

同行专家业内评价意见书编号： 20250854383

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： 张睿昊

学号： 22260468

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03月16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)

在专业实践过程中，我系统掌握了工程实践所需的基础理论与专业技术知识，尤其在计算机技术与混合现实领域形成了完整的知识体系。首先，数学与自然科学知识为项目开发提供了理论支撑。例如，在三维场景构建中，线性代数的矩阵运算和坐标变换理论被应用于虚拟物体与现实场景的精准对齐；物理学中的光学原理帮助理解混合现实设备（如HoloLens 2）的成像机制，确保虚拟物体与真实环境的融合效果。其次，计算机专业知识是技术实现的核心。我深入学习了计算机图形学、Unity引擎开发、C#编程语言以及混合现实开发工具包（MRTK），掌握了虚拟场景开发、用户交互设计、实时渲染等关键技术。通过参与液压马达装配项目，我深入了解了工业装配流程、设备标准及安全生产规范，同时将计算机视觉、机器学习等跨学科知识应用于零部件识别与定位问题，实现了智能化的装配指导。此外，我还探索了少样本目标检测和分类技术，并成功发表了一篇相关论文，申请了一项专利，进一步巩固了我对专业知识的掌握和实践能力的提升。

2. 工程实践的经历(不少于200字)

在宁波沙塔信息技术有限公司的专业实践中，本人作为软件工程师深度参与了“基于混合现实的虚拟装配指导软件”项目开发。该项目的核心目标是利用混合现实技术解决液压马达装配过程中操作复杂、培训成本高等问题。实践期间的主要工作包括：

需求分析与系统设计：与工厂工程师深入沟通，明确装配流程中的痛点和需求，设计虚实融合的交互界面，并制定系统架构方案。

技术开发与实现：基于Unity引擎和HoloLens

2开发平台，构建虚拟装配场景，集成图像识别算法，实现零部件的实时定位与动画演示。

算法优化与创新：针对工业场景数据稀缺问题，研究小样本学习算法，提出基于CLIP模型的改进方法，显著提升了零部件分类精度。

系统测试与迭代：通过单元测试、集成测试及用户实地测试，优化交互逻辑与算法性能，最终完成支持文字、图片、视频及动画多模态指导的系统开发。

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)

一、项目背景

在企业工程实践过程中，我主要负责的项目是基于混合现实（MR）的液压马达装配指导系统的研发。液压马达是工业生产中关键的动力传输部件，其装配过程涉及多个复杂的步骤，对精度要求极高。传统的装配培训方式依赖于纸质手册或电子文档，存在信息获取不便、学习成本高、装配错误率高等问题。因此，希望开发一款智能化的装配指导系统，通过MR技术提供沉浸式、交互式的指导，提升装配效率和准确性。

二、主要研究目标：

1. 开发基于混合现实的虚拟装配软件

该软件需要具备虚实融合的能力，界面友好，接口规范，兼容HoloLens2等混合现实设备，满足软件开发需求，支持装配过程教学。

2. 实现液压马达虚拟装配操作指导

结合实际液压马达装配场景，以手势方式实现发动机零部件的虚拟装配，同时支持对实际装配的操作指导，并配有文字、图片、视频多种形式的说明以及操作步骤动画提示。

3. 提升装配效率和效果

利用虚拟装配技术降低成本，提高装配质量和效率。同时，通过虚拟环境进行装配操作培训，提升学员的实际操作能力。通过实时显示装配步骤和操作指南，减少工人培训时间，提高装配效率。

三、技术难点

1. 混合现实软件的实现

如何将虚拟装配指导与现实场景无缝结合，提供准确且稳定的增强现实体验。同时，还需解决虚拟场景的碰撞检测、装配约束等技术难题。

2. 智能识别

确保系统能够识别装配零件类别，以及识别装配工具和部件的位置与状态。具备较高的识别算法精度和响应速度，以确保识别与交互的准确性。

3. 装配流程设计

完成一套完善的装配流程指导设计和开发，确保用户能够按照标准操作流程进行装配。同时满足较高效的用户交互和指导，提供多种模态的信息引导。

四、本人承担任务及完成情况

1. 需求分析与系统设计

参与系统需求分析和设计，对需求进行整理，并进行技术调研，提出技术解决方案。

2. 虚拟软件系统开发

基于Unity游戏引擎，在HoloLens2上进行系统开发，满足装配流程指导功能。系统实现装配流程约束，提供包含文字、图片、视频和动画等多种形式的操作指导。最终，完成软件基础功能开发。

3. 图像分类和目标检测算法研究

针对工业场景下，图片等数据稀缺的问题，传统的图像分类和目标检测模型不能很好的检测零件图像。为了提高图像分类和目标检测的精度，考虑工业零件目标检测数据集的图像特性，研究小样本学习下的图像分类和目标检测算法。

五、成果与效益

技术成果：申请了一项发明专利，专利名称为“一种基于混合现实技术的液压马达装配指导系统和方法”；基于该项目，撰写了一篇论文并被PRCV 2024会议录用，论文题目为“Multi-Layer Tuning CLIP for Few-shot Image Classification”。

行业价值：该系统针对制造业装配环节的“三高三难”痛点，通过混合现实技术实现了装配流程的智能化与可视化，显著提升操作效率与准确性。

六、创新与总结

算法层面：将CLIP模型应用于工业小样本识别，提出多层次特征融合方法，基于小样本学习解决零件识别问题。

工程层面：通过“客户端-

服务器架构”平衡性能与资源限制，完成装配流程指导设计和开发，并融合多模态反馈与智能指导，确保用户能够按照标准操作流程进行装配。

工程实践经验：通过该项目，积累了多个领域的工程经验。提高了与企业工程师的沟通能力、团队协作能力，并学会了如何将科研成果转化为实际工程应用。

（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
Multi-Layer Tuning CLIP for Few-shot Image Classification	会议论文	2024年10月20日	PRCV 2024	1/7	EI收录
一种基于混合现实技术的液压马达装配指导系统和方法	发明专利申请	2023年11月17日	申请号：202311541120.8	1/4	

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: 86 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 年(要求1年及以上) 考核成绩: 83 分
本人承诺	
<p>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!</p> <p style="text-align: right;">申报人签名: 张睿昊</p>	

浙 江 大 学 研 究 生 院
攻读硕士学位研究生成绩表

学号：22260468		姓名：张睿昊		性别：男		学院：工程师学院		专业：计算机技术				学制：2.5年			
毕业时最低应获：27.0学分				已获得：29.0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		工程伦理			2.0	84	公共学位课	2022-2023学年春季学期		新时代中国特色社会主义思想理论与实践			2.0	90	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学年春季学期		自然辩证法概论			1.0	82	公共学位课
2022-2023学年秋季学期		工程数值分析			2.0	88	专业选修课	2022-2023学年春季学期		港口设施规划与布局			2.0	84	专业学位课
2022-2023学年秋季学期		创新设计方法			2.0	通过	专业选修课	2022-2023学年春季学期		研究生英语基础技能			1.0	72	公共学位课
2022-2023学年冬季学期		机器视觉及其应用			2.0	90	专业学位课	2022-2023学年春季学期		研究生论文写作指导			1.0	89	专业学位课
2022-2023学年冬季学期		海洋管理学			2.0	90	专业选修课	2022-2023学年夏季学期		数据科学前沿技术导论			2.0	92	专业学位课
2022-2023学年冬季学期		产业技术发展前沿			1.5	93	专业学位课	2022-2023学年春夏学期		高阶工程认知实践			3.0	84	专业学位课
2022-2023学年秋冬学期		研究生英语			2.0	92	公共学位课			硕士生读书报告			2.0	通过	

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：2025-03-20

22260468

二、日常表现考核评价及申报材料审核公示结果

日常表现 考核评价	非定向生由德育导师考核评价、定向生由所在工作单位考核评价： <input checked="" type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格 德育导师/定向生所在工作单位分管领导签字（公章）：  2024年3月17日
申报材料 审核公示	根据评审条件，工程师学院已对申报人员进行材料审核（学位课程成绩、专业实践训练时间及考核、学位论文、代表作等情况），并将符合要求的申报材料在学院网站公示不少于5个工作日，具体公示结果如下： <input type="checkbox"/> 通过 <input type="checkbox"/> 不通过（具体原因： 工程师学院教学管理办公室审核签字（公章）：) 年 月 日

[Home](#) > [Pattern Recognition and Computer Vision](#) > Conference paper

Multi-layer Tuning CLIP for Few-Shot Image Classification

| Conference paper | First Online: 20 October 2024

| pp 173–186 | [Cite this conference paper](#)



Pattern Recognition and Computer Vision
(PRCV 2024)

[Ruihao Zhang](#), [Jinsong Geng](#), [Cenyu Liu](#), [Wei Zhang](#), [Zunlei Feng](#), [Liang xue](#) & [Yijun Bei](#)

Part of the book series: [Lecture Notes in Computer Science](#) ((LNCS, volume 15035))

Included in the following conference series:
[Chinese Conference on Pattern Recognition and Computer Vision \(PRCV\)](#)

225 Accesses

Abstract

CLIP bridges the gap between visual and language by learning both image and text representations simultaneously. As a large pre-trained visual language model, CLIP is highly generalisable and has demonstrated excellent few-shot learning capabilities. Numerous studies have been conducted on CLIP models for few-shot learning of

Multi-layer Tuning CLIP for Few-Shot Image Classification

Ruihao Zhang¹, Jinsong Geng², Cenyu Liu², Wei Zhang², Zunlei Feng^{2,3},
Liang xue⁴, and Yijun Bei²(✉)

¹ Polytechnic Institute, Zhejiang University, China

² School of Software Technology, Zhejiang University, China beiyj@zju.edu.cn

³ Key Laboratory of Visual Perception (Zhejiang University) , Ministry of Education
and Microsoft

⁴ Suzhou City University, China

Abstract. CLIP bridges the gap between visual and language by learning both image and text representations simultaneously. As a large pre-trained visual language model, CLIP is highly generalisable and has demonstrated excellent few-shot learning capabilities. Numerous studies have been conducted on CLIP models for few-shot learning of downstream visual tasks, all of which have demonstrated excellent results. However, current research methods, such as the Adapter and Prompts methods, still fall short in extracting visual features for CLIP. Many methods only fine-tune the adapter after feature extraction, failing to fully utilise the feature extraction potential of CLIP. In addition, the fine-tuning approach using key-value cache improves performance significantly, but it requires careful tuning of the model’s hyperparameters for a specific dataset. Considering these issues, we propose a new approach: fine-tuning the multi-layer features with side adapters for adaptive selection in the visual backbone network. This approach efficiently extracts effective visual features for different layers of the task. Additionally, we propose augmenting the original features using dynamic feature fusion to reduce reliance on hyper-parameter tuning. Extensive experiments are conducted on 11 datasets to verify the superiority of the proposed method over existing methods.

Keywords: Few-Shot · CLIP · Multi-layer · Dynamic feature fusion

1 Introduction

Deep learning technologies have achieved significant success in various computer vision tasks, including image classification, object detection, and semantic segmentation. However, these methods rely heavily on training with large-scale datasets, which can be challenging to obtain in real-world scenarios, such as in industrial and military domains. Consequently, few-shot learning has emerged as a crucial solution to enable models to learn generalized feature representations from a limited amount of labeled data [1].

经检索“Engineering Village”，下述论文被《Ei Compendex》收录。（检索时间：2024年12月12日）。

<RECORD 1>

Accession number:20244517329410

Title:Multi-layer Tuning CLIP for Few-Shot Image Classification

Authors:Zhang, Ruihao (1); Geng, Jinsong (2); Liu, Cenyu (2); Zhang, Wei (2); Feng, Zunlei (2, 3); xue, Liang (4); Bei, Yijun (2)

Author affiliation:(1) Polytechnic Institute, Zhejiang University, Hangzhou, China; (2) School of Software Technology, Zhejiang University, Hangzhou, China; (3) Key Laboratory of Visual Perception, Zhejiang University, Ministry of Education and Microsoft, Hangzhou, China; (4) Suzhou City University, Suzhou, China

Corresponding author:Bei, Yijun(beiylj@zju.edu.cn)

Source title:Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

Abbreviated source title:Lect. Notes Comput. Sci.

Volume:15035 LNCS

Part number:5 of 15

Issue title:Pattern Recognition and Computer Vision - 7th Chinese Conference, PRCV 2024, Proceedings

Issue date:2025

Publication year:2025

Pages:173-186

Language:English

ISSN:03029743

E-ISSN:16113349

ISBN-13:9789819786190

Document type:Conference article (CA)

Conference name:7th Chinese Conference on Pattern Recognition and Computer Vision, PRCV 2024

Conference date:October 18, 2024 - October 20, 2024

Conference location:Urumqi, China

Conference code:321879

Publisher:Springer Science and Business Media Deutschland GmbH

Number of references:39

Main heading:Zero-shot learning

Controlled terms:Visual languages

Uncontrolled terms:CLIP - Dynamic feature fusion - Dynamic features - Features extraction - Features fusions - Few-shot -

Fine tuning - Hyper-parameter - Multi-layers - Visual feature

Classification code:1101.2 - 1106.1.1

DOI:10.1007/978-981-97-8620-6_12

Funding details: Number: 2023Z057, Acronym: -, Sponsor: Key Research and Development Program of Ningxia;Number:

-, Acronym: -, Sponsor: Key Research and Development Program of Ningxia;Number: 2023J281, Acronym: -, Sponsor:

Natural Science Foundation of Ningbo Municipality;Number: -, Acronym: -, Sponsor: Natural Science Foundation of

Ningbo Municipality;Number: 226-2024-00058, Acronym: -, Sponsor: Fundamental Research Funds for the Central

Universities;Number: -, Acronym: -, Sponsor: Fundamental Research Funds for the Central Universities;

Funding text:This work is supported by the Ningbo Key Research and Development Program (Grant No. 2023Z057),

Ningbo Natural Science Foundation (2023J281), and the Fundamental Research Funds for the Central Universities

(226-2024-00058).

Database:Compendex

Compilation and indexing terms, Copyright 2024 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。





(12) 发明专利申请



(10) 申请公布号 CN 117541754 A

(43) 申请公布日 2024. 02. 09

(21) 申请号 202311541120.8

(22) 申请日 2023.11.17

(71) 申请人 浙江宇丰信息技术有限公司

地址 315200 浙江省宁波市镇海区庄市街
道西陆路288号萌恒大厦中楼4楼

(72) 发明人 张睿昊 刘二腾 吴连秋 钟钊瑜

(74) 专利代理机构 常州智慧腾达专利代理事务
所(普通合伙) 32328

专利代理师 杜杰

(51) Int.Cl.

G06T 19/00 (2011.01)

G06F 9/451 (2018.01)

G06V 10/25 (2022.01)

G06V 10/774 (2022.01)

G06V 10/776 (2022.01)

G06V 20/20 (2022.01)

G06T 7/73 (2017.01)

G06V 10/82 (2022.01)

G06N 3/0464 (2023.01)

G06N 3/096 (2023.01)

权利要求书3页 说明书8页 附图4页

(54) 发明名称

一种基于混合现实技术的液压马达装配指
导系统和方法

(57) 摘要

本发明公开了一种基于混合现实技术的液压马达装配指导系统,所述系统包括服务器端和设备终端,在服务器端通过建立液压马达零部件目标检测模型可以基于获取到的液压马达零部件的图像信息直接生成与其对应的目标检测结果,在设备终端接收到该目标检测结果,通过在设备终端对该目标检测结果进行坐标变换可以得到其的真实位置,基于该真实位置可以对液压马达零部件的装配位置进行装配指导。针对液压马达装配过程中零件多样、装配过程复杂等特点,克服传统纸质或电子操作指导的不便,结合混合现实技术,实现液压马达装配指导系统。

