2022 10th International Conference on Information Systems and Computing Technology (ISCTech) | 979-8-3503-3293-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISCTECH58360.2022.00093

A Multi-head Two-level Attention-based Network for Plant-part Segmentation on 3D Point Cloud

1st Lin Luo
School of Computer Science and Technology
Guangdong University of Technology
Guangzhou, China
2845211183@qq.com

2nd An Zeng*
School of Computer Science and Technology
Guangdong University of Technology
Guangzhou, China
zengan@gdut.edu.cn

3rd Dan Pan
School of Electronics and Information
Guangdong Polytechnic Normal University
Guangzhou, China
pandan@gpnu.edu.cn

***Abstract*—The automated and accurate segmentation of structural plant parts from 3D point cloud has becoming a promising approach for non-destructive and high-throughput measurement of sought-after phenotypic traits in plant phenotyping. However, previous studies towards automatic segmentation of plants relied heavily on environmental circumstances, vulnerability toward occlusion, sensor set-up, hand-crafted features, as well as manually tuned thresholds. The purpose of this study was to establish a complete framework for plant-part segmentation of the point cloud in 3D space. To this end, we built an image capturing system to capture Caladium bicolor images for 3D reconstruction, and then hand-labeled the point cloud data. Furthermore, based on attention mechanism, we proposed a Muti-head Two-level Attention-based Network (MTANet) to hierarchically capture geometric features and predict per-point semantic annotations directly on the basis of the fully labelled point cloud data. The experimental results outperformed other main-stream point-based deep learning architectures with Overall Accuracy (OAcc), mean Accuracy (mAcc), and the mean Intersection over Union (mIoU) reached 98.73%, 94.67% and 92.58%, respectively. In addition, we also conducted ablation experiments to evaluate its efficiency and examine specific decisions of the network design. This study may contribute to automated measurements of plant phenotypic traits in breeding programs.**

Keywords- Plant phenotyping, Plant-part segmentation, Deep learning, 3D point cloud, Attention mechanism;

I. Introduction

Plant phenotyping is a set of physical, physiological and biochemical traits formed by the dynamics of genotypes and environmental factors during plant growth and development, and can be used to visualize the growth of a plant [1]. Due to the slow, costly and typically destructive nature of traditional plant phenotype data acquisition, automated access to plant phenotype data in a non-destructive, high-throughput and high-precision way has become a bottleneck in plant phenotyping. The major plant phenotypic traits are usually involved with plant organs, including roots, leaves, stems and fruits in most plants. In order to obtain quantitative phenotypic data on plant organs or parts, plant organ segmentation is an important prerequisite step towards automated plant phenotypic measurements in agriculture.

Previously, computer vision, machine learning and deep learning techniques based on 2D images were important technical tools to solve the plant-part segmentation problem. However, methods based on 2D images have limitations as they cannot handle the overlap and occlusion between leaves well. In contrast to 2D images, 3D reconstruction of plants and organ-level segmentation based on 3D point cloud have gradually become a hot research issue in modern agricultural information technology in recent years. The current segmentation methods mainly include edge-based methods[2][3], region-based methods[4], model-based methods[5][6][7], attribute-based methods[8][9][10] and graph optimization-based methods[11][12]. However, the above-mentioned traditional point cloud segmentation methods still lack adaptivity in various plant species, and most of them require empirical setting of reasonable threshold parameters.

Deep learning (DL) techniques have been revitalized and perform vast agricultural applications as a standard approach to semantic segmentation for 2D images, but the use of 3D DL-based techniques for plant segmentation at organ level is still in fancy stage. Different from regular input representations that can be processed with classical convolution, a point cloud is conceptualized as an unorganized collection of vectors scattered in 3D space and it's difficult to apply general CNNs. To deal with the irregular data, one common solution is to adopt multi-view projection-based methods[13][14]. However, the performance is affected by viewpoint selection and occluded parts which result in data loss. Another straightforward way is 3D voxelization in a grid structure[15][16], but the process can incur massive computation cost with 3D convolution and memory cost due to data sparsity. Since introducing the PointNet[17], researches on point-based approaches that directly deal with unordered 3D points has exploded.

However, the current DL-based approaches on 3D point cloud are limited to a few object datasets such as the ModelNet40[18], the indoor scene dataset S3DIS[19] and the outdoor large-scale dataset KITTI[20] for outdoor objects, the application on 3D point clouds of plant-part segmentation has rarely previously emerged. The main factors hindering the application of DL technology in plant phenotyping is the lack of annotated point cloud dataset of real plant model, CNNs that must be properly built for

979-8-3503-3293-3/22/$31.00 ©2022 IEEE
DOI 10.1109/ISCTech58360.2022.00093

Authorized licensed use limited to: University College Dublin. Downloaded on August 02,2024 at 18:02:53 UTC from IEEE Xplore. Restrictions apply.

unstructured and unordered point cloud data and challenging for network design. Based on attention mechanism, we proposed a hierarchical Muti-head Two-level Attention-based Network (MTANet) to perform the organ-level plant segmentation task. We also adopted other mainstream point-based DL architectures such as PointNet, PointNet++, etc. as baseline methods.

The main contributions are as follows:

(1) A well-labeled plant dataset on 3D point cloud which is built using 3D reconstruction of Caladium bicolor plant imagery data.

(2) A point-based network MTANet which directly dealed with our fully annotated 3D point cloud dataset of Caladium bicolor was designed to obtain the semantic segmentation result.

(3) A Two-level Attention-based module (TAM) was proposed to explore the feature dependencies between points in the local and global regions based on the attention mechanism.

II. Related works

In previous studies, traditional point cloud segmentation methods mainly design feature descriptors based on spatial information, colour texture, etc. Elnashef[2] performed stem leaf segmentation of plant point clouds utilizing the first tensor and second tensor, and segmented individual leaves using the density-based spatial clustering of applications with noise (DBSCAN) algorithm. Hu[8] captured RGB images, depth data and RGB-D fusion data of poplar saplings in a large field for the complex planting environment, and for the fusion data first used the SegNet to segment the leaves and trunks of poplar saplings, and then used KD-Tree based on the set distance to segment the individual leaves of poplar saplings. Kuo[9] collected 2D images of plants from multiple views to construct 3D models, and first used the watershed algorithm to segment the leaves in the top view image, then generate seed points for each leaf, and extend the seed points to grow on the 3D image to segment individual leaves. Based on the Multi-view Stereo (MVS) method, Junjie[21] reconstructed the rape point cloud, extending the existing Euclidean distance and spectral clustering algorithms and segmenting the rape point cloud organs iteratively. However, their application is limited by tedious and laborious parameter adjustment during the segmentation process.

In recent years, advances in high-performance hardware systems and neural network architectures, 3D DL-based methods have shown great potential for point cloud segmentation. Current researches have focused on projection-based methods, voxel-based methods and point-based methods. For example, Ni[22] first used the Mask R-CNN model to segment ripe blueberries from 2D images, then projected the instance mask onto 3D space to establish 2D-3D correspondence, enabling to segment individual blueberries. Jin[16] proposed a voxel-based convolutional neural network (VCNN) for the classification and segmentation of maize point clouds. However, projection-based methods are sensitive to projection angles and occlusions, and the projection step inevitability of lossing information about the geometry of the 3D point cloud. Voxel-based methods typically voxelise the disordered original point cloud and then perform voxel-level segmentation using standard 3D convolution. The voxelisation step itself introduces discretization errors and information loss, and the high resolution of the input voxel mesh usually implies high memory resources and computational costs. Point-based methods directly manipulate irregular point cloud data and are able to perform classification and segmentation at the point level. In PointNet[17], a symmetric function applied to 3D coordinates in a permutation-invariant way is the fundamental element. PointNet++[23] was presented as a way to augment the original PointNet by capturing local features. Subsequently, a dynamic graph convolutional neural network (DGCNN)[24] was proposed to dynamically update the graph. Point-based DL techniques have been applied in botany and agriculture such as for distinguishing leaves and woody components, Wu[25] enhanced the PointNet. Gong L[26] designed the Panicle-3D network for segmentation of rice spikes and stalks using a produced rice spike point cloud dataset for training.

III. Material and method

A. Dataset

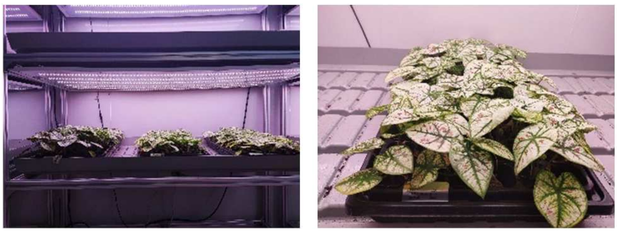

Figure 1. Part of pictures of Caladium bicolor as an example.

Caladium bicolor, its peltate arrow-shaped or heart-shaped leaves are very brightly coloured and there are a great number of variants as Fig. 1 shown. Its high value and low maintenance make it a very popular foliage plant for indoor use. In this study, Caladium bicolor was selected from a normal growth cycle, with 3-7 leaves per plant, of varying heights and uneven leaf sizes. An image capturing platform was set up to obtain highly accurate plant imagery data. It is composed of a frame, a turntable, a holder, three LED lights, three digital cameras, a light controller and a computer as a camera controller. The digital camera at each position was controlled by a python program to capture 60 images and a total of 180 images were collected, all which were set the same pixel of 2048×1536 and saved as jpg files. Fig.2 displays the system to capture 2D images of Caladium bicolor.

2D images obtained from different positions were input into RealityCapture[27] and reconstructed in 3D using the SFM/MVS algorithm. In order to ensure accuracy and improve computational efficiency, the reconstructed plant model need to be pre-processed by filtering. Considering that the point clouds of Caladium bicolor have redundant point clouds such as soil, we first used colour filtering to remove points that are not relevant to the plant part. Then we used statistical filtering to remove outliers and combines voxel filtering to downsample the point cloud. Also, the data was augmented with a series of rotation,

Authorized licensed use limited to: University College Dublin. Downloaded on August 02,2024 at 18:02:53 UTC from IEEE Xplore. Restrictions apply.

scaling and Gaussian noise addition operations to obtain a total of 3300 Caladium bicolor point clouds, which are separated in a 2:1 ratio into training and test sets. Finally, we annotated the point clouds using the segmentation tool in CloudCompare[28] for three classes: $NonPlant$, $Leaf$ and $Stem$, which is as shown in Fig. 3.

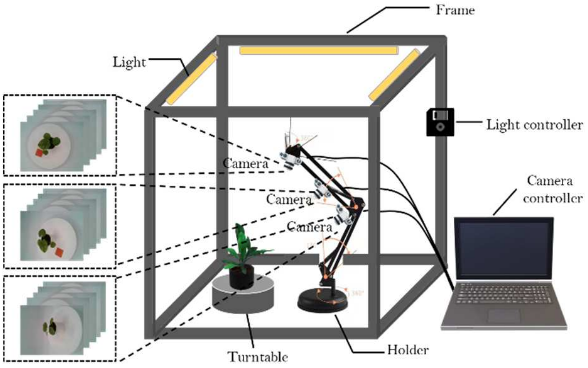

Figure 2. The image capturing platform.

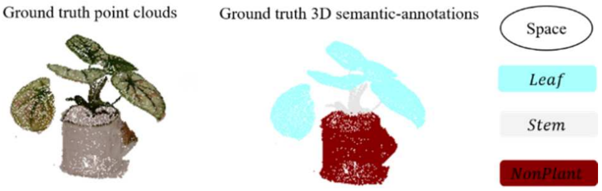
Ground truth point clouds
Ground truth 3D semantic-annotations
Space
Leaf
Stem
NonPlant

Figure 3. Example of the point cloud with ground-truth and semantic annotations.

B. Model

A U-net design was adopted for semantic segmentation following a widely used encoder - decoder architecture with skip connections. The proposed MTANet is mainly composed of three modules: Multi-head Two-level Attention Module (MTAM), Down-Sampling module (DS) and Up-Sampling module (US). The detailed architecture of our MTANet used for point-wise semantic segmentation is shown in Fig. 4. The input cloud was first fed to a shared multilayer perceptrons (MLP) layer, then we employed four encoding layers to reduce the size of points while increasing the per-point feature dimensions. A MTAM and a DS module were included in each encoding layer. The point cloud was downsampled with four-fold rates and retaining only 25% points after each layer, resulting in gradually reducing the carnality of the point set, i.e., (N → N/4 → N/16 → N/64 → N/256). Simultaneously, the feature dimension of each layer continuously rised to get more information after each layer, i.e., (32→64→128→256→512), where N represents the number of points.

Following the encoder, four decoders were utilized to restore the number of the points to N, using an US module and a MLP layer for each layer of the decoder. By skip connection, the upsampled feature maps of low-level features generated from the encoder stage were fused with the intermediate feature map of high-level features generated from the decoder stage. Final semantic predictions were obtained through three shared Fully Connected (FC) layers, i.e., (N, 128) →(N, 32) →(N, C). To avoid overfiting, we applyed a dropout layer after the FC layer. The network produced a matrix with a size of $N \times C$, where C represents the number of categories.

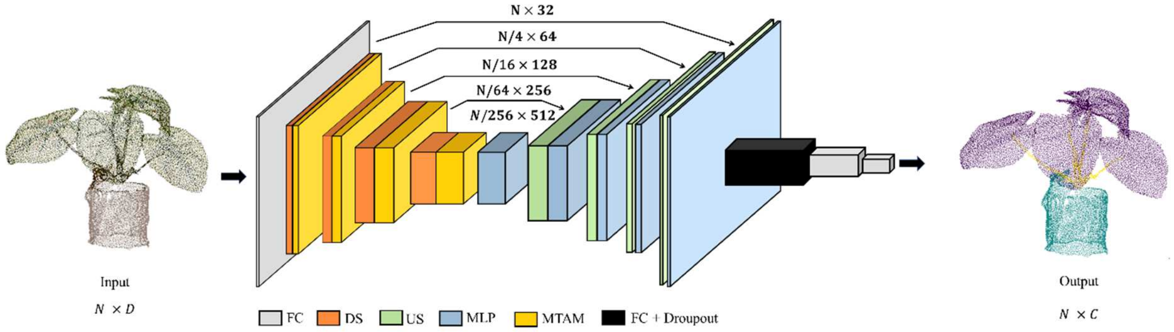

Figure 4. The architecture of MTANet. It includes four encoders and four decoders.

We proposed the MTAM which is shown in Fig. 5, each head of TAM contains two sequential sub-modules: Local Attention Module and Non-Local Attention Module. The Local Attention Module focuses on the feature interdependencies of point hierarchically organized in local regions. Firstly, we find local regions definited by Knearest neighbour (KNN) and introduced a position embedding modul(PEM) which explicitly encode the spatial position as formula (1):

$$pos_{ik} = x_i \oplus x_{ik} \oplus x_i - x_{ik} \oplus ||x_i - x_{ik}|| \qquad (1)$$

where x_i and x_{ik} represent the xyz coordinates of the centroid point and k neighboring points, $\oplus$ is the feature concatenation, ||.|| computes the Euclidean distance. Then concatenated with the features of its k nearest points as formula (2):

$$f_{ik} = pos_{ik} \oplus (p_{ik} - p_i) \qquad (2)$$

Where p_{ik} represents the features of k neighboring points, p_i represents the features of the centroid.

In the Local Attention Module, we used the powerful attention mechanism instead of max/mean pooling to automatically aggregate the features. In detail, we employed a shared MLP layer with a ReLU nonlinear layer, followed by Softmax, where W is a learnable parameter, to learn a attention score a_{ik} and weighted summed these features as formula (3) and formula (4):

$$a_{ik} = MLP(f_{ik}, W) \qquad (3)$$

$$f_i = \sum_{k=1}^{K} (a_{ik} \cdot f_{ik}) \qquad (4)$$

After aggregating the local features, we designed a Lon-Local Attention Module to update global features based on self-attention and use matrix dot-product to calculate the

Authorized licensed use limited to: University College Dublin. Downloaded on August 02,2024 at 18:02:53 UTC from IEEE Xplore. Restrictions apply.

attention scores for all points. Let Q, K, V represent query, key, and value, respectively, generated by a shared MLP layer of input features as formula (5), W_q, W_k, W_v are the learnable weights. To begin, we apply the matrix dot-product to determine the weights a_{ij} using the Q, K matrices, then apply the softmax operator to normalized the first dimension of attention map as formula (6). To enhance normalization, we apply $l1-norm$ to normalize the second dimension as formula (7):

$$Q, K, V = F_l \cdot (W_q, W_k, W_v) \tag{5}$$

$$a_{ij} = softmax(Q \cdot K^T) \tag{6}$$

$$a'_{ij} = \frac{a_{ij}}{\sum_k a_{ik}} \tag{7}$$

the weighted sums of the value vector utilizing the normalized attention weights a'_{ij} is denoted as f'_i:

$$f'_i = a'_{ij} \cdot V \tag{8}$$

Then we concatenated loacl and non-local features as formula (9):

$$F_i = f_i \oplus f'_i \tag{9}$$

In addition, we introduced multi-head mechanism to obtain more comprehensive information and further enhance the generalization ability of the network which is calculated as formula (10). Where m is the feature of the m^{th} head of point p_i and M is the number of attention heads, we set M to 4.

$$F'_i = F_i^1 \oplus F_i^m \oplus \cdots \oplus F_i^M \tag{10}$$

In summary, for the i^{th} point p_i, the MTAM learn to aggregate the features of its K nearest points and all points, finally generate a discriminative feature vector F_i'.

For each DS module, iterative farthest point sampling (FPS) algorithm is performed on the input point set P0 to acquire the subsampled point set P1. In order to map the features from original point set onto the downsampled subset, we adopt KNN to constitute local regions. And the features in a local region goes through a shared MLP, followed by normalization and ReLU layer, then max pooling is performed on each point in P1 from its neighbors in P0.

For each US module in the decoder, the primary function is to restore the point numbers and map the features from P1 onto its superset P0. Similar to the deconvolution in CNN, the point cloud is upsampled, and the feature is transferred from shape level to point level. Firstly, we employed the hierarchical propagation strategy with distance-nearest-neighbor interpolation to propagate the point features from downsampled points to original points. Then these interpolated features from the preceding decoding layer were concatenated with the features generated by the relevant encoding layer to obtain concatenated feature maps through skip connections, after which were processed by a shared MLP layer followed by batch normalization and ReLU.

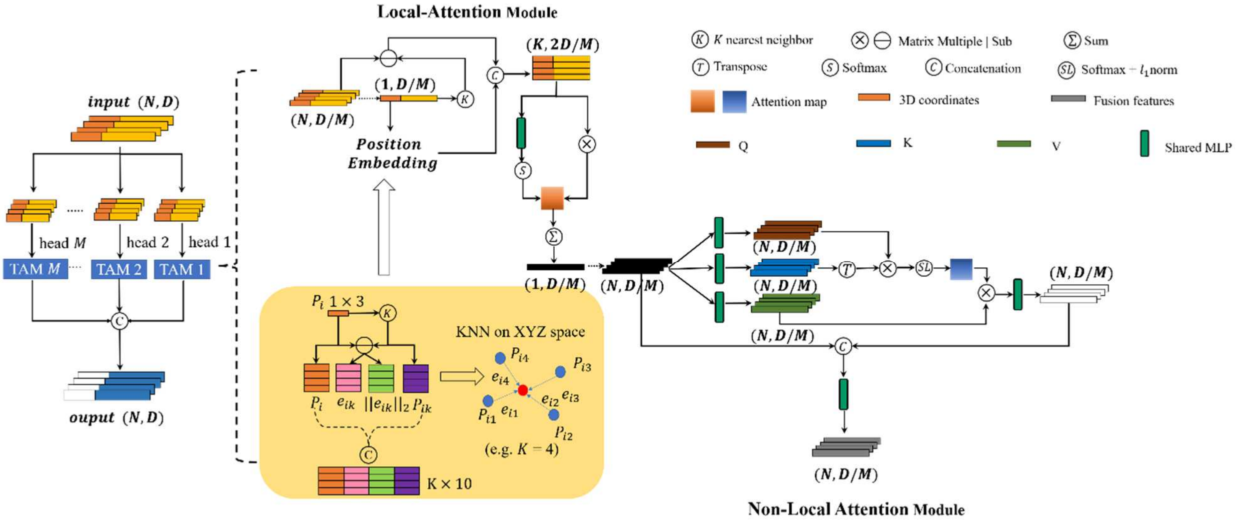

Figure 5. Architecture of Multi-head Two-level Attention Module.

IV. Experiment and Result

A. Training

Due to feed the entire Caladium bicolor plant model to the point-based DL architectures entails a high subsampling rate, we followed the strategy in PointNet which divided the input Caladium bicolor model into overlapping fixed-size blocks. Our proposed network receives a sub-point cloud of 1024 points for each block. With momentum of 0.9 and weight decay of 0.0005, the network was trained using the encapsulated Stochastic Gradient Descent (SGD) optimizer. The initial learning rate was set to 1e-3, and after 20 epochs, the learning decay rate was set to 0.5. The batch size was all set to 16, and the network was trained using a total of 100 epochs.

The code was implemented in Pytorch 1.6 with AMD EPYC 7302 CPU @ 3.00 GHz processors which contains 16 processing cores, 24 GB RAM and used two NVIDIA GeForce RTX 3090 GPU devices on a 64-bit server of Linux CentOS 8. We used the cross-entropy loss function as usual, the loss vs.epochs curve is shown in Fig. 6.

Authorized licensed use limited to: University College Dublin. Downloaded on August 02,2024 at 18:02:53 UTC from IEEE Xplore. Restrictions apply.

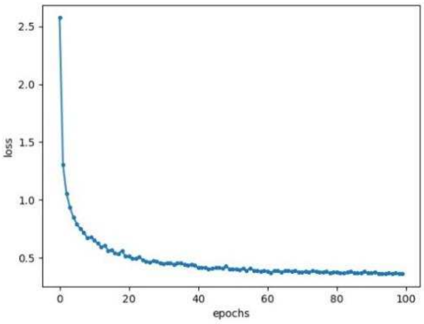

Figure 6. Loss vs. epochs curve.

B. Evaluation Metrics

We used the commonly Overall Accuracy ($OAcc$), mean Accuracy ($mAcc$), and mean of the Intersection over Union ($mIoU$) over all classes as evaluation metrics. The formulas for the $OAcc$, $mAcc$ and $mIoU$ are given as follows. where c is the category of structural parts of Caladium bicolor plant, here we set $c = 3$, p_{ij} is a point that is expected to belong to class i but is predicted to class j, p_{ji} is a point that is expected to belong to class j but is predicted to class i, both of which are misclassified, and p_{ii} is a point that is correctly classified.

$$OAcc = \frac{\sum_{i=0}^{c} p_{ii}}{\sum_{i=0}^{c}\sum_{j=0}^{c} p_{ij}} \quad (11)$$

$$mAcc = \frac{1}{c+1}\sum_{i=0}^{c}\frac{p_{ii}}{\sum_{j=0}^{c} p_{ij}} \quad (12)$$

$$mIoU = \frac{1}{c+1}\sum_{i=0}^{c}\frac{p_{ii}}{\sum_{j=0}^{c} p_{ij} + \sum_{j=0}^{c} p_{ji} - p_{ii}} \quad (13)$$

C. Result

We comprehensively conducted quantitative and qualitative assessment of our proposed MTANet and compared it with other mainstream DL-based architectures: (1) PointNet[17]; (2) PointNet++[23]; (3) DGCNN[24]; (4) ShellNet[28]; (5) PointWeb[29]. The quantitative semantic segmentation performance based on the evaluation metrics ($OAcc$, $mAcc$ and $mIoU$) on the testing set comparing with these DL-based networks is shown in Table 1. Our proposed MTANet achieves the best segmentation performance in terms of 99.16% $OAcc$, 95.73% $mAcc$ and 93.64% $mIoU$ which outperforms other existing models in terms of local information perception and segmentation precision.

In addition, Table 1 shows that among all DL-based models, the accuracy of $Stem$ is lower than the other two classes. There are two possible reasons: (1) the $Stem$ is more difficult to be segmented by the network than $Leaf$; (2) the number of points for $Stem$ is fewer than the $Leaf$.

Fig.7 shows the visualisation results obtained by MTANet for a selection of Caladium bicolor examples in different plant growth cycles and spatial structures. It can be seen that the junction between the leaf and stem can be almost exactly segmented, and our method can sensitively detect the newly grown leaflets and the dying leaves to correctly segment them. As it shown, MTANet successfully detected the newly grown leaflets at the bifurcation of the stem, including dying leaves mixed with the pot in Caladium bicolor point cloud.

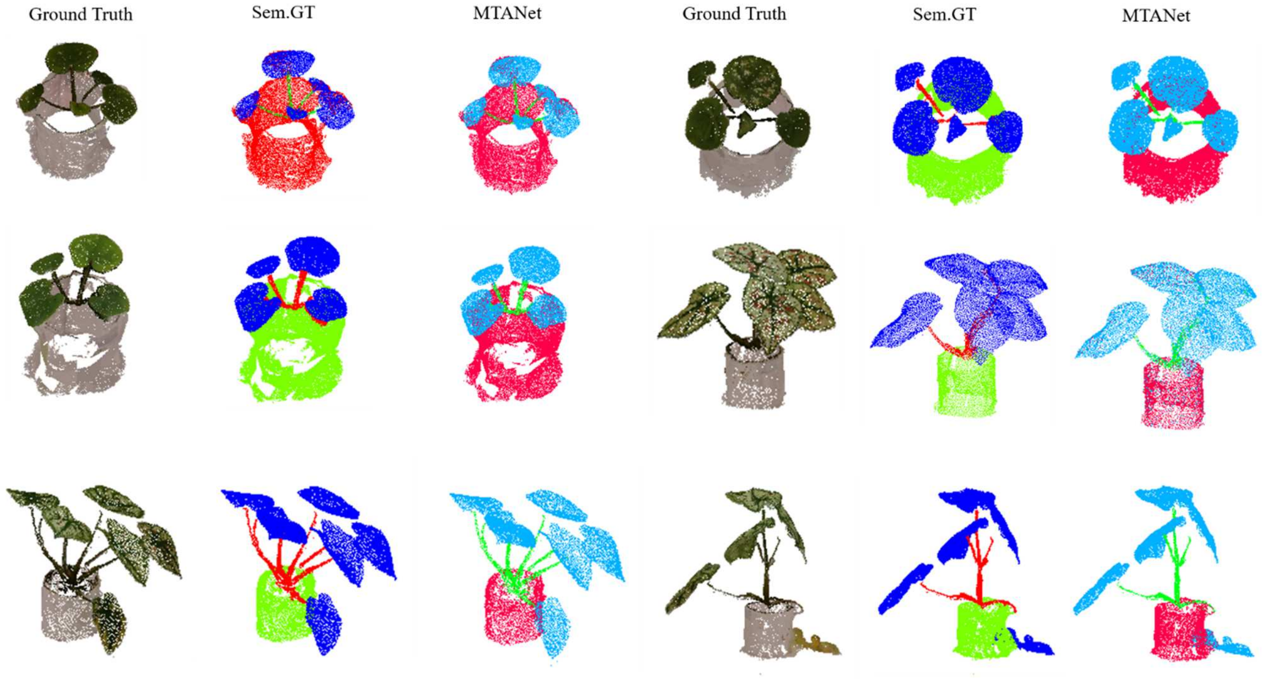

Figure 7. Example of visualization results for plant-part segmentation of Caladium bicolor by MTANet, where Ground Truth is the original Caladium bicolor, Sem.GT is the Caladium bicolor with semantic labels, and MTANet is the semantic segmentation results predicted by the network.

Authorized licensed use limited to: University College Dublin. Downloaded on August 02,2024 at 18:02:53 UTC from IEEE Xplore. Restrictions apply.

TABLE I. QUANTITATIVE RESULTS OF THE PLANT-PART SEGMENTATION PERFORMANCE OF THE DL-BASED NETWORK.

Methods	Acc			IoU			$OAcc$	$mAcc$	$mIoU$
	$Leaf$	$Stem$	$NonPlant$	$Leaf$	$Stem$	$NonPlant$			
PointNet	91.71	72.49	98.12	88.76	67.75	86.51	91.38	87.44	81.01
PointNet++	98.45	84.44	99.94	95.74	80.40	99.57	96.76	94.28	91.90
DGCNN	97.80	**85.09**	97.91	94.03	77.79	95.06	95.43	93.60	88.96
ShellNet	98.13	79.96	99.68	**97.68**	66.33	98.21	98.19	92.59	87.41
PointWeb	98.48	82.43	99.74	94.61	79.77	99.43	93.52	93.55	91.27
MTANet(ours)	**99.70**	84.36	**99.95**	96.55	**81.35**	**99.83**	**98.73**	**94.67**	**92.58**

V. DISCUSSION

A. Ablated Modules of NetWork.

we did ablation experiments to study the effectiveness of modules of MTANet, including the PEM、the TAM including the Local and the Non-Local Attention Module. The ablation experimental results for plant-part segmentation are shown in the TABLE 2, where each row indicates the removal of an existing module from MTANet, we compared the three versions of MTANet named A1-A3 with the full MTANet-A4.

By analysing the experimental results, we can see that the Non-Local Attention Module has the greatest impact on segmentation performance among them. According to our analysis, this is mainly due to the inability of the network to capture important global features, resulting in approximate 5.0%、5.6%、8.0% reduction in the $OAcc$, $mAcc$ and $mIoU$ of the network. The Local Attention Module, which may capture the feature dependencies of locations in the local area, is the next more crucial component of MTANet. When the PEM, which encodes the location of 3D points in space, was removed, the network was unable to learn information about the local geometry of the plant, the average drop in $mIoU$ reach about 2. 5%.

TABLE II. EXPERIMENTAL ANALYSIS OF ABLATION OF MTANET FOR SEMANTIC SEGMENTATION OF POINT CLOUDS OF CALADIUM BICOLOR.

	PEM	TAM		$OAcc$	$mAcc$	$mIoU$
		$Local$	$Non\text{-}Loacl$			
$A1$		√	√	97.81	91.77	90.06
$A2$	√		√	97.17	91.26	87.63
$A3$	√	√		93.71	89.09	84.64
$A4$	√	√	√	**98.73**	**94.67**	**92.58**

B. Neighbourhood k setting.

Based on the full MTANet, i.e. version A4, we investigated the performance of the neighbourhood k setting used to determine the local neighbourhood size around each local region of the network. The plant-part segmentation results obtained by setting different k are shown in TABLE 3, with the best segmentation performance obtained when k is set to 16.

When the k is set relatively small, the model may not capture enough contextual information for prediction. When the k is set too large, it may contain points that are far from the centre and have low relevance, inevitably introducing too much noise into the process, leading to high computational costs and reducing the accuracy of the model.

TABLE III. THE PERFORMANCE OF THE K-NEIGHBOURHOOD SETTING ON MTANET. THE K WAS SET TO 4, 8, 16, 24 AND 32.

		$k = 4$	$k = 8$	$k = 16$	$k = 24$	$k = 32$
Acc	$Leaf$	93.04	95.43	**99.70**	99.06	97.40
	$Stem$	79.01	80.72	**84.36**	82.06	81.45
	$NonPlant$	98.15	**99.97**	99.95	99.37	98.92
IoU	$Leaf$	91.16	92.60	**96.55**	95.91	91.19
	$Stem$	71.74	72.59	**81.35**	80.18	72.92
	$NonPlant$	95.05	98.82	**99.83**	96.79	98.97
$OAcc$		92.76	94.20	**98.73**	96.80	92.64
$mAcc$		90.06	92.04	**94.67**	93.50	92.59
$mIoU$		85.98	88.00	**92.58**	90.96	87.69

VI. CONCLUSION

To address the challenging problems of plant organ-level segmentation in modern agricultural application, we first build an image capturing platform to obtain multi-view 2D images of Caladium bicolor cultivated under different growth environments, and obtains 3D point clouds through 3D reconstruction. The point clouds then were pre-processed, and through manual annotation and data augmentation, a dataset of 3300 Caladium bicolor point clouds was finally obtained. Compared with other mainstream DL-based architectures, MTANet achieved

Authorized licensed use limited to: University College Dublin. Downloaded on August 02,2024 at 18:02:53 UTC from IEEE Xplore. Restrictions apply.

98.73% $OAcc$, 94.67% $mAcc$ and 92.58% $mIoU$, achieving the best segmentation performance on our fully annotated dataset. With the addition of the MTAM, our MTANet can focus on important local and global features, tap into the deep and fine-grained information. It can solve the under-segmentation problem on the partial segmentation and boundary segmentation where the plant biomass is weak, and obtain more accurate segmentation results.

However, the MTANet's performance is inversely proportional to the size of the dataset, and in practice, labeling the point clouds takes a lot of time and is prone to mistakes. In addition, when new leaves are wrapped by stems, the categories are often not correctly labelled and blurred edge labelling of individual plant parts also contributes to segmentation errors. In the future we will focus on introducing more plant species into the dataset and using simulator synthetic data methods to help increase plant 3D point cloud data for network training.

References

[1] Pan Y H. Analysis of concepts and categories of plant phenome and phenomics[J]. Acta agronomica sinica, 2015, 41(2): 175-186.

[2] Elnashef B, Filin S, Lati R N. Tensor-based classification and segmentation of three-dimensional point clouds for organ-level plant phenotyping and growth analysis[J]. Computers and electronics in agriculture, 2019, 156: 51-61.

[3] Boulch A, Guerry J, Le Saux B, et al. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks[J]. Computers & Graphics, 2018, 71: 189-198.

[4] Paulus S, Dupuis J, Riedel S, et al. Automated analysis of barley organs using 3D laser scanning: An approach for high throughput phenotyping[J]. Sensors, 2014, 14(7): 12670-12686.

[5] Lin Chengd, Han Jing, Xie Liangyi,Hu Fangzheng. Cylinder space segmentation method for field crop population using 3D point cloud[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE),2021,37(7):175-182.

[6] Gélard W, Devy M, Herbulot A, et al. Model-based segmentation of 3D point clouds for phenotyping sunflower plants[C]//12. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. 2017.

[7] Adam A, Chatzilari E, Nikolopoulos S, et al. H-RANSAC: A hybrid point cloud segmentation combining 2D and 3D data[J]. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci, 2018, 4(2): 1-8.

[8] Hu C, Pan Z, Zhong T. Leaf and wood separation of poplar seedlings combining locally convex connected patches and K-means++ clustering from terrestrial laser scanning data[J]. Journal of Applied Remote Sensing, 2020, 14(1): 018502.

[9] Kuo K, Itakura K, Hosoi F. Leaf segmentation based on k-means algorithm to obtain leaf angle distribution using terrestrial LiDAR[J]. Remote Sensing, 2019, 11(21): 2536.

[10] Ferrara R, Virdis S G P, Ventura A, et al. An automated approach for wood-leaf separation from terrestrial LIDAR point clouds using the density based clustering algorithm DBSCAN[J]. Agricultural and forest meteorology, 2018, 262: 434-444.

[11] Hétroy-Wheeler F, Casella E, Boltcheva D. Segmentation of tree seedling point clouds into elementary units[J]. International Journal of Remote Sensing, 2016, 37(13): 2881-2907.

[12] Santos T T, Koenigkan L V, Barbedo J G A, et al. 3D plant modeling: localization, mapping and segmentation for plant phenotyping using a single hand-held camera[C]//European Conference on Computer Vision. Springer, Cham, 2014: 247-263.

[13] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, “Multi-view convolutional neural networks for 3d shape recognition,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 945–953.

[14] W. Shi, R. van de Zedde, H. Jiang, and G. Kootstra, “Plant-part segmentation using deep learning and multi-view vision,” Biosystems Engineering, vol. 187, pp. 81–95, 2019.

[15] D. Maturana and S. Scherer, “Voxnet: A 3d convolutional neural network for real-time object recognition,” in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2015, pp. 922–928.

[16] Jin S, Su Y, Gao S, et al. Separating the structural components of maize for field phenotyping using terrestrial LiDAR data and deep convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(4): 2644-2658.

[17] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 652-660.

[18] Z. Wu et al., “3d shapenets: A deep representation for volumetric shapes,” in Proceedings of 579 the IEEE conference on computer vision and pattern recognition, 2015, pp. 1912–1920.

[19] I. Armeni et al., “3d semantic parsing of large-scale indoor spaces,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1534–1543.

[20] J. Behley et al., “Semantickitti: A dataset for semantic scene understanding of lidar sequences,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9297–9307.

[21] Junjie L Y L, John D. Point cloud based iterative segmentation technique for 3D plant phenotyping[C]//International Conference on Information and Automation, 2018:1072-1077.

[22] Ni X, Li C, Jiang H, et al. Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171: 297-309.

[23] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.

[24] Wang Y, Sun Y, Liu Z, et al. Dynamic graph cnn for learning on point clouds[J]. Acm Transactions On Graphics (tog), 2019, 38(5): 1-12.

[25] Wu B, Zheng G, Chen Y. An improved convolution neural network-based model for classifying foliage and woody components from terrestrial laser scanning data[J]. Remote Sensing, 2020, 12(6): 1010.

[26] Gong L, Du X, Zhu K, et al. Panicle-3D: Efficient Phenotyping Tool for Precise Semantic Segmentation of Rice Panicle Point Cloud[J]. Plant Phenomics, 2021, 2021.

[27] Girardeau-Montaut D. CloudCompare[J]. France: EDF R&D Telecom ParisTech, 2016, 11.

[28] Zhang Z, Hua B S, Yeung S K. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1607-1616.

[29] Zhao H, Jiang L, Fu C W, et al. Pointweb: Enhancing local neighborhood features for point cloud processing[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 5565-5573.

Authorized licensed use limited to: University College Dublin. Downloaded on August 02,2024 at 18:02:53 UTC from IEEE Xplore. Restrictions apply.