中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2023)06-1741-26

论文引用格式:Gong J Y, Lou Y J, Liu F Q, Zhang Z W, Chen H M, Zhang Z Z, Tan X, Xie Y and Ma L Z. 2023. Scene point cloud understanding and reconstruction technologies in 3D space. Journal of Image and Graphics, 28(06):1741-1766(龚靖渝, 楼雨京, 柳奉奇, 张志伟, 陈豪明, 张志忠, 谭鑫, 谢源, 马利庄. 2023. 三维场景点云理解与重建技术. 中国图象图形学报, 28(06):1741-1766)[DOI:10.11834/jig.230004]

三维场景点云理解与重建技术

龚靖渝[1],楼雨京[1],柳奉奇[1],张志伟[1],陈豪明[2],张志忠[2],
谭鑫[2],谢源[2],马利庄[1,2*]

1. 上海交通大学计算机科学与工程系,上海　200240; 2. 华东师范大学计算机科学与技术学院,上海　200062

摘　要:3维场景理解与重建技术能够使计算机对真实场景进行高精度复现并引导机器以3维空间的思维理解整个真实世界,从而使机器拥有足够智能参与到真实世界的生产与建设,并能通过场景的模拟为人类的决策和生活提供服务。3维场景理解与重建技术主要包含场景点云特征提取、扫描点云配准与融合、场景理解与语义分割、扫描物体点云补全与细粒度重建等,在处理真实扫描场景时,受到扫描设备、角度、距离以及场景复杂程度的影响,对技术的精准度和稳定性提出了更高的要求,相关的技术也十分具有挑战性。其中,原始扫描点云特征提取与配准融合旨在将同场景下多个扫描区域进行特征匹配,从而融合得到完整的场景点云,是理解与重建技术的基石;场景点云的理解与语义分割的目的在于对场景模型进行整体感知并根据语义特征划分为功能性物体甚至是部件的点云,是整套技术的核心组成部分;后续的物体点云细粒度补全主要研究扫描物体的结构恢复和残缺部分补全,是场景物体点云细粒度重建的关键性技术。本文围绕上述系列技术,详细分析了基于3维点云的场景理解与重建技术相关的应用领域和研究方向,归结总结了国内外的前沿进展与研究成果,对未来的研究方向和技术发展进行了展望。

关键词:3维场景;点云融合;场景分割;物体形状补全;深度学习

Scene point cloud understanding and reconstruction technologies in 3D space

Gong Jingyu[1], Lou Yujing[1], Liu Fengqi[1], Zhang Zhiwei[1], Chen Haoming[2], Zhang Zhizhong[2], Tan Xin[2], Xie Yuan[2], Ma Lizhuang[1,2*]

1. Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai　200240, China;
2. School of Computer Science and Technology, East China Normal University, Shanghai　200062, China

Abstract: 3D scene understanding and reconstruction are essential for machine vision and intelligence, which aim to reconstruct completed models of real scenes from multiple scene scans and understand the semantic meanings of each functional component in the scene. This technique is indispensable for real world digitalization and simulation, which can be widely used in related domains like robots, navigation system and virtual tourism. Its key challenges are required to be resolved on the three aspects: 1) to recognize the same area in multiple real scans and fuse all the scans into an integrated scene point cloud; 2) to make sense of the whole scene and recognize the semantics of multiple functional components; 3)

收稿日期:2023-01-03;**修回日期**:2023-02-21;**预印本日期**:2023-02-28

*通信作者:马利庄　ma-lz@cs.sjtu.edu.cn

基金项目:国家自然科学基金项目(61972157,72192821);上海市科技创新行动计划人工智能科技支撑项目(21511101200)

Supported by: National Natural Science Foundation of China(61972157, 72192821); Shanghai Municipal Science and Technology Commission(21511101200)

to complete the missing region in the original point cloud caused by occlusion during scanning. It is necessary to extract point cloud feature in order to fuse multiple real scene scans into an integrated point cloud, which can be invariant to scanning position and rotation. Thus, intrinsic geometry features like point distance and singular value in neighborhood covariance matrix are often involved in rotation-invariant feature design. Contrastive learning scheme is usually taken to help the learned features from the same area to be close to each other, while extracted features from different areas to be far away. To get generalization ability better, data augmentation of scanned point cloud can also be used during feature learning process. Features-learnt pose estimation of scanning device can be configured to calculate the transformation matrix between point cloud pairs. After the transformation relationship is sorted out, the following point cloud fusion can be implemented using the raw point cloud scans. To further understand raw point cloud-based whole scene and segment the whole scene into functional parts on the basis of multiple semantics, an effective and efficient network with appropriate 3D convolution operation is required to parse entire points-based scene hierarchically, and specific learning schemes are necessary as well to adapt to various situation. The definition and formulation of basic convolution operation in 3D space is recognized as the core of pattern recognition for 3D scene point cloud. It is highly correlated to the approximated convolution kernel in 3D space where feature extraction can be developed in terms of appropriate point cloud grouping and down/up-sampling. The discrete approximation of 3D continuous convolution pursues being capable of recognizing various geometry pattern while keeping as few parameters as possible. Network design based on these elementary 3D convolution operations is also a fundamental part of outstanding scene parsing. Furthermore, point-level semantic segmentation of scanned scene can be linked mutually in relevance to such aspects of boundary detection, instance segmentation, and scene coloring, where network parameters are supervised through more auxiliary regularization. Semi-supervised methods and weak-supervised methods are required to overcome the lack of data annotation for real data. The segmentation results and semantic hints can be used to strengthen the fine-grained completion of object point cloud from scanned scene, in which the segmented objects can be handled separately, and semantics can be used to provide the structure and geometry prior when occlusion-derived missing region is completed. For the learning of object point cloud completion, it is crucial to learn a compact latent code space to represent all the complete shapes and design versatile decoder to reconstruct the structure and fine-grained geometry details of object point cloud. The learnt latent code space should contain complete shapes as much as possible, thus requiring large-scale synthetic model dataset for training to ensure the generalization ability. The encoder should be designed to recognize the structure of original point cloud and extract specific geometry pattern which preserves this information in latent code, while the decoder is used to recover the overall skeleton of original scanned objects and complete all the details according to the existing local geometry hints. For real scanned object completion, it is required to optimize the integration of latent code space further for synthetic models and real scanned point cloud. A cross-domain learning scheme is used to apply the knowledge of completion to real object scans, whereas the details of real scanned object can be preserved in the completed version. We analyze the current situation about scene understanding and reconstruction, including point cloud fusion, 3D convolution operation, entire scene segmentation, and fine-grained object completion. We analyze the frontier technologies and predict promising future research trends. It is significant for the following research to pay more attention on more open space with further challenges on computing efficiency, handling out-of-domain knowledge, and more complex situation with human-scene interaction. The 3D scene understanding and reconstruction technology will help the machine to understand the real world in a more natural way which can facilitate such various application domains like robots and navigation. It also potential to conduct plausible simulation of real world based on the reconstruction and parsing of real scenes, making it a useful tool in making various decisions.

Key words: 3D scenes; point could fusion; scene segmentation; object shape completion; deep learning

0 引 言

3维场景模型是真实世界在计算机中进行数字化后的具体表征方式,对3维场景模型的研究不仅能够使机器模仿人类通过3维空间的思维来理解周围环境,更能够使机器以智能体的方式参与到真实3维世界的工业生产、城市与交通规划以及与人体

的交互中。基于3维场景模型研究的核心在于对3维场景的语义理解以及细粒度重建，而点云作为采集设备通用的3维数据形式，常被主流的工作用于表征3维场景模型进行相关研究。

场景点云通常可以通过色彩深度（RGB-depth，RGB-D）相机、激光雷达等设备对场景进行扫描来获得。除此之外，室内场景点云也可以利用即时定位与地图构建（simultaneous localization and mapping，SLAM）的技术（Hosseinzadeh 等，2019），通过相机拍摄的相邻帧间的图像估计相机运动，并恢复场景的空间结构来得到。但是通过扫描方法得到的原始点云往往并不完整，需要后续的处理，而后续的扫描点云特征提取与融合主要包括对扫描的原始点云进行点级别的几何特征提取，以及根据点的特征进行配准从而完成点云的融合。其中的点云配准是从扫描3维数据到完整点云场景模型的核心技术模块（李建微和占家旺，2022）。3维场景扫描与配准系列技术可以广泛应用于真实场景的3维建模以及虚拟混合现实等信息化生产与数字娱乐的应用中。针对不同点云提取特征的主要挑战在于探索局部点云几何特征的平移旋转不变性，找到不同扫描数据中的匹配区域。然而由于扫描设备扫描角度距离存在差异，同时受到离群噪声点的影响，同一区域的点云也有不同，这会大幅提升特征提取与匹配的难度。

3维场景语义理解的目的是根据语义信息识别场景中不同功能的物体，从而对整个场景进行物体甚至是部件级别的划分。对场景点云进行语义分割的技术也能直接在机器人与场景物体的交互以及自动驾驶这些场景中得到很好的运用。这个任务所包含的场景特征识别、网络结构设计、多任务协同以及面对极少标注样本时的应对技术也都是国内外的研究热点。然而，3维点云的结构不规则性、不同扫描设备以及距离角度导致的不均匀性使得鲁棒的3维特征提取变得十分困难，而对不同3维场景进行精准的语义分割甚至是实例分割也成为了一项十分具有挑战性的任务（龙霄潇 等，2021）。

物体扫描点云补全的核心作用在于针对遮挡所导致的点云残缺问题利用已有的大型完整点云数据库学习完整点云的先验知识，从而将残缺的物体点云修复为完整的点云。该类方法可有效修复物体扫描时出现的残缺，同时能够在机器人应用中对不可视区域做出合理推理。针对大型合成3维模型数据集以及真实扫描物体的点云补全受到了国内外的广泛关注，吸引了众多学者。此类技术重点研究点云编解码方式以及跨域跨数据集之间的统一特征学习方法。由于点云本身的非结构化表征方式，仍然缺乏细粒度点云解码与重建的方法。针对无完整点云的扫描数据，也很难仅凭借合成的完整模型数据集进行扫描点云补全。

1 点云特征提取与匹配

随着深度学习在2维图像上的广泛应用，其在3维数据上的拓展取得了不错的成就。3维数据有多种表示方式，例如体素、网格以及点云等方式。传统的深度学习框架得益于2维卷积架构。结合现代并行计算硬件，卷积操作能够高效地处理规则的数据结构，但图像缺失的深度信息往往会导致语义歧义性，特别是在极端光照（Tan 等，2021）或特殊光路（Tan 等，2023）的条件下。作为在3维数据的拓展，3维卷积应运而生，能够自然地处理规则化的体素数据。然而，相较于2维图像，处理体素这种表示方式需要的计算资源呈指数级增长。并且，3维结构是稀疏的，这导致体素这一类表示方式会造成大量的计算资源浪费。面对大场景分析任务时，体素将不再适合。相反，点云这种无规则表征能够简单有效地表示稀疏的3维数据结构，在3维场景理解任务中发挥了重要作用。因此，针对点云的特征提取是面向3维场景分析流程中的重要一环。点云特征提取的技术取得了前所未有的发展。本节围绕传统点云特征提取、深度学习在点云上的初应用、点云卷积、稀疏卷积和点云 Transformer 介绍点云特征提取的相关研究以及点云特征提取在点云匹配任务下的应用。

1.1 传统点云特征提取

传统点云特征提取借助3维点云的局部几何信息进行编码生成几何算子，作为点云局部几何特征。一个好的3维算子具有一些优秀的性质，如可描述性、紧密性和鲁棒性等。其中，可描述性以及鲁棒性被认为是3维局部特征算子最重要的属性。算子是可描述的是指其能够封装3维表面中的主导信息内容。换句话说，算子能够提供充足的可描述内容来区分两个不同的表面。算子的鲁棒性是指其对于模型引入的噪声和变化不敏感。在过去几十年的发展

中，研究人员提出了针对不同特性的3维几何算子。大多数3维局部特征算子都是对局部3维表面的几何信息进行编码。在这些算子中，一部分利用局部几何统计量来表示局部表面不同的性质。具体来说，通过累计特定域（例如点坐标、几何属性）中几何的或拓扑的量化值（例如点的数量）构建统计直方图，用于表示几何特征。这些方法依据统计的类型可以分为空间分布统计算子和几何属性统计算子。

基于空间分布统计算子统计了局部区域内点云分布状态。自旋图像（spin image，SI）算法（Johnson和Hebert，1999）利用给定关键点与其法向量构建局部参考坐标轴，并记录局部区域中任意点到关键点切平面内、外的距离作为算子的统计量。3维形状上下文特征（3D shape context，3DSC）方法（Frome等，2004）同样构建参考坐标轴。不同的是，其将局部空间划分为3维球形网格，通过统计每一个网格中的点数量作为该区域的算子。唯一形状上下文特征（unique shape context，USC）算法（Tombari等，2010）作为3DSC的一个拓展，通过构建局部坐标参考系，锁定了参考坐标轴存在的绕轴旋转的自由度，从而排除了算子歧义性。旋转投影统计量（rotational projection statistics，RoPS）构建局部参考系（Guo等，2013）。针对每一坐标轴，RoPS都将点云绕轴旋转多个离散角度值，并统计点云沿坐标轴的分布图来得到最终算子。

基于几何属性统计算子计算局部表面上点的几何属性（例如法向量，曲率）统计直方图来表示特征。局部表面补丁（local surface patches，LSP）算法（Chen和Bhanu，2007）通过统计区域内每一点和关键点法向量夹角的余弦值来表征几何特征。Thrift算法（Flint等，2007）根据与关键点之间的偏移角进行划分，统计不同偏移角度区间点分布情况作为几何特征。持久特征直方图（persistent feature histograms，PFH）算法（Rusu等，2008）依据局部区域任意两点构成的点对的表面法向量来构建Darboux参考系，通过统计参考系中的距离角度信息作为局部几何特征。快速点特征直方图（fast point feature histograms，FPFH）算法（Rusu等，2009）作为PFH的改进，仅采用中心点与区域中任一点构成的点对计算特征值，降低了计算复杂度。方向直方图特征（signature of histograms of orientations，SHOT）算法（Salti等，2014）首先构建局部参考系，并将局部空间根据半径、方位角以及仰角划分为球形网格，统计了每一个网格中点法向量分布，构成最终的算子。点对特征（point pair feature，PPF）方法（Drost等，2010）依据计算任意点对的距离、法向量之间的夹角以及两点连线与法向量之间的夹角构成4维特征来表示几何结构。

1.2 点云深度学习

点云特征提取的先驱是Qi等人（2017a）提出的PointNet。点云数据由于其离散以及不规则性，传统需要权重共享的卷积操作无法直接应用到点云数据上。传统的研究方法将点云转换到对应的3维体素网格或多视角下的图像数据，从而可以间接使用卷积操作构建深度网络结构，提取特征。然而，这种方式会生成庞大而不必要的冗余数据，并引入了许多量化计算，会改变数据原本包含的信息内容。其实，点云本身是一种简单统一的表示方式，直接从点云提取特征可以避免不必要以及不规则的组合计算，又可以降低3维结构表征的复杂度。PointNet是一个统一的点云处理架构，直接以点云数据的3维坐标作为输入，可以预测完整点云的类别标签用于点云分类任务，还能够输出逐点的语义标签用于物体部件分割以及场景语意分割等任务。PointNet方法的关键技术是利用一个简单的对称函数max-pooling，使模型网络能够有效学习到一组优化指标。这些指标可以挑选出表示完整点云信息的关键特征。同时，对称函数可以确保输出的结果与输入点的排列顺序无关。PointNet最后的全连接层将这些学习优化后的特征值汇聚到一个全局的描述子中用于表示整个点云，可以进一步用于预测逐点的语义标签。点云的另一个优势是模型可以轻易地对其进行刚体或仿射变化，因为每一个点的变换是独立的。于是，PointNet引入一个独立于数据的空间变换网络（spatial transformer network，STN），使PointNet开始处理输入数据之前，先将输入数据标准化，从而进一步提升实验结果。研究人员对于处理点云数据对称函数也展开了许多相关研究（Ravanbakhsh等，2017；Zaheer等，2017；Li等，2018a）。

PointNet开创了点云特征提取的先河，学习得到每一个输入点的空间编码，然后将各个单独的点汇总成一个全局点云标志。PointNet设计的全局对称函数造成其无法捕获局部的结构信息。然而，对于局部信息的探索被证实为卷积神经网络的重要成

功因素。一个标准卷积神经网络可以在逐渐增加的尺度上不断地提取特征，从而形成一个多尺度的分层架构来获取不同分辨率下的局部特征。在低层的神经一般具有较小的感受野，在高层的则具有更大的感受野。为了点云特征提取结构也能够继承2维卷积神经网络的特点，获取局部几何信息，Qi等人（2017b）在PointNet基础上进一步提出了分层结构的PointNet++，首先利用最远点采样（farthest point sampling，FPS）将输入的点云根据距离标准划分为若干互相重叠的球形局部区域。与卷积神经网络相似，每一个小的局部区域都会用PointNet提取特征，作为细粒度的局部几何结构表征，同时不同区域可以贡献特征提取的权重。类似的局部特征会聚集组合到一个更大的几何单元中，从而处理得到更高层的特征。该步骤会不断重复，直至得到完整点云的特征。PointNet++最显著的贡献在于其利用在不同尺度下的邻域几何信息来实现鲁棒的细粒度特征提取。

1.3 点云卷积

PointNet++提供了分层和多尺度提取点云局部特征的范式。不过与2维卷积操作相比，其特征提取方式与2维卷积操作仍存在差异。传统卷积操作针对邻域中不同区域赋予了相互独立的权重用于区分各自的相对位置。PointNet++对局部邻域中的每一个点都赋予相同的权重进行特征提取，未区分各自点的在邻域中的相对位置。后续研究均利用该信息进一步改进，并提出了点云的卷积操作。

Li等人（2018b）提出了点卷积神经网络（point convolutional neural network，PointCNN），通过在点云上的卷积操作实现了点云卷积神经网络架构的搭建。2维卷积依据每一个像素在规则的局部网格中的位置，按顺序赋予权重进行卷积操作。与2维规则的网格数据不同，3维数据局部邻域中点云的空间分布是不规则的，其排序方式有多种方式。根据不同的排序方式，点云卷积得到的结果往往是不相等的。因此，确定邻域中点的顺序使之与权重顺序相对应是PointCNN解决的一个难题。PointCNN提出了χ-卷积操作。χ-卷积首先依据输入邻域中心点，并将邻域中点相对中心的位置用多层感知器进行编码。结合位置编码以及输入点特征，再次利用多层感知器生成χ-变换矩阵，将邻域中点依据变换矩阵进行排序使之产生固定的顺序，从而赋予对应位置下的卷积权重。χ-卷积操作确保了点云卷积结果不受输入点排列顺序改变的影响。PointCNN最早实现了2维卷积到点云卷积的拓展，为点云特征提取打下了良好的基石。后续研究工作提出了各种各样点云卷积的变体。SpiderCNN（Xu等，2018）利用邻域中点的测地信息以及三线性插值方式来生成给定邻域中的滤波器，依据该滤波器便可对邻域点进行卷积操作。Hua等人（2018）设计了逐点的卷积操作，通过核支持区域划分局部空间用于卷积操作。动态图卷积神经网络（dynamic graph convolution neural network，DGCNN）算法（Wang等，2019）利用邻域点中心点位置和邻域点的相对位置生成对应点特征每一维的权重来实现卷积操作。形状关系卷积神经网络（relation-shape convolutional neural network，RSCNN）算法（Liu等，2019b）也利用邻域中点的相对位置生成权重来实现卷积操作。

不同于利用多层感知器生成卷积权重的方法，核心点卷积（kernel point convolution，KPConv）算法（Thomas等，2019）是一种全新的点云卷积运算。KPConv的灵感同样来自于基于2维图像的卷积。类似2维卷积使用像素网格作为卷积核，KPConv定义了一系列固定位置的核点用于卷积操作。卷积的权重分别由这些核点生成。每一个核点所辐射的空间根据相关函数来确定。输入的点通过寻找与其相邻最近的核点，利用其核点对应的权重以及到核点的距离的计算结果作为卷积操作的输出值。值得注意的是，邻域中核点的数量是不确定的，使得KPConv能够灵活地适应不同的输入点云，并且不受输入点云密度变化的影响。在此基础上，KPConv还拓展出可形变的形式。对于每一个核点，网络可以针对每一个卷积区域生成对应的偏移向量去改变核点的空间位置，使其更好地适应输入点云结构。KPConv可以构建出非常深的网络架构，同时保持快速的训练以及推理时间。

将点云构建成图结构，在图上进行卷积操作也是提取点云特征的一种方式。这种方式与直接在点云上卷积具有一定的相似性。在图上的卷积可以使用在其光谱表示上的乘法来实现（Defferrard等，2016；Yi等，2017），也可以利用在点云表面上所构建得到的图来实现（Masci等，2015；Bronstein等，2017；Simonovsky和Komodakis，2017；Monti等，2017）。

1.4 稀疏卷积

由于3维数据的稀疏性，完整的3维物体体素表示并不适合作为3维特征提取的输入。借鉴点云离散表示3维场景表面的方式，将点云进一步转化为稀疏体素，并利用稀疏卷积网络来提取特征成为研究的热点。稀疏卷积网络仅对空间中非空的体素进行卷积操作，从而避免了传统3维卷积在非空体素上浪费大量的计算和存储资源。

Graham等人（2018）率先提出了稀疏卷积的概念，并设计了子流形稀疏卷积和网络框架来处理稀疏的3维场景数据。Choy等人（2019a）提出了稀疏卷积网络架构Minkowski Engine。对于传统语音、文字以及图像数据，特征往往是稠密地提取。然而对于3维扫描数据，甚至在更高维度的空间中，这种稠密的表示效率十分低。原因是数据在高维空间的分布往往是稀疏的。因此，Choy等人（2019a）认为可以仅保存空间中非空的部分作为其坐标以及关联的特征，即稀疏矩阵在高维空间的拓展，名为稀疏张量。在稀疏张量上卷积的定义与传统卷积操作一样，仅需要给出卷积步长、稀疏张量坐标以及点云转换成稀疏张量时的步长。实现稀疏卷积最重要的步骤就是确定输入张量和输出张量的映射。不同于2维卷积输出坐标容易计算得到，稀疏张量是点任意聚集在一起的。因此，稀疏卷积需要给出输入张量到输出的映射来实现输入和卷积核的卷积操作，该映射定义为核映射。最后，给定核映射、权重以及输入输出坐标即可实现稀疏卷积操作，从而对稀疏体素进行特征提取。

Liu等人（2019c）提出点—体素卷积（point-voxel convolution，PVC），同时在点云和体素两种表示下提取特征。PVC利用点云表示输入3维数据来减少内存消耗，同时利用体素表示减少组合不规则数据带来的不必要的计算浪费。对于点云处理分支，PVC对每一个点进行单独处理；对于体素处理分支，PVC对体素化的点进行卷积处理。尽管基于PVC的神经网络（PVC neural network，PVCNN）能够处理大体积的体素数据。单个体素包含大范围的实际区域，但是PVCNN对于小个体（例如行人）的识别能力很差。因为小个体仅占用了少量的体素从而增强了识别难度。一种解决方案是将大范围场景用滑动窗口划分为不同的子区域，在子区域进行特征提取，然而子区域划分的操作并不适用于实时的应用。针对PVCNN的缺陷，Tang等人（2020）在PVCNN的基础上提出了稀疏的点—体素卷积（sparse point-voxel convolution，SPVC）。对于点云分支，SPVC依旧保持高精度的特征提取。而对于体素分支，SPVC则借鉴稀疏卷积，在不同尺度进行特征提取。两个分支之间的信息传递所需要的资源是可以忽略的。

1.5 点云Transformer

基于自注意力机制的Transformer（Vaswani等，2017）网络结构在自然语言处理任务上引发了巨大变革，确立了大模型在自然语言处理各项任务上的领先地位。与此同时，自注意力机制在图像分析任务上的拓展也取得了不错的成就。参考Transformer在自然语言处理和图像分析领域上的成功，研究人员展开了在点云数据处理上的Transformer拓展。

Zhao等人（2021）和Guo等人（2021）提出了用于点云特征提取的Transformer架构。Guo等人（2021）所设计的点云Transformer架构将自注意力机制应用到全局的点云上，即输入点云任意点之间均计算关联度。这种全局方式受限于内存和计算资源，只能应用在点数量较少的单个物体或小场景，而无法处理大场景点云数据。Zhao等人（2020）基于向量注意力机制实现了针对局部点的Transformer架构。向量注意力机制主要计算给定点与其相邻点之间的关联度，从而对每个点均复用该权重。关注局部信息的Transformer显著降低了内存资源的占用。不足的是，基于向量注意力机制网络结构的参数量随着深度的增加而大幅度增加，将导致严重的过拟合以及深度限制问题。并且，点云的坐标位置相对于2维图像的像素位置提供了更复杂的几何信息，对点云特征提取至关重要。传统的用于图像Transformer的位置编码不再适用于点云数据。针对以上问题，Wu等人（2022）提出改进版本的点云Transformer（point transformer V2，PTv2），利用分组向量注意力机制有效降低了模型参数量，同时设计了专门针对3维点云的位置编码机制，提升了模型框架对点之间的几何关联的敏感程度。

1.6 点云旋转不变特征提取

上述点云特征提取方法与2维卷积相似，仅具有平移不变性。但是对于3维点云，其在现实空间中会处在不同的姿态之下。同时旋转变换会给点云特征提取带来一定程度的影响。因此，许多工作专门针对提取点云旋转不变特征展开研究。点云旋转

不变特征提取大致分为3类。第1类利用旋转不变几何特征作为模型的输入，代替受旋转变换影响的坐标输入；第2类寻找表示点云旋转不变的局部参考系来避免旋转变化带来的影响；第3类则是估计输入点云的姿态并将其调整到标准姿态再提取特征。

旋转不变卷积（rotation invariant convolution，RIConv）算法（Zhang等，2019）、ClusterNet（Chen等，2019）和排序Gram矩阵网络（sorted gram matrix，SGMNet）算法（Xu等，2021a）通过计算输入点云点之间的相对距离和角度作为特征来代替点坐标作为网络结构的输入。由于旋转变换为刚体变换，在整体点云经过旋转后，点云内部几何仍旧保持相对不变。因此局部几何中点之间的相对距离以及角度等信息可以作为低层旋转不变特征，从而利用神经网络进一步提取高层特征。然而，在将点坐标转换为这些底层特征的过程中伴随着重要几何信息的损失，所以这类方法面临不同程度的结果下降。集成位置关系特征的旋转不变网络（positional & relational feature embedding block-based rotation-invariant network，PR-invNet）（Yu等，2020）方法和Li等人（2021a）提出的方法首先利用主成分分析（principal component analysis，PCA）选取最代表点云几何结构的3个坐标轴作为参考系表示点云的标准姿态。但是PCA存在歧义性，点云的标准姿态并不唯一。因此，这类方法利用固定数量的旋转增强构建一个姿态空间来涵盖所有存在歧义的标准姿态，并利用姿态选择器挑选一个最终姿态表示该点云的旋转不变表示。旋转不变图卷积网络（rotation invariant graph convolution network，RI-GCN）算法（Kim等，2020）和边缘对齐卷积神经网络（aligned edge convolutional neural network，AECNN）算法（Zhang等，2020a）则设计不同的局部参考系提取局部的旋转不变特征，最终汇聚得到全局的旋转不变特征。RI-GCN利用PCA构建局部邻域点对应的参考系，而AECNN则利用局部邻域中心点以及请求点之间的相对位置构建局部参考系。局部参考系之所以能够作为点云的旋转不变特征，是因为旋转变换不改变点云的局部几何结构。

1.7 点云匹配

点云特征提取方法将无规则的点云结构抽取为高维包含各种结构信息的几何特征。这些特征可以用于相似几何结构的匹配任务，构建其对应关系，并依据对应关系实现点云的配准。在现实场景应用中，扫描得到的点云往往不是完整的，拍摄得到的点云序列需要拼接才能得到完整的场景点云数据。找到合适的点云特征用于匹配不同扫描点云之间的几何关系极为关键。深度点云特征提取方式为场景点云匹配提供了新的思路。

利用预训练的方式进行点云特征匹配是一种常用的方式。首先分别提取输入两块点云的特征。接着利用对比学习，在特征空间中拉近存在对应关系点的特征对，将几何结构相差较大的特征对互相推远，从而使提取的点云特征能够将相似几何结构的区域匹配上。3DMatch（Zeng等，2017）提出了用于场景匹配的数据集，并将场景中任意两块互相有重叠区域的点云构建匹配对用于训练得到匹配特征，利用3维卷积实现场景的特征提取。3DSmoothNet（Gojcic等，2019）在3DMatch基础上引入了旋转不变局部参考系，使提取的特征与旋转变换不相关。全卷积几何特征（fully convolutional geometric features，FCGF）算法（Choy等，2019b）利用系数卷积提取点云特征，并提出了最困难样本对比学习，使点云特征彼此更具区分度，更容易学习得到不相关特征之间的边界。稠密3维局部特征检测与描述（dense detection and description of 3D local feature，D3Feat）算法（Bai等，2020）利用稠密特征提取获取更精细的点云特征，并利用关键点预测筛选出更具代表性的候选匹配点。SpinNet（spin network）（Ao等，2021）使用柱形卷积提取点云特征来提升匹配表现。

另一种点云匹配的方法是结合点云特征提取和点云匹配进行端到端的训练。借助2维图像端到端匹配的思路（Sarlin等，2020；Sun等，2021），首先提取场景点云从粗到细的特征，接着根据粗特征生成相似度矩阵进行粗匹配，再根据得到的匹配点周围的细粒度特征进一步进行细匹配。端到端训练的方式（Yu等，2021a；Qin等，2022；Yew和Lee，2022）在点云匹配任务上取得了不错的成就。

点云匹配成功地实现了将真实场景下拍摄得到的离散的点云碎片拼接成完整的场景点云。

2 场景点云语义分割

基于点云场景的语义分割技术是对3维场景精

细化、智能化理解的关键技术之一。语义分割任务早先源于对数字图像进行逐像素分类的需求(Long等,2015),后逐渐向3维视觉领域拓展。由于点云是3维场景中常用的离散化表征方式,因此逐点的语义类别预测成为3维视觉中的一项重要研究方向。与特征稠密分布的数字图像相比,3维点云场景数据规模大、覆盖空间广、特征分布稀疏以及缺乏顺序性,使得点云语义分割任务成为一大挑战。本节从点云场景表征与数据集、点云语义分割方法分类、多模态融合的分割方法与场景点云的实例分割方法四方面综述国内外研究趋势。

2.1 场景表征与数据集

点云场景表征方式可分为室内场景表征与室外场景表征。

2.1.1 室内场景表征与相关数据集

早期点云场景分割任务大多定义在室内场景中。室内传感器采集到的点云数据通常分布相对稠密,具备良好的几何结构特征,适合神经网络进行细粒度的分割。室内场景表征方法主要包括基于点特征的表征方法、基于图网络的表征方法和基于注意力机制的表征方法(Ye等,2022)。

基于点的特征提取网络PointNet与PointNet++(Qi等,2017a,b)是早期的点云特征提取网络。在此基础上,后续工作针对室内场景分割任务特点对网络进行优化改进。例如,为进一步挖掘点云局部区域间的上下文信息,PointWeb网络(Zhao等,2019)在PointNet++基础上提出自适应特征调整模块,利用局部区域中点对点的交互改变其在特征空间中的位置,以获取更好的区域特征向量。PointCNN(Li等,2018b)与PointConv(Wu等,2019b)等网络致力于定义基于点特征的卷积操作,根据空间密度、距离权重等设计卷积核,并构建深度点卷积网络提取特征等。Liu等人(2020)针对点云局部特征聚合操作,总结了基于多层感知机(multi-layer perceptron,MLP)、基于伪网格特征和基于相对位置加权的3种改进方式。基于点特征提取网络能较好地捕捉点云局部信息,但是对于全场景特征提取有欠缺,且在大规模点云数据集上存储与计算资源占用较大,不够高效。

基于图网络的表征方式充分考虑空间中点、边缘和区域等元素之间的邻接关系,是对3维几何结构的近似刻画。如Wang等人(2018a)提出的谱图卷积网络,对局部区域内的邻近点子集构建完全图,通过图傅里叶变换将特征映射到频域空间中再进行谱滤波,增强了提取空间结构特征的能力。与之类似的正则图卷积神经网络(regularized graph convolutional neural network,RGCNN)算法(Te等,2018)对点云的图卷积网络的监督函数增加了基于平滑性先验的正则项约束,使图卷积网络学习到的空间特征具有更好的几何连续性。Wang等人(2019)提出DGCNN,在每一层动态图上增加对边卷积网络层,能更好地学习室内物体的形状特征与潜在语义特征。然而,基于图网络的表征方式同样面临在大规模点云数据集上的存储开销和计算速度问题。

基于注意力网络的表征方式通过注意力机制建模3维空间中点之间或区域之间的上下文关系。Feng等人(2020)针对卷积网络难以充分提取不规则点云分布的特征的缺陷,提出了使用基于点的局部注意力和边缘卷积网络,通过空间注意力机制构建大范围内长距离的关系信息。在此基础上,之后的研究工作开始利用基于Transformer的自注意力机制来提取点云表征,进而获取丰富的局部邻域信息和区域之间的上下文关系。Park等人(2022)提出由轻量级的自注意力层组成的快速点云Transformer网络,通过编码连续的点云坐标和基于体素哈希的架构来有效地提升网络的计算效率。Yu等人(2022)设计了一种基于掩码Transformer的点云预训练方法,首先将整个输入点云切分为若干区域块并随机掩盖掉部分区域块,然后使用基于Transformer的点云网络来恢复缺失的点云数据,从而达到预训练的目的。除此之外,为了解决自注意力机制在大规模点云数据集上空间和时间复杂度较的问题,Zhang等人(2022)提出了基于块注意力的点云Transformer网络来自适应地学习更小点集的特征,并设计了轻量级的多尺度注意力网络来构建不同场景规模下的区域注意力关系。此类基于Transformer的点云特征提取网络利用注意力机制来获取3维空间中点之间或区域之间的上下文关系,同样存在对存储空间占用高的问题。

室内点云场景数据集主要以RGB-D相机扫描得到的数据为主,包括NYUv2(New York University version 2)数据集(Silberman等,2012)、SUN RGB-D(scene understanding RGB-D)数据集(Song等,2015)、S3DIS(Stanford large-scale 3D indoor spaces

dataset）数据集（Armeni 等，2016）和 ScanNet 数据集（Dai 等，2017）等。这些数据集涵盖多种室内场景，包含从物体级别语义标注到全场景的高层次标注，有力支持了室内点云场景分割的研究发展。

2.1.2　室外场景表征与相关数据集

随着智慧城市建设、自动驾驶感知等应用任务需求增加，室外场景表征方法受到广泛关注。室外场景与室内场景相比，场景类型更加复杂，点云密度更加稀疏，室外天气与光照影响更加明显，各类别物体长尾分布现象更加严重，使得室外点云场景分割成为一项极具挑战性的任务。

目前的室外场景表征方法大致包括基于环视图（range view）的分割方法、基于稀疏卷积（sparse voxel）的方法、基于鸟瞰图（bird-eye-view，BEV）的方法和基于神经辐射场（neural radiance filed，NeRF）的方法。基于环视图的方法（Milioto 等，2019；Cortinhal 等，2020）将点云数据 360°投影到预设半径的环视面（range view）上，形成 2 维环视图，然后使用图像卷积网络提取特征并预测分割结果。最后通过相关后处理算法（k 近邻采样、双线性插值等）将环视图的分割结果传播到点云上。该类方法的优势在于可以用 2 维卷积网络提取 3 维点云投影降维后的图像特征，较好满足实时性需求。缺点是将 2 维分割结果传播到 3 维点云数据时会造成较大的精度损失。基于稀疏卷积的方法（Graham 等，2018）通过将卷积计算限制在活跃区域（active region）中，避免纳入空区域的计算操作，从而大幅减少计算量。在此基础上，针对室外激光雷达数据集环形分布特点，Zhu 等人（2021）采用扇形卷积的方式划分点云，更好地满足近密远疏的分布特性。近年基于鸟瞰图的场景特征提取方法日渐兴起。点云场景感知中的鸟瞰图概念源于 2020 年特斯拉公司公布的全自动驾驶算法，但该方案是纯视觉方案，具体做法是将多视角相机拍摄的数字图像转化为鸟瞰图特征。后续有很多研究者尝试使用鸟瞰图类似地表征激光点云场景，如 Zhang 等人（2020c）提出的 PolarNet 网络，在极坐标系下，通过池化层将点云特征投影到固定大小的俯视图平面上，使用卷积网络得到 2D 特征并获得预测结果，最后同一俯视图栅格里不同高度的点云赋予相同的预测类别。虽然基于鸟瞰图特征的表征方式在实时分割的前提下也能获得不错的精度，但是对于悬吊物体的预测结果通常较差。基于神经辐射场的相关表征方法（Kundu 等，2022）使用多层感知机构建了从 3 维场景中的位置坐标（视角+距离）到语义特征（颜色+反射率）的映射函数，作为 3 维场景的神经辐射场用于辅助下游语义分割任务。该类方法可直接用于下游的 3 维场景分割任务，亦可以作为点云—图像融合的上游特征提取器，在未来有较大的研究与应用前景。

室外场景数据集根据传感器不同，主要分为激光雷达（light detection and ranging，LiDAR）数据集和毫米波雷达（radio detection and ranging，RADAR）数据集。室外静态 LiDAR 数据集如 Semantic3D（Hackel 等，2017）提供了包括城市、乡村、广场以及街景建筑等多种场景的 3 维语义数据。室外自动驾驶场景 LiDAR 数据集，如 SemanticKitti（Behley 等，2019）、nuScenes（Caesar 等，2020）、Waymo Open Dataset（Sun 等，2020）和 Lyft L5（Houston 等，2020）等提供了自动驾驶场景下的大规模点云—图像多模态数据集，包含行人、非机动车、机动车以及各类交通标注物等类别。此外，nuScenes 与 Waymo Open Dataset 数据集亦提供毫米波雷达的相关数据，可有效支持在雨天、雪天和雾天等极端天气下较准确地探测到移动物体。

2.2　点云场景语义分割

2.2.1　全监督分割方法

点云场景语义分割任务需要神经网络学习到多种场景下的 3 维特征表示。由于 3 维场景的复杂性，仅依靠数据集自身提供的全量标签直接训练特征提取器，难以使神经网络快速学习到有价值的信息。因此，很多全监督分割方法会挖掘点云的先验信息，如空间分布特征和时序特征，增强网络对点云的识别与分割能力。Gong 等人（2021b）提出了边缘预测模块（boundary prediction module）和边缘几何特征编码模块（boundary-aware geometry encoding module），使得神经网络对物体的边缘特征更加敏感，从而提升分割准确率。Chen 等人（2022）利用激光点云中的中心对称性分布特征，提出极角正则化数据增强操作，将不同水平角下划分的点云区域旋转到相同的角度，减小了因角度多样性给点云网络训练带来的困难，在多种点云语义分割基线网络中得到分割精度的提升。Schutt 等人（2022）借鉴光流法的思想，提出基于多级循环神经网络连接的前后点云帧时序融合方法，使点云网络能够更有效地区分静止

物体与运动物体。

此外，点云场景的表征方式多种多样，如何充分利用不同的表征方式融合点云各项信息，从而降低语义分割的训练难度也是研究者关注的内容。Xu等人(2021b)提出环视图—点—体素三位一体的融合模块，增强了同一个点在不同表征下的特征交互的能力。Ye等人(2021)在点—体素双路感知网络的基础上，提出了交替转换的训练方法，将原先双分支各自独立训练的方式改为从点云到稀疏体素，从稀疏体素到点云两种融合模块，并在这两种融合模块间进行多轮循环迭代，充分提取各个层次上的体素级与点级语义信息。Gong等人(2021a)首次提出一种层次化感受野因果推理模块，将场景分割问题转化成多种类别所在的子区域感受野成分分解和编码问题。Li等人(2022b)提出了基于特征金字塔和注意力感知的点—网格融合插件模块，对环视图—鸟瞰图双路点云感知网络进行增强，在多种数据集上达到了领先的性能。

2.2.2 有限标注条件下的分割方法

相比全监督学习，有限标注信息下的点云语义分割方法有更加丰富的应用场景和工业界落地需求，在实现精度上接近全监督方法的同时，尽可能减少人工标注的成本。根据标签利用方式的不同，可大致分为半监督学习和弱监督学习。半监督学习的目标是在只给定部分场景标注的条件下训练神经网络(被选定场景下的点云标注是完整的)，强化其在不同场景下的泛化能力。而弱监督学习的目标是在给定不完整标注的条件下(例如每帧点云场景只随机挑选1%的点标注)，通过学习有限区域的监督信息，传播并习得所有区域的点云特征。

针对室内半监督分割，Li等人(2021b)提出一种基于伪标签置信度预测的半监督分割方法，以减少对大规模高质量人工标注的依赖，在分割网络的基础上，额外设计判别网络(discriminator network)，该网络目标是区分预测结果和真实标注，并对无标注点云的预测结果输出置信度预测，对判别网络的训练更好地促进了整个网络对无标注数据的分割与预测能力。面向室外激光点云数据集，Kong等人(2022)基于激光点云扫描线环视分布的特点，提出一种有标注场景和无标注场景的点云环形混合增强方法(LaserMix)，在多种现有半监督方法上均取得较大的分割精度提升。

Xu和Lee(2020)首次在点云上提出弱监督语义分割任务，在理论上说明了使用不完整标签的数据集训练的网络权重的梯度与全监督梯度基本近似，在室内点云数据集中，提出的基线方法在只使用约10%的点云标注条件下，精度可达到全监督方法的95%左右。此后，更多研究者开始关注如何使用更少的点云标注获得与全监督基线更接近的分割性能。Zhang等人(2021b)提出通过加入点云排列增强模块监督预测结果的拓扑一致性，在室内场景中使用约1%的真值获得的mIoU(mean intersection over union)与全监督基线的结果仅相差近2%。基于混合对比学习正则化约束的增强方法，Li等人(2022a)使用极少标注(0.03%)在室内点云数据集上获得的分割精度为全监督方法的78.3%。面向室外点云弱监督分割任务，Unal等人(2022)提出了首个室外激光雷达弱监督非精确标注数据集Scribble-Kitti，并在该数据集上使用基于教师—学生网络(Tarvainen和Valpola，2017)改进的弱监督方法，使用约8%的真值标签获得的精度可达到全监督方法的96%左右。目前，已有研究工作(Sautier等，2022)在室外激光点云数据集上使用约0.8%的真值标签获得的精度达到了全监督方法的90%左右。

2.2.3 无监督分割方法

无标注的分割方法主要聚焦在点云自监督学习和无监督域迁移方向。鉴于点云标注非常耗费时间与人力资源，只对部分场景进行部分标注也难以适应海量增长的3维点云数据量。因此，采用自监督学习的方式对海量点云进行预训练是一个值得深入探讨的问题。Sautier等人(2022)首次提出一种室外场景下的图像预训练权重向点云网络知识蒸馏的方法，在不需要任何点云与图像标注的条件下，通过提取超级像素(super pixel)建模图像与点云间高相似度区域间的对应关系，并通过基于对比学习的蒸馏损失函数进行监督。Afham等人(2022)在室内场景物体上提出一种简单的跨模态3维—2维区域对应模块，分别将点云模态和图像模态提取的特征向量重新投影到一个公共的特征空间中，并基于最大化与模态无关的互信息的思想设计对比学习损失函数。总体来看，目前的点云自监督学习方法与全监督方法仍有巨大差距，预训练权重对下游全监督任务的提升效果有限，有待进一步研究发掘点云自监督学习的潜力。

除了在无标注信息的条件下做网络自监督预训练外，另一个工业界与学术界的重大需求是克服不同域/数据集之间的特征分布差距，使模型在源域数据集上训练达到很好的精度时，迁移到无标注的目标域上能缩小目标域特征分布与源域之间的“距离”。Wu等人（2019a）研究从大规模道路场景仿真数据集向真实数据集域迁移，通过提出的邻域特征聚合模块和渐进式域校正算法有效克服跨域噪声干扰与信息丢失问题。此后，许多研究工作，如跨模态无监督域适应（cross-modal unsupervised domain adaptation，xMUDA）算法（Jaritz 等，2020）和点无监督域适应（point unsupervised domain adaptation，PointUDA）算法（Bian 等，2022），围绕该方向提出一系列改进算法，促进了无监督分割的研究进展。

2.3 多模态融合的分割方法

单一模态的场景分割方法虽然已达到较高的精度性能，但也面临着与模态相关的固有缺陷。如纯图像的场景分割容易受光照、遮挡因素影响；RGB-D点云数据受限于室内小规模场景扫描；激光点云数据在室外容易受极端天气的干扰；超声波雷达数据探测精度相对激光点云会差等。因此，研究跨传感器多模态融合的分割方法，可以较好地实现模态间信息互补，使网络更容易学习到鲁棒性强的场景特征表示。依据融合方式，目前多模态点云分割方法大致可以分为前融合、深度特征融合，后融合、非对称融合四种（Ma等，2022）。依据使用的主流传感器类型，可分为激光点云—相机融合（Zhuang 等，2021）和毫米波点云—相机融合两类（Zhou 等，2022）。虽然目前多模态融合方法在许多数据集上取得领先的性能，但仍有许多问题须待解决。例如，克服跨模态特征错位对应问题、多模态数据集跨域迁移时模态失配问题等。该方法仍有很大提升空间。

2.4 场景点云的实例分割方法

在场景理解中，语义分割虽然能够提供每个点的类别属性，但是无法区分出每个实例的边界，即缺乏对场景内的3D点云进行实例级别的感知。相比于语义分割，实例分割的着眼点在于区分不同的实例，需要对场景内的点进行额外的身份标识。因此，实例分割的研究，能够使环境感知系统具备理解3维真实世界中每个独立物体或个体的能力，直接影响着与3维场景中每个实例的交互活动。依照流程，目前的实例分割方法可分为以3D-BoNet（Yang 等，2019a）、生成形状提议网络（generative shape proposal network，GSPN）算法（Yi 等，2019）为代表的基于Proposal的方法和以PointGroup（Jiang 等，2020）、层次化聚合3维样例分割（hierarchical aggregation for 3D instance segmentation，HAIS）算法（Chen 等，2021）为代表的Proposal-free的方法。基于Proposal的方法遵循自上而下的流程，首先生成众多的实例候选区域，并在每个区域内预测实例的掩码；Proposal-free的方法则采用自底向上的方式，通过计算点之间的相似度或距离，将点聚类至不同的实例之中。从当前的研究工作来看，Proposal-free的实例分割方法在ScanNet和S3DIS等数据集上取得了不错的性能。

3 扫描点云物体补全

点云作为一种表征3维物体的基础数据形式，具备高纬度信息量的优势，在自动驾驶和场景感知等领域有着广泛的应用。但是在点云数据采集的过程中，由于遮挡、噪声干扰和视角变换等问题，真实扫描到的3维点云通常会出现残缺和数据不完整的问题，严重阻碍了下游的点云分析和处理任务的性能。因此，通过残缺点云数据恢复出3维物体的整体形状的3维点云补全任务逐渐成为一个新的研究热点。

本节首先总结3维点云补全任务中常用的数据集，然后从全监督点云补全和真实扫描点云跨域补全两方面介绍3维点云补全任务。

3.1 点云补全数据集

对于3维点云补全任务，常用的数据集主要分为人工生成的点云数据集和真实扫描的点云数据集两种类别。人工生成的数据集是通过在某个固定视角下均匀采集3维面片模型的表面点云，得到具有残缺几何形状的3维点云数据。真实扫描的点云数据集则是通过激光雷达等采集设备从真实环境中直接扫描得到不完整的3维点云数据。

3.1.1 人工生成的点云补全数据集

ShapeNet数据集（Chang等，2015）是一个大规模的3维模型数据集，具有丰富的注释信息，共包含55种常见的物体类别和220 000个计算机辅助设计（computer aided design，CAD）模型，每个模型对应的3维点云大概包含15 000个数据点。对点云补全任

务来说，选取8个类别的物体，共30 974个3维CAD。其中，完整的点云数据通过在每个3维模型的表面均匀采样2 048个点组成，对应的残缺点云数据则是将这个3维模型随机视图下的深度图反投影到3维空间来获得，残缺点云的点数也是2 048个。

ModelNet40数据集（Wu等，2015）是一个综合的3维CAD模型数据集，包含40个类别和13 356个模型。残缺点云数据和完整点云数据的获得方法与ShapeNet数据集相同。

3.1.2　真实扫描的点云补全数据集

KITTI（Karlsruhe Institute of Technology and Toyota Technological Institute）数据集（Geiger等，2012）是通过激光扫描仪收集的。该数据集最初是为了评估立体匹配的性能，由雷达点云、点云数据序列和标注信息组成，包含22个点云数据序列，其中训练集包括11个具有标注信息的点云数据序列，评估集包含11个没有标注的点云数据序列。对于3维点云补全任务来说，只选取了其中的汽车类别作为训练和测试数据。其中，残缺的3维点云数据是通过均匀选取2 048个数据点获得。KITTI数据集中的3维点云数据是非常稀疏的，且物体的几何结构往往是不完整的，因此在这个数据集上进行点云补全非常具有挑战性。

3.2　全监督点云补全

3维点云补全任务旨在从输入的残缺点云数据中恢复物体完整的几何形状。全监督3维点云补全是在有完整点云数据作为监督标签的情况下，训练点云补全网络，达到预测完整补全结果的目的。根据3维点云补全任务中采用的网络结构，全监督点云补全方法可以分为基于点、基于图、基于生成对抗模型和基于变分自动编码器的点云补全方法。

3.2.1　基于点的全监督点云补全

基于点的点云补全方法通常采用编码器—解码器方式设计网络架构。在编码器—解码器结构中，补全分支中的编码器旨在提取全局的3维几何特征和每个点的区域局部特征。而解码器负责预测3维物体完整的点云并对其进行细化处理。

Xia等人（2020）设计了端到端的3维点云补全网络，从车辆应用中的稀疏点云重建更均匀和更精细的结构，同时采用上采样方法生成更均匀的点云。此外，提出一种非对称的连体特征匹配网络（Xia等，2021），其中，非对称连体自动编码器生成粗略但完整的点云数据，随后的细化单元旨在恢复具有细粒度细节的最终点云预测结果。Mendoza等人（2020）提出一个由缺失部分预测模块和合并细化模块共同组成的端到端补全网络，在保留现有几何形状和细化细节的同时预测点云数据的残缺部分。Peng等人（2020）提出一种端到端的稀疏到密集多编码器神经网络来补全残缺点云数据，同时可以有效保留原始3维物体的形状细节。残缺的输入点云分两个阶段补全和细化。在第1阶段，基于两层感知机网络生成粗略但完整的结果；在第2阶段，使用新的网络对第1阶段的稀疏结果进行编码和解码，以产生高密度和高保真点云数据。Miao等人（2021）提出一种具有形状保持功能的补全网络，通过设计编码器—解码器的方式来保持物体的3维形状并恢复重建物体的精细信息。这种形状保持网络可以学习全局特征并整合具有不同方向和尺度的相邻点的区域信息。在解码过程中，信息将融合到潜在向量中。

3.2.2　基于图的全监督点云补全

由于点云和图都可以视为非欧几里得的结构化数据，因此将点或局部区域作为某些图的顶点来探索点或局部区域之间的关系是很有潜力的方法。基于图的网络可以将输入中的每个点都视为顶点，同时利用相邻点的信息来生成边。因此，图卷积网络可以适用于点云的处理和补全任务。

Wang等人（2019）开创性地提出DGCNN，成功地将动态图卷积结构引入3维点云补全任务。在动态图卷积中，相邻矩阵可以通过来自潜在空间的顶点关系计算，该图是在特征空间中建立的，可以在网络训练过程中动态更新。Hassani和Haley（2019）引入多级网络来利用点和形状特征进行自监督的3维点云补全。Wu等人（2021a，b）提出一种基于学习的图卷积方法，对部分输入的局部区域进行采样，对其特征进行编码，并将它们与全局特征相结合。建立图后，收集所有区域特征，并用多头注意力机制对图进行卷积。图注意机制使每个局部特征向量能够跨区域搜索，并根据高维特征空间中的关系选择性地吸收其他局部特征。同时，设计了一个基于图注意力的跨区域注意力单元，该模块量化了特定背景下区域特征之间的潜在联系，并通过全局特征进行解释。因此，每个条件区域特征向量都可以作为图注意力进行搜索。Zhang等人（2021c）设计了一个图神经网络模块，通过局部—全局注意机制和基于多尺

度图的上下文聚合，全面捕捉点之间的关系，大幅增强了图网络编码特征。

3.2.3　基于生成对抗模型的全监督点云补全

与传统的卷积网络相比，生成对抗网络（generative adverserial network，GAN）利用判别器的隐式学习来估计生成器预测的完整点云的准确性。本节将从端到端机制和点云精细化模块两部分介绍基于生成对抗模型的全监督点云补全。

围绕端到端机制，Wang等人（2017）利用编码器将体素化的3维形状映射到概率潜在空间中，并使用生成对抗学习来帮助解码器借助潜在特征表示生成完整的点云形状。Achlioptas等人（2018）则使用全连接层设计了具有生成器和判别器的生成对抗网络，自动编码器被训练来学习潜在空间，然后在固定的潜在表示中训练生成模型。 这种网络在潜在空间中进行训练，比普通的生成对抗网络更容易训练，从而可以更好地恢复残缺的物体的几何结构。

点云的精细化模块常常作为一项关键性的技术集成到生成对抗学习中。Wang等人（2020b）提出一种用于学习先验形状的特征对齐方法。同时，设计了一种从粗到细的方法，将形状先验与从粗到细的策略相结合。除此之外，还设计了一个点云补全网络（Wang等，2020a），以级联细化网络作为生成器，通过利用输入的细节高质量地生成点云残缺的几何结构。同时，设计了一个分片化处理的判别器，使用对抗训练来精确地学习点云分布，并约束预测点云与完整点云之间不同的几何结构。

3.2.4　基于变分自动编码器的全监督点云补全

Spurek等人（2021）首次利用变分自动编码器架构来补全输入的残缺点云的完整几何结构。其中，点云处理被分成两个未连接的数据流，并利用超网络范式来恢复丢失部分留下的空间结构。Pan等人（2021）设计了一种变分关系补全网络，利用双路径单元和基于变分编码器的关系增强模块进行概率建模，同时还设计了多个关系模块，可以有效地利用和集成多级的点云特征，包括点自注意力内核和关键点选择内核单元。Zamorski等人（2020）提出了3种生成建模方法的应用，并定量和定性地测试了自动编码器 、变分自动编码器和对抗性自动编码器的架构特点。

3.3　真实扫描点云跨域补全

目前主流点云补全网络依赖于成对的数据监督，即对每一个残缺的点云扫描需要一个相应的完整点云。成对数据通过扫描虚拟3维物体很容易获得，但在现实世界中难以获取，且由于虚拟与现实域间的数据分布差异，使用虚拟成对数据训练的补全网络难以推广到真实数据。因此，真实扫描的点云跨域补全成为一个新的研究热点。

3.3.1　基于生成对抗模型的跨域补全

Chen等人（2020）首先提出在不需要成对数据的情况下以无监督方式进行点云补全，该方法训练两个独立的自动编码器，分别用于重建虚拟完整点云和真实残缺点云，并训练生成器将残缺点云的潜在空间映射到完整点云潜在空间，同时引入判别器约束目标样本的潜变量与源样本的分布相同。Wen等人（2021）设计了残缺输入和完整点云的潜码之间的双向循环转换框架。正向循环将点云从残缺域转换到完整域，然后再将其投射回残缺域。该循环学习完整点云的几何特征，并保持完整预测和残缺输入点云之间的形状一致性。反向循环转换从完整域转换到残缺域，然后投射回完整域来学习残缺点云的特征。由于神经网络无法将单个完整点云表示映射为多个残缺点云表示（目标混淆问题），故提出缺失区域编码以表达目标残缺点云信息，原始残缺点云的编码表示分解为相应完整点云的表示和缺失区域表示。当从残缺点云预测完整点云时，只需考虑完整点云表示的部分；而当从完整点云中预测残缺点云时，则需同时考虑两个编码表示。该框架不足之处在于双向循环过程需各自单独建模，尤其完全到残缺的映射过程难以学习。如果一个方向没有学好，另一个方向也会受到性能制约。

Zhang等人（2021a）首次在点云补全任务中引入GAN逆映射。利用在完整点云上预训练GAN得到的点云形状先验，通过GAN逆映射寻找最佳匹配的潜码。具体而言，一个潜码通过预训练GAN生成一个完整点云，再通过一个3维降采样模块将完整点云转化为残缺点云，进而与输入残缺点云计算损失。该框架利用梯度下降方法反传损失以更新潜码并微调预训练的GAN网络，从而使生成的完整点云与输入的残缺点云在可见部分最接近。3维降采样模块寻找输入的残缺点云与任意生成的完整点云间的对应关系。具体而言，对残缺点云中每一个点寻找完整点云中欧氏距离下最近邻点，所有邻点的并集构成了与输入残缺点云对应的输出残缺点云。该方法

在保证泛化能力的同时，对残缺输入的不确定性可提供多解，并且保证各解都合理地反映残缺物体的可见部分。且由于GAN的引入，该框架能够很好地实现对已知点云形状的编辑。然而，与基于学习的方法相比，这种基于GAN逆映射反转优化的方式效率极低，且补全性能非常依赖于潜码的初始值。

3.3.2　基于解耦的跨域补全

Cai等人(2022)提出了一个统一的结构化潜空间以增强残缺—完整点云的几何一致性，并提高补全精度。该方法将残缺点云表示解耦为完整形状因子和遮挡因子。两者逐元素乘积用以重建残缺点云，补全过程仅使用完整形状因子。为学习该结构化潜空间提出了一系列约束条件，包括结构化排名正则、潜码交换以及潜码分布监督。具体而言，对某输入残缺点云进行下采样得到一系列残缺点云，该系列点云完整形状因子相同，遮挡因子满足不等式关系。同时，该方法引入潜码判别器使得从残缺点云学习得到的完整形状因子与从完整点云学习得到的完整形状因子相匹配。

Gong等人(2022)结合回归与优化两个阶段提高补全点云与输入残缺点云间的一致性，加速模型推理速度。第1阶段特征解耦进行域级别的对齐，残缺点云特征被解耦为域、形状和遮挡3个因子。其中，残缺点云的遮挡因子与观察视角强相关，故设计自监督视点预测任务以学习遮挡因子；域因子与形状因子分别代表域风格与点云形状，故使用域判别器结合梯度反转同时训练域因子与形状因子；设计因子排列一致性正则以确保因子间相互独立，随机交换样本间因子用以重建特征并约束重建特征一致。第2阶段推理优化过程进行实例级别的对齐，第1阶段预训练编码器产生的潜码并不直接生成点云，而只是作为解码器的初始输入。使用输入残缺点云与预测完整点云间的距离作为监督，在多轮迭代中微调潜码以寻找最佳点云生成效果。

4　国内研究进展

4.1　3维特征提取方式与旋转不变性

3维特征提取在近几年取得了飞速发展，国内对于点云特征提取的研究也产出了优秀的成果。Li等人(2018b)提出了PointCNN，设计了χ-卷积初步实现对离散点集进行卷积操作，为之后点云卷积的发展铺下了良好的基石。Liu等人(2019b)提出形状关系卷积神经网络(relation-shape convolutional neural network，RSCNN)，利用点云几何形状的特征生成对应卷积核的权重来实现点云卷积，带来了显著的效果提升。Yan等人(2020)设计了点适应性采样与局部非局部模块(point adaptive sample and local-nonlocal module，PointASNL)，在点云卷积神经网络中引入注意力机制。PointASNL利用注意力机制提出自适应采样，使得降采样点具有偏移能力，从而提升其代表能力。同时，引入局部与非局部模块提升不同局部模块之间的关联程度，提升特征的全局表达能力。马利庄团队(Liu等，2022)提出了Scatter-Net(scatter network)，利用散布探索模块代替传统的最近邻搜索和球形搜索算法，实现更长、更广范围的局部邻域点组合，使卷积操作能够从更详细的局部几何信息中提取特征。

Guo等人(2021)以及Zhao等人(2021)率先在点云上拓展了Transformer框架。前者利用自注意力机制通过挖掘输入点云整体点之间的关联度来提取逐点的特征。但是全局的方式会占用大量的内存资源，导致无法适用于大规模的场景点云特征提取任务。后者则将自注意力机制运用到局部点云上，并在不同局部几何上复用自注意力模块。该方式有效减少了计算资源的浪费，并且使得点云Transformer达到相当的效果。Wu等人(2022)在Point Transformer v1(2021)的基础上拓展了Point Transformer v2。PTv2提出了分组向量注意力机制，改善了深度模型过拟合等问题，使得点云Transformer模型也可以部署足够深度的神经网络结构。

针对点云旋转不变特征提取，国内也展开了研究。Chen等人(2019)提出ClusterNet，利用局部邻域中点之间的相对角度和相对距离代替坐标作为神经网络的输入来提取点云的特征。由于旋转变换是刚体变换，不会改变点云局部的几何结构。相对距离和相对角度作为局部几何的一种衡量标准可以作为低层的旋转不变特征。因此，ClusterNet能够进一步将低层特征提取为高层的旋转不变特征。You等人(2020)提出逐点旋转不变网络(pointwise rotation-invariant network，PRIN)算法以及稀疏逐点旋转不变网络(Sparse PRIN，SPRIN)算法(You等，2022)来提取点云旋转不变特征。PRIN将旋转空间划分为离散的球形体素，并利用球形体素卷积提取逐点的

旋转不变特征。Yu等人(2020)设计PR-invNet(positional & relational feature embedding block-based rotation-invariant network)，利用PCA初步计算一种输入点云的参考系，并在此基础上用固定角度的旋转增强来构建旋转空间。PR-invNet借助提出的姿态选择器从旋转空间中挑选输入点云的标准姿态后，将其作为神经网络的输入，从而提取旋转不变特征。Zhao等人(2022a)同样借助局部相对信息，提出局部全局表征网络(local global representation network，LGR-Net)，利用更精细的8维相对距离角度特征来代替坐标输入，在实现旋转不变特征提取的同时，提升了实验结果。

4.2 场景点云语义分割

点云场景分割在3维视觉感知中具有关键作用。目前国内点云场景理分割的相关技术在快速发展，在多个子方向与赛道上均有许多出色研究工作涌现。其中，马利庄团队在全监督和弱监督点云场景分割任务上有重要研究进展。

点云场景中对物体边缘的识别能力对分割效果有着重要影响。基于此，马利庄团队(Gong等，2021b)提出边缘预测模块(boundry prediction module)对不同类别物体的边缘进行预测。其中，边缘预测模块预测结果如图1所示。同时，提出边缘感知的几何特征编码模块(boundary-aware geometry encoding module)从局部区域里挖掘边缘敏感的几何特征。相比现有的基于点特征的表征方式PointNet++(Qi等，2017b)和PointCNN(Li等，2018b)、基于图卷积表征的方法分割图卷积网络(graph convolution network for segmentation，SegGCN)算法(Lei等，2020)以及基于注意力机制的表征方式点注意力转化器(point attention transformer，PAT)算法(Yang等，2019b)等多种现有分割方法，均得到了显著的分割精度提升。

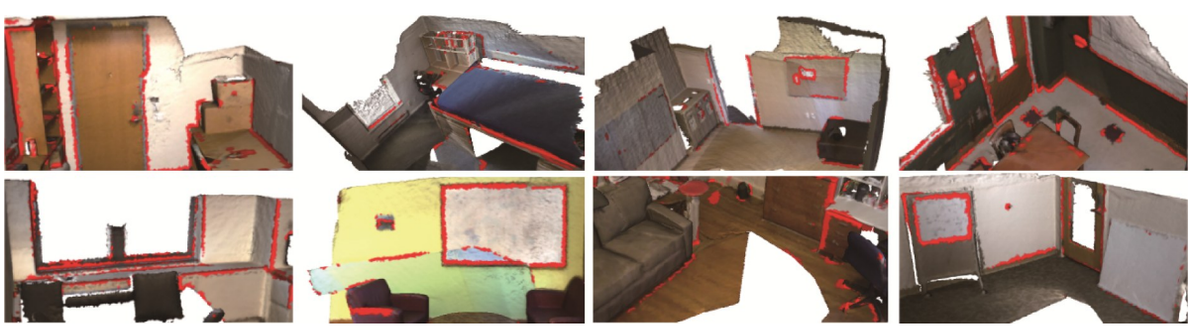

图1 ScanNet场景边缘预测结果(Gong等，2021b)

Fig. 1 Boundary prediction results for scenes in ScanNet (Gong et al., 2021b)

面向场景点云分割中的细粒度学习与因果推理，马利庄团队(Gong等，2021a)首次提出一种层次化场景感受野成分推理模块，将场景分割问题转化成多种类别所在的子区域感受野成分分解问题。基于感受野的子区域成分编码(receptive field component code)很好地刻画了区域语义类别信息，将不同层次的感受野成分编码从粗粒度向细粒度分解，最后得到逐点的语义类别推理结果。此外，在网络训练阶段亦可对全层次的中间层编码进行多尺度监督。相关研究成果(Gong等，2021a)在室内点云数据集S3DIS和室外点云数据集Semantic3D上均取得领先的分割效果。刘盛等人(2021)设计了空间深度残差网络(spatial depthwise residual network，SDRNet)，结合空间深度卷积与残差结构以及扩张特征整合模块有效减少了计算量，保持较快的分割速率。

在弱监督点云分割中，马利庄团队提出一种混合对比学习正则化约束的增强方法(Li等，2022a)。现有基于对比学习的弱监督点云分割方法通过对真实点云做数据增强(如随机旋转、随机翻转等)形成参照样本，通过构建原始点云和参照样本之间的正负样本对，从而使用对比损失函数训练。此外，该方法进一步考虑点与其近邻区域间语义类别应具有局部连续性的特点，结合伪标签和一致性约束的相关技术，提出一种混合对比学习的网络结构，如图2所示。在局部区域里，每个视角的点云与另一视角下的邻域空间满足一致性约束；在全局层次里，每个视角的预测结果与另一视角下的全局类原型特征通过对比学习建立约束。在S3DIS数据集上成功实现每帧点云场景只使用0.03%标注获得的分割精度为全监督方法的78.3%左右。

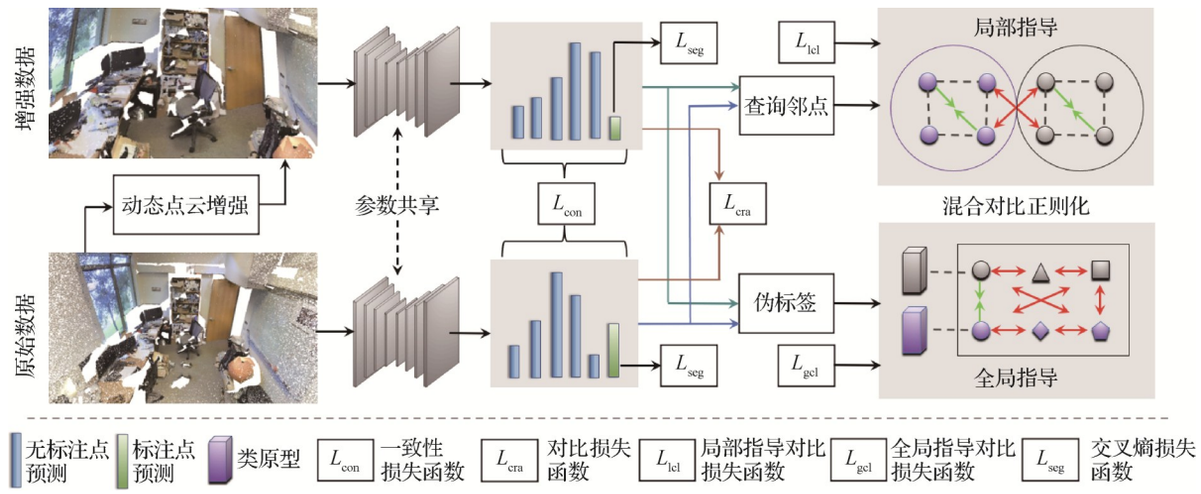

图2　混合对比学习正则化约束的增强方法框架(Li等,2022a)

Fig. 2　Framework of hybrid contrastive regularization (Li et al., 2022a)

4.3　场景点云样例分割与检测

在场景理解中,3D点云实例分割是一项具有挑战性的任务。相比于语义分割,实例分割需要对场景内的点进行更为细粒度的推理。具体来说,实例分割除了需要区分不同语义类别的点,还需要进一步分离属于同一语义类别的单独实例。现有的研究方法可以归纳为两类,即基于Proposal的方法和Proposal-free的方法。

基于Proposal的方法遵循一种自上而下的策略,通过生成一系列的proposal来检测出每个实例,并在每个proposal内分割出实例掩码。Yang等人(2019a)提出3D-BoNet来直接回归点云中所有实例的3D边界框,并同时预测出每个实例掩码。对于目标proposal的生成,3D边界框是对物体的一种简单几何近似形式。然而,对于大部分物体3D边界框是不可靠的,因为3D边界框不依赖于对目标物体几何形状的深入理解,导致单个proposal内会包含多个对象或仅包含对象的某个部分。因此,Yi等人(2019)提出了GSPN方法,没有将目标proposal的生成视为一个直接的边界框回归问题,而是采用综合分析的策略,通过场景内的噪音观察重建形状以生成优质的目标proposal。

Proposal-free的方法摒弃了对Proposal的依赖,将实例分割作为语义分割的后续聚类步骤。Wang等人(2018b)提出了相似性群提议网络(similarity group proposal network,SGPN),以PointNet作为骨干网络来提取点的特征,并设置了相似度矩阵模块来学习所有点对在特征空间上的相似度,从而将相似的点融合为实例。然而,构造点对的相似矩阵需要占用大量内存,且相似矩阵存在较多的冗余信息,难以拓展到大规模的点云数据中。因此,Liu等人(2019a)提出了基于稀疏卷积的多尺度亲和度(multi-scale affinity with sparse convolution,MASC),该方法首先对点云做体素化处理,并在子流形稀疏卷积的基础上预测每个非空体素的语义得分,同时生成不同尺度下相邻体素的亲和度,最后根据语义预测和亲和度大小来生成实例。除了通过相似度矩阵和亲和度来进行实例聚类外,许多现有方法计算点的中心偏移量,并依据偏移点之间的空间距离来进行实例分组。Jiang等人(2020)提出PointGroup方法,在预测点语义标签的同时估计点到对应实例中心的偏移量,并用该偏移量来生成一个偏移点集。然后,PointGroup在原始点集和偏移点集内均进行实例聚类。对于点集内的每个点,PointGroup以点的坐标作为参考,将点与其邻近且有着相同语义的点进行分组,并渐进地扩大每个实例组。在PointGroup的基础上,Chen等人(2021)提出分层聚合的HAIS(hierarchical aggregation for 3D instance segmentation)方法,首先将点聚合至距离阈值较低的点集中,以避免过分割,然后再用动态的距离阈值合并点集以形成完整的实例。考虑到点集的聚合会将噪声点吸收至实例中,HAIS设计了针对实例内部的子网络,用于去除实例内部的噪点并对实例掩码的质量进行评分。PointGroup和HAIS在区分前景点和背景点时,均采用了硬语义分割的形式,即一个点仅被分配单个语义类别,然而在大多数情况下,点云物体

的局部通常都是模糊的，这使得同一个物体的不同部分易被预测为不同的类别，此时使用硬语义分割的结果进行后续的实例聚类将导致语义分割的错误预测被传播至实例分割结果。因此，Vu等人（2022）提出了SoftGroup模型，允许每个点关联多个类别，以缓解语义预测错误对实例分割的影响，并将假阳性的实例预测视为背景类来进一步提高语义分割的性能。

4.4 扫描场景与物体点云补全

4.4.1 全监督点云补全

3维点云补全任务旨在从输入的残缺点云中预测完整的几何形状。随着点云处理方法的快速发展，全监督点云补全任务不断取得性能上的提升。Zhang等人（2020b）提出两种特征组装策略进行3维点云补全，利用多尺度特征的功能并整合不同的信息来分别表示给定的部分和缺失的部分。同时，借助全局和局部特征聚合和残差特征聚合来恢复完整的点云几何结构。此外，还设计了一个细化模块，以防止生成的点云分布不均和异常值。Zhao等人（2021）设计了一种缺失点云部分的补全方法，主要强调两个点云非常接近且上下文相关的配对场景，还设计了一个网络来编码单个的几何形状以及成对场景中不同点云之间的空间关系，使用不同点云序列之间的一致性损失作为监督来训练双路径网络，这种方法可以处理点云之间严重相互遮挡的复杂情况。Yu等人（2021b）首先将基于Transformer的编码器—解码器网络集成到点云完成任务中，并通过解决集合到集合的转换问题完成残缺点云的补全。刘心溥等人（2022）提出多尺度的嵌入注意力模块，通过特征嵌入层与Transformer层提取融合不同尺度特征，优化细节补全效果。除此之外，受经典几何建模理论的启发，马利庄团队（Tang等，2022）提出一种创新性的关键点—骨架—形状的点云补全网络，利用3维物体的几何和结构化拓扑信息来辅助点云完整结构的恢复。该方法包括关键点定位、骨架生成和形状细化3个步骤，这种递进式的网络结构有效提升了点云补全的准确性和精度。

4.4.2 真实扫描点云跨域补全

Chen等人（2020）利用两个自动编码器来重构虚拟完整的点云和真实的残缺点云，并使用映射函数将真实点云的编码映射到虚拟完整空间中来补全点云。然后，设计了对抗性损失以确保目标样本的映射隐藏编码与源样本共享相同的分布。Wen等人（2021）在输入的潜在空间编码和完整点云的空间编码之间设计了双向循环转换机制，并引入了从完整分支到残缺分支的反向映射功能，以进一步保持形状一致性。Cai等人（2022）提出一种统一的结构化网络，将部分点云解耦为完整的形状因子和遮挡因子，可以有效提高形状完成精度，完整形状因子和遮挡因子两者逐元素乘积用以重建残缺点云，补全过程仅使用完整形状因子，为学习该结构化潜空间提出了一系列约束条件，包括结构化排名正则、潜码交换以及潜码分布监督。马利庄团队（Gong等，2022）结合回归与优化两个阶段提高补全点云与输入残缺点云间的一致性，加速模型推理速度。其中，特征解耦进行域级别的对齐，残缺点云特征被解耦为域、形状和遮挡3个因子。残缺点云的遮挡因子与观察视角强相关，故设计自监督视点预测任务以学习遮挡因子；域因子与形状因子分别代表域风格与点云形状，故使用域判别器结合梯度反转同时训练域因子与形状因子。

5 发展趋势与展望

得益于激光雷达等远距离传感器和结构光等近距离传感器的发展，3维点云场景数据的获取变得愈发便利。相比于2维图像，点云数据受外界光照和成像距离的影响较小，并能够更为有效地反映3维真实世界的空间结构，呈现出更为丰富的几何信息、形状信息和尺度信息。凭借这些优势，3维场景理解与重建技术能够使机器以3维空间的思维来记录和理解真实世界，这对于工业生产自动化、城市管理信息化以及生活娱乐智能化有着重要意义。3维场景理解与重建系列技术可广泛应用于场景模型重建、SLAM、机器人感知、路况分析和历史文物保护等场景中。为此，众多研究聚焦3维点云的场景理解与重建中点云特征提取与匹配融合、场景理解与语义分割以及扫描点云补全等关键问题，取得了一系列重大进展。但是，目前仍然存在扫描场景差距大、高精度3维场景计算开销大的问题，极大程度影响真实场景应用精度；点云数据表征非结构化、真实物体形态多种多样，要求补全方法具有极强的鲁棒性和泛化能力；对于3维场景中存在的人物，要求进一步探索场景与人物行为之间的联系。为进一步

发展相关技术，促进落地应用，仍需针对室外点云有限标注下的分割、大规模场景形状与纹理补全以及3维场景下人物行为理解生成等问题进行更深层次的探索。

在场景点云分割领域中，虽然现有方法模型已经展现出了优秀的性能，但依旧存在许多挑战。例如，在基于激光雷达扫描的室外场景语义分割中，点云的特征较弱，大多仅包含3维坐标和反射强度，加剧了算法区分点语义类别的难度；在真实应用场景下，不同物体所对应的点云规模差别很大，对模型分割不同尺度的点云物体提出了极高要求；由于点云非结构化的性质，催生了多视图、2D/3D投影等多种点云的数据表征类型，每种数据类别有着各自的优势，但也存在着各式各样的缺点；相比于图像分割模型，训练点云分割模型需要更大的计算开销，对模型训练时长和硬件资源有着更高的要求。此外，由于分割任务的定义，对3D点云的数据标注要求较为严格，需要进行逐点的标签标注，然而3D点云的标注是昂贵、费力且易出错的。因此，在有限标签数据的条件下，研究快速且精准的点云分割算法和框架是该领域的研究重点。

在场景重建领域，随着人工智能技术的发展，场景重建的真实还原度和纹理细节方面得到了明显的提升，但在基于图像视频的场景重建、大规模场景点云补全等任务内还存在许多有待完善的问题。首先，在基于图像视频的场景重建中，不同相机或不同场景条件下的场景深度估计精度难以得到保障，尤其是被遮挡的物体轮廓部分，虽然在图像中往往占比较小，却是场景重建的重要线索；当针对视频数据进行场景重建时，需要关注如何解决视频帧数据对应的问题；对于点云的稠密化，需要解决的不仅是如何从原本稀疏的点云来生成稠密的点云，更重要的是如何保证生成的点能够均匀且准确地附着在物体的表面。其次，在大规模场景点云补全中，需要关注如何解决大规模点云场景整体特征提取与物体间信息传递的问题；如何解决扫描数据中密度差异巨大的问题，以及如何处理大规模点云中细粒度特征重建的问题。这些问题都是值得未来研究的重要方向。

在3维场景理解与重建的基础上，对真实世界的数字化建模更要求能够探索3维场景与人之间的关系，对场景中人的行为进行理解甚至能够对场景中的人物进行模拟和动作生成。但点云场景的非结构化表征与人体行为的多样性都使得人体与场景之间的关联很难通过简单的显式表达式进行定义。因此，在基于场景的人物行为理解与生成中，如何更好地建模3维点云场景与人物行为之间的关联性和一致性；如何在3维场景下生成人物长时间且真实的行为动作；在保证生成的人物行为在3D场景中是自然且合理的同时如何提升动作合成的效率以实现分钟级别的动作生成速度等仍需要后续的工作进行进一步探索。

综上所述，基于3维点云的场景理解与重建的相关技术面临着许多亟待解决的问题和挑战。在未来，场景点云语义分割的研究应当综合考虑3D真实物理世界在不同视角下的映射，并设计对硬件资源更为友好的算法框架；场景重建领域的研究重心应在于重建出细致化且更为真实的大规模场景；对于3维场景和人的关系，重心在于理解和遵循两者之间存在的规律，建模人与3维场景之间更为精细化的联系，以及探索快速生成自然且合理的人物行为的模型。毫无疑问的是，3维点云的场景理解与重建对国民日常生活、工业生产和国防建设有着巨大的经济和社会价值。期待点云特征提取、场景点云分割和扫描点云补全等相关领域得到进一步发展，在数据集建设、模型计算优化以及鲁棒性和可解释性上取得更大的前进，为实现自动驾驶、数字工厂和智慧城市等方面提供持续且可靠的动力。

致　谢　本文由中国图象图形学学会动画与数字娱乐专业委员会组织撰写，该专委会更多详情请见链接：http://www.csig.org.cn/detail/2387。

参考文献（References）

Achlioptas P, Diamanti O, Mitliagkas I and Guibas L. 2018. Learning representations and generative models for 3D point clouds//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR: 40-49

Afham M, Dissanayake I, Dissanayake D, Dharmasiri A, Thilakarathna K and Rodrigo R. 2022. CrossPoint: self-supervised cross-modal contrastive learning for 3D point cloud understanding//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 9892-9902 [DOI: 10.1109/CVPR52688.2022.00967]

Ao S, Hu Q Y, Yang B, Markham A and Guo Y L. 2021. SpinNet:

learning a general surface descriptor for 3D point cloud registration//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11748-11757 [DOI: 10.1109/CVPR46437.2021.01158]

Armeni I, Sener O, Zamir A R, Jiang H, Brilakis I, Fischer M and Savarese S. 2016. 3D semantic parsing of large-scale indoor spaces//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1534-1543 [DOI: 10.1109/CVPR.2016.170]

Bai X Y, Luo Z X, Zhou L, Fu H B, Quan L and Tai C L. 2020. D3Feat: joint learning of dense detection and description of 3D local features//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 6358-6366 [DOI: 10.1109/CVPR42600.2020.00639]

Behley J, Garbade M, Milioto A, Quenzel J, Behnke S, Stachniss C and Gall J. 2019. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9296-9306 [DOI: 10.1109/ICCV.2019.00939]

Bian Y K, Hui L, Qian J J and Xie J. 2022. Unsupervised domain adaptation for point cloud semantic segmentation via graph matching//Proceedings of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems. Kyoto, Japan: IEEE: 9899-9904 [DOI: 10.1109/IROS47612.2022.9981603]

Bronstein M M, Bruna J, LeCun Y, Szlam A and Vandergheynst P. 2017. Geometric deep learning: going beyond Euclidean data. IEEE Signal Processing Magazine, 34(4): 18-42 [DOI: 10.1109/MSP.2017.2693418]

Caesar H, Bankiti V, Lang A H, Vora S, Liong V E, Xu Q, Krishnan A, Pan Y, Baldan G and Beijbom O. 2020. nuScenes: a multimodal dataset for autonomous driving//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11618-11628 [DOI: 10.1109/CVPR42600.2020.01164]

Cai Y J, Lin K Y, Zhang C, Wang Q, Wang X G and Li H S. 2022. Learning a structured latent space for unsupervised point cloud completion//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5533-5543 [DOI: 10.1109/CVPR52688.2022.00546]

Chang A X, Funkhouser T, Guibas L, Hanrahan P, Huang Q X, Li Z M, Savarese S, Savva M, Song S R, Su H, Xiao J X, Yi L and Yu F. 2015. ShapeNet: an information-rich 3D model repository [EB/OL]. [2023-01-03]. https://arxiv.org/pdf/1512.03012.pdf

Chen C, Li G B, Xu R J, Chen T S, Wang M and Lin L. 2019. ClusterNet: deep hierarchical cluster network with rigorously rotation-invariant representation for point cloud analysis//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4989-4997 [DOI: 10.1109/CVPR.2019.00513]

Chen H and Bhanu B. 2007. 3D free-form object recognition in range images using local surface patches. Pattern Recognition Letters, 28(10): 1252-1262 [DOI: 10.1016/j.patrec.2007.02.009]

Chen S Y, Fang J M, Zhang Q, Liu W Y and Wang X G. 2021. Hierarchical aggregation for 3D instance segmentation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 15447-15456 [DOI: 10.1109/ICCV48922.2021.01518]

Chen S Y, Wang X G, Cheng T H, Zhang W Q, Zhang Q, Huang C and Liu W Y. 2022. AziNorm: exploiting the radial symmetry of point cloud for azimuth-normalized 3D perception//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 6377-6386 [DOI: 10.1109/CVPR52688.2022.00628]

Chen X L, Chen B Q and Mitra N J. 2020. Unpaired point cloud completion on real scans using adversarial training//Proceedings of the 8th International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net

Choy C, Gwak J and Savarese S. 2019a. 4D spatio-temporal ConvNets: minkowski convolutional neural networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3070-3079 [DOI: 10.1109/CVPR.2019.00319]

Choy C, Park J and Koltun V. 2019b. Fully convolutional geometric features//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 8957-8965 [DOI: 10.1109/ICCV.2019.00905]

Cortinhal T, Tzelepis G and Erdal Aksoy E. 2020. SalsaNext: fast, uncertainty-aware semantic segmentation of LiDAR point clouds//Proceedings of the 15th International Symposium on Visual Computing. San Diego, USA: Springer: 207-222 [DOI: 10.1007/978-3-030-64559-5_16]

Dai A, Chang A X, Savva M, Halber M, Funkhouser T and Nießner M. 2017. ScanNet: richly-annotated 3D reconstructions of indoor scenes//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2432-2443 [DOI: 10.1109/CVPR.2017.261]

Defferrard M, Bresson X and Vandergheynst P. 2016. Convolutional neural networks on graphs with fast localized spectral filtering//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc.: 3844-3852

Drost B, Ulrich M, Navab N and Ilic S. 2010. Model globally, match locally: efficient and robust 3D object recognition//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE: 998-1005 [DOI: 10.1109/CVPR.2010.5540108]

Feng M T, Zhang L, Lin X F, Gilani S Z and Mian A. 2020. Point attention network for semantic segmentation of 3D point clouds. Pattern

Recognition, 107: #107446 [DOI: 10.1016/j. patcog. 2020. 107446]

Flint A, Dick A and van den Hengel A. 2007. Thrift: local 3D structure recognition//Proceedings of the 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007). Glenelg, Australia: IEEE: 182-188 [DOI: 10.1109/DICTA.2007.4426794]

Frome A, Huber D, Kolluri R, Bülow T and Malik J. 2004. Recognizing objects in range data using regional point descriptors//Proceedings of the 8th European Conference on Computer Vision. Prague, Czech Republic: Springer: 224-237 [DOI: 10.1007/978-3-540-24672-5_18]

Geiger A, Lenz P and Urtasun R. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 3354-3361 [DOI: 10.1109/CVPR.2012.6248074]

Gojcic Z, Zhou C F, Wegner J D and Wieser A. 2019. The perfect match: 3D point cloud matching with smoothed densities//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5540-5549 [DOI: 10.1109/CVPR.2019.00569]

Gong J Y, Liu F Q, Xu J C, Wang M, Tan X, Zhang Z Z, Yi R, Song H C, Xie Y and Ma L Z. 2022. Optimization over disentangled encoding: unsupervised cross-domain point cloud completion via occlusion factor manipulation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 517-533 [DOI: 10.1007/978-3-031-20086-1_30]

Gong J Y, Xu J C, Tan X, Song H C, Qu Y Y, Xie Y and Ma L Z. 2021a. Omni-supervised point cloud segmentation via gradual receptive field component reasoning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11668-11677 [DOI: 10.1109/CVPR46437.2021.01150]

Gong J Y, Xu J C, Tan X, Zhou J, Qu Y Y, Xie Y and Ma L Z. 2021b. Boundary-aware geometric encoding for semantic segmentation of point clouds//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI: 1424-1432 [DOI: 10.1609/aaai.v35i2.16232]

Graham B, Engelcke M and van der Maaten L. 2018. 3D semantic segmentation with submanifold sparse convolutional networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 9224-9232 [DOI: 10.1109/CVPR.2018.00961]

Guo M H, Cai J X, Liu Z N, Mu T J, Martin R R and Hu S M. 2021. PCT: point cloud transformer. Computational Visual Media, 7(2): 187-199 [DOI: 10.1007/s41095-021-0229-5]

Guo Y L, Sohel F, Bennamoun M, Lu M and Wan J W. 2013. Rotational projection statistics for 3D local surface description and object recognition. International Journal of Computer Vision, 105(4): 63-86 [DOI: 10.1007/s11263-013-0627-y]

Hackel T, Savinov N, Ladicky L, Wegner J D, Schindler K and Pollefeys M. 2017. Semantic3D.net: a new large-scale point cloud classification benchmark [EB/OL]. [2023-01-03]. http://arxiv.org/pdf/1704.03847.pdf

Hassani K and Haley M. 2019. Unsupervised multi-task feature learning on point clouds//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 8159-8170 [DOI: 10.1109/ICCV.2019.00825]

Hosseinzadeh M, Li K J, Latif Y and Reid I. 2019. Real-time monocular object-model aware sparse SLAM//Proceedings of 2019 International Conference on Robotics and Automation. Montreal, Canada: IEEE: 7123-7129 [DOI: 10.1109/ICRA.2019.8793728]

Houston J, Zuidhof G, Bergamini L, Ye Y W, Chen L, Jain A, Omari S, Iglovikov V and Ondruska P. 2020. One thousand and one hours: self-driving motion prediction dataset//Proceedings of the 4th Conference on Robot Learning. Cambridge, USA: PMLR: 409-418

Hua B S, Tran M K and Yeung S K. 2018. Pointwise convolutional neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 984-993 [DOI: 10.1109/CVPR.2018.00109]

Jaritz M, Vu T H, de Charette R, Wirbel E and Pérez P. 2020. xMUDA: cross-modal unsupervised domain adaptation for 3D semantic segmentation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 12602-12611 [DOI: 10.1109/CVPR42600.2020.01262]

Jiang L, Zhao H S, Shi S S, Liu S, Fu C W and Jia J Y. 2020. PointGroup: dual-set point grouping for 3D instance segmentation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4866-4875 [DOI: 10.1109/CVPR42600.2020.00492]

Johnson A E and Hebert M. 1999. Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(5): 433-449 [DOI: 10.1109/34.765655]

Kim S, Park J and Han B. 2020. Rotation-invariant local-to-global representation learning for 3D point cloud//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #685

Kong L D, Ren J W, Pan L and Liu Z W. 2022. LaserMix for semi-supervised LiDAR semantic segmentation [EB/OL]. [2022-06-30]. https://arxiv.org/pdf/2207.00026.pdf

Kundu A, Genova K, Yin X Q, Fathi A, Pantofaru C, Guibas L, Tagliasacchi A, Dellaert F and Funkhouser T. 2022. Panoptic neural fields: a semantic object-aware neural scene representation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12861-12871

［DOI：10.1109/CVPR52688.2022.01253］

Lei H，Akhtar N and Mian A. 2020. SegGCN：efficient 3D point cloud segmentation with fuzzy spherical kernel//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：11608-11617［DOI：10.1109/CVPR42600.2020.0116］

Li F R，Fujiwara K，Okura F and Matsushita Y. 2021a. A closer look at rotation-invariant deep point cloud analysis//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：16198-16207［DOI：10.1109/ICCV48922.2021.01591］

Li H Y，Sun Z X，Wu Y J and Song Y C. 2021b. Semi-supervised point cloud segmentation using self-training with label confidence prediction. Neurocomputing，437：227-237［DOI：10.1016/j. neucom.2021.01.091］

Li J W and Zhan J W. 2022. Review on 3D point cloud registration method. Journal of Image and Graphics，27(2)：349-367（李建微，占家旺. 2022. 3维点云配准方法研究进展. 中国图象图形学报，27(2)：349-367）［DOI：10.11834/jig.210243］

Li J X，Chen B M and Lee G H. 2018a. SO-Net：self-organizing network for point cloud analysis//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，USA：IEEE：9397-9406［DOI：10.1109/CVPR.2018.00979］

Li M T，Xie Y，Shen Y H，Ke B，Qiao R Z，Ren B，Lin S H and Ma L Z. 2022a. HybridCR：weakly-supervised 3D point cloud semantic segmentation via hybrid contrastive regularization//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，USA：IEEE：14910-14919［DOI：10.1109/CVPR52688.2022.01451］

Li X Y，Zhang G，Pan H Y and Wang Z H. 2022b. CPGNet：cascade point-grid fusion network for real-time LiDAR semantic segmentation//Proceedings of 2022 International Conference on Robotics and Automation. Philadelphia，USA：IEEE：11117-11123［DOI：10.1109/ICRA46639.2022.9811767］

Li Y Y，Bu R，Sun M C，Wu W，Di X H and Chen B Q. 2018b. PointCNN：convolution on *X*-transformed points//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal，Canada：Curran Associates Inc.：828-838

Liu C and Furukawa Y. 2019a. MASC：multi-scale affinity with sparse convolution for 3D instance segmentation［EB/OL］.［2023-01-03］. https://arxiv.org/pdf/1902.04478.pdf

Liu Q，Jiang N J，Lu J B，Chen M G，Yi R and Ma L Z. 2022. ScatterNet：point cloud learning via scatters//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa，Portugal：ACM：5611-5619［DOI：10.1145/3503161.3548354］

Liu S，Huang S Y，Cheng H H，Shen J Y and Chen S Y. 2021. A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation. Journal of Image and Graphics，26(12)：2848-2859（刘盛，黄圣跃，程豪豪，沈家瑜，陈胜勇. 2021. 结合空间深度卷积和残差的大尺度点云场景分割. 中国图象图形学报，26(12)：2848-2859）［DOI：10.11834/jig.200477］

Liu X P，Ma Y X，Xu K，Wan J W and Guo Y L. 2022. Multi-scale transformer based point cloud completion network. Journal of Image and Graphics，27(2)：538-549（刘心溥，马燕新，许可，万建伟，郭裕兰. 2022. 嵌入Transformer结构的多尺度点云补全. 中国图象图形学报，27(2)：538-549）［DOI：10.11834/jig.210510］

Liu Y C，Fan B，Xiang S M and Pan C H. 2019b. Relation-shape convolutional neural network for point cloud analysis//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，USA：IEEE：8887-8896［DOI：10.1109/CVPR.2019.00910］

Liu Z，Hu H，Cao Y，Zhang Z and Tong X. 2020. A closer look at local aggregation operators in point cloud analysis//Proceedings of the 16th European Conference on Computer Vision. Glasgow，UK：Springer：326-342［DOI：10.1007/978-3-030-58592-1_20］

Liu Z J，Tang H T，Lin Y J and Han S. 2019c. Point-voxel CNN for efficient 3D deep learning//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver，Canada：Curran Associates Inc.：#87

Long J，Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston，USA：IEEE：3431-3440［DOI：10.1109/CVPR.2015.7298965］

Long X X，Cheng X J，Zhu H，Zhang P J，Liu H M，Li J，Zheng L T，Hu Q Y，Liu H，Cao X，Yang R G，Wu Y H，Zhang G F，Liu Y B，Xu K，Guo Y L and Chen B Q. 2021. Recent progress in 3D vision. Journal of Image and Graphics，26(6)：1389-1428（龙霄潇，程新景，朱昊，张朋举，刘浩敏，李俊，郑林涛，胡庆拥，刘浩，曹汛，杨睿刚，吴毅红，章国锋，刘烨斌，徐凯，郭裕兰，陈宝权. 2021. 3维视觉前沿进展. 中国图象图形学报，26(6)：1389-1428）［DOI：10.11834/jig.210043］

Ma Y X，Wang T，Bai X Y，Yang H T，Hou Y N，Wang Y M，Qiao Y，Yang R G，Manocha D and Zhu X G. 2022. Vision-centric BEV perception：a survey［EB/OL］.［2022-08-04］. https://arxiv.org/pdf/2208.02797.pdf

Masci J，Boscaini D，Bronstein M M and Vandergheynst P. 2015. Geodesic convolutional neural networks on riemannian manifolds//Proceedings of 2015 IEEE International Conference on Computer Vision Workshops. Santiago，Chile：IEEE：832-840［DOI：10.1109/ICCVW.2015.112］

Mendoza A，Apaza A，Sipiran I and López C. 2020. Refinement of predicted missing parts enhance point cloud completion［EB/OL］.［2023-01-03］. https://arxiv.org/pdf/2010.04278.pdf

Miao Y W，Zhang L，Liu J Z，Wang J R and Liu F C. 2021. An end-to-end shape-preserving point completion network. IEEE Computer Graphics and Applications，41(3)：20-33［DOI：10.1109/MCG.2021.3065533］

Milioto A, Vizzo I, Behley J and Stachniss C. 2019. RangeNet ++: fast and accurate LiDAR semantic segmentation//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE: 4213-4220 [DOI: 10.1109/IROS40897.2019.8967762]

Monti F, Boscaini D, Masci J, Rodolà E, Svoboda J and Bronstein M M. 2017. Geometric deep learning on graphs and manifolds using mixture model CNNs//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5425-5434 [DOI: 10.1109/CVPR.2017.576]

Pan L, Chen X Y, Cai Z G, Zhang J Z, Zhao H Y, Yi S and Liu Z W. 2021. Variational relational point completion network//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 8520-8529 [DOI: 10.1109/CVPR46437.2021.00842]

Park C, Jeong Y, Cho M and Park J. 2022. Fast point transformer//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16928-16937 [DOI: 10.1109/CVPR52688.2022.01644]

Peng Y J, Chang M, Wang Q, Qian Y L, Zhang Y K, Wei M Q and Liu X Y. 2020. Sparse-to-dense multi-encoder shape completion of unstructured point cloud. IEEE Access, 8: 30969-30978 [DOI: 10.1109/ACCESS.2020.2973003]

Qi C R, Su H, Mo K and Guibas L J. 2017a. PointNet: deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 77-85 [DOI: 10.1109/CVPR.2017.16]

Qi C R, Yi L, Su H and Guibas L J. 2017b. PointNet++: deep hierarchical feature learning on point sets in a metric space//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 5105-5114

Qin Z, Yu H, Wang C J, Guo Y L, Peng Y X and Xu K. 2022. Geometric transformer for fast and robust point cloud registration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11133-11142 [DOI: 10.1109/CVPR52688.2022.01086]

Ravanbakhsh S, Schneider J G and Póczos B. 2017. Deep learning with sets and point clouds//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: OpenReview.net: 1-12

Rusu R B, Blodow N and Beetz M. 2009. Fast point feature histograms (FPFH) for 3D registration//Proceedings of 2009 IEEE International Conference on Robotics and Automation. Kobe, Japan: IEEE: 3212-3217 [DOI: 10.1109/ROBOT.2009.5152473]

Rusu R B, Blodow N, Marton Z S and Beetz M. 2008. Aligning point cloud views using persistent feature histograms//Proceedings of 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Nice, France: IEEE: 3384-3391 [DOI: 10.1109/IROS.2008.4650967]

Salti S, Tombari F and Di Stefano L. 2014. SHOT: unique signatures of histograms for surface and texture description. Computer Vision and Image Understanding, 125: 251-264 [DOI: 10.1016/j.cviu.2014.04.011]

Sarlin P E, DeTone D, Malisiewicz T and Rabinovich A. 2020. SuperGlue: learning feature matching with graph neural networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4937-4946 [DOI: 10.1109/CVPR42600.2020.00499]

Sautier C, Puy G, Gidaris S, Boulch A, Bursuc A and Marlet R. 2022. Image-to-lidar self-supervised distillation for autonomous driving data//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 9891-9901 [DOI: 10.1109/CVPR52688.2022.00966]

Schutt P, Rosu R A and Behnke S. 2022. Abstract flow for temporal semantic segmentation on the permutohedral lattice//Proceedings of 2022 International Conference on Robotics and Automation. Philadelphia, USA: IEEE: 5139-5145 [DOI: 10.1109/ICRA46639.2022.9811818]

Silberman N, Hoiem D, Kohli P and Fergus R. 2012. Indoor segmentation and support inference from RGBD images//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer: 746-760 [DOI: 10.1007/978-3-642-33715-4_54]

Simonovsky M and Komodakis N. 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 29-38 [DOI: 10.1109/CVPR.2017.11]

Song S R, Lichtenberg S P and Xiao J X. 2015. SUN RGB-D: a RGB-D scene understanding benchmark suite//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 567-576 [DOI: 10.1109/CVPR.2015.7298655]

Spurek P, Kasymov A, Mazur M, Janik D, Tadeja S, Struski Ł, Tabor J and Trzciński T. 2021. HyperPocket: generative point cloud completion//Proceedings of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems. Kyoto, Japan: IEEE: 6848-6853 [DOI: 10.1109/IROS47612.2022.9981829]

Sun J M, Shen Z H, Wang Y, Bao H J and Zhou X W. 2021. LoFTR: detector-free local feature matching with transformers//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 8918-8927 [DOI: 10.1109/CVPR46437.2021.00881]

Sun P, Kretzschmar H, Dotiwalla X, Chouard A, Patnaik V, Tsui P, Guo J, Zhou Y, Chai Y, Caine B, Vasudevan V, Han W, Ngiam J, Zhao H, Timofeev A, Ettinger S, Krivokon M, Gao A, Joshi A, Zhang Y, Shlens J, Chen Z F and Anguelov D. 2020. Scalability in perception for autonomous driving: waymo open dataset//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and

Pattern Recognition. Seattle, USA: IEEE: 2443-2451 [DOI: 10.1109/CVPR42600.2020.00252]

Tan X, Lin J Y, Xu K, Chen P, Ma L Z and Lau R W H. 2023. Mirror detection with the visual chirality cue. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3): 3492-3504 [DOI: 10.1109/TPAMI.2022.3181030]

Tan X, Xu K, Cao Y, Zhang Y H, Ma L Z and Lau R W H. 2021. Night-time scene parsing with a large real dataset. IEEE Transactions on Image Processing, 30: 9085-9098 [DOI: 10.1109/TIP.2021.3122004]

Tang H T, Liu Z J, Zhao S Y, Lin Y J, Lin J, Wang H R and Han S. 2020. Searching efficient 3D architectures with sparse point-voxel convolution//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 685-702 [DOI: 10.1007/978-3-030-58604-1_41]

Tang J S, Gong Z J, Yi R, Xie Y and Ma L Z. 2022. LAKe-Net: topology-aware point cloud completion by localizing aligned keypoints//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: 1716-1725 [DOI: 10.1109/cvpr52688.2022.00177]

Tarvainen A and Valpola H. 2017. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 1195-1204

Te G S, Hu W, Zheng A M and Guo Z M. 2018. RGCNN: regularized graph CNN for point cloud segmentation//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 746-754 [DOI: 10.1145/3240508.3240621]

Thomas H, Qi C R, Deschaud J E, Marcotegui B, Goulette F and Guibas L. 2019. KPConv: flexible and deformable convolution for point clouds//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6410-6419 [DOI: 10.1109/ICCV.2019.00651]

Tombari F, Salti S and Di Stefano L. 2010. Unique shape context for 3D data description//Proceedings of 2010 ACM Workshop on 3D Object Retrieval. Firenze, Italy: ACM: 57-62 [DOI: 10.1145/1877808.1877821]

Unal O, Dai D X and Van Gool L. 2022. Scribble-supervised LiDAR semantic segmentation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 2687-2697 [DOI: 10.1109/CVPR52688.2022.00272]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010

Vu T, Kim K, Luu T M, Nguyen T and Yoo C D. 2022. SoftGroup for 3D instance segmentation on point clouds//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 2698-2707 [DOI: 10.1109/CVPR52688.2022.00273]

Wang C, Samari B and Siddiqi K. 2018a. Local spectral graph convolution for point set feature learning//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 56-71 [DOI: 10.1007/978-3-030-01225-0_4]

Wang W Y, Huang Q G, You S Y, Yang C and Neumann U. 2017. Shape inpainting using 3D generative adversarial network and recurrent convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2317-2325 [DOI: 10.1109/ICCV.2017.252]

Wang W Y, Yu R, Huang Q G and Neumann U. 2018b. SGPN: similarity group proposal network for 3D point cloud instance segmentation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2569-2578 [DOI: 10.1109/CVPR.2018.00272]

Wang X G, Ang M H and Lee G H. 2020a. Cascaded refinement network for point cloud completion//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 787-796 [DOI: 10.1109/cvpr42600.2020.00087]

Wang X G, Ang M H and Lee G H. 2020b. Point cloud completion by learning shape priors//Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, USA: IEEE: 10719-10726 [DOI: 10.1109/IROS45743.2020.9340862]

Wang Y, Sun Y B, Liu Z W, Sarma S E, Bronstein M M and Solomon J M. 2019. Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics, 38(5): #146 [DOI: 10.1145/3326362]

Wen X, Han Z Z, Cao Y P, Wan P F, Zheng W and Liu Y S. 2021. Cycle4Completion: unpaired point cloud completion using cycle transformation with missing region coding//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13075-13084 [DOI: 10.1109/CVPR46437.2021.01288]

Wu B C, Zhou X Y, Zhao S C, Yue X Y and Keutzer K. 2019a. SqueezeSegV2: improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud//Proceedings of 2019 International Conference on Robotics and Automation. Montreal, Canada: IEEE: 4376-4382 [DOI: 10.1109/ICRA.2019.8793495]

Wu H and Miao Y B. 2021a. Cross-regional attention network for point cloud completion//Proceedings of the 25th International Conference on Pattern Recognition. Milan, Italy: IEEE: 10274-10280 [DOI: 10.1109/ICPR48806.2021.9413104]

Wu H, Miao Y B and Fu R C. 2021b. Point cloud completion using multiscale feature fusion and cross-regional attention. Image and Vision Computing, 111: #104193 [DOI: 10.1016/J. IMAVIS. 2021.104193]

Wu W X, Qi Z G and Li F X. 2019b. PointConv: deep convolutional net-

works on 3D point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 9613-9622 [DOI: 10.1109/CVPR.2019.00985]

Wu X Y, Lao Y X, Jiang L, Liu X H and Zhao H S. 2022. Point transformer V2: grouped vector attention and partition-based pooling// Proceedings of the 36th Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.

Wu Z R, Song S R, Khosla A, Yu F, Zhang L G, Tang X O and Xiao J X. 2015. 3D ShapeNets: a deep representation for volumetric shapes//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1912-1920 [DOI: 10.1109/CVPR.2015.7298801]

Xia Y, Liu W, Luo Z, Xu Y and Stilla U. 2020. Completion of sparse and partial point clouds of vehicles using a novel end-to-end network//Proceedings of 2020 ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. [s.l.]: ISPRS: 933-940 [DOI: 10.5194/isprs-annals-v-2-2020-933-2020]

Xia Y Q, Xia Y, Li W, Song R, Cao K L and Stilla U. 2021. ASFM-Net: asymmetrical Siamese feature matching network for point completion//Proceedings of the 29th ACM International Conference on Multimedia. Chengdu, China: ACM: 1938-1947 [DOI: 10.1145/3474085.3475348]

Xu J Y, Tang X, Zhu Y S, Sun J and Pu S L. 2021a. SGMNet: learning rotation-invariant point cloud representations via sorted Gram matrix//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 10448-10457 [DOI: 10.1109/ICCV48922.2021.01030]

Xu J Y, Zhang R X, Dou J, Zhu Y S, Sun J and Pu S L. 2021b. RPVNet: a deep and efficient range-point-voxel fusion network for LiDAR point cloud segmentation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 16004-16013 [DOI: 10.1109/ICCV48922.2021.01572]

Xu X and Lee G H. 2020. Weakly supervised semantic point cloud segmentation: towards 10×fewer labels//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13703-13712 [DOI: 10.1109/CVPR42600.2020.01372]

Xu Y F, Fan T Q, Xu M Y, Zeng L and Qiao Y. 2018. SpiderCNN: deep learning on point sets with parameterized convolutional filters// Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 90-105 [DOI: 10.1007/978-3-030-01237-3_6]

Yan X, Zheng C D, Li Z, Wang S and Cui S G. 2020. PointASNL: robust point clouds processing using nonlocal neural networks with adaptive sampling//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5588-5597 [DOI: 10.1109/CVPR42600.2020.00563]

Yang B, Wang J N, Clark R, Hu Q Y, Wang S, Markham A and Trigoni N. 2019a. Learning object bounding boxes for 3D instance segmentation on point clouds//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #650

Yang J C, Zhang Q, Ni B B, Li L G, Liu J X, Zhou M D and Tian Q. 2019b. Modeling point clouds with self-attention and gumbel subset sampling//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3318-3327 [DOI: 10.1109/CVPR.2019.00344]

Ye M S, Wan R, Xu S J, Cao T Y and Chen Q F. 2022. Efficient point cloud segmentation with geometry-aware sparse networks//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 196-212 [DOI: 10.1007/978-3-031-19842-7_12]

Ye M S, Xu S J, Cao T Y and Chen Q F. 2021. DRINet: a dual-representation iterative learning network for point cloud segmentation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 7427-7436 [DOI: 10.1109/ICCV48922.2021.00735]

Yew Z J and Lee G H. 2022. REGTR: end-to-end point cloud correspondences with transformers//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 6667-6676 [DOI: 10.1109/CVPR52688.2022.00656]

Yi L, Su H, Guo X W and Guibas L. 2017. SyncSpecCNN: synchronized spectral CNN for 3D shape segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6584-6592 [DOI: 10.1109/CVPR.2017.697]

Yi L, Zhao W, Wang H, Sung M and Guibas L J. 2019. GSPN: generative shape proposal network for 3D instance segmentation in point cloud//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3942-3951 [DOI: 10.1109/CVPR.2019.00407]

You Y, Lou Y J, Liu Q, Tai Y W, Ma L Z, Lu C W and Wang W M. 2020. Pointwise rotation-invariant network with adaptive sampling and 3D spherical voxel convolution//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press: 12717-12724 [DOI: 10.1609/aaai.v34i07.6965]

You Y, Lou Y J, Shi R X, Liu Q, Tai Y W, Ma L Z, Wang W M and Lu C W. 2022. PRIN/SPRIN: on extracting point-wise rotation invariant features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44 (12): 9489-9502 [DOI: 10.1109/TPAMI.2021.3130590]

Yu H, Li F, Saleh M, Busam B and Ilic S. 2021a. CoFiNet: reliable coarse-to-fine correspondences for robust pointcloud registration// Proceedings of the 35th Advances in Neural Information Processing Systems. [s.l.]: Curran Associates, Inc.: 23872-23884

Yu R X, Wei X, Tombari F and Sun J. 2020. Deep positional and relational feature learning for rotation-invariant point cloud analysis//

Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 217-233 [DOI: 10.1007/978-3-030-58607-2_13]

Yu X M, Rao Y M, Wang Z Y, Liu Z Y, Lu J W and Zhou J. 2021b. PoinTr: diverse point cloud completion with geometry-aware transformers//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 12478-12487 [DOI: 10.1109/ICCV48922.2021.01227]

Yu X M, Tang L L, Rao Y M, Huang T J, Zhou J and Lu J W. 2022. Point-BERT: pre-training 3D point cloud transformers with masked point modeling//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 19291-19300 [DOI: 10.1109/CVPR52688.2022.01871]

Zaheer M, Kottur S, Ravanbhakhsh S, Póczos B, Salakhutdinov R and Smola A J. 2017. Deep sets//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 3394-3404

Zamorski M, Zięba M and Świątek J. 2020. Generative modeling in application to point cloud completion//Proceedings of the 19th International Conference on Artificial Intelligence and Soft Computing. Zakopane, Poland: Springer: 292-302 [DOI: 10.1007/978-3-030-61401-0_28]

Zeng A, Song S R, Nießner M, Fisher M, Xiao J X and Funkhouser T. 2017. 3DMatch: learning local geometric descriptors from RGB-d reconstructions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 199-208 [DOI: 10.1109/CVPR.2017.29]

Zhang C, Wan H C, Shen X Y and Wu Z Z. 2022. PatchFormer: an efficient point transformer with patch attention//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11789-11798 [DOI: 10.1109/CVPR52688.2022.01150]

Zhang J M, Yu M Y, Vasudevan R and Johnson-Roberson M. 2020a. Learning rotation-invariant representations of point clouds using aligned edge convolutional neural networks//Proceedings of 2020 International Conference on 3D Vision (3DV). Fukuoka, Japan: IEEE: 200-209 [DOI: 10.1109/3DV50981.2020.00030]

Zhang J Z, Chen X Y, Cai Z G, Pan L, Zhao H Y, Yi S, Yeo C K, Dai B and Loy C C. 2021a. Unsupervised 3D shape completion through GAN inversion//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 1768-1777 [DOI: 10.1109/CVPR46437.2021.00181]

Zhang W X, Yan Q G and Xiao C X. 2020b. Detail preserved point cloud completion via separated feature aggregation//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 512-528 [DOI: 10.1007/978-3-030-58595-2_31]

Zhang Y, Zhou Z X, David P, Yue X Y, Xi Z R, Gong B Q and Foroosh H. 2020c. PolarNet: an improved grid representation for online LiDAR point clouds semantic segmentation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 9598-9607 [DOI: 10.1109/CVPR42600.2020.00962]

Zhang Y C, Qu Y Y, Xie Y, Li Z H, Zheng S S and Li C H. 2021b. Perturbed self-distillation: weakly supervised large-scale point cloud semantic segmentation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 15500-15508 [DOI: 10.1109/ICCV48922.2021.01523]

Zhang Y N, Huang D and Wang Y H. 2021c. PC-RGNN: point cloud completion and graph neural network for 3D object detection//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI: 3430-3437 [DOI: 10.1609/AAAI.v35i4.16456]

Zhang Z Y, Hua B S, Rosen D W and Yeung S K. 2019. Rotation invariant convolutions for 3D point clouds deep learning//Proceedings of 2019 International Conference on 3D Vision (3DV). Quebec City, Canada: IEEE: 204-213 [DOI: 10.1109/3DV.2019.00031]

Zhao C, Yang J Q, Xiong X, Zhu A F, Cao Z G and Li X. 2022a. Rotation invariant point cloud analysis: where local geometry meets global topology. Pattern Recognition, 127: #108626 [DOI: 10.1016/j.patcog.2022.108626]

Zhao H S, Jia J Y and Koltun V. 2020. Exploring self-attention for image recognition//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10073-10082 [DOI: 10.1109/CVPR42600.2020.01009]

Zhao H S, Jiang L, Fu C W and Jia J Y. 2019. PointWeb: enhancing local neighborhood features for point cloud processing//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5560-5568 [DOI: 10.1109/CVPR.2019.00571]

Zhao H S, Jiang L, Jia J Y, Torr P and Koltun V. 2021. Point transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 16239-16248 [DOI: 10.1109/ICCV48922.2021.01595]

Zhao X, Zhang B W, Wu J J, Hu R Z and Komura T. 2022b. Relationship-based point cloud completion. IEEE Transactions on Visualization and Computer Graphics, 28(12): 4940-4950 [DOI: 10.1109/TVCG.2021.3109392]

Zhou T H, Chen J N, Shi Y N, Jiang K, Yang M M and Yang D G. 2022. Bridging the view disparity between radar and camera features for multi-modal fusion 3D object detection. IEEE Transactions on Intelligent Vehicles, 8(2): 1523-1535 [DOI: 10.1109/TIV.2023.3240287]

Zhu X G, Zhou H, Wang T, Hong F Z, Ma Y X, Li W, Li H S and Lin D H. 2021. Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 9934-9943 [DOI: 10.1109/cvpr46437.2021.00981]

Zhuang Z W, Li R, Jia K, Wang Q C, Li Y Q and Tan M K. 2021.

Perception-aware multi-sensor fusion for 3D LiDAR semantic segmentation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 16260-16270 [DOI: 10.1109/ICCV48922.2021.01597]

作者简介

龚靖渝，男，博士研究生，主要研究方向为3维场景分割与物体重建、3维动作生成。E-mail: gongjingyu@sjtu.edu.cn

马利庄，通信作者，男，教授，主要研究方向为计算机图形图像、计算机视觉、智能信息处理。E-mail: ma-lz@cs.sjtu.edu.cn

楼雨京，男，博士研究生，主要研究方向为3维场景物体理解。E-mail: louyujing@sjtu.edu.cn

柳奉奇，男，博士研究生，主要研究方向为3维物体检测与补全。E-mail: liufengqi@sjtu.edu.cn

张志伟，男，博士研究生，主要研究方向为3维场景理解与语义分割。E-mail: zhangzhiwei@sjtu.edu.cn

陈豪明，男，博士研究生，主要研究方向为场景点云全景分割。E-mail: chenhaomingbob@gmail.com

张志忠，男，副研究员，主要研究方向为机器学习与计算机视觉。E-mail: zzzhang@cs.ecnu.edu.cn

谭鑫，男，副研究员，主要研究方向为计算机视觉与机器学习。E-mail: xtan@cs.ecnu.edu.cn

谢源，男，教授，主要研究方向为图像处理、计算机视觉、机器学习与模式识别。E-mail: yxie@cs.ecnu.edu.cn