计算机科学与探索
Journal of Frontiers of Computer Science and Technology
ISSN 1673-9418,CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目：基于深度学习的点云语义分割研究综述
作者：景庄伟，管海燕，臧玉府，倪欢，李迪龙，于永涛
网络首发日期：2020-08-28
引用格式：景庄伟，管海燕，臧玉府，倪欢，李迪龙，于永涛．基于深度学习的点云语义分割研究综述．计算机科学与探索．
https://kns.cnki.net/kcms/detail/11.5602.tp.20200827.1544.004.html

 doi: 10.3778/j.issn.1673-9418.2006025

基于深度学习的点云语义分割研究综述*

景庄伟[1]，管海燕[2+]，臧玉府[2]，倪　欢[2]，李迪龙[3]，于永涛[4]
1. 南京信息工程大学 地理科学学院，南京 210044
2. 南京信息工程大学 遥感与测绘工程学院，南京 210044
3. 武汉大学 测绘遥感信息工程国家重点实验室，武汉 430079
4. 淮阴工学院 计算机与软件学院，江苏 淮安 223003
+ 通信作者 E-mail：guanhy.nj@nuist.edu.cn

摘　要：近年来，深度传感器和三维激光扫描仪的普及推动了三维点云处理方法的快速发展。点云语义分割作为理解三维场景的关键步骤，受到了研究者的广泛关注。随着深度学习的迅速发展并广泛应用到三维语义分割领域，点云语义分割效果得到了显著提升。本文主要对基于深度学习的点云语义分割方法和研究现状进行了详细的综述。将基于深度学习的点云语义分割方法分为间接语义分割方法和直接语义分割方法，根据各方法的研究内容进一步细分，对每类方法中代表性算法进行分析介绍，总结每类方法的基本思想和优缺点，并系统地阐述了深度学习对语义分割领域的贡献。然后，归纳了当前主流的公共数据集和遥感数据集，并在此基础对比主流点云语义分割方法的实验结果。最后，对语义分割技术面临的挑战以及未来工作的发展方向进行了展望。

关键词：深度学习；语义分割；点云；计算机视觉；综述
文献标志码：A　　**中图分类号：**TP391

景庄伟，管海燕，臧玉府，等. 基于深度学习的点云语义分割研究综述[J]. 计算机科学与探索
JING Z W, GUAN H Y, ZANG Y F, et al. Survey of Poind Cloud Semantic Segmentation Based on Deep Learning[J]. Journal of Frontiers of Computer Science and Technology

Survey of Poind Cloud Semantic Segmentation Based on Deep Learning

JING Zhuangwei[1], GUAN Haiyan[2+], ZANG Yufu[2], NI Huan[2], LI Dilong[3], YU Yongtao[4]
1. School of Geographical Science, Nanjing University of Information Science & Technology, Nanjing 210044, China
2. School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China
3. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
4. Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, Jiangsu 223003, China

Abstract: In recent years, the popularity of depth sensors and 3D laser scanners has led to a rapid development of 3D point clouds processing methods. Semantic segmentation of point cloud, as a key step in understanding 3D scenes, has attracted extensive attention of researchers. With the rapid development of deep learning and its

*The National Natural Science Foundation of China under Grant Nos. 41671454, 41971414 (国家自然科学基金).

widespread applications in 3D semantic segmentation, the quality of point cloud semantic segmentation has been significantly improved. This paper mainly reviews the mainstream deep learning-based methods in point cloud semantic segmentation. We catergoried these deep learning-based methods for point clouds into two groups: indirect and direct semantic segmentation methods. In terms of the characteristics of the algorithm, each of groups are further subdivided. The representative algorithms are analyzed and introduced. The paper systematically details the following aspects of the presented algorithms in semantic segmentation -theories, princioles, advantages and disadvantages, contributions. Moreover, the current mainstream datasets and remote sensing datasets are summarized and the experimental results of some algorithms are compare. Finally, the summaries and future trends are given.

Key words: Deep learning; semantic segmentation; point cloud; Computer vision; review

1 引言

近年来，随着计算机视觉、人工智能以及遥感测绘的发展，SLAM(Simultaneous Localization And Mapping)技术、Kinect 技术以及激光扫描等技术日渐成熟，点云的数据量迅速增长，针对描述点云数据空间信息的高层语义理解也越来越受到关注。语义分割作为点云数据处理与分析的基础技术，成为自动驾驶、导航定位、智慧城市、医学影像分割等领域的研究热点，具有广泛的应用前景。语义分割是一种典型的计算机视觉问题，也称为场景标签，是指将一些原始数据（例如：二维(two-dimensional, 2D)图像，三维(three-dimensional, 3D)点云）作为输入并通过一系列技术操作转换为具有突出显示的感兴趣区域的掩模。

点云语义分割是把点云分为若干个特定的、具有独特性质的区域并识别出点云内容的技术。由于初期三维数据模型库可用数据量较少以及深度网络由二维转到三维的复杂性，传统的点云语义分割方法大多是通过提取三维形状几何属性的空间分布或者直方图统计等方法得到手工提取特征，构建相应的判别模型（如：支持向量机(Support Vector Machine, SVM)[1]、随机森林(Random Forest, RF)[2]、条件随机场(Conditional Random Field, CRF)[3]、马尔可夫随机场(Markov Random Field, MRF)[4]等）实现分割。由于手工提取的特征主要依靠设计者的先验知识以及手工调动参数，限制了大数据的使用。伴随着大型三维模型数据的出现和 GPU 计算能力的不断迭代更新，深度学习在点云语义分割领域逐渐占据了绝对主导地位。深度学习模型的核心思想是采用数据驱动的方式，通过多层非线性运算单元，将低层运算单元的输出作为高层运算单元的输入，从原始数据中提取由一般到抽象的特征。初期，研究者们借鉴二维图像语义分割模型的经验，对输入点云形状进行规范化，将不规则的点云或者网格数据转换为常规的 3D 体素网格或者多视图，将他们提供给深层的网络体系结构。然而，丢失几何结构信息和数据稀疏性等问题限制了多视图方法和体素化方法的发展。于是，研究者开始从三维数据源头着手，斯坦福大学 Qi 等人[5]提出的 PointNet 网络模型，直接从点云数据中提取特征信息，在没有向体素转换的情况下，体系结构保留原始点内的固有信息以预测点级语义。随后，直接处理点云的网络模型方法逐渐发展起来。

目前已有一些综述性论文[6-9]对基于深度学习的点云语义分割研究进行了总结和分析。文献[6]是基于深度学习和遥感数据背景下进行的分类研究进展综述；文献[7]从遥感和计算机视觉的角度概述了三维点云数据的获取和演化，对传统的和先进的点云语义分割技术进行了比较和总结；文献[8]详细介绍了一些较为突出的点云分割算法及常见数据集；文献[9]所做的综述工作涵盖了不同的应用，包括点云数据的形状分类，目标检测和跟踪以及语义和实例分割，涉及的方面较为广泛。本文对前人工作进行了完善，在算法内容上，本文添加了最近提出的新方法，总结了 50 多种三维语义分割算法，根据三维点云数据处理方式，将他们分为两类：间接语义

分割方法和直接语义分割方法。数据集内容上，本文在新增最新公共数据集的同时，增加了常用的三维遥感数据集。未来研究方向上，本文在基于深度学习的语义分割技术评述基础上，对语义分割领域未来研究方向进行了展望并给出各类技术的参考性价值。

2 点云介绍

点云(point cloud)是在同一空间参考系下表达目标空间分布和目标表面特性的海量点集合，其独立描述每个点的相关属性信息，点与点之间没有显著的联系。点云数据主要使用非接触式的技术进行获取，如：图像衍生方法从光谱图像间接生成点云、机载激光雷达扫描仪进行扫描采集、对 CAD 模型进行虚拟扫描等。相对于二维图像，点云有其不可替代的优势——深度信息，点云数据不仅规避了图像采集过程中遇到的姿态、光照等问题，而且其本身具有丰富的空间信息，能够有效地表达空间中物体的大小、形状、位置和方向。相比于体素数据，点云数据空间利用率更高，更加关注于描述对象本身的外表面形状，不会为描述空间的占用情况而保存无用的冗余信息。因此，点云已成为三维数据模型的研究重点，并应用于多种领域，如：大规模场景重建、车载激光雷达、虚拟现实、数字高程模型制作等。然而点云数据自身存在的无序性、密度不一致性、非结构性、信息不完整性等特性使得点云的语义分割充满挑战。因此，有效处理并运用点云的特性是现今研究者应当关注的重点。本章节将点云特性进行简单整理阐述，希望能够为研究者们的研究提供方便。

（1）点云无序性

从数据结构的角度来讲，点云数据只是一组无序的向量集合，若不考虑其他诸如颜色等因素，只考虑点的坐标，则点云数据只是一组 n×3 的点集合。那么当对这 n 个点进行不同顺序的读入时，点的输入组合中共有 n !种，如图 1 所示，图左f_a, f_b, f_c为输入的三个点组成的点云，图右为点云直接输入网络存在的 6 种顺序情况。因此，解决点云的无序性是必不可少的。为了使模型对于输入排列不变，PiontNet[5]使用简单的对称函数汇总来自每个点的信息和特征进行语义分割。PointSIFT[10]使用编码八个方位信息的逐点局部特征描述符保留了无序点云更多的信息，同时仍然保持输入点顺序的不变性。SO-Net[11]网络使用 SOM 模块对归一化后的点云进行批处理解决了点云的无序性。HDGCN[12]提出了图卷积来处理无序点云数据，并且具有强大的提取局部形状信息的能力。RSNet[13]通过切片池层将无序和无结构的输入点的特征投影到特征向量的有序和结构化的序列上。PointCNN[14]学习 x -变换卷积算子，将无序的点云转换为相应的规范顺序。ShellNet[15]将 ShellConv 定义在可由同心球壳划分的区域上，并通过从内壳到外壳的卷积顺序解决了点云的无序性。

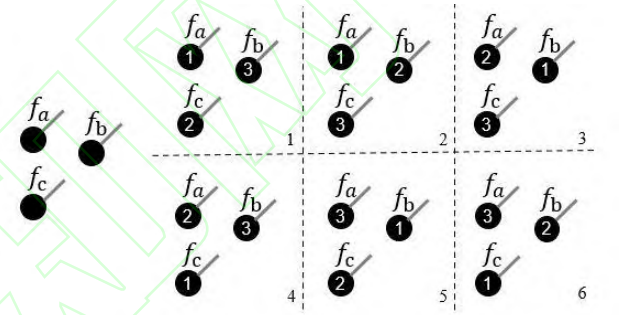

Fig.1 Example of point cloud unordered input

图 1 点云的无序输入示例

（2）点云密度不一致性

实际场景所包含的物体多种多样，相应点云数据也具有不同空间属性。不同点云数据获取方式下，物体的点云的空间距离、密集程度以及点数量差距都很大，如图 2。在密集数据中学习的特征可能不能推广到稀疏采样区域，用稀疏点云训练的模型可能无法识别细粒度的局部结构。因此，能否处理不同密度的点云对分割模型来说具有非常大的挑战性[16]。PointNet++[17]模型中提出的密度自适应点网层，该层可在输入采样密度发生变化时学会组合来自不同尺度区域的特征。RandLA-Net[18]采用随机点采样的方法进行点的选择，以解决高密度大规模的点云场景。GACNet[19]构造了有向图$G(V,E)$，其中K_G邻域是通过在半径 ρ 内随机采样的，相比于K_G的最近邻域查询方法，该方法不受点云稀疏性的影响。3P-RNN[20]通过考虑多尺度邻域，逐点金字塔池化模块以捕获各种密度条件下的局部特征。KPConv[21]

通过结合半径邻域和常规下采样，确保了 KPConv 对不同密度数据的鲁棒性。InterpConv[22]在每个核权值向量的邻域内对点进行归一化，保证其网络具备稀疏不变性。PointConv[23]通过学习 MLP 以近似权重函数，并对学习的权重应用反密度标度补偿非均匀采样。

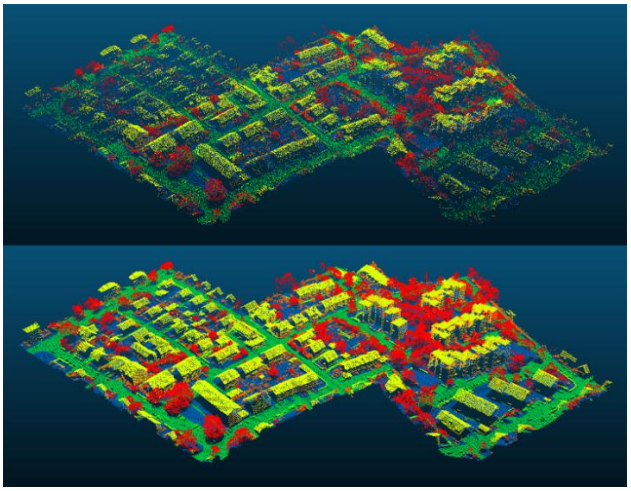

Fig.2 Point cloud scenes with different densities

图 2 不同密度的点云场景

（3）点云非结构性

二维图像是结构化的数据，可以使用一个二维矩阵进行表示。而点云数据是非结构化的，想要直接输入到神经网络模型中是非常困难的。如果将点云数据体素化，利用深度学习模型进行特征提取可以取得较好的分割结果，但是这种方法由于内存限制，只能使用比较小分辨率的体素网格，从而造成信息的丢失，所以其整体性能与精度仍然无法得到显著提高。点云本质上缺乏拓扑信息，因此设计恢复拓扑的模型（如 DGCNN[24]、RGCNN[25]、DPAM[26]等基于图卷积的方法）可以丰富点云的表示能力。另外，ConvPoint[27]中设计了一种针对非结构化数据的连续卷积公式。

（4）点云信息不完整性

点云是一群三维空间点坐标构成的点集。由于本质上是对三维世界中物体几何形状进行低分辨率重采样，因此点云数据提供的几何信息是不完整的；另外，点云数据采集时由于遮挡等原因，无法获取目标物体完整的三维描述。而且在模型训练过程中也存在这样的问题，如 PointNet[5]的全局特征仅汇总了单个块的上下文，汇总信息仅在同一个块中的各个点之间传递，但是每个块之外的上下文信息也同样重要。因此，CU&RCU[28]引入了两种添加上下文的机制：输入级上下文（直接在输入点云上运行）和输出级上下文（用于合并输入级上下文的输出）。图神经网络(GNNs)也被广泛用于处理不规则点云数据，这些方法[14、23、24、29、30]在欧几里得或特征空间的邻域中构建局部图，通过加权和或从邻域到中心的池化来聚合局部特征，处理不规则点云数据。

3 基于深度学习的三维点云语义分割方法

随着深度学习技术的出现，点云语义分割领域实现了巨大的改进。近年来，研究者们提出了大量的基于深度学习的分割模型以处理点云。与传统算法相比，此类模型性能更优，达到了更高的基准。本章根据三维点云数据处理方式，将基于深度学习的三维点云语义分割方法分为两大类，即间接语义分割方法和直接语义分割方法。间接语义分割方法是将原始点云数据转换为常规的 3D 体素网格或者多视图，通过数据转变的方式间接地从三维点云数据中提取特征，从而达到语义分割的目的。直接语义分割方法是直接从点云数据中提取特征信息，在没有向体素和多视图转换的情况下，体系结构保留原始点内的固有信息以预测点级语义。

3.1 间接语义分割方法

借鉴二维图像语义分割模型的经验，研究者们首先将不规则的点云数据转换为常规的 3D 体素网格或者多视图，输入到深层网络体系结构以实现点云的语义分割。本节整理总结了 20 篇具有代表性的文献，将间接语义分割方法再分为基于二维多视图方法和基于三维体素化方法两个子类，并分别进行了总结与分析。图 3 为 2015 年起间接语义分割方法的发展，不同颜色代表不同间接语义分割方法类别。

doi: 10.3778/j.issn.1673-9418.2006025

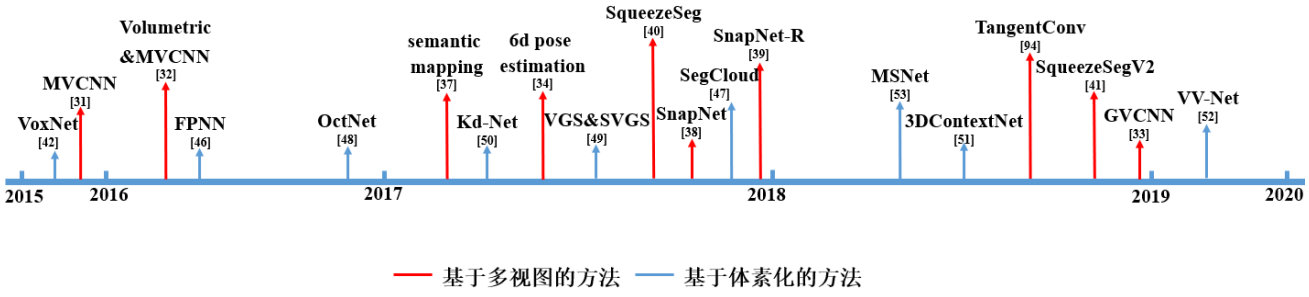

Fig.3 Timeline of indirect semantic segmentation

图 3 间接语义分割方法发展时间轴

3.1.1 基于二维多视图方法

早期研究者们在点云数据上应用深度学习是将点云投影到多个视图的二维图像中，在投影的二维图像上使用卷积等常规处理技术，从而实现点云数据语义分割。多视图 CNN(Multi-view Convolutional Neural Network, MVCNN)处理三维点云数据的方法由 Su 等人[31]首次提出，该类方法的具体步骤如图 4 所示，首先获取三维目标形状在不同视角下的二维图像，对每个视图进行图像特征提取，最后通过池化层和完全连接层将不同视角的图像进行聚合得到最终的语义分割结果。

虽然 MVCNN 能很好地整合不同视角下影像特征从而获得较好的三维物体的描述，但是该方法并不能有效地利用每张视图的局部特征信息，也不能动态地选择视图；同时，将三维物体投影到二维图像会丢失大量关键的几何空间信息，导致其最终语义分割精度不高。因此，Qi 等人[32]通过引入多分辨率三维滤波来捕获目标多尺度信息以提高其语义分割性能。Feng 等人[33]在 MVCNN 的基础上提出 GVCNN(group-view convolutional neural network)框架，将不同视图下 CNN 提取的视觉描述子进行分组，可有效利用多视图状态下特征之间的关系。

随着 RGB-D 传感器（微软 Kinect 等）的发展，RGB-D 数据也逐渐被广泛应用。RGB-D 数据除了提供颜色信息外，还提供额外的深度信息，有利于语义分割任务。Zeng 等人[34]使用机械臂获取多视角 RGB-D 图像并输入 FCN 网络中，通过训练多个网络（AlexNet[35]和 VGG-16[36]）提取特征，同时评估了使用 RGB-D 图像深度信息的优势。随后，Ma 等人[37]使用 SLAM 技术获取相机轨迹，并将 RGB-D 图像转换到真实标注数据相同尺度，保证模型训练中多个视角的一致性。SnapNet[38]围绕三维场景生成一系列二维快照，对每对二维快照进行完全卷积网络的像素标记后，再将像素标记反投影到原始点云上。SnapNet-R[39]改进了 SnapNet 网络，对多个视图直接处理以实现密集的三维点标记，从而改善分割效果。然而，二维快照破坏了三维数据的内在几何关系，无法充分利用三维空间上下文的全部信息。

SqueezeNet 作为轻量级网络结构，能够减少模型参数量并且保持精度，因而在计算机视觉领域得到越来越广泛应用。Wu 等人[40]借鉴 SqueezeNet 的思想，提出了 SqueezeSeg 网络。SqueezeSeg 利用球面投影将稀疏的三维点云转换为二维图像输入到基于 SqueezeNet 的 CNN 模型中进行语义分割，利用条件随机场(Conditional Random Field, CRF)作为递归层对语义分割结果进一步优化，并通过传统的聚类算法获得最终标签。但是该方法语义分割准确率受到点云采集过程中产生的失调噪声(dropout noise)影响，随后该团队[41]提出 SqueezeSegV2，添加了上下文聚合模块(the Context Aggregation Module, CAM)，该模块可以从更大的接收域中聚合上下文信息，从而增强网络对失调噪声的鲁棒性，提高了语义分割的准确率。

尽管基于多视图的语义分割方法存在三维空间信息不完整性和投影角度的问题，但其解决了点云数据的结构化问题，又可依赖于较多成熟的二维算法和丰富的数据资源，可用于许多特定和小型的场景，具有较强实用性。

doi: 10.3778/j.issn.1673-9418.2006025

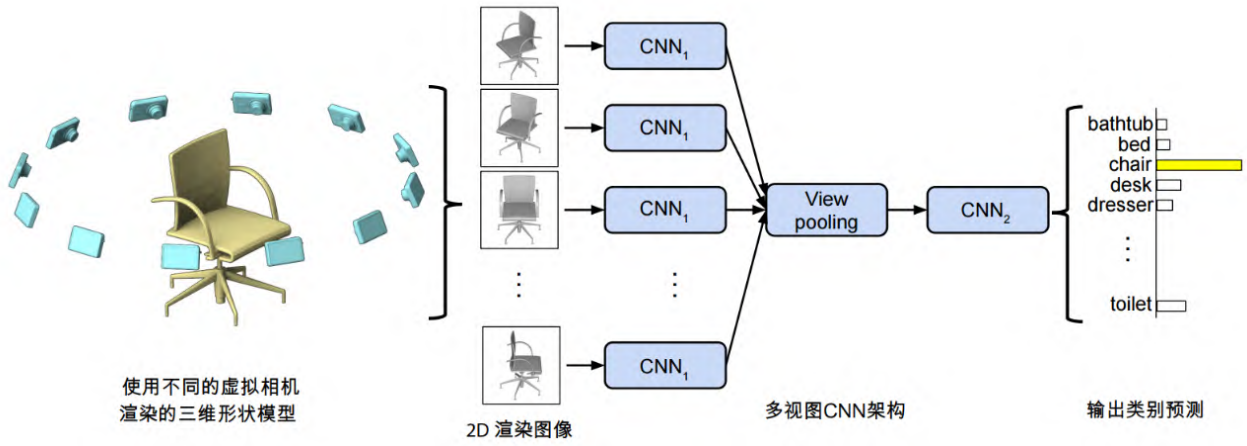

Fig.4 Workflow for MVCNN

图 4 MVCNN 网络的处理流程

3.1.2 基于三维体素化方法

鉴于 CNN 在图像语义分割中取得的有效成果以及体素与图像在数据组织形式上的相似性，研究者们将原始点云数据转换为体积离散（即体素）数据，提出了基于三维的神经网络模型，以实现点云的语义分割。体素化操作是利用占用网格将环境状态表示为随机变量的 3D 网格（每个网格对应于一个体素），并根据传入的传感器数据和先验知识维持其占用率的概率估计[42]。目前，基于体素数据的各种深度网络已被应用于形状分类[43]，室内场景的语义分割[44]和生物医学记录[45]。VoxNet 模型[42]是最早基于体素数据的三维 CNN 模型，该模型展示了三维卷积算子从体素占用网格学习特征的潜力。虽然体素模型的提出解决了点云无序性和非结构化的问题，但三维数据的稀疏性与空间信息不完整性导致语义分割效率低。此外，相较于二维图像数据，点云数据体素化由于增加了一个维度，其计算开销更大，并且限制了体素模型的分辨率。

针对三维数据的稀疏性，Li 等人[46]采用场探测滤波器(field probing filter)代替卷积神经网络中的卷积层从点云体素中提取特征。但是，该方法会降低语义分割输出结果的分辨率。针对体素网格低分辨率的限制，SegCloud[47]网络放弃了基于体素的 CRF 方法，转而使用原始 3D 点作为节点来运行 CRF 推理。该网络将 3D-FCNN 生成的粗体素预测通过三线性插值返回到原始点云，然后使用全连接条件随机场(Fully Connected CRFs, FCCRF)增强预测结果的全局一致性并在这些点上提供细粒度语义。

为了减少不必要的计算和内存消耗，有些学者提出了基于八叉树结构的分割模型，如 OctNet[48]和 VGS&SVGS[49]模型。OctNet[48]模型中，每个八叉树根据数据的密度分割三维空间，将存储器分配和计算集中到相关的密集区域，在不影响分辨率的情况下实现更深层的网络。VGS(voxel- and graph-based segmentation)&SVGS(supervoxel- and graph-based segmentation)[49]模型采用基于八叉树的体素化方法组织点云以方便邻域遍历，利用图论(graph theory)在局部上下文信息的基础上进行体素和超体素的聚类，并使用感知定律(perceptual laws)以纯几何的方式进行分割。Kd-tree 结构也被应用到基于深度学习的语义分割模型中，如 Kd-Net[50]和 3DContextNet[51]模型。Kd-Net[50]提出使用 Kd-tree 组织点云数据，规则化深度网络输入结构，提高了点云计算和存储效率。3DcontextNet[51]利用 Kd-tree 结构提供的点云局部和全局上下文线索进行特征学习并聚合点特征。与 Kd-Net 不同，3DContextNet 不改变空间关系，可用于三维语义分割。以上基于树结构的方法虽然减少了计算和内存消耗，但此类方法依赖体素边界，没有充分利用其局部几何结构。因此，Meng 等人[52]

利用基于径向基函数(RadialBasis Functions,RBF)的变分自动编码器(Variational Autoencoder, VAE)网络对体素结构进行扩展，编码每个体素内的局部几何结构从而提高分割精度。MSNet[53]网络围绕每个点，将不同尺度的空间上下文划分为不同尺度的体素，以自适应地学习局部几何特征，该方法在遥感、测绘数据获得不错的语义分割结果。

以上研究从不同角度解决了点云体素化带来的不足，减少三维体素输入的信息丢失和计算需求，但由于体素化算法的空间复杂度高，存储和运算过程中均需较大的开销，因此实用性相对较低。不过随着计算性能和存储方法的不断升级，该类方法还是具有一定潜在的发展空间。

3.2 直接语义分割方法

为了降低预处理过程中的计算复杂度与噪音误差影响，研究者开始从三维数据源头着手，直接从点云数据中提取特征信息，因而逐渐发展出一些直接处理点云的网络模型方法。PointNet 网络[5]架构是该类方法的开拓者，该网络直接处理点云数据的分类与分割任务，如图 5 所示。PointNet 在语义分割时，以点云中每一个点作为输入，输出每个点的语义类标签。PointNet 网络主要解决三个核心问题：点云无序性、置换不变性和旋转不变性。针对点云的无序性，PointNet 使用简单的对称函数聚合每一个点的信息。针对点云的置换不变性，PointNet 采用多层感知机(MLP)对每个点进行独立的特征提取，并将所有点信息聚合得到全局特征。此外，PointNet 网络参考了二维深度学习中的 STN 网络，在网络架构中加入 T-Net 网络架构，对输入的点云进行空间变换，使其尽可能满足旋转不变性。

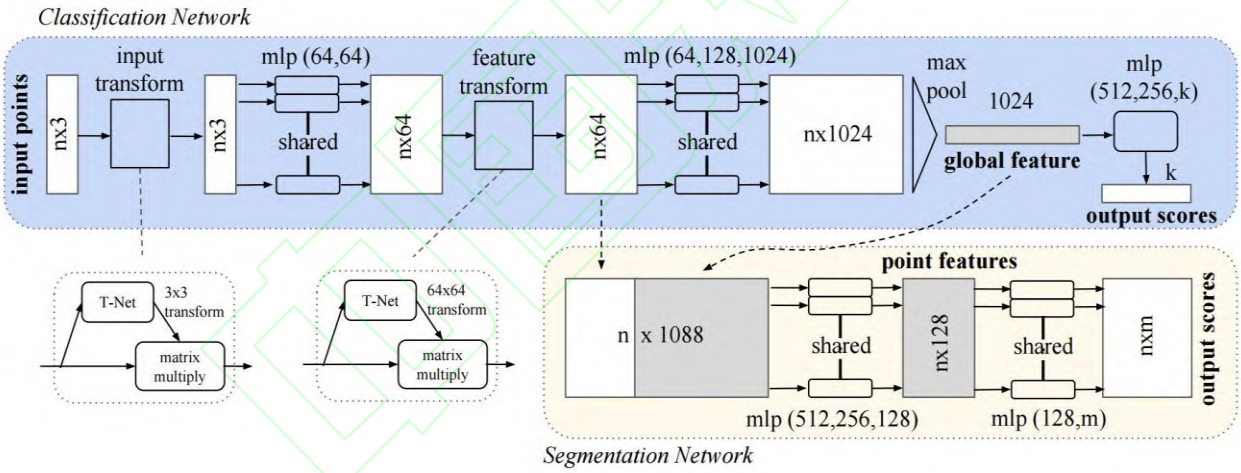

Fig.5 Network framework for PointNet

图 5 PointNet 网络架构

PointNet 网络依旧存在着很多的缺陷：无法很好地捕捉由度量空间引起的局部结构问题，欠缺对局部特征的提取及处理；每个点操作过于独立，其没有考虑到邻近点的交互关系，而无法高效刻画相关区域的语义结构；统一的模板无法有效地解决密度不均一的数据。为了解决这些问题，研究者们基于 PointNet 算法提出了一系列解决方案，本节整理总结了 30 篇具有代表性的文献，从算法特点的角度分为六大类：基于邻域特征学习的方法、基于图卷积的方法、基于 RNN 的方法、基于优化 CNN 的方法、基于注意力机制的方法和结合实例分割的方法，并分别进行总结和分析。图 6 为 2017 年起直接语义分割方法发展的时间轴，图中不同颜色代表不同的直接语义分割方法类别。

doi: 10.3778/j.issn.1673-9418.2006025

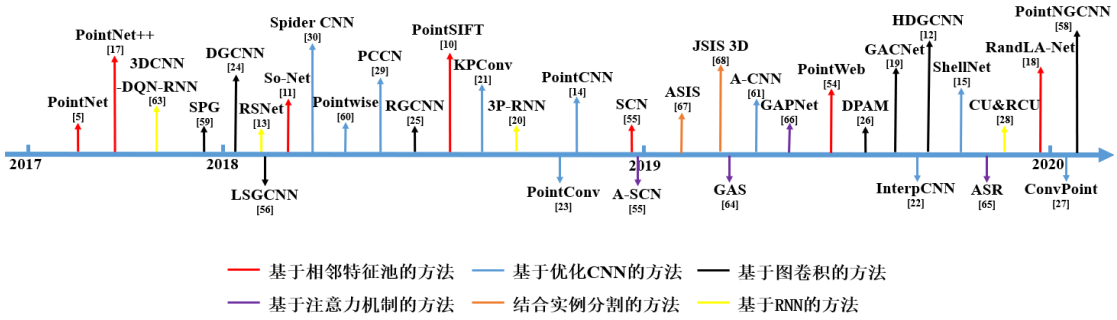

Fig.6 Timeline of direct semantic segmentation

图 6 直接语义分割方法发展时间轴

3.2.1 基于邻域特征学习的方法

PointNet 没有捕获由度量空间点引起的局部结构特征，限制了细粒度图案识别和复杂场景泛化能力。目前，为了捕获局部特征，已有大量基于邻域特征学习的网络模型通过聚集来自局部相邻点的信息或融合不同层次区域特征来捕获点云中的上下文信息，将获取的全局特征与局部特征有效结合以提高语义分割的性能。

PointNet++[17]是 PointNet 的分层版本，它的每个图层都有三个子阶段：采样，分组和特征提取。图 7 为 Pointnet++的整体网络架构。采样层中，在输入点云中使用迭代最远点采样(Farthest Point Sampling, FPS)方法选择一系列局部区域的中心点。分组层中，通过查找中心点周围的“邻近”点，创建多个点云子集。最后采用 PointNet 网络进行卷积和池化来获得这些点云子集的高阶特征表示。此外，作者还提出了密度自适应切入点网层，当输入采样密度发生变化时，则学习不同尺度区域的特征。

Pointnet++网络不仅解决了点云数据采样不均匀的问题，而且考虑了点与点之间的距离度量。它通过层级结构学习局部区域特征，使得网络结构更有效、更稳健。虽然该模型有效改善了局部特征提取问题，但 PointNet++和 PointNet 模型一样，单独提取点的特征，依然没有建立点与点之间关系（如方向性等），对于局部特征的学习仍然不够充分。为了模拟点之间的交互关系，Zhao 等人[54]提出了 PointWeb，通过自适应特征调整(Adaptive Feature Adjustment,AFA)模块实现信息交换和点的局部特征学习，构建局部完全链接网络来探索局部区域中所有点对之间的关系。该方法充分利用点的局部特征，并形成聚合特征进行三维点云语义分割。另外，为了解决 PointNet++中 K-邻域搜索可能处于一个方向的问题，PointSIFT 模块[10]的方向编码单元在 8 个方向上对最近点(nearest point)的特征进行卷积，从而能够提取更可靠和稳定的表征点。

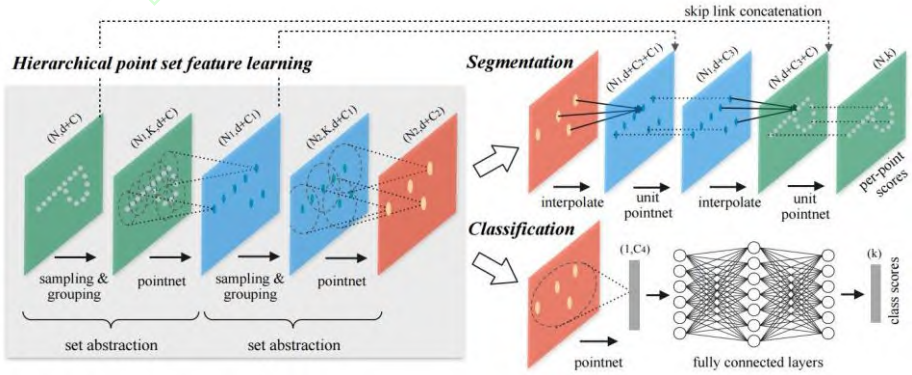

Fig.7 Network framework for PointNet++

图 7 PointNet++网络架构

为了更加有效的利用点云的局部特征信息，研究者们基于 PointNet++网络架构提出了许多点云语义分割的网络模型，如：SO-Net[11]、SCN[55]、RandLA-Net[18]等。

SO-Net[11]网络通过建立自组织映射(Self-Organizing Map, SOM)模拟点云的空间分布，对单个点和 SOM 节点进行分层特征提取，最终用单个特征向量来表示输入点云，从而固定点的位置以实现点云高效分割。虽然 SO-Net 网络架构对于大规模点云数据处理还具有一定的局限性，但其为后续的大规模点云语义分割提供了重要基础。与 SO-Net 不同，ShapeContextNet(SCN)[55]采用形状上下文作为基本构建块开发了一种分层结构，通过捕获并传播局部和全局形状信息来表示对象点的内在属性。RandLA-Net[18]是一种用于大规模点云处理的轻量级网络，该网络使用随机点采样法替代 PointNet++的最远点采样法，通过局部特征聚集模块以捕获和保留局部几何特征，在存储和计算方面实现了显着的提高。

3.2.2 基于图卷积的方法

图卷积方法将卷积运算与图结构表示相结合。图卷积神经网络是一种直接在图结构上运行且能够依靠图中节点之间的信息传递来捕获图中依赖关系的卷积神经网络，在计算机视觉领域的应用越来越广泛。

针对 Pointnet++框架[17]中以孤立方式进行特征学习的局限性，Wang 等人[56]提出一种局部谱图卷积(Local Spectral Graph Convolution)，它从点的邻域构造局部图，利用谱图卷积结合新的图池策略学习相邻点的相对布局及特征。与上述方法不同，Martin 等人[57]在空间域中对图形信号进行了类似于卷积的运算，并使用非对称边缘函数来描述局部点之间的关系。但是，边缘标签是动态生成的，没有考虑局部点的分布不规则性。于是，RGCNN[25]基于谱图理论，将点的特征作为图上的一个节点以克服点云的不规则性。Wang 等人[24]改进文献[57]的方法，提出了动态图卷积神经网络 DGCNN。DGCNN 通过构造局部邻域图并利用边缘卷积(EdgeConv)操作提取中心点的特征和中心点与 K 近邻域(KNN)点的边缘向量以获得点云的局部特征。EdgeConv 考虑了点的坐标与邻域点的距离，却忽视了相邻点之间的向量方向，最终还是损失了一部分局部几何信息。

随后，在 DGCNN 的研究基础上发展了一系列基于图卷积的算法，如 GACNet[19]、HDGCN[12]、DPAM[26]和 PointNGCNN[58]等。其中，GACNet[44]提出了一种具有可学习内核形状的图注意力卷积(graph attention convolution,GAC)，用于 3D 点云的结构化特征学习。受深度卷积和图卷积的启发，Wang 等人[12]提出由深度图卷积(Depthwise Graph Convolutional,DGConv)块组成的层次结构网络：HDGCN，以提取点云局部特征和全局特征。Liu 等人[26]认为以往的点聚集方法仅在欧几里得空间中进行点采样和分组，严重限制了他们适应更多场景的能力。于是提出了一种基于图卷积的动态点聚集模块(dynamic points agglomeration module, DPAM)，将点聚集（采样，分组和合并）的过程简化为聚集矩阵和点特征矩阵相乘。PointNGCNN[58]构造邻域图来描述邻域点之间的关系，并使用切比雪夫多项式作为邻域图滤波器提取邻域几何特征。在此基础上，将每个邻域的特征矩阵和拉普拉斯矩阵(Laplacian matrix)放入网络中，利用最大池化操作获得到每个中心的特征。

此外，为了处理大规模点云的语义分割，Landrieu 等人[59]在 2017 年提出了超点图(superpoints graph, SPG)。SPG 将几何分割后的每一个几何形状看作一个超点(superpoint)构建超点图，利用 PointNet 对超点图进行超点嵌入以及图卷积处理，分类得到语义标签。SPG 能够详细描述相邻目标之间的关系，可有效解决每个点操作过于独立，点与点之间缺乏联系等问题。

3.2.3 基于优化 CNN 的方法

卷积神经网络(Convolutional Neural Network, CNN/ ConvNets)是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元，目前对于大型图像处理有着出色的表现。卷积神经网络由一个或多个卷积层和顶端的全连接层组成，同时也包括关联权重和池化层。这一结构使得卷积神经网络能够利用输入数据的三维结构，将特征从低级

特征提取到高级特征。近年来，一些研究者对 CNN 进行了优化，并将他们应用在点云语义分割的模型中。上文提到的图卷积也算优化 CNN 方法中的一类。

由于点云数据的无序性，导致输入点云数据时的排列顺序千差万别，使得卷积操作很难直接应用到点云数据上。为了进一步解决这个问题并利用标准 CNN 操作的优势，PointCNN[14]尝试学习 x -变换卷积算子，将无序的点云转换为相应的规范顺序，之后再使用典型的 CNN 架构来提取局部特征。 x -变换可以实现“随机应变”，即当输入点的顺序变化时, x 能够相应地变化，使加权和排列之后的特征近似不变，输入特征在经过 x -变换的处理之后能够变成与输入点顺序无关同时也编码了输入点形状信息的归一化的特征。不同于 PointCNN，PCCN[29]提出一种参数连续卷积(Parametric Continuous Convolution)，使用点来承载内核权重并利用参数化的核函数跨越整个连续向量空间，由于其不使用任何形式的邻域，导致该网络不可再扩展。同样，在解决缺乏空间卷积的过程中，Thomas 等人[21]提出了提供可变形卷积算子的核点卷积(Kernel Point Convolution, KPConv)，通过应用邻域中最近距离内核点的权重，对每个局部邻域进行卷积。KPConv 的卷积权重由到核点的欧几里德距离确定，并且核点的数量不是固定的，因此 KPConv 比固定网格卷积灵活性更强。随后，ConvPoint[27]使用多层感知器(MLP)学习关联函数替代 KPConv 使用的 RBF 高斯函数关联输入和内核。ConvPoint[27]提出离散卷积神经网络的泛化，通过使用连续核替换离散核以处理点云。Pointwise[60]利用逐点卷积(Pointwise Convolution)获取点的局部特征信息实现语义分割。但是，逐点卷积使用体素容器定位内核权重，因此缺乏像 KPConv 一样的灵活性。SpiderCNN[30]通过对一系列的卷积滤波器进行参数化，将卷积运算从常规网格扩展到可嵌入$\mathbb{R}^n$的不规则点集，并捕获复杂的局部几何变化。SpiderCNN 继承了经典 CNN 的多尺度层次结构，进而能够提取语义深层特征。InterpConv[22](Interpolated Convolution)利用一组离散的内核权重，并通过插值函数将点特征插值到相邻的内核权重坐标上进行卷积。在 InterpConv 基础上提出内插卷积神经网络(InterpCNN)，以处理点云的室内场景语义解析任务。ShellConv[15]使用同心球壳的统计信息来定义有代表性的特征并解决了点的无序性输入，使得传统的卷积运算可以直接处理这些特征。Wu 等人[23]将动态滤波器扩展到一个新的卷积运算，命名为 PointConv。PointConv 在局部点坐标上训练多层感知器来逼近卷积滤波器中的连续权函数和密度函数，使其具有置换不变性和平移不变性。此外，将 PointConv 扩展为反卷积运算符(PointDeconv)，将特征从子采样点云传播回原始分辨率。A-CNN[61]在分层神经网络中应用环形卷积(Annular Convolution)以实现大场景的语义分割。环形卷积可提取每个点周围局部邻域的几何特征，并在后续的点云处理中，使用特征融合方法将全局特征与局部特征结合以改善分割效果。

3.2.4 基于 RNN 的方法

循环神经网络(recurrent neural network, RNN)[62]是目前深度学习中另一种主流模型，RNN 不仅可以学习当前时刻的信息，还可以依赖之前的序列信息，有利于建模全局内容和保存历史信息，促进上下文信息的利用。Engelmann 等人[28]在 PointNet 网络的基础上提出了输入级上下文和输出级上下文两个扩展。输入级上下文是将点块转换为多尺度块和网络块；输出级上下文是将 PointNet 提取的分块特征依次送入合并单元(Consolidation Units, CU)或循环合并单元(Recurrent Consolidation Units, RCU)。实验结果表明，将网络架构扩展到更大尺度的空间上下文中有助于提高语义分割性能。Liu 等人[63]融合三维卷积神经网络(CNN)，深层 Q 网络(DQN)和残差递归神经网络(RNN)，提出了 3DCNN-DQN-RNN 用于大规模点云的语义解析。3D CNN 网络学习点的空间分布和形状颜色特征；DQN 网络定位类对象；残差 RNN 处理输入的级联特征向量获得最终的分割结果。该方法利用残差 RNN 进一步提取了点的识别性特征，从而提高了大规模点云的解析精度。

为了进一步优化 Pointnet++网络，并且考虑点与点之间方向性关系，RSNet[13]模型通过 x、y、z 三个方向的切片池化层将无序点云转换为有序序列并提取全局特征，采用双向 RNN(bidirectional RNN)处理点云有序序列，提取局部相关性特征，利用切

片解析层将序列中的特征分配回各个点，最终输出每个点的语义预测标签。相比其他为了得到局部信息需要复杂计算的模型，RSNet 简化了获取局部信息的计算。同样，Ye 等人[20]从 x，y 方向连续地扫描三维空间提取信息，并构建一个逐点金字塔池化模块(pyramid pooling module)提取三维点云不同密度的局部特征，同时使用分层的双向 RNN 学习空间上下文信息，从而实现多层次的语义特征融合。

3.2.5 基于注意力机制的方法

注意力机制基本思想是让系统能够忽略无关信息而关注重点信息。注意力机制通过神经网络算出梯度并且前向传播和后向反馈来学习得到注意力的权重。为进一步提升分割精度，一些研究者将注意力机制引入至语义分割算法中。Yang 等人[64]开发了一个基于点云推理的点注意力变压器(Point Attention Transformer, PAT)，并提出了群洗牌注意力(Group Shuffle Attention, GSA)用于建模点之间的关系。同时，Yang 等人[64]还提出了一种端到端、置换不变性、可微的 Gumbel 子集采样(Gumbel Subset Sampling, GSS)替代广泛使用的最远点采样(FPS)，以选择具有代表性的点子集。Zhao 等人[65]考虑通过利用相邻点的初始分割分数来改善三维点云分割结果，提出了一种基于注意力的分数细化(Attention-based Score Refinement, ASR)模块，该模块根据各个点的初始分割分数计算权重，再根据计算出的权重合并每个点及其邻近点的分数，从而对分数进行优化。该模块可以轻松集成到现有的深度网络中，以提高最终的分割效果。GACNet[19]通过建立每个点与周围点的图结构，并引入注意力机制计算中心点与每一个邻接点的边缘权重，最后通过对权重加权计算出每个点的特征后再进行图池化(Graph Pooling)和下采样，从而使得网络能在分割的目标的边缘部分取得更好的效果。

借鉴 Mnih 等人提出的自注意力机制(self-attention)，GAPNet[66]将其与 GCNN 结合，通过在堆叠的多层感知器(MLP)层中嵌入图形注意机制以学习局部几何表示，GAPNet 可将 GAPLayer 和注意力池层集成到堆叠的多层感知器(MLP)层或现有管道（例如 PointNet)中，以更好地从无序点云中提取局部上下文特征。SCN[55]受基于自注意力模型的启发，在 SCN 基础上提出 A-SCN(Attentional ShapeContextNet)模型，以自动完成上下文区域选择、特征聚合和特征转换等过程。

3.2.6 结合实例分割的方法

语义分割和实例分割相结合方法能取长补短，既不重复操作，减小计算的复杂度，又可以增加分割精度，实现双赢。

Wang 等人[67]提出了一个 ASIS（实例和语义的关联分割）框架，通过学习语义感知的点级实例嵌入，使实例分割从语义分割中受益。同时，融合属于同一实例的点的语义特征，可自动分离属于不同语义类的点嵌入，以进行更准确的基于点的语义预测。

与此同时，Pham 等人[68]基于 PointNet 网络开发了一个多任务逐点网络，它同时执行两项任务:预测三维点的语义信息，并将这些点嵌入高维向量中，使相同对象实例的点相似嵌入表示。然后，利用一个多值条件随机场模型，将语义和实例标签结合起来，将语义和实例分割问题表述为场模型中标签的联合优化问题。作者所提出的联合语义实例分割方案对单个构件具有较强的鲁棒性，实验结果相对于 ASIS 来说更好一些。

4 语义分割实验分析与对比

本章首先梳理了下测试阶段价值较高的 RGB-D 和三维公开数据集，然后在此基础上对现有语义分割算法的性能进行了综合性对比和讨论。

4.1 公共数据集

为了验证研究者们提出算法对语义分割的效果，有效的数据集是不可或缺的一环。随着深度学习在三维语义分割中的发展，三维数据集的地位愈加重要。目前，为了促进三维点云语义分割的研究，许多研究机构提供了一些可靠且开放的三维数据集，见表 1，下面对点云语义常用的数据集按类别以及

时间顺序进行简要地描述。

4.1.1 RGB-D 数据集

（1）RGB-D Object[69]：（https://rgbd-dataset.cs.washington.edu/）该数据集 2011 年由美国华盛顿大学的研究小组开发，由 11427 幅人工手动分割的 RGB-D 图像组成，整个数据集包含 300 个常见的室内物体，并将这些物体分为了 51 个类。该数据集使 Kinect 型三维摄像机获取图像，对于每一帧，数据集提供了 RGB 及深度信息，这其中包含了物体、位置及像素级别的标注。另外，还提供了 22 个带注释的自然场景视频序列，用于验证过程以评估性能。

（2）NYUDv2[70]：（https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html）该数据集 2012 年由美国纽约大学的研究小组开发，包含 1449 张由微软 Ki-nect 设备捕获的室内场景的 RGB-D 图像，其中训练集 795 张，测试集 654 张，对象被分为 40 个类，每个对象都标有类和实例号。但是由于其相对于其他数据集规模较小，所以该数据集主要用于辅助机器人导航的训练任务。

（3）SUN3D[71]：（http://sun3d.cs.princeton.edu/）该数据集 2013 年由美国普林斯顿大学的研究小组开发，其中包含使用 Asus Xtion 传感器捕获的 415 个 RGB-D 序列，是一个具有摄像机姿态和物体标签的大型 RGB-D 视频数据库。每一帧均包含场景中物体的语义分割信息以及摄像机位态信息。

（4）Bigbird[72]：（http://rll.berkeley.edu/bigbird/）该数据集 2014 年由美国加州大学伯克利分校的研究小组开发，使用计算机控制的光平台和静态校准的成像设备对 125 个对象进行 3D 扫描，每个对象由 600 个 3D 点云和 600 个跨越所有视图的高分辨率(1200 万像素)图像组成。

（5）ViDRILO[73]：（http://www.rovit.ua.es/dataset/vidrilo）该数据集 2015 年由西班牙卡斯蒂利亚大学和阿利坎特大学的研究小组共同开发，包含其使用 Microsoft Kinect v1 传感器在五个室内场景中捕获的 22454 个 RGB-D 图像。每个 RGB-D 图像都标有场景的语义类别（走廊、教授办公室等）。该数据集被发布用于基准测试多个问题，如多模式地点分类、目标识别，三维重建或点云数据压缩。

（6）SUN RGB-D[74]：（http://rgbd.cs.princeton.edu/）该数据集与 SUN3D 数据集由美国普林斯顿大学的同一研究小组开发，数据由四个不同的传感器捕获，包含 10000 张 RGB-D 图像，其尺寸与 Pasc-al VOC 相当。整个数据集是密集注释的，包括 146617 个 2D 多边形和 58657 个具有精确对象定位的 3D 包围框，以及一个三维房间布局和场景类别，适用于场景理解任务。

（7）ScanNet[44]：（http://www.scan-net.org/）该数据集 2017 年由美国普林斯顿大学、斯坦福大学以及德国慕尼黑工业大学的研究者共同开发，是一个 RGB-D 视频的室内场景数据集。在 1513 次扫描中获得 250 万次视图，附加了 3D 相机姿态、表面重建和实例级语义分割的注释。该数据集的对象被分为 20 个类，包含各种各样的空间，范围从小（例如，浴室，壁橱，杂物间）到大（例如，公寓，教室和图书馆）。该数据被广泛应用于三维对象分类、语义体素标记和 CAD 模型检索等三维场景理解任务上。

（8）Matterport3D[75]：（https://niessner.github.io/Matterport/）如图 8，该数据集 2017 年由美国普林斯顿大学、斯坦福大学以及德国慕尼黑工业大学的研究者共同开发，包含来自 90 多个建筑规模场景的 194 400 个 RGB-D 图像和 10800 个全景。注释提供了表面重建，相机姿态以及 2D 和 3D 语义分割内容。精确的全局校准和全面的、多样的全景视图覆盖了整个建筑，从而支持各种监督的计算机视觉任务，如：关键点匹配，视图重叠预测，根据颜色进行的正常预测，语义分割和区域分类。

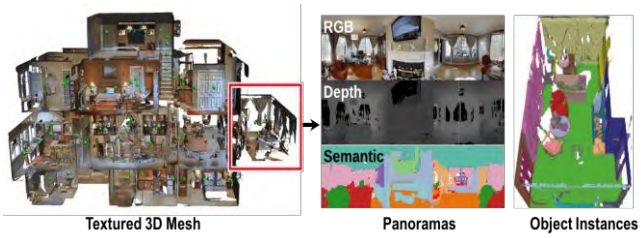

Fig.8 Example image of Matterport3D dataset

图 8 Matterport3D 数据集示例图

4.1.2 室内三维数据集

（1）A Benchmark for 3D Mesh Segmentation[76]：（http://segeval.cs.princeton.edu/）该数据集 2009 年

由美国普林斯顿大学的研究小组开发，包含 380 个网格，被分为 19 个常见对象类别（如桌子、椅子等），每个网格手动的被分割为不同的功能区域，旨在帮助研究三维零件的语义分割和人类如何将对象分解为各个有意义的部分。

（2）PrincentonModelNet[43]：（http://modelnet.cs.princeton.edu/）该数据 2015 年由美国普林斯顿大学、麻省理工学院以及中国香港中文大学的研究人员共同开发，该数据集是一个为计算机视觉，计算机图形学，机器人和认知科学的研究者提供的清晰物体 3D CAD 模型，ModelNet 总共有 662 中目标分类，127915 个 CAD，以及十类标记过方向朝向的数据。其中包括三个子数据集：ModelNet10（10 个标记朝向的子集数据）、ModelNet40（40 个类别的三维模型）、Aligned40（40 类标记的三维模型）。

（3）ShapeNet Part[77]：（https://cs.stanford.edu/~ericyi/project_page/part_annotation/）该数据集 2016 年由美国斯坦福大学，普林斯顿大学和芝加哥丰田技术学院的研究人员共同开发，该数据集是 ShapeNet 数据集的子集，一个由 3D CAD 模型对象表示的丰富注释的大型形状存储库，关注于细粒度的三维物体分割。包含 16 个类别的 16881 个形状 31693 个网格，每个形状类被标注为二到五个部分，整个数据集共有 50 个物体部分。

（4）S3DIS[78]：（http://buildingparser.stanford.edu/dataset.html）如图 9，该数据 2016 年由美国斯坦福大学的研究小组开发，是一个多模态、大规模室内空间数据集，具有实例级语义和几何注释。S3DIS 数据集覆盖超过 6,000 平方米，包含超过 70,000 个 RGB 图像，以及相应的深度，表面法线，语义注释，全局 XYZ 图像以及相机信息。收集在 6 个大型室内区域 272 个 3D 房间场景内。共有 13 个类别（墙、桌子、椅子、柜子等）。该数据集能够利用大规模室内空间中存在的规律来开发联合跨模式学习模型和潜在的无监督方法。

（5）Multisensorial Indoor Mapping and Positioning Dataset[79]：（http://mi3dmap.net/dataset.jsp）该数据集 2018 年由厦门大学的研究小组开发，数据通过多传感器获取，例如激光扫描仪，照相机，WIFI 和蓝牙等。该数据集提供了密集的激光扫描点云，用于室内制图和定位。同时，他们还提供基于多传感器校准和 SLAM 映射过程的彩色激光扫描。

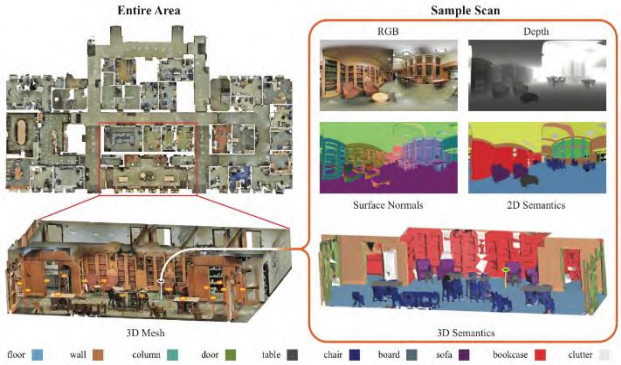

Fig.9 Example image of S3DIS dataset

图 9 S3DIS 数据集示例图

4.1.3 室外三维数据集

自 2009 年以来，已有多个室外三维数据集可用于三维点云的语义分割研究，然而，早期的数据集有很多缺点。例如：the Oakland outdoor MLS dataset[80], the Sydney Urban Objects MLS dataset[81], the Paris-rue-Madame MLS dataset[82], the IQmulus&TerraMobilita Contest MLS dataset[83] 和 ETHZ CVL RueMonge 2014 multiview stereo dataset[84]无法同时提供不同的对象表示和标注点。为了克服早期数据集的缺点，近年来已提供了新的基准数据。下面对这些数据集进行简单的描述。

（1）TUMCity Campus[85]：（https://www.iosb.fraunhofer.de/servlet/is/71820/）该数据集 2016 年由德国慕尼黑技术大学的 Fraunhofer IOSB 开发，在“TUM 城市校园”试验场(48.1493 °N，11.5685 °E)获得了移动激光扫描(MLS)数据，所有点的 x，y，z 都被地理参照到一个局部欧氏坐标系中。该数据集包含 17 亿多个点，9 个类别。随后，2017 年新增了一个红外图像序列来扩展数据集；2018 年对“MLS1-TUM 城市校园”三维测试数据集的一部分进行了手动标记；2019 年对“TUM 城市校园”试验场进行了重新扫描更新；2020 年新增了 2009 年的机载激光扫描(ALS)数据。

（2）vKITTI(Virtual KITTI)[86]：（http://www.europe.naverlabs.com/Research/Computer-Vision/Proxy-Virtual-Worlds）该数据集 2016 年由法国欧洲施乐研究中心计算机视觉小组和美国亚利桑那州立大学研究小

组共同开发，vKITTI 数据集是从真实世界场景的 KITTI 数据集模拟形成的大规模户外场景数据集，包含 13 个语义类别，35 个合成视频，总共约 17,000 个高分辨率帧，旨在学习和评估几个视频理解任务的计算机视觉模型：对象检测和多对象跟踪，场景级和实例级语义分割，光流和深度估计。2020 年研究人员对该数据集又进行了更新。

（3）Semantic3D[87]：（http://semantic3d.net/）如图 10，该数据集 2017 年由瑞士苏黎世联邦理工学院的研究小组开发，Semantic3D 提供了一个大型标记的三维点云数据集，其自然场景总数超过 40 亿个点。它还涵盖了一系列不同的城市场景：教堂、街道、铁轨、广场、村庄、城堡、足球场等。训练集和测试集各包含 15 个大规模的点云，8 个具体的语义类，扫描范围还包括各种场景类型，包括城市、次城市和农村，是目前最大的可用激光雷达数据集。

（4）Paris-Lille-3D[88]：（http://npm3d.fr/paris-lille-3d）该数据集 2018 年由巴黎高等矿业学院的研究小组开发，是一个城市 MLS 数据集，包含 1431 万个标记点，涵盖 50 个不同的城市对象类。整个数据集由三个子集组成，分别为 713 万、268 万和 457 万个点。作为 MLS 数据集，它也可以用于自动驾驶研究。

（5）Apollo[89]：（http://apolloscape.auto/car_instance.html）该数据集 2019 年由百度的研究小组开发，是一个大规模的自动驾驶数据集，提供了三维汽车的实例理解，LiDAR 点云对象检测和跟踪以及基于 LiDAR 的定位的标记数据。该数据集包含 5,277 个驾驶图像和超过 6 万个的汽车实例，其中每辆汽车都配备了具有绝对模型尺寸和语义标记关键点的行业级 3D CAD 模型。该数据集比 PASCAL3D 和 KITTI（现有技术水平）大 20 倍以上。

（6）SemanticKITTI[90]：（http://semantic-kitti.org/）如图 11，该数据集 2019 年由德国波恩大学的研究小组开发，是一个基于汽车 LiDAR 的大型户外场景数据集，SemancKITTI 由属于 21 个序列的 43552 个密集注释的激光雷达扫描组成，其中包含 19 个对象类别，序列 00-07 和 09-10 用于训练，序列 08 用于验证，序列 11-21 用于在线测试。该数据的原始 3D 点仅具有 3D 坐标，而没有颜色信息。

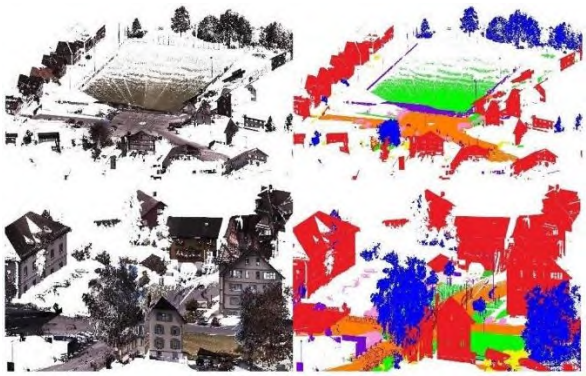

Fig.10 Point cloud scene and semantic segmentation diagram inSemantic3D dataset

图 10 Semantic3D 数据集中点云场景语义分割图

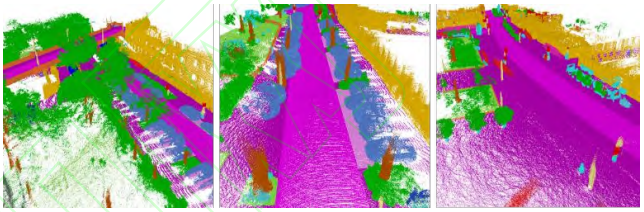

Fig.11 Semantic segmentation diagram in SemanticKITTI

图 11 SemanticKITTI 数据集中的语义分割图

4.1.4 遥感三维数据集

（1）Vaihingen point cloud semantic labeling dataset[91]：（http://www2.isprs.org/commissions/comm3/wg4/3d-semantic-labeling.html）该数据集 2014 年由德国汉诺威大学和达姆施塔特工业大学的研究者共同开发，它是遥感领域中第一个发布的基准数据集。该数据集是 ALS 点云的集合，由 Leica ALS50 系统捕获的 10 个条带组成，该条带的视场角为 45°，在德国 Vaihingen 的平均飞行高度为 500m。两个相邻条带之间平均重叠率为 30%左右，中点密度为 6.7 点每平方米。目前，该数据标记的点云被分为 9 个类别作为算法评估标准。

（2）The US3D Dataset[92]：（http://www.grss-ieee.org/community/technical-committees/data-fusion/2019-ieee-grss-data-fusion-contest/）如图 12，该数据集 2019 年由美国约翰 霍普金斯大学的研究小组开发，包括多视点、多波段卫星图像和两个大城市的地面真相、几何和语义标签的大规模公共数据集，超过 320GB 的数据用于训练和测试，覆盖了佛罗里达州杰克逊维尔和内布拉斯加州奥马哈的城区约 100 平方

公里，该数据集被用于 2019 年 IEEE GRSS 数据融合竞赛——大规模语义三维重建，比赛中的语义类包括建筑物、高架道路和桥梁、高植被、地面、水等。

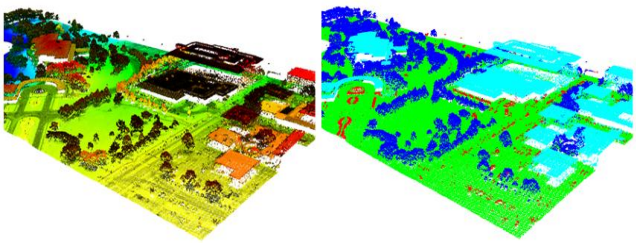

Fig.12 Point cloud scene and semantic segmentation diagram in The US3D dataset

图 12 The US3D 数据集中点云场景和语义分割图

（3）WHU-TLS[93]：（http://3s.whu.edu.cn/ybs/en/benchmark.htm）该数据集 2020 年由中国武汉大学、德国慕尼黑工业大学、芬兰大地所、挪威科技大学以及荷兰代尔夫特理工大学的研究小组共同开发。WHU-TLS 是全球最大规模和最多样化场景类型的 TLS 点云配准基准数据集，涵盖了地铁站、高铁站、山地、森林、公园、校园、住宅、河岸、文化遗产建筑、地下矿道、隧道等 11 种不同的环境，其中包含 115 个测站、17.4 亿个三维点以及点云之间的真实转换矩阵。该基准数据集也为铁路安全运营、河流勘测和治理、森林结构评估、文化遗产保护、滑坡监测和地下资产管理等应用提供了典型有效数据。

Table 1 Common 3D datasets of point cloud semantic segmentation.

表 1 点云语义分割常用的 3D 数据集

数据集名称	时间	应用	标注	类别	数据总量	传感器
ModelNet10[43]	2015	多种应用	3D CAD	10	4899models	----
ModelNet40[43]	2015	多种应用	3D CAD	40	12311models	----
ShapeNetPart[77]	2016	物体零件	pointwise	16	16881shapes	----
S3DIS[78]	2016	室内场景	pointwise	13	700million	----
Oakland3d[80]	2009	城市街道	pointwise	5	1.6million	SICK LMS
TUMCity[85]	2016	城市街道	voxel	9	1700million	Velodyne HDL-64E
KITTI[86]	2016	城市街道	bounding box	3	1799million	Velodyne HDL-64E
Semantic3D[87]	2017	城市街道	pointwise	8	4009million	Terrestrial Laser Scanner
Paris-Lille-3D[88]	2018	城市街道	pointwise	50	143miliion	Velodyne HDL-32E
SemanticKITTI[90]	2019	自动驾驶	pointwise	25	4549million	Velodyne HDL-64E
Vaihingen[91]	2014	户外场景	pointwise	9	1165598	Leica ALS50
US3D[92]	2019	户外场景	pointwise	6	320GB	----
WHU-TLS[93]	2020	多种应用	pointwise	----	1740million	VZ-400

4.2 实验结果分析与对比

为了评估三维语义分割算法的性能，需要借助通用的客观评价指标来保证算法评价的公正性。语义分割算法的实验性能评价标准主要分为以下几个方面：精确度、时间复杂度和内存损耗（空间复杂度）。

4.2.1 精确度

精确度是其中最为关键的指标，虽然现有的文献对语义分割成果采用了许多不同精度衡量的方法，如：平均准确率(mean accuracy,MA)、总体准确率(Overall Accuracy, OA)、平均交并比(mean intersection over union, mIoU)和带权交并比(frequency weighted intersection over union, FWIoU)，但本质上他们都是准确率及交并比(IoU)的变体。在精确度结果评价时，一般选取总体准确率(OA)和均

交并比(mIoU)两种评价指标综合分析，其中，mIoU表示数据分割的预测值与其真实值这两个集合的交集和并集之比，是目前语义分割领域使用频率最高和最常见的标准评价指标，其具体计算方法如公式(1)所示。假设共有 k+1 个类别（包括一个背景类），记P_{ij}是将i类预测为j类的点数，则P_{ii}表示真实值为i，预测值为i的点数；P_{ji}表示真实值为j，预测值为i的点数。

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

为便于对比实验结果和说明算法效果，本节将按照图 3 和图 6 中的分类基于深度学习的三维点云语义分割方法的实验结果进行分析与对比。表 2 列举了在具代表性的三维点云数据集上进行语义分割的方法的 mIoU(%)实验结果对比。主要比较各算法在五大类三维公共数据集的评价指标结果。“----”表示该方法未提供相应的结果。

Table 2　Experimental comparison of mIoU for methods of point cloud semantic segmentation

表 2　点云语义分割方法的 mIoU(%)实验结果对比

模型	年份	S3DIS (6-fold)	ScanNet	ShapeNet Part	Seman-tic3D	Semantic KITTI
SnapNet[38]	2017	----	----	----	59.1	----
TangentConv[94]	2018	52.8	43.8	----	----	40.9
GVCNN[33]	2018	----	----	----	----	----
SqueezeSeg[40]	2017	----	----	----	----	29.5
Squeezesegv2[41]	2018	----	----	----	----	39.7
SegCloud[47]	2017	48.92	----	79.4	61.3	----
Kd-Net[50]	2017	----	----	82.3	----	----
3DContextNet[51]	2018	55.6	----	84.3	----	----
VV-NET[52]	2018	78.22	----	87.4	----	----
PointNet[5]	2017	47.6	----	83.7	----	14.6
PointNet++[17]	2017	54.5	33.9	85.1	63.1	20.1
PointWeb[54]	2019	66.7	----	----	----	----
PointSIFT[10]	2018	70.23	41.5	----	----	----
SO-Net[11]	2018	----	----	84.6	----	----
SCN[55]	2018	52.72	----	84.6	----	----
RandLA-Net[18]	2020	68.5	----	----	76	50.3
DGCNN[24]	2018	56.1	----	85.1	----	----
RGCNN[25]	2018	----	----	84.3	----	----
HDGCN[12]	2019	66.9	----	----	----	----
GACNet[19]	2019	62.85	----	----	70.8	----
DPAM[26]	2019	64.5	----	86.1	----	----
PointNGCNN[58]	2020	----	----	85.6	----	----
SPG[59]	2017	62.1	----	----	73.2	17.4
PointCNN[14]	2018	65.39	45.8	86.1	----	----
PCCN[29]	2018	58.3	----	----	----	----
KPConv[21]	2019	70.6	68.4	86.4	74.6	----
ConvPoint[27]	2019	68.2	----	85.8	76.5	----
SpiderCNN[30]	2018	----	----	85.3	----	----
InterpCNN[22]	2019	66.7	----	86.3	----	----
ShellNet[15]	2019	66.8	----	82.8	69.4	----

PointConv[23]	2019	----	55.6	85.7	----	----
A-CNN[61]	2019	62.9	----	86.1	----	----
CU&RCU[28]	2018	49.7	----	----	----	----
RSNet[13]	2018	56.47	39.35	84.9	----	----
3P-RNN[20]	2018	56.3	----	----	----	
GSA[64]	2019	64.28	----	----	----	----
ASR[65]	2019	----	----	85.6	----	----
ASIS[67]	2019	51.1	----	----	----	----
JSIS3D[68]	2019	----	----	----	----	----

表 2 中可以发现，三维公共数据集中 ShapeNetPart 和 S3DIS 这两个数据集运用得最多，ShapeNet Part 是一个由 3D CAD 模型对象表示的丰富注释的大型形状存储库，关注于细粒度的三维物体分割。S3DIS 是一个多模态、大规模的室内空间数据集，具有实例级语义和几何注释。

选用 ShapeNet Part 数据集的算法中，分割效果都很好，mIoU 基本均在 80%以上，说明目前已有的算法对细粒度的三维物体有较好的识别效果，物体分割结果能够接近真实的分割。由于 S3DIS 数据集的数据量庞大，所以大部分算法的分割效果不明显，mIoU 都普遍较低，其中将点云体素化的 VV-NET 网络表现突出，该网络使用基于内核的内插变分自动编码器(VAE)结构对每个提速中的局部几何进行编码，同时利用径向基函数(RBF)计算每个体素内的局部连续表示以处理点的稀疏分布。此外，作者将 RBF-VAE 与 group-conv 相结合发现该方法比仅使用 group-conv 或仅使用 RBF-VAE 取得了更好的性能。

表 2 中，SnapNet、SegCloud、PointNet++、GACNet、KPConv、ConvPoint、RandLA-Net 和 SPG 等算法均选用了 Semantic3D 城市场景数据集，这些算法可运用在大场景中进行语义分割，其中 2017 年提出的 SPG 网络表现突出，在几亿点的场景下，评价指标可达到 73.2%，是目前运用于大场景分割中最有效的分割网络之一。不难发现，近些年提出的基于优化 CNN 的算法在各类公共数据集上的表现均较为优异，进一步优化卷积，并将其集成到各种优秀的网络架构中，将会是未来研究的一个热点方向。

SemanticKITTI 作为一个基于汽车 LiDAR 的大型户外场景数据集，可运用于汽车的无人驾驶中，目前实现 SemanticKITTI 数据集语义分割的算法中，RandLA-Net 的表现最为突出。RandLA-Net 网络不需要任何前/后处理步骤（如体素化、块分割或图形构建），能够直接处理大规模三维点云，相比于现有的大规模点云语义分割方法，其分割速率提升近 200 倍。

4.2.2 复杂度

复杂度是对模型性能检测的另一个有价值且重要的度量指标，包括时间复杂度和空间复杂度。随着语义分割技术的发展和数据处理能力的提高，该技术应用面更加广泛，除了运用复杂的网络提高算法的分割准确率外，现实中的应用程序（如行人检测、自动驾驶等）更需要实时高效的分割网络。因此，本节从时间复杂度（运行速率）和空间复杂度（参数数量）两方面考察了部分网络的实时性。

表 3 中根据参数数量和转发时间评估了模型的复杂度。该实验对比在 1080X GPU 的硬件环境下进行，针对 ModelNet40 数据集，批次大小设置为 8。对于参数数量指标，ShellNet 优于现有的方法，虽然在空间上没有那么复杂，但是 ShellNet 仍然可以非常有效地收敛到最先进的精度。另外，从表 3 中不难发现，RGCNN 具有最快的推算时间和可接受的模型大小，适用于实时任务。为了进一步减少模型大小和推断时间，在 PointNet 和 DPAM 模型中均尝试删除了模型使用的 T-net（即表 3 中以 vanilla 表示），其中 DPAM 仅在模型精度降低 0.5%的情况下，即可实现更小的模型尺寸和更快的推算时间。

doi: 10.3778/j.issn.1673-9418.2006025

Table 3　Time and Space Complexity Analysis on ModelNet40

表 3　各类算法在 ModelNet40 数据集上的时空复杂度分析

Method	Parameters (M)	Infer (ms)	Acc (%)
PointNet(Vanilla)[5]	0.8	11.6	87.2
PointNet[5]	3.48	25.3	89.2
MVCNN[31]	60	----	90.1
PointNet++[17]	12.4	163.2	90.7
SO-Net[11]	----	59.6	90.9
SpecGCN[56]	2.05	----	91.5
RGCNN[25]	----	7.5	90.5
DGCNN[24]	1.84	94.6	92.2
DPAM(Vanilla)[26]	----	18.4	91.4
DPAM[26]	----	36.6	91.9
PointCNN[14]	0.6	----	92.2
ShellNet[15]	0.48	----	93.1
RSNet[13]	0.76	----	----
PAT[64]	----	88.6	91.7

表中的 Vanilla 表示在不使用 T-Net 的情况下训练的模型，表中 Parameters 表示参数量，以百万（million, M）为单位；infer 表示推算时间（inference time），以毫秒（millisecond, ms）为单位；Acc 表示精确度（accuracy）。

表 4 定量地显示了不同方法的总时间和内存消耗。该实验对比在 RTX2080Ti GPU 的硬件环境下进行，针对 SemanticKITTI 数据集。从表 4 中可以看出，SPG 网络参数最少，但处理点云的时间最长，原因是几何划分和超图构造步骤繁琐；PointNet++ 和 PointCNN 的计算开销也很大，主要是由于 FPS 的采样操作；PointNet 和 KPConv 由于内存操作效率低，无法一次通过获取超大规模的点云；而 RandLA-Net 基于简单的随机抽样和高效的局部特征聚合器，实现了用较短的时间来推断每个大规模点云的语义标签。

Table 4　Time and Space Complexity Analysis on SemanticKITTI dataset

表 4　各类算法在 SemanticKITTI 数据集上的时空复杂度分析

Method	Total time(s)	Parameters (M)	Max Infer (M)
PointNet(Vanilla)[5]	192	0.8	0.49
PointNet++(SSG)[17]	9831	0.97	0.98
RandLA-Net[18]	176	11	1.15
SPG[59]	43584	0.25	----
PointCNN[14]	8142	14.9	0.05
KPConv[21]	717	0.95	0.54

表中的 Total time 表示总运行时间，以秒（second, s）为单位；Max Infer 表示最大的输入点数（Maximum inference points），以百万（million, M）为单位。

5　展望

现有的方法在很大程度上提高了语义分割的精度，但仍存在一定局限性，因此，如何解决这些局限性是未来研究的热点，本章基于前面章节对应用深度学习技术解决语义分割问题的研究评述，对语义分割领域未来研究方向进行了展望。

（1）训练数据库和应用场景

基于深度学习的语义分割方法需要海量的数据库作为支撑，目前已有的数据集并不能满足语义分割发展的需求，所以构建数据量丰富、有效且全面的数据集是目前语义分割的首要条件。而且，现有的三维数据集大部分局限在室内场景以及城市街道场景，对于有标注且内容丰富的户外点云场景数据集及遥感三维数据集相对较少，建立一整套作为基准点的数据集十分重要。另外，SqueezeSeg V2[41] 算法为了避免收集和注释的成本，使用诸如 GTA-V 之类的模拟器来创建无限数量的标记的合成数据，为补充预训练数据集的方法提供了思路，但是这类合成的仿真数据仍需解决域迁移的问题。

（2）序列数据集

三维大规模数据集缺乏的问题同样影响到了视频序列分割，目前基于序列的可用数据集较少，导致针对视频数据的语义分割方法研究进展缓慢。带

有时间序列的视频数据在语义分割过程中可以利用其时空序列信息提供高阶特征，进而提高准确率和效率。

（3）全景分割

全景分割由 Kirillov 等人[95]提出，全景分割是将前景和背景分开来分割的，对目标区域（前景对象）做实例分割，对背景区域做语义分割。2019 年，Kirillov 等人[96]将分别用于语义分割和实例分割的 FCN 和 Mask R-CNN 结合起来，设计了 Panoptic FPN，实验证明 Panoptic FPN 对语义分割和实例分割两个任务都有效，同时兼具稳健性和准确性。但是在合并过程中，如果没有足够的上下文信息，很难确定对象实例之间的重叠关系。针对这一问题，Liu[97]等人提出了一种端到端的遮挡感知网络(OcclusionAware Network,OANet)用于全景分割，该网络可有效地预测单个网络的实例分割和实体分割。DeeperLab[98]是一种单镜头，自下而上的图像解析器，该网络使用全卷积网络生成每像素的语义和实例预测，然后通过合并启发式算法将这些预测融合到最终的图像解析结果中。虽然上述几种方法在 Cityscapes[99]、COCO Stuff[100]等数据集上获得了较为可观的精度，但分割过程中仍然需要进行复杂的实例掩码预测(instance mask predictions)或合并启发式算法(merging heuristics)，很难实现模型的实时性需求。FPSNet[101]的提出有效地解决了这个问题，该网络使用自定义的密集像素分类任务（为每个像素分配一个类标签或一个实例 id）代替复杂的全景任务，实现了分割速度的提升。上述的全景分割操作主要是针对图像进行的，目前对点云数据进行全景分割的研究很少，如 ASIS[67]、JSIS-Net[68]使用两个并行的分支分别进行实例分割和语义分割，然后融合两个结果作为输出。另外，3D 全景分割数据集 SemanticKITTI[90]的提出，将高质量的全景分割引入至机器人和智能车辆的实时应用方面迈出了重要一步。全景分割作为计算视觉一个新的任务场景，其在三维数据的应用前景仍有待挖掘与探索。

（4）实时分割

目前提出的语义分割网络模型在分割精度上已经取得了很大的进展，然而却增加了模型的复杂度和运行速率。随着自动驾驶、行人检测和环境感知等应用领域的发展，对语义分割实时性的要求也越来越高。因此，在维持高准确率的同时，降低模型复杂度，缩短响应时间，实现实时分割，是未来重要的工作方向。

（5）遥感领域

在过去的十年里，深度学习推动了遥感影像语义分割的进步，但遥感点云语义分割的发展还相对不太成熟。目前已发表的计算机视觉算法通常在对象类别有限的小区域数据集上进行测试，但是对于遥感应用，需要具有更复杂和特定地面对象类别的大面积数据。而且，计算机视觉算法的精度评价体系并不完全适用于遥感应用，遥感应用更关心特定目标的精度。例如：在城市管理监测中，对于建筑物语义分割的准确性至关重要。随着三维遥感语义分割应用需求的不断提升，能够学习对象语义特征和分类三维遥感数据的算法成为研究者们未来的一个研究热点。

（6）弱监督或无监督语义分割技术

弱监督方法使用轻量级的弱监督标注数据进行训练，减少了标注成本和标注时间，在图像语义分割中已经有了很大的进展。目前，三维数据库需求量大，标注困难，若弱监督或无监督的语义分割技术能够应用到三维点云语义分割中，不仅能解决数据问题，而且在提高网络模型的精度的同时实现速率的提升，将会是未来发展的趋势。

（7）迁移学习

一个完整的语义分割深度神经网络训练需要足够数量的数据集，初始化权重的调试以及长时间的收敛过程。通过继续训练过程来微调预训练网络的权重是主要的迁移学习方法之一，因此，为了提高效率，部分学者会选择预先训练的权重而不是随机初始化的权重。另外，PointNet[5]、PointNet++[17]网络的提出为点云语义分割提供了完整的体系结构，为实现迁移学习提供了前提条件，PointSIFT[10]是一个通用模块，可以集成到各种基于 PointNet 的体系结构中以改善 3D 形状表示；DPAM[26]可以插入大多数现有体系结构中构建分层的学习体系结构；ASR[65]模块可以轻松集成到现有的深度网络中，通过将相邻点的分数与学习的注意力权重合并在一起，对网络产生的分割结果进行后处理，与 CRF 的功能

类似；Francis[102]等人提出的扩张点卷积（Dilated Point Convolutions, DPC）运算代替 K-最近邻域方法，以汇总扩张的邻近要素，此操作不仅增加了接收范围，并且可以轻松地集成到现有的基于聚合的网络中。迁移学习已在点云语义分割领域得到了广泛的应用，未来对迁移学习的研究可以关注以下几点：（1）通过半监督学习减少对标注数据的依赖，应对标注数据的不对称性；（2）使用迁移学习来做到持续学习，让神经网络得以保留在旧任务中所学到的能力；（3）使用迁移学习来提高模型的稳定性和可泛化性等。

（8）各类技术的参考性价值

从边缘特征的角度：利用有意义的边缘特征，并将边缘特征馈送到点特征中以提供上下文信息，有助于点云语义理解。如：PCCN[29]自适应地从边缘学习权重以融合点特征；KCNet[103]定义点集内核和内核相关性以沿边缘聚合局部特征；Jiang 等人[104]设计了一种分层点--边缘的交互网络，通过将每个点特征与最大池化相对应的边缘特征连接在一起。

从自动编码器的角度：自动编码器(Autoencoders, AEs)是一种无监督的神经网络模型，目前，自动编码器已被广泛应用于生成图像语义分割模型来表示数据，一些研究者发现，自动编码器对于不规则的三维点云同样适用，并且可在上采样阶段解决点云的稀疏性问题。Zhao 等人[105]基于 2D 胶囊网络(CN)提出了一种无监督的自动编码器 3DPointCapsNet，用于处理稀疏 3D 点云，同时保留输入数据的空间排列，并在零件分割中取得了不错的进展。

从零样本学习(zero-shot learning)的角度：零样本学习[106]具有识别训练数据集中未观察到的类别的能力。获取特征图后，零样本学习可以将语义嵌入用于诸如对象检测之类的应用程序。在特征融合的方法中，模型提取了点云的局部特征和全局特征，而这些模型可用作零镜头学习中的特征提取器，这将有助于使用稀缺的数据集学习权重。

从过分割(oversegmentation)的角度：过分割可作为点云语义分割中的一种预分割算法，其具有降低数据量和光精度损失的作用。Landrieu 等人[107]提出了第一个将三维点云过分割为超点的监督学习框架，将点云过分割表述为一个由邻接图构造的深度度量学习问题。利用一种图形结构的对比损失，学习将三维点均匀地嵌入对象中，从而使对象的边界呈现出高对比度。

从多形态融合的角度：目前的语义分割可以将不规则的点云或者网格数据转换为常规的三维体素网格或者多视图。也可以直接在点云数据上进行分割。为了进一步利用可用信息，可通过多形态融合的方式从不同形态的数据中分别提取点云特征。Jaritz 等人[108]提出多视图点网(MVPNet)，以聚合二维多视图图像的外观特征和规范点云空间中的空间几何特征。

从 RNN 中长短时间记忆(LSTM)的角度：LSTM 具有几个语义分割模型所需的属性，如：可以端到端进行微调，并且允许输入和输出中的可变长度。二维图像语义分割中，Li 等人[109]提出的 LSTM-CF(Long Short-Term Memorized Context Fusion)网络，该网络利用基于 LSTM 的融合层整合竖直方向上的光度和深度通道的上下文信息，完成网络端到端的训练和测试。

从时空信息的角度：目前已有研究开始从动态点云中学习时空信息，未来可以尝试通过时空信息提高点云语义分割模型的性能。Liu 等人[110]提出 MeteorNet，直接对动态点云进行处理，学习从时空相邻点聚合信息。

6 结束语

本文综述了基于深度学习的点云语义分割的研究现状，虽然三维深度学习是一个相对较新的领域，但综述的内容显示了一个快速增长和高效的群体。虽然三维深度学习没有二维深度学习成熟，但不难发现，这一差距正在缩小。本文从语义分割的应用和深度学习的发展出发，对三维点云进行了详细的介绍，将三维深度学习语义分割方法分为间接语义分割方法和直接语义分割方法两大类，从算法特点以及模型结构方面梳理了一些较为突出的方法，并进行了较为细致的分类、介绍和评估。此外，本文回顾了用于网络评估的现代基准数据集。最后，本文结合上述内容，对未来工作方向以及该领域一些

开放问题提出了一些展望。深度学习技术被证明可有效解决语义分割问题，并且在语义分割领域许多优秀的方法也不断地推进。因此，我们期待在未来几年各种创新的研究思路不断涌现。

References:

[1] Zhang J, Lin X, Ning X. SVM-based classification of segmented airborne LiDAR point clouds in urban areas[J]. Remote Sensing, 2013, 5(8): 3749-3775.

[2] Sun J, Lai ZL. Airbone LiDAR feature selection for urban classification using random forests[J]. Geomatics and Information Science of Wuhan University, 2014, 39(11): 1310-1313.

[3] Zhuang Y, Liu Y, He G, et al. Contextual classification of 3D laser points with conditional random fields in urban environments[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015: 3908-3913.

[4] Lu Y, Rasmussen C. Simplified Markov random fields for efficient semantic labeling of 3D point clouds[C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2012: 2690-2697.

[5] 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.

[6] Griffiths D, Boehm J. A Review on deep learning techniques for 3D sensed data classification[J]. Remote Sensing, 2019, 11(12): 1499.

[7] Xie Y, Tian J, Zhu X X. A review of point cloud semantic segmentation[J]. arXiv preprint arXiv:1908.08854, 2019.

[8] Zhang JY, Zhao XL, Chen Z. Review of Semantic Segmentation of Point Cloud Based on Deep Learning[J]. Laser & Optoelectronics Progress, 2020, 57(4): 040002.

[9] 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.

[10] Jiang M, Wu Y, Zhao T, et al. Pointsift: A sift-like network module for 3d point cloud semantic segmentation[J]. arXiv preprint arXiv:1807.00652, 2018.

[11] Li J, Chen B M, Hee Lee G. So-net: Self-organizing network for point cloud analysis[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 9397-9406.

[12] Liang Z, Yang M, Deng L, et al. Hierarchical depthwise graph convolutional neural network for 3d semantic segmentation of point clouds[C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 8152-8158.

[13] Huang Q, Wang W, Neumann U. Recurrent slice networks for 3d segmentation of point clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2626-2635.

[14] L Li Y, Bu R, Sun M, et al. Pointcnn: Convolution on x-transformed points[C]//Advances in neural information processing systems, 2018: 820-830.

[15] Zhang Z, Hua B S, Yeung S K. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019: 1607-1616.

[16] Zhang J, Zhao X, Chen Z, et al. A Review of Deep Learning-based Semantic Segmentation for Point Cloud (November 2019)[J]. IEEE Access, 2019.

[17] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[C]// Advances in neural information processing systems, 2017: 5099-5108.

[18] 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.

[19] Wang L, Huang Y, Hou Y, et al. Graph attention convolution for point cloud semantic segmentation[C]//

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 10296-10305.

[20] Ye X, Li J, Huang H, et al. 3d recurrent neural networks with context fusion for point cloud semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 403-417.

[21] Thomas H, Qi C R, Deschaud J E, et al. Kpconv: Flexible and deformable convolution for point clouds[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019: 6411-6420.

[22] Mao J, Wang X, Li H. Interpolated convolutional networks for 3d point cloud understanding[C]// Proceedings of the IEEE International Conference on Computer Vision, 2019: 1578-1587.

[23] Wu W, Qi Z, Fuxin L. Pointconv: Deep convolutional networks on 3d point clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 9621-9630.

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

[25] Te G, Hu W, Zheng A, et al. Rgcnn: Regularized graph cnn for point cloud segmentation[C]//Proceedings of the 26th ACM international conference on Multimedia, 2018: 746-754.

[26] Liu J, Ni B, Li C, et al. Dynamic points agglomeration for hierarchical point sets learning[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019: 7546-7555.

[27] Boulch A. ConvPoint: Continuous convolutions for point cloud processing[J]. Computers & Graphics, 2020.

[28] Engelmann F, Kontogianni T, Hermans A, et al. Exploring spatial context for 3d semantic segmentation of point clouds[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 716-724.

[29] Wang S, Suo S, Ma W C, et al. Deep parametric continuous convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2589-2597.

[30] Xu Y, Fan T, Xu M, et al. Spidercnn: Deep learning on point sets with parameterized convolutional filters[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 87-102.

[31] 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, 2015: 945-953.

[32] Qi C R, Su H, Nießner M, et al. Volumetric and multi-view cnns for object classification on 3d data[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 5648-5656.

[33] Feng Y, Zhang Z, Zhao X, et al. GVCNN: Group-view convolutional neural networks for 3D shape recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 264-272.

[34] Zeng A, Yu K T, Song S, et al. Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge[C]//2017 IEEE international conference on robotics and automation (ICRA). IEEE, 2017: 1386-1383.

[35] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems, 2012: 1097-1105..

[36] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[37] Ma L, Stückler J, Kerl C, et al. Multi-view deep learning for consistent semantic mapping with rgb-d cameras[C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017: 598-605.

[38] 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.

[39] Guerry J, Boulch A, Le Saux B, et al. Snapnet-r: Consistent 3d multi-view semantic labeling for robotics[C]// Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 669-678.

[40] Wu B, Wan A, Yue X, et al. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 1887-1893.

[41] Wu B, Zhou X, Zhao S, et al. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud[C]// 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 4376-4382.

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

[43] Wu Z, Song S, Khosla A, et al. 3d shapenets: A deep representation for volumetric shapes[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1912-1920..

[44] Dai A, Chang A X, Savva M, et al. Scannet: Richly-annotated 3d reconstructions of indoor scenes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5828-5839.

[45] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2016: 424-432.

[46] Li Y, Pirk S, Su H, et al. Fpnn: Field probing neural networks for 3d data[C]//Advances in Neural Information Processing Systems, 2016: 307-315.

[47] Tchapmi L, Choy C, Armeni I, et al. Segcloud: Semantic segmentation of 3d point clouds[C]//2017 international conference on 3D vision (3DV). IEEE, 2017: 537-547.

[48] Riegler G, Osman Ulusoy A, Geiger A. Octnet: Learning deep 3d representations at high resolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3577-3586.

[49] Xu Y, Hoegner L, Tuttas S, et al. Voxel- and Graph-Based Point Cloud Segmentation of 3D Scenes Using Perceptual Grouping Laws[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017: 43-50.

[50] Klokov R, Lempitsky V. Escape from cells: Deep kd-networks for the recognition of 3d point cloud models[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 863-872.

[51] Zeng W, Gevers T. 3DContextNet: Kd tree guided hierarchical learning of point clouds using local and global contextual cues[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 0-0.

[52] Meng H Y, Gao L, Lai Y K, et al. VV-Net: Voxel vae net with group convolutions for point cloud segmentation[C]// Proceedings of the IEEE International Conference on Computer Vision, 2019: 8500-8508.

[53] Wang L, Huang Y, Shan J, et al. MSNet: Multi-scale convolutional network for point cloud classification[J]. Remote Sensing, 2018, 10(4): 612.

[54] Zhao H, Jiang L, Fu C W, et al. PointWeb: Enhancing local neighborhood features for point cloud processing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5565-5573.

[55] Xie S, Liu S, Chen Z, et al. Attentional shapecontextnet for point cloud recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4606-4615.

[56] Wang C, Samari B, Siddiqi K. Local spectral graph convolution for point set feature learning[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 52-66.

[57] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3693-3702.

[58] Lu Q, Chen C, Xie W, et al. PointNGCNN: Deep convolutional networks on 3D point clouds with neighborhood graph filters[J]. Computers & Graphics, 2020, 86: 42-51.

[59] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4558-4567.

[60] Hua B S, Tran M K, Yeung S K. Pointwise convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 984-993.

[61] Komarichev A, Zhong Z, Hua J. A-CNN: Annularly convolutional neural networks on point clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7421-7430.

[62] Himmelsbach M, Hundelshausen F V, Wuensche H J. Fast segmentation of 3D point clouds for ground vehicles[C]// 2010 IEEE Intelligent Vehicles Symposium. IEEE, 2010: 560-565.

[63] Liu F, Li S, Zhang L, et al. 3DCNN-DQN-RNN: A deep reinforcement learning framework for semantic parsing of large-scale 3D point clouds[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 5678-5687.

[64] Yang J, Zhang Q, Ni B, et al. Modeling point clouds with self-attention and gumbel subset sampling[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3323-3332.

[65] Zhao C, Zhou W, Lu L, et al. Pooling scores of neighboring points for improved 3D point cloud segmentation[C]//2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 1475-1479.

[66] Chen C, Fragonara L Z, Tsourdos A. GAPNet: Graph attention based point neural network for exploiting local feature of point cloud[J]. arXiv preprint arXiv:1905.08705, 2019.

[67] Wang X, Liu S, Shen X, et al. Associatively segmenting instances and semantics in point clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 4096-4105.

[68] Pham Q H, Nguyen T, Hua B S, et al. JSIS3D: joint semantic-instance segmentation of 3d point clouds with multi-task pointwise networks and multi-value conditional random fields[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 8827-8836.

[69] Lai K, Bo L, Ren X, et al. A large-scale hierarchical multi-view rgb-d object dataset[C]//2011 IEEE international conference on robotics and automation, IEEE, 2011: 1817-1824.

[70] Silberman N, Hoiem D, Kohli P, et al. Indoor segmentation and support inference from rgbd images[C]// European conference on computer vision. Springer, Berlin, Heidelberg, 2012: 746-760.

[71] Xiao J, Owens A, Torralba A. Sun3d: A database of big spaces reconstructed using sfm and object labels[C]//Proceedings of the IEEE International Conference on Computer Vision. 2013: 1625-1632.

[72] Singh A, Sha J, Narayan K S, et al. Bigbird: A large-scale 3d database of object instances[C]//2014 IEEE international conference on robotics and automation (ICRA). IEEE, 2014: 509-516.

[73] Martínez-Gómez J, García-Varea I, Cazorla M, et al. ViDRILO: The visual and depth robot indoor localization with objects information dataset[J]. The International Journal of Robotics Research, 2015, 34(14): 1681-1687.

[74] Song S, Lichtenberg S P, Xiao J. Sun rgb-d: A rgb-d scene understanding benchmark suite[C]//Proceedings of the IEEE conference on computer vision and pattern recog-

nition. 2015: 567-576.

[75] Chang A, Dai A, Funkhouser T, et al. Matterport3d: Learning from rgb-d data in indoor environments[J]. arXiv preprint arXiv:1709.06158, 2017.

[76] Chen X, Golovinskiy A, Funkhouser T. A benchmark for 3D mesh segmentation[J]. Acm transactions on graphics (tog), 2009, 28(3): 1-12.

[77] 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 (TOG), 2016, 35(6): 1-12.

[78] Armeni I, Sener O, Zamir A R, et al. 3d semantic parsing of large-scale indoor spaces[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1534-1543.

[79] Wang C, Hou S, Wen C, et al. Semantic line framework-based indoor building modeling using backpacked laser scanning point cloud[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 143: 150-166.

[80] Munoz D, Bagnell J A, Vandapel N, et al. Contextual classification with functional max-margin markov networks[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 975-982.

[81] De Deuge M, Quadros A, Hung C, et al. Unsupervised feature learning for classification of outdoor 3d scans[C]// Australasian Conference on Robitics and Automation. 2013, 2: 1.

[82] Serna A, Marcotegui B, Goulette F, et al. Paris-rue-Madame database: a 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods[C]//international conference on pattern recognition applications and methods, 2014: 819-824.

[83] Brédif M, Vallet B, Serna A, et al. TerraMobilita/IQmulus urban point cloud classification benchmark[C]//in Workshop on Processing Large Geospatial Data, 2014.

[84] Riemenschneider H, Bódis-Szomorú A, Weissenberg J, et al. Learning where to classify in multi-view semantic segmentation[C]//European Conference on Computer Vision. Springer, Cham, 2014: 516-532.

[85] Gehrung J, Hebel M, Arens M, et al. An approach to extract moving objects from MLS data using a volumetric background representation[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, 4: 107.

[86] Gaidon A, Wang Q, Cabon Y, et al. Virtual worlds as proxy for multi-object tracking analysis[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4340-4349.

[87] Hackel T, Savinov N, Ladicky L, et al. Semantic3d. net: A new large-scale point cloud classification benchmark[J]. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017: 91-98.

[88] Roynard X, Deschaud J E, Goulette F. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification[J]. The International Journal of Robotics Research, 2018, 37(6): 545-557.

[89] Song X, Wang P, Zhou D, et al. Apollocar3d: A large 3d car instance understanding benchmark for autonomous driving[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5452-5462.

[90] Behley J, Garbade M, Milioto A, et al. SemanticKITTI: A dataset for semantic scene understanding of lidar sequences[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 9297-9307.

[91] Niemeyer J, Rottensteiner F, Soergel U. Contextual classification of lidar data and building object detection in urban areas[J]. ISPRS journal of photogrammetry and remote sensing, 2014, 87: 152-165.

[92] Bosch M, Foster K, Christie G, et al. Semantic stereo for incidental satellite images[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019: 1524-1532.

[93] Dong Z, Liang F, Yang B, et al. Registration of large-scale

terrestrial laser scanner point clouds: A review and benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 163: 327-342.

[94] Tatarchenko M, Park J, Koltun V, et al. Tangent convolutions for dense prediction in 3d[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3887-3896.

[95] Kirillov A, He K, Girshick R, et al. Panoptic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2019: 9404-9413.

[96] Kirillov A, Girshick R, He K, et al. Panoptic feature pyramid networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 6399-6408.

[97] Liu H, Peng C, Yu C, et al. An end-to-end network for panoptic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 6172-6181.

[98] Yang T, Collins M D, Zhu Y, et al. DeeperLab: Single-Shot Image Parser[J]. arXiv: Computer Vision and Pattern Recognition, 2019.

[99] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 3213-3223.

[100] Caesar H, Uijlings J, Ferrari V. Coco-stuff: Thing and stuff classes in context[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1209-1218.

[101] de Geus D, Meletis P, Dubbelman G. Fast panoptic segmentation network[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 1742-1749.

[102] Engelmann F, Kontogianni T, Leibe B. Dilated point convolutions: On the receptive field of point convolutions[J]. arXiv preprint arXiv:1907.12046, 2019.

[103] Shen Y, Feng C, Yang Y, et al. Mining point cloud local structures by kernel correlation and graph pooling[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 4548-4557.

[104] Jiang L, Zhao H, Liu S, et al. Hierarchical point-edge interaction network for point cloud semantic segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 10433-10441.

[105] Zhao Y, Birdal T, Deng H, et al. 3D point capsule networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 1009-1018.

[106] Cheraghian A, Rahman S, Petersson L. Zero-shot learning of 3d point cloud objects[C]//2019 16th International Conference on Machine Vision Applications (MVA). IEEE, 2019: 1-6.

[107] Landrieu L, Boussaha M. Point cloud oversegmentation with graph-structured deep metric learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7440-7449.

[108] Jaritz M, Gu J, Su H. Multi-view pointnet for 3D scene understanding[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. 2019: 0-0.

[109] Li Z, Gan Y, Liang X, et al. Lstm-cf: Unifying context modeling and fusion with lstms for rgb-d scene labeling[C]//European conference on computer vision. Springer, Cham, 2016: 541-557.

[110] Liu X, Yan M, Bohg J. MeteorNet: Deep learning on dynamic 3D point cloud sequences[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 9246-9255.

附中文参考文献:

[2]孙杰, 赖祖龙. 利用随机森林的城区机载 LiDAR 数据特征选择与分类. 武汉大学学报信息科学版, 2014, 39(11): 1310-1313.

[8]张佳颖, 赵晓丽, 陈正. 基于深度学习的点云语义分割综述. 激光与光电子学进展, 2020, 57(4):040002.

附表：

模型	年份	简介	数据集对象	关键技术
PointNet[5]	2016	基于深度学习直接处理点云语义分割方法的开拓者	ShapeNet、S3DIS	对称函数、多层感知机、T-net 网络
PointNet++[17]	2017	PointNet 的分层版本	ScanNet、ModelNet	最远点采样、球查询、多分辨率组合
PointWeb[54]	2019	从点云的局部邻域中提取上下文特征的局部完全链接网络	S3DIS、ScanNet、ModelNet	自适应特性调整（AFA）模块
PointSIFT[10]	2018	可嵌入各种网络框架的模块	S3DIS	SIFT（方向编码和尺度感知）
SO-Net[11]	2018	有效利用点云空间分布的置换不变性网络	ModelNet、ShapeNet	自组织映射
SCN[55]	2018	基于形状上下文的端到端可训练架构	ModelNet、ShapeNet、S3DIS	形状上下文
RandLA [18]	2020	用于大规模点云处理的轻量级网络	SemanticKITTI	随机点采样法、局部特征聚集模块
DGCNN[24]	2018	网络架构与 PointNet 几乎类似，将堆叠多层感知机转换为边缘卷积	ModelNet、ShapeNet、S3DIS	EdgeConv
RGCNN[25]	2018	直接使用不规则点云的正则化图卷积神经网络	ShapeNet、ModelNet	谱图理论、切比雪夫多项式、图信号平滑先验
HDGCN[12]	2019	用于点云语义分割的分层深度图卷积神经网络	S3DIS、Paris-Lille	深度图卷积
GACNet[19]	2019	端到端的图形注意卷积网络	S3DIS、Semantic3D	图形注意卷积
DPAM[26]	2019	通过堆叠多个动态点聚集模块构建的分层点集学习体系结构	ModelNet、ShapeNet、S3DIS	动态点聚集模块
PointNGCNN[58]	2020	利用点云进行三维目标识别和分割的端到端深度学习网络	ModelNet、ScanNetShapeNet、S3DIS、	邻域图、切比雪夫多项式、拉普拉斯矩阵
SPG[59]	2017	用于百万点的大规模点云语义分割的深度学习框架	Semantic3D、S3DIS	超点图
CU&RCU[28]	2018	基于 PointNet 可扩展到包含更大尺度的空间环境的网络	Virtual KITTI、S3DIS	输出级上下文、输入级上下文
3DCNN-DQN-RNN[63]	2018	融合三维卷积神经网络（CNN），深层 Q-网络（DQN）和残差递归神经网络（RNN）的大规模点云的语义解析网络	S3DIS、SUNCG	3D CNN、DQN、RNN
RSNet[13]	2018	对点云中的局部结构进行建模的三维分割框架	S3DIS、ScanNet、ShapeNet	双向 RNN 单元
3P-RNN[20]	2018	融合局部空间结构和长依赖上下文的端到端的非结构化点云语义分割框架	S3DIS、ScanNet、vKITTI	金字塔池化模型、分层的双向 RNN 模块
PointCNN[14]	2018	从不规则和无序的点云中学习特性的语义分割框架	ShapeNet、S3DIS 、ScanNet	χ-Conv
ConvPoint[27]	2019	离散卷积神经网络的泛化	ShapeNet、S3DIS	连续卷积
PCCN[29]	2018	可应用于室外和室内三维点云分割任务以及驾驶场景中的激光雷达运动估计的参数连续卷积算法	S3DIS	参数连续卷积
InterpCNN[22]	2019	直接处理不规则输入的插值卷积神经网络	ModelNet40、S3DIS 、ShapeNet Parts、	插值卷积
Pointwise[60]	2017	基于卷积神经网络的三维点云语义分割和目标识别方法	S3DIS、SceneNN	逐点卷积
ShellNet[15]	2019	使用同心壳统计信息的高效点云卷积神经网络	ShapeNet、ScanNet Semantic3D 、S3DIS	ShellConv
PointConv[23]	2019	实现点云置换不变性和平移不变性的插值卷积神经网络	ModelNet、ShapeNet、ScanNet	PointConv、PointDeconv
SpiderCNN[30]	2018	有效地提取点云几何特征语义分割框架	SHREC15、ShapeNet	SpiderConv
A-CNN[61]	2019	基于环形卷积的多层次分层网络架构	ModelNet、ScanNet、ShapeNet、S3DIS、	环形卷积
GSA[64]	2019	基于点云关系推理的点注意力变压器	ModelNet、S3DIS	PAT、CAS、GSS
ASR[65]	2019	可嵌入各种网络框架的基于注意力的分数细化模块	ShapeNet、ScanNet	ASR
A-SCN[55]	2019	基于自注意力机制和形状上下文的端到端可训练架构	ModelNet、ShapeNet、S3DIS	形状上下文，注意力机制
ASIS[67]	2019	实例和语义的关联分割框架	ShapeNet、S3DIS	语义和实例结合
JSIS3D[68]	2019	多任务逐点网络	S3DIS、SceneNN	语义和实例结合、多值条件随机场模型

 doi: 10.3778/j.issn.1673-9418.2006025

JING Zhuangwei was born in 1996. He is a student at Nanjing University of Information Science & Technology. His research interests include Laser point cloud data processing, etc.

景庄伟（1996-），男，江苏盐城人，南京信息工程大学地理科学学院硕士研究生，主要研究领域为激光点云数据处理。

GUAN Haiyan was born in1976. She received the Ph.D. degrees in University of Waterloo, Canada and Wuhan University. Now he is a professor at School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science & Technology. Her research interests include Intelligent Processing of Remote Sensing Data, Target Recognition and Extraction and Laser point cloud data processing, etc.

管海燕（1976-），女，江苏南京人，加拿大滑铁卢大学、武汉大学双博士学位，现为南京信息工程大学教授，主要研究领域为遥感数据智能化处理、目标识别与提取、激光扫描数据等。

ZANG Yufu was born in 1987. He received the Ph.D. degree in photogrammetry and remote sensing at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) from Wuhan University, in 2016. He is currently the lecturer of Nanjing University of Information Science and Technology. His interests include computer vision, deep learning, and urban scene understanding.

臧玉府（1987-），男，江苏徐州人，2016 年武汉大学测绘遥感信息工程国家重点实验室获博士学位，代尔夫特理工大学博士后，目前任南京信息工程大学讲师，主要研究领域为计算机视觉、深度学习和城市场景理解。

NI Huan was born in Liaoning, China, in 1989. He received the Ph.D. degree in geographical information sciences from the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China, in 2017. He is currently a Lecturer with Nanjing University of Information Science and Technology, Nanjing, China. His current research focuses on computer vision, machine learning, pattern recognition, and their application on remote sensing data.

倪欢（1989-），男，辽宁人，南京信息工程大学讲师，于 2017 年获得武汉大学博士学位。主要研究领域为计算机视觉、机器学习、模式识别及其在遥感数据中的应用。

LI Dilong was born in 1989. He is currently working toward the Ph.D. degree in Photogrammetry and Remote Sensing at Wuhan University, China. His current research interests include pattern recognition, machine learning, information extraction from LiDAR point clouds and remotely sensed imagery.

李迪龙（1989-），男，福建龙岩人，武汉大学摄影测量与遥感专业博士研究生，主要研究领域为模式识别，机器学习，LiDAR 点云和遥感图像信息提取。

YU Yongtao was born in 1986. He received the Ph.D. degree in computer science and technology from Xiamen University in 2015. He is a lecturer at Huaiyin Institute of Technology. His research interests include intelligent transportation systems, deep learning, and intelligent interpretation of 3D point clouds and remote sensing images.

于永涛（1986-），男，内蒙古自治区赤峰市人，2015 年毕业于厦门大学并获得博士学位，目前为淮阴工学院讲师，主要研究领域为智能交通系统，深度学习，以及三维点云与遥感影像智能化解译。