

重庆邮电大学  
CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

硕士学位论文  
MASTER THESIS



论文题目 面向三维点云语义分割的  
主动域自适应方法研究

学科专业 计算机科学与技术  
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

## 面向三维点云语义分割的主动域自适应方法研究

指导教师

申请学位级别 硕士 学科专业 计算机科学与技术

答辩委员会主席 论文答辩日期 2025 年 5 月 20 日

学位授予单位和日期 重庆邮电大学 2025 年 6 月

# **Research on Active Domain Adaptive Method for Semantic Segmentation of 3D Point Clouds**

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

Discipline **Computer Science and Technology**

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

点云语义分割是三维场景理解的核心任务，其目标在于将场景中每个点准确地映射到预设的语义类别上，广泛应用于无人驾驶、智能机器人以及虚拟和增强现实等领域。随着深度学习技术的迅速发展，大规模点云数据的高效处理成为可能，并催生了大量优秀的开源模型。然而，尽管点云数据采集变得更加便捷，但全监督训练要求逐点标注的数据依然难以获取，因为大规模点云逐点标注是一项耗时耗力的工作。此外，由于不同采集设备和场景的差异，将源数据集上训练的模型直接应用于新目标数据集时常会出现显著的域间隙，从而导致性能下降。域自适应作为解决域间隙问题和减少对标注依赖的重要策略，目前主要采用无监督或半监督模式，虽然节省了标注成本，但往往会牺牲部分模型性能。相比之下，主动域自适应在人工成本与分割精度之间提供了更具性价比的平衡方式。然而，目前针对三维点云语义分割主动域适应尚未提出专门的主动查询策略，也缺乏将主动选择的目标域点与已标注的源域数据有效结合的方案。针对上述问题，本文对主动学习和域自适应在视觉语义分割中的研究进展和方法进行了归类总结，阐述了现有方法的优缺点。此外，进一步提出了有效的针对三维点云语义分割的域自适应的主动学习方法以及充分发挥两域潜能的主动混合方法，方法如下所示：

1. 基于点云语义分割域适应的主动学习方法，其主要贡献如下：1) 首创原型指导的域差异感知查询策略，通过计算目标域点云特征与源域类别原型中心的余弦距离，构建“特征距离-预测熵”联合评估指标，实现跨域场景下高迁移价值点的选择。2) 首次将主动学习与混合策略进行结合并用于点云语义分割域适应任务，通过混合策略将主动学习选择的目标点与源域点进行混合，构建出强健的中间域数据，进一步缩减域间隙。
2. 基于点云语义分割域适应的主动混合策略，其主要贡献如下：1) 提出一种源-目标数量平衡算法，混合标注数量相同的源和目标域点，使得模型可以学到更加平衡的两域知识，解决模型学习过程中的域偏移累积问题。2) 在源-目标数量平衡模块的基础上提出类别平衡主动混合算法，解决混合中间域类别不平衡问题，进一步提升模型的性能。

本文提出的方法在合成到真实，真实到真实的跨域任务上做了大量实验和验证。通过与现有的同类优秀域自适应方法进行对比，实验结果和可视化结果充分证明了方法的优越性，在极少数标签的情况下达到了超越全监督的效果。

**关键词：**点云处理，三维视觉，语义分割，域适应，主动学习，混合方法

## ABSTRACT

Point cloud semantic segmentation is the core task of 3D scene understanding, and its goal is to accurately map each point in the scene to a preset semantic category, which is widely used in the fields of unmanned driving, intelligent robotics, and virtual and augmented reality. With the rapid development of deep learning technology, efficient processing of large-scale point cloud data has become possible and has given rise to a large number of excellent open-source models. However, although point cloud data acquisition has become more convenient, fully supervised training requiring point-by-point labeling is still difficult to obtain, because point-by-point labeling of large-scale point clouds is a time-consuming and labor-intensive task. In addition, due to the differences in different acquisition devices and scenarios, there are often significant domain gaps when directly applying models trained on source datasets to new target datasets, resulting in performance degradation. As an important strategy to solve the domain gap problem and reduce the dependence on annotation, domain adaptation currently mainly adopts unsupervised or semi-supervised modes, which saves the annotation cost but often sacrifices part of the model performance. In contrast, active domain adaptation provides a more cost-effective way to balance between labor cost and segmentation accuracy. However, no specialized active query strategy has yet been proposed for active domain adaptation for semantic segmentation of 3D point clouds, and there is a lack of schemes to effectively combine the actively selected target domain points with the annotated source domain data. To address the above issues, this thesis categorizes and summarizes the research progress and methods of active learning and domain adaptation in visual semantic segmentation, and describes the advantages and disadvantages of the existing methods. In addition, effective domain-adaptive active learning methods for semantic segmentation of 3D point clouds as well as active mixing methods that fully utilize the potential of both domains are further proposed as shown below:

1. A Domain discrepancy aware active learning for cross-domain segmentation, its main contributions are as follows: 1) The first prototype-guided domain discrepancy awareness query strategy, by calculating the cosine distance between the point cloud features of the target domain and the prototype centers of the source domain categories, and constructing a joint evaluation index of “feature distance-predictive entropy” to achieve the selection of points with high migration value in cross-domain scenarios. 2) For the

---

## ABSTRACT

---

first time, this thesis combined the active learning with the mixing strategy and used it for the domain adaptation task of point cloud semantic segmentation, and constructed a robust intermediate domain data to further reduce the domain gap by mixing the target points selected by active learning with source domain points.

2. Active mixing method for cross-domain semantic segmentation, its main contributions are as follows: 1) Propose a source-target amount balance algorithm with the source-target amount balance algorithm, mixing the source and target domain points with the same number of labeling, so that the model can learn a more balanced knowledge of the two domains, and solving the problem of accumulating domain bias in the process of model learning. 2) Propose the class-balanced active mixing algorithm on the basis of the source-target quantity balance module, solving the mixing intermediate domain class imbalance problem and further improve the performance of the model.

The method proposed in this thesis has been extensively experimented and validated on synthetic-to-real and real-to-real cross-domain tasks. The experimental and visualization results fully demonstrate the superiority of the method by comparing with similar existing excellent domain adaptive methods, even achieving results beyond full supervision in very few labeled cases.

**Keywords:** Point Cloud Processing, 3D Vision, Semantic Segmentation, Domain Adaptation, Active Learning, Mixing Methods

## 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 基于深度学习的点云语义分割方法 .....	2
1.2.2 视觉语义分割中的主动学习方法 .....	5
1.2.3 视觉语义分割域自适应方法 .....	8
1.2.4 视觉语义分割中的 Mixng 方法 .....	10
1.3 论文研究主要内容 .....	11
1.4 论文组织结构 .....	12
第2章 点云语义分割主动域适应相关理论基础 .....	14
2.1 本章引言 .....	14
2.2 点云语义分割相关知识 .....	14
2.2.1 三维点云理论基础 .....	14
2.2.2 点云语义分割方法 .....	15
2.2.3 评价指标 .....	17
2.3 主动学习基础知识 .....	18
2.4 域适应基础知识 .....	18
2.4.1 域对齐的典型方法 .....	19
2.5 跨域点云语义分割数据集 .....	20
2.5.1 SynLiDAR .....	21
2.5.2 SemanticKITTI .....	21
2.5.3 SemanticPOSS .....	23
2.5.4 nuScenes .....	23
2.6 本章小结 .....	24
第3章 基于点云语义分割域适应的主动学习方法 .....	25
3.1 本章引言 .....	25
3.2 研究动机及贡献 .....	25

3.3 基于点云语义分割域适应的主动学习方法 .....	26
3.3.1 问题陈述.....	26
3.3.2 方法概述.....	27
3.3.3 源域原型构建.....	27
3.3.4 源域原型指导的数据选择 .....	29
3.3.5 动态混合中间域构建 .....	31
3.4 实验评估 .....	31
3.4.1 实验设置.....	31
3.4.2 实验结果.....	32
3.4.3 消融对比实验.....	38
3.5 本章小结 .....	39
<b>第 4 章 基于点云语义分割域适应的主动混合方法 .....</b>	<b>41</b>
4.1 本章引言 .....	41
4.2 研究动机及贡献.....	41
4.3 基于点云语义分割域适应的主动混合方法 .....	42
4.3.1 方法概述.....	42
4.3.2 源-目标数量平衡算法 .....	43
4.3.3 类别平衡主动混合算法.....	44
4.3.4 损失函数.....	46
4.4 实验评估 .....	46
4.4.1 实验设置.....	46
4.4.2 实验结果.....	46
4.4.3 消融对比实验.....	51
4.5 本章小结 .....	54
<b>第 5 章 总结与展望 .....</b>	<b>55</b>
5.1 主要结论 .....	55
5.2 研究展望 .....	56
<b>参考文献 .....</b>	<b>58</b>
<b>作者简介 .....</b>	<b>65</b>
1. 攻读学位期间的研究成果 .....	65
1.1 发表的学术论文和著作 .....	65
1.2 申请（授权）专利 .....	65
1.3 参与的科研项目及获奖 .....	65
<b>致 谢 .....</b>	<b>66</b>

## 图目录

图 2-1 3D 点云 .....	15
图 2-2 主动学习 .....	18
图 2-3 域适应 .....	19
图 2-4 SynLiDAR 数据集 <sup>[53]</sup> .....	21
图 2-5 SemanticKITTI 数据集 <sup>[6]</sup> .....	22
图 2-6 SemanticPOSS 数据集 <sup>[8]</sup> .....	23
图 2-7 nuScenes 数据集 <sup>[7]</sup> .....	24
图 3-1 基于点云语义分割域适应的主动学习方法框架 .....	27
图 3-2 源域原型构建 .....	28
图 3-3 源域原型指导的数据选择 .....	29
图 3-4 动态混合中间域构建 .....	31
图 3-5 第三章 SynLiDAR→SemanticKITTI 分割可视化图 .....	34
图 3-6 第三章 SynLiDAR→SemanticPOSS 分割可视化图 .....	35
图 3-7 第三章 SemanticKITTI→nuScenes 分割可视化图 .....	36
图 3-8 第三章 nuScenes→SemanticKITTI 分割可视化图 .....	37
图 4-1 基于点云语义分割域适应的主动混合方法框架 .....	43
图 4-2 源-目标数量平衡模块 .....	44
图 4-3 类别平衡主动混合模块 .....	45
图 4-4 第四章方法在合成到真实场景分割可视化图 .....	50
图 4-5 第四章方法在真实到真实场景分割可视化图 .....	51
图 4-6 第四章方法结合不同主动学习在多个混合比例下的对比结果 .....	52

## 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 .....	33
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 .....	34
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 .....	36
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 .....	37
表 3-5 第三章主动学习方法与其他传统主动学习方法对比 .....	38
表 3-6 第三章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 .....	39
表 3-7 第三章方法消融实验 .....	39
表 4-1 第四章方法与其他域适应方法在 SynLiDAR→SemanticKITTI 数据上 的比较 .....	47
表 4-2 第四章方法与其他域适应方法在 SynLiDAR→SemanticPOSS 数据上 的比较 .....	48
表 4-3 第四章方法与其他域适应方法在 SemanticKITTI→nuScenes 数据上的 比较 .....	49
表 4-4 第四章方法与其他域适应方法在 nuScenes→SemanticKITTI 数据上的 比较 .....	50
表 4-5 第四章方法与其他 Mixing 方法在 SynLiDAR→SemanticKITTI 数据上 的比较 .....	51
表 4-6 第四章方法结合不同主动学习方法结果对比 .....	53
表 4-7 第四章方法消融实验 .....	54

# 第1章 绪论

## 1.1 研究背景及意义

随着科技进步和经济发展，人工智能已悄然融入人们的日常生活，智能产品正潜移默化地改变着生活方式、交通出行和娱乐方式。其中，无人驾驶<sup>[1]</sup>、机器人<sup>[2]</sup>以及增强现实<sup>[3]</sup>等领域备受关注。这些技术的出现不仅使人们的生活更加便捷，出行更为高效，同时也丰富了娱乐体验。然而，在机器人和自动驾驶等应用中，环境感知能力是实现高级智能的核心，一个可靠的感知系统对于自动避障和环境识别至关重要。为确保实际应用中的安全性与准确性，优秀的感知系统往往需要集成多种传感器进行多模态数据采集，而激光雷达正是其中的一个关键传感器。激光雷达通过固定频率和角度发射激光，接收返回信号以采集周围环境信息，生成包含三维空间几何信息的点云数据，这一点是二维图像所无法提供的。此外，其采集的数据不受光照变化影响，能够真实还原物体的尺寸和形态，基于这些优势，激光雷达已被广泛应用于各种智能系统中。因此，如何高效利用三维点云数据实现精准感知，已成为当前非常重要的研究课题。点云语义分割是实现三维场景感知的关键任务。在无人驾驶和机器人领域，通过对周围环境和道路进行语义分割，可以为不同类别的点赋予相应的语义标签，从而将原始点云数据转化为含有语义信息的三维场景数据。基于此，系统能够依据环境中的语义标签有效区分不同物体，并据此调整导航路径，确保车辆或机器人的正确行进方向。在增强现实中，准确的语义分割则是实现虚拟内容与真实环境无缝融合的前提，只有对真实世界中的物体及其边界进行精确感知和分类，才能将虚拟场景合理地叠加到现实世界中，达到增强现实的效果。尽管点云语义分割的应用价值巨大，早在深度学习兴起之前便已引起学术界广泛关注，但受限于传统机器学习方法在处理大规模点云数据时的效率和性能不足，其研究进展较为缓慢。随着深度学习技术的发展，端到端的点云语义分割方法在大规模数据处理和分割精度上均取得了显著提升，大量优秀模型如雨后春笋般涌现，为该领域注入了新的动力。值得注意的是，由于点云数据集根据获取场景的不同又可以分为室内点云数据集<sup>[4,5]</sup> 和室外点云数据集<sup>[6-8]</sup>，它们各具特点。室内数据集密度均匀且规模相对较小，多用于增强现实和机器人等应用；而室外数据集则呈现近密远疏的特性，覆盖的场景更为广阔，通常用于无人驾驶等领域。尽管基于深度学习的算法在这两类数据集上均取得了显著进展，并为实际应用带来了美好前景，但这些算法普遍依赖于逐点标注数据进行全监督训练。然而现实中，一帧点云通常包含十多万个点，一个场景可能包含数千帧点云，而一个完整的数据集则可能涉及十多个不同场景。由于点云本质上是无序的几何数据，其标注工作需

要专业人士操作，因此逐点标注不仅复杂而且极为耗时<sup>[6]</sup>。高昂的标注成本已成为制约这些先进算法落地应用的重要挑战。因此，如何在标注成本与模型性能之间取得平衡，以及如何利用已有数据或仅用少量标注数据训练出高性能的分割模型，成为亟待解决的关键问题。域适应作为解决上述问题的重要研究分支，其基本思路是利用已有逐点标注的全监督数据训练预训练模型，再将该模型迁移到未标注的目标域中，同时解决由于场景差异或传感器差异所引起的域偏移问题。常见的域适应方法包括无监督域适应、半监督域适应和主动域适应。无监督域适应虽然不依赖目标域标注，但其效果与全监督方法仍存在较大差距；半监督域适应虽然优于无监督方法，但由于目标域中标注数据通常为随机采样，受域偏移影响较大，其作用未能充分发挥。相比之下，主动域适应通过结合主动学习方法，从目标域中选择对模型迁移最有价值的样本进行标注，既有效缩小域间隙、提升了模型性能又降低了标注成本。然而，目前基于语义分割的主动域适应研究主要集中在二维图像领域，而在三维点云领域尚缺乏系统性的探索。此外，传统的主动学习方法难以直接应用于域适应场景；同时，由于主动学习选择的样本数量较少，如何充分利用源域和目标域的标注数据，既发挥各自潜力又有效防止域偏移累积和类别不平衡，仍是需要解决的问题。因此，本文的研究以主动学习和域适应为出发点，首先通过大量文献调研，基于域适应的特点寻找适用于三维点云语义分割任务的主动查询策略；随后，深入探索主动标注样本与源域数据的有效融合的主动混合方法，以期获得更加稳固的跨域特征表示。最后，通过大量实验对提出方法进行验证分析。

## 1.2 国内外研究现状

本文主要研究方向为点云语义分割任务下的主动域适应，因此重点是探索适合点云语义分割域适应任务的主动学习方法，并在此基础上进一步研究适合主动学习的混合（Mixing）方法，因此相关工作涉及到点云语义分割、主动学习、域适应以及 Mixing 数据增强等相关领域。所以本节将分别对这四个领域中最近几年的相关研究和进展进行分析和总结。

### 1.2.1 基于深度学习的点云语义分割方法

由于传统机器学习方法难以高效且精确地对大规模点云数据实现逐点分割，基于端到端的深度学习方法因其大批量数据处理能力已成为点云语义分割任务的主流方法。这些方法大致可以分为三类：基于二维投影的点云语义分割方法、基于体素的点云语义分割方法和基于点的点云语义分割方法。基于二维投影的方法通过不同的投影视角将无序非结构化的点云投影到结构化的二维图像中，投影方法

包括鸟瞰图<sup>[9]</sup>，球形<sup>[10-14]</sup>，多视角<sup>[15,16]</sup>等。然后利用更加成熟稳定的二维卷积网络去对这些点云图片进行特征提取和语义分割，并将分割后的结果映射回点云中。此类方法的核心思想是模态转变，将点云从三维转变为图像，然后就可以利用二维卷积网络处理。但是其不足也非常的明显，二维和三维模态差异导致其表征信息的方式也有所不同，模态的转换导致三维几何信息被破坏，进而造成信息的丢失。因此，完全基于投影的方法在随后的研究中，逐渐被人遗忘，研究者们转而开始寻找更加有效的点云语义分割的处理方式。有研究者认为图像模态下的点云可以提取不同于三维模态下的特征，为了充分利用图像提取到的信息，提出了混合点云三维特征和二维特征的多模态方法，这种方法将两者的处理结果相结合，形成了优势互补<sup>[17]</sup>，并成功取得非常可观的分割效果。基于体素的方法首先将三维点云数据转换为体素网格，即将原本离散且稀疏的点云数据规整成规则的三维格状结构，接着利用稀疏卷积技术在该结构上进行特征提取和表示学习，从而获得较为紧凑且具有表达能力的点云特征。这种方法在处理大规模室外场景时表现出较高的效率，因为规则化的体素结构便于卷积操作的实现，并可显著降低计算复杂度。然而，体素化过程中由于数据离散化的固有属性，往往会导致细节信息的损失，尤其是在高分辨率特征提取方面存在一定局限。这些方法中有一些是比较优秀且至今仍被作为基础框架研究的。MinkowskiNet<sup>[18]</sup> 提出了一种高可用到的稀疏卷积方法，并构建了一个专门用于稀疏张量的自动微分库。基于这一方法与工具，构建出一种四维卷积神经网络，用于时空感知任务，其设计能够同时捕捉空间与时间维度的信息，实现更为精准的特征提取。而之后提出的一些稀疏卷积方法<sup>[19,20]</sup> 则是对许多无意义的计算消耗进行了改进，进一步提高了计算效率。最新的基于体素的方法 OA-CNNs<sup>[21]</sup> 则提出了一种全自适应三维卷积神经网络，该网络由动态感受野和自适应关系映射组成：空间动态感受野能够根据输入数据的分布灵活调整感受区域；自适应关系卷积，用于捕捉并建模局部特征之间的复杂关联。因此其在一定程度上展现了超越 Transformer 架构的潜能。由于点到体素的转换，一定程度上缓解了计算消耗，因此这些方法更适用于大规模的室外点云场景处理。基于点的方法直接将点云输入到分割网络而不进行任何处理。在提出 PointNet<sup>[22]</sup> 架构之前，三维点云特征的提取都是手动完成的。PointNet 是第一个能够直接使用三维点云提取深度学习特征的架构。其通过提取点云特征中的最大值来处理点云的无序特性，然而这也会导致其无法利用点云中的局部结构信息，无法运用到大规模数据集上。而它的改进版本 PointNet++<sup>[23]</sup> 的主要贡献在于提出了分层结构，使用多个抽象层进行特征提取，从而实现邻域局部特征和全局特征的提取。受 PointNet 的影响和启发，更多的基于点的方法<sup>[24,25]</sup> 被提出。其中，KPCov<sup>[24]</sup> 提出了一种新型点卷积算法，其利用一组核点来确定卷积核权重的作用区域，从而在保留点云

完整几何信息的同时克服了传统点卷积的局限性。该方法借鉴了图像卷积的思想，但通过灵活定义核点的数量和位置，实现了对局部特征的高效提取，并以相关函数确定其影响范围。采用半径邻域和规则下采样策略，有效应对了非均匀采样问题，从而在大规模点云场景下保持较高鲁棒性和计算效率。此后，RandLA-Net<sup>[26]</sup>引入了一个局部特征聚合模块，有效地保留来自大范围邻域的有用特征，并在其方法中使用随机采样，以显著减少内存占用和计算成本。SCF-Net<sup>[27]</sup>提出了一个可学习的模块（Spatial Contextual Features, SCF），由三个部分组成：局部极坐标表示模块（Local Polar Representation, LPR）、双重距离注意力池化模块（Dual-Distance Attentive Pooling, DDAP）和全局上下文特征模块（Global Contextual Feature, GCF）。LPR 模块在极坐标系中构建 z 轴旋转不变的表示，以捕捉每个 3D 点的局部上下文。DDAP 模块通过利用几何距离和特征距离学习的权重，整合邻域点的表示，学习有效的局部特征。GCF 模块利用邻域的位置和体积比，学习每个 3D 点的全局上下文。SCF 模块可以嵌入到各种网络架构中用于点云分割，但其网络结构比较复杂。BBAF-Net<sup>[28]</sup>通过引入密集区域来增强局部上下文解决相邻点的模糊性，同时自适应地融合多分辨率特征，以获取关于点云的全面知识。而随着注意力机制下大模型的爆发，Transformer<sup>[29]</sup>首次将注意力机应用到点云语义分割任务中。随后的一些工作也将其加入到自己的方法中<sup>[30,31]</sup>，虽然注意力机制完美适配了点云的无序特性，然而这些方法大多需要更多的计算时间。TransformerV3<sup>[32]</sup>探索了一种新的序列化邻域机制，将点云按照特定模式组织起来。并采用简化方法取代了更复杂的注意力区域交互机制，从而更适用于序列化点云处理。同时消除了对相对位置编码的依赖，并采用了一种更简单的前置稀疏卷积层进行特征提取，从而提高计算效率。基于点的方法能够完整保留点云中每个点的原始几何信息和局部特征，这一优势使其在三维语义分割任务中展现出独特的潜能。当前虽已有不少相关方法取得了显著进展，但在深层次特征利用以及复杂场景的细粒度信息捕捉方面，仍存在大量未被充分挖掘的可能性。随着计算机视觉与深度学习技术的不断演进，未来在这一领域内定将涌现出更多创新方法，为三维数据的高效处理和精准理解提供更加完善的解决方案。虽然上述算法都在基于深度学习的点云语义分割任务中取得了优异的结果，但是这些方法仍然需要基于全标注的数据进行训练才能实现最佳的效果，而对于点云语义分割任务来说，逐点标注是极其复杂的一项工程。如何平衡性能与标注是一个非常重要的研究方向，主动域适应是其中一个方法，也正是本文的研究目的所在。

## 1.2.2 视觉语义分割中的主动学习方法

随着深度学习的发展，模型高性能与数据标注困难以及无标注但性能低的问题也逐渐显现，因此一些研究者开始将目光转向能够平衡性能与标注花费的主动学习上来。在视觉语义分割任务中，根据其筛选的基本单元的不同，大致可以将其分为基于样本的主动学习方法，基于区域的主动学习方法和基于最小语义单元的主动学习方法。

### 1.2.2.1 基于样本的主动学习方法

在主动学习的早期阶段，研究者们主要采用以样本作为基本查询单位的方法。这一策略源于主动学习最初在视觉分类任务中的应用，而在这些任务中，数据通常是以单个样本为单位进行划分和标注的。因此，许多针对语义分割任务的早期方法也从分类任务中获得了启发，采用在每一轮数据选择中筛选出信息量丰富的样本并进行逐点或者逐像素的标注。Tan 等人<sup>[33]</sup> 的研究重点在于充分挖掘类别的边缘信息。他们提出了一种方法，首先借助边缘检测器提取图像中的边缘特征，然后将这些边缘信息与利用 KL 散度评估的不确定性及数据代表性相融合，以实现主动样本选择。该方法将传统的人工设计理念与任务的实际需求紧密结合，取得了显著成效。Sinha 等人<sup>[34]</sup> 出了一种新的池化主动学习算法，该算法基于变分自编码器（VAE）和对抗网络，通过隐式学习采样机制来选择最具代表性的查询样本进行标注，从而提高模型性能并降低标注成本。Zhang 等人<sup>[35]</sup> 提出了一种状态重新标记对抗性主动学习模型，模型充分利用标注信息和状态信息，用于推导出对于当前模型来说信息最丰富的未标记样本。其设计了一种在线不确定性指示器，重新标记未标记数据的状态，赋予其不同的重要性，同时构建了一个无监督的图像重建器和一个监督的目标学习器，生成图像的统一表示，迭代嵌入标注信息。最后结合所提出的 k-center 方法的初始采样算法，使后续采样更加高效。但这些方法通常以全局不确定性或多样性为准则筛选样本，忽略了语义分割任务中不同区域的内在难度差异，导致标注资源浪费且模型在困难区域表现不足。为了解决这问题，Xie 等人<sup>[36]</sup> 提出了难度感知的主动学习方法 DEAL。该方法通过引入语义难度分支和设计像素级概率注意力模块动态学习不同语义区域的难度分数，并利用分割误差差作为监督信号，使模型准确捕捉那些易混淆或低置信度的像素区域，从而将主动学习的焦点从整体信息量转向局部难度感知。为了解决困难区域样本稀缺的问题，该方法进一步提出了一种双阶段样本获取策略，基于语义难度分数筛选出包含高难度区域的图像，并结合区域不确定性和多样性对像素级标注进行优先级排序，实现了从图像级粗筛选到像素级细粒度标注的渐进式优化。Huang 等人<sup>[37]</sup> 利用一种

新颖的无标签数据采样策略，结合半监督训练方案进行数据注释，从而利用无标签数据提升任务模型的性能。基于样本的主动学习方法虽然能够降低标注成本，但在视觉语义分割任务中也存在一些不足。由于一幅图像往往包含多个类别，而各类别的区域大小和分布差异明显，所选样本中的大部分信息可能仅集中在部分类别，其他区域则容易成为噪声。而且标注过程中常见的类别不平衡问题可能导致模型在训练时偏重于主流类别，从而忽略细节部分。此外，语义分割需要对每个像素或者点进行精细判断，单纯依赖样本级的选择往往难以捕捉到局部的复杂信息，这在处理复杂语义区域时尤为明显。

### 1.2.2.2 基于区域的主动学习方法

在语义分割任务中，基于样本的主动学习方法问题逐渐显现，一些研究者开始寻找缓解问题的方法。由于不同的区域对样本的影响程度不同，因此基于区域的主动学习方法被提出，这种方法将样本划分为不同的区域并作为查询和标注的基本单元。Qiao 等人<sup>[38]</sup> 提出了一种基于区域的主动学习方法框架 CRPA，通过结合全景和区域信息选择策略，有效平衡了标注工作量和模型性能。该框架引入了区域高斯注意力模块来缓解类别不平衡问题，并通过上下文标签扩展模块区域标注，进一步提高了标注效率。但是这种方法的划分是基于规则区域的，即将图像规则划分为同等大小的区域，然而语义类别在图像上的分布往往是不规则的，因此规则区域虽然有效降低了标注成本，却依然存在区域内类别不平衡或者分布不均匀的问题，导致少量的语义类别学习困难。Siddiqui 等人<sup>[39]</sup> 提出了一种基于不规则区域面向深度图像的主动学习方法。其设计了一种基于 KL 散度的最优视角选择标准，用于衡量预测概率分布在不同视角之间的变化。通过分析预测在不同视角下的不一致性来估计模型的不确定性，并将其定义为视角熵。最后将视角熵与超像素区域选择相结合，使得模型能够高效地筛选出高价值的样本区域。针对普通图像，Cai 等人<sup>[40]</sup> 也提出了一种类别平衡的采样策略，以提升超像素方法的性能，该策略能够优先选择来自低占比类别的高信息量样本。方法通过计算每个区域的最小边际不确定性，再依据每个区域主导类别所占比例对不确定性进行调整，从而获得更精准的不确定性度量。同时该文章也通过一系列的实验证明了不规则超像素在降低标注成本方面优于规则化的区域。而在三维点云语义分割的主动学习方法中，区域的划分根据点云的非结构化几何特征而呈现出多样性。Lin 等人<sup>[41]</sup> 提出了一种主动增量学习框架，通过迭代地选择高信息量样本，逐步丰富模型知识，从而有效降低标注需求。该方法将点云数据被划分为多个小块，并分为已标注和未标注两组。已标注的小块用于训练，在初始迭代中，网络从零开始训练。在后续迭代中，采用微调策略，使模型在已有知识的基础上不断扩展。每轮训练完成后，网络会根

据点熵、分割熵和互信息评估未标注点云小块的信息量，并选择最具信息价值的小块进行标注。Wang 等人<sup>[42]</sup>提出了一种结合准场景级弱标签的主动弱监督框架，用具有固定半径的球形子点云作为查询单位，每个子点云类别仅需一个点级标签，通过主动策略识别并选择最具信息量的子点云进行标注，从而同时兼顾场景级和点级语义信息。Xie 等人<sup>[43]</sup>则是将点云划分到同等大小的体素网格中，然后对网格中的点的类别进行预测，计算其内部类别点的熵程度，值越高说明体素网格中包含的不同的类别的点越多，体素网格中的点类别分布越平衡，包含信息越丰富。除了上述规则的方式以外，其实更多的则是通过不规则区域即超点的形式进行的区域划分。Shao 等人<sup>[44]</sup>提出了一种结合类平衡不确定性和多样性的采样策略。该策略引入加权超点不确定性估计，根据超点中类别数量的多少赋予这些不同类别点不同的权重，以更准确地衡量超点的不确定性。同时，通过空间结构多样性推理机制，构建出超点图和图聚合操作，将超点特征投影到多样性空间中，利用最远点采样选择最具代表性的超点进行标注。Hu 等人<sup>[45]</sup>提出了一种新颖的基于视角一致性的不确定性度量方法，以利用帧间约束信息。该方法基于连续 LiDAR 帧中预测分数函数的方差，设计了新的帧间散度和熵公式，对于一个未标注目标，如果其预测标签在不同帧之间存在较大差异，便认为模型预测不可靠，并选择这些最具不确定性的区域进行人工标注。与 Hu 利用时序帧间一致性的思路不同，Liu 等人<sup>[46]</sup>提出了一种基于多版本预测一致性的主动学习策略，通过数据增强构建输入点云的多个变体，以评估模型预测的稳定性。针对每个点在不同增强版本中的预测结果，计算其对应最高置信度类别的概率分布标准差作为不确定性度量，为缺乏连续帧数据的场景下提供了更灵活的标注策略。而 Wei 等人<sup>[47]</sup>则提出了一种基于尺寸平衡的主动学习方法，旨在解决点云语义分割中的类别不平衡与冷启动问题。该方法主要通过物体尺寸特征间接平衡类别分布，将不同尺寸的物体聚类为同一类别，在选择的时候优先选择各类别分区中信息量最大的簇进行标注。基于区域的方法由于在标注和性能上都有所突破，因此仍是语义分割任务中比较主流的方法。

### 1.2.2.3 基于最小语义单元的主动学习方法

语义分割任务中基于区域的主动学习方法显著优于基于样本主动学习的方法。为进一步优化标注效率，部分学者提出将选择粒度细化至最小语义单元，旨在通过精准标注关键局部区域，实现以更低成本的更高表现。在二维图像中，最小的语义单元是像素，因此在面向图像的基于最小语义单元的主动学习方法中，都是以像素为基本选择和标注单位。Shin 等人<sup>[48]</sup>提出了第一个基于像素的主动学习框架，该方法将基于边界的不确定性策略运用到了不同的图片像素间，并比较不同像素间的差异作为衡量指标。该方法证明了仅在少量像素点上提供标注的情况下，深度

神经网络能否依然取得良好表现，并证明在语义单元层面的标注，模型仍然能够有效学习语义信息。Rückin 等人<sup>[49]</sup>则是进一步将基于像素的主动学习方法与半监督学习相结合，并运用到了实际场景中。这一举措从实际运用层面证明了基于像素的主动学习标注的有效性和高效性。而在三维点云语义分割中，最小语义单元是点。Xu 等人<sup>[50]</sup>提出了第一个基于点的主动学习方法，该方通过加权计算每个点与不同下采样层次的周围领域信息的熵值总和，然后选择出不确定性排名靠前的点，同时通过一种特征距离抑制的方法来过滤空间距离相近且类别相似的冗余点。该方法不仅提高了模型性能，也大大增加了有效标注率。以上这些主动学习方法不仅显著降低了所需的标注比例，还能在保持模型性能达到全监督水平约 90% 的前提下，大幅降低人力成本。但是，传统的主动学习方法主要适用于单一数据集，对于存在显著域间差异的域适应任务则难以直接应用。

### 1.2.3 视觉语义分割域自适应方法

殊途同归，与主动学习不同的是域自适应方法则通过模型迁移的方式来解决目标域数据需要大量标注的问题。其核心思想是将在有标注的数据上训练好的模型迁移到无标注的目标域上去，并解决因域偏差而导致的模型性能下降问题。基于视觉语义分割的域自适应方法大致可以分为三类：无监督域自适应方法、半监督域自适应方法以及主动域自适应方法。

#### 1.2.3.1 无监督域适应

目前，在三维点云领域，研究者们更多地关注基于无监督的域自适应方法。当聚焦于点云语义分割任务时，根据跨域数据集的差异，这些方法又可以分为两类：真实到真实的场景，以及合成到真实的场景。真实到真实的无监督域自适应方法，通过使用 LiDAR 传感器捕获的真实场景数据进行深度网络训练，随后在不同 LiDAR 传感器捕获的未知场景中进行测试。在这种情况下，Yi 等人<sup>[51]</sup>将域适应形式化为 3D 表面补全任务，该方法通过设计并训练一个上采样网络，将雷达线束补全到统一的 64 线，然后再将补全后的数据输入到分割模型进行分割。而 Langer 等人<sup>[52]</sup>则通过射线投射将目标域的传感器模式转移到源域。在合成到真实的域自适应中，源数据通过模拟 LiDAR 传感器获得非真实场景的合成数据，而目标数据则由真实 LiDAR 传感器采集。在这种情况下，域偏移是由于采样噪声、环境结构和类别分布的差异引起的。Wu 等人<sup>[12]</sup>认为注意力模型可以用于聚合上下文信息，并且可以采用逐步域校准的测地线相关对齐来改善域自适应。在后续的文章中，Wu 等人<sup>[13]</sup>则是通过生成对抗网络在合成数据上模拟真实的点云噪声。类似地，Xiao 等人<sup>[53]</sup>将域偏移分解为外观差异和稀疏度差异，然后应用生成网络来减轻每个差异。Saltori

等人<sup>[54]</sup>则是通过对源域和目标域分别进行语义选择而后进行混合再放入到师生模型中进行训练，从而提高模型的性能。Ding 等人<sup>[55]</sup>提出了一个可以运用在室内数据集上的方法，其注意到了长尾类问题，设计了一个长尾类别注意机制，把已经打好伪标签的目标域长尾类别数据和源域进行混合后一同放入模型中训练。Bian 等人<sup>[56]</sup>提出了一个基于图的框架来探索两个域之间的局部级特征对齐，在两个域上动态构建局部特征图，并利用源域的图建立记忆库，使用最优传输来生成图匹配对，最后基于图的局部特征损失来调整两个域之间的特征分布。Jiang 等人<sup>[57]</sup>设计了一个可以提取领域共享特征和领域私有特征的模型，然后利用 GatedSCNN 网络使领域共享特征提取器能够在领域共享特征中保留边界信息，并利用学到的边界来完善分割结果。然而这些方法与在全监督下测得的目标数据集的性能上界仍存在很大的差异。

### 1.2.3.2 半监督域适应

为了解决无监督域适应与全监督性能差距过大问题，一些学者提出了基于半监督以及弱监督的域适应方法，但是这些方法目前在三维点云中出现的并不多，尤其是在三维语义分割领域中。Jaritz 等人<sup>[58]</sup>提出了第一篇点云语义分割领域的半监督域自适应。其使用跨模态的方式进行学习，利用点云几何先验信息以及图像的纹理颜色信息，并取二者的优点进行互补，使得提取的特征更加鲁棒，其具体实现是设计了一个双流、双头架构，一端是图像输入，一端是点云输入，在三维语义分割任务中对图像和点云模态应用了跨模态损失。跨模态损失包括应用于两种模态预测之间的 KL 散度，从而加强一致性。Chen 等人<sup>[59]</sup>提出了双层面域混合架构，从样本和区域层面混合两个域，生成两个老师模型，并通过这两个老师模型蒸馏知识给一个学生模型，使得学生模型更加强大，而其生成的伪标签将继续用于下一次双域混合中训练老师模型。Saito 等人<sup>[60]</sup>提出了一个由特征编码网络组成模型，使用一个分类层，用来计算特征与一组估计类别代表的相似性。通过交替地最大化未标记目标数据相对于分类器的条件熵和最小化其相对于特征编码器的条件熵来实现自适应。Yu 等人<sup>[61]</sup>则是从数据的角度来看问题，认为原始源标签可能存在噪声，因此将域自适应作为一个噪声标签学习问题，并利用具有伪中心的质子集的预测来修正源标签。Saltori 等人<sup>[62]</sup>设计了一种双分支对称网络架构，能够同时处理来自源域和目标域的点云，并通过跨域语义混合来减少域间分布差异。该方法基于教师-学生框架，在源域和目标域之间混合点云片段，然后利用源标签和目标伪标签进行监督，同时结合少量人工标注数据进一步提升性能。尽管在少量真实标注的支持下，半监督域自适应方法相较于无监督方法已经取得了显著的性能提升，但与目标域下全监督的结果仍存在一定差距。同时，半监督方法中所采用的

标注往往具有较大的随机性，无法确保每个标注点都能有效地促进模型性能提升，这在一定程度上暴露了标注效率的问题。

### 1.2.3.3 主动域适应

基于半监督和弱监督的方法在一定程度上提升了模型性能，但其标注点的选取往往是随机进行的，这使得数据利用效率和标注效果存在一定局限。为实现更高效的标注并充分挖掘数据的潜在价值，部分学者开始探索将主动学习与域适应技术相融合，通过主动学习策略在目标域中主动选取那些信息量大、价值高的点，而非被动地随机标注低价值数据，同时利用域适应技术最大化这些数据的效用。然而，传统主动学习策略通常依赖于模型对数据的筛选，而在域适应场景中，模型往往基于源域数据进行训练，由于存在域间隙，这种筛选过程可能导致所选标注点失效，最终引发无效标注的问题。早期的主动域适应研究提出通过利用对抗训练的领域判别器来衡量每个目标实例的不确定性和域特性。不同于之前的方法，CLUE<sup>[63]</sup> 则提出了一种熵加权聚类算法，用于查询目标域中不确定和多样化的样本。而 SDM<sup>[64]</sup> 被提出来优化边际损失函数，以探索与源领域中潜在硬样本相似的目标实例。然而，现有的主动域适应方法大多刻意设计手工制作的查询函数来评估样本的注释值，并对所有目标数据一视同仁地采用相同的适应策略。这种僵化的标准使它们很容易过度适应某些领域的适应情况，从而限制了主动域适应方法的泛化。一些具有创新性的主动域适应方法，类似 Huang 等人<sup>[65]</sup> 采取了分而自治的思想，把目标域上的数据分为 4 类，分别对这类数据进行不同的定制学习，但是这个方法是用在分类任务上的对于语义分割来说难以直接利用。MHPL<sup>[66]</sup> 发现满足邻近混沌区域、个体差异和来源不相似属性的样本是信息量最大的样本，并将其定义为最小快乐点，并设计了最小快乐点学习方法（Minimum Happy Point Learning, MHPL）以很好地探索和利用最小快乐点，有效提高模型性能。Annotator<sup>[43]</sup> 是首个将主动域适应方法应用于点云语义分割任务的方案。该方法将点划分至更大的体素块中，并采用专门设计的体素混淆度选择策略，在有限预算下实现了高效的样本选择与性能优化。然而，其主动学习策略并非专门为域适应场景设计，类似于传统主动学习方法，仍忽略了域差异可能导致所选样本在目标域上无法实现最佳提升的问题。

### 1.2.4 视觉语义分割中的 Mixng 方法

Mixng 方法在深度学习中，一直被作为一种数据增强的方式进行使用。其核心思想是通过混合不同的样本或特征来增强模型的泛化能力和鲁棒性，然而不同的任务和模态下的数据其 Mixing 的方式也有所不同。2018 年，Zhang 等人<sup>[67]</sup> 首次

提出 mixup 方法，用于超越经验风险最小化。这是第一篇关于 Mixing 的文章，其通过线性插值的方式将训练样本及其对应的标签进行凸组合，将组合后的数据用以训练模型，从而正则化神经网络，使其在训练样本之间表现出更简单的线性行为。该方法被用于分类任务上，并显著提高了模型的泛化能力，其在面对对抗样本和标签噪声时表现出很强的鲁棒性。然而，mixup 在图像分类任务中虽然表现优异，但在语义分割这种需要细粒度分类的任务中存在局限性。为了解决这一问题，CutMix 方法<sup>[68]</sup> 被提出。该方法通过切割和粘贴图像区域，结合标签的线性插值，不仅保留了 mixup 的正则化效果，还能够更好地处理局部特征，从而在弱监督目标定位和图像分类任务中取得了更好的性能。随着研究的深入，Mixing 方法逐渐被应用到三维点云数据处理领域。Mix3D 方法<sup>[69]</sup> 首次被提出，并用于 3D 场景的上下文数据增强。其通过混合两个增强场景，将对象实例隐式地放置在新的上下文中，从而帮助模型减少对场景上下文的依赖，更多地关注局部结构和特征。为了进一步提升模型在不同领域数据上的适应能力，CoSMix 方法<sup>[54]</sup> 被提出，用于 3D 点云分割的领域自适应。该方法通过组合语义信息和几何信息的样本混合策略，有效地平衡了全局上下文和局部几何结构，从而在不同领域数据上取得了更好的性能。随后，LaserMix 方法<sup>[70]</sup> 被提出，用于半监督的 LiDAR 语义分割任务。该方法通过混合来自不同 LiDAR 扫描的激光束，鼓励模型在混合前后做出一致且正确的预测。该方法在多个主流 LiDAR 分割数据集上展示了其有效性和优越性。尽管这些 Mixing 方法在解决特定问题上展现了卓越的性能提升，但目前它们主要与无监督、半监督乃至弱监督策略相结合，而在主动学习领域的应用仍未得到充分探索，而这正是本文关注和研究的重点之一。

### 1.3 论文研究主要内容

本研究聚焦于三维点云语义主动域适应问题，旨在探索和设计一种主动学习方法，以选择出对模型迁移最具价值的目标样本。同时，为了充分挖掘源域数据和主动标注样本的潜能，本文设计了一种高效的数据混合方法，将两者有机结合，从而进一步提升模型性能。主要研究内容如下：

- 1) 大量查阅相关领域文献，对基于深度学习的点云语义分割算法、视觉语义分割中的主动学习方法以及域自适应技术的最新进展及存在问题进行了系统梳理。通过总结上述研究成果，本文阐明了现有方法的优势与不足，并探讨了可改进的方向，为新方法的提出提供了坚实的理论基础。
- 2) 提出了基于点云语义分割域适应的主动学习方法。该方法由三个模块构成：源域原型构建模块、源域原型指导的数据选择模块和动态混合中间域构建模块。首

先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，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，该方法首次将主动学习方法与混合（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。

3) 提出了深度结合混合方法与主动学习的主动混合方法。该方法在基于点云语义分割域适应的主动学习框架上进行了深入探索，旨在解决由于数据量问题导致的域偏移累积以及主动学习选点不平衡问题。具体而言，该方法将动态混合中间域构建模块细分为两个模块：源-目标数量平衡模块和类别平衡主动混合模块。在每一轮主动学习过程中，通过动态匹配源域与目标域标注点的数量，确保混合数据中双域信息的均衡融合；同时，类别平衡主动混合模块保证了混合后数据的类别分布相对平衡，从而增强了主动学习与混合方法之间的协同作用。

## 1.4 论文组织结构

本文一共分为五个章节，各章节安排如下：

第1章为绪论，主要针对点云分割的研究背景、意义以及国内外的现状做详细阐述。首先介绍点云语义分割主动域适应方法的研究背景和研究意义。其次分别介绍了基于深度学习的点云语义分割方法、视觉语义分割相关的主动学习方法和域自适应方法以及视觉语义分割中的混合方法（Mixing）。最后简述本文的主要研究内容以及论文的主要组织架构。

第2章为点云语义分割主动域适应相关理论基础，对点云语义分割，主动学习及域自适应等本文研究涉及领域的基础知识做详细阐述。首先介绍了点云表示及点云语义分割模型基础，并介绍了点云语义分割中的常用评价指标；接着介绍了主动学习的相关基础概念和流程；随后又介绍了域自适应的基础概念和常用域对齐方法；最后对跨域点云语义分割任务中常用的主流公开数据集进行了介绍。

第3章提出基于点的主动半监督点云语义分割方法。首先介绍了该方法的研究动机和贡献，接着详细的阐述了提出的方法框架和基于原型指导的学习方法，包括动态源域原型构建模块、原型指导的数据选择模块和动态中间域构建模块，最后在合成到真实以及真实到真实的跨域场景下进行了大量实验验证以及可视化分析证明该方法的有效性。

第4章提出基于点云语义分割域适应的主动混合方法。同样首先介绍了该方

法的研究动机和研究贡献，接着详细介绍了方法中的主要模块，包括源-目标数量平衡方法和类别平衡主动混合方法。最后在合成到真实以及真实到真实的两个跨域场景、四个跨域数据集下进行了大量实验验证以及可视化分析证明该方法的有效性。

第5章为总结和展望。对全文研究内容进行了全面总结和深入分析，总结了本文在三维点云语义分割主动域自适应中的相关问题、研究内容和实验结果。并基于文章总结，对点云语义分割主动域适应任务未来的研究方向提出了展望。

## 第2章 点云语义分割主动域适应相关理论基础

### 2.1 本章引言

本章节对本文研究相关的各个领域理论基础进行介绍，包含点云语义分割基础知识、主动学习基础知识以及域适应相关知识。对于点云语义分割，主要介绍常用点云表示方式，点云语义分割任务的基本任务模型以及评价指标；对于主动学习，介绍主动学习的基本概念及其流程框架；对于域适应部分，主要对域适应概念及域对齐的常见方法进行了介绍。最后，对点云语义分割域适应相关任务中常用公开数据集进行了介绍。

### 2.2 点云语义分割相关知识

#### 2.2.1 三维点云理论基础

三维点云是三维空间中离散点的集合，每个点通过坐标  $(x,y,z)$  描述位置，部分还包含颜色、反射强度等附加信息。这些点通常由激光雷达（LiDAR）、深度相机等设备采集而来，能够高精度还原物体表面的几何特征。与传统的二维图像不同，点云直接记录三维空间信息，因此在机器人导航、自动驾驶、虚拟现实等领域有不可替代的作用。如图2-1所示，点云数据根据采集场景可分为室内与室外两类。室内点云多由深度相机或近距离激光雷达获取，特点是密度高、遮挡少，适合精细建模。例如扫地机器人通过室内点云构建房间地图，避开桌椅等障碍物。室外点云则依赖车载或机载激光雷达，场景尺度大，但点云稀疏且包含动态物体。自动驾驶汽车利用这类数据识别远处的车辆、行人，但需要处理树木遮挡或雨天噪声的干扰。尽管场景不同，两类点云均需应对数据量大、无序排列和非结构化的共性挑战。点云的优势在于其真实的三维几何表达能力。例如在建筑测绘中，点云能直接输出墙壁的倾斜角度或梁柱的尺寸，而无需从二维图像中推算。同时，点云对光照变化不敏感，在黑暗环境中仍可依靠几何信息工作。然而，点云数据也存在明显缺陷。单帧激光雷达点云可能包含数十万个点，存储与计算成本高昂；传感器噪声或物体遮挡会导致数据缺失，影响后续处理。因此，在数据处理层面，点云通常经历去噪、配准、特征提取等步骤。去噪用于消除传感器误差或环境干扰产生的离群点；配准将多视角点云对齐，形成完整场景模型；特征提取则从点云中识别平面、边缘等结构，为物体检测提供基础。随着技术进步，轻量化传感器与实时处理算法逐渐成为趋势。例如自动驾驶系统需在毫秒级时间内完成点云分割，区分道路、车辆与行人。点云与其他三维表示方法各有优劣，相比网格（Mesh）模型，点云保留

原始采集数据，但缺乏明确的表面连接关系；与体素（Voxel）相比，点云内存占用低且分辨率高，但难以直接应用卷积操作。因此，一般在大规模场景中，往往更多的是选择体素方式进行点云处理。

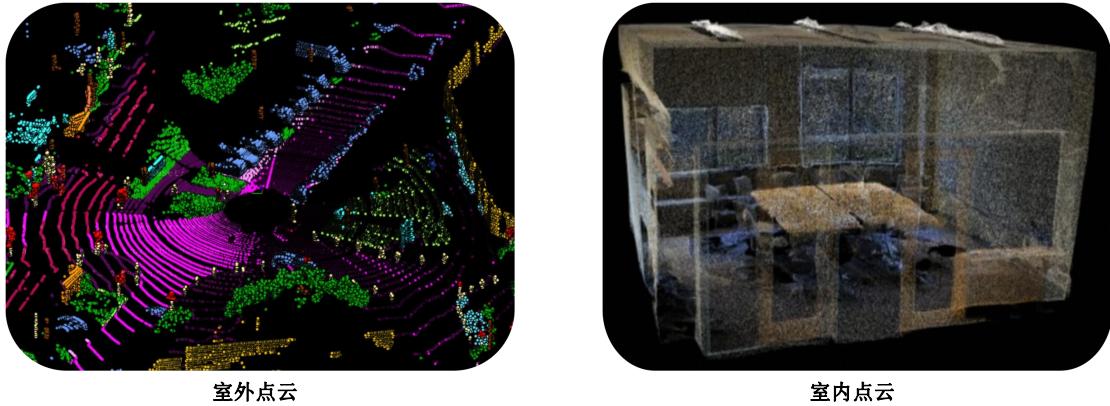


图 2-1 3D 点云

Fig. 2-1 3D point cloud

## 2.2.2 点云语义分割方法

### 2.2.2.1 目标与输入数据

点云语义分割的目标是为三维空间中的每个点赋予特定语义标签。在自动驾驶场景中区分道路、车辆、行人等类别，在室内建模中识别墙面、家具等物体。这一过程本质上是将无序的三维点云转化为带有语义信息的数据。输入的数据通常是由激光雷达（LiDAR）或深度相机采集的原始点云，其基础形式为包含  $N$  个点的坐标矩阵，形状为  $N \times 3$ ，对应每个点的三维坐标  $(x, y, z)$ 。在一些特殊的场景中，点云可能还包含颜色（RGB 值）、反射强度（LiDAR 返回信号强度）或时间戳（动态场景中的时序信息）等附加特征，此时数据维度扩展为  $N \times D$ ，其中  $D \geq 3$ 。数据预处理阶段通常会对坐标进行归一化，以消除传感器位姿的影响，并对离群点进行滤波以提高后续处理稳定性。

### 2.2.2.2 特征提取方法

特征提取是点云语义分割的核心，其目的是从原始点云中挖掘具有判别性的局部与全局特征。根据输入表示方式的不同，主流方法可分为三类：

1) 基于点的方法。此类方法直接处理原始点云，避免因体素化或投影导致的信息损失。典型代表如 PointNet++<sup>[23]</sup>，其核心思想是通过多层感知机（MLP）逐点提取特征，并通过最大池化聚合全局信息。具体而言，对于每个点  $\mathbf{p}_i$ ，模型不仅编码其自身坐标，还通过邻域查询（如  $k$  近邻或球查询）收集周围点集  $\{\mathbf{p}_j \mid j \in \mathbf{N}(i)\}$ ，

进而利用共享权重的 MLP 提取局部几何模式，如公式(2-1)所示：

$$\mathbf{F}_i = \text{MLP}(\mathbf{p}_i - \mathbf{p}_j, \|\mathbf{p}_i - \mathbf{p}_j\|_2) \quad \forall j \in \mathcal{N}(i) \quad (2-1)$$

其中  $\mathbf{p}_i - \mathbf{p}_j$  表示相对位置， $\|\cdot\|_2$  为欧氏距离。通过堆叠多个局部特征提取层，模型可逐步扩大感受野，捕获多尺度几何结构。

2) 基于体素的方法。基于体素的方法通过将点云转换为规则三维网格实现高效计算。MinkowskiNet<sup>[18]</sup> 是该领域的一个代表性方法，其核心是通过稀疏卷积网络，仅对非空体素进行计算，大幅减少内存与计算开销。稀疏卷积的表达式如公式(2-2)所示：

$$\mathbf{F}_{\text{out}}(x, y, z) = \sum_{(dx, dy, dz) \in \mathbf{K}} \mathbf{W}(dx, dy, dz) \cdot \mathbf{F}_{\text{in}}(x + dx, y + dy, z + dz) \quad (2-2)$$

其中  $\mathbf{K}$  为卷积核覆盖的偏移范围， $\mathbf{W}$  为卷积核权重。与传统三维卷积不同，稀疏卷积通过哈希表管理非空体素坐标，仅对有效位置执行计算。

3) 基于投影的方法。此类方法将三维点云投影至二维平面，复用成熟的图像处理网络。其中 RangeNet++<sup>[11]</sup> 将 LiDAR 点云转换为球面距离图像 (Range Image)，每个像素对应点的深度与方位角。投影后的二维图像通过改进的二维卷积网络提取特征，如公式(2-3)所示：

$$\mathbf{F}_{l+1}(u, v) = \text{ReLU} \left( \sum_{m, n} \mathbf{K}(m, n) \cdot \mathbf{F}_l(u + m, v + n) \right) \quad (2-3)$$

其中 ReLU 为激活函数。投影方法的计算效率较高，但可能因遮挡或投影畸变导致部分三维信息丢失。

### 2.2.2.3 分类与损失函数

在特征提取后，模型通过全连接层将  $N \times F$  维特征映射至  $N \times C$  维类别得分矩阵，其中  $C$  为类别总数。Softmax 函数将得分转换为概率分布，为每个点分配类别标签，如公式(2-4)所示：

$$P(y_i = c) = \frac{e^{\mathbf{z}_{i,c}}}{\sum_{c'=1}^C e^{\mathbf{z}_{i,c'}}} \quad (2-4)$$

其中  $\mathbf{z}_{i,c}$  表示第  $i$  个点在第  $c$  类上的得分。损失函数采用交叉熵损失，衡量预测概率与真实标签的差异，其数学表达式如公式(2-5)所示：

$$\mathbf{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log P(y_i = c) \quad (2-5)$$

式中  $y_{i,c}$  为 one-hot 编码的真实标签。

### 2.2.3 评价指标

在点云语义分割任务中，常用的评估指标有两个，一个是总体精度（Overall Accuracy, OA），另一个则是平均交并比（mean Intersection over Union, mIoU）。

总体精度通过计算正确预测点数占总点数的比例，直观反映模型的整体分类能力，其表示式如公式(2-6)所示：

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-6)$$

其中  $TP$  表示正确预测的正例， $TN$  为正确预测的反例， $FP$  为误判的正例， $FN$  为误判的反例。尽管 OA 计算简单，但在点云语义分割中，由于场景中不同类别点数差异显著，如城市道路点占比可能超过 50%，而交通灯可能不足 1%，OA 易被多数类别主导，这一缺陷使得 OA 难以真实评估模型性能。

平均交并比则通过衡量每个类别的预测区域与真实区域的重叠程度，提供更均衡的评估。对于类别  $c$ ，其交并比计算公式如(2-7)所示：

$$\text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}, \quad (2-7)$$

其中  $TP_c$  为类别  $c$  的正确预测点数， $FP_c$  为误判为  $c$  的点数， $FN_c$  为漏判的  $c$  类点数。mIoU 取所有类别 IoU 的平均值，其表达式如公式(2-8)所示：

$$\text{mIoU} = \frac{1}{C} \sum_{c=1}^C \text{IoU}_c \quad (2-8)$$

式中  $C$  为类别总数，mIoU 的核心优势在于其对每个类别的平等关注，即使某一类别的点数极少，其 IoU 值仍能直接影响整体得分，从而迫使模型兼顾所有类别。

当前研究普遍以 mIoU 为核心评价指标，主要因其能够克服类别不平衡带来的评估偏差。点云数据中，高频率类别与低频率类别的点数差异可能达到数百倍，若依赖 OA 这类整体指标，模型可能仅通过优化高频类别即可获得高评分，而忽视低频类别的学习，然而对于语义分割任务来说，实现对每一个类别点的精准分割是其目的所在，所以无论低频率类别还是高频率类别都是同等重要。mIoU 通过独立计算每个类别的重叠率，确保模型在各类别上的表现均被量化，从而更真实地反映其实际分割能力。此外，mIoU 对边界误差的敏感性也优于 OA，物体边缘点的误分分类会同时增加  $FP$  和  $FN$ ，导致 IoU 显著下降，而 OA 可能因整体正确率高而掩盖此类局部缺陷。

### 2.3 主动学习基础知识

主动学习作为传统机器学习中的一个分支，其核心思想是让模型在训练过程中主动选择对提升性能最有价值的数据进行标注，而非被动接受随机标注的数据。该方法的核心目的是在有限标注成本的约束下，通过主动的数据选择策略，最大化模型的性能。如图2-2所示，主动学习的流程可概括为迭代式的“选择-标注-训练”循环。首先，利用初始标注数据训练一个基础的目标模型。随后，模型对未标注数据进行推理预测，生成预测结果。基于这些预测，通过预设的选择策略，从未标注数据中筛选出对当前模型提升最有价值的候选数据。筛选出的候选数据由标注者即该领域相关专家进行标注，新标注数据与原有标注数据合并后，用于更新模型参数。这一过程持续迭代，直至标注预算耗尽或模型性能趋于稳定。流程中的核心模块包括选择策略设计以及模型更新机制，选择策略的优劣直接决定主动学习效率，常见策略包括不确定性采样、多样性采样和委员会查询。不确定性采样：选择模型预测置信度最低的样本；多样性采样：选择代表数据分布多样性的样本，覆盖不同特征空间区域；委员会查询：训练多个模型，选择各模型预测差异最大的样本。模型更新机制一般考虑如何利用这些已标注的高价值目标点来对模型进行微调以充分发挥数据与模型的潜力。

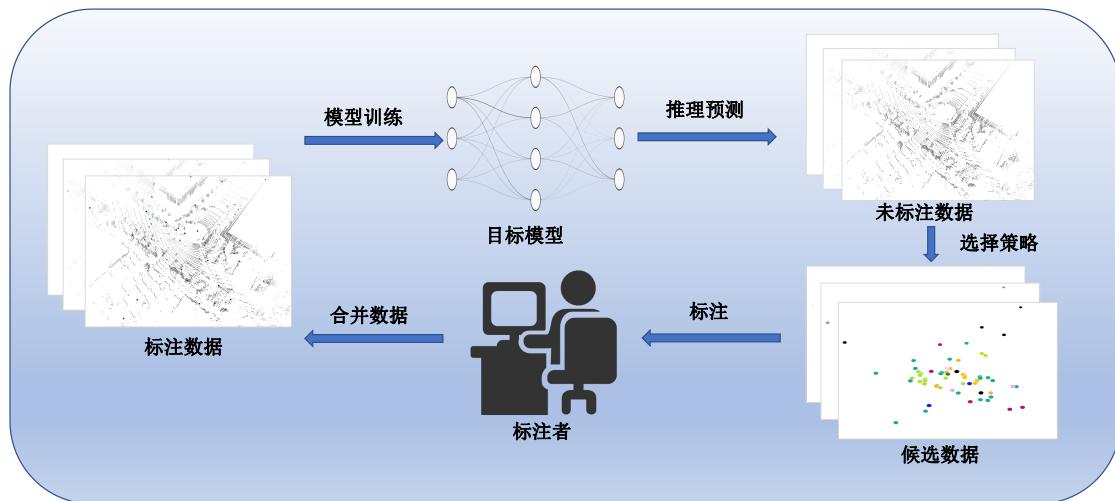


图 2-2 主动学习

Fig. 2-2 Active learning

### 2.4 域适应基础知识

域适应是机器学习中一种重要的技术，主要用于解决模型在相同任务不同数据分布场景下的泛化问题。如图2-3所示，当模型在一个数据集即源域上训练得非常好，能够正确的将数据分类，但在另一个数据集称为目标域上表现却不佳，无法对

目标域中的数据进行正确的分类，而造成这一现象的主要原因是不同数据域之间分布不同，存在域间差异，从而导致模型无法对新数据域中的数据进行正确分类，而域适应的目标就是帮助模型适应目标域的数据分布，从而提升其在新数据中的性能。域适应的核心思想是找到源域和目标域之间的共同特征，并通过调整模型或数据，缩小两者之间的分布差异。造成这些差异的原因有很多，比如点云的密度、物体形状的细微变化，或者传感器采集数据的方式不同。域适应的目标就是让模型学会忽略这些差异，而专注于识别共同的特征信息。

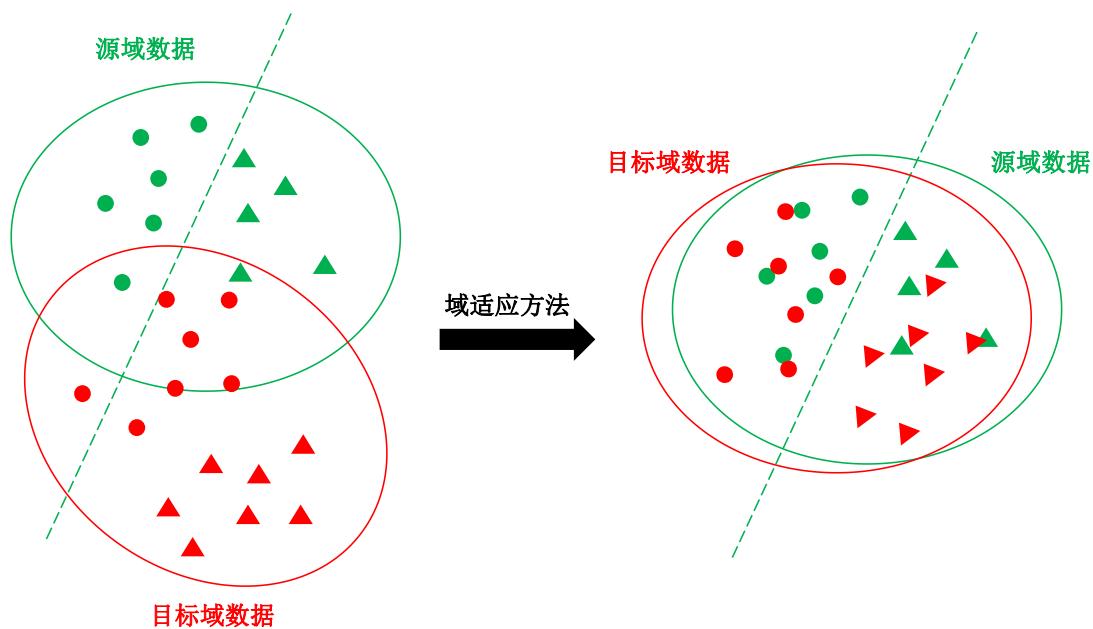


图 2-3 域适应

Fig. 2-3 Domain adaptation

## 2.4.1 域对齐的典型方法

### 2.4.1.1 基于差异度量的域对齐

最大均值差异（Maximum Mean Discrepancy, MMD）是一种衡量两个分布之间差异的常用方法，尤其在域适应与生成对抗网络等任务中应用广泛。它的核心思想是先将数据映射到一个称为可再生核希尔伯特空间（RKHS）的高维特征空间中，再比较两个分布在该特征空间中平均嵌入（mean embedding）之间的距离。如果两个分布越相似，它们在该空间中的平均嵌入就越接近；反之则越远。MMD 表达式如公式(2-9)所示：

$$\text{MMD}^2 = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(x_i^s) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(x_j^t) \right\|_{\mathbf{H}}^2, \quad (2-9)$$

其中  $\phi(\cdot)$  是将数据映射到高维空间的核函数,  $N_s$ ,  $N_t$  分别是源域和目标域的样本数。通过最小化 MMD, 模型特征提取器  $g_\theta$  被约束以生成域不变特征。

#### 2.4.1.2 对抗式域适应

受生成对抗网络 (GAN) 启发, 对抗训练通过域判别器  $D$  引导特征生成器  $g_\theta$  混淆域特征, 其目标函数为极小极大博弈, 如公式(2-10)所示:

$$\min_{g_\theta} \max_D \mathbb{E}_{x^s} [\log D(g_\theta(x^s))] + \mathbb{E}_{x^t} [\log (1 - D(g_\theta(x^t)))] , \quad (2-10)$$

其中  $D$  试图区分源域与目标域特征, 而  $g_\theta$  试图生成  $D$  无法区分的特征。而在经典模型对抗域适应网络 DANN<sup>[71]</sup> 中, 则是将对抗损失与任务损失结合, 如公式(2-11)所示:

$$\mathbf{L}_{\text{DANN}} = \mathbf{L}_{\text{task}} - \lambda \cdot \mathbf{L}_{\text{adv.}} \quad (2-11)$$

#### 2.4.1.3 自训练与伪标签

自训练利用模型对目标域的预测生成伪标签, 逐步迭代优化。定义伪标签为  $\hat{y}_j^t = \arg \max_c f_\theta(x_j^t)$ , 目标域损失表达是如公式(2-12)所示:

$$\mathbf{L}_{\text{self}} = \sum_{j=1}^{N_t} \mathbb{I}(\max f_\theta(x_j^t) > \tau) \cdot \mathbf{L}(f_\theta(x_j^t), \hat{y}_j^t) , \quad (2-12)$$

其中  $\tau$  为置信度阈值,  $\mathbb{I}(\cdot)$  为指示函数。此方法需谨慎设计阈值以避免噪声累积, 常与对抗训练结合使用。

## 2.5 跨域点云语义分割数据集

虽然目前已有许多开源的点云语义分割数据集, 但它们往往针对不同的应用场景设计, 具有各自的特点和局限性。在点云语义分割的域适应任务中, 通常会选择一些流行且广受认可的特定数据集作为实验标准, 以确保结果具有可比性和公平性。本研究同时涉及合成到真实 (Synthetic-to-Real) 和真实到真实 (Real-to-Real) 两个跨域场景。在合成到真实的跨域场景下, 通常使用合成数据 SynLiDAR<sup>[53]</sup> 作为源域数据, 将其迁移至 SemanticPOSS<sup>[8]</sup> 或 SemanticKITTI<sup>[6]</sup> 等真实数据集; 而在真实到真实的跨域场景中, 则一般选择 SemanticKITTI 与 nuScenes<sup>[7]</sup> 这两个主流的真实数据集进行互相迁移适应。这些数据集均主要反映室外场景的特点, 如数据密度、采集角度及环境复杂性等, 具有较高的代表性和挑战性。接下来, 本小节将对上述数据集的采集背景、数据规模以及在跨域任务中的应用情况进行详细介绍, 以便为后续实验提供坚实的数据基础。

### 2.5.1 SynLiDAR

SynLiDAR 是一个大规模的合成 LiDAR 点云数据集，专为促进从合成数据到真实数据的跨域语义分割研究而创建。该数据集包含超过 190 亿个逐点标注的点，涵盖 32 个语义类别，具有丰富的语义信息和高度的几何准确性。与现有的真实 LiDAR 数据集相比，SynLiDAR 的规模更大，质量更高，语义类别更丰富，且标注更加细致和全面。SynLiDAR 的点云数据是在多个由专业 3D 设计师构建的虚拟环境中生成的，这些虚拟环境包含大量与真实世界数据在几何形状和布局上相似的物体模型。如图2-4所示，数据集中的每个虚拟场景都经过精心设计，以确保合成数据的高质量和真实性。SynLiDAR 的点云数据具有逐点标注，标注类别包括汽车、自行车、摩托车、卡车、其他车辆、行人、骑自行车者、骑摩托车者、道路、停车场、人行道、其他地面、建筑物、围栏、植被、树干、地形、杆、交通标志等，涵盖了丰富的动静态对象，适用于多种语义分割任务。



图 2-4 SynLiDAR 数据集<sup>[53]</sup>

Fig. 2-4 SynLiDAR datasets<sup>[53]</sup>

### 2.5.2 SemanticKITTI

SemanticKITTI 是一个大规模的室外数据集，是基于真实道路场景构建的激光雷达点云数据集，可用于多种点云分割任务和语义场景补全，与 SemanticPOSS 相比，该数据集场景更广泛且庞大。其数据通过车载激光雷达采集，包含三维坐标与反射强度信息，但未包含颜色特征。数据集基于 KITTI 里程计基准扩展，通过逐点标注实现了 360 度全景点云的精细化语义标注，涵盖动态与静态对象 19 个类目标，同时包含地面与可行驶区域等特殊语义类别。如图2-5所示，数据集共由 22 个点云序列组成，每个序列都包含一段录制点云。其中，序列 00~10 和 08 总共提供了 23201 个有标注信息的点云，分别用于训练和验证；11~21 提供了 20351 个没有标注信息的点云，用于测试。点云语义分割域适应中，将 SynLiDAR 数据集上训练

的模型迁移到 SemanticKITTI 数据集上，并将类别映射为 19 个共有的语义类别，包括汽车、自行车、摩托车、卡车、其他车辆、行人、骑自行车者、骑摩托车者、道路、停车场、人行道、其他地面、建筑物、栅栏、植被、树干、地形、杆状物、交通标志。

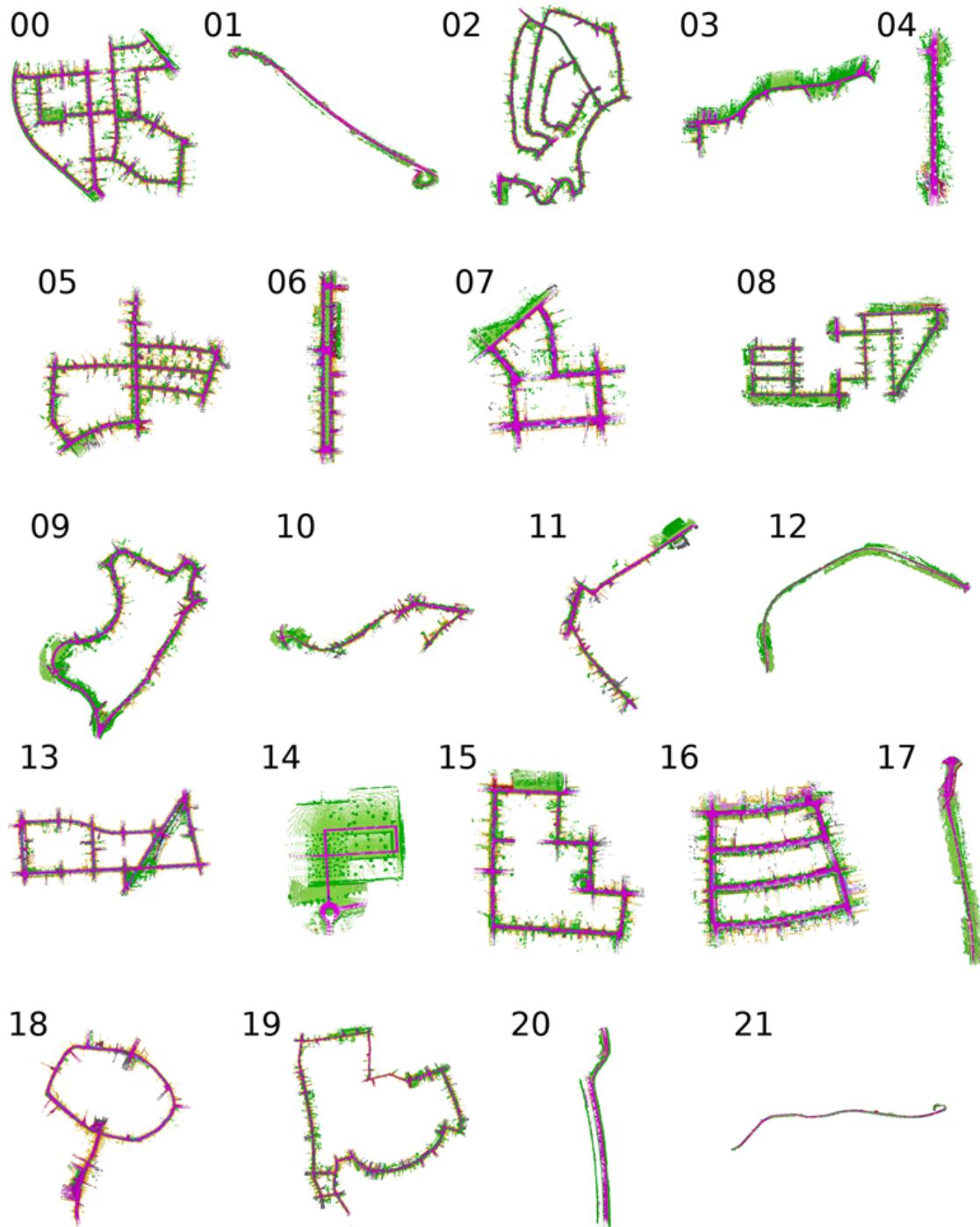


图 2-5 SemanticKITTI 数据集<sup>[6]</sup>  
Fig. 2-5 SemanticKITTI datasets<sup>[6]</sup>

### 2.5.3 SemanticPOSS

SemanticPOSS 是一个真实世界的小规模数据集，使用 Pandora 40 线激光雷达传感器在北京大学采集，包含 2988 个扫描样本。每个扫描样本包含约 80,000 到 100,000 个点，捕捉了典型的都市户外场景。该数据集包含 13 个语义类别，为研究者提供了丰富的数据资源。在数据集划分方面，遵循先验研究的方法，将包含 500 个点云样本的 03 序列分配用于验证，而剩余的 2488 个点云样本则用于训练。在点云语义分割域适应中，将 SynLiDAR 数据集上训练的模型迁移到 SemanticPOSS 数据集上，并将类别映射为 13 个共有的语义类别，包括汽车、自行车、人、骑行者、地面、建筑物、栅栏、植物、树干、杆、交通标志、垃圾桶、路锥/石块。其数据集展示图如图2-6所示。

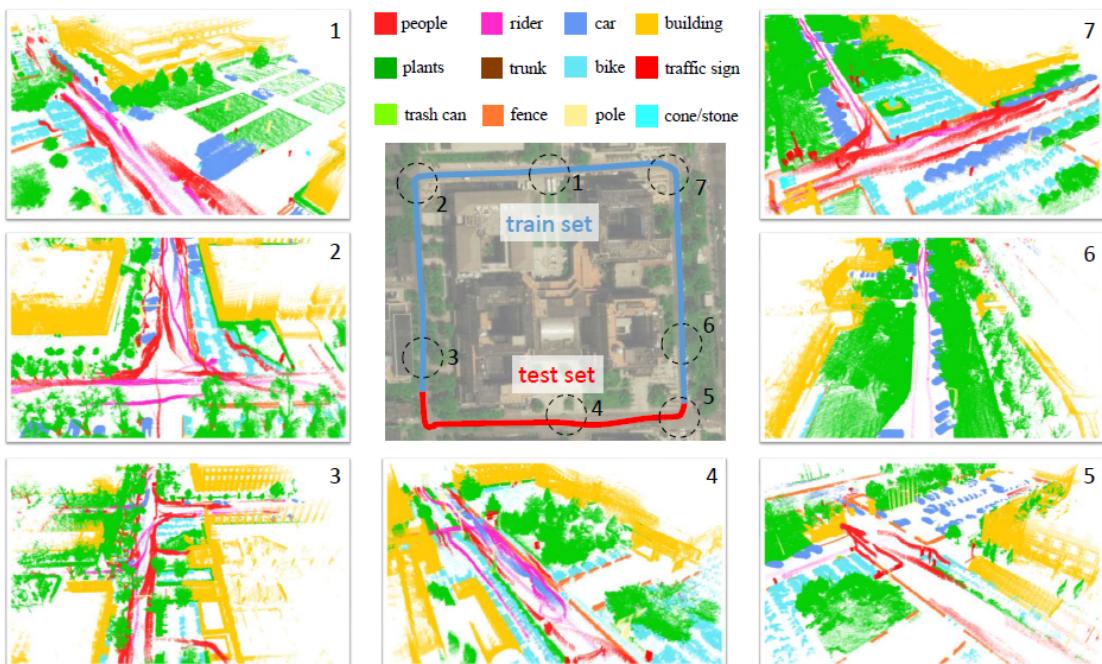


图 2-6 SemanticPOSS 数据集<sup>[8]</sup>

Fig. 2-6 SemanticPOSS datasets<sup>[8]</sup>

### 2.5.4 nuScenes

nuScenes 是多模态自动驾驶数据集的代表性资源，整合了激光雷达、摄像头、毫米波雷达及定位系统的同步采集数据，覆盖波士顿与新加坡的 1,000 个复杂驾驶场景。其激光雷达语义分割子集（nuscenes-lidarseg）包含约 14 亿个点云，总计 40,000 帧数据，划分为 850 个训练验证场景与 150 个测试场景。语义标注涵盖 16 类关键目标，包括自行车、公交车、轿车、工程车、摩托车、行人、卡车、拖车等交通参与者，可行驶区域、其他人造地面、人行道、自然地面等道路要素，以及障

碍物、锥形路标、植被和人造物体等环境元素。类别设计注重实际驾驶场景的多样性，尤其强调对临时障碍物的精细化标注，例如锥形路标与非铺装路面，为复杂环境下的多目标感知与路径规划提供高精度数据支持。其数据集展示图如图2-7所示。在点云语义分割域适应中，SemanticKITTI与nuScenes数据进行模型的相互迁移，最终将类别映射为7个共有的语义类别，包括车辆、行人、道路、人行道、地形、人造物体、植被。

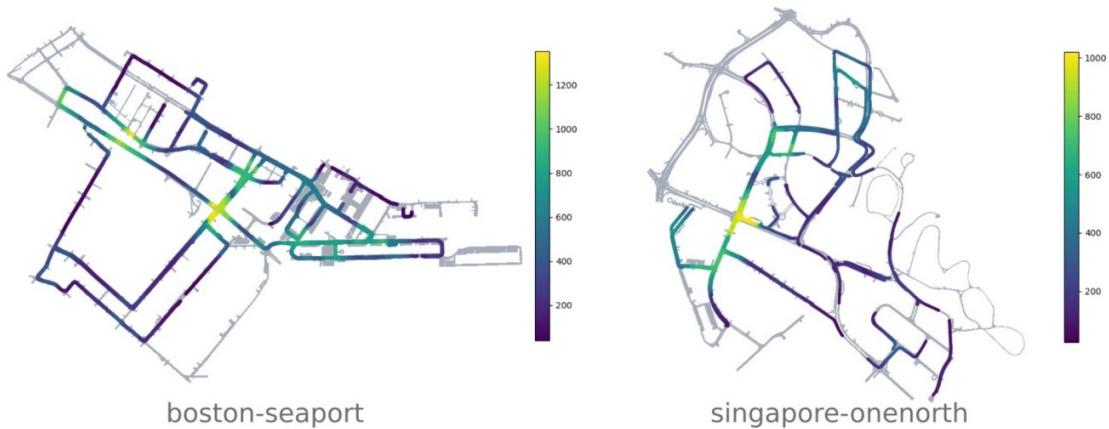


图 2-7 nuScenes 数据集<sup>[7]</sup>

Fig. 2-7 nuScenes datasets<sup>[7]</sup>

## 2.6 本章小结

本章主要对点云语义分割主动域适应相关领域的基础知识做了全面介绍，包含对点云语义分割，主动学习，域适应在内的多个不同领域分别做了介绍。对于点云语义分割的相关基础知识，首先详细介绍了点云的获取方式、点云的特征以及室内外点云的差异。根据特征提取方式的不同，进一步对点云语义分割任务的基本模型和基本实现方式做了基础介绍，最后还介绍了语义分割任务中常用的评价指标 mIoU。对于主动学习相关基础知识，主要对主动学习方法的基本概念和基本流程做了详细阐述；针对域适应基础知识，首先介绍了域适应的概念、目的以及任务，接着又对常见的域对齐方法进行了梳理和解释。最后还对点云语义分割域适应任务中常用的室外跨域数据集分别进行了介绍。本章介绍的各方面基础知识对后续章节中算法设计的阅读理解起到重要铺垫作用。

## 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，并结合混合（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，接着对每个方法子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：昂贵的标注成本与数据间分布差异。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<sup>[6]</sup>。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<sup>[72]</sup>。

为缓解上述问题，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<sup>[73,74]</sup>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<sup>[65,75]</sup>。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<sup>[63,76,77]</sup>，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，然而这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，直接套用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<sup>[43]</sup>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) 提出了一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性的目标点。
- 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 基于点云语义分割域适应的主动学习方法

#### 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  $\mathbf{S} = \{(\mathbf{X}_i^s, \mathbf{Y}_i^s)\}_{i=1}^{N^s}$ ，和一个目标域数据  $\mathbf{T} = \{(\mathbf{X}_i^t)\}_{i=1}^{N^t}$ ，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  $\mathbf{G}$  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  $\mathbf{T}$ ，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  $\mathbf{G}$  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  $\mathbf{T}_l$  进行标注，并利用其参与分割模型  $\mathbf{G}$  的调优，同时更新未标注目标数据集  $\mathbf{T} = \mathbf{T} - \mathbf{T}_l$ 。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源原型引导的数据选择：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，构建出兼具双域信息的中间域数据，将此数据用于分割模型的微调，使得模型可以学习到更加稳定可靠的域不变特征，从而进一步缩减域间隙。

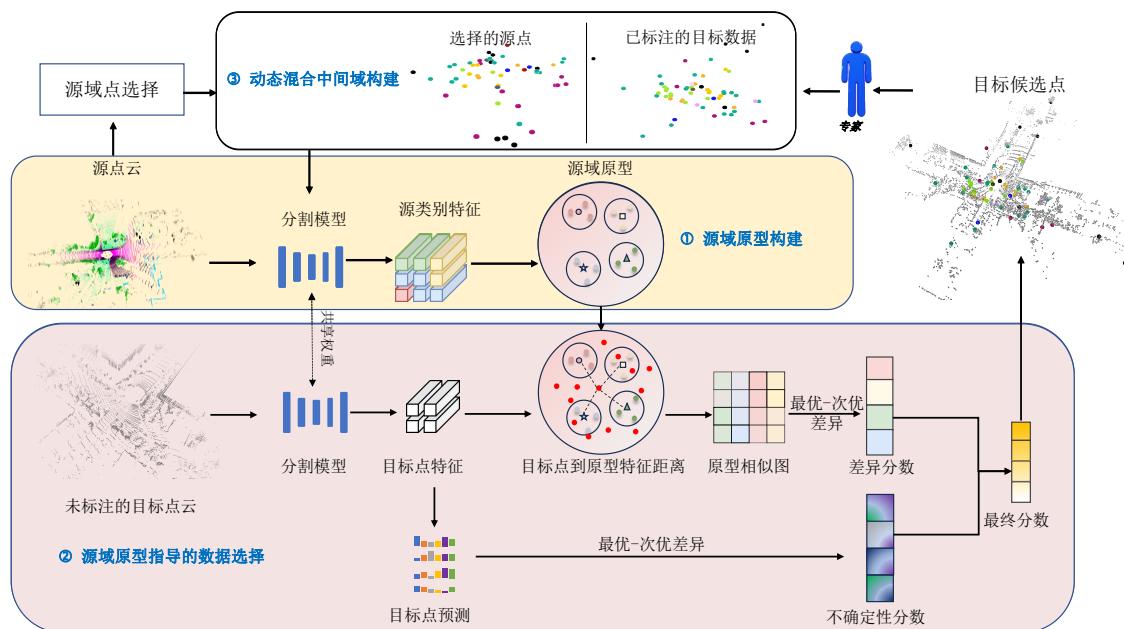


图 3-1 基于点云语义分割域适应的主动学习方法框架

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  $\mathbf{G}$ ，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  $\mathbf{S}$  输入当前网络  $\mathbf{G}$ , 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<sup>[78]</sup> 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个类别的质心, 蕴含该类别的语义特征。由于分割模型在不断微调, 这些源域原型将在每一轮的主动学习阶段动态更新, 根据新的模型动态的构建新的源域原型, 并为后续的跨域数据选择提供指导。

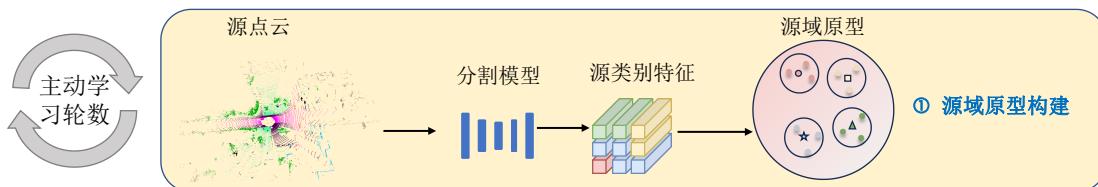


图 3-2 源域原型构建

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，其表达式公式如(3-4)所示：

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

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距离，其值越小代表未标注点在特征空间中距离源域类别  $c$  的聚类中心越近；越大则表明未标注点在特征空间中距离源域类别  $c$  的聚类中心越远。因此， $d_i^c$  将作为一个重要的域差异计算依据，并用于后续的域差异评分的一个重要衡量指标

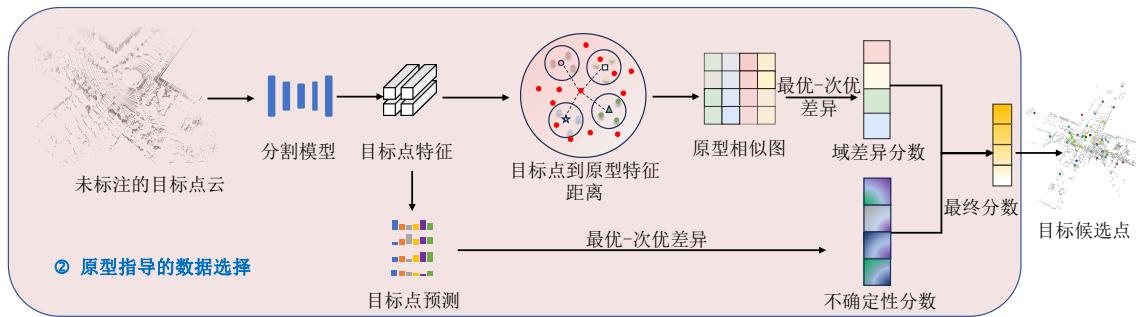


图 3-3 源域原型指导的数据选择

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  $p_i$  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，确保不同类别的相似值在同一尺度下比较。

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

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

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：

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

式中， $S_{R1}$  和  $S_{R2}$  分别代表最大和次大类概率值。该评分反映模型对目标点不确定性：当  $M_{us}^i$  较小时，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  $\alpha$  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  $\mathbf{T}_l$  提交至专家（Oracle）进行人工标注，同时更新未标注目标数据集  $\mathbf{T} = \mathbf{T} - \mathbf{T}_l$ ，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  $R$ 、标注总预算  $B$  以及目标点云的帧数  $N^T$  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更加可靠的域不变特征，进一步缩减域间隙<sup>[59, 79]</sup>。

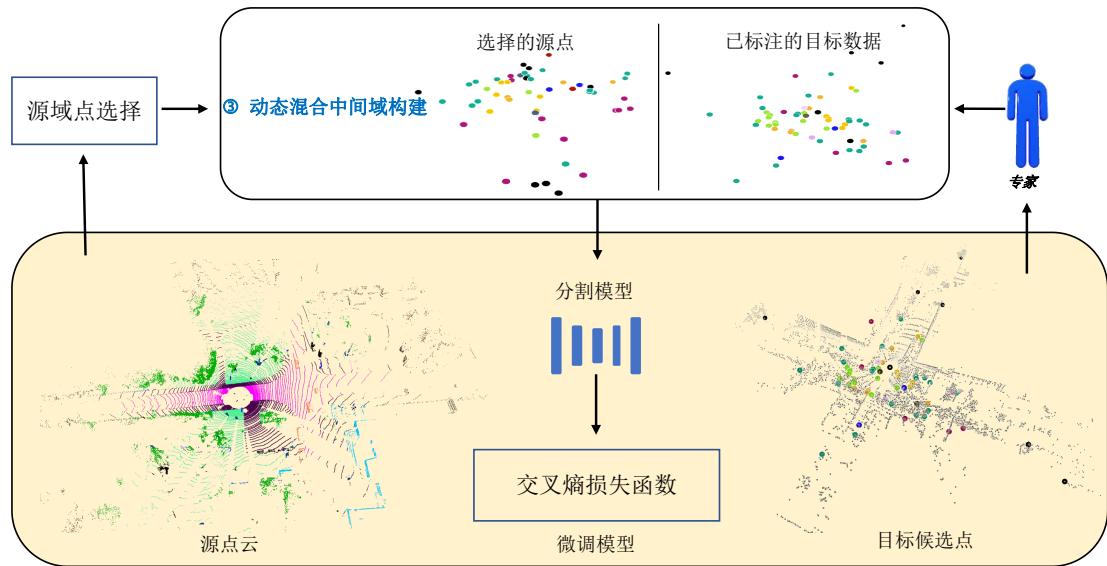


图 3-4 动态混合中间域构建

Fig. 3-4 Dynamic mixed intermediate domain construction

## 3.4 实验评估

### 3.4.1 实验设置

在实验中，使用 MinkowskiNet<sup>[18]</sup> 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)<sup>[80]</sup> 作为学习优化器，将动量设置为 0.9，在第一轮的训练中使用线性预热策略，使学习率线性增加到基础学习率 0.01，同时使用权重衰减系数为 0.0001 的余弦衰减调度器动态调整学习率，进一步稳定模型收敛，所有实验都是在单张 NVIDIA RTX A6000 GPU 上进行的训练。本章方法在真实到真实及合成到真实的跨域场景下均进行了大量充分的实验，并通过详细的可视化分析验证其有效性。对于所有场景，主动学习过程共执行 5 次迭代，以达到预设的标注预算。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  $\alpha$  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，SemanticKITTI→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  $\alpha$  为 0.6。其中 Mixing 方法的随机混合比例为 30，即混合后的中间域数据，源域点数是目标域点数的 30 倍。

### 3.4.2 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。

#### 3.4.2.1 合成到真实场景

为了能够与现有的点云语义分割域适应方法进行公平公正的比较，在合成到真实跨域场景的实验中，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，并通过实验结果分析及可视化证明了本章方法的泛化性和有效性。

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 <sup>[76]</sup>	UDA	-	23.0
AdvEnt <sup>[81]</sup>	UDA	-	25.8
CRST <sup>[82]</sup>	UDA	-	26.5
ST-PCT <sup>[53]</sup>	UDA	-	28.9
PolarMix <sup>[83]</sup>	UDA	-	32.2
CoSMix <sup>[54]</sup>	UDA	-	31.0
DGT-ST <sup>[72]</sup>	UDA	-	43.1
Annotator <sup>[43]</sup>	ADA	0.1%	57.7
<b>本章方法</b>	<b>ADA</b>	<b>0.1%</b>	<b>58.7</b>

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。本数据集的分割可视化如图3-5所示，使用红框进行突出对比显示，可以看到本章方法相比其他主动学习方法在语义类别交界处以及一些细节处分割表现得更好。

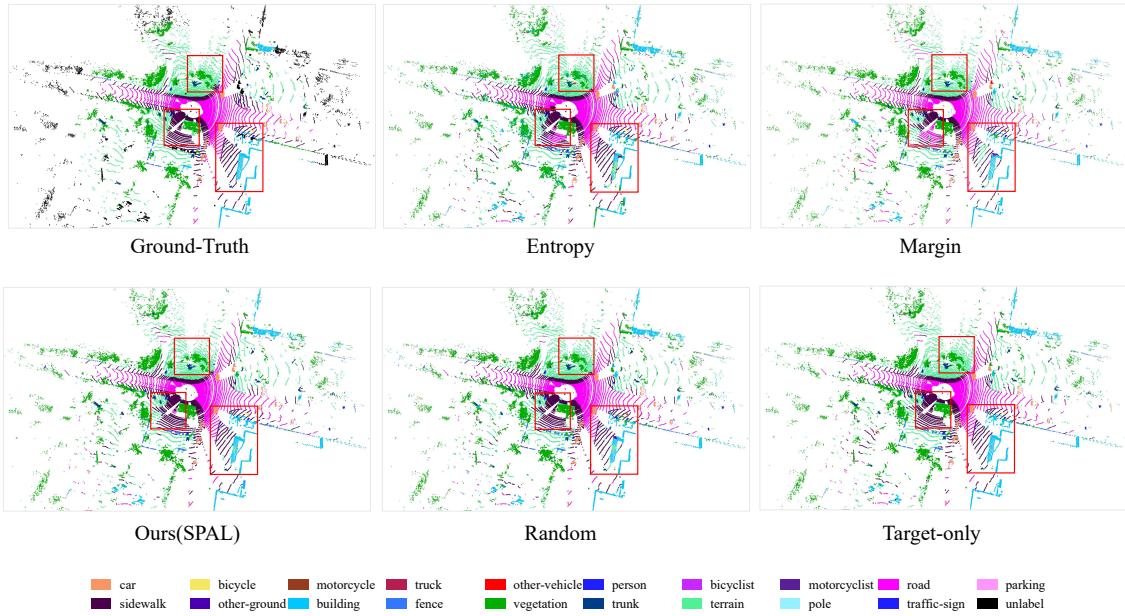


图 3-5 本章 SynLiDAR→SemanticKITTI 分割可视化图

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 <sup>[82]</sup>	UDA	-	27.1
ST-PCT <sup>[53]</sup>	UDA	-	29.6
PolarMix <sup>[83]</sup>	UDA	-	30.4
CoSMix <sup>[54]</sup>	UDA	-	40.4
DGT-ST <sup>[72]</sup>	UDA	-	50.8
Annotator <sup>[43]</sup>	ADA	0.1%	52.0
<b>本章方法</b>	ADA	0.1%	<b>56.6</b>

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，相比全监督（Target-Only）结果仅差 1.4 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。分割可视化结果如图3-6所示，在红色方框所突出显示的区域中可以观察到，本章方法在一些易混淆的语义类别比如植物和建筑中的分割效果依然优于其他主动学习方法。

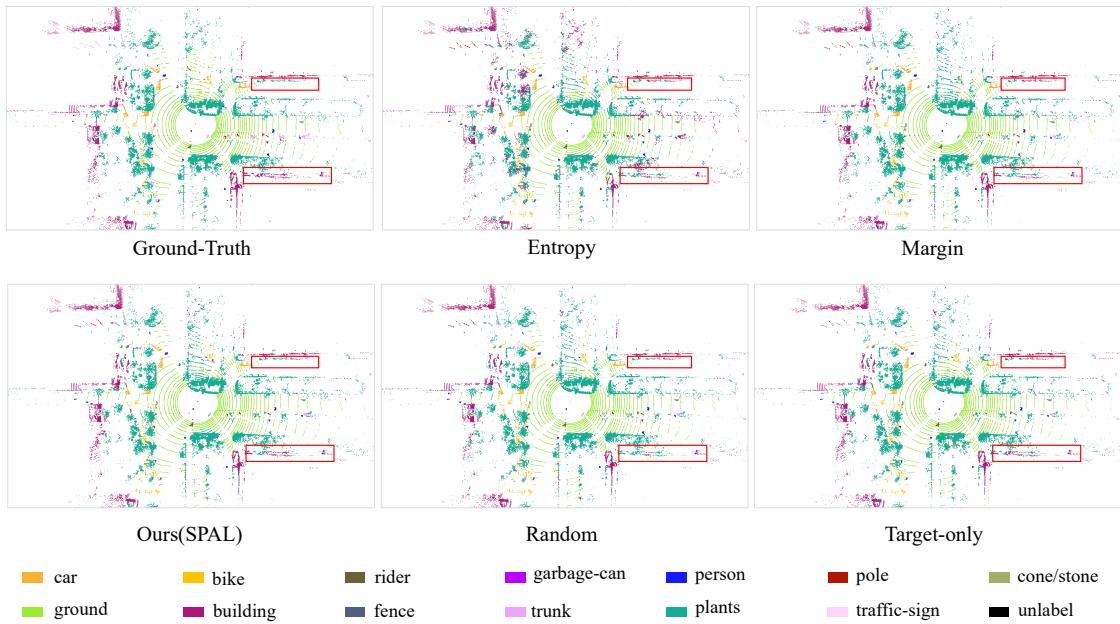


图 3-6 本章 SynLiDAR→SemanticPOSS 分割可视化图

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

### 3.4.2.2 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，因此其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，虽然不用进行标注但其性能仍然离全监督非常远。在主动域适应方法中（ADA），本章方法以 0.1% 标注量达到目标域全监督的 92.5%，较 Annotator 提升 0.9 个百分点，验证了本章方法的有效性。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，本章方法在一些易错类别如人行道上的分割效果依然优于其他主动学习方法，再一次验证了该方法的有效性。

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 <sup>[69]</sup>	UDA	-	31.5
CoSMix <sup>[54]</sup>	UDA	-	29.8
SN <sup>[84]</sup>	UDA	-	25.8
RayCast <sup>[52]</sup>	UDA	-	30.9
LiDOG	UDA	-	34.9
Annotator <sup>[43]</sup>	ADA	0.1%	75.9
本章方法	ADA	0.1%	<b>76.5</b>

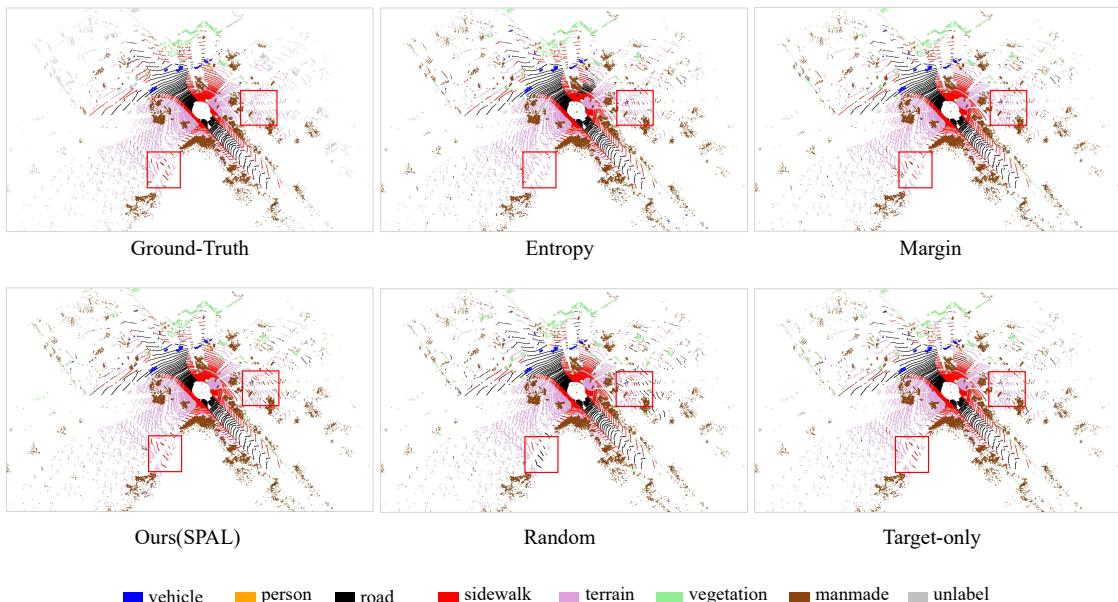


图 3-7 本章 nuScenes→SemanticKITTI 分割可视化图

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模型性能差距进一步扩大至 52.9 个百分点，这说明跨域任务的难度也进一步增加。在本

实验结果中，UDA 最佳方法达到了 41.2% 的性能，但是仍距离全监督有 44.2 个百分点的性能差距。值得注意的是，本章方法在 0.1% 标注下取得 83.0% 的结果，达到全监督性能的 97%，相较同标注量下的 Annotator 提升 1.2 个百分点。本实验也提供了分割可视化结果如图3-8，分割证明了本章方法在此数据集上的有效性。通过实验结果可知，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。

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 <sup>[69]</sup>	UDA	-	32.4
CoSMix <sup>[54]</sup>	UDA	-	36.8
SN <sup>[84]</sup>	UDA	-	23.6
RayCast <sup>[52]</sup>	UDA	-	31.5
LiDOG <sup>[85]</sup>	UDA	-	41.2
Annotator <sup>[43]</sup>	ADA	0.1%	81.8
<b>本章方法</b>	ADA	0.1%	<b>83.0</b>

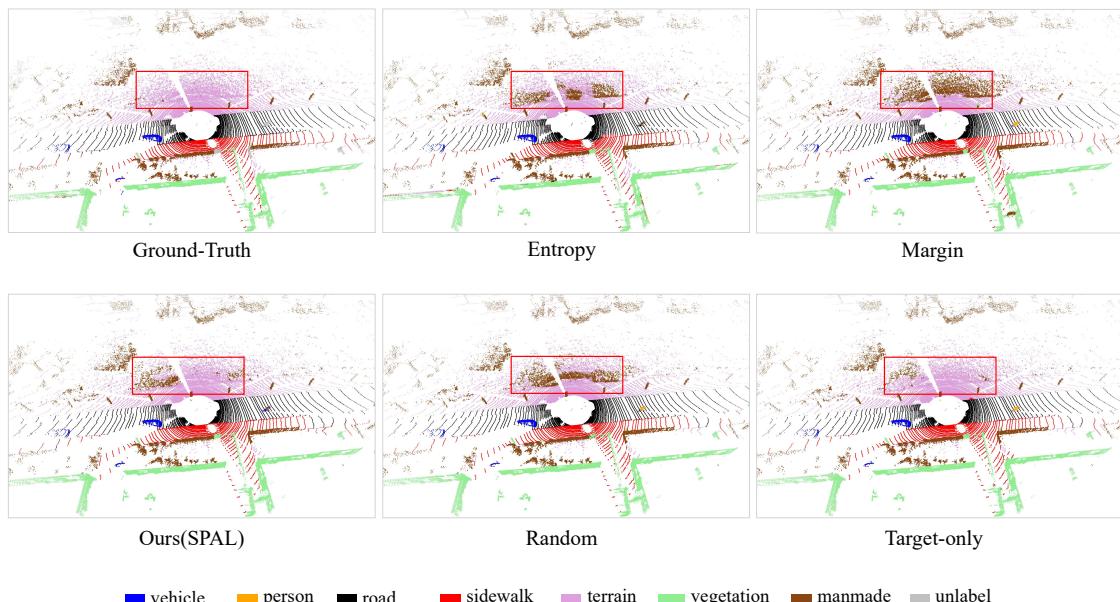


图 3-8 本章 nuScenes→SemanticKITTI 分割可视化图

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

### 3.4.3 消融对比实验

#### 3.4.3.1 与其他主动学习方法对比

在未引入 Mixing 方法时，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 Random 0.2 个百分点 Entropy 2.8 个百分点。而在 SynLiDAR→SemanticPOSS 任务中，SPAL 低于 Entropy 0.7 个百分点，却也超过了 Random 3.2 个百分点，Margin 0.2 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 <sup>[73]</sup>	0.1%	51.7
	<b>Margin</b> <sup>[74]</sup>	0.1%	<b>54.4</b>
	Ours(SPAL)	0.1%	53.5
SynLiDAR→SemanticPOSS	Random	0.1%	47.5
	Entropy <sup>[73]</sup>	0.1%	<b>51.4</b>
	Margin <sup>[74]</sup>	0.1%	50.5
	Ours(SPAL)	0.1%	50.7

而在结合 Mixing 方法后如表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 <sup>[73]</sup>	0.1%	55.1
	Margin <sup>[74]</sup>	0.1%	57.6
	<b>Ours(SPAL)</b>	0.1%	<b>58.7</b>

### 3.4.3.2 消融实验

为证明方法的有效性，本章在 SynLiDAR→SemanticKITTI 数据上进行了消融实验，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（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 SPAL 时，模型性能提升至 50.7%，较基线增加 27.9 个百分点，证明了原型指导的数据选择主动学习方法对目标域关键样本筛选的优势。而当仅使用 Mixing，模型性能为 51.2%，较无任何模块的基线提升 28.3 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% 的结果达到最优，较单一模块性能分别提升 8.0 和 7.5 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
	✓	51.2
✓	✓	<b>58.7</b>

## 3.5 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 方法结合并用于点云语义分割域适应任务中，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 基于点云语义分割域适应的主动混合方法

### 4.1 本章引言

本章主要介绍基于点云语义分割域适应的主动混合方法。该方法通过将基于原型指导的主动学习方法和本节所提出的主动混合方法进行深度结合，进一步提升了主动域适应方法的结果。接下来，下文将分别从方法提出的研究动机及主要贡献、方法的具体组成和实现细节，以及实验分析评估等方面，对所提出的方法进行全面阐述。

### 4.2 研究动机及贡献

科技的进步使得获取点云雷达数据变得更加容易，而深度学习技术的发展则推动了点云语义分割任务的迅速提升，涌现出一系列优秀模型并取得了显著成果。然而，这些模型大多基于全监督模式，需要对点云样本进行逐点标注，这是一项耗费大量人力物力的艰巨任务。为缓解这一问题，研究者们提出了多种方法，而域适应就是其中一种有效的策略。在无监督和半监督域适应中，一些方法巧妙地使用混合方法（Mixing）构建中间域，以学习域不变特征，进而解决域偏差问题。Polarmix<sup>[83]</sup> 方法在  $x, y$  组成的水平面上，按照指定的角度分割并混合不同扫描区域的点云数据，在域适应中混合的则是不同数据集下的同角度点云区域。LaserMix<sup>[70]</sup> 方法则是沿方位角，裁剪并混合不同点云中相同角度的雷达线束，在域适应中混合的则是不同数据集下相同线束点云。此外，主动域适应方法<sup>[63, 64, 66]</sup> 通过主动选择对减小域偏差最有帮助的目标点进行标注，从而提升模型性能。

尽管上述算法各自以不同方式缓解了标注问题，提升了分割模型的性能，但在域适应中，主动学习与 Mixing 方法的结合尚未得到深入研究。无论是在图像还是三维点云领域，混合方法和主动学习通常独立应用，分别为减少标注需求做出贡献。然而，现有的混合方法若直接应用于主动域适应，可能存在以下问题：1) 数量不均衡导致的域偏差问题。现有的混合方法多基于同一数据集或大量伪标签，特点是场景连续且点数丰富。然而在域适应中，源域和目标域存在域偏移，主动学习预算下选择的点数量有限。若直接应用这些混合方法，可能导致模型过度学习源域信息，阻碍对目标域信息的提取，最终导致次优结果。因此，在主动域适应中，保持源域和目标域信息的平衡可能比场景连续性更为重要。2) 主动学习选点导致的语义类别不平衡问题。大多数主动学习选点策略基于不确定性，在实际标注前，无法确定所选子集的类别分布是否均衡。尽管一些方法通过伪标签预判类别，但

这基于模型不可靠的预测，在域适应任务中可能不适用。此外，现有的混合方法未考虑逐点的语义类别平衡问题。因此，如何使主动学习选择的目标子集与源数据混合后的训练子集在类别分布上尽可能平衡，仍是一个普遍存在的问题。针对上述问题，本章进一步探讨了主动学习方法与混合方法在域适应任务中的深度结合。在该方法中，延续前章使用基于原型指导的主动学习与 Mixing 相结合的基础框架。在此基础上，提出了面向点云语义分割域适应的主动混合方法。其包括两个模块：源-目标数量平衡模块旨在解决数量不均衡导致的域偏差问题；类别平衡主动混合模块用于解决主动学习选点导致的语义类别不平衡问题。这两个模块有效地实现了主动学习与 Mixing 方法的深度结合，进一步提升了模型的分割性能。

本章研究的主要贡献如下：

- 1) 提出了源-目标数量平衡模块，通过选择与目标域样本中标注数量相同的源点进行混合构建中间域，解决了因数量不平衡导致的域偏差积累问题，进一步提高了混合中间域的有效性。
- 2) 提出了类别平衡主动混合模块，利用在源-目标数量平衡算法得到多个候选源点混合子集，计算并选择与标注的目标域混合后类别分布最为均衡的子集进行 Mixing，借助源域数据实现主动选点的类别平衡，进一步提升了主动学习的有效性。
- 3) 根据提出的两个算法模块实现了主动学习与 Mixing 的深入结合，得到了深度融合主动混合的点云语义分割域适应框架，分割表现领先于所有现存的点云语义分割域适应方法。

## 4.3 基于点云语义分割域适应的主动混合方法

### 4.3.1 方法概述

在第三章中，设计了一个基于原型的主动学习方法，并初步尝试结合混合方法构建强健的中间域数据，以缩减域偏差，提升模型性能。然而，混合方法与主动学习的结合仍有广阔的探索空间。如何根据域偏差以及主动学习的特点来有效结合混合方法，仍是一个值得深入研究的问题。

为了深入探讨 Mixing 方法与主动学习在域适应任务中的深度融合，本章在第三章提出的总体框架基础上进行了改进，改进后的框架如图4-1所示。该算法的基本流程主要由三个主要模块构成：①源原型引导的数据选择模块（SPAL）；②源-目标数量平衡模块（STNB）；③类别平衡主动混合模块（CBAM）。为进一步实现两者的高效协同作用，从而获得更优的分割表现，本章对第三章的动态中间域构建模块进行了改进，分为两个模块：一是源-目标数量平衡模块，选择类别数量大于预

设定值的源域帧，并从中筛选与所匹配的目标域帧标注点数平衡的候选子集；二是类别平衡主动混合模块，计算每个候选子集与目标域帧标注点混合后的类别分布熵值，选择熵值最大的候选源域子集，以构建类别相对平衡的中间域数据。通过上述改进，实现混合方法与主动学习的深度融合，进一步提升点云语义分割模型的性能。

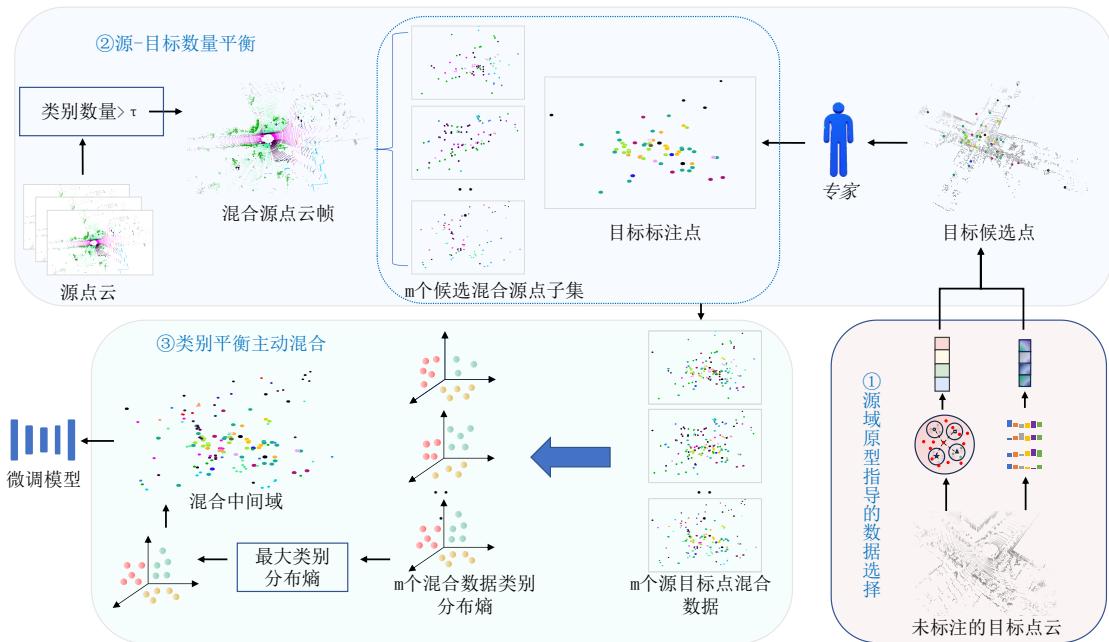


图 4-1 基于点云语义分割域适应的主动混合方法框架

Fig. 4-1 Framework of active mixing for domain adaptation in point cloud semantic segmentation

### 4.3.2 源-目标数量平衡算法

前一章节的实验结果证明了主动学习与 Mixing 方法的结合取得了显著的效果。然而这只是对于两种方法结合的初步探索，对于他们的深度结合仍然有很大的探索空间。在主动域适应中，主动学习标注点的数量极少，又加之域间隙的存在，使得之前的连续场景、或者基于伪标签的 Mixing 方法与主动学习结合并不能发挥其最大的效果。如图4-2所示，本章的源-目标平衡算法旨在数量层面平衡混合的源域点和标注的目标点，对主动学习与 Mixing 的深入结合进行探索。

在每一轮主动学习迭代中，将最新标注的目标点加入到已标注数据集中，更新数据集  $\mathbf{T}_r^{al}$ ，更新公式如(4-1)所示：

$$\mathbf{T}_r^{al} = \mathbf{T}_{r-1}^{al} \cup \mathbf{T}_r^{al} \quad (4-1)$$

其中， $\mathbf{T}_r^{al}$  是当前主动学第  $r$  轮的主动标注最新数据集，而  $\mathbf{T}_{r-1}^{al}$  是上一轮即第  $r-1$  轮主动标注的数据集，其中  $r$  从 1 开始计算，并且当  $r=0$  时， $\mathbf{T}_0^{al}$  为空集  $\emptyset$ 。在主

动学习阶段结束后，接着就是将已标注的最新的目标域数据与源数据混合。选择一帧带有标注的目标域数据，然后再随机匹配一帧源域数据，为了保证数据的质量，源域数据的筛选是有条件的，即仅当某一帧点云的类别数量大于预设的阈值 $\tau$ 时，该帧数据才有资格与当前目标域数据进行混合，筛选过程如(4-2)所示，其中 $S_i = \{\mathbf{X}_i^S, \mathbf{Y}_i^S\}$ ，表示源域中的一帧点云数据。

$$\mathbf{M}_{mix} = \{\mathbf{S}_i | unique(\mathbf{Y}_i^S) > \tau\}, \quad \mathbf{Y}_i^S \in \mathbf{S}_i \quad (4-2)$$

接着，从 $\mathbf{M}_{mix}$ 中随机筛选一个源域帧 $\mathbf{M}_i \in \mathbf{M}_{mix}$ 与目标域帧 $\mathbf{p}_i$ 进行匹配，在匹配成功后，将从匹配的源域帧中，候选 $m$ 个混合子集，这些候选子集中的点数与匹配的目标域中主动标注的点数量相同，候选点的公式如(4-3)所示：

$$\mathbf{Q} = \{\mathbf{q}_1, \dots, \mathbf{q}_m\}, \quad \|\mathbf{q}_i\| = \|\mathbf{p}_i\|, \quad \mathbf{q}_i \subset \mathbf{Q}, \quad \mathbf{p}_i \subset \mathbf{T}_r^{al} \quad (4-3)$$

式中， $\mathbf{Q} \subset \mathbf{M}_i$ 为 $m$ 个源域候选混合子集的集合， $\mathbf{q}_i$ 代表第*i*个候选子集， $\mathbf{p}_i$ 代表一个目标域点云帧中已标注点的集合。 $\|\mathbf{q}_i\|$ 和 $\|\mathbf{p}_i\|$ 分别代表候选子集 $\mathbf{q}_i$ 和目标点云已标注点 $\mathbf{p}_i$ 的数量。候选子集 $\mathbf{Q}$ 将在下一个模块中进行筛选，以进一步缓解类别不平衡问题。

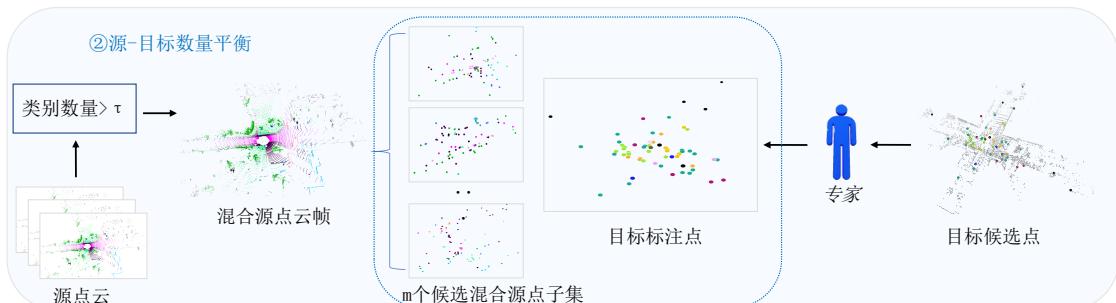


图 4-2 源-目标数量平衡模块

Fig. 4-2 Source-target amount balance module

### 4.3.3 类别平衡主动混合算法

针对主动学习在目标域中引发的类别不平衡问题，本章提出了类别平衡主动混合算法。在主动学习中，由于所选取点的类别无法提前预知，因此主动选择的点往往会导致目标域每一帧点云中各类别之间所含标注点数存在显著差异。值得关注的是，在跨域点云分割任务中，源域通常具备大量完整标注且可直接利用的语义类别点。而这为本章解决主动学习选择的类别不平衡提供了一个思路。如图4-3所示，从源域数据上选择候选点并与目标域标注点混合，使得混合后的数据中类别分布相对平衡。混合后的中间域数据不仅有助于学习域不变特征，也能有效缓解类别不平衡的问题。

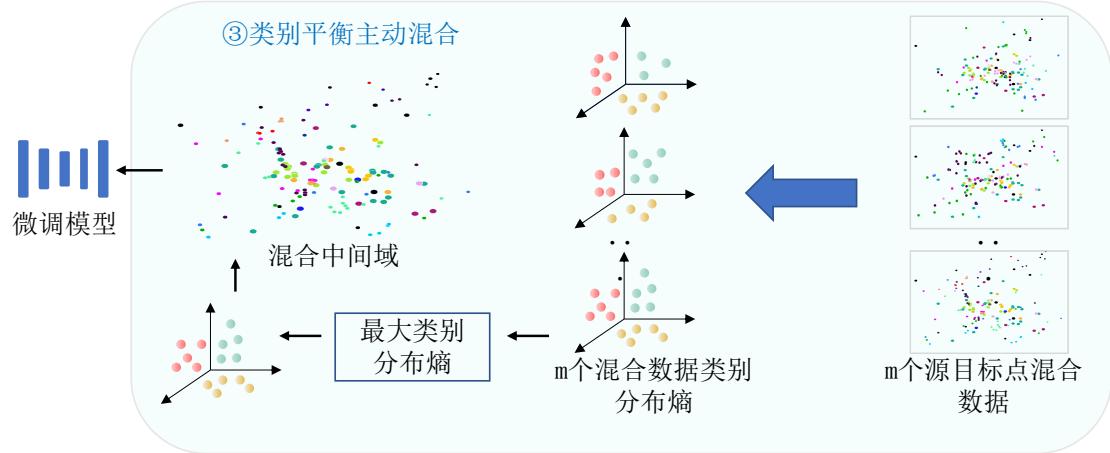


图 4-3 类别平衡主动混合模块

Fig. 4-3 Class-balanced active mixing module

基于源-目标数量平衡模块生成的  $m$  候选子集，平衡模块的核心任务是从集合  $\mathbf{Q}$  中选取最优子集  $\mathbf{q}_{max}$ ，使得该子集与已选数据组合后的整体类别分布达到最优平衡状态。为实现这一目标，本模块采用信息熵作为量化指标，即通过计算候选子集与目标域数据混合后的真实标签分布熵值，筛选出能够最大程度缓解类别不平衡的候选子集。具体实施过程定义如下：设定已经构建的动态中间域训练数据集包含  $N_{all}$  个已标注点，其中类别  $c$  的标注点数量为  $N_c$ 。对于混合源域帧中的候选子集  $\mathbf{q}_i$ ，将其与对应目标域标注数据进行混合，混合后形成的数据子集包含  $N_{mix}$  个点，其中类别  $c$  的标注点数量为  $N_{mix}^c$ 。为评估候选子集  $\mathbf{q}_i$  的平衡效果，需计算其引入后各类别的相对分布比例。具体方法是将已标注数据总量  $N_{all}$  与候选混合数据量  $N_{mix}$  进行叠加，按公式(4-4)逐类计算混合后的分布比例  $\{R_{i,c}\}_{c=1}^C$ 。

$$R_{i,c} = \frac{N_c + N_{mix}^c}{N_{all} + N_{mix}} \quad (4-4)$$

同时，为了保证在同一个量纲下比较，对  $R_{i,c}$  使用 softmax 函数进行归一化处理，得到  $\hat{R}_{i,c}$ ，如公式(4-5)所示：

$$\hat{R}_{i,c} = \frac{e^{R_{i,c}}}{\sum_{c=1}^C e^{R_{i,c}}} \quad (4-5)$$

其表示选择候选子集  $\mathbf{q}_i$  与当前已标注的目标点云帧混合后，各类别所占点数的归一化概率。再然后，通过计算类别分布的熵值来可作为候选子集的选择评判指标，如公式(4-6)所示：

$$H_i(\hat{R}_i) = - \sum_{c=1}^C \hat{R}_{i,c} \log(\hat{R}_{i,c}) \quad (4-6)$$

式中， $H_i$  表示选择候选子集  $\mathbf{q}_i$  混合后类别分布的熵值结果，并选择熵值最大的一个候选子集  $\mathbf{q}_{max} = \{\mathbf{X}_{\mathbf{q}_{max}}, \mathbf{Y}_{\mathbf{q}_{max}}\}$  作为最终混合的源域集，与目标域标注点集

$\mathbf{p}_i = \{\mathbf{X}_{p_i}, \mathbf{Y}_{p_i}\}$  组成新的中间域数据集  $\mathbf{I} = \{\hat{\mathbf{X}}, \hat{\mathbf{Y}}\}$ , 如公式(4-7)所示:

$$\begin{aligned}\hat{\mathbf{X}} &= concat\{\mathbf{X}_{q_{max}}, \mathbf{X}_{p_i}\} \\ \hat{\mathbf{Y}} &= concat\{\mathbf{Y}_{q_{max}}, \mathbf{Y}_{p_i}\}\end{aligned}\quad (4-7)$$

其中,  $\hat{\mathbf{X}}$  表示混合的点数据,  $\hat{\mathbf{Y}}$  则表示对应的标签。最后更新训练集总点数以及类别点数, 公式如(4-8)所示:

$$N_c = N_c + N_{mix}^c, \quad N_{all} = N_{all} + N_{mix} \quad (4-8)$$

#### 4.3.4 损失函数

在后续的训练中, 模型仅采用混合后的中间域数据进行微调优化。由于这些数据均包含标注信息, 因此所有实验均采用交叉熵损失函数作为优化策略, 如公式(4-9)所示, 其中  $\hat{x}_i \in \hat{\mathbf{X}}$ ,  $\hat{y}_i \in \hat{\mathbf{Y}}$ :

$$\mathbf{L}_{CE} = -\frac{1}{|I|} \sum_{i \in I} \sum_{c=1}^C \hat{y}_i^c \log h(\hat{x}_i^c) \quad (4-9)$$

### 4.4 实验评估

#### 4.4.1 实验设置

在本章实验中, 同样使用 MinkowskiNet<sup>[18]</sup> 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)<sup>[80]</sup> 作为学习优化器。所有实验都在单张 NVIDIA RTX A6000 GPU 上进行训练。本章方法同样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。由于在四个跨域数据集上的映射类别数量不同, 参数  $\tau$  在 SynLiDAR→SemanticKITTI、SynLiDAR→SemanticPOSS、SemanticKITTI→nuScenes 和 nuScenes→SemanticKITTI 分别设置为 13、9、6、6, 候选参数  $m$  统一设为 10。此外, 对于其他实验参数的配置仍延续第三章的实验设定。

#### 4.4.2 实验结果

本章方法同样在合成到真实场景下的跨域数据集 SynLiDAR→SemanticKITTI 和 SynLiDAR→SemanticPOSS, 以及真实到真实场景下的跨域数据集 SemanticKITTI →nuScenes 和 nuScenes→SemanticKITTI 上做了相应的实验, 并与最先进的方法以及前一章的方法进行了对比以验证其有效性。为了公平比较, 如果没有特殊标明标注预算的情况下, 默认都是在 0.1% 标注预算下做出的实验结果。

#### 4.4.2.1 合成到真实场景

表 4-1 本章方法与其他域适应方法在 SynLiDAR→SemanticKITTI 数据上的比较  
Table 4-1 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 <sup>[76]</sup>	UDA	-	23.0
AdvEnt <sup>[81]</sup>	UDA	-	25.8
CRST <sup>[82]</sup>	UDA	-	26.5
ST-PCT <sup>[53]</sup>	UDA	-	28.9
PolarMix <sup>[83]</sup>	UDA	-	32.2
CoSMix <sup>[54]</sup>	UDA	-	31.0
DGT-ST <sup>[72]</sup>	UDA	-	43.1
MME <sup>[60]</sup>	SSDA	0.04%	24.5
APE <sup>[86]</sup>	SSDA	0.04%	25.1
APE-PCT <sup>[53]</sup>	SSDA	0.04%	27.0
CoSMix-SSDA <sup>[62]</sup>	SSDA	0.04%	34.3
本章方法	SSDA	0.04%	50.5
Annotator <sup>[43]</sup>	ADA	0.1%	57.7
前章方法	ADA	0.1%	58.7
本章方法	ADA	0.1%	<b>65.5</b>

本章在 SynLiDAR→SemanticKITTI 跨域数据集上的实验如表4-1所示，分析对比不同的结果可知，在没有任何域适应的情况下，Source-Only 模型在目标域上的性能非常低，仅有 22.8%，这说明直接将源域训练的模型用到目标域效果很差。而使用全部目标域数据训练的模型（Target-Only）达到了 60.1% 的结果，这说明域间隙的存在会导致在源域训练好的模型在目标域上性能急剧下降。

在不需要目标域标注的无监督域适应（UDA）方法中，DGT-ST 的效果最好，达到 43.1%，比 Source-Only 提升了近 20 个百分点，但相比 Target-Only 仍有较大差距。这说明即使不标注目标域数据，通过域适应也能部分缩小域间差异，但还不能完全解决问题。为了与半监督域适应方法进行公平比较，本章将主动学习的标注预算降低到与半监督同一量级即 0.04%，即使在这种极少量的标注下，本章方法仍达到了 50.5% 的结果，比同样标注量的 CoSMix-SSDA 高出了近 16 个百分点。这

说明在标注极少的情况下，本章方法能更有效地利用有限的标注信息。

在与主动域适应方法进行比较时，标注比例提升到 0.1%，本章方法的结果达到了 65.5%，不仅比前章方法提高了近 7 个百分点，甚至超越了 Target-Only，这说明本章方法构建的中间域增强了模型的性能，学习到了源和目标域的信息，缓解了类不平衡，因此超越了全监督结果。当然，超越 Target-Only 本方法并非先例，在之前的一些域适应文章中<sup>[87]</sup> 也有做到。另外，值得注意的是，虽然标注量增加了 2.5 倍，但性能提升幅度远高于标注量的增长比例，这说明在前一章的基础上，结合本章主动混合方法，充分发挥了它们的最佳性能。

表 4-2 本章方法与其他域适应方法在 SynLiDAR→SemanticPOSS 数据上的比较

Table 4-2 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 <sup>[82]</sup>	UDA	-	27.1
ST-PCT <sup>[53]</sup>	UDA	-	29.6
PolarMix <sup>[83]</sup>	UDA	-	30.4
CoSMix <sup>[54]</sup>	UDA	-	40.4
DGT-ST <sup>[72]</sup>	UDA	-	50.8
MME <sup>[60]</sup>	SSDA	0.01%	33.2
APE <sup>[86]</sup>	SSDA	0.01%	30.3
APE-PCT <sup>[53]</sup>	SSDA	0.01%	31.2
CoSMix-SSDA <sup>[62]</sup>	SSDA	0.01%	41.0
本章方法	SSDA	0.01%	57.5
Annotator <sup>[43]</sup>	ADA	0.1%	52.0
前章方法	SSDA	0.1%	56.6
本章方法	ADA	0.1%	<b>60.2</b>

如表4-2结果显示，在 SynLiDAR→SemanticPOSS 数据集上域差异依然显著。在无监督域适应（UDA）中，DGT-ST 以 50.8% 的表现达到最优，但距离全标注仍有 7 个百分点的差距。本章方法在 0.01% 的极低标注预算下达到了 57.5% 的结果，远超 CoSMix-SSDA，甚至接近 Target-Only 水平。当与主动域适应方法（ADA）进行比较时，标注量提升至 0.1%，本章方法以 60.2% 的结果超越前章方法和 Annotator 分别 3.6 和 7.8 个百分点，超过 Targe-Only 2.2 个百分点，这证明本章的主动混合方

法是有效果的，在结合第三章的方法后可以大幅度提升模型性能。

#### 4.4.2.2 真实到真实场景

SemanticKITTI→nuScenes 的结果如表4-3所示，Source-Only 模型在目标域的结果仅为 33.7%，而全标注的 Target-Only 模型则达到了 82.7%，这验证了真实到真实跨域场景中源域与目标域的巨大差异。在无监督域适应（UDA）方法中，LiDOG 以 34.9% 的表现最优，但仍远低于全标注结果，表明纯无监督策略在复杂域差异下的局限性。在主动域适应（ADA）场景下，本章方法以 81.0% 的结果显著超越 Annotator 的 75.9%，提升幅度达 5.1 分，且与 Target-Only 的差距缩小至 1.7 个百分点，同时与前章方法相比，提升了 4.5 个百分点。从上述实验结果可知本章方法通过主动混合方法有效利用有限标注信息与源域数据，在极低标注成本下逼近全标注性能，证明了其在真实跨域场景中的有效性。

表 4-3 本章方法与其他域适应方法在 SemanticKITTI→nuScenes 数据上的比较  
Table 4-3 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 <sup>[69]</sup>	UDA	-	31.5
CoSMix <sup>[54]</sup>	UDA	-	29.8
SN <sup>[84]</sup>	UDA	-	25.8
RayCast <sup>[52]</sup>	UDA	-	30.9
LiDOG <sup>[85]</sup>	UDA	-	34.9
前章方法	ADA	0.1%	76.5
Annotator <sup>[43]</sup>	ADA	0.1%	75.9
本章方法	ADA	0.1%	<b>81.0</b>

如表4-4数据显示，在 nuScenes→SemanticKITTI 的实验结果中，Source-Only 模型仅得 32.5%，而 Target-Only 模型通过全标注数据达到 85.4%，再次印证域差异的显著影响。在无监督域适应（UDA）方法中，LiDOG 以 41.2% 的表现依然取得最优，但仍与全标注结果相差 44.2 个百分点。在标注量仅 0.1% 的主动域适应（ADA）场景下，本章方法以 86.2% 的结果超越目标域全监督（Target-Only）0.8 个百分点，远超 Annotator 的 81.8%，超越前章方法 3.2 个百分点。本章方法的突破性结果表明，通过主动混合方法在极低标注量下超越了全标注的性能，进一步证明了本章方

法的真实有效性。

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

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

方法	域适应	标注结果	
Source-Only	-	-	32.5
Target-Only	-	100%	85.4
Mix3D <sup>[69]</sup>	UDA	-	32.4
CoSMix <sup>[54]</sup>	UDA	-	36.8
SN <sup>[84]</sup>	UDA	-	23.6
RayCast <sup>[52]</sup>	UDA	-	31.5
LiDOG <sup>[85]</sup>	UDA	-	41.2
前章方法	ADA	0.1%	83.0
Annotator <sup>[43]</sup>	ADA	0.1%	81.8
<b>本章方法</b>	ADA	0.1%	<b>86.2</b>

#### 4.4.2.3 分割可视化结果

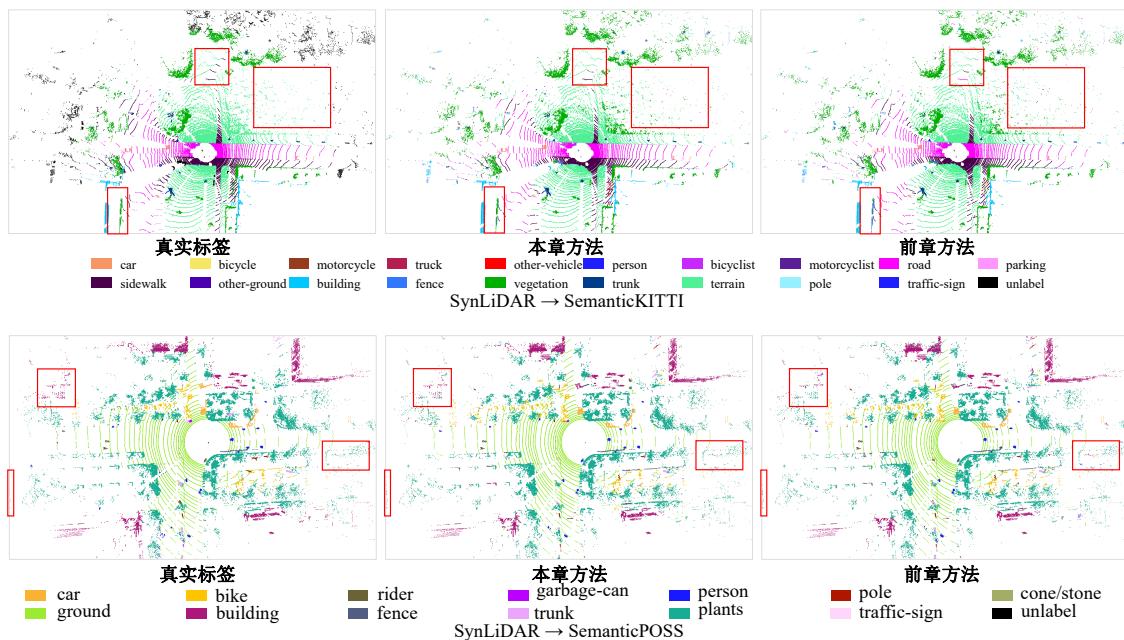


图 4-4 本章方法在合成到真实场景分割可视化图

Fig. 4-4 Visualization of the segmentation results in the synthetic-to-real

为了进一步证明本章方法的有效性，提供了合成到真实场景下的可视化结果如图4-4所示。以真实标签为参照，与前章方法进行对比，使用红框标出差异区域，

可以看到本章方法在一些语义边界处有着更好的辨别效果。同样的，本章也提供了真实到真实场景下的可视化结果如图4-5所示。本章方法在人造物品、人行道等边缘类别相比前章均取得了更好的分割效果。

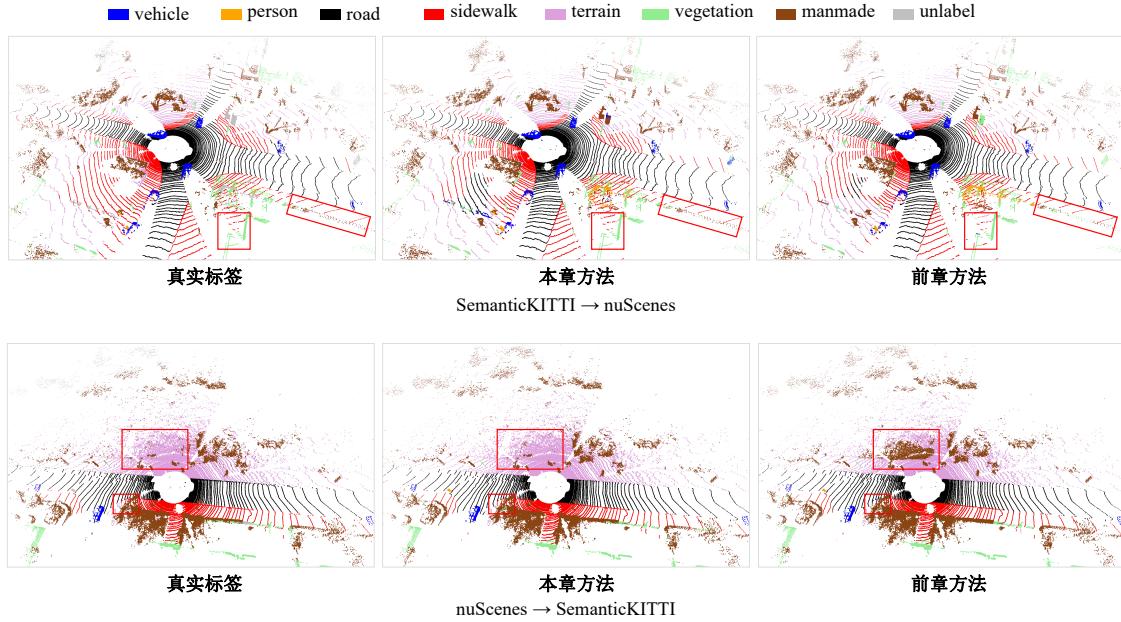


图 4-5 本章方法在真实到真实场景分割可视化图

Fig. 4-5 Visualization of the segmentation results in the real-to-real

### 4.4.3 消融对比实验

#### 4.4.3.1 对比实验

表 4-5 本章方法与其他 Mixing 方法在 SynLiDAR→SemanticKITTI 数据上的比较

Table 4-5 Comparison with other mixing methods on SynLiDAR→SemanticKITTI

主动学习	Mixing 方法	标注	结果
SPAL	LaserMix <sup>[70]</sup>	0.1%	39.6
	PolarMix <sup>[83]</sup>	0.1%	46.6
	<b>Ours</b>	0.1%	<b>65.5</b>

为证明本章方法的有效性，在 SynLiDAR→SemanticKITTI 上进行了与其他传统 Mixing 方法的对比实验，在 SPAL 框架下统一采用 0.1% 标注量进行公平比较，实验结果如表4-5所示。实验中，PolarMix 和 LaerMix 都是按照 1/4 比例进行混合，即源和目标交换自身 1/4 圆周角度内的点或者 1/4 线束角度内的点进行混合。在相同标注条件下，LaserMix 和 PolarMix 方法分别取得 39.6% 和 46.6% 的性能结

果，而本章提出的方法以 65.5% 的显著优势超越了它们，性能提升幅度达 25.9% 和 18.9%。这一结果验证了传统混合方法在跨域数据融合时对类别分布不敏感，数量过多导致模型学习的知识更偏向源域等问题，而本章方法通过双域数量平衡算法以及熵驱动的源域筛选动态平衡算法，能够精准选择与目标域互补的源域样本，有效缓解混合过程中的类别不均衡问题。

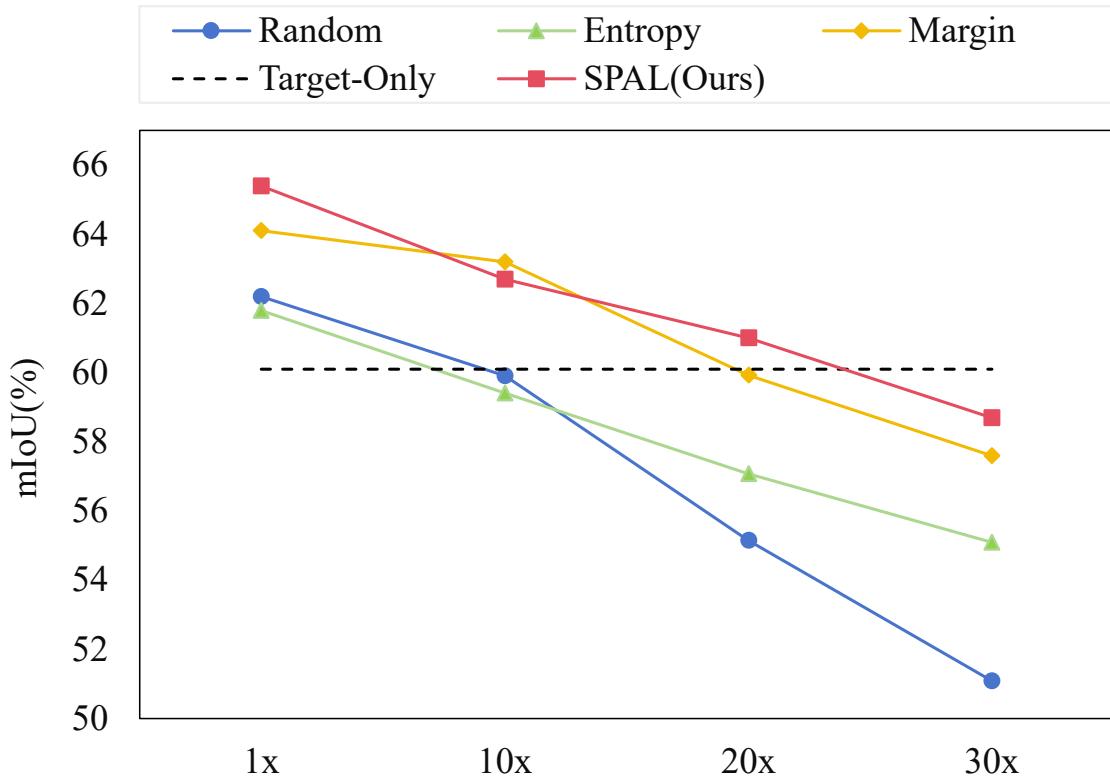


图 4-6 本章方法结合不同主动学习在多个混合比例下的对比结果

Fig. 4-6 Comparison results of different active learning methods under various mixing ratios

如图4-6所示，本章在 SynLiDAR→SemanticKITTI 数据集上进行了对比实验，展示了不同主动学习策略在不同源-目标混合比例下的性能对比。其中，1x 代表的是本章提出的源-目标域数量平衡算法（即源域与目标域数据按 1:1 比例动态混合），10x 代表源域与目标域数据按 10:1 比例动态混合，20x 和 30x 则分别代表源域与目标域数据按 20:1 和 30:1 比例进行动态混合。实验结果显示，在默认 0.1% 的标注预算下，本章源-目标数量平衡方法在多种主动学习方法中均取得了最优的结果，证明了该方法的有效性，根据域适应以及主动学习的特点，简单高效的提升了性能。而在源-目标域数量平衡算法下，与前一章提出的 SPAL 方法的结合又是最高的，达到了 65.5% 的结果，这证明了 SPAL 与本章方法结合的有效性。

为了进一步证明本章方法结合前章 SPAL 方法的有效性，在四个跨域数据集（SynLiDAR→SemanticKITTI、SynLiDAR→SemanticPOSS、SemanticKITTI→nuScenes、

nuScenes→SemanticKITTI) 上均进行了实验。如表4-6所示，在四个跨域数据实验中，结合两章的算法均取得最高结果，且稳定超越其他主动学习方法。这验证了第四章提出的主动混合方法与第三章 SPAL 框架结合的有效性和普适性。这一跨数据集的全面优势表明，源-目标平衡机制能够适配不同域差异场景，显著提升模型的性能。

表 4-6 本章方法结合不同主动学习方法结果对比

Table 4-6 Comparison with different active learning methods that incorporate this chapter's modules

数据集	方法	标注	结果
SynLiDAR→SemanticKITTI	Random	0.1%	62.2
	Entropy <sup>[73]</sup>	0.1%	61.8
	Margin <sup>[74]</sup>	0.1%	64.0
SynLiDAR→SemanticPOSS	<b>SPAL</b>	0.1%	<b>65.5</b>
	Random	0.1%	58.3
	Entropy <sup>[73]</sup>	0.1%	55.6
SemanticKITTI→nuScenes	Margin <sup>[74]</sup>	0.1%	60.0
	<b>SPAL</b>	0.1%	<b>60.2</b>
	Random	0.1%	85.7
nuScenes→SemanticKITTI	Entropy <sup>[73]</sup>	0.1%	83.2
	Margin <sup>[74]</sup>	0.1%	85.9
	<b>SPAL</b>	0.1%	<b>86.2</b>
	Random	0.1%	77.7
	Entropy <sup>[73]</sup>	0.1%	76.5
	Margin <sup>[74]</sup>	0.1%	80.7
	<b>SPAL</b>	0.1%	<b>81.0</b>

#### 4.4.3.2 消融实验

最后，在 SynLiDAR→SemanticKITTI 跨域数据上进行了本章的消融实验，主动学习默认使用 0.1% 的标注预算。如表4-7所示，结果证明本章提出的两个模块对性能提升均有贡献。当任何模块都不用时，基线结果仅为 22.84%。当仅使用第三章的 SPAL 模块时，性能提升至 58.70%。在引入源-目标平衡模块（STNB）后，性能跃升至 65.46%，表明在域差异场景下，数量平衡的混合能帮助模型学习到更有效的域不变特征，显著缓解域差异问题。最终完整方法以 65.51% 达到最优，CBAM 进一步优化了 0.05 个百分点，说明其在解决类别平衡问题上也有所贡献。

表 4-7 本章方法模块消融实验

Table 4-7 Ablation experiments on modules

SPAL	STNB	CBAM	结果
			22.84
✓			58.70
✓	✓		65.46
✓	✓	✓	<b>65.51</b>

## 4.5 本章小结

本章提出了一种面向点云语义分割域适应任务的主动混合方法，重点探索主动学习与 Mixing 方法协同作用对域适应性能的提升。通过设计源-目标数量平衡算法（STNB），在每轮主动学习中动态匹配源域与目标域标注点的数量，确保混合数据中双域信息均衡融合。同时，类别平衡主动混合算法（CBAM）通过计算混合后类别分布的熵值，评估多个候选源域子集与目标域标注集混合后的分布平衡性，从而选择最为均衡的集合。这种方法增强了主动学习与混合策略的协同效率。最后，在两个跨域场景下的四个跨域数据集上的实验表明，在 0.1% 标注量下，本章完整方法在四个跨域数据集中均达最优的性能，并且显著超越了传统混合方法，最后的消融实验显示模块协同作用时性能提升最大，充分验证了本章算法的有效性。

## 第 5 章 总结与展望

### 5.1 主要结论

点云语义分割是实现三维场景感知的关键技术，也是自动驾驶、机器人和增强现实等领域实现高级智能的重要基础。然而，目前多数优秀深度学习模型均依赖于逐点标注数据进行训练，而点云逐点标注工作极为繁琐且耗时，这成为制约这些模型实际应用的主要障碍。因此，在无标注或仅有少量标注数据的条件下获得高性能模型，显得尤为重要。域适应方法旨在将基于全标注源域训练得到的模型迁移至未标注的目标域，以实现对新数据的学习并降低标注成本。主动学习则通过选择最具价值的样本进行标注以优化模型训练效果，而主动域适应则融合了二者的优势，提供了一种更为有效的解决方案。尽管该方法在二维图像语义分割领域已取得一定成果，但针对三维点云的研究仍显不足，主要存在以下问题：第一，目前多数主动域适应方法集中于二维图像领域，因三维点云数据在结构和数量级上均存在显著差异，直接移植图像方法往往难以奏效。第二，传统主动学习方法通常针对单一数据域设计，未能充分考虑源域与目标域之间的域偏移问题，因此在域适应场景下直接应用可能导致样本选择次优，影响最终效果。第三，现有方法未能充分整合标注后的目标数据与全标注的源域数据，因主动学习选取的目标点数量远低于源域数据量，容易在迁移过程中积累域偏移，使模型倾向于源域，从而难以获得最优性能。

基于上述问题，通过大量文献阅读和实验验证，本文的主要研究内容可归纳如下：

- 1) 通过大量阅读相关文献，对点云语义分割方法、主动学习、域自适应以及分割中的混合方法等相关领域的研究现状和进展进行了系统总结和分析。基于个人的理解与观点，对现有方法进行了归类，总结了各自的优势与不足，并提出了一些具有实际意义的研究思路和解决方案。
- 2) 对点云语义分割、主动学习和域自适应相关的基础背景知识做了详细总结。包括点云的常见表示方式，点云语义分割任务的概念、常见模型以及评估指标，主动学习的基本概念和流程以及域自适应任务的概念、目标以及常见的域对齐方式等。并对跨域点云语义分割相关任务中常用公开的数据集进行了介绍整理。
- 3) 提出了一种基于点云语义分割域适应的主动学习方法。该方法通过构建出的源域原型，指导目标数据的选择，同时结合不确定性，筛选出兼具高不确定性和域差异性的目标点，用以高效提升模型的性能。随后将这些选择后的目标点进行标注，同时结合 Mixing 方法，构建兼具双域信息的中间域数据，帮助模型学习到

更加鲁棒的域不变特征，进一步缩小域间隙。文中首先详细的介绍了该方法的研究动机和贡献，接着详细的阐述了该方法中的三个模块，包括源域原型构建模块、原型指导的数据选择模块、动态中间域构建模块。最后在合成到真实以及真实到真实的 4 个跨域数据集上，对提出的方法进行大量实验和可视化分析证明了该方法的有效性。

4) 提出了一个基于点云语义分割域适应的主动混合方法。该方法在第三章方法框架基础上，深入探索了域适应场景下主动学习与 Mixing 方法的结合，并将动态中间域构建模块更改为两个模块：源-目标数量平衡模块和类别平衡动态混合模块。该方法改善了模型迁移过程中因源域数据过多而导致的模型学到的知识偏向源域的问题，同时减缓了主动学习类别不平衡问题。最后，在合成到真实以及真实到真实的四个跨域数据集上，对该方法进行了大量实验和可视化分析，从而充分验证了其有效性。

## 5.2 研究展望

本文的研究中，对于所提出的两个方法分别解决了不同的问题。基于点云语义分割域适应的主动学习方法中，提出了一种适用于点云语义分割域适应的主动学习方法，在结合 Mixing 后发挥最大优势，超越传统主动学习方法。基于点云语义分割域适应的主动混合方法，则是在上个方法的框架下，对点云语义分割域适应场景下主动学习与 Mixing 的结合进行了深入研究和探索，通过构建数量及类别平衡的中间域数据缓解了迁移过程中的域偏移累积问题和类别不平稳问题。虽然上述研究在与同领域方法的比较中取得了先进的表现，但是受限于时间、设施等多种因素，仍然存在很大的局限和提升空间。同时，基于对目前前沿技术的理解以及所做的大量实验验证分析，归结了一些点云语义分割域自适应领域中仍存在的需要进一步解决或者缓解的问题。为此，本文对未来研究做出了如下展望：

1) 模型泛化能力提升。目前，本研究的跨域实验虽然涵盖了合成到真实和真实到真实两个场景，但均基于室外数据集进行的测试，尚未在室内场景下验证。由于室外与室内数据在密度和场景规模上存在显著差异，本文方法在室内跨域任务中的适用性尚无保障。因此，未来研究需着力提升模型的泛化能力，以确保其在室内外跨域数据集上均能获得良好表现，这将是一个重要的探索方向。

2) 进一步挖掘未标注数据的潜在价值。尽管本研究通过主动学习策略筛选出极少的高价值目标域样本进行标注，但占总量 99% 的未标注点云数据仍未被有效利用。为解决这一资源浪费问题，未来研究可探索主动学习与伪标签技术的协同优化框架：首先通过构建基于不确定性和域间差异的双重评估机制，对未标注点云

进行可信度分层，筛选出置信度高于阈值的点生成伪标签；同时设计对抗噪声干扰的自适应校正模块，利用点云局部几何一致性特征对伪标签进行拓扑纠偏。此外，可建立动态迭代优化机制，在每轮主动学习标注新增样本后，同步更新伪标签生成模型，形成标注数据增强与伪标签质量提升的正向循环。但也需要特别关注伪标签噪声传播与域偏移累积的耦合效应。同时，通过在特征空间构建源域标注数据、目标域标注数据、目标域伪标签数据的三体平衡约束，从而在扩大训练数据规模的同时维持模型的域不变特征表达能力。

3) 语义选择的主动混合。尽管本文提出了一种基于类别平衡的主动混合方法，但在域适应条件下，对混合方法的探索仍远远不够。目前的策略仅对源域数据进行了约束，而目标域数据仍有较大改进空间。未来可以在混合过程中采用更具选择性的策略，分别从源域和目标域中依据语义平衡挑选数据，然后进行有针对性的混合，从而更充分地利用目标域信息，进一步提升模型在跨域任务中的性能。

4) 大模型融合。大模型融合是当前的研究热点，因其卓越的文本理解和逻辑推理能力，已在多个领域取得显著成效。在最近的研究中，虽然有越来越多的方法开始引入大模型以提升整体表现。然而，在语义分割域适应任务中，大模型的应用仍然缺乏深入探索。如何根据域自适应的特点，充分利用大模型的优势，构建更高效的跨域融合策略，并为提升模型性能提供新的方向，这也是未来值得重点关注和研究的课题。

## 参考文献

- [1] LEVINSON J, ASKELAND J, BECKER J, et al. Towards fully autonomous driving: systems and algorithms[C]. IEEE Intelligent Vehicles Symposium, Baden-Baden, Germany, 2011: 163-168.
- [2] ZHOU Y, WANG Y, POIESI F, et al. Loop closure detection using local 3d deep descriptors[J]. IEEE Robotics and Automation Letters.2022, 7(3): 6335-6342.
- [3] CARMIGNIANI J, FURHT B. Augmented reality: an overview[J]. Handbook of Augmented Reality.2011: 3-46.
- [4] ARMENI I, SENER O, ZAMIR A R, et al. 3d semantic parsing of large-scale indoor spaces[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 1534-1543.
- [5] DAI A, CHANG A X, SAVVA M, et al. Scannet: richly-annotated 3d reconstructions of indoor scenes[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 5828-5839.
- [6] BEHLEY J, GARBADE M, MILIOTO A, et al. Semantickitti: a dataset for semantic scene understanding of lidar sequences[C]. IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019: 9297-9307.
- [7] CAESAR H, BANKITI V, LANG A H, et al. Nuscenes: a multimodal dataset for autonomous driving[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 2020.
- [8] PAN Y, GAO B, MEI J, et al. Semanticposs: a point cloud dataset with large quantity of dynamic instances[C]. IEEE Intelligent Vehicles Symposium, Aachen, Germany, 2020: 687-693.
- [9] AKSOY E E, BACI S, CAVDAR S. Salsanet: fast road and vehicle segmentation in lidar point clouds for autonomous driving[C]. IEEE Intelligent Vehicles Symposium, Aachen, Germany, 2020: 926-932.
- [10] CORTINHAL T, TZELEPIS G, ERDAL AKSOY E. Salsanext: fast, uncertainty-aware semantic segmentation of lidar point clouds[C]. Advances in Visual Computing: 15th International Symposium, ISVC, San Diego, CA, USA, 2020: 207-222.
- [11] MILIOTO A, VIZZO I, BEHLEY J, et al. Rangenet++: fast and accurate lidar semantic segmentation[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, Macau, China, 2019: 4213-4220.

- [12] WU B, WAN A, YUE X, et al. Squeezeseg: convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud[C]. IEEE International Conference on Robotics and Automation, Brisbane, Australia, 2018: 1887-1893.
- [13] WU B, ZHOU X, ZHAO S, et al. Squeezesegv2: improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud[C]. IEEE International Conference on Robotics and Automation, Montreal, Canada, 2019: 4376-4382.
- [14] XU C, WU B, WANG Z, et al. Squeezesegv3: spatially-adaptive convolution for efficient point-cloud segmentation[C]. European Conference on Computer Vision, Glasgow, UK, 2020: 1-19.
- [15] BOULCH A, LE SAUX B, AUDEBERT N, et al. Unstructured point cloud semantic labeling using deep segmentation networks[J]. 3dor@ eurographics.2017, 3: 1-8.
- [16] KUNDU A, YIN X, FATHI A, et al. Virtual multi-view fusion for 3d semantic segmentation[C]. European Conference on Computer Vision, Glasgow, UK, 2020: 518-535.
- [17] ROBERT D, VALLET B, LANDRIEU L. Learning multi-view aggregation in the wild for large-scale 3d semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, USA, 2022: 5575-5584.
- [18] CHOY C, GWAK J, SAVARESE S. 4d spatio-temporal convnets: minkowski convolutional neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, California, USA, 2019: 3075-3084.
- [19] GRAHAM B, ENGELCKE M, VAN DER MAATEN L. 3d semantic segmentation with submanifold sparse convolutional networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 9224-9232.
- [20] TANG H, LIU Z, ZHAO S, et al. Searching efficient 3d architectures with sparse point-voxel convolution[C]. European Conference on Computer Vision, Glasgow, UK, 2020: 685-702.
- [21] PENG B, WU X, JIANG L, et al. Oa-cnns: omni-adaptive sparse cnns for 3d semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2024: 21305-21315.
- [22] QI C R, SU H, MO K, et al. Pointnet: deep learning on point sets for 3d classification and segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 652-660.
- [23] QI C 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.
- [24] 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, South Korea, 2019: 6411-6420.
- [25] LI Y, BU R, SUN M, et al. Pointcnn: convolution on x-transformed points[J]. Advances in Neural Information Processing Systems.2018, 31.

- [26] HU Q, YANG B, XIE L, et al. Randla-net: efficient semantic segmentation of large-scale point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 11108-11117.
- [27] FAN S, DONG Q, ZHU F, et al. Scf-net: learning spatial contextual features for large-scale point cloud segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 14504-14513.
- [28] QIU S, ANWAR S, BARNES N. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 1757-1767.
- [29] ZHAO H, JIANG L, JIA J, et al. Point transformer[C]. IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 16259-16268.
- [30] 吴军, 崔玥, 赵雪梅, et al. Ssa-pointnet++: 空间自注意力机制下的 3d 点云语义分割网络 [J]. 计算机辅助设计与图形学学报.2022, 34(03): 437-448.
- [31] 杨军, 李博赞. 基于自注意力特征融合组卷积神经网络的三维点云语义分割 [J]. 光学精密工程.2022, 30(07): 840-853.
- [32] WU X, JIANG L, WANG P-S, et al. Point transformer v3: simpler faster stronger[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA, 2024: 4840-4851.
- [33] TAN Y, YANG L, HU Q, et al. Batch mode active learning for semantic segmentation based on multi-clue sample selection[C]. ACM International Conference on Information and Knowledge Management, Beijing China, 2019: 831-840.
- [34] SINHA S, EBRAHIMI S, DARRELL T. Variational adversarial active learning[C]. IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019: 5972-5981.
- [35] ZHANG B, LI L, YANG S, et al. State-relabeling adversarial active learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 8756-8765.
- [36] XIE S, FENG Z, CHEN Y, et al. Deal: difficulty-aware active learning for semantic segmentation[C]. Asian Conference on Computer Vision, Virtual Kyoto, 2020: 672-688.
- [37] HUANG S, WANG T, XIONG H, et al. Semi-supervised active learning with temporal output discrepancy[C]. IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 3447-3456.
- [38] QIAO Y, ZHU J, LONG C, et al. Cpral: collaborative panoptic-regional active learning for semantic segmentation[C]. AAAI Conference on Artificial Intelligence, Virtual, 2022: 2108-2116.
- [39] SIDDIQUI Y, VALENTIN J, NIESSNER M. Viewal: active learning with viewpoint entropy for semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 9433-9443.

- [40] CAI L, XU X, LIEW J H, et al. Revisiting superpixels for active learning in semantic segmentation with realistic annotation costs[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 10988-10997.
- [41] LIN Y, VOSSELMAN G, CAO Y, et al. Active and incremental learning for semantic als point cloud segmentation[J]. ISPRS Journal of Photogrammetry and Remote Sensing.2020, 169: 73-92.
- [42] WANG P, YAO W, SHAO J. One class one click: quasi scene-level weakly supervised point cloud semantic segmentation with active learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing.2023, 204: 89-104.
- [43] 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.
- [44] SHAO F, LUO Y, LIU P, et al. Active learning for point cloud semantic segmentation via spatial-structural diversity reasoning[C]. ACM International Conference on Multimedia, Lisboa Portugal, 2022: 2575-2585.
- [45] HU Z, BAI X, ZHANG R, et al. Lidal: inter-frame uncertainty based active learning for 3d lidar semantic segmentation[C]. European Conference on Computer Vision, Tel Aviv, Israel, 2022: 248-265.
- [46] LIU G, van KAICK O, HUANG H, et al. Active self-training for weakly supervised 3d scene semantic segmentation[J]. Computational Visual Media.2024, 10(3): 425-438.
- [47] WEI J, LIN Y, CAESAR H. Basal: size-balanced warm start active learning for lidar semantic segmentation[C]. IEEE International Conference on Robotics and Automation, Yokohama, Japan, 2024: 18258-18264.
- [48] SHIN G, XIE W, ALBANIE S. All you need are a few pixels: semantic segmentation with pixelpick[C]. IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 1687-1697.
- [49] RÜCKIN J, MAGISTRI F, STACHNISS C, et al. Semi-supervised active learning for semantic segmentation in unknown environments using informative path planning[J]. IEEE Robotics and Automation Letters.2024, 9(3): 2662-2669.
- [50] XU Z, YUAN B, ZHAO S, et al. Hierarchical point-based active learning for semi-supervised point cloud semantic segmentation[C]. IEEE/CVF International Conference on Computer Vision, Paris, France, 2023: 18098-18108.
- [51] YI L, GONG B, FUNKHOUSER T. Complete & label: a domain adaptation approach to semantic segmentation of lidar point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 15363-15373.

- [52] LANGER F, MILIOTO A, HAAG A, et al. Domain transfer for semantic segmentation of lidar data using deep neural networks[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 2020: 8263-8270.
- [53] XIAO A, HUANG J, GUAN D, et al. Transfer learning from synthetic to real lidar point cloud for semantic segmentation[C]. AAAI Conference on Artificial Intelligence, Virtual, 2022: 2795-2803.
- [54] SALTORI C, GALASSO F, FIAMENI G, et al. Cosmix: compositional semantic mix for domain adaptation in 3d lidar segmentation[C]. European Conference on Computer Vision, Tel Aviv, Israel, 2022: 586-602.
- [55] DING R, YANG J, JIANG L, et al. Doda: data-oriented sim-to-real domain adaptation for 3d semantic segmentation[C]. European Conference on Computer Vision, Tel Aviv, Israel, 2022: 284-303.
- [56] BIAN Y, HUI L, QIAN J, et al. Unsupervised domain adaptation for point cloud semantic segmentation via graph matching[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, Kyoto, Japan, 2022: 9899-9904.
- [57] JIANG P, SARIPALLI S. Lidarnet: a boundary-aware domain adaptation model for point cloud semantic segmentation[C]. IEEE International Conference on Robotics and Automation, Xi'an, China, 2021: 2457-2464.
- [58] JARITZ M, VU T-H, DE CHARETTE R, et al. Cross-modal learning for domain adaptation in 3d semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1533-1544.
- [59] CHEN S, JIA X, HE J, et al. Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 11018-11027.
- [60] SAITO K, KIM D, SCLAROFF S, et al. Semi-supervised domain adaptation via minimax entropy[C]. IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019: 8050-8058.
- [61] YU Y-C, LIN H-T. Semi-supervised domain adaptation with source label adaptation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 2023: 24100-24109.
- [62] SALTORI C, GALASSO F, FIAMENI G, et al. Compositional semantic mix for domain adaptation in point cloud segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 14234-14247.

- [63] PRABHU V, CHANDRASEKARAN A, SAENKO K, et al. Active domain adaptation via clustering uncertainty-weighted embeddings[C]. IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 8505-8514.
- [64] XIE M, LI Y, WANG Y, et al. Learning distinctive margin toward active domain adaptation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 2022: 7993-8002.
- [65] HUANG D, LI J, CHEN W, et al. Divide and adapt: active domain adaptation via customized learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 2023: 7651-7660.
- [66] WANG F, HAN Z, ZHANG Z, et al. Mhpl: minimum happy points learning for active source free domain adaptation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 2023: 20008-20018.
- [67] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization[C]. International Conference on Learning Representations, Vancouver, Canada, 2018: 1-11.
- [68] YUN S, HAN D, OH S J, et al. Cutmix: regularization strategy to train strong classifiers with localizable features[C]. IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019: 6023-6032.
- [69] NEKRASOV A, SCHULT J, LITANY O, et al. Mix3d: out-of-context data augmentation for 3d scenes[C]. International Conference on 3d Vision, Online, 2021: 116-125.
- [70] KONG L, REN J, PAN L, et al. Lasermix for semi-supervised lidar semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 2023: 21705-21715.
- [71] AJAKAN H, GERMAIN P, LAROCHELLE H, et al. Domain-adversarial neural networks[EB/OL], 2014-12-15 [2025-02-27], <https://arxiv.org/abs/1412.4446>.
- [72] YUAN Z, ZENG W, SU Y, et al. Density-guided translator boosts synthetic-to-real unsupervised domain adaptive segmentation of 3d point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA, 2024: 23303-23312.
- [73] WANG D, SHANG Y. A new active labeling method for deep learning[C]. International Joint Conference on Neural Networks, Beijing, China, 2014: 112-119.
- [74] ROTH D, SMALL K. Margin-based active learning for structured output spaces[C]. European Conference on Machine Learning, Berlin, Germany, 2006: 413-424.
- [75] NING M, LU D, WEI D, et al. Multi-anchor active domain adaptation for semantic segmentation[C]. IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 9112-9122.
- [76] SU J-C, TSAI Y-H, SOHN K, et al. Active adversarial domain adaptation[C]. IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 2020: 739-748.

- [77] XIE B, YUAN L, LI S, et al. Active learning for domain adaptation: an energy-based approach[C]. AAAI Conference on Artificial Intelligence, Virtual, 2022: 8708-8716.
- [78] WELFORD B P. Note on a method for calculating corrected sums of squares and products[J]. Technometrics.1962, 4(3): 419-420.
- [79] WANG Y, YIN J, LI W, et al. Ssda3d: semi-supervised domain adaptation for 3d object detection from point cloud[C]. AAAI Conference on Artificial Intelligence, Washington, USA, 2023: 2707-2715.
- [80] GAO Y, LI J, ZHOU Y, et al. Optimization methods for large-scale machine learning[C]. International Computer Conference on Wavelet Active Media Technology and Information Processing, Chengdu, China, 2021: 304-308.
- [81] VU T-H, JAIN H, BUCHER M, et al. Advent: adversarial entropy minimization for domain adaptation in semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, California, USA, 2019: 2517-2526.
- [82] ZOU Y, YU Z, LIU X, et al. Confidence regularized self-training[C]. IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019: 5982-5991.
- [83] XIAO A, HUANG J, GUAN D, et al. Polarmix: a general data augmentation technique for lidar point clouds[J]. Advances in Neural Information Processing Systems.2022, 35: 11035-11048.
- [84] WANG Y, CHEN X, YOU Y, et al. Train in germany, test in the usa: making 3d object detectors generalize[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 11713-11723.
- [85] SALTORI C, OSEP A, RICCI E, et al. Walking your lidog: a journey through multiple domains for lidar semantic segmentation[C]. IEEE/CVF International Conference on Computer Vision, Paris, France, 2023: 196-206.
- [86] KIM T, KIM C. Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation[C]. European Conference on Computer Vision, Glasgow, UK, 2020: 591-607.
- [87] SCHACHTSIEK M, ROSSI S, HANNAGAN T. Class balanced dynamic acquisition for domain adaptive semantic segmentation using active learning[C]. NeurIPS 2023 Workshop on Adaptive Experimental Design and Active Learning in the Real World, New Orleans, Louisiana, United States, 2023: 1-6.

## 作者简介

### 1. 攻读学位期间的研究成果

#### 1.1 发表的学术论文和著作

[1] 第 2 作者 (导师为第 1 作者) . ACM International Conference on Multimedia(在投).

#### 1.2 申请 (授权) 专利

[1] 第 2 作者 (导师为第 1 作者) . 2025.02.18.

[2] 第 3 作者 (导师为第 1 作者) . 2023.08.04.

[3] 第 3 作者 (导师为第 1 作者) . 2023.08.04.

#### 1.3 参与的科研项目及获奖

[1] 国家自然科学基金青年科学基金项目, 2023.01.01-2025.12.31, 参与

致 谢

---

## 致 谢