

基于双重注意力和匹配矩阵优化的点云配准算法

姬 硕¹, 胡立华¹, 张素兰¹, 胡建华^{2,3}, 王欣波^{2,3}

(1. 太原科技大学 计算机科学与技术学院, 山西 太原 030024;

2. 中国科学院 自动化研究所, 北京 100190;

3. 中科锐智(洛阳)数码科技有限公司, 河南 洛阳 471000)

摘要:针对点云配准过程中由于噪声、误匹配和漏匹配导致点云配准算法的配准精度低、鲁棒性差的问题,提出了一种融合双重注意力和匹配矩阵优化的点云配准算法。首先,设计了结合通道注意力和空间注意力的双重注意力模块,对噪声部分赋予较低的权重,使模型能够更专注于重要或相关的信息,从而减少噪声对模型的影响。其次,结合特征点的局部信息和全局信息设计了匹配矩阵优化模块,模型可以充分利用点云数据的多层次特征,从而提高配准的准确性。最后,以人工合成数据集 ModelNet40、真实室内场景数据集 7Scenes 和真实室外场景数据集 KITTI 为对象进行验证,在 ModelNet40 高噪声、7Scenes 和 KITTI 的点云配准实验中,旋转矩阵和平移向量的均方根误差分别降低至 0.665 7 和 0.001 7、0.079 6 和 0.000 9、2.061 7 和 0.041 7。实验结果表明,该方法可以在降低噪声对模型影响的同时,有效地减少漏匹配以及剔除误匹配,提高点云配准的精度和鲁棒性。

关键词:点云配准;通道注意力;空间注意力;匹配矩阵优化;深度学习

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2025)05-0097-09

doi:10.20165/j.cnki.ISSN1673-629X.2025.0002

Point Cloud Registration Algorithm Based on Dual Attention and Matching Matrix Optimization

Ji Shuo¹, Hu Li-hua¹, Zhang Su-lan¹, Hu Jian-hua^{2,3}, Wang Xin-bo^{2,3}

(1. School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China;

2. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;

3. Zhongke Ruizhi (Luoyang) Digital Technology Co., Ltd., Luoyang 471000, China)

Abstract: To address the issues of low registration accuracy and poor robustness in point cloud registration algorithms caused by noise, mismatches, and missing correspondences, a point cloud registration algorithm integrating dual attention and matching matrix optimization is proposed. Firstly, to improve the registration results caused by noise points, a dual attention module combining channel attention and spatial attention is designed, where noisy points are assigned lower weights, this allows the model to focus more on important or relevant information. Secondly, a matching matrix optimization module is designed by incorporating both local and global information of feature points, enabling the model to fully utilize the multi-level features of the point cloud data, thereby improving the registration accuracy. Finally, validation is conducted on the synthetic dataset ModelNet40, the real-world indoor dataset 7Scenes, and the real-world outdoor dataset KITTI. In the point cloud registration experiments on ModelNet40 with high noise, 7Scenes, and KITTI, the root mean square errors of the rotation matrix and translation vector were reduced to 0.665 7 and 0.001 7, 0.079 6 and 0.000 9, 2.061 7 and 0.041 7, respectively. The experimental results demonstrate that the proposed method effectively reduces missed matches and eliminates mismatches while minimizing the influence of noise on the model, thereby improving the accuracy and robustness of point cloud registration.

Key words: point cloud registration; channel attention; spatial attention; matching matrix optimization; deep learning

收稿日期:2024-09-26

修回日期:2025-01-21

基金项目:国家自然科学基金资助项目(62273248);山西省自然科学基金资助项目(202103021224285);中科院科技服务网络计划(STS-HP-202202)

作者简介:姬 硕(1998-),男,硕士研究生,研究方向为点云配准;通信作者:胡立华(1982-),女,博士,教授,研究方向为计算机视觉、人工智能、模式识别;张素兰(1971-),女,博士,教授,研究方向为计算机视觉、机器学习。

0 引言

点云配准主要是估计两个或多个三维点云之间变换矩阵的问题,即将两个或多个三维点云正确配准到同一个坐标系下,最终形成更完整的点云^[1]。点云配准在三维重建^[2]、自动驾驶^[3]、高精地图^[4]、医学图像处理^[5]和增强现实^[6]等领域具有至关重要的作用。然而在实际场景下,受到采集设备、遮挡等因素的影响,采集到的点云数据往往存在大量噪声,进而导致点云配准过程中存在不确定性和误差,从而影响点云配准的精度。另外,在配准过程中出现一个点云中的特征点错误地匹配到另一个点云中不正确的特征点上以及应当匹配的点未被成功匹配,同样会影响点云配准的精度。因此,设计高精度、鲁棒性强的点云配准算法有着重要的研究意义和价值。

目前,点云配准算法主要包括传统点云配准方法和基于深度学习的配准方法。传统点云配准方法中最经典的方法是迭代最近点算法^[7](Iterative Closest Point, ICP),其原理是通过迭代地修改点云中的刚性变换以最小化对应点之间的误差。然而,ICP 算法严重依赖于初始点云的选择,若初始点云选择不当,会使算法陷入局部最优而无法达到全局最优。因此,为了解决局部最优的问题,Go-ICP^[8]使用分支定界优化方法,通过对搜索空间进行分割和筛选,有效地寻找全局最优的配准结果。传统的点云配准方法中不仅有上述基于数学迭代法的算法,还有一些方法利用随机采样的策略来估计点云之间的变换关系,如 SDRSAC^[9]和 OPRANSAC^[10]等。这些方法往往与 RANSAC^[11](random sample consensus)算法相关,它们都尝试找到一个全局最优解,以确保配准过程的准确性并且这些方法不需要对刚体运动参数进行初始估计,这简化了配准过程。但是,相较于基于深度学习的点云配准算法,传统点云配准方法的配准精度通常较低。

由于深度学习模型具有强大的特征学习能力,可以从原始的点云数据中学习更高级别的特征表示,包括局部特征、全局特征以及语义信息等。近些年来,基于深度学习的点云配准方法得到了广泛的关注,并且相较于传统点云配准方法在精度上也有很大的提高。DCP^[12](Deep Closest Point)结合了动态图卷积网络(Dynamic Graph Convolutional Neural Network, DGCNN)和注意力机制来提取特征,并采用指针网络来预测点云之间的软匹配,这种模块化设计使得算法灵活且易于扩展,但是输入数据中的噪声和异常值会影响配准的准确性。DeepBBS^[13]通过学习一种在特征空间中能够识别“最佳匹配”即互为最近邻的点对的表示,并引入可微分的 SoftBBS 度量,提高了点云配准的精度。此外,该模型设计了一个新颖的损失函数,

不仅可以惩罚不准确的变换估计,还通过跳跃连接加速反向传播,使得学习过程更快收敛。VRNet^[14](VCPS to RCPS Network)模型通过学习矫正虚拟对应点和采用混合损失函数,在提高 3D 点云配准精度的同时保持了高时间效率,有效解决了异常点问题,并具有良好的泛化能力。MFGNet^[15]通过独立利用形状特征和空间坐标来引导对应点的搜索,并将匹配结果融合为最终的匹配矩阵。同时,它引入了一个对应点可信度计算模块,通过分析特征匹配矩阵和坐标匹配矩阵之间的冲突关系,来评估对应点的可信度。该方法有效提高了配准的准确性和鲁棒性。

CMIGNet^[16](Cross-Modal Information-Guided Network)通过融合跨模态信息和对比学习策略,在点云配准任务中实现了更高的精度和鲁棒性。AMGNet^[17]模型采用多重引导策略,独立处理点云匹配点概率矩阵和点云空间信息特征矩阵。模型结合注意力机制,共同搜索对应关系并确定对应点的权重。通过这种方式,实现了精确的部分到部分点云配准。EGST^[18]模型通过提出增强的几何结构变换器,利用显式的几何结构描述符和结构注意力机制来捕获点云间的结构一致性,从而提取可靠的对应关系,并实现准确的点云配准。SDT^[19]模型通过计算局部空间关系和提取多级分辨率特征,增强了点云空间特征的表达能力。另外,利用边缘松弛块降低了不可行匹配对配准鲁棒性的影响,有效提高了下点云配准的精度和鲁棒性。DBDNet^[20]模型通过采用双分支结构独立计算旋转和平移,结合多分辨率特征提取网络和注意力机制来处理配准精度的问题。

基于深度学习的点云配准方法在提取特征前,需要使用设备或技术来采集点云数据。然而,无论使用哪种设备或技术,采集到的点云数据都无法完全避免噪声的干扰。因此,特征提取器在提取特征时往往会受到噪声点的影响。此外,大多数现有模型过分依赖于单点特征匹配,这导致了匹配过程中的漏匹配问题,进而影响了配准的精度和鲁棒性。

因此,针对上述问题,该文结合通道注意力机制和空间注意力机制、特征点的全局信息和局部信息,设计了一种融合双重注意力和匹配矩阵优化的点云配准算法。由于噪声往往以较小、较分散的形式出现,而重要的点云特征往往具有较强的连续性和局部集中性。因此算法首先采用双重注意力机制,通过关注重要的点云特征从而降低了噪声的影响;然后采用匹配矩阵优化方法有效减少了漏匹配以及剔除模型中存在的误匹配,提高了点云配准的精度;最后采用数据集 ModelNet40^[21]、7Scenes^[22]和 KITTI^[23]来验证该算法的性能和稳定性。

1 网络模型

该方法的整体网络架构如图 1 所示。首先 DGCNN 对源点云 $x = \{x_i \in \mathbb{R}^3 \mid i = 1, 2, \dots, N\}$ 和目标点云 $y = \{y_j \in \mathbb{R}^3 \mid j = 1, 2, \dots, M\}$ 提取特征,然后通过 Transformer 模块得到融合的特征后引入了 DCAM 模

块以降低噪声对模型的影响,匹配矩阵优化模块能够全面地利用点云中的信息,得到一个更细致、更全面的匹配矩阵。最后,通过奇异值分解^[24] (Singular Value Decomposition, SVD) 层分解来提取到刚性变换 $T = \{R, t\}$,将两个点云进行对齐。

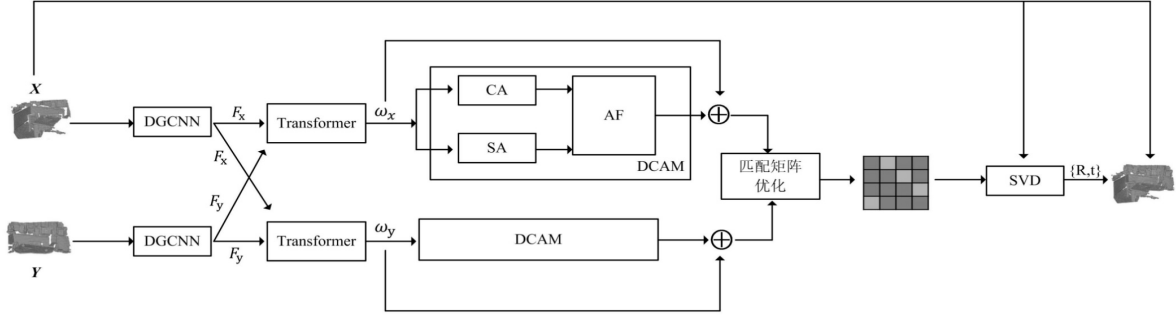


图 1 整体模型结构

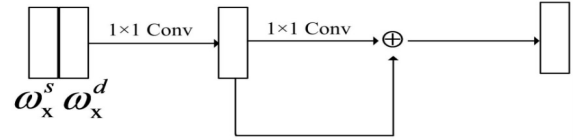
1.1 DGCNN 和 Transformer 模块

对于给定的源点云 x 和目标点云 y ,首先使用动态图卷积(DGCNN)提取逐点特征 F_x 和 F_y ,由于提取到的特征是相互独立的,会导致一些重要的结构或形状特征被忽略掉。因此,为了学习两个点云的上下文信息并获得融合特征,采用了 Transformer 来学习一个函数 μ ,这里的函数 μ 表示通过 Transformer 模块应用的一系列操作,包括自注意力、多头注意力、前馈网络和层归一化等。其能使模型学习点云特征 F_x 和 F_y 之间的关联性,更好地理解点云上下文信息,并提取到融合后的特征 ω_x 和 ω_y ,表示为:

$$\begin{aligned} \omega_x &= F_x + \mu(F_x, F_y) \\ \omega_y &= F_y + \mu(F_y, F_x) \end{aligned} \quad (1)$$

1.2 DCAM 模块

Transformer 的自注意力机制主要专注于捕捉序列内部的依赖关系,旨在捕获长距离依赖。但未专门处理噪声,导致融合后的点云特征中噪声累积,影响配准的质量和精度。因此,为了降低噪声的影响,该文提出了双通道注意力机制 (Dual Channel Attention Mechanism, DCAM),图 2 为 DCAM 模块的示意图。



(c) 注意力融合

图 2 DCAM 结构

为了降低噪声对模型的影响,引入了空间注意力模块。通过结合点的局部特征和空间全局特征,并为特征图中的不同区域分配不同权重,使模型更专注于重要的空间区域,并忽略受噪声影响的区域。此外,通过考虑空间位置间的关系,模型能够利用上下文信息来抑制噪声的影响。空间注意力模块如图 2(a)所示。首先,对于提取到的特征 ω_{x_i} ,将其输入到三个不同的 1×1 的卷积层得到 $\omega_{x_i}^a$ 、 $\omega_{x_i}^b$ 和 $\omega_{x_i}^c$ 。然后, $\omega_{x_i}^a$ 和 $\omega_{x_i}^b$ 通过 Softmax 函数获得空间注意力图 S 。

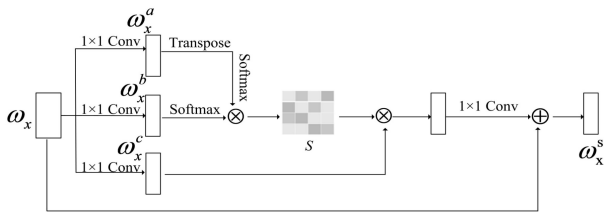
$$S_{ji} = \frac{\exp(\omega_{x_i}^b \cdot \omega_{x_j}^a)}{\sum_{i=1}^N \exp(\omega_{x_i}^b \cdot \omega_{x_j}^a)} \quad (2)$$

其中, S_{ji} 表示点 X_i 与点 X_j 之间相互关系的强度。最后,根据空间注意力图和原始的特征向量得到空间注意力模块的输出 $\omega_{x_i}^s$ 。

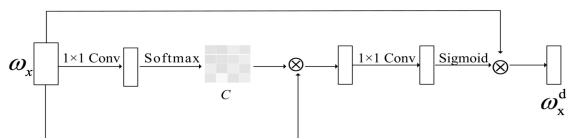
$$\omega_{x_i}^s = g_{\mu} \left(\sum_{i=1}^N (S_{ji} \cdot \omega_{x_j}^c) \right) + \omega_{x_i} \quad (3)$$

其中, $g_{\mu}(\cdot)$ 表示一个 1×1 的卷积层, μ 是一个可学习的参数。这样,对于每个点,既包含了点本身的信息,也融合了来自其他点的全局空间信息。因此,每个点的特征就不再是孤立的,而是包含了整个空间的上下文信息。

通道注意力模块通过学习每个通道的重要性权重,可以对特征通道进行重标定,强调重要的通道并抑制不重要的通道。使模型能够更加集中于那些对于当前任务最具区分性的特征,抑制那些包含噪声和不重



(a) 空间注意力



(b) 通道注意力

要信息的通道。通道注意力模块如图 2(b) 所示。首先将特征图进行注意力池化,注意力池化的步骤是先对输入的特征张量进行卷积操作,生成一个上下文掩码 Z_i 。这个掩码的作用是根据输入数据的实际特征情况动态地生成权重值,使得模型可以更好地适应不同样本的特征分布和结构,提高了模型的自适应能力。然后使用 Softmax 函数对每个通道内的上下文特征进行归一化,从而可以使得生成的权重值之间的差异更加明显,增强了模型对不同位置的关注度的区分能力。

$$C_{ji} = \frac{\exp(Z_{ij})}{\sum_{k=1}^N \exp(Z_{ik})} \quad (4)$$

最后,将输入特征张量和上下文掩码进行矩阵乘法操作,得到一个加权的上下文张量,其中每个位置的值表示该位置的特征值乘以对应位置的权重值。

经过注意力池化后,再使用两个卷积进行信息的处理并通过 Sigmoid 函数进行归一化。通过将特征图的每个通道的特征响应与对应通道的权重进行逐元素相乘,得到经过信息校准过的通道特征图 $\omega_{x_i}^d$ 。

$$\omega_{x_i}^d = \sigma(f(\sum_{i=1}^N (C_{ji} \cdot \omega_{x_i}))) \cdot \omega_{x_i} \quad (5)$$

其中, $f(\cdot)$ 表示一维卷积运算, σ 表示 Sigmoid 函数。

图 2(c) 展示了注意力特征融合的结构,通道注意力的输出 $\omega_{x_i}^s$ 和空间注意力的输出 $\omega_{x_i}^d$ 经过级联和残差连接等操作得到最终的特征 Φ_{x_i} 。

$$\Phi_{x_i} = \sigma(\omega_{x_i}^s \oplus \omega_{x_i}^d) + \omega_{x_i} \quad (6)$$

其中, \oplus 表示沿着特征维度的拼接操作, σ 由一个残差结构和 1×1 卷积组成。

1.3 匹配矩阵优化

在特征预处理后,模型虽能减少噪声对匹配矩阵的影响,但要获得理想的匹配矩阵还需考虑更多因素。首先,需要确保匹配矩阵考虑了整体的匹配关系,其次,要能更精细地识别相似的特征点对。因此,基于上述考量,该文提出了匹配矩阵优化模块,其目的是为了生成一个较为可靠的匹配矩阵,图 3 为匹配矩阵优化模块的示意图。

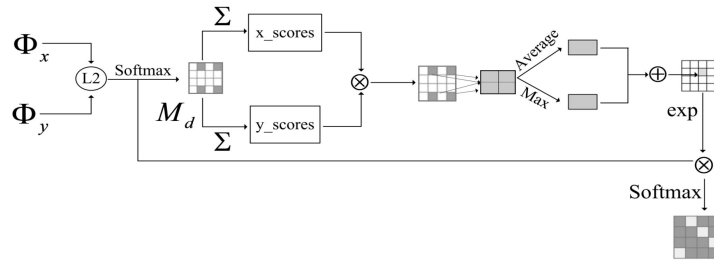


图 3 匹配矩阵优化结构

首先,将特征 Φ_x 和 Φ_y 通过特征差异来求得初始匹配矩阵 $M_{d,i,j}$ 。

$$M_{d,i,j} = \text{softmax}([-D_{i,1}, -D_{i,2}, \dots, -D_{i,m}])_j \quad (7)$$

其中, D_{ij} 代表了点云 x 中第 i 个点 x_i 和点云 y 中第 j 个点 y_j 之间的特征距离,这个特征距离是通过计算两个点的特征向量之间的欧氏距离得到的。

基于匹配矩阵 $M_{d,i,j}$,计算基于目标点云和源点云的匹配得分矩阵 $\mathbf{x_scores}_{i,j}$ 和 $\mathbf{y_scores}_{i,j}$,目的是对原矩阵进行归一化处理,从而缓解由于特征匹配得分的变化幅度较大而导致的不稳定现象以及更好地指导优化算法的收敛方向,从而提高优化的效果和速度。

$$\begin{aligned} \mathbf{x_scores}_{i,j} &= \frac{M_{d,i,j}}{\sum_{k=1}^n M_{d,i,k}} \\ \mathbf{y_scores}_{i,j} &= \frac{M_{d,i,j}}{\sum_{k=1}^m M_{d,i,k}} \end{aligned} \quad (8)$$

其中, m 和 n 代表目标点云和源点云特征点的总数。

然后将目标点云的匹配得分矩阵和源点云的匹配得分矩阵相乘,综合利用两种匹配得分的信息得到逐点匹配分数 $M_{a,i,j}$,从而可以更全面地评估特征点之间的匹配关系。

基于估计出的逐点匹配分数,通过对其周围点的对应分数求平均和最大来计算得到邻域分数矩阵 $M_{b,i,j}$,从而可以同时实现对局部结构的整体把握和关键特征的强化,优化点云配准的鲁棒性和准确性。

$$\begin{aligned} M_{b,i,j} &= \frac{1}{k} \sum_{x_i \in N_{x_i}, y_j \in N_{y_j}} M_{a,i,j} + \\ &\frac{1}{k} \text{Max}_{x_i \in N_{x_i}, y_j \in N_{y_j}} M_{a,i,j} \end{aligned} \quad (9)$$

其中, N_{x_i} 表示围绕点 x_i 的 k 个最近邻点的集合, k 表示邻域中的点的总数,即 k 近邻的数量, x_i 表示源点云中点 x_i 的邻居点,即 x_i 的 k 个最近邻点之一。这样计算的邻域匹配得分反映了点 x_i 和点 y_j 周围点对之间的一致性程度。如果这个邻域匹配得分较高,则表示该点对的周围点对之间的匹配概率较大,说明该点对的匹配是可靠的。下面得到精细化的特征距离 $M'_{b,i,j}$,目的是为了改进匹配映射。

$$\mathbf{M}'_{b_{i,j}} = \exp(\alpha - \mathbf{M}_{b_{i,j}}) * D_{ij} \quad (10)$$

其中, α 用于调节相邻分数之间的影响,其值设为0.7。 $\mathbf{M}'_{b_{i,j}}$ 是与邻域得分 $\mathbf{M}_{b_{i,j}}$ 负相关的,这意味着如果邻域得分 $\mathbf{M}_{b_{i,j}}$ 较高,则 $\mathbf{M}'_{b_{i,j}}$ 会相应较低,反之亦然。这样设计是因为真实对应点对通常具有较高的邻域相似性,而错误对应点对的邻域相似性往往较低。最后,使用 Softmax 来计算出最终的匹配矩阵 $\mathbf{M}_{i,j}$ 。

$$\mathbf{M}_{i,j} = \text{softmax}([-\mathbf{M}'_{b_{i,1}}, -\mathbf{M}'_{b_{i,2}}, \dots, -\mathbf{M}'_{b_{i,m}}])_j \quad (11)$$

1.4 SVD 模块

在这里假设上述源点云 x 和目标点云 y 中的点 x_i, y_i 是一一对应的,并且设 y 是 x 通过刚性变换得到的,将刚性变换表示为 $[\mathbf{R}_{xy}, \mathbf{t}_{xy}]$, 其中旋转矩阵 $\mathbf{R}_{xy} \in SO(3)$, 平移向量 $\mathbf{t}_{xy} \in \mathbb{R}^3$ 。利用最小二乘法思想最小化误差 $E(\mathbf{R}_{xy}, \mathbf{t}_{xy})$ 来求得刚性变换矩阵,其中 $E(\mathbf{R}_{xy}, \mathbf{t}_{xy})$ 的表达式为:

$$E(\mathbf{R}_{xy}, \mathbf{t}_{xy}) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{R}_{xy}x_i + \mathbf{t}_{xy} - y_i\|^2 \quad (12)$$

其中, N 代表源点云和目标点云中点的数量。

定义 x 和 y 的质心 \bar{x}, \bar{y} :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (13)$$

定义协方差矩阵 \mathbf{H} :

$$\mathbf{H} = \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})^T \quad (14)$$

使用 SVD 将协方差矩阵分解为 $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, 并通过最小化 $E(\mathbf{R}_{xy}, \mathbf{t}_{xy})$ 得到 \mathbf{R}_{xy} 和 \mathbf{t}_{xy} 。

$$\mathbf{R}_{xy} = \mathbf{V}\mathbf{U}^T, \quad \mathbf{t}_{xy} = -\mathbf{R}_{xy}\bar{x} + \bar{y} \quad (15)$$

1.5 损失函数

该文使用的损失函数和文献[13]一致,该损失函数由两部分组成。

$$L = \|\mathbf{R}_{xy}^{-1}\mathbf{R}_{gt} - \mathbf{I}\|^2 + \|\mathbf{t}_{xy}^{-1} - \mathbf{t}_{gt}\|^2 + \rho^n \frac{1}{N} \sum_{i=1}^N \delta_{gt,i} \|y'_i - (\mathbf{R}_{gt}x_i + \mathbf{t}_{gt})\|^2 \quad (16)$$

其中,前两项变换损失用于惩罚预测的旋转矩阵 \mathbf{R}_{xy} 和平移向量 \mathbf{t}_{xy} 与真实值 \mathbf{R}_{gt} 和 \mathbf{t}_{gt} 之间的不准确性。第三项点匹配权重损失促使网络学习权重 y'_i , 从而生成更准确的点映射。

这里 $\delta_{gt,i}$ 是一个指示器,它的作用是确保只有在点 x_i 在目标点云 Y 的坐标系中有匹配点时,则其值等于1并将该点对的匹配误差计入损失函数中。 ρ 是衰减常数, n 是训练的周期数,权重衰减用于平衡损失函数中的不同部分,使预测变换的准确性在训练过程中变得更加重要。这个损失函数结合了最佳匹配点的相似性度量,并通过迭代优化过程来最小化配准误差,从而提高点云配准的准确性。

2 实验

2.1 实验环境及参数

(1)参数设置:实验使用 PyTorch 框架实现并在一台配备 4090 GPU 的设备上进行端到端的训练。所有实验采用 Adam 优化器,具体实验参数为:学习率为 0.001,衰减率为 0.000 1。该文使用 ModelNet40 数据集、7Scenes 数据集和 KITTI 数据集。

(2)实验数据:ModelNet40 数据集由 40 个不同类别的 12 311 个 CAD 模型组成,80% 用于训练,20% 用于测试。从每个模型的外表面随机选择 1 024 个点。初始刚性变换从以下间隔中随机采样:在 $[0^\circ, 45^\circ]$ 范围内采样三个欧拉角旋转来应用旋转,在 $[-0.5, 0.5]$ 范围内对每个轴应用平移。7Scenes 数据集是广泛使用的室内环境配准的数据集,该数据集包含 7 个场景 (Chess、Fire、Heads、Office、Pumpkin、Redkitchen、Stairs) 的 353 个样本。其中 6 个场景 (Chess、Fire、Heads、Pumpkin、Stairs 和 Redkitchen) 的 296 个样本作为训练集; Office 场景的 57 个样本作为测试集。KITTI 是典型的室外场景数据集,采集过程使用配备 Velodyne-64 LiDAR 传感器的车辆,汽车在城市、乡村和高速公路等室外环境中行驶,捕获了 14 999 个户外场景数据,其中 7 481 组用于训练,7 518 组用于测试。

(3)评价指标:实验评价指标为旋转矩阵和平移向量的均方根误差^[25] (Root Mean Square Error, RMSE) 与平均绝对误差^[25] (Mean Absolute Error, MAE)。RMSE(\mathbf{R}) 表示旋转矩阵预测值与真值的均方根误差, RMSE(\mathbf{t}) 表示平移向量预测值与真值的均方根误差,其计算方法如式 17 所示。

$$\text{RMSE}(\mathbf{R}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{R}_{xy} - \mathbf{R}_{gt})^2} \quad (17)$$

$$\text{RMSE}(\mathbf{t}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{t}_{xy} - \mathbf{t}_{gt})^2}$$

MAE(\mathbf{R}) 表示旋转矩阵预测值与真值的平均绝对误差, MAE(\mathbf{t}) 表示平移向量预测值与真值的平均绝对误差,其计算方法如式 18 所示。

$$\text{MAE}(\mathbf{R}) = \frac{1}{N} \sum_{i=1}^N |\mathbf{R}_{xy} - \mathbf{R}_{gt}| \quad (18)$$

$$\text{MAE}(\mathbf{t}) = \frac{1}{N} \sum_{i=1}^N |\mathbf{t}_{xy} - \mathbf{t}_{gt}|$$

在理想配准情况下,点云变换误差应当完全消除,误差度量值都应趋近于零。因此, RMSE 和 MAE 值越小,表示配准精度越高。

2.2 实验结果

2.2.1 ModelNet40 配准实验

将该文提出的模型与传统点云配准算法 (ICP^[7]、Go-ICP^[8]、SDRSAC^[9]) 和基于深度学习的点云配准

网络 (包括 DCP^[13]、DeepBBS^[16]、VRNet^[17]、MFGNet^[18]、CMIGNet^[19]、DBDNet^[23]) 在同等硬件条件下针对以下三种场景进行了比较:(1)无噪声点云实验;(2)低噪声点云实验;(3)高噪声点云实验。

(1)无噪声点云实验。

在无噪声点云的实验中,使用了完整的数据集和测试集,在训练集上的 9 843 个数据用于训练网络,在测试集上的 2 468 个数据用于评估模型的性能。将该模型与其他多个模型进行了比较,对比结果如表 1(a)所示。结果表明文中方法在各个指标上都表现出了很好的性能。图 5(a)展示了模型的配准效果。

(2)低噪声点云实验。

为了验证模型在噪声下的抗干扰性,该文在无噪声数据的基础上向各个对比模型中加入同等条件的高斯噪声进行点云配准实验。图 5(b)展示了低噪声点

云配准的结果图。实验将均差为 0、标准差为 0.01 的低噪声加到点云中,并将噪声剪裁到 $[-0.05, 0.05]$ 的范围内。实验结果如表 1(b)所示。从表中可以看出,文中方法明显优于其他方法。这说明该模型具备一定的噪声抗干扰能力。

(3)高噪声点云实验。

高噪声点云的实验同样是通过向对比模型数据中加入同等条件的高斯噪声进行点云配准实验。与低噪声实验不同的是,高噪声实验是将均差为 0、标准差为 0.05 的高噪声加到点云中并将噪声剪裁到 $[-0.5, 0.5]$ 的范围内,其它实验设置与低噪声实验相同。图 5(c)表示高噪声点云配准效果。高噪声的实验结果如表 1(c)所示,从表中可以看出文中方法仍然优于其他方法。

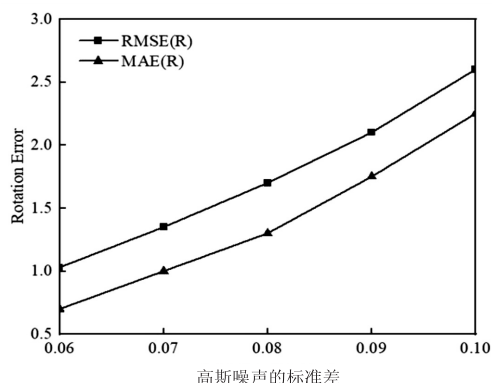
表 1 不同方法在 ModelNet40 无噪声、低噪声和高噪声上的配准结果

模型	(a) 无噪声点云				(b) 低噪声点云				(c) 高噪声点云			
	RMSE (<i>R</i>)	MAE (<i>R</i>)	RMSE (<i>t</i>)	MAE (<i>t</i>)	RMSE (<i>R</i>)	MAE (<i>R</i>)	RMSE (<i>t</i>)	MAE (<i>t</i>)	RMSE (<i>R</i>)	MAE (<i>R</i>)	RMSE (<i>t</i>)	MAE (<i>t</i>)
ICP	33.682 1	25.050 2	0.289 3	0.250 4	35.075 2	25.563 1	0.291 7	0.248 2	36.821 2	26.123 7	0.295 5	0.245 7
Go-ICP	14.054 6	3.173 2	0.033 6	0.012 7	12.264 1	2.852 2	0.028 7	0.029 4	11.435 1	2.753 9	0.024 2	0.044 6
SDRSAC	3.922 4	2.781 5	0.011 7	0.010 2	3.026 7	2.921 2	0.012 8	0.011 9	3.645 2	3.017 2	0.013 9	0.012 5
DCP	6.709 7	4.852 2	0.027 1	0.021 5	6.882 1	4.471 6	0.023 3	0.017 6	10.207 1	7.934 6	0.054 2	0.030 7
DeepBBS	0.043 7	0.022 8	0.000 7	0.000 4	4.201 2	1.770 3	0.044 7	0.023 5	7.056 6	4.285 1	0.063 4	0.048 1
VRNet	0.982 9	0.496 3	0.006 1	0.003 9	3.615 2	1.637 3	0.010 1	0.006 3	7.147 7	5.014 7	0.025 4	0.012 2
MFGNet	1.576 2	0.770 4	0.006 2	0.002 1	3.564 7	1.523 2	0.019 6	0.008 3	6.643 3	3.978 5	0.040 7	0.026 3
CMIGNet	0.772 5	0.408 2	0.004 8	0.003 0	1.472 1	0.632 4	0.005 8	0.003 6	4.793 2	1.952 0	0.015 7	0.005 9
DBDNet	4.921 2	2.011 6	0.078 7	0.037 2	5.684 2	2.221 9	0.080 7	0.040 2	7.176 3	3.002 4	0.122 5	0.071 5
文中算法	0.041 3	0.023 6	0.000 3	0.000 2	0.092 8	0.073 8	0.000 8	0.000 6	0.665 7	0.340 2	0.001 7	0.001 3

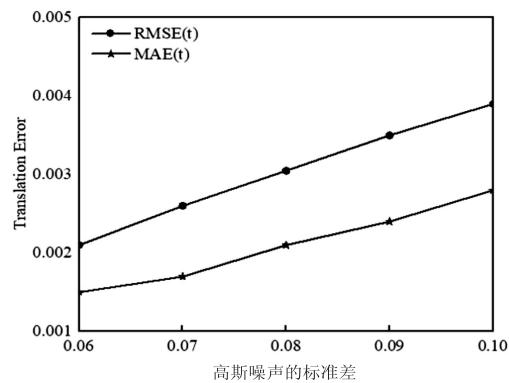
(4)噪声鲁棒性分析。

为了进一步证明模型对噪声的鲁棒性,在更高噪声水平下进行了测试,如图 4 所示。图 4(a)和图 4(b) X 轴表示高斯噪声的标准差,图 4(a) Y 轴表示旋转矩阵的两种不同的误差度量,分别是旋转矩阵的均方根误差 RMSE(*R*)和平均绝对误差 MAE(*R*)。图

4(b) Y 轴代表平移向量的两种不同的误差度量,分别是平移向量的均方根误差 RMSE(*t*)和平均绝对误差 MAE(*t*)。从图中可以看出,文中模型在噪声持续增长的情况下始终具有相当好的性能并且依然保持了稳定的配准精度。



(a) 旋转矩阵误差



(b) 平移向量误差

图 4 不同噪声水平下的误差

2.2.2 7Scenes 配准实验

7Scenes 数据集模拟了实际的复杂室内环境,这对于评估点云配准算法在现实世界中的应用潜力非常重要。7Scenes 数据集上的配准结果图如图 5(d)所示。实验环境、实验参数以及评价指标与 ModelNet40 一致,实验结果如表 2 所示,其中 ICP^[7]、SDRSAC^[9]为传统的点云配准方法,DCP^[13]、DeepBBS^[16]、VRNet^[17]、MFGNet^[18]、CMIGNet^[19]、DBDNet^[23]为基于深度学习的点云配准方法。实验结果表明,文中方法在 RMSE(R)、MAE(R)、RMSE(t)、MAE(t) 这四项指标上相较于其它方法都有较好的结果。

2.2.3 KITTI 配准实验

KITTI 数据集包含 11 个真实位姿的序列。使用序列的 00-05 进行训练,06-10 进行测试。实验设置与 ModelNet40 配准实验一致。其中,传统点云配准方法包括 ICP^[7]、SDRSAC^[9],基于深度学习的点云配准方法包括 DCP^[13]、DeepBBS^[16]、VRNet^[17]、MFGNet^[18]、CMIGNet^[19]、DBDNet^[23]。KITTI 数据集上的实验结果如表 3 所示,配准结果如图 5(e)所示。虽然文中方法在 MAE(R) 指标上不如 SDRSAC,但是仍能在与其它方法的比较中实现较为理想的结果。

表 2 7Scenes 数据集的实验结果

模型	RMSE(R)	MAE(R)	RMSE(t)	MAE(t)
ICP	19.789 4	7.505 2	0.106 3	0.029 8
SDRSAC	0.347 2	0.290 6	0.471 9	0.471 6
DCP	7.544 6	5.668 1	0.039 7	0.029 0
DeepBBS	1.853 8	0.055 9	0.001 8	0.001 3
VRNet	2.252 1	0.593 5	0.004 6	0.002 4
MFGNet	2.034 7	0.851 3	0.008 6	0.004 5
CMIGNet	0.804 7	0.488 3	0.003 2	0.001 8
DBDNet	3.352 2	1.657 5	0.051 6	0.028 7
文中算法	0.079 6	0.044 6	0.000 9	0.000 6

表 3 KITTI 数据集的实验结果

模型	RMSE(R)	MAE(R)	RMSE(t)	MAE(t)
ICP	19.554 3	3.972 0	1.825 6	0.901 4
SDRSAC	7.282 3	1.442 7	0.059 3	0.027 1
DCP	9.931 3	2.155 7	0.087 9	0.053 2
DeepBBS	2.538 2	1.744 6	0.052 5	0.022 8
VRNet	7.561 3	3.424 7	1.726 4	1.185 2
MFGNet	4.314 7	3.866 2	1.210 8	0.745 4
CMIGNet	2.875 6	1.537 9	0.076 2	0.048 7
DBDNet	6.714 1	2.722 3	1.028 6	0.645 5
文中算法	2.061 7	1.475 3	0.041 7	0.019 8

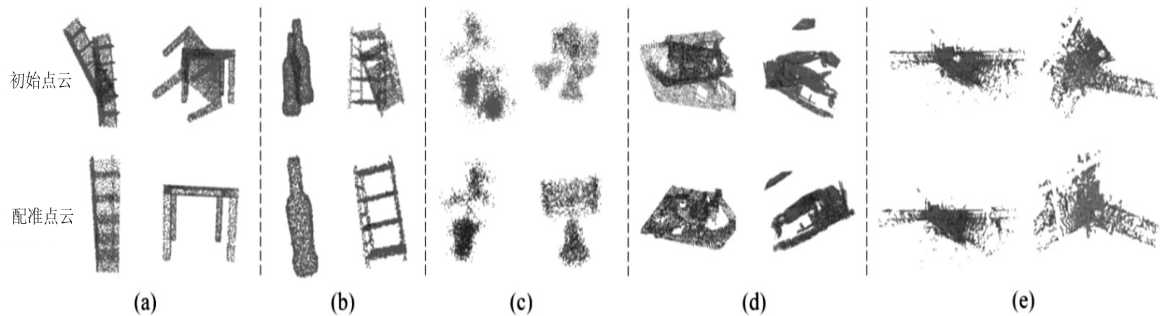


图 5 ModelNet40(a, b, c)、7Scenes(d)、KITTI(e)的定性配准结果

2.3 消融实验

在现实应用中,如自动驾驶、增强现实等,点云数据会因为运动、环境变化而发生大幅度的旋转。因此,配准算法需要能够处理这些大旋转的情况。为了验证文中模型对大旋转的鲁棒性,使用 $[0^\circ, 45^\circ]$ 范围内的样本进行训练,并使用 $[60^\circ, 180^\circ]$ 范围内的样本进行测试。不同方法的旋转误差在图6中显示,其中图6横坐标代表不同模型初始旋转的角度,纵坐标代表旋转矩阵的平均绝对误差。可以看到,文中模型对大变换是具有鲁棒性的。此外,SDRSAC依赖于旋转不变的FPFH,因此性能没有显著下降。

将文中方法与对比方法在ModelNet40数据集上进行了推理时间的比较。其中,文中方法的推理时间为52 ms,其它方法推理时间分别为ICP(8 ms),Go-ICP(2 740 ms),SDRSAC(22 416 ms),DCP(26 ms),DeepBBS(57 ms),VRNet(28 ms),MFGNet(36 ms),CMIGNet(102 ms),DBDNet(43 ms)。文中模型在推理时间方面处于中等水平,但是配准精度在可接受的计算成本内要优于其它方法。因此,文中方法在精度和算法效率之间取得了较好的平衡。

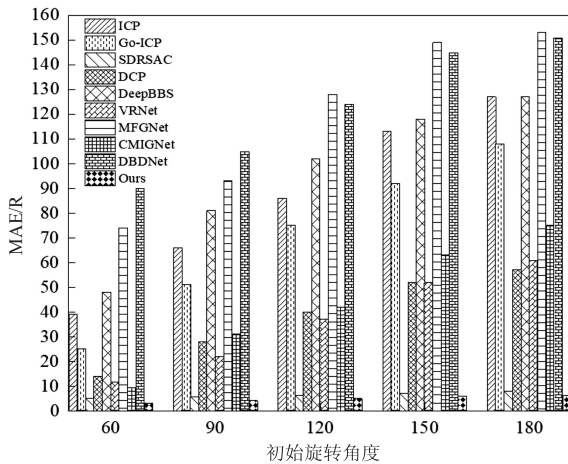


图6 在ModelNet40数据集上不同初始旋转角度的旋转误差

为了验证提出的模块对于整体模型的作用,对注意力机制和匹配矩阵进行了消融实验。该实验保持了与高噪声点云实验相同的参数设置,实验结果如表4所示。表中第一行表示既不加入注意力模块也不加入匹配矩阵优化模块,可以看出模型精度最低;模型中加入匹配矩阵优化模块(MMO)而缺少双重注意力模块时,配准的精度有所改善;模型中加入双重注意力模块而缺少匹配矩阵优化时,配准的精度相较于只加入匹配矩阵优化模块时的精度有一定的提高;当把双重注意力模块和匹配矩阵优化模块相结合时的模型配准效果最佳,这也说明了双重注意力模块和匹配矩阵优化模块对配准的精度改善最有效。

表4 高噪声点云配准消融实验结果

MMO	DCAM	RMSE (R)	MAE (R)	RMSE (t)	MAE (t)
		4.933 5	3.242 6	0.020 9	0.016 5
✓		2.327 6	2.082 6	0.010 6	0.007 4
	✓	1.513 8	1.145 2	0.004 6	0.003 8
✓	✓	0.665 7	0.340 2	0.001 7	0.001 3

3 结束语

该文提出了一种基于双重注意力和匹配矩阵优化的点云配准算法。通过DGCNN提取特征,使用Transformer获得具有融合性的特征,DCAM模块降低模型中存在的噪声,然后通过匹配矩阵优化模块可以降低配准过程中的误差从而使得点云配准过程更加精确,最后通过SVD分解层得到刚性变换。在ModelNet40、7Scenes和KITTI数据集上进行了多次实验。结果表明该模型相较于其他模型表现出了更好的性能,这也说明了模型具有更强的鲁棒性以及更高的配准精度。后续的工作考虑在保持配准精度的同时使模型可以更加轻量化。

参考文献:

- [1] CHEN L, FENG C, MA Y, et al. A review of rigid point cloud registration based on deep learning [J]. *Frontiers in Neurobotics*, 2024, 17: 1281332.
- [2] 翁弘毅. 基于激光雷达点云配准的三维重建技术研究 [D]. 北京: 北京邮电大学, 2024.
- [3] CHEN Z, LI C, ZUO W, et al. Study on improved point cloud registration algorithm enhanced by double down-sampling [J]. *Automot Eng*, 2023, 45(4): 572-578.
- [4] XU X, WANG P, GAN X, et al. Automatic marker-free registration of single tree point-cloud data based on rotating projection [J]. *Artificial Intelligence in Agriculture*, 2022, 6(1): 176-188.
- [5] 王 宾, 刘 林, 侯榆青, 等. 应用改进迭代最近点方法的三维心脏点云配准 [J]. *光学精密工程*, 2020, 28(2): 474-484.
- [6] 陆卫刚, 周治平. 面向增强现实的点云配准算法 [J]. *激光与光电子学进展*, 2019, 56(19): 263-269.
- [7] BESL P J, MCKAY N D. A method for registration of 3-D shapes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(2): 239-256.
- [8] YANG J, LI H, CAMPBELL D, et al. Go-icp: a globally optimal solution to 3D ICP point-set registration [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(11): 2241-2254.
- [9] LE H M, DO T, HOANG T, et al. Sdrsac: semidefinite-based randomized approach for robust point cloud registration with-

- out correspondences [C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. California: IEEE, 2019: 124–133.
- [10] LI J, HU Q, AI M. Point cloud registration based on one-point ransac and scale-annealing biweight estimation [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11): 9716–9729.
- [11] FISCHLER M A, BOLLSE 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.
- [12] 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. Seoul: IEEE, 2019: 3523–3532.
- [13] HEZRONI I, DRORY A, GIRYES R, et al. Deepbbs: deep best buddies for point cloud registration [C] // Proceedings of the 2021 international conference on 3D vision (3DV). London: IEEE, 2021: 342–351.
- [14] ZHANG Z, SUN J, DAI Y, et al. Vmnet: learning the rectified virtual corresponding points for 3D point cloud registration [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(8): 4997–5010.
- [15] WANG H, LIU X, KANG W, et al. Multi-features guidance network for partial-to-partial point cloud registration [J]. Neural Computing and Applications, 2022, 34(2): 1623–1634.
- [16] XIE Y, ZHU J, LI S, et al. Cross-modal information-guided network using contrastive learning for point cloud registration [J]. IEEE Robotics and Automation Letters, 2023, 9(1): 103–110.
- [17] 刘旭珩, 柏正尧, 许 祝, 等. 结合注意力机制的多重引导点云配准网络 [J]. 计算机科学, 2024, 51(2): 142–150.
- [18] YUAN Y, WU Y, FAN X, et al. Egst: enhanced geometric structure transformer for point cloud registration [J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 30(9): 6222–6234.
- [19] 谢帅康, 熊风光, 朱新杰, 等. 基于空间可变形 Transformer 的三维点云配准方法 [J]. 计算机工程, 2024, 50(3): 224–232.
- [20] LI S, ZHU J, XIE Y. Dbdnet: partial-to-partial point cloud registration with dual branches decoupling [J]. Knowledge-Based Systems, 2024, 296: 111864.
- [21] WU Z, SONG S, KHOSLA A, et al. 3D shapenets: a deep representation for volumetric shapes [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 1912–1920.
- [22] SHOTTON J, GLOCKER B, ZACH C, et al. Scene coordinate regression forests for camera relocalization in RGB-D images [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. Oregon: IEEE, 2013: 2930–2937.
- [23] GEIGER A, LENZ P, URTASUN R. "Are we ready for autonomous driving? The KITTI vision benchmark suite" [C] // Proceedings of the 2012 IEEE conference on computer vision and pattern recognition. Rhode Island: IEEE, 2012: 3354–3361.
- [24] 赵夫群, 耿国华. 基于图像特征和奇异值分解的点云配准算法 [J]. 激光与光电子学进展, 2020, 57(10): 237–243.
- [25] HODSON T O. Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not [J]. Geoscientific Model Development Discussions, 2022, 15: 5481–5487.