同行专家业内评价意见书编号：20250854371

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： 李森杰

学号： 22260105

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14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)

本人系统掌握了机器视觉、物联网及安防监控领域的专业理论和技术知识，包括深度学习、计算机视觉算法、图像语义表示等关键技术方法。具体而言，在深度学习方面，精通卷积神经网络（CNN）的基本结构和优化策略，如卷积运算、批量归一化、残差连接、深度可分离卷积等。掌握了基于图像语义表征的视频监控异常检测方法，熟悉自监督学习、对比学习以及知识蒸馏理论。此外，还深入理解了多模态学习方法（如CLIP、VIST），掌握了如何从多模态大模型中有效提取语义信息并用于实际的图像表征优化。本人还熟悉资源受限设备上的模型轻量化设计与优化部署的相关技术，能够针对不同硬件平台高效实现视觉算法部署和应用。

2.工程实践的经历(不少于200字)

在工程实践方面，本人积极参与了浙江大学与相关企业的合作项目，承担了物联网终端设备上异常事件自动检测的关键研发工作。首先，参与了针对低成本视频监控设备（如IoT终端芯片K210）的异常检测技术研发，独立完成了基于图像语义表征的轻量化模型开发与优化。其次，在实验室环境与实际场景中，搭建并调试了轻量级卷积神经网络，基于图像语义特征实现异常检测，解决了在资源受限终端设备上高效部署的技术难题。此外，通过自主开发的图像编辑管道生成训练数据，增强了模型泛化能力。工程实践使本人深入理解从需求分析、方案制定、模型设计到落地部署的完整流程，并积累了丰富的物联网与安防监控领域的项目经验。

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）

本人在实际工程工作中承担并解决了“面向低成本视频监控终端的异常事件智能检测系统”的研发工作。传统视频监控系统虽然广泛应用，但在异常事件的自动化实时监测上存在明显不足，依靠人工长时间观察视频流效率低下且容易疏漏关键事件。针对这一问题，本人综合运用机器视觉、物联网、深度学习等技术知识，开发了适合低成本IoT终端的轻量级视频监控异常检测方法。

在项目启动阶段，我充分分析了视频监控设备受限的算力和存储能力等限制条件，选择适用于实际场景的轻量级CNN模型架构。在具体研发中，设计了一种基于图像语义表征的异常检测方法，将视频帧转换为特征嵌入向量，并与预定义的正常样本进行比较，从而判断异常事件的发生。这种方法避免了传统监督学习对特定异常事件的依赖，具备较强的通用性和可扩展性。同时，通过自监督学习与知识蒸馏方法，将多模态大模型的语义特征隐式地融入轻量化单模态CNN模型中，使其在低成本、资源受限的IoT芯片（如K210）上运行，最终实现了异常检测精度达98.7%，检测速度达到每秒9.4帧，满足了企业实际部署的严格要求。

此外，为了克服数据不足和多样性问题，开发了一种创新的数据生成管道，结合多模态大模型理解场景语义，自动生成各种潜在异常图像进行训练数据扩增。这有效提高了模型对不同场景的泛化性和可靠性。

在企业实际应用场景中，如何快速、准确地自动检测异常事件（如火灾、入侵或违规停车）一直是困扰行业的复杂工程问题。以往的解决方案依赖于预定义的异常事件检测模型，这种

方案的通用性差、模型复杂度高，难以部署在低成本IoT设备上。本人的研究旨在解决这一现实难题，并已在实际工作中取得了显著成效。

本项目首先面临的复杂问题是如何在资源有限的低成本终端设备上，实现高精度、实时的视频异常检测。本人在深入分析需求后，选择了基于图像语义表征的异常检测框架。这种方法不仅能够满足实时检测的要求，还能适应不同的异常场景，解决了以往方法普适性不足的问题。

研发初期，本人发现传统基于对比学习的图像嵌入方法难以明确包含语义信息，严重影响异常检测的准确性。为此，提出了一种创新的轻量级图像语义表征训练方法，通过知识蒸馏技术，将多模态大型模型（如CLIP、VIST）所具有的丰富语义信息转化为单模态的CNN模型权重。这一过程有效提升了CNN模型在有限计算资源下的语义表征能力。具体而言，本人设计了一个由Triplet Loss和均方误差（MSE）损失联合组成的损失函数框架，确保模型能够有效区分正常与异常事件。这一方法经实验证明在本企业的真实场景中取得了98.7%的异常检测准确率，远高于传统的SIFT特征匹配（56.2%）和单纯的CNN分类模型（77.6%），有效地解决了传统方法泛化能力差的问题。

在实际实施中，为了增强模型泛化性和对未知异常事件的检测能力，本人进一步开发了一种图像到图像的自动生成技术。通过多模态大模型（如GPT-4o、LLAVA）理解图像语义场景，再利用图像编辑模型生成异常图像，构建了超过1000张高难度负样本数据集，解决了数据不足且收集困难的复杂工程难题。这种创新的数据增强策略大幅提高了模型在实际场景下的表现，使得模型在各种复杂场景中依然保持高度准确的检测能力。

同时，针对企业实际需求，本人还优化设计了一种适合IoT终端设备的轻量级CNN网络结构。该网络仅采用基础的卷积、池化、批归一化和残差连接操作，避免了复杂算子的应用。通过精确的模型结构设计，减少参数量，仅为6.1M参数，运行速度和精度均优于同类传统架构（如ResNet、MobileNet）。在资源受限的企业级IoT设备（K210芯片）上成功部署后，企业反馈表明，系统实际应用效果显著，异常事件漏检率低、误报少，显著提高了安全管理效率，并节约了大量的人力成本，获得了显著的经济和社会效益。

综上所述，本人通过全面、深入地掌握的机器视觉、物联网与人工智能专业知识，创造性地提出并成功落地了一套低成本、高效的异常事件检测系统，有效地解决了企业视频监控中面临的实际复杂工程问题，实现了显著的技术创新和应用价值。

（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
A Lightweight Video Surveillance Anomaly Detection Method Based on Image Semantic Representation	会议论文	2024年10月14日	International Conference on Artificial Intelligence of Things and Systems (AIoTSys 2024)	1/3	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： 88 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： 1.6 年（要求1年及以上） 考核成绩： 83 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：**李蕤杰	

22260105

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： ☑优秀　　☐良好　　☐合格　　☐不合格 德育导师/定向生所在工作单位分管领导签字（公章）：　　　　年 3 月 14 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： ☐通过　　☐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

浙 江 大 学 研 究 生 院

攻读硕士学位研究生成绩表

学号：22260105	姓名：李森杰	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：30.0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	工程技术创新前沿		1.5	91	专业学位课	2022-2023学年春季学期	研究生英语基础技能		1.0	77	公共学位课
2022-2023学年秋季学期	人工智能算法与系统		2.0	100	跨专业课	2022-2023学年春夏学期	高阶工程认知实践		3.0	89	专业学位课
2022-2023学年秋季学期	新时代中国特色社会主义理论与实践		2.0	90	专业学位课	2022-2023学年夏季学期	物联网信息安全技术与应用基础		2.0	95	专业选修课
2022-2023学年秋冬学期	数据分析的概率统计基础		3.0	93	专业选修课	2022-2023学年春夏学期	工程伦理		2.0	85	专业学位课
2022-2023学年冬季学期	物联网操作系统与边缘计算		2.0	92	专业选修课	2022-2023学年夏季学期	自然辩证法概论		1.0	81	专业学位课
2022-2023学年秋冬学期	研究生论文写作指导		1.0	79	专业选修课	2022-2023学年春夏学期	移动互联网智能设备应用设计与实践		3.0	93	专业学位课
2022-2023学年冬季学期	产业技术发展前沿		1.5	96	专业学位课	2023-2024学年秋季学期	歌唱艺术		1.0	73	跨专业课
2022-2023学年秋冬学期	研究生英语		2.0	87	专业学位课		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。

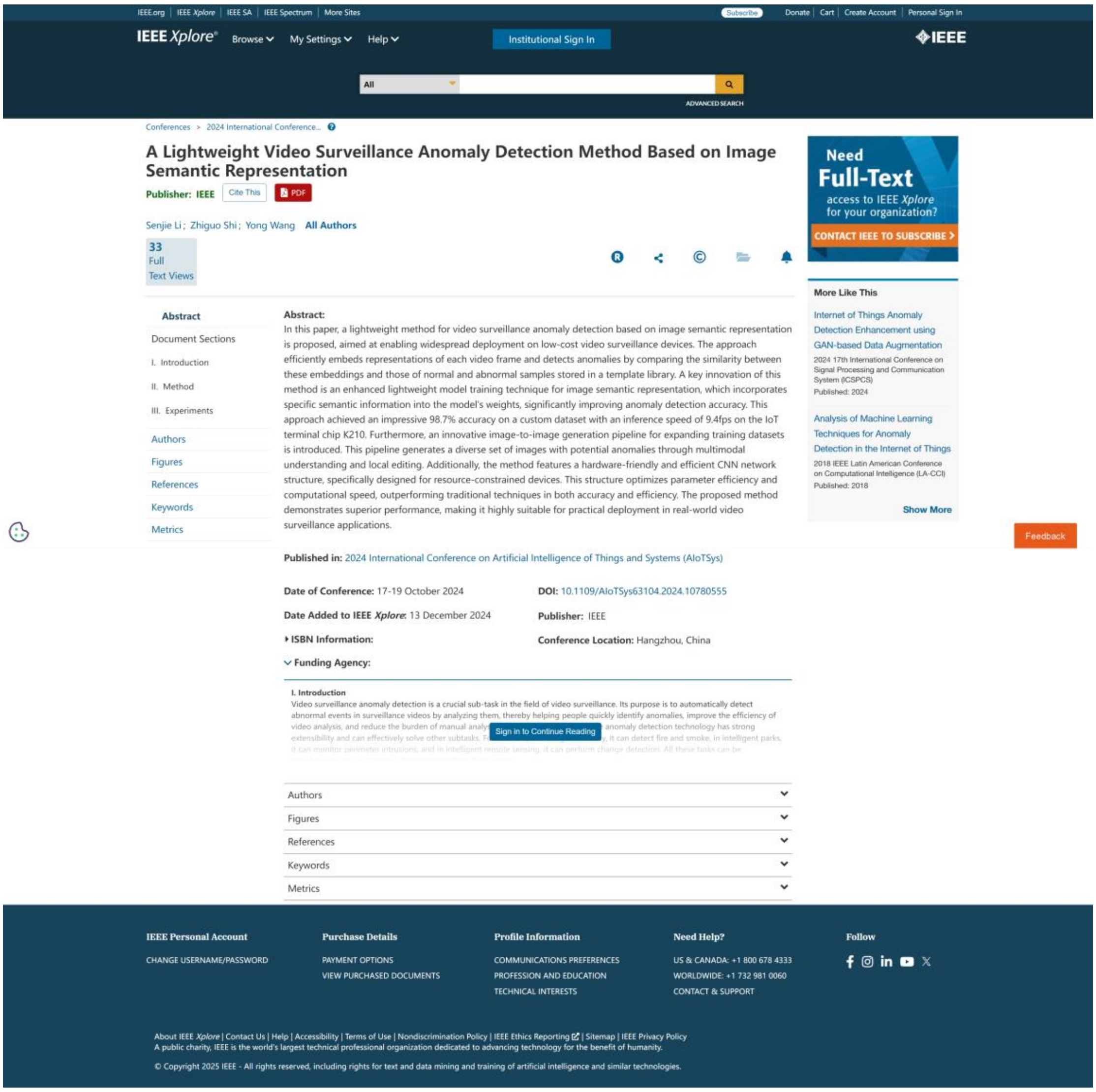
学院成绩校核章

成绩校核人：张梦依

打印日期：2025-03-20

浙江大学研究生（60）成绩校核章

1. 论文网络搜索页截图：https://ieeexplore.ieee.org/document/10780555



2. 论文首页：
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10780555

A Lightweight Video Surveillance Anomaly Detection Method Based on Image Semantic Representation

Senjie Li[1], Zhiguo Shi[1], and Yong Wang[1,*]

[1]College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[*]Corresponding author, Email: wangy@zju.edu.cn

2024 International Conference on Artificial Intelligence of Things and Systems (AIoTSys) | 979-8-3315-0788-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIOTSYS63104.2024.10780555

***Abstract*—In this paper, a lightweight method for video surveillance anomaly detection based on image semantic representation is proposed, aimed at enabling widespread deployment on low-cost video surveillance devices. The approach efficiently embeds representations of each video frame and detects anomalies by comparing the similarity between these embeddings and those of normal and abnormal samples stored in a template library. A key innovation of this method is an enhanced lightweight model training technique for image semantic representation, which incorporates specific semantic information into the model's weights, significantly improving anomaly detection accuracy. This approach achieved an impressive 98.7% accuracy on a custom dataset with an inference speed of 9.4fps on the IoT terminal chip K210. Furthermore, an innovative image-to-image generation pipeline for expanding training datasets is introduced. This pipeline generates a diverse set of images with potential anomalies through multimodal understanding and local editing. Additionally, the method features a hardware-friendly and efficient CNN network structure, specifically designed for resource-constrained devices. This structure optimizes parameter efficiency and computational speed, outperforming traditional techniques in both accuracy and efficiency. The proposed method demonstrates superior performance, making it highly suitable for practical deployment in real-world video surveillance applications.**

***Index Terms*—Video Surveillance, Anomaly Detection, Image Semantic Representation, Lightweight Model**

I. Introduction

Video surveillance anomaly detection is a crucial subtask in the field of video surveillance. Its purpose is to automatically detect abnormal events in surveillance videos by analyzing them, thereby helping people quickly identify anomalies, improve the efficiency of video analysis, and reduce the burden of manual analysis. Moreover, video surveillance anomaly detection technology has strong extensibility and can effectively solve other subtasks. For instance, in intelligent security, it can detect fire and smoke, in intelligent parks, it can monitor perimeter intrusions, and in intelligent remote sensing, it can perform change detection. All these tasks can be transformed into an anomaly detection paradigm for solution.

The core of video surveillance anomaly detection lies in defining normal and abnormal events in surveillance videos, which can generally be divided into two approaches: recognition and comparison. Recognition refers to detecting anomalies by identifying predefined abnormal events in surveillance videos, such as fires, intrusions, and illegal parking. This approach typically involves developing recognition algorithms for each predefined abnormal event, such as classification and detection algorithms based on deep learning [1]–[4]. The advantage of this method is that it can output specific types of anomalies. However, its drawbacks include high complexity and a lack of universality, as each type of anomaly requires a targeted recognition algorithm, and many anomalies in real life cannot be predefined.

Consequently, the current mainstream trend is to solve anomaly detection problems through comparison. This approach involves converting images into vector representations and detecting anomalies by comparing the representation of the current frame with predefined normal frames. The benefits of this method include strong generalization and extensibility. Generalization is reflected in its ability to adapt to different environments and scenes; anomaly detection can be performed by simply collecting a normal frame from the current environment, minimizing the impact of domain distribution shifts. Extensibility is demonstrated by the ability to set multiple normal frame templates simultaneously; false positives can be reduced by adding comparison templates without modifying the model and algorithm. The disadvantage is the lack of specific anomaly types, but with the continuous development of contrastive learning techniques and accuracy improvements, specific anomaly types can also be identified by adding anomaly frame templates.

There are various methods based on comparison, including image feature operator methods like SIFT [5] and SSIM [6], deep learning-based image feature methods like LPIPS [7], and methods based on Siamese networks [8]–[10]. Recently, with the rise of self-supervised pretraining paradigms for neural networks based on contrastive learning [11]–[15], the representation capabilities of images have been gradually improved, and neural network model architectures have evolved from convolutional neural networks (CNNs) [16] to Transformers [17]. Moreover, multimodal representation learning has become a research hotspot, with various methods for image-text comparison

3. EI 收录证明：
https://www.engineeringvillage.com/search/quick.url?SEARCHID=68aec7cd992345839acdd333644d4b74&COUNT=1&usageOrigin=recordpage&usageZone=backtoresultstop