

## 3D-based precise evaluation pipeline for maize ear rot using multi-view stereo reconstruction and point cloud semantic segmentation

Rui Yang<sup>a</sup>, Yong He<sup>a</sup>, Xiangyu Lu<sup>a</sup>, Yiyi Zhao<sup>b</sup>, Yanmei Li<sup>c</sup>, Yinhui Yang<sup>d</sup>, Wenwen Kong<sup>d</sup>, Fei Liu<sup>a,\*</sup>

<sup>a</sup> College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

<sup>b</sup> Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

<sup>c</sup> National Maize Improvement Centre of China, China Agricultural University, Beijing 100193, China

<sup>d</sup> College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China



### ARTICLE INFO

**Keywords:**

Maize ear rot  
Multi-view stereo reconstruction  
Point cloud  
3D semantic segmentation  
Deep learning

### ABSTRACT

Maize ear rot poses a severe threat to maize yield and quality. Breeding and cultivating highly resistant maize varieties is a crucial approach for preventing and controlling maize ear rot. However, traditional methods of visually grading the severity of maize ear infection and resistance lack objectivity and repeatability. To meet the requirement of precise breeding and resistance assessment scenarios, a novel pipeline based on three-dimensional (3D) point clouds of maize ear was developed for ear rot precise evaluation. First, multi-view stereo (MVS) reconstruction was employed to obtain high-precision dense point clouds of maize ears. And the coordinate correction and circular sampling approaches were proposed to optimize the data structure of the input maize ear samples. Next, a specialized network called the ear rot segmentation network (**ERSegNet**) was proposed to detect the infected area of maize ears. This network incorporated an orientation-encoding (OE) module and point transformer (PT) attention, which effectively boosted the performance of **PointNet++**. The proposed ERSegNet achieved impressive results, including a mean intersection over union (mIoU) of 85.83%, a mean precision (mPrec) of 92.34%, a mean recall (mRec) of 92.23%, a mean F1-score of 92.28%, and an overall accuracy (OA) of 93.76%. This demonstrated the feasibility of using semantic segmentation algorithms to predict 3D point clouds of maize ears. Furthermore, a point cloud resampling method was suggested to enhance the spatial uniformity of maize ear point clouds and a point-level quantitative assessment approach based on the 3D point cloud data was provided for evaluating the severity of ear rot. The results showed an average evaluation error of 1.55% in the testing set, indicating the accuracy of the proposed method. This study provides a reliable and objective method for maize ear rot precise assessment, offering potential and valuable support for the identification of resistant varieties in breeding programs.

### 1. Introduction

Maize ear rot, a fungal disease prevalent in maize-growing regions worldwide, poses a significant threat to maize production (Naz et al., 2021). It is caused by various fungi that infect maize ears and kernels, leading to mold, rot, and the production of harmful toxins by pathogenic bacteria, thereby compromising human and animal health and safety (Dong et al., 2018; Tran et al., 2021). Controlling maize ear rot solely through pesticide spraying is challenging due to its diverse infection pathways (Thompson and Raizada, 2018). Currently, breeding and cultivating resistant maize varieties have proven to be the most effective

and economical approach for prevention and control, making it a primary goal in maize breeding (Rose et al., 2017). The assessment of maize ear rot severity, crucial for breeding outcomes, traditionally relies on visual estimation methods in intensive breeding programs (Zhu et al., 2021). This involves visually estimating the percentage of the ear surface or kernels exhibiting disease symptoms and rating the severity on ordinal scales, such as 1–5, 1–7, or 0–9 (Mesterhazy et al., 2012). However, relying solely on severity levels for ear incidence evaluation can be challenging, as it may not provide accurate results or sufficient precision for breeding reference (Reynolds et al., 2020). Despite this, the evaluation based on the severity levels remains a primary method for

\* Corresponding author.

E-mail address: [fliu@zju.edu.cn](mailto:fliu@zju.edu.cn) (F. Liu).

disease analysis and resistance gene identification in various studies (Dong et al., 2018; Gaikpa et al., 2021; Yao et al., 2020).

With increasing demands for accuracy in breeding resistance identification, some studies have conducted scoring from 0 to 100% based on visual evaluation and infected ear area proportion (Galiano-Carneiro et al., 2020; Liu et al., 2021). Still, based on visual evaluation methods, there are significant subjective factors that make it difficult to avoid errors caused by human sensory factors. Wen et al. (2021a) proposed an image analysis method, photographing the most severely affected parts of maize ears. They used Photoshop to analyze infected and uninfected areas, and the percentage of infected area in the total area was calculated. Hudson et al. (2023) explored the feasibility of quantitative evaluation of maize ear rot by concatenating single row pixels from multi-view images of rotated maize and calculating the number of pixels. The methods based on two-dimensional (2D) images promote objectivity and repeatability in the evaluation of ear rot disease. The accuracy of pixel-level evaluation on images depends on the mapping of three-dimensional (3D) maize ears to 2D images. However, the representation accuracy of 2D images for 3D objects has dimensional limitations, which makes it difficult to further improve the evaluation accuracy in precise phenotyping (Zhang et al., 2023).

3D reconstruction technology enables efficiently construct 3D models based on real-world objects (Long et al., 2021). Phenotype analysis based on 3D reconstruction has become a popular trend in plant digitization (Sun et al., 2022), making the assessment accuracy from the pixel-level of a plane to the point-level of 3D. In 3D phenotype evaluation of plants, 3D laser scanners (Miao et al., 2022), multi-view stereo (MVS) reconstruction (Wu et al., 2020), and time-of-flight (TOF) cameras (Bao et al., 2019) are commonly used to acquire 3D point cloud data of plants. Compared to lower dimensional data, 3D data provides clear details and intuitively represents objects' shape, connectivity, and spatial coordinates (Wen et al., 2021b), providing strong support for plant phenotype analysis. Wang et al. (2019b) compared three 3D data acquisition methods for maize plant 3D phenotyping: 3D laser scanners, MVS, and 3D digitizers. The results showed that MVS was suitable for 3D reconstruction of small and medium-sized plants, and the reconstruction method based on 2D images was low-cost, efficient, and practical. Wu et al. (2020) developed an affordable MVS-based platform for individual maize plant phenotyping. The extracted maize plant height, leaf width, and leaf area achieved  $R^2$  values of 0.99, 0.87, and 0.93 respectively, validating the application potential of MVS for plant phenotypic extraction.

Benefiting from rapid progress in deep learning for 3D point cloud data processing, the analysis path for plant point cloud data has been greatly expanded and the model prediction accuracy and robustness have also been improved effectively (Ding et al., 2023; Fei et al., 2022). Guo et al. (2023) implemented organ-level point cloud segmentation on cabbage point cloud datasets obtained with multi-view images and structure from motion (SFM) method based on an improved PointNet++ algorithm. The accuracy and intersection over union (IoU) of semantic segmentation reached 95% and 86%, respectively. Turgut et al. (2022) applied an attention-based deep learning framework for the segmentation of rosebush plant point clouds, improving mIoU by 4% over PointNet++. Du et al. (2023) used handheld laser scanning to obtain high-resolution rapeseed plant point clouds. They proposed a plant segmentation transformer network for tiny siliques point cloud semantic segmentation, achieving an mIoU of 93.96% and overall accuracy (OA) of 97.07%. Maize ears differ from most plants in that plant point clouds cover entire plant morphology, often with intuitive spatial structure differences between leaves and stems. Maize ear point clouds cover ear surfaces with high point density, requiring comprehensive scanning from multiple perspectives for accurate details. Despite regular kernel arrangement, samples infected with maize ear rot exhibit complex information features such as kernel discoloration and surface mycelia coverage (Dong et al., 2018). Furthermore, the morphological variations among different maize varieties pose challenges in employing 3D

semantic segmentation for recognizing infected areas in maize ear point clouds. Additionally, determining how to convert the semantic segmented point cloud into susceptibility assessment for ear rot is also a crucial consideration.

Therefore, in response to current limitations and challenges in evaluating maize ear rot, this study proposes a novel maize ear rot precise evaluation pipeline based on 3D semantic segmentation for maize ear point clouds. The main contributions are: (1) Based on MVS reconstruction, a well-labeled 3D point cloud dataset containing 100 maize ears with varying shapes and infected severity is constructed; (2) To optimize the input data structure, coordinate correction and circular sampling approaches are developed based on maize ear morphology and structure. Furthermore, local feature aggregation and self-attention modules are introduced to design ear rot segmentation network (ERSegNet) for enhancing 3D semantic segmentation performance of maize ears. And the feasibility of using the 3D semantic segmentation algorithm for predicting maize ear point cloud is also verified; (3) A point cloud resampling method is proposed to address the issue of evaluating maize ear rot using point clouds. This method improves the spatial structure uniformity of maize ear point clouds and provides support for accurate evaluation of maize ear rot at the point-level.

## 2. Materials and methods

### 2.1. Maize ear samples

The maize samples infected with ear rot disease were collected from the Hainan winter nursery (Sanya, Hainan, N18°36', E109°21'), which were provided by Professor Mingliang Xu (China Agricultural University). They were sown on November 12, 2022, and harvested on March 4, 2023. The plants were the progeny population of the resistant inbred line CML304 and the susceptible inbred line B73. And the plants were artificially inoculated on their kernels with *Fusarium verticillioides* at approximately 10–15 d after pollination in the field. The inoculated maize ear samples containing a range of infected severity, different ear sizes and shapes were manually selected, ensuring the diversity of the final dataset (Fig. 1).

### 2.2. Maize ear rot precise evaluation pipeline

The maize ear rot evaluation pipeline in this study consists of four components, including data acquisition, data preprocessing, maize ear semantic segmentation, and quantitative assessment of ear rot, as shown in Fig. 2.

#### 2.2.1. Data acquisition

The data collection and MVS reconstruction were completed on March 14, 2023. To ensure accurate reconstruction of the maize ear point clouds, the maize ears were positioned securely at a fixed height using a support base. Then, a smartphone capable of collecting at least  $1920 \times 1080$  resolution was used to capture 50 images as equidistant as possible by circling around the maize ears twice, covering both the top and bottom views. This image acquisition process took approximately 30 s. Once the multi-view images of the maize ears were obtained, high-precision MVS reconstruction was performed using Agisoft Metashape (Agisoft LLC., St. Petersburg, Russia). This software generated 3D point cloud data with 9 channels, including coordinates (X, Y, Z), colors (R, G, B), and surface normals ( $N_x$ ,  $N_y$ ,  $N_z$ ). The reconstruction of a single maize ear point cloud took no more than 60 s. The raw 3D point cloud data was then processed with algorithms such as statistical filtering and Random sample consensus (RANSAC) (Fischler and Bolles, 1981) for noise reduction. Finally, CloudCompare (<https://cloudcompare.org/>) was employed for annotating the maize ear point cloud data.

The maize ear points were categorized into three classes, including non-kernel, uninfected kernel, and infected kernel classes. The non-kernel class represented points that did not contain any maize kernels.



Fig. 1. Representative maize ear rot samples.

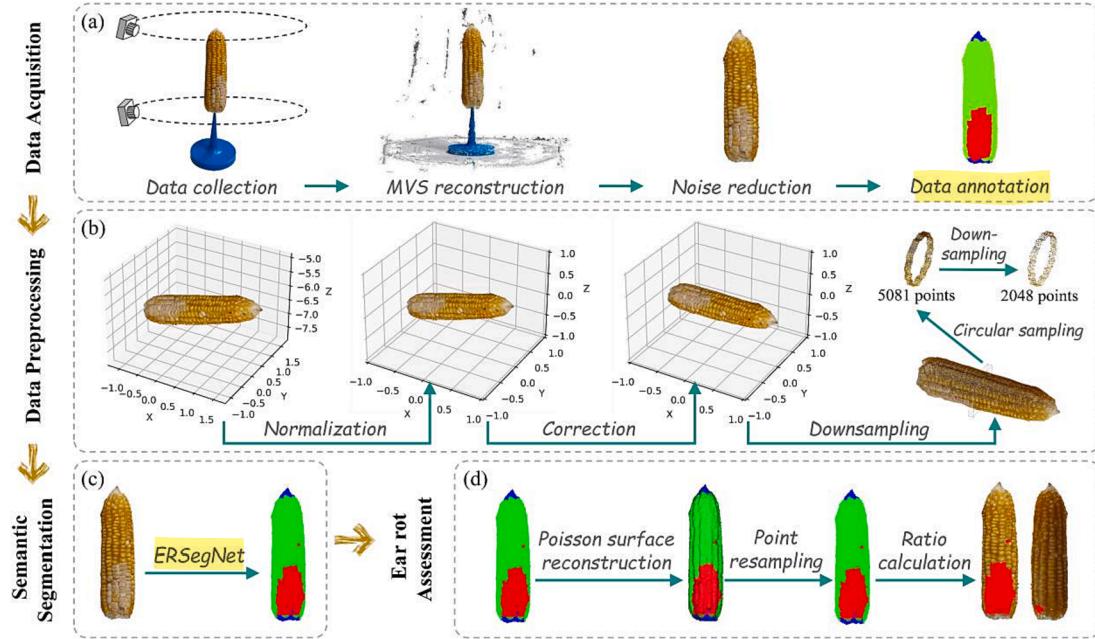


Fig. 2. Illustration of maize ear rot precise evaluation pipeline. (a) Workflow of data acquisition; (b) Data preprocessing; (c) Maize ear semantic segmentation; (d) Maize ear rot quantitative assessment.

All maize kernels without any infected symptoms were considered as the uninfected kernel class. Points within the kernel area covered by mycelium or with color changes were labeled as infected kernel class (Mesterhazy et al., 2012).

### 2.2.2. Data preprocessing

Data preprocessing mainly includes coordinate normalization, coordinate correction, and data downsampling, as shown in Fig. 2(b). Due to the different sizes and shapes of maize, coordinate normalization was first adopted to efficiently learn features and fit labels. This centered the sample at the origin and scaled the maize point cloud coordinates to the range [-1, 1]. In terms of the inconsistent orientation of samples during the maize ear data acquisition process, the point cloud was rotated through coordinate correction to align the direction of all maize ear samples with the x-axis. The key to coordinate correction of maize ear point clouds lies in the rotation matrix.

Given a 3D point cloud  $P$  containing 3D coordinates and normals,  $P$  is initially projected onto the xy plane and xz plane respectively to obtain two 2D point sets,  $P_{xy}$  and  $P_{xz}$ . Subsequently, the rotation angles  $\theta_{xy}$  and  $\theta_{xz}$  are calculated to rotate the two 2D point sets towards the x-axis direction based on their respective minimum bounding boxes. With these two rotation angles, the 3D rotation matrix  $R$  can be obtained to achieve x-axis coordinate correction of the 3D point cloud  $P$ . The point cloud  $P_R$  after coordinate correction can be calculated as:

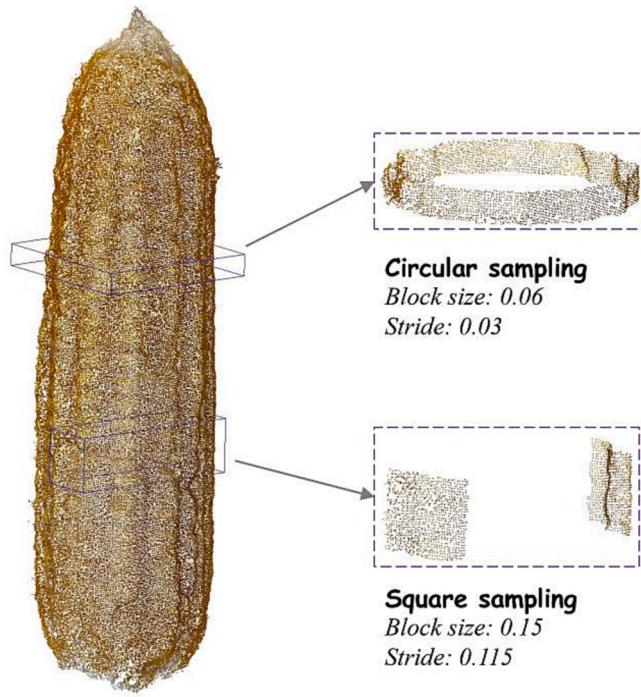
$$R = \begin{bmatrix} \cos\theta_{xy} & 0 & \sin\theta_{xy} \\ 0 & 1 & 0 \\ -\sin\theta_{xy} & 0 & \cos\theta_{xy} \end{bmatrix} \begin{bmatrix} \cos\theta_{xz} & -\sin\theta_{xz} & 0 \\ \sin\theta_{xz} & \cos\theta_{xz} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$P_R = R \bullet P \quad (2)$$

where  $R$  donates the 3D rotation matrix,  $P$  donates the original point cloud, and  $P_R$  donates the point cloud after coordinate correction.  $\theta_{xy}$  and  $\theta_{xz}$  are the rotation angles between  $P_{xy}$  and the x-axis,  $P_{xz}$  and the x-axis, respectively.

Considering the unique characteristics of maize ear point clouds, which exclusively capture the surface of the maize ear without including internal points, the utilization of typical square sampling methods may result in generating two distinct sets of points on the maize ear surface or even empty point sets (Fig. 3). This instability in the data structure of the input sample can significantly affect the fitting performance of models. Consequently, this study proposes a circular sampling approach along the x-axis for the coordinate-corrected maize ear point cloud, as depicted in Fig. 3.

The size of the circular block was set larger than the width of a single maize kernel to ensure sufficient coverage and a more comprehensive capture of the maize kernel structure. The circular block width was set to 0.06 with a sliding stride of 0.03. By overlapping local regions, the circular sampling approach increased the number of samples and achieved data augmentation. The block obtained from circular sampling



**Fig. 3.** Comparison of square sampling and circular sampling proposed in this study.

represented a complete set of points in a contiguous region, which could ensure the stability and consistency of the data structure when sliding sampling along the  $x$ -axis. In comparison, a square sampling approach was also utilized with a size of 0.15 and a sliding stride of 0.115. Both approaches generated a similar number of blocks, ensuring consistency in data quantities in training sets.

After segmenting the blocks, each block needs to be downsampled to ensure a consistent number of points received by the model. To maximize data utilization, points were randomly duplicated within each

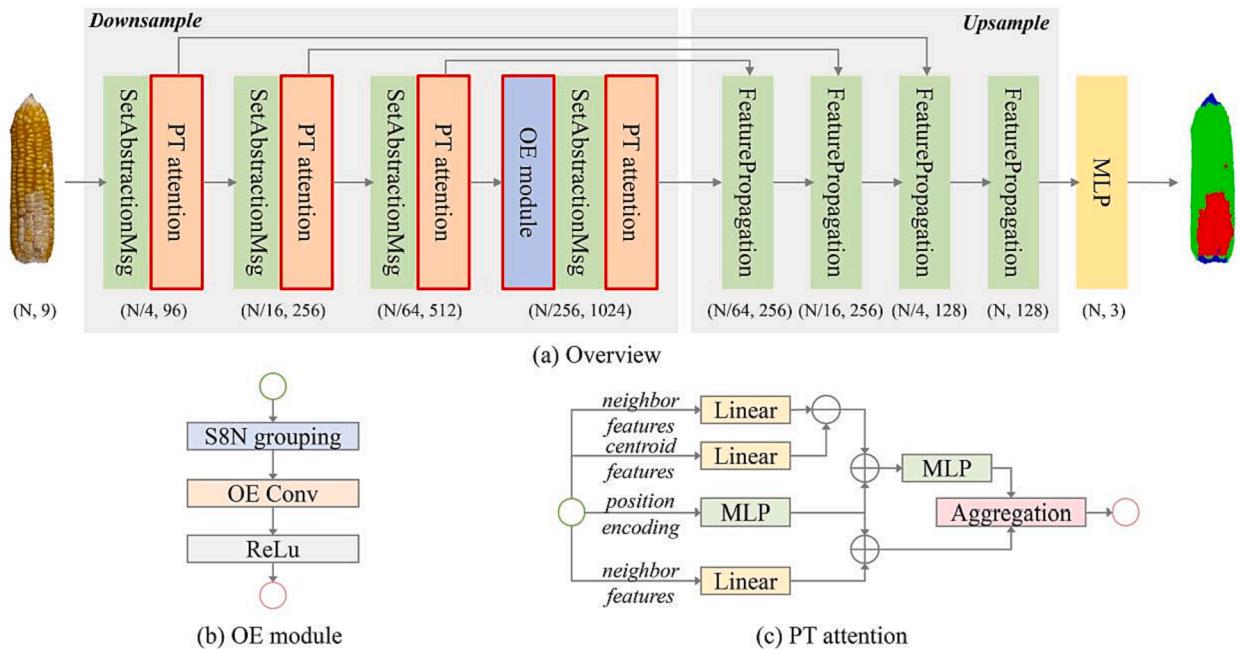
block, increasing the number of points in each block to a multiple of 2048 ( $k \times 2048$ ). Next, the processed points from each block were randomly divided into  $k$  sets of 2048 points each, which served as the final input data for the model.

### 2.2.3. Network architecture

The proposed ERSegNet, follows the standard encoder-decoder framework commonly used in point cloud semantic segmentation tasks. It is built upon the classic PointNet++ architecture (Qi et al., 2017), which serves as the baseline for semantic segmentation of maize ear rot. PointNet++ lacks efficiency in utilizing local neighborhood information on highly curved surfaces or regions (Sheshappanavar and Kambhamettu, 2020). Considering the geometric structure of maize ears and the characteristics of maize ear rot, local information is often more detailed and crucial. To address this issue, ERSegNet introduces local feature aggregation and self-attention modules during the down-sampling stage, improving the original structure by enhancing neighborhood information interaction (Fig. 4(a)).

The Orientation-encoding (OE) module is incorporated into the network to enhance the integration of neighborhood information. This module is inspired by the work of Jiang et al. (2018) and is a fundamental component of the PointSIFT network. It acts as a point-wise local feature descriptor, encoding information from eight orientations. The module employs stacked 8-neighborhood (S8N) search to identify the nearest neighbors in each of the eight octants, resulting in a  $2 \times 2 \times 2$  cube that describes the local pattern. Subsequently, the cube undergoes a three-stage OE convolution along the X, Y, and Z axes to integrate information from the eight spatial orientations, generating a representation that encodes orientation information. Unlike PointSIFT, which incorporates the OE module in both set abstraction (SA) and feature propagation (FP) layers, OE module is strategically inserted only before the fourth stage of the SA encoder to balance computational costs and facilitate integration of deep-level features (Fig. 4(b)).

Secondly, the self-attention operator proposed in Point Transformer is introduced to further enhance the model's ability to learn local geometric relationships between the center point and its neighborhood (Zhao et al., 2021). By integrating neighborhood features, center point features, and local positional features, the self-attention mechanism



**Fig. 4.** Architecture of ERSegNet. (a) Network overview. (b) Main structure of OE module. (c) Main structure of Point-Transformer Attention. The red borders highlight the differences between the proposed ERSegNet and PointNet++. OE: Orientation-encoding; PT: Point-Transformer. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

enables the model to focus on relevant information and capture intricate spatial dependencies (Fig. 4(c)). The Point-Transformer attention (PT attention) is seamlessly integrated throughout the entire downsampling process (Qiu et al., 2021). After each downsampling operation, new feature vectors are generated for all data points, allowing the network to refine its understanding of the 3D scene and improve information fusion. This enables the model to effectively capture and leverage point information for more accurate and detailed segmentation results.

#### 2.2.4. Quantitative assessment of maize ear rot

After performing semantic segmentation and predicting the class of each point in the maize ear, accurately evaluating ear rot still requires quantifying the proportion of infected areas. Due to significant infection differences of maize ears with various infected severity, as well as scattered infections of local or individual kernels, the evaluation method of converting point clouds of each class into triangulated meshes and then sequentially counting the area of each discrete area is relatively cumbersome. To achieve accurate point-level evaluation, a precise method based on point cloud semantic segmentation was proposed (Fig. 2(d)). This method characterizes the severity of ear rot by calculating the proportion of infected class points.

The density and uniformity of the maize ear point cloud are prerequisites for the feasibility of this method. Based on this, a point cloud resampling method was proposed for obtaining an optimized surface point distribution. This is achieved by performing Poisson surface reconstruction (Kazhdan et al., 2006) on the point cloud with predicted label and then relatively uniformly resampling 100 k points on the 3D surface by sample elimination (Yuksel, 2015). Based on the label information carried by the resampled point cloud, the ratio of the number of points predicted by the model as the infected class to the number of points predicted by the model as the kernel area is calculated. This ratio effectively characterizes the severity of infection of ear rot, which can be described as:

$$S = \frac{N_{\text{infected}}}{N_{\text{infected}} + N_{\text{uninfected}}} \times 100\% \quad (3)$$

where  $S$  is the severity of maize ear rot,  $N_{\text{infected}}$  and  $N_{\text{uninfected}}$  represent the number of points predicted as the infected kernel class and uninfected kernel class, respectively.

#### 2.3. Experiment settings

The preprocessed 100 maize ear point cloud data was divided into training, validating, and testing sets at a ratio of 7:1:2 (Table 1). All experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU with the PyTorch library on the Python platform. The models were trained from scratch for 30 epochs using AdamW optimizer and Cross-Entropy loss with label smoothing (Qian et al., 2022). According to the different number of points in each class, class weights were added during loss calculation to avoid the negative effect of class imbalance. The batch size was 64, and the learning rate was initially set to 0.1 for the first 15 epochs and 0.01 for epochs 15–30. During model training, the model parameters corresponding to the epoch with the highest IoU on the validating set were saved for performance testing. The data augmentation method of point random jitter was applied to the training set samples. Also, random seeds were fixed to ensure the reproducibility of the experiments.

**Table 1**  
Dataset of maize ear rot.

Number of maize ears	Number of points per maize ear	Number of blocks Training set	Validating set
100	80–200 k	12 k	1.6 k

#### 2.4. Evaluation metrics

The IoU, precision (Prec), recall (Rec), F1-score, and OA were used as evaluation metrics to evaluate the performance of semantic segmentation. IoU measures the overlap between the predicted points and the ground truth points. Prec is the proportion of the points correctly classified in this semantic class to all the points predicted by the network. Rec the proportion of the points correctly classified in this semantic class to the total points of this class in the ground truth. F1 score is the harmonic mean of precision and recall. OA measures the fraction of points that are correctly predicted overall. A higher IoU, Prec, Rec, F1 score, and OA indicate better performance. These metrics evaluate a model from different perspectives to provide a reliable assessment of its effectiveness. The five evaluation metrics can be calculated as:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (4)$$

$$\text{Prec} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Rec} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-score} = 2 \frac{\text{Prec} \times \text{Rec}}{\text{Prec} + \text{Rec}} \quad (7)$$

$$\text{OA} = \frac{TP + TN}{TP + FP + FN + TN} \quad (8)$$

where  $TP$ ,  $FP$ ,  $FN$ , and  $TN$  denote the number of true positives, false positives, false negatives, and true negatives, respectively.

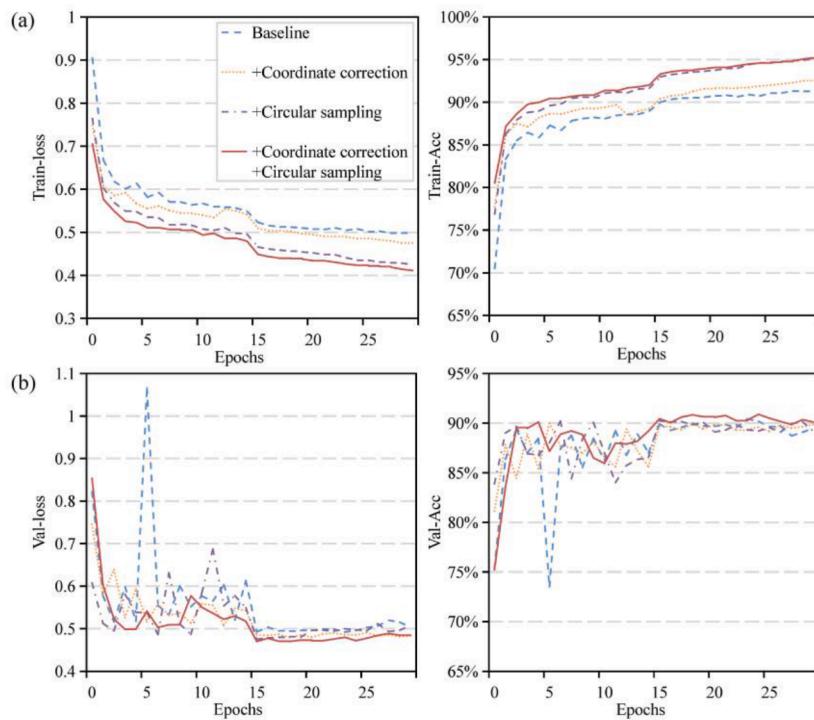
### 3. Results

#### 3.1. Performance of coordinate correction and circular sampling

The performance of PointNet++ on the testing sets using coordinate correction and circular sampling was compared to verify the effectiveness of the proposed data preprocessing approaches. After 30 epochs of training, all four models were well fitted, with the baseline combining with coordinate correction and circular sampling demonstrating the best fitting ability and generalization (Fig. 5). As shown in Table 2, the baseline achieves a mean Intersection over Union (mIoU) of 84.10%, mean Precision (mPrec) of 91.94%, mean Recall (mRec) of 90.60%, mean F1-score of 91.23%, and OA of 93.04%. This was a decent performance and validated the feasibility of semantic segmentation using PointNet++ on the 3D point cloud data of maize ears.

With the adoption of coordinate correction and circular sampling, PointNet++ exhibited overall enhancements. The model achieved a slight improvement in mIoU (84.33%), mRec (91.03%), mean F1-score (91.36%), and OA (93.07%) when training on the dataset with coordinate correction, indicating the beneficial impact of coordinate correction on maize ear point cloud with square sampling. The inclusion of circular sampling alone resulted in slightly better outcomes compared to using coordinate correction alone, with the highest mRec (91.07%), primarily attributed to the improved Recall (86.07%) for the infected kernel class. However, the increase in OA achieved by solely using circular sampling was minimal, possibly due to the remaining differences in the block spatial structure obtained by circular sampling along the x-axis for samples without coordinate correction.

In comparison to the baseline, the combined utilization of coordinate correction and circular sampling demonstrated the most significant performance enhancement, with the best results in mIoU (84.67%), mPrec (92.40%), mean F1-score (91.57%), and OA (93.23%), showing stable improvement in point predictions. Compared to the individual result of circular sampling alone for OA, the input block spatial structure



**Fig. 5.** Loss and accuracy map of four models on the (a) training and (b) validating sets during the training process.

**Table 2**

Performance comparison of PointNet++ on the testing sets with coordinate correction and circular sampling. The best results are in boldface, and the second best results are underlined.

Method		IoU (%)	Prec (%)	Rec (%)	F1-score (%)	OA (%)
Baseline (PointNet++)	Non-kernel	84.62	91.83	91.51	91.67	93.04
	Uninfected	91.71	94.53	96.85	95.68	
	Infected	75.97	89.45	83.45	86.35	
	Mean	84.10	<u>91.94</u>	90.60	91.23	
+Coordinate correction	Non-kernel	85.31	91.94	92.21	92.07	<u>93.07</u>
	Uninfected	91.68	95.01	96.32	95.66	
	Infected	75.99	88.23	84.56	86.36	
	Mean	84.33	91.73	91.03	91.36	
+Circular sampling	Non-kernel	85.66	93.52	91.07	92.28	93.04
	Uninfected	91.42	94.98	96.07	95.52	
	Infected	76.28	87.02	86.07	86.55	
	Mean	<u>84.46</u>	91.84	<b>91.07</b>	<u>91.45</u>	
+Coordinate correction +Circular sampling	Non-kernel	86.23	91.52	93.73	92.61	<u>93.23</u>
	Uninfected	91.63	94.28	97.02	95.63	
	Infected	76.15	91.41	82.02	86.46	
	Mean	<b>84.67</b>	<b>92.40</b>	<u>90.92</u>	<b>91.57</b>	

based on coordinate correction and circular sampling showed greater consistency, resulting in a 0.19% increase in OA.

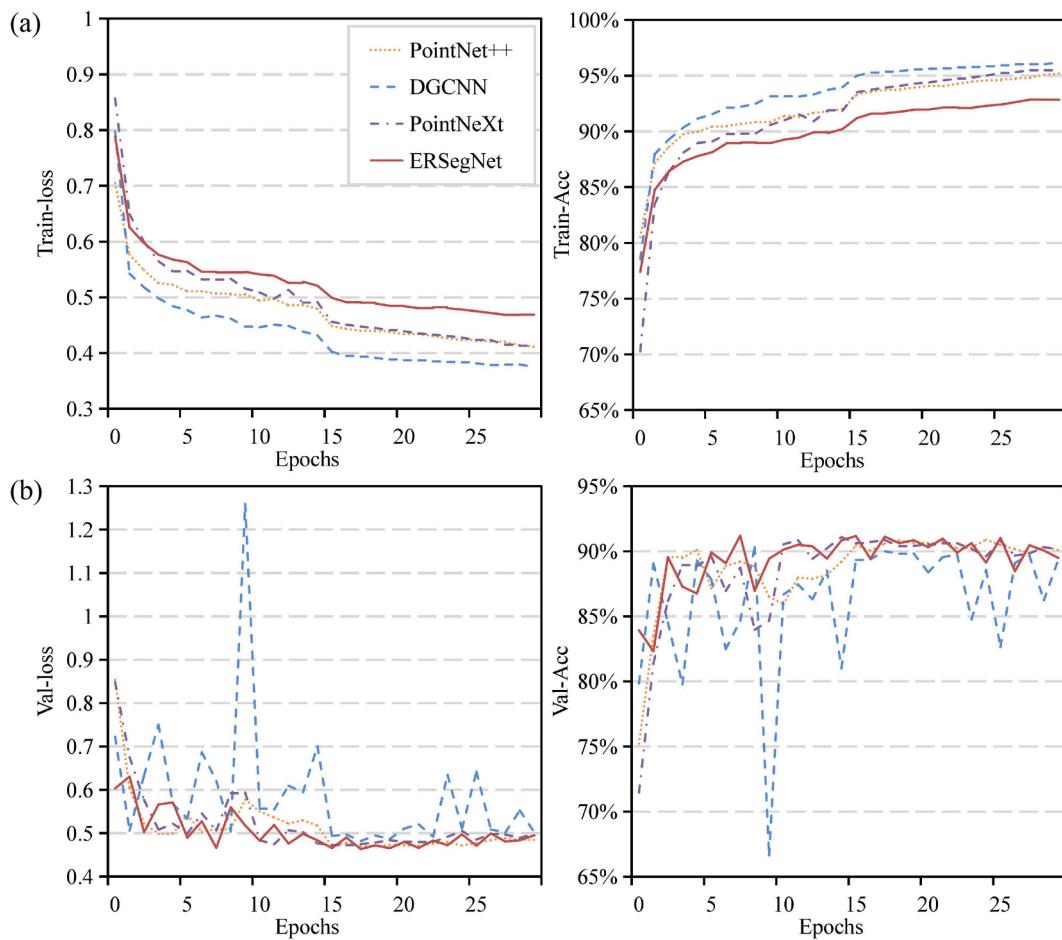
### 3.2. Semantic segmentation results and comparison

#### 3.2.1. Methods comparison

In this section, several popular point cloud segmentation networks were compared with ERSegNet, including PointNet++, DGCNN (Wang et al., 2019a), and PointNeXt (Qian et al., 2022). All models were trained on the same dataset using coordinate correction and circular sampling and were fitted after 30 epochs of training (Fig. 6). As shown in Fig. 6, ERSegNet demonstrates a better generalization performance on the validating set compared to the other three models. Despite the superior performance of the DGCNN model on the training set, it lacks stability

on the validating set. The results of the four models on the testing set are presented in Table 3.

ERSegNet achieved the best performance with the highest mIoU (85.83%), mRec (92.23%), mean F1-score (92.28%), and OA (93.76%). Compared to PointNet++, ERSegNet improved by 1.16%, 1.31%, 0.71%, and 0.53% in mIoU, mRec, mean F1-score, and OA, respectively. PointNet++ achieved the second-best results, with the highest mPrec (92.40%). As for PointNeXt, its original training involved various data augmentations and training strategies. However, some data augmentation methods such as point cloud rotation and color drop may disrupt the structural stability of the input data and lose key information about the features of maize ear rot. Therefore, these methods were not adopted in this study. This might have slightly limited the effectiveness of PointNeXt in maize ear point cloud segmentation, with the results of 84.35%



**Fig. 6.** Loss and accuracy map of the models on the (a) training and (b) validating sets during the training process.

mIoU and 92.97% OA. As a graph-based method, DGCNN showed the worst performance, with the mIoU of 93.13% and the OA of 92.47%.

### 3.2.2. Maize ear rot evaluation

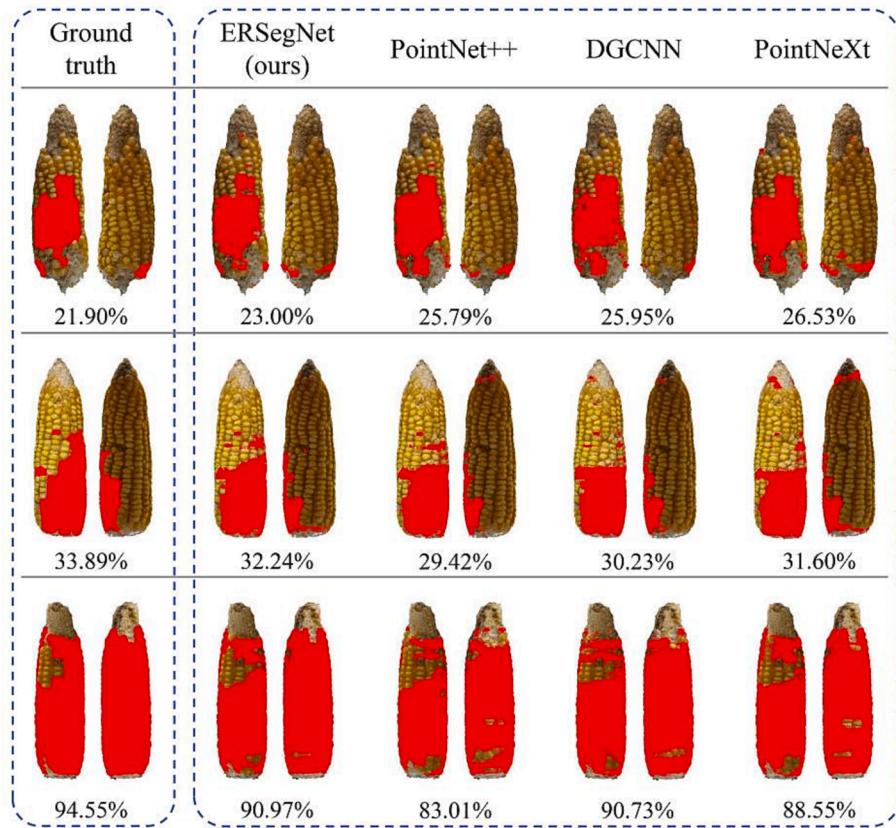
Fig. 7 displays the prediction results of the four methods on maize ear rot from the testing set. Intuitively, it could be observed that compared with the other three methods, ERSegNet's prediction results were closest

to the ground truth. ERSegNet had a more acute perception and obvious advantages in predicting some detail areas and edge areas. DGCNN's predictions for infected points were relatively scattered and showed poor sensitivity, which was consistent with its semantic segmentation performance in Table 3. The quantitative evaluation of the infected areas of maize ear rot is also illustrated below each example sample in Fig. 7. The quantitative values were calculated based on the method

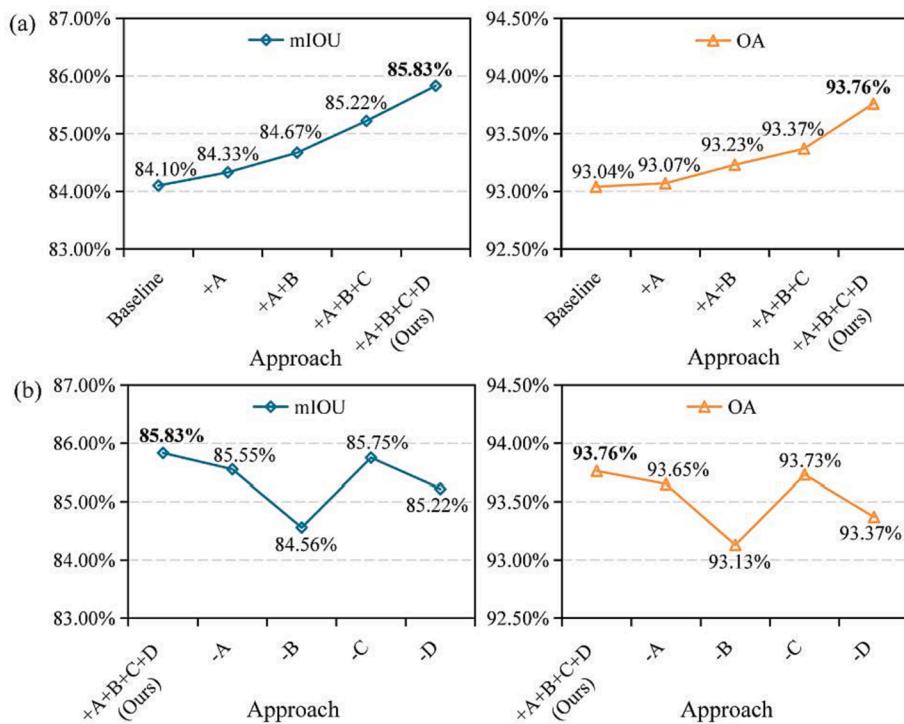
**Table 3**

Performance comparison for maize ear semantic segmentation of four methods. The best results are in boldface, and the second best results are underlined.

Method		IoU (%)	Prec (%)	Rec (%)	F1-score (%)	OA (%)
PointNet++	Non-kernel	86.23	91.52	93.73	92.61	<u>93.23</u>
	Uninfected	91.63	94.28	97.02	95.63	
	Infected	76.15	91.41	82.02	86.46	
	Mean	<u>84.67</u>	<b>92.40</b>	90.92	<u>91.57</u>	
DGCNN	Non-kernel	84.26	89.87	93.10	91.46	<u>92.47</u>
	Uninfected	90.97	95.02	95.53	95.27	
	Infected	74.17	87.18	83.24	85.17	
	Mean	83.13	90.69	90.62	90.63	
PointNeXt	Non-kernel	85.91	92.79	92.06	92.42	<u>92.97</u>
	Uninfected	91.24	94.90	95.94	95.42	
	Infected	75.91	87.39	85.24	86.30	
	Mean	84.35	91.70	<u>91.08</u>	91.38	
ERSegNet (Ours)	Non-kernel	86.28	91.89	93.39	92.64	<u>93.76</u>
	Uninfected	92.35	95.84	96.21	96.02	
	Infected	78.87	89.30	87.09	88.18	
	Mean	<b>85.83</b>	<u>92.34</u>	<b>92.23</b>	<b>92.28</b>	



**Fig. 7.** Demonstration of semantic segmentation results for the infected areas of maize ear rot. The infected points are marked in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 8.** Ablation study. (a) Results of adding components. (b) Results of removing components. A: Coordinate correction; B: Circular sampling; C: Orientation-encoding module; D: Point-Transformer attention. The best results are in boldface.

**Table 4**  
Performance of ERSegNet on different number of points.

Number of points	mIoU (%)	OA (%)	Params. (M)	FLOPs (G)	Throughput (items/s)
1024	83.97	92.87	11.80	7.38	98.00
2048	85.83	93.76	11.80	7.45	81.84
3072	86.13	93.92	11.80	7.52	71.48

proposed in Section 2.2.4. ERSegNet obtained the infected area ratios that were closest to the ground truth. The average evaluation errors of ERSegNet, PointNet++, DGCNN, and PointNeXt on the whole testing set were 1.55%, 2.91%, 2.45%, and 2.29%, respectively.

### 3.3. Ablation study for method effectiveness verification

To verify the effectiveness of each component in our method, ablation studies were conducted, and the results are shown in Fig. 8. Fig. 8(a) compares the improvement of mIoU and OA by adding one method or module each time. Fig. 8(b) shows the results of removing one method or module each time from the final model to evaluate the effectiveness of a single component (Lai et al., 2022).

As the components were added cumulatively, the model's performance improved steadily (Fig. 8). Coordinate correction resulted in gains of 0.23% mIoU and 0.03% OA. circular sampling further enhanced mIoU by 0.34% and OA by 0.16% on this basis. The addition of the two feature aggregation modules led to more substantial progress. The OE module increased mIoU to over 85% and boosted mIoU by 0.55% and OA by 0.14%. PT attention achieved the largest improvement, increasing mIoU by 0.61% and OA by 0.39%. The addition of the four components ultimately enabled ERSegNet to attain 85.83% mIoU and 93.76% OA in maize ear rot semantic segmentation.

When removing methods or modules, it was found that reducing any component did not perform better than the final model ERSegNet, indicating the effectiveness of each enhancement proposed. Among all components, removing circular sampling resulted in the largest performance drop, leading to a decrease of 1.27% in mIoU and 0.63% in OA. Secondly, the removal of PT attention caused mIoU to drop 0.61% and OA to drop 0.39%. Canceling coordinate correction led to a 0.28% decrease in mIoU and a 0.11% decrease in OA. Removing the OE module had the least impact on maize ear rot semantic segmentation performance, possibly due to the offset from PT attention's substantial performance gains. This further validated the effectiveness of PT attention.

## 4. Discussion

### 4.1. The number of downsampled points

As is well known, the number of points input into the model has a significant impact on its fitting results. The performance of ERSegNet was compared using different numbers of downsampled points, specifically 1024, 2048, and 3072, as inputs (Table 4). To ensure a similar number of samples in the training set, the strides of circular sampling were adjusted during these experiments. The throughput was tested on GPU with a batch size of 64, to simulate the application on a single maize ear.

The results demonstrated that increasing the number of downsampled points had a positive impact on the model's performance. This improvement could be attributed to the enhanced information and feature density (Wang et al., 2019b), enabling the model to capture more intricate details. Due to computational limitations, the input of a higher number of points was not tested. It is worth noting that as the number of points increases, the inference speed of the model tends to decrease. The inference time for a single maize ear with 2048 input points on the GPU is approximately 3 s, which is acceptable for

applications. Consequently, in practical scenarios, it is crucial to choose the number of downsampling points to optimize model prediction accuracy and efficiency, taking into account the available computing power determined by hardware configuration.

### 4.2. Maize ear rot evaluation method based on resampling

After obtaining the semantic segmentation results of maize ears, this study focused on optimizing the uniformity of reconstructing dense point clouds from MVS data. To achieve this, a point cloud resampling method was employed. The maize ear point cloud carrying segmentation labels first underwent surface reconstruction and then performed relatively uniform resampling to enhance the uniformity of point cloud distribution. This process enables the representation of the infection rate of ear rot disease through the proportion of infected points. The objective of this point cloud resampling operation was to reach a more accurate quantification of maize ear rot.

The resampling approach was achieved through sample elimination based on greedy algorithms (Yuksel, 2015), which can enhance the blue noise characteristics of the sampled point cloud and achieve more uniform and natural sampling results (Yan et al., 2015). The mean uniformity of point clouds on the testing set was compared between original and resampled point clouds. The uniformity is quantified by the standard deviation of the nearest neighbor distance for each point in the target point cloud. The mean nearest neighbor distances of original and resampled point clouds on testing set were  $3.04e^{-3}$  and  $4.23e^{-3}$ , respectively. And the mean standard deviation of the nearest neighbor distance of point clouds on testing set had been significantly reduced to  $3.94e^{-4}$  from  $1.01e^{-3}$  after resampling, which supported the validity of the proposed resampling approach.

### 4.3. Limitations and improvements

In this study, the feasibility of detecting and evaluating maize ear rot based on 3D semantic segmentation algorithms was verified. And the proposed ERSegNet effectively improves the semantic segmentation accuracy of the network by incorporating feature aggregation and self-attention modules. Future improvements may consider the following three aspects:

- (1) Extraction of point cloud features from maize ears. Inspired by the results of PointNet++ on non-Euclidean metric spaces (Qi et al., 2017), considering that the point cloud data of maize ears is also a form of 2D surface embedded in 3D space, it is possible that using intrinsic point features extracted from non-Euclidean metric space as the inputs can facilitate the capture of multi-scale intrinsic structures of maize ears. This approach may help mitigate the influence of shape factors and further enhance the robustness and effectiveness of the segmentation model.
- (2) Optimization of model structure and dataset. In terms of the requirements of hardware constraints and real-time assessment scenes, exploring lightweight functional modules could be considered to enhance the inference speed of the models in future research. And augmenting the training dataset with more samples is an effective approach to enhance prediction performance and improve network generalization and robustness.
- (3) Improvement of point cloud resampling. Improving the uniformity performance of point cloud can further enhance the accuracy of point-level evaluation of maize ear rot, so effective and easy-to-operate resampling methods can be explored. In addition, this study applied the data resampling approach to the point cloud of maize ears after semantic segmentation. It is possible to consider integrating the resampling step into the 3D reconstruction stage to improve the overall detection efficiency.

The maize ear rot evaluation pipeline established in this study encompasses maize ear data collection, MVS reconstruction, point cloud preprocessing, semantic segmentation, and precise evaluation of maize ear rot. It is worth noting that this pipeline can be integrated to achieve end-to-end automatic and precise assessment for maize ear rot, offering great potential and value for field applications.

## 5. Conclusion

To overcome the challenge of rough manual assessment of maize ear rot, this study first proposes the pipeline of using the 3D point cloud data for ear rot precise evaluation. First, the feasibility of 3D semantic segmentation using PointNet++ on maize ear point clouds obtained through MVS reconstruction was investigated, resulting in 84.10% mIoU and 93.04% OA. Combining with the characteristics of maize ear point cloud data structure, effective data preprocessing approaches including point cloud coordinate correction and circular sampling were proposed. These improvements led to a stable enhancement of 0.57% in mIoU and 0.19% in OA for the prediction results. Furthermore, the integration of the OE module and PT attention into PointNet++ significantly enhanced local feature extraction capability and semantic segmentation performance, achieving an mIoU of 85.83% and OA of 93.76%. To achieve the point-level quantification of maize ear rot in 3D point clouds, a point cloud resampling approach was introduced to promote spatial uniformity, supporting precise assessment of maize ear rot. Based on the semantic segmentation results of ERSegNet, the average evaluation error of the infection proportion of maize ear rot in the testing set was 1.55%, demonstrating superior detection capability. The developed pipeline for quantifying maize ear rot provides a novel, reliable, and objective analysis approach, providing valuable technical support and reference for resistance identification and precision breeding.

## CRediT authorship contribution statement

**Rui Yang:** Methodology, Software, Visualization, Writing – original draft. **Yong He:** Methodology, Writing – review & editing, Supervision. **Xiangyu Lu:** Methodology, Validation. **Yiying Zhao:** Investigation, Resources. **Yanmei Li:** Investigation, Resources. **Yinhui Yang:** Conceptualization, Resources. **Wenwen Kong:** Resources, Methodology. **Fei Liu:** Conceptualization, Writing – review & editing, Supervision.

## Declaration of Competing Interest

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

## Data availability

The authors do not have permission to share data.

## Acknowledgements

This work was supported by the Science and Technology Department of Zhejiang Province (2022C02034), and Science and Technology Department of Shenzhen (CJGJZD20210408092401004).

## References

- Bao, Y., Tang, L., Srinivasan, S., Schnable, P.S., 2019. Field-based architectural traits characterisation of maize plant using time-of-flight 3D imaging. *Biosyst. Eng.* 178, 86–101. <https://doi.org/10/gmtnzs>.
- Ding, Z., Sun, Y., Xu, S., Pan, Y., Peng, Y., Mao, Z., 2023. Recent advances and perspectives in deep learning techniques for 3D point cloud data processing. *Robotics* 12, 100. <https://doi.org/10.3390/robotics12040100>.
- Dong, C., Wu, Y., Gao, J., Zhou, Z., Mu, C., Ma, P., Chen, J., Wu, J., 2018. Field inoculation and classification of maize ear rot caused by *Fusarium verticillioides*. <https://doi.org/10.21769/BioProtoc.3099 Bio Protoc. 8, e3099>.
- Du, R., Ma, Z., Xie, P., He, Y., Cen, H., 2023. PST: Plant segmentation transformer for 3D point clouds of rapeseed plants at the podding stage. *Isprs J. Photogramm.* 195, 380–392. <https://doi.org/10.1016/j.isprsjprs.2022.11.022>.
- Fei, B., Yang, W., Chen, W., Li, Z., Li, Y., Ma, T., Hu, X., Ma, L., 2022. Comprehensive review of deep learning-based 3D point cloud completion processing and analysis. *IEEE Trans. Intell. Transport. Syst.* 23, 22862–22883. <https://doi.org/10.1109/TITS.2022.3195555>.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 381–395. <https://doi.org/10.1145/358669.358692>.
- Gaikpa, D.S., Kessel, B., Presterl, T., Ouzunova, M., Galiano-Carneiro, A.L., Mayer, M., Melchingher, A.E., Schon, C.-C., Miedaner, T., 2021. Exploiting genetic diversity in two European maize landraces for improving Gibberella ear rot resistance using genomic tools. *Theor. Appl. Genet.* 134, 793–805. <https://doi.org/10.1007/s00122-020-03731-9>.
- Galiano-Carneiro, A.L., Kessel, B., Presterl, T., Gaikpa, D.S., Kistner, M.B., Miedaner, T., 2020. Multi-parent QTL mapping reveals stable QTL conferring resistance to Gibberella ear rot in maize. *Euphytica* 217, 2. <https://doi.org/10/gm95g5>.
- Guo, R., Xie, J., Zhu, J., Cheng, R., Zhang, Y., Zhang, X., Gong, X., Zhang, R., Wang, H., Meng, F., 2023. Improved 3D point cloud segmentation for accurate phenotypic analysis of cabbage plants using deep learning and clustering algorithms. *Comput. Electron. Agr.* 211, 108014. <https://doi.org/10.1016/j.compag.2023.108014>.
- Hudson, O., Hudson, D., Brahmstedt, C., Brawner, J., 2023. The ear unwrapper: a maize ear image acquisition pipeline for disease severity phenotyping. *AgriEng.* 5, 1216–1225. <https://doi.org/10.3390/agriengineering5030077>.
- Jiang, M., Wu, Y., Zhao, T., Zhao, Z., Lu, C., 2018. PointSIFT: A SIFT-like network module for 3D point cloud semantic segmentation. <<https://doi.org/10.48550/arXiv.1807.00652>>.
- Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. *Eurograph. Assoc.* <https://doi.org/10.2312/SGP/SGP06/061-070>.
- Liu, Y., Hu, G., Zhang, A., Loladze, A., Hu, Y., Wang, H., Qu, J., Zhang, X., Olsen, M., San Vicente, F., Crossa, J., Lin, F., Prasanna, B.M., 2021. Genome-wide association study and genomic prediction of *Fusarium* ear rot resistance in tropical maize germplasm. *Crop. J.* 9, 325–341. <https://doi.org/10/gm95g6>.
- Long, W., Wang, T., Lu, Y., Xu, H., Gao, Q., Jiang, L., 2021. Research progress of virtual reality modeling based on three-dimensional reconstruction technology. *J. Phys. Conf. Ser.* 1769, 012014. <https://doi.org/10.1088/1742-6596/1769/1/012014>.
- Mesterhazy, A., Lemmens, M., Reid, L.M., 2012. Breeding for resistance to ear rots caused by *Fusarium* spp. in maize—a review. *Plant Breed.* 131, 1–19. <https://doi.org/10.1111/j.1439-0523.2011.01936.x>.
- Miao, Y., Peng, C., Wang, L., Qiu, R., Li, H., Zhang, M., 2022. Measurement method of maize morphological parameters based on point cloud image conversion. *Comput. Electron. Agr.* 199, 107174. <https://doi.org/10.1016/j.compag.2022.107174>.
- Naz, R., Bano, A., Nosheen, A., Yasmin, H., Keyani, R., Shah, S.T.A., Anwar, Z., Roberts, T.H., 2021. Induction of defense-related enzymes and enhanced disease resistance in maize against *Fusarium verticillioides* by seed treatment with *Jacaranda mimosifolia* formulations. *Sci. Rep.* 11, 59. <https://doi.org/10.1038/s41598-020-79306-x>.
- Lai, X., Liu, J., Jiang, L., Wang, L., Zhao, H., Liu, S., Qi, X., Jia, J., 2021. Stratified transformer for 3D point cloud segmentation. <https://doi.org/10.1109/CVPR52688.2022.000831>.
- Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. PointNet++: deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17. Curran Associates Inc., Red Hook, NY, USA, pp. 5105–5114. <https://doi.org/10.48550/arXiv.1706.02413>.
- Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., Ghanem, B., 2022. PointNeXt: revisiting PointNet++ with improved training and scaling strategies. *Adv. Neural Inf. Proces. Syst.* 35, 23192–23204.
- Qiu, S., Wu, Y., Anwar, S., Li, C., 2021. Investigating attention mechanism in 3D point cloud object detection. In: 2021 International Conference on 3D Vision (3DV). Presented at the 2021 International Conference on 3D Vision (3DV), pp. 403–412. <https://doi.org/10.1109/3DV53792.2021.00050>.
- Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D.N.L., Pinto, F., Pinera-Chavez, F.J., Poland, J., Rivera-Amado, C., Saint Pierre, C., Sukumaran, S., 2020. Breeder friendly phenotyping. *Plant Sci.* 295, 110396. <https://doi.org/10.1016/j.plantsci.2019.110396>.
- Rose, L.J., Okoth, S., Beukes, I., Ouko, A., Mouton, M., Flett, B.C., Makumbi, D., Viljoen, A., 2017. Determining resistance to *Fusarium verticillioides* and fumonisin accumulation in African maize inbred lines resistant to *Aspergillus flavus* and aflatoxins. *Euphytica* 213, 93. <https://doi.org/10.1007/s10681-017-1883-7>.
- Sheshappanavar, S.V., Kambhamettu, C., 2020. A novel local geometry capture in PointNet++ for 3D classification. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, pp. 1059–1068. <https://doi.org/10.1109/CVPRW50498.2020.00139>.
- Sun, D., Robbins, K., Morales, N., Shu, Q., Cen, H., 2022. Advances in optical phenotyping of cereal crops. *Trends Plant Sci.* 27, 191–208. <https://doi.org/10.gncsqt>.
- Thompson, M.E.H., Raizada, M.N., 2018. Fungal pathogens of maize gaining free passage along the silk road. *Pathogens* 7, 81. <https://doi.org/10.3390/pathogens7040081>.
- Tran, T.M., Ameye, M., Devlieghere, F., De Saeger, S., Eeckhout, M., Audenaert, K., 2021. *Streptomyces* strains promote plant growth and induce resistance against *Fusarium verticillioides* via transient regulation of auxin signaling and archetypal defense.

- pathways in maize plants. *Front. Plant Sci.* 12. <https://doi.org/10.3389/fpls.2021.755733>.
- Turgut, K., Dutagaci, H., Rousseau, D., 2022. RoseSegNet: an attention-based deep learning architecture for organ segmentation of plants. *Biosyst. Eng.* 221, 138–153. <https://doi.org/10.1016/j.biosystemseng.2022.06.016>.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019a. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.* 38, 146: 1–146, 12. <https://doi.org/10.1145/3326362>.
- Wang, Y., Wen, W., Wu, S., Wang, C., Yu, Z., Guo, X., Zhao, C., 2019b. Maize plant phenotyping: comparing 3D laser scanning, multi-view stereo reconstruction, and 3D digitizing estimates. *Remote Sens.* 11, 63 <https://doi.org/10/ggbnrj>.
- Wen, J., Shen, Y., Wang, Z., Li, S., Mo, L., Lei, Y., Zhang, Y., Han, S., 2021a. QTL mapping of resistance to *Fusarium* ear rot in maize based on image analysis. *Scientia Agric. Sin.* 54, 2724–2736. <https://doi.org/10.3864/j.issn.0578-1752.2021.13.003>.
- Wen, W., Wang, Y., Wu, S., Liu, K., Gu, S., Guo, X., 2021b. 3D phytomer-based geometric modelling method for plants—the case of maize. *Aob Plants* 13, plab055 <https://doi.org/10/gnzfcn>.
- Wu, S., Wen, W., Wang, Y., Fan, J., Wang, C., Gou, W., Guo, X., 2020. MVS-pheno: a portable and low-cost phenotyping platform for maize shoots using multiview stereo 3D reconstruction. *Plant Phenom.* 2020, 1–17 <https://doi.org/10/gmtnzt>.
- Yan, D.-M., Guo, J.-W., Wang, B., Zhang, X.-P., Wonka, P., 2015. A survey of blue-noise sampling and its applications. *J. Comput. Sci. Technol.* 30, 439–452. <https://doi.org/10.1007/s11390-015-1535-0>.
- Yao, L., Li, Y., Ma, C., Tong, L., Du, F., Xu, M., 2020. Combined genome-wide association study and transcriptome analysis reveal candidate genes for resistance to *Fusarium* ear rot in maize. *J. Integr. Plant Biol.* 62, 1535–1551. <https://doi.org/10.1111/jipb.12911>.
- Yuksel, C., 2015. Sample elimination for generating Poisson disk sample sets. *Comput. Graph. Forum.* 34, 25–32. <https://doi.org/10.1111/cgf.12538>.
- Zhang, W., Wu, S., Wen, W., Lu, X., Wang, C., Gou, W., Li, Y., Guo, X., Zhao, C., 2023. Three-dimensional branch segmentation and phenotype extraction of maize tassel based on deep learning. *Plant Methods* 19, 76. <https://doi.org/10.1186/s13007-023-01051-9>.
- Zhao, H., Jiang, L., Jia, J., Torr, P.H.S., Koltun, V., 2021. Point transformer. In: Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16259–16268. doi: 10.1109/ICCV48922.2021.01595.
- Zhu, M., Tong, L., Xu, M., Zhong, T., 2021. Genetic dissection of maize disease resistance and its applications in molecular breeding. *Mol. Breed.* 41, 32. <https://doi.org/10.1007/s11032-021-01219-y>.