

## 一种使用伪对应点生成的3D点云配准方法

柏正尧, 许祝, 张奕涵

### 引用本文

柏正尧, 许祝, 张奕涵. 一种使用伪对应点生成的3D点云配准方法[J]. 计算机科学, 2023, 50(9): 210-219.

BAI Zhengyao, XU Zhu, ZHANG Yihan. [Deep Artificial Correspondence Generation for 3D Point Cloud Registration](#) [J]. Computer Science, 2023, 50(9): 210-219.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于CT图像语义的COVID-19实例分割与分类网络](#)

COVID-19 Instance Segmentation and Classification Network Based on CT Image Semantics  
计算机科学, 2023, 50(6A): 220600142-9. <https://doi.org/10.11896/jsjx.220600142>

#### [基于全局属性注意力神经过程模型的数据补全研究](#)

Study on Data Filling Based on Global-attributes Attention Neural Process Model  
计算机科学, 2022, 49(10): 111-117. <https://doi.org/10.11896/jsjx.210800038>

#### [融合交叉注意力机制的图像任意风格迁移](#)

Image Arbitrary Style Transfer via Criss-cross Attention  
计算机科学, 2022, 49(6A): 345-352. <https://doi.org/10.11896/jsjx.210700236>

#### [基于Transformer交叉注意力的文本生成图像技术](#)

Text-to-Image Generation Technology Based on Transformer Cross Attention  
计算机科学, 2022, 49(2): 107-115. <https://doi.org/10.11896/jsjx.210600085>

# 一种使用伪对应点生成的 3D 点云配准方法

柏正尧 许 祝 张奕涵

云南大学信息学院 昆明 650500

**摘 要** 针对三维重建过程中点云配准存在的挑战性问题(如寻找对应点困难等)展开研究,充分利用源点云和目标点云的几何信息,提出了一种基于交叉注意力和伪对应点生成机制的点云配准方法——深度伪对应点生成(DeepACG)。该方法采用三级网络模型,第一级是深度特征编码模块,利用交叉注意力机制交换和增强两片待配准点云之间的上下文和结构信息;第二级是伪对应点生成模块,基于软映射关系加权合成伪对应点;第三级为对应点加权和离群点过滤模块,赋予每个对应点对不同的权重值并剔除概率较低的离群点。在合成和真实数据集上进行大量实验,DeepACG 方法在室内真实数据集 3DMatch 上的配准召回率达到 92.61%;在数据集 ModelNet40 上进行目标未知的局部点云配准实验,旋转矩阵和平移向量的均方根误差分别降至 0.016 和 0.00009。实验结果表明,DeepACG 配准精度高,鲁棒性强,配准误差低于当前主流的配准方法。

**关键词:** 交叉注意力;伪对应点生成;离群点过滤;三维点云配准

中图法分类号 TP391

## Deep Artificial Correspondence Generation for 3D Point Cloud Registration

BAI Zhengyao, XU Zhu and ZHANG Yihan

School of Information Science and Engineering, Yunnan University, Kunming 650500, China

**Abstract** To address the challenging problems of point cloud registration in 3D reconstruction(e. g., difficulty in finding corresponding points, etc.), this paper proposes a point cloud registration method based on cross-attention and artificial correspondence generation mechanism, Deep Artificial Correspondence Generation(DeepACG), by fully utilizing the geometric information of the source and target point clouds. Our method adopts a three-stage network model. The first stage is the deep feature encoding module, which exchanges and enhances the contextual and structural information between two unaligned point clouds using the cross-attention mechanism. The second stage is the artificial correspondence generation module, which synthesizes the artificial correspondences by weighting the soft mapping. The third one is the correspondence weighting and outlier filtering module, which assigns different weights to the correspondence pairs and rejects them with a small probability. Extensive experiments are conducted on both synthetic and real-world datasets. Our method achieves a registration recall of 92.61% on the real-world indoor dataset 3DMatch, and we execute unseen partial registration experiments on ModelNet40, reducing the root mean square error of the rotation matrix and translation vector to 0.016 and 0.00009, respectively. Experimental results show that DeepACG has higher registration accuracy and robustness, and its alignment error is lower than that of the existing mainstream registration approaches.

**Keywords** Cross-attention, Artificial correspondence generation, Outlier filtering, 3D point cloud registration

## 1 引言

随着激光雷达技术(Light Detection and Ranging, LiDAR)的飞速发展,三维点云数据采集成本较低、数据精度较高等方面的优势逐步显现。在自动驾驶<sup>[1]</sup>、移动机器人的同步定位和制图(Simultaneous Localization and Mapping, SLAM)<sup>[2]</sup>、目标位姿估计<sup>[3]</sup>等应用领域,点云配准是其中的重要一环。最基础的配准过程涉及预测源点云和目标点云间

的相对刚性变换,点云数据的许多特性令配准过程充满挑战。首先,点云数据具有稀疏和局部重叠特性,导致在目标点云中难以寻找到真实准确的对应点;其次,处理点云数据中数以万计的点需要有快速高效的算法及较强的计算资源;最后,处理好点云数据的非结构化特性和噪声对于精确的位姿估计来说非常重要。

传统的点云配准算法<sup>[4-5]</sup>高度依赖源点云和目标点云之间的初始变换估计,如果两片待配准点云之间的初始位置

到稿日期:2022-07-04 返修日期:2022-12-09

基金项目:云南省重大科技专项课题(202002AD080001);云南大学专业学位研究生实践创新基金(2021Y168)

This work was supported by the Yunnan Provincial Major Science and Technology Special Plan Projects(202002AD080001) and Practice & Innovation Foundation for Professional Degree Graduates of Yunnan University(2021Y168).

通信作者:柏正尧(baizhy@ynu.edu.cn)

信息较为粗糙,那么算法就容易陷入局部最优值,即非凸问题。目前,深度学习方法在目标分类、检测和分割等计算机视觉领域取得了显著的进步。为了能够最大化地提高配准精度,研究者提出了基于深度神经网络的特征描述符<sup>[6-7]</sup>、关键点检测<sup>[8]</sup>和端对端的配准网络<sup>[9-10]</sup>等方法。然而,采用深度学习方法解决点云配准问题仍存在不少挑战。一方面,尽管基于深度学习的方法已经被应用于点云配准场景中,但是大多数方法直接忽略了点云间丰富的高层次上下文信息;另一方面,根据特征匹配生成的部分对应点容易成为离群点,这限制了算法在大尺寸点云场景中的应用。

针对以上问题,本文提出了一种估计两片点云间相对刚性变换的端到端的配准方法——深度伪对应点生成网络(Deep Artificial Correspondence Generation, DeepACG)。DeepACG充分考虑了源点云和目标点云的结构信息,并在特征提取阶段交换源、目标点云之间的上下文信息,以提高配准的性能。DeepACG将两片待配准的源点云和目标点云作为输入,并预测出两片点云间的相对刚性变换。在3DMatch等公开数据集上的实验结果表明,本文方法具有较强的鲁棒性,能有效地将两片点云对齐。

## 2 相关工作

### 2.1 离群点过滤

由于设备存在固有噪声以及人为操作的误差,数据集合中会有少量的噪声<sup>[11]</sup>,根据深度特征匹配得到的对应点不可避免地包含少量离群点,若要实现更加精确的配准,则必须将这些离群点过滤掉。目前,传统的随机采样一致性(Random Sample Consensus, RANSAC)<sup>[12]</sup>和它的改进方法<sup>[13-14]</sup>依然是最为广泛采用的离群点过滤方法。RANSAC以迭代的方式采样小样本对应点子集,而这些小样本对应点子集在理想情况下是不包含离群点的,从而实现了剔除离群点的目的。其他过滤方法还有基于分支定界(Branch-and-bound)<sup>[5]</sup>、半正定规划(Semi-definite Programming)<sup>[15]</sup>和最大团(Maximal clique)<sup>[16]</sup>。以上3种方法虽然精度高,但往往需要很高的迭代采样次数或者较高的资源成本。此外,快速全局配准(Fast Global Registration, FGR)<sup>[17]</sup>、截断最小二乘估计和半定松弛(Truncated least squares Estimation and SEMidefinite Relaxation, TEASER)<sup>[18]</sup>方法对离群点的适应能力较强,这是因为FGR和TEASER构建了稳健的代价体,排除了离群点的干扰。

基于学习的离群点过滤方法首先在2D图像配准中得到应用,2D对应点被分类为内点/离群点。基于学习的3D离群点过滤方法,如深度全局配准(Deep Global Registration, DGR)<sup>[19]</sup>、3DRegNet<sup>[20]</sup>同样遵循上述分类思路。DGR<sup>[19]</sup>使用了一个6D的卷积网络来辨别内点和离群点,该网络在测试阶段将迭代次数降为一次,仅需要一个前向反馈通道。3DRegNet<sup>[20]</sup>以匹配点为输入,由分类模块输出所有匹配点的置信度参数,置信度参数表征匹配点成为内点的概率大小。

### 2.2 点云配准方法

迭代最近点(Iterative Closest Point, ICP)<sup>[4]</sup>是目前解决刚性配准问题最有名的传统算法,它交替执行寻找对应点和

最小二乘优化算法这两个步骤来更新配准状态。然而,ICP算法对初始变换值极为敏感。为了寻找全局最优解并解决非凸问题,Yang等<sup>[5]</sup>提出基于分支定界的Go-ICP算法来确定全局最优位姿,在提供全局最优解决方案时,该算法优于ICP算法。其他求解全局最优位姿估计的方法有凸松弛<sup>[21]</sup>、黎曼优化<sup>[22]</sup>和混合整数规划<sup>[23]</sup>等算法,但是这些方法的计算开销大,不能很好地满足实际应用的需求。

随着深度学习在点云配准中的应用发展,传统方法中存在的问题有了一些新颖的解决方案。PointNetLK<sup>[24]</sup>结合基于PointNet<sup>[25]</sup>的全局特征描述符和Lucas/Kanade(LK)<sup>[26]</sup>优化算法,以迭代方式求解相对刚性变换,从而使得算法的时间复杂度不再随着点数正相关变化。深度最近点(Deep Closest Point, DCP)<sup>[10]</sup>使用DGCNN<sup>[27]</sup>来提取局部特征,并用奇异值分解方法来求解刚性变换估计参数。文献<sup>[28]</sup>使用ISS方法对输入点云数据进行特征点提取,以提高待匹配点的重复性,同时采用深度特征描述符来改善待匹配点的相似度。RIENet<sup>[29]</sup>提出了一种基于可靠内点估计的无监督深度点云配准网络。

## 3 问题描述

点云数据一般由三维激光扫描笔扫描采集得到。通过估计待配准点云之间的相对刚性变换来对齐两片或者多片点云的任务称为点云配准。给定两片点云:

$$\mathcal{P} = \{x_i, i=1, 2, \dots, M\} \quad (1)$$

$$\mathcal{Q} = \{y_j, j=1, 2, \dots, N\} \quad (2)$$

其中, $\mathcal{P} \in \mathbb{R}^{M \times 3}$ 和 $\mathcal{Q} \in \mathbb{R}^{N \times 3}$ 分别表示源、目标点云中的所有点, $x_i \in \mathbb{R}^3$ 和 $y_j \in \mathbb{R}^3$ 是点云扫描片段中的关键点。这里先考虑点云配准中最简单的一种情况: $M=N$ 且 $\{(x_i, y_{\mathcal{M}(x_i)})\}_{i=1}^M$ ,即 $x_i$ 和 $y_{\mathcal{M}(x_i)}$ 是对应点对, $\mathcal{M}(\cdot)$ 表示 $\mathcal{P}$ 和 $\mathcal{Q}$ 之间的匹配关系。点云配准的目标是计算出能够使两片点云正确对齐的相对刚性变换:旋转矩阵 $\mathbf{R}^* \in \text{SO}(3)$ 、平移向量 $\mathbf{t}^* \in \mathbb{R}^3$ 。该变换可以通过最小化均方误差函数 $E$ 来求解,即:

$$\mathbf{R}^*, \mathbf{t}^* = \arg \min_{\mathbf{R}, \mathbf{t}} E \quad (3)$$

$$E = \frac{1}{M} \sum_{i=1}^M e_i \|\mathbf{R}x_i + \mathbf{t} - y_{\mathcal{M}(x_i)}\|^2 \quad (4)$$

其中, $e_i$ 表示 $(x_i, y_{\mathcal{M}(x_i)})$ 成为内点的概率。采用加权奇异值分解(Singular Value Decomposition, SVD)方法可以求得上述优化问题的闭式解。具体地,定义点云 $\mathcal{P}$ 和 $\mathcal{Q}$ 的加权形心:

$$\bar{x} = \frac{\sum_{i=1}^M e_i x_i}{\sum_{i=1}^M e_i}, \bar{y} = \frac{\sum_{i=1}^N e_i y_{\mathcal{M}(x_i)}}{\sum_{i=1}^N e_i} \quad (5)$$

然后,计算协方差矩阵 $\mathbf{H}$ :

$$\mathbf{H} = \sum_{i=1}^M e_i (x_i - \bar{x})(y_{\mathcal{M}(x_i)} - \bar{y})^T \quad (6)$$

最后,对矩阵 $\mathbf{H}$ 进行奇异值分解: $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ 。如果源、目标点云之间的匹配关系已知,那么旋转矩阵 $\mathbf{R}^*$ 和平移向量 $\mathbf{t}^*$ 可以通过式(7)和式(8)得到近似预测值:

$$\mathbf{R}^* = \mathbf{V} \text{diag}(1, 1, \dots, \det(\mathbf{V}\mathbf{U}^T)) \mathbf{U}^T \quad (7)$$

$$\mathbf{t}^* = -\mathbf{R}^* \bar{x} + \bar{y} \quad (8)$$

解决点云配准问题的挑战在于能否设计出准确反映 $\mathcal{P}$

和  $\mathcal{Q}$  之间匹配关系的映射函数  $\mathcal{M}(\cdot)$ 。在给定当前映射关系  $\mathcal{M}^n(\cdot)$  的条件下, ICP 交替地无限接近式(4)中函数  $E$  的临界点。ICP 算法首先基于之前的映射关系  $\mathcal{M}^{n-1}(\cdot)$  估计出当前的最优刚性变换, 然后根据得到的最优刚性变换更新映射关系  $\mathcal{M}^n(\cdot)$ ,  $n$  表示当前的迭代次数, 该算法达到最优条件时会自动终止。这个过程很容易实现, 同时也相对高效。但 ICP 算法容易陷入局部优化, 差异较大的初始变换容易造成对映射关系的不良估计, 从而导致 ICP 算法的鲁棒性较差。

寻找三维对应点是三维图形学中一项比较有挑战性的研究, 这是因为三维点没有明显特别的特征。为寻找位于目标点云中的对应点, DGR 学习点云中每个点的 FCGF 特征描述符, 寻找对应点的任务转变为匹配它们对应的深度特征。然而, 由于点云数据具有稀疏特性和局部重叠的特点,  $\mathcal{P}$  和  $\mathcal{Q}$  之间并不存在严格意义上完全匹配的对应点对。

为了解决寻找对应点的问题, 本文构造了伪对应点  $\hat{y}_i, \hat{y}_j$ , 是通过加权目标点云中一系列候选关键点产生的。如图 1 所示,  $x_i \in \mathcal{P}$  和  $\hat{y}_i \in \mathcal{Q}$  形成了伪对应点对, 已知一系列的伪对应点对  $\{(x_i, \hat{y}_i)\}_{i=1}^M$  和其相应的内点概率  $e_i$ , 相对刚性变换可以通过加权奇异值分解方法求解。

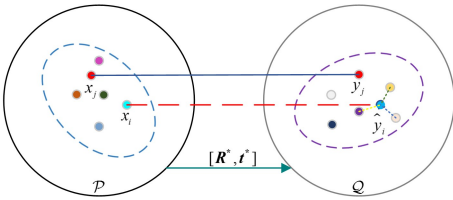


图 1 点云配准问题分析

Fig. 1 Analysis of point cloud registration problem

## 4 网络结构

本文提出了融合交叉注意力和伪对应点生成机制的配准方法——深度伪对应点生成方法 DeepACG。该方法提供了一种在缺少初始刚性运动信息的场景下用于恢复出正确映射关系  $\mathcal{M}(\cdot)$  和分类内点/离群点的配准网络。DeepACG 的网络结构如图 2 所示, 它由深度特征编码模块、伪对应点生成模块、对应点加权和离群点过滤模块组成。伪对应点生成模块构建了  $\mathcal{P} \rightarrow \mathcal{Q}$  的软映射关系, 从而确定  $\mathcal{P}$  和  $\mathcal{Q}$  之间的对应点。对应点加权模块和离群点过滤模块为对应点对赋予不同的权重并将权重值较低的离群点剔除, 避免增大刚性变换估计的误差。

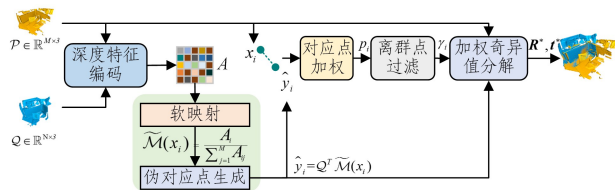


图 2 DeepACG 的网络结构

Fig. 2 Network architecture of DepACG

### 4.1 深度特征编码

深度特征编码模块的总体结构如图 3 所示, 它由  $L$  层结构一样的交叉注意力卷积网络组成,  $L$  的取值在后续实验中

具体讨论。每层交叉注意力卷积网络包含一个残差网络块  $\mathcal{B}$ ,  $\mathcal{B}$  由 4 层完全相同的残差网络基本单元顺序连接组成<sup>[30-31]</sup>。残差网络基本单元建立在核心点卷积(Kernel Point Convolution, KPConv)<sup>[32]</sup> 和特征核对齐(Feature-Kernel Alignment, FKACConv)<sup>[33]</sup> 的基础上, 其结构如图 4 所示, 依次由一层卷积 Conv1D、一层  $k$ -近邻算法( $k$ -Nearest Neighbor,  $k$ -NN)和 KPConv、两层 FKACConv 构成, 其中  $k$  等于 40。为了提高网络训练的稳定性, Conv1D 和 FKACConv 卷积层后面加入了实例归一化(Instance Normalization, IN)和激活函数 ReLU(Rectified Linear Units), KPConv 卷积层后面加入了 IN 和激活函数 LeakyReLU。

第  $l(1 \leq l \leq L)$  层交叉注意力卷积网络的输入为上一层的输出特征  $\mathcal{F}_{\mathcal{P}}^{(l-1)} \in \mathbb{R}^{M \times 2C^{(l-1)}}$  和  $\mathcal{F}_{\mathcal{Q}}^{(l-1)} \in \mathbb{R}^{N \times 2C^{(l-1)}}$ 。首先, 残差网络块  $\mathcal{B}$  将上一层的输出特征向量映射到高维特征空间中, 即  $\mathcal{F}_{\mathcal{P}}^{(l-1)}$  和  $\mathcal{F}_{\mathcal{Q}}^{(l-1)}$  分别经过相同的残差网络块  $\mathcal{B}$ , 从而得到新的特征向量  $\mathcal{H}_{\mathcal{P}}^{(l)}$  和  $\mathcal{H}_{\mathcal{Q}}^{(l)}$ 。为了方便描述, 将该过程简述为:

$$\mathcal{F}_{\mathcal{P}}^{(l-1)} \in \mathbb{R}^{M \times 2C^{(l-1)}} \xrightarrow{\mathcal{B}} \mathcal{H}_{\mathcal{P}}^{(l)} \in \mathbb{R}^{M \times C^{(l)}} \quad (l \neq 1) \quad (9)$$

其中,  $C^{(l-1)}$  是第  $(l-1)$  层中残差网络块  $\mathcal{B}$  的输出特征  $\mathcal{H}^{(l)}$  的通道数量, 每个点特征都可以用通道数为  $C^{(l-1)}$  的特征向量来表示。特别地, 第 1 层( $l=1$ )交叉注意力卷积网络的输入为源点云  $\mathcal{P}$  和目标点云  $\mathcal{Q}$ , 即  $\mathcal{F}_{\mathcal{P}}^{(0)} = \mathcal{P} \in \mathbb{R}^{M \times 3}$ ,  $\mathcal{F}_{\mathcal{Q}}^{(0)} = \mathcal{Q} \in \mathbb{R}^{N \times 3}$ , 此时  $C^{(0)} = 3$ 。

考虑到在配准任务中寻找两片点云之间的匹配关系的特殊性, 本文引入了交叉注意力机制, 将目标点云  $\mathcal{Q}$  中的结构信息聚合到点云  $\mathcal{P}$  中, 以此提高寻找对应点的准确率。特别地, 本文构建可微的最优相似度匹配矩阵  $\mathbf{A}_{(l)}$  来实现交叉注意力机制:

$$(\mathbf{A}_{(l)})_{ij} = \frac{e^{\left[\frac{a_{ij}}{\sigma^2}\right]_+}}{\sum_{k=1}^M e^{\left[\frac{a_{ik}}{\sigma^2}\right]_+}} \quad (10)$$

$$a_{ij} = \overline{\mathcal{H}_{\mathcal{P}}^{(l)}(x_i)} \cdot \overline{\mathcal{H}_{\mathcal{Q}}^{(l)}(y_j)} \quad (11)$$

其中,  $\overline{\mathcal{H}_{\mathcal{P}}^{(l)}(x_i)}$  和  $\overline{\mathcal{H}_{\mathcal{Q}}^{(l)}(y_j)}$  分别是  $\mathcal{H}_{\mathcal{P}}^{(l)}(x_i)$  和  $\mathcal{H}_{\mathcal{Q}}^{(l)}(y_j)$  的 L2-归一化(L2-normalization)特征向量,  $\mathcal{H}_{\mathcal{P}}^{(l)}(x_i)$  和  $\mathcal{H}_{\mathcal{Q}}^{(l)}(y_j)$  分别表示三维点  $x_i$  和  $y_j$  通过残差网络块  $\mathcal{B}$  的输出特征。 $a_{ij}$  用于度量关键点  $x_i$  和  $y_j$  之间的特征相似度、交换  $x_i$  和  $y_j$  之间的高层次上下文信息。 $\sigma$  取 0.2, 用于控制对特征相似度的敏感度。 $[\cdot]_+$  是  $\max(\cdot, 0)$  运算符, 保证  $a_{ij}$  的非负性。

其次, 使用最优相似度匹配矩阵  $\mathbf{A}_{(l)}$  交换点云  $\mathcal{P}$  和  $\mathcal{Q}$  之间的信息, 得到点云  $\mathcal{P}$  的新特征描述:

$$\mathcal{F}_{\mathcal{P}}^{(l)} = \text{cat}[\mathcal{H}_{\mathcal{P}}^{(l)}, \mathbf{A}_{(l)} \mathcal{H}_{\mathcal{Q}}^{(l)}] \in \mathbb{R}^{M \times 2C^{(l)}} \quad (12)$$

其中,  $\text{cat}[\cdot, \cdot, \cdot]$  表示按通道融合函数(channel-wise concatenation function),  $\mathcal{F}_{\mathcal{P}}^{(l)} \in \mathbb{R}^{M \times 2C^{(l)}}$  同样可以通过与式(12)相同的运算得到。构建新特征  $\mathcal{F}^{(l)}$  的过程有利于在下一层挖掘并交换两片点云之间的结构信息, 这取决于当前层的两片点云之间的几何相似度水平。 $\mathcal{F}_{\mathcal{P}}^{(l)}$  和  $\mathcal{F}_{\mathcal{Q}}^{(l)}$  表示第  $l$  层交叉注意力卷积网络的输出特征, 同时也是第  $(l+1)$  层交叉注意力卷积网络的输入特征。



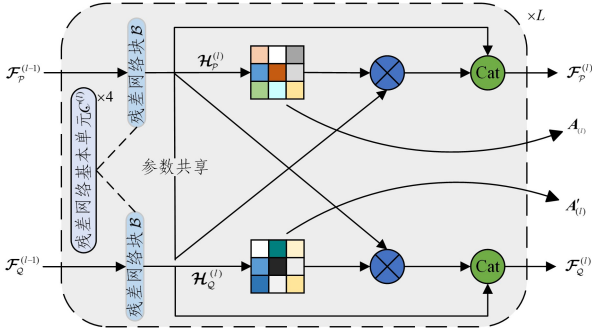


图3 深度特征编码模块

Fig. 3 Deep feature encoder module

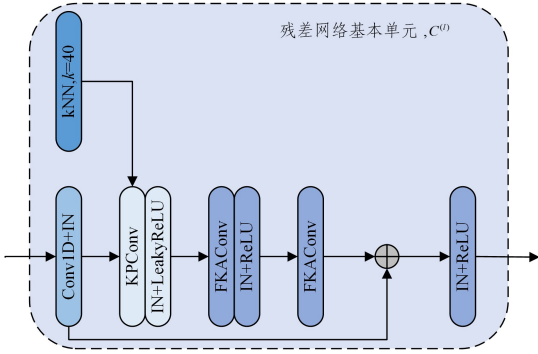


图4 残差网络基本单元

Fig. 4 Basic unit of residual network

特征编码模块的输入是待配准的源、目标点云  $\mathcal{P}$  和  $\mathcal{Q}$ , 第  $l$  层交叉注意力卷积网络会产生最优相似度匹配矩阵  $\mathbf{A}^{(l)} \in \mathbb{R}^{M \times N}$ 。最后, 通过一个简单的累加运算得到全局相似度匹配矩阵  $\mathbf{A}$ :

$$\mathbf{A} = \sum_{l=1}^L \mathbf{A}^{(l)} \quad (13)$$

其中, 矩阵  $\mathbf{A}$  的第  $i$  行第  $j$  列元素表示 3D 点  $x_i$  和  $y_j$  之间匹配成功的概率, 若  $x_i$  和  $y_j$  完全匹配, 则  $A_{ij} = 1$ , 否则  $0 \leq A_{ij} < 1$ 。

#### 4.2 伪对应点生成

伪对应点生成模块的结构如图 2 所示, 已知全局相似度匹配矩阵  $\mathbf{A}$ , 定义可微的软映射关系:

$$\tilde{\mathcal{M}}(x_i) = \frac{A_i}{\sum_{j=1}^N A_{ij}} \quad (14)$$

其中,  $\tilde{\mathcal{M}}(x_i) \in \mathbb{R}^N$ , 具有归一化指针的重要作用, 它规定了  $x_i$  指向伪对应点  $\hat{y}_i$  的映射关系;  $A_i$  是矩阵  $\mathbf{A}$  的第  $i$  行。

$\tilde{\mathcal{M}}(x_i)$  表征目标点云  $\mathcal{Q}$  中所有关键点  $y_j |_{j=1}^N$  与  $x_i$  匹配成功的归一化概率大小,  $\tilde{\mathcal{M}}_j$  越大,  $y_j$  与  $x_i$  的匹配成功可能性就越高。式(14)通过对矩阵  $\mathbf{A}$  按行进行归一化可以进一步扩大特征相似度的差异, 避免目标点云中特征相似度接近的两个关键点对合成的伪对应点造成干扰, 从而有效保证了得到高质量的伪对应点  $\hat{y}_i$ 。

DeepACG 使用加权运算方法得到伪目标点云  $\hat{\mathcal{Q}} = \{\hat{y}_i\}_{i=1}^N$ :

$$\hat{y}_i = \mathcal{Q}^T \tilde{\mathcal{M}}(x_i) \quad (15)$$

预测的伪目标点云  $\hat{\mathcal{Q}}$  用一个大小为  $N \times 3$  的张量表示。式(15)得到  $x_i$  的伪对应点  $\hat{y}_i$ ,  $\hat{y}_i$  由位于目标点云  $\mathcal{Q}$  中的真实 3D 点的加权和产生。理想情况下, 加权和的结果  $\hat{y}_i$  应该和  $x_i$  完全匹配。

传统 ICP 算法依赖迭代优化的方式, 其他基于学习的算法<sup>[19, 34]</sup>从目标点云中已存在的真实点中搜索源点云的对应点, 然后使用 RANSAC 算法来过滤离群点。与上述 ICP 等算法相比, DeepACG 方法充分利用了 KPConv 和 FKACConv 网络强大的泛化能力来学习关键点之间的高层次信息, 进而直接估计出对应点的位置。这减少了 RANSAC 算法的使用, 网络迭代次数减少到 1 次, 显著缩短了运行时间, 实现了较好的配准效果。

#### 4.3 对应点加权和离群点过滤

伪对应点生成模块为源点云  $\mathcal{P}$  中的所有关键点生成了对应点对  $(x_i, \hat{y}_i) \in \mathbb{R}^6$ 。点云  $\mathcal{P}$  和  $\mathcal{Q}$  并非完全重叠, 这意味着  $\mathcal{P}$  和  $\mathcal{Q}$  中仅有一部分关键点能够成功匹配, 因此需要采取措施过滤掉匹配失败的对应点对(离群点)。

本文首先构建对应点加权模块, 用于预测对应点对  $(x_i, \hat{y}_i)$  的内点置信度  $p_i$ ,  $p_i \in (0, 1)$  表征对应点对  $(x_i, \hat{y}_i)$  成为内点的置信度。

然后, 为实现剔除离群点的目标, 将  $p_i$  与阈值  $\tau$  进行比较, 当概率值  $p_i$  小于给定阈值  $\tau$  时,  $\gamma_i$  的值被置为 0(此时相应的对应点对分类为离群点), 反之,  $\gamma_i$  的值则等于  $p_i$  本身, 对应点对  $\{(x_i, \hat{y}_i)\}_{i=1}^M$  被区分为内点/离群点:

$$\gamma_i = p_i \llbracket p_i \geq \tau \rrbracket \quad (16)$$

其中,  $\llbracket \cdot \rrbracket$  是艾弗森括号(Iverson bracket)。所有的对应点对  $(x_i, \hat{y}_i)$  并行地执行上述比较运算, 可将置信度较低的对点滤除, 有效提高配准的精确度。阈值  $\tau$  在训练期间设定为 0.5, 这样约有 50% 的对应点对被预处理为离群点, 测试期间设置为 0, 在提高配准精确度的同时保证了网络训练的稳定性。

对应点加权模块的输入为对应点对  $\{(x_i, \hat{y}_i)\}_{i=1}^M \in \mathbb{R}^{M \times 6}$ , 输出为对应点对的内点概率  $\{p_i \in (0, 1)\}_{i=1}^M \in \mathbb{R}^{M \times 1}$ ,  $p_i$  的值越大, 对应点对  $(x_i, \hat{y}_i)$  成为内点的可能性就越大。该模块的结构如图 5 所示, 由 9 层输出通道数不同的残差网络基本单元、多层感知机 (Multi-Layer Perception, MLP) 和 Softmax 函数依次组成。

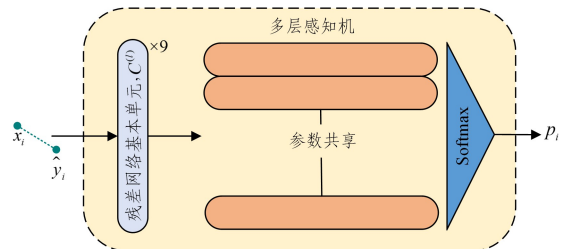


图5 对应点加权模块

Fig. 5 Correspondence weighting module

除输出通道数不同外, 每个残差网络基本单元的内部结构和图 4 完全一样, 每层残差网络基本单元的输出通道数

$C^{(l)}$  ( $1 \leq l \leq 9$ ) 依次为  $[16, 16, 32, 32, 64, 64, 64, 128, 128]$ , 其中第 1 层残差网络基本单元的输入通道数为 6, 表示对应点  $(x_i, y_i) \in \mathbb{R}^6$  的输入维度。多层感知机由 3 层全连接层构成, 前两层由卷积函数 Conv1D 和激活函数 ReLU 组成, 第三层仅由卷积函数 Conv1D 组成。Softmax 函数将多层感知机的输出结果转化为内点概率预测值。

#### 4.4 加权奇异值分解

已知对应点对的阈值权重  $\gamma_i$ , 本文使用最小二乘拟合法来估计旋转矩阵  $\mathbf{R}^*$  和平移向量  $\mathbf{t}^*$ 。

$$\mathbf{R}^*, \mathbf{t}^* = \arg \min_{\mathbf{R}, \mathbf{t}} \frac{1}{M} \sum_{i=1}^M \gamma_i \|\mathbf{R}x_i + \mathbf{t} - \hat{y}_i\|_2^2 \quad (17)$$

式(17)的解决方式类似于式(3), 即采用加权奇异值分解方法来求解相对刚性变换的预测值。

#### 4.5 损失函数

本文的总体损失函数由 3 个独立损失函数的加权和组成, 即分类损失、配准损失和对比损失。

针对源点云  $\mathcal{P}$  中的每个关键点  $x_i$ , 可以根据真实刚性变换  $[\mathbf{R}^{\text{GT}}, \mathbf{t}^{\text{GT}}]$  计算出相应的真实对应点  $y_i^{\text{GT}} = \mathbf{R}^{\text{GT}}x_i + \mathbf{t}^{\text{GT}}$ 。定义  $\mathbf{w}^{\text{GT}} = \{w_i^{\text{GT}}\}_{i=1}^M$ ,  $w_i^{\text{GT}}$  (等于 0 或 1) 是真实内点/离群点标签, 当  $\|\mathbf{R}^{\text{GT}}x_i + \mathbf{t}^{\text{GT}} - \hat{y}_i\| < \mathcal{T}$  时,  $w_i^{\text{GT}} = 1$ , 对应点对  $(x_i, \hat{y}_i)$  是真实内点:

$$w_i^{\text{GT}} = \|\mathbf{R}^{\text{GT}}x_i + \mathbf{t}^{\text{GT}} - \hat{y}_i\| < \mathcal{T} \quad (18)$$

其中,  $\mathbf{R}^{\text{GT}}$  和  $\mathbf{t}^{\text{GT}}$  分别表示真实旋转矩阵和真实平移向量,  $\mathcal{T}$  表示内点阈值。

分类损失 (Classification Loss) 使用二进制交叉熵来惩罚匹配失败的对应点。

$$\mathcal{L}_1 = \text{BCE}(\mathbf{p}, \mathbf{w}^{\text{GT}}) \quad (19)$$

其中, BCE 表示二进制交叉熵 (binary cross entropy)。 $\mathbf{p} = \{p_i\}_{i=1}^M$ ,  $p_i$  是 3.3 节中得到的内点概率。

配准损失 (Registration Loss) 使用目标点云中的伪对应点  $\hat{y}_i$  和真实点  $y_i^{\text{GT}}$  之间的欧氏距离来惩罚匹配错误的点。

$$\mathcal{L}_2 = \frac{1}{M} \sum_i p_i \|\mathbf{y}_i^{\text{GT}} - \hat{y}_i\|_2 \quad (20)$$

理想情况下, 伪对应点  $\hat{y}_i$  和相应的理想真实点  $y_i^{\text{GT}}$  之间的欧氏距离足够近时, 内点概率  $p_i$  的值将无限趋近于 1。

如果仅使用式(20)中的损失函数  $\mathcal{L}_2$ , 那么寻找关键点的匹配过程是相互独立的。因此, 配准过程中只考虑了局部邻近信息。针对这一特殊情况, 需要设计一种新的损失函数来平衡真实映射关系 ( $x_p \rightarrow y_q$ ) 和伪映射关系 ( $x_i \rightarrow \hat{y}_i$ ) 之间的一致性。受文献[18, 35]的启发, 本文考虑使用对比损失函数。真实对应点表决矩阵  $\mathbf{T}^{\text{GT}}$  是一个评价  $\mathcal{P}$  和  $\mathcal{Q}$  中的任意两个 3D 点之间的真实对应关系的二进制矩阵, 当  $(\mathbf{T}^{\text{GT}})_{pq} = 1$  时,  $x_p$  和  $y_q$  是成功匹配的真实对应点对。

$$\begin{aligned} (\mathbf{T}^{\text{GT}})_{pq} &= p = \arg \min_i \|\mathbf{R}^{\text{GT}}x_i' + \mathbf{t}^{\text{GT}} - y_q\| \cdot \\ q &= \arg \min_j \|\mathbf{R}^{\text{GT}}x_p + \mathbf{t}^{\text{GT}} - y_j'\| \end{aligned} \quad (21)$$

对比损失 (Contrastive Loss) 可以增强  $\mathcal{P}$  和  $\mathcal{Q}$  之间可靠的映射关系, 使离群点过滤模块学习如何滤除匹配失败的对应点。

$$\mathcal{L}_3 = \frac{1}{M} \sum_{(p,q)} \frac{1}{\log[\mathbf{A}_{pq} \cdot (\mathbf{T}^{\text{GT}})_{pq}]} \quad (22)$$

总损失函数是上述 3 种损失函数  $\mathcal{L}_1, \mathcal{L}_2$  和  $\mathcal{L}_3$  的加权和。

$$\mathcal{L} = \mathcal{L}_1 + \alpha \mathcal{L}_2 + \beta \mathcal{L}_3 \quad (23)$$

其中, 系数  $\alpha$  和  $\beta$  是用于平衡这 3 种损失函数的超参数。

## 5 实验设计

### 5.1 基准数据集

为了正确评估 DeepACG 方法的性能, 本文在室内、室外真实场景数据集和合成数据集上设计了多组实验。

对于室内真实场景数据, 本文使用了 3DMatch<sup>[34]</sup>, 该数据集共包含 62 个真实世界场景。本文按照标准的训练/验证/测试划分流程来制作重叠率超过 30% 的成对点云扫描图<sup>[6-7, 35]</sup>, 3DMatch 的 48 个场景用于训练, 6 个场景用于验证, 剩余的 8 个场景用于测试网络模型。训练和验证期间, 本文对 3DMatch 点云数据进行数据增强, 围绕每个空间坐标轴进行  $[-180^\circ, 180^\circ]$  范围内的随机旋转和  $[0, 0.5 \text{ m}]$  范围内的随机缩放。

对于大规模室外真实场景数据, 本文使用 KITTI Odometry<sup>[36]</sup>。该数据集包含了使用激光扫描仪在中型城市卡尔斯鲁厄周围、农村地区 and 高速公路上拍摄的 11 个序列模型, 真实位姿是由 GPS/IMU 系统提供的。为了公平比较, 本文按照文献[7, 19]中的划分策略将该数据集划分为训练/验证/测试集, 即立体序列编号为 0-5, 6-7, 和 8-10 的点云数据分别用于训练/验证/测试。在具体划分过程中, 本文使用 ICP 算法对原始真实位姿进行优化处理, 并使用 GPS-IMU 来建立相距 10m 以上的成对点云, 该过程分别产生了 1358 对、180 对和 555 对点云用于训练/验证/测试。

对于合成数据集, 本文使用 ModelNet40<sup>[37]</sup>, 该数据集包含 40 个目标类别的合成数据, 共计 12311 个 CAD 模型。为了平滑地模拟局部点云, 本文首先从每个 CAD 模型的外部表面均匀采样 1024 ( $M=1024$ ) 个点构建源点云  $\mathcal{P}$ , 然后对各片点云执行随机刚性变换以产生相应的目标点云  $\mathcal{Q}$ 。具体地, 沿各空间轴进行  $[0, 45^\circ]$  范围内的随机旋转和  $[-0.5 \text{ m}, 0.5 \text{ m}]$  范围内的随机平移, 最后随机放置一个 3D 点在点云空间中, 并使用最远点采样 (Farthest-Point Sampling, FPS) 方法采样其周围最近的 768 个点来模拟  $\mathcal{P}$  和  $\mathcal{Q}$  的局部点云扫描图。

### 5.2 评价指标

针对 ModelNet40 数据集, 本文采用均方根误差 (Root Mean Square Error, RMSE) 和平均绝对误差 (Mean Absolute Error, MAE) 两项评价指标, 用于评估旋转矩阵和平移向量的预测值与真实值之间的误差偏移。

针对 3DMatch 和 KITTI Odometry 数据集, 本文采用旋转误差、平移误差和配准召回率这 3 种指标来评估和比较性能表现。

旋转误差 (Rotation Errors, RE) 和平移误差 (Translation Errors, TE) 用于评估刚性变换预测值和真实值之间的误差。

$$\text{RE} = \arccos\left(\frac{\text{Tr}[(\mathbf{R}^*)^T \mathbf{R}^{\text{GT}}] - 1}{2}\right) \quad (24)$$

$$\text{TE} = \|\mathbf{t}^* - \mathbf{t}^{\text{GT}}\|_2$$

其中,  $Tr(\cdot)$  表示矩阵的迹。

配准召回率(Registration Recall, RR)表示成功配准的成对点云的比例,当 RE 和 TE 低于给定阈值时,点云配准的结果是成功的。评估误差 RE 和 TE 时,只有成功配对的点云才列入计算,这是因为配准失败的点云会返回一个与真实变换差异较大的刚性变换预测值,使得误差评估变得不可靠。

### 5.3 实验细节

训练阶段,本文使用深度学习框架 PyTorch 来实现 DeepACG 网络。所有的训练模型均使用 AdamW 优化器,权重衰减和初始学习率均设为 0.001,批次(batchsize)设置为 1,超参数  $\alpha$  和  $\beta$  均设置为 1。本文的实验环境基于 Ubuntu 系统,3DMatch 和 KITTI Odometry 数据集的硬件平台是 Intel i9-10900X CPU 和 GeForce GTX3090GPU,训练过程采用较大的学习率逐步衰减方案,两个数据集各自训练 120 轮(epochs),学习率在第 48,72 和 96 个初始轮次时分别衰减 90%;而 ModelNet40 数据集的硬件平台是 Intel Xeon E5-26530 v3 CPU 和 GTX 2080Ti GPU,该数据集共训练 20 轮,学习率在第 8,12 和 16 个初始轮次时分别衰减为原来的 1/10。

测试阶段,本文在 3DMatch 的测试集上提出联合防御机制来改善 DeepACG 的性能,联合防御机制由 RANSAC<sup>[12]</sup> 和 ICP<sup>[4]</sup> 算法组成。

## 6 实验结果及分析

### 6.1 室内场景配准

对于室内 3DMatch 点云数据,本文首先使用 5cm 体素网格对成对点云数据进行下采样,从而生成密度均匀的体素化点云。然后从体素化点云中随机采样 3000 个点构建源、目标点云( $M=N=3000$ )。4.1 节中深度特征编码模块的交叉注意力卷积网络层数  $L=6$ ,每层交叉注意力卷积网络中残差网络块  $\mathcal{B}$  的具体输出通道数如表 1 所列。式(18)中的内点阈值  $\mathcal{T}$  设为 10 cm。当 RE 和 TE 同时满足  $RE < 15^\circ$ ,  $TE < 30$  cm 时,3DMatch 的成对配准结果是成功的。3DMatch 的测试集共包含 8 个室内场景的局部扫描点云,如图 6 所示,本文提出的 DeepACG 方法能够出色地将两片室内场景点云对齐。

表 1 每层交叉注意力卷积网络中残差网络块  $\mathcal{B}$  的具体输出通道数  $C^{(l)}$

Table 1 Different output channels  $C^{(l)}$  of residual block  $\mathcal{B}$  in each cross-attention convolutional network

$l$	$C^{(l)}$		
	3DMatch	KITTI	ModelNet40
1	32	32	32
2	32	64	64
3	64	128	128
4	64	256	256
5	128	—	—
6	256	—	—

本文选取了几种经典的传统配准算法作为比较基准,即 point-to-point ICP, point-to-plane ICP, RANSAC<sup>[12]</sup>, FGR<sup>[17]</sup>, GC-RANSAC<sup>[13]</sup>。这 5 种算法的输入均是从体素下采样点云中提取的全卷积几何特征(Fully Convolutional Geometric Features, FCGF)<sup>[7]</sup>,其中前 4 种算法使用 Open3D 库<sup>[38]</sup>来实现。各种方法的比较结果如表 2 所列,

具体配准可视化对比如图 7 所示。

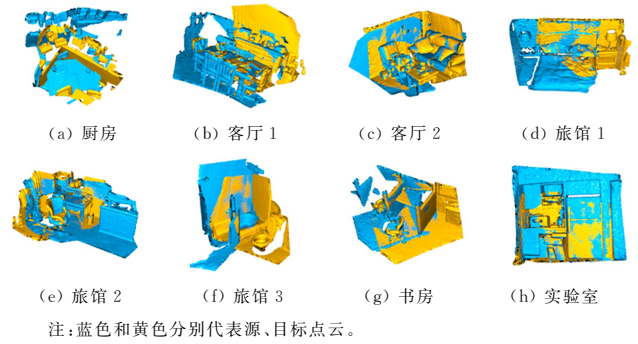


图 6 3DMatch 中 8 个不同测试场景的可视化结果 (电子版为彩图)

Fig. 6 Visualization results of 8 different test scenes on 3DMatch

表 2 3DMatch 数据集上的配准结果

Table 2 Registration results on 3DMatch

方法	RE(deg ↓)	TE(m ↓)	RR(% ↑)	时间/s
FGR <sup>[17]</sup>	2.91	0.84	78.62	0.64
ICP(P2Point) <sup>[38]</sup>	4.06	0.10	10.10	0.10
ICP(P2Plane) <sup>[38]</sup>	2.40	0.68	11.34	0.71
GC-RANSAC-1M <sup>[13]</sup>	2.29	0.07	91.68	0.42
RANSAC-2k <sup>[12]</sup>	3.02	0.09	88.42	0.15
RANSAC-20k	2.67	0.08	91.13	1.05
RANSAC-200k	2.47	0.08	91.99	10.89
DGR w/osafeguard <sup>[19]</sup>	2.58	0.07	85.20	0.70
DGR	2.43	0.07	91.30	1.21
PCAM-Sparse <sup>[39]</sup>	2.33	0.07	92.48	3.47
PCAM-Soft <sup>[39]</sup>	2.44	0.08	91.86	3.50
DeepACG(本文)	1.57	0.06	92.61	1.72

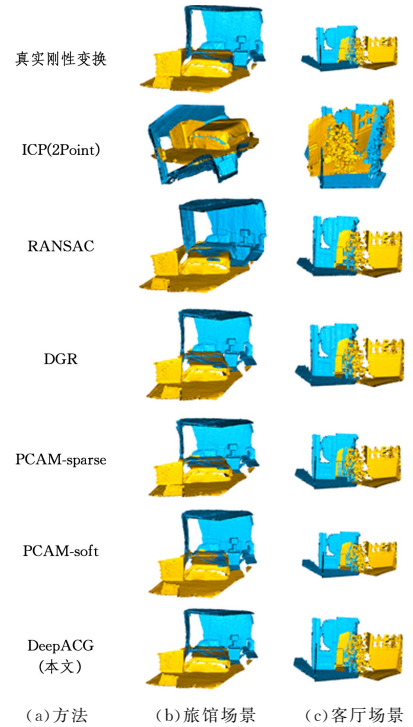


图 7 不同方法的可视化结果对比

Fig. 7 Visualization comparisons of different methods

ICP 的两种改进方法(P2Point 和 P2Plane)在大多数测试场景中配准失败,这是因为 3DMatch 数据集中存在大量的刚性运动且未提供可靠的先验知识。而基于特征匹配的方法 FGR, RANSAC 和 GC-RANSAC 的配准效果比 ICP 改进



方法的表现更为优秀。值得注意的是, RANSAC 算法在结合 FCGF 特征描述符后, 其配准性能较好, 召回率达到了 91.99%。DeepACG 的召回率高于 RANSAC 算法, 运行时间仅为 RANSAC-200k 的 1/5。

此外, 本文选取了 DGR<sup>[19]</sup> 和交叉注意力矩阵乘积(Product of Cross-Attention Matrices, PCAM)<sup>[39]</sup> 这两种基于深度学习的算法作为比较基准, DGR 的各项评测指标结果引用自文献[19]。本文探索了 DGR 缺少保护机制(即 RANSAC)时的性能表现, 当 DGR 缺少保护机制时, 其召回率仅为 85.20%。在待配准点云初始旋转角度较大时, 本文提出的 DeepACG 方法的配准召回率等评测指标优于基于深度学习的 DGR 和 PCAM 方法, 能够有效降低刚性变换的预测误差, 并精确地将两片局部点云扫描图对齐。总之, 在面对多目标、局部重叠、遮挡等极端情况时, 与其他主流方法相比, 本文提出的 DeepACG 方法同时保证了高精度的配准召回率和较短的运行时间。

## 6.2 大尺寸室外场景配准

本文进一步在 KITTI Odometry 数据集上评估了 DeepACG 在大规模室外稀疏点云场景中的配准性能表现。针对 KITTI 点云较为稠密的特点, 本文使用更大的 30 cm 体素网格对点云进行下采样, 并从体素化点云中随机选择 2048 个点, 即  $M=N=2048$ 。内点阈值  $\mathcal{T}=60$  cm, 交叉注意力卷积网络层数  $L=4$ , 每层交叉注意力卷积网络中残差网络块  $\mathcal{B}$  的具体输出通道数如表 1 所列。RE 和 TE 的阈值分别设置为  $5^\circ$  和 60 cm, 即  $RE < 5^\circ$  且  $TE < 60$  cm 时, 成对点云配准成功。

KITTI 数据集可视化示例如图 8 所示, 其中深粉红色和绿色分别代表源点云和目标点云, 比较结果如表 3 所列。

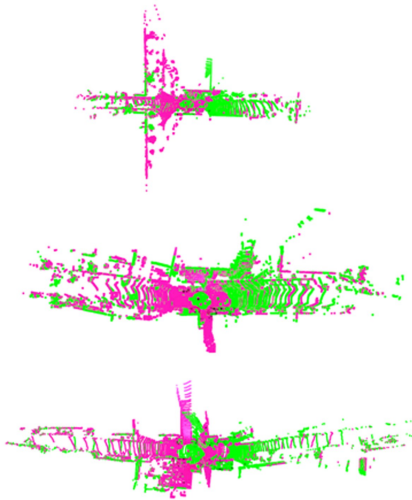


图 8 KITTI 配准可视化示例(电子版为彩图)

Fig. 8 Visualization examples of KITTI registration

可以看出, 相比其他配准方法, 本文提出的 DeepACG 方法能够有效降低刚性变换预测误差。如果使用 ICP 算法对初始刚性变换预测值进行优化处理, 本文方法的性能表现能达到最佳状态。得益于 FCGF 算法设计的全卷积几何特征体, 该算法在 KITTI 数据集上实现了最高的召回率, 且召回率趋于饱和状态, 但运行时间较长, 远高于 DeepACG。考虑到

KITTI 此类大尺寸点云场景的实际应用需求, 设计高效且精确的配准算法是非常必要的。DeepACG 实现了最低的旋转误差 RE 和平移误差 TE, 且运行时间极短, 这表明 DeepACG 在刚性变换预测误差和运行时间两者之间实现了更好的平衡。

表 3 KITTI 数据集上的配准结果

Table 3 Registration results on KITTI

方法	RE(deg ↓)	TE(m ↓)	RR(% ↑)	时间/s
RANSAC-1k <sup>[12]</sup>	0.48	0.23	96.58	0.30
RANSAC-20k	0.38	0.23	97.48	3.59
RANSAC-200k	0.35	0.22	97.12	37.44
GC-RANSAC <sup>[13]</sup>	0.44	0.23	96.22	0.51
FCGF <sup>[7]</sup>	0.33	0.10	<b>98.20</b>	6.38
DGR <sup>[19]</sup>	0.34	0.22	96.90	2.29
DeepACG(本文)	0.39	0.09	94.05	0.19
DeepACG+ICP	<b>0.32</b>	<b>0.03</b>	96.21	0.45

此外, 表 2 和表 3 的对比结果表明, DeepACG 方法在 3DMatch 上的所有评测指标均逊色于在 KITTI 上的配准表现, 这主要是因为两个数据集中待配准点云对的遮挡程度、重叠率、噪声强度和分布不一样。由于采集设备和环境的不同, 3DMatch 和 KITTI 中点云数据的噪声强度和分布情况也不同; 同时, 3DMatch 为室内场景点云数据, 遮挡较为严重。

## 6.3 合成点云数据配准

Wang 等<sup>[9]</sup> 比较了 ICP 等算法在合成数据集 ModelNet40 上的性能表现, 本文直接引用文献[9]中的实验结果。交叉注意力卷积网络层数  $L$  以及残差网络块  $\mathcal{B}$  的具体输出通道数和 KITTI Odometry 数据集完全一样。

### 6.3.1 目标未知的局部点云配准

本文首先在 ModelNet40 上分析目标位置的局部点云配准, 使用 9843 个 CAD 目标模型用于训练网络, 剩下的 2468 个目标模型用于测试。飞机、吉他等目标物体的点云配准可视化过程如图 9 所示, 其中深蓝色和红色分别代表源点云和目标点云。从表 4 可以看出, PCAM-Soft 方法实现了最低值的 MAE, DeepACG 方法的 RMSE 评测指标优于其他对比方法。

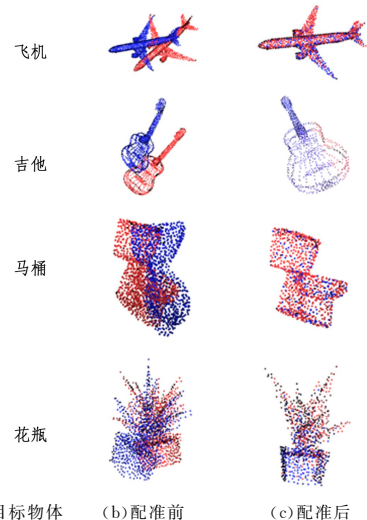


图 9 目标未知的局部点云的配准可视化结果(电子版为彩图)

Fig. 9 Registration visualization results on unseen partial point clouds



表4 目标未知的局部点云配准结果

Table 4 Registration results on unseen partial point clouds

方法	RMSE( $\mathbf{R}^*$ )	MAE( $\mathbf{R}^*$ )	RMSE( $t^*$ )	MAE( $t^*$ )
FGR <sup>[17]</sup>	11.238	2.832	0.03000	0.00800
ICP <sup>[4]</sup>	33.683	25.045	0.29300	0.25000
Go-ICP <sup>[5]</sup>	13.999	3.165	0.03300	0.01200
SymmetricICP <sup>[40]</sup>	20.312	6.242	0.09000	0.03700
PointNetLK <sup>[24]</sup>	16.375	7.550	0.04500	0.02500
DCP-v2 <sup>[10]</sup>	6.709	4.448	0.02700	0.02000
PRNet <sup>[9]</sup>	3.199	1.454	0.01600	0.01000
PCAM-Sparse <sup>[39]</sup>	0.023	0.012	0.00020	0.00009
PCAM-Soft <sup>[39]</sup>	0.023	<b>0.006</b>	0.00010	<b>0.00004</b>
DeepBBS <sup>[41]</sup>	0.041	0.021	0.00070	0.00040
DeepACG(本文)	<b>0.016</b>	0.008	<b>0.00009</b>	0.00005

### 6.3.2 类别未知的点云配准

为了评估 DeepACG 对类别未知的点云的泛化能力,本文将 ModelNet40 数据集按照目标类别分类为训练集和测试集<sup>[9-10]</sup>。DCP, PRNet, PCAM, DeepBBS, PointNetLK 和本文提出的 DeepACG 使用 ModelNet40 的前 20 个类别的 CAD 模型进行训练,所有方法均使用剩余 20 个类别的 CAD 模型进行测试。由表 5 可以看出,DeepACG 对类别未知的点云的泛化能力比其他比较基准方法更具竞争力,能够有效对齐类别未知的局部点云。

表5 点云类别未知的配准结果

Table 5 Registration results on unseen point cloud categories

方法	RMSE( $\mathbf{R}^*$ )	MAE( $\mathbf{R}^*$ )	RMSE( $t^*$ )	MAE( $t^*$ )
FGR <sup>[17]</sup>	9.932	1.952	0.0380	0.00700
ICP <sup>[4]</sup>	34.894	25.455	0.2930	0.25100
Go-ICP <sup>[5]</sup>	12.533	2.940	0.0310	0.01000
SymmetricICP <sup>[40]</sup>	19.058	5.847	0.0900	0.03800
PointNetLK <sup>[24]</sup>	22.943	9.655	0.0610	0.03300
DCP-v2 <sup>[10]</sup>	9.769	6.954	0.0340	0.02500
PRNet <sup>[9]</sup>	4.986	2.329	0.0210	0.01500
PCAM-Sparse <sup>[39]</sup>	0.072	0.018	0.0002	0.00010
PCAM-Soft <sup>[39]</sup>	<b>0.056</b>	0.013	0.0006	0.00010
DeepBBS <sup>[41]</sup>	0.075	0.040	0.0011	0.00060
DeepACG(本文)	0.064	<b>0.012</b>	<b>0.0002</b>	<b>0.00008</b>

### 6.3.3 含高斯白噪声的局部点云配准

本文进一步研究了 DeepACG 对噪声点云的鲁棒性,在 6.3.1 节中目标未知的局部点云配准实验基础上对点云添加均值为 0、标准差为 0.01 的高斯白噪声,超出 $[-0.05, 0.05]$

范围内的噪声会被剪掉。表 6 列出了各种配准方法的比较结果,可以看出,ICP 显然收敛于局部最优值,FGR 对噪声较为敏感。Go-ICP, PointNetLK, DCP 和本文提出的 DeepACG 对噪声的鲁棒性比较强劲。DeepACG 方法的 RMSE 等 4 项评价指标均为最低值,这表明 DeepACG 对噪声点云有着很强的鲁棒性。

表6 含高斯白噪声的局部点云配准结果

Table 6 Registration results on partial point clouds corrupted with white Gaussian noise

方法	RMSE( $\mathbf{R}^*$ )	MAE( $\mathbf{R}^*$ )	RMSE( $t^*$ )	MAE( $t^*$ )
FGR <sup>[17]</sup>	27.65	13.79	0.070	0.0390
ICP <sup>[4]</sup>	35.07	25.56	0.294	0.2500
Go-ICP <sup>[5]</sup>	12.26	2.85	0.028	0.0290
SymmetricICP <sup>[40]</sup>	20.69	6.77	0.092	0.0440
PointNetLK <sup>[24]</sup>	19.94	9.08	0.057	0.0320
DCP-v2 <sup>[10]</sup>	6.88	4.53	0.028	0.0210
PRNet <sup>[9]</sup>	4.32	2.05	0.017	0.0120
PCAM-Sparse <sup>[39]</sup>	0.58	0.15	0.002	0.0010
PCAM-Soft <sup>[39]</sup>	0.47	0.15	0.002	0.0010
DeepBBS <sup>[41]</sup>	4.20	1.72	0.044	0.0230
DeepACG(本文)	<b>0.37</b>	<b>0.14</b>	<b>0.001</b>	<b>0.0008</b>

## 6.4 消融实验与分析

本文设计了消融实验,以验证 DeepACG 网络不同模块的有效性,将 DeepACG 网络分割成独立的功能模块并使用可替代方案来代替模块。除待验证模块外,实验的其他参数设置与 6.3.1 节中目标未知和类别未知的局部点云配准实验完全一样。

### 6.4.1 伪对应点生成

为阐述伪对应点生成模块的有效性,本文直接在目标点云中搜索最优的真实对应点。具体地,对于源点云中的一个关键点  $x_i$ ,目标点云中与  $x_i$  相似度概率值最高的 3D 真实点  $y_j$  被认定为  $x_i$  的对应点:

$$j' = \arg \max_j (\mathbf{A})_{ij} \quad (25)$$

具体比较结果如表 7 中的第 1、4 行所示,使用伪对应点  $\hat{y}_i$  的效果比真实对应点更佳。这是因为点云数据自身的稀疏特性导致目标点云中不一定存在确切的真实对应点,而伪对应点可以有效地提高对应点匹配成功的可能性。

表7 伪对应点生成和离群点过滤的消融实验

Table 7 Ablation studies on artificial correspondence generation and outlier filtering

伪对应点生成	离群点过滤	目标未知的局部点云配准				类别未知的点云配准			
		RMSE( $\mathbf{R}^*$ )	MAE( $\mathbf{R}^*$ )	RMSE( $t^*$ )	MAE( $t^*$ )	RMSE( $\mathbf{R}^*$ )	MAE( $\mathbf{R}^*$ )	RMSE( $t^*$ )	MAE( $t^*$ )
×	✓	0.031	0.0100	0.00020	0.00007	0.106	0.035	0.0007	0.00027
✓	×	0.023	0.009	0.00020	0.00006	0.101	0.021	0.0005	0.00015
×	×	0.027	0.0100	0.00010	0.00007	0.070	0.015	0.0003	0.00011
✓	✓	<b>0.016</b>	<b>0.008</b>	<b>0.00009</b>	<b>0.00005</b>	<b>0.064</b>	<b>0.012</b>	<b>0.0002</b>	<b>0.00008</b>

### 6.4.2 离群点过滤

为验证离群点过滤模块的有效性,本文将离群点过滤模块直接剔除,即 4.3 节中的式(16)被修改为:

$$\gamma_i = p_i \quad (26)$$

此时,源点云  $\mathcal{P}$  和目标点云  $\mathcal{Q}$  中所有的关键点均参与刚性变换预测的求解过程。从表 7 中的第 2、4 行的对比结果

可以看出,结合离群点过滤模块后,旋转矩阵和平移向量的 RMSE 指标同时达到了最低水平,这表明离群点过滤模块的设计能够增强 DeepACG 网络训练的稳定性,提高配准精确度。

**结束语** 本文提出了一种基于交叉注意力和伪对应点生成的配准方法,即 DeepACG。交叉注意力机制可以关注点云

数据的高层次语义信息,增强配准网络对点云上下文信息的学习能力;伪对应点生成模块可以充分保留目标点云中不同候选关键点的细节信息。在合成和真实数据集上的实验结果表明,本文提出的 DeepACG 方法能够快速准确地将两片待配准点云对齐,配准精度高且泛化能力强。后续研究工作将聚焦探索新方法拓展 DeepACG 在更多实际应用场景下的泛化能力,并进一步研究更有效的深度特征编码策略,以帮助 DeepACG 提高大规模密集点云数据的配准精度。

## 参 考 文 献

- [1] ZERMAS D, IZZAT I, PAPANIKOLOPOULOS N. Fast segmentation of 3D point clouds: A paradigm on LiDAR data for autonomous vehicle applications[C]// Proceedings of the IEEE International Conference on Robotics and Automation(ICRA). New York: IEEE, 2017: 5067-5073.
- [2] DESCHAUD J E. IMLS-SLAM: Scan-to-Model Matching Based on 3D Data[C]// Proceedings of the IEEE International Conference on Robotics and Automation(ICRA). Los Alamitos: IEEE Computer Society, 2018: 2480-2485.
- [3] WONG J M, KEE V, LE T, et al. SegICP: Integrated deep semantic segmentation and pose estimation[C]// Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS) / Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics. New York: IEEE, 2017: 5784-5789.
- [4] BESL P J, MCKAY N D. A method for registration of 3-D shapes[C]// Proceedings of the Sensor Fusion Conf: Control Paradigms and Data Structures. Washington: The International Society for Optical Engineering(SPIE), 1992: 586-606.
- [5] YANG J L, LI H D, CAMPBELL D, et al. Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11): 2241-2254.
- [6] DENG H, BIRDAL T, ILIC S. PPFNet: Global Context Aware Local Features for Robust 3D Point Matching[C]// Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). New York: IEEE, 2018: 195-205.
- [7] CHOY C, PARK J, KOLTUN V. Fully Convolutional Geometric Features[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 8957-8965.
- [8] LI J, LEE G H. Usip: Unsupervised stable interest point detection from 3d point clouds[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Los Alamitos: IEEE Computer Society, 2019: 361-370.
- [9] WANG Y, SOLOMON J. PRNet: self-supervised learning for partial-to-partial registration[C]// Proceedings of the 33rd Conference on Neural Information Processing Systems(NeurIPS). La Jolla, California: Neural Information Processing Systems (NIPS), 2019: 8814-8826.
- [10] WANG Y, SOLOMON J M. Deep Closest Point: Learning Representations for Point Cloud Registration[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Los Alamitos: IEEE Computer Society, 2019: 3522-3531.
- [11] FENG L, ZHU D M, LI Z X, et al. Sparse Point Cloud Filtering Algorithm Based on Mask[J]. Computer Science, 2020, 47(12): 252-257.
- [12] 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.
- [13] BARATH D, MATAS J. Graph-cut RANSAC[C]// Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). New York: IEEE, 2018: 6733-6741.
- [14] LI J Y, HU Q W, AI M Y. GESAC: Robust graph enhanced sample consensus for point cloud registration[J]. Isprs Journal of Photogrammetry and Remote Sensing, 2020, 167: 363-374.
- [15] MACIEL J, COSTEIRA J P. A global solution to sparse correspondence problems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(2): 187-199.
- [16] PERERA S, BARNES N. Maximal Cliques Based Rigid Body Motion Segmentation with a RGB-D Camera[C]// Proceedings of the 11th Asian Conference on Computer Vision. Berlin: Springer, 2013: 120-133.
- [17] ZHOU Q Y, PARK J, KOLTUN V. Fast Global Registration [C]// Proceedings of the 14th European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2016: 766-782.
- [18] YANG H, SHI J N, CARLONE L. TEASER: Fast and Certifiable Point Cloud Registration[J]. IEEE Transactions on Robotics, 2021, 37(2): 314-333.
- [19] CHOY C, DONG W, KOLTUN V. Deep Global Registration [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 2511-2520.
- [20] PAIS G D, RAMALINGAM S, GOVINDU V M, et al. 3DRegNet: A Deep Neural Network for 3D Point Registration [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE: 7191-7201.
- [21] MARON H, DYM N, KEZURER I, et al. Point Registration via Efficient Convex Relaxation[J]. ACM Transactions on Graphics (TOG), 2016, 35(4): 1-12.
- [22] ROSEN D M, CARLONE L, BANDEIRA A S, et al. SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group[J]. The International Journal of Robotics Research, 2019, 38(2/3): 95-125.
- [23] IZATT G, DAI H, FEDRAKE R. Globally Optimal Object Pose Estimation in Point Clouds with Mixed-Integer Programming [C]// Proceedings of the 18th International Symposium on Robotics Research(ISRR). Cham: Springer, 2020: 695-710.
- [24] AOKI Y, GOFORTH H, SRIVATSAN R A, et al. PointNetLK: Robust & Efficient Point Cloud Registration using PointNet [C]// Proceedings of the IEEE/CVF Conference on Computer

- Vision and Pattern Recognition (CVPR). New York: IEEE, 2019;7156-7165.
- [25] CHARLES R Q, SU H, KAICHUN M, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 77-85.
- [26] LUCAS B D, KANADE T. An Iterative Image Registration Technique with an Application to Stereo Vision [C]// Proceedings of the 7th International joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers Inc, 1981;674-679.
- [27] WANG Y, SUN Y B, LIU Z W, et al. Dynamic Graph CNN for Learning on Point Clouds[J]. Acm Transactions on Graphics, 2019, 38(5):1-12.
- [28] SHI W K, ZHANG Z C, YU M J, et al. Point Cloud Coarse Alignment Algorithm Based on Feature Detection and Depth Feature Description[J]. Computer Science, 2020, 47(12): 252-257.
- [29] SHEN Y Q, HUI L, JIANG H B, et al. Reliable Inlier Evaluation for Unsupervised Point Cloud Registration [C]// Proceedings of the AAAI Conference on Artificial Intelligence. California: Association for the Advancement of Artificial Intelligence (AAAI), 2022;2198-2206.
- [30] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016;770-778.
- [31] BAI Z Y, TAO J Y. Super-Resolution Reconstruction of Brain MR Images Using Pseudo-3D Convolutional Network[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(2):208-216.
- [32] THOMAS H, QI C R, DESCHAUD J E, et al. KPConv: Flexible and Deformable Convolution for Point Clouds [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019;6420-6429.
- [33] BOULCH A, PUY G, MARLET R. FKConv: Feature-kernel alignment for point cloud convolution [C]// Proceedings of the Asian Conference on Computer Vision (ACCV) 2020. Cham: Springer, 2020;381-399.
- [34] ZENG A, SONG S R, NIESSNER M, et al. 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions [C]// Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017;199-208.
- [35] DENG H, BIRDAL T, ILIC S. PPF-FoldNet: Unsupervised Learning of Rotation Invariant 3D Local Descriptors [C]// Proceedings of the 15th European Conference on Computer Vision (ECCV). Berlin: Springer-Verlag, 2018;620-638.
- [36] GEIGER A, LENZ P, URTASUN R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite [C]// Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2012;3354-3361.
- [37] WU Z R, SONG S R, KHOSLA A, et al. 3D ShapeNets: A Deep Representation for Volumetric Shapes [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2015;1912-1920.
- [38] ZHOU Q Y, PARK J, KOLTUN V. Open3D: A Modern Library for 3D Data Processing [J]. arXiv:1801.09847, 2018.
- [39] CAO A Q, PUY G, BOULCH A, et al. PCAM: Product of Cross-Attention Matrices for Rigid Registration of Point Clouds [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Los Alamitos: IEEE Computer Society, 2021;13209-13218.
- [40] RUSINKIEWICZ S. A symmetric objective function for ICP [J]. ACM Transactions on Graphics, 2019, 38(4):1-7.
- [41] HEZRONI I, DRORY A, GIRYES R, et al. DeepBBS: Deep Best Buddies for Point Cloud Registration [C]// Proceedings of 2021 International Conference on 3D Vision (3DV). Los Alamitos: IEEE Computer Society, 2021;342-351.



**BAI Zhengyao**, born in 1967, Ph.D, professor, master supervisor. His main research interests include signal processing, image processing, pattern recognition and machine learning, etc.

(责任编辑:何杨)