Computers and Electronics in Agriculture 234 (2025) 110293

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Cotton3DGaussians: Multiview 3D Gaussian Splatting for boll mapping and plant architecture analysis

Lizhi Jiang [a,b], Jin Sun [c], Peng W. Chee [d], Changying Li [b,*], Longsheng Fu [a,*]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Bio-Sensing, Automation, and Intelligence Laboratory, Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611, USA
[c] College of Computing, University of Georgia, Athens, GA 30602, USA
[d] Department of Crop and Soil Science and Institute of Plant Breeding, Genetics, and Genomics, University of Georgia-Tifton Campus, University of Georgia, Tifton, GA 31793, USA

ARTICLE INFO

Keywords:
3D reconstruction
3D segmentation
3D Gaussian Splatting
SAM
YOLOv11x

ABSTRACT

Cotton is an economically important crop cultivated worldwide for textile production. Breeding programs focus on selecting genotypes with favorable traits for high yields. This study introduced 3D Gaussian Splatting (3DGS) to reconstruct high-fidelity three-dimensional (3D) models and developed a segmentation workflow, Cotton3DGaussians, to analyze cotton bolls and extract architectural traits from single plants. Cotton plants were scanned 360° using a smartphone, and photogrammetry was used to estimate camera parameters and reconstruct a sparse point cloud, which was then optimized into a 3DGS model. In Cotton3DGaussians, 2D masks of bolls segmented from four views were mapped to 3D space, and redundant bolls were removed through cross-view clustering. YOLOv11x and a foundation model, segment anything model (SAM), were compared to obtain 2D masks, with YOLOv11x achieving an F1-score 5.9 % higher than SAM. Phenotypic traits such as boll number, volume, plant height, and canopy size were estimated. The 3DGS model exhibited superior rendering quality, achieving a peak signal-to-noise ratio (PSNR) that was 6.91 higher than NeRF. Cotton3DGaussians effectively segmented 3D bolls from multiple views, with mean absolute percentage errors (MAPE) of 9.23 % for boll number, 3.66 % for canopy size, 2.38 % for plant height, and 8.17 % for boll volume compared to LiDAR ground truth. The regression analysis between convex boll volume and boll weight showed a 19.3 % weight error per plant. This study demonstrates the potential of 3DGS for low-cost, high-fidelity 3D modeling, enabling high-resolution phenotyping and advancing cotton breeding programs. The methodology can also be applied to other crops for improved 3D trait measurement research and enhanced productivity.

1. Introduction

Cotton (Gossypium hirsutum L.) is an important economic crop, with its fibers widely used in the textile industry (Zhao et al., 2023). The global production value of raw cotton is estimated at $50 billion annually, with international trade contributing approximately $20 billion (Devoto et al., 2024). In cotton phenotyping, high-resolution RGB images capture fine details such as textures and color variations, enabling the observation of subtle features in plants (Li et al., 2021). However, plants are inherently three-dimensional (3D), and occlusions severely limit the ability of 2D images to characterize the complex spatial structure (Miao et al., 2021). To overcome these limitations and meet the growing demand for 3D high-throughput phenotyping, high-fidelity 3D models of cotton plants are essential. 3D models preserve the spatial architecture of crops, enabling phenotypic analysis at various levels—from populations to individual plants and even down to specific organs. These models provide breeders with valuable insights into the intricate relationships between phenotypic traits and genetic variations, facilitating more informed breeding decisions.

3D imaging technology has been applied in crop phenotyping studies, with some instruments capable of directly acquiring 3D point cloud data. For example, 3D laser scanners, which operate by emitting laser pulses and measuring their return time to determine distances, offer high precision and are unaffected by ambient lighting conditions (Dong et al., 2020). These scanners have been widely adopted in agricultural research for plant phenotyping and structural analysis. A

* Corresponding authors.
E-mail addresses: cli2@ufl.edu (C. Li), fulsh@nwafu.edu.cn (L. Fu).

https://doi.org/10.1016/j.compag.2025.110293
Received 11 January 2025; Received in revised form 6 March 2025; Accepted 12 March 2025
Available online 24 March 2025

terrestrial LiDAR sensor has been used to capture high-density point clouds of cotton plants in both indoor and field environments, enabling the analysis of phenotypic traits such as boll count, plant height, node count, and boll distribution (Jiang et al., 2022; Saeed et al., 2023b; Sun et al., 2021). Similarly, handheld LiDAR devices, due to their portability, are suitable for field applications, such as capturing maize plant point clouds. Using the Minkowski distance field method, these point clouds were segmented to extract individual plant traits (Wang et al., 2023). Backpack LiDAR systems have been utilized to acquire apple tree point clouds, enabling branch segmentation using TreeQSM and calculating branch length (Zhang et al., 2020). Handheld mobile laser scanners have been used to scan blueberry plants, extracting size and shape information from point cloud data (Jiang et al., 2019). Airborne-LiDAR allows high-throughput field-level data acquisition, facilitating the extraction of traits such as plant height and canopy width (Liu et al., 2024). However, the high cost of LiDAR systems has prompted many researchers to explore RGB-D depth cameras as a more affordable alternative for acquiring point clouds and phenotypic traits (Jiang et al., 2025). Depth cameras were used to capture point clouds from multiple angles, which were then registered into a complete plant model using registration algorithms such as iterative closest point (ICP), enabling the extraction of structural characteristics (Liu et al., 2023). Despite their affordability, RGB-D sensors generate low-resolution point cloud data, constraining their capability to capture fine-scale plant structures.

Multiview 2D images are used to reconstruct 3D models through photogrammetric methods such as structure-from-motion (SfM) to generate 3D point clouds from a series of images taken from different perspectives (Schonberger and Frahm, 2016). Various software tools have been developed to support this 3D reconstruction process. For instance, 2D images of soybean plants captured indoors have been reconstructed into 3D point clouds using VisualSFM, allowing segmentation into individual plant parts for phenotypic analysis (He et al., 2023). Similarly, tomato plants grown in greenhouses have been reconstructed with VisualSFM and analyzed for phenotypic traits using the 3DPhenoMVS pipeline (Wang et al., 2022). For field-based studies, multiview image acquisition systems have been employed to capture images of cotton plants, which are then processed with Agisoft Metashape to reconstruct point clouds for extracting traits such as boll count and spatial distribution (Sun et al., 2020). Additionally, UAVs equipped with cameras facilitate rapid field data collection, with color images used to reconstruct 3D point clouds for analyzing traits such as plant height and canopy width (Jamil et al., 2022; Xiao et al., 2023). While these 3D reconstructions address the occlusion issues inherent to 2D images and facilitate the extraction of spatial traits, they often lose the high-resolution color and texture details provided by 2D images.

The Neural Radiance Fields (NeRF) method is a breakthrough in implicit 3D scene representation, allowing the reconstruction of 3D models from 2D images. This approach addresses common limitations in traditional 3D reconstruction, such as holes, texture blending, and detail loss due to voxel resolution constraints (Mildenhall et al., 2022). More importantly, NeRF enables the reconstruction of 3D models using 2D deep learning techniques, expanding its applications, particularly in phenotyping. There are two primary applications of NeRF in plant phenotyping. First, multiview images can be used to train a NeRF model, which can then synthesize novel views and reconstruct point clouds, enabling trait extraction. Second, NeRF can be used to segment the object of interest from multiview images, integrate the segmented object into NeRF training, and reconstruct 3D point clouds of the object. This approach has been successfully applied to reconstruct 3D models of various fruits (e.g., pitahaya, grapes, litchis, oranges) with fine geometric details, generating point clouds and meshes (Hu et al., 2024). For corn plants, the point clouds reconstructed with NeRF were comparable in quality to those generated using expensive LiDAR systems (Arshad et al., 2024). NeRF has also demonstrated robust performance in outdoor environments, where strawberry plants of varying scales were reconstructed (Zhang et al., 2024). In studies involving crops such as peanuts, tomatoes, and walnuts, NeRF was employed for point cloud reconstruction to estimate traits such as fruit quantity and size (Choi et al., 2024; Huang et al., 2024; Saeed et al., 2023a; Zheng et al., 2024). One study developed PAg-NeRF which integrates Mask2Former to achieve 3D panoramic scene understanding for agricultural robots (Smitt et al., 2024). Another example is PanicleNeRF, which segments the panicle region in 2D images, trains NeRF with these images, and exports point clouds to estimate panicle length and volume (Yang et al., 2024a). FruitNeRF uses RGB images and fruit semantic masks as inputs to train the NeRF model, ultimately extracting both scene and fruit point clouds (Meyer et al., 2024). Despite its successes, NeRF currently faces challenges with computational efficiency. Its use of implicit representation means that the resulting 3D model is stored within the neural network, which is not ideal for downstream tasks.

3D Gaussian Splatting (3DGS) is a state-of-the-art 3D reconstruction method that offers fast training speeds and real-time, high-quality rendering with an explicitly expressed model (Kerbl et al., 2023). This provides a new approach for the rapid reconstruction of high-quality 3D crop models. Recent advancements in 3DGS have introduced segmentation techniques such as SAGA (Cen et al., 2023), Gaussian Grouping (Ye et al., 2023), and Omniseg3D (Ying et al., 2023). Studies have shown that 3DGS can effectively reconstruct plants such as canola, wheat, and beans with high-quality geometric accuracy and architectural detail (Ojo et al., 2024). Despite its potential, the application of 3DGS in the agricultural domain remains underexplored, particularly in plant segmentation and phenotypic trait extraction, with a notable lack of research on its use for cotton plants. Integrating high-resolution 2D cotton plant images with 3DGS models to capture spatial architecture and extract phenotypic traits presents a promising avenue for future research.

To fill this gap, this study aimed to explore the application of 3DGS for high-fidelity 3D reconstruction of cotton plants, 3D segmentation of cotton bolls, and evaluation of plant architectural traits. The specific objectives of this study were to: (1) reconstruct high-fidelity 3DGS models of individual cotton plants; (2) compare the performance of a foundation model (SAM) and YOLOv11x in obtaining 2D instance masks from RGB images; (3) segment bolls in the 3DGS model and cluster the segmentation results from multiple views; (4) evaluate 3D boll mapping and architectural traits.

2. Materials and methods

2.1. Data collection

All the cotton plant data used in this experiment were collected indoors. The leaves of these plants had already fallen off, with only a few scattered ones remaining on the branch. They were cut from the field and brought indoors, where they were secured on a table (120 cm × 60 cm) using two short metal tubes (4 cm in height). These tubes were placed parallel to each other on the table, clamping the plant securely in place during data collection. A consumer-grade smartphone (iPhone 11) was used to capture a 360–degree video of a single plant. The resolution was set to 3840 × 2160 pixels, with a frame rate of 60 frames per second (fps). A complete plant was consistently present within the frame throughout the data collection process. Data collection was completed in December 2023, and 50 plants were tested in this experiment.

Cotton plant point cloud data was also collected using a LiDAR sensor in this experiment as ground truth. The plants were nailed to a stake without overlapping. There were 12 plants in each batch. A 3D terrestrial LiDAR sensor (FARO Focus S70, FARO Technologies, USA) was used to collect point clouds. The resolution and quality of LiDAR scans were set as ½ and 2x, respectively. One scan took 11.04 min. For each batch, eleven scans were selected.

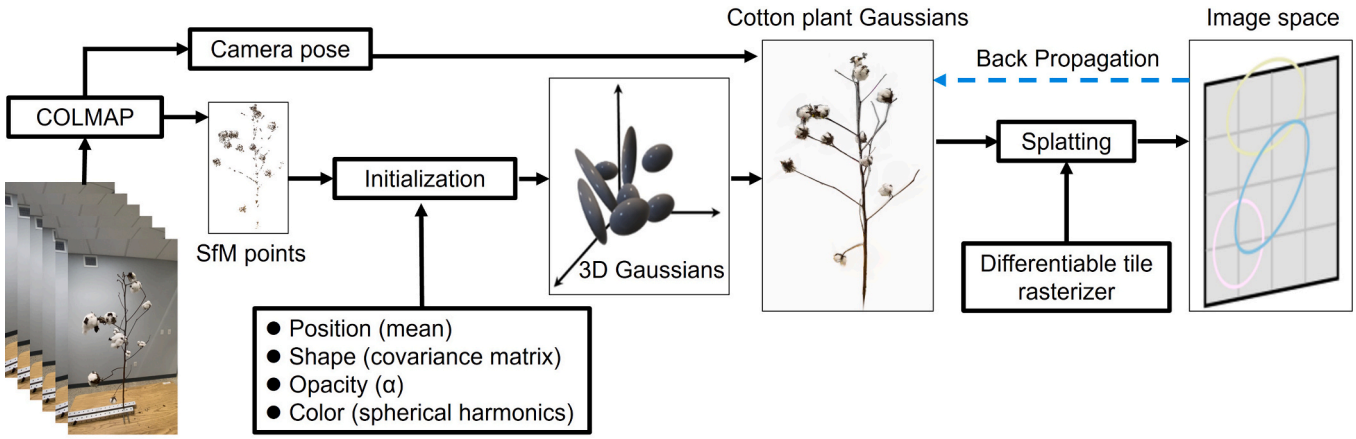

Fig. 1. 3DGS cotton plant reconstruction workflow.

2.2. Data preprocessing

2.2.1. Camera parameter estimation

The 3D Gaussian Splatting (3DGS) model requires precise camera parameters to establish the projection relationship between 2D images and their corresponding 3D spatial positions. These parameters include intrinsic (focal length, principal point) and extrinsic (rotation, translation) data, which are critical for accurate mapping in 3D space. COLMAP (Schonberger and Frahm, 2016) was used to compute these camera parameters through a SfM method, providing the necessary data for each image. The process began with evenly extracting images from the collected videos using FFmpeg to create a training set of cotton plant scenes. Then, COLMAP was employed for feature extraction, feature matching, and sparse point cloud reconstruction to estimate the camera parameters. Finally, this information was converted into the Local Light Field Fusion (LLFF) data format (Mildenhall et al., 2019).

2.2.2. Data annotation for YOLOv11x instance segmentation

The 2D images were annotated to create a dataset for training a YOLOv11x model to predict a 2D mask for each cotton boll. Roboflow (https://app.roboflow.com) was used to annotate the polygons of the cotton bolls. It is a 2D data annotation tool that directly uses the Segment Anything Model (SAM) model (Kirillov et al., 2023) to improve the annotation efficiency for cotton bolls. This study annotated 157 images: 109 for training and 48 for validation. To enhance the model's generalization ability, data augmentation techniques such as brightness adjustment, blurring, rotation, and scaling were applied to the training set.

2.3. Overview of the proposed multiview 3DGS segmentation algorithm

This study developed a new method to estimate 3D traits of cotton plants based on 3DGS. The pipeline takes in 2D images collected from individual cotton plants, which can be used to train a 3DGS model (Fig. 1). Simultaneously, the 2D masks of the cotton boll can be obtained from the 2D images through various 2D segmentation methods (such as YOLOv11x or SAM). Next, the 3DGS model was segmented into individual cotton bolls using a multiview 3DGS segmentation algorithm. Finally, the cotton boll number, volume, and architectural traits were estimated.

2.3.1. 3D Gaussian Splatting

3DGS (Kerbl et al., 2023) employs trainable 3D Gaussian distributions to represent 3D scenes, where each Gaussian encapsulates attributes such as position, shape, color, and opacity. 3DGS introduces an efficient differentiable rasterization algorithm to enable effective training and rendering. The properties of 3D Gaussian distributions encompass several key aspects. The position of a 3D Gaussian distribution is determined by its mean value, which represents its location in 3D space. Its anisotropic covariance describes the distribution's shape, size, and orientation, with the covariance matrix being required to be semi-positive definite. Additionally, spherical harmonic (SH) coefficients, a set of 48 values, are used to represent the color and texture variations across different views on the object's surface. Finally, the opacity of a 3D Gaussian, denoted by the parameter α, defines its transparency level and governs its visibility during rendering.

Given a training dataset $\mathfrak{T}$ consisting of multi-view 2D images, COLMAP's SfM algorithm estimated the camera parameters for each image and generated a sparse point cloud. The sparse point cloud was initialized as a 3D Gaussian distribution. Initialization was the process of assigning initial values to the properties of the Gaussians. For example, a Gaussian was placed at the location of each point in the sparse point cloud, the SH coefficients were computed from the point cloud color, the opacity was set to 1, and the initial covariance matrix was set to the identity matrix. After initialization, gradient descent was used to further optimize the 3D Gaussian parameters.

3DGS learns a set of colored 3D Gaussians $\mathscr{G} = \{g_1, g_2, \cdots g_N\}$, where N represents the number of 3D Gaussians in the scene. The mean of each Gaussian defines its position, and its covariance determines its scale. Given a specific camera pose, 3DGS projects the 3D Gaussians onto 2D and computes the color C of a pixel by blending N ordered set of Gaussians that overlap the pixel (Eq. (1).

$$C = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \tag{1}$$

Where c_i is the color of each Gaussian, and α_i is calculated by evaluating the 2D Gaussian with covariance Σ and multiplied it by the learnable opacity of each Gaussian.In the gradient descent process, the loss function was a combination of L_1 loss and a D-SSIM term (Eq. (2). The captured image was used as the ground truth (GT), and the 3D Gaussian distribution was rendered into a 2D image from the same viewpoint as the training image. The rendering of a 3D Gaussian distribution into a 2D image is called "Splatting." At a specific view, the difference between the corresponding pixels of the ground truth 2D image and the rendered 2D image was calculated. The average of these pixel differences was then computed for each image. The optimization goal in 3DGS was to iteratively minimize this average difference to improve the accuracy of the 3DGS model.

$$L = (1 - \lambda) L_1 + \lambda L_{D-SSIM} \tag{2}$$

λ is an adjustment coefficient, with a default value of 0.2.

To reconstruct a high-quality 3DGS model, this study analyzed various conditions. During data collection, the impact of different

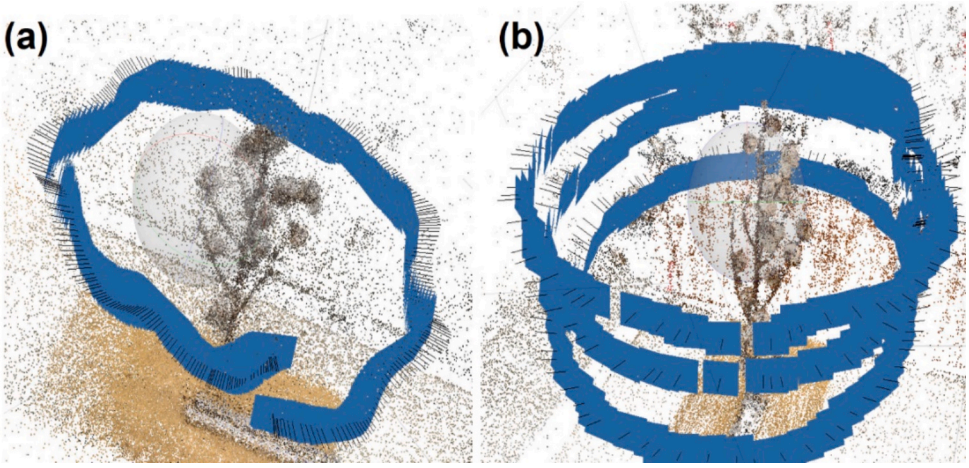

Fig. 2. Point cloud sparse reconstruction and camera pose. (a) Global shooting data collection, where the entire plant was in view. (b) Local shooting data collection, where the whole plant was scanned at three heights.

acquisition methods, such as global shooting and local shooting was considered (Fig. 2). Global shooting ensures the entire plant remains within the camera's field of view, whereas local shooting captures only parts of the plant in each image. During model training, the effects of image resolution, the number of training images, and training iterations were evaluated. Additionally, the reconstruction quality of 3DGS was compared with that of NeRF.

2.3.2. Multi-view 3DGS segmentation algorithm

Segment Any 3D GAussians (SAGA, Cen et al., 2023) is an interactive 3D segmentation method based on 3DGS. It leverages contrastive training with 2D segmentation models (e.g., SAM) to embed 2D segmentation information into 3D Gaussian feature vectors. Specifically, SAGA is trained using a SAM-guidance loss and a correspondence loss to refine 3D Gaussian features. The SAM-guidance loss helps guide 3D feature learning for segmentation by incorporating SAM-derived features, while the correspondence loss captures point-to-point relationships from segmentation masks and distills them into the feature space, enhancing feature compactness.

Building upon the SAGA framework, we propose an improved multiview 3DGS-based cotton boll segmentation algorithm (Cotton3DGaussians, Fig. 3). In this approach, high-precision 2D masks predicted by YOLOv11x replace the manually provided prompts in SAGA that were used to generate SAM-based masks. To address the issue of incomplete mask detection caused by occlusion in 2D views, we selected four views per plant to generate masks. Finally, cotton boll counting was achieved through cross-view DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering.

Given a pre-trained 3DGS model $\mathscr{G}$ and its training dataset $\mathfrak{T}$, the Cotton3DGaussians process started by using SAM's encoder to extract feature maps F_I^{SAM} and a set of multi-scale masks from each 2D image M_I^{SAM}. Next, these 2D masks were leveraged to train the low-dimensional feature f_g of each Gaussian g, aggregating multi-scale segmentation information across views (i.e., segmenting the same object from different views into a unified category). Training Gaussian features combined the training images I, their corresponding camera poses v, and a carefully designed SAM-guidance loss. First, the pre-trained 3DGS model was used to render the feature map for each image in the training set. The rendered feature $F_{I,p}^r$ of a pixel p was computed as shown in Eq. (3), where $\mathscr{N}$ is the ordered set of Gaussians overlapping the pixel.

$$F_{I,p}^r = \sum_{i\in\mathscr{N}} f_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_j) \tag{3}$$

Where f_i is the segmentation attribute value of the i-th Gaussian obtained training. α_i is the opacity of the i-th Gaussian. $\prod_{j=1}^{i-1}(1-\alpha_j)$ is a cumulative product of opacities, indicating how the i-th Gaussian is influenced by the opacities of all previous Gaussians.

Since the 2D masks automatically extracted by SAM were complex and often confusing, points belonging to the same object in 3D space may be segmented as different objects across different views. To address this, SAM-extracted features were used as guidance. First, an MLP φ projected the SAM features into the same low-dimensional space as the 3D features (Eq. (4). Then, average pooling was performed on each extracted mask M in M_I^{SAM} to generate the corresponding query T_M (Eq. (5). Next, T_M was used to segment the rendered feature map F_I^r through a softmaxed dot product (Eq. (6). Consequently, the SAM-guidance loss was defined as the binary cross-entropy between the segmentation result P_M and the corresponding SAM-extracted mask M (Eq. (7).

$$F_I' = \varphi(F_I^{SAM}) \tag{4}$$

$$T_M = \frac{1}{\|M\|_1}\sum_{p=1}^{HW} \mathbb{I}(M_p = 1) F_{I,p}' \tag{5}$$

$$P_M = \sigma(softmax(T_M \bullet F_I^r)) \tag{6}$$

$$\mathscr{L}_{SAM} = -\sum_{I\in\mathfrak{T}}\sum_{M\in M_I}\sum_{p}^{HW}\left[M_p log P_{M,p} + (1-M_p) log(1-P_{M,p})\right] \tag{7}$$

Where $\mathbb{I}$ denotes the indicators function, H and W represent the height and width of the image, and σ denotes the element-wise sigmoid function.

The features learned by the SAM-guidance loss were not sufficiently compact, so Correspondence Loss was introduced. For each image I in the training set $\mathfrak{T}$, a series of masks M_I were extracted by SAM. Considering two pixels p_1 and p_2 in I, they may belong to many masks in M_I. Let $M_I^{p_1}$ and $M_I^{p_2}$ represent the sets of masks that include $p1$ and $p2$, respectively. If the intersection over union (IoU) of these two sets is larger, the two pixels share more similar features. So the mask correspondence $K_I(p_1,p_2)$ was defined as Eq. (8). The feature correspondence $S_I(p_1,p_2)$ between two pixels p_1,p_2 was defined as the cosine similarity between their rendered features (Eq. (9). Then, the correspondence loss was defined as Eq. (10).

$$K_I(p_1,p_2) = \frac{|M_I^{p_1} \cap M_I^{p_2}|}{|M_I^{p_1} \cup M_I^{p_2}|} \tag{8}$$

$$S_I(p_1,p_2) = < F_{I,p_1}^r, F_{I,p_2}^r > \tag{9}$$

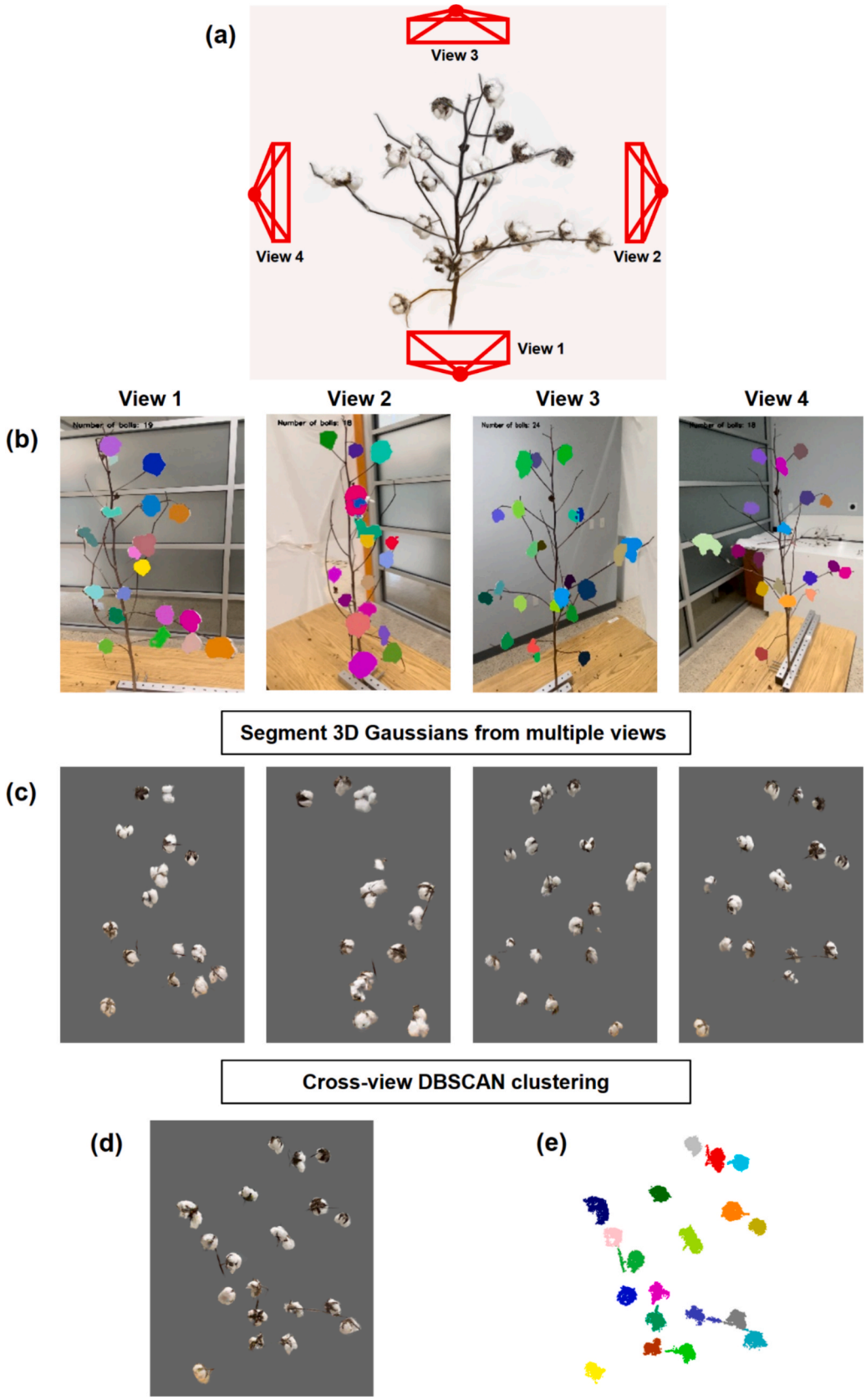

Fig. 3. Multiview cotton boll instance segmentation pipeline. (a) Four views of a single cotton plant, each 90° apart, for boll segmentation. (b) 2D instance segmentation results of YOLOv11x from different views. (c) 3D Gaussians of cotton boll segmented from different views. (d) Cross-view clustering results of cotton bolls from four views. (e) Each cotton boll instance was assigned a unique color.

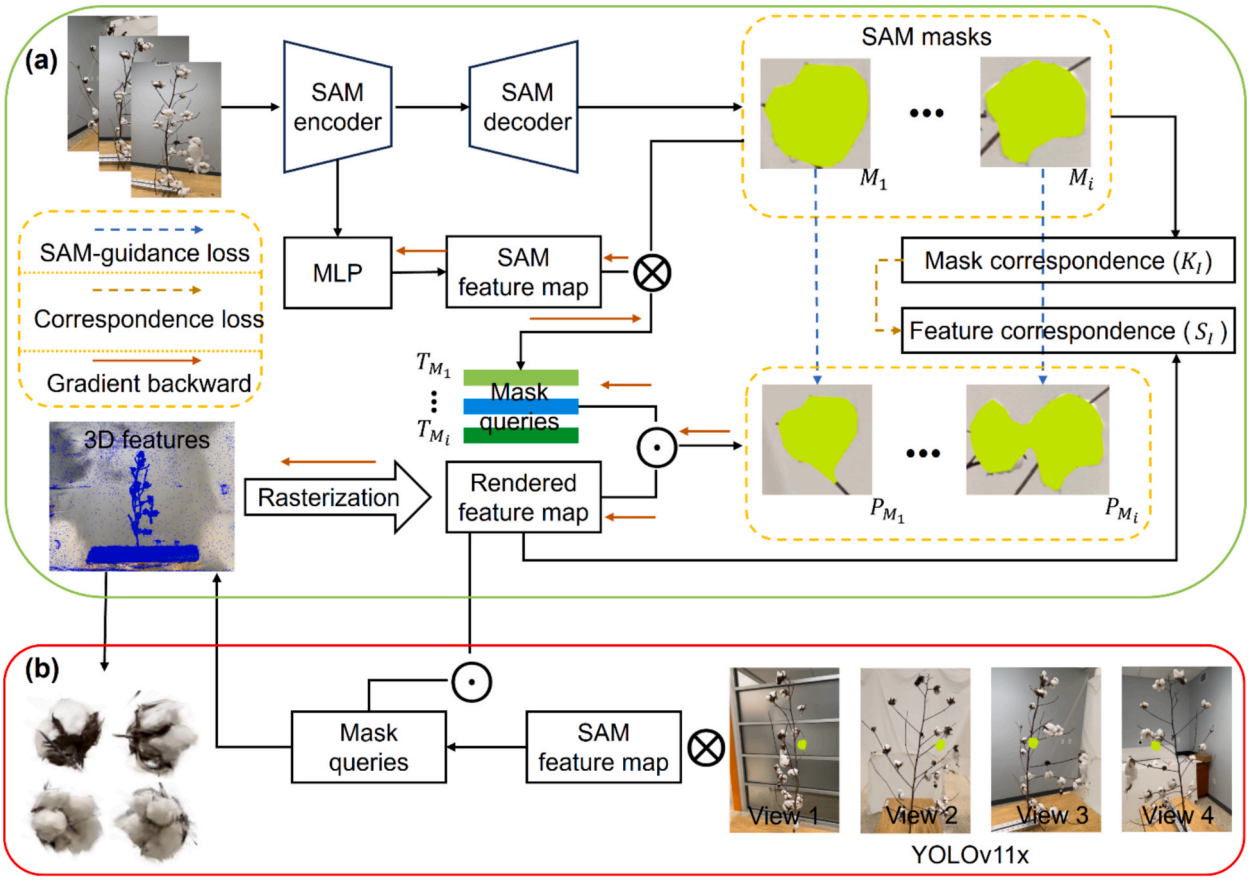

Fig. 4. Multiview 3DGS segmentation algorithm workflow. (a) The training stage of the model, where segmentation features are retrained for the pre-trained 3D Gaussian model. (b) The inference stage of the model, where the 2D masks of cotton bolls segmented by YOLOv11x are used to segment the corresponding Gaussian set for each cotton boll. ⊗ is masked average pooling. ⊙ is point product.

$$\mathscr{L}_{corr} = -\sum_{I\in\mathfrak{T}}\sum_{p_1}^{HW}\sum_{p_2}^{HW} K_I(p_1,p_2)S_I(p_1,p_2) \tag{10}$$

The final loss function was a combination of the SAM-guidance loss and the correspondence loss (Eq. (11) for the multiview 3DGS segmentation algorithm, where β is the hyperparameter balancing the two loss terms. Although the training was conducted on the rendered feature maps, the rasterization operation was linear, ensuring that the features in 3D space align with those in the 2D rendered images. Therefore, the features from 2D rendering were used to segment the 3D Gaussians.

$$\mathscr{L} = \mathscr{L}_{SAM} + \beta\mathscr{L}_{corr} \tag{11}$$

2D instance masks of the cotton bolls were first segmented from the initial view images (Fig. 4b). In this experiment, the 2D masks of the cotton bolls were automatically segmented by YOLOv11x, which overcame the issue of manually provided prompts in the original SAGA method. Then, the mask was matched with the learned features to extract the 3D Gaussian representation of the cotton bolls. If cotton bolls were segmented from a single view, occlusion issues in the 2D image may result in missed boll segments. In this study, four views around one plant were selected for segmenting the 2D masks of the cotton bolls, with each view approximately 90° apart.

Selecting four views for segmentation may lead to redundant counting of the same cotton boll. During the segmentation process, some cotton bolls were incorrectly segmented together with branches or the main stem as a single instance. Before cross-view clustering, a convex hull-based volume threshold was introduced to filter out objects exceeding 25,000 cm^3. Next, cross-view DBSCAN was applied to cluster cotton bolls, with the neighborhood radius set to 0.3 m, representing the maximum distance between points within a cluster. The minimum number of samples was set to 10, indicating the least number of points required to form a cluster. Cross-view clustering refers to clustering only the 3D bolls obtained from different views, without clustering bolls within the same view. It should be noted that since 3D reconstruction directly from 2D images does not provide the physical size of objects in the real world, the thresholds in this experiment were determined based on the trained 3D model.

2.4. Two methods of generating 2D masks

2.4.1. Label-based method—YOLOv11x

YOLOv11 (https://docs.ultralytics.com/models/yolo11/) was used in this study to predict the 2D mask of each cotton boll. YOLOv11 is an enhanced version of the original YOLO (You Only Look Once). As a supervised learning model, YOLOv11x achieved state-of-the-art results for instance segmentation tasks on the COCO dataset. Therefore, YOLOv11x was selected for accurate cotton boll segmentation.

2.4.2. Label-free method—foundation model

SAM (Kirillov et al., 2023) was also used to generate 2D masks for cotton bolls. It is a foundation model with zero-shot learning capabilities. Unlike traditional segmentation models that rely solely on pixel-level annotations, SAM employs a prompt-based approach. SAM takes an image I and a set of prompts P as input, generating the corresponding 2D segmentation mask M (Eq. (12). In this experiment, the point prompts were manually assigned. A random point was selected at the center of each cotton boll, serving as the point prompt for segmentation.

$$M = SAM(I,P) \tag{12}$$

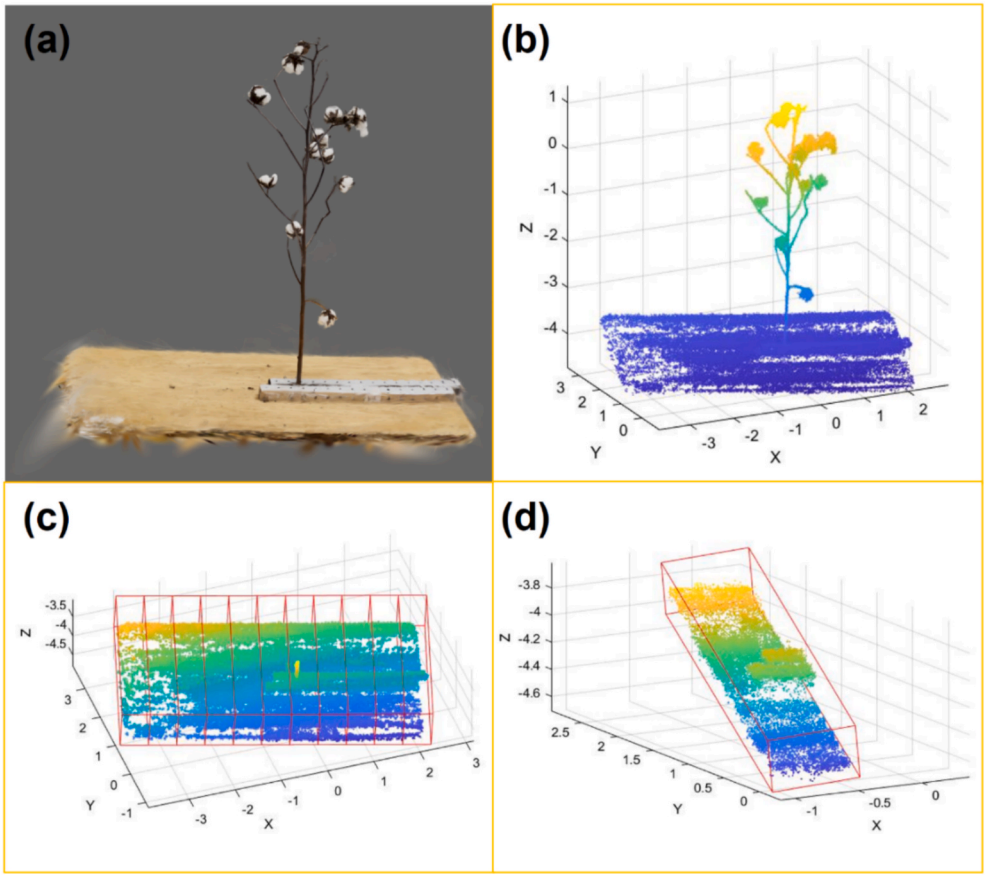

Fig. 5. Scale relationship between 3DGS and the physical world. (a) 3DGS model after background removal. (b) Point cloud composed of Gaussian function means after denoising. (c) Table divided into 11 sections. (d) Table width in the 3DGS model.

2.5. Relationship between the generated 3DGS model and the true size

Due to the difference in scale between the trained 3DGS model and the actual size of the plants, this experiment established the scaling relationship between the model and real plants based on the physical dimensions of a table as a reference object (Fig. 5). The physical width of the table used in this experiment is 60 cm, and the height of the tube is approximately 4 cm. The process involved the following steps.

First, the table was segmented manually from the 3DGS cotton plant model. The Z-axis of the plant point cloud (the mean of 3D Gaussians) corresponded to the height direction of the plant. Next, an oriented bounding box (OBB1) was calculated for the table point cloud and divided into 11 equal sections (Fig. 5c). The point cloud within the central section (Fig. 5d) was selected, and a second oriented bounding box (OBB2) was recalculated. The length (L) of OBB2 was then used as the table's width in the 3DGS model. Finally, the scaling factor was determined as $\beta = L/60cm$, where 60 cm represented the actual width of the table in the real world. Using this scaling factor β, the 3DGS model's point cloud was adjusted back to its original physical size.

2.6. Acquisition of phenotypic traits

Cotton plant height, canopy size, cotton boll number, and boll volume were estimated from the reconstructed 3DGS model. For the 3DGS model, the background was manually removed, leaving only individual cotton plants. Each plant was enclosed in an OBB box, with the height of the box representing the plant's height, and the average of the box's length and width representing the canopy size. The number of cotton bolls was determined using the automated segmentation results from Cotton3DGaussians. Convex volumes were calculated for the automatically segmented bolls, and the total boll volume for each plant was summed and regressed against the boll weight of the individual plant. To assess whether the reconstructed 3DGS model accurately estimates boll volume, 55 bolls with good segmentation results were manually selected to minimize the impact of segmentation errors. All traits were validated against ground truth obtained from LiDAR-captured point clouds.

2.7. Implementation details

All experiments were conducted on the HiPerGator high-performance computing cluster, which was equipped with 8 AMD EPYC Rome CPU cores, a single NVIDIA DGX A100 GPU node (80 GB), and ran on the Linux operating system. The YOLOv11x model was trained using software libraries including Python 3.10, PyTorch 1.12.1, and CUDA 11.3. The training parameters of the model were as follows: the input image size was 1280 × 1280, the batch size was 16, the total number of training epochs was 500, and the remaining parameters were set to default values.

To compare the effects of different conditions on the reconstruction of the 3DGS model, a large number of comparative experiments were conducted, which were mainly completed in nerfstudio-1.0.0. The NeRF model used was nerfacto model, and the 3DGS model used was splatfacto model. The segmented cotton bolls in 3D generated 3DGS model were visualized could be displayed on Super Splat (https://playcanvas.com).

2.8. Evaluation metrics

Precision (P), Recall (R), F1-Score, and mean Intersection over Union

(mIoU) were used to evaluate the performance of the 2D segmentation results of YOLOv11x and SAM. P and R represent the ratio of correctly predicted pixels to the total predicted pixels and the total ground truth pixels, respectively. The F1-score is the harmonic mean of P and R, providing a balance between the two metrics. mIoU was used to measure the overlap between the number of predicted pixels and the number of ground truth pixels.

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \tag{13}$$

$$mIoU = \frac{\sum_i^N IoU_i}{N} \tag{14}$$

$$P = \frac{TP}{TP + FP} \tag{15}$$

$$R = \frac{TP}{TP + FN} \tag{16}$$

$$F_1 - Score = \frac{2PR}{P + R} \tag{17}$$

In addition, the trait extraction algorithms were evaluated by the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The coefficient of determination (R^2) was also calculated to assess the performance. All the evaluation metrics were defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(N_i - m_i)^2} \tag{18}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|N_i - m_i| \tag{19}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{N_i - m_i}{m_i}\right| \tag{20}$$

Where TP, TN, FP, and FN represent the number of true positive, true negative, false positive, and false negative points of a certain class, respectively. N_i is the predicted value of the i-th sample by the algorithm, m_i is the true value of the i-th sample, and n is the number of plants.

Mean Squared Error (MSE) was used to evaluate the difference between two images, with smaller MSE values indicating greater similarity (Eq. (21). Here, m represents the number of rows (height) in the image, and n represents the number of columns (width). $I(i,j)$ refers to the pixel value of the original image at position (i,j), while $K(i,j)$ represents the pixel value of the processed image at the same position. Peak Signal-to-Noise Ratio (PSNR) is widely used for quantitatively evaluating image quality, with higher values indicating a better match to the reference image (Eq. (22). PSNR was calculated based on *MSE*, where *MAX* represents the maximum feasible pixel intensity value of image *I*.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{21}$$

$$PSNR = 10 \bullet log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{22}$$

The Structural Similarity Index (SSIM) was used to measure image similarity from brightness, contrast, and structure. μ_x and μ_y are the mean of images x and y. σ_x and σ_y are the variances. σ_{xy} is the covariance of images x and y. C_1, C_2 and C_3 are smoothness constant, contrast constant and structure constant respectively. Learned Perceptual Image Patch Similarity (LPIPS) is also called perceptual loss (Zhang et al., 2018). It works by comparing deep features of image patches that capture the perception of image quality in the human visual system.

Table 1
Effect of image resolution on the quality of 3D models from Cotton3DGaussions. The original resolution was x (4320 × 2160), and it is downsampled by factors of 2, 4, and 8, respectively.

Resolutions	PSNR ↑	SSIM ↑	LPIPS ↓	FPS ↑
x	33.99	0.9560	0.2095	10.88
$x/2$	34.25	0.9551	0.1486	35.73
$x/4$	35.19	0.9709	0.0639	90.33
$x/8$	36.28	0.9816	0.0158	135.88
NeRF (x)	27.08	0.9131	0.2337	0.1759

Table 2
Effects of different training steps on the 3DGS model.

# of iteration steps	PSNR ↑	SSIM ↑	LPIPS ↓
50	16.49	0.7273	0.5860
100	18.30	0.7896	0.4684
1000	30.00	0.9366	0.1387
5000	33.94	0.9650	0.0712
10,000	**35.02**	**0.9690**	**0.0658**
20,000	35.99	0.9712	0.0619
30,000	**36.23**	**0.9717**	**0.0598**
50,000	36.36	0.9718	0.0589
80,000	36.40	0.9718	0.0596
100,000	35.38	0.9702	0.0616

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)(\sigma_{xy} + c_3)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)(\sigma_x\sigma_y + c_3)} \tag{23}$$

3. Results

3.1. Factors affecting the quality of cotton plant 3DGS model

The resolution of the input image data significantly affected the quality of the 3DGS model. After selecting 150 images and training for 30,000 iterations, the highest PSNR was achieved with the lowest resolution of the input images. This was primarily because low-resolution images suppressed high-frequency details, reducing the model's demands for reconstructing textures and edges. The choice of image resolution should therefore be tailored to the specific requirements of the task. In subsequent comparative experiments, the image resolution was set to 960 × 540. A notable advantage of 3DGS was its efficient rendering capability, achieving 35.73 FPS at 1080p resolution (Table 1). Furthermore, as the image resolution decreased, rendering speed improved significantly.

The 3DGS model demonstrated clear advantages over NeRF when tested on cotton plants. A direct comparison of 3DGS and NeRF under the same conditions revealed that 3DGS outperformed NeRF in all metrics, with PSNR being 6.91 higher. Additionally, 3DGS achieved a significant improvement in rendering speed, with a frame rate 62 times faster than NeRF. These findings highlight the superior performance of 3DGS in both 3D model quality and computational efficiency, making it an ideal choice for applications that require high-quality 3D models along with fast processing capabilities.

The performance of 3DGS did not improve further when the optimization reached 30,000 iterations. Training with 150 images, the model achieved high-quality results at 10,000 iterations (PSNR = 35.02). After 20,000 and 30,000 iterations, all metrics had stabilized, with PSNR reaching 36 (Table 2). Excessive optimization not only fails to enhance the reconstruction performance of 3DGS but also results in significant time and resource consumption, potentially introducing noise and degrading rendering quality. In subsequent tests, 30,000 iterations were chosen.

Experiments with varying numbers of training images (at a resolution of 960 × 540 and 30,000 training steps) revealed that for reconstructing a single cotton plant, 100 images were sufficient to achieve

Table 3
Effects of different numbers of training images on 3DGS.

# of images	PSNR ↑	SSIM ↑	LPIPS ↓
50	31.23	0.9430	0.0802
100	35.39	0.9643	0.0641
150	35.61	0.9654	0.0662
300	36.63	0.9692	0.0664
600	36.76	0.9703	0.0680

Table 4
Effects of different data collection methods on 3DGS.

Data collection	# of images	PSNR↑	SSIM ↑	LPIPS ↓
Global shot	300	36.63	0.9692	0.0664
	600	36.76	0.9703	0.0680
Local shot	300	31.94	0.9598	0.0637
	600	32.85	0.9647	0.0578

high-quality results. The number of training images is closely related to the scale of the reconstruction object. When using more than 100 images, the PSNR improvement was marginal, being only 1.37 higher using 600 images than using 100 images (Table 3). However, using 600 images significantly increased computational demands during training and posed challenges when estimating camera parameters with COLMAP. Therefore, we chose about 100 images for the training data set in this study.

Among the various data collection methods, the best results were achieved when the entire plant was within the camera's field of view (i.e., the global shooting collection method). Table 4 presents the outcomes of training with an image resolution of 960 × 540 and 30,000 steps. While different data acquisition approaches can produce good 3D reconstructions, images obtained through the global shooting method demonstrated better performance than the local shot method across evaluation metrics. Specifically, when using 300 and 600 training images, the PSNR of local shooting was lower than that of global shot. Local shooting may capture finer details and textures, but global shooting encompasses a broader range of the scene. The complexity and level of detail in the scene can impact the signal-to-noise ratio of the images, leading to variations in PSNR values.

In comparisons of different cotton 3D models (Fig. 6), 3DGS maintained higher fidelity at the image level than other methods. While the point cloud collected by LiDAR surpasses COLMAP reconstruction results, LiDAR is expensive and slow to collect. Its advantage lies in directly capturing the physical dimensions of the target. However, the plant represented by the point cloud consists of discrete points, often resulting in broken branches. COLMAP's reconstruction was the least effective; even after denoising, significant noise remained on the branches. The point cloud generated by the NeRF model is superior to both COLMAP and LiDAR but requires longer training times and greater computational resources. Additionally, NeRF's point cloud is implicitly stored in a neural network, making it an implicit 3D representation that cannot be directly edited like the 3DGS model. The 3DGS provides a more accurate and editable 3D scene representation, balancing fidelity, computational efficiency, and practicality in cotton plant reconstruction. Each point represents the mean value of a 3D Gaussian, encompassing not only spatial position but also color and opacity. This Gaussian distribution represents the scene more accurately than a point cloud.

3.2. *Performance of YOLOv11x and SAM in cotton boll instance segmentation*

Both YOLOv11x and SAM achieved promising results in instance segmentation, and the evaluation of these two methods under different

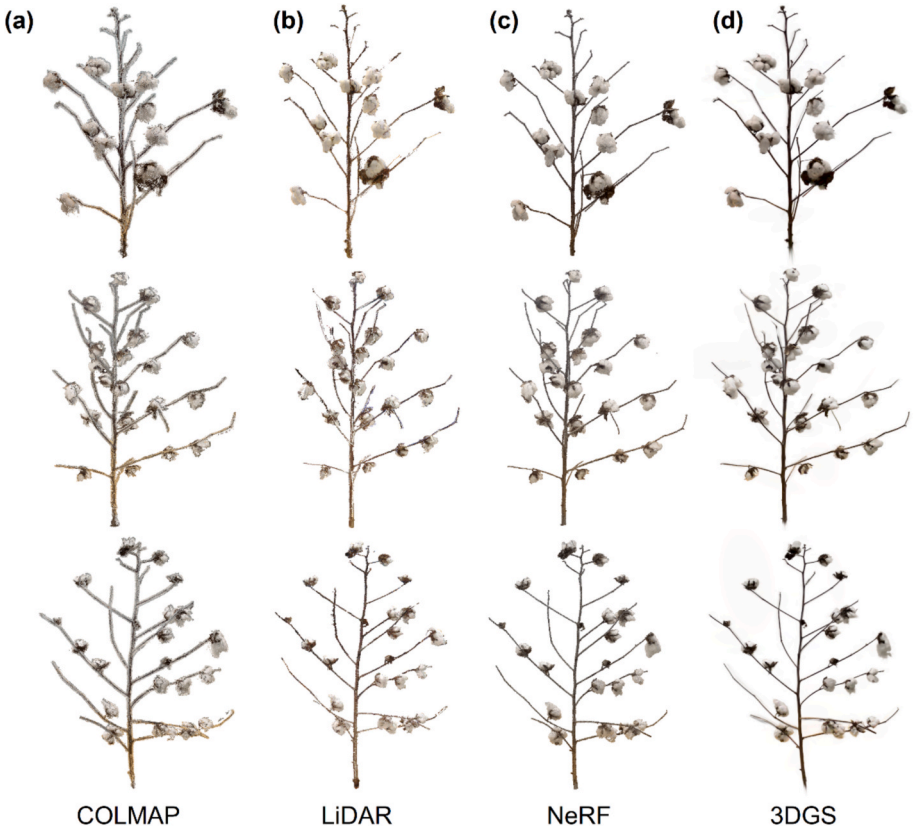

Fig. 6. Comparison of point cloud and 3DGS model of cotton plant. (a) Reconstructed point cloud by COLMAP. (b) Point cloud collected by LiDAR. (c) Reconstructed point cloud by NeRF. (d) 3DGS model.

Table 5
Comparison between YOLOv11x and a foundation model (SAM) on cotton boll instance segmentation in 2D images.

Model	IoU threshold	P	R	F1-Score	mIoU
YOLOv11x	0.25	0.870	0.950	0.903	0.831
	0.50	0.851	0.930	0.884	
	0.75	0.748	0.814	0.775	
SAM	0.25	0.943	0.923	0.931	0.839
	0.50	0.835	0.818	0.825	
	0.75	0.706	0.696	0.700	

IoU thresholds revealed varying outcomes (Table 5). At an IoU threshold of 0.25, SAM achieved an F1-Score that was 2.8 % higher than YOLOv11x. This relatively lenient threshold considers partially segmented masks as correct (Fig. 7). Since SAM's results were generated based on manually provided prompts, it ensured that every cotton boll was assigned a corresponding mask, thereby eliminating the risk of missed detections. This characteristic contributed to SAM's superior performance at this lower IoU threshold. In contrast, at higher IoU thresholds of 0.5 and 0.75, YOLOv11x outperformed SAM, with F1-scores exceeding SAM by 5.9 % and 7.5 %, respectively. These results demonstrate that YOLOv11x can produce more complete and reliable masks for cotton boll segmentation. However, YOLOv11x encounters challenges related to missed detections. For mIoU, both methods achieve comparable results, with SAM showing a slight advantage. This shows that the zero-shot foundation model SAM has great potential.

3.3. 3D cotton boll segmentation

The Cotton3DGaussians model successfully achieved 3D instance segmentation of cotton bolls. The segmented cotton bolls retained high fidelity comparable to image-level details, with each boll and even individual cotton petals clearly visible (Fig. 8b). The 3DGS model preserved significantly more detailed information compared to point clouds captured by LiDAR (Jiang et al., 2022). To better distinguish individual instances, each cotton boll's 3D Gaussian mean was assigned a unique color for visualization (Fig. 8c).

Cotton3DGaussians, utilizing multiviews, enabled a more comprehensive and complete segmentation of cotton bolls from the plant. Experiments demonstrated that using a single view could result in undetected cotton bolls, as shown in View 1 and View 2 of Fig. 9. Additionally, due to the limitations of 3D segmentation performance, some bolls might be partially missing, such as Boll 1 in View 4 of Fig. 9. Introducing four different views effectively addressed these issues.

3.4. Evaluation of boll and plant architectural traits

The Cotton3DGaussians workflow successfully derived 3D phenotypic traits based on the boll segmentation results. The overall accuracy of the cotton boll number in the multiview 3DGS segmentation results was 90.77 %, and the predicted boll number shows a high correlation with the manually counted boll numbers (Fig. 10a). The volume of 3D cotton bolls showed a significant positive correlation with individual plant boll weight (Fig. 10b). By using the 3DGS model for boll segmentation and estimating boll weight based on the convex hull volume, the feasibility of this method was demonstrated, providing an efficient, rapid, and cost-effective solution for non-destructive measurement of boll weight. There were three main sources of error in volume estimation. The first was introduced by the scaling factor using the table as a reference as the 3DGS model did not represent the true size. The second source was from the method of volume calculation, where the convex hull volume was affected by noise points. The third source was from the segmentation method for the 3D cotton boll. Although multiview integration improved the segmentation quality, it still could not achieve perfect segmentation accuracy. We manually selected 55 cotton bolls with excellent segmentation results to eliminate the third source of

Fig. 7. Visualize the segmented 2D instance masks. Each column represents a different plant sample. The first row is the raw image, the second is the SAM segmentation result, and the third is the YOLOv11x segmentation result.

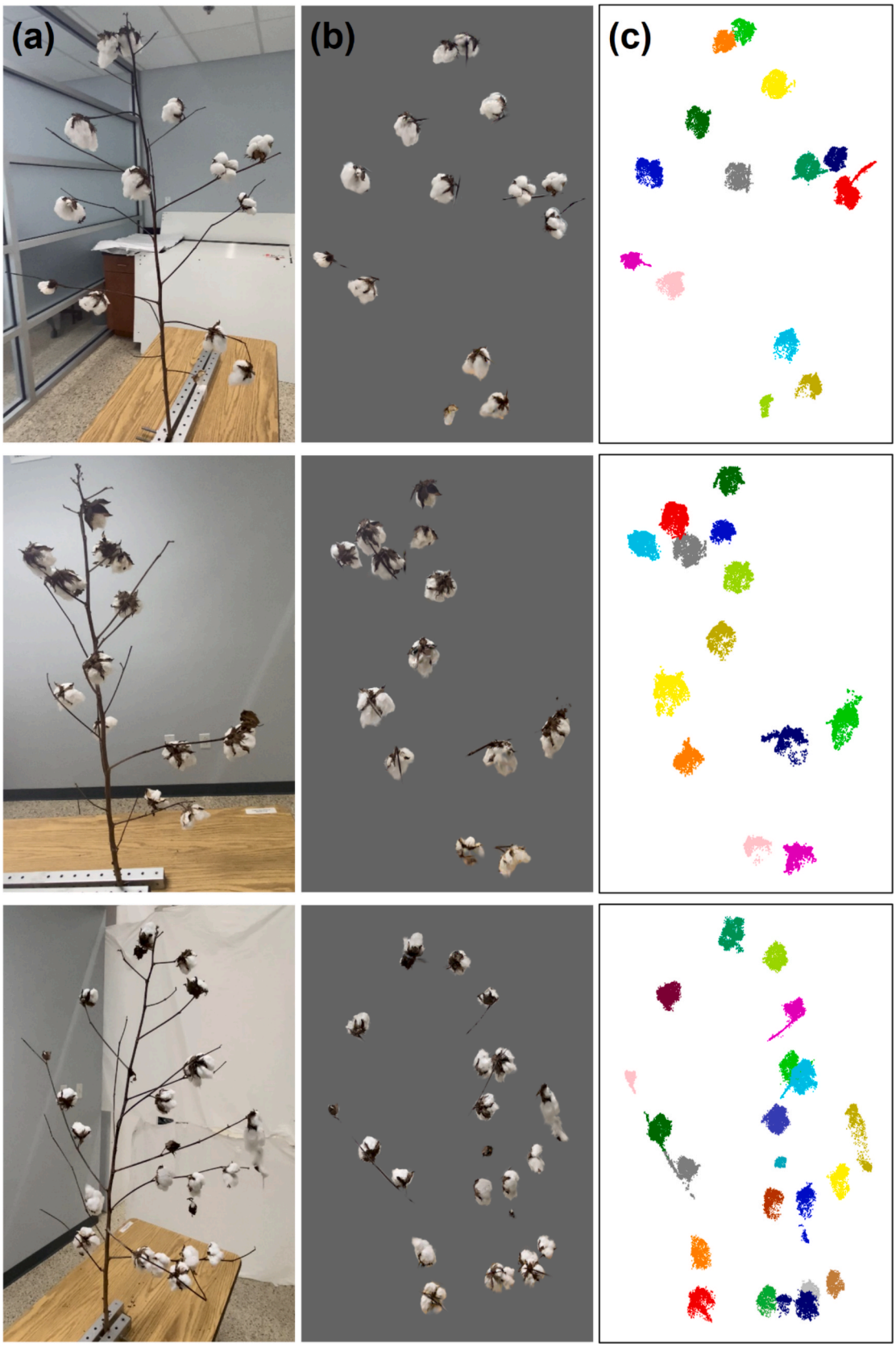

Fig. 8. Cotton boll segmentation results of three different samples (one sample per row). (a) 2D image of a cotton plant from a single view. (b) Segmented 3D Gaussian cotton bolls. All boll instances are shown in one image. (c) Each boll instance was assigned a unique color and displayed as point cloud.

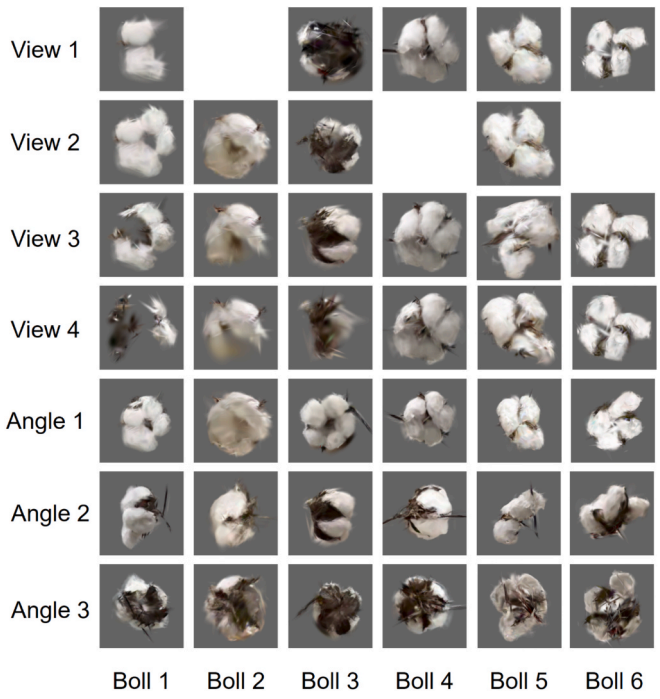

Fig. 9. Segmented 3DGS cotton bolls. Each column represents the same boll. The first four rows show bolls segmented from single views, with missing cells indicating undetected bolls or those exceeding the volume threshold. The last three rows present cross-view clustering results, with each row showing a different angle of the boll.

error, achieving a high correlation with the ground truth obtained by LiDAR (Fig. 10c). This suggests that the primary source of error stems from the quality of cotton boll segmentation. It also demonstrates that the reconstructed 3DGS model can be used to accurately estimate cotton boll volume.

The plant height and canopy size derived from the 3DGS model showed a high degree of correlation with the ground truth measurements obtained through LiDAR. By rapidly reconstructing detailed 3D models, the 3DGS approach offers a fast and cost-effective alternative to traditional methods. Compared to point clouds generated with expensive LiDAR technology, 3DGS not only produced high-quality 3D model reconstructions but also delivered results that were comparable to LiDAR in terms of accuracy for capturing architectural traits (Fig. 11). The ability to accurately measure plant height and canopy size, without the high cost and complexity of LiDAR (Rodriguez-Sanchez et al., 2024), highlights the effectiveness of the 3DGS model as an efficient method for capturing plant architectural traits.

4. Discussion

Our Cotton3DGaussions workflow has several notable advantages for reconstructing high-resolution 3D cotton plants compared to other methods. The traditional Structure-from-Motion (SfM) approach reconstructs 3D models using multiview stereo (MVS), with accuracy relying heavily on image quality and feature extraction. However, it often struggles with complex textures and dense scenes, leading to incomplete reconstructions and a cumbersome workflow (Arief et al., 2024). NeRF as a deep learning-based 3D reconstruction method replaces the MVS pipeline, surface reconstruction, and texture mapping with a neural network, enabling high-fidelity 3D reconstruction. NeRF, however, demands significant training time and high-performance hardware, and its reliance on a single neural network to implicitly store the entire scene limits its utility for downstream tasks (Wang et al., 2024). LiDAR directly captures point clouds with high precision and is resilient to lighting conditions, but its resolution cannot match that of NeRF or 3DGS, and additionally it is costly and less accessible for widespread use. In contrast, 3DGS technology balances cost, efficiency, and accuracy. Once trained, it stores the entire 3D scene in a point cloud format, enabling straightforward editing and manipulation of the reconstructed environment. The explicit representation of 3DGS facilitates the future design of a segmentation model based on 3DGS, similar to how PointNet++ (Qi et al., 2017) performs segmentation on point clouds.

2D segmentation plays a crucial role in extracting masks for 3D segmentation in our workflow. In the comparison between two 2D methods: YOLOv11x and a foundation model (SAM), both achieved relatively high-quality 2D masks, although the former demonstrated a slightly higher accuracy than SAM. However, SAM shows greater potential as it can generate masks without requiring annotations. One major challenge for SAM is replacing manual prompts with automated prompt generation. Some studies aim to achieve SAM segmentation without manual prompts, such as providing uniformly distributed point prompts for the entire image. After SAM performs segmentation, irrelevant backgrounds can be removed based on the size of the masks (Williams et al., 2024). Although the automatically generated point prompt did not produce quality masks for our cotton plant dataset, this method shows promise and warrants further exploration.

In SAM-based segmentation, only a single point prompt was provided for each cotton boll in this study, resulting in incomplete segmentation for some cotton bolls (Fig. 12a, c). This issue can be addressed by providing additional point prompts or using box prompts. Another problem, similar to YOLOv11x, occurred when cotton bolls were in

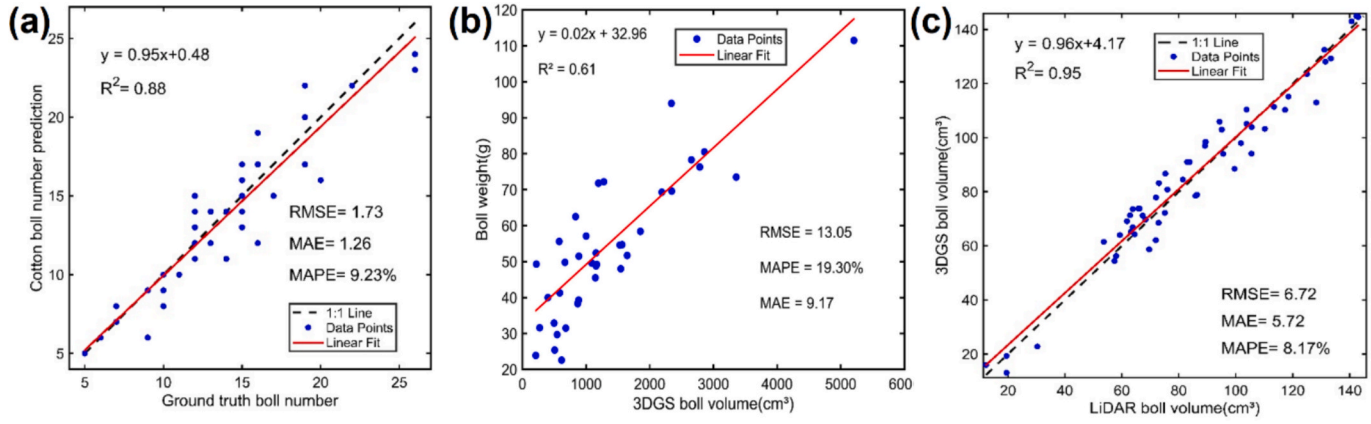

Fig. 10. Evaluation of cotton boll traits. (a) Regression between predicted boll number and the ground truth. (b) Regression between predicted boll volume and boll weight. (c) Regression between the boll volumes measured by Cotton3DGaussions and LiDAR.

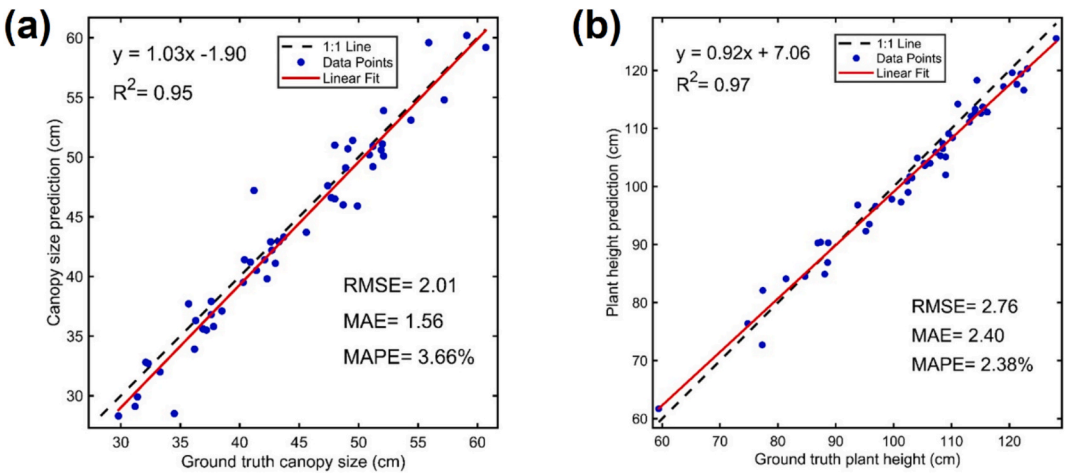

Fig. 11. Phenotypic measurement performance evaluation. (a) Linear regression of cotton plant canopy size. (b) Linear regression of cotton plant height.

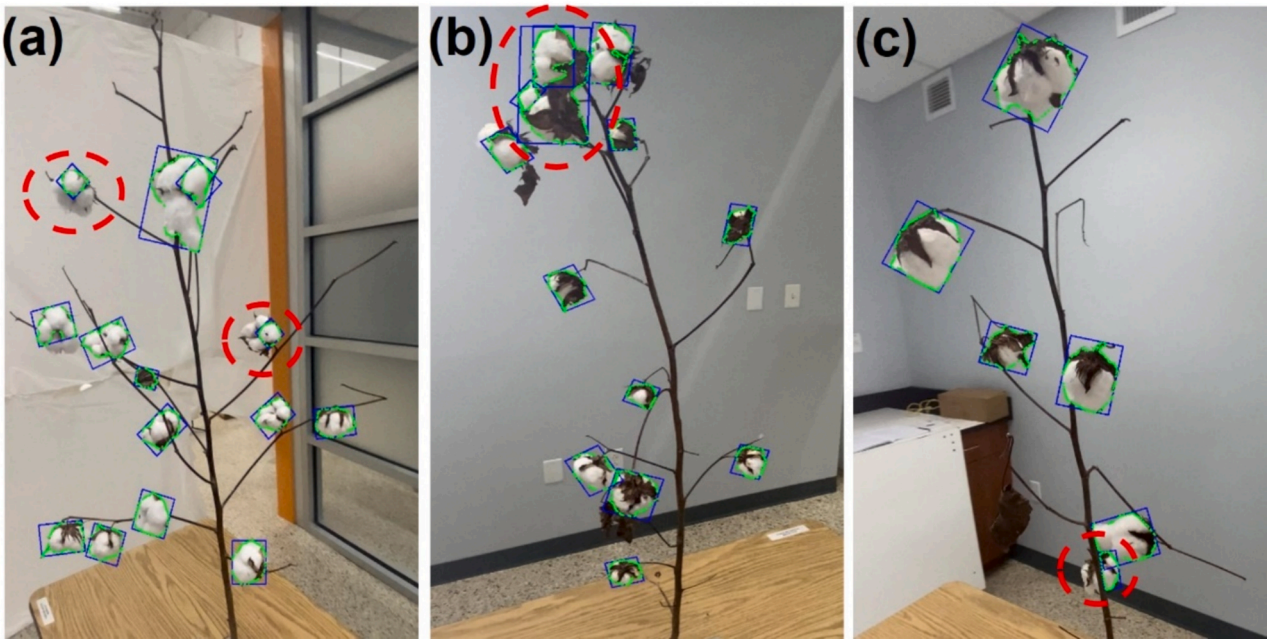

Fig. 12. Error analysis of SAM segmentation. (a) Incomplete boll segmentation. (b) Overlapping segmentation results. (c) Segmentation of bolls occluded by the main stem.

contact with each other, causing two bolls to be segmented as a single entity (Fig. 12b). For SAM, automatically determining a reliable prompt is a bottleneck. In the future, background removal, such as by combining Depth Anything v2 (Yang et al., 2024b), and then providing prompts, will be one research direction.

Multiview segmentation addressed the occlusion problem when obtaining 2D masks. The four different views selected in this experiment were sufficient to segment the masks of all the cotton bolls on the plant. However, the multiview 3DGS segmentation method designed in this study has some limitations. Its performance was affected by certain unique boll growth structures. For example, some bolls grew in clusters at the top of the cotton plant (Fig. 13), and the method struggled to segment these dense areas effectively. This growth pattern was better suited for semantic segmentation. Another segmentation issue arose from the multiview segmentation model. During the process of mapping 2D masks to 3DGS, there were instances where the 3D model did not align with the 2D masks (Fig. 14). The reason for this misalignment was that the 3D segmentation features of the 3DGS model was obtained through training, which had certain deviations.

Compared to estimate traits from LiDAR (Saeed et al., 2023b), the Cotton3DGaussions developed in this study achieved comparable results (plant height) while offering a more cost-effective and rapid data collection method. In the estimation of cotton boll volume, there was some deviation between the predicted and actual values. Although Cotton3DGaussians improved segmentation results by combining multiple views, some cotton bolls may have branches attached, which could affect the volume estimation.

Boll volume was a very sensitive trait, with small changes in the points could lead to significant variations. We considered the red boll in Fig. 15 to be a good segmentation result; however, it differed from the ground truth by 15 cm^3 simply due to the addition of a small branch. This was, of course, also related to the method used for volume calculation. We also presented some other good segmentation results along with their volumes, further confirming the potential of 3DGS in estimating cotton boll volume.

Although the data for this study was collected indoors, which may be

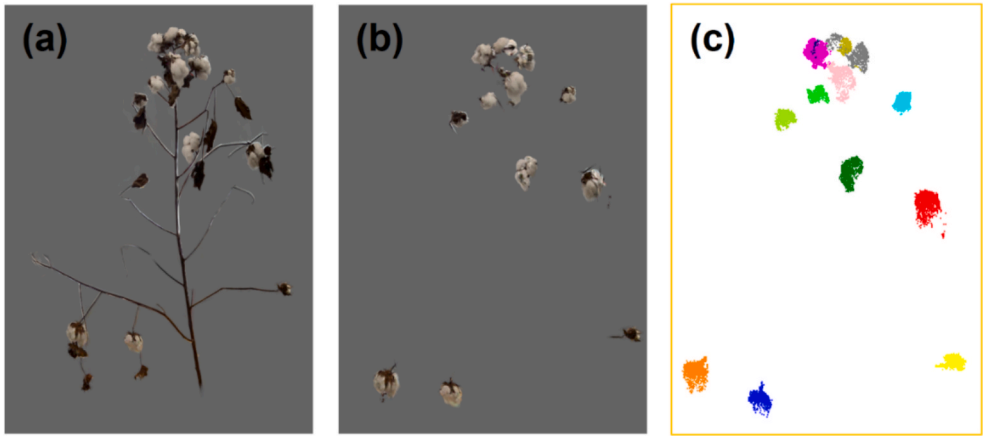

Fig. 13. Analysis of segmentation errors. (a) 3DGS cotton plant model. (b) Segmented 3DGS boll. (c) Each boll instance displayed in color. Errors occur in the top dense region.

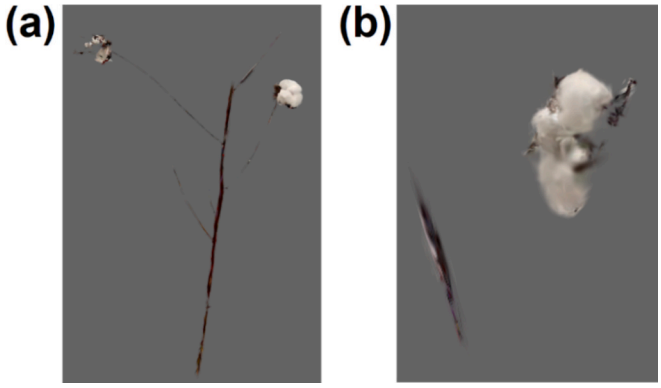

Fig. 14. Example of a boll with poor segmentation. (a) Cotton bolls and a large portion of segmented non-cotton bolls. (b) Cotton bolls and a small section of the main stem.

a limitation, this controlled environment provided a critical foundation for future field applications. Indoor studies allowed precise control over variables such as lighting and background, ensuring reliable data collection and model training. This step was essential for optimizing methods and algorithms before applying them to more complex and variable field conditions. The experiments conducted demonstrated that 3DGS is not only feasible but also effective in accurately analyzing the architectural traits of cotton plants, as shown by its ability to reliably estimate traits such as plant height, canopy size, and boll volume. In the future, data collection will primarily occur under field conditions, and the model will be improved to better segment small objects, such as thin branches.

5. Conclusions

This study presented a novel Cotton3DGaussians workflow that achieved high-fidelity 3D Gaussian Splatting model reconstruction of individual cotton plants, and successfully segmented 3D boll instances by leveraging 2D masks from four views. Our approach derived cotton 3D phenotypic traits such as the number of cotton bolls, boll volume, and other architectural traits with high accuracies in comparison to ground truth. This innovative method provides a novel tool for 3D plant phenotyping and high-resolution plant mapping, which could be extended to other crops. Our future work will involve collecting plot-level data from the field for high-fidelity 3D reconstruction.

CRediT authorship contribution statement

Lizhi Jiang: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Jin Sun:** Writing – review & editing, Supervision, Methodology. **Peng W. Chee:** Writing – review & editing, Supervision, Resources. **Changying Li:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition. **Longsheng**

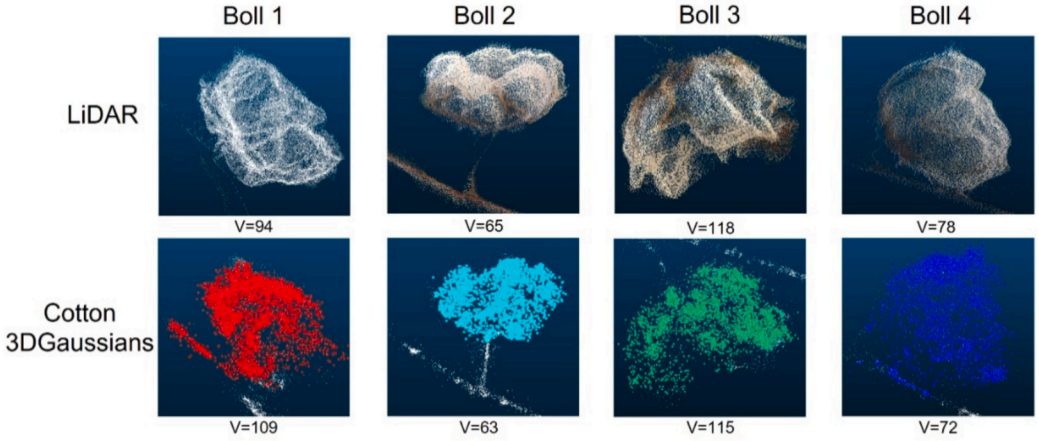

Fig. 15. Pairwise comparisons of boll volumes measured using Cotton3DGaussians and ground truth from LiDAR. The first row shows the LiDAR-captured point cloud, and the colored points in the second row represent segmented cotton bolls (each point is the mean of a Gaussian). V represents the convex volume (cm^3).

Fu: Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the USDA National Institute of Food and Agriculture (Award No. 2023-67021-40646), Cotton Incorporated and the Hatch Project (FLA-ABE-006451) from the University of Florida. The authors sincerely acknowledge Dalton West for his assistance in plant sample collection.

Data availability

Data will be made available on request.

References

Arief, M.A.A., Nugroho, A.P., Putro, A.W., Sutiarso, L., Cho, B.K., Okayasu, T., 2024. Development and application of a low-cost 3-dimensional (3D) reconstruction system based on the structure from motion (SfM) approach for plant phenotyping. J. Biosyst. Eng. 49, 326–336. https://doi.org/10.1007/s42853-024-00237-w.

M.A. Arshad, T. Jubery, J. Afful, A. Jignasu, A. Balu, B. Ganapathysubramanian, S. Sarkar, A. Krishnamurthy, 2024. Evaluating NeRFs for 3D Plant Geometry Reconstruction in Field Conditions 1–17. https://doi.org/10.34133/plantphenomics.0235.

J. Cen, J. Fang, C. Yang, L. Xie, X. Zhang, W. Shen, Q. Tian, Segment Any 3D Gaussians. arXiv:2312.00860 1–10. https://doi.org/10.48550/arXiv.2312.00860.

Choi, H.B., Park, J.K., Park, S.H., Lee, T.S., 2024. NeRF-based 3D reconstruction pipeline for acquisition and analysis of tomato crop morphology. Front. Plant Sci. 15, 1–12. https://doi.org/10.3389/fpls.2024.1439086.

F. Devoto, S. Reynolds-Massey-Reed, P.C. Segura, M. Bell, T. Mclaren, R. Awale, C. Camino, M. Bange, W. Woodgate, S. Chapman, A.B. Potgieter, 2024. Insights in the Ability of High-Resolution Narrow Band Multispectral and Thermal Sensors to Estimate Cotton Production in Australia. IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium 1510–1513. https://doi.org/10.1109/igarss53475.2024.10642663.

Dong, Z., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppäb, J., Stilla, U., 2020. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. ISPRS J. Photogramm. Remote Sens. 163, 327–342. https://doi.org/10.1016/j.isprsjprs.2020.03.013.

He, W., Ye, Z., Li, M., Yan, Y., Lu, W., Xing, G., 2023. Extraction of soybean plant trait parameters based on SfM-MVS algorithm combined with GRNN. Front. Plant Sci. 14, 1–16. https://doi.org/10.3389/fpls.2023.1181322.

Hu, K., Wei, Y., Pan, Y., Kang, H., Chen, C., 2024. High-fidelity 3D reconstruction of plants using neural radiance field. Comput. Electron. Agric. 220, 108848. https://doi.org/10.1016/j.compag.2024.108848.

Huang, T., Bian, Y., Niu, Z., Taha, M.F., He, Y., Qiu, Z., 2024. Fast neural distance field-based three-dimensional reconstruction method for geometrical parameter extraction of walnut shell from multiview images. Comput. Electron. Agric. 224, 109189. https://doi.org/10.1016/j.compag.2024.109189.

Jamil, N., Kootstra, G., Kooistra, L., 2022. Evaluation of individual plant growth estimation in an intercropping field with UAV imagery. Agriculture (Switzerland) 12, 1–23. https://doi.org/10.3390/agriculture12010102.

Jiang, L., Li, C., Fu, L., 2025. Apple tree architectural trait phenotyping with organ-level instance segmentation from point cloud. Comput. Electron. Agric. 229, 109708. https://doi.org/10.1016/j.compag.2024.109708.

Jiang, L., Li, C., Fu, L., 2022. 3D deep learning-based segmentation to reveal the spatial distribution of cotton bolls. In: In: 2022 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, p. 1.

Jiang, Y., Li, C., Takeda, F., Kramer, E.A., Ashrafi, H., Hunter, J., 2019. 3D point cloud data to quantitatively characterize size and shape of shrub crops. Hortic. Res. 6. https://doi.org/10.1038/s41438-019-0123-9.

Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G., 2023. 3D gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. 42, 1–14. https://doi.org/10.1145/3592433.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R., 2023. Segment Anything. In: Proceedings of the IEEE International Conference on Computer Vision 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371.

Li, L., Zhang, S., Wang, B., 2021. Plant disease detection and classification by deep learning—A review. IEEE Access 9, 56683–56698. https://doi.org/10.1109/ACCESS.2021.3069646.

Liu, T., Zhu, S., Yang, T., Zhang, W., Xu, Y., Zhou, K., Wu, W., Zhao, Y., Yao, Z., Yang, G., Wang, Y., Sun, C., Sun, J., 2024. Maize height estimation using combined unmanned aerial vehicle oblique photography and LIDAR canopy dynamic characteristics. Comput. Electron. Agric. 218, 108685. https://doi.org/10.1016/j.compag.2024.108685.

Liu, Y., Yuan, H., Zhao, X., Fan, C., Cheng, M., 2023. Fast reconstruction method of three-dimension model based on dual RGB-D cameras for peanut plant. Plant Methods 19, 1–16. https://doi.org/10.1186/s13007-023-00998-z.

L. Meyer, A. Gilson, U. Schmid, M. Stamminger, 2024. FruitNeRF: A Unified Neural Radiance Field based Fruit Counting Framework. arXiv preprint arXiv:2408.06190.

Miao, T., Zhu, C., Xu, T., Yang, T., Li, N., Zhou, Y., Deng, H., 2021. Automatic stem-leaf segmentation of maize shoots using three-dimensional point cloud. Comput. Electron. Agric. 187, 106310. https://doi.org/10.1016/j.compag.2021.106310.

Mildenhall, B., Srinivasan, P.P., Ortiz-Cayon, R., Kalantari, N.K., Ramamoorthi, R., Ng, R., Kar, A., 2019. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. ACM Trans. Graph. 38, 1–14. https://doi.org/10.1145/3306346.3322980.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2022. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. Commun. ACM 65, 99–106. https://doi.org/10.1145/3503250.

Ojo, T., La, T., Morton, A., Stavness, I., 2024. Splanting: 3D plant capture with gaussian splatting. Proceedings - SIGGRAPH Asia 2024 Technical Communications. SA 2024, 1–4. https://doi.org/10.1145/3681758.3698009.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems 2017-Decem, 5100–5109. https://doi.org/10.48550/arXiv.1706.02413.

Rodriguez-Sanchez, J., Snider, J.L., Johnsen, K., Li, C., 2024. Cotton morphological traits tracking through spatiotemporal registration of terrestrial laser scanning time-series data. Front. Plant Sci. 15, 1–23. https://doi.org/10.3389/fpls.2024.1436120.

Saeed, F., Sun, J., Ozias-Akins, P., Chu, Y.J., Li, C.C., 2023a. PeanutNeRF: 3D radiance field for peanuts. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. https://doi.org/10.1109/CVPRW59228.2023.00665.

Saeed, F., Sun, S., Sanchez, J.R., Snider, J., Liu, T., Li, C., 2023b. Cotton plant part 3D segmentation and architectural trait extraction using point voxel convolutional neural networks. Plant Methods 19, 1–23. https://doi.org/10.1186/s13007-023-00996-1.

Schonberger, J.L., Frahm, J.M., 2016. Structure-from-motion revisited. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2016.445.

Smitt, C., Halstead, M., Zimmer, P., Labe, T., Guclu, E., Stachniss, C., Mccool, C., 2024. PAg-NeRF: Towards fast and efficient end-to-end panoptic 3D representations for agricultural robotics. IEEE Rob. Autom. Lett. 9, 907–914. https://doi.org/10.1109/LRA.2023.3338515.

Sun, S., Li, C., Chee, P.W., Paterson, A.H., Jiang, Y., Xu, R., Robertson, J.S., Adhikari, J., Shehzad, T., 2020. Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. ISPRS J. Photogramm. Remote Sens. 160, 195–207. https://doi.org/10.1016/j.isprsjprs.2019.12.011.

Sun, S., Li, C., Chee, P.W., Paterson, A.H., Meng, C., Zhang, J., Ma, P., Robertson, J.S., Adhikari, J., 2021. High resolution 3D terrestrial LiDAR for cotton plant main stalk and node detection. Comput. Electron. Agric. 187, 106276. https://doi.org/10.1016/j.compag.2021.106276.

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. Front. Plant Sci. 14, 1–13. https://doi.org/10.3389/fpls.2023.1109314.

G. Wang, L. Pan, S. Peng, S. Liu, C. Xu, Y. Miao, W. Zhan, M. Tomizuka, M. Pollefeys, H. Wang, NeRF in Robotics: A Survey. arXiv:2405.01333 1–21. https://doi.org/10.48550/arXiv.2405.01333.

Wang, Y., Hu, S., Ren, H., Yang, W., Zhai, R., 2022. 3DPhenoMVS: A Low-Cost 3D tomato phenotyping pipeline using 3D reconstruction point cloud based on multiview images. Agronomy 12, 1865. https://doi.org/10.3390/agronomy12081865.

Williams, D., Macfarlane, F., Britten, A., 2024. Leaf only SAM: A segment anything pipeline for zero-shot automated leaf segmentation. Smart Agric. Technol. 8, 100515. https://doi.org/10.1016/j.atech.2024.100515.

Xiao, S., Ye, Y., Fei, S., Chen, H., Zhang, B., li, Q., Cai, Z., Che, Y., Wang, Q., Ghafoor, A. Z., Bi, K., Shao, K., Wang, R., Guo, Y., Li, B., Zhang, R., Chen, Z., Ma, Y., 2023. ISPRS J. Photogramm. Remote Sens. 201, 104–122. https://doi.org/10.1016/j.isprsjprs.2023.05.016.

L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, H. Zhao, 2024. Depth Anything V2. arXiv 1–30. https://doi.org/10.48550/arXiv.2406.09414.

X. Yang, X. Lu, P. Xie, Z. Guo, H. Fang, H. Fu, X. Hu, Z. Sun, H. Cen, 2024. PanicleNeRF : Low-Cost , High-Precision In-Field Phenotyping of Rice Panicles with Smartphone 1–14. https://doi.org/10.34133/plantphenomics.0279.

Ye, M., Danelljan, M., Yu, F., Ke, L., 2023. Gaussian grouping: Segment and Edit anything in, 3D Scenes.

H. Ying, Y. Yin, J. Zhang, F. Wang, T. Yu, R. Huang, L. Fang, 2023. OmniSeg3D: Omniversal 3D Segmentation via Hierarchical Contrastive Learning 20612–20622.

Zhang, C., Jiang, Y., Xu, B., Li, X., Zhu, Y., Lei, L., Chen, R., Dong, Z., Yang, H., Yang, G., 2020. Apple tree branch information extraction from terrestrial laser scanning and backpack-LiDAR. Remote Sens. (Basel) 12, 1–17. https://doi.org/10.3390/rs12213592.

Zhang, J., Wang, X., Ni, X., Dong, F., Tang, L., Sun, J., Wang, Y., 2024. Neural radiance fields for multi-scale constraint-free 3D reconstruction and rendering in orchard scenes. Comput. Electron. Agric. 217, 108629. https://doi.org/10.1016/j.compag.2024.108629.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE

Computer Society Conference on Computer Vision and Pattern Recognition 586–595. https://doi.org/10.1109/CVPR.2018.00068.
Zhao, H., Chen, Y., Liu, J., Wang, Z., Li, F., Ge, X., 2023. Recent advances and future perspectives in early-maturing cotton research. New Phytol. 237, 1100–1114. https://doi.org/10.1111/nph.18611.
Zheng, X., Ai, X., Qin, H., Rong, J., Zhang, Z., Yang, Y., Yuan, T., Li, W., 2024. Tomato-nerf: Advancing tomato model reconstruction with improved neural radiance fields. IEEE Access 12, 184206–184215. https://doi.org/10.1109/ACCESS.2024.3424908.