

3D 点云的深度学习:综述

郭*玉兰, 王*汉云, 胡*庆勇, 刘*浩, 刘丽, Mohammed Bennamoun

由于在计算机视觉、自动驾驶和机器人等许多领域的广泛应用, 抽象点云学习最近受到了越来越多的关注。作为 AI 领域的主导技术, 深度学习已经被成功地用于解决各种 2D 视觉问题。然而, 由于深度神经网络处理点云所面临的独特挑战, 点云上的深度学习仍处于起步阶段。最近, 点云上的深度学习甚至开始蓬勃发展, 人们提出了许多方法来解决这一领域的不同问题。为了刺激未来的研究, 本文对点云的深度学习方法的最新进展进行了全面的综述。它涵盖了三个主要任务, 包括 3D 形状分类、3D 目标检测与跟踪、3D 点云分割。它还展示了几个公开可用数据集的比较结果, 以及有见地的观察和鼓舞人心的未来研究方向。

索引术语-深度学习, 点云, 3D 数据, 形状分类, 形状检索, 目标检测, 目标跟踪, 场景流, 实例分割, 语义分割, 部分分割。

1 介绍

随着 3D 采集技术的快速发展, 3D 传感器越来越普及和实惠, 包括各种类型的 3D 扫描仪、激光雷达和 RGB-D 相机(如 Kinect、RealSense 和苹果深度相机)[W1W]。这些传感器获取的 3D 数据可以提供丰富的几何、形状和尺度信息[W2W]、[W3W]。与 2D 图像相辅相成, 3D 数据为机器更好地了解周围环境提供了机会。3D 数据在不同领域有大量的应用, 包括自动驾驶、机器人、遥感和医疗[W4W]。

3D 数据通常可以用不同的格式表示, 包括深度图像、点云、网格和体积网格。作为一种常用的格式, 点云表示保留了 3D 空间中原始的几何信息, 无需进行任何离散化。因此, 它是自动驾驶、机器人等许多场景理解相关应用的首选表示形式。最近, 深度学习技术已经主导了许多研究领域, 如计算机视觉、语音识别和自然语言处理。然而, 3D 点云上的深度学习仍然面临着几个重大的挑战[5], 如数据集规模小, 三维点云的高维性和非结构化。在此基础上, 重点分析目前已用于处理三维点云的深度学习方法。

点云上的深度学习吸引了越来越多的关注, 尤其是在最近 5 年。同时也发布了几个公开可用的数据集, 比如

•Guo Y 和刘就职于中国中山大学电子与通信工程学院。王洪就职于中国信息工程大学测绘学院。胡麒就职于英国牛津大学计算机科学系。L. 刘就职于中国国防科技大学系统工程学院, 同时就职于芬兰奥卢大学机器视觉与信号分析中心。M. Bennamoun 就职于澳大利亚西澳大学计算机科学与软件工程系。

•*Y. Guo Y、王洪、胡麒、刘对本工作贡献相同, 为共同第一作者。

通讯作者:Guo Y(yulan.guo@nudt.edu.cn)。

F

ModelNet [6], ScanObjectNN [7], ShapeNet [8], PartNet [9], S3DIS [10], ScanNet [11], Semantic3D [12], ApolloCar3D[13], 以及 KITTI Vision 基准套件[14], [15]。这些数据集进一步推动了深度学习在 3D 点云上的研究, 越来越多的方法被提出来解决与点云处理相关的各种问题, 包括 3D 形状分类、3D 目标检测与跟踪、3D 点云分割、3D 点云配准、6 自由度姿态估计以及 3D 重建[16]、[17]、[18]。关于 3D 数据上的深度学习的调研也很少, 如[19]、[20]、[21]、[22]。然而, 我们的论文是第一次专门关注点云理解的深度学习方法。现有的 3D 点云深度学习方法的分类如图 1 所示。

与现有文献相比, 本工作的主要贡献可以归纳为以下几点:

- 1)据我们所知, 这是第一篇全面覆盖了几个重点云理解任务的深度学习方法的综述论文, 包括 3D 形状分类、3D 目标检测和跟踪以及 3D 点云分割。
- 2)与现有评论[19]、[20]不同, 我们特别关注 3D 点云的深度学习方法, 而不是所有类型的 3D 数据。
- 3)本文涵盖了深度学习在点云上的最新和先进进展。因此, 为读者提供了最先进的方法。
- 4)在几个公开可用的数据集上提供了现有方法的综合比较(例如, 在表 2、3、4、5 中), 并给出了简要的总结和有见地的讨论。

本文的结构如下。第 2 节介绍了各自任务的数据集和评估指标。第 3 节回顾了 3D 形状分类的方法。第 4 节提供了 3D 目标检测和跟踪的现有方法的综述。第五节

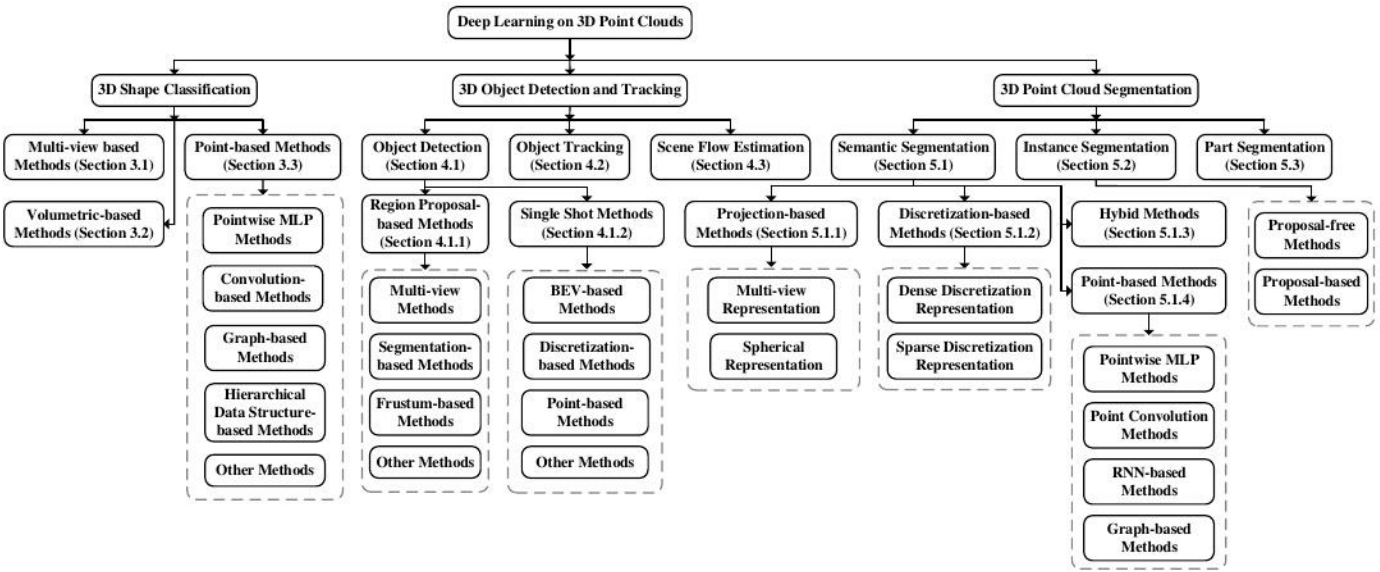


图 1:3D 点云深度学习方法分类。

对点云分割方法进行了综述，包括语义分割、实例分割和局部分割。最后，第 6 节总结了本文。我们还提供定期更新的项目页面:<https://github.com/QingyongHu/SoTA-Point-Cloud>。

2 背景

2.1 数据集

已经收集了大量的数据集来评估不同 3D 点云应用的深度学习算法的性能。表 1 列出了一些用于 3D 形状分类、3D 目标检测和跟踪以及 3D 点云分割的典型数据集。特别地，对这些数据集的属性也进行了总结。

对于 3D 形状分类，有两类数据集:合成数据集[6]、[8]和真实数据集[7]、[11]。合成数据集中的对象是完整的，没有任何遮挡和背景。相比之下，真实世界数据集中的物体在不同程度上被遮挡，一些物体被背景噪声污染。

对于 3D 目标检测和跟踪，有两类数据集:室内场景[11]、[25]和室外城市场景[14]、[28]、[30]、[31]。室内数据集中的点云要么是从密集深度图转换而来，要么是从 3D 网格中采样而来。室外城市数据集是为自动驾驶设计的，其中对象在空间上很好地分离，这些点云是稀疏的。

对于三维点云分割，这些数据集是由不同类型的传感器获取的，包括移动激光扫描仪(MLS)[15]、[34]、[36]、空中激光扫描仪(ALS)[33]、[38]、静态地面激光扫描仪(TLS)[12]、RGB- D 相机[11]和其他 3D 扫描仪[10]。这些数据集可用于开发算法，以应对各种挑战，包括类似的干扰物、形状不完整和类别不平衡。

2.2 评价指标

已经提出了不同的评估指标来测试这些方法在各种点云理解任务中的应用。对于 3D 形状分类，总体精度(OA)和平均类精度(mAcc)是最常用的性能标准。“OA”表示所有测试实例的平均精度，“mAcc”表示所有形状类的平均精度。对于 3D 目标检测，平均精度(AP)是最常用的标准。它被计算为准确率-召回率曲线下的面积。精确度和成功率通常用来评估 3D 单目标跟踪器的整体性能。平均多目标跟踪精度(AMOTA)和平均多目标跟踪精度(AMOTP)是评价三维多目标跟踪效果最常用的标准。对于三维点云分割，OA、平均交联(Intersection over Union, mIoU)和平均类精度(mean class Accuracy, mAcc)[10]、[12]、[15]、[36]、[37]是最常用的性能评价标准。特别地，平均精度均值(mAP)[39]也被用于 3D 点云的实例分割。

3 三维形状分类

用于此任务的方法通常首先学习每个点的嵌入，然后使用聚合方法从整个点云中提取全局形状嵌入。最后通过将全局嵌入馈送到几个完全连接的层中来实现分类。根据神经网络输入的数据类型，现有的 3D 形状分类方法可以分为基于多视图的方法、基于体的方法和基于点的方法。几个里程碑式的方法如图 2 所示。

基于多视图的方法将非结构化点云投影到 2D 图像中，而基于体的方法将点云转换为 3D 体表示。然后，利用完善的 2D 或 3D 卷积网络来实现形状分类。相比之下，基于点的方法直接在原始点云上工作，不需要任何体素化或投影。基于点的方法则不然

表 1:3D 形状分类、3D 物体检测和跟踪、3D 点云分割的现有数据集总结。¹用于评估的类的数量和标注类的数量(括号中显示)。

Datasets for 3D Shape Classification							
Name and Reference	Year	#Samples	#Classes	#Training	#Test	Type	Representation
McGill Benchmark [23]	2008	456	19	304	152	Synthetic	Mesh
Sydney Urban Objects [24]	2013	588	14	-	-	Real-World	Point Clouds
ModelNet10 [6]	2015	4899	10	3991	605	Synthetic	Mesh
ModelNet40 [6]	2015	12311	40	9843	2468	Synthetic	Mesh
ShapeNet [8]	2015	51190	55	-	-	Synthetic	Mesh
ScanNet [11]	2017	12283	17	9677	2606	Real-World	RGB-D
ScanObjectNN [7]	2019	2902	15	2321	581	Real-World	Point Clouds
Datasets for 3D Object Detection and Tracking							
Name and Reference	Year	#Scenes	#Classes	#Annotated Frames	#3D Boxes	Secne Type	Sensors
KITTI [14]	2012	22	8	15K	200K	Urban (Driving)	RGB & LiDAR
SUN RGB-D [25]	2015	47	37	5K	65K	Indoor	RGB-D
ScanNetV2 [11]	2018	1.5K	18	-	-	Indoor	RGB-D & Mesh
H3D [26]	2019	160	8	27K	1.1M	Urban (Driving)	RGB & LiDAR
Argoverse [27]	2019	113	15	44K	993K	Urban (Driving)	RGB & LiDAR
Lyft L5 [28]	2019	366	9	46K	1.3M	Urban (Driving)	RGB & LiDAR
A*3D [29]	2019	-	7	39K	230K	Urban (Driving)	RGB & LiDAR
Waymo Open [30]	2020	1K	4	200K	12M	Urban (Driving)	RGB & LiDAR
nuScenes [31]	2020	1K	23	40K	1.4M	Urban (Driving)	RGB & LiDAR
Datasets for 3D Point Cloud Segmentation							
Name and Reference	Year	#Points	#Classes ¹	#Scans	Spatial Size	RGB	Sensors
Oakland [32]	2009	1.6M	5(44)	17	-	N/A	MLS
ISPRS [33]	2012	1.2M	9	-	-	N/A	ALS
Paris-rue-Madame [34]	2014	20M	17	2	-	N/A	MLS
IQmulus [35]	2015	300M	8(22)	10	-	N/A	MLS
ScanNet [11]	2017	-	20(20)	1513	8×4×4	Yes	RGB-D
S3DIS [10]	2017	273M	13(13)	272	10×5×5	Yes	Matterport
Semantic3D [12]	2017	4000M	8(9)	15/15	250×260×80	Yes	TLS
Paris-Lille-3D [36]	2018	143M	9(50)	3	200×280×30	N/A	MLS
SemanticKITTI [15]	2019	4549M	25(28)	23201/20351	150×100×10	N/A	MLS
Toronto-3D [37]	2020	78.3M	8(9)	4	260×350×40	Yes	MLS
DALES [38]	2020	505M	8(9)	40	500×500×65	N/A	ALS

引入显式信息损失，变得越来越受欢迎。请注意，本文主要关注基于点的方法，但也包括少数基于多视图和基于体的完整性方法。

3.1 基于多视图的方法

这些方法首先将 3D 形状投影到多个视图中并提取视图级特征，然后将这些特征进行融合以进行精确的形状分类。如何将多个视图特征聚合成具有判别力的全局表示，是这些方法面临的关键挑战。

MVCNN[40]是一个开创性的工作，它简单地将多视图特征最大池化为一个全局描述子。然而，max-pooling 只保留特定视角的最大元素，导致信息丢失。MHBN[41]通过协调双线性池化集成局部卷积特征，产生紧凑的全局描述子。Yang et al.[42]首先利用关系网络来利用一组视图上的相互关系(例如，区域-区域关系和视图-视图关系)，然后聚合这些视图以获得具有判别性的 3D 对象表示。此外，还提出了其他几种方法[43]、[44]、[45]、[46]来提高识别精度。与之前的方法不同，Wei 等人[47]通过考虑多个视图作为图形节点，在 View-GCN 中使用了一个有向图。然后将由局部图卷积、非局部消息传递和选择性视图采样组成的核心层应用于构建的图。的

所有级别的最大池化节点特征的连接最终用于形成全局形状描述符。

3.2 基于体积的方法

这些方法通常将点云体素化成三维网格，然后在体积表示上应用三维卷积神经网络(CNN)进行形状分类。

Maturana 等人[48]引入了一种称为 VoxNet 的体占用网络来实现鲁棒的 3D 物体识别。Wu et al.[6]提出了一种基于卷积深度信念的 3D ShapeNets 来学习来自各种 3D 形状(由体素网格上二进制变量的概率分布表示)的点的分布。虽然已经取得了令人鼓舞的性能，但这些方法无法很好地扩展到密集的 3D 数据，因为计算和内存占用随分辨率呈立方增长。

为此，引入了一种分层和紧凑的结构(如八叉树)来减少这些方法的计算和内存成本。OctNet[49]首先使用混合网格-八叉树结构分层划分点云，该结构用沿规则网格的几个浅八叉树表示场景。利用位串表示法对八叉树结构进行高效编码，并通过简单的算法对每个体素的特征向量进行索引。Wang 等人[50]提出了一种基于 octree 的 CNN 三维形状分类方法。在最精细的叶八边形中采样的 3D 模型的平均法向量被馈送到网络中，3D- cnn 被应用于占用的八边形上

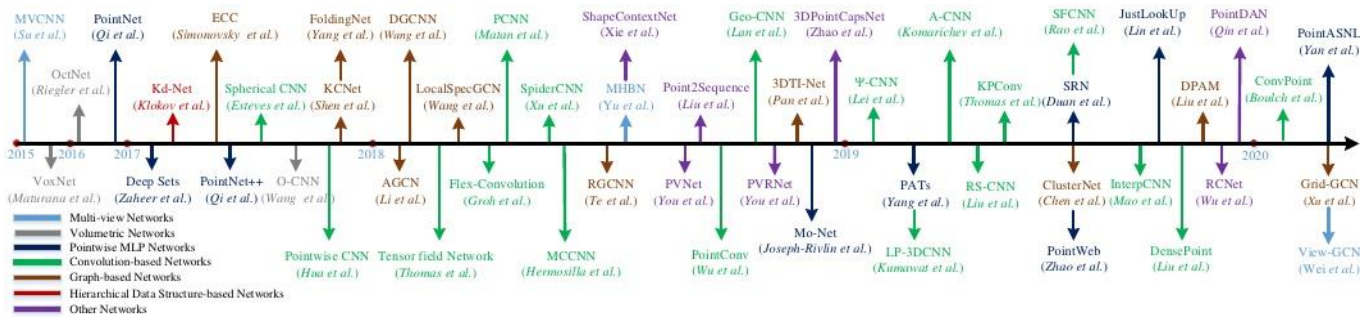


图 2:最相关的基于深度学习的 3D 形状分类方法的时间综述。

由 3D 形状表面。与基于密集输入网络的基线网络相比，OctNet 对高分辨率点云所需的内存和运行时间要少得多。Le et al.[51]提出了一种称为 PointGrid 的混合网络，它集成了点和网格表示，以高效地处理点云。在每个嵌入体网格单元内采样恒定数量的点，这允许网络通过使用 3D 卷积来提取几何细节。Ben-Shabat 等人[52]将输入点云转换成三维网格，再用 3D modified Fisher Vector (3DmFV)方法表示三维网格，然后通过传统的 CNN 架构学习全局表示。

3.3 点基方法

根据每个点的特征学习所使用的网络架构，这类方法可以分为基于点的 MLP、基于卷积的、基于图的、基于分层数据结构的方法等典型方法。

3.3.1 逐点 MLP方法

这些方法使用几个共享的多层感知器(mlp)独立地对每个点进行建模，然后使用对称聚合函数聚合全局特征，如图 3 所示。

由于 2D 图像固有的数据不规则性，典型的 2D 图像深度学习无法直接应用于 3D 点云。作为一项开创性的工作，点网[5]直接将点云作为其输入并实现

对称函数的置换不变性。具体来说，PointNet 通过几个 MLP 层独立学习点特征，并通过最大池化层提取全局特征。深度集[53]通过汇总所有表示和应用非线性变换来实现排列不变性。由于 PointNet 中每个点的特征都是独立学习的[5]，因此无法捕获点之间的局部结构信息。因此，Qi 等[54]提出了一种分层网络 PointNet++，从每个点的邻域捕获精细的几何结构。作为 PointNet++结构的核心，其集合抽象层由三层组成:采样层、分组层和基于 PointNet 的学习层。通过堆叠几个抽象层，PointNet++从局部几何结构中学习特征，并逐层抽象局部特征。

由于其简单性和较强的表示能力，基于 PointNet 开发了许多网络[5]。Mo-Net[55]的架构类似于 PointNet[5]，但它采用一组有限的矩作为输入。点注意力 transformer (PATs)[56]通过其自身相对于邻居的绝对位置和相对位置来表示每个点，并通过 MLPs 学习高维特征。然后，使用 Group Shuffle Attention (GSA)来捕获点之间的关系，并开发一个置换不变、可微和可训练的端到端 Gumbel 子集采样(GSS)层来学习分层特征。在 PointNet++[54]的基础上，PointWeb[57]利用局部邻域的上下文，利用自适应特征调整(AFA)来改进点特征。Duan 等[58]提出了一种结构关系网络(Structural Relational Network, SRN)，利用 MLP 学习不同局部结构之间的结构关系特征。Lin 等人[59]通过为 PointNet 学习的输入空间和函数空间构建查找表，加速了推理过程。在中等机器上，ModelNet 和 ShapeNet 数据集上的推理时间比 PointNet 加快了 1.5 ms 和 32 倍。SRINet[60]首先投影点云获得旋转不变表示，然后利用基于 pointnet 的骨干提取全局特征，利用基于图的聚合提取局部特征。在 PointASNL 中，Yan 等人[61]利用自适应采样(AS)模块自适应调整通过最远点采样(FPS)算法采样的点的坐标和特征，并提出了局部-非局部(L-NL)模块来捕获这些采样点的局部和长程依赖关系。

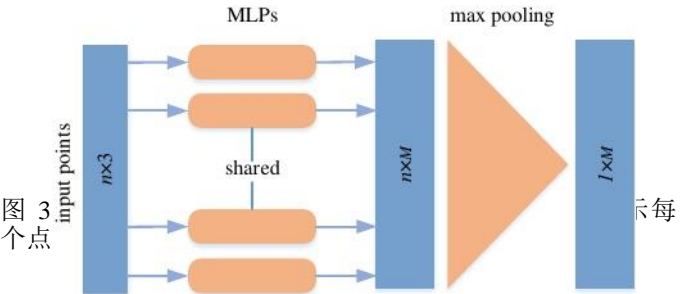


图 3 个点

3.3.2 基于卷积的方法

与定义在 2D 网格结构(如图像)上的卷积核相比，由于点云的不规则性，3D 点云的卷积核很难设计。根据卷积核的类型，目前的 3D 卷积方法可以分为连续卷积和离散卷积方法，如图 4 所示。

三维连续卷积方法。这些方法在连续空间上定义卷积核，其中邻近点的权重与相对于中心点的空间分布有关。

3D 卷积可以被解释为给定子集上的加权和。RS-cnn[62]作为 RS- cnn 的核心层，RS- Conv 以某一点周围的点的局部子集作为输入，通过学习局部子集中点之间从低级关系(如欧几里得距离和相对位置)到高级关系的映射，使用 MLP 实现卷积。在[63]中，核元素是在单位球体中随机选择的。然后利用基于 MLP 的连续函数来建立核元素位置与点云之间的关系。在 DensePoint[64]中，卷积被定义为具有非线性激活器的单层感知器(SLP)。通过连接所有先前层的特征来学习特征，以充分利用上下文信息。Thomas 等人[65]使用一组可学习的核点为 3D 点云提出了刚性和可变形核点卷积(KPConv)算子。ConvPoint[66]将卷积核分离为空间部分和特征部分。空间部分的位置从单位球随机选择，并通过简单的 MLP 学习权重函数。

有些方法还使用现有的算法来执行卷积。在 PointConv[67]中，卷积被定义为相对于一个重要采样的连续三维卷积的蒙特卡罗估计。卷积核由加权函数(通过 MLP 层学习)和密度函数(通过核化密度估计和 MLP 层学习)组成。为了提高记忆和计算效率，3D 卷积被进一步简化为两个操作:矩阵乘法和 2D 卷积。在相同的参数设置下，其内存消耗可以降低约 64 倍。在 MCCNN[68]中，卷积被认为是依赖于样本密度函数的蒙特卡罗估计过程(用 MLP 实现)。然后使用泊松圆盘采样来构建点云层次结构。这个卷积算子可以用来执行两个或多个采样方法之间的卷积，并且可以处理不同的采样密度。在 SpiderCNN

[69]， SpiderConv 提出将卷积定义为阶跃函数与 k 个最近邻上定义的泰勒展开的乘积。阶跃函数通过编码局部测地线距离来捕获粗几何，泰勒展开通过在立方体顶点插入任意值来捕获固有的局部几何变化。此外，还建立了卷积网络 PCNN

[70]也提出了基于径向基函数的三维点云。

已经提出了几种方法来解决 3D 卷积网络面临的旋转等变问题。

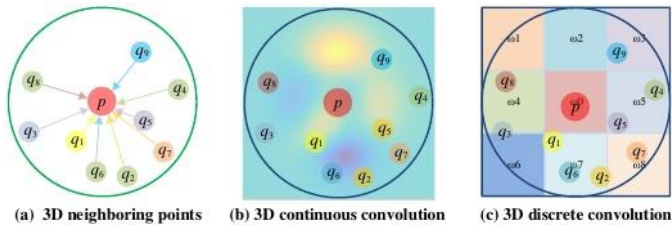


图 4:一个点的局部邻域的连续和离散卷积的示意图。(a)表示一个以点 p 为中心的局部邻域 q_i p;(b)和(c)分别表示 3D 连续卷积和离散卷积。

Esteves 等人[71]提出了 3D 球面 CNN 来学习 3D 形状的旋转等变表示，它以多值球面函数作为输入。通过在球面调和域内用锚点对谱进行参数化来获得局部卷积滤波器。提出了张量场网络[72]，将点卷积运算定义为可学习的径向函数和球调和函数的乘积，它们对 3D 旋转、平移和排列是局部等变的。[73]中的卷积是基于球面互相关定义的，并使用广义快速傅里叶变换(FFT)算法实现。基于 PCNN, SPHNet[74]通过在体积函数的卷积过程中加入球谐核实现旋转不变性。

为了加快计算速度，Flex-Convolution[75]将卷积核的权重定义为 k 近邻上的标准标量积，可以使用 CUDA 进行加速。实验结果表明，它在参数更少、内存消耗更低的小型数据集上具有竞争力的性能。

3D 离散卷积方法。这些方法在规则网格上定义卷积核，其中邻近点的权重与相对于中心点的偏移量相关。

Hua 等人[76]将非均匀的 3D 点云转换为均匀的网格，并在每个网格上定义卷积核。所提出的 3D 核为落入同一网格的所有点分配相同的权重。对于一个给定的点，位于同一网格上的所有相邻点的平均特征都是从前一层计算出来的。然后，对所有网格的平均特征进行加权求和，生成当前层的输出。Lei 等人[77]定义了一个球面卷积核，方法是将 3D 球面邻域区域划分为多个体积 bins，并将每个 bin 与一个可学习的加权矩阵相关联。一个点的球面卷积核的输出由其邻域点的加权激活值的均值的非线性激活决定。在 GeoConv[78]中，一个点与其邻域点之间的几何关系是基于 6 个基显式建模的。沿着基的每个方向的边缘特征由方向相关的可学习矩阵独立加权。然后根据给定点及其邻近点形成的角度，对这些方向相关的特征进行聚合。对于给定的点，它在当前层的特征被定义为特征的总和

给定点及其相邻边缘特征的前一层。

PointCNN[79]通过 x -conv 变换(通过 MLP 实现)将输入点转换为潜在的和潜在的正则阶, 然后对转换后的特征应用典型卷积算子。

[80]提出了一种插值卷积算子 Inter- pConv 来度量输入点云和核坐标之间的几何关系。Zhang 等人。

[81]提出了一种实现旋转不变性的 RConv 算子, 它将低层次的旋转不变性几何特征作为输入, 然后通过简单的分箱方法将卷积转化为 1D。a - cnn[82]通过在查询点的每个环上围绕核的大小对近邻数组进行循环来定义环形卷积, 并学习局部子集中邻近点之间的关系。

为了减少 3D cnn 的计算和存储成本, Kumawat 等人[83]提出了一种整流局部相位体积(ReLPV)块, 以基于 3D 短期傅里叶变换(STFT)的 3D 局部领域中提取相位, 这大大减少了参数的数量。在 SFCNN[84]中, 点云被投影到具有对齐球坐标的正二十面体晶格上。然后通过 convolutional -maxpooling-convolution 结构对从球面晶格顶点及其邻居连接而来的特征进行卷积。SFCNN 具有抗旋转和摄动的能力。

3.3.3 基于图的方法

基于图的网络将点云中的每个点视为图的一个顶点, 并基于每个点的邻居为图生成有向边。然后在空间域或谱域中进行特征学习[85]。一个典型的基于图的网络如图 5 所示。

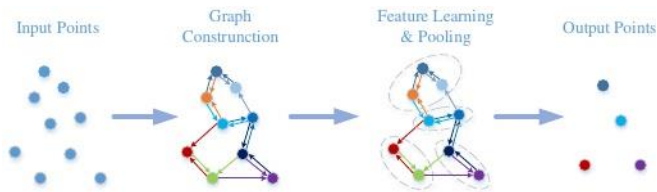


图 5:基于图的网络的图示。

基于图的空间域方法。这些方法定义了空间域的操作(例如, 卷积和池化)。具体来说, 卷积通常是通过对空间邻居的 MLP 来实现的, 并采用池化的方法通过聚合每个点的邻居信息来生成新的粗化图。每个顶点上的特征通常被赋予坐标、激光强度或颜色, 而每个边缘上的特征通常被赋予两个连通点之间的几何属性。

作为一项开创性的工作, Simonovsky 等人[85]将每个点视为图的一个顶点, 并通过有向边将每个顶点与其所有邻居连接起来。然后, 利用滤波器生成网络(如 MLP)提出了边缘条件卷积(ECC)。采用最大池化来

基于 VoxelGrid 实现聚合邻域信息和图粗化 [86]。在 DGCNN[87]中, 在特征空间中构造一个图, 并在网络的每一层之后动态更新。作为 EdgeConv 的核心层, 使用 MLP 作为每个边缘的特征学习函数, 并对与每个点的邻居相关联的边缘特征进行通道对称聚合。进一步, LDGCNN[88]去除了变换网络, 将 DGCNN[87]中不同层的分层特征链接起来, 以提高其性能并减小模型尺寸。还提出了图式无监督深度 AutoEncoder 网络(即 FoldingNet[89]), 该网络使用向量化的局部协方差矩阵和点坐标的连接作为其输入。受盗梦空间 [90]和 DGCNN[87]的启发, Hassani 和 Haley[91]提出了一种无监督多任务自编码器来学习点和形状特征。编码器是基于多尺度图构建的。解码器使用三个无监督任务构建, 包括聚类、自监督分类和重构, 这些任务与多任务损失联合训练。Liu 等[92]提出了一种基于图卷积的动态点集聚模块(Dynamic Points Agglomeration Module, DPAM), 将点集聚(采样、分组、池化)过程简化为一个简单的步骤, 通过集聚矩阵与点特征矩阵的乘法实现。在点网体系结构的基础上, 通过堆叠多个 dpm 构建了一个分层学习体系结构。与点网++的层次策略相比[54], DPAM 动态地利用了语义空间中点的关系和聚类点。

为了利用局部几何结构, KCNet[93]基于核相关性学习特征。具体来说, 一组表征局部结构几何类型的可学习点被定义为核。然后, 计算核与给定点的邻域之间的亲和度。在 G3D[94]中, 将卷积定义为邻接矩阵多项式的一种变体, 将池化定义为将拉普拉斯矩阵和顶点矩阵用一个粗化矩阵相乘。ClusterNet[95]利用严格的旋转不变模块从每个点的 k 个最近邻中提取旋转不变特征, 并基于具有 ward-linkage 准则的无监督凝聚层次聚类方法构建点云的层次结构[96]。每个子簇中的特征首先通过 EdgeConv 块进行学习, 然后通过 max pooling 进行聚合。

为了解决当前数据结构化方法(如 FPS 和近邻点查询)的耗时问题, Xu 等人[97]提出融合基于体和基于点方法的优点来提高计算效率。在 ModelNet 分类任务上的实验表明, 所提出的 Grid-GCN 网络的计算效率平均比其他模型高 $5 \times$ faster。

谱域基于图的方法。这些方法将卷积定义为频谱滤波, 通过图上的信号与图拉普拉斯矩阵的特征向量相乘来实现[98], [99]。

RGCNN[100]通过将每个点与点云中的所有其他点连接起来并进行更新来构建一个图

每一层的图拉普拉斯矩阵。为了使相邻顶点的特征更加相似，在损失函数中添加了图-信号平滑先验。为了解决数据的多样化图拓扑结构带来的挑战，AGCN[101]中的SGC-LL 层利用可学习的距离度量来参数化图上两个顶点之间的相似性。从图中得到的邻接矩阵使用高斯核和学习距离进行归一化。HGNN[102]通过在超图上应用谱卷积来构建超边卷积层。

上述方法在全图上运行。为了利用局部结构信息，Wang 等人 [103] 提出了一种端到端的谱卷积网络 LocalSpecGCN，用于处理局部图(由 k 个最近邻构建)。该方法不需要任何图拉普拉斯矩阵和图粗化层次的离线计算。在 PointGCN[104]中，基于点云的 k 个近邻构建图，并使用高斯核对每条边进行加权。卷积滤波器被定义为图谱域中的切比雪夫多项式。使用全局池化和多分辨率池化来捕获点云的全局和局部特征。Pan 等[105]通过在谱域的 k 个最近邻图上应用卷积提出了 3DTI-Net。几何变换的不变性是通过从相对欧氏距离和方向距离学习来实现的。

3.3.4 基于分层数据结构的方法

这些网络是基于不同的层次数据结构(例如，八叉树和 kd 树)构建的。在这些方法中，点特征是沿着树从叶节点到根节点分层学习的。

Lei 等人[77]提出了一种使用球面卷积核的八叉树引导 CNN(如第 3.3.2 节所述)。网络的每一层对应于八叉树的一层，每一层应用一个球面卷积核。当前层神经元的值被确定为前一层所有相关子节点的平均值。与基于八叉树的 OctNet[49]不同，Kd-Net[106]是使用多个 K-d 树构建的，每个树在每次迭代中具有不同的分裂方向。遵循自底向上的方法，使用 MLP 从其子节点的表示计算非叶节点的表示。根节点(描述整个点云)的特征最终被馈送到全连接层以预测分类分数。请注意，Kd-Net 根据节点的分裂类型在每个级别共享参数。3DContextNet[107]使用标准平衡 K-d 树来实现特征学习和聚合。在每个级别上，首先通过基于局部线索(模拟局部区域中点之间的相互依赖关系)和全局上下文线索(模拟一个位置相对于所有其他位置的关系)的 MLP 来学习点特征。然后，使用 MLP 从其子节点计算非叶节点的特征，并通过最大池化进行聚合。对于分类，重复上述过程，直到获得根节点。

SO-Net 网络的层次结构是通过执行点到节点的 k 近邻搜索来构建的[108]。具体来说，利用改进的排列不变自组织映射(SOM)来模拟空间分布

点云的分布。通过一系列全连接层，从规范化的点到节点坐标中学习单个点的特征。SOM 中每个节点的特征是使用通道最大池化从与该节点相关的点特征中提取的。然后使用类似于 PointNet[5]的方法从节点特征中学习最终特征。相对于 PointNet++ [54]，SOM 的层次结构更加高效，对点云的空间分布进行了充分的探索。

3.3.5 其他方法

此外，还提出了许多其他方案。RBFNet[113]通过聚合具有可学习核位置和大小稀疏分布的径向基函数(RBF)核的特征，明确地建模点的空间分布。3DPointCapsNet[112]通过点向 MLP 和卷积层学习点无关特征，并通过多个最大池化层提取全局潜在表示。然后基于无监督动态路由，学习强大的代表性潜在胶囊。Qin 等人[116]提出了一种用于 3D 点云表示的端到端无监督域适应网络 PointDAN。为了捕获点云的语义属性，提出了一种自监督方法来重建点云，点云的部分已被随机重排[117]。Li 等人[118]提出了一种自动增强框架 PointAugment，用于自动优化和增强用于网络训练的点云样本。具体来说，自动学习每个输入样本的形状变换和逐点位移，并通过交替优化和更新其增强器和分类器的可学习参数来训练网络。受 shape context[119]的启发，Xie 等人[109]提出了一种 ShapeContextNet 架构，将亲和点选择和紧凑的特征聚合结合到使用点积自注意力的软对齐操作中[120]。为了处理 3D 点云中的噪声和遮挡，Bobkov 等人[121]将基于 4D 旋转不变描述子的手工点对函数馈送到 4D 卷积神经网络中。Prokudin 等人[122]首先从单位球上随机采样一个均匀分布的基点集，然后将点云编码为到基点集的最小距离。因此，点云被转换为一个相对较小的固定长度的向量。编码后的表示可以用现有的机器学习方法进行处理。

RCNet[115]利用标准 RNN 和 2D CNN 构建了用于三维点云处理的排列不变网络。首先将点云划分成平行光束，并沿特定维度排序，然后将每个光束馈送到共享 RNN 中。将学习到的特征进一步馈送到高效的 2D CNN 中进行分层特征聚合。为了增强其描述能力，RCNet-E 被提出，以沿着不同的划分和排序方向集成多个 rcnet。Point2Sequences[114]是另一个基于 rnn 的模型，它捕获点云局部区域中不同区域之间的相关性。它将从多个尺度的局部区域学习到的特征作为序列，并将这些序列从所有局部区域输入到基于 rnn 的编码器-解码器结构中以聚合局部区域特征。

表 2:在 ModelNet10/40 基准上比较 3D 形状分类结果。在这里，我们只关注基于点的网络。`#params` 表示模型的参数数量，`OA` 表示所有测试实例的平均精度，`mAcc` 表示表中所有形状类的平均精度。符号“-”表示结果不可用。

Methods		Input	#params (M)	ModelNet40 (OA)	ModelNet40 (mAcc)	ModelNet10 (OA)	ModelNet10 (mAcc)
Pointwise MLP Methods	PointNet [5]	Coordinates	3.48	89.2%	86.2%	-	-
	PointNet++ [54]	Coordinates	1.48	90.7%	-	-	-
	MO-Net [55]	Coordinates	3.1	89.3%	86.1%	-	-
	Deep Sets [53]	Coordinates	-	87.1%	-	-	-
	PAT [56]	Coordinates	-	91.7%	-	-	-
	PointWeb [57]	Coordinates	-	92.3%	89.4%	-	-
	SRN-PointNet++ [58]	Coordinates	-	91.5%	-	-	-
	JUSTLOOKUP [59]	Coordinates	-	89.5%	86.4%	92.9%	92.1%
	PointASNL [61]	Coordinates	-	92.9%	-	95.7%	-
Convolution-based Methods	PointASNL [61]	Coordinates+Normals	-	93.2%	-	95.9%	-
	Pointwise-CNN [76]	Coordinates	-	86.1%	81.4%	-	-
	PointConv [67]	Coordinates+Normals	-	92.5%	-	-	-
	MC Convolution [68]	Coordinates	-	90.9%	-	-	-
	SpiderCNN [69]	Coordinates+Normals	-	92.4%	-	-	-
	PointCNN [79]	Coordinates	0.45	92.2%	88.1%	-	-
	Flex-Convolution [75]	Coordinates	-	90.2%	-	-	-
	PCNN [70]	Coordinates	1.4	92.3%	-	94.9%	-
	Boulch [63]	Coordinates	-	91.6%	88.1%	-	-
	RS-CNN [62]	Coordinates	-	93.6%	-	-	-
	Spherical CNNs [71]	Coordinates	0.5	88.9%	-	-	-
	GeoCNN [78]	Coordinates	-	93.4%	91.1%	-	-
	Ψ -CNN [77]	Coordinates	-	92.0%	88.7%	94.6%	94.4%
	A-CNN [82]	Coordinates	-	92.6%	90.3%	95.5%	95.3%
	SFCNN [84]	Coordinates	-	91.4%	-	-	-
	SFCNN [84]	Coordinates+Normals	-	92.3%	-	-	-
	DensePoint [64]	Coordinates	0.53	93.2%	-	96.6%	-
	KPConv rigid [65]	Coordinates	-	92.9%	-	-	-
	KPConv deform [65]	Coordinates	-	92.7%	-	-	-
	InterpCNN [80]	Coordinates	12.8	93.0%	-	-	-
Graph-based Methods	ConvPoint [66]	Coordinates	-	91.8%	88.5%	-	-
	ECC [85]	Coordinates	-	87.4%	83.2%	90.8%	90.0%
	KCNet [93]	Coordinates	0.9	91.0%	-	94.4%	-
	DGCNN [87]	Coordinates	1.84	92.2%	90.2%	-	-
	LocalSpecGCN [103]	Coordinates+Normals	-	92.1%	-	-	-
	RGCNN [100]	Coordinates+Normals	2.24	90.5%	87.3%	-	-
	LDGCNN [88]	Coordinates	-	92.9%	90.3%	-	-
	3DII-Net [105]	Coordinates	2.6	91.7%	-	-	-
	PointGCN [104]	Coordinates	-	89.5%	86.1%	91.9%	91.6%
	ClusterNet [95]	Coordinates	-	87.1%	-	-	-
	Hassani et al. [91]	Coordinates	-	89.1%	-	-	-
	DPAM [92]	Coordinates	-	91.9%	89.9%	94.6%	94.3%
Hierarchical Data Structure-based Methods	Grid-GCN [97]	Coordinates	-	93.1%	91.3%	97.5%	97.4%
	KD-Net [106]	Coordinates	2.0	91.8%	88.5%	94.0%	93.5%
	SO-Net [108]	Coordinates	-	90.9%	87.3%	94.1%	93.9%
	SCN [109]	Coordinates	-	90.0%	87.6%	-	-
	A-SCN [109]	Coordinates	-	89.8%	87.4%	-	-
	3DContextNet [107]	Coordinates	-	90.2%	-	-	-
	3DContextNet [107]	Coordinates+Normals	-	91.1%	-	-	-
Other Methods	3DmFV-Net [52]	Coordinates	4.6	91.6%	-	95.2%	-
	PVNet [110]	Coordinates+Views	-	93.2%	-	-	-
	PVRNet [111]	Coordinates+Views	-	93.6%	-	-	-
	3DPointCapsNet [112]	Coordinates	-	89.3%	-	-	-
	DeepRBFNet [113]	Coordinates	3.2	90.2%	87.8%	-	-
	DeepRBFNet [113]	Coordinates+Normals	3.2	92.1%	88.8%	-	-
	Point2Sequences [114]	Coordinates	-	92.6%	90.4%	95.3%	95.1%
	RCNet [115]	Coordinates	-	91.6%	-	94.7%	-
	RCNet-E [115]	Coordinates	-	92.3%	-	95.6%	-

一些方法也从 3D 点云和 2D 图像中学习。在 PVNet[110]中，从多视图图像中提取的高层全局特征通过嵌入网络投影到点云的子空间中，并通过软注意力掩膜与点云特征进行融合。最后，对融合的特征和多视图特征采用残差连接来进行形状识别。后来，进一步提出 PVRNet[111]，通过关系分数模块来利用 3D 点云与其多个视图之间的关系。基于关系得分，对原始的 2D 全局视图特征进行点-单视图融合和点-多视图融合增强。

3.4 总结

ModelNet10/40[6]数据集是 3D 形状分类最常用的数据集。表 2 展示了不同基于点的网络所取得的结果。可以得出几个观察结果:

- 点态 MLP 网络通常被用作其他类型网络学习点态特征的基本构建块。
- 作为标准的深度学习架构，基于卷积的网络可以在不规则 3D 点云上实现优越的性能。对于不规则数据，无论是离散卷积网络还是连续卷积网络，都应该给予更多的关注。

- 由于其固有的处理不规则数据的强大能力，近年来基于图的神经网络吸引了越来越多的关注。然而，将谱域中的基于图的神经网络扩展到各种图结构仍然是具有挑战性的。

4 3d 目标检测与跟踪

在本节中，我们将回顾现有的 3D 目标检测、3D 目标跟踪和 3D 场景流估计方法。

4.1 3D 目标检测

典型的 3D 物体检测器将场景的点云作为其输入，并在每个检测到的物体周围产生一个有方向的 3D 包围盒，如图 6 所示。与图像中的目标检测类似[123]，3D 目标检测方法可以分为两类：基于区域建议框的方法和单镜头方法。图 7 给出了几个里程碑式的方法。

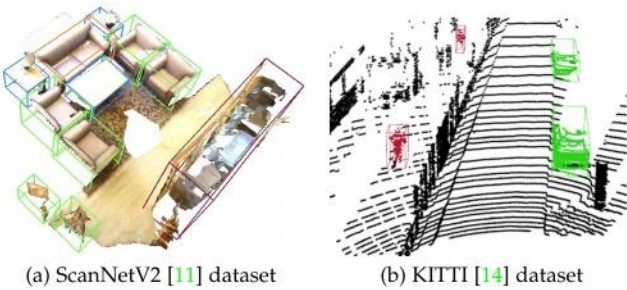


图 6:3D 目标检测示意图。(a)和(b)最初分别显示在[124]和[125]中。

4.1.1 基于区域提议的方法

这些方法首先提出几个包含对象的可能区域(也称为建议框)，然后提取区域特征来确定每个建议框的类别标签。根据它们的对象建议生成方法，这些方法可以进一步分为三类：基于多视图的方法、基于分割的方法和基于视锥图的方法。

基于多视角的方法。这些方法融合了来自不同视点地图(例如，LiDAR 前视图、鸟瞰图(BEV)和图像)的建议特征，以获得 3D 旋转框，如图 8(a)所示。这些方法的计算成本通常很高。

Chen 等人[4]从 BEV 地图中生成一组高精度的 3D 候选框，并将其投影到多视图(如激光雷达前视图图像、RGB 图像)的特征图中。然后他们结合这些来自不同视角的区域特征来预测有方向的 3D 边界框，如图 8(a)所示。虽然这种方法在只有 300 个建议的情况下，在交并比(IoU)为 0.25 时实现了 99.1%的召回率，但其速度对于实际应用来说太慢了。随后，一些方法被开发出来，从两个方面改进多视角 3D 目标检测方法。

首先，提出了几种方法来有效地融合不同模态的信息。生成

Ku 等人[126]提出了一种基于多模态融合的区域建议网络。他们首先使用裁剪和调整大小操作从 BEV 和图像视图中提取相等大小的特征，然后使用元素平均池将这些特征融合在一起。Liang 等人[127]利用连续卷积实现了不同分辨率下图像和 3D 激光雷达特征图的有效融合。具体来说，他们提取 BEV 空间中每个点最近的对应图像特征，然后通过将图像特征投影到 BEV 平面上，使用双线性插值得到密集的 BEV 特征映射。实验结果表明，密集 BEV 特征图比离散图像特征图和稀疏激光雷达特征图更适合于三维目标检测。Liang 等[128]提出了一种多任务多传感器 3D 目标检测网络，用于端到端训练。具体来说，利用多个任务(例如，2D 目标检测、地面估计和深度补全)来帮助网络学习更好的特征表示。进一步利用学习到的跨模态表示来产生高度准确的目标检测结果。实验结果表明，该方法在 2D、3D 和 BEV 检测任务上取得了显著的改进，在 TOR4D 基准上优于以往最先进的方法[129]，[130]。

其次，已经研究了不同的方法来提取输入数据的鲁棒表示。Lu 等人[39]通过引入空间通道注意力(SCA)模块探索了多尺度上下文信息，该模块捕获了场景的全局和多尺度上下文并突出了有用的特征。他们还提出了一种可扩展的空间 Unsample (ESU)模块，通过结合多尺度的低层特征来获取具有丰富空间信息的高层特征，从而生成可靠的 3D 物体建议。虽然可以实现更好的检测性能，但前述的多视图方法需要很长的运行时间，因为它们对每个候选框进行特征池化。随后，Zeng 等人[131]使用 pre-RoI 池化卷积来提高[4]的效率。具体来说，他们将大部分卷积操作移动到 RoI 池模块之前。因此，对所有目标建议执行一次 RoI 卷积。实验结果表明，该方法能够以 11.1 fps 的速度运行，比 MV3D[4]快 5 倍。

有效方法。这些方法首先利用现有的语义分割技术去除大部分背景点，然后在前景点上生成大量高质量的建议以节省计算，如图 8(b)所示。与多视角方法[4]、[126]、[131]相比，这些方法实现了更高的对象召回率，更适用于物体高度遮挡和拥挤的复杂场景。

Yang 等[132]使用 2D 分割网络预测前景像素，并将其投影到点云中，以去除大部分背景点。然后，他们在预测的前景点上生成建议，并设计了一个名为 PointsIoU 的新准则，以减少建议的冗余和歧义。在[132]之后，Shi 等人[133]提出了一个 PointRCNN 框架。具体来说，他们直接分割 3D 点云获取前景点，然后融合语义特征和局部空间特征

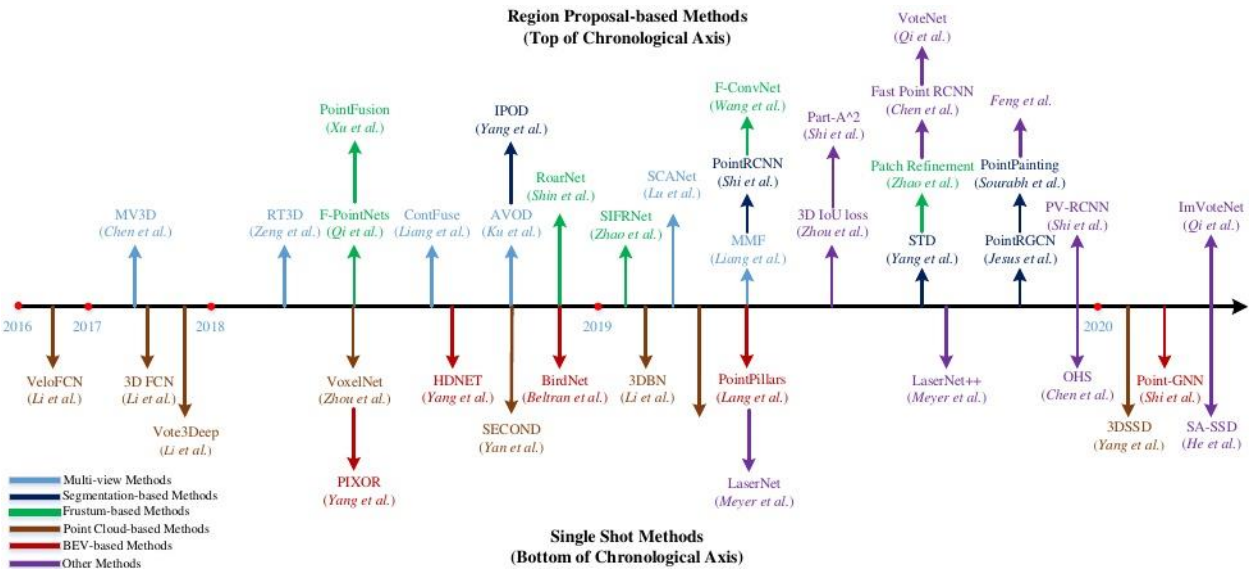


图 7:最相关的基于深度学习的 3D 物体检测方法的时间综述。

制作高质量 3D 盒子的功能。继[133]的区域建议网络(RPN)阶段之后, Jesus 等人[134]提出了一项利用图卷积网络(GCN)进行 3D 目标检测的开创性工作。具体来说, 引入了两个模块来使用图卷积来细化目标建议。第一个模块 R-GCN 利用候选框中包含的所有点来实现每个候选框的特征聚合。第二个模块 C-GCN 融合所有候选框的每帧信息, 通过利用上下文回归准确的目标框。Sourabh 等人[135]将点云投影到基于图像的分割网络的输出中, 并将语义预测分数附加到这些点上。绘制的点被输入到现有的检测器中[133]、[136]、[137], 以实现显著的性能提升。Yang 等[138]将每个点与球形锚相关联。然后使用每个点的语义分数来去除冗余锚点。因此, 与之前的方法相比, 该方法以较低的计算成本实现了较高的召回率[132], [133]。此外, 还提出了 PointsPool 层来学习提案内部点的紧凑特征, 并引入了并行 IoU 分支来提高定位精度和检测性能。

Frustum-based 方法。这些方法首先利用现有的 2D 物体检测器来生成物体的 2D 候选区域, 然后为每个 2D 候选区域提取一个 3D 视锥建议框, 如图 8(c)所示。虽然这些方法可以有效地提出 3D 物体的可能位置, 但逐步的管道使它们的性能受到 2D 图像检测器的限制。

F-PointNets[139]是这方面的先驱工作。它为每个二维区域生成一个载体建议, 并应用 PointNet[5](或 PointNet++[54])学习每个三维截体的点云特征, 进行模态三维盒估计。在后续工作中, Zhao 等人[140]提出了一个 Point-SENet 模块来预测一组缩放因子, 这些因子被进一步用于自适应地突出有用特征并抑制无信息特征。他们还将 PointSIFT[141]模块整合到网络中

来捕获点云的方向信息, 实现了对形状缩放的强大鲁棒性。与 F-PointNets[139]相比, 该方法在室内和室外数据集上都取得了显著的改进[14], [25]。

Xu 等人[142]同时利用了 2D 图像区域及其对应的视锥点来精确回归 3D 盒子。为了融合点云的图像特征和全局特征, 他们提出了一个用于直接回归盒角点位置的全局融合网络。他们还提出了一个密集融合网络, 用于预测每个角点的逐点偏移量。Shin 等人[143]首先从 2D 图像中估计物体的 2D 边界框和 3D 姿态, 然后提取多个几何上可行的物体候选。这些 3D 候选物体被输入到一个盒子回归网络中, 以预测准确的 3D 物体盒子。Wang 等[144]为每个 2D 区域沿截锥体轴生成一系列截锥体, 并应用 PointNet[5]提取每个截锥体的特征。对视锥层特征进行改造以生成 2D 特征图, 然后将其输入到全卷积网络中进行 3D 盒估计。该方法在基于 2D 图像的方法中实现了最先进的性能, 并在 KITTI 官方排行榜上排名最高。Johannes 等人[145]首先在 BEV 图上获得初步检测结果, 然后根据 BEV 预测提取小点子集(也称为 patch)。应用局部细化网络来学习 patch 的局部特征, 以预测高精度的 3D 包围盒。

其他方法。受轴向 IoU 在图像目标检测中的成功启发, Zhou 等人[146]将两个 3D 旋转边界盒的 IoU 集成到几个最先进的检测器中[133], [137], [158], 以实现一致的性能改进。Chen 等人[147]提出了一种两阶段的网络架构, 以同时使用点云和体素表示。首先, 点云被体素化, 并馈送到 3D 骨干网络以产生初始检测结果。其次, 进一步利用初始预测的内部点特征进行盒子细化。虽然这个设计在概念上很简单, 但实现了可比性

表 3:在 KITTI 测试三维检测基准上对比三维物体检测结果。3D 边界框 IoU 阈值对于汽车为 0.7，对于行人和骑自行车的人为 0.5。模式是 LiDAR (L)和 image (I)。“E”、“M”和“H”分别代表简单、中等和困难类别的物体。为简单起见，我们省略了值后面的“%”。符号“-”表示结果不可用。

Method			Modality	Speed (fps)	Cars			Pedestrians			Cyclists		
					E	M	H	E	M	H	E	M	H
Region Proposal -based Methods	Multi-view Methods	MV3D [4]	L & I	2.8	74.97	63.63	54.00	-	-	-	-	-	-
		AVOD [126]	L & I	12.5	76.39	66.47	60.23	36.10	27.86	25.76	57.19	42.08	38.29
		ContFuse [127]	L & I	16.7	83.68	68.78	61.67	-	-	-	-	-	-
		MMF [128]	L & I	12.5	88.40	77.43	70.22	-	-	-	-	-	-
		SCANet [39]	L & I	11.1	79.22	67.13	60.65	-	-	-	-	-	-
		RT3D [131]	L & I	11.1	23.74	19.14	18.86	-	-	-	-	-	-
	Segmentation -based Methods	IPOD [132]	L & I	5.0	80.30	73.04	68.73	55.07	44.37	40.05	71.99	52.23	46.50
		PointRCNN [133]	L	10.0	86.96	75.64	70.70	47.98	39.37	36.01	74.96	58.82	52.53
		PointRGCN [134]	L	3.8	85.97	75.73	70.60	-	-	-	-	-	-
		PointPainting [135]	L & I	2.5	82.11	71.70	67.08	50.32	40.97	37.87	77.63	63.78	55.89
		STD [138]	L	12.5	87.95	79.71	75.09	53.29	42.47	38.35	78.69	61.59	55.30
	Frustum -based Methods	F-PointNets [139]	L & I	5.9	82.19	69.79	60.59	50.53	42.15	38.08	72.27	56.12	49.01
		SIFRNet [140]	L & I	-	-	-	-	-	-	-	-	-	-
		PointFusion [142]	L & I	-	77.92	63.00	53.27	33.36	28.04	23.38	49.34	29.42	26.98
		RoarNet [143]	L & I	10.0	83.71	73.04	59.16	-	-	-	-	-	-
		F-ConvNet [144]	L & I	2.1	87.36	76.39	66.69	52.16	43.38	38.80	81.98	65.07	56.54
		Patch Refinement [145]	L	6.7	88.67	77.20	71.82	-	-	-	-	-	-
	Other Methods	3D IoU loss [146]	L	12.5	86.16	76.50	71.39	-	-	-	-	-	-
		Fast Point R-CNN [147]	L	16.7	84.80	74.59	67.27	-	-	-	-	-	-
		PV-RCNN [148]	L	12.5	90.25	81.43	76.82	-	-	-	-	-	-
		VoteNet [124]	L	-	-	-	-	-	-	-	-	-	-
		Feng et al. [149]	L	-	-	-	-	-	-	-	-	-	-
		ImVoteNet [150]	L & I	-	-	-	-	-	-	-	-	-	-
		Part-A*2 [151]	L	12.5	87.81	78.49	73.51	-	-	-	-	-	-
Single Shot Methods	BEV-based Methods	PIXOR [129]	L	28.6	-	-	-	-	-	-	-	-	-
		HDNET [152]	L	20.0	-	-	-	-	-	-	-	-	-
		BirdNet [153]	L	9.1	13.53	9.47	8.49	12.25	8.99	8.06	16.63	10.46	9.53
	Discretization -based Methods	VeloFCN [154]	L	1.0	-	-	-	-	-	-	-	-	-
		3D FCN [155]	L	<0.2	-	-	-	-	-	-	-	-	-
		Vote3Deep [156]	L	-	-	-	-	-	-	-	-	-	-
		3DBN [157]	L	7.7	83.77	73.53	66.23	-	-	-	-	-	-
		VoxelNet [136]	L	2.0	77.47	65.11	57.73	39.48	33.69	31.51	61.22	48.36	44.37
		SECOND [158]	L	26.3	83.34	72.55	65.82	48.96	38.78	34.91	71.33	52.08	45.83
		MVX-Net [159]	L & I	16.7	84.99	71.95	64.88	-	-	-	-	-	-
		PointPillars [137]	L	62.0	82.58	74.31	68.99	51.45	41.92	38.89	77.10	58.65	51.92
		SA-SSD [160]	L	25.0	88.75	79.79	74.16	-	-	-	-	-	-
	Point-based Methods	3DSSD [161]	L	25.0	88.36	79.57	74.55	54.64	44.27	40.23	82.48	64.10	56.90
	Other Methods	LaserNet [162]	L	83.3	-	-	-	-	-	-	-	-	-
		LaserNet++ [163]	L & I	26.3	-	-	-	-	-	-	-	-	-
		OHS-Dense [164]	L	33.3	88.12	78.34	73.49	47.14	39.72	37.25	79.09	62.72	56.76
		OHS-Direct [164]	L	33.3	86.40	77.74	72.97	51.29	44.81	41.13	77.70	63.16	57.16
		Point-GNN [125]	L	1.7	88.33	79.47	72.29	51.92	43.77	40.14	78.60	63.48	57.08

性能提高到[133]，同时保持 16.7 fps 的速度。Shi 等人[148]提出了 PointVoxel-RCNN (PV-RCNN)，利用 3D 卷积网络和基于 PointNet 的集合抽象来学习点云特征。具体来说，首先对输入的点云进行体素化，然后送入 3D 稀疏卷积网络以生成高质量的候选框。然后通过体素集抽象模块将学习到的体素特征编码为一组小的关键点。此外，他们还提出了一个关键点到网格的 ROI 抽象模块，以捕获丰富的上下文信息，用于盒子细化。实验结果表明，该方法明显优于以往的方法，在 KITTI 三维检测基准的 Car 类中排名第一¹。

Qi 等人[124]受基于 Hough 投票的 2D 物体检测器的启发，提出了 VoteNet 从点云中直接投票选出物体的虚拟中心点，并通过聚合投票特征生成一组高质量的 3D 物体提案。VoteNet 显著优于先前仅使用几何信息的方法，并且实现了

在两个大型室内基准(即 ScanNet[11]和 SUN RGB-D[25])上的最先进性能。然而，对于部分遮挡的物体，虚拟中心点的预测是不稳定的。进一步，Feng 等[149]增加了一个方向向量的辅助分支，以提高虚拟中心点和 3D 候选框的预测精度。此外，还建立了候选框之间的 3D 对象-对象关系图，以强调用于精确目标检测的有用特征。Qi 等人[150]提出了一种 ImVoteNet 检测器，该检测器将 2D 物体检测线索(例如几何和语义/纹理线索)融合到 3D 投票管道中。Shi 等人[151]受三维物体地面真实盒提供物体内部零件精确位置的启发，提出了由零件感知阶段和零件聚合阶段组成的 Part-²ANet。部分感知阶段应用具有稀疏卷积和稀疏反卷积的类 unet[165]网络来学习逐点特征，用于预测和粗粒度生成物体内部部件位置。部分聚合阶段采用 roi 感知池化来聚合预测的部件位置以进行盒子细化。

1.排名参照提交时间:2020 年 6 月 12 日

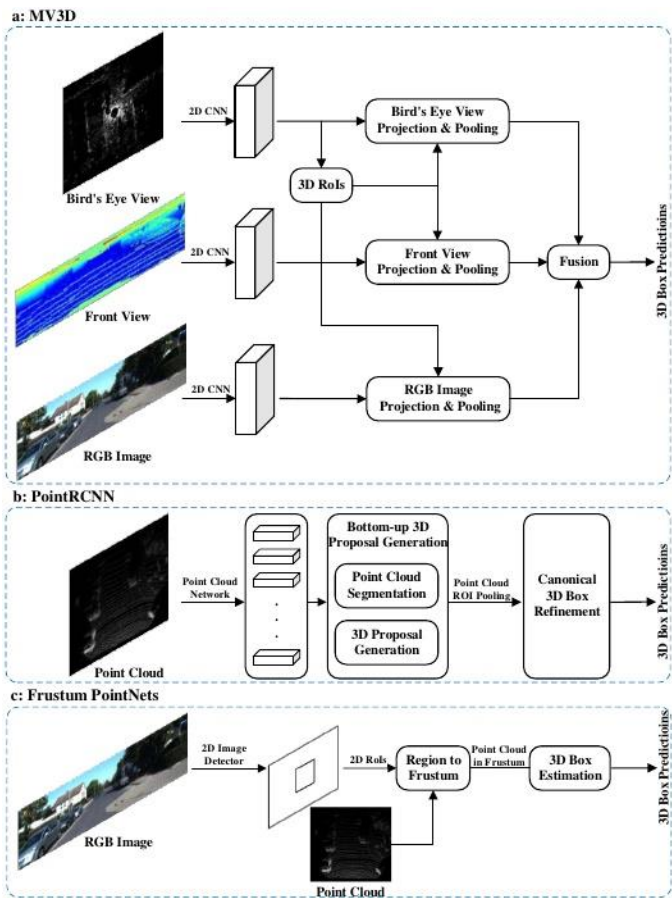


图 8:三类基于区域提议的 3D 目标检测方法的典型网络。从上到下:(a)基于多视图的方法, (b)基于分割的方法和(c)基于视锥图的方法。

4.1.2 单镜头方法

这些方法直接预测类概率, 并使用单阶段网络回归物体的 3D 边界框。它们不需要区域候选框的生成和后处理。因此, 它们可以高速运行。根据输入数据的类型, 单次射击方法可分为基于 bev 的方法、基于离散化的方法和基于点的方法。

基于 bev 的方法。这些方法主要以 BEV 表示作为输入。Yang 等人[129]用等间距的单元离散场景的点云, 并以类似的方式编码反射率, 从而得到规则表示。然后应用全卷积网络(FCN)网络来估计物体的位置和航向角。该方法优于大多数单镜头方法(包括 VeloFCN [154], 3D-FCN[155]和 Vote3Deep[156]), 同时运行速度为 28.6 fps。后来, Yang 等人[152]利用高清(HD)地图提供的几何和语义先验信息, 提高了[129]的鲁棒性和检测性能。具体来说, 他们从 HD 地图中获得地面点的坐标, 然后使用相对于地面的距离进行 BEV 表示, 以弥补道路坡度造成的平移方差。此外, 他们将一个二进制道路掩码与 BEV 表示相连接

沿通道维度聚焦移动对象。由于高清地图并非无处不在, 他们还提出了一个在线地图预测模块, 从单个激光雷达点云估计地图先验。这种地图感知方法在 TOR4D[129]、[130]和 KITTI[14]数据集上的表现明显优于其基线。然而, 其对不同密度点云的泛化性能较差。为了解决这一问题, Beltr' an 等人[153]提出了一种考虑不同激光雷达传感器之间差异的归一化图。归一化图是与 BEV 图具有相同分辨率的 2D 网格, 它对每个单元中包含的最大点数进行编码。结果表明, 该归一化映射显著提高了基于 bev 的检测器的泛化能力。

Discretization-based 方法。这些方法将点云转换为常规的离散表示, 然后应用 CNN 来预测类别和物体的 3D 盒子。

Li 等人[154]提出了第一种使用 FCN 进行三维目标检测的方法。他们将点云转换成二维点图, 并使用二维 FCN 来预测物体的边界框和置信度。后来, 他们[155]将点云离散为具有长、宽、高和通道维度的 4D 张量, 并扩展了

基于二维 fcn 的检测技术到三维领域进行三维目标检测。与[154]相比, 基于 3D fcn 的方法[155]获得了超过 20%的精度增益, 但由于 3D 卷积和数据的稀疏性, 不可避免地要消耗更多的计算资源。为了解决体素的稀疏性问题, Engelcke 等人[156]利用以特征为中心的投票方案为每个非空体素生成一组投票, 并通过累积投票来获得卷积结果。其计算复杂度与被占用的体素数量成正比。Li 等[157]通过堆叠多个稀疏 3D cnn 构建了一个 3D 骨干网。该方法旨在通过充分利用体素的稀疏性来节省内存和加速计算。这

3D backbone 网络提取丰富的 3D 特征用于目标检测, 而不引入繁重的计算负担。

Zhou 等人[136]提出了一个基于体素的端到端可训练框架 VoxelNet。他们将点云划分为等间隔的体素, 并将每个体素内的特征编码为 4D 张量。然后连接一个区域建议网络以产生检测结果。虽然它的性能很强, 但这种方法由于体素和 3D 卷积的稀疏性, 速度非常慢。后来, Yan 等人[158]利用稀疏卷积网络[166]来提高[136]的推理效率。他们还提出了一种正弦误差角度损失来解决 0 和 π 方向之间的歧义性。Sindagi 等人[159]在早期阶段通过融合图像和点云特征来扩展 VoxelNet。具体来说, 他们将[136]生成的非空体素投影到图像中, 并使用预训练网络为每个投影体素提取图像特征。然后将这些图像特征与体素特征连接起来, 生成精确的 3D 盒子。与[136]、[158]相比, 该方法可以有效地利用多模态信息来减少假阳性和假阴性。Lang 等人[137]提出了一种名为 PointPillars 的 3D 目标检测器。该方法利用 PointNet[5]来学习以垂直柱(pillar)形式组织的点云的特征并进行编码

将学习到的特征作为伪图像。然后应用一个 2D 对象检测管道来预测 3D 边框。在平均精度(AP)方面, PointPillars 优于大多数融合方法(包括 MV3D[4]、RoarNet[143]和 AVOD[126])。此外, PointPillars 在 3D 和 BEV KITTI[14]基准测试中都可以以 62 fps 的速度运行, 非常适合实际应用。

He 等人[160]观察到点云的部分空间信息不可避免地会在现有单发探测器逐渐缩小的特征图中丢失, 受到启发, 提出了一种 SA-SSD 探测器, 利用细粒度结构信息来提高定位精度。具体来说, 他们首先将点云转换为张量, 并将其输入到骨干网络中以提取多阶段特征。此外, 还使用了一个具有点级监督的辅助网络来指导特征学习点云的结构。实验结果表明, SA-SSD 在 KITTI² BEV 检测基准的 Car 类中排名第一。

积分方法。 这些方法直接将原始点云作为输入。3DSSD[161]是这一方向的开创性工作。它引入了距离-fps (D-FPS)和特征- fps (F-FPS)的融合采样策略, 以去除耗时的特征传播(FP)层和[133]中的细化模块。然后, 使用候选生成(CG)层来充分利用代表点, 这些代表点被进一步送入带有 3D 中心标签的无锚框回归头来预测 3D 对象框。实验结果表明, 3DSSD 在保持 25 fps 的速度的同时, 优于两阶段基于点的方法 PointRCNN[133]。

其他方法。 Meyer 等人[162]提出了一种高效的三维物体探测器, 称为 LaserNet。该方法预测每个点边界框上的概率分布, 然后结合这些逐点分布来生成最终的 3D 物体框。进一步, 将点云的密集范围视图(RV)表示作为输入, 并提出一种快速的 mean-shift 算法来减少逐点预测产生的噪声。LaserNet 在 0 到 50 米的范围内实现了最先进的性能, 其运行时间明显低于现有方法。Meyer 等人[163]随后扩展了 LaserNet[162], 以利用 RGB 图像提供的密集纹理(例如 50 至 70 米)。具体来说, 他们通过将 3D 点云投影到 2D 图像上, 将激光雷达点与图像像素关联起来, 并利用这种关联将 RGB 信息融合到 3D 点中。他们还将 3D 语义分割作为学习更好表示的辅助任务。该方法在保持 LaserNet 的高效率的同时, 在远距离(如 50 ~ 70 米)目标检测和语义分割方面都取得了显著的进步。Chen 等人[164]观察到孤立物体部分上的点可以提供关于物体位置和方向的丰富信息, 受此启发, 提出了一种新颖的热点表示和第一个基于热点的无锚点检测器。具体来说, 首先对原始点云进行体素化, 然后将其送入骨干网络以生成 3D 特征图。这些特征图被用于对热点进行分类, 并同时预测 3D 边界框。

2.排名参照提交时间:2020 年 6 月 12 日

注意, 热点分配在骨干网的最后一层卷积层。实验结果表明, 该方法取得了相当的性能, 对稀疏点云具有鲁棒性。Shi 等人[125]提出了一种图神经网络 point - gnn 来检测激光雷达点云中的三维物体。他们首先将输入点云编码为具有固定半径的近邻图, 然后将该图输入 point - gnn, 以预测物体的类别和盒子。

4.2 3D 物体跟踪

给定目标在第一帧中的位置, 目标跟踪的任务是估计其在后续帧中的状态[167], [168]。由于 3D 目标跟踪可以利用点云中丰富的几何信息, 因此有望克服基于图像的跟踪所面临的几个缺点, 包括遮挡、光照和尺度变化。

受暹罗网络[169]在基于图像的目标跟踪方面的成功启发, Giancola 等人[170]提出了一种具有形状补全正则化的三维暹罗网络。具体来说, 他们首先使用卡尔曼滤波器生成候选项, 并使用形状正则化将模型和候选项编码为紧凑的表示。然后利用余弦相似度搜索下一帧中被跟踪目标的位置。这种方法可以作为目标跟踪的替代方法, 并且显著优于大多数 2D 目标跟踪方法, 包括 STAPLE_{CA} [171]和 SiamFC[169]。为了高效地搜索目标物体, Zarzar 等人[172]利用 2D 暹罗网络在 BEV 表示上生成大量粗候选物体。然后, 他们通过利用三维暹罗网络中的余弦相似性来改进候选对象。这种方法在精度(即提高 18%)和成功率(即提高 12%)方面都明显优于[170]。Simon 等人[173]提出了一种面向语义点云的 3D 目标检测和跟踪架构。他们首先通过融合 2D 视觉语义信息生成体素化语义点云, 然后利用时间信息提高多目标跟踪的准确性和鲁棒性。此外, 他们还引入了一个强大且简化的评估指标(即尺度-旋转-平移分数(SRFs))来加速训练和推理。Complexer- YOLO 实现了有希望的跟踪性能, 并且仍然可以实时运行。此外, Qi 等人[174]提出了一种 P2B (Point-to-Box)网络。他们将模板和搜索区域输入骨干网络, 以获得它们的种子。搜索区域的种子用特定于目标的特征进行增强, 然后通过霍夫投票回归出潜在的目标中心。实验结果表明, 当以 40 fps 运行时, P2B 的性能优于[170]10%以上。

4.3 3D 场景流估计

给定两个点云 X 和 Y, 三维场景流 $D = \{D_i\}^N$ 描述每个点 x_{in_i} X 到对应位置 x_{in} Y 的运动, 使得 ${}_i X = X + D$ 。图 9 显示 ${}_i$ 了 ${}_i$ 个 KITTI 点云之间的三维场景 ${}_i$ 流。类似于 2D 视觉中的光流估计, 一些方法已经开始从点云序列中学习有用的信息(例如 3D 场景流, 空间-临时信息)。

表 4:在 KITTI 测试 BEV 检测基准上对比三维物体检测结果。3D 边界框 IoU 阈值对于汽车为 0.7，对于行人和骑自行车的人为 0.5。模式是 LiDAR (L)和 image (I)。“E”、“M”和“H”分别代表简单、中等和困难类别的物体。为简单起见，我们省略了值后面的“%”。符号“-”表示结果不可用。

Method			Modality	Speed (fps)	Cars			Pedestrians			Cyclists		
					E	M	H	E	M	H	E	M	H
Region Proposal-based Methods	Multi-view Methods	MV3D [4]	L & I	2.8	86.62	78.93	69.80	-	-	-	-	-	-
		AVOD [126]	L & I	12.5	89.75	84.95	78.32	42.58	33.57	30.14	64.11	48.15	42.37
		ContFuse [127]	L & I	16.7	94.07	85.35	75.88	-	-	-	-	-	-
		MMF [128]	L & I	12.5	93.67	88.21	81.99	-	-	-	-	-	-
		SCANet [39]	L & I	11.1	90.33	82.85	76.06	-	-	-	-	-	-
		RT3D [131]	L & I	11.1	56.44	44.00	42.34	-	-	-	-	-	-
	Segmentation-based Methods	IPOD [132]	L & I	5.0	89.64	84.62	79.96	60.88	49.79	45.43	78.19	59.40	51.38
		PointRCNN [133]	L	10.0	92.13	87.39	82.72	54.77	46.13	42.84	82.56	67.24	60.28
		PointRGCN [134]	L	3.8	91.63	87.49	80.73	-	-	-	-	-	-
		PointPainting [135]	L & I	2.5	92.45	88.11	83.36	58.70	49.93	46.29	83.91	71.54	62.97
		STD [138]	L	12.5	94.74	89.19	86.42	60.02	48.72	44.55	81.36	67.23	59.35
	Frustum-based Methods	F-PointNets [139]	L & I	5.9	91.17	84.67	74.77	57.13	49.57	45.48	77.26	61.37	53.78
		SIFRNet [140]	L & I	-	-	-	-	-	-	-	-	-	-
		PointFusion [142]	L & I	-	-	-	-	-	-	-	-	-	-
		RoarNet [143]	L & I	10.0	88.20	79.41	70.02	-	-	-	-	-	-
		F-ConvNet [144]	L & I	2.1	91.51	85.84	76.11	57.04	48.96	44.33	84.16	68.88	60.05
		Patch Refinement [145]	L	6.7	92.72	88.39	83.19	-	-	-	-	-	-
	Other Methods	3D IoU loss [146]	L	12.5	91.36	86.22	81.20	-	-	-	-	-	-
		Fast Point R-CNN [147]	L	16.7	90.76	85.61	79.99	-	-	-	-	-	-
		PV-RCNN [148]	L	12.5	94.98	90.65	86.14	-	-	-	82.49	68.89	62.41
		VoteNet [124]	L	-	-	-	-	-	-	-	-	-	-
		Feng et al. [149]	L	-	-	-	-	-	-	-	-	-	-
		ImVoteNet [150]	L & I	-	-	-	-	-	-	-	-	-	-
		Part-A*2 [151]	L	12.5	91.70	87.79	84.61	-	-	-	81.91	68.12	61.92
		PIXOR [129]	L	28.6	83.97	80.01	74.31	-	-	-	-	-	-
Single Shot Methods	BEV-based Methods	HDNET [152]	L	20.0	89.14	86.57	78.32	-	-	-	-	-	-
		BirdNet [153]	L	9.1	76.88	51.51	50.27	20.73	15.80	14.59	36.01	23.78	21.09
	Discretization-based Methods	VeloFCN [154]	L	1.0	0.02	0.14	0.21	-	-	-	-	-	-
		3D FCN [155]	L	<0.2	70.62	61.67	55.61	-	-	-	-	-	-
		Vote3Deep [156]	L	-	-	-	-	-	-	-	-	-	-
		3DBN [157]	L	7.7	89.66	83.94	76.50	-	-	-	-	-	-
		VoxelNet [136]	L	2.0	89.35	79.26	77.39	46.13	40.74	38.11	66.70	54.76	50.55
		SECOND [158]	L	26.3	89.39	83.77	78.59	55.99	45.02	40.93	76.50	56.05	49.45
		MX-Net [159]	L & I	16.7	92.13	86.05	78.68	-	-	-	-	-	-
		PointPillars [137]	L	62.0	90.07	86.56	82.81	57.60	48.64	45.78	79.90	62.73	55.58
		SA-SSD [160]	L	25.0	95.03	91.03	85.96	-	-	-	-	-	-
	Point-based Methods	3DSSD [161]	L	25.0	92.66	89.02	85.86	60.54	49.94	45.73	85.04	67.62	61.14
	Other Methods	LaserNet [162]	L	83.3	79.19	74.52	68.45	-	-	-	-	-	-
		LaserNet++ [163]	L & I	26.3	-	-	-	-	-	-	-	-	-
		OHS-Dense [164]	L	33.3	93.73	88.11	84.98	50.87	44.59	42.14	82.13	66.86	60.86
		OHS-Direct [164]	L	33.3	93.59	87.95	83.21	55.90	49.48	45.79	79.66	67.20	61.04
		Point-GNN [125]	L	1.7	93.11	89.17	83.90	55.36	47.07	44.61	81.17	67.28	59.67

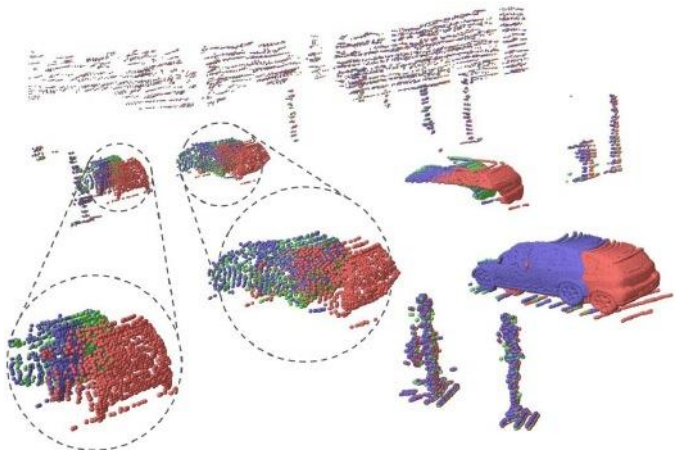


图 9:两个 KITTI 点云之间的 3D 场景流，最初显示在[175]中。点云 X、Y 和平移后的 X 点云分别用红色、绿色和蓝色高亮显示。

Liu 等人[175]提出 FlowNet3D 从一对连续的点云中直接学习场景流。FlowNet3D 通过流嵌入层学习点级特征和运动特征。然而，FlowNet3D 有两个问题。首先，一些预测的运动矢量在它们的方向上与地面真实值显著不同。其次，FlowNet 难以应用于非静态场景，特别是以可变形物体为主的场景。为了解决这个问题，Wang 等人[176]引入了余弦距离损失来最小化预测和真实值之间的夹角。此外，他们还提出了点到平面的距离损失，以提高刚性场景和动态场景的精度。实验结果表明，这两个损失项将 FlowNet3D 的训练精度从 57.85%提高到 63.43%，并加快和稳定了训练过程。Gu 等人[177]提出了一种分层多面体晶格 FlowNet (HPLFlowNet)来直接估计大规模点云的场景流。提出了几个双边卷积层，用于从原始点云中恢复结构信息，同时降低了计算成本。

为了有效处理顺序点云，Fan 和 Yang[178]提出了 PointRNN、PointGRU 和 PointLSTM

网络和序列到序列模型跟踪移动点。PointRNN、PointGRU 和 PointLSTM 能够捕捉空间-临时信息并对动态点云进行建模。同样，Liu 等人[179]提出了 MeteorNet，直接从动态点云中学习一种表示。这种方法学习从时空邻近点聚合信息。进一步引入直接分组和链流分组来确定时间邻居。然而，上述方法的性能受到数据集规模的限制。Mittal 等人[180]提出了两种自监督损失来在大型无标签数据集上训练他们的网络。他们的主要思想是，鲁棒的场景流估计方法应该在前向和后向预测中都是有效的。由于无法获得场景流注释，因此将预测的变换点的最近邻视为伪真值。然而，真实的 ground truth 可能与最近的点并不相同。为了避免这个问题，他们反向计算了场景流，并提出了一个循环一致性损失来将点平移到原始位置。实验结果表明，这种自监督方法超过了基于监督学习的方法的最先进性能。

4.4 总结

KITTI 基准[14]是自动驾驶领域最具影响力的数据集之一，在学术界和工业界都得到了广泛的应用。表 3 和表 4 给出了不同探测器在 KITTI 测试 3D 基准上获得的结果。可以得出以下观察结果：

- 基于区域提议的方法是这两类方法中最常被研究的方法，并且在 KITTI 测试 3D 和 BEV 基准测试中都大大优于单镜头方法。
- 现有的 3D 目标检测器有两个限制。第一，现有方法的远程检测能力相对较差。其次，如何充分利用图像中的纹理信息仍然是一个开放问题。
- 多任务学习是 3D 目标检测的未来方向。例如，MMF[128]通过合并多个任务来学习跨模态表示，以实现最先进的检测性能。
- 3D 目标跟踪和场景流估计是新兴的研究主题，自 2019 年以来逐渐受到越来越多的关注。

5 3d 点云分割

3D 点云分割既需要了解全局几何结构，也需要了解每个点的细粒度细节。根据分割粒度，三维点云分割方法可以分为三类：语义分割(场景级)、实例分割(对象级)和部分分割(部件级)。

5.1 三维语义分割

给定一个点云，语义分割的目标是根据点的语义将其分成几个子集。与 3D 形状分类的分类法(第 3 节)类似，语义分割有四种范式：基于投影的、基于离散化的、基于点的和混合的方法。

无论是基于投影的方法还是基于离散化的方法，第一步都是将点云转换为中间正则表示，如多视图[181]、[182]、球面[183]、[184]、[185]、体积[166]、[186]、[187]、permutohedral lattice[188]、[189]和混合表示[190]、[191]，如图 11 所示。然后，将中间分割结果投影回原始点云。相比之下，基于点的方法直接在不规则点云上工作。图 10 展示了几种代表性的方法。

5.1.1 基于投影的方法

这些方法通常将 3D 点云投影到 2D 图像中，包括多视图和球面图像。

多视点表示。Lawin 等人[181]首先从多个虚拟摄像机视角将 3D 点云投影到 2D 平面上。然后，使用多流 FCN 来预测合成图像的像素级分数。通过融合不同视角上的重投影得分，得到每个点的最终语义标签。类似地，Boulch 等人[182]首先使用多个相机位置生成点云的多个 RGB 和深度快照。然后，他们使用 2D 分割网络对这些快照进行像素级标记。利用残差校正进一步融合 RGB 和深度图像预测的分数[192]。基于点云采样自局部欧氏曲面的假设，Tatarchenko 等人[193]引入了切线卷积用于密集点云分割。该方法首先将每个点周围的局部曲面几何投影到一个虚拟切平面。然后在曲面几何上直接操作切线卷积。这种方法显示出了极大的可扩展性，能够处理具有数百万点的大规模点云。总的来说，多视角分割方法的性能对视点选择和遮挡很敏感。此外，这些方法没有充分利用底层的几何和结构信息，因为投影步骤不可避免地引入了信息损失。

球形表示。为了实现 3D 点云的快速准确分割，Wu 等[183]提出了一种基于 SqueezeNet[194]和条件随机场(Conditional Random Field, CRF)的端到端网络。为了进一步提高分割精度，引入 SqueezeSegV2[184]，通过利用无监督域适应管道来解决域偏移问题。Milioto 等人[185]提出 RangeNet++用于激光雷达点云的实时语义分割。首先将 2D 深度图像的语义标签转移到 3D 点云，进一步使用高效的 gpu 支持的基于 knn 的后处理步骤来缓解离散化误差和模糊推理输出的问题。与单视图投影相比，球面投影保留了更多的信息，适合激光雷达点云的标记。然而，这种中间表示

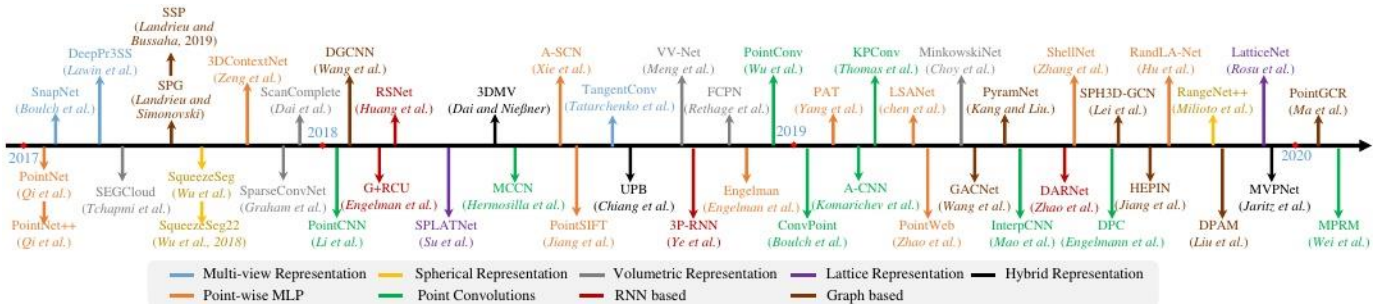


图 10:最相关的基于深度学习的 3D 语义分割方法的时间综述。

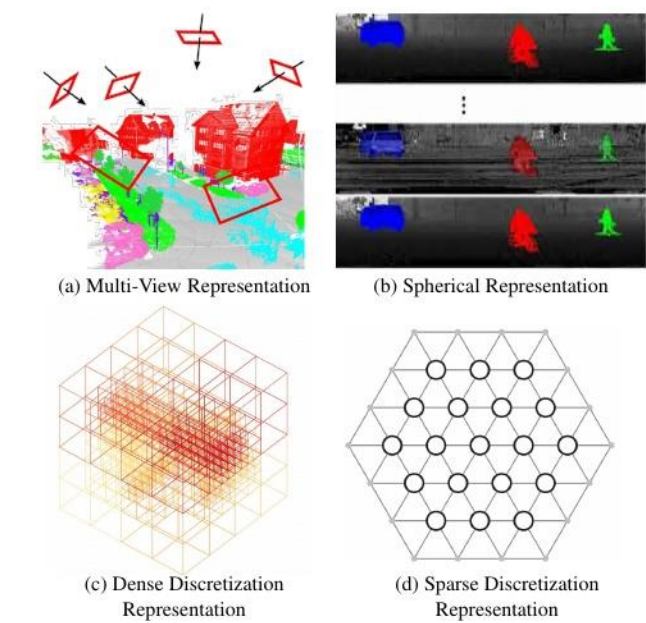


图 11:中间表示的示意图。(a)和(b)最初分别显示在[182]和[183]中。

不可避免地带来离散化误差和遮挡等几个问题。

5.1.2 基于离散化的方法

这些方法通常将点云转换为密集/稀疏的离散表示，如体和稀疏的 permutohedral lattice。

密集离散表示。早期的方法通常将点云体素化为密集网格，然后利用标准的 3D 卷积。Huang 等人[195]首先将点云划分为一组占用体素，然后将这些中间数据馈送到全 3d CNN 进行体素分割。最后，一个体素内的所有点都被分配与体素相同的语义标签。这种方法的性能受到体素粒度和点云划分造成的边界伪影的严重限制。进一步，Tchapmi 等人[196]提出了 SEGCloud 来实现细粒度和全局一致的语义分割。该方法引入了确定性的三线插值，将 3D-FCNN[197]生成的粗体素预测映射回

到点云，然后使用全连接 CRF (FC- CRF)来强制这些推断出的每个点标签的空间一致性。孟等人[186]引入了一种基于核的插值变分自编码器架构来编码每个体素内的局部几何结构。对于每个体素，采用 RBFs 来获得连续的表示，并捕获每个体素中点的分布，而不是二进制占用表示。VAE 进一步用于将每个体素内的点分布映射到一个紧凑的潜空间。然后，利用对称群和等价 CNN 实现鲁棒特征学习。

由于 3D CNN 具有良好的可扩展性，基于体积的网络可以在不同空间大小的点云上进行训练和测试。在全卷积点网络(FCPN)[187]中，首先从点云中分层抽象出不同层次的几何关系，然后使用 3D 卷积和加权平均池化来提取特征并合并长程依赖关系。这种方法可以处理大规模的点云，并且在推理过程中具有良好的可扩展性。Dai 等人[198]提出 ScanComplete 来实现 3D 扫描补全和每个体素语义标注。这种方法利用了全卷积神经网络的可扩展性，可以在训练和测试期间适应不同的输入数据大小。采用由粗到细的策略，分层次地提高预测结果的分辨率。

总的来说，体表示自然地保留了 3D 点云的邻域结构。它的常规数据格式也允许直接应用标准 3D 卷积。这些因素导致了在这一领域的稳定性能提升。然而，体素化步骤内在地引入了离散化伪影和信息损失。通常，高分辨率会导致高内存和计算成本，而低分辨率会引入细节损失。在实践中，选择一个合适的网格分辨率并非易事。

稀疏离散化表示。体表示自然是稀疏的，因为非零值的数量只占很小的百分比。因此，在空间稀疏的数据上应用密集的卷积神经网络是低效的。为此，Graham 等人[166]提出了基于索引结构的子流形稀疏卷积网络。这种方法通过限制卷积的输出只与被占用的体素相关，显著降低了内存和计算成本。同时，它的稀疏卷积也可以控制

提取特征的稀疏性。这种子流形稀疏卷积适用于高效处理高维和空间稀疏数据。进一步，Choy 等人[199]提出了一种 4D 时空卷积神经网络 MinkowskiNet 用于 3D 视频感知。为了有效地处理高维数据，提出了一种广义稀疏卷积。进一步应用三边平稳条件随机场来加强一致性。

另一方面，Su 等人[188]提出了基于双边卷积层(BCLs)的稀疏格网(SPLATNet)。该方法首先将原始点云插值到 permutohedral sparse lattice，然后应用 BCL 对稀疏填充 lattice 中已占据的部分进行卷积。然后将滤波后的输出插值回原始点云。此外，这种方法允许对多视图图像和点云进行灵活的联合处理。进一步，Rosu 等人[189]提出了 LatticeNet 来实现对大型点云的高效处理。还引入了一个名为 DeformsSlice 的数据依赖插值模块，将格特征反向投影到点云。

5.1.3 混合方法

为了进一步利用所有可用信息，人们提出了几种方法来从 3D 扫描中学习多模态特征。Dai 和 Nießner[190]提出了一种联合 3d -多视图网络，以结合 RGB 特征和几何特征。利用 3D CNN 流和若干二维流提取特征，提出了一个可微的反投影层，将学习到的二维嵌入和三维几何特征联合融合。进一步，Chiang 等人[200]提出了一个统一的基于点的框架，从点云中学习 2D 纹理外观、3D 结构和全局上下文特征。该方法直接应用基于点的网络从稀疏采样的点集中提取局部几何特征和全局上下文，不需要任何体素化。Jaritz 等人[191]提出了多视图点网(MVPNet)来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。

5.1.4 基于点的方法

基于点的网络直接在不规则的点云上工作。然而，点云是无序的、无结构的，直接应用标准的 cnn 是不可行的。为此，开创性的点网[5]提出了使用共享 mlp 学习每点特征和使用对称池化函数学习全局特征的方法。在 Point-Net 的基础上，最近又提出了一系列基于点的网络。总的来说，这些方法大致可以分为逐点 MLP 方法、点卷积方法、基于 RNN 的方法和基于图的方法。

点 MLP 方法。 这些方法通常使用共享 MLP 作为其网络的基本单元，以提高其效率。然而，通过共享 MLP 提取的逐点特征无法捕捉点云中的局部几何以及点之间的相互作用[5]。为了为每个点捕获更广泛的上下文并学习更丰富的局部结构，已经引入了几个专用网络，包括基于邻域特征池化、基于注意力的聚合和局部-全局特征连接的方法。

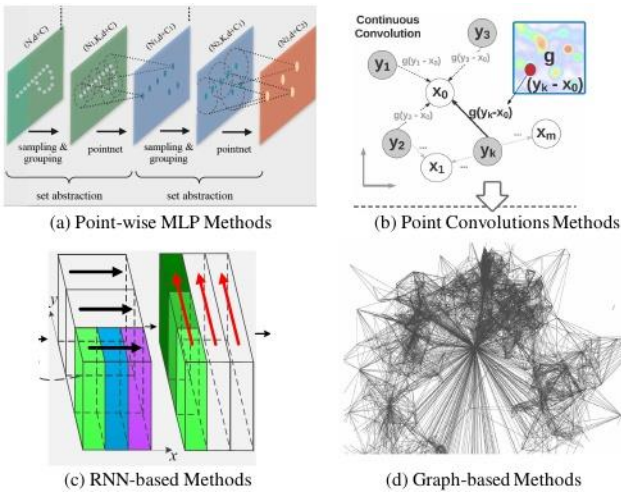


图 12:基于点的方法的自适应。(a)-(d)最初分别显示在[54]、[201]、[202]、[203]中。

邻近特征池化:为了捕获局部几何模式，这些方法通过聚合来自局部邻近点的信息来学习每个点的一个特征。特别是，PointNet++[54]分层分组点，并从更大的局部区域逐步学习，如图 12(a)所示。还提出了多尺度分组和多分辨率分组，以克服点云不均匀和密度变化带来的问题。后来，Jiang 等人[141]提出了 PointSIFT 模块，以实现方向编码和尺度感知。这个模块通过一个三阶段有序卷积，将来自八个空间方向的信息进行堆叠和编码。将多尺度特征串联起来，实现对不同尺度的自适应。与 PointNet++中使用的分组技术(即球查询)不同，Engelmann 等[204]利用 K-means 聚类和 KNN 分别在世界空间和特征空间中定义两个邻域。基于来自同一类的点预计在特征空间中更接近的假设，引入了成对距离损失和质心损失来进一步正则化特征学习。为了模拟不同点之间的相互作用，Zhao 等[57]提出了 PointWeb，通过密集构建局部全链接网络来探索局部区域内所有点对之间的关系。提出了一种自适应特征调整模块来实现信息交换和特征细化。这种聚合操作有助于网络学习具有判别力的特征表示。Zhang et al.[205]基于来自同心球壳的统计数据，提出了一种称为 Shellconv 的置换不变卷积。该方法首先查询一组多尺度同心球，然后在不同的壳内使用 max-pooling 操作来总结统计，使用 MLPs 和 1D 卷积来获得最终的卷积输出。Hu 等人[206]提出了一种名为 RandLA-Net 的高效轻量级网络，用于大规模点云分割。该网络利用随机点采样，在内存和计算方面实现了非常高的效率。进一步提出了一个局部特征聚合模块，以捕获和保留几何特征。

注意力的聚合:为了进一步提高分割精度，将注意力机制[120]引入到点云分割中。Yang 等人[56]提出了一种 group shuffle attention 来对点之间的关系进行建模，并提出了一种排列不变、任务无关和可微的 Gumbel 子集采样(GSS)来取代广泛使用的 FPS 方法。该模块对异常值不太敏感，可以选择一个具有代表性的点子集。为了更好地捕捉点云的空间分布，Chen 等人[207]提出了一种局部空间感知(LSA)层，以基于点云的空间布局和局部结构来学习空间感知权重。与 CRF 类似，Zhao 等人[208]提出了一种基于注意力的分数细化(Attention-based Score Refinement, ASR)模块，对网络产生的分割结果进行后处理。通过将邻域点的分数与学习到的注意力权重进行池化，对初始分割结果进行细化。该模块可以很容易地集成到现有的深度网络中，以提高分割性能。

局部-全局连接:Zhao 等人[112]提出了一种排列不变的 PS-²Net，以结合点云的局部结构和全局上下文。Edgeconv[87]和 NetVLAD[209]被反复堆叠，以捕获局部信息和场景级全局特征。

点卷积方法。这些方法倾向于为点云提出有效的卷积算子。Hua 等人[76]提出了一种逐点卷积算子，将相邻的点分箱成核单元，然后与核权重进行卷积。如图 12(b)所示，Wang et al.[201]提出了一种基于参数化连续卷积层的 PCCN 网络。该层的核函数通过 MLPs 进行参数化，并跨越连续向量空间。Thomas 等人[65]提出了一种基于核点卷积(KPConv)的核点全卷积网络(Kernel Point full Convolutional Network, KP-FCNN)。具体来说，KPConv 的卷积权值是由到核点的欧氏距离决定的，核点的个数是不固定的。将核点的位置表示为球面空间中最佳覆盖的优化问题。注意，半径邻域用于保持一致的感受野，而在每一层中使用网格下采样，以实现在不同密度的点云下的高鲁棒性。在[211]中，Engelmann 等人提供了丰富的消融实验和可视化结果来展示感受野对基于聚合的方法性能的影响。他们还提出了一种扩展点卷积(DPC)操作来聚合扩展的相邻特征，而不是 K 个最近邻。这种操作被证明在增加感受野方面非常有效，并且可以很容易地集成到现有的基于聚合的网络中。

基于 rnn 的方法 为了从点云中捕获固有的上下文特征，递归神经网络(RNN)也被用于点云的语义分割。Engelmann 等人[213]首先基于点网[5]，将点块转换成多尺度块和网格块，获得输入级上下文。然后，将点网提取的块特征依次输入到整合单元(CU)或循环整合单元(RCU)中，以获得输出级上下文。实验结果表明，合并空间上下文对于

提高分割性能。Huang 等[212]提出了一个轻量级的局部依赖建模模块，并利用切片池化层将无序的点特征集转换为有序的特征向量序列。如图 12(c)所示，Ye 等人[202]首先提出了逐点金字塔池化(3P)模块来捕获由粗到细的局部结构，然后利用双向分层 mn 进一步获取长程空间依赖关系。然后应用 RNN 实现端到端学习。然而，这些方法在聚合具有全局结构特征的局部邻域特征时，会失去点云丰富的几何特征和密度分布[220]。为了缓解僵化和静态池化操作带来的问题，Zhao 等人[220]提出了一种动态聚合网络(DAR-Net)来同时考虑全局场景复杂度和局部几何特征。利用自适应感受野和节点权重动态聚合媒介间特征。Liu 等[221]提出 3DCNN-DQN-RNN 用于大规模点云的高效语义解析。该网络首先使用 3D CNN 网络学习空间分布和颜色特征，DQN 进一步用于定位属于特定类别的物体。将最终拼接的特征向量输入残差 RNN 中，得到最终的分割结果。

图论方法。为了捕捉 3D 点云的底层形状和几何结构，一些方法求助于图网络。如图 12(d)所示，Landrieu 等人[203]将点云表示为相互连接的简单形状和叠加点的集合，并使用有属性有向图(即叠加点图)捕获结构和上下文信息。然后，将大规模点云分割问题分解为三个子问题，即几何同质划分、超点嵌入和上下文分割。为了进一步改进分割步骤，Landrieu 和 Boussaha[214]提出了一种监督框架，将点云过度分割为纯超点。这个问题被表述为一个由邻接图构造的深度度量学习问题。此外，还提出了一个图结构的对比损失，以帮助识别对象之间的边界。

为了更好地捕提高维空间中的局部几何关系，Kang 等[222]提出了一种基于图嵌入模块(Graph Embedding Module, GEM)和金字塔注意力网络(Pyramid Attention Network, PAN)的 Pyra-mNet。GEM 模块将点云模拟为有向无环图，并利用协方差矩阵代替欧几里得距离来构建相邻的相似矩阵。PAN 模块使用四种不同大小的卷积核来提取不同语义强度的特征。在[215]中，提出了图注意卷积(Graph Attention Convolution, GAC)，从局部邻近集中选择性地学习相关特征。该操作是通过根据不同邻近点和特征通道的空间位置和特征差异，动态分配注意力权重来实现的。GAC 可以学习捕捉判别特征进行分割，与常用的 CRF 模型具有相似的特征。Ma 等人[223]提出了一个点全局上下文推理(Point Global Context Reasoning, PointGCR)模块，沿通道维度捕获全局上下文信息

表 5:在 S3DIS(包括 Area5 和 6-fold 交叉验证)[10]、Semantic3D(包括 semantic-8 和 reduced8 子集)[12]、ScanNet[11]和 SemanticKITTI[15]数据集上的比较语义分割结果。总体准确率(OA)、平均交并比(mIoU)是主要的评价指标。为简单起见，我们省略了值后面的“%”。符号“-”表示结果不可用。

Method			S3DIS				Semantic3D				ScanNet(v2)		Sem. KITTI (mIoU)
			Area5 (OA)	Area5 (mIoU)	6-fold (mIoU)	6-fold (mIoU)	sem. (OA)	sem. (mIoU)	red. (OA)	red. (mIoU)	OA	mIoU	
Projection-based Methods	Multi-view	DeePr3SS [181]	-	-	-	-	-	-	88.9	58.5	-	-	-
		SnapNet [182]	-	-	-	-	91.0	67.4	88.6	59.1	-	-	-
		TangentConv [193]	82.5	52.8	-	-	-	-	-	-	80.1	40.9	40.9
	Spherical	SqueezeSeg [183]	-	-	-	-	-	-	-	-	-	-	29.5
		SqueezeSegV2 [184]	-	-	-	-	-	-	-	-	-	-	39.7
		RangeNet++ [185]	-	-	-	-	-	-	-	-	-	-	52.2
Discretization-based Methods	Volumetric	SEGCloud [196]	-	48.9	-	-	-	-	88.1	61.3	-	-	-
		SparseConvNet [166]	-	-	-	-	-	-	-	-	-	72.5	-
		MinkowskiNet [199]	-	-	-	-	-	-	-	-	-	73.6	-
	Permutohedral lattice	VV-Net [186]	-	-	87.8	78.2	-	-	-	-	-	-	-
		SPLATNet [188]	-	-	-	-	-	-	-	-	-	39.3	18.4
		LatticeNet [189]	-	-	-	-	-	-	-	-	-	64.0	52.2
Hybrid Methods	Hybrid	3DMV [190]	-	-	-	-	-	-	-	-	-	48.4	-
		UPB [200]	-	-	-	-	-	-	-	-	-	63.4	-
		MVPNet [191]	-	-	-	-	-	-	-	-	-	64.1	-
Point-based Methods	Point-wise MLP	PointNet [5]	-	41.1	78.6	47.6	-	-	-	-	-	-	14.6
		PointNet++ [54]	-	-	81.0	54.5	85.7	63.1	-	-	84.5	33.9	20.1
		PointSIFT [141]	-	-	88.7	70.2	-	-	-	-	86.2	41.5	-
		Engelmann [210]	84.2	52.2	84.0	58.3	-	-	-	-	-	-	-
		3DContextNet [107]	-	-	84.9	55.6	-	-	-	-	-	-	-
		A-SCN [109]	-	-	81.6	52.7	-	-	-	-	-	-	-
		PointWeb [57]	87.0	60.3	87.3	66.7	-	-	-	-	85.9	-	-
		PAT [56]	-	60.1	-	64.3	-	-	-	-	-	-	-
		LSANet [207]	-	-	86.8	62.2	-	-	-	-	85.1	-	-
		ShellNet [205]	-	-	87.1	66.8	-	-	93.2	69.3	85.2	-	-
		RandLA-Net [206]	-	-	88.0	70.0	94.6	74.8	94.8	77.4	-	-	55.9
	Point convolution	PointCNN [79]	85.9	57.3	88.1	65.4	-	-	-	-	85.1	45.8	-
		PCCN [201]	-	58.3	-	-	-	-	-	-	-	-	-
		A-CNN [82]	-	-	87.3	-	-	-	-	-	85.4	-	-
		ConvPoint [66]	-	-	88.8	68.2	93.4	76.5	-	-	-	-	-
		KPConv [65]	-	67.1	-	70.6	-	-	92.9	74.6	-	68.4	-
		DPC [211]	86.8	61.3	-	-	-	-	-	-	-	59.2	-
		InterpCNN [80]	-	-	88.7	66.7	-	-	-	-	-	-	-
	RNN-based	RSNet [212]	-	51.9	-	56.5	-	-	-	-	84.9	39.4	-
		G+RCU [213]	-	45.1	81.1	49.7	-	-	-	-	-	-	-
		3P-RNN [202]	85.7	53.4	86.9	56.3	-	-	-	-	-	-	-
		DGCNN [87]	-	-	84.1	56.1	-	-	-	-	-	-	-
	Graph-based	SPG [203]	86.4	58.0	85.5	62.1	92.9	76.2	94.0	73.2	-	-	17.4
		SSP+SPG [214]	87.9	61.7	87.9	68.4	-	-	-	-	-	-	-
		GACNet [215]	87.8	62.9	-	-	-	-	91.9	70.8	-	-	-
		PAG [216]	86.8	59.3	88.1	65.9	-	-	-	-	-	-	-
		HDGCN [217]	-	59.3	-	66.9	-	-	-	-	-	-	-
		HPEIN [218]	87.2	61.9	88.2	67.8	-	-	-	-	-	61.8	-
		SPH3D-GCN [219]	87.7	59.5	88.6	68.9	-	-	-	-	-	61.0	-
		DPAM [92]	86.1	60.0	87.6	64.5	-	-	-	-	-	-	-

使用无向图表示。PointGCR 是一个即插即用和端到端可训练的模块。它可以很容易地集成到现有的分割网络中，以实现性能提升。

此外，最近的几项工作试图在弱监督下实现点云的语义分割。Wei 等人[224]提出了一种两阶段的方法来训练具有子云级别标签的分割网络。Xu 等[225]对点云语义分割的几种不精确监督方案进行了研究。他们还提出了一种网络，该网络能够仅用部分标记的点(例如 10%)进行训练。

5.2 实例分割

与语义分割相比，实例分割更具挑战性，因为它需要对点进行更准确和细粒度的推理。特别是，它不仅需要区分具有不同语义的点，还需要区分具有相同语义的实例。总的来说，现有的方法可以分为两类:基于提议的方法和无提议的方法。图 13 说明了几个里程碑式方法。

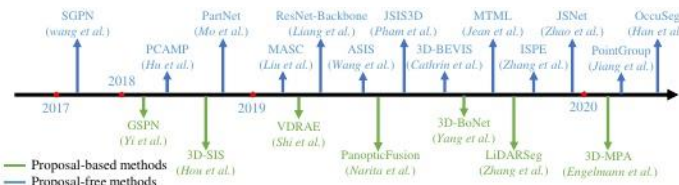


图 13:最相关的基于深度学习的 3D 实例分割方法的时间综述。

5.2.1 基于提议的方法

这些方法将实例分割问题转换为两个子任务:3D 目标检测和实例掩码预测。

Hou 等人[226]提出了一种 3D 全卷积语义实例分割(3D-sis)网络，在 RGB-D 扫描上实现语义实例分割。该网络同时从颜色和几何特征中进行学习。与 3D 物体检测类似，3D 区域提案网络(3D- RPN)和 3D 区域(3D- roi)层用于预测边界框位置、物体类别标签和目标

实例的面具。根据综合分析策略，Yi 等人[227]提出了一种生成形状提案网络(GSPN)来生成高客观度的 3D 提案。这些建议由基于区域的点网络(R-PointNet)进一步细化。最终的标签是通过预测每个类标签的逐点二进制掩码获得的。与从点云直接回归 3D 边界框不同，该方法通过加强几何理解来删除大量无意义的建议框。

Narita 等[228]通过将二维全景分割扩展到三维制图，提出了一种在线体三维制图系统，以共同实现大规模三维重建、语义标记和实例分割。他们首先利用二维语义和实例分割网络来获得像素级全景标签，然后将这些标签集成到体地图中。进一步使用全连接 CRF 来实现精确分割。这个语义映射系统可以实现高质量的语义映射和判别性的物体识别。Yang 等[229]提出了一种称为 3D-BoNet 的单阶段、无锚点和端到端可训练网络，以实现点云上的实例分割。该方法直接回归所有潜在实例的粗糙 3D 边界框，然后利用点级别的二进制分类器来获得实例标签。特别地，边界框生成任务被表述为一个最优分配问题。此外，还提出了一个多准则损失函数来正则化生成的边界框。这种方法不需要任何后处理，计算效率高。Zhang et al.[230]提出了一种网络，用于对大规模室外激光雷达点云进行实例分割。该方法使用自注意力块学习点云鸟瞰图上的特征表示。最终的实例标签是根据预测的水平中心和高度限制获得的。Shi 等人[231]提出了一种分层感知的变分降噪递归自编码器(VDRAE)来预测室内 3D 空间的布局。对象建议通过递归的上下文聚合和传播迭代生成和细化。

总的来说，基于提议的方法[226]、[227]、[229]、[232]直观而直接，实例分割结果通常具有良好的对象化性。然而，这些方法需要多阶段的训练和冗余建议框的修剪。因此，它们通常很耗时，计算成本很高。

5.2.2 无提案方法

无提议方法[233]、[234]、[235]、[236]、[237]、[238]、[239]、[240]没有目标检测模块。相反，它们通常将实例分割视为语义分割之后的后续聚类步骤。特别是，大多数现有的方法都是基于属于同一个实例的点应该具有非常相似的特征的假设。因此，这些方法主要集中在判别性特征学习和点分组上。

在一项开创性的工作中，Wang 等人[233]首次引入了相似组建议网络(SGPN)。该方法首先学习每个点的特征和语义图，然后引入一个相似度矩阵来表示每个配对特征之间的相似度。为了学习更多的判别性特征，他们使用双铰链损失来相互调整

相似度矩阵和语义分割结果。最后，采用启发式非极大值抑制方法将相似点合并为实例。由于构造相似矩阵需要较大的内存消耗，这种方法的可扩展性受到限制。类似地，Liu 等人[237]首先利用子流形稀疏卷积[166]来预测每个体素的语义分数和相邻体素之间的亲和度。然后，他们引入了一种聚类算法，根据预测的亲和度和网格拓扑结构将点分组为实例。Mo 等人[241]在 PartNet 中引入了一个基于分割的检测网络来实现实例分割。pointnet++被用作主干来预测每个点的语义标签和不相交的实例掩码。此外，Liang 等人[238]提出了一种结构感知的损失，用于学习判别式嵌入。这种损失同时考虑了特征的相似性和点之间的几何关系。进一步使用基于注意力的图 CNN，通过聚合来自邻居的不同信息来自适应地细化学习到的特征。

由于点的语义类别和实例标记通常是相互依赖的，已有多方法将这两个任务耦合为一个任务。Wang 等人[234]通过引入端到端可学习的关联分段实例和语义(associative segmentation Instances and Semantics, ASIS)模块整合了这两个任务。实验表明，通过该 ASIS 模块，语义特征和实例特征可以相互支持，从而达到提高性能的目的。同样，Zhao 等[242]提出了 JSNet 来同时实现语义和实例分割。进一步，Pham 等人[235]首先引入了多任务逐点网络(MT-PNet)来为每个点分配标签，并通过引入判别损失来正则化特征空间中的嵌入[243]。然后，他们将预测的语义标签和嵌入融合到一个多值条件随机场(MV-CRF)模型中进行联合优化。最后，使用平均场变分推理来产生语义标签和实例标签。Hu 等人[244]首先提出了一种动态区域生长(DRG)方法来动态地将点云分割成一组不相连的 patch，然后使用无监督的 K- mean++算法对所有这些 patch 进行分组。然后在 patch 之间的上下文信息的指导下进行多尺度 patch 分割。最后，将这些标记的块合并到对象级别，以获得最终的语义和实例标签。

为了在全 3D 场景上实现实例分割，Elich 等人[236]提出了一种混合 2D-3D 网络，从点云的 BEV 表示和局部几何特征中联合学习全局一致的实例特征。然后将学习到的特征进行组合，以实现语义和实例分割。注意，不是启发式的 groupmerge 算法[233]，而是使用更灵活的均值偏移[245]算法将这些点分组到实例中。或者，也引入多任务学习来进行实例分割。Lahoud 等人[246]学习了每个实例的独特特征嵌入和估计对象中心的方向信息。提出了特征嵌入损失和方向损失来调整潜特征空间中学习到的特征嵌入。均值偏移聚类和非最大值抑制

采用 Sion 将体素分组为实例。该方法在 ScanNet[11]基准上取得了最先进的性能。此外，预测的方向信息对于确定实例的边界特别有用。Zhang et al.[247]将概率嵌入引入点云的实例分割。该方法还纳入了不确定性估计，并为聚类步骤提出了一个新的损失函数。Jiang 等[240]提出了一种 PointGroup 网络，它由一个语义分割分支和一个偏移预测分支组成。进一步利用双集聚类算法和 ScoreNet 来实现更好的分组结果。

总而言之，无候选框方法不需要计算昂贵的区域候选框组件。然而，这些方法分组的实例段的对象性通常很低，因为这些方法不显式检测对象边界。

5.3 部分分割

三维形状的部件分割有两方面的困难。首先，具有相同语义标签的形状部件具有较大的几何变异性和歧义性。第二，具有相同语义的物体中的零件数量可能不同。

VoxSegNet[248]被提出，在有限的解决方案下，在 3D 体素化数据上实现细粒度的部件分割。提出了一个空间密集提取(SDE)模块(由堆叠的空洞残差块组成)，用于从稀疏体数据中提取多尺度判别特征。通过逐步应用注意力特征聚合(AFA)模块，进一步重新加权和融合学习到的特征。Kalogerakis 等人[249]结合了 FCNs 和基于表面的 CRFs 来实现端到端的 3D 零件分割。他们首先从多个视图生成图像以实现最优表面覆盖，并将这些图像输入 2D 网络以生成置信度图。然后，这些置信度图由一个基于表面的 CRF 聚合，CRF 负责对整个场景进行一致的标记。Yi 等人[250]引入了同步谱 CNN (SyncSpecCNN)来对不规则和非同构的形状图进行卷积。引入了膨胀卷积核的光谱参数化和光谱 transformer 网络，以解决零件的多尺度分析和跨形状的信息共享问题。

Wang 等[251]首先通过引入形状全卷积网络(shape Fully Convolutional Networks, SFCN)并以三个低级几何特征作为输入，对三维网格进行形状分割。然后，他们利用基于投票的多标签图割来进一步细化分割结果。Zhu et al.[252]提出了一种用于 3D 形状协同分割的弱监督 CoSegNet。该网络将未分割的 3D 点云形状集合作为输入，并通过迭代最小化组一致性损失来产生形状部分标签。与 CRF 类似，提出了一个预训练的部分细化网络，以进一步细化和去噪部分建议。Chen 等人[253]提出了一种分支自编码器网络(BAE-NET)，用于无监督、单次和弱监督的 3D 形状协同分割。该方法将形状协同分割任务表述为表示学习问题，并旨在寻找

通过最小化形状重建损失来实现最简单的部分表示。基于编码器-解码器架构，该网络的每个分支都可以学习特定零件形状的紧凑表示。然后将从每个分支学习到的特征和点坐标馈送给解码器，以产生一个二进制值(表示该点是否属于该部分)。这种方法具有良好的泛化能力，可以处理大型 3D 形状集合(多达 5000+形状)。但是，它对初始参数敏感，并且没有将形状语义纳入到网络中，这阻碍了该方法在每次迭代中获得鲁棒和稳定的估计。Yu et al.[254]提出了一种自上而下递归的部分分解网络(PartNet)用于分层形状分割。与现有的将形状分割为固定标签集的方法不同，该网络将零件分割制定为级联二进制标记问题，并根据几何结构将输入点云分解为任意数量的零件。Luo 等人[255]针对零样本 3D 零件分割任务引入了一种基于学习的分组框架。为了提高跨类别泛化能力，该方法倾向于学习一种分组策略，限制网络在部件局部上下文内学习部件级特征。

5.4 总结

表 5 给出了现有方法在公共基准上的结果，包括 S3DIS[10]、Semantic3D[12]、ScanNet[39]和 SemanticKITTI[15]。以下问题需要进一步研究:

- 由于常规的数据表示，基于投影的方法和基于离散化的方法都可以利用其 2D 图像对应的成熟网络架构。然而，基于投影的方法的主要限制在于 3D-2D 投影造成的信息损失，而基于离散的方法的主要瓶颈是由于分辨率的增加而导致的三倍增加的计算和内存成本。为此，建立在索引结构上的稀疏卷积将是一个可行的解决方案，值得进一步探索。
- 基于点的网络是最常被研究的方法。然而，点表示自然没有明确的邻域信息，大多数现有的基于点的方法求助于昂贵的邻域搜索机制(例如 KNN[79]或球查询[54])。这内在限制了这些方法的效率，最近提出的点-体素联合表示[256]将是一个值得进一步研究的有趣方向。
- 在点云分割中，从不平衡数据中学习仍然是一个具有挑战性的问题。虽然几种方法[65]、[203]、[205]取得了显著的整体性能，但它们在少数类上的表现仍然有限。例如，RandLA-Net[206]在 Semantic3D 的 reduce -8 子集上实现了 76.0%的总体 IoU，但在硬景观类上实现了 41.1%的非常低的IoU。

对标城市检测、分割和分类方法”，《ICRA》，2014 年。

[35]张晓明，张晓明，张晓明，“基于 gis 的城市点云分析”，《城市信息学报》，第 4 卷，第 3 期，第 1 - 3 页，2015。

[36]王志强，刘志强。Deschaud 和 F. Goulette，“Paris-lille-3d:用于自动分割和分类的大型高质量地真城市点云数据集”，IJRR, 2018。

[37]谭文，秦宁，马展，李旸，杜军，蔡国光，杨，李旸，“基于 gis 的城市道路语义分割”，arxi, 2003.08 - 284,2020。

[38]张晓明，张晓明，张晓明，“基于多尺度的航空激光雷达数据分割方法研究”，中国航空学报，2004,21(1):1 - 3。

[39]陈晓霞，张之，卢慧，周清，马展，赵，“基于空间通道注意力网络的三维目标检测”，计算机工程学报，2019。

[40]李晓明，苏竑，李晓明，“基于多视图卷积神经网络的三维形状识别”，计算机工程学报，2015。

[41]于涛，孟军，袁军，“多视图协调双线性网络在三维物体识别中的应用”，CVPR, 2018。

[42]杨，王，“多视图三维物体识别的学习关系”，计算机工程，2019。

[43]祁春荣，苏竑，聂明，戴安，闫明，gul . J.，“基于多视角的 CNNs 图像分类方法”，CVPR, 2016。

[44]冯勇，张之，赵晓，纪仁，高勇，“基于群视图卷积神经网络的三维形状识别”，CVPR, 2018。

[45]王，王晓明，王晓明，“基于聚类的多视角三维物体识别”，计算机工程，2017。

[46]马展，郭，杨，安文，“基于 LSTM 的三维形状识别与检索”，机械工程学报，2018。

[47]于涛，魏新，孙军，“基于图像卷积神经网络的三维形状分析”，CVPR, 2020。

[48]张晓明，张晓明，“一种基于卷积神经网络的实时目标识别方法”，计算机应用学报，2015。

[49]张建军，张建军，张建军，“基于深度学习的高分辨率三维图像识别技术”，CVPR, 2017。

[50] p - s。王，刘玉霞，刘玉霞。郭。孙、童，“O- CNN:基于八叉树的卷积神经网络的三维形状分析”，中国机械工程学报，2017。

[51]李涛，段勇，“基于点云的三维形状识别”，CVPR, 2018。

[52]张晓明，张晓明，张晓明，等。基于卷积神经网络的三维点云图像分割方法研究[j].计算机工程，计算机工程，2017。

[53]张晓明，张晓明，张晓明，“深度集”，《神经科学》，2017。

[54]齐正荣，易磊，苏慧，李建奎，“基于点集的深度层次特征学习”，《神经科学》，2017。

[55]张晓明，张晓明，张晓明，“基于模糊神经网络的图形分类方法研究”，中文信息学报，2018。

[56]杨军，张琪，倪波，李丽，刘勇，周，田强，“基于自关注和 gumbel 子集采样的点云建模”，CVPR, 2019。

[57]赵红华，蒋丽文。傅，贾杰，“点网络:增强点云处理的低局部邻域特征”，CVPR, 2019。

[58]段勇，郑旸，鲁京，周，田强，“点云的结构关系推理”，CVPR, 2019。

[59]林辉，肖之，谭勇，丁生，“基于查找表的点云特征提取方法”，计算机工程学报，2019。

[60]孙晓霞，李志强，肖，“基于旋转不变性的点云分类和分割方法”，计算机工程学报，2019。

[61]郑旸，李丽，闫晓霞，崔顺生，“基于自适应采样的非局部神经网络的点云处理”，CVPR, 2020。

[62]刘勇，范斌，向生，潘翀，“关系形卷积神经网络在点云分析中的应用”，CVPR, 2019。

[63]李志刚，“基于离散卷积的非结构化点云算法”，《计算机工程》，计算机工程第 1 卷第 1 期，2019。

[64]刘勇，范斌，孟刚，鲁京，向生，潘朝旭，“基于深度学习的点云处理方法”，计算机工程学报，2019。

[65]齐正荣，李建平。Deschaud, B. Marcotegui, F. Goulette 和 L. J. gu，“KPConv:点云的柔性和可变形卷积”，ICCV, 2019。

[66]李志刚，“基于深度卷积的点云处理”，《计算机工程学报》，2012。

[67]吴伟，齐正荣，吴伟。基于深度卷积神经网络的三维点云研究，《CVPR》，2019。

[68]李建平，李建平。V' azquez, A. Vinacia 和 T. Ropinski，“在非均匀采样点云上学习的蒙特卡罗卷积”，ACM TOG, 2018。

[69]徐旸，范涛，徐旸，“基于参数化卷积滤波器的蜘蛛 cnn 深度学习”，《ECCV》，2018。

[70]张晓明，张晓明，张晓明，“基于扩展算子的点卷积神经网络”，ACM TOG, 2018。

[71]张晓明，张晓明，张晓明，“基于球面 cnn 网络的学习方法”，《ECCV》，2017。

[72]杨丽娟，李理，杨丽娟，张晓明，张晓明，“三维点云的旋转和平移等变神经网络”，物理学报，物理学报，第 1 卷第 1 期，2018。

[73]张晓明，张晓明，张晓明，“球面 cnn 网络”，中国生物医学工程学报，2018。

[74]王志强，王志强。Rakotosaona, Y. Ponty 和 M. Ovsjanikov，“具有球面谐波核的有效旋转不变量点 CNN”，3DV, 2019。

[75]张晓明，张晓明，“柔性卷积算法”，计算机应用学报，2018。

[76] b - s。华，m.k. Tran 和 s - k. Yeung，“点卷积神经网络”，CVPR, 2018。

[77]李洪波，张建军，张建军，“基于八叉树的三维点云神经网络”，《CVPR》，2019。

[78]于国，于国，兰淑娟，“基于地理-cnn 的三维点云局部几何结构建模”，CVPR, 2019。

[79]李勇，孙明，吴伟，陈斌，“基于卷积神经网络的 x 变换点卷积算法”，计算机应用学报，2018。

[80]王晓明，王晓明，李理，“基于插值卷积网络的三维点云识别”，计算机工程学报，2019。

[81]张志强。华，D. W.罗森，s.k. 杨，“3D 点云深度学习的旋转不变卷积”，《3DV》，2019 年。

[82]钟志强，华杰，“基于点云的卷积神经网络”，《CVPR》，2019。

[83]张晓明，张晓明，“三维卷积神经网络的局部相位分析”，《CVPR》，2019。

[84]刘建军，周建军，周建军，“球面分形卷积神经网络在点云识别中的应用”，CVPR, 2019。

[85]张晓明，张晓明，“基于卷积神经网络的动态边缘条件滤波”，《CVPR》，2017。

[86]刘志刚，“三维空间的研究:点云库(PCL)”，《计算机科学》，2011。

[87]王，孙勇，刘杰，孙志刚，s.e. Sarma, M. M. Bronstein, J. M. Solomon，“基于点云的动态图 CNN 学习”，中国计算机学报，2019。

[88]张凯，郝明，王，付春华，“基于链接层次特征的点云 CNN 学习”，arXiv 预印本 arXiv:1904 - 10014, 2019。

[89]杨，沈勇，田东，“基于深度网格变形的点云自编码器”，CVPR, 2018。

[90]陈晓明，李晓明，李晓明，李晓明，李晓明，“基于卷积的深度学习”，《CVPR》，2015。

[91]王晓明，王晓明，“基于点云的无监督多任务特征学习”，计算机应用学报，2019。

[92]刘杰，倪波，李光，杨，田东，“基于点集的动态聚类学习”，计算机工程学报，2019。

[93]沈勇，冯超，杨，田东，“基于核关联和图池的点云局部结构挖掘”，CVPR, 2018。

[94]张晓明, 张晓明, 张晓明, “基于深度点云的特征提取方法”, 计算机应用学报, 2018。

[95]陈涛, 李光, 徐强, 陈涛, 王, 林玲, “基于深度层次聚类网络的点云分析”, CVPR, 2019。

[96]张晓明, 张晓明, “基于聚类的聚类算法”, arXiv, arXiv, 2011,25(4):559 - 559。

[97]徐强, 孙晓明, 孙春英. Wu, P.王和 U. Neumann, “网格-gcn 用于快速和可扩展的点云学习”, CVPR, 2020。

[98]张晓明, 张晓明, 张晓明, “基于局部连通网络的频谱网络研究”, 《计算机科学》, 2014。

[99]张晓明, 张晓明, “基于卷积神经网络的快速局部谱滤波”, 《神经网络学报》, 2016。

[100]郭忠, 胡伟, 郑安, “基于 RGCNN 的点云分割”, 中国计算机工程学报, 2018。

[101]李劼, 王生, 朱峰, 黄军, “自适应图卷积神经网络”, AAAI, 2018。

[102]冯勇, 游红, 张之, 日然, 高翼, “超图神经网络”, AAAI, 2019。

[103]王晓明, 王晓明, “基于局部谱图卷积的点集特征学习”, 《ECCV》, 2018。

[104]张之, 张晓明, 张晓明, “一种基于 cnn 的三维点云分类方法”, 计算机应用学报, 2018。

[105]潘国光, 王生, 应然, 刘, “基于动态 GCN 的三维几何特征学习方法”, arXiv 预印本 arXiv:1812.06254, 2018。

[106]王晓明, 张晓明, “基于深度 kd 网络的三维点云模型识别”, 计算机工程学报, 2017。

[107]陈晓明, “基于局部和全局上下文线索的点云学习方法研究”, 《ECCV》, 2018。

[108]李劼, 陈之, 李国辉, “基于自组织网络的点云分析”, CVPR, 2018。

[109]谢生, 刘, 陈之, 杜志强, “基于图像的点云识别”, CVPR, 2018。

[110]游红, 冯勇, 日然, 高翼, “基于多视角的三维图像识别方法”, 计算机工程学报, 2018。

[111]游红, 冯勇, 赵旸, 邹翀, 日然, 高翼, “基于 PVRNet 的三维形状识别方法”, AAAI, 2019。

[112]张晓明, 赵旸, 邓海燕, 张晓明, “三维点囊网络”, CVPR, 2019。

[113]陈之, 韩新, 李劼, 陈之, 邢建军, 赵旸, 李劼, “基于径向基函数的深度 RBFNet 点云特征学习”, arXiv 预印本 arXiv:1812.04302, 2018。

[114]刘, 韩新, 杨永生. 刘和 M. Zwicker, “Point2Sequence:使用基于注意力的序列到序列网络学习 3D 点云的形状表示”, AAAI, 2019。

[115]吴鹏, 陈之, 易建军, “基于循环集编码的点云处理”, 《AAAI》, 2019。

[116]秦春, 尤海华, 王利, 等。郭杰, 傅, “PointDAN:用于点云表示的多尺度 3D 域自适应网络”, NIPS, 2019。

[117]王晓明, “基于点云的自监督深度学习”, 《NIPS》, 2019。

[118] R. Li, X. Li, p.a. 恒, 陈志伟. 傅, “点增强:点云分类的自动增强框架”, CVPR, 2020。

[119]张晓明, 张晓明, “一种基于图像识别的图像识别方法”, 计算机应用学报, 2002。

[120]张晓明, 张晓明, 张晓明, 等。Kaiser, and I. Polosukhin, 《Attention is all you need》, 2017 年, NeurIPS。

[121]陈建军, 陈建军, 陈建军, “基于点对描述符的三维点云深度学习算法”, 《电子工程学报》, 2018。

[122]张晓明, 张晓明, 张晓明, “基于点云的深度学习方法”, 计算机应用学报, 2019。

[123]刘生, 欧阳文, 王晓, 陈, 刘晓, “基于深度学习的通用目标检测方法研究”, 中国计算机工程, 2020。

[124]王晓明, 王晓明, 王晓明, “基于深度投票的点云三维目标检测”, 计算机工程学报, 2019。

[125]石生, 李建军, “基于点云的三维目标检测方法研究”, CVPR, 2020。

[126]张晓明, 张晓明, 李晓明, “基于图像聚合的三维目标识别方法”, 计算机应用学报, 2018。

[127]梁明, 杨 b, 王, R. Urtasun, “基于深度连续融合的多传感器三维目标检测”, 中国机械工程, 2018。

[128]梁明, 杨 b, 陈, 胡, Urtasun, “多任务多传感器融合三维目标检测”, CVPR, 2019。

[129]杨 b, 罗文, R. Urtasun, “基于像素点云的实时 3D 目标检测”, CVPR, 2018。

[130]罗伟, 杨 b, R. Urtasun, “速度与激情:基于单卷积网络的实时端到端 3D 检测”, CVPR, 2018。

[131]曾艳, 胡, 刘生, 叶建军, 韩勇, 李辉, 孙宁, “面向自动驾驶的激光雷达点云实时 3D 车辆检测”, 电子工程学报, 2018。

[132]杨 b, 孙勇, 刘生, 沈晓霞, 贾建军, “基于点云的目标检测方法”, arXiv, 预印本 arXiv, 2018。

[133]石生, 王晓, 李辉, “基于点云的三维目标识别方法”, CVPR, 2019。

[134]张晓明, 张晓明, 张晓明, “基于图像卷积神经网络的三维车辆检测方法研究”, 预印版 arXiv, 2016。

[135]张晓明, 张晓明, 张晓明, “基于图像融合的三维目标检测方法”, 计算机工程学报, 2016。

[136]张晓明, 周. 基于点云的三维物体检测方法[j]. 计算机工程, 2018。

[137]王小明, 王小明, 王小明, 周, 杨然, “一种基于点云的快速目标检测方法”, CVPR, 2019。

[138]杨然, 刘生, 沈晓, 贾佳, “基于点云的三维目标检测方法”, 计算机工程学报, 2019。

[139]刘生, 吴翀, 齐长荣, 苏宏, 桂英杰, “基于 RGB-D 数据的截锥体点网”, CVPR, 2018。

[140]赵晓霞, 刘志, 胡仁, 黄凯, “基于尺度不变性和特征重加权的三维目标检测”, 中国机械工程, 2019。

[141]江明, 吴翀, 陆超, “一种基于点云语义分割的网络模型”, arXiv 预印本 arXiv: 187.00652, 2018。

[142]张晓明, 张晓明, “基于深度传感器融合的三维边界盒估计”, CVPR, 2018。

[143]王小明, 王小明, 王小明, “基于区域逼近的三维目标检测方法”, 计算机应用学报, 2019。

[144]王勇, 贾佳, “截锥体卷积神经网络:基于局部点特征的模态三维目标检测”, 中国机械工程, 2019。

[145]张晓明, 张晓明, 张晓明, 等. 三维目标识别技术的研究进展 [J]. arXiv, 2016, (4):1 - 4。

[146]周, 方军, 宋晓霞, 关超, 殷建军, 戴勇, 杨然, “二维/三维目标检测中的相位损失”, 电子学报, 2019。

[147]陈勇, 刘生, 沈晓, 贾佳, “Fast point r-cnn”, 中国机械工程, 2019。

[148]石生, 郭超, 江明, 王勇, 石生, 王勇, 李华, “基于点体素的三维目标检测方法”, CVPR, 2020。

[149]冯明, 王勇, 张丽丽, 米安, “基于点云的三维目标检测方法”, arXiv, 预印本 arXiv, 2001,12 - 12, 2019。

[150]齐长荣, 陈晓, 张志强, L. J. gu, “基于图像投票的点云三维目标检测”, CVPR, 2020。

[151]王勇, 王勇, 石生, 李宏, “基于点云的三维物体识别方法”, 机械工程学报, 2020。

[152]杨斌, 梁明, R. Urtasun, “基于高清地图的三维目标检测”, 测绘科学, 2018。

[153]张晓明, 张晓明, 张晓明, “基于激光雷达的三维目标识别技术”, 计算机应用, 2018。

[154]李, 张涛, 夏涛, “基于全卷积神经网络的三维激光雷达车辆检测”, arXiv 预印本 arXiv: 168.07916, 2016。

[155]李, “基于卷积神经网络的车辆识别”, 计算机应用, 2017。

[156]王志强, 王志强, 王志强, Engelcke, “基于卷积神经网络的三维点云快速目标检测”, 《ICRA》, 2017。

[157]李旭, 郭宁, 徐勇, 李, “三维目标检测的主干网”, 计算机工程学报, 2019。

[158]李斌, 闫燕, 毛燕, 李, “基于稀疏嵌入的卷积检测”, 计算机工程, 2018。

[159]张建军, 周勇, 张晓明, “基于多模态体素网的三维目标检测”, 《ICRA》, 2019。

[160]何超, 曾宏华, 黄建军. 张涛, “基于点云的结构感知单阶段3D目标检测”, CVPR, 2020。

[161]杨志强, 孙勇, 刘顺生, 贾锴, “基于点的三维单级目标检测”, CVPR, 2020。

[162]李建军, 李建军, 李建军, g.p. Meyer, A. Laddha, E. Kee, C. vallespiz - gonzalez, C. K. Wellington, “激光网络的高效概率三维目标检测”, CVPR, 2019。

[163]张晓明, 张晓明, 张晓明, “基于图像融合的三维目标检测方法”, 计算机应用, 2016。

[164]陈庆, 孙磊, 王志, 贾锴, 王宇, “一种基于热点的三维目标检测方法”, arxi, 预印本 arxi, 2001,12 - 12, 2019。

[165]张晓明, 张晓明, “基于卷积神经网络的生物医学图像分割”, 中国生物医学工程学报, 2015,pp。

[166]张晓明, 张晓明, 张晓明, “基于卷积神经网络的三维语义分割”, 计算机工程学报, 2018。

[167]胡清, 陈旸, 肖, 安文, “基于开环系统的相关滤波跟踪”, 计算机工程学报, 2017。

[168]刘新, 胡清, 李索, 郭英, “基于实例的鲁棒长期跟踪”, 电子学报, 2019。

[169]张晓明, 张晓明, 张晓明, “基于卷积神经网络的目标跟踪”, 计算机工程学报, 2016。

[170]张晓明, 张晓明, 张晓明, “基于几何形状补全的三维连体跟踪”, CVPR, 2019。

[171]张晓明, 王小明, “情境感知关联滤波跟踪”, 《CVPR》, 2017。

[172]张晓明, 张晓明, 张晓明, 等. 基于三维连体网络的激光雷达跟踪方法研究[j]. 计算机工程学报, 2016,33(4):557 - 557,2019。

[173]张晓明, 张晓明, 张晓明, 张晓明, “基于语义点云的实时三维目标检测与跟踪”, 计算机工程学报, 2019。

[174]齐竑, 冯春春, 曹志军, 赵峰, 肖, “基于点对盒网络的三维目标跟踪”, CVPR, 2020。

[175]刘新, 齐竑, 桂英杰, “FlowNet3D: 三维点云的场景流学习”, CVPR, 2019。

[176]王志, 李索, 陈明, “基于 FlowNet3D++ 的深度场景流估计方法”, 计算机应用学报, 2020。

[177]顾晓霞, 王勇, 吴春春, 李彦杰, 王鹏, “基于点云的场景流估计方法研究”, CVPR, 2019。

[178]范辉, 杨勇, “基于点回归神经网络的移动点云处理”, 预印本预印本, 预印本, 2019。

[179]刘晓东, 闫明, 刘建军, “基于深度学习的三维点云模型”, 计算机应用学报, 2019。

[180]张晓明, 张晓明, 张晓明, “基于自监督的场景流估计”, CVPR, 2010。

[181]张晓明, 张晓明, 张晓明, “深度投影3D语义分割”, 中文信息学报, 2017。

[182]张晓明, 张晓明, 张晓明. 基于深度分割网络的非结构化点云语义标注。“在

[183]吴波, 万安, 余晓, K. Keutzer, “基于卷积神经网络的三维激光雷达点云实时道路目标分割”, ICRA, 2018。

[184]吴波, 周晓霞, 赵树生, 岳晓, K. Keutzer, “基于激光雷达点云的道路目标分割方法研究”, ICRA, 2019。

[185]张晓明, 张晓明, 张晓明, “激光雷达语义分割技术研究进展”, 中国激光学报, 2019。

[186]张志强. 孟, 高丽丽, 杨凯. Lai 和 D. Manocha, “VV-Net:用于点云分割的组卷积体素vae网络”, ICCV, 2019。

[187]张晓明, 张晓明, 张晓明, “基于卷积神经网络的大尺度点云”, 《ECCV》, 2018。

[188]李建军, 孙立军, 孙立军, 苏竑. Yang 和 J. Kautz, “SplatNet:用于点云处理的稀疏晶格网络”, CVPR, 2018。

[189]张晓明, 张晓明, 张晓明, 等. 基于多面体网格的点云分割方法[j]. 预印版, 预印版, 2001,12(4):559 - 559。

[190]戴晓明, 陈晓明, “基于多视角的三维语义场景分割方法”, 《ECCV》, 2018。

[191]顾建军, 苏洪波, 李建军, “基于多视角的三维场景识别技术”, 计算机应用学报, 2019。

[192]张晓明, 张晓明, 张晓明, “基于多模态和多尺度深度网络的遥感数据语义分割”, 遥感学报, 2016。

[193]张建军, 张建军, 张建军. 周, “三维密集预测的正切卷积”, 《CVPR》, 2018。

[194]张晓明, 张晓明, 张晓明, 张晓明, “一种基于数据集的数据集优化算法”, 应用计算机学报, 2016。

[195]黄建军, 尤生, “基于卷积神经网络的点云标记”, 计算机工程学报, 2016。

[196]张晓明, 张晓明, 张晓明, “基于数据集的三维点云语义分割方法”, 计算机应用学报, 2017。

[197]王小明, 王小明, “基于深度卷积神经网络的语义分割”, 中文信息学报, 2015。

[198]李晓明, 李晓明, 李晓明, “基于图像的三维图像补全方法研究”, 计算机工程学报, 2018。

[199]郭志强, 蔡志强, “基于卷积神经网络的时空卷积神经网络”, 中国计算机工程学报, 2019。

[200]王志强. 蒋, y.l. 林. 研究. 刘和徐文辉, “3D 分割的统一基于点的框架”, 3DV, 2019。

[201]王士生, 索士生. Ma, A. Pokrovsky 和 R. Urtasun, “深度参数连续卷积神经网络”, CVPR, 2018。

[202]叶晓霞, 李璟, 黄晴, 杜磊, 张, “基于上下文融合的三维递归神经网络在点云语义分割中的应用”, ECCV, 2018。

[203]张晓明, 王小明, “基于点云的语义分割方法”, 《CVPR》, 2018。

[204]张晓明, 张晓明, 张晓明, F. Engelmann, T. Kontogianni, J. Schult, B. leibbe, “基于点云的三维语义分割”, 计算机工程, 2018。

[205]张. 华志强. 杨, “ShellNet:使用同心壳统计的高效点云卷积神经网络”, 在 ICCV, 2019 年。

[206]胡强, 杨斌, 谢磊, 郭勇, 王志, “基于 RandLA-Net 的点云语义分割”, CVPR, 2020。

[207]李志强. 陈晓宇. 李璟, d.p. 风扇, m m. 程凯, 王伟, Lu S.P, “LSANet:基于局部空间注意力的点集特征学习”, arXiv 预印本 arXiv:1905.05442, 2019。

[208]赵青, 周文伟, Lu S.P, 赵青, “基于点云的三维点云分割方法”, 计算机工程学报, 2019。

[209]张晓明, 张晓明, 张晓明, “一种基于神经网络的弱监督位置识别方法”, CVPR, 2016。

[210] F. Engelmann, T. Kontogianni, J. Schult, B. leibbe, “基于点的三维语义分割”云, ” 《ECCV》, 2018

[211] F. Engelmann, T. Kontogianni, B. leibbe, “扩展的点卷积:点卷积的接受场”, 中国计算机学报, 2020。

[212]黄晴, 王伟, “基于回归切片网络的点云三维分割”, CVPR, 2018。

[213]张晓明, 张晓明, 张晓明, “基于空间上下文的点云空间语义分割方法研究”, 计算机工程学报, 2017。

[214]王晓明, 王晓明, “基于深度度量学习的点云过度分割”, CVPR, 2019。

[215]王磊, 黄燕, 张淑琴, 单军, “基于图像关注卷积的点云语义分割”, 中文信息学报, 2019。

[216]李建军, 李建军. Chew 和 g.h. Lee, “Pointatrousgraph:点云的深度分层编码器-解码器”, arXiv 预印本 arXiv:1907.09798, 2019。”

[217]梁志强, 杨明, 邓丽丽, 王超, 王超, “基于层次深度图卷积神经网络的点云三维语义分割”, 中文信息学报, 2019。

[218]蒋亮, 赵伟, 刘峰, 沈晓文. 傅, 贾杰, “用于点云语义分割的分层点-边缘交互网络”, ICCV, 2019。

[219]雷伊, 张晓明, 张晓明, “基于卷积神经网络的三维点云计算”, arXiv, arXiv, 第 4 卷第 1 期, 2018。

[220]赵志, 刘峰, 张志强, “基于语义的场景分割”, arXiv 预印本 arXiv: 2020,2019。

[221]刘峰, 李生, 张列, 周春春, 叶仁, 王, 陆军, “基于深度强化学习的三维点云语义分析”, 中文信息学报, 2017。

[222]康志强, 李生, “基于点云的金字塔关注网络和图像嵌入模型”, 计算机应用学报, 2019。

[223]马艳, 郭艳, 刘峰, 雷伊, 温国光, “三维点云语义分割的全局上下文推理”, 中文信息学报, 2020。

[224]林国强, 刘建军.是的, t - y. 洪磊, 谢磊, “基于点云的弱监督 3D 语义分割的多路径区域挖掘”, CVPR, 2020。

[225]徐锴, 李国辉, “弱监督语义点云分割方法的研究”, 《CVPR》, 2020,pp: 13706 - 13715。

[226]王晓明, 王晓明, “基于 RGB-D 图像的三维语义实例分割”, 《CVPR》, 2019。

[227]李磊, 赵伟, 王, 宋明明, gu, “基于 GSPN 的点云三维实例分割”, CVPR, 2019。

[228]陈晓明, 陈晓明, 陈晓明, “基于图像的三维立体语义映射”, 计算机应用学报, 2019。

[229]杨晓明, 王, 胡清, 王, A. Markham, N. Trigoni, “基于点云的三维实例分割方法”, 计算机工程学报, 2019。

[230]张飞, 关超, 方建军, 白淑娟, 杨然, “激光雷达点云的实例分割”, 中国激光学报, 2020。

[231]史勇, 张爱祥, 吴振华, 徐锴, “基于层次去噪的递归自编码器的三维场景布局预测”, CVPR, 2019。

[232]张晓明, 张晓明, 张晓明, “一种基于深度学习的三维语义实例分割方法”, 《CVPR》, 2012。

[233]王, 余仁杰, 黄强, 王勇, “基于相似群建议的三维点云实例分割”, CVPR, 2018。

[234]王新, 刘生, 沈新, 沈新, 贾杰, “点云的关联分割实例和语义”, CVPR, 2019。

[235]王志强. Pham, T. Nguyen, b.s. 华, G.罗伊和 s . k .。杨, “JSIS3D:基于多任务点向网络和多值条件随机场的三维点云联合语义实例分割”, CVPR, 2019。

[236]张晓明, 张晓明, 张晓明, “三维立体图像分割技术”, 计算机应用, 2019。

[237]刘生, 古川元, “基于稀疏卷积的多尺度图像分割方法”, arXiv, 第 1 卷第 1 期, 2019。

[238] z 梁、m • 杨和王, “3 d 图形嵌入学习支持结构损失函数的点云分割语义实例”, “arXiv 预印本 arXiv: 1902.05247, 2019。

[239]韩磊, 郑涛, 徐磊, 方磊, “基于占位的三维实例分割”, CVPR, 2020。

[240]蒋丽丽, 赵, 石生, 刘生. 傅, 贾杰, “点组:三维实例分割的双设定点分组”, CVPR, 2020。

[241]朱淑霞, 张爱祥, 李丽娟, 苏兹, “基于 3D 模型的零件识别方法研究”, CVPR, 2019。

[242]赵磊, 赵, 陶伟, “基于 JSNet 的三维点云语义分割”, 计算机工程学报, 2020。

[243]王晓明, 王晓明, 王晓明, “基于语义的语义实例分割”, 中文信息学报, 2017。

[244]李小明. 胡, J.-X. 蔡、y.k. Lai, “基于补丁上下文分析和多尺度处理的 3D 点云语义标记和实例分割”, IEEE TVCG, 2018。

[245]张志强, “基于数据的特征空间分析方法”, IEEE, 2002。

[246]张晓明, 张晓明, 张晓明, “基于多任务度量学习的三维实例分割”, 计算机工程学报, 2019。

[247]张晓明, 张晓明, “基于概率嵌入的点云实例分割”, arXiv 预印本 arXiv: 2020,2019。

[248]王振, 卢峰, “基于本体的三维几何图形语义分割”, IEEE TVCG, 2019。

[249]张晓明, 张晓明, 张晓明, “基于投影卷积神经网络的三维图像分割”, 《CVPR》, 2017。

[250]Yi l, 苏慧, 郭旭, 桂巴斯, “基于同步光谱 CNN 的三维形状分割”, CVPR, 2017。

[251]王鹏, 甘艳, 水平, 于峰, 张, 陈生, 孙忠, “基于全卷积网络的三维形状分割”, 计算机与图形学, 2018。

[252]徐, S. Chaudhuri, Yi l, gu, 张, “基于群一致性损失的三维图形深度共分割方法”, arXiv 预印本 arXiv: 193.10297, 2019。

[253]陈志, 殷建军, S. Chaudhuri, 张, “基于 bp - NET 的自编码器的形状共分割”, 计算机工程学报, 2019。

[254]于峰, 刘康, 张, 朱翀, 徐, “基于递归的零件分割网络”, CVPR, 2019。

[255]罗涛, 莫昆, 黄志强, 徐, 胡索, 王鹏玲, 苏慧, “基于自适应的三维零件分类发现”, 中国机械工程, 2020。

[256]刘志强, 唐辉, 刘志强, “基于点体元的 CNN 深度学习方法”, 应用科学学报, 2019。



郭旭, 现任副教授。2008 年获国防科技大学工学学士学位, 2015 年获 NUDT 博士学位。2011 年至 2014 年在西澳大利亚大学访问博士研究生。2016 年至 2018 年在中国科学院计算技术研究所任博士后研究员。他在 IEEE TPAMI 和 IJCV 等期刊和会议上发表了 90 多篇文章。他的。

目前的研究兴趣集中在 3D 视觉, 特别是 3D 特征学习, 3D 建模, 3D 物体识别和场景理解。郭旭获得 2019 年 ACM 中国 SIGAI 新星奖、2019 年吴文俊优秀 AI 青年奖、2016 年中国计算机学会优秀博士论文奖。他曾担任 IET 计算机视觉和 IET 图像处理的副编辑, IEEE TPAMI 的客座编辑, CVPR 2021 和 ICPR 2020 的区域主席。



IET 图像处理。

王鹏，2015 年毕业于国防科技大学，获博士学位。2011 - 2014 年厦门大学访问博士生。他在 IEEE TGRS 和 IEEE TITS 等期刊和会议上发表了 20 多篇文章。他的研究兴趣包括移动激光扫描数据分析和 3D 计算机视觉，特别是在 3D 目标检测和 3D 场景理解方面。他还担任了许多期刊的审稿人，如 IEEE TGRS, IEEE GRSL 和



Mohammed Benmamoun 是 UWA 计算机科学和软件工程系的温斯洛普教授，是计算机视觉、机器/深度学习、机器人和信号/语音处理的研究员。他出版了 4 本书(在亚马逊上可获得)，1 本编辑书籍，1 篇百科全书文章，14 本书章节，120+篇期刊论文，250+会议出版物，16 篇邀请和主题出版物。他的 h 指数为 50，被引次数为 11000 +(谷歌学者)。他是



胡索获得硕士学位。2018 年毕业于国防科技大学(NUDT)信息与通信工程专业。现任英国牛津大学计算机科学系博士研究生。他的研究兴趣在于 3D 计算机视觉、大规模点云处理和视觉跟踪。

获得了 65+竞争性研究资助，来自澳大利亚研究委员会，以及许多其他政府，UWA 和行业研究资助。成功指导 26+名博士研究生完成学业。他曾获得昆士兰科技大学年度最佳导师奖(1998 年)，并获得 UWA 研究指导奖(2008 年和 2016 年)和副校长导师奖(2016 年)。曾在 IEEE 计算机视觉与模式识别(CVPR 2016)、Interspeech 2014、IEEE 国际声学语音与信号处理会议(ICASSP)和欧洲计算机视觉会议(ECCV)等重要会议上发表会议指导。他还受邀在一个国际暑期学校(DeepLearn 2017)做了一个关于深度学习的教程。



刘康，2016 年毕业于中国电子科技大学工学学士，2018 年毕业于中国国防科技大学(NUDT)理学硕士。他目前在中山大学电子与通信工程学院攻读博士学位。他的研究兴趣在于三维计算机视觉和点云处理。

李刘于 2003 年、2005 年和 2012 年分别获得国防科技大学通信工程学士学位、摄影测量与遥感硕士学位和信息与通信工程博士学位。她于 2012 年加入 NUDT，现任系统工程学院副教授。博士期间，她在北京科技大学做了两年多的访问学生



滑铁卢，加拿大，2008 年至 2010 年。2015 年至 2016 年，她花了十个月时间参观香港中文大学多媒体实验室。2016.12 - 2018.11，在芬兰奥卢大学机器视觉组担任高级研究员。她是 CVPR、ICCV 和 ECCV 的九个国际研讨会的联合主席。她是 IEEE TPAMI 和 IJCV 特刊的客座编辑。她目前的研究兴趣包括计算机视觉、模式识别和机器学习。她的论文目前在谷歌 Scholar 上被引用超过 2300 次。她目前担任《视觉计算机杂志》和《模式识别快报》的副主编。她担任 ICME 2020 的区域主席。