



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象学报

2022
02
VOL.27

ISSN1006-8961
CN11-3758/TB



三维视觉与智能图形



第27卷第2期（总第310期）
2022年2月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

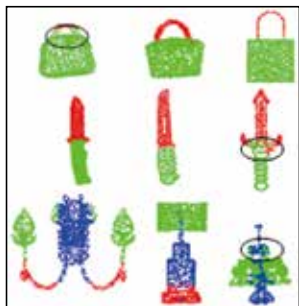
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



面向本征图像分解的高质量渲染数据集与非局部卷积网络(第0404页)



高效鲁棒三维结构化重建(第0421页)



利用隐式解码器的三维模型簇协同分割(第0550页)

三维视觉与智能图形专刊简介

陈宝权, 虞晶怡, 周昆, 郭裕兰, 黄惠, 刘利刚, 刘焯斌, 徐凯, 章国锋, 周晓巍 0289

综述

基于深度学习的单目深度估计技术综述

宋巍, 朱孟飞, 张明华, 赵丹枫, 贺琪 0292

深度学习刚性点云配准前沿进展

秦红星, 刘镇涛, 谭博元 0329

三维点云配准方法研究进展

李建微, 占家旺 0349

多源融合SLAM的现状与挑战

王金科, 左星星, 赵祥瑞, 吕佳俊, 刘勇 0368

深度学习单目深度估计研究进展

罗会兰, 周逸风 0390

数据集论文

面向本征图像分解的高质量渲染数据集与非局部卷积网络

王玉洁, 樊庆楠, 李坤, 陈冬冬, 杨敬钰, 卢健智, Dani Lischinski, 陈宝权 0404

深度估计与三维重建

高效鲁棒三维结构化重建

潘珊珊, 吕佳辉, 方昊, 黄惠 0421

结合LiDAR与RGB数据构建稠密深度图的多阶段指导网络

贾迪, 王子滔, 李宇扬, 金志杨, 刘泽洋, 吴思 0435

多尺度相似性迭代查找的可靠双目视差估计

晏敏, 王军政, 李静 0447

显微光学模糊图像的景物深度提取及应用

苗国超, 魏阳杰, 侯威翰 0461

融合注意力机制和多层U-Net的多视图立体重建

刘会杰, 柏正尧, 程威, 李俊杰, 许祝 0475

单目相机轨迹的真实尺度恢复

刘思博, 房立金 0486

三维形状分析

基于显著性图的点云替换对抗攻击

刘复昌, 南博, 缪永伟 0500

边收缩池化的网格变分自编码器

袁宇杰, 来煜坤, 杨洁, 段琦, 傅红波, 高林 0511

雅可比矩阵引导的无翻转体映射生成

徐茂峰, 刘利刚 0525

嵌入Transformer结构的多尺度点云补全

刘心溥, 马燕新, 许可, 万建伟, 郭裕兰 0538

三维点云分割

利用隐式解码器的三维模型簇协同分割

杨军, 张敏敏 0550

面向形状特征的多维度多层级点云分析

徐嘉利, 方志军, 伍世虞 0562

多特征融合与几何卷积的机载LiDAR点云地物分类

戴莫凡, 邢帅, 徐青, 李鹏程, 陈坤 0574

图像视频分析

聚合细粒度特征的深度注意力自动截图

方玉明, 钟裕, 鄢杰斌, 刘丽霞 0586

单幅人脸图像的全景纹理图生成方法

刘洋, 樊养余, 郭哲, 吕国云, 刘诗雅 0602

时空联合视差优化的立体视频重定向

金康俊, 柴雄力, 邵枫 0614

形状的全尺度可视化表示与识别

闵睿朋, 李一凡, 黄瑶, 杨剑宇, 钟宝江 0628

结合掩码定位和漏斗网络的6D姿态估计

李冬冬, 郑河荣, 刘复昌, 潘翔 0642

CONTENTS

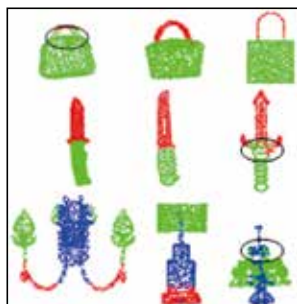
JOURNAL OF IMAGE AND GRAPHICS



High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition(P0404)



Efficient and robust 3D structure-aware reconstruction(P0421)



Co-segmentation of 3D shape clusters based on implicit decoder(P0550)

Review

A review of monocular depth estimation techniques based on deep learning	
Song Wei, Zhu Mengfei, Zhang Minghua, Zhao Danfeng, He Qi	0292
Review on deep learning rigid point cloud registration	
Qin Hongxing, Liu Zhenhao, Tan Boyuan	0329
Review on 3D point cloud registration method	
Li Jianwei, Zhan Jiawang	0349
Review of multi-source fusion SLAM: current status and challenges	
Wang Jinke, Zuo Xingxing, Zhao Xiangrui, Lyu Jiajun, Liu Yong	0368
Review of monocular depth estimation based on deep learning	
Luo Huilan, Zhou Yifeng	0390

Dataset

High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition	
Wang Yujie, Fan Qingnan, Li Kun, Chen Dongdong, Yang Jingyu, Lu Jianzhi, Dani Lischinski, Chen Baoquan	0404

Depth Estimation & 3D Reconstruction

Efficient and robust 3D structure-aware reconstruction	
Pan Shanshan, Lyu Jiahui, Fang Hao, Huang Hui	0421
Multi-stage guidance network for constructing dense depth map based on LiDAR and RGB data	
Jia Di, Wang Zitao, Li Yuyang, Jin Zhiyang, Liu Zeyang, Wu Si	0435
Reliable binocular disparity estimation based on multi-scale similarity recursive search	
Yan Min, Wang Junzheng, Li Jing	0447
Scene depth extraction and application of microscopic optical blur image	
Miao Guochao, Wei Yangjie, Hou Weihai	0461
Fusion attention mechanism and multilayer U-Net for multiview stereo	
Liu Huijie, Bai Zhengyao, Cheng Wei, Li Junjie, Xu Zhu	0475
Monocular camera trajectory recovery with real scale	
Liu Sibao, Fang Lijin	0486

3D Shape Analysis

Point cloud replacement adversarial attack based on saliency map	
Liu Fuchang, Nan Bo, Miao Yongwei	0500
Mesh variational auto-encoders with edge contraction pooling	
Yuan Yujie, Lai Yukun, Yang Jie, Duan Qi, Fu Hongbo, Gao Lin	0511
Generation of foldover-free volumetric mapping guided by Jacobian matrix	
Xu Maofeng, Liu Ligang	0525
Multi-scale Transformer based point cloud completion network	
Liu Xinpu, Ma Yanxin, Xu Ke, Wan Jianwei, Guo Yulan	0538

3D Point Cloud Segmentation

Co-segmentation of 3D shape clusters based on implicit decoder	
Yang Jun, Zhang Minmin	0550
Multi-dimensional multi-layer point cloud analysis for shape features	
Xu Jiali, Fang Zhijun, Wu Shiqian	0562
Semantic segmentation of airborne LiDAR point cloud based on multi-feature fusion and geometric convolution	
Dai Mofan, Xing Shuai, Xu Qing, Li Pengcheng, Chen Kun	0574

Image & Video Analysis

Deep attention guided image cropping with fine-grained feature aggregation	
Fang Yuming, Zhong Yu, Yan Jiebin, Liu Lixia	0586
Solo face image-based panoramic texture map generation	
Liu Yang, Fan Yangyu, Guo Zhe, Lyu Guoyun, Liu Shiya	0602
Optimizing spatiotemporal disparities for stereoscopic video retargeting	
Jin Kangjun, Chai Xiongli, Shao Feng	0614
Visualized all-scale shape representation and recognition	
Min Ruipeng, Li Yifan, Huang Yao, Yang Jianyu, Zhong Baojiang	0628
6D pose estimation based on mask location and hourglass network	
Li Dongdong, Zheng Herong, Liu Fuchang, Pan Xiang	0642

中图法分类号: TP391.7 文献标识码: A 文章编号: 1006-8961(2022)02-0329-20

论文引用格式: 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(02): 0329-0348 (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(02): 0329-0348) [DOI: 10.11834/jig.210556]

深度学习刚性点云配准前沿进展

秦红星^{1,2}, 刘镇涛^{1*}, 谭博元¹

1. 重庆邮电大学计算机科学与技术学院, 重庆 400065; 2. 重庆大学计算机学院, 重庆 400030

摘要: 随着3维采集设备的日渐推广,点云配准在越来越多的领域得到应用。然而,传统方法在低重叠、大量噪声、多异常点和大场景等方面表现不佳,这限制了点云配准在真实场景中的应用。面对传统方法的局限性,结合深度学习技术的点云配准方法开始出现,本文将这种方法称为深度点云配准,并对深度点云配准方法研究进展予以综述。首先,根据有无对应关系对目前的深度学习点云配准方法进行区分,分为无对应关系配准和基于对应关系的点云配准。针对基于对应关系的配准,根据各类方法的主要功能进行详细的分类与总结,其中包括几何特征提取、关键点检测、点对离群值去除、姿态估计和端到端配准,并重点介绍了最新出现的一些方法;针对无对应配准方法,详细介绍了各类方法的特点并对无对应与有对应方法的特点进行了总结。在性能评估中,首先对现有主要的评价指标进行了详细的分类与总结,给出其适用场景。对于真实数据集,给出了特征匹配、点对离群值去除的对比数据,并进行了总结。在合成数据集中,给出了相关方法在部分重叠、实时性和全局配准场景下的对比数据。最后讨论了当前深度点云配准面临的挑战并给出对未来研究方向的展望。

关键词: 点云配准;深度学习;无对应配准;端到端配准;对应关系;几何特征提取;离群值去除;综述

Review on deep learning rigid point cloud registration

Qin Hongxing^{1,2}, Liu Zhentao^{1*}, Tan Boyuan¹

1. School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. School of Computer Science, Chongqing University, Chongqing 400030, China

Abstract: A sharp increase in point cloud data past decade, which has facilitated to point cloud data processing algorithms. Point cloud registration is the process of converting point cloud data in two or more camera coordinate systems to the world coordinate system to complete the stitching process. In respect of 3D reconstruction, scanning equipment is used to obtain partial information of the scene in common, and the whole scene is reconstructed based on point cloud registration. In respect of high-precision map and positioning, the local point clouds fragments obtained in driving vehicle are registered to the scene map in advance to complete the high-precision positioning of the vehicle. In addition, point cloud registration is also widely used in pose estimation, robotics, medical and other fields. In the real-world point cloud data collection process, there are a lot of noise, abnormal points and low overlap, which brings great challenges to traditional methods. Currently, deep learning has been widely used in the field of point cloud registration and has achieved remarkable results.

收稿日期: 2021-07-05; 修回日期: 2021-10-01; 预印本日期: 2021-10-08

* 通信作者: 刘镇涛 liuzhentao1996@163.com

基金项目: 国家自然科学基金项目(61772097)

Supported by: National Natural Science Foundation of China(61772097)

In order to solve the limitations of traditional methods, some researchers have developed some point cloud registration methods integrated with deep learning technology, which is called deep point cloud registration. First of all, this analysis distinguishes the current deep learning point cloud registration methods according to the presence or absence of correspondence, which is divided into correspondence-free registration and point cloud registration based on correspondence. The main functions of various methods are classified as follows: 1) geometric feature extraction; 2) key point detection; 3) outlier removal; 4) pose estimation; and 5) end-to-end registration. The geometric feature extraction module aims to learn the coding method of the local geometric structure of the point cloud to generate discriminative features based on the network. Key point detection is used to detect points that are essential to the registration task in a large number of input points, and eliminate potential outliers while reducing computational complexity. Point-to-outliers are the final checking step before estimating the motion parameters to ensure the accuracy and efficiency of the solution. In the correspondence-free point cloud registration, a network structure similar to PointNet is used to obtain the global features of the perceived point cloud pose, and the rigid transformation parameters are estimated from the global features. In the performance of evaluation, the feature matching and registration error performance evaluation indicators are illustrated in detail. Feature matching performance metrics mainly include inlier ratio (IR) and feature matching recall (FMR). Registration error performance metrics include root mean square error (RMSE), mean square error (MSE), and mean Absolute error (MAE), relative translation error (RTE), relative rotation error (RRE), chamfer distance (CD) and registration recall (RR). RMSE, MSE and MAE are the most widely used metrics, but they have the disadvantage of Anisotropic. Isotropic RRE and RTE are indicators that actually measure the differences amongst the angle, the translation distance. The above five metrics all have unequal penalties for the registration of axisymmetric point clouds, and CD is the most fair metric. Meanwhile, real data sets registration tends to focus on the success rate of registration. With respect of real data sets, this research provides comparative data for feature matching and outlier removal. In the synthetic data set, this demonstration presents the comparative data of related methods in partial overlap, realtime, and global registration scenarios. At the end, the future research is from the current challenges in this field. 1) The application scenarios faced by point cloud registration are diverse, and it is difficult to develop general algorithms. Therefore, lightweight and efficient dedicated modules are more popular. 2) By dividing the overlap area, partial overlap can be converted into no overlap problem. This method is expected to lift the restrictions on the overlap rate requirements and fundamentally solve the problem of overlapping point cloud registration, so it has greater application value and prospects. 3) Most mainstream methods use multilayer perceptrons (MLPs) to learn saliency from data. 4) Some researchers introduced the random sample consensus (RANSAC) algorithm idea into the neural network, and achieved advanced results, but also led to higher complexity. Therefore, the balance performance and complexity is an issue to be considered in this sub-field. 5) The correspondence-free registration method is based on learning global features related to poses. The global features extracted by existing methods are more sensitive to noise and partial overlap, which is mainly caused by the fusion of some messy information in the global features. Meanwhile, the correspondence-free method has not been widely used in real data, and its robustness is still questioned by some researchers. Robust extraction of the global features for pose perception is also one of the main research issues further.

Key words: point cloud registration; deep learning; registration without corresponding; end-to-end registration; correspondence; geometric feature extraction; outliers removal; review

0 引言

2010年至2020年期间,3维扫描设备的普及呈现出加速趋势,造成点云数据的急剧增长,推动了深度学习3维视觉领域的迅速发展。学者们从不同角度对该领域的发展做出了总结。Guo等人(2021)针对深度学习在3D点云领域的应用给出了综述。

龙霄潇等人(2021)对3维视觉的前沿领域进行了系统的综述。点云配准是将两个或多个相机坐标系下的点云数据转换到世界坐标系完成拼接的过程,是3维视觉中的一项重要任务。比如:在3维重建中,通常利用扫描设备获取场景的局部信息,通过点云配准完成对整个场景的重建。在高精度地图与定位中,通过车辆行驶过程中采集到的局部点云片段配准到提前制作好的场景地图中,可以完成车辆的

高精度定位。此外,点云配准还在位姿估计、机器人和医疗等领域得到了大量的应用。

以ICP(iterative closest point)(Besl 和 McKay, 1992)、NDT(normal distributions transform)(Biber 和 Strasser, 2003)和4PCS(4-points congruent sets)(Aiger 等,2008)等为代表的传统方法已经在各个领域得到广泛应用,然而这些方法大多对噪声、异常点、低重叠和初始位姿敏感。另外,Johnson 和 Hebert(1999)、Frome 等人(2004)、Rusu 等人(2008,2009)、Salti 等人(2014)设计了人工编码特征用于全局配准,这些方法通常统计空间坐标、曲率、法向量等几何属性得到直方图,然后通过人工编码得到几何特征,这种方式容易受到噪声、异常点的影响,特征匹配效率不高。对于传统方法,Cheng 等人(2018)、Saiti 和 Theoharis(2020)给出了更为详尽的总结。在现实世界的点云数据采集过程中,存在着大量的噪声、异常点和较低的重叠,对传统方法带来了极大的挑战。近年来,深度学习在点云配准领域广泛应用并取得显著成效,引起了研究者广泛的兴趣。

本文旨在为深度学习刚性点云配准领域提供全面的综述。为方便起见,将基于深度学习的刚性点云配准称为深度点云配准。目前仅有少量综述性文章对深度学习点云配准研究进行了分析和总结。Huang 等人(2021b)将点云配准分为同源配准和跨源配准,对传统方法、基于深度学习的方法进行了综述。Zhang 等人(2020)对近年来深度学习方法进行了综述,但没有对算法进行明确的区分,并且缺乏完整的统一基准的对比参考数据。另外,Bello 等人(2020)对体素点云数据、原始点云上的学习进行了综述,但只涉及点云特征提取部分。与前人的工作相比,本文的主要贡献在于:聚焦于深度学习点云配准,综合最新出现的参考文献,从算法的主要功能出发,对算法重新进行了分类和总结,重点阐述了最新出现的方法和发展趋势;将配准过程划分为不同的阶段,对不同算法在相同阶段中的处理方式进行了详细阐述并使用表格进行归纳,然后总结其优劣;对以往综述工作做出补充,详细介绍了多种度量指标并总结同类指标的差异,汇总了在不同基准下的较为详细的对比数据,并提供分析和总结。

1 点云配准问题概述

给定两个点云:源点云 $X = \{x_i \in \mathbf{R}^3 \mid i = 1, \dots, N\}$ 和目标点云 $Y = \{y_j \in \mathbf{R}^3 \mid j = 1, \dots, M\}$ 。其中, N 与 M 分别为源点云和目标点云点的数量。点云配准的目标是求解源点云 X 到目标点云 Y 在世界坐标系下的相对变换,包括旋转矩阵 $R \in \mathbf{SO}(3)$ 和平移向量 $t \in \mathbf{R}^3$, 其中 $\mathbf{SO}(3)$ 为3维旋转群。点云配准可以描述为一个均方误差最小化问题

$$\operatorname{argmin}_{R, t} \frac{1}{N} \sum_{i=1}^N \|R x_i + t - y_m\|^2 \quad (1)$$

式中, y_m 表示任意点 $x_i \in X$ 在目标点云 Y 中的对应点(correspondence)。式(1)可以通过奇异值分解(singular value decomposition, SVD)求解 R 和 t , 然而对应点 y_m 通常是未知的,在求解变换之前,需要建立逐点的对应关系

$$m = \operatorname{argmin}_{j \in \{1, \dots, M\}} \|R x_i + t - y_j\| \quad (2)$$

因此,式(1)与式(2)是一个典型的鸡和蛋的问题,式(2)的求解依赖于已知的 R 和 t , 而这恰好是式(1)求解的目标。在传统方法中,典型的ICP算法假设初始 $R = I$ 并且 $t = 0$, 首先在欧氏空间中使用最近邻操作来建立对应关系 m , 然后利用式(1)求解刚性变换,循环以上两个过程直到收敛。这样的解决方案通常会导致算法对初始位姿敏感。

为了解除初始位姿的限制,一些研究者开发了全局配准方法,通过设计人工编码的特征用于描述点的局部几何结构,然后利用真实点对中,两点的邻域具有几何相似性的特点构建点与点的对应关系。现有深度学习方法大多沿用了该思路,其关键问题在于如何利用网络从数据中学习几何特征编码函数 $f: \mathbf{R}^3 \rightarrow \mathbf{R}^C$, 使得 $\|f(x) - f(y)\| \approx 0$, 其中 C 表示几何特征的维度,点 y 表示点 x 的真实对应点。

随着深度学习在点云配准领域的应用,出现了一批不需要依靠对应关系的方法,本文将其称为无对应配准方法。无对应配准方法的关键问题在于如何利用网络学习一个从点云到全局特征的映射 $\varphi: \mathbf{R}^{3 \times N} \rightarrow \mathbf{R}^K$, 并且寻找合适的旋转 R 和平移 t 使得 $\varphi(Y) = \varphi(RX + t)$, 其中 K 表示全局特征的维度。

点云配准是一个典型的流水线处理过程,其流程如图1所示。在图1中,预处理过程通常用于对原始点云数据进行降噪、去异常值、去除非重叠区域

和采样处理;另外,基于对应关系的点云配准在特征提取后还需要进行额外的特征匹配步骤来获得点对。

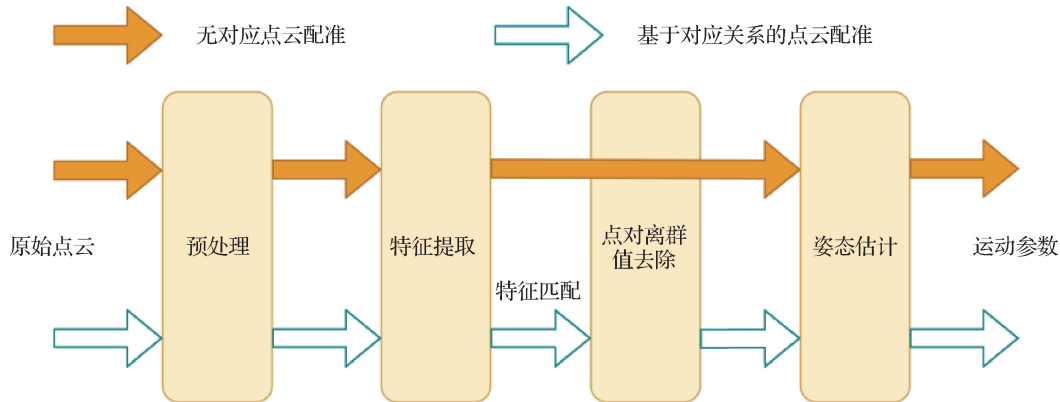


图1 深度点云配准流水线过程

Fig. 1 Pipeline process of deep point cloud registration

根据有无借助对应关系,本文将现有研究分为基于对应关系的深度点云配准和无对应配准分别进行介绍。目前该领域对跨源点云的研究较少,因此不再对其进行划分。

2 基于对应关系的深度点云配准

目前研究者在合成数据与真实数据上都展开了研究,相比合成数据而言,真实数据的表面点拓扑结构更加复杂,并且点的数量更多,对鲁棒性的要求更高,许多研究者将真实数据上的点云配准拆分为多个子问题进行研究。而在合成数据中,现有的研究大多采用了端到端方法。为了方便对比,本文按照各个方法的主要功能,即特征提取、关键点检测、离群点对去除、姿态估计和端到端点云配准进行分类。相关代表方法整理如表1所示。

2.1 特征提取

在基于对应关系的点云配准中,学习鉴别力高的特征表示是配准好坏的关键。点云中包含丰富的空间几何信息,以合理的方式组织点云,从点云中提取更多的信息,才能丰富特征的辨识度。因此,如何从无组织的点云中提取更多的具有辨识度的信息是特征提取最受关注的问题。近年来,研究者们提出大量基于深度学习的特征提取方法,按照它们的特点,分为以下两种类型:1)基于局部块(local patch)

的特征提取;2)基于卷积的特征提取。

Qi 等人(2017a)提出了 PointNet,这是第1个直接在输入点云上提取特征的网络。PointNet 主要解决了点云的无序性、置换不变性和旋转不变性问题。对于输入的无序点云,使用多层感知机(multilayer perceptrons, MLPs)分别对每个点提取特征,然后使用对称函数(最大池化)来达到置换不变的目的,最后使用 T-net(transformation network)预测一个刚性变换矩阵,用于满足旋转不变性。

PointNet 无法捕获空间点的语义信息,限制了特征的通用性。此外,由 MLPs 学习而来的网络无法处理密度不均匀的点云数据。Qi 等人(2017b)进一步提出了 PointNet++,针对以上问题进行了改进。PointNet++ 引入了一个由最远点采样层、分组层以及 PointNet 层组成的层次化结构来捕捉不同尺度的上下文信息。为了提高在不同密度下采样的性能,提出了多尺度聚合和多分辨率聚合,用于提取多个尺度的局部模式,并根据点密度进行自适应组合,最终得到密度自适应特征。

PointNet 与 PointNet++ 没有关注点与点之间的几何关系,限制了特征的表示能力,因而不能直接应用于点云配准领域。但其为解决点云特征提取中存在的无序性、密度变化、置换不变性和旋转不变性等问题提供了有效参考,促进了后续研究的开展。

表 1 基于对应关系配准的典型方法

Table 1 Typical methods of registration based on correspondence

阶段	分类	参考文献
几何特征提取	基于局部块的特征提取	Qi 等人(2017a,b)、Khoury 等人(2017)、Zeng 等人(2017)、Deng 等人(2018b)、Deng 等人(2018a)、Yew 和 Lee(2018)、Gojcic 等人(2019)、Ao 等人(2021)
	基于卷积的特征提取	Wang 等人(2019)、Choy 等人(2019b)、Horache 等人(2021)
关键点提取	—	Lu 等人(2020)、Bai 等人(2020)、Huang 等人(2021a)
点对离群值去除	—	Pais 等人(2020)、Yang 等人(2020)、Bai 等人(2021)
端到端配准	递归配准	Wang 和 Solomon(2019b)、Yew 和 Lee(2020)、Li 等人(2020)
	非递归配准	Lu 等人(2019)、Wang 和 Solomon(2019a)、Yuan 等人(2020)、Choy 等人(2020)、Ginzburg 和 Raviv(2021)

注:“—”表示对该阶段无分类。

2.1.1 基于局部块的特征提取

为了解决点云的无组织性问题,一些研究者利用体素、包围球和 K 最近邻(K-nearest neighbor, KNN)等方式对点云进行组织,以提取更为丰富的局部几何特征,这些方法通常称为基于局部块的方法。

Khoury 等人(2017)提出一种紧凑几何特征(compact geometric features, CGF),通过在点上建立一个包围球,建立局部参考框架(local reference frame, LRF)得到描述局部邻域中点分布的直方图,然后训练一个神经网络将输入的直方图映射到一个低维的欧氏空间得到更紧凑的几何特征描述符。

Deng 等人(2018b)提出了 PPFNet。首先对采样点的局部邻域计算点对特征(point pair feature, PPF)(Rusu 等,2008,2009),将其作为网络的输入,利用多个 PointNet 网络将不同尺度的局部特征和全局特征融合,最后经过 MLPs 编码得到最终的特征。PPFNet 利用了全局上下文感知和编码特征,提高了特征的旋转不变性和对噪声的鲁棒性,但是 PPF 特征的计算需要大量的最近邻标注数据,并且局部参考框架的建立依赖于法向量的估计,导致了噪声敏感。另外,寻找固定的 K 近邻点导致对点云稀疏程度的变化敏感。

为改进 PPFNet 的缺点,Deng(2018a)进一步提出了 PPF-FoldNet。该方法是一种无监督的方法,首先提取 PPF 特征,使用了一个包含具有跳连接的类 PointNet 编码器和类 FoldingNet(Yang 等,2018)的解码器的框架,通过度量输入 PPF 特征和输出点对特征之间的差异,来获取最终点云的特征表示。

PPF-FoldNet 在噪声下取得了较好的效果,但仍然对点密度变化较为敏感。

Yew 和 Lee(2018)提出了 3DFeatNet,针对网络训练中难以获取精确标注的问题,提出一种弱监督深度学习框架。从两个输入点云的所有描述符对之间的相似性度量中选择正样本以及置信度最高的负样本,然后使用注意力层来学习一个权重,用于衡量每个输入描述符对三元组损失的贡献,最后通过最小化三元组损失来训练网络。

使用类似 PointNet 的结构可以实现直接从原始点云中提取特征,但这也限制了卷积操作的使用,不利于捕获局部拓扑信息。针对这个问题,Gojcic 等人(2019)提出了 3DSmoothNet,设计了可以适用于卷积操作的平滑密度值(smoothed density value voxelization, SDV)体素格子,来编码原始点云,利用 Siamese 网络架构学习最终的特征。该方法增强了泛化能力,在单一的场景下进行训练,再扩展到其他场景中时仍然可以取得较好的结果。

2.1.2 基于卷积的特征提取

目前基于卷积的方法主要包括两种,一种是直接在点云上进行卷积操作来提取特征,另一种是通过将点云体素化后,使用体素卷积方法提取特征。

受 PointNet 和卷积操作的启发,Wang 等人(2019)设计了边卷积(edge convolution, EdgeConv)操作。首先在点云中构建局部邻域图,然后在中心点与相邻点构成的边上应用类卷积操作,称为边卷积。进一步,在网络中多次叠加包含边卷积的层,得到动态图卷积神经网络(dynamic graph convolutional

neural network, DGCNN)。与普通的图卷积神经网络不同,DGCNN 中的图节点不是固定的,而是在网络的每一层之后动态更新,也就是说,一个点的近邻点集随着网络的加深逐层变化。因此,EdgeConv 非常灵活,在网络中加入一定的 EdgeConv 层有利于捕获点之间的拓扑信息,但这样也引入了额外的 K 近邻计算开销。

为了解决不规则点云上的卷积问题,并更好地捕获局部几何信息,Thomas 等人(2019)提出了核点卷积(kernel point convolution, KPConv),使用携带卷积权值的核点(kernel point)来模拟 2 维卷积中的核像素,进而定义原始点云上的卷积操作。KPConv 通过点云位置与核点的关系来生成卷积核,其组合权重矩阵是人为设定的,不一定能得到最优结果,且灵活性有限。同时,针对不同的点云数据,核点空间需要专门进行定制,这使得其对超参数敏感。为此,Xu 等人(2021b)提出了位置自适应卷积(position adaptive convolution, PAConv),通过网络从点的位置中自适应地学习权重矩阵,然后通过动态组合权重矩阵来构造卷积核。在现有的点云处理框架中,可以使用 PAConv 替换 MLPs 模块,无需改变框架的结构和参数,因此具备更好的灵活性。

Zeng 等人(2017)提出了 3DMatch 网络,以体素为输入,首先需要把点云量化成体素表示,然后利用 3D 卷积神经网络来学习局部几何模式,得到 512 维

度的特征描述符(descriptor)。

由于点云的稀疏性,传统 3D 卷积操作会造成计算资源的浪费,并且 3D 卷积本身计算的时间复杂度也很高,因此 Choy 等人(2019b)提出全卷积几何特征(fully convolutional geometric features, FCGF),使用稀疏 3D 卷积代替传统的 3D 卷积,以缓解点云稀疏性带来的问题。通过使用稀疏卷积构建包含跳连接和残差块 ResBlock 的 ResUNet 网络架构来提取点云的局部几何特征,其网络结构如图 2 所示,图中卷积块参数分别为卷积核大小、步长和通道数。FCGF 特征输出的维度仅为 32 维,较为紧凑,并且其运行效率相较以往有着巨大的提高,可以拓展到真实场景。另外,FCGF 需要通过旋转增强来实现特征的旋转不变性。

FCGF 对数据的采样出现过拟合,导致其泛化效果较差。Horache 等人(2021)在 FCGF 的基础上提出了多尺度架构与自监督细化(multi-scale architecture and self-supervised fine-tuning, MS-SVConv)卷积神经网络。该方法首先体素化点云,通过选取不同大小的体素格子来获得不同尺度下的体素数据,然后将它们输入到一个共享权重的 U-Net 结构的网络中,分别提取特征。最后使用一个全连接层对这些特征进行融合,最终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。

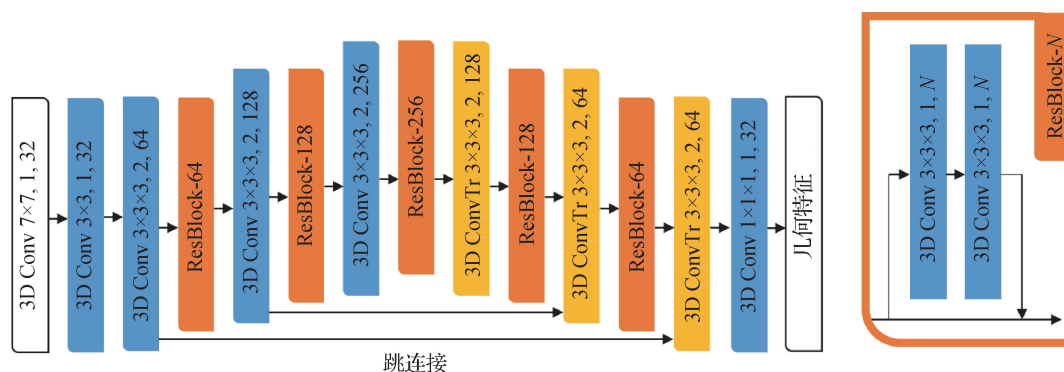


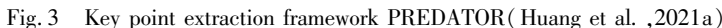
图 2 特征提取框架 FCGF(Choy 等,2019b)

Fig. 2 Feature extraction framework FCGF(Choy et al.,2019b)

Ao 等人(2021)提出了一种新的神经网络结构 SpinNet。SpinNet 包含两个模块:1)空间点变换模块。首先估计出一个参考轴,使其与全局参考系的 Z 轴对齐,以消除这个方向上的自由度;然后利用球形体素化和 XY 平面变换操作消除 XY 平面的旋转

自由度;最后将球形体素投影到一个圆柱形容器中。2)特征提取模块。使用一个共享权重的 MLPs 和最大池化聚合函数提取每个体素的初始特征,然后使用 3D 圆柱卷积(Joung 等,2020)操作提取最终的特征。SpinNet 保证了特征的旋转不变性,具有较好的

Huang 等人(2021a)对以往的关键点提取方法进行了补充,指出提取关键点的前提是必须保证关键点应该在两个点云的重叠区域内。基于此观点将问题延伸到低重叠场景,提出 PREDATOR 用于低重叠的点云配准,其网络框架如图 3 所示。PREDATOR 通过 PConv 和 EdgeConv 分别捕获密度不变的局部几何信息与上下文信息进而得到超点(super point),通过使用交叉注意力块(Sarlin 等,2020)提取两点云的特征编码之间的上下文信息得到上下文特征,来预测超点位于重叠部分的概率。然后通过共享权重的解码器对超点进行上采样,最终输出点的特征和点位于重叠区域的概率以及点的显著性打分。值得注意的是,与以往的显著性定义不同,PREDATOR 将显著性定义为某个点能找到另一与之匹配的点的可能性。



另外一些端到端配准方法中也嵌入了关键点检测模块。比如: PRNet (partial-to-partial registration network) (Wang 和 Solomon, 2019b) 通过特征的二范数距离定义点的显著性, 然后根据二范数的大小来选取显著性高的若干个关键点。为了提高计算效率, 避免随机采样导致的网络性能下降, IDAM (iterative distance-aware similarity matrix) (Li 等, 2020) 使用 MLPs 预测点的显著性来获取固定数量的关键点。

关键点的显著性通常与特征进行联合学习, 由于显著性往往来自于独特的特征, 这种显著性选取方式依赖于特征学习模块的编码, 由网络自适应选取对于配准任务更为重要的点。另外, 由于配准过程往往更加关注于重叠区域, 适当地交换两个点云之间的信息可以使得关键点的检测更加精准。在一些端到端网络中也嵌入了关键点模块, 对关键点检测技术的总结如表 2 所示。

表 2 关键点检测方法概要总结
Table 2 Summary of key point detection methods

模型	方法概要
DeepVCP (Lu 等, 2019)	使用 MLPs 预测点的显著性
PRNet (Wang 和 Solomon, 2019b)	将关键点定义为特征二范数最大的点, 由网络自学习特征的编码
RSKDD-Net (Lu 等, 2020)	随机采样后, 使用随机拓展群策略和注意力机制获取关键点
D3feat (Bai 等, 2020)	结合显著性打分与特征通道最大打分得到点的最终显著性打分
IDAM (Li 等, 2020)	使用 MLPs 预测点的显著性
PREDATOR (Huang 等, 2021a)	使用交叉注意力提取两点云的共同上下文, 预测位于重叠区域内的关键点

注: DeepVCP (deep virtual corresponding point) 是深度虚拟对应点。

2.3 点对离群值去除

点对离群值去除 (outliers removal) 是点云配准最为关键的任务之一。事实上, 通过几何特征匹配得到的对应关系并不总是可靠的。造成这种情况的原因主要有 3 个: 1) 噪声和异常点。噪声和异常点会降低特征的辨识度。2) 部分重叠问题。位于重叠区域之外的点显然没有对应点与之匹配。3) 局部点云不显著问题。某些点云局部区域非常“平坦”, 显著性较低, 容易导致特征的误匹配。通常, 特征匹配后得到的点对包含较多离群值, 很难用于直接求解刚性变换, 因此需要对其进行去除。

在 3DRegNet (3D point registration network) (Pais 等, 2020) 中, 使用一个包含 ResNet 块的深度神经网络作为分类模块, 用以预测某一个点对属于内点的置信度。3D 点对之间存在丰富的几何信息, 单纯地将点对离群值去除视为一个二分类模块并不能取得很好的效果。Yang 等人 (2020) 观察到正确的点对之间应该满足几何上的兼容性, 因此提出使用兼容性特征 (compatibility features, CF) 来表示点对。首先, 针对点对的长度与角度进行兼容性检查, 得到点对的兼容性分数。然后, 聚合兼容性打分最高的点的信息, 得到兼容性特征。最后, 将提取的兼

容性特征输入到一个 MLPs 中, 进行密集的二分类, 将点对区分为内点 (inlier) 和离群点 (outlier)。与 Yang 等人 (2020) 的工作相似, PointDSC (point deep spatial consistency) (Bai 等, 2021) 中将传统方法的空间几何一致性约束引入了深度学习领域, 通过类似于 Nonlocal 网络 (Wang 等, 2018) 的 SCNonlocal 模块提取空间一致性特征。不同的是, PointDSC 使用 MLPs 输出的置信度选择种子点, 对每个种子点进行 K 近邻查找以提取满足空间一致性的点对的集合, 最后从若干集合中选择最优的集合求解刚性变换, 其网络框架如图 4 所示。

3DRegNet (Pais 等, 2020)、DGR (deep global registration) (Choy 等, 2020) 方法将点对离群值去除考虑为二分类问题, 利用网络预测点对的置信度。然而点对中可能包含非常多的离群值, 这对网络的拟合效果造成影响, 因此这类方法并没有达到理想中的性能, 相比随机采样一致性 (random sample consensus, RANSAC) (Fischler 和 Bolles 等, 1981) 仍然存在一定差距。CF 与 PointDSC 的工作本质上都是利用网络学习点对之间的旋转不变约束, 因此对离群值的筛选变得更为严格, 更加贴合真实数据配准需求。此外, PointDSC 还通过类似于空间播种的方

法选择多个一致性集合进一步加强鲁棒性,相比直接对点对进行分类的回归方法而言,离群值去除的性能

有了较大的提高。在一些端到端网络中也嵌入了点对离群值去除模块,对上述方法的总结如表 3 所示。

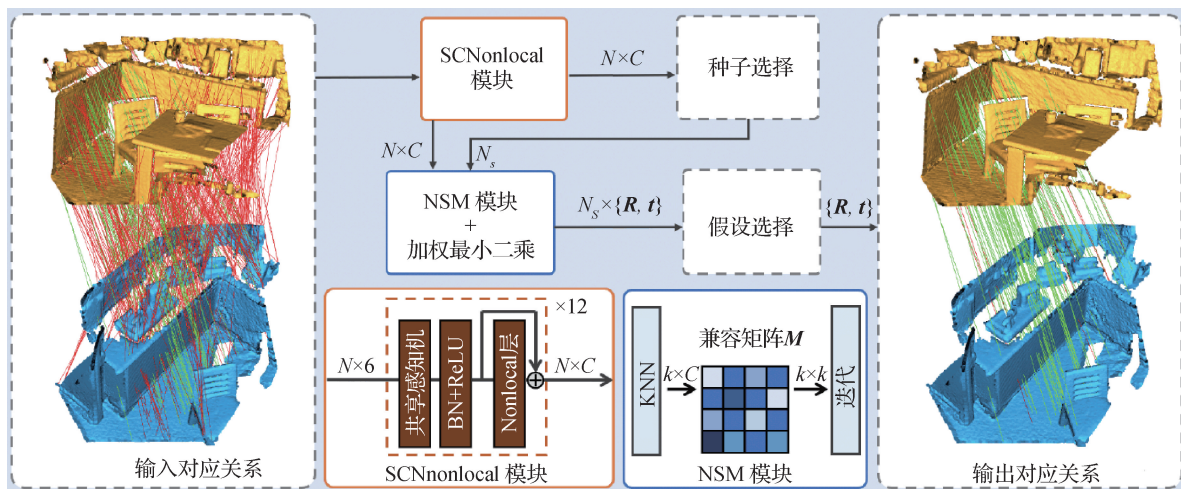


图 4 点对离群值去除框架 PointDSC (Bai 等, 2021)

Fig. 4 Outlier removal framework PointDSC (Bai et al., 2021)

表 3 点对离群值去除方法概要总结

Table 3 Summary of outliers removal methods

模型	方法概要
RPMNet (Yew 和 Lee, 2020)	引入 sinkhorn 层, 通过迭代归一化得到双随机矩阵
3DRegNet (Pais 等, 2020)	使用二分类网络预测点对权重
DGR (Choy 等, 2020)	使用二分类网络预测点对权重
CF (Yang 等, 2020)	提取几何兼容性特征表示点对, 然后使用 MLPs 进行二分类
IDAM (Li 等, 2020)	计算混合消除权重用于预测点对的置信度
PointDSC (Bai 等, 2021)	使用空间一致性特征表示点对, 使用空间播种方法划分若干个一致性集合

2.4 姿态估计

在基于对应关系的点云配准中, 姿态估计即通过对对应关系计算源点云与目标点云之间的刚性变换 R 和 t , 是配准的最后阶段。SVD 方法计算的是解析解, 并且提供了可微分的实现 (Papadopoulos 和 Lourakis, 2000), 因此在深度点云配准中得到广泛应用。大部分方法采用了常规的 SVD 求解方法, 但在对 SVD 的使用上有所不同, 因此本文仅从使用的角度对各个方法进行总结与讨论。

DCP (deep closest point) (Wang 和 Solomon, 2019a)、DeepVCP (Lu 等, 2019) 通过置信度的加权 and 计算对应点的位置, 进而估计相对姿态。PRNet (Wang 和 Solomon, 2019b) 选取置信度较高的点对, 用于 SVD 的计算。RPMNet (robust point matching network) (Yew 和 Lee, 2020)、IDAM (Li 等, 2020)、DGR (Choy 等, 2020) 通

过置信度选择点对, 但同时保留了置信度, 将其作为权重, 使用加权 SVD 求解刚性变换。PointDSC (Bai 等, 2021) 在网络的训练阶段使用加权 SVD 方法, 在测试阶段使用加权最小二乘法估计求解刚性变换。

仅对 SVD 求解模块输入点对坐标意味着丢弃网络中计算的点对置信度, 因此网络的反向传播需要通过坐标值进行传递, 而坐标的值通常是确定的数值, 因此可能对网络中梯度的传递造成阻碍, 而保留权重则意味着网络可以通过权重值来传播梯度。使用加权和计算的对应点并不真实存在于目标点云中, 因此更加依赖于网络对置信度的估计。PointDSC 在测试阶段使用传统优化技术对加权 SVD 进行替代, 并且取得了成功, 提供了新的思路。

2.5 端到端点云配准方法

点云数据配准结果主要取决于对应关系的估

计,好的对应关系可以通过高效的特征提取、匹配与离群值去除模块来获取,一些研究者对这些流程进行了整合,通过训练一个深度神经网络来直接解决配准问题,输入两个点云,输出运动参数,这种网络称为端到端网络(end-to-end)。通常,端到端网络将各个流程的处理模块集成到一个网络中,对显存的需求较大,更适合数据量较小的配准任务。另外,一些网络并不是直接输入原始点云,而是输入处理后的点云相关数据,为叙述方便,本文统一将其归类为端到端网络。

Wang 和 Solomon (2019a) 提出了深度最近点(DCP)方法,用于实现端到端的点云配准。DCP 使用由 Wang 等人(2019)提出的 DGCNN 学习点云的嵌入(embedding),使用 transformer (Vaswani 等, 2017)对上下文信息进行编码,实现对嵌入的微调。为了解决特征匹配的不可微分问题,使用嵌入的点积来度量相似度。DCP 对噪声取得了较好的鲁棒性,但由于其采用单随机矩阵确定对应关系,这意味着点云 X 中的所有点在目标点云中都有一个对应,这种假设在异常值和低重叠情况下会引入错误的对应关系。

Wang 和 Solomon(2019b)进一步提出了 PRNet,用来改进 DCP。与 DCP 考虑所有输入点的对应不同,PRNet 采用部分关键点进行配准。此外,为了获取高质量的对应关系,PRNet 使用了一种演员—评论家(actor-critic)模式,使用全局池化聚合逐点特征来得到全局特征,然后通过一个子网络预测退火参数用于控制匹配的锐化程度。在匹配阶段,使用了近似可微的 Gumbel-softmax (Jang 等, 2017) 函数代替不可微 argmax 函数,确保反向传播时可以获得梯度。最后,使用迭代优化提高配准的精度。尽管在低重叠和噪声场景下 PRNet 取得了更好的效果,但是 PRNet 网络模型较为庞大,不利于现实场景应用。

Yew 和 Lee(2020)基于鲁棒点匹配算法(robust point matching, RPM) (Gold 等, 1998) 提出 RPMNet。在特征提取部分,采用与 PPFNet 相似的结构,同时在特征中加入点的 3 维坐标得到混合特征;使用参数预测网络来估计离群值参数与退火参数,然后结合混合特征计算点对匹配。在点对离群值去除阶段,引入一个可微的 sinkhorn 层,对匹配矩阵进行拓展,然后进行迭代归一化得到置信度更高的对应关

系。RPMNet 在噪声和部分重叠下取得了优秀的性能,但模型的初始输入需要额外的标签数据,因此不适用于密集的点云输入。与 PRNet 类似, RPMNet 需要在迭代中重复计算特征,这增加了计算成本。

由于在网络模型中 K 近邻的计算效率不高,大多数模型通常使用特征的内积或者特征的二范数距离来表示两点之间的匹配相似度,这种方式忽略了特征在某个通道上的差异。为了改善这种情况, IDAM (Li 等, 2020) 使用距离感知相似矩阵卷积学习特征匹配的相似性度量,以获取更好的匹配关系。该网络支持使用快速点特征直方图(fast point feature histograms, FPFH) 或图卷积网络(graph neural network, GNN)来提取旋转不变的几何特征,并将其与一个简单的点对 4 维欧氏特征进行拼接得到距离增广特征张量,然后在该张量上应用 1 维卷积来学习特征匹配度量。为了提高相似矩阵卷积的运算效率和匹配的准确率,使用两阶段的点消除策略分别输出逐点的显著性打分和预测点对的置信度,进而计算每个点对的权重。IDAM 在密集输入点的场景下有显著的运算优势,但网络仅仅依靠关键点的选择来避免离群点对,因此其配准结果依赖于网络对关键点的预测。

深度高斯混合模型配准(deep Gaussian mixture model registration, DeepGMR) (Yuan 等, 2020) 将神经网络嵌入到极大似然(maximum likelihood estimate, MLE) 框架,将目标点云用高斯混合模型(Gaussian mixed model, GMM) 进行建模,利用网络预测任意两个点对的对应概率,然后利用可微分模块计算 GMM 参数,然后从 GMM 参数中估计最优变换。DeepGMR 利用神经网络代替期望最大化算法中的 E 步(expectation-step),对初始位置和噪声不敏感,并且不需要迭代优化,是一种全局的配准方法。

Ginzburg 和 Raviv (2021) 提出深度加权一致性(deep weighted consensus, DWC) 用于全局配准。DWC 是一种无监督网络,首先在点的局部邻域内提取线性变换不变特征(rotation-invariant, RI),然后使用两个 DGCNN (Wang 等, 2019) 分别提取 RI 的全局和局部特征,通过全局—局部融合网络得到最终特征。然后使用余弦相似性来定义两个点云概率匹配矩阵,采样包含 K 个对应关系的集合并分别计算刚性变换,然后选取满足内点数量最多的刚性变换,

最后通过采用加权一致性损失优化对应关系。

Lu 等人 (2019) 提出了深度虚拟对应点方法 (deep virtual corresponding point, DeepVCP) 以便在 LiDAR (light detection and ranging) 点云配准 (真实数据) 中能够避免动态对象, 并采样有利于配准的稳定的、独特的特征, 采用 PointNet++ 提取点的语义特征。在关键点提取阶段, 受 3DFeatNet 启发, 设计了 1 个包含 3 个叠加完全连接层的多层感知器和一个 Topk 操作组成的点加权层, 用于进一步提取关键点。在匹配阶段, 为了在目标点云中找到对应点进行最终配准, 使用了一个 mini-PointNet 提取逐个关键点的几何特征。然后通过候选点坐标与特征相似性的加权和来计算虚拟对应点, 进而估计刚性变换。最后在损失中加入全局几何约束以保证网络预测的点的一致性。

深度全局配准 (deep global registration, DGR) (Choy 等, 2020) 是用于真实数据的端到端配准网络。在特征提取阶段, DGR 采用 FCGF (Choy 等, 2019b)。在离群值去除阶段, 采用了类似 3DRegNet 的思路, 但有所不同的是, DGR 利用 Minkowski 引擎 (Choy 等, 2019a) 实现 6D 卷积操作, 进而构建一个 Res-UNet 网络用于预测逐个点对的置信度。为了进一步提高算法的鲁棒性, 在得到初步参数后, 通过能量最小化函数对位姿进行微调。

一些方法为了提高配准精度, 采用了递归网络的设计, 例如: PRNet (Wang 和 Solomon, 2019b)、PRMNet (Yew 和 Lee, 2020)、IDAM (Li 等, 2020), 但这也带来重复计算特征的开销, 本质上是因为特征对于旋转的变化不够鲁棒; 同时, 递归的方式也导致了网络在训练过程中的不稳定, 这可能是特征匹配模块在不同迭代过程中网络所关注到的信息不同所

导致的。另外, 由于配准过程中存在大量的干扰因素, 例如: 异常点、部分重叠等, 这些因素导致了一些点不可能存在一一对应关系, 因此考虑所有点的对应关系必然会导致算法应用的场景受到限制。

2.6 小结

在基于对应关系的深度点云配准中, 核心任务是获取有效的对应关系。现有方法基本上都是通过特征提取和匹配来获取对应关系。通常经过特征提取和匹配阶段的对应关系并不一定是可靠的, 其中包含大量离群点对, 不能直接用于估计刚性变换, 因此需要去除错误匹配的点对。关键点检测通常用于采样对于配准任务有效的点, 这潜在地去除了一部分离群值的影响, 但对于大规模离群值去除, 专用的点对离群值去除模块能发挥更强大的功效。另外, 对于不同的干扰而言: 1) 点云中的噪声、异常值主要影响的是特征的辨识度, 直接的结果是导致误匹配。2) 对于存在部分重叠的情况, 主要的解决方案包括两种, 一种是“锐化”对应关系, 即选择置信度最高的一部分点; 另一种是在特征匹配前去除非重叠区域的点以彻底排除非重叠点的影响。3) 对于大旋转场景下, 要求特征对旋转的变化具有较好鲁棒性。

3 无对应配准

3.1 无对应配准方法

主流的深度点云配准方法使用对应关系求解刚性变换, 研究者们探索出了一条不依赖于对应关系的道路。相关方法概要整理如表 4 所示。

借鉴 2 维图像中的算法思路, Aoki 等人 (2019) 提出 PointNetLK, 利用去除了 T-net 模块的 PointNet

表 4 无对应配准方法概要总结

Table 4 Summary of correspondence-free registration methods

模型	采用方法	主要贡献
PointNetLK (Aoki 等, 2019)	在网络中嵌入优化方法	首个提出无对应配准的算法
Deng 等人 (2019)	回归	利用 PPF-FoldNet 从不同输入产生旋转感知特征
FMR (Huang 等, 2020)	在网络中嵌入优化方法	使用编码器对全局特征进行监督
PCRNet (Sarode 等, 2019)	回归	使用 Saimese 架构预测运动参数
AlignNet-3D (Groß 等, 2019)	回归	将运动参数回归考虑为分类任务
OMNet (Xu 等, 2021a)	回归	通过迭代预测重叠区域以降低对全局特征的影响

注: FMR (feature-metric registration) 是特征度量配准。

网络从两个点云 X 和 Y 中提取相对位姿信息。然后利用逆合成 (inverse compositional, IC) 公式计算目标点云全局特征的雅可比矩阵。最后,利用一

个可微的 Lucas & Kanade (LK) 算法优化全局特征之间的差异计算刚性变换,其网络框架如图 5 所示。

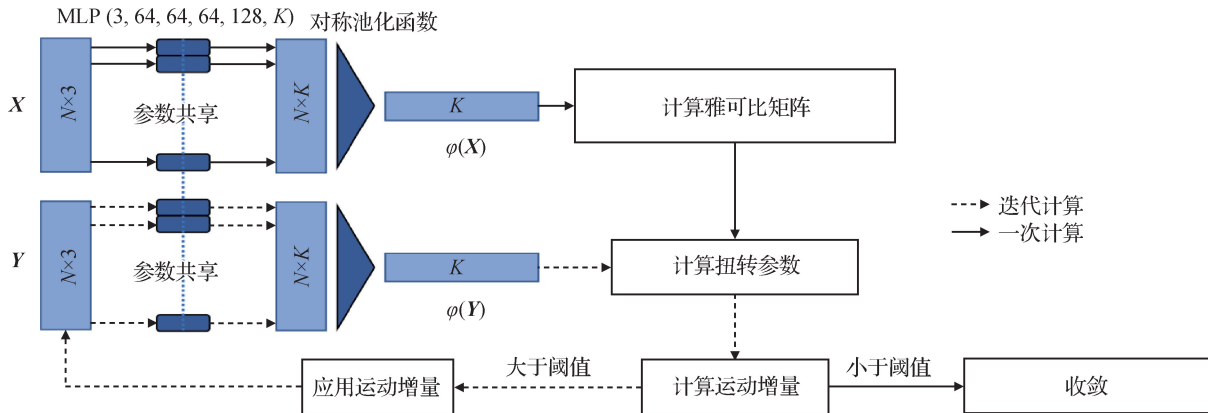


图 5 无对应点云配准 PointNetLK (Aoki 等, 2019)

Fig. 5 Point cloud registration without corresponding PointNetLK (Aoki et al., 2019)

Deng 等人 (2019) 提出一种数据驱动的点云配准网络。利用不同输入但网络结构相同的 PPF-FoldNet (输入为 PPF 姿态不变特征) 和 PC-FoldNet (输入为点云) 输出特征的差异产生包含结构和姿态信息的新特征。在此基础上, 设计了 RelativeNet 直接预测相对姿态。该方法在真实数据集中得到了较高的召回率, 但配准精度较低。

PCRNet (Sarode 等, 2019) 提出一种数据驱动的方法。与 PointNetLK 不同, PCRNet (point cloud registration network) 将两个点云的全局特征进行拼接, 直接利用类似 Siamese (Held 等, 2016) 的网络架构预测运动参数。Groß 等人 (2019) 设计了一种网络 AlignNet-3D 用于 3 维轨道状态估计。该方法将旋转角度划分为若干个子区间, 由网络预测旋转所属区间, 最后预测一个角度差得到最终的旋转角度参数。特征度量配准 (FMR) (Huang 等, 2020) 沿用了 PointNetLK 的思路, 但不同的是, FMR 利用刚性变换可逆的特性, 使用编码器—解码器的模型对全局特征进行间接监督。编码器模块生成独特的特征后, 使用解码器模块将特征映射回 3 维点云。这种方法可以通过监督或无监督的方式训练网络。FMR 在真实数据集中取得了较高的精度, 同时噪声、密度和低重叠具有一定鲁棒性。此外, FMR 在跨源点云配准中也取得了最先进的结果。

Xu 等人 (2021a) 提出了 OMNet (overlapping mask network), 为了避免非重叠点的负面影响, 在每

次迭代中分别预测两个源点云和目标点云的重叠掩码, 对非重叠区域进行过滤, 然后再通过 MLPs 从两个点云的全局特征中预测相对运动参数。OMNet 通过移除重叠区域来避免全局特征受到干扰, 取得了目前最先进的结果。

目前无对应方法通常都采用了与 PointNet 类似的网络结构用于提取感知空间位姿的全局特征, 在姿态估计阶段, PointNetLK (Aoki 等, 2019)、FMR (Huang 等, 2020) 使用传统优化方法, 从特征中计算雅可比矩阵, 进而估计运动参数, 造成了较大的计算开销。其他的无对应方法使用了回归的方式预测运动参数, 这样的好处在于避免计算雅可比矩阵的同时, 通过网络学习从全局特征到运动参数的映射, 从而使得全局特征更好地感知空间位姿的变化, 但是这样的方式更加依赖于网络对训练数据的学习。

3.2 无对应方法与有对应方法的比较

相比于基于对应关系的方法, 无对应方法直接关注于两个点云的相对位姿信息而不是局部几何信息, 避免了计算对应关系的开销, 降低了后续处理的难度, 这直接导致了两者在参数估计上的差别。在有对应方法中, SVD 是更有效、流行的求解方式; 而无对应方法中, 使用回归的方式更为有利。这主要是因为表示相对位姿信息的全局特征具有抽象性, 通过网络的自适应学习更加合理。而采用 SVD 方法本质上是对对应关系进行监督。在配准的鲁棒性方面, 无对应配准方法更容易受到非重叠区域的干

扰,主要的原因在于缺乏明确的监督机制来保证全局特征与相对位姿信息唯一相关。对于合成数据的配准,无对应方法中的 OMNet (Xu 等, 2021a) 通过优化网络结构、引入重叠区域分割技术增强了全局特征的抗干扰能力,进而增强了在未知类别模型 (unseen-categories) 数据上的泛化性能,取得了先进的结果,展示了在泛化上的潜力。另外,目前无对应方法很少应用于真实场景的配准,尽管无对应方法能够避免计算对应关系带来的一些问题,但其更可能受到复杂场景数据的影响,其跨数据集的泛化仍然有待研究。

4 性能评估

4.1 公开数据集

本文总结了 3 维点云配准领域的通用数据集。用于配准任务的 3 维点云数据集可以分为两类:合成数据集和真实场景数据集。真实场景数据集通常由激光雷达或 RGBD 相机等设备获得,包含户外、室内场景。合成数据通常包含人工合成的各类物体的 3 维模型。Zhang 等人 (2020) 已经对常用的数据集进行了详细的总结,本文仅对目前最具代表性的数据集进行介绍。

1) ModelNet40 (Wu 等, 2015)。由 12 311 个不包含噪声和异常点的 CAD (computer aided design) 网格模型组成,总共包含 40 个常见类别,该数据集主要用于点云分类和检索,同时也是近年来深度学习点云配准领域中主要的合成数据集。获取地址: <http://modelnet.cs.princeton.edu/>。

2) 3DMatch (Zeng 等, 2017)。该数据集由 SUN3D (Xiao 等, 2013)、7-Scene (Shotton 等, 2013) 等 3 维重建数据集中拆分而来,包括 62 个不同室内场景的 RGB-D 帧序列。获取地址: <http://3dmatch.cs.princeton.edu>。

4.2 性能评价指标

为了评估深度学习点云配准算法的性能,通常需要借助通用的、客观的评价指标进行度量。深度学习点云配准的实验性能评价标准从处理阶段来看,主要分为:特征匹配度量和配准误差度量。本小节对常用的度量指标进行了详细的阐述,分别给出其适用的场景并比较了其差别。

4.2.1 特征匹配度量

特征匹配度量主要用于直接衡量特征匹配模块

性能的好坏,其主要包含以下两个指标:

1) 内点比例 (inlier ratio, IR)。该值表示特征匹配后有效的对应关系占有对应关系的比例。对于一组假定的对应关系 $(p, q) \in M_{i,j}$, 内点比例为

$$R_1 = \frac{1}{|M_{i,j}|} \sum_{(p,q) \in M_{i,j}} 1[\|R^*p + t^* - q\| < \tau_1] \quad (3)$$

式中, $1[\]$ 表示 Iverson 括号,括号内为真取 1,否则取 0。 R^* 与 t^* 表示真实的旋转、平移标签, τ_1 表示最小距离误差阈值。

2) 特征匹配召回率 (feature match recall, FMR)。其表示配准任务中成功置信度高的任务占总配准任务的比例,计算为

$$R_2 = \frac{1}{K} \sum_{i=1}^K 1[R_{i1} > \tau_2] \quad (4)$$

式中, K 表示数据集中用于配准的点云对数量, R_{i1} 表示第 i 个点云对中的内点比例, τ_2 表示内点比例的最小阈值。

IR 与 FMR 的主要区别在于, IR 用于直接度量两个点云的特征匹配的性能,而 FMR 是用于整个数据集中的配准好坏的预示性度量 (假设当 IR 超过一定阈值后就可以完成配准)。

4.2.2 配准误差度量

1) 均方根误差 (root mean squared error, RMSE)、均方误差 (mean squared error, MSE) 与平均绝对误差 (mean absolute deviation, MAE) 分别记为 E_1 、 E_2 、 E_3 , 即

$$E_1 = \sqrt{\frac{1}{|M_{i,j}^*|} \sum_{(p,q) \in M_{i,j}^*} \|Rp + t - q\|^2} \quad (5)$$

$$E_2 = \frac{1}{|M_{i,j}^*|} \sum_{(p,q) \in M_{i,j}^*} \|Rp + t - q\|^2 \quad (6)$$

$$E_3 = \frac{1}{|M_{i,j}^*|} \sum_{(p,q) \in M_{i,j}^*} |Rp + t - q| \quad (7)$$

式中, R 与 t 表示估计的刚性变换,对应关系 $(p, q) \in M_{i,j}^*$ 表示真实的对应关系。通过引入真实标签,以上所有度量也可以用于单独度量旋转和平移单项性能。

2) 相对平移误差 (relative translation error, RTE) 和相对旋转误差 (relative rotation error, RRE) 用于度量平移、旋转的估计值与真实标签的差值,其单位分别为厘米 (cm)、度 ($^\circ$), 计算公式分别为

$$E_4 = \|t - t^*\|_2 \quad (8)$$

$$E_5 = \arccos\left(\frac{\text{tr}(\mathbf{R}^T \mathbf{R}^*) - 1}{2}\right) \quad (9)$$

式中, $\mathbf{R}^*$ 与 $\mathbf{t}^*$ 分别表示真实的旋转、平移标签, $\text{tr}()$ 表示矩阵的迹。

3) 倒角距离 (chamfer distance, CD) 用于公平地度量存在轴对称场景下的配准精度, 即

$$D_1 = \frac{1}{|\mathbf{X}|} \sum_{\mathbf{x} \in \mathbf{X}} \min_{\mathbf{y} \in \mathbf{Y}_c} \|\mathbf{x} - \mathbf{y}\|^2 + \frac{1}{|\mathbf{Y}|} \sum_{\mathbf{y} \in \mathbf{Y}} \min_{\mathbf{x} \in \mathbf{X}_c} \|\mathbf{x} - \mathbf{y}\|^2 \quad (10)$$

式中, $\mathbf{X}_c$ 与 $\mathbf{Y}_c$ 分别表示初始的源、目标点云, $\mathbf{X}$ 、 $\mathbf{Y}$ 表示将经过变换后的源、目标点云。该公式由 Yew 和 Lee (2020) 提出。

4) 配准召回率 (registration recall, RR) 表示整个数据集中配准误差小于一定阈值的点云对的比例, 即

$$R_3 = \frac{1}{C} \sum_{i=1}^C 1[E_1 < \tau_3] \quad (11)$$

式中, C 为数据集中用于配准的点云对数量, E_1 表示 RMSE 误差, τ_3 表示点云对的 RMSE 误差阈值。

RMSE、MSE 和 MAE 是使用最为广泛的度量指标, 但具有各向异性 (anisotropic) 的缺点。RRE 与

RTE 是实际度量角度与平移距离差值的指标, 是各向同性的 (isotropic)。以上 5 种度量都对轴对称点云的配准存在不公平的惩罚, 而 CD 是最公正的度量。此外, 在真实数据集中的配准, 往往更加关注配准的成功率, 因此 RR 在真实数据配准中使用更为广泛。

4.3 合成点云数据配准性能评估

在合成点云数据上, 需要人工处理原始数据集, 但不同的处理方式都会对算法造成影响。此外, 一些算法只针对某个特点的场景而设计, 比如低重叠、大旋转等。因此, 本小节对比了各个场景中主流相关方法的性能。

在部分重叠场景下, 性能对比如表 5 所示。其中, 无对应方法 FMR (Huang 等, 2020)、OMNet (Xu 等, 2021a) 在性能上相较于 PointNetLK (Aoki 等, 2019) 出现了较大的提升, 这得益于网络结构的优化。另外, OMNet 引入了重叠区域分割技术, 相较于 FMR 效果更好。在基于对应关系的方法中, DeepGMR (Yuan 等, 2020) 考虑了所有输入点的对应关系, 因此无法适应部分重叠场景的配准任务。

表 5 ModelNet40 基准下的部分重叠点云配准

Table 5 Partial-to-partial registration on ModelNet40 benchmark

模型	RMSE($\mathbf{R}$)	MAE($\mathbf{R}$)	RMSE($\mathbf{t}$)	MAE($\mathbf{t}$)	迭代次数
PointNetLK (Aoki 等, 2019)	27.589	16.047	0.151 6	0.095 5	10
DCP (Wang 和 Solomon, 2019a)	7.353	4.923	0.065 7	0.045 1	1
PRNet (Wang 和 Solomon, 2019b)	3.241	1.632	0.018 1	0.012 7	3
FMR (Huang 等, 2020)	4.819	2.488	0.034 5	0.015 8	10
DeepGMR (Yuan 等, 2020)	19.720	10.227	0.091 5	0.059 6	1
IDAM (Li 等, 2020)	5.188	3.114	0.037 7	0.020 8	3
OMNet (Xu 等, 2021a)	2.203	0.910	0.015 5	0.007 8	4

注: 实验大部分数据来自 Xu 等人 (2021a) 的工作; 加粗字体为每列最优值; RMSE($\mathbf{R}$)、RMSE($\mathbf{t}$) 表示使用均方根误差度量旋转矩阵 $\mathbf{R}$ 和平移向量 $\mathbf{t}$; MAE($\mathbf{R}$)、MAE($\mathbf{t}$) 表示使用平均绝对误差度量旋转矩阵 $\mathbf{R}$ 和平移向量 $\mathbf{t}$ 。

在不同采样点下的执行速度如表 6 所示。DCP (Wang 和 Solomon, 2019a) 与 DeepGMR 没有使用迭代的方式, 因此其执行速度相比于其他方法更快。PointNetLK 与 FMR 需要重复计算全局特征和雅可比矩阵, 而 OMNet 采用了回归的方式预测相对姿态, 避免使用复杂度较高的逆合成方法计算雅可比矩阵, 从而显著提高了算法的运行速度。相比于 RPM-Net (Yew 和 Lee, 2020)、PRNet (Wang 和 Solomon,

2019b)、IDAM (Li 等, 2020) 在每次迭代中重复使用了特征提取模块提取到的特征, 避免了重复计算特征的开销, 并且使用关键点技术去除了一部分点对, 因此运行速度更快。尽管 PRNet 也采用了关键点技术, 但因其网络过于庞大, 执行速度仍然较慢。另外, RPM-Net 迭代地归一化对应关系矩阵的行和列, 这种方法在点数量较多的情况下, 计算时间出现了大幅增加。

全局配准性能对比如表 7 所示。全局配准是指

表6 ModelNet40 基准下各算法在不同采样点的执行速度

Table 6 The execution speed of each algorithm at different sampling points on ModelNet40 Benchmark

模型	采样点数				备注
	512	1 024	2 048	4 096	
PointNetLK(Aoki 等,2019)	73	77	83	89	递归
DCP(Wang 和 Solomon,2019a)	15	17	26	88	非递归
PRNet(Wang 和 Solomon,2019b)	79	84	114	—	递归
FMR(Huang 等,2020)	138	158	295	764	递归
RPMNet(Yew 和 Lee,2020)	58	115	271	726	递归
IDAM(Li 等,2020)	27	28	33	62	递归
DeepGMR(Yuan 等,2020)	9	9	9	10	非递归
OMNet(Xu 等,2021a)	24	25	27	32	递归

注:实验设置以及数据来自 Xu 等人(2021a)的工作,执行速度以 ms 为单位,“—”表示无数据,加粗字体为每列最优值。

表7 ModelNet40 基准下全局配准召回率

Table 7 Global registration recall on ModelNet40 benchmark

模型	RR		
	原始测试集	添加噪声	未知类别测试集
DCP(Wang 和 Solomon,2019a)	0.99	0.94	0.54
PointNetLK(Aoki 等,2019)	0.44	0.38	0.13
PRNet(Wang 和 Solomon,2019b)	0.64	0.57	0.30
DeepGMR(Yuan 等,2020)	1.00	0.99	0.99

注:数据来自 Yuan 等人(2020)的工作,加粗字体为每列最优值。

任意的初始位姿下的配准,DeepGMR 和 DCP 在这种实验设置下取得了更为优秀的性能,但 DCP 泛化效果较差,这可能是因为其特征提取模块的泛化性能较差。

综上所述,随着深度点云配准领域的发展,各方法分别在对初始位姿、部分重叠等干扰上的鲁棒性得到了显著的提升,但很多算法仍然存在一定的局限性,例如:DeepGMR 在全局配准中取得了优秀的结果,但其无法配准部分重叠的点云数据;此外,在合成数据集上的不同处理方法产生的数据很难为研究者提供统一、公正的对比,针对不同任务的统一的基准亟需研究人员的关注。

4.4 真实点云数据配准性能评估

目前,在真实数据中的端到端方法较少,许多工作在特征提取、离群值去除两个主要的阶段上建立了比较基准。特征提取的工作相对较多,离群值去除在近两年得到了发展。

表8展示了在3DMatch比较基准上主流特征提

取算法的性能。从时间尺度上看,近年来特征提取算法的研究呈现出了高速发展的趋势,已经在匹配回调率、旋转不变性、紧凑性和泛化能力上都有着巨大的提升。3DMatch(Zeng 等,2017)率先开发了在体素上提取点云特征的方法。CGF(Khoury 等,2017)、PPFNet(Deng 等,2018b)和 PPF-FoldNet(Deng 等,2018a)依赖于对手工特征进行二次编码,限制了网络的学习能力,性能较差。近年来,随着卷积网络的应用,特征辨识度和旋转鲁棒性得到了很大的改善。FCGF(Choy 等,2019b)依赖于固定大小的体素进行稀疏体素卷积,导致泛化性能较差。Ms-SVConv(Horache 等,2021)在 FCGF 的基础上,通过多尺度的稀疏体素卷积大幅提高了泛化能力,达到了目前先进的效果。而 SpinNet(Ao 等,2021)将输入点云转化为圆柱体素,在不需旋转增强的情况下取得了较好的泛化效果。SpinNet 的成功在一定程度上说明:相比于纯粹地依赖于网络的学习,通过合适的引导和监督,可以让网络学习到更加普适的描述点特征的方式。

表 8 3DMatch 基准下的特征匹配
Table 8 Feature match results on 3DMatch benchmark

模型	FMR/%			时间/s	特征维度	旋转增强
	原始测试集	旋转后的数据	在 ETH 上泛化			
3DMatch(Zeng 等,2017)	59.6	1.1	16.9	3.21	512	否
CGF(Khoury 等,2017)	58.2	58.5	20.2	1.463	32	否
PPFNet(Deng 等,2018b)	62.3	0.3	—	2.257	64	否
PPF-FoldNet(Deng 等,2018a)	71.8	73.1	—	0.794	512	否
FCGF(Choy 等,2019b)	95.2	95.3	16.1	0.019	32	是
3DSmoothNet(Gojcic 等,2019)	94.7	94.9	79.0	5.515	32	否
D3Feat-rand(Bai 等,2020)	95.3	95.2	26.2	—	32	是
D3Feat-pred(Bai 等,2020)	95.8	95.5	61.6	—	32	是
SpinNet(Ao 等,2021)	97.6	97.5	92.8	—	32	否
Ms-SVConv_3(Horache 等,2021)	98.4	—	98.9	—	32	是

注：“—”表示未获得相关数据；旋转后的数据表示在原始数据应用旋转群 $SO(3)$ 组上的任意旋转。旋转增强表示对训练数据使用随机旋转进行处理。

表 9 展示了在 3DMatch 比较基准上点对离群值去除性能。端到端的 DGR(Choy 等,2020)使用离群值去除网络模块并未取得理想的效果,其效果仍然依赖 RANSAC(Fischler 和 Bolles,1981)算法作为保护措施(safeguard)。3DRegNet(Pais 等,2020)采用了与 DGR 一样的特征提取模块 FCGF 和二分类网络,尽管运行速度更快,但是效果较

差,这与 3DRegNet 采用单独训练的方式和网络构成有关。最近提出的 PointDSC(Bai 等,2021)取得了优于 RANSAC 的效果,展示了神经网络在离群值去除上的潜力,促进端到端网络在真实数据点云配准上的发展。总的来说,基于深度学习的真实点云数据配准研究仍然有着巨大的发展空间。

表 9 3DMatch 基准上的离群值去除
Table 9 Outliers removal on 3DMatch benchmark

模型	RR/%	RE/(°)	TE/cm	IR/%	时间/s
FCGF + RANSAC-100k(Fischler 和 Bolles,1981)	91.5	2.49	7.54	85.30	5.50
DGR(Choy 等,2020)	86.5	2.33	7.36	78.94	0.56
DGR + safeguard(Choy 等,2020)	91.3	2.40	7.48	78.94	1.36
FCGF + 3DRegNet(Pais 等,2020)	77.76	2.74	8.13	56.28	0.05
FCGF + PointDSC(Bai 等,2021)	93.28	2.06	6.55	86.54	0.09

注:该数据来自 Bai 等人(2021)的工作,表中所有模型均以 FCGF 作为默认特征提取器,加粗字体为每列最优值。

5 结 语

点云配准是诸多计算机视觉应用的重要组成部分。本文对深度点云配准领域的研究进行了综述,首先根据现有深度点云配准方法的特点将其划分为两大类,在不同类别下,进行了详细的分组阐述和对

比总结。然后列举了主要方法在不同测试基准上的性能。目前,尽管深度点云配准技术取得了巨大的进步,但是仍然难以满足现实需求,主要的困难表现在鲁棒性和泛化两个方向,导致这些问题的因素有很多,比如:大量噪声和异常值、低重叠、不同的初始位姿、对称性、大尺度场景下的内存负担和计算开销等。针对这些问题,研究者开发了各种各样的算法,

然而,这些算法大多具有局限性。最后,本文从目前点云配准面临的挑战这一点出发,对未来的研究趋势进行展望。

1)在不同的应用场景中,算法面临的挑战不同,这对点云配准算法的通用性提出了要求。然而从目前研究阶段来看,开发通用的算法是困难的。并且,基于对应关系的深度点云配准是一个流水线式的处理流程,在计算运动参数之前,通常需要经过多个模块的处理。因此,开发轻量、高效的专用模块是更受欢迎的研究热点。

2)现实中的传感器获取点云通常由于视角的限制而得到部分重叠的点云数据,直接对这些数据进行配准通常是困难的,尽管目前一部分研究者开发了能够在部分重叠下配准的网络,但通常对重叠率有一定的要求。目前出现了一个更先进的思路:一些研究人员开发了用于分割重叠区域的网络(Sarode等,2020;Huang等,2021a;Xu等,2021a),将部分重叠问题转化为完全重叠问题。这种方法有望解除对重叠率要求的限制,从而在根本上解决部分重叠点云配准的问题,因此具有较大的应用价值和前景空间。

3)真实场景中会获得海量的点,通常使用降采样对点云进行处理,然而降采样可能会导致局部几何信息的丢失,不利于局部几何特征的提取。关键点技术能有效弥补这个缺点,通过寻找独特的、对配准任务有效的少量点用于下游任务,可以显著降低内存负担并减少计算开销,对于在真实应用场景中的配准具有重大意义。目前,关键点检测技术在点云配准领域中得到了广泛的应用,但得到的关注却较少。因为显著性的定义并不明确,现阶段主流方法大多采用MLPs从数据中学习显著性。因此,开发更高效、明确的关键点检测方法仍然是现阶段亟待解决的问题。

4)基于回归的离群值去除方法在真实数据点云配准中未能取得理想的效果,目前仍然依赖传统的RANSAC(Fischler和Bolles,1981)算法,然而该算法具有随机性,且迭代次数随离群值数目的增加出现指数级增长。一些研究者将RANSAC算法的思想引入到神经网络中,取得了更优的效果,这表明神经网络与传统技术的结合具有较大的潜力,通常来说,传统方法具有透明的特点,而神经网络则有强大的拟合能力,如何将两者的优势进行结合是目前

研究的热点。

5)无对应配准方法依赖于学习和位姿相关的全局特征,然而神经网络学习出来的全局特征是非常抽象的,很难准确地施加约束。现有方法提取出的全局特征对噪声和部分重叠比较敏感,这主要是全局特征中融合了一些杂乱的信息导致的。另外,无对应方法尚未广泛应用于真实数据,其鲁棒性仍然受到一些研究者的质疑。综合来说,如何鲁棒地提取位姿感知的全局特征也是未来的主要研究方向之一。

参考文献 (References)

- Aiger D, Mitra N J and Cohen-Or D. 2008. 4-points congruent sets for robust pairwise surface registration. *ACM Transactions on Graphics*, 27(3): 1-10 [DOI: 10.1145/1360612.1360684]
- 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]
- 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, USA: IEEE: 7156-7165 [DOI: 10.1109/cvpr.2019.00733]
- Bai X Y, Luo Z X, Zhou L, Chen H K, Li L, Hu Z Y, Fu H B and Tai C L. 2021. PointDSC: robust point cloud registration using deep spatial consistency// *Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 15854-15864 [DOI: 10.1109/CVPR46437.2021.01560]
- 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]
- Bello S A, Yu S S, Wang C, Adam J M and Li J. 2020. Review: deep learning on 3D point clouds. *Remote Sensing*, 12(11): #1729 [DOI: 10.3390/rs12111729]
- Besl P J and McKay N D. 1992. Method for registration of 3-D shapes// *Proceedings of SPIE 1611, Sensor Fusion IV: Control Paradigms and Data Structures*. Boston, USA: SPIE: 586-606 [DOI: 10.1117/12.57955]
- Biber P and Strasser W. 2003. The normal distributions transform: a new approach to laser scan matching// *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*. Las Vegas, USA: IEEE: 2743-2748 [DOI: 10.1109/iroso.

2003. 1249285]
- 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]
- Choy C, Dong W and Koltun V. 2020. Deep global registration//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA; IEEE: 2511-2520 [DOI: 10.1109/CVPR42600.2020.00259]
- 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 (CVPR)*. 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]
- Deng H W, Birdal T and Ilıc S. 2018a. PPF-FoldNet: unsupervised learning of rotation invariant 3D local descriptors//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich, Germany; Springer: 620-638 [DOI: 10.1007/978-3-030-01228-1_37]
- Deng H W, Birdal T and Ilıc 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, USA; IEEE: 195-205 [DOI: 10.1109/CVPR.2018.00028]
- Deng H W, Birdal T and Ilıc S. 2019. 3D local features for direct pairwise registration//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA; IEEE: 3239-3248 [DOI: 10.1109/CVPR.2019.00336]
- 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, Czech Republic; Springer: 224-237 [DOI: 10.1007/978-3-540-24672-5_18]
- Ginzburg D and Raviv D. 2021. Deep weighted consensus (DWC): dense correspondence confidence maps for 3D shape registration [EB/OL]. [2021-05-06]. <https://arxiv.org/pdf/2105.02714.pdf>
- 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]
- Gold S, Rangarajan A, Lu C P, Pappu S and Mjolsness E. 1998. New algorithms for 2D and 3D point matching: pose estimation and correspondence. *Pattern Recognition*, 31(8): 1019-1031 [DOI: 10.1016/S0031-3203(98)80010-1]
- Groß J, Ošep A and Leibe B. 2019. AlignNet-3D: fast point cloud registration of partially observed objects//*Proceedings of 2019 International Conference on 3D Vision (3DV)*. Quebec City, Canada; IEEE: 623-632 [DOI: 10.1109/3DV.2019.00074]
- Guo Y L, Wang H Y, Hu Q Y, Liu H, Liu L and Bennamoun M. 2021. Deep learning for 3D point clouds: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12): 4338-4364 [DOI: 10.1109/TPAMI.2020.3005434]
- Held D, Thrun S and Savarese S. 2016. Learning to track at 100 FPS with deep regression networks//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, the Netherlands; Springer: 749-765 [DOI: 10.1007/978-3-319-46448-0_45]
- Horache S, Deschaud J E and Goulette F. 2021. 3D point cloud registration with multi-scale architecture and unsupervised transfer learning [EB/OL]. [2021-03-26]. <https://arxiv.org/pdf/2103.14533.pdf>
- 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 (CVPR)*. Nashville, USA; 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, USA; 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 [EB/OL]. [2021-03-03]. <https://arxiv.org/pdf/2103.02690.pdf>
- Jang E, Gu S X and Poole B. 2017. Categorical reparameterization with gumbel-softmax [EB/OL]. [2021-03-03]. <https://arxiv.org/pdf/1611.01144.pdf>
- 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]
- Joung S, Kim S, Kim H, Kim M, Kim I J, Cho J and Sohn K. 2020. Cylindrical convolutional networks for joint object detection and viewpoint estimation//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA; IEEE: 14151-14160 [DOI: 10.1109/CVPR42600.2020.01417]
- Khoury M, Zhou Q Y and Koltun V. 2017. Learning compact geometric features//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy; IEEE: 153-161 [DOI: 10.1109/ICCV.2017.26]

- Li J H, Zhang C H, Xu Z Y, Zhou H N and Zhang C. 2020. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration//Proceedings of the 16th European Conference on Computer Vision. Glasgow, United Kingdom: Springer; 378-394 [DOI: 10.1007/978-3-030-58586-0_23]
- 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. 三维视觉前沿进展. *中国图象图形学报*, 26(6): 1389-1428) [DOI: 10.11834/jig.210043]
- Lu F, Chen G, Liu Y L, Qu Z N and Knoll A. 2020. RSKDD-Net: random sample-based keypoint detector and descriptor//Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS 2020). Thiele, Lothar: Curran Associates Inc.; 21297-21308 [DOI: 10.1146/annurev-nips-2020-07-0113]
- 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 Conference on Computer Vision. Seoul, Korea(South); IEEE; 12-21 [DOI: 10.1109/ICCV.2019.00010]
- Pais G D, Ramalingam S, Govindu V M, Nascimento J C, Chellappa R and Miraldo P. 2020. 3DRegNet: a deep neural network for 3D point registration//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE; 7191-7201 [DOI: 10.1109/CVPR42600.2020.00722]
- Papadopoulos T and Lourakis M I A. 2000. Estimating the jacobian of the singular value decomposition: theory and applications//Proceedings of the 6th European Conference on Computer Vision. Ireland: Springer; 554-570 [DOI: 10.1007/3-540-45054-8_36]
- 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
- 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 C 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]
- Saiti E and Theoharis T. 2020. An application independent review of multimodal 3D registration methods. *Computers and Graphics*, 91: 153-178 [DOI: 10.1016/j.cag.2020.07.012]
- 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]
- Sarode V, Dhagat A, Srivatsan R A, Zavallos N, Lucey S and Choset H. 2020. MaskNet: a fully-convolutional network to estimate inlier points//Proceedings of 2020 International Conference on 3D Vision (3DV). Fukuoka, Japan; IEEE; 1029-1038 [DOI: 10.1109/3DV50981.2020.00113]
- Sarode V, Li X Q, Goforth H, Aoki Y, Srivatsan R A, Lucey S and Choset H. 2019. PCRNet: point cloud registration network using PointNet encoding [EB/OL]. [2021-03-03]. <https://arxiv.org/pdf/1908.07906.pdf>
- Shotton J, Glocker B, Zach C, Izadi S, Criminisi A and Fitzgibbon A. 2013. Scene coordinate regression forests for camera relocation in RGB-D images//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA; IEEE; 2930-2937 [DOI: 10.1109/CVPR.2013.377]
- 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]
- 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
- Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE; 7794-7803 [DOI: 10.1109/CVPR.2018.00813]
- Wang Y and Solomon J M. 2019a. Deep closest point: learning representations for point cloud registration//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE; 3522-3531 [DOI: 10.1109/ICCV.2019.00362]
- Wang Y and Solomon J M. 2019b. PRNet: self-supervised learning for partial-to-partial registration//Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver, Canada; Curran Associates Inc.; 8814-8826

- 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]
- 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]
- Xiao J X, Owens A and Torralba A. 2013. SUN3D: a database of big spaces reconstructed using SfM and object labels//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia; IEEE: 1625-1632 [DOI: 10.1109/ICCV.2013.458]
- Xu H, Liu S C, Wang G F, Liu G H and Zeng B. 2021a. OMNet: learning overlapping mask for partial-to-partial point cloud registration [EB/OL]. [2021-03-03]. <https://arxiv.org/pdf/2103.00937.pdf>
- Xu M T, Ding R Y, Zhao H S and Qi X J. 2021b. PAConv: position adaptive convolution with dynamic kernel assembling on point clouds//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA; IEEE: 3172-3181 [DOI: 10.1109/CVPR46437.2021.00319]
- Yang J Q, Chen J H, Huang Z Q, Quan S W, Zhang Y N and Cao Z G. 2020. 3D correspondence grouping with compatibility features [EB/OL]. [2021-03-03]. <https://arxiv.org/pdf/2007.10570.pdf>
- Yang Y Q, Feng C, Shen Y R and Tian D. 2018. FoldingNet: point cloud auto-encoder via deep grid deformation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 206-215 [DOI: 10.1109/cvpr.2018.00029]
- Yew Z J and Lee G H. 2018. 3DFeat-Net: weakly supervised local 3D features for point cloud registration//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: 630-646 [DOI: 10.1007/978-3-030-01267-0_37]
- 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, USA; IEEE: 11824-11833 [DOI: 10.1109/CVPR42600.2020.01184]
- 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, United Kingdom; 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, USA; IEEE: 199-208 [DOI: 10.1109/CVPR.2017.29]
- Zhang Z Y, Dai Y C and Sun J D. 2020. 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]

作者简介



秦红星, 1977年生, 男, 教授, 主要研究方向为计算机图形学、计算机视觉、可视化。

E-mail: qin_hx@foxmail.com



刘镇涛, 通信作者, 男, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: liuzhentao1996@163.com

谭博元, 男, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: tanboyuan20@163.com