remote sensing

MDPI

Review

Delving into the Potential of Deep Learning Algorithms for Point Cloud Segmentation at Organ Level in Plant Phenotyping

Kai Xie [1], Jianzhong Zhu [1], He Ren [1], Yinghua Wang [2,3], Wanneng Yang [2,3], Gang Chen [4], Chengda Lin [5] and Ruifang Zhai [1,6,*]

[1] College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[2] National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China
[3] College of Plant Science, Huazhong Agricultural University, Wuhan 430070, China
[4] Department of Earth, Environmental and Geographical Sciences, University of North Carolina, Charlotte, NC 28223, USA
[5] College of Resource and Environment, Huazhong Agricultural University, Wuhan 430070, China
[6] Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China
* Correspondence: rfzhai@mail.hzau.edu.cn; Tel.: +86-27-87288509

Abstract: Three-dimensional point clouds, as an advanced imaging technique, enable researchers to capture plant traits more precisely and comprehensively. The task of plant segmentation is crucial in plant phenotyping, yet current methods face limitations in computational cost, accuracy, and high-throughput capabilities. Consequently, many researchers have adopted 3D point cloud technology for organ-level segmentation, extending beyond manual and 2D visual measurement methods. However, analyzing plant phenotypic traits using 3D point cloud technology is influenced by various factors such as data acquisition environment, sensors, research subjects, and model selection. Although the existing literature has summarized the application of this technology in plant phenotyping, there has been a lack of in-depth comparison and analysis at the algorithm model level. This paper evaluates the segmentation performance of various deep learning models on point clouds collected or generated under different scenarios. These methods include outdoor real planting scenarios and indoor controlled environments, employing both active and passive acquisition methods. Nine classical point cloud segmentation models were comprehensively evaluated: PointNet, PointNet++, PointMLP, DGCNN, PointCNN, PAConv, CurveNet, Point Transformer (PT), and Stratified Transformer (ST). The results indicate that ST achieved optimal performance across almost all environments and sensors, albeit at a significant computational cost. The transformer architecture for points has demonstrated considerable advantages over traditional feature extractors by accommodating features over longer ranges. Additionally, PAConv constructs weight matrices in a data-driven manner, enabling better adaptation to various scales of plant organs. Finally, a thorough analysis and discussion of the models were conducted from multiple perspectives, including model construction, data collection environments, and platforms.

Keywords: organ segmentation; plant phenotyping; deep learning; plant point cloud datasets

check for updates

Citation: 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. https://doi.org/10.3390/rs16173290

Academic Editor: Jochem Verrelst

Received: 22 June 2024
Revised: 2 September 2024
Accepted: 2 September 2024
Published: 4 September 2024

Copyright: © 2024 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) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The automatic phenotypic analysis of plants based on computer vision technology has become a crucial component in facilitating high-throughput experimentation in both botany and agricultural research. Compared to traditional manual measurement methods, computer vision-based phenotyping methodologies offer advantages in reducing manual labor, promoting non-invasiveness, and efficiently managing large volumes of data [1]. As a result, these methodologies have been widely recognized and applied over the past several years. This morphological analysis technique can be divided into 2D image-based

phenotyping methods and 3D point cloud-based phenotyping, depending on the data dimensionality used. While 2D phenotyping technology has been actively pursued by many researchers due to its convenient data acquisition and high-throughput capabilities, the persisting challenge of plant occlusion, primarily due to the lack of spatial structural information within 2D images, remains. To overcome these challenges, the research community has turned to three-dimensional data processing techniques in phenotypic studies, with a notable focus on employing 3D point cloud data for detailed phenotypic analysis [2–5]. Unlike 2D imaging, 3D imaging technologies capture and convey spatial information inherently in three dimensions. These methodologies have the potential to significantly reduce information loss from data occlusion and overlap, offering a promising outlook for phenotypic data acquisition. Consequently, the integration of 3D vision technologies into plant phenotyping research has become more prevalent, enabling more detailed and comprehensive phenotypic evaluations.

Methods for obtaining plant point clouds can be systematically categorized as either active or passive. Active measurement techniques involve the direct manipulation of sensors to capture the 3D structure of plants [6]. Common 3D sensor measurement techniques include laser triangulation [7], time of flight [8], terrestrial laser scanning (TLS) [9], structured light [10], and tomographic methods [11,12]. The application of UAV laser scanning (UAV-LS) technology in the agricultural field has steadily increased with the recent advancements in UAV technology [13]. In addition, full-waveform LiDAR data can store the entire echo scattered by an illuminated object with different temporal resolutions [14–19]. In [20], the DJI Zenmuse L1 multi-temporal point cloud was utilized to simulate ULS, and the vertical stratification of various crops was evaluated. Passive techniques, on the other hand, principally revolve around reconstructing 3D point clouds from 2D images collected from various vantage points. Within the realm of passive measurement technologies, the primary theologies and algorithms include the Structure from Motion (SfM) technique [21,22], the Multi-View Stereo (MVS) technique [23], and space carving [24,25].

Discerning various attributes associated with plant phenotypes often necessitates an initial and accurate segmentation of the plant structures. Precisely segmenting plant organs is essential for high-throughput and high-precision phenotype analysis. However, due to the intrinsic complexity of plant structures and the morphological variability that occurs throughout the plant life cycle, achieving precise segmentation is challenging. Over the last decade, numerous scientists have utilized machine learning algorithms and constructed an array of models to intensively research plant organ segmentation. Standard practice involves pre-computing the local features of plant point clouds, including surface normal, curvature, smoothness [26–28], local covariance matrices [29], tensor [30], and point feature histograms [31–33]. The foundation of these methods lies in the noticeable heterogeneity in the morphological structures of different plant organs.

Alternatively, another potent strategy involves modeling the point cloud data to specific geometric structures, such as ellipsoids [34], tubular structures [23], cylinders [35], or rings [36]. These techniques find particular relevance for organs exhibiting significant differential morphological structures. Apart from these direct methodologies that capitalize on morphological differences, an efficacious approach involves the use of spectral clustering for aggregating points of the same organ by calculating inter-point similarities [37]; however, this approach often incurs considerable computational costs. The combination of 2D imagery and 3D point clouds has also demonstrated commendable results in some plant studies [38,39]. Mature plants often display organ overlap due to dense canopies, prompting noteworthy attempts to adopt skeletonization algorithms to extract the "skeleton" of the plant to simplify its morphological structural complexity [40]. Despite the success of these methodologies in some studies, they frequently necessitate the pre-computation of complex handcrafted features and are typically relevant only to specific morphological structures or closely related plant species, thereby limiting their widespread applicability.

In contrast to traditional processing methods, deep learning approaches enable a data-driven feature extraction from the input data. These techniques eliminate the constraints of

prior plant knowledge and facilitate organ segmentation of crop objects at varying growth stages without the need for intricate parameter tuning. Over recent years, a multitude of point cloud processing solutions based on deep learning have been broadly applied within the plant phenotyping domain. According to the diverse data input methods, these can be categorized into projection-based methods, voxel-based methods, and point-based methods. Projection-based methods involve the projection of 3D point clouds onto 2D planes to form images, subsequently analyzing these 2D images utilizing Convolutional Neural Networks (CNNs), and eventually mapping the segmentation results back into the 3D point space [41]. Some strategies establish correlations between 2D images and 3D point clouds via projection to enhance understanding of point clouds [42,43]. However, these methods can be impeded by projection angles and occlusions, potentially leading to a loss of crucial structural information.

An alternative approach involves mapping points into a 3D voxel grid. The voxelized data are then processed by 3D Convolutional Neural Networks (CNNs). For example, Jin et al. presented a voxel-based Convolutional Neural Network (VCNN), specifically designed for the semantic segmentation of maize stems and leaves across various growth stages [44]. To address potential information loss during the mapping process, some researchers have proposed the concept of dynamic voxels [45], applying it to plant point cloud segmentation [46]. However, even with this approach, some point cloud information may be lost, and the size of the voxel significantly impacts network performance. Larger voxels often lose more detailed information, while smaller voxels usually lead to considerable computational costs.

Point-based methods, on the other hand, extract features directly from individual points for segmentation. Outstanding advancements in point cloud data processing have been marked by frameworks such as PointNet [47] and PointNet++ [48], which both enable end-to-end learning. These frameworks can learn features directly from raw point cloud data without manual intervention. The algorithms PointNet and PointNet++ have seen substantial application in the field of plant phenotyping [49–53], with a host of scholars refining them further to enhance their performance. Main directions for improvement include adding local feature extraction modules [54], constructing point connection graphs [55,56], applying the method of kernel point convolution [57,58], implementing down-sampling strategies [56,59,60], increasing residual layers [61], and promoting multimodal feature fusion [62,63]. Despite these advancements, these methods still utilize traditional featurizers that allow extensive progress in local feature extraction but fail to consider information from further points, thus hampering performance. In response, Xiang et al. [64] proposed the CurveNet model, a model pivoting on curves, which continually refines its trajectory to connect with more distant points and, as such, has demonstrated remarkable results.

Acknowledging the escalating importance of attention mechanisms, certain researchers have adopted these strategies to foster enhanced interactions within point cloud data. For instance, Li et al. advocated for the PSegNet approach [59], which capitalized on channel and spatial attention to amplify the model's extraction capabilities. Furthermore, the advent of self-attention mechanisms and transformers has paved the way for exhilarating opportunities in the realm of point cloud data processing. Transformers inherently capture long-range relations via self-attention mechanisms [65]. Models such as the Point Transformer (PT) [66], Voxel Transformer (VT) [67], and Stratified Transformer (ST) [68] have demonstrated exceptional performance on numerous datasets, thereby drawing considerable interest in plant phenotyping research.

Moreover, significant strides have been made by researchers such as Guo et al., who integrated an Adaptive Self-Attention module into the PointNet framework, resulting in substantial improvements in performance [69]. Compellingly, some scholars have employed novel position encoding strategies [70] and stratified feature extraction architectures [71] in the classic transformer framework to fortify the model's attribute extraction ability. Undoubtedly, the emergence of self-attention mechanisms and transformer frameworks has significantly amplified the performance of point cloud segmentation models.

However, this enhancement also incurs an appreciable computational cost. Moreover, models stemming from the transformer framework usually necessitate a substantial volume of training data. To address this predicament, certain researchers have gravitated towards weakly supervised segmentation techniques such as Eff-3DPSeg [72]. Not only do these weakly supervised approaches reduce the reliance on extensive data, but they also provide fresh perspectives to enhance network segmentation performance.

Despite the relatively comprehensive application of various point cloud segmentation algorithms in the realm of plant phenotyping, there are numerous issues necessitating further exploration. The intricate morphological structure of plants poses significant challenges; while some models demonstrate commendable segmentation performance on simpler plant entities, they struggle with segmenting complex mature crops or smaller organ categories. This predicament significantly impedes the progression of plant phenomics research. Therefore, whether these novel network models are suited to a variety of crop structures remains a worthwhile topic for exploration. Additionally, the majority of studies on plant organ segmentation methods are conducted on crop objects in controlled environments, with few extending to field environments. In plant cultivation and crop breeding, the ultimate aim of phenotyping is to promote vigorous plant growth in fields. Thus, technologies suitable for field phenotyping are needed. Finally, the quality of point clouds generated by various point cloud acquisition techniques differs; one cannot simply test model performance on point clouds acquired from a single platform. To some extent, data disparities also influence model performance. Therefore, a comprehensive evaluation of models on point cloud datasets acquired from various platforms is necessary to meet the requirements of current practical applications.

In our view, it is essential to explore the application of some classical point cloud data processing models on point cloud datasets of various plant entities collected under diverse environments and on numerous platforms. By analyzing these methods, we hope to inspire the design of superior models and point cloud processing strategies and to promote the application of actual 3D point cloud technology in agricultural production.

2. Materials and Methods

2.1. Experimental Design

To thoroughly investigate deep learning methods for point cloud segmentation, a data acquisition experiment was designed to test the diversity across platforms and environments, in addition to the publicly available datasets. In this study, the segmentation performance of the model was sequentially testified on five types of crops: corn, cotton, tomato, potato, and rapeseed. These crops exhibit significantly different morphological structures, providing a solid basis for assessing the model's generalization ability. To evaluate the impact of different types of point clouds on model performance, three data collection methods were employed: TLS-based point cloud, UAV-LS-based point cloud, and image-generated point cloud.

This study was carried out on the crop phenotyping platform at Huazhong Agricultural University, Wuhan, China (30°28′29.47″N, 114°21′24.37″E), between September 2023 and August 2024. In 2023, point cloud data were gathered from mature cotton and corn crops in an open-field environment, as well as rapeseed during the seedling stage in an outdoor potted environment using a terrestrial laser scanner, FARO FocusS70 (Figure 1). Regarding UAV-based point clouds, the UAV RGB camera and UAV laser scanner of DJI Matrix 350 RTK were used to collect data from tasseling-stage maize grown in pots, yielding point clouds and RGB images (Figure 2a–f).

2.2. Point Cloud Data Collection and Generation

2.2.1. TLS-Based Point Cloud

The application of terrestrial laser scanning (TLS) technology in the agricultural field has a robust history with the capability to capture accurate three-dimensional models of plants [73,74]. Increasingly, it has been used for measuring plant canopy structure [75],

plant height [76], biomass [77], leaf area index [78], and vertical leaf area profile [79]. The present study implemented a terrestrial laser scanner, FARO FocusS70 (Figure 1a,b). The FARO FocusS70 is a short-range instrument that emits laser pulses at a wavelength of 1550 nm within a maximum field of view of 360° × 300°. Its maximum measurement range is 70 m with a ranging error of 1 mm within the 10–25 m range. Scanning was carried out during weather conditions with wind speeds below 5 m/s. The point clouds for three plant species, rapeseed and maize at two growth points, and cotton at a mature stage, were collected using TLS technology, as depicted in Figure 1.

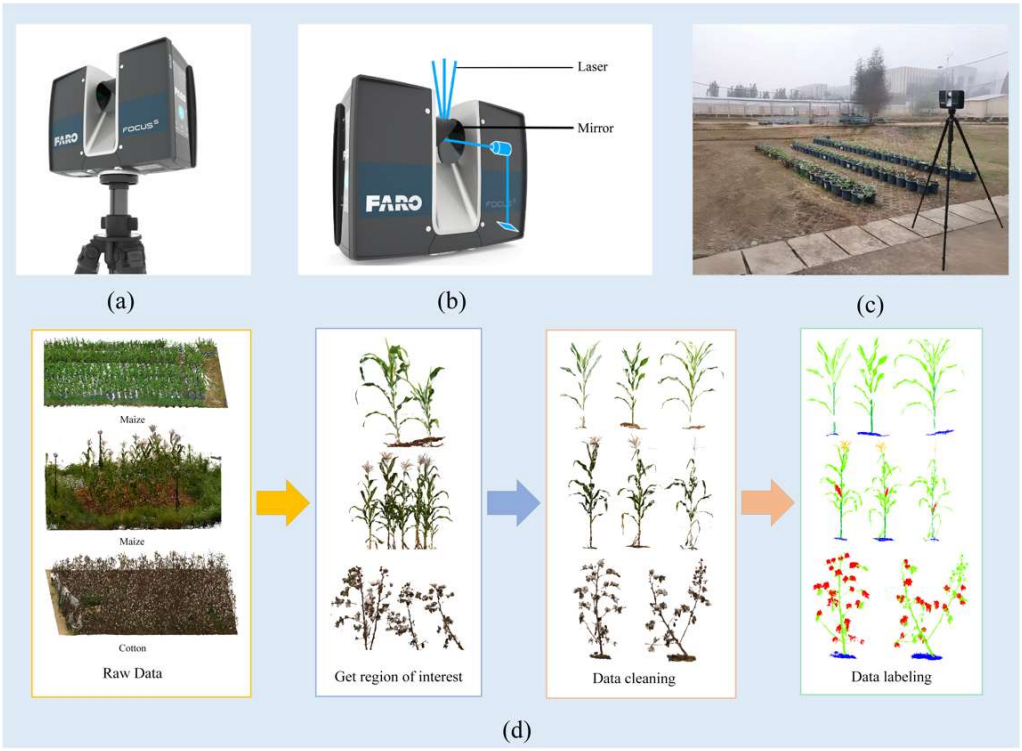

Figure 1. Data collection of terrestrial laser scanning. (**a**,**b**) FARO FocusS70 scanner. (**c**) Data collection scenario (rapeseed). (**d**) Point cloud data processing pipeline (maize and cotton).

The data collection scenario was as follows: The scanning device was mounted on a tripod approximately 1.8 m above ground level and one scan was carried out every 8 m, about 1.5 m away from the nearest plant, resulting in 15 TLS scans in total. Under potted conditions, five scans were conducted at matching height and distance for rapeseed (from the four corners and the center of the planting area, Figure 1c). During the scan, the ranges of vertical and horizontal scanning angles of the scanner were −60°~90° and 0°~360°, respectively. Owing to minor wind influence during our measurements, some plant replicas exhibited subtle shaking and shadowing. However, the overall plant morphology was not compromised. Each scanning required approximately 11 min to complete.

Owing to differing coordinate systems of the return TLS data obtained from the four positions, six target spheres were positioned at fixed locations. The FARO SCENE (version 2022.2.0.10355) software was used to collect and preprocess the TLS data, including denoising and point cloud registration. Subsequent steps included further partitioning, individual plant segmentation, and semantic annotation of the point cloud data, as exhibited in Figure 1d.

2.2.2. UAV-LS-Based Point Cloud

In comparison to TLS technology, UAV-LS technology is more suited for large-field scenarios, allowing for multi-scale, real-time, and close-range data collection of field crops [13]. This technology is extensively used for tasks such as crop biomass assessment [80], height

estimation [81], and canopy structure analysis [82,83]. We carried out data collection using the DJI Matrice 350 RTK (made by DJI, Shenzhen, China) (Figure 2d) equipped with the DJI Zenmuse L2 (made by DJI, Shenzhen, China) (Figure 2e) laser scanner. The DJI Zenmuse L2's ranging accuracy at a distance of 150 m is 2 cm, with a laser wavelength of 902 nm and a laser frequency of 204 kz. To capture complete structural information of plants, data collection was executed in weather conditions with wind speeds below 5 m/s. The UAV-LS operated in a non-repeat scanning mode, moving from north to south and east to west. The flight altitude was set at 12 m, with a runway overlap of 65%, a flight speed of 1.5 m/s, and a gimbal pitch angle of -60°. In this scenario, we collected UAV-LS-based point clouds of maize plants as depicted in Figure 2.

Figure 2 outlines the process of our UAV survey. To generate an authentic maize point cloud, preprocessing was necessary for the data collected by the UAV laser scanner. This involved three primary tasks: registration, stitching, and filtering of the point cloud. DJI Terra software was employed to handle the raw data collected, opting for the CGCS2000 coordinate system. Subsequently, the point cloud in LAS format was exported. To achieve individual plant segmentation in the field environment, LAS-formatted point clouds were divided into 2.5 m $\times$ 1.3 m sections using the CloudCompare (version 2.13.1) software, with each plot containing approximately 15 maize plants. Following Successive Over-Relaxation (SOR) filtering for each plot, these were manually annotated. The annotation work for the point cloud instance segmentation dataset involved two primary attributes: semantic attributes and instance attributes. The annotation process began by manually segmenting the plants and ground in each small area, followed by assigning these two types of scalar field properties to every isolated instance object. In semantic category division, the ground was labeled as 0, and plants were labeled as 1. Finally, a unique identifier starting from 0 was assigned to each instance to complete the instance segmentation annotation.

2.2.3. UAV Image-Based Point Cloud Generation

In addition to directly acquiring point clouds using the LiDAR system equipped on the UAV, we also reconstructed a 3D point cloud of the same area using images obtained from the high-resolution camera mounted on the UAV. Specifically, we used the DJI Matrice 350 RTK drone fitted with a SHAR 100M PRO optical camera (made by SHARE, Shenzhen, China) (Figure 2f). This RGB camera boasts a resolution of 11,648 $\times$ 8736 pixels and a 70 mm focal length. To obtain clear RGB images of the maize plants, we employed a five-direction flying method that included a nadir and four oblique photos. The flight altitude for capturing both nadir and oblique images was set at 25 m (Figure 2g). For nadir images, the longitudinal and lateral overlap was set to 80% and 70%, respectively, achieving a ground sampling distance (GSD) of 0.13 mm per pixel. The gimbal angle was set at 45°, with both longitudinal and lateral overlap at 70%, and a GSD of 0.19 mm per pixel.

Additionally, Metashape (version 2.10) software was utilized to generate 3D point clouds from these 2D images and to export these point clouds for analysis of maize plants. To assess the potential suitability of UAV-derived point clouds for segmentation tasks, the point clouds were denoised using CloudCompare (version 2.13.1) software. Semantic annotation was then manually performed on the structurally intact maize point clouds at the organ level.

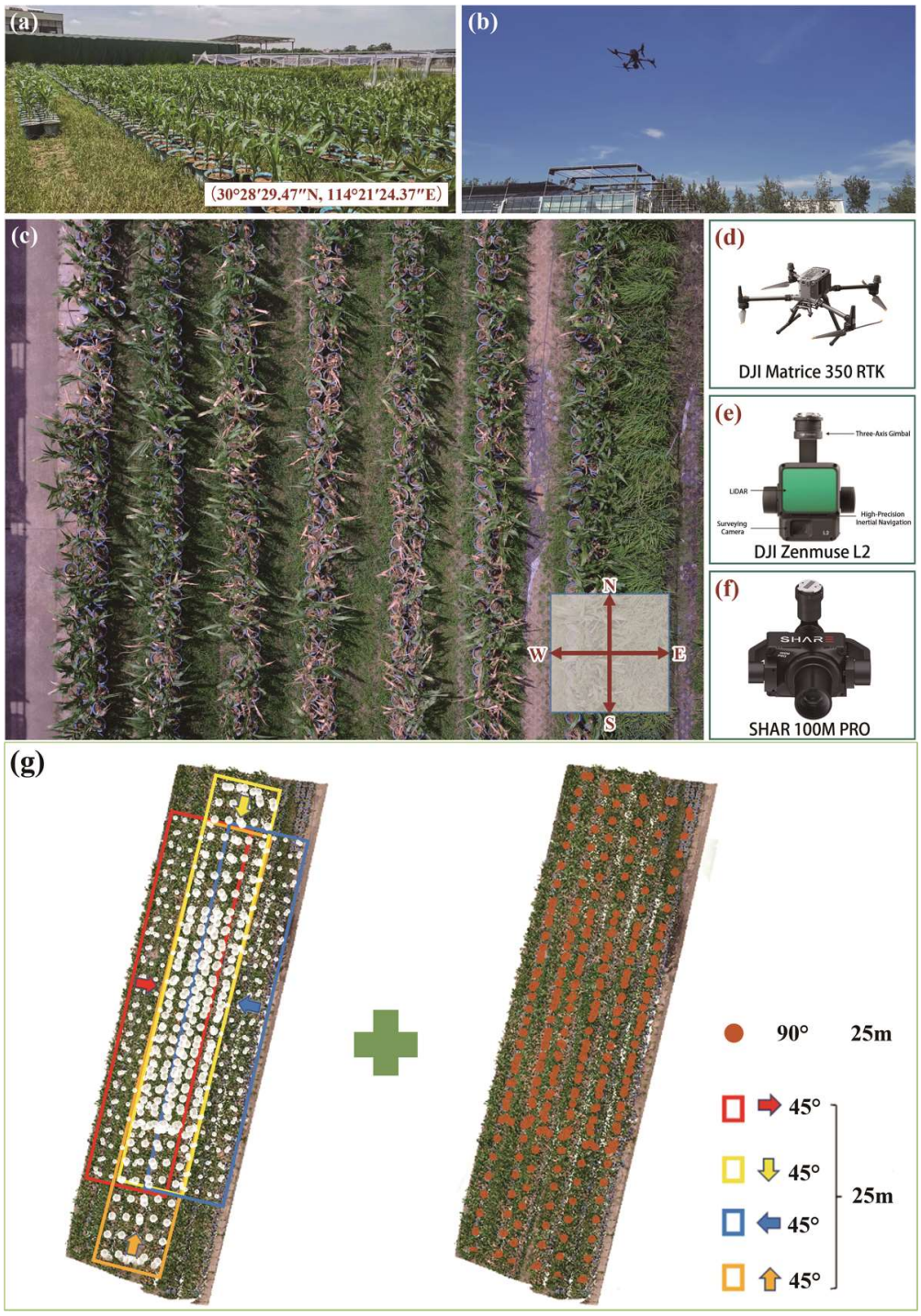

Figure 2. UAV data acquisition. (**a**) Data collection site. (**b**) UAV collection. (**c**) Coverage area and flight path for the maize plot. (**d**) DJI Matrice 350 RTK. (**e**) DJI Zenmuse L2. (**f**) SHAR 100M PRO. (**g**) Five-direction flight route.

2.3. The Datasets

As is well known, dataset creation is vital for deep learning algorithms. The creation of high-quality annotated datasets is labor-intensive and requires skilled annotation efforts. At the individual plant level, there is still a shortage of point cloud datasets clipped from plots, thus limiting the application of deep learning to such tasks. Although plant point cloud datasets at the organ level are relatively more abundant, they face several challenges. Specifically, semantic annotation of the plant point cloud at the organ level is also hindered

by the complex morphological structure of the plants. Despite the introduction of tools designed to simplify this process, like Label3DMaize [84], they are typically specialized for certain plant species, making them unsuitable for varied crops with diverse morphological characteristics. As plants mature, their morphological structures become more complex, which contributes to the intricacy of the annotation process.

While there has been a considerable upward trend in the generation of 3D plant datasets in recent years, the availability of suitable datasets for plant organ segmentation tasks is relatively limited. We acknowledge several notable datasets that excel in ultra-high spatial resolution and comprehensive annotations, including Plant3D [85], ROSE-X [11], Pheno4D [86], Soybean-MVS [87], and LAST-Straw [88], each of which is detailed in Table 1. However, these datasets were procured from controlled environments with simplistic backdrops, contain a limited variety of crops, and the provision of labeled datasets is still sparse. A significant deficit persists for high-quality datasets derived from natural planting environments.

Table 1. Details of plant point cloud datasets.

Dataset	Species	Semantic Level	Total Number/ Labeled Number	Sensor	Scenario	Methods
Plant3D	Tomato, tobacco, sorghum	Organ level	557/0	Laser triangulation	Indoor	Active
ROSE-X	Rose	Organ level	11/11	X-Ray tomography	Indoor	Active
Pheno4D	Tomato, maize	Organ level	224/126	Laser triangulation	Indoor	Active
Soybean-MVS	Soybean	Organ level	102/102	RGB camera	Indoor	Passive
LAST Straw	Strawberry	Organ level	84/13	Structured light	Indoor	Active
Potato	Potato	Organ level	118/118	TLS	Indoor	Active
Tomato	Tomato	Organ level	83/83	RGB camera	Indoor	Passive
Our	Maize	Individual plant	47/47	UAV-LiDAR	Outdoor Pot	Active
		Organ level	134/134	UAV-RGB camera	Outdoor Pot	Passive
		Organ level	225/225	TLS	Field	Active
	Cotton	Organ level	176/176	TLS	Field	Active
	Rapeseed	Organ level	150/150	TLS	Outdoor Pot	Active

In comparison, the data annotation work in this study addresses some of these shortfalls to an extent. We have conducted outdoor data collection in field and potted environments, accurately depicting real-life planting scenarios. Furthermore, we have acquired diverse plant point cloud data utilizing both active methods, including TLS and UAV-LS, and passive methods, which involve image-generated point clouds. The validation of data collected from different platforms can verify the potential of existing deep learning models in agricultural applications.

Figure 3 presents a subset of our data. During the data annotation process, it was observed that the maize point cloud data collected by UAV-LS was relatively sparse, as illustrated in Figure 3d,e, making it challenging to distinguish partly morphological structures of maize. Consequently, it was deemed that such data would be inadequate for organ-level segmentation. UAV-LS technology has been widely applied in field crop scenarios, typically to extract phenotypic parameters of individual plants. Therefore, an attempt is being made to segment these point clouds into individual plants. The point cloud derived from UAV RGB images (Figure 3c) was utilized for the organ segmentation task. Moreover, to augment the point cloud data acquired from various sensors, we utilized point cloud datasets of climbing-stage potato plants [89] and fruit-bearing tomato plants [90] provided by Huazhong Agricultural University's phenotyping platform. These datasets were collected from greenhouse environments using the FARO FocusS70 scanner and the EOS 77D single-lens reflex camera, respectively. Notably, a Multi-View Stereo (MVS) algorithm was employed for the 3D reconstruction of the tomato plants. Table 1 presents the aforementioned point cloud datasets alongside the point cloud datasets generated in this study.

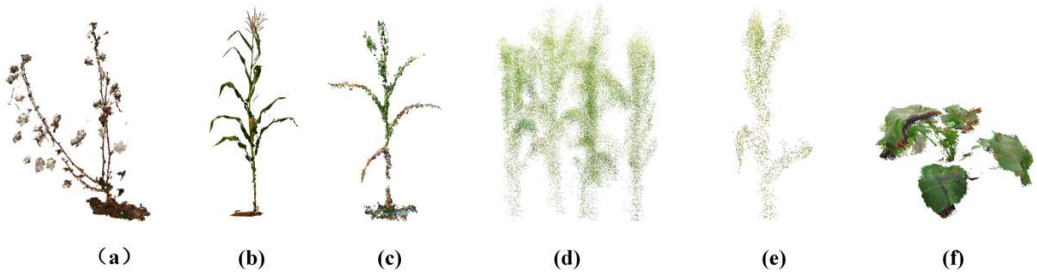

Figure 3. Visualization of selected point clouds from our dataset. (**a**) Cotton point cloud scanned by TLS in the field. (**b**) Maize point cloud scanned by TLS in the field. (**c**) Maize point cloud derived from UAV images. (**d**) Maize plot point cloud collected by UAV-LS. (**e**) Individual maize plant segmented from (**d**). (**f**) Rapeseed point cloud obtained by TLS outdoors.

3. Methodology

Distinct segmentation methods were employed for segmentation tasks at organ levels. In this study, our intent was to leverage the end-to-end point cloud segmentation model to simplify segmentation tasks. Ideally, in phenotypic practice, a trained model would be used to perform segmentation on numerous unlabeled point clouds directly, thus eliminating the need for extensive manual intervention or intricate specialized knowledge. Consequently, models based on deep learning were chosen for comparison. Specifically, nine classic network architectures were considered to adapt to the problem, with their respective merits and drawbacks being compared. These architectures include PointNet [47], PointNet++ [48], DGCNN [55], PointCNN [57], Curvenet [64], PointMLP [61], PAConv [91], Point Transformer (PT) [66], and Stratified Transformer (ST) [68]. Notably, PAConv, a flexible convolution module, was incorporated into DGCNN for evaluation.

The process of segregating individual plants from the filed point cloud is known as instance segmentation. A classic end-to-end scene instance segmentation model, Mask3D [92], is used to perform instance segmentation on the plot point cloud. The following subsections briefly explain the characteristic extraction methods and structures of these architectures.

3.1. PointNet Series Model

The PointNet series encapsulates several network architectures built on the PointNet [47] framework, specifically encompassing PointNet, PointNet++ [48], and PointMLP [61]. These three network models bear profound similarities. PointNet was the inaugural deep learning framework to accept points directly as input, aptly mapping the XYZ coordinate points in the spatial coordinate system to high-dimensional features through multiple traditional MLP layers, followed by maximum pooling to finish feature aggregation. Its successor, PointNet++, enhances this base with a hierarchical Set Abstraction (SA) layer, uses farthest-point sampling to establish the anchor points of the point cloud, and then adopts a spherical query approach to determine each point's k-nearest points. This approach segments the point cloud into repeating point plots, each conducting local feature extraction independently. The PointMLP architecture employs multiple layers of shared Multi-Layer Perceptrons (MLPs) to construct the model. Its core principle involves the progressive extraction of local features through successive residual layers, resulting in a computationally efficient feature extractor. Prior to inputting points into the feature extractor, a geometric affine module is applied. This module addresses the challenges associated with the sparsity and irregularity of geometric structures in local regions.

3.2. DGCNN, PointCNN, and PAConv

Dynamic Graph CNN (DGCNN) [55] architecture constructs the point cloud data into a graph structure, where points serve as nodes, and the edges are dynamically built based on spatial relations between points. This method effectively tackles challenges such as the disordered nature and non-Euclidean structure of point cloud data. Its core operation, EdgeConv, can generate effective local neighborhood feature representation for each point.

Specifically, the EdgeConv operation encompasses two key steps: computation of edge features and aggregation of these edge features. Edge feature computation defines edge features as a function of the relationships between a point and its k-Nearest Neighbors. The arguments of these functions are learnable, permitting the precise capture of local structure information in the point cloud. Edge feature aggregation utilizes a symmetrical function across a channel (max pooling) to aggregate the edge features and re-represents the centroid's features based on these aggregated characteristics. This aggregation mechanism enables the model to incorporate the information in local neighborhoods, thus enriching and improving feature representation accuracy. A noteworthy trait of DGCNN is the dynamic recalculation of the graph structure in each layer's feature space, meaning the model builds the graph based not only on initial spatial relationships but also dynamically updates the graph structure in each layer according to nearest-neighbor relationships in the feature space.

Next, the PointCNN [57] architecture emulates the characteristics of 2D convolution, constructing a core X-Conv operation, and mimicking the convolution layer in two-dimensional models. Consequently, 3D point cloud data can conduct feature extraction similarly to 2D images under the operation of X-Conv. Like PointCNN, PAConv [91] also follows the core point convolution concept, a plug-and-play convolution module. Its key idea is to construct its convolution kernel by dynamically combining the fundamental weight matrix in the weight library, where the assembly coefficients are learned adaptively from the relative position. Particularly, a ScoreNet assigns scores to the primary weight matrix, thereby assembling the new weight matrix; hence, it can adaptively assemble the corresponding weight matrix for different points, overcoming the limitation of poor flexibility in traditional core point convolution kernels. We amalgamate PAConv and DGCNN to validate the performance enhancement effect of PAConv on DGCNN.

3.3. CurveNet

The CurveNet architecture [64] introduces a novel approach for aggregating local features by generating continuous point cloud sequences, referred to as curves, to capture long-range information within point clouds. The model's performance is mainly influenced by two modules: the Curve Grouping Block and the Curve Aggregation Block. The Curve Grouping Block generates continuous point sequences (curves) from the point cloud. To prevent the formation of loops and intersections, a guided walking strategy is employed, which traverses along geometric curves within the point cloud structure, thus capturing the relationships between local points. Specifically, a point is selected as the starting point, and neighboring points are dynamically chosen according to the walking strategy to gradually form the curve. The Curve Aggregation Block integrates the generated curve features back into each point's features. Through this method, local geometric information is effectively captured and integrated, thereby enhancing the representation capability of each point.

3.4. Point Transformer and Stratified Transformer

The Point Transformer [66] is a classical point cloud self-attention model. The core idea of this architecture is vector self-attention, which accumulates local features through a subtraction relationship, unlike the scalar self-attention mechanism. Local features are extracted through a vector self-attention module, and features are merged and restored through the Transition Down and Transition Up modules, which serve as a down-sampling and up-sampling function, respectively.

The Stratified Transformer [68] architecture integrates a stratified feature extraction structure into the transformer framework, composed of multiple down-sampling layers, scalar attention modules, and up-sampling layers. At the model's first layer, the point embedding module employs the KPconv method to effectively aggregate local information of the point cloud, fostering expedited model convergence and performance enhancement. Concurrently, adopting a stratified method, it densely samples neighboring points and sparsely samples distant points as keys in the attention mechanism query, thereby signifi-

cantly expanding the effective receptive field while maintaining lower computational cost and enhancing long-range context capture ability.

3.5. The Composition of Organ-Level Segmentation Models

Based on the research conducted by Ren et al. [93], the segmentation performance of the models is primarily influenced by three components: local operations, grouping strategies, and feature extractors. Local operations were categorized into three classes: non-operational, k-Nearest Neighbors (k-NNs), and ball query techniques. For data grouping, a standard procedure involves initially down-sampling key points through Farthest Point Sampling (FPS) and subsequent grouping. Feature extraction methods include conventional operators (such as shared Multi-Layer Perceptrons and convolutional layers), adaptive kernels, and attention mechanisms. A summary of recent architectural decisions in segmentation networks is detailed in Table 2.

Table 2. Analysis of the nine network model architectures.

	Local Operation	Grouping	Featurizer
PointNet	None	None	Conventional
PointNet++	Ball query	Conventional	Conventional
PointMLP	k-NN	Conventional	Conventional
DGCNN	k-NN	Conventional	Conventional
PointCNN	k-NN	Conventional	Conventional
PAConv	k-NN	Conventional	Adaptive kernel
CurveNet	k-NN	Curve	Conventional
PT	k-NN	Conventional	Self-attention
ST	Ball query+ k-NN	Stratified window	Self-attention

3.6. Mask3D

Mask3D [92] is an innovative model for 3D instance segmentation, marking the first successful application of transformer architecture to large-scale 3D semantic instance segmentation tasks. Several advantages are offered by Mask3D, as it does not rely on (1) voting schemes requiring the manual selection of geometric properties (such as centroids), nor (2) geometric grouping mechanisms necessitating manual hyperparameter tuning (such as radius). Additionally, (3) the model is capable of directly optimizing the loss of instance masks. The Mask3D network model principally comprises three core components: a feature backbone network, a transformer decoder, and a mask generation module. The feature backbone network employs a 3D sparse convolutional network based on MinkowskiEngine, enabling the model to extract features efficiently from 3D point clouds. By performing multiscale voxel feature extraction on the point cloud, this network gains a profound understanding of the point cloud geometric shapes. Following this, the transformer decoder further processes these features via a succession of decoding layers. Using cross-attention and self-attention mechanisms, it accomplishes fine-grained feature extraction and localization for each instance. Ultimately, the mask generation module transforms the decoder's output into concrete instance masks, substantially enhancing segmentation accuracy by directly optimizing the quality of instance masks.

3.7. Evaluation Metrics

When evaluating the segmentation performance of models on point cloud data, four fundamental metrics are usually considered: True Positives (TPs), False Positives (FPs), False Negatives (FNs), and True Negatives (TNs). A TP denotes the number of points correctly identified as class i, an FP denotes the number of points from other classes incorrectly identified as class i, an FN denotes the number of points from class i incorrectly classified into other classes, and a TN denotes the number of points from other classes correctly not classified as class i.

3.7.1. Evaluation Metrics for Organ-Level Segmentation

In the organ segmentation, we employ several metrics to assess the performance of the models. These metrics include the following:

Intersection over Union (IoU): This metric measures the overlap between the predicted positive class and the actual positive class. The intersection ratio for the *i*-th class among N classes is represented by Equation (1):

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \tag{1}$$

Mean Intersection over Union (mIoU): This is the average IoU across all classes, providing a single performance measure for multi-class classification problems. The intersection ratio for all classes is represented by Equation (2):

$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoU_i \tag{2}$$

Accuracy: This is the ratio of correctly identified instances (both TP and TN) to all instances in the dataset (TP + TN + FP + FN). The precision for the i-th organ part among N is calculated as shown in Equation (3):

$$Acc_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{3}$$

Mean Accuracy (mACC): This is the average per-class accuracy, which considers the accuracy for each class separately and then computes the average. The mean precision for all classes is calculated using Equation (4):

$$mACC = \frac{1}{N}\sum_{i=1}^{N} Acc_i \tag{4}$$

Global Accuracy (OA): This metric measures the percentage of correctly classified points in the entire dataset. The global accuracy is calculated using Equation (5), and N represents the number of organ parts:

$$0A = \frac{\sum_{i}^{N}(TP_i + TN_i)}{\sum_{i}^{N}(FP_i + FN_i + TP_i + TN_i)} \tag{5}$$

3.7.2. Evaluation Metrics for Individual Plant-Level Segmentation

In the instance segmentation experiments at the individual plant level, model performance is measured using Average Precision (AP). AP is a comprehensive metric that considers performance across various confidence thresholds. It is calculated by sampling multiple Intersection over Union (IoU) thresholds and averaging the precision at these thresholds to assess overall model performance. AP_{50} refers to AP calculated at a specific IoU threshold of 0.5, while AP_{25} refers to AP calculated at an IoU threshold of 0.25. AP is calculated from the precision–recall (P-R) curve, where precision is defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Recall is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

AP is obtained by plotting precision and recall at various thresholds to form the $P(R)$ curve, where $P(R)$ denotes precision at a given recall R.

$$AP = \int_0^1 P(R)dR \tag{8}$$

The calculation methods for AP_{50} and AP_{25} are similar to those for AP. However, a fixed IoU threshold is used to determine precision and recall, plot the $P(R)$ curve, and compute the area under the curve.

4. Results

For the task of organ-level point cloud segmentation, we conducted a comprehensive assessment of the segmentation performance of the aforementioned nine models on point cloud data acquired under different scenarios and from various sensor platforms. The segmentation experiments were classified into three categories according to the method of obtaining point clouds: (1) experiments using point clouds acquired through laser triangulation, (2) experiments utilizing TLS point clouds, and (3) experiments on image-generated point clouds. Furthermore, in this section, the performance of the Mask3D model in performing instance segmentation on plot point clouds was also validated.

All experiments were conducted on a server with a 24-core, 48-thread CPU, 256 GB memory, and 4 NVIDIA Tesla V100 SXM3 GPUs, running on the Ubuntu operating system. PyTorch was used as the training framework. Each network model underwent single-GPU training, with 150 epochs for Stratified Transformer and Point Transformer training, 4000 epochs for Mask3D, and 250 epochs for the remaining models.

For the task of organ-level point cloud segmentation, to ensure consistent evaluation of the model, all crop point clouds were down-sampled to 10,000 points using the Farthest Point Sampling method during organ-level validation. Before the data were fed into the network, shuffling and normalization were performed. In doing so, we aimed to validate the efficacy of these classical deep learning models in dealing with point clouds collected from various environments and platforms. For the segmentation of individual plants, the network required a large sub-sampling rate when the entire regional point cloud was fed into it, which resulted in a significant loss of geometric information. Therefore, the regional maize point cloud was partitioned into fixed-size voxels, and a traversal of the voxels was performed, enabling feature extraction on all points within each voxel.

4.1. Organ-Level Segmentation

4.1.1. Experiments Using Point Clouds Acquired through Laser Triangulation

The Pheno4D dataset [86] was utilized as the source of laser scanning data under controlled indoor conditions to evaluate the performance of various semantic segmentation models. Table 3 presents the segmentation results of these models on the Pheno4D dataset, where OA denotes the overall accuracy, mACC denotes the average accuracy of every class, and mIoU represents the average IoU of every class. The majority of the models achieved an mIoU of over 80%, with the exception of PointNet and DGCNN. The ST and PT models exhibited the best segmentation performance, with overall accuracies surpassing 95%. The segmentation outcomes are depicted in Figure 4.

PointNet and DGCNN exhibited subpar performance on maize and tomato crops, with notable inaccuracies, especially in segmenting the upper stems and leaves of tomato canopies. While other models successfully segmented corn plant stems and leaves, only PointCNN, PT, and ST managed to effectively segment the structurally intricate tomato plants. Challenges persist for other networks in accurately segmenting the upper stems of tomato plants.

4.1.2. Experiments Utilizing TLS Point Clouds

The datasets of maize, cotton, rapeseed, and potatoes were utilized to validate the model's application potential for point cloud data acquired using TLS technology. Table 4 presents the segmentation results of these models on TLS point clouds. Focusing on maize and cotton as representative examples, the segmentation results are illustrated in Figure 5.

Compared to the Pheno4D dataset, there was a general decline in performance metrics. Point cloud data collected in real growing scenarios are typically affected by various environmental factors, such as background clutter and wind. These point clouds often contain noise points that are difficult to filter and occlusions caused by dense planting. Compared to Pheno4D, these datasets present greater challenges for segmentation. ST achieved the best segmentation results across all subjects, while PT secured sub-optimal results in most cases. PointCNN and PAConv yielded comparable segmentation outcomes for both crops, effectively separating the stems and leaves. However, more detailed areas, such as the maize cobs and obscured sections of cotton plants, showed varying levels of misclassification.

Table 3. Segmentation results on Pheno4D. The bolded text indicates optimal performance, while the underlined text denotes the second-best results.

	Maize			**Tomato**		
	OA	**mACC**	**mIoU**	**OA**	**mACC**	**mIoU**
PointNet	0.917	0.803	0.716	0.945	0.822	0.767
PointNet++	0.967	0.916	0.863	0.962	0.888	0.833
PointMLP	0.974	0.950	0.896	0.963	0.892	0.839
DGCNN	0.935	0.827	0.753	0.951	0.854	0.798
PointCNN	0.966	0.933	0.868	<u>0.970</u>	0.903	0.862
PAConv	0.972	0.938	0.887	0.963	0.891	0.838
CurveNet	0.968	0.928	0.874	0.964	0.891	0.842
PT	<u>0.981</u>	<u>0.959</u>	<u>0.919</u>	**0.971**	**0.914**	**0.868**
ST	**0.983**	**0.963**	**0.926**	0.964	<u>0.908</u>	<u>0.865</u>

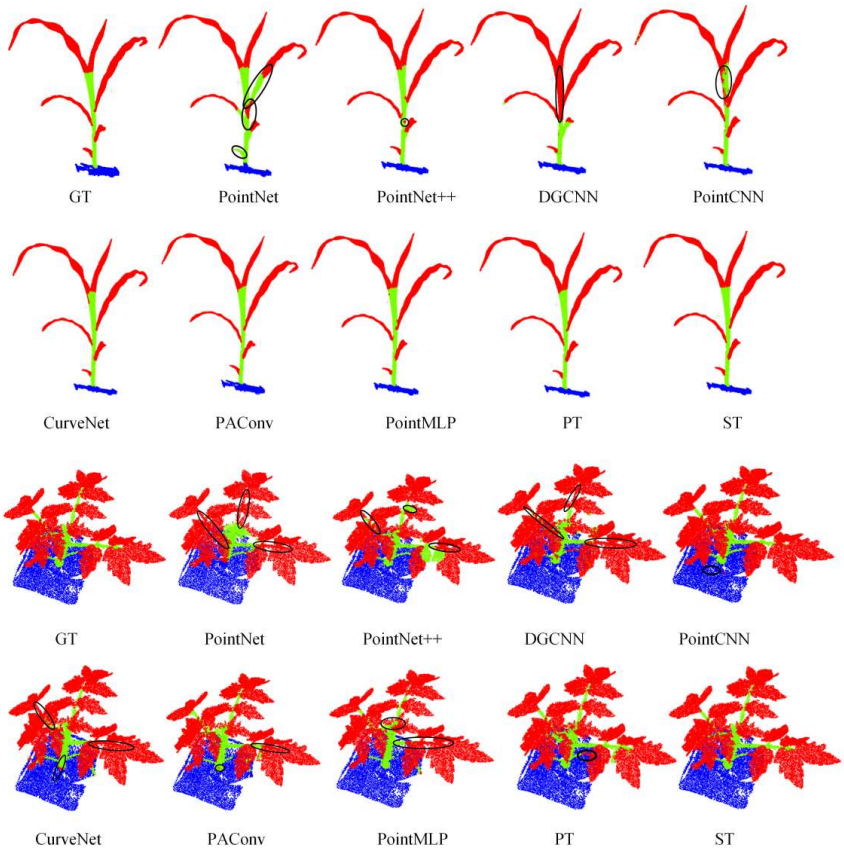

Figure 4. Segmentation results on Pheno4D. The circled areas indicate incorrectly segmented regions; GT represents the ground truth.

Table 4. Segmentation results on plants from agricultural scenarios. Maize_1 refers to the maize data collected in field scenarios using TLS. The bolded text indicates optimal performance, while the underlined text denotes the second-best results.

	Maize_1			Cotton			Potato			Rapeseed		
	OA	**mACC**	**mIoU**	**OA**	**mACC**	**mIoU**	**OA**	**mACC**	**mIoU**	**OA**	**mACC**	**mIoU**
PointNet	0.857	0.598	0.527	0.743	0.746	0.626	0.933	0.792	0.702	0.927	0.797	0.727
PointNet++	0.929	0.731	0.676	0.811	0.813	0.705	0.957	0.881	0.799	0.936	0.805	0.742
PointMLP	0.909	0.766	0.633	0.821	0.822	0.716	0.954	0.835	0.773	0.922	0.809	0.721
DGCNN	0.912	0.701	0.620	0.804	0.806	0.696	0.974	0.914	0.865	0.938	0.845	0.768
PointCNN	0.934	0.744	0.683	0.845	0.846	0.749	0.968	0.880	0.836	0.938	0.819	0.750
PAConv	0.925	0.736	0.668	0.824	0.825	0.719	0.972	0.901	0.855	0.950	0.844	0.777
CurveNet	0.926	0.721	0.660	0.818	0.820	0.713	0.959	0.869	0.800	0.939	0.830	0.764
PT	0.945	0.806	0.751	0.837	0.839	0.739	0.974	0.912	0.867	0.947	**0.859**	0.785
ST	**0.950**	**0.890**	**0.806**	**0.878**	**0.879**	**0.794**	**0.988**	**0.948**	**0.923**	**0.959**	0.857	**0.799**

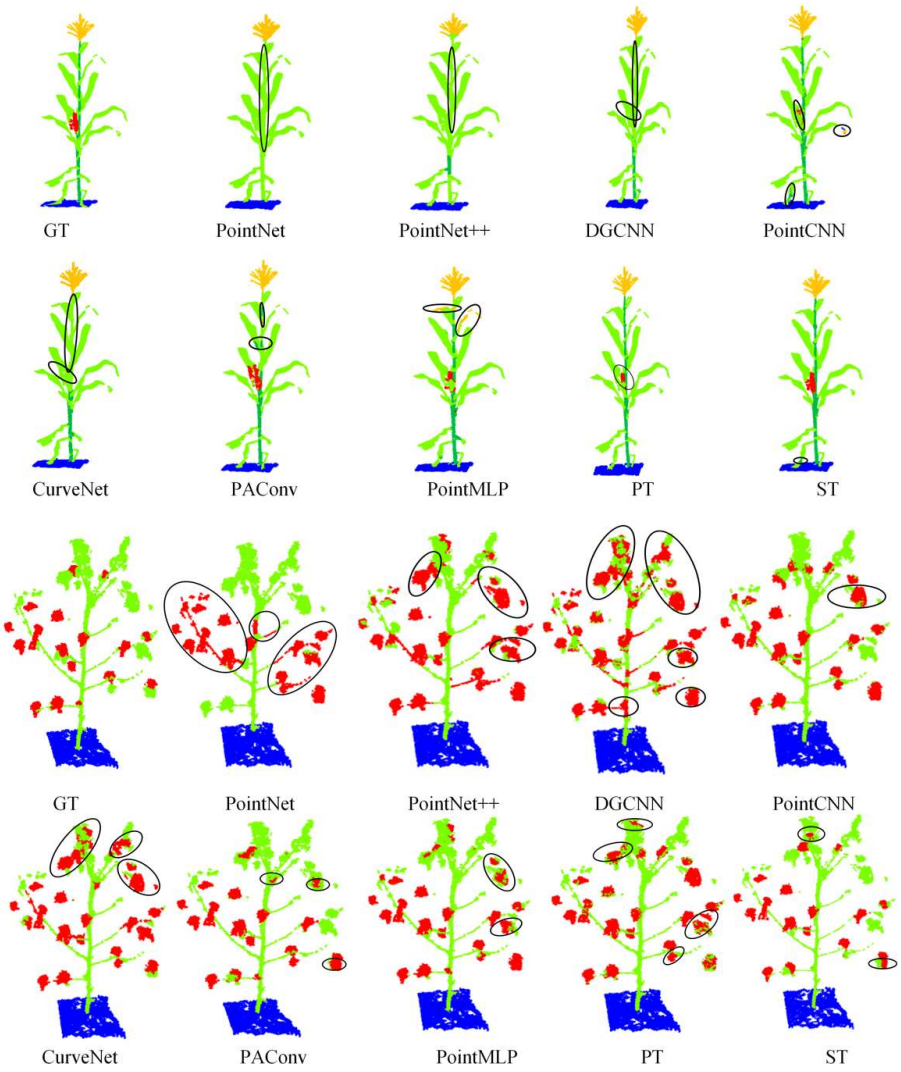

Figure 5. Segmentation results of maize and cotton collected from real planting scenarios. The circled areas indicate incorrectly segmented regions; GT represents the ground truth.

4.1.3. Experiments on Image-Generated Point Cloud

The model's segmentation performance was validated using point clouds of reconstructed tomato plants as the dataset. This dataset was obtained under greenhouse conditions using DSLR cameras to capture RGB images, which were then reconstructed.

Table 5 presents the segmentation results of the models on a dataset of mature tomatoes, with the visualization results displayed in Figure 6. The morphological structure of the tomato plants is highly complex, compounded by a few persistent noise points that increase the difficulty of organ segmentation. The mean Intersection over Union (mIoU) for all models was below 70%. ST and PT achieved the best and second-best results, respectively, but the segmentation outcomes were still unsatisfactory. The models continue to struggle with the segmentation of tomato fruits, and the segmentation of stems and leaves in densely leafed areas also poses significant challenges. Other models even failed to identify the majority of the tomato stems.

Table 5. Segmentation results on tomato point clouds derived from DSLR images. The bolded text indicates optimal performance, while the underlined text denotes the second-best results.

	OA	mACC	mIoU	Stem		Leaf		Fruit	
				IoU	Acc	IoU	Acc	IoU	Acc
PointNet	0.812	0.559	0.450	0.324	0.384	0.796	0.943	0.230	0.351
PointNet++	0.823	0.504	0.426	0.366	0.435	0.806	0.945	0.101	0.132
PointMLP	0.806	<u>0.721</u>	0.466	0.431	0.573	0.782	0.874	0.185	**0.716**
DGCNN	0.852	0.639	0.538	<u>0.490</u>	<u>0.603</u>	0.829	0.931	0.296	0.382
PointCNN	0.861	0.662	0.567	0.512	0.620	<u>0.839</u>	0.937	0.351	0.427
PAConv	0.855	0.687	0.563	**0.525**	**0.678**	0.831	0.912	0.335	0.469
CurveNet	0.842	0.592	0.502	0.454	0.554	0.821	0.934	0.230	0.289
PT	<u>0.856</u>	0.708	<u>0.629</u>	0.464	0.546	0.835	<u>0.949</u>	**0.588**	0.628
ST	**0.869**	**0.741**	**0.643**	0.504	0.573	**0.850**	**0.950**	<u>0.576</u>	<u>0.693</u>

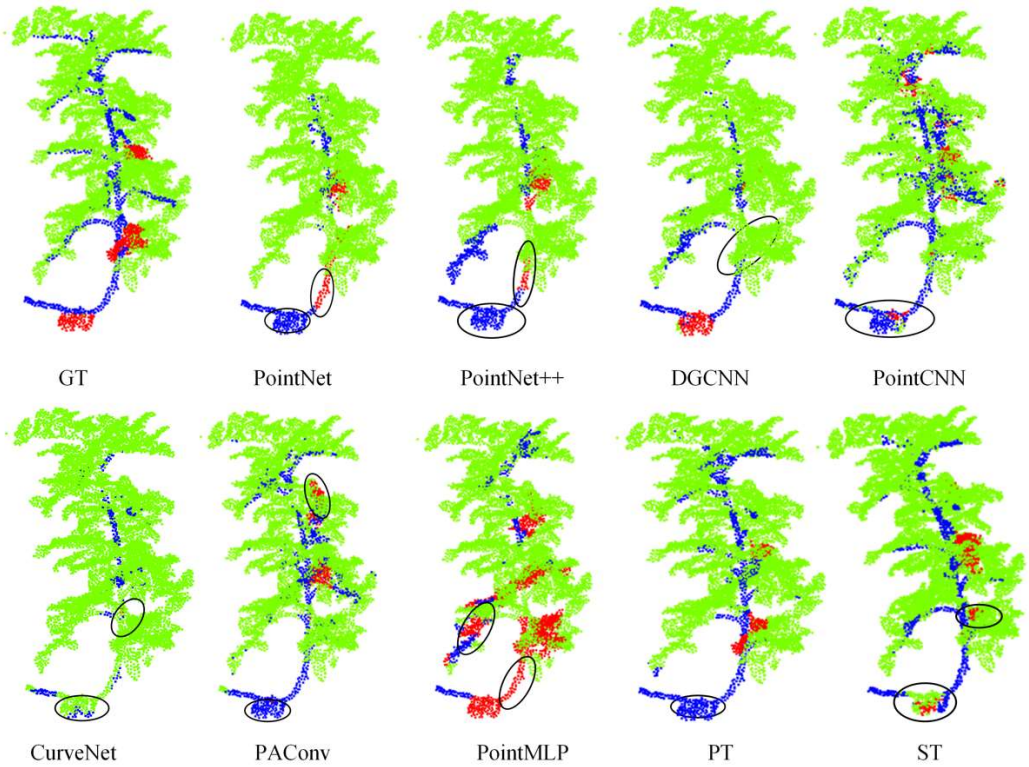

Figure 6. Segmentation results of tomatoes reconstructed by close-range DSLR camera RGB images. The circled areas indicate incorrectly segmented regions; GT represents the ground truth.

Additionally, to verify the potential of the model in the segmentation task of point clouds derived from UAV-based RGB images, we utilized reconstructed maize point clouds as an example dataset. The model performance is summarized in Table 6, with the visualization results displayed in Figure 7. All models achieved an accuracy above 0.8.

CurveNet exhibited the poorest performance, with an mIoU of 0.608, indicating significant degradation compared to data collected from other platforms. Moreover, PointMLP's performance was notably impacted, falling below that of its predecessor, PointNet. In contrast, DGCNN showed relative performance improvements over previous datasets. Both ST and PT maintained their positions as the top-performing and second-best models, respectively.

Table 6. Segmentation results on maize point clouds derived from UAV images. The bolded text indicates optimal performance, while the underlined text denotes the second-best results.

	OA	mACC	mIoU	Soil		Stem		Leaf	
				IoU	Acc	IoU	Acc	IoU	Acc
PointNet	0.890	0.864	0.761	0.960	0.997	0.529	0.747	0.795	0.849
PointNet++	0.928	0.896	0.827	0.971	**0.998**	0.647	0.764	0.864	0.926
PointMLP	0.871	0.837	0.732	0.917	0.939	0.506	0.701	0.772	0.872
DGCNN	0.926	0.906	0.828	0.967	0.990	0.656	0.817	0.861	0.910
PointCNN	0.918	0.890	0.810	0.980	0.994	0.606	0.770	0.844	0.907
PAConv	0.928	0.902	0.828	0.980	0.995	0.644	0.792	0.860	0.917
CurveNet	0.838	0.695	0.608	0.952	1.000	0.128	0.144	0.743	**0.940**
PT	0.938	0.922	0.851	0.982	0.996	0.692	0.849	0.880	0.920
ST	**0.949**	**0.944**	**0.877**	**0.984**	0.995	**0.746**	**0.915**	**0.900**	0.922

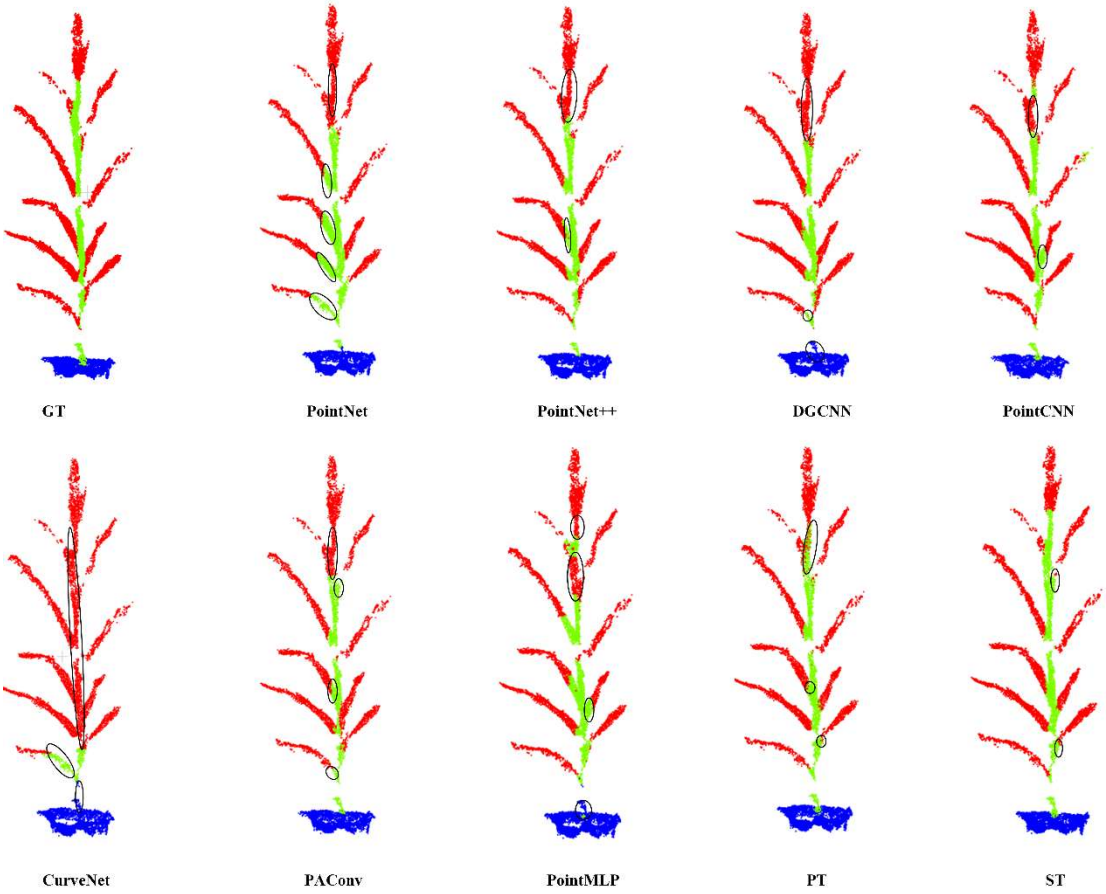

Figure 7. Segmentation results of maize reconstructed by airborne RGB images. The circled areas indicate incorrectly segmented regions; GT represents the ground truth.

4.1.4. Comprehensive Evaluation

The measurement approach comprehensively evaluates the models' performance on both the ideal Pheno4D data and datasets collected under real-world conditions. We aim to use this test to assess the segmentation performance of the models across various types of

datasets. As shown in Table 7, ST and PT achieved the best and second-best results, respectively. PointCNN and PAConv reached an mIoU of over 75%, followed by PointNet++ and DGCNN. CurveNet and PointMLP exhibited similar segmentation outcomes, performing poorly on crop subjects with insufficient precision in detailed segmentation. PointNet was unable to handle the organ segmentation tasks for the aforementioned subjects adequately.

Table 7. The summary of segmentation results on all crops. The bolded text indicates optimal performance, while the underlined text denotes the second-best results.

	OA	**mACC**	**mIoU**
PointNet	0.868	0.748	0.659
PointNet++	0.910	0.804	0.734
PointMLP	0.897	0.829	0.722
DGCNN	0.906	0.811	0.733
PointCNN	0.921	0.835	0.766
PAConv	0.918	0.841	0.767
CurveNet	0.902	0.793	0.720
PT	<u>0.927</u>	<u>0.865</u>	<u>0.801</u>
ST	**0.940**	**0.891**	**0.829**

PointNet, a pivotal model in point cloud data processing, utilizes a unique point-wise approach and architecture as detailed in Section 3.2. Despite its innovation, PointNet does not consider inter-point dynamics and primarily depends on global feature extraction through max pooling at the terminal stage. Nonetheless, PointNet lacks power in tasks requiring both local and global feature integration for point cloud segmentation. Its inability to simultaneously capture detailed local structures and overarching global patterns makes it less effective for segmenting plant organs, where intricate local–global feature extraction is crucial.

PointNet++ significantly enhances local feature extraction compared to its predecessor, PointNet. It adopts an encoder–decoder framework that leverages FPS to identify key central points. These points anchor a ball search that divides the point cloud into overlapping local areas via a grouping process. Within these areas, a mini-PointNet was used to encode local region patterns into feature vectors. This design enables PointNet++ to efficiently gather a more comprehensive set of local data, improving its utility in detailed segmentation tasks. PointMLP is noted for its efficient design, incorporating residual point MLP modules for local feature extraction and max pooling for feature aggregation. This model, inspired by residual networks, progressively expands its receptive field, similar to PointNet++, which also uses k-NN for local tasks and MLP for feature processing. Although PointMLP successfully enlarges its receptive field by layering MLP modules to enhance global feature capture, it falls short in integrating local group interconnections. DGCNN utilizes a graph-based framework to delve into complex point relationships within a graph. DGCNN incorporates k-NN aggregation in its EdgeConv module to capture local details. Post the final EdgeConv operation, it aggregates global features via max pooling. Despite its strengths, DGCNN's local information capture is limited by the scope of neighboring points it considers, omitting more distal points.

CurveNet starts with the farthest point, employing a k-NN method and an MLP-based scoring system to select the starting point for curve analysis. It then uses a walking policy for curve construction and an MLP to extract and integrate curve features. CurveNet's advanced curve aggregation technique and walking strategy effectively incorporate data from distant points, enhancing connectivity across extended regions of the point cloud. Our comparative experiments indicate that CurveNet matches the performance of PointCNN and PAConv on the clean Pheno4D datasets, but its effectiveness declines on crop types collected outdoors. Xiang et al. have noted that CurveNet's curve detection is particularly vulnerable to noise, which can alter the initial curve point and cause inaccuracies [64]. In real-world conditions, the crop point clouds we collected contain noise and partial occlu-

sions, significantly affecting CurveNet's ability to accurately group data. We conjecture that the robustness of CurveNet's performance is strongly dependent on dataset quality, with noise and missing data notably degrading its ability to segment plant objects.

PointCNN and PAConv demonstrate acceptable segmentation abilities for various plants, including rapeseed, potatoes, and cotton, with similar performance outcomes. PointCNN features a unique convolutional operator called X-Conv, which utilizes k-NN to collect nearby points for feature aggregation, adapting the principles of traditional CNNs for 3D point cloud data. The X-Conv layers gather and centralize features from adjacent points, and through their sequential arrangement, they progressively widen their receptive field to cover the entire point cloud. While this method effectively captures and combines local features, it struggles to integrate attributes from more distant points. This shortcoming is notably problematic when analyzing crops with complex morphologies, where the model's reliance on stacking convolutional layers might fail to capture detailed feature interactions, possibly resulting in less accurate segmentation. The Position-Adaptive Convolution (PAConv) method optimizes point cloud processing by employing k-NN to select neighboring points and dynamically creating a weight matrix using a network named ScoreNet. This adaptability allows the operator to conform to diverse point cloud geometries effectively. Integrating PAConv into the DGCNN framework, and replacing the EdgeConv operator, significantly enhances the model's ability to capture local features. This integration reduces the computational load of measuring feature distances and facilitates the inclusion of additional neighboring points, thereby improving local feature representation. As a result, there is a notable increase in the average Intersection over Union (IoU) across various crops, from 0.733 to 0.767. The application of PAConv within the DGCNN framework enables superior segmentation, particularly in relatively complex crops like cotton, marking a substantial improvement in the detailed capture of local structures within point clouds.

The Point Transformer employs a self-attention mechanism to effectively extract features from point clouds, using k-NN during its Transition Down phase to collect neighboring points and max pooling to integrate this local information. Building on this, the Stratified Transformer enhances the design by incorporating the KPConv method for point embedding and adopting stratified sampling to capture both dense and sparse points. This methodology ensures a thorough integration of both local and global information. The Stratified Transformer, in contrast to the Point Transformer, features a broader receptive field that allows it to collect data from more distant points. This is particularly demonstrated by its robust segmentation capabilities across a variety of crop types. Excelling in segmenting plant organs within seven distinct crops, the Stratified Transformer proves highly effective and versatile in handling complex point cloud data for agricultural applications. The excellent segmentation performance of both the Point Transformer and Stratified Transformer demonstrates the significant potential of self-attention mechanisms in the task of plant organ segmentation.

4.1.5. Evaluation of Computational Costs

During our performance evaluation, the Stratified Transformer emerged as the most accurate model for segmentation. To thoroughly assess each model's effectiveness, we recorded their training times, which reflect computational costs. Each model underwent training on a single GPU, optimizing the batch size to match the GPU's capacity. Table 8 details the time efficiency and segmentation accuracy of each model. Notably, the Stratified Transformer, while delivering superior segmentation results, required the most extensive training time at 88.16 h. Conversely, the PointMLP model, with its efficient architecture, logged the shortest training period of just 2.86 h.

Table 8. Training duration for nine segmentation models across seven crop datasets. Maize represents the maize data sourced from Pheno4D; Maize_1 refers to the maize data collected in field scenarios using TLS; Maize_2 represents the maize point cloud derived from UAV RGB images; and Tomato_ represents the tomato point clouds derived from DSLR images.

	Tomato	Maize	Tomato_	Maize_1	Potato	Cotton	Rapeseed	Maize_2	Total
PointNet	0.67	0.62	0.17	0.35	0.18	0.28	0.25	0.43	2.95
PointNet++	0.70	0.63	0.35	0.63	0.50	0.67	0.38	0.28	4.15
PointMLP	0.42	0.47	0.28	0.33	0.23	0.37	0.33	0.43	**2.86**
DGCNN	1.00	1.17	1.17	2.80	1.58	2.22	1.92	1.67	13.52
PointCNN	1.10	1.33	1.00	2.73	1.50	2.17	1.90	2.26	13.99
PAConv	2.00	2.63	1.92	5.38	4.12	5.70	5.02	6.22	32.99
CurveNet	4.00	4.50	4.63	4.65	11.30	7.00	5.63	5.83	47.55
PT	4.07	4.20	7.68	10.00	4.80	8.32	6.20	1.41	46.68
ST	5.00	4.67	7.63	12.00	6.20	11.00	11.63	30.03	88.16

Figure 8 provides a comprehensive comparative analysis of various segmentation models, illustrating their performance and computational efficiency. Additionally, the figure summarizes key performance metrics—overall accuracy (OA) and mean IoU (mIoU)—for all evaluated crop types. The comparison highlights the high computational demands of the Stratified Transformer. In contrast, models like PointCNN and PAConv offer impressive accuracy with much shorter training times, making them suitable for simpler crops.

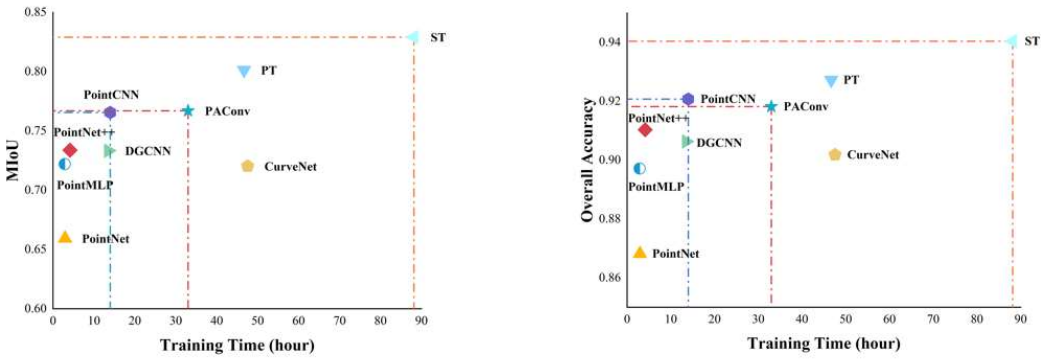

Figure 8. Performance and training time comparison among segmentation models for various crop types. The figure on the left depicts the analysis of mIoU against training time, while the figure on the right illustrates the analysis of OA against training time.

The field of point cloud data processing has become a focal point in both academic and industrial circles, fueled by innovative models that have significantly enhanced segmentation technologies. These models excel at segmentation, utilizing a range of techniques including convolutional operations, graph-based methods, and attention mechanisms to extract local features effectively.

However, despite these advancements, the task of accurately segmenting point clouds, particularly for plant species with complex morphologies, continues to be a pressing challenge. Models like the Stratified Transformer demonstrate exceptional segmentation abilities but are hindered by high computational demands. There is an evident need for models that optimally balance computational efficiency and segmentation accuracy. This balance is crucial and represents a key direction for future research in point cloud segmentation. The ongoing efforts to improve the precision and efficiency of these models are expected to catalyze the next generation of innovations in this field.

4.2. Individual Plant-Level Segmentation

The dataset of the plot maize was utilized to validate the Mask3D's application potential for point cloud data acquired using UAV-LS technology. Table 9 presents the

segmentation results of Mask3D on the plot maize dataset. Figure 9 illustrates the visualization of instance segmentation results produced by Mask3D. Overall, Mask3D achieved satisfactory segmentation performance on the plot maize, with maize accuracy of 0.817 and a mean Average Precision (mAP) of 0.909. The visualization of the Mask3D results demonstrates that Mask3D effectively achieves instance segmentation of plot maize point clouds, except for one plant that was not distinguished. Remarkable segmentation results have been achieved on other individual maize plants. More importantly, compared to traditional segmentation algorithms, Mask3D does not require the configuration of complex parameters, showcasing significant application potential in point cloud segmentation at the individual plant level.

Table 9. Performance metrics of Mask3D.

	AP	**AP_{50}**	**AP_{25}**
Soil	1.000	1.000	1.000
Maize	0.817	0.994	1.000
mAP	0.909	0.997	1.000

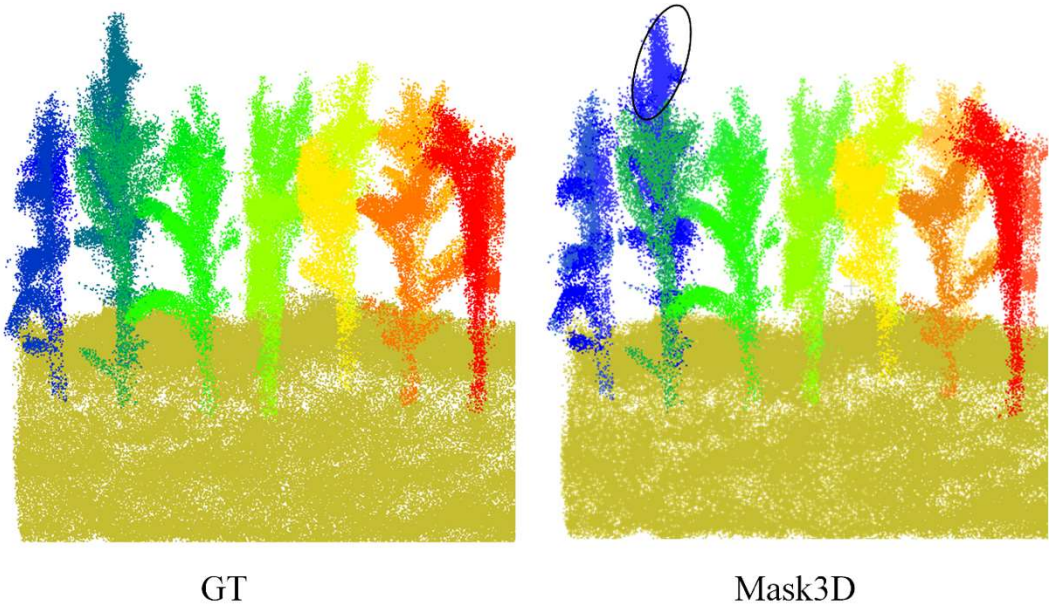

Figure 9. Visualization of instance segmentation results by Mask3D. The circled areas indicate incorrectly segmented regions; GT represents the ground truth.

5. Discussion

5.1. The Impact of Class Imbalance

Our experiments revealed a persistent issue of class imbalance in plant datasets, which notably impacts the segmentation accuracy across various plant organ categories. This imbalance primarily stems from the disproportionate distribution of point counts among different organs. For instance, leaves, which are larger and more abundant, dominate the point count, whereas the main stem, being smaller and structurally simpler, contributes fewer points. This disparity often leads to segmentation errors, particularly at the junctures between stems and leaves, where models tend to misclassify boundary points or those in smaller proportions as leaves. As an example, the segmentation results of models on tomato point clouds derived from images were utilized, as illustrated in Table 5 and Figure 6. Our analysis indicates significant performance variations among segmentation models when processing different organ categories of the same crop. These results highlight the critical need for segmentation models to effectively address class imbalance, ensuring accurate representation and segmentation of all plant organs, regardless of their point count or

structural complexity. Our analysis shows that segmentation models for tomato plants perform less effectively in identifying stems and fruits, with notably lower Intersection over Union (IoU) and accuracy (acc) values, compared to leaves. To explore this, we analyzed the total point counts and their proportions across all labeled categories in the tomato point cloud dataset derived from images. The results show leaves holding the highest point proportion at 0.734, stems at 0.198, and fruits the lowest at 0.068. This marked disparity suggests that imbalances in point distribution significantly impair model performance. However, altering these proportions to balance point counts is not feasible without compromising the natural morphological integrity of the plants, as different structural components inherently vary in size and complexity.

5.2. The Impact of Model Composition

In Table 2, three critical components of organ segmentation models, namely local operations, grouping methods, and featurizers, are summarized. Based on these components, an analysis of the performance differences among the models was conducted. Feasible recommendations for constructing segmentation models in the field of plant phenotyping were proposed.

Discussing from the perspective of local operation, the goal of local operations is to encode information representation utilizing local geometric features. Two commonly used local search approaches are ball query and k-Nearest Neighbor (k-NN). Ball query entails random sampling of adjacent points within a predefined radius, while k-NN focuses on the nearest neighboring points. Consequently, for crops with significant local morphological deficiencies, random sampling within a fixed radius could often result in point loss around the local region, such as with PointNet++. This model exhibits commendable precision on point clouds with relatively intact morphology, such as Pheno4D, while it underperforms on objects with missing morphological structures. In facing this challenge, k-NN sustains favorable local features by opting for neighbors from the remaining points, similar to PAConv. However, in situations with an abundance of noise points, ball query often establishes recognition power superior to that of k-NN. This is because, for a point on an object, outliers are unlikely to fall within the querying sphere rather than to be its nearest neighbors [93]. In comparison, ST employs a method combining ball query and k-NN to acquire local features. It selects points based on their proximity to neighbor points within a set local radius, maintaining robustness and the completeness of local features simultaneously. Moreover, the dynamic assembly method was also noteworthy. PAConv demonstrated commendable performance in this validation due to its distinctive method of constructing convolution kernels. These were dynamically data-driven, with a weight matrix adaptively learned from points' positional relationships, enabling the model to manage plants' issues of inconsistent organ scale and irregularity in shape. A substantial performance enhancement was clearly observed in DGCNN following the integration of PAConv.

The absence of point cloud morphological structures during data collection often poses challenges to achieving a comprehensive understanding of the models. However, two classical transformer-based models were less impacted by this issue. Compared to other models, transformer-based architectures, with their self-attention mechanism for feature aggregation, are capable of considering information from longer-range points. Specifically, even with partial local region occlusion, the transformer architecture can still capture points from other regions, maintaining the overall structural integrity. Other architectures have employed various strategies to achieve a wide receptive field. For instance, PointCNN maintains structural information integrity through multilayer stacked X-Conv kernels, which provide a global perspective. While CurveNet also attempts to capture more feature information by establishing long-range point-to-point relationships, it is hampered by the robustness of its scoring function, performing poorly on objects with significant structural information loss, such as maize point clouds reconstructed from UAV RGB images. Therefore, when designing an architecture to handle plant object data from

3D point clouds, the model's robustness and long-distance feature extraction capabilities still require careful consideration.

Moreover, as indicated in Section 5.1, point cloud data of plant objects often present substantial class imbalance. Certain studies have attempted to rectify this problem via novel down-sampling approaches [60], yet the issue is not so simplistic. As plants grow, their organ scales also change, rendering the multiscale characteristics of plants another vital consideration. The model architecture should be capable of adaptively adjusting its spatial receptive field to accommodate the continually changing scale information of plants.

5.3. The Influence of Collection Environment

We conducted a comprehensive evaluation of various classical models using point clouds of several crops collected from different platforms and environments. Our findings indicate that the environmental context of data collection and the platform used substantially affect model performance. The environmental context of data collection significantly impacts model application. Various elements of data collection in natural conditions, such as weed interference and wind, create challenges and introduce a considerable amount of noise, which is difficult to eliminate from the point cloud data. Some models, such as CurveNet, perform well in controlled environments but show poor performance in real-world agricultural settings due to these noise factors. Additionally, real-world agricultural environments are often densely planted, making it difficult for sensors to capture the morphological structure of the inner layers of plants. This results in severe structural information loss within the point cloud data, adversely affecting the performance of nearly all models.

Moreover, data collection using UAV platforms has been widely applied to large-scale crop objects. However, it remains challenging to acquire plants with intact structures at the organ level. Compared with terrestrial laser scanning, this method seems more suited for trait extraction at individual plant levels. High-resolution RGB images also pose a significant challenge to 3D reconstruction methods, requiring substantial computational resources. There is a long way to go in the application of drones in field settings and the extraction of organ-level traits.

5.4. Individual Plant-Level Segmentation Based on UAV-LS

Phenotypic analysis technology based on UAV-LS has garnered significant interest from researchers due to its capability to collect large-scale field crop data in a multi-scale and real-time manner. Upon obtaining the crop point cloud data from the field scene, efficiently extracting individual plants is a crucial step for acquiring their phenotypic traits. Unlike organ-level segmentation tasks, current methods for segmenting plots into individual plants largely depend on traditional machine learning algorithms, with end-to-end deep learning models being relatively scarce.

Conventional methodologies, such as region-growing algorithms [94], Euclidean clustering algorithms [95], and k-means clustering algorithms [96], have served as important facilitators in the individual plant point cloud segmentation process. In this regard, Lin et al. proposed a columnar space clustering segmentation method for singularly segmenting the plant species of maize, cotton, and oilseed rape [97]. These tactics tend to be predicated on certain predetermined feature information. With the growing adoption of deep learning, certain deep learning-based models have reduced the reliance on such a priori features. For instance, Jin et al. introduced a segmentation strategy that integrates Faster R-CNN with the aforementioned region-growing algorithms to segregate individual maize plants existing in a field context [98]. Despite this advancement, the algorithm still requires synchronization with conventional methodologies to achieve optimal performance, warranting a need for a comprehensive end-to-end network model to independently handle individual plant segmentation. In this experiment, the potential of Mask3D for segmenting plot maize was validated, demonstrating its effectiveness in segmenting individual plant-level point clouds. Compared to organ-level segmentation, extracting individual plants often necessitates a

comprehensive exploration of the entire regional point cloud information. By employing sparse convolutions, Mask3D reduces the computational cost, enabling feature extraction from the entire plot point cloud. However, this experiment was limited to maize, and further exploration is required for crops with more complex morphological structures.

6. Conclusions

We conducted in-depth comparative experimental analyses on several typical deep learning models. To evaluate the models' segmentation performance across diverse environments and data generated by different platforms, we have summarized the existing public datasets, provided detailed descriptions, and identified the shortcomings of the current datasets: limited plant species, small sample size, single data collection method, and overly ideal collection environment. We conducted data acquisition using terrestrial laser scanning and UAV laser scanning, and reconstructed point clouds from RGB images. The datasets comprised point cloud data from various crops, including maize, cotton, potato, and rapeseed. These datasets were subsequently utilized to assess the models' practical application effectiveness.

At the organ-level segmentation, point cloud data acquisition and generation can be divided into five categories: indoor laser scanning, indoor 3D reconstruction from images, outdoor potted terrestrial laser scanning, field terrestrial laser scanning, and 3D reconstruction from UAV RGB images. The experiments show that ST and PT have achieved the best and second-best results, respectively, but require significant computational resources. PAConv and PointCNN have great potential on crop objects with simpler morphological structures. Based on the experimental results of the nine typical models, we have conducted in-depth experiments and comparative analyses of the models' performance, focusing on three key steps of model composition: local operations, grouping methods, and feature aggregation methods. In this way, we have explored the reasons for the advantages and disadvantages of model performance and identified the inherent challenges of plant organ segmentation tasks: imbalanced class data, data loss caused by occlusion, and unavoidable noise.

Additionally, the potential application of the classic scene instance segmentation model Mask3D on maize point clouds collected by UAV-LS was validated. Mask3D achieved commendable performance, with a mean Average Precision (mAP) of 0.909. This finding supports the development of automated segmentation methods from plot point clouds to individual plant point clouds. Unlike traditional point cloud instance segmentation methods, Mask3D does not require the pre-computation of complex features, which greatly facilitates the segmentation of field maize. Research on end-to-end algorithms for segmenting individual point clouds from plot point clouds remains relatively scarce and warrants further exploration.

In future research, several key points warrant attention. First, the lack of automated methods for segmenting individual plant point clouds from plot point clouds limits high-throughput phenotyping in field scenarios. Second, robust point cloud segmentation models with strong feature extraction capabilities and the ability to adapt to multi-scale information are required. Additionally, in the realm of deep learning models, the downsampling of point clouds presents a critical issue. The number of input points is linearly related to the model's computational cost. Therefore, determining an appropriate threshold is a topic that merits investigation.

Author Contributions: R.Z. designed the work and wrote the manuscript. K.X. performed experiments, analyzed the data, and wrote the manuscript. J.Z., H.R. and Y.W. performed experiments and analyzed the data. W.Y., G.C. and C.L. revised the manuscript. R.Z. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (2023YFF1000100, 2022YFD2002304).

Data Availability Statement: The point cloud data in this study are available upon request by contacting the corresponding author.

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

References

1. Thiébaut, M. A foliar morphometric approach to the study of salicaceae. *Bot. Rev.* **2000**, *66*, 423–439. [CrossRef]
2. Gaertner, H.; Wagner, B.; Heinrich, I.; Denier, C. 3D-laser scanning: A new method to analyze coarse tree root systems. *For. Snow Landsc. Res.* **2009**, *82*, 95–106.
3. Omasa, K.; Hosoi, F.; Konishi, A. 3D lidar imaging for detecting and understanding plant responses and canopy structure. *J. Exp. Bot.* **2007**, *58*, 881–898. [CrossRef]
4. Vázquez-Arellano, M.; Reiser, D.; Paraforos, D.S.; Garrido-Izard, M.; Burce, M.E.C.; Griepentrog, H.W. 3-D reconstruction of maize plants using a time-of-flight camera. *Comput. Electron. Agric.* **2018**, *145*, 235–247. [CrossRef]
5. Ziamtsov, I.; Navlakha, S. Plant 3D (P3D): A plant phenotyping toolkit for 3D point clouds. *Bioinformatics* **2020**, *36*, 3949–3950. [CrossRef]
6. Li, Z.; Guo, R.; Li, M.; Chen, Y.; Li, G. A review of computer vision technologies for plant phenotyping. *Comput. Electron. Agric.* **2020**, *176*, 105672. [CrossRef]
7. Sirault, X.; Fripp, J.; Paproki, A.; Kuffner, P.; Nguyen, C.V.; Li, R.; Daily, H.; Guo, J.; Furbank, R.T. PlantScan: A three-dimensional phenotyping platform for capturing the structural dynamic of plant development and growth. In Proceedings of the 7th International Conference on Functional-Structural Plant Models, Saariselkä, Finland, 9–14 June 2013; pp. 45–48.
8. Xiang, L.; Bao, Y.; Tang, L.; Ortiz, D.; Salas-Fernandez, M.G. Automated morphological traits extraction for sorghum plants via 3D point cloud data analysis. *Comput. Electron. Agric.* **2019**, *162*, 951–961. [CrossRef]
9. Ao, Z.; Wu, F.; Hu, S.; Sun, Y.; Su, Y.; Guo, Q.; Xin, Q. Automatic segmentation of stem and leaf components and individual maize plants in field terrestrial LiDAR data using convolutional neural networks. *Crop J.* **2022**, *10*, 1239–1250. [CrossRef]
10. Li, Y.; Fan, X.; Mitra, N.J.; Chamovitz, D.; Cohen-Or, D.; Chen, B. Analyzing Growing Plants from 4D Point Cloud Data. *ACM Trans. Graph.* **2013**, *32*, 157. [CrossRef]
11. Dutagaci, H.; Rasti, P.; Galopin, G.; Rousseau, D. ROSE-X: An annotated data set for evaluation of 3D plant organ segmentation methods. *Plant Methods* **2020**, *16*, 28. [CrossRef]
12. Dowd, T.; McInturf, S.; Li, M.; Topp, C.N. Rated-M for mesocosm: Allowing the multimodal analysis of mature root systems in 3D. *Emerg. Top. Life Sci.* **2021**, *5*, 249–260. [CrossRef] [PubMed]
13. Gao, M.; Yang, F.; Wei, H.; Liu, X. Individual Maize Location and Height Estimation in Field from UAV-Borne LiDAR and RGB Images. *Remote Sens.* **2022**, *14*, 2292. [CrossRef]
14. Nie, S.; Wang, C.; Zeng, H.; Xi, X.; Li, G. Above-ground biomass estimation using airborne discrete-return and full-waveform LiDAR data in a coniferous forest. *Ecol. Indic.* **2017**, *78*, 221–228. [CrossRef]
15. Ben Hmida, S.; Kallel, A.; Gastellu-Etchegorry, J.-P.; Roujean, J.-L. Crop biophysical properties estimation based on LiDAR full-waveform inversion using the DART RTM. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4853–4868. [CrossRef]
16. Ben Hmida, S.; Kallel, A.; Gastellu-Etchegorry, J.-P.; Roujean, J.-L.; Zribi, M. Lidar full waveform inversion to estimate maize and wheat crops biophysical properties. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 1316–1319.
17. Nie, S.; Wang, C.; Dong, P.; Xi, X. Estimating leaf area index of maize using airborne full-waveform lidar data. *Remote Sens. Lett.* **2016**, *7*, 111–120. [CrossRef]
18. Qin, H.; Wang, C.; Pan, F.; Lin, Y.; Xi, X.; Luo, S. Estimation of FPAR and FPAR profile for maize canopies using airborne LiDAR. *Ecol. Indic.* **2017**, *83*, 53–61. [CrossRef]
19. Gao, S.; Niu, Z.; Sun, G.; Zhao, D.; Jia, K.; Qin, Y. Height extraction of maize using airborne full-waveform LIDAR data and a deconvolution algorithm. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1978–1982. [CrossRef]
20. Fareed, N.; Das, A.K.; Flores, J.P.; Mathew, J.J.; Mukaila, T.; Numata, I.; Janjua, U.U.R. UAS Quality Control and Crop Three-Dimensional Characterization Framework Using Multi-Temporal LiDAR Data. *Remote Sens.* **2024**, *16*, 699. [CrossRef]
21. Dey, D.; Mummert, L.; Sukthankar, R. Classification of plant structures from uncalibrated image sequences. In Proceedings of the 2012 IEEE Workshop on Applications of Computer Vision (WACV), Breckenridge, CO, USA, 9–11 January 2012; pp. 329–336. [CrossRef]
22. Santos, T.T.; Koenigkan, L.V.; Barbedo, J.G.A.; Rodrigues, G.C. 3D Plant modeling: Localization, mapping and segmentation for plant phenotyping using a single hand-held camera. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 247–263. [CrossRef]
23. Paproki, A.; Sirault, X.; Berry, S.; Furbank, R.; Fripp, J. A novel mesh processing based technique for 3D plant analysis. *BMC Plant Biol.* **2012**, *12*, 63. [CrossRef]
24. Gaillard, M.; Miao, C.; Schnable, J.C.; Benes, B. Voxel carving-based 3D reconstruction of sorghum identifies genetic determinants of light interception efficiency. *Plant Direct* **2020**, *4*, e00255. [CrossRef]

25. Scharr, H.; Briese, C.; Embgenbroich, P.; Fischbach, A.; Fiorani, F.; Müller-Linow, M. Fast High Resolution Volume Carving for 3D Plant Shoot Reconstruction. *Front. Plant Sci.* **2017**, *8*, 1680. [CrossRef] [PubMed]
26. Guo, J.; Xu, L. Automatic Segmentation for plant leaves via multiview stereo reconstruction. *Math. Probl. Eng.* **2017**, *2017*, 9845815. [CrossRef]
27. Jin, S.; Su, Y.; Wu, F.; Pang, S.; Gao, S.; Hu, T.; Liu, J.; Guo, Q. Stem–leaf segmentation and phenotypic trait extraction of individual maize using terrestrial LiDAR data. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1336–1346. [CrossRef]
28. Li, D.; Cao, Y.; Tang, X.-S.; Yan, S.; Cai, X. Leaf segmentation on dense plant point clouds with facet region growing. *Sensors* **2018**, *18*, 3625. [CrossRef]
29. Fang, L.; He, N.; Li, S.; Plaza, A.J.; Plaza, J. A new spatial–spectral feature extraction method for hyperspectral images using local covariance matrix representation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3534–3546. [CrossRef]
30. Elnashef, B.; Filin, S.; Lati, R.N. Tensor-based classification and segmentation of three-dimensional point clouds for organ-level plant phenotyping and growth analysis. *Comput. Electron. Agric.* **2019**, *156*, 51–61. [CrossRef]
31. Paulus, S.; Dupuis, J.; Mahlein, A.-K.; Kuhlmann, H. Surface feature based classification of plant organs from 3D laserscanned point clouds for plant phenotyping. *BMC Bioinform.* **2013**, *14*, 238. [CrossRef]
32. Paulus, S.; Dupuis, J.; Riedel, S.; Kuhlmann, H. Automated analysis of barley organs using 3d laser scanning: An approach for high throughput phenotyping. *Sensors* **2014**, *14*, 12670–12686. [CrossRef] [PubMed]
33. Wahabzada, M.; Paulus, S.; Kersting, K.; Mahlein, A.-K. Automated interpretation of 3D laserscanned point clouds for plant organ segmentation. *BMC Bioinform.* **2015**, *16*, 248. [CrossRef]
34. Gelard, W.; Devy, M.; Herbulot, A.; Burger, P. Model-based segmentation of 3D point clouds for phenotyping sunflower plants. In Proceedings of the 12th International Conference on Computer Vision Theory and Applications, Porto, Portugal, 27 February–1 March 2017; pp. 459–467. [CrossRef]
35. Gélard, W.; Herbulot, A.; Devy, M.; Debaeke, P.P.; McCormick, R.F.; Truong, S.K.; Mullet, J.E. Leaves segmentation in 3D point cloud. In Proceedings of the Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, 18–21 September 2017; pp. 664–674. [CrossRef]
36. Sun, S.; Li, C.; Paterson, A.; Chee, P. Three-dimensional cotton plant shoot architecture segmentation and phenotypic trait characterization using terrestrial LiDAR point cloud data. In Proceedings of the 2020 ASABE Annual International Virtual Meeting, Virtual, 13–15 July 2020; p. 1.
37. Liu, J.; Liu, Y.; Doonan, J. Point cloud based iterative segmentation technique for 3D plant phenotyping. In Proceedings of the 2018 IEEE International Conference on Information and Automation (ICIA), Wuyishan, China, 11–13 August 2018; pp. 1072–1077. [CrossRef]
38. Itakura, K.; Hosoi, F. automatic leaf segmentation for estimating leaf area and leaf inclination angle in 3D plant images. *Sensors* **2018**, *18*, 3576. [CrossRef] [PubMed]
39. Kuo, Y.-T.; Teng, C.-H.; Chen, Y.-S. Leaf segmentation, classification, and three-dimensional recovery from a few images with close viewpoints. *Opt. Eng.* **2011**, *50*, 037003. [CrossRef]
40. Ma, Z.; Du, R.; Xie, J.; Sun, D.; Fang, H.; Jiang, L.; Cen, H. Phenotyping of silique morphology in oilseed rape using skeletonization with hierarchical segmentation. *Plant Phenom.* **2023**, *5*, 0027. [CrossRef]
41. Japes, B.; Mack, J.; Rist, F.; Herzog, K.; Töpfer, R.; Steinhage, V. Multi-View semantic labeling of 3D point clouds for auto-mated plant phenotyping. *arXiv* **2018**, arXiv:1805.03994.
42. Shi, W.; van de Zedde, R.; Jiang, H.; Kootstra, G. Plant-part segmentation using deep learning and multi-view vision. *Biosyst. Eng.* **2019**, *187*, 81–95. [CrossRef]
43. Ni, X.; Li, C.; Jiang, H.; Takeda, F. Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 297–309. [CrossRef]
44. Jin, S.; Su, Y.; Zhao, X.; Hu, T.; Guo, Q. A point-based fully convolutional neural network for airborne LiDAR ground point filtering in forested environments. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3958–3974. [CrossRef]
45. Zhou, Y.; Sun, P.; Zhang, Y.; Anguelov, D.; Gao, J.; Ouyang, T.Y.; Guo, J.; Ngiam, J.; Vasudevan, V. End-to-End multi-view fusion for 3D object detection in LiDAR point clouds. In Proceedings of the Conference on Robot Learning, PMLR 100, Cambridge, MA, USA, 16–18 November 2020; pp. 923–932.
46. 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]
47. Charles, R.Q.; Su, H.; Kaichun, M.; 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]
48. Qi, C.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5105–5114.
49. Louedec, J.; Li, B.; Cielniak, G. Evaluation of 3D vision systems for detection of small objects in agricultural environments. In Proceedings of the 15th International Conference on Computer Vision Theory and Applications, Valetta, Malta, 27–29 February 2020; pp. 628–689. [CrossRef]

50. Kang, H.; Zhou, H.; Wang, X.; Chen, C. Real-time fruit recognition and grasping estimation for robotic apple harvesting. *Sensors* **2020**, *20*, 5670. [CrossRef]
51. 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), Virtual, 11–17 October 2021; pp. 1381–1389. [CrossRef]
52. Heiwolt, K.; Duckett, T.; Cielniak, G. Deep semantic segmentation of 3D plant point clouds. In Proceedings of the Towards Autonomous Robotic Systems: 22nd Annual Conference, Lincoln, UK, 8–10 September 2021; pp. 36–45. [CrossRef]
53. Li, Y.; Wen, W.; Miao, T.; Wu, S.; Yu, Z.; Wang, X.; Guo, X.; Zhao, C. Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning. *Comput. Electron. Agric.* **2022**, *193*, 106702. [CrossRef]
54. Lai, Y.; Lu, S.; Qian, T.; Chen, M.; Zhen, S.; Guo, L. Segmentation of plant point cloud based on deep learning method. *Comput. Des. Appl.* **2022**, *19*, 1117–1129. [CrossRef]
55. 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*, 146. [CrossRef]
56. 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]
57. Li, Y.; Bu, R.; Sun, M.; Wu, W.; Di, X.; Chen, B. PointCNN: Convolution on X-transformed points. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 2–8 December 2018; pp. 828–838.
58. 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 Phenom.* **2021**, *2021*, 9838929. [CrossRef]
59. Li, D.; Li, J.; Xiang, S.; Pan, A. PSegNet: Simultaneous semantic and instance segmentation for point clouds of plants. *Plant Phenom.* **2022**, *2022*, 9787643. [CrossRef]
60. 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]
61. Ma, X.; Qin, C.; You, H.; Ran, H.; Fu, Y.R. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. *arXiv* **2022**, arXiv:2202.07123.
62. Boogaard, F.P.; van Henten, E.J.; Kootstra, G. Boosting plant-part segmentation of cucumber plants by enriching incomplete 3D point clouds with spectral data. *Biosyst. Eng.* **2021**, *211*, 167–182. [CrossRef]
63. Guo, X.; Sun, Y.; Yang, H. FF-Net: Feature-fusion-based network for semantic segmentation of 3D plant point cloud. *Plants* **2023**, *12*, 1867. [CrossRef]
64. Xiang, T.; Zhang, C.; Song, Y.; Yu, J.; Cai, W. Walk in the cloud: Learning curves for point clouds shape analysis. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 895–904. [CrossRef]
65. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
66. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.; Koltun, V. Point transformer. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 16239–16248. [CrossRef]
67. Mao, J.; Xue, Y.; Niu, M.; Bai, H.; Feng, J.; Liang, X.; Xu, H.; Xu, C. Voxel Transformer for 3D Object Detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 3144–3153. [CrossRef]
68. Lai, X.; Liu, J.; Jiang, L.; Wang, L.; Zhao, H.; Liu, S.; Qi, X.; Jia, J. Stratified Transformer for 3D Point Cloud Segmentation. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 8490–8499. [CrossRef]
69. Guo, R.; Xie, J.; Zhu, J.; Cheng, R.; Zhang, Y.; Zhang, X.; Gong, X.; Zhang, R.; Wang, H.; Meng, F. Improved 3D point cloud segmentation for accurate phenotypic analysis of cabbage plants using deep learning and clustering algorithms. *Comput. Electron. Agric.* **2023**, *211*, 108014. [CrossRef]
70. 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]
71. Sun, Y.; Guo, X.; Yang, H. Win-Former: Window-based transformer for maize plant point cloud semantic segmentation. *Agronomy* **2023**, *13*, 2723. [CrossRef]
72. 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 Phenom.* **2023**, *5*, 0080. [CrossRef] [PubMed]
73. Lin, Y. LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics? *Comput. Electron. Agric.* **2015**, *119*, 61–73. [CrossRef]
74. Hosoi, F.; Nakabayashi, K.; Omasa, K. 3-D modeling of tomato canopies using a high-resolution portable scanning lidar for extracting structural information. *Sensors* **2011**, *11*, 2166–2174. [CrossRef] [PubMed]
75. Tilly, N.; Hoffmeister, D.; Cao, Q.; Huang, S.; Lenz-Wiedemann, V.; Miao, Y.; Bareth, G. Multitemporal crop surface models: Accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice. *J. Appl. Remote Sens.* **2014**, *8*, 083671. [CrossRef]

76. Tilly, N.; Hoffmeister, D.; Schiedung, H.; Hütt, C.; Brands, J.; Bareth, G. Terrestrial laser scanning for plant height measurement and biomass estimation of maize. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 181–187. [CrossRef]
77. Eitel, J.U.; Magney, T.S.; Vierling, L.A.; Brown, T.T.; Huggins, D.R. LiDAR based biomass and crop nitrogen estimates for rapid, non-destructive assessment of wheat nitrogen status. *Field Crop Res.* **2014**, *159*, 21–32. [CrossRef]
78. Zheng, G.; Moskal, L.M.; Kim, S.-H. Retrieval of effective leaf area index in heterogeneous forests with terrestrial laser scanning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 777–786. [CrossRef]
79. Su, W.; Zhu, D.; Huang, J.; Guo, H. Estimation of the vertical leaf area profile of corn (*Zea mays*) plants using terrestrial laser scanning (TLS). *Comput. Electron. Agric.* **2018**, *150*, 5–13. [CrossRef]
80. Shendryk, Y.; Sofonia, J.; Garrard, R.; Rist, Y.; Skocaj, D.; Thorburn, P. Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102177. [CrossRef]
81. Luo, S.; Liu, W.; Zhang, Y.; Wang, C.; Xi, X.; Nie, S.; Ma, D.; Lin, Y.; Zhou, G. Maize and soybean heights estimation from unmanned aerial vehicle (UAV) LiDAR data. *Comput. Electron. Agric.* **2021**, *182*, 106005. [CrossRef]
82. Zhou, L.; Gu, X.; Cheng, S.; Yang, G.; Shu, M.; Sun, Q. Analysis of Plant Height Changes of Lodged Maize Using UAV-LiDAR Data. *Agriculture* **2020**, *10*, 146. [CrossRef]
83. Hu, X.; Sun, L.; Gu, X.; Sun, Q.; Wei, Z.; Pan, Y.; Chen, L. Assessing the Self-Recovery Ability of Maize after Lodging Using UAV-LiDAR Data. *Remote Sens.* **2021**, *13*, 2270. [CrossRef]
84. Miao, T.; Wen, W.; Li, Y.; Wu, S.; Zhu, C.; Guo, X. Label3DMaize: Toolkit for 3D point cloud data annotation of maize shoots. *GigaScience* **2021**, *10*, giab031. [CrossRef]
85. Conn, A.; Pedmale, U.V.; Chory, J.; Stevens, C.F.; Navlakha, S. A Statistical Description of Plant Shoot Architecture. *Curr. Biol.* **2017**, *27*, 2078–2088.e3. [CrossRef]
86. Schunck, D.; Magistri, F.; Rosu, R.A.; Cornelißen, A.; Chebrolu, N.; Paulus, S.; Léon, J.; Behnke, S.; Stachniss, C.; Kuhlmann, H.; et al. Pheno4D: A spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis. *PLoS ONE* **2021**, *16*, e0256340. [CrossRef]
87. 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]
88. James, K.M.; Heiwolt, K.; Sargent, D.J.; Cielniak, G. Lincoln's Annotated Spatio-Temporal Strawberry Dataset (LAST-Straw). *arXiv* **2024**, arXiv:abs/2403.00566.
89. Hu, S.; Zhai, R.; Wang, Y.; Liu, Z.; Zhu, J.; Ren, H.; Yang, W.; Song, P. Extraction of Potato Plant Phenotypic Parameters Based on Multi-Source Data. *Smart Agric.* **2023**, *5*, 132–145. [CrossRef]
90. Wang, Y.; Hu, S.; Ren, H.; Yang, W.; Zhai, R. 3DPhenoMVS: A Low-Cost 3D Tomato Phenotyping Pipeline Using 3D Reconstruction Point Cloud Based on Multiview Images. *Agronomy* **2022**, *12*, 1865. [CrossRef]
91. Xu, M.; Ding, R.; Zhao, H.; Qi, X. PAConv: Position adaptive convolution with dynamic kernel assembling on point clouds. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 3172–3181. [CrossRef]
92. Schult, J.; Engelmann, F.; Hermans, A.; Litany, O.; Tang, S.; Leibe, B. Mask3D: Mask Transformer for 3D Semantic Instance Segmentation. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 8216–8223.
93. Ren, J.; Pan, L.; Liu, Z. Benchmarking and analyzing point cloud classification under corruptions. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 18559–18575.
94. Li, W.; Guo, Q.; Jakubowski, M.K.; Kelly, M. A new method for segmenting individual trees from the LiDAR point cloud. *Photogramm. Eng. Remote Sens.* **2012**, *78*, 75–84. [CrossRef]
95. Liang, X.; Zhou, F.; Chen, H.; Liang, B.; Xu, X.; Yang, W. Three-dimensional maize plants reconstruction and traits extraction based on structure from motion. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 209–219.
96. Miao, Y.; Wang, L.; Peng, C.; Li, H.; Li, X.; Zhang, M. Banana plant counting and morphological parameters measurement based on terrestrial laser scanning. *Plant Methods* **2022**, *18*, 66. [CrossRef]
97. Lin, C.; Han, J.; Xie, L.; Hu, F. Cylinder space segmentation method for field crop population using 3D point cloud. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 175–182.
98. Jin, S.; Su, Y.; Gao, S.; Wu, F.; Hu, T.; Liu, J.; Li, W.; Wang, D.; Chen, S.; Jiang, Y.; et al. Deep Learning: Individual Maize Segmentation from Terrestrial Lidar Data Using Faster R-CNN and Regional Growth Algorithms. *Front. Plant Sci.* **2018**, *9*, 866. [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.