

基于注意力机制的植物三维点云语义分割方法

邹一波^{1,2} 周泽政¹ 陈明^{1,2} 葛艳^{1,2} 王文娟^{1,2}

(1.上海海洋大学信息学院,上海 201306; 2.农业农村部渔业信息重点实验室,上海 201306)

摘要: 在植物表型分析中,植物器官分割是实现自动、准确、无损、高通量表型参数测量的关键。传统的植物器官分割方法凭借经验手动设置参数和调整算法,而现有的基于深度学习的分割方法存在对局部特征和全局特征表达能力不足的缺陷。针对以上问题,本文提出一个基于注意力机制的植物三维点云语义分割网络(APSNet)。在编码阶段提出了一种基于注意力机制的局部(邻域)特征提取方法,充分利用多级点云特征,提高了网络提取点云局部(邻域)特征的能力。在解码阶段提出了一种结合特征距离和空间距离的双近邻插值上采样方法,更准确地恢复下采样时丢失的点云特征,进一步增强了网络对局部特征的表达。同时引入通道和多头空间自注意力机制,增强网络对某些重要通道的关注和全局几何结构的捕捉能力,提高了网络对全局特征的表达。在多种植物点云数据集上的实验结果表明,该方法语义分割平均交并比分别达到 87.32%、79.68%、94.73%、91.43%、95.02%,均优于 DGCNN、PointCNN、ShellNet 等目前流行的深度学习网络。通过交叉验证实验和消融实验,证实了网络泛化性和有效性。在 ShapeNet 数据集上进行了相关实验,该网络在其他非植物三维点云目标语义分割任务上也取得了较好的分割结果。

关键词: 三维点云; 注意力机制; 植物器官分割; 深度学习

中图分类号: S126

文献标识码: A

文章编号: 1000-1298(2025)03-0129-11

OSID:



3D Plant Point Cloud Semantic Segmentation Method APSNet Based on Attention Mechanism

ZOU Yibo^{1,2} ZHOU Zezheng¹ CHEN Ming^{1,2} GE Yan^{1,2} WANG Wenjuan^{1,2}

(1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China

2. Key Laboratory of Fisheries Information, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China)

Abstract: In plant phenotypic analysis, plant organ segmentation is the key to achieve automatic, accurate, non-destructive and high-throughput phenotypic parameter measurement. Traditional plant organ segmentation methods rely on experience to manually set parameters and adjust algorithms, but the existing deep learning based segmentation methods have insufficient ability to express local and global features. In order to solve these shortcomings, a semantic segmentation network for three-dimensional point clouds was proposed based on attention mechanism. In the coding stage, a local feature extraction method based on attention mechanism was proposed, which made full use of multilevel point cloud features and improved the ability of network to extract local feature of point cloud. In the decoding stage, a double nearest neighbor interpolation upsampling method combining feature distance and spatial distance was proposed to recover the lost point cloud features in downsampling more accurately, and further enhanced the expression ability of local features. At the same time, the channel and multi-head spatial self-attention mechanism were introduced to enhance the attention of the network to some important channels and ability to capture the global geometric structure, and improve the expression ability of the network for global features. Experimental results on a variety of plant point cloud datasets showed that the mean intersection over union of semantic segmentation of the proposed method reached 87.32%, 79.68%, 94.73%, 91.43%, 95.02%, respectively, which were better than those of popular deep

收稿日期: 2024-10-22 修回日期: 2024-12-12

基金项目: 上海市科技创新计划项目(20dz1203800)

作者简介: 邹一波(1983—),男,讲师,博士,主要从事区块链技术与机器视觉研究,E-mail: ybzou@shou.edu.cn

通信作者: 陈明(1966—),男,教授,博士,主要从事智慧水产养殖和水产品追溯技术研究,E-mail: mchen@shou.edu.cn

learning networks such as DGCNN, PointCNN, ShellNet and so on. Cross-validation experiments and ablation experiments were carried out to confirm the generalization and effectiveness of the network. Relevant experiments were carried out on ShapeNet dataset, and the network also achieved good segmentation results on other non-plant 3D point cloud target semantic segmentation tasks.

Key words: 3D point cloud; attention mechanism; plant organ segmentation; deep learning

0 引言

植物表型是指植物对其基因型和生长环境所表现出的独特生理生化形态参数或性状^[1]。在现代农业中,植物表型是分析单株生长发育的基础。实时监测和分析植物叶片形状、颜色和大小等表型参数,有助于检测病虫害、预测作物产量,选择优质作物品种^[3]。传统表型参数获取方法依赖人工手动测量,耗费大量劳动和时间成本,且易受主观影响。一些性状衍生方法甚至对植物具有破坏性^[4]。此前,基于二维图像的深度学习技术一直是植物器官分割问题的主要解决方案^[5-6]。然而,基于二维图像的植物器官分割方法存在缺乏深度信息导致空间结构难以区分的缺陷。

近年来基于三维点云的植物器官分割逐步成为现代农业信息技术研究的热点问题^[7]。基于三维点云的植物器官分割方法为自动、准确、无损、高通量测量植物形态表型参数提供了关键技术手段。文献[8-12]所述点云分割算法常需凭借经验手动设置参数和调整算法,自动化程度低,难以大规模应用;并且在处理复杂的植物形态时,很难保持稳定的性能。使用深度学习进行三维点云分割的优点显著,尤其在处理复杂三维结构和实现高精度分割任务方面具有独特优势^[13]。因此随着深度学习技术的不断发展和硬件计算能力的提升,基于深度学习的植物点云器官分割成为当前的研究热点。目前分割方法主要分为基于多视图投影的方法^[14-15]、基于体素化的方法^[16]和基于点的直接处理方法^[17-19]。

目前影响植物点云分割网络性能的因素包括3个方面:①大部分网络只是在特征空间对邻域信息进行聚合,而未考虑到坐标空间信息,未融合坐标空间信息会导致网络对植物局部几何结构理解不充分,进而引发分割精度和鲁棒性降低的问题。②在当前植物分割网络解码阶段,仅根据几何距离选择最近的一个点进行赋值,可能导致细节丢失或插值不准确。③过分聚焦局部特征,缺乏对全局信息的利用,可能导致网络对整体结构理解不足,无法识别植物器官边界和位置关系,导致分割不准确。

针对上述问题,本文提出一种基于注意力机制的植物三维点云语义分割网络,首先针对现有植物

分割网络邻域特征提取不充分的缺陷,构建一个注意力局部(邻域)特征提取模块(Attention local feature extraction module, ALEFM)。为了使网络能够更好地理解和处理点云细节部分,在解码阶段提出一个双近邻插值上采样模块(Double nearest neighbor interpolation module, DNNIM)。提出多头空间自注意力模块(Multi-head spatial self-attention module, MHSSAM)和通道注意力模块(Channel attention module, CAM),以增强网络在空间和通道维度上提取全局信息的能力,以期实现植株语义分割。

1 数据

1.1 数据获取

为了评估所提出网络的分割性能,选择公开数据集 Pheno4D^[20]和 Planest3D^[21]进行训练和测试。Pheno4D数据集共有224个点云,在7株玉米和7株番茄植株第1次发芽后每天捕获得到这些点云。Planest3D数据集包含10株辣椒点云、10株玫瑰点云和14株红醋栗点云。

1.2 数据预处理

Pheno4D数据集将点云标记为“土壤”、“茎”和“叶”点,其中每片叶子在整个测量期间被分配一个唯一的标签,如图1a所示。为了便于进行植物语义分割,使用CloudCompare工具手动将所有叶子点重新标记为相同的“叶”标签。图1b为重新标记后的植物点云。Pheno4D数据集收集12 d玉米点云和

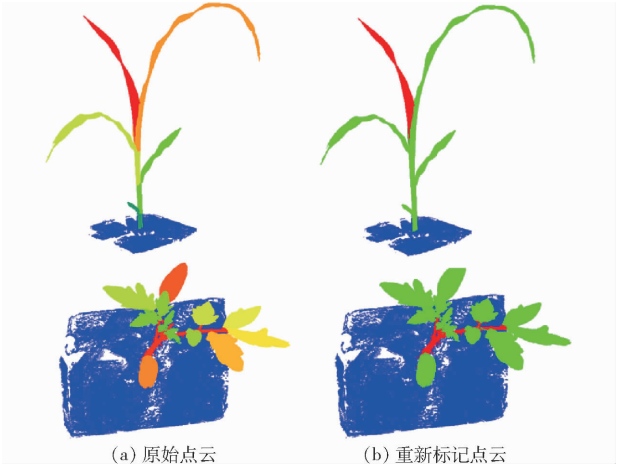


图1 点云重新标记示意图

Fig. 1 Illustration of point cloud relabeling

原始 Pheno4D 和 Planest3D 数据集采用 TXT 文件格式存储数据,读取效率较低,特别是对于大规模数据,I/O 操作时间较长。为此,将点云数据转换为 HDF 格式以提高数据读取效率。由于原始点云通常包含大量的点,这会导致消耗巨大的计算资源。为了提高处理效率、降低计算复杂度,使用与 DGCNN^[22] 相同的数据策略来处理输入玉米和番茄点云。将植物点云划分为固定大小的长方体块,然后由网络独立处理每个长方体块。并且以固定的 2 048 个数对块中的点进行最远点采样操作,以确保采样点均匀分布,保留点云的全局几何结构。对于块中不足 2 048 个点的点云块,进行重复随机采样,凑够 2 048 个点。Pheno4D 和 Planest3D 数据集分配情况如表 1、2 所示。最后将同一植物点云预测块进行组合,得到完整分割。

Tab. 1 Pheno4D data set allocation**Tab. 2 Planest3D data set allocation**

Fig. 2 Network structure diagram

可使网络更准确地理解和处理点云中的复杂几何结构。因此本文引入注意力局部特征提取模块,模块结构如图 3 所示。它通过池化、多层感知机和 Sigmoid 操作将邻域点的低级特征融合成一个注意力分数,作为权重与每个邻域点的当前特征相

乘得到加权特征。然后在编码阶段添加邻域点相对于中心点的局部笛卡尔积坐标信息和局部极坐标信息,以帮助网络更好地捕捉局部几何形状和拓扑结构,最后通过共享多层感知机得到最终输出。

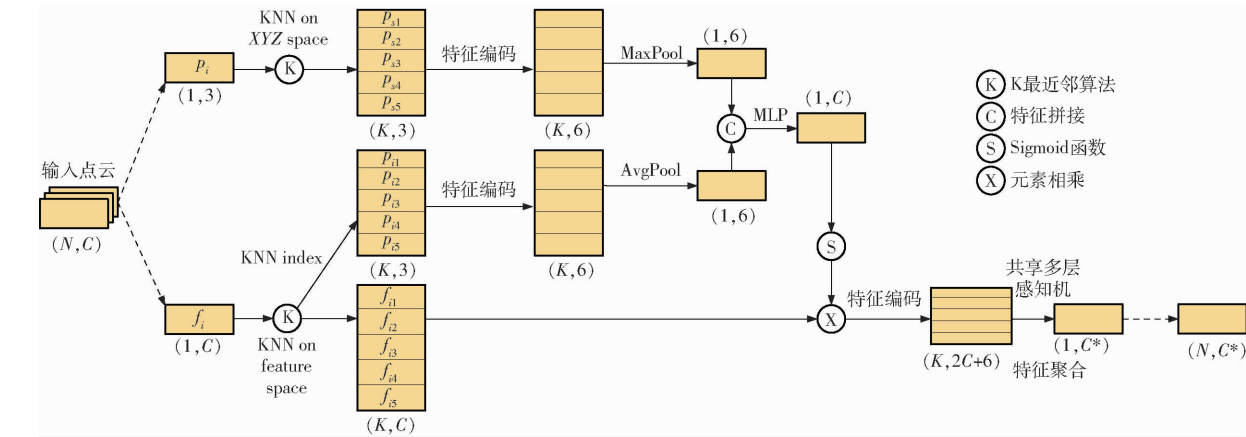


图 3 注意力局部特征提取模块结构图
Fig. 3 Attention local feature extraction module

以点 i 为例, p_i 为其三维坐标向量, f_i 为其特征向量, C 为特征维数。首先以点 i 为中心点, 在坐标空间 (XYZ space) 通过 K 最近邻算法计算中心点与剩余点之间的几何距离。选择最近的 k 个点作为最近邻点, 记作 $\{p_{s1}, p_{s2}, \dots, p_{sk}\}$ 。在特征空间 (Feature space) 通过 K 最近邻算法计算中心点与剩余点之间的特征距离。选择最近的 k 个点作为最近邻点, 记作 $\{f_{i1}, f_{i2}, \dots, f_{ik}\}$ 。然后将特征空间邻域点的索引 (index) 信息传入坐标空间, 记作 $\{p_{i1}, p_{i2}, \dots, p_{ik}\}$ 。由于极坐标含有距离和方位信息, 是笛卡尔积坐标不具备的, 因此将坐标空间的邻域点的笛卡尔积坐标转换为极坐标再与原坐标进行拼接。计算式为

$$f_q^k = \sigma(\text{MLP}(\text{MaxPool}(\mathbf{p}_s^k) \oplus \text{AvgPool}(\mathbf{p}_i^k))) \otimes f_i^k \quad (1)$$

其中 $\mathbf{p}_s^k$ 和 $\mathbf{p}_i^k$ 表示以点 i 为中心点, 不同 index 信息, k 个近邻点的笛卡尔积坐标和极坐标的拼接向量, σ 为 Sigmoid 函数, $\oplus$ 表示特征拼接操作, $\otimes$ 表示逐元素相乘, MLP 表示多层感知机, MaxPool 表示最大池化, AvgPool 表示平均池化。将 $\mathbf{p}_s^k$ 最大池化, 将 $\mathbf{p}_i^k$ 平均池化后拼接在一起。再将拼接得到的特征经过多层感知机后再经过一层 Sigmoid 函数, 与当前特征空间邻域特征信息 f_i^k 相乘得到模块前半部分的输出。

将得到的结果 f_q^k 进行特征编码操作, 再与邻域点的局部笛卡尔积坐标和局部极坐标拼接, 得到高级邻域特征集合, 最后经过共享多层感知机和池化得到最终输出, 即

$$f_{out} = \text{MaxPool}(\text{MLP}(f_z^k \oplus \hat{p}_i^k \oplus \hat{p}_i^k)) \quad (2)$$

其中 $f_z^k = f_q^k \oplus (f_i - f_q^k)$ (3)
式中 f_i ——中心点 i 特征向量
 f_z^k ——经过特征编码操作后近邻点特征向量
 $\hat{p}_i^k, \hat{p}_i^k$ —— k 个邻域点局部笛卡尔积坐标和局部极坐标

对于邻域点位置表示, 图 4 为组内点 p_{ik} 相对于中心点 p_i 的位置关系。如图 4 所示, p_i 为中心点, p_{ik} 为它的第 k 个邻域点, r 表示极径长度, θ 表示直角坐标系中向量 (x, y, z) 与 Z 轴夹角, Φ 表示向量 (x, y, z) 在 XY 平面上投影向量与 X 轴正向夹角。

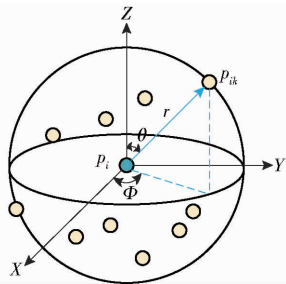


图 4 近邻点相对位置
Fig. 4 Relative position of the nearest neighbors

p_{ik} 相对于 p_i 的局部笛卡尔积坐标表示为 $\hat{p}_{ik} = (\hat{x}_{ik}, \hat{y}_{ik}, \hat{z}_{ik})$, 计算式为

$$\hat{x}_{ik} = x_{ik} - x_i \quad (4)$$

$$\hat{y}_{ik} = y_{ik} - y_i \quad (5)$$

$$\hat{z}_{ik} = z_{ik} - z_i \quad (6)$$

式中 (x_i, y_i, z_i) —— p_i 全局笛卡尔积坐标
 (x_{ik}, y_{ik}, z_{ik}) —— p_{ik} 全局笛卡尔积坐标
 p_{ik} 局部极坐标表示为 $\tilde{p}_{ik} = (\hat{r}_{ik}, \hat{\theta}_{ik}, \hat{\Phi}_{ik})$, 计算式为

$$\hat{r}_{ik} = \sqrt{\hat{x}_{ik}^2 + \hat{y}_{ik}^2 + \hat{z}_{ik}^2} \quad (7)$$

$$\hat{\theta}_{ik} = \arccos \frac{\hat{y}_{ik}}{\hat{r}_{ik}} \quad (8)$$

$$\hat{\phi}_{ik} = \arctan \frac{\hat{z}_{ik}}{\hat{r}_{ik}} \quad (9)$$

2.3 双近邻插值上采样模块

通过下采样操作后,可用更少的点概括原始点信息,使网络更好地捕捉全局特征。在解码部分,要将这些融合了大量点特征的采样点恢复到原始点中。现有的上采样方法大多是基于几何距离的,很少利用到特征距离,例如选择最近采样点的特征值作为原始点特征值的最近邻插值,虽简单快速,但可能导致特征不连续,丢失点云细节。为了在恢复原始空间分辨率的过程中,尽可能保留和恢复更多的细节信息,设计双近邻插值上采样模块。

该模块结构如图5所示,以经过池化的点云 $(N/64, C)$ 为例,在坐标空间根据几何距离获得最近索引,而后将特征值映射到原始点上。同理,在特征空间根据特征距离获得最近索引,将特征值映射到原始点上。最后,将两者拼接再经过一个共享多层感知机得到最终输出。

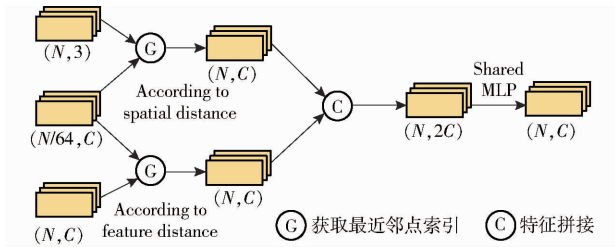


图5 双近邻插值上采样模块结构图

Fig. 5 Double nearest neighbor interpolation upsampling module

2.4 多头空间自注意力模块

多头空间自注意力模块如图6所示,通过将输入特征划分为多个子空间,每个子空间对应一个独立的注意力头,增强了特征的表达能力。每个注意力头在其特征子空间专注不同点之间的依赖关系,使得网络能够在全局范围内捕获点云中的长距离依赖关系,同时也能在局部范围内捕捉点与点之间的细粒度关系。

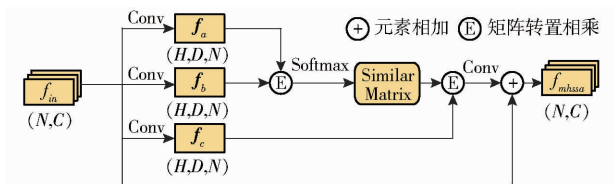


图6 多头空间自注意力模块

Fig. 6 Multi-head spatial self-attention module

输入特征向量形状为 (N, C) ,使用3个一维卷

积操作生成查询 f_a 、键 f_b 和值 f_c 特征向量,即

$$f_a = \text{Conv1D}(f_{in}) \quad (10)$$

$$f_b = \text{Conv1D}(f_{in}) \quad (11)$$

$$f_c = \text{Conv1D}(f_{in}) \quad (12)$$

将结果向量调整为形状 (H, D, N) ,其中 H 为注意力头数量, $D = C/H$ 表示每个头维度。在每个注意力头中,使用查询特征和键特征计算每个头的点对相似性矩阵 S ,其形状为 (H, N, N) 。通过Softmax函数对相似性矩阵的最后一个维度进行归一化,得到注意力权重 A ,即

$$S = f_a f_b^T \quad (13)$$

$$A = \sigma(S) \quad (14)$$

利用注意力权重 A 对值特征 f_c 进行加权求和得到加权特征,然后将各个头的结果拼接并恢复到原始通道维度。拼接后的特征通过一个1D卷积层映射回输入特征维度,并通过残差连接与输入特征 f_{in} 相加得到最终输出,即

$$f_{mhssa} = \text{Conv1D}(A f_c^T) + f_{in} \quad (15)$$

2.5 通道注意力模块

通道注意力机制可用于提高模型对输入特征中关键信息的学习能力。通道注意力模块如图7所示,通过分配较高的权重给特征向量中重要的维度(通道),强化这些维度特征,通过分配较低的权重给不重要的特征维度(通道),抑制不重要特征。

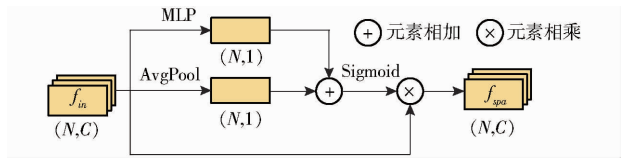


图7 通道注意力模块

Fig. 7 Channel attention module

首先通过平均池化和最大池化操作对输入点云特征 f_{mhssa} 进行聚合,生成两个 $1 \times C$ 特征向量 f_{avg} 和 f_{max} ,分别表示输入点云在通道维度上的平均水平和最大水平,再将这两个向量分别输入到多层感知机,将两个输出特征向量通过向量相加求和,最后通过Sigmoid操作获取一个长度 $1 \times C$ 的通道权重向量 w ,计算式为

$$w = \sigma(\text{MLP}(\text{Avg}(f_{mhssa})) \oplus \text{MLP}(\text{Max}(f_{mhssa}))) \quad (16)$$

将该权重向量与经空间注意力机制强化后特征对位相乘,得到通道注意力筛选后增强特征 f_{out} ,即

$$f_{out} = f_{mhssa} \otimes w \quad (17)$$

3 实验结果与分析

3.1 实验评价指标

使用平均交并比(Mean intersection over union,

mIoU) 和每个类交并比 (Intersection over union, IoU)来评价点云语义分割效果。

3.2 实验设置

所有实验均在 Intel (R) Xeon (R) Platinum 8255C CPU、主频 2.50 GHz, NVIDIA GeForce RTX 3090 GPU 的 64 位服务器上进行,服务器系统为 Linux Ubuntu 18.04,运算平台为 CUDA - Toolkit 11.0, GPU 加速库为 Cudnn 8.0.5,深度学习框架为 Pytorch 1.7.1。网络训练时批量大小 (Batch size) 设置为 32,测试时 Batch size 设置为 16,初始学习率设置为 0.001,训练迭代次数设为 200。训练过程使用随机梯度下降 (SGD) 进行训练,设置动量为 0.9,权重衰减系数为 0.000 1。

3.3 番茄数据集实验结果

番茄语义分割结果如表 3 所示,土壤类 IoU 均可达到 99% 以上, PiontCNN^[23] 和 ShellNet^[24] 的叶类 IoU 略大于本文方法。本文方法茎类 IoU 达到最高 (70.84%), 高于另外 3 种网络,分割效果提升明显。与 DGCNN、PiontCNN 和 ShellNet 相比, mIoU 分别提高 6.73、2.14、1.06 个百分点。

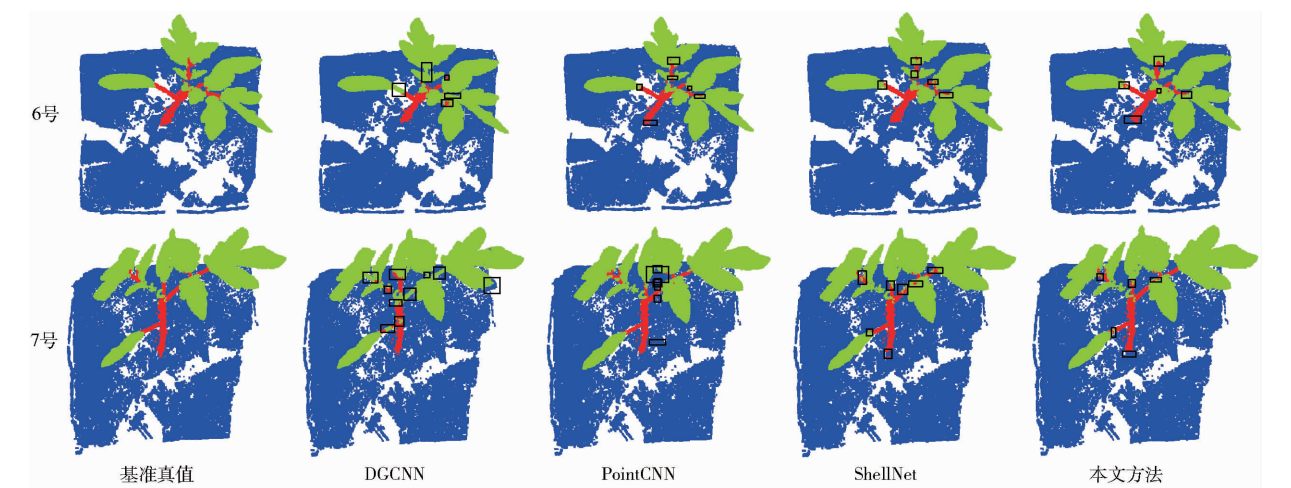


图 8 番茄数据集语义分割结果可视化

Fig. 8 Visualization of semantic segmentation results for tomato dataset

3.4 玉米数据集实验结果

玉米语义分割结果如表 4 所示, DGCNN 的土壤类 IoU 略大于本文方法。除了土壤类 IoU 不是最优值, 其余类别和 mIoU 均是最优值。与 DGCNN、PiontCNN 和 ShellNet 相比, 本文方法 mIoU 分别提高 7.55、5.01、2.23 个百分点。

玉米分割可视化结果如图 9 所示, DGCNN、PiontCNN、ShellNet 和本文方法对于玉米茎叶均不能进行很好的分割。由表 4 可知, 在玉米数据集上茎的交并比也远低于其他 2 种, 有以下原因: ①玉米茎通常位于植株中心位置, 容易被周围的叶片遮挡, 且存在断续不完整的现象。这种不完整性会使得模型

表 3 不同网络在番茄数据集上结果对比

Tab.3 Comparison of results of different networks				
on tomato dataset				%
网络	mIoU	IoU		
		土壤	茎	叶
DGCNN	80.59	99.35	51.92	90.49
PointCNN	85.18	99.13	63.89	92.53
ShellNet	86.26	99.25	65.64	93.90
本文方法	87.32	99.36	70.84	91.77

为了直观感受本文方法分割效果, 将 4 种网络分割结果进行可视化, 结果如图 8 所示。由图 8 可知, DGCNN 几乎只能识别主要的茎, 对于分茎不能很好地识别。PiontCNN、ShellNet 和本文方法可以正确识别大部分茎叶, 但是对于茎叶连接处还是不能很好地识别。由表 3 可知, 茎的交并比远低于其他两种, 有 2 个原因: ①茎空间结构细长且形态不规则, 因此预测区域与真实区域之间差异较大。②茎的点所占比例远低于叶子, 这种数据不均衡问题会导致模型在训练过程中更加关注占比大的器官, 而错误分割占比小的茎。

表 4 不同网络在玉米数据集上结果对比

Tab.4 Comparison of results of different networks				
on corn dataset				%
网络	mIoU	IoU		
		土壤	茎	叶
DGCNN ^[22]	72.13	99.81	27.63	88.95
PointCNN ^[23]	74.67	99.52	35.16	89.31
ShellNet ^[24]	77.45	99.13	40.27	92.96
本文方法	79.68	99.61	45.84	93.58

难以准确预测茎的区域, 导致交并比降低。②玉米的茎在点云中所占的点数较少, 数据量显著小于叶片。③茎与叶在空间结构上极为相似, 这种相似性使得模型在区分茎与叶时容易混淆, 导致预测不准确。

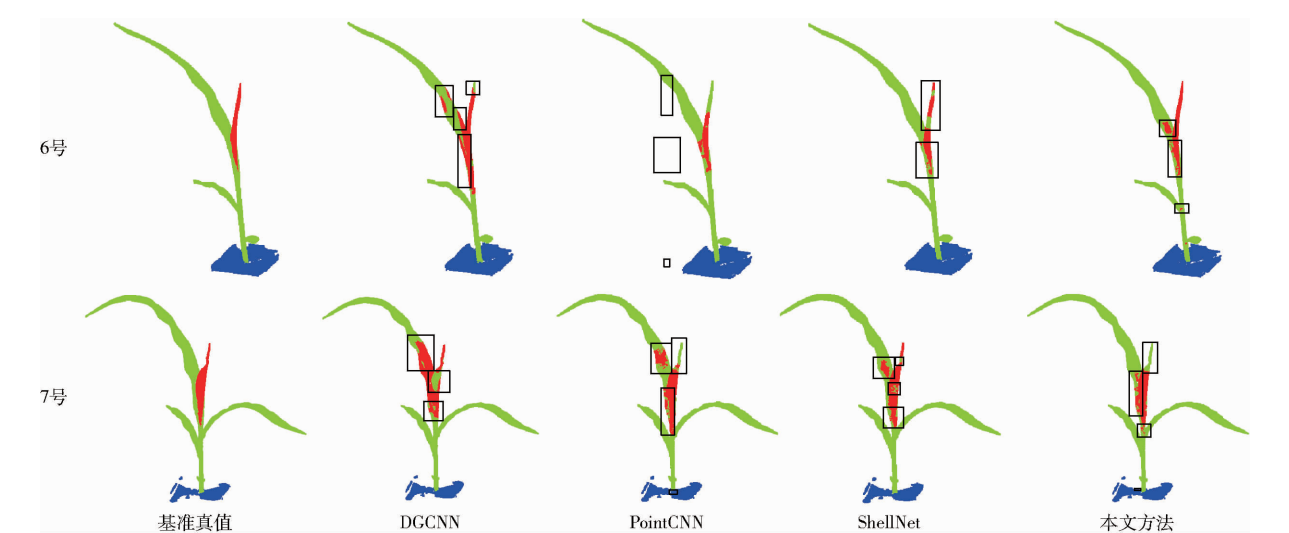


图 9 玉米数据集语义分割结果可视化

Fig. 9 Visualization of semantic segmentation results for corn dataset

3.5 Planest3D 数据集实验结果

基于 Planest3D 数据集在语义分割任务中评估本文方法的分割性能,并与其他主流的深度学习点云分割网络进行对比。由表 5 可知,本文方法 3 种

植物 mIoU 分别达到 94. 73%、91. 43%、95. 02%,玫瑰叶类 IoU 和红醋栗茎类 IoU 低于 ShellNet,多数情况下实现了最佳语义分割性能。

3 种植物分割可视化结果如图 10 所示。在辣

表 5 不同网络语义分割性能对比

Tab. 5 Comparison of semantic segmentation performance of different networks									%
网络	辣椒			玫瑰			红醋栗		
	茎 IoU	叶 IoU	mIoU	茎 IoU	叶 IoU	mIoU	茎 IoU	叶 IoU	mIoU
DGCNN	83. 48	93. 45	88. 47	81. 23	93. 46	87. 35	87. 48	94. 75	91. 16
PointCNN	89. 49	94. 12	91. 8	83. 25	95. 19	89. 22	89. 68	95. 74	92. 71
ShellNet	89. 75	96. 61	93. 18	85. 37	96. 29	90. 83	92. 46	95. 27	93. 87
本文方法	91. 84	97. 62	94. 73	87. 36	95. 49	91. 43	91. 85	98. 19	95. 02

椒点云中,分割错误主要集中在叶片和茎的连接处。而在玫瑰和红醋栗点云中,除茎叶连接处外,叶片中叶存在少量分割错误的情况。原因为:①在植物的茎叶连接部位,往往存在复杂的几何形态和结构过渡,这些部位的点云数据可能具有相似的局部特征,导致模型难以准确识别它们边界。②有些叶片的表面结构可能与茎的表面相似,尤其是当植物的茎较为光滑,或者叶片与茎形态相似时,并且点云数据几何特征可能无明显差异,造成模型识别错误。

3.6 交叉实验

为了评估网络泛化能力,对 Pheno4D 中两个植物数据集进行了交叉验证实验。首先,评估在番茄数据集上训练网络,在玉米数据集上测试网络结果。然后,评估在玉米数据集上训练网络,在番茄数据集上测试网络结果。为了保证实验公平,采用与上述实验相同的参数设置和相同评价指标。

3.6.1 番茄数据集交叉验证

在玉米数据集上训练网络并在番茄数据集上测试网络结果如表 6 所示。所有网络土壤类 IoU 都达

到 98% 以上,只有本文方法叶类 IoU 达到 90% 以上。然而与在番茄数据集上训练的网络相比,所有网络茎类 IoU 均下降到将近 0。图 11 为 4 种网络在番茄数据集上的交叉验证可视化结果,其中 A 行表示模型在番茄数据集上进行训练和测试结果,B 行表示模型在玉米数据集上进行训练,在番茄数据集上进行测试的结果。当从玉米数据集推广到番茄数据集时,所有网络几乎都正确地分类了土壤和叶子,然而所有网络都无法准备识别茎点。

3.6.2 玉米数据集交叉验证

表 7 为在番茄数据集上训练和在玉米数据集上测试的结果。由于两个数据集中的地面具有相似的特征,所有模型土壤类均保持较高 IoU (> 96%)。本文方法叶类 IoU 低于 ShellNet,茎类 IoU 最高(>14%),优于其他 3 类网络模型。图 12 为 4 种网络在玉米数据集上的交叉验证可视化结果,其中 C 行表示网络在玉米数据集上进行训练和测试的结果,D 行表示网络在番茄数据集上进行训练,在玉米数据集上进行测试的结果。当从番茄数据集推广到玉米数据集时,所有网络将大部分茎识别为叶,只有

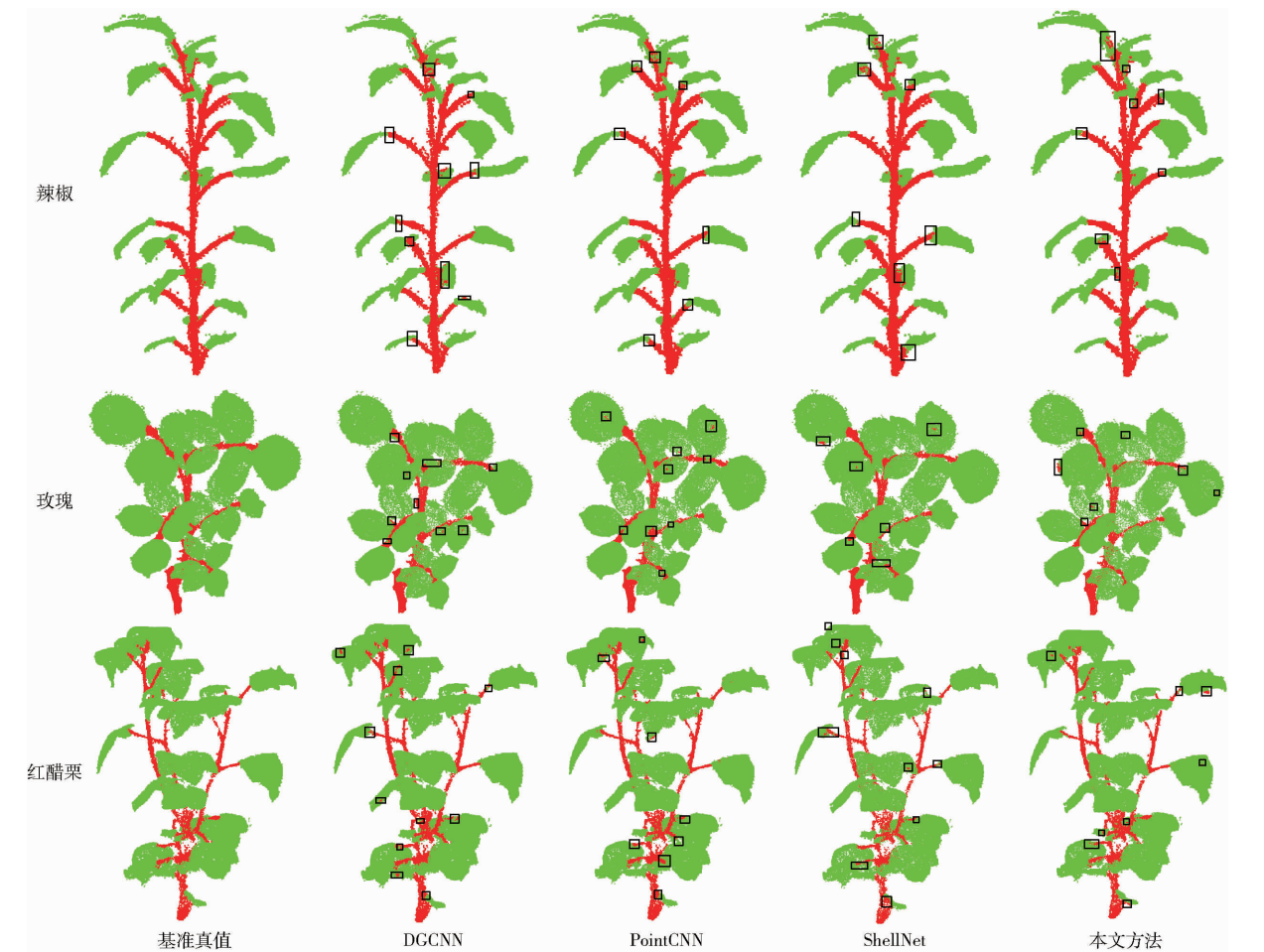


图 10 Planest3D 数据集语义分割结果可视化

Fig. 10 Planest3D data set semantic segmentation results visualization

表 6 玉米数据集训练和番茄数据集测试结果

Tab. 6 Training on corn dataset and testing results

on tomato dataset

%

网络	mIoU	IoU		
		土壤	茎	叶
DGCNN	61. 97	99. 22	2. 59	84. 11
PointCNN	63. 76	98. 49	4. 91	87. 88
ShellNet	64. 16	98. 78	3. 85	89. 85
本文方法	65. 81	99. 64	5. 18	92. 61

少部分茎识别正确。

3. 7 在 ShapeNet 数据集上的实验结果

为了进一步评估本文方法泛化性和鲁棒性,选用 ShapeNet^[25]数据集进行语义分割,并与其他方法进行定量对比。ShapeNet 数据集有 16 880 个 3D 点云,其中训练集有 14 006 个模型,测试集有 2 874 个模型。该数据集有 16 个类,注释 50 个部分,每个点

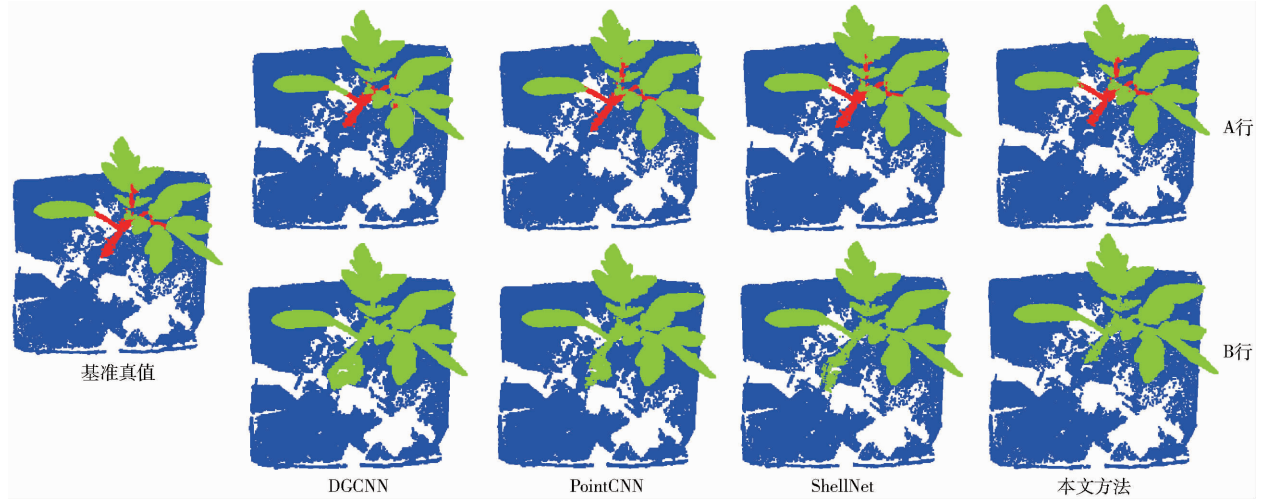


图 11 玉米数据集训练番茄数据集测试分割结果可视化

Fig. 11 Corn dataset training tomato dataset test segmentation results visualization

表 7 番茄数据集训练和玉米数据集测试结果

Tab.7 Training on tomato dataset and testing results on corn dataset

网络	mIoU	IoU			%
		土壤	茎	叶	
DGCNN	60. 86	96. 83	3. 79	81. 95	
PointCNN	62. 42	97. 51	7. 54	82. 21	
ShellNet	62. 91	96. 75	8. 81	83. 17	
本文方法	64. 75	97. 16	14. 17	82. 92	

云包含 2~6 个部分。由于点云数据包含大量的点，直接将其送入网络训练会占用大量计算资源，降低训练速度，对于提高网络的分割性能意义不大，因此采用前文提到的最远点降采样策略将所有点云点数降到 2 048 后送入到网络中进行训练。图 13 为本文方法在 ShapeNet 数据集上的分割可视化结果，分割错误主要集中在不同语义类的边界处，其他区域识别正确。本文方法对 ShapeNet 数据集中每种物

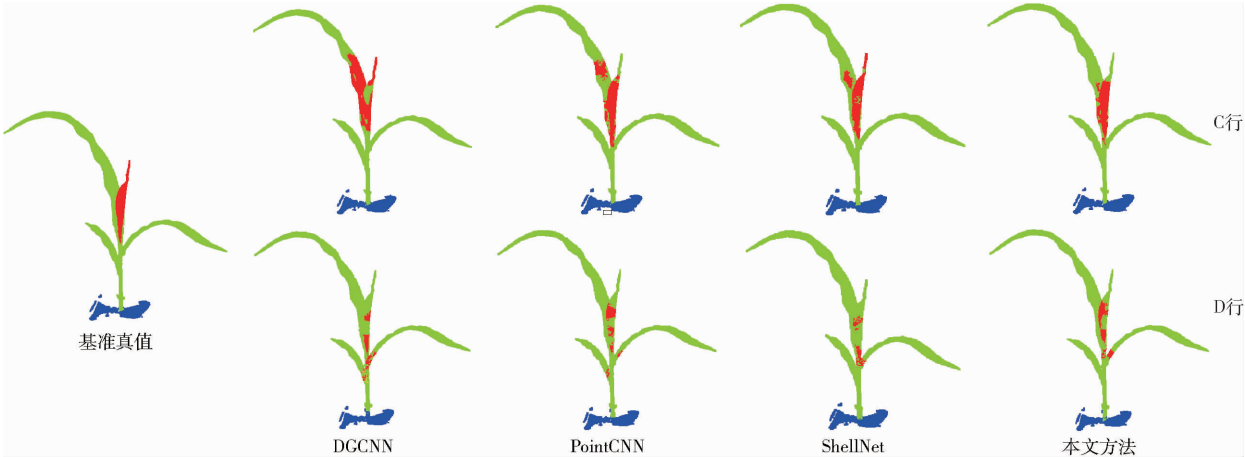


图 12 番茄数据集训练玉米数据集测试分割结果可视化

Fig. 12 Tomato dataset training corn dataset evaluation segmentation results visualization

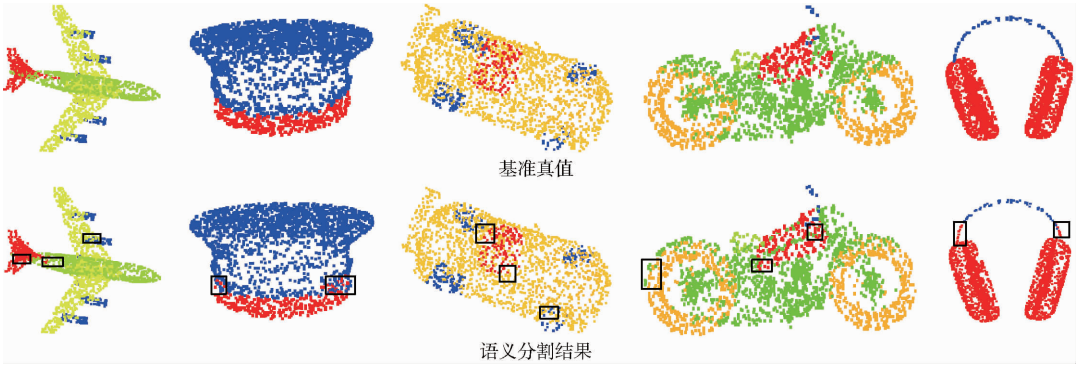


图 13 ShapeNet 数据集分割结果可视化

Fig. 13 Visualization of segmentation results of ShapeNet dataset

件语义分割定量结果如表 8 所示,展示了每个类别 IoU、平均类 IoU (mcIoU) 和平均实例 IoU (mIoU)。mcIoU 为所有形状类别的平均 IoU, mIoU 通过计算所有测试实例的平均 IoU 得到。

由表 8 可知,本文方法 mcIoU 和 mIoU 分别达到 83.2% 和 85.9%,其中 mcIoU 优于其他网络, mIoU 仅次于 PointASNL^[26] 和 PointCNN。PointASNL 引入非局部神经网络,这种架构能够捕捉点云中远距离点之间的依赖关系。使其能够从更广的上下文中理解点云的全局结构。在处理复杂形状或稀疏点云的数据时更为突出。PointCNN 通过 X-Transformation 将不规则点云数据转换为适合卷积操作的结构。这使得它能够充分利用卷积神经网络

优势,在处理不规则点云数据时更加出色。由于 ShapeNet 数据集点云涵盖了多种物体类别,不同类别之间的几何复杂度差异大,本文方法处理复杂物体能力和泛化能力不及 PointASNL 和 PointCNN。综上所述,这是这两类网络 mIoU 优于本文方法的原因。

3.8 消融实验

设计 5 种组合来验证各模块对植物分割任务的有效性,选取前文中的番茄点云作为数据集。未采用注意力局部(邻域)特征提取模块时使用的是图卷积模块,未采用双近邻插值模块时使用的是最近邻插值模块。表 9 为 5 种组合 mIoU 定量结果。

表 8 ShapeNet 数据集分割结果

Tab. 8 Segmentation results of ShapeNet dataset																		%
网络	mIoU	mIoU	aero	bag	cap	car	chair	Ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	Rocket	Skate board	table
PointASNL ^[26]		86.1	84.1	84.7	87.9	79.7	92.2	73.7	91.0	87.2	84.2	95.8	74.4	95.2	81.0	63.0	76.3	83.2
PointNet ^[27]	80.4	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
3D-GCN ^[28]	82.1	85.1	83.1	84.0	86.6	77.5	90.3	74.1	90.9	86.4	83.8	95.6	66.8	94.8	81.3	59.6	75.7	82.8
So-Net ^[29]	81.0	84.9	82.8	77.8	88.0	77.3	90.6	73.5	90.7	83.9	82.8	94.8	69.1	94.2	80.9	53.1	72.9	83.0
PointNet++ ^[30]	81.9	85.1	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6
PointCNN		86.1	84.1	86.5	86.0	80.8	90.6	79.7	92.3	88.4	85.3	96.1	77.2	95.2	84.2	64.2	80.0	83.0
PCNN ^[31]		85.1	82.4	80.1	85.5	79.5	90.8	73.2	91.3	86.0	85.0	95.7	73.2	94.8	83.3	51.0	75.0	81.8
DGCNN ^[22]	82.3	85.2	84.0	83.4	86.7	77.8	90.6	74.7	91.2	87.5	82.8	95.7	66.3	94.9	81.1	63.5	74.5	82.6
Kd-Net ^[32]	77.4	82.3	80.1	74.6	74.3	70.3	88.6	73.5	90.2	87.2	81.0	94.9	87.4	86.7	78.1	51.8	69.9	80.3
本文方法	83.2	85.9	83.8	84.6	87.2	80.8	90.5	77.9	90.8	85.2	86.6	93.9	72.6	91.4	84.4	62.9	74.4	83.7

表 9 消融实验结果

Tab. 9 Results of ablation experiments				%
组合	ALEFM	MHSSAM&CAM	DNNIM	mIoU
A	×	×	×	82.25
B	√	×	×	85.29
C	√	√	×	86.84
D	√	×	√	85.66
E	√	√	√	87.32

注：√表示使用此模块，×表示不使用此模块。

由表 9 可知,随着模块添加,网络分割性能也在不断提高,其中添加注意力局部(邻域)特征提取模块对网络分割性能提升比较明显,mIoU 提升 3 个百分点。多头空间自注意注意力和通道注意力模块对网络的性能影响较大,mIoU 提升 1.55 个百分点。对网络性能影响最小的是双近邻插值上采样模块,语义分割指标 mIoU 提升 0.4 个百分点。通过叠加这 3 个模块,比单独使用其中一种或两种的分割效果要好,进一步证明所提各模块的有效性。

3.9 注意力局部特征提取模块中邻域 k 设置

注意力局部特征提取模块中邻域 k 值的设置对于能否有效捕获局部信息至关重要。为了研究其对网络性能的影响,基于完整的网络即 E 组合,以番茄点云作为数据集,设置不同数值进行测试。

当 k 发生变化时,语义分割指标变化如表 10 所示。随着 k 不断增大,评价指标先是不断升高,然后逐渐下降,当 k 为 24 时,各项 IoU 达到最高。当 k 设置过小(k=6)时,网络将无法充分捕捉点云中的整体结构和上下文信息,而当 k 设置过大(k=96)时,网络会导致局部特征的细节丢失、引入无关信

息,并增加计算复杂度,降低网络对细节的敏感度和性能。

表 10 不同邻域 k 设置下本文方法性能结果

Tab. 10 Quantitative analysis of performance of proposed method under different neighborhood k settings					%
k	mIoU	IoU			
		土壤	茎	叶	
6	85.13	99.23	65.73	90.42	
12	86.18	99.27	68.71	90.57	
24	87.32	99.36	70.84	91.76	
48	85.73	98.72	68.12	90.34	
96	84.82	98.57	66.48	89.42	

4 结束语

植物器官语义分割对于植物表型研究、农业自动化、植物生长监测等领域具有重要意义。根据语义类别对大规模植物数据集 Pheno4D 进行重新标记。提出了一个基于注意力机制的植物三维点云语义分割网络。该网络模型集成了注意力局部(邻域)特征提取模块、双近邻插值上采样模块、多头空间自注意力和通道注意力模块,以充分提取局部特征、恢复点云特征和利用全局信息。本文方法在 Pheno4D 和 Planest3D 植物点云数据集上 mIoU 分别达到 87.32%、79.68%、94.73%、91.43%、95.02%, 优于当前流行的 3 种网络,较好地完成了植物器官语义分割任务。通过相关交叉验证实验,证实了网络具有较好的泛化能力。在 ShapeNet 数据集上的实验结果进一步验证了网络泛化性和鲁棒性。通过消融实验证明了所提出各模块的有效性。

参 考 文 献

[1] ARAUS J L, CAIRNS J E. Field high-throughput phenotyping: the new crop breeding frontier[J]. Trends Plant Sci., 2014, 19(1): 52-61.

[2] CHAIVIVATRAKUL S, TANG L, DAILEY M N, et al. Automatic morphological trait characterization for corn plants via 3D

- holographic reconstruction[J]. *Computers and Electronics in Agriculture*, 2014, 109: 109 – 203.
- [3] 张慧春,周宏平,郑加强,等. 植物表型平台与图像分析技术研究进展与展望[J]. *农业机械学报*, 2020, 51(3): 1 – 17.
ZHANG Huichun, ZHOU Hongping, ZHENG Jiaqiang, et al. Research progress and prospect in plant phenotyping platform and image analysis technology[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(3): 1 – 17. (in Chinese)
- [4] JAY S. In-field crop row phenotyping from 3D modeling performed using structure from motion[J]. *Computers and Electronics in Agriculture*, 2015, 110: 70 – 77.
- [5] 宋彦,杨帅,郑子秋,等. 基于多头自注意力机制的茶叶采摘点语义分割算法[J]. *农业机械学报*, 2023, 54(9): 297 – 305.
SONG Yan, YANG Shuai, ZHENG Ziqiu, et al. Semantic segmentation algorithm based multi-headed self-attention for tea picking points[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(9): 297 – 305. (in Chinese)
- [6] 顾文娟,魏金,阴艳超,等. 基于改进 DeepLabv3 + 的番茄图像多类别分割方法[J]. *农业机械学报*, 2023, 54(12): 261 – 271.
GU Wenjuan, WEI Jin, YIN Yanchao, et al. Multi-category segmentation method of tomato image based on improved DeepLabv3 + [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(12): 261 – 271. (in Chinese)
- [7] PATURKAR A, GUPTA G S, BAILEY D. Making use of 3D models for plant physiognomic analysis: a review[J]. *Remote Sensing*, 2021, 13(11): 2232.
- [8] 梁秀英,周风燃,陈欢,等. 基于运动恢复结构的玉米植株三维重建与性状提取[J]. *农业机械学报*, 2020, 51(6): 209 – 219.
LIANG Xiuying, ZHOU Fengran, CHEN Huan, et al. Three-dimensional maize plants reconstruction and traits extraction based on structure from motion[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(6): 209 – 219. (in Chinese)
- [9] MIAO T, ZHU C, XU T, et al. Automatic stem-leaf segmentation of maize shoots using three-dimensional point cloud[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106310.
- [10] 杨琳,翟瑞芳,阳旭,等. 结合超体素和区域增长的植物器官点云分割[J]. *计算机工程与应用*, 2019, 55(16): 197 – 203.
YANG Lin, ZHAI Ruifang, YANG Xu, et al. Segmentation of plant organs point clouds through super voxel-based region growing methodology[J]. *Computer Engineering and Applications*, 2019, 55(16): 197 – 203. (in Chinese)
- [11] 肖奕同,刘帅,侯晨连,等. 基于三维点云的大豆植株器官分割及表型分析[J]. *中国农业科技导报*, 2023, 25(8): 115 – 125.
XIAO Yitong, LIU Shuai, HOU Chenlian, et al. Organ segmentation and phenotypic analysis of soybean plants based on three-dimensional point clouds[J]. *Journal of Agricultural Science and Technology*, 2023, 25(8): 115 – 125. (in Chinese)
- [12] 彭程,李帅,苗艳龙,等. 基于三维点云的番茄植株茎叶分割与表型特征提取[J]. *农业工程学报*, 2022, 38(9): 187 – 194.
PENG Cheng, LI Shuai, MIAO Yanlong, et al. Stem-leaf segmentation and phenotypic trait extraction of tomatoes using three-dimensional point cloud[J]. *Transactions of the CSAE*, 2022, 38(9): 187 – 149. (in Chinese)
- [13] CHEN Y, XIONG Y, ZHANG B, et al. 3D point cloud semantic segmentation toward large-scale unstructured agricultural scene classification[J]. *Computers and Electronics in Agriculture*, 2021, 190: 106445.
- [14] GUO Y, WANG H, HU Q, et al. Deep learning for 3D point clouds: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(12): 4338 – 4364.
- [15] 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.
- [16] GUO X, SUN Y, YANG H. FF – Net: feature-fusion-based network for semantic segmentation of 3D plant point cloud[J]. *Plants*, 2023, 12(9): 1867.
- [17] LI D, LI J, XIANG S, et al. PSegNet: simultaneous semantic and instance segmentation for point clouds of plants[J]. *Plant Phenomics*, 2022: 9787643.
- [18] HAN B, LI Y, BIE Z, et al. MIX – NET: deep learning-based point cloud processing method for segmentation and occlusion leaf restoration of seedlings[J]. *Plants*, 2022, 11(23): 3342.
- [19] SUN Y, GUO X, YANG H. Win-Former: window-based transformer for maize plant point cloud semantic segmentation[J]. *Agronomy*, 2023, 13: 2723.
- [20] SCHUNCK D, MAGISTRI F, ROSU R A, et al. Pheno4D: a spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis[J]. *PLoS One*, 2021, 16(8): e0256340.
- [21] KEREM M, YUSUF Ş, SERVER K S, et al. PLANesT – 3D: a new annotated dataset for segmentation of 3D plant point clouds[J]. *arXiv Preprint, arXiv:2407.21150*, 2024.
- [22] WANG Y, SUN Y, LIU Z, et al. Dynamic graph CNN for learning on point clouds[J]. *ACM Transactions on Graphics*, 2018, 38(5): 1 – 12.
- [23] LI Y, BU R, SUN M, et al. PointCNN: convolution on x-transformed points[J]. *Advances in Neural Information Processing Systems*, 2018, 31: 820 – 830.
- [24] 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.

- Intelligence, 1992, 14(2): 239 – 256.
- [28] DDOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]//9th International Conference on Learning Representations, ICLR 2021, Virtual Event, 2021.
- [29] CHARLES R Q, SU H, KAICHUN M, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 77 – 85.
- [30] QI C R, YI L, SU H, et al. Pointnet + + : deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), 2017: 5105 – 5114.
- [31] TAUNK K, DE S, VERMA S, et al. A brief review of nearest neighbor algorithm for learning and classification[C]//2019 International Conference on Intelligent Computing and Control Systems (ICCS). Madurai: IEEE, 2019: 1255 – 1260.
- [32] 曾安, 彭杰威, 刘畅, 等. 基于多尺度几何感知 Transformer 的植物点云补全网络[J]. 农业工程学报, 2022, 38(4): 198 – 205.
- ZENG An, PENG Jiewei, LIU Chang, et al. Plant point cloud completion network based on multi-scale geometry-aware point Transformer[J]. Transactions of the CSAE, 2022, 38(4): 198 – 205. (in Chinese)
- [33] KAZHDAN M, BOLITHO M, HOPPE H. Poisson surface reconstruction[C]//Proceedings of the Fourth Eurographics Symposium on Geometry, 2006: 61 – 70.
- [34] XIE H, YAO H, ZHOU S, et al. GRNet: gridding residual network for dense point cloud completion[M]//VEDALDI A, BISCHOF H, BROX T, et al. Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 365 – 381.
- [35] TCHAPMI L P, KOSARAJU V, REZATOFIGHI H, et al. TopNet: structural point cloud decoder[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 383 – 392.
-

(上接第 139 页)

- [25] YI L, KIM V G, CEYLAN D, et al. A scalable active framework for region annotation in 3D shape collections[J]. ACM Transactions on Graphics, 2016, 35(6): 1 – 12.
- [26] YAN X, ZHENG C, LI Z, et al. PointASNL: robust point clouds processing using nonlocal neural networks with adaptive sampling[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 5589 – 5598.
- [27] 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.
- [28] LIN Z, HAUNG S, WANG Y F. Convolution in the cloud: learning deformable kernels in 3D graph convolution networks for point cloud analysis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 1800 – 1809.
- [29] LI J, CHEN B M, LEE G H. So-net: self-organizing network for point cloud analysis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9397 – 9406.
- [30] 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: 5099 – 5108.
- [31] ATZMON M, MARON H, LIPMAN Y, et al. Point convolutional neural networks by extension operators[J]. arXiv Preprint, arXiv:1803.10091, 2018.
- [32] KLOKOV R, LEMPITSKY V. Escape from cells: deep kd-networks for the recognition of 3D point cloud models[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 863 – 872.