

## Article

# Plant-GeoAT: Geometry-Aware Organ Identity Parsing of 3D Plant Point Clouds for Organ-Level Phenotyping

Junjie Wu <sup>1</sup>, Bei Zhou <sup>1,2,\*</sup>, Jie Liu <sup>3</sup>, Zhuyang Xie <sup>1,2</sup>, Jie Luo <sup>3</sup>, Jiajun Liu <sup>1</sup>, Jiahui Zhu <sup>1</sup>, Mingju Li <sup>1</sup>, Yan Ai <sup>1</sup>, Mingwei Liu <sup>1</sup> and Yu Chen <sup>1</sup>

<sup>1</sup> College of Information Engineering, Sichuan Agricultural University, Ya'an 625014, China

<sup>2</sup> Sichuan Key Laboratory of Agricultural Information Engineering, Sichuan Agricultural University, Ya'an 625014, China

<sup>3</sup> College of Agronomy, Sichuan Agricultural University, Chengdu 611130, China

\* Correspondence: 12801@sicau.edu.cn; Tel.: +86-189-8000-0216

## Abstract

Three-dimensional plant point clouds retain crop architecture, but organ-level phenotyping requires reliable semantic separation of leaves and stems. Plant-GeoAT is a geometry-aware parsing network that encodes local relative-XYZ neighbourhoods before RGB fusion and couples spatial neighbourhoods with feature-space relations for dense point prediction. We evaluated the model separately within the native protocol of a self-built structure-from-motion rapeseed dataset, an image-based soybean dataset, and laser-scanned Pheno4D maize and tomato datasets; these are within-dataset train/test experiments, not cross-dataset transfer or domain-generalisation tests. Across five seeds, mIoU was  $92.26 \pm 0.28\%$ ,  $82.50 \pm 0.24\%$ ,  $99.74 \pm 0.05\%$ , and  $94.75 \pm 0.15\%$ , respectively. Maize Stem IoU reached  $99.57 \pm 0.09\%$ . Adding LLGE increased the four-dataset average mIoU from 66.38% to 85.84%, and the complete LLGE + SSCA model reached 92.31%. On the fixed six-sample test set, exploratory semantic-guided clustering achieved  $86.67 \pm 7.45\%$  Count Accuracy; structural correctness ranged from 3/6 to 4/6 samples across seeds. Plant-height consistency was assessed independently on all 60 reconstructed rapeseed samples. The results indicate that geometry-to-context encoding produces organ-level semantic units for subsequent phenotyping, while independent instance-labelled datasets and additional growth stages are still needed for trait-level validation.

**Keywords:** 3D point-cloud; precision agriculture; high-throughput phenotyping; organ identity parsing; semantic segmentation; plant structure analysis



Academic Editor: Ainong Shi

Received: 12 August 2026

Revised: 30 August 2026

Accepted: 31 August 2026

Published: 15 September 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

High-throughput plant phenotyping platforms (HTPPs) and plant-phenomics programmes require structural measurements that can be repeated across samples and interpreted in biological terms [1–6]. Leaf number, stem continuity, and leaf–stem attachment are architectural traits, but single-view 2D images represent them through a projection. Overlap, occlusion, and viewpoint dependence can obscure organ boundaries. Three-dimensional point clouds retain coordinates and colour, providing a more direct representation of plant structure for quantitative phenotyping and a structured input for automated HTPP pipelines [5,6].

Three-dimensional reconstruction supplies plant geometry, but organ-level measurements also require the biological identity of each point. Structure-from-motion, multi-view

stereo, and laser scanning produce geometrically informative point sets without assigning leaf, stem, main-stem, or junction labels [4,7–13]. Dense organ parsing is therefore the computational link between reconstructed coordinates and phenotyping-ready leaf and stem units.

Existing plant point-cloud research has established complementary learning-based routes to organ analysis. PlantNet and graph-based semantic-instance networks jointly recover semantic and instance structure [14,15]. Eff-3DPSeg addresses annotation efficiency [11], PST demonstrates attention-based rapeseed segmentation [16], and DFSP uses explicit geometry for maize stem–leaf separation [17]. Recent graph-attention and reconstruction-to-phenotyping studies further connect point-cloud learning with crop analysis [13,18].

Recent Agronomy studies provide particularly relevant crop-specific context. Geometry-driven pipelines have segmented rapeseed seedlings and extracted morphological traits, evaluated rapeseed leaf segmentation from binocular point clouds, and separated corn stems and leaves through octree voxelization and multi-stage region growing [19–21]. Point-cloud completion has also been used to recover occluded corn geometry [22]. These studies establish the value of crop-specific reconstruction, geometric segmentation, and trait-extraction pipelines; they do not, however, resolve how a learned organ-identity representation should be shared across crop structures and acquisition sources.

Therefore, three gaps remain. First, generic point transformers such as PointTransformerV3 [23] generalise across scenes but do not explicitly encode relative-XYZ evidence before colour fusion, so thin stems and leaf–stem junctions can be confused when RGB appearance is similar. Second, organ-specialised models such as PointLIBR rely on super-point priors and EMA branches that are sensitive to prior quality, and they are not designed to operate jointly across COLMAP, image-based MVS, and laser-scanned inputs. Third, most studies report organ segmentation in isolation and do not show whether the semantic output is geometrically clean enough to feed a simple, learning-free downstream tool. Together, these gaps motivate a focused question: how should local geometry be combined with contextual relations to obtain reliable leaf/stem identity across crop structures and point-cloud sources, and does that identity transfer directly to a lightweight phenotyping step?

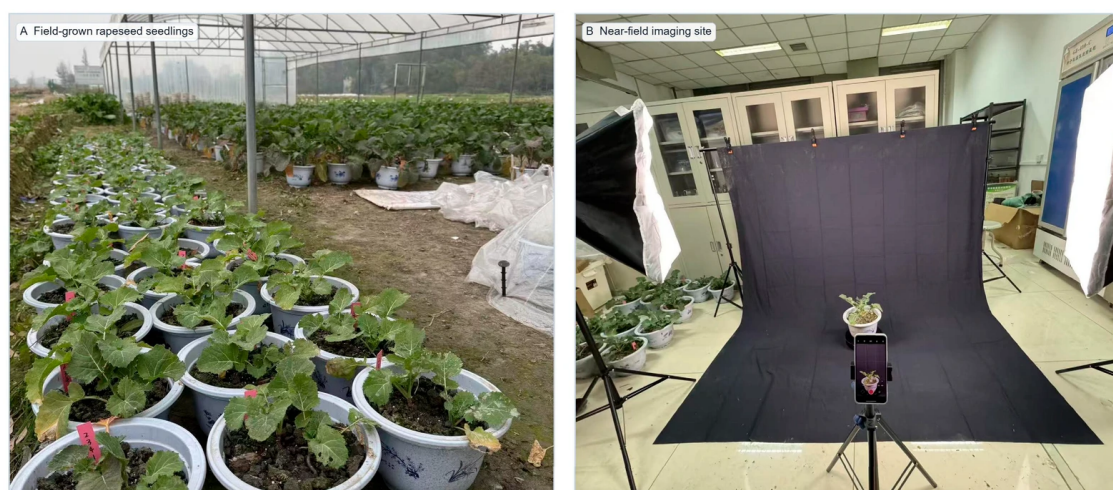
In this paper, plant architecture is used in the limited sense of the observable spatial arrangement and continuity of leaf and stem organs. We do not infer organ connections, developmental hierarchy, or a complete plant topology; those relations would require additional instance and graph annotations. Leaf/stem identity depends on both local shape and surrounding structure. A thin leaf edge, a petiole-like leaf base, a young stem, and a leaf–stem junction can have similar local width or curvature. Local geometry is needed to distinguish these structures, while spatial and feature-space context is needed when organs overlap or share similar point-wise appearance. This motivates a geometry-to-context representation in which relative coordinates are encoded before contextual relation modelling.

Plant-GeoAT implements this geometry-to-context representation in two coupled modules. The learnable local geometry extractor (LLGE) encodes relative-XYZ neighbourhoods before RGB fusion, preserving local shape cues at leaf edges, stems, and junctions. The spatial–semantic coupled attention module (SSCA) then combines local spatial neighbourhoods with feature–space EdgeConv relations. The novelty is the ordering and coupling for organ identity rather than a claim that relative coordinates, attention, or EdgeConv are individually new: unlike conventional early-feature concatenation, LLGE keeps the relative-XYZ stream separate until neighbourhood aggregation, and SSCA fuses a radius-defined physical branch with a learned feature-neighbour branch at every encoder stage.

This differs from generic point transformers that fuse coordinates and appearance in a shared token pathway and from EdgeConv-only networks that model feature neighbours without a parallel physical-context branch.

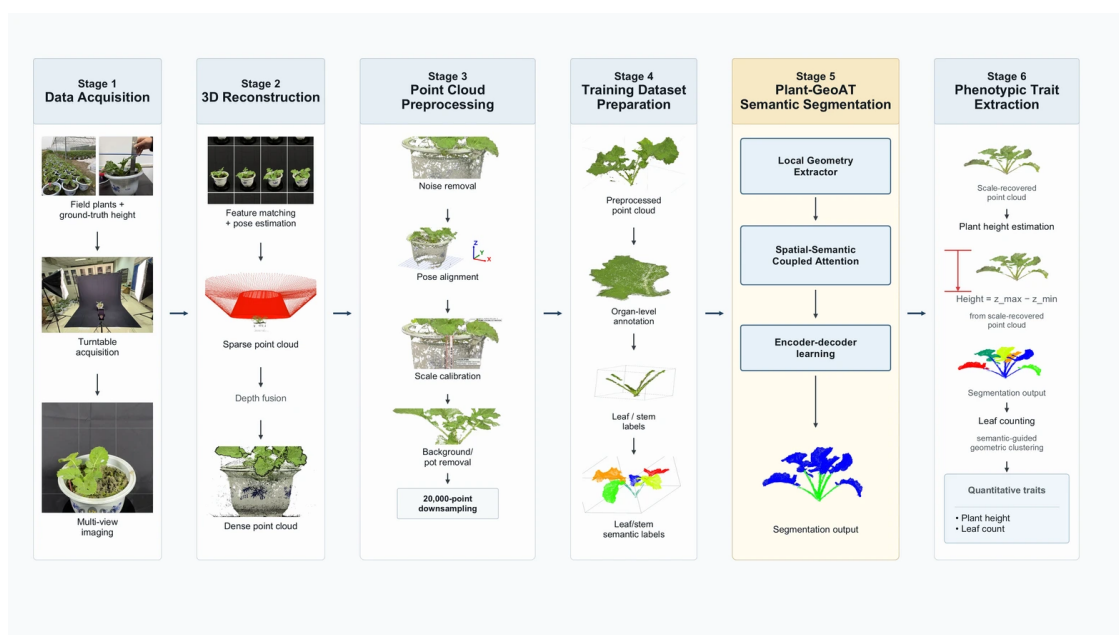
The design was evaluated on four datasets that differ in crop architecture and acquisition source: self-built COLMAP-based rapeseed, an image-based soybean dataset, and laser-scanned Pheno4D maize and tomato point clouds. Each dataset retained its native semantic label space and was trained and tested separately; the experiment therefore measures robustness across separately fitted dataset-specific models, not cross-dataset generalisation. The evaluation combines dataset-level mIoU, class-wise stem performance, qualitative visualisation, module ablation, reconstruction quality checks, and an exploratory downstream leaf-counting test.

The self-built dataset supports this application framing at the seedling stage of rapeseed growth. Rapeseed seedlings of cultivar Rongyou 13 were sown in late September 2024 and grown under normal field conditions in the experimental field in Wenjiang District, Chengdu, Sichuan Province, China. They grew in the native agricultural soil, which is clay-dominant in this part of Sichuan, and were watered approximately once every 7 days when rainfall did not make irrigation unnecessary. Public ERA5 reanalysis data for the approximate site (approximate Wenjiang site) were obtained through the Open-Meteo Historical Weather API with the model explicitly set to ERA5 (accessed on 27 August 2026). ERA5 is produced by the European Centre for Medium-Range Weather Forecasts and distributed through the Copernicus Climate Change Service Climate Data Store [24–26]. The resulting mean daily temperature was 18.2 degrees C from 25 September to 12 November 2024; the corresponding means were 23.8 degrees C for 25–30 September, 17.8 degrees C in October, and 16.4 degrees C for 1–12 November. These gridded reanalysis values are background estimates rather than measurements from an on-site weather station. The plants were photographed in a simple room adjacent to the field. Near-field controlled acquisition reduced background clutter, wind-induced motion, and rapid illumination variation during multi-view reconstruction while retaining field-grown seedling structure. The workflow used a consumer camera, a turntable, and two fill lights to provide an offline, low-cost research setting for organ-level point-cloud analysis. An example of the plant material and imaging setup is shown in Figure 1.



**Figure 1.** Plant material and near-field image acquisition for the self-built seedling-stage rapeseed dataset. (A) Field-grown seedling-stage rapeseed plants used for dataset construction. (B) Near-field imaging site set up in a small room adjacent to the field. The cultivar was Rongyou 13.

This study makes three contributions. First, it formulates post-reconstruction leaf/stem parsing as the organ-identity step that makes 3D plant geometry usable for structural phenotyping, and defines a geometry-to-context representation that addresses the relative-XYZ-before-colour gap left by generic point transformers. Second, it introduces LLGE + SSCA coupling, in which LLGE extracts relative-XYZ cues before RGB fusion and SSCA couples spatial and feature-space relations, providing an organ-specialised alternative to superpoint-prior models such as PointLIBR. Third, it tests this pathway across four crop datasets and two acquisition classes, and links the semantic output to reconstruction quality assessment and a semantic-guided leaf-counting test that uses only a simple, learning-free clustering step, closing the loop from reconstruction to a downstream phenotyping operation. The resulting organ-resolved measurements are intended as a digital-phenotyping component for future crop-breeding and cultivation studies; the present experiments do not evaluate genetic response, yield, stress tolerance, or management effects. An overview of this workflow is shown in Figure 2.



**Figure 2.** Workflow for Plant-GeoAT-based 3D plant point-cloud segmentation and organ-level phenotyping. The workflow proceeds from field-plant and multi-view image acquisition through 3D reconstruction, point-cloud preprocessing, training-dataset preparation, Plant-GeoAT semantic segmentation, and downstream phenotypic trait extraction. Solid blue arrows indicate the main stage-to-stage processing flow. Point-cloud preprocessing follows the order of denoising, pose alignment, scale recovery, plant/background separation, Statistical Outlier Removal (SOR), and 20,000-point downsampling; the exact SOR settings are given in Section 2.2. Within Stage 5, the blue rounded boxes denote local geometry extraction, Spatial-Semantic Coupled Attention, and encoder-decoder learning; the pale-yellow panel highlights the Plant-GeoAT semantic-segmentation stage. In Stage 6, the red bracket marks the vertical extent used for plant-height calculation from the scale-recovered point cloud; the colored point clouds illustrate semantic-segmentation output, and semantic-guided geometric clustering is used for leaf counting. The final panel summarizes the quantitative traits considered: plant height and leaf count.

## 2. Materials and Methods

### 2.1. Datasets, Label Definitions, and Random Rapeseed Split

The evaluation used one self-built seedling-stage rapeseed dataset and three public plant point-cloud datasets: Soybean-MVS, Pheno4D-Maize, and Pheno4D-Tomato. Their plant structures and annotation protocols differ, so each dataset was trained and evaluated

within its native semantic label space. For all three public datasets, the source train/test partitions and training-time augmentation strategies followed the corresponding source publications. Where a validation subset was used for checkpoint selection, it was drawn only from the corresponding training partition with plant/model-level isolation; no test plant, model, or date entered training or validation, so no data leakage occurred. Augmentation was applied only to training data, while validation and test data were not augmented. Plant-GeoAT-specific point sampling, input construction, and optimisation settings were then applied consistently within each dataset. No model was trained on one crop and evaluated on another; the four-dataset summary is a descriptive collection of independent within-dataset experiments, not a cross-dataset generalisation score.

The study used two linked evaluation components. Model evaluation measured organ parsing on self-built rapeseed, Soybean-MVS, Pheno4D-Maize, and Pheno4D-Tomato. Application-oriented analyses used the self-built rapeseed data to assess reconstruction and preprocessing quality and to explore whether predicted semantic leaf points could support conventional clustering and counting. The first component evaluates the network; the second examines how its output can enter a phenotyping workflow. Table 1 summarises the dataset-specific label spaces, splits, structural focus, and evaluation context.

**Table 1.** Plant point-cloud datasets used in this study.

| Dataset                              | Source     | Semantic Label Space | Samples and Split                                                                                   | Main Structural Focus                                                      | Evaluation Context                                                                                       |
|--------------------------------------|------------|----------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| Seedling-stage rapeseed (Rongyou 13) | Self-built | Leaf, Stem           | 60 original samples: 42 train, 12 validation, and 6 test; fixed random allocation                   | Dense rosette-like canopy, leaf–stem occlusion, low geometric separability | Random allocation was fixed for all reported experiments; validation and test samples were not augmented |
| Soybean-MVS                          | Public     | Leaf, MainStem, Stem | 89 plant models for training and 13 plant models for testing                                        | Branching structure, stem subclass discrimination, organ overlap           | MainStem and Stem are dataset-specific subclasses                                                        |
| Pheno4D-Maize                        | Public     | Leaf, Stem           | Plant-level split: 5 plants for training and 2 for testing; all dates from each plant kept together | Elongated structure and slender stem continuity                            | Leaf/Stem labels interpreted within the maize annotation protocol                                        |
| Pheno4D-Tomato                       | Public     | Leaf, Stem           | Plant-level split: 5 plants for training and 2 for testing; all dates from each plant kept together | Intersecting leaf–stem structures and organ overlap                        | Leaf/Stem labels interpreted within the tomato annotation protocol                                       |

Table note: Soybean-MVS retained the source–publication split of 89 training models and 13 test models. Pheno4D-Maize and Pheno4D-Tomato were trained separately using the source–publication plant-level split of five plants for training and two for testing. The source training-time augmentation strategies were retained and applied only to training data; validation and test data were not augmented. Any validation subset used for checkpoint selection remained inside the corresponding training partition, and no test plant, model, or date entered training or validation. Plant-GeoAT input sampling and optimisation settings are reported separately.

For the self-built seedling-stage rapeseed dataset, 60 original point clouds were used. The planting site was in Wenjiang District, Chengdu, Sichuan Province, China, and the cultivar was Rongyou 13. Each sample was stored as a  $20,000 \times 7$  array with  $x$ ,  $y$ ,  $z$ ,  $r$ ,  $g$ ,  $b$ , and semantic label fields. The semantic classes were Stem = 0 and Leaf = 1. The original samples were randomly allocated into 42 training samples, 12 validation samples, and six test samples, and this allocation was fixed for all reported experiments. The six test-sample IDs were 11, 16, 22, 24, 26, and 50. The 42 original training samples were expanded offline into 840 training files by generating 20 variants per sample. Validation and test samples were not augmented.

The seedling-stage rapeseed annotations focused on leaf–stem junctions, petiole-like structures, and young stem continuity, where local geometry has low separability. Soybean-MVS retained Leaf, MainStem, and Stem as separate classes according to its source protocol. Pheno4D provides high-accuracy laser-scanned maize and tomato point clouds with manually labelled points [27], whereas Soybean-MVS was reconstructed from multi-view images and supports soybean organ-segmentation analysis [28]. Cross-dataset summaries therefore describe performance under the native protocol of each dataset.

## 2.2. Multi-View Reconstruction and Preprocessing

The self-built seedling-stage rapeseed point clouds were generated from multi-view images of field-grown plants. To reduce uncontrolled background and motion during reconstruction, images were acquired in a simple near-field shooting setup adjacent to the field. The plants retained their field-grown architecture, while the acquisition environment provided more stable conditions for 3D reconstruction.

Dense point clouds were reconstructed from multi-view image sets using COLMAP version 3.10 [29,30]. Following the order summarised in Figure 2, preprocessing comprised denoising by colour-filtering black points, pose alignment using the pot bottom as the reference plane, coordinate transformation and scale recovery using the known pot height, manual plant/background separation to obtain a plant-only point-cloud, Statistical Outlier Removal (SOR), and 20,000-point farthest point sampling (FPS) downsampling. SOR used 30 points for mean-distance estimation and a standard-deviation multiplier threshold ( $n\sigma$ ) of 1.5. Organ-level semantic annotation and subsequent analyses used the resulting plant-only point clouds. Pose alignment used the pot bottom as a reference plane so that the vertical coordinate could support later height calculation. On the local workstation used for dataset preparation, one COLMAP reconstruction required approximately 1 h and preprocessing of one plant required approximately 0.5 h; these are approximate workflow times rather than fixed model-inference benchmarks and depend on hardware, image count, and I/O.

Scale recovery was performed after coordinate correction using the known pot height as the physical reference. For each sample, points were binned along the corrected vertical  $z$ -axis, and the number of points in each height bin was counted. Because the pot rim forms a dense circular structure, it produced a high-density band or peak in the height-point-count profile. This band was used to locate the reconstructed pot rim height, while the pot bottom reference plane provided the lower boundary. A fixed height-bin setting was used consistently for all samples. The identified pot-rim band was checked against the aligned pot point-cloud to reduce the risk of confusing the pot rim with dense plant layers. The reconstructed pot height was then compared with the measured physical pot height to obtain an individual scale factor for that plant:

$$s_m = \frac{H_{\text{pot}}^{\text{real}}}{H_{\text{pot},m}^{\text{rec}}}. \quad (1)$$

Here,  $s_m$  is the scale factor for plant  $m$ ,  $H_{\text{pot}}^{\text{real}}$  is the measured physical pot height, and  $H_{\text{pot},m}^{\text{rec}}$  is the reconstructed pot height for that sample. The scale factor was applied only to spatial coordinates:

$$\mathbf{x}_{i,m}^* = s_m \mathbf{x}_{i,m}^{\text{rec}}, \quad (2)$$

where  $\mathbf{x}_{i,m}^{\text{rec}}$  and  $\mathbf{x}_{i,m}^*$  denote the reconstructed and scale-recovered XYZ coordinates, respectively. RGB values and semantic labels were not scaled. This factor was applied separately to each reconstructed plant point-cloud before plant-only 20,000-point FPS downsampling and subsequent analysis.

Reconstruction and preprocessing quality was assessed before the phenotyping analyses. Manual comparison with source images checked retention of the main plant structure, effective leaves, visible leaf–stem topology, and foreground separation. Plant height from all 60 scale-recovered point clouds provided a quantitative check of vertical metric consistency. The manual reference height was measured with a long ruler from the root-/pot-base level to the highest visible point. After denoising by black-point colour filtering, pose alignment, coordinate transformation and scale recovery, manual plant-only segmentation, and SOR filtering (30 points for mean-distance estimation; nSigma = 1.5), all before the final 20,000-point downsampling, the resulting plant-only point-cloud was used for the height calculation; no semantic label or learned height model was used. Any isolated point surviving the preprocessing filter would affect the simple maximum/minimum calculation, which is noted as a limitation.

During model training, each network input was a 4096-point random-centre KNN block sampled from the 20,000-point-cloud. This 4096-point block is not input-level farthest point sampling. Farthest point sampling is used inside the encoder set-abstraction stages to select centroids for hierarchical feature abstraction. During inference, the complete 20,000-point plant scene was passed through the network; the decoder propagated predictions back to all input points, so no block-level voting or overlapping-block aggregation was used for the reported plant-level metrics.

### 2.3. Problem Formulation

An input point-cloud block is represented as:

$$P = \{\mathbf{p}_i\}_{i=1}^N, \quad \mathbf{p}_i = [\mathbf{x}_i, \mathbf{c}_i], \quad (3)$$

$$\mathbf{x}_i = (x_i, y_i, z_i), \quad \mathbf{c}_i = (r_i, g_i, b_i), \quad (4)$$

where  $\mathbf{x}_i$  is the XYZ coordinate and  $\mathbf{c}_i$  is the RGB colour. Plant-GeoAT uses XYZRGB and an auxiliary relative-XYZ neighbourhood tensor and  $\Delta X$  as an input. Normals, handcrafted curvature, height, and PCA-derived features are not used by the segmentation network.

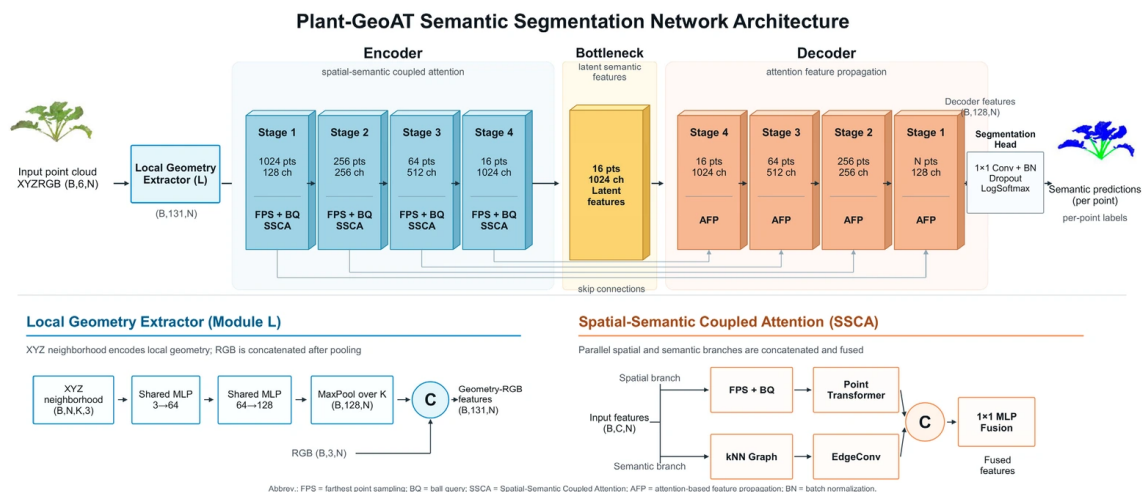
The model predicts dataset-specific semantic labels:

$$\hat{y}_i = f_{\theta}(P, \Delta X)_i, \quad \hat{y}_i \in \mathcal{Y}_D. \quad (5)$$

The dataset-specific label space is:

$$\mathcal{Y}_D = \begin{cases} \{\text{Leaf, Stem}\}, & D \in \{\text{Rapeseed, Maize, Tomato}\}, \\ \{\text{Leaf, MainStem, Stem}\}, & D = \text{Soybean-MVS}. \end{cases} \quad (6)$$

This formulation keeps the segmentation task aligned with each dataset's annotation protocol. The complete encoder–decoder data flow and point-wise prediction path are shown in Figure 3.



**Figure 3.** Overall architecture of Plant-GeoAT for geometry-aware organ identity parsing of 3D plant point clouds. The network receives an XYZRGB point cloud, followed by the Learnable Local Geometry Extractor (LLGE) that produces Geometry-RGB features. The encoder contains four progressively downsampled stages; the blue blocks show the point and channel dimensions, and FPS, BQ, and SSCA denote farthest point sampling, ball query, and Spatial-Semantic Coupled Attention, respectively. The yellow bottleneck stores latent semantic features, while the peach decoder restores point resolution using attention-based feature propagation (AFP); the bottom skip-connection lines transfer features from the corresponding encoder stages. The final segmentation head applies  $1 \times 1$  convolution, batch normalization, dropout, and log-softmax to produce per-point semantic predictions. The lower panels expand the LLGE and SSCA operations. Arrows indicate the direction of tensor flow, the circled letter “C” indicates feature concatenation, and the colors distinguish the encoder, bottleneck, decoder, local-geometry encoding, and spatial and feature-space contextual-coupling components.

#### 2.4. Learnable Local Geometry Extractor and Geometry-RGB Features

For each point  $i$ , KNN in XYZ space defines a local neighbourhood:

$$\mathcal{N}_K(i) = \text{kNN}_K(x_i; \{x_j\}_{j=1}^N), \quad K = 32. \quad (7)$$

The local geometry evidence is represented by relative-XYZ offsets:

$$\Delta x_{ij} = x_j - x_i, \quad j \in \mathcal{N}_K(i). \quad (8)$$

LLGE encodes these relative vectors through shared one-dimensional convolutions:

$$g_{ij} = \phi_{\text{geo}}(\Delta x_{ij}), \quad (9)$$

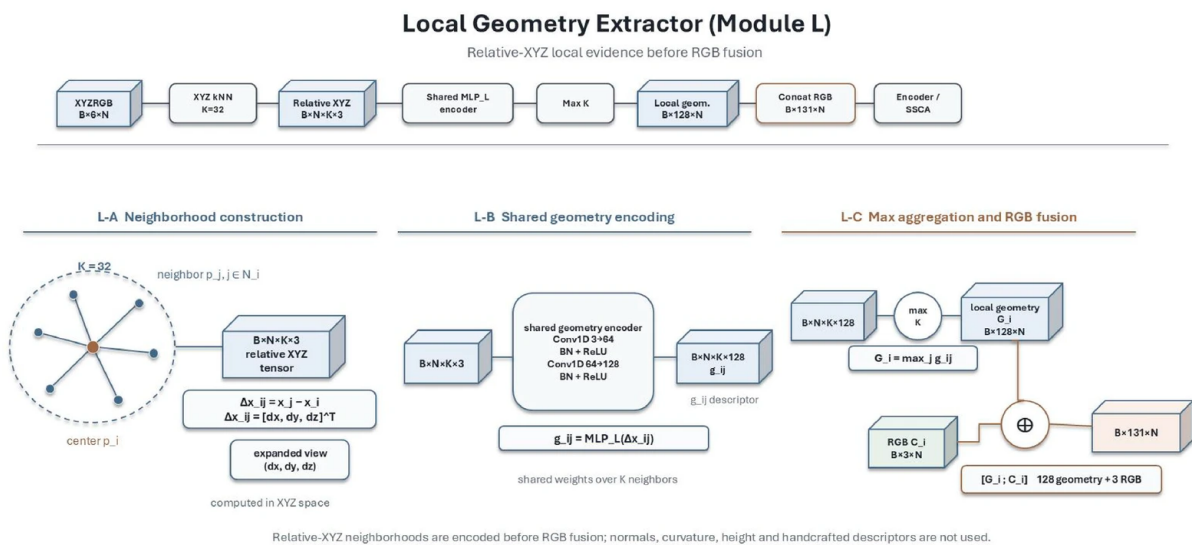
where  $\phi_{\text{geo}}$  is implemented as Conv1d(3, 64), BatchNorm, ReLU, Conv1d(64, 128), BatchNorm, and ReLU. A max operation over the  $K$  neighbours produces a 128-dimensional local geometry descriptor:

$$l_i = \max_{j \in \mathcal{N}_K(i)} g_{ij}. \quad (10)$$

The descriptor  $l_i$  captures local relative-XYZ evidence before RGB fusion. It is intended to make the early representation geometry-aware in regions such as thin stems, leaf-stem junctions, and local discontinuities. RGB is concatenated after local geometry aggregation:

$$h_i^0 = [l_i; c_i]. \quad (11)$$

This Geometry-RGB feature keeps geometry extraction separate from colour while still allowing colour to support subsequent organ prediction; the tensor flow is illustrated in Figure 4.



**Figure 4.** Learnable Local Geometry Extractor (LLGE) for relative-XYZ evidence encoding. The top row shows the tensor flow from the XYZRGB input through XYZ k-nearest-neighbor (KNN) neighborhood construction, relative-XYZ encoding, shared multilayer perceptron (MLP) layers, maximum pooling over neighbors, RGB concatenation, and the encoder/SSCA input. Here,  $K$  denotes the number of neighboring points and is set to 32. In the lower panels, L-A denotes relative-coordinate construction from a center point to its  $K$  neighbors, L-B denotes shared geometry encoding of each relative-XYZ vector, and L-C denotes maximum aggregation followed by concatenation of the local geometry descriptor with RGB. Arrows indicate the processing direction, the circled letter denotes feature concatenation, and the blue and orange headings distinguish the three processing steps. The displayed tensor dimensions describe the intermediate feature shapes; normals, curvature, height, and other handcrafted descriptors are not used.

## 2.5. Spatial–Semantic Coupled Attention

SSCA places Geometry-RGB features into a spatial-feature context. The term “semantic” refers to learned feature–space relations, not to ground-truth semantic labels or predicted class feedback inside the attention module.

At encoder stage  $s$ , the input feature is first lifted by a shared mapping:

$$u_i^s = \phi_s(h_i^{s-1}). \quad (12)$$

The spatial branch uses radius-based local neighbourhoods around sampled centroids. A local point-transformer block models the arrangement of grouped features and relative spatial offsets:

$$S_i^s = T_s\left(\left\{u_j^s, x_j - x_i \mid j \in \mathcal{B}_r(i)\right\}\right), \quad (13)$$

where  $\mathcal{B}_r(i)$  is the radius-based spatial neighbourhood and  $T_s$  denotes local transformer-based spatial context. The branch models radius-defined local context without applying global attention over the full point-cloud.

The feature–space branch constructs a KNN graph among learned features and applies EdgeConv-style coupling:

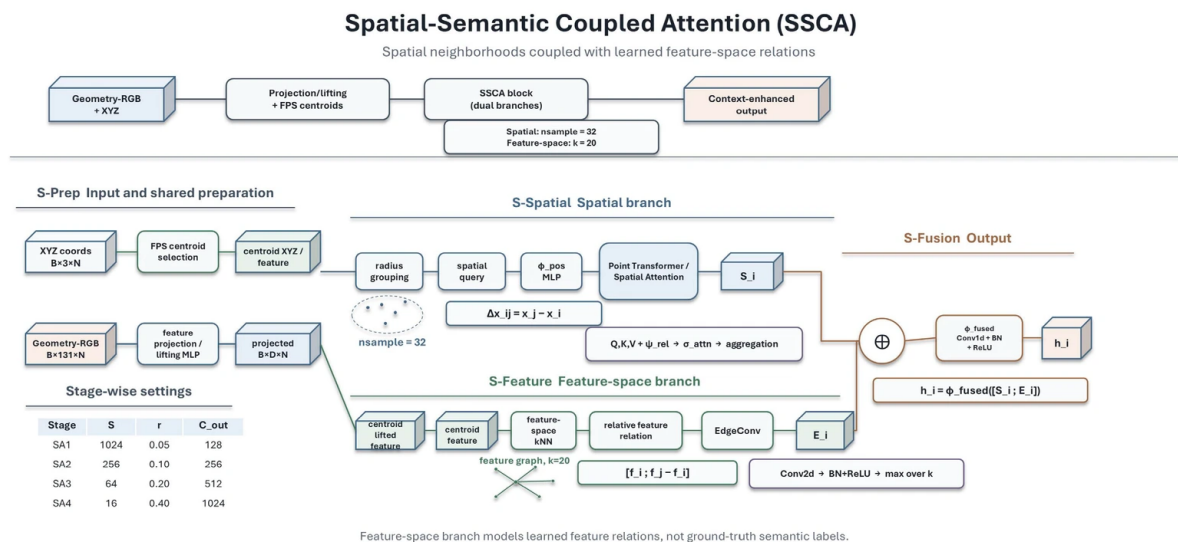
$$E_i^s = \max_{j \in \mathcal{G}_k(i)} \eta_s\left(\left[u_i^s, u_j^s - u_i^s\right]\right), \quad (14)$$

where  $\mathcal{G}_k(i)$  is the feature-space KNN neighbourhood and  $\eta_s$  denotes the EdgeConv mapping. This branch links points with similar learned responses and does not use semantic labels as inputs.

The two branches are fused as:

$$h_i^s = \psi_s([S_i^s; E_i^s]). \quad (15)$$

This fusion is the core of SSCA: local geometry evidence is interpreted under both spatial arrangement and feature-space relations before point-wise prediction. Figure 5 summarises the two branches and their fusion.



**Figure 5.** Spatial-Semantic Coupled Attention (SSCA) for spatial-feature contextual interpretation. The top row summarizes the pathway from Geometry-RGB and XYZ features through projection and lifting with FPS centroids, the dual-branch SSCA block, and the context-enhanced output. In the lower panels, S-Prep denotes shared input preparation. S-Spatial denotes the physical-neighborhood branch, where radius grouping, spatial querying, positional encoding, and Point Transformer spatial attention produce the spatial feature  $\backslash(S_i)\backslash$ . S-Feature denotes the learned feature-space branch, where feature-space k-nearest-neighbor search, relative feature relations, and EdgeConv produce the feature-space representation  $\backslash(E_i)\backslash$ . The two branch outputs are combined at the circled fusion symbol and passed through a fused  $1 \times 1$  convolution, batch normalization, and ReLU to obtain the context-enhanced output  $\backslash(h_i)\backslash$ . Arrows indicate the direction of data flow; three-dimensional rectangular boxes denote feature tensors; rounded rectangular boxes denote processing operations; the dashed ellipse represents a spatial neighborhood; the node-link diagram represents a feature-space neighborhood; and the circled plus symbol denotes feature fusion. Blue, green, and orange headings distinguish shared preparation, the spatial branch, the feature-space branch, and fusion/output operations. The displayed settings are  $\backslash(n_{\mathrm{sample}} = 32)\backslash$  for spatial grouping and  $\backslash(k = 20)\backslash$  for the feature-space graph; the stage-wise table lists the number of centroid points, radius, and output channels.

## 2.6. Decoder Feature Propagation and Semantic Prediction

The encoder contains four abstraction stages with centroid counts of 1024, 256, 64, and 16. The corresponding spatial radii are 0.05, 0.10, 0.20, and 0.40; the spatial neighbourhood size is 32; the feature-space neighbourhood size is 20; and the stage output channels are 128, 256, 512, and 1024. These settings are listed in Table 2.

**Table 2.** Encoder-stage settings used in the Plant-GeoAT architecture.

| Encoder Stage | Centroids | Radius | Spatial Neighbours | Feature-Space Neighbours | Output Channels |
|---------------|-----------|--------|--------------------|--------------------------|-----------------|
| Stage 1       | 1024      | 0.05   | 32                 | 20                       | 128             |
| Stage 2       | 256       | 0.10   | 32                 | 20                       | 256             |
| Stage 3       | 64        | 0.20   | 32                 | 20                       | 512             |
| Stage 4       | 16        | 0.40   | 32                 | 20                       | 1024            |

The decoder uses attention-based feature propagation and skip connections to restore dense point-wise features. For each finer-resolution point, the nearest three coarser points provide the propagation neighbourhood. The attended coarse feature is fused with the skip feature and refined by shared MLP layers. This step restores context-conditioned features to the original point resolution; it is not an instance segmentation branch.

The final prediction head maps the decoded 128-channel feature to per-point semantic logits, followed by softmax classification. In Equation (16), the logits are the unnormalised vector  $z_i = W h_i^{\text{dec}}$ ,  $p_{ic}$  is the softmax probability for class  $c$ , and  $\hat{y}_i$  is the class index with the largest  $p_{ic}$ .

$$z_i = W h_i^{\text{dec}}, p_{ic} = \text{softmax}(z_{ic}), \hat{y}_i = \underset{c}{\operatorname{argmax}} p_{ic} \quad (16)$$

Plant-GeoAT outputs semantic labels only. Leaf-instance recovery for counting and plant height calculation are model-external phenotyping steps.

### 2.7. Training and Baseline Protocol

Plant-GeoAT was implemented in PyTorch. The seedling-stage rapeseed training configuration used XYZRGB input, relative-XYZ neighbourhoods with  $K = 32$ , 4096-point random-centre KNN blocks, AdamW optimisation, an initial learning rate of  $1 \times 10^{-4}$ , a cosine annealing learning-rate schedule, and checkpoint selection based on best validation mIoU. The model was configured for a maximum of 3000 epochs with early stopping after 100 validation epochs without improvement. Across the checked archived training logs, the mean run duration was approximately 25 h 29 min 28 s, with a mean of 240 completed epochs before early stopping and approximately 6.37 min per completed epoch; the 3000-epoch setting was only the maximum allowance. Five repeated runs (S1–S5) were evaluated under the stated fixed splits and configurations.

Plant-GeoAT was compared with PointNet++, DGCNN, PointTransformerV3, PointVector, and PointLIBR [23,31–42]. Each baseline retained its native loss function and model-specific configuration rather than being forced to use Plant-GeoAT's hybrid loss. The common controls were the same dataset-specific FPS preprocessing and input convention where supported, the same fixed split and label space, the best observed configuration archived for that model, and five independent seed runs (S1–S5) where the multi-seed table is reported. Checkpoints were selected by the highest validation mIoU before test evaluation. Epoch ceilings and batch sizes were not identical for all implementations, so this is a controlled protocol comparison rather than a claim of equal wall-clock or equal-compute budgets.

### 2.8. Phenotyping-Related Analyses

Application-oriented analyses used the self-built seedling-stage rapeseed data. Reconstruction checks establish the quality of the image-based point clouds. Semantic-guided leaf counting was evaluated across S1–S5 on the same fixed six-sample test set; checkpoint

selection used the separate validation split. HDBSCAN and fragment-fusion parameters were fixed from prior empirical implementation experience and held constant across S1–S5; they were not systematically tuned on these six samples. The counting evaluation was independent of checkpoint selection; however, it remains an exploratory phenotyping analysis because it contains only six test plants and lacks external instance-labelled replication.

Leaf counting was examined as an exploratory analysis on the same six-sample test set; checkpoint selection used the separate validation split. Plant-GeoAT semantic predictions first provide leaf candidate points:

$$P_{\text{leaf}} = \{p_i \mid \hat{y}_i = \text{Leaf}\}. \quad (17)$$

A model-external semantic-guided geometric clustering step then groups candidate leaf points:

$$C = G(P_{\text{leaf}}; \theta_g), \quad (18)$$

where  $G$  denotes the downstream clustering procedure and  $\theta_g$  denotes its geometric clustering parameters.

In this study,  $G$  was implemented as a conventional density/geometric clustering procedure. This model-external choice tests whether Plant-GeoAT semantic leaf points can feed an ordinary grouping operation without adding a learned instance module.

The predicted leaf count is:

$$\hat{N}_{\text{leaf}} = |C|. \quad (19)$$

This model-external procedure evaluates semantic-guided counting with conventional geometric clustering; it does not introduce a learned leaf-instance head.

Plant height was computed separately from the denoised, pose-aligned, scale-recovered point-cloud  $P^*$ :

$$H = z_{\max}(P^*) - z_{\min}(P^*). \quad (20)$$

Height estimation uses no semantic labels. It quantifies vertical metric consistency after COLMAP reconstruction, preprocessing, pose alignment, and per-plant pot-height scale recovery.

Semantic segmentation performance was evaluated using mIoU and class-wise IoU. For a dataset-specific class  $c$  in  $\mathcal{Y}_D$ ,  $TP_c$ ,  $FP_c$ , and  $FN_c$  denote true-positive, false-positive, and false-negative point counts for class  $c$ , respectively. IoU is defined as:

$$\text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}. \quad (21)$$

Dataset-specific mIoU is then:

$$\text{mIoU} = \frac{1}{|\mathcal{Y}_D|} \sum_{c \in \mathcal{Y}_D} \text{IoU}_c. \quad (22)$$

Precision, Recall, and F1-score were used for class-wise segmentation summaries and for the leaf-clustering structural review where applicable:

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}, \quad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c}, \quad (23)$$

$$\text{F1}_c = \frac{2\text{Precision}_c\text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}. \quad (24)$$

Leaf-counting error was summarised using Count Accuracy, MAE, and RMSE. For  $n$  evaluated plants with ground-truth leaf counts  $N_t$  and predicted leaf counts  $\hat{N}_t$ , Count Accuracy is:

$$\text{CountAccuracy} = \frac{1}{n} \sum_{t=1}^n 1(\hat{N}_t = N_t) \times 100\%. \quad (25)$$

$$\text{MAE}_{\text{count}} = \frac{1}{n} \sum_{t=1}^n |\hat{N}_t - N_t|, \quad (26)$$

$$\text{RMSE}_{\text{count}} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{N}_t - N_t)^2}. \quad (27)$$

Height estimation was summarised using  $R^2$ , RMSE, MAE, and MAPE. For measured height  $H_t$ , estimated height  $\hat{H}_t$ , and mean measured height  $\bar{H}$ , these metrics are:

$$R^2 = 1 - \frac{\sum_{t=1}^n (H_t - \hat{H}_t)^2}{\sum_{t=1}^n (H_t - \bar{H})^2}, \quad (28)$$

$$\text{RMSE}_{\text{height}} = \sqrt{\frac{1}{n} \sum_{t=1}^n (H_t - \bar{H})^2}, \text{MAE}_{\text{height}} = \frac{1}{n} \sum_{t=1}^n |H_t - \bar{H}|. \quad (29)$$

$$\text{MAPE}_{\text{height}} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{H_t - \hat{H}_t}{H_t} \right|. \quad (30)$$

### 3. Results

#### 3.1. Reconstruction and Preprocessing Quality of Self-Built Seedling-Stage Rapeseed Point Clouds

Manual inspection confirmed that cropping, filtering, and plant-only downsampling retained the main plant structure, effective leaves used for annotation and counting, and visible leaf–stem topology. Major pot and background artefacts were removed. The approximate preparation time was 1 h per plant for COLMAP reconstruction and 0.5 h per plant for preprocessing on the local workstation; these timings describe data preparation rather than network inference.

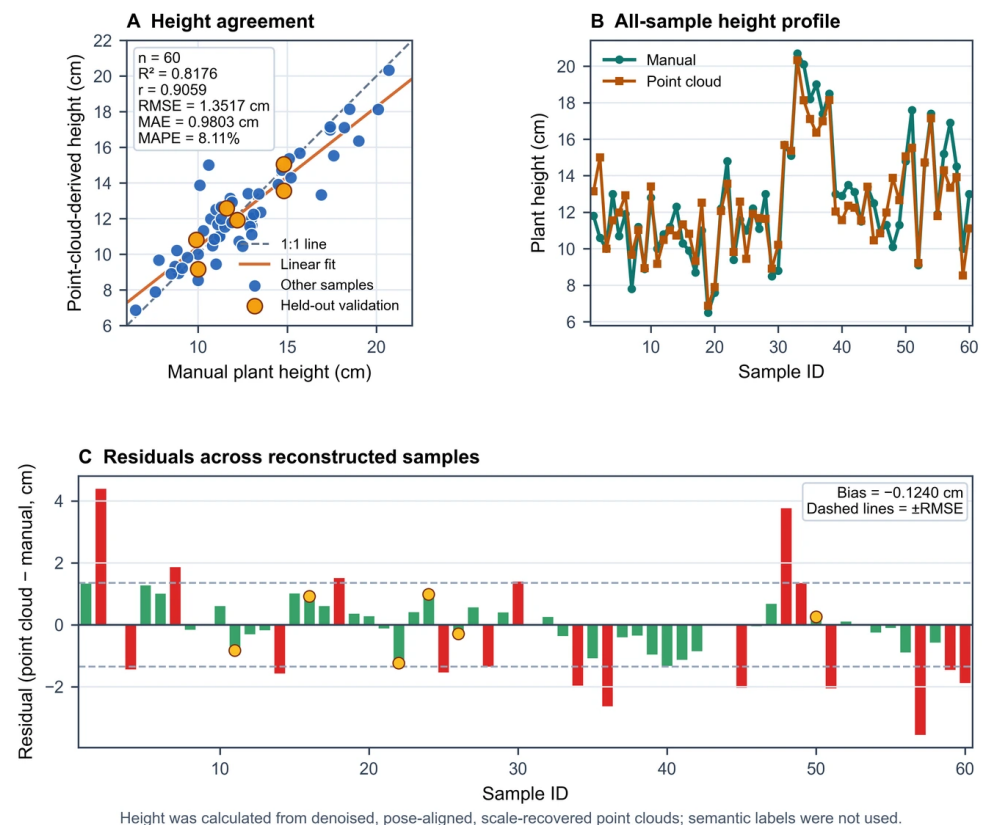
Height values were converted from the reconstructed coordinate unit to centimetres after per-plant scale recovery; RMSE and MAE are reported in cm in Table 3. The quantitative height component was then calculated on all 60 reconstructed and preprocessed seedling-stage rapeseed samples. The calculation uses denoised, pose-aligned, per-plant scale-recovered point clouds without semantic labels, so it quantifies vertical metric consistency of the reconstruction and scale-recovery workflow. The resulting metrics are  $\text{RMSE} = 1.3517$  cm,  $\text{MAE} = 0.9803$  cm, and  $\text{MAPE} = 8.1133\%$ , consistent with the values reported in Table 3. The corresponding all-sample visualisation is shown in Figure 6.  $H = z_{\max} - z_{\min} R^2 = 0.8176$ .

**Table 3.** Plant-height quality validation based on scale-recovered point clouds.

| Metric Category        | Metric    | Result |
|------------------------|-----------|--------|
| Regression consistency | $R^2$     | 0.8176 |
| Absolute error         | RMSE (cm) | 1.3517 |
| Absolute error         | MAE (cm)  | 0.9803 |
| Relative error         | MAPE (%)  | 8.1133 |

Table note: Plant height was calculated from all 60 denoised, pose-aligned, and per-plant scale-recovered seedling-stage rapeseed point clouds. The metrics quantify vertical consistency of the reconstruction, preprocessing, and scale-recovery workflow; no semantic labels are used.

## Plant height quality validation for reconstructed and preprocessed rapeseed point clouds



**Figure 6.** Plant-height quality validation after point-cloud reconstruction and preprocessing. Plant height was calculated for all 60 self-built seedling-stage rapeseed samples from denoised, pose-aligned, and per-plant scale-recovered point clouds as the vertical coordinate range. The figure reports all-sample agreement between manual height and point-cloud-derived height, sample-wise height comparison, and residual distribution. This analysis evaluates whether the COLMAP reconstruction, preprocessing, pose alignment, and pot-height scale recovery preserve sufficient metric structure for a simple geometric trait. The analysis is independent of semantic labels and quantifies reconstruction and preprocessing quality. The dashed lines indicate  $\pm$ RMSE, the yellow circles identify the held-out validation samples, and the green and red bars indicate residuals within and beyond  $\pm$ RMSE, respectively.

The input point clouds were assessed in an application-realistic near-field setting. Pheno4D reports high-accuracy laser-scanned point clouds, with scanner measuring accuracy of  $\sigma = 0.012$  mm and measuring-arm tip accuracy of  $45 \mu\text{m}$ ; its reports also note occlusion-related missing measurements and plant-surface reconstruction challenges [27]. Soybean-MVS uses 60 images per plant group from four height levels and assesses point-cloud consistency and segmentation usability [28]. The self-built COLMAP point clouds retained plant completeness and metric vertical structure while representing the occlusion, local missing regions, and reconstruction noise encountered in near-field agricultural photogrammetry. They therefore provide a practical input condition for testing organ parsing.

### 3.2. Multi-Dataset Semantic Segmentation Performance

Plant-GeoAT produced the largest mean mIoU on each evaluated dataset across five seeds:  $92.26 \pm 0.28\%$  on rapeseed,  $82.50 \pm 0.24\%$  on Soybean-MVS,  $99.74 \pm 0.05\%$  on Pheno4D-Maize, and  $94.75 \pm 0.15\%$  on Pheno4D-Tomato. The four datasets retained their native label spaces and were trained and tested independently; the results are not cross-dataset transfer tests. The equal-weight average of the four means was 92.31%, providing a

descriptive summary across these heterogeneous dataset-specific protocols. The complete comparison is reported in Table 4. The corresponding qualitative examples are shown in Figure 7.

Table 4. Quantitative comparison of semantic segmentation performance across crop datasets.

| Model              | Params (M) | FLOPs (G) | Rapeseed mIoU (%) | Soybean mIoU (%) | Maize mIoU (%) | Tomato mIoU (%) | Dataset-Specific Average mIoU (%) |
|--------------------|------------|-----------|-------------------|------------------|----------------|-----------------|-----------------------------------|
| PointNet++         | 0.97       | 1.00      | 77.09             | 50.87            | 67.10          | 70.45           | 66.38                             |
| DGCNN              | 1.80       | 2.40      | 87.48             | 69.47            | 93.82          | 88.79           | 84.89                             |
| PointTransformerV3 | 10.50      | 6.20      | 83.81             | 51.21            | 81.58          | 75.44           | 73.01                             |
| PointVector        | 3.20       | 3.80      | 77.37             | 51.68            | 63.42          | 86.27           | 69.69                             |
| PointLiBR          | 9.80       | 5.90      | 89.99             | 50.56            | 84.09          | 86.70           | 77.84                             |
| Plant-GeoAT        | 10.90      | 7.02      | 92.26             | 82.50            | 99.74          | 94.75           | 92.31                             |

Table note: Average mIoU summarises the four dataset-specific evaluations while retaining each native semantic label definition. Params and FLOPs provide a common description of model scale; runtime also depends on hardware, implementation, batch size, and I/O conditions.

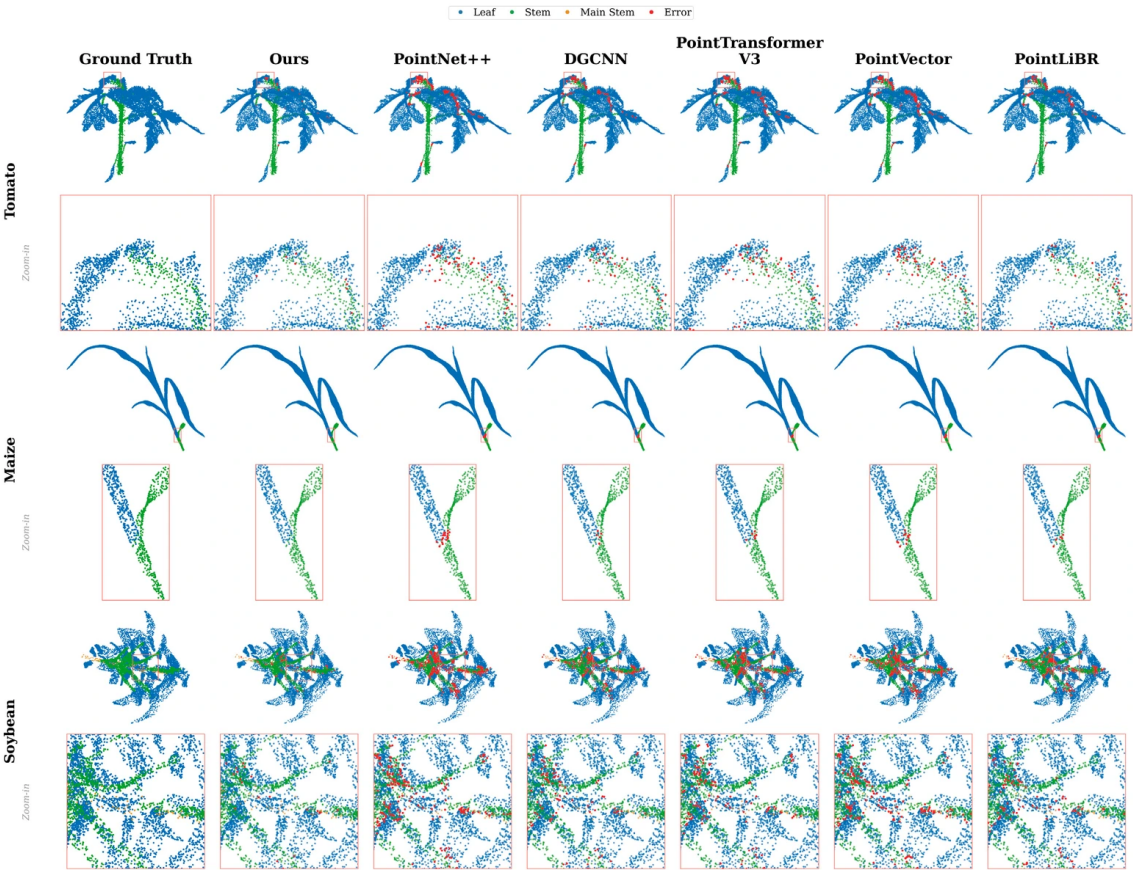


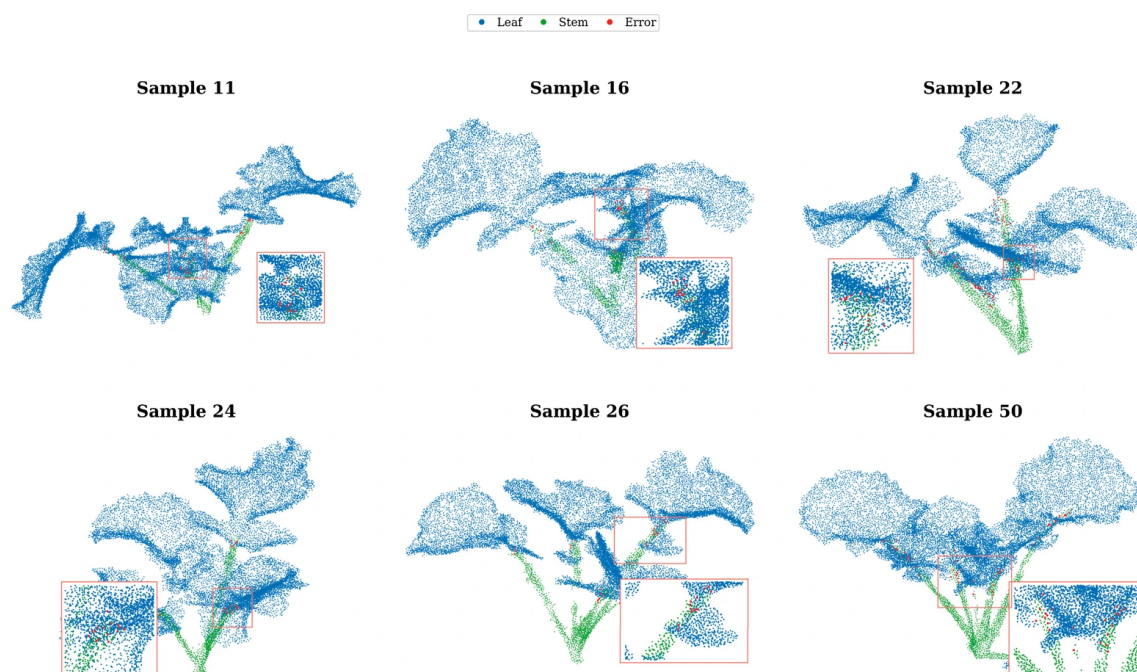
Figure 7. Qualitative semantic-segmentation results on public plant point-cloud datasets. Rows correspond to tomato, maize, and soybean scenes; columns compare Ground Truth, Ours, PointNet++, DGCNN, PointTransformer V3, PointVector, and PointLiBR. Blue, green, and red points denote leaf, stem, and prediction-error classes, respectively. The red boxes identify local regions shown as enlarged views in the second row of each crop block, enabling closer inspection of boundary leakage and missed stem/leaf points. Results follow the class definitions and annotation protocols of each public dataset. The identified soybean main-stem failures are examined separately in Supplementary Figure S1, and the class-wise tables provide the dataset-level quantitative results.

The four datasets span rosette-like rapeseed, branching soybean, elongated maize, and intersecting tomato structures, as well as COLMAP, image-based MVS, and laser-scanned point clouds. Plant-GeoAT retained the highest observed mIoU among the compared models in all four settings. This consistency links the LLGE-SSCA pathway to organ parsing across heterogeneous protocols, but it should not be interpreted as evidence that a model trained on one dataset will transfer without retraining. The near-perfect Pheno4D-Maize result is aided by high-quality laser scans, a binary Leaf/Stem label space, elongated organs, and a plant-level split with all dates from a biological plant kept together. The five training plants and two test plants were disjoint, and no plant-level or time-level information leakage occurred.

### 3.3. Class-Wise Organ Parsing and Stem Performance

Class-wise results are important because stems are often thin, partially occluded, and structurally central for plant architecture. Across five seeds, rapeseed Leaf IoU was  $99.28 \pm 0.16\%$  and Stem IoU was  $85.24 \pm 0.48\%$ . Maize Stem IoU was  $99.57 \pm 0.09\%$ , compared with DGCNN at 90.32% in the reference comparison, corresponding to a +9.25 percentage-point difference between the reported means. Tomato Stem IoU was  $91.68 \pm 0.36\%$ ; the best recovered reference baseline was DGCNN at 82.82%.

Stem IoU complements overall mIoU because stem continuity and leaf–stem attachment define the structural scaffold used in plant-architecture analysis. The rapeseed, maize, and tomato class-wise results show that Plant-GeoAT preserves this scaffold across rosette-like, elongated, and intersecting organ arrangements, which is the main biological value of the stem-oriented comparison. Representative self-built rapeseed predictions are shown in Figure 8.



**Figure 8.** Qualitative semantic-segmentation results on the self-built seedling-stage rapeseed dataset. Each column shows a representative sample (Samples 11, 16, 22, 24, 26, and 50), with the original point cloud above and the corresponding Plant-GeoAT prediction below. Blue, green, and red points denote leaf, stem, and segmentation-error points, respectively. The red boxes identify local leaf–stem boundary regions shown as enlarged views, enabling closer inspection of fine-scale segmentation errors under overlapping rosette-like foliage.

Soybean-MVS uses three semantic classes: MainStem, Stem, and Leaf. Plant-GeoAT reached 88.72% Leaf IoU, 70.42% MainStem IoU, and 88.36% Stem IoU. PointNet++ reached 83.35% MainStem IoU, identifying main-axis versus lateral-stem topology as the remaining soybean-specific challenge. We therefore performed a frozen-checkpoint diagnostic on all 13 held-out Area 2 scenes and retained two identified failure cases rather than selecting only visually successful plants (Supplementary Figure S1). In the moderate case (Area\_2\_20180817\_DN251), scene mIoU was 0.646 and MainStem IoU was 0.545; in the severe case (Area\_2\_20180908\_DN252), scene mIoU was 0.389 and MainStem IoU was 0.005 because most annotated main-stem points were assigned to Leaf or Stem. The error maps localise this failure along the principal axis and at branch junctions. This diagnostic is a failure analysis of the final checkpoint, not a replacement for the five-seed dataset-level results.

### 3.4. Ablation Analysis

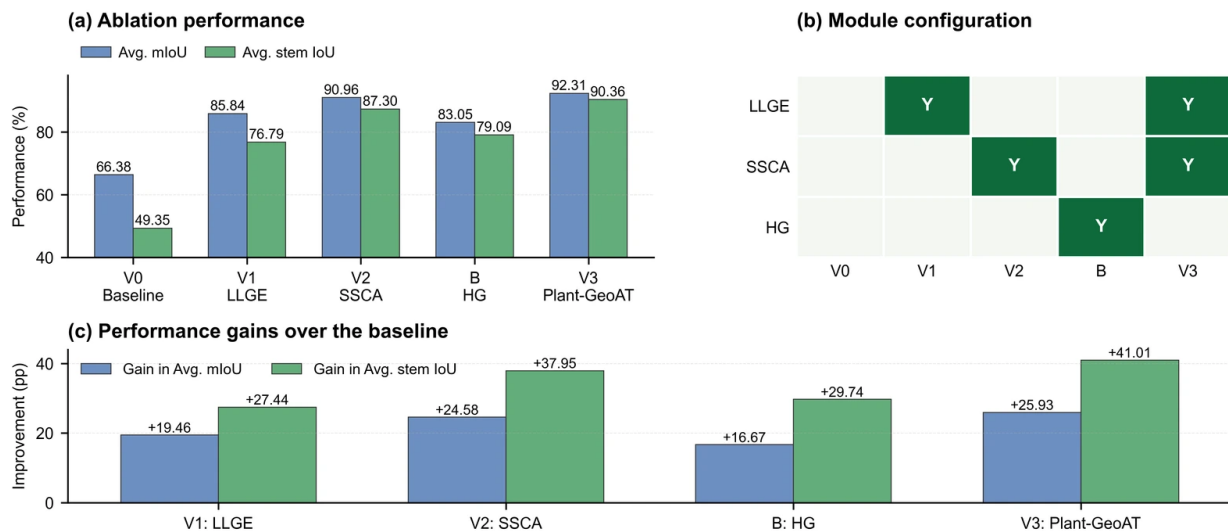
The ablation study evaluated how learnable local geometry and spatial-feature coupling contribute to segmentation performance (Table 5). V0 represents the baseline configuration, V1 adds LLGE, V2 adds SSCA, B uses handcrafted geometry, and V3 is the full Plant-GeoAT configuration.

**Table 5.** Ablation study of learnable local geometry extraction and spatial–semantic coupled attention.

| Variant | LLGE | SSCA | Handcrafted Geometry | Params (M) | FLOPs (G) | Dataset-Specific Avg. mIoU (%) | Dataset-Specific Avg. Stem IoU (%) |
|---------|------|------|----------------------|------------|-----------|--------------------------------|------------------------------------|
| V0      | No   | No   | No                   | 0.97       | 1.00      | 66.38                          | 49.35                              |
| V1      | Yes  | No   | No                   | 2.87       | 4.04      | 85.84                          | 76.79                              |
| V2      | No   | Yes  | No                   | 10.90      | 5.87      | 90.96                          | 87.30                              |
| B       | No   | No   | Yes                  | 10.88      | 5.79      | 83.05                          | 79.09                              |
| V3      | Yes  | Yes  | No                   | 10.90      | 7.02      | 92.31                          | 90.36                              |

Adding LLGE increased average mIoU from 66.38% to 85.84% and average Stem IoU from 49.35% to 76.79%. The SSCA-only variant reached 90.96% average mIoU and 87.30% average Stem IoU. These changes follow the proposed division of labour: LLGE contributes relative-XYZ shape cues, and SSCA contributes spatial-feature contextual coupling. Because V0 has 0.97 M parameters and V2/V3 has about 10.90 M, the V0 comparison measures the combined change in architecture and capacity.

The near-matched B, V2, and V3 variants provide the more specific module comparison. Their parameter counts are 10.88–10.90 M, while average mIoU was 83.05% for handcrafted geometry, 90.96% for SSCA alone, and 92.31% for the full LLGE + SSCA model. V2 and V3 have the same reported parameter scale; adding LLGE increased average mIoU by 1.35 percentage points and average Stem IoU from 87.30% to 90.36%. These within-scale results support learned local geometry and spatial-feature coupling as complementary parts of the geometry-to-context pathway. The ablation trends are visualised in Figure 9.



**Figure 9.** Ablation analysis of local-geometry extraction and spatial-feature contextual coupling. (a) Average mIoU and average stem IoU for the baseline (V0), LLGE-only (V1), SSCA-only (V2), handcrafted-geometry (B), and complete Plant-GeoAT (V3) configurations; blue and green bars denote average mIoU and average stem IoU, respectively, and the labels above the bars give the corresponding percentages. (b) Module-configuration matrix, where “Y” indicates that the corresponding module is enabled; dark-green cells denote enabled modules, while light-colored cells denote modules that are not enabled. LLGE, SSCA, and HG denote the Learnable Local Geometry Extractor, Spatial-Semantic Coupled Attention, and handcrafted geometry, respectively. (c) Improvements over the baseline in percentage points; blue and green bars denote gains in average mIoU and average stem IoU, respectively. The results show the separate and combined effects of learnable local geometry and spatial-feature coupling.

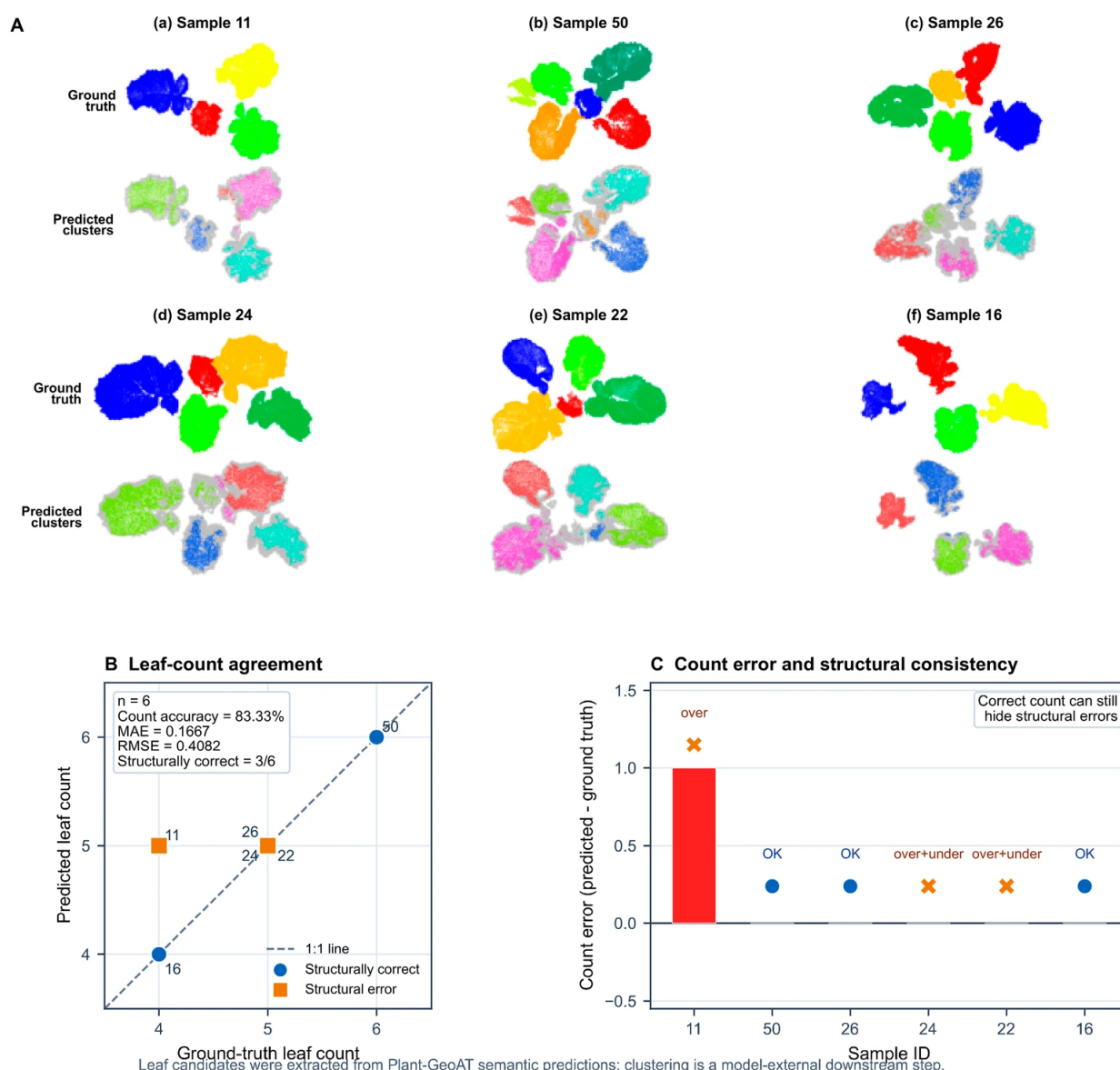
### 3.5. Semantic-Guided Leaf Counting

Across S1–S5, Plant-GeoAT predictions supplied leaf candidate points to the same conventional, model-external density/geometric clustering procedure on the fixed six-sample test set. The HDBSCAN and fragment-fusion parameters were fixed empirically before the five-seed comparison and held constant across runs rather than tuned on these six samples.

Correctness is reported as the mean number of structurally correct samples per run and its observed range. Precision, Recall, and F1-score are point-level leaf-candidate metrics. The counting evaluation was independent of checkpoint selection; however, it remains exploratory because it contains only six test plants and lacks external instance-labelled replication.

Across the five seed runs, mean Count Accuracy was  $86.67 \pm 7.45\%$ , MAE was  $0.1333 \pm 0.0745$ , and RMSE was  $0.3266 \pm 0.1826$ . S1 and S3–S5 each achieved 83.33% Count Accuracy, whereas S2 achieved 100.00%. Structural correctness averaged 3.20/6 samples and ranged from 3/6 to 4/6. Precision, Recall, and F1-score were  $0.9625 \pm 0.0006$ ,  $0.9747 \pm 0.0008$ , and  $0.9686 \pm 0.0012$ , respectively; these are point-level leaf-candidate metrics rather than cluster-level instance-recovery scores. Because the separate validation split was used for checkpoint selection, while the six test samples were reserved for counting evaluation, this result remains exploratory; an independent instance-labelled set is required for trait-level validation. Representative S3 cluster assignments and structural errors are shown in Figure 10.

# Semantic-guided leaf clustering and counting analysis



**Figure 10.** Semantic-guided leaf clustering and counting validation for reference run S3. (A) Ground-truth leaf instances and predicted clusters for Samples 11, 50, 26, 24, 22, and 16. Different colors identify individual ground-truth leaf instances or predicted clusters, while gray points indicate unlabeled or background points. (B) Agreement between ground-truth and predicted leaf counts. Blue circles denote structurally correct samples, orange squares denote samples with structural errors, and the dashed line indicates 1:1 count agreement. (C) Sample-level count error and structural consistency. The red bar indicates a positive count error, blue circles denote structurally correct samples, and orange crosses denote samples containing over-segmentation, under-segmentation, or both. Samples 24 and 22 retain the correct total leaf count but contain one over-segmentation and one under-segmentation event, demonstrating that count agreement can conceal structural errors. Table 6 reports the corresponding five-seed summary across S1–S5.

**Table 6.** Five-seed summary of semantic-guided leaf-counting analysis.

| Metric Category               | Metric                       | Result                  |
|-------------------------------|------------------------------|-------------------------|
| Counting accuracy             | Count Accuracy (%)           | 86.67 ± 7.45            |
| Counting error                | MAE                          | 0.1333 ± 0.0745         |
| Counting error                | RMSE                         | 0.3266 ± 0.1826         |
| Leaf-candidate classification | Precision                    | 0.9625 ± 0.0006         |
| Leaf-candidate classification | Recall                       | 0.9747 ± 0.0008         |
| Leaf-candidate classification | F1-score                     | 0.9686 ± 0.0012         |
| Structural consistency        | Structurally correct samples | 3.20/6 (range: 3/6–4/6) |

Table note: Results are mean ± sample SD across five seed runs (S1–S5;  $n = 5$ ) on the fixed six-sample test set. Leaf candidate points were extracted from Plant-GeoAT predictions and clustered using a conventional, model-external density/geometric procedure.

3.6. Application-Oriented Analyses for Practical 3D Phenotyping

These analyses form a connected 3D phenotyping workflow. Reconstruction checks establish input-cloud quality, semantic segmentation converts the point clouds into leaf/stem units, and semantic-guided clustering explores one downstream use of those units. Table 7 states the sample set, dependency, and contribution of each analysis.

**Table 7.** Analysis roles of semantic segmentation, leaf counting, and reconstruction/preprocessing quality assessment.

| Analysis                             | Samples                                                                      | Depends on Plant-GeoAT Semantic Labels? | Main Purpose                                             | Analysis Role                                                     |
|--------------------------------------|------------------------------------------------------------------------------|-----------------------------------------|----------------------------------------------------------|-------------------------------------------------------------------|
| Semantic segmentation                | Dataset-specific test sets                                                   | Yes                                     | Organ identity parsing                                   | Primary model performance under dataset-specific protocols        |
| Leaf counting                        | Six-sample test set; separate validation split used for checkpoint selection | Yes                                     | Semantic-guided counting feasibility                     | Tests transfer of semantic leaf points to conventional clustering |
| Reconstruction/preprocessing quality | All 60 reconstructed seedling-stage rapeseed samples                         | No                                      | Input-cloud completeness and vertical metric consistency | Establishes reconstruction and preprocessing quality              |

The full model has 10.90 M parameters and 7.02 G FLOPs. The checked archived training logs recorded a mean wall-clock duration of approximately 25 h 29 min 28 s for 240 completed epochs before early stopping (approximately 6.37 min per completed epoch); this is a training-time reference separate from data preparation and inference profiling. We profiled the frozen rapeseed checkpoint on all six held-out 20,000-point test scenes using one NVIDIA A100-SXM4-40GB GPU, PyTorch 2.3.1, CUDA 12.1, FP32, and batch size 1. After three warm-up passes, mean network-forward time was 0.260 s per complete scene (median 0.260 s). CPU-based KNN neighbourhood construction required 0.221 s per scene on average, giving 0.480 s for neighbourhood construction plus network forward, excluding disc I/O. The combined value was calculated from the per-scene total time rather than by adding the separately rounded component means. Peak CUDA memory was 1226 MiB allocated and 1626 MiB reserved. A 100-server verification recorded the

reported inference timing and GPU-memory usage under the documented hardware and software configuration.

## 4. Discussion

### 4.1. Reconstructed Point Clouds Need an Organ-Identity Layer

Multi-view reconstruction provides plant geometry, and Plant-GeoAT adds the organ identity required for structural analysis. By converting XYZRGB points into leaf/stem units, the model supplies the biological organisation needed to measure leaf number, stem continuity, and organ arrangement. Stem continuity matters because the stem is the support and transport axis linking leaves; organ identity is needed to separate leaf area or count from stem biomass and to define attachment neighbourhoods. The present model can support leaf number, leaf-area proxies, stem length/continuity, and leaf–stem spatial relationships, but it does not infer organ hierarchy or developmental connectivity.

The semantic-segmentation results directly test this contribution. Dataset-specific mIoU and class-wise Stem IoU measure organ parsing, while the reconstruction review verifies input-cloud quality and the exploratory counting analysis demonstrates one downstream use of predicted leaf points. These analyses connect acquisition, semantic interpretation, and downstream phenotyping without treating segmentation as independent trait validation.

Public image-based and laser-scanned datasets extend model validation beyond the self-built acquisition route. The self-built field-grown rapeseed data then place the model in a low-cost near-field workflow that includes conventional photogrammetric occlusion, local missing regions, and reconstruction noise. Plant completeness, vertical consistency, and organ parsing under these conditions show that the semantic layer remains usable with practical image-based inputs [27,28].

The closest Agronomy studies clarify the boundary of the present contribution. Sun et al. [19] coupled rapeseed stem–leaf processing with direct morphological-trait validation, whereas Zhang et al. [20] evaluated rule-based leaf segmentation across rapeseed stages. Zhu and Yu [21] used octree voxelization and region growing on Pheno4D corn; because its split, label definition, and metrics are not identical to the protocol used here, the reported values are for context rather than a direct numerical benchmark. Plant-GeoAT's contribution is therefore not the first use of 3D point clouds for rapeseed or stem–leaf analysis, but the learned relative-XYZ-before-colour and spatial-feature coupling evaluated separately across four datasets. Occlusion-aware completion [22] remains a complementary next step for improving structural recovery before trait validation.

### 4.2. LLGE for Local Geometry Encoding

LLGE extracts plant-oriented local geometry from KNN relative-XYZ vectors using shared Conv1d layers and neighbourhood max pooling. RGB is concatenated only after this aggregation. The ordering preserves shape changes around slender stems, leaf bases, and junctions before colour and broader context enter the representation.

The V1 result follows this mechanism: adding LLGE raised average mIoU by 19.46 percentage points and average Stem IoU by 27.44 points over V0. The comparison combines module and capacity changes, while the equal-scale V2-to-V3 comparison isolates an additional 1.35-point average-mIoU gain when learned local geometry is added to SSCA. Both comparisons connect the observed improvement to the intended first stage of the geometry-to-context pathway.

#### 4.3. SSCA as Spatial-Feature Contextual Coupling

SSCA supplies the contextual stage of the pathway. Its spatial branch models radius-defined physical neighbourhoods, and its feature-space branch models EdgeConv relations among learned responses. Their fusion lets the network compare local shape with both organ arrangement and feature similarity.

#### 4.4. Comparison with Related Point-Cloud Models

Plant-GeoAT's stem-focused gains are consistent with the proposed ordering in which LLGE encodes relative-XYZ evidence before RGB fusion and SSCA then couples spatial and feature-space context at stems and organ junctions. The near-matched B, V2, and V3 ablations provide the clearest mechanism evidence while keeping parameter counts close. Baseline implementation provenance, native losses, sampling choices, and checkpoint rules are documented in Supplementary Table S11.

This coupling is useful where leaves overlap or touch stems: Euclidean adjacency preserves physical organisation, while feature similarity links points with related learned responses. The SSCA-only variant reached 90.96% average mIoU, and the full model reached 92.31% after LLGE was added. The result supports contextual coupling as the second stage of the proposed representation.

#### 4.5. From Semantic Parsing to Phenotyping

Semantic parsing enables phenotyping by defining candidate organ points for downstream algorithms. In the exploratory counting analysis, a simple geometric clustering step converted predicted leaf points into counts, while structural correctness remained lower than numerical count agreement. This separation shows that semantic leaf candidates can enter a conventional workflow, but one-to-one instance recovery requires an independent instance-labelled evaluation.

The reconstruction analysis establishes the quality of the point clouds entering this pipeline. Manual review checks plant completeness and organ visibility, while height on all 60 pose-aligned, scale-recovered samples measures vertical metric consistency. These input-quality checks complement segmentation mIoU, class-wise IoU, ablation, and downstream counting, giving each stage a distinct measurable responsibility.

#### 4.6. Scope and Next Validation

The self-built dataset contains 60 original seedling-stage rapeseed point clouds, with 42 training, 12 validation, and six test samples. Offline augmentation broadens geometric variation within the training set. The five-seed analysis quantifies run-to-run variation under the stated fixed split and training configuration, while additional acquisition sessions, growth stages, and independent biological samples remain necessary to quantify trait-level reliability beyond this seedling-stage setting.

The Supplementary Materials report baseline configurations, recovered implementation details, and documented split deviations. The comparison therefore represents the stated implementations under a common dataset-specific protocol. Releasing these settings makes the benchmark reproducible and identifies implementation harmonisation as a useful target for future comparisons.

Model capacity is the main constraint on interpreting the V0 comparison: V0 has 0.97 M parameters, while V2 and V3 have about 10.90 M. The similar-scale B/V2/V3 comparisons and the equal-scale V2-to-V3 comparison provide the clearest module evidence. A capacity-matched LLGE/SSCA study would further separate representation design from parameter count.

The multi-dataset experiment spans COLMAP-based rapeseed, an image-based soybean dataset, and laser-scanned Pheno4D maize and tomato, covering different crop structures, acquisition routes, artefacts, and label protocols. Each dataset was trained and evaluated within its native label space, so the four-dataset average is a descriptive equal-weight summary rather than a cross-dataset transfer score. Soybean MainStem IoU identifies topology-aware branch-order modelling as the remaining challenge, whereas the Pheno4D-Maize result is aided by clean binary labels and a plant/time-consistent split.

Semantic-guided counting was an exploratory analysis on the same six-sample test set; checkpoint selection used the separate validation split. The five-seed summary quantifies run-to-run counting variation, but checkpoint selection used the separate validation split, while downstream evaluation used the six test samples. A larger independent instance-labelled set is therefore needed to estimate structural recovery reliability at scale.

Plant height was computed independently from semantic labels on all 60 denoised, pose-aligned, and scale-recovered rapeseed point clouds. It is not a Plant-GeoAT result: it quantifies the reconstruction and preprocessing pipeline only. Together with manual completeness checks, the height metrics quantify the input quality of the reconstruction and preprocessing workflow.

The fixed-hardware profile now provides an offline A100 reference for complete-scene inference. The next operational validation should repeat profiling on deployment-class hardware while testing additional growth stages, acquisition environments, and repeated batches. These experiments would extend the present organ-parsing results from controlled near-field inputs towards open-field phenotyping.

## 5. Conclusions

Plant-GeoAT converts reconstructed XYZRGB point clouds into leaf/stem semantic units through a geometry-to-context pathway. LLGE encodes relative-XYZ neighbourhoods before RGB fusion, and SSCA couples spatial neighbourhoods with feature-space relations. This ordering links local organ shape to the context needed for dense point-wise identity prediction and provides an explicit organ-identity layer after reconstruction.

Across five seeds on self-built rapeseed, Soybean-MVS, Pheno4D-Maize, and Pheno4D-Tomato, Plant-GeoAT achieved the strongest reported dataset-specific mIoU in the compared protocols, with a descriptive four-dataset average of 92.31%. Near-matched ablations support complementary roles for learned local geometry and contextual coupling, while the capacity difference between V0 and the full model limits claims from the smallest baseline comparison.

On the self-built rapeseed data, all-sample height consistency established the metric quality of the reconstructed inputs independently of semantic labels. Exploratory semantic-guided clustering showed that predicted leaf points can enter a conventional phenotyping operation, but the small six-sample test set and incomplete instance recovery prevent a claim of independent trait-validation accuracy. The main limitations are the absence of cross-dataset transfer tests, capacity-matched ablations, deployment-hardware comparisons, and broader growth-stage and biological replication. Future work should add independent instance-labelled datasets, explicit organ-connection or hierarchy annotations, capacity-matched baselines, cross-platform profiling, and open-field multi-stage evaluation.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy16181809/s1>, Figure S1: Representative soybean failure cases on held-out Area 2 scenes. For each scene, the panels show the ground truth, Plant-GeoAT prediction, error localization, and a main-stem failure zoom. Blue, green, and orange points denote leaf, lateral-stem, and main-stem classes, respectively; red points mark other errors, magenta points

mark missed main-stem points, and the zoom outlines the ground-truth main stem. Table S1: Overall ablation results with model complexity across crop datasets; Table S2: Class-wise IoU of ablation variants across crop datasets; Table S3: Overall segmentation performance and model complexity of comparison methods; Table S4: Full class-wise segmentation metrics on Rapeseed and Soybean-MVS; Table S5: Full class-wise segmentation metrics on Maize and Tomato; Table S6: Stem IoU improvement over the best baseline; Table S7: Exploratory leaf-counting and structural-consistency results on the six-sample test set; Table S8: Feature design for semantic-guided leaf counting; Table S9: Clustering parameters for semantic-guided leaf counting; Table S10: Fragment attachment and merge rules; Table S11: Baseline configuration and dataset protocol reporting; Table S12: Definitions used in structural-correctness review; Table S13: Public-dataset partition, augmentation, and source reporting; Table S14: Plant-GeoAT multi-seed dataset-level mIoU; Table S15: Multi-seed dataset-level mIoU of all compared models; Table S16: Plant-GeoAT multi-seed class-wise IoU; Table S17: Multi-seed ablation average mIoU; Table S18: Multi-seed ablation average stem IoU; Table S19: Multi-seed ablation per-dataset mIoU; Table S20: Multi-seed ablation per-dataset stem IoU; Table S21: Multi-seed predicted leaf counts per test sample; Table S22: Multi-seed leaf-counting summary metrics; Table S23: Plant-height validation metrics and multi-seed behavior; Method S1: HDBSCAN parameter design for semantic-guided leaf counting; Method S2: Baseline configuration reporting; Method S3: Structural-correctness protocol for exploratory leaf-counting analysis; Method S4: Public-dataset partition and augmentation reporting; Method S5: Self-built rapeseed annotation protocol; Method S6: Plant-GeoAT implementation details.

**Author Contributions:** J.W.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualisation, Writing—original draft. B.Z.: Conceptualization, Methodology, Supervision, Project administration, Writing—review and editing. J.L. (Jie Liu): Investigation, Resources, Writing—review and editing. Z.X.: Methodology, Investigation, Data curation, Writing—review and editing. J.L. (Jie Luo): Investigation, Resources, Data curation. J.L. (Jiajun Liu): Investigation, Data curation. J.Z.: Investigation, Data curation. M.L. (Mingju Li): Methodology, Formal analysis, Data curation, Writing—review and editing. Y.A.: Investigation, Data curation. M.L. (Mingwei Liu): Investigation, Data curation. Y.C.: Investigation, Data curation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project is supported by the National Natural Science Foundation of China, grant number 32301762.

**Data Availability Statement:** The self-built rapeseed point-cloud data generated in this study, including reconstructed XYZRGB point clouds, semantic annotations, fixed train/validation/test split files, and processed 20,000-point samples, are not publicly deposited at initial submission because the dataset contains unpublished self-built annotation files and requires institutional approval for research-use sharing. The data are available from the corresponding author (Bei Zhou; 12801@sicau.edu.cn) upon reasonable e-mail request and approval for non-commercial research use. Public datasets used in this study, including Soybean-MVS and Pheno4D-Maize/Pheno4D-Tomato, should be obtained from their original providers and publications; this article does not redistribute third-party datasets. The processed summary metrics required to evaluate the reported results are included in the main article and Supplementary Material. The reviewed Plant-GeoAT research-preview code is available at the immutable historical commit <https://github.com/kimevans111/Plant-GeoAT/commit/e0aba5b94e26ba4306d0f7f9901185883fe8d4c4> (accessed on 30 August 2026). This snapshot is distinct from the repository's current default branch. It contains the model implementation, training and evaluation scripts, data-format and release documentation, command templates, citation metadata, and licence/notice files. Datasets, trained checkpoints, private logs, server scripts, and local machine paths are not included.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

| Symbol or Abbreviation | Definition                                                 | Unit             |
|------------------------|------------------------------------------------------------|------------------|
| 2D                     | two-dimensional                                            | -                |
| 3D                     | three-dimensional                                          | -                |
| COLMAP                 | structure-from-motion and multi-view stereo software       | -                |
| Conv1d                 | one-dimensional convolution                                | -                |
| DGCNN                  | dynamic graph convolutional neural network                 | -                |
| EdgeConv               | edge convolution                                           | -                |
| F1                     | harmonic mean of precision and recall                      | -                |
| FN                     | false negatives                                            | count            |
| FP                     | false positives                                            | count            |
| FPS                    | farthest point sampling                                    | -                |
| IoU                    | intersection over union                                    | -                |
| KNN                    | k-nearest neighbours                                       | -                |
| LLGE                   | learnable local geometry extractor                         | -                |
| MAE                    | mean absolute error                                        | metric-dependent |
| MAPE                   | mean absolute percentage error                             | %                |
| mIoU                   | mean intersection over union                               | -                |
| MVS                    | multi-view stereo                                          | -                |
| RGB                    | red-green-blue colour channels                             | -                |
| RMSE                   | root mean square error                                     | metric-dependent |
| SSCA                   | spatial-semantic coupled attention                         | -                |
| TP                     | true positives                                             | count            |
| XYZRGB                 | 3D coordinates with red-green-blue colour                  | -                |
| Roman symbols          |                                                            |                  |
| $B_{r(i)}$             | radius-based neighbourhood of point i                      | -                |
| $c_i$                  | RGB colour vector of point i                               | -                |
| $C$                    | set of detected leaf clusters                              | -                |
| $D$                    | dataset identifier                                         | -                |
| $E_i^s$                | feature-space branch output at encoder stage s             | -                |
| $G$                    | semantic-guided clustering function                        | -                |
| $g_{ij}$               | encoded relative-geometry feature                          | -                |
| $H$                    | plant height                                               | cm               |
| $H_t, \hat{H}_t$       | measured and estimated plant heights                       | cm               |
| $H_{pot}^{real}$       | measured physical pot height                               | mm               |
| $H_{pot}, m^{rec}$     | reconstructed pot height for plant m before metric scaling | -                |
| $h_i^s$                | feature of point i at encoder stage s                      | -                |
| $K, k$                 | neighbour counts                                           | count            |
| $l_i$                  | local geometry descriptor of point i                       | -                |
| $N$                    | number of points in a point-cloud block                    | count            |
| $N_{K(i)}$             | KNN neighbourhood of point i                               | -                |
| $N_t, \hat{N}_t$       | observed and predicted leaf counts                         | count            |
| $P, P^*$               | input and scale-recovered point clouds                     | -                |
| $P_{leaf}$             | predicted leaf-candidate point set                         | -                |
| $p_i$                  | point i                                                    | -                |
| $s_m$                  | scale factor for plant m                                   | mm/model unit    |
| $S_i^s$                | spatial branch output at encoder stage s                   | -                |
| $T_s$                  | local point-transformer mapping at stage s                 | -                |
| $u_i^s$                | lifted feature at encoder stage s                          | -                |
| $W$                    | semantic prediction-head weight matrix                     | -                |
| $x_i$                  | XYZ coordinate vector of point i                           | -                |

|                    |                                                  |    |
|--------------------|--------------------------------------------------|----|
| $x_{i,m}^*$        | scale-recovered coordinate of point i in plant m | mm |
| $\hat{y}_i$        | predicted semantic label of point i              | -  |
| $Y_D$              | dataset-specific semantic label space            | -  |
| $z_{max}, z_{min}$ | maximum and minimum vertical coordinates         | mm |
| Greek symbols      |                                                  |    |
| $\Delta x_{ij}$    | relative-XYZ offset from point i to neighbour j  | -  |
| $\eta_s$           | EdgeConv-style mapping at stage s                | -  |
| $\varphi_{geo}$    | local geometry encoding function                 | -  |
| $\varphi_s$        | shared feature-lifting mapping at stage s        | -  |
| $\psi_s$           | spatial-feature fusion mapping at stage s        | -  |
| $\theta$           | trainable Plant-GeoAT parameters                 | -  |
| $\theta_g$         | semantic-guided clustering parameters            | -  |

## References

- Jin, S.; Li, D.; Yun, T.; Tang, J.; Wang, K.; Li, S.; Yang, H.; Yang, S.; Xu, S.; Cao, L.; et al. Deep learning for three-dimensional (3D) plant phenomics. *Plant Phenomics* **2025**, *7*, 100107. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yang, W.; Feng, H.; Zhang, X.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L.; Yan, J. Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* **2020**, *13*, 187–214. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, D.; Quan, C.; Song, Z.; Li, X.; Yu, G.; Li, C.; Muhammad, A. High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Front. Bioeng. Biotechnol.* **2021**, *8*, 623705. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, Y.; Liang, Z.; Liu, B.; Yin, L.; Wan, F.; Qian, W.; Qiao, X. Applications of 3D reconstruction techniques in crop canopy phenotyping: A review. *Agronomy* **2025**, *15*, 2518. [\[CrossRef\]](#)
- Poorter, H.; Hummel, G.M.; Nagel, K.A.; Fiorani, F.; von Gillhaussen, P.; Virnich, O.; Schurr, U.; Postma, J.A.; van de Zedde, R.; Wiese-Klinkenberg, A. Pitfalls and potential of high-throughput plant phenotyping platforms. *Front. Plant Sci.* **2023**, *14*, 1233794. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gano, B.; Bhadra, S.; Vilbig, J.M.; Ahmed, N.; Sagan, V.; Shakoar, N. Drone-based imaging sensors, techniques, and applications in plant phenotyping for crop breeding: A comprehensive review. *Plant Phenome J.* **2024**, *7*, e20100. [\[CrossRef\]](#)
- Song, H.; Wen, W.; Wu, S.; Guo, X. Comprehensive review on 3D point cloud segmentation in plants. *Artif. Intell. Agric.* **2025**, *15*, 296–315. [\[CrossRef\]](#)
- Zhu, J.; Zhai, R.; Ren, H.; Xie, K.; Du, A.; He, X.; Cui, C.; Wang, Y.; Ye, J.; Wang, J.; et al. Crops3D: A diverse 3D crop dataset for realistic perception and segmentation toward agricultural applications. *Sci. Data* **2024**, *11*, 1438. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lou, L.; Liu, Y.; Han, J.; Doonan, J.H. Accurate multi-view stereo 3D reconstruction for cost-effective plant phenotyping. In *Image Analysis and Recognition*; Springer International Publishing: Cham, Switzerland, 2014; pp. 349–356. [\[CrossRef\]](#)
- Gong, L.; Du, X.; Zhu, K.; Lin, K.; Lou, Q.; Yuan, Z.; Huang, G.; Liu, C. Panicle-3D: Efficient phenotyping tool for precise semantic segmentation of rice panicle point cloud. *Plant Phenomics* **2021**, *2021*, 9838929. [\[CrossRef\]](#) [\[PubMed\]](#)
- Luo, L.; Jiang, X.; Yang, Y.; Samy, E.R.A.; Lefsrud, M.; Hoyos-Villegas, V.; Sun, S. Eff-3DPSeg: 3D organ-level plant shoot segmentation using annotation-efficient deep learning. *Plant Phenomics* **2023**, *5*, 0080. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, W.; Wu, S.; Wen, W.; Lu, X.; Liu, H.; Zhang, M.; Xiao, P.; Guo, X.; Zhao, C. Using high-throughput phenotype platform MVS-Pheno to reconstruct the 3D morphological structure of wheat. *AoB Plants* **2024**, *16*, plae019. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mirande, K.; Godin, C.; Tisserand, M.; Charlaix, J.; Besnard, F.; Hetroy-Wheeler, F. A graph-based approach for simultaneous semantic and instance segmentation of plant 3D point clouds. *Front. Plant Sci.* **2022**, *13*, 1012669. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, D.; Shi, G.; Li, J.; Chen, Y.; Zhang, S.; Xiang, S.; Jin, S. PlantNet: A dual-function point cloud segmentation network for multiple plant species. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 243–263. [\[CrossRef\]](#)
- Du, R.; Ma, Z.; Xie, P.; He, Y.; Cen, H. PST: Plant segmentation transformer for 3D point clouds of rapeseed plants at the podding stage. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 380–392. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zou, Y.; Wang, H.; Zhang, F.; Ge, Y.; Wang, W.; Chen, M. GCASSN: A graph convolutional attention synergistic segmentation network for 3D plant point cloud segmentation. *Front. Plant Sci.* **2025**, *16*, 1621934. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sun, B.; Zain, M.; Zhang, L.; Han, D.; Sun, C. Stem-leaf segmentation and morphological traits extraction in rapeseed seedlings using a three-dimensional point cloud. *Agronomy* **2025**, *15*, 276. [\[CrossRef\]](#)

20. Zhang, L.; Shi, S.; Zain, M.; Sun, B.; Han, D.; Sun, C. Evaluation of rapeseed leaf segmentation accuracy using binocular stereo vision 3D point clouds. *Agronomy* **2025**, *15*, 245. [[CrossRef](#)]
21. Zhu, Q.; Yu, M. A corn point cloud stem-leaf segmentation method based on octree voxelization and region growing. *Agronomy* **2025**, *15*, 740. [[CrossRef](#)]
22. Gao, Y.; Li, Z.; Liu, T.; Li, B.; Zhang, L. Point cloud completion of occluded corn with a 3D positional gated multilayer perceptron and prior shape encoder. *Agronomy* **2025**, *15*, 1155. [[CrossRef](#)]
23. Wu, X.; Jiang, L.; Wang, P.-S.; Liu, Z.; Liu, X.; Qiao, Y.; Ouyang, W.; He, T.; Zhao, H. Point transformer V3: Simpler, faster, stronger. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024; pp. 4840–4851.
24. Copernicus Climate Change Service (C3S). *ERA5 Hourly Data on Single Levels from 1940 to Present*; Copernicus Climate Change Service Climate Data Store: Reading, UK, 2023. [[CrossRef](#)]
25. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
26. Zippenfenig, P. *Open-Meteo.com Weather API*; Zenodo: Geneva, Switzerland, 2024. [[CrossRef](#)]
27. Schunck, D.; Magistri, F.; Rosu, R.A.; Cornelißen, A.; Chebrolu, N.; Paulus, S.; León, 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](#)]
28. Sun, Y.; Zhang, Z.; Sun, K.; Li, S.; Yu, J.; Miao, L.; Zhang, Z.; Li, Y.; Zhao, H.; Hu, Z.; et al. Soybean-MVS: Annotated three-dimensional model dataset of whole growth period soybeans for 3D plant organ segmentation. *Agriculture* **2023**, *13*, 1321. [[CrossRef](#)]
29. Schönberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113. [[CrossRef](#)]
30. Schönberger, J.L.; Zheng, E.; Pollefeys, M.; Frahm, J.M. Pixelwise view selection for unstructured multi-view stereo. In *Computer Vision—ECCV 2016*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9907, pp. 501–518. [[CrossRef](#)]
31. 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](#)]
32. Masuda, T. Leaf area estimation by semantic segmentation of point cloud of tomato plants. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 1381–1389. [[CrossRef](#)]
33. 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](#)]
34. Xie, K.; Zhu, J.; Ren, H.; Wang, Y.; Yang, W.; Chen, G.; Lin, C.; Zhai, R. Delving into the potential of deep learning algorithms for point cloud segmentation at organ level in plant phenotyping. *Remote Sens.* **2024**, *16*, 3290. [[CrossRef](#)]
35. Zhang, C.; Zhang, K.; Ge, L.; Zou, K.; Wang, S.; Zhang, J.; Li, W. A method for organs classification and fruit counting on pomegranate trees based on multi-features fusion and support vector machine by 3D point cloud. *Sci. Hortic.* **2021**, *278*, 109791. [[CrossRef](#)]
36. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85. [[CrossRef](#)]
37. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
38. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph.* **2019**, *38*, 1–12. [[CrossRef](#)]
39. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.H.S.; Koltun, V. Point transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 16259–16268. [[CrossRef](#)]
40. Deng, X.; Zhang, W.; Ding, Q.; Zhang, X. PointVector: A vector representation in point cloud analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 9455–9465.
41. Li, B.; Guo, C. MASPC\_Transform: A plant point cloud segmentation network based on multi-head attention separation and position code. *Sensors* **2022**, *22*, 9225. [[CrossRef](#)] [[PubMed](#)]
42. 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](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.