

单位代码: 10293 密 级:

南京邮电大学

硕 士 学 位 论 文



论文题目: 基于深度学习的高精度
 点云配准算法研究

学	号	<u>1019020913</u>
姓	名	<u>梅青</u>
导	师	<u>肖建</u>
学 科 专 业		<u>电路与系统</u>
研 究 方 向		<u>智能信息处理</u>
申请学位类别		<u>工学硕士</u>
论文提交日期		<u>2022.04</u>

Research of High-precision Point Cloud Registration Algorithm Based on Deep Learning

Thesis Submitted to Nanjing University of Posts and
Telecommunications for the Degree of
Master of Science in Engineering



By

Qing Mei

Supervisor: Prof. Jian Xiao

April 2022

南京邮电大学学位论文原创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得南京邮电大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

本人学位论文及涉及相关资料若有不实，愿意承担一切相关的法律责任。

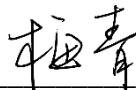
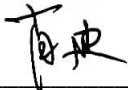
研究生学号：1019020913 研究生签名：  日期：2022.04.12

南京邮电大学学位论文使用授权声明

本人承诺所呈交的学位论文不涉及任何国家秘密，本人及导师为本论文的涉密责任并列第一责任人。

本人授权南京邮电大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档；允许论文被查阅和借阅；可以将学位论文的全部或部分内容编入有关数据库进行检索；可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文。本文电子文档的内容和纸质论文的内容相一致。论文的公布（包括刊登）授权南京邮电大学研究生院办理。

非国家秘密类涉密学位论文在解密后适用本授权书。

研究生签名：  导师签名：  日期：2022.04.12

摘要

点云配准是三维计算机视觉和图形学中一个重要的研究课题，在三维重建、三维数据融合、即时定位与地图构建（Simultaneous Localization and Mapping, SLAM）等领域中都有重要的应用。随着自动驾驶、人工智能、虚拟现实技术等领域的兴起，将深度学习技术应用于点云配准也取得了很多极具启发意义的成果。但外部传感器采集到的点云数据往往存在冗余性与质量缺陷，这一方面使得配准过程占用大量的内存资源和处理时间，另一方面噪声干扰也会导致特征匹配产生的对应关系出现异常值，从而影响点云配准的精度。在深度学习领域中，以点云配准经典网络 PCRNet（Point Cloud Registration Network）为代表的点云配准方法在特征提取时只关注点云的全局信息，并没有考虑到点云的局部特征，难以表征点云空间复杂的变化关系，一定程度上限制了其对场景的理解能力，从而降低点云配准的精度。同时，PCRNet 网络在特征融合模块仅仅以数组拼接的方式将两组点云的特征在维度上进行连接，不能充分利用不同层次特征的互补性，无法很好地融合点云的局部特征和全局信息，从而破坏点云配准的精度。本文针对上述点云配准任务中存在的问题，主要对以下三个方面进行研究与改进：

（1）针对外部传感器采集到的点云数据存在冗余性与质量缺陷的问题，本文提出一种基于几何特征点云数据预处理方案。该方案主要分为点云滤波和点云分割两个部分。其中点云滤波部分使用随机降采样（Random Sampling）方法对采集到的原始点云数据完成指定数量的采集，再使用统计滤波器对降采样后的点云进行离群点去除。点云分割部分使用欧式聚类算法对离群点去除后的点云进行分割，可以较好地分离出目标物体，为后续利用有效点云信息进行点云配准奠定基础。

（2）针对以点云配准经典网络 PCRNet 为代表的点云配准方法在特征提取模块只关注点云的全局信息而忽略局部特征的问题，本文在基础网络 PCRNet 的特征提取模块进行了改进，提出了一种基于位置自适应卷积的点云配准网络 PACNet（Point Cloud Registration Network Based on Position Adaptive Convolution）。实验结果表明，PACNet 对比 PCRNet，在评价旋转矩阵和平移向量的均方误差（Mean Squared Error, MSE）、均方根误差（Root Mean Square Error, RMSE）以及平均绝对误差（Mean Absolute Error, MAE）等 6 个指标上分别提升了 25.6%、13.7%、21.6%、39.4%、22.1%、19.1%。与同类主流算法相比，本文提出的方法在点云配准任务中也具有更高的精度。

（3）针对 PCRNet 在特征融合模块将两组点云的特征在维度上进行连接而无法很好地融合不同层次特征的问题，本文对 PCRNet 的特征融合模块进行了改进，提出了一种基于双重

注意力机制的点云配准网络 PACNet-Att (Point Cloud Registration Network Based on Position Adaptive Convolution and Attention Mechanism)。实验结果表明, PACNet-Att 对比 PCRNet, 在评价旋转矩阵和平移向量的 MSE、RMSE 以及 MAE 等 6 个指标上分别提升了 30.3%、16.5%、23.4%、40.7%、22.9%、24%。与同类主流算法相比, 本文提出的方法在点云配准任务中具有更高的精度、更高的泛化性能以及更稳健的噪声鲁棒性。

关键词: 深度学习, 点云配准, 欧式聚类, 动态卷积, 注意力机制

Abstract

Point cloud registration is essential in 3D computer vision and graphics and has important applications in 3D reconstruction, 3D data fusion, simultaneous localization and mapping (SLAM), and many other research fields. With the development of autonomous driving, artificial intelligence, virtual reality technology etc., deep learning in point cloud registration has also achieved many inspiring research findings. However, point cloud data collected by external sensors often have redundancy and quality defects. On the one hand, the registration process takes up a lot of memory resources and processing time. On the other hand, noise interference will also lead to abnormal correspondences generated by feature matching, thereby influencing the accuracy of point cloud registration. In the field of deep learning, point cloud registration methods represented by PCNet only focus on the global information of point cloud without considering local features, which makes it difficult to characterize the complex space and restricts the ability of scene understanding, thereby reducing the accuracy of point cloud registration. At the same time, PCNet only connects the features of two groups of point cloud data in dimensions based on concatenation, which cannot fully use the complementarity of different levels and cannot well fuse the local features with global information, thereby destroying the accuracy of point cloud registration. Aiming at the aforementioned problems, this thesis mainly focuses on studying and improving the following three issues:

(1) Aiming at the redundancy and quality defects in point cloud data collected by external sensors, this thesis proposes a point cloud data preprocessing scheme based on geometric features, mainly containing point cloud filtering and point cloud segmentation. The point cloud filtering part utilizes the random sampling method to complete a specified number of collections for the original point cloud data and then utilizes a statistical filter to remove outliers from the downsampled point cloud. The point cloud segmentation part utilizes the euclidean clustering algorithm to separate the target object better, which can lay the foundation for the next point cloud registration using effective point cloud information.

(2) Aiming at the problem that the deep-learning-based methods for point cloud registration represented by PCNet only focus on the global information of point cloud without considering local features, this thesis improves the feature extraction module of the primary PCNet, and uses position adaptive convolution to construct our network PACNet. Compared with the primary PCNet, the

experimental results demonstrate that PACNet has brought 25.6%, 13.7%, 21.6%, 39.4%, 22.1%, and 19.1% improvements in the six criteria: mean squared error (MSE), root mean square error (RMSE) and mean absolute error (MAE) of the rotation matrix and translation vector, respectively. Compared with other state-of-the-art algorithms, the method proposed in the thesis also has higher accuracy in point cloud registration.

(3) Aiming at the problem that PCRNet only connects the features of two groups of point cloud data in dimension based on concatenation, which cannot well fuse the features of different levels, this thesis improves the feature extraction module of the primary PCRNet, and utilizes the dual attention mechanism to form our point cloud registration network PACNet-att. Compared with the primary PCRNet, the experimental results show that PACNet-Att brings 30.3%, 16.5%, 23.4%, 40.7%, 22.9%, and 24% improvement in the six criteria mentioned in issue (2). Compared with other state-of-the-art algorithms, the method proposed in the thesis has higher accuracy, better generalization performance, and stronger noise robustness in point cloud registration tasks.

Keywords: deep learning, point cloud registration, euclidean clustering, dynamic convolution, attention mechanism

目录

专用术语注释表	VII
插图清单	VIII
插表清单	IX
第一章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	1
1.3 论文研究内容	4
1.4 论文组织结构	5
第二章 点云配准基本理论概述	7
2.1 点云数据	7
2.1.1 点云数据特点	7
2.1.2 点云特征描述	8
2.2 刚体运动参数估计	8
2.2.1 基于 SVD 的线性代数求解	8
2.2.2 基于 Levenberg-Marquardt 的非线性优化	12
2.3 标准数据集	15
2.3.1 ShapeNet	15
2.3.2 ModelNet	16
2.3.3 S3DIS	17
2.4 点云配准评价指标	17
2.4.1 基于特征提取与匹配的评价指标	18
2.4.2 基于刚体运动参数估计的评价指标	19
2.5 本章小结	19
第三章 基于几何特征的点云数据预处理	21
3.1 点云滤波	21
3.1.1 降采样	21
3.1.2 离群点去除	23
3.2 点云分割	24
3.2.1 欧式聚类算法	24
3.2.2 区域生长算法	25
3.3 实验结果与对比分析	27
3.3.1 实验环境与硬件配置	27
3.3.2 对比实验分析	28
3.4 本章小结	32
第四章 基于位置自适应卷积提取特征点云配准优化	33
4.1 基础网络 PCRNet	33
4.1.1 PointNet	33
4.1.2 EMD 损失函数	34
4.1.3 PCRNet 网络设计	35
4.2 基于位置自适应卷积提取特征点云配准的 PACNet	37
4.2.1 位置自适应卷积	37
4.2.2 PACNet 网络设计	38
4.3 实验结果与分析	39
4.3.1 数据集与硬件配置	39

4.3.2 训练参数设置	40
4.3.3 整体配准效果	40
4.3.4 测试对比	41
4.4 本章小结	42
第五章 基于双重注意力机制融合特征的点云配准优化.....	43
5.1 注意力机制	43
5.2 基于双重注意力机制融合特征的 PACNet-Att.....	43
5.2.1 空间注意力机制	43
5.2.2 通道注意力机制	44
5.2.3 PACNet-Att 网络设计	45
5.3 实验结果与对比分析	46
5.3.1 实验环境和训练参数设置	46
5.3.2 整体配准效果	46
5.3.3 测试对比	47
5.3.4 消融实验	49
5.4 本章小结	49
第六章 总结与展望	51
6.1 本文总结	51
6.2 研究展望	52
参考文献	53
附录 1 攻读硕士学位期间撰写的论文	57
附录 2 攻读硕士学位期间申请的专利	58
附录 3 攻读硕士学位期间参加的科研项目	59
附录 4 攻读硕士学位期间参加的学科竞赛	60
致谢	61

专用术语注释表

缩略词说明:

AP	Average Precision	平均精度
AMOTA	Average Multi-Object Tracking Accuracy	平均多目标跟踪精度
AMOTP	Average Multi-Object Tracking PreciSion	平均多目标跟踪精度
BnB	Branch and Bound	分枝定界
EMD	Earth Mover Distance	搬土距离
FPFH	Fast Point Feature Histogram	快速点特征直方图
FPS	Farthest Point Sampling	最远点采样
GPU	Graphics Processing Unit	图形处理器
GPS	Global Positioning System	全球定位系统
ICP	Iterative Closest Point	迭代最近点算法
INS	Inertia Navigation System	惯性导航系统
LM	Levenberg-Marquardt	列文伯格·马夸尔特方法
mAcc	mean class Accuracy	平均类别准确率
mIoU	mean Intersection over Union	平均交叉联合
mAP	mean Average Precision	平均精度
MSE	Mean Squared Erro	均方误差
MAE	Mean Absolute Error	平均绝对误差
MLP	Multi-layer Perceptron	多层感知机
NDT	Normal Distributions Transform	正态分布变换
OA	Overall Accuracy	总体准确率
RMSE	Root Mean Square Error	均方根误差
SHOT	Signature of Histograms of OrienTations	方向直方图
SVD	Singular Value Decomposition	奇异值分解
SLAM	Simultaneous Localization and Mapping	即时定位与地图构建
Seq2Seq	Sequence to Sequence	序列到序列模型

插图清单

图 2.1	城市点云数据示例图	7
图 2.2	根据对应点求解刚性坐标变换的简单示例图.....	9
图 2.3	最小化 point-to-plane 误差示意图	13
图 2.4	ShapeNet 数据集的部分可视化效果图	16
图 2.5	ModelNet 数据集的部分可视化效果图	17
图 2.6	S3DIS 数据集的部分可视化效果图.....	17
图 3.1	点云预处理方案示意图	21
图 3.2	均匀降采样算法过程示意图	22
图 3.3	真实场景二维图像	27
图 3.4	真实场景点云可视化效果图	27
图 3.5	三种降采样结果对比图	28
图 3.6	统计滤波器不同参数效果对比图.....	29
图 3.7	半径滤波器不同参数效果对比图.....	29
图 3.8	点云分割效果对比图	31
图 4.1	PointNet 的轻量级架构示意图	34
图 4.2	EMD 原理示意图	34
图 4.3	PCNet 网络模型整体框架示意图	36
图 4.4	PCNet 网络预测坐标变换示意图	37
图 4.5	位置自适应卷积原理示意图	38
图 4.6	PACNet 网络整体框架示意图	39
图 4.7	不同评价指标的训练曲线和测试曲线对比图.....	40
图 4.8	配准效果展示图	41
图 5.1	PACNet-Att 网络模型整体框架示意图	45
图 5.2	双重注意力机制结构示意图	45
图 5.3	不同评价指标的训练曲线和测试曲线对比图.....	46
图 5.4	配准效果展示图	47

插表清单

表 2.1 点云信息处理任务中常见的数据集列表..... 15

表 3.1 降采样实验结果对比表 28

表 3.2 离群点去除实验结果对比表 30

表 3.3 点云分割实验结果对比表 31

表 3.4 点云预处理实验结果汇总表 32

表 4.1 不同卷积层数测试结果对比表 41

表 4.2 精度测试结果对比表 42

表 5.1 各算法精度测试结果对比表 48

表 5.2 泛化性测试结果对比表 48

表 5.3 噪声鲁棒性测试结果对比表 49

表 5.4 消融实验测试结果对比表 49

第一章 绪论

1.1 研究背景与意义

计算机视觉是人工智能领域具有前沿性的研究方向，经过数十年的发展和众多科学工作者的共同努力，计算机系统可以成功地感知信号，并从视频、图像以及其他输入信息中获得有价值的信息^[1]。计算机视觉已经跨越了不同层次的发展，从边缘识别等^[2]低级任务到场景理解等^[3]高级任务，都产生了优异的成效。其中，这些发展离不开以下三个要素：一是机器学习的发展让深度学习走向成熟，能够在原始信息中对复杂的算法执行端到端学习。二是通过图形处理器（Graphics Processing Unit, GPU）提高算力，极大地提升了算法的运行效率^[4]。三是大量标记精准的开源数据集为算法的训练提供了基础性的数据支撑。

随着高精度传感器的快速发展，点云数据的获取变得越来越方便，在数字城市^[5]、无人驾驶^[6]、人机交互^[7]等领域都有着广泛的应用。不同的三维传感器有其独特的优势和局限性，例如，三维相机可以生成稠密的点云数据，但是视野范围通常被限制在 5 米以内。激光雷达具有较长的视野范围，但生成的点云数据比较稀疏^[8]。由于传感器只能在有限视野范围内对物体进行扫描，因此，大型三维场景需要点云配准算法来生成完整的点云模型。点云配准是三维计算机视觉和图形学中的一个重要的基础领域，在三维重建^[9]、三维数据融合^[10]、SLAM^[11]等应用中都有着特殊而重要的意义。点云配准旨在估计源点云和模板点云之间的刚性运动变换参数，应用变换参数后实现两片点云的对齐，从而将不同视角采集的点云数据拼接成一个完整的三维点云模型^[12]。一般来说，采集来自不同视角的点云之间会存在一定的重叠区域，如何从重叠区域找到用于配准的“匹配点对”是匹配算法的关键^[13]。因此，点云的配准问题可以转化成点对的配准问题。传统方法中，查找匹配点对时会计算几何特征的相似性，但是对点云的初始位置要求较为严格，需要在空间上有一定的重叠部分才能获得较好的配准效果。随着深度学习的发展，点云配准网络可以直接处理无序的原始点云数据，逐渐对点云的初始位置不敏感，并获得了高精度并且健壮的鲁棒性效果^[14]。最近，基于深度学习方法和优化方法的组合进一步提高了点云配准的性能，属于当前点云配准领域研究的热点问题^[15]。

1.2 国内外研究现状

目前，点云配准根据其实现方法主要可以分为两类，一类是基于传统方法的点云配准，

通常分为两个阶段,即粗配准和精配准^[16]。一般首先使用粗配准算法获得一个良好的初始值,其次使用精配准算法最小化两片点云之间的刚性变换误差。另一类是基于深度学习的点云配准,一般分为端到端的学习方法和基于特征学习的方法。端到端的学习方法一次性完成所有特征点的学习和刚性变换的估计,更关注点云之间的全局特征。基于特征学习的方法集中关注描述特征点的学习,而刚性变换则由一些鲁棒的姿态估计器获得^[17]。

传统方法中应用最广泛的点云配准算法就是 1992 年 Besl 等^[18]提出的迭代最近点算法 (Iterative Closest Point, ICP), 该算法作为开山之作对点云精配准来说具有开创性的意义。ICP 算法原理简单,实现起来也很容易,但是该算法对初始值的要求很高,并且非常容易陷入局部最优,算法效率不高,因此点云配准领域科研工作者们在此基础上做出了许多改进,并衍生出了各种 ICP 算法的变形。2009 年, M.Magnusson 等^[19]提出了 3D-NDT (Normal Distributions Transform) 算法,该算法对局部点的统计特性进行了表征,将点云分割为多个同尺寸的栅格,每个栅格用正态分布表示,并使用 ICP 算法计算了每个栅格中所有点的均值和方差,通过匹配每个栅格的正态分布来匹配点集。NDT 算法对点集进行了抽象,并不需要逐点找对应点,但它本质上仍是基于最近点假设,仍然容易陷入局部极小值。2015 年, J.Yang 等^[20]提出了 Go-ICP 算法, Go-ICP 算法基于搜索整个三维运动空间 $SE(3)$ 的分枝定界 (Branch and Bound, BnB) 方案,利用 $SE(3)$ 的特殊几何结构,推导出配准误差函数的上下界,之后将局部 ICP 应用于 BnB 方案中,能够保证算法获得全局最优,同时加快了配准的速度。2016 年, Zhou 等^[21]提出了 FGR (Fast Global Registration) 算法,该算法利用损失函数将非凸的问题进行凸化,通过引入 GM (Geman-McClure) 估计量和回归损失函数,在一定程度上都可以将函数光滑化,但是 GM 估计量下降更快,所以它对噪点的限制作用更大。传统方法中,还有一类方法基于手动设计的特征获得匹配关系。这类方法通常首先从输入的原始点云中提取关键点和特征描述符。其次利用异常值剔除法和最近邻算法选择内部关键点,获得更准确的关键点对应关系。最后估计从源点云到目标点云的转换。现有的相对较好的手工特征,如快速点特征直方图^[22] (Fast Point Feature Histogram, FPFH)、方向直方图^[23] (Signature of Histograms of Orientations, SHOT) 等已经取得了显著的效果,可以对特殊场景下的点云实现特征提取与匹配。然而,这类方法对异常值和重复几何结构很敏感,缺乏点云的语义信息,因此精度较低且鲁棒性较差。

2017 年, Charles 等^[24]人提出首个可以直接处理点云数据的深度学习模型框架 PointNet, 该网络通过一个简单的对称函数即最大池化来处理每个点,消除了输出结果因点云数据的无序性所受到的影响。之后,基于深度学习的点云配准研究也取得了一些重要成果。Zeng 等^[25]提出了 3DMatch 深度网络模型,通过学习自定义局部空间域的描述子,该网络可以建立局部

三维数据的对应关系。实验结果表明,该方法不仅在重建的局部几何匹配上表现很好,并且可以扩展到不同的任务和尺度中。2018年,Deng等^[26]提出了PPFNet(Point Pair Feature Network)模型,将一些简单的几何特征属性组合成原始特征,生成易区分且抗旋转的三维局部特征子。实验结果表明,该方法可以得到一个理想的三维局部特征子,但对于旋转操作很敏感。Yew等^[27]提出了3DFeat-Net网络模型,使用弱监督学习来学习三维特征检测器和描述子,对来自GPS/INS(Global Positioning System/ Inertia Navigation System)组合导航系统标记的三维点云使用对齐操作和注意力机制来学习特征点匹配。2019年,Aoki等^[28]提出了PointNetLK网络模型,把PointNet网络看作是一个成像函数,并提取两片输入点云的全局特征,然后使用雅可比矩阵的逆形式估计变换矩阵。通过估计变换矩阵,目标是最小化两个特征之间的特征差异。Wang等^[29]提出了一种新的DCP(Deep Closest Point)网络模型,该网络使用动态图卷积提取特征,并使用注意力模块将点云配准问题视为序列到序列模型(Sequence to Sequence, Seq2Seq)问题,之后利用奇异值分解(Singular Value Decomposition, SVD)线性求解旋转矩阵和平移向量。Lu等^[30]提出了DeepVCP(Virtual Corresponding Points)网络模型,通过引入语义特征自动避开动态目标并选取易于匹配的关键点。在关键点的候选匹配区域内,通过计算特征相似概率的方式来生成同名点,最后网络损失综合了关键点的局部和全局匹配效果。实验结果表明,该方法达到了与传统ICP算法相当的匹配精度以及更高的稳定性。Sarode等^[31]提出了PCRNet,其使用PointNet对点云提取全局特征,再将两个全局特征进行连接,然后使用5层全连接网络回归相对位姿。由于不需要计算点云之间的一一对应关系,PCRNet能够快速地进行点云配准。2020年,Pairs等^[32]通过使用深度残差层和卷积层构建了3DRegNet网络模型,将匹配点的对应关系区分为正常或错误,并回归了当前扫描帧对齐到参考帧的运动参数。实验结果显示,该算法效果远远优于传统模型,并已推广至多视图场景。2021年,Fu等^[33]在点云中构建图,再通过基于深度图匹配的方法建立更好的对应关系。实验结果表明该方法虽然配准效果很好,但完全无法应用于真实场景,且泛化性能较差。目前,主流的基于深度学习的点云配准方法往往集成了用于建立两组未对齐点云对应关系的深度神经网络和用于计算两组未对齐点云刚性运动变换的线性或非线性算法,从而构建用于点云配准的端到端可训练网络。这些方法探索深层抽象特征来建立对应关系,尽管它们具有良好的性能,但从点云中提取的所有关键特征辨别能力较差,而输入点云中的噪声干扰也会导致特征匹配产生的对应关系容易出现异常值,从而破坏了配准精度^[34]。因此,基于深度学习的点云配准还有待国内外学者们的进一步研究探索。

1.3 论文研究内容

由于外部传感器采集到的点云数据往往存在冗余性与质量缺陷,这一方面使得配准过程需要占用大量的内存资源和处理时间,另一方面输入点云中的噪声干扰也会导致特征匹配产生的对应关系出现异常值,从而影响点云配准的精度。因此,研究并实现有效的点云数据预处理方案是后续进行特征提取与匹配的基础,也是利用关键特征进行点云配准的基础。

在深度学习领域,以 **PCNet** 为代表的点云配准网络在特征提取模块只关注了点云的全局信息,并没有考虑到点云的局部特征,难以表征点云空间复杂的变化关系,一定程度限制了其对场景的理解能力。同时,**PCNet** 网络在特征融合模块仅以数组拼接的方式将两组点云的特征在维度上进行连接,不能充分利用不同层次特征的互补性,无法很好地融合点云的局部特征和全局信息。在特征提取模块和特征融合模块的局限性都导致了 **PCNet** 网络在点云配准任务中精度的欠缺。因此,研究并改进点云配准网络中的特征提取模块和特征融合模块是提高点云配准精度的有效方法。

本文主要有三个研究点,主要内容如下:

(1) 针对外部传感器采集到的点云数据存在冗余性与质量缺陷的问题,本文提出一种基于几何特征点云数据预处理方案,该方案主要分为点云滤波和点云分割两个部分。通过分析和对比不同方法以及不同参数的算法,本文确定了点云预处理的有效方案,即使用随机降采样对采集到的原始点云数据完成指定数量的采集,再使用统计滤波器对降采样后的点云进行离群点去除,最后使用欧式聚类算法对离群点去除后的点云进行分割。相比于均匀降采样和体素降采样,随机降采样在三种降采样方法中算法复杂度最低,运行速度最快,同时可以指定采样点的数量。统计滤波器根据全局距离平均值和方差来定义阈值标准,无需人工指定阈值标准,更为高效,并且更能适应不同的点云场景。欧式聚类算法以邻域内点与点之间的距离作为判定标准,算法复杂度更低,可以较好地分离出目标物体,为后续利用有效点云信息进行点云配准奠定基础。

(2) 针对以点云配准经典网络 **PCNet** 为代表的点云配准方法在特征提取模块只关注点云的全局信息而忽略局部特征的问题,本文在基础网络 **PCNet** 的特征提取模块进行了改进,提出了一种新的基于深度学习的点云配准网络模型 **PACNet**。本网络引入了位置自适应卷积,其根据每个点与其邻域点的位置关系学习权重系数,并组合权重矩阵自适应地构建卷积内核,可以灵活地捕获局部区域的信息,获得更好的场景理解能力,从而提高点云配准的精度。

(3) 针对 **PCNet** 在特征融合模块仅仅以数组拼接方式将两组点云的特征在维度上进行连接而无法很好地融合不同层次特征的问题,本文在基础网络 **PCNet** 的特征融合模块进行

了改进,提出了一种新的基于深度学习的点云配准网络模型 PACNet-Att。该网络引入了结合空间注意力机制和通道注意力机制的双重注意力机制,其中的空间注意力机制通过对空间维度上的特征进行聚焦,可以加深网络结构更深层次的信息传递。通道注意力机制通过对通道维度上的特定语义进行响应增强,可以提高特征的可分辨性。将两者结合的双重注意力机制可以捕获不同层次特征的长期语义依赖关系,有效融合局部特征和全局信息,从而提高点云配准的精度。

1.4 论文组织结构

第一章绪论阐述了本文基于深度学习的点云配准的研究背景与意义,并分析了传统方法和深度学习方法这两个方面在点云配准课题中的研究现状,最后明确了本文的主要内容和组织结构。

第二章介绍了与点云配准相关的基本理论。首先介绍了点云数据具有的特点,其次阐述了点云数据的两类特征描述方式,分别是局部特征描述和全局特征描述。然后详细分析了点云配准任务中基于 SVD 的线性代数法和基于列文伯格·马夸尔特算法(Levenberg-Marquardt, LM)的非线性优化法估计刚体运动参数的过程,接着介绍了点云配准中的两类数据集,即合成数据集和真实场景数据集,并展示了 ShapeNet、ModelNet 以及 S3DIS 这三种常用的标准数据集。最后介绍了点云信息处理任务中常用的评价标准,并重点介绍了点云配准任务中基于特征提取与匹配以及基于刚体运动参数估计的两类评价指标。

第三章主要研究了基于几何特征点云数据预处理方案。针对外部传感器采集到的点云数据存在冗余性与质量缺陷的问题,本文提出一种基于几何特征点云数据预处理方案,分析并对比不同的点云滤波方法以及不同的点云分割方法处理点云的效果,确定了点云预处理的有效方案,为后续利用有效点云信息进行点云配准奠定了基础。

第四章提出了基于位置自适应卷积提取特征点云配准优化方法。针对以点云配准经典网络 PCRNNet 为代表的点云配准方法在特征提取模块只关注点云的全局信息而忽略局部特征的问题,本章基于 PCRNNet 网络提出了一种基于位置自适应卷积提取特征点云配准网络 PACNet。首先阐述并分析了 PCRNNet 网络的工作机制和存在的局限性,其次解释了位置自适应卷积的工作原理,并介绍了 PACNet 网络在特征提取模块的改进思路,最后在标准数据集上,将本章提出的方法与现有的同类主流算法进行对比,验证了本章提出的方法在点云配准任务中具有更高的精度。

第五章提出了基于双重注意力机制融合特征点云配准优化方法。针对 PCRNNet 在特征

融合模块仅仅以数组拼接方式将两组点云的特征在维度上进行连接而无法很好地融合不同层次特征的问题，本章基于 PCRNet 深度学习网络模型提出了一种基于双重注意力机制融合特征的点云配准网络 PACNet-Att。首先介绍了与注意力机制有关的背景知识，其次解释了双重注意力机制中空间注意力机制和通道注意力机制的工作原理，并介绍了 PACNet-Att 网络在特征融合模块的改进思路。最后在标准数据集上，将本章提出的方法与现有的同类主流算法进行对比，证明了本章提出的方法在点云配准任务中具有更高的精度、更高的泛化性能以及更稳健的噪声鲁棒性。同时，通过消融实验，验证了位置自适应卷积和双重注意力机制在点云配准任务中的有效性。

第六章对本文的研究工作进行了总结与展望。本章首先对本文所做出的研究工作进行了详细的总结，其次指出了本文提出的方法可进一步探索的空间，并提出了可行的改进思路与方向，最后对今后的研究进行了展望。

第二章 点云配准基本理论概述

本章首先介绍了点云数据的特点，并且介绍了点云数据的两类特征描述方式，其次分析了点云配准任务中刚体运动参数估计的方法，最后详细阐述了点云配准任务中常用的标准数据集以及评价指标。

2.1 点云数据

三维数据的表征方式可以有多种类型，例如点云、体素^[35]、网格^[36]和深度图像等。点云作为其中最为常用的格式之一，保留了三维空间中原始的几何信息，不需要任何离散化。因此，点云是许多与场景理解相关的应用领域首选的三维数据表征方法，如图 2.1 所示为激光雷达采集到的城市点云数据。

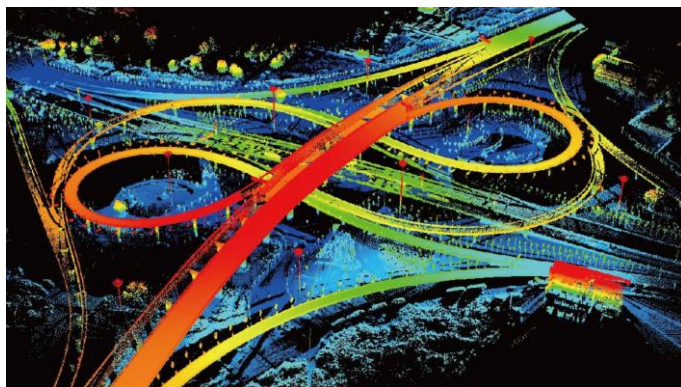


图 2.1 城市点云数据示例图

2.1.1 点云数据特点

点云数据的本质是一个点集，它具有以下特征：（1）无序性（Unordered）：在欧式空间的度量中，点和点之间并不像一维和二维空间有固定的顺序。对于点云数据信息，每个点的顺序不应当影响到结果。（2）局部相关性（Interaction Among Points）：在空间中的点并不独立，与其邻域点之间的位置关系信息是有意义的，它代表了物体的形状特征。（3）不变性（Invariance Under Transformation）：对于点云数据应该满足一些空间变换的不变性，例如平移和旋转，这些都不会影响输出的结果。卷积网络通常需要规则的数据形式作为输入，但由于点云是非规则数据类型，所以通常的做法大多都是先对点云进行规则化的处理，将空间划分成栅格的体素^[37]就是一种典型的做法。但这类方法会造成有很多不必要的体积划分，使得

输入数据变得稀疏,同时会影响点云数据的不变性。之后,PointNet 作为首个可以直接处理点云数据的深度学习网络模型,很好地解决了上述的问题。

2.1.2 点云特征描述

三维点云特征描述在点云信息处理任务中具有重要的作用,点云特征描述的精确度直接影响到后续点云特征提取的准确性。以尺度为分类标准,三维点云特征描述一般可以分为局部特征描述和全局特征描述。局部特征描述包含法向量、曲率等几何特征^[38],全局特征描述包含点云的拓扑结构^[39]等。

2.2 刚体运动参数估计

刚性运动参数估计是点云配准任务的最后阶段。运动参数有不同的表示形式,如四元数^[40]、欧拉角^[41]、旋转矩阵和平移向量等。其中,旋转矩阵和平移向量是最常用的,并且已被证明是基于对应的奇异值分解的最优解。

2.2.1 基于 SVD 的线性代数求解

奇异值分解^[42](SVD)是线性代数中的特征分解在任意矩阵上的推广,在机器学习领域有着重要应用。

在刚体对齐问题中,假设点云 $P = \{p_i | i = 1, 2, \dots, n\} \subset \mathbb{R}^3$ 从空间中的一个姿态刚体运动到另一个姿态,运动后的点云为 $Q = \{q_j | j = 1, 2, \dots, m\} \subset \mathbb{R}^3$,不包括拉伸、弯曲、变形等仿射变换。刚体运动参数由旋转矩阵 $R \in SO(3)$ 和平移向量 $t \in \mathbb{R}^3$ 组成。通常情况下,刚体是如何运动的无法确定,但知道刚体在运动前后的对应点坐标,因此可以通过这些对应点求解出旋转矩阵 R 和平移向量 t 。如下图 2.2 所示为简单的 3 组对应点的例子,对应点具有相同的颜色。点云配准任务希望找到最佳的旋转矩阵 R 和平移向量 t ,将数据集 A 中的点与数据集 B 中的点对齐。

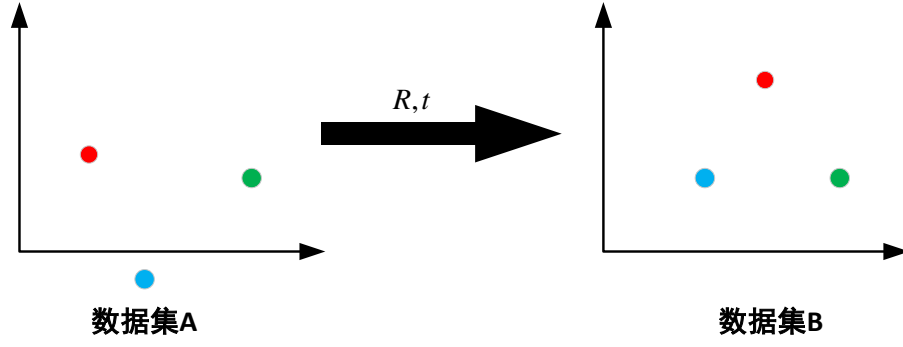


图 2.2 根据对应点求解刚性坐标变换的简单示例图

假设 $n = m$ ，则目标函数可表示为：

$$F(R, t) = \operatorname{argmin} \sum_{i=1}^n \|Rp_i + t - q_i\|^2 \quad (2.1)$$

首先对公式 2.1 求导，可以得到：

$$0 = \frac{\partial F}{\partial t} = \sum_{i=1}^n 2(Rp_i + t - q_i) = 2nt + 2R \left(\sum_{i=1}^n p_i \right) - 2 \sum_{i=1}^n q_i \quad (2.2)$$

再对公式 2.2 两边同时除以 n ，可以得到：

$$\frac{0}{n} = 2t + \frac{2R(\sum_{i=1}^n p_i)}{n} - \frac{2 \sum_{i=1}^n q_i}{n} \quad (2.3)$$

其次定义点云 P 的质心 \hat{p} 和点云 Q 的质心 \hat{q} ，质心就是点云中所有点的平均值，分别表示为：

$$\hat{p} = \frac{\sum_{i=1}^n p_i}{n} \quad (2.4)$$

$$\hat{q} = \frac{\sum_{i=1}^n q_i}{n} \quad (2.5)$$

将其带入公式 2.3 中，可以得到：

$$0 = 2t + 2R\hat{p} - 2\hat{q} \quad (2.6)$$

化简后得到：

$$t = \hat{q} - R\hat{p} \quad (2.7)$$

将公式 2.7 带入公式 2.1，可以得到：

$$\sum_{i=1}^n \|Rp_i + t - q_i\|^2 = \sum_{i=1}^n \|Rp_i + \hat{q} - R\hat{p} - q_i\|^2 \quad (2.8)$$

$$= \sum_{i=1}^n \|R(p_i - \hat{p}) - (q_i - \hat{q})\|^2 \quad (2.9)$$

假设集合 $P' = \{p_i - \hat{p} | i = 1, 2, \dots, n\} \subset R^3$ ，集合 $Q' = \{q_j - \hat{q} | j = 1, 2, \dots, m\} \subset R^3$ ， p'_i 和 q'_j

分别表示新数据集中的点：

$$p'_i = p_i - \hat{p} \quad (2.10)$$

$$q'_i = q_i - \hat{q} \quad (2.11)$$

此时，求解旋转矩阵 R 的目标函数可以等价为：

$$F(R) = \argmin \sum_{i=1}^n \|Rp'_i - q'_i\|^2 \quad (2.12)$$

接下来，利用 SVD 求解出最佳的旋转矩阵 R ，由于矩阵 Frobenius 范数（F-范数）的定义如下表示：

$$\|A\|_F = \sqrt{\text{tr}(A^T A)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2} \quad (2.13)$$

所以矩阵 F-范数的平方可以转化为矩阵的内积，再转化为矩阵的迹，即：

$$\|A\|_F^2 = \text{tr}(A^T A) \quad (2.14)$$

将此性质应用于公式 2.12，可以得到：

$$\sum_{i=1}^n \|Rp'_i - q'_i\|^2 = (Rp'_i - q'_i)^T (Rp'_i - q'_i) = (p_i^T R^T - q_i^T) (Rp'_i - q'_i) \quad (2.15)$$

$$= p_i^T R^T Rp'_i - q_i^T Rp'_i - p_i^T R^T q'_i + q_i^T q'_i \quad (2.16)$$

由正交矩阵的性质 $R^T R = R R^T = I$ ，可以得到：

$$\sum_{i=1}^n \|Rp'_i - q'_i\|^2 = p_i^T p'_i - q_i^T Rp'_i - p_i^T R^T q'_i + q_i^T q'_i \quad (2.17)$$

$$= p_i^T R^T q'_i = (p_i^T R^T q'_i)^T = q_i^T Rp'_i \quad (2.18)$$

由 $x_i^T R^T q'_i = (x_i^T R^T q'_i)^T = q_i^T Rp'_i$ ，公式 2.17 可以简写为：

$$\sum_{i=1}^n \|Rp'_i - q'_i\|^2 = p_i^T p'_i - 2q_i^T Rp'_i + q_i^T q'_i \quad (2.19)$$

将公式 2.19 带入公式 2.12，可以得到：

$$F(R) = \argmin \sum_{i=1}^n \|Rp'_i - q'_i\|^2 = \argmin \sum_{i=1}^n (p_i^T p'_i - 2q_i^T Rp'_i + q_i^T q'_i) \quad (2.20)$$

$$= \argmin \left(\sum_{i=1}^n p_i^T p'_i - \sum_{i=1}^n 2q_i^T Rp'_i + \sum_{i=1}^n q_i^T q'_i \right) \quad (2.21)$$

其中 $\sum_{i=1}^n p_i^T p'_i$ 和 $\sum_{i=1}^n q_i^T q'_i$ 与旋转矩阵 R 无关，因此求解旋转矩阵 R 的目标函数可以简

化为下式:

$$F(R) = \operatorname{argmin} \left(- \sum_{i=1}^n 2q'^T_i R p'_i \right) = \operatorname{argmax} \left(\sum_{i=1}^n q'^T_i R p'_i \right) \quad (2.22)$$

将集合 P' 和集合 Q' 分别表示为矩阵形式:

$$P' = [p'_1 \quad p'_2 \quad \cdots \quad p'_n] \quad (2.23)$$

$$Q' = [q'_1 \quad q'_2 \quad \cdots \quad q'_n] \quad (2.24)$$

其中 p'_i 、 q'_i 分别为矩阵 P' 和矩阵 Q' 的列向量。那么:

$$\sum_{i=1}^n q'^T_i R p'_i = \begin{bmatrix} q'^T_1 \\ q'^T_2 \\ \vdots \\ q'^T_n \end{bmatrix} R [p'_1 \quad p'_2 \quad \cdots \quad p'_n] \quad (2.25)$$

$$= \begin{bmatrix} w_1 q'^T_1 \\ w_2 q'^T_2 \\ \vdots \\ w_n q'^T_n \end{bmatrix} [R p'_1 \quad R p'_2 \quad \cdots \quad R p'_n] \quad (2.26)$$

$$= \begin{bmatrix} q'^T_1 R p'_1 & & & \\ & q'^T_2 R p'_2 & & \\ & & \ddots & \\ & & & q'^T_n R p'_n \end{bmatrix} \quad (2.27)$$

$$= \operatorname{tr}(Q'^T R P') \quad (2.28)$$

由矩阵的迹的性质 $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ ，可以得到:

$$\sum_{i=1}^n q'^T_i R p'_i = \operatorname{tr}(Q'^T R P') = \operatorname{tr}(Q'^T (R P')) \quad (2.29)$$

$$= \operatorname{tr}((R P') Q'^T) = \operatorname{tr}(R P' Q'^T) \quad (2.30)$$

记矩阵 $C = P' Q'^T$ ，利用 SVD 分解矩阵 C ，即:

$$C = U \Sigma V^T \quad (2.31)$$

其中 U 是 $n \times 3$ 阶酉矩阵， Σ 是半正定 3×3 阶对角矩阵，其对角线上的元素 σ_i 为矩阵 C 的奇异值，而 V^T 是 $3 \times n$ 阶酉矩阵，将公式 2.31 带入公式 2.22，可以得到:

$$\sum_{i=1}^n q'^T_i R p'_i = \operatorname{tr}(R U \Sigma V^T) = \operatorname{tr}(R U (\Sigma V^T)) \quad (2.32)$$

$$= \operatorname{tr}((\Sigma V^T) R U) = \operatorname{tr}(\Sigma V^T R U) \quad (2.33)$$

由于 V 、 R 、 U 都是正交矩阵，记矩阵 $D = V^T R U$ ，不难得出矩阵 D 同样也是正交矩阵，综上所述可以得到：

$$F(R) = \operatorname{argmax} \left(\sum_{i=1}^n q'^T_i R p'_i \right) = \operatorname{argmax} \operatorname{tr}(\Sigma D) \quad (2.34)$$

$$= \operatorname{argmax} \operatorname{tr} \left(\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{31} & d_{33} \end{pmatrix} \right) \quad (2.35)$$

$$= \max(\sigma_1 d_{11} + \sigma_1 d_{22} + \sigma_1 d_{33}) \quad (2.36)$$

由于矩阵 D 是正交矩阵，正交矩阵中每个元素的绝对值都小于等于 1，且矩阵的奇异值总是大于等于 0，不难得出当 $d_{11} = d_{22} = d_{33} = 1$ ，即 $D = V^T R U = I$ 时， $F(R)$ 取最大值。继而可以得到：

$$R = V U^T \quad (2.37)$$

计算出旋转矩阵 R 之后，便由公式 2.7 可以计算出平移向量 t ：

$$t = \hat{q} - R \hat{p} \quad (2.38)$$

2.2.2 基于 Levenberg-Marquardt 的非线性优化

列文伯格·马夸尔特（Levenberg-Marquardt）算法^[43]是一种主流的信任域方法，用于在参数空间上找到非线性函数的最小值。本质上，目标函数的可信区域是在非线性函数的内部建模，当找到合适的拟合时，信任区域就会扩大。

求解空间点云刚性变化运动参数通常在源点云和模板点云之间建立点的对应，通过最小化对应点之间的误差，可以得到源点云和模板点云之间的变换矩阵。度量对应点之间的误差主要有点到点（point-to-point）、点到面（point-to-plane）、面到面（plane-to-plane）这几类方式。其中，最小化 point-to-plane 误差就是最小化源点云中的点到模板点云中对应点的切平面距离的平方和^[44]，如下图 2.3 所示，这类问题实质上是一个非线性最小二乘问题，常常使用 LM 算法求解最优变换矩阵。

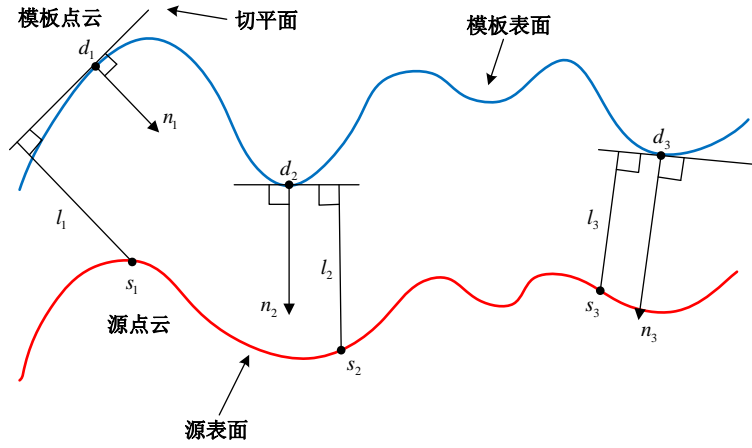


图 2.3 最小化 point-to-plane 误差示意图

LM 算法是在高斯牛顿法^[45]的基础上迭代求解非线性最优目标函数。设 $\mathbf{x} \in \mathcal{R}^n$ 是待优化的向量参数，LM 希望找到最优的点 \mathbf{x}^* ，从而最小化目标函数 $F(\cdot)$ ：

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} F(\mathbf{x}) \quad (2.39)$$

其中，目标函数 $F(\mathbf{x})$ 可以表示为：

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{f}(\mathbf{x})\|^2 = \frac{1}{2} \mathbf{f}^T(\mathbf{x}) \mathbf{f}(\mathbf{x}) \quad (2.40)$$

其中， $\mathbf{f}: \mathcal{R}^n \rightarrow \mathcal{R}^m$ 表示从 n 维欧式空间 \mathcal{R}^n 到 m 维欧式空间 \mathcal{R}^m 的映射，有时可能包括与某些参考数据或者观察到的数据的比较。一个非常简单的线性示例是 $\mathbf{f}(\mathbf{x}) = \mathbf{b} - \mathbf{a}\mathbf{x}$ ，但是 $\mathbf{f}(\mathbf{x})$ 可以有任何形式：

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}) \quad \dots \quad f_m(\mathbf{x}))^T \quad (2.41)$$

LM 将目标函数的雅可比矩阵^[46]定义为 $m \times n$ 矩阵：

$$J_{ij}(\mathbf{x}) = \frac{\partial f_i}{\partial x_j}, i = 1, 2, \dots, m \quad j = 1, 2, \dots, n \quad (2.42)$$

目标函数的 *Hessian* 矩阵^[47] 为 $n \times n$ 二阶导数矩阵，它近似为：

$$\mathbf{H}(\mathbf{x}) = \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) \quad (2.43)$$

如果没有雅可比矩阵所需导数的封闭形式表达式，LM 可以使用每个单独变量 Δx_j 的增量从有限差分中估计它们：

$$J_{ij}(\mathbf{x}) \simeq \frac{f_i(\mathbf{x} + \Delta \mathbf{x}) - f_i(\mathbf{x} - \Delta \mathbf{x})}{2\Delta x_j} \quad (2.44)$$

之后，LM 方法通过线性近似最小化目标函数：

$$F(\mathbf{x} + \mathbf{h}) \simeq L(\mathbf{h}) = F(\mathbf{x}) + \mathbf{h}g(\mathbf{x}) + \frac{1}{2} \mathbf{h}^T \mathbf{H}(\mathbf{x}) \mathbf{h} \quad (2.45)$$

其中， $g(\mathbf{x})$ 是 $\mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x})$ 的梯度。

假设 x_t^* 是对最优参数集 x^* 的迭代逼近序列, 给定初始量 x_0^* , LM 算法每次迭代的过程可以表示为:

$$x_{t+1}^* = x_t^* + h_{lm} \quad (2.46)$$

其中每个步骤来自:

$$h_{lm} = -(H(x) + \lambda I)^{-1} g(x) \quad (2.47)$$

阻尼系数 λ 根据启发式规则动态调整, 如果一次迭代后, 目标函数值有所下降, 则缩小阻尼因子 λ , 否则, 增大阻尼因子 λ 。基本上, 当达到最大迭代次数或向量参数的变化非常小时, 迭代终止。

在刚体运动参数估计任务中, LM 算法的主要思想为:

(1) 搜索运动前的点云数据中的特征点 p_i 在运动后的点云数据中对应的特征点 q_j 以及特征面 S_l ;

(2) 计算特征点 p_i 到特征面 S_l 的欧氏距离 d_e ;

(3) 根据特征点到特征面的距离度量, 建立匹配对的约束关系;

$$f(p_i, T) = d_e \quad (2.48)$$

其中, T 表示刚体运动变换参数。

将变换矩阵作为待估计参数, 利用 LM 非线性优化方法求解位姿变换:

$$\hat{T} = T - (J^T J + \lambda I)^{-1} J^T f \quad (2.49)$$

由链式法求出雅可比矩阵:

$$J = \frac{\partial f}{\partial T} = \frac{\partial f}{\partial p_i} \frac{\partial p_i}{\partial T} \quad (2.50)$$

将其带入公式 2.49 可求解得到刚体运动变换参数 T 。

(4) 判断变换参数是否属于最优估计, 若满足:

$$0 < \frac{f(p_i, \hat{T}) - f(p_i, T)}{|f(p_i, T)|} < T_d \quad (2.51)$$

则当前变换参数为最优估计, 否则, 增大阻尼因子 λ , 继续迭代求解最优变换参数。其中, T_d 为距离阈值。

(5) 将求解的最优变换参数作用于运动前的特征点 p_i , 得到 q'_j , 若满足:

$$|p_i - q_j| > |p_i - q'_j| \quad (2.52)$$

则认为 q_j 为点 p_i 的最佳匹配点, 否则, 缩小阻尼因子 λ , 继续迭代求解最优匹配点对。若达到最大迭代次数仍未得到最优匹配点对, 则将点 p_i 作为噪点从原始点云数据中剔除, 转向其它

特征点的匹配与运动参数估计。

2.3 标准数据集

本节将介绍用于三维点云配准的标准数据集。在评估不同指标的性能时，数据集必不可少。配准任务的点云数据集可以分为两类，即合成数据集和真实场景数据集。合成数据集中的对象是完整的，不存在任何遮挡以及无关背景的干扰。真实场景数据集在不同层次上存在遮挡以及无关背景的干扰。真实场景数据集包括室内场景数据集和室外场景数据集，可以通过激光雷达直接获取，或者通过 RGB-D（RGB-Depth Map）相机间接获取的多视角深度图像转换而来。室外场景数据集专为自动驾驶而设计，其中的对象在空间上的分离性较好，并且点云数据分布均匀。下表 2.1 所示为点云配准任务中常见的数据集，这些数据集可用于研究并解决点云配准任务中存在的相似信息干扰、形状不完整以及类别不平衡等各种具有挑战性的课题。

表 2.1 点云信息处理任务中常见的数据集列表

数据集名称	年份	模型/场景数量	类别数量	训练集模型数量	测试集模型数量	类型	结构
McGill Benchmark ^[48]	2008	456	19	304	152	合成	Mesh
Sydney Urban Objects ^[49]	2013	588	14	—	—	真实场景	Point Clouds
ModelNet10 ^[50]	2015	4899	10	3991	605	合成	Mesh
ModelNet40 ^[50]	2015	12311	40	9843	3	合成	Mesh
ShapeNet ^[51]	2015	51190	55	—	—	合成	Mesh
ScanNet ^[52]	2017	12283	17	9677	2606	真实场景	RGB-D
ScanObjectNN ^[53]	2016	2902	15	2321	581	真实场景	Point Clouds
S3DIS ^[54]	2017	271	13	—	—	真实场景	Point Clouds

2.3.1 ShapeNet

ShapeNet 数据集是由斯坦福大学、普林斯顿大学和丰田工业大学芝加哥分校合作开发的一个大型三维模型存储库，其中的信息非常丰富，包含跨越多种语义类别的模型，总共有 16 个物体种类，每个物体又可以分成若干个部件，例如飞机可以分成机身、机翼等部件，总共有 55 个部件种类，51190 个三维模型。每个模型都有相对的注释集，注释包括一致的刚性对齐、零件和关键点以及以真实世界为单位的物理尺寸等几何属性。通过可视化模型对象，可

以驱动数据的几何分析，并且能够给计算机视觉研究带来可参考的定量标准。ShapeNet 数据集的部分可视化效果如下图 2.4 所示。



图 2.4 ShapeNet 数据集的部分可视化效果图

2.3.2 ModelNet

ModelNet 数据集由普林斯顿视觉与机器人实验室开发完成，该数据集共有 662 个对象类别，127915 个三维模型，以及 10 类标记方向的数据，并且包含三个子集：（1）ModelNet10 包含 10 个类别的三维模型，类别分别为浴缸、床、椅子、服务台、梳妆台、监控、床头柜、沙发、桌子和卫生间。（2）ModelNet40 包含 40 个类别的三维模型，其中像桌子、花瓶、飞机这样具有规则对称结构的点云，可以利用对称信息来识别物体的关键信息，而对于吉他、花、人这样具有复杂结构的点云，很难获得识别物体的关键信息，在点云处理任务中具有一定的挑战性。（3）Aligned40 包含 40 个类别对齐的三维模型。ModelNet 数据集旨在为计算机视觉、机器人自动化领域和认知科学领域的研究人员提供大规模的三维物体模型。ModelNet 数据集的部分可视化效果如下图 2.5 所示。

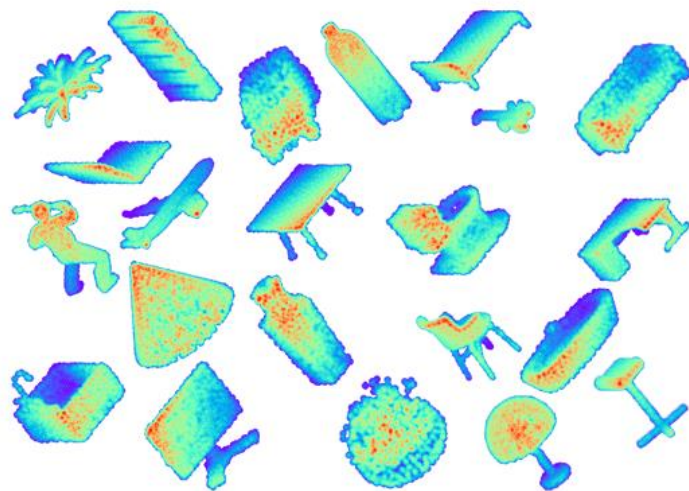


图 2.5 ModelNet 数据集的部分可视化效果图

2.3.3 S3DIS

S3DIS 数据集由斯坦福大学开发完成, 包含 13 类带有像素级语义标注的室内场景对象, 它们包括具有结构元素的天花板、地板、墙壁、梁、柱子、窗户和门以及桌子、椅子、沙发和书架这些常见的物品和家具。数据共有 9 个维度, 分别为 XYZ 坐标、RGB 颜色和关于它所属空间的归一化坐标。该数据集由来自三个不同建筑物的大型室内区域组成, 场景包括会议室、个人办公室、洗手间、开放空间、大堂、楼梯和走廊等。其中一个区域包括多个楼层, 而其余区域则只有一个楼层。S3DIS 数据集的部分可视化效果如下图 2.6 所示。

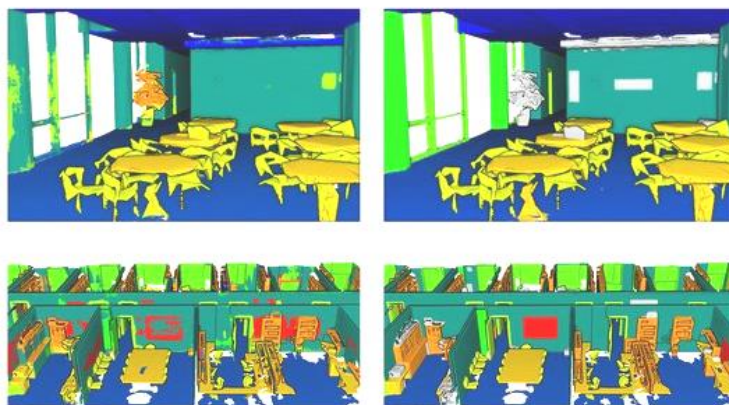


图 2.6 S3DIS 数据集的部分可视化效果图

2.4 点云配准评价指标

在点云信息处理任务中, 前沿学者已经提出并总结了不同的评价指标用于性能测试。对于三维形状分类任务, 总体准确率 (Overall Accuracy, OA) 和平均类别准确率 (Mean Class

Accuracy, mAcc) 是最常用的评价指标。OA 代表所有测试实例的平均精度, mAcc 代表所有形状类别的平均精度。对于三维对象识别任务, 平均精度 (Average Precision, AP) 是最常用的评价指标。AP 计算的是精确召回曲线下的面积。对于三维多目标跟踪任务, 平均多目标跟踪精度 (Average Multi-Object Tracking Accuracy, AMOTA) 和平均多目标跟踪精度 (Average Multi-Object Tracking Precision, AMOTP) 是最常用的评价指标。对于三维点云分割任务, 平均交叉联合 (Mean Intersection over Union, mIoU)、平均精度 (Mean Average Precision, mAP) 是最常用的评价指标^[55]。

本节将介绍用于三维点云配准的评价指标, 主要可以分为两类, 第一类用于特征提取与匹配, 第二类用于刚体运动参数估计。

2.4.1 基于特征提取与匹配的评价指标

(1) 特征匹配召回率 (Feature-Match Recall)

特征匹配召回率是指正确提取出的匹配点对数量和所有点对数量的比率, 衡量的是匹配算法的查全率。从数学上来说, 特征匹配召回率可以表示为:

$$R_{fa} = \frac{1}{n} \sum_{s=1}^n 1 \left(\left[\frac{1}{|\Omega|} \sum_{(i, j) \in \Omega} (\|Tp_i - q_j\| < \tau_1) \right] > \tau_2 \right) \quad (2.53)$$

其中 n 是所有点对的数量, Ω 是一组匹配点对 (p, q) 之间的对应关系, $p = (x_p, y_p, z_p)$, $q = (x_q, y_q, z_q)$, $T \in SE_3$ 是地面真实姿态变换。此外, τ_1 是局内距离阈值, τ_2 是局内召回率阈值。

(2) 配准召回率 (Registration Recall)

配准召回率采用两组具有地面真实姿态变换且存在重叠部分的点云, 衡量有多少存在重叠部分的点云组可以通过匹配算法被正确恢复。具体地说, 配准召回率使用如下误差矩阵来定义真阳性:

$$E = \sqrt{\frac{1}{\Omega^*} \sum_{(p^*, q^*) \in \Omega^*} \|\hat{T}_{i,j} p^* - q^*\|^2} < \tau_3 \quad (2.54)$$

其中 Ω^* 是重叠部分一组匹配点对 (p^*, q^*) 之间的对应关系, $p^* = (x_p^*, y_p^*, z_p^*)$, $q^* = (x_q^*, y_q^*, z_q^*)$, 对于重叠部分, τ_3 是用于判断匹配点对是否正确的阈值。

2.4.2 基于刚体运动参数估计的评价指标

旋转矩阵和平移向量是点云配准任务中最常用的目标参数。最常见的评价指标有均方误差 (MSE)、均方根误差 (RMSE) 以及平均绝对误差 (MAE)，其中 $MSE(R)$ 和 $MSE(t)$ 分别计算的是旋转矩阵和平移向量的真实值和预测值平方误差的平均值，可以表示为：

$$MSE(R) = \frac{1}{n} \sum_{i=1}^n (R_{pre} - R_{gt})^2 \quad (2.55)$$

$$MSE(t) = \frac{1}{n} \sum_{i=1}^n (t_{pre} - t_{gt})^2 \quad (2.56)$$

$RMSE(R)$ 和 $RMSE(t)$ 分别计算的是旋转矩阵和平移向量的真实值和预测值平方根误差的平均值，可以表示为：

$$RMSE(R) = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_{pre} - R_{gt})^2} \quad (2.57)$$

$$RMSE(t) = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_{pre} - t_{gt})^2} \quad (2.58)$$

$MAE(R)$ 和 $MAE(t)$ 分别计算的是旋转矩阵和平移向量的真实值和预测值绝对误差的平均值，可以表示为：

$$MAE(R) = \frac{1}{n} \sum_{i=1}^n |R_{pre} - R_{gt}| \quad (2.59)$$

$$MAE(t) = \frac{1}{n} \sum_{i=1}^n |t_{pre} - t_{gt}| \quad (2.60)$$

这些误差值越小，表示配准的精度越高。理想情况下，如果刚性运动变换是完美的，则上述所有误差度量都应该为 0。

2.5 本章小结

本章首先介绍了点云数据具有的无序性、局部相关性以及不变性这三大特点，其次介绍了点云数据的两类特征描述方式，即局部特征描述和全局特征描述。然后分析了求解刚体运动最优变换参数的两种方法，一是基于 SVD 的线性代数求解法，二是基于 Levenberg-Marquardt 的非线性优化法。接着阐述了点云配准任务中常用的标准数据集，包括 ShapeNet、

ModelNet 以及 S3DIS，并阐述了用于衡量点云配准算法性能的两类评价指标，分别是基于特征提取与匹配的评价指标和基于刚体运动参数估计的评价指标。

第三章 基于几何特征的点云数据预处理

本章设计了一种基于几何特征的点云数据预处理方案，主要可以分为两个部分：点云滤波和点云分割。其中点云滤波包括降采样和离群点去除，点云分割包括地面点数据过滤和无关背景去除，如下图 3.1 所示。通过分析并对比不同的点云滤波方法以及不同的点云分割方法处理点云的效果，本章确定了点云预处理的有效方案，为后续利用有效点云信息进行点云配准奠定了基础。

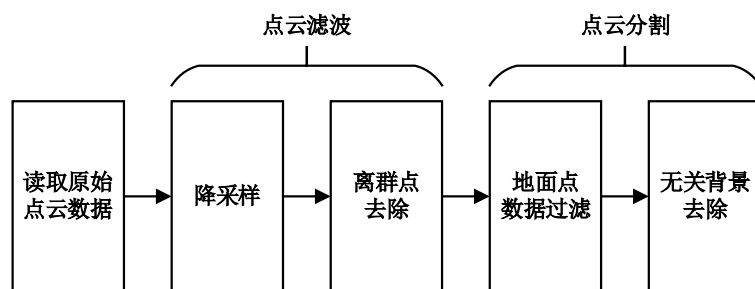


图 3.1 点云预处理方案示意图

3.1 点云滤波

由于外部传感器采集的原始点云数据比较稠密，为避免较大数据量对后续的算法处理造成一定的计算压力，需要对原始点云数据进行降采样，即在不损失点云重要特征的前提下对点云进行删减，从而提高算法执行的效率。

由于设备测量误差的影响，外部传感器获得的数据不可避免地会受到环境噪声的干扰。通常，点云数据的密度分布是不均匀的，并且会产生少量从固有数据中分离出来的离群点。这些问题可能会增加曲面法向量或曲率等局部点云特征的计算复杂度，从而导致后续的算法执行失败^[56]。因此，在点云分割之前，需要对点云数据进行去噪处理。

3.1.1 降采样

(1) 随机降采样

随机降采样的思想比较简单，就是从点云数据中随机选择 k 个点，每个点被选到的概率相同，可以得到指定采样点数量的点云。如原始点云中有 10000 个点，若从中随机采样 5000 个点，那么每个点被选择的概率为 $\frac{1}{2}$ ，这种随机性使得采样后的点云分布也具有随机性。如果原始点云是不均匀的，比如有的区域密度大，有的区域密度小，虽然每个点被选到的概率一样，

但密度大的区域点比较多，这个区域中的点被采样的机会就更多，因此原来密度大的地方，采样后密度还是大，这也就导致随机降采样后的点云同样也是不均匀的。不仅如此，随机降采样由于其随机性可能会过滤掉一些包含重要信息的特征点，影响后续算法的精度。在实际应用中，由于计算机的随机数都是伪随机数，这意味着它们有随机数的表现和属性，但实际上并不是随机生成的，而是由算法生成的。该算法每生成一个随机数序列都有一个初始值，即随机种子。算法可以通过固定随机种子生成固定的随机数序列，以此来保证随机降采样的科学性。同时，随机降采样的时间复杂度很低，仅为 $O(k)$ ，其中 k 为指定采样点的数量，这使得随机降采样在点云预处理中也得到了广泛应用。

(2) 均匀降采样

均匀降采样（Uniform Sampling）有多种不同的采样方式，其中最远点采样^[57]（Farthest Point Sampling, FPS）是常用的一种。假设一组点云数据 $P = \{p_i | i = 1, 2, \dots, n\} \subset R^3$ ，FPS 首先设置一个初始状态为空的内点集合 $N = \emptyset$ ，并在点云中随机选取一个种子点 p_{k_1} 作为起始点并放入集合 $N = \{p_{k_1}\}$ ，其次分别计算剩余的 $n - 1$ 个点到集合 N 中的点 p_{k_1} 的距离，选择距离最大的点 p_{k_2} 放入集合 $N = \{p_{k_1}, p_{k_2}\}$ ，然后计算 $n - 2$ 个点到集合 N 中 p_{k_1} 的距离，保留距离最小的点 p_{k_3} ，并记录此时的距离值 d ，再计算 $n - 2$ 个点到集合 N 中 p_{k_2} 的距离，保留距离最小的点 p'_{k_3} ，并记录此时的距离值 d' ，假设 $d > d'$ ，则将点 p_{k_3} 放入集合 $N = \{p_{k_1}, p_{k_2}, p_{k_3}\}$ ，以此类推，重复上述步骤，直到选取出需要的采样点个数。这个过程如下图 3.2 所示，其中黑色为原始点云数据，红色为内点集合。FPS 这种方式的降采样点云分布均匀，但是算法复杂度较高且效率低。

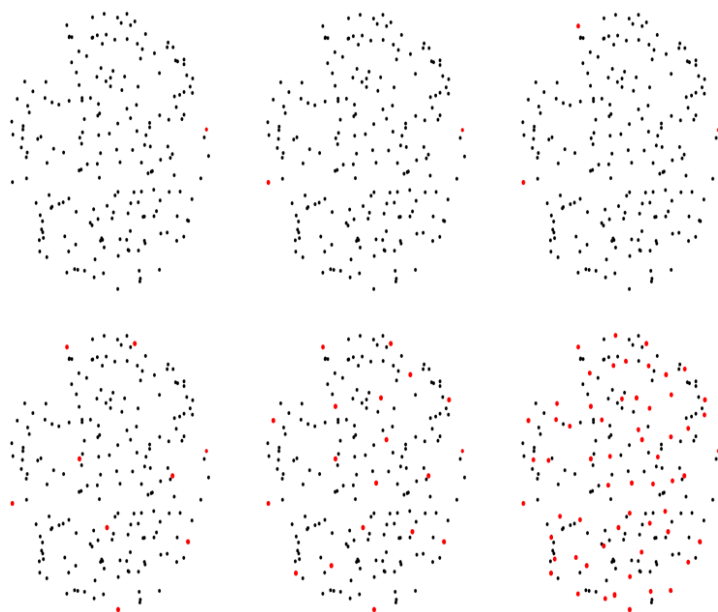


图 3.2 均匀降采样算法过程示意图

(3) 体素降采样

体素降采样 (Voxel Sampling) 的优势在于其能保持点云本身的几何结构。体素滤波器首先根据输入的点云数据 $P = \{p_i | i = 1, 2, \dots, n\} \subset R^3$ 计算出该点云的最小外接立方体, 其次根据设定的分辨率 $\gamma = (r, r, r)$ 建立最小三维体素栅格, 并将最小外接立方体分割成数个三维体素栅格, 此时点云数据填充至对应的三维体素栅格中。然后分别计算每个三维体素栅格中的质心 M , 以此质心的坐标 $M = (x_{center}, y_{center}, z_{center})$ 可以近似表示该三维体素栅格中的所有点。每个三维体素栅格中, 质心的计算公式为:

$$\begin{cases} x_{center} = \frac{\sum_{i=1}^h x_i}{h} \\ y_{center} = \frac{\sum_{i=1}^h y_i}{h} \\ z_{center} = \frac{\sum_{i=1}^h z_i}{h} \end{cases} \quad (3.1)$$

其中 x_i 、 y_i 、 z_i 分别表示三维体素栅格中任一点 p_i 的三维坐标信息, h 为三维体素栅格中点云的数量。

由此可见, 体素降采样无法指定采样后点云的数量, 但是这种方法简洁高效, 并不需要建立复杂的拓扑结构。同时, 体素降采样后的点云分布比较均匀, 并且可以通过调整分辨率来控制分割的体素栅格数量, 从而控制采样后点云的规模大小。分辨率设置得越大, 体素栅格越大, 体素栅格数量越小, 降采样后点的个数则越少。

3.1.2 离群点去除

(1) 统计滤波器

统计滤波器的思想为对每一个点的邻域进行一个统计分析, 计算它到近邻域内所有点的平均距离。假设点云中任一点与其近邻域 k 个点的平均距离满足高斯分布, 根据全局距离平均值和方差定义的标准范围确定距离阈值 σ , 当某个点与其近邻域 k 个点的平均距离大于阈值 σ 时, 则判定该点为离群点并从数据中剔除。

(2) 半径滤波器

半径滤波器通过搜索以查询点为原点, 半径为 r 的邻域, 并统计该邻域内点的数量, 如果数量大于给定阈值, 则将其视为局内点保留, 否则视为局外点剔除。此算法运行速度很快, 但是查询点邻域的半径以及判别是否为局内点的阈值标准都需要人工指定。

3.2 点云分割

点云的有效分割是点云处理的关键环节，往往直接影响后续点云处理的质量，并为后续的点云配准提供了重要的前提条件。目前比较常用的点云分割算法是鲁棒性较好的聚类分割算法和基于随机采样一致性的分割算法。

在传感器采集的点云数据中，绝大部分数据属于地面点数据，这对后续点云分割算法会带来干扰，因此需要先进行地面点数据过滤。同时，考虑到非理想状态，周围环境可能存在非目标物体的杂物，需要通过点云分割算法对点云数据进行聚类，从而实现无关背景去除。由于真实场景较小，采集的点云数据滤波之后数量较少，直接将地面作为一个分割的类别更为高效。因此，本节并没有进行地面点数据过滤，而是直接使用了欧式聚类和区域生长这两类点云分割算法对滤波后的点云进行分割，从而分离出目标物体点云。

3.2.1 欧式聚类算法

欧式聚类^[58]是一种简单的聚类方法。该算法首先在整个点云上构建 k -d tree，然后将近邻域内的所有点组成一个实例。

k -d 树 (k -dimensional tree, k -d tree) 是一种高效的数据结构，用于组织多维数据，可以快速搜索最近邻，是点云配准任务中广泛需要的一种基本操作。 k -d tree 是一种轻量级的二叉树，其原理是用一个垂直于坐标轴的超平面连续划分 k 维空间，可以得到多个超矩形区域。 k -d tree 首先是一个包含所有 k 维空间的根节点，然后通过垂直于坐标轴的超平面获得两个左右节点。同时，将 k 维空间划分为两个区域，重复上述过程，直到左、右节点不再有样本，即可得到一个 k -d tree。

在整个 k -d tree 中，每个节点为 T_r ，左侧和右侧的子节点分别为 T_{rl} 和 T_{rr} ，可以得到：

$$T_r = T_{rl} \cup T_{rr} \quad (3.2)$$

$$T_{rl} \cap T_{rr} = \emptyset \quad (3.3)$$

当 T_{rl} 和 T_{rr} 都为空时，节点 T_r 为叶节点。

欧式聚类的算法步骤大致如下：

(1) 随机定义空间中任意一点 p_{11} 为查询点，通过 k -d tree 算法找到 p_{11} 的近邻域内的 k 个点，判断这 k 个点到查询点 p_{11} 的距离，并将距离小于阈值的点 p_{12} , p_{13} , ...放在点集 $C = \{p_{11}, p_{12}, p_{13}, \dots\}$ 中。

(2) 以集合 C 中除 p_{11} 以外的任意一点作为查询点，假设为点 p_{12} ，重复步骤 1，得到点集

$C = \{p_{11}, p_{12}, p_{13}, \dots, p_{22}, p_{23}, p_{24} \dots\}$ 。

(3) 当点集 C 中所有的点都已被作为查询点, 且没有满足条件的新点可以加入点集 C 时, 即完成类别 I 点云的搜索与分割。

(4) 随机定义空间中除类别 I 点云外的任意一点为查询点, 重复步骤 1~3, 完成类别 II 点云的搜索与分割。

(5) 以此类推, 当原始点云中所有的点都被搜索与分割完毕时, 算法终止。

3.2.2 区域生长算法

区域生长算法^[59]的基本思路是把具有相同特性的点结合在一起, 首先对每一区域确定一个种子点为生长的起点, 然后将种子点周围邻域的点与种子点进行特征比较, 并将与种子点具备相同特性的点结合在一起不断地往外生长, 直至没有满足条件的点便停止生长。区域生长的过程中主要有以下三个要点, 一是种子点的选取, 二是区域生长的生长准则, 三是生长的停止条件。

(1) 种子点的选取

种子点的选取是否合理直接影响分割结果, 通常利用法线、曲率等三维几何特征作为选取标准。通过计算点云中每个点的平均曲率和高斯曲率, 并选择点云数据中曲率最小的点作为初始种子点开始生长, 即在最平坦的区域选择种子点, 可以减少总线段数, 避免重叠分割, 提高分割的稳定性。选择种子点的计算步骤如下:

将曲面 $r(u, v)$ 的偏导数 $\frac{\partial r}{\partial u}$ 、 $\frac{\partial r}{\partial v}$ 、 $\frac{\partial^2 r}{\partial u \partial u}$ 、 $\frac{\partial^2 r}{\partial u \partial v}$ 、 $\frac{\partial^2 r}{\partial v \partial v}$ 记为 r_u 、 r_v 、 r_{uu} 、 r_{uv} 、 r_{vv} , 那么曲线的单位法向量 \bar{N} 可以表示为:

$$\bar{N} = \frac{r_u \times r_v}{\|r_u \times r_v\|} \quad (3.4)$$

平均曲率和高斯曲率是分析三维曲面的两个重要几何特征, 将两者结合起来选取最小曲率点, 可以更好地提取点云曲面的局部特征。曲面上点 p 的平均曲率 Z 和高斯曲率 H 的计算公式如下:

$$Z = \frac{LN - M^2}{EG - F^2} \quad (3.5)$$

$$H = \frac{EN + GL - 2FM}{2(EG - F^2)} \quad (3.6)$$

其中各变量可以由以下公式得到:

$$E = r_u \times r_u \quad (3.7)$$

$$F = r_u \times r_v \quad (3.8)$$

$$G = r_v \times r_v \quad (3.9)$$

$$L = r_{uu} \times N \quad (3.10)$$

$$M = r_{uv} \times N \quad (3.11)$$

$$N = r_v \times \bar{N} \quad (3.12)$$

计算曲面上点的最大曲率和最小曲率：

$$K_{1,2} = H \pm \sqrt{H^2 - K} \quad (3.13)$$

其中， K_1 表示该点所在曲面在所有方向上的最大曲率， K_2 表示该点所在曲面在所有方向上的最小曲率， K_1 和 K_2 可以为后续的曲率阈值设置提供参考。

(2) 区域生长的生长准则

区域生长方法用最小二乘法拟合估计点的切平面法向量，即分析协方差矩阵的特征向量和特征值。对于曲面上的一个点 p_i ，该点的法向量可以是协方差矩阵 C 的最小特征值相对应的特征向量。协方差矩阵 C 的计算公式如下：

$$C = \frac{1}{k} \sum_{i=1}^k (p_i - \bar{p}) \cdot (p_i - \bar{p})^T \quad (3.14)$$

其中 k 是点 p_i 邻域内点的个数， \bar{p} 是 p_i 邻域内所有点的质心。计算当前种子点的法向量与其邻域内所有点的法向量之间的夹角：

$$\alpha = \arccos \frac{\vec{n}_p \cdot \vec{n}_i}{|\vec{n}_p| |\vec{n}_i|} \quad (3.15)$$

其中 \vec{n}_p 表示当前种子点的法向量， \vec{n}_i 表示其邻域内任一点的法向量。当夹角在阈值范围内时，则将该点添加到种子点。

阈值的设置也是区域生长中一个非常重要的步骤，甚至直接影响分割结果。首先需要根据点云数据设置点云聚类的最小聚类点数和最大聚类点数，最大聚类点数一般选择远大于点云数量的值，这使得分割后的点数小于最小聚类点数。其次设置种子点的法向量与其邻域内任一点法向量的夹角偏差允许范围，如果两条法向量之间的夹角在阈值范围内，则将这两个点归为同一类。最后根据主曲率的最小值和最大值，调整点云的曲率阈值。如果曲面上点的曲率小于阈值则也将该点加入种子点集。合理的曲率阈值设置可以在一定程度上避免点云数据的欠分割和过分割。

(3) 生长的停止条件。

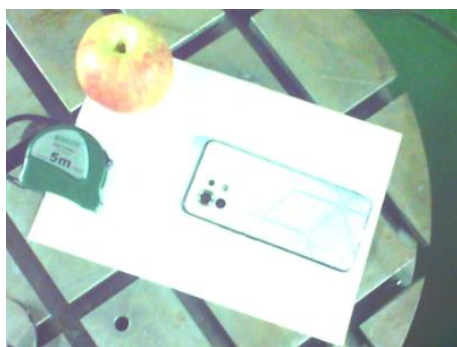
通常，当没有点满足法向量和曲率两个几何特征属性的约束时，即没有点能够被加入某个区域时，区域生长就会停止。

3.3 实验结果与对比分析

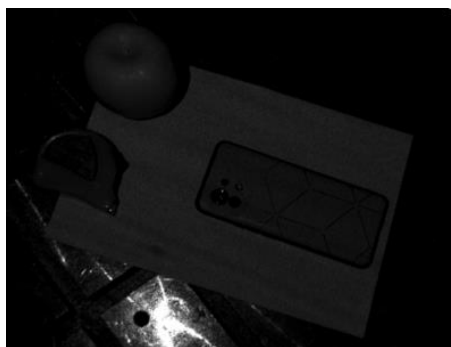
本节主要在真实场景中测试了基于几何特征的点云数据预处理方案，首先使用外部传感器采集稠密点云数据，其次分析并对比了随机降采样、均匀降采样以及体素降采样三类降采样方法对原始点云的处理效果，然后分析并对比了使用不同参数的统计滤波器以及半径滤波器对降采样后的点云数据进行离群点去除的效果，之后分析并对比了欧式聚类算法和区域生长算法对点云进行分割的效果，最后确定了最优预处理方案。

3.3.1 实验环境与硬件配置

本节对真实场景进行点云数据采集与点云数据预处理。在真实场景中，本节将手机作为目标物体放置于工作台，同时在手机周围放置了苹果、卷尺等杂物作为无关背景。通过手动操控深度相机对真实场景进行扫描，获得了目标物体及其周围环境的点云数据，采集的原始点云共有 271114 个点，可视化结果如下图所示。

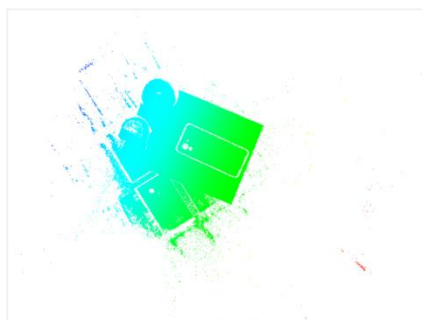


(a) RGB 图像

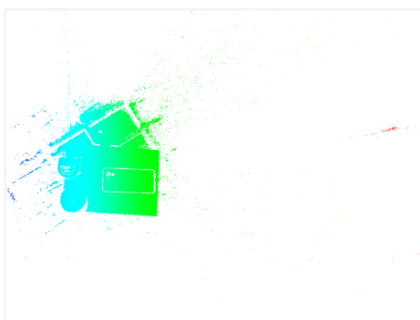


(b) 灰度图像

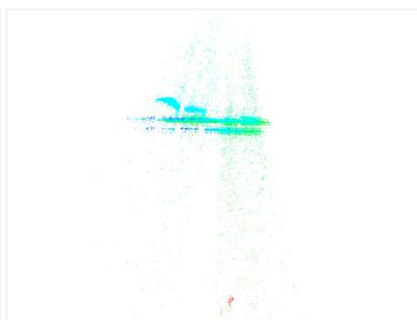
图 3.3 真实场景二维图像



(a) 正面



(b) 背面



(c) 侧面

图 3.4 真实场景点云可视化效果图

硬件配置包括 Intel® Core™ i9-9900k CPU@3.60GHz 和两块 NVIDIA RTX 2080 Ti GPU，内存为 32GB，总显存为 22GB。本节在 Ubuntu18.04 环境下运行，采用 PCL 1.8.0 三维点云处

理跨平台开源库对真实场景采集的点云数据进行预处理，并利用 CloudCompare 三维点云编辑和处理软件可视化点云处理的效果。

3.3.2 对比实验分析

(1) 降采样不同方法效果对比

为了比较不同降采样方法处理点云的效果，本节对原始点云数据分别进行了随机降采样、均匀降采样和体素降采样。通过调整均匀降采样近邻搜索半径的大小以及体素降采样中栅格的分辨率，保证均匀降采样和体素降采样后点云的个数一致，均为 48977。之后指定采样个数为 48977 对原始点云进行随机降采样。实验效果与数据如下图 3.5 与下表 3.1 所示。

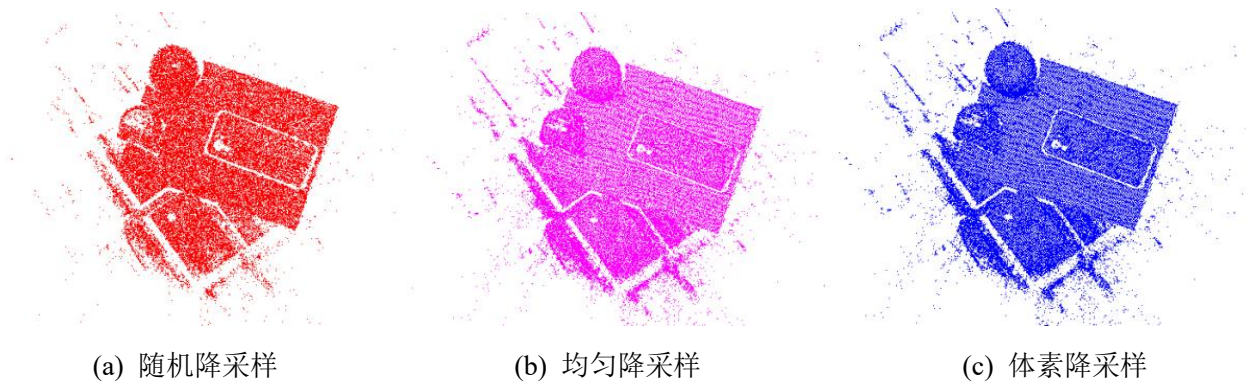


图 3.5 三种降采样结果对比图

从可视化结果可以看出，均匀降采样和体素降采样基本保持了原始点云本身的几何结构，随机降采样并不能保持原始点云本身的几何结构。在外部传感器采集过程，由于受到光照、光强等外界环境的影响，原始点云数据中除了目标物体及其周围杂物外，还存在很多干扰噪声。这些干扰噪声在进行降采样时也被认为是点云本身的几何结构，因此在对原始点云分别进行均匀降采样和体素降采样之后仍保留了大量干扰噪声。在对原始点云进行随机降采样时，对于目标物体及其周围杂物这样密度大的区域，其中的点被采样到的概率更大，而对于干扰噪声这样密度小的区域，其中的点被采样到的概率更小，反而使得随机降采样后的点云数据过滤了绝大多数干扰噪声，也能保证目标物体及其周围杂物的重要信息。

表 3.1 降采样实验结果对比表

方法	原始点云数量（个）	降采样后点云数量（个）	用时（s）
随机降采样	271114	48977	0.104379
均匀降采样			0.122201
体素降采样			0.155843

从表格数据可以看出, 同样从 271114 个点降采样到 48977 个点, 随机降采样用时最短, 仅为 0.104379s, 比均匀降采样和体素降采样分别快了 15%和 33%。

(2) 离群点去除不同方法和参数效果对比

为了比较不同的离群点去除方法和不同的参数处理点云的效果, 本节分别使用统计滤波器和半径滤波器对降采样后的点云进行去噪。当使用统计滤波器时, 通过调整近邻搜索点的个数 k_1 以寻求最优参数, 获得最佳效果。当使用半径滤波器时, 通过调整近邻搜索半径的大小 r_1 和近邻点集数阈值的大小 k_2 以寻求最优参数, 获得最佳效果。实验效果与数据如下图 3.6 和 3.7 以及下表 3.2 所示。

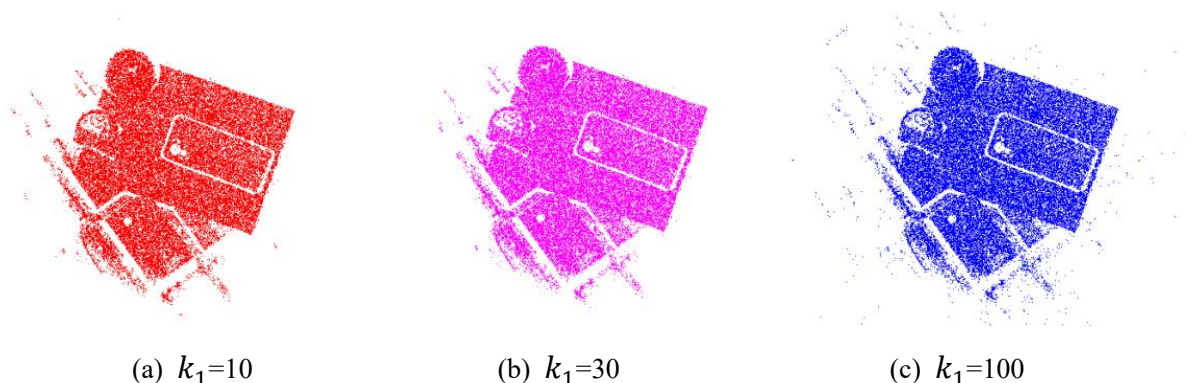


图 3.6 统计滤波器不同参数效果对比图

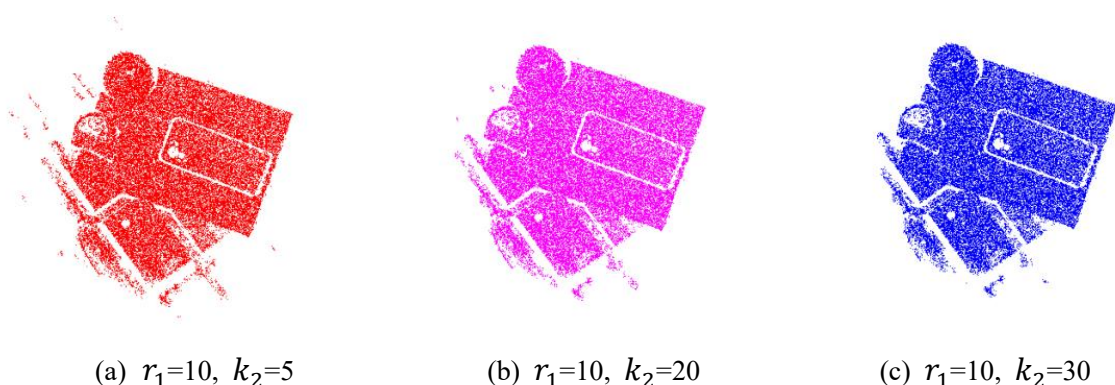


图 3.7 半径滤波器不同参数效果对比图

从可视化结果可以看出, 在使用统计滤波器进行离群点去除时, 当近邻搜索点的个数设为 30 时, 效果最好。当近邻搜索点的个数设为 10 时, 近邻域太小, 有一部分噪声密集区域的查询点到近邻点集的平均距离恰好在标准范围内, 故被认为是局内点, 从而导致一部分噪声没有从数据中剔除。当近邻搜索点的个数设为 100 时, 近邻域内的噪声点和非噪声的分布情况和原始点云类似, 噪声区域查询点到近邻点集的平均距离恰好与全局平均距离相当, 故被认为是局内点, 从而导致大量的噪声没有从数据中剔除。在使用半径滤波器进行离群点去除时, 当近邻搜索半径的大小设为 10, 查询点的近邻点集数阈值设为 30 时, 效果最好。当查询点的近邻点集数阈值设为 5 时, 阈值太小, 噪声区域查询点的近邻域内点的个数远大于 5,

故被认为是局内点，从而导致大部分噪声没有从数据中剔除。当查询点的近邻点集数阈值设为 20 时，有一部分噪声区域查询点的近邻域内点的个数小于 20 被认为是离群点从数据中剔除，但仍有一部分噪声区域查询点的近邻域内点的个数大于 20 被认为是局内点，无法从数据中剔除。当使用半径滤波器进行离群点去除时，需要指定近邻搜索半径的大小以及近邻点集数阈值的大小，且它们两者之间并不独立，相互影响，需要通过先验知识或反复调整来获得最优参数。而统计滤波器根据全局距离平均值和方差来定义阈值标准，相较而言更为高效。

表 3.2 离群点去除实验结果对比表

方法	参数	降采样后 点云数量（个）	离群点去除后 点云数量（个）	用时（s）
统计滤波	$k_1=10$	48977	48177	0.104379
	$k_1=30$		48105	0.122201
	$k_1=100$		48189	0.155843
半径滤波	$r_1=10, k_2=5$		47994	0.099722
	$r_1=10, k_2=20$		47362	0.135366
	$r_1=10, k_2=30$		46889	0.161095

从表格数据可以看出，当使用统计滤波器进行离群点去除时，近邻搜索点的个数与去除离群点后点的数量无线性关系，但与用时存在线性关系，即近邻搜索点的个数越大用时越久。其中效果最好的实验，点云个数从 48977 减少为 48105，用时为 0.122201s。当使用半径滤波器进行离群点去除时，近邻点集数阈值与去除离群点后点的数量以及用时都存在线性关系，即近邻点集数阈值越大，去除离群点后点的数量越少，用时也越久。其中效果最好的实验，点云个数从 48977 减少为 46889，用时为 0.161095s。

（3）背景去除不同方法效果对比

为了比较不同的背景去除效果，本节分别使用了欧式聚类算法和区域生长算法对点云进行分割。当使用欧式聚类算法时，将近邻搜索半径的大小设为 5 可以获得较好的分割效果。当使用区域生长算法时，将近邻搜索点的个数设为 50，法向量角度阈值设为 $\frac{5}{180}\pi$ ，曲率阈值设为 1.0，可以获得较好的分割效果。实验效果与数据如下图 3.8 与下表 3.3 所示。

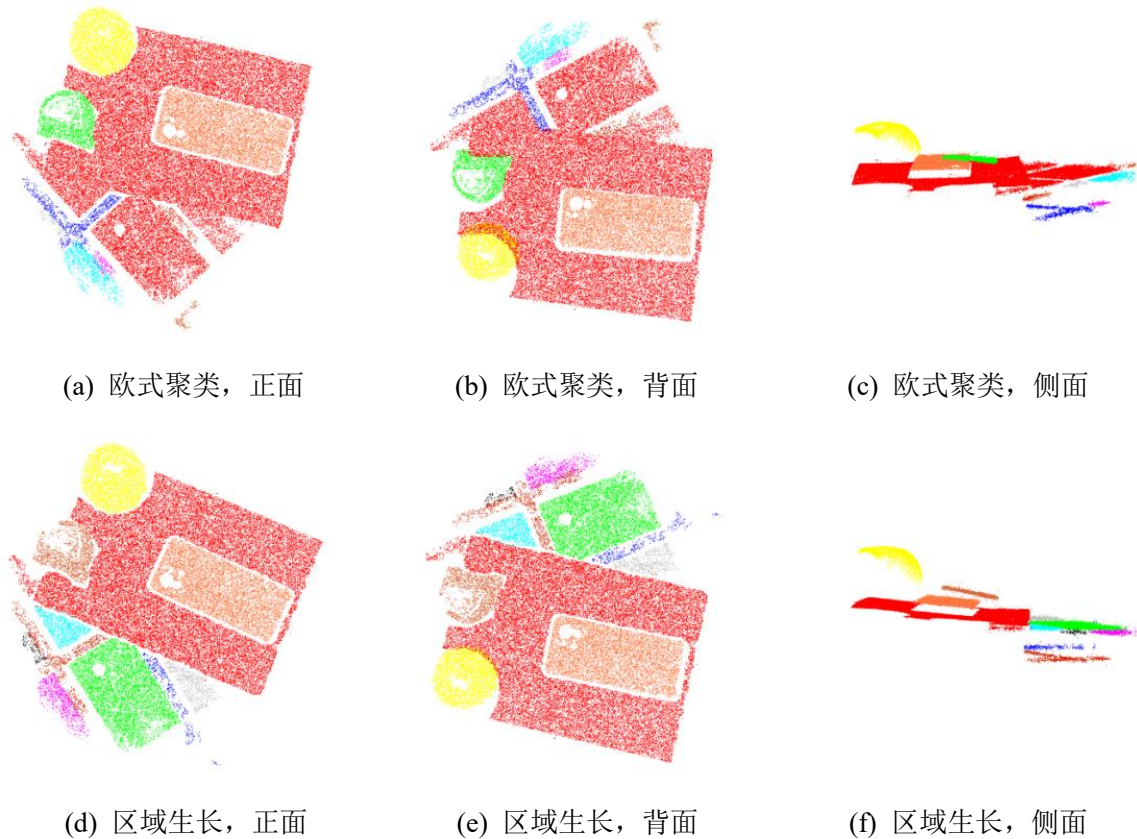


图 3.8 点云分割效果对比图

从可视化结果可以看出，当使用欧式聚类算法对离群点去除后的点云做无关背景去除时，大部分属于工作台类别的点云都能被认为是地面点聚为一类，且目标物体（手机）以及杂物（苹果、卷尺）都能被聚类正确。当使用区域生长算法对离群点去除后的点云做无关背景去除时，虽然有大部分属于工件台类别的点云被认为是地面点，但仍有一部分属于工件台类别的点云被认为是其它类别。但目标物体（手机）以及杂物（苹果、卷尺）都能被聚类正确。这是因为真实场景的工作台不是一个完整的平面，存在很多凹槽，区域生长算法中的法向量和曲率通过最小二乘法对近邻域点集拟合平面求得，凹槽处在没有点云的情况下无法计算出精确的法向量和曲率。而欧式聚类计算的是查询点与近邻域内所有点的平均距离，由于凹槽间距并没有很大，取平均之后，这一部分带来的影响可以有所减少。

表 3.3 点云分割实验结果对比表

方法	离群点去除后 点云数量（个）	点云分割 类别（类）	无关背景去除后 点云数量（个）	用时（s）
欧式聚类	48105	10	6067	0.295816
区域生长		12	6001	0.499993

从表格数据可以看出，当使用欧式聚类算法分割点云时，点云被分为 10 个类别，且无关背景去除后点云数量为 6067，即目标物体（手机）的点云数量为 6067，用时 0.295816s。当使用区域生长算法分割点云时，点云被分为 12 个类别，且无关背景去除后点云数量为 6001，

即目标物体（手机）的点云数量为 6061，用时 0.499993s。从用时来看，由于欧式聚类算法以邻域内点与点之间的距离作为判定标准，而区域生长算法利用了法向量和曲率这两种几何特征作为判定标准，从算法复杂度来说欧式聚类远低于区域生长算法，因此处理时长也就更短。

（4）点云预处理方案

综上所述，本章确定了点云预处理的有效方案，实验数据如下表 3.4 所示。首先使用随机降采样这种方法来降低原始点云的冗余性，并指定采样数量 m 为 48977。然后使用统计滤波器进行离群点去除，并取近邻搜索点的个数 k_1 为 30。最后使用欧式聚类算法进行无关背景去除，并取近邻搜索半径的大小 r_2 为 5。测试下来，降采样后，点云数据从 271114 个点减少到 48977 个点，用时为 0.104379s，且在保证点云结构与重要特征的同时过滤掉了绝大多数干扰噪声。离群点去除后，点云个数又从 48977 减少为 48105，用时为 0.122201s，且去除了绝大多数离群点。在进行无关背景去除时，点云被分为 10 个类别，分离出目标物体（手机）后，点云个数从 48105 减少为 6067，即目标物体（手机）的点云数量为 6067，用时为 0.295816s。总用时 0.522397s。

表 3.4 点云预处理实验结果汇总表

方法	参数	处理前点云数量（个）	处理后点云数量（个）	用时（s）
随机降采样	$m=48977$	271114	48977	0.104379
统计滤波器	$k_1=30$	48977	48105	0.122201
欧式聚类	$r_2=30$	48105	6067	0.295816
总用时（s）	0.522397			

3.4 本章小结

本章设计了一种基于几何特征的点云数据预处理方案，主要包括点云滤波和点云分割两个部分。通过分析和对比不同方法以及不同参数的算法，本章确定了点云配准的有效方案，即使用随机降采样对采集到的原始点云数据完成指定数量的采集，再使用统计滤波器对降采样后的点云进行离群点去除，最后使用欧式聚类算法对离群点去除后的点云进行分割，可以较好地分离出目标物体，为后续利用有效点云信息进行点云配准奠定基础。

第四章 基于位置自适应卷积提取特征的点云配准优化

本章基于 PCRNet 深度学习网络模型提出一种基于位置自适应卷积提取特征的点云配准优化方法。首先阐述了 PCRNet 网络的工作机制，并分析了 PCRNet 网络在特征提取方面存在的局限性。然后针对这些局限性对原网络提出了改进，构建了 PACNet 网络模型。最后通过实验将本章提出的网络模型与当前主流的点云配准算法进行了多项测试对比，证明了 PCRNet 网络具有更高的精度。

4.1 基础网络 PCRNet

本节阐述了 PCRNet 网络中用于提取点云全局信息的 PointNet 框架，分析了搬土距离^[60]（Earth Mover Distance, EMD）损失函数的原理，介绍了 PCRNet 网络模型的整体框架，并对其存在的局限性做出了分析。

4.1.1 PointNet

在点云处理任务中，被广泛应用的共享权值的多层感知机^[61]（Multi-Layer Perceptron, MLP）可以将输入点云的所有点映射到一个冗余的高维空间，目的是减少最大池化层带来的特征损失。其本质是用 1×1 大小的卷积核来做卷积操作，可以减少大量的参数。MLP 的结构可以简单看成：

$$Output = Input \times Weight + Bias \quad (4.1)$$

其中 $Input$ 、 $Weight$ 、 $Bias$ 、 $Output$ 的尺寸分别为 $N \times C_1$ 、 $C_1 \times C_2$ 、 $C_2 \times 1$ 、 $N \times C_2$ ，其能够将输入的 C_1 维特征转换成 C_2 维特征，这 C_2 维中每一维都整合了原来 C_1 维的全部信息。

PointNet 是首个可以直接处理点云数据的深度学习模型架构，输入对象是含有 n 个点的三维点云，并使用 T-Net 网络预测三维空间刚性运动变换矩阵，将其作用于原始点云以完成数据的对齐。对齐后的数据以点为基本单元，采用 MLP 对每个点进行独立建模，从而完成特征提取。然后使用对称函数聚合全局特征，即在特征空间的维度上进行最大池化(max pooling)，提取出点云的全局特征向量。PointNet 的轻量级架构如图 4.1 所示，其中 n 表示输入点的数量， M 表示每个点学习特征的维度。

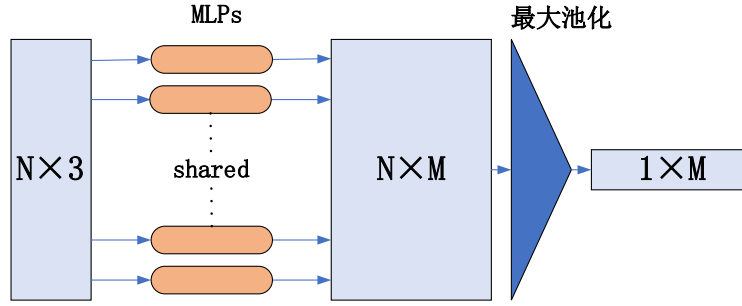


图 4.1 PointNet 的轻量级架构示意图

4.1.2 EMD 损失函数

搬土距离（EMD）是一种距离度量的定义，可以用来评估某一特征空间中两个多维分布之间的差异性^[60]。

EMD 实际上是线性规划中运输问题的最优解。假设有若干个大小不一且分布不同的土堆，同样还有若干个大小不一且分布不同的土坑。每对土堆-土坑的运输成本用距离表示，EMD 度量的就是用土堆填满土坑所消耗的最小总运输成本。

如下图 4.2 所示，土堆和土坑分别用两个直方图来表示，两个直方图分别被描述为 $P = \{(p_1, w_{p1}), (p_2, w_{p2}), \dots, (p_m, w_{pm})\}$ 和 $Q = \{(q_1, w_{q1}), (q_2, w_{q2}), \dots, (q_n, w_{qn})\}$ ，其中 p_i 和 q_j 表示直方的位置，即土堆和土坑的分布位置， w_{p_i} 和 w_{q_j} 表示相应直方的高度，即土堆和土坑的大小。 $F = [f_{ij}]$ 表示从 p_i 到 q_j 的规划运输量。 $D = [d_{ij}]$ 表示 p_i 和 q_j 之间的运输距离。将土堆填满土坑有很多种运输的方式，EMD 的目标是使得总运输成本最小化，这个过程可以表示为：

$$WORK(P, Q, F) = \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij} \quad (4.2)$$

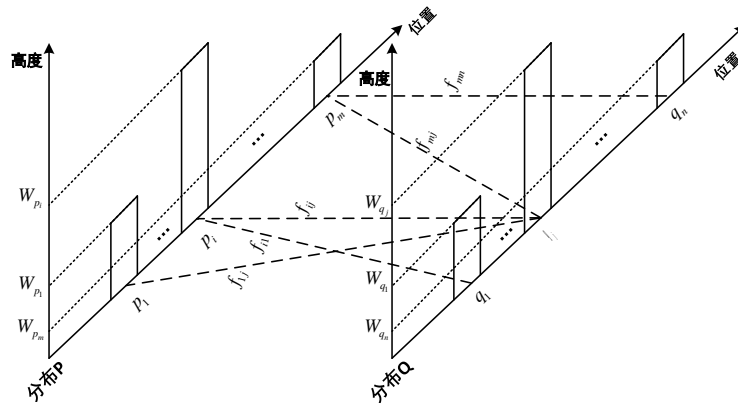


图 4.2 EMD 原理示意图

上式受以下约束：

$$f_{ij} \geq 0 \quad 1 \leq i \leq m, 1 \leq j \leq n \quad (4.3)$$

$$\sum_{j=1}^n f_{ij} \leq w_{p_i} \quad 1 \leq i \leq m \quad (4.4)$$

$$\sum_{i=1}^m f_{ij} \leq w_{q_j} \quad 1 \leq j \leq n \quad (4.5)$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} \quad (4.6)$$

其中公式 4.3 表示从 p_i 到 q_j 的运输量大于零。公式 4.3 和公式 4.5 表示从 p_i 到 q_j 运输总量不能大于土堆的总大小，也不能大于土坑的总大小。公式 4.5 表示总运输量不能大于土堆的总大小，也不能大于土坑的总大小。当运输问题得到解决并求解得到使得总运输成本最小的调度方式 $F^* = \operatorname{argmax}_F \operatorname{WORK}(P, Q, F)$ 时，EMD 可以定义为总运输成本归一化后的表达：

$$\operatorname{EMD}(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (4.7)$$

归一化的目的是为了消除在部分匹配的情况下受到奇异样本数据的影响。

当度量点云距离时，可以将 $P = \{(p_1, 1), (p_2, 1), \dots, (p_m, 1)\}$ 表示一组点云数据，其中 p_i 表示索引为 i 的点云坐标， w_{p_i} 表示每个点的权重，这里都用 1 来表示。 $Q = \{(q_1, 1), (q_2, 1), \dots, (q_m, 1)\}$ 表示另一组点云数据，其中 q_j 表示索引为 j 的点云坐标，同样 w_{q_j} 表示每个点的权重，这里都用 1 来表示。 $D = [d_{ij}]$ 表示从 P 第 i 个点到 Q 第 j 个点的距离，这里用欧氏距离表示。 $F = [f_{ij}]$ 表示是否将 P 第 i 个点移动到 Q 第 j 个点，因此 f_{ij} 的值为 0 或 1。此时， $\operatorname{WORK}(P, Q, F)$ 就表示将 P 中所有点移动到 Q 中所有点位置的总移动成本， $F^* = \operatorname{argmax}_F \operatorname{WORK}(P, Q, F)$ 就表示使得总移动成本最小的调度方式。

为了最小化模板点云和源点云对应点之间的距离，PCRNNet 通过 EMD 损失函数在网络训练过程中实现反向传播，这个过程可以表示为：

$$\operatorname{EMD}(P_S^{\text{est}}, P_T) = \min_{\psi: P_S^{\text{est}} \rightarrow P_T} \frac{1}{|P_S^{\text{est}}|} \sum_{x \in P_S^{\text{est}}} \|x - \psi(x)\|_2 \quad (4.8)$$

其中， P_T 表示模板点云， P_S^{est} 表示将网络预测得到的位姿变换作用于原始源点云后的变换源点云， $\psi(\cdot)$ 表示源点云到模板点云的映射关系。

4.1.3 PCRNNet 网络设计

PCRNNet 网络模型的整体框架如图 4.3 所示。该网络模型的总体结构由特征提取模块、特征融合模块以及位姿回归模块组成。

在特征提取模块，该模型使用 PointNet 分别对输入的模板点云和源点云提取全局信息，即使用 5 层 MLP 对输入的点云进行特征抽象与提取，然后经过 1 层最大池化对抽象到高维

空间后的点云特征进行降维操作，从而保证输入点云的顺序保持不变。在特征融合模块，将两个全局信息以深度学习中的 `concatenate`^[62] 数组拼接操作进行连接，即将 2 个 1×1024 维的全局特征拼接成 1 个 1×2048 维的特征。在位姿回归模块，该模型使用不同维度的全连接层预测相对位姿，获得旋转量和平移量。全连接层共有 5 个隐藏层，维度分别为 1024、1024、512、512、256，输出层的维度为 7，即四元数和平移向量的组合。

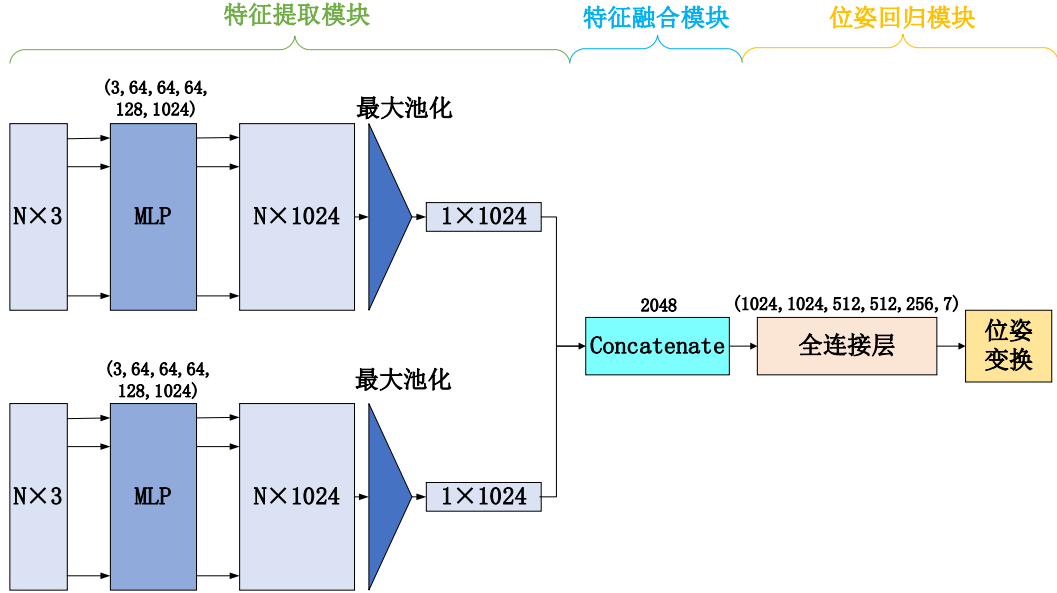


图 4.3 PCRNet 网络模型整体框架示意图

由此可见，PCRNet 的网络模型结构十分简单。由于并没有考虑到点云的局部特征，因此相比于其他点云配准方法，PCRNet 的配准速度非常快，但同时精度上也有所妥协。为了提高点云配准的精度，PCRNet 网络的作者 Sarode 又通过迭代方案改进了基础网络，从而提出了迭代的 PCRNet。如图 4.4 所示，迭代的 PCRNet 将一次 PCRNet 预测得到的坐标变换 $T(1)$ 作用于输入的源点云，再将原始模板点云和变换一次后的源点云作为下一次 PCRNet 的输入。以此类推，在执行 n 次迭代时，将原始模板点云和变换 $n - 1$ 次后的源点云作为输入，通过 PCRNet 预测得到坐标变换 $T(n)$ 。通过组合每次预测得到的坐标变换，获得原始模板点云和源点云之间的整体变换。这个过程可以表示为：

$$T = T(n) \times T(n - 1) \times \cdots \times T(1) \quad (4.9)$$

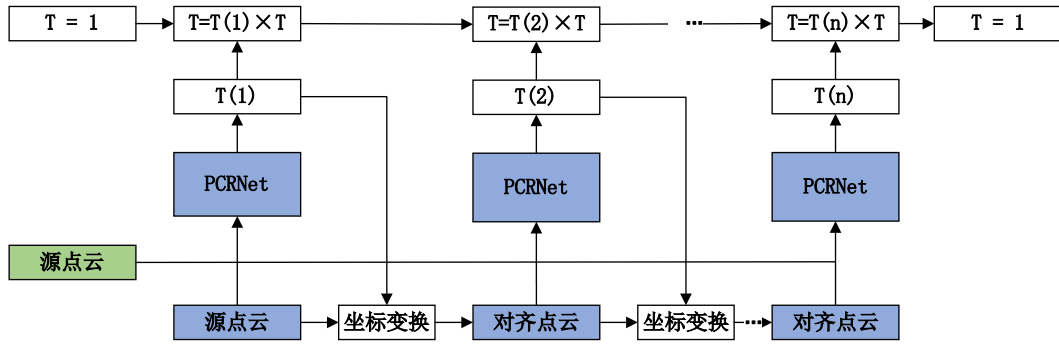


图 4.4 PCRNet 网络预测坐标变换示意图

为了提高迭代的效率，迭代的 PCRNet 仅使用 3 层全连接网络预测相对位姿。与 PCRNet 相比，即使使用了更少数量的隐藏层，仍然能获得更稳健的性能。但由于 PCRNet 网络在特征提取模块中仅关注了点云的全局信息，忽略了点云的局部特征，难以表征点云空间复杂的变化关系，一定程度上限制了其对场景的理解能力，所以在配准精度上有所欠缺。

4.2 基于位置自适应卷积提取特征的 PACNet

为了解决上述问题，本节引入位置自适应卷积，基于 PCRNet 网络在特征提取模块提出了改进，构建了 PACNet 点云配准网络模型，该网络可以灵活地对点云的局部区域进行建模。

4.2.1 位置自适应卷积

如图 4.5 所示，位置自适应卷积部分首先初始化了一个由 K 个大小为 $C_{in} \times C_{out}$ 的权重矩阵组成的权重库 $W = \{W_k | k = 1, 2, \dots, K\}$ ，其中 C_{in} 表示网络在当前层的输入维度， C_{out} 表示网络在当前层的输出维度。较大的 K 可以保证卷积内核的多样性，但也会增加模型的负担。因此，在本章的网络模型中， K 取值为 $16^{[63]}$ 。接下来计算输入点云中每一个点 p_i 与其邻域点 p_j 的相对位置关系，并学习不同位置的权重系数 $E_{ij} = \{E_{k_{ij}} | k = 1, 2, \dots, K\}$ ，这个过程可以表示为：

$$E_{ij} = \text{Softmax}(\theta(p_i, p_j)) \quad (4.10)$$

其中 θ 是核大小为 1×1 的卷积实现的非线性函数。使用 Softmax 对其进行规范化操作可以确保输出分数在范围 $(0, 1)$ 之间^[64]。较高的分数意味着对应的位置具有更重要的局部信息。

位置自适应卷积的内核是通过组合权重矩阵 W_m 和从点位置学习到的权重系数 $E_{k_{ij}}$ 构成的，可以表示为：

$$\mathcal{K}(p_i, p_j) = \sum_k^K E_{kij} W_k \quad (4.11)$$

至此，位置自适应卷积完成了动态构建卷积内核的工作^[65]。它可以灵活地捕获输入特征局部区域的信息，输出具有局部关联性的特征。

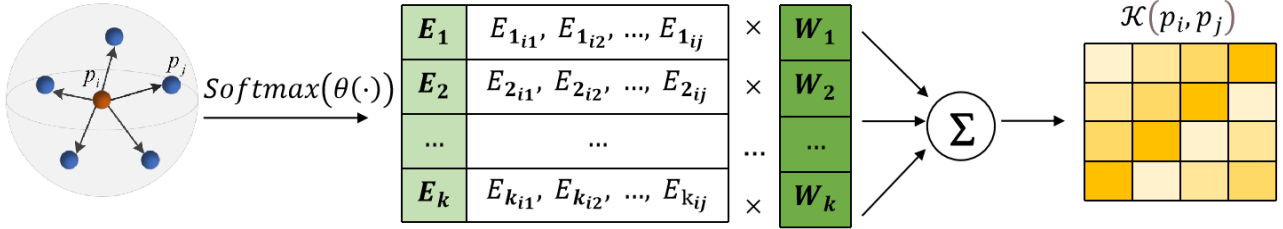


图 4.5 位置自适应卷积原理示意图

由于权重矩阵是随机初始化的并且可能会收敛的彼此非常相似，在其他网络学习任务中，位置自适应卷积往往会对数据进行权重正则化^[66]来惩罚不同权重矩阵之间的相关性。但在点云配准任务中，其目标是将源点云与模板点云对齐，当网络逐渐收敛，权重矩阵彼此之间也将更相似，这是必然的也是必需的。因此本章取消了位置自适应卷积对权重的正则化，使其能更好地适用于点云配准任务。

4.2.2 PACNet 网络设计

通过引入位置自适应卷积，本章基于 PCRNet 网络在特征提取模块提出了改进，构建了 PACNet 点云配准网络模型，网络模型整体框架如图 4.6 所示。使用 $P_T = \{P_{T_i} | i = 1, 2, \dots, n\} \subset \mathbb{R}^3$ 来表示输入的模板点云，其中每个点 P_{T_i} 都包含了坐标 (x, y, z) 以及强度、颜色等信息，并初始化一个旋转矩阵 $R_0 \in SO(3)$ 和平移向量 $t_0 \in \mathbb{R}^3$ ，将其作用于模板点云 P_T 生成源点云 $P_S = \{P_{S_j} | j = 1, 2, \dots, n\} \subset \mathbb{R}^3$ ，该过程可以由公式 (4.12) 表示：

$$P_S = R_0 P_T + t_0 \quad (4.12)$$

至此得到了待配准的点云对。

在特征提取模块，首先通过 2 层 MLP 将 n 个输入的点云由 3 维变为 64 维，再通过 2 层位置自适应卷积对 64 维的高维全局特征进行特征学习，并输出 128 维的总结了关键点局部关联性和全局信息的高维特征，再通过 1 层 MLP 将 128 维的特征变为 1024 维，最后通过最大池化，将 n 个 1024 维的特征整合为一个 1 个 1024 维的特征，同时，这一步也可以消除点云数据顺序的影响。

特征融合模块和 PCRNet 网络模型相同，直接将两个点云的特征合并为 1 个 2048 维的特

征。假设 P_T 和 P_S 有 n 对对应关系，配准的目标是找到最能将源点云 P_S 对齐到模板点云 P_T 的刚性变换矩阵，包括旋转矩阵 $R \in SO(3)$ 和平移向量 $t \in \mathbb{R}^3$ 。这个过程可以表示为：

$$\underset{R \in SO(3), t \in \mathbb{R}^3}{\operatorname{argmin}} \quad \|P_S - (RP_T + t)\|_2^2 \quad (4.13)$$

在位姿回归层模块，不再使用一个全连接层预测四元数和平移向量的组合，而是通过两个不同通道数的全连接层分别回归旋转矩阵 $R \in SO(3)$ 和平移向量 $t \in \mathbb{R}^3$ 。在这个模块中，全连接层共有 5 个隐藏层，分别为 1024、1024、512、512、256。在回归旋转矩阵时，输出层的大小为 4，此时网络的输出为一个四元数，还需要将其进一步转化为旋转矩阵。在回归平移向量时，输出层的大小为 3。将网络输出的旋转矩阵 $R \in SO(3)$ 和平移向量 $t \in \mathbb{R}^3$ 作用于源点云，即可将源点云与模板点云对齐，从而完成点云配准。

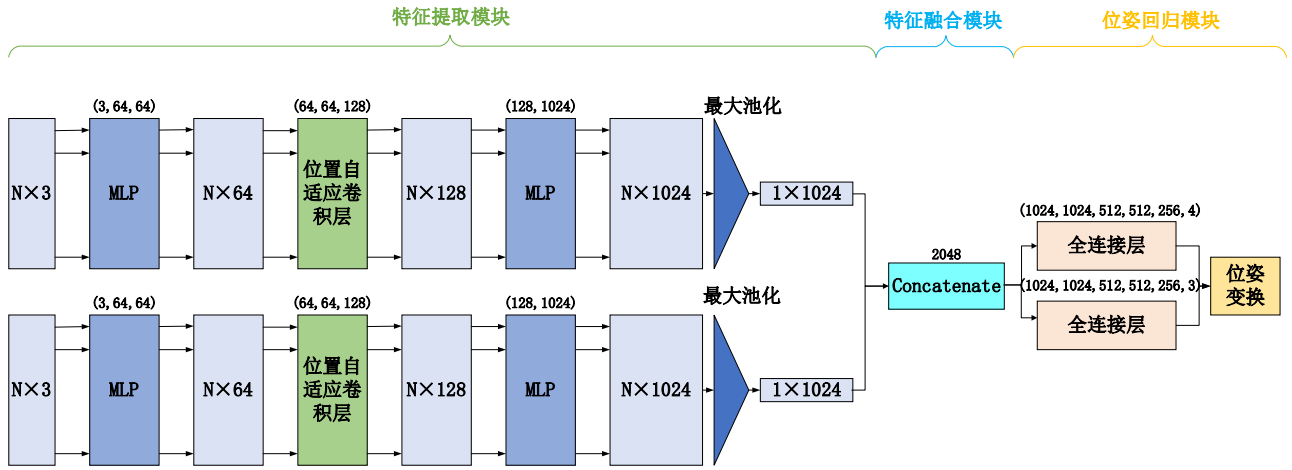


图 4.6 PACNet 网络整体框架示意图

4.3 实验结果与分析

本节将本文提出的网络模型与当前比较主流的点云配准算法 ICP、Go-ICP、FGR、PointNetLK、PCRNNet 进行比较，使用均方误差、均方根误差以及平均绝对误差来评估真实值和预测值之间的误差。其中 ICP、Go-ICP、FGR、PointNetLK 的各项测试性能指标本节参考了文献^[29]中的数据，而 PCRNNet 以及 PACNet 的各项测试性能指标均在实验室服务器上完成，并保证了不同网络模型训练和测试采用的数据集相同以及方式相同。

4.3.1 数据集与硬件配置

本节采用 ModelNet40 标准数据集对本文提出的网络模型进行训练与测试。本节使用 9843 个模型作为训练集，2468 个模型作为测试集。硬件配置为 Intel®Xeon®Silver4210

CPU@2.20GHz 和两块 TeslaP100-PCI-E GPU，内存为 12GB，总显存为 32GB。本章实验在 Ubuntu18.04 环境下运行，采用 PyTorch 1.9.0 作为深度学习框架。

4.3.2 训练参数设置

对于训练集和测试集中的每个样本随机选择 1024 个点，即为模板点云，并随机生成一个旋转矩阵和平移向量作为初始变换矩阵。将初始变换矩阵作用于模板点云生成源点云，这样就得到了待配准的点云对。为了保证实验的科学性，本节设置了随机种子，确保每次随机生成的变换矩阵相同。本节采用 Adam^[67] (Adaptive Moment Estimation) 梯度下降优化器训练网络，初始学习率设置为 0.001，训练集和测试集的批量大小均设置为 16，共训练 250 步。

如图 4.7 所示分别为本章的模型在各项评价指标上的训练曲线和测试曲线。从图中可以看出，训练曲线和测试曲线相似，误差指标在前 75 步下降剧烈，但当训练步数达到 75 时，误差指标曲线趋于平缓，表明模型已经收敛并进入了全局最优。

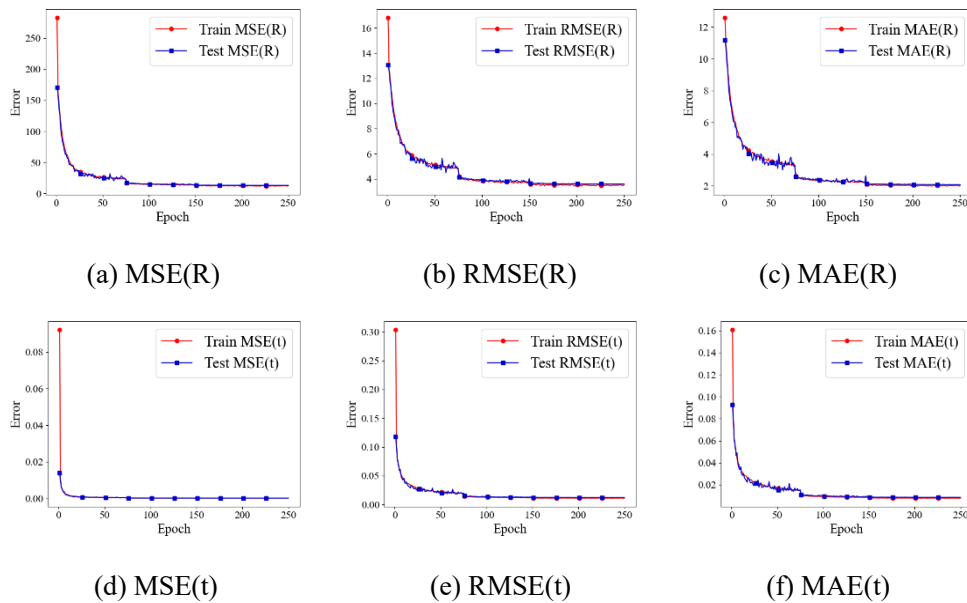


图 4.7 不同评价指标的训练曲线和测试曲线对比图

4.3.3 整体配准效果

本节将模型预测得到的旋转矩阵和平移向量作用到源点云上得到配准后的点云，并同时可视化模板点云、源点云以及配准后的点云来查看配准的效果，如图 4.8 所示，其中蓝色为源点云，绿色为模板点云，红色为配准后的点云。从可视化结果可以看出，本章的配准方法不论是在像桌子、花瓶、飞机这样具有规则对称结构的点云中，还是在吉他、花束、人这样具有复杂结构的点云中，均可以获得较好的配准效果。

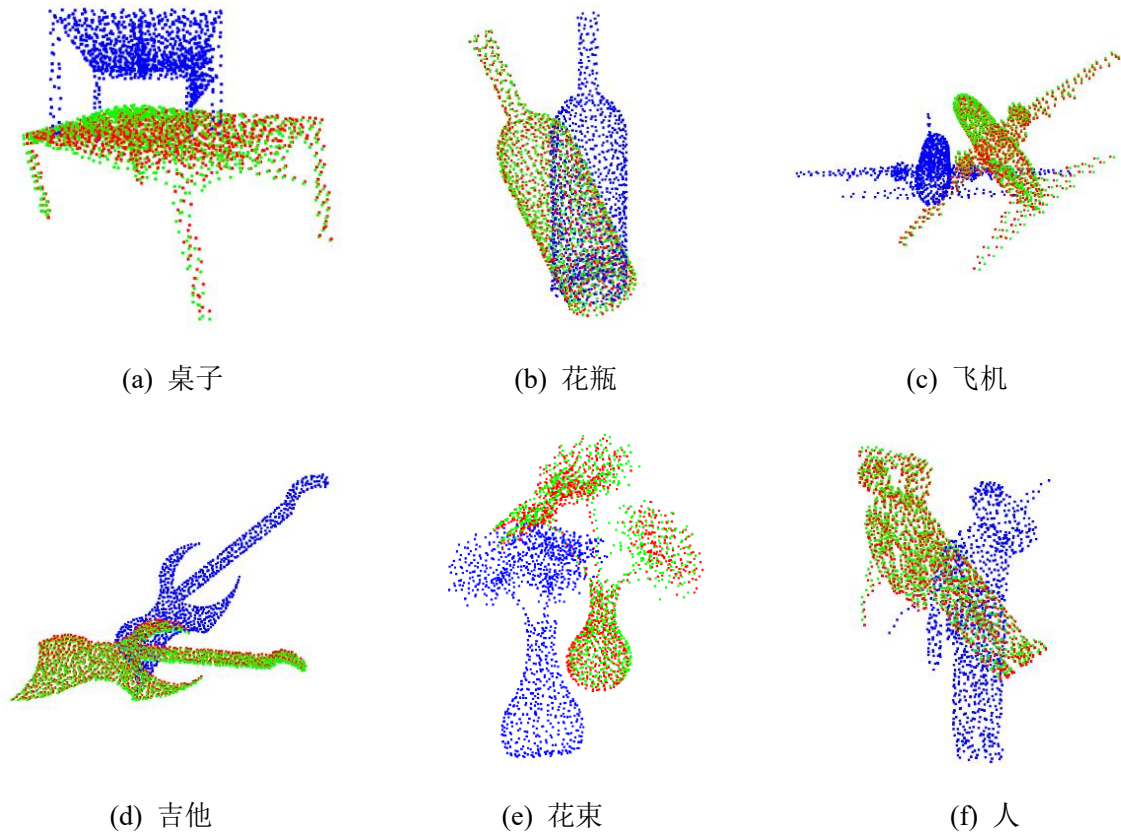


图 4.8 配准效果展示图

4.3.4 测试对比

(1) 不同卷积层数对比

在实验的过程中，本节分别训练并测试了嵌入 1 到 3 层位置自适应卷积的模型，并对这些模型在原始数据集上进行了性能评估，有趣地发现嵌入不同层数的位置自适应卷积会产生不同的结果，实验结果如表 4.1 所示。

表 4.1 不同卷积层数测试结果对比表

卷积层数	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
1	14.113483	3.756792	2.166226	0.000149	0.012212	0.008324
2	12.930086	3.595843	2.072818	0.000143	0.011957	0.008473
3	13.466946	3.669734	2.148391	0.000137	0.011701	0.008457

从表格数据可以看出，嵌入 2 层位置自适应卷积的模型具有最优性能。在旋转矩阵上的性能评估上，嵌入 2 层位置自适应卷积的模型均优于嵌入其他数量的模型。在平移矩阵的性能评估上，嵌入 2 层位置自适应卷积的模型没有明显的优势，但性能均衡。这说明过多的嵌入位置自适应卷积并不会对模型带来更好的结果。这是因为过多的卷积层将导致提取出的

局部特征变得冗余，从而影响了重要信息的提取，反而降低配准精度^[68]。同时，冗余的数据也加重了网络的负担，增加了算法的时间复杂度和空间复杂度。

(2) 精度测试

表 4.2 评估了本章的方法及其同行的性能。在所有性能指标中，本章的方法在 MSE(R)、RMSE(R)、MSE(t)和 RMSE(t)均最低。在 MAE(R)这个指标上，略高于 FGR。在 MAE(t)这个指标上，本章的方法没有明显的优势。这是由于 MAE 计算的是真实值和预测值绝对误差的平均值，大部分情况下梯度都相等，即使是较小的损失值其梯度也可能很大，因此不利于模型参数的收敛与学习。对比 PCRNet，本章的方法嵌入了两层位置自适应卷积，分别带来了 25.6%、13.7%、21.6%、39.4%、22.1%、19.1%的提升。同时，本章的方法预测出的旋转矩阵和平移向量与地面值的误差有四个指标最小，优于其他主流方法，具有更高的配准精度。

表 4.2 精度测试结果对比表

模型	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
ICP	894.897339	29.914835	23.544817	0.084643	0.290935	0.248755
Go-ICP	140.477325	11.852313	2.588463	0.000659	0.025665	0.007092
FGR	87.661491	9.362772	1.999290	0.000194	0.013939	0.002839
PointNetLK	227.870331	15.095374	4.225304	0.000487	0.022065	0.005404
PCRNet	17.447983	4.177078	2.628245	0.000241	0.015526	0.010579
本章方法	12.930086	3.595843	2.072818	0.000143	0.000143	0.008473

4.4 本章小结

综上所述，本章以 PCRNet 深度学习网络模型为基础进行了改进与优化，构建了一种使用位置自适应卷积提取特征的 PACNet 点云配准网络模型。在特征提取模块，利用具有动态内核组装的位置自适应卷积对输入的点云进行特征学习，可以灵活地对三维点云的复杂空间变化和几何结构进行建模，从而提高对场景的理解能力。通过实验将 PACNet 网络模型在标准数据集 ModelNet-40 上进行了各项测试对比。实验结果表明，与 ICP、Go-ICP、FGR、PCRNet 这些同类主流算法相比，本章提出的方法在点云配准任务中具有更高的精度

第五章 基于双重注意力机制融合特征的点云配准优化

基于 PCRNet 深度学习网络模型, 本章提出了一种基于双重注意力机制融合特征的点云配准的点云配准优化方法。首先介绍了注意力机制相关的背景知识, 并分析了 PCRNet 网络在特征融合方面存在的局限性。然后针对这些局限性提出改进, 结合上一章改进的网络构建了 PACNet-Att 网络模型。最后通过实验将 PACNet-Att 网络模型与当前主流的点云配准算法进行了测试对比以及消融实验, 证明了此优化方法的先进性以及两个改进方法的有效性。

5.1 注意力机制

注意力是人类不可或缺的复杂认知功能, 人类的视觉感知系统在处理大量信息时并不会一次性处理全部信息, 而是在需要的时间和地点选择性地专注于信息独特的部分, 同时忽略其他可感知的信息^[69]。这是人类利用有限的处理资源从海量信息中快速选择高价值信息的一种手段。受此启发, 研究学者们将注意力机制运用于计算机视觉领域, 让其能根据具体的任务处理关键信息。目前, 注意力机制已经成为深度学习领域最重要的概念之一^[70]。由于注意力机制可以学习特征之间的映射关系, 并捕获特征之间的相互依赖性^[71], 因此本章通过引入注意力机制学习不同层次特征的映射关系, 实现点云局部特征和全局信息的有效融合。

5.2 基于双重注意力机制融合特征的 PACNet-Att

由于 PCRNet 网络在特征融合模块仅以数组拼接的方式将两组点云的特征在维度上进行连接, 不能充分利用不同层次特征的互补性, 无法很好地融合点云的局部特征和全局信息。为了解决上述问题, 本章构建了一种结合空间注意力机制和通道注意力机制融合特征^[72, 73]的点云配准网络模型, 该网络可以捕获不同层次特征的长期语义依赖关系, 有效融合局部特征和全局信息, 从而提高点云配准的精度。

5.2.1 空间注意力机制

空间注意力机制^[74, 75]可以实现空间维度上的特征聚焦, 加深网络中深层信息的传递。设 $A^{G \times N \times M}$ 表示输入点云特征矩阵, 其中 G 表示通道数, N 和 M 分别表示点云的数量和维度。将 A 分别输入三个通道数为 64 的卷积层进行降维处理, 得到 $B^{G \times N \times M}$ 、 $C^{G \times N \times M}$ 和 $D^{G \times N \times M}$ 。将 B 和

C 进行维度转换得到 $B^{G \times J}$ 和 $C^{G \times J}$, 此时 $J = N \times M$ 。对 B 进行转置得到 B^T , 并与 C 进行矩阵相乘, 经过 softmax 归一化得到空间注意力系数矩阵 S_{ji} , 如公式 5.1 所示:

$$S_{ji} = \frac{\exp(B_i^* C_j)}{\sum_i^J \sum_j^J \exp(B_i^* C_j)} \quad (5.1)$$

其中 S_{ji} 表示 i 位置特征对 j 位置特征的影响。相似位置的特征系数更高, 点的依赖性关系更强。

同时, 将特征 A 输入到卷积层得到特征矩阵 $D^{G \times N \times M}$, 通过维度转换得到 $D^{G \times J}$ 。加入一个缩放因子 a , 每个位置的特征 $E^{G \times N \times M}$ 是所有位置的特征和原始特征的加权和, 这个过程如公式 5.2 所示:

$$E_j = a \sum_{i=1}^J (S_{ji} D_i) + A_j \quad (5.2)$$

a 初始化为 0, 通过逐渐地学习并更新权重, 最后输出加权后的每个位置的特征。空间注意机制可以选择性地聚合特征, 并提高语义一致性^[76]。

5.2.2 通道注意力机制

通道注意力机制^[74, 75]也通过类似空间注意力机制的方式来学习任意两个通道特征之间的映射关系, 利用通道之间的相互依赖性来更新某一个通道, 并通过通道之间的关联, 增强特定语义的响应能力。为了避免打破通道之间的关系, 通道注意力机制最开始不进行卷积运算, 而是直接将原始特征矩阵 $A^{G \times N \times M}$ 通过矩阵计算和 softmax 函数得到通道注意力系数矩阵 $X^{G \times G}$, 如公式 5.3 所示:

$$X_{ji} = \frac{\exp(A_i^* A_j)}{\sum_{i=1}^G \exp(A_i^* A_j)} \quad (5.3)$$

其中 X_{ji} 表示 i 通道特征对 j 通道特征的影响。

加入一个缩放因子 b , 每个通道的输出特征是所有通道的特征和原始特征的加权和, 这个过程如公式 5.4 所示:

$$E_j = b \sum_{i=1}^G (X_{ji} A_i) + A_j \quad (5.4)$$

b 初始化为 0, 通过逐渐地学习并更新权重, 最后输出加权后的每个通道的特征, 表明特征映射之间的长期语义依赖关系, 有助于提高特征的可分辨性。

5.2.3 PACNet-Att 网络设计

本节引入空间注意力机制和通道注意力机制，基于 PCRNet 网络在特征融合模块提出了改进，并结合上一章改进的网络 PACNet 构建了 PACNet-Att 点云配准网络模型，网络模型整体框架如图 5.1 所示。该网络模型的整体结构同样是由特征提取模块、特征融合以及位姿回归模块组成。

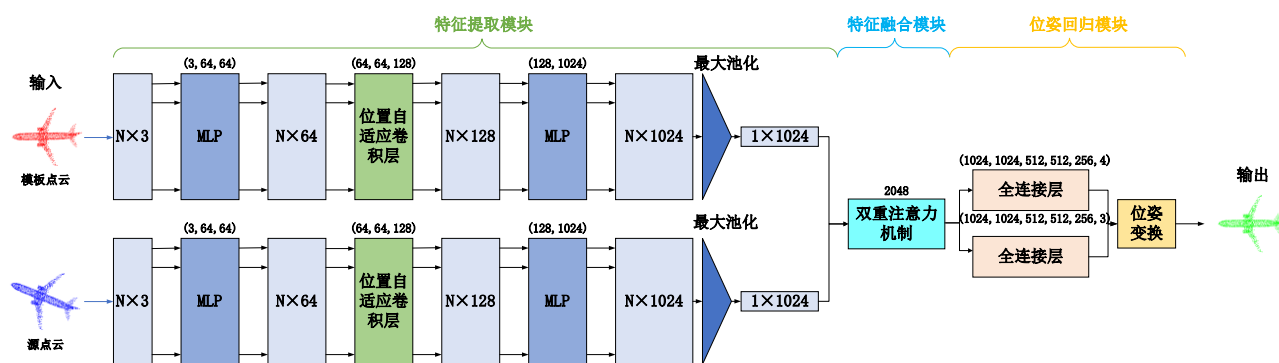


图 5.1 PACNet-Att 网络模型整体框架示意图

特征融合模块使用了结合空间注意力机制和通道注意力机制的双重注意力机制，结构如图 5.2 所示。首先使用空间注意力机制学习综合了局部特征和全局信息的 1024 维抽象特征，实现空间维度上的特征聚焦，输出加权后每个位置的特征。其次使用通道注意力机制学习综合了局部特征和全局信息的 1024 维抽象特征，捕获任意两个通道特征之间的映射关系，输出加权后的每个通道的特征。最后以深度学习中的 concatenate 数组拼接操作将 1×1024 维的抽象特征拼接成 1 个 1×2048 维的特征。至此，本章完成了基于双重注意力机制的特征融合。

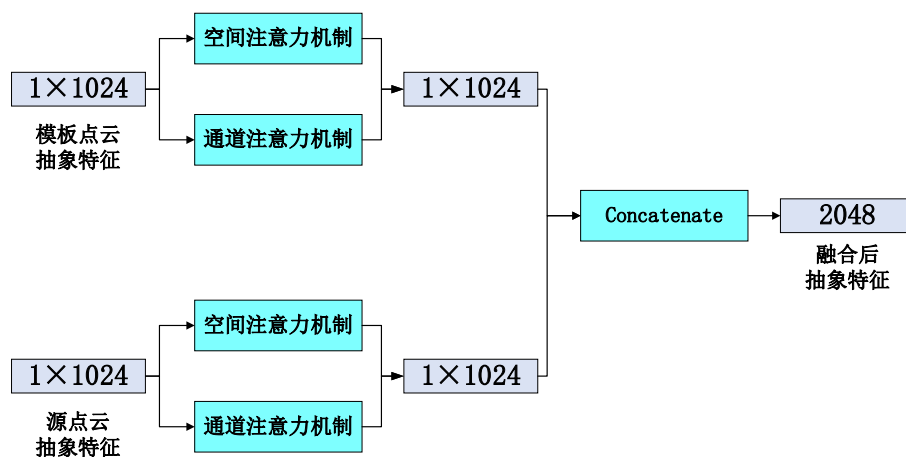


图 5.2 双重注意力机制结构示意图

5.3 实验结果与对比分析

本节将本章构建的网络模型 PACNet-Att 与当前比较主流的点云配准算法 ICP、Go-ICP、FGR、PointNetLK、PCRNet 分别在精度、泛化性以及噪声鲁棒性三个方面进行了比较，并同样使用均方误差、均方根误差以及平均绝对误差来评估真实值和预测值之间的误差。其中，PACNet-Att 的各项测试性能指标均在自己的电脑上完成。

5.3.1 实验环境和训练参数设置

本节实验环境和训练参数和第四章完全相同，因此不再赘述。如图 5.3 所示分别为本章的模型在不同评价指标上的训练曲线和测试曲线。从图中可以看出，同样地，训练曲线和测试曲线相似，但训练时的初始误差比测试时的初始误差大得多。误差曲线在前 75 步下降幅度很大，尤其在 25~50 步之间抖动剧烈，之后趋于平缓，表明模型已经收敛并进入了全局最优。

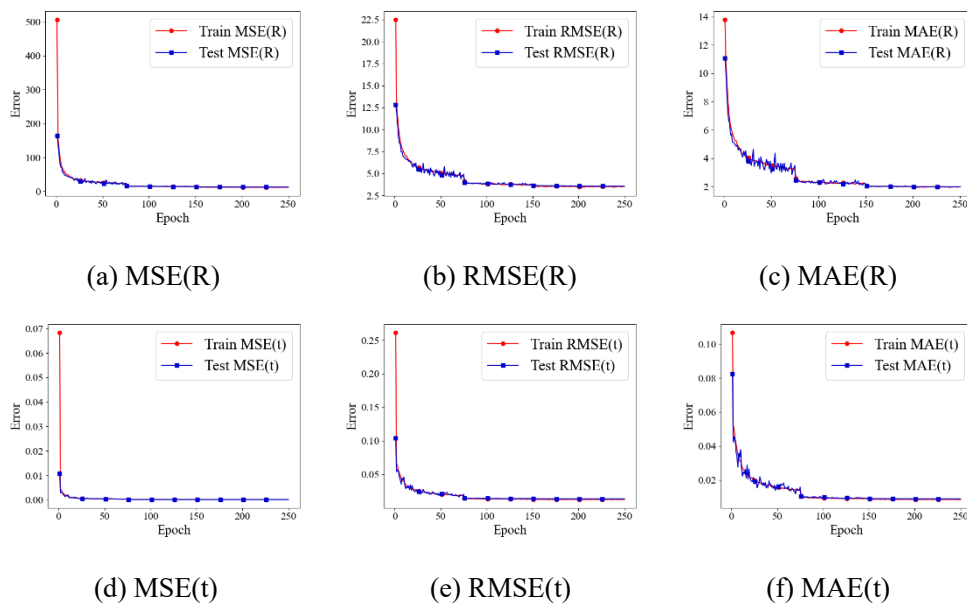


图 5.3 不同评价指标的训练曲线和测试曲线对比图

5.3.2 整体配准效果

同样地，本章将模型预测得到的旋转矩阵和平移向量作用到源点云上得到配准后的点云，并同时可视化模板点云、源点云以及配准后的点云来查看配准的效果，如图 5.4 所示，其中蓝色为源点云，绿色为模板点云，红色为配准后的点云。从可视化结果可以看出，本章配准方法不论是在像门框、转椅、显示器这样具有规则对称结构的点云中，还是在沙发、树枝、钢琴这样具有复杂结构的点云中均可以获得较好的配准效果。

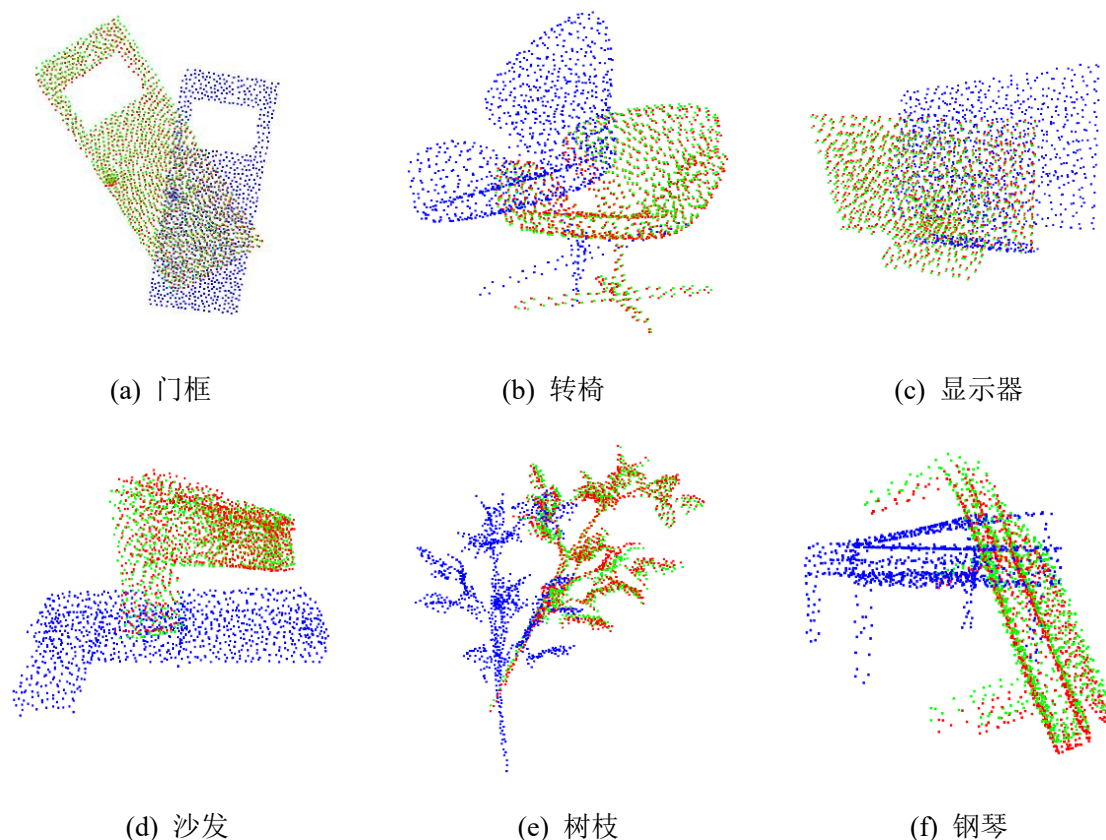


图 5.4 配准效果展示图

5.3.3 测试对比

(1) 精度测试

表 5.1 评估了本章的方法及其同行的性能。在所有性能指标中，同样地，本章的方法在 $MSE(R)$ 、 $RMSE(R)$ 、 $MSE(t)$ 和 $RMSE(t)$ 三个指标上均最低。在 $MAE(R)$ 这个指标上，仅次于 FGR。在 $MAE(t)$ 这个指标上，本章的方法没有明显的优势。对比 PCRNet，PACNet-Att 网络模型在嵌入了两层位置自适应卷积和双重注意力机制后，在 6 个评价指标上分别带来了 30.3%、16.5%、23.4%、40.7%、22.9%、24% 的提升。本章的方法预测出的旋转矩阵和平移向量与地面值的误差有四个指标最小，其余两个指标表现均衡。总体来看，本章的方法性能优于其他主流方法，具有更高的配准精度。

表 5.1 各算法精度测试结果对比表

模型	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
ICP	894.897339	29.914835	23.544817	0.084643	0.290935	0.248755
Go-ICP	140.477325	11.852313	2.588463	0.000659	0.025665	0.007092
FGR	87.661491	9.362772	1.999290	0.000194	0.013939	0.002839
PointNetLK	227.870331	15.095374	4.225304	0.000487	0.022065	0.005404
PCNet	17.447983	4.177078	2.628245	0.000241	0.015526	0.010579
本章方法	12.411096	3.522938	2.049789	0.000140	0.011842	0.008301

(2) 泛化性对比

为了测试不同模型的泛化能力，本节在看不见的类别中对本章方法及其同行进行了测试，结果如表 5.2 所示。同样地，本章的方法在 MSE(R)、RMSE(R)、MSE(t)和 RMSE(t)均最低。在 MAE(R)这个指标上，略高于 FGR。在 MAE(t)这个指标上，本章的方法没有明显的优势。本章的方法预测出的旋转矩阵和平移向量与地面值的误差有四个指标最小，其余两个指标表现均衡。总体来看，本章的方法在看不见的类别中依然可以保持更高的配准精度，这也体现了本章方法具有更好的泛化性能。

表 5.2 泛化性测试结果对比表

模型	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
ICP	892.601135	29.876431	23.626110	0.086005	0.293266	0.251916
Go-ICP	192.258636	13.865736	2.914169	0.000491	0.022154	0.006219
FGR	97.002747	9.848997	1.445460	0.000182	0.013503	0.002231
PointNetLK	306.323975	17.502113	5.280545	0.000784	0.028007	0.007203
PCNet	18.333996	4.281822	2.693274	0.000194	0.013912	0.009988
本章方法	12.666718	3.559033	1.982284	0.000182	0.013489	0.008896

(3) 噪声鲁棒性对比

为了测试噪声对不同模型的影响，本节在输入点云中添加期望值为 0、方差为 0.01 的高斯噪声进行测试，结果如表 5.3 所示。本章的方法在 MSE(R)、RMSE(R)、MAE(R)、MSE(t)和 RMSE(t)这五个指标上均优于其他所有方法。唯一不足的 MAE(t)指标，本章的方法同样仅仅优于 ICP 和 PCNet，没有明显的优势。但总体看来，本章方法对噪声的鲁棒性更强。

表 5.3 噪声鲁棒性测试结果对比表

模型	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
ICP	882.564209	29.707983	23.557217	0.084537	0.290752	0.249092
Go-ICP	131.182495	11.453493	2.534873	0.000531	0.023051	0.004192
FGR	607.694885	24.651468	10.055918	0.011876	0.027393	0.002231
PointNetLK	256.155548	16.004860	4.595617	0.000465	0.021558	0.005652
PCRNet	17.806925	4.219825	2.675704	0.000236	0.015368	0.010528
本章方法	12.837771	3.582983	2.016258	0.000179	0.013361	0.008862

5.3.4 消融实验

为了验证位置自适应卷积和双重注意力机制在基础网络 PCRNet 中的有效性, 本节又对单独使用双重注意机制融合特征的网络进行了消融实验, 并将其和基础网络 PCRNet、单独使用位置自适应卷积提取特征的网络 PACNet、同时使用位置自适应卷积提取特征和双重注意力机制融合特征的网络 PACNet-Att 进行了对比, 结果如表 5.4 所示。

实验数据表明, 在 PCRNet 中单独使用位置自适应卷积提取特征, 6 项评价指标分别带来了 25.6%、13.7%、21.6%、39.4%、22.1%、19.1%的提升。在 PCRNet 中单独使用双重注意力机制融合特征, 6 项评价指标分别带来了 27.3%、14.8%、23.3%、27.1%、14.6%、17.4%的提升。在 PCRNet 中同时使用位置自适应卷积提取特征以及双重注意力机制融合特征, 6 项评价指标分别带来了 30.3%、16.5%、23.4%、40.7%、22.9%、24%的提升。综上所述, 位置自适应卷积和双重注意力机制各自都能对网络起到实质的作用, 从而提高点云配准的精度。

表 5.4 消融实验测试结果对比表

模型	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
PCRNet	17.806925	4.219825	2.675704	0.000236	0.015368	0.010528
PACNet	13.254151	3.640625	2.098704	0.000143	0.011975	0.008520
PCRNet-Att	12.938107	3.596958	2.052865	0.000172	0.013123	0.008699
本章方法	12.411096	3.522938	2.049789	0.000140	0.011842	0.008301

5.4 本章小结

综上所述, 本章以 PCRNet 深度学习网络模型为基础进行了改进与优化, 并结合上一章改进的网络, 构建了一种基于双重注意力机制融合特征的 PACNet-Att 点云配准网络模型。在特征融合模块, 结合空间注意力机制和通道注意力捕获不同层次特征的长期语义依赖关系,

可以有效融合点云的局部特征和全局信息。通过实验将 PACNet-Att 网络模型与当前主流的点云配准方法进行了各项测试对比。实验结果表明,与 ICP、Go-ICP、FGR、PCRNNet 这些同类主流算法相比,本章提出的方法在点云配准任务中具有更高的精度、更高的泛化性能以及更稳健的噪声鲁棒性。同时通过消融实验,验证了本文提出的两个改进点,即使用位置自适应卷积提取特征,并使用双重注意力机制融合特征,各自都能对网络起到实质性的作用,从而提高点云配准的精度。

第六章 总结与展望

6.1 本文总结

点云配准是三维计算机视觉领域的关键课题之一，探索点云配准对理论基础研究与实际工程应用中都有重要意义。但外部传感器采集到的点云数据往往存在冗余性与质量缺陷，这一方面使得配准过程占用大量的内存资源和处理时间，另一方面噪声干扰也会导致特征匹配产生的对应关系出现异常值，从而影响点云配准的精度。在深度学习领域中，以点云配准经典网络 **PCNet** 为代表的点云配准方法在特征提取模块只关注点云的全局信息，并没有考虑到点云的局部特征，难以表征点云空间复杂的变化关系，一定程度限制了其对场景的理解能力，从而降低点云配准的精度。同时，**PCNet** 在特征融合模块仅仅以数组拼接方式将两组点云的特征在维度上进行连接，不能充分利用不同层次特征的互补性，无法很好地融合点云的局部特征和全局信息，从而破坏点云配准的精度。本文针对以上点云配准任务中存在的问题，主要对以下三个部分进行研究与改进：

第一部分，针对外部传感器采集到的点云数据存在冗余性与质量缺陷的问题，本文提出一种基于几何特征点云数据预处理方案，为后续利用有效点云信息进行点云配准奠定基础。该方案首先使用随机降采样这种方法来降低原始点云的冗余性，并指定采样数量为 48977。然后使用统计滤波器进行离群点去除，并取近邻搜索点的个数为 30。最后使用欧式聚类算法进行无关背景去除，并取近邻搜索半径的大小为 5。测试下来，降采样后，点云数据从 271114 个点减少到 48977 个点，用时为 0.104379s，且在保证点云结构与重要特征的同时过滤掉了绝大多数干扰噪声。离群点去除后，点云个数又从 48977 减少为 48105，用时为 0.122201s，且去除了绝大多数离群点。在进行无关背景去除时，点云被分为 10 个类别，分离出目标物体（手机）后，点云个数从 48105 减少为 6067，即目标物体（手机）的点云数量为 6067，用时为 0.295816s。总用时 0.522397s。

第二部分，针对以点云配准经典网络 **PCNet** 为代表的点云配准方法在特征提取模块只关注点云的全局信息而忽略局部特征的问题，本文在基础网络 **PCNet** 的特征提取模块进行了改进，提出了一种使用位置自适应卷积提取特征点云配准网络模型 **PACNet**。在特征提取模块，利用具有动态内核组装的位置自适应卷积层对输入的点云进行特征学习，根据每个点与其邻域点的位置关系学习权重系数，并组合权重矩阵自适应地构建卷积内核，可以灵活地捕获局部区域的信息，从而提高对场景的理解能力。本文通过实验将 **PACNet** 网络模型在标

准数据集 ModelNet-40 上进行了各项测试对比。实验结果表明, PACNet 对比基础网络 PCRNet, 在 6 个评价指标上分别带来了 25.6%、13.7%、21.6%、39.4%、22.1%、19.1% 的提升。与 ICP、Go-ICP、FGR、PCRNet 这些同类主流算法相比, 本文提出的方法在点云配准任务中也具有更高的精度。

第三部分, 针对 PCRNet 在特征融合模块仅仅以数据拼接方式将两组点云的特征在维度上进行连接而无法很好地融合不同层次特征的问题, 本文在基础网络 PCRNet 的特征融合模块进行了改进, 并结合 PACNet 网络构建了一种基于双重注意力机制融合特征的点云配准网络模型 PACNet-Att。在特征融合模块, 双重注意力机制可以捕获不同层次特征的长期语义依赖关系, 有效融合点云的局部特征和全局信息。本文通过实验将 PACNet-Att 网络模型与当前主流的点云配准算法进行了各项测试对比以及消融实验。实验结果表明, PACNet-Att 对比基础网络 PCRNet, 在 6 个评价指标上分别带来了 30.3%、16.5%、23.4%、40.7%、22.9%、24% 的提升。与 ICP、Go-ICP、FGR、PCRNet 这些同类主流算法相比, 本文提出的方法在点云配准任务中也具有更高的精度、更高的泛化性能以及更稳健的噪声鲁棒性。同时本文提出的两个改进点, 即使用位置自适应卷积提取特征, 并使用双重注意力机制融合特征, 各自都能对网络起到实质的作用, 从而提高点云配准的精度。

6.2 研究展望

本文对基于深度学习的点云配准算法研究取得了一定的成果, 但仍存在可进一步探索的空间, 对此提出一些可行的改进思路与方向, 并对今后的研究进行展望:

(1) 在使用外部传感器对真实场景进行点云数据采集时, 目标物体(手机)与周围无关背景(苹果、卷尺等杂物)的摆放比较理想, 并没有重叠的部分, 这使得本文可以很好地分离出目标物体。未来, 具有重叠部分的点云分割将成为接下来的研究重点之一。

(2) 在对外部传感器采集到的稠密点云数据进行点云预处理时, 本文采用的是基于几何特征的方案, 其中近邻搜索点的个数、近邻搜索半径的大小等参数都需要根据实际场景来调整, 这使得算法的自适应能力较差。未来可考虑使用点云的语义特征, 实现更高层次特征点云预处理, 更加智能化地处理点云数据。

(3) 在使用 PACNet-Att 网络对两片不对齐的点云进行配准时, 以 MSE(R)这个指标为例, 误差仍有 12.411096, 还有很大的提升空间。本文仅在特征提取模块与特征融合模块做出了有效地改进, 如何更好地解算位姿将成为接下来的研究内容之一。

参考文献

- [1] Tian H, Wang T, Liu Y, et al. Computer Vision Technology in Agricultural Automation—A Review[J]. Information Processing in Agriculture, 2020, 7(1): 1-19.
- [2] 赵跃东, 孙多润, 郭警中, et al. 基于 3D 视觉传感器边缘识别和追踪算法的研究[J]. 锻压装备与制造技术, 2020, 55(02): 91-94.
- [3] Pawar P G, Devendran V. Scene Understanding: A Survey to See the World at a Single Glance[C]. 2019 2nd International Conference on Intelligent Communication and Computational Techniques, 2019: 182-186.
- [4] Mittal S, Vaishay S. A Survey of Techniques for Optimizing Deep Learning on GPUs[J]. Journal of Systems Architecture, 2019, 99: 101635.
- [5] 徐锡磊. 数字城市的应用与展望[J]. 智能城市, 2020, 6(13): 42-43.
- [6] Yurtsever E, Lambert J, Carballo A, et al. A Survey of Autonomous Driving: Common Practices and Emerging Technologies[J]. IEEE Access, 2020, 8: 58443-58469.
- [7] 张辉, 王盼, 肖军浩, et al. 一种基于三维建图和虚拟现实的人机交互系统[J]. 控制与决策, 2018, 33(11): 1975-1982.
- [8] Wang X, Chan T O, Liu K, et al. A Robust Segmentation Framework for Closely Packed Buildings from Airborne Lidar Point Clouds[J]. International Journal of Remote Sensing, 2020, 41(14): 5147-5165.
- [9] Phang J T S, Lim K H, Chiong R C W. A Review of Three Dimensional Reconstruction Techniques[J]. Multimedia Tools and Applications, 2021, 80(12): 17879-17891.
- [10] 邹娟茹, 孙兴华. 多源数据融合构建精细三维模型的关键技术[J]. 地理空间信息, 2021, 19(12): 131-134+137.
- [11] Campos C, Elvira R, Rodríguez J J G, et al. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap Slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [12] Huang X, Mei G, Zhang J, et al. A Comprehensive Survey on Point Cloud Registration[J]. arXiv preprint arXiv: 210302690, 2021.
- [13] 王珊, 王春阳. 基于特征点匹配的三维点云配准算法[J]. 国外电子测量技术, 2020, 39(12): 23-28.
- [14] Zhang Z, Dai Y, Sun J. Deep Learning Based Point Cloud Registration: An Overview[J]. Virtual Reality & Intelligent Hardware, 2020, 2(3): 222-246.
- [15] Huang X, Mei G, Zhang J. Feature-Metric Registration: A Fast Semi-Supervised Approach for Robust Point Cloud Registration without Correspondences[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11366-11374.
- [16] Makovetskii A, Voronin S, Kober V, et al. An Algorithm for Rough Alignment of Point Clouds in Three-Dimensional Space[C]. 2020 International Conference on Information Technology and Nanotechnology, 2020: 1-4.
- [17] Lu H, Shi H. Deep Learning for 3D Point Cloud Understanding: A Survey[J]. arXiv preprint arXiv: 200908920, 2020.
- [18] Makay B P. A Method for Registration of 3D Shape[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14: 239-256.
- [19] Magnusson M. The Three-Dimensional Normal-Distributions Transform: An Efficient Representation for Registration, Surface Analysis, and Loop Detection[D]. Örebro universitet, 2009.
- [20] Yang J, Li H, Campbell D, et al. Go-Icp: A Globally Optimal Solution to 3D ICP Point-Set Registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(11): 2241-2254.
- [21] Zhou Q-Y, Park J, Koltun V. Fast Global Registration[C]. European Conference On Computer Vision, 2016: 766-782.
- [22] 祝瑞红, 黄昶. 基于快速点特征直方图的三维点云配准算法[J]. 长江信息通信, 2021, 34(02): 72-74.

- [23] Shen Z, Ma X, Li Y. A Hybrid 3D Descriptor with Global Structural Frames and Local Signatures of Histograms[J]. IEEE Access, 2018, 6: 39261-39272.
- [24] Qi C R, Su H, Mo K, et al. Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [25] Zeng A, Song S, Niessner M, et al. 3DMatch: Learning Local Geometric Descriptors from Rgb-D Reconstructions[M]. 30th IEEE Conference on Computer Vision and Pattern Recognition. 2017: 199-208.
- [26] Deng H, Birdal T, Ilic S. Ppfnet: Global Context Aware Local Features for Robust 3D Point Matching[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 195-205.
- [27] Yew Z J, Lee G H. 3DFeat-Net: Weakly Supervised Local 3D Features for Point Cloud Registration[C]. Proceedings of the European Conference on Computer Vision, 2018: 607-623.
- [28] Aoki Y, Goforth H, Srivatsan R A, et al. Pointnetlk: Robust & Efficient Point Cloud Registration Using Pointnet[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7163-7172.
- [29] Wang Y, Solomon J M. Deep Closest Point: Learning Representations for Point Cloud Registration[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3523-3532.
- [30] Lu W, Wan G, Zhou Y, et al. DeepVCP: An End-to-End Deep Neural Network for Point Cloud Registration[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 12-21.
- [31] Sarode V, Li X, Goforth H, et al. PCRNNet: Point Cloud Registration Network Using Pointnet Encoding[J]. arXiv preprint arXiv:190807906, 2019.
- [32] Pais G D, Ramalingam S, Govindu V M, et al. 3DRegNet: A Deep Neural Network for 3D Point Registration[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7193-7203.
- [33] Fu K, Liu S, Luo X, et al. Robust Point Cloud Registration Framework Based on Deep Graph Matching[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8893-8902.
- [34] Bustos A P, Chin T-J. Guaranteed Outlier Removal for Point Cloud Registration with Correspondences[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(12): 2868-2882.
- [35] Guan B, Lin S, Wang R, et al. Voxel-Based Quadrilateral Mesh Generation from Point Cloud[J]. Multimedia Tools and Applications, 2020, 79(29), 20561-20578.
- [36] Jiang C, Sud A, Makadia A, et al. Local Implicit Grid Representations for 3D Scenes[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6001-6010.
- [37] Zhou Y, Tuzel O. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.
- [38] Thabet A, Alwassel H, Ghanem B. Self-Supervised Learning of Local Features in 3D Point Clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 938-939.
- [39] 张学典, 方慧. BTDCNN: 面向三维点云拓扑结构的 BallTree 动态图卷积神经网络[J]. 小型微型计算机系统, 2021: 1-8.
- [40] Sarabandi S, Thomas F. A Survey on the Computation of Quaternions from Rotation Matrices[J]. Journal of Mechanisms and Robotics, 2019, 11(2): 021006
- [41] 野莹莹, 张艳珠, 邹心宇, et al. 欧拉角姿态解算的改进[J]. 装备制造技术, 2018, (03): 104-105+134.
- [42] Levinson J, Esteves C, Chen K, et al. An Analysis of SVD for Deep Rotation Estimation[J]. Advances in Neural Information Processing Systems, 2020, 33: 22554-22565.
- [43] Hu G, Zhou Z, Cao J, et al. Non-Linear Calibration Optimisation Based on the Levenberg-Marquardt Algorithm[J]. IET Image Processing, 2020, 14(7): 1402-1414.
- [44] 马洁莹, 田暄, 翟庆, et al. 基于点到面度量的多视角点云配准方法[J]. 西安交通大学学报, 2022, (06): 1-12.

- [45] Chen Y, Hosseini B, Owaldi H, et al. Solving and Learning Nonlinear PDEs with Gaussian Processes[J]. *Journal of Computational Physics*, 2021, 447: 110668.
- [46] 纳文, 张世强, 曹越琦, et al. 雅可比矩阵在机器人运动中的应用[J]. *北京理工大学学报自然版*, 2020, 40(5): 576-580.
- [47] 兰林, 朱春钢. 基于 Newton 迭代法的最小二乘渐进迭代逼近[J]. *数值计算与计算机应用*, 2022, 43(1): 88-111.
- [48] Siddiqi K, Zhang J, Macrini D, et al. Retrieving Articulated 3D Models Using Medial Surfaces[J]. *Machine Vision and Applications*, 2008, 19(4): 261-275.
- [49] De Deuge M, Quadros A, Hung C, et al. Unsupervised Feature Learning for Classification of Outdoor 3D Scans[C]. *Australasian Conference on Robotics and Automation*, 2013, 2: 1-9.
- [50] Wu Z, Song S, Khosla A, et al. 3D Shapenets: A Deep Representation for Volumetric Shapes[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1912-1920.
- [51] Chang A X, Funkhouser T, Guibas L, et al. ShapeNet: An Information-Rich 3D Model Repository[J]. *arXiv preprint arXiv: 151203012*, 2015.
- [52] Dai A, Chang A X, Savva M, et al. ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 5828-5839.
- [53] Uy M A, Pham Q-H, Hua B-S, et al. Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 1588-1597.
- [54] Armeni I, Sener O, Zamir A R, et al. 3D Semantic Parsing of Large-Scale Indoor Spaces[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1534-1543.
- [55] Guo Y, Wang H, Hu Q, et al. Deep Learning for 3D Point Clouds: A Survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(12): 4338-4364.
- [56] Ruchay A, Dorofeev K, Kalschikov V. Accuracy Analysis of 3D Object Reconstruction Using Point Cloud Filtering Algorithms[J]. *Proceedings of the 5th Information Technology and Nanotechnology, ITNT-2019, Samara, Russia*, 2019: 21-24.
- [57] Chen C, Fragonara L Z, Tsourdos A. Fast Hierarchical Neural Network for Feature Learning on Point Cloud[J]. *arXiv preprint arXiv: 190604117*, 2019.
- [58] Liu H, Song R, Zhang X, et al. Point Cloud Segmentation Based on Euclidean Clustering and Multi-Plane Extraction in Rugged Field[J]. *Measurement Science and Technology*, 2021, 32(9): 095106.
- [59] Ge Y, Tang H, Xia D, et al. Automated Measurements of Discontinuity Geometric Properties from a 3D-Point Cloud Based on a Modified Region Growing Algorithm[J]. *Engineering Geology*, 2018, 242: 44-54.
- [60] Andoni A, Indyk P, Krauthgamer R. Earth Mover Distance over High-Dimensional Spaces[C]. *SODA*, 2008, 8: 343-352.
- [61] 袁冰清, 陆悦斌, 张杰. 神经网络与深度学习基础[J]. *数字通信世界*, 2018, 5: 59-62.
- [62] Ketkar N, Moolayil J. Introduction to Pytorch[M]. *Deep Learning with Python*. Springer. 2021: 27-91.
- [63] Chen Y, Dai X, Liu M, et al. Dynamic Convolution: Attention over Convolution Kernels[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11030-11039.
- [64] Mei J, Xiao C, Szepesvari C, et al. On the Global Convergence Rates of Softmax Policy Gradient Methods[C]. *International Conference on Machine Learning*, 2020: 6820-6829.
- [65] Xu M, Ding R, Zhao H, et al. Paconv: Position Adaptive Convolution with Dynamic Kernel Assembling on Point Clouds[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 3173-3182.
- [66] Lubana E S, Dick R P, Tanaka H. Beyond Batchnorm: Towards a General Understanding of Normalization in Deep Learning[J]. *arXiv preprint arXiv: 210605956*, 2021.
- [67] Singarimbun R N, Nababan E B, Sitompul O S. Adaptive Moment Estimation to Minimize Square Error in

- Backpropagation Algorithm[C]. 2019 International Conference of Computer Science and Information Technology, 2019: 1-7.
- [68] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [69] Niu Z, Zhong G, Yu H. A Review on the Attention Mechanism of Deep Learning[J]. Neurocomputing, 2021, 452: 48-62.
- [70] Jian L, Jun L, Xiong L. Research Progress in Attention Mechanism in Deep Learning[J]. 工程科学学报, 2021, 43(11): 1499-1511.
- [71] Cai Z. Interflow: Aggregating Multi-Layer Feature Mappings with Attention Mechanism[J]. arXiv preprint arXiv:2106.14073, 2021.
- [72] Ju M, Luo J, Wang Z, et al. Adaptive Feature Fusion with Attention Mechanism for Multi-Scale Target Detection[J]. Neural Computing and Applications, 2021, 33(7): 2769-2781.
- [73] Zhou H, Fang Z, Gao Y, et al. Feature Fusion Network Based on Attention Mechanism for 3D Semantic Segmentation of Point Clouds[J]. Pattern Recognition Letters, 2020, 133: 327-333.
- [74] Fu J, Liu J, Tian H, et al. Dual Attention Network for Scene Segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3146-3154.
- [75] Wang X, Jiang B, Zhang Z, et al. DeepGCNs-Att for Point Cloud Semantic Segmentation[C]. Journal of Physics: Conference Series, 2021, 2025(1): 012059.
- [76] Lu H, Chen X, Zhang G, et al. SCANet: Spatial-Channel Attention Network for 3D Object Detection[C]. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing, 2019: 1992-1996.

附录 1 攻读硕士学位期间撰写的论文

- [1] **Mei Q**, Wang F, Tong C, Zhang J, Jiang B, *Xiao J. PACNet: A High-precision Point Cloud Registration Network Based on Deep Learning[C]. 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2021: 1-5.
- [2] Wang F, **Mei Q**, Liu X, Xiao Y. Optimized Spatial Matching for Visual Object Tracking[C] 2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2021: 1-6.
- [3] Wang X, Jiang B, Zhang Z, Tong C, **Mei Q**, *Xiao J, *Tong, Y. DeepGCNs-Att for Point Cloud Semantic Segmentation[C]. Journal of Physics: Conference Series. IOP Publishing, 2021, 2025(1): 012059.
- [4] 许杰, 张子恒, 王新宇, 佟诚, **梅青**, *肖建. 一种基于 Zynq 的 CNN 加速器设计与实现[J]. 计算机技术与发展, 2021, 31(11): 108-113+121.

附录 2 攻读硕士学位期间申请的专利

- [1] 肖建, **梅青**, 佟诚, 王新宇, 张子恒, 江斌. 一种手写成绩识别显示器, ZL202021317576.8, 2021.03, 授权;
- [2] 肖建, 佟诚, **梅青**, 王新宇, 张子恒, 江斌. 一种基于仿生图像增强算法和 FPGA 硬件加速的手写成绩录入系统, CN111914683A, 2020.01, 公开;
- [3] 肖建, 王新宇, 张子恒, **梅青**, 佟诚, 张雷, 陈文勤. 一种基于 CNN 检测遵章率的路口交通规划系统, CN111145551A, 2020.05, 公开;
- [4] 肖建, 张子恒, 王新宇, 佟诚, **梅青**, 张雷, 许杰. 一种基于 ANN 的路口交通警力资源调度系统, CN111160537A, 2020.05, 公开。

附录 3 攻读硕士学位期间参加的科研项目

- (1) 2020 年江苏省研究生科研创新计划, 基于 GAAS 无人机的三维建模方法(KYCX20_0801), 主持人;
- (2) 横向项目, 基于 3D 扫描的高精度焊接路径规划方法研发(KH1060321192), 南京欧睿三维科技有限公司, 主要参与人员;
- (3) 横向项目, 励磁 CPU 开发服务(KH1060321191), 国电南瑞科技股份有限公司, 主要参与人员。

附录 4 攻读硕士学位期间参加的学科竞赛

- [1] 佟诚, **梅青**, 张子恒. “兆易创新杯”第十五届中国研究生电子设计竞赛, TI 企业命题二等奖, 2020.08;
- [2] 佟诚, **梅青**, 张子恒. “兆易创新杯”第十五届中国研究生电子设计竞赛, 全国总决赛团队三等奖, 2020.08;
- [3] 佟诚, **梅青**, 张子恒. “兆易创新杯”第十五届中国研究生电子设计竞赛, 全国总决赛华为专项赛三等奖, 2020.08;
- [4] 佟诚, **梅青**, 张子恒, 王新宇, 张欣. 第六届 3S 杯全国大学生物联网技术与应用“三创”大赛, 全国一等奖, 2020.12;
- [5] 章佳琪, **梅青**, 江斌. 第七届 3S 杯全国大学生物联网技术与应用“三创”大赛, 全国二等奖, 2021.10;
- [6] 江斌, 章佳琪, **梅青**. “兆易创新杯”第十五届中国研究生电子设计竞赛, 全国总决赛团队二等奖, 2021.08;
- [7] 王新宇, 张子恒, 张雷, 许杰, **梅青**. 第五届 3S 杯全国大学生物联网技术与应用“三创”大赛, 全国三等奖, 2019.11。

致谢

两年半的研究生生活即将结束，在本论文完成之际，谨向两年多来给予我关爱的良师益友和亲人们致以最崇高的敬意和衷心的感谢。

首先，我要感谢我的研究生导师肖建教授对我的指导与培养。在理论基础积累的阶段，肖老师教会我高效的文献调研能力、科学的文献管理能力以及全面的综述评价能力。在完成毕业课题阶段，从论文选题、实验方案确定、理论分析、数据处理直至论文撰写与定稿，每一步肖老师都给予了指导性意见与建议，倾注了肖老师大量的心血。肖老师渊博的学术积累、深邃的思想维度、严谨的治学态度、务实的工作作风，无一不深深地影响着我。

其次，我要感谢我的父母对我的付出与理解。感谢他们一直都是我最坚实的后盾，在我的身后给予我支持与信任，给予我勇气和力量，祝愿他们身体健康，平安如意。

再次，我要感谢课题组的同学们对我的关心与帮助。在研究生学习期间，我们相互分享学习心得体会，共同探讨学术研究问题，祝愿他们前程似锦，未来可期。

最后，我要向即将审阅本论文的专家们，以及评议和参与本论文答辩的老师表示诚挚的谢意。