

同行专家业内评价意见书编号： 20250854436

附件1

浙江工程师学院（浙江大学工程师学院） 同行专家业内评价意见书

姓名： 彭昊龙

学号： 22260035

申报工程师职称专业类别（领域）： 电子信息

浙江工程师学院（浙江大学工程师学院）制

2025年05月09日

填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位+申报工程师职称专业类别(领域)4位+流水号3位，共11位。

一、个人申报

（一）基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

1. 对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

对本专业基础理论知识和专业技术知识掌握情况

在余姚市机器人研究中心的学习和研究过程中，我系统掌握了控制工程相关的基础理论和专业技术，特别是在智能物流机器人方向的核心关键技术，包括但不限于以下方面：

自动控制理论：掌握了经典控制理论（PID控制、状态空间方法）和现代控制方法（MPC、鲁棒控制、自适应控制等），并将其应用于机器人路径规划和导航控制。

机器人学：深入理解机器人运动学、动力学建模，包括正逆运动学求解、轨迹规划、路径优化等。

人工智能与深度学习：熟练运用LSTM（长短时记忆网络）、自注意力机制、强化学习等前沿AI算法，提高机器人自主决策能力。

多传感器融合：掌握扩展卡尔曼滤波（EKF）、粒子滤波（PF）、信息滤波等融合技术，实现机器人在复杂环境中的精准定位与感知。

点云数据处理：熟练使用PCL（Point Cloud

Library）等工具进行点云预处理、降噪、特征提取等，提升SLAM（同步定位与建图）的准确性。

优化算法：应用斯坦因变分梯度（SVGD）和概率补偿方法，提高机器人导航算法的稳定性和鲁棒性。

2. 工程实践的经历(不少于200字)

工程实践的经历

在机器人研究中心的工程实践中，我主要参与了智能物流机器人在医院环境中的研发与部署，涵盖以下关键实践环节：

（1）路径规划与导航系统设计

设计并优化基于A*、D*、DWA（动态窗口法）的全局路径规划与局部避障算法，提升机器人在复杂环境中的移动效率。

结合强化学习，使机器人能够在动态环境中适应不同路径选择需求，实现自主避障。

（2）传感器融合与环境感知

设计多传感器信息融合系统，整合激光雷达（LiDAR）、RGB-D相机、惯性测量单元（IMU）、超声波等数据源。

采用扩展卡尔曼滤波（EKF）和粒子滤波（PF）进行数据融合，提升定位精度，确保机器人在医院环境中的稳定导航。

（3）点云数据处理

处理大规模点云数据，实现环境建模和实时障碍物检测：

采用Voxel Grid滤波降低数据冗余，提高计算效率。

使用RANSAC算法去除噪声，提高环境建模精度。

采用KD-Tree搜索加速点云匹配，提高SLAM算法的实时性。

（4）深度学习优化导航算法

结合LSTM、自注意力机制、斯坦因变分梯度（SVGD）优化机器人导航路径预测，提高机器人在复杂环境下的规划能力。

（5）医院场景测试与实际部署

经过实验室测试和模拟医院环境测试后，智能物流机器人成功部署于医院，实现自动药品配

送、样本收集、医疗器械运输等任务，提高了医疗物流效率。

3. 在实际工作中综合运用所学知识解决复杂工程问题的案例（不少于1000字）

在实际工作中综合运用所学知识解决复杂工程问题的案例

智能物流机器人在医院环境中的路径规划与导航优化

1、背景

传统医院物流主要依赖人工运输药品、样本和医疗设备，但效率低、易出错，且在疫情等特殊情况下存在安全隐患。智能物流机器人在医院环境中应用时，面临诸多挑战：

复杂动态环境：医院内部存在大量动态障碍（如病人、医护人员、移动设备等）。

高精度路径规划需求：需要精确控制机器人在狭窄通道内移动，并精准送达目的地。

稳定的导航算法：需要结合多传感器数据，确保机器人在不同光照、障碍物等环境变化下仍能稳定运行。

2、解决方案

本研究针对医院环境的特殊需求，综合运用了控制工程、人工智能、点云处理、多传感器融合等技术，优化智能物流机器人的路径规划与导航系统。

(1) 路径规划优化

全局路径规划：采用A*算法结合Dijkstra算法，在医院2D地图上计算最短路径，并对路径进行平滑处理，确保机器人移动稳定。

局部避障规划：结合DWA（动态窗口法），使机器人能够在动态环境中自主避障，提高行驶安全性。

强化学习优化：通过深度强化学习（DRL），让机器人在多次模拟训练中学习最优路径选择策略，使其能适应不同的医院环境。

(2) 多模态传感器融合

激光雷达（LiDAR）+ RGB-D相机：用于构建3D环境模型，检测障碍物。

惯性测量单元（IMU）：用于机器人姿态估计，避免倾倒或路径偏移。

超声波传感器：提供近距离避障能力，确保机器人在狭窄空间内稳定运行。

融合方法：

采用扩展卡尔曼滤波（EKF）融合激光雷达和IMU数据，提高定位精度。

使用概率补偿方法，修正传感器误差，确保环境感知的准确性。

(3) 点云数据处理

机器人通过LiDAR采集的大规模点云数据进行实时环境建模：

采用Voxel Grid滤波降低点云冗余，减少计算开销。

使用RANSAC算法去除异常点，提高数据精度。

结合KD-Tree搜索进行点云匹配，加速SLAM（同步定位与建图）。

(4) 基于深度学习的导航优化

LSTM路径预测：基于历史轨迹数据训练LSTM模型，使机器人能预测最优路径，提高路径规划的稳定性。

自注意力机制：优化路径规划过程中关键特征提取能力，提高障碍物识别精准度。

斯坦因变分梯度（SVGD）优化：用于概率路径采样，减少路径偏移，提高导航精度。

3、实施与应用

在医院环境中部署的智能物流机器人经过实验室测试、模拟环境测试、实际应用测试三个阶段，最终成功应用：

任务执行效率提高30%，大幅减少人工配送时间。

自主避障能力增强，机器人能够在病房、手术室、走廊等复杂环境中稳定运行。

降低物流错误率，提高医院物流管理的自动化程度。

4、总结

本案例综合运用了控制工程、机器人学、深度学习、传感器融合、点云处理等技术，实现了医院环境下智能物流机器人的高效路径规划与稳定导航。通过优化路径规划、传感器融合、导航算法，使机器人能够适应复杂动态环境，提高医院物流的智能化水平。本研究的成果不仅提升了机器人在医疗领域的应用价值，也为智能物流机器人在更多行业（如仓储、智能制造）提供了技术借鉴

(二)取得的业绩(代表作)【限填3项,须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等)供核实,并提供复印件一份】

1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】

成果名称	成果类别 [含论文、授权专利（含发明专利申请）、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/ 授权或申请时间等	刊物名称 /专利授权或申请号等	本人排名/ 总人数	备注
CGA-MAE: Camera-Guided Alignment to Auxiliary Point Cloud MaskedAuto-encoders for 3D Object Detection	会议论文	2024年05月17日	2024 IEEE 13th Data Driven Control and Learning Systems Conference	1/5	EI会议收录

2. 其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】

(三) 在校期间课程、专业实践训练及学位论文相关情况	
课程成绩情况	按课程学分核算的平均成绩： 84 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 1.7 年（要求1年及以上） 考核成绩： 86 分
本人承诺	
个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！ 申报人签名：彭昊友	

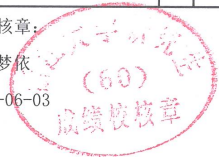
浙江大学研究生院
攻读硕士学位研究生成绩表

学号：22260035		姓名：彭昊龙		性别：男		学院：工程师学院			专业：电子信息			学制：2.5年	
毕业时最低应获：26.0学分				已获得：28.0学分					入学年月：2022-09			毕业年月：	
学位证书号：					毕业证书号：					授予学位：			
学习时间	课程名称		备注	学分	成绩	课程性质	学习时间	课程名称		备注	学分	成绩	课程性质
2022-2023学年秋季学期	新时代中国特色社会主义思想理论与实践			2.0	87	公共学位课	2022-2023学年春季学期	研究生英语基础技能			1.0	76	公共学位课
2022-2023学年秋季学期	工程技术创新前沿			1.5	85	专业学位课	2022-2023学年春夏学期	工程伦理			2.0	84	公共学位课
2022-2023学年秋冬学期	研究生论文写作指导			1.0	91	专业学位课	2022-2023学年夏季学期	自然辩证法概论			1.0	65	公共学位课
2022-2023学年秋冬学期	智能工业机器人及其应用			3.0	75	专业选修课	2022-2023学年春夏学期	高阶工程认知实践			3.0	79	专业学位课
2022-2023学年冬学期	产业技术发展前沿			1.5	85	专业学位课	2022-2023学年春夏学期	人工智能制造技术			3.0	82	专业学位课
2022-2023学年冬学期	研究生英语			2.0	87	公共学位课	2022-2023学年夏季学期	工程师创新创业思维			2.0	97	专业选修课
2022-2023学年春季学期	“四史”专题			1.0	94	公共选修课		硕士生读书报告			2.0	通过	
2022-2023学年春季学期	数学建模			2.0	92	专业选修课							

说明：1. 研究生课程按三种方法计分：百分制，两级制（通过、不通过），五级制（优、良、中、及格、不及格）。

2. 备注中“*”表示重修课程。

学院成绩校核章：
成绩校核人：张梦依
打印日期：2025-06-03



CGA-MAE: Camera-Guided Alignment to Auxiliary Point Cloud Masked Autoencoders for 3D Object Detection

Haolong Peng ^{†1,3}, Yibo Chen ^{†2,3}, Yuxi Chen ^{2,3}, Ke Zhu ³, Jianming Zhang ^{*2,3}

1. Zhejiang Polytechnic institute, Polytechnic institute, Zhejiang University, Hangzhou, China

2. College of control science and engineering, Zhejiang University, Hangzhou, China

3. Zhejiang University Robotics Institute, Yuyao, China

E-mail: hollingworth, cybh, 22332147@zju.edu.cn, 2572873948@qq.com, jmzhang@iipc.zju.edu.cn

Abstract: Masked Autoencoders have demonstrated remarkable proficiency in acquiring robust visual representations and achieving cutting-edge outcomes across diverse independent modalities. However, only a handful of studies have delved into their efficacy within multi-modality environments. In this work, we focus on LiDAR point cloud and Camera image, two modalities commonly coexistent in real-world scenarios, aiming to explore their intricate interactions. To improve upon the cross-modal synergy in existing works, we project the feature in the bird's eye view (BEV) plane to receive more geometry information and propose CGA-MAE, a simple and effective cross-modal fusion self-supervised pretraining method based on Camera-Guided Alignment (CGA), which can bridge the BEV feature learning gap between LiDAR and Camera. We initially recognize the significance of masking strategies between these sources, leveraging a projection module, we harmonize the alignment of masks and visible tokens inherent in both modalities. Then, we propose a new Modality Feature Fusion Model called MFFM to promote the interaction between the LiDAR feature and the Camera Feature. Through extensive experiments performed on large NuScenes Dataset, we discover it is nontrivial to interactively learn LiDAR-Camera features, where Our approach significantly advances the State-Of-The-Art (SOTA), achieving a mean Average Precision (mAP) of 67.8% and a nuScenes Detection Score (NDS) of 71.0%.

Key Words: Masked Autoencoder, LiDAR-Camera, Feature Alignment, Multi-Modality and Feature Fusion

1 Introduction

With the proliferation of autonomous driving technology, the need for effectively sensing 3D environments has become paramount. While LiDAR-based methods have made significant advancements, outdoor point cloud data acquired via radar still lacks several crucial geometric attributes at a fine-grained level. Furthermore, unlike pixel data, the number of points within different data frames is variable, which results in a mismatch in the degree of correlation between discrete point clouds due to temporal shifts in spatial alignment. This, in turn, leads to a reduction in the model's generalization capacity and a subsequent decrease in the accuracy of the validation set. As a result, there is ample room for improvement in current point cloud target detection algorithms.

A straightforward approach to improving LiDAR-based 3D object detection involves enhancing feature extraction through methods such as Farthest Point Sampling (FPS) and Pillar Feature Extracting method, etc. These techniques aim to constrain the correlation between features in point clouds, providing a more robust representation of point cloud sequences depicting the same object with varying morphologies. However, these algorithms will have to increase the computational effort of the whole work, while the geometry-level clues will still not be fully captured. Another strategy involves fusing pixel information from cameras to match point cloud features, utilizing pixel data to correlate with the spatial position information of the point cloud. Unfortunately, this approach is both more resource-intensive and less effective. Relying solely on a single feature alignment or fusion method restricts the model's ability to fully comprehend the potential features or representations of objects.

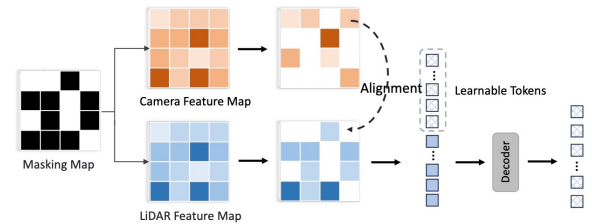


Fig. 1: **Camera-Guided Alignment (CGA) mask strategy.** Compared with the traditional MAE masking method, CGA-MAE can utilize the special pixel information and feature geometry information brought by the image to obtain better reconstruction results for the point cloud.

As an alternative, the Bird's-Eye-View (BEV) based framework has gained considerable attention for providing a comprehensive feature representation space. This approach framework is drawing extensive attention to offer a holistic feature representation space from multi-camera images and has made substantial improvements. The BEV-based approach is capable of fusing information from multi-camera images into the BEV space, and many works have already utilized BEV for the analysis of multi-camera images and LiDAR points. Camera images and LiDAR point clouds, and the results have verified that the method can indeed make up for the better spatial correlation of LiDAR point clouds. Importantly, the output from the BEV-based framework proves highly applicable to downstream task predictions. Despite the advancement achieved in this field, a distinct performance gap remains between multi-camera BEV and LiDAR-based 3D object detection. The original intention of doing multi-modal fusion at the BEV level is to obtain better feature representation and associated point cloud objects How-

This work was supported by Robotics Institute of Zhejiang University under Grant K12201. [†]: Equal first authorship. ^{*}: Corresponding author

经检索“Engineering Village”，下述论文被《Ei Compendex》收录。（检索时间：2025年3月21日）。

<RECORD 1>

Accession number:20243516969447

Title:CGA-MAE: Camera-Guided Alignment to Auxiliary Point Cloud Masked Autoencoders for 3D Object Detection

Authors: Peng, Haolong (1, 2); Chen, Yibo (2, 3); Chen, Yuxi (2, 3); Zhu, Ke (2); Zhang, Jianming (2, 3)

Author affiliation: (1) Zhejiang Polytechnic Institute, Polytechnic Institute, Zhejiang University, Hangzhou, China; (2) Zhejiang University Robotics Institute, Yuyao, China; (3) College of Control Science and Engineering, Zhejiang University, Hangzhou, China

Corresponding author: Zhang, Jianming(jmzhang@iipc.zju.edu.cn)

Source title: Proceedings of 2024 IEEE 13th Data Driven Control and Learning Systems Conference, DDCLS 2024

Abbreviated source title: Proc. IEEE Data Driven Control Learn. Syst. Conf., DDCLS

Part number: 1 of 1

Issue title: Proceedings of 2024 IEEE 13th Data Driven Control and Learning Systems Conference, DDCLS 2024

Issue date: 2024

Publication year: 2024

Pages: 79-84

Language: English

ISBN-13: 9798350361674

Document type: Conference article (CA)

Conference name: 13th IEEE Data Driven Control and Learning Systems Conference, DDCLS 2024

Conference date: May 17, 2024 - May 19, 2024

Conference location: Kaifeng, China

Conference code: 201823

Sponsor: Chinese Association of Automation (CAA); DCLOD; Henan University; IEEE; IEEE Beijing Section; Qingdao University

Publisher: Institute of Electrical and Electronics Engineers Inc.

Number of references: 27

Main heading: Object detection

Controlled terms: Alignment - Image coding - Image enhancement

Uncontrolled terms: 3D object - Auto encoders - Bird's eye view - Cross-modal - Feature alignment - LiDAR-camera - Masked autoencoder - Multi-feature fusion - Multi-modality fusion - Point-clouds

Classification code: 1106.3.1 Image Processing - 1106.8 Computer Vision - 601.1 Mechanical Devices

Numerical data indexing: Percentage 6.78E+01%, Percentage 7.10E+01%

DOI: 10.1109/DDCLS61622.2024.10606845

Funding details: Number: K12201, Acronym: -, Sponsor: -;

Funding text: This work was supported by Robotics Institute of Zhejiang University under Grant K12201.

Database: Compendex

Compilation and indexing terms, Copyright 2025 Elsevier Inc.

注:

1. 以上检索结果来自 CALIS 查收查引系统。
2. 以上检索结果均得到委托人及被检索作者的确认。

