2024 IEEE 14th International Conference on Electronics Information and Emergency Communication (ICEIEC) | 979-8-3503-6189-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICEIEC61773.2024.10561642

Multi-Layer Attention Fusion and Edge-Guided Structural Enrichment Completion Network for Leaf Point Cloud

Qingqing Liao, An Zeng*, Yuzhu Ji, Yiqun Zhang
School of Computer Science and Technology
Guangdong University and Technology
Guangzhou, Guangdong Province, 510006, China
**Corresponding author, e-mail: zengan@gdut.edu.cn*

Abstract—**Three-dimensional (3D) reconstruction is an important task to capture the spatial structure and morphological features of plants. However, the occlusion between leaves resulted in an incomplete point cloud. It thus affects the accuracy of plant phenotype analysis. To obtain complete leaf point clouds, a Multi-Layer Attention Fusion and Edge-Guided Structural Enrichment Network (MAEGNet) was proposed for the Leaf Point Cloud Completion. It designed multiple layers of attention fusion modules and an edge-preserving loss function, using a Transformer to process and fuse local and global features of leaves at varying receptive fields. It encouraged the model to recover the leaf edge structure by calculating the loss between the edges extracted from the completed and ground-truth leaf point clouds. Experimental results show that our model can achieve a better result on our constructed leaf point cloud dataset. The Earth Mover's Distance (EMD) is reduced by 7.44 cm.**

Keywords-point cloud completion; deep learning; attention mechanism; plant leaves

I. INTRODUCTION

Phenotypic traits in plants reflect the combined influence of genetics and environment [1], and their measurement through plant phenotyping reveals crucial ecological processes[2]. Leaves are central to this, yet traditional assessments rely on manual methods, restricting precision and scalability[3]. 3D laser scanning technology provides point cloud data, advancing plant research, though data may be incomplete, noisy, or obstructed.

Deep learning-based 3D point cloud completion methods address this by automatically learning to fill gaps, improving leaf structure analysis. Yuan et al.[4] introduce PCN, a two-step system that converts sparse to dense, detailed point clouds. Tchapmi et al.[5] developed TopNet, a hierarchical network is generates point clouds without needing predefined topology, enhancing adaptability and generality. The attention mechanism, influential in natural language processing, computer vision, and speech recognition[8], has been harnessed for point cloud processing. Zhao et al.[9] first applied Transformers to point cloud segmentation, while Yu et al.[10] employed a Transformer-based encoder-decoder for completion, though local detail retention remains a challenge.

Beyond classical architectures, GANs are also used for point cloud completion. Huang et al.'s PF-Net[6] innovatively predicts missing parts hierarchically, whereas Zhang et al.'s ShapeInversion[7] uses GAN inversion to reconstruct complete shapes from partial data, optimizing the completion workflow.

Previous methods excel on regular, structured datasets but struggle with the complexities of plant leaf point clouds, characterized by irregularity and sheet-like structures. This paper presents MAEGNet, a novel completion network tailored for leaf point clouds, showing improved performance. MAEGNet consists of an encoder-decoder design where encoders use dual attention modules to blend multi-scale point cloud features. This cross-level fusion strengthens the model's ability to extract rich features. Decoders employ FC layers with skip connections to preserve feature info throughout decoding. The main contributions of this work can be summarized as follows:

- A multi-layer attention fusion module has been devised that skillfully extracts and consolidates multi-level features tailored to leaf shapes, effectively integrating both high-dimensional and low-dimensional spatial details. This design innovation significantly enhances the network's learning potential.
- An edge-aware loss function has been introduced, which contrasts the disparities between the edges of the input and reconstructed point clouds. By doing so, it acts as a guiding force for the network to give precedence to the learning of the leaf point cloud's structural edge details. This strategic addition ensures that the network can more effectively preserve and rebuild the intricate boundaries of leaves during the completion process.

II. DATA COLLECTION AND PROCESSING

A. Data Collection

This study targets Caladium plants, with leaf counts varying from 3 to 7. Economical 3D point cloud data is obtained using SFM[11] for 3D reconstruction. A custom platform with a white background, LED lights, a camera rig, a

979-8-3503-6189-6/24/$31.00©2024 IEEE

rotating turntable, and synced triple cameras (Figure 1) captures 180 multi-angle 2D images per plant (60 shots each).

These 2D images are transformed into a 3D Caladium model and sampled into a point cloud. Noisy data is filtered using color and mean filters, resulting in clean point clouds. K-Means clustering alongside color filtering isolates and removes non-leaf components. We processed real Caladium plants through multi-view imaging, 3D reconstruction, surface sampling, clustering, and augmentation to generate 8,000 leaf point cloud samples 4,000 for training, 2,000 for validation, and 2,000 for testing. Each dataset entry contains a complete and a synthetically incomplete leaf point cloud for comparative analysis.

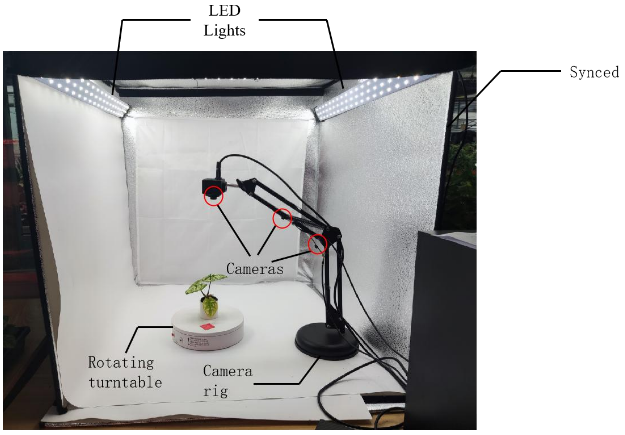

Figure 1. Image acquisition platform

B. Data preprocessing

Preprocessing normalizes point cloud coordinates to [-1, 1] for computational consistency across varying scales. For standardization and dataset expansion, raw data is calibrated and augmented. PCA identifies principal axes, and the primary eigenvector is used as a rotation matrix for alignment. Augmentation creates random planar rotations by multiplying the first two axes with a rotation matrix, yielding five extra samples per original point cloud.

During training, we simulate missing points by randomly translating the point cloud on the xy-plane, computing Euclidean distances from the translated centroid to all points, and eliminating the k-nearest points. This simulates a point cloud with missing data.

III. Method

This study's network structure adopts an encoder-decoder architecture. The encoder extracts multi-level features from point cloud data, focusing on learning the relationship between sparse and complete representations. Conversely, the decoder transforms the encoded, high-level features back into a fully realized point cloud format. Notably, the encoder's construction integrates down-sampling stages and multiple layers of attention fusion modules. For a clear illustration of the overall network design, refer to Figure 2.

A. Encoder

1) Downsampling Module

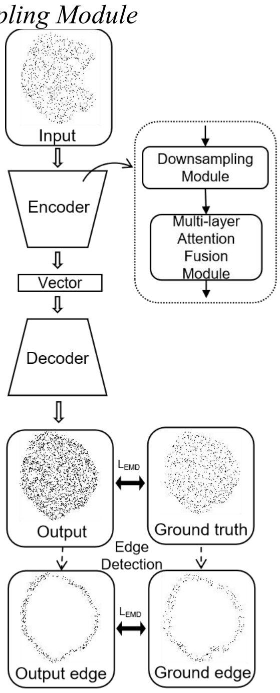

Figure 2. Overview of our proposed MAEGNet

Input to the network is a point cloud with x, y, and z coordinates. Initially, convolutional operations amplify the thefeature depth. Then, the point cloud undergoes downsampling. Here, we leverage the Set Abstraction (SA) technique from PointNet++[12] which progressively extracts features at different resolutions. In SA, the process begins with Farthest Point Sampling (FPS) selecting N_1 representative points from the raw point cloud. Next, for each sampled point, K-Nearest Neighbor (KNN) is used to gather K nearby points which can be represented as $\{f_i^1, f_i^2, ..., f_i^K\}$, segmenting the cloud into N_1 clusters: $\{F_1, F_2, ..., F_{N1}\}$. An MLP then embeds these point clusters to derive local features. In this paper, the value of K is set to 32 throughout the entire network.

Hierarchical local feature extraction is a key aspect of this study, involving two cascaded phases. In the first phase, 512 central points are picked, creating a first-tier set of locally grouped points. In the subsequent phase, further refinement occurs by sampling anew within these 512 points, reducing them to 256 secondary central points, denoted as F_{512} and F_{256}. This sequential extraction enables the network to capture features at multiple granularities within the point cloud data.

2) Multi-Layer Attention Fusion Module

The network faces difficulty learning intricate details, particularly at the base and tip of leaves. To tackle this, we integrate the OffSet-Transformer module from PCT[13], which deviates from traditional Transformers. This module introduces an "Offset-Attention" scheme, utilizing point displacement to infer the topology and geometry of the point cloud, thereby accurately capturing subtle spatial details.

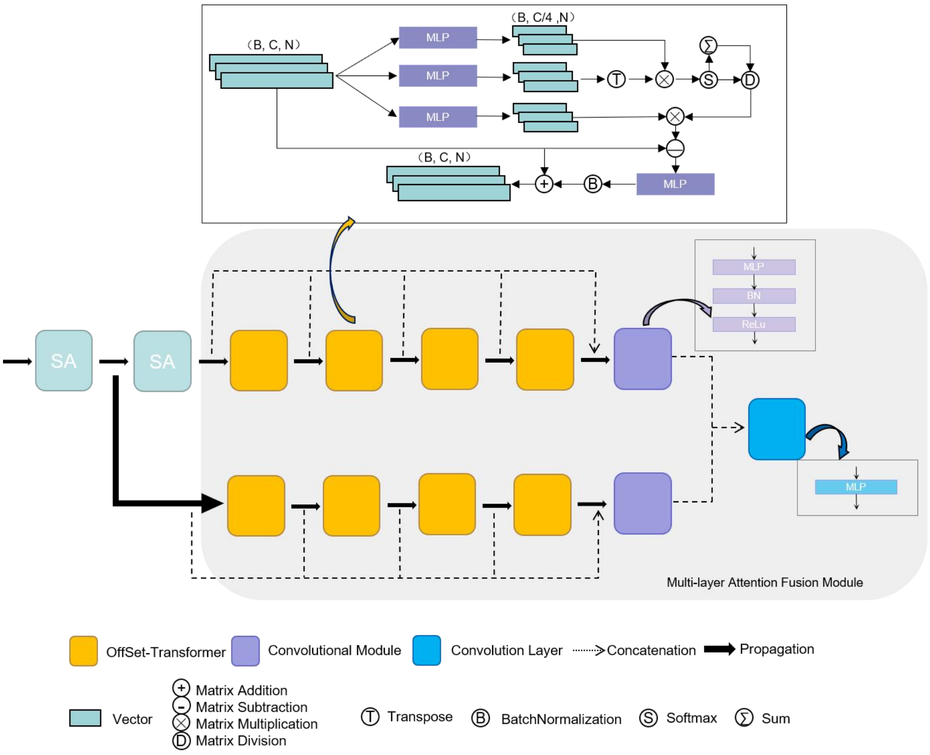

Figure 3. The architecture of the encoder

In this study, after passing the point cloud through two layers of downsampling with differing receptive fields, we stack four attention layers to enrich the network's ability to produce detailed point cloud representations. This step improves the network's expressiveness, generalizability, and overall performance. The multi-layer attention mechanism gradually aggregates and fuses information across different levels within the input features, allowing the model to discern both local and global properties while attending to crucial information. The key equations are shown in (1).

$$\begin{gathered}F_{11}, \ F_{21} = \mathrm{OST}(F_{512}), \mathrm{OST}(F_{256}) \\ F_{1i}, \ F_{2i} = \mathrm{OST}(F_{1(i-1)}), \mathrm{OST}(F_{2(i-1)}), i = 2,3,4 \\ F_{o1} = concat(F_{512}, F_{11}, F_{12}, F_{13}, F_{14}) \\ F_{o2} = concat(F_{256}, F_{21}, F_{22}, F_{23}, F_{24}) \\ F_{o} = \mathrm{mlp}(concat(\mathrm{MLP}(F_{o1}), \mathrm{MLP}(F_{o2})))\end{gathered} \tag{1}$$

where OST denotes Offset Attention Mechanism. F_{1i} and F_{2i} represent the features extracted by the first and second layers of attention, respectively. MLP stands for Multi-Layer Perceptron, which is composed of an MLP layer, a Batch Normalization (BN) layer, and a Rectified Linear Unit (ReLU) activation function. Here, mlp refers to a single MLP layer. F_{oi} represent output vectors.

Finally, the outputs from the two attention modules are concatenated via convolutional layers, ensuring the model draws insights from various receptive fields in the data. Figure 3 illustrates the structure of this multi-layer attention fusion submodule.

B. Decoder

This study's decoder uses a fully connected network (FCN) to decode the compressed, encoded representation of the point cloud into its coordinate space. Taking a 1024-dim feature vector reflecting the global attributes of the point cloud post-encoder, the process begins with a linear transformation, followed by batch normalization to stabilize training and enhance feature distribution. ReLU activation expands the dimension to 2048, and dropout (probability 0.5) regularizes to prevent overfitting.

Next, a similar computational block doubles the feature dimension to 4096. This expanded vector is concatenated with the original 1024-dim input, integrating low- and high-level features. After passing through a final linear layer, this merged vector is projected into an N × 3 matrix, where each row represents the 3D coordinates of a single point in the reconstructed point cloud.

C. Edge Detection

For plant leaf point cloud data, this paper proposes a novel edge detection method where edge loss is computed based on deviations between a completed leaf edge and an intact one, using nearest neighbor points. The process involves:

(1) Calculating the point count within a radius r for each point pi.

(2) Sorting the counts of points within the r-radius vicinity of pi in ascending order, then selecting the top N_2 points as

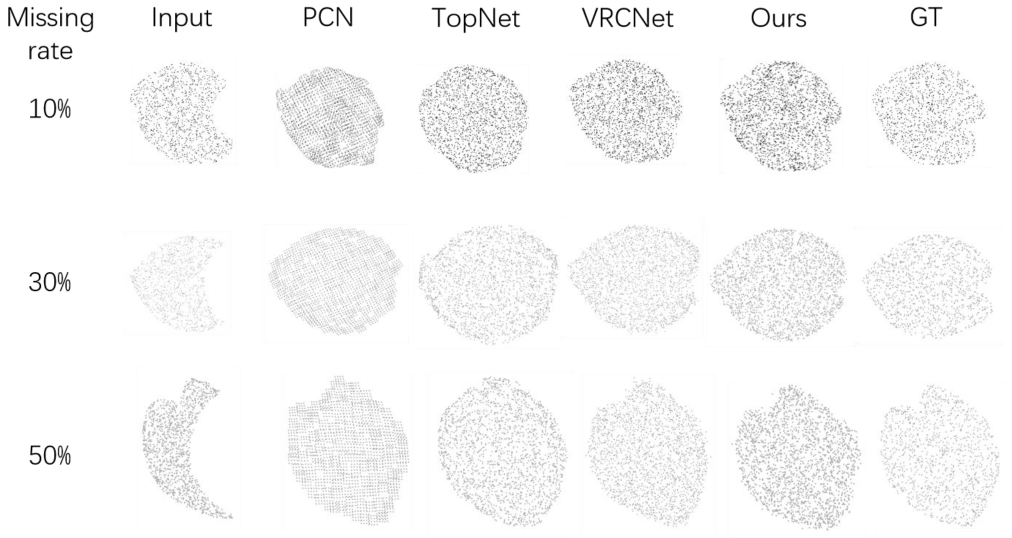

Figure 4. Visualization of completion results produced by different counterparts

edge points in the point cloud. For each point pi, calculate the number of points within its radius r

D. Loss Function

In point cloud analysis, Earth Mover's Distance (EMD) and Chamfer Distance (CD) serve as prevalent similarity metrics [14]. EMD calculates the minimum effort needed to morph one point cloud into another, meticulously accounting for point correspondences and distances, thereby offering precise insight into local distributions and shape variations. Contrarily, CD gauges the overall geometric mismatch by summing up the shortest distances from each point in one cloud to its closest counterpart in the other. Given the significance of nuanced local shapes in plant leaf point cloud completion, our study opts for EMD as the loss function. The calculation processes for EMD and CD can be respectively represented by formula (2):

$$EMD(P,Q) = \min_{\phi} \frac{1}{N} \sum_{\mathrm{i}=1}^{N} \left\| p_i - q_{\phi(i)} \right\|_2$$

$$CD(P,Q) = \sum_{p\in P} \min_{q\in Q} \|p-q\|_2^2 + \sum_{q\in Q} \min_{p\in P} \|p-q\|_2^2 \qquad (2)$$

In this context, ∅ represents a point matching from set A to set B, where each point in A is mapped to a corresponding point in B. The objective of the Earth Mover's Distance (EMD) loss is to identify the optimal point matching that minimizes the distance between the two point clouds. This process can be metaphorically understood as moving each point in set A to some corresponding point in set B such that the total distance traveled by all points is minimized. In essence, it seeks to find the most efficient way to 'transport' the mass distribution of points from one cloud to another, minimizing the cumulative movement cost.

$$Loss = (1 - \alpha)\ EMD\,(P_{gt}, P_{fake}) + \alpha EMD\,(P_{\mathrm{g\_edge}}, P_{f\ \_edge}) \qquad (3)$$

In the equation, P_{gt} and P_{fake} represent the complete point cloud and the point cloud outputted by the completion network, respectively. P_{gt_edge} and P_{f_edge} denote the edges of the complete point cloud and the edges of the point cloud generated by the completion network. EMD(P_{gt}, P_{fake}) indicates the Earth Mover's Distance between the complete point cloud and the completed point cloud, while EMD(P_{gt_edge}, P_{f_edge}) stands for the Earth Mover's Distance between the edges of the complete point cloud and the edges of the completed point cloud. α is the weighting factor.

IV. Experimental and Results Analysis

A. Experimental Environment Parameters

The experiment ran on CentOS 8 OS, powered by an AMD EPYC 7302 16-Core Processor (3.0 GHz) and an NVIDIA GeForce RTX 3090 GPU. The software stack comprised Python 3.9.10, PyTorch 2.0.0, CUDA Toolkit 11.2, cuDNN 8.0.5, and Open3D 0.9.0.

The neural net model was trained for 1000 iterations with a batch size of 64, using Adam optimizer initialized at a learning rate of 0.0001, which was adaptively tuned across epochs. The weight α was fixed at 0.5, playing a role in the Edge Loss function by determining the EMD impact between whole and edge point cloud structures in the total loss computation. Although specifics weren't provided, the learning rate schedule potentially included strategies like step or exponential decay based on model convergence.

B. Evaluation Metrics

In this study, we evaluate the network's performance on point cloud reconstruction using three metrics: Chamfer Distance (CD), Earth Mover's Distance (EMD), and Hausdorff Distance (HD). Lower values in these metrics indicate superior performance. CD gauges the overall geometric resemblance, EMD assesses the local structure similarities, whereas HD signifies the greatest disparity between two point clouds. The mathematical expression for HD is presented in Equation (4):

$$HD(P_{gt}, P_{fake}) = \max\{\min\|a - b\|\}, (a \in P_{gt}, b \in P_{fake}) \qquad (4)$$

C. *Experimental Results*

To analyze the performance of the multi-level attention fusion-based plant leaf completion network proposed in this paper, comparative experiments were conducted against representative point cloud completion networks, including PCN, TopNet, and VRCNet. The quantitative results of these methods on the current dataset are presented in Table 1.

TABLE I. MODEL COMPARISON

Method	***Missing rate***	***CD （$\times10^{-3}$cm）***	***EMD(cm)***	***HD(cm)***
PCN	10％	0.165	48.52	0.056
	30％	0.162	50.35	0.055
	50％	0.161	56.56	0.048
TopNet	10％	0.165	50.32	0.049
	30％	0.164	52.38	0.047
	50％	0.162	60.61	0.044
VRCNet	10％	**0.162**	45.18	**0.046**
	30％	**0.161**	46.43	0.046
	50％	**0.158**	47.59	0.043
Ours	10％	0.163	**42.92**	**0.046**
	30％	**0.161**	**43.99**	**0.045**
	50％	0.159	**44.85**	**0.042**

Quantitative assessments show our method consistently outperforms others in EMD and HD, indicating superior accuracy in reconstructing point cloud shapes and details across all missing data ratios. Though competitive in CD, our method scores slightly higher than VRCNet, suggesting less ideal preservation of point cloud density and distribution due to our EMD-focused training, which emphasizes shape over density/distribution learning.

CloudCompare inspections confirm these results. Figure 4 compares completion outcomes where the first column shows input partial point clouds, followed by PCN, TopNet, VRCNet, and our method's completions. PCN outputs evenly distributed clouds yet lack edge precision. TopNet creates blurred edges and scattered points, and VRCNet misses some leaf base details. However, our method distinctly retains crisp leaf edges and a clearer overall leaf shape and lamina structure. Overall, the experiments prove that our method is tailored to the unique morphological aspects of plant leaves, excelling at retaining sharp edges and significantly boosting completion quality. These results underscore the enhanced capability of our proposed method to capture delicate leaf details and uphold structural integrity during point cloud completion.

V. CONCLUSION

Our specialized MAEGN network targets leaf point cloud completion, significantly enhancing its ability to identify leaf shapes. The innovative architecture incorporates an attention-driven encoder-decoder setup, featuring a fusion module that merges multi-scale attention-weighted features, refining the model's grasp of leaf structures and boosting its performance on leaf reconstruction tasks. Emphasizing leaves' planarity, we added an edge-aware loss function to guide the model in precisely capturing edge details. Experiments on a test dataset showed impressive results: Chamfer Distance (CD) of 0.163 × 10^{-3} cm, Earth Mover's Distance (EMD) of 42.92 cm, and Hausdorff Distance (HD) of 0.046 cm.

ACKNOWLEDGMENT

This research was supported by the Key-Area R&D Program of Guangdong Province under grant 2021B0101220006, the Science and Technology Planning Project of Guangdong Province under grant 2019A050510041, the Natural Science Foundation of Guangdong Province under grant 2021A1515012300, the National Natural Science Foundation of China (NSFC) under grants 61976058, 92267107, the Science and Technology Planning Project of Guangzhou under grants 202103000034, 202002020090.

REFERENCES

[1] Y He, X. Y. Li, G. F. Yang, et al. Research progress and prospect of indoor high-throughput germplasm phenotyping platforms[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(17): 127-141.

[2] G. Bai, Y. F. Ge. Crop Stress Sensing and Plant Phenotyping Systems: A Review[J]. Smart Agriculture, 2023, 5(1): 66-81.

[3] M. F. Ren , G. L. Mao, S. Z. Liu, et al.Research progress on the effects of light quality on plant growth and development, photosynthesis, and carbon and nitrogen metabolism[J].Plant Physiology Journal,2023,59(07):1211-1228.DOI:10.13592/j.cnki.ppj.300151.

[4] W. Yuan, T. Khot, D. Held, et al. Pcn: Point completion network[C]//2018 International Conference on 3D Vision (3DV). IEEE, 2018: 728-737.

[5] L. P. Tchapmi, V. Kosaraju, H. Rezatofighi, et al. Topnet: Structural point cloud decoder[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 383-392.

[6] Z. Huang, Y. Yu, J. Xu, et al. Pf-net: Point fractal network for 3d point cloud completion[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 7662-7670.

[7] J. Zhang, X. Chen, Z. Cai, et al. Unsupervised 3d shape completion through gan inversion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 1768-1777.

[8] A. VASWANI, N. SHAZEER, N. PARMAR, et al. Attention is All You Need[J]. Neural Information Processing Systems, Neural Information Processing Systems, 2017.

[9] H. Zhao, L. Jiang, J. Jia, et al. Point transformer[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 16259-16268.

[10] X. Yu, Y. Rao, Z. Wang, et al. Pointr: Diverse point cloud completion with geometry-aware transformers[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 12498-12507.

[11] M. J. Westoby, J. Brasington, N. F. Glasser, et al. ＇Structure-from-Motion＇photogrammetry: A low-cost, effective tool for geoscience applications[J]. Geomorphology, 2012, 179: 300-314.

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

[13] M. H. Guo, J. X. Cai, Z. N. Liu, et al. Pct: Point cloud transformer[J]. Computational Visual Media, 2021, 7: 187-199.

[14] Y. Rubner, C. Tomasi, L. J. Guibas . The earth mover's distance as a metric for image retrieval[J]. International journal of computer vision, 2000, 40: 99-121.

[15] L. Pan, X. Chen, Z. Cai, et al. Variational relational point completion network[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 8524-8533.

[16] GIRARDEAU-MONTAUT D. CloudCompare[J]. France:EDF R&D Telecom ParisTech, 2016,