

DOI: 10.19650/j.cnki.cjsi.J2311115

顾及长尾分布的机载 LiDAR 点云 CNN 语义分割*

陈睿星¹, 吴 军¹, 赵雪梅¹, 徐 刚²

(1. 桂林电子科技大学电子工程与自动化学院 桂林 541000; 2. 中科院宁波材料技术与工程研究所 宁波 315201)

摘要:针对目前 PointNet++ 系列网络模型倾向于牺牲尾类分割精度以保证全局分割精度这一现象, 构建顾及数据长尾分布的机载 LiDAR 点云语义分割网络, 主要涉及两方面内容, 聚类最远点采样和空间自注意力机制下的局部特征学习。聚类最远点采样通过类内点云最远点采样、划分区域最远点采样以及基于置信度的均值漂移 (Meanshift) 聚类组合策略, 最大程度保留尾类样本并通过循环赋权方式使每类样本均能被网络充分学习; 空间自注意力机制下的局部特征学习为结合不同空间编码方式增强采样点邻域拓扑结构的学习, 以利于从稀疏样本数据中完整学习目标空间结构。公开数据集实验表明, 本文网络模型整体分割精度和平均 F_1 较 PointNet++ 分别提升 6.3% 和 6.6%, 并优于其它 6 种 PointNet++ 系列网络模型及新公布的 10 种网络模型, 具有良好的泛化性能与应用价值。

关键词: 点云语义分割; 卷积神经网络; 长尾分布; 自注意力机制; 聚类最远点采样

中图分类号: TP391.41 TH74 **文献标识码:** A **国家标准学科分类代码:** 420.20

CNN semantic segmentation of airborne LiDAR point cloud considering long-tailed distribution

Chen Ruixing¹, Wu Jun¹, Zhao Xuemei¹, Xu Gang²

(1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541000, China;

2. Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, China)

Abstract: Traditional 3D point semantic segmentation networks based on PointNet++ tend to sacrifice the accuracy of minority classes to maintain the overall accuracy dominated by majority classes. A new CNN is proposed to improve the segmentation accuracy of PointNet++ when processing airborne LiDAR point clouds with long-tailed distribution, which mainly consists of two aspects. The first is cluster-based farthest point sampling (FPS). Through intra-class FPS under proportional constraints, meanshift clustering based on confidence and zoning FPS combined with neighborhood compensation, the samples of minority classes in airborne LiDAR point clouds can be retained to the maximum extent, and can be well learned by the network through re-weighting. The second is local feature learning under the spatial self-attention mechanism. By using different spatial encoding methods, a new spatial self-attention mechanism is constructed to facilitate learning the complete structure of the target from sparse sample data. Therefore, the learning ability of the network model for minority classes is improved while ensuring the good learning ability of the majority classes. Experiments on public data set show that the overall accuracy (OA) and F_1 score in this article have a significant improvement, which is 6.3% and 6.6% higher than those of PointNet++. Compared with other 6 networks based on PointNet++ and the top 10 network model in recent publications, the proposed algorithm has the best performance, good generalization ability and application value.

Keywords: point cloud semantic segmentation; CCN; long-tailed distribution; self-attention mechanism; cluster-based FPS

收稿日期: 2023-03-01 Received Date: 2023-03-01

* 基金项目: 国家自然科学基金项目 (41801233)、桂林电子科技大学研究生教育创新计划资助项目 (2021YCXB07)、宁波市科技创新重大专项 (2020Z013) 资助

0 引言

近年来,机载激光雷达(laser radar, LiDAR)点云数据被广泛应用于大型复杂场景分析^[1]、铁路检测^[2]、地面识别^[3]等众多领域。然而,由于场景复杂性、地物形态分布不规则性以及3D点云数据自身特点,机载LiDAR点云高精度语义分割仍是一项极具挑战性的任务。得益于卷积神经网络(convolutional neural networks, CNN)在众多图像视觉任务中的大获成功,设计针对现实场景、具有端到端输出能力的点云语义分割CNN成为当前研究热点^[4]。

依据是否对点云进行结构化预处理,现有点云语义分割CNN可概略分为两类^[5],间接点云分割和直接点云分割。前者基本思想是对原始点云进行网格结构化处理以利于利用成熟的2D卷积神经网络。如Su等^[6]利用传统CNN提取点云数据在不同视角下的二维投影图像特征,并聚合不同投影图像特征得到语义分割结果;Guerry等^[7]和Kalogerakis等^[8]则将投影图像标记结果反投影到原始点云上以实现多视图特征聚合与联合处理;Whu Y4^[9]和NANJ2^[10]设计CNN从投影图像提取多尺度局部特征并进行预分割,然后计算各点类别概率并构建决策树以引导进行再训练、优化初始分割结果。上述多视图投影方式虽然较好地实现了点云网格化处理,但破坏了点云的立体结构且会丢失大量空间细节信息。以VoxNet^[11]为代表的、基于体素的3D卷积神经网络虽能充分学习点云3D空间信息和上下文依赖关系,但体素方块太大丢失细节,太小则增加计算负担。总体上,相比于前者,直接点云分割方法信息利用更充分且易于捕捉细粒度局部特征用于精确分割目的。本文方法属于后者,这里对该类方法作进一步回顾。

由Qi等^[12]构建的PointNet开创性地解决了三维点云的无序性、置换不变性等问题,并通过最远点采样构建网络PointNet++^[13]成功解决大规模3D点云数据网络实施计算问题,但过于独立的点操作不利于捕捉局部空间结构。针对这一问题,后续研究提出一系列网络改进方案,包括:PointSIFT^[14]受SIFT算子启发在xyz空间8个方向上对最近点特征进行编码以解决PointNet++中K-邻域搜索局限于同一方向的问题;PointWeb^[15]通过自适应特征调整模块对局部点云信息进行聚合以实现特征强化;SSA-Pointnet++^[5]采用注意力机制强化网络对点云局部空间关系的精细学习,并通过差异性池化函数来整合注意力池化与最大池化提取的多个全局特征;RIT_1^[16]从地物空间分布连续特性出发,采用近似分块策略对训练/测试数据进行处理并通过最邻近内插得到待分割点所属类别;HDA-PointNet++^[17]和DPE^[18]通过分层数据

增强方式分别提升网络对大尺度类别特征、深层点特征的学习能力。总体上,上述算法通过点卷积聚集局部邻域点信息,再通过融合不同层次特征来捕获不同空间感受野的点云上下文信息,相比于原PointNet++均有所提升。但它们普遍采用的聚类最远点(farthest point sampling, FPS)是一种“无差别”点云采样方式,对于分布相对稀疏、样本数占比小的少数类点云而言,会使其类别样本愈加不平衡;同时,训练过程中无偏好的学习方法也会使网络倾向于牺牲少数类分割精度以保证多数类主导的全局分割精度。一些研究者尝试将关注重点信息的自注意力机制^[5]引入网络以增强其学习点云细粒度特征的能力,如文献[19]在图卷积网络中引入全局-局部注意力机制模块,根据相邻点空间位置动态学习卷积权重并利用局部区域密度对该权重进行加权更新以利于捕获多尺度特征;Hu等^[20]将自注意力机制与随机采样算法相结合设计网络RandLA-Net,通过增强从稀疏样本数据中充分学习目标空间结构特征的能力而有效应对大规模3D点云处理需求。其它网络模型或以投影到二维的感受野作为方向感知模式进行局部上下文编码^[21],或获取局部细粒度信息用于邻域空间关系学习^[22-23],或通过强化全局-局部特征关注捕捉多尺度特征^[24-26],虽可在某些类别目标分割方面显示出特定优势,但总体分割精度OA和平均 F_1 分数间难以同时取得较好的成绩。

现实世界采集的机载LiDAR点云数据通常表现出长尾(long-tailed)分布^[27],即少数几个类别(称为“头类”)包含大量样本,而大多数类别(称为“尾类”)只有非常少的样本。目前,PointNet++系列网络普遍采用FPS,这将导致信息损失并对点云局部精细结构的学习带来影响,尤其是分布相对稀疏、样本数占比小的尾类目标;此外,尾类样本也难以提供充分一致的信息用于CNN学习,在训练学习过程中往往被头类样本信息所淹没^[28]。针对上述问题,本文以PointNet++为基础,构建网络SE-PointNet++用于机载LiDAR点云高精度语义分割并能有效适应长尾分布,其创新之处在于两方面:1)设计FPS聚类采样过程,最大程度保留尾类样本并通过循环赋权方式使每类样本均能被网络充分学习;2)受文献[15、20]启发,引入自注意力机制并将其划分为采样点中心注意力机制和邻域注意力机制,并结合不同空间编码方式以增强网络从稀疏样本中学习类别目标空间结构的能力。

1 网络模型 SE-PointNet++

传统PointNet++系列网络在点云采样阶段并未充分考虑长尾分布问题,同时其训练过程倾向于牺牲少数类分割精度以保证多数类主导的全局分割精度,这不仅会影响到少数类目标的精确分割,一定程度上也会影响到

整体分割精度。本文旨在通过顾及长尾分布的 FPS 聚类点云采样和空间自注意力机制下的局部结构增强学习进一步改善 PointNet++点云语义分割性能。

1.1 FPS 聚类点云采样

为应对大规模复杂场景的点云数据处理需求, PointNet 系列网络模型通常先将大场景按平面位置划分成不同区域(格网),再利用 FPS 对划分区域内点云数据降采样并分批次输入网络模型进行处理^[13]。

令 M 、 N 分别表示划分区域内的原始点数与设定采样点数, PointNet 系列网络点云采样实际上分 3 种情形: 1) 当 M 远小于 N , 直接舍弃该划分区域及其包含的点云; 2) 当 M 大于 N 时, 利用 FPS 将区域内的样本点数稀释至 N 个; 3) 当 M 小于 N 且不符合情形 1) 时, 对子区域内的点云进行随机复制至 N 个。 PointNet 系列网络点云采样示意如图 1 所示, 其中实线方框表示划分区域, 虚线方框所在区域因其点数远小于设定的采样点数被舍弃, 圆点虚线方框所在区域因其点数大于设定采样点数将利用 FPS 进行稀释, 方点虚线方框所在区域则需通过点云随机复制以达到设定的采样点数。令 A 表示某一划分区域中原始点集, B 表示采样后的点集, FPS 实施过程总体可分为 3 步:

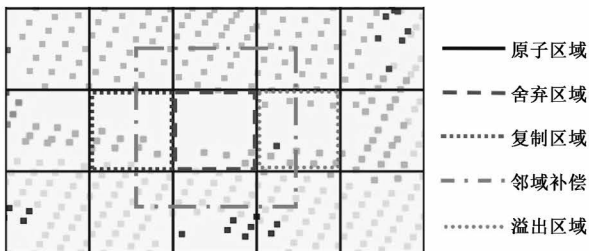


图 1 PointNet 系列网络模型点云采样示意

Fig. 1 Depict of point cloud sampling in PointNet series network models

1) 从 A 中随机选取某一点作为采样点并放入集合 B ;

2) 以集合 B 中点为参考点并计算它与 A 集合内剩余点的距离, 获得 A 中距离最远的点加入集合 B ;

3) 重复步骤 2) 直至集合 B 中采样点数达到给定阈值。因集合 B 中采样点数大于 1, 此时下一采样点选取过程与步骤 2) 略有不同, 具体为: ①对于集合 A 中任一点, 计算该点到集合 B 中每一采样点的距离并取距离最小值作为该点到参考点的距离; ②计算出集合 A 中每个点到参考点的距离, 取距离最大值所对应的点放入集合 B 中。

由图 1 可知, 对于分布相对稀疏、样本占比小的尾类目标, 由于其所在划分区域点较少, 网络模型将难以从该划分区域获得有效的训练样本, 丢弃该划分区域则给

尾类带来严重的样本信息损失; FPS 虽能保证采样点足够分散已覆盖整个空间, 并尽可能维持原始点云目标结构关系, 但该算法对点云的无差别采样处理会加剧不同类间的样本不平衡程度。此外, 采用随机复制方式扩充划分区域内点云数量破坏了划分区域内的类别(特征)多样性, 数据驱动的深度学习模型难以从有限、重复样本中学习目标本质特征, 存在特定类别(特征)学习过拟合风险^[29]。考虑到目前 PointNet 系列网络模型从点云数据中抽取特征、预测分割结果并据此构建目标函数进行训练学习的过程并未考虑到类别间的样本均衡问题, 本文对传统 PointNet 系列网络的点云采样过程重新进行设计, 通过等比例约束下的类内 FPS、基于置信度的 Meanshift 聚类及结合邻域补偿的划分区域 FPS 组合策略, 以增强网络模型对尾类样本的学习能力, 避免少数类样本信息在训练学习过程中被多数类样本信息淹没。有别于传统 PointNet 系列网络点云 FPS 处理, 这里将新点云采样过程称为 FPS 聚类采样, 如图 2 所示, 其实施主要涉及 3 方面内容:

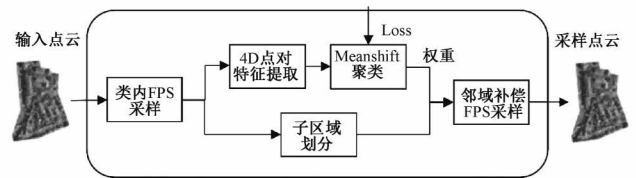


图 2 FPS 聚类采样过程示意

Fig. 2 Diagram of clustered FPS

1) 等比例约束下的类内 FPS。基本思想是利用 FPS 对头类点云数据(不划分区域)逐类别进行降采样以间接提高尾类样本占比, 主要通过约束头类的采样点数目来实现。令 N 表示样本点云总数, D 表示样本类别总数, 则理想情况下各类别样本均衡占比可简单视为 N/D , 从这一角度出发, 本文以单个类样本为对象进行点云采样并规定: 仅对样本占比超过 N/D 的多数类进行 FPS 处理且将其样本数稀释至 N/D 个, 其余类别样本则全部保留。显然, 由样本类别总数给出的等比例约束可有效降低头类的样本数, 而尾类的样本数则保持不变, 从而一定程度上起到均衡类别样本占比的效果。

2) 基于置信度的 Meanshift 聚类。由于 FPS 并不能改善点云数据长尾分布问题, 这里采用基于置信度的 Meanshift 聚类策略以削弱长尾分布影响, 基本思想是首先通过 4D 点对特征提取(point pair features, PPF)^[30] 获取点云局部几何结构信息, 得到包含邻域结构信息的中心点特征集合 $F_n = \{f_1, \dots, f_n\}$, 然后引入 Meanshift^[31] 对特征集合进行聚类, 当其均值漂移矢量收敛时输出各类别的聚类信息 $y_n = (f_{n,1}, \dots, f_{n,d})$, 其中 $d=1, 2, \dots, D$ 表示该特征所属类别。显然, 由于训练样本存在长尾分布

及边缘点特征相似性,直接由 Meanshift 给出的聚类信息将存在偏差,受文献[32-33]启发这里采取置信度函数对该聚类信息进行校准,以削弱网络模型对头类特征的学习偏好、促进对不同类别特征的充分学习,具体计算过程(伪码)如算法 1 所示。

算法 1 基于置信度的 Meanshift 聚类算法

输入: Meanshift 聚类信息 y_n , 点云数据集 X_n , 标签分布 $q_r(d|f)$, 最大批次(Max epoch)。

初始化: 权重 $\omega(i)$: for $i \in [n]$, $\omega(i) = 0$; epoch = 0.

1: 从 X_n 学习校准参数 α_d, β_d 和置信度得分函数 $\sigma(f)$ 。

2: **repeat**

3: 利用分类器头函数计算各点特征的类别得分:

$$h_d = h(f, y_n) \in R^K$$

4: 利用校准参数调整类别得分: $t_d = \alpha_d \times h_d + \beta_d$

5: 利用 $\sigma(f)$ 对 t_d 和 h_d 校准得到自适应类别得分函数:

$$h'_d = [1 + (\alpha_d - 1) \times \sigma(f)] \times h_d + \sigma(f) \beta_d$$

6: 估算总体预测分布:

$$q_m(d|f) = \text{Softmax}(h'_d) = \frac{\exp(h'_d)}{\sum_{k=1}^D \exp(h'_k)}$$

7: 计算基于 KL 散度的损失函数信息:

$$L_{KL}(q_r, q_m) = -1/N \sum_{i=1}^N [\sum_{d \in D} q_r(d|f_i) \log(q_m(d|f_i))] + D$$

8: **For** $i \in [n]$:

9: **If**

$$L_{KL}(q_{ri} | q_{mi}) = 0$$

11: $\omega(i) = 1$

12: **Else**

$$L_{KL}(q_{ri} | q_{mi}) \neq 0$$

14: $\omega(i) = 0.5$

15: **repeat**

16: 训练数据根据 $\omega(i)$ 大小顺序和邻域补偿下的划分区域 FPS 得到。

17: 网络对训练数据进行迭代训练。

18: 被学习过的点云 $\omega(i) = \omega(i) - 1$ 。

19: **until** $0 = \sum_{i=0}^n \omega(i)$

20: **until** epoch = Max epoch

3) 结合邻域补偿的划分区域 FPS。基本思想对未达到设定采样点数的划分区域利用其周围区域(邻域)点云样本进行补偿,以避免因丢弃该划分区域而使少数类样本信息发生严重损失及因采用随机复制方式扩充点云数量而导致目标特征学习过拟合。以图 1 中应舍弃区域(虚线方框)为例,令 M_i 表示该区域内原始点数,因 $M_i < N$,此时可以舍弃区域中心平面位置为基准逐步将虚

线方框按比例 $\lambda \in [1.2, 2.0]$ 扩大,直至扩大区域(点线方框)覆盖下的点云数不少于 N 。当扩大区域点云数大于 N 时仍采用 FPS 以符合预设采样点数,但与类内点云 FPS 过程略微不同的是,此时集合 B 中已包含应舍弃区域内全部点云,由邻域补偿获得的点则包含在集合 A 中,故可利用 FPS 将集合 A 中的样本点合理地补充到集合 B 。显然,相比于扩充舍弃区域样本点数的传统点云随机复制过程,这种邻域补偿方式下能最大限度地保留原始点云的空间结构、保证该划分区域内的样本类别(特征)多样性;当少数类位于划分区域边缘时,将能够有效避免此类数据被丢弃而发生的样本信息严重损失现象,从而促进模型对不同类别特征的充分学习,进而整体提高模型训练精度。

1.2 空间自注意力机制

尽管聚类 FPS 过程能一定程度上提高尾类样本占比,但由于绝对数量少,其点云采样后样本占比仍非常小;此外, FPS 会使划分区域中的点云数据变得稀疏,因破坏了点云间的邻接关系而影响到目标局部几何结构特征的完整表达。为使网络模型仍能从相对稀疏的点云数据中充分学习不同类别特征,这里采用空间自注意力机制来增强网络模型学习局部细粒度结构(特征)的能力。

近年来,深度学习网络模型尝试利用自注意力机制从输入数据中自主学习不同特征及其权重,并发展为用于增强点云语义分割网络模型捕捉细粒度局部特征能力的空间自注意力机制^[20]。目前,空间自注意力机制或倾向学习以采样点为中心的局部空间结构特征而忽视对其邻域点间空间结构关系的学习^[20],或虽然同时考虑到了对两者空间结构关系的学习,但邻域点间相关性的简单表示不足以描述复杂点间关系^[5]。对于稀疏点云数据,尤其是样本占比很小的少数类,其邻域点相互结构关系的学习不应被忽视。其原因在于,当场景感受野较小时,来自同一目标结构的采样点邻域点间同样存在有助于网络模型更好地理解该目标结构特征的、较强的几何结构关系。受文献[5]启发并从完整学习点云局部空间关系角度出发,本文将采样点中心、邻域自注意力机制引入 PointNet++ 基础框架,并采用更为新颖的邻域点间关系学习方式突显采样点邻域几何结构(特征)的细粒度,最后集合两者信息来构建局部空间特征编码模块。

对于 FPS 聚类采样下的点云 $O = \{f_i | i=1, \dots, N\}$, 其中: N 为点云个数, $f_i = [u_i, v_i] \in R^{3+d}$ 表示坐标与属性特征的组合,利用近邻点采样算法(k-nearest neighbor, KNN)获取采样点的 K 个邻域点后^[34],本文空间自注意力机制下的点云局部特征学习过程如图 3 所示,具体实施如下:

1) 采样点中心自注意力机制。与文献[5]保持一致,旨在通过欧氏距离、方向矢量及各点空间位置表达采

样(中心)点与其邻域点之间的几何关系,其空间编码公式如下所示:

$$\varphi_i^k = M(\mathbf{u}_i \oplus \mathbf{u}_i^k \oplus \sqrt{(\mathbf{u}_i - \mathbf{u}_i^k)^2} \oplus (\mathbf{u}_i - \mathbf{u}_i^k), (k \in [0, K])) \quad (1)$$

式中: $\mathbf{u}_i (i \in N)$ 和 \mathbf{u}_i^k 分别表示采样(中心)点及其邻域点的空间坐标, \oplus 为拼接符号, $\sqrt{(\mathbf{u}_i - \mathbf{u}_i^k)^2}$ 表示采样点与邻域点之间的欧氏距离, $(\mathbf{u}_i - \mathbf{u}_i^k)$ 表示相对位移(方向)矢量, $M(\cdot)$ 表示利用多层感知器(multilayer perceptron, MLP)将多个空间信息映射成特征矢量 φ_i^k 。进一步的,按式(2)计算特征 φ_i^k 的自注意力得分并聚合:

$$F(\mathbf{u}_i^k) = g(\varphi_i^k) \times \left[\text{softmax} \left(\frac{g(\varphi_i^k)^T \times g(\varphi_i^k)}{\sqrt{C}} \right) + 1 \right] \quad (2)$$

式中: $g(\cdot)$ 表示线性变换操作(这里为 1×1 的卷积操作), C 代表输出通道数。图3给出了采样点中心自注意力机制下的局部空间位置编码过程示意,该空间编码同时考虑了邻域点相对采样中心点的偏移量和偏移方向,从而能准确构建采样中心点与其邻域点间的空间关系。

2) 采样点邻域自注意力机制。邻域自注意力机制旨在观察采样点 K 个邻域点间的相互影响关系,受文献[15]启发,采用以下公式进行空间编码:

$$\gamma_i^k = g(\mathbf{u}_i^k) + \sum_{j=1, j \neq i}^N [-M(\mathbf{u}_i^k - \mathbf{u}_j^k)] (\mathbf{u}_j^k - \mathbf{u}_i^k), (k \in [0, K]) \quad (3)$$

式中: $g(\cdot)$ 为线性变换操作, $(\mathbf{u}_i^k - \mathbf{u}_j^k)$ 与 $(\mathbf{u}_j^k - \mathbf{u}_i^k)$ 分别表示邻域点 i 与邻域点集中其他点之间的相互位移矢量, 可视为邻域点间的正向、反向作用力, 用于增强采样点邻域拓扑结构关系表达。图3给出了采样点邻域自注意力机制下的局部空间位置编码过程示意, 其输出维度为3。需要指出的是, 本文采样点邻域自注意力机制与文献[5]的区别在于 SSA-PointNet++ 仅考虑各邻域点与最小值点的位置差异性, 其单向相关性不足以表示邻域空间内复杂的点间关系; 而本文方法通过 MLP 将邻域点间相互位移矢量映射成特征矢量, 并将其赋值给自身的编码方式能更细致地表达采样点邻域拓扑结构。

级联两种局部空间编码方式(过程)即可构建本文空间自注意力机制下的点云局部特征学习模块, 其输出结果 $S(\mathbf{u}_i)$ 隐含表达了采样中心所在邻域的局部空间几何结构。需要指出的是, 就类样本不均衡点云数据而言, PointNet++ 网络模型在卷积核参数一致情况下, 由于头类目标样本占比大, 其在卷积层叠加过程中贡献会明显多于尾类目标, 极有可能导致尾类目标信息被头类信息淹没, 尤其是与头类目标邻接的边缘区域。理论上, 本文空间自注意力机制可同时作用于点云数据中不同类样本, 但采样点邻域自注意力机制的构建更有利于从稀疏的少

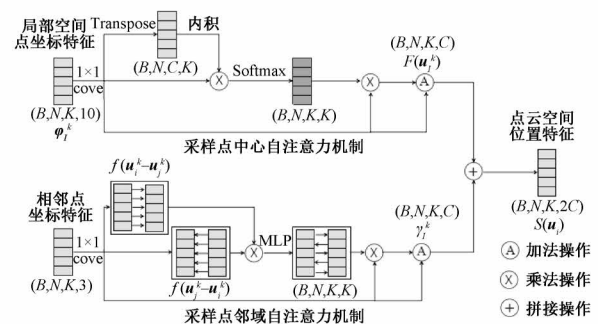


图3 空间自注意力机制下的点云局部特征学习示意
Fig. 3 Diagram of point clouds Local feature learning under spatial self-attention mechanism

数类样本数据中完整学习对应目标的空间分布结构, 从而提高网络模型对少数类目标的学习能力, 有效避免网络模型为保持全局最优而牺牲少数类语义分割精度。

1.3 网络结构设计

本文网络模型以 PointNet++ 为基础框架构建, 取名为 SE-PointNet++, 其结构设计如图4所示, 主要有 FPS 聚类采样、下采样、上采样和全连接4部分构成。每个下采样部分均包含空间自注意力层和最大池化层, 空间自注意力层在关联各邻域点的同时维护点云数据几何轮廓, 促进局部点云特征被网络充分学习; 最大池化层在有效降低计算成本的同时提高空间自注意力机制影响范围, 以提升网络对总体结构的理解能力; 上采样层的跳跃连接和3D-插值使网络能够综合考虑不同层次的特征表示; 最后通过全连接层实现点云语义分割。为防止过拟合, 批标准化和修正线性单元(rectified linear unit, ReLU) 激活函数内嵌于每一层, 且各全连接层都添加了下降速度为0.4的随机失活(dropout)层。

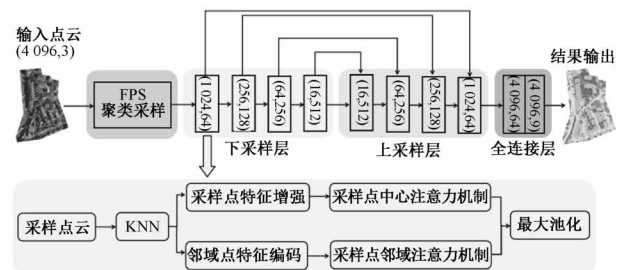


图4 SE-PointNet++点云语义分割网络结构示意图
Fig. 4 Depict of SE-PointNet++ network architecture

为能输出更大感受野下的全局特征向量, 现有网络通常使用池化操作来集成邻域特征以确保输出特征具有平移不变性^[13]。需要指出的是, 传统最大池化的信息传递主要取决于池化窗口内特征值大小, 受文献[5]启发, 本文给出一种空间自注意力机制下的最大池化操作以确

保表征细粒度空间结构特征的重要信息尽可能得到传递,从而为降低少数类样本特征在信息传输过程中的损失。如图 5 所示,采用式(4)将点云空间位置特征与其对应的属性特征进行 MLP 映射与拼接,经多层卷积后即可输出如下最大池化特征:

$$\text{MaxPool}(f_i) = \max_k (M(S(\mathbf{u}_i^k \oplus \mathbf{v}_i^k) \oplus M(f_i))),$$

$$(k \in [0, K]) \quad (4)$$

式中: \max 表示最大池化操作, $S(\cdot)$ 为空间注意力机制输出结果, \mathbf{u}_i^k 和 \mathbf{v}_i^k 分别表示采样中心点的空间位置特征和属性特征。

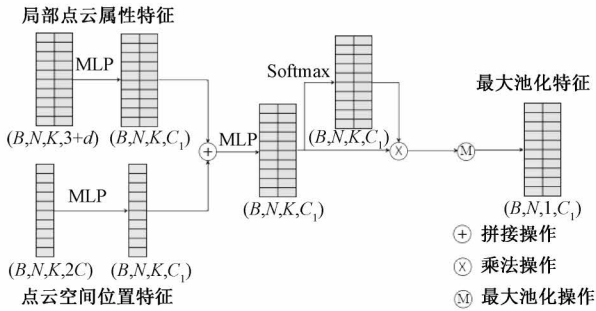


图 5 空间自注意力机制下的最大池化示意

Fig. 5 Diagram of Max pooling under spatial self-attention mechanism

2 实验与分析

本文网络实施计算环境配置为: GPU TITAN XP 12 G 的深度学习工作站, 工作站系统为 ubuntu18.04, 显卡驱动版本 cuda10.0。各项参数指标为: 批量规格设为 10, 动量值取 0.9, 采用十万个衰减周期和 0.96 衰减率, 优化器选用 Adam, 将 0.001 作为初始学习率。本文网络模型性能测试采用 F_1 分数和总体分割精度 OA , 具体公式如下:

$$F_1 = \frac{2p_{ii}}{\sum_{j=0}^k (p_{ij} + p_{ji})}, OA = \frac{p_{ii}}{p} \quad (5)$$

表 1 训练数据类别样本点数及占比信息

Table 1 Training data class points and proportion

类别	电力线	车辆	树篱	建筑立面	灌木丛	树木	建筑顶面	低矮植被	道路表面	总和
点云数量	546	4 614	12 070	27 250	47 605	135 173	152 045	180 850	193 723	753 876
百分比/%	0.072	0.612	1.601	3.615	6.315	17.931	20.168	23.989	25.697	100

2.1 模块增益分析

根据 1.1 节首先实施等比例约束下的类内点云 FPS。这里根据 Vaihingen 3D 点云数据类别总数 $D=9$, 将类内点云 FPS 采样点数设为总点数的 $1/9$, 类内点云

式中: p_{ji} 表示属于类 j 但被预测为类 i 的点云数量, 而 p_{ij} 表示属于类 i 但被预测为类 j 的点云数量, p_{ii} 表示所有正确预测的类的数量, F_1 分数定义为精准率和召回率的调和平均数, F_1 的值域属于 $0 \sim 1$ 的范围, F_1 越趋近于 1 则 i 类的分割效果越好; OA 为模型的总体分割评估指标, 即标记正确的点云个数与点云总数的比例, p 表示点云模型的所有点数, OA 值越大, 分割效果越好。

本文网络模型训练与测试采用国际摄影测量与遥感学会(International Society for Photogrammetry and Remote Sensing, ISPRS) Vaihingen 3D 语义标记竞赛标准点云数据。2008 年 8 月该数据集通过 ALS50 系统在 500 m 航高、 45° 视场角下采集的。该数据集具有较高的点云密度, 每平方米内有 4~8 个点云, 同时各类地物被人工分割为 9 类, 包括建筑外表面、道路表面、灌木丛等, 如图 6 所示, 其中训练数据由 5 个子区域拼接而成, 测试数据包含商品房区域与别墅房区域。训练数据中各类地物点数及占比信息如表 1 所示, 由表 1 可看出, 该训练数据类别样本不平衡特点显著, 其中树木、建筑顶面、低矮植被以及道路表面类别总点数占比达到 87.785%, 而电力线、车辆和树篱等少数类样本占比分别为 0.072%、0.612% 和 1.601%, 远低于其它类别。因此, 对现有 PointNet++ 网络模型进行改进以使其适应点云数据长尾分布, 在保证头类目标分割性能的同时尽可能提高尾类目标分割精度, 将具有实际意义。

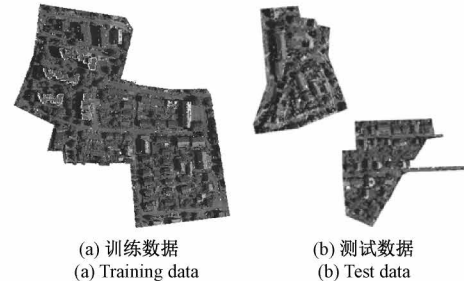


图 6 LiDAR 实验数据

Fig. 6 Experimental data of LiDAR

FPS 前后的测试点云数据点数统计如表 2 所示, 类内点云 FPS 前后的各类地物样本占比示意及其方差统计结果如图 7 所示。由表 2 可知, 等比例约束下的类内点云 FPS 使样本点数从采样前的 753 876 个大幅减少为采样

后的 427 141 个(约减少 40%),少数类样本不平衡问题也得到明显改善,其原因在于:类内点云 FPS 极大地稀释了低矮植被、建筑顶面和道路表面等多数类样本点数,并适当地削减了树木点数,由于减少训练样本总点数的同时少数类样本点数保持不变,从而间接使其样本占比获

得不同程度提升,如图 7 所示。若以类别样本占比方差来表征样本总体平衡状况,则由图 7 可看出,各类地物样本比例方差由类内点云 FPS 前的 1.01% 减少为类内点云 FPS 前的 0.668%,证明了等比例约束下的类内点云 FPS 对于改善类别样本占比不平衡的有效性。

表 2 类内点云 FPS 前后的测试点云数据点数统计

Table 2 Statistics of number of point clouds before and after intra-class FPS

类别	电力线	车辆	树篱	建筑立面	灌木丛	树木	建筑顶面	低矮植被	道路表面	总和
采样前点数	546	4 614	12 070	27 250	47 605	135 173	152 045	180 850	193 723	753 876
采样后点数	546	4 614	12 070	27 250	47 605	83 764	83 764	83 764	83 764	427 141

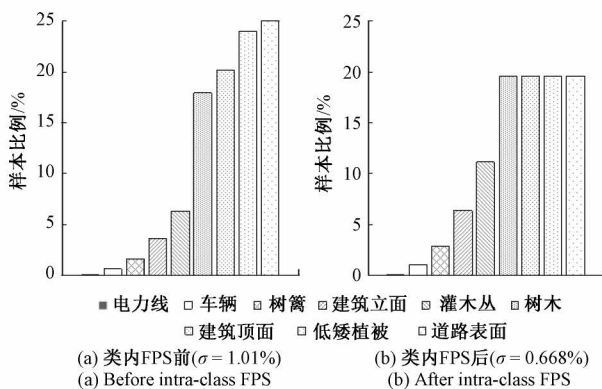


图 7 类内点云 FPS 前后各类别样本占比示意及其方差统计
Fig. 7 Depict of sample proportion of each category and its variance statistic before and after intra-class FPS

进一步的,按平面位置对完成类内 FPS 的采样点云进行区域划分并对其进行邻域补偿下的 FPS,这里划分区域设为 $10\text{ m} \times 10\text{ m}$ 大小的正方形,区域内采样点个数设为 1 024,除去空区域(域内无点云的子区域)共计可得 1 000 个划分区域。1 000 个划分区域 FPS 后各地物类别的样本总体占比及其方差如图 8 所示,其中图 8(a)为传统 PointNet++ 直接对划分区域实施 FPS 后的统计结果,图 8(b)为本文邻域补偿下的划分区域 FPS 统计结果。由图 8 可以看出,相对于传统 PointNet++,本文邻域补偿下的划分区域 FPS 更有助于改善少数类样本占比不平衡问题,体现为多数类样本占比进一步减小,少数类样本占比则有所增加,相应的,类别样本占比方差也更小,证明了邻域补偿对于改善类别样本占比不平衡的有效性。有趣的是,对比图 8(a)和图 7(b)可发现,前者各地物类别样本占比方差(0.731%)大于后者(0.668%),这表明传统 PointNet++ 直接对划分区域实施 FPS 会加剧少数类样本占比的不平衡性,其原因在于 PointNet++ 对划分区域实施 FPS 时,采用点云随机复制方式对点数极少的区域进行扩充会因道路表面、低矮植被、建筑顶面等多

数类目标在空间分布中占主要地位而增大其点云被复制的概率,使得多数类与少数类地物样本占比间的差距被拉大。相反的,对比图 8(b)和图 7(b)可发现,前者各地物类别样本占比方差(0.599%)更小,这表明顾及目标空间分布的邻域补偿策略可避免上述点云随机复制方式之不足,通过确保位于划分区域边缘的少数类目标点云获得“合理”扩充而能有效维护划分区域 FPS 前后的样本占比一致性,甚至于进一步改善车辆、树篱、建筑立面等少数类目标样本占比不平衡性。

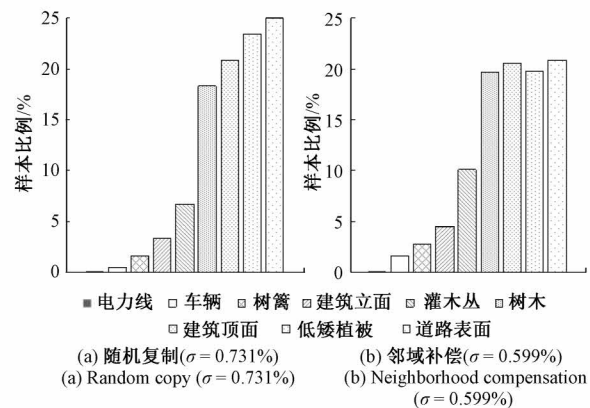


图 8 划分区域 FPS 后各类别样本占比示意及其方差统计
Fig. 8 Depict of sample proportion of each category and its variance statistic after inter-block FPS

总体上,等比例约束下的类内 FPS 和邻域补偿下的划分区域 FPS 虽均能一定程度上提高少数类样本占比,但多数类样本仍占据绝对数量,可进一步通过基于置信度的 Meanshift 聚类对点云数据进行权重赋值、按权重大小顺序进行邻域补偿采样来削弱多数类样本对网络模型学习偏好的影响。通过 Meanshift 聚类输出的错误分类点(标记为白色)分布图及 3 个局部区域(白方框、黑方框和白色虚线方框)放大显示的结果如图 9 所示,其中第 1 行为循环赋权下的 Meanshift 聚类最后结果,第 6 行为 3 个局部区域的真实标签(ground truth),第 2~5 行为不

同次数迭代(赋权)时的 Meanshift 聚类结果。由图 9 中 3 个局部错误分布图可看出,初次 Meanshift 聚类能较好地识别道路表面、车辆、建筑顶面和树木类别,而灌木丛、树篱、建筑立面和低矮植被之间则存在大量错误分类点,其原因在于道路表面、车辆、建筑顶面和树木虽然相互交错堆叠,但是它们的特征差异较大且易于分辨;灌木丛、树篱和低矮植被之间不仅互相包含、堆叠且三者之间的特征相似度较高,建筑立面与顶部衔接区域的特征相似性极高,建筑立面的底部区域与周围环境(灌木丛,树篱和低矮植被)相结合也使得它们之间的特征难以区分,最终导致这 4 类目标存在大量错误分类点。然而,随着循环赋权下的 Meanshift 聚类次数增加,大量错误分类点获得了“纠正”,其原因在于,对分类正确的点赋予更高的权重值有助于其优先被网络选中进行学习,同时权重值高的点每被选中一次就会降低权重值的处理方式也确保了不忽视对其它低权重值点的学习,从而提高网络的总体分割精度。本文一个权重周期(所有点云权重从非 0 值到 0 值)内各批次(epoch)的训练精度变化情况如表 3 所示,由该表可看出,循环赋权方式使得每类样本均能被网络充分学习,不仅头类目标(建筑顶面、树木、低矮植被和道路表面)的分割精度不断提高,尾类目标(电力线、车辆、灌木丛、树篱)整体分割精度也获得提升,从而缓解点云数据长尾分布特性所带来的网络模型侧重头类目标学习而忽略尾类目标的问题。

为了进一步验证所提模块的有效性,本文在 PointNet++ 基础上设计了样本均衡化处理(+SE)、空间自注意力机制(+ATT)等不同模块组合下的消融实验,以分析各模块对测试区域分割精度的直接影响。不同模块组合下的分割结果如表 4 所示,其中: D_1 为原 PointNet++ 网络模型, D_2 为仅以+SE 替换 D_1 中点云 FPS 采样及随机置换批次组合获得的网络模型, D_3 为在 D_1 中仅添加+ATT 获得的网络模型, D_4 为本文 SE-PointNet++ 网络模型。由表 4

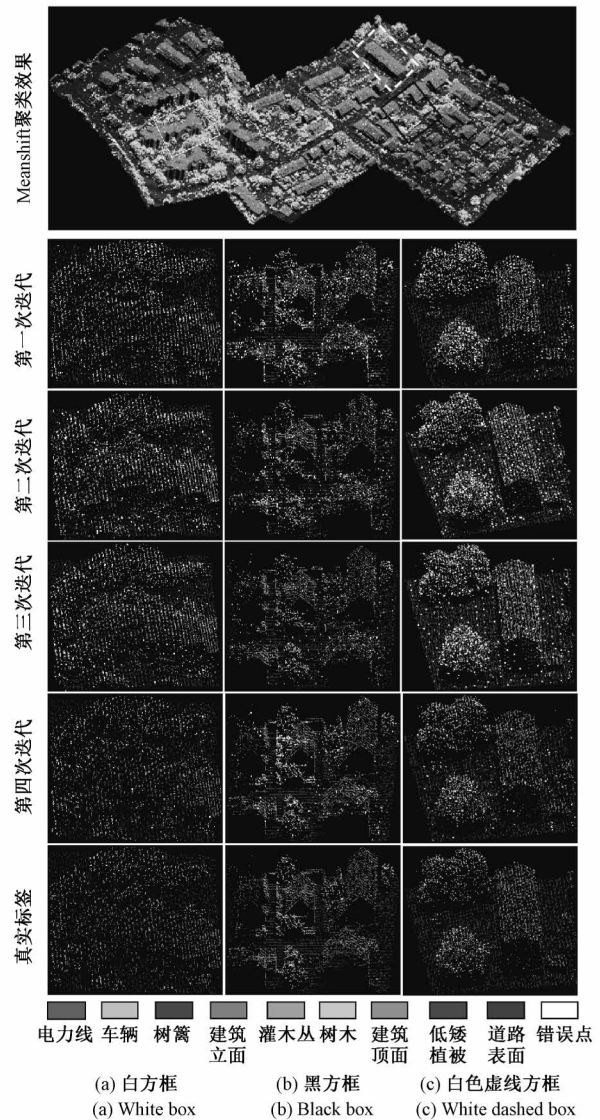


图 9 循环赋权下的 Meanshift 聚类错误点变化情况
Fig. 9 Depict of Meanshift clustering error change under iterative weighting

表 3 权重周期内各 epoch 训练精度结果统计

Table 3 Statistics of training accuracy of each epoch within an weight cycle

Epoch	电力线	车辆	树篱	建筑立面	灌木丛	树木	建筑顶面	低矮植被	道路表面	OA	平均 F_1
1	19.6	53.3	11.5	73.6	48.3	84.6	92.6	81.0	89.5	83.5	61.6
2	42.4	65.0	31.4	72.1	45.9	87.9	92.6	83.9	91.0	85.2	64.5
3	50.6	68.0	18.0	79.3	48.5	88.2	95.0	85.2	91.0	86.3	69.3
4	58.9	71.2	33.8	80.3	50.7	88.9	95.6	85.3	91.8	87.0	71.5

可知,除电力线外,设计模块的引入均能不同程度提高类别 F_1 分数,其中: D_2 在提升车辆、建筑立面、树篱等少数类目标分割精度方面效果显著并优于 D_1 和 D_3 ,而 D_3 则在提升建筑顶面、灌木丛、树木、低矮植被、道路表面等多数类目标分割精度方面效果显著并优于 D_1 和 D_2 , D_2 和

D_3 两者叠加还可进一步提升网络模型点云语义分割整体性能,使得 OA 和平均 F_1 分别提升 6.3% 和 6.6%。不同模块组合下的地物类别分割错误结果如图 10 所示,其中:图 10(a) 为测试数据类别标签(ground truth),图 10(b) 为不同模块组合下的地物类别分割错误局部

表4 模块消融实验分割精度统计

Table 4 Statistics of segmentation accuracy of Module ablation experiment

%

方法	电力线	车辆	树篱	建筑立面	灌木丛	树木	建筑顶面	低矮植被	道路表面	OA	平均 F_1
D_1	57.9	66.1	31.5	54.3	41.6	77.0	91.6	79.6	90.6	81.2	65.6
D_2	40.6	75.8	34.1	59.9	44.8	80.2	94.4	76.6	91.5	82.2	66.3
D_3	51.3	66.5	32.9	58.4	49.1	82.9	94.7	81.2	92.2	86.0	67.8
D_4	69.6	77.4	35.8	58.1	50.1	86.6	95.3	83.7	93.6	87.5	72.2

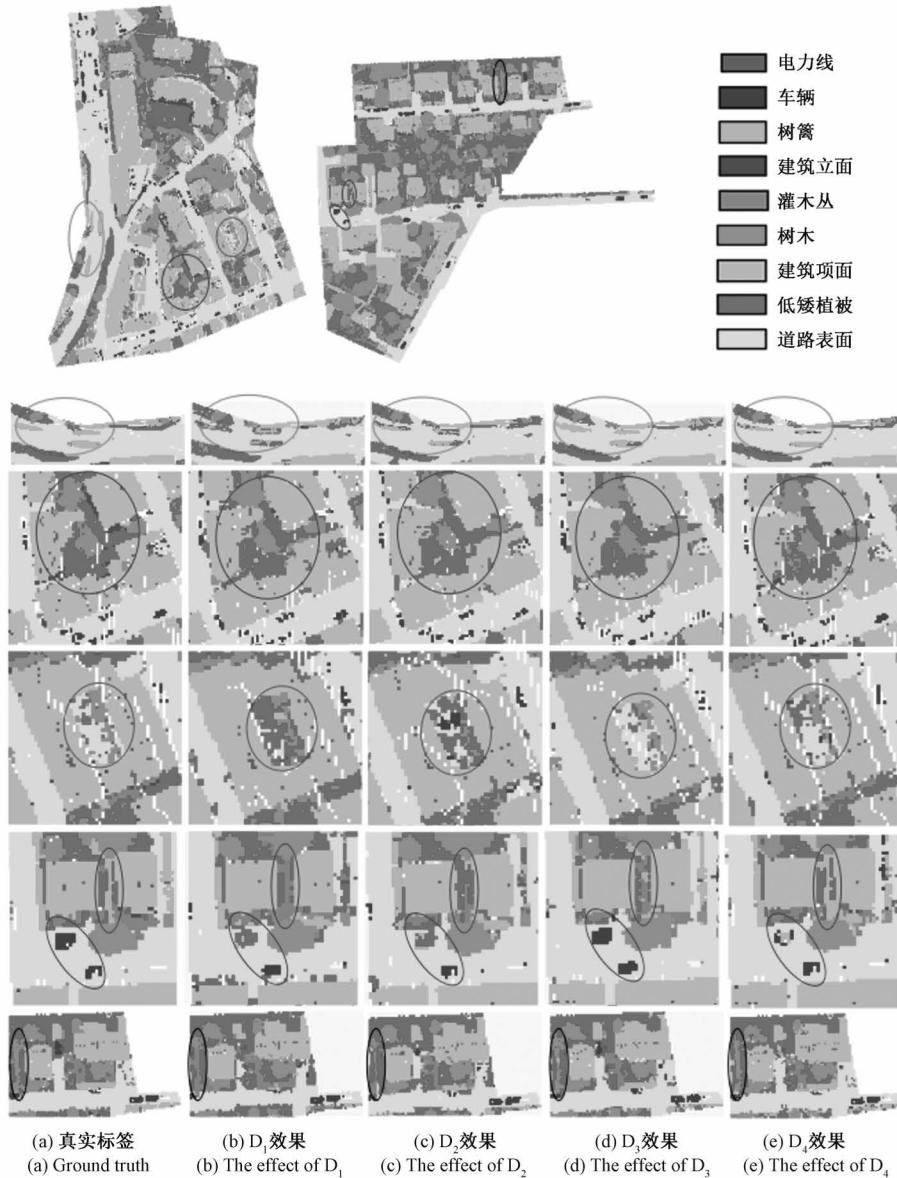


图10 模块消融实验错误分割点分布图

Fig. 10 Depict of wrongly segmented points from module ablation experiment

放大示意。结合表4中九类地物 F_1 分数和图10中地物类别分割错误可看出,无论是引入+SE还是+ATT,均能有效提升原 PointNet++语义分割整体性能,但具体到某

一地物类别时,则会因其点云空间分布特性及样本占比而呈现出不同分割效果:

1) D_4 的车辆和电力线 F_1 分数提升幅度最大(约

11.3%~11.7%)。观察图 10(b) 中第 4 行图片可发现,斜框内的车辆在 D_1 中完全被分割为低矮植被与树篱,单纯引入+ATT 可大幅提高车辆点分割精度,但单纯引入+SE 未能明显改善车辆点分割结果,其原因在于 D_3 虽能正确分割车辆但其部分边缘被误分割为道路表面, D_4 则可对车辆边缘进行了正确分割,虽仍将内部少量点误分割为树篱,但车辆整体分割精度获得了进一步提升并优于 D_2 。另一方面,观察图 10(b) 中第 2、3、4 行图片可发现,建筑顶面上零星分布的电力线点唯有 D_4 最接近真实情况,其中 D_1 在图 10(b) 第 2 行图片中将部分建筑顶面误分割为电力线点, D_2 在图 10(b) 第 3 行图片中将部分电力线点误分割为建筑顶面点, D_4 在图 10(b) 第 3 行图片中对电力线的分割与真实标签一致。这表明,在网络全局特征学习能力方面,+ATT 与+SE 相结合优于其中任意单个模块的引入。

2) 树篱、灌木丛、低矮植及树木类别间的点云特征相似性高,易存在误分割且存在少数类样本占比不均衡问题,本文网络模型可显著提升对这些相近地物类别的区分能力。观察图 10(b) 第 1 行框内可发现, D_1 在树篱与低矮植被间存在大量误分割,+SE 的引入使得这一状况得到了有效改善,+ATT 的引入也减少了树篱与低矮植被间的误分割;观察 10(b) 第 4 行图片竖框内可发现,树篱在 D_1 中完全被分割为灌木丛,+SE 的引入使得部分树篱被分割正确但依然存在被误分割为低矮植被与灌木丛的现象,+ATT 的引入则进一步提升被正确分割的树篱点数;观察 10(b) 第 5 行图片横框可发现, D_1 仅有小部分树篱点分割正确, D_2 的分割正确率在+SE 作用下得到一定程度提升, D_3 在+ATT 的作用下仅树篱与灌木丛衔接处出现误分割, D_4 则进一步提升衔接处的分割正确率;类似的,通过观察竖框内树木点可发现,其几何结构与相邻灌木丛极为相似, D_1 与 D_2 完全将树木分割成了低矮植被与灌木丛, D_3 可以正确分割部分树木点, D_4 分割效果基本与标签结果相吻合。这表明,+SE 可有效处理不均衡样本中少数类信息难以被捕捉和学习的问题,而+ATT 能显著增强模型对目标局部特征的描述能力,从而有效提升了网络对上述相近地物特征的区分与学习能力。

3) D_4 在建筑顶面、道路表面的 F_1 分数相比 D_1 提升幅度不大(3%~4%),其根本原因在于两类别样本点云占比较大且与其它类别具有较高区分度, D_1 已可取得较好的 F_1 分数(分别为 91.6%和 90.6%)。观察图 10(b) 中第 2、3 行图片可看出, D_1 与 D_2 将框内的建筑立面误分割为建筑顶面,同时低矮植被、树木与建筑顶面之间也存在误分割现象,而 D_3 能较好地区分低矮植被与建筑顶面, D_4 在 D_3 基础上则能分割出部分建筑立面; D_1 将框内道路表面误分割为低矮植被, D_2 将部分道路表面误分割

为低矮植被与车辆, D_3 与 D_4 则能对空间分布区域存在较大程度重叠的类别进行正确分割。这表明,当不同地物类别聚集于局部重叠区域时,本文网络模型能较好地捕捉各类地物特征并进行分割,从而进一步提升其类别分割精度;另一方面,由于有效减少建筑顶面与建筑立面、道路表面与低矮植被间的误分割, D_4 在建筑立面、低矮植被的 F_1 分数也获得了明显提升(4%)。

总体上,+SE 和+ATT 可从不同角度改善网络模型对不均衡样本的特征学习偏好,且两者结合能够显著提升点云整体分割精度。

2.2 分割性能对比

选取 ISPRS 官网及近期刊物中名列前茅的 10 种网络模型进行对比,本文方法(D_4)及 10 种网络模型在测试数据集上的总体分割精度 OA 及各类地物 F_1 分数如表 5 所示,由该表可看出,本文方法树木、建筑顶面和道路表面类别 F_1 分数居首,分别为 86.6%、95.3%和 93.6%,高于第 2 名分别为 3.5%、0.9%和 0.8%;低矮植被、灌木丛、车辆和电力线类别 F_1 分数居第 2 名,分别为 83.7%、50.1%、77.4%和 69.6%,低于第 1 名分别为 5.1%、5.8%、0.7%和 0.8%;建筑立面类别 F_1 分数居第 5 名,处于所列文献网络模型中等水平,仅树篱类别 F_1 分数居第 7 名,处于所列文献网络模型中等以下水平。通常,由于 Training 数据集上存在类别点数极不均匀,难以同时获得最佳 OA 和平均 F_1 分数,原因在于,网络模型在训练过程倾向于牺牲少数类目标以获得整体最优分割精度 OA ,从而会降低平均 F_1 分数,而提高 F_1 平均分数需要网络模型充分学习少数类样本特征,会因模型过拟合而限制整体分割精度,如 PointConv^[25] 虽在部分少数类地物(电力线、车辆)获得较好的分割精度,但过度关注少数类样本特征导致模型陷入过拟合状态,使得 OA 仅有 79.6%;LUH^[35] 虽在特征区分度较高的地物(树木、建筑顶面和道路表面)具有良好的分割精度,但特征学习模块缺乏局部细粒度特征的学习能力,其 OA 和平均 F_1 分数仅有 81.6%和 68.4%;DGCNN^[21] 的特征编码模式过于单一,虽考虑了全局-局部特征的融合,但对于大规模复杂城市场景存在限制,其 OA 和平均 F_1 分数仅有 81%和 64.1%;D-FCN^[24] 以投影到二维的感受野作为方向感知模式进行局部上下文编码,虽然获得了较高的平均 F_1 分数(70.7%)但却抑制了 OA (82.2%);WhuY4^[9] 和 NANJ2^[10] 将三维点云投影到二维图像过程中丢失了局部空间结构信息,使得平均 F_1 分数与 OA 均难以进一步提升;PointCNN^[23] 虽能充分学习采样点局部空间特征,但忽略了长尾分布现象下的少数类信息缺失问题,也不能在 OA (83.3%)与平均 F_1 分数(69.5%)间取得平衡;DANCE-Net^[26] 考虑到了全局-局部特征的重要性,但其

表5 本文及文献方法 OA 及 F_1 分数统计
Table 5 Statistics of OA and F_1 score in this paper and literature

方法	电力线	车辆	树篱	建筑立面	灌木丛	树木	建筑顶面	低矮植被	道路表面	OA	平均 F_1
DGCNN	67.6	54.5	26.8	48.8	41.5	77.3	89.8	80.4	90.6	81.0	64.1
PointConv	65.5	72.1	25.0	54.2	45.6	75.8	90.5	79.9	88.5	79.6	66.3
LUH	59.6	73.1	34.0	56.3	46.6	83.1	94.2	77.5	91.1	81.6	68.4
KPConv	63.1	72.5	25.2	60.3	44.9	81.2	94.4	82.3	91.4	83.7	68.4
WhuY4	42.5	74.7	53.7	53.1	47.9	82.8	94.3	82.7	91.4	84.9	69.2
NANJ2	62.0	66.7	40.7	42.6	55.9	82.6	93.6	88.8	91.2	85.2	69.3
PointCNN	61.5	75.8	35.9	57.8	49.1	78.1	92.7	82.7	91.8	83.3	69.5
D-FCN	70.4	78.1	37.0	60.5	46.0	79.4	93.0	80.2	91.4	82.2	70.7
Randla-Net	68.8	76.6	43.8	61.9	45.2	77.4	91.1	82.1	91.3	82.1	70.9
DANCE-Net	68.4	77.2	38.6	60.2	47.2	81.4	93.9	81.6	92.8	83.9	71.2
本文(D ₄)	69.6	77.4	35.8	58.1	50.1	86.6	95.3	83.7	93.6	87.5	72.2

密度感知卷积模块不能很好地区分特征相似性与重叠度较高的树篱与灌木丛类别,未能在 OA (83.9%) 与平均 F_1 分数 (71.2%) 取得进一步突破。本文采样点中心注意力机制空间编码灵感源于 Randla-Net^[20] 并在此基础上进一步提出采样点邻域注意力机制,对比表5中各项分割指标可看出,Randla-Net 采用的随机采样策略虽能提升网络对少数类样本特征的捕捉能力,但不利于该网络模型对多数类样本特征的充分学习,尤其是道路表面、灌木丛、树木与低矮植被这4类存在空间重叠、特征相似度高的地物,从而一定程度上损失总体分割精度,仅为82.1%,处于表5中所列文献网络模型中等以下水平(后四),与本文相差约5%。总体上,本文通过+ATT对点云局部空间关系及上下文特征的充分学习,提高了网络模型对相似地物类别的辨识度;+SE的引入则兼顾到了少数类样本特点,从而提高了网络模型对各类地物的特征提取与学习效率,两者使得本文模型在 OA 和平均 F_1 分数间取得了令人满意的平衡。

本文网络模型与其它6种以 PointNet++ 为基础框架构建的网络模型作对比,包括 HDA-PointNet++^[17]、RIT_1^[16]、DPE^[18]、SSA-Pointnet++^[5]、PointSift^[14] 和 PointWeb^[15],6种网络模型在测试数据集上的 OA 及各类地物 F_1 分数如表6所示。由表6可以看出,HDA-PointNet++ 在部分多数类地物如低矮植被 (85.1%)、道路表面 (99.2%) 上表现出较好结果,但该网络分层数据增强方式不利于学习全局特征信息,其 OA 和平均 F_1 分数没有竞争力;DPE 与前者类似,以牺牲他类地物分割精度来提升道路表面和低矮植被分割精度,在 OA 与平均 F_1 分数上有待提升;RIT_1 对于特征区分明显的地物如车辆 (73.4%)、建筑顶面 (94%) 和道路表面 (91.5%) 具

有良好的分割效果,但缺乏有效手段处理存在空间重叠、特征相似度高的地物,故模型在 OA 和平均 F_1 分数表现不佳;PointSIFT 采用了与其他 PointNet 系列网络不同的 xyz 邻域点搜索方式,利于从稀疏邻域点中学习局部空间结构,在车辆 (77.8%)、电力线 (55.7%)、建筑立面 (56.9%) 等少数类地物分割中具有优势,但由于缺乏对采样中心点-邻域点间的空间关系的精细学习,该网络在 OA 与平均 F_1 分数仍有提升空间;PointWeb 和 SSA-Pointnet++ 在网络设计中均采用了注意力机制策略,以增强网络对点云局部空间关系的精细学习能力,但由于忽略了少数类样本占比不均衡问题(直接利用原 PointNet 中 FPS) 导致相应地物分割不具优势,两种模型平均 F_1 分数均不高 (<68%),同时由于采样点邻域空间编码时对采样中心点与邻域点、邻域点与邻域点间空间关系学习的侧重点不同,两种模型的 OA 仍存在提升空间。总体上,对比表5中 D₄ 可以看出,本文网络整体分割性能优于表6中6种 PointNet++ 系列模型,在 OA 和平均 F_1 分数方面均达最高,其原因主要在于两方面:1) 点云采样虽利于网络语义分割计算实施,但其对 LiDAR 点云密度、空间分布的潜在改变无疑会影响后续对各类地物类别特征的充分学习,本文+SE 顾及到了点云采样对类别样本不均衡的影响,在保证多数类地物分割结果的同时增强网络对少数类地物分割精度,从而能进一步提升平均 F_1 分数;2) 本文+ATT 实施充分考虑到了采样中心点与邻域点、邻域点与邻域点间的空间关系表达差异,并据此采用不同空间编码方式以利于从采样后的稀疏点中完整、精细地学习其邻域空间结构,从而有效滤除无关特征的同时增强网络局部细粒度表达能力,并进一步提升 OA 。

表6 基于 PointNet++基础框架的6种网络模型 OA 及 F_1 分数统计Table 6 Statistics of OA and F_1 score of six networks based on PointNet++ framework

方法	电力线	车辆	树篱	建筑立面	灌木丛	树木	建筑顶面	低矮植被	道路表面	OA	平均 F_1
HDA-PointNet++	64.2	68.9	19.2	36.5	37.7	69.2	88.2	85.1	99.2	81.2	63.1
RIT_1	37.5	73.4	18.0	49.3	45.9	82.5	94.0	77.9	91.5	81.6	63.3
DPE	68.1	75.2	19.5	44.2	39.4	72.6	91.1	86.5	99.3	83.2	66.2
SSA-Pointnet++	42.2	71.8	39.7	56.6	50.1	78.7	94.9	80.4	91.8	84.6	67.4
PointSIFT	55.7	77.8	30.5	56.9	44.4	79.6	92.5	80.7	90.9	82.2	67.7
PointWeb	49.4	70.6	34.4	52.1	52.5	83.7	93.6	82.0	92.7	82.4	67.9

本文研究表明:

1) FPS 聚类采样通过等比例约束下的类内点云 FPS、基于置信度的 Meanshift 聚类及结合邻域补偿的划分区域 FPS 组合策略,避免尾类样本信息在训练学习过程中被头类样本信息淹没是可行的,从而在保持头类目标点云分割精度的同时提高尾类目标点云分割精度,并通过大幅减少样本数据量而显著提高网络模型学习效率。

2) 将点云语义分割网络中的空间自注意力机制明确分为采样点中心自注意力和邻域自注意力机制,综合两者并通过不同的空间编码方式来增强采样点邻域拓扑结构学习能力是有效的,有助于网络从采样后的稀疏点云中充分捕捉细粒度局部特征用于精确分割目的;

3) 从空间自注意力机制下的空间结构精细学习、顾及长尾分布的点云采样过程两方面出发,本文 SE-PointNet++可进一步提升点云语义分割类别的准确性,对于解决 3D 点云语义分割深度学习网络面临的场景复杂、类别样本不均衡等问题,具有重要意义与应用价值。

3 结 论

CNN 本身并无法区分训练数据集中的有“益”和有“害”信息,故难以从信息角度控制网络的学习偏好,导致 CNN 性能严重依赖于训练集质量。目前 PointNet++ 系列网络模型的 FPS 并未考虑机载 LiDAR 点云数据普遍存在长尾分布现象,其训练过程倾向于获得头类目标主导的整体最优分割,既不利于尾类目标精确分割,一定程度上也会影响到整体分割精度。本文 SE-PointNet++ 网络克服了上述不足,通过 FPS 聚类采样过程可最大程度保留尾类样本并确保网络模型对尾类目标特征的充分学习;通过更完整、细致地观察采样点邻域内的空间关系,有效增强了网络从稀疏样本邻域中学习获得细粒度结构(特征)的能力,相比于所列文献,在 ISPRS 公共数据集上获得了最高 OA 及平均 F_1 分数,并在两者间取得了较好地平衡。理论上,采用点中心自注意力与邻域自

注意力机制在采样点邻域特征选取及其信息池化传递方面的作用应有所不同,本文尚未对此作深入讨论,后续将作深入研究并将结合更多场景机载 LiDAR 点云数据进行网络性能测试,以使之实用化。

参考文献

- [1] 王溪波,曹士彭,赵怀慈,等. 双边特征聚合与注意力机制点云语义分割[J]. 仪器仪表学报, 2021, 42(12): 175-183.
WANG X B, CAO SH P, ZHAO H C, et al. Semantic segmentation of point cloud via bilateral feature aggregation and attention mechanism [J]. Chinese Journal of Scientific Instrument, 2021, 42 (12): 175-183.
- [2] 郭保青,余祖俊,张楠,等. 铁路场景三维点云分割与分类识别算法[J]. 仪器仪表学报, 2017, 38(9): 2103-2111.
GUO B Q, YU Z J, ZHANG N, et al. 3D point cloud segmentation, classification and recognition algorithm of railway scene [J]. Chinese Journal of Scientific Instrument, 2017, 38(9): 2103-2111.
- [3] 邱佳月,赖际舟,李志敏,等. 面向复杂场景的激光雷达地面分割算法[J]. 仪器仪表学报, 2020, 41(11): 244-251.
QIU J Y, LAI J ZH, LI ZH M, et al. A lidar ground segmentation algorithm for complex scenes [J]. Chinese Journal of Scientific Instrument, 2020, 41 (11): 244-251.
- [4] 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.
- [5] 吴军,崔玥,赵雪梅,等. SSA-PointNet++: 空间自注意力机制下的 3D 点云语义分割网络[J]. 计算机辅助设计与图形学学报, 2022, 34(3): 437-448.
WU J, CUI Y, ZHAO X M, et al. SSA-PointNet++: A

- space self-attention CNN for the semantic segmentation of 3D point cloud[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2022, 34(3): 437-448.
- [6] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3d shape recognition[C]. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015: 945-953.
- [7] GUERRY J, BOULCH A, SAUX B L, et al. SnapNet-R: Consistent 3D multi-view semantic labeling for robotics[C]. *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2017: 669-678.
- [8] KALOGERAKIS E, AVERKIOU M, MAJI S, et al. 3D shape segmentation with projective convolutional networks[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 3779-3788.
- [9] YANG Z, TAN B, PEI H, et al. Segmentation and multi-scale convolutional neural network-based classification of airborne laser scanner data[J]. *Sensors*, 2018, 18(10): 3347.
- [10] ZHAO R, PANG M, WANG J. Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network [J]. *International Journal of Geographical Information Science*, 2018, 32(5): 960-979.
- [11] MATURANA D, SCHERER S. VoxNet: A 3D convolutional neural network for real-time object recognition[C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015: 922-928.
- [12] QI C R, SU H, MO K, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 77-85.
- [13] QI C R, YI L, SU H, et al. Pointnet ++: Deep hierarchical feature learning on point sets in a metric space [C]. *The 24th Annual Conference on Neural Information Processing Systems*, 2017: 5105-5114.
- [14] JIANG M, WU Y, ZHAO T, et al. Pointsift: A sift-like network module for 3d point cloud semantic segmentation[J]. *ArXiv Preprint*, 2018, ArXiv: 1807.00652.
- [15] 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.
- [16] YOUSEFHUSSIEN M, KELBE D J, IENTILUCCI E J, et al. A multi-scale fully convolutional network for semantic labeling of 3D point clouds[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 143: 191-204.
- [17] YE Z, XU Y, HUANG R, et al. LASDU: A large-scale aerial LiDAR dataset for semantic labeling in dense urban areas [J]. *International Journal of Geo-Information*, 2020, 9(7): 450.
- [18] HUANG R, XU Y, HONG D, et al. Deep point embedding for urban classification using ALS point clouds: A new perspective from local to global [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 163(5): 62-81.
- [19] CWAB C, XIANG L C, XY A, et al. Airborne LiDAR point cloud classification with global-local graph attention convolution neural network [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 173: 181-194.
- [20] HU Q, YANG B, XIE L, et al. Randla-net: Efficient semantic segmentation of large-scale point clouds [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11108-11117.
- [21] 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.
- [22] THOMAS H, QI C R, DESCHAUD J E, et al. Kpconv: Flexible and deformable convolution for point clouds[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 6411-6420.
- [23] LI Y, BU R, SUN M, et al. PointCNN: Convolution on x-transformed points[J]. *Advances in Neural Information Processing Systems*, 2018, 31.
- [24] WEN C, YANG L, LI X, et al. Directionally constrained fully convolutional neural network for airborne LiDAR point cloud classification [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 162: 50-62.
- [25] WU W, QI Z, FUXIN L. Pointconv: Deep convolutional networks on 3d point clouds [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 9621-9630.
- [26] LI X, WANG L, WANG M, et al. DANCE-NET: Density-aware convolution networks with context encoding for airborne LiDAR point cloud classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020,

- 166(12): 128-139.
- [27] ZHANG Y, ZHOU Z, DAVID P, et al. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9601-9610.
- [28] ZHAO X M, WU J, WANG H J, et al. Injecting spectral indices to transferable convolutional neural network under imbalanced and noisy labels for Landsat image classification [J]. International Journal of Digital Earth, 2022, 15(1): 437-462.
- [29] 张永清, 卢荣钊, 乔少杰, 等. 一种基于样本空间的类别不平衡数据采样方法 [J]. 自动化学报, 2022, 48(10): 2549-2563.
- ZHANG Y Q, LU R ZH, QIAO SH J, et al. A sampling method of imbalanced data based on sample space [J]. Acta Automatica Sinica, 2022, 48(10): 2549-2563.
- [30] YE W Z J, LEE G H. Rpm-net: Robust point matching using learned features [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11824-11833.
- [31] COMANICIU D, MEER P. Mean shift analysis and applications [C]. Proceedings of the Seventh IEEE International Conference on Computer Vision, IEEE, 1999, 2: 1197-1203.
- [32] KANG B, XIE S, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition [J]. ArXiv Preprint, 2019, ArXiv: 1910.09217.
- [33] ZHANG S, LI Z, YAN S, et al. Distribution alignment: A unified framework for long-tail visual recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 2361-2370.
- [34] COVER T, HART P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
- [35] NIEMEYER J, ROTTENSTEINER F, SÖRCEL U, et al. Hierarchical higher order crf for the classification of airborne lidar point clouds in urban areas [J]. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives, 2016, 41: 655-662.

作者简介



陈睿星, 2014年于广州大学获学士学位, 现为桂林电子科技大学博士生, 主要研究方向为人工智能、深度学习、三维点云处理。

E-mail: chenruixing163@163.com

Chen Ruixing received his B. Sc. degree from Guangzhou University in 2014. He is currently a Ph. D. candidate at Guilin University of Electronic Technology. His main research interests include deep learning-based 3D point cloud processing techniques.



吴军 (通信作者), 2007年于武汉大学获工学博士学位, 现为桂林电子科技大学教授、博士生导师, 主要研究方向为数字摄影测量、计算机视觉、图像处理和人工智能。

E-mail: wujun93161@163.com

Wu Jun (Corresponding author) received his Ph. D. degree from Wuhan University in 2007. He is currently a professor and a Ph. D. advisor at Guilin University of Electronic Technology. His main research interests include digital photogrammetry, computer vision, image process and artificial intelligence.



赵雪梅, 2017年于辽宁工程技术大学获工学博士学位, 现为桂林电子科技大学副教授, 主要研究方向为信息几何、人工智能、深度学习遥感数据处理。

E-mail: zhaoxm@guet.edu.cn

Zhao Xuemei received her Ph. D. degree from Liaoning Technical University in 2017. She is currently an associate professor at Guilin University of Electronic Technology. Her main research interests include information geometry, deep learning, and their applications in image processing.



徐刚, 2011年于桂林电子科技大学获硕士学位, 现为中国科学院宁波材料技术与工程研究所副研究员, 主要从事机器视觉、人工智能和视频信息处理。

E-mail: xugang@nimte.ac.cn

Xu Gang received his M. Sc. degree from Guilin University of Electronic Technology in 2011. He is currently an associate professor at Ningbo Institute of Materials Technology and Engineering CAS. His main research interests include machine vision, and image stitching.