

清 华 大 学

# 高年级本科生研究员 (SURF) 结题报告

题目：前端自定义的点线联合 VIO 框架

班 级：行健-力 0

姓 名：于舒昂

实习企业：美团

企业导师：黄鹏辉

2024 年 3 月 3 日

## 中文摘要

视觉定位是无人系统中最基础也最重要的问题之一，视觉惯性里程计（Visual-Inertial Odometry, VIO）是无人机、机器人等无人设备中应用广泛的视觉定位系统。传统的 VIO 系统主要基于点特征，通过提取和匹配图像中的角点或特征点来估计相机的位姿和轨迹。然而，点特征在低纹理场景下存在局限，将导致 VIO 系统的性能下降。线特征的加入可以进一步利用图像中存在的边缘信息，使 VIO 系统在更广泛的环境中具有更好的性能。但是，线特征存在参数化复杂且耗时的问题，将导致 VIO 系统在嵌入式设备上难以达到实时性需求。

为了构建一个可以在嵌入式小型无人设备上实时运行的点线 VIO 系统，本课题首先基于神经网络的特征推理网络，提出了点线融合的 SP-SOLD2 网络，并利用自监督训练使得该网络可以通过一次推理得到特征点、特征线以及同时适用于点线匹配的描述子；为了将各类点线提取和匹配方法灵活应用于 VIO 任务中，本工作构建了一个前端方法可自定义的 NN-PL-VIO 框架，并利用 SP-SOLD2 点线联合网络以及一系列速度优化方法，得到了一个可以在嵌入式设备 Orin NX 上实时运行的点线联合 VIO 系统。

**关键词：**计算机视觉；点线特征；视觉里程计

## ABSTRACT

Quadrotors are agile and can fly at high speed in complex environments. They have been widely used in rescue, logistics, agriculture and military fields, and have high research value and application value. However, their autonomous flight is still a challenging task due to the nonlinearity of dynamics and the complexity of the environments. So far, only well-trained human pilots can exert their extreme performance.

There are two main types of existing drone autonomous flight methods. The optimization-based method has advantages of strong interpretability, stable flight, and high planning reliability, but it also faces the problem of slow mapping and planning on-board. Another type of algorithm based on deep learning attempts to establish an end-to-end planning controller. Although it solves the problem of slow calculation speed, it still has the problem of relying on a large number of manually collected expert data and labels. Some methods are insufficient generalization and are difficult to adapt to autonomous navigation tasks under different speeds and scenarios.

This study is oriented towards the autonomous navigation task of quadrotors, and proposes an end-to-end autonomous navigation algorithm framework based on reinforcement learning. Using deep reinforcement learning to fit highly dynamic and nonlinear capabilities, an autonomous navigation algorithm with better performance is obtained. This study also proposes a three-stage training method, which improves the convergence efficiency of the algorithm by an order of magnitude. This research also improves the existing simulator according to the algorithm requirements, so that it can increase the simulation speed by 2~6 times. Finally, this study also developed a deployment platform with a common interface. In the capability test, this deployment platform also achieved better capabilities than similar platforms. In this study, the trained algorithm was tested on the flight platform, which can effectively complete the navigation task and has the ability of application and deployment.

**Keywords:** Autonomous vehicle navigation; Motion and path planning; Reinforcement learning

# 目 录

第 1 章 研究背景 .....	1
1.1 点特征提取及描述方法 .....	2
1.2 线特征提取描述方法 .....	3
1.2.1 基于全局哈夫的方法 .....	4
1.2.2 基于局部信息的方法 .....	4
1.2.3 基于深度学习的方法 .....	5
1.3 线特征描述子提取方法 .....	6
1.3.1 基于统计的方法 .....	7
1.3.2 基于深度学习的方法 .....	7
1.4 各章节概述 .....	8
插图索引 .....	9
表格索引 .....	10
参考文献 .....	11
致 谢 .....	12
声 明 .....	13

## 第 1 章 研究背景

在机器人学研究中，视觉惯性测距（VIO）在导航、定位和感知任务中发挥着举足轻重的作用。VIO 系统通过融合来自视觉和惯性传感器的数据来估计机器人的姿态（位置和方向）。近年来，人们提出了许多 VIO 方法，例如直接使用像素强度估算摄像机运动的直接方法，以及基于特征的方法等。其中，基于特征的方法使用图像特征（点、线或平面）来估计传感器位姿和周围环境的地图。基于点的

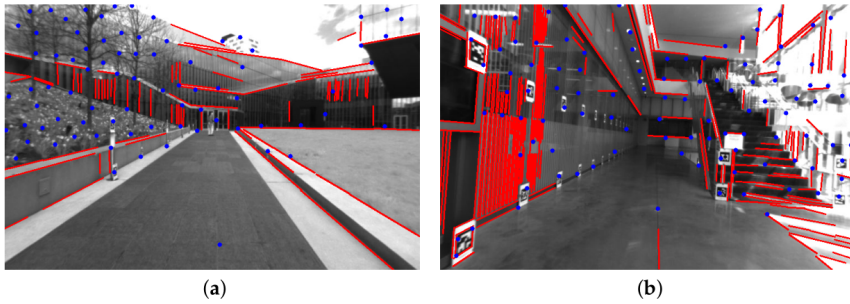


图 1.1 图像中的点线特征

VIO 算法发展最为成熟，现有点 VIO 算法在精度和速度方面都表现出了卓越的性能，可以应用于需要实时跟踪的场景。然而，在低纹理环境或重复纹理环境中（例如走廊、道路、开阔水域等），由于点特征信息匮乏，基于点的 VIO 算法会出现性能下降的问题。为了利用环境中丰富的边缘信息，克服点信息失效的问题，研究人员进一步开发了基于线特征的 VIO 算法，并将线特征作为辅助构建了点线联合的 VIO 框架，现有框架包括 PL-SLAM[4]，PL-VINS[5] 等。但是，由于线特征存在参数化复杂且耗时的问题，现有的点线联合 VIO 框架难以在嵌入式设备上实时运行，使得线特征的精度增益难以在小型无人设备（比如无人机）上得到实现。为了构建一个可以在嵌入式小型无人设备上实时运行的点线 VIO 系统，本工作首先基于神经网络的特征推理网络，提出了点线融合的 SP-SOLD2 网络，并利用自监督训练使得该网络可以通过一次推理得到特征点、特征线以及同时适用于点线匹配的描述子；为了将各类点线提取和匹配方法灵活应用于 VIO 任务中，本工作构建了一个前端方法可自定义的 NN-PL-VIO 框架，并利用 SP-SOLD2 点线联合网络以及一系列速度优化方法，得到了一个可以在嵌入式设备 Orin NX 上实时运行的点线联合 VIO 系统。

## 1.1 点特征提取及描述方法

点特征提取方法可以简单地分类为传统方法和深度学习方法两大类，其中传统方法是通过一系列数学计算得到特征点及其描述子，具有数学可解释性；而深度学习方法则是通过设计网络来端到端推理定位特征点，和描述子可以不耦合。一些常见的传统点特征方法如下：

1. Harris 角点检测算法 (Harris Corner Detection)：该算法通过计算图像中每个像素点的角点响应函数来检测角点。角点响应函数基于图像中局部区域的亮度变化，对角点位置具有较高的响应值。Harris 角点检测算法具有旋转不变性和一定程度的尺度不变性。
2. Shi-Tomasi 角点检测算法 (Shi-Tomasi Corner Detection)：该算法是对 Harris 角点检测算法的改进。它使用了一个更加鲁棒的角点响应函数，能够在保持高角点检测率的同时提供更好的角点定位精度。
3. SIFT (尺度不变特征变换, Scale-Invariant Feature Transform)：SIFT 是一种广泛应用的特征点提取算法，能够在不同尺度和旋转条件下提取稳定的特征点。SIFT 算法通过在图像中检测局部极值点，并在多个尺度上计算尺度空间特征向量，得到具有尺度不变性的特征描述子。
4. SURF (加速稳健特征, Speeded-Up Robust Features)：SURF 算法是对 SIFT 算法的改进，通过提出一种快速的特征点检测和描述子计算方法，实现了更高的计算效率。SURF 算法利用图像的 Hessian 矩阵来检测特征点，并使用 Haar 小波响应描述子来描述特征。
5. FAST (Features from Accelerated Segment Test)：FAST 算法是一种快速的角点检测算法，适用于实时应用。该算法基于像素灰度值的变化来检测角点，通过高效的特征点筛选策略提高了检测速度。
6. ORB (Oriented FAST and Rotated BRIEF)：ORB 算法是对 FAST 和 BRIEF (Binary Robust Independent Elementary Features) 算法的结合。它在 FAST 算法的基础上引入了旋转不变性，并使用二进制描述子来进行特征匹配，具有较高的计算效率和良好的性能。

而基于神经网络的点特征提取方法可以分为两种类别：检测和描述分离的方法与端到端匹配的方法。其中检测和描述分离的方法是设计两个网络或者一个联合网络，分别进行点特征的检测和点描述子的提取；而端到端匹配的方法则是直接设计和训练一个网络，此网络可以接受两个输入图像，直接给出特征匹配结果。基

于神经网络的方法总结如表1.1:

名称	类型	描述
ASLFeat[6]	点检测	使用可变形卷积网络，利用固有的特征层次，提出了一种多级检测机制，该机制不仅在没有额外学习权重的情况下可以恢复低层细节以进行精确的关键点定位
SuperPoint[7]	点检测和描述子提取联合	介绍了一个自监督框架，利用运行在全尺寸图像上的全卷积模型，在一个前向传递中同时计算像素级兴趣点位置和关联描述子
R2D2[8]	点检测和描述子提取联合	在 D2-Net 基础上考虑了特征检测的可重复性和可靠性，为检测得分设计了损失函数，使用全卷积网络推理特征点及其描述子
LoFTR[9]	端到端匹配	使用了 Transformer 中的自我和交叉注意力层 (self and cross attention layers) 来获取两个图像的特征描述符，实现无检测器的局部特征匹配。
Patch2Pix[10]	端到端匹配	建立了一种新的匹配细化网络，首先得到 patch-level 的匹配，再细化到 pixel-level 的匹配，网络可以同时细化匹配并排除错误匹配，且训练不需要像素级的 GT 对应关系。
SuperGlue[11]	描述子提取	提出了一种能够同时进行特征匹配以及滤除外点的网络，基于注意力机制提出了一种灵活的内容聚合机制，这其能够同时感知潜在的 3D 场景以及进行特征匹配。

表 1.1 基于神经网络的点特征提取方法

## 1.2 线特征提取描述方法

线特征的提取方法可以分为三种类别：基于全局哈夫的方法、基于局部信息的方法，以及基于深度学习的方法，下面将分别介绍。

### 1.2.1 基于全局哈夫的方法

基于任何一个点都可能是线上的点的假设，此方法对所有点按照下式进行 Hough 变换：

$$\rho = x_0 \cos \theta + y_0 \sin \theta$$

则一个点  $(x_0, y_0)$  将对应 Hough 平面下的一系列曲线；Hough 平面上产生交点的曲线，则对应原平面上同线的一系列点，如图??所示。利用 Hough 变换，此方法利

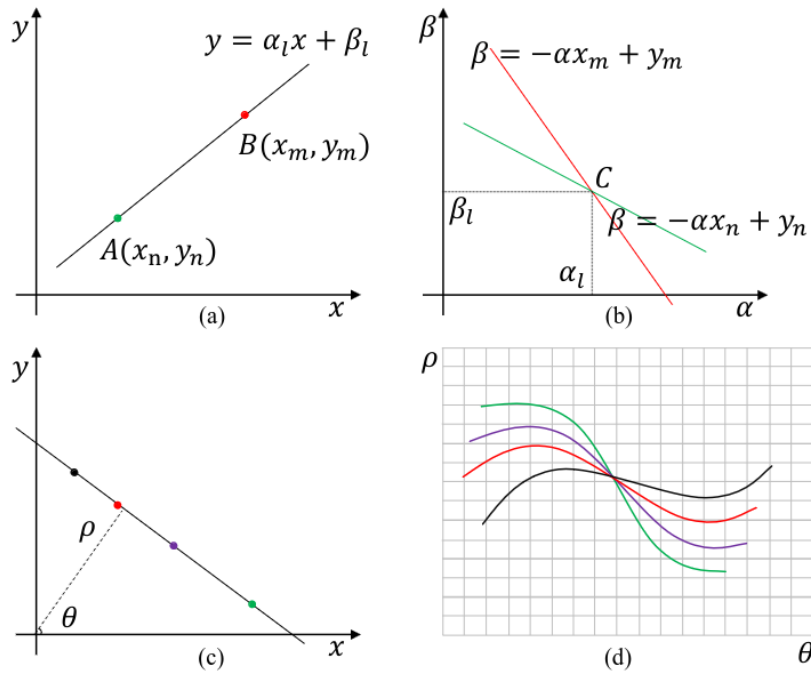


图 1.2 哈夫变换示意图

用严格的数学推导得到共线点关系，但是作为一个全局方法，需要对所有像素进行操作（包括预处理和变换），复杂度较高，并且空间定量搜索方式提取的数量和质量一般，不能定位线段端点。

### 1.2.2 基于局部信息的方法

此类方法利用局部信息（图像梯度和边缘检测结果等）探寻可能在线段的位置，提取线段。LSD[12] 是一种典型的基于局部信息的线段提取方法，如图??所示。该算法首先对输入图像进行多尺度的高斯滤波，以便检测不同尺度下的线段。接着，LSD 会计算图像的梯度信息，找出梯度变化显著的区域，作为线段潜在区域。然后，通过区域合并方法，相邻的梯度变化显著的像素点会被组合成候选线段，计算线段位置、长度、方向和梯度等参数。为了去除冗余线段，LSD 在最后会



使用非极大值抑制和滤波的方法来筛选得到最终的线段提取结果。基于局部信息

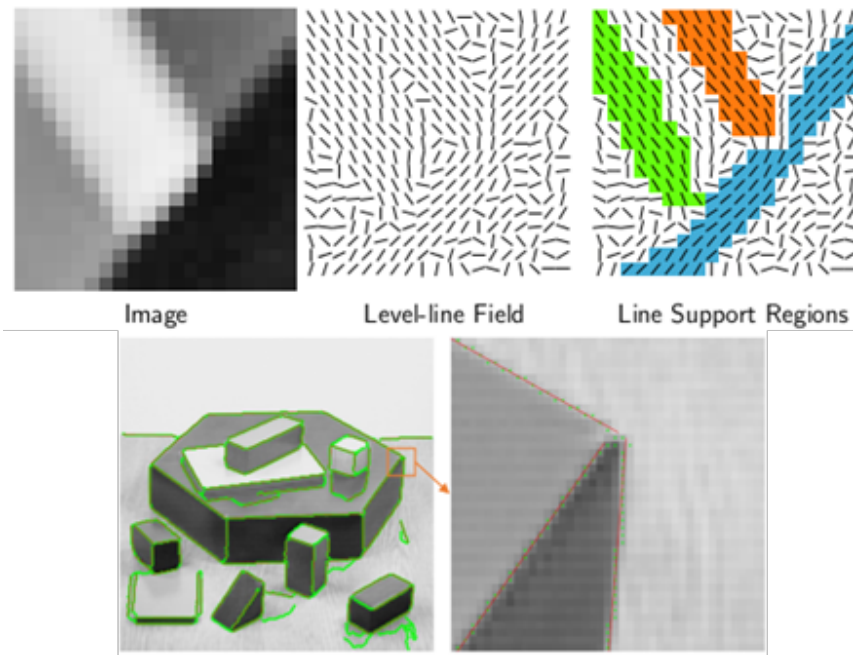


图 1.3 LSD 线提取方法

的方法虽然一定程度上降低了线段提取的复杂度，但是算法仍为像素级操作，空间复杂度高，并且缺乏语义和全局约束信息，此类方法易受到噪声干扰。

1.2.3 基于深度学习的方法

基于深度学习的线段提取方法利用神经网络来自动检测和提取图像中的线段，能够通过网络设计更灵活地融合全局和局部信息，得到准确度和实用性更好的线段提取结果。根据网络设计的思路，可以分为 Top-Down 策略、Bottom-Up 策略和 Tri-Point 策略。

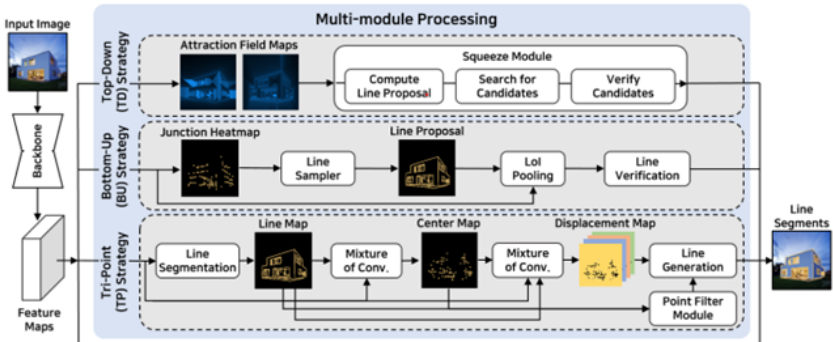


图 1.4 基于深度学习的线特征提取方法

1. Top-Down 策略：自上而下的策略核心是从线段可能存在的区域“挤压”得到线段提取结果，故而网络提取的目标是与线段相关的概率图。
2. Bottom-Up 策略：自下而上的策略核心是先找到线段的端点，再从端点定位线段。
3. Tri-Point 策略：Tri-Point 策略核心是定位线上的某些点，用多个点和线段之间的距离关系定位线段。

线提取方法总结如表1.2所示：基于深度学习的方法虽然有着高精度和端到端

类别	典型工作	优点	缺点	典型应用
基于全局哈夫的方法	哈夫变换 哈夫概率变换	原理简单，数学可解释；有一定抗干扰能力	缺少局部特征和语义约束；没有做线段切分	灭点检测 地平线检测
基于局部信息的方法	LSD[12] EDLines[13]	计算量较低；数学可解释；有效利用局部连接信息	对噪声敏感；缺乏全局感知和语义信息	VO 视觉定位 VSLAM
基于深度学习的方法	MLSD[14] L-CNN[15] SOLD2[16] DeepLSD[17]	端到端处理；融合了局部和全局信息	受网络设计、训练数据干扰较大；GPU 计算需要实时优化	三维重建 图像分割

表 1.2 线特征提取方法总结

运行的优势，但效果依赖于数据集和网络设计，速度依赖于硬件性能和网络复杂度，尚需泛化性和实时性优化。

### 1.3 线特征描述子提取方法

线的描述子是一种用来表示和描述检测到的线段特征的数据结构，也即线段的参数化形式。描述子利用线段的几何特性和其他相关信息，进行线段的匹配、跟踪等任务。根据线特征描述子所利用的信息，可以将其分为基于统计的方法、基于结构的方法和基于深度学习的方法。

### 1.3.1 基于统计的方法

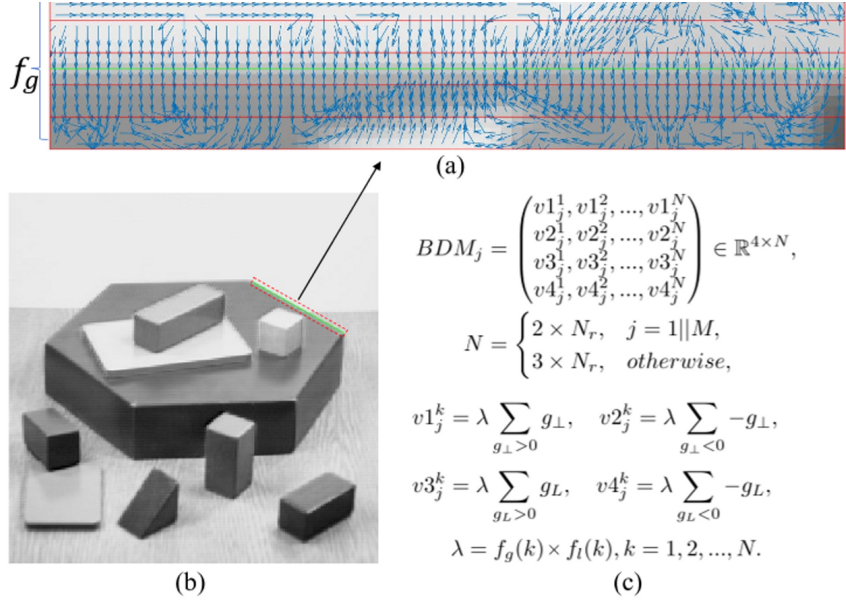


图 1.5 LBD 方法示意图

基于统计的方法通过计算统计线相关的特征来描述线段，比如线梯度和线强度等。LBD 是一种典型的基于统计描述方法。它采用了二进制编码，以紧凑的方式表示线段的局部特征，这使得它在快速匹配和检索任务中表现出色。LBD 描述子具有旋转不变性，可以处理不同方向的线段，而无需进行旋转归一化，如图??所示。此外，它还具有尺度不变性，可适应不同尺度的线段。LBD 描述子对抗噪声，适应线段形状的小变化，因此在实际应用中非常稳定。由于其二进制性质，LBD 描述子可通过位运算进行高效匹配，这对于需要实时性能和大规模图像数据库的应用非常有帮助。LBD 描述子关注线段周围的局部信息，从而能够有效地区分不同的线段模式。它通常用于线段匹配任务，如基于线段的目标识别、基于线段的姿态估计以及基于线段的视觉 SLAM。因此，LBD 描述子在计算机视觉领域扮演着重要的角色，为线段相关的应用提供了强大的工具。

### 1.3.2 基于深度学习的方法

基于深度学习的方法通过设计描述子网络来提取线段的描述子，其设计损失函数的理念是使得匹配的线段在描述空间中尽可能相近，使得不匹配的线段在描述空间中尽可能远离。根据网络推理的策略，可以将深度学习方法划分为基于线段支持区域的和不基于线段支持区域的两个种类。

1. 基于线段支持区域的方法：在网络设计和训练中都用到线段周围的局部特

征，并在此基础上进一步加入局部特征的描述子。

2. 基于线段支持区域的方法：在网络设计和训练中都用到线段周围的局部特征，并在此基础上进一步加入局部特征的描述子。

线段描述子提取方法总结如表1.3所示。

类别	典型工作	优点	缺点	典型应用
Statistic-based	LBD[18] MSLD[19]	一定程度抗干扰；数学可解释	对低纹理场景敏感；缺乏全局语义信息	视觉定位 VSLAM
Structure-based	NLD[20] LS[21]	利用局部信息；可以处理一些低纹理场景	要求线段共面；丢失全局和语义信息	地图配准
Learning-based	DLD[22] WLD[23]	端到端方法；可以通过设计融合局部和全局信息	缺乏通用训练集和表示方法，且泛化性有待考证；GPU 计算	场景解析和抽象

表 1.3 线段描述子提取方法总结

## 1.4 各章节概述

本研究报告共分为六章，本章主要介绍了本研究的背景、现存挑战，从应用和学术研究的角度阐述了本研究的必要性和挑战性。第二章至第四章分别介绍了仿真器的改进工作、强化学习算法和部署平台的结构设计，并简要介绍这些设计对总体飞行性能的影响。第五章介绍了本研究的实验结果，包括仿真平台参数的设置，算法调试的结果和实机部署的结果。第六章对本研究做一个整体总结并提出该项目未来可继续改进的方向。

## 插图索引

图 1.1	图像中的点线特征 .....	1
图 1.2	哈夫变换示意图 .....	4
图 1.3	LSD 线提取方法 .....	5
图 1.4	基于深度学习的线特征提取方法 .....	5
图 1.5	LBD 方法示意图.....	7

## 表格索引

表 1.1	基于神经网络的点特征提取方法 .....	3
表 1.2	线特征提取方法总结 .....	6
表 1.3	线段描述子提取方法总结 .....	8

## 参考文献

## 致 谢

首先感谢汪玉老师一年以来对我的指导和帮助，汪老师提出的不少针对性建议是本项目得以完善至今不可或缺的支持。

本项目是在课题组师兄、师姐的帮助下完成的。本人负责的部分是仿真器改进的部分工作、主动感知算法的大部分工作、以及部署系统设计的全部工作。其中仿真器改进的代码是我与高枫学长共同完成的，部署平台的硬件设计是我与宋宏宇学长一同完成的。感谢整个研究过程中高枫学长、宋宏宇学长、于超学姐和余金城学长的帮助。

最后，感谢钱班项目组的老师们，尤其是负责 ORIC 项目的徐芦平老师对我的关心和支持。也感谢助教一年来对我的帮助。本项目在研究过程中参与过多次书院、学校的科研交流活动，同学们提出的建议和改进方向也对本项目起到了非常大的帮助作用。



## 声 明

本人郑重声明：所呈交的结题报告，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本报告所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：\_\_\_\_\_ 日 期：\_\_\_\_\_