiscale feature embedding and attention mechanisms (Jin et al., 2020b; Li, D et al., 2022; Yu et al., 2022).

Third, the detected branches may contain petioles that are directly connected to the stem. Discriminating petioles and branches that are both connected to stems is challenging based on the precision of the current data. Once a plant point cloud containing leaves, petioles, branches, and stems can be collected completely, we could use graph theory and technology to discriminate among different organs (Li et al., 2020). Finally, although high-throughput phenotyping has been widely used in genetic analysis, we should be careful about the impact of phenotypic measurement errors on genetic analysis, which are not only random effects but also may be partially under genetic control (Freitas Moreira et al., 2021; Morota et al., 2022).

Conclusions

Branch detection and angle calculations are important prerequisites for quantifying plant structure and selecting ideal plant types that are important for achieving the soybean Green Revolution. In this work, a novel algorithm (SCAG) for branch detection and angle calculation is proposed. The proposed SCAG algorithm achieves good results in terms of the *recall* (0.81), *precision* (0.73), and *F-score* (0.77) for branch detection, as well as *r* (0.84) for angle calculation. The proposed SCAG algorithm outperforms the SVM (*F-score*=0.53) and DB (*F-score*=0.55) methods, while their angle calculation accuracies are similar. Additionally, the proposed SCAG algorithm has proven to be insensitive to parameter *N* within the range of 20 to 140 cm, as well as parameter *D* within the range of 1.5 cm to 4.0 cm, which are used for node point detection and angle calculation, respectively. Although the proposed SCAG algorithm is affected by parameter *H*, the influence pattern is clear, and the optimal values are usually located around the branch diameter. Additionally, the generalization ability of the proposed SCAG algorithm for other crops was verified by selecting representative maize (*r*=0.95) and tomato (*r*=0.94) point clouds. Finally, we proposed new indices (i.e., AHR and ALR) for screening density-tolerant soybean varieties using SCAG-predicted angles. While we emphasized the contributions of the proposed novel algorithm, we also hope to promote future studies that focus on data quality restoration and the development of artificial general intelligence algorithms by sharing our research dataset, algorithms, and software.

Acknowledgments: This work was supported in part by the Science and Technology Innovation 2030-Major Project (2023ZD04034), the Fundamental Research Funds for the Central Universities (KYCYXT2022017; KYQN2023021), Hainan Yazhou Bay Seed Laboratory (B21HJ1005), Jiangsu Province Key Research and Development Program (BE2023369), the Natural Science Foundation of Jiangsu Province (BK20231469), and the National Natural Science Foundation of China (32201656). J. Zhang was supported by the Sanya Yazhou Bay Science and Technology City (SCKJ-JYRC-2022-20). J. Wu was supported by the HKU Seed Funding for Strategic Interdisciplinary Research Scheme and the Innovation and Technology Fund (funding support to State Key Laboratories in Hong Kong of Agrobiotechnology) of the HKSAR, China. S. Jin would like to thank Prof. Junyi Gai for providing soybean seeds.

Competing interests: The authors declare no conflicts of interest.

Author contributions: S. Jin designed the study. S. Zhang, Y. Song, R. Ou, X. Lu, and J. Zhang carried out the field experiments. S. Zhang and S. Jin designed the SCAG algorithm, wrote the paper, and crafted all figures and tables. Y. Liu conducted the script testing and distribution. J. Wu, Y. Su, S. Xu, D. Jiang, Y. Ding, H. Xia, Q. Guo, and J. Zhang helped to revise the manuscript. All the authors have read and approved the final manuscript.

Data availability: The *Soybean3D* datasets, source code, software, and other supporting data are openly available on GitHub (https://github.com/Jinlab-9AiPhenomics/SCAG_PlantAngleExtractor).

References

- Abdel-Khalek S, Ben Ishak A, Omer OA, Obada ASF. 2017.** A two-dimensional image segmentation method based on genetic algorithm and entropy. *Optik* **131**: 414-422.
- Ao X, Zhao M-h, Zhu Q, Li J, Zhang H-j, Wang H-y, Yu C-m, Li C-h, Yao X-d, Xie F-t, et al. 2013.** Study on Plant Morphological Traits and Production Characteristics of Super High-Yielding Soybean. *Journal of Integrative Agriculture* **12**(7): 1173-1182.
- Basu U, Parida SK. 2023.** Restructuring plant types for developing tailor - made crops. *Plant Biotechnology Journal* **21**(6): 1106-1122.
- Bentley JL. 1975.** Multidimensional binary search trees used for associative searching. *Communications of the ACM* **18**(9): 509-517.
- Clark CB, Wang W, Wang Y, Fear GJ, Wen Z, Wang D, Ren B, Ma J. 2022.** Identification and molecular mapping of a major quantitative trait locus underlying branch angle in soybean. *Theoretical and Applied Genetics* **135**(3): 777-784.
- Coupel - Ledru A, Pallas B, Delalande M, Segura V, Guitton B, Muranty H, Durel CE, Regnard JI, Costes E. 2022.** Tree architecture, light interception and water - use related traits are controlled by different genomic regions in an apple tree core collection. *New Phytologist* **234**(1): 209-226.
- Dutagaci H, Rasti P, Galopin G, Rousseau D. 2020.** ROSE-X: an annotated data set for evaluation of 3D plant organ segmentation methods. *Plant methods* **16**: 28.
- Freitas Moreira F, Rojas de Oliveira H, Lopez MA, Abughali BJ, Gomes G, Cherkauer KA, Brito LF, Rainey KM. 2021.** High-throughput phenotyping and random regression models reveal temporal genetic control of soybean biomass production. *frontiers in plant science* **12**: 715983.
- Furukawa Y, Hernández C 2015.** Multi-View Stereo: A Tutorial, Foundations and Trends® in Computer Graphics and Vision: Hanover, MA: now Publishers Inc.
- Gao J, Yang S, Cheng W, Fu Y, Leng J, Yuan X, Jiang N, Ma J, Feng X. 2017.** GmILPA1, encoding an APC8-like protein, controls leaf petiole angle in soybean. *Plant physiology* **174**(2): 1167-1176.
- Ghahremani M, Williams K, Corke F, Tiddeman B, Liu Y, Wang X, Doonan JH. 2021.** Direct and accurate feature extraction from 3D point clouds of plants using RANSAC. *Computers and Electronics in Agriculture* **187**: 106240.
- Goutte C, Gaussier E 2005.** A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. *European conference on information retrieval*: Springer. 345-359.
- Gratani L, Ghia E. 2002.** Changes in morphological and physiological traits during leaf expansion of *Arbutus unedo*. *Environmental and experimental botany* **48**(1): 51-60.
- Guo B, Sun L, Jiang S, Ren H, Sun R, Wei Z, Hong H, Luan X, Wang J, Wang X, et al. 2022.** Soybean genetic resources contributing to sustainable protein production. *Theoretical and Applied Genetics* **135**(11): 4095-4121.
- Guo Q, Jin S, Li M, Yang Q, Xu K, Ju Y, Zhang J, Xuan J, Liu J, Su Y. 2020.** Application of deep learning in ecological resource research: Theories, methods, and challenges. *Science China Earth Sciences* **63**: 1457-1474.
- Guo Q, Wu F, Pang S, Zhao X, Chen L, Liu J, Xue B, Xu G, Li L, Jing H, et al. 2018.**

- Crop 3D—a LiDAR based platform for 3D high-throughput crop phenotyping. *Science China Life Sciences* **61**(3): 328-339.
- Guo W, Chen L, Herrera-Estrella L, Cao D, Tran L-SP. 2020.** Altering plant architecture to improve performance and resistance. *Trends in plant science* **25**(11): 1154-1170.
- Huang K, Wang Y, Kneip L 2021.** Dynamic Event Camera Calibration. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*: IEEE. 7021-7028.
- Huang X, Hilscher J, Stoger E, Christou P, Zhu C. 2021.** Modification of cereal plant architecture by genome editing to improve yields. *Plant Cell Reports* **40**(6): 953-978.
- Huang Z, Yu Y, Xu J, Ni F, Le X 2020.** Pf-net: Point fractal network for 3d point cloud completion. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7662-7670.
- Ingale AK, J DU. 2021.** Real-time 3D reconstruction techniques applied in dynamic scenes: A systematic literature review. *Computer Science Review* **39**: 100338.
- Jayakumari R, Nidamanuri RR, Ramiya AM. 2021.** Object-level classification of vegetable crops in 3D LiDAR point cloud using deep learning convolutional neural networks. *Precision Agriculture* **22**(5): 1617-1633.
- Jin S, Su Y, Gao S, Wu F, Hu T, Liu J, Li W, Wang D, Chen S, Jiang Y. 2018.** Deep learning: individual maize segmentation from terrestrial lidar data using faster R-CNN and regional growth algorithms. *frontiers in plant science* **9**: 866.
- Jin S, Su Y, Song S, Xu K, Hu T, Yang Q, Wu F, Xu G, Ma Q, Guan H. 2020a.** Non-destructive estimation of field maize biomass using terrestrial lidar: an evaluation from plot level to individual leaf level. *Plant methods* **16**: 1-19.
- Jin S, Su Y, Wu F, Pang S, Gao S, Hu T, Liu J, Guo Q. 2019.** Stem–Leaf Segmentation and Phenotypic Trait Extraction of Individual Maize Using Terrestrial LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing* **57**(3): 1336-1346.
- Jin S, Su Y, Zhang Y, Song S, Li Q, Liu Z, Ma Q, Ge Y, Liu L, Ding Y. 2021a.** Exploring seasonal and circadian rhythms in structural traits of field maize from LiDAR time series. *Plant Phenomics* **2021**.
- Jin S, Su Y, Zhao X, Hu T, Guo Q. 2020b.** A point-based fully convolutional neural network for airborne LiDAR ground point filtering in forested environments. *IEEE journal of selected topics in applied earth observations and remote sensing* **13**: 3958-3974.
- Jin S, Sun X, Wu F, Su Y, Li Y, Song S, Xu K, Ma Q, Baret F, Jiang D, et al. 2021b.** Lidar sheds new light on plant phenomics for plant breeding and management: Recent advances and future prospects. *ISPRS Journal of Photogrammetry and Remote Sensing* **171**: 202-223.
- Kar A, Tulsiani S, Carreira J, Malik J 2015.** Category-specific object reconstruction from a single image. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1966-1974.
- Kenchanmane Raju SK, Adkins M, Enersen A, Santana de Carvalho D, Studer AJ, Ganapathysubramanian B, Schnable PS, Schnable JC. 2020.** Leaf Angle eXtractor: A high-throughput image processing framework for leaf angle measurements in maize and sorghum. *Appl Plant Sci* **8**(8): e11385.

- Li D, Shi G, Li J, Chen Y, Zhang S, Xiang S, Jin S. 2022.** PlantNet: A dual-function point cloud segmentation network for multiple plant species. *ISPRS Journal of Photogrammetry and Remote Sensing* **184**: 243-263.
- Li Q, Jin S, Zang J, Wang X, Sun Z, Li Z, Xu S, Ma Q, Su Y, Guo Q. 2022.** Deciphering the contributions of spectral and structural data to wheat yield estimation from proximal sensing. *The Crop Journal* **10**(5): 1334-1345.
- Li Y, Su Y, Zhao X, Yang M, Hu T, Zhang J, Liu J, Liu M, Guo Q. 2020.** Retrieval of tree branch architecture attributes from terrestrial laser scan data using a Laplacian algorithm. *Agricultural and Forest Meteorology* **284**: 107874.
- Liu S, Zhang M, Feng F, Tian Z. 2020.** Toward a “Green Revolution” for Soybean. *Molecular Plant* **13**(5): 688-697.
- Ma Z, Liu S. 2018.** A review of 3D reconstruction techniques in civil engineering and their applications. *Advanced Engineering Informatics* **37**: 163-174.
- Mantilla-Perez MB, Bao Y, Tang L, Schnable PS, Salas-Fernandez MG. 2020.** Toward “smart canopy” sorghum: Discovery of the genetic control of leaf angle across layers. *Plant physiology* **184**(4): 1927-1940.
- Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R. 2021.** Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1): 99-106.
- Moeinizade S, Kusmec A, Hu G, Wang L, Schnable PS. 2020.** Multi-trait Genomic Selection Methods for Crop Improvement. *Genetics* **215**(4): 931-945.
- Morota G, Jarquin D, Campbell MT, Iwata H. 2022.** Statistical methods for the quantitative genetic analysis of high-throughput phenotyping data. *High-Throughput Plant Phenotyping: Methods and Protocols*: Springer, 269-296.
- Muda L, Begam M, Elamvazuthi I. 2010.** Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv preprint arXiv:1003.4083*.
- Noble WS. 2006.** What is a support vector machine? *Nature biotechnology* **24**(12): 1565-1567.
- Nuijten RJG, Kooistra L, De Deyn GB. 2019.** Using Unmanned Aerial Systems (UAS) and Object-Based Image Analysis (OBIA) for Measuring Plant-Soil Feedback Effects on Crop Productivity. *Drones* **3**(3): 54.
- Park H, Stefanski LA. 1998.** Relative-error prediction. *Statistics & Probability Letters* **40**(3): 227-236.
- Pearcy RW, Muraoka H, Valladares F. 2005.** Crown architecture in sun and shade environments: assessing function and trade - offs with a three - dimensional simulation model. *New Phytologist* **166**(3): 791-800.
- Pound MP, French AP, Murchie EH, Pridmore TP. 2014.** Automated recovery of three-dimensional models of plant shoots from multiple color images. *Plant physiology* **166**(4): 1688-1698.
- Schubert E, Sander J, Ester M, Kriegel HP, Xu X. 2017.** DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)* **42**(3): 1-21.
- Schunck D, Magistri F, Rosu RA, Cornelißen A, Chebrolu N, Paulus S, Léon J, Behnke S, Stachniss C, Kuhlmann H. 2021.** Pheno4D: A spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis.

- Plos one* 16(8): e0256340.
- Singh RK, Prasad A, Muthamilarasan M, Parida SK, Prasad M. 2020.** Breeding and biotechnological interventions for trait improvement: status and prospects. *Planta* 252(4): 54.
- Su W, Jiang K, Guo H, Liu Z, Zhu D, Zhang X. 2019.** Extraction of phenotypic information of maize plants in field by terrestrial laser scanning. *Transactions of the Chinese Society of Agricultural Engineering* 35(10): 125-130.
- Sun Z, Li Q, Jin S, Song Y, Xu S, Wang X, Cai J, Zhou Q, Ge Y, Zhang R. 2022.** Simultaneous prediction of wheat yield and grain protein content using multitask deep learning from time-series proximal sensing. *Plant Phenomics* 2022.
- Takahashi Y, Shiojiri K, Yamawo A. 2021.** Aboveground plant-to-plant communication reduces root nodule symbiosis and soil nutrient concentrations. *Scientific Reports* 11(1): 12675.
- Tao H, Xu S, Tian Y, Li Z, Ge Y, Zhang J, Wang Y, Zhou G, Deng X, Zhang Z. 2022.** Proximal and remote sensing in plant phenomics: Twenty years of progress, challenges and perspectives. *Plant Communications*: 100344.
- Thapa S, Zhu F, Walia H, Yu H, Ge Y. 2018.** A Novel LiDAR-Based Instrument for High-Throughput, 3D Measurement of Morphological Traits in Maize and Sorghum. *Sensors (Basel)* 18(4).
- Tian J, Wang C, Xia J, Wu L, Xu G, Wu W, Li D, Qin W, Han X, Chen Q. 2019.** Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* 365(6454): 658-664.
- Tsai D, Worrall S, Shan M, Lohr A, Nebot E 2021.** Optimising the selection of samples for robust lidar camera calibration. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*: IEEE. 2631-2638.
- Tsang IW, Kwok JT, Cheung P-M, Cristianini N. 2005.** Core vector machines: Fast SVM training on very large data sets. *Journal of Machine Learning Research* 6(4).
- van Esse GW. 2022.** The quest for optimal plant architecture. *Science* 376(6589): 133-134.
- Verbin D, Zickler T 2020.** Toward a universal model for shape from texture. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 422-430.
- Visscher PM, Hill WG, Wray NR. 2008.** Heritability in the genomics era—concepts and misconceptions. *Nature reviews genetics* 9(4): 255-266.
- Wang D, Song Z, Miao T, Zhu C, Yang X, Yang T, Zhou Y, Den H, Xu T. 2023.** DFSP: A fast and automatic distance field-based stem-leaf segmentation pipeline for point cloud of maize shoot. *frontiers in plant science* 14.
- Wang H, Wang H. 2015.** The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Molecular Plant* 8(5): 677-688.
- Wang S, Wang K, Li Z, Li Y, He J, Li H, Wang B, Xin T, Tian H, Tian J. 2022.** Architecture design of cucurbit crops for enhanced productivity by a natural allele. *Nature Plants*: 1-14.
- Wu S, Wen W, Xiao B, Guo X, Du J, Wang C, Wang Y. 2019.** An accurate skeleton extraction approach from 3D point clouds of maize plants. *frontiers in plant science* 10: 248.
- Xiong X, Yu L, Yang W, Liu M, Jiang N, Wu D, Chen G, Xiong L, Liu K, Liu Q.**

2017. A high-throughput stereo-imaging system for quantifying rape leaf traits during the seedling stage. *Plant methods* **13**(1): 1-17.
- Xu R, Li C. 2022.** A review of high-throughput field phenotyping systems: Focusing on ground robots. *Plant Phenomics* **2022**.
- Yidan Y, Chengda L, Ruifang Z, Kun C 2016.** Rapeseed 3D reconstruction and morphological parameter measurement based on laser point cloud. *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*: IEEE. 1-6.
- You L, Sun Z. 2022.** Mapping global cropping system: Challenges, opportunities, and future perspectives. *Crop and Environment* **1**(1): 68-73.
- Yu C, Liu Y, Wu S, Hu Z, Xia X, Lan D, Liu X. 2022.** Infrared small target detection based on multiscale local contrast learning networks. *Infrared Physics & Technology* **123**: 104107.
- Zang J, Jin S, Zhang S, Li Q, Mu Y, Li Z, Li S, Wang X, Su Y, Jiang D. 2023.** Field-measured canopy height may not be as accurate and heritable as believed—Evidence from advanced 3D sensing. *Plant methods*: 19, 39.
- Zhang C, Yang G, Jiang Y, Xu B, Li X, Zhu Y, Lei L, Chen R, Dong Z, Yang H. 2020.** Apple Tree Branch Information Extraction from Terrestrial Laser Scanning and Backpack-LiDAR. *Remote Sensing* **12**(21): 3592.
- Zhang Y, Zhao X, Qian D. 2022.** Learning-Based Framework for Camera Calibration with Distortion Correction and High Precision Feature Detection. *arXiv preprint arXiv:2202.00158*.