

基于深度学习的成对点云刚性配准现状与进展

周汝琴¹, 王鹏¹, 戴晨光¹, 汪汉云¹, 江万寿², 张永生¹

1. 信息工程大学 地理空间信息学院, 郑州 450001;

2. 武汉大学 测绘遥感信息工程国家重点实验室, 武汉 430079

摘要: 点云刚性配准作为三维点云数据处理的一项基础任务, 在自动驾驶、机器人、测绘遥感、医疗、工业设计以及文物保护等方面具有广泛应用。基于深度学习的点云配准方法能够自动学习高判别力的点云特征, 取得较高的配准精度。为对点云配准进行更深入有效的探索, 本文对基于深度学习的成对点云刚性配准相关技术研究进行了系统的综述与分析。首先介绍了基于深度学习的局部特征提取方法; 其次, 介绍对应关系解算、姿态回归和场景流估计等3类基于深度学习的配准方法相关研究进展; 再次, 对现有的可用于点云刚性配准的公开数据集进行了总结与归纳; 最后, 对点云配准研究现状进行了总结, 并对该领域未来研究方向进行展望。

关键词: 遥感, 三维点云, 刚性配准, 深度学习, 点云特征, 对应关系解算, 姿态回归, 场景流估计, 配准数据集
中图分类号: TP391.7/P2

引用格式: 周汝琴, 王鹏, 戴晨光, 汪汉云, 江万寿, 张永生. 2024. 基于深度学习的成对点云刚性配准现状与进展. 遥感学报, 28(12): 3074-3093

Zhou R Q, Wang P, Dai C G, Wang H Y, Jiang W S and Zhang Y S. 2024. Status and progress of deep learning-based pairwise point cloud rigid registration. National Remote Sensing Bulletin, 28(12): 3074-3093[DOI: 10.11834/jrs.20243396]

1 引言

随着LiDAR (Light Detection and Ranging)、立体相机和结构光传感器等设备的轻量化和低成本化, 三维点云的获取越来越方便。与此同时, 国内外对三维点云处理展开了深入的理论与应用研究, 研究成果已经广泛的应用于自动驾驶、机器人、测绘遥感、医疗、工业设计以及文物保护等领域(范文辉等, 2024; 王濮等, 2024; 杨必胜等, 2021)。然而, 由于目标间的遮挡、设备扫描距离有限等因素影响, 在数据获取时单个传感器单次扫描仅能完成对场景(或目标)部分表面数据的获取。为了获取场景(或目标)的完整数据, 通常需要进行多站点、多角度地数据采集, 以获得多个既相互独立、又局部相互重叠的子区域点云。点云刚性配准就是将每部分三维点云看作是一个刚体, 根据一些预定义的规则, 对不同站点、不同视角下采集到的三维点云进行旋转、平移,

从而将多个点云拼接成一个完整的三维点云(熊风光, 2018)。点云刚性配准作为三维点云数据处理的基础任务, 为后续三维建模、目标识别、场景理解奠定数据基础, 有着重要的研究价值。

关于点云刚性配准的研究起步较早。1992年, Stein和Medioni(1992)提出了基于splash特征描述符的点云粗配准方法。同年, Besl和McKay(1992)提出了迭代最近邻点ICP (Iterative Closest Point) 算法来解决点云刚性精配准问题。近几十年里, 研究学者们提出了各种特征描述符(例如Spin Image (Johnson和Hebert, 1999)、3DSC (Frome等, 2004)、FPFH (Rusu等, 2009)、RoPS (Guo等, 2013)、B-SHOT (Prakhya等, 2015)、BSC (Dong等, 2017)、KDD (Zhang等, 2021))和一系列基于ICP (Besl和McKay, 1992)的变种算法(例如Generalized-ICP (Segal等, 2009)、Sparse ICP (Bouaziz等, 2013)、Go-ICP (Yang等, 2016))来提高点云配准的性能。随着神经

收稿日期: 2023-09-15; 预印本: 2024-05-08

基金项目: 国家自然科学基金(编号: 42271457); 武汉大学测绘遥感信息工程国家重点实验室开放基金(编号: 22E03)

第一作者简介: 周汝琴, 研究方向为三维点云配准与重建。E-mail: zhouruqin@whu.edu.cn

通信作者简介: 王鹏, 研究方向为激光点云配准。E-mail: wpu2023@163.com

网络的快速发展,深度学习为点云配准开辟了新的道路。一系列基于深度学习的点云配准网络(如DeepVCP(Lu等,2019)、DCP(Wang等,2019c)、PointNetLK(Aoki等,2019)、PCRNNet(Sarode等,2019)、PREDATOR(Huang等,2021a)、FlowNet3D++(Wang等,2020c))层出不穷,并在配准精度和计算效率等方面显示出巨大的潜力。

尽管已有有点云配准技术的综述,然而,由于基于深度学习的点云配准在近几年发展迅速,已有的综述都未能全面覆盖现有的研究成果。其中:Guo等(2016)仅对2014年以前的传统点云配准方法中的局部特征描述符进行了系统的总结与分析;Cheng等(2018)仅聚焦于2018年以前摄影测量与遥感领域传统的点云配准方法;Zhang等(2020b)和秦红星等(2022)仅限于对基于对应关系和无对应关系的配准方法进行了重点总结与分析;Huang等(2021b)从同源点云配准和异源点云配准的角度出发,对2021年前的部分点云配准文献进行分

类与总结,但仅对基于优化的配准、基于特征学习的配准和端到端的配准方法进行了介绍。

与以上综述不同,本文从深度学习方法出发,对成对点云刚性配准最新文献资料进行系统的分析与整理,从局部特征提取、3类点云配准策略和数据集3个方面依次展开进行详细介绍,并对该领域未来研究方向进行展望。

2 基于深度学习的特征提取

基于深度学习的点云配准方法最为关键的步骤之一是利用神经网络自动学习出具有高判别力和鲁棒性的特征描述。本文按照输入点云的表示形式,如图1所示,总体上将点云特征提取的网络分为6大类:基于体素的神经网络、基于点的神经网络、基于图的神经网络、基于混合的神经网络、注意力机制和Transformer模型和其他网络。深度学习特征提取网络及对应的配准方法统计如表1所示。

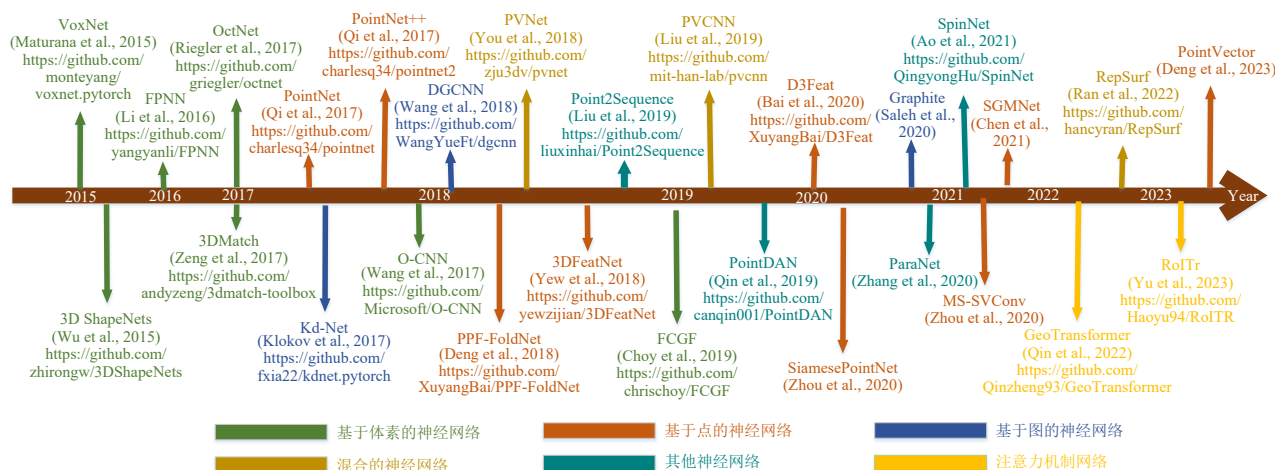


图1 基于深度学习的特征描述符学习网络时间发展脉络

Fig.1 The timeline of the development of feature descriptor learning networks based on deep learning

2.1 基于体素的方法

即将点云转化为规则的体素形式作为网络输入。借鉴于规则排列的二维影像, Maturana和Scherer(2015)提出了针对规则三维体素的卷积神经网络VoxNet。该网络首先对点云进行三维体素化,然后在规则化的体素网格上构建三维卷积神经网络来学习点的特征。Wu等(2015)将点云表面形状看作是三维体素的概率分布问题,提出了3D ShapeNets网络。该网络先对点云进行体素化,然后使用神经网络学习复杂3D形状在不同类别和

任意姿态下的分布,并自动学习多层次的特征描述。上述规则体素化的方法虽能直接使用三维卷积神经网络学习出点云丰富的几何结构特征,然而,计算和显存占用随着体素数量的增加呈立方的增长,无法扩展到大规模密集的三维数据。为此, Riegler等(2017)提出了基于八叉树的网络OctNet来降低计算和存储开销。该网络首先使用混合网格八叉树结构对点云进行层次划分,然后采用位串表示法对八叉树结构进行编码,并通过简单的算法对每个体素的特征向量进行索引。为

了提高稀疏点云的特征提取效率，Choy 等（2019）提出了 FCGF 网络。该网络使用稀疏张量来表示三维点云数据，并用 Minkowski 卷积代替传统卷积来快速提取紧凑的包含上下文信息几何特征。Horache 等（2021）在 FCGF（Choy 等，2019）的基础上提出了多尺度架构与自监督微调卷积神经

网络 MS-SVConv。该网络将不同大小的体素输入到共享权重的 U-Net 网络结构中分别提取不同尺度特征，然后使用全连接层对这些特征进行融合得到每个点的特征描述符。MS-SVConv 延续了 FCGF 运算速度快、对旋转鲁棒的特点，同时大幅加强了泛化性能。

表 1 深度学习特征提取技术及对应配准方法统计
Table 1 Deep learning feature extraction technology and corresponding registration method

类型	配准方法			
基于体素的方法	RoReg (FCGF)	DGR (FCGF)	DHVR (FCGF)	PointFlowNet (Voxelnet)
	RegTR (KPCnv)	PCAM (KAConv)	PREDATOR (KPCnv)	UDPReg (KPCnv)
基于点的方法	GeoTransformer (KAConv)	OGSF-Net (PointConv)	Bi-PointFlowNet (PointConv)	OMNet (MLP)
	DeepVCP (PointNet++)	SPFlowNet (PointNet++)	FINet (MLP)	PointNetLK (PointNet)
	DeepGMR (PointNet)	RoTr (PPFs)	LPD-Registration (PointNet)	PCRNet (PointNet)
	SCANet (MLP)	FlowNet3D (PointNet++)	Feature-metric Registration (MLP)	
基于图的方法	STORM (DGCNN)	DCP (DGCNN)	GAFAR (DGCNN)	
混合的方法	GAFAR (DGCNN, MLP)	RGM (KPCnv, PCA-GM)	Stickypillars (PointPillar, MLP)	
注意力机制和 Transformer	RegTR	DCP	PCAM	DeepVCP
	PREDATOR	UDPReg	GeoTransformer	RoTr
	LPD-Registration	RGM	SCANet	Stickypillars
其他神经网络	LO-Net (projection)	DeLORA (projection)	EfficientLO-Net (projection)	HPLFlowNet (BCL)
	DCA-SRSFE (Laplacian Mesh)			

2.2 基于点的方法

直接采用最原始的点作为网络输入。针对基于体素化的特征学习网络无法平衡分辨率与效率的问题，Qi 等（2017a）创新性地提出了能直接提取无序点云特征的网络 PointNet。该网络首先通过 T-Net 来学习变换矩阵以保证网络对空间几何变换的不变性；然后通过多层感知机进行特征提取，再用 T-Net 对在特征空间进行特征对齐；最后，通过最大池化（Max Pooling）操作来得到最终的全局特征。但是，PointNet 忽略了点云局部邻域信息。针对此问题，Qi 等（2017b）在 PointNet（2017a）的基础上提出了 PointNet++，通过采样（Sampling）、组合（Grouping）和 PointNet（2017a）3 个步骤来聚合局部邻域信息。PPFNet（Deng 等，2018b）和 PPF-FoldNet（Deng 等，2018a）将几何特征属性如点的坐标、法线以及点对特征 PPF（Point Pair Features）组合成原始特征，然后输入到 PointNet（Qi 等，2017a）网络结构中来提取具有全局上下文信息的描述符。SiamesePointNet（Zhou 等，2020）通过一系列基于 PointNet（Qi 等，2017a）的分层编码器—解码器网络结构生成关键点的描述

符。由于附加了形状对比约束和层次局部算子，SiamesePointNet 对全局上下文和局部上下文都有很强的感知能力。此外，通过引入特征变换操作，将点特征变换为紧凑的描述空间，使得 SiamesePointNet 提取的描述符不受形状几何差异的影响。Thomas 等（2019）提出核点卷积 KPConv（Kernel Point Convolution）。该网络在中心点的球形邻域点集选择若干核点，给每个核点一个可学习得权重矩阵，根据核点与邻域点的距离由核函数计算邻域点的特征，将球体内的所有邻域点的特征累加得到中心点特征。面对局部几何提取器逐渐复杂，性能趋于饱和，Ma 等（2022）重新审视复杂局部描述的必要性，构建了纯残差 MLP 网络 PointMLP。该网络由残差 MLP 模块和几何仿射模块为基础，通过多层残差 MLP 模块逐步扩大感受野提取深层特征，几何仿射模块对局部点云进行仿射变换，弥补局部区域差异导致的网络不稳定。Deng 等（2023）提出了向量为导向的点集抽象方法 PointVector，通过高维向量聚合相邻的特征，用于解决标准 MLP 在提取局部特征方面受到限制的问题。

2.3 基于图的方法

以点云的坐标为节点、以点云邻域特征构建边形成图作为网络输入。针对点云的无序性和非结构性,研究者尝试以图的方式将点云为重建结构数据,用图神经网络提取点云的结构信息。Wang等(2019c)提出了可以直接在点云上卷积的模块EdgeConv,在保证置换不变性的同时捕获局部几何信息。Shen等(2018)将核相关性和图池化操作引入网络中,提出了KCNet网络。该网络将点集核定义为一组根据邻域点核相关性来学习的三维点。然而,由于涉及到邻域点的访问,KCNet网络的效率较低。Saleh等(2020)提出了简单有效的特征提取和关键点检测的网络结构Graphite,通过不同尺度的图卷积操作来描述点云区域并提取显著点。实验证明该网络学习出的轻量级描述符在特征匹配和点云配准方面具有明显的精度和效率优势。Fu等(2023)在RGM网络将Transformer引入图的构建中而不用最KNN等方式,增强了图特征感知结构信息和点云对重叠区域的能力。并且,受PCA-GM(Wang等,2019a)启发,RGM引入图内卷积(Intra-graph Conv)和图间卷积(Cross-graph Conv)增强节点特征的图间相关性。

2.4 混合的方法

使用两种及两种以上网络结构进行结合,以提高学习出的特征的多样性。You等(2018)提出的点云目标识别网络PVNet融合了点云和多视图数据。Liu等(2019d)提出的PVCNN网络结合了基于点的网络结构PointNet(Qi等,2017a)和基于体素的网络结构VoxNet(Maturana和Scherer,2015),实现了无动态核的高效局部几何信息提取。Zhou等(2019)通过分别在鸟瞰图和透视图上进行动态体素化,提出端到端多视图融合MVF(Multi-View Fusion)网络以获取互补信息。该方法比单视点方法具有更高的精度。

2.5 注意力机制和Transformer模型

近年来,自注意力、交叉注意力、多头注意力和Transformer等新兴网络结构深受研究学者们的喜爱。Qin等(2022)设计了GeoTransformer。该方法首先通过KPCov(Thomas等,2019)模块得到超点,对超点之间的距离信息和角度信息进行编码并嵌入到Transformer中进行特征信息交换。因为距离和角度信息都是位姿不变的,因此在特

征学习阶段能够显式的捕捉到源点云和目标点云中包含全局上下文信息的位姿不变几何特征。Yu等(2023)提出了基于点对特征(PPF)和点对特征注意力机制(PAM)的编码—解码架构RoITr。RoITr使用PPFs(Drost等,2010)的方法构造局部坐标,并以位姿无关的局部坐标三元组(坐标,法向量,特征)代替点云坐标,嵌入注意力模块更新位姿不变的局部几何特征。然后使用Transformer模型对三元组的坐标及其特征进行跨点云的上下文信息学习,捕获全局结构特征,并将其与局部几何特征结合。RoITr可以获得高显著性和姿态无关的特征。

2.6 其他神经网络

为了增强特征描述能力,Liu等(2019a)将从局部区域学习的多尺度特征视为序列,提出了序列处理网络Point2Sequence。该网络将这些序列输入到基于RNN的编解码器结构中来聚合局部区域特征。PointDAN(Qin等,2019)采用具有可调节感受野的自适应模块和对抗训练策略分别在多层次上联合对齐局部和全局特征表达。Zhang等(2020a)提出了将不规则三维点云编码为规则的二维彩色图像的深度学习框架ParaNet。该网络由表面嵌入模块和网格重采样模块组成,以完全规则且几乎无损的方式将不规则的三维点云转换为规则的图像。Ao等(2021)提出了能学习旋转不变性特征的网络SpinNet,通过三维圆柱卷积将三维曲面转化为圆柱体提取出紧凑且具有较强描述能力的局部特征。然而,SpinNet需要预先计算出每个局部邻域参考轴。Ran等(2022)提出了新颖的点云表示RepSurf,通过在点云中建立三角网格曲面或伞形曲面描述其局部几何结构,将三角形特征(即法向量、曲面位置、归一化坐标)经可学习的变换函数计算得到RepSurf。通过结合位置信息和几何信息,点表示可以更具表现力,可以显式地描述局部的结构。

总的来说,通过神经网络学习点云深层次的信息得到的特征描述符,具有较强的描述性。然而,基于点的方法需要建模点之间的关系,存储过多权重矩阵可能会造成网络冗余;基于体素的方法难以平衡分辨率和精细特征之间的问题;基于图的方法在点云数据量大时效率较低;额外添加注意力机制会导致效率低、算力消耗大、占用空间多。如何能快速、高效的学习点云的丰富多

样且紧凑的特征,以满足硬件设备的处理能力和不同应用的实时需求,是该类方法主要的研究方向之一。

3 基于深度学习的点云配准

3.1 基于对应关系解算的点云配准

基于对应关系解算的点云配准方法总体流程如图2所示。该类方法利用神经网络学习逐点的特

征表达,然后采用可微的优化方法或者神经网络来求解三维点的对应关系,最后利用可微的奇异值分解来估计最优变换。在该类方法中,最为关键的是局部特征表示与点对应关系求解。如图3所示,总体上可以将对应关系的配准网络分为针对点云密度分布不均匀的网络、针对局部重叠和低重叠的网络、提高鲁棒性的网络和提高泛化性的网络和提高效率的网络共5大类。

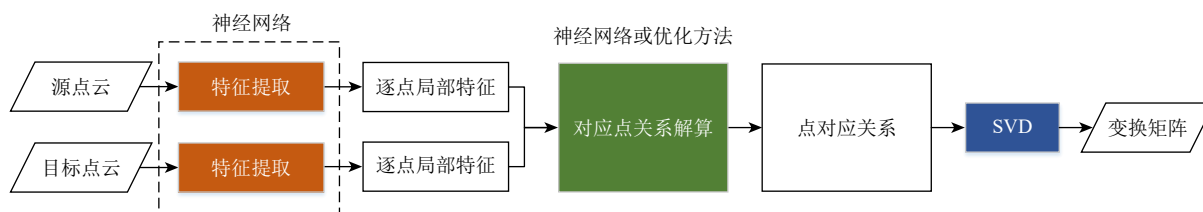


图2 基于对应关系的点云配准基本流程

Fig.2 The basic workflow of correspondence-based point cloud registration

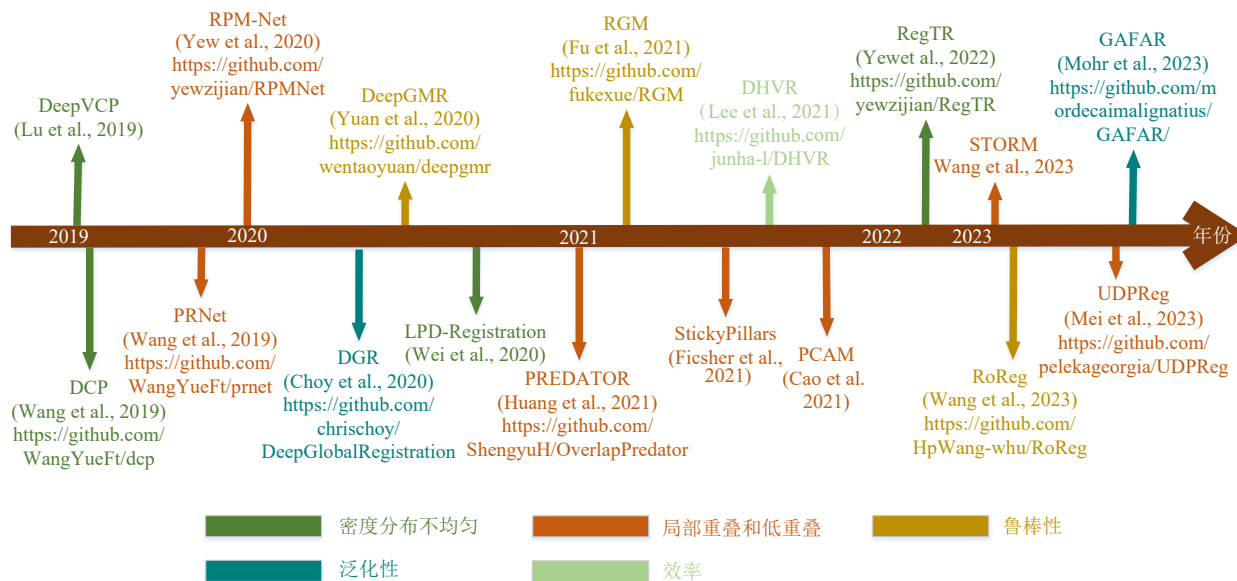


图3 基于对应关系解算的点云配准时间发展脉络

Fig. 3 The development timeline of correspondence-based point cloud registration methods

3.1.1 针对密度不均匀的网络

针对点云密度不均匀导致的相邻两帧点云中的对应点并不一一精确对应的问题,已有研究提出了基于虚拟匹配点生成的网络。该类网络不需要剔除错误匹配对,实现了与最先进的基于几何的方法相当的配准精度和运行效率。Lu等(2019)提出了第一个端到端的、基于虚拟匹配点的点云配准网络DeepVCP。该网络首先将PointNet++(Qi等,2017b)学习到的逐点特征输入点加权层选择

特征最显著的点作为关键点;然后以关键点为中心,对其邻域使用小型PointNet(Qi等,2017a)聚合得到嵌入特征;最后,用初始位姿将源点云的关键点变换到目标点云,得到候选虚拟对应点,并将源点云和目标点云的特征输入由3个3DCNN层组成的对应点生成模块,通过Softmax函数得到匹配矩阵,使之与候选虚拟对应点相乘即得到最终虚拟对应点。考虑到对配准最有用的特征主要是从局部和全局信息中联合学习的特点,Wang等

(2019c) 提出了基于 Transformer 的配准网络 DCP。与 DeepVCP (Lu 等, 2019) 不同的是, DCP 首先利用深度图卷积网络学习点云刚性不变特征, 然后利用 Transformer 模块学习两帧点云的互相关特征, 然后采用对应点生成模块来生成虚拟点。类似的, Wei 等 (2020) 提出了基于注意力的点云配准网络 LPD-Registration。LPD-Registration 首先引入改进的 LPD-Net (Liu 等, 2019c) 进行特征提取, 并利用图卷积进行特征聚合; 然后分别通过自注意力和互注意力机制来增强感知点云的结构信息和源点云和目标点云之间的关联信息; 最后将特征输入对应关系指示器得到两帧点云的点匹配概率, 从源点云中取概率最高的 K 个点, 对于其中的每个匹配点都从目标点云取 J 个候选对应点, 然后此 J 个候选点加权聚合即得到 K 个虚拟匹配点。Yew 和 Lee (2022) 提出了同时估计重叠分数的配准网络 RegTR。该网络首先使用 KPConv 边降采样边提取邻域特征, 然后将关键点特征直接输入多头自注意力和互注意力层获取约束性强的条件特征, 最后经过对应点生成模块直接生成虚拟点并估计点的重叠分数。

3.1.2 针对局部重叠和低重叠的网络

针对局部重叠点云配准的非凸性和匹配点部分对应的问题, 学者们大多采用注意力机制来减少非重叠区域信息的干扰。这类基于注意力机制的方法通过动态的学习使注意力集中在重叠区域, 能有效克服局部重叠点云配准的非凸性和匹配点部分对应的问题。Cao 等 (2021) 提出了基于互注意力的网络结构 PCAM。该结构允许两帧点云在每一层局部特征提取时交换上下文信息, 从而可以在重叠区域找到最佳匹配点, 然后将匹配点对连接成向量输入点过滤模块得到所有匹配点对的置信估计, 以此过滤不可靠的点对。Huang 等 (2021a) 提出了顾及两帧点云重叠分数的网络 PREDATOR。该网络首先通过 KPConv 编码器将点云降采样编码为带有特征的超点 (superpoints), 然后通过重叠注意力模块提取两帧点云间共同的上下文信息, 并估计每个超点的重叠分数, 最后通过解码器将共同的上下文信息和重叠分数进行解码, 得到重叠分数和匹配概率。PREDATOR 在低重叠度 (小于 30%) 的情况下能显著提高配准精度。Fischer 等 (2021) 提出了高效、精确的点

云配准网络 StickyPillars。该网络直接在稀疏的关键点进行操作, 首先通过圆柱编码器和位置编码器对关键点进行不同层次特征提取, 然后通过多个多头自注意力和交叉注意力模块对上下文信息进行整合, 最后通过优化传输层得到关键点匹配关系。StickyPillars 在高速场景和帧丢失的情况下依旧有效。Wang 等 (2023c) 提出了基于结构的重叠匹配框架 STORM, 该框架由结构特征提取模块和重叠预测模块组成。为了学习到的逐点特征具有更强的判别性, 结构特征提取模块以多个 EdgeConv 层和两个 Transformer 层作为特征提取和细化方法, 将不同 EdgeConv 层生成的上下文结构逐点特征聚合, 促进重叠预测模块识别重叠点。基于输入点云的上下文结构信息, 重叠预测模块使用 Gumbel-Softmax 构建描述点云潜在共同结构的分布, 并从共同结构中采样点集。然后 STORM 依据特征相似性构建对应点对, 使用加权 SVD 解算变换矩阵。实验表明, STORM 在不同程度的重叠率时, 相比 PREDATOR (Huang 等, 2021a) 取得优势。针对低重叠点云配准, Mei 等 (2023) 提出了无监督深度概率配准框架 UDPReg。该框架首先使用自注意力和互注意力配合 KPConv-FPN 来获取点云的逐点特征和逐点重叠分数, 然后采用 MLP 和 Softmax 以逐点特征和点云坐标为输入学习点云的高斯混合模型 (GMMs) 分布, 然后应用 Sinkhorn (Cuturi, 2013) 算法获取源和目标点云 GMM 的分布级匹配来估计聚类点集对应关系, 最后采用 RANSAC (Fischler 和 Bolles, 1981) 的变体来计算变换矩阵。UDPReg 通过分布一致性损失和局部对比损失来训练网络来学习更多的判别特征, 在 3DLoMatch 低重叠数据上, 相对旋转误差、相对平移误差和召回率等指标与 RegTR (Yew 和 Lee, 2022) 等有监督的方法同样先进。

3.1.3 提高鲁棒性的网络

针对现有配准方法在初始位置相差较大、噪音干扰表现较差的问题。Yuan 等 (2020) 将配准问题定义成两个高斯混合模型 (GMM) 分布的 KL (Kullback-Leibler) 散度最小值求解问题, 并提出了 DeepGMR 框架。该框架首先提出了可学习的置换不变点网络, 向其输入点云数据来估计每个点云 Point-to-Component 的对应关系矩阵, 然后通过两个可微分的计算模块分别以对应关系矩阵和点

云坐标、两帧点云对应的GMM为输入计算混合高斯模型参数、变换参数。DeepGMR在噪声影响下能达到和未加噪声时相同的配准均方根误差指标(RMSE),对数据质量差的稀疏区域也能得到正确的配准矩阵,但没有明确考虑重叠区问题。Fu等(2023)提出了基于图匹配的点云配准框架RGM。该框架首先用KPCov提取点云的逐点特征;然后以点云坐标为节点、以transformer学习节点特征之间的关联构建边形成图,并进行多层intra-graph和cross-graph卷积;最后使用图匹配来计算匹配概率矩阵。RGM在建立对应关系时不仅考虑到每个点的局部几何形状,还考虑到更大范围的结构信息,在对数据加噪声和随机位移后仍能相比对照方法达到最优结果。Wang等(2023b)提出了基于群旋转等变特征及局部旋转估计的配准框架RoReg。在关键点检测、特征描述、特征匹配和变换估计中引入旋转群特征及局部旋转,显著提高了点云配准的鲁棒性。RoReg在3DMatch上训练,在低重叠的3DLoMatch和ETH数据上测试相比CoFiNet(Yu等,2021)等基线方法显著提升了配准召回率和降低了旋转平移误差。

3.1.4 提高泛化性的网络

泛化性一直是深度学习应用中面临的重要问题之一,研究学者们也从不同的角度进行了尝试来提高配准网络的泛化性。Choy等(2020)提出了精度高的配准网络DGR。该网络首先使用FCGF(Choy等,2019)提取逐点特征,并使用6D卷积神经网络来评估点匹配关系的置信度,然后利用可微加权Procrustes(Gower,1975)算法估计姿态参数,最后利用鲁棒优化模块对姿态参数进行微调优化。DGR仅在3DMatch上训练,在3DMatch(Zeng等,2017)室内数据集和KITTI室外数据集上测试均能达到比已有先进方法更高的配准精度。Mohr等(2023)结合局部结构信息、全局信息和重叠区域内的拓扑结构,提出了基于图注意力特征增强的快速轻量的网络架构GAFAR。该网络由特征头和注意力特征增强模块组成。首先,特征头包使用DGCNN和MLP获取局部特征和位置编码并连接得到逐点局部特征描述符。然后,图注意力特征增强模块使用多层自注意力和交叉注意力层,将源点云和目标点云重叠区域的关联拓扑信息嵌入描述符,促进匹配搜索。最后,使用

Sinkhorn算法和SVD建立点对应关系、计算刚性变换矩阵。GAFAR仅在ModelNet40上训练,在KITTI测试的结果表明,泛化性能胜过RGM、RPM-Net等对比方法。

3.1.5 提高效率的方法

Lee等(2021)提出了基于Hough投票的配准框架DHVR。该框架首先使用FCGF(Choy等,2019)网络进行局部特征提取并计算假定的对应关系,然后构造一系列三元对应关系,并通过在6D Hough空间上投票来判断对应关系的一致性,从而估计6个变换参数,最后通过一个全卷积Hough空间精化模块来抑制Hough投票空间的噪声,输出优化的变换参数。DHVR能处理真实室内场景数据集,并且相对DGR(Choy等,2020)等方法具有更高的运算效率。

总的来说,基于对应关系解算的方法利用神经网络学习逐点特征并采用可微的优化方法或者神经网络来求解三维点的对应关系。该类方法对初始位置不敏感,且在数据质量较好的情况下能达到较高的精度。然而,该类方法在密度分布不均匀、重叠区域变化较大、遮挡或噪声等情况下表现较差。因此,如何提高基于对应关系解算的方法在局部重叠的真实复杂场景中的精度、效率、鲁棒性和泛化性,减少对训练样本的需求量,是需要重点解决的问题。

3.2 基于姿态回归的点云配准

如图4所示,基于姿态回归的配准方法将配准视为几何变换参数回归的问题。首先使用神经网络提取点云的全局特征,然后使用回归方法或神经网络来回归姿态参数。该类方法中间过程不涉及任何对应关系的解算,直接输出变换参数。由于该方法原理简单,流程较短,效率较高,近几年来受到许多研究学者的关注,并提出了一系列创新性成果。如图5所示,总体上可以将姿态回归的配准网络分为5大类:提高鲁棒性的网络、针对局部重叠的网络、提高泛化性的网络、克服数据标记问题的半监督和无监督网络和提高效率的网络。

3.2.1 提高鲁棒性的网络

PointNetLK(Aoki等,2019)作为第一个基于姿态回归的方法,为点云配准开辟了新的思路。

PointNetLK 首先通过 PointNet (Qi 等, 2017a) 来提取点云全局特征, 然后通过改进的 LK (Lucas Kanade) (Lucas 和 Kanade, 1981) 算法估计变换参数。该方法不需要计算点对应关系, 在 ModelNet40 和 Stanford bunny 上测试表明, PointNetLK 在配准精度、对初始位置的鲁棒性和计算效率方面相比 ICP 基线方法带来了巨大的优势。然而, PointNetLK 严重依赖于通过有限差分对梯度进行估计, 对步长的选择非常敏感 (Li 等, 2021c)。受 PointNetLK (Aoki 等, 2019) 启发, Sarode 等 (2019) 提出了完全基于神经网络的姿态回归配准框架 PCRNet。该框架首先使用 PointNet (Qi 等, 2017a) 来提取

点云的全局信息, 然后通过多个全连接层进行姿态回归。PCRNet 对高斯噪声具有比 PointNetLK 更强的鲁棒性, 且很好地解决了 PointNetLK (Aoki 等, 2019) 固有的内存和效率问题。Li 等 (2019) 提出了能同时估计动态区域掩码和相对姿态的网络 LO-Net。该网络首先将点云通过圆柱投影转换为二维图像, 然后使用编码器对带有法向量信息的点云进行特征编码, 最后将源点云和目标点云的特征连接起来输入多个卷积模块和全连接模块, 在最后一层分别输出变换参数。LO-Net 在估计变换参数的同时, 可对动态对象进行滤除, 从而减小动态对象对配准的干扰。

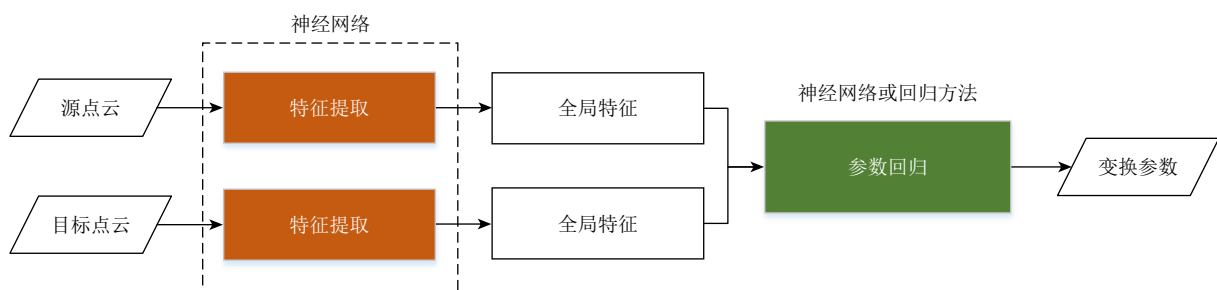


图4 基于姿态回归的点云配准基本流程

Fig.4 The basic workflow of point cloud registration based on pose regression

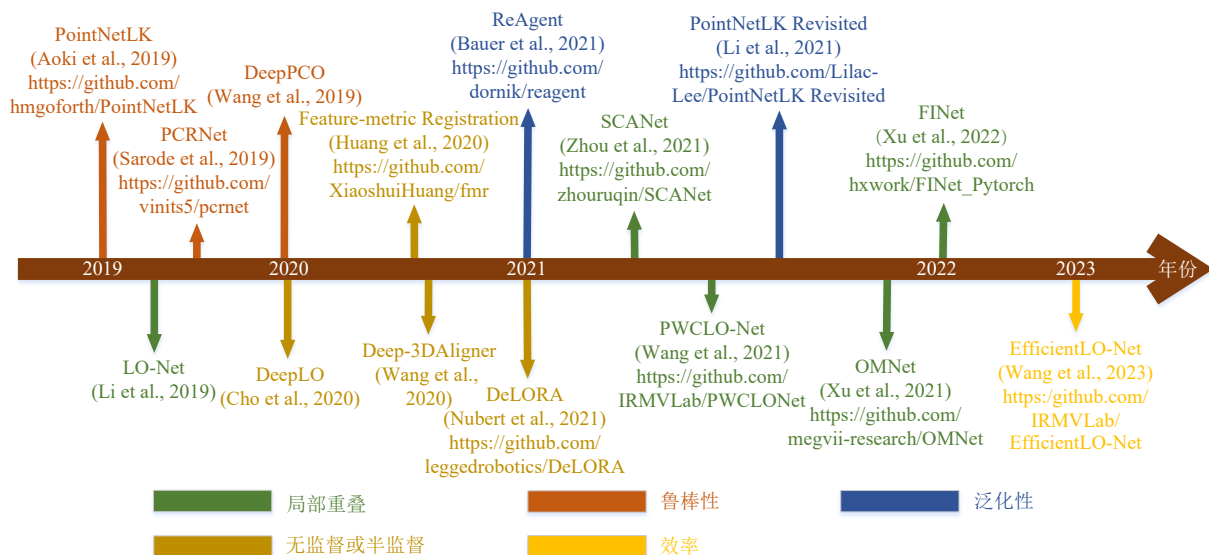


图5 基于姿态回归的点云配准时间发展脉络

Fig.5 The timeline of the development of pose regression-based point cloud registration

3.2.2 针对局部重叠的网络

针对现有的姿态回归网络不适用于局部重叠点云配准的问题, 学者们大多采用源点云和目标点云全局特征交互的形式来学习点云之间的相关性进而感知重叠区域。Zhou 等 (2021) 提出了基

于注意机制的网络结构 SCANet。该网络采用多层空间自注意力模块聚合点云不同层次的局部特征并经过最大池化获取全局信息; 然后采用通道互注意力实现源点云和目标点云全局特征向量之间的信息交互, 增强特征重叠区域的能力, 最后通

过全连接层回归到7个变换参数。SCANet能同时兼顾精度和效率,且对重叠率变化具有一定的鲁棒性。Xu等(2021)提出可同时学习重叠区域掩码和配准参数的网络OMNet。该网络首先提取源点云和目标点云的逐点特征和全局特征,然后将两帧点云的特征输入重叠掩模预测模块得到重叠掩膜(包含对应点云中所有点是否在重叠区域的标签向量),掩膜与逐点特征相乘以分割重叠区域。将两帧点云重叠区域的全局特征和逐点特征互相连接输入变换回归模块得到配准参数。OMNet对重叠率变化具有较强的鲁棒性。在OMNet(Xu等,2021)的基础上,Xu等(2022)提出了双分支特征交互点云配准网络FINet。该网络分为两个部分:双分支编码器和双分支变换回归器。双分支编码器通过在特征提取阶段进行源点云和目标点云之间的多层特征交互,加强了源点云和目标点云数据的关联,从而突显了非重叠区域给特征带来的差异性,可使网络具备重叠区域感知能力,同时顾及旋转和平移属于不同的解空间的事实。双分支回归器结构分别回归旋转和平移的参数,以减轻旋转和平移之间的解空间差异的影响。

3.2.3 无监督或半监督的网络

针对现有配准网络依赖庞大标记数据的问题,研究学者们通常以最小化点云对齐损失作为监督信号进行无监督或半监督训练。Huang等(2020)提出了半监督、通过最小化特征度量投影误差来实现配准的框架Feature-metric Registration。该框架包含两个模块:编码器和多任务分支。编码器模块用于点云的特征学习。对于多任务分支模块:编码器—解码器分支可以从有旋转差异的点云中学习生成特异性的特征,并将不同的特征恢复到对应的位姿状态点云。Feature-metric Registration以目标点云和经过变换对齐点云的Chamfer损失为优化目标用无监督的方式训练解码器分支;特征度量配准分支通过最小化特征差异来估计两帧点云间的变换矩阵。Wang等(2020b)提出了非监督的配准网络Deep-3Daligner。该网络主要包含3个核心:SCR(Spatial Correlation Representation)优化器、解码器和自适应Chamfer损失。Deep-3Daligner首先通过SCR优化器优化随机初始化的潜在空间相关表示特征,然后通过解码器将SCR特征解码为所需的几何变换以对齐源点云和目标

点云。自适应Chamfer损失作为监督信号,用来衡量转换后的源点云和目标点云的配准效果,通过梯度反向传播优化SCR和解码器。Deep-3Daligner虽然能取得较高的配准精度,但是牺牲了计算效率。Nubert等(2021)提出了无监督学习的里程估计框架DeLORA。该框架首先将源和目标点云映射成二维图像并堆叠在一起,然后经卷积层得到特征向量,接着用MLP分别回归特征向量得到平移参数和四元数旋转参数。该网络以两帧点云的对应点之间的点云误差和法向量误差作为损失函数进行无监督训练。DeLORA简单计算量小,能够在移动类CPU上实时运行。

3.2.4 提高泛化性的网络

针对现有配准网络泛化性较差的问题,Li等(2021c)在PointNetLK(Aoki等,2019)的基础上提出了具有较高泛化性的配准框架PointNetLK Revisited。该框架在LK(Lucas和Kanade,1981)算法的基础上,采用可分解为“feature”和“warp”分量的雅可比矩阵。这种分解能够修改可分析的扭曲雅可比矩阵,而无须重新训练全局特征。实验证明,该网络在ModelNet40数据集上训练的模型可以直接用在3DMatch数据集上,具有较强的泛化性。Bauer等(2021)将迭代的点云配准作为强化学习任务,提出新的配准框架ReAgent。该框架首先通过特征嵌入模块来提取每个点云全局特征,然后将源点云和目标点云的全局特征串联起来作为Agent状态信息进行编码,最后通过模拟专家策略网络使用状态信息来分别估计旋转和平移。ReAgent在ModelNet40前20类数据上训练,在ModelNet40后20类数据和ScanObjectNN上测试相比对比方法能分别取得最优和次优的配准结果。然而,ReAgent需要不断的迭代多次执行全局特征提取和变换参数估计两个步骤来取得更高的精度。

3.2.5 提高效率的方法

EfficientLO-Net(Wang等,2023a)使用基于柱面投影的点云投影感知表示将原始点云组织成有序的数据形式,再通过分组后使用多层set conv进行每个点的局部特征聚合;然后引入投影感知特征的注意成本体(Attentive Cost volume),在欧氏空间建立源点云的每个点和目标点云的关联,并用注意力网络生成源点云对目标点云的注意力

流嵌入特征。将注意力流嵌入特征输入 MLP 回归整体姿态参数。同时, 以注意力流嵌入特征和局部特征为输入经 MPL 生成掩膜, 来处理各种异常点和动态对象。EfficientLO-Net 在 KITTI 测试集、M2DGR (Yin 等, 2022) 数据集、Argoverse (Chang 等, 2019) 数据集等不同序列、不同场景上测试实现了相对 HALFlow (Wang 等, 2021a)、A-LOAM (Zheng 等, 2020) 等先进方法在平移旋转指标上更好的性能。在时间效率上, 与 LO-Net (Li 等, 2019)、Full LOAM (Zhang 和 Singh, 2017) 比较具有更高的效率, 能够实现 20 Hz 的 LiDAR 里程计, 并比以往所有基于学习的方法和基于几何的 LOAM (Zhang 和 Singh, 2017) 方法具有更高的性能。

总的来说, 基于姿态回归的方法利用神经网络来学习全局特征并直接回归姿态参数, 不需要计算点对应关系, 在精度、鲁棒性和计算效率方面带来了巨大的优势。然而, 该类方法目前大多聚焦于完全重叠的小目标点云配准, 且采用迭代

的方式来达到较高的精度。如何用尽可能紧凑的全局信息来表示复杂场景点云的特征, 尤其是重叠区域的特征, 在保证其配准精度的同时提高其运行效率、鲁棒性和泛化性, 仍是该类算法后续需要进一步研究的重要内容。

3.3 基于场景流估计的点云配准

与基于对应关系和姿态回归的配准方法不同, 基于场景流估计的方法将点云配准看作是两个连续时刻的点云空间运动问题。如图 6 所示, 给定两个连续时刻 $t, t+1$ 的点云 S 和 T , 该类方法通过关联不同时刻的局部特征, 通过流估计网络输出源点云中每个点的平移运动参数。基于场景流的点云配准不仅可以估计两帧点云间的刚性运动, 还可以估计场景中动态物体运动的场景流。目前来说, 基于场景流的点云配准研究主要集中在自动驾驶点云处理上。如图 7 所示, 总体上可以将场景流的配准网络分为 4 大类: 提高鲁棒性的网络、提高泛化性的网络、克服数据标记问题的半监督和无监督网络和提高效率的网络。

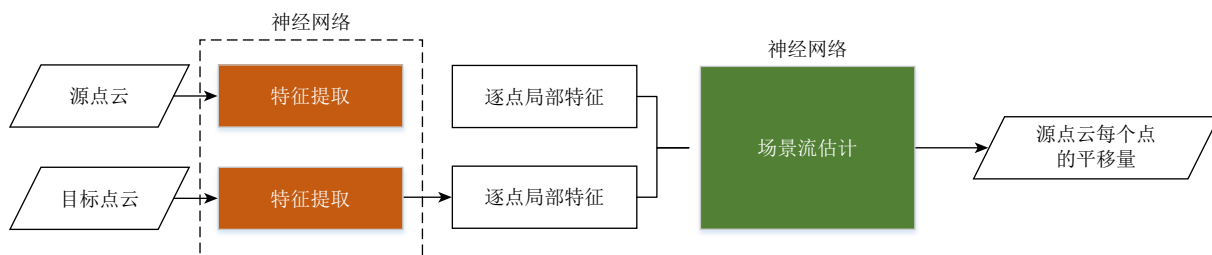


图 6 基于场景流的点云配准基本流程

Fig. 6 The basic workflow of point cloud registration based on scene flow

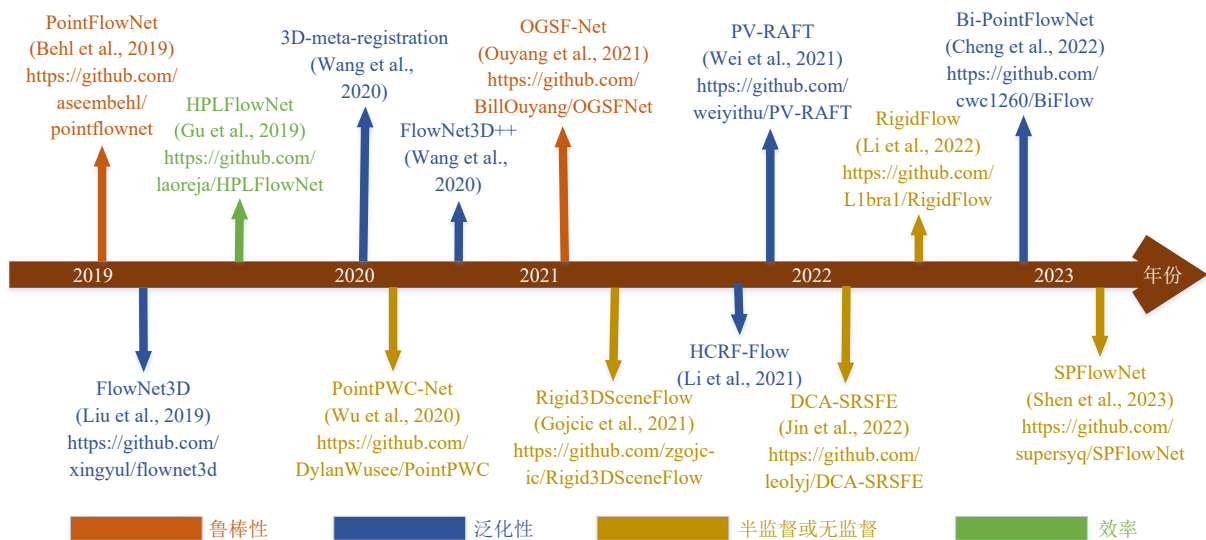


图 7 基于场景流的点云配准时间发展脉络

Fig. 7 The timeline of the development of point cloud registration based on scene flow

3.3.1 提高泛化性的网络

Liu 等 (2019b) 在 FlowNet (Dosovitskiy 等, 2015) 基础上改进, 提出了直接对三维点云进行场景流估计的网络 FlowNet3D。该网络由 3 个模块组成: 点特征学习模块、点混合模块和流优化模块。该网络在点特征学习模块采用 set conv 层下采样学习逐点局部特征; 在点混合模块用流嵌入层学习两个点云之间的几何相似性和空间关系; 在流优化模块, set up-conv 层将嵌入层特征向上采样赋予源点云, 得逐点传播特征; 最后将得到的特征输入 MLP 回归出源点云中每个点的平移运动矢量。FlowNet3D 在 FlyingThings3D 合成数据集上训练, 在 KITTI 数据集上测试获得最优结果。Wang 等 (2020c) 在 FlowNet3D (Liu 等, 2019b) 基础上, 提出了加入几何损失的场景流估计网络 FlowNet3D++。该网络将流场中点到面距离和各个向量间的角度等几何约束合并到 FlowNet3D (Liu 等, 2019b) 中。和 FlowNet3D 类似, FlowNet3D++ 只需要在 FlyingThings3D 合成数据集进行训练便可在 KITTI 数据集上微调使用。鉴于数据分布的不同会降低网络的泛化性, 为了提高现有方法对未训练点云的泛化性, Wang 等 (2020a) 首次引入元学习 (Meta-learner) 策略 (指参数预测, 以元学习网络的输出作为另一个网络的参数) 来训练模型, 并设计了基于元学习的三维配准框架 3D-Meta-Registration。该框架由三维配准学习器和三维配准元学习器两个模块组成。元学习模块将每一对源点云和目标点云的配准看作单独的配准任务, 并学习众多任务的潜在空间分布, 生成对于给定源点云和目标点云最优的网络参数来更新配准学习器参数, 以快速适应不同情况的配准任务。3D-Meta-Registration 在 FlyingThings3D 数据训练后直接在 KITTI 数据集测试, 配准效果相比 FlowNet3D 显著提高。考虑到单向特征传播限制了场景流估计精度, Cheng 和 Ko (2022) 提出了基于双向流嵌入层的场景流估计新架构 Bi-PointFlowNet。首先, 应用分层特征提取器在两帧点云中提取多层次的局部特征; 然后, 应用双向流嵌入 (BFE) 层进行两帧点云双向的分组—串联—MLP—池化操作, 学习具有丰富的上下文信息和相关信息的流嵌入特征来回归输出场景流。Bi-PointFlowNet 在 FlyingThings3D 和 KITTI 基准测试集上, 在遮挡和

非遮挡条件下都实现了最先进的性能和泛化性。

3.3.2 提高鲁棒性的网络

利用局部几何信息和上下文信息增强特征描述性能有效提升网络的鲁棒性。针对自动驾驶场景点云的稀疏性、密度不均匀和外观信息缺失导致的数据关联困难的问题, Behl 等 (2019) 提出多任务网络 PointFlowNet 来联合预测三维场景流以及场景中物体的三维边界盒和刚体运动。该网络首先将点云体素化, 然后通过特征编码器学习局部特征和上下文编码器学习聚合源点云和目标点云上下文信息的特征; 然后将不同层次的上下文特征输入卷积网络组成的场景流解码器输出三维场景流。PointFlowNet 整合多任务结果输出, 使用场景流真值、刚体运动真值和相应的预测值计算损失函数, 增强特征提取模块和上下文编码模块的参数训练。在添加遮挡、噪音的增强 KITTI 数据集上测试, 该方法与 Dewan 等 (2016) 基线方法相比, 具有更高的精度和更快的效率。Ouyang 和 Raviv (2021) 提出了能同时估计场景流和检测遮挡区域的网络 OGSF-Net。该方法对场景流和遮挡检测耦合学习, 利用预测的遮挡掩码对流估计进行精化。OGSF-Net 采用特征金字塔结构, 对点云降采样并使用 PointConv (Wu 等, 2019) 学习每个点的局部特征, 在每层金字塔中, 使用上一层的上采样流将目标点云向源点云扭曲, 然后使用来自点云的特征和上一层的上采样遮挡掩膜输入 PointConv (Wu 等, 2019) 和 MLP 组成的 Estimator 分别输出点云场景流和遮挡掩码, 以由粗到精的方式进行更精细的预测。在有遮挡的 Flyingthings3D 上训练测试, OGSF-Net 比现有先进场景流估计方法 PointPWC-Net (Wu 等, 2020), HPLFlowNet (Gu 等, 2019), FLOT (Puy 等, 2020) 具有更高的精度, 表现出对遮挡的鲁棒性。

3.3.3 无监督或半监督的网络

由于缺少真值场景流标签, 使用变换后的源点云和目标点云的对齐损失函数作为监督信号是常用方法。Li 等 (2022) 提出了用于场景流估计的无监督学习框架 RigidFlow。该框架采用分割方法将点云分解为超体素, 并求解每个超体素的独立刚性配准。基于刚性变换估计, 生成每个超体素的刚性流, 从而近似生成整个场景流自监督学

习的伪标签。该方法以 FLOT (Puy 等, 2020) 估计的场景流作为基准, 以其与本方法估计标签的 L1 范数损失函数作为监督信号进行自监督学习。但 RigidFlow 的伪标签生成方法是基于局部刚性假设, 对于具有强烈非刚性运动的场景, 局部刚性假设可能不适用。针对现有用于场景流估计的合成数据集与真实场景差异较大, 造成模型迁移后性能退化的问题, Jin 等 (2022) 用游戏引擎构建了大规模逼真场景合成数据集, 并提出了用于点云场景流估计的无监督域适应框架 DCA-SRSFE。该框架采用教师—学生的学习范式。首先将合成数据输入学生模型进行有监督训练, 用得到的学生模型权重以指数移动平均 EMA (Exponential Moving Average) 的方式更新教师模型参数。然后, 教师模型为无标注真实点云数据生成伪场景流, 同时学生模型为无标注真实点云数据生成预测场景流, 以伪场景流和预测场景流的一致性约束损失不断更新模型参数, 提高学生模型对真实数据的认识, 从而获得对真实数据集的泛化性。针对现有方法使用离线聚类方法生成的固定超点不可更新导致具有不同流动模式的点聚类到相同的超点中的问题, Shen 等 (2023) 提出了基于超点引导的无监督迭代场景流估计框架 SPFlowNet。该框架由流引导超点生成模块和超点引导流细化模块组成, 两个模块交替执行迭代多次。超点是具有相似几何特征的点集 (包括坐标、流和特征信息)。流细化模块的功能是由超点中心的流值和关联图得到超点层面新的流, 使用反向内插由超点层面新的流得到点层面新的流。在超点生成模块, 使用前次迭代的点层面和超点层面的流计算超点中心和点云中所有点的关联图, 之后使用关联图更新超点中心 (新的坐标、流和特征信息)。通过最小化流变换后的源和目标点云的倒角距离、约束局部区域的流值的相似性、增强源点云的流和目标点云的反向流的一致性, 实现该网络的无监督训练。在 KITTI 和 FT3D 数据集的实验表明, SPFlowNet 精度优于 RigidFlow (Li 等, 2022) 和 Self-Point-Flow (Li 等, 2021b) 等最近的先进方法, 尤其在 KITTI 数据上得到实质的提升。

3.3.4 提高效率的网络

Gu 等 (2019) 提出了基于六面体结构的网络 HPLFlowNet。鉴于 FlowNet3D (Liu 等, 2019b) 受

到其单一流相关性的制约, HPLFlowNet 使用双边卷积层 BCL (Bilateral Convolutional Layers) 捕获多尺度相关性, 并提出了 DownBCL、UpBCL 和 CorrBCL 来恢复非结构化点云的结构信息。其中 CorrBCL 接收多尺度的源点云和目标点云特征并融合以学习两帧点云的几何信息和关联特征, 然后在上采样阶段通过多层不同尺度的双边卷积得到与源点云数目相同的点及其特征, 最后将特征输入 1×1 卷积模块输出源点云向目标点云的场景流估计。HPLFlowNet 的处理时间不随输入点数的增加而线性增加, 在点数达 8.6 万的情况下依旧保持较高的效率。

总的来说, 基于场景流估计的方法通过关联两帧点云的局部特征来估计源点云中每个点的平移量。该类方法适合应用于场景中的目标有不同运动参数的情况。然而, 该类方法大多忽略了每个目标内部间的运动一致性问题, 且对遮挡和噪声敏感。面对复杂的自动驾驶场景, 如何同时考虑非刚性运动的行人、刚性运动的车辆和传感器自身的运动情况, 以及每个目标内在的运动约束, 是该类方法需要研究的关键内容之一。

4 数据集

本文根据数据集范围大小与复杂程度, 如表 2 所示, 将现有的点云配准数据集主要分为小目标、室内场景和室外场景 3 大类。

小目标数据集通常为单个目标不同角度的 2.5 维扫描数据。这类数据集几何形状较简单, 数量较小, 点云密集且分布较均匀, 噪声较少。然而, 该类数据存在以下问题: (1) 局部曲面特征不明显, 特征可区分性不高; (2) 初始位置相差较大, 易陷入局部最优; (3) 常常包含对称目标, 易导致匹配的歧义性。这给点云配准任务带来了较大的挑战。典型的数据集有 Stanford 3D Scanning Repository、UWA Dataset 和 Kinect Views Dataset。随着深度学习的发展, 小样本的数据集无法满足基于深度学习的网络训练要求, 因此, 研究学者们尝试着用大型的模型数据集来模拟生成不同角度的配准数据, 例如 ModelNet40、ScanObjectNN 和 ShapeNetCore, 来对配准网络进行训练。

室内场景数据集通常为单个房间不同角度的扫描数据。该类数据集每帧点云数量适中, 点云密度较大且分布均匀。然而, 该类场景存在以下

问题：(1) 平面结构较多，从而导致以点为匹配基元的方法错误匹配较多且效率较低；(2) 自遮挡现象较严重，导致部分局部结构严重缺失；(3) 数据质量一般，噪声较多，从而导致局部结构被破坏。典型数据集有 3DMatch、ICL-NUIM、Indoor LiDAR RGB-D dataset 和 Stanford S3DIS indoor dataset。

表2 点云配准数据集
Table 2 Data set of point cloud registration

类型	数据集名称	获取设备	真值获取	具体描述	典型算法	网址
小目标	Stanford 3D Scanning Repository	Cyberware 3030 scanner	ICP	包含8个模型,10、12、60、70、114、47、18和36个2.5维扫描数据	PRNet(Wang等,2019d)	http://graphics.stanford.edu/data/3Dscanrep/
	UWA Dataset	Minolta Vivid 910 scanner	无	包含4个模型22、16、16和21个2.5维扫描数据	TOLDI(Yang等,2017)	http://www.csse.uwa.edu.au/~ajmal/3Dmodeling
	Kinect Views Dataset	Microsoft Kinect sensor	无	包含6个模型15、16、20、13、16和15个2.5维扫描数据	LoVS(Quan等,2018)	http://vision.deis.unibo.it/research/78-cvlab/80-shot
	ModelNet40	Synthetic	模拟	包含40个CAD类别,9843个模型作为训练数据和2468个模型作为测试数据	RPM-Net(Yew和Lee,2020);PointNetLK(Aoki等,2019)	https://modelnet.cs.princeton.edu/
	ModelNet10	Synthetic	模拟	包含10个CAD类别,3991个模型作为训练数据和605个模型作为测试数据	—	https://modelnet.cs.princeton.edu/
	ScanObjectNN	RGB-D sensor	模拟	从SceneNN和ScanNet数据集中采样生成15类约15000个对象	ReAgent(Bauer等,2021)	https://hkust-vgd.github.io/scanobjectnn/
室内场景	ShapeNetCore	Synthetic	模拟	包含55个CAD类别51300个模型,12个类别作为测试数据	PRNet(Wang等,2019d)	https://shapenet.org/
	ALS Dataset	Hokuyo UTM-30LX	IMU/GPS/Theodolite	包含3个场景(Stairs,ETH Hauptgebaude,Apartment)31、36和45个扫描数据	TOLDI(Yang等,2017)	https://projects.asl.ethz.ch/datasets
	3DMatch	Microsoft Kinect, Structure Sensor, Asus Xtion Pro Live, Intel RealSense	ICP	包含7Scenes,SUN3D,and RGB-D Scenes v.2等62个RGB-D场景,54个场景作为训练,8个场景作为测试	PREDATOR(Huang等,2021a);RoReg(Wang等,2023b);DGR(Choy等,2020)	https://3dmatch.cs.princeton.edu/
	ICL-NUIM	Synthetic	Kintinuuous	包含4601个卧室扫描数据和4601个办公室扫描数据	DGR(Choy等,2020)	https://www.doc.ic.ac.uk/~ahanda/VaFRIC/iclnuim.html
	ScanNet	Structure sensor	手工对齐和算法优化	在707个不同空间采集的21类1513个2.5维RGB-D图像,1201个用于训练和312个用于测试	—	http://www.scan-net.org/#code-and-data
	S3DIS Dataset	FARO Focus 3D X330 HDR scanner	生厂商提供的软件	包含在3个不同建筑物采集的5个大规模室内场景(Apartment, Bedroom, Boardroom, Lobby, and Loft)	PointNetLK(Aoki等,2019)	http://redwood-data.org/indoor_lidar_rgbd/download.html
	FlyingThings3D	Synthetic	模拟	从ShapeNet数据集中生成的多个随机移动对象场景,19640个作为训练,3824个作为测试	FlowNet3D(Wang等,2021);Bi-PointFlow-Net(Cheng和Ko,2022)	https://lmb.informatik.uni-freiburg.de/resources/datasets/SceneFlowDatasets.en.html

续表

类型	数据集名称	获取设备	真值获取	具体描述	典型算法	网址
室外场景	KITTI Scene Flow 2015	Velodyne lidar sensor	半自动	由 KITTI 原始数据构建动态运动生成, 由 200 个训练场景和 200 个测试场景组成	PV-RAFT(Wei 等, 2021); Bi-PointFlow-Net(Cheng 和 Ko, 2022)	http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php
	odometryKITTI	Velodyne lidar sensor	GPS/OXTS	序列 00—10(23201 个扫描数据)用于训练, 序列 11—21(20351 个扫描数据)用于测试	LO-Net(Li 等, 2019); DeepVCP(Lu 等, 2019)	http://www.cvlibs.net/datasets/kitti/eval_odometry.php
	Apollo-SouthBay Dataset	Velodyne HDL-64E LiDAR	GNSS/IMU	包括住宅区、城市市中心区和高速公路等各种情况	DeepVCP(Lu 等, 2019)	https://apollo.auto/southbay.html
	WHU-TLS	IMAGER 5010C, VZ-400, ScanStation C5, Leica HDS6100, Leica P40	HMMR (Dong 等, 2018)	来自 11 个不同环境(地铁站、高速铁路平台、山地、森林、公园、校园、住所、河岸、遗产建筑、地下挖掘和隧道)的 115 个扫描数据	BSC(Dong 等, 2017)	http://3s.whu.edu.cn/ybs/en/benchmark.htm
	FORD	Velodyne lidar scanner, Riegl lidars	Applanix POS LV/GPS/IMU	包含不同城市环境的 2 个闭环序列, 比 KITTI 数据集有更多的移动车辆	LO-Net(Li 等, 2019)	http://robots.engine.umich.edu/SoftwareData/Ford
	ALS Dataset	Hokuyo UTM-30LX	IMU/GPS/Theodolite	分别对夏季和冬季的 Gazebo 和冬季和秋季的 Wood 进行了 31、32、37 和 32 次扫描	—	https://projects.asl.ethz.ch/datasets

室外场景通常为车载激光扫描的大规模道路场景。该类场景具有较大的挑战性, 主要体现在: (1) 点云数量巨大, 复杂的网络结构和普通的设备算力无法满足时效性的要求; (2) 密度分布不均匀且存在“近密远疏”的特点, 导致重叠区域的点并不是一一对应的; (3) 常存在与扫描仪器的刚性运动不一致的动态移动的目标(如行人, 车辆等); (4) 自遮挡及孔洞现象, 导致局部结构的严重缺失。典型的数据有 odometryKITTI、Apollo-SouthBay Dataset 和 FORD 数据集。

5 结 语

本文从深度学习方法角度对点云配准技术相关研究进行了阐述。总的来说, 基于深度学习的方法能够自动学习出高判别力的点云特征, 以及自动实现匹配点查找, 但是配准效果依赖于大量标注的训练数据。具体来说: (1) 基于深度学习的点云配准方法研究起步较晚, 大多集中在基于对应关系解算的点云配准方面。基于对应关系解算的配准方法精度高但效率低。(2) 基于姿态回归的方法在鲁棒性和效率方面有很大的优势, 但大多采用迭代来追求更高的精度。(3) 基于场景流估计的方法不仅可以处理刚性变换, 也可以处理非刚性变换, 但由于输出的维度较高, 参数量

较大, 效率较低。

通过本文的总结分析可以看出, 在三维点云配准领域, 基于深度学习的点云配准方法近年来发展迅速, 已成为当前的研究热点, 并且在可预见的未来仍会成为主流的研究方向。展望未来, 三维点云配准可以在以下 3 个方面进行探索:

(1) 现有的基于深度学习的点云配准方法需要大量的标注数据, 然而数据标注时间和人力成本高。小样本学习、自监督、弱监督以及无监督学习等方法在一定程度上可以缓解深度学习技术对样本标注量的需求。未来可考虑将这些方法应用于三维点云配准问题。

(2) 现有的基于深度学习的点云配准方法中匹配基元仍然是以三维点为主。相比于三维线段、平面等几何结构元素, 三维点的歧义性大, 误匹配率高。目前, 已有学者利用深度霍夫变换技术从二维图像中提取语义线段(Zhao 等, 2022)。未来也可考虑利用深度学习技术从三维点云数据中自动提取场景的几何结构元素, 然后进一步基于几何结构元素解决真实场景下三维点云的配准问题。

(3) 无论是传统方法还是基于深度学习的方法, 主要利用的信息还是场景的空间几何结构特征, 对场景语义信息的利用较少。近年来, 基于

深度学习的点云语义理解技术进展迅速, 场景语义信息获取越来越方便, 精度也越来越高。因此, 未来可考虑利用结合空间几何结构信息和语义信息来解决复杂场景的三维点云配准问题。

参考文献(References)

- 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: IEEE: 11748-11757 [DOI: 10.1109/CVPR46437.2021.01158]
- Aoki Y, Goforth H, Srivatsan R A and Lucey S. 2019. PointNetLK: robust and efficient point cloud registration using PointNet//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 7163-7172 [DOI: 10.1109/CVPR.2019.00733]
- Bauer D, Patten T and Vincze M. 2021. ReAgent: point cloud registration using imitation and reinforcement learning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 14581-14589 [DOI: 10.1109/CVPR46437.2021.01435]
- Behl A, Paschalidou D, Donné S and Geiger A. 2019. PointFlowNet: learning representations for rigid motion estimation from point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 7954-7963 [DOI: 10.1109/CVPR.2019.00815]
- Besl P J and McKay H D. 1992. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2): 239-256 [DOI: 10.1109/34.121791]
- Bouaziz S, Tagliasacchi A and Pauly M. 2013. Sparse iterative closest point//Proceedings of the 11th Eurographics/ACMSIGGRAPH Symposium on Geometry Processing. Genova: Eurographics Association: 113-123 [DOI: 10.1111/cgf.12178]
- Cao A Q, Puy G, Boulch A and Marlet R. 2021. PCAM: product of cross-attention matrices for rigid registration of point clouds//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 13229-13238 [DOI: 10.1109/ICCV48922.2021.01298]
- Chang M F, Lambert J, Sangkloy P, Singh J, Bak S, Hartnett A, Wang D, Carr P, Lucey S, Ramanan D and Hays J. 2019. Argoverse: 3D tracking and forecasting with rich maps//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 8740-8749 [DOI: 10.1109/CVPR.2019.00895]
- Cheng L, Chen S, Liu X Q, Xu H, Wu Y, Li M C and Chen Y M. 2018. Registration of laser scanning point clouds: a review. Sensors, 18(5): 1641 [DOI: 10.3390/s18051641]
- Cheng W C and Ko J H. 2022. Bi-PointFlowNet: bidirectional learning for point cloud based scene flow estimation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer: 108-124 [DOI: 10.1007/978-3-031-19815-1_7]
- Choy C, Dong W and Koltun V. 2020. Deep global registration//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 2514-2523 [DOI: 10.1109/CVPR42600.2020.00259]
- Choy C, Park J and Koltun V. 2019. Fully convolutional geometric features//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 8957-8965 [DOI: 10.1109/ICCV.2019.00905]
- Cuturi M. 2013. Sinkhorn distances: lightspeed computation of optimal transport//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc.: 2292-2300
- Deng H W, Birdal T and Ilic S. 2018a. PPF-FoldNet: unsupervised learning of rotation invariant 3D local descriptors//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer: 620-638 [DOI: 10.1007/978-3-030-01228-1_37]
- Deng H W, Birdal T and Ilic S. 2018b. PPFNet: global context aware local features for robust 3D point matching//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 195-205 [DOI: 10.1109/CVPR.2018.00028]
- Deng X, Zhang W Y, Ding Q and Zhang X M. 2023. PointVector: a vector representation in point cloud analysis//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 9455-9465 [DOI: 10.1109/CVPR52729.2023.00912]
- Dewan A, Caselitz T, Tipaldi G D and Burgard W. 2016. Rigid scene flow for 3D LiDAR scans//Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. Daejeon: IEEE: 1765-1770 [DIO: 10.1109/IROS.2016.7759282]
- Dong Z, Yang B S, Liang F X, Huang R G and Scherer S. 2018. Hierarchical registration of unordered TLS point clouds based on binary shape context descriptor. ISPRS Journal of Photogrammetry and Remote Sensing, 144: 61-79 [DOI: 10.1016/j.isprsjprs.2018.06.018]
- Dong Z, Yang B S, Liu Y, Liang F X, Li B J and Zang Y F. 2017. A novel binary shape context for 3D local surface description. ISPRS Journal of Photogrammetry and Remote Sensing, 130: 431-452 [DOI: 10.1016/j.isprsjprs.2017.06.012]
- Dosovitskiy A, Fischer P, Ilg E, Häusser P, Hazirbas C, Golkov V, van der Smagt P, Cremers D and Brox T. 2015. FlowNet: learning optical flow with convolutional networks//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE: 2758-2766 [DOI: 10.1109/ICCV.2015.316]
- 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: IEEE: 998-1005 [DOI: 10.1109/CVPR.2010.5540108]
- Fan W H, Lin X, Luo H, Guo W Z, Wang H Y and Dai C G. 2024. Domain adaptation learning for 3D point clouds: A survey. National Remote Sensing Bulletin, 28(4): 825-842 (范文辉, 林茜, 罗欢, 郭文忠, 汪汉云, 戴晨光. 2024. 面向三维点云的域自适应学习. 遥感学报, 28(4): 825-842) [DOI: 10.11834/jrs.20233140]
- Fischer K, Simon M, Ölsner F, Milz S, Groß H M and Mäder P. 2021.

- StickyPillars: robust and efficient feature matching on point clouds using graph neural networks//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 313-323 [DOI: 10.1109/CVPR46437.2021.00038]
- Fischler M A and Bolles R C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381-395 [DOI: 10.1145/358669.358692]
- 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: Springer: 224-237 [DOI: 10.1007/978-3-540-24672-5_18]
- Fu K X, Luo J Z, Luo X Y, Liu S L, Zhang C X and Wang M N. 2023. Robust point cloud registration framework based on deep graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 6183-6195 [DOI: 10.1109/TPAMI.2022.3204713]
- Gower J C. 1975. Generalized procrustes analysis. *Psychometrika*, 40(1): 33-51 [DOI: 10.1007/BF02291478]
- Gu X Y, Wang Y J, Wu C R, Lee Y J and Wang P Q. 2019. HPLFlowNet: hierarchical permutohedral lattice FlowNet for scene flow estimation on large-scale point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 3249-3258 [DOI: 10.1109/CVPR.2019.00337]
- Guo Y L, Bennamoun M, Soheli F, Lu M, Wan J W and Kwok N M. 2016. A comprehensive performance evaluation of 3D local feature descriptors. *International Journal of Computer Vision*, 116(1): 66-89 [DOI: 10.1007/s11263-015-0824-y]
- Guo Y L, Soheli 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(1): 63-86 [DOI: 10.1007/s11263-013-0627-y]
- Horache S, Deschaud J E and Goulette F. 2021. 3D point cloud registration with multi-scale architecture and unsupervised transfer learning//Proceedings of 2021 International Conference on 3D Vision. London: IEEE: 1351-1361 [DOI: 10.1109/3DV53792.2021.00142]
- Huang S Y, Gojcic Z, Usvyatsov M, Wieser A and Schindler K. 2021a. PREDATOR: registration of 3D point clouds with low overlap//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 4265-4274 [DOI: 10.1109/CVPR46437.2021.00425]
- Huang X S, Mei G F and Zhang J. 2020. Feature-metric registration: a fast semi-supervised approach for robust point cloud registration without correspondences//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 11363-11371 [DOI: 10.1109/CVPR42600.2020.01138]
- Huang X S, Mei G F, Zhang J and Abbas R. 2021b. A comprehensive survey on point cloud registration. *arXiv*: 2103.02690
- Jin Z, Lei Y J, Akhtar N, Li H F and Hayat M. 2022. Deformation and correspondence aware unsupervised synthetic-to-real scene flow estimation for point clouds//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 7223-7233 [DOI: 10.1109/CVPR52688.2022.00709]
- 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]
- Lee J, Kim S, Cho M and Park J. 2021. Deep Hough voting for robust global registration//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 15994-16003 [DOI: 10.1109/ICCV48922.2021.01569]
- Li Q, Chen S Y, Wang C, Li X, Wen C L, Cheng M and Li J. 2019. LO-Net: deep real-time lidar odometry//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 8465-8474 [DOI: 10.1109/CVPR.2019.00867]
- Li R B, Lin G S, He T, Liu F Y and Shen C H. 2021a. HCRF-Flow: scene flow from point clouds with continuous high-order CRFs and position-aware flow embedding//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 364-373 [DOI: 10.1109/CVPR46437.2021.00043]
- Li R B, Lin G S and Xie L H. 2021b. Self-point-flow: self-supervised scene flow estimation from point clouds with optimal transport and random walk//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 15572-15581 [DOI: 10.1109/CVPR46437.2021.01532]
- Li R B, Zhang C, Lin G S, Wang Z and Shen C H. 2022. RigidFlow: self-supervised scene flow learning on point clouds by local rigidity prior//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 16938-16947 [DOI: 10.1109/CVPR52688.2022.01645]
- Li X Q, Pontes J K and Lucey S. 2021c. PointNetLK Revisited//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 12758-12767 [DOI: 10.1109/CVPR46437.2021.01257]
- Liu X H, Han Z Z, Liu Y S and Zwicker M. 2019a. Point2Sequence: learning the shape representation of 3D point clouds with an attention-based sequence to sequence network//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu: AAAI Press: 8770-8777 [DOI: 10.1609/aaai.v33i01.33018778]
- Liu X Y, Qi C R and Guibas L J. 2019b. FlowNet3D: learning scene flow in 3D point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 529-537 [DOI: 10.1109/CVPR.2019.00062]
- Liu Z, Zhou S B, Suo C Z, Yin P, Chen W, Wang H S, Li H A and Liu Y H. 2019c. LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 2831-2840 [DOI: 10.1109/ICCV.2019.00292]
- Liu Z J, Tang H T, Lin Y J and Han S. 2019d. Point-voxel CNN for efficient 3D deep learning//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc.: 87
- Lu W X, Wan G W, Zhou Y, Fu X Y, Yuan P F and Song S Y. 2019. DeepVCP: an end-to-end deep neural network for point cloud registration//Proceedings of 2019 IEEE/CVF International Confer-

- ence on Computer Vision. Seoul: IEEE: 12-21 [DOI: 10.1109/IC-CV.2019.00010]
- Lucas B D and Kanade T. 1981. An iterative image registration technique with an application to stereo vision//Proceedings of the 7th International Joint Conference on Artificial Intelligence. Vancouver: Morgan Kaufmann Publishers Inc.: 674-679
- Ma X, Qin C, You H X, Ran H X and Fu Y. 2022. Rethinking network design and local geometry in point cloud: a simple residual MLP framework//Proceedings of the 10th International Conference on Learning Representations. [s.l.]: OpenReview.net
- Maturana D and Scherer S. 2015. VoxNet: a 3D convolutional neural network for real-time object recognition//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg: IEEE: 922-928 [DOI: 10.1109/IROS.2015.7353481]
- Mei G F, Tang H, Huang X S, Wang W J, Liu J, Zhang J, Van Gool L and Wu Q. 2023. Unsupervised deep probabilistic approach for partial point cloud registration//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 13611-13620 [DOI: 10.1109/CVPR52729.2023.01308]
- Mohr L, Geles I and Fraundorfer F. 2023. GAFAR: graph-attention feature-augmentation for registration a fast and light-weight point set registration algorithm//Proceedings of 2023 European Conference on Mobile Robots. Coimbra: IEEE: 1-8 [DOI: 10.1109/ECMR59166.2023.10256277]
- Nubert J, Khattak S and Hutter M. 2021. Self-supervised learning of LiDAR odometry for robotic applications//Proceedings of 2021 IEEE International Conference on Robotics and Automation. Xi'an: IEEE: 9601-9607 [DOI: 10.1109/ICRA48506.2021.9561063]
- Ouyang B J and Raviv D. 2021. Occlusion guided scene flow estimation on 3D point clouds//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville: IEEE: 2799-2808 [DOI: 10.1109/CVPRW53098.2021.00315]
- Prakhya S M, Liu B B and Lin W S. 2015. B-SHOT: a binary feature descriptor for fast and efficient keypoint matching on 3D point clouds//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg: IEEE: 1929-1934 [DOI: 10.1109/IROS.2015.7353630]
- Puy G, Boulch A and Marlet R. 2020. FLOT: scene flow on point clouds guided by optimal transport//Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer: 527-544 [DIO: 10.1007/978-3-030-58604-1_32]
- Qi C R, Su H, Mo K C 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: 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: Curran Associates Inc.: 5105-5114
- Qin C, You H X, Wang L C, Kuo C C J and Fu Y. 2019. PointDAN: a multi-scale 3D domain adaption network for point cloud representation//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc.: 646
- Qin H X, Liu Z T and Tan B Y. 2022. Review on deep learning rigid point cloud registration. *Journal of Image and Graphics*, 27(2): 329-348 (秦红星, 刘镇涛, 谭博元. 2022. 深度学习刚性点云配准前沿进展. *中国图象图形学报*, 27(2): 329-348) [DOI: 10.11834/jig.210556]
- Qin Z, Yu H, Wang C, 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: IEEE: 11133-11142 [DOI: 10.1109/CVPR52688.2022.01086]
- Quan S W, Ma J, Hu F Y, Fang B and Ma T. 2018. Local voxelized structure for 3D binary feature representation and robust registration of point clouds from low-cost sensors. *Information Sciences*, 444: 153-171 [DOI: 10.1016/j.ins.2018.02.070]
- Ran H X, Liu J and Wang C J. 2022. Surface representation for point clouds//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 18942-18952 [DOI: 10.1109/CVPR52688.2022.01837]
- Riegler G, Ulusoy A O and Geiger A. 2017. OctNet: learning deep 3D representations at high resolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 6620-6629 [DOI: 10.1109/CVPR.2017.701]
- 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: IEEE: 3212-3217 [DOI: 10.1109/ROBOT.2009.5152473]
- Saleh M, Dehghani S, Busam B, Navab N and Tombari F. 2020. Graphite: graph-induced feature extraction for point cloud registration//Proceedings of 2020 International Conference on 3D Vision. Fukuoka: IEEE: 241-251 [DOI: 10.1109/3DV50981.2020.00034]
- Sarode V, Li X Q, Goforth H, Aoki Y, Srivatsan R A, Lucey S and Cho-set H. 2019. PCNet: point cloud registration network using pointnet encoding. arXiv: 1908.07906
- Segal A V, Haehnel D and Thrun S. 2009. Generalized-ICP//Proceedings of the Robotics: Science and Systems V, University of Washington. Seattle: [s.n.]: 435 [DOI: 10.15607/RSS.2009.V.021]
- Shen Y Q, Hui L, Xie J and Yang J. 2023. Self-supervised 3D scene flow estimation guided by superpoints//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 5271-5280 [DOI: 10.1109/CVPR52729.2023.00510]
- Shen Y R, Feng C, Yang Y Q and Tian D. 2018. Mining point cloud local structures by kernel correlation and graph pooling//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 4548-4557 [DOI: 10.1109/CVPR.2018.00478]
- Stein F and Medioni G. 1992. Structural indexing: efficient 3-D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2): 125-145 [DOI: 10.1109/34.121785]
- 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: IEEE: 6410-6419 [DOI: 10.1109/ICCV.2019.00651]
- Wang G M, Wu X R, Jiang S Y, Liu Z and Wang H S. 2023a. Efficient 3D deep LiDAR odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5749-5765 [DOI: 10.1109/TPAMI.2022.3207015]
- Wang G M, Wu X R, Liu Z and Wang H S. 2021a. Hierarchical attention learning of scene flow in 3D point clouds. *IEEE Transactions on Image Processing*, 30: 5168-5181 [DOI: 10.1109/TIP.2021.3079796]
- Wang G M, Wu X R, Liu Z and Wang H S. 2021b. PWCLO-Net: deep LiDAR odometry in 3D point clouds using hierarchical embedding mask optimization//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 15905-15914 [DOI: 10.1109/CVPR46437.2021.01565]
- Wang H P, Liu Y, Hu Q Y, Wang B, Chen J G, Dong Z, Guo Y L, Wang W P and Yang B S. 2023b. RoReg: pairwise point cloud registration with oriented descriptors and local rotations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10376-10393 [DOI: 10.1109/TPAMI.2023.3244951]
- Wang L J, Hao Y, Li X and Fang Y. 2020a. 3D meta-registration: learning to learn registration of 3D point clouds. *arXiv*: 2010.11504
- Wang L J, Li X and Fang Y. 2020b. Deep-3DAligner: unsupervised 3D point set registration network with optimizable latent vector. *arXiv*: 2010.00321
- Wang P, Wang C, Xi X H, Nie S and Du M. 2024. Automatic extraction of high-voltage transmission pylons with multifeature constraints. *National Remote Sensing Bulletin*, 28(10): 2651-2660 (王濮, 王成, 习晓环, 聂胜, 杜蒙. 2024. 多特征约束的输电通道杆塔点云提取. *遥感学报*, 28(10): 2651-2660) [DOI: 10.11834/jrs.20232684]
- Wang R Z, Yan J C and Yang X K. 2019a. Learning combinatorial embedding networks for deep graph matching//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 3056-3065 [DOI: 10.1109/ICCV.2019.00315]
- Wang W, Saputra M R U, Zhao P J, Gusmao P, Yang B, Chen C H, Markham A and Trigoni N. 2019b. DeepPCO: end-to-end point cloud odometry through deep parallel neural network//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE: 3248-3254 [DOI: 10.1109/IROS40897.2019.8967756]
- Wang Y and Solomon J. 2019c. Deep closest point: learning representations for point cloud registration//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 3522-3531 [DOI: 10.1109/ICCV.2019.00362]
- Wang Y and Solomon J. 2019d. PRNet: self-supervised learning for partial-to-partial registration//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc.: 791
- Wang Y, Sun Y B, Liu Z W, Sarma S E, Bronstein M M and Solomon J M. 2019e. Dynamic graph CNN for learning on point clouds. *ACM Transactions on Graphics*, 38(5): 146 [DOI: 10.1145/3326362]
- Wang Y J, Yan C G, Feng Y T, Du S Y, Dai Q H and Gao Y. 2023c. STORM: structure-based overlap matching for partial point cloud registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 1135-1149 [DOI: 10.1109/TPAMI.2022.3148308]
- Wang Z R, Li S D, Howard-Jenkins H, Prisacariu V A and Chen M. 2020c. FlowNet3D++: geometric losses for deep scene flow estimation//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass: IEEE: 91-98 [DOI: 10.1109/WACV45572.2020.9093302]
- Wei H S, Qiao Z J, Liu Z, Suo C Z, Yin P, Shen Y L, Li H A and Wang H S. 2020. End-to-End 3D point cloud learning for registration task using virtual correspondences//Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas: IEEE: 2678-2683 [DOI: 10.1109/IROS45743.2020.9341249]
- Wei Y, Wang Z Y, Rao Y M, Lu J W and Zhou J. 2021. PV-RAFT: point-voxel correlation fields for scene flow estimation of point clouds//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 6954-6963 [DOI: 10.1109/CVPR46437.2021.00688]
- Wu W X, Qi Z G, and Li F X. 2019. PointConv: deep convolutional networks on 3D point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 9613-9622 [DOI: 10.1109/CVPR.2019.00985]
- Wu W X, Wang Z Y, Li Z W, Liu W and Li F X. 2020. PointPWC-Net: cost volume on point clouds for (self-) supervised scene flow estimation//Proceedings of the 16th European Conference. Glasgow: Springer: 88-107 [DOI: 10.1007/978-3-030-58558-7_6]
- 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: IEEE: 1912-1920 [DOI: 10.1109/CVPR.2015.7298801]
- Xiong F G. 2018. Research on Registration Technology of 3D Point Cloud. Taiyuan: North University of China (熊风光. 2018. 三维点云配准技术研究. 太原: 中北大学)
- Xu H, Liu S C, Wang G F, Liu G H and Zeng B. 2021. OMNet: learning overlapping mask for partial-to-partial point cloud registration//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 3112-3121 [DOI: 10.1109/ICCV48922.2021.00312]
- Xu H, Ye N J, Liu G H, Zeng B and Liu S C. 2022. FINet: dual branches feature interaction for partial-to-partial point cloud registration//Proceedings of the 36th AAAI Conference on Artificial Intelligence. [s. l.]: AAAI Press: 2848-2856 [DOI: 10.1609/aaai.v36i3.20189]
- Yang B S, Han X and Dong Z. 2021. Point cloud benchmark dataset WHU-TLS and WHU-MLS for deep learning. *National Remote Sensing Bulletin*, 25(1): 231-240 (杨必胜, 韩旭, 董震. 2021. 点云深度学习基准数据集. *遥感学报*, 25(1): 231-240) [DOI: 10.11834/jrs.20210542]
- Yang J L, Li H D, Campbell D and Jia Y D. 2016. Go-ICP: a globally optimal solution to 3D ICP point-set registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11): 2241-

- 2254 [DOI: 10.1109/TPAMI.2015.2513405]
- Yang J Q, Zhang Q, Xiao Y and Cao Z G. 2017. TOLDI: an effective and robust approach for 3D local shape description. *Pattern Recognition*, 65: 175-187 [DOI: 10.1016/j.patcog.2016.11.019]
- Yew Z J and Lee G H. 2020. RPM-Net: robust point matching using learned features//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE: 11824-11833 [DOI: 10.1109/CVPR42600.2020.01184]
- 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: IEEE: 6667-6676 [DOI: 10.1109/CVPR52688.2022.00656]
- Yin J, Li A, Li T, Yu W X and Zou D P. 2022. M2DGR: a multi-sensor and multi-scenario SLAM dataset for ground robots. *IEEE Robotics and Automation Letters*, 7(2): 2266-2273 [DOI: 10.1109/LRA.2021.3138527]
- You H X, Feng Y F, Ji R R and Gao Y. 2018. PVNet: a joint convolutional network of point cloud and multi-view for 3D shape recognition//*Proceedings of the 26th ACM International Conference on Multimedia*. Seoul: Association for Computing Machinery: 1310-1318 [DOI: 10.1145/3240508.3240702]
- Yu H, Li F, Saleh M, Busam B and Ilic S. 2021. CoFiNet: reliable coarse-to-fine correspondences for robust point cloud registration//*Proceedings of the 35th International Conference on Neural Information Processing System*. [s.l.]: Curran Associates Inc.: 1828
- Yu H, Qin Z, Hou J, Saleh M, Li D S, Busam B and Ilic S. 2023. Rotation-invariant transformer for point cloud matching//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE: 5384-5393 [DOI: 10.1109/CVPR52729.2023.00521]
- Yuan W T, Eckart B, Kim K, Jampani V, Fox D and Kautz J. 2020. DeepGMR: learning latent Gaussian mixture models for registration//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer: 733-750 [DOI: 10.1007/978-3-030-58558-7_43]
- 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: IEEE: 1802-1811 [DOI: 10.1109/CVPR.2017.29]
- Zhang J and Singh S. 2017. Low-drift and real-time lidar odometry and mapping. *Autonomous Robots*, 41(2): 401-416 [DOI: 10.1007/s10514-016-9548-2]
- Zhang Q J, Hou J H, Qian Y, Zhang J Y and He Y. 2020a. ParaNet: deep regular representation for 3D point clouds. *arXiv*: 2012.03028
- Zhang Y H, Li C H, Guo B, Guo C H and Zhang S L. 2021. KDD: a kernel density based descriptor for 3D point clouds. *Pattern Recognition*, 111: 107691 [DOI: 10.1016/j.patcog.2020.107691]
- Zhang Z Y, Dai Y C and Sun J D. 2020b. Deep learning based point cloud registration: an overview. *Virtual Reality and Intelligent Hardware*, 2(3): 222-246 [DOI: 10.1016/j.vrih.2020.05.002]
- Zhao K, Han Q, Zhang C B, Xu J and Cheng M M. 2022. Deep Hough transform for semantic line detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 4793-4806 [DOI: 10.1109/TPAMI.2021.3077129]
- Zheng C, Lyu Y C, Li M and Zhang Z M. 2020. LodoNet: a deep neural network with 2D keypoint matching for 3D LiDAR odometry estimation//*Proceedings of the 28th ACM International Conference on Multimedia*. Seattle: ACM: 2391-2399 [DIO: 10.1145/3394171.3413771]
- Zhou J, Wang M J, Mao W D, Gong M L and Liu X P. 2020. Siamese-PointNet: a siamese point network architecture for learning 3D shape descriptor. *Computer Graphics Forum*, 39(1): 309-321 [DOI: 10.1111/cgf.13804]
- Zhou R Q, Li X X and Jiang W S. 2021. SCANet: a spatial and channel attention based network for partial-to-partial point cloud registration. *Pattern Recognition Letters*, 151: 120-126 [DOI: 10.1016/j.patrec.2021.08.002]
- Zhou Y, Sun P, Zhang Y, Angelov D, Gao J Y, Ouyang T, Guo J, Ngiam J and Vasudevan V. 2019. End-to-end multi-view fusion for 3D object detection in LiDAR point clouds//*Proceedings of the 3rd Annual Conference on Robot Learning*. Osaka: PMLR: 923-932

Status and progress of deep learning-based pairwise point cloud rigid registration

ZHOU Ruqin¹, WANG Peng¹, DAI Chenguang¹, WANG Hanyun¹,
JIANG Wanshou², ZHANG Yongsheng¹

1. Institute of Geospatial Information, Information Engineering University, Zhengzhou 450001, China;
2. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract: Point cloud registration is the spatial alignment of two or more point clouds through geometric transformations. As a fundamental task in 3D point cloud data processing, point cloud registration is an important preprocessing operation for tasks, such as 3D modeling, object recognition, and scene understanding. Given the nonstructural, sparse, and uneven characteristics of point cloud data, point

cloud registration remains one of the hotspots and challenges in computer vision, mapping, and remote sensing, although many studies have been conducted on it. With the emergence and rapid development of neural networks, deep learning demonstrates huge utilization potential in applications, such as point cloud classification, recognition, detection, and reconstruction. In recent years, many researchers have attempted to apply deep learning techniques to point cloud registration. Deep learning-based point cloud registration methods can automatically learn highly discriminative and robust point cloud features that contain geometric structural and semantic information, which are crucial for achieving high registration accuracy. In this study, the research on pairwise point cloud rigid registration technology based on deep learning is systematically reviewed and analyzed. First, feature extraction networks based on deep learning are introduced. Second, the progress of research on registration methods, namely, correspondence estimation, pose regression, and scene flow estimation, is reviewed, and the characteristics, advantages, and disadvantages of the three methods are summarized. Third, this study systematically summarizes and categorizes existing publicly available datasets that can be used for rigid point cloud registration. Last, the status of current research on point cloud registration is determined; the advantages and limitations of existing methods in terms of feature learning, registration accuracy, registration efficiency, and other aspects are explained; and future research directions are proposed. Specifically, three exploration directions are presented. (1) Existing deep learning-based point cloud registration methods require a large amount of annotated data, and the data annotation time and manpower costs are high. Methods, such as few-shot learning, self-supervised learning, weakly supervised learning, and unsupervised learning, can alleviate deep learning technology's need for large amounts of labeled data to some extent. (2) The matching primitives in existing deep learning-based point cloud registration methods still focus on 3D points. Compared with geometric elements, such as 3D line segments and planes, 3D points have high ambiguity and mismatch rates. In the future, the use of deep learning techniques to automatically extract the geometric structural elements of scenes from 3D point cloud data should be considered, and the registration problem of 3D point clouds in large scenes should be solved based on these geometric structural elements. (3) Existing deep learning-based registration methods mainly utilize the spatial geometric features of scenes, with minimal consideration of semantic information. In recent years, rapid advances in deep learning-based point cloud semantic understanding techniques have been achieved. Therefore, in the future, the combination of spatial geometric structural information with semantic information to solve the registration problem of complex scenes in 3D point clouds should be considered.

Key words: remote sensing, three-dimensional point cloud, rigid registration, deep learning, point cloud feature, correspondences estimation, pose regression, scene flow estimation, registration datasets

Supported by National Natural Science Foundation of China (No. 42271457); Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (No. 22E03)