

重庆邮电大学
CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

硕士学位论文
MASTER THESIS



论文题目 基于显著锚点几何嵌入的点云配准
方法研究

学科专业 _____
学号 _____
作者姓名 _____
指导教师 _____
学院 _____

学校代码 10617 UDC 004.93
分 类 号 TP391.4 密级 公开

学 位 论 文

基于显著锚点几何嵌入的点云配准方法研究

指导教师

申请学位级别 硕士 学科专业 电子信息
答辩委员会主席 论文答辩日期 2023 年 5 月 20 日
学位授予单位和日期 重庆邮电大学 2023 年 6 月

Research on Point Cloud Registration Method Based on Geometric Embedding of Significant Anchor Points

A Master Thesis Submitted to
School of Chongqing University of Posts and Telecommunications

Discipline **Electronic and Information Engineering**
Student ID
Author
Supervisor
School

重庆邮电大学

学位论文独创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文中不包含其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在论文中以明确方式标明。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：

日期： 年 月 日

重庆邮电大学

学位论文使用授权书

本人同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于：

公开论文

涉密论文，保密____年，过保密期后适用本授权书。

(请在以上方框内选择打“√”)

作者签名：

导师签名：

日期： 年 月 日

摘要

点云是一种用于表示三维空间中的对象的数据结构。它由许多离散的点组成，可以通过激光扫描仪、三维相机或其他传感器来捕捉和记录。由于这些传感器的视野有限，因此需要将多个点云合并成一个更大的点云，或将点云与先前的点云模型对齐以进行比较或更新，这个过程就是点云配准。点云配准是计算机视觉和机器人视觉中的重要问题，它们可以用于许多应用，例如三维建模、机器人导航、虚拟现实和医学影像分析。早期的点云配准主要集中于计算机合成数据集，然而随着社会的发展，越来越多的研究开始关注真实场景下的点云配准。但真实场景中普遍存在图案重复、几何形状较弱的困难区域，这些困难区域往往会由于特征相似导致点匹配的错误，影响变换矩阵的估计结果。

为了解决该问题，本文提出了一种基于显著锚点几何嵌入的点云配准方法。该方法嵌入显著锚点与超点之间的几何结构，以增强点特征的差异性和区分度。即使在源点云和目标点云中存在大量图案重复和弱几何区域，也能分辨出相似非重叠区域找到正确的点匹配。具体来说，首先通过锚点定位模块，在源点云和目标点云中定位识别能力最强、几何信息最丰富的超点对应作为显著锚点对应。采用非最大抑制算法，保证选定的一组显著锚点在点云中稀疏分布并且具有一定的几何结构。针对显著锚点，提出了一种基于锚点距离和角度的选择性几何结构嵌入算法，用于超点特征增强。这种显著锚点与超点之间的几何一致性，可以提高几何挑战性区域的特征区分度。然后，迭代更新以增强特征和锚点位置，获得最有效的显著锚点和超点特征。最后，通过在超点对应区域内寻找最近的相邻点来实现精确的点对应。

另外，本文提出了一种基于多模态融合的锚点定位点云配准方法，该方法通过将点云的结构特征和图像的纹理特征融合以提高几何挑战性区域的特征差异性。首先本文利用对齐模块将点云和图像数据对齐以找到超点与像素之间的对应关系。然后，利用融合模块将超点与对应像素之间的特征进行融合。该融合模块将点云特征和图像特征分别投影至模态无关和模态相关的两个子空间中，并先后在两个子空间中融合两种模态特征以达到减小域差异影响和防止信息丢失的作用。

关键词：点云配准，几何嵌入，弱几何区域，重复图案，多模态融合

ABSTRACT

A point cloud is a data structure used to represent objects in three-dimensional space. It consists of many discrete points that can be captured and recorded by laser scanners, 3D cameras, or other sensors. Due to the limited field of view of these sensors, multiple point clouds need to be merged into a larger point cloud, or point clouds are aligned with previous point cloud models for comparison or updating. This process is point cloud registration. Point cloud registration is an important problem in computer vision and robot vision, and it can be used in many applications, such as 3D modeling, robot navigation, virtual reality, and medical image analysis. Early point cloud registration mainly focused on computer-synthesized datasets. However, with the development of society, more and more researchers began to focus on point cloud registration in real scenes. However, the geometrically challenging areas with repetitive patterns and low geometry commonly exist in real scenes, causing failure in point matching followed by inaccurate point cloud registration.

In this thesis, this thesis propose a robust point cloud registration approach that embeds the geometry of salient anchors to enhance the discriminative ability of the point features even in the presence of a large number of repetitive patterns and low-geometry areas in the source and target point clouds. Specifically, an anchor location module is designed to locate corresponding superpoints with the most discriminative and the richest geometric information as salient anchors in the source and target. Non-maximum suppression is adopted to ensure the salient anchors are structure-preserved and sparsely distributed. With salient anchors, a selectively geometric structure embedding of anchorsuperpoint distances and angles is proposed for superpoint feature enhancement. This integration of geometry consistency between the salient anchors and superpoints can improve the distinction of features in those geometrically challenging areas. Afterwards, the enhanced features and anchor positions are updated in an iterative manner to acquire the most effective salient anchors and descriptive superpoint features. The updated features allow for accurate superpoint matches. Finally, accurate point correspondences are achieved by finding the nearest neighbour points within superpoints.

In addition, this thesis proposes a point cloud registration method based on multi-modal fusion for anchor location, which improves the feature diversity of geometrically challenging regions by fusing the structural features of the point cloud and the texture fea-

ABSTRACT

tures of the image. First, this paper uses the alignment module to align the point cloud and image data to find the correspondence between superpoints and pixels. Then, a fusion module is used to fuse the features between the superpoints and the corresponding pixels. The fusion module projects the point cloud features and image features into two subspaces that are modality-independent and modality-dependent, and fuses the two modality features successively in the two subspaces to reduce the impact of domain differences and prevent information loss role.

Keywords: Point cloud registration, Geometry embedding, Low-geometry area, Repetitive patterns, Multimodal fusion

目 录

| | |
|------------------------------|-----|
| 摘 要 | I |
| ABSTRACT | II |
| 图目录 | VI |
| 表目录 | VII |
| 第1章 绪论 | 1 |
| 1.1 研究背景及意义 | 1 |
| 1.2 国内外研究现状 | 2 |
| 1.2.1 传统点云配准方法 | 3 |
| 1.2.2 基于深度学习的点云配准方法 | 4 |
| 1.2.3 存在的问题 | 6 |
| 1.3 论文研究的主要内容 | 6 |
| 1.4 论文组织结构 | 7 |
| 第2章 点云配准相关理论 | 9 |
| 2.1 本章引言 | 9 |
| 2.2 刚体变换基础 | 9 |
| 2.2.1 表示形式 | 9 |
| 2.2.2 变换矩阵求解 | 12 |
| 2.3 卷积神经网络 | 15 |
| 2.4 数据集介绍 | 16 |
| 2.5 评价指标 | 17 |
| 2.6 本章小结 | 18 |
| 第3章 基于点云语义分割域适应的主动学习方法 | 19 |
| 3.1 本章引言 | 19 |
| 3.2 研究动机及贡献 | 19 |
| 3.3 基于原型指导的主动学习方法 | 20 |
| 3.3.1 问题陈述 | 20 |
| 3.3.2 方法概述 | 21 |
| 3.3.3 源域原型构建 | 21 |
| 3.3.4 源域原型指导的数据选择 | 22 |
| 3.3.5 动态混合中间域构建 | 24 |
| 3.3.6 实验评估 | 25 |

| | |
|---------------------------------------|-----------|
| 3.3.7 实验结果..... | 26 |
| 3.3.8 与其他主动学习方法对比 | 31 |
| 3.3.9 消融实验..... | 32 |
| 3.3.10 本章小结..... | 32 |
| 第 4 章 基于多模态特征融合的锚点定位点云配准 | 33 |
| 4.1 本章引言 | 33 |
| 4.2 基于多模态特征融合的锚点定位点云配准方法 | 34 |
| 4.2.1 对齐模块..... | 35 |
| 4.2.2 多模态特征提取 | 36 |
| 4.2.3 融合模块..... | 36 |
| 4.2.4 损失函数..... | 38 |
| 4.3 实验结果与分析..... | 39 |
| 4.3.1 数据集预处理..... | 39 |
| 4.3.2 3DMatch 和 3DLoMatch 实验..... | 39 |
| 4.3.3 消融实验..... | 44 |
| 4.4 本章小结 | 45 |
| 第 5 章 总结与展望 | 46 |
| 5.1 主要结论 | 46 |
| 5.2 研究展望 | 46 |
| 参考文献 | 48 |
| 作者简介 | 55 |
| 1. 攻读学位期间的研究成果 | 55 |
| (一) 发表的学术论文和著作..... | 55 |
| (二) 申请(授权)专利..... | 55 |
| (三) 参与的科研项目及获奖..... | 55 |
| 致 谢 | 56 |

图目录

| | |
|---|----|
| 图 1-1 不同时间同一场景的点云与图像对比 | 1 |
| 图 1-2 端到端的点云配准方法流程图 | 4 |
| 图 1-3 基于对应关系的点云配准方法流程图 | 5 |
| 图 3-1 基于点云语义分割域适应的主动学习方法框架 | 21 |
| 图 3-2 源域原型构建 | 22 |
| 图 3-3 源域原型指导的数据选择 | 23 |
| 图 3-4 动态混合中间域构建 | 25 |
| 图 3-5 第三章 SynLiDAR→SemanticKITTI 分割可视化图 | 27 |
| 图 3-6 第三章 SynLiDAR→SemanticPOSS 分割可视化图 | 28 |
| 图 3-7 第三章 SemanticKITTI→nuScenes 分割可视化图 | 29 |
| 图 3-8 第三章 nuScenes→SemanticKITTI 分割可视化图 | 30 |
| 图 4-1 第四章方法示意图 | 34 |
| 图 4-2 第四章方法框架图 | 35 |
| 图 4-3 特征融合模块流程图 | 37 |
| 图 4-4 配准结果可视化 | 43 |

表目录

| | |
|--|----|
| 表 3-1 第三章方法与其他域适应方法在 SynLiDAR→SemanticKITTI 数据上 的比较 | 26 |
| 表 3-2 第三章方法与其他域适应方法在 SynLiDAR→SemanticPOSS 数据上 的比较 | 27 |
| 表 3-3 第三章方法与其他域适应方法在 SemanticKITTI→nuScenes 数据上的 比较 | 29 |
| 表 3-4 第三章方法与其他域适应方法在 nuScenes→SemanticKITTI 数据上的 比较 | 30 |
| 表 3-5 SAPL | 31 |
| 表 3-6 本章的主动学习方法与其他传统主动学习方法在结合 Mixing 后的对 比 | 32 |
| 表 3-7 第三章方法消融实验 | 32 |
| 表 4-1 第四章方法与先进方法的内点率 | 40 |
| 表 4-2 第四章方法与先进方法的匹配召回率的比较 | 40 |
| 表 4-3 本方法与先进方法的配准召回率的比较 | 41 |
| 表 4-4 在 3DMatch 和 3DLoMatch 上使用不同姿态估计器的配准结果 | 42 |
| 表 4-5 3DMatch 和 3DLoMatch 的相对平移误差和相对旋转误差比较 | 42 |
| 表 4-6 对齐模块消融实验 | 44 |
| 表 4-7 融合模块消融实验 | 44 |

表目录

第1章 绪论

1.1 研究背景及意义

近些年来，计算机视觉的发展让我们的生活越来越便利。与此同时，随着我国信息化水平和自动化进程的不断提高与推进，国家对计算机视觉的应用和发展也提出了更高要求。相比于二维图像，三维点云数据能更加清晰的表示我们所处的三维世界。图 1-1 显示了两对激光雷达点云和图像的例子，它们分别取自于不同时间的同一场景。可以清晰的观察到，点云的几何结构在光照和季节变化的情况下能够保持基本不变，而图像的变化使得人眼也难以分辨出这对图像来自于同一场景。由于点云数据具有光照不变性，能够有效避免图像处理过程中的问题，因此越来越多的研究人员开始研究点云并从中受益。但是三维视觉传感器的视野范围是有限的，因此为了感知全局的环境，在应用中经常需要将若干不同位置采集的点云数据对齐到世界坐标系下。因此，点云配准已成为许多任务的基础问题，近年来备受关注。

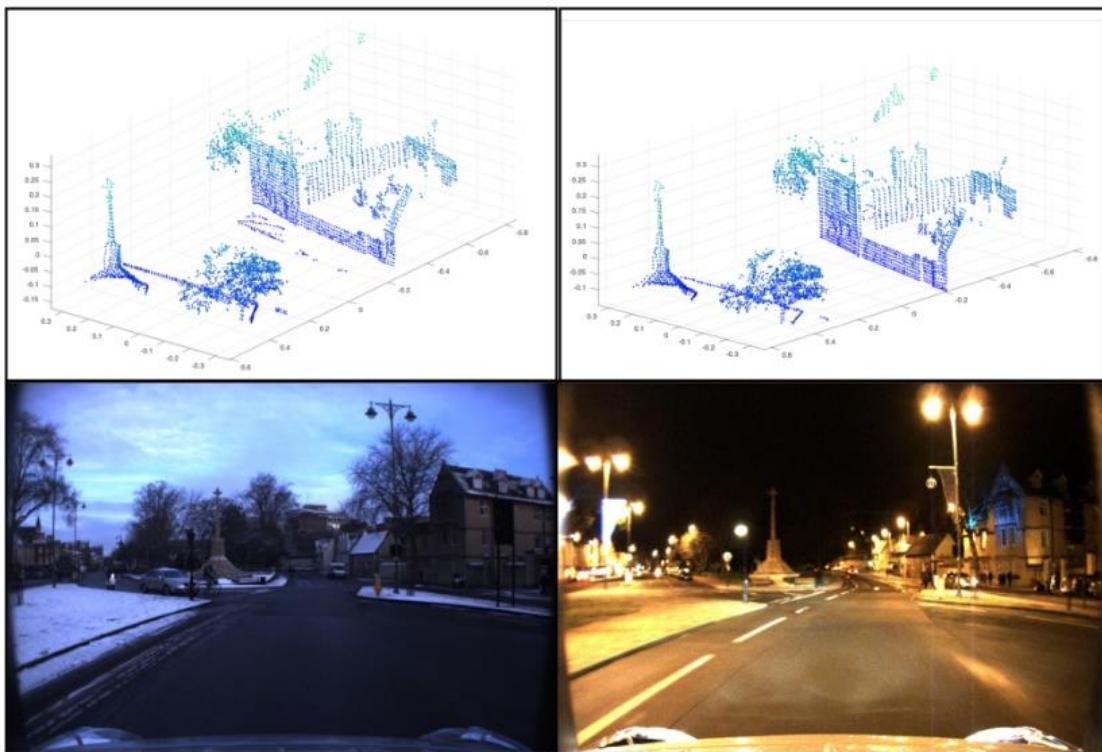


图 1-1 不同时间同一场景的点云与图像对比^[1]

Fig. 1-1 Comparison between point cloud and image of the same scene at different time^[1]

在医疗领域，使用点云配准能够将患者的器官及内脏等组织融合成一个整体

构建三维模型，辅助医生诊断。在文物修复领域，研究人员能够对大型文物多次扫描后将其配准生成完整的文物模型，不仅能够存储文物的三维数据以实现文物的数字化存储，还能够为后续文物的保护和修复提供可靠的数据^[2]。在工业领域，点云配准是逆向工程的重要一环，能够为研究人员提供产品模型的三维数据^[3]。更重要的是，点云配准是许多机器人任务的关键组成部分，是机器人对环境感知的重要一环^[4]。在同时定位与建图（Simultaneous Localization And Mapping, SLAM）中，点云配准可以构建用于自动驾驶的路径规划和决策的三维地图^[5]。点云配准也广泛应用于位置识别，它能够将实时的三维视图匹配到所属的三维地图中，以实现机器人对自身位置的定位^[6]。同时点云配准在机器人的姿态估计中也发挥着不可或缺的作用。通过对齐视图与环境，可以获取机器人手臂的姿态信息，进而决定下一步该如何移动以抓取物体^[7]。

点云配准是三维图形学研究的一个重要研究课题，也是计算机视觉的重要分支，其目的是将两个具有部分重叠区域的点云，经过一个变换矩阵在同一个坐标系下对齐。早期，不少学者提出了许多不同的方法来解决点云配准问题，这些方法主要针对于实验室理想环境下的合成数据集，而这些数据集也往往由单一的物体模型而不是场景构成。随着社会的发展，这些方法已经不能够满足生产生活需要。近年来越来越多的工作开始关注真实环境下的点云配准，其中基于深度学习的点云配准方法有着突出的表现。最近，针对低重叠率情况的点云配准得到了学界的广泛关注。所谓低重叠率的点云配准指的是将两个至多只存在 30% 重叠区域的点云进行对齐。相比于一般情况的点云配准，低重叠率的点云对之间存在许多相似非重叠区域，这会很大程度上增加特征搜索寻找正确点对应的难度，导致大量非匹配区域的误匹配。由此可见低重叠情况下的点云配准的研究重点在于如何将点的对应关系聚集在重叠区域和如何增加相似区域的特征差异。

综上所述，点云配准是处理点云数据的一项基本任务，是推进自动化进程的关键一环，在生产生活中扮演着重要的角色。因此，研究点云配准算法，提高配准精度和减少算法时间复杂度具有重大的研究价值。同时，随着近年来人工智能深度学习的快速发展，基于深度学习的点云配准方法也取得了巨大成功。本文通过对现有方法研究进行分析并改进，提高在低重叠度的情况下点云配准算法的成功率。

1.2 国内外研究现状

本节将从传统点云配准方法和基于深度学习的点云配准方法两个方向介绍点云配准方法的研究现状。其中传统点云配准方法早在上世纪 90 年代就得到了初步发展，而后在研究人员的不懈努力下传统方法的点云配准现在已经广泛应用于工

业生产领域。随着近些年来的人工智能的发展，将点云配准与与深度学习融合也逐渐受到越来越多的研究人员的关注，并且其性能上已经超过传统的点云配准方法。

1.2.1 传统点云配准方法

迭代式最近点法^[8] (Iterative Closet Points, ICP) 是传统点云配准方法中最典型的一类，它由 BESL 等人在 1992 年提出。ICP 方法通过计算源点云和目标点云原始点之间的欧氏距离，以最临近点作为对应点确定两点云间点的对应关系。然后，在已知点的对应关系时，通过基于奇异值分解法求解两点云之间的变换矩阵，并进行单次对齐。重复上述两个步骤直至满足预设要求完成整个配准过程。当源点云与目标点云之间具有良好的初始位姿时，ICP 方法可以取得良好的配准结果。但是，当二者之间的距离较大时，该算法往往会在局部最优点处收敛，导致最终的结果不能满足实际生产生活需要。文献 [9] 提出一种模拟退火算法将 ICP 算法中根据点到点的距离确定的“硬”匹配关系转化为一种“软”匹配方式。这种方法虽然不能完全避免局部最优解问题，但是能够使算法在一定程度上得到缓解。YANG 等人^[10] 为了解决上述问题，提出一种全局最优的迭代式最近点算法 (Globally Optimal Iterative Closet Point, Go-ICP)，其基本思想是通过分支界定法跳出局部最优解，以实现全局最优解，但与此同时算法的速度严重下降。

基于图的配准是另一类常见的方法，它主要是寻找更加准确的对应关系。相比于 ICP 方法中直接选取最邻近点作为对应点，基于图的配准方法将同时考虑点和边的关系。具体而言，图匹配算法不仅要求匹配点的点相似度高，而且要求节点之间的连线即边的相似度也要高。这种关系能够找到更准确的对应关系，而精确的对应关系有助于更好的变换估计。图匹配的优化属于二次分配问题，是一个典型的 NP 难问题，解决思想主要采取近似策略逼近。文献 [11] 和文献 [12] 采用线性规划来解决图匹配问题。文献 [13] 则将较大的相似矩阵分解为若干较小矩阵，这些矩阵对每个图的局部结构和相似性进行编码，解耦节点和边之间的相似性，使得求解过程简化。LERDEANU 等人^[14] 提出了一种谱松弛的方法来近似二次分配问题，它指出正确的对应能够形成强关联的集群，而错误的对应只是一种偶然，因此不太可能产生强关联的簇，根据这一特性能够有效找出正确的对应关系。

高斯混合模型 (Gaussian Mixture Models, GMM) 是另一种常见的点云配准方法，它的核心思想是将配准问题中变换估计问题转化为求解点云数据的最大似然估计问题。任意两点之间的对应关系，将由原来的“硬”匹配转化为了由置信度表示的“软”匹配，但也因此其时间复杂度大大增加了。JRMPC 等人^[15] 提出了一个 EM 算法，它估计了 GMM 参数以及将每个独立集合映射到“中心”模型上的旋转和平移。文献 [16] 通过最大似然估将源点云数据拟合到目标点云。使源点云作为

一个整体移动，以保持点集的拓扑结构。通过对具有刚性参数的 GMM 质心位置的重新参数化来施加一致性约束，并推导出 EM 算法的最大步骤的封闭解。文献 [17] 提出了一种新的凸包索引高斯混合模型。该模型通过计算每个点集凸包上的加权高斯混合模型响应来工作。

传统的点云配准算法的优点有两个方面：（1）严格的数学理论可以保证算法的收敛性；（2）不需要训练数据。然而这类方法的局限性也较为明显：ICP 算法需要一个较好的初始位置才能有良好的表现，这在真实场景下尤其是低重叠情况下难以达到要求；这类方法对数据的离群值、噪声和点云密度等特点较为敏感，在真实场景下的表现远没有在合成数据上的表现好。

1.2.2 基于深度学习的点云配准方法

基于深度学习的点云配准方法大致可以分为两类：端到端的点云配准方法和基于对应关系的点云配准方法。

1.2.2.1 端到端的点云配准方法

文献 [18] 提出的 PointNetLK 可以被认为是一个可学习的函数。因此，将用于图像对齐的经典视觉算法 Lucas-Kanade^[19] (LK) 算法与其相结合，并融合为一个循环深度神经网络。文献 [20] 赋予 PPF-FoldNet 自动编码器 (Auto Encoder, AE) 一个姿态差异结构，其中两者之间的差异产生特定于姿态的描述符。在此基础上，引入了相对姿态估计网络 RelativeNet，为关键点分配对应特定的方向。最后，利用一个简单而有效的假设-验证算法来快速预测和对齐两个点云。FMR^[21] 借鉴了 PointNetLK 的思想，利用刚性变换的可逆特性，采用编解码器结构监督全局特征。文献 [22] 提出了一种基于全局特征的迭代网络 OMNet，用于部分重叠点云的配准。OMNet 以由粗到细的方式学习掩码来拒绝非重叠区域，这将部分重叠的配准转换为相同形状的配准。此外，它提出了一种更实用的数据生成方式，其中 CAD 模型对源点云和参考点云进行两次采样，避免了普遍存在的过拟合问题。

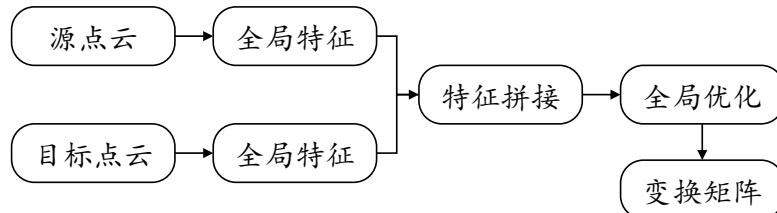


图 1-2 端到端的点云配准方法流程图

Fig. 1-2 Flowchart of the end-to-end point cloud registration

1.2.2.2 基于对应关系的点云配准方法

基于对应关系的配准方法主要集中在四个方面：特征提取、关键点检测、异常值去除和位姿估计。QI 等人先后提出了 PointNet^[23] 和 pointnet++^[24]。这两种方法虽然为点云的特征提取提供了参考，但都没有考虑点云的几何结构特征。文献 [25] 提出了一种利用弱监督学习三维特征检测器和描述子进行点云匹配的 3DFeat-Net。与许多以往的工作不同，该方法不需要手动标注匹配的点。相反，可以利用对齐和注意力机制从全球定位系统（Global Positioning System, GPS）标记的 3D 点云中学习特征对应关系，而无需人工标注。文献 [26] 提出了 3DSmoothNet，它利用暹罗网络架构匹配 3D 点云，并使用体素化平滑密度值表示实现全卷积层，并与局部参考系对齐以实现旋转不变性。文献 [27] 提出了一种新的神经网络模块 EdgeConv，构造了 DGCNN 来捕获点之间的拓扑信息。EdgeConv 作用于网络每一层中动态计算的图，其中包含了局部邻域信息，并可以叠加应用于学习全局形状属性。文献 [28] 提出 KPConv 来模拟二维卷积中的运算，以更好地捕获局部几何信息。文献 [29] 提出的 3DMatch 网络以体素为输入，利用三维卷积神经网络学习局部几何特征。文献 [30] 提出全卷积几何特征采用稀疏三维卷积代替传统的三维卷积来缓解点云稀疏性带来的问题。SpinNet^[31] 通过估计的参考轴约束 z 轴自由度，并使用球面体素化消除 XY 平面旋转自由度，提取具有高鲁棒性的特征。文献 [32] 提出的 D3feat 在提取点云特征时使用 KPConv 组成的 U-Net 网络来检测关键点，并使用密度不变显著性评分来缓解密度对显著性的影响。文献 [33] 提出了一种点云配准模型 Predator，该模型对重叠区域进行了深度关注。与以前的工作不同，该模型是专门设计来处理低重叠的点云对的。其核心思想是在两个点云的潜在编码之间进行早期信息交换的重叠注意块，以预测哪些点不仅是显著的，而且还位于两个点云之间的重叠区域。

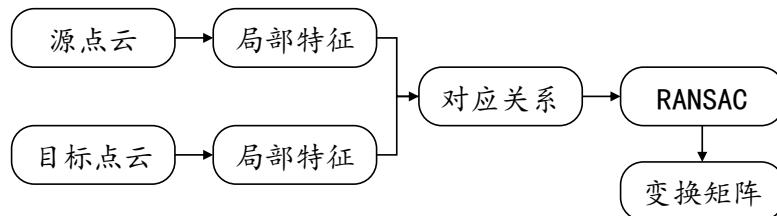


图 1-3 基于对应关系的点云配准方法流程图

Fig. 1-3 Flowchart of point cloud registration method based on correspondence

文献 [34] 提出 PointDSC，将传统方法中的空间几何一致性约束添加到网络中，利用神经网络提取对应关系的特征，通过可微的光谱匹配模块，对成对的空间一致性监督，以估计每个对应的嵌入特征的置信度。文献 [35] 提出一个由点云嵌入网

络、结合注意力模块近似组合匹配层、可微奇异值分解层三部分组成的 DCP 网络，解决局部最优和 ICP 方法中的其他问题。IDAM^[36] 包括一个迭代的距离感知相似矩阵卷积模块，将来自特征和欧几里得空间的信息合并到成对点匹配过程中。这些卷积层学习基于整个几何特征的联合信息和每个点对的欧几里得偏移来匹配点，克服了通过简单地使用特征向量的内积来匹配的缺点。文献 [37] 提出了一个可微分的框架 DGR。它由三个模块组成：用于对应置信度预测的 6 维卷积网络，用于封闭姿态估计的可微分加权 Procrustes 算法，以及用于姿态细化的鲁棒基于梯度的 SE(3) 优化器。

1.2.3 存在的问题

点云配准方法的研究已经有了二十多年的发展并且取得了一系列的成就，特别是在深度学习流行起来之后，许多研究利用深度神经网络取得了不错的成果。然而，在最近的相关文献 [38] 中已经提到这些基于深度学习的方法的主干网络往往遇到高层特征的过度平滑和结构模糊性相关的难以区分的特征问题，这是点云配准的一个关键瓶颈。它们忽略了特征提取的一个关键因素，这可能严重影响配准精度：源点云和目标点云中每个点特征的独特性；也就是说，为了获得精确的点对应关系，以估计最优刚性变换，所需的点特征应该充分表示任何给定点附近的几何模式，同时仍然与同一点云中围绕其它点的局部结构特征有足够的差异性。然而，许多工作^[39,40] 使用的骨干网络容易导致特征的超平滑和结构性的模糊问题，导致点特征难以区分。同时，将图像数据中的颜色和纹理信息引入进来通过多模态的融合，增加点特征之间的差异性是一个简单有效的想法。但是在相关研究中，许多多模态融合的方法并没有显示出比单模态的方法更加优异的表现。如何更有效的融合点云和图像两种模态的特征也是一个值得研究的问题。

1.3 论文研究的主要内容

本研究分别从嵌入几何结构和多模态融合两个方面入手，提高特征间的差异性，提出如下两个点云配准的方法。

1. 基于显著锚点几何嵌入的点云配准方法。首先使用共享参数的骨干网络来提取源点云和目标点云的局部特征，在特征提取过程中同时对点云下采样进行超点聚合。之后在超点层面选取若干分布于重叠区域的特征显著的锚点。同时，提出了一种新颖的基于注意力机制的几何嵌入方法。它的核心思想是将每个超点与锚点间的距离与角度信息进行编码，由于选取的多个锚点在空间中保持了一定的几何结构，因而使得这些超点与锚点之间的几何编码能够给每个超点带来各不相同的差异性特征。然后通过一个迭代优化模块，选取更加显著的锚点进一步作结

构嵌入增加特征差异性，并形成最终的超点对应。这些超点在物理空间中表示一个连续空间的区域，通过上采样能够建立超点与原始点之间的包含关系。这种由粗到细的配准方法可以在对应区域内部寻找点的对应关系而无需在全局点云中寻找，能够有效缓解特征平滑带来的误匹配，进而在变换估计中产生更加准确的变换矩阵。

2. 基于多模态特征融合的锚点定位点云配准方法。现有的点云和图像两种模态融合方法往往通过对图像进行特征提取之后，将其与点云进行简单拼接并送入神经网络完成特征融合。与这些方法不同的是，本方法采用一种对齐策略，利用相机参数将点云和图像完成点与像素的对齐之后分别提取点云特征和图像特征。并利用点与像素之间的对应关系，通过交叉注意力机制完成像素到点的选择性融合。在此过程中，两种模态的特征均会被映射至模态无关与模态相关的两个特征子空间。在模态无关子空间中，点云和图像完成模态间特征的融合，随后将融合后的特征与点云在模态相关子空间中的投影相融合。这种方法能够有效减少点云和图像之间的域间隙，使得融合过程既不过多的引入噪声也不丢失互补信息，形成最终的超点特征中。

1.4 论文组织结构

为了更加清晰地阐述本文的主要工作，本文结构安排如下：

第1章为绪论部分。首先对本文的选题背景和研究意义进行了介绍。然后，对国内外学者在点云配准研究领域取得的一些成就和研究进展进行了简单的陈述，并对目前点云配准方法中存在的问题进行了分析探讨。最后，对本文的主要研究做了简单的介绍，并阐述了本文的组织结构。

第2章为相关技术理论基础。首先对点云配准任务进行介绍。然后，介绍了点云配准任务中常用的用于求解刚体变换矩阵的基于SVD的线性代数法和基于RANSAC^[41]的随机一致性采样法。接着，对点云配准的数据集做了详细介绍，主要包括合成数据集和真实场景数据集。最后，介绍了用于评估点云配准算法的性能的评价指标。

第3章是基于显著锚点几何嵌入的点云配准方法。首先介绍了本文提出方法的主要动机。接下来介绍了基于显著锚点几何嵌入的框架网络，介绍了如何从下采样后的超点中选取出位于重叠区域的保持一定几何结构的显著锚点。接下来介绍了如何利用自注意力和交叉注意力机制嵌入超点与超点之间以及超点与锚点之间的几何结构特征。之后对基于由粗到细框架的点云配准方法的点匹配阶段和变换估计方法做了简单介绍。接着介绍了该网络分别用来监督粗匹配阶段产生的超点

匹配结果与细匹配阶段产生的点匹配结果的两个损失函数。然后是实验部分，对实验用到的数据集，实验设置以及实验结果及分析做了详细的描述。最后对第三章进行了总结。

第4章是基于多模态的锚点定位点云配准。首先介绍了当前方法将点云与图像两种模态融合的一些问题和多模态融合之前加入对齐模块的动机。然后简要介绍了本文对点云和图像两种模态数据进行特征提取的网络结构。之后介绍了对齐模块在整个网络中的作用与功能，主要是消除在数据增强情况下点云和图像数据的错位。然后介绍了多模态融合模块，利用一个简单的映射网络将两种模态的特征在模态无关子空间进行融合，并在模态相关子空间进一步补充相关信息的过程。接下来是实验部分，从数据集的预处理，实验设计和结果分析三个方面来进行阐述。最后对本章进行了总结。

第5章是总结与展望。首先对本文所作的工作进行了总结，然后根据本文的实验结果指出了方法中存在的问题和未来的研究方向。

第2章 点云配准相关理论

2.1 本章引言

首先，本章将对数字空间如何表示三维空间的物体及运动做出简要介绍，这是研究点云配准方法的前提。同时，当前点云配准方法与人工智能相关技术深度融合，因此本章将对深度神经网络作出必要介绍，这是研究点云配准方法的基础。之后，本章将介绍若干用于评估算法性能的公开数据集，包括合成数据集和真实场景数据集，这是验证点云配准算法的基础。最后，本章将对用于评价算法优劣的各项指标做出简要介绍。

2.2 刚体变换基础

2.2.1 表示形式

2.2.1.1 变换矩阵

三维变换最常见的表示是变换矩阵，三维变换矩阵由旋转和平移两部分组成。首先在只考虑旋转变换的情况下，对于任意向量 \mathbf{a} 而言，已知它在两个单位正交基 $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, $(\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)$ 下的坐标分别是 (a_1, a_2, a_3) , (a'_1, a'_2, a'_3) ，其中 $(\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)$ 由 $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ 经过旋转得到。因为向量本身并没有发生变化，所以可以用公式 (2-1) 表示 \mathbf{a} 与 \mathbf{a}' 之间的关系：

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 & \mathbf{e}'_3 \end{bmatrix} \begin{bmatrix} a'_1 \\ a'_2 \\ a'_3 \end{bmatrix} \quad (2-1)$$

对公式 (2-1) 的左右两边同时左乘 $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)^T$ ，将左边系数矩阵化简为单位矩阵可以得到公式 (2-2)，它表示了坐标 (a_1, a_2, a_3) 与坐标 (a'_1, a'_2, a'_3) 之间的转换关系：

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1^T \mathbf{e}'_1 & \mathbf{e}_1^T \mathbf{e}'_2 & \mathbf{e}_1^T \mathbf{e}'_3 \\ \mathbf{e}_2^T \mathbf{e}'_1 & \mathbf{e}_2^T \mathbf{e}'_2 & \mathbf{e}_2^T \mathbf{e}'_3 \\ \mathbf{e}_3^T \mathbf{e}'_1 & \mathbf{e}_3^T \mathbf{e}'_2 & \mathbf{e}_3^T \mathbf{e}'_3 \end{bmatrix} \begin{bmatrix} a'_1 \\ a'_2 \\ a'_3 \end{bmatrix} \quad (2-2)$$

将公式 (2-2) 右边的系数矩阵称为旋转矩阵 \mathbf{R} 。旋转矩阵由两组基之间的内积组成，能够表示任意向量在旋转前后的坐标变化关系。可以注意到旋转矩阵是正交矩阵，

因此上述旋转变换的逆变换可由公式 (2-3) 表示:

$$\mathbf{a}' = \mathbf{R}^{-1}\mathbf{a} = \mathbf{R}^T\mathbf{a} \quad (2-3)$$

为了更准确的描述刚体在三维空间中的运动，现在将平移分量 \mathbf{t} 引入进来，公式 (2-4) 表示了物体的旋转和平移过程:

$$\mathbf{a} = \mathbf{R}\mathbf{a}' + \mathbf{t} \quad (2-4)$$

公式 (2-4) 虽然能够准确的表示刚体在三维空间中的单次运动变换，但是当需要连续表示多次变换时往往就会包含多个括号并显得不够简洁。故引入齐次坐标和变换矩阵，公式 (2-4) 的齐次表达式为公式 (2-5):

$$\begin{bmatrix} \mathbf{a} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}' \\ 1 \end{bmatrix} \quad (2-5)$$

公式 (2-5) 等式左边第一个矩阵被定义为变换矩阵 **Trans**，它主要由表示旋转的旋转矩阵 \mathbf{R} 和表示平移的平移向量 \mathbf{t} 两部分组成。

变换矩阵虽然能够精确的表示刚体在三维空间中的运动，但是这种表现形式也存在着某些局限性：(1) 对于旋转矩阵而言，需要有 9 个变量描述物体在三维空间中的旋转，而旋转只存在 3 个自由度；对变换矩阵而言，则需要 16 变量表示物体的旋转和平移，即使这些运动只存在 6 个自由度。因此，以变换矩阵来表示物体在三维空间中的运动是冗余的，这也引出旋转矩阵的另一个局限性；(2) 变换矩阵变量之间存在约束关系，由上述公式 (2-2) 可以看到变换矩阵的旋转矩阵部分是一个行列式为 1 的正交矩阵。这种约束关系对于变换矩阵的计算求解和优化而言会使得问题变得麻烦复杂。

2.2.1.2 旋转向量

旋转向量是表示刚体在三维空间中运动的另一种常见形式。它通过两个变量将三维空间中的旋转参数化，单位向量 \mathbf{n} 表示旋转轴的方向，角度 θ 描述围绕旋转轴的旋转幅度。通过罗格里格斯公式可以完成旋转向量和变换矩阵之间的转化。对于任意向量 \mathbf{v} 绕单位向量 \mathbf{n} 旋转 θ 角度后得到 \mathbf{v}' 。可以将向量 \mathbf{v} 相对于 \mathbf{n} 分解成平行分量 \mathbf{v}_{\parallel} 和垂直分量 \mathbf{v}_{\perp} ，由公式 (2-6) 表示:

$$\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp} \quad (2-6)$$

式中，平行分量 \mathbf{v}_{\parallel} 和垂直分量 \mathbf{v}_{\perp} 分别可以用公式(2-7)和公式(2-8)表示：

$$\mathbf{v}_{\parallel} = (\mathbf{v} \cdot \mathbf{n})\mathbf{n} \quad (2-7)$$

$$\mathbf{v}_{\perp} = \mathbf{v} - \mathbf{v}_{\parallel} = \mathbf{v} - (\mathbf{v} \cdot \mathbf{n})\mathbf{n} = -\mathbf{n} \times (\mathbf{n} \times \mathbf{v}) \quad (2-8)$$

根据投影关系可以得到 \mathbf{v}' 和 \mathbf{v} 的平行分量和垂直分量之间的关系，分别用公式(2-9)和公式(2-10)表示：

$$\mathbf{v}'_{\parallel} = \mathbf{v}_{\parallel}, \quad (2-9)$$

$$\mathbf{v}'_{\perp} = \cos(\theta)\mathbf{v}_{\perp} + \sin(\theta)\mathbf{n} \times \mathbf{v}_{\perp} \quad (2-10)$$

又因为 \mathbf{n} 与 \mathbf{v}' 平行，因此可以进一步得到公式(2-11)：

$$\mathbf{n} \times \mathbf{v}'_{\perp} = \mathbf{n} \times (\mathbf{v} - \mathbf{v}_{\parallel}) = \mathbf{n} \times \mathbf{v} - \mathbf{n} \times \mathbf{v}_{\parallel} = \mathbf{n} \times \mathbf{v} \quad (2-11)$$

将公式(2-11)代入公式(2-10)可知公式(2-12)：

$$\mathbf{v}'_{\perp} = \cos(\theta)\mathbf{v}_{\perp} + \sin(\theta)\mathbf{n} \times \mathbf{v} \quad (2-12)$$

进而 \mathbf{v}' 可由公式(2-13)表示：

$$\begin{aligned} \mathbf{v}' &= \mathbf{v}_{\parallel} + \cos(\theta)\mathbf{v}_{\perp} + \sin(\theta)\mathbf{n} \times \mathbf{v} \\ &= \mathbf{v}_{\parallel} + \cos(\theta)(\mathbf{v} - \mathbf{v}_{\perp}) + \sin(\theta)\mathbf{n} \times \mathbf{v} \\ &= \cos(\theta)\mathbf{v} + (1 - \cos(\theta))\mathbf{v}_{\perp} + \sin(\theta)\mathbf{n} \times \mathbf{v} \\ &= \cos(\theta)\mathbf{v} + (1 - \cos(\theta))(\mathbf{n} \cdot \mathbf{v})\mathbf{n} + \sin(\theta)\mathbf{n} \times \mathbf{v} \end{aligned} \quad (2-13)$$

将公式(2-13)重新组合并化简可得到最终旋转向量与变换矩阵之间的关系，可以得到公式(2-14)：

$$\mathbf{T} = \cos(\theta)\mathbf{I} + (1 - \cos(\theta))\mathbf{n}\mathbf{n}^T + \sin(\theta)\mathbf{n}^\wedge \quad (2-14)$$

2.2.1.3 四元数

与变换矩阵和旋转向量相比，单位四元数是一种更加紧凑、高效且数值稳定的描述刚体在三维空间运动变换的表达形式。虽然，它并不直观，并且由于三角函数的周期性，不同的旋转角度可能会编码为相同的四元数。但是只要将其弧度限制在 $[0, 2\pi]$ ，四元数不失为一种良好的形式。单位四元数可以通过引入抽象符号 $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 来定义，它们满足规则 $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$ ，同时满足除乘法交换律以

外的常用代数规则。设 $v = [0, \mathbf{v}]$, $u = [0, \mathbf{u}]$, 易得公式 (2-15) 与公式 (2-16):

$$vu = [-\mathbf{v} \cdot \mathbf{u}, \mathbf{v} \times \mathbf{u}], \quad (2-15)$$

$$uv_{\perp} = [-\mathbf{u} \cdot \mathbf{v}_{\perp}, \mathbf{u} \times \mathbf{v}_{\perp}] \quad (2-16)$$

又因为 \mathbf{v}_{\perp} 与 \mathbf{u} 正交, 所以 $\mathbf{u} \cdot \mathbf{v}_{\perp} = 0$, 将之带入公式 (2-16), 可得公式 (2-17):

$$uv_{\perp} = [0, \mathbf{u} \times \mathbf{v}_{\perp}] = \mathbf{u} \times \mathbf{v}_{\perp} \quad (2-17)$$

进一步可得到公式 (2-18):

$$\begin{aligned} v'_{\perp} &= \cos(\theta)v_{\perp} + \sin(\theta)(uv_{\perp}) \\ &= (\cos(\theta) + \sin(\theta)u)v_{\perp} \end{aligned} \quad (2-18)$$

令 $q = \cos(\theta) + \sin(\theta)u$, 可得 $v'_{\perp} = qv_{\perp}$, 令 $q = p^2$, 有 $p = [\cos(0.5\theta), \sin(0.5\theta)\mathbf{u}]$ 。又因为 $qq^{-1} = qq^* = 1$, $qv_{\parallel} = v_{\parallel}q$, $qv_{\perp} = v_{\perp}q^*$, 将其带入公式 (2-13) 可得公式 (2-19):

$$v' = pvp^* = pvp^{-1} \quad (2-19)$$

2.2.2 变换矩阵求解

对于基于对应关系的点云配准方法而言, 在得到源点云和目标点云之间的点对应关系之后, 需要将这种点的对应关系转化成点云之间的变换矩阵, 以得到的最终的配准结果。由于本文在不同阶段利用了奇异值分解法 (SVD) 和随机一致性估计法 (RANSAC), 本节将对这两种方法进行介绍。

2.2.2.1 奇异值分解法

设集合 $\mathcal{C} = \{(\mathbf{p}_i, \mathbf{q}_i) | i = 1, \dots, n\}$, 其中 \mathbf{p}_i 是源点云 \mathcal{P} 中与目标点云 \mathcal{Q} 中的 \mathbf{q}_i 点相对应的点。在已知所有对应点的对应关系 \mathcal{C} 后, 目标点云与源点云之间的变换问题可以表示成公式 (2-20):

$$\mathbf{R}, \mathbf{t} = \operatorname{argmin}_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}} \sum \|\mathbf{q}_i - (\mathbf{R} \cdot \mathbf{p}_i + \mathbf{t})\|^2 \quad (2-20)$$

对公式 (2-20) 右边求导并令其等于 0 有公式 (2-21):

$$\begin{aligned} 0 &= \sum_{i=1}^n 2(\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i) \\ &= 2n\mathbf{t} + 2\mathbf{R}(\sum_{i=1}^n \mathbf{p}_i) - 2 \sum_{i=1}^n \mathbf{q}_i \end{aligned} \quad (2-21)$$

公式(2-21)左右边同时除以 n 并化简可以得到公式(2-22):

$$0 = 2\mathbf{t} + \frac{2\mathbf{R}(\sum_{i=1}^n \mathbf{p}_i)}{n} - \frac{2\sum_{i=1}^n \mathbf{q}_i}{n} = 2\mathbf{t} + 2\mathbf{R}\bar{\mathbf{p}} - 2\bar{\mathbf{q}} \quad (2-22)$$

式中, $\bar{\mathbf{p}}$ 和 $\bar{\mathbf{q}}$ 分别是源点云和目标点云的形心。将公式(2-22)化简可得:

$$\mathbf{t} = \bar{\mathbf{q}} - \mathbf{R}\bar{\mathbf{p}} \quad (2-23)$$

将公式(2-23)带入(2-20)可知公式(2-24):

$$\begin{aligned} \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}} \|\mathbf{q}_i - (\mathbf{R} \cdot \mathbf{p}_i + \mathbf{t})\|^2 &= \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}} \|(\mathbf{q}_i - \bar{\mathbf{q}}) - \mathbf{R}(\mathbf{p}_i - \bar{\mathbf{p}})\|^2 \\ &= \sum \|\mathbf{R}\mathbf{p}'_i - \mathbf{q}'_i\|^2 \end{aligned} \quad (2-24)$$

根据矩阵 Frobenius 范数(F-范数)的定义,矩阵F-范数的平方可以转化成矩阵的内积形式,进而得到矩阵迹的表达形式,然后再带入公式(2-24)化简得到公式(2-25):

$$\sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}} \|\mathbf{R}\mathbf{p}'_i - \mathbf{q}'_i\|^2 = \mathbf{p}'_i{}^T \mathbf{R}^T \mathbf{R} \mathbf{p}'_i - \mathbf{q}'_i{}^T \mathbf{R} \mathbf{p}'_i - \mathbf{p}'_i{}^T \mathbf{R}^T \mathbf{q}'_i + \mathbf{q}'_i{}^T \mathbf{q}'_i \quad (2-25)$$

又因为 \mathbf{R} 是正交矩阵且 $\mathbf{q}'_i{}^T \mathbf{R} \mathbf{p}'_i = (\mathbf{q}'_i{}^T \mathbf{R} \mathbf{p}'_i)^T = \mathbf{p}'_i{}^T \mathbf{R}^T \mathbf{q}'_i$, 所以公式(2-25)可化简为公式(2-26):

$$\sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}} \|\mathbf{R}\mathbf{p}'_i - \mathbf{q}'_i\|^2 = \mathbf{p}'_i{}^T \mathbf{p}'_i - 2\mathbf{q}'_i{}^T \mathbf{R} \mathbf{p}'_i + \mathbf{q}'_i{}^T \mathbf{q}'_i \quad (2-26)$$

将公式(2-26)代入公式(2-24)并化简有公式(2-27):

$$\mathbf{R} = \operatorname{argmin} \left(\sum_{i=1}^n \mathbf{p}'_i{}^T \mathbf{p}'_i - \sum_{i=1}^n 2\mathbf{q}'_i{}^T \mathbf{R} \mathbf{p}'_i + \sum_{i=1}^n \mathbf{q}'_i{}^T \mathbf{q}'_i \right) \quad (2-27)$$

又因为 $\sum_{i=1}^n p'_i{}^T p'_i$ 和 $\sum_{i=1}^n q'_i{}^T q'_i$ 对旋转矩阵 \mathbf{R} 求导恒等于 0, 故可是公式(2-28):

$$\mathbf{R} = \operatorname{argmin} \left(\sum_{i=1}^n \mathbf{q}'_i{}^T \mathbf{R} \mathbf{p}'_i \right) = \operatorname{argmin} (tr(\mathbf{R} \mathbf{P}' \mathbf{Q}'{}^T)) \quad (2-28)$$

记矩阵 $\mathbf{S} = \mathbf{P}' \mathbf{Q}'{}^T$, 对之进行奇异值分解可得公式(2-29):

$$\mathbf{S} = \mathbf{U} \Sigma \mathbf{V}^T \quad (2-29)$$

式中, \mathbf{U} 和 \mathbf{V} 是 $n \times 3$ 阶酉矩阵, Σ 是 3×3 阶半正定对角矩阵, 且其对角线上的元素是 \mathbf{S} 的奇异值。将公式 (2-29) 带入公式 (2-28) 有公式 (2-30):

$$\mathbf{R} = \operatorname{argmin}(tr(\mathbf{R}\mathbf{U}\Sigma\mathbf{V}^T)) = \operatorname{argmin}(tr(\Sigma\mathbf{V}^T\mathbf{R}\mathbf{U})) \quad (2-30)$$

又因为 $\mathbf{V}, \mathbf{R}, \mathbf{U}$ 均为正交矩阵, 因此 $\mathbf{V}^T\mathbf{R}\mathbf{U}$ 也为正交阵, 又因为正交阵每个元素的绝对值小于等于 1, 奇异值大于等于 0, 因此上式当且仅当对角线上的元素均为 1 时取最大值, 可以得到公式 (2-31):

$$\mathbf{R} = \mathbf{V}\mathbf{U}^T \quad (2-31)$$

将公式 (2-31) 带入公式 (2-23) 可以得到 \mathbf{t} 的解:

$$\mathbf{t} = \bar{\mathbf{q}} - \mathbf{V}\mathbf{U}^T\bar{\mathbf{p}} \quad (2-32)$$

2.2.2.2 随机一致性估计法

随机一致性估计是一种迭代方法, 用于从一组包含离群值的观察数据中利用随机抽样来估计数学模型的参数。它是一种非确定性算法, 它仅以一定的概率产生正确的结果, 并且随着迭代次数的增加, 产生正确结果概率也会随之增加。使用 RANSAC 方法从一组包含离群值的对应关系中估计出集合所对应的变换矩阵, 具体流程如下:

- (1) 从源点云和目标点云之间的点的对应关系集合 \mathcal{C} 中选取距离大于阈值的三对点对应。
- (2) 利用三对点对应关系求解变换矩阵 \mathbf{T}_i
- (3) 将变换矩阵 \mathbf{T}_i 应用至源点云, 并计算变换后的源点云与目标点云之间所有对应点的距离之和, 记为该次估计的误差。

不断重复上述三个步骤直至满足迭代次数要求, 选取出误差最小的变换矩阵 \mathbf{T} 作为最终的估计结果, 其中参数迭代次数 k 主要由内点率 w 确定。假设 3 对点对应的选择是独立同分布的, 那么在一次选择中 3 对对应关系均正确的概率为 w^3 。所以一次选择中至少存在一个异常对应的概率为 $1 - w^3$, 这意味着估计出错误的变换矩阵的概率。在经过 k 次估计之后, 在这 k 次预测中至少存在一次成功估计的概率为:

$$p = 1 - (1 - w^3)^k \quad (2-33)$$

化简公式 (2-33), 那么可以得出公式 (2-34) 表示 k :

$$k = \frac{\log(1 - p)}{\log(1 - w^3)} \quad (2-34)$$

RANSAC 的一个优点是它能够对模型参数进行鲁棒估计，即使存在大量异常值的情况下，它也可以高精度地估计变换矩阵。RANSAC 的一个缺点是计算这些参数所需的时间没有上限。当计算的迭代次数有限时，获得的解决方案可能不是最优的，甚至可能无法良好拟合数据。

2.3 卷积神经网络

卷积神经网络，它的输入层和输出层之间存在多个非线性映射的隐藏层。通过若干非线性映射的隐藏层的叠加，深度神经网络能过够有效的拟合任意的函数，能够对真实世界的问题进行数学建模。通过计算输出层产生的结果与实际结果之间的差距，并最小化数据集的整体差距，网络能够以反向传播的方式优化整个网络的参数并完成网络的整体训练。当前，卷积神经网络主要由卷积层、全连接层、激活函数和池化层等一些基本结构组成。同时，无论是在 2 维图像领域还是在 3D 点云领域，关注像素或者点之间的上下文关系的注意力机制已经被广泛应用。综上，本节将逐一介绍相关基本结构。

(1) 卷积层

卷积层指的是使用预先固定尺寸的卷积核与输入特征进行卷积滤波操作的网络结构，整个输入共享同一组卷积核，卷积核通过在输入特征上进行滑窗操作以遍历整个输入特征。卷积层主要由如下参数：卷积核个数、卷积核尺寸和卷积步长。卷积核个数决定卷积层输出特征的维度，而卷积核尺寸决定该层卷积的感受野大小，卷积步长决定滑窗操作的步长。

(2) 全连接层

全连接层指的是将来自上一层的所有输入都连接到下一层的每一个激活单元的网络结构，下一层输出的特征的每一个特征值均为上一层的全部特征通过权重加权以及添加偏移值之后得到。在常见的二维图像处理任务中，全连接层常被应用于网络结构的最后几层，它将之前的网络结构提取的数据特征进行降维或升维以形成最终输出；在常见的三维图形任务中，全连接层常被用于直接特征提取或是作为基本模块来提取三维图形输入的局部特征。

(3) 激活函数

激活函数是穿插在多层全连接层或卷积层之间的关键函数。如果不存在激活函数，那么多层全连接层或卷积层的叠加与一层全连接层或者卷积层的效果是一样的。主要原因是失去了激活函数的非线性映射能力，多层线性映射的叠加本质上就是一层线性映射。因此在深度神经网络中激活函数是不可或缺的，同时，设计更加有效的激活函数也是深度神经网络的关键。当前，研究领域主要采用的激活

函数包括：Sigmoid、Tanh、ReLU、Leaky ReLU 和 ELU 等。

(4) 注意力机制

在现实世界当中，通过眼睛我们可以观察到各种各样的事物，从而能够感知到大量的信息。此外，因为我们具备对信息进行筛选的能力，所以可以根据实际情况来选择重要的信息，而忽略不重要的信息，以避免受海量信息的干扰。从这一角度出发，深度学习研究者希望网络也能够具备与我们相同的能力，所以在网络当中引入了注意力机制。通过注意力机制的方式，网络可以对输入特征进行加权之后再输出，以希望网络对重要的特征给较大的权重，对不太重要的特征给较小的权重，使得网络具备了对特征进行筛选的能力。

2.4 数据集介绍

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-G 相机获取的深度图通过三维成像得到。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好，并且点云数据分布均匀。当前常见的用于点云配准的数据集包括：ModelNet40、3DMatch 和 KITTI。

(1) ModelNet40

普林斯顿大学提出的 ModelNet40 数据集包含来自 40 个类别的 12311 个对齐 CAD 模型，其中有 80% 共 9843 个数据用作训练，剩余 20% 的数据用于测试。ModelNet40 数据集包含 40 个类别的三维模型，其中既有桌子、花瓶、飞机这样具有规则对称结构的点云，也有吉他、花、人这样结构复杂的点云。ModelNet40 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。

(2) 3DMatch

3DMatch 数据集包含 62 个不同场景的 RGB-D 数据。每个场景被分成几个片段，每个片段使用 TSDF 融合算法从 50 个深度图中重建三维点云。最终整个数据集有 54 个场景用于训练，8 个场景用于测试。

(3) KITTI

KITTI 数据集最初设计用于立体匹配性能评估，包括立体序列、激光雷达点云和地面真实姿态。数据集包含 10 条完整采集轨迹，市中心的交通、住宅区，以及德国卡尔斯鲁厄周围的高速公路场景和乡村道路，共标注 28 个类，包括区分非移动

对象和移动对象的类，即地面、建筑、车、人、物体等大类。原始数据包括 22 个序列组成，序列 00 到 10 作为训练集共 23201 个数据，11 到 21 作为测试集共 20351 个数据。

2.5 评价指标

根据点云配准方法的基本步骤，为了评估点云配准不同阶段性能因采取不同的评价指标。同时，针对不同数据集的特点不同，在某些评价方式上也存在些许差异。本节将介绍针对特征提取的评价指标和针对刚体运动估计的评价指标两大类指标进行介绍。其中，针对特征提取的评价指标包括内点率（IR），特征匹配召回率（FMR）和匹配召回率（RR）；针对刚体运动估计的评价指标包括均方根误差（RMSE）和相对平移误差（RTE）、相对旋转误差（RRE）。

(1) 内点率 (Inlier Ratio, IR) 测量通过网络预测的假定点对应集合中正确对应所占的比例。所谓正确对应，即源点云中的点经过真实变换运动后，与目标点云中的对应点的距离小于某一阈值 τ_1 。给定源点云 \mathcal{P} 和目标点云 \mathcal{Q} 的待评价的对应集合 \mathcal{C} ，内点率的数学表达式为公式 (2-36):

$$\text{IR}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}} [\|\bar{\mathbf{T}}_{\mathcal{P}}^{\mathcal{Q}} \mathbf{p}_i - \mathbf{q}_i\|_2 < \tau_1] \quad (2-35)$$

式中， $\bar{\mathbf{T}}_{\mathcal{P}}^{\mathcal{Q}}$ 表示源点云与目标点云的之间的真实变换矩阵， \mathbf{p}_i 和 \mathbf{q}_i 是一对对应点。

(2) 特征匹配召回率 (Feature Matching Recall, FMR) 测量的是内点率大于某一阈值 τ_2 的点云对的比例。它表明了整个数据集中可以通过鲁棒姿态估计器 RANSAC 恢复两个点云之间的变换矩阵的点云对的比例。给定一个数据集的所有测试集的点云对集合 \mathcal{D} ，特征匹配召回可以用公式 (2-37) 表示：

$$\text{FMR} = \frac{1}{|\mathcal{D}|} \sum_{\mathcal{D}} [\text{IR} > \tau_2] \quad (2-36)$$

(3) 均方根误差 (Root Mean Square Error, RMSE) 与配准召回率 (Registration Recall, RR) 与上述度量对应关系质量的指标不同，RR 直接度量点云配准目标任务的性能。它测量的是均方根误差在某一阈值 τ_3 内的点云对的比例。给定一个数据集的所有测试集的点云对集合 \mathcal{D} ，配准召回定义为公式 (2-38):

$$\text{RR} = \frac{1}{|\mathcal{D}|} \sum_{\mathcal{D}} [\text{RMSE} < \tau_3] \quad (2-37)$$

式中，RMSE 由公式 (2-39) 表示为：

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{C}|} \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}} \|\mathbf{T}_{\mathcal{P}}^Q(\mathbf{p}_i) - \mathbf{q}_i\|^2} \quad (2-38)$$

(4) 相对平移误差 (RTE) 和相对旋转误差 (RRE)

相对平移和旋转误差 (RTE/RRE) 测量与真实变换矩阵之间的偏差。给定预测的变换 $\mathbf{T}_{\mathcal{P}}$ ，其平移向量和旋转矩阵分别是 \mathbf{t} 和 \mathbf{R} 。其相对平移误差 (RTE) 和相对旋转误差 (RRE) 相对于真实位姿 $\mathbf{T}_{\mathcal{P}}^Q$ 的表示分别为公式 (2-40)，公式 (2-41)：

$$\text{RTE} = \|\mathbf{t} - \bar{\mathbf{t}}\| \quad (2-39)$$

$$\text{RRE} = \arccos\left(\frac{\text{tr}(\mathbf{R}^T \bar{\mathbf{R}}) - 1}{2}\right) \quad (2-40)$$

式中， $\bar{\mathbf{t}}$ 与 $\bar{\mathbf{R}}$ 分别是 $\mathbf{T}_{\mathcal{P}}^Q$ 中真实的平移分量和旋转分量。

2.6 本章小结

本章主要介绍了点云配准方法中所用到的相关技术的基础知识。首先介绍了三维空间中的物体运动在数字空间中的三种表现形式以及在已知对应关系的条件下求解变换矩阵的常见的两种方法。接下来介绍了对本文将要使用的深度神经网络的基础知识做了简要介绍。然后介绍了常见的用于评估点云配准算法性能的数据集，包括合成数据集与真是数据集，室内数据集与室外数据集。最后介绍了点云配准方法研究中常使用的几种评价指标，分别是内点率、特征匹配召回率、均方根误差与配准召回率以及相对平移误差和相对旋转误差。

第3章 基于点云语义分割域适应的主动学习方法

3.1 本章引言

本章主要介绍一种新的用于点云语义分割域适应的主动学习方法，该方法提出了一种基于源域原型指导的主动查询策略，并结合 Mixing 方法构建出强壮的中间域数据，极大的提高了点云语义分割域适应模型的性能。接下来，本章节将首先介绍方法的研究动机和贡献，接着对每个子方法模块的原理及实现细节进行详细的介绍，最后本文将通过在主流公开数据集上取得的实验结果以及对消融实验的分析展示方法的有效性。

3.2 研究动机及贡献

近年来，三维点云语义分割技术在自动驾驶、智能机器人等领域的应用需求日益迫切。尽管基于深度学习的全监督方法在点云语义分割上取得了显著性能，但其实际部署面临两大核心挑战：标注成本瓶颈与域间分布差异。一方面，点云的逐点标注需耗费大量人力物力，标注单帧车载激光雷达点云需约 2 小时^[42]。而真实场景中目标域数据因传感器配置、环境动态变化等因素，与源域存在显著分布偏移，导致模型泛化性能急剧下降^[43]。

为缓解标注压力，现有研究主要沿两条路径展开：主动学习通过选择最具代表性的样本进行标注，以最小标注代价提升模型性能；无监督域适应则尝试在无目标域标注下对齐源域与目标域特征分布。然而，两者在跨域场景中均存在固有局限。传统主动学习方法^[44-46]的样本选择策略都假定在单模态源域分布上，忽略了潜在的多模态分布，因此选择的样本无法有效指导域间特征对齐，这会导致标注资源浪费并影响模型性能^[47,48]。无监督虽无需目标域标注，但其依赖于大量的伪标签，而伪标签噪声会随迭代过程累积，限制性能提升，导致其与全监督基线仍存在很大的差异。

针对上述存在的问题，一些学者开始致力于主动域适应方法研究^[49-51]，其结合主动学习和域适应方法的优势，并在图像语义分割领域取得了一定的成果，然而些方法大多不能直接应用于点云语义分割。一方面，点云数据具有独特的几何结构和稀疏性，与图像数据的网格结构有本质区别；另一方面，点云的数据量庞大且无序，直接迁移图像领域的主动域适应方法无法有效捕捉点云的关键特征。在三维点云中，Annotator^[52]提出了一种以体素为中心的主动学习方法，用以选择显著且具有代表性的体素，并随后对这些体素内的所有点进行标注。它第一次将主动学习

运用到三维点云语义分割域适应中，并取得了超越其他传统主动学习方法的效果，然而它只考虑了点云特性依然忽略了域间差异。

通过上述分析，本章提出基于点云语义分割域适应的主动学习方法，核心思想是通过源域原型指导目标域上点的选择，并将标注后的目标点与源域点进行混合，组成中间域数据，实现标注效率与域对齐能力的协同优化。构建以下两个模块：1) 域差异感知的主动查询：通过动态构建源域类别原型以代表源域，计算目标域中候选点与源域的偏离程度，筛选同时具备高不确定性与高域差异的目标点。这些样本能够精准暴露域间分布边界，指导模型聚焦于域偏移敏感区域。2) 动态中间域构建：引入 Mixing 方法，随机从源域中采样一定比例的标注点与已标注的目标点云进行混合增强，生成兼具双域信息的中间域数据，该方法增强模型对域不变特征的提取能力，可以进一步缩减域间隙。

本章节研究内容的主要贡献如下：

- 1) 提出了以个面向点云语义分割域适应的主动学习方法，超过传统主动学习方法下的点云语义分割域适应结果，并在极少量的标注下取得了超过最先进方法的结果。
- 2) 提出了一种源域指导的目标点主动选择策略，筛选出兼具高不确定性和高域差异的目标点。
- 3) 首次将主动学习与 Mixing 方法进行结合并运用到域适应领域，动态构建包含双域信息的中间域数据，进一步缩减域间隙。

3.3 基于原型指导的主动学习方法

3.3.1 问题陈述

在点云语义分割域适应任务中，给定一个标注的源域数据 $\mathcal{S} = \{(\mathbf{X}_i^s, \mathbf{Y}_i^s)\}_{i=1}^{N^s}$ ，和一个目标域数据 $\mathcal{T} = \{(\mathbf{X}_i^t)\}_{i=1}^{N^t}$ ，其中 N^t 和 N^s 分别代表目标域和源域点云帧的数量， $\mathbf{X}_i \in \mathbb{R}^{n_i \times 4}$ 代表含有 (x, y, z, i) 三维坐标点和反射强度的一帧点云集合，并且 n_i 代表在第 i 帧点云中点的数量。域适应的目标则是在主动学习方法的帮助下，将从标注的源域上训练好的模型 \mathcal{G} 迁移泛化到目标域数据上来，并在目标域上实现准确的点云语义分割。

在主动域适应场景中，给定一个未标注的目标域数据集 \mathcal{T} ，需要从中筛选出能够代表目标域且信息量最大的数据子集进行标注。利用新标注的目标域数据，源域训练的分割模型可逐步迭代调整以适应目标域分布，最终实现目标域上的精确语义分割。具体流程如下：首先，基于预训练模型 \mathcal{G} 对目标域数据的预测结果，设计查询策略以计算每个未标注目标点的代表性度量得分；随后，选择最具代表性的

目标点子集 \mathcal{T}_l 进行标注，并利用其参与分割模型 \mathcal{G} 的调优，同时更新未标注目标数据集 $\mathcal{T} = \mathcal{T} - \mathcal{T}_l$ 。该过程循环迭代直至达到预设的主动学习预算 B 。

3.3.2 方法概述

本方法的总体框架如图3-1所示。其算法流程主要由三个模块构成：①源域原型构建：通过分割模型从源点云中提取特征，并基于这些特征构建源域原型，并将作为源域的语义表征代表源域。②源域原型指导的数据选择：计算未标记的目标点到原型的特征空间距离，得到原型相似度特征图，并根据最优-次优差异算法生成域差异分数，并将此分数与模型预测的不确定性分数结合，生成最终评分以指导标注候选目标点的选择。③动态混合中间域构建：使用 Mixing 方法，随机从源域中采样一定比例的标注点与已标注的目标点进行混合增强，生成兼具双域信息的中间域数据，并将此数据用于分割模型的微调。

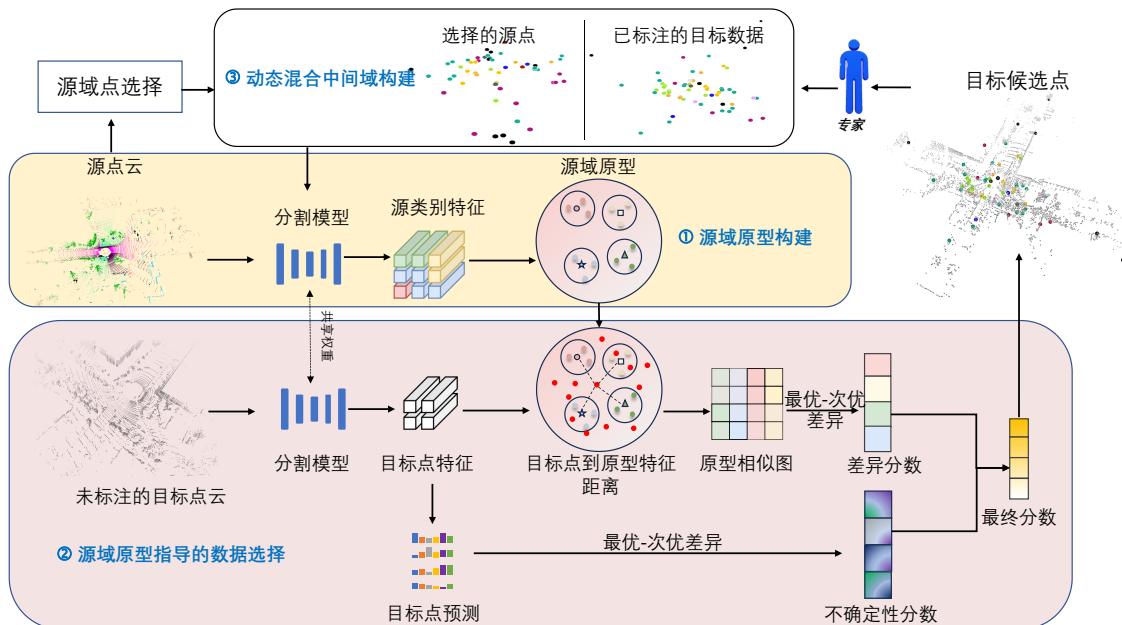


图 3-1 基于点云语义分割域适应的主动学习方法框架

Fig. 3-1 Framework of the active learning method for domain adaptation in point cloud semantic segmentation

3.3.3 源域原型构建

为表征源域数据分布的结构特征，首先需要基于特征中心构建源域类别原型。具体而言，对于源域中每个类别的点云样本，利用分割模型 \mathcal{G} ，提取其特征向量，并将同类特征向量的均值作为该类的原型表征。如图3-2所示，将标注的源域数据集 \mathcal{S} 输入当前网络 \mathcal{G} ，提取特征矩阵 $\mathbf{F} \in \mathbb{R}^{N_P^S \times d_f}$ ，其中 N_P^S 表示源域中所有标注类

别点的数量, d_f 为特征维度, 基于源数据类别信息, 通过公式(3-1)所示计算类别原型 \mathbf{p}^c :

$$\mathbf{p}^c = \frac{\sum_{i=1}^{N_c} \mathbf{f}_i^c}{N_c} \quad (3-1)$$

其中, $c \in [1, C]$ 表示类别索引, C 为源域类别总数; \mathbf{f}_i^c 为类别 c 中第 i 个点的特征向量; N_c 为类别 c 的样本数量, 对应的公式如(3-2)所示:

$$N_c = \sum_{i=1}^{N_P^S} \mathbb{I}(y_i^s = c) \quad (3-2)$$

其中 y_i^s 表示第 i 个源点特征 \mathbf{f}_i 对应的类别标签, $\mathbb{I}(y_i^s = c)$ 为指示函数, 当 y_i^s 属于类别 c 时取值为 1, 否则为 0。

由于点云数据集庞大, 因此常规服务器设备无法一次性将所有的数据都加载到内存并进计算, 而如果设计全局变量累加多次迭代的结果, 可能会造成一定的精度损失和内存消耗, 为了节省内存资源和保证结果的准确性, 本章采用 Welford 增量均值算法^[53] 进行渐进式的原型计算, 如公式(3-3)所示:

$$\mathbf{p}_{b+1}^c = \mathbf{p}_b^c + \frac{\mathbf{p}_{b+1}^c - \mathbf{p}_b^c}{x_{b+1}} \quad (3-3)$$

式中, \mathbf{p}_{b+1}^c 代表在第 $b+1$ 训练批次时类别 c 的原型结果, \mathbf{p}_b^c 代表上一批次时类别 c 的原型结果, x_{b+1} 则代表在第 $b+1$ 训练批次时类别 c 的总数量。通过该算法最终得到源域原型矩阵 $\mathbf{P} \in \mathbb{R}^{C \times d_f} = \{\mathbf{p}^i\}_{i=1}^C$, 其中每个原型向量 \mathbf{p}^i 对应特征空间中源域某类别的质心, 蕴含该类别的语义特征。这些原型将在每一轮的主动学习阶段动态更新, 为跨域数据选择提供指导。

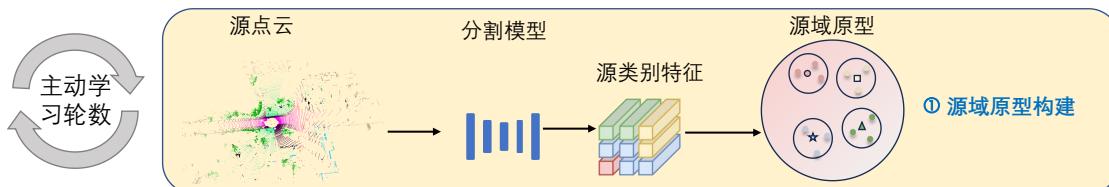


图 3-2 源域原型构建

Fig. 3-2 Source prototype construction

3.3.4 源域原型指导的数据选择

源域原型构建完成后, 可将其作为基准指导目标域数据的筛选。如图3-3所示, 将未标注的目标域点云输入预训练分割网络, 提取其特征矩阵 $\mathbf{F}^T \in \mathbb{R}^{N^T \times d_f}$, 其中 N^T 为目标域点数。随后逐点计算其与源域各类别原型的欧氏距离, 其表达式公式

为(3-4):

$$\mathbf{d}_i^c = \|\mathbf{f}_i^t - \mathbf{p}^c\| \quad (3-4)$$

式中 $\mathbf{f}_i^t \in \mathbf{F}^T$ 表示目标域第 i 个未标注点的特征向量, \mathbf{p}^c 为源域类别 c 的原型向量。 d_i^c 代表点到源域类别 c 的欧式距离, 该距离度量反映了目标点特征与源域类别质心的空间匹配度: 距离越小, 表明目标点特征在源域特征空间中越接近类别 c 的聚类中心。

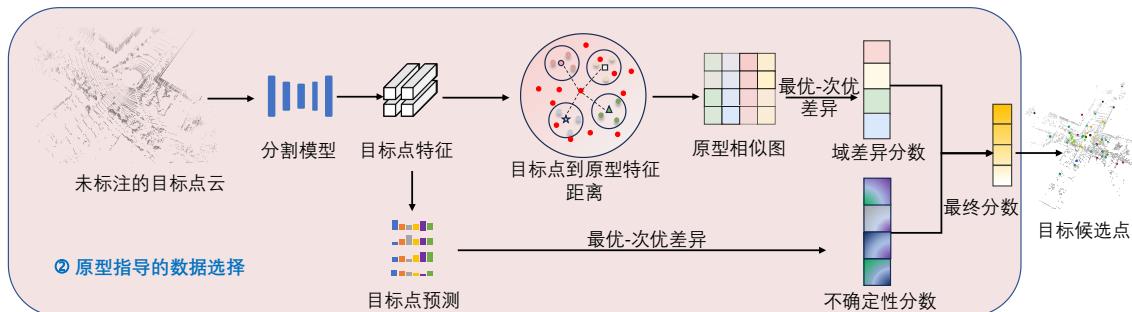


图 3-3 源域原型指导的数据选择

Fig. 3-3 Source-Prototype guided data selection

3.3.4.1 计算域差异评分

对每个目标点 i 遍历计算与所有类别原型的特征距离, 生成距离向量 $\mathbf{D}_i = [d_i^1, d_i^2, \dots, d_i^C]$ 。为量化其域间分布特性, 需将距离向量转换为相似性度量并进行归一化处理:

1) 相似性转换: 对 \mathbf{D}_i 逐元素取倒数, 得到相似性向量 $\mathbf{D}'_i = [1/d_i^1, 1/d_i^2, \dots, 1/d_i^C]$, 使得距离越近的类别相似度值越高。

2) 概率归一化: 通过 Softmax 函数 $f(\cdot)$ 将 \mathbf{D}'_i 映射为概率分布, 如公式(3-5)所示, 消除量纲差异增强判别性, 确保不同类别的距离在同一尺度下比较。

$$f(\mathbf{D}'_i) = \text{softmax}(\mathbf{D}'_i) = \left[\frac{e^{d_i^{1'}}}{\sum_{c=1}^C e^{d_i^{c'}}}, \dots, \frac{e^{d_i^{C'}}}{\sum_{c=1}^C e^{d_i^{c'}}} \right] \quad (3-5)$$

最终通过最优-次优差异算法计算域差异评分 M_{ds}^i , 公式如(3-6)所示:

$$M_{ds}^i = S_{R1}(f(\mathbf{D}'_i)) - S_{R2}(f(\mathbf{D}'_i)) \quad (3-6)$$

其中 $S_{R1}(\cdot)$ 和 $S_{R2}(\cdot)$ 分别表示最大概率值与次大概率值。 M_{ds}^i 越小, 则表明目标点与两个源域类别的相似度接近, 意味着其处于源域类别边界区域, 这样的目标点对缓解域偏移具有更高价值。

3.3.4.2 融合不确定性评分

计算得到域差异评分后，为了避免选择的点都是同类别的点，融合不确定性评分进行最终筛选。通过分割头 $h(\cdot)$ 获取目标点的预测概率分布 $\mathbf{p}_i^t \in \mathbb{R}^{1 \times C}$ ，并计算其不确定性评分 M_{us}^i ，其表达公式如公式(3-7)所示：

$$M_{us}^i = S_{R1}(\mathbf{p}_i^t) - S_{R2}(\mathbf{p}_i^t) \quad (3-7)$$

该评分反映模型对目标点类别判定的置信度：当最大概率与次大概率差值较小时（即 M_{us}^i 较小），表明模型对该点的预测存在较高不确定性，其位于类别边界上无法区分，此类样本的标注可有效提升模型性能。为平衡域差异特性与模型不确定性，采用加权融合策略生成最终评分，其表达式如公式(3-8)所示：

$$M_{final}^i = \alpha \times M_{ds}^i + (1 - \alpha) \times M_{us}^i \quad (3-8)$$

其中， $\alpha \in [0, 1]$ 可调节的超参数。当 $\alpha = 0.5$ 时，两类评分贡献均等；当目标域与源域分布差异显著时，可增大以强化域差异指导作用，因此在不同的场景下 α 可能会有所不同。

3.3.4.3 筛选候选样本

基于最终目标评分 M_{final}^i ，对所有目标点进行升序排列，评分越低优先级越高，选取每帧点云中排名前 k 的点组成候选样本点集 \mathcal{T}_l 提交至专家（Oracle）进行人工标注，同时更新未标注目标数据集 $\mathcal{T} = \mathcal{T} - \mathcal{T}_l$ ，在下一轮的主动学习中，已标注的目标点将不参与筛选，其中 k 与主动学习轮数 R 以及标注总预算 B 的关系如公式(3-9)所示：

$$k = \frac{B}{R \times N^T} \quad (3-9)$$

3.3.5 动态混合中间域构建

最后，为了进一步增强模型的泛化能力，本算法引入了 Mixing 混合方法。如图3-4所示，在每一轮训练中任意一帧目标点云都会随机匹配一帧源点云，并随机从源点云中选择一定比例的源点 \mathbf{S}_{s_i} ，将这些选择后的源点与已标注的目标点 \mathbf{T}_{a_i} 进行混合，组成含两域信息的中间域数据 $\mathbf{I}_i = concat(\mathbf{T}_{a_i}, \mathbf{S}_{s_i})$ ，其中 $concat(\cdot)$ 代表拼接混合。这些中间域数据可以帮助模型学习到域不变特征，进一步缩减域间隙^[54,55]。

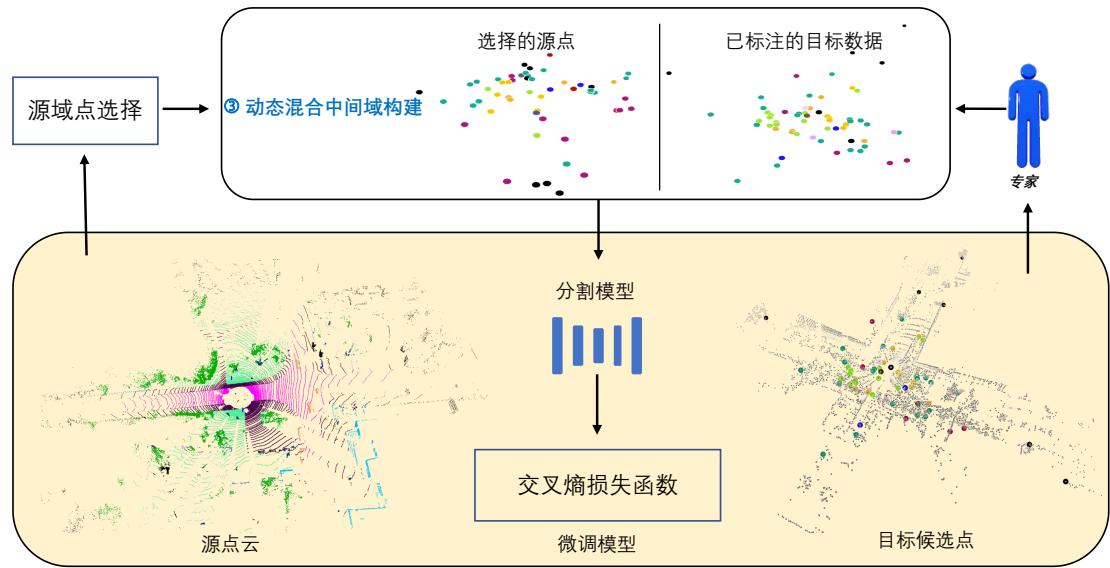


图 3-4 动态混合中间域构建

Fig. 3-4 Dynamic mixed intermediate domain construction

3.3.6 实验评估

3.3.6.1 实验设置

在实验中，使用 MinkowskiNet^[56] 在 Annotator 中的 PyTorch 实现版本作为目标分割网络模型，并使用随机梯度下降 (SGD)^[57] 作为学习优化器，动量为 0.01，权重衰减系数为 0.0001，在第一轮训练中使用线性预热，将学习率线性增加到基础学习率 0.01，并使用初始学习率为 0.01 的余弦衰减调度器动态调整学习率。所有实验都在单张 NVIDIA RTX A6000 GPU 上进行训练。本章方法在真实到真实以及合成到真实的跨域场景下分别进行了实验。对于所有的场景，主动学习全程总共执行 5 次迭代并达到预设标注预算。在合成到真实的跨域场景实验中，预训练模型 (Source-only) 和全监督模型 (Target-only) 阶段的批量大小设为 16；在 SynLiDAR→SemanticKITTI 实验中训练 10 轮，在 SynLiDAR→SemanticPOSS 实验中训练 20 轮。而在域适应阶段，每一步的批量大小设为 14 并训练 50 轮。权重参数 α 分别在 SynLiDAR→SemanticKITTI 和 SynLiDAR→SemanticPOSS 的实验中设置为 0.4 和 0.6。在真实到真实的跨域场景实验中，SynLiDAR→nuScenes 和 nuScenes→SemanticKITTI 的实验配置相同，预训练模型 (Source-only) 和全监督模型 (Target-only) 阶段的每一步批量大小设为 16，并训练 10 轮；在域适应阶段的每一步批量大小设置为 10 并训练 50 轮，权重参数 α 为 0.6。

3.3.7 实验结果

为了证明本章方法的有效性，分别在“合成到真实”和“真实到真实”这两个跨域场景下，对四个数据集进行了实验。随后，通过可视化手段对实验结果进行展示，从而更直观地呈现所提出方法在不同场景中的具体效果。

3.3.7.1 合成到真实场景

在“合成到真实”跨域场景的实验中，为确保与现有域适应方法的公平比较，选择 SynLiDAR→SemanticKITTI 和 SynLiDAR→SemanticPOSS 这两个主流的合成到真实跨域数据集进行了实验。

表 3-1 本方法与其他域适应方法在 SynLiDAR→SemanticKITTI 数据上的比较

Table 3-1 Comparison with other domain adaptation methods on SynLiDAR→SemanticKITTI

| 方法 | 域适应 | 标注 | 结果 |
|-------------|-----|------|-------------|
| Source-Only | - | - | 22.8 |
| Target-Only | - | 100% | 60.1 |
| AADA | UDA | - | 23.0 |
| AdvEnt | UDA | - | 25.8 |
| CRST | UDA | - | 26.5 |
| ST-PCt | UDA | - | 28.9 |
| PolarMix | UDA | - | 32.2 |
| CoSMix | UDA | - | 31.0 |
| DGT-ST | UDA | - | 43.1 |
| Annotator | ADA | 0.1% | 57.7 |
| 本章方法 | ADA | 0.1% | 58.7 |

实验 SynLiDAR→SemanticKITTI 的结果如表3-1所示：实验结果表明，在全监督基准测试中，Source-Only 模型与 Target-Only 模型间的性能差达到 37.3 个百分点，直观的反映了合成数据与真实场景间的域间分布差异，表明直接迁移模型到目标域会因为域偏移而导致严重的性能下降。在无监督域适应（UDA）的方法中，各方法性能分布在 22.8% 至 43.1% 区间，其中 DGT-ST 方法以 43.1% 取得最高结果，但相较目标域全监督性能仍存在 17 个百分点的差距。在三维语义分割主动域适应（ADA）方法中，Annotator 是唯一可比较的方法，为保证公平比较，主动学习预算参考其设置为 0.1%，Annotator 与本章 ADA 方法的性能分别达到目标域全

监督基准的 96% 与 97.7%，证明本章方法有效性的同时也表明了主动学习域适应的高效性。提供本数据集下分割可视化如图3-5所示：

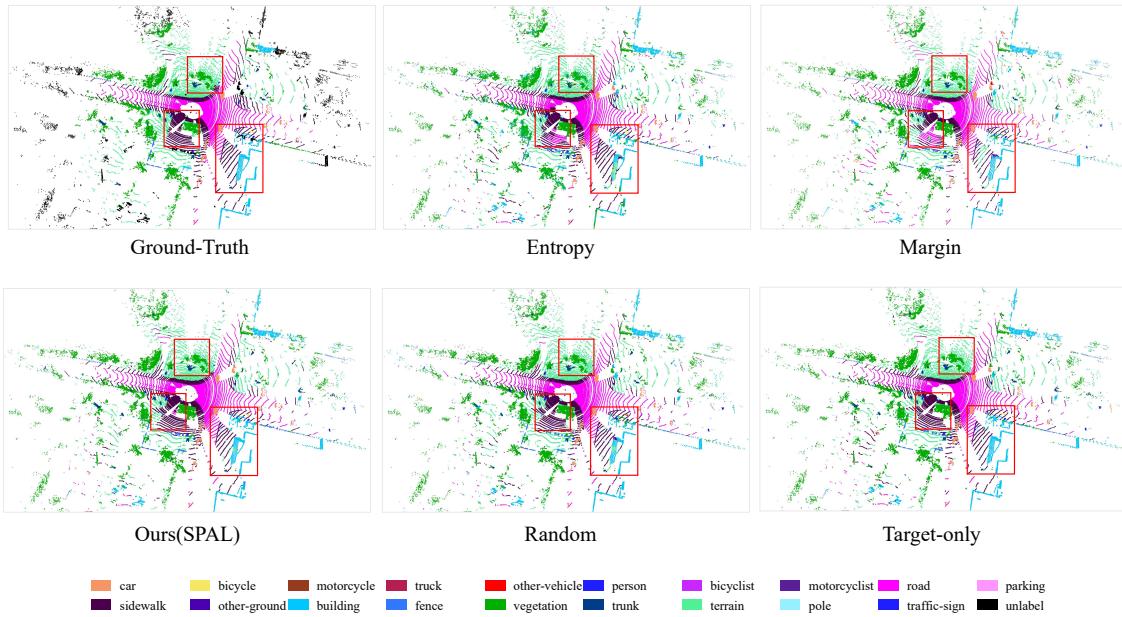


图 3-5 本章 SynLiDAR→SemanticKITTI 分割可视化图

Fig. 3-5 Visualization of the segmentation results in the SynLiDAR→SemanticKITTI

表 3-2 本方法与其他域适应方法在 SynLiDAR→SemanticPOSS 数据上的比较

Table 3-2 Comparison with other domain adaptation methods on SynLiDAR→SemanticPOSS

| 方法 | 域适应 | 标注 | 结果 |
|-------------|-----|------|-------------|
| Source-Only | - | - | 34.6 |
| Target-Only | - | 100% | 58.0 |
| CRST | UDA | - | 27.1 |
| ST-PCT | UDA | - | 29.6 |
| PolarMix | UDA | - | 30.4 |
| CoSMix | UDA | - | 40.4 |
| DGT-ST | UDA | - | 50.8 |
| Annotator | ADA | 0.1% | 52.0 |
| 本章方法 | ADA | 0.1% | 56.6 |

实验 SynLiDAR→SemanticKITTI 的结果如表3-2所示：在无监督域适应（UDA）方法中，DGT-ST 仍以 50.8% 的性能占据首位；而在主动域适应（ADA）方法中，

本章方法在 0.1% 的标注下取得 56.6% 的性能超过 Annotator4.6 个百分点，再一次证明了本章方法在“合成到真实”的跨域场景的有效性。可视化结果如图3-6所示：

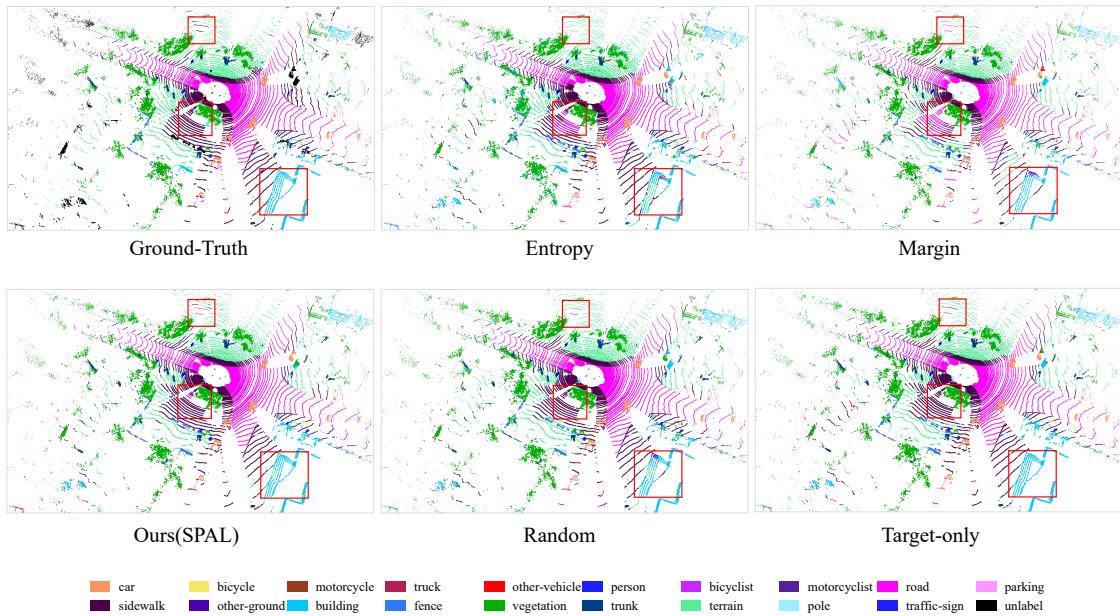


图 3-6 本章 SynLiDAR→SemanticPOSS 分割可视化图

Fig. 3-6 Visualization of the segmentation results in the SynLiDAR→SemanticPOSS

3.3.7.2 真实到真实场景

在“合成到真实”的跨域任务中，合成的数据集一般都有着标注精度更高，噪音度低的特性，其仍然与真实数据集有所差异。为进一步验证方法的泛化能力，本章将方法拓展至“真实到真实”跨域场景，并选择 nuScenes→SemanticKITTI 和 SemanticKITTI→nuScenes 跨域数据集进行了实验，在此场景下，本章方法展现出与合成到真实任务相似的性能优势，进一步证明了本章方法的有效性。

如表3-3所示，在 SemanticKITTI→nuScenes 实验中，目标域全监督基准结果与源域模型间存在 49.0 个百分点的性能差距，这说明在真实数据间，跨域任务要比合成更难。无监督域适应（UDA）方法中，LiDOG 以 34.9% 领先，但其性能仅为目标域基准的 42.2%，说明了无监督域适应在真实场景中的局限性，虽然不用进行标记但其性能仍然离全监督非常远。而本章 ADA 方法以 0.1% 标注量达到目标域全监督的 97.9%，较 Annotator 提升 5.1 个百分点，验证了本方法的有效性。同“合成到真实”数据集一样，本实验依然提供与 Ground-Truth、Target-Only 以及其他主动学习的可视化对比图，如图3-7所示：

表3-3 本方法与其他域适应方法在 SemanticKITTI→nuScenes 数据上的比较

Table 3-3 Comparison with other domain adaptation methods on SemanticKITTI→nuScenes

| 方法 | 域适应 | 标注 | 结果 |
|-------------|-----|------|----------------|
| Source-Only | - | - | 33.7 |
| Target-Only | - | 100% | 82.7 |
| Mix3D | UDA | - | 31.5 |
| CoSMix | UDA | - | 29.8 |
| SN | UDA | - | 25.8 |
| RayCast | UDA | - | 30.9 |
| LiDOG | UDA | - | 34.9 |
| Annotator | ADA | 0.1% | 75.9 |
| 本章方法 | ADA | 0.1% | 81.0(改) |

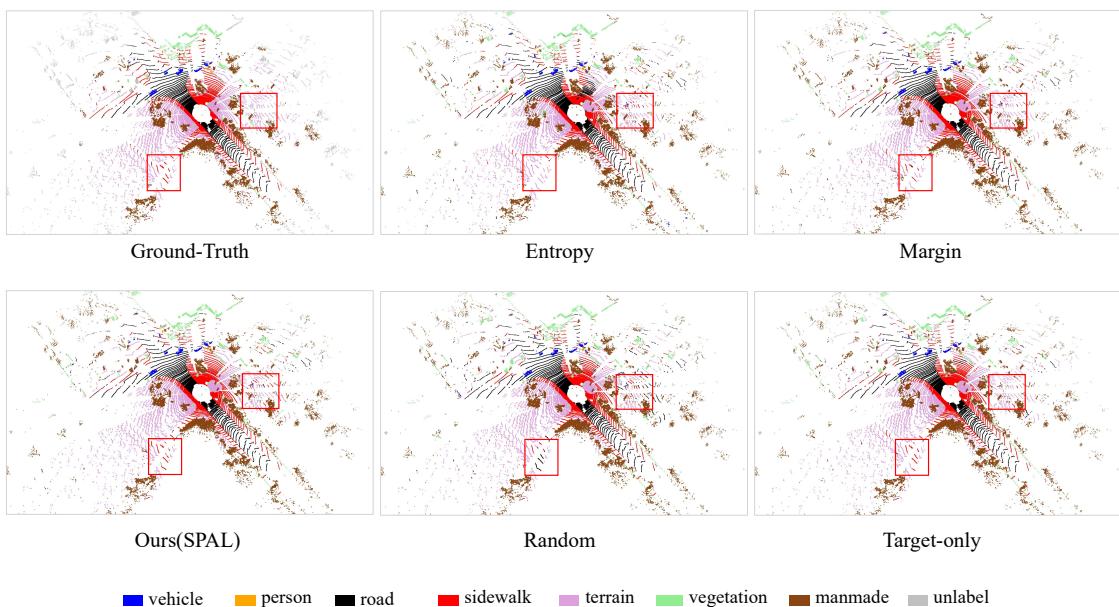


图3-7 本章 nuScenes→SemanticKITTI 分割可视化图

Fig. 3-7 Visualization of the segmentation results in the SemanticKITTI→nuScenes

在 nuScenes→SemanticKITTI 中如表3-4所示，目标域全监督性能与源域差距进一步扩大至 52.9 个百分点，任务难度进一步增加。值得注意的是，本章 ADA 方法在 0.1% 标注下完全复现目标域全监督性能，较 Annotator 提升 3.6 个百分点，首次实现极低标注量下的无损迁移。对比两类场景，UDA 方法 LiDOG 在 nuScenes→SemanticKITTI 任务中的性能依然强势，但是仍距离全监督有接近 44.2

个百分点的性能差距。综合而言，本章方法在真实到真实场景中均以 0.1% 标注量实现超 95% 全监督性能。可视化结果如图3-8所示：

表 3-4 本方法与其他域适应方法在 nuScenes→SemanticKITTI 数据上的比较

Table 3-4 Comparison with other domain adaptation methods on nuScenes→SemanticKITTI

| 方法 | 域适应 | 标注结果 |
|-------------|----------|----------------|
| Target-Only | - 100% | 85.4 |
| Source-Only | -- | 32.5 |
| Mix3D | UDA - | 32.4 |
| CoSMix | UDA - | 36.8 |
| SN | UDA - | 23.6 |
| RayCast | UDA - | 31.5 |
| LiDOG | UDA - | 41.2 |
| Annotator | ADA 0.1% | 81.8 |
| 本章方法 | ADA 0.1% | 85.4(改) |

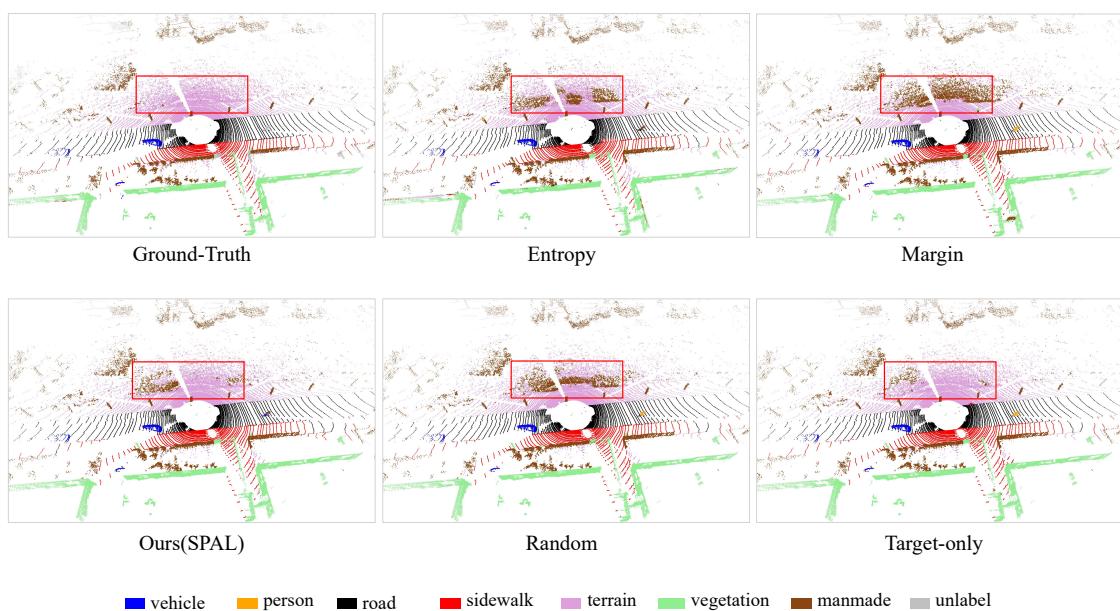


图 3-8 本章 nuScenes→SemanticKITTI 分割可视化图

Fig. 3-8 Visualization of the segmentation results in the nuScenes→SemanticKITTI

3.3.8 与其他主动学习方法对比

在未引入 Mixing 策略时，传统主动学习方法在两类合成到真实跨域任务中表现出了一定的优势如表3-5所示。本章的方法是通过源域而构建起的原型，因此比较依赖源域信息，在没有 mixing 的情况下无法发挥最大的效果，在 SynLiDAR→SemanticKITTI 任务中，Margin 方法以 54.4% 的性能取得最优结果，而本章方法（SPAL）以 53.5% 略低 1.1 个百分点，但仍然超过了 Random0.2 个百分点 Entropy2.8 个百分点。而在 SynLiDAR→SemanticPOSS 任务中，本章的主动学习策略（SPAL）低于熵采样（Entropy）0.7 个百分点，却也超过了 Random3.2 个百分点，Margin0.2 个百分点，这一结果表明，传统主动学习策略在特定场景下仍具竞争力，但单一采样准则难以适应跨域任务的复杂性。

表 3-5 本章的主动学习方法与其他传统主动学习方法对比

Table 3-5 Comparison with other active learning methods

| 数据集 | 方法 | 标注 | 结果 |
|------------------------|------------|------|-------------|
| SynLiDAR→SemanticKITTI | Random | 0.1% | 53.2 |
| | Entropy | 0.1% | 51.7 |
| | Margin | 0.1% | 54.4 |
| SynLiDAR→SemanticPOSS | Ours(SPAL) | 0.1% | 53.5 |
| | Random | 0.1% | 47.5 |
| | Entropy | 0.1% | 51.4 |
| | Margin | 0.1% | 50.5 |
| | Ours(SPAL) | 0.1% | 50.7 |

而在结合 Mixing 策略后如表3-6所示，除 Random 以外各主动学习方法在 SynLiDAR→SemanticKITTI 任务中的性能均显著提升，其中本章方法（SPAL）以 58.7% 的性能达到最优，较次优的 Margin 方法提升 1.1 个百分点，较未结合 Mixing 时的自身结果提升 5.2 个百分点且显著超越传统方法 Margin 的 57.6%，Margin、Entropy 则分别较自身提升 3.2 和 3.4 个百分点。这表明结合主动学习和 Mixing 策略的有效性，通过主动筛选出的标注目标点和源点动态构建中间域数据可以有效提升模型的性能，进一步缓解域间分布差异。但 Random 方法下降了 2.1 个百分点，这说明性能增益也与主动学习方法有关系，混合中间域信息包含域差异信息越丰富，其提升程度越高，而本文方法差异选择的是差异性和不确定性最高的点，因此提升幅度最大。

表 3-6 本章的主动学习方法与其他传统主动学习方法在结合 Mixing 后的对比

Table 3-6 Comparison with other active learning methods after integrating Mixing

| 数据集 | 方法 | 标注 | 结果 |
|------------------------|------------|------|-------------|
| SynLiDAR→SemanticKITTI | Random | 0.1% | 51.1 |
| | Entropy | 0.1% | 55.1 |
| | Margin | 0.1% | 57.6 |
| | Ours(SPAL) | 0.1% | 58.7 |

3.3.9 消融实验

为证明方法的有效性，本章消融实验结果如表3-7所示，本章提出的源域指导的主动学习方法（SPAL）与混合增强（Mixing）策略构建的中间域数据对模型性能具有显著协同优化作用。仅使用主动学习（SPAL）时，模型性能提升至 54.3%，较基线增加 31.5 个百分点，证明了语义感知采样机制对目标域关键样本筛选的优势。而当仅使用混合增强时（Mixing），模型性能为（running），较无任何模块的基线提升 (...) 个百分点，验证了跨域数据融合对缓解域间分布差异的有效性。而当二者联合使用时，模型以 58.7% 的达到最优，较单一模块性能分别提升 8.0 和 4.4 个百分点，表明本章方法的真实有效性。

表 3-7 本章方法模块消融实验

Table 3-7 Ablation experiments on modules

| 主动学习 (SPAL) | Mixing | 结果 |
|-------------|--------|-------------|
| | | 22.8 |
| ✓ | | 50.7 |
| | ✓ | running |
| ✓ | ✓ | 58.7 |

3.3.10 本章小结

本章主要研究适用于点云语义分割域适应任务的主动学习方法。为了解决传统主动学习方法中的不足，提出了一种原型指导的主动学习策略，该策略通过动态构建源域原型来代表源域类别质心，并在每一轮主动学习阶段实时更新原型，在进行目标域候选点筛选时，计算每个目标域中未标注的点与每个源域类别原型的相似度，通过最优-次优差异算法获取归一化后的类别概率的差值得到域差异性评

分，值越小说明该点域差异性越高，同时结合不确定性评分得到最终候选评分，升序排列后选取前 k 个同时兼备高不确定性和高域差异性的目标点。此外，本章首次将主动学习方法与 Mixing 策略结合，构建包含目标域信息和源域信息的中间域数据，帮助模型学习到更稳定的域不变特征，进一步缩小域间隙。在本章中，首先介绍了方法的主要框架和流程，并分别对方法中的三个模块做了详细介绍，这三个模块共同组成了本章的方法，大幅度提升了模型的跨域性能。同时，为了验证方法的有效性，在两个跨域场景四个数据集上进行了大量实验，并与此前最有效的点云语义分割无监督域适应和主动域适应进行了对比，通过实验分析验证了本章方法的有效性，最后进一步对 Mixing 和 SPAL 模块进行消融实验，充分验证了两个模块的有效性。

第 4 章 基于多模态特征融合的锚点定位点云配准

4.1 本章引言

点云中广泛存在可重复且模糊的结构，例如地板、墙壁和天花板都是平面。这些可重复且不明确的结构信息将在很大程度上影响特征的独特性、差异性。因此，只考虑点云结构信息的神经网络预测出来的对应关系中包含大量的离群值。如何提高点云特征的区分度是当前点云配准方法的主要研究方向。

在早期研究阶段，大多数方法要么对几何结构进行充分的发掘和利用，要么参考图像配准方法开发新的配准框架，往往忽略了来自图片的纹理信息。二维图像虽然缺乏距离和角度等立体空间信息，但是其提供的颜色纹理等内容是人类理解世界的重要组成部分。近年来在 3D 目标检测和位置识别等领域越来越多的研究人员考虑将图像信息与点云信息融合起来。多模态信息融合的基本理论是利用模态信息间的互补性实现信息的相互补充提高特征的鲁棒性。就点云和图像而言，点云数据缺乏颜色信息和纹理信息但是对物体与环境的拓扑结构有着良好的表示，图像数据便能够作为有效的补充数据对点云数据进行补充。

文献 [58, 59] 通过用逐点的 2D 分割特征增强 3D 坐标来进行数据级融合；文献 [60, 61] 通过简单的拼接或特定模块实现来自单个网络的 2D 和 3D 表示的特征级融合。与仅使用激光雷达的方法不同，这种方法在单点云模式下不断更新更复杂的设计模型和更合适的训练方案，多模态替代方案努力利用更多样化的信息，并显示出巨大的潜力。

尽管使用多模态的方法来进行特征提取受到越来越多的研究者的青睐，但是随之带来的挑战是不同模态之间存在着模态差异，如何更好的融合来自多种模态的信息是重点的研究方向。据调查发现，在 3D 目标检测的同行研究中，虽然研究人员期望这两种传感器的组合能够提供更好的性能，但事实证明，大多数最先进的 3D 物体探测器仅使用激光雷达作为输入。这表明如何有效地融合来自这两个传感器的数据仍然具有挑战性。虽然激光雷达点云和 RGB 图像具有互补的信息，但是由于两种模态的数据存在较大的域间隙，实现信息的互补并不容易。

这种差距的出现主要由三方面造成：（1）点云和图像特征提取的网络之间存在较大差异，用于提取点云特征的网络往往针对点云数据的无序性，不规则性和稀疏性设计，而图像特征提取网络主要利用卷积对图像的结构和纹理信息进行提取，因此两种模态数据的特征之间存在着较大鸿沟，导致了融合过程中信息的丢失。（2）当前算法往往使用卷积神经网络提取图像特征之后，将像素特征与原始点云进行简单拼接并输入点云骨干网络以完成特征融合，这进一步限制了融合效果。

这使得不同模态特征的相关性被忽略了，关键信息没有有效的突出。(3) 数据增强技术被广泛应用于各种任务当中，但是对于多模态融合来说这种简单的机制可能不会有效的提高算法性能，这主要是由于对多模态而言对齐两种模态的数据是非常重要的，通过旋转平移等数据增强操作往往会造成模态间数据的错位。

综上所述，本章为了提高点特征的区分度设计了一个点云图像融合的点云配准框架。提出了一个对齐模块将数据增强后的两种模态数据进行像素与超点间的对齐。提出一种新的多模态融合方法，先后在模态无关和模态相关两个子空间对点云特征和图像特征进行融合，减少模态间的域间隙和信息丢失。

4.2 基于多模态特征融合的锚点定位点云配准方法

为了更有效的提高源点云和目标点云相似非重叠区域特征间的特征差异性，本章提出了一个基于多模态特征融合点云配准新框架，整个方法遵循图4-1所示的基于多模态融合的点云配准新框架结构。

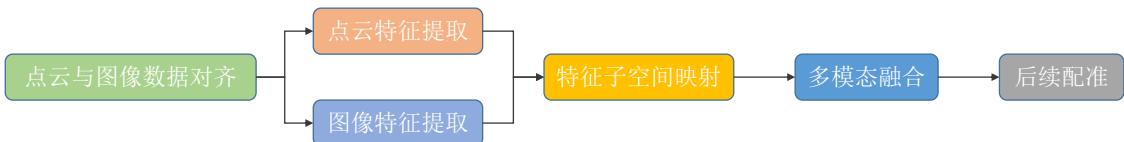


图 4-1 本方法示意图

Fig. 4-1 Schematic diagram of the method

图4-2展示了整个方法的示意图。该方法主要包括三个步骤：(1) 在特征提取之前进行数据对齐，将点云和对应图像间的空间位置对齐即将点云中点的坐标投影至图像中，寻找点与像素之间的对应关系；(2) 从点云中提取出超点的局部几何特征，并从各自对应的图像中提取二维纹理特征；(3) 最后使用基于注意力机制的融合模块，将点云中的超点的特征与对应像素的纹理特征相融合，得到融合特征。随后操作与第三章中的方法一致，通过融合后的特征寻找具有显著特征且位于重叠区域的锚点，然后利用自注意力和交叉注意力机制对点云的几何结构特征进行编码，接着通过迭代优化更新显著超点与超点特征，最终将超点匹配扩充为点匹配并估计变换矩阵。

基于多模态特征融合的锚点定位点云配准方法的重点在强调点云和图像间数据的对齐作为多模态融合的前置操作，经过对齐模块能够寻找到点云中的点与对应图像中像素的对应关系，相比于使用全局融合，这种融合方法是能够更加准确的有效的融合点的颜色纹理信息。许多以前的方法都没有在多模态融合之前进行额外的操作，这使得他们的融合效果受到限制；同时该方法在融合之前将模态数据投影到两个子空间中，通过减少域间隙融合两种不同模态的数据能够减少冗余信息

对特征的干扰，使得多模态融合的效果更好。

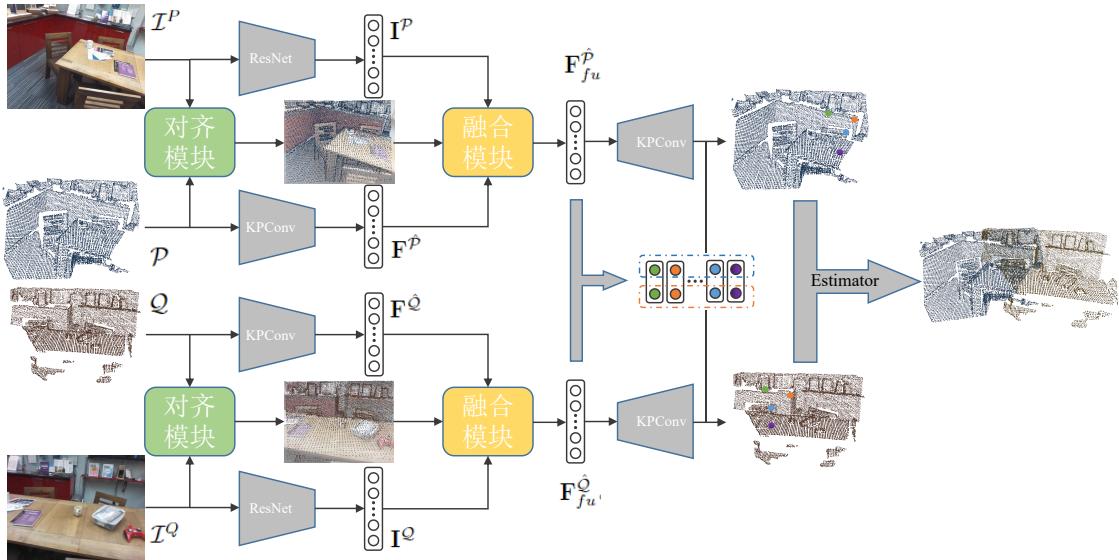


图 4-2 本方法框架图

Fig. 4-2 Overview of the proposed registration method

该方法的贡献如下：(1) 对于点云和图像模态，使用对齐模块将点云和图像之间的数据对齐，具体而言将点云投影至二维平面并与图像对齐，寻找到超点与像素之间的一种一对多的对应关系；(2) 在多模态融合之前，将投影两种模态的数据到正交的子空间，并先后融合两种模态数据学习跨模态的信息。

4.2.1 对齐模块

为了训练图像和点云数据，本文提出了一个融合模块，用于将图像的纹理颜色和点云的几何结构进行融合。因此，需要找到点云中的点和图像中的像素之间的对应关系。从三维空间到二维图像有一个成像过程，本模块通过将三维点云投影到二维图像，以便精确的寻找到点和像素之间的映射关系，促进后续两种模态特征的融合。对于源点云 \mathcal{P} 和源点云对应的图像 I^P 而言，它们之间的成像过程可以由公式 (4-1) 表示：

$$I^P = \mathbf{K}_{in} \mathbf{T}_{ex} \mathcal{P} \quad (4-1)$$

式中， \mathbf{K}_{in} 表示相机的内参矩阵； \mathbf{T}_{ex} 表示相机的外参矩阵。对于未经过数据增强的点云和图像而言外参矩阵 \mathbf{T}_{ex} 为 4 阶单位矩阵， \mathbf{K}_{in} 则与相机本身相关且不受到数据增强的影响。

一种处理方式是不对点云数据进行数据增强处理，然而这种简单的处理方法会使模型在训练过程中陷入过拟合。本方法为了避免模型过拟合依旧采用随机旋转和增强噪音等策略。然而这时如果不采用额外的处理，依旧使用 4 阶单位矩阵作

为外参矩阵 \mathbf{T}_{ex} , 那么点云的投影和图像之间将会存在偏差, 这将导致点的颜色信息被错误的融合了, 进一步到整体模型性能的下降。文献 [62] 指出数据增强对模型的促进作用将会随着随机旋转角度的增大而下降。这也进一步表明在特征融合之前对齐点云和图像两种模态的数据是至关重要的。

为了解决数据增强带来的对齐偏差问题, 本文提出了一个模态对齐模块。为了使重定位可行, 对点云进行数据增强后, 首先保存数据增强相关参数比如旋转角度。在对齐阶段, 它将这些数据增强进行反转, 得到输入点云中的三维空间点的原始坐标, 然后在相机空间中找到其对应的二维坐标。由于本方法采用的是又粗到细的配准框架, 即在提取点云特征过程中将点云下采样为超点。超点与原始点之间存在一对多的关系, 而原始点与像素存在一对一的关系, 因此最终本模块将得到超点与像素之间的一对多的对应关系。这种对应关系表示一个超点是图像中某块区域。最终, 本节将得到超点与像素之间的对应关系矩阵 $\mathbf{C}^{\hat{P}} \in \mathbb{R}^{N' \times L}$ 。同理本节会得到目标点云与其图像间的对应关系 $\mathbf{C}^{\hat{Q}} \in \mathbb{R}^{M' \times L}$ 。

4.2.2 多模态特征提取

该方法的另一模块是多模态特征提取, 而多模态特征的提取的核心是从点云和是图像当中提取出能够很好地代表点云局部结构和图像纹理信息的特征。本研究将源点云和目标点云输入到 KPConv 网络中, 该网络能过够在提取特征的同时将点云下采样为超点。另一方面, 本方法将对应的图像输入到用于提取像素特征的标准的 ResUNet 骨干网络中。选择 ResUNet 主要是因为可以加载该网络的最流行的预训练模型。这种由大数据集训练而来预训练模型, 能够使网络起初就拥有良好的图像特征, 同时使网络在训练过程中更加稳定。

具体的网络结构遵循 CofiNet 使用 KPConv 稀疏卷积网络。输入源点云 $\mathcal{P} \in \mathbb{R}^{N \times 3}$ 和目标点云 $\mathcal{Q} \in \mathbb{R}^{M \times 3}$, 输出 $\mathbf{F}^{\hat{P}} \in \mathbb{R}^{N' \times d}$ 和 $\mathbf{F}^{\hat{Q}} \in \mathbb{R}^{M' \times d}$ 。输入是源点云和目标点云对应的图像 $\mathcal{I}^P \in \mathbb{R}^{W \times H \times 3}$ 和 $\mathcal{I}^Q \in \mathbb{R}^{W \times H \times 3}$, 输出是其特征 $\mathbf{I}^P \in \mathbb{R}^{L \times d}$ 和 $\mathbf{I}^Q \in \mathbb{R}^{L \times d}$ 。其中图像特征的维度 $L = H \times W$ 。

4.2.3 融合模块

在经过模态数据的对齐和多模态特征提取之后, 本节将对不同模态的特征进行融合。首先将不同模态特征提取到的点云和图像特征投影到不同的子空间中捕获模态相关和模态无关的信息, 以获得更全面的特征表示。不失一般性, 以源点云为例。在得到点云特征 $\mathbf{F}^{\hat{P}}$ 和图像特征 \mathbf{I}^P 之后, 点云中第 k 个超点的特征以符号 $\mathbf{F}^{\hat{P}}(k)$ 表示, 第 k 个超点在图像中的对应区域内所有像素的特征构成一个集合以符号 $\{\mathbf{I}^P(k)\}$ 表示。具体如图4-3所示, 图中紫色点代表超点, 而图像中的紫色区

域代表该超点的对应区域。利用解耦器 \mathbf{E}_{ir}^P 和解耦器 \mathbf{E}_{co}^P 对 $\mathbf{F}^{\hat{P}}(k)$ 进行解码，将其映射到模态无关和模态相关两个子空间中。用 $\mathbf{F}_{ir}^{\hat{P}}(k)$ 和 $\mathbf{F}_{co}^{\hat{P}}(k)$ 分别表示点云的模态无关特征和模态相关特征。

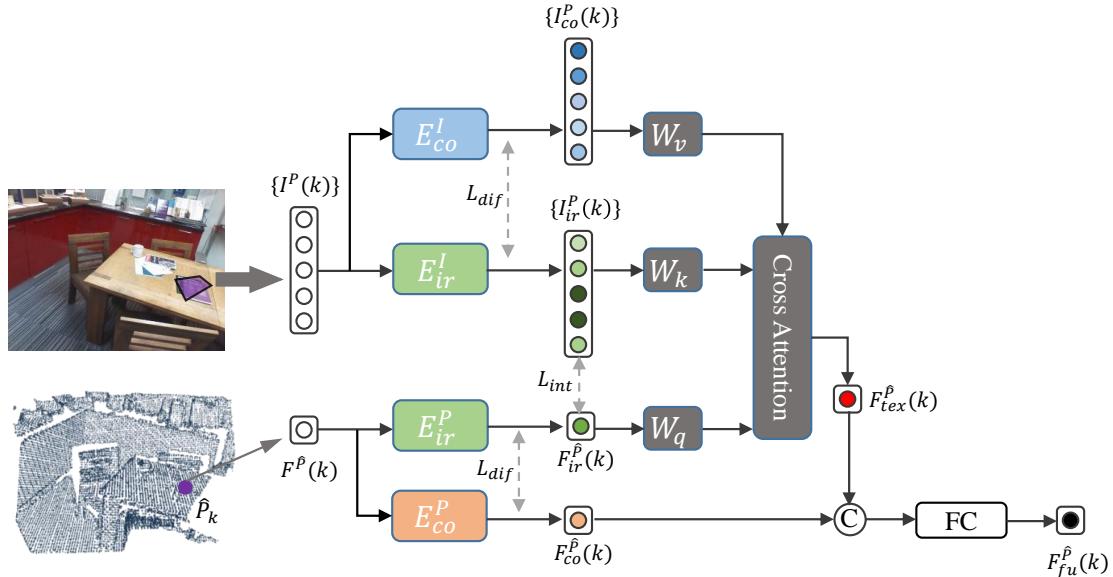


图 4-3 特征融合模块流程图

Fig. 4-3 Flowchart of feature fusion module

类似的，利用解耦器 \mathbf{E}_{ir}^I 和解耦器 \mathbf{E}_{co}^I 将 $\{\mathbf{I}^P(k)\}$ 解码为图像模态无关特征 $\{\mathbf{I}_{ir}^P(k)\}$ 和图像模态相关特征 $\{\mathbf{I}_{co}^P(k)\}$ 。可用公式 (4-2) 表示上述解码过程：

$$\begin{cases} \{\mathbf{I}_{ir}^P(k)\} = \mathbf{E}_{ir}^I(\{\mathbf{I}^P(k)\}) \\ \{\mathbf{I}_{co}^P(k)\} = \mathbf{E}_{co}^I(\{\mathbf{I}^P(k)\}) \\ \mathbf{F}_{ir}^{\hat{P}}(k) = \mathbf{E}_{ir}^P(\mathbf{F}^{\hat{P}}(k)) \\ \mathbf{F}_{co}^{\hat{P}}(k) = \mathbf{E}_{co}^P(\mathbf{F}^{\hat{P}}(k)) \end{cases} \quad (4-2)$$

提取模态无关特征的目的在于提取潜在的公共特征，从而缩小不同模态之间的域间隙。在得到 $\mathbf{F}_{ir}^{\hat{P}}(k)$ 和 $\{\mathbf{I}_{ir}^P(k)\}$ 之后，本模块利用可学习矩阵 \mathbf{W}_q 和 \mathbf{W}_k 分别对两种模态无关特征进行映射后计算得到第 k 超点与它所对应图像区域像素的相似度矩阵 $\mathbf{S}^P(k)$ 。具体来说，第 k 个超点与它的第 i 个像素的相似度分数 $s^P(k, i)$ 可由公式 (4-3) 计算：

$$\mathbf{s}^P(k, i) = (\mathbf{W}_q \mathbf{F}_{ir}^{\hat{P}}(k)) (\mathbf{W}_k \mathbf{I}_{ir}^P(k, i))^T \quad (4-3)$$

式中， $\mathbf{I}_{ir}^P(k, i)$ 是第 k 个超点的第 i 个对应像素特征 $\mathbf{I}^P(k, i)$ 在模态无关子空间中的投影。随后 $\{\mathbf{I}_{co}^P(k)\}$ 经过 \mathbf{W}_v 的映射并结合相似度矩阵进行加权求和可以得

到第 k 超点的纹理特征 $\mathbf{F}_{tex}^P(k)$:

$$\mathbf{F}_{tex}^P(k) = \sum_i \mathbf{s}^P(k, i) \mathbf{I}_{co}^P(k, i) \quad (4-4)$$

式中, $\mathbf{I}_{co}^P(k, i)$ 第 k 个超点的第 i 个对应像素特征 $\mathbf{I}^P(k, i)$ 在模态相关子空间中的投影。相比于直接利用 $\mathbf{F}^{\hat{P}}(k)$ 与 $\{\mathbf{I}^P(k)\}$ 进行特征融合, 上述方法由于在特征无关子空间中进行不同模态特征的查询, 缩小了两种模态的域间隙减少了独有信息的影响。同理可以得到源点云中所有超点的纹理特征 $\mathbf{F}_{tex}^{\hat{P}}$ 。

在得到超点的纹理特征 \mathbf{F}_{tex}^P 之后, 本文再将超点的模态相关特征 $\mathbf{F}_{co}^{\hat{P}}$ 与之在通道维度上拼接, 并将其送入最后的映射层完成特征融合。可由公式 (4-5) 表示:

$$\mathbf{F}_{fu}^{\hat{P}} = f(\mathbf{F}_{co}^{\hat{P}} \oplus \mathbf{F}_{tex}^{\hat{P}}) \quad (4-5)$$

式中, $f(\cdot)$ 表示映射层, 由全连接层和激活函数构成。将 \mathbf{F}_{tex}^P 和 $\mathbf{F}_{co}^{\hat{P}}$ 融合主要是为了避免两种模态独有信息的丢失。同理可以得到目标点云融合后的超点特征 \mathbf{F}_{fu}^Q 。

4.2.4 损失函数

为了使融合模块中的四个解耦器 \mathbf{E}_{ir}^I 、 \mathbf{E}_{co}^I 、 \mathbf{E}_{ir}^P 和 \mathbf{E}_{co}^P 达到设计要求。本文利用差异损失 \mathcal{L}_{dif} 和三元组损失 \mathcal{L}_{int} 分别监督同一模态特征和不同模态特征。

4.2.4.1 差异损失

差异损失主要是为了使同一模态的两种解码器能够捕获输入特征的不同方面的信息。通过解耦模态相关特征和模态无关特征之间存在一个正交约束。具体的, 对于第 k 个超点的第 i 个像素特征在两个子空间的投影 $\mathbf{I}_{ir}^P(k, i)$ 和 $\mathbf{I}_{co}^P(k, i)$ 来说, 理论上它们存在公式 (4-6) 的正交约束:

$$\|\mathbf{I}_{ir}^P(k, i)(\mathbf{I}_{co}^P(k, i))^T\|_F^2 = 0 \quad (4-6)$$

式中, $\|\cdot\|_F^2$ 表示 Feobenius 范数的平方。同时, 这种约束对于第 k 个超点在两个子空间中的投影依旧成立。为此, 本文使用的差异损失 \mathcal{L}_{dif} 如公式 (4-7) 所示:

$$\mathcal{L}_{dif} = \sum_k (\|\mathbf{F}_{ir}^{\hat{P}}(k)(\mathbf{F}_{co}^{\hat{P}}(k))^T\|_F^2) + \sum_{k,i} (\|\mathbf{I}_{ir}^P(k, i)(\mathbf{I}_{co}^P(k, i))^T\|_F^2) \quad (4-7)$$

4.2.4.2 三元组损失

模态无关特征指的是同一事物的结构纹理等信息不随表示形式的变化而变化的那部分。将点云和图像两种模态的特征同时投影至模态无关子空间中, 得到每个

模态自身的模态无关向量。模态间的损失函数的作用就是让不同模态的模态无关向量更加相似，这有助于通过一种模态特征查询另一种模态特征。本方法使用三元组损失函数监督两种模态在模态无关子空间中的投影。三元组损失在拉近锚点和正样本之间距离的同时，增加锚点与负样本之间的距离。具体而言，任一超点的模态无关特征和它对应像素的模态无关特征之间的距离要小于它与其他像素之间的特征距离；同时，对应像素集合中的特征距离要小于与其他像素之间的特征距离。因此，本文使用的三元损失 \mathcal{L}_{int} 如公式 (4-8) 所示：

$$\begin{aligned} \mathcal{L}_{int} = & \sum_{k,x} \max(\lambda - d(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(k), \{\mathbf{I}_{ir}^{\mathcal{P}}(k)\}) + d(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(k), \{\mathbf{I}_{ir}^{\mathcal{P}}(x)\})) \\ & + \sum_{k,x} \max(\lambda - d(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(k), \{\mathbf{I}_{ir}^{\mathcal{P}}(k)\}) + d(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(x), \{\mathbf{I}_{ir}^{\mathcal{P}}(k)\})) \end{aligned} \quad (4-8)$$

式中， λ 用于控制正负样本之间的距离； $(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(k), \mathbf{I}_{ir}^{\mathcal{P}}(k))$ 对应三元损失中的正样本对； $(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(k), \mathbf{I}_{ir}^{\mathcal{P}}(x))$ 和 $(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(x), \mathbf{I}_{ir}^{\mathcal{P}}(k))$ 对应三元函数的负样本对。 $d(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(x), \mathbf{I}_{ir}^{\mathcal{P}}(y))$ 表示超点 \mathcal{P}_x 与超点 \mathcal{P}_y 所对应像素集之间的距离和，由公式 (4-9) 定义：

$$d(\mathbf{F}_{ir}^{\hat{\mathcal{P}}}(x), \{\mathbf{I}_{ir}^{\mathcal{P}}(y)\}) = \sum_i -\| \max(0, \mathbf{F}_{ir}^{\hat{\mathcal{P}}}(x) - \mathbf{I}_{ir}^{\mathcal{P}}(y, i)) \|^2 \quad (4-9)$$

4.3 实验结果与分析

4.3.1 数据集预处理

本章采用的数据集 3DMatch 和 3DLoMatch 与第三章略有差别。在上一章所用数据集的基础上本章为每个点云配对了 RGB 图像数据。RGB 图像数据来自于最原始的 3DMatch 数据集，每 50 张图像合成一个点云数据。为了减少时间花销，本方法没有使用全部的 50 张图片而是选择第一张图片用于多模态融合。同时，本方法记录了第一帧图片的相机位姿和相机内参并加入了数据集，以保证本章中的对其模块能够正常使用。至此，整个数据集的预处理完成，整个数据集由点云、图像和相机的内参外参矩阵组成。

4.3.2 3DMatch 和 3DLoMatch 实验

在本节中，本文对多模态融合的点云配准任务在 3DMatch 和 3DLoMatch 上进行了实验分析。实验表明，通过融合图像特征本方法的点云配准方法能够达到较好水平。如表4-1所示，本方法在内点率上远高于其他方法。尤其当采样数为 250 时，在 3DMatch 和 3DLoMatch 上比 CoFiNet 分别高了 32.4% 和 30.2%。这表明本方法能够在相同情况下比其他方法能够更加有效的找到正确的点对应。

表4-1 本方法与先进方法的内点率和匹配召回率的比较

Table 4-1 Comparison of Inlier Ratio with advanced methods

| Sample | 3DMatch | | | | | | 3DLoMatch | | | |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 5000 | 2500 | 1000 | 500 | 250 | 5000 | 2500 | 1000 | 500 | 250 |
| PerfectMatch | 36.0 | 32.5 | 26.4 | 21.5 | 16.4 | 11.4 | 10.1 | 8.1 | 6.4 | 4.8 |
| FCGF | 56.8 | 54.1 | 48.7 | 42.5 | 34.1 | 21.4 | 20.0 | 17.2 | 14.8 | 11.6 |
| D3Feat | 39.0 | 38.8 | 40.4 | 41.5 | 41.8 | 13.2 | 13.1 | 14.0 | 14.6 | 15.0 |
| SpinNet | 47.5 | 44.7 | 39.4 | 33.9 | 27.6 | 20.5 | 19.0 | 16.3 | 13.8 | 11.1 |
| Predator | 58.0 | 58.4 | 57.1 | 54.1 | 49.3 | 26.7 | 28.1 | 28.3 | 27.5 | 25.8 |
| YOHO | 64.4 | 60.7 | 55.7 | 46.4 | 41.2 | 25.9 | 23.3 | 22.6 | 18.2 | 15.0 |
| CoFiNet | 49.8 | 51.2 | 51.9 | 52.2 | 52.2 | 24.4 | 25.9 | 26.7 | 26.8 | 26.9 |
| Ours | 71.9 | 78.5 | 83.0 | 84.8 | 85.8 | 41.4 | 47.0 | 52.9 | 55.5 | 57.1 |

对于特征召回率而言,本方法虽然在3DMatch上的表现一般,在采样数为5000、2500和250时分别比最优模型低了0.2%、0.3%和0.4%。但是如表4-2所示,本方法在3DLoMatch的表现都达到了最优性能,分别比次优模型高了4.5%、4.4%、4.1%、4.3%和4.2%。这表明本方法相比于其他方法能够更好的在低重叠率情况下工作,能够使源点云和目标点云之间的点对应更多,从而达到RANSAC算法发挥最用的最低要求。

表4-2 本方法与先进方法的匹配召回率的比较

Table 4-2 Comparison of Feature Matching Recall with advanced methods

| Sample | 3DMatch | | | | | | 3DLoMatch | | | |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 5000 | 2500 | 1000 | 500 | 250 | 5000 | 2500 | 1000 | 500 | 250 |
| PerfectMatch | 95.0 | 94.3 | 92.9 | 90.1 | 82.9 | 63.6 | 61.7 | 53.6 | 45.2 | 34.2 |
| FCGF | 97.4 | 97.3 | 97.0 | 96.7 | 96.6 | 76.6 | 75.4 | 74.2 | 71.7 | 67.3 |
| D3Feat | 95.6 | 95.4 | 94.5 | 94.1 | 93.1 | 67.3 | 66.7 | 67.0 | 66.7 | 66.5 |
| SpinNet | 97.6 | 97.2 | 96.8 | 95.5 | 94.3 | 75.3 | 74.9 | 72.5 | 70.0 | 63.6 |
| Predator | 96.6 | 96.6 | 96.5 | 96.3 | 96.5 | 78.6 | 77.4 | 76.3 | 75.7 | 75.3 |
| YOHO | 98.2 | 97.6 | 97.5 | 97.7 | 96.0 | 79.4 | 78.1 | 76.3 | 73.8 | 69.1 |
| CoFiNet | <u>98.1</u> | 98.3 | <u>98.1</u> | 98.2 | 98.3 | <u>83.1</u> | <u>83.5</u> | <u>83.3</u> | <u>83.1</u> | <u>82.6</u> |
| Ours | 98.0 | <u>98.0</u> | 98.4 | 98.2 | 98.3 | 87.6 | 87.9 | 87.4 | 87.4 | 86.8 |

本方法进一步与各个先进方法在匹配召回率上做了对比。如表4-3所示，本方法在 3DLoMatch 和 3DMatch 数据集上的性能均达到了当前最优的性能。具体来说，在 3DMatch 上当采样数为 5000、2500、1000、500 和 250 时，本模型的新能分别比次优模型高了 0.2%、1%、0.1%、1.4% 和 2.3%；在 3DLoMatch 上则分别高出了 2.5%、3.3%、6%、6.4% 和 7.5%。这表明本方法生成的点对应关系能够有效的实现点云间的配准。同时可以注意到，本方法的性能的变化幅度远远低于其他方法，其他模型的性能随着采样数的减少有明显下降。这表明本方法能够有效提取空间一致性的点对应，即使在低重叠率的情况下，也能够通过少数高置信度的点对应恢复点云间的变换矩阵。这也表明通过融合图像特征，本模型受到的噪声影响更小，模型更加稳定。

表 4-3 本方法与先进方法的配准召回率的比较

Table 4-3 Comparison of registration recall rates of this method with advanced methods

| Sample | 3DMatch | | | | | 3DLoMatch | | | | |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 5000 | 2500 | 1000 | 500 | 250 | 5000 | 2500 | 1000 | 500 | 250 |
| PerfectMatch | 78.4 | 76.2 | 71.4 | 67.6 | 50.8 | 33.0 | 29.0 | 23.3 | 17.0 | 11.0 |
| FCGF | 85.1 | 84.7 | 83.3 | 81.6 | 71.4 | 40.1 | 41.7 | 38.2 | 35.4 | 26.8 |
| D3Feat | 81.6 | 84.5 | 83.4 | 82.4 | 77.9 | 37.2 | 42.7 | 46.9 | 43.8 | 39.1 |
| SpinNet | 88.6 | 86.6 | 85.5 | 83.5 | 70.2 | 59.8 | 54.9 | 48.3 | 39.8 | 26.8 |
| Predator | 89.0 | 89.9 | <u>90.6</u> | 88.5 | 86.6 | 59.8 | 61.2 | 62.4 | 60.8 | 58.1 |
| YOHO | <u>90.8</u> | <u>90.3</u> | 89.1 | <u>88.6</u> | 84.5 | 65.2 | 65.5 | 63.2 | 56.5 | 48.0 |
| CoFiNet | 89.3 | 88.9 | 88.4 | 87.4 | <u>87.0</u> | <u>67.5</u> | <u>66.2</u> | <u>64.2</u> | <u>63.1</u> | <u>61.0</u> |
| Ours | 91.0 | 91.3 | 90.7 | 90.0 | 89.3 | 70.0 | 69.5 | 70.2 | 69.5 | 68.5 |

本章同样使用加权奇异值分解法和局部到全局变换估计两种算法估计变换矩阵，表4-4的实验结果进一步验证了上述表述。本实验首先使用加权奇异值分解算法求解变换矩阵，其他模型要么不能达到合理的结果，要么性能严重下降。比如 CoFiNet 模型使用加权奇异值分解进行变换估计时比基于 RANSAC 的变换估计性能在 3DMatch 和 3DLoMatch 上分别下降了 24.7% 和 49.5%。相比之下，本方法在 3DMatch 上的配准召回率达到了 84.7% 仅比利用 RANSAC 得到的结果下降了 5.3%。在 3DLoMatch 上虽然性能下降了 14.1%，但是也接近了 Predator 方法性能。这得益于本方法能够有效找到可靠和分布良好的超点对应。当使用局部到全局配准进行变换估计时，本方法在 3DMatch 上的配准召回率达到 90.6%，3DLoMatch 的配准召回率达到 73.1%，基本达到了其他模型使用 RANSAC 的结果。

表4-4 在3DMatch和3DLoMatch上使用不同姿态估计器的配准结果

Table 4-4 Registration results with different pose estimators on 3DMatch and 3DLoMatch

| Model | Estimator | Sample | RR(%) | | Times(s) | | |
|----------|--------------|--------|-------------|-------------|----------|--------|--------|
| | | | 3DMatch | 3DLoMatch | Model | Pose | Total |
| FCGF | RANSAC-50k | 5000 | 85.1 | 40.1 | 0.098 | 9.015 | 9.113 |
| Predator | RANSAC-50k | 5000 | 89.0 | 59.8 | 0.079 | 15.434 | 15.513 |
| CoFiNet | RANSAC-50k | 5000 | <u>89.3</u> | <u>67.5</u> | 0.259 | 5.321 | 5.580 |
| Ours | RANSAC-50k | 5000 | 91.0 | 70.0 | 0.141 | 4.215 | 4.356 |
| FCGF | weighted SVD | 250 | 42.1 | 3.9 | 0.098 | 0.008 | 0.106 |
| Predator | weighted SVD | 250 | 50.0 | 6.4 | 0.079 | 0.010 | 0.089 |
| CoFiNet | weighted SVD | 250 | <u>64.6</u> | <u>21.6</u> | 0.259 | 0.004 | 0.263 |
| Ours | weighted SVD | 250 | 84.7 | 55.9 | 0.141 | 0.004 | 0.145 |
| CoFiNet | LGR | all | <u>87.6</u> | <u>64.8</u> | 0.259 | 0.242 | 0.501 |
| Ours | LGR | all | 90.6 | 73.1 | 0.141 | 0.317 | 0.458 |

本文进一步在3DMatch和3DLoMatch数据集上测试了相对旋转误差和相对平移误差。如4-5所示，本方法相比于其他方法有明显优势，在3DMatch和3DLoMatch中都获得了最低的相对旋转误差和相对平移误差。在3DMatch上，本方法的相对旋转误差和相对平移误差分别比次优模型的性能低了 0.138° 和0.001cm。尤其在3DLoMatch上，本方法的相对旋转误差比次优的Predator小了 0.202° 。该实验结果表明通过将图像中的纹理信息和点云的结构信息相融合，能够提高点对应的数量和质量，能够使模型在配得“多”的同时配得“好”，这也从侧面验证了本方法融合方式的有效性。

表4-5 3DMatch和3DLoMatch的相对平移误差和相对旋转误差比较

Table 4-5 Comparison of RRE and RTE on 3DMatch and 3DLoMatch

| Model | 3DMatch | | 3DLoMatch | |
|----------|-----------------|--------------|-----------------|--------------|
| | RRE($^\circ$) | RTE(cm) | RRE($^\circ$) | RTE(cm) |
| FCGF | <u>1.949</u> | 0.066 | 3.146 | 0.100 |
| D3Feat | 2.161 | 0.067 | 3.361 | 0.103 |
| Predator | 2.029 | <u>0.064</u> | <u>3.048</u> | <u>0.093</u> |
| CoFiNet | 2.002 | <u>0.064</u> | 3.271 | 0.090 |
| Ours | 1.811 | 0.063 | 2.846 | 0.090 |

图4-4中展示了当前方法和CoFiNet方法两种方法配准结果的可视化。如图4-4所示，本方法不仅在重叠率高于30%时实现配准，而且在重叠率低至14%时也能准确配准点云。

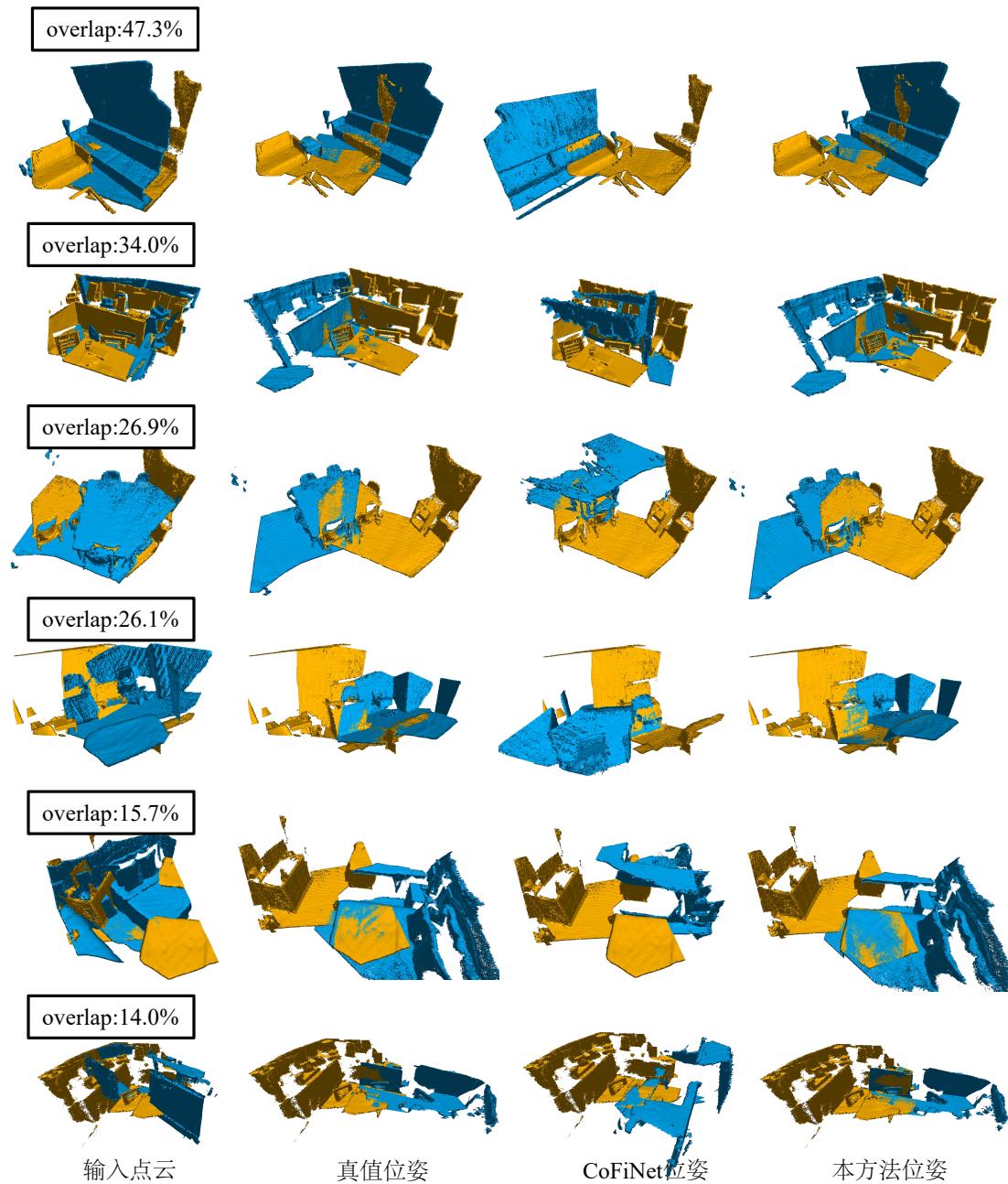


图 4-4 配准结果可视化

Fig. 4-4 Visualization of registration results

4.3.3 消融实验

本文提出的基于多模态融合的锚点定位点云配准方法为解决点云配准提供了一种新的思路，为了验证对齐模块和融合模块的有效性，本节将对这两个模块进行消融实验。

4.3.3.1 对齐模块的效果

表4-6展示了对齐模块的对本点云配准框架的作用，实验结果显示了超点匹配率、特征召回率、内点率和匹配召回率四个性能指标。从表4-6可以看出，在使用对齐模块之后所有的性能指标都得到了有效提高。这说明在多模态特征融合之前进行模态间的数据对齐，寻找点云中的点和图像中的像素的对应有助于更好的特征融合。

表 4-6 对齐模块消融实验

Table 4-6 Ablation experiments of alignment module

| Model | 3DMatch | | | | 3DLoMatch | | | |
|---------------|---------|------|------|------|-----------|------|------|------|
| | PIR | FMR | IR | RR | PIR | FMR | IR | RR |
| w/o Alignment | 83.4 | 98.0 | 68.8 | 89.0 | 53.1 | 85.6 | 42.4 | 71.8 |
| Ours | 85.4 | 98.0 | 70.2 | 90.6 | 51.4 | 87.6 | 41.2 | 73.1 |

4.3.3.2 融合模块的效果

在特征融合过程中，与以往的特征融合方式不同，本方法在两个子空间将超点特征和它所对应的像素特征进行了融合，表4-7验证本融合方式的有效性。其中cFusion表示直接将点云特征与图像特征进行拼接，aFusion表示将点云和图像特征直接利用注意力进行融合，iFusion表示仅在模态无关子空间进行特征融合。可以

表 4-7 融合模块消融实验

Table 4-7 Ablation experiment of fusion module

| Model | 3DMatch | | | | 3DLoMatch | | | |
|---------|---------|------|------|------|-----------|------|------|------|
| | PIR | FMR | IR | RR | PIR | FMR | IR | RR |
| cFusion | 82.7 | 97.6 | 68.3 | 87.8 | 48.5 | 84.5 | 39.3 | 69.7 |
| aFusion | 83.4 | 97.6 | 68.6 | 88.8 | 49.5 | 83.9 | 39.7 | 70.3 |
| iFusion | 84.4 | 97.8 | 70.1 | 89.9 | 51.8 | 85.8 | 42.2 | 72.2 |
| Ours | 85.4 | 98.0 | 70.2 | 90.6 | 51.4 | 87.6 | 41.2 | 73.1 |

看到融合模块能够有效提高各个阶段的实验性能。这说明通过将两种模态特征投影到模态无关子空间能够有效减少模态间的域差异，同时在模态相关子空间的特征融合能够减少信息的丢失。

4.4 本章小结

本章提出了一个基于多模态特征融合的锚点定位点云配准方法，在第三章的基于锚点几何嵌入点云配准基础框架中的锚点定位阶段引入了多模态融合辅助定位。一方面，受到3D目标检测同行多模态融合方法的启发，该方法利用一个对齐模块寻找点云的点与图像的像素之间的对应关系，实现像素到点的准确映射。另一方面，该方法加入模态相关与无关的特征学习，旨在在模态无关子空间中减少模态间特征差异，同时利用注意力机制融合不同模态的信息，并在模态相关子空间中融合最终特征防止信息丢失。实验表明，本章提出的方法对如何融合点云和图像信息并最终受益于点云配准方法提供了很好的解决思路，同时提高了点云配准的性能。相比于其他方法，本方法在多模态融合之前进行了模态间的对齐，提升了多模态融合的效果，取得了更好的点云配准任务的实验结果。

第5章 总结与展望

5.1 主要结论

随着科技的发展，计算机三维视觉与生产生活联系的越来越紧密，对点云的自动化处理成为了一个急切的需求。在自动驾驶的城市建图，在机器人领域的姿态估计以及虚拟现实中的三维建模，点云配准的应用广泛分布于各个生产生活环节。如何快速高效的实现各种复杂真实场景的点云配准是一项具有挑战和意义的研究。针对点云配准任务，本文提出了一种基于显著锚点的点云配准框架，它通过第三章的几何嵌入增加弱几何区域的特征差异性，通过引入第四章的多模态融合模块增加锚点的显著性。其中第四章可以看成是

针对第三章方法锚点选择的局限性做的进一步改进。两个方法都在真实场景的点云配准数据集上进行了实验和评估，实验结果表明本文提出的方法能够有效地在低重叠率得场景中，增强特征间的差异性提高相似区域重复模式的匹配成功率，并实现了当前点云配准方法的先进水平。两个方法得贡献如下：

(1) 在基于显著锚点的点云配准方法中，通过提取显著锚点利用锚点与超点、超点与超点间的距离和角度等几何结构信息进行特征嵌入，增加了相似不重叠区域的差异性，能够有效提高超点匹配的内点率。该方法使用 KPConv 网络来提取点云局部区域的超点特征。利用一个锚点定位模块选取若干保持一定几何结构的高置信度的锚点，并通过注意力机制对点云中的超点进行结构嵌入并寻找超点间的对应关系。最后，在经过将超点对应扩充为点对应之后，利用一个局部到全局的姿态估计得到最终的变换矩阵。

(2) 在基于多模态特征融合的锚点定位点云配准方法中，通过融合点云和图像两种模态的信息，提高锚点选择的可靠性。该方法首先利用对其模块，将不同模态的两种数据进行对齐，寻找到点云到像素的对应关系。在多模态融合模块，将两种模态的特征分解为模态相关和模态无关的特征，并在模态无关子空间中缩小特征间的域间隙减少噪声干扰。并最终与模态相关的特征融合以减少信息的丢失形成最终的特征，实现锚点定位。

5.2 研究展望

本文提出了基于显著锚点几何嵌入的点云配准和基于多模态融合的锚点定位点云配准方法，虽然两个方法都在低重叠率的真实场景数据集上取得了不错的效果，但是仍然存在一些可以改进的地方。本文的第三章提出了一种基于显著锚点

几何嵌入的点云配准方法，其核心思想是通过多个保持一定几何机构的锚点缓解相似不重叠区域特征过度平滑问题。虽然该方法取得了一定的效果，但是仍然存在一些尚需改进的地方：(1) 首先该方法虽然设计了一个锚点定位模块并利用迭代优化更新显著锚点，但是这种设计产生的锚点在某些场景中依然会失败，并导致最终的结果相较于一般方法较差。为此，需要设计一个更加鲁棒的锚点定位模块使得整体网络更加稳定，可以考虑设计一个损失函数来有效监督锚点对应。(2) 其次在使用由粗到细的点云配准框架之后，整个网络的模型较大导致训练时间较长，如何有效轻量化模型是一个急需解决的问题。后续可以通过提高下采样倍率减少超点个数，进而减少几何嵌入过程的时间开销。

本文的第四章提出以一种基于多模态融合的锚点定位点云配准方法，其核心思想是通过将图像信息和点云信息进行特征融合提高锚点对应的准确性。该方案针对第三章框架做出一点改进并取得了一定效果，但也存在一些问题：(1) 点云数据集中每个点云实际是由 50 张 RGB-D 图像合成，而文章中仅采取某一视角下的一张图片进行融合导致某些点在该视角下被遮挡找不到准确的图像信息，但是如果使用全部图像又会导致时间花销较大，故需要有效实验对图像数量对模态融合的影响进行分析。(2) 在多模态融合模块中，将两种模态的特征投影到模态相关和模态无关的两个子空间并分别学习模态相关与无关的特征表示，后续工作可以重新设计更加适应点云和图像融合的损失函数对其进行监督以达到预设效果。同时在最终的融合过程中使用多头注意力机制完成，后续工作可以考虑使用其他的多模态融合方法。

如何增加点特征间的差异性缓解特征的过渡平滑是点云配准任务的关键，提出的两个方法都是借助锚点嵌入几何信息来提取最终特征。但是由于点云点的个数较多，在做特征提取时导致时间和空间花销较大，影响了整个模型的训练效率。在后续工作中，可以考虑使用与训练好的特征提取网络来提前提取好特征，训练时则读取相应的特征，以减少训练过程中的时间花销是整个网路的参数量大大减少。同时随着点云和图像多模态融合的研究越发深入，对于这两种模态如何更好地融合有了更多的考量。相比于直接利用图像特征修饰点云，使用特征间的融合更加有效；相比于特征间的隐式融合，通过对齐两种模态的显式特征融合更加优越，后续工作可以对点云和图像特征的融合方式开展广泛的实验研究。同时为了能够将现有网络模型部署到例如火星探测等相关实验平台，设计轻量化的网络结构也是一种新的研究方向。

参考文献

- [1] UY M A, LEE G H. PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018 : 4470–4479.
- [2] 赵夫群, 周明全. 文物点云模型的优化配准算法 [J]. 计算机应用研究, 2017, 34(12) : 4.
- [3] 宋丽梅. 双目立体机器视觉检测系统及其应用 [J]. 西南科技大学学报, 2006, 21(1) : 30–34.
- [4] SOBREIRA H, COSTA C M, SOUSA I, et al. Map-matching algorithms for robot self-localization: a comparison between perfect match, iterative closest point and normal distributions transform[J]. Intelligent & Robotic Systems, 2019, 93 : 533 – 546.
- [5] CATTANEO D, VAGHI M, VALADA A. LCDNet: Deep loop closure detection and point cloud registration for LiDAR SLAM[J]. IEEE Transactions on Robotics, 2022, 38(4) : 2074 – 2093.
- [6] LIU Z, ZHOU S, SUO C, et al. LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis[C]. IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019 : 2831 – 2840.
- [7] POMERLEAU F, COLAS F, SIEGWART R, et al. A review of point cloud registration algorithms for mobile robotics[J]. Foundations and Trends® in Robotics, 2015, 4(1) : 1 – 104.
- [8] BESL P, MCKAY N D. A method for registration of 3-D shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(2) : 239 – 256.
- [9] GOLD S, RANGARAJAN A, LU C-P, et al. New algorithms for 2D and 3D point matching: pose estimation and correspondence[J]. Pattern Recognition, 1998, 31(8) : 1019 – 1031.

- [10] YANG J, LI H, JIA Y. Go-ICP: Solving 3D registration efficiently and globally optimally[C]. IEEE/CVF International Conference on Computer Vision, Sydney, Australia, 2013 : 1457–1464.
- [11] ALMOHAMAD H, DUFFUAA S O. A linear programming approach for the weighted graph matching problem[J]. IEEE Transactions on pattern analysis and machine intelligence, 1993, 15(5) : 522–525.
- [12] HUANG X, ZHANG J, FAN L, et al. A systematic approach for cross-source point cloud registration by preserving macro and micro structures[J]. IEEE Transactions on Image Processing, 2017, 26(7) : 3261–3276.
- [13] ZHOU F, De la TORRE F. Factorized graph matching[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012 : 127–134.
- [14] LEORDEANU M, HEBERT M. A spectral technique for correspondence problems using pairwise constraints[C]. IEEE/CVF International Conference on Computer Vision, Beijing, China, 2005 : 1482–1489.
- [15] EVANGELIDIS G D, KOUNADES-BASTIAN D, HORAUD R, et al. A generative model for the joint registration of multiple point sets[C]. European Conference on Computer Vision, Zurich, Switzerland, 2014.
- [16] MYRONENKO A, SONG X. Point set registration: coherent point drift[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(12) : 2262–2275.
- [17] FAN J, YANG J, AI D, et al. Convex hull indexed Gaussian mixture model for 3D point set registration[J]. Pattern Recognition, 2016, 59 : 126–141.
- [18] AOKI Y, GOFORTH H, SRIVATSAN R A, et al. PointNetLK: robust & efficient point cloud registration ssing pointNet[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seoul, Korea, 2019 : 7156–7165.
- [19] BAKER S, MATTHEWS I. Lucas-Kanade 20 years on: A unifying framework[J]. International Journal of Computer Vision, 2004, 56 : 221–255.
- [20] DENG H, BIRDAL T, ILIC S. PPFNet: Global context aware local features for robust 3D point matching[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018 : 195–205.

- [21] HUANG X, MEI G, ZHANG J. Feature-Metric registration: A fast semi-Supervised approach for robust point cloud registration without correspondences[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020 : 11363 – 11371.
- [22] XU H, LIU S, WANG G, et al. OMNet: Learning overlapping mask for partial-to-partial point cloud registration[C]. IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021 : 3112 – 3121.
- [23] CHARLES R Q, SU H, KAICHUN M, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017 : 77 – 85.
- [24] QI R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in Neural Information Processing Systems, 2017, 30 : 5099 – 5108.
- [25] YEW Z J, LEE G H. 3DFeat-Net: Weakly supervised local 3D features for point cloud registration[C]. European Conference on Computer Vision, Munich, Germany, 2018 : 630 – 646.
- [26] GOJCIC Z, ZHOU C, WEGNER J D, et al. The Perfect Match: 3D Point Cloud Matching With Smoothed Densities[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019 : 5540 – 5549.
- [27] WANG Y, SUN Y, LIU Z, et al. Dynamic graph cnn for learning on point clouds[J]. Acm Transactions On Graphics, 2019, 38(5) : 1 – 12.
- [28] THOMAS H, QI C R, DESCHAUD J-E, et al. KPConv: Flexible and deformable convolution for point clouds[C]. IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019 : 6410 – 6419.
- [29] ZENG A, SONG S, NIEßNER M, et al. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017 : 199 – 208.
- [30] CHOY C, PARK J, KOLTUN V. Fully convolutional geometric features[C]. IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019 : 8957 – 8965.

- [31] AO S, HU Q, YANG B, et al. Spinnet: Learning a general surface descriptor for 3d point cloud registration[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, virtual, 2021 : 11753 – 11762.
- [32] BAI X, LUO Z, ZHOU L, et al. D3Feat: Joint learning of dense detection and description of 3D local features[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020 : 6358 – 6366.
- [33] HUANG S, GOJCIC Z, USVYATSOV M, et al. PREDATOR: Registration of 3D point clouds with low overlap[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, virtual, 2021 : 4265 – 4274.
- [34] BAI X, LUO Z, ZHOU L, et al. PointDSC: Robust point cloud registration using deep spatial consistency[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, virtual, 2021 : 15854 – 15864.
- [35] WANG Y, SOLOMON J. Deep Closest Point: Learning representations for point cloud registration[C]. IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019 : 3522 – 3531.
- [36] LI J, ZHANG C, XU Z, et al. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration[C]. European Conference on Computer Vision, Glasgow, UK, 2020 : 378 – 394.
- [37] CHOY C, DONG W, KOLTUN V. Deep global registration[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020 : 2511 – 2520.
- [38] MIN T, SONG C, KIM E, et al. Distinctiveness oriented positional equilibrium for point cloud registration[C]. IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021 : 5490 – 5498.
- [39] LI Q, HAN Z, WU X-M. Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning[C]. Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018 : 3538 – 3545.
- [40] CHEN D, LIN Y, LI W, et al. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(04) : 3438 – 3445.

- [41] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6) : 381–395.
- [42] BEHLEY J, GARBADE M, MILIOTO A, et al. Semantickitti: A dataset for semantic scene understanding of lidar sequences[C]. Proceedings of the IEEE/CVF international conference on computer vision, 2019 : 9297–9307.
- [43] YUAN Z, ZENG W, SU Y, et al. Density-guided Translator Boosts Synthetic-to-Real Unsupervised Domain Adaptive Segmentation of 3D Point Clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024 : 23303–23312.
- [44] WANG D, SHANG Y. A new active labeling method for deep learning[C]. 2014 International joint conference on neural networks (IJCNN), 2014 : 112–119.
- [45] ROTH D, SMALL K. Margin-based active learning for structured output spaces[C]. Machine Learning: ECML 2006: 17th European Conference on Machine Learning Berlin, Germany, September 18-22, 2006 Proceedings 17, 2006 : 413–424.
- [46] SETTLES B. Active learning literature survey[J], 2009.
- [47] NING M, LU D, WEI D, et al. Multi-anchor active domain adaptation for semantic segmentation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021 : 9112–9122.
- [48] HUANG D, LI J, CHEN W, et al. Divide and adapt: Active domain adaptation via customized learning[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023 : 7651–7660.
- [49] SU J-C, TSAI Y-H, SOHN K, et al. Active adversarial domain adaptation[C]. Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2020 : 739–748.
- [50] PRABHU V, CHANDRASEKARAN A, SAENKO K, et al. Active domain adaptation via clustering uncertainty-weighted embeddings[C]. Proceedings of the IEEE/CVF international conference on computer vision, 2021 : 8505–8514.

- [51] XIE B, YUAN L, LI S, et al. Active learning for domain adaptation: An energy-based approach[C]. Proceedings of the AAAI conference on artificial intelligence : Vol 36, 2022 : 8708 – 8716.
- [52] XIE B, LI S, GUO Q, et al. Annotator: A generic active learning baseline for lidar semantic segmentation[J]. Advances in Neural Information Processing Systems, 2023, 36.
- [53] WELFORD B P. Note on a method for calculating corrected sums of squares and products[J]. Technometrics, 1962, 4(3) : 419 – 420.
- [54] CHEN S, JIA X, HE J, et al. Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021 : 11018 – 11027.
- [55] WANG Y, YIN J, LI W, et al. Ssda3d: Semi-supervised domain adaptation for 3d object detection from point cloud[C]. Proceedings of the AAAI Conference on Artificial Intelligence : Vol 37, 2023 : 2707 – 2715.
- [56] CHOY C, GWAK J, SAVARESE S. 4d spatio-temporal convnets: Minkowski convolutional neural networks[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019 : 3075 – 3084.
- [57] GAO Y, LI J, ZHOU Y, et al. Optimization methods for large-scale machine learning[C]. 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2021 : 304 – 308.
- [58] VORA S, LANG A H, HELOU B, et al. PointPainting: Sequential fusion for 3D object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020 : 4603 – 4611.
- [59] CHEN X, KUNDU K, ZHANG Z, et al. Monocular 3D object detection for autonomous driving[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016 : 2147 – 2156.
- [60] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017 : 6526 – 6534.

- [61] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Spain, 2018 : 1 – 8.
- [62] LI Y, YU A W, MENG T, et al. Deepfusion: Lidar-camera deep fusion for multi-modal 3d object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022 : 17182 – 17191.

作者简介

1. 攻读学位期间的研究成果

(一) 发表的学术论文和著作

[1] 第 2 作者 (导师为第 1 作者) . IEEE Transactions on Circuits and Systems for Video Technology.(在投)

(二) 申请 (授权) 专利

[1] 第 2 作者 (导师为第 1 作者) . 2023.01.29.

[2] 第 2 作者 (导师为第 1 作者) . 2023.02.24.

(三) 参与的科研项目及获奖

致 谢

致 谢