



**RESEARCH ARTICLE**

# Phenotyping of Siliques Morphology in Oilseed Rape Using Skeletonization with Hierarchical Segmentation

Zhihong Ma<sup>1,2</sup>, Ruiming Du<sup>1,2</sup>, Jiayang Xie<sup>1,2</sup>, Dawei Sun<sup>1,2</sup>, Hui Fang<sup>1,2</sup>, Lixi Jiang<sup>3</sup>, and Haiyan Cen<sup>1,2\*</sup>

<sup>1</sup>College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, P.R. China.

<sup>2</sup>Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, P.R.

China. <sup>3</sup>Institute of Crop Science and Zhejiang Key Laboratory of Crop Germplasm, Zhejiang University, Hangzhou 310058, P.R. China.

\*Address correspondence to: [hycen@zju.edu.cn](mailto:hycen@zju.edu.cn)

Siliques morphology is an important trait that determines the yield output of oilseed rape (*Brassica napus L.*). Segmenting siliques and quantifying traits are challenging because of the complicated structure of an oilseed rape plant at the reproductive stage. This study aims to develop an accurate method in which a skeletonization algorithm was combined with the hierarchical segmentation (SHS) algorithm to separate siliques from the whole plant using 3-dimensional (3D) point clouds. We combined the L1-median skeleton with the random sample consensus for iteratively extracting skeleton points and optimized the skeleton based on information such as distance, angle, and direction from neighborhood points. Density-based spatial clustering of applications with noise and weighted unidirectional graph were used to achieve hierarchical segmentation of siliques. Using the SHS, we quantified the siliques number (SN), siliques length (SL), and siliques volume (SV) automatically based on the geometric rules. The proposed method was tested with the oilseed rape plants at the mature stage grown in a greenhouse and field. We found that our method showed good performance in siliques segmentation and phenotypic extraction with  $R^2$  values of 0.922 and 0.934 for SN and total SL, respectively. Additionally, SN, total SL, and total SV had the statistical significance of correlations with the yield of a plant, with  $R$  values of 0.935, 0.916, and 0.897, respectively. Overall, the SHS algorithm is accurate, efficient, and robust for the segmentation of siliques and extraction of siliques morphological parameters, which is promising for high-throughput siliques phenotyping in oilseed rape breeding.

## Introduction

Oilseed rape (*Brassica napus L.*) as an important oil source of food [1,2] is one of the most important oil industrial crops worldwide. The yield and quality of oilseed rape are largely determined by the development of siliques, i.e., the number, length, and volume of siliques at the mature stage, because the siliques are one of the important organs of an oilseed rape plant, playing a role not only in its bearing of oilseeds but also in photosynthesis and delivering developmental signals to maturing seeds [3–5]. Therefore, breeding a genotype with a large number of siliques and an ideal siliques appearance is a prioritized goal for high-yield breeding. The number of siliques varies significantly among different genotypes and under different cultivation practices; therefore, the accurate quantification of the parameters regarding siliques development is critical for yield prediction and breeding high-yielding oilseed rape varieties [6,7].

Currently, the traditional measurement of siliques development parameters such as siliques length (SL) and siliques number

(SN) depends largely on manual work, which is invasive, time-consuming, and inaccurate [8]. With the development of computer vision technologies, effective and image-based approaches have emerged to cope with the above problems. To date, methods have been developed to determine siliques parameters based on 2-dimensional (2D) or 3-dimensional (3D) data. Of these, 2D images have been the most widely used data thus far to phenotype a plant. Liu et al. [8] cut off the siliques from the plants and placed them on the conveyor belt; then, an RGB camera was used to obtain 2D images from the topside. The siliques were detected by the concave point detection segmentation method and measured by the pose estimation method. Wang et al. [9] cut and took photos of stems with and without siliques. The morphological appearances were then marked, and the angle between the siliques and main stem was measured using AutoCAD software. However, because they are restricted by illumination problems and the one-sided perspective inherent to 2D imaging [10], methods based on 2D image data are less robust and difficult to use to obtain the complete spatial information of plants. The above methods are limited

to obtaining simple morphological parameters and cannot be widely applied. In contrast, 3D data acquired using different 3D sensors, such as time-of-flight cameras and laser scanners [11], can represent a more detailed spatial information of the targets, ensuring more high-throughput plant phenotyping.

Clustering-based algorithms are the most widely used methods for 3D plant phenotyping. Xu et al. [12] used an RGB-D camera to obtain reconstructed point clouds of oilseed rapes. They used morphological operations (dilation and erosion) to segment stems and siliques and applied the clustering algorithm to extract the siliques. Lin et al. [13] obtained oilseed rape point clouds of high quality using a laser scanner. The region growing-based clustering algorithm with normal information was applied to segment siliques and stems. However, clustering-based methods are generally applied to plants with population scales or simple structures [14–16], and the performance is highly influenced by the initial value setting and tedious parameter tuning. Moreover, clustering-based algorithms have a limited understanding of the contextual information of the data, making it less reliable to analyze the topological relation between plant skeletons and plant organs.

The skeleton extraction algorithm has been proposed to tackle the above challenge in 3D plant phenotyping analysis due to its ability to depict the plant skeleton and the details of a plant structure [17]. Many researchers have applied skeleton data to analyze morphological parameters. Wu et al. [18] used Laplacian contraction to extract the skeleton of maize data by multiple-view stereo reconstruction and the nearest neighbor clustering method to separate the leaves and calculate the height, leaf length, and leaf tilt angle. Gaillard et al. [19] obtained the skeleton of sorghum by a thinning algorithm based on voxel data. On the basis of the skeleton model, they compared the classification performance between support vector machine, support vector machine with radial-based function, and multi-layer perceptron. Because maize and sorghum have simple structures, these methods ignore the influence of occlusion caused by complex structures with multiple branches on skeleton extraction. For plants with complex structures such as trees, skeleton algorithms have also been widely applied. Chaudhury and Godin [20] used B-spline curve lines with an expectation–maximization algorithm to extract and optimize the skeletons of trees. Preuksakarn et al. [21] extracted the skeleton of apple trees by the space colonization algorithm. Li et al. [22] used the *k*-means method and clustering method to separate the branches and main trunk and consequently calculated the skeleton points by breadth-first search in the undirected graph of *Toona* trees and peach trees. They constructed a skeleton model fitting trees well. However, the current methods for skeleton extraction vary because of the experimental conditions and the complexity of the plant structure, and some of them require prior knowledge and labor-intensive manual optimization. In addition, there is less occlusion between tree branches, which is significantly different from oilseed rape. The current skeleton methods cannot be generalized to analyze the oilseed rape directly. Therefore, developing an accurate method for skeleton extraction and siliques segmentation of oilseed rape point clouds is necessary for phenotyping studies.

In this study, we proposed a novel algorithm by skeletonization combined with hierarchical segmentation (SHS) on point clouds of a complete oilseed rape at the mature stage, which has not been reported to our knowledge. The goals were to (a) develop the skeleton model of oilseed rape, (b) achieve

accurate segmentation of siliques, and (c) obtain the morphological parameters of siliques for yield prediction. Eventually, we hope to introduce a noninvasive phenotyping method with morphological parameters that could be applied to real production.

## Materials and Methods

### Plant materials

Eight oilseed rape (*Brassica napus* L.) cultivars were selected for the study, with 3 biological replicates in this study. Three cultivars (c1 to c3) were grown in pots in the greenhouse at the Zijingang Campus of Zhejiang University, Zhejiang Province, China (30°18'N, 120°6'E). Five cultivars (c4 to c8) were grown in the field at the Changxing Agricultural Experiment Station of Zhejiang University, Zhejiang Province, China (30°53'N, 119°38'E). Basically, these oilseed rapes were divided into 3 types based on the structure of the branches: the plants of few-branch broom shape (FBBS), which only had few branches, and the canopy of the plant was spread out and looked like a broom; the plants of multibranch broom shape (MBBS), which had more branches than FBBS; and the plants of multibranch cylinder shape (MBCS), which had multiple branches, and the canopy of the plant looked like a cylinder. 3D point clouds of the whole plants at the siliques mature stage were acquired using a handheld 3D laser scanner. The SN, SL, and seed weight (SW) were measured. The SN was counted manually, and the SL was measured with a tape measure. The seeds of each siliques were naturally dried before weighing using an electronic balance (BSA224S, Sartorius, Germany). The details of the cultivars and measured parameters are summarized in Table 1.

**Table 1.** Summary of varieties and parameters of oilseed rape.

Growth condition	Cultivar name	Origin	Replicate #	Manually measured parameters
Green-house	ZD619 (c1)	China	3	SN, SL, and SW
	ZD622 (c2)	China	3	
	ZD630 (c3)	China	3	
	SI512 (c4)	Czechoslovakia	3	
	Bnw1.61/83 (c5)	Germany	3	
Field	8426016 (c6)	China	3	SN and SW
	CR3168 (c7)	Germany	3	
	Hu135 (c8)	China	3	

SN, SL, and SW represent the siliques number, siliques length, and seed weight, respectively.

## Point cloud data acquisition

A handheld laser scanner (PRINCE775 Inc., SCAN Tech., Hangzhou, China) with 15 red laser beams was used to harvest 3D point cloud data (Fig. 1), and the spatial resolution was 0.05 mm. The plants in pots were scanned in the laboratory, while the field-grown plants were transported to the laboratory for scanning to avoid interference from wind and other plants. The acquisition of raw data involved continuous manual scans from different perspectives. Two blackboards with marker dots provided a coordinate reference for the scanner (Fig. 1B). Because of the limited field of view, the acquired data covered the horizontal range of 240° (Fig. 1C).

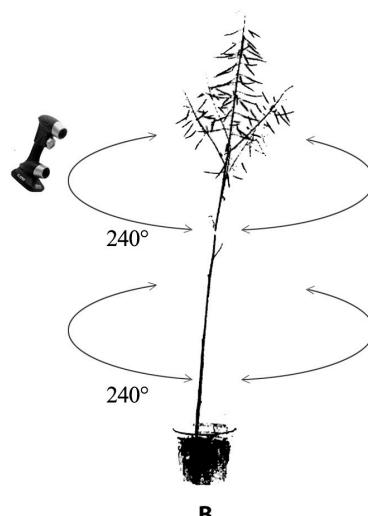
## Data processing

As we only focused on plant points for subsequent analysis, the data preprocessing was first implemented by removing the noise and pot points within the raw data (Fig. 3B). The skeleton point extraction, connection, and skeleton optimization were then performed using SHS, followed by the segmentation of siliques. Finally, siliques morphological parameters such as SN, SL, and SW were extracted based on the segmentation results. The overall workflow is shown in Fig. 2.

### Preprocessing

Preprocessing of point clouds was performed to remove outlier noise points using statistical filtering [23,24]. Because the distribution of the pot points approximated an inverted frustum of a cone, the least-squares method (LSM) was employed to fit the circle [25]. On the basis of the fitting circle with the maximum radius found by LSM, we removed the points from up to the bottom inside the fitting circles. In Fig. 3B and C,  $O_{\max}$  and  $O_1$  are the center points of the top and bottom fitting circles of the pot, respectively, while  $r_{\max}$  and  $r_1$  are the corresponding radii.  $h$  is the distance between the top and bottom as described in Eq. 1:

$$h = z_{O_{\max}} - z_{O_1} \quad (1)$$



**Fig. 1.** The acquisition of point cloud data. (A) The acquisition environment (plant and 2 blackboards with marker dots). (B) The acquisition process with the laser scanner. (C) The field-grown plants.

The fitting processing began from the bottom point  $P_1(x_1, y_1, z_{\min})$ , which had the smallest  $z$  value. The points inside the range of  $[z_{\min}, z_{\min} + \Delta d]$  ( $\Delta d$  was the step distance) were projected into the plane  $xP_1y$ , and the point  $O_1$  was calculated by LSM. This procedure was then repeated in the range of  $[z_{\min} + (n - 1)*\Delta d, z_{\min} + n*\Delta d]$  to obtain the fitting circle points  $O_n$  ( $n = 1, 2, 3, \dots$ ) until the largest fitting circle center point  $O_{\max}$  and radius  $r_{\max}$  were found. The pot data in the range of  $[z_{\min}, z_{\min} + h]$  were then removed.

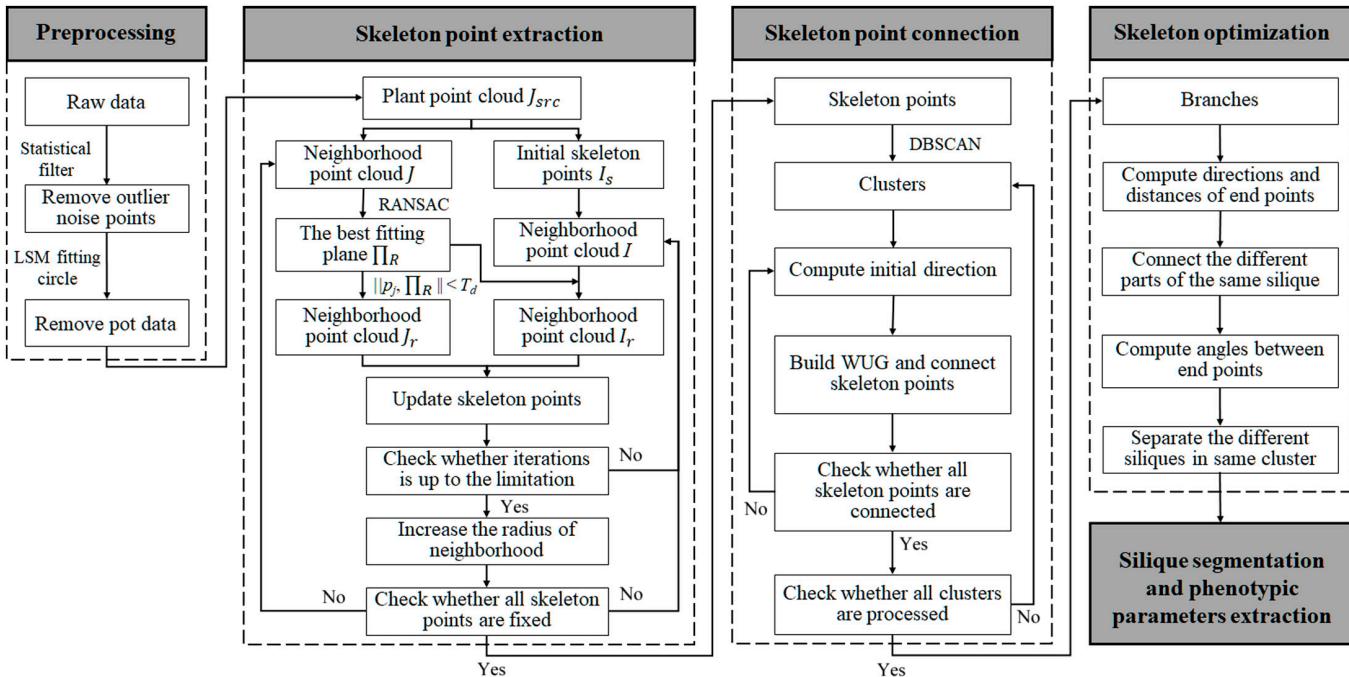
### Skeleton point extraction

The L1-median point was defined as any point that minimized the sum of Euclidean distances to all points in the dataset. It had a good antinoise performance, and the L1-median point cloud could be set as the only global center of a given point set [26]. In this study, we took the L1-median point as a local skeleton point. For a given local point set  $J = \{p_j\}$  with  $n$  points,  $j = 1, 2, 3, \dots, n$ ,  $J \subset R^3$ , the weighted center point of these points was computed iteratively by Eq. 2 to obtain the L1-median point  $p_{L1}$ .

$$p_{L1} = A(J) = \sum_{j \in J} \frac{p_j w_j}{\sum_{j \in J} w_j} \quad (2)$$

where  $w_j$  is the weight factor. As the L1-median points could present a sparse distribution where the local center points were excessively clustered together [27,28], the conditional regularization was used to shift each point into its respective geometric center position to avoid sparse distribution [18,27,28]. A repulsive force was applied between the L1-median points to keep these points at an appropriate distance from each other. The position of each skeleton point was then calculated by attractive item  $A(J)$  and repulsive item  $R(I)$  (Eq. 3). The Gaussian kernel function  $\theta(\bullet)$  and the distance function  $\|\bullet\|$  were used to give the weights of neighborhood points (Eqs. 4 and 5) so that the impact of random and uneven distribution of points could be reduced.

$$p_{L1}^{k+1} = A(J) + R(I) = \frac{\sum_{j \in J} p_j \alpha_j^k}{\sum_{j \in J} \alpha_j^k} + \tau \lambda_i^k \frac{\sum_{i \in I} (p_{L1}^k - p_i^k) \beta_i^k}{\sum_{i' \in I} \beta_i^k} \quad (3)$$



**Fig. 2.** The overall workflow of SHS.  $J_{src}$  is the source point cloud without a pot.  $J$  is the neighborhood point set in  $J_{src}$ , while  $I$  is the neighborhood point set in the skeleton point cloud  $I_s$ .  $J_r$  and  $I_r$  are the neighborhood point sets after constraint.  $\|p_j, \Pi_R\|$  is the distance between point  $p_j$  and plane  $\Pi_R$ .  $T_d$  is the distance threshold.

$$\alpha_j^k = \frac{\theta(\|p_{L1}^k - p_j\|)}{\|p_{L1}^k - p_j\|}, j = 1, 2, 3, \dots, n, p_j \in J \quad (4)$$

$$\beta_i^k = \frac{\theta(\|p_{L1}^k - p_i^k\|)}{\|p_{L1}^k - p_i^k\|^2}, i = 1, 2, 3, \dots, m, p_i^k \in I \quad (5)$$

where  $p_{L1}^{k+1}$  is the selected L1-median point, and  $k$  is the number of iterations.  $J$  is the source neighborhood points of  $p_{L1}^{k+1}$ , while  $I$  is the neighborhood L1-median points of  $p_{L1}^{k+1}$ .  $\alpha_j^k$  is the attractive factor, while  $\beta_i^k$  is the repulsion factor.  $\tau$  is a factor for balancing attraction and repulsion.  $\lambda_i^k$  is the directionality degree of  $p_{L1}^k$ , which is used to give the repulsion direction.

In each iteration, the radius of the neighborhood was increased to avoid the position deviation of the L1-median point (red dot). However, the increased radius (blue circle) could introduce interferential points (black points) from other objects, which were not in the same object as the green points and thus caused its position deflection (Fig. 4A to C). To avoid interference of the points from different objects, the calculating range was limited by the fitting plane  $\Pi_R$  (Fig. 4D). Because of the flat and long shape of a silique, we applied the random sample consensus algorithm (RANSAC) for detecting the fitting plane  $\Pi_R$  [29]. Only the points whose Euclidean distance from plane  $\Pi_R$  was less than the limited distance threshold  $T_d$  were used to calculate the skeleton point's position. The neighborhood points were updated by Eqs. 6 and 7:

$$J_R = \left\{ p_j \mid \|p_j, \Pi_R\| < T_d, p_j \in J \right\} \quad (6)$$

$$I_R = \left\{ p_i \mid \|p_i, \Pi_R\| < T_d, p_i \in I \right\} \quad (7)$$

where  $J_R$  is the point set constrained by  $\Pi_R$  from source  $J$ .  $I_R$  is the point set constrained by  $\Pi_R$  from  $I$ .  $\|p_j, \Pi_R\|$  and  $\|p_i, \Pi_R\|$  are the distances between the points and  $\Pi_R$ .

### Skeleton point connection

Because there were no topological relationships between discrete skeleton points, it was necessary to connect the skeleton points for silique segmentation. The local partition and growing algorithm were used to connect the skeleton points. The density-based spatial clustering of applications with noise (DBSCAN) performed well in finding various shapes [30], and it was first employed to roughly cluster skeleton points into several classes. A graph theoretical algorithm was proposed to connect the points in the same class. Here, we used the weighted unidirectional graph (WUG) to connect the skeleton points of the same class. WUG was used to solve the shortest-path problem, and it built connections between the skeleton points after giving an initial searching direction with the Euclidean distance as the weight factor. In general, the tangent vector of a skeleton point can represent the extension or searching direction. Because the skeleton changes smoothly and the position of the initial point  $p_i$  within the class is random, the tangent vector function can be simplified as Eq. 8. If  $p_i(x_i, y_i, z_i)$  is the end point, then  $p_i$  and the nearest point  $p_{i-1}(x_{i-1}, y_{i-1}, z_{i-1})$  are used to get a tangent vector. Otherwise, the nearest 2 points  $p_{i-1}(x_{i-1}, y_{i-1}, z_{i-1})$  and  $p_{i+1}(x_{i+1}, y_{i+1}, z_{i+1})$  are used. We set the initial direction of the initial point to be positive.

$$\begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} = \frac{y - y_{i-1}}{y_i - y_{i-1}} = \frac{z - z_{i-1}}{z_i - z_{i-1}} & \text{if } p_i = \text{end point} \\ \frac{x - x_{i-1}}{x_{i+1} - x_{i-1}} = \frac{y - y_{i-1}}{y_{i+1} - y_{i-1}} = \frac{z - z_{i-1}}{z_{i+1} - z_{i-1}} & \text{if } p_i \neq \text{end point} \end{cases} \quad (8)$$

For skeleton points in the same class, the WUG stores the weight factor between 2 points in the form of an adjacency list. Without correcting the searching direction, all the weights are positive, with  $d_{ij} = d_{ji}$  (Fig. 5A), indicating that the connection of skeleton points depends on the absolute value of distance, which may cause the wrong connection lines of skeleton points. For example, if weight  $d_{A_1 D_1}$  is smaller than  $d_{A_1 B_1}$  in sub-skeleton 1 (Fig. 5C), then  $A_1$  would be wrongly connected with  $D_1$ . Therefore, the searching directions of each point are corrected by previous connecting points, and some weights in the adjacency list changed to be negative (Fig. 5B). We defined the out-degree (OD) and in-degree (ID) to limit the searching direction. Given a point  $p$ , OD is the number of lines coming out of  $p$ , while ID is the number of lines coming into  $p$ . End points only have one edge, indicating  $OD = 1$  or  $ID = 1$ . For non-end points, we limited the points to have only 2 edges, indicating  $OD = 1$  and  $ID = 1$ . As shown in Fig. 5C, sub-skeleton 1 with yellow points and sub-skeleton 2 with green points are 2 classes divided by DBSCAN. They are adjacent in the 3D space, and the WUG searches only the points within the same class. In sub-skeleton 1, the initial point  $A_1$  is a non-end point and only connects the nearest 2 points,  $B_1$  and  $C_1$ . If the initial point is an end point such as  $A_2$  in sub-skeleton 2, then it would only connect to the nearest point  $B_2$ , and then  $C_2$  would be connected by  $B_2$ .

### Skeleton optimization

Some siliques skeletons were still incorrectly connected because of the limited clustering performance of DBSCAN. These incorrect cases would reduce the accuracy of siliques segmentation. As shown in Fig. 6A, one siliques or stem is clustered into 2 classes (yellow and green). Because the skeleton changes smoothly, the distance ( $d_e$ ) between end points and the angles ( $\gamma_e^1, \gamma_e^2$ ) between end points' tangent vectors and end points' connection lines (red dotted) are small. Thus, we calculated the average distance ( $d_{sk}$ ) between connected skeleton points and

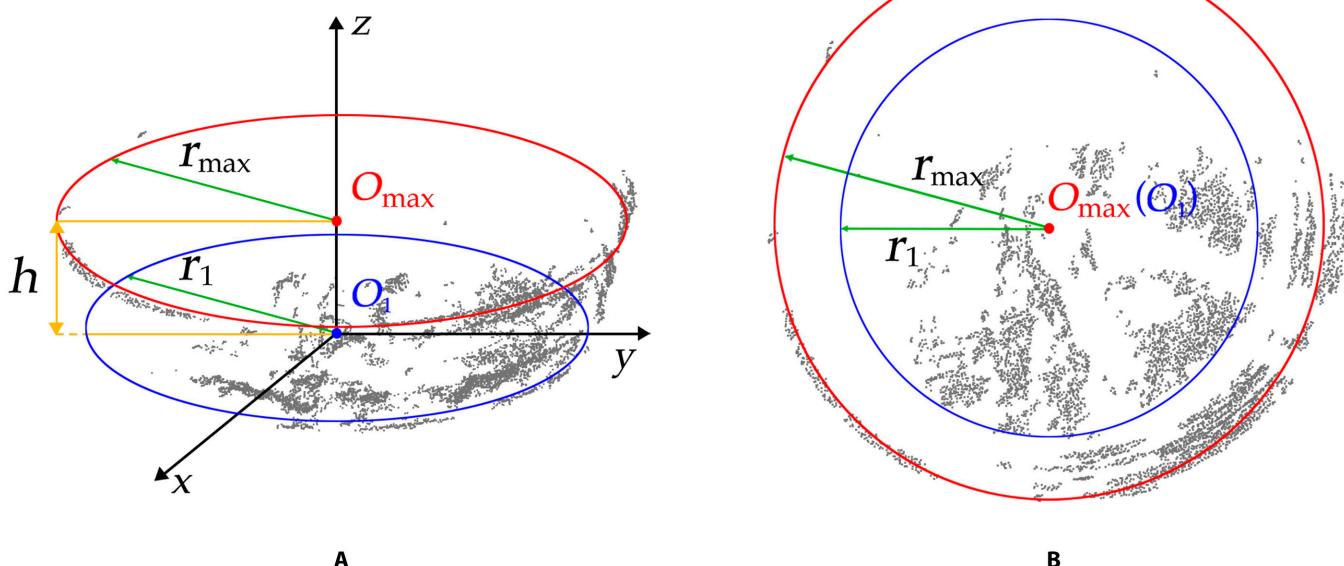
connected the 2 adjacent classes that met the following 2 requirements: (a) Both  $\gamma_e^1$  and  $\gamma_e^2$  were smaller than  $15^\circ$ ; (b)  $d_e$  was smaller than  $5 \times d_{sk}$ . In addition, different siliques could also be clustered into the same class. Thus, we set the angles ( $\gamma_{sk}^1, \gamma_{sk}^2$ ) between each point's 2 edges larger than  $165^\circ$  to divide nearby siliques into different classes; only if  $\gamma_{sk}^1$  and  $\gamma_{sk}^2$  were both larger than  $165^\circ$ , then the sub-skeletons were considered as the same siliques class (Fig. 6B).

### Siliques segmentation and morphological parameter extraction

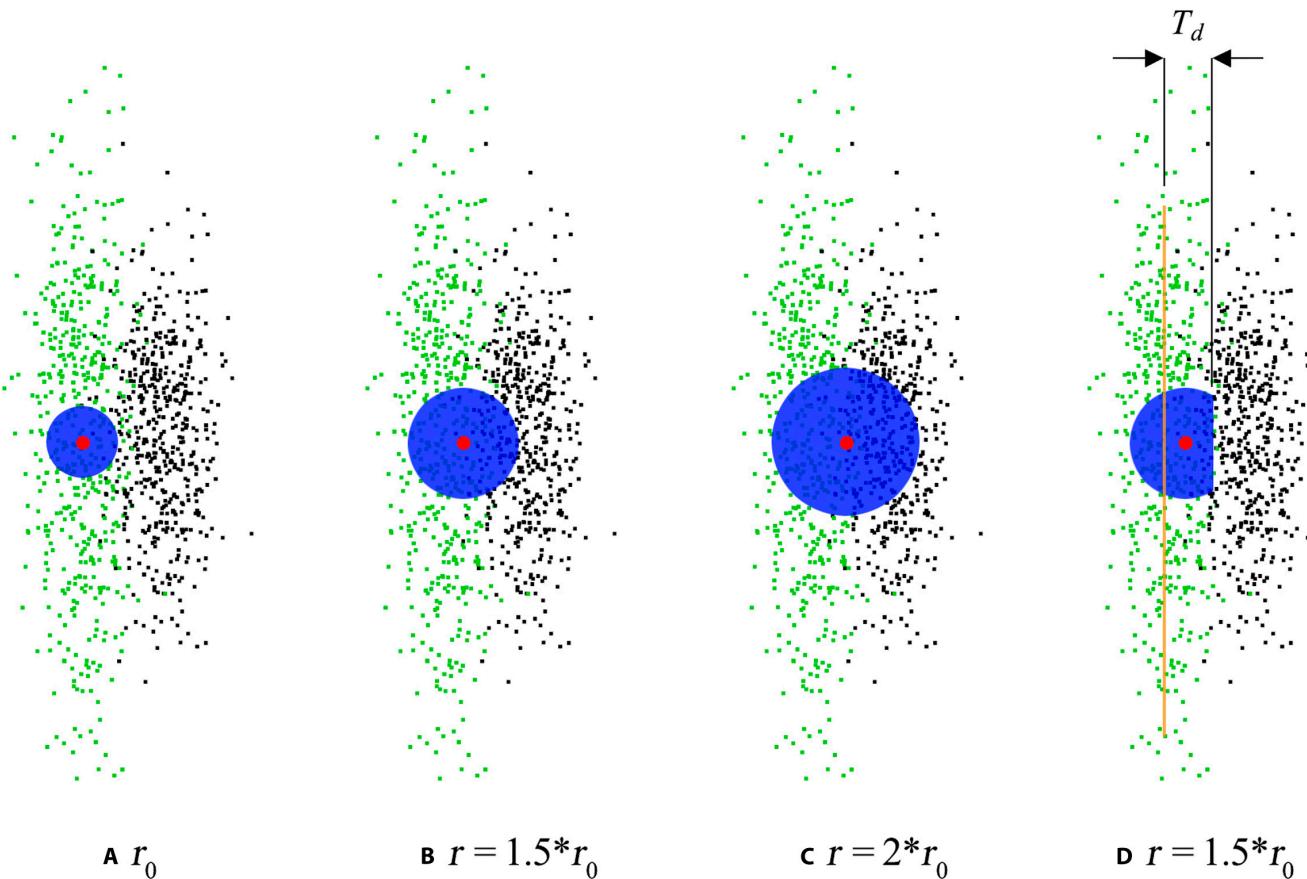
After skeletonization and optimization, each siliques or stem had its own unique corresponding sub-skeleton, and the skeleton points were ordered under the given direction in the sub-skeletons. On the basis of the assumption of the normal distribution or near-normal distribution of SL [6,31,32], we used siliques with SL within the 95% confidence interval as the effective siliques ( $SL_{ef}$ ) and counted the SN. Additionally, siliques abortion was a normal phenomenon for oilseed rape [33], and the stunted siliques were tenuous or small in shape and without seeds. Thus, siliques shorter than 15 mm were removed. Here, the sub-skeleton length  $l_{\{bj\}}$  was calculated as the sum of the distances of all adjacent skeleton points in this sub-skeleton  $\{b_j\}$  (Eq. 9). Because of the influence of extreme values such as long stems, only 60% of the middle length data was used to calculate the mean value  $\mu$  of the normal distribution of SL (Eq. 10). After removing aborted siliques,  $SL_{ef}$  was defined as Eq. 11. The total SL was the sum of  $SL_{ef}$  (Eq. 12).

$$l_{\{b_j\}} = \left\{ \sum \|p_i - p_{i+1}\|, p_i \in \{b_j\}, i = 1, 2, 3, \dots, m-1; j = 1, 2, 3, \dots, n \right\} \quad (9)$$

$$\mu = \frac{\sum_{j=int(0.2m)}^{0.8m} l_{\{b_j\}}}{0.6m} \quad (10)$$



**Fig. 3.** The pot point cloud and fitting circles, including (A) the front view and (B) the top view.  $r_{max}$  is the maximum radius of the fitting circle with center point  $O_{max}$ .  $r_1$  is the radius of the bottom circle with center point  $O_1$ .  $h$  is the distance between fitting circles.



**Fig. 4.** L1-median points of local regions with different radii.  $r_0$  is the initial radius. The red dot is the L1-median point. The blue circle is the radius of the local region, and green and black points represent 2 different objects. The yellow line represents the RANSAC fitting plane.  $T_d$  is the limited distance threshold.

$$\text{SL}_{ef} \in \left[ \max\left( 15 \text{ mm}, \mu - 1.96 * \frac{\sigma}{\sqrt{n}} \right), \mu - 1.96 * \frac{\sigma}{\sqrt{n}} \right] \quad (11)$$

$$\text{Total SL} = \sum_{i=0}^{m'} \text{SL}_{ef} \quad (12)$$

where  $m$  is the number of skeleton points in the sub-skeleton, and  $p_i$  is a skeleton point.  $\|p_i - p_{i+1}\|$  represents the Euclidean distance between  $p_i$  and  $p_{i+1}$ .  $\sigma$  is the standard deviation (SD).  $m'$  is the number of effective siliques.

Furthermore, the classified siliques were obtained by searching the neighborhood source points of the RANSAC constraint plane, and the siliques volume (SV) was calculated based on the siliques points. As shown in Fig. 7, the green points are the source points of the siliques, and the red points are the skeleton points. In Fig. 7A and B,  $\vec{v}$  is the direction of a skeleton point  $A$ , and plane  $\alpha$  is the normal plane of  $A$ . Points  $B$  and  $C$  are the adjacent skeleton points of  $A$  (Fig. 7B and D).  $d_1$  and  $d_2$  are the corresponding distances. The source points inside the range of  $[A - d_1/2, A + d_2/2]$  were projected onto plane  $\alpha$  (Fig. 7C). Assuming that the distribution of the projected points was rectangular,  $W$  and  $L$  were the width and length of the rectangle, and the partial volume of skeleton point  $A$  was calculated (Eq. 13). All the partial volumes were added to obtain a single

siliques (Eq. 14). Similarly, we also calculated the total SV by Eq. 15:

$$V_A = L * W * \frac{d_1 + d_2}{2} \quad (13)$$

$$\text{SV} = \sum_{i=1}^n L_i * W_i * \frac{d_1^i + d_2^i}{2} \quad (14)$$

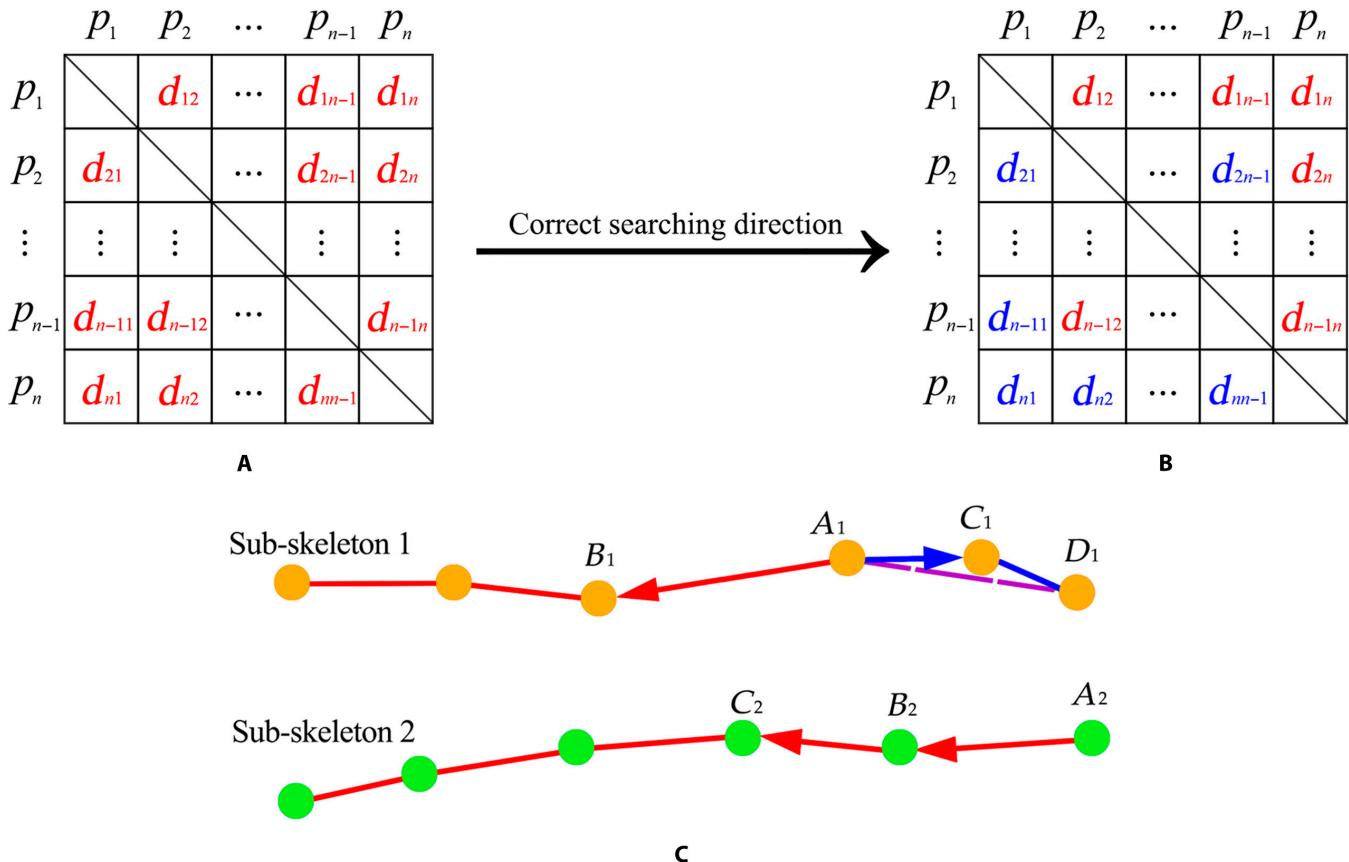
$$\text{Total SV} = \sum_{j=1}^{m'} V_j \quad (15)$$

where  $n$  is the skeleton point number and  $m'$  is the number of effective siliques.

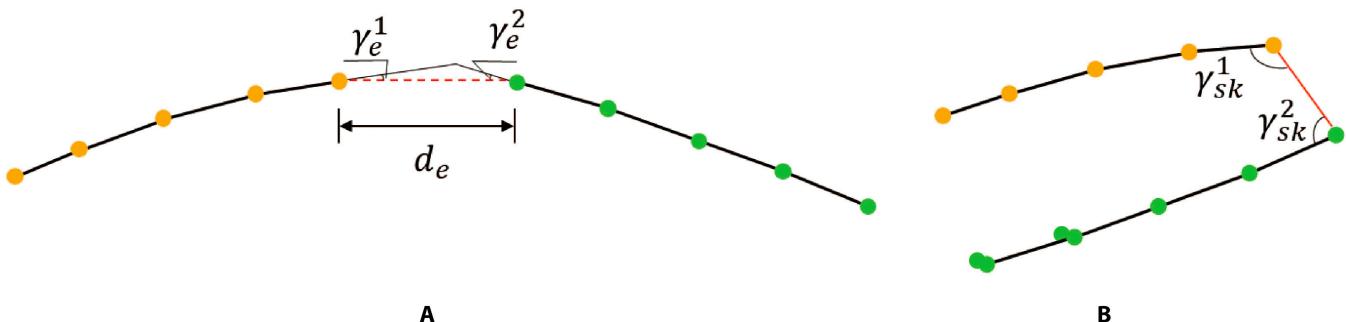
### Validation and statistical analysis

The architecture of oilseed rape was related to the branch. We denoted 3 types of branches as follows: the first branch ( $B_{1st}$ ), also called the main stem; the second branch ( $B_{2nd}$ ), grown on the first branch; and the third branch ( $B_{3rd}$ ), grown on the second branch [34].

To validate the performance of the SHS algorithm, the recall ( $Re_{EL}$ ) was computed by Eq. 16, representing the proportion of estimated siliques ( $SN_E$ ) to siliques counted by the laser point



**Fig. 5.**(A) The adjacency list without the given initial direction, and all the weights are positive (red). (B) The adjacency list with the given initial direction, and the weights are negative (blue).  $p_1, p_2, \dots, p_n$  are the skeleton points.  $d_{ij}, i, j=1, 2, \dots, n, i \neq j$ , is the Euclidean distance weight factor. (C) The connection of skeleton points of different organ skeletons by DBSCAN and WUG. Sub-skeleton 1 and sub-skeleton 2 are 2 classes. The red lines indicate that the weights are positive, while the blue lines indicate that the weights were negative. In sub-skeleton 1,  $A_1$  is the non-end point, and WUG search directions are  $\overrightarrow{A_1C_1}$  and  $\overrightarrow{A_1B_1}$ ;  $A_1D_1$  (purple dotted line) is the wrong connection line. In sub-skeleton 2,  $A_2$  is the end point, and the WUG search directions are  $\overrightarrow{A_2B_2}$  and  $\overrightarrow{B_2C_2}$ .



**Fig. 6.** Two incorrect cases after skeletonization. (A) One siliques or stem is clustered into 2 classes (yellow and green), and the red dashed line is the wrong unconnected line. (B) Two siliques or stems (yellow and green) are clustered into one class, and the red solid line is the wrong connected line.  $\gamma_e^1$  and  $\gamma_e^2$  are the angles between end points' tangent vectors and end points' connection lines.  $d_e$  is the distance between 2 end points from different classes.  $\gamma_{sk}^1$  and  $\gamma_{sk}^2$  are the angles between each point's 2 edges.

cloud ( $SN_L$ ). In addition, the SN manually measured ( $SN_M$ ) was taken as the ground truth, and the recall ( $Re_{LM}$ ) was computed by Eq. 17 to evaluate the quality of the laser-scanned point cloud, representing the ratio of  $SN_L$  to  $SN_M$ . The recall ( $Re_{EM}$ ) computed by Eq. 18 represents the ratio of  $SN_E$  to  $SN_M$ .

$$Re_{EL} = \frac{SN_E}{SN_L} \times 100\% \quad (16)$$

$$Re_{LM} = \frac{SN_L}{SN_M} \times 100\% \quad (17)$$

$$Re_{EM} = \frac{SN_E}{SN_M} \times 100\% \quad (18)$$

Additionally, to test the correlation between SHS estimations and manual measurements, the coefficient of determination

( $R^2$ ) and the root mean squared errors (RMSE) were calculated. Because mature siliques were easily broken during the transportation of oilseed rape, measuring the length of these siliques was difficult. Therefore, we used the total SL of a single plant grown in the greenhouse to verify the accuracy of SL estimation. According to linear regression analyses, we used the correlation

coefficient  $R$  to explore the relationship between SL, SN, and SV and the yield of a plant (YOP), where YOP was denoted as the total SW in siliques.

The proposed algorithm was developed by Point Cloud Library [35] and Open3D Library [36] in the Visual Studio 2015 and PyCharm 2011 development platforms. The algorithm was implemented on a desktop workstation with the configuration of an Intel Core i7 processor and 16 GB of memory.

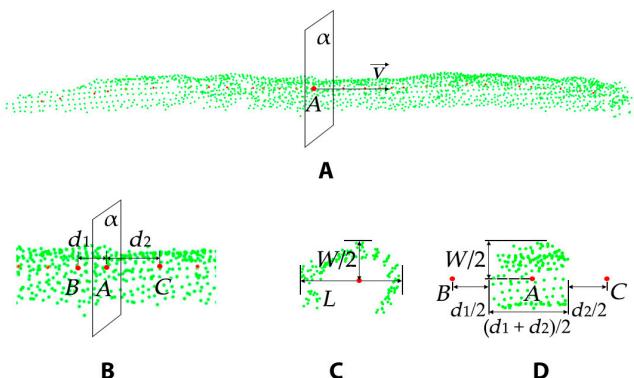
## Results

### The visualization of plant architecture

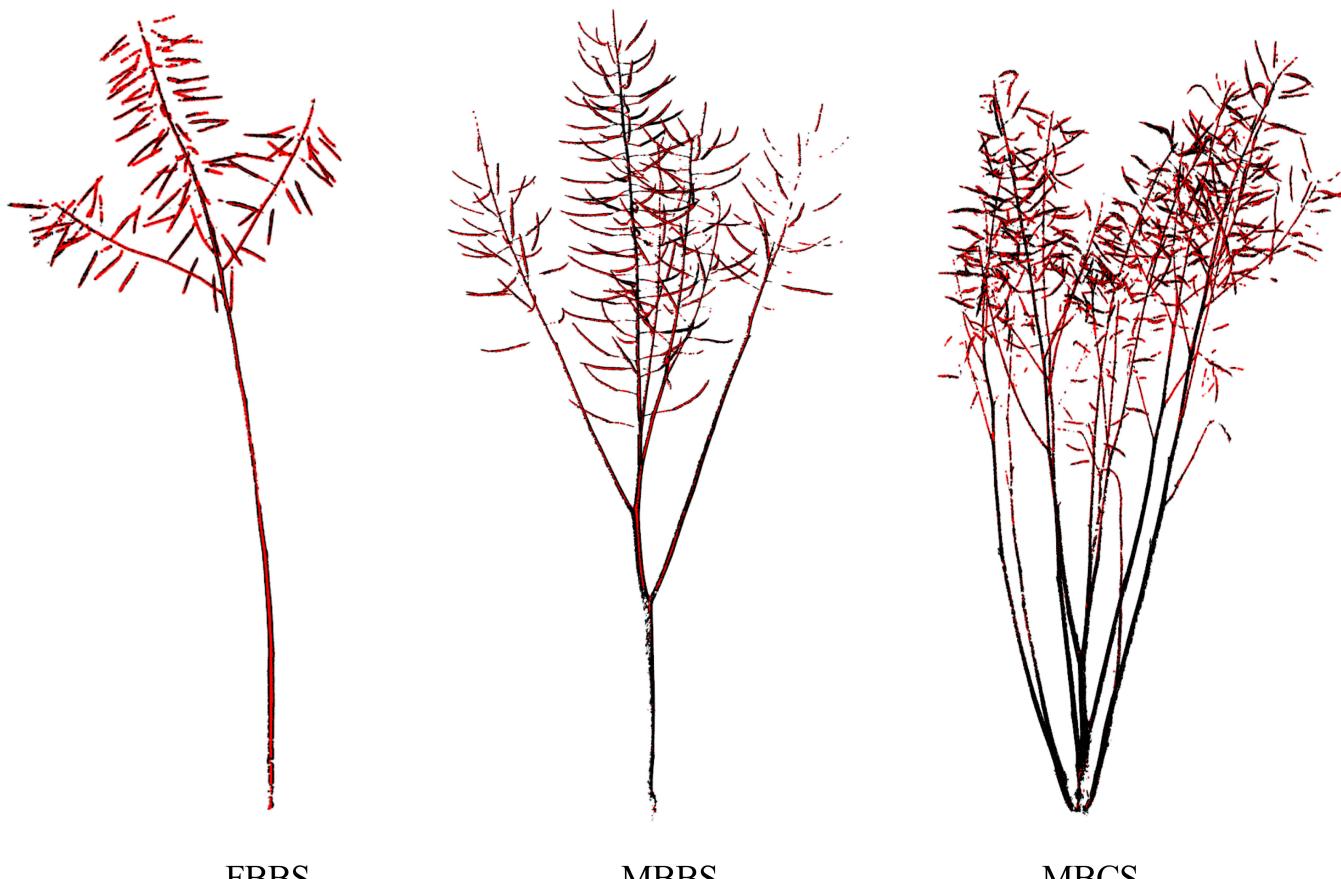
Figure 8 presents 3 typical plant architectures of oilseed rapes. There were obvious differences among the 3 types. The plants of FBBS only had one  $B_{1st}$  and few or no  $B_{2nd}$ s and  $B_{3rd}$ s, and the canopy was sparse. The plants of MBBS also only had 1 or 2  $B_{1st}$ s but more  $B_{2nd}$ s and  $B_{3rd}$ s with a dense canopy. Compared with FBBS and MBBS, the plants of MBCS had more  $B_{1st}$ s with multiple second and third branches. The siliques canopy was evenly distributed and cylindrical.

### Siliques segmentation

We counted the lengths of sub-skeletons of the 3 plant architectures, and the sub-skeletons included stems/branches and siliques. Through manual verification, 95% of the sub-skeletons over 200 mm were long branches, which was significantly different from that of siliques (Fig. 9A). After removing the branch



**Fig.7.** The process of volume calculation. (A) Source point (green) cloud and skeleton points (red). (B) Partial data around point A. (C) The distribution of projected neighborhood points of A. (D) Points inside the space that point A represents. Above all, A, B, and C are skeleton points.  $\vec{v}$  is the direction of skeleton point A, and plane  $\alpha$  is the normal plane of A.  $d_1$  is the distance between A and B, while  $d_2$  is the distance between A and C.  $W$  is the width, and  $L$  is the length.



**Fig.8.** Examples of source point clouds and skeleton points (red dots) of 3 plant architectures of oilseed rape. FBBS, MBBS, and MBCS represent the few-branch broom shape, the multibranch broom shape, and the multibranch cylinder shape, respectively.

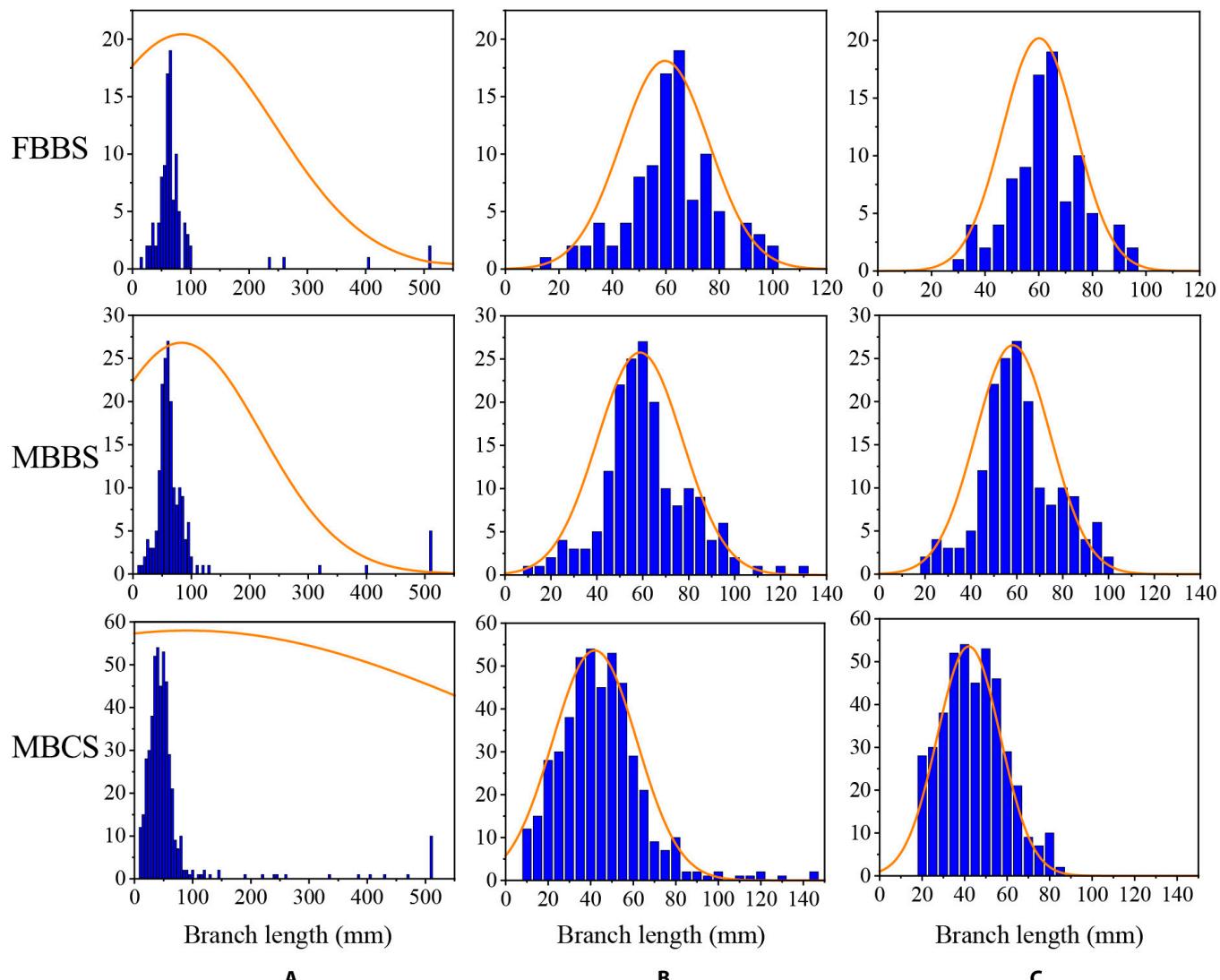
sub-skeleton, the length distribution of all sub-skeletons of a single plant followed the assumption of a normal distribution or near-normal distribution (Fig. 9B). In addition, the aborted siliques without seeds were removed. Finally, we obtained the effective siliques in a 95% confidence interval of length distribution, which also followed the above assumption (Fig. 9C).

At the same time, these results proved that our materials were conventional, and the results were universal. The effective siliques were rendered by different colors (Fig. 10).

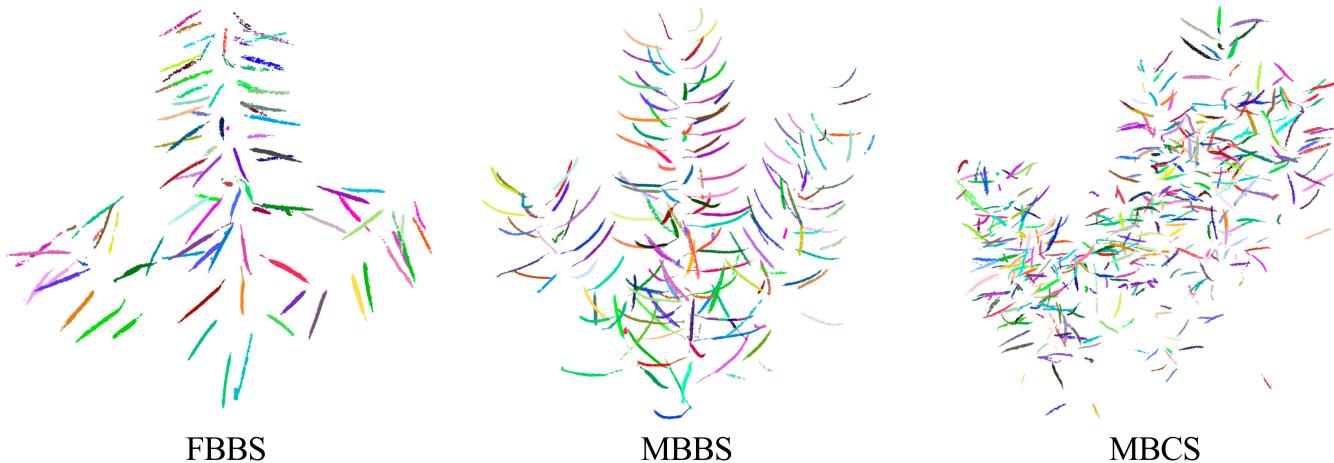
On the basis of the differences in branches, we obtained 13, 6, and 5 plants of FBBS, MBBS, and MBCS, respectively. There were significant differences between these 3 plant architectures (Table 2). The plants of FBBS only had one average  $B_{1\text{st}}$ , 3.69 average  $B_{2\text{nd}}$ s, and almost no  $B_{3\text{rd}}$ s. There were more second and third branches in the MBBS plants, which had 6.33 average  $B_{2\text{nd}}$ s and 4.33  $B_{3\text{rd}}$ s. The numbers of the first, second, and third branches of MBCS plants was the highest, with the average values of 4.20, 14.20, and 6.20, respectively. In addition, the

average SN of MBCS plants were the largest at 549.60, while those of MBBS and FBBS were 316.17 and 115.54, respectively. More details about the siliques and branches of each plant are presented in the Supplementary Materials.

Because of the SHS algorithm being carried on the laser point cloud, the estimation of the SN was affected by the data quality of the point cloud and the performance of the algorithm. Therefore, the results were evaluated from these 2 perspectives. For all cultivars, the average recall ( $Re_{LM}$ ) between  $SN_L$  and  $SN_M$  was at a high level of 92.23% with an SD of 8.82, indicating that most of the plants were well scanned, and the laser point cloud of siliques was complete. The average recall ( $Re_{EL}$ ) between  $SN_E$  and  $SN_L$  was 90.90% with a low SD of 6.07, indicating the accurate and stable performance of the SHS algorithm. On the basis of the data of good quality and algorithm performance, the  $SN_E$  was well correlated with  $SN_M$ , and the average recall ( $Re_{EM}$ ) between  $SN_E$  and  $SN_M$  was 84.23% with an  $R^2$  of 0.922 (Fig. 11A).



**Fig. 9.** Histograms of the sub-skeleton length of 3 plant architectures. FBBS, MBBS, and MBCS represent the few-branch broom shape, the multibranch broom shape, and the multibranch cylinder shape, respectively. (A) The length distribution of all sub-skeletons, including branches and siliques, and the sub-skeletons longer than 500 mm were combined into one group for display convenience. (B) The length distribution of sub-skeletons without long stems. (C) The length distribution of effective siliques.



**Fig.10.** The segmented and classified siliques point cloud of 3 plant architectures. Each siliques was rendered by a unique color. The FBBS, MBBS, and MBCS represented 3 plant architectures: the few-branch broom shape, the multibranch broom shape, and the multibranch cylinder shape, respectively.

**Table 2.** Segmentation results of siliques with different plant architectures.

Plant architecture		$NB_{1st}$	$NB_{2nd}$	$NB_{3rd}$	$SN_M$	$SN_L$	$SN_E$	$Re_{LM} (\%)$	$Re_{EL} (\%)$	$Re_{EM} (\%)$
FBBS	Average	1	3.69	0.08	115.54	108.08	102.23	94.97	93.91	82.50
	SD	0	1.64	0.27	65.61	60.23	58.81	5.85	4.85	7.42
MBBS	Average	1.33	6.33	4.33	316.17	295.33	265.00	93.45	90.27	85.43
	SD	0.47	0.75	2.29	114.06	100.50	88.89	3.69	3.80	6.13
MBCS	Average	4.20	14.20	6.20	549.60	446.8	374.00	82.33	83.81	69.22
	SD	1.17	3.66	2.14	93.80	76.41	63.83	12.02	4.86	12.38
Average	/	/	/	/	/	/	/	92.23	90.90	84.23
SD	/	/	/	/	/	/	/	8.82	6.07	11.54

$SN_M$ , siliques number measured manually;  $SN_L$ , siliques number counted by laser point cloud;  $SN_E$ , siliques number estimated by our method; SD, standard deviation;  $Re_{LM}$ , recall =  $SN_L/SN_M$ ;  $Re_{EL}$ , recall =  $SN_E/SN_L$ ;  $Re_{EM}$ , recall =  $SN_E/SN_M$ ;  $NB_{1st}$ , the number of first branches;  $NB_{2nd}$ , the number of second branches;  $NB_{3rd}$ , the number of the third branches.

Moreover, the structure of plants affected the data quality and performance of the algorithm. The plants of FBBS and MBBS architectures were well scanned with the point cloud of high completeness, and the siliques were well-segmented, so the average  $Re_{LM}$ s were of high value (FBBS was 94.47%; MBBS was 93.45%), as well as average  $Re_{EL}$ s (FBBS was 93.91%; MBBS was 90.27%). In contrast, the plants of MBCS architecture had significant differences, with more second branches, third branches, and siliques. Many siliques data were missing in the scanning process, causing low  $Re_{LM}$ s, which affected the data quality and the performance of the algorithm, decreasing the average  $Re_{EL}$  to 83.81%. Additionally, the total SL was accurately estimated with an  $R^2$  of 0.934 (Fig. 11B).

### The relationship between siliques morphological parameters and the yield of a plant

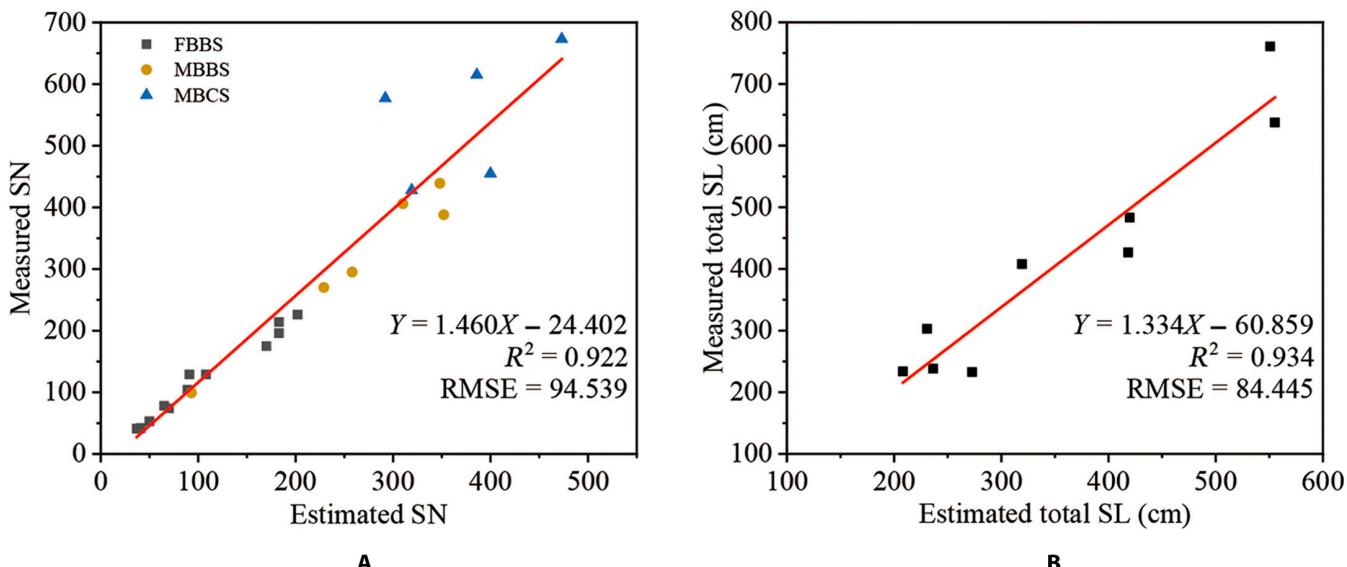
The differences in plant architectures affected YOP (Fig. 12A to C). The structures of plants of FBBS and MBBS were similar, and the main difference was the branch number, so the YOP

of MBBS was higher than that of FBBS. The structure of MBBS was significantly different and its yield was higher. The SN, total SL, and total SV estimated by our method were significantly positively correlated with YOP, and the  $R$  values were 0.935, 0.916, and 0.897, respectively.

## Discussion

### Acquisition of high-quality point cloud data

Currently, the phenotypic identification related to oilseed rape yield is still dependent on manual methods, which are time-consuming and labor-intensive, and the results are greatly influenced by subjective factors. In contrast, methods based on 3D imaging have the advantages of high automation and more objective results. Among them, the laser point cloud is accurate and reliable and has been widely applied for agricultural research, such as plant reconstruction and parameter extraction [21,22]. However, the laser point cloud scanning process is tedious. Because the plants are nonrigid objects that are irregularly



**Fig. 11.** The relationship between morphological parameters derived by manual measurement and the SHS. The morphological parameters include SN and total SL. (A) The relationship of SN of all cultivars, including 3 plant architectures: FBBS, MBBS, and MBCS. (B) Comparison of the total SL of cultivars (c1 to c3) grown in a greenhouse.

deformed by the wind, the scanning process requires a windless environment, which greatly affects the acquisition efficiency. In addition, siliques breakage and falling occurring during plant transfer and siliques abortion also affect the results of morphological analysis, resulting in incorrect or missing data. Therefore, the optimization of point cloud data acquisition is essential. RGB-D cameras and stereo cameras have shown potential for field-based plant phenotyping [17,37,38], but these sensors have shortcomings in accuracy and stability. It is necessary to improve the resolution and accuracy of these sensors and develop corresponding 3D imaging algorithms for the acquisition performance. To achieve accurate, efficient, and invasive yield estimation in the field, a field phenotyping platform with integrated multiple sensors needs to be further investigated.

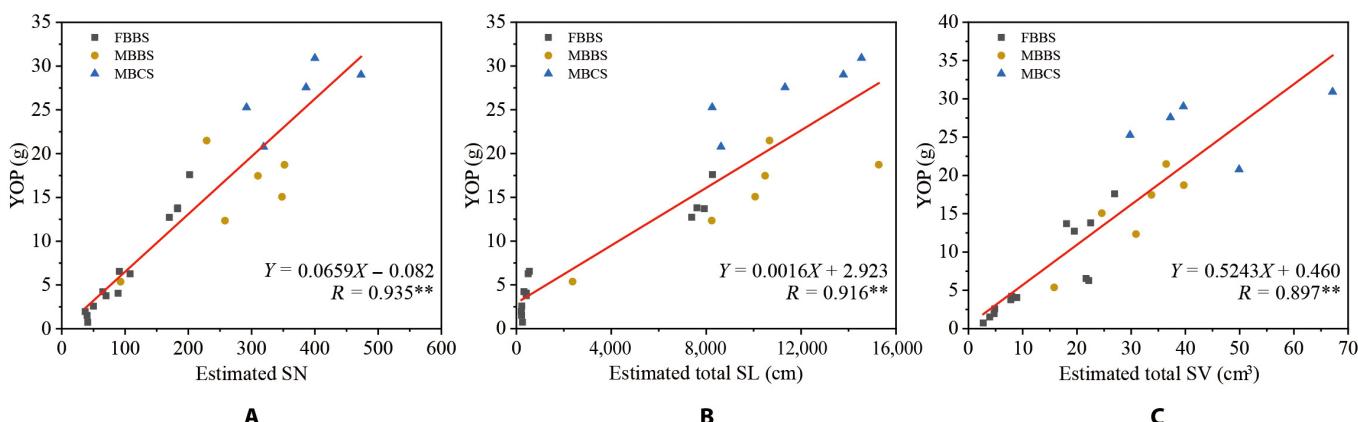
Occlusion has an important effect on the quality of the point clouds. It is one of the main limitations of acquiring plant parameters by computer vision technology and is always accompanied by missing data. The probability of occlusion is different for plants with different structures, and complex structures are more likely to introduce the occlusion problem. There is less occlusion and more space between organs in plants with spread-out structures. In our study, MBBS and FBBS both incorporate spread-out structures. Therefore, although the plants of MBBS have more branches, the data quality was similar to that of FBBS plants. For the MBCS plants, the situation was more complicated. If the branches were clustered (Fig. 13A), then there would be serious occlusion between different organs, and the quality of the laser point cloud was poor. In contrast, if the branches were spread out (Fig. 13B), then the occlusion was not serious, and we could still obtain complete data. The  $Re_{LM}$  of c7-1 was only 60.01%, while that of c8-1 was 94.29%. More parameters need to be applied for more detailed morphological quantitative analysis, such as the dispersion of different levels of branches. Similar problems have also been found in other studies, in which occlusion makes it challenging to segment leaves and other organs [39,40], thereby affecting the inversion of the vegetation index [41,42] and the estimation of yield [43]. The occlusion effect is an emergent problem that needs to be

approached for the future application of 3D phenotyping for plant populations or field scales.

### Applicability of the SHS algorithm in 3D phenotyping

The skeleton has been proven to be a piece of effective structural information for plant organ segmentation [44,45], and it has been applied to monocot plants such as sorghum and maize [44,45]. However, these methods demand the manual inclusion of prior knowledge, such as height information and initial growth points [46,47]. In addition, the plant material used in such earlier studies has been simple in structure, generally with only one main stem and no multiple branches. In contrast, oilseed rape is a dicotyledonous plant, and its structure is complex and varied. The SHS algorithm proposed in this study achieves skeleton extraction and optimization by combining the ideas of the local clustering segmentation and the growing algorithms for skeleton points and realizes the segmentation of siliques on this basis. Our method uses DBSCAN to cluster the skeleton into multiple classes and then applies the similarity of local geometric information of the same skeleton for optimization, so it can reduce the effect of other organs on its skeleton and has good performance on plants with different structures (see the Supplementary Materials). Moreover, our method requires no manual prior knowledge and is suitable for oilseed rapes of various postures. In addition, for the oilseed rapes grown in the field, there are more factors such as the interference between different plants and organs, which will be the limitation of the application of our method. It is necessary to establish the relationship model between the individual scale and the group scale to better apply our method.

In recent years, deep learning has been applied for the segmentation of plant organ point clouds with good results [48–50]. Compared with deep learning algorithms, SHS is driven by geometric rules rather than data, so it has no requirement of training a large amount of data. However, the methods based on geometric rules require researchers to have a comprehensive



**Fig.12.** The relationship between the yield of a plant and estimated morphological parameters of all cultivars with 3 plant architectures, including the FBBS, the MBBS, and the MBCS. The parameters include SN (A), total SL (B), and total SV (C). \*\* indicates statistical significance at 0.01.

understanding of object morphology. For the structural diversity of oilseed rape, many special cases have been easy to ignore. In our study, the method is based on a common assumption that siliques are flat and long. Because of genetic and environmental effects, some siliques become short and wide, which will be treated as aborted siliques by our method [51]. For other plants, we need to adjust the method according to the morphological characteristics of their organs. For example, the rice tillers are slim and long, which are similar to oilseed rape branches, and our method can be applied directly for extracting tillers. As for the cotton boll, we need to design strategies to figure out the globular organs. Deep learning is driven by data, which makes it easier to understand geometric rules and extract features of morphological structures. Researchers' tasks are to build big datasets and design robust frameworks. It is necessary to collect data on various morphological architectures and cultivars of oilseed rape plants for further research. Nevertheless, on the basis of the performance of the SHS algorithm, it can also be applied to the task of data labeling in deep learning research.

### Structural parameter relationships with oil seed yield

Seed yield is directly determined by the SN, seed number in siliques and thousand SW, and is also indirectly affected by siliques traits [52–55]. For example, SL is positively correlated with seed number but poorly correlated with seed yield [52,53]. Plants differ greatly in their structures of branches and the numbers of branches, and the SN of lateral branches of a single plant has a high correlation with yield [56,57]. Moreover, many studies have shown that siliques structural parameters are influenced by quantitative trait loci and the environment [6,32,51,58–61]. In our study, we reached a similar conclusion that the environment had a great influence on the plant structure, and the structural differences further affected the yield of each plant. The greenhouse environment was controllable, but the cultivation and fertilization treatments were not optimal. Therefore, the greenhouse cultivars (c1 to c3) were not only small in SN and yield but also had few differences in structure. In contrast, there were many factors in the field environment, resulting in large differences in plant structure. There were 3 architectures of FBBS, MBBS, and MBCS in field cultivation. However, the reasonable control of fertilization and cultivation in the field

resulted in a higher YOP. Moreover, the number of siliques, and branches and the levels of branches were different in different architectures. Among them, the plants of MBCS architecture had the most multiple branches, which resulted in more siliques and higher yield. The architecture of MBBS plants was similar to that of FBBS plants, but the former had more multiple branches and, correspondingly, more siliques and higher yields. Despite the influence of the environment, the traits of siliques (SN, SL, and SV) showed a high correlation with yield. Moreover, affected by environmental factors, such as high temperature, seed abortion can occur easily in effective siliques [62,63]. It is difficult to provide information on seed abortion inside siliques by non-invasive 3D methods at the plant scale, which affects the application of siliques parameters (especially length and volume) for yield estimation. In addition, the breakage of mature siliques resulting in seed loss also affects yield estimation.

It is worth mentioning that, because of the limitations of traditional artificial or 2D image methods, the current research on oilseed rape mainly focuses on the relationship between SN, SL, and yield, while the correlation analysis of other phenotypic traits, such as SV, is limited. In this study, 3D imaging technology was used for the digitalization of plants, which could show the structural characteristics of plants more intuitively and provide for the possibility of quantitative analysis of plant structural phenotypes. In addition to SN, SL, and SV, we can further explore the phenotype information of oilseed rape to quantify traits such as the thickness of the siliques canopy (see the Supplementary Materials), which are also parameters of particular interest to breeders [64]. Furthermore, 3D imaging of the total plant can retain the complete physical structure information of the plant, which provides possibilities for traceability and future exploration of potential traits.

### Conclusions

Conventional skeleton extraction requires prior knowledge or ignores the occlusion caused by complex structure limits to simple plants, which is not suitable for obtaining siliques traits of oilseed rape with multiple branches. To overcome these problems, we proposed an accurate SHS method for morphological parameter extraction of oilseed rape siliques. After removing noise and non-plant data, the point cloud was skeletonized and hierarchically segmented to determine the siliques followed by morphological trait extraction. The performance



**Fig.13.** The plants of MBCS architecture. (A) CR3168 (c7)-1; (B) Hul35 (c8)-1.

of SHS has good accuracy, adaptability, and robustness for the cultivars of oilseed rape plants with 3 different structures: FBBS, MBBS, and MBCS. The average recall ( $Re_{EL}$ ) between  $SN_E$  and  $SN_L$  was 93.91%, 90.27%, and 83.81%, respectively. For all cultivars, the  $R^2$  of  $SN_E$  and ground truth was 0.922. In addition, the traits SN, total SL, and total SV of a single plant extracted by SHS were significantly correlated with YOP, and the  $R$  values were 0.935, 0.916, and 0.897, respectively. Therefore, this non-invasive method could be applied for yield estimation and phenotyping analysis and provide technical support for oilseed rape breeding. Further work should focus on enriching the 3D data of oilseed rape with a larger number of different structures and combining deep learning for research.

## Acknowledgments

**Funding:** This work was funded by the Key R&D Program of Zhejiang Province, China (2021C02057); the National Key R&D Program of China (2021YFD2000104); the Key R&D projects in Huzhou City (2021ZD2037); and the Fundamental Research Funds for the Central Universities (226-2022-00217).

**Author contributions:** Z.M. and H.C. designed the experiment. Z.M. developed the algorithm and analyzed the data. Z.M. and R.D. performed the experiments. Z.M., R.D., J.X., D.S., H.F., L.J., and H.C. drafted the manuscript. All authors provided critical comments on the results and discussion. **Competing interests:** The authors declare that there is no conflict of interest regarding the publication of this article.

## Data Availability

The datasets during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## Supplementary Materials

Fig. S1. Source point clouds and skeleton points (red dots) of 8 cultivars.

Fig. S2. Branch length histograms of 8 cultivars.

Fig. S3. The partial region data of the plants.

Fig. S4. The thickness of siliques canopy.

Table S1. Siliques segmentation results of all plants with different plant architectures.

## References

1. Friedt W, Snowdon R, Ordon F, Ahlemeyer J. Plant breeding: Assessment of genetic diversity in crop plants and its exploitation in breeding. *Prog Bot.* 2007;68:151–178.
2. Nowosad K, Liersch A, Poplawska W, Bocianowski J. Genotype by environment interaction for seed yield in rapeseed (*Brassica napus* L.) using additive main effects and multiplicative interaction model. *Euphytica.* 2016;208(1):187–194.
3. Oleksy A, Zajac T, Klimek-Kopyra A, Pustkowiak H, Jankowski K. Relative siliques position in a crop layer as an indicator of yield and quality in winter rape. *Pak J Agric Sci.* 2018;55(4):727–738.
4. Wang C, Hai J, Yang J, Tian J, Chen W, Chen T, Luo H, Wang H. Influence of leaf and siliques photosynthesis on seeds yield and seeds oil quality of oilseed rape (*Brassica napus* L.). *Eur J Agron.* 2016;74:112–118.
5. Wolko J, Dobrzycka A, Bocianowski J, Bartkowiak-Broda I: Estimation of heterosis for yield-related traits for single cross and three-way cross hybrids of oilseed rape (*Brassica napus* L.). *Euphytica.* 2019;215(10):156.
6. Yang P, Shu C, Chen L, Xu J, Wu J, Liu K. Identification of a major QTL for siliques length and seed weight in oilseed rape (*Brassica napus* L.). *Theor Appl Genet.* 2012;125(2):285–296.
7. Mingliang W, Lun T, Chunyun G, Chuzhou T. Optimization analysis on parameters of cleaning sieve of rape combine of "Bi Lang 4LZ(Y)-1.0". *Adv J Food Sci Technol.* 2014;6(1):144–148.
8. Liu R, Huang S, Nie Y, Xu S. Automated detection research for number and key phenotypic parameters of rapeseed siliques. *Chin J Oil Crop Sci.* 2020;42(1):71–77.
9. Wang W, Hu Q, Mei D, Li Y, Wang H, Wang J, Fu L, Liu J. Evaluation of branch and pod angle measurement based on digital images from *Brassica napus* L. *Chin J Oil Crop Sci.* 2015;37(4):566–570.
10. Li L, Zhang Q, Huang DF. A review of imaging techniques for plant phenotyping. *Sensors.* 2014;14(11):20078–20111.
11. Sun D, Robbins K, Morales N, Shu Q, Cen H. Advances in optical phenotyping of cereal crops. *Trends Plant Sci.* 2022;27(2):191–208.
12. Xu S, Lu K, Pan L, Liu T, Zhou Y, Wang B. 3D reconstruction of rape branch and pod recognition based on RGB-D camera. *Trans Chin Soc Agric Mach.* 2019;50(2):21–27.
13. Lin Y, Ruifang Z, Pujuan S, Pengfei W. Segmentation of crop organs through region growing in 3D space. Paper presented at: Agro-Geoinformatics 2016. Proceedings of the 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics); 2016 Jul 18–20; Tianjin, China.
14. Amiri N, Yao W, Heurich M, Krzystek P, Skidmore AK. Estimation of regeneration coverage in a temperate forest by 3D segmentation using airborne laser scanning data. *Int J Appl Earth Obs Geoinf.* 2016;52:252–262.
15. Lindberg E, Holmgren J, Olofsson K, Wallerman J, Olsson H. Estimation of tree lists from airborne laser scanning using tree model clustering and k-MSN imputation. *Remote Sens.* 2013;5(4):1932–1955.

16. Liu WT, Wang CL, Yan D, Chen WL, Luo LF. Estimation of characteristic parameters of grape clusters based on point cloud data. *Front Plant Sci.* 2022;13:885167.
17. Bao Y, Tang L, Srinivasan S, Schnable PS. Field-based architectural traits characterisation of maize plant using time-of-flight 3D imaging. *Biosyst Eng.* 2019;178:86–101.
18. Wu S, Wen W, Xiao B, Guo X, Du J, Wang C, Wang Y. An accurate skeleton extraction approach from 3D point clouds of maize plants. *Front Plant Sci.* 2019;10:248.
19. Gaillard M, Miao C, Schnable J, Benes B. Sorghum segmentation by skeleton extraction. In: Bartoli A, Fusiello A, editors. *European conference on computer vision: 2020*. Cham (Switzerland): Springer; 2020. p. 296–311.
20. Chaudhury A, Godin C. Skeletonization of plant point cloud data using stochastic optimization framework. *Front Plant Sci.* 2020;11:773.
21. Preuksakarn C, Boudon F, Ferraro P, Durand J-B, Nikinmaa E, Godin C. Reconstructing plant architecture from 3D laser scanner data. Paper presented at: FSPM 2010. Proceedings of the 6th International Workshop on Functional-Structural Plant Models; 2010 Sep 12–17; Davis, CA.
22. Li R, Bu G, Wang P. An automatic tree skeleton extracting method based on point cloud of terrestrial laser scanner. *Int J Opt.* 2017;2017:5408503.
23. Han X-F, Jin JS, Wang M-J, Jiang W, Gao L, Xiao L. A review of algorithms for filtering the 3D point cloud. *Signal Process Image Commun.* 2017;57:103–112.
24. Yang C, Wang Z, He W, Li Z. Development of a fast transmission method for 3D point cloud. *Multimed Tools Appl.* 2018;77(19):25369–25387.
25. Hu Y, Wang L, Xiang L, Wu Q, Jiang H. Automatic non-destructive growth measurement of leafy vegetables based on kinect. *Sensors.* 2018;18(3):306.
26. Daszykowski M, Kaczmarek K, Heyden YV, Walczak B. Robust statistics in data analysis—A review basic concepts. *Chemom Intell Lab Syst.* 2007;85(2):203–219.
27. Huang H, Wu S, Cohen-Or D, Gong M, Zhang H, Li G, Chen B.  $L_1$ -medial skeleton of point cloud. *ACM Trans Graph.* 2013;(4):32, 65.
28. Ozbay E, Cinar A, Guler Z. A hybrid method for skeleton extraction on Kinect sensor data: Combination of  $L_1$ -median and Laplacian shrinking algorithms. *Measurement.* 2018;125:535–544.
29. Fischler MA, Bolles RC. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM.* 1981;24(6):381–395.
30. Bäcklund H, Hedblom A, Neijman N. A density-based spatial clustering of application with noise. *Data Mining TNM033.* 2011;33:11–30.
31. Li N, Shi J, Wang X, Liu G, Wang H. A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and siliques length in rapeseed (*Brassica napus* L.). *BMC Plant Biol.* 2014;14.
32. Fu Y, Wei D, Dong H, He Y, Cui Y, Mei J, Wan H, Li J, Snowdon R, Friedt W, et al. Comparative quantitative trait loci for siliques length and seed weight in *Brassica napus*. *Sci Rep.* 2015;5:14407.
33. Bennett E, Roberts JA, Wagstaff C. Manipulating resource allocation in plants. *J Exp Bot.* 2012;63(9):3391–3400.
34. Sylvester-Bradley R, Lunn G, Foulkes J, Shearman V, Spink J, Ingram J. Management strategies for high yields of cereals and oilseed rape. Paper presented at: HGCA 2002. Proceedings of the HGCA Conference: Agronomic Intelligence: The Basis for Profitable Production; 2002 Jan 16 to 17; London, UK.
35. Rusu RB, Cousins S. 3D is here: Point Cloud Library (PCL). Paper presented at: ICRA 2011. Proceedings of the 2011 IEEE International Conference on Robotics and Automation; 2011 May 9–13; Shanghai, China.
36. Zhou Q-Y, Park J, Koltun V. Open3D: A modern library for 3D data processing. arXiv. 2018. <https://doi.org/10.48550/arXiv.1801.09847>
37. Bao Y, Tang L, Breitzman MW, Fernandez MGS, Schnable PS. Field-based robotic phenotyping of sorghum plant architecture using stereo vision. *J Field Robot.* 2019;36(2):397–415.
38. Dandrifosse S, Bouvry A, Leemans V, Dumont B, Mercatoris B. Imaging wheat canopy through stereo vision: Overcoming the challenges of the laboratory to field transition for morphological features extraction. *Front Plant Sci.* 2020;11:96.
39. Scharr H, Minervini M, French AP, Klukas C, Kramer DM, Liu XM, Luengo I, Pape JM, Polder G, Vukadinovic D, et al. Leaf segmentation in plant phenotyping: A collation study. *Mach Vis Appl.* 2016;27(4):585–606.
40. Xia CL, Wang LT, Chung BK, Lee JM. In situ 3D segmentation of individual plant leaves using a RGB-D camera for agricultural automation. *Sensors.* 2015;15(8):20463–20479.
41. Lei L, Qiu C, Li Z, Han D, Han L, Zhu Y, Wu J, Xu B, Feng H, Yang H, et al. Effect of leaf occlusion on leaf area index inversion of maize using UAV-LiDAR data. *Remote Sens.* 2019;11(9):1067.
42. Zermas D, Morellas V, Mulla D, Papanikopoulos N. Estimating the leaf area index of crops through the evaluation of 3D models. Paper presented at: IROS 2017. Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2017 Sep 24–28; Vancouver, BC, Canada.
43. Iniguez R, Palacios F, Barrio I, Hernandez I, Gutierrez S, Tardaguila J. Impact of leaf occlusions on yield assessment by computer vision in commercial vineyards. *Agronomy.* 2021;11(5):1003.
44. Jin SC, Su YJ, Wu FF, Pang SX, Gao S, Hu TY, Liu J, Guo QH. Stem-leaf segmentation and phenotypic trait extraction of individual maize using terrestrial LiDAR data. *IEEE Trans Geosci Remote Sens.* 2019;57(3):1336–1346.
45. Miao T, Zhu C, Xu TY, Yang T, Li N, Zhou YC, Deng HB. Automatic stem-leaf segmentation of maize shoots using three-dimensional point cloud. *Comput Electron Agric.* 2021;187.
46. Miao T, Wen WL, Li YL, Wu S, Zhu C, Guo XY. Label3DMaize: Toolkit for 3D point cloud data annotation of maize shoots. *GigaScience.* 2021;10(5): 10.1093/gigascience/giab031.
47. 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;162:951–961.
48. Ghahremani M, Williams K, Corke FMK, Tiddeman B, Liu Y, Doonan JH. Deep segmentation of point clouds of wheat. *Front Plant Sci.* 2021;12:608732.
49. Li Y, Wen W, Miao T, Wu S, Yu Z, Wang X, Guo X, Zhao C. Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning. *Comput Electron Agric.* 2022;193:106702.
50. Turgut K, Dutagaci H, Galopin G, Rousseau D. Segmentation of structural parts of rosebush plants with 3D point-based deep learning methods. *Plant Methods.* 2022;18:20.
51. Tyagi S, Mazumdar PA, Mayee P, Shivaraj SM, Anand S, Singh A, Madhurantakam C, Sharma P, Das S, Kumar A, et al. Natural variation in Brassica FT homeologs influences multiple

- agronomic traits including flowering time, siliques shape, oil profile, stomatal morphology and plant height in *B. juncea*. *Plant Sci.* 2018;277:251–266.
52. Shi JQ, Li RY, Qiu D, Jiang CC, Long Y, Morgan C, Bancroft I, Zhao JY, Meng JL. Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics*. 2009;182(3):851–861.
  53. Zhang LW, Yang GS, Liu PW, Hong DF, Li SP, He QB. Genetic and correlation analysis of siliques-traits in *Brassica napus* L. by quantitative trait locus mapping. *Theor Appl Genet*. 2011;122(1):21–31.
  54. Wang XD, Chen L, Wang AN, Wang H, Tian JH, Zhao XP, Chao HB, Zhao YJ, Zhao WG, Xiang J, et al. Quantitative trait loci analysis and genome-wide comparison for siliques related traits in *Brassica napus*. *BMC Plant Biol*. 2016;16:Article 71.
  55. Raboanatahiry N, Chao HB, Dalin H, Pu S, Yan W, Yu LJ, Wang BS, Li MT. QTL alignment for seed yield and yield related traits in *Brassica napus*. *Front Plant Sci*. 2018;9:Article 1127.
  56. Zajac T, Oleksy A, Stoklosa A, Klimek-Kopyra A. Comparison of morphological traits, productivity and canopy architecture of winter oilseed rape (*Brassica napus* L.) and white mustard (*Sinapis alba* L.). *J Appl Bot Food Qual*. 2011;84(2):183–191.
  57. Andelić E, Antunović M, Stošić M, Iljkić D, Varga I. Yield components of winter oilseed rape regard to plant population. *Columella-J Agric Environ Sci*. 2018;5(2):33–41.
  58. Diepenbrock W. Yield analysis of winter oilseed rape (*Brassica napus* L.): A review. *Field Crop Res*. 2000;67(1):35–49.
  59. Paulus S. Measuring crops in 3D: Using geometry for plant phenotyping. *Plant Methods*. 2019;15(1):Article 103.
  60. Bouchet A-S, Nesi N, Bissuel C, Bregeon M, Lariepe A, Navier H, Ribiere N, Orsel M, Grezes-Besset B, Renard M, et al. Genetic control of yield and yield components in winter oilseed rape (*Brassica napus* L.) grown under nitrogen limitation. *Euphytica*. 2014;199(1–2):183–205.
  61. Liu J, Hua W, Hu Z, Yang H, Zhang L, Li R, Deng L, Sun X, Wang X, Wang H. Natural variation in ARF18 gene simultaneously affects seed weight and siliques length in polyploid rapeseed. *Proc Natl Acad Sci USA*. 2015;112(37):E5123–E5132.
  62. Elster R, Bommert P, Sheridan WF, Werr W. Analysis of four embryo-specific mutants in *Zea mays* reveals that incomplete radial organization of the proembryo interferes with subsequent development. *Dev Genes Evol*. 2000;210(6):300–310.
  63. Young LW, Wilen RW, Bonham-Smith PC. High temperature stress of *Brassica napus* during flowering reduces micro- and megagametophyte fertility, induces fruit abortion, and disrupts seed production. *J Exp Bot*. 2004;55(396):485–495.
  64. Chen Z, Huo Q, Yang H, Jian H, Qu C, Lu K, Li J. Joint RNA-seq and miRNA profiling analyses to reveal molecular mechanisms in regulating thickness of pod canopy in *Brassica napus*. *Genes*. 2019;10(8):591.