Semantic Segmentation Network of Noisy Plant Point Cloud based on Self-Attention Feature Fusion

1st Yang Chen
School of Computer Science
Guangdong University of Technology
Guangzhou, China
2112005180@mail2.gdut.edu.cn

2nd An Zeng*
School of Computer Science
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

4th Yuzhu Ji
School of Computer Science
Guangdong University of Technology
Guangzhou, China
yuzhu.ji@gdut.edu.cn

2022 10th International Conference on Information Systems and Computing Technology (ISCTech) | 979-8-3503-3293-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISCTECH58360.2022.00066

***Abstract*—When 3D reconstruction of plant seedlings is performed to obtain plant point clouds, there are many noisy points between the leaves and stems of plant point clouds due to the influence of ambient light and the limitation of the camera vision, which affects the automatic measurement of plant phenotypes.In order to achieve plant point cloud stem and leaf segmentation with a large amount of noise, We propose a network for semantic segmentation of noisy plant point clouds based on self-attentive feature fusion (abbreviated as SAFF-Net).The network first extracts shallow spatial features and higher-level semantic features between points by using correlations between pairs of points in the neighborhood through the local feature fusion module, and enhances the ability to extract plant shape features;The dual-branch attention pooling module is used to aggregate features, and one branch uses the channel attention mechanism to adaptively filter low-correlation features, avoiding redundancy of features and mitigating the bias effect of noise points.Another branch performs max pooling for the highest level feature map to obtain global contextual features, and finally merges local and global contextual features to learn more discriminative features.The point cloud dataset for the experiment is derived from multi-angle photographs of the plant taken by a high-definition camera and reconstructed in 3D by Structure from Motion(SFM) algorithm.The experimental results indicate that SAFF-Net performs better than the mainstream semantic segmentation networks and extracts more fine-grained local features. OA and (mIoU) of this method on the plant dataset are 93.7% and 83.4%, respectively. It outperforms existing methods in the segmentation of noise points and leaves, and the segmentation results have higher precision and recall.**

Keywords-Domain-Specific AI Applications ; point cloud semantic segmentation; plant point cloud; self-attention mechanism;

I. INTRODUCTION

Recently, the total global population has exceeded 8 billion, global warming has led to frequent mountain fires around the world, extreme and severe weather is becoming more frequent, natural resources such as fresh water and land are at risk of loss and scarcity, and with the impact of COVID-19 and war, human demand for food is increasing[1] .In order to address the growing demand for food, breeders need to use efficient breeding programs to produce new varieties of plants with high quality and high yields.While recent advances in plant gene sequence research have led to the rapid development of new plant varieties, plant phenotypic analysis, i.e., the quantitative evaluation of plant morphological traits, is a time-consuming and labor-intensive process, thus limiting the periodicity of plant breeding in modern agriculture[2]. Therefore, the automated measurement and monitoring technology of plant phenotypic parameters has attracted more and more attention in the field of modern smart agriculture.

Traditional measurements of phenotypic parameters rely on manual or 2D images for measurement[3].Due to the complex spatial morphology of the plant, manual measurement of parameters such as plant height, leaf area and number of branches is not only highly subjective, but also requires contact measurement, which is likely to cause certain damage to the plant surface and affect the final 3D measurement results.In comparison, 2D image-based measurements are often not commonly used to measure phenotypic parameters such as leaf area and minimum bounding box volume because of dimensional constraints and shading between leaves.In order to obtain the phenotypic parameters of the plant, the plant needs to be first divided into individual plant organs..Existing plant point cloud segmentation algorithms rely heavily on high quality clean point clouds, complex calibration procedures and manually designed thresholds for segmentation using Euclidean distance clustering algorithms or region growing algorithms.These methods are very sensitive to noise, therefore, they cannot be applied to segmentation of plant point clouds with complex backgrounds and large amounts of noise.

Due to the influence of ambient light and the limitation of camera vision, there are many noisy points between the stems and leaves of plant point clouds obtained by 3D reconstruction, which are continuously distributed with each other, making the traditional stem and leaf segmentation method based on Euclidean distance clustering unfeasible.Therefore, it is important to design noise-resistant, high-throughput and high-precision plant semantic segmentation models and improve the generalization of the models for the study of automated measurement of plant phenotypic parameters[4]. We propose to use supervised deep learning algorithms to directly process disordered 3D point clouds, to predict the semantic label of each point, and to automatically and

979-8-3503-3293-3/22/$31.00 ©2022 IEEE
DOI 10.1109/ISCTech58360.2022.00066

efficiently perform organ-level segmentation of plant 3D point clouds..The existing point cloud semantic segmentation methods[9][10][11][12][13][14][15][16][17][18] based on deep learning are good for segmenting point cloud data with regular shape, uniform density and clear boundaries (such as airplanes, windows, furniture, etc.), but the plant point cloud has complex shape and uneven spatial density. There are many noise points between stems and leaves of the plant point cloud acquired by 3D reconstruction, and there is no obvious demarcation line between plant organs and noise points.Therefore, the segmentation effect of these segmentation methods still need to be improved.

In order to achieve plant point cloud stem and leaf segmentation with a large amount of noise, We propose a network for semantic segmentation of noisy plant point clouds based on self-attentive feature fusion (abbreviated as SAFF-Net).A comparison experiment between this network and the classical point cloud semantic segmentation network is performed on the plant dataset collected by this paper to verify the effectiveness, hoping to provide an effective method for automated plant stem and leaf segmentation with noise for plant phenotype measurement.Our main contributions are as follows:

- A plant multi-view image acquisition platform is constructed, which can realize fully automatic and efficient image acquisition work.
- We propose the local feature fusion module, which can extract shallow spatial features and higher level semantic features between neighboring points.
- We propose the dual-branch attention pooling module, which merges local and global contextual features to learn more discriminative features.
- We propose a network for semantic segmentation of noisy plant point clouds based on self-attentive feature fusion ,which can achieve plant point cloud stem and leaf segmentation with a large amount of noise.

II. Related Work

In previous work, researchers have mostly used traditional Euclidean distance-based clustering algorithms or region growth algorithms to segment plant stems and leaves.Specifically, [5] reconstructs a 3D point cloud of plants from images by means of the Multiple View Stereo (MVS) method, segments the leaves and extracts a series of phenotypic parameters via region growth algorithms.In [6], a 3D model of the plant was constructed from multi-scene images taken at different locations, and the distance transform and watershed algorithm were used to segment the leaves in the images.[7] utilize a spectral clustering algorithm based on Euclidean distance to segment the plant point cloud into individual plant organs (including leaves, stems, branches, etc.) by an iterative method.However, traditional algorithms rely on high quality clean point clouds, which are captured by expensive equipment such as 3D laser scanners or depth cameras.This paper uses an inexpensive high-definition camera to take photos from different angles and reconstructs the plant point cloud in three dimensions by the Structure From Motion (SFM) algorithm. This method obtains plant point clouds with many noise points between leaves and stems, and the traditional segmentation method with manually designed thresholds can no longer meet the plant point cloud segmentation in this complex scene.

Deep learning techniques have a widespread application in agriculture due to their excellent data representation learning capability and powerful generality[8].Depending on the processing methods, semantic segmentation methods for 3D point clouds based on deep learning can be classified as projection-based, discretization-based, and direct point-based[13][14][15][16][17][18] methods.The projection-based method[9][10] projects 3D point clouds to multiple 2D views and utilizes advanced graphical segmentation algorithms to segment and then converge the scores of different views to draw conclusions.However, the method is sensitive to view selection and occlusion, and the projection step is prone to information loss.The discretization-based method[11][12] converts the disordered point cloud into a normalized discrete representation for voxel segmentation using 3D convolution.This method preserves the neighborhood structure of the 3D point cloud very well, but its complex data structure leads to a large computational cost in memory. The point cloud-based method directly processes the original point cloud datat, fully preserving the spatial location and multi-dimensional feature information of point clouds.Among them, PointNet[13] obtains information of features for the whole point cloud using SharedMLP, and solves the point cloud disorder problem by symmetric function.However, because feature extraction is performed independently for each point, PointNet ignores the correlation between point pairs, resulting in the network with no knowledge of local information..PointNet++[14] improves on the base of PointNet by adding stratified sampling, which has some improvement for the local feature extraction ability of point clouds.DGCNN[15] proposes EdgeConv, which is able to extract local features well while maintaining the invariance of point cloud arrangement.The AFA module of PointWeb[16] associates each point in the local neighborhood and adaptively learns the influence weights between point pairs to adjust each point feature.RandLA-Net[17] utilizes random sampling to enhance the computational efficiency, and increases the perceptual field of each point by stacking Shared Multilayer Perceptron (SharedMLP) and attention pooling modules.

III. Methodology

In this section, we first propose the SAFF module for plant point cloud segmentation with noise, which consists of two blocks, LFF and DBAP. Then, we propose the SAFF-network, which uses PointNet++[14] as the backbone

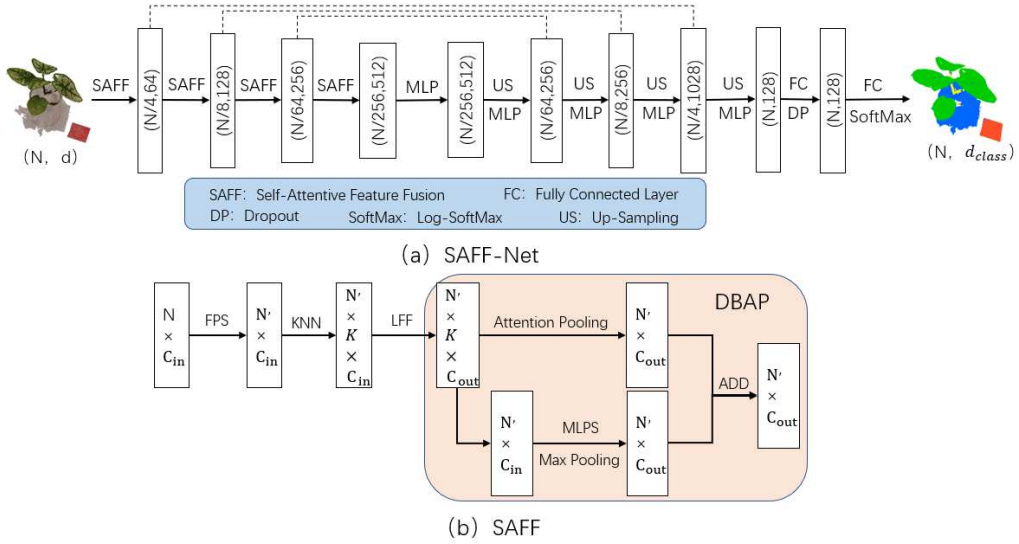


Figure 1. Architecture of the SAFF-Net and the SAFF module.

network and mainly adopts the structure of encoding before decoding with SAFF module.

A. *Architecture of SAFF-Net*

From Figure 1. (a), the input of the SAFF-net network is a point cloud with a scale of N×d, where N denotes the number of points and d denotes the feature dimension of the input points. The four encoders encode the features through SAFF module in multiple levels, with a gradual decrease in the number of points from N to N/256 and an increase in the feature dimension from d to 512, and then, the features are decoded layer by layer using the four decoders. Each decoder upsamples point features by the nearest neighbor interpolation algorithm and splices them with intermediate features of the coding layer in the corresponding dimension using skip connections.Finally the prediction semantics of all points is obtained using two consecutive fully connected layers, a dropout layer and a Log-SoftMax layer, which has a size of N × d_{class} .Among them, the dropout ratio is 0.5, and class is the number of semantic tag classes.

B. *Architecture of SAFF module*

From Figure 1. (b), it can be seen that the structure of the SAFF module, assuming that the input point cloud has N points and the feature dimension is C_{in} .First, N' centroids are selected using the farthest point sampling (FPS) algorithm, and then, for each selected centroid, the K nearest neighbors of Euclidean distance from the centroid are calculated using the K nearest neighbors (KNN) algorithm.The local neighborhood map is constructed by fusing centroid and neighbor point features to achieve a hierarchical processing of features, and then local contextual features are extracted using the local feature fusion module, and the feature dimensions are mapped to C_{out} .Then the local context features are processed in two branches, one of which filters out low relevance features by attention pooling to avoid redundancy of features.Another channel acquires the high-level feature maps through several shared perceptron layers, and then obtains the global contextual features by max pooling.Finally, the two branch features are fused to get the output point cloud features of this module, and the number of points of the output point cloud is N' ,and the feature dimension is C_{out} .

1) Local Feature Fusion: Plant point clouds have complex spatial structures, and the correlation between the centroids and their neighbors needs to be fully explored to retain more fine-grained features.Inspired by Rand-La[17] and PointWeb[16], Local Spatial Encoding (LocSE)[17] adequately fuses shallow features such as spatial locations and distances of centroids and neighboring points, but does not deal with point features in the neighborhood, and only simply splices it with the output of (LocSE), so Local feature aggregation (LFA)[17] has insufficient ability to extract higher-level features such as shape and semantics between point features.In contrast, Adaptive feature adjustment (AFA)[16] Module uses the deviation between each point and other point features in the neighborhood to learn the influence weight between them, and each point feature is adapted by other point features according to the influence weight, allowing the network to learn more high-level features such as shape and semantics.Therefore, we propose the local feature fusion module that processes 3D coordinates and point features separately, merges shallow features and high-level semantic features, reduces the loss of regional semantic information, and better describes local regions.the architecture of the LFF is shown in Figure 2. .

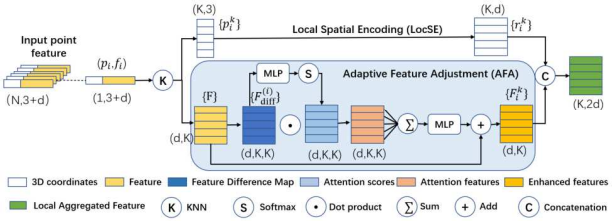


Figure 2. Architecture of the local feature fusion block.

The scale entered by this module is N × (3 + d), in which N represents the number of points, 3 represents the location coordinates in 3D space, and d represents the feature dimension. For each point, use the KNN algorithm once to find the K points with the closest Euclidean distance. Suppose the coordinates of the ith point in the input are P_i and the K neighbor coordinates are $\left\{p_i^1,\cdots,p_i^k,\cdots,p_i^K\right\}$ respectively.Some valuable information between p_i and its neighbor point coordinates (centroid coordinates, neighbor point coordinates, coordinate difference, distance between centroid and neighbor point) is concatenated by LocSE[17] and subsequently adjusted by MLP to obtain a new feature r_i^k that aggregates Euclidean distance space information.The equation is as follows:

$$r_i^K = MLP\left(p_i \oplus p_i^K \oplus \left(p_i - p_i^K\right) \oplus \left\|p_i - p_i^K\right\|\right) \quad (1)$$

Where $\oplus$ denotes the concatenation operation, and $\| \ \|$ denotes L2 norm. This Local Spatial Encoding unit makes the corresponding point features perceive the relative spatial positions between them by concatenating various coordinate information of neighboring points and the center point, which allows this encoding unit to learn shallow features such as the spatial positions and distances between the center point and the neighboring points.

For the processing of point features, AFA is different from the EdgeConv[15] that uses all neighboring points to augment information on the center point. It calculates, for each point in the neighborhood, the influence score of all other points to that point, and finally adjusts the features of each point according to the score.The input of this component is a feature vector of size d × K, which is repeatedly expanded to d × K × K dimensions first.Suppose a local neighborhood R has M points, and we denote the set of point features in R by F, like this $F=\left\{F_1,F_{2,}\cdots,F_m\right\}$, where $F \in R^C$, C represents the number of features' channels..The feature difference map is calculated according to Equation 2.

$$F_{diff}\left(F_i,F_j\right)=\begin{cases}F_i-F_j & if \ \mathrm{i}\neq \mathrm{j}\\ F_i & if \ \mathrm{i}=\mathrm{j}\end{cases} \quad (2)$$

The feature difference map contains the association information between all point pairs, using the self-attention mechanism to let it get the attention scores between all point pairs in the neighborhood through an MLP layer and a Softmax function. The attention scores are multiplied by the feature difference maps to produce the adaptively adjusted attention features..Then, the augmented features F_i^k containing information such as shape and semantics are obtained by residual concatenation and weighted summation.Finally, concatenating r_i^k and F_i^k together yields a local aggregation feature that merges shallow spatial features with higher-level semantic features. This gives each point feature the potential to represent the surrounding space, resulting in the entire network efficiently learning complex local structures.

2) Dual-Branch Attention Pooling: The pooling neural units are usually used for the aggregation of neighborhood point features.The existing works [13][14] generally use max/mean pooling for simple integration of neighborhood features, which preserves key information to a certain extent, but usually leads to local information loss.Inspired by SCF net, distance can measure the correlation between points to some extent. The smaller the distance, the higher the correlation and the richer the local information.Consequently, this paper propose the dual-branch attention pooling module.On the one hand we extract global contextual features by using max pooling to highlight and reinforce the core information of the neighborhood point features;On the other hand we use the channel attention mechanism based on geometric spatial distance, feature spatial distance and coordinate difference between centroid and neighboring points to adaptively aggregate highly correlated features, filter out redundant features and guide the network to focus on the shape structure information of the plant;Finally, the features of the two branches are combined to obtain discriminative features and improve the feature representation capability of the model, and the module structure as illustrated in Figure 3. .

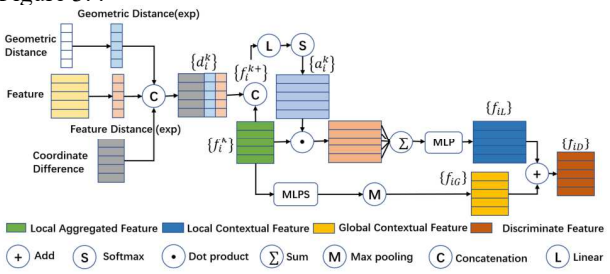


Figure 3. Architecture of the dual-branch attention pooling block.

As shown in Figure 3. , the module has four inputs, which are geometric distance, neighboring point features, coordinate difference and local aggregation features, where coordinate difference is the vector of center point coordinates minus neighboring point coordinates, and local aggregation features are the output of the local feature fusion module.To calculate the feature distance in the abstract feature space, Suppose f(i) and f(k) are any two input feature vectors in the neighborhood point feature, and we define the feature distance d_{if}^k between f(i) and f(k) as:

$$d_{if}^k = mean\left(\left|f\left(i\right)-f\left(k\right)\right|\right) \quad (3)$$

where $\left|\ \right|$ denotes the absolute value function and mean is the average function.In order to conform to the characteristics that distance and correlation are inversely proportional, we use the negative exponent of distance to represent the weight of attention pooling.In addition, since the abstract feature distance is automatically learned through the network and has instability, we introduce the parameter λ to control the weight of the feature distance, and usually λ is set to 0.1.

$$d_i^k = \exp\left(-d_{ig}^k\right) \oplus \lambda \exp\left(-d_{if}^k\right) \oplus d_{iD}^k \quad (4)$$

where d_{ig}^{k} denotes the geometric distance and d_{iD}^{k} denotes the coordinate difference between the center point and the neighboring points.

For one branch of the two-branch attention pooling module, the feature impact factor d_i^k and the locally aggregated feature f_i^k are merged via concatenation.

$$f_i^{k+} = d_i^k \oplus f_i^k \tag{5}$$

Then, f_i^{k+} is calculated by a linear layer and a Softmax function to obtain the attention weight a_i^k.

$$a_i^k = soft\max\left(Linear\left(f_i^{k+}\right)\right) \tag{6}$$

Finally, the learned attention weights are used to weight and sum the local aggregated features to derive local contextual features.

$$f_{iL} = MLP\left(\sum_{k=1}^{K}\left(a_i^k \cdot f_i^k\right)\right) \tag{7}$$

For another branch, the formula of global context feature is:

$$f_{iG} = Max\left(MLP\left(f_i^k\right)\right) \tag{8}$$

Finally, the discriminative features are obtained by adding the two parts of features, which effectively merge local contextual features and global contextual features and improve the feature characterization capability.

IV. EXPERIMENTS

In this section, we evaluate SAFF-Net on the plant point cloud dataset and compare it with existing networks. The experiments are implemented with Pytorch and run on a NVIDIA GeForce RTX 3090 GPU.

A. Data collection and point cloud annotation

In this paper, we selected healthy and uniformly growing seedlings of caladium bicolor with a canopy height of 5-10 cm and a number of leaves of 3-7.In order to acquire high-precision 2D plant sequence images, we built a low-cost platform specifically for acquiring 3D target plant data, as shown in Figure 4. , which consists of the following parts.

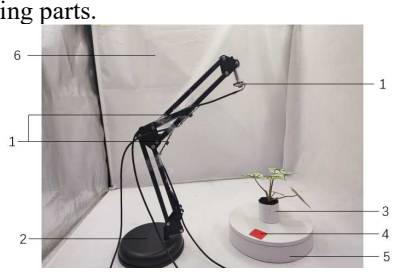
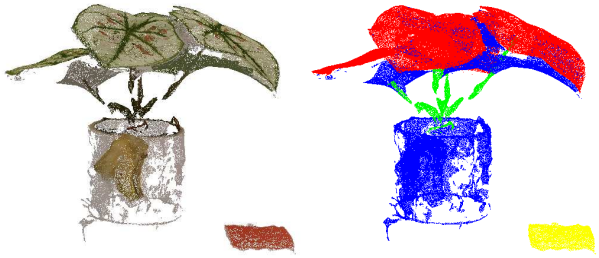


Figure 4. Data collection platform.

- 3 cameras, 8 million pixel USB driver-free autofocus camera.
- Bracket: Used to fix the position of the camera.
- Plant.
- Red paper, the size is 3 × 3cm.
- White circular turntable: the plants revolve with the turntable on a specific movement path.
- The photo studio (0.8m X 0.8m) includes a white backdrop of the same size and lamp equipment to provide light when light is insufficient.

The test plant was placed in the center of a white turntable and the plant was rotated along with the turntable, and three cameras fixed at different heights were focused on the plant.The camera is set to autofocus mode during the entire shooting process, while ensuring that the camera parameters remain the same.With plants as the center, a two-dimensional image is taken every 10-20 degrees. The three cameras obtain 60 images from different viewing angles, and a total of 180 pictures are collected for each caladium bicolor plant sample.

In this paper, the 3D reconstruction of seedling plants was performed by the Structure From Motion (SFM) algorithm with the Reality Capture software.Due to the influence of ambient light and the limitation of camera vision, there are many noise points between the reconstructed plant point cloud leaves and stems. We obtained 297 plant point cloud data by some pre-processing steps (e.g., PassThrough, statistical outlier removal, manual segmentation, annotation, etc.).Among them, 246 plants are used as the training set and 51 plants are used as the test set,and the original samples have six dimensions (XYZ and RGB).All points of each sample are classified into 4 classes of semantic labels(leaf,stem,red and cluster), as shown in Figure 5. .Among them, the red part represents the leaf, the green part represents the stem, the yellow part represents the red, and the blue part represents the cluster.



(a) Original plant cloud. (b) Labeled plant cloud.

Figure 5. Semantic segmentation annotation of plants.

B. Implementation Detail and Evaluation Metrics

In our experiments, Adam is used as the parameter optimizer during training. The number of iteration rounds of the model is 100 and the batch size is 4. The loss function uses nllloss, and the neighbor point K is set to 32 in the experiment.

The performance of semantic segmentation networks is usually measured by mean of class-wise intersection over union (mIoU), and overall point-wise accuracy (OA). In addition, we also evaluate network models using precision and recall.

C. Performance Comparison

To examine the performance of SAFF-Net network, we implement some point cloud semantic segmentation networks such as pointnet[13], pointnet++[14],

DGCNN[15], and pointWeb[16].Semantic segmentation comparison experiments were performed on the plant point cloud test dataset, and the mIoU, OA and IoU comparison results of models for each semantic class are shown in TABLE I. . The precision and recall comparison results for each semantic tag class can be found in Table 2 and Table 3, respectively.

TABLE I. SEMANTIC SEGMENTATION RESULTS ON PLANT DATASET EVALUATED ON TEST DATASET.

Models	OA	mIoU	cluster	leaf	red	stem
PointNet	91.6	80.5	84.7	85.8	**95.7**	52.0
PointNet++	91.5	79.4	84.9	85.8	84.5	52.3
DGCNN	92.6	80.4	87.0	87.0	95.3	52.1
PointWeb	90.2	76.9	84.8	85.5	87.9	49.6
Ours	**93.7**	**83.4**	**87.5**	**89.4**	88.1	**52.7**

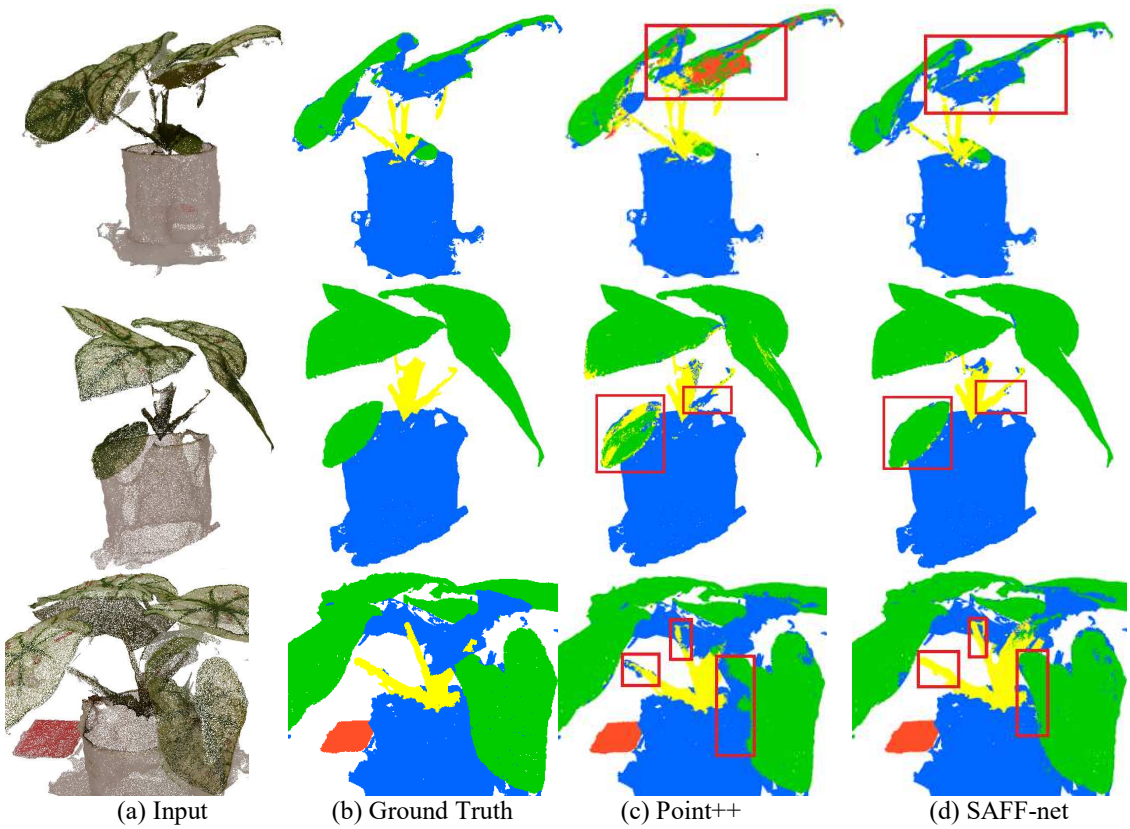
TABLE II. SEGMENTATION PRECISION ON PLANT DATASET

Models	cluster	leaf	red	stem
PointNet	95.2	86.4	95.1	49.4
PointNet++	92.4	90.4	93.9	83.3
DGCNN	**96.3**	90.2	**96.6**	72.1
PointWeb	92.0	90.9	97.5	75.1
Ours	95.3	**91.3**	95.8	**87.4**

TABLE III. SEGMENTATION RECALL ON PLANT DATASET

Models	cluster	leaf	red	stem
PointNet	83.7	97.3	**99.4**	**88.0**
PointNet++	90.7	95.2	84.3	49.6
DGCNN	89.5	95.3	98.2	72.6
PointWeb	91.2	93.5	93.6	59.4
Ours	**91.5**	**97.7**	91.6	57.0

As seen from TABLE I. , the OA and mIoU of SAFF-Net network are 93.7% and 83.4%, respectively, and the model is 2.1% higher than PointNet, 2.2% higher than PointNet++, 1.5% higher than DGCNN, and 3.5% higher than PointWeb in terms of overall accuracy.In terms of mean of intersection over union, it is 2.9% higher than PointNet, 4.0% higher than PointNet++, 3.0% higher than DGCNN, and 6.5% higher than PointWeb.The model of this paper achieves the highest IoU accuracy on cluster, leaf and stem.As shown in Table 2 and Table 3, the SAFF-Net network achieved good precision and recall in all semantic classes.In particular, the precision on the stem class is 38% higher than PointNet, 4.1% higher than PointNet++, 15.3% higher than DGCNN, and 12.3% higher than PointWeb.Since the network has high precision and recall in both cluster and leaf, it indicates that the model has excellent segmentation ability for closely connected miscellaneous points and blades.The above analysis shows that the introduction of self-attention and local feature fusion can effectively learn the contextual features, and also learn the shape structure information of the point cloud, which makes the plant segmentation more accurate.



(a) Input (b) Ground Truth (c) Point++ (d) SAFF-net

Figure 6. Visualization of semantic segmentation results on Plant dataset.

D. Visualization Comparison

To compare the performance between the models more intuitively, we selected three plant samples to visualize the segmentation results, as shown in Figure 6. .The first column shows the original input point cloud, the second column shows the real segmentation result, the third column shows the segmentation result of PointNet++ model, and the last column is the segmentation result of SAFF-Net network.In the first plant sample (the first row in the Figure 6.), PointNet++ is confusing for the segmentation of the noise points below the leaves, while the segmentation results of this paper for the leaf and the cluster are highly consistent with the real results.In the second plant sample (the second row in the Figure 6.), PointNet++ mis-segmented part of the leaf with the lowest height into stem, and the boundary segmentation between stem and cluster was not accurate, but the model in this paper almost perfectly segmented the leaf, and The prediction results for stems are also more accurate.In the third plant sample (the third row in the Figure 6.), it can be seen that PointNet++ has a poor ability to segment the local details of the red box,while the network in this paper can segment the clean stem and can clearly segment the boundary between the stray points and the leaves.Compared with PointNet++, the SAFF-Net network has a better segmentation effect on our plant dataset and is more suitable for the plant shape structure.This is because the network in this paper can extract more local context features, and further integrate point cloud features through dual-branch attention pooling to learn point cloud structure information, thereby obtaining better segmentation accuracy.

V. CONCLUSION

Aiming at the problem that there are many noise points between plant point clouds due to the influence of ambient light and the limitation of the camera vision during 3D reconstruction, this paper proposes a plant point cloud semantic segmentation network SAFF-Net based on self-attention feature fusion. The network uses the local feature fusion module to mine the correlation between all point pairs in the neighborhood, which improves the network's ability to represent the local features of the plant point cloud;The dual-channel attention pooling module is used to adaptively extract high-relevance features, avoid feature redundancy, and directs the network to focus on the shape structure information of the plants.The OA and the mIoU on the plant dataset are 93.7% and 83.4%, respectively. The segmentation of noise points and blades is better than existing methods, and the segmentation results have high precision and recall.Compared with other networks, SAFF-Net performs better in the semantic segmentation of noisy plant point clouds, which provides a feasible method and new ideas for stem and leaf segmentation of noisy plant point clouds.Since the plant data set is collected in a studio, the actual environment often has a more complex background, and the plant species are diverse and the structure is complex and disorderly.Therefore, how to improve the universality and robustness of the neural network and fully learn more fine-grained local context features is the focus of our next research.

REFERENCES

[1] Fukase, E., & Martin, W. (2017). Economic Growth, Convergence, and World Food Demand and Supply. World Bank Policy Research Working Paper Series.

[2] Ma, X., Zhu, K., Guan, H., Feng, J., Yu, S., & Liu, G. (2019). High-Throughput Phenotyping Analysis of Potted Soybean Plants Using Colorized Depth Images Based on A Proximal Platform. Remote. Sens., 11, 1085.

[3] Hayashi, E., Amagai, Y., Maruo, T., & Kozai, T. (2020). Phenotypic Analysis of Germination Time of Individual Seeds Affected by Microenvironment and Management Factors for Cohort Research in Plant Factory. Agronomy.

[4] Boogaard, F.P., van Henten, E.J., & Kootstra, G. (2022). Improved Point-Cloud Segmentation for Plant Phenotyping Through Class-Dependent Sampling of Training Data to Battle Class Imbalance. Frontiers in Plant Science, 13.

[5] Hui, F., Zhu, J., Hu, P., Meng, L., Zhu, B., Guo, Y., Li, B.G., & Ma, Y. (2018). Image-based dynamic quantification and high-accuracy 3D evaluation of canopy structure of plant populations. Annals of Botany, 121, 1079–1088.

[6] Itakura, K., & Hosoi, F. (2018). Automatic Leaf Segmentation for Estimating Leaf Area and Leaf Inclination Angle in 3D Plant Images. Sensors (Basel, Switzerland), 18.

[7] Liu, J., Liu, Y., & Doonan, J.H. (2018). Point cloud based iterative segmentation technique for 3D plant phenotyping. 2018 IEEE International Conference on Information and Automation (ICIA), 1072-1077.

[8] Schmoldt, D.L., & Symons, S.J. (2000). Computers and Electronics in Agriculture.

[9] Lawin, F.J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F.S., & Felsberg, M. (2017). Deep Projective 3D Semantic Segmentation. International Conference on Computer Analysis of Images and Patterns.

[10] Boulch, A., Saux, B.L., & Audebert, N. (2017). Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. 3DOR@Eurographics.

[11] Huang, J., & You, S. (2016). Point cloud labeling using 3D Convolutional Neural Network. 2016 23rd International Conference on Pattern Recognition (ICPR), 2670-2675.

[12] Graham, B., Engelcke, M., & Maaten, L.V. (2017). 3D Semantic Segmentation with Submanifold Sparse Convolutional Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9224-9232.

[13] Qi, C., Su, H., Mo, K., & Guibas, L.J. (2016). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 77-85.

[14] Qi, C., Yi, L., Su, H., & Guibas, L.J. (2017). PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. NIPS.

[15] Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., & Solomon, J.M. (2018). Dynamic Graph CNN for Learning on Point Clouds. ACM Transactions on Graphics (TOG), 38, 1 - 12.

[16] Zhao, H., Jiang, L., Fu, C., & Jia, J. (2019). PointWeb: Enhancing Local Neighborhood Features for Point Cloud Processing. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5560-5568.

[17] Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, A., & Markham, A. (2019). RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11105-11114.

[18] Fan, S., Dong, Q., Zhu, F., Lv, Y., Ye, P., & Wang, F. (2021). SCF-Net: Learning Spatial Contextual Features for Large-Scale Point Cloud Segmentation. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 14499-14508.