Expert Systems With Applications 229 (2023) 120497

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Towards end-to-end deep RNN based networks to precisely regress of the lettuce plant height by single perspective sparse 3D point cloud

Jingsong Li [a], Ying Wang [a], LiHua Zheng [a], Man Zhang [a], Minjuan Wang [a,b,c,*]

[a] Key Lab of Smart Agriculture Systems, Ministry of Education, China Agricultural University, Beijing 100083, China
[b] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
[c] Yantai Institute of China Agricultural University, Yantai 264670, China

ARTICLE INFO

Keywords:
3D Point cloud
Remote sensing
Deep RNN
Plant height
Regression network
Vegetation structural parameters

ABSTRACT

Nowadays, 3D point cloud is supposed to be the most direct and effective data form for studying plant morphology structure. However, automatic and high-throughput acquisition of accurate individual plant height traits from 3D point cloud remains an urgent challenging problem. Summarizing the related research results in recent years, the factors limiting its application mainly come from these aspects: (1) Many existing methods require spatial auxiliary information such as ground control points (GCP), digital terrain models (DTM) and digital surface models (DSM) to obtain accurate plant height; (2) For 3D point cloud data in different environments, specialized modeling and careful parameter fine-tuning are usually required; (3) Sometimes, the point cloud processing involves the combined utilization of multiple programming languages and software, which is difficult for system integration. Focusing on these challenges, firstly, we proposed a novel end-to-end deep Recurrent Neural Network (RNN) based regression network framework called DRN, which consists of three parts: point cloud feature extraction network, deep RNN and regression network. The convolution operations-based point cloud feature extraction network is function as filtering noise, outliers and redundant information; The deep RNN network with long and short-term memory (LSTM) ability is used to learn the relationships between the feature points on the high-dimensional feature sequence separated by a certain distance; regression network is used to regress the output from deep RNN to plant height value. Experiments results on the 3rd Greenhouse Growing Challenge datasets show that DRN can directly regress the plant height of a single plant effectively without manual operations and the participation of spatial auxiliary information with an R^2 of 0.948 and a relative root mean square error (RRMSE) of 10.06% in four different varieties of lettuce at different growth period. After studying the influence of the weights of the x, y, z coordinate of the input 3D point cloud on the regression result, then, we design a Dimension Attention (DA) module at the front end of the feature extraction network to learning the characteristic coordinate weight for every input point cloud sample. The DRN network with a DA module is called D-DRN, experiment results indicate D-DRN tend to achieve better result (R^2 = 0.960; RRMSE = 8.680%) than DRN. Considering the end-to-end-based DRN and D-DRN network capable of ease of integration and their considerable prediction accuracy on public datasets, we believe they has a certain complementary effect on the existing study methods of obtaining plant morphological structure phenotype by point cloud data.

1. Introduction

Plant height is one of the main plant morphological structural traits, which is often used for biomass inversion (Grenzdorffer, 2014), population structure prediction (Tian et al., 2019), variety breeding (Malambo et al., 2018), and autonomous farming systems (Andujar et al., 2016), etc. Traditional field-measured methods are easy to operate and can obtain accurate plant heights. However, field measurements are time-consuming and laborious with low efficiency. Hence, it is increasingly difficult to meet the requirements of high-throughput, high-efficiency, non-destructive and repeatable measurement of phenotype acquisition (Araus and Cairns, 2014). Earlier studies show that obtaining

* Corresponding author at: Key Lab of Smart Agriculture Systems, Ministry of Education, China Agricultural University, Beijing 100083, China.
E-mail addresses: jingsong@cau.edu.cn (J. Li), wyleo@cau.edu.cn (Y. Wang), zhenglh@cau.edu.cn (L. Zheng), cauzm@cau.edu.cn (M. Zhang), minjuan@cau.edu.cn (M. Wang).

https://doi.org/10.1016/j.eswa.2023.120497
Received 12 July 2022; Received in revised form 2 May 2023; Accepted 13 May 2023
Available online 18 May 2023
0957-4174/

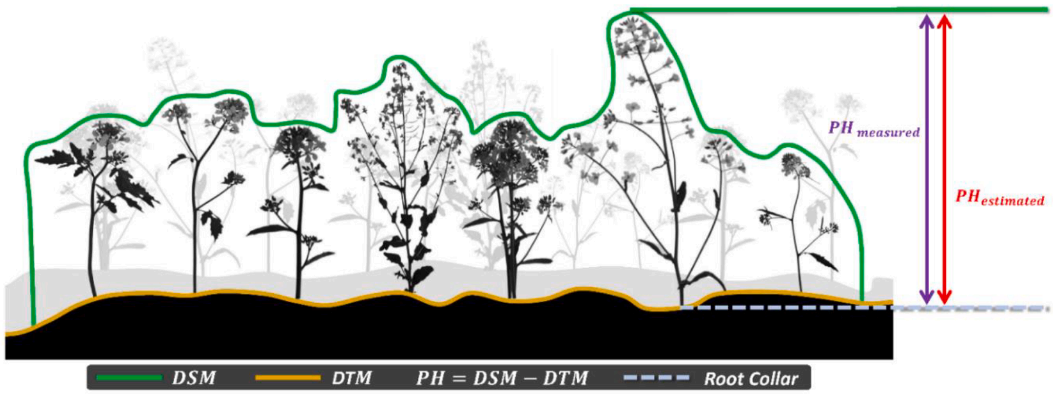

Fig. 1. Difference method: Extracting PH by subtracting the pre-acquired DTM from the DSM.

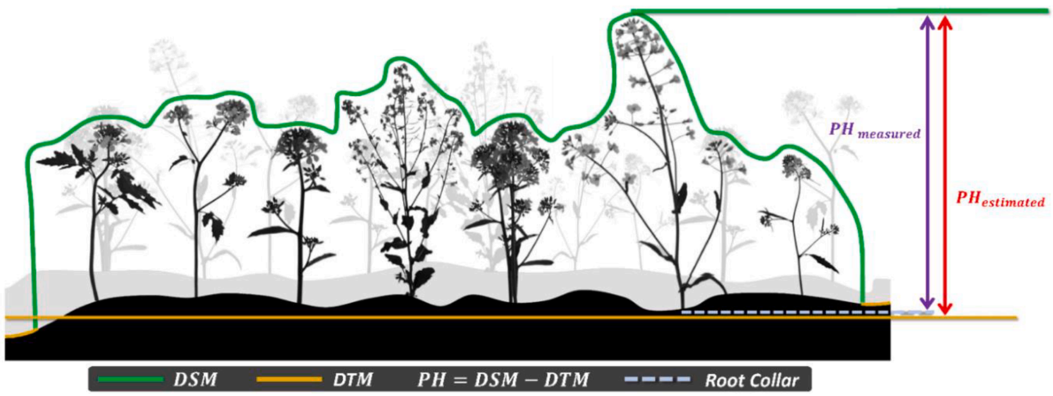

Fig. 2. 3D point cloud method: Extracting PH by subtracting the estimated DTM from DSM.

plant image data to estimate phenotypic parameters, such as plant height is obviously an effective and potential method that can replace manual measurement in parallel (Kataoka et al., 2003). The 3D information of the plant object in space is lost while acquiring the 2D image, which means that the process of estimating plant phenotype through 2D images often involve a series of tedious tasks, such as camera parameter acquisition, image correction, and calibration object selection (Kataoka et al., 2003).

The 3D point cloud data contains real 3D coordinate information of the object. Compared with ranging by 2D image, point cloud ranging usually needs no various perspective corrections or special shooting conditions sampling due to the effect from shooting angle of view and shooting distance. Moreover, point cloud is supposed to be the most direct and effective data form for studying the morphology and structure of plants in current times(Miao et al., 2021). With the development of sensor technology and small drone technology, in the past two decades, a large number of researches involving in estimation of the plant heights based on point cloud data have emerged. The hardware platforms for obtaining point cloud data is constantly enriched, these platforms include ground base station lidar (Wu et al., 2019), drone airborne lidar (Wang et al., 2021), time of fly (TOF) depth camera (Wang et al., 2021), structure from motion (SFM)based drone RGB camera (Bendig et al., 2014) and so on. The research objects range from the model plant Arabidopsis thaliana (Bernotas et al., 2019) to common field crops such as barley (Bendig et al., 2014), wheat (Tao et al., 2019), corn (Lu et al., 2021), sugarcane (Sumesh et al., 2021), soybeans (Maimaitijiang et al., 2019) to lettuce (Mortensen et al., 2018), cauliflower (Andujar et al., 2016), and trees in the forest (Tian et al., 2019), etc. These above works have proved that acquisition of plant height based on 3D point cloud data is widely effective in a variety of platforms and objects. Judging from previous research, the existing research results really provide technical guidance and theoretical basis for people to choose more cost-effective and suitable devices for various demand scenarios in the automatic acquisition of plant phenotype solutions. The methods of using 3D models to estimate crop height can be divided into two categories: difference method and 3D point cloud method.

The difference method is commonly used in unmanned aerial vehicle (UAV) platform equipped with a high-resolution RGB camera to obtain plant phenotypic parameters (De Souza et al., 2017; Huett et al., 2016). When using the difference method, lots of extra spatial assistant information are required, such as ground control points (GCPs), digital terrestrial model (DTM) before planting or just after the plant is harvested, and digital surface model (DSM), and etc(Bendig et al., 2015). The steps of the difference method to estimate plant height by point cloud data mainly include three parts: (1) Get the digital terrestrial model (DTM); (2) Obtain the digital surface model (DSM); (3) DSM minus DTM to get Plant Height (PH), as shows in Fig. 1. Considering the requirement of collecting ground images data without plant covering, it has to assume that the ground conditions remain unchanged during data acquisition.

Compared with lidar, the costs of RGB sensors are relatively low, and there are mature auxiliary software such as Agisoft Photoscan (Grenzdorffer, 2014) and relatively complete calculation process for difference method to implementing. Most of all, the DTM that the difference method relies on was obtained before planting or after the crop is harvested, even when the canopy or leaves of the plant are severely obscured by each other, the difference method can still effectively predict the plant height parameters. At present, the difference method is still an important method to obtain accurate plant height and other phenotypic parameters from 3D models.

The most obvious advantage of the point cloud method compared with the difference method is that it does not need to obtain DTM before planting or just after the plant is removed. In most cases, the existing point cloud method is to separate the crop part from the ground soil part in the point cloud data through some statistical related algorithms, then, obtain the plant height information by the difference between the separated crop part and the point cloud of the soil part (Hyyppa et al., 2020). When using the point cloud method to measure plant height, it is usually required that the occlusion of the leaves of the plant is not so severe and a good transmission capability of the sensor, so that the sensor can simultaneously obtain the point cloud data of both the plants and the bare ground. Compared with passive sensors such as RGB cameras, active sensors can show stronger transmission capabilities (Wang et al., 2021), which can explain the fact that point cloud method is commonly used in active sensors (Wang et al., 2021; Yang et al., 2020).

Due to the sparsity of the point cloud, occlusion and sensor noise, it is difficult to obtain accurate point coordinates at the top and bottom of the target plant for direct plant height measurement. For the 3D point cloud method, after dividing the plant point cloud and the ground point cloud into categories, DTM and DSM can be obtained, and the remaining work is to calculate the PH model through the difference between them, as shown in Fig. 2. Cloud clustering is a common method for Segmentation of plants and ground in point cloud method (Gao et al., 2015; Yang et al., 2020). Separating the plant point cloud from the soil point cloud can essentially be regarded as a binary classification of the point cloud. (Song and Wang, 2019)suggested that the well-known Otsu method in the 2D image binary processing is effective in 3D point cloud segmentation by a threshold. It is also based on the idea of two classifications, (Vazquez-Arellano et al., 2018)firstly used the RANSAC algorithm to remove the "invalid sample" noise point cloud from the corn plant point cloud data collected by the TOF camera, then used the RANSAC algorithm to successfully separate the corn plant point cloud from the soil point cloud. Currently, deep learning algorithms are widely regarded as having an overwhelming advantage over other machine learning algorithms in machine vision classification tasks. (Bernotas et al., 2019) successfully separated the point cloud of potted Arabidopsis thaliana from the soil point cloud through a deep learning algorithm to obtain 3D phenotypic information. For some plants with relatively simple canopy structures like corn, the plant morphology structure information can be obtained by skeleton extraction after segmenting the 3D point cloud (Wu et al., 2019; Zhu et al., 2021). Compared with difference method, the point cloud method usually relies on more penetrating but more expensive active sensors, which reduces the practical value of the point cloud method (Wang et al., 2021; Zolkos et al., 2013). With the development of point cloud technology and the improvement of point cloud quality, people's demanding for more accurate and high-throughput plant height estimation might be more urgent.

Although some plant height estimation systems based on 3D point cloud data could achieve good results in terms of estimation accuracy and robustness (Bernotas et al., 2019), these pipelines usually require professional and careful adjustment of parameters to ensure its good performance, and even using a commercial software to process data in some period (Han et al., 2018). What's more, existing studies indicate that the mainstream method of using plant point cloud data to estimate plant height usually rely on DTM as the reference ground, and the height of the z coordinate of the plant point cloud relative to the ground is obtained through the difference between DTM and DSM to estimate plant height, which essentially uses the z coordinate of the point cloud to calculate the plant height. For existing plant height estimation methods based on point cloud data, accurate plant height estimation not only requires high-density and low-error point cloud data, an accurate and reliable DTM model is also needed, which obviously leads to more time and labor costs (Xie et al., 2021).

The deep learning network has achieved extraordinary results in machine vision-related tasks with its strong expressive ability and good robustness in both 2D images and 3D point clouds datasets.With the continuous development of deep learning platforms, deep learning networks could be easily embedded in mobile devices or the web for practical use. Moreover, due to the versatility and ease of transplantation of the deep learning network, a phenotypic parameter extraction network designed for a certain plant object can easily be used to train other's 3D point cloud data. So, isn't it an excellent idea to design an end-to-end training deep learning network that can automatically extract the 3D phenotype of the plant from the 3D point cloud data? In recent years, researchers have made amazing achievements on 3D point cloud data sets using deep learning networks. The well-known deep learning network models for 3D point cloud data processing tasks include PointNet series (Qi et al., 2017a; Qi et al., 2017b), PointCNN (Li et al., 2018), DGCNN (Wang et al., 2019) and C2Fnet (Lei et al., 2022). PointNet proposed to perform high-dimensional mapping of the input point cloud data before maximizing pooling to reduce the loss of information as much as possible, so as to solve the problem of disordered input of point cloud data by pooling layer. PointCNN attempts to general the convolutional neural network algorithm shines in 2D image processing to point cloud, which also achieve excellent performance in point cloud feature extraction without a rotation matrix to align the input point cloud. In response to the unstructured nature of unordered point cloud data, DGCNN utilizes its distinct EdgeConv and dynamic graph update methods, as opposed to GCN (Graph Convolutional Network), to learn local point-to-point relationships, further enhancing the network model's ability to extract local features, resulting in impressive performance in point cloud classification and segmentation tasks.These currently research results confirm the superiority of deep learning network in extracting features from 3D point cloud data. The introduction of point cloud processing deep learning networks into the regression task for plant height estimation has the potential to avoid a series of tedious manual tasks in the traditional 3D point cloud processing pipeline, such as noise reduction, outlier deletion, modeling and background removal. However, most of them mainly focus on classification, segmentation and multi-classification tasks (Li et al., 2018; Thomas et al., 2019), not on regression tasks.

In response to the challenges and bottlenecks encountered in estimating plant height by the existing methods, in this paper, we propose a novel deep learning network structure named DRN which can be trained end-to-end to predict lettuce plant height. This new framework can directly estimate the plant height value through the input point cloud data without extra spatial auxiliary information such as DTM and DSM. DRN network is composed of three parts: feature extraction module, RNN module and regression module. On the basis of DRN, we continue to explore the improvement of the network structure to reduce the error between the estimated plant height and the measured plant height. Inspired by the experiment results that the weights of the point cloud coordinates could affect the estimation results of plant height, we set up a Dimension Attention (DA) structure specifically for the DRN model, this structure can learn individual weight values for its x, y, and z coordinates according to the characteristics of every input point cloud sample. The DRN with the Dimension Attention module is called D-DRN, compared with the DRN model, D-DRN achieves a reduction in the prediction error of plant height. More specifically, the main contributions of our work can be summarized as follows.

(1) Proposed a novel end-to-end framework for plant height prediction based on a point cloud feature extraction network and a deep recurrent neural network, which can use the x, y, and z coordinates value of the input 3D point cloud sample to directly regress of the individual plant height.
(2) A series of experiments were designed to compare the influence from the x and y coordinate value of the input point cloud under different weights on the estimation height. It was found that when the weight of x and y coordinates are both 1, the plant height estimation result is generally stable, but the estimation accuracy is not always the best in 500 iterations of training.

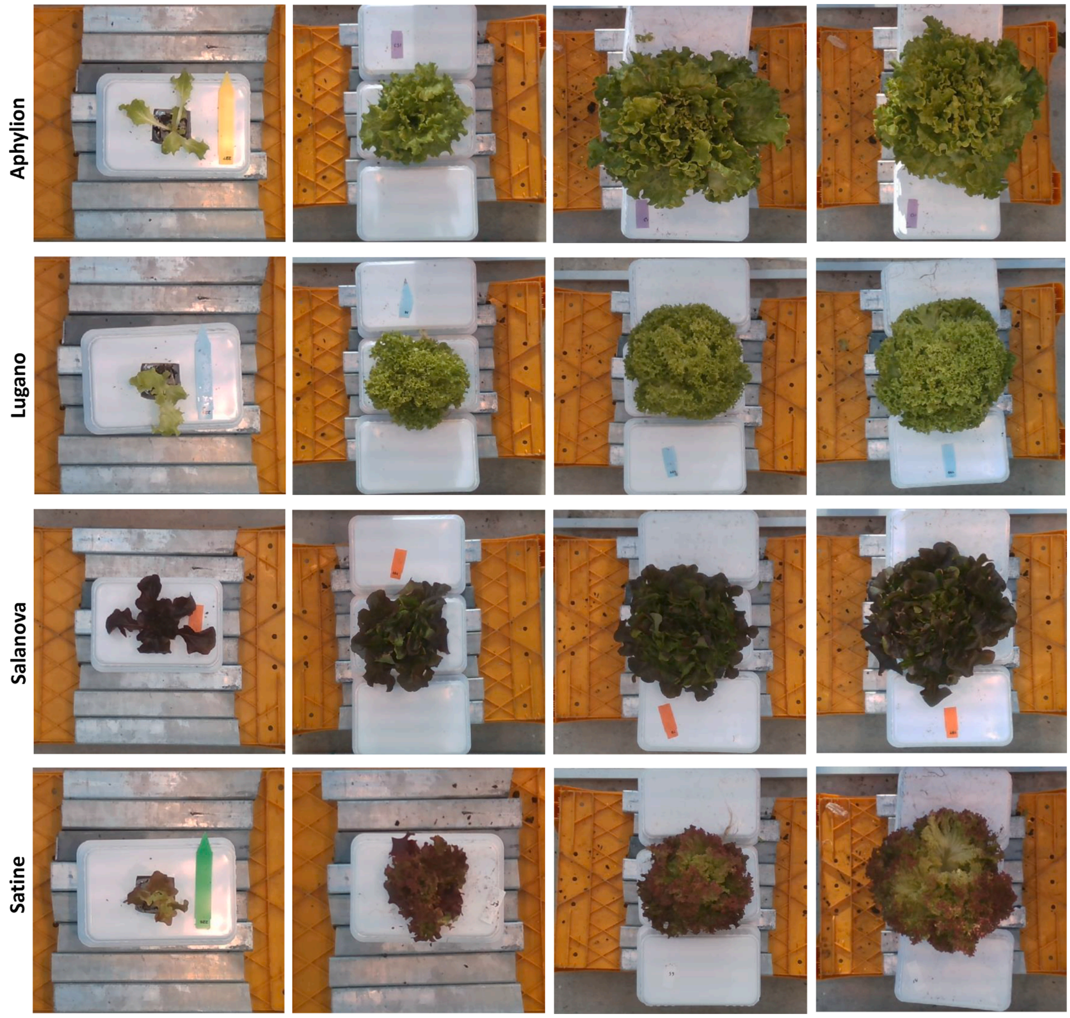

Fig. 3. RGB images of 4 varieties of lettuce samples.

(3) Inspired by the results obtained in (2), We present the Dimension Attention module, which can learn the specific coordinate weights for each input sample point cloud according to their characteristics. The model with a DA module (D-DRN) shows better test results when compared with DRN.

The rest of this paper is organized as follows. The data sets used in this research and the structure of the proposed deep RNN-based neural network along with the dimension attention module in Section2. Section 3 gives experimental results, while Section 5 provides a discussion about the performance of the proposed models. Conclusion is drawn in section 4.

2. Materials and methods

2.1. Dataset preparation

The experimental data are from the Third Greenhouse Growing Challenge (Hemming et al., 2021), that were taken with a RealSense camera (D435) under defined conditions and contained images of individual lettuce plants of 4 varieties in different growth periods and different growing conditions, including AphyLion, Lugano, Salanova and Satine, partially displayed in Fig. 3. Each lettuce samples are connected with information on the ground truth plant height from manual measurement by tape. we transformed these 1080×1920 resolution images and corresponding depth value into point cloud data efficiently by Matrix Operations and CUDA acceleration, then got the sparse point cloud (1024 dimension) by Voxel down sampling and filtering, as is shown in Fig. 4. The depth camera captured RGB and depth image data from the top of the target plant. The point cloud data reconstructed from the bottom of the plant was missing and noisy due to the occlusion of the top of the target plant and the noise of the sensor. A total of 391 samples of lettuce point clouds of 4 types in different time periods were used in the experiment.

2.2. Architecture of the proposed DRN

Since these point cloud that relevant to plant height may be a set of points compose apart from each other, it is obviously unreasonable to directly apply the currently network to point cloud regression tasks such as the prediction of individual plant height. The network that can regress of lettuce plant height requires the ability to learn the relationship between distant point cloud nodes. A network with strong point cloud feature extraction ability is also essential. Hence, the above two factors should be taken into consideration when designing the regression network of lettuce plant height.

The end-to-end lettuce plant height prediction network DRN proposed in this paper is shown in Fig. 5. DRN consists of three parts: a feature extraction network, an RNN, and a regression network. These networks process the input raw sparse plant point cloud data to regress the plant height. In more detail, first the 1024×3-dimension point cloud is processed by a feature extraction network to obtain $1024 \times$

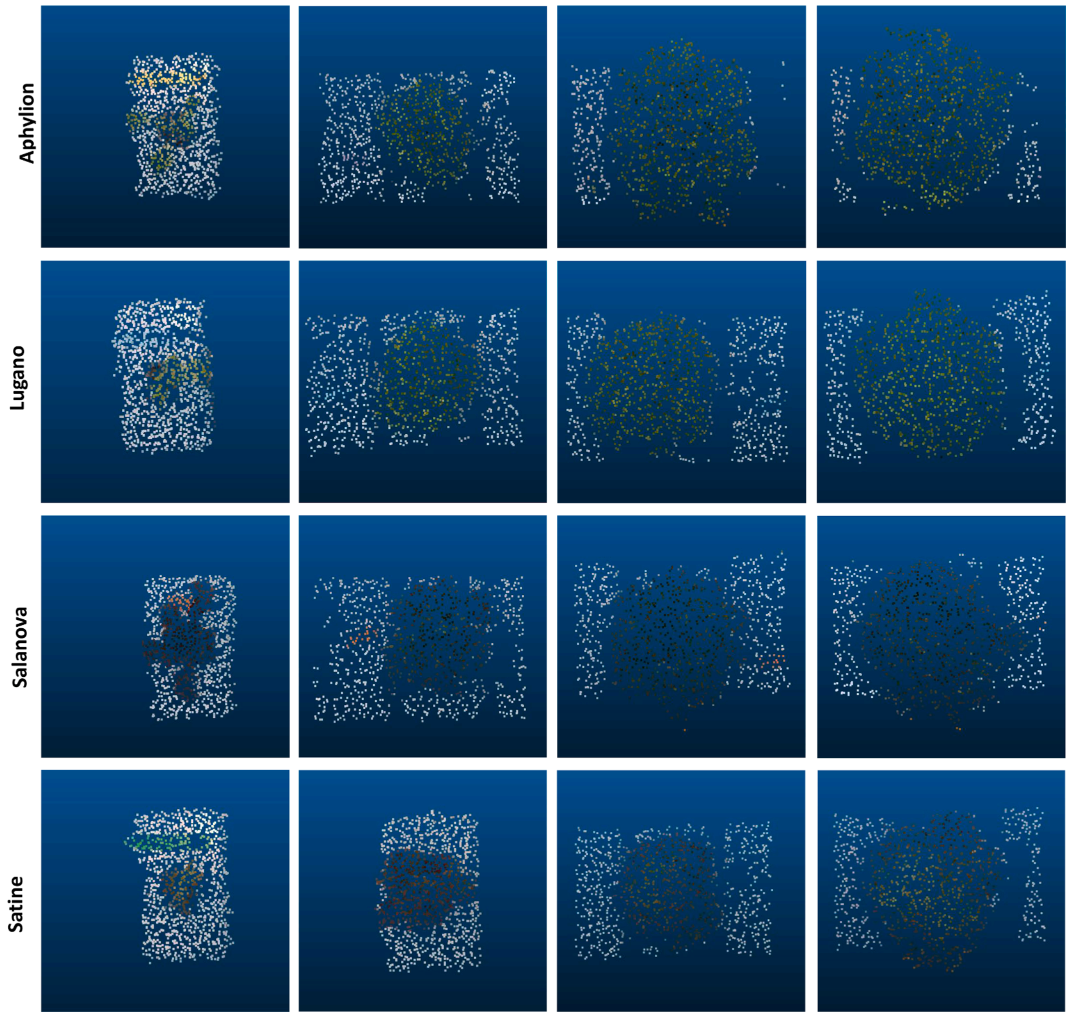

Fig. 4. 3D Point cloud images of lettuce sample after preparation.

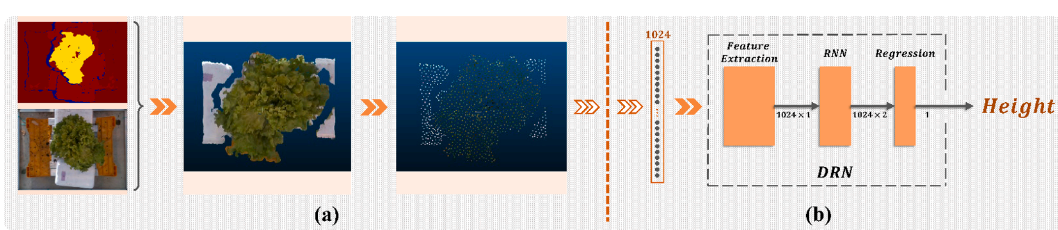

Fig. 5. Overview of end-to-end method. (a) Data and preparation: transforming the given images into point cloud and down-sampling. (b) Regressing lettuce plant height from single perspective point cloud of 1024 x3-dimension by proposed DRN model.

Table 1
Parameter configuration of Feature extraction network.

Layer		Feature Points	Ratio	R	Maximum of Neighbors
SA1	Sampling	512x3	0.5	–	–
	Grouping	512x64x3	–	0.2	64
	PointConv	512x128	–	–	–
SA2	Sampling	128x128	0.25	–	–
	Grouping	128x64x128	–	0.4	64
	PointConv	128x56	–	–	–

1-dimension embedding. Next, we input these embedding into an RNN moudle with input size of 1 and output size of 1 to learn the relationship between them. From RNN moudle we obtain a new embedding with a dimension of 1024×2.We then use a pooling layer to process these new embeddings and send the resulting into regression network to regress of the plant height value. Compared with traditional plant height prediction algorithms based on point cloud data, the proposed DRN network structure combining the powerful point cloud feature extraction model and the RNN model can be trained end-to-end, which means that we can directly regress of the geometric information by raw 3D point cloud without extra handcrafted features. Next, we will introduce the DRN proposed in this article and the upgraded version of DRN, D-DRN in further detail.

2.2.1. *Feature extraction network for point cloud*

For the plant height regression task, the original 3D point cloud data usually contains lots of noise, outliers and redundant information. The

Table 2
The network structure of regression layer.

Layer	Input Size	Output Size
Linear1	1024	512
Linear2	512	1

point cloud feature extraction network module is designed to automatically filtering unnecessary information from raw point cloud data, simultaneously to extract features for effective regression of plant height. The point cloud feature extraction network structure is inspired by PointNet++, which has achieved good results in single classification, multiple classification and segmentation tasks of point cloud data.

The sampling layer is used to select a certain proportion of points from the input feature points as the center point, and the number of selections is determined by the parameter **Ratio**. When the input lettuce point cloud scale is 1024 × 3, the point cloud feature extraction network parameters are shown in Table 1, where **Ratio** represents the ratio of the center points selected by the Sampling layer, and **R** is the normalized radius. The point cloud feature extraction network consists of two set abstractions named SA1 and SA2, each one includes a sample layer, a grouping layer and a PoitConv layer. The Grouping layer takes each selected center point as the cluster center, takes the parameter **R** as the radius, and uses "**Maximum of Neighbors**" as the upper limit of the number of points in the cluster to generate multiple local point clusters. The PointConv layer inputs each local point cluster into the PointNet network to extract local features (see Table 2).

2.2.2. RNN based modeling

As mentioned above, the set of points that determines the height of the lettuce plant might be at different positions in the 3D lettuce point cloud sequence. After local feature extraction by feature extraction network, the factors that determine the height of lettuce are still scattered among the different positions of the high-dimensional extracted feature sequence. Here we use an RNN connected to the feature extraction network to learn the relationship between feature points in different sequence positions to better nonlinearly fit the mathematical relationship between the feature and the height of lettuce. The feature extraction network outputs a 1024-dimensional feature sequence, the ordinary RNN network lacks long-distance memory function, so it is not good at learning the relationship between these feature points that are far away in the sequence position. Inspired by the BERT (Devlin et al., 2018) network structure which has a pivotal position in the Natural Language Processing (NLP) field, we chose the LSTM network (Hochreiter and Schmidhuber, 1997), which performs well in long-distance memory ability to learn the point cloud feature sequence output by the feature extraction network.

The continued transmission of the sequence state at the previous stage and the gate structure formed by the activation function inspired by Neurons endows LSTM with the ability to effectively find the relationship between the features of different positions in the long sequence. Fig. 6 describes the detailed process of information transmission of feature points at different linear positions in the LSTM network space. x_t represents the feature point input at time t, h_{t-1} represents the hidden layer state of the feature point at t-1, c_{t-1} represents the cell state transmitted from the last feature point, the LSTM network updates and feature point p as shown in the following formula:

$$\begin{aligned} i_t &= \sigma(W_i x_t + W_i h_{t-1} + B_i) \\ f_t &= \sigma(W_f x_t + W_f h_{t-1} + B_f) \\ g_t &= \tanh(W_g x_t + W_g h_{t-1} + B_g) \\ o_t &= \sigma(W_o x_t + W_o h_{t-1} + B_o) \\ c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{1}$$

In **Formula 1** above,i_t represents the state of the input gate and o_p represents the state of the output gate,f_t represents the state of the important forget gate in the LSTM network. The state of these doors will determine the degree to which the corresponding information passes, when the status value is 0, the information is blocked, and when the status value is 1, the information is completely passed. W and B represents the weight value and the bias of the connection in the LSTM network, g_t, c_t and h_t respectively represent the input regulatory gate, the state and hidden state of memory cells passed down at time t, $\odot$ denote an element-wise production operation by two feature vectors. In this research, the output size and input size of the LSTM are both set to 1.

2.2.3. Regression layer

After the LSTM network, connect a 2-layer MLP (Multilayer Perception) module to generate the height of the lettuce by using the output of the LSTM network. The parameters of the 2-layer MLP network structure are shown in the following table:

The MLP structure in the model consists of 1024 input layer neurons, 512 hidden layer neurons and one output neuron. In order to avoid overfitting in the MLP network, when the model is trained, the connections between the input layer neurons and the hidden layer neurons are randomly drop out by 50%.

2.2.4. Loss function

If we take the DRN model structure we proposed for lettuce plant height prediction as an encoder, then we need to make the parameters constituting the encoder able to encode the input single perspective point cloud with the same height into a value as close as possible to the real plant height. We use Euclidean distance to describe the deviation between the predicted value of the model and the label value of the plant height corresponding to the point cloud data to guide the adjustment of the model parameters when training. The MSE (Mean Square Error) between the predicted plant height and the actual plant height is used as the loss function as shown in the following formula:

$$L(l,p) = \frac{\sum_{k=1}^{N}(l_k - p_k)^2}{N} \tag{2}$$

In the **Formula 2**, p represents the predicted value of lettuce plant height by DRN model, l represents the label value of lettuce plant height, N represents the total number of samples, $L(l,p)$ represents the MSE of all predicted and true values of the entire data set.

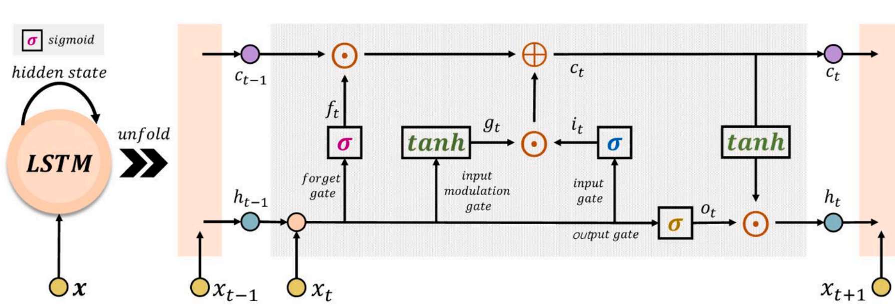

Fig. 6. The structure of LSTM. ⊙ represents product, while ⊕ denote addition.

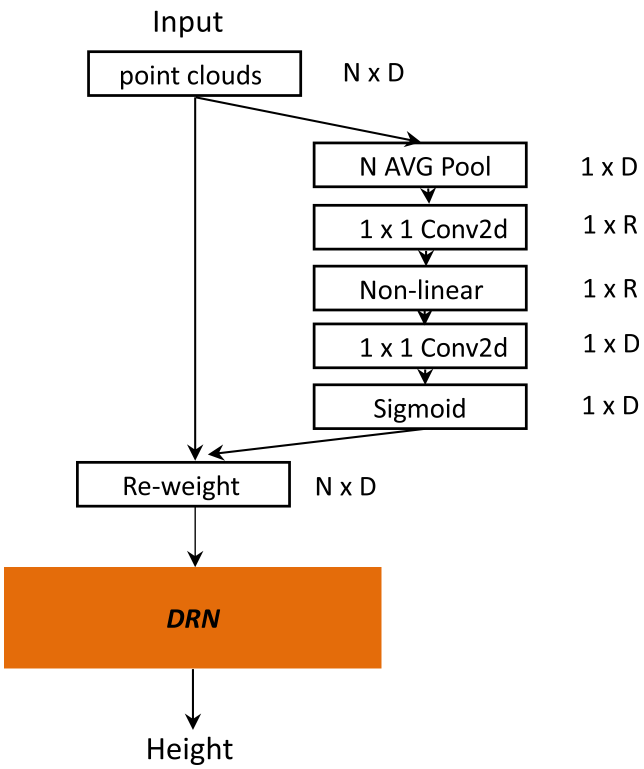

Fig. 7. The structure of D-DRN module.

2.3. D-DRN model

Inspired by the influence of the input of x and y coordinates in different weighted on the estimation results by DRN model, we designed a Dimension Attention module for DRN based on the self-attention mechanism, which can automatically learn the weight of the x, y, and z coordinates of the point cloud according to the characteristics of the input point cloud sample. The new model with a dimension attention module is called D-DRN. The structure of Dimension attention in the D-DRN is shown in Fig. 7 below:

3. Results

We used python programming language, OpenCV and Open3D libraries to read, process and save the original RGB-D data. The hardware platform for model training is an AMD 3900XT CPU and two NVIDIA GeForce RTX3090 graphics cards. The optimizer used to train the model is Adam, the initial learning rate is 0.001, the batch size is 8, and the total epoch is 500. A total of 391 lettuce point cloud samples of 4 types in different periods were used for the experiment, of which 341 samples were selected as the training data, and 50 samples were used as the testing data set. We used MSE as the loss function of the model, simultaneously, the RMSE value between the predicted value and ground truth is used as the evaluation index of the accuracy, the lower the RMSE, the more accurate the prediction of the lettuce plant height by the model.

3.1. A more suitable network structure to regress of the lettuce plant height by point cloud data

As mentioned above, the proposed DRN model has three parts, including feature extraction network, deep RNN and a regression module. As a comparison with DRN, we designed a new network called Baseline that only contains feature extraction network and regression module, to initially demonstrate the effectiveness of deep RNN structure. The original point cloud data is synthesized from the RGB image and the depth image provided in the 3rd Green House Challenge. In order to better show the effect of x and y coordinates on plant height estimation, both x and y in the plant point cloud are multiplied by 0 and input into the network model. The testing RMSE result of Baseline model and DRN model is shown in Fig. 8 below.

The blue line represents the RMSE result estimated by Baseline on a test set composed of 50 lettuce point cloud samples during each training epoch, the green one represents that result from DRN. With the LSTM network structure, the model can better learn the relationship between feature points at a longer distance, therefore, the accuracy of the estimated lettuce plant height is improved. The following experiments will use the DRN model as the baseline model.

3.2. Different weight choice for input point cloud data in x coordinate and y coordinate

In order to make full use of the spatial information contained in the point cloud, we directly use original point cloud data (x, y and z coordinates value) to regress of the plant height of lettuce. Input the above point cloud coordinate value with different weight gradient values into the built DRN model to train 500 epochs, after each training epoch, the accuracy of the model is tested on the test set of 50 lettuce point cloud samples for the current model. The test set is invisible to the model during training, therefore, the performance of the model on the test set has certain guiding significance for the estimation accuracy of the model in practical application in the future.

The results of the x, y coordinate weight gradient experiments are shown in Fig. 9 above, the abscissa is the number of iterations in the

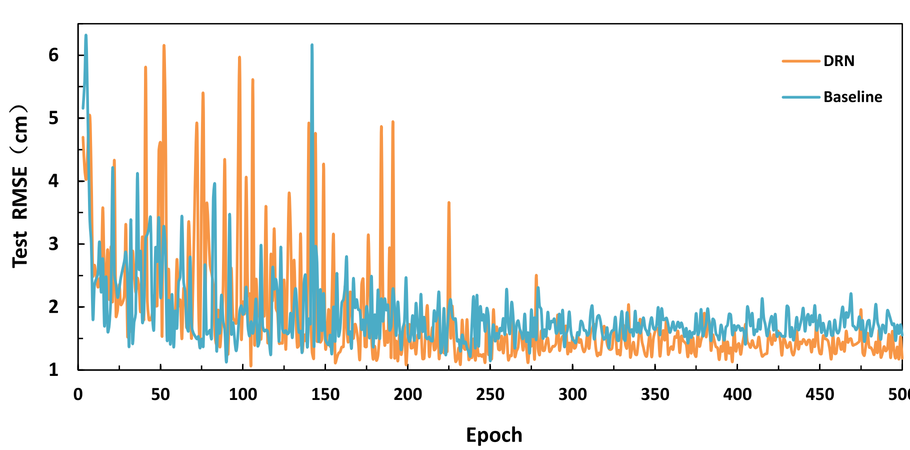

Fig. 8. RMSE of Baseline and DRN in testing data.

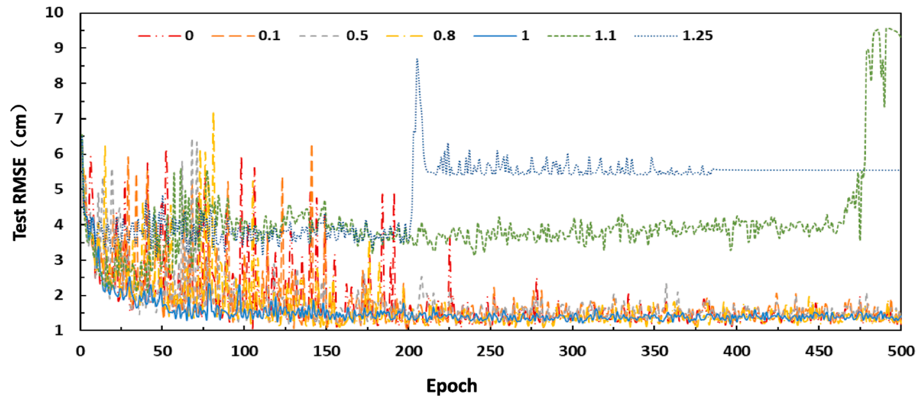

Fig. 9. Testing RMSE curve from DRN model by every epoch of different x, y coordinate weight, the 0, 0.1,0.5.0.8,1,1.1 and 1.25 represents the weights of the x, y coordinate of 3D point cloud.

Table 3
Testing RMSE value from DRN model by every 50 epochs of different x, y coordinate weight.

RMSE (CM)		Epoch Number									
		0	50	100	150	200	250	300	350	400	500
weight	**0.0**	4.07	4.13	3.29	1.92	1.57	1.46	1.41	1.42	1.37	1.37
	0.1	4.11	2.99	1.92	2.20	1.48	1.34	1.51	1.49	1.44	1.44
	0.5	4.14	2.21	1.85	1.66	1.51	1.46	1.56	1.51	1.44	1.44
	0.8	3.97	2.81	1.67	1.56	1.45	1.30	1.34	1.39	1.36	1.37
	1.0	4.15	1.92	1.59	1.54	1.48	1.37	1.37	1.41	1.35	1.36
	1.1	3.90	3.15	3.98	4.14	3.63	3.63	3.73	3.80	3.87	9.49
	1.25	4.24	3.95	3.76	3.96	3.75	5.59	5.58	5.54	5.56	5.55

training process, the ordinate is the RMSE value between the measured plant height and estimated plant height in 50 test samples. The RMSE calculation formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{N-1}} \tag{3}$$

The results of gradient experiments with different weights are represented by line segments of different colors. Since we found that the model no longer converges when the weight value is greater than or equal to 1.1, the experiments with the weight values of 1.5, 1.8 and 2.0 were abandoned. As can be seen from the above Fig. 9, in addition to the weight values of 1.1 and 1.25, the model use input point cloud with the weights from 0 to 1 in x, y coordinate can all converge. Carefully observing these convergent curves, you can find that the RMSE curve has the least volatility when the weight of the input point cloud x and y coordinates is 1, and the others all showed a large degree of volatility, which is particularly evident before the 250th epoch. Although the above Fig. 9 can well visualize the overall performance of each experimental group's plant height prediction under epoch by epoch during the entire training process, there are many overlaps between the curves of the experimental results, which is not convenient for quantitatively analyzing the experimental results. In order to better show the performance of each experimental group on plant height estimation, we averaged the RMSE of the test results from the 1st to the 10th epoch of each experimental group, then, starting from the 50th epoch, the RMSE values obtained from the test of 10 consecutive epochs before them are averaged by Every 50 epochs, the results shown in Table 3.

When the weight of the point cloud x and y coordinates exceeds 1, the RMSE between the estimated plant height and the measured plant height is as high as 9.49 cm (weight = 1.1) and 5.55 cm (weight = 1.25). The results of the model under different weights for the input point cloud in x coordinate and y coordinate are also quite interesting, When the weight of x and y coordinates in point cloud is 0, it showed great volatility when estimating plant height in each epoch of the training phase. The accuracy of plant height estimation in the experimental group with a weight of 0 is not the worst, which is different from our previous expectations. We speculate that when the weight of the x and y coordinates of the point cloud is 0, the 3D point cloud of the lettuce plant becomes a straight line that is perpendicular to the x-o-y plane and coincides with the z-axis in the point cloud 3D coordinate system with the optical center of the sensor as the origin o, which makes the model show overfitting on such point cloud data. When the weight of x and y is 1, the model has better stability and accuracy when estimating the height of lettuce in each epoch. When the weight of input point cloud x and y coordinates is between 0 and 1, the performance of the model is not unworthy of mentioning. Although the estimated RMSE results during the entire training test show poor stability compared to the weight of 1, it showed best performance during a certain epoch of testing. Take the test result of the experimental group with the input point cloud x and y coordinate weight of 0.8 as an example, the average RMSE measured in the first 10 epochs of the 250th epoch reached 1.30 cm, which exceeds the final test result when the input point cloud x, y coordinate weight is 1. From the experimental results of different point cloud x and y coordinate weight gradients, the x and y coordinates of the point cloud do play an important role in the plant height estimation model, it can not only maintain the stability of the plant height regressed by the model, but also improves the accuracy of the estimation results. The above experimental results show that when using the x and y coordinate information of the input point cloud to estimate the height of lettuce, the weight of the x and y coordinates has an impact on the accuracy of the plant height predicted by the model.

3.3. Get the characteristic weight of point cloud coordinates by self-attention module

The experimental results in 3.2 show that the x and y coordinate weights of the input point cloud have an impact on the estimated plant height results of the model. Therefore, it is necessary to find a set of

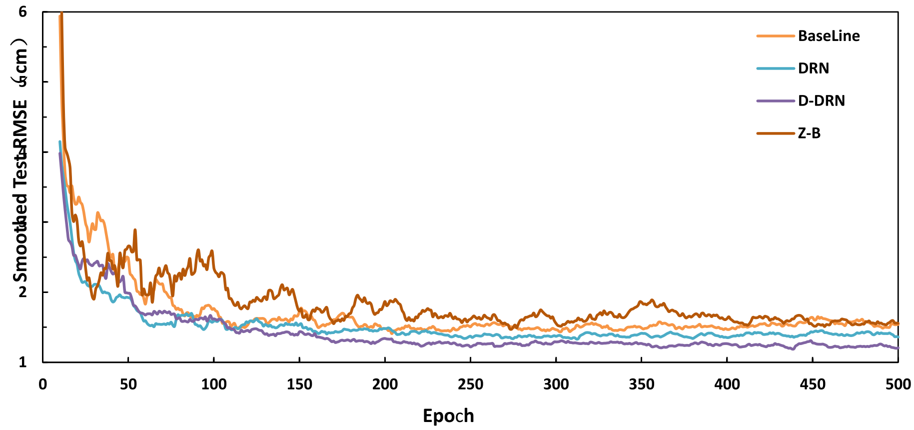

Fig. 10. Moving average RMSE of Z-B, Baseline, DRN and D_DRN.

Table 4
Validation statistic of Z-B, Baseline, DRN and D-DRN model for plant height estimation on testing data.

Model	X, Y Coordinate	RNN Module	DA Module	Regression Model	R^2	RMSE (CM)	RRMSE (%)
Z-B				$Y = 0.8004 \times x + 1.5715$	0.956	1.697	13.379
Baseline	√			$Y = 0.8654 \times x + 1.1209$	0.941	1.480	11.668
DRN	√	√		$Y = 0.888 \times x + 1.1995$	0.948	1.276	10.060
D-DRN	√	√	√	$\mathbf{Y = 0.9073 \times x + 1.108}$	**0.960**	**1.101**	**8.680**

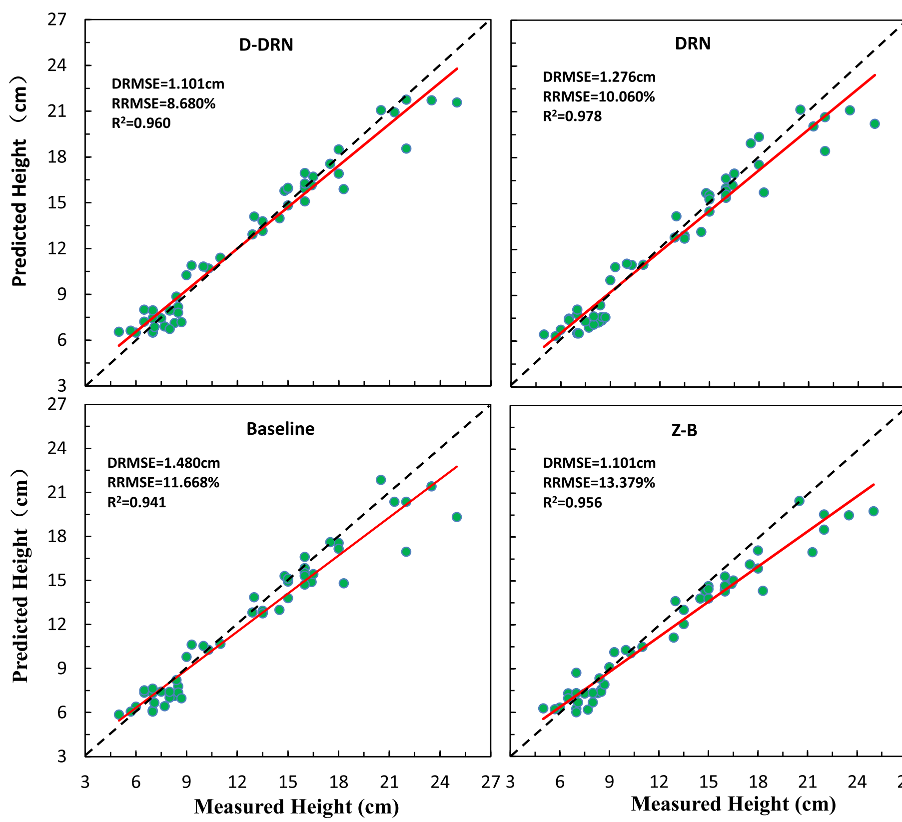

Fig. 11. Scatter plot of cross-validation between measured plant height and predicted plant height of proposed model.

appropriate weight values for the coordinates of the point cloud. A simple but time-consuming method is to set different weight gradient combinations between 0 and 1 for x, y and z coordinates to find the optimal weight value, in addition to time-consuming and labor-intensive, this method has another drawback, that is, different plant point clouds may have their own weight combinations. Wouldn't it be a good idea to let the model learn a set of point cloud x, y, z coordinate weights based on the input point cloud data? On the basis of the existing DRN, we added a x, y, z coordinate weight learning module based on self-attention structure to obtain appropriate coordinate weights

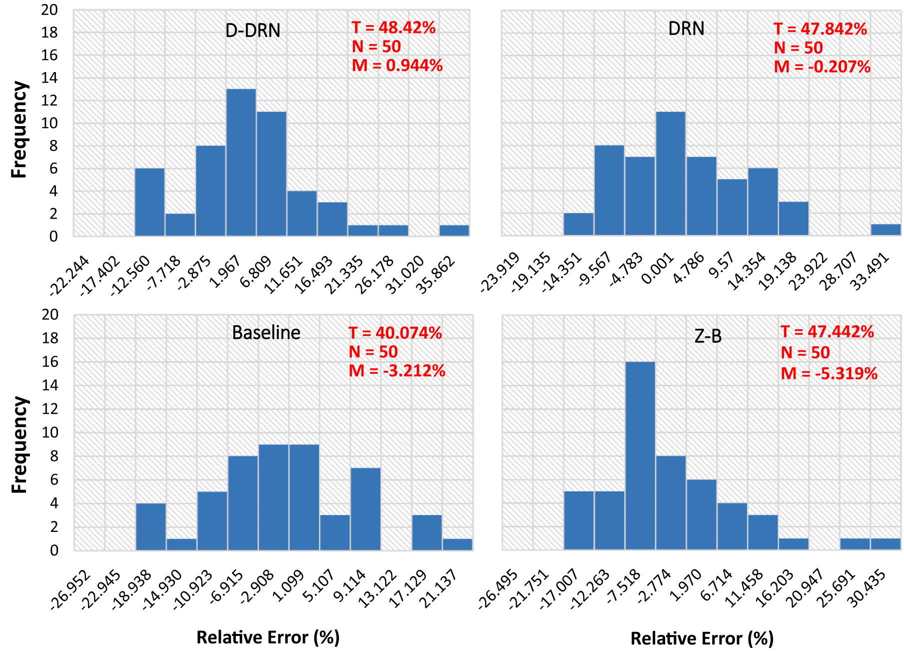

Fig. 12. Frequency distribution histogram of relative error between predicted plant height and measured plant height from proposed models.

Table 5
Describe statistics of the measured plant height of four lettuce varieties on the testing data.

variety	Type of Height	Min (cm)	Max (cm)	Mean (cm)	SD (cm)	CV (%)
Aphylion	Measured	5	25	16.65	7.008	42.090
Lugano	Measured	6.5	18	12.275	4.112	33.499
Salanova	Measured	5.7	16	10.677	3.811	35.694
Satine	Measured	6	18.3	11.408	4.536	39.762

according to the input point cloud data of each plant, we called the new model structure D-DRN. To better evaluate the effectiveness of the designed self-attention based D-DRN, We introduced a new model Z-B (Set the input point cloud x-coordinate and y-coordinate weights of the baseline model to zero). After setting the x and y coordinate values of the input point cloud to zero, the Z-B model relies mainly on the z coordinate values of the point cloud for feature extraction and plant height regression, which was used to simulate the situation in the traditional pipeline for plant height prediction where the z coordinate values of the three-dimensional spatial coordinates of the point cloud are mainly used for plant height calculation. The results of D-DRN comparing with the Z-B, Baseline and DRN models are shown in Fig. 10 below.

The abscissa indicates the number of iterations, the ordinate table is the moving average of RMSE obtained from each epoch test, and the average is taken every 10 epochs. After moving average processing, the test result curve is smoother, which can more clearly show the difference of each model in RMSE between the estimated value and the measured value on the testing data. The experimental results on Fig. 10 show that the model, with a self-attention module to learn the point cloud coordinate weights, starting from the 100th epoch, the RMSE sliding average on the test set is lower than Baseline and DRN. The performance of D-DRN model on the test set before the 100th epoch is worse than that of DRN. We speculate that it may be because the model with the added weight learning module increases the learning burden, so that relatively more epochs are required to train the model parameters. The amplitude of oscillation and RMSE values of the brown curve in the above figure are significantly higher than the other three curves, saying that the Z-B model that did not use the point cloud x and y coordinate values to regress the plant height values showed relatively low accuracy and relatively high instability. We analyze that this may be due to the fact that only one dimension (z-axis) of the point cloud data input to the model is involved in the regression of plant height using only the point cloud z-coordinate values, which may lead to underfitting.

Table 6
Validation statistic of proposed regression models for plant height estimation of 4 lettuce varieties.

Variety	Estimation Model	Regression Model	R2	RMSE (CM)	RRMSE (%)
Aphylion	Z-B	$Y = 0.7594 \times x + 2.242$	0.949	2.670	16.035
	Baseline	$Y = 0.8064 \times x + 1.8832$	0.917	2.476	14.872
	DRN	$Y = 0.8156 \times x + 1.9304$	0.954	2.065	12.403
	D-DRN	$\mathbf{Y = 0.8598 \times x + 1.5859}$	**0.962**	**1.666**	**10.009**
Lugano	Z-B	$Y = 0.8075 \times x + 1.8524$	0.950	1.173	9.558
	Baseline	$Y = 0.9579 \times x + 0.4598$	0.949	**0.896**	**7.299**
	DRN	$Y = 1.026 \times x + 0.0463$	0.955	0.960	7.819
	D-DRN	$\mathbf{Y = 1.0006 \times x + 0.3878}$	**0.956**	0.931	7.583
Salanova	Z-B	$Y = 0.8147 \times x + 1.5215$	0.941	1.126	10.546
	Baseline	$Y = 0.971 \times x - 0.055$	0.962	0.801	7.500
	DRN	$Y = 0.9698 \times x + 0.315$	**0.966**	**0.674**	**6.317**
	D-DRN	$\mathbf{Y = 1.0156 \times x - 0.0728}$	0.957	0.727	6.805
Satine	Z-B	$Y = 0.827 \times x + 1.06$	0.959	1.399	12.260
	Baseline	$Y = 0.8845 \times x + 0.7169$	0.951	1.174	10.292
	DRN	$\mathbf{Y = 0.9824 \times x + 0.0725}$	0.949	1.008	8.835
	D-DRN	$Y = 0.923 \times x + 0.8404$	**0.961**	**0.880**	**7.711**

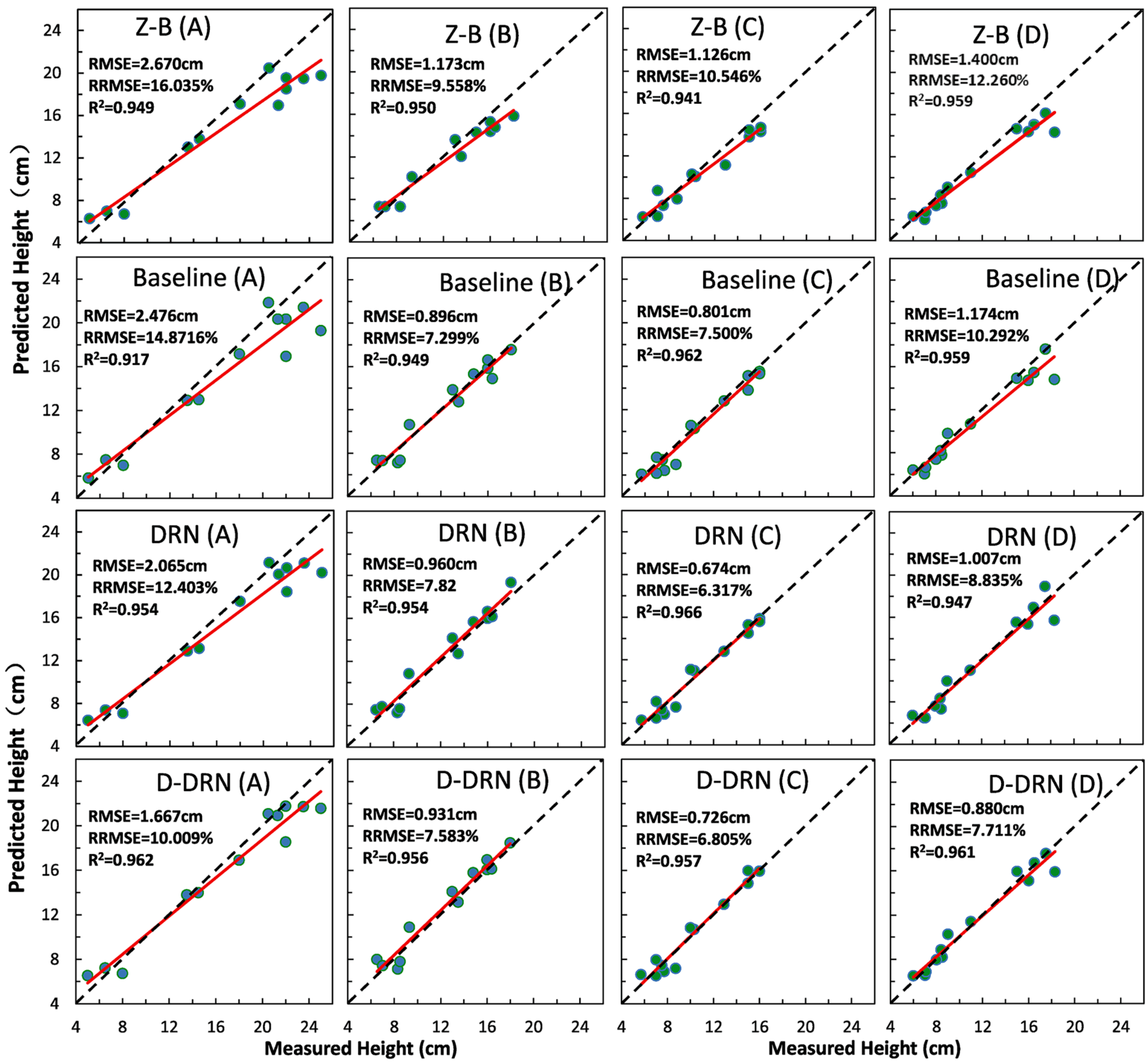

Fig. 13. Scatter plot of cross-validation between predicted and measured plant height in four lettuce varieties of proposed methods. A, B, C and D represent Aphylion, Lugano, Salanova and Satine, respectively.

4. Discussion

From 4.1 to 4.3, we display and analyze the estimation results of the proposed models on testing set when trained in each epoch. This part will analyze and discuss the prediction results of the D-DRN, DRN, Baseline and Z-B models on testing data after the training is completed.

4.1. Regression of the final prediction results and true values of plant height on the test set of each model

The estimated lettuce plant height value predicted by the model constructed by the deep learning network has certain volatility, we perform 10 times estimation on each lettuce point cloud sample in the test set and take the average as the final estimated plant height value. As shown in the following Table 4, the RMSE of the model Z-B is 1.697 cm, and the RRMSE is 13.379%, both of which are the highest among the above four models. Compared with Baseline model, RMSE is 14.26% higher and RRMSE is 14.66% higher. This result indicates that when using point cloud data to estimate the height of lettuce plants, the participation of the point cloud's x-coordinate and y-coordinate values in modeling could improve the accuracy of estimated plant height. Compared with the Baseline, the DRN model has a decrease of about 13.78% in both RMSE and RRMSE on the testing date, which illustrates the effectiveness of the proposed DRN structure in predicting the height of lettuce. The RMSE value of D-DRN has dropped by about 13.71% on the basis of DRN, which means that the introduction of the Dimension Attention (DA) module to automatically assign weights to the x, y, and z coordinates of the input point cloud helps the model to fit the height of lettuce.

The best regression model for plant height estimation is highlighted in boldface.

The RMSE and RRMSE indicators mainly reflect the difference between the predicted value and the true value. In order to further analyze the relationship between the estimated plant height by these models and the corresponding measured value, we performed regression analysis on 50 test samples to obtain the coefficient determination (R^2), as shown in column R^2 in Table 4. R^2 is known as the goodness of fit, which is used to indicate the closeness of the relationship between the independent variable and the dependent variable in the regression. Under certain conditions, the larger its value, the better the fit of the regression model. This result is different from the performance of RMSE and RRMSE indicators on these models, which also shows that only RMSE and RRMSE are used to evaluate the quality of the plant height prediction model is limited.

We made a cross-validation scatter plot of the predicted lettuce plant height of the above models on the test set, as shown in the above Fig. 11. The black dashed line in the figure is composed of the function: $y = x$, which represents the state when the predicted value and the measured value are completely equal, and the red solid line is the linear curve regressed by measured plant height as the independent variable and the corresponding predicted value as the dependent variable. We supposed that the model proposed in this study can well handle the nonlinear relationship between lettuce point cloud data and lettuce plant height. Therefore, we use a simple regression function to describe the relationship between the predicted plant height of the model and the true plant height, namely $h_{estimated} = k \times h_{measured} + b$. When the measured value of lettuce plant height is very close to the predicted value, the slope k of the linear regression equation should tend to 1, while the intercept b tends to 0. It can be seen from the regression model column in the above Table 4, the regression linear equation slope k value obtained by the Z-B model is 0.8004, which is the smallest in each model, and the intercept b is 1.5715, which is the largest among models. Through the cross-validation scatter plot, it can also be seen that the deviation angle between the regression fitting curve of the Z-B model and the $y = x$ curve is the largest. Therefore, to evaluate from the slope and intercept of the regression equation, the Z-B model performs the worst among the models. We speculate that after setting the x-coordinate value and y-coordinate value of the input point cloud to zero, the lettuce point cloud data becomes a series of discrete points distributed on the z-axis in the 3D space. This makes the DRN model framework based on the deep learning network overfitting on it, which increases the systematic deviation between the lettuce plant height predicted by the model and the measured value. The D-DRN model, which integrates the DRN network framework and the DA module, can better nonlinearly fit the relationship between lettuce point cloud and lettuce plant height. Perhaps because of this, the D-DRN model has achieved the best results in RMSE, RRMSE and regression analysis indicators.

4.2. Histogram analysis of relative error and regression analysis by variety

When constructing the lettuce plant height prediction model, we used the MSE loss function which is commonly used in regression tasks. Therefore, we chose RMSE and RRMSE indicators when evaluating the trained model. Since both RMSE and RRMSE are positive values, which only reflect the degree of difference between the predicted value and the measured value of plant height, but cannot describe whether the predicted value is more or less than the measured one. To make up for this, we first calculate the error between the predicted plant height and the measured plant height in each test sample, and then divide it with the measured plant height to obtain the relative error. The relative error can not only reflect the degree of difference between the predicted value and the measured value of each lettuce point cloud test sample (Similar to RRMSE), but also can see whether the predicted plant height value is larger or smaller than the measured plant height value.

After calculating the relative errors of the above models on the testing data, draw a histogram of the relative error frequency distribution as shown in Fig. 12 above. It can be seen that the relative error between the predicted lettuce plant height and the measured plant height is distributed between around −25% and 35%, the relative error groups with higher frequencies are basically distributed near both sides of M (Mean of relative error). Looking further, in addition to the Z-B model, the relative error grouping point corresponding to the highest frequency of the D-DRN, DRN, and Baseline models is near by 0%. The relative error grouping point corresponding to the highest frequency 13 of D-DRN 1 is 1.967%; the error grouping point corresponding to the highest frequency 11 of the DRN model is 0.001%. There are two error grouping point corresponding to the highest frequency 9 for baseline model, respectively − 2.908 and 1.099. The frequency distribution histograms obtained by models do not signal a standard normal distribution, they all show a discrete distribution (Doane, 1976). The existence of abnormal measurement values is usually considered as one of the reasons for the discrete or island distribution. Our experimental data are provided by the organizer of the Third Greenhouse Growing Challenge (Hemming et al. 2021) in multiple batches. After the first few batches of data were given, the organizer has corrected some obviously inconsequent measured plant height value. The 50 testing samples were provided in the final batch; therefore, we do not rule out the possibility that the discrete distribution of the histogram comes from measurement errors. The model's sensitivity to certain specific test samples (such as lack of training on specific samples) may also lead to discrete distribution. The relative error histogram obtained by the Z-B model shows a skewed distribution. The relative error histograms obtained by the D-DRN model and the DRN model show multimodal distribution, while the Baseline model is closer to flat-topped distribution. The existence of more than one distribution in the data is usually considered a factor for flat-topped and multimodal distribution. As mentioned above, the training and testing data of our experiments consist of 4 lettuce varieties at different growth periods. Some of the RGB and point cloud images from them are shown in Fig. 3 and Fig. 4. We speculate that the difference in the nonlinear fitting ability of the model on the point cloud data of different varieties of lettuce results in multiple distributions on relative error data. Next, we divide the test set samples by varieties and analyze the performance of each model on them. The testing date contains 50 lettuce sample data from 4 varieties of lettuce, of which Aphylion and Lugano each account for 12, and Salanova and Satine each account for 13 samples. The descriptive statistical results of the measured plant height of lettuce Aphylion, Lugano, Salanova and Satine in the test set shown in the following Table 5. Among the four lettuce varieties, the average measured plant height of Aphylion is the largest, around 16.65 cm; Lugano comes next, around 12.275 cm; measured average plant height of Satine is the smallest, around 11.408 cm.

The best regression model for plant height estimation is highlighted in boldface.

From the view of different lettuce varieties, all of models achieved the highest RMSE and RRMSE values on Aphylion, among them the highest RMSE value is 2.670 cm which was obtained by Z-B model (shown in Table 4). Another interesting phenomenon is that the RRMSE of the Baseline, DRN and D-DRN in the varieties Lugano, Salanova and Satine is relatively small and mostly less than 10%. This may be able to explain why the relative error frequency histogram of above models shows multimodal and flat-topped distribution. Besides the differences in the point cloud morphology of different lettuce varieties, the descriptive statistical results in Table 5 show that the average plant height and the CV of Aphylion are larger than other varieties. We supposed it is these differences that make the model's estimation ability differ among different varieties. From the perspective of models, as the results shown in Table 6, the prediction accuracy of the D-DRN model for lettuce plant height is more prominent in all varieties, especially in Aphylion, all indicators are better than the other models. This shows that the dimension attention module we proposed has an advantage in handling lettuce point clouds with more complex plant height distribution. Without using the point cloud x and y coordinates, the Z-B model achieved the largest RMSE values in all four varieties, which indirectly shows the important role of the point cloud x and y coordinates in the estimation of lettuce plant height. Except for Aphylion, the RRMSE of the DRN and D-DRN models with RNN structure is lower than 10% in other varieties. As for Aphylion, the RRMSE of the Baseline model is 19.906% higher than that of the DRN model. These results reveal that the RNN structure also helps to deal with relatively more complex samples (see Fig. 13).

The algorithm model proposed in this study can directly predict plant height using the original point cloud of individual plants as input, which does not involve manual tuning or designing parameters to extract features compared with typically traditional pipelines of plant height prediction based on point cloud data (Bendig et al., 2014; Bendig et al., 2015; Maimaitijiang et al., 2019). Due to the robust nature of the deep learning model, when predicting the height of a single plant within a dense vegetation zone, the original point cloud data from an individual segmented plant can be input into the pre-trained model to predict height phenotype metrics at a single plant scale within the region. The existing literature shows that the feature extraction module Pointnet++ in DRN model has been successfully used in dense plant point cloud feature extraction based on SFM and Lidar data (Herrero-Huerta et al., 2021; Jin et al., 2020; Jing et al., 2021).We believe that the algorithm model proposed in this study is expected to be used in other 3D point cloud data for plant height prediction.

5. Conclusion

In this study, we proposed a novel deep RNN and convolution operations-based end-to-end training network called DRN for accuracy single plant height prediction. The testing results show that DRN can regress of lettuce height effectively by raw perspective sparse point clouds without extra SA information such as DTM, DSM and GCP, etc. This deep learning base network is supposed to be easily for transplanting to other kinds of plant for height estimation and system integration with the back-end system. Afterwards, we further explored the model improvement strategies. Based on the experimental result about how the weights of the x, y, z coordinate influences the regression result of plant height, we designed a DA module at the front end of the feature extraction network to learning the characteristic coordinate weights for every input point cloud sample. Extensive experiment results show that the model can achieve more accurate prediction plant height with the coordinate weights-learning attention module.Combined with a large-scale plant point cloud acquisition system and a single plant point cloud segmentation algorithm, we suppose that these proposed models might be helpful to achieve high-throughput acquisition of plant height trait at a single plant scale. This research has only used lettuce point cloud data as the research object, therefore, the validity of the model proposed in this study on other kinds of plant point cloud data still needs further study. In summary, our work explores a new plant height estimation algorithm system, mainly include the idea of constructing plant height prediction model based on point cloud data and a direction to improve the accuracy of plant height estimation by adjusting the model network structure. The operators we used to build model are mainly from pytorch and pytorch-geometric. The source code and the download link of dataset of our research can be found at https://github.com/LeJson/DRN, We hope it helpful to encourage further development in plant height estimated algorithm based on point cloud data.

CRediT authorship contribution statement

Jingsong Li: Methodology, Writing – original draft. **Ying Wang:** Visualization, Investigation. **LiHua Zheng:** Data curation, Writing – review & editing. **Man Zhang:** Supervision. **Minjuan Wang:** Supervision, Writing – review & editing, Conceptualization.

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

Data will be made available on request.

Acknowledgement

We thank the Key Lab of Smart Agriculture Systems for providing an experimental platform. We especially thank the official holding the Third Greenhouse Growing Challenge for giving out a fantastic dataset.

This work was supported in part by National Natural Science Foundation of China (32201654), the Natural Science Foundation of Shandong Province (ZR2021MC021) and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (No. VRLAB2021A02).

Fundings

This work was supported in part by National Natural Science Foundation of China (32201654), the Natural Science Foundation of Shandong Province (ZR2021MC021) and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (No. VRLAB2021A02).

References

Andujar, D., Ribeiro, A., Fernandez-Quintanilla, C., & Dorado, J. (2016). Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. *Computers and Electronics in Agriculture, 122*, 67–73.
Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science, 19*(1), 52–61.
Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., & Bareth, G. (2014). Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. *Remote Sensing, 6*(11), 10395–10412.
Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., … Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation, 39*, 79–87.
Bernotas, G., Scorza, L. C. T., Hansen, M. F., Hales, I. J., Halliday, K. J., Smith, L. N., … McCormick, A. J. (2019). A photometric stereo-based 3D imaging system using computer vision and deep learning for tracking plant growth. *Gigascience, 8*(5).
De Souza, C. H. W., Lamparelli, R. A. C., Rocha, J. V., & Magalhaes, P. S. G. (2017). Height estimation of sugarcane using an unmanned aerial system (UAS) based on structure from motion (SfM) point clouds. *International Journal of Remote Sensing, 38* (8–10), 2218–2230.
Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.J.a.e.-p., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, pp. arXiv: 1810.04805.
Doane, D. P. (1976). *AESTHETIC FREQUENCY CLASSIFICATIONS. AMERICAN STATISTICIAN, 30*(4), 181–183.
Gao, S., Niu, Z., Sun, G., Zhao, D., Jia, K., & Qin, Y. C. (2015). Height Extraction of Maize Using Airborne Full-Waveform LIDAR Data and a Deconvolution Algorithm. *Ieee Geoscience and Remote Sensing Letters, 12*(9), 1978–1982.
Grenzdorffer, G. J. (2014). Crop Height Determination with Uas Point Clouds. *Isprs Technical Commission I Symposium, 40–1*, 135–140.
Han, L., Yang, G. J., Yang, H., Xu, B., Li, Z. H., & Yang, X. D. (2018). Clustering Field-Based Maize Phenotyping of Plant-Height Growth and Canopy Spectral Dynamics Using a UAV Remote-Sensing Approach. Frontiers. *Plant Science, 9*.
Hemming, S., de Zwart, H.F., Elings, A., bijlaard, m., Marrewijk, v., Bart and Petropoulou, A., 2021. 3rd Autonomous Greenhouse Challenge: Online Challenge Lettuce Images. 4TU.ResearchData.
Herrero-Huerta, M., Tolley, S., Tuinstra, M.R. and Yang, Y., 2021. Individual maize extraction from UAS imagery-based point clouds by 3D deep learning, Conference on Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping VI. Proceedings of SPIE, Electr Network.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
Huett, C., Tilly, N., Schiedung, H., & Bareth, G. (2016). *Potential of multitemporal tandem-x derived crop surface models for maize growth monitoring* (pp. 803–808). XXIII ISPRS CONGRESS, COMMISSION VII.
Hyyppa, E., Kukko, A., Kaijaluoto, R., White, J. C., Wulder, M. A., Pyorala, J., … Hyyppa, J. (2020). Accurate derivation of stem curve and volume using backpack mobile laser scanning. *Isprs Journal of Photogrammetry and Remote Sensing, 161*, 246–262.
Jin, S., Su, Y., Gao, S., Wu, F., Ma, Q., Xu, K., … Guo, Q. (2020). Separating the Structural Components of Maize for Field Phenotyping Using Terrestrial LiDAR Data and Deep Convolutional Neural Networks. *Ieee Transactions on Geoscience and Remote Sensing, 58*(4), 2644–2658.
Jing, Z., Guan, H., Zhao, P., Li, D., Yu, Y., Zang, Y., … Li, J. (2021). Multispectral LiDAR Point Cloud Classification Using SE-PointNet plus. *Remote Sensing, 13*(13).
Kataoka, T., Kaneko, T., Okamoto, H. and Hata, S., 2003. Crop growth estimation system using machine vision. Proceedings of the 2003 Ieee/Asme International Conference on Advanced Intelligent Mechatronics (Aim 2003), Vols 1 and 2: 1079-1083.
Lei, J., Song, J., Peng, B., Li, W., Pan, Z., & Huang, Q. (2022). C2FNet: A Coarse-to-Fine Network for Multi-View 3D Point Cloud Generation. *Ieee Transactions on Image Processing, 31*, 6707–6718.
Li, Y.Y., Bu, R., Sun, M.C., Wu, W., Di, X.H. and Chen, B.Q., 2018. PointCNN: Convolution On X -Transformed Points. Advances in Neural Information Processing Systems 31 (Nips 2018), 31.

Lu, J., Cheng, D., Geng, C., Zhang, Z., Xiang, Y., & Hu, T. (2021). Combining plant height, canopy coverage and vegetation index from UAV-based RGB images to estimate leaf nitrogen concentration of summer maize. *BIOSYSTEMS ENGINEERING, 202*, 42–54.

Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K. T., … Fritschi, F. B. (2019). Vegetation Index Weighted Canopy Volume Model (CVMVI) for soybean biomass estimation from Unmanned Aerial System-based RGB imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 151*, 27–41.

Malambo, L., Popescu, S. C., Murray, S. C., Putman, E., Pugh, N. A., Horne, D. W., … Bishop, M. (2018). Multitemporal field-based plant height estimation using 3D point clouds generated from small unmanned aerial systems high-resolution imagery. *International Journal of Applied Earth Observation and Geoinformation, 64*, 31–42.

Miao, T., Wen, W. L., Li, Y. L., Wu, S., Zhu, C., & Guo, X. Y. (2021). & para; Label3DMaize: Toolkit for 3D point cloud data annotation of maize shoots. *Gigascience, 10*(5).

Mortensen, A. K., Bender, A., Whelan, B., Barbour, M. M., Sukkarieh, S., Karstoft, H., & Gislum, R. (2018). Segmentation of lettuce in coloured 3D point clouds for fresh weight estimation. *Computers and Electronics in Agriculture, 154*, 373–381.

Qi, C.R., Su, H., Mo, K., Guibas, L.J. and Ieee, 2017a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, 30TH IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), pp. 77-85.

Qi, C.R., Yi, L., Su, H. and Guibas, L.J., 2017b. PointNet plus plus : Deep Hierarchical Feature Learning on Point Sets in a Metric Space, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30 (NIPS 2017).

Song, Y., & Wang, J. F. (2019). Winter Wheat Canopy Height Extraction from UAV-Based Point Cloud Data with a Moving Cuboid Filter. *Remote Sensing, 11*(10).

Sumesh, K.C., Ninsawat, S. and Som-ard, J., 2021. Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. COMPUTERS AND ELECTRONICS IN AGRICULTURE, 180.

Tao, H., Xu, L., Feng, H., Yang, G., Yang, X., Miao, M., & Dai, Y. (2019). Estimation of plant height and biomass of winter wheat based on UAV digital image. *Transactions of the Chinese Society of Agricultural Engineering, 35*(19), 107–116.

Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F. and Guibas, L.J., 2019. KPConv: Flexible and Deformable Convolution for Point Clouds. 2019 Ieee/Cvf International Conference on Computer Vision (Iccv 2019): 6420-6429.

Tian, J. R., Dai, T. T., Li, H. D., Liao, C. R., Teng, W. X., Hu, Q. W., … Xu, Y. N. (2019). A Novel Tree Height Extraction Approach for Individual Trees by Combining TLS and UAV Image-Based Point Cloud Integration. *Forests, 10*(7).

Vazquez-Arellano, M., Reiser, D., Paraforos, D. S., Garrido-Izard, M., Burce, M. E. C., & Griepentrog, H. W. (2018). 3-D reconstruction of maize plants using a time-of-flight camera. *Computers and Electronics in Agriculture, 145*, 235–247.

Wang, Q., Che, Y., Chai, H., Shao, K., Li, B., & Ma, Y. (2021). Quantitative Evaluation of Sugar Beet Plant Height Based on UAV-RGB and UAV-LiDAR. *Transactions of the Chinese Society for Agricultural Machinery, 52*(3), 178–184.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic Graph CNN for Learning on Point Clouds. *ACM Transactions on Graphics, 38*(5).

Wu, S., Wen, W. L., Xiao, B. X., Guo, X. Y., Du, J. J., Wang, C. Y., & Wang, Y. J. (2019). An Accurate Skeleton Extraction Approach From 3D Point Clouds of Maize Plants. Frontiers. *Plant Science, 10*.

Xie, T., Li, J., Yang, C., Jiang, Z., Chen, Y., Guo, L. and Zhang, J., 2021. Crop height estimation based on UAV images: Methods, errors, and strategies. Computers and Electronics in Agriculture, 185.

Yang, S., Zheng, L. H., Gao, W. L., Wang, B. B., Hao, X., Mi, J. Q., & Wang, M. J. (2020). An Efficient Processing Approach for Colored Point Cloud-Based High-Throughput Seedling Phenotyping. *Remote Sensing, 12*(10).

Zhu, C., Miao, T., Xu, T., Li, N., Deng, H., & Zhou, Y. (2021). Ear segmentation and phenotypic trait extraction of maize based on three-dimensional point cloud skeleton. *Transactions of the Chinese Society of Agricultural Engineering, 37*(6), 295–301.

Zolkos, S. G., Goetz, S. J., & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment, 128*, 289–298.