

基于深度神经网络的无监督点云表示学习:综述

肖奥然、黄嘉兴、达扬·关、张晓琴, *IEEE 高级会员*, 陆世健, 邵玲, *IEEE Fellow*

抽象点云数据由于在各种不利情况下具有优越的精度和鲁棒性, 已经被广泛探索。与此同时, 深度神经网络(dnn)在监视和自动驾驶等各种应用中取得了令人印象深刻的成功。点云和深度神经网络的收敛导致了深度点云模型, 这些模型主要是在大规模和密集标记的点云数据的监督下训练的。无监督点云表示学习旨在从无标记的点云数据中学习通用和有用的点云表示, 由于大规模点云标记中的约束, 最近引起了越来越多的关注。本文对使用深度神经网络的无监督点云表示学习进行了全面的回顾。它首先描述了最近研究的动机、一般管道以及术语。然后简要介绍相关背景, 包括广泛采用的点云数据集和 DNN 架构。随后根据其技术方法对现有的无监督点云表示学习方法进行了广泛的讨论。我们还在多个广泛采用的点云数据集上定量基准测试和讨论了所审查的方法。最后, 我们分享了对无监督点云表示学习未来研究中可以追求的几个挑战和问题

索引术语-3d 视觉, 深度学习, 深度神经网络, 点云, 预训练, 自监督学习, 迁移学习, 无监督表示学习。

Manuscript received 11 June 2022; revised 20 March 2023; accepted 26 March 2023. 出版日期 2023 年 3 月 29 日; 当前版本 4 的发布日期 2023 年 8 月。这项工作部分由新加坡教育部资助, 在一级计划下, 项目在补助金 RG18/22 下。它还得到了 RIE2020 行业协调基金-行业合作项目(IAF-ICP)资助计划的部分支持, 以及新加坡电信有限公司(新加坡电信)通过新加坡电信企业认知和人工智能实验室(SCALE@NTU)提供的现金和实物捐助。由 M. Salzmann 推荐接受。(肖奥然、黄嘉兴为共同第一作者(通讯作者: 陆世健; 张晓琴。))

肖奥然、黄嘉兴、陆世健, 就职于南洋理工大学计算机科学与工程学院, 新加坡 639798 (e-mail: aoran.xiao@ntu.edu.sg; jiaxing.huang@ntu.edu.sg; shi-jian.lu@ntu.edu.sg)。

达扬·关就职于穆罕默德·本·扎耶德人工智能大学, 马斯达尔城, 阿联酋阿布扎比 7909(电子邮件: Dayan)。guan@mbzuai.ac.ae)。

张晓琴, 温州大学浙江省安全与应急智能信息重点实验室, 浙江温州 325035 (e-mail: zhangxiaojinnan@gmail.com)。

邵玲, 中国科学院大学 ucas 终端人工智能实验室, 北京 101408 (e-mail: ling.shao@ieee.org)

与此调查相关的一个项目已在 https://github.com/xiaoaoran/3d_url_survey 建立。

数字对象标识 10.1109/TPAMI.2023.3262786

0162-8828©2023 IEEE。允许个人使用, 但重新发布/再分发需要 IEEE 许可。更多信息请参见 <https://www.ieee.org/publications/rights/index.html>。

授权授权使用仅限于: 东北大学。下载于 2023 年 11 月 21 日 06:00:44 UTC 从 IEEE 探索。限制适用。

我的介绍。
3d 采集技术近年来发展迅速。这可以从在工业和我们日常生活中越来越流行的不同 3D 传感器中得到见证, 例如自动驾驶汽车中的激光雷达传感器, Kinect 和苹果设备中的 RGB-D 相机, 各种重建任务中的 3D 扫描仪等。与此同时, 网格、点云、深度图像和体积网格等不同模态的 3D 数据, 捕捉了物体和场景的精确几何信息, 已经被收集并广泛应用于自动驾驶、机器人、医疗、遥感等不同领域。

点云作为普遍存在和广泛使用的 3D 数据的一个来源, 可以直接用入门级深度传感器捕获, 然后再三角化为网格或转换为体素。这使得它很容易适用于各种 3D 场景理解任务[1], 如 3D 目标检测和形状分析、语义分割等。随着深度神经网络(DNNs)的发展, 近年来开发的大量深度架构和深度模型[2]观察到, 点云理解吸引了越来越多的关注。另一方面, 深度网络的有效训练需要大规模的人工标注训练数据, 如用于物体检测的 3D 包围盒和用于语义分割的点标注, 由于 3D 视图变化和人类感知与点云显示之间的视觉不一致, 这些数据通常是费力且耗时的收集。在处理各种真实世界任务[3]的同时, 高效收集大规模标注点云已成为深度网络有效设计、评估和部署的瓶颈之一。

无监督表示学习(Unsupervised representation learning, URL)旨在从未标记的数据中学习鲁棒和通用的特征表示, 最近得到了广泛的研究, 以减轻费力和耗时的数据标注挑战。如图 1 所示, URL 的工作方式与预训练类似, 从未标记的数据中学习有用的知识, 并将学习到的知识转移到各种下游任务中[4]。更具体地说, URL 可以提供有用的网络初始化, 与随机初始化训练相比, 使用少量标记和特定于任务的训练数据可以训练性能良好的网络模型, 而不会出现过度拟合的问题。因此, URL 可以帮助减少训练数据和注释, 这在

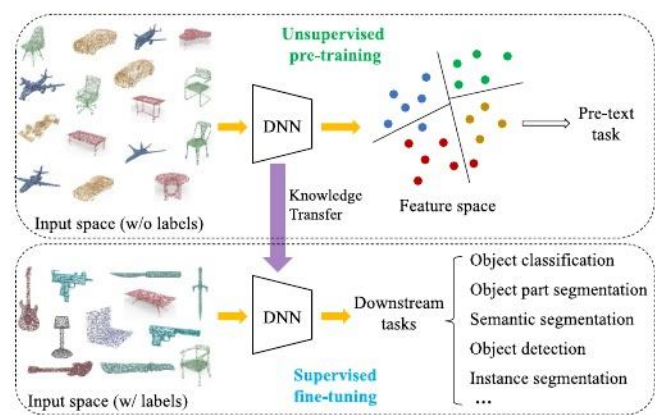


图 1 所示。点云上无监督表示学习的一般管道:首先通过在某些前置文本任务上进行无监督学习,用无标注的点云对深度神经网络进行预训练。然后将学习到的无监督点云表示转移到各种下游任务中,以提供网络初始化,用少量标注的特定任务点云数据对预训练的网络进行微调。

自然语言处理(NLP)[5]、[6]、二维计算机视觉[7]、[8]、[9]、[10]等领域。

类似于 URL 从其他类型的数据,如文本和二维图像、点云的 URL 近年来越来越受到计算机视觉研究界的关注。已经报道了许多 URL 技术,这些技术通常是通过设计不同的前置文本任务来实现的,例如

3D 物体重建[11],部分物体补全[12],3D 拼图解决[13]等。然而,点云的 URL 在 NLP 和二维计算机视觉任务中仍然远远落后。就目前而言,在各种目标新数据上从头开始训练仍然是大多数现有 3D 场景理解开发中的流行方法。另一方面,由于缺乏大规模、高质量的点云数据、统一的深度骨干架构、通用的技术途径以及全面的公共基准,点云数据 URL 面临越来越多的问题和挑战。

此外,点云的 URL 仍然缺乏系统的调查,这可以为这个新的但具有挑战性的任务提供一个清晰的大局。为填补这一空白,本文从数据集、网络架构、技术途径、性能基准和未来研究方向等角度,对无监督点云表示学习的近期进展进行了全面综述。如图 2 所示,我们将现有方法根据其借口任务大致分为四类,包括使用数据生成的 URL 方法、全局和本地上下文、多模态数据和本地描述符,更多细节将在随后的小节中讨论。

这项工作的主要贡献有三个方面:

- 1)对无监督点云表示学习的最新进展进行了全面综述。据我们所知,这是第一个为这个令人兴奋的研究主题提供概述和大图的综述。

- 2)研究了无监督点云表示学习的最新进展,包括在多个公开数据集上对现有方法进行全面的基准测试和讨论。

- 3)它分享了几个在无监督点云表示学习中可以追求的研究挑战和潜在的研究方向。

本文介绍了无监督点云学习的背景知识,包括术语定义、点云理解的常见任务和这项工作的相关调查。第三节介绍了广泛使用的数据集及其特点。第四节介绍了常用的深度点云架构和点云 URL 常用的典型模型。在第五节中,我们系统地回顾了点云 URL 的方法。第 VI 节总结并比较了现有方法在多个基准数据集上的性能。最后,我们在第 VII 节中列出了无监督点云表示学习的几个有希望的未来方向。

2. 背景

A. 基本概念

我们首先定义所有相关的术语和概念,将在随后的章节中使用。

点云数据:点云 P 是向量 $P = \{p_1, \dots, p_N\}$ 的集合,其中每个向量代表一个点 $p_i = [C_i, A_i]$ 。其中, $C_i \in \mathbb{R}^{1 \times 3}$ 表示点的三维坐标 (x_i, y_i, z_i) , A_i 表示点的特征属性,如 RGB 值、LiDAR 强度、法线值等,根据 3D 传感器和应用的不同,这些属性是可选的,是可变的。

监督学习(Supervised learning):在深度学习的范式下,监督学习旨在通过使用标记训练数据来训练深度网络模型。

无监督学习:无监督学习旨在通过使用未标记的训练数据来训练网络。

无监督表示学习:URL 是无监督学习的一个子集。它旨在不使用任何数据标签/注释从数据中学习有意义的表示,其中学习到的表示可以转移到不同的下游任务。一些文献交替使用术语“自监督学习”(self-supervised learning)。

半监督学习:在半监督学习中,深度网络是用少量标记数据和大量未标记数据进行训练的。它旨在通过从少量标记数据和大量分布相似的未标记数据中学习来缓解数据注释约束。

预训练:网络预训练在其他数据集上通过特定的前置任务进行学习。学习到的参数通常用于模型初始化,以便对各种特定任务的数据进行进一步微调。

迁移学习:迁移学习旨在跨任务、模态或数据集迁移知识。与本调查相关的一个典型场景是为进行无监督学习

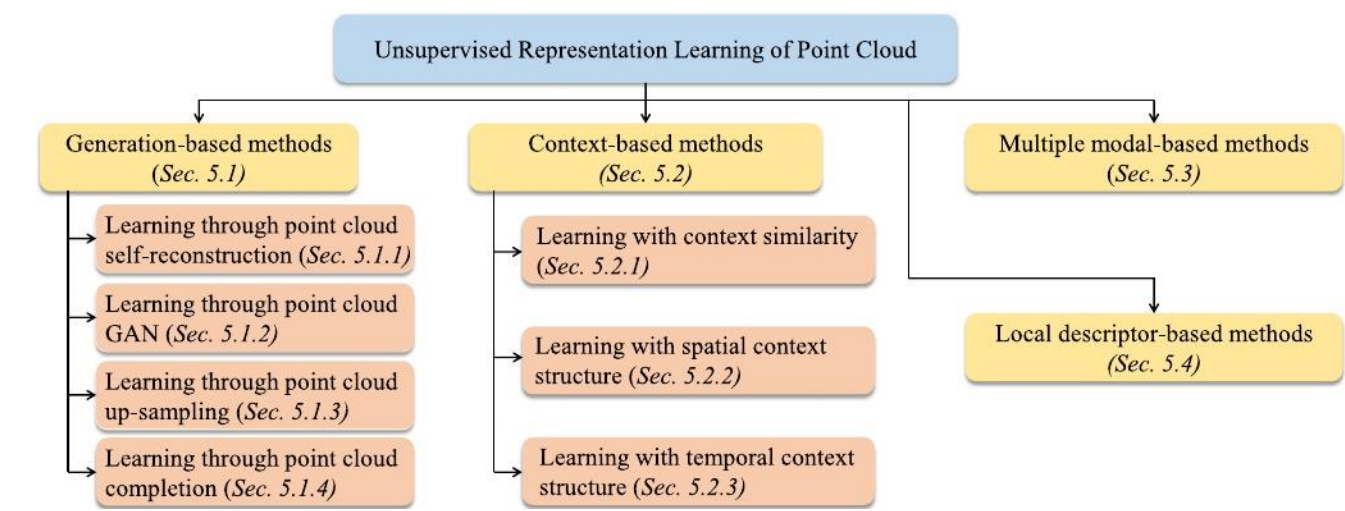


图 2 所示。点云表示学习中现有无监督方法的分类。

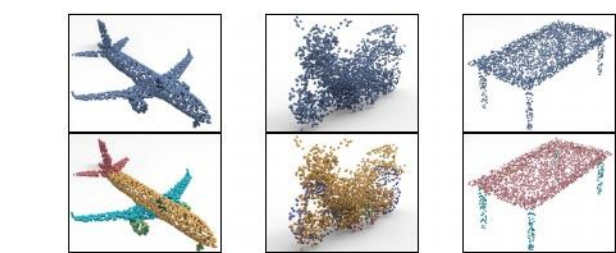


图 3 所示。对象部分分割插图:第一行显示了一些对象样本,包括飞机、摩托车和 ShapeNetPart 数据集[14]中的表格。第二行显示了不同部分的分割 ground truth,用不同的颜色突出显示。

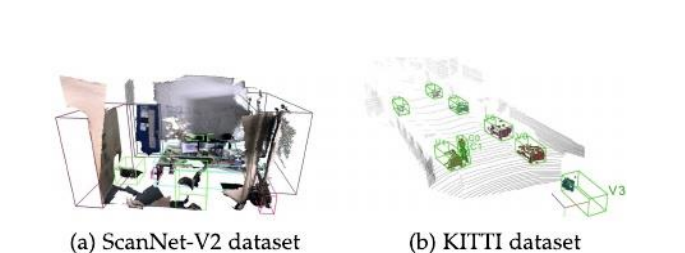


图 4 所示。点云目标检测中的三维边界框示意图:这两张图分别显示了数据集 ScanNet-V2[18]和 KITTI[19]中的三维边界框,它们分别是[16]和[20]中裁剪出来的。

预训练, 将学习到的知识从未标记的数据转移到各种下游网络。

B. 常见的 3D 理解任务

本小节介绍常见的 3D 理解任务, 包括对象分类和对象部分分割中的对象级任务和 3D 目标检测、语义分割和实例分割中的场景级任务。这些任务已被广泛用于评估通过各种无监督学习方法学习的点云表示的质量, 这将在第 VI 节中详细讨论。

1)对象分类:对象分类旨在将点云对象分类为许多预定义的类别。两个最常用的评价指标:总体精度(OA)代表测试集中所有实例的平均精度; 平均类精度(m Acc)代表测试集所有对象类的平均精度。

2)对象部分分割:对象部分分割是点云表示学习的重要任务。它旨在为图 3 所示的每个点分配一个部件类别标签(例如飞机机翼、桌腿等)。mIoU (Intersection over Union)均值[15]是采用最广泛的评价指标。对于每个实例, 为属于该对象类别的每个部件计算借据。部分借据的平均值表示

该对象实例的借据。总体借据计算为所有测试实例的借据的平均值, 而分类借据(或类借据)计算为该类别下实例的平均值。

3) 3D 目标检测:点云上的 3D 目标检测对于许多现实世界的应用来说是至关重要和不可或缺的任务, 例如自动驾驶和家用机器人。该任务旨在定位 3D 空间中的物体, 如图 4 所示的 3D 物体边界框。平均精度(AP)指标在 3D 目标检测[16], [17]中被广泛用于评估。

4)三维语义分割:点云上的三维语义分割是 3D 理解的另一项关键任务, 如图 5 所示。与分割点云对象的对象部分分割不同, 3D 语义分割旨在为复杂度高得多的场景级点云中的每个点分配一个类别标签。广泛采用的评估指标包括 OA、语义类别 mIoU 和 mAcc。

5) 3D 实例分割:3D 实例分割旨在检测和描绘场景级点云中每个不同的感兴趣对象, 如图 6 所示。在只考虑语义类别的语义分割之上, 实例分割为每个对象分配一个唯一的标识。平均精度均值(m AP)已被广泛采用用于这一任务的定量评估。

表我

之前用点云进行ud 研究中用于训练和评估的常用数据集总结

Dataset	Year	#Samples	#Classes	Type	Representation	Label
KITTI [19]	2013	15K frames	8	Outdoor driving	RGB & LiDAR	Bounding box
ModelNet10 [27]	2015	4,899 objects	10	Synthetic object	Mesh	Object category label
ModelNet40 [27]	2015	12,311 objects	40	Synthetic object	Mesh	Object category label
ShapeNet [14]	2015	51,190 objects	55	Synthetic object	Mesh	Object/part category label
SUN RGB-D [28]	2015	5K frames	37	Indoor scene	RGB-D	Bounding box
S3DIS [21]	2016	272 scans	13	Indoor scene	RGB-D	Point category label
ScanNet [18]	2017	1,513 scans	20	Indoor scene	RGB-D & mesh	Point category label & Bounding box
ScanObjectNN [29]	2019	2,902 objects	15	Real-world object	Points	Object category label
ONCE [30]	2021	1M scenes	5	Outdoor driving	RGB & LiDAR	Bounding box

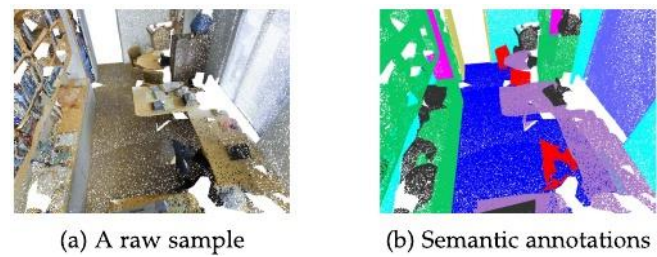


图 5 所示。语义点云分割示意图:对于左边 S3DIS[21]的点云样本,右边的图显示了对应的 ground truth,不同类别用不同颜色突出显示。

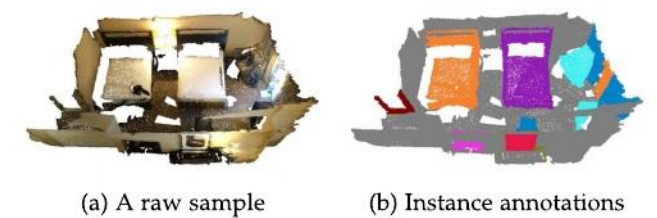


图 6 所示。点云实例分割示意图:左图为 ScanNet-V2[18]的点云样本,右图为对应的 ground truth,不同的实例用不同的颜色突出显示。

C. 相关调查

据我们所知,这篇文章是第一篇全面回顾无监督点云学习的综述。已经进行了几个相关但不同的调查。例如,几篇论文回顾了点云上深度监督学习的最新进展:Ioannidou 等人。[22]回顾了 3D 数据上的深度学习方法;Xie et al.[23]提供了点云分割任务的文献综述;Guo et al.[2]针对分类、检测、跟踪和分割等多个任务,对点云的深度学习进行了全面而详细的综述。此外,几项工作回顾了其他数据模态上的无监督表示学习:Jing et al.[24]介绍了二维计算机视觉中无监督表示学习的进展;Liu et al.[25]介绍了无监督表示学习方法在 2D 计算机视觉、NLP 和图学习方面的最新进展;Qi et al.[26]介绍了包括无监督和半监督方法在内的小数据学习的最新进展。

3. 点云数据集

在本节中,我们总结用于训练和评估无监督点云表示学习的常用数据集。如表 1 所示,现有工作主要从 1)合成对象数据集(如 ModelNet[27]和 ShapeNet[14])或 2)真实场景数据集(如 ScanNet[18]和 KITTI[19])中学习无监督点云表示。此外,还收集了各种特定于任务的数据集,这些数据集可用于对下游模型进行微调,例如用于点云分类的 ScanObjectNN[29]、ModelNet40[27]和 ShapeNet[14],用于零件分割的 ShapeNetPart[14],用于语义分割的 S3DIS[21]、ScanNet[18]或 Synthia4D[31],用于对象检测的室内数据集 SUNRGB-D[28]和 ScanNet[18]以及用于对象检测的室外数据集 ONCE[30]。

•ModelNet10/ModelNet40 [27]: ModelNet 是用于 3D 分类的合成对象级数据集。最初的 ModelNet 提供了由顶点和面表示的 CAD 模型。点云是通过从模型中均匀采样生成的。ModelNet40 包含 40 个类别的 13834 个对象,其中 9843 个对象构成训练集,其余为测试集。ModelNet10 由 10 类 3377 个样本组成,分为 2468 个训练样本和 909 个测试样本。

•ShapeNet [14]: ShapeNet 包含 55 个类别的合成 3D 对象。它是通过从在线开源 3D 存储库收集 CAD 模型来策划的。与 ModelNet 类似,ShapeNet 中的合成对象是完整的,对齐的,没有遮挡或背景。它的扩展 ShapeNetPart 包含 16 个类别的 16,881 个对象,并由点云表示。每个对象由 2 到 6 个部分组成,在数据集中总共有 50 个部分类别。

•ScanObjectNN [29]: ScanObjectNN 是一个真实的对象级数据集,其中从真实室内场景中捕获的扫描中构建了 15 个类别的 2902 个 3D 点云对象。与合成对象数据集不同,ScanObjectNN 中的点云对象是有噪声的(包括背景点、遮挡和对象中的孔),并且不是轴对齐的。

•S3DIS[21]:斯坦福大尺度 3D 室内空间(Stanford Large-Scale 3D Indoor Spaces, S3DIS)数据集包含 3 栋办公楼的 6 个大尺度室内区域的超过 2.15 亿个点,每个区域的面积为 6000 平方米。这些扫描被表示为带有 13 个对象类别逐点语义标签的点云。

•ScanNet-V2 [18]: ScanNet-V2 是一个 RGB-D 视频数据集,包含超过 1500 次扫描的 250 万次观看,其中

都是在办公室和客厅等室内场景中拍摄的，并标注了 3D 相机姿态、表面重建以及用于分割的语义和实例标签。

•**SUN RGB-D [28]:** SUN RGB-D 数据集是从室内环境中收集的单视图 RGB-D 图像的集合。总共有 10335 张 RGB-D 图像标注了模态和面向 3D 的 37 类对象边界框。

•**KITTI [19]:** KITTI 是一个先锋户外数据集，提供来自激光雷达传感器的密集点云，以及其他模式，包括正面立体图像和 GPS/IMU 数据。它提供了超过 22 个场景的 200k 3D 盒子，用于 3D 物体检测。

•**ONCE [30]:** ONCE 数据集有 100 万个激光雷达场景和 700 万张相应的相机图像。总共有 581 个序列，其中 560 个序列是未标记的并用于无监督学习，10 个序列是标注的并用于测试。它为户外环境中的目标检测提供了一个无监督学习基准。

公开的点云 URL 数据集在数据大小和场景种类方面仍然有限，特别是与用于二维计算机视觉和 NLP 研究的图像和文本数据集相比。例如，BooksCorpus 有 8 亿字，英文维基百科有 25 亿字，能够为 NLP 中的无监督表示学习提供全面的数据源[32];ImageNet[33]拥有超过 1000 万张图像用于无监督视觉表示学习。大规模和高质量的点云数据对于未来关于这一主题的研究是非常需要的，我们在第七节对此问题进行了详细的讨论。

Iv.常见的 DEEP 架构

在过去的十年中，深度学习在点云处理和理解中发挥了越来越重要的作用。这一点可以从近年来开发的大量深度架构中观察到。与将点云转换为八叉树[34]或哈希体素列表[35]等结构的传统 3D 视觉不同，深度学习倾向于更具可微性的结构和/或高效的神经处理，这些结构已经在各种 3D 任务中取得了非常令人印象深刻的性能。

另一方面，基于 dnn 的点云处理和理解远远落后于 NLP 和 2D 计算机视觉的对应领域。这对于无监督表示学习的任务尤其如此，很大程度上是由于点云数据中缺乏规则的代表。具体来说，词嵌入和 2D 图像具有规则且定义明确的结构，但由无序点集表示的点云却没有这样通用且结构化的数据格式。

在本节中，我们将介绍为点云的 URL 探索的深度架构。点云的深度学习在过去十年中取得了重大进展，我们看到大量的 3D 深度架构和 3D 模型被提出。然而，在 2D 计算机视觉中，我们没有像 VGG[36]或 ResNet[37]那样通用和无处不在的“3D 骨干”。因此，在本次调查中，我们将重点关注点云 URL 中经常使用的架构。

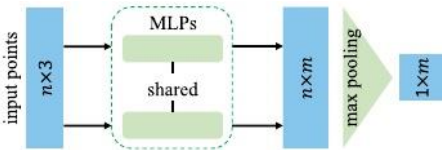


图 7 所示。用于点云对象分类的简化的 PointNet 架构[15]，其中参数 n 和 m 分别表示点个数和特征维度。

为了描述的清晰性，我们将它们大致分为五类，即基于点的架构、基于图的架构、基于稀疏体素的架构、基于空间 cnn 的架构和基于 transformer 的架构。注意[2]中讨论的各种 3D 任务也存在其他深度架构，如基于投影的网络[38]，[39]，[40]，[41]，[42]，[43]，循环神经网络[44]，[45]，[46]，3D 胶囊网络[47]等。然而，他们并不经常被用于 URL 任务，因此在本调查中没有详细说明。

A.基于点的深度架构

基于点的网络被设计用来直接处理原始点云，而无需事先进行点数据转换。通常首先通过多层感知器(MLPs)堆叠网络来提取独立的点特征，然后使用对称聚合函数将其聚合为全局特征。

PointNet[15]是一种先锋的基于点的网络，如图 7 所示。它堆叠几个 MLP 层来独立学习逐点特征，并将学习到的特征转发到最大池化层来提取全局特征以实现排列不变性。为了改进 PointNet，Qi 等人提出了 PointNet++[48]从点的邻域学习局部几何细节，其中集合的抽象层包括采样层、分组层和 PointNet 层，用于学习局部和分层特征。PointNet++在包括对象分类和语义分割在内的多个 3D 任务中取得了巨大的成功。Qi 等人以 PointNet++为骨干，设计了首个基于点的三维目标检测网络 VoteNet[16]。VoteNet 采用 Hough 投票策略，在物体中心周围生成新的点，并将其与周围的点组合在一起，生成 3D 盒子提案。

B.基于图的深度架构

基于图的网络将点云视为欧氏空间中的图，顶点是点，边捕获相邻点关系，如图 8 所示。它与图卷积一起工作，其中滤波器权重以边标签为条件，并为单个输入样本动态生成。这允许通过强制权重共享和提取可以捕获邻近点之间依赖关系的局部特征来减少学习模型中的自由度。

动态图卷积神经网络(Dynamic Graph Convolutional Neural Network, DGCNN)[49]是一种典型的基于图的架构，经常用于点云的 URL。它与一个名为 EdgeConv 的图卷积模块堆叠在一起，该模块在特征空间中动态地对图进行卷积。DGCNN 将 EdgeConv 集成到 PointNet 结构的基本版本中

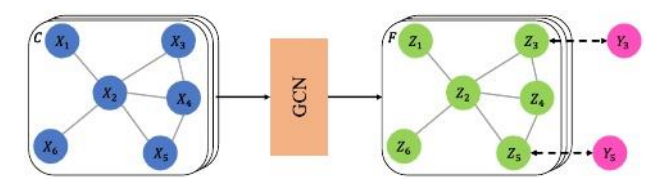


图 8 所示。图卷积网络(GCN)示意图:每个图由代表点 X_i or 特征 Z_i 的多个顶点组成(用圆点突出显示), 以及连接代表点关系的顶点的边(如图黑线所示)。C表示输入通道, F表示输出特征维度, y_i 表示标签。

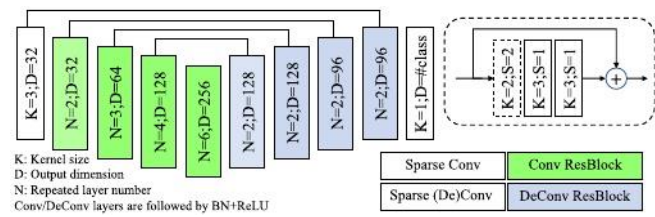


图 9 所示。采用统一 U-Net[55]架构进行稀疏卷积的 SR-UNet[54]示意图。图基于[54]进行复现。

用于学习点云理解的全局形状属性和语义特征。

C. 基于稀疏体素的深度架构

基于体素的架构 v 在将 3D CNN 应用于体表示之前将点云oxel化为 3D 网格。由于点云数据的稀疏性, 在处理大量的同时往往涉及巨大的计算冗余或牺牲表示精度。为了克服这种约束, [50], [51], [52], [53]采用稀疏张量作为基本单位, 点云用数据列表和索引列表表示。与使用滑动窗口(PyTorch 和 TensorFlow 中的 *im2col* 函数)来构建计算管道的标准卷积操作不同, 稀疏卷积[50]收集包括卷积核元素在内的所有原子操作, 并将它们作为计算指令保存在规则手册中。

最近, Choy 等人提出了 Minkowski 引擎[51], 引入了广义稀疏卷积和稀疏张量的自动微分库。在此基础上, Xie 等[54]采用了统一的 U-Net[55]架构, 构建了一个骨干网(SR-UNet, 如图 9 所示)进行无监督预训练。学习到的编码器可以被迁移到不同的下游任务中, 如分类、目标检测和语义分割。

D. 基于空间cnn 的深度架构

基于空间 cnn 的网络已经被开发出来, 以扩展规则网格cnn 分析不规则间隔点云的能力。根据卷积核[2], 它们可以分为连续卷积网络和离散卷积网络。如图 10 所示, 连续卷积网络在连续空间中定义卷积核,

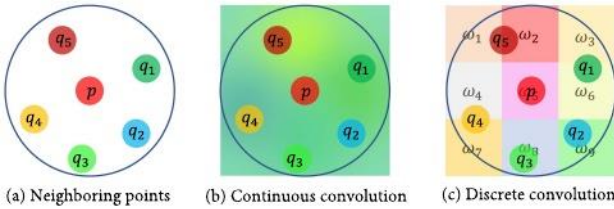


图 10 所示。3D 空间卷积的图解, 包括连续卷积和离散卷积。参数 p 和 q 分别表示中心点及其邻近点。图是基于[2]复制的。

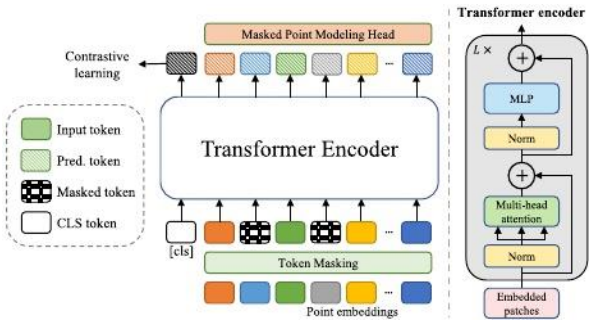


图 11 所示。point - bert[57]中用于无监督预训练的点云 Transformer 的架构。更多的网络细节可以在[57]中找到。图根据[57]、[58]进行转载。

其中邻近点的权重由它们相对于中心点的空间分布决定。不同的是, 离散卷积网络在规则网格上运行, 并在邻近点相对于中心点具有固定偏移的离散空间中定义卷积核。连续卷积模型的一个典型例子是 RS-CNN[56], 它已被广泛用于点云的 URL。具体而言, RS-CNN 提取局部中心与其周围点之间的几何拓扑关系, 并学习卷积的动态权值。

E. 基于 transformer 的深度架构

在过去的几年里, transformer 由于其结构优势和通用性, 在 NLP[32]、[59]和 2D 图像处理[58]、[60]的研究领域取得了惊人的进展。它们也被引入点云处理领域[57], [61]。图 11 为点云 URL 的标准 Transformer 架构[57], 其中包含一堆 Transformer 块[59], 每个块由多头自关注层和前馈网络组成。无监督的预训练 Transformer 编码器可用于微调下游任务, 如对象分类和语义分割等。

V. 无监督点云表示学习

在本节中, 我们将回顾现有的点云 URL 方法。如图 2 所示, 我们根据现有的前置任务将现有方法大致分为四类, 包括基于生成的方法、基于上下文的方法、多重方法

表二世

基于生成的 POINT 云 UNSUPERVISED 表示学习方法综述

Method	Published in	Category	Contribution
VConv-DAE [62]	ECCV 2016	Completion	Learning by predicting missing parts in 3D grids
TL-Net [63]	ECCV 2016	Reconstruction	Learning by 3D generation and 2D prediction
3D-GAN [64]	NeurIPS 2016	GAN	Pioneer GAN for 3D voxels
3D-DescriptorNet [65]	CVPR 2018	Completion	learning with energy-based models for point cloud completion
FoldingNet [66]	CVPR 2018	Reconstruction	learning by folding 3D object surfaces
SO-Net [67]	CVPR 2018	Reconstruction	Performing hierarchical feature extraction on individual points and SOM nodes
Latent-GAN [68]	ICML 2018	GAN	Pioneer GAN for raw point clouds and latent embeddings
MRT [69]	ECCV 2018	Reconstruction	A new point cloud autoencoder with multi-grid architecture
VIP-GAN [70]	AAAI 2019	GAN	Learning by solving multi-views inter-prediction tasks for objects
G-GAN [11]	ICLR 2019	GAN	Pioneer GAN with graph convolution for point clouds
3DCapsuleNet [47]	CVPR 2019	Reconstruction	Learning with 3D point-capsule network
L2G-AE [71]	ACM MM 2019	Reconstruction	Learning by global and local reconstruction of point clouds
MAP-VAE [72]	ICCV 2019	Reconstruction	Learning by 3D reconstruction and half-to-half prediction
PointFlow [73]	ICCV 2019	Reconstruction	Learning by modeling point clouds as a distribution of distributions
PDL [74]	CVPR 2020	reconstruction	A probabilistic framework for point distribution learning
GraphTER [75]	CVPR 2020	Reconstruction	Proposed a graph-based autoencoder for point clouds
SA-Net [76]	CVPR 2020	Completion	Learning by completing point cloud objects with a skip-attention mechanism
PointGrow [77]	WACV 2020	Reconstruction	An autoregressive model that can recurrently generate point cloud samples.
PSG-Net [78]	ICCV 2021	Reconstruction	Learning by reconstruct point cloud objects with seed generation
OcCo [12]	ICCV 2021	Completion	Learning by completing occluded point cloud objects
Point-Bert [57]	CVPR 2022	Reconstruction	Learning for Transformers by recovering masked tokens of 3D objects
Point-MAE [79]	ECCV 2022	Reconstruction	Autoencoder transformer recovers masked parts from input data
Point-M2AE [80]	NeurIPS 2022	Reconstruction	Masked autoencoder with hierarchical point cloud encoding and reconstruction.

modal-based methods and local descriptor-based methods。通过这种分类法，我们对现有的方法进行了梳理，并在本节接下来的小节中系统地介绍了它们。

A. 基于生成的方法

基于生成的点云 URL 方法涉及在训练中生成点云对象的过程。根据采用的文本前任务，它们可以进一步分为四个子类别，包括点云自重建(用于生成与输入相同的点云对象)，点云 GAN(用于生成假点云对象)，点云上采样(用于生成具有更密集的点云但形状相似的对象)和点云补全(用于从不完整的点云对象中预测缺失部分)。这些 URL 方法的基本原理是点云本身。因此，这些方法不需要人工标注，可以以无监督的方式进行学习。表 II 列出了基于生成的方法。

1)通过点云自重建进行学习:用于自重建的网络通常将点云样本编码为表示向量，并将其解码回原始输入数据，在此过程中提取形状信息和语义结构。它属于一种典型的 URL 方法，因为它不涉及任何人工注释。一个代表性的网络是 autoencoder[81]，它有一个编码器网络和一个解码器网络，如图 12 所示。编码器将点云对象压缩并编码为低维嵌入向量(即码字[66])，然后由解码器解码回 3D 空间。

通过强制最终输出与输入相同来优化模型。在这个过程中，编码通过尝试从编码中重新生成输入进行验证和学习，而自编码器通过训练网络忽略无关紧要的数据(“噪声”)来学习低维表示[82]。排列不变损失[83]广泛存在

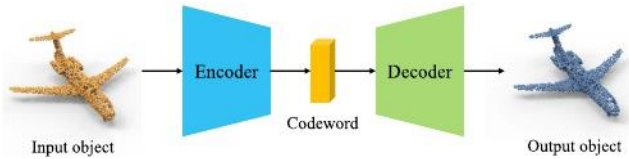


图 12 所示。自动编码器在无监督点云表示学习中的一个例子:编码器学习通过码字向量来表示点云对象，而解码器从码字重构输出对象。

采用作为训练目标来描述输入和输出的点云对象如何彼此相似。它们可以通过 Chamfer Distance L_{CD} Earth Mover’s Distance L_{EMD} 来测量 EMD：

$$L_{CD} = \sum_{p \in P} \min_{p' \in P'} ||p - p'||^2 + \sum_{p' \in P'} \min_{p \in P} ||p' - p||^2 \tag{1}$$

$$L_{EMD} = \min_{\phi: P \rightarrow P'} \sum_{x \in P} ||p - \phi(p)||_2 \tag{2}$$

何处 P 和 P' 表示输入和输出大小相同的点云， $\phi: P \rightarrow P'$ 是双射，而 p & p' 是点。

在过去的十年中，自重构已经成为最广泛采用的点云 URL 预文本任务之一。通过假设点云表示应该在 3D 空间中生成并从 2D 空间中可预测，Girdhar 等人提出了 TL-Net[63]，该网络采用 3D 自编码器来重建 3D 体积网格和 2D 卷积网络来从投影图像中学习 2D 特征。Yang 等人设计了 FoldingNet[66]，它引入了一个基于折叠的解码器，将规范的 2D 网格变形到点云对象的底层 3D 对象表面。Li 等人提出了 SO-Net[67]，引入了自组织映射，通过自重建来学习点云的分层特征。Zhao 等人[47]将胶囊网络[84]扩展到 3D 点云处理 and

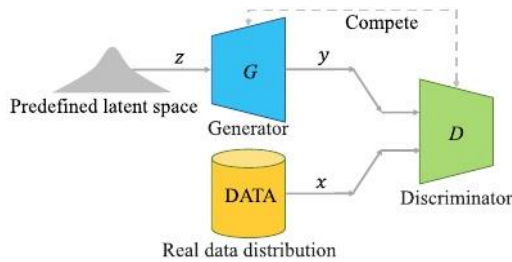


图 13 所示。GAN 的示意图，通常由生成器 G 和鉴别器 D 组成，它们在训练过程中相互争斗(以零和游戏的形式，其中一个代理的收益是另一个代理的损失)。

设计的 3D 胶囊网络可以从非结构化 3D 数据中学习通用表示。Gao 等人[75]提出了一种基于图的自动编码器，可以在全局和局部变换下学习点云结构的内在模式。Chen 等人[85]设计了一个 deepautoencoder，利用图拓扑推理和过滤从 3D 点云中提取紧凑的表示。

一些研究探索了全局和局部几何来从点云对象中学习鲁棒的表示[71], [72]。例如，[71]在编码器中引入了层次自关注，用于信息聚合，并将递归神经网络(RNN)作为解码器，用于局部和全局点云重建。[72]提出了 MAP-VAE，它引入了一种半对半的预测任务，首先将点云对象分成前半部分和后半部分，然后训练 RNN 从相应的前半部分序列预测后半部分序列。一些研究反而将点云重建表述为点分布学习任务[73]、[74]、[77]。例如，[73]提出了 PointFlow，它通过对形状的分布和给定形状的点的分布进行建模来生成 3D 点云。[74]提出了一个概率框架，通过点分布学习来提取无监督形状描述符，它将每个点与高斯相关联，并将点云模型作为点的分布。[77]提出了一种自回归模型 Pointgrow，它可以从零开始或以语义上下文为条件生成多样化和逼真的点云样本。

此外，一些研究从不同的对象分辨率[69]、[78]、[86]中学习点云表示。例如，Gadelha 等人[69]设计了一个具有多分辨率树结构的自动编码器，通过从粗到细的分析来学习点云表示。Yang 等人[78]提出了一种带有种子生成模块的自动编码器，允许在多个阶段提取依赖输入的逐点特征，并逐步提高分辨率。Chen 等人[86]提出通过重建不同分辨率的点云对象并最小化它们之间的倒角距离来学习采样不变特征。

2) 通过点云学习 GAN: 生成与对抗网络(Generative and Adversarial Network, GAN)[87]是一种典型的深度生成网络。如图 13 所示，它由一个生成器和一个判别器组成。生成器旨在合成尽可能逼真的数据样本，而鉴别器则试图区分真实样本和合成样本。因此 GAN 学会

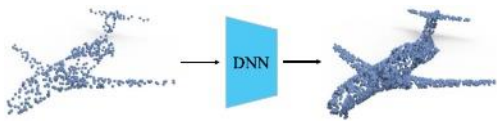


图 14 所示。点云上采样的示意图:网络 DNN 通过解决一个前文本任务来学习点云表示，该任务再现了具有相同几何形状但点分布更密集的对象。

生成具有与训练集相同统计量的新数据，建模可以表述为 min-max 问题:

$$\min_G \max_D L_{GAN} = \log D(x) + \log(1 - D(G(z))), \quad (3)$$

其中 G 是生成器， D 代表鉴别器。 x 和 z 分别表示来自分布 $p(z)$ 的真实样本和随机采样的噪声向量。

当对点云的 URL 进行 gan 训练时，生成器从采样向量或潜在嵌入中学习生成点云实例，而鉴别器则试图区分输入点云是来自真实数据分布还是来自生成的数据分布。两个子网络在训练过程中相互争斗，鉴别器学习提取有用的特征表示用于点云对象识别。学习过程不涉及人工标注，因此网络可以以无监督学习的方式进行训练。在此之后，通过微调模型，学习到的鉴别器被扩展到各种下游任务中，如对象分类或部分分割。

一些网络成功地将 GAN 用于点云的 URL[11], [64], [68], [88]。例如，Wu 等人[64]提出了第一个应用于 3D 体素的 GAN 模型。然而，体素化过程要么牺牲了表示精度，要么招致了巨大的冗余。Achlioptas 等人提出了 Latent-GAN[68]作为原始点云的第一个 GAN 模型。Li 等人[88]进一步提出了一种具有分层采样和推理网络的点云 GAN 模型，该模型学习随机过程以生成新的点云对象。Valsesia 等人[11]设计了第一个基于图的 GAN 模型，从点云中提取局部特征。这些方法通过微调学习到的表示到高级下游 3D 任务来评估学习到的表示的泛化性。

3) 通过点云学习学习:如图 14 所示，给定一组点，点云上采样的目的是生成具有相似几何形状的更密集的点集。该任务需要深度点云网络在没有任何监督的情况下学习 3D 形状的底层几何形状，并且学习到的表示可以用于 3D 下游任务中的微调。

Lieta1. [89]将 ganto 引入 pointcloudup-sampling 任务，并提出了 PU-GAN，通过在物体表面补丁上的上采样点来从潜空间学习各种点分布。生成器旨在产生上采样点云，而鉴别器试图区分其输入的点云是由生成器产生的还是真实的点云。与 V-A2 节中介绍的 GANs 类似，学习到的鉴别器可以在下游任务中迁移。Remelli

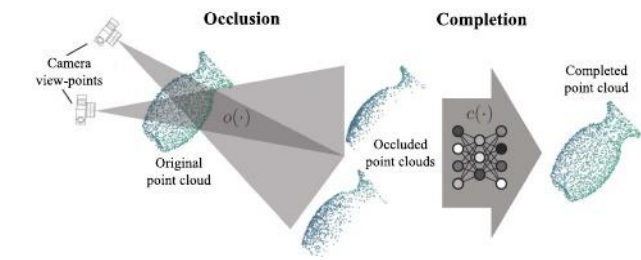


图 15 所示。OcCo[12]的管道。将被遮挡的点云对象作为输入，训练一个编码器-解码器模型来完成被遮挡的点云，其中编码器学习点云表示，解码器学习生成完整的对象。学习到的编码器权重可用于下游任务的网络初始化。该图来自[12]，已获作者许可。

等人[90]设计了一个自动编码器，可以将稀疏点云上采样成稠密表示。编码器的学习权重也可以用作第 V-A1 节所述的下游任务的初始化权重。虽然点云上采样近年来受到越来越多的关注[89]、[91]、[92]、[93]、[94]、[95]，但它在很大程度上是通过生成点云的质量来评估的，而其在迁移学习中的性能还没有得到很好的研究。

4)通过点云补全进行学习:点云补全是一项基于其余 3D 点云预测任意缺失部分的任务。为了实现这一目标，深度网络需要学习 3D 对象的内部几何结构和语义知识，从而正确预测缺失的部件。最重要的是，学习到的表示可以转移到下游任务。整个过程不涉及人工标注，因此属于无监督表示学习。

在过去十年中，点云补全一直是一个活跃的研究领域[76]，[112]，[113]，[114]，[115]，[116]，[117]，并在不同的 URL 基准中进行了评估[12]，[62]，[65]，[76]。一项先驱工作 VConv-DAE[62]体素化点云对象到体网格中，并通过从其余部分预测缺失的体素，用自动编码器学习对象形状分布。Xie 等人[65]设计了 3D-DescriptorNet 用于体形状模式的概率建模。Achlioptas 等人[68]介绍了第一个用于原始点云补全的 DNN，它是具有编码器-解码器结构的基于点的网络。Yuan 等人[113]提出了一种点补全网络，这是一种自动编码器结构网络，通过修复不完整的点云对象来学习有用的表示。Wen 等人[76]提出了 SA-Net，它在编码器中引入了一种跳过注意力机制，选择性地几何信息从局部区域传输到解码器，以生成完整的点云对象。Wang 等人[12]提出学习一个编码器-解码器模型，该模型可以通过不同的摄像机视角恢复被遮挡的点，如图 15 所示。编码器参数被用作包括分类、部分分割和语义分割在内的下游任务的初始化。

最近，在 NLP[5]，[6]和 2D 计算机视觉[10]中，从被屏蔽的输入中恢复缺失部分作为 URL 的 pre-text 任务已经被证明是非常成功的。这样的想法也在 3D 点云学习[57]中进行了研究，[79]，

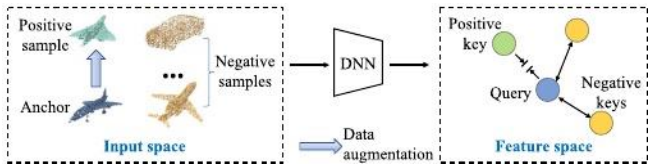


图 16 所示。实例对比学习的例子，通过自我辨别学习局部平滑表示，它将查询(来自锚样本)拉到正键(来自正样本)附近，并将其推离负键(来自负样本)。

[110]、[118]。例如，Yu 等人[57]提出了一种点伯特范式，通过掩蔽点建模任务预训练点云变压器。他们使用离散变分自编码器为物体补丁生成 token，并随机屏蔽 token 以训练 Transformer 恢复原始的完整点 token。点伯特学习到的表征可以很好地转移到新的任务和领域，如对象分类和对象部分分割。

B.基于上下文的方法

基于上下文的方法是另一类重要的点云 URL，近年来受到越来越多的关注。与以生成方式学习表示的基于生成的方法不同，这些方法采用判别式的前文本任务来学习点云的不同上下文，包括上下文相似性、空间上下文结构和时间上下文结构。所设计的前文本任务不需要人工标注，表 III 列出了最近的方法。

1)具有上下文相似性的学习:这种类型的方法通过探索样本之间的潜在上下文相似性来学习点云的无监督表示。图 16 提供了实例对比学习的说明。给定一个输入点云对象实例作为锚点，它的增强视图被定义为正样本，而其他不同的实例是负样本。网络通过优化一个自我判别任务来学习点云的表示，即查询(锚点的特征)应该靠近正关键字(正样本的特征)和远离其负关键字(负样本的特征)。这种学习策略以无监督的方式将相似样本的表示分组在一起，并帮助网络从无标记的数据分布中学习语义结构。下面定义的 InfoNCE 损失[120]及其变体通常被用作训练中的目标函数：

$$\mathcal{L}_{\text{InfoNCE}} = -\log \frac{\exp(q \cdot k_+/\tau)}{\sum_{i=0}^K \exp(q \cdot k_i/\tau)}, \tag{4}$$

其中 q 是编码查询， $\{k_0, k_1, k_2, \dots\}$ 是键， k_+ 是正键， τ 是控制分布如何集中的温度超参数。

与基于生成的方法类似，人们提出了不同的对比学习方法 [99]，[100]，[102]，[121]，[122]来学习合成对象上的表示。为

表 3

基于上下文的点云无监督表示学习方法总结

Method	Published in	Category	Contribution
MultiTask [96]	ICCV 2019	Hybrid	Learning by clustering, reconstruction, and self-supervised classification
Jigsaw3D [13]	NeurIPS 2019	Spatial-context	Learning by solving 3D jigsaws
Constrast&Cluster [97]	3DV 2019	Hybrid	Learning by contrasting and clustering with GNN
GLR [98]	CVPR 2020	Hybrid	Learning by global-local reasoning for 3D objects
Info3D [99]	ECCV 2020	Context-similarity	Learning by contrasting global and local parts of objects
PointContrast [54]	ECCV 2020	Context-similarity	Learning by contrasting different views of scene point clouds
ACD [100]	ECCV 2020	Context-similarity	Learning by contrasting convex components decomposed from 3D objects
Rotation3D [101]	3DV 2020	Spatial-context	Learning by predicting rotation angles
HNS [102]	ACM MM 2021	Context-similarity	Learning by contrasting local patches of 3D objects with hard negative sampling
CSC [3]	CVPR 2021	Context-similarity	Techniques to improve contrasting scene point cloud views
STRL [1]	ICCV 2021	Temporal-context	Learning spatio-temporal data invariance from point cloud sequences
RandomRooms [103]	ICCV 2021	Context-similarity	Constructing pseudo scenes with synthetic objects for contrastive learning
DepthContrast [104]	ICCV 2021	Context-similarity	Joint contrastive learning with points and voxels
SelfCorrection [105]	ICCV 2021	Hybrid	Learning by distinguishing and restoring destroyed objects
PC-FractalDB [106]	CVPR 2022	Context-similarity	Leveraging fractal geometry to generate high-quality pre-training data
4dcontrast [107]	ECCV 2022	Temporal-context	Learning by contrasting dynamic correspondences from 3D scene sequences
DPCo [108]	ECCV 2022	Context-similarity	A unified contrastive-learning framework for point cloud pre-training
ProposalContrast [109]	ECCV 2022	Context-similarity	Pre-training 3D detectors by contrasting region proposals
MaskPoint [110]	ECCV 2022	Context-similarity	Learning by discriminating masked object points and sampled noise points
FAC [111]	CVPR 2023	Context-similarity	Learning by contrasting between grouped foreground and background

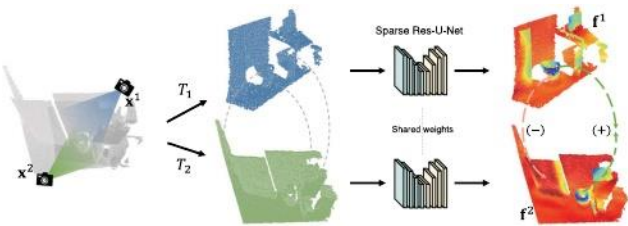


图 17。PointContrast 的流水线[54]:对两个不同视点捕获的同一场景的两次扫描 x_1 和 x_2 进行 T_1 和 T_2 变换,进行数据增强。计算两个视图之间的对应映射,以最小化匹配点特征的距离,最大化未匹配点特征的距离,以进行对比学习。图是在作者许可下从[54]中提取的。

例如, Sanghi 等人[99]提出通过最大化合成对象与其局部分之间的互信息来学习有用的特征表示。Wang et al.[121]提出了一种混合对比学习策略,使用不同分辨率的对象进行实例级对比以捕获分层的全局表示,同时对比点和实例以学习局部特征。Gadelha 等人[100]将 3D 对象分解为凸组件,并在相同组件之间构建正对,在不同组件之间构建负对,用于对比学习。Du 等人[102]在实例和局部部件之间的对比学习中引入了一种硬负采样策略。此外, Rao 等[98]将对比学习、法向量估计和自重建统一到同一个框架中,并制定了多任务学习方法。

最近, Xie 等人提出了 PointContrast[54],这是一个学习场景点云表示的对比学习框架,如图 17 所示。该工作首次表明,在 3D 场景部分帧上预训练的网络权重,在对多个 3D 高级任务(包括目标分类、语义分割和目标检测)进行微调时,可以导致性能提升。首先,在 ScanNet[18]的两个对齐视图之间提取密集对应关系以构建点对,然后进行点级别的对比学习

具有统一骨干网(SR-UNet)。最后,将学习到的模型迁移到多个下游 3D 任务,包括分类、语义分割和目标检测,并获得一致的性能增益。

由于 PointContrast 带来了新的见解,即从场景级点云学习的无监督表示可以跨域推广并促进高级场景理解任务,因此针对场景级 3D 任务提出了几种无监督预训练工作。考虑到 PointContrast 侧重于点级对齐,而没有捕捉场景中的空间配置和上下文, Hou 等[3]通过将空间划分为空间非均匀的单元进行对应匹配,将空间上下文整合到预训练目标中。Hou 等人[123]构建了一个多模态对比学习框架,该框架可以模拟 2D 多视图对应以及几何到图像对齐的 2D- 3d 对应。虽然上述工作[3], [54], [123]需要从多个相机视角捕获的 3D 数据,但 Zhang 等人[104]提出了可以与单视角数据一起工作的 DepthContrast。Rao 等人[103]没有使用真实的点云作为之前的方法,而是从 ShapeNet[14]生成合成场景和物体用于网络预训练。

另一种学习上下文相似性的无监督方法是聚类。在这种方法中,首先通过 K-Means[124]等聚类算法将样本分成簇,并为每个样本分配一个簇 ID 作为伪标签。然后以有监督的方式训练网络,以学习数据分布的语义结构。学习到的参数用于模型初始化,用于微调各种下游任务。一个典型的例子是 DeepClustering[125],它是 2d 视觉表示学习的第一个监督聚类方法。然而,对于点云的 URL,目前还没有研究采用纯粹的聚类策略。相反,通过将聚类与其他无监督学习方法(例如,自重建[96]或对比学习[97])集成,以学习更鲁棒的表示,混合方法被提出。

2)具有空间上下文结构的学习:具有空间坐标的点云提供了精确的几何描述

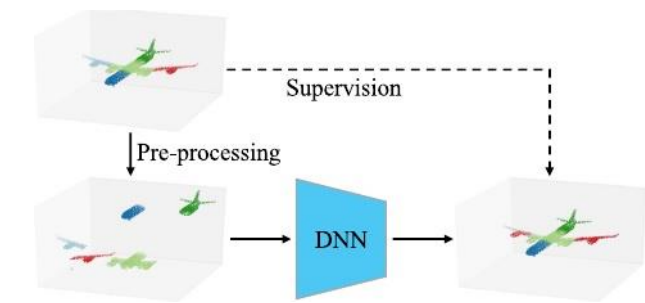


图 18 所示。3DJigsaw[13]的流水线:一个对象被分割成体素，每个点被分配一个体素标签。分割后的体素通过预处理进行随机重排，训练一个深度神经网络来预测每个点的体素标签。图是基于[13]重现的。

物体和场景的 3D 形状。点云中丰富的空间上下文可以用于 URL 的前置文本任务。例如，可以训练网络来整理不同对象部分的关系。同样，学习到的参数可以用于下游任务的模型初始化。由于在训练中不需要人工注释，所以关键是设计有效的 pre-text 任务来利用空间上下文作为 URL 目标。

Sauder 等人提出的 Jigsaw3D[13]方法是将空间上下文用于点云 URL 的先驱作品之一。如图 18 所示，对象首先被分割成体素，每个点被分配一个体素标签。然后用随机重新排列的点云填充网络，并通过预测每个点的正确体素标签来优化网络。在训练过程中，网络旨在从点云中提取空间关系和几何信息。在他们接下来的工作中[126]，设计了另一个前文本任务，以预测来自同一对象的两个局部部分的十个空间关系中的一个。受预测图像旋转的 2D 方法[127]的启发，Poursaeed 等人[101]提出通过预测 3D 对象的旋转角度来学习表示。Thabet et al.[128]设计了一个预文本任务，预测由 Morton-order 空间填充曲线定义的点序列中的下一个点。Chen 等[105]提出通过区分形状中扭曲的部分和正确的部分来学习物体的空间上下文。Sun 等人[129]引入了一种混合-解缠任务来利用空间上下文线索。

3)具有时间上下文结构的学习:点云序列是一种常见的点云数据类型，由连续的点云帧组成。例如，有由 RGB-D 视频帧[18]和激光雷达序列数据[19]、[130]、[131]转换而成的室内点云序列，连续点云扫描，每次扫描由激光雷达传感器一次扫描收集。点云序列包含丰富的时间信息，可以通过设计前置文本任务来提取这些信息，并用作训练 dnn 的监督信号。学习到的表示可以转移到下游任务。

最近，Huang 等人[1]提出了一个时空表征学习(spatial-temporal Representation Learning, STRL)框架，如图 19 所示。STRL 将 BYOL[8]从 2D 视觉扩展到 3D 视觉，从点云中提取时空表征。它将两个相邻的点云帧视为正对，并最小化样本对学习到的特征表示之间的均方误差。Chen 等人[107]利用

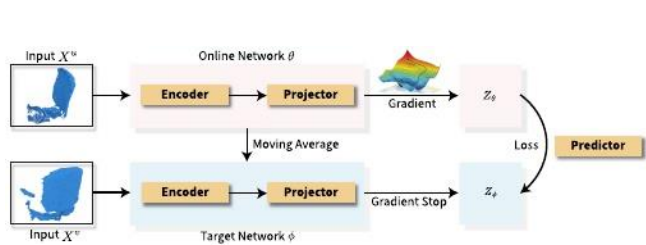


图 19 所示。STRL 的流水线[1]:在线网络从相邻的两个点云框架 x^u 和 x^v 中学习时空结构。图来自[1]，并经作者许可。

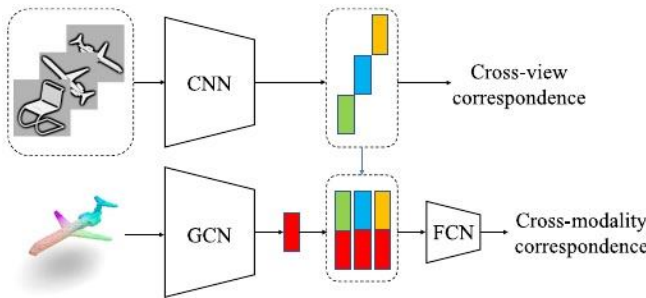


图 20。流水线 CMCV [4]:CMCV 使用 2D CNN 从 3D 物体的渲染视图中提取 2D 特征，使用 3D GCN 从点云中提取 3D 特征。这两种类型的特征由一个两层的全连接网络(FCN)连接起来，以预测跨模态对应。图是基于[4]进行复制的。

在静态 3D 环境中移动的合成 3D 形状，以时间顺序创建动态场景和样本对。他们进行对比学习，以学习具有动态理解的 3D 表示。

具有时间上下文结构的无监督学习在 2D 计算机视觉任务[132]、[133]、[134]、[135]和 3D 计算机视觉任务[1]、[107]中都证明了其有效性。如第 VII 节所讨论的，这个方向非常有希望，但还需要更多的研究来更好地获取时间上下文信息。

C.多种基于模态的方法

不同的模态如图像[19]和自然语言描述[136]可以为点云数据提供额外的信息。跨模态的关系建模可以设计为 URL 的 pre-text 任务，这有助于网络学习更健壮和全面的表示。同样，学习到的参数可以用作各种下游任务的初始化权重。

最近的一些工作[4]，[137]利用了 URL 在 3D 点云对象和 2D 图像之间的对应关系。例如，Jing 等人[4]将具有不同相机视角的 3D 对象渲染为 2D 图像，用于从多模态数据中学习。如图 20 所示，他们分别使用 2D CNN 和 3D GCN 提取图像特征和点云特征，然后对模态内对应和跨模态对应进行对比学习。他们的研究表明，与随机初始化相比，预训练的 2D CNN 和 3D GCN 都取得了更好的分类效果。不同的是，Wang 等人[138]将点云投影到彩色图像中，然后将它们输入到具有冻结权重的图像预训练模型中

表 4

比较 modelnet10 和 modelnet40[27] 上的线性形状分类:线性 SVM 分类器使用不同的无监督方法学习到的表示进行训练。*突出显示的精度是通过多模态数据预训练得到的。
[T]表示

使用改进的 transformer 的模型。[st]表示使用标准 transformer 的模型

Method	Year	Pre-text task	Backbone	Pre-train dataset	ModelNet10	ModelNet40
Supervised learning	2017	N.A.	PointNet [15]	N.A.	-	89.2
	2017		PointNet++ [48]		-	90.7
	2019		DGCNN [49]		-	93.5
	2019		RSCNN [56]		-	93.6
	2021		[T]PointTransformer [144]		-	93.7
	2022		[ST]Transformer [57]		-	91.4
SPH [145]	2003	Generation	-	ShapeNet	79.8	68.2
LFD [146]	2003	Generation	-	ShapeNet	79.9	75.5
TL-Net [63]	2016	Generation	-	ShapeNet	-	74.4
VConv-DAE [62]	2016	Generation	-	ShapeNet	80.5	75.5
3D-GAN [64]	2016	Generation	-	ShapeNet	91.0	83.3
3D DescriptorNet [65]	2018	Generation	-	ShapeNet	-	92.4
FoldingNet [66]	2018	Generation	-	ModelNet40	91.9	84.4
FoldingNet [66]	2018	Generation	-	ShapeNet	94.4	88.4
Latent-GAN [68]	2018	Generation	-	ModelNet40	92.2	87.3
Latent-GAN [68]	2018	Generation	-	ShapeNet	95.3	85.7
MRTNet [69]	2018	Generation	-	ShapeNet	86.4	-
VIP-GAN [70]	2019	Generation	-	ShapeNet	94.1	92.0
3DCapsuleNet [47]	2019	Generation	-	ShapeNet	-	88.9
PC-GAN [88]	2019	Generation	-	ModelNet40	-	87.8
L2G-AE [71]	2019	Generation	-	ShapeNet	95.4	90.6
MAP-VAE [72]	2019	Generation	-	ShapeNet	94.8	90.2
PointFlow [73]	2019	Generation	-	ShapeNet	93.7	86.8
MultiTask [96]	2019	Hybrid	-	ShapeNet	-	89.1
Jigsaw3D [13]	2019	Context	PointNet	ShapeNet	91.6	87.3
Jigsaw3D [13]	2019	Context	DGCNN	ShapeNet	94.5	90.6
ClusterNet [97]	2019	Context	DGCNN	ShapeNet	93.8	86.8
CloudContext [126]	2019	Context	DGCNN	ShapeNet	94.5	89.3
NeuralSampler [90]	2019	Generation	-	ShapeNet	95.3	88.7
PointGrow [77]	2020	Generation	-	ShapeNet	85.8	-
Info3D [99]	2020	Context	PointNet	ShapeNet	-	89.8
Info3D [99]	2020	Context	DGCNN	ShapeNet	-	91.6
ACD [100]	2020	Context	PointNet++	ShapeNet	-	89.8
PDL [74]	2020	Generation	-	ShapeNet	-	84.7
GLR [98]	2020	Hybrid	PointNet++	ShapeNet	94.8	92.2
GLR [98]	2020	Hybrid	RSCNN	ShapeNet	94.6	92.2
SA-Net-cls [76]	2020	Generation	-	ShapeNet	-	90.6
GraphTER [75]	2020	Generation	-	ModelNet40	-	89.1
Rotation3D [101]	2020	Context	PointNet	ShapeNet	-	88.6
Rotation3D [101]	2020	Context	DGCNN	ShapeNet	-	90.8
MID [121]	2020	Context	HRNet	ShapeNet	-	90.3
GTIF [85]	2020	Generation	HRNet	ShapeNet	95.9	89.6
HNS [102]	2021	Context	DGCNN	ShapeNet	-	89.6
ParAE [147]	2021	Generation	PointNet	ShapeNet	-	90.3
ParAE [147]	2021	Generation	DGCNN	ShapeNet	-	91.6
CMCV [4]	2021	Multi-modal	DGCNN	ShapeNet	-	89.8*
GSIR [86]	2021	Context	DGCNN	ModelNet40	-	90.4
STRL [1]	2021	Context	PointNet	ShapeNet	-	88.3
STRL [1]	2021	Context	DGCNN	ShapeNet	-	90.9
PSG-Net [78]	2021	Generation	PointNet++	ShapeNet	-	90.9
SelfCorrection [105]	2021	Hybrid	PointNet	ShapeNet	93.3	89.9
SelfCorrection [105]	2021	Hybrid	RSCNN	ShapeNet	95.0	92.4
OcCo [12]	2021	Generation	[ST]Transformer	ShapeNet	-	92.1
CrossPoint [137]	2022	Multi-modal	PointNet	ShapeNet	-	89.1*
CrossPoint [137]	2022	Multi-modal	DGCNN	ShapeNet	-	91.2*
Point-BERT [57]	2022	Generation	[ST]Transformer	ShapeNet	-	93.2
Point-MAE [79]	2022	Generation	[ST]Transformer	ShapeNet	-	93.8

为下游任务提取代表性特征。然而，如何学习无监督点云表示与其他模态，如文本描述和音频数据，仍然是一个未被探索的领域。我们期待在这个有前途的研究方向上有更多的研究。

D. 基于局部描述符的方法

上述方法旨在学习点云的语义结构以进行高层理解，而局部

基于描述符的方法专注于学习底层任务的表示。例如，Deng 等人[139]介绍了 PPF- FoldNet，它提取旋转不变的 3D 局部描述符进行 3D 匹配[140]。一些作品[141]，[142]利用非刚性形状对应提取作为点云 URL 的文本前任务，旨在找到两个可变形的 3D 形状的对点对应关系。Jiang 等人[143]探索无监督 3D 配准，以找到最佳的刚性变换，使源点云与目标精确对齐。

现有基于局部描述符的方法的性能主要在低层次任务上进行评估。然而，如何将学习到的特征表示适应于其他高层任务却很少被讨论。我们期待未来会有更多相关的研究。

E. 利弊

*生成的方法。*由于能够在不承担任何下游任务的情况下恢复原始数据分布，因此在 3D URL 中得到了广泛的研究。然而，现有的大多数研究都集中在对象级点云上，其特点是点数量有限和数据可变性，限制了它们对对象分类和部分分割任务的适用性。此外，由于难以生成分布复杂、噪声和稀疏变化丰富、遮挡多样的场景级点云，这些方法在场景级任务(如 3D 目标检测和语义分割)中表现出有限的有效性。尽管如此，基于生成的方法最近在 2D 图像[10]中取得了非常令人印象深刻的进展，展示了它们在处理 3D 点云数据方面的巨大潜力。预计在场景级任务以及各种下游应用中会有更多的努力。

*基于上下文的方法。*最近已成为场景级任务中的一种流行方法，如 3D 语义分割、3D 实例分割和 3D 目标检测，这得益于它们在处理复杂现实世界数据方面的能力。然而，它们仍然面临着几个挑战。第一个是难例挖掘，这对有效的对比学习至关重要。除此之外，设计有效的自我监督对于基于上下文的方法也是具有挑战性的，特别是在考虑各种任务和应用的泛化时。

*多种基于模态的方法。*允许利用额外的数据模态来丰富点云的分布。点云和其他数据模态之间的成对对应也提供了额外的监督，从而增强了学习到的无监督点云表示。然而，多模态方法仍然面临着几个挑战。例如，获取大规模成对数据往往是一项非平凡的任务，设计有效的跨域任务也是如此。此外，如何跨多种模态学习有效的同质表示空间仍然是一个非常开放的研究问题。

*基于局部描述符的方法。*在捕捉详细的空间线索和利用低级位置信息方面提供了明显的优势。然而，这些方法将学习到的表示迁移到高级识别模型的能力有限，这限制了它们在更复杂和抽象的识别任务中的应用范围。

六、基准性能

我们用两种广泛采用的评价指标对具有代表性的 3D URL 方法进行了基准测试。基准测试是在公共点云数据上进行的，其中所有性能都是从相应的论文中提取的。

A. 评价标准

有两个指标被广泛采用来评估学习的无监督点云表示的质量。

- 线性分类*首先应用预训练的无监督模型从某些标记数据中提取特征。然后它用提取的特征和相应的标签一起训练一个有监督的线性分类器，其中预学习的无监督表示的质量是通过训练的线性分类器对测试数据的性能进行评估。因此，线性分类可以被视为一种以隐式方式提供聚类分析的表示学习度量。

*微调*使用来自下游任务的标记数据来优化预训练的无监督模型。它可以通过评估微调模型在下游测试数据上的性能来评估预学习的无监督表示的质量，*即与随机初始化相比*，无监督预训练可以获得多少性能增益。

注意 URL 可以用其他定量指标来评估。例如，*重构误差*[66]可以告诉我们学习到的表示如何编码原始点云。不同的聚类指标，如*归一化互信息*[96]，可以补充线性分类指标。然而，这些指标大多是特定于任务的，例如，由于其固有的噪声、遮挡和稀疏性，重建误差可能无法很好地评估场景级点云的表示。事实上，很少有通用指标可以直接和明确地评估学习的 3D 无监督表示的质量，尽管它对 3D URL 研究至关重要。需要沿着这个方向进行更多的研究，以进一步推进这一研究领域。

除了定量的指标，无监督的特征表示可以以定性的方式进行评估。例如，t-SNE (t 分布随机邻域嵌入)[148]被广泛采用来压缩学习到的特征表示的维度，并将压缩后的特征嵌入可视化。

B. 对象级任务

1)*对象分类*:对象分类是评估中使用最广泛的任务，因为大多数现有工作都在对象级点云数据集上学习点云表示。如 VI-A 节所述，两种类型的协议都被广泛采用，包括线性分类协议和微调协议。

表 IV 总结了线性分类协议通过现有方法的性能。具体来说，使用不同的无监督方法在 ShapeNet 或 ModelNet40 数据集上学习到的表示来训练线性分类器，并报告在 ModelNet10 和 ModelNet40 上测试集上的分类结果。为了进行比较，我们还列出了相同骨干模型在相同数据集上的监督学习性能。可以看出，无监督学习方法的性能在不断提高，有些方法甚至已经超越了有监督学习方法，这说明了点云 URL 的有效性和巨大的潜力。

表五世

scanobjcnn 中无监督预训练性能在对象分类数据集 modelnet40 和 obj-b gsplit 上的比较。性能数字以 “a /b” 的格式呈现, “a” 表示用随机初始化从头开始训练分类模型, “b” 表示用无监督初始化的微调分类模型

PRE-TRAINED 模型。 “a” 下的性能可能会因相关论文中报告的不同实现而有所不同

Method	Backbone	ModelNet40	ScanObjectNN
Jigsaw3D [13]	PointNet [15]	89.2/89.6(+0.4)	73.5/76.5(+3.0)
Info3D [99]	PointNet [15]	89.2/90.2(+1.0)	-/-
SelfCorrection [105]	PointNet [15]	89.1/90.0(+0.9)	-/-
OcCo [12]	PointNet [15]	89.2/90.1(+0.9)	73.5/80.0(+6.5)
ParAE [147]	PointNet [15]	89.2/90.5(+1.3)	-/-
Jigsaw3D [13]	PCN [113]	89.3/89.6(+0.3)	78.3/78.2(-0.1)
OcCo [12]	PCN [113]	89.3/90.3(+1.0)	78.3/80.4(+2.1)
GLR [98]	RSCNN [56]	91.8/92.2(+0.5)	-/-
SelfCorrection [105]	RSCNN [56]	91.7/93.0(+1.3)	-/-
Jigsaw3D [13]	DGCNN [49]	92.2/92.4(+0.2)	82.4/82.7(+0.3)
Info3D [99]	DGCNN [49]	93.5/93.0(-0.5)	-/-
OcCo [12]	DGCNN [49]	92.5/93.0(+0.5)	82.4/83.9(+1.6)
ParAE [147]	DGCNN [49]	92.2/92.9(+0.7)	-/-
STRL [1]	DGCNN [49]	92.2/93.1(+0.9)	-/-
OcCo [12]	Transformer [57]	91.2/92.2(+1.0)	79.9/84.9(+5.0)
Point-BERT [57]	Transformer [57]	91.2/93.4(+2.2)	79.9/87.4(+7.5)

表 V 列出了 ModelNet40 和 ScanObjectNN 数据集上的微调性能。我们可以看到, 无论骨干架构如何, 与随机初始化相比, 用无监督预训练权重初始化的分类模型总是能取得更好的分类性能。另一方面, 性能差距仍然有限, 这主要是由于预训练数据集(即 ShapeNet 和 ModelNet40)的大小和多样性有限, 以及现有骨干模型的简单性。相比之下, 由于更大的预训练数据集 ImageNet[33]和更强大的骨干网络 ResNet[37], 最先进的 2D 图像无监督预训练方法能够在分类任务中实现更显著的性能提升。正如在第七节中讨论的那样, 我们期待更多样化的数据集和更先进、更慷慨的骨干模型, 可以为这个领域奠定更强大的基础。

2)对象部分分割:表 VI 给出了使用线性分类协议(即表 VI 中的 “Un- sup”)和微调协议(即 “Trans”)在 ShapeNetPart 数据集[14]上对对象部分分割的基准测试。(见表六), 如 VI-A 节所述。如表所示, 无监督学习和有监督学习(即 “Unsup. “vs. Sup.”)的差距正在缩小。此外, 在大多数情况下, 在微调协议(即 “Trans”)下, 无监督预训练可以获得更好的性能。” vs. “Sup.”), 尽管改进仍然有限。

C.场景级任务

如 V-B 节所讨论的, 由于其在各种应用中的巨大潜力, 场景级任务中的无监督预训练最近变得流行起来。这伴随着一系列 3D URL 研究, 研究了不同场景级点云数据集预训练的有效性。我们提供了这些方法在不同 3D 任务方面的全面基准。

表六世

形状部分三维 URL 分割方法与 shapenetpart 的比较[14]。 “UNSUP.” 表示学习到的无监督点特征的线性分类。 “反式”。以 “a /b” 的格式呈现, 其中 “a” 是通过分割模型获得的

从头开始随机初始化训练, “b” 是通过微调分割模型得到的, 这些分割模型是用无监督预训练模型初始化的。我们还用 random 提供了不同骨干模型的监督性能 (“sup .”))

初始化(从原始论文中提取)

URL Method	Type	Backbone	class mIoU	instance mIoU
N.A.	Sup.	PointNet	80.4	83.7
	Sup.	PointNet++	81.9	85.1
	Sup.	DGCNN	82.3	85.1
	Sup.	RSCNN	84.0	86.2
	Sup.	Transformer	83.4	85.1
Latent-GAN [68]	Unsup.	-	57.0	-
MAP-VAE [72]	Unsup.	-	68.0	-
CloudContext [126]	Unsup.	DGCNN	-	81.5
GraphTER [75]	Unsup.	-	78.1	81.9
MID [121]	Unsup.	HRNet	83.4	84.6
HNS [102]	Unsup.	DGCNN	79.9	82.3
CMCV [4]	Unsup.	DGCNN	74.7	80.8
SO-Net [67]	Trans.	SO-Net	-/-	84.6/84.9(+0.3)
Jigsaw3D [13]	Trans.	DGCNN	82.3/83.1(+0.8)	85.1/85.3(+0.2)
MID [121]	Trans.	HRNet	84.6/85.2(+0.6)	85.5/85.8(+0.3)
CMCV [4]	Trans.	DGCNN	77.6/79.1(+1.5)	83.0/83.7(+0.7)
OcCo [12]	Trans.	PointNet	82.2/83.4(+1.2)	-/-
OcCo [12]	Trans.	DGCNN	84.4/85.0(+0.6)	-/-
OcCo [12]	Trans.	Transformer	83.4/83.4(+0.0)	85.1/85.1(+0.0)
Point-BERT [57]	Trans.	Transformer	83.4/84.1(+0.7)	85.1/85.6(+0.5)

表 VII 和表 VIII 显示了在 S3DIS[21]数据集上的语义分割性能。由于在之前的工作中使用了不同的微调设置, 我们分别对它们进行了汇总。在表 7 中, 无监督预训练的 DGCNN 在 S3DIS 的每个区域上进行微调, 并在区域 5(表的上半部分)或区域 6(表的下半部分)上进行测试。相反, 表 VIII 通过遵循一折(表上半部分)和六折交叉验证设置(表下半部分), 分别展示了用整个数据集对不同分割网络进行微调的性能。

我们还总结了现有的处理目标检测的无监督预训练的工作。表 IX 和表 X 分别展示了它们在室内数据集 SUN RGB-D[28]和 ScanNet-V2[18]以及室外 LiDAR 数据集 ONCE[30]上的性能。此外, 一些工作研究了实例分割的无监督预训练。我们在表 11 中总结了它们在 S3DIS[21]和 ScanNet-V2[18]上的性能。

令人鼓舞的是, 与从零开始训练相比, 无监督学习表示可以跨域泛化, 并提高多个高级 3D 任务的性能。这些实验证明了点云 URL 在节省昂贵的人工注释方面的巨大潜力。然而, 改进仍然有限, 我们期待在这一领域进行更多的研究。

7. 未来的方向

近十年来, 点云的 URL 研究取得了重大进展。在本节中, 我们分享了这个研究领域的几个潜在的未来研究方向。

表七世

S3 DIS 上的语义分割[21]:将监督训练与随机权值初始化进行比较, 将微调与从无监督预训练任务中学习到的预训练权值进行比较。它使用 DGCNN 作为分割模型, 并对其进行训练

不同的单一区域, 并在 5 区(上部)和 6 区(下部)进行测试

Method	OA on area 5 with different train area					mIoU on area 5 with different train area				
	Area1	Area2	Area3	Area4	Area6	Area1	Area2	Area3	Area4	Area6
from scratch	82.9	81.2	82.8	82.8	83.1	43.6	34.6	39.9	39.4	43.9
Jigsaw3D [13]	83.5(+0.6)	81.2(+0.0)	84.0(+1.2)	82.9(+0.1)	83.3(+0.2)	44.7(+1.1)	34.9(+0.3)	42.4(+2.5)	39.9(+0.5)	43.9(+0.0)
ParAE [147]	91.8(+8.9)	82.3(+1.1)	89.5(+6.7)	88.2(+5.4)	86.4(+3.3)	53.5(+9.9)	38.5(+3.9)	48.4(+8.5)	45.0(+5.6)	49.2(+5.3)

Method	OA on area 6 with different train area					mIoU on area 6 with different train area				
	Area1	Area2	Area3	Area4	Area5	Area1	Area2	Area3	Area4	Area5
from scratch	84.6	70.6	77.7	73.6	76.9	57.9	38.9	49.5	38.5	48.6
STRL [1]	85.3(+0.7)	72.4(+1.8)	79.1(+1.4)	73.8(+0.2)	77.3(+0.4)	59.2(+1.3)	39.2(+0.8)	51.9(+2.4)	39.3(+0.8)	49.5(+0.9)

表八世

在 s3dis[21]上的语义分割性能。上部:

模型在区域 5(折叠#1)上进行测试, 其余部分进行训练

数据。下半部分:三次运行的六折交叉验证

Method	Backbone	mACC	mIoU
from scratch	SR-UNet	75.5	68.2
PointConstrast [54]		77.0	70.9
DepthContrast [104]		-	70.6
Method	Backbone	OA	mIoU
from scratch	PointNet	78.2	47.0
Jigsaw3D [13]		80.1	52.6
OcCo [12]		82.0	54.9
from scratch	PCN	82.9	51.1
Jigsaw3D [13]		83.7	52.2
OcCo [12]		85.1	53.4
from scratch	DGCNN	83.7	54.9
Jigsaw3D [13]		84.1	55.6
OcCo [12]		84.6	58.0

表九世

不同无监督学习方法的预训练效果比较。基准测试是 sun rgb-d[28]和 scannet-v2[18]数据集上的 3d 目标检测任务。“@0.25”和“@0.5”代表每个类别的平均精度(ap)结果与 iou 阈值 0.25(map @0.25)和 0.5 (map @0.5),

分别

Method	Backbone	SUN RGB-D		ScanNet-V2	
		@0.5	@0.25	@0.5	@0.25
from scratch	SR-UNet	31.7	55.6	35.4	56.7
PointConstrast [54]		34.8	57.5	38.0	58.5
PC-FractalDB [106]		35.9	57.1	37.0	59.4
from scratch	VoteNet	32.9	57.7	33.5	58.6
STRL [1]		-	58.2	-	-
RandRooms [103]		35.4	59.2	36.2	61.3
DepthContrast [104]		-	-	-	62.2
CSC [3]		33.6	-	-	-
PointContrast [54]		34.0	-	38.0	-
4DContrast [107]		34.4	-	39.3	-
from scratch	PointNet++	-	57.5	-	58.6
PointContrast [54]		-	57.9	-	58.5
RandRooms [103]		-	59.2	-	61.3
DepthContrast [104]		-	60.7	-	-
PC-FractalDB [106]		33.9	59.4	38.3	61.9
DPCo [108]		35.6	59.8	41.5	64.2
from scratch	H3DNet	39.0	60.1	48.1	67.3
RandRooms [103]		43.1	61.6	51.5	68.6

需要统一的 3D 主干:深度学习在 2D 计算机视觉领域取得巨大成功的一个主要原因是 VGG[36]、ResNet[37]等 CNN 架构的标准化。例如, 统一的骨干网结构极大地促进了跨不同数据集和任务的知识迁移。对于 3D 点云, 尽管最近报道了各种 3D 深度架构, 但类似的发展还远远没有得到充分的探索。这可以从第 VI 节表中的 URL 方法中观察到, 其中大多数采用了非常不同的主干

表 X

数据集上的对象检测性能 once[30]。基线是从头开始训练的。无监督学习方法用于预训练模型。Usmall、Umedian 和 ullarge 表示小、中、大数量的未标记数据

分别用于无监督学习

Method	Vehicle	Pedestrian	Cyclist	mAP
Baseline [149]	69.7	26.1	59.9	51.9

U_{small}

BYOL [8]	67.6	17.2	53.4	46.1 (-5.8)
PointContrast [54]	71.5	22.7	58.0	50.8 (-0.1)
SwAV [150]	72.3	25.1	60.7	52.7 (+0.8)
DeepCluster [125]	72.1	27.6	50.3	53.3 (+1.4)

U_{median}

BYOL [8]	69.7	27.3	57.2	51.4 (-0.5)
PointContrast [54]	70.2	29.2	58.9	52.8 (+0.9)
SwAV [150]	72.1	28.0	60.2	53.4 (+1.5)
DeepCluster [125]	72.1	30.1	60.5	54.2 (+2.3)

U_{large}

BYOL [8]	72.2	23.6	60.5	52.1 (+0.2)
PointContrast [54]	73.2	27.5	58.3	53.0 (+1.1)
SwAV [150]	72.0	30.6	60.3	54.3 (+2.4)
DeepCluster [125]	71.9	30.5	60.4	54.3 (+2.4)

TABLE XI

在数据集 s3dis[21]和 scannet-v2[18]上的实例分割性能。它报告了所有语义类的平均精度均值(map), 3d iou 阈值为 0.25

Method	Backbone	S3DIS	ScanNet
from scratch	SR-UNet	59.3	53.4
PointContrast [54]		60.5	55.8
CSC [3]		63.4	56.5
4DContrast [107]		-	57.6

模型。这阻碍了 3D 点云网络在可扩展设计和各种新任务中的高效部署的发展。设计某些在 2D 计算机视觉中可以像 ResNet 一样普遍存在的通用主干, 对于包括无监督点云表示学习在内的 3D 点云网络的进步至关重要。

需要更大的数据集:如第三节所述, 大多数现有的 URL 数据集最初是为监督学习的任务收集的。由于点云标注费力且耗时, 这些数据集在数据规模和数据多样性方面受到严重限制, 不适合用于通常需要大量良好规模和多样性的点云的 URL。这个问题很好地解释了第六节表格中 URL 的微小改进。因此, 迫切需要在对象级和场景级点云、室内和室外点云等方面收集足够多样性的大规模、高质量的未标记点云数据集。

场景级任务的无监督预训练:如 V-B 节所述, 尽管一些先驱研究[1], [3], [54], [103], [123]探索了如何在场景级点云上预训练 dnn, 以改进各种场景级下游任务, 如目标检测和实例分割。之前的研究表明, 学习到的无监督表示可以有效地跨域和跨任务泛化。因此, 场景级点云的 URL 作为一个新的研究方向, 具有广阔的应用前景, 值得关注。另一方面, 这条路线上的研究仍处于初期阶段, 主要是由于网络架构和数据集的限制。我们预见, 在不久的将来, 将会进行更多的相关研究。

从多模态数据中学习表示:3D 传感器通常配备其他传感器, 可以捕获额外和互补的信息, 以点云。例如, 深度相机通常配备光学传感器, 以捕获更好的外观信息。激光雷达传感器、光学传感器、GPU 和 IMU 通常作为传感器套件安装在一起, 以捕获互补信息, 并在自动驾驶车辆和移动机器人导航中提供一定的冗余。近年来, 从这样的多模态数据中进行无监督学习越来越受到关注。例如, 如 V-C 节所述, 学习多模态数据之间的对应关系已被探索为无监督学习的前文本任务。然而, 沿着这条研究路线的研究仍然缺乏调查, 我们期待更多相关的研究点云、RGB 图像、深度图等。

学习时空表示:支持捕获顺序点云的 3D 传感器如今变得越来越受欢迎。从点云流中提取丰富的时间信息可以作为有用的监督信号用于无监督学习, 而现有的大多数工作仍然专注于静态点云。我们预计将设计更有效的前置任务, 可以有效地从无标记的连续点云框架中学习时空表示。

8. 结论

无监督表示学习旨在从未标注的数据中学习有效的表示, 这在点云数据的研究中已经展示了令人印象深刻的进展。本文介绍了点云的无监督表示学习的当代综述。首先介绍了广泛采用的数据集和深度网络架构。然后对方法进行全面的分类和详细的回顾。然后, 讨论了代表性的方法, 并在多个 3D 点云任务上进行了基准测试。最后, 分享了我们几个潜在的未来研究方向的拙见。我们希望这项工作可以为未来从点云数据中进行无监督表示学习的研究奠定坚实的基础。

参考文献。

[1] S.Huang、Y.Xie S.-C.Zhu,安迪。朱, “三维点云的时空自监督表示学习”, *IEEE/CVF Int 项目。相依第一版。Vis.*, 2021, pp. 6535-6545。

[2]郭勇, 王红红, 胡强, 刘林, 刘林, “三维点云的深度学习研究综述”, *IEEE, 第2 版。模式肛门。马赫。智能。*, 第 32 卷, no. 12, pp. 4338-4364, 2021 年 12 月。

[3]侯杰, 谢世杰, 张建军, “基于场景背景的三维场景识别”, *IEEE/CVF Conf. 计算, 程序。视觉模式识别。*, 2021, 第 15 587-15 597 页。

[4]景亮, 张亮, 田勇, “基于交叉模态和交叉视图对应的自监督特征学习”, *IEEE/CVF Conf 计算程序。视觉模式识别。*, 2021, pp. 1581-1591。

[5] A. Radford 等, “语言模型的无监督多任务学习”, *OpenAI Blog*, vol. 1, no. 5. 2019 年第 8 期, 艺术。不。9。

[6]张志强。C. Kenton 和 L. K. Toutanova, “BERT:深度双向转换器用于语言理解的预训练”, *《Conf 程序》, 北美。章 Assoc. 计算。语言学——哼。韵。抛光工艺。*, 2019, 第 4171-4186 页。

[7]何开, 范辉, 吴勇, 谢淑, R. Girshick, “无监督视觉表征学习的动力对比”, *程序。IEEE/CVF Conf. 计算机学报。视觉模式识别。*, 2020, 第 9729-9738 页。

[8]王志强。Grill 等人, “引导你自己的潜在:一种自我监督学习的新方法”, *《神经网络过程》程序。系统。*, 2020, pp. 21271-21284。

[9]陈晓, 范辉, R. Girshick, 何开, “基于动量对比学习的改进基线”, 2020, *数学学报:2003.04* - 297。

[10]何开, 陈晓, 谢生, 李辉, P. Dollár, R. Girshick, “可扩展视觉学习的掩码自编码器”, *程序。IEEE/CVF Conf. 计算机学报。视觉模式识别。*, 2022, pp. 16 000-16 009。

[11]王晓明, 张晓明, “基于图像卷积的三维点云生成模型研究”, *《程序》, 集成。Conf. 学习。《表征》*, 2018 年, 第 1-11 页。

[12]王辉, 刘庆, 刘晓明, 王晓明, “基于遮挡补全的无监督点云预训练”, *IEEE/CVF 集成程序。Conf 计算。Vis.*, 2021, pp. 9782-9792。

[13]王志强, “自监督深度学习在点云中的应用”, *程序学报, 神经网络。系统。*, 2019, 第 12 962-12 972 页。

[14]张晓霞, 张晓明, “三维模型库的构建与实现”, 2015, *数学学报*, 32(4):555 - 557。

[15]齐长荣, 苏慧, 莫锴, L. J. gu, “基于点集的深度学习三维分类与分割”, *IEEE Conf 计算, 程序。视觉模式识别。*, 2017, 第 652-660 页。

[16]齐长荣, 何开, L. J. gu, “基于深度投票的点云三维目标检测”, *程序, IEEE/CVF. Conf 计算。Vis.*, 2019, pp. 9277-9286。

[17]史生, 王晓, 李辉, “基于点云的三维目标识别方法”, *IEEE/CVF Conf, 计算, 程序。视觉模式识别。*, 2019, pp. 770-779。

[18] A. Dai, 张晓霞, M. Savva, M. Halber, T. Funkhouser, M. niener, “Scannet:丰富注释的室内场景三维重建”, *IEEE Conf, 计算, 程序。视觉模式识别。*, 2017, pp. 5828-5839。

[19]张晓明, 张晓明, 张晓明, “基于视觉的机器人学研究”, *计算机科学, 2011. J.机器人。Res.*, 第 32 卷, 第 1 期。11, pp. 1231-1237, 2013。

[20]齐, 刘伟, 吴, 苏宏, gul. J., “基于 RGB-D 数据的三维目标检测方法”, *IEEE 会议计算, 进程。视觉模式识别。*, 2018, pp. 918-927。

[21]陈志强, “空间空间的三维语义分析”, *IEEE 会议, 计算, 进程。视觉模式识别。*, 2016, pp. 1534-1543。

[22]张晓明, 张晓明, 张晓明, “基于深度学习的计算机视觉技术研究进展”, *《计算机计算》。测量员*, 第 50 卷, no. 2, pp. 1-38, 2017。

[23]谢勇, 田健, 朱晓霞, “基于点云的点云语义分割方法研究”, *IEEE。Remote Sens. Mag.*, 第 8 卷, 第 1 期。4, 第 38-59 页, 2020 年 12 月。

[24]景亮, 田健, “基于深度神经网络的自监督视觉特征学习”, *《IEEE Trans》。模式肛门。马赫。智能。*, 卷八, no. 4, 第 38-59 页, 2020 年 12 月。

[25]刘晓, “自监督学习:生成式或对比式”, *《IEEE Trans》。数据中。*, 第 35 卷, 第 2 期。1, 第 857-876 页, 2023 年 1 月。

[26]王志强。齐和罗杰, “大数据时代的小数据挑战:无监督和半监督方法的最新进展调查”, *IEEE Trans. 模式肛门。马赫。智能。*, 第 44 卷, no. 1, 第 2168-2187 页, 2022 年 4 月。

[27]吴志等, “三维几何形状的深度表示”, *IEEE 会议, 计算, 进程。视觉模式识别。*, 2015, 第 1912-1920 页。

[28]宋绍鹏, 肖, “基于 RGB-D 的 RGB-D 场景理解测试”, *IEEE 会议计算程序. Vis. 模式识别.*, 2015, pp. 567-576.

[29]刘志强, 刘志强. 范, b.s. Hua, T. Nguyen 和 s - k. 杨, “重新审视点云分类:现实世界数据的新基准数据集和分类模型”, *IEEE/CVF 国际计算机学报程序. 会议计算.* Vis., 2019, pp. 1588-1597.

[30]毛志军, “自动驾驶的百万场景:一次数据集”, *arXiv.* 2021:2106.11037.

[31]张晓明, 张晓明, 张晓明, “基于图像的城市场景语义分割方法研究”, *《IEEE 会议, 计算》程序. Vis. 模式识别.*, 2016, pp. 3234-3243.

[32]王志强, 王志强. 张凯, 李和 K. Toutanova, “BERT:深度双向变换的语言理解预训练”, 2018,*arXiv:1810.04805*.

[33]邓建军, 董伟. 李, 李, 李飞飞, “Imagenet:一个大规模的分层图像数据库”, *IEEE 会议计算程序. Vis. 模式识别.*, 2009, pp. 248-255.

[34]张晓明, 张晓明, 张晓明, “基于八叉树的概率三维地图构建方法研究”, *北京交通大学学报.《机器人》*, 第 34 卷, 第 7 期. 3, 第 189-206 页, 2013.

[35]王晓明, 王晓明, 王晓明, “基于体素哈希的实时三维重建技术”, *计算机工程学报*, 2001,21(5):448 - 448. *图.*, 第 32 卷, no. 6, pp. 1-11, 2013.

[36]张晓明, 张晓明, “基于深度卷积神经网络的图像识别”, *程序, 国际计算机学报. 会议学习.《表征》*, 2015 年, 第 1-10 页.

[37]何凯, 张晓明, 任淑娟, 孙建军, “基于深度残差学习的图像识别”, *IEEE 会议, 计算, 程序. Vis. 模式识别.*, 2016, pp. 770-778.

[38]张晓明, 张晓明, 张晓明, “基于多视图卷积神经网络的三维形状识别”, *IEEE 国际计算机学报, 程序. 会议计算.* Vis., 2015, pp. 945-953.

[39]余天, 孟建军, 袁军, “多视图协调双线性网络在三维物体识别中的应用”, *IEEE 会议, 计算, 程序. Vis. 模式识别.*, 2018, pp. 186-194.

[40]杨斌, 罗文杰, “基于点云的实时三维目标检测”, *IEEE 会议, 计算, 程序. Vis. 模式识别.*, 2018, pp. 7652-7660.

[41]杨新, 王磊, “多视图三维物体识别的学习关系”, *IEEE/CVF 程序国际计算机学报. 会议计算.* Vis., 2019, pp. 7505 - 7514.

[42]魏晓明, 余天, 孙俊, “基于图像卷积网络的三维形状分析”, *IEEE/CVF 会议计算, 程序. Vis. 模式识别.*, 2020, pp. 1850-1859.

[43]肖安, 杨新, 陆生, 关东, 黄军, “基于卷积融合网络的大规模激光雷达点云分割方法”, *遥感学报*, vol 131, pp. 357 - 357, 2013.

[44]王伟, 黄家, “基于循环切片网络的点云三维分割”, *IEEE 配置, 计算, 进程. Vis. 模式识别.*, 2018, pp. 2626-2635.

[45]叶晓霞, 李军, 黄家, 杜丽丽, 张晓霞, “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”, *进程, Eur. 配置计算.* Vis., 2018, pp. 403-417.

[46]杨建军, 邹建军, 杨建军, “基于神经网络的三维 prnn 算法”, *IEEE, 进程. 配置计算.* Vis., 2017, pp. 900-909.

[47] y 赵, t . Birdal h • 邓和 f . Tombari “3 d 点胶囊网络” in *Proc.IEEE / CVFConf.Comput. Vis. Pattern Recognit.*, 2019, pp. 1009 - 1018.

[48]祁春荣, 李磊, 苏慧, gul., “基于点集的深度层次特征学习”, *进程, 神经网络学报. 系统.*, 2017, 第 1-10 页.

[49] Y.Wang, Y.Sun、Z.Liu S.E.Sarma, M.M.Bronstein and J.M。Solomon, “Dynamic graph CNN for learning on point clouds,” *ACM 译. 图.*, 卷三十八, no. 5, 2019 年第 1-12 页.

[50]张晓明, 张晓明, “基于卷积神经网络的三维语义分割”, *IEEE 配置, 计算, 进程. Vis. 模式识别.*, 2018, pp. 9224-9232.

[51]郭志强, 张志强, “基于卷积神经网络的四维时空卷积神经网络”, *进程, IEEE/CVF 配置, 计算.* Vis. 模式识别., 2019, pp. 3075-3084.

[52]唐辉, 刘志强, 李军, 林勇, 韩生, “基于点云的高效点云推理引擎”, *进程, 配置. 学习. 系统.*, 2022, pp. 302-315.

[53]唐洪等, “基于稀疏点体素卷积的高效三维结构搜索”, *进程, Eur. 配置计算.* Vis., 2020, pp. 685-702.

[54]顾军, 郭德明, 谢军, 齐传荣, L. gu, O. Litany, “基于点云的三维点云识别”, *程序, 欧洲. 配置计算.* Vis., 2020, pp. 574-591.

[55]王晓明, 张晓明, “基于卷积神经网络的生物医学图像分割”, *程序 国际版. 配置 Med. 图像计算. 计算. ——协助.《干预》*, 2015 年, 第 234-241 页.

[56] Y.Liu、B.Fan S.Xiang, 并且. 着潘, “用于点云分析的关系-形状卷积神经网络”, *IEEE/CVF 配置程序: 计算.* Vis. 模式识别., 2019, pp. 8895-8904.

[57]于晓霞, 唐亮, 饶勇, 黄涛, 周建军, 卢军, “基于点云模型的三维点云变形预训练”, *IEEE/CVF 配置程序, 计算.* Vis. 模式识别., 2022, pp. 19 313-19 322.

[58]张志强, “一幅图像值为 16x16 个单词的图像识别方法”, *《程序》, 国际版. 配置学习.《表征》*, 2021 年, 第 1-12 页.

[59]王志强, “注意力是你所需要的”, *《程序》,《神经网络》. 系统.*, 2017, 第 1-11 页.

[60]刘忠等, “Swin 变压器:基于移位窗口的分层视觉变压器”, *IEEE/CVF 国际版程序. 配置计算.* Vis., 2021, pp. 10012 - 10022.

[61]赵红红, 蒋丽丽, 贾建军, 赵洪辉, 张晓明, “点变压器”程序, *IEEE/CVF 国际版. 配置计算.* Vis., 2021, pp. 16 259-16 268.

[62]王晓明, 王晓明, “基于 vconvo - dae 的深度体积形状学习”, *程序, 欧洲. 配置计算.* Vis., 2016, pp. 236-250.

[63]王晓明, 王晓明, “面向对象的可预测生成向量表示”, *程序, 欧洲. 配置计算.* Vis., 2016, pp. 484-499.

[64]吴杰, 张志强, 张志强, “基于三维生成-对抗模型的物体形状概率潜在空间学习”, *《程序》, 第 30 期, 国际版. 配置. 神经网络. Process. 系统.*, 2016, 第 82-90 页.

[65]谢军, 郑志刚, 高仁, 王伟. 朱和吴旻, “用于三维形状综合和分析的学习描述符网络”, *IEEE 配置程序计算.* Vis. 模式识别., 2018, pp. 8629-8638.

[66]杨勇, 冯春春, 沈勇, 田东, “基于深度网格变形的点云自编码器”, *IEEE 配置, 计算, 程序.* Vis. 模式识别., 2018, pp. 206-215.

[67]李建军, 陈宝明, 李光辉, “基于自组织网络的点云分析”, *计算机工程学报. IEEE. conf . computer . vis . pattem recognition .*, 2018, pp. 9397-9406.

[68]张晓明, 张晓明, “三维点云的学习表征与生成模型”, *《程序》, 国际版. 配置. 学习.*, 2018, 第 40-49 页.

[69]王晓明, 王伟, 马吉, “多分辨率树网络在三维点云处理中的应用”, *程序, 欧洲. 配置计算.* Vis., 2018, pp. 103-118.

[70]韩志强, 尚明. 刘, 和 M. Zwicker, “视图间预测 GAN:通过学习全局形状记忆来支持局部视图预测的 3D 形状的无监督表示学习”, *AAAI 会议 Artif.进程. 智能.*, 2019, pp. 8376-8384.

[71]刘晓东, 韩志强, 温晓东. 刘, 和 M. Zwicker, “L2G 自编码器:通过分层自关注的局部到全局重建来理解点云”, *进程, 第 27 期 ACM Int. 会议. 多媒体*, 2019, pp. 989-997.

[72]韩志强, 王晓明, 王玉生. 刘和 M. Zwicker, “多角度点云-VAE:基于联合自重建和半对半预测的多角度三维点云无监督特征学习”, *IEEE/CVF Int 进程. 会议计算.* Vis., 2019, pp. 10 441-10 450.

[73]杨国光, 黄晓明, 郝振宇. 刘, S. Belongie 和 B. Hariharan, “PointFlow:三维点云生成与连续归一化流”, *IEEE/CVF Int 进程. 会议计算.* Vis., 2019, pp. 4541-4550.

[74]施勇, 徐明, 袁生, 方勇, “基于点分布学习的无监督深度形状描述符”, *IEEE/CVF 会议计算, 进程.* Vis. 模式识别., 2020, pp. 9353-9362.

[75]高欣, 胡伟, G.-J.:Qi, “GraphTERr:通过自动编码节点智能转换的图变换等变表示的无监督学习”, *IEEE/CVF 进程会议 计算.* Vis. 模式识别., 2020, pp. 7163-7172.

[76]李涛, 韩忠. 刘, “基于分层折叠的跳跃注意网络的点云补全”, *IEEE/CVF 会议进程, 计算.* Vis. 模式识别., 2020, 1939-1948 页.

[77]孙勇, 王勇, 刘志, “基于自关注的点云生成方法”, *IEEE/CVF 学报, 程序. 计算. Vis.*, 2020, 第 61-70 页。

[78]金, 李洪波, 李志强, 杨, 金, “无监督点云学习的渐进种子生成自编码器”, *IEEE/CVF Int 程序. 配置计算. Vis.*, 2021, pp. 6413-6422。

[79]庞勇, 王勇, 戴凤英, 刘伟, 田东, 袁磊, “点云自监督学习的自编码器”, *程序, 第 7 期. 配置计算. Vis.*, 2022, pp. 604-621。

[80]张等, “基于多尺度自编码器的点云预训练方法”, *程序, Adv. 神经网络. 系统.*, 2022, 第 1-12 页。

[81]张晓明, “基于自关联神经网络的非线性主成分分析”, *中国机械工程学报*, vol.37, no. 31. 2, pp. 233-243, 1991。

[82]自动编码器, “自动编码器-维基百科, 免费的百科全书”, 2022, 访问 2 月 16 日, 2022. (在线). 可用 <https://en.wikipedia.org/wiki/自动编码器>

[83]范宏, 苏宏, 李军, gu, “基于单图像的三维物体重建的点集生成网络”, *IEEE 配置, 计算, 程序. Vis. 模式识别.*, 2017, 第 605-613 页。

[84]王晓明, 王晓明, “动态路由”, *程序, Adv. 神经网络. 系统.*, 2017, 第 3859-3869 页。

[85]陈辉, 段正成, 杨, 李, 冯正成, 田东, “基于图拓扑推理和滤波的三维点云深度无监督学习”, *IEEE, 译. 图像的过程.*, vol. 29, pp. 3183-3198, 2020。

[86]陈辉, 高晓霞, 胡伟, “三维点云几何采样不变量表示的无监督学习”, *IEEE/CVF Int 程序. 配置计算. Vis. workshop*, 2021, pp. 893-903。

[87] I. Goodfellow et al., “生成对抗网络”, 在程序中, *ad. Neural Inf. Process. 系统.*, 2014 年。

[88]张志强. 李, M. Zaheer, Y.张, B. Póczos, R. Salakhutdinov, “点云 GAN”, 2018, *arXiv:1810.05795*。

[89]李, 李. D. Cohen-Or 和 p. a. Fu. 恒, “PU-GAN: 一个点云上采样对抗网络”, *IEEE/CVF Int 程序. 配置计算. Vis.*, 2019, pp. 7203-7212。

[90]张建军, 张建军, 张建军, 等. 基于点云自编码的神经网络采样器[j]. *计算机工程学报*, 2016, 32(4): 559 - 559。

[91]于磊, 李, 李长文. D. Cohen-Or 和 p. a. Fu. 恒, “PU-Net: 点云上采样网络”, *IEEE 配置程序, 计算. Vis. 模式识别.*, 2018, pp. 2790-2799。

[92]吴淑娟, 黄洪波, 吴志强, “基于图像拼接的三维点设置上采样”, *IEEE/CVF 配置计算, 程序. Vis. 模式识别.*, 2019, pp. 5958-5967。

[93]侯建军, 钱元, 何勇, “一种基于几何中心的三维点云采样网络”, *计算机工程学报. 施普林格*, 2020, pp. 752-769。

[94]钱国强, 李新, 张勇, “基于图像卷积网络的点云上采样方法”, *IEEE/CVF 配置 computing, 程序. 视觉模式识别.*, 2021, pp. 11 683-11 692。

[95]李新, 李新, 李鹏. 恒, 陈志伟. 傅, “通过解纠缠细化的点云上采样”, *IEEE/CVF 配置程序, 计算. 视觉模式识别.*, 2021, pp. 344-353。

[96]王志强, “基于点云的无监督多任务特征学习”, *程序, IEEE/CVF 计算机学报. 配置计算. 视觉*, 2019, pp. 8160-8171。

[97]张磊, 朱振民, “基于图像卷积神经网络的无监督特征学习的点云理解”, *IEEE 计算机学报, 程序. 配置 3D 视觉*, 2019, pp. 395-404。

[98]卢建军, 周建军, “三维点云的无监督表示学习的全局-局部双向推理”, *IEEE/CVF 配置程序, 计算. 视觉模式识别.*, 2020, pp. 5376-5385。

[99]张志刚, “基于信息交互最大化的三维物体表征学习”, *程序, 欧洲. 配置计算. 视觉*, 2020, pp. 626-642。

[100]张志强, “基于凸分解的点云标记高效学习”, *程序, 欧洲. 配置计算. 视觉*, 2020, 第 473-491 页。

[101]蒋涛, 徐宁, 金卫国, “基于方向估计的点云自监督学习”, *IEEE, 程序, 计算机学报. 配置 3D 视觉*, 2020, pp. 1018-1028。

[102]杜波, 高晓明, 胡伟, 李新, “基于硬负抽样的自监督点云学习的自对比学习”, *程序, 第 29 期, 计算机学报. 配置. 多媒体*, 2021, pp. 3133-3142。

[103]饶勇, 刘波, 魏勇, 鲁, 陈建军. Hsieh 和 J. Zhou, “随机房间: 用于 3D 对象检测的合成形状和随机布局的无监督预训练”, *IEEE/CVF 程序集成. 配置计算. 视觉*, 2021, pp. 3283-3292。

[104]张志强, 张志强. Misra, “三维特征共振任意点云的自我监督预训练”, 在 *proc. ieee / cvfint. conf. computer. 视觉*, 2021, pp. 10 252-10 263。

[105]陈等, “无监督点云识别的形状自校正”, *IEEE/CVF 程序集成. 配置计算. 视觉*, 2021, pp. 8382 - 8391。

[106]张晓明, 张晓明, 张晓明, “基于自然三维结构的点云预训练”, *程序, IEEE/CVF 配置, 计算. 视觉模式识别.*, 2022, pp. 21283 - 21293。

[107]陈毅, 戴毅, “基于动态对应的 3D 场景理解的对比学习”, *《程序》, 第 17 期, 欧洲. 配置计算. 视觉*, 2022, 第 543-560 页。

[108]李丽娟, “基于自监督的三维视觉预训练方法研究”, *《程序》, 第 17 期, 欧洲. 配置计算. 《视觉》*, 2022, 第 656-673 页。

[109]张建军, “基于无监督预训练的激光雷达三维目标检测方法”, *程序, 第 17 期, 欧洲. 配置计算. 《视觉》*, 2022, 第 17-33 页。

[110]蔡明, 李彦杰, 刘波, “基于点云的自监督学习的模糊识别”, *程序, 第 17 期, 欧洲. 配置计算. 视觉*, 2022, 第 657-675 页。

[111]刘波, 肖, 张伟, 鲁, 邵, “基于前景感知特征对比的三维表征学习”, *程序. IEEE/CVF Conf. 计算机学报. 视觉模式识别.*, 2023, pp. 1-11。

[112]张晓明, 张晓明, 张晓明, “一种三维曲面生成的学习方法”, *IEEE 配置, 计算, 程序. 视觉模式识别.*, 2018, pp. 216-224。

[113]袁文杰, 张晓明, 张晓明, “基于点补全的 PCN 网络”, *IEEE, 程序, 集成. 配置 3D 视觉*, 2018, pp. 728-737。

[114]黄振宇, 徐军, 倪峰, 李晓霞, “基于分形网络的三维点云补全”, *IEEE/CVF 配置程序, 计算. 视觉模式识别.*, 2020, pp. 7659-7667。

[115]刘波, 盛磊, 杨, 邵, 王绍明. 胡, “密集点云补全的变形和采样网络”, *AAAI 配置, 程序. 智能.*, 2020, 第 11 596-11 603 页。

[116]张伟, 闫庆, 肖, “基于分离特征聚合的点云补全”, *程序, 第 16 期欧洲. 配置计算. 视觉*, 2020, pp. 512-528。

[117]谢超, 王翀, 张伟, 杨辉, 陈, 温峰, “基于风格的点云补全方法”, *程序, IEEE/CVF Conf, 计算机学报. 视觉模式识别.*, 2021, 第 4619 - 4628 页。

[118]付科, 高平, 刘波, 张伟, 乔勇, 王翀, “pos - bert: 点云的单阶段 bert 预训练”, 2022, vol. 39(4): 998 - 998。

[119]陈晓明, 张晓明, 张晓明, “一种基于视觉表征的对比学习方法”, *《程序》, Int. 配置. 学习.*, 2020, 第 1597-1607 页。

[120]王晓明, 王晓明, 王晓明, “基于深度学习的深度学习研究方法”, *中文信息学报*, 2018, 32(4): 557 - 557。

[121]陈志强. 王, y - q. 杨. Q - F. 邹, 吴之, 刘彦, 童新, “基于多分辨率实例识别的形状分析无监督 3D 学习”, *AAAI 配置, 程序. 智能.*, 2021, 第 2773-2781 页。

[122]蒋建军, 卢晓霞, 欧阳伟, 王, “基于无监督表示学习的三维点云数据学习”, 2018, *arXiv: 210.06632*。

[123]侯杰, 谢世生, 戴晓明, “三维先验在三维表征学习中的应用”。, *《IEEE/CVF Int 程序》. 配置计算. Vis.*, 2021, pp. 5693-5702。

[124]王明明, “基于 K-means 聚类算法的 as136 算法”, *中国科学: 自然科学 Stat. 社会. 爵士. C. 中央集权.*, 第 28 卷, 第 6 期. 1, 页 100-108, 1979 年。

[125]张晓明, 张晓明, 张晓明, “基于深度聚类的非监督学习视觉特征”, *《程序》. 配置计算. Vis.*, 2018, pp. 132-149。

[126]张晓明, 张晓明, 张晓明, “基于点云的深度学习预测方法”, *《计算机工程学报》*, 2019, 32(4): 557 - 557。

[127]张晓明, 张晓明, “基于无监督表示学习的图像旋转预测”, *《科学进展》, Int. 配置学习. 《表征》*, 2018 年, 第 1-14 页。

[128]张晓明, 张晓明, “三维点云的局部特征自监督学习”, *IEEE/CVF 配置程序, 计算. Vis. 模式识别. workshop*, 2020, pp. 938-939。

[129]孙晨, 郑志强, 王晓, 徐明, 杨, “基于混合和解纠缠的点云预训练方法”, 2013.vol . 39(4): 559 - 559。

[130]张志强, “基于语义的激光雷达场景识别方法”, *IEEE/CVF 程序集成。配置计算。视觉*, 2019.pp. 9297-9307。

[131]肖安, 黄建军, 关东, 詹峰, 路生, “基于合成到真实激光雷达点云的语义分割”, *中国计算机工程学报*, 程序, *配置, 智能*, 第 36 卷, no. 第 3 期, 2022, 第 2795-2803 页。

[132] C.Feichtenhofer, H.Fan、B.Xiong R.Girshick and K.他, “无监督时空表征学习的大规模研究”, *IEEE/CVF 配置, 计算, 程序。视觉模式识别*, 2021, pp. 3299-3309。

[133]宋旭等, “动作识别的时空对比域自适应”, *IEEE/CVF 配置, 程序, 计算。视觉模式识别*, 2021, 第 9787-9795 页。

[134]胡凯, 邵建军, 刘彦杰, 沈勇, “基于视频自监督学习的对比和顺序表示”, *IEEE/CVF 集成, 程序。配置计算。视觉*, 2021, pp. 7939-7949。

[135]王志强, “基于全局语境的视频对比学习”, 程序, *IEEE/CVF 集成。配置计算。视觉*, 2021,pp. 3195-3204。

[136]陈德志, 张爱祥, 陈志强, “基于自然语言的 RGB-D 扫描三维目标定位”, 程序, *Eur. 配置计算。视觉*, 2020, pp. 202-221。

[137] M. Afham, D. Dissanayake, A. Dharmasiri, K. Thilakaratna, R. Rodrigo, “基于自监督的三维点云交叉模态对比学习”, *IEEE/CVF 配置计算程序。视觉模式识别*, 2022, 第 9902-9912 页。

[138]王志, 余晓霞, 周建军, 路生, “基于点对像素提示的点云图像模型的优化”, *计算机工程学报*, 程序。系统, 2022, 第 1-17 页。

[139]邓海涛, 张建军, “基于无监督学习的三维局部描述子的 PPF-foldNet”, 程序学习, *Eur. 配置计算。视觉*, 2018,pp. 602-618。

[140]陈晓明, 宋松生, 陈晓明, “基于 RGB-D 重构的局部几何描述符学习方法”, *中国计算机工程学报*, 程序, *IEEE 配置 computer。视觉模式识别*, 2017, pp. 1802-1811。

[141]曾翼, 钱, 朱正杰, 侯军, 袁红华, 何勇, “三维点云密集匹配的无监督端到端学习”, 程序, *IEEE/CVF Conf computer。视觉模式识别*, 2021, pp. 6052-6061。

[142]张晓明, 张晓明, “基于无监督深度点匹配的交叉和自构造方法”, *IEEE 集成, 程序。配置 3D 视觉*, 2021,pp. 1442-1451。

[143]沈勇, 蒋辉, 谢军, 李军, 钱, 杨, “基于采样网络的无监督点云配准方法”, *IEEE/CVF 程序。配置计算。视觉*, 2021,pp. 6128-6137。

[144]赵红红, 蒋丽丽, 贾建军, 赵洪辉, 张晓明, "点变压器"程序, *IEEE/CVF vol. 11 - vol. 11 配置计算。Vis.*, 2021, pp. 16 259-16 268。

[145]王晓明, 王晓明, “三维形状描述子的旋转不变性球谐表示”, 《程序》, 第 2 卷。*几何图形的过程*, 2003, pp. 156-164。

[146]李建平. 陈晓平. 田, y - t. 沈, M. ouhyung, “On visual similarity based 3D model retrieval”, *计算机图形论坛*. Hoboken, NJ, USA: Wiley, 2003, pp. 223-232。

[147]刘建军, 张建军, 张建军, “基于离散生成模型的三维点云自监督学习”, *IEEE/CVF 配置, 程序。计算。Vis. 模式识别*, 2021,pp. 8248-8257。

[148]王晓明, “基于 tsne 的数据可视化”, *中国科学技术研究院。学习。Res.*, 第 9 卷, no. 86, pp. 2579-2605, 2008。

[149]李斌, 闫燕, “稀疏嵌入卷积检测”, vol. 18, no. 14. 2018 年第 10 期, 艺术。不。s3337。

[150]张晓明, 张晓明, 张晓明, “一种基于非监督学习的视觉特征识别方法”, *程序学报, 中文信息学报。系统*, 第 9912-9924 页, 2020。



肖安然 分别于 2016 年和 2019 年获得中国武汉大学理学学士和理学硕士学位。他目前在新加坡南洋理工大学计算机科学与工程学院攻读博士学位。他的研究方向包括点云处理、计算机视觉和遥感。



2010 年毕业于中国中国科学院自动化研究所模式识别国家实验室，获模式识别与智能系统博士学位。现为中国温州大学教授。他在国际和国内期刊和国际会议上发表了 100 多篇论文，包括 *IEEE Transactions on Pattern*

《分析与机器智能》、《国际计算机视觉杂志》、《IEEE 神经网络与学习系统学报》、《IEEE 神经网络与学习系统学报》、《IEEE 神经网络与学习系统学报》、ICCV、CVPR、NIPS、IJCAI、AAAI 等。主要研究方向包括模式识别、计算机视觉和机器学习。

黄嘉兴 毕业于英国格拉斯哥大学电子与电气工程学士学位，以及新加坡南洋理工大学(NTU)电子与电气工程硕士学位。他目前是 NTU 新加坡计算机科学与工程学院的研究助理，目前正在攻读博士学位。他的研究兴趣包括计算机视觉和机器学习。



卢世坚，新加坡国立大学电气与计算机工程专业博士。他是新加坡南洋理工大学计算机科学与工程学院的副教授。主要研究方向包括图像与视频分析、视觉智能、机器学习。



关大燕，2019 年 9 月毕业于中国浙江大学，获博士学位。他目前是阿拉伯联合酋长国穆罕默德·本·扎耶德人工智能大学的研究科学家。在此之前，他于 2019 年 11 月至 2022 年 3 月在南洋理工大学担任研究员。主要研究方向为计算机视觉、模式识别、深度学习。



邵凌，IEEE 院士，中国科学院大学 ucas 终端人工智能实验室特聘教授。他是阿联酋阿布扎比盗梦人工智能研究所的创始人、首席执行官和首席科学家。他的研究兴趣包括计算机视觉、深度学习、医学成像和视觉与语言。他是 IAPR、BCS 和 IET 的研究员。



张晓琴，IEEE 高级会员，2005 年毕业于中国中南大学电子信息科学与技术专业，获理学学士学位。