

北京理工大学

本科生毕业设计（论文）

基于深度学习的端到端多实例点云配准

Deep Learning Based End-To-End Multi-instance Point Cloud Registration

学 院：	自动化学院
专 业：	自动化
班 级：	06111902
学生姓名：	杨润一
学 号：	1120191211
指导教师：	由育阳

2023 年 5 月 2 日

原创性声明

本人郑重声明：所呈交的毕业设计（论文），是本人在指导老师的指导下独立进行研究所取得的成果。除文中已经注明引用的内容外，本文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

特此申明。

本人签名：

日期：

年

月

日

关于使用授权的声明

本人完全了解北京理工大学有关保管、使用毕业设计（论文）的规定，其中包括：①学校有权保管、并向有关部门送交本毕业设计（论文）的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存本毕业设计（论文）；③学校可允许本毕业设计（论文）被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换本毕业设计（论文）；⑤学校可以公布本毕业设计（论文）的全部或部分内容。

本人签名：

日期：

年

月

日

指导老师签名：

日期：

年

月

日

基于深度学习的端到端多实例点云配准

摘 要

本文……。

摘要正文选用模板中的样式所定义的“正文”，每段落首行缩进 2 个字符；或者手动设置成每段落首行缩进 2 个汉字，字体：宋体，字号：小四，行距：固定值 22 磅，间距：段前、段后均为 0 行。阅后删除此段。

摘要是一篇具有独立性和完整性的短文，应概括而扼要地反映出本论文的主要内容。包括研究目的、研究方法、研究结果和结论等，特别要突出研究结果和结论。中文摘要力求语言精炼准确，本科生毕业设计（论文）摘要建议 300-500 字。摘要中不可出现参考文献、图、表、化学结构式、非公知公用的符号和术语。英文摘要与中文摘要的内容应一致。阅后删除此段。

关键词：点云配准；多实例；聚类；对应聚类；深度学习

Deep Learning Based End-To-End Multi-instance Point Cloud Registration

Abstract

In order to study……

Abstract 正文设置成每段落首行缩进 2 字符，字体：Times New Roman，字号：小四，行距：固定值 22 磅，间距：段前、段后均为 0 行。阅后删除此段。

Key Words: Point Cloud Registration; Multi-instance; Clustering; Correspondence Clustering; Deep Learning

目 录

摘 要	I
Abstract	II
第 1 章 绪论	1
1.1 研究背景和意义	1
1.1.1 三维点云配准	1
1.1.2 多实例点云配准	2
1.2 国内外研究现状	3
1.2.1 点云配准	3
1.2.2 三维目标检测和实例分割	4
1.2.3 多模型拟合	4
1.2.4 发展趋势	5
1.3 论文结构安排	5
1.4 小结	5
第 2 章 点云处理与配准	6
2.1 点云处理	6
2.2 点云配准	6
2.3 深度学习方法在点云中的应用	6
2.4	6
结 论	8
参考文献	9
附 录	11
附录 A L ^A T _E X 环境的安装	11
附录 B BITHesis 使用说明	11
致 谢	12

第 1 章 绪论

1.1 研究背景和意义

1.1.1 三维点云配准

21 世纪以来，人工智能技术的发展对于社会有着重大的影响，智能化成为工程技术突破的内核。机器能够进行快速计算、存储和处理大量数据，并通过互联网将社会连为一体。现在由人工智能驱动的新一代机器，它们可以越来越自主地解决复杂的任务，其中以视觉为核心的机器技术快速发展，机械臂、自动驾驶、自主运动机器人等进入了人们的视野。随着 2012 年 AlexNet^[1]问世以来，深度学习方法打开了计算机视觉的新大门。越来越多的深度学习方法比如 VGG^[2]、ResNet^[3]、ViT^[4]被用在了图像分类、分割、场景理解等任务中。为了更好的理解真实世界，人们开始尝试将深度学习方法用于三维数据中，随着激光雷达和 Kinect 等高精度传感器的快速发展，点云已经成为表示三维世界的主要数据格式。2017 年 PointNet^[5]出现后，深度学习方法也同样被广泛应用在了点云处理中。

三维点云配准是点云处理中的一项基本任务^[5-7]，其在机械臂、自动驾驶、自主运动机器人等众多基于视觉方法的应用中起着关键的作用。首先是三维重建，生成完整的三维场景是各种计算机视觉应用的基础和重要技术，包括自动驾驶中的高精度三维地图重建、机器人技术中的三维环境重建等。例如，配准可以为机器人应用程序中的路线规划和决策构建三维环境。

其次，三维场景中的定位。三维场景中的定位和重定位对于机器人技术尤其重要。例如，无人驾驶汽车会估计其在地图上的位置及其与道路边界线的距离。点云配准可以将当前的实时三维视图与其所属的三维环境准确匹配，提供高精度定位服务。此应用表明，点云配准提供了机器和三维环境交互一种解决方案。

第三，位姿估计。将点云 A 与另一个点云 B 对齐可以生成与点云 B 相关的点云 A 的位姿信息。这个位姿信息可用于机器人决策。例如，点云配准可以获取环境中物体的位姿信息，以决定机械臂移动到哪里以准确抓取并移动物体。位姿估计为机器人三维环境理解提供了重要信息。

在国防安全、信息安全、环境安全等领域，无人机系统、自主导航、环境感知等技术应用愈发广泛。在这些应用中，点云配准也发挥着重要作用。例如，无人机系统

需要对目标进行跟踪，而点云配准可以提供目标的位姿信息，来实现目标跟踪。点云配准可以用于环境感知，用于分割、检测、识别等任务，从而实现环境感知。在自动驾驶、机器人自主导航中，高精度的点云配准算法可以提供高精度的 3D 地图场景重建，为自主机器提供视觉定位、路径规划、障碍物检测等技术保障^[8]。

1.1.2 多实例点云配准

点云配准旨在通过对源点云和目标点云之间进行刚性变换，使得源点云和目标点云尽可能重合。点云配准的输入是两个点云，输出是一个刚性变换矩阵。点云配准的目标是找到一个刚性变换矩阵，使得源点云和目标点云之间的距离最小。传统方法中，一般流程为查找匹配点，通过 SVD 等方法求解出变换矩阵。随着机器学习和深度学习算法的广泛使用，基于深度学习方法和组合优化方法进一步提高了点云配准的准确率^[9-11]。

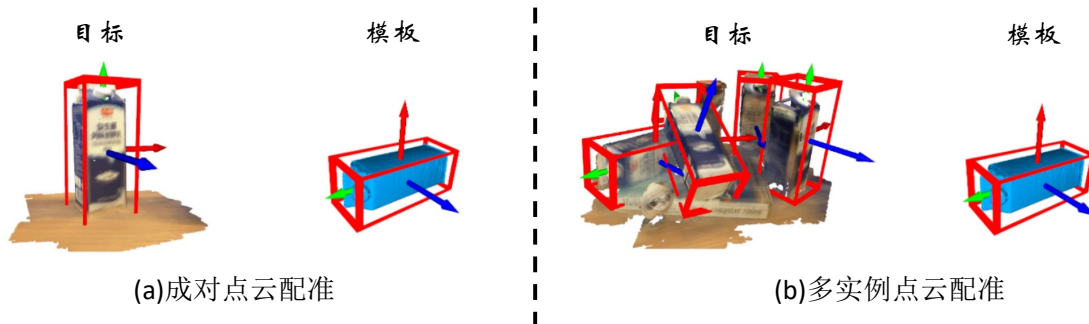


图 1-1 多实例点云配准：给定目标的模板点云，成对点云配准（左）侧重于估计模板点云和目标点云之间的单个刚性变换，而多实例点云配准（右）旨在估计目标点云中相同物体的 6D 位姿。

目前大多数点云配准任务研究主要集中在成对配准上。然而，在实际应用中，目标场景可能包含多个重复实例，我们需要估计模板点云与目标点云中这些重复实例之间的多个刚性变换。比如说在室内场景中，我们希望机器人能够将屋子中所有的椅子摆正，那么首先需要将多个椅子点云和模板椅子点云进行配准，求的目标椅子的位姿，通过机械运动来达到位姿改变的效果。图1-1展示了一个示例。这个问题被命名为多实例点云配准，它比成对点云配准更具挑战性。针对该任务已有的现有文献研究较少，扩展现有的点云配准方法来解决这个问题并非易事。多实例点云配

准不仅需要从嘈杂的对应中拒绝异常值，还需要识别单个实例的异常值集，这使得它比传统的配准问题更具挑战性。

与传统的两两配准方法相比，多实例点云配准需要解决更复杂的问题，同时也具有更广泛的应用价值。比如，在大规模场景重建任务中，通常需要处理成千上万个点云数据。单纯采用两两配准的方法可能导致累积误差，从而影响重建结果的精度。因此，研究多实例点云配准算法具有重要的实际意义。在机械臂抓取任务中，多实例点云配准算法可以在全局范围内考虑点云之间的约束关系，有助于消除局部误差和噪声的影响，从而提高配准结果的鲁棒性^[12]。

尽管多实例点云配准技术在近年来取得了显著的进展，仍然存在许多亟待解决的问题。例如，现存的多实例点云配准一般采用多任务的方式，也就是先对点云分割或者三维目标检测，然后进行两两点云配准，这样的方式需要先训练点云分割或者目标检测网络，泛化性差。并且如果见到了不存在先验的点云，下游的配准任务仍然会失效。所以，本文我们会对多实例点云配准进行研究，通过点云直接进行多实例点云配准，不需要先进行点云分割或者目标检测，从而提高多实例点云配准的泛化性。

1.2 国内外研究现状

1.2.1 点云配准

点云配准长期以来一直是计算机视觉和机器人领域的一项基本任务，大致可分为直接方法 [33, 34, 35] 和基于特征的方法 [5, 26, 36]。近年来，由于深度学习的发展，许多基于特征的方法取得了最先进的性能。这些方法通常通过特征匹配产生对应关系，然后移除异常值以稳健地估计转换。尽管深度特征 [5, 26, 36, 37] 发展迅速，但特征匹配生成的对应关系仍然包含异常值。因此，去除异常点在点云配准中具有重要意义。过去，已经提出了许多传统方法来去除异常值，包括基于 RANSAC 的方法 [15, 16, 19]、基于分支和边界的方法 [38] 以及许多其他方法 [26, 39]。最近，一系列基于学习的方法 [36, 40] 被提出，并在异常值去除方面取得了显著的效果。以上的方法都是基于成对点云配准来完成的。然而，与成对配准不同，一个实例的内点构成多实例点云配准中所有其他实例的异常值。这种伪异常值使得很难将上述二元分类模型直接推广到多实例点云配准的情况。现有该问题解决方案包括采用目标检测方法或对目标点云应用实例分割，将多实例点云配准问题转化为多个成对点云配准问

题，但是这种方法需要预先训练一个目标检测或者点云分割网络，这样的方法对于已有点云类别是有效的，但是对于未知的类别是不适用的。另一种解决方案是通过多模型拟合，但是现有的多模型拟合方法依赖于抽样有效假设，当模型数量或离群率变高时，会涉及大量的抽样步骤，使得这些算法的效率和鲁棒性急剧下降。

1.2.2 三维目标检测和实例分割

三维物体的目标检测和实例分割与多实例点云配准有着密切的关系。输入一帧点云，目标检测模型^[13]可以用来对获取每个目标对象的边界框，三维实例分割^[14-15]为每个点生成实例标签。

这样的方法产生的结果类似于多实例注册的结果，但是它们需要将特定对象或类别的先验训练到网络中。基于点云匹配的筛选和聚类方法来进行多实例点云配准通过直接将模板点云和目标点云中的多个实例对齐来处理两组点云，而不使用任何关于输入的点云的先验信息。

1.2.3 多模型拟合

多实例配准也可以通过多模型拟合来实现，其目的是根据多个模型生成的数据点来进行建模。例如在点云中拟合多个平面^[16]，在运动分割中估计基本矩阵^[17]，在多实例点云配准中计算刚性变换^[18]等。但是由于一个实例的正常值构成所有其他实例的离群值，所以多模型拟合比单模型拟合更具挑战性。

现有的多模型拟合方法大致可以分为两类。第一类按顺序拟合模型^[19-22]，通过重复采样和筛选模型来进行建模。比如，Progressive-X^[19]和 Progressive-X+^[20]使用了表现更好的 Graph-cut RANSAC^[23]作为采样方法来生成假设。CONSAC^[22]首次将深度模型引入多模型拟合中，使用类似 PointNet^[5]的网络来引导采样。通过重复采样来恢复单个实例，从输入中删除正常值和，以顺序的方式来检测实例。

第二类模型同时拟合多个模型^[18,24-27]。许多基于偏好分析的方法^[24,27]最初对一系列假设进行采样，然后根据假设的残差对输入点进行聚类。ECC^[18]利用点云刚性变换空间一致性^[28]和以自下而上的方式基于距离不变矩阵对对应关系进行聚类。PointCLM^[29]使用了一种新的深层表示方法来与空间一致性相结合，得到了更好的结果。

1.2.4 发展趋势

多实例点云配准目前的方法主要集中于多模型拟合。检测/分割 + 多对点云配准的方法在泛化性和未知类别中有着较大的技术弱势。国内外的学者在多模型拟合的方法中取得了很好的成果，特别是国内上海交通大学提出的 EEC^[18]和复旦大学提出的 PointCLM^[29]方法在这个任务中取得了目前最好的结果。

1.3 论文结构安排

1.4 小结

本章主要介绍了点云配准任务以及多实例点云配准任务的研究背景和研究意义、国内外研究现状。然后介绍了本文的研究内容和结构安排。

第 2 章 点云处理与配准

2.1 点云处理

2.2 点云配准

2.3 深度学习方法在点云中的应用

2.4

```
1 import numpy as np
2
3 def incmatrix(genl1,genl2):
4     m = len(genl1)
5     n = len(genl2)
6     M = None #to become the incidence matrix
7     VT = np.zeros((n*m,1), int) #dummy variable
8
9     #compute the bitwise xor matrix
10    M1 = bitxormatrix(genl1)
11    M2 = np.triu(bitxormatrix(genl2),1)
12
13    for i in range(m-1):
14        for j in range(i+1, m):
15            [r,c] = np.where(M2 == M1[i,j])
16            for k in range(len(r)):
17                VT[(i)*n + r[k]] = 1;
18                VT[(i)*n + c[k]] = 1;
19                VT[(j)*n + r[k]] = 1;
20                VT[(j)*n + c[k]] = 1;
21
22            if M is None:
23                M = np.copy(VT)
24            else:
25                M = np.concatenate((M, VT), 1)
26
27            VT = np.zeros((n*m,1), int)
28
```

```
return M
```

代码 2.1: Python Code

结 论

参考文献

- [1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv preprint arXiv:1409.1556, 2014.
- [3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [4] 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.
- [5] Qi C R, Su H, Mo K, 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. 2017: 652-660.
- [6] Huang X, Mei G, Zhang J, et al. A comprehensive survey on point cloud registration[J]. ArXiv preprint arXiv:2103.02690, 2021.
- [7] Besl P J, McKay N D. Method for registration of 3-D shapes[C]//Sensor fusion IV: control paradigms and data structures: vol. 1611. 1992: 586-606.
- [8] 周慧子, 胡学敏, 陈龙, 等. 面向自动驾驶的动态路径规划避障算法[J]. 计算机应用, 2017, 37(883-888).
- [9] Deng H, Birdal T, Ilic S. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 602-618.
- [10] Deng H, Birdal T, Ilic S. Ppfnet: Global context aware local features for robust 3d point matching[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 195-205.
- [11] Qin Z, Yu H, Wang C, et al. Geometric transformer for fast and robust point cloud registration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 11143-11152.
- [12] Stücker J, Behnke S. Robust real-time registration of RGB-D images using multi-resolution surfel representations[C]//ROBOTIK 2012; 7th German Conference on Robotics. 2012: 1-4.
- [13] Qi C R, Litany O, He K, et al. Deep hough voting for 3d object detection in point clouds[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9277-9286.
- [14] Wang W, Yu R, Huang Q, et al. Sgpn: Similarity group proposal network for 3d point cloud instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2569-2578.
- [15] Han L, Zheng T, Xu L, et al. Occuseg: Occupancy-aware 3d instance segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 2940-2949.
- [16] Barath D, Matas J. Multi-class model fitting by energy minimization and mode-seeking[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 221-236.
- [17] Hartley R I. In defense of the eight-point algorithm[J]. IEEE Transactions on pattern analysis and machine intelligence, 1997, 19(6): 580-593.

- [18] Tang W, Zou D. Multi-instance point cloud registration by efficient correspondence clustering[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 6667-6676.
- [19] Barath D, Matas J. Progressive-x: Efficient, anytime, multi-model fitting algorithm[C]// Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3780-3788.
- [20] Barath D, Rozumny D, Eichhardt I, et al. Progressive-x+: Clustering in the consensus space[J]. ArXiv preprint arXiv:2103.13875, 2021.
- [21] Kanazawa Y, Kawakami H. Detection of planar regions with uncalibrated stereo using distributions of feature points.[C]//BMVC. 2004: 1-10.
- [22] Kluger F, Brachmann E, Ackermann H, et al. Consac: Robust multi-model fitting by conditional sample consensus[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 4634-4643.
- [23] Barath D, Matas J. Graph-cut RANSAC[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6733-6741.
- [24] Toldo R, Fusiello A. Robust multiple structures estimation with j-linkage[C]//Computer Vision—ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10. 2008: 537-547.
- [25] Magri L, Fusiello A. Multiple model fitting as a set coverage problem[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3318-3326.
- [26] Magri L, Fusiello A. T-linkage: A continuous relaxation of j-linkage for multi-model fitting[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 3954-3961.
- [27] Magri L, Andrea F, et al. Robust multiple model fitting with preference analysis and low-rank approximation[C]//Proceedings of the British Machine Vision Conference 2015. 2015: 20-1.
- [28] Leordeanu M, Hebert M. A spectral technique for correspondence problems using pairwise constraints[C]//Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1: vol. 2. 2005: 1482-1489.
- [29] Yuan M, Li Z, Jin Q, et al. PointCLM: A Contrastive Learning-based Framework for Multi-instance Point Cloud Registration[C]//Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX. 2022: 595-611.

附 录

附录相关内容…

附录 A L^AT_EX 环境的安装

L^AT_EX 环境的安装。

附录 B B^IThesis 使用说明

B^IThesis 使用说明。

附录是毕业设计（论文）主体的补充项目，为了体现整篇文章的完整性，写入正文又可能有损于论文的条理性、逻辑性和精炼性，这些材料可以写入附录段，但对于每一篇文章并不是必须的。附录依次用大写正体英文字母 A、B、C……编序号，如附录 A、附录 B。阅后删除此段。

附录正文样式与文章正文相同：宋体、小四；行距：22 磅；间距段前段后均为 0 行。阅后删除此段。

致 谢

值此论文完成之际，首先向我的导师……

致谢正文样式与文章正文相同：宋体、小四；行距：22 磅；间距段前段后均为 0 行。阅后删除此段。