

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 8 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 点云配准相关背景知识介绍	6
2.1 点云数据	6
2.1.1 点云数据特点	6
2.1.2 点云特征描述	7
2.2 刚体运动参数估计	8
2.2.1 旋转矩阵	8
2.2.2 欧拉角	9
2.2.3 四元数	11
2.2.4 基于奇异值分解的线性代数求解	12
2.2.5 基于 Levenberg-Marquardt 算法的非线性优化求解	13
2.3 标准数据集	15
2.3.1 ModelNet40	15
2.3.2 ShapeNet	15
2.3.3 Scan2CAD	15
2.4 点云配准评价指标	15
2.4.1 基于特征提取与匹配的评价指标	15
2.4.2 基于刚体运动的评价指标	15
2.4.3 多实例点云配准评价指标	15

2.5 点云处理与配准相关深度学习技术	15
2.5.1 深度学习概述	15
2.5.2 多层感知机	15
2.5.3 卷积神经网络	15
2.5.4 注意力机制	15
2.5.5 谱聚类算法	15
2.6 小结	15
结 论	17
参考文献	18
附 录	21
附录 A \LaTeX 环境的安装	21
附录 B BITHesis 使用说明	21
致 谢	22

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]。

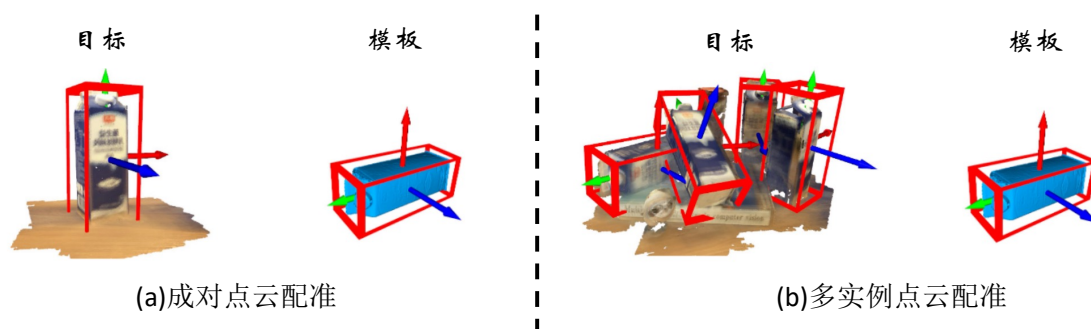


图 1-1 多实例点云配准：给定目标的模板点云，成对点云配准（左）侧重于估计模板点云和目标点云之间的单个刚性变换，而多实例点云配准（右）旨在估计目标点云中相同物体的 6D 位姿。

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 点云配准相关背景知识介绍

本章将对点云配准的基本理论和相关背景知识进行系统阐述，包括配准的运动学解析、常用数据集、常用配准算法以及相关深度学习方法，并且我们会基于点云配准的评价指标，扩展到多实例点云配准的评价指标。

2.1 点云数据

三维世界中的数据表征有多种类型，如点云数据 (Point Clouds)^[30]、三角网格 (Triangle Mesh)^[31]、体素 (Voxel Grids)^[32]、深度相机数据 (RGB-D Camera)^[33]等，可视化数据如图2-1所示，本文主要研究的是点云数据，因此本节将对点云数据进行简要介绍。

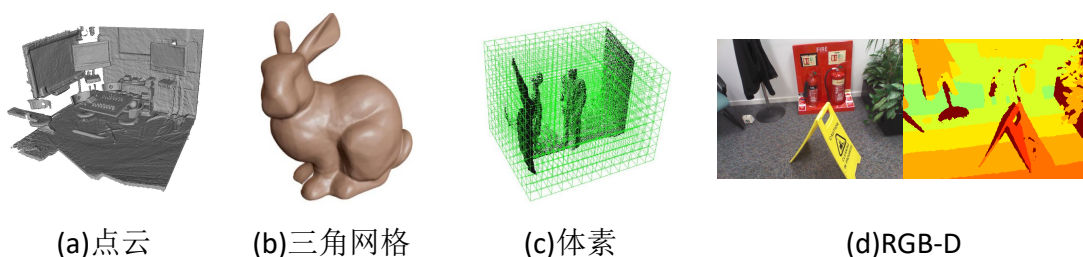


图 2-1 点云、三角网格、体素、深度相机数据可视化示意图

2.1.1 点云数据特点

点云数据是一种常用的三维数据表征形式，由一系列具有三维坐标 (X, Y, Z) 的点组成。这些点可以通过激光雷达、结构光传感器、多视角立体视觉等方式采集。点云数据具有以下特点：

1. 无序性：点云数据通常是无序的，即点在数据结构中的顺序并不表示它们在空间中的相对位置。这使得处理点云数据时需要额外关注邻域搜索等问题。
2. 不完整性：由于采集设备的视野限制以及物体遮挡等因素，点云数据往往只能捕捉到物体表面的部分信息，无法完整地描述物体的几何结构。
3. 稀疏性：点云数据在空间中分布可能是不均匀的，有些区域可能点密集，而有些区域可能点稀疏。这会对点云处理算法的性能产生影响。

4. 噪声敏感性：点云数据容易受到测量噪声、环境光照等因素的影响。为了提高数据质量，通常需要对点云进行预处理，如滤波、降采样等。

5. 缺乏拓扑信息：点云数据仅包含物体表面的几何信息，不包含拓扑信息。在需要考虑物体结构的应用场景中，点云数据需要进行进一步处理，如重建三角网格或提取骨架结构等。

6. 可扩展性：点云数据可以方便地扩展以包含其他属性信息，如颜色、法向量、强度等。这有助于提高点云处理算法的性能和鲁棒性。

7. 易于处理和存储：由于点云数据直接表示了物体表面的几何信息，其数据结构简单，便于处理和存储。此外，点云数据可以通过各种数据结构（如 KD 树、八叉树等）进行高效的组织和检索。

在应用深度学习算法时，由于点云数据的无序性，深度学习算法，比如全连接层、卷积层等无法直接应用于点云数据。因此，需要对点云数据进行预处理，将其转换为有序的数据表征形式，如三角网格^[31]、体素^[32]等。但是这样的方法会造成空间中很多点的浪费，使得输入的数据变得更加稀疏，使得训练数据变少。PointNet^[5]是一种用于处理点云数据的深度学习网络结构，于 2017 年首次提出。它是一个端到端的神经网络，可以直接从原始点云数据中学习特征表示。PointNet 通过使用对称函数（如最大池化）处理输入点云的无序性，同时具有对输入点顺序的不变性。

2.1.2 点云特征描述

三维点云的特征描述，也叫描述子 (Descriptor)，是一种用于表示点云数据中每个点周围的局部几何特征的向量。描述子捕捉了点云中每个点的几何结构和形状信息，这对于解决诸如点云配准、物体识别、分类和分割等问题至关重要。点云描述子应具有以下特性：鲁棒性、区分性、旋转不变性、尺度不变性和噪声不敏感性。点云描述子对于点云的后处理有着非常巨大的影响，对于不同的任务和数据特征，应该选用合适的描述子来作为网络的预处理。

有许多不同类型的点云描述子，它们根据计算方法和考虑的几何属性而有所不同。以下是一些常见的点云描述子：

1. Spin Images (旋转图像)^[34]：通过在点云中每个点周围投影二维图像来表示局部形状信息。

2. Normal Aligned Radial Features (NARF)^[35]：基于局部表面法线的方向和强度

来描述点云中的局部特征。

3. Fast Point Feature Histograms (FPFH)^[36]: 通过计算每个点周围的点对的几何特征直方图来表示局部特征。

4. Signature of Histograms of Orientations (SHOT)^[37]: 结合局部点的颜色信息和表面法线分布, 为每个点计算描述子。

5. Point Pair Features (PPF)^[10]: 描述点云中两个点之间的几何关系, PPF 是一个四元组 $(\alpha, d, \theta, \phi)$, 其中 α 是两点之间的距离, d 是两点的法向量之差, θ 是两点的法线之间的角度, ϕ 是两点的法线在两点之间连线所确定的平面上的角度。PPF 特征在处理点云配准问题时具有很高的鲁棒性, 因为它仅依赖于点云的几何信息。

2.2 刚体运动参数估计

刚体运动参数估计是点云配准任务的最后一个阶段的任务。对于位移来说, 常用的是位移向量 (Transition); 旋转的运动参数有不同的表示形式, 如欧拉角^[38]、四元数^[39]、旋转矩阵^[40]、轴角^[41]等。下面我们对旋转的表示进行简要介绍。

2.2.1 旋转矩阵

旋转矩阵是一种与向量相乘时旋转向量同时保持其长度的矩阵。所有 3×3 旋转矩阵的特殊正交群表示为 $SO(3)$ 。因此, 如果 $\mathbf{R} \in SO(3)$, 那么

$$\det(\mathbf{R}) = \pm 1 \quad \text{且} \quad \mathbf{R}^{-1} = \mathbf{R}^T \quad (2-1)$$

对于满足 $\det(\mathbf{R}) = 1$ 的旋转矩阵, 称为正规旋转; 对于满足 $\det(\mathbf{R}) = -1$ 的旋转矩阵, 称为非正规旋转。非正规旋转也称为旋转倒数, 由旋转后接反演操作组成。我们将分析限制在正规旋转上, 因为非正规旋转不是刚体变换。我们按如下方式引用旋转矩阵的元素:

$$\mathbf{R} = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} \quad (2-2)$$

$$= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (2-3)$$

对于正规旋转矩阵，我们可以使用群的定义来进行更精准的定义。

$$SO(n) = \{\mathbf{R} \in \mathbf{R}^{n \times n} | \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = 1\} \quad (2-4)$$

其中， $SO(n)$ 是特殊正交群 (Special Orthogonal Group)，这个集合由 n 维空间的旋转矩阵组成。 $SO(n)$ 是一个群，因为它满足群的四个条件：封闭性、结合律、单位元和逆元。 $SO(n)$ 的单位元是 \mathbf{I} ，逆元是 \mathbf{R}^T 。 $SO(n)$ 的元素称为正规旋转矩阵。通过旋转矩阵可以直接描述相机的旋转。

为了描述两个坐标之间的相对旋转，我们假设某个单位正交基 $\mathbf{i}, \mathbf{j}, \mathbf{k}$ ，并且我们希望将其旋转到另一个正交基 $\mathbf{i}', \mathbf{j}', \mathbf{k}'$ 。假设对于同一个向量 $\mathbf{a} = [a_1, a_2, a_3]^T$ ，在两个坐标系中的表示分别为 $\mathbf{a} = a_1 \mathbf{i} + a_2 \mathbf{j} + a_3 \mathbf{k}$ 和 $\mathbf{a}' = a'_1 \mathbf{i}' + a'_2 \mathbf{j}' + a'_3 \mathbf{k}'$ ，根据坐标的定义，有

$$[\mathbf{i}, \mathbf{j}, \mathbf{k}] \mathbf{a} = [\mathbf{i}', \mathbf{j}', \mathbf{k}'] \mathbf{a}' \quad (2-5)$$

我们对上述等式同时左成 $[\mathbf{i}, \mathbf{j}, \mathbf{k}]^T$ ，我们可以通过旋转矩阵 \mathbf{R} 来表示两个坐标系之间的旋转关系，即

$$\mathbf{a}' = \begin{bmatrix} \mathbf{i}^T \mathbf{i}' & \mathbf{i}^T \mathbf{j}' & \mathbf{i}^T \mathbf{k}' \\ \mathbf{j}^T \mathbf{i}' & \mathbf{j}^T \mathbf{j}' & \mathbf{j}^T \mathbf{k}' \\ \mathbf{k}^T \mathbf{i}' & \mathbf{k}^T \mathbf{j}' & \mathbf{k}^T \mathbf{k}' \end{bmatrix} \mathbf{a} = \mathbf{R} \mathbf{a} \quad (2-6)$$

因为旋转矩阵是正交矩阵，它的逆就可以用来描述一个相反的旋转，按照上面的推到，则有：

$$\mathbf{a} = \mathbf{R}^T \mathbf{a}' = \mathbf{R}^{-1} \mathbf{a}' \quad (2-7)$$

2.2.2 欧拉角

三次坐标旋转依次可以描述任意旋转。我们考虑三次旋转，其中第一次旋转是关于 \mathbf{k} 轴的角度 ψ ，第二次旋转是关于 \mathbf{j} 轴的角度 θ ，第三次旋转是关于 \mathbf{i} 轴的角度 ϕ ，如图 2-2 (a, b, c) 所示。为了简化符号，我们将这些角度排列成一个三维向量，称为欧拉角向量，定义为

$$\mathbf{u} := [\phi, \theta, \psi]^T \quad (2-8)$$

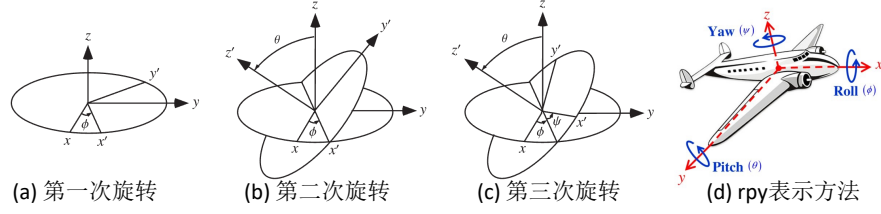


图 2-2 欧拉角旋转表示

将欧拉角向量映射到其对应的旋转矩阵的函数， $\mathbf{R}_{ijk} : \mathbf{R}^3 \rightarrow SO(3)$ ，定义为

$$\mathbf{R}_{ijk}(\phi, \theta, \psi) := \mathbf{R}_i(\phi)\mathbf{R}_j(\theta)\mathbf{R}_k(\psi) \quad (2-9)$$

与一般情况相同，如果 $\mathbf{z} \in \mathbf{R}^3$ 是世界坐标系中的一个向量， $\mathbf{z}_0 \in \mathbf{R}^3$ 是以体固定坐标表示的相同向量，那么以下关系成立：

$$\mathbf{z}_0 = \mathbf{R}_{ijk}(\mathbf{u})\mathbf{z} \quad (2-10)$$

$$\mathbf{z} = \mathbf{R}_{ijk}(\mathbf{u})^T \mathbf{z}_0 \quad (2-11)$$

也就是说，欧拉角并不是定义在三维线性空间的群，欧拉角的相互转换需要借助旋转矩阵来进行表达。

在自动化或者航空航天领域比较常用的欧拉角的分解方式是 $Z - Y - X$ ，也就是说，先绕 z 轴旋转 ψ 角，然后绕 y 轴旋转 θ 角，最后绕 x 轴旋转 ϕ 角。这种分解方式也被称为航空分解（Aerospace decomposition），通常这种方式中的 $Z - Y - X$ 会被称为“偏航 - 俯仰 - 旋转 (yaw - pitch - roll)”。在这种分解方式中，欧拉角向量的顺序是 $\mathbf{u} = [\psi, \theta, \phi]^T$ ，如图 2-2 (d) 所示。

欧拉角的一个缺点是会碰到著名的万象锁问题^[42]，也就是说，当 $\theta = \pm \frac{\pi}{2}$ 时，旋转矩阵 $\mathbf{R}_{ijk}(\mathbf{u})$ 的值会变成奇异的。理论上可以证明，只要想要用 3 个数来表达三位旋转时，基本都会碰到奇异性问题^[43]。由于这种原理，欧拉角不适用于迭代和插值，在传统的 SLAM 中和深度学习中几乎都不会是用这种方式来进行迭代，但是欧拉角对于人机交互是友好的。

2.2.3 四元数

旋转矩阵使用了 9 个未知数来描述 3 自由度的旋转，这在数学表达上具有冗余性了；欧拉角使用了 3 个未知数来描述 3 自由度的旋转，在数学表达上是紧凑的，但是具有奇异性。因为我们找不到不带奇异性，且使用 3 个未知数的方法来描述 3 维旋转的描述方式^[43]。这个性质，可以类比于用两个坐标描述地球表面（二维流形）上的点，我们可以使用经纬度来描述地球表面上的点，但是在极点处，经纬度的表达就会出现奇异性，即纬度为 $\pm 90^\circ$ 时，经度无意义。

三位旋转是一个三维流形，因此，为了最优平衡无奇异性性和数学表达的紧凑性，我们需要引入一个新的数学工具来描述 3 自由度的旋转，这个数学工具就是四元数。四元数是 Hamilton 找到的一种超复数，它在数学形式上既是紧凑的，又是非奇异的。四元数的定义如下：

$$\mathbf{q} = q_0 + q_1i + q_2j + q_3k \quad (2-12)$$

其中， $q_0, q_1, q_2, q_3 \in R$ ， i, j, k 是四元数的三个虚数单位，满足如下关系：

$$\begin{cases} i^2 = j^2 = k^2 = ijk = -1, \\ ij = k, jk = i, ki = j, \\ ji = -k, kj = -i, ik = -j \end{cases} \quad (2-13)$$

假设某个旋转是绕着单位向量 $\mathbf{u} = [u_x, u_y, u_z]^T$ 旋转 θ 角，那么这个旋转可以用四元数来表示为：

$$\mathbf{q} = \cos \frac{\theta}{2} + \sin \frac{\theta}{2} (u_x i + u_y j + u_z k) \quad (2-14)$$

同时，我们也可以从单位四元数中计算出对应的旋转轴和旋转角：

$$\begin{cases} \theta = 2 \arccos q_0, \\ u_x = \frac{q_1}{\sin \frac{\theta}{2}}, u_y = \frac{q_2}{\sin \frac{\theta}{2}}, u_z = \frac{q_3}{\sin \frac{\theta}{2}} \end{cases} \quad (2-15)$$

因为四元数是定义在复向量空间中而非旋转流形中，并且四元数没有死锁的特性，所以在深度学习的梯度下降算法中，常常用四元数作为回归的目标。并且，因为三维流形的性质，我们在四元数中同乘一个常数，可以得到相同的旋转，即

$[q_0, q_1, q_2, q_3] \iff k[q_0, q_1, q_2, q_3]$ 。由于这个性质，在回归任务中，我们通常回归单位四元数，即 $\sum_0^3 q_i^2 = 1$ 。

用四元数表示旋转。用四元数同样也可以表达对一个点的旋转，假设一个空间中的三维坐标点为 $\mathbf{a} = [x, y, z] \in \mathbb{R}^3$ 。以及一个旋转轴 \mathbf{n} 和旋转角 θ ，这个点绕着旋转轴旋转 θ 角后的坐标为 \mathbf{a}' ，则：

$$\mathbf{p}' = \mathbf{R}\mathbf{p} = \mathbf{q}\mathbf{p}\mathbf{q}^{-1} \quad (2-16)$$

其中， $\mathbf{q} = [\cos \frac{\theta}{2}, \mathbf{n} \sin \frac{\theta}{2}]$ 。

2.2.4 基于奇异值分解的线性代数求解

奇异值分解^[44] (Singular Value Decomposition, SVD) 是线性代数中一种重要的矩阵分解，在信号处理、计算机视觉、运动估计中有着非常广泛的运用。下面，我们对奇异值分解的线性代数部分进行简要介绍。

奇异值分解对一个矩形数据矩阵（定义为 \mathbf{A} ，其中 \mathbf{A} 是一个 $n \times p$ 矩阵）进行处理，其中 n 行代表观察值， p 列代表变量。SVD 定理如下：

$$\mathbf{A}_{n \times p} = \mathbf{U}_{n \times n} \mathbf{S}_{n \times p} \mathbf{V}_{p \times p}^T \quad (2-17)$$

其中

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}_{n \times n} \quad (2-18)$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{I}_{p \times p} \quad (2-19)$$

即 \mathbf{U} 和 \mathbf{V} 是正交的。 \mathbf{U} 的列是左奇异向量（观察值系数向量）； \mathbf{S} （与 \mathbf{A} 的维度相同）具有奇异值，并且是对角的（模振幅）； \mathbf{V}^T 的行是右奇异向量（变量水平向量）。SVD 表示了原始数据在协方差矩阵为对角线的坐标系中的扩展。

计算 SVD 包括寻找 $\mathbf{A}\mathbf{A}^T$ 和 $\mathbf{A}^T\mathbf{A}$ 的特征值和特征向量。 $\mathbf{A}^T\mathbf{A}$ 的特征向量构成 \mathbf{V} 的列， $\mathbf{A}\mathbf{A}^T$ 的特征向量构成 \mathbf{U} 的列。此外， \mathbf{S} 中的奇异值是来自 $\mathbf{A}\mathbf{A}^T$ 或 $\mathbf{A}^T\mathbf{A}$ 的特征值的平方根。奇异值是 \mathbf{S} 矩阵的对角线条目，并按降序排列。奇异值总是实

数。如果矩阵 \mathbf{A} 是实数矩阵，那么 \mathbf{U} 和 \mathbf{V} 也是实数。

奇异值分解（SVD）在点云配准问题中也具有重要应用。点云配准是将两个或多个点云数据集合并为一个统一坐标系的过程。在这种情况下，我们的目标是找到一个最优的刚体变换，包括旋转和平移，使得两个点云之间的距离最小化。

假设我们有两个点云数据集 \mathbf{P} 和 \mathbf{Q} ，每个点云包含 n 个点。我们首先计算两个点云的质心 \mathbf{p}_c 和 \mathbf{q}_c ，然后将点云平移到原点。接下来，我们计算点云 \mathbf{P} 和 \mathbf{Q} 之间的距离矩阵 \mathbf{H} ，其中 $\mathbf{H} = \mathbf{P}^T \mathbf{Q}$ 。

在这种情况下，我们可以使用 SVD 来求解最优的旋转矩阵 \mathbf{R} ，使得两个点云之间的距离最小化。具体来说，我们对距离矩阵 \mathbf{H} 进行奇异值分解：

$$\mathbf{H} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (2-20)$$

然后，我们计算旋转矩阵 \mathbf{R} ：

$$\mathbf{R} = \mathbf{U} \mathbf{V}^T \quad (2-21)$$

如果 \mathbf{R} 是一个不合适的旋转矩阵（即 $\det(\mathbf{R}) \neq 1$ ），我们可以通过修改 \mathbf{S} 矩阵来修复它。具体而言，我们将 \mathbf{S} 的最小奇异值设为 -1 ，然后重新计算旋转矩阵 \mathbf{R} 。

最后，我们可以计算最优的平移向量 \mathbf{t} ：

$$\mathbf{t} = \mathbf{q}_c - \mathbf{R} \mathbf{p}_c \quad (2-22)$$

通过应用旋转矩阵 \mathbf{R} 和平移向量 \mathbf{t} ，我们可以将点云 \mathbf{P} 对齐到点云 \mathbf{Q} ，从而实现点云配准。

2.2.5 基于 Levenberg-Marquardt 算法的非线性优化求解

在数学和计算领域，Levenberg-Marquardt 算法（LMA 或简称 LM）^[45]，也称为阻尼最小二乘法（DLS），用于求解非线性最小二乘问题。这些最小化问题尤其出现在最小二乘曲线拟合中。LMA 在高斯-牛顿算法（GNA）和梯度下降法之间进行插值。LMA 比 GNA 更稳健，这意味着在许多情况下，即使它从距离最终最小值非常远的地方开始，也能找到解决方案。对于行为良好的函数和合理的初始参数，LMA 往往比 GNA 慢。LMA 也可以被视为使用信任区域方法的高斯-牛顿算法。

需要 LM 算法求解的问题被称为非线性最小二乘最小化问题。这意味着要最小化的函数具有以下特殊形式：

$$f(x) = \frac{1}{2} \sum_{j=1}^m r_j^2(x) \quad (2-23)$$

其中 $x = (x_1, x_2, \dots, x_n)^T$ 是一个向量，每个 r_j 是一个从 \mathbb{R}^n 到 \mathbb{R} 的函数。 r_j 被称为残差，并且假定 $m \geq n$ 。

为了简化问题，我们用残差向量 $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 来表示 f ，定义为：

$$r(x) = \begin{bmatrix} r_1(x) & r_2(x) & \cdots & r_m(x) \end{bmatrix} \quad (2-24)$$

现在， f 可以重写为 $f(x) = \frac{1}{2} \|r(x)\|^2$ 。 f 的导数可以用相对于 x 的 r 的雅可比矩阵 $J(x)$ 来表示，定义为 $J(x) = \frac{\partial r_j}{\partial x_i}$ ，其中 $1 \leq j \leq m$ ， $1 \leq i \leq n$ 。

首先考虑线性情况，其中每个 r_i 函数都是线性的。在这里，雅可比矩阵是常数，我们可以将 r 表示为空间中的超平面，这样 f 给出二次型：

$$f(x) = \frac{1}{2} \|Jx - r_0\|^2 \quad (2-25)$$

我们还得到 $\nabla f(x) = J^T(Jx - r_0)$ 和 $\nabla^2 f(x) = J^T J$ 。通过将 $\nabla f(x) = 0$ 求解最小值，我们得到 $x_{\min} = (J^T J)^{-1} J^T r$ ，这是一组无约束优化问题的解。

回到一般的非线性情况，我们有：

$$\nabla f(x) = J(x)^T r(x) \quad (2-26)$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x) \quad (2-27)$$

最小二乘问题的独特属性是，给定雅可比矩阵 J ，如果可以将 r_j 用线性函数近似（ $\nabla^2 r_j(x)$ 较小）或残差 $r_j(x)$ 本身较小，我们可以实质上免费获得海森矩阵（ $\nabla^2 f(x)$ ）。在这种情况下，海森矩阵简化为：

$$\nabla^2 f(x) = J(x)^T J(x) \quad (2-28)$$

这与线性情况相同。这里通常使用的近似是在解附近的 r_i 近似线性，这样 $\nabla^2 r_j(x)$ 较小。还要注意，只有在残差较小时，式 2-28 才有效。对于大残差问题，不能使用二次近似求解^[46]。

考虑到点云配准问题，我们可以将 Levenberg-Marquardt 算法应用于点云之间的非线性最小二乘优化问题。在这种情况下，我们寻求最小化点云之间的距离，同时考虑旋转和平移变换。通过迭代地优化变换参数，Levenberg-Marquardt 算法可以在点云配准问题中找到一个稳定的解决方案。

2.3 标准数据集

2.3.1 ModelNet40

2.3.2 ShapeNet

2.3.3 Scan2CAD

2.4 点云配准评价指标

2.4.1 基于特征提取与匹配的评价指标

2.4.2 基于刚体运动的评价指标

2.4.3 多实例点云配准评价指标

2.5 点云处理与配准相关深度学习技术

2.5.1 深度学习概述

2.5.2 多层感知机

2.5.3 卷积神经网络

2.5.3.1 三维卷积神经网络

2.5.3.2 点云卷积神经网络

2.5.4 注意力机制

2.5.5 谱聚类算法

2.6 小结

```
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
29    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.
- [30] Leberl F, Irschara A, Pock T, et al. Point clouds[J]. Photogrammetric Engineering & Remote Sensing, 2010, 76(10): 1123-1134.
- [31] Jiang C, Sud A, Makadia A, et al. Local implicit grid representations for 3d scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 6001-6010.
- [32] Guan B, Lin S, Wang R, et al. Voxel-based quadrilateral mesh generation from point cloud[J]. Multimedia Tools and Applications, 2020, 79: 20561-20578.
- [33] Cruz L, Lucio D, Velho L. Kinect and rgbd images: Challenges and applications[C]//2012 25th SIBGRAPI conference on graphics, patterns and images tutorials. 2012: 36-49.
- [34] Johnson A E. Spin-images: a representation for 3-D surface matching[J]., 1997.
- [35] Steder B, Rusu R B, Konolige K, et al. NARF: 3D range image features for object recognition[C]// Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS): vol. 44. 2010: 2.
- [36] Rusu R B, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3D registration[C]//2009 IEEE international conference on robotics and automation. 2009: 3212-3217.

- [37] 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.
- [38] Pío R. Euler angle transformations[J]. IEEE Transactions on automatic control, 1966, 11(4): 707-715.
- [39] Shoemake K. Animating rotation with quaternion curves[C]//Proceedings of the 12th annual conference on Computer graphics and interactive techniques. 1985: 245-254.
- [40] Horn A. Doubly stochastic matrices and the diagonal of a rotation matrix[J]. American Journal of Mathematics, 1954, 76(3): 620-630.
- [41] Diebel J, et al. Representing attitude: Euler angles, unit quaternions, and rotation vectors[J]. Matrix, 2006, 58(15-16): 1-35.
- [42] Šenk M, Cheze L. Rotation sequence as an important factor in shoulder kinematics[J]. Clinical biomechanics, 2006, 21: S3-S8.
- [43] Stuelpnagel J. On the parametrization of the three-dimensional rotation group[J]. SIAM review, 1964, 6(4): 422-430.
- [44] Levinson J, Esteves C, Chen K, et al. An analysis of svd for deep rotation estimation[J]. Advances in Neural Information Processing Systems, 2020, 33: 22554-22565.
- [45] Levenberg K. A method for the solution of certain non-linear problems in least squares[J]. Quarterly of applied mathematics, 1944, 2(2): 164-168.
- [46] 李娇娇, 孙红岩, 董雨, 等. 基于深度学习的 3 维点云处理综述[J]. 计算机研究与发展, 2022, 59(5): 20.

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。