



兰州大学

本科毕业论文

论文题目 (中文) 低重叠率的三维点云配准技术研究

论文题目 (英文) A Research on Registration of 3D

Point Cloud with Low Overlap

学生姓名 张艺欣

指导教师 赵志立

学 院 信息科学与工程学院

专 业 计算机科学与技术专业

年 级 2019 级

兰州大学教务处

诚信责任书

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或在网上发表的论文。

本声明的法律责任由本人承担。

论文作者签名：_____ 日 期：_____

关于毕业论文（设计）使用授权的声明

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文（设计）的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文（设计）。本人离校后发表、使用毕业论文（设计）或与该毕业论文（设计）直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文（设计）研究内容：

☒ 可以公开

☐ 不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名：_____

导师签名：_____

日 期：_____

日 期：_____

低重叠率的三维点云配准技术研究

中文摘要

点云是三维图像的主要表征形式，其中包含了丰富的数据信息，广泛应用于逆向工程、机器人导航定位等领域。受到采集角度及采集设备等因素的影响，通常点云数据只能表示一部分的三维物体或三维场景，因此需要点云配准技术通过融合多组点云数据来增强、还原三维物体和场景的点云表示。点云配准的任务目标是找到源点云和目标点云之间的变换矩阵，使得变换后的源点云尽可能与目标点云重叠，其在计算机视觉中有着广泛的应用场景。点云配准技术已经取得了相对成熟的发展，但由于评价指标的局限性等原因，目前的点云配准方法大多忽略了在低重叠率的三维点云上的配准表现。本文探究了低重叠率三维点云上的配准技术，提出了基于 Transformer 架构的低重叠率点云配准模型，从而提升点云配准性能及泛化能力。本文的主要工作有以下两点：

1. 针对低重叠率点云上的配准任务研究，搭建了基于 Transformer 架构的点云配准模型，首先通过对点云使用全卷积几何特征进行特征提取，实现数据降维和去噪，然后分别对源点云和目标点云进行位置编码并传入两个 Transformer 网络，提取点云内的局部上下文信息来丰富点云特征。同时将特征编码一同传入交叉注意力网络中进行点云之间的信息交换，获取点云重叠区域信息，进一步用点云间的信息丰富特征表示。在编码-解码网络中引入一个 EnhanceBlock 来进一步融合局部和全局信息，最后将得到的点云特征输入到解码器中获取特征描述。
2. 在 3DMatch 和 ModelNet 两个点云数据集上进行实验并对比分析模型和其他基线模型的性能优劣。为探究低重叠率点云上的配准性能，过滤一部分点云重叠率较高的数据并进一步探究模型性能与泛化能力。进行消融实验探究模型各个部件的作用以及必要性，探讨进一步提升低重叠率点云配准性能的可行方案。

关键词：点云配准；深度学习；交叉注意力机制；Transformer

A Research on Registration of 3D Point Cloud with Low Overlap

Abstract

Point Cloud is the primary representational form of 3D image, which contains rich data information and is widely used in fields such as Reverse Engineering, Robot Navigation. Due to the influence of acquisition angle and acquisition equipment, point cloud data usually only represents a part of 3D objects or 3D scenes. Therefore, point cloud registration techniques are needed to enhance and restore the point cloud representation of 3D objects and scenes by fusing multiple sets of point cloud data. The task of point cloud registration is to find the transformation matrix between source cloud and target point cloud, so that the transformed source point cloud overlaps with the target point cloud as much as possible. Point cloud registration has a wide range of applications in computer vision. Although point cloud alignment techniques have achieved relatively mature development, most of the current methods ignore the registration performance on 3D point clouds with low overlap rates due to the limitations of evaluation indicators. In this paper, we explore the registration techniques on 3D point clouds with low overlap and propose a low overlap point cloud registration model based on Transformer architecture to improve the point cloud registration performance and generalization ability. The main work of this paper has the following two points:

1. To address the registration task on point clouds with low overlap, a point cloud registration model based on the Transformer architecture is built in this study. First, feature extraction is performed on the point cloud using fully convolutional geometric features to achieve data dimensionality reduction and denoising. Then, the position encoding is applied to the source and target point clouds, and they are passed into two Transformer networks to extract the local contextual information within the point clouds to enrich the point cloud features. Simultaneously, the feature encodings are passed into a cross-attention network for information exchange between point clouds, obtaining the overlapping area information and further enriching the feature representation. Finally, we introduce an EnhanceBlock in the encoder-decoder network to further fuse local and global information, and then the obtained point cloud features are input to the decoder to obtain the feature description.
2. Experiments are conducted on two point cloud datasets, 3DMatch and ModelNet, to compare and analyze the performance of the proposed model and other baseline models. To explore the registration performance on low overlap point clouds, a portion of high overlap point cloud data is filtered out, and further investigations are conducted on the model's performance and generalization ability. An ablation study is conducted to explore the roles and

necessity of each component of the model and to investigate feasible solutions for further improving the registration performance on point clouds with low overlap.

Key Words: Point Cloud Registration; Deep Learning; Cross-attention Mechanism; Transformer

目 录

中文摘要	I
英文摘要	II
第一章 绪论	1
1.1 研究背景及意义	1
1.2 三维点云配准任务研究的发展及现状	2
1.3 本文主要研究内容	3
1.4 论文组织结构	3
第二章 相关理论知识与技术	4
2.1 深度点云配准	4
2.2 点云局部特征描述	5
2.3 Transformer 框架在三维视觉上的应用	6
2.4 本章小结	8
第三章 基于 Transformer 架构的低重叠率点云配准模型	9
3.1 任务描述	10
3.2 特征提取	10
3.3 局部 Transformer 模块	11
3.4 交叉注意力模块	12
3.5 EnhanceBlock 与解码器	13
3.6 本章小结	14
第四章 实验与分析	15
4.1 数据收集及处理	15
4.1.1 数据集描述	15
4.1.2 数据预处理	15
4.2 评价指标	16

4.2.1	3DLoMatch 评价指标.....	16
4.2.2	ModelNet 评价指标	17
4.3	实验设计与结果分析.....	18
4.3.1	实验设置	18
4.3.2	基线数据集	18
4.3.3	实验结果分析	18
4.4	消融实验.....	19
4.5	本章小结.....	20
第五章	总结与展望.....	21
5.1	工作总结.....	21
5.2	工作展望.....	21
参考文献	22
致 谢	25

第一章 绪论

1.1 研究背景及意义

计算机视觉在发展初期的主要研究数据是二维图像，而其表示的对象实体通常是三维的，以二维图像作为数据媒介对三维物体进行处理会丢失物体本身的深度特征和尺度信息。相较于二维数据，三维图像能够更加精确地反应三维物体的几何属性和空间特征，以三维图像为研究对象的三维视觉领域近年来也取得了巨大的关注和进展。点云（Point Cloud）是三维图像的主要表征形式，是海量离散、无序、无拓扑结构的三维点的集合，其中包含了丰富的信息，如三维坐标 (x,y,z) 、颜色（RGB）、激光反射强度（Intensity）等，点云数据通过三维传感器采集得到，在最大程度上保留了三维空间的原始几何信息，广泛应用于逆向工程、远程遥感、文化遗产管理等领域。随着如 LiDAR（Light Detection And Ranging）、结构光、SFM（structure from motion）系统等获取技术和硬件设备的发展，三维数据的采集难度和成本大幅下降，进一步推动了三维视觉的发展。

在三维图像的实际应用中，有滤波、特征提取、配准、识别等技术。由于数据采集过程中的遮挡、采集设备的扫描角度等因素导致了点云数据的遮挡、错位、噪声等问题^[1]，因此需要通过对多个不同采集角度或者不同采集时间的点云数据进行配准融合来还原完整点云。点云配准的任务目标是找出两点云之间的最佳刚体变换，建立点云间的点对对应关系，使得经由变换后，两点云尽可能重叠。点云配准在三维重建、三维定位、位姿估计等计算机视觉领域有着广泛的应用。

随着科技革新与发展，三维点云的数据体量正在呈指数型增长，对于点云配准模型的性能表现也有了更高的要求。目前在三维点云配准领域的研究取得了实质性的进展，深度学习也广泛应用于点云配准任务。然而目前的大部分工作只在点云数据的重叠率较高时能够达到理想表现，在点云重叠率较低的情况下则可能会出现陷入局部最优等情况，不容易给出精准的变换矩阵。在实际应用中，由于点云天然的稀疏性和采集设备的角度、性能等原因，两个点云之间的重叠区域可能很小，同时为了降低数据采集和模型训练的成本，研究如何在低重叠率点云上高效、准确地进行点云配准任务非常具有现实意义。

因此，本文从探究提升低重叠率点云配准任务的角度出发，总结相关点云配准的基础知识和技术，通过多次提取融合点云内局部信息和点云间重叠信息来针对点云重叠区域进行高效配准，并进一步在低重叠率的点云数据集上验证模型的有效性。

1.2 三维点云配准任务研究的发展及现状

三维点云配准在计算机视觉、遥感、计算机图形学等领域都有着广泛的应用,点云配准是三维模型重建过程中的关键步骤,点云配准的精度对后续的模型建立过程至关重要。目前关于点云配准任务的研究已经有了一定的发展,由探索过渡到了优化性能的阶段。将待配准的两点云命名为源点云和目标点云,点云配准的任务是找到源点云和目标点云之间的变换矩阵,使得经由矩阵变换后的源点云尽可能多地与目标点云重叠。迭代最近点算法 ICP (Iterative Closet Point) ^[2] 是点云配准的经典算法,通过找到源点云和目标点云中的对应点,针对对应点用最小二乘估计求出变换矩阵,然后进行迭代优化。传统的 ICP 算法在寻找对应点时存在着计算量大的问题,此外,初始位置的选择也可能导致模型陷入局部最优。另一个常见的配准算法,正态分布变换 NDT (Normal Distribution Transform) ^[3],通过正态分布对参考点云进行划分,计算各个区块的正态分布模型,而非直接对点云进行匹配,提升了模型对微小偏差的容忍度,但仍需要目标点云和源点云已基本对齐或者需要粗配准提供初始位姿。

随着深度学习的发展,诸多算法在深度学习的加成下性能得到了进一步提升,深度学习也在点云配准领域有了许多应用。PointNet ^[4] 使用多层感知机提取点云特征,再通过池化操作提取全局特征。PointnetLK ^[5] 利用 LK 算法 ^[6] 将循环神经网络嵌入到传统的 PointNet 架构之中,相较于传统的 ICP 及其变式省去了计算点云对应点的步骤,提升了计算效率和稳健性。FCGF ^[7] 将三维语义分割领域 ^{[8] [9]} 里提出的稀疏张量表示应用到了三维点云数据上,仅需全卷积网络的一个通道即可进行特征计算,可以快速生成高分辨率特征。D3feat ^[10] 使用基于 KPConv ^[11] 的全卷积网络,通过多任务学习来完成三维局部特征提取和描述,能够实现快速推断。SpinNet ^[12] 提出了用于三维曲面配准的具有旋转不变性、高描述性且泛化能力强的特征描述子。

目前对于点云配准任务的研究基本都是建立在点云重叠率 $\geq 30\%$ 的数据上的,然而实际上受到采集设备和角度遮挡等因素的影响,三维点云多数是低重叠率的。目前的很多模型由于对点云初始位置的要求等原因,会在低重叠率点云上出现性能下降的情况。PREDATOR ^[13] 提出通过交叉注意力机制来交换两点云之间的信息来将后续配准步骤聚焦在点云重叠区域上,从而提升低重叠率点云上的配准表现。Liu 等人 ^[14] 认为以 ICP 为代表的点云精配准方法为了避免陷入局部最优的情况,对源点云和目标点云初始位置要求较高,提出了通过聚焦于对点云重叠区域的提取和配准来解决低重叠率点云配准问题的方法。对于低重叠率点云上配准任务的性能提升,可以从关注点云间的信息交换、提取两点云重叠区域的信息等角度进行探索。

1.3 本文主要研究内容

低重叠率点云上的配准任务研究对于节约数据采集和训练成本、提升配准模型的计算效率和泛化能力都有着实际意义。因此,本文旨在探究低重叠率三维点云上的配准技术,主要内容如下:

1. 总结回顾现有的三维点云配准模型的方法和技术,分析现有模型在低重叠率点云上配准性能下降的可能原因,阐述研究低重叠率点云配准技术的必要性与重要性。
2. 学习总结三维点云配准相关理论知识和研究技术,汇总深度点云配准方向、点云局部特征描述子和 Transformer 框架在三维视觉上的应用的理论知识与技术。
3. 对点云数据进行特征提取和特征编码,搭建基于 Transformer 架构的低重叠率点云配准模型,在融合点云内上下文信息的同时,交换源点云和目标点云间的信息,以获取点云重叠区域信息,提升模型配准效率与性能。在两个点云数据集上与基线模型进行对比实验和消融实验,验证模型的可行性与有效性。

1.4 论文组织结构

本文旨在研究低重叠率三维点云上的配准任务,探究分析了现有的大部分点云配准模型在低重叠率点云数据上表现性能下降的原因,总结了低重叠率点云配准相关的技术与模型,然后搭建了基于 Transformer 架构的低重叠率三维点云配准模型,并进行对比实验验证模型的有效性。本文组织安排如下:

第一章 绪论。本章主要介绍三维点云配准的研究背景、目的、研究现状及意义,引出本文的主要研究内容。

第二章 相关理论知识与技术。本章详细介绍点云配准任务的相关概念和理论方法,主要包括:深度学习、点云局部特征描述以及 Transformer 框架在三维视觉上的应用,为后续的模型搭建奠定理论知识基础。

第三章 基于 Transformer 架构的低重叠率三维点云配准模型。基于对相关技术的学习与总结,针对低重叠率点云配准任务,本章详细阐述点云配准的任务目标并详细展开了 Transformer 架构的点云配准模型的结构。

第四章 实验及结果分析。本章选取两个点云数据集并进行了数据预处理,针对不同的数据集选取相应评价指标并进行了数理推导,介绍与本文模型进行对比试验的基线模型,详细阐述实验设置及流程,进行消融实验以验证本文模型各网络组件的有效性,并对实验结果进行了分析与总结。

第五章 总结与展望。本章是对本文的主要工作的总结概括。再次阐述低重叠率点云配准技术研究的现实意义与可行方案,总结文章不足并对未来工作做出展望。

第二章 相关理论知识与技术

2.1 深度点云配准

点云是目前表示三维图像的主要形式,广泛应用于逆向工程、机器人导航定位等领域。点云处理可以大致分为三个层次,低层次的处理方式包括滤波、关键点、边缘检测等,中层次则有特征描述、点云分割及分类等处理,高层次的处理则有点云配准和三维重建等应用,通常对于点云数据的处理需要多个层次的协作完成。点云配准在三维目标检测、三维重建、三维定位以及位姿估计等应用上都发挥着至关重要的作用。三维点云数据是通过三维传感器采集得到的,随着传感器不断的更新迭代,点云数据的精度也逐渐提升。由于在采集数据时传感器与扫描仪只能捕捉部分的点云数据,因此需要配准算法来根据有限的点云数据生成更大范围的场景。通过点云配准技术,可以将关于同一个三维场景或物体的多个部分点云融合为一个完整的三维点云。点云配准的任务目标是估计两点云之间的变换矩阵,通过找出一个能够将相似性指标的误差最小化的最佳变换来将两点云对齐。

点云配准分为粗配准(Coarse Registration)和精配准(Fine Registration)两个阶段^[15]。粗配准可以在两个点云之间的变换矩阵完全未知的情况下更快速地完成配准工作,并为精配准提供变换初值。经典的粗配准算法 RANSAC (RANdom SAMple Consensus, 采样一致性)^[16] 是一种随机参数估计算法,其基本思想为迭代选择数据用来估计模型,直到得到较好的模型。具体而言, RANSAC 通过从两点云内随机抽样两个点并对其计算配准模型参数,然后用所有点云数据计算该模型的偏差,若偏差小于模型偏差阈值则认为选取的点为内点(Inlier Points),否则则视为外点,通过迭代选取对应内点最多的模型为最佳配准模型。而精配准则侧重于模型之间的精细配准,进一步降低误差。迭代最近点算法 ICP(iterative closest point)^[2] 是一种最经典且广泛应用的点云配准算法, ICP 将点云配准问题分为两个子问题,即寻找最近点和最优变换,通过迭代进行点云之间的对应点搜寻以及最小化对应点距离来估计变换矩阵。

深度学习在诸多计算机视觉任务上都取得了成功,但由于三维数据的稀疏性、无序排列等特质给深度学习在三维点云配准任务上的应用带来了一定的困难。PointNet^[4] 将多层感知器应用到每个点上来提取特征,然后通过池化操作提取全局表示。由于 PointNet 没有捕捉点云周围的局部结构, PointNet 无法处理一些几何细节,因此 PointNet++^[17] 通过将 PointNet 递归应用到点云上来提取点云的局部信息。PointNetLK^[5] 则将 PointNet 看作成像函数,利用 Lucas/Kanade 优化法^[6] 的反向组合形式对模版图像进行逆变换得到原图像。DGCNN^[18] 基于卷积神经网络,在每层特征空间上重新使用邻近算法计算采样点的邻域并更新图模型。DCP^[19] 利用 Transformer 中的 Attention 机制^[20] 计算得到一个假想的目标点

云, 通过损失函数约束使得其与真正的目标点云不断逼近, 最终实现点云配准。FCGF^[7] 通过将卷积操作替换为全卷积网络, 将三维卷积应用到整个输入上, 这种全卷积预处理^[21] 可以达到更佳快速高效、内存效率更高的效果。PREDATOR^[13] 基于编码器-解码器(编码-解码)的模型架构, 通过两个独立的图神经网络分别对源点云和目标点云进行局部上下文信息的融合, 然后再使用交叉注意力机制^[22] 基于消息传递机制进行两点云之间的信息交换, 最后再次使用图神经网络来整合点云内部的上下文信息与点云间的交换信息, 从而实现对低重叠率点云的重叠区域的信息充分利用。

2.2 点云局部特征描述

绝大部分的基于点对点关系的点云配准模型依赖基于点云局部特征描述和特征匹配建立的点对点匹配^[23]。由于待配准点云仅具有一部分重叠区域, 需要 RANSAC 等算法剔除误匹配并从匹配子集中计算变换矩阵。其中点云局部特征描述和特征匹配是关键的两步。点云局部特征描述用于对特征点的局部曲面内的空间信息和几何信息进行编码^[24], 对三维点云模型的质量与分辨率有着关键的影响。特征匹配则是通过度量两组点云之间任意两个特征的相似度来确定初始对应关系并筛选出正确的匹配子集。点云局部特征描述子根据特征提取的方式不同主要分为两类, 即基于手工设计(hand-crafted)和基于深度学习的描述子。基于手工设计的局部描述子通过使用人为定义的特征来描述局部几何信息, 常见的方法有 FPFH^[25], RoPS^[26], SHOT^[27], 和 RoPS^[26]等。然而这类基于手工设计的描述子普遍存在难以处理分辨率变化、遮挡和杂乱以及噪声数据等情况。

随着深度学习方法的广泛应用, 出现了许多基于深度学习的局部描述子。其中, 3DMatch^[28] 开创性地提供了一个利用 Siamese 3D CNN^[29] 在三维数据上计算局部描述子的方法。PPFNet^[30] 通过 Point-Net 结构^[4] 来结合点对特征, 从而实现了局部特征提取, 提取的特征具有旋转不变性且能够融合其他点集的局部特征。PPF-FoldNet^[31] 结合了 PointNet 的编码器与 FoldingNet 的解码器, 解决了 PPFNet 对旋转操作敏感的问题。FCGF^[7] 的数据表示和卷积操作主要基于 MinkowskiEngine^[32] (支持稀疏卷积的自动微分库), 在编码和解码之间通过跳跃连接来解决梯度爆炸和梯度消失的问题, 增强对稀疏卷积特征的表达能力。D3Feat^[10] 基于 KPCov^[11] 的全卷积网络和联合优化框架, 通过多任务学习对每个点同时输出显著性得分和局部描述子, 提出了一种新的密度不变关键点选择策略。

表 2.1 点云局部特征描述子的经典方法汇总

	名称	主要贡献
基于手工设计的局部描述子	SHOT ^[27]	通过对局部球域进行空间划分, 然后对每个子空间计算特征以及特征对应的串联向量。
	RoPS ^[26]	通过将局部曲面在局部坐标系中不断旋转并统计投影后点密度, 然后将点密度量串接得到点云特征。
基于机器学习的局部描述子	FCGF ^[7]	数据表示和卷积操作主要基于 MinkowskiEngine ^[32] , 编码解码之间通过跳跃连接来增强特征的表达能力。
	PPFNet ^[30]	实现提取具有旋转不变性的局部特征, 构建单个点集的特征时可以融合其他点集的局部特征。
	D3Feat ^[10]	使用基于 KPConv 的全卷积网络和联合优化框架进行局部特征描述; 提出密度不变关键点选择策略。

2.3 Transformer 框架在三维视觉上的应用

在自注意力机制中, 输入序列 X , 对其进行线性变换得到查询 Q 、键 K 和值 V , 通过计算 Q 和 K 的相似度得到注意力权重, 再基于权重对 V 进行加权求和运算。每个查询都会关注所有的键值对并且生成一个注意力输出。公式2.1为注意力机制的具体运算过程, 首先对 Q 和 K 进行点积运算, 使用键 K 的维度 d_k 进行缩放后再进行 softmax 运算, 最后再与 V 求内积。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.1)$$

由于 Q 和 K 来自同一序列 X 因此称为自注意力 (self-attention) 机制, 反映的是序列 X 对自身的关注。另一种交叉注意力 (cross-attention) 机制的输入则是两个维度不同的序列, 一个序列用作 Q 输入, 另一个序列用作 K 和 V 输入, 以捕捉两个序列之间的依赖关系。

Transformer^[20] 是目前自然语言处理 (NLP) 领域的主流架构, 广泛应用于语音识别、问答系统、机器翻译等方向。Transformer 是一个编码-解码 (encoder-decoder) 的网络结构, 并通过位置编码 (positional encoding) 对数据进行预处理。其中编码由一个多头注意力 (Multi-head Attention) 层和一个前馈神经网络组成, 如图2.1所示, 多头注意力层使用多个自注意力机制学习源语句的多种内部关系, 再通过前馈神经网络将编码传给解码器。而解码器将这些编码与其包含的上下文信息相融合来生成输出序列。解码器由两个多头注意力层和一个前馈神经网络组成, 第一个多头注意力层通过自注意力机制学习目标语句的内部关系, 输出结果与编码的输出一同传入第二个多头注意力层, 学习源语句和目标语句之间的关系,

再经由一个前馈神经网络得到输出。**Transformer** 的经典模型 **BERT** [33] 在考虑每个词的上下文信息的同时对未标记的文本进行预训练, 采用双向 **Transformer** 组件, 生成深度的双向语言表征。

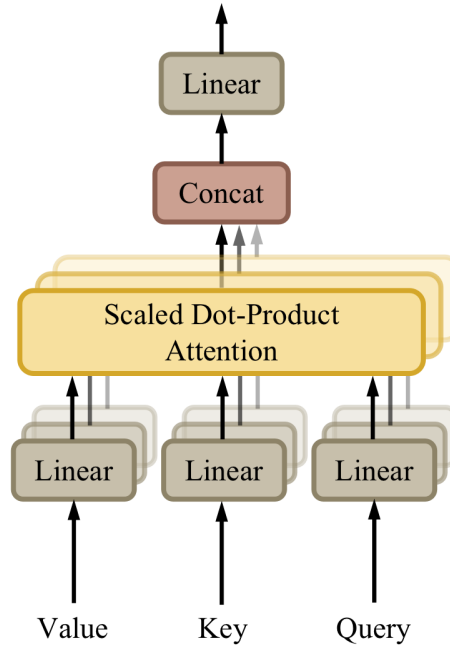


图 2.1 多头注意力机制

受 **Transformer** 在 **NLP** 领域成功应用的启发, 计算机视觉领域里的诸多任务也进行了对 **Transformer** 架构的迁移与应用, 如自动驾驶, 工业检测, 可视计算等 [34] [35]。**Transformer** 在 **NLP** 领域上的应用是对源语句和目标语句进行位置编码与处理后输入到模型中, 学习语句内的词语的内部关系和源语句与目标语句之间的关系。在自然语言处理任务中, 单词在语句中的排列顺序对语句和词语的含义有着至关重要。由于 **Transformer** 不具有 **RNN** 主动学习语序的能力, 因此为了消除输入顺序对于数据排列顺序及含义的影响, **Transformer** 提出位置编码 (**positional encoding**) 给不同位置的词赋予唯一的编码, 从而建模文本中的顺序关系。位置编码的方式可分为绝对位置编码和相对位置编码, 前者是给不同位置随机初始化一个编码后输入模型进行训练, 而后者是通过三角函数 *sine*, *consine* 等方式来进行编码 [20] [36]。

将 **Transformer** 架构从 **NLP** 领域迁移到三维点云配准任务中, 首先可以进行类比迁移, 具体来说, 即将点云对应到语句, 点云内的点对应语句中的词语。实际上, **Transformer** 架构非常适用于处理三维点云数据, 因为其核心的自注意力 (**self-attention**) 操作本质上是一个集合操作, 对于输入具有序列不变性和基数不变性 [20], 而点云数据本质上就是位置信息的集合 [36]。相较于传统点云配准任务中常用到的卷积神经网络, **Transformer** 更加擅长于捕捉长期依赖 [35]。此外与多层感知器 (**MLP**) 相比, 在推断过程中 **Transformer** 对于输入的依赖更加灵活, 并且 **Transformer** 具有置换不变性, 而在 **MLP** 中权重矩阵中编码了

输入输出的顺序。**Transformer** 架构在三维视觉领域已经有了一定的应用工作。**PT**^[37] 采用了 **PointNet++**^[17] 的分层结构进行点云分类和分割,在此基础上,**PT** 还将 **PointNet++** 中的 **shared MLP** 模块替换为了局部 **Transformer** 模块,用于处理点云局部信息。**Pointformer**^[38] 通过 **Transformer** 模块提取并结合局部特征和全局特征,对于输入的点云具有置换不变性。**Pointformer** 通过两个 **Transformer** 模块分别提取局部点云特征和全局上下文信息,最后再使用一个新的交叉注意力模块将局部特征整合到全局信息当中。

2.4 本章小结

本章内容为点云配准相关技术和理论基础的介绍。本章首先回顾并汇总了深度学习在点云配准上的应用,并具体地介绍了描述点云局部特征的相关方法,此外还介绍了 **Transformer** 架构的基本原理以及其在三维视觉领域的应用。通过总结相关工作,介绍了点云配准的相关知识和技术,为后续的模型搭建环节奠定了理论基础。

第三章 基于 Transformer 架构的低重叠率点云配准模型

点云是三维图像的主要表现形式，其中包含了三维坐标、分类值、颜色等丰富的信息。给定两个三维点云，点云配准的任务是找出源点云和目标点云之间的变换矩阵，使得变换后的源点云与目标点云尽可能重叠。点云配准领域目前已经有了许多研究并取得了优异的配准表现，但由于评价指标的限制，目前的点云配准方法大多忽略了在低重叠率的点云上的配准表现。在实际的数据收集过程中，获取高重叠率的点云的成本很高，因此提升模型在低重叠率点云上的配准性能将显著降低点云数据采集的成本并更加充分利用点云信息。本文建立了基于 Transformer 架构的点云配准模型来提升在低重叠率点云上的配准性能。

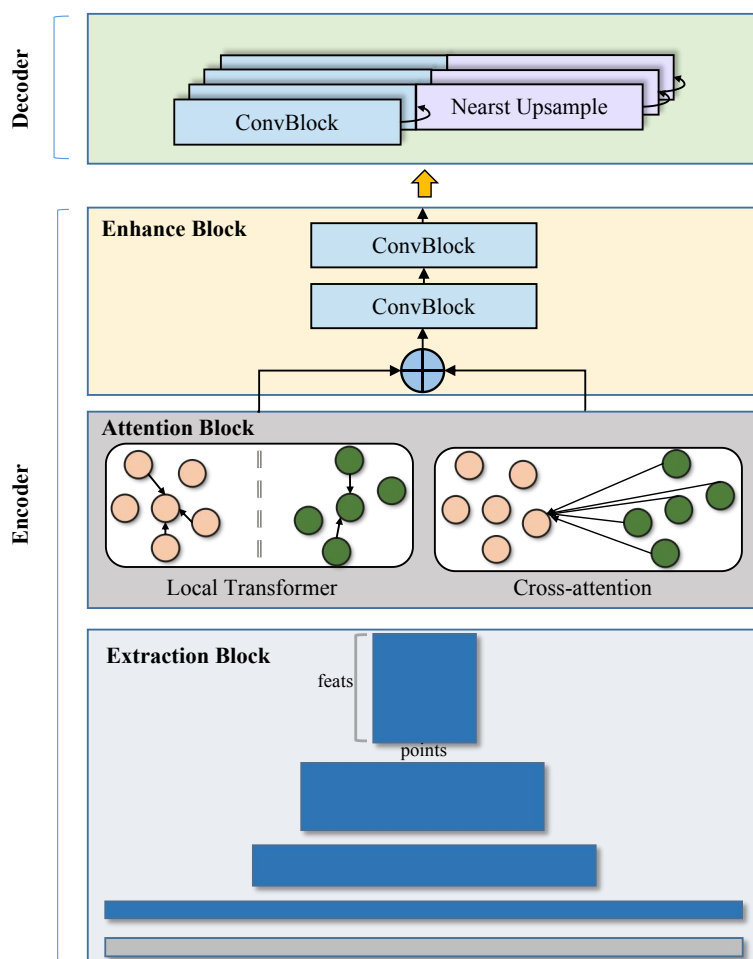


图 3.1 基于 Transformer 架构的低重叠率点云配准模型

在本章中，针对低重叠率点云配准问题，搭建了基于 Transformer 架构的点云配准模型。模型基于编码-解码网络架构，首先使用 KPConv 网络对点云数据进行特征提取，降低数据维度并方便后续特征匹配的展开，提升计算效率，然后通过两个独立的局部 Transformer 架

构对源点云和目标点云分别进行点云内部上下文信息的融合, 增强特征表示, 同时通过交叉注意力机制来交换点云间的信息, 获取点云重叠区域信息。在注意力模块前加入 **DGCNN** 模块来进行特征提取并保留原始特征, 将注意力模块得到的特征表示与原始特征进行拼接融合以丰富特征空间, 通过关注于点云重叠区域的配准来提升配准效率从而提升低重叠率点云上的配准表现。模型在编码-解码网络中间加入一个 **EnhanceBlock** 来进一步融合编码中提取到的点云局部信息和点云间的交互信息, 然后输入到解码器中进行上采样还原特征并获取特征描述子。

3.1 任务描述

给定两个低重叠率的点云 \mathbf{P} 和 \mathbf{Q} , 其中源点云 $\mathbf{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid i = 1 \dots N\}$, 目标点云 $\mathbf{Q} = \{\mathbf{q}_i \in \mathbb{R}^3 \mid i = 1 \dots N\}$, 模型目标是找出变换矩阵 $T_{\mathbf{P}}^{\mathbf{Q}}$ 来将源点云 \mathbf{P} 映射到目标点云 \mathbf{Q} 上并且使得二者尽可能重叠。模型通过判断经过矩阵 $T_{\mathbf{P}}^{\mathbf{Q}}$ 变换后, 点云 \mathbf{P} 和 \mathbf{Q} 间的重叠率 $r_{overlap}$ 是否达到设定的阈值来训练 $T_{\mathbf{P}}^{\mathbf{Q}}$, 其中重叠率 $r_{overlap}$ 通过计算变换后的源点云 \mathbf{P} 上与目标点云 \mathbf{Q} 中的对应点足够相近的点的个数占总点云数量的比例得出, 经过变换后的两点云上的重叠点对个数越多, 重叠率 $r_{overlap}$ 越高, 模型配准效果越好。由于模型背景是在低重叠率点云上配准任务的探索, 点云 \mathbf{P} 与 \mathbf{Q} 原本就是低重叠率的, 因此重叠率阈值不会设置过高。

3.2 特征提取

点云特征按照空间尺度分类可以分为单点特征, 局部特征和全局特征。其中单点特征主要包括点云本身的三维坐标等信息, 局部特征则包含局部坐标系内一系列点的信息, 而全局特征是指整个点云的信息。特征提取是从原始数据中选出一部分具有代表意义的特征, 对每个点提供相关的局部特征和全局特征, 从而来实现数据降维并方便特征匹配和后续处理。由于点云数据具有稀疏性, 因此通过特征提取可以显著降低数据量并提升计算效率。作为后续点云配准的第一步, 特征提取工作非常重要。给定一个点云 $\mathbf{P} \in \mathbb{R}^{N \times 3}$, 特征提取模型对点云中的每个点 $\mathbf{p}_i = [x_i, y_i, z_i] \in \mathbb{R}^3$ 提取特征 $\mathbf{x}_i \in \mathbb{R}^C$, 并尽可能地保留其在原始点云空间中相邻的上下文信息, 同理对点云 $\mathbf{Q} \in \mathbb{R}^{N \times 3}$ 上的每个点 $\mathbf{q}_j = [x_j, y_j, z_j] \in \mathbb{R}^3$ 提取特征 $\mathbf{y}_j \in \mathbb{R}^C$ 。在点云配准任务中, 特征提取模型则需要保证经过变换后的点尽可能接近另一个点云中的对应点。常见的特征提取模型可以分为 **hand-crafted** 和基于深度学习的两种, 前者存在着难以处理分辨率变化以及噪声数据等情况, 后者则在计算速度和准确度上更胜一筹。因此本文选取其中一种基于深度学习的特征提取模型, 全卷积几何特征 (**FCGF**^[7]) 来进行点云的特征提取。**FCGF** 采用稀疏卷积来构建深度网络, 将输入的点云体素化为离散的三维网络, 采用 **Minkowski** 卷积^[32] 代替传统卷积, 当体素分辨率较高时, 通常会有几个具有不同标签/颜色的点落在同一个体素中, **MinkowskiEngine** 随机选择一个点并将其标

签分配给该体素，提升了内存效率。

3.3 局部 Transformer 模块

为了分别融合源点云和目标点云内部的局部上下文信息，通过邻居结点来丰富点云特征信息，从而进一步提升点云配准效果，本文引入 Transformer^[20] 架构来融合点云局部特征。为了消除输入顺序对数据含义的影响，Transformer 首先对输入数据进行位置编码来建模数据的顺序关系。由于三维点云的坐标直接表示了点云的位置信息，因此可以直接使用点云坐标信息来进行相对位置编码，同时提升 Transformer 模块对于点云局部信息的学习表现^[37]。

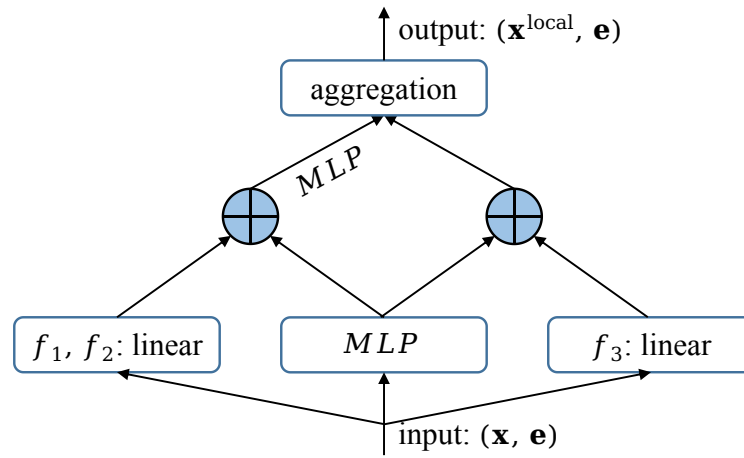


图 3.2 局部 Transformer 网络训练流程

以点云 \mathbf{P} 为例（点云 \mathbf{Q} 同理），对于点云 \mathbf{P} 上的每个点 \mathbf{p}_i ，本文首先通过 K 邻近算法选取距离点 \mathbf{p}_i 最近的 k 个邻居点 \mathbf{p}_j ， $j: (i, j) \in \mathcal{E}$ ，点 \mathbf{p}_i 和邻居点 \mathbf{p}_j 的三维坐标分别为 $\mathbf{e}_i, \mathbf{e}_j \in \mathbb{R}^3$ 。首先对点 \mathbf{p}_i 进行相对位置编码：

$$\delta_i = \text{MLP}(\mathbf{e}_i - \mathbf{e}_j) \quad (3.1)$$

为了融合点云内的局部上下文信息，本文分别对源点云 \mathbf{P} 和目标点云 \mathbf{Q} 建立两个结构相同的 Transformer 网络，以源点云 \mathbf{P} 为例（点云 \mathbf{Q} 的网络训练流程类似），点 \mathbf{p}_i 和邻居点 \mathbf{p}_j 在特征提取环节 3.2 中得到的特征编码为 \mathbf{x}_i 和 \mathbf{x}_j ，点 \mathbf{p}_i 的局部特征融合流程为：

$$\mathbf{x}_i^{\text{local}} = \sum_{j: (i, j) \in \mathcal{E}} \langle \text{softmax}(\text{MLP}(f_1(\mathbf{x}_i) - f_2(\mathbf{x}_j) + \delta_i)), (f_3(\mathbf{x}_j) + \delta_i) \rangle \quad (3.2)$$

这里 \langle, \rangle 表示内积操作； f_1, f_2, f_3 分别为三个线性映射函数；MLP 层使用了两个线性层和 ReLU 激活函数； δ_i 为公式 3.1 中得到的点 \mathbf{p}_i 的相对位置编码。

3.4 交叉注意力模块

为了找出两点云的重叠区域，需要交换两点云之间的信息，因为仅对一个点云单独进行训练无法得到另一个点云上的信息，特征 \mathbf{x}^P 是不包含关于目标点云 \mathbf{Q} 的信息，同样的，特征 \mathbf{x}^Q 也不包含源点云 \mathbf{P} 的信息。为此，本文采用了一个基于消息传递机制^[39]的交叉注意力（Cross-attention）^[22]模块，交叉注意力机制将两个不同的序列不对称地组合在一起，其中一个序列用作查询输入，定义输出的序列长度，另一个序列用作键和值输入。

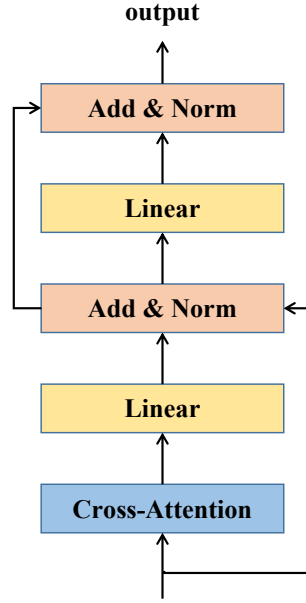


图 3.3 交叉注意力模块

以 $\mathbf{P} \leftarrow \mathbf{Q}$ 方向上的消息传递为例，同理 $\mathbf{P} \rightarrow \mathbf{Q}$ 方向上的消息传递是类似的。对于点云 \mathbf{P} 上的每个点 \mathbf{p}_i ，其在特征提取环节3.2中得到的特征编码为 \mathbf{x}_i ，点云 \mathbf{Q} 上的每个点 \mathbf{q}_j 的特征编码为 \mathbf{y}_j 。首先，点云 \mathbf{P} 中的每个点与 \mathbf{Q} 中的对应点相连形成一个二部图。对点云 \mathbf{P} 上的每个点 \mathbf{p}_i ，使用点云 \mathbf{Q} 上与 \mathbf{p}_i 相关联的所有点 \mathbf{q}_j 来对 \mathbf{p}_i 进行更新，从而实现两点云之间的信息交流。

$$\mathbf{m}_{i \leftarrow} = \sum_{j: (i,j) \in \mathcal{E}} \text{softmax} \frac{\mathbf{s}_i^\top \mathbf{k}_j}{\sqrt{b}} \mathbf{v}_j \quad (3.3)$$

这里将点 \mathbf{p}_i 的查询 $\mathbf{s}_i \in \mathbb{R}^b$ 和 \mathbf{q}_j 的键 $\mathbf{k}_j \in \mathbb{R}^b$ 和值 $\mathbf{v}_j \in \mathbb{R}^b$ 传入交叉注意力网络来对点 \mathbf{p}_i 进行更新。

$$\mathbf{x}_i^{\text{cross}} = \mathbf{x}_i + \text{MLP}(\text{cat}[\mathbf{x}_i, \mathbf{m}_{i \leftarrow}]) \quad (3.4)$$

这里 $\text{MLP}(\cdot)$ 为含有实例正则化和 ReLU 激活函数的三层感知机； $\mathbf{x}_i^{\text{cross}}$ 为融合了两点云重叠信息的点 \mathbf{p}_i 特征编码，同理点 \mathbf{q}_i 的特征编码 $\mathbf{y}_i^{\text{cross}}$ 的计算流程类似。通过局部 Transformer 和交叉注意力网络，模型得到了融合了局部上下文和点云重叠信息的特征编码 \mathbf{F}^P 、 \mathbf{F}^Q ，特征编码经过线性映射可以得到两个重叠度得分 $\mathbf{o}^P, \mathbf{o}^Q$ ，反映了某个点落在两点云重叠区域

上的概率。此外，点 \mathbf{p}_i 的交叉重叠得分，即其在点云 \mathbf{Q} 上的对应点，位于重叠区域的概率可以如下计算得到：

$$\hat{o}_i^{\mathbf{P}} := \mathbf{w}_i^T \mathbf{o}^{\mathbf{Q}}, \quad w_{ij} := \text{softmax} \left(\frac{1}{t} \langle \mathbf{f}_i^{\mathbf{P}}, \mathbf{f}_j^{\mathbf{Q}} \rangle \right) \quad (3.5)$$

此外，在注意力模块中，为了加快计算速度，减少计算量，并引入更加丰富的特征空间，本文使用 CSPNet^[40] 架构对注意力模块进行改进。本文首先对特征提取模块输出的特征进行特征降维操作，为了使得点云中的信息更好的保留，本文引入了 DGCNN^[18] 模块，根据点云中点的坐标对其进行卷积操作，得到降维后的特征，降维后的特征一方面作为注意力模块的输入，另一方面将会被保留，在输入至解码器前，经过注意力计算的特征将会与原保留的特征合并，以此保留两个不同的特征图，获取更加丰富的特征空间。

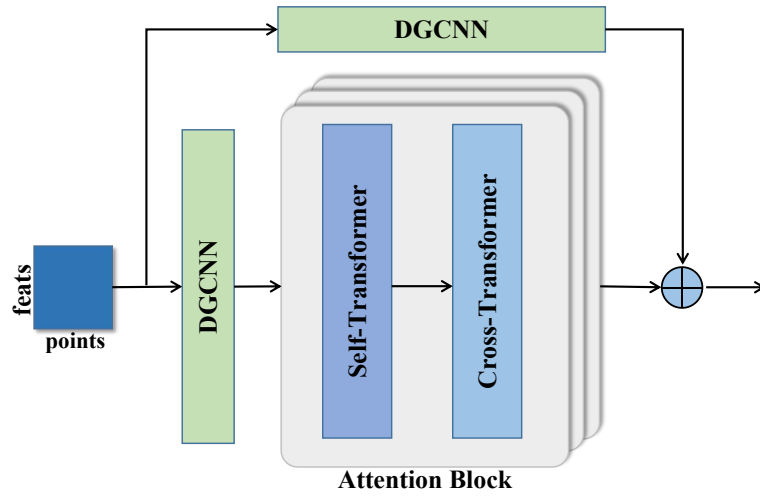


图 3.4 DGCNN 模块丰富特征空间

3.5 EnhanceBlock 与解码器

模型在编码-解码网络中间引入一个 EnhanceBlock^[41]，用于汇总并增强上一步编码器中局部 Transformer 网络和交叉注意力网络得到的特征，以便更好地融合局部信息和全局特征。单独对两点云分别进行训练无法交换点云之间的信息，而获取点云重叠区域需要同时用到两点云的信息，并且进行点云配准也需要综合考虑源点云和目标点云的信息。因此，EnhanceBlock 通过进一步融合局部信息和点云重叠信息来丰富点云特征，首先将从两个局部 Transformer 网络得到的点云特征 $\mathbf{X}^{\text{local}}, \mathbf{Y}^{\text{local}}$ 进行拼接，同时拼接从两个交叉注意力网络得到的点云特征 $\mathbf{X}^{\text{cross}}, \mathbf{Y}^{\text{cross}}$ ，然后将拼接后的点云特征输入到两个依次相连的卷积神经网络中，并通过激活函数 ReLU 映射给解码器。

在训练深度神经网络时，模型的性能可能会随着网络深度的增加而下降，这种现象称为退化问题。导致退化问题的原因可能为模型过拟合，还有另外一种原因是梯度消失和梯度爆炸问题。跳跃连接（Skip Connections）在训练时会跳过神经网络中的某些层，直接将

一层的输出作为下一层的输入。跳跃连接缓解了非线性变化难以收敛的问题，提升了深度神经网络的收敛性。**Transformer** 在模型内就包含了跳跃连接和层标准化操作，跳跃连接也可以用于编码-解码结构之间的跨网络连接 [42]。

解码部分将上采样与 4 个 **PointNet-style**^[4] 的 **MLP** 层相结合，并对相应的编码层进行跳跃连接。上采样使用最邻近插值，即使用上一层网络中最邻近点的相关特征进行上采样。通过将 $\mathbf{F}^{\mathbf{P}}$ 与重叠得分 $\mathbf{o}^{\mathbf{P}}, \delta^{\mathbf{P}}$ 拼接后得到特征描述子 $\mathbf{F}^{\mathbf{P}} \in \mathbb{R}^{N \times 32}$ 和新的重叠得分和匹配度 $\mathbf{o}^{\mathbf{P}}, \mathbf{m}^{\mathbf{P}} \in \mathbb{R}^N$ 。重叠得分 $\mathbf{o}^{\mathbf{P}}$ 反映了经过矩阵变换后的两点云的重叠部分占整个点云的比例， $\mathbf{o}^{\mathbf{P}}$ 值越高，两点云重叠部分越大，重叠率越高。匹配度 $\mathbf{m}^{\mathbf{P}} \in \mathbb{R}^N$ 则反映了给定某一点云中的点，另一点云中的点匹配正确的可能性。

3.6 本章小结

本章在点云配准的理论基础之上，针对低重叠率点云配准问题，提出了通过关注局部信息和点云重叠区域的配准来提升配准效率。具体流程为在使用交叉注意力机制关注点云间的信息交换来获取点云重叠信息的同时，分别对源点云和目标点云建立局部 **Transformer** 网络提取上下文信息。在注意力模块前引入 **DGCNN** 模块来保留原始特征并丰富特征空间，然后通过 **EnhanceBlock** 来进一步融合局部信息和全局信息，最后解码输出点云特征。本章对模型框架及训练流程进行了详细的阐述，为后续实验环节的展开奠定了理论基础。

第四章 实验与分析

为了评估模型性能并探究其在低重叠率点云上的配准性能，本文选取了两个点云数据集 3DMatch^[28] 和 ModelNet40^[43] 并为其分别选取了三个不同的评价指标来对模型性能进行评估，同时本文选出了 FCGF、D3Feat、RPM-Net、PREDATOR 等四个基线模型进行对比。此外，本文还进行了消融实验来验证模型的每个模块的有效性。实验结果证明，本文的模型在点云配准任务尤其是低重叠率的点云配准任务上的表现良好。

4.1 数据收集及处理

4.1.1 数据集描述

本文选用 3DMatch^[28] 和 ModelNet40^[43] 两个数据集来评估模型性能，首先整理了两个数据集的相关基础信息，然后对数据分别进行了预处理，以便后续点云配准工作的开展以及模型在低重叠率点云上配准性能的探究。

3DMatch 数据集中包含了 62 个场景的数据，其中包含 54 个训练场景和 8 个测试场景。官方的 3DMatch 数据集只考虑了重叠率大于 30% 的数据，为了探究低重叠率点云上的模型配准性能，本文参照 PREDATOR^[13]，选取重叠率在 10% 到 30% 之间的数据并命名为 3DLoMatch，依旧保存 54 个训练场景和 8 个测试场景的设置。

ModelNet40 数据集中包含 40 个不同对象类别（如飞机，汽车，沙发等）的 12,311 个 CAD 模型。数据分为 9843 个训练样本和 2468 个测试样本。参照 Rpm-Net^[44]，为了验证模型对不同类别的物体的泛化能力，本文分别从不同类别数据选取训练和测试样本，从前 20 个类别中选取 5,112 个训练样本和 1,202 个验证样本，再从后 20 个类别中选取 1,266 个测试样本。从前 20 个类别中选取训练集和验证集，再从后 20 个类别中选取测试集，样本数量如表格4.1所示。

表 4.1 ModelNet40 数据集采样数量

	训练集	验证集	测试集
样本数量	5,112	1,202	1,266

4.1.2 数据预处理

1. 3DLoMatch 数据预处理

点云数据中存在着许多噪声，会影响后续的数据处理和模型配准性能，因此在正

式进行特征提取前需要用滤波器对点云数据进行去噪。我们选取点云体素降采样法 (Voxel Grid Downsampling) 对 3DLoMatch 数据集进行降噪处理, 每个体素空间保留一个点, 如果多个点同时落在一个体素中则随机保留一个。

2. ModelNet40 数据预处理

首先我们对每个 CAD 模型均匀采样 2048 个点并将他们缩放到一个球面上。然后参照 Rpm-net^[44], 对于源点云, 均匀采样一个穿过原点的平面将球面分成两部分, 将平面沿着法向量方向旋转直到有 $[2048 \cdot p_v]$ 个点在同一边, 然后舍弃掉另外一边的所有点。对目标点云同样操作。然后对两点云进行随机的旋转、平移操作和高斯噪声抖动。随机从 $[0, 45^\circ]$ 选取旋转角度, 在范围 $[-0.5, 0.5]$ 内选取平移系数。最终去掉了 30% 的点, 得到了平均重叠率为 73.5% 的点云数据。

4.2 评价指标

为了与不同的基线模型评价方式保持一致, 本文分别为 3DLoMatch 和 ModelNet 两个数据集各自选取了三个不同的评价指标。对于 3DLoMatch 数据集, 本文选取了配准召回率、内点率和特征匹配召回率作为评价指标, 而在 ModelNet 上的模型比较则选择了倒角距离, 相对平移误差和相对旋转误差三个评价指标。

4.2.1 3DLoMatch 评价指标

参照 3DMatch^[28] 数据集的原始评价指标, 选取三个评价指标来比较模型与其他基线在 3DLoMatch 数据集上的表现性能, 评价指标包含配准召回率 (Registration Recall, RR)、内点率 (Inlier Ratio, IR) 以及特征匹配召回率 (Feature Match Recall, FMR)。

1. 配准召回率 RR

配准召回率 RR 衡量的是模型预测的位姿变换在真实匹配对上的表现质量。给定误差允许阈值, RR 值越高, 说明预测的变换在越多的真实匹配对上估计正确, 模型预测表现越好。

$$RR = E_{RMSE} = \sqrt{\frac{1}{|\Omega^*|} \sum_{(p,q) \in \Omega^*} \|T_p^Q(p) - q\|_2^2} \quad (4.1)$$

这里 T_p^Q 为预测的位姿变换, (p, q) 为真实点对关系 Ω^* 中的一组匹配对。RR 计算的是所有 $RMSE < 0.2$ 的匹配对占全部匹配对的比例。

2. 内点率 IR

RANSAC 算法中假设数据样本中包含两种数据, 一种称为内点 (inliers) 是可以被模型描述的, 另一种为异常数据 (outliers) 可能是由错误的数据采集、计算造成的, 无法适应模型。内点率 IR 则是指在给定误差范围内正确匹配的点对 (p, q) 数量所占

有匹配 Ω 的比例。

$$IR = \frac{1}{|\Omega|} \sum_{(p,q) \in \Omega} \left[\left\| \bar{T}_P^Q(p) - q \right\|_2 < \tau_1 \right] \quad (4.2)$$

这里 $[\cdot]$ 为艾佛森括号，满足括号内的条件值为 1，不满足则为 0；本文设置 $\tau_1 = 10\text{cm}$ 用来判断 (p, q) 是否匹配正确，即距离真值变换 \bar{T}_P^Q 后的位置不超过 10cm 。

3. 特征匹配召回率 FMR

特征匹配召回率 FMR 量化了给定内点率阈值，正确选取点云对的比例，这里的正确是指选取的点云对可以经由合适的变换匹配成功（通过内点率 IR 是否达到阈值进行判断）。

$$FMR = \frac{1}{M} \sum_{s=1}^M [IR > \tau_2] \quad (4.3)$$

本文设置 $\tau_2 = 0.05$ ，即当某组点云对的内点率 $IR > 0.05$ 时才被判断为选择正确； $[\cdot]$ 为艾佛森括号，满足括号内的条件值为 1，不满足则为 0。

4.2.2 ModelNet 评价指标

为了与基线模型的评价指标保持一致，本文选用三个评价指标来对模型在 ModelNet 上的表现性能进行评估比较，分别为倒角距离（Chamfer Distance, CD），相对平移误差（Relative Transformation Error, RTE）和相对旋转误差（Relative Rotation Error, RRE）。

1. 倒角距离 CD:

倒角距离 CD 衡量了两点云的配准效果，通过计算源点云 P 到目标点云 Q 中每个点的距离并选取最小距离来反应两组点云间的区别，倒角距离越大，源点云和目标点云的区别越大，点云配准效果越差。本文参照 Rpm-Net^[44] 对原始倒角距离的公式进行的修改，如公式4.4所示。

$$\tilde{CD}(P, Q) = \frac{1}{|P|} \sum_{p \in P} \min_{q \in Q_{\text{clean}}} \left\| T_P^Q(p) - q \right\|_2^2 + \frac{1}{|Q|} \sum_{q \in Q} \min_{p \in P_{\text{clean}}} \left\| q - T_P^Q(p) \right\|_2^2 \quad (4.4)$$

这里本文参照 Rpm-Net^[44]，从 ModelNet40 数据集中分别对源点云和目标点云都采样 1024 个点，以此保证源点云 P 上的每个点都和目标点云 Q 上的点一一对应。

2. 相对平移误差和相对旋转误差 RTE, RRE:

本文使用相对平移误差（Relative Transformation Error, RTE）和相对旋转误差（Relative Rotation Error, RRE）来评估与真值之间的差距。

$$\begin{aligned} RTE &= \|\hat{\mathbf{t}} - \mathbf{t}\| \\ RRE &= \arccos \left(\frac{\text{Tr}(\hat{\mathbf{R}}^T \mathbf{R}) - 1}{2} \right) \end{aligned} \quad (4.5)$$

这里 \mathbf{t} 代表平移矩阵； \mathbf{R} 代表旋转矩阵。

4.3 实验设计与结果分析

4.3.1 实验设置

本文实验采用的编程语言为 Python，版本号 3.8.5，使用深度学习 PyTorch 框架，版本号 1.7.1，实验使用的 GPU 型号为 Tesla V100-SXM2。

4.3.2 基线数据集

- **FCGF**^[7]: 使用稀疏张量表示三维点云数据，通过全卷积网络一次计算点云特征，采用 Minkowski 卷积代替传统卷积，快速高效地进行特征计算。
- **D3Feat**^[10]: 使用基于 KPConv^[11] 的全卷积网络，通过多任务学习来完成提取和描述局部特征，实现快速推断。
- **PREDATOR**^[13]: 基于编码-解码网络架构，分别对源点云和目标点云进行局部信息提取，并使用交叉注意力机制^[22]进行点云之间的信息交换，从而充分利用点云重叠区域信息。
- **DCP-v2**^[19]: 使用可微分奇异值分解（SVD）求解刚性变换矩阵，首次将 Transformer 模型引入到点云配准任务当中来提升特征提取表现。
- **RPM-Net**^[44]: 使用可微分的 Sinkhorn 层和退火技术，利用从空间坐标和局部几何信息学习到的特征实现点对应的软分配。此外，RPM-Net 还提出了一种改进的倒角距离度量，以提高点云局部可见情况下的配准质量。

4.3.3 实验结果分析

在 3DLoMatch 数据集和 ModelNet 数据集上分别对本文的模型与不同的基线模型进行了对比实验，分别汇总在表4.2和表4.3当中并进行了进一步的结果分析。

表 4.2 3DLoMatch 数据集上实验结果对比

	RR	IR	FMR
FCGF	38.2	17.2	<u>74.2</u>
D3Feat	46.9	14.0	67.0
PREDATOR	<u>56.7</u>	<u>21.0</u>	73.8
Ours	58.6	23.1	77.9

如表4.2所示，本文的模型在低重叠率的 3DLoMatch 数据集上的表现显著优于现有的方法，说明本文模型通过关注重叠区域并融合上下文信息来进行点云配准是可行且有效的。此外，实验还说明描述子的选择对配准性能影响不大，反而是能够更加准确地选取位于重叠区域上的兴趣点会显著提升配准性能。本文模型在三个评价指标上的表现取得了 3.35% –

10.0% 不等的提升，尤其是在内点率 IR 上的表现较之最优的基线模型 PREDATOR 高出了 10.0%，说明本文模型能够给出更多正确的匹配点云对。此外，本文模型在配准召回率 RR 上较 PREDATOR 高出了 3.35%，在特征匹配召回率 FMR 上较 PREDATOR 高出了 5.56%，说明本文模型预测的位姿变换在真实匹配对上的表现和正确选取点云对的能力均优于基线模型。

表 4.3 ModelNet 数据集上实验结果对比

	RRE	RTE	CD
DCP-v2	11.975	0.171	0.01170
RPM-Net	1.712	0.018	0.00085
PREDATOR	1.856	0.019	0.00088
Ours	1.867	0.021	0.00089

如表4.3所示，本文模型在与专门针对 ModelNet 的模型的性能表现上相差不大。此外还观察到相较于随机采样，对具有重叠注意力特征点进行采样可以极大地提高配准性能。

根据两个数据集上的对比实验结果，可以得出本文模型在低重叠率点云上的配准表现能够达到甚至优于现有模型的配准水平，同时实验结果表明，本文模型能够通过关注点云重叠区域的信息，提取重叠区域上的兴趣点来进一步提升模型配准表现。

4.4 消融实验

为了验证章节三中介绍的模型各个组件的有效性，本文去掉了局部 Transformer 架构并命名为 $\text{model}_{\text{cross}}$ ， $\text{model}_{\text{cross}}$ 的训练流程为将两点云特征传入交叉注意力模块后将输出特征重新分为源点云特征和目标点云特征，再传入 EnhanceBlock 中进行融合后输入到解码中进行解码。将去掉交叉注意力模块的模型命名为 $\text{model}_{\text{local}}$ ，训练流程为将两点云特征分别传入局部 Transformer 架构中直接输入到 EnhanceBlock 中然后再将输出传入解码中。此外，本文还设置了单纯使用传统编码-解码网络架构，即去掉 EnhanceBlock 的模型并命名为 $\text{model}_{\text{noEnhance}}$ 。本文通过将三个分别去掉了部分模块的模型与原始模型 $\text{model}_{\text{origin}}$ 在 3DLoMatch 数据集上以同样的模型参数设置同样训练 30 个 epoch 来比较各个模型的表现，并进一步探讨各模块的有效性与作用。

表 4.4 消融实验探究模型各模块的有效性

	RRE	RTE	CD
model _{cross}	2.112	0.024	0.00121
model _{local}	2.268	0.024	0.00133
model _{noEnhance}	2.142	0.025	0.00131
model _{origin}	1.867	0.021	0.00089

综合表4.4中的实验结果，模型在去掉相应模块后的性能都出现了一定程度的下降，据此本文模型各个模块的有效性。其中，局部 Transformer 模块通过分别对源点云和目标点云提取点云特征来融合点云局部上下文信息，交叉注意力模块则实现了点云之间的信息交换，提取了点云重叠区域的信息，EnhanceBlock 则用于进一步融合点云局部和全局信息，在提升模型配准性能中发挥着关键的作用。

4.5 本章小结

为探究提升低重叠率点云上的配准性能，本文构建了基于 Transformer 架构的低重叠率点云配准模型。为了进一步探究模型在低重叠率点云上的配准表现与泛用性，本章选取了两个数据集 3DMatch 和 Modelnet 来验证模型性能，并为了更加精准的对比模型表现，在不同的数据集上分别选取不同的评价指标与基线模型与本文模型进行评估比较。此外，本章还进行了消融实验来验证模型各个模块的有效性与作用。实验结果表明，模型在点云配准上的表现可以达到甚至优于业界主流模型的水平，并且能够通过关注点云重叠区域的兴趣点采样来提升模型在低重叠率点云配准任务上的表现。

第五章 总结与展望

5.1 工作总结

低重叠率点云上的配准技术研究对于节约三维数据采集和训练成本、提升配准模型的泛化能力等方面有着重要的意义。本文从探究低重叠率点云上的配准技术出发，总结了点云配准相关的技术和理论知识，分析了现有方法在低重叠率点云上配准性能下降的可能原因，并进一步提出了基于 Transformer 架构的低重叠率点云配准模型，然后进行实验对比验证了模型的优异表现与模型各模块的有效性。本文的工作贡献可具体分为以下两点：

1. 通过使用 Transformer 和交叉注意力机制分别捕捉点云内部的局部上下文信息和点云之间的重叠区域信息来丰富特征表示，并引入 EnhanceBlock 来进一步融合局部信息和全局信息，提升模型对于点云重叠区域的提取和配准的关注度，从而解决低重叠率点云上的配准问题。
2. 通过实验探究得到了通过提升位于点云重叠区域兴趣点的采样比例能够提升模型配准效率的结论，为低重叠率点云上配准技术的性能提升提供思路。

5.2 工作展望

本文对低重叠率点云上的配准问题，提出了基于 Transformer 架构的点云配准模型，并进行了相关实验，验证了模型的有效性，取得了一定的成果。基于目前的研究，探究如何充分利用点云局部信息与全局信息，探索更加精确的点云重叠区域信息的提取方式，提升模型在不同点云数据集上的泛化能力和配准性能，都是本文未来的研究方向。

参考文献

- [1] Duan Y, Yang C, Chen H, et al. Low-complexity point cloud denoising for lidar by pca-based dimension reduction[J]. *Optics Communications*, 2021, 482:126567.
- [2] Besl P J, McKay N D. Method for registration of 3-d shapes[C]. 1992. 586–606.
- [3] Magnusson M, Lilienthal A, Duckett T. Scan registration for autonomous mining vehicles using 3d-ndt[J]. *Journal of Field Robotics*, 2007, 24(10):803–827.
- [4] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]. 2017. 652–660.
- [5] Aoki Y, Goforth H, Srivatsan R A, et al. Pointnetlk: Robust & efficient point cloud registration using pointnet[C]. 2019. 7163–7172.
- [6] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]. 1981. 674–679.
- [7] Choy C, Park J, Koltun V. Fully convolutional geometric features[C]. 2019. 8958–8966.
- [8] Choy C, Gwak J, Savarese S. 4d spatio-temporal convnets: Minkowski convolutional neural networks[C]. 2019. 3075–3084.
- [9] Graham B, Engelcke M, Van Der Maaten L. 3d semantic segmentation with submanifold sparse convolutional networks[C]. 2018. 9224–9232.
- [10] Bai X, Luo Z, Zhou L, et al. D3feat: Joint learning of dense detection and description of 3d local features[C]. 2020. 6359–6367.
- [11] Thomas H, Qi C R, Deschaud J E, et al. Kpconv: Flexible and deformable convolution for point clouds[C]. 2019. 6411–6420.
- [12] Ao S, Hu Q, Yang B, et al. Spinnet: Learning a general surface descriptor for 3d point cloud registration[C]. 2021. 11753–11762.
- [13] Huang S, Gojcic Z, Usvyatsov M, et al. Predator: Registration of 3d point clouds with low overlap[C]. 2021. 4267–4276.
- [14] Liu W, Sun W, Wang S, et al. Coarse registration of point clouds with low overlap rate on feature regions[J]. *Signal Processing: Image Communication*, 2021, 98:116428.
- [15] Cheng L, Chen S, Liu X, et al. Registration of laser scanning point clouds: A review[J]. *Sensors*, 2018, 18(5):1641.
- [16] 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.
- [17] Qi C, Yi L, Su H P, et al. Deep hierarchical feature learning on point sets in a metric space. arxiv 2017[J]. *arXiv preprint arXiv:1706.02413*.
- [18] Phan A V, Le Nguyen M, Nguyen Y L H, et al. Dgcnn: A convolutional neural network over large-scale labeled graphs[J]. *Neural Networks*, 2018, 108:533–543.

- [19] Wang Y, Solomon J M. Deep closest point: Learning representations for point cloud registration[C]. 2019. 3523–3532.
- [20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30.
- [21] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. 2015. 3431–3440.
- [22] Sarlin P E, DeTone D, Malisiewicz T, et al. Superglue: Learning feature matching with graph neural networks[C]. 2020. 4938–4947.
- [23] 杨佳琪. 点云局部特征描述与匹配研究 [D]. 博士学位论文, 武汉: 华中科技大学, 2019.
- [24] Yang J, Cao Z, Zhang Q. A fast and robust local descriptor for 3d point cloud registration[J]. *Information Sciences*, 2016, 346:163–179.
- [25] Rusu R B, Blodow N, Beetz M. Fast point feature histograms (fpfh) for 3d registration[C]. 2009. 3212–3217.
- [26] Guo Y, Sohel F, Bennamoun M, et al. Rotational projection statistics for 3d local surface description and object recognition[J]. *International journal of computer vision*, 2013, 105:63–86.
- [27] Salti S, Tombari F, Di Stefano L. Shot: Unique signatures of histograms for surface and texture description[J]. *Computer Vision and Image Understanding*, 2014, 125:251–264.
- [28] Zeng A, Song S, Nießner M, et al. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions[C]. 2017. 1802–1811.
- [29] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification[C]. 2005. 539–546.
- [30] Deng H, Birdal T, Ilic S. Ppfnet: Global context aware local features for robust 3d point matching[C]. 2018. 195–205.
- [31] Deng H, Birdal T, Ilic S. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors[C]. 2018. 602–618.
- [32] Graham B, Van der Maaten L. Submanifold sparse convolutional networks[J]. *arXiv preprint arXiv:1706.01307*, 2017.
- [33] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.
- [34] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. *arXiv preprint arXiv:2010.11929*, 2020.
- [35] Lu D, Xie Q, Wei M, et al. Transformers in 3d point clouds: A survey[J]. *arXiv preprint arXiv:2205.07417*, 2022.
- [36] Zhao H, Jia J, Koltun V. Exploring self-attention for image recognition[C]. 2020. 10076–10085.
- [37] Zhao H, Jiang L, Jia J, et al. Point transformer[C]. 2021. 16259–16268.
- [38] Pan X, Xia Z, Song S, et al. 3d object detection with pointformer[C]. 2021. 7463–7472.
- [39] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry[C]. 2017. 1263–1272.

- [40] Wang C Y, Liao H Y M, Wu Y H, et al. Cspnet: A new backbone that can enhance learning capability of cnn[C]. 2020. 390–391.
- [41] Qu L, Liu S, Wang M, et al. Transmef: A transformer-based multi-exposure image fusion framework using self-supervised multi-task learning[C]. 2022. 2126–2134.
- [42] Liu F, Ren X, Zhang Z, et al. Rethinking skip connection with layer normalization in transformers and resnets[J]. *arXiv preprint arXiv:2105.07205*, 2021.
- [43] Wu Z, Song S, Khosla A, et al. 3d shapenets: A deep representation for volumetric shapes[C]. 2015. 1912–1920.
- [44] Yew Z J, Lee G H. Rpm-net: Robust point matching using learned features[C]. 2020. 11824–11833.

致 谢

四年的本科学习生涯转瞬即逝，积石堂夏日的绿荫满窗、毓秀湖上的彩虹垂柳，回想大学校园的种种场景还历历在目，如今我们都即将要开启到人生新的篇章。

首先要感谢我的父母，他们陪伴我在懵懂中成长，在人生每个重要阶段都帮助我做出了正确的抉择。

感谢赵志立老师，我本科的科研工作和毕设论文都是在他的指导下进行的，带领我初涉科研领域，锻炼了我的科研素养和写作能力。

感谢我的导师和师兄，带领我进入了三维视觉的研究领域，耐心为我的毕业设计提出指导建议。今后的日子里我会努力坚持、脚踏实地，不断向前。

感谢我的朋友们，所有或是开心或是焦虑的情绪都有你们听我分享倾诉，愿我们未来也能一同探索这世间的美好。

凡是过往，皆为序章。祝愿我们天高海阔，万事胜意。

毕业论文（设计）成绩表

导师评语

建议成绩_____

指导教师（签字）_____

答辩小组意见

答辩委员会负责人（签字）_____

成绩_____

学院（盖章）_____

年 月 日