agronomy

MDPI

*Article*

# RQ-PointNeXt: An End-to-End 3D Point Cloud Instance Segmentation Method for Field Cotton Boll Phenotyping

**Haoyuan Niu [1], Yuxiang Wang [2], Xiaoyan Meng [1], Xi Cheng [1], Wenbin Zhang [3], Hao Qiu [1] and Yunjie Zhao [1,*]**

[1] College of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China
[2] College of Agronomy, Xinjiang Agricultural University, Urumqi 830052, China
[3] Xinjiang Jinfengyuan Seed Industry Co., Ltd., Aksu 843100, China
* Correspondence: zyj@xjau.edu.cn

**Abstract**

Accurate point-cloud segmentation of cotton organs is essential for precise phenotypic characterization. However, reliable instance segmentation of cotton bolls in field-derived point clouds remains challenging because foliage occlusion, contact between adjacent bolls and incomplete reconstruction obscure instance boundaries. Here we present RQ-PointNeXt, an end-to-end framework that directly maps input point clouds to boll instance masks within a unified trainable network. Built on PointNeXt, it incorporates relative-elevation geometric channel attention in the shallow encoder to fuse global channel context with elevation and surface-normal cues. Its Query–mask branch integrates semantic guidance, center-seeded queries and a center-aware mask prior to suppress background responses, localize instances and constrain mask extent. Hungarian matching and multitask optimization establish one-to-one query–instance assignments, whereas query-based decoding produces instance masks without external geometric clustering. We evaluated the framework on 226 field-grown cotton plants containing 720 annotated boll instances reconstructed from UAV multi-view imagery using neural radiance fields. On the held-out test set, overall accuracy, mean class accuracy and mean intersection over union reached 0.8942, 0.8980 and 0.8076, respectively. AP25, AP50 and AP75 were 0.7359, 0.5585 and 0.2777, yielding an mAP25/50/75 of 0.5240. The framework provides instance-level outputs for boll counting and spatial analysis in high-throughput field phenotyping.

**Keywords:** cotton boll; three-dimensional point cloud; query-based instance segmentation; PointNeXt; neural radiance field reconstruction; field phenotyping



Academic Editor: Francis Drummond

Received: 20 August 2026
Revised: 8 September 2026
Accepted: 9 September 2026
Published: 11 September 2026



## 1. Introduction

Cotton is a major global cash crop and a key source of natural fiber for the textile industry [1]. Drought and heat stress, whether occurring individually or concurrently, inhibit cotton growth and reduce both yield and fiber quality, with their combined effects generally being more severe than those of either stress alone [2]. Rapid and accurate quantification of multidimensional phenotypic traits, including plant height and canopy volume, is therefore essential for monitoring cotton growth and predicting yield [3]. Conventional field phenotyping relies largely on manual observation, measurement, and sampling. These procedures are time-consuming, labor-intensive, and costly and are susceptible to operator-induced errors, limiting their suitability for large-scale germplasm screening and continuous monitoring throughout the crop growth cycle [4,5]. UAV-based phenotyping platforms, by contrast, offer flexible operation, straightforward deployment, broad spatial coverage, and efficient data acquisition, thereby enabling rapid, non-destructive, and high-throughput monitoring of field crops [4].

Advances in digital imaging, three-dimensional (3D) sensing, and computer vision have established two-dimensional (2D) imagery and 3D point clouds as important data modalities for non-contact crop phenotyping [6,7]. Image-based high-throughput phenotyping supports non-destructive data acquisition while reducing reliance on manual measurements, thereby facilitating the large-scale collection and analysis of plant phenotypic data [8]. Plant image segmentation encompasses conventional approaches based on image attributes, such as color indices and thresholding, as well as deep learning approaches built primarily on convolutional neural networks. In recent years, the latter have become a major focus of agricultural image segmentation research [9]. For example, Yi et al. [10] incorporated multiscale feature fusion, atrous spatial pyramid pooling, and a cross-attention mechanism into U-Net to achieve pixel-level semantic segmentation of stems and leaves in Betula luminifera seedlings. Nevertheless, a single-view 2D image cannot fully capture the depth and 3D spatial relationships among plant organs, while structures occluded by foreground organs remain unobserved. Consequently, complete plant architecture is difficult to reconstruct when stems and leaves intersect, occlude one another, or overlap [7]. Field imagery is also affected by uneven illumination and distinct shadow boundaries, which can result in segmentation errors; thus, the robustness of existing agricultural image segmentation methods under complex field conditions requires further improvement [9].

Supported by advances in three-dimensional (3D) imaging, computer vision, and deep learning, 3D point clouds have become an important data modality for crop phenotyping [11,12]. Compared with two-dimensional (2D) images, point clouds encode the 3D coordinates and spatial organization of plant surfaces, providing a basis for analyzing plant architecture, segmenting organs, and quantifying phenotypic traits such as canopy structure, leaf area, and stem diameter [11,12]. Nevertheless, point-cloud acquisition under field conditions remains susceptible to organ occlusion, wind-induced motion, variable illumination, and complex terrain. It would therefore be inaccurate to assume that point clouds can completely eliminate occlusion or are inherently robust to environmental variability [13,14]. Deep learning has increasingly been applied to organ segmentation and phenotypic trait extraction from point clouds of cotton and other crops. Qiu et al. [15] developed ELGCot3D, which integrates EdgeConv, local attention, graph convolution, and a semantic feature enhancement module for organ-level semantic segmentation of cotton point clouds; the model achieved an mIoU of 76.7% on the cotton validation set of Crops3D, with 0.86 million parameters. Liu et al. [16] incorporated a Transformer-based attention module into PointNet++ to construct TPointNetPlus for semantic segmentation of cotton organs, after which HDBSCAN was applied to separate leaves, cotton bolls, and branches. Chen et al. [3] acquired field cotton point clouds using UAV LiDAR, employed PointNet++ to extract cotton point clouds at the individual-plant and plot scales, and subsequently calculated phenotypic traits including plant height, porosity, and canopy volume. Yao et al. [17] reconstructed field tobacco point clouds from UAV-based multi-view images and incorporated local spatial encoding and density-aware pooling into PointNet++, enabling semantic segmentation of stems and leaves and automated measurement of plant height, leaf length, leaf width, leaf number, and internode length. Dong et al. [18] combined PointNeXt-based semantic segmentation with Quickshift++ leaf clustering in a two-stage framework and evaluated organ-level instance segmentation on point clouds of sugarcane, maize, and tomato.

A growing number of network architectures have also been developed specifically for plant point-cloud segmentation. RoseSegNet was designed for organ-level semantic segmentation of leaves, stems, and flowers in rose plants [19], whereas Pattern-Net was applied to the semantic separation of wheat ears from non-ear regions [20]. PlaneSegNet was developed to distinguish plant from non-plant points in agricultural scenes containing multiple plants [21]. Despite these advances, much of the existing research on plant point-cloud segmentation has relied on data acquired from controlled environments, isolated plants, or individual organ samples, while studies and validation involving plants growing in situ under field conditions remain relatively limited [13,22]. In actual field settings, overlapping and touching leaves complicate the separation of individual plants and can produce substantial errors along segmentation boundaries. Soil-related background features, environmental noise, and nonuniform point distributions further increase the difficulty of background removal and point-cloud preprocessing [23]. Moreover, agricultural point-cloud datasets collected under real field conditions remain relatively scarce, and discrepancies between simulated or controlled environments and actual fields pose challenges to the practical deployment of existing methods [24]. Given these limitations, this study investigates cotton boll instance segmentation using point clouds of individual cotton plants acquired in situ under field conditions.

For field-acquired point clouds of individual cotton plants, foreground recognition and individual-boll separation are closely related. Foreground recognition is affected by class imbalance [25] and the loss of local structure during downsampling, whose effects vary across network architectures [26] and can weaken boundary cues in overlapping canopy regions. At the instance level, PointGroup groups semantic predictions and center offsets through geometric clustering [27], while Dong et al. combined PointNeXt with Quickshift++ clustering for plant organ instance segmentation [18]. These two-stage methods determine instance IDs after semantic prediction, making their results dependent on clustering parameters, point-cloud connectivity, and offset distributions, which may cause adjacent instances to merge or produce incomplete point aggregation [28]. PlantSegNet addressed adjacent plant organs with identical semantic labels [22]. PointNeXt is primarily a point-cloud feature extraction backbone [29]. When used for semantic segmentation, an additional clustering method such as DBSCAN is required to generate instance IDs, making the results sensitive to clustering parameters and point-cloud connectivity. By contrast, Mask3D uses object queries to predict instance masks in general 3D scenes [30]. However, existing approaches do not specifically combine relative elevation, surface-normal information, and center constraints for cotton boll segmentation in field-reconstructed point clouds. To address this gap, this study extends PointNeXt with R-EGCA and a center-constrained Query–mask decoder. R-EGCA introduces these geometric cues into shallow feature extraction, while the Query–mask decoder directly predicts individual boll masks without clustering-based local instance generation. The principal contributions of this study are summarized as follows:

1. An R-EGCA module is introduced at the initial high-resolution stage of the PointNeXt encoder. It combines global channel context with normalized relative elevation and surface-normal information to generate channel-wise attention weights, and integrates the reweighted features through a learnable residual connection. This design incorporates vertical position and surface orientation into shallow feature extraction for cotton boll foreground recognition and boundary representation.
2. A center-constrained Query–mask branch is integrated with PointNeXt for end-to-end cotton boll instance segmentation. Center-seeded and learned queries represent candidate bolls, while a center-aware mask prior introduces point-to-center distances as a spatial bias in mask prediction. The branch is trained through Hungarian matching

and query classification and mask losses, generating local instance masks directly rather than clustering semantic predictions.

3. A field dataset containing 226 individual cotton plants and 720 annotated bolls is constructed from UAV imagery using NeRF reconstruction. Comparative and ablation experiments evaluate the proposed method and its components on this dataset, with additional evaluation on the public UGA-BSAIL dataset to examine performance under different point-cloud acquisition conditions.

## 2. Materials and Methods

### *2.1. Study Site and Plant Materials*

The field cotton data used in this study were collected at Huaxing Farm in Changji, Xinjiang Uygur Autonomous Region, China (44°02′ N, 87°30′ E), to assess the applicability of the proposed model under actual cotton-field conditions. Cotton plants were imaged at maturity and following the first mechanical harvesting operation. The raw data comprised close-range, multi-view images acquired using an unmanned aerial vehicle (UAV). Individual cotton plant point clouds were subsequently extracted from the NeRF-based reconstructions, after which semantic and instance labels were assigned to the cotton bolls. Close-range oblique images were acquired using a DJI Mavic 3 Multispectral (Mavic 3M) UAV equipped with an integrated RGB camera (SZ DJI Technology Co., Ltd., Shenzhen, China) along circular and rectilinear-grid flight paths to capture the three-dimensional (3D) canopy structure. The relative flight altitude was 5 m, the radius of the circular path was 2.5 m, the flight speed was 0.5 m $s^{-1}$, the gimbal pitch angle was 45°, and the forward overlap was 45%. This flight design increased the overlap among images captured from different viewpoints and provided coverage of both the upper and lateral canopy structures, thereby improving image coverage and view redundancy for subsequent NeRF reconstruction and high-quality 3D point-cloud generation.

Mature field-grown cotton has a complex canopy architecture, and intertwined branches and leaves, together with interplant occlusion, make it difficult to observe local plant structures adequately from a single viewpoint. Three-dimensional plant point clouds are commonly acquired using multi-view photogrammetry or active sensing technologies such as LiDAR. Multi-view photogrammetry typically combines Structure from Motion (SfM) and Multi-View Stereo (MVS) for camera pose estimation and dense 3D reconstruction [31]. When internal canopy structures are occluded or insufficiently covered by the input views, the reconstructed point clouds may exhibit local sparsity, missing details, or structural discontinuities [32]. Given these limitations, UAV-based multi-view RGB images were combined with Neural Radiance Fields (NeRF) in this study for 3D reconstruction. NeRF learns a continuous implicit representation from multi-view images with known camera poses, modeling scene color and volumetric density and synthesizing views through volume rendering [33]. The trained implicit representation can subsequently be sampled to derive plant point clouds, providing the data required for point-cloud preprocessing, surface-normal estimation, and instance annotation [34]. Following coordinate standardization, denoising, and surface-normal estimation, the reconstructed point clouds were used to construct a cotton boll instance segmentation dataset. The complete workflow is illustrated in Figure 1.

During 3D reconstruction, the multi-view RGB images acquired by the UAV were first screened for suitability. COLMAP was then used to perform image feature matching, camera pose estimation, and sparse scene reconstruction, yielding the camera parameters required for Nerfacto training [35]. The Nerfacto model was subsequently trained within the Nerfstudio framework, and cotton plant point clouds were exported from the resulting neural radiance field [35]. Reconstruction quality depended primarily on image sharpness, the reliability of feature matching, the accuracy of the estimated camera poses, viewpoint coverage, and the visibility of plant structures. Variable illumination, wind-induced motion, complex soil backgrounds, and occlusion between neighboring plants introduced additional challenges under actual field conditions [36]. Moreover, conventional NeRF methods generally assume a static scene; plant movement during image acquisition can introduce image blur and inconsistencies among views, thereby compromising camera pose estimation and 3D reconstruction [35,36]. In the present dataset, wind-driven movement of branches and leaves appeared as local blurring and missing geometry around some cotton bolls. After export, the point clouds were denoised and coordinate-normalized, and their surface normals were estimated. Semantic and instance labels were then manually assigned to the cotton bolls for training and testing the subsequent instance segmentation model.

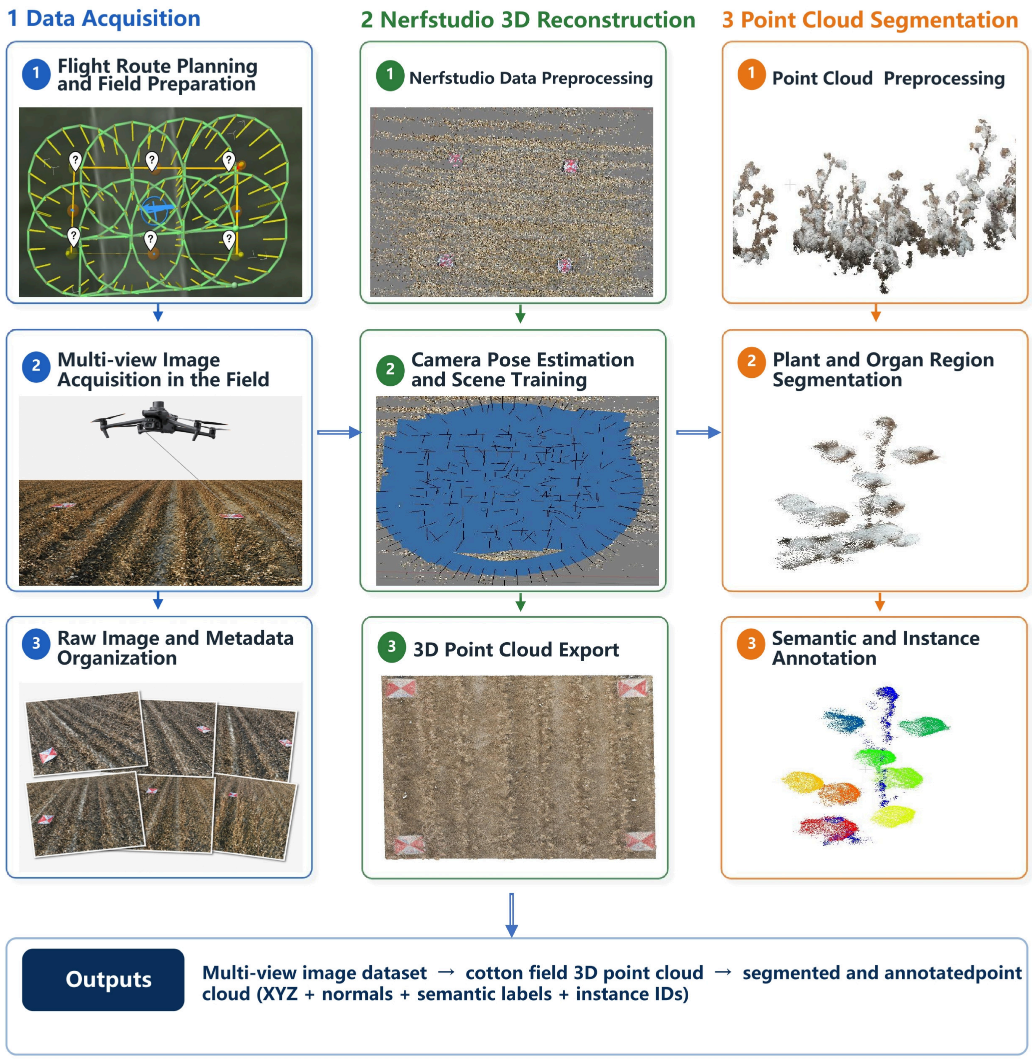


**Figure 1.** Cotton point cloud data segmentation process.

### 2.2. Data Preprocessing and Instance Annotation

After the raw 3D field point clouds had been generated, they were preprocessed in CloudCompare (version 2.14 alpha, Windows 64-bit). Ground points, weeds, and isolated noise points were removed, and the dense field point clouds were segmented into individual cotton plant point clouds. To construct the dataset for network training, each individual-plant point cloud was annotated at the point level using both semantic and instance labels. For semantic annotation, each point was assigned to one of two classes: cotton boll or stems and leaves (background). For instance annotation, a unique instance

ID was assigned to each cotton boll within the same plant. Examples of individual cotton plant point clouds acquired at maturity and after the first mechanical harvesting operation, together with their semantic and instance annotations, are presented in Figure 2.

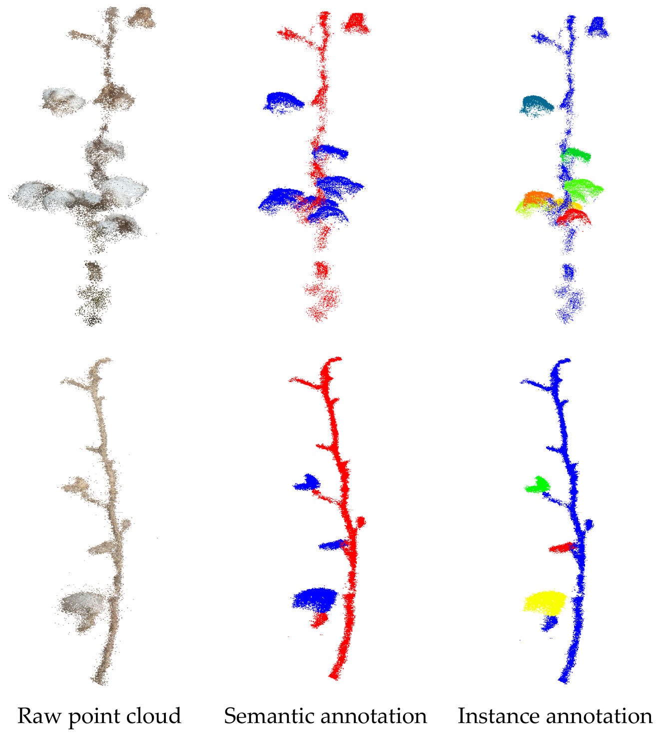


**Figure 2.** Annotation examples of individual cotton plant point clouds. The first row shows a plant before the first harvest, and the second row shows a plant after the first harvest.

*2.3. Class Distribution in the Cotton Boll Instance Segmentation Dataset*

To increase variation in plant morphology, cotton boll number, and occlusion severity, multi-view images were acquired from mature cotton plants before and after the first mechanical harvesting operation. Before harvesting, the plants generally had denser branches and leaves, more cotton bolls, and more extensive occlusion and contact among canopy organs. After harvesting, the removal of cotton bolls altered their number, spatial distribution, and local occlusion relationships. Combining data from these two acquisition periods increased the structural diversity of the dataset and allowed the model to be trained and evaluated using plants with different canopy conditions, boll distributions, and degrees of occlusion.

Before the first mechanical harvesting operation, the mature cotton plants had dense branches and leaves, with extensive contact and occlusion among cotton bolls. Previous studies have shown that organ self-occlusion in dense and structurally complex canopies can impede the acquisition of complete 3D data, leaving gaps in the reconstructed point clouds even when multi-view imaging is employed [37]. In the present study, the limited visibility of some cotton bolls in the multi-view images resulted in missing local geometry or structural discontinuities in the reconstructed point clouds. Consequently, relatively few preharvest plants satisfied the reconstruction and annotation criteria, yielding 44 valid individual-plant point clouds. After the first mechanical harvesting operation, most cotton bolls had been removed, reducing occlusion and contact among the remaining plant organs. A larger proportion of the plant surfaces was therefore visible in the multi-view images, resulting in more complete point-cloud reconstruction and yielding 182 valid individual-plant point clouds. In total, 226 cotton plant point clouds were obtained across the two acquisition periods.

The class distribution and spatial characteristics of the individual-plant point clouds were subsequently quantified. Table 1 summarizes the training, validation, and test sets constructed by combining the data from both acquisition periods. The dataset contained 720 cotton boll instances, with 164, 31, and 31 cotton plants assigned to the training, validation, and test sets, respectively. Cotton boll points accounted for 39.89%, 42.56%, and 43.71% of all points in these three subsets, respectively, and for 40.65% of the complete dataset. The proportion of cotton boll foreground points was therefore lower than that of stem-and-leaf background points, indicating a moderate foreground–background class imbalance.

The greater proportion of stem-and-leaf background points, together with semantic ambiguity at the interfaces between cotton bolls and branches, may impede the extraction of local cotton boll features [38]. For cotton boll instances represented by relatively few or spatially scattered points, or those partially affected by occlusion, geometric features may become progressively attenuated during successive downsampling stages, potentially resulting in missed detections or incomplete instance boundaries [38,39]. Furthermore, variations in plant growth, harvesting stage, and observation viewpoint can cause neighboring cotton bolls to touch or occlude one another, producing locally continuous or difficult-to-distinguish point distributions. Clustering methods that separate instances primarily according to spatial distance or local point density are sensitive to the selected distance thresholds and neighborhood scales. Consequently, multiple cotton bolls within contact regions may be incorrectly merged into a single instance [40,41]. Reducing interference from the complex stem-and-leaf background while improving the discrimination of adjacent cotton bolls therefore constitutes a central challenge addressed in this study.

**Table 1.** Analysis of the cotton plant point cloud data.

| Data Split | Number of Plants | Total Points | Cotton Boll Points | Proportion of Cotton Boll Points | Total Number of Cotton Bolls |
|---|---|---|---|---|---|
| Train | 164 | 6,299,210 | 2,512,722 | 39.89% | 534 |
| Val | 31 | 1,205,789 | 513,126 | 42.56% | 97 |
| Test | 31 | 821,157 | 358,888 | 43.71% | 89 |
| Total | 226 | 8,326,156 | 3,384,736 | 40.65% | 720 |

*2.4. Overall Architecture of RQ-PointNeXt*

PointNeXt was adopted as the backbone for pointwise feature extraction, and the overall architecture of the proposed RQ-PointNeXt network is illustrated in Figure 3. Developed from PointNet++, PointNeXt retains the hierarchical encoder–decoder architecture while introducing improved training and architectural strategies for point-cloud learning [29,42]. The encoder constructs local neighborhoods using farthest point sampling and ball query and extracts multiscale features through local aggregation blocks. The decoder progressively upsamples the deep features through feature propagation and combines them with the corresponding high-resolution features from the encoder, producing pointwise representations for subsequent prediction. Although the standard PointNeXt semantic-segmentation configuration can capture local geometric information and predict a semantic class for each point, it does not assign instance identities and therefore cannot directly separate individual cotton bolls. In the present dataset, stem-and-leaf background points outnumber cotton boll foreground points. This class imbalance may bias feature learning toward the background and increase the risk of missed cotton boll detections or incomplete boundaries. Moreover, generating instances from semantic predictions using DBSCAN or cKDTree-based neighborhood grouping introduces sensitivity to inter-boll distance, point density, and local connectivity. Such procedures may merge adjacent cotton bolls or fragment a single boll when their point distributions are spatially connected or locally discontinuous.

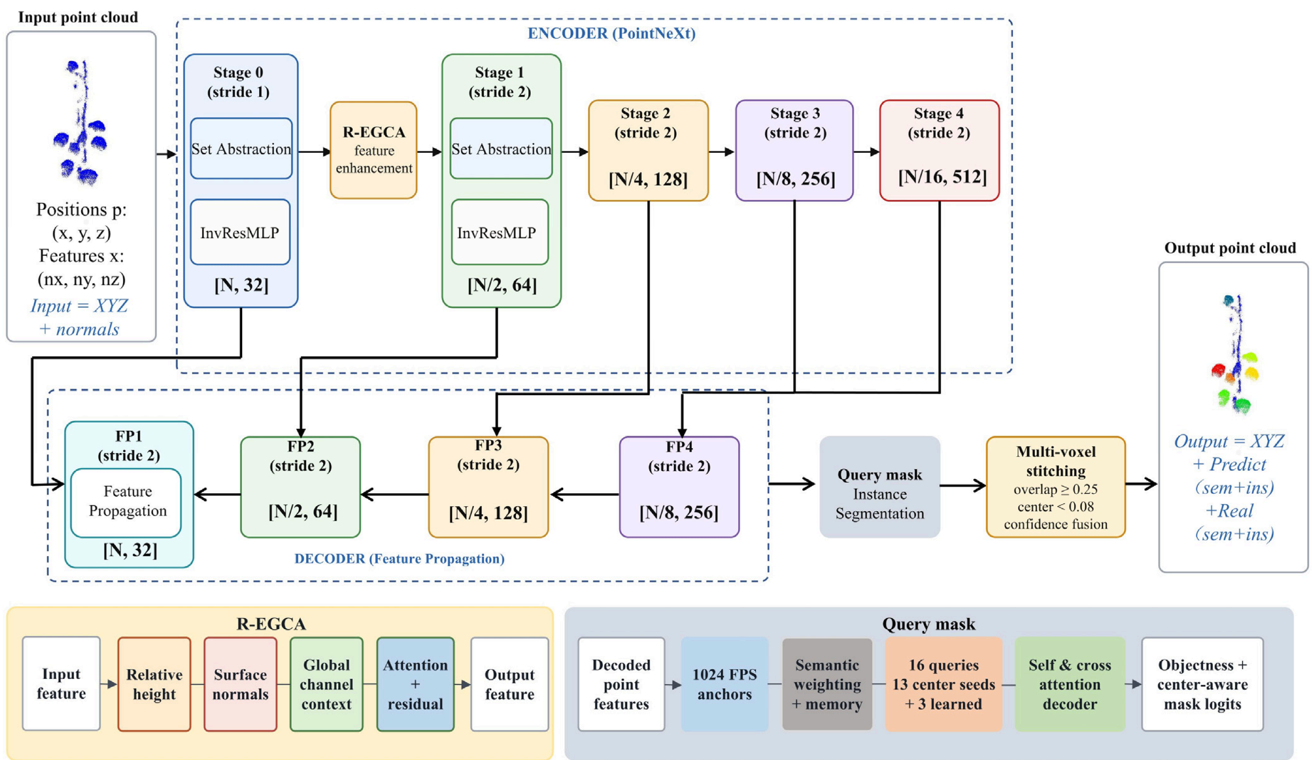


**Figure 3.** RQ-PointNeXt network architecture.

Building on the PointNeXt backbone, RQ-PointNeXt introduces several components specifically designed for cotton boll instance segmentation. First, the R-EGCA module is inserted after the initial high-resolution encoder stage to strengthen local feature representation. It combines global channel context with normalized pointwise relative height and surface-normal information to generate channel-wise attention weights and integrates the reweighted features through a learnable residual connection. Second, a Query–mask decoding branch is constructed to represent candidate cotton bolls using instance queries and to predict both query-level objectness scores and corresponding pointwise mask logits. This formulation enables the network to generate cotton boll instances directly rather than deriving them from an independent geometric clustering procedure. Third, a center-aware mask prior introduces the distances between points and predicted query centers as a differentiable spatial bias on the mask logits, suppressing activations far from the corresponding centers and constraining excessive mask expansion around adjacent cotton bolls. During training, Hungarian matching establishes the correspondence between predicted queries and ground-truth instances, after which query classification, mask BCE, and Dice losses provide direct instance-level supervision. These modifications extend PointNeXt from a pointwise feature backbone into an end-to-end instance segmentation network for separating individual cotton bolls, as shown in Figure 3.

### *2.5. R-EGCA-Based Feature Enhancement Module*

After the surface normals were provided as input and the relative height of each point was calculated, the R-EGCA module was inserted after the initial high-resolution stage of the PointNeXt encoder. The module adaptively reweights the point-level features produced at this stage. Let the resulting point-level feature map be $F \in \mathrm{R}^{B \times C \times N}$, where $B$, $C$, and $N$ represent the batch size, number of feature channels, and number of points, respectively. R-EGCA constructs the feature weights by combining three components: the global channel response, relative-height response, and surface-normal response. First, global max pooling is applied to the point-level feature $F$ along the point dimension, and the pooled feature is passed through the channel branch to obtain the global channel response, as expressed in Equation (1). The normalized relative-height feature $Z^r$ and surface normals $N_m$ are then processed by their respective branches to produce the relative-height and surface-normal responses, as expressed in Equations (2) and (3). The three responses are summed and passed through a Sigmoid function to generate the fused feature weights, as expressed in Equation (4). The resulting weights are then applied to the original point-level features through a scaled residual connection, producing the enhanced features described in Equation (5). These enhanced features are subsequently propagated through the remaining PointNeXt encoder stages and the decoder, providing the feature representation used by the Query–mask instance segmentation branch.

$$A_c = MLP_c(GMP(F)) \tag{1}$$

$$A_z = MLP_z(Z^r) \tag{2}$$

$$A_n = MLP_n(N_m),\ N_m = [n_x, n_y, n_z] \tag{3}$$

$$A = \sigma(A_c + A_z + A_n) \tag{4}$$

$$F' = F + \alpha \cdot (F \odot A) \tag{5}$$

Here, $GMP(F)$ represents the global max-pooling operation applied to the point-level feature $F$; $MLP_c$ is the multilayer perceptron used in the channel branch; and $A_c$ is the resulting global channel response. $Z^r$ is the normalized relative-height feature of the input point cloud; $N_m = [n_x, n_y, n_z]$ contains the surface normals of the input point cloud; $MLP_z$ and $MLP_n$ are the multilayer perceptrons used in the relative-height and surface-normal branches, respectively. The corresponding outputs, $A_z$ and $A_n$ are the relative-height and surface-normal responses, respectively. $\sigma()$ is the Sigmoid activation function, and $A$ represents the fused feature weights. The symbol $\odot$ indicates elementwise multiplication, $\alpha$ is a learnable residual scaling coefficient; and $F'$ is the enhanced point-level feature.

### *2.6. Query–Mask Branch for Cotton Boll Instance Segmentation*

After the PointNeXt decoder restores the point-level resolution, a linear projection maps the decoded features to a common hidden dimension. To reduce the computational cost of interactions between instance queries and all input points, the model samples 1024 superpoint anchors using farthest point sampling and assigns each point to its nearest anchor. The point-level features assigned to each superpoint are then aggregated by weighted averaging, yielding the superpoint feature described in Equation (6). Positional encodings derived from the superpoint coordinates are added to the aggregated features to construct the superpoint memory. Within this aggregation path, the detached semantic foreground probability provides an auxiliary weight, preventing gradients from the instance branch from propagating back to the semantic head through this operation. The semantic prediction therefore assists feature aggregation but does not directly generate instance IDs. With 20,000 input points, the use of 1024 superpoint anchors reduces the cross-attention input to 1024 tokens, corresponding to a nominal average of approximately 20 points per anchor. This setting reduces the computational cost while preserving spatial coverage through farthest point sampling.

A fixed set of $K$ instance-query slots is subsequently introduced, as expressed in Equation (7). The dataset contains 720 annotated cotton bolls from 226 plants, corresponding to an average of 3.19 bolls per plant, while the maximum number of bolls in a single-plant point cloud is 13. Accordingly, $K$ is set to 16, comprising 13 center-seeded query slots and 3 learned query slots. The three learned queries provide additional candidates for instances not adequately represented by the center seeds. Larger $K$ values produced more redundant local candidates and increased duplicate predictions during subcloud stitching. The queries first exchange information through self-attention and then interact with the superpoint memory through a multilayer cross-attention decoder, producing the updated query features described in Equation (8). These features are passed to the classification and mask prediction heads. The classification head generates query-level logits indicating whether each query represents a valid cotton boll instance, as expressed in Equation (9). The mask head computes the base pointwise mask logits from the inner product between the projected point and query features, as given in Equation (10). These logits are subsequently adjusted by the model-internal semantic guidance and center-aware mask prior.

During training, pointwise binary masks are constructed from the ground-truth cotton boll instance IDs. Hungarian matching establishes a one-to-one correspondence between valid predicted queries and ground-truth instances using a matching cost composed of classification, *BCE*, and *Dice* terms, as expressed in Equation (11). During inference, queries with objectness scores below 0.28 are discarded, and redundant predictions are suppressed using mask-IoU-based Query NMS with an IoU threshold of 0.65. For the *i*-th point and *k*-th retained query, the query objectness probability is multiplied by the corresponding mask probability to obtain the joint score described in Equation (12). Each point is assigned to the query with the highest joint score if the score exceeds 0.35, as expressed in Equation (13); otherwise, it remains unassigned. Local instances containing fewer than 50 points are removed before stitching to reduce fragmented predictions. These inference thresholds were selected based on validation-set behavior to balance missed instances, duplicate predictions, and small prediction fragments, and were then kept fixed for all test evaluations. For full-cloud multi-voxel inference, predictions from each subcloud are mapped back to the original point indices. Let $I_l$ and $I_g$ denote the point sets of local instance $l$ and global instance $g$, respectively. Their point-overlap ratio is defined in Equation (14).

The local instance is matched to the global instance with the highest overlap ratio when the ratio is not less than 0.25. Otherwise, it is matched to the nearest global instance when the center distance is less than 0.08 in the normalized coordinate system. The local center is calculated as the mean of its pointwise center votes. For points predicted in multiple subclouds, only the prediction with the higher joint score defined in Equation (12) is retained. A matched local prediction must contribute at least 40 newly accepted, higher-confidence points to update an existing global ID, whereas an unmatched prediction requires at least 180 points to create a new ID. The lower threshold permits incremental updates to an existing instance, while the higher threshold reduces new IDs generated from small fragments. After stitching, instances containing fewer than 100 retained points are removed. The final instance confidence is calculated as the mean joint score of its retained points. This stitching procedure reconciles repeated subcloud predictions without using DBSCAN or cKDTree clustering.

$$\mathrm{M_s} = (\Sigma_{\mathrm{i}\in\Omega_\mathrm{s}} \mathrm{w_i F_{p,i}}) / (\Sigma_{\mathrm{i}\in\Omega_\mathrm{s}} \mathrm{w_i} + \varepsilon) \tag{6}$$

$$Q = q_{k}{}_{k=1}^{K}, q_k \in R^D \tag{7}$$

$$Q' = Decoder(Q, M) \tag{8}$$

$$\hat{c}_k = MLP_{cls}(q'_k) \tag{9}$$

$$\hat{m}_{i,k} = F_{m,i}^T Q_{m,k} \tag{10}$$

$$C_{k,j} = \lambda_{cls} C_{cls} + \lambda_{bce} C_{bce} + \lambda_{dice} C_{dice} \tag{11}$$

$$s_{i,k} = p_k \cdot \sigma(\hat{m}_{i,k}) \tag{12}$$

$$\hat{y}_i = argmax_k s_{i,k} \tag{13}$$

$$R_{ov(l,g)} = \frac{|I_l \cap I_g|}{|I_l|} \tag{14}$$

The terms $M_s$ and $\Omega_s$ represent the aggregated feature of the $s$-th superpoint and the set of points assigned to it, respectively; $F_{p,i}$ and $w_i$ are the projected feature and aggregation weight of the $i$-th point, respectively, while $\varepsilon$ is a small positive constant added to the denominator for numerical stability. $q_k$ is the k-th instance query, and $K$ is the total number of query slots. M is the superpoint memory, and $Q'$ contains the updated query features. $\hat{c}_k$ represents the classification logits of the k-th query, whereas $\hat{m}_{i,k}$ is the mask logit indicating the membership of the $i$-th point in the instance represented by that query. The terms $F_{m,i}$ and $Q_{m,k}$ are the mask-projected features of the $i$-th point and the $k$-th query, respectively. $C_{k,j}$ is the matching cost between the $k$-th predicted query and the $j$-th ground-truth instance. The classification, mask *BCE*, and *Dice* components of this cost are $C_{cls}$, $C_{bce}$, and $C_{dice}$, respectively, with corresponding weights $\lambda_{cls}$, $\lambda_{bce}$, and $\lambda_{dice}$. The joint assignment score between the $i$-th point and the $k$-th query is $s_{i,k}$. In this score, $p_k$ is the probability that the $k$-th query represents a cotton boll instance, while $\sigma(\hat{m}_{i,k})$ is the mask probability that the $i$-th point belongs to the instance represented by that query. The term $\hat{y}_i$ is the predicted instance ID assigned to the $i$-th point. $R_{ov(l,g)}$ denotes the point-overlap ratio between local instance $l$ and global instance $g$, while $I_l$ and $I_g$ denote their respective sets of original point indices. The numerator represents the number of points shared by the two instances, and the denominator represents the total number of points in local instance $l$.

### 2.7. Experiments

#### 2.7.1. Experimental Environment and Implementation Details

The experiments were implemented in PyTorch 2.6. The network was optimized using AdamW with a weight decay of 0.01 and an initial learning rate of 0.0007. A linear learning-rate warm-up was applied during the first 10 epochs, followed by a multistep decay schedule. The learning rate was reduced by a factor of 0.3 at epochs 70, 120, and 175. Training was conducted for 200 epochs with a micro-batch size of 4. Gradients were accumulated over four consecutive iterations, resulting in an effective batch size of 16.

To align the physical scales of the point clouds acquired during the two periods, the raw coordinates of the maturity-stage point clouds were multiplied offline by the physical scale factor $W_{scale}$. The scaled point clouds were subsequently saved as preprocessed dataset files. This operation was performed only once, and the physical scaling was not repeated during data loading. Each individual-plant point cloud from both acquisition periods was then centered independently by subtracting its centroid. The centered coordinates were subsequently multiplied by the fixed coordinate scale factor $S_{coord}$ to obtain a consistent input scale for the network. The complete coordinate transformation is expressed as follows:

$$p_i^{in} = S_{coord}\left(p_i^{phys} - \frac{1}{N}\sum_{j=1}^{N} p_j^{phys}\right) \tag{15}$$

For the maturity-stage point clouds acquired before the first mechanical harvesting operation, the offline physical scaling is defined by $p_i^{phys} = W_{scale} p_i^{raw}$, whereas the post-harvest point clouds were already represented at the target physical scale and therefore required no additional offline scaling. This preprocessing reduces variation in the network inputs arising from differences in absolute point-cloud position and source scale while preserving the relative spatial arrangement of organs within each plant.

During training, each individual-plant point cloud was sampled to 20,000 points. Foreground–background balancing and instance-aware sampling were applied to preserve small or sparsely represented cotton boll instances. Sampling with replacement was used when the number of available points in a sampling pool was smaller than its assigned quota. During validation and testing, no upper limit was imposed on the number of input points. Each original point cloud was processed through multiple voxel-based subcloud passes, after which the subcloud predictions were mapped to the original point indices and combined before the evaluation metrics were calculated. This strategy limited the memory required for each inference pass while retaining all points for full-cloud evaluation.

#### 2.7.2. Evaluation Metrics

Model performance was evaluated at two levels: pointwise semantic segmentation and cotton boll instance segmentation. The semantic segmentation metrics quantify the classification of cotton boll foreground points and stem-and-leaf background points, whereas the instance segmentation metrics assess the detection and delineation of individual cotton bolls. Because cotton boll instance segmentation is the primary task of this study, the instance-level metrics constitute the principal evaluation criteria, while the semantic segmentation metrics provide complementary results.

For semantic segmentation, overall accuracy (OA) is the proportion of correctly classified points among all evaluated points, as expressed in Equation (16). Mean class accuracy (mAcc) is the arithmetic mean of the recall values calculated separately for each semantic class, as expressed in Equation (17). Mean intersection over union (mIoU) averages the intersection-over-union value across the semantic classes and quantifies the agreement between the predicted and ground-truth regions, as expressed in Equation (18).

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{16}$$

$$\mathrm{mAcc} = \frac{1}{\mathrm{k}} \sum_{\mathrm{i}=1}^{\mathrm{k}} \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FN_i}} \tag{17}$$

$$\mathrm{mIoU} = \frac{1}{\mathrm{k}} \sum\nolimits_{\mathrm{i}=1}^{\mathrm{k}} \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FP_i} + \mathrm{FN_i}} \tag{18}$$

In the above equations, TP and TN are the numbers of correctly classified cotton boll and background points, respectively. FP is the number of background points incorrectly classified as cotton boll, whereas FN is the number of cotton boll points incorrectly classified as background. The total number of semantic classes is represented by k.

Cotton boll instance segmentation was the primary evaluation task, and its performance was quantified using $\mathrm{AP}_{25}$, $\mathrm{AP}_{50}$, $\mathrm{AP}_{75}$, and $mAP_{25/50/75}$. $\mathrm{AP}_{25}$, $\mathrm{AP}_{50}$, and $\mathrm{AP}_{75}$ are the average precision values calculated at instance-level IoU thresholds of 0.25, 0.50, and 0.75, respectively. The metric $mAP_{25/50/75}$ is their arithmetic mean and summarizes instance segmentation performance across these three matching thresholds. $\mathrm{AP}_{25}$ applies a relatively lenient overlap criterion and primarily indicates whether the model can coarsely localize and segment individual cotton bolls. $\mathrm{AP}_{50}$ provides an intermediate evaluation criterion, whereas $\mathrm{AP}_{75}$ requires substantially greater agreement between the predicted and ground-truth instance masks and therefore provides a more stringent assessment of pointwise assignment and boundary delineation, particularly for adjacent or touching cotton bolls.

For instance segmentation evaluation, the point-set intersection over union (IoU) is calculated for each pair of predicted and ground-truth instances. Let $P_a$ and $G_b$ represent the point sets of the predicted cotton boll instance indexed by a and the ground-truth cotton boll instance indexed by b, respectively. The instance IoU between these two point sets is defined in Equation (19). A higher instance IoU indicates greater pointwise overlap and stronger spatial agreement between the predicted and ground-truth instances. At a given IoU threshold $\tau$, the predicted instances within each plant point cloud are first ranked in descending order of confidence. Each prediction is then greedily matched to the unmatched ground-truth instance with which it has the highest IoU. A prediction is recorded as a true positive if this maximum IoU is greater than or equal to $\tau$, otherwise, it is recorded as a false positive. This decision rule is given in Equation (20).

For each IoU threshold $\tau$, the true-positive and false-positive records obtained from all test point clouds are pooled and ranked globally by confidence. Cumulative TP and FP values are then used to calculate precision and recall at each position in the ranked list, as expressed in Equations (21) and (22), respectively. To construct the precision envelope, the precision at each recall level is replaced by the maximum precision observed at that level or any higher recall level. AP is subsequently obtained by integrating the resulting precision–recall curve over recall, using all recall-change points, as expressed in Equation (23). This dataset-level calculation accounts for both instance-mask matching accuracy and confidence ranking. Finally, $mAP_{25/50/75}$ is used as the overall cotton boll instance segmentation metric, as expressed in Equation (24).

$$\mathrm{IoU}(\mathrm{P_a}, \mathrm{G_b}) = \frac{|\mathrm{P_a} \cap \mathrm{G_b}|}{|\mathrm{P_a} \cup \mathrm{G_b}|} \tag{19}$$

$$\mathrm{TP}_\mathrm{a}^\tau = \begin{cases} 1, \max\limits_\mathrm{b} \mathrm{IoU}(\mathrm{P_a}, \mathrm{G_b}) \geq \tau \\ 0, \max\limits_\mathrm{b} \mathrm{IoU}(\mathrm{P_a}, \mathrm{G_b}) < \tau \end{cases} \tag{20}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{21}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{N_{gt}}} \tag{22}$$

$$\mathrm{AP}_\tau = \sum\nolimits_\mathrm{r} (\mathrm{R_r} - \mathrm{R_{r-1}}) \mathrm{P_{interp}}(\mathrm{R_r}) \tag{23}$$

$$\mathrm{mAP}_{25/50/75} = \frac{\mathrm{AP}_{25} + \mathrm{AP}_{50} + \mathrm{AP}_{75}}{3} \tag{24}$$

Here, $P_a$ denotes the set of points contained in the a-th predicted instance, and $G_b$ denotes the set of points contained in the b-th ground-truth instance; $|P_a \cap G_b|$ denotes the number of overlapping points between the predicted and ground-truth instances, and $|P_a \cup G_b|$ denotes the number of points in their union. $\tau$ denotes the IoU threshold for instance matching. In this study, $\tau$ = 0.25, $\tau$ = 0.50, and $\tau$ = 0.75 were set to calculate $\mathrm{AP}_{25}$, $\mathrm{AP}_{50}$, and $\mathrm{AP}_{75}$, respectively. TP denotes the cumulative number of true-positive predictions before the current confidence cutoff position, FP denotes the cumulative number of false-positive predictions, and $N_{gt}$ denotes the total number of ground-truth cotton boll instances in the test set. $R_r$ denotes the r-th recall position, and $P_{interp}(R_r)$ denotes the maximum Precision at and after that recall position. $\mathrm{AP}_{25}$, $\mathrm{AP}_{50}$, and $\mathrm{AP}_{75}$ denote the average precision values obtained at IoU thresholds of 0.25, 0.50, and 0.75, respectively. This study ultimately reports $\mathrm{AP}_{25}$, $\mathrm{AP}_{50}$, $\mathrm{AP}_{75}$, and $\mathrm{mAP}_{25/50/75}$, where $\mathrm{mAP}_{25/50/75}$ serves as the overall performance metric for instance segmentation, while $\mathrm{AP}_{50}$ and $\mathrm{AP}_{75}$ are used to further analyze cotton boll instance segmentation performance under moderate and relatively strict matching conditions.

## 3. Results

### *3.1. Comparison with Representative Methods*

As shown in Table 2, RQ-PointNeXt achieved an $\mathrm{mAP}_{25/50/75}$ of 0.5240 on the test set, with $\mathrm{AP}_{25}$, $\mathrm{AP}_{50}$, and $\mathrm{AP}_{75}$ values of 0.7359, 0.5585, and 0.2777, respectively. The updated PointNet++ + DBSCAN results were 0.2300, 0.3022, 0.2759, and 0.1118 for $\mathrm{mAP}_{25/50/75}$,

$AP_{25}$, $AP_{50}$, and $AP_{75}$, respectively. Accordingly, RQ-PointNeXt achieved absolute improvements of 0.2940, 0.4337, 0.2826, and 0.1659 over PointNet++ + DBSCAN for these four metrics. Among all competing methods, Mask3D achieved the highest $mAP_{25/50/75}$, $AP_{50}$, and $AP_{75}$ values of 0.4052, 0.4394, and 0.2119, respectively, whereas Point Transformer V3 obtained the highest $AP_{25}$ of 0.7091. Compared with these metric-specific best-performing baselines, RQ-PointNeXt improved $mAP_{25/50/75}$, $AP_{25}$, $AP_{50}$, and $AP_{75}$ by 0.1188, 0.0268, 0.1191, and 0.0658, respectively. Overall, RQ-PointNeXt outperformed all comparison methods across the three IoU thresholds, demonstrating particularly strong performance in intermediate- and high-precision instance mask prediction.

**Table 2.** Comparison of Cotton Boll Segmentation Performance.

| Model | OA | mAcc | mIoU | IoU_ Cotton Boll | IoU_ Background | $mAP_{25/50/75}$ | $AP_{25}$ | $AP_{50}$ | $AP_{75}$ |
|---|---|---|---|---|---|---|---|---|---|
| PointNet++ + DBSCAN | 0.8453 | 0.8386 | 0.7269 | 0.6893 | 0.7645 | 0.2300 | 0.3022 | 0.2759 | 0.1118 |
| PointNeXt + DBSCAN | 0.7523 | 0.7261 | 0.5786 | 0.4774 | 0.6798 | 0.1459 | 0.2325 | 0.1311 | 0.0740 |
| Point Transformer V3 | 0.8930 | 0.8880 | 0.8030 | 0.7760 | 0.8301 | 0.4047 | 0.7091 | 0.3855 | 0.1195 |
| DGCNN + DBSCAN | 0.7620 | 0.7468 | 0.6036 | 0.5349 | 0.6722 | 0.2899 | 0.5305 | 0.2951 | 0.0442 |
| PointGroup | 0.6415 | 0.6076 | 0.4357 | 0.2920 | 0.5793 | 0.1590 | 0.2263 | 0.1716 | 0.0791 |
| Mask3D | 0.6853 | 0.6514 | 0.4844 | 0.3465 | 0.6223 | 0.4052 | 0.5642 | 0.4394 | 0.2119 |
| SPFormer | 0.7487 | 0.7287 | 0.5815 | 0.4977 | 0.6654 | 0.3589 | 0.5025 | 0.3778 | 0.1963 |
| RQ-PointNeXt | 0.8942 | 0.8980 | 0.8076 | 0.7931 | 0.8221 | 0.5240 | 0.7359 | 0.5585 | 0.2777 |

RQ-PointNeXt predicts cotton boll instances and their pointwise masks through the Query–mask branch. During inference, query objectness scores and pointwise mask probabilities determine local point-to-instance assignment, while predictions from multiple voxel-based subcloud passes are stitched into global instance IDs. Neither DBSCAN nor cKDTree-based clustering is used for local instance generation, maintaining consistency between the query-based training objective and inference procedure. For downstream tasks focused on cotton boll counting or spatial-distribution analysis, $AP_{25}$ and $AP_{50}$ provide practical measures of whether individual bolls are detected and separated, even when local boundary alignment is imperfect. $AP_{75}$ applies a substantially stricter overlap criterion and therefore provides a more demanding assessment of pointwise assignment and boundary delineation under challenging conditions such as occlusion, contact between adjacent bolls, and incomplete point-cloud reconstruction.

Although RQ-PointNeXt achieved a substantially higher $AP_{75}$ than the comparison methods, its $AP_{75}$ remained below its $AP_{25}$ and $AP_{50}$, as expected under the more stringent IoU criterion. The gap indicates that many predictions achieved sufficient overlap for coarse or intermediate instance matching but did not satisfy the higher pointwise agreement required at an IoU threshold of 0.75. Occlusion by branches and leaves, missing or sparsely reconstructed points, and contact between adjacent cotton bolls can make instance boundaries difficult to distinguish. Under these conditions, even relatively small point-assignment errors may prevent a predicted mask from reaching the $AP_{75}$ matching threshold. Because AP also depends on the confidence ordering of predicted instances, ranking errors may further affect the resulting precision–recall curve. In addition, the test set contained 31 plants; consequently, $AP_{75}$ may be sensitive to a small number of particularly challenging samples. Future work will focus on improving feature representation for partially observed cotton bolls and refining point assignment along boundaries between adjacent instances.

### 3.2. Ablation Experiments

Ablation experiments were conducted to assess the contribution of each component to cotton boll instance segmentation. The baseline combined the PointNeXt backbone with the Query–mask decoder, after which R-EGCA, semantic guidance, center-seeded queries, and the center-aware mask prior were introduced sequentially. The results are summarized in Table 3. The Query–mask decoder represents candidate instances using queries and directly predicts their pointwise masks, without requiring DBSCAN or cKDTree-based clustering for local instance generation. R-EGCA integrates global channel context with relative-height and surface-normal information to enhance shallow geometric features. Semantic guidance incorporates cotton boll foreground responses into the mask logits to reduce the assignment of stem-and-leaf background points to candidate instances.

Adding center-seeded queries increased $mAP_{25/50/75}$ from 0.4652 to 0.5059, corresponding to an absolute gain of 0.0407. $AP_{25}$ and $AP_{75}$ increased by 0.0506 and 0.0628, respectively. These improvements suggest that center-informed query initialization facilitates candidate-instance localization and increases the proportion of predictions satisfying both lenient and stringent IoU criteria. The subsequent addition of the center-aware mask prior, which adjusts the mask logits according to the distance between each point and its predicted query center, further increased $mAP_{25/50/75}$ to 0.5240. The corresponding $AP_{25}$, $AP_{50}$, and $AP_{75}$ values reached 0.7359, 0.5585, and 0.2777, respectively. Overall, the ablation results suggest that R-EGCA strengthens shallow geometric representation, while center-seeded queries and the center-aware mask prior improve instance localization and constrain mask extent. These components therefore provide complementary contributions to cotton boll instance segmentation.

**Table 3.** Results of the ablation experiments.

| Model | OA | mAcc | mIoU | IoU_ Cotton Boll | IoU_ Background | $mAP_{25/50/75}$ | $AP_{25}$ | $AP_{50}$ | $AP_{75}$ |
|---|---|---|---|---|---|---|---|---|---|
| A0: PointNeXt + DBSCAN | 0.7523 | 0.7261 | 0.5786 | 0.4774 | 0.6798 | 0.1459 | 0.2325 | 0.1311 | 0.0740 |
| A1: A0 + R-EGCA+Query mask | 0.9241 | 0.9245 | 0.8573 | 0.8423 | 0.8722 | 0.4643 | 0.5431 | 0.4888 | 0.3611 |
| A2: A1 + semantic guidance | 0.8842 | 0.8897 | 0.7917 | 0.7789 | 0.8045 | 0.4652 | 0.6580 | 0.5251 | 0.2124 |
| A3: A2 + center-seeded queries | 0.8973 | 0.8974 | 0.8118 | 0.7927 | 0.8310 | 0.5059 | 0.7086 | 0.5340 | 0.2752 |
| A4: A3 + center-aware mask prior (RQ-PointNeXt) | 0.8942 | 0.8980 | 0.8076 | 0.7931 | 0.8221 | 0.5240 | 0.7359 | 0.5585 | 0.2777 |

### 3.3. Qualitative Analysis of Cotton Boll Instance Segmentation

To examine the segmentation of individual cotton bolls in greater detail, the prediction results for representative whole-plant point clouds from the test set were visualized, as shown in Figure 4. Different colors represent different predicted cotton boll instances, whereas background points are not assigned instance IDs. The proposed method identifies most cotton bolls within each plant and assigns distinct instance IDs to different bolls. For spatially separated and lightly occluded cotton bolls, the predicted point sets are relatively complete and their instance boundaries are clearly delineated. In accordance with the network design, the Query–mask branch directly generates these instances from query classification scores and pointwise mask assignment probabilities, without using DBSCAN or cKDTree-based neighborhood grouping as the principal instance-generation procedure. Nevertheless, occasional boundary-assignment errors remain in regions affected by severe occlusion or contact between adjacent cotton bolls. For example, points within a contact region may be incorrectly assigned to the same instance, while some peripheral points may remain unassigned to their corresponding bolls. Such errors have a greater effect on $AP_{75}$ because this metric applies a stricter requirement to the pointwise overlap between predicted and ground-truth instances.

To further evaluate the proposed method under conditions of relatively high point-cloud completeness, experiments were conducted using the public UGA-BSAIL Cotton Plants with Foliage dataset. The point-cloud data are publicly available through Figshare (https://figshare.com/projects/Cotton_plant_with_foliage/258065, accessed on 8 September 2026), while the associated code and documentation are hosted on GitHub (https://github.com/UGA-BSAIL/Cotton_plants_with_foliage, accessed on 8 September 2026). The dataset contains relatively complete cotton plant point clouds, surface-normal attributes, and semantic labels distinguishing cotton bolls from stem-and-leaf structures. For this study, the data were converted into a unified format comprising three-dimensional coordinates, surface normals, and semantic labels. Cotton boll instance annotations were subsequently added, and the resulting data were divided into training, validation, and test sets. Compared with the self-constructed field dataset, the UGA-BSAIL dataset contains more complete cotton boll structures and fewer reconstruction gaps, providing a complementary setting for evaluating the model's ability to segment individual cotton bolls when organ geometry is more completely represented.

The performance of RQ-PointNeXt on the test set of the public dataset after 100 epochs of training is presented in Table 4. Owing to the relatively complete point clouds, low noise levels, and clearly defined cotton boll boundaries, several models effectively learned the semantic and instance features of cotton bolls and achieved high segmentation performance. RQ-PointNeXt obtained AP25, AP50, and AP75 values of 0.9577, 0.9270, and 0.8675, respectively, with an mAP25/50/75 of 0.9174. Although these results demonstrate strong overall performance, RQ-PointNeXt did not show a clear advantage over PointNeXt combined with DBSCAN, Point Transformer V3, or PointGroup. This finding suggests that performance differences among competitive methods become relatively small when point-cloud structures are complete and individual instances are readily distinguishable, making it difficult to fully demonstrate the benefits of RQ-PointNeXt in complex scenes. Therefore, experiments on this public dataset primarily validated the model's ability to segment cotton bolls in high-quality point clouds. Its robustness to occlusion, reconstruction gaps, background interference, and contact between adjacent cotton bolls was further evaluated using the self-constructed field dataset. Representative instance segmentation results are shown in Figure 5.

**Table 4.** Experimental results on the public UGA-BSAIL cotton point cloud dataset.

| Model | OA | mAcc | mIoU | IoU-Cotton Boll | IoU-Background | $mAP_{25/50/75}$ | $AP_{25}$ | $AP_{50}$ | $AP_{75}$ |
|---|---|---|---|---|---|---|---|---|---|
| PointNet++ + DBSCAN | 0.8659 | 0.8055 | 0.6596 | 0.7149 | 0.8962 | 0.5048 | 0.7304 | 0.5219 | 0.2622 |
| PointNeXt + DBSCAN | 0.9877 | 0.9794 | 0.9654 | 0.9466 | 0.9842 | 0.9353 | 0.9583 | 0.9414 | 0.9062 |
| Point Transformer V3 | 0.9876 | 0.9891 | 0.9658 | 0.9476 | 0.9840 | 0.9385 | 0.9580 | 0.9430 | 0.9144 |
| DGCNN + DBSCAN | 0.9432 | 0.9048 | 0.8497 | 0.7693 | 0.9301 | 0.8037 | 0.9101 | 0.8146 | 0.6861 |
| PointGroup | 0.9822 | 0.9844 | 0.9516 | 0.9262 | 0.9770 | 0.9354 | 0.9583 | 0.9424 | 0.9056 |
| Mask3D | 0.9397 | 0.8751 | 0.8337 | 0.7401 | 0.9273 | 0.1830 | 0.3966 | 0.1179 | 0.0345 |
| SPFormer | 0.9610 | 0.9233 | 0.8919 | 0.8323 | 0.9517 | 0.6171 | 0.7865 | 0.6759 | 0.3890 |
| RQ-PointNeXt | 0.9832 | 0.9846 | 0.9541 | 0.9297 | 0.9784 | 0.9174 | 0.9577 | 0.9270 | 0.8675 |

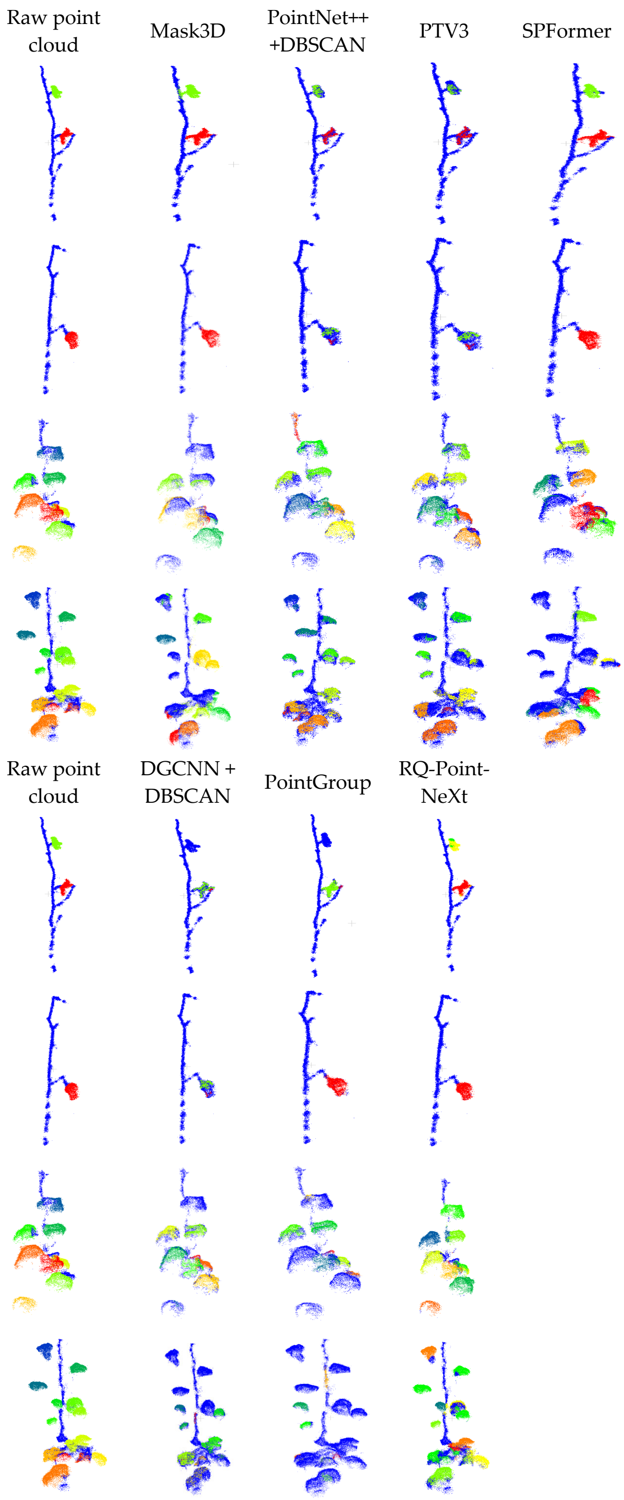


**Figure 4.** Qualitative comparison of instance segmentation results on the self-constructed field dataset.

Figure 4 shows several differences among the methods. PointNet++ and DGCNN combined with DBSCAN produce fragmented or incomplete instances in some sparsely reconstructed boll regions. Mask3D and SPFormer identify several major boll regions but miss or incompletely segment some small or occluded bolls. PointGroup separates many visible bolls but shows local assignment errors where adjacent bolls are spatially connected. RQ-PointNeXt generally produces more coherent masks and fewer small fragments, although incomplete boundaries and occasional merging remain in regions affected by boll contact or sparse reconstruction. These errors are consistent with the lower $AP_{75}$ value relative to $AP_{25}$ and $AP_{50}$.

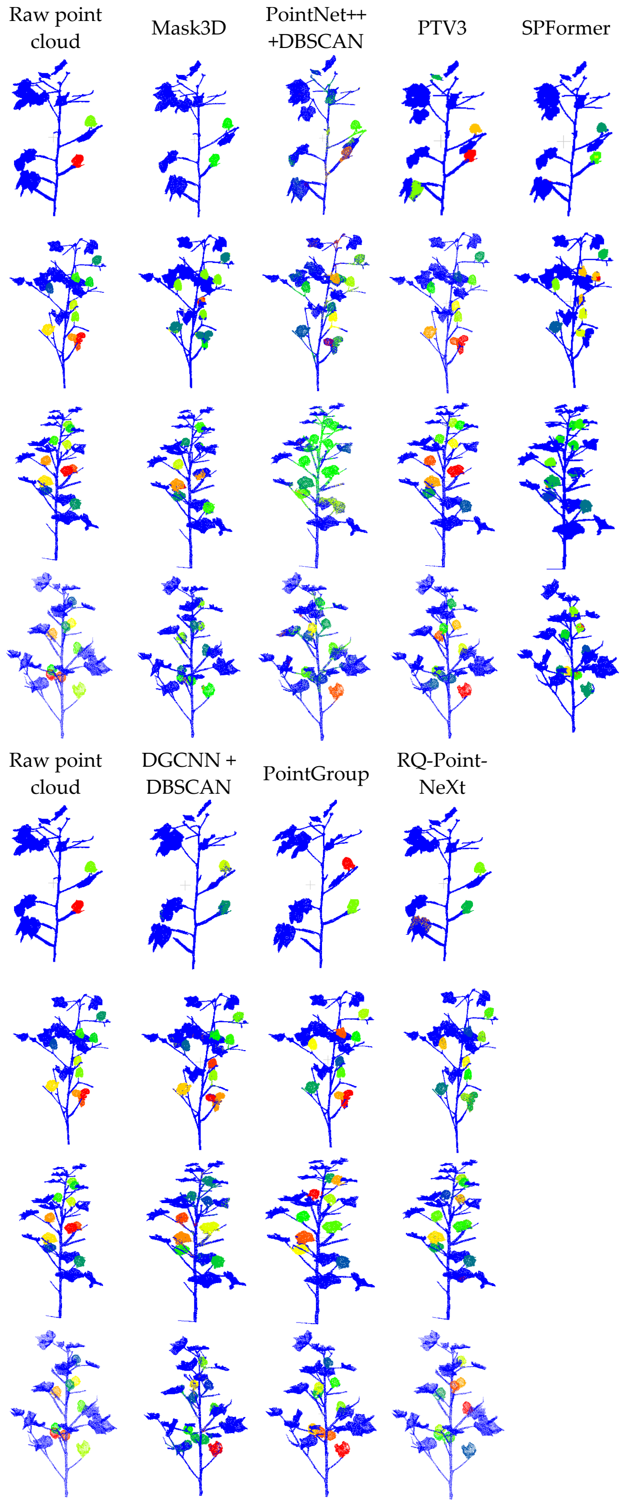


**Figure 5.** Qualitative comparison of instance segmentation results on the UGA-BSAIL dataset. The differences among the methods are less pronounced in Figure 5 because the UGA-BSAIL point clouds contain more complete boll structures and clearer boundaries. PointNeXt combined with DBSCAN, Point Transformer V3, PointGroup, and RQ-PointNeXt separate most visible bolls, which is consistent with their $mAP_{25/50/75}$ values of 0.9353, 0.9385, 0.9354, and 0.9174, respectively. RQ-PointNeXt therefore does not provide a numerical advantage on this dataset. Compared with the self-constructed field dataset, these results indicate that conventional clustering is effective when point-cloud connectivity is complete, whereas the differences among methods become more apparent under occlusion, reconstruction gaps, and contact between adjacent bolls.

### 3.4. Robustness and Failure-Mode Analysis

To assess the robustness of the results to dataset partitioning, five-fold cross-validation was conducted on the 226 individual-plant point clouds. The dataset was divided into five mutually exclusive folds, with each fold used once as the test set. The same model configuration, training settings, and evaluation procedure were used for all folds. The fold-wise results and their mean, standard deviation (SD), and 95% confidence interval (CI) are reported in Table 5.

**Table 5.** Five-fold cross-validation results of RQ-PointNeXt.

| Fold | OA | mAcc | mIoU | IoU-Cotton Boll | IoU-Background | $mAP_{25/50/75}$ | $AP_{25}$ | $AP_{50}$ | $AP_{75}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.8664 | 0.8717 | 0.7591 | 0.7236 | 0.7945 | 0.5036 | 0.7572 | 0.5664 | 0.1873 |
| 2 | 0.8623 | 0.8708 | 0.7562 | 0.7354 | 0.7769 | 0.5283 | 0.7774 | 0.5523 | 0.2551 |
| 3 | 0.8641 | 0.8676 | 0.7599 | 0.7454 | 0.7744 | 0.5100 | 0.7470 | 0.5425 | 0.2404 |
| 4 | 0.8673 | 0.8721 | 0.7640 | 0.7443 | 0.7838 | 0.5252 | 0.7740 | 0.5456 | 0.2560 |
| 5 | 0.8551 | 0.8582 | 0.7369 | 0.6856 | 0.7882 | 0.5476 | 0.7722 | 0.5998 | 0.2708 |
| Mean ± SD | 0.8630 ± 0.0049 | 0.8681 ± 0.0058 | 0.7552 ± 0.0106 | 0.7269 ± 0.0247 | 0.7836 ± 0.0082 | 0.5229 ± 0.0172 | 0.7656 ± 0.0129 | 0.5613 ± 0.0234 | 0.2419 ± 0.0324 |
| 95% CI | 0.8570–0.8691 | 0.8609–0.8753 | 0.7420–0.7684 | 0.6962–0.7575 | 0.7734–0.7938 | 0.5016–0.5443 | 0.7495–0.7816 | 0.5323–0.5904 | 0.2017–0.2821 |

Across the five folds, RQ-PointNeXt achieved mean OA, mAcc, and mIoU values of 0.8630, 0.8681, and 0.7552, respectively. The IoU of cotton boll points varied more across folds than that of background points, with standard deviations of 0.0247 and 0.0082, respectively. For instance segmentation, the mean mAP25/50/75 was 0.5229 ± 0.0172, with a 95% confidence interval of 0.5016–0.5443, while AP25 reached 0.7656 ± 0.0129. The cross-validation mean mAP25/50/75 was close to the value of 0.5240 obtained from the original fixed split, indicating that the overall result was not strongly dependent on a single dataset partition.

A quantitative failure-mode analysis was further conducted using the 31 plants in the original fixed test set. The results are presented in Table 6.

**Table 6.** Quantitative analysis of RQ-PointNeXt failure modes on the original fixed test set.

| Failure Mode | Number | Rate (%) |
|---|---|---|
| Missed instances | 21 | 23.60 |
| False-positive instances | 22 | 24.44 |
| Merged instances | 16 | 17.78 |
| Split instances | 9 | 10.11 |
| Detected instances failing to reach IoU 0.75 | 34 | 50.00 |

Missed and split-instance rates were calculated relative to the 89 ground-truth instances. False-positive and merged-instance rates were calculated relative to the 90 predicted instances. The final row represents the proportion of the 68 instances detected at IoU $\geq$ 0.25 that failed to reach IoU 0.75. The failure categories are not mutually exclusive. RQ-PointNeXt missed 21 ground-truth instances and produced 22 false-positive instances. Among the 68 instances detected at IoU $\geq$ 0.25, 34 did not reach IoU 0.75. Together with the lower cross-validation AP75, this result shows that accurate mask delineation remains more difficult than coarse instance detection. Merged and split predictions occurred mainly for adjacent cotton bolls with weak boundaries and for partially occluded or incompletely reconstructed instances.

## 4. Discussion

### *4.1. Role of NeRF-Based 3D Point Cloud Reconstruction in Field Cotton Point Cloud Acquisition*

At maturity, cotton plants exhibit dense, interwoven branches and leaves, with cotton bolls distributed across multiple canopy layers. Images captured from a single viewpoint provide limited information about the three-dimensional relationships among plant organs because of occlusion and projection ambiguity. Close-range UAV images were therefore acquired from multiple viewpoints, followed by NeRF-based reconstruction and point-cloud extraction for individual cotton plants. By integrating complementary observations, this procedure provides a more complete representation of plant geometry than any individual image and recovers spatial information concerning cotton boll position, surface morphology, and relative arrangement. The exported three-dimensional coordinates, together with subsequently calculated relative-height and surface-normal features, provide geometric cues for distinguishing cotton bolls that overlap in two-dimensional projections and for determining pointwise instance assignments. Nevertheless, reconstruction quality depends strongly on image sharpness, viewpoint coverage, camera-pose accuracy, and scene stability. Variations in illumination, wind-induced plant motion, and severe organ occlusion may produce point-cloud gaps, floating outliers, or distorted organ surfaces. Following point-cloud export, coordinate standardization, outlier removal, surface-normal estimation, and relative-height calculation were performed to reduce the effects of reconstruction artifacts and ensure consistent feature representation. Pointwise semantic labels and cotton boll instance IDs were then added manually. These procedures produced standardized input data for RQ-PointNeXt, in which the relative-height and surface-normal features support R-EGCA feature enhancement, while the semantic and instance annotations provide supervision for the Query–mask branch.

### *4.2. Effect of R-EGCA on Feature Representation of Small Cotton Boll Instances*

In the ablation baseline comprising the original PointNeXt backbone and the Query–mask branch but excluding R-EGCA, instance predictions were less stable, particularly near cotton boll boundaries and in regions where adjacent bolls were in contact. Cotton boll points accounted for a smaller proportion of the whole-plant point clouds than stem-and-leaf background points, while some small or partially occluded boll instances contained relatively few and spatially discontinuous points. Consequently, their local geometric features could be weakened during hierarchical sampling and feature aggregation or confused with nearby background structures. R-EGCA does not directly generate or partition cotton boll instances. Instead, it reweights shallow pointwise features by combining channel responses with relative-elevation and surface-normal information. These additional geometric cues emphasize feature responses associated with cotton boll surface morphology and spatial position during the early stages of feature extraction. The enhanced features are subsequently propagated through the remaining encoder and decoder layers, providing the Query–mask branch with more discriminative pointwise representations for query classification and mask prediction. The improvements in the instance segmentation metrics after introducing R-EGCA support the contribution of shallow geometric feature enhancement to cotton boll instance separation. These findings suggest that, for structurally complex crop point clouds containing small or partially occluded target instances, features extracted solely by a general-purpose backbone may not fully preserve the geometric cues required for instance discrimination. Incorporating relative elevation and surface normals into feature learning can therefore improve the separation of individual cotton bolls.

### 4.3. Role of Query–Mask Instance Segmentation in Separating Individual Cotton Bolls

Cotton boll instance segmentation requires not only the identification of foreground points but also their assignment to individual bolls. The Query–mask branch incorporates this assignment process into the network by representing candidate cotton boll instances with a fixed set of queries and directly predicting the class confidence and pointwise mask of each candidate. Unlike pipelines that generate instances by applying DBSCAN or other geometric grouping methods after semantic segmentation, this design makes the training objective consistent with the query-based output produced during inference and reduces sensitivity to manually specified distance and connectivity thresholds. Moreover, instance classification, mask completeness, and pointwise overlap can be directly optimized through classification, binary cross-entropy, Dice, and center-related constraints. During whole-plant inference, point assignments are determined from the joint confidence of query classification and mask probabilities, after which predictions from overlapping subclouds are consolidated into globally consistent instance IDs. The experimental and visualization results indicate that this strategy can identify and separate most cotton bolls without external clustering as the principal instance-generation procedure. A current limitation of RQ-PointNeXt is its lower performance under the strict AP75 criterion because severe occlusion, incomplete reconstruction, and contact between adjacent bolls can cause inaccurate boundary and point assignments. Under such conditions, neighboring instances may be locally merged, or peripheral boll points may remain unassigned. These errors have a greater effect on $AP_{75}$ because it imposes stricter requirements on pointwise mask overlap. Further improvements should therefore focus on stronger boundary supervision, greater discrimination among instance queries, and targeted augmentation of samples containing severe occlusion and boll contact. Future work will investigate explainable AI methods to determine how geometric features and query attention contribute to instance predictions. The resulting boll counts and spatial distributions can support plot comparison, growth monitoring, and harvest assessment after further field validation.

### 4.4. Comparison with Existing Instance Segmentation Methods

Under the same self-constructed dataset and evaluation protocol, Mask3D achieved an $mAP_{25/50/75}$ of 0.4052 and an $AP_{75}$ of 0.2119, while the corresponding values of RQ-PointNeXt were 0.5240 and 0.2777. Point Transformer V3 obtained the highest baseline $AP_{25}$ of 0.7091, compared with 0.7359 for RQ-PointNeXt. These results correspond to absolute improvements of 0.1188 in $mAP_{25/50/75}$, 0.0268 in $AP_{25}$, and 0.0658 in $AP_{75}$. However, on the more complete UGA-BSAIL dataset, Point Transformer V3, PointGroup, and Point-NeXt combined with DBSCAN achieved $mAP_{25/50/75}$ values of 0.9385, 0.9354, and 0.9353, respectively, compared with 0.9174 for RQ-PointNeXt. The comparison indicates that RQ-PointNeXt is not consistently superior under relatively complete point-cloud conditions. Its main benefit is observed in the self-constructed field data, where occlusion, missing points, and adjacent boll contact increase the difficulty of clustering-based instance separation.

## 5. Conclusions

To address cotton boll identification and separation in point clouds of individual field-grown cotton plants, this study developed RQ-PointNeXt for cotton boll instance segmentation from field-reconstructed point clouds and constructed an annotated dataset containing 226 cotton plants and 720 cotton bolls. The proposed method combines relative elevation and surface-normal information with a center-constrained Query–mask decoder to generate cotton boll instance masks without clustering-based local instance generation. On the self-constructed test set, RQ-PointNeXt achieved an mIoU of 0.8076 and an mAP25/50/75 of 0.5240. The five-fold cross-validation mean mAP25/50/75 was

0.5229 ± 0.0172, close to the fixed-split result. On the more complete UGA-BSAIL point clouds, the model achieved an mAP25/50/75 of 0.9174. These results demonstrate the feasibility of directly generating cotton boll masks using instance queries and provide instance-level data for phenotypic analyses, including boll counting, position determination, and within-plant spatial distribution analysis. The difference between the two datasets also shows that point-cloud completeness substantially affects segmentation accuracy. The main remaining limitation is precise mask delineation under severe occlusion, incomplete reconstruction, and contact between adjacent bolls, as reflected by the lower AP75 on the field dataset. Future work will focus on boundary modeling and validation across additional cultivars, growth stages, and field environments.

**Author Contributions:** Conceptualization, Y.Z., H.N., Y.W., X.M. and X.C.; methodology, H.N., Y.Z. and H.Q.; software, H.N.; validation, H.N., H.Q. and Y.Z.; formal analysis, H.N.; investigation, H.N., H.Q. and W.Z.; resources, Y.Z., Y.W., X.M., X.C. and W.Z.; data curation, H.N., H.Q. and Y.Z.; writing—original draft preparation, H.N.; writing—review and editing, Y.Z., Y.W., X.M., X.C. and H.N.; visualization, H.N. and H.Q.; supervision, Y.Z., Y.W., X.M., X.C. and W.Z.; project administration, Y.Z., Y.W., X.M., X.C. and W.Z.; funding acquisition, Y.Z. and Y.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number 32560419; the Xinjiang Talent Development Fund Project, grant number XJRC-2025-JY-YJ-GX-QNQZ-047; and the Major Science and Technology Special Project of the Xinjiang Uygur Autonomous Region, grant number 2024A02002-3.

**Data Availability Statement:** The self-constructed field cotton point-cloud dataset is available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors gratefully acknowledge the faculty members and colleagues who provided guidance and assistance throughout this research.

**Conflicts of Interest:** Author Wenbin Zhang was employed by the Xinjiang Jinfengyuan Seed Industry Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Yang, Z.-Y.; Xia, W.-K.; Chu, H.-Q.; Su, W.-H.; Wang, R.-F.; Wang, H. A comprehensive review of deep learning applications in cotton industry: From field monitoring to smart processing. *Plants* **2025**, *14*, 1481. [CrossRef] [PubMed]
2. Zafar, M.M.; Chattha, W.S.; Khan, A.I.; Zafar, S.; Subhan, M.; Saleem, H.; Ali, A.; Ijaz, A.; Anwar, Z.; Qiao, F.; et al. Drought and heat stress on cotton genotypes suggested agro-physiological and biochemical features for climate resilience. *Front. Plant Sci.* **2023**, *14*, 1265700. [CrossRef] [PubMed]
3. Chen, X.; Wen, S.; Zhang, L.; Lan, Y.; Ge, Y.; Hu, Y.; Luo, S. A calculation method for cotton phenotypic traits based on unmanned aerial vehicle LiDAR combined with a three-dimensional deep neural network. *Comput. Electron. Agric.* **2025**, *230*, 109857. [CrossRef]
4. Ayankojo, I.T.; Thorp, K.R.; Thompson, A.L. Advances in the application of small unoccupied aircraft systems (sUAS) for high-throughput plant phenotyping. *Remote Sens.* **2023**, *15*, 2623. [CrossRef]
5. Cudjoe, D.K.; Virlet, N.; Castle, M.; Riche, A.B.; Mhada, M.; Waine, T.W.; Mohareb, F.; Hawkesford, M.J. Field phenotyping for African crops: Overview and perspectives. *Front. Plant Sci.* **2023**, *14*, 1219673. [CrossRef] [PubMed]
6. Akhtar, M.S.; Zafar, Z.; Nawaz, R.; Fraz, M.M. Unlocking plant secrets: A systematic review of 3D imaging in plant phenotyping techniques. *Comput. Electron. Agric.* **2024**, *222*, 109033. [CrossRef]
7. Harandi, N.; Vandenberghe, B.; Vankerschaver, J.; Depuydt, S.; Van Messem, A. How to make sense of 3D representations for plant phenotyping: A compendium of processing and analysis techniques. *Plant Methods* **2023**, *19*, 60. [CrossRef] [PubMed]
8. Murphy, K.M.; Ludwig, E.; Gutierrez, J.; Gehan, M.A. Deep learning in image-based plant phenotyping. *Annu. Rev. Plant Biol.* **2024**, *75*, 771–795. [CrossRef] [PubMed]
9. Luo, Z.; Yang, W.; Yuan, Y.; Gou, R.; Li, X. Semantic segmentation of agricultural images: A survey. *Inf. Process. Agric.* **2024**, *11*, 172–186. [CrossRef]

10. Yi, X.; Wang, J.; Wu, P.; Wang, G.; Mo, L.; Lou, X.; Liang, H.; Huang, H.; Lin, E.; Maponde, B.T.; et al. AC-UNet: An improved UNet-based method for stem and leaf segmentation in *Betula luminifera*. *Front. Plant Sci.* **2023**, *14*, 1268098. [CrossRef] [PubMed]
11. Schunck, D.; Magistri, F.; Rosu, R.A.; Cornelißen, A.; Chebrolu, N.; Paulus, S.; Léon, J.; Behnke, S.; Stachniss, C.; Kuhlmann, H.; et al. Pheno4D: A spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis. *PLoS ONE* **2021**, *16*, e0256340. [CrossRef] [PubMed]
12. Zhao, Y.; Liu, D.; Wang, Z.; Yao, J.; Yu, J.; Wang, Z.; Jia, Z. Automated phenotyping of maize from 3D point clouds using an optimized deep learning approach. *Agriculture* **2025**, *15*, 2430. [CrossRef]
13. Omia, E.E.; Park, E.; Semyalo, D.; Joshi, R.; Cho, B.-K. Advancements in 3D field-crop phenotyping using point clouds: A comparative review of sensor technology, target traits, and challenges under controlled and field conditions. *Front. Plant Sci.* **2026**, *17*, 1731852. [CrossRef] [PubMed]
14. Zhou, J.; Tang, Y.; Cui, M.; Zou, W.; Gao, Y.; Wu, Y.; Wu, M.; Jiang, B.; Zhong, Z.; Zou, Y.; et al. A 3D stem diameter measurement method for field maize at jointing stage: Combining RLRSA-PointNet++ and structural feature fitting. *Front. Plant Sci.* **2026**, *16*, 1724096. [CrossRef] [PubMed]
15. Qiu, H.; Meng, X.; Li, Y.; Zhao, Y.; Li, X.; Yin, S.; Niu, H. ELGCot3D: A lightweight 3D cotton point cloud segmentation model based on EdgeConv-Local Attention-GCN and semantic feature enhancement. *Front. Plant Sci.* **2026**, *17*, 1765604. [CrossRef] [PubMed]
16. Liu, F.-Y.; Geng, H.; Shang, L.-Y.; Si, C.-J.; Shen, S.-Q. A cotton organ segmentation method with phenotypic measurements from a point cloud using a transformer. *Plant Methods* **2025**, *21*, 37. [CrossRef] [PubMed]
17. Yao, J.; Wang, W.; Fu, H.; Deng, Z.; Cui, G.; Wang, S.; Wang, D.; She, W.; Cao, X. Automated measurement of field crop phenotypic traits using UAV 3D point clouds and an improved PointNet++. *Front. Plant Sci.* **2025**, *16*, 1654232. [CrossRef] [PubMed]
18. Dong, S.; Fan, X.; Li, X.; Liang, Y.; Zhang, M.; Yao, W.; Yang, X.; Wang, Z. Automatic 3D plant organ instance segmentation method based on PointNeXt and Quickshift++. *Plant Phenomics* **2025**, *7*, 100065. [CrossRef] [PubMed]
19. Turgut, K.; Dutagaci, H.; Rousseau, D. RoseSegNet: An attention-based deep learning architecture for organ segmentation of plants. *Biosyst. Eng.* **2022**, *221*, 138–153. [CrossRef]
20. Ghahremani, M.; Williams, K.; Corke, F.M.K.; Tiddeman, B.; Liu, Y.; Doonan, J.H. Deep segmentation of point clouds of wheat. *Front. Plant Sci.* **2021**, *12*, 608732. [CrossRef] [PubMed]
21. Yang, X.; Xu, C.; Wang, Y.; Feng, R.; Yu, J.; Su, Z.; Miao, T.; Xu, T. PlaneSegNet: A deep learning network with plane attention for plant point cloud segmentation in agricultural environments. *Artif. Intell. Agric.* **2026**, *16*, 284–299. [CrossRef]
22. Zarei, A.; Li, B.; Schnable, J.C.; Lyons, E.; Pauli, D.; Barnard, K.; Benes, B. PlantSegNet: 3D point cloud instance segmentation of nearby plant organs with identical semantics. *Comput. Electron. Agric.* **2024**, *221*, 108922. [CrossRef]
23. Shi, B.; Guo, L.; Yu, L. Accurate LAI estimation of soybean plants in the field using deep learning and clustering algorithms. *Front. Plant Sci.* **2025**, *15*, 1501612. [CrossRef] [PubMed]
24. Marks, E.A.; Bömer, J.; Magistri, F.; Sah, A.; Behley, J.; Stachniss, C. BonnBeetClouds3D: A dataset towards point cloud-based organ-level phenotyping of sugar beet plants under real field conditions. In Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Abu Dhabi, United Arab Emirates, 14–18 October 2024; pp. 1804–1811. [CrossRef]
25. Boogaard, F.P.; van Henten, E.J.; Kootstra, G. Improved point-cloud segmentation for plant phenotyping through class-dependent sampling of training data to battle class imbalance. *Front. Plant Sci.* **2022**, *13*, 838190. [CrossRef] [PubMed]
26. Li, D.; Wei, Y.; Zhu, R. A comparative study on point cloud down-sampling strategies for deep learning-based crop organ segmentation. *Plant Methods* **2023**, *19*, 124. [CrossRef] [PubMed]
27. Jiang, L.; Zhao, H.; Shi, S.; Liu, S.; Fu, C.-W.; Jia, J. PointGroup: Dual-set point grouping for 3D instance segmentation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4866–4875. [CrossRef]
28. Zhao, W.; Yan, Y.; Yang, C.; Ye, J.; Yang, X.; Huang, K. Divide and conquer: 3D point cloud instance segmentation with point-wise binarization. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 562–571. [CrossRef]
29. Qian, G.; Li, Y.; Peng, H.; Mai, J.; Hammoud, H.A.A.K.; Elhoseiny, M.; Ghanem, B. PointNeXt: Revisiting PointNet++ with improved training and scaling strategies. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 23192–23204. [CrossRef]
30. Schult, J.; Engelmann, F.; Hermans, A.; Litany, O.; Tang, S.; Leibe, B. Mask3D: Mask transformer for 3D semantic instance segmentation. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 8216–8223. [CrossRef]
31. Arshad, M.A.; Jubery, T.; Afful, J.; Jignasu, A.; Balu, A.; Ganapathysubramanian, B.; Sarkar, S.; Krishnamurthy, A. Evaluating neural radiance fields (NeRFs) for 3D plant geometry reconstruction in field conditions. *Plant Phenomics* **2024**, *6*, 0235. [CrossRef] [PubMed]

32. Choi, H.-B.; Park, J.-K.; Park, S.H.; Lee, T.S. NeRF-based 3D reconstruction pipeline for acquisition and analysis of tomato crop morphology. *Front. Plant Sci.* **2024**, *15*, 1439086. [CrossRef] [PubMed]
33. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Springer: Cham, Switzerland, 2020; pp. 405–421. [CrossRef]
34. Qiao, G.; Zhang, Z.; Niu, B.; Han, S.; Yang, E. Plant stem and leaf segmentation and phenotypic parameter extraction using neural radiance fields and lightweight point cloud segmentation networks. *Front. Plant Sci.* **2025**, *16*, 1491170. [CrossRef] [PubMed]
35. Ku, K.; Jubery, T.Z.; Rodriguez, E.; Balu, A.; Sarkar, S.; Krishnamurthy, A.; Ganapathysubramanian, B. SC-NeRF: NeRF-based point cloud reconstruction using a stationary camera for agricultural applications. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Nashville, TN, USA, 11–15 June 2025; pp. 5511–5520. [CrossRef]
36. Fang, J.; Xie, N.; Duan, Y.; Fan, J.; Wang, X.; Liu, H.; Zheng, H.; Wu, H.; Meng, Z.; Wang, X.; et al. F2DMAS: A smartphone video-based 3D phenotyping workflow for potted plants in complex backgrounds. *Front. Plant Sci.* **2026**, *17*, 1885579. [CrossRef] [PubMed]
37. Wang, Z.; Zhang, H.; Bian, L.; Zhou, L.; Ge, Y. Accurate plant 3D reconstruction and phenotypic traits extraction via stereo imaging and multi-view point cloud alignment. *Front. Plant Sci.* **2025**, *16*, 1642388. [CrossRef] [PubMed]
38. Qiu, H.; Zhao, Y.; Meng, X.; Wang, Y.; Li, X.; Niu, H.; Yu, L.; Mu, Q.; Meng, S.; Yin, S. A scalable framework for UAV-based point cloud reconstruction and organ segmentation of field-grown cotton in complex field environments. *Front. Plant Sci.* **2026**, *17*, 1854261. [CrossRef] [PubMed]
39. Wang, D.; Song, Z.; Miao, T.; Zhu, C.; Yang, X.; Yang, T.; Zhou, Y.; Den, H.; Xu, T. DFSP: A fast and automatic distance field-based stem-leaf segmentation pipeline for point cloud of maize shoot. *Front. Plant Sci.* **2023**, *14*, 1109314. [CrossRef] [PubMed]
40. Jiang, L.; Rodriguez-Sanchez, J.; Snider, J.L.; Chee, P.W.; Fu, L.; Li, C. Mapping of cotton bolls and branches with high-granularity through point cloud segmentation. *Plant Methods* **2025**, *21*, 66. [CrossRef] [PubMed]
41. Lei, Z.; Zeng, D.; Zheng, L. PointNeXt-DBSCAN: A hybrid point cloud deep learning framework for multi-stage cotton leaf instance segmentation. *Front. Plant Sci.* **2026**, *16*, 1705564. [CrossRef] [PubMed]
42. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems 30*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 5099–5108. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/hash/d8bf84be3800d12f74d8b05e9b89836f-Abstract.html (accessed on 6 September 2026).