

学校代码: 10284

分类号: TP181

密级: 公开

U D C: 004.93

学号: MG20330007



南京大學

硕士学位论文

论文题目 基于变化对比检测的巡

检系统设计与实现

作者姓名 陈昊

专业名称 计算机科学与技术

研究方向 人工智能与模式识别

导师姓名 申富饶教授

2023年5月24日

答辩委员会主席 戴新宇 教授

评 阅 人 张 荆 高工

徐明华 教授

论文答辩日期 2023年5月22日

研究生签名: 陈昊

导师签名: 申嘉屹

Change Contrast Detection Based Inspection System Design and Implementation

by

Chen Hao

Supervised by

Professor Furao Shen

A dissertation submitted to

the graduate school of Nanjing University

in partial fulfilment of the requirements for the degree of

MASTER

in

Computer Science and Technology



Department of Computer Science and Technology

Nanjing University

May 24, 2023

南京大学学位论文原创性声明

本人郑重声明，所提交的学位论文是本人在导师指导下独立进行科学研究工作所取得的成果。除本论文中已经注明引用的内容外，本论文不包含其他个人或集体已经发表或撰写过的研究成果，也不包含为获得南京大学或其他教育机构的学位证书而使用过的材料。对本文的研究做出重要贡献的个人和集体，均已在论文的致谢部分明确标明。本人郑重声明愿承担本声明的法律责任。

研究生签名：

日期：

南京大学研究生毕业论文中文摘要首页用纸

毕业论文题目：基于变化对比检测的巡检系统设计与实现

计算机科学与技术 专业 2020 级硕士生姓名：陈昊

指导教师（姓名、职称）：申富饶 教授

摘 要

随着现代工业的不断发展，工业生产的自动化程度越来越高，为了确保生产的正常进行，设备以及设施的日常检视维护显得尤为重要。过去，这些工作往往由巡检工人来完成。人工作业方式不但需要支付高昂的用人成本，而且不可避免一些由于人员疏忽、漏检带来的损失，甚至在一些高危极端环境下人工巡检有极大的安全隐患。近年来，智能巡检越来越被应用于替换人工作业巡检，然而，这些系统往往基于目标检测模型，不仅需要大量人工标注的训练数据，存在冷启动问题，而且模型参数量大，需要的计算设备成本居高不下。针对上述问题，本文提出一种基于对比变化检测的智能巡检系统，使用对比基准图和实时影像图片的方式，对于设备或者设施异常进行检测识别。本文的主要研究工作如下：

1. 基于前人研究基础上，提出一种更为鲁棒的变化检测方法。该方法使用对光照条件和视角变化泛化能力更强的关键点检测模型 SuperPoint，对基准图和实时影像图片提取关键点和对应的特征描述子，并利用快速最近邻搜索算法对两张图片中的关键点进行匹配。由于初步得到的匹配点对存在不少误匹配，再使用基于向量场平滑一致性的 VFC 算法对匹配点进行筛选，得到质量较好的匹配点对。由于单个单应性矩阵的局限性，本文采用多单应性矩阵估计的方式，对图片的中不同部分选取不同的变换矩阵对其进行变换，然后比对相应位置的特征描述子，经过后续处理得到变化区域。在公开数据集和实验室环境数据下，新方法对比传统方法，在不损失速度的情况下，检测效果更好。
2. 为了识别变化区域是否是某个已知异常类别，本文基于原型学习和对比学习提出一种开集识别方法 ConSim-CNN。为了使模型既能在已知类别上正

确识别实例，又能拒绝从未知类中采样的样本，ConSim-CNN 去除了经典深度网络的线性分类层和 Softmax 函数，直接使用特征向量计算与每个原型向量的相似度，如果最大相似度大于阈值，则接受为对应已知类别，否则为未知类别。我们使用有监督对比学习损失，以及原型向量正则损失，在仅有已知类的训练数据上进行训练和估计阈值，在加入了未知类的测试集上进行测试。在变电站异常检测数据集以及 Tiny-Imagenet 等公开数据集上相比传统方法有一定提升。

3. 本文基于上述算法实现了基于对比变化检测的智能巡检系统，系统采用边缘端、云服务器端、用户端的空间布局方式，尽可能地将流量和计算保留在边缘端进行。系统使用 ZMQ 消息队列实现了低延迟、低流量、高可用的实时视频传输功能，结合 Flask 框架和 gRPC 框架实现了算法调用和功能接口，基于前端框架设计实现了用户界面，方便对于巡检过程进行监视和操作。系统还实现了用户登录，基准图管理等功能，使得系统完整可靠。

关键词：变化检测；开集识别；巡检系统

南京大学研究生毕业论文英文摘要首页用纸

THESIS: Change Contrast Detection Based Inspection System Design
and Implementation

SPECIALIZATION: Computer Science and Technology

POSTGRADUATE: Chen Hao

MENTOR: Professor Furao Shen

ABSTRACT

With the continuous development of modern industry, the degree of automation in industrial production is increasing, and the daily inspection and maintenance of equipment and facilities are particularly important to ensure the normal operation of production. In the past, these tasks were often completed by inspection workers. The manual work not only requires high labor costs but also inevitably incurs losses due to human negligence and missed inspections, and even has great safety hazards in some high-risk extreme environments. In recent years, intelligent inspection has been increasingly applied to replace manual inspection. However, these systems are often based on target detection models, requiring a large amount of manually annotated training data and suffering from cold start problems. Moreover, the model's parameters are large, and the required computing equipment costs are high. To address these issues, this paper proposes an intelligent inspection system based on contrastive change detection, which detects and identifies equipment or facility abnormalities using a comparison of baseline image and real-time video images. The main research work of this paper is as follows:

1. Based on previous research, a more robust change detection method is proposed. This method uses the SuperPoint keypoint detection model, which has stronger generalization ability to illumination conditions and viewpoint changes, to extract keypoints and corresponding feature descriptors from the baseline image and real-time video images. The Fastest Approximate Nearest Neighbor Search algorithm is used to match the keypoints in the two images. Due to the existence of many

false matches in the preliminary matching point pairs, the VFC algorithm based on vector field smoothing consistency is used to filter the matching point pairs to obtain better quality matching point pairs. Due to the limitations of a single homography matrix, this paper adopts the method of estimating multiple homography matrices to select different transformation matrices for different parts of the image to perform transformations and then compare the feature descriptors at the corresponding positions to obtain the change region. Under public dataset and laboratory environment data, the new method has better detection performance compared to traditional methods without sacrificing speed.

2. In order to identify whether the change region belongs to a known abnormal category, this paper proposes an open-set recognition method ConSim-CNN based on prototype learning and contrastive learning. In order to enable the model to correctly identify instances on known categories and reject samples sampled from unknown categories, ConSim-CNN removes the linear classification layer and Softmax function of the classic deep network and directly calculates the similarity between the feature vector and each prototype vector. If the maximum similarity is greater than the threshold, it is accepted as the corresponding known category, otherwise it is an unknown category. We use supervised contrastive learning loss and prototype vector regularization loss to train and estimate the threshold on training data with only known categories and test it on the test set with unknown categories. It has a certain improvement compared to traditional methods on the abnormal detection dataset of substations and public datasets such as Tiny-Imagenet.
3. Based on the above algorithm, this paper has implemented an intelligent inspection system based on change contrastive detection. The system adopts the spatial layout of edge-side, cloud server-side, and user-side to reserve traffic and computing as much as possible at the edge-side. The system uses ZMQ message queues to achieve low-latency, low-traffic, and high-availability real-time video transmission function, combines Flask and gRPC frameworks to achieve algorithm calling and functional interfaces, and designs and implements a user interface based on

the front-end framework, making it convenient for monitoring and operating the inspection process. The system also implements user login, baseline image management, and other functions, making the system complete and reliable.

KEYWORDS: change detection, open set recognition, inspection system

目 录

中文摘要	I
ABSTRACT	III
目 录	VII
插图目录	XI
表格目录	XIII
第一章 绪论	1
1.1 研究背景	1
1.2 研究现状	3
1.2.1 变化检测研究现状	3
1.2.2 基于图像的智能巡检研究现状	3
1.2.3 现有问题和不足	5
1.3 本文工作	5
1.4 章节安排	8
第二章 相关工作	11
2.1 变化检测部分相关算法	11
2.1.1 关键点检测算法	12
2.1.2 RANSAC	16
2.1.3 VFC 匹配算法	17
2.1.4 单应性矩阵	19
2.2 开集识别部分相关算法	21
2.2.1 开集识别	21

2.2.2	模型鲁棒性：原型学习和对比学习	23
2.3	系统相关技术	25
2.3.1	视频推流	25
2.3.2	Flask 框架	26
2.3.3	MongoDB	27
第三章	需求分析	29
3.1	可行性分析	29
3.2	功能需求分析	30
3.2.1	算法功能需求	32
3.2.2	系统功能需求	33
3.3	非功能需求分析	34
3.4	本章小结	36
第四章	系统设计	37
4.1	系统总体设计	37
4.2	算法设计	39
4.2.1	变化检测算法	40
4.2.2	开集识别算法设计	47
4.2.3	前处理算法设计	51
4.2.4	后处理算法设计	56
4.3	基于 React 的客户端设计	57
4.4	本章小结	59
第五章	模块实现	61
5.1	前后处理模块实现	61
5.1.1	前处理模块实现	61
5.1.2	后处理模块实现	64
5.2	异常检测模块实现	65
5.2.1	变化检测部分	66
5.2.2	开集识别部分	69

5.3	视频推流模块实现	73
5.4	基准图管理模块实现	74
5.5	本章总结	76
第六章	系统测试	77
6.1	测试环境	77
6.2	模块测试	78
6.2.1	变化检测模块测试	78
6.2.2	开集识别模块测试	82
6.3	集成测试	87
6.3.1	测试准备	88
6.3.2	初始化阶段测试	89
6.3.3	运行时阶段测试	89
6.4	非功能性测试	91
6.4.1	负载测试	91
6.4.2	性能测试	92
6.5	运行结果	92
6.6	本章小结	94
第七章	结语	95
7.1	全文总结	95
7.2	未来展望	96
	参考文献	99
	致 谢	115
	简历与科研成果	117
	学位论文出版授权书	119

插图目录

1-1	模型精度与每类别图片数的关系	4
1-2	本文章节结构	8
2-1	高斯金字塔	13
2-2	局部极值点	14
2-3	SuperPoint 模型结构图	15
2-4	同态适应过程过程	16
2-5	RANSAC 拟合	17
2-6	不同图像和视角中的相同点	21
3-1	巡检系统流程图	31
4-1	系统总体框架	38
4-2	向量流插值	41
4-3	理想情况下内点集合确定的区域	46
4-4	实际情况下多单应性矩阵	46
4-5	CNN 在 MNIST 数据集上学到的特征表示	48
4-6	径向畸变	51
4-7	偏离示意图 1	52
4-8	偏离示意图 2	52
4-9	灰度图像的区域划分	54
4-10	遮挡判定：当摄像头被遮挡时，某些区域的特征点数量会大幅降低	54
5-1	标定板	62
5-2	标定结果示意	62
5-3	前处理模块示意图	64

5-4	异常检测模块结构图	66
5-5	变化检测示意图	66
5-6	Con-SimCNN 推理示意图	70
5-7	非异常数据构造示例	71
6-1	基准图	82
6-2	基准方法变化检测效果示意	83
6-3	使用 VFC 算法变化检测效果示意	83
6-4	t-sne 可视化	86
6-5	机器人平台	88
6-6	室内场景巡检示意图	90
6-7	测试场景示例	90
6-8	JMeter 负载测试	92
6-9	系统登录界面	93
6-10	监视界面	93
6-11	图片管理界面	93
6-12	异常检测结果展示	94

表格目录

4-1	模糊检测方法比较	56
5-1	数据增强方法列表	72
5-2	基准图类成员	75
5-3	枚举类 StatusEnum 成员	76
6-1	服务器硬件配置	77
6-2	边缘设备硬件配置	78
6-3	个人主机测试环境	78
6-4	Hpatches 关键点匹配实验, 以不同光照场景为测试对象	80
6-5	Hpatches 关键点匹配实验, 以不同视点场景为测试对象	81
6-6	CD2014 视频变化检测实验	81
6-7	开集识别结果	85
6-8	开集识别: 消融实验	85
6-9	异常与非异常判别	87
6-10	异常分类	87
6-11	初始化阶段: 直接上传基准图片测试	89
6-12	初始化阶段: 巡检过程上传基准图片测试	89
6-13	运行时阶段: 监视、检测功能测试	91
6-14	异常检测模块性能测试结果	92

第一章 绪论

1.1 研究背景

近年来，深度学习已经成为人工智能领域的热点话题之一，其在图像识别、自然语言处理、语音识别、机器翻译等任务上都取得了巨大的成功^[1]。深度学习是一种模拟人类大脑神经网络的机器学习技术，其核心是多层神经网络，通过反向传播算法优化网络参数，从而实现对复杂数据的高效处理和学习^[2]。深度学习技术的发展得益于计算硬件、数据资源和算法的不断提升。特别是在计算硬件方面，GPU 和 TPU 等专门的硬件加速器的出现，为深度学习的高效实现提供了重要支持。数据是深度学习算法的前提，源于图像和视频数据的增长，深度学习算法被广泛用于解决计算机视觉领域的任务，这其中最令人感兴趣的任务之一变化检测，它是指识别两张或多张在同一场景拍摄的图像，在不同时间、光照、视点或更复杂条件下的状态差异的过程^[3]。

根据目标区域相对传感器的距离不同，变化检测^[4]任务可以细分为遥感影像变化检测和（近）场景变化检测^[5]。遥感影像变化检测得益于遥感技术的不断发展，高分辨率、多光谱和高光谱遥感数据的获取和应用已经成为了现实。源于多时相遥感数据、合成孔径雷达数据以及医学图像的发展，变化检测在遥感领域和医学领域已经取得了巨大的成功，目前已经提出了很多深度学习方法用于遥感和医学变化检测，如 STANet^[6]、ChangeNet^[7]等基于孪生网络的算法。场景变化检测（Scene Change Detection,）是指在给定两个场景的情况下，定位变化并识别变化类别的任务，这两个场景在同一地点拍摄，时间相差一定^[8]。通常情况下，先拍摄的场景被称为参考场景（在 t_0 时刻拍摄），后拍摄的场景被称为查询场景（在 t_1 时刻拍摄）。场景数据可以是（深度）彩色图像、视频或 3D 点云。如果场景数据是图像，则场景变化检测是一种像素级预测，因为图像中的每个像素都被分类为某个类别。另一方面，如果场景数据是点云，则场景变化检测是

一种点级预测，因为点云中的每个点都被分类为某个类别。

不同于遥感影像变化检测，场景变化检测的研究还处于非常初步的阶段。这一方面是由于场景变化检测数据严重缺失，另一方面也是由于相应应用场景稀缺，需求不足。遥感影像变化检测有 AirChange^[9]、AICD^[10]、WHU^[11]、S2Looking^[12]、BANDON^[13]、GETNET^[14] 等大量数据集，时间跨度从 2008 到 2023，数据涵盖多光谱和高光谱等，种类丰富，研究热度不减。而场景变化检测，目前研究比较多的还是基于室外街景变化的较为粗粒度的数据，例如 CD2014^[15]、PCD^[16]、VL-CMU-CD^[17]。室内场景几乎是点云数据，直到最近才有 ChangeSim^[8] 这样使用模拟数据高精度仿真真实场景的数据，同时提供了图片和点云数据，相应的研究也没有形成系统。

除了上述数据集问题，应用场景的稀缺也是一大原因。本文是首次将变化检测的方法引入到现代工业制造中的智能巡检领域。在工业制造领域，一个很重要的工作就是对于机器设备的日常管理。机器发生老化、故障甚至意外损毁，轻则影响生产效率，重则危害生产安全甚至酿成重大生产事故，造成生命财产的损失。过去，这种设备监测主要是依靠人工巡检作业方式。在高压、超高压以及恶劣气象条件下，人工作业方式不仅人身危险性大，而且对设备安全运行也带来一定隐患。智能巡检系统利用巡检机器人沿着既定导航路线进行设备自动巡检，利用其自带的可见光和热成像摄像头实现工业设备或者物流仓储的无人值守或少人值守。智能巡检系统具有以下优势：减少人员疏忽、漏检等带来的设备损失，提高设备的运行质量；减少人员投入，降低人员成本；减少辐射等环境对于人的慢性健康危害。

目前，市面上存在相应的成熟系统，但是这些系统往往依赖需要大规模标注数据的目标检测模型，对计算设备的要求也比较高，企业应用和维护成本比较高，阻碍了工业智能化的发展。本文提出一种基于对比检测的巡检系统，通过事先设定的基准图与实时图进行比对，在两者存在光照和视角差异的情况下准确地检测出变化区域，再利用有限数据训练出的模型对变化区域进行识别，给出是否有异常以及异常类别的判断。系统实现了对目标区域的异常检测识别，其研究及开发的意义在于其初始化成本和运行维护成本非常低廉，既能赋能企业实现智能化生产，又能大大降低企业使用门槛，推动智能制造发展。

1.2 研究现状

1.2.1 变化检测研究现状

变化检测目前的应用包括视频监控^[18-20]、遥测感知^[21-23]、医学诊断以及治疗^[24-28]、民用基础设施^[29-30]、水下感知^[31-33]以及辅助驾驶系统^[34-35]。尽管应用非常多样，变化检测的研究中有许多共通的处理步骤和核心算法。

变化检测技术可分为基于像素和基于对象的方法两种，其中基于像素的方法又可分为基于差异和基于变化检测指数（Change Detection Index, CDI）的方法^[21]。基于像素的方法主要是对两幅或多幅遥感图像进行像素级别的比较，以找出其中的差异和变化。而基于对象的方法则是将图像中的像素按照一定规则组成对象，再对这些对象进行比较和分析，以实现变化检测。

目前，常用的基于像素的变化检测方法主要包括差值法、比率法、指数法和基于特征的方法等。其中，差值法是最简单的一种方法，其基本思想是对两幅图像进行减法运算，然后通过设置阈值来判断变化像素^[36]。比率法则是将两幅图像相除，得到的结果作为变化检测的依据，其优点是对比度变化不敏感，缺点是对灰度比较低的像素检测效果较差^[37]。指数法则是将两幅图像分别转换成指数图像，然后进行减法运算和阈值分割，得到变化图像。基于特征的方法则是将图像转换成特征空间，然后利用不同的特征比较方法进行变化检测^[21]。

基于对象的变化检测方法则是将像素按照一定规则组成对象，然后通过对象之间的比较来实现变化检测。对象可以是建筑物、道路、河流等，也可以是任意形状的区域，这些对象可以通过图像分割、目标识别和特征提取等方法得到^[38]。与基于像素的方法相比，基于对象的方法更加适用于复杂的遥感图像，能够更好地区分变化和不变化的区域，但是需要更多的先验知识和处理步骤，计算量也相对较大。

1.2.2 基于图像的智能巡检研究现状

因为本文基于对比检测的巡检系统尚属首创，国内外还没有相关具体研究。故这部分，只介绍基于目标检测等图像方法的智能巡检研究现状。

早期目标检测大都基于特征集成，特征是手工设计而成的，例如 Viola-Jones

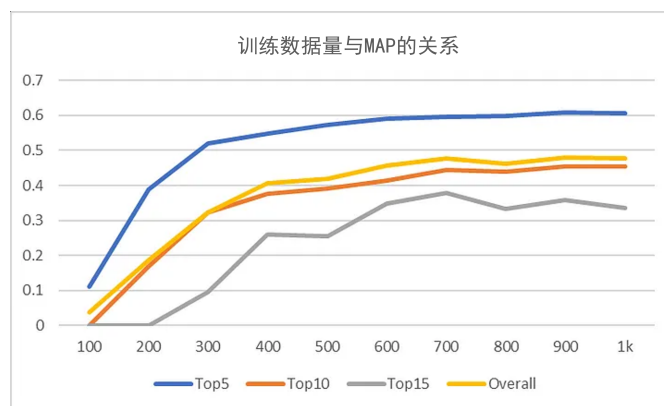


图 1-1 模型精度与每类别图片数的关系

检测器^[39]、HOG^[40]等。这些方法往往速度慢、错误率高以及泛化性差。随着卷积神经网络的引入，图像领域的深度学习开启了新篇章，于此同时基于深度学习的目标检测模型也井喷式发展。Girshick 等人提出了基于区域的卷积神经网络 R-CNN^[41]，将滑动窗口与支持向量机结合，奠定了区域建议方法（Region Proposal）与深度卷积网络结合的开山之作。不久，He 等人提出 SPP-Net^[42]，压缩了特征提取过程，大大提高了 R-CNN 的效率。Girshick 后来又在前作基础上提出 Fast R-CNN^[43]，利用感兴趣池化（Interest Region Pooling）对动态区域大小进行归一化，并且用 Softmax 函数替换支持向量机对候选区进行分类，同时使用了回归对候选框进行回归，最后以一种多任务的损失函数进行学习。Ren 等人，进一步提出 Faster R-CNN^[44]，并成为目标检测的代表模型。以上模型都包含区域建议和候选框位置回归两个分支，统称为两阶段模型。尽管两阶段模型检测精度高，但是效率低，为此 SSD^[45]和 YOLO^[46]等端到端模型被提出来。时至今日 YOLO 系列已开发了多个版本，加入了很多最先进的方法以及一些小技巧，检测速度和精度都有着很大的提高。

尽管目标检测任务上，现有模型取得了非常大的成功，但是不可忽视的是，一个好的检测模型往往需要大量的数据进行训练^[47]。例如对于 YOLOv5^[48]，有研究表明^[49]，在 Korean Sidewalk 数据集上，至少每种类别需要 300-500 张图片模型才具有较为优秀的性能，如图 1-1^[49]所示。由于每张图片中的目标实例往往不止一个，将如此数量的图片进行仔细标注将耗费大量的人力物力。

另一方面，模型的性能和模型的大小、运算效率往往需要一些取舍，实际应用中往往需要部署在专门的深度学习运算芯片或者昂贵的 GPU 设备上，这些成本给基于目标检测的巡检系统的推广应用带来了阻碍^[50]。并且，基于目标检测

的巡检系统在工程应用时，由于需要不停的对当前视频图像进行检测，如果算法模型部署在云端，将会占用大量的带宽，而且实时性也比较差，尽管大多数时候检测区域是没有异常的。而本文提出的系统将采用边云结合的方式，只在边缘端检测到可疑变化时才将当前图片交由云端研判，并且云端模型也相当的轻量化。

1.2.3 现有问题和不足

本文首次将变化检测方法引入智能巡检领域，然而现有的变化检测方法，由于过去没有这样的需求，现有的方法或多或少不能很好地适用到智能巡检上。过去，在遥感影像对比检测中方法，由于遥感影像数据的特殊性，例如受光照影像较小，目标区域近似是一个平面，视像头位姿已知等等，对于场景甚至室内场景粒度的变化检测就显得泛化性不足。而深度模型方向，又由于场景数据集的极度稀缺，目前的研究也才处于刚刚起步的阶段。

目前市面上一些已有的智能巡检系统，大都基于目标检测模型进行构建。一方面为了模型能够达到预期精度，需要采集大量的图片进行人工标注，费时费力，而且一旦切换到不同领域，同样的工作需要重复进行，对于产品和系统的交付非常不友好。另一方面，模型对计算设备有着很高的要求，对网络带宽需求也很大，需要企业配备相应的设施，应用成本居高不下。虽然，目前模型向着轻量化、高效化方向发展，甚至出现过依托专用边缘计算硬件的解决方案，但是依然不能很好满足成本低效果好的要求。

站在系统角度，尚未出现类似本文的工作。而本文提出的方法，结合系统设计，将算法和工程同步优化，可以很好的解决巡检任务。并且系统设计中采用模块化设计，对于后续算法调整，功能更新，系统部署升级都能灵活高效地完成。

1.3 本文工作

智能巡检对于基础设施维护以及油气、电力行业生产运输的安全运行有着巨大的应用场景，但由于现有方法和系统存在冷启动和部署维护成本高昂等原因尚不能广泛应用。本文提出一种基于变化对比检测的巡检系统，其核心是利用基准图与实时拍摄的影像，将有着视角和光照差异的两者进行对比，找出其中的差异部分，再利用开集识别模型判断差异部分是否是异常，如果是异常，给出

异常种类。从框架布局角度，本文提出与算法相适应的边云结合系统构建，即将变化对比检测部分部署在边缘设备上，将开集识别部分部署在云服务器端，由于大部分情况巡检的目标区域是不会发生明显变化的，此时边缘计算设备就可自主决策，当出现可疑变化时，边缘设备可以请求云端服务进行判别，这样可以极大减少云端负载和网络带宽，从而系统与算法一同优化达到巡检目标。本文的研究内容主要包含两部分：一是针对巡检任务，提出新的异常检测方法，分为变化检测和开集识别两部分，变化检测部分基于巡检场景对现有的变化检测方法进行调整优化，由于存在部分非异常变化和误检测，本文提出使用开集识别的方法作后续的异常识别，并且将对对比学习和原型学习结合应用到模型设计和训练中；二是从需求分析、系统设计、系统实现及系统测试四个环节完整实现了基于变化检测的巡检系统，并根据算法特点做系统优化，该系统具有完整的功能，能提供可靠的巡检服务，具体如下：

1. 本文首先提出对现有的变化检测方法的优化。该方法首先使用关键点检测方法 SuperPoint 提取基准图和实时拍摄影像的关键点，该方法对光照和视角变化更为鲁棒且高效的，然后使用快速最近邻估计算法对两张图片中提取的关键点进行匹配，接着利用向量场一致 (Vector Field Consensus, VFC) 算法剔除掉不满足一致性约束的错配点。对于已经匹配好的关键点，传统的方法会直接使用随机抽样一致 (RANDOM SAMPLE CONSENSUS, RANSAC) 算法计算与多数的匹配关键点对符合的单应性变换矩阵，作为两张图片的变换关系。然而传统方法只有在关键点位于同一个视平面时才会有比较好的效果，且容易出现不能完全配准的情况，给比对带来困难。本文对基准图片进行格点采样，将基准图片的格点映射到实时影像图片上，映射函数或单应性变换矩阵通过最近邻方法，由相邻的关键点投票得到。将映射位置的点的特征描述子与基准图对应格点的特征描述子进行比较，大于一定的阈值则认为变化点。随后将变化点和非变化点进行扩散，得到像素级别的变化区域，框选出矩形区域，作为算法的返回值。此方法相比于传统方法，更适合智能巡检场景，具有很好的鲁棒性和准确性。
2. 其次，本文提出使用开集识别的方法对变化区域进行异常识别。对于本文提出的基于变化检测的异常检测方法，由于前置步骤中变化检测得到的变化区域可能是非异常变化甚至是误检测（后文统称为非异常变化），而这

样的一个非异常变化的情况多种多样，对于一般的输出为类别数 +1 的识别算法（1 表示不在类别之中），很难收集到足够的训练数据帮助模型准确地识别不在目标类别中的图片，或者即使采集到足够多的非异常图片，也将面临类别及其不平衡的问题（异常图片往往数据量比较少）。本文结合对比学习和原型学习设计了 ConSim-CNN，不再需要在特征提取器上再训练一个带 Softmax 函数的线性层，直接计算原型向量与样本向量之间的相似度即可估计是否是异常类别。并且，本方法在开集识别范式下，仅在已知类别构成的数据集上进行训练，并能很好地同时检测未见类别样本和识别已知类别样本地具体类别。且模型的 backbone 参数量只有 45MB，可以灵活地选择在边缘设备或者云端部署。

3. 与基于变化对比检测的巡检方案相适应，本文提出边云结合的系统架构方案。过去基于目标检测的巡检方案，要么部署在边缘端但是需要算力强大却昂贵的边缘计算硬件，要么部署在云服务器端但是对带宽要求高，实际巡检时的延迟由于模型计算费时也较高。本文将上文提出的方法中的模型分别放置在边缘端和云服务器，在大多数情况下，巡检目标区域是没有变化的，这时边缘设备可以将计算检测在本地进行，当检测到变化时，再请求云端的识别服务。这样既分担了计算压力，又节省了带宽。除了算法主体部分，系统还加入了前处理和后处理部分，使得系统更加健壮和鲁棒。
4. 在系统实现方面，系统由前处理模块、异常检测模块、后处理模块、基准图管理模块组成。前处理模块实现对运动模糊识别、视像头遮挡等识别。异常检测模块实现核心算法，分为变化检测部分和开集识别部分，对比基准图和实时图进行变化区域检测以及异常类别识别。后处理模块平滑异常检测框，使得视频流的检测结果更为鲁棒。基准图管理模块，实现基准图的上载和更新。系统的实现过程围绕需求分析、系统设计、系统实现以系统测试四方面进行，涵盖了完整的软件开发测试过程，系统功能健全，具有良好的扩展性与内聚性。

1.4 章节安排

本文的研究围绕巡检任务展开，通过用变化对比检测和开集识别方法来代替原有的基于目标检测的巡检系统，以克服其存在的对数据量要求大以及对计算硬件要求高的问题。同时，本文还通过在系统实现层面对算法进行联合优化，使得算法与系统相互配合。系统遵循需求分析、设计、实现以及测试的过程，逐步完成系统的构建。全文共有七个章节，整体结构关系如图 1-2 所示。

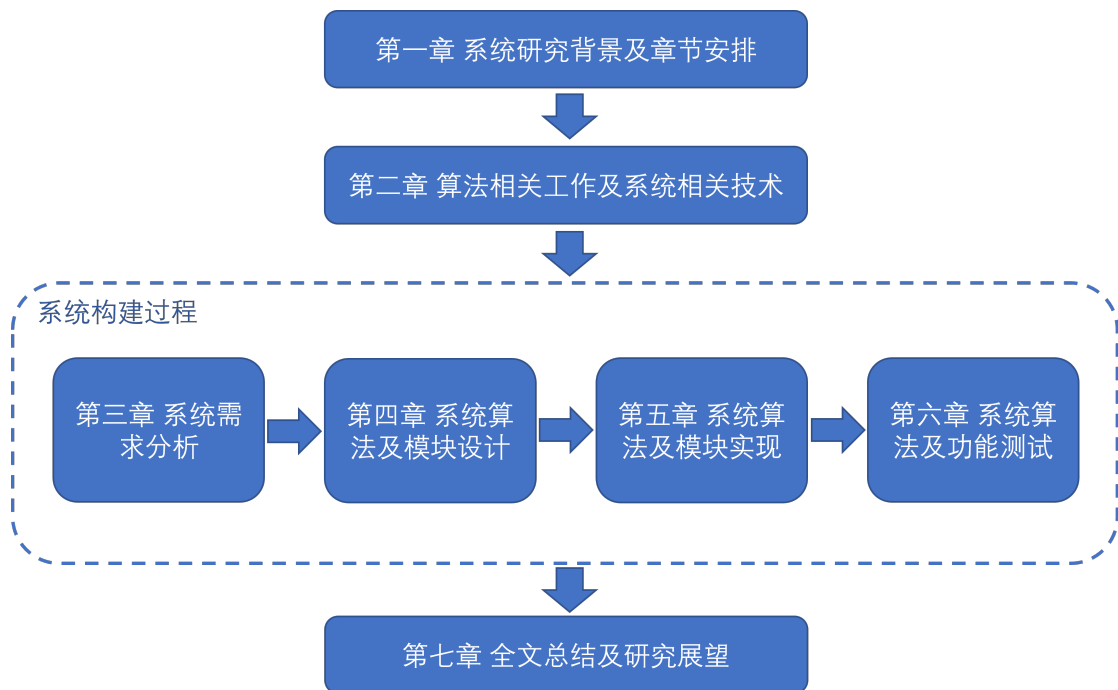


图 1-2 本文章节结构

第一章绪论。针对当前变化检测与智能巡检技术发展现状进行阐述与总结，介绍研究现状以及现有研究的不足，并对本文工作进行概括。

第二章相关工作。对所用的变化检测、开集识别技术和相关系统开发工具进行罗列和介绍，并说明这些技术的特点，以及在该系统中所体现的作用。

第三章需求分析。对智能巡检任务的市场需求、需要解决的问题等进行详尽的分析，准确定义系统整体及各模块所需的功能、性能以及可靠性。

第四章系统设计。首先提出基于变化检测的异常检测方案设计，分为变化检测设计和开集识别设计两部分，使得设计的检测方案适配巡检场景。其次提出和算法相辅相成的系统架构方案，具有低成本和高效率的特点，满足可扩展性、低耦合性、安全性要求。在前后端设计以及各模块设计中，分别对系统各组成模

块作进一步规划与设计。

第五章模块实现。系统各功能模块的算法及工程实现，系统层面包括前处理模块、异常检测模块、后处理模块、基准图管理模块，其中算法又分为变化检测算法和开集识别算法实现。

第六章系统测试。定义系统质量标准并对系统进行模块测试、集成测试、性能测试等，并展示系统运行效果。

第七章结语。对本文工作进行总结，分析系统及算法所存在的不足，并对未来研究提出展望。

第二章 相关工作

本章主要介绍系统设计实现过程中所需要的相关背景知识，以及具有代表性的工作内容。首先介绍算法部分系统所使用和借鉴的相关工作，分为变化检测部分和开集识别部分。变化检测部分按照算法流程相关的关键点检测、匹配点筛选、单应性估计的顺序介绍；开集识别部分首先介绍开集识别任务所面临的问题和相关研究，接着介绍与之相关并对本文模型设计有重要启发的关联领域研究。最后介绍系统实现所使用的相关框架技术。

2.1 变化检测部分相关算法

本系统的目的是为了检测识别设备及设施运行过程中发生的异常变化，与一般的异常检测问题有所不同。异常检测是指在常规的数据集合中，发现未知模式或者有别于数据集中其他数据的特征，数据的形式可以是图片、视频、音频、文本以及时间序列。异常检测依赖于一个参考数据集合，来理解和定义正常情形下数据的样貌。而本系统所研究的变化检测，是将单独的一对图片进行比较，只针对两张图片中的信息进行比对。两张图片一方面具有高度的相关性，即一张图片中出现的物体和背景绝大部分在另一张图片也存在；另一方面具有一定的差异性，这种差异性即包括不太关键的变化，例如由于视角变化带来位姿变化、取景边缘物体的出现消失，以及光照变化带来明暗不同，还包括关键性变化，即可视范围内物体的出现和消失。在变化检测过程中需要找到两张图片中不同位置的对应关系，这个过程中需要使用关键点检测、匹配点筛选、单应性估计等方法，下面逐一进行介绍。

2.1.1 关键点检测算法

基于特征的图像配准是一个低级处理过程，它是计算机视觉在遥感观测、辅助航空和天文观测应用的重要组成部分。图像配准首先在两幅或多幅图像中检测并描述局部特征区域，然后对这些局部特征区域进行匹配，再根据匹配好的特征区域来估计两幅图之间的变换模型，最后对齐两个图像。检测和描述局部特征区域这一步很关键，因为它决定了是否可以后续的处理。一个好的检测器旨在从图像中找到稳定且不同的局部区域。即使图像通过经过了视点、光照、比例、模糊和压缩等变换，这些检测器仍然可以找到或识别局部区域。目前常见的描述局部特征的方式包括点、线、曲线、边缘和轮廓五种类型。其中点，又名关键点或兴趣点，与其他特征相比更易于定位和描述，因此它是图像配准中应用最多的局部特征。传统特征检测方法是根据数学理论设计的，也被称为手工算子，常见的关键点类型有角点和斑点两种。角点通常是指图像中两条或多条边线的交叉部分。与角点不同，虽然斑点名字里含有点，但它其实主要描述的是一个区域，这个区域内的像素在颜色或者灰度上与其邻域的像素有明显区别，在纯色的背景中，里面单一的物体也可以称之为斑点。

SIFT

尺度不变特征变换（Scale-Invariant Feature Transform, SIFT）是一种斑点特征点检测器，由 Lowe 在 2004 年提出，用于解决特征匹配中的图像旋转、仿射变换、强度和视点变化等问题^[51]。SIFT 算法由以下四个步骤组成：第一步是使用多尺度高斯函数来构建图像的尺度空间表示。一副图像的尺度空间表示由可变高斯函数和输入图像进行卷积得到^[51]，见式 (2-1)。

$$L(x, y, \sigma) = G(x, y, \sigma) \cdot I(x, y) \quad (2-1)$$

其中 $I(x, y)$ 为输入的二维图像， x, y 表示图片上点的横纵坐标， $G(x, y, \sigma)$ 为可变高斯函数， σ 是高斯函数的平滑系数。

将相邻高斯函数平滑系数比例为 k 的高斯函数组和输入图像进行卷积后得到高斯金字塔的一组尺度表示，高斯金字塔由多组尺度表示构成，不同组的输入图像尺寸不同，第 i 组的尺寸是第 $i+1$ 组尺寸的两倍。高斯金字塔如图 2-1 左

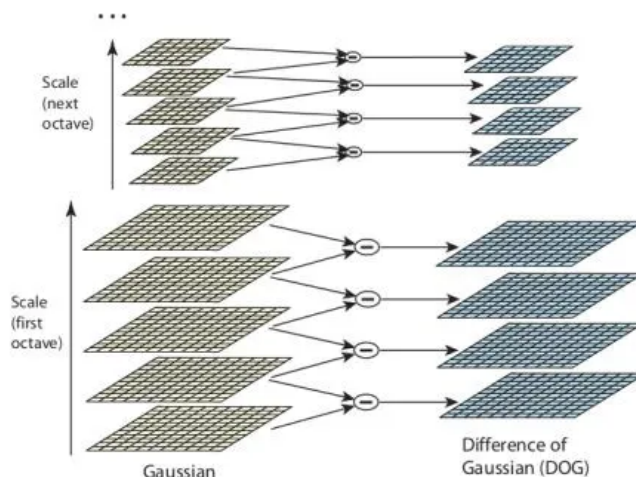


图 2-1 高斯金字塔

侧所示。

计算高斯差分金字塔 (Difference of Gaussian, DOG) 是检测关键点的重要步骤, DOG 的一层表示被定义为高斯差分函数与图像的卷积, 见式 (2-2)。我们不难发现高斯差分金字塔的一层表示也可以由高斯金字塔的每组尺度表示的相邻两层作差得到, 因此我们可以直接根据高斯金字塔来构建高斯差分金字塔。高斯差分金字塔如图 2-1 所示。

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) \cdot I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (2-2)$$

第二步需要在高斯差分金字塔中寻找局部极值点。每个像素需要与其相邻的 26 个像素进行比较, 26 个像素包括当前层的八个像素和上下两个层各九个像素, 如果该像素为局部极大值或者局部极小值, 那么将它标记为关键点, 如图 2-2 所示。

第三步需要计算关键点的方向 $\theta(x, y)$, 关键点方向定义见式 2-3。

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (2-3)$$

最后, 将关键点的局部区域划分为 16 个子区域, 每个子区域有 8 个方向直方图值, 根据这 16 个子区域的方向直方图值得到当前关键点的描述符。

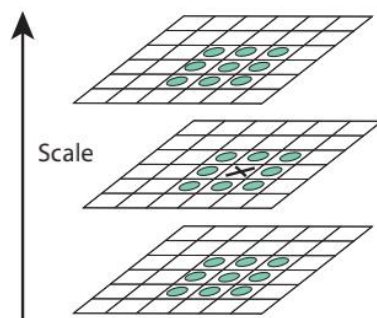


图 2-2 局部极值点

SuperPoint

SuperPoint 是一种自监督的全卷积神经网络，用于提取关键点及其描述符。其架构首先对名为 MagicPoint 的基础检测器进行预训练，该检测器使用合成数据集（包括矩形、星形和立方体等简单形状）提取关键点。使用初始基础检测器检测简单几何形状的关键点即使加入噪声也能表现良好 [5]。然而，在真实图像中，基础检测器错过了许多重要的关键点。为了更好地泛化，它与单应性适应相结合，从 MS-COCO 数据集中的每个图像提供更多的训练样本。使用这种方法，为每个图像生成了伪地面实况关键点。最终的网络具有共享的基于 VGG 的编码器和两个解码器。第一个解码器用于点检测，而第二个解码器用于点描述。两者之间有多个共享参数。与传统的特征点提取器（如 LIFT、ORB 或 SIFT）相比，SuperPoint 模型最近在几个研究项目中取得了卓越成果。

SuperPoint

SuperPoint^[52]是由 DeTone 等人在 2018 年提出来的一个自监督网络，用于关键点和描述子提取，对光照和季节变化具有鲁棒性。与常见的基于块的神经网络不同，SuperPoint 全卷积网络的输入是全尺寸图像，并且在一次前向传播中计算像素级别的关键点位置和关键点描述子。

SuperPoint 网络主要由编码器和解码器组成，首先是一个共享的编码器，用来处理和减少输入的全尺寸图像的维数。图像经过编码器处理之后输入到解码器，解码器部分又分为关键点解码器和描述子解码器，分别用于关键点检测和描述子求解。传统检测方法通常先检测关键点然后计算描述子，这种方式缺乏

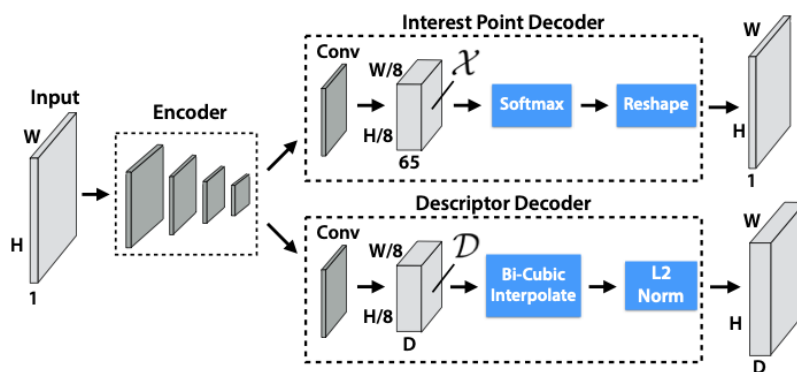


图 2-3 SuperPoint 模型结构图

两个任务之间的共享和表示能力，与传统方法不同的是 SuperPoint 的关键点解码器和描述子解码器之间共享大多数网络参数。模型结构见图 2-3。

SuperPoint 网络的编码器风格类似于 VGG^[53]，将输入图像的尺寸一步步减小，包括卷积层，池化层和非线性激活函数几部分。编码器中使用了三个最大池化层用来减小图像的尺寸，全尺寸图像经过编码器后尺寸变为原来的 1/8，即

$$H_c = \frac{H}{8}, W_c = \frac{W}{8} \quad (2-4)$$

如图 2-3 所示，关键点解码器的输入经过卷积后的尺寸为 $H_c \times W_c \times 65$ ，其中 65 个通道包括 64 个不重叠的局部像素网格区域通道和一个额外的没有关键点的通道。解码器的输出尺寸为 $H \times W$ ，每个像素的输出代表着这个像素成为关键点的概率。

描述子解码器的输入经过卷积后大小为 $H_c \times W_c \times D$ ，并输出一个 $H \times W \times D$ 的张量，其中 D 是描述子的维数。描述子解码器中使用了一个类似 UCN^[54] 的网络得到半稠密的描述子，半稠密的描述子是指多个像素共享一个描述子^[52]。然后对半稠密描述子进行双三次多项式插值得到所有像素的描述子，最后对描述子进行 L2 归一化到单位长度。

SuperPoint 的自监督主要体现在它对于真实场景图像的关键点是通过关键点检测器标注的，而非人工标注。这个关键点检测器是在一个由数百万个简单几何形状（如三角形、矩形等）组成的合成数据集上训练得到的，称为 MagicPoint。由于简单几何形状的关键点位置比较明显，没有太大的歧义，而现实世界中的图像纹理和图案都比较复杂，所以 MagicPoint 在虚拟场景中变现很好，在真实

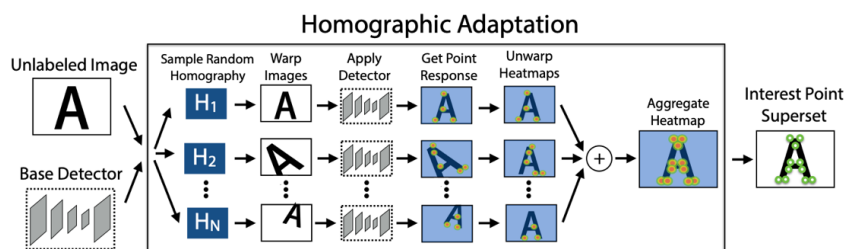


图 2-4 同态适应过程过程

场景图像上会遗漏很多潜在的关键点。同态适应 (Homographic Adaptation) 则是 MagicPoint 检测器实现自我监督训练的方式, 它通过对图像进行多次随机单应变换 (旋转和平移) 来帮助 MagicPoint 从不同视点、不同尺度来检测关键点, 从而提高 MagicPoint 在真实图像上的泛化能力, 过程如图 2-4 所示。

2.1.2 RANSAC

在通过求解最小二乘问题估计单应性矩阵时, 使用随机抽样一致性 (Random Sample Consensus, RANSAC) 算法来减少数据中异常值的影响是一种很好的做法^[55]。RANSAC 算法通常用于图像中的模型拟合, 它的基本思想是通过随机在数据集进行抽样得到数据子集, 然后使用数据子集来求解拟合问题, 多次重复这个过程然后选择表现最好的拟合。

RANSAC 有以下三条假设:

- (1) 数据由内点组成, 内点是指距离当前拟合小于阈值的点。
- (2) 离群点是不能适应该模型的数据, 离群点是指距离当前拟合小于阈值的点。
- (3) 内点和离群点之外的数据属于噪声。

在图 2-5 中展示了一个通过线性模型来拟合数据点的例子, 我们可以看到图中的蓝色数据点大致呈线性分布, 但是周围红色数据点偏离了蓝色数据点, 如果仍然将红色数据点和蓝色数据点一起拟合, 将会显著影响拟合的准确性。而 RANSAC 算法进行拟合时, 它会将红色数据识别为离群点, 仅保留蓝色数据点进行拟合。RANSAC 算法迭代地对原始数据点进行抽样得到数据点子集, 然后在数据点子集上进行模型拟合, 并在原始数据点上计算该拟合模型对应的内点数量, 如果大部分原始数据点都为内点, 则该认为拟合是一次成功拟合。根据这次拟合对数

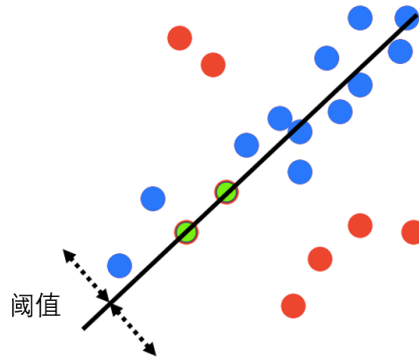


图 2-5 RANSAC 拟合

据点的分类结果，丢掉异常点，并使用所有内点对模型进行重新拟合。RANSAC 算法用于估计单应性矩阵的流程如下：

- (1) 从所有的匹配点中随机抽样得到匹配点子集
- (2) 计算当前匹配点子集的单应性变换矩阵。
- (3) 通过给定的阈值在所有匹配点中计算内点的数量。
- (4) 重复上述步骤多次并选择内点数量最多的匹配点子集。
- (5) 使用所有内点重新求解最小二乘问题得到最终的单应性矩阵。

2.1.3 VFC 匹配算法

向量场一致 (Vector Field Consensus, VFC) 算法是 Ma 等人在 2014 年提出的^[56]，可以在两组图片中的关键点之间建立鲁棒的关键点联系。算法首先从一组预先假定的匹配关键点出发，这些假定的关联关键点可能由很大一部分假的关联关键点和有限数量的真的关联关键点组成。算法从这组关联关键点中通过拟合插值向量场的方式解算出联系，即估算出一组满足无参数几何约束的一致性内点。算法通过混合模型来描述这个问题，引入了显式的潜在变量来识别和拒绝异常值，并对几何结构施加先验，即对向量场施加非参数光滑性约束。这导致了一个最大后验概率 (Maximum a Posteriori, MAP) 估计问题，存在许多局部最小值，算法求解过程可能会困在某个局部极值点上。为了解决这个问题，算法使用最大期望演算法 (Expectation Maximization, EM) 来估计先验方差，并通过初始化为一个大值，来快速地获得良好的估计。

令 $\mathbf{X} \in \mathbb{R}^{N \times P}$ 和 $\mathbf{Y} \in \mathbb{R}^{N \times D}$ 分别表示观测到的输入和输出数据，第 n 行分

别表示输入点 $x_n \in \mathcal{X}$ 和输出点 $y_n \in \mathcal{Y}$ ，其中 N 表示数据量大小， P 表示输入的维度， D 表示输出的维度， $\mathcal{X} \subset \mathbb{R}^P$ 是输入空间， $\mathcal{Y} \subset \mathbb{R}^D$ 是输出空间。对于一组观测输入输出对，目标是要拟合映射函数 f ，将样本点插值。这个问题一般而言是病态问题，因为有无穷多满足问题的解。为了得到有意义的解，一种可行的方法是引入一个正则项，这个正则项一般作用在向量形式的可再生希尔伯特空间 (Reproducing Kernel Hilbert Space)。特别的，可再生希尔伯特空间 \mathcal{H} 中，由核 $\Gamma : \mathbb{R}^P \times \mathbb{R}^P \rightarrow \mathbb{R}^D \times \mathbb{R}^D$ 定义的 Tikhonov 正则^[57]，其最小化如下的损失函数：

$$\epsilon(f) = \min_{f \in \mathcal{H}} \left\{ \sum_{n=1}^N \|y_n - f(x_n)\|^2 + \lambda \|f\|_{\mathcal{H}}^2 \right\} \quad (2-5)$$

$x_n \in \mathbb{R}^P$ 是输入， $y_n \in \mathbb{R}^D$ 是输出， f 是待求的映射函数。其中第一项是经验损失函数，强迫拟合数据；第二项是正则项，迫使向量场更加的平滑。 λ 是正则系数，权衡两项的比重。 $\|\cdot\|_{\mathcal{H}}$ 表示可再生希尔伯特空间 \mathcal{H} 中的范数。

根据表示定理^[58]，正则化损失函数 (2-5) 的解可以写成如下形式：

$$f(x) = \sum_{n=1}^N \Gamma(x, x_n) c_n, \quad c_n \in \mathcal{Y} \quad (2-6)$$

其中系数集合 $\{c_n : n \in \mathbb{N}\}$ 由以下线性系统决定：

$$(\tilde{\Gamma} + \lambda \mathbf{I}) \tilde{\mathbf{C}} = \tilde{\mathbf{Y}} \quad (2-7)$$

其中 Gram 矩阵 $\tilde{\Gamma} \in \mathbb{R}^{DN \times DN}$ 是一个 $N \times N$ 的块矩阵，第 (i, j) 块是 $\Gamma(x_i, x_j)$ ， \mathbf{I} 是单位矩阵， $\tilde{\mathbf{Y}} = (y_1^T, \dots, y_N^T)$ ， $\tilde{\mathbf{C}} = (c_1^T, \dots, c_N^T)$ 。

Tikhonov 正则的问题在于，它将所有的样本点视作内点，在有离群点时会有问题。VFC 算法假设，对于内点，它的噪声是高斯分布的，并且每个部分都是均值为 0，方差为 δ ；对于离群点，输出空间是一个有界区域 \mathbb{R}^D ，并且分布是参数为 $\frac{1}{a}$ 的均匀分布。算法将每个样本点与一个隐变量 $z_n \in \{0, 1\}$ 相联系，当 $z_n = 1$ 时表示服从高斯分布的内点，而当 $z_n = 0$ 时表示服从均匀分布的离群点。

则似然函数可以表示成如下形式：

$$\begin{aligned} \mathbf{p}(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) &= \prod_{n=1}^N \sum_{z_n} \mathbf{p}(y_n, z_n | x_n, \boldsymbol{\theta}) \\ &= \prod_{n=1}^N \left(\frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|y_n - f(x_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \right) \end{aligned} \quad (2-8)$$

其中 \mathbf{p} 表示概率， $\boldsymbol{\theta} = \{f, \sigma^2, \gamma\}$ 是一组未知的参数， γ 是说明隐变量边际分布的混合系数，即 $\forall z_n, \mathbf{p}(z_n = 1) = \gamma$ 。这里的均匀分布只在有界区域内非零（此处省略了归一化的指示函数）。

VFC 假设向量场 f 是一个后验概率已知的随机场的实例，这里的后验是用来约束 f 的选择，只有那些满足约束的函数才具有较大的概率。VFC 采用低速平滑模型，可以适用很大一部分运动场景，则 f 的后验可以写作：

$$\mathbf{p}(f) \propto e^{-\frac{\lambda}{2}\phi(f)} \quad (2-9)$$

其中 $\phi(f)$ 是一个平滑函数， λ 是一个正实数。

根据贝叶斯定理，可以求解 $\boldsymbol{\theta}$ 的最大似然估计，即 $\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta}} \mathbf{p}(\boldsymbol{\theta} | \mathbf{X}, \mathbf{Y}) = \arg \max_{\boldsymbol{\theta}} \mathbf{p}(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) \mathbf{p}(f)$ 。这等价于寻找最小能量，如式 2-10 所示：

$$E(\boldsymbol{\theta}) = -\ln \mathbf{p}(f) - \sum_{n=1}^N \ln \sum_{z_n} \mathbf{p}(y_n, z_n | x_n, \boldsymbol{\theta}) \quad (2-10)$$

向量场 f 可以直接从最优解 $\boldsymbol{\theta}^*$ 中求解，而隐变量集合 $\{z_n : n \in \mathbb{N}\}$ 决定哪些样本是离群点。

2.1.4 单应性矩阵

单应性矩阵描述了两张图片上的点的转换关系，它是一个 3×3 的矩阵，如式 2-11 所示，可以将齐次坐标 $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ 进行变换。齐次坐标坐标可以把平移变换，旋转变换，缩放变换统一成矩阵乘法^[59]，所以计算单应性矩阵时通常需要将非齐次坐标转为齐次坐标。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-11)$$

由于齐次坐标只依赖于尺度定义，即 $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} kx \\ ky \\ k1 \end{bmatrix}$ ，它们都表示同一个二维空间的点。所以单应性矩阵的参数可以进行缩放，这意味着它只有八个自由度，它通常是一个标准化的矩阵，可以通过以下两种方式来标准化：

(1) 将 h_{33} 置为 1，其他系数等比例缩放。

$$h_{33} = 1 \quad (2-12)$$

(2) 所有系数的平方和相加为 1。

$$h_{11}^2 + h_{12}^2 + h_{13}^2 + h_{21}^2 + h_{22}^2 + h_{23}^2 + h_{31}^2 + h_{32}^2 + h_{33}^2 = 1 \quad (2-13)$$

比如在图 2-6 中，在相应的视角中分别用坐标标注出来了相同的点。用相同颜色标注出来的点的坐标的转换关系可以表示为式 (2-14)：

$$\begin{bmatrix} wx' \\ wy' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-14)$$

通常不存在一个单应性矩阵 H 能够将一张图片上的所有点转换到另一视角的图片的对应点，但是可以通过最小二乘法估计这样一个单应性矩阵，使单应性矩阵能够描述大部分点的变换关系。给定 N 个匹配的关键点对 X_1 和 X_2 (X_1 和 X_2 都是行是齐次坐标的 $N \times 3$ 的矩阵)，单应性矩阵 H 可以通过求解最小二乘问题来估计：

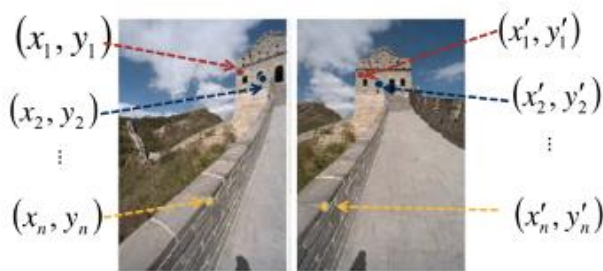


图 2-6 不同图像和视角中的相同点

$$X_2 H = X_1 \quad (2-15)$$

2.2 开集识别部分相关算法

2.2.1 开集识别

传统的识别算法基于一个基本假设：训练数据和测试数据的类别标签以及特征空间都是相同的，且属于同一个封闭集合中的不同实例。然而，在实际应用中，无法保证在使用过程中输入的样本与训练数据的类别标签完全一致，并且测试时的样本的潜在类别数是无穷的。因此，即使测试数据所代表的类别在训练阶段从未被学习过，传统的分类器仍会将其划分为训练集类别标签的一种。这种实际识别场景中的挑战被称为开集识别（Open Set Recognition, OSR）问题。在开集识别问题中，模型需要能够检测出未知类别的样本，同时保持对已知类别的高效分类能力。有研究者提出了一种具有拒绝选项的分类器^[60-62]，启发了 Scheirer 等人^[63]首次定义了 OSR 问题，并提出了一个基本框架进行训练和评估。近年来，尽管 OSR 比常见的闭集设置具有更多的实用价值，但却受到了惊人的忽视。目前，这个领域中的研究工作可以广泛地分为两类：判别模型和生成模型。

判别模型：在深度学习时代之前，就已经有一些利用传统机器学习方法的开集识别工作被提出。例如，Scheirer 等人^[64]和 Jain 等人^[65]考虑使用支持向量机结合极值模型，对未知类别的检测得分的分布进行建模。Rudd 等人提出了极值机器^[66]，利用基于极值理论的密度函数对类包含概率进行建模。Mendes Júnior

等人^[67]提出了一种开放集最近邻方法，通过计算最近邻之间距离的比率来确定任何具有低相似性的测试样本。Zhang 等人^[68]提出了一种基于稀疏表示的开集识别方法，同样使用极值理论来通过残差误差识别未知样本。需要注意的是，这些方法通常需要仔细设计特征工程才能很好地扩展。最近，Bendale 等人^[69]将深度神经网络引入了开集识别任务。他们证明了对 Softmax 概率的阈值设置不会产生稳健的开集识别模型。Openmax^[69]随后被提出，通过对激活向量的距离建模来检测未知类别。Shu 等人^[70]提出了一种基于 K-Sigmoid 激活函数的方法，通过消除网络外的离群点检测器，使得端到端训练成为可能。需要注意的是，这些工作中的 Sigmoid 函数没有紧凑消退的性质^[64]。该性质可以通过添加来自所有训练数据的无限远输入来激活，因此其开放空间风险没有被限定^[71]。ARPL^[72]引入了一种新颖的概念，称为互补点，以在特征空间中对每个已知类别建模其潜在的开放空间。PROSER^[73]在特征空间中为未知类别分配占位符，试图模仿开放集类别并预测未知数据的分布。PROSER 还使用特征混合来生成虚拟样本作为占位符。CVAECapOSR^[74]使用胶囊网络作为特征编码模型，以学习已知类别的紧凑特征表示。

生成式方法：与判别式模型不同，生成式方法使用 GANs^[75]、自编码器^[76]和流模型^[77]生成未知或已知样本，以帮助分类器学习已知和未知样本之间的决策边界。Ge 等人^[78]提出了 GOpenmax，它是 Openmax 的直接扩展，使用生成模型合成未知样本来训练网络。类似于 Zhang 等人^[68]提出的思想，Yoshihashi 等人^[71]提出了 CROSR 模型，它将监督学习的预测和无监督重构的潜在表示相结合，重新分配概率分布。Oza 等人^[79]提出了 C2AE 模型用于 OSR，利用类条件自编码器通过极值理论从重构误差中获得决策边界。Sun 等人^[76]提供了一个条件高斯分布学习的变分自编码器来检测未知类，并通过强制不同的潜在特征逼近不同的高斯模型来分类已知样本。Zhang 等人^[77]提出了一个由分类器和基于流的密度估计器组成的联合嵌入空间。GFROSR^[80]使用重建模型作为数据增强，迫使网络学习捕捉对象结构的特征。生成方法通过对数据分布进行建模为识别系统提供更多背景信息，但训练生成模型会显著增加识别系统的总训练成本。

除此之外开集识别还与其他一些领域的问题密切相关，如分布外（Out of Distribution, OOD）检测^[81]、异常检测^[82]和新颖性检测^[83]等。考虑到 AI 系统的安全性，Hendrycks 等人^[81]首次引入了 OOD 样本的检测。在这里，OOD 检测是

指检测不属于训练集但可能在测试中出现的样本^[81]。Hendrycks 等人^[81]证明了异常样本的最大 Softmax 概率比分布内样本低。Liang 等人^[84]提出了 ODIN 来使用温度缩放和对输入添加小扰动，以实现更有效的检测。Lee 等人^[85]利用生成模型从 OOD 样本中生成最有效的样本，并从中导出新的 OOD 评分。Hendrycks 等人^[86]提出了异常样本曝光，通过使用辅助数据集来教授网络更好的异常检测表示。OOD 检测类似于 OSR 中未知类别的拒绝，因为它们都研究了分布内（已知）和分布外（未知）样本的分离^[63]，^[81]，并且不需要针对已知类别的鉴别器功率。

2.2.2 模型鲁棒性：原型学习和对比学习

神经网络在泛化到从未见过的样本输入以及简单的图像变换（如旋转）方面的表现差强人意^[87-88]。这种欠鲁棒性表现在：在输入空间中具有不显著 l_p 距离的两个输入可能会产生显著不同的特征表示，从而导致误分类；另一方面，从感知角度与训练样本完全不同的输入（如随机噪声）可能会导致特征表示与训练样本几乎相同。原型学习（Prototype Learning）和对比学习（Contrastive Learning）方法是两个相互关联的研究方向，致力于提升深度神经网络的鲁棒性。

原型学习

原型是某个类别的平均或最佳范例，因此可以为整个实例类别提供简明的表示。最著名的原型学习方法是 K 最近邻（K-Nearest Neighbor, KNN）算法^[89]。Kuncheva 等人^[90]提出了学习向量量化（Learning Vector Quantization, LVQ）来节省存储空间并提高 KNN 的计算效率。在大多数先前的工作中，原型是通过优化自定义的目标函数来学习的^[91]。最近，一些方法还将原型学习与概率模型和神经网络结合起来进行分类任务。在神经网络框架下，原型是以每个类别的一个或多个潜在向量的形式进行可学习表示的。Bonilla 等人^[92]将输入实例表示为一个 K 维向量，将每个分量建模为概率混合，并最终应用概率模型通过最大化似然来参数化 K 原型模式。Wen 等人^[93]提出了一个中心损失来学习每个身份的深层特征中心，并使用中心来减少类内差异。Yang 等人^[94-95]提出了广义卷积原型学习（Generalized Convolutional Prototype Learning, GCPL），其中原型损失用

作正则化方法，以改善特征表示的类内紧密性。对于开放空间识别问题，原型有助于减少已知类别的类内距离，但它忽略了未知数据的潜在特征，导致在减少开放空间风险方面效果较弱。

对比学习

对比学习是最近受到广泛研究的表征学习领域。大多数对比学习方法都是自监督的^[96-99]，不依赖于任务特定的监督。自监督对比学习中的一个主要问题是如何在没有监督的情况下获取正负样本对。MOCO^[97]和 SimCLR^[98]使用单个训练示例的多个视图作为正样本对，不同的训练示例作为负样本对，但需要大量的负样本对才能达到良好的性能。BYOL^[88]和 SimSiam^[99]使用孪生网络结构和 stop-gradient 避免使用负样本对，因此它们可以处理较小批量的训练数据。在 ImageNet 分类任务中，最近的自监督对比学习方法已经取得了与监督学习可比的结果。

在损失函数设计方面，三元组损失^[100]使用每个样本的一个负样本对和一个正样本对来学习度量，不过容易对噪声敏感，从而导致过拟合。N-pair 损失^[101]减少了多个负样本对之间的相似度，并增加了正样本对之间的相似度。当前的对比学习方法如果存在质量差的样本对，会生成无信息的子梯度，从而导致模型收敛差。然而挖掘有信息的对（例如难负样本对）可能会有计算成本过高的问题。目前已经提出了几种方法来改善样本挖掘问题。Duan 等人^[102]训练一个额外得负样本生成器，与深度神经网络一同训练，为其增强容易的负样本，将其变为难负样本。LSL^[103]在重要性抽样方法中使用多个正负样本对来改善难负对的样本挖掘。然而，样本挖掘技术具有计算成本高的问题。相比之下，本文使用可学习的原型来表示正负对，从而避免了样本挖掘问题。

除了原本的自监督邻域，在有监督邻域，对比学习也有应用。Mao 等人^[104]在对抗训练设置中使用三元组损失作为交叉熵损失的辅助训练目标。类似 SimCLR^[98]、MoCo^[97]和 BYOL^[88]等对比学习框架，使用一个广义版本的 N-pair loss，即 InfoNCE，在增强样本的特征表示上创建一个正样本，并将批处理中的所有其他样本视为负样本。Li 等人^[105]的工作在 NCE 损失中添加了一个额外的损失项，该损失项将原始样本与相同聚类的原型和 r 个负原型进行比较。正负原型都是使用期望最大化聚类框架在每次训练迭代中确定的。相反，本文在监督设置

中为每个相应的类别分配原型。监督对比学习 (Supervised Contrastive Learning, SupCon)^[106] 基于有标签数据, 学习使得同类正样本对之间的相似度最大化, 不同类负样本对之间的差异最大化的表征。在分类准确率方面, SupCon 比自监督方法表现要好得多。在多个闭集分类任务中, SupCon 也优于普通 CNN 网络。然而, 由于不能处理未标记的数据, SupCon 没有像自监督方法一样受到太多的研究关注。

2.3 系统相关技术

2.3.1 视频推流

在将边缘设备采集到的实时影像传输回客户端或中央服务器时, 需要一种快速便捷的流传输协议。常用的流媒体协议主要有 HTTP 渐进下载和基于 RT-SP/RTP 的实时流媒体协议两类。在流式传输的实现方案中, 一般采用 HTTP/TCP 来传输控制信息, 而用 RTP/UDP 来传输实时多媒体数据。

实时传输控制协 (Real-time Transport Protocol, RTP) 是用于 Internet 上针对多媒体数据流的一种传输协议。RTP 由两个紧密联系的部分组成: RTP—传送具有实时属性的数据; RTP 控制协议—监控服务质量并传送正在进行的会话参与者的相关信息。RTP 协议是建立在 UDP 协议上的。RTP 协议详细说明了在互联网上传递音频和视频的标准数据包格式。RTP 协议常用于流媒体系统 (配合 RTCP 协议)、视频会议和视频电话系统 (配合 H.263 或 SIP)。RTP 本身并没有提供按时发送机制或其他服务质量 (QoS) 保证, 它依赖于底层服务去实现这一过程。RTP 并不保证传送或防止无序传送, 也不确定底层网络的可靠性。RTP 实行有序传送, RTP 中的序列号允许接收方重组发送方的包序列, 同时序列号也能用于决定适当的包位置, 例如: 在视频解码中, 就不需要顺序解码^[107]。

实时传输控制协议 (Real-time Transport Control Protocol, RTCP) 是实时传输协议 (RTP) 的一个姐妹协议。RTCP 为 RTP 媒体流提供信道外控制。RTCP 定期在流多媒体会话参加者之间传输控制数据。RTCP 的主要功能是为 RTP 所提供的服务质量提供反馈。RTCP 收集相关媒体连接的统计信息, 例如: 传输字节数, 传输分组数, 丢失分组数, 时延抖动, 单向和双向网络延迟等等。网络应用

程序可以利用 RTCP 所提供的信息试图提高服务质量，比如限制信息流量或改用压缩比较小的编解码器。RTCP 本身不提供数据加密或身份认证，其伴生协议安全实时传输控制协议则可用于此类用途^[107]。

实时流协议 RTSP 定义了一对多应用程序如何有效通过 IP 网络传送多媒体数据。RTSP 在体系结构上位于 RTP 和 RTCP 之上，它使用 TCP 或 RTP 完成数据传输。HTTP 与 RTSP 相比，HTTP 传送 HTML，而 RTP 传送的是多媒体数据。HTTP 请求由客户机发出，服务器做出响应；RTSP 可以是双向的，即客户机和服务器都可以发出请求。RTSP 与 RTP 最大的区别在于：RTSP 是一种双向实时数据传输协议，它允许客户端向服务器端发送请求，如回放、快进、倒退等操作。当然 RTSP 可基于 RTP 来传送数据，还可以选择 TCP、UDP、组播 UDP 等通道来发送数据，具有很好的扩展性。它是一种类似于 HTTP 协议的网络应用协议。

imageZMQ 是一组 Python 类对象，使用 PyZMQ 的消息传递机制将 OpenCV 类型的图片从一个主机传输到另一个主机。imageZMQ 实现了 ZMQ 中的两种消息传递方式：1. REQ/REP：每一个采集端向中央收集端发送一张图片并等待回复（REPLY）。采集端只有当接收到回复时才发送新的图片。在 REQ/REP 模式中，每个图片发送者必须等待回执信息才能继续执行。对于发送者来说，这是一种阻塞的模式。2. PUB/SUB：每个采集端发送一张图片，但并不预设会从中央收集端收到回复^[108]。

2.3.2 Flask 框架

Flask 是一种 Python 编写的轻量级 Web 框架，其具有核心简单、开发便捷、可扩展性强等特点，因此被广泛用于深度学习框架的算法执行系统中。与其他大型框架如 Django 相比，Flask 不需要使用特定的工具或库，也不对数据库、模板引擎等进行限制，因此开发者可以更加灵活地编写应用程序和扩展，并自由选择配置。Flask 本身不包含数据库抽象层、表单验证工具或其他通用功能组件，在使用时可根据需要自行添加第三方组件或扩展，例如用户帐户管理和身份验证、静态代码检查、表单处理、会话管理、缓存支持等。Flask 主要依赖于五个组件，其中 Werkzeug 提供了路由和 Web 服务器网关接口等功能；Jinja 提供了用于渲染页面的模板语言，并且能够在沙盒环境中执行不受信任的模板，具有模板集成功能、即时编译、异常调试方便、语法可扩展等特点，同时在 HTML 模板中

编写的代码会编译为 Python 代码；MarkupSafe 用于在渲染模板时将不受信任的文本生成转义字符以避免注入式攻击，保证系统运行的安全性；ItsDangerous 进行数据签名以保护会话 Cookie；Click 提供了命令行集成功能^[109]。

尽管 Flask 具有内置的 Web 服务器，但其需放置于真正的 Web 服务器之后，并通过 WSGI 协议与 Flask 进行通信。因其在生产环境中的适用性受到限制，因此一般采用 gunicorn 作为 Flask 的扩展。gunicorn 同样基于 WSGI 的 HTTP 服务器，使用 pre-fork 模式，可在多个线程上运行多个服务器。为扩展关系型数据库，可使用 SQLAlchemy，并使用相应数据库驱动程序，例如 psycopg2 用于 PostgreSQL，MySQLdb 用于 MySQL 等。此外，还可以添加 Alembic 扩展作为轻量级数据库迁移管理工具。对于非关系型数据库，可使用 PyMongo、dynamo 等扩展作为 MongoDB 及 DynamoDB 等非关系型数据库的接口。此外，Flask 还可以与 Redis 结合使用 celery 创建分布式异步任务队列，并使用 Flask-WTF 进行表单验证，使用 Flask-limiter 限制 Web 请求速率等。与 Django 相比，Flask 更适用于快速构建中小型项目，具有易于配置、灵活部署等特点。而 Django 则更适用于大型项目，易于管理和维护，但其学习曲线上升较为缓慢。因此，在不需要使用 Django 的大多数功能时，选择 Flask 框架更为合适。使用 Python 和 Flask 的优点在于其能够进行快速原型设计和动态语义，并且还可以将新模块添加到 Flask 中以扩展其核心功能^[110]。

2.3.3 MongoDB

MongoDB 是一种备受欢迎的面向文档的非关系型数据库，其在应用程序开发和运行中与 AWS、Azure 以及其他众多数据源协作广泛使用。相较于关系型数据库，MongoDB 具有更好的可扩展性，可以存储和查询大量数据，并提供了诸如二级索引、区间查询、排序、分析框架和地理空间索引等功能。通过适当的索引和处理，可以更好地执行查询，在实时分析和优化数据处理方面使用临时查询，而强大的复制功能则有助于提高数据的可用性和灵活性。数据分片机制使得大数据块可以拆分，从而实现了分布式和更快速的查询执行过程。传统的关系型数据库在处理有限数量的数据和需要复杂查询时表现良好，但在海量数据的持久性储存方面则不如非关系型数据库。与关系型数据库不同，非关系型数据库不使用关系数据库管理系统，其数据不像关系型数据库那样存储在表中，

也没有固定的模式。而是采用丰富的数据模型，同时使用标识键，以便根据分配的键找到数据^[111]。

MongoDB 是一种非关系型数据库系统，其数据模型采用文档型结构，大部分数据存储在内存在中，能够通过直接从内存中读取数据来提高查询效率，并减少磁盘读写的时间消耗。相比于一般的关系型数据库系统，MongoDB 的速度有着显著的提升。此外，MongoDB 的环境配置简便，快速搭建和运行，具有动态的示意图架构，支持非结构性数据和存储。在处理大量数据集时，MongoDB 采取分片机制，将数据集分为多部分，并分布式存储到多台服务器中。每个分片环境中的操作由一个轻量级进程 mongos 负责，该进程会根据分片键直接查询到正确的分片，从而实现数据库的良好水平扩展性和更好的负载均衡。此外，MongoDB 对即席查询进行了优化，同时支持字段查询、区间查询、正则表达式查询等功能，使得用户拥有更灵活的选择。

此外，MongoDB 具有内置的大规模负载均衡功能，使得开发人员不必添加额外的负载均衡服务器。MongoDB 实现了数据的并发读写，通过使用并发控制和读写锁来处理多个并发读写请求，以确保数据的一致性。此外，MongoDB 还通过主从复制机制在多个服务器上同步存储数据，以实现读写分离、高可用性和安全性，并在发生故障时允许数据的恢复。数据恢复是通过一个主节点与多个副本节点服务器来实现的，用户对数据的交互是由主节点完成的。此外，MongoDB 还实现了副本集机制，当主节点失效后，副本节点接管主节点成为新的主节点，进一步提高了数据库的可用性。

第三章 需求分析

随着工业智能化的不断发展，智能制造不断探索深入人心。目前，中国的许多厂房、仓库和基础设施的监测维护主要还是依靠人工巡检的作业方式。例如电力行业的变电站、物流行业的仓储等。在高压、超高压以及恶劣气象条件下，人工作业方式不仅人身危险性大，而且对设施的安全运行也带来一定隐患。智能巡检系统利用巡检机器人沿着既定的导航路线进行设备自动巡检，利用其自带的可见光和热成像摄像头实现设施监测维护的无人值守或者少人值守。智能巡检系统具有以下优势：减少人员疏忽、漏检等带来的设备损失，提高设施的运行质量；减少系统的人员投入，降低人员成本；减少不利环境对人的慢性健康危害。

当前的巡检机器人对于设备异常的定位识别的自动化，一般基于通用或者在特定数据集上的特质化检测深度模型，它的特点是需要大量的训练数据集和人工标注，成本巨大，如果出现了新的设备或者新的需求，需要进行新的数据标注和重新训练；另外要想深度模型具有好的效果，往往需要参数量较大的模型，对于机器人的计算设备成本有不小的挑战。而本文基于变化对比检测方法的巡检系统则可以在克服上述局限性的同时，又能达到造价低、准确性高等要求。

本文所设计的系统要实现两方面的需求：从设施或者设备异常检测的角度出发的算法设计要求，以及系统设计层面出发的系统需求。本章首先探讨该系统的可行性，然后分别从功能性需求和非功能性需求两方面对实现系统所需要满足的条件进行分析和归纳。

3.1 可行性分析

本文所实现智能巡检系统应具备，在有限的资源下检测变化区域并识别缺陷种类的能力，在变化检测算法上，本文对传统算法做了适当改进，并在开发系统前进行了实验验证，证明了其有效性。在识别算法上，在前人工作的基础上进

行了调整并在公开数据集和国网数据上进行了测试，因此系统算法功能可达到预期效果。

在经济可行性方面，本文所实现的巡检系统以桌面客户端的形式实现，算法服务由于其轻量性特点既可以部署在云端服务又可以部署在边缘计算设备上，并基于 github 进行版本控制，能很方便地部署；系统采用模块化设计，其高内聚低耦合性使得后期对系统维护也不需要作过多改动，因此系统地运营成本核维护成本都会很低。系统的数据吞吐中占比较大的一部分是图片的上传，以及实时视频流，因为视频流经过高效的压缩处理，因此一般的民用带宽就能满足要求。

在技术可行性方面，系统基于 Python 语言开发，使用 flask 作为后端框架，gRPC 作为算法服务框架，桌面前端基于 PyQt5 框架，数据库使用 MongoDB，所有技术均广泛使用且形成成熟的技术生态，前后端技术框架均包含完整的文档、官方或者第三方开发的模块或插件，这使得开发过程中不会因遇到关键的技术瓶颈或者技术难点而导致开发成本急剧增加。

综上所述，该系统具有良好的算法功能可行性、经济可行性、技术可行性。下面进一步分析系统的功能性需求与非功能性需求。

3.2 功能需求分析

随着工业生产的不断发展，工业智能化的要求越来越高，为了对各种工业基础设施进行检测维护，智能巡检系统变得非常需要。传统的巡检系统，基于目标检测进行工作，需要大量的人工标注数据，模型本身参数量也非常可观，计算设备要求高。前期需要投入巨量的成本，后期的维护升级工作也繁琐低效。这就需要一种新的智能巡检系统，克服上述缺点，同时满足一定的巡检精度要求。从整体上，本文所设计的系统需要实现的主要功能有以下几个方面：

- 实现异常检测功能。不同于传统的基于目标检测的方法，本文的异常检测算法分为两步：变化检测和开集识别。变化检测部分将当前拍摄的图像与设定的正常图片进行比较，找出当前图片的变化区域。开集识别部分对变化区域进行识别。由于可能存在误检测的情况，识别模块需要首先判断是否是异常，然后在是异常的基础上判断异常类别。本文将变化检测模块置于边缘设备，而将开集识别模块置于服务器端。

- 实现视觉感知功能。视觉感知除了需要稳定实时地传输摄像头拍摄的图片，还需要进行运动模糊检测和摄像头遮挡检测，提高后续巡检模块中算法的稳健性。还需要对摄像头事先做一定的标定，减少摄像头畸变带来的误差。
- 实现基准图管理功能。基准图是实现异常检测的重要组成部分，需要管理必要的信息以备算法和客户端调用。
- 实现异常日志功能。包含对异常信息的记录和报警，以及对异常的处理记录。

基于上述功能需求，系统采用边云结合的形式，边缘设备负责实时监视影像向客户端的传输，以及对于异常检测的前半部分变换检测，如果出现了变化区域（这种情形一般比较少，所以可以将大部分计算和流量保留在边缘端）则进一步调用云端的异常检测的后半部分开集识别。云端除了负责算法，还有基准图和异常日志的管理功能。客户端负责视频和信息的展示以及对边云的功能操作。

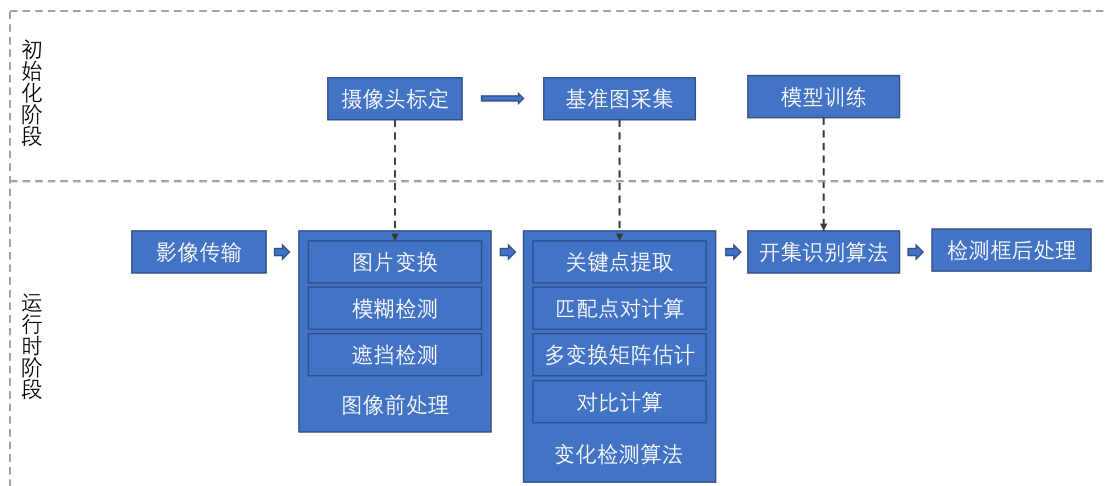


图 3-1 巡检系统流程图

图 3-1展示了使用系统进行巡检所需包含的过程。巡检包含两个阶段，初始化阶段和运行时阶段：

- 初始化阶段，用户首先需要对摄像头进行标定，避免摄像头畸变对后续算法的影响。然后对于目标区域，需要用摄像头拍摄一张光线充足条件下的图片，作为基准图，并上传到服务器中。除此之外还需要对模型在事先采集和标注的少量训练集上进行训练。
- 运行时阶段，用户客户端会不断的接收边缘设备端摄像头端传来的图片（图片经过畸变消除变换），用户这时可以选择开始检测操作。检测操作执

行时，边缘设备端会首先进行运动检测和摄像头遮挡检测，如果不通过，则检测算法不工作并提示摄像头工作异常，如果全部通过，这时边缘端会请求服务器获得最新的基准图，随后结合基准图和当前通过检测的实时图，利用变化检测模块提取出实时图和基准图重合区域有变化的部分，若无变化区域则直接返回无异常结果。如果出现变化区域（多个变化区域，需要进行合并），系统将变化区域裁切出来，调用云端的远程调用接口，使用开集识别模块确定是否是缺陷，如果是缺陷，则同时返回缺陷类别。视频流中的变化区域会进入后处理模块，可以稳定视频流中的变化区域，不至于出现过大波动。客户端接收之后，将变化区域和缺陷类别标志在检测界面上。

下面进一步对系统的算法方面的功能与其他辅助性系统功能作详细的需求分析。

3.2.1 算法功能需求

在算法功能方面，为了实现智能巡检，系统需要实现异常检测功能，即需要对于指定区域实时检测巡视，如果出现了目标异常，可以识别并进行日志登记和告警。系统最核心的部分是变化检测和开集识别。变化检测部分，需要比较基准图和识别图，对于两者重合区域，如果发生了变化，需要能检测出变化区域。由于实时图和基准图之间存在光照、视角等差异，算法需要能克服这些差异准确找出实时图和基准图之间对应区域的差异，这些差异只考察两张图片的相交区域，不相交区域不需要考虑。开集识别部分，因为变化检测算法可能存在误检测的情况，算法除了能在存在异常的情况下分类出异常类别，还需要识别是否是异常，而非异常可能是各种背景或者前景物体，算法需要对训练集中没有出现过的类别具有很好的鉴别能力。

由于算法在准确率和召回率上需要一定的折中，为了满足客户对于两个指标的不同需求，需要一种机制可以让用户调节算法的超参数，从而获得特质化的服务体验。需要面向一般使用者和专业使用者提供不同的接口和界面来满足其调节算法表现的需求。对于一般使用者，提供高召回、默认、高准确率三种不同的选项，让其选择模型事先设定的参数组合。而对于专业使用者，提供文件配

置选项，让其可以更为精细地调整模型地效果。

由于算法运行在实时监控视频上，对于算法的运算效率提出一定要求。在边缘端的变化检测算法，需要采用不需要大量张量计算的算法，并且计算复杂度和内存复杂度要控制在一定程度。在云端的开集识别算法，需要保证一定的吞吐量和低延迟，需要对模型的算子进行优化和蒸馏或者采用轻量化的模型。对于一些核心的算法，需要将其使用 C++ 进行实现，然后在用 Pybind11 包装，既提高效率，又和整体开发框架相适应。

为了进一步提高智能巡检的效果，系统应该包含对视频进行前处理和后处理的算法。前处理的算法需要对各种不需要进行检测的情形进行剔除，如摄像头因为运动而模糊、摄像头被意外遮挡不能正常工作，对于这些情形，将他们从算法的请求队列中排除，既可以提高算法的质量，又能减轻服务器和网络的负担，提升用户体验。后处理的算法需要稳定检测框，避免出现检测框剧烈变化的情况，检测框剧烈变化往往是由于算法的稳定性和泛化性存在问题，后处理可以大大减轻这类问题。可以将前处理、后处理等算法功能以多种接口形式对外提供，使得用户端有更自由的形式选择。

3.2.2 系统功能需求

系统采用云端、用户端和边缘设备端的空间布局方式。既可以最大程度地利用边缘设备的计算能力，又能灵活地使用云端的计算服务。其中用户端完成交互操作，接收边缘设备的实时监控影像，以及操作边云协同工作，完成巡检任务。边缘设备包括前处理模块和变化检测模块，云端包括开集识别模块、后处理模块、基准图管理模块、异常记录管理模块。

系统应将所实现算法以服务形式对外开发，包括前处理服务、变化检测服务、开集识别服务、后处理服务，可实现巡检实时影像的异常检测功能，以及摄像头非正常工作报警功能。包含对实时影像图片的传输、展示，基准图的获取、更新，以及是否发现异常的标志提醒。当发现异常时，需要用矩形框框出目标区域并标志上异常种类。用户可以依据自身需求，对算法的超参数进行调节，指定平滑系数、阈值等。系统应该能时暂停或者继续视频监控、异常检测，上传或者更新基准图。功能操作之间应该满足一定的约束关系：只有视频监控运行时，才可以进行基准图的上传或者更新；存在基准图时，才可以进行开始检测算法；当

检测算法暂停时，才可以将监视视频暂停。

系统需要实现对基准图的管理。基准图除了可以统一上传，也可以在实时监视的时候上传。基准图需要载明的信息包括：图片文件 id、图片分辨率、图片拍摄地点、图片拍摄时间、图片拍摄设备等。图片拍摄地点和图片拍摄时间，用于巡检到设定位置时，通过当前位置和时间，查询基准图数据库，选择相同时刻和季节、相同位置的可选图片，从而获得当前应该使用的基准图。

系统需要实现对异常记录的管理。巡检系统的最终目的是为了对于设施或者设备进行日常维护和检修。所以需要对于检测到的异常进行记载、报警以及处理。系统需要对于预定位置的待检测设施或设备记录是否有异常，如果发现了异常需要通过客户端警示、邮件或者短信提醒等方式进行告警，对于已有的异常经过人工干预处理后，需要进行已处理标志、或者进行待处理标志，方便进行维护和维修任务安排。

系统需要实现低带宽、低延迟、高质量的视频推流功能。视频推流是巡检系统的前提条件，推流需要满足低带宽的要求，否则不能满足多个摄像头的监视，网络会出现干扰卡顿。推流需要满足低延迟，如果延迟太高，会影像用户的体验和异常检测的时效性。推流还需要有较高的质量，需要尽可能地保留图像的细节信息，已保证后续算法处理的准确性。除此之外，推流要保证可以同时支持多个用户端连接而不影响其效率。

除了主要算法功能外，系统还应具备一定的运营功能，包括用户身份机制、数据管理、设备管理机制。用户身份认知机制由于实现用户的注册、登入、登出功能，实现用户权限管理功能、实现账号管理机制、通过维护用户级别以实现用户访问控制，增强系统安全性。只有经过身份验证的用户才能进行系统资源、服务的访问和使用，以及个人信息的管理。对于管理员权限的用户可以对用户、系统资源和系统数据进行管理和修改。

3.3 非功能需求分析

系统的非功能性需求包括系统的运行性能、安全性、健壮性、可用性、可拓展性、可维护性等方面。

(1) 系统性能。在系统性能方面，该系统需要尽可能地实现实时性要求，保证

算法的计算效率。系统采用云端、用户端、边缘设备的空间布局方式，尽可能地利用边缘设备本身的算力，降低系统部署的成本。在实际场景中，监视画面往往长时间地保持不变，边缘设备通过自身算力运行变化检测模块，只有当检测到出现变化区域（疑似异常）时，才进一步调用开集识别模块，这样设计，可以极大减少不必要的网络传输。

- (2) 可扩展性。统需保证其可扩展性，当会话数量或计算数量增加时能根据需要增加相应计算节点数以满足服务需求。系统结构清晰，使用命名规范，并将参数配置进行集中化。系统应采取模块化设计，每个系统事件应分别由某一特定模块完成，每个模块应具有良好的输入输出数据定义以及功能定义，所定义的数据对模块运行应当是必要的。设计时应尽可能抽象化实现细节，以最小化其他模块的调整对这一模块的影响。系统中不同算法分别使用不同的模块，应做到灵活加入新的算法作为新的模块，为系统提供更高的扩展性。
- (3) 健壮性。系统的健壮性通常由异常处理机制来完成，需要使得系统能在包含无效输入的场景或严苛运行环境条件下仍然能正常运行，包括系统资源不足、用户高并发场景、服务器宕机等情况，当请求出错或发生异常时系统应具备异常处理能力，模块单元在检测到无法处理的异常时应抛出异常，由上层模块处理，使得进程不发生崩溃。对于无效的用户请求也需根据不同的错误情况分别返回不同的错误码。最后，还需对系统进行大量测试，测试样例需包含各种容易出错的情形，从而使得系统能应对最大限度地不同场景并能正常运行不崩溃，以保证系统的健壮性。
- (4) 可维护性。要增加系统的可维护性，在设计模块功能时需要保证各模块的高内聚性和低耦合性。模块的内聚性表示模块内部实现的功能的单一性，模块耦合性表示模块之间联系的密切性。一个易于维护的系统需要通过松耦合来减少模块或类之间的互连性，使得在修改一个模块时最大限度减小对其他模块的修改。同时还需保证系统各模块的高内聚性，一个设计良好的模块应该有一定的用途，所有元素都应与单个任务相关联，使得模块具有高内聚性。在系统模块化设计中，每个模块具备逻辑上单一与独立的功能，并具备良好定义的接口。设计时需保持每个模块的大小适中，易于维护。所有模块构成有向无环图，形成底层模块相互不具有依赖关系，上层

模块依赖于下层模块的关系。应尽可能减少其下游模块调用的深度，对顶层模块保持高扇出，即尽可能多的直接下属模块，下游模块保持低扇出以及高扇入。每个模块功能应可预测，应具有单一入口和出口，每个模块的作用域仅限于该模块中。

3.4 本章小结

本章首先从功能可行性、经济可行性和技术可行性三个角度对系统可行性进行分析，接着分析了系统各组成部分的功能性需求、非功能性需求对系统的需求进行综合性的分析，从多角度描述了系统整体的设计目标及注意点，为系统的后续设计与开发工作奠定基础。

第四章 系统设计

文章分别从系统总体设计，算法设计，前后端设计、各模块设计进行系统性的设计。系统总体设计部分介绍系统框架及各组成部分的功能。算法设计分为变化检测算法设计、开放集识别算法设计以及前后处理算法设计。在前后端设计中分别介绍前端与后端的设计思路与数据流程，在模块设计中根据系统需求分析所提出的各项要求，分别对异常检测模块、视频流模块、基准图管理模块、日志管理模块进行详细设计。

4.1 系统总体设计

本系统除了客户端和服务端，还引入了边缘设备作为额外的计算资源。边缘设备是指在距离数据源更近的位置进行数据处理的设备，这种设备可以让数据更快地响应请求并处理，从而提供更低的延迟和更高的数据隐私性。相比传统的云计算模式，边缘设备的优势在于其能够更加灵活地处理数据，实现更加实时的响应。边缘设备通常包括智能手机、物联网设备、车载设备、家庭网络设备等等，这些设备可以与传感器、摄像头、其他设备等通信，从而实现数据的实时处理。边缘计算可以帮助解决传统云计算架构面临的问题，如高延迟、网络拥塞和数据隐私性问题。边缘设备的优势主要体现在以下几个方面：首先，边缘设备具有更短的响应时间，可以实现更快的数据处理和响应，提高应用程序的实时性和用户体验；其次，边缘设备可以更好地保护数据隐私，由于数据在离用户更近的位置处理，可以避免数据在网络传输过程中被窃取或被黑客攻击；最后，边缘设备可以更加灵活地进行数据处理，可以根据需要对数据进行预处理和过滤，从而减少网络流量和减轻云端压力。

本系统提出的基于对比变化检测的巡检方法，可以很好地融合到边缘计算设备中。一方面，本方法可以分解为两个阶段：变化检测阶段和开放集识别阶

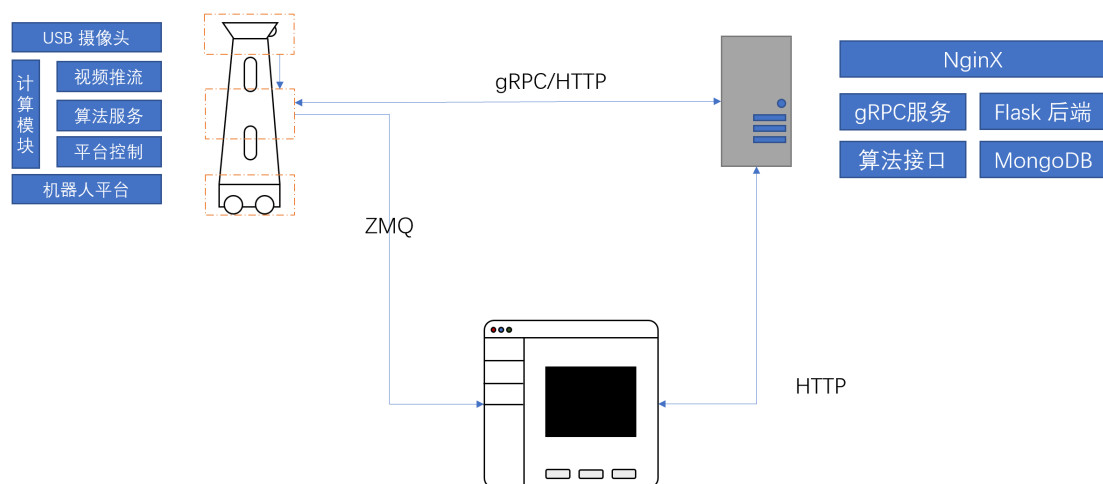


图 4-1 系统总体框架

段，变化检测算法部分可以部署在边缘设备，而开放集识别部署在云端服务器上。变化检测阶段只有在检测到变化区域时才会进行后续开放集识别阶段，在绝大多数情况下，巡讲场景是基本不会发生变化的，故而计算和数据流很大部分停留在边缘计算设备进行；另一方面，本方法高效且模型轻量，可以很好地部署在边缘设备上，对云端计算能力要求也不高。具体的，整个系统地拓扑结构如图 4-1。

在拓扑图右侧，根据系统要求，系统采用 B/S 架构来组织，并遵循前后端分离原则，由于系统地算法部分涉及深度学习框架，如果使用其他语言作为后端语言将需要对 Python 再做一层封装，而且也不方便模型地一次加载和多次调用，因此后端框架采用轻量级 Python 框架 Flask，对外提供算法接口、基准图管理接口等；前端基于 PyQt5 框架进行开发，负责对接后端服务地功能使用。系统在启动后端时将一次性地将模型加载，每次收到算法 API 调用请求时对已经预加载地模型进行调用，以减少模型每次启动地开销。前端为用户桌面应用，基于 Python 框架处理用户请求，并使用 PyQt 框架进行桌面应用设计与开发。

后端对外开放的 API 接口有两种：基于 gPRC 的服务与基于 HTTP 的 RestFul API，其中基于 gPRC 的服务使用 Google 的 Protobuf 协议，拥有高速序列化的能力，可应用于桌面端应用程序、移动端应用中基于视频数据流或图片数据的服务，以及一些带宽受限场景等。基于 HTTP 的 RestFul API 则很常用，可应用于网页前端、桌面端应用程序等更广泛的场景。后端的模块包括变化检测模块（开集识别部分）、基准图管理模块、异常日志管理模块等。算法模块负责实现变化

检测方法中的开集识别部分。基准图管理模块实现基准图的上传与更新功能，以及获取功能。异常日志模块负责对异常的记录和处理，方便相关人员跟踪维护巡检区域的设备设施。

在拓补图的左侧，系统主要包含三个部分：视频推流、算法服务、(机器人)平台控制。机器人平台在本系统中并不是必须的，也可以采用轨道摄像头或者固定摄像头的形式，故本文不做深入讨论，它的作用是路线规划并按照指定位置停靠，并告知系统其他部分，所处的位置以向后端获取对应的记住图影像。视频推流部分，没有采用 IP 摄像头所使用的 RTSP 等协议，由于协议需要复杂编码，协议还包括许多控制信息，经过测试延迟很大，质量也欠佳。故本系统采用在 PyZMQ 上开发的 ImageZMQ，最轻量化的传输视频流。算法部分主要包含前处理模块、异常检测模块（变化检测部分）、后处理模块。前处理模块会对摄像头进行畸变矫正、摄像头模糊检测和遮挡检测，当出现不利于算法运行的质量较差的实时影像时，算法会拒绝运行。异常检测模块会首先检测是否发生了变化，如果没有，则直接进行后处理，否则会进行 gPRC 调用，调用服务器的开集识别算法，于此同时后处理同步进行（已经有检测框），当两者同时完成，返回图片进行传输。

4.2 算法设计

目前多数基于目标检测的异常检测方法都需要有大量的训练数据对模型进行训练，使用的目标检测模型往往也参数巨大，对计算资源的要求也比较高，进一步阻碍了智能巡检的应用。另一方面，传统的图像算法，计算简单，可解释性强，但是泛化性不足，准确率不高。本文将两者进行结合，将整个算法拆分成两部分，变化检测和开集识别。变化检测部分，根据比较事先准备的基准图和实时采集的图片的不同，检测出图片的变化区域并进行适当裁切。开集识别部分，将变化区域的图片进行识别，区分具体的异常类别或者判断是非异常。下面对这两个算法作进一步的说明。

4.2.1 变化检测算法

本方法来自于遥感影像的变化检测方法，即先配准、再图像变换、最后计算差值。这一方法在遥感影像中工作的很好，但是在巡检领域，经常出现配准不到位的情况，甚至变换矩阵容易出现奇异值。究其原因，一是由于提取关键点阶段，遥感影像数据的光照变化和视角变化几乎不存在，传统关键点检测 ORB 等可以很好地胜任，而在巡检场景，这样的前提往往是不成立的；二是由于计算变换矩阵时，关键点需要满足极线约束^[56]，RANSAC 等方法才能工作的比较好，在遥感这种长焦距的影像中是很容易满足的，但是巡检场景使用的摄像头常常是短焦距的。三是计算差异时，由于图像变换时需要重采样和插值，会出现轻微的差异，给直接利用像素信息计算差异时带来困难。

针对上述问题，本文设计了如下的变化检测方法

算法 1: ChangeDetection

Input: 基准图 I_{base} , 实时图序列 $[I_0, I_1, \dots, I_i, \dots]$

Output: 变化区域序列 $[T_0, T_1, \dots, T_i, \dots]$, T_i 形如

$[rect_i^0, rect_i^1, \dots, rect_i^j, \dots]$, $rect_i^j$ 表示矩形区域，包含左上和右下角点坐标

```

1 对  $I_{base}$  做灰度化处理;
2 for  $I_i \in [I_0, I_1, \dots, I_i, \dots]$  do
3    $I_i$  做灰度化处理;
4    $kp_{base}, des_{base} \leftarrow \text{SuperPoint}(I_{base});$ 
5    $kp_i, des_i \leftarrow \text{SuperPoint}(I_i);$ 
6    $M_{init} \leftarrow \text{FlannBasedMatcher}(des_{base}, des_i);$ 
7    $M_{good} \leftarrow \text{VFC}(kp_{base}, kp_i, M_{init});$ 
8    $T_i \leftarrow \text{KNN-Difference}(I_{base}, I_i, M_{good});$ 
9 end

```

其中，算法1所示第4行和第5行分别计算基准图和实时图的关键点坐标 kp ，和相应特征描述子 des ，接着使用 FlannBasedMatch 计算匹配关键点。最后两步利用粗匹配关键点进行过滤，以及变化区域计算。下面分别对后两步 VFC 和 KNN-Difference 进行详细的说明。

VFC 算法

VFC 算法假设算法1中第6行所得到的匹配关键点集合 $M_{init} = \{(x_n, x'_n) : n \in \mathbb{N}\}$ 满足其中只有少部分是正确的匹配关键点，大部分是错配的匹配关键点。

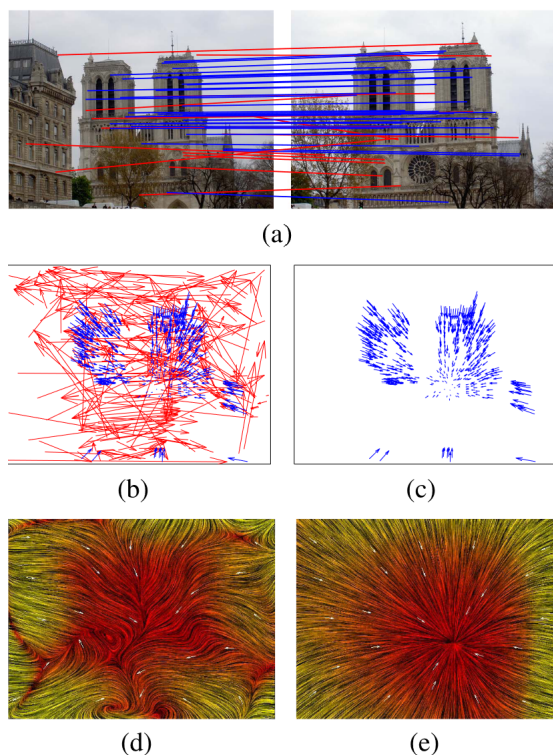


图 4-2 向量流插值

在这样的假设下（在巡检场景下是很常见的），传统方法 RANSAC 基于极线约束的配准方案效果就会大打折扣。VFC 基于向量流插值（Vector Flow Interpolation）提出向量场一致性约束。令位移矢量 $y_n = x_n - x'_n$ ，基准图与实时图之间的映射关系 $x \mapsto f$ 满足 $\forall x_n, y_n = f(x_n)$ 。映射 f 中引入了一个隐变量，它显示了对于集合 $S = \{(x_n, y_n) : n \in \mathbb{N}\}$ 哪些元素是正确匹配的点，哪些是错误匹配的点，以及一个反映几何关系的先验作为向量场的无参数平滑约束。形象的解释如图 4-2^[56]所示，图 4-2 (a) 是两张图片，图上用线连接的点是预估匹配关键点，关键点的特征使用 SIFT 算法计算得到。图 4-2 (b) 显示出这些预估的匹配关键点以他们的匹配关系画位移矢量的结果，其中蓝色箭头表示内点（正确匹配的关键点及其位移矢量），红色箭头表示外点（错误匹配的关键点及其位移矢量），图 4-2 (c) 单独将内点画出。如果不将外点筛选出来，整个图片的位移矢量场就会如图 4-2 (d) 所示，而图 4-2 (f) 是图像上的点都匹配正确的情况下的位移矢量图，整幅图像更加平滑和有序。所以反过来讲，如果加上平滑约束，可以很好地排除掉错配点。

如式 2-10 所示，由于存在隐变量，要求解此式需要使用 EM 算法。首先需

要计算后验概率，如式 4-1

$$\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = & -\frac{1}{2\sigma^2} \sum_{n=1}^N \mathbf{p}(z_n = 1 | x_n, y_n, \boldsymbol{\theta}^{\text{old}}) \|y_n - f(x_n)\|^2 \\
& -\frac{D}{2} \ln \sigma^2 \sum_{n=1}^N \mathbf{p}(z_n = 1 | x_n, y_n, \boldsymbol{\theta}^{\text{old}}) \\
& + \ln(1 - \gamma) \sum_{n=1}^N \mathbf{p}(z_n = 0 | x_n, y_n, \boldsymbol{\theta}^{\text{old}}) \\
& + \ln \gamma \sum_{n=1}^N \mathbf{p}(z_n = 1 | x_n, y_n, \boldsymbol{\theta}^{\text{old}}) - \frac{\lambda}{2} \phi(f).
\end{aligned} \tag{4-1}$$

接着需要基于 $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ ，不断基于旧的参数来估计分布，从而更新新的参数值，具体步骤如下：

E 步骤：使用现有的参数值 $\boldsymbol{\theta}^{\text{old}}$ 计算隐变量的后验分布。令 $\mathbf{P} = \text{diag}(p_1, \dots, p_N)$ ，是一个对角矩阵，其中 $p_n = \mathbf{p}(z_n = 1 | \mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\text{old}})$, $n \in \{1, \dots, N\}$ 可以由贝叶斯公式得到：

$$p_n = \frac{\gamma e^{-\frac{\|y_n - f(x_n)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|y_n - f(x_n)\|^2}{2\sigma^2}} + (1 - \gamma) \frac{(2\pi\sigma^2)^{D/2}}{a}} \tag{4-2}$$

后验概率 p_n 是一种软间隔，决定了第 n 个样本和目前估计的向量场 f 有多一致。

M 步骤：使用公式 $\boldsymbol{\theta}^{\text{new}} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{new}})$ 计算修正后的参数 $\boldsymbol{\theta}^{\text{new}}$ 。考虑到 \mathbf{P} 是一个对角阵，将 $Q(\boldsymbol{\theta})$ 分别关于 σ^2 和 γ 求偏微分，并置结果为 0，可以得到：

$$\sigma^2 = \frac{(\tilde{\mathbf{Y}} - \tilde{\mathbf{V}})^T \tilde{\mathbf{P}} (\tilde{\mathbf{Y}} - \tilde{\mathbf{V}})}{D \cdot \text{tr}(\mathbf{P})} \tag{4-3}$$

$$\gamma = \text{tr}(\mathbf{P})/N \tag{4-4}$$

其中 $\tilde{\mathbf{V}} = (f(x_1)^T, \dots, f(x_N)^T)^T$ ， $\tilde{\mathbf{P}} = \mathbf{P} \otimes \mathbf{I}_{D \times D}$ ， \otimes 表示克罗内克乘积， $\text{tr}(\cdot)$ 表示求迹。

对于参数 f ，我们提取出式 4-1 中只关于 f 的项，可以得到正则化后的损

失函数的形式：

$$\mathcal{E}(f) = \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|y_n - f(x_n)\|^2 + \frac{\lambda}{2} \phi(f) \quad (4-5)$$

这是 Tikhonov 正则的一种，第一项是加权经验误差。则求使得 \mathcal{Q} 最大化的 f 变成最小化正则化损失函数 (4-5) 的问题。

对于上述问题的求解，可以将 f 约束在一个可再生核希尔伯特空间 \mathcal{H} 中，这个空间由矩阵值核 $\Gamma : \mathbb{R}^P \times \mathbb{R}^P \rightarrow \mathbb{R}^{D \times D}$ 定义。对于平滑函数 $\phi(f)$ ，使用二范数，即 $\phi(f) = \|f\|_{\mathcal{H}}^2$ 。则由式 2-6 可以给出问题解，其中系数集合由以下线性系统给出：

$$\left(\tilde{\Gamma} + \lambda \sigma^2 \tilde{\mathbf{P}}^{-1} \right) \tilde{\mathbf{C}} = \tilde{\mathbf{Y}} \quad (4-6)$$

一旦完成了 EM 算法的步骤，得到了位移向量场 f ，就可以估计样本配对集合 M_i 中哪些是内点，哪些是外点。可以使用如下两种方法：

1. 预定义阈值 τ ，得到内点集合 \mathcal{I}

$$\mathcal{I} = \{n : p_n > \tau, n \in \mathbb{N}\} \quad (4-7)$$

2. 通过得到的向量场 f ，检验每个点是否和该向量场一致。

基于效率考虑，本文使用第一种方案。全部步骤，如算法2所示。

KNN-Difference 算法

得到两张图片中较为准确的关键点之间的匹配关系后，我们寻求将这种部分的对应关系推广到整张图片上来。一个常见的做法是利用已有的匹配关系，结合图片之间的视角几何假设，估计出单应性矩阵。这样的单应性矩阵虽然反应了两个平面之间的映射关系，但忽略了现实世界中存在深度差异以及场景中存在物体移动等复杂情况。传统的单应性矩阵要求，两个图像对应的区域必须在同一平面或者两个图像之间只有旋转视角变化而没有平移视角变化^[112]。

考虑到现实世界特别是巡检任务下场景的复杂性，本文提出一种对图像中的不同区域采取不同单应性矩阵的多单应性矩阵方法。在背景介绍中提到通过

算法 2: VFC

Input: 匹配关键点坐标集合 $M = \{(\mathbf{x}_n, \mathbf{x}'_n) : n \in \mathbb{N}\}$ 。核 Γ , 正则系数 λ , 内点阈值 τ (具有默认值)

Output: 内点集合 I

- 1 计算样本集合 $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{N}\}$ 初始化 $\gamma, \mathbf{V} = \mathbf{0}_{2N \times 1}, \mathbf{P} = \mathbf{I}_{N \times N}$;
- 2 设置 a 为图片长宽的最小值;
- 3 根据式 4-3 初始化 σ^2 ;
- 4 根据 Γ 的定义构造 Gram 矩阵 $\tilde{\Gamma}$;
- 5 **repeat**
- 6 E 步骤:
- 7 根据式 4-2 更新 $\mathbf{P} = \text{diag}(p_1, \dots, p_N)$;
- 8 M 步骤:
- 9 求解线性系统式 4-6 跟新 $\tilde{\mathbf{C}}$;
- 10 根据式 2-6 更新 $\tilde{\mathbf{V}}$;
- 11 根据式 4-3 和式 4-4 更新 σ^2 和 γ ;
- 12 **until** Q 收敛;
- 13 内点集合 I 由式 4-7 给出。

RANSAC 算法计算单应性矩阵时, 会区分“内点”和“离群点”, 算法最后会在内点的比例达到一定阈值时, 剔除掉离群点, 再在内点上拟合作为最后的输出单应性矩阵结果。本文利用 RANSAC 算法得到离群点和内点, 但是考虑当离群点的比例大于 $threshold_{outlier}$ 时, 认为整个特征点匹配几何需要多于一个单应性矩阵进行准确表达, 需要继续计算单应性矩阵, 但是计算下一个单应性矩阵时, 需要从上一轮的匹配点中将上一轮的内点去除, 即计算第 $i + 1$ 个单应性矩阵时使用的点是初始匹配点减去第 1 次到第 i 次的所有内点。计算完成后, 我们会得到多个内点几何以及多个单应性矩阵, 每个内点集合对应一个单应性矩阵。多单应性矩阵算法如下

对于经过过滤的匹配关键点, 我们已经将他们分割为不同的集合, 每个点集合覆盖图片中的一片区域, 这个区域内的非关键点也可以用对应的对应关系即单应性矩阵找到对应到另一张图片中的点, 这样的区域可以用如下的方法计算:

理想情况下, 每个内点集合对应图片中的一个物体, 因此也就可以确定一个矩阵区域且各个内点集合确定的区域互不重合。这样我们就可以根据内点集合确定多个区域, 每个矩形区域内的关键点使用对应的单应性矩阵进行转换。图像中距离较近的区域单应性矩阵变换关系相似, 所以对于没有被内点集合确定的矩形区域包含的关键点则使用距离该点最近的矩阵区域的单应性矩阵进行

算法 3: ComputeHomographyMatrixs**Input:** 关键点匹配集合 $M = \{(x_n, x'_n) : n \in \mathbb{N}\}$, 阈值 threshold**Output:** 单应性矩阵集合 $\mathcal{H} = \{H_1, \dots\}$, 内点集合

$$\mathcal{I} = \{I_1, \dots\}, I_i \subseteq M, I_i \cap I_j = \emptyset, \forall i, j \in \mathbb{N}$$

```

1 初始化  $\mathbb{H}, \mathbb{I}$ , total =  $\|M\|$ , count = total;
2 repeat
3   计算内点和相应单应性矩阵:  $I, H = \text{RANSAC}(M)$ ;
4   count = count -  $\|I\|$ ;
5    $\mathcal{H} = \mathcal{H} \cup \{H\}$ ;
6    $\mathcal{I} = \mathcal{I} \cup \{I\}$ ;
7    $M = M - I$ ;
8 until count/total > threshold;
```

算法 4: 关键点覆盖区域计算**Input:** 内点集合 $\mathcal{I} = \{I_1, \dots\}, I_i \subseteq M, I_i \cap I_j = \emptyset, \forall i, j \in \mathbb{N}$ **Output:** 关键点覆盖区域 $\mathcal{R} = \{rect_1, \dots\}$

```

1 初始化  $\mathcal{R} = \emptyset$ ;
2 for  $I_i \in \mathcal{I}$  do
3    $X_{min} \leftarrow \text{width}$ ;
4    $X_{max} \leftarrow 0$ ;
5    $Y_{min} \leftarrow \text{height}$ ;
6    $Y_{max} \leftarrow 0$ ;
7   for  $(x_k, y_k) \in I_i$  do
8      $X_{min} \leftarrow \min(X_{min}, x_k)$ ;
9      $X_{max} \leftarrow \min(X_{max}, x_k)$ ;
10     $Y_{min} \leftarrow \min(Y_{min}, y_k)$ ;
11     $Y_{max} \leftarrow \min(Y_{max}, y_k)$ ;
12  end
13  $\mathcal{R} \leftarrow \mathcal{R} \cup (X_{min}, Y_{min}, X_{max}, Y_{max})$ ;
14 end
```

转换。理想情况下由内点集合确定的区域如图 4-3 所示。

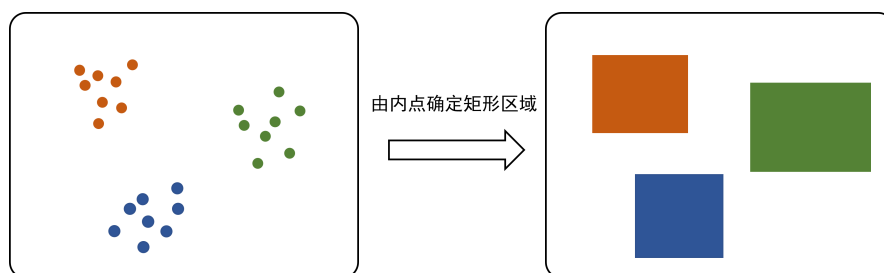


图 4-3 理想情况下内点集合确定的区域

然而实际情况下各个内点集合并不是聚集的，而是分散的，因此内点集合确定的矩形区域也是相互交叉的，实际情况下由内点集合确定的矩形区域如图 4-4 所示。容易发现对于交叉区域内的像素点无法确定应该使用哪个矩形区域对应的单应性矩阵。特别是存在一种极端情况，某个内点集合确定的区域特别大，覆盖了其他所有区域，如果所有像素点都使用这个比较大的区域对应的单应性矩阵得到效果一定不如使用各个小区域的单应性矩阵的效果。

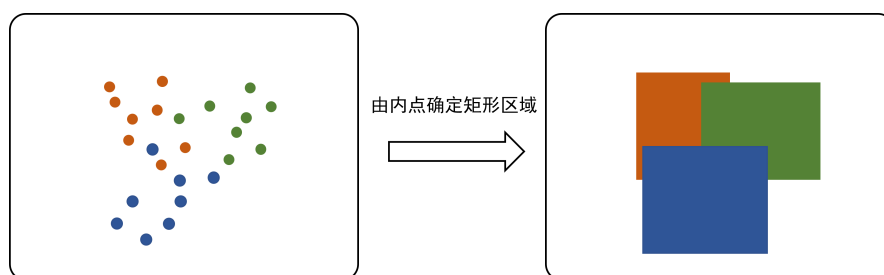


图 4-4 实际情况下多单应性矩阵

所以这种简单的做法不能满足实际场景中的需求。故而，我们转向为每个像素点决定应该使用何种单应性变换。由于图片的总的像素是 $O(NM)$ 数量级， N ， M 分别表示图片的宽和高。对每个点寻找它对应的变换矩阵复杂度太高，所以我们使用固定间隔做格点采样，然后对这样规则选取的代表点决定其应使用何种单应性变换矩阵。由于图片具有局部平滑性，可以设想内点以均匀的速度扩散，影响周围的像素点。这其实是一种带权最近邻估计，每个内点以距离的函数作为权重对代表点最后使用哪种单应性矩阵作投票。距离的函数的形式参考热传导方程：

$$\frac{\partial T(x, t)}{\partial t} = D \frac{\partial^2 T(x, t)}{\partial x^2} \quad (4-8)$$

这是一维情况下的热传导方程， T 代表温度，在初始条件 $T(x, t = 0) = \delta(x - x_0)$

下， δ 是狄拉克函数，它的解的形式如下：

$$T(x, t) = \left(\frac{1}{4\pi Dt} \right)^{1/2} e^{-\frac{(x-x_0)^2}{4Dt}} \quad (4-9)$$

故 KNN 算法中的权重方程使用 $e^{-(x-x_0)^2}$ 的形式。

这样我们对于格点形式的代表点得到了其相应的单应性变换矩阵，我们对每个代表点计算其变换后的位置，对于变换前后位置，分别计算其特征描述子，如果描述子差异太大，则认为对应位置发生了变化，即对应格点记作一个变化点。对所有的代表点对比完后，我们得到变化点集合。由于格点是以固定间隔采样的，我们讲变化点膨胀对应间隔大小的一半，这样如果变化点是格点上相邻的点，则格点之间的区域都纳入到变化区域了，否则是不同的变化区域。得到变化区域后，如果变化区域的面积大于某个事先设定的阈值，则认为这个变化区域是有效的。最后再利用联通分量算法，求出每个有效变化区域对应的最小矩形框。矩形框由于会存在重叠或者非常接近的情况，这里我们认为可以将两个区域合并，一起作为变化区域矩形框。

4.2.2 开集识别算法设计

有研究表明，尽管 CNN 网络有很高的准确率，但是 CNN 很容易被对抗样本欺骗，说明 CNN 对于模式识别的鲁棒性还不够。Yang 等人^[113]认为 CNN 鲁棒性不足主要是由于 Softmax 层，CNN 完全是一个判别模型，并且基于闭空间的假设（类别数量固定），它只学习了整个特征空间的一部分，所以对于未见类的样本，CNN 仍然将其预测到这部分空间中的某个区域中，并依旧以比较高的置信度认为其是已知类。这说明 CNN 的拒绝能力比较弱；另一方面，从表示学习的角度，CNN 学到的表示是线性可分的，如图 4-5^[113]所示，在这样的表示下，类间距离甚至比类内距离还要小，这极大的降低了 CNN 在现实复杂环境中的鲁棒性。

为了提高 CNN 的鲁棒性，Yang 等人^[113]提出一个新的学习框架—卷积原型学习（Convolutional Prototype Learning, CPL）。在 CPL 的底层，和传统的 CNN 一样，卷积层被用来提取判别式的特征，而在 CPL 的顶层，则使用多个原型来表示不同的类别。分类任务则变成在特征空间中基于欧氏距离计算最近邻原型。

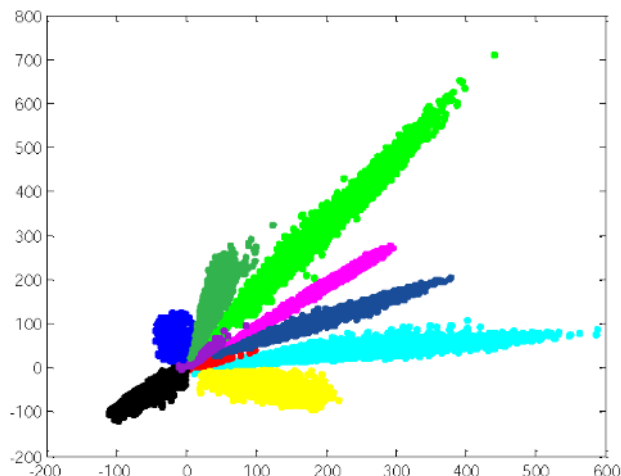


图 4-5 CNN 在 MNIST 数据集上学到的特征表示

具体的，模型使用 CNN 作为特征提取器 $f(x; \theta)$ ， x 和 θ 分别表示原始输入和 CNN 的参数。不同于传统 CNN 使用 Softmax 层在学到的特征上进行线性分类，原型学习方法保存并学习一些特征原型，并使用原型匹配来做分类。原型记作 m_{ij} ，其中 $i \in \{1, 2, \dots, C\}$ 表示类别的索引，而 $j \in \{1, 2, \dots, K\}$ 表示每个类别的原型的索引。这里我们假定每个类被有相同数目的 K 个原型，这个假设可以在实际应用中很容易放宽。

前向预测阶段，对于给定输入模式 x ，我们首先得到 CNN 特征提取器计算的抽象特征，接着我们将其与所有的原型进行比较，通过最近原型将其归入所属类别，即：

$$x \in \text{class} \arg \max_{i=1}^C g_i(x) \quad (4-10)$$

其中 $g_i(x)$ 是类别 i 的判别函数：

$$g_i(x) = - \min_{j=1}^K \|f(x; \theta) - m_{ij}\|_2^2 \quad (4-11)$$

训练阶段，要同时训练两部分参数，一个是 CNN 特征提取器，记为 θ ，另一个是每个类别的原型，记为 $M = \{m_{ij} | i = 1, \dots, C; j = 1, \dots, K\}$ 。需要将两者端到端地一起训练，这样才能得到好地分类性能。对于损失含函数，既要对于 θ 和 M 是可微分的，同时损失函数要尽可能地接近分类准确率。下面介绍损失函

数，首先样本的误分类类别 y 的程度由以下公式给出：

$$\mu_y(x) = -g_y(x) + \left[\frac{1}{C-1} \sum_{j \neq y} g_j(x)^\eta \right]^{1/\eta} \quad (4-12)$$

当 η 趋近无穷时，上式变成：

$$\mu_y(x) = -g_y(x) + g_r(x) \quad (4-13)$$

其中 $g_r(x)$ 表示最优竞争力的类别，即：

$$g_r(x) = \max_{k \neq y} g_k(x) \quad (4-14)$$

则误分类程度度量可以写成：

$$\mu_y(x) = \|f(x) - m_{y_i}\|_2^2 + \|f(x) - m_{r_j}\|_2^2 \quad (4-15)$$

其中 m_{y_i} 是真实类别中最近的原型特征， m_{r_j} 是错误类别中的最近的原型特征。则损失函数定义如下：

$$l((x, y); \theta, M) = \frac{1}{1 + e^{-\xi \mu_y}} \quad (4-16)$$

然而，这一类的损失函数会有过拟合的情况。结合对比学习，本文提出使用对比损失作为优化目标的方法。一般而言对比损失是指形如式 4-17 的损失函数，

$$\mathcal{L} = \sum_i -\frac{1}{|P_i|} \sum_{j \in P_i} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_j / \tau)}{\sum_{k \neq i} \exp(\mathbf{z}_i \cdot \mathbf{z}_k / \tau)} \quad (4-17)$$

其中 \mathbf{z}_i 是样本经过特征提取层之后输出， P_i 是与样本 i 构成负样本的下标集合。对比学习的本质是希望，在特征空间，语义相似的样本在模型学习过程中逐渐拉近，而与其他样本相互远离。

将原型向量 M （这里我们只考虑每个类比仅有一个原型向量的情况）和批次中样本的特征向量一起考虑时，我们发现，可以将原型向量也纳入到对比学习的框架中，迫使和原型向量同标签的样本向量彼此相近，从而将式 4-17 改写

如式 4-18 的形式。其中为了方便，我们令 $\bar{z}_i \in \{z_i\} \cup \{m_i\}$ ，即将原型向量加入原本需要对比学习的样本向量中，其标签和样本一样，有对应的 one-hot 向量

$$\mathcal{L} = \sum_i -\frac{1}{|P_i|} \sum_{j \in P_i} \log \frac{\exp(\bar{z}_i \cdot \bar{z}_j / \tau)}{\sum_{k \neq i} \exp(\bar{z}_i \cdot \bar{z}_k / \tau)} \quad (4-18)$$

在图像识别领域，有很多增强方法，比如 mixup，它将两张图片以一定的比例混合，同时将两者对应的标签向量也按照同样的比例进行混合，这种增强方法是为了使得模型的特征空间具有良好的线性运算性质。然而在对比学习中，样本向量之间的关系是离散的，对于特定样本，其他样本要么是正样本，要么是负样本。为此我们继续改写式 4-18，使得其可以适应样本标签非 one-hot 向量的情况。对于两个样本 (x_i, y_i) 和 (x_j, y_j) ，其中 y_i, y_j 是样本标签，考虑函数 $s(\cdot, \cdot)$ ：

$$s(y_i, y_j) = \frac{y_i \cdot y_j}{\|y_i\| \|y_j\|} \quad (4-19)$$

$s(\cdot, \cdot)$ 当且仅当标签完全一致时为 1，正交时为 0，和 one-hot 标签下一致。利用 $s(\cdot, \cdot)$ ，将式 4-18 改写如下：

$$\mathcal{L} = - \sum_i \sum_{j \neq i} \frac{s(\bar{y}_i, \bar{y}_j)}{\sum_{k \neq i} s(\bar{y}_i, \bar{y}_k)} \log \frac{\exp(\bar{z}_i \cdot \bar{z}_j / \tau)}{\sum_{k \neq i} \exp(\bar{z}_i \cdot \bar{z}_k / \tau)} \quad (4-20)$$

注意这里的 \bar{z}_i, \bar{y}_i 是从加入了原型向量和对应标签的批数据中的其中一个样本，模型基于此损失，会将相近的向量彼此拉近的同时，使得原型向量也和同一类别的样本的特征向量彼此靠近。除此之外考虑到原型彼此之间需要有较低的相似度，故本文又设计了原型向量相关度正则损失：

$$\mathcal{L}^{corr} = \sum_{i \neq j} sim(m_i, m_j) \quad (4-21)$$

相关度函数 sim 可以是 L2 范数或者是余弦相似度。本文称结合上述对比损失和正则损失的模型为 ConSim-CNN。

4.2.3 前处理算法设计

在巡检场景中，利用摄像头进行图像拍摄时，受限于摄像头的硬件条件，常常有额外的干扰因素制约后续的计算处理。这里我们主要考虑三种常见情形：摄像头图像失真、摄像头遮挡和摄像头运动模糊。

摄像头图像失真

图像失真来源于镜头畸变，镜头畸变包括径向畸变和切向畸变两类。

径向畸变是沿着透镜半径方向分布的畸变，产生原因是光线在原理透镜中心的地方比靠近中心的地方更加弯曲，这种畸变在普通廉价的镜头中表现更加明显，径向畸变主要包括桶形畸变和枕形畸变两种，如图 4-6所示：成像仪光轴

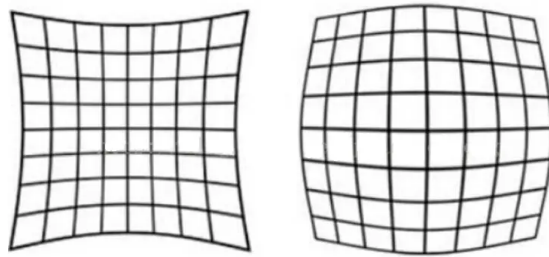


图 4-6 径向畸变

中心的畸变为 0，沿着镜头半径方向向边缘移动，畸变越来越严重。畸变的数学模型可以用主点（principle point）周围的泰勒级数展开式的前几项进行描述，调节公式为：

$$x_0 = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

$$y_0 = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

式里 (x_0, y_0) 是畸变点在成像仪上的原始位置， (x, y) 是畸变矫正后新的位置， $k_i, i = 1, 2, 3$ 是拟合系数。图 4-7是距离光心不同距离上的点经过透镜径向畸变后点位的偏移示意图，距离光心越远，径向位移越大，表示畸变也越大，在光心附近，几乎没有偏移。

切向畸变：切向畸变是由于透镜本身与相机传感器平面（成像平面）或图像平面不平行而产生的，这种情况多是由于透镜被粘贴到镜头模组上的安装偏差

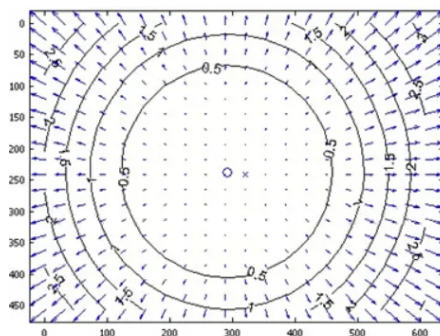


图 4-7 偏离示意图 1

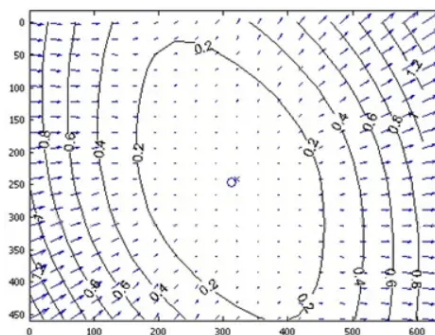


图 4-8 偏离示意图 2

导致。畸变模型可以用两个额外的参数 p_1 和 p_2 来描述：

$$x_0 = x + [2p_1xy + p_2(r^2 + 2x^2)]$$

$$y_0 = y + [2p_2xy + p_1(r^2 + 2y^2)]$$

大体上畸变位移相对于左下、右上角的连线是对称的，说明该镜头在垂直于该方向上有一个旋转角度。根据上述两式，可以得出矫正畸变后的坐标和矫正前的坐标关系为：

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = (1 + k_1r^2 + k_2r^4 + k_3r^6) \begin{bmatrix} x_p \\ y_p \end{bmatrix} + \begin{bmatrix} 2p_1x_p y_p + p_2(r^2 + 2x_p^2) \\ 2p_2x_p y_p + p_1(r^2 + 2y_p^2) \end{bmatrix} \quad (4-22)$$

其中 x_p, y_p 是原始坐标， x_c, y_c 是标定后坐标， $k_i, i = 1, 2, 3$ 和 p_1, p_2 是标定所需的相关系数。

。摄像机标定误差包括内参（4个）、畸变参数（径向和切向共5个）、外参

(平移和旋转共 6 个)。

摄像机标定过程：可以简单的描述为通过标定板，可以得到 n 个对应的世界坐标三维点 x_i 和对应的图像坐标二维点 x_i ，这些三维点到二维点的转换都可以通过上面提到的相机内参 K ，相机外参 R 和 t ，以及畸变参数 D ，经过一系列的矩阵变换得到。

假设我们提供 K 个棋盘图像，每个棋盘有 N 个角点，于是我们拥有 $2KN$ 个约束方程（乘以 2 是因为每个点都由 x 和 y 两个坐标值组成）。与此同时，忽略畸变的情况下，我们就需要求解 4 个内参和 $6K$ 个外参，也就是说，只有当 $2KN \geq 4 + 6K$ 的时候，也即 $K(N - 3) \geq 2$ 时，才能求出内外参矩阵。同时，无论在一张棋盘上检测到多少角点，由于棋盘上角点的规则布置使得真正能利用上的角点只有 4 个（在四个方向上可延展成不同的矩形），于是有当 $N = 4$ 时， $K(4 - 3) \geq 2$ ，即 $K \geq 2$ ，也就是说，我们至少需要两张棋盘在不同方位的图像才能求解出无畸变条件下的内参和外参。实际上我们往往会在一张棋盘上布置更多的角点，因为这样就可以通过最小二乘法求得最优解了，同样地，我们会需要至少 10 张以上的棋盘图像，目的是考虑数值稳定性和提高信噪比，得到更高质量的结果。

摄像头遮挡

巡检终端需要通过摄像头采集设备的图像信息，但是摄像头可能被淤泥、水滴等物体局部遮挡，影响巡检终端正常工作，也可能因为其他特殊原因被意外遮挡，巡检终端需要及时发现这种情况并进行处理。现有的遮挡检测方法虽然可以判断出长期处理摄像头上的遮挡区域，但需要处理一段时间的连续帧，速度较慢，不适用于对实时性要求较高的应用场景。并且对于具有移动性的遮挡物，如恶意的人为操作等不具有识别性。

本文提出一种准确率高、判定速度快、可以用于对实时性要求较高的应用场景的摄像头遮挡状态判定方法。此方法包括如下步骤：

1. 实时读取摄像头拍摄的一帧 RGB 图像，将图像先缩放到目标尺寸，再矫正去畸变；
2. 矫正去畸变后的 RGB 图像进行色度空间转换，得到灰度图像，再使用角点检测法提取灰度图像中的特征点，得到灰度图像的特征点集合；

3. 将所述灰度图像划分为 4 个区域并编号，分别计算每个区域的特征点个数，区域划分法如图 4-9 所示：
4. 若 4 个区域中任一区域的特征点个数小于预设个数阈值 t ，则输出摄像头被遮挡的判定结果；若 4 个区域中所有区域的特征点个数均大于预设个数阈值 t ，则输出摄像头未被遮挡的判定结果。如图 4-10 所示，当摄像头被遮挡时，某个区域的特征点数量会大幅度减少。

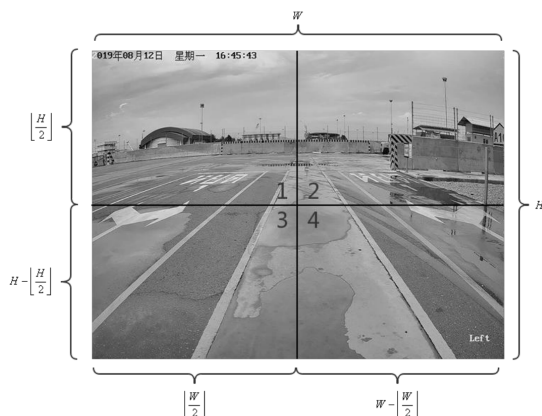


图 4-9 灰度图像的区域划分



图 4-10 遮挡判定：当摄像头被遮挡时，某些区域的特征点数量会大幅降低

本文提出的这一方法具有以下优点：可适用于配有摄像头的任何设备，仅需单个摄像头即可实现遮挡检测；基于角点检测提取特征点，特征点提取速度快；使用单帧图像即可实现检测，不依赖于连续视频图像帧信息，也不依赖于预存储信息；通过对图像划分区域，能很好地适应动态场景、全遮挡和部分遮挡。

摄像头运动模糊检测

图像质量可能会因获取和处理中的失真而降低。影响图像质量的一些常见因素包括对比度、噪声、伪影和模糊^[114]。为了解决这个问题，图像识别技术正

在不断地被进行着改进^[115]。

图像模糊是由于不完美的图像构建过程而导致的理想图像带宽降低的一种形式^[116]。在拍摄照片时,图像模糊是典型的图像失真问题。图像模糊在大多数情况下是由于聚焦不准或手持相机不稳造成的物体运动导致的图像退化问题^[117]。导致图像模糊输出的原因包括:摄像机镜头在拍摄过程中发生动态运动,物体运动,镜头无法设置正确的角度和焦距导致聚焦不准,相机质量较差^[114,118]。

有许多图像模糊检测技术可以检测图像是否模糊或清晰。其中一些技术包括:

1. 快速傅里叶变换 (Fast Fourier Transform, FFT)。在傅里叶变换中,该方法计算图像在不同点的频率,并基于设置的频率水平来决定图像是否模糊。当频率水平低于设定的频率水平时,它会判定图像为模糊的,否则,如果计算的频率较高,则图像为清晰的。关于低频和高频的阈值是由程序员来决定的^[119]。
2. 哈尔小波变换 (HaarWavelet Transform, HWT)。在这种方法中,通过对二维离散小波变换的每个小块进行迭代,将图像分割成 $N \times N$ 。然后将对角线、垂直或水平连接的小块分组成簇,并声明包含这些簇的图像为模糊图像^[120]。
3. 拉普拉斯算子 (Laplacian Operator, LAP)。这种方法用于发现图片中的边缘。它也是一个导数运算符,但与其他运算符(如 Sobel、Kirsch 运算符)的基本区别在于,所有其他导数都是一阶导数掩模。Laplacian 运算符进一步分为两类,即正 Laplacian 运算符和负 Laplacian 运算符。
4. 改进拉普拉斯 (Modified LAPlacian, MLAP)。改进的 Laplacian 算子被开发出来用于计算图像焦点质量的局部度量。通过获取 x 和 y 方向上的二阶导数的绝对值^[121]。
5. 特南梯度 (TENengrad, TEN)。基于图像梯度的著名焦点度量是通过将图像与 Sobel 算子卷积得到的,该算子也可以被视为模糊度量算子^[121]。

为了定量评估不同的模糊检测技术的性能,我们从 Tang 等人^[122]研究中提供的数据集中随机选择了 200 张模糊和清晰图像。RGB 图像大小为 640 x 480 像素。这些模糊图像可能包括运动模糊、失焦模糊和合成模糊。对于阈值,选择适当的

阈值完全取决于领域。如果所选阈值过高或过低，那么图像将被错误标记，例如，如果一张图像很清晰，但阈值设置过高，则该图像将被标记为模糊。最后的结果如表 4-1 所示：

表 4-1 模糊检测方法比较

方法名	精确率	召回率	F1	耗时 (单位: 秒)
FFT	1.0	0.87	0.93048	6.2001
LAP	0.784	0.98	0.87111	1.1482
MLAP	0.9358	0.73	0.82022	0.8951
TEN	0.94	0.94	0.94	5.6921
HWT	0.9895	0.95	0.96938	6.0370

通过权衡性能的各项指标，特别是处理速度，我们选择 MLAP 作为最后的算法选择。

4.2.4 后处理算法设计

变化检测和其他检测问题如目标检测存在一个问题，那就是在静态的图像数据上和动态的视频数据上，检测方法的效果存在巨大的差距，静态数据的检测效果要显著优于动态数据，这主要是因为视频数据由于众多的伪影、遮挡或罕见的物体姿势等困难而更具挑战性。

除了提高检测算法的性能，如专门用来对视频进行检测的算法，它会利用相邻帧的图像进行信息融合一起作为算法的输入，后处理方法也能一定程度上缓解视频数据上的预测问题。前者往往算法复杂，计算复杂度大，对于边缘设备等不够友好，而后者如 Seq-NMS^[123]和 Seq-Bbox-Matching^[124]只需要很小的改动就能达到不错的效果。他们主要将跨越多个视频帧的预测目标联系起来，并利用链接信息改进检测结果。这种策略通常比特定的视频对象检测方法快得多。

后处理方法的关键是它们如何将视频中的连续帧中检测到的对象关联起来。这种关联通常基于手工制定的启发式算法，常用的一种方法是使用目标检测算法中的概念-交并比^[125]来对关联度作估计。交并比是指两个检测框相交部分的面积与合并后总面积的比值，当两个检测框高度重合时，交并比为 1，反之完全不相交时为 0，反应了检测框的相似程度。这种方法有几个限制：首先，基础检测器并不总是能够准确预测边界框坐标；其次，交并比值强烈依赖于由于摄像

机或对象运动引起的位移，例如，快速移动的物体可能没有足够的重叠来进行可靠的关联，而当帧率降低（例如由于计算约束）时，同样的效果也会发生；最后，图像场景中可能同时存在多个对象，而这些对象有可能是重叠的。

具体到本系统的变化检测器，目前为止可以认为是静态变化检测器，虽然静态变化检测器可能在一对图片上的变化检测效果很好，但是静态变化检测器并不是专门为视频设计的，所以使用静态变化检测器对视频进行变化检测有一些限制：

1. 静态变化检测器的检测框可能包含较大的时间波动，因为它们不包含时间一致性约束。
2. 静态变化检测器仅根据单帧信息产生误检（FP）和漏检（FN），而这些误检和漏检可以通过考虑整个视频的上下文信息来区分。

视频的一个关键要素是上下文信息，因为视频中的对象的位置和外观在时间上应该是一致的，所以变化检测框在位置上不应该随时间发生剧烈变化。提高时间一致性的一种方法是将检测结果传播到相邻帧，从而减少变化检测框位置的突然变化，如果在某帧中存在某个位置发生变化，那么很可能相邻帧在相邻位置也发生了变化。因此可以通过将变化检测结果传播到相邻帧来减少检测框的突然变化以及漏检误检。

基于上述思想，我们通过缓冲区来将变化检测结果传播到相邻帧。缓冲区是指设置一个变化检测框对象数组，每个框对象存储当前框出现的频数，框的显示标示，以及框的基本属性即左上角和右下角的横纵坐标，只有连续出现一定次数的检测框才会显示出来，已经显示出的框只有连续一定次数不出现才会从缓冲区中移除。通过这种方法来平滑连续时间上的检测结果。

4.3 基于 React 的客户端设计

客户端为用户提供和系统功能交互的界面，通过设计桌面客户端，将应用界面更好地呈现给用户，并为用户提供与后端服务地交互接口。通过前端可进行视频监视、算法功能调用、基准图管理、异常日志管理等操作。

React 是一个用于构建用户界面的 JavaScript 库。它由 Facebook 开发并于 2013 年首次发布。React 采用组件化的开发方式，将 UI 拆分成一系列的组件，

每个组件都有自己的状态和生命周期。这种组件化的方式使得开发者可以更加高效地构建大规模复杂的应用程序。React 的核心思想是使用声明式编程来描述用户界面。声明式编程是一种编程范式，开发者只需要定义组件的状态和属性，React 会根据这些定义自动计算出需要更新的 UI，而不需要手动操作 DOM。这样可以大大减少手动操作 DOM 的代码量，同时提高开发效率和应用性能。React 还提供了一些重要的特性，如虚拟 DOM 和单向数据流。虚拟 DOM 是一种在内存中构建的轻量级 DOM 结构，React 会使用它来计算出需要更新的部分，然后批量更新到真实的 DOM 中，从而提高应用性能。单向数据流是指 React 应用程序中的数据只能从父组件向子组件传递，而子组件不能直接修改父组件的数据。这种数据流的限制可以使得应用程序更加易于理解和维护^[126]。

系统的前端总体分为三个部分，也是三个主要的功能模块：视频监控（包括检测算法功能调用）、摄像头管理、基准图管理。不同功能可以通过页面左侧的导航栏进行切换，切换的时候，已经进行的任务会转到后台运行。

视频监控界面是系统的主要服务，它由显示区域和控制区域组成，两个区域以垂直布局分开。显示区域水平布局，分别有三个区块：一个监视显示区块、一个基准图显示区块、一个检测结果显示区块。监视显示区块负责接受边缘设备的视频推流，基准图显示区块负责展示当前检测环境下所使用的基准图，检测结果显示区块负责展示当算法应用时检测结果的展示，包括用矩形框表示的异常区域和其中的异常类别。控制区域也是水平布局，分别由四个按钮控件组成：摄像头选择下拉菜单、监视开关按钮控件、基准图上传/更新控件、检测开关控件。摄像头选择下拉菜单可以选择当前需要控制的摄像头。监视开关控件负责监视的开启与关闭，基准图上传/更新控件将当前的监视画面作为基准图上传至服务器或者替换掉已有的基准图，检测开关控件控制异常检测服务的启动和关闭。这些控件之间彼此有制约关系，监视开关只有打开时，才能操作基准图的修改，只有基准图存在时，才能操作算法的运行。下拉菜单切换摄像头时，会保存现有的工作状态，而切换到另一个摄像头之前保存的状态。

摄像头管理界面负责摄像头的录入，页面整体由列表展示区和控制区组成。列表展示显示已经注册的设备，包含设备网络地址，设备类别，设备工作状态，设备 id 等。每一列还有删除和修改控件按钮负责摄像头信息的删除和修改。在列表的最上方提供增加摄像头控件，负责呼出摄像头增加弹框，让用户填写摄

像头信息，并记录到数据库中。基准图管理界面，负责管理基准的删除和信息修改。基准图的信息包括拍摄设备，拍摄地点、拍摄时间以及基准图 id。异常日志管理界面，负责异常日志的展示以及处理，异常展示的信息包括，发现异常的地点、时间、异常的种类、关联基准图 id、异常实时图片、以及异常处理状态。异常处理状态包括待处理、已清除、误检测三种。

4.4 本章小结

本章分别从系统的整体框架设计、核心算法设计、前后端设计、各模块设计等方面对系统整体架构与组成部分进行规划、方案设计，为后续系统的实现打下基础。其中整体框架设计结合算法特性充分利用了边云协同的优势，兼顾了效率和性能；核心算法设计，变化检测部分从检测的鲁棒性和性能平衡角度对传统算法进行了改进，开集识别部分结合原型学习和对比学习，将训练损失目标推广到了更一般的情况，有利于和其他训练方法结合提升性能；前后端以及其他模块设计，为系统的使用和平稳运行提供保障。

第五章 模块实现

本章首先介绍系统整体框架的实现方案，再依次介绍系统各模块的实现方案与细节，包括视频推流模块、前后处理模块、异常检测模块以及基准图管理模块。

系统部署在一台或多台边缘设备以及一台服务器上，所需的技术框架包括 Flask、Pybind11^[127]、gRPC、MongoDB、PyQt5、ZMQ 以及 Nginx。基于 Python 和 PyQt5 框架及其他所需组件开发实现用户桌面端界面，包括巡检功能、基准图和异常日志管理等，对于不同的操作页面，利用 Tab 组件实现切换，并保留原有操作状态。基于 gRPC 框架实现了云端的异常检测模块（开集识别部分），基于 Flask 实现了基准图管理模块以及异常记录管理模块，并完成与数据库 MongoDB 的交互，利用 Nginx 将请求 gRPC 和 Flask 的流量分别转发相应的服务器端口，设置防火墙只开放 NginX 端口。基于 Python、C++、Pybind11、ZMQ 实现前后处理模块、视频推流模块、异常检测模块（变化检测部分），其中变化检测部分的部分核心代码使用 C++ 实现，并利用 Pybind11 包装，基于 ZMQ 实现了视频推流和边缘端与用户端的异常检测信息传输。

5.1 前后处理模块实现

前后处理模块关系着系统核心算法的输入和输出，对于算法的应用效果和呈现效果起到辅助和增强作用，是不可或缺的一部分。下面分别就前处理模块和后处理模块的实现进行详细描述。

5.1.1 前处理模块实现

前处理模块主要负责对摄像头图像进行处理，包括图像校准和图片过滤。图像校准部分主要是摄像头标定，即对由于视像头镜头畸变原因造成的图像失真

进行校准；图片过滤即过滤掉不在算法正常工况状态下允许的输入图片，例如由于摄像头遮挡和摄像头运动模糊。

图像校准

图像校准需要事先对摄像头进行标定，然后在系统运行时阶段利用标定结果对图像进行变换。标定过程分为以下几个步骤：

1. 选择标定板，本文选取黑白相间的棋盘图片作为标定板，行列数选择奇数 \times 偶数，本文选择 7×10 ，如图 5-1 所示。

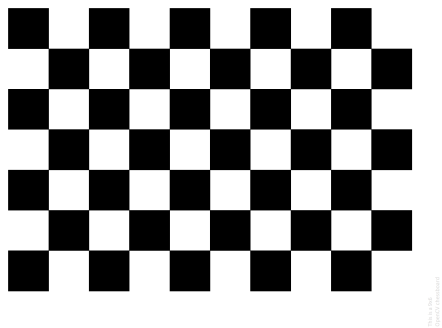


图 5-1 标定板

2. 将标定板图片置于平面上，用待标定的摄像头拍摄多组图片，平面需要和摄像头的成像平面有一定的角度，多组图片需要角度各不相同，拍摄 10 组
3. 利用 matlab 程序进行参数估计，使用 cameraCalibrator 命令，将拍摄图片导入，摄像头类型选择对应类型，本文为标准，然后导出相机参数，如图 5-2 所示，其中对涉及隐私部分做了模糊处理。

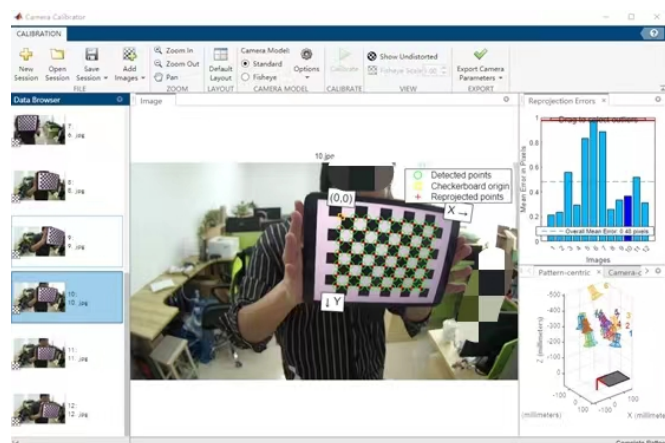


图 5-2 标定结果示意

最终得到的 `IntrinsicMatrix(3 × 3)` 矩阵和 `RadialDistortion(5 × 1)` 矩阵，分别代表摄像头的内参矩阵和畸变参数，使用 `opencv` 的 `initUndistortRectifyMap` 函数可以从这两个矩阵中计算映射变换矩阵 `map1` 和 `map2`。图像校准部分提供一个函数，它的输入是一帧实时影像图，输出是一张校准过的图片，内部用 `opencv` 的 `remap` 函数，使用前述的变换矩阵 `map1` 和 `map2` 对输入图像进行矫正。

图像过滤

图片过滤部分提供两种并联的过滤方式，遮挡检测进程和模糊检测进程。只有当两者都给出 `False` 结果，即既无遮挡也无模糊时，后续的检测算法才会实质进行，否则将会通知客户端程序，摄像头工况异常需要人为确认并干预调整。

模糊检测函数，接收一个 `numpy` 数组形式的图片，放缩到目标尺寸 `400 × 300`，初始化一个 `[-1, 2 - 1]` 的向量算子，分别从 `X`，`Y` 两个方向进行拉普拉斯度量：先使用 `filter2D` 方法在图片上利用初始化的计算核进行滑动窗口计算，最后对计算结果取绝对值，再另一个方向上重复计算，只是需要将计算核转置。对两个方向计算的结果相加，然后在维度上进行求均值即为最后的度量结果。为了确定阈值，我们在数据集上进行统计，最后选取 20 最为判断是否模糊的阈值。如果超过阈值，函数返回 `True`，图像存在模糊，否则返回 `False`。

视频遮挡函数，同样接收一个 `numpy` 数组形式的图片，放缩到目标尺寸 `400 × 300`，再经过色度空间转换为灰度图像，利用 `opencv` 库函数中的 `harris` 角点检测方法对图片进行特征点提取。接着在横纵坐标方向，对特征点集合进行分组，分成四个方位的特征点子集，并同时计数，如果任何一个方位的特征点个数小于阈值，则判定为发生遮挡，函数返回 `True`，否则返回 `False`。

模块实现

由于图像校准、图像过滤都需要消耗一定的时间，但同时它们的任务之间不是彼此依赖的，故而前处理模块我们采用并联的设计。如图 5-3 所示：

图像校准处于主进程中，不断接收实时图片，进行处理并输出处理好图片给后续模块，遮挡检测和模糊检测分别各自运行在独立的进程中，当两者任意一个返回 `False` 时，这部分算法立刻终止，未完成的算法也即刻停止，模块返回摄像头不可用状态，后续模块将暂停异常检测模块的继续使用。前处理模块会间隔一

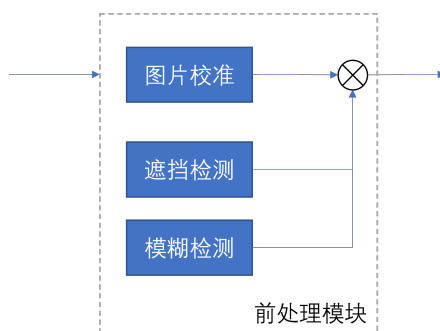


图 5-3 前处理模块示意图

段时间使用图片筛选部分的算法，同时会有一个累计计数，如果发生摄像头不可用的次数到达一定值，会认为摄像头短时间内无法恢复，则会通知前端桌面程序对使用人员进行告警提示。当摄像头恢复后，计数器清零，系统继续工作。

5.1.2 后处理模块实现

后处理模块基于缓冲区，它通过缓冲区机制将相邻图片的检测框进行平滑。我们定义两个检测框之间的相关性度量 IOU^[125]：

$$\text{IOU} = \frac{\|U_1 \cap U_2\|}{\|U_1 \cup U_2\|} \quad (5-1)$$

其中 U_1 , U_2 表示检测框所围成的区域， \cap 和 \cup 分别求检测框区域之间的重合部分以及合并部分， $\|\cdot\|$ 表示求区域面积。平滑后处理的具体步骤如下：

- 当缓冲区为空时，直接存入当前帧的变化检测框，每个检测框的出现次数置为 1。
- 当缓冲区不为空时，遍历缓冲区中的变化检测框，通过将当前帧的检测框与缓冲区的检测框的 IOU 与预先设置的 IOU 阈值进行比较从而判断当前帧是否存在检测框与缓冲区的检测框匹配。如果匹配成功，则将缓冲区的检测框的出现次数加一，如果 $\text{frequency} > \text{len}$ ，将该框的 flag 置为 True，并通过平均滤波更新框的基本属性，方法如下：

$$\overline{\text{box}_t} = \frac{\text{box}_t + \text{box}_{t-1}}{2} \quad (5-2)$$

其中 box_t 是一个四维向量，分别包含 t 时刻检测框的左上角和右下角的横

纵坐标。如果未匹配成功，则将缓冲区的当前框的出现次数减一，如果框的 $\text{frequency} < 0$ ，则将框从缓冲区中移除。

- 考虑到存在缓冲区的框与当前帧多个框匹配的情况以及某个框被另外一个框包含的情况，对于当前帧未被匹配到的框，首先判断该框是否被缓冲区中的框所包含，如果是则将该框标记为匹配。然后再计算一次与缓冲区中的框的 IOU，如果 IOU 大于预先设置的 IOU 阈值，则通过平均滤波更新缓冲区的框的属性。最后将剩下的框作为新的检测框加入到缓冲区中。
- 最后返回缓冲区中所有 flag 为布尔值真的变化框。

在异常检测模块的变化检测部分输出变化区域数组时，由于这部分矩形框可能存在误检测的情况，所以需要开集识别部分识别筛选完后在进行后处理，但是如果这些检测框全部都是正确的异常，则会浪费等待时间。故而，实现中，后处理模块会先进行对于变化区域数组的后处理，当开集识别部分通过远程调用返回时，比较变化区域数组与变化区域标签，如果判定为无修正，则直接使用预处理的结果，否则再次进行后处理，这样可以以一定概率减少处理时间。

5.2 异常检测模块实现

异常检测模块是系统的核心算法部件，我们将它拆分为两阶段的算法，即变化检测部分和开集识别部分。得益于这样得设计，在具体实现时，可以将变化检测部分部署在边缘设备上，将开集识别部分部署在云端服务器上，尽可能地节省计算和网络资源。具体的模块结构示意图如下所示：

异常检测模块接收两种输入，一种是 RGB 实时图，格式是 numpy 的矩阵形式 (`np.ndarray`)，提供实时图异常检测服务；一种是超参数，格式是 json 字典格式，用来设置算法模型参数，调节模型性能，一次设置完，后续沿用此设置。在提供实时图异常检测服务时，如果当前的基准图缺省，模块使用 HTTP 服务基于时间和位置 id 向云端查询此时应该使用基准图，同样处理成 numpy 矩阵格式，作为模块输入。如果基准图已经存在，则会重复利用基准图在模块中的中间处理结果，提高效率。切换地点或者时间时，模块会将基准图清空，等待下次调用。

如图 5-4 所示的边缘设备侧，经过变化检测部分处理后，会输出变化区域数组，数组元素是一个四维元组，分别是变化区域矩阵框的左上端点和右下端点的

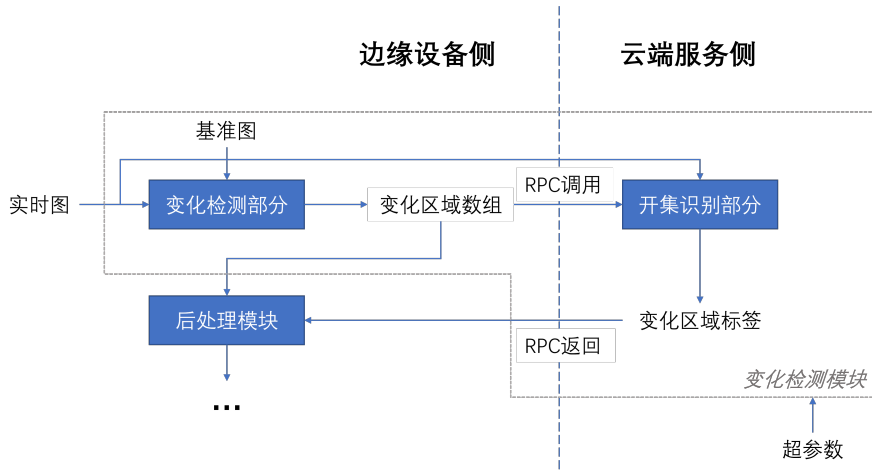


图 5-4 异常检测模块结构图

坐标（以百分比为单位）。如果数组为空，则代表实时图与基准图相比较不存在变化，本模块直接转移至系统后续的处理模块；如果数组不为空，则利用 gRPC 服务调用云端服务器的开集识别部分，输入是变化区域数组和实时图 numpy 矩阵，输出是对应于变化区域数组的变化区域标签，gRPC 返回结果后，本模块将变化区域数组和变化区域标签交予系统后续模块进一步处理。

下面分别对于模块的边缘侧部分和云端服务部分的实现细节进行描述。

5.2.1 变化检测部分

变化检测部分主要完成给定基准图和实时图，在实时图上找到相比较于基准图，出现了明显变化的部分，组成变化检测部分的各个分算法以及相互关系如下图所示：

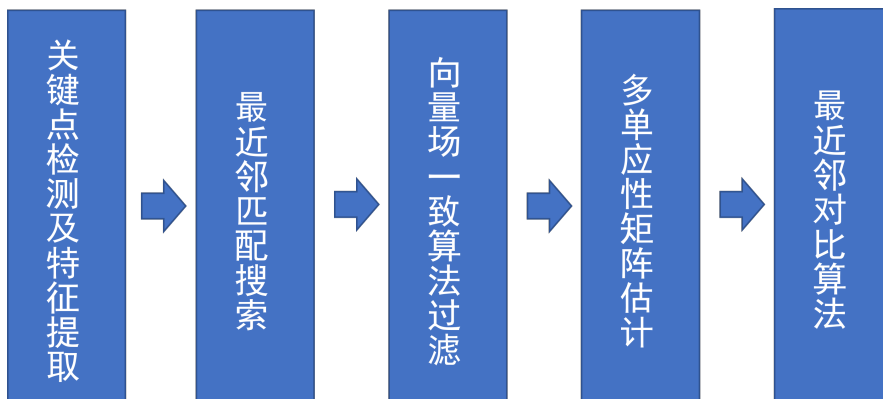


图 5-5 变化检测示意图

关键点检测及特征提取部分计算实现两个功能，一种是输入只有一张图片，

则输出图片中的关键点坐标和对应的特征描述子，关键点有多个，输出是一个数组；另一种是输入为图片以及坐标点数组，输出这些坐标点位置图片的特征描述子。前一个功能的提取器基于 SuperPoint，它的输入必须是灰度图像，故而需要转成灰度图像格式，后一种功能基于 OpponentSIFT，它的输入是彩色图片，可以不需要准换。图片需要事先转成固定大小的宽 800，高 600 的图片。

变化检测方法首先包括对基准图和实时图进行关键点提取，以及对应关键点特征描述子的计算。先将输入的 RGB 图像转化成灰度度，然后再归一化，作为模型 SuperPoint 的输入，然后将模型输出从张量形式变为适合 python 处理的数组形式。基准图的关键点和对应特征描述子可以保存下来，以供后续使用，只有当基准图重新设置时才重新计算。模型会在模块初始化时就预加载，以供后续直接调用。

得到基准图和实时图的关键点和特征描述子后，会计算粗略的匹配关键点，这部分由最近邻匹配搜索部分实现。算法会使用 opencv 中的快速最近邻方法进行计算，算法使用 KD 树来进行检索，树的深度设置为 6，最近邻数目 k 设置为 2。算法会寻找与被查询点的特征差异最小的前两个点，即有两个匹配，应用 Lowe's algorithm 可以对此进行粗略的筛选。它会对两个匹配计算得到的特征差异在向量空间进行比较，如果两者相差太小，则意味着被查询点和不止一个点的特征有相似，也就是说这个点不够具有代表性，故而应该从匹配集中筛选出去，当特征向量距离差异很大，则取第一个最相似点作为匹配的点。

对于得到粗筛匹配关键点，我们使用 VFC 算法进行精筛，这部分由向量场一致算法过滤实现。VFC 算法由于需要效率，将其以 c++ 语言进行实现，并用 pybind11 对其进行包装，以供 python 进行调用。由于要解算匹配点之间的矢量位移向量场 f ，需要求解式 4-6，而对于匹配点数目 N 较多的情况下，计算复杂度是 $O(N^3)$ ，存储复杂度是 $O(N^2)$ ，所以我们求解一个次优解但是能以更有效率的方式求解。相比于在空间 \mathcal{H}_N 中寻找最优解，我们使用一个稀疏近似，并在一个使用更少基函数定义的空间 $\mathcal{H}_M = \left\{ \sum_{m=1}^M \Gamma(\cdot, \tilde{x}_m) c_m : c_m \in \mathcal{Y} \right\}$ 中寻找近似解，并在所有样本集合上最小化正则化后的损失函数。其中， \mathcal{Y} 表示输出空间， Γ 是与可再生希尔伯特空间绑定的核，这里我们使用的 $M \ll N$ ，我们选取的集合 $\{\tilde{x}_m : m \in \mathbb{N}_M\}$ 是集合 $\{\tilde{x}_n : n \in \mathbb{N}_N\}$ 随机采样的子集^[128]。我们发现任意选取原训练集合中的子集的方法并不比原先得复杂方法表现差。根据稀疏近似，

我们搜索具有如下形式得解：

$$f(x) = \sum_{m=1}^M \Gamma(x, \tilde{x}_m) c_m, \quad c_m \in \mathcal{Y} \quad (5-3)$$

其中的系数 $\{c_m : m \in \mathbb{N}_M\}$ 由下面的线性方程决定：

$$\left(\tilde{\mathbf{U}}^T \tilde{\mathbf{U}} + \lambda \sigma^2 \tilde{\Gamma}_s \right) \tilde{\mathbf{C}}_s = \tilde{\mathbf{U}}^T \tilde{\mathbf{P}} \tilde{\mathbf{Y}} \quad (5-4)$$

其中 $\tilde{\mathbf{C}}_s = (c_1^T, \dots, c_M^T)^T$ 是系数向量， $\tilde{\Gamma}_s$ 是一个 $M \times M$ 的 Gram 矩阵，其 (i, j) 个块是 $\Gamma(\tilde{x}_i, \tilde{x}_j)$ ， $\tilde{\mathbf{U}}$ 是一个 $N \times M$ 的 Gram 矩阵，其 (i, j) 个块是 $\Gamma(x_i, \tilde{x}_j)$ 。与由表示定理给出的最优解不同，不再是由基函数 $\{\Gamma(\cdot, x_n) : n \in \mathbb{N}_N\}$ 的线性组合，而是其随机的 M 元组线性组合。总的来说，这样的稀疏近似在速度和内存消耗上有着极大的性能提升，而在准确度上的损耗可以近似忽略不计。我们将这样的实现叫 SparseVFC，和算法2中第10行求解不一样的线性方程。

得到精筛后的匹配点后，接下来算法求解多单应性矩阵并为每个代表格点分配其所属单应性矩阵。计算多个单应性矩阵时，循环利用 RANSAC 算法将匹配点分割为内点和外点，并将外点更新为新的匹配点重复操作，直到剩余外点数目低于最初的匹配点数目的一定比值。将每次 RANSAC 得到的内点和对应的单应性矩阵用数组存储起来，作为 KNN 算法的训练集。对基准图图片进行格点采样，然后对每个格点通过带权 KNN 计算其应该使用的单应性矩阵，即前文得到的内点依据近邻关系和权重函数对其投票，选择投票最多的单应性矩阵。由于对每个格点进行单应性变换效率太低，故将使用相同单应性矩阵的格点收集起来，同一应用单应性变换。变换后，计算变换后的靶点位置，如果超出实时图的范围，则忽略，否则计算变换前后点的位置处的特征描述子差异，如果差异超过一定阈值，则标记为变化点。得到所有的变化点之后，对其进行膨胀，使得直接相邻格点能像素上联通，最后通过图联通分量找到变化区域并用矩形框框出。对得到的矩阵框进行处理：如果矩阵框之间有重叠，则将两者合并成覆盖两者区域的大矩形框。算法最后返回在实时图上的变化区域矩形框。

5.2.2 开集识别部分

开集识别部分部署在云端服务器上，使用 gRPC 提供服务。当然算法也是可以部署在边缘端的，因为模型的轻量化设计以及计算优化可以让其也能很好在计算资源紧张的设备上使用。gRPC 分为客户端和服务端，位于客户端的程序可以像调用本地对象一样直接调用另一台不同的机器上服务端应用的方法。首先需要定义一个开集识别算法服务，指定其能够被远程调用的方法 `inference`，这个方法包含参数 `InferenceRequest` 以及返回类型 `InferenceReply`。

其中 `InferenceRequest` 表示模型的输入，即实时图和变化区域矩形框数组，矩阵框信息包含左上角和右下角的端点的横纵坐标，以浮点数百分比表示。`InferenceReply` 表示模型的输出，即对应于变化区域矩形框数组的变化区域类别数组，类别由无符号整数表示，从数字 0 开始。开集识别算法服务的定义包含在 `inference.proto` 文件中。

接着需要将写有服务定义的 `inference.proto` 文件通过 `protoc` 命令解析并生成脚手架代码：`inference_pb2_grpc.py`、`inference_pb2.py`、`inference_pb2.pyi`，分别包含 gRPC 的服务器定义、信息定义、信息类型定义，为后续实现提供 python 调用接口。之后需要在服务端实现这个接口，除此之外还需要一个辅助函数，可以将字节码流形式的图像数据转换成 `numpy` 矩阵的形式，和变化区域矩阵框数组一起输入进模型 `ConSim-CNN`。

最后运行一个 gRPC 服务器来处理客户端调用，我们设置服务端口号为 50052，并将前述实现的接口注入到服务器中。在服务启动上，我们将其包裹在 `asyncio` 的 `async/await` 原语下，在 io 密集型场景下，可以更好的发挥服务器性能。`asyncio`，是一个 Python 内置的标准库，它提供了一种基于协程的异步编程模型，使得编写高效且可伸缩的异步应用程序变得更加容易。

模型 `ConSim-CNN` 实质负责算法的运行，在整个服务初始化阶段，会进行 GPU 资源请求、资源分配、模型预加载、模型预热等步骤，其中模型预加载和预热是为了让 `cpu`、`gpu` 的缓存调度向模型调用倾斜，减小后续调用延迟。`simcnn` 的 `infer` 方法会将图片 `image` 依据变化区域矩形框 `boxes` 裁切出不同的块，合并成数组形式，这样可以让模型一次性处理，加快处理速度。然后经过 `transform` 变换：重新设置大小为 256x256、中心裁剪为 224x224、转变成张量，接着传入

ConSim-CNN 中。ConSim-CNN 模型结构采用一个特征提取层与一个特征投影层的形式，具体如下：

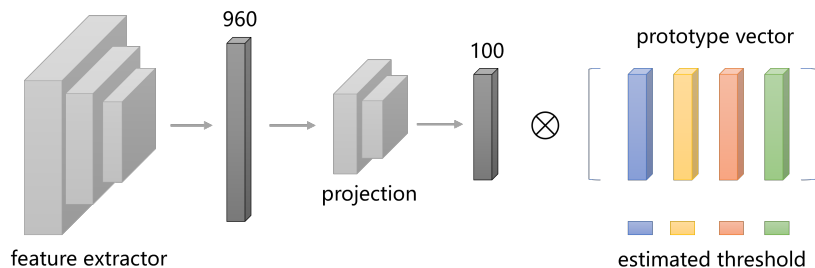


图 5-6 Con-SimCNN 推理示意图

模型的特征提取器部分采用 ghostnet 的结构，ghostnet 是一种轻量化的网络结构，在相同的计算成本下可以比 MobileNetV3 获得更高的识别性能，这里只使用其中的特征提取层，模型会输出一个 960 维的特征，后续采用一个级联 480、100 维的线性投影层，将特征降维到 100 维。降维到 100 维的好处是即尽可能地保持了特征信息，又缓解了维度太大的距离计算问题。得到输出特征向量后，将其与原型特征向量计算欧几里得距离（先各自归一化），距离最近的原型向量即为可能的类别，如果相应距离大于估计阈值，则表明是未知类别，否则为该类别。估计阈值是在训练数据集上取相应类别数据的特征向量与原型向量之间的距离经过均值方差估计而来，在实际推理过程中，使用均值加 0.5 倍的方差作为阈值。得到变化区域的类别后，将数据从 gpu 上的张量转为到 cpu 上的数组，并由 RPC 服务器经过数据包装后返回给客户端。

ConSim-CNN 的训练阶段，首先是数据集的准备，这里我们以国家电网变电站异常数据为例。变电站异常数据本身是一种目标检测数据集，一共有 4 种检测类别，每种类别有 100 张图片，标注信息文件中写有检测框，一张图片中存在一个或者多个目标。本文利用检测框信息将对应异常裁剪出来，并在原检测框的基础上扩大 20% 的大小，引入一定的背景信息，最后一共获得了 431 张异常图片。除了异常图片，在模型测试时，还需要一些非异常图片测试模型对未见类别的识别能力。如图 5-7 所示，我们取标准大小的矩形框作为非异常图片裁切大小，矩形框中心在图片中随机采样，如果候选矩形框与图片中异常区域矩形框重合

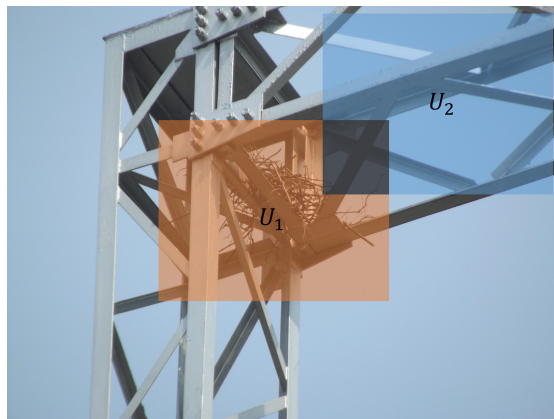


图 5-7 非异常数据构造示例

率低于 15%，则认为非异常区域，选入非异常图像集合。重合率定义如下：

$$\rho = \frac{\|U_1 \cap U_2\|}{\|U_2\|} \quad (5-5)$$

其中 U_1 表示数据标注的矩形框， U_2 表示图片中任意的非异常区域矩形框， $\|\cdot\|$ 表示矩形框面积， \cap 表示对两个矩形框求重合区域。我们取 15% 的异常图片作为测试集的一部分，再从非异常数据中取相同数目的图片一起补充进去，形成最后的测试集，剩下的异常图片作为训练集，这里注意训练过程中只用到了异常图片。

训练过程中，模型 ConSim-CNN 的特征提取层使用 ghostnet 的官方预训练参数^[129]的 feature 部分，对于降维投影层以及原型向量组的参数则使用标准正态分布方法进行初始化。对于训练集，选取批大小为 20。由于训练数据非常少，需要对其做数据增强。我们不使用固定的一个或者几种数据增强方法，而是对于一个批次的数据图片，对其进行 RandAug 增强。RandAug 随机从一些数据增强方法列表选取 n 个，然后对图像进行不改变图片语义的变换，并利用取值参数 m 调节各个方法的程度范围。RandAug 包含的方法如表 5-1 所示。图像经过 ConSim-CNN 后会得到一个 100 维度的向量，为了计算相似度，虽然可以直接点乘，但是有研究表明^[106]将向量先归一化到单位球上，效果会更好。模型的对比学习误差为

$$\mathcal{L}^{con} = - \sum_i \sum_{j \neq i} \frac{s(\bar{\mathbf{y}}_i, \bar{\mathbf{y}}_j)}{\sum_{k \neq i} s(\bar{\mathbf{y}}_i, \bar{\mathbf{y}}_k)} \log \frac{\exp(\bar{\mathbf{z}}_i \cdot \bar{\mathbf{z}}_j / \tau)}{\sum_{k \neq i} \exp(\bar{\mathbf{z}}_i \cdot \bar{\mathbf{z}}_k / \tau)} \quad (5-6)$$

表 5-1 数据增强方法列表

方法名	作用	参数范围
AutoContrast	最大化图像对比度	-
Equalize	图像直方图均匀化	-
Invert	对图像取反	-
Rotate	旋转图像	0, 30
Posterize	色调分离	0, 4
Solarize	像素值高于阈值取反	0, 256
SolarizeAdd	像素值增加再 solarize	0, 110
Color	色彩空间转移	0.1, 1.9
Contrast	改变对比度	0.1, 1.9
Brightness	改变亮度	0.1, 1.9
Sharpness	改变锐度	0.1, 1.9
ShearX	沿 X 轴仿射变换	0, 0.3
ShearY	沿 Y 轴仿射变换	0, 0.3
CutoutAbs	图片填充	0, 40
TranslateXabs	沿 X 轴平移变换	0, 100
TranslateYabs	沿 Y 轴平移变换	0, 100

其中 $\bar{\mathbf{z}}_i$ 表示第 i 个扩展后样本的特征向量（包含进了原型向量）， $\bar{\mathbf{y}}_i$ 表示第 i 个扩展后样本的标签（包含进了原型向量对应的标签）。这里我们不采用一般方法使用的 one-hot 向量，而是经过平滑后的标签向量

$$\mathbf{y}^{smooth} = s \times \mathbf{y}^{one-hot} + \frac{1-s}{m-1} \times (\mathbf{1} - \mathbf{y}^{one-hot}) \quad (5-7)$$

其中 $\mathbf{y}^{one-hot}$ 是 one-hot 向量， \mathbf{y}^{smooth} 是平滑后向量， $\mathbf{1}$ 是全 1 向量， s 是平滑系数， m 是类别个数，本文 s 取 0.8， m 取 4。

在进行对比损失函数的计算上为了效率，不能采用遍历下标的循环方式进行，而是需要将其转化称矩阵运算形式。首先是标签相似度矩阵 $S \in \mathbb{R}^{N \times N}$ ， N 表示批次大小， $S_{i,j}$ 表示样本对 i 和 j 的标签向量 $\bar{\mathbf{y}}_i$ 和 $\bar{\mathbf{y}}_j$ 之间的相似度，注意，这里的样本已经将原型向量也包含进去，并且标签向量已经不再是 one-hot 向量，经过了标签平滑处理。令 $mask \in \mathbb{R}^{N \times N}$ 为对角线全为 0，其他都为 1 的矩阵， $mask \times S$ 并在横轴上求和，即可求得式 5-6 中的标签相似度的分母部分。特征相似度可以进行相同的处理。

我们希望学习的原型向量能够在特征空间中尽可能地互相远离彼此，这样在各自的阈值范围内作分类判断时，不会因为判别空间有很大重叠而造成混淆，用相关性误差来表示，即

$$\mathcal{L}^{corr} = \sum_{0 < i < j < m} \frac{1}{\|m^i - m^j\|} \quad (5-8)$$

这里利用了原型向量在特征空间的势能函数，当彼此之间因为斥力相距最远时，势能最小。最终的损失函数是分类损失与相关性损失之和。

$$\mathcal{L} = \mathcal{L}^{con} + \mathcal{L}^{corr} \quad (5-9)$$

对于模型的优化器部分，我们采用 Adam 优化方法，学习率取 0.001，权重衰减系数取 0.01，学习轮数取 20，预热轮数比例选择为总轮数的 20%，设置模型早停，并采用线性规划器。最后，我们需要在训练数据集上估计对每个类别的预测阈值，阈值基于如下方法：收集每个类比数据与对应类别的原型向量的距离，然后对这些距离计算均值 mean 和方差 var，取 $\text{mean} + \lambda \cdot \text{var}$ 作为最后的类别阈值， λ 取值为 0 到 1，本文取 0.5。

模型训练好后，为了使得推理速度加快，利用 onnxruntime 库对模型进行进一步的算子优化和算子融合，并利用 TensorRT 作为最后的推理框架。

5.3 视频推流模块实现

对于 IP 摄像头，例如海康威视 smart265，设备自身具有流媒体传输接口，客户端使用 opencv 的 VideoCapture 直接读取。VideoCapture 基于 RTSP 协议与 IP 摄像头建立连接，通过 isOpened() 方法判断连接建立情况，如果返回值为 False，则再次尝试连接，如果再次失败，则弹出报错对话框，提醒用户确认设备状况。VideoCapture 连接正常后，通过 read() 方法读取缓冲区队列的队首的一帧。

对于配备了 usb 摄像头的边缘设备，边缘设备使用 opencv 的 VideoCapture 读取 usb 接口上的视频帧，然后使用 ImageZMQ 将图像传输到客户端。由于 opencv 使用 uvc 格式在边缘设备上读取 usb 摄像机流，而 VideoCapture 默认为未压缩的流，需要将流格式更改为 MJPG 格式（需要 usb 摄像头支持），否则会造成图像

数据无法很好处理，抛出“select timeout”错误。读取 usb 接口的视频帧时，为了保证无延迟，画面无拖曳感，需要将 VideoCapture 的 CAP_PROP_BUFFERSIZE 属性设置为 1，即对 usb 视频帧不做缓存，边缘设备直接转发。ImageZMQ 作图像转发时，采用的连接模式是 PUB/SUB 模式，即边缘设备作为一个发布者，客户端侧则是订阅者，可以支持多个客户端订阅同一个视频发布者。在图片转发时，需要将图片编码成 jpeg 格式，可以极大地降低网络带宽。

在客户端读取传输过来地图片帧时，由于有缓存机制以及图片传输过程中不可控地差错，不能简单的以固定时间读取图片，否则会出现花屏跳帧的现象。StreamSubscriber 抽象类实现了 receive 方法，会阻塞等待数据指示信号量，并返回缓存图片成员变量，将图片返回。还实现了 close 方法，其作用是设置停止标记成员变量为真。抽象类有一个需要继承类实现的抽象方法线程运行程序 run，类初始化时会新建一个线程，并设置其参数变量为 run 方法，run 方法需要子类具体实现读取视频流的功能。基于 RTSP 和 ImageZMQ 的视频串流分别继承了 StreamSubscriber 抽象类，即 RTSPStreamSubscriber 和 ZMQStreamSubscriber。他们都有相似的循环结构，不断阻塞读取接收缓冲队列中的图片，直到缓冲为空阻塞等待，每次读取都会更新读取的图片到缓存图片成员变量，并设置线程信号量数据指示信号量。循环的停止条件是停止标记成员变量值为假。循环停止后将释放协议连接所占用的资源。两个实现类所不同的方面在于连接的建立方法和读取新图片的方式。

5.4 基准图管理模块实现

基准图是整个系统中算法良好运行颇为重要的一部分，基准图管理模块是为了方便算法和用户随时调取以及管理用户上传的基准图片，包括基准图上传、基准图更新、基准图查询以及基准图更新等功能。基于 flask-sqlalchemy 的 ORM (Object Relation Mapping) 框架，将与图片相关的详细信息存储在数据库中，以对类的成员变量进行调用的方式获取其具体数据，flask-sqlalchemy 会将其转变成具体的 SQL 查询语句到数据库中进行查询，框架还提供会话机制，可以处理诸如数据库事务失败、状态回滚等等异常情况。

在基准图类的设计中，每个基准图对象包含以下成员：基准图文件 id、基准

图上传时间、基准图拍摄地址、基准图拍摄季节、基准图拍摄时间段，如表 5-2 所示。

表 5-2 基准图类成员

名称	成员变量名	成员类型	说明
基准图文件 id	file_id	String	图片文件唯一标识符
基准图文件存储路径	file_url	String	
上传者	username	String	
基准图上传时间	create_time	datetime.datetime	日期类型
基准图拍摄地址	create_address	String	用于位置定位
基准图拍摄季节	create_season	Integer	以 24 节气为单位，从 0 到 23
基准图拍摄时间段	create_span	Integer	以小时为单位，从 0 到 23
基准图备份状态	backup_status	StatusEnum	备份状态、枚举类型

基准图上传有两个入口：可以在实时巡检时，在不打开检测算法的情况下，在每个目标位置，选择当前实时影像的一帧图片作为当前目标位置的基准图，系统会根据自动获取的时间、地址、授权信息调用本模块接口；也可以在基准图管理界面，选择文件上传的方式，这时需要手动设置时间、地址信息。在系统运行时阶段，每当到达指定的目标位置，系统会根据当前的地址和时间信息，向本模块请求查询服务，然后进行后续的计算检测。另外在基准图管理界面，也可以对基准图进行信息修改甚至删除，本模块确保了基准图的全面使用。

除了管理和保存与基准图有关的详细信息，对于图片文件本身，模块采用本体文件存储加分布式备份的方式。本地文件存储使用文件 id 作为文件名，存储在机械硬盘上，由于巡检之前的路线和时间是已知的，可以利用这个信息将相关的基准图预读取到内存中，加快访问响应。由于基准图非常重要，是宝贵的图像数据，可以为以后更多的相关研究提供支持，故而本系统还实现了冗余备份方案，利用 FastDFS 进行分布式文件系统的搭建，由于将众多的基准图备份是一个耗时的过程，故而需要实现一个异步调用接口，本系统利用 python-celery 来实现这样的任务系统，在基准图管理页面确定好需要备份的基准图范围后，点击备份，本模块会调用此异步接口，当备份任务完成后，会根据备份成功与否状态，修改每张图片的备份状态，即 backup_status，它由枚举类实现，包含以下状态：未备份、备份中、备份失败、备份成功，对应表 5-3 所示，在界面刷新时，可以看到图片对应的状态。

表 5-3 枚举类 StatusEnum 成员

名称	成员命名	成员数值	说明
未备份	UNBACKED	0	图片没有备份过
备份中	BACKING	1	备份任务正在执行
备份失败	FAIL	2	备份过程报错
备份成功	SUCCESS	3	备份任务完毕且无报错

5.5 本章总结

本章对系统各模块的具体实现方法作了详细的阐述，包括前后处理模块、异常检测模块、视频推流模块以及基准图管理模块的实现。其中对于前后处理模块包括前处理与后处理部分，并对其中多线程并行运行进行了介绍。异常检测算法模块实现了跨设备框架的构造，并对于模型的构造、训练、实际推理等方面进行了完整阐述。

第六章 系统测试

本章分别从系统各模块测试、集成测试及非功能性测试对系统功能完整性、系统性能、系统质量等方面进行测试，并展示系统实际运行效果。系统测试是在软件开发过程结束时对完整实现的集成系统上进行测试，以验证系统是否符合需求分析所列出的实现要素，达到系统功能确认、系统风险控制的目的。系统测试分别针对功能性需求与非功能性需求进行测试，分为模块测试和集成测试，其中模块测试对系统每一子模块分别编写测试样例进行功能性测试；集成测试检测集成在一起的模块的协同行为是否存在缺陷，同时对集成单元性能负载、可靠性等非功能性需求进行测试。

6.1 测试环境

测试环境为一台服务器主机、一台个人电脑，还有一台边缘设备，其中服务器包含 Nginx、Python、MongoDB、Flask 等组件框架，服务器的配置如表 6-1 所示。

表 6-1 服务器硬件配置

硬件名	硬件信息
操作系统	Ubuntu 18.04.5 LTS
处理器	Intel Xeon E5-2678 v3@2.50GHz
内存	128GB
显卡	NVIDIA GeForce RTX 2080 Ti
网卡	Intel I350

服务器主机中设置 gRPC 服务监听端口 41400，Flask 后端监听端口 41401，MongoDB 监听端口 41402，NginX 监听端口 41404，防火墙设置对外开放 41404 端口，NginX 反向代理规则中将/api 转发至端口 41401，其余流量转发至端口

41403。

边缘设备是 Nvidia Jetson Xavier NX，其上配置如表 6-2 所示，除此之外还有一个罗技的 USB 摄像头 Logitech HD Pro C920，可以支持 1080p 30FPS 或者 720p 60FPS 的摄像。

表 6-2 边缘设备硬件配置

硬件名	硬件信息
操作系统	Ubuntu 20.04.5 LTS
处理器	64 位 ARM Carmel 架构，6 核，6MB L2 缓存，4MB L3 缓存
显卡	Volta 架构，384 CUDA 核心，48 张量核心
内存	8GB（与显卡共享）
网卡	Gigabit Ethernet

个人主机测试环境如下表 6-3 所示：

表 6-3 个人主机测试环境

硬件名	硬件信息
操作系统	win11
处理器	Intel Core i5-9300H@2.40GHz
内存	16GB
显卡	集显，512MB 显存

6.2 模块测试

模块测试用于验证构成系统的每个模块是否能正确执行其各个功能，模块测试中又包含更细粒度的单元测试，用于验证组成程序的相对较小的单元是否正确执行其各个功能

6.2.1 变化检测模块测试

本节对变化检测模块进行测试，分为多单应性矩阵方法和 VFC 方法。

对于多单应性矩阵方法，我们在 Hpatches 和 CDNet Dataset 2014 两个图像数据集上进行了实验。HPatches 是 2016 年提出的一个局部描述子评价指标，它包含 57 个光照文件夹和 59 个视点文件夹。每个文件夹都包含 1 个参考图像和 5 个

在不同光照或不同视点下对于同一场景拍摄的图像，也提供了参考图像和变化图像之间的单应性矩阵。CDNet Dataset 2014(CD2014)是在 CDNet Dataset 2012 的基础上改进得到的。CDNet Dataset 2012 是一个包含 31 个视频，大约 90000 个手动注释帧的数据集，CDNet Dataset 2014 新增了五个类别。每个类别包含 900 到 7000 帧图像，对于每一帧图像，CD2014 数据集都提供了像素级别的 GroundTruth 标签。

关键点检测和多单应性矩阵的实验在 Hpatches 数据集上进行，性能评价指标是关键点可重复性 (R) 和匹配成功的关键点比例 (M-Ratio)。可重复匹配定义如下：给定一个阈值 ϵ ，一对关键点 a, b ，可重复匹配表示为 a 经过单应性矩阵 H 转换后与 b 的距离小于 ϵ ：

$$\text{corr}(a, b) = \|b - a \times H\| \leq \epsilon \quad (6-1)$$

两张图像的可重复性可以表示为两张图的可重复匹配点数与两张图的总匹配点数的比例：

$$R = \frac{1}{N} \sum_i \text{corr}(a_i, b_i) \quad (6-2)$$

匹配成功的关键点是指计算单应性矩阵时的内点，单应性矩阵是在关键点检测部分匹配成功的关键点的基础上计算的，所以这里匹配成功的关键点的比例计算如式：

$$\text{M-Ratio} = \frac{\text{inliers}}{\text{good Matches}} \quad (6-3)$$

此外我们也添加了一个综合指标 F ，它为 Repeatability 和 MatchRatio 的加权平均值：

$$F = \frac{R + \text{M-Ratio}}{2} \quad (6-4)$$

变化检测是一个复杂的过程，需要通过关键点检测、关键点匹配、单应性估计和差异检测等多个步骤来实现。最终目标是通过在变化的区域周围绘制矩形边界框来标记出地图或图像中的变化。虽然现有的一些遥感变化数据集和街景数据集提供了 Ground Truth 标签，但它们对于微小的变化不够敏感，只能标记出较大幅度的变化。因此，我们选择了 CDNet 2014 数据集，该数据集提供了更细致的 Ground Truth 标签，能够标记出图像中发生的变化。我们使用算法检测到的

表 6-4 Hpatches 关键点匹配实验，以不同光照场景为测试对象

	单个单应性矩阵			多个单应性矩阵		
	R	M-Ratio	F	R	M-Ratio	F
SIFT	0.968	0.744	0.856	0.956	0.834	0.895
SuperPoint	0.953	0.879	0.916	0.952	0.890	0.921
Key.Net	0.962	0.914	0.938	0.962	0.920	0.940

变化框与 Ground Truth 标签中处理得到的变化框进行比较，以评估变化检测的效果。由于时间和资源的限制，我们只选取了 CDNet 2014 数据集中的 Baseline、BadWeather 和 Dynamic-Background 等三种场景的一部分数据进行实验。评价指标是变化区域的准确率 (Precision) 和召回率 (Recall)。具体定义如下：假设通过变化检测框架预测的变化框集合为 a_1, \dots, a_k ，通过 Ground Truth (GT) 得到的变化框集合为 b_1, \dots, b_n ，预测框的总面积为 S_{pred} ，GT 框的总面积为 S_{gt} ，预测框与 GT 框的交叉面积为 S_{iou} 。则 *Precision* 在此定义为预测变化的地方 S_{pred} 中有多少为真正变化的地方 S_{iou} ，*Recall* 定义为真正变化的地方 S_{gt} 中有多少被预测出来了。

$$Precision = \frac{S_{iou}}{S_{pred}} \quad (6-5)$$

$$Recall = \frac{S_{iou}}{S_{gt}} \quad (6-6)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6-7)$$

本算法没有限定检测图像关键点的方法，因此我们采用了 SIFT、Key.Net、SuperPoint 三种不同方法分别在视点变化和光照变化下进行单个单应性矩阵与多个单应性矩阵的实验，这里我们选取每个方法的 1000 个关键点来测试，结果如表 6-4、表 6-5 所示。由于 CD2014 数据集每个类别的每个场景的图像都是在同一个视角拍摄的，而且图像的物体不存在相对运动，而多单应性矩阵主要用于解决复杂视角变化以及有相对运动物体场景下的单应性估计，所以这里没有进行单个单应性矩阵和多单应性矩阵的对比实验，多单应性矩阵的实验结果如表 6-6 所示。

从表 6-4 中我们可以看到在光照变化场景下，使用一个单应性矩阵时 Super-

表 6-5 Hpatches 关键点匹配实验，以不同视点场景为测试对象

	单个单应性矩阵			多个单应性矩阵		
	R	M-Ratio	F	R	M-Ratio	F
SIFT	0.939	0.743	0.841	0.938	0.838	0.888
SuperPoint	0.943	0.798	0.870	0.941	0.840	0.890
Key.Net	0.956	0.869	0.913	0.955	0.899	0.927

表 6-6 CD2014 视频变化检测实验

	多单应性矩阵		
	Precision	Recall	F1
SIFT	0.880	0.832	0.848
SuperPoint	0.877	0.807	0.831
Key.Net	0.880	0.827	0.846

Point 的单应性估计效果远远好于 SIFT，使用多个单应性矩阵对 SIFT 的提升效果较大，SuperPoint 只有轻微的提升。从表 6-6 我们可以发现在视点变化场景中使用多个单应性矩阵对 SIFT 和 SuperPoint 方法都有很大的提升，这也与我们预期是相符的。因为视点变化可能会导致图像深度的变化，打破了传统单应性估计的约束，因此视点变化下多单应性矩阵的效果会提升更多。使用多单应性矩阵可能会导致关键点可重复性有小幅度的下降，但会大幅度提升匹配成功关键点的比例。图 表 6-6 是变化检测实验结果，从中我们可以看出 SIFT、SuperPoint 的 Precision、Recall 和 F1 三个指标相差不大，反而是 SIFT 的效果最好。我们分析了导致这个结果的原因，这是 CD2014 数据集的限制，因为数据集的图像既没有视点变化也没有光照变化，用传统方法来进行关键点检测就可以达到不错的效果，以及图像本身就是对齐的，使用一个单应性矩阵就可以很好地描述两幅图之间的映射关系。在我们实验室拍摄的测试视频上进行的定性实验上，SuperPoint、多单应性矩阵以及缓冲区都使变化检测效果变得更好，SuperPoint 使关键点覆盖的范围变得更大，匹配成功的关键点变得更多，多单应性矩阵使我们的算法能够应对图像中出现运动物体场景下的单应性估计，缓冲区减少了变化检测的误检和漏检。上述实验结果也显示了多单应性矩阵大幅缩小了传统关键点检测方法和深度学习方法在配准性能上的差距，使我们可以根据设备的算力和任务的实时性需求来灵活选择不同的关键点检测方法。

对于 VFC 方法，我们进行了定性实验的对比，如下

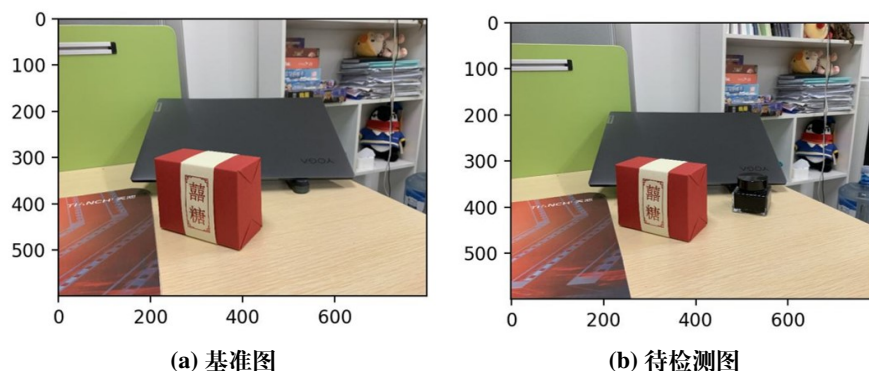


图 6-1 基准图

图 6-1a 是基准图，图 6-1b 是含有变化的待检测图，从图中可以看出变化区域是墨水瓶。示例是一种非常典型且有挑战性的情况，图片中存在大片的重复元素，各种材料的纹理存在相似的地方，只有少部分区域特征明显。我们分别对不使用 VFC 和使用 VFC 的情况进行对比，并且每种情况都采用两种方法进行单应性矩阵估计：RANSAC 和 LMDES，区分对应区域的是否变化的阈值分别设置 0.005 和 0.01，图中的红色矩形框表示算法得到的变化区域。我们可以看到，对于不同的单应性估计方法，VFC 算法都有效改进了变化框的效果。对于不适用 VFC 算法的情况，当阈值设置比较小时，由于错配等原因，图 6-2a 中会出现一些不改定义为变化区域的框，当阈值较大时图 6-2b 甚至本该标记为异常的区域都不显示了，相比图 6-2c、图 6-2c 则矩形框效果不好。当使用 VFC 算法时，由于算法过滤了大量的错配关键点。对于图 6-3 不管使用何种单应性矩阵估计算法和采用的两种阈值，都能有不错的变化检测效果。

6.2.2 开集识别模块测试

本节进行开集识别检测模块的测试，开集识别会检测出未知类的同时，对已知类进行分类。最近开集识别得相关研究工作都基本遵循 Neal 等人^[130]的设置。将包含多个类别的分类数据集通过随机选取 k 个类作为已知类，其他类别作为未知类，将数据集分成两个不相交的子集，用未知类别样本所形成的空间模拟开集识别场景下的开放空间。数据集的切分可以明显地影响开集识别实验的结果，并且骨干模型的学习能力也和深度开集识别网络的表现正相关。所以为了公平比较，我们使用和 ARPL^[72]工作中一样的骨干网络和数据集切分。

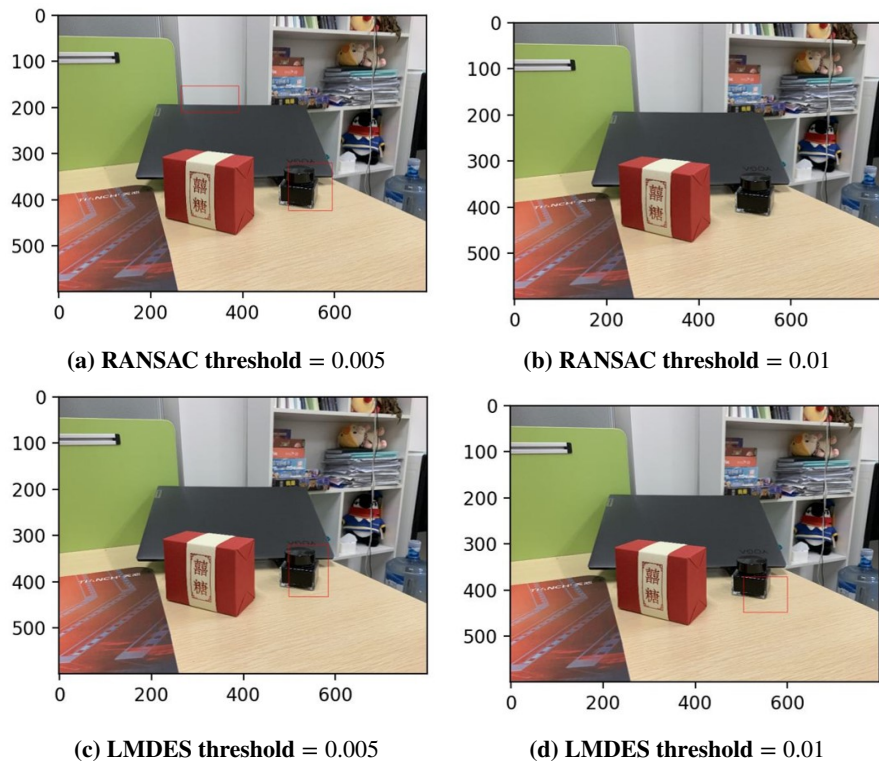


图 6-2 基准方法变化检测效果示意

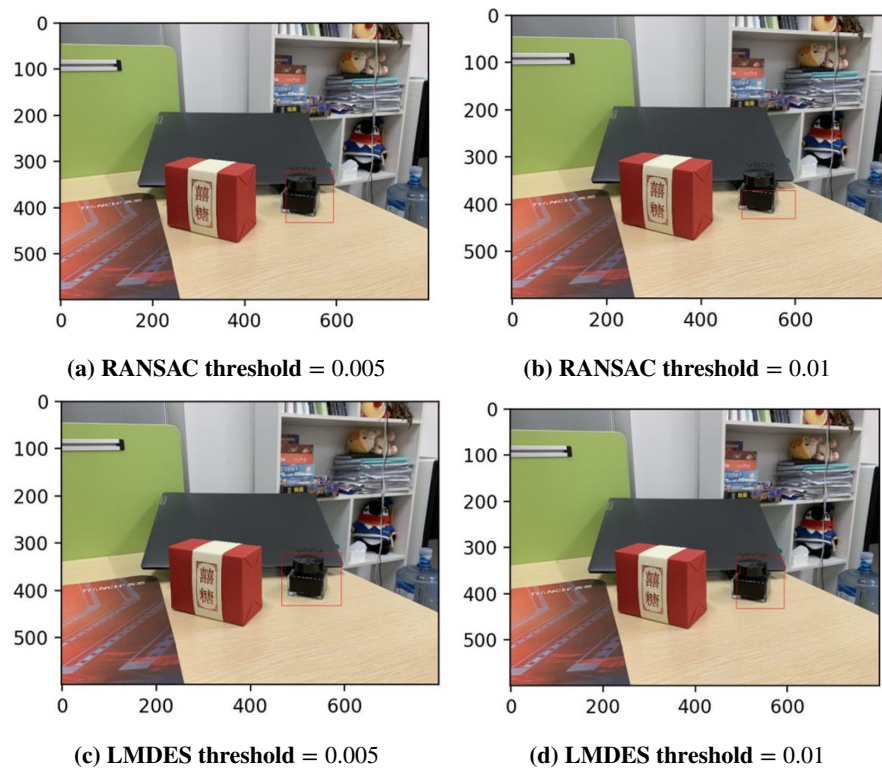


图 6-3 使用 VFC 算法变化检测效果示意

下面介绍实验所使用的基准数据集，以及所使用的设置详细信息。

我们使用 MNIST^[131]、SVHN^[132]、CIFAR-10^[133]、Tiny-Imagenet^[134]作为本测试的公开数据集。MNIST 是一个大规模的手写数字集合。它包含一个训练集，共有 60,000 个样本，以及一个测试集，共有 10,000 个样本，数字已经进行了大小归一化并居中在一个固定大小的图像中。SVHN 是一个数字分类基准数据集，包含了 60 万个 32×32 的 RGB 图像，这些图像是从房屋号牌的照片中裁剪出来的数字（从 0 到 9）。这些裁剪后的图像以数字为中心，但周围的其他数字和干扰物仍然保留在图像中。SVHN 包含三个数据集：训练集、测试集和一个包含 53 万张图像的额外数据集。Tiny-Imagenet 包含了 200 个类别，共 100000 张图片（每个类别 500 张），这些图片被缩小为 64×64 像素的彩色图像。每个类别都有 500 张训练图片、50 张验证图片和 50 张测试图片。CIFAR-10 是 Tiny-Imagenet 数据集的一个子集，由 60000 个 32x32 彩色图像组成。这些图像被标记为 10 个互斥的类别之一，每个类别有 6000 张图像，每个类别分别有 5000 张训练图像和 1000 张测试图像。

具体的实验设置如下：

- **MNIST \ SVHN \ CIFAR-10:** 这些数据集集中有 10 个类别，将其中的 6 个类别作为已知类别，剩余 4 类作为未知类别。
- **CIFAR+10 \ CIFAR+50:** 从 CIFAR-10 中选择 4 类作为已知类别，从 CIFAR-100 中选择 10/50 张图片作为未知类别。
- **Tiny-Imagenet** Tiny-Imagenet 有 200 种类别，我们选择其中的 20 个类别作为已知类别，剩余的 180 个类别作为未知类别。

在每个基准数据集上，我们采用和 Chen 等人^[72]工作中一样的数据集划分方式，并进行实验，实验结果取平均值。我们使用 AUROC (Area Under the ROC) 作为评估指标。AUROC 是一种阈值无关的度量指标，可以解释成正样本比负样本被模型指派更高的检测得分的概率^[135]。开集识别实验的复杂度称之为开放度，定义为开放度 $1 - \sqrt{K/M}$ ，其中 K 和 M 分别表示训练集合和测试集合的类别数目。在公开数据集上，本文将 ConSim-CNN 分别于各种基线模型进行比较，包括：Softmax Thresholding^[81]、OpenMax^[78]、OSRCI^[130]、C2AE^[79]、RPL++^[136]、GFROSR^[80]、PROSER^[73]和 ARPL^[72]

表 6-7 开集识别结果

数据集	MNIST	SVHN	CIFAR-10	CIFAR+10	CIFAR+50	Tiny-ImageNet
开放程度	22.54%	22.54%	22.54%	46.55 %	72.78%	68.38%
Softmax	97.8	88.6	67.7	81.6	80.5	57.7
OpenMax	98.1	89.4	69.5	81.7	79.6	57.6
G-OpenMax	98.4	89.6	67.5	82.7	81.9	58.0
OSRCI	98.9	91.0	69.9	83.8	82.7	58.6
C2AE	98.9	89.2	71.1	81.0	80.3	58.1
RPL++	99.3	95.1	86.1	85.6	85.0	70.2
GFROSR	N.R	93.5	80.7	92.8	92.6	60.8
PROSER	N.R	94.3	89.1	96.0	95.3	69.3
ARPL	99.7	96.7	91.0	97.1	95.1	78.2
ConSim-CNN	99.6	97.8	92.9	98.1	96.3	79.7

表 6-8 开集识别: 消融实验

Method	AUROC	UnKnown ACC	Known ACC	ACC
cls	73.8	88.0	46.3	83.8
cls + con	76.5	91.5	44.2	86.8
soft	76.9	95.4	35.5	89.4
soft + mixup	78.5	92	47.1	87.5
soft + mixup + two aug	79.5	95.7	37.4	89.9
corr	80.1	96.0	37.8	90.2

表 6-7 是公开数据集上的实验结果。基准模型的表现数据引自 Chen 等人^[72], Zhou 等人^[73]的工作。N.R 表示原论文没有报告相关的实验结果。从表 6-7, 我们可以看大盘, 大部分的实验在数字图像数据集 MNIST 和 SVHN 上都具有非常好的结果。特别的, 在 MNIST 数据集上的表现, 几乎是饱和了的。尽管如此, 我们的方法仍然在 SVHN 上将 AUROC 指标提升到了 97.8。在真实物体数据集上, 我们的方法比 SOTA 方法 PROSER 和 ARPL 取得了更好的结果。在 Tiny-Imagenet 数据集上, ConSim-CNN 方法比 ARPL 效果提升了 1.4 个百分点。

针对损失函数不同部分的作用, 我们还做了消融实验, 如表 6-8 所示, 其中第一部分比较了只包含分类损失和引入对比损失后模型的性能变化, 可以看出对比损失的引入使得模型在 AUROC 和总体准确度 ACC 以及未知类准确度 UnKnown ACC 上都有接近 3 个点的提高。第二部分, 在第一部分的分类损失和

对比损失的基础上，使用软标签形式的损失，并逐步引入更多的数据增强方法，软标签、mixup 增强以及做两次 RandAug 增强。可以看出软标签给数据增强带来了方便，同时数据增强可以叠加使用，模型都有大约 1 个点的提升。最后，我们还加入了原型向量相关性损失，可以看出，模型性能又提高了大约 0.5 个点，说明原型向量的先验分布对模型训练具有正向作用。这些实验充分说明了损失函数各个部分的有效性和必要性。

为了验证模型学习的特征的有效性，我们还进行了可视化实验，如图 6-4 所示。我们在 Tiny-Imagenet 数据集的一种类别设置上进行实验，为了演示方便，取已知类中的 5 种以及未知类别中的 10 种进行可视化。其中，红色的点表示所有的未知类别，紫色的点表示原型向量，其他的 5 种颜色分别表示 5 种已知类。从图中可以看出，已知类内部，同种类别的样本点之间的间距小，已知类之间，不同种类别之间的样本点之间的间距大。而未知类别，均匀地分布在中间区域，原型向量分布在各自的类别团簇中，且与未知类别远离。这说明，我们的模型学得的特征具有很好的表示能力，既满足了分类已知类的要求，又具有泛化到未见样本的能力。

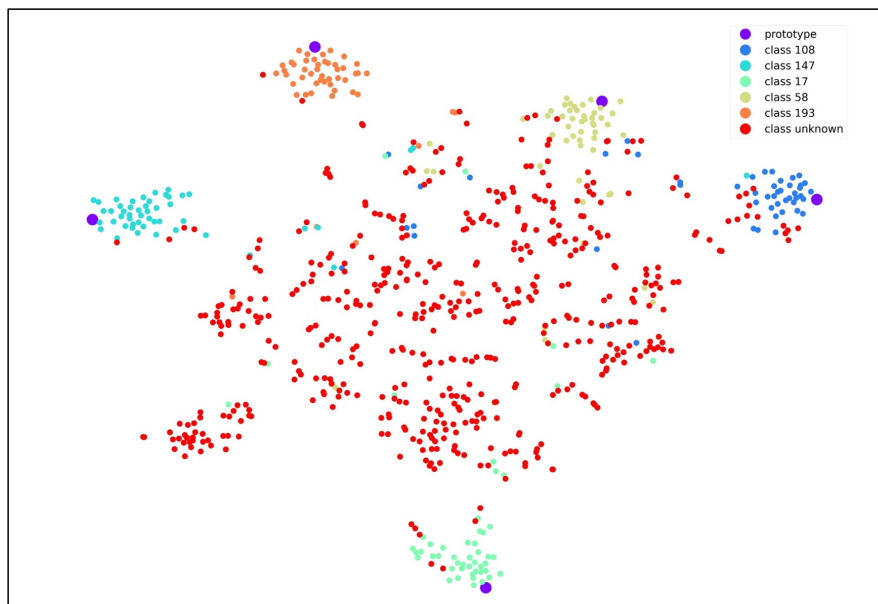


图 6-4 t-sne 可视化

除了在公开数据集上进行测试，我们还在国家电网项目采集的变电站异常检测数据上进行测试。数据集共包括四种异常类别：部件表面油污、高空悬浮物、地面油污、鸟巢。每种异常类别分别有 100 张图片，每张图片有标注信息

表 6-9 异常与非异常判别

	ARPL			ConSim-CNN		
	F1	Precision	Recall	F1	Precision	Recall
异常	79.8	73.2	87.7	82.3	75.6	90.3
正常	71.6	64.1	81.0	72.9	63.7	85.2

表 6-10 异常分类

	ARPL			ConSim-CNN		
	F1	Precision	Recall	F1	Precision	Recall
部件表面油污	79.3	68.7	93.8	79.5	67.3	97.2
高空悬浮物	74.9	72.1	78.0	79.5	78.6	83.4
地面油污	84.1	81.4	86.9	84.5	80.3	89.1
鸟巢	79.2	70.6	92.1	83.2	76.2	91.5

包含异常位置和异常类别。对于非异常类别类别，使用固定大小的矩形随机采样背景图，当矩形与异常区域的交集区域大小除以异常区域大小小于 15%，即认为是合适的非异常类，总共采样 160 张非异常数据，其中 80 张作为训练数据，80 张作为测试数据。在变电站异常数据上，我们测试了 ARPL 和我们的方法 ConSim-CNN 的效果，采用的评价指标是准确率、召回率以及 F1。其中总的异常和正常识别结果采用平均值的方式。结果如表 6-9 和表 6-10 所示。

可以看出在 F1 的指标上 ConSim-CNN 要显著优于 ARPL，但是在正常背景图片上的准确度要略微差一些。在异常图片上也有这样的情况，但总的来说 ConSim-CNN 相对于 ARPL 有不小的提升。

6.3 集成测试

集成测试将各个软件模块组合起来并作为一个整体进行测试，执行集成测试以评估系统和组件是否符合指定的功能需求。本节对系统架构运作进行集成测试，由于系统包含了两个主要阶段：初始化阶段和运行时阶段，这两个阶段的作用各不相同，所以下面将分别就这两个阶段进行系统集成测试。



图 6-5 机器人平台

6.3.1 测试准备

在云服务器端，后端框架 Flask 监听 41401 端口，数据库 MongoDB 监听 41402 端口、gRPC 异常检测开集识别部分服务监听 41403 端口。在边缘设备端，罗技摄像头插在设备的第一号 usb 接口，zmq 视频推流服务在 5100 端口广播视频流，zmq 双端信息通道监听在 5101 端口。桌面客户端登录后，与边缘设备连接的 zmq 双端信息通道服务也同样监听在 5101 端口。

除了上述软件的准备，还有实验场地的准备。为了模拟巡检场景，我们利用教研室作为系统巡检的外部环境，同时我们还使用了机器人平台，如图 6-5 所示。机器人在头有固定摄像头的位置，同时在底部是一个受独立硬件设备树莓派控制的四向移动平台，可以进行前进、后退、向右转和向左转的操作，满足测试的巡检移动要求。机器人中柱体内部有独立电池可以给边缘设备提供供电，也有位置可以固定边缘计算硬件。机器人平台可以进行实现的运行路线设置，自动地完成行走过程。我们设计机器人绕着教室内行走一圈，并设置了 3 个停靠点位，机器人将在停靠点位暂留 10s。同时我们实现了一个查询端口，可以访问到当前机器人是否处在停靠状态以及停靠地点是第几个位置。

6.3.2 初始化阶段测试

系统初始阶段的任务，是将目标检测位置的基准图进行上传。系统采用了两种方式进行基准图的上传，分别是直接上传和在巡检过程中上传。

首先进行直接上传的测试，直接上传需要提前准备好图片，并通过客户端上传文件，同时需要填写图片信息，并进行保存或者修改。客户端需要对用户上传的文件的格式进行检查，对于不符合要求的文件予以拒绝和提示，具体测试结果如表 6-11 所示。

表 6-11 初始化阶段：直接上传基准图片测试

操作	测试结果	预期结果
上传单个图片	通过	客户端提示上传成功、并能查看结果
上传 TXT 文件	通过	客户端提示文件格式错误，请求用户重新上传
编辑图片信息并保存	通过	客户端提示基准图保存成功，并能查看结果
编辑图片部分信息并保存	通过	客户端提示信息不完整，请求继续填写
对已有的基准图更新信息	通过	客户端提示更新成功，并能查看结果

然后进行巡检过程中上传测试，巡检过程中上传需要在客户端打开监视功能，同时机器人平台同步按照规划的路线进行行走，在监视界面，当机器人停靠目标地点时将当前监视界面进行基准图上传，不同于直接上传基准图，巡检过程中因为地点信息是已知，不要进行信息的手动填写。具体测试结果如表 6-12 所示。

表 6-12 初始化阶段：巡检过程上传基准图片测试

操作	测试结果	预期结果
打开监视功能	通过	客户端显示实时监视图像
设备未连接，打开监视	通过	客户端提示设备断开，请求检查
监视下，目标地点上传图片	通过	客户端提示基准图保存成功，并能查看结果
监视下，非目标地点上传图片	通过	客户端提示未抵达目标位置，提示等待
监视下修改上传图片信息	通过	客户端提示更新成功，并能查看结果

6.3.3 运行时阶段测试

对于教研室场景的异常检测，我们从 ImageNet 数据集中选取了四种类别作为训练数据训练模型，分别是玩具、杯子、书本、球，并将这四种类别作为运行

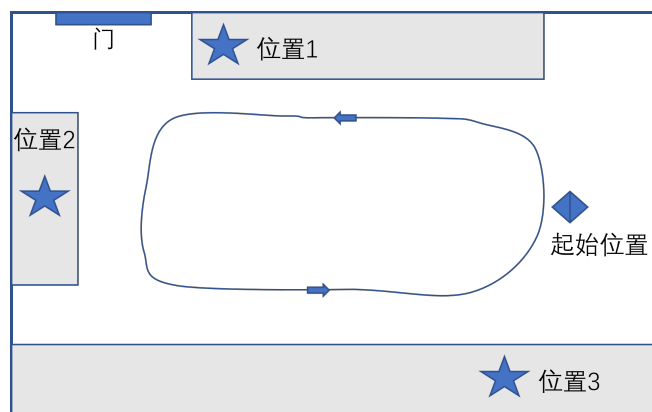
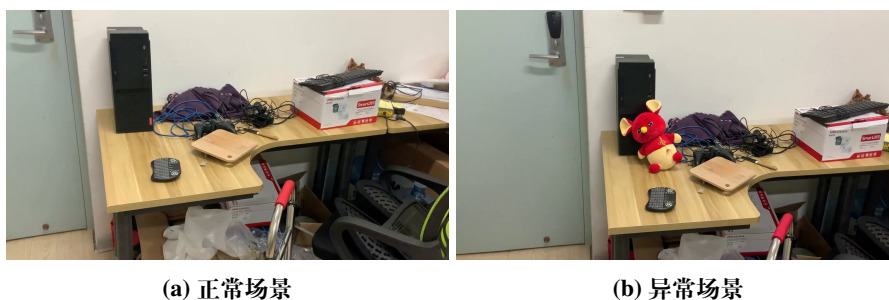


图 6-6 室内场景巡检示意图

时阶段测试的目标异常类别。在机器人平台停靠的三个位置，随机选择两个位置在其中放置目标异常，如图 6-7b所示，异常的类别从四种目标异常中随机选择，剩下的位置设置为正常，保持原样。

室内场景巡检的环境设置如图 6-6所示，图中灰色的区域是办公桌和书架，可以用来放置异常物体，机器人每次固定从房间的右侧开始逆时针进行定速行走，到达设定位置后，停留 10s，完成一轮循环后再回到出发点停止，作为一个完整循环。室内光线充足，各种背景物体的放置没有固定规律，符合一般的巡检场景。

在进行运行时阶段测试之前，需要先按照初始化阶段测试那样，先进行基准图上传，我们直接采用巡检过程中上传的方式。然后再次让机器人平台进行固定路线行走，我们同时在客户端打开监视，并打开检测功能，同时为了测试在光照条件变化的情况的巡检情况，选择关闭室内的一盏灯来模拟光照条件的改变。我们分别对基本功能进行测试，还在改变环境光照下重复和光照条件较好时一样的测试，具体的测试结果如表 6-13所示。



(a) 正常场景

(b) 异常场景

图 6-7 测试场景示例

表 6-13 运行时阶段：监视、检测功能测试

操作	测试结果	预期结果
打开监视功能	通过	客户端显示实时监视图像
检测功能打开，目标位置未到达	通过	客户端提示为达到位置
检测功能打开，达到异常位置	通过	客户端实时检测出变化区域，并正确分辨异常
检测功能打开，达到正常位置	通过	客户端检测视图无变化
检测功能打开，光照变化，达到异常位置	通过	客户端实时检测出变化区域，并正确分辨异常
检测功能打开，光照变化，达到正常位置	通过	客户端检测视图无变化

6.4 非功能性测试

非功能性测试用于测试系统是否满足各项非功能性需求，包括系统是否具有有良好的扩展性、系统数据安全性、是否易受攻击、系统数据吞吐量大小，在高并发请求下的表现等多项非功能性评价标准。下面分别对后端进行负载压力测试，以及对变化检测模块的性能进行测试，以反映系统整体的抗压能力与计算能力。

6.4.1 负载测试

负载测试通过所用多个线程并发地向服务器发送大量请求，能了解服务器的吞吐量、最大请求数等信息。这里使用 Apache JMeter 对服务器进行负载测试，设置线程数为 100，每个线程循环发送请求 10 次，从首个线程到最后一个线程启动的时间间隔为 100 秒，单次请求向服务器 41404 端口发送 HTTP GET 请求，设置跟随重定向、持久连接，并获取所有嵌入资源，根据每次请求的响应时间绘制散点图，并计算平均响应时间。吞吐量等性能指标。个人主机发出的请求将经过若干次跳转后对服务端登录页面进行请求，其测试结果如图 6-8 所示。图中数据显示服务器吞吐量为每分钟 867 次请求，单次请求最低时延约 400ms，高负载时平均时延 2733ms，偏差为 2689ms。这表明系统在用户并发数量少的情况下有较高的响应速度和较低的延迟，这随用户并发数量增多而呈近似线性增长的趋势，即在用户并发数量超过系统所能承受的吞吐量情况下请求具有较低的响应速度和较高的延迟。

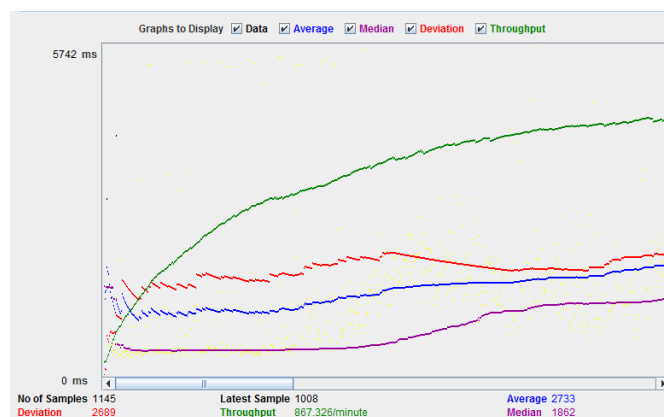


图 6-8 JMeter 负载测试

6.4.2 性能测试

本节对异常检测模块进行性能测试，通过向检测模块发送基准图和测试用实时图，测试算法的执行效率。测试使用 1 个 GPU，并且在模块的各个分支环节设置计时点，通过多轮测试，最后取平均值，测试结果如表 6-14 所示。

表 6-14 异常检测模块性能测试结果

边缘设备数	测试部件	平均耗时
1	前处理	22.1ms
1	前处理 + 变化检测	89.2ms
1	开集识别	48.6 ms
1	前处理 + 变化检测 + 开集识别 + 后处理	142.7ms
2	前处理 + 变化检测 + 开集识别 + 后处理	151.3ms
4	前处理 + 变化检测 + 开集识别 + 后处理	175.5ms

由数据可知，随着边缘设备的增加，模块总体的耗时也随之增多，但是总体增加的量还是比较小的，可以忽略。同时，从数据也可知，模块的一部分耗时消耗在网络上了，所以图片的大小和解码编码非常重要。并且由于系统的优化，使得一部分的处理工作使用并行的方式，模块总体耗时还处在可控的范围。

6.5 运行结果

本节展示系统的运行结果：

图 6-9 是用户的访问登录界面，用户输入用户名和密码实现对功能的授权访问。

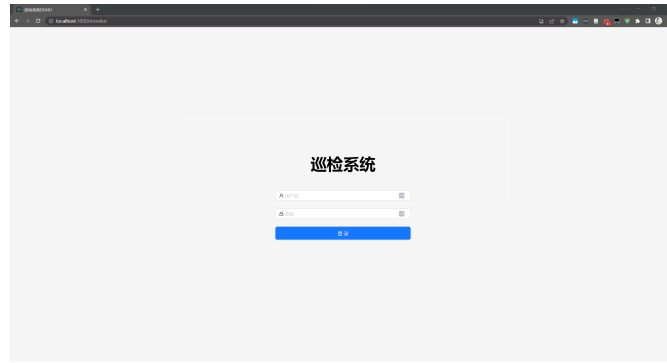


图 6-9 系统登录界面

图 6-10 登录系统后的监视检测界面，用户可以选择需要监视的边缘设备，点击开始监视按钮，可以看到实时影像，以及在目标地点相应的基准图。用户可以在此界面进行更细基准图，也可以开启检测功能，检测框会出现在实时图所在的视图上。

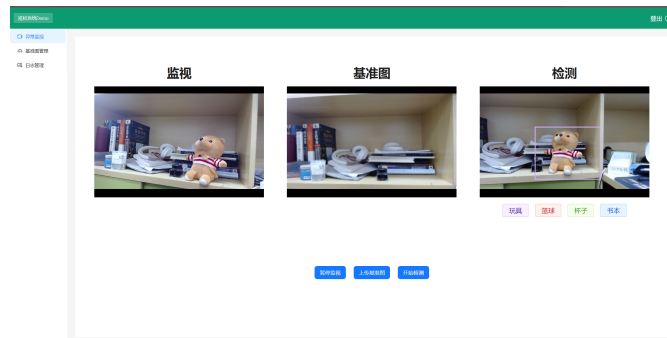


图 6-10 监视界面



图 6-11 图片管理界面

图 6-11 显示通过点击左侧 tab 切换到基准图管理界面，用户在此界面可以看到关于基准图的详细信息，同时可以点选按钮，切换不同的基准图进行查看。除此之外，用户可以上传和备份图片，以及对于图片信息进行修改甚至删除。

图 6-12展示了异常检测的结果示例，图中大致框出了和原图相变化的区域，相对于目标检测而言虽然不是特别精确，但是已经符合其使用目的。并且开集识别模型也能正确识别其类别。

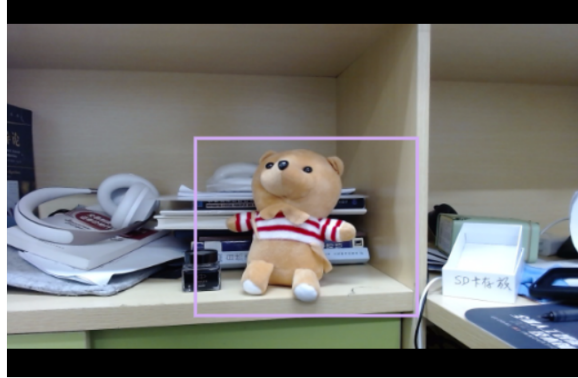


图 6-12 异常检测结果展示

6.6 本章小结

本章对系统分别进行模块测试、接口测试、集成测试、负载测试，系统功能完整，实现了巡检功能，各模块均正常运行，在压力环境下系统可容纳一定的边缘设备同时访问，但是边缘设备过多时，各个设备的平均耗时还是有所增加，说明对于多设备同时使用场景还有优化的空间。

第七章 结语

7.1 全文总结

智能巡检系统是一种基于人工智能技术的自动化检测和监控系统，可以对各种设备、机器和设施进行实时的监测和分析，以及预测和防范可能的故障和问题。智能巡检系统的优点包括减少人力成本、提高效率和准确性、实现更可靠和安全的运营等方面。智能巡检系统通常包括多个传感器和数据采集设备，通过数据的采集和分析，可以及时发现设备的异常情况和故障，减少因设备故障而造成的生产停顿和损失。同时，智能巡检系统还可以实现设备的预测性维护，通过对历史数据的分析和建模，可以预测设备未来的运行状态和故障可能性，从而提前采取维护措施，避免不必要的损失。智能巡检系统在工业、交通、能源等领域中得到广泛应用，可以帮助企业和机构实现智能化运营和管理，提高生产效率和质量，减少人员安全风险，优化资源利用等方面。

然而智能巡检系统的建设和实现需要较高的技术投入和专业知识，除此之外高昂的数据标注和计算设备成本，也阻碍了智能巡检系统的进一步普及和使用。本文提出一种基于对比变化检测的智能巡检系统，可以大大减少对于标注数据的需求。系统通过对比事先采集的标准图片和巡检过程中的实时图片，检测出变化区域，再对变化区域进行识别，从而完成异常检测任务。在系统算法层面，对现有的变化检测方法和识别算法进行改进，以满足巡检场景的业务要求。在软件架构层面，通过边缘、云端、用户端的设计，使得大部分的计算和流量保留在边缘端，降低了成本，提高了效率，算法也与这样的架构设计相适应，做到了算法系统一体优化。最后，在系统测试阶段通过模块测试、集成测试对系统整体功能实现与系统性能进行验证，得出了系统在正常访问情况下系统响应时间满足正常访问需求的结论。

本文的主要工作包括：

1. 提出一种更为鲁棒的变化检测方法，该方法使用 SuperPoint 模型提取对于光照和视角变化更为鲁棒的关键点，再对基准图和目标图的关键点利用快速最近邻方法计算匹配点对，由于匹配点对存在误匹配的情况，使用向量场一致性约束对匹配点对进行过滤。得到匹配点对后，采用多个单应性变换矩阵估计的方法，对图片的采样点进行变化对比，经过处理后得到变化区域，避免了传统方法由于物体不在同一平面而无法共用变换矩阵的问题。

2. 提出一种基于原型学习的开集识别方法，该方法不再使用 Softmax 作为类别预测的输出，而是为每个类别学习多个原型向量，通过计算待识别图片的特征向量和类别的原型向量的相似度以及是否满足阈值，来决定图片是未知类别还是具体的某一已知类别。方法还聚合使用了多个图像增强方法，以及 MixUp 软标签方法，进一步提升了算法的鲁棒性。

3. 设计并实现了完整的智能巡检系统。采用边云结合的空间布局设计，充分利用了边缘设备以及算法特性，降低了成本。系统基于消息队列实现了高效的视频传输服务，具有低延迟和低误差率的特性。系统以 Flask 为后端，PyQt 为前端，结合 MongoDB 数据库和 NginX 反向代理来共同实现功能。

7.2 未来展望

随着生产力的不断提高，工业生产自动化、智能化程度会越来越高，对于设备的巡检维护也将由机器代替人来进行。本文从工业智能巡检任务出发，设计了相应算法和系统，然而本系统还存在一定的局限性，未来工作可以从如下几个方面继续深入：

1. 本系统对于工作环境的光照条件有一定的要求，对于光照不均匀或者光照条件太差的情形下，系统不能很好的对比出变化区域。未来工作可以研究基于特征语义而不是像素特征的图片对比方法，从而能在极端条件下也具有较好的效果。

2. 本系统只对于巡检任务中的异常检测的一方面，即异常目标的出现进行了检测，而对于比如物体移动、物体交换等变化情形，还没有继续深入研究。未来工作可以将多个识别结果结合进行决策考虑，从而完整地判断更为广阔地异常情形。

3. 本系统提出地开集识别方法，在已知类中识别准确度还有一定的提升空间。近年来，对比学习等自监督方法有了很多新进展，模型具有较强的泛化性，未来工作可以考虑引入这些方法进行模型训练。

参考文献

- [1] DONG S, WANG P, ABBAS K. A survey on deep learning and its applications[J/OL]. Computer Science Review, 2021, 40: 100379. <https://www.sciencedirect.com/science/article/pii/S1574013721000198>. DOI: <https://doi.org/10.1016/j.cosrev.2021.100379>.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [3] RADKE R J, ANDRA S, AL-KOFAHI O, et al. Image change detection algorithms: a systematic survey[J]. IEEE transactions on image processing, 2005, 14(3): 294-307.
- [4] RENSINK R A. Change detection[J]. Annual review of psychology, 2002, 53(1): 245-277.
- [5] SHI W, ZHANG M, ZHANG R, et al. Change detection based on artificial intelligence: State-of-the-art and challenges[J]. Remote Sensing, 2020, 12(10): 1688.
- [6] BI H B, LU D, ZHU H H, et al. STA-Net: spatial-temporal attention network for video salient object detection[J]. Applied Intelligence, 2021, 51: 3450-3459.
- [7] VARGHESE A, GUBBIJ, RAMASWAMY A, et al. ChangeNet: A Deep Learning Architecture for Visual Change Detection. LNCS[J]. Springer Verlag. doi, 2019, 10(1): 007.
- [8] PARK J M, JANG J H, YOO S M, et al. ChangeSim: towards end-to-end online scene change detection in industrial indoor environments[C] // 2021 IEEE/RSJ

- International Conference on Intelligent Robots and Systems (IROS). 2021: 8578-8585.
- [9] BENEDEK C, SZIRANYI T. Change Detection in Optical Aerial Images by a Multilayer Conditional Mixed Markov Model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(10): 3416-3430. DOI: 10.1109/TGRS.2009.2022633.
- [10] BOURDIS N, MARRAUD D, SAHBI H. Constrained optical flow for aerial image change detection[C]//2011 IEEE International Geoscience and Remote Sensing Symposium. 2011: 4176-4179. DOI: 10.1109/IGARSS.2011.6050150.
- [11] JI S, WEI S, LU M. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(1): 574-586. DOI: 10.1109/TGRS.2018.2858817.
- [12] SHEN L, LU Y, CHEN H, et al. S2Looking: A Satellite Side-Looking Dataset for Building Change Detection[J/OL]. Remote Sensing, 2021, 13(24). <https://www.mdpi.com/2072-4292/13/24/5094>. DOI: 10.3390/rs13245094.
- [13] PANG C, WU J, DING J, et al. Detecting Building Changes with Off-Nadir Aerial Images[Z]. 2023. arXiv: 2301.10922 [cs.CV].
- [14] WANG Q, YUAN Z, DU Q, et al. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(1): 3-13. DOI: 10.1109/TGRS.2018.2849692.
- [15] GOYETTE N, JODOIN P M, PORIKLI F, et al. Changedetection.net: A new change detection benchmark dataset[C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2012: 1-8. DOI: 10.1109/CVPRW.2012.6238919.
- [16] JST C. Change detection from a street image pair using cnn features and superpixel segmentation[J]., 2015.

- [17] ALCANTARILLA P F, STENT S, ROS G, et al. Street-view change detection with deconvolutional networks[J]. *Autonomous Robots*, 2018, 42: 1301-1322.
- [18] COLLINS R T, LIPTON A J, KANADE T. Introduction to the special section on video surveillance[J]. *IEEE Transactions on pattern analysis and machine intelligence*, 2000, 22(8): 745-746.
- [19] STAUFFER C, GRIMSON W E L. Learning patterns of activity using real-time tracking[J]. *IEEE Transactions on pattern analysis and machine intelligence*, 2000, 22(8): 747-757.
- [20] WREN C R, AZARBAYEJANI A, DARRELL T, et al. Pfinder: Real-time tracking of the human body[J]. *IEEE Transactions on pattern analysis and machine intelligence*, 1997, 19(7): 780-785.
- [21] BRUZZONE L, PRIETO D. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images[J]. *IEEE Transactions on Image Processing*, 2002, 11(4): 452-466. DOI: 10.1109/TIP.2002.999678.
- [22] COLLINS J B, WOODCOCK C E. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal landsat TM data[J/OL]. *Remote Sensing of Environment*, 1996, 56(1): 66-77. <https://www.sciencedirect.com/science/article/pii/0034425795002332>. DOI: [https://doi.org/10.1016/0034-4257\(95\)00233-2](https://doi.org/10.1016/0034-4257(95)00233-2).
- [23] HUERTAS A, NEVATLA R. Detecting changes in aerial views of man-made structures[C]//Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271). 1998: 73-80. DOI: 10.1109/ICCV.1998.710703.
- [24] BOSCH M, HEITZ F, ARMSPACH J P, et al. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution[J/OL]. *NeuroImage*, 2003, 20(2): 643-656. <https://www.sciencedirect.com/science/article/pii/S1053811903004063>. DOI: [https://doi.org/10.1016/S1053-8119\(03\)0406-3](https://doi.org/10.1016/S1053-8119(03)0406-3).

- [25] DUMSKYJ M J, ALDINGTON S J, DORÉ C J, et al. The accurate assessment of changes in retinal vessel diameter using multiple frame electrocardiograph synchronised fundus photography.[J]. *Current eye research*, 1996, 15 6: 625-32.
- [26] LEMIEUX L, WIESHMANN U C, MORAN N F, et al. The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization[J/OL]. *Medical Image Analysis*, 1998, 2(3): 227-242. <https://www.sciencedirect.com/science/article/pii/S1361841598800212>. DOI: [https://doi.org/10.1016/S1361-8415\(98\)80021-2](https://doi.org/10.1016/S1361-8415(98)80021-2).
- [27] REY D, SUBSOL G, DELINGETTE H, et al. Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis[J/OL]. *Medical Image Analysis*, 2002, 6(2): 163-179. <https://www.sciencedirect.com/science/article/pii/S1361841502000567>. DOI: [https://doi.org/10.1016/S1361-8415\(02\)00056-7](https://doi.org/10.1016/S1361-8415(02)00056-7).
- [28] THIRION J P, CALMON G. Deformation Analysis to Detect and Quantify Active Lesions in Three-Dimensional Medical Image Sequences[J/OL]. *IEEE Transactions on Medical Imaging*, 1999, 18(5): 429-441. <https://hal.inria.fr/inria-00615095>.
- [29] LANDIS E, NAGY E, KEANE D, et al. Technique to Measure 3D Work-of-Fracture of Concrete in Compression[J]. *Journal of Engineering Mechanics-asce - J ENG MECH-ASCE*, 1999, 125. DOI: 10.1061/(ASCE)0733-9399(1999)125:6(599).
- [30] NAGY G, ZHANG T, FRANKLIN W, et al. Volume and Surface Area Distributions of Cracks in Concrete[C]//: vol. 2059. 2001: 759-768. DOI: 10.1007/3-540-45129-3_70.
- [31] EDGINGTON D, SALAMY K, RISI M, et al. Automated event detection in underwater video[C]// *Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492)*: vol. 5. 2003: P2749-P2753 Vol.5. DOI: 10.1109/OCEANS.2003.178344.

- [32] LEBART K, TRUCCO E, LANE D. Real-time automatic sea-floor change detection from video[C]// OCEANS 2000 MTS/IEEE Conference and Exhibition. Conference Proceedings (Cat. No.00CH37158): vol. 2. 2000: 1337-1343 vol.2. DOI: 10.1109/OCEANS.2000.881789.
- [33] WHORFF J, GRIFFING L. A video recording and analysis system used to sample intertidal communities[J/OL]. *Journal of Experimental Marine Biology and Ecology*, 1992, 160(1): 1-12. <https://www.sciencedirect.com/science/article/pii/002209819290106K>. DOI: [https://doi.org/10.1016/0022-0981\(92\)90106-K](https://doi.org/10.1016/0022-0981(92)90106-K).
- [34] FANG C Y, CHEN S W, FUH C S. Automatic change detection of driving environments in a vision-based driver assistance system[J]. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 2003, 14: 646-57. DOI: 10.1109/TNN.2003.811353.
- [35] KAN W Y, KROGMEIER J V, DOERSCHUK P C. Model-based vehicle tracking from image sequences with an application to road surveillance[J]. *Optical Engineering*, 1996, 35: 1723-1729. DOI: 10.1117/1.600747.
- [36] ROSIN P L. Thresholding for Change Detection[J/OL]. *Computer Vision and Image Understanding*, 2002, 86(2): 79-95. <https://www.sciencedirect.com/science/article/pii/S1077314202909604>. DOI: <https://doi.org/10.1006/cviu.2002.0960>.
- [37] MAHMOUDZADEH H. Digital Change Detection Using Remotely Sensed Data for Monitoring Green Space Destruction in Tabriz[J]. *Int. J. Environ. Res*, 2007, 1: 35-411735.
- [38] SHAO P, SHI W, HE P, et al. Novel Approach to Unsupervised Change Detection Based on a Robust Semi-Supervised FCM Clustering Algorithm[J/OL]. *Remote Sensing*, 2016, 8(3). <https://www.mdpi.com/2072-4292/8/3/264>. DOI: 10.3390/rs8030264.
- [39] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]// Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001: vol. 1. 2001: I-I.

- [40] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05): vol. 1. 2005: 886-893.
- [41] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [42] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.
- [43] GIRSHICK R. Fast R-CNN[Z]. 2015. arXiv: 1504.08083 [cs.CV].
- [44] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[Z]. 2016. arXiv: 1506.01497 [cs.CV].
- [45] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[G/OL]//Computer Vision – ECCV 2016. Springer International Publishing, 2016: 21-37. https://doi.org/10.1007/978-3-319-46448-0_2. DOI: 10.1007/978-3-319-46448-0_2.
- [46] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[Z]. 2016. arXiv: 1506.02640 [cs.CV].
- [47] SHAHINFAR S, MEEK P, FALZON G, 2020. <https://arxiv.org/abs/2010.08186>. DOI: 10.48550/ARXIV.2010.08186.
- [48] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios[Z]. 2021. arXiv: 2108.11539 [cs.CV].
- [49] LEE C. How many images do you need for object detection?[Z]. <https://changesin.medium.com/how-many-images-do-you-need-for-object-detection-d33185629843>.

- [50] YANG Z, LI Z, JIANG X, et al. Focal and Global Knowledge Distillation for Detectors[Z]. 2022. arXiv: 2111.11837 [cs.CV].
- [51] LOWE G. Sift-the scale invariant feature transform[J]. Int. J, 2004, 2(91-110): 2.
- [52] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: Self-supervised interest point detection and description[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 224-236.
- [53] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[Z]. 2015. arXiv: 1409.1556 [cs.CV].
- [54] CHOY C B, GWAK J, SAVARESE S, et al. Universal Correspondence Network[Z]. 2016. arXiv: 1606.03558 [cs.CV].
- [55] DERPANIS K G. Overview of the RANSAC Algorithm[J]. Image Rochester NY, 2010, 4(1): 2-3.
- [56] MA J, ZHAO J, TIAN J, et al. Robust Point Matching via Vector Field Consensus[J/OL]. IEEE Transactions on Image Processing, 2014, 23(4): 1706-1721 [2022-10-06]. <https://ieeexplore.ieee.org/document/6746218/>. DOI: 10.1109/TIP.2014.2307478.
- [57] GOLUB G H, HANSEN P C, O'LEARY D P. Tikhonov regularization and total least squares[J]. SIAM journal on matrix analysis and applications, 1999, 21(1): 185-194.
- [58] MICCHELLI C A, PONTIL M. On learning vector-valued functions[J]. Neural computation, 2005, 17(1): 177-204.
- [59] DUBROFSKY E. Homography estimation[J]. Diplomová práce. Vancouver: Univerzita Britské Kolumbie, 2009, 5.
- [60] BARTLETT P L, WEGKAMP M H. Classification with a Reject Option using a Hinge Loss[J/OL]. Journal of Machine Learning Research, 2008, 9(59): 1823-1840. <http://jmlr.org/papers/v9/bartlett08a.html>.

- [61] DA Q, YU Y, ZHOU Z H. Learning with Augmented Class by Exploiting Unlabeled Data[J/OL]. Proceedings of the AAAI Conference on Artificial Intelligence, 2014, 28(1) [2023-04-07]. <https://ojs.aaai.org/index.php/AAAI/article/view/8997>. DOI: 10.1609/aaai.v28i1.8997.
- [62] YUAN M, WEGKAMP M. Classification Methods with Reject Option Based on Convex Risk Minimization.[J]. Journal of Machine Learning Research, 2010, 11(1).
- [63] SCHEIRER W J, de REZENDE ROCHA A, SAPKOTA A, et al. Toward open set recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 35(7): 1757-1772.
- [64] SCHEIRER W J, JAIN L P, BOULT T E. Probability models for open set recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(11): 2317-2324.
- [65] JAIN L P, SCHEIRER W J, BOULT T E. Multi-class open set recognition using probability of inclusion[C]//Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13. 2014: 393-409.
- [66] RUDD E M, JAIN L P, SCHEIRER W J, et al. The extreme value machine[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(3): 762-768.
- [67] MENDES JÚNIOR P R, DE SOUZA R M, WERNECK R D O, et al. Nearest neighbors distance ratio open-set classifier[J]. Machine Learning, 2017, 106(3): 359-386.
- [68] ZHANG H, PATEL V M. Sparse representation-based open set recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(8): 1690-1696.

- [69] BENDALE A, BOULT T E. Towards open set deep networks[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1563-1572.
- [70] SHU L, XU H, LIU B. Doc: Deep open classification of text documents[J]. ArXiv preprint arXiv:1709.08716, 2017.
- [71] YOSHIHASHI R, SHAO W, KAWAKAMI R, et al. Classification-reconstruction learning for open-set recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4016-4025.
- [72] CHEN G, PENG P, WANG X, et al. Adversarial Reciprocal Points Learning for Open Set Recognition[EB/OL]. 2021 [2023-03-29]. <http://arxiv.org/abs/2103.00953>.
- [73] ZHOU D W, YE H J, ZHAN D C. Learning placeholders for open-set recognition[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 4401-4410.
- [74] GUO Y, CAMPORESE G, YANG W, et al. Conditional variational capsule network for open set recognition[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 103-111.
- [75] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [76] SUN X, YANG Z, ZHANG C, et al. Conditional gaussian distribution learning for open set recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 13480-13489.
- [77] ZHANG H, LI A, GUO J, et al. Hybrid models for open set recognition[C] // Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. 2020: 102-117.
- [78] GE Z, DEMYANOV S, CHEN Z, et al. Generative openmax for multi-class open set classification[J]. ArXiv preprint arXiv:1707.07418, 2017.

- [79] OZA P, PATEL V M. C2ae: Class conditioned auto-encoder for open-set recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2307-2316.
- [80] PERERA P, MORARIU V I, JAIN R, et al. Generative-discriminative feature representations for open-set recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11814-11823.
- [81] HENDRYCKS D, GIMPEL K. A baseline for detecting misclassified and out-of-distribution examples in neural networks[J]. ArXiv preprint arXiv:1610.02136, 2016.
- [82] ROZSA A, GÜNTHER M, BOULT T E. Adversarial robustness: Softmax versus openmax[J]. ArXiv preprint arXiv:1708.01697, 2017.
- [83] PERERA P, NALLAPATI R, XIANG B. Ocgan: One-class novelty detection using gans with constrained latent representations[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2898-2906.
- [84] LIANG S, LI Y, SRIKANT R. Enhancing the reliability of out-of-distribution image detection in neural networks[J]. ArXiv preprint arXiv:1706.02690, 2017.
- [85] LEE K, LEE H, LEE K, et al. Training confidence-calibrated classifiers for detecting out-of-distribution samples[J]. ArXiv preprint arXiv:1711.09325, 2017.
- [86] HENDRYCKS D, MAZEIKA M, DIETTERICH T. Deep anomaly detection with outlier exposure[J]. ArXiv preprint arXiv:1812.04606, 2018.
- [87] NGUYEN A, YOSINSKI J, CLUNE J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 427-436.
- [88] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent-a new approach to self-supervised learning[J]. Advances in neural information processing systems, 2020, 33: 21271-21284.

- [89] GUO G, WANG H, BELL D, et al. KNN model-based approach in classification[C]//On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings. 2003: 986-996.
- [90] KUNCHEVA L I, BEZDEK J C. Nearest prototype classification: Clustering, genetic algorithms, or random search?[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 1998, 28(1): 160-164.
- [91] SATO A, YAMADA K. A formulation of learning vector quantization using a new misclassification measure[C]//Proceedings. Fourteenth international conference on pattern recognition (Cat. No. 98EX170): vol. 1. 1998: 322-325.
- [92] BONILLA E, ROBLES-KELLY A. Discriminative probabilistic prototype learning[J]. ArXiv preprint arXiv:1206.4686, 2012.
- [93] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14. 2016: 499-515.
- [94] YANG H M, ZHANG X Y, YIN F, et al. Robust classification with convolutional prototype learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3474-3482.
- [95] YANG H M, ZHANG X Y, YIN F, et al. Convolutional prototype network for open set recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(5): 2358-2370.
- [96] OORD A V D, LI Y, VINYALS O. Representation learning with contrastive predictive coding[J]. ArXiv preprint arXiv:1807.03748, 2018.
- [97] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9729-9738.

- [98] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C] // International conference on machine learning. 2020: 1597-1607.
- [99] CHEN X, HE K. Exploring simple siamese representation learning[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 15750-15758.
- [100] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[J]. ArXiv preprint arXiv:1703.07737, 2017.
- [101] SOHN K. Improved deep metric learning with multi-class n-pair loss objective[J]. Advances in neural information processing systems, 2016, 29.
- [102] DUAN Y, ZHENG W, LIN X, et al. Deep adversarial metric learning[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2780-2789.
- [103] OH SONG H, XIANG Y, JEGELKA S, et al. Deep metric learning via lifted structured feature embedding[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4004-4012.
- [104] MAO C, ZHONG Z, YANG J, et al. Metric learning for adversarial robustness[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [105] LI J, ZHOU P, XIONG C, et al. Prototypical contrastive learning of unsupervised representations[J]. ArXiv preprint arXiv:2005.04966, 2020.
- [106] KHOSLA P, TETERWAK P, WANG C, et al. Supervised Contrastive Learning[EB/OL]. arXiv. 2021 [2023-04-01]. <http://arxiv.org/abs/2004.11362>.
- [107] HUITEMA C. Real time control protocol (RTCP) attribute in session description protocol (SDP)[R]. 2003.
- [108] HINTJENS P. ZeroMQ: messaging for many applications[M]. " O'Reilly Media, Inc.", 2013.
- [109] D G. Comparative study on Python web frameworks: Flask and Django[J][M]. 2020.

- [110] IDRIS N. A Generic Review of Web Technology: Django and Flask[J]., 2020, 2(1).
- [111] GYŐRÖDI C, GYŐRÖDI R, PECHERLE G, et al. A comparative study: MongoDB vs. MySQL[C]//2015 13th International Conference on Engineering of Modern Electric Systems (EMES). 2015: 1-6. DOI: 10.1109/EMES.2015.7158433.
- [112] ZHANG J, WANG C, LIU S, et al. Content-Aware Unsupervised Deep Homography Estimation[EB/OL]. arXiv. 2019. <https://arxiv.org/abs/1909.05983>.
- [113] YANG H M, ZHANG X Y, YIN F, et al. Robust Classification with Convolutional Prototype Learning[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2018: 3474-3482. arXiv: 1805.03438. DOI: 10.1109/CVPR.2018.00366.
- [114] SU B, LU S, TAN C L. Blurred image region detection and classification[C]//Proceedings of the 19th ACM international conference on Multimedia. 2011: 1397-1400.
- [115] SPRAWLS P. Physical principles of medical imaging[J]., 1993.
- [116] BOVIK A C. Handbook of image and video processing[M]. Academic press, 2010.
- [117] YANG F W, LIN H J, CHUANG H. Image deblurring[C]//2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CB-DCOM/IOP/SCI). 2017: 1-4. DOI: 10.1109/UIC-ATC.2017.8397472.
- [118] DHARAVATH K, AMARNATH G, TALUKDAR F A, et al. Impact of image preprocessing on face recognition: A comparative analysis[C]//2014 International Conference on Communication and Signal Processing. 2014: 631-635.
- [119] PERTUZ S, PUIG D, GARCIA M A. Analysis of focus measure operators for shape-from-focus[J]. Pattern Recognition, 2013, 46(5): 1415-1432.

- [120] TONG H, LI M, ZHANG H, et al. Blur detection for digital images using wavelet transform[C]//2004 IEEE international conference on multimedia and expo (ICME)(IEEE Cat. No. 04TH8763): vol. 1. 2004: 17-20.
- [121] PECH-PACHECO J L, CRISTÓBAL G, CHAMORRO-MARTINEZ J, et al. Diatom autofocusing in brightfield microscopy: a comparative study[C]//Proceedings 15th International Conference on Pattern Recognition. ICPR-2000: vol. 3. 2000: 314-317.
- [122] TANG C, HOU C, SONG Z. Defocus map estimation from a single image via spectrum contrast[J]. Optics letters, 2013, 38(10): 1706-1708.
- [123] HAN W, KHORRAMI P, PAINE T L, et al. Seq-nms for video object detection[J]. ArXiv preprint arXiv:1602.08465, 2016.
- [124] BELHASSEN H, ZHANG H, FRESSE V, et al. Improving Video Object Detection by Seq-Bbox Matching.[C]//VISIGRAPP (5: VISAPP). 2019: 226-233.
- [125] SABATER A, MONTESANO L, MURILLO A C. Robust and efficient post-processing for video object detection[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2020: 10536-10542.
- [126] GACKENHEIMER C. Introduction to React[M]. Apress, 2015.
- [127] IZZO D, BISCANI F. Darioizzo/audi: Pybind11[J]. Zenodo,
- [128] RIFKIN R, YEO G, POGGIO T, et al. Regularized least-squares classification[J]. Nato Science Series Sub Series III Computer and Systems Sciences, 2003, 190: 131-154.
- [129] Huawei-noah. Efficient-AI-Backbones[Z]. <https://github.com/huawei-noah/Efficient-AI-Backbones>. 2022.
- [130] NEAL L, OLSON M, FERN X, et al. Open set learning with counterfactual images[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 613-628.
- [131] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

-
- [132] NETZER Y, WANG T, COATES A, et al. Reading digits in natural images with unsupervised feature learning[J]., 2011.
- [133] KRIZHEVSKY A, HINTON G, et al. Learning multiple layers of features from tiny images[J]., 2009.
- [134] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115: 211-252.
- [135] GENG C, HUANG S J, CHEN S. Recent advances in open set recognition: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 43(10): 3614-3631.
- [136] CHEN G, QIAO L, SHI Y, et al. Learning open set network with discriminative reciprocal points[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. 2020: 507-522.

致 谢

在我完成本篇论文之际，我想向许多人表达我的感激之情。他们的支持和帮助是我完成这项研究的关键。因此，在此我要向以下人员致以衷心的感谢：

首先，我要感谢我的导师申富饶教授。他不仅在学术上给予了我无私的指导和帮助，还在生活上给予了我鼓励和支持。在整个研究过程中，他提供了许多宝贵的意见和建议，不断地鼓励我，帮助我克服各种困难。他的指导让我更加深入地了解研究领域，同时也提高了我的学术水平。

其次，我要感谢实验室的所有成员，感谢他们在我实验和数据处理过程中的帮助和支持。在这个大家庭中，我得到了许多有益的交流 and 启示，这对我的研究起到了至关重要的作用。

我还要感谢我的同学们和朋友们。在整个研究过程中，他们不仅与我分享了自己的知识和经验，还给予了我许多鼓励和支持。在我疲惫时，他们为我打气，让我重新振作起来。在此，我要感谢他们一直以来的支持和鼓励。

最后，我要感谢我的家人。在我求学过程中，他们一直支持我，鼓励我，帮助我克服各种困难。他们的无私奉献和理解支持，让我在学术上更加专注，也更有信心面对未来的挑战。

在此，我要再次感谢所有帮助和支持我的人。这篇论文的完成离不开你们的支持和鼓励。我会一直怀着感激之心，回报社会，回馈家人和朋友们对我的支持和帮助。

简历与科研成果

基本信息

陈昊，男，汉族，1995年3月出生，江苏省兴化人。

教育背景

2020年9月—2023年6月 南京大学计算机科学与技术系 硕士

2013年9月—2017年6月 南京大学物理学院 本科

攻读硕士学位期间的发明专利

1. 陈昊, 杨锁荣, 杨洪朝, 卢保金, 张凌茗, 刘佩涵, 李若彤, 《一种基于对比检测的机器人巡检系统》(202310111632.4)

攻读硕士学位期间参与的科研课题

1. 国家电网“基于多维巡检影像匹配和对比技术的变电设备缺陷分析技术研究”(课题年限2019年1月—2022年12月), 负责神经网络模型相关研究。

学位论文出版授权书

本人完全同意《中国优秀博硕士学位论文全文数据库出版章程》（以下简称“章程”），愿意将本人的学位论文提交“中国学术期刊（光盘版）电子杂志社”在《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》中全文发表。《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》可以以电子、网络及其他数字媒体形式公开出版，并同意编入《中国知识资源总库》，在《中国博硕士学位论文评价数据库》中使用和在互联网上传播，同意按“章程”规定享受相关权益。

作者签名：_____

_____年__月__日

论文题名	基于变化对比检测的巡检系统设计与实现				
研究生学号	MG20330007	所在院系	计算机科学与技术系	学位年度	2023
论文级别	<input checked="" type="checkbox"/> 学术学位硕士 <input type="checkbox"/> 专业学位硕士 <input type="checkbox"/> 学术学位博士 <input type="checkbox"/> 专业学位博士				
作者 Email	harveechen@smail.nju.edu.cn				
导师姓名	申富饶				

论文涉密情况：

不保密

保密，保密期（_____年__月__日至_____年__月__日）

