



# HaMonitor Sentry

## 高层智能监测系统

第二十六届中国机器人及人工智能大赛

# HaMonitorSentry-高层智能监测系统

## 项目研究报告- L<sup>A</sup>T<sub>E</sub>X 版

作者：程子洋 & 冉丰源 & 陆一杰  
组织：武汉大学  
时间：June 9, 2024  
版本：1.0  
组别：人工智能创新赛



# 目录

<b>第 1 章 作品概述</b>	<b>1</b>
1.1 作品简介	1
1.1.1 作品概述	1
1.1.2 功能设计	2
1.1.3 目标用户	4
1.2 背景分析	5
1.2.1 国家政策背景	5
1.2.2 社会背景	7
1.3 特色描写	9
1.3.1 高空抛物检测与高层危险行为原创数据集	9
1.3.2 类间距离蒸馏的语义分割算法	9
1.3.3 融合 MA 模块以增强前景运动特征检测	10
1.3.4 双流网络结构	10
1.4 创新性说明	11
<b>第 2 章 需求分析</b>	<b>12</b>
2.1 现有解决方案	12
2.1.1 高空抛物检测	12
2.1.2 高层危险行为检测	14
2.2 现有产品的不足之处	15
2.2.1 现有产品内蕴算法准确性与鲁棒性不强	15
2.2.2 现有产品实时监测效果不佳	16
2.2.3 现有的平台承载能力弱	16
2.2.4 我国安防普及率较低	16
2.3 解决问题的思路	17
2.3.1 小目标跟踪及复杂环境适应性强化	18
2.3.2 提升监测系统的灵敏度与准确性	18
2.3.3 构建高空抛物与高层危险行为的专用数据集	18
<b>第 3 章 技术方案</b>	<b>20</b>
3.1 技术路线总览	20
3.2 高空抛/坠物检测核心算法	21
3.2.1 背景提取	21
3.2.2 小目标运动检测	22
3.2.3 数据去噪与膨化	24
3.2.4 目标跟踪	25
3.3 高层危险行为检测与识别核心算法	25
3.3.1 提取时序特征和空间特征	25
3.3.2 特征融合	26
3.3.3 建模时空特征	27
3.4 监测数据收集与处理	28
3.5 训练与测试	29

<b>第4章 系统实现</b>	<b>30</b>
4.1 数据集搭建	30
4.1.1 概述	30
4.1.2 高空抛物数据集	31
4.1.2.1 数据集的属性	31
4.1.2.2 数据收集与设备配置	31
4.1.2.3 数据预处理	32
4.1.2.4 数据集的划分和统计	33
4.1.3 高层危险行为数据集	35
4.1.3.1 数据集的基本情况	35
4.1.3.2 数据集获取方式	36
4.1.3.3 数据预处理	37
4.2 系统部署方法	38
4.2.1 系统总体框架	38
4.2.2 软件开发环境	39
4.2.3 硬件部署方案	39
4.3 多功能集成、高度迁移性的平台综合管理	39
4.3.1 软件界面设计	39
4.3.2 实时视频检测模块	40
4.3.3 历史数据统计模块	41
4.3.4 视频智能分析模块	42
<b>第5章 作品测试与分析</b>	<b>43</b>
5.1 性能评估指标	43
5.1.1 检出率 (Detection Rate)	43
5.1.2 误报率 (False Alarm Rate)	43
5.2 高空抛/坠“物”检测	43
5.3 高层建筑边缘侧“人”的危险行为检测	44
<b>第6章 作品总结与展望</b>	<b>45</b>
6.1 作品特色与创新点	45
6.1.1 高度集成的算法与功能	45
6.1.2 高层安全双重监测技术	45
6.1.3 高质量的原创数据集	45
6.2 成果支撑	46
6.3 应用推广	47
6.3.1 深圳英飞拓智能技术有限公司合作落地试点	47
6.3.2 应用方向	47
6.3.3 应用效益分析	47
6.4 作品展望	48
<b>参考文献</b>	<b>49</b>

# 第1章 作品概述

## 内容提要

- ❑ 作品简介
- ❑ 背景分析
- ❑ 特色描述
- ❑ 创新性说明

## 1.1 作品简介

### 1.1.1 作品概述

安防行业是构建立体化社会治安防控体系、维护国家安全及社会稳定重要的安全保障性行业。随着光电信息技术、微电子技术、微计算机技术与视频图像处理技术等的发展，传统的安防系统正由数字化、网络化，而逐步走向智能化。同时，我们关注到近年来高层安全事故频发，公民的生命财产安全愈加难以得到保障；我们竞赛小组从“智能化”解决当今高层安全问题的思路出发，经过一系列探索和钻研，最终成功研发出了一款名为 HaMonitorSentry (High Altitude Monitoring Sentry, 高层监测哨兵) 的高层智能监测系统，其立足于国家新兴技术赋能行业\* (如智慧安防)，应用于各类高空安全场景 (如高空抛物监测、高层安全隐患行为监测)，并与物业、企业、政府合作，最终致力以保障公民的生命财产安全。

本作品是一体化的高层智能监测系统，主要结合超清监控摄像头，基于事件中人和人、人和物之间的关系，对高层场景中发生的意外事件，进行**针对性、高精度、低误报**的检测。目前本产品集成了两套针对高层事故的算法：**复杂状态下高空抛/坠“物”检测系统**，以及**高层建筑物边缘侧“人”的危险行为检测与识别系统**。

本着产学研结合的理念，我们与英飞拓-武汉大学智慧城市时空信息技术研究中心展开科创合作，共同推动作品发展。同时，我们与深圳英飞拓智能技术有限公司合作以寻求实际落地应用，共同构建一体化高层智能监测系统，主要设备包含高空智能摄像机、网络视频录像机 (NVR) 和综合管理平台。目前，该作品部分功能已经在深圳湾超级总部基地的实地场景中部署和测试，并筹备进行进一步的落地和推广。

我们的作品亮点从教育维度、创新维度、团队维度、社会价值、应用维度 5 个角度展开，具体总结如下：



图 1.1: 作品概况

### 1.1.2 功能设计

HaMonitorSentry 高层智能监测系统专为各类人群密集场所设计，适用于景点、小区、科技园区等高层建筑区域。系统通过高效率摄像头实时监控，一旦侦测到投掷物或其他危险行为，便能迅速、精准地锁定事故地点，并立即启动报警程序，有效地维护了公共安全及财产保护。系统采用的高空抛物和高层危险行为监测算法具有高度鲁棒性，即使在光照波动、背景噪音和空中遮挡等不利条件下，也能保持低误报率，稳定运行。

本团队所开发作品的最显著特点是双角度监测、一体化通讯与数据治理交叉融合。通过高空抛物和高层危险行为的监测，实现在事件发生前，根据监测事件的危险程度触发警报。及时预警，相关人员就可以快速做出反应，有效阻止悲剧的发生。在事件发生后，可以对相关事件进行回溯取证，有效应对高空抛物的三难问题。即作品重点功能可概括为先预警，后追责。此外，我们的作品还提供信息建模处理功能。

其中本系统中高空抛物监测功能模块构造如图 1.1 所示：

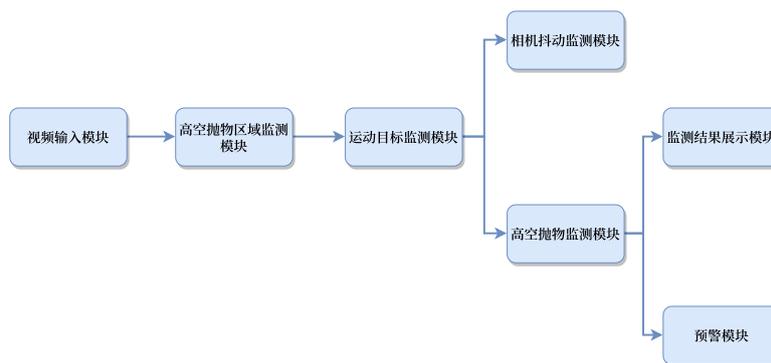


图 1.2: 高空抛物监测功能模块

其中本系统中高层危险行为监测功能模块构造如图 1.3 所示：

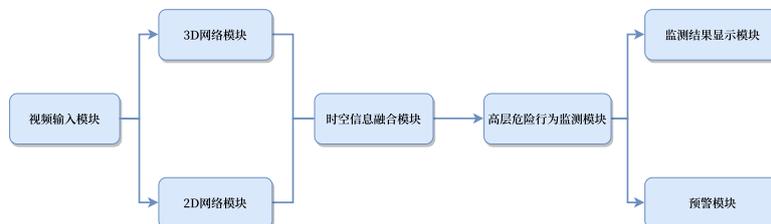


图 1.3: 高层危险行为监测功能模块

其中具体的模块介绍将在第 4 节中做进一步详细的阐述。基于本作品集成的监测系统模块的实现，本作品的应用功能可概括如下：

#### (1) 高空抛物监测与预警

现如今，越来越多的小区建设为高层小区，与此同时带来了频发的各类高层安全问题，写字楼和商场多为高楼和超高楼，进出人员复杂繁多，其悬空的建筑构造导致安全隐患颇多，通过本系统，我们可以对高空抛物行为进行判断监测并预警，以最大程度减少人民生命财产损失。

依靠本系统集成的高鲁棒性，准确性高空抛物监测算法，可以满足高空抛物的即时性监测要求，同时在抛物发生的最短时间内进行预警，警示可能抛物范围中的民众，尽可能的保护抛物范围内的人民生命财产安全。

#### (2) 高层危险行为监测与预警

与高空抛物发生的背景类似，如对企业员工在工作中可能存在的高风险行为进行实时监测和分析，及时发现和预警潜在的危险行为，帮助企业及时采取措施防范事故的发生。同时可以针对一般性的高空危险行为，如跳楼等轻生事件或无意识坠楼等。

以景区为例，景区包括山区、动物园等，山区通常大而广，人工对危险区域进行安全监管耗时耗力；动物园内有野兽区域，多设置下沉式，需要监测游客的不文明投喂行为，以及防止游客不慎坠入。在本系统的帮助下，

景区物业可以及时有效地对各类危险事件做出预警。

应用于建筑工地、高楼清洁等定制化场景，定制实现对高层工作人员的监督，及时发现违反企业安全规章的行为并做出预警，以协助安全监管工作。

本作品能够发现可能存在的风险因素，及时识别出危险行为的发生，并进行预防和干预，从而减少事故和安全风险。此外本作品集成高空抛物，高层危险行为的监测与预警两大功能，弥补了市场空缺，能够考虑高层安全的一体化维护。

### **(3) 高层危险事故的事后追责**

现阶段高层危险事故的事后追责问题是当今安全维护的难题之一，事故发生后，如何精确到个人的责任追究，如何最大程度上实现责任公平化，本作品的高精确度高可信度算法为该问题的解决提供了成熟的方案。

本作品记录高层危险事故过程中的各种参数数据，如抛物起点，落点信息，危险行为发生时间等，这些数据可以作为证据，帮助调查人员了解事故发生的原因和过程。即可以对已经发生恶劣影响的高空抛物行为做出辅助的追责工作，对即将发生的高层危险行为及时做出判断，第一时间提供救援，具有极强的社会意义。

### **(4) 高层安全监管的决策规划**

本作品在综合以上功能的基础上，还重点涵盖了数据分析功能，针对已经监测到的高层安全事故，我们会重点分析其发生时间，发生地点，当日环境数据，抛物类型，危险行为种类等关键信息，用来进行进一步的数据建模与分析，从而分析出对应的事故发生特点，针对分析出的特点采取对应的防护措施，达到正向的预防与保护作用。

依据数据分析，本作品可为高层安全监管部门提供详细的情报和决策支持，从事件的发生到事件的预防，本作品都起到了关键性作用。

### 1.1.3 目标用户

立足于“高层安全”问题，表1.1指出了本作品的目标用户及其选择本作品的原因。

目标用户	特点与需求分析	作品推广建议
公共部门（如学校、政府机构）	由于这些机构的运营不以盈利为目的，且往往面对大规模采购需求，因此对于产品的性能与可靠性有较高要求。	可通过挖掘和利用政府相关政策的支持，着重强调产品的稳定性和长期维护保障，以促进与政府部门的长期合作关系。
小区物业及居民	考虑到小区覆盖面积广、居民需求量大，同时社区居民对高层建筑的安全性持续关注，特别是高层安全问题。	在推广时需精准定位用户的痛点，强调产品在提高居住安全方面的价值，并考虑价格因素对决策的重要影响。
商业高层建筑用户（如企业、高级物业管理）	此类用户的需求相对分散且具体，每个企业或物业对产品的功能和效能需求可能都有所不同，对于高端定制化产品较为开放。	需要进行细致的市场调研，了解具体需求，并提供定制化的解决方案。同时，宣传产品的独特性和高端定制的优势。
高科技园区等特殊区域	这些区域通常面积广阔，安全需求高，对于高层安全隐患持续关注，需求功能上比较统一，但对性价比的要求也相对较高。	推广策略需要考虑到这些区域的具体特点和环境，提供性价比高的解决方案，并强调产品如何适应特殊环境的能力。

表 1.1: 高层智能监测系统目标用户

## 1.2 背景分析

### 1.2.1 国家政策背景

近年来，中国智能安防行业相关国家政策频出，主要政策内容围绕鼓励安防行业与互联网、5G 领域技术结合，推广智能安防应用等。具体政策如表1.2所示。

发布时间	政策名称	机构	重点内容解读
2021 年 12 月	《“十四五”城乡社区服务体系建设规划》	国务院办公厅	推广社区级的高端议事、政务、养老、家政及托育网络服务；促进社区内物业设施及安防系统的智能化升级。
2021 年 7 月	《5G 应用“扬帆”行动计划 (2021-2023)》	十部门	推动 5G 基础智能家居产品如安防监控、智能音箱、穿戴设备及服务机器人的发展；加速智能穿戴如云 AR/VR 设备、5G+4K 视频设备及全景 VR 相机的市场推广，刺激新型消费和体验式消费增长。
2021 年 3 月	《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》	中共中央	关注高端芯片、操作系统、人工智能算法、传感器等关键技术领域；推动人工智能、大数据、区块链、云计算及网络安全等数字产业的成长。
2020 年 12 月	《关于推动物业服务企业加快发展线上线下生活服务的意见》	住建部、工信部等六部门	强调构建智能化安防系统，创建智慧型安防小区，确保居民的居住安全。
2020 年 7 月	《关于开展 2020 年网络安全技术应用试点示范工作的通知》	工业和信息化部办公厅	针对智慧家庭、智能抄表、零售、智能安防、智慧物流及智慧农业等应用场景，进行物联网卡、芯片、联网终端、网关及应用平台的网络安全管理和威胁监测。

表 1.2: 国家性政策汇总

随着我国国家安全的内涵和外延日益丰富，保障各领域安全的任务变得愈加繁重，相应的经济投入需求也在逐年上升。

根据最新统计数据，从 2021 年到 2023 年，广州市本级财政共投入 7.05 亿元用于社会治安视频监控系统的

建设。这一投资旨在建立一个高标准、高水平、高效率的广州市社会治安与城市管理智能化视频系统。具体来说，2021年，广州市财政首次重点关注社会治安视频监控系统，当年即投入了1.81亿元，以期构建一个共建共治共享的社会治理新格局。



图 1.4: 公共安全领域投入逐步扩大

此外，《“十四五”规划和2035年远景目标纲要》提出聚焦智能安防上游高端芯片、操作系统、人工智能关键算法、传感器等关键领域；同时提出培育壮大人工智能、大数据、区块链、云计算、网络安全等智能安防相关新兴数字产业。

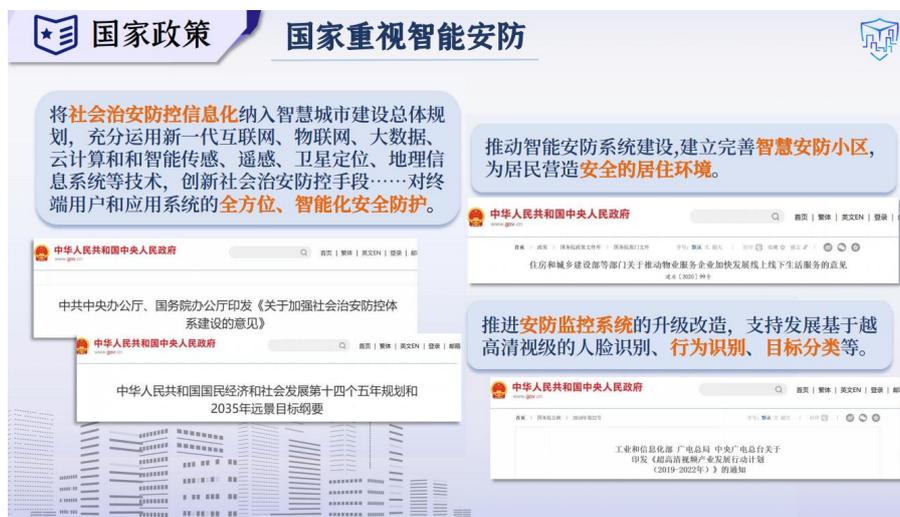


图 1.5: 国家重视智能安防

在国内经济持续增长和安防需求日益上升的大环境下，中国的安防市场呈现出快速的发展态势。根据中安网的统计数据显示，2022年中国安防行业的总产值已经达到8510亿元，相比前一年增长了2.9%。到了2023年，该行业的产值进一步增至9020亿元。

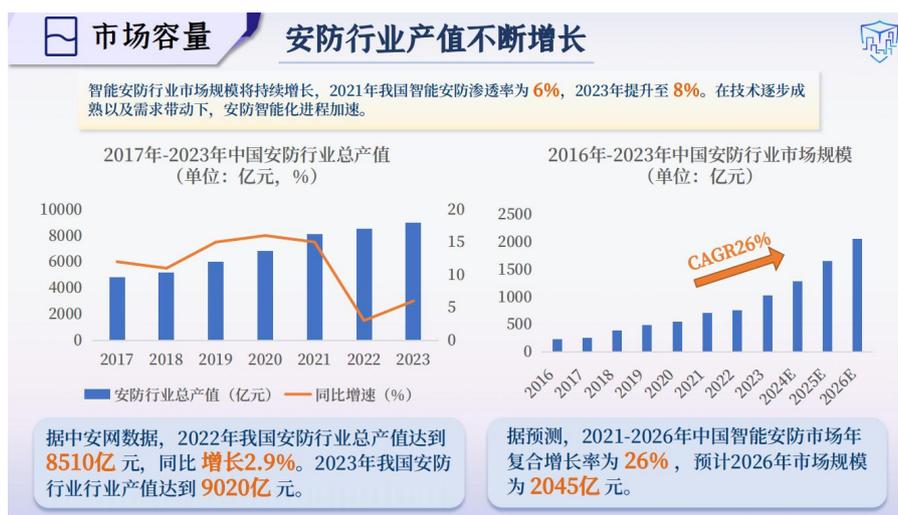


图 1.6: 安防行业产值不断增长

## 1.2.2 社会背景

高空抛物被视为城市中的一大顽疾，常与“随手丢垃圾”并列，位居“城市不良行为榜”第二位。这种行为不仅反映了不文明，同时也对社会造成了严重的危害。在我国，高空抛物问题首次作为法律案件出现在2000年的“重庆烟灰缸事件”中。



图 1.7: 高空抛物事故新闻

2019年11月，最高人民法院发布了《关于依法妥善审理高空抛物、坠物案件的意见》，对故意抛物行为依据情况进行法律处置。2021年3月1日，《刑法修正案（十一）》将高空抛物纳入作为独立罪名的“妨害社会管理秩序罪”，明确了其刑事责任。然而，高空抛物事件发生后，诸如确定抛物楼层和抛物人身份等问题常常困扰执法机关，影响追责效率，有时甚至导致整栋楼的居民共同承担责任。

一些高空抛物案件发生在目击者较少的情况下，如果当时又没有相关视频证据，或无法通过视频推测抛物路径，揪出“肇事者”的概率就会很小。**取证难**

高空抛物达到“情节严重”才构罪。哪些情形属于情节严重？具体司法实践中，需要结合抛物的重量、形状，抛掷的场所、次数以及行为人的主观心态等因素综合判断。比如抛掷大件、锋利物品，在人多的地方抛掷物品，多次抛掷物品，或行为人明知可能砸中人仍为之的，更容易被判断为情节严重。**判定难**

但并非所有的高空抛物行为都以高空抛物罪来追究刑事责任，如果同时构成其他犯罪，将依照处罚较重的规定定罪处罚。“比如从高处向公共场所抛掷物品，可能构成以危险方法危害公共安全罪；造成他人重伤或死亡的，根据不同情形可能会构成过失致人重伤罪、过失致人死亡罪或故意伤害罪、故意杀人罪。”**定罪难**



图 1.8: 法律困境

研究表明，未能及时处理的高空抛物事件会显著增加类似行为的发生概率。因此，对高空抛物的实时监控和预警处理显得尤为重要。目前，常用的检测方法存在实时性不足、漏检和误检率高等问题。为此，一些社区物业考虑安装监控摄像头，这些摄像头不仅可能产生威慑效果，还能在再次发生抛物事件时提供有力的视频证据，帮助警方侦破案件。我们的团队正在研究如何通过社区常见的监控摄像头，部署高质量高效率的高层智能监测系统，实现高层危险事件的实时监测。

据《世界预防儿童伤害报告》统计表示，高空坠落是常见的儿童意外伤害死亡五大原因之一。数十万计儿童坠落，其中有近 4.7 万名儿童意外身亡，而在 1—4 岁的调查人群中，有超过 80% 的的坠楼儿童大多发生在窗户或阳台。如图 1.9 所示。

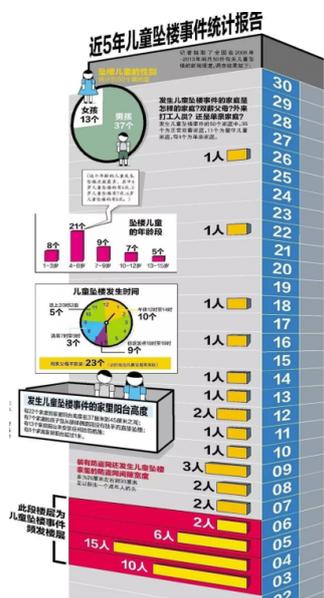


图 1.9: 近 5 年儿童坠楼事件统计报告

不完全统计显示，全国每年累计发生的儿童坠楼事件不少于 2000 起，而近年来，事件更是高发。而且高处坠落的儿童多为低幼儿！对于社会而言，“掉下来的不仅仅是一个孩子，更是一个家庭的幸福和希望”。

“窗户边的隐患”触目惊心，一次次地引发公众对安全的思考。

2020 年全国共发生房屋市政工程生产安全事故 734 起、死亡 840 人。其中高处坠落事故 383 起，占总数的 52.2%。



图 1.10: 近三年高层事故数量变化趋势

高空坠落形成的冲击力对人体的伤害是毁灭性的、不可逆转的。坠落伤是高能量创伤之一，也是引起严重复合型创伤的常见原因。高空坠落极易危及人体头脑、肾脏、心肺等多个组织器官系统，因此失去生命者更不在少数。

社会的稳定安宁、家庭的美满幸福、个人的人身安全都系于“保障高空安全”这一重要命题，缓解上述危险情况，刻不容缓。

## 1.3 特色描写

### 1.3.1 高空抛物检测与高层危险行为原创数据集

为了有效克服这一障碍并提升我们网络的性能以及算法的实际效果，我们在获得必要的审批和合规之后，投入了大量资源去构建一个专门的数据集。整个过程包括了为期两个月的数据采集、预处理和分析，详细的工作流程和技术细节已在文档的“数据集的采集与预处理”部分（见 4.1）进行了全面阐述。

具体来说，我们自主构建的数据集包含了两大类：高空抛/坠物和高层危险行为。**高层人体行为数据集**包含 1,160 个视频，覆盖了 16 个不同的动作类别，这些动作类别进一步分为三大类：单人动作，人对物的动作，和人对人的动作。每个视频均在四种不同的光照强度和分辨率下拍摄，并涵盖了 16 个场景，共计 54,045 个精细标记的帧，以确保数据的多样性和丰富性。

同时，**高空抛物数据集**涵盖了 1,157 个不同的场景，包括 125,616 个视频帧。这些视频记录了四种天气状况、两种光线情况下的事件，覆盖了 17 种场景，并涉及 10 种不同抛物类别。此外，这些数据从三种摄像机拍摄角度和四种视频分辨率中捕获，确保了高度的实用性和适用性。

### 1.3.2 类间距离蒸馏的语义分割算法

在应用语义分割方法提取监控视频中建筑物的区域的过程中，我们特别采用了团队独创的类间距离蒸馏 (IDD) 方法结合知识蒸馏，这一策略显著地提高了学生网络从教师网络学习的效率，尤其是在特征空间中的类间距离的迁移上。此外，鉴于语义分割任务对位置信息的高依赖性，我们还引入了一个专门的位置信息蒸馏模块，它特别处理水平和垂直方向的信息，以进一步增强学生网络对位置细节的捕捉能力。

这一系列技术的融合不仅提高了语义分割模型的准确性，还极大地增强了整个系统的技术优势，使其在处理高空抛物监测任务时更为高效和可靠。通过这样的技术创新，我们能够在保持高性能的同时，也保证了系统的实时响应能力。

### 1.3.3 融合 MA 模块以增强前景运动特征检测

在高空抛物监测中，尤其是针对高层建筑，背景的变化通常较为有限，而抛物速度较快，这对实时检测系统提出了特别的要求。基于这种需求，我们引入了一种新型的注意力机制——Mask-Attention (MA) 模块，专门设计用来增强视频监控系统对前景运动特征的识别能力。

MA 模块的设计借鉴了现有的注意力机制研究 [2, 10]，并将其与卷积神经网络 (CNN) 的特征提取功能结合，实现了在特征图中自适应调整各个空间位置的特征权重。这一机制有效地聚焦于视频帧中的动态前景区域，同时抑制背景噪声，从而优化了检测性能。网络结构的具体实施如图 1.11 所示。

具体操作流程包括：首先对视频帧进行深度特征提取，获取特征图；接着通过全局平均池化操作计算出特征图每个位置的注意力权重，并利用这些权重对特征图进行加权处理，合成最终的特征表示。这个特征表示进一步在全连接层中用于前景运动物体的分类与回归任务。

此外，为了进一步提升检测的精度，MA 模块还融入了前景运动信息，通过突出动态物体的关键特征并抑制无关信息，提高了网络的响应速度和学习效率。这种方法的引入不仅加速了网络的收敛，还提高了抛物轨迹预测的准确度。

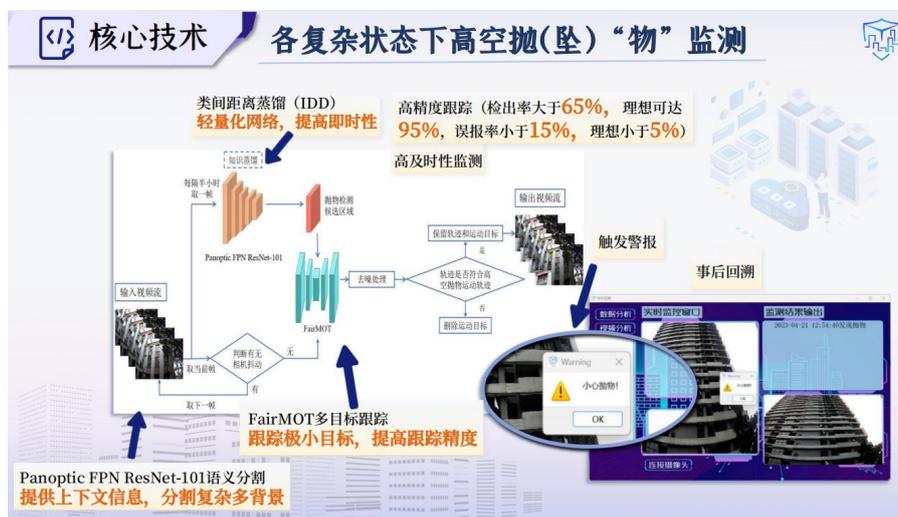


图 1.11: Mask-Attention

通过这种创新的融合方式，MA 模块有效地指导了高空抛物检测网络学习在视觉空间中应关注的关键信息，实现了对运动相关特征的加速识别，从而显著提高了系统对高空抛物行为的识别和预测能力。

### 1.3.4 双流网络结构

现有的研究主要集中于直接移植目标检测方法进行单帧检测，这样的做法虽然简单易行，却完全忽略了时间维度上的变化，导致了动作时空定位精度较低。所以在我们的核心算法中，有效地结合时序信息和空间信息是提高检测和识别精度的关键。

为了从视频帧中提取动作的**时序特征**，我们采用了改良的 3D ResNeXt101 网络。这个网络模型通过替换二维卷积为三维卷积，能够捕捉时间上的连续性和空间上的特征。同时，关键帧还需输入到基于 CSPDarkNet-53 的 YOLO-X 目标检测网络中进行**空间特征**的提取。

我们使用这种 2D 与 3D 网络结合的双流结构，通过自注意力机制融合时空信息，保证了动作时空特征的一致性，提高了时空定位精度。

## 1.4 创新性说明

### 高空抛物检测：

1. 采用了创新的应用类间距离蒸馏语义分割算法，有效地在监控视频中识别出建筑物区域，进而预测物体可能的坠落位置，成功克服了复杂环境中高空抛物检测的挑战。此外，作品还利用知识蒸馏技术开发了一种轻量级的语义分割网络，显著缩短了推理时间，提升了系统的实时响应能力。
2. 在背景变化较小且光照条件良好的环境中，作品通过背景差分法快速解决了目标物辨识问题。对于背景变化频繁的场景，采用帧差分法更新模型，利用帧间差异提高目标检测的准确度。
3. 创新性地结合多种技术手段减少干扰，包括使用相机抖动检测消除一部分干扰、采用非极大值抑制技术清除不完整的检测框，并通过分析目标运动轨迹验证其是否符合抛物轨迹，这些方法的综合应用大幅降低了误检率。
4. 引入的 MA 模块使得检测网络能够在监控前景运动信息的指导下学习关注空间维度中的关键信息，通过强调重要动态信息并抑制不重要内容，提高了信息处理效率，加速了网络学习过程，并在一定程度上提升了预测的准确率。
5. 鉴于现有数据集大多只适用于单一视角和无遮挡的目标检测，不满足项目对特定条件的需求，我们通过网络搜集与实地拍摄相结合的方式，扩充了数据集，构建了一个多视角、适应不同光照、天气变化等环境因素的高空抛物数据集。

### 高层危险行为检测：

1. 算法更精确快速。本作品是属于动作时空定位范畴的技术衍生物。本作品采用的算法使用了 2D 与 3D 网络结合的双流结构，通过自注意力机制融合时空信息，保证了动作时空特征的一致性，提高了时空定位精度。此外，该算法处理视频的帧速率为 25FPS，在这种帧速率下高层人体检测、高层动作识别过程较为流畅，达到实时计算的水平，保证了危险动作报警的及时性。
2. 结果更形象直观。本作品对于检测结果有两方面的输出。一方面，本算法预测结束后，会在原始视频中标示出人体目标的位置框，并在边框左上角标注动作类别和置信度，将结果实时展示在监控的输出界面，方便相关人员及时定位高层危险动作的发生位置。另一方面，本算法还会根据具体动作类别输出高层行为的危险系数，在终端输出不同等级的报警信息，有利于相关人员及时采取应对措施，阻止危险的发生。
3. 高层行为数据库的构建提高了特定场景下的检测精度。本团队提出了第一个专门面向高层场景的动作识别数据集，使用该数据集对原始网络进行微调后，本作品在高层动作识别和检测任务上的精度将高于其他现有算法。同时，该数据集作为高层行为识别领域的第一个正规数据集，具有开拓性意义，为往后该领域的发展奠定了基础。

## 第 2 章 需求分析

### 内容提要

- 现有解决方案
- 解决问题的思路
- 现有产品的不足之处

### 2.1 现有解决方案

在当前的高层监测技术应用中，尽管取得了一定的进展，但仍面临着一系列挑战。首先，**准确性弱**，这意味着系统经常无法正确识别或分类高层中的特定行为或事件。其次，**鲁棒性差**，在多变的环境条件下，系统的性能往往不稳定。此外，**实时性弱**，在处理和响应速度方面还远远达不到实际应用的需求。最后，现有的监测解决方案**未针对高层复杂场景针对优化**，这在很大程度上限制了它们在复杂高层环境中的有效性。这些问题突显了现有技术在应对高层建筑特有的安全监测需求时的不足。

#### 2.1.1 高空抛物检测

GPk-YOLOv5s 方法 [21] 是一种高空抛物检测和跟踪的智能检测网络，利用 C3Ghost 模块提取有效特征并简化网络结构，C3Transformer 模块加强特征提取并捕捉全局上下文信息。CARAFE 用于上采样以提高像素级预测效果，而新的大规模特征映射检测层能获取更小的感受野，便于检测小型目标。多尺度检测头结合 CBAM 增强目标的表达能力，且卡尔曼滤波器的引入有助于克服实时检测中的帧损失，实现更精确的目标跟踪。

然而，该方法存在一些不足，包括模型训练方式需要进一步优化以提高有限样本的预测准确性，以及当前模型在数据较少的情况下会表现不佳，显示出对大数据的依赖，迫切需要大规模的高空抛物优质数据集。此外，尽管总体计算消耗降低，模型的复杂性使其在资源有限的环境中难以部署，且在极高速度和复杂环境中的实时处理能力仍面临挑战。**对高质量抛物数据依赖高，资源有限环境难以部署**

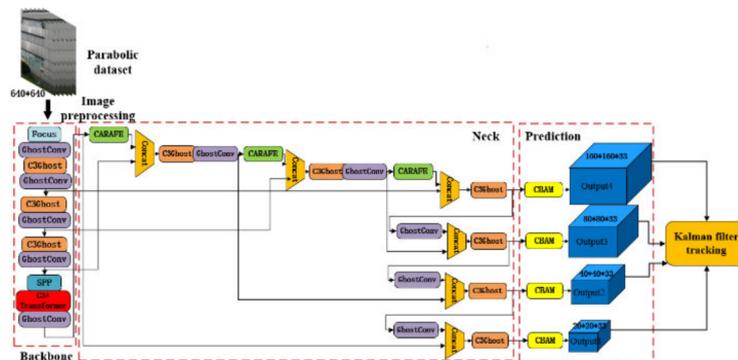


图 2.1: GPk-YOLOv5s 高空抛物检测和跟踪系统示意图

转向另一种基于高斯混合模型 (GMM) 和简易在线实时跟踪 (SORT) 算法的技术 [14]。首先，使用基于 GMM 的背景建模方法从视频序列中分离背景和前景，以获取动态目标的运动图像。通过数学形态学去噪，进一步确定动态目标是否为抛掷物体。此外，结合 SORT 算法实现对抛掷物体的跟踪，确保检测的连续性和稳定性。实验结果表明，该方法能够有效识别和跟踪高空抛掷物体，具有较好的稳定性和较低的误检率，检测精度、召回率和 F1 值均有显著提升。

然而，该方法在处理极端情况如极高速抛掷物或环境复杂度极高的场景时的有效性尚未得到充分验证。此外，依赖于视频监控设备的质量和角度，检测效果可能会有所不同，存在一定的局限性。未来的研究需要针对

算法的泛化能力和适应性进行优化，减少对高质量视频数据的依赖，以应对更多样化的实际应用场景。**复杂情况下的有效性未验证，对监控设备质量和角度敏感，算法泛化能力和适应性不足。**

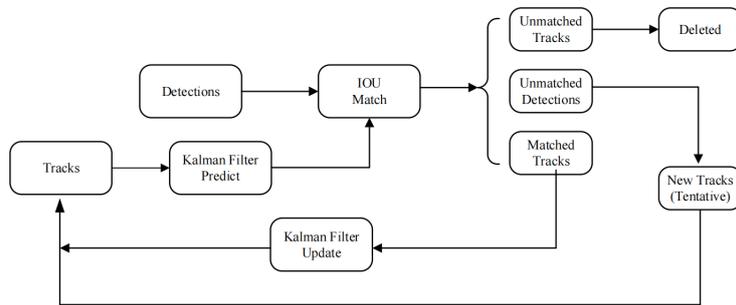


图 2.2: 基于 GMM 和 SORT 算法的高空抛物检测系统示意图

接下来，基于机器视觉的高空抛物智能检测与预警系统则结合了传统视频监控和先进的深度学习技术 [5]。系统首先通过灰度和高斯滤波方法对初始图像进行形态处理，然后通过 5G 通信模块快速传输数据。接收端使用卷积神经网络算法识别从高空坠落的物体类型，并通过双摄像头接收的数据，利用卡尔曼滤波方法和建立的数学模型共同预测坠落物的落点坐标。声光预警模块接收到坐标后，迅速处理数据，在物体落地前使用引人注目的激光标记落点，并配备声音警报提醒行人，同时部署阻挡网以防止行人受伤。

尽管该系统具备高精度和低成本的优点，并能及时发现高空抛物源头，最大限度地避免因坠落物体造成的重大伤亡，但仍存在一些不足。首先，系统的有效性高度依赖于视频监控设备的质量和安装角度，这限制了其在各种环境中的普适性和准确性。其次，虽然系统结合了深度学习进行物体识别和落点预测，但在极端天气条件或复杂光照环境下，其性能可能会受到影响。未来的研究需进一步优化算法的鲁棒性，提高系统在多变环境中的适应性和准确性。**高度依赖视频监控设备质量和安装角度，在极端天气或复杂光照下性能受影响，算法鲁棒性和适应性不足**

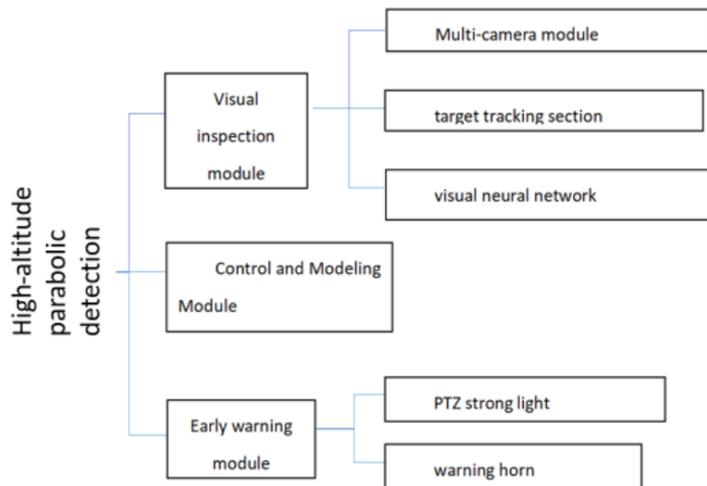


图 2.3: 基于机器视觉的高空抛物智能检测与预警系统组成图

最后，基于改进帧差法 (IFD) 的高空坠落自动检测方法，以提高监控视频中高空坠落行为的检测效率 [12]。首先，通过去抖处理对视频数据进行预处理，然后确定最优间隔帧数，使用这一帧数获取运动前景。接着，提出了一种坠落判断标准，用于判断运动前景是否为高空坠落。实验结果表明，相较于传统的帧差法和三帧法，该方法能够更高效地处理监控视频，有效解决了其他方法在检测高空坠落物体时效率低下的问题。

然而，该方法尚处于实验阶段，并未广泛应用于实际场景中，其在实际环境中的表现和适应性还有待验证。此外，尽管该方法提高了检测效率，但在复杂的环境条件下，如极端天气或光线变化大的场景，其稳定性和准确

性会受到影响。未来的工作需要将此方法应用于实际场景，并根据具体应用场景进一步优化算法，以增强其鲁棒性和适应性。**尚处于实验阶段，适应性待验证，在复杂环境下稳定性和准确性受影响，无法迁移到多个场景**



图 2.4: 基于改进帧差法 (IFD) 的高空坠落自动检测系统示意图

### 2.1.2 高层危险行为检测

在人体行为检测领域，虽然当前存在多种方法可通过解析视频数据识别行为模式，但这些方法在特定的高层危险行为检测方面仍显示出明显的局限性。特别是针对此类高层行为的检测，现在尚未有明显的解决方案，这表明此领域的潜力仍未被充分开发。

3D Convolutional Neural Networks (3D-CNNs) [20] 通过在视频帧序列上应用三维卷积来直接提取时间和空间特征，使得它们特别适合处理视频数据。这种网络结构在理解视频内容中的动态变化方面，较传统的 2D CNNs 有明显优势。然而，3D-CNNs 在实际应用中可能因训练数据的局限性而难以泛化到未见过的场景或新的行为类型上，影响其应用的广泛性和有效性。**泛化性差**

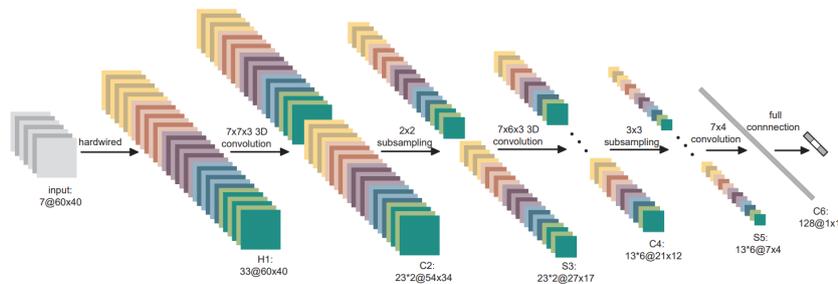


图 2.5: 3D Convolutional Neural Networks

Temporal Segment Networks (TSNs) [18] 通过将视频分成若干短片段，并独立地从每个片段中提取特征，之后将这些特征整合来进行行为分类。此方法有效利用了视频中的全局信息，并在减少计算量的同时保持了较好的性能。然而，片段的选择可能导致重要信息的丢失，尤其是当危险行为恰好在片段的边界发生时。对于需要连续理解的长序列危险行为，TSN 的性能可能会受到限制。**监测不确定性高**

Inflated 3D ConvNets (I3D) [8] 通过将 2D CNN 架构扩展为 3D，并“膨胀”卷积核，使得网络可以从预训练的 2D 网络转换为 3D，以此学习更丰富的时间特征。I3D 能较好地捕获时间信息，但在处理细微的动作变化时，其精确度可能不及专门的时序分析模型。**监测细微动作变化时精确度不足**

这些模型虽然提供了视频行为检测的有效手段，但未对高层危险行为进行特别优化，限制了它们在该特定领域的应用效果。此外，高质量的专用数据集缺乏也是影响这些技术进一步发展的重要因素。针对高层危险行为检测的研究和开发，需要更精细的方法和更丰富的数据支持，以实现在此类复杂环境中的有效应用。



图 2.6: 现有技术短板分析

## 2.2 现有产品的不足之处

竞品分析		与现有产品对比: 功能全面, 独具优势		
现有产品分析	海康视视——高空抛物自清洁智能 800W定焦筒机 (集成)	本产品 (集成) (产品商业化后预计) 算法服务: ¥5,000-60,000 / 每年 定制服务+摄像机: ¥500,000-5,000,000 / 每年	上海智眸科技——“牛盾”系列高空抛物智能侦测仪 (非集成, 自配摄像机)	大华——智能检测报警带抛物轨迹线 (仅检测仪)
售价	算法服务: ¥160,000 / 每年 摄像机: ¥199,00 / 每台 定制服务: ¥800,000起		算法服务: ¥144,000 / 每年 侦测仪: ¥9999 / 每台	侦测仪: ¥5190-10,390 / 每台
产品性能	场景适应性强, 分辨率为3840 × 2160, 画面清晰	场景适应性强, 分辨率为3840 × 2160, 画面清晰	场景适应性一般, 分辨率为2560*1440, 相对较低	多场景适用, 但分辨率为2688 × 1520, 相对较低
兼容性	只兼容海康播放器	多算法相互结合, 使算法具备演进能力, 兼容性极佳	兼容性良好, 但需自配摄像机	兼容性好, 支持多段远程查看
检出率	可同时检测110个目标, 检出率大于95%	检出率理想可达95%, 误报率小于15%	白天大于97%, 夜晚大于95%	检出率理想可达95%以上
高空危险行为检测 (本产品的创新与优势)	不具备高空危险行为的自动检测功能	快速检测人体 (检测率大于70%, 理想可达90%), 快速识别行为及危险系数确认 (误报率小于15%, 理想可达5%)	不具备高空危险行为检测功能	不具备事前智能判断预警预报功能和高空危险行为检测功能

图 2.7: 竞品分析

### 2.2.1 现有产品内蕴算法准确性与鲁棒性不强

现有高层建筑智能监测系统内蕴算法鲁棒性不强的问题主要表现在以下方面:

首先, 面对不同的场景和复杂环境, 算法的准确率和鲁棒性不高, 易受到外界因素的影响。例如, 天气、灯光、建筑物等因素会影响算法的检测结果, 而一般情况下, 高层监测的环境是复杂的, 建筑物的不同建筑风格与外界环境变化都是十分常见的, 因此如何解决复杂场景中的关键信息提取问题变得非常重要。

目前有部分产品主要采用的算法模型是传统的计算机视觉算法或单一的目标检测算法, 对于一些复杂的场景和情况, 算法的效果不佳。此外, 传统算法的优化空间有限, 无法充分挖掘监测数据的特征和规律, 从而影响算法的准确性, 导致漏检、错检比例高。

相比之下, 本作品内蕴算法具有更高的准确率和鲁棒性, 能够更好地适应不同的监测场景和建筑物需求, 能够更好地挖掘监测数据中的特征和规律, 从而提高算法的准确性和鲁棒性, 具有极大的优化空间。

### 2.2.2 现有产品实时监测效果不佳

高层安全监测的即时性要求非常高，主要是因为高层危险事件是一种突发事件，发生的时间和地点都是不确定的，需要及时发现并采取措施。如果监测系统的响应时间过长，可能会导致伤害或财产损失等严重后果。

另外，高空抛物与高层危险行为发生的速度通常较快，如果监测系统的响应速度不够快，可能会导致监测数据的不准确，进而影响监测系统的效果。因此，高空抛物监测系统需要具备较高的实时性能，能够在短时间内对高空抛物进行快速检测和识别，并及时报警。

但现存的高层智能监测系统，一方面为了追求高精确性，会采用较为复杂的监测模型与算法，导致实时监测的速度有欠缺，难以完成实时监测要求；另一方面为了追求高灵敏度，导致算法的精确性不够，对应了上述讨论的准确性问题。

而我们提供的监测方法，则针对高层安全问题（如高空抛物、危险动作预警等），解决当前的高层危险事件监测系统的性能不理想的问题（例如通常面临实时性不足，漏检、错检比例高等）。本方法抗干扰能力强，能充分适用于小区，办公楼等复杂场景，可检测到体积较小的目标且跟踪效果较好，并能有效处理高层危险事件检测中目标小、算法时效性差和场景复杂三大难点问题。

### 2.2.3 现有的平台承载能力弱

随着智慧城市、智慧社区、智能安防等项目的大规模开展以及跨区域性联网项目的兴起，智能监测系统的建设规模也越来越大。如果作为系统核心的管理平台软件能力不够，那么系统就会面临着持续建设的问题。

现存的监测平台往往以企业垄断性为开发前提，自身开发的系统与其他企业开发的系统无法联通，形成“孤岛效应”。如大华——智能检测报警带抛物轨迹线，其只兼容海康播放器。对于一些大型建筑物，可能涉及多个智能监测系统的数据交互和共享，这就需要建立更加开放、灵活的数据交互和共享机制，以提高系统的整体效能。

因而，建立共享的管理平台，集成多个监测系统的数据和功能，以提高整体效能和管理效率是非常重要的问题之一。本作品在开发初期，便秉持着联通，共享的原则，计划采用标准化接口，尝试实现数据的格式化统一处理，以实现系统间的数据交互和共享，提高系统的互通性。

### 2.2.4 我国安防普及率较低

据统计，目前全球民用安防普及率整体水平 10%，而美国民用安防普及率在 50%，中国民用安防普及率仅约 11%。目前我国智慧社区安防市场大约占安防市场总产值的 6%-10%，与国外成熟市场相比，我国安防社区市场比例很低。从以上数字来看我国的安防产业市场在民用领域还有很大的上升空间。

视频监控管理平台行业与众多安防产品类似，都缺少相应的技术规范和管理标准，行业内各前端设备厂家的标准、协议、接口等不一致，导致视频监控管理平台软件应用于市场的难度增大。平台厂商不可避免的重要工作就是与主流前端设备厂家进行大量的协调沟通，由此增加了平台开发的技术难度和技术投入，从而也大大提升了平台软件的开发成本。

此外高空人体危险行为检测功能则作为创新点填补了市场空白，当前高层危险行为检测有关产品市场仍处于蓝海阶段，从而迫切需要新产品的进入以填补空白。



图 2.8: 现有产品短板分析

## 2.3 解决问题的思路

结合当前的背景分析和行业挑战，本作品的核心优势在于实施双重角度监控和融合一体化的通信与数据管理。我们通过监控高空抛物和其他高层危险行为，能够在事故发生之前根据事件的危险级别主动触发警报系统。这样的及时预警机制使得相关人员能迅速响应，极大地阻止了可能的悲剧。事后，系统还能进行事件的回溯取证，有效地解决高空抛物案件的三大难题：定位、认证和追责。因此，系统的主要功能可以总结为“预警先行，事后追责”。此外，我们的解决方案还包括高级信息建模处理。

本团队致力于利用深度学习和计算机视觉技术，在现有深度学习框架的基础上进行优化和提升，特别是解决光线变化和微小目标跟踪的难题。我们设计的算法具有高度的鲁棒性，能够在光照变化、背景噪音、空中遮挡等复杂环境下有效运作，减少因环境变化引起的误报。

从单一功能到多功能集成，实现全场景、跨领域的高层安全监控，全面保障公众的人身和财产安全。



图 2.9: 解决方案

### 2.3.1 小目标跟踪及复杂环境适应性强化

在应用高层智能监测系统于实地场景中时，经常会遇到树叶晃动、飞鸟等自然因素干扰摄像头判断，尤其是在社区这样背景复杂的环境中，楼层推算尤为困难。

为了克服这些挑战，我们的团队计划综合当前主流的目标检测技术，开发一种多场景适用的检测模型。我们特别引入了类间距离蒸馏的语义分割算法，以此来识别监控视频中的建筑物背景，明确潜在的抛物区域，有效减少了复杂环境对监测的干扰。

鉴于抛物目标通常较小，小目标的追踪难度大，并容易在复杂环境中误检，我们采取了融合多种算法优势的策略。在背景变化小且亮度明显的环境中，尽管背景差分法能有效识别目标，但其对计算资源和时间的需求较高，不适宜于实时场景。因此，我们选择了自主开发的 FairMOT+MA 模型，这一模型不仅考虑了小目标检测的挑战，也适应于复杂环境的监测需求。

### 2.3.2 提升监测系统的灵敏度与准确性

为了提升抛物检测的灵敏度和准确度，我们引入了卡尔曼滤波算法进行目标物体的状态估计和预测。通过结合物体的运动模型和实时测量数据，我们能够准确地估计目标物体的位置、速度和加速度，以及预测其未来状态，从而优化目标跟踪和检测过程。

此外，我们也实施了多种干扰消除策略以提高监测准确度。首先，通过相机抖动检测消除初步干扰，接着使用非极大值抑制方法清除不完整的检测框，最后通过分析目标轨迹的抛物性质，确保在抛物事件的各个阶段—发生前、发生时、以及事后—均有效地减少干扰。

在高层危险行为的监测上，我们采用了结合 2D 与 3D 网络的双流结构算法，利用自注意力机制融合时空信息，确保动作特征的时空一致性，提升了时空定位的精确度。该算法支持高帧率视频处理，能够流畅进行高层人体检测和动作识别，实现实时计算，确保及时发出危险动作警报。

### 2.3.3 构建高空抛物与高层危险行为的专用数据集

#### 1. 数据集构思

对于高层危险事件监测的深度学习模型，构建一个多元化和全面的数据集是至关重要的。详细的数据集构建将在第 4.1 节详述。

我们计划在多种环境设置下收集高空抛物的视频数据，涵盖 4 种天气状况、2 种光照条件、17 种不同场景、10 种抛物种类、3 种摄像机角度和 4 种视频分辨率，拍摄总数高达十万帧的视频。对于高层危险行为的数据集，我们将包括 1160 个视频，16 种不同的动作类型（包括个人动作、人与物的互动、人与人之间的动作）、4 种光照强度、4 种视频分辨率和 16 种场景，总共 54045 个标注帧。

#### 2. 数据集获取

在获得学校的实验许可并确保安全措施到位后，我们将在校园内的高层建筑进行物体投掷和高层危险动作演示。数据的收集将使用宽动态范围（WDR）摄像机，这种摄像机配备了逐行扫描的 CMOS 传感器和点阵 LED 红外灯，以确保高质量的视频记录。

#### 3. 数据处理

我们将收集的视频进行两阶段的标注处理。第一阶段，将视频分帧，并利用 labeling 软件对每一帧进行精确的矩形框标注，完整覆盖抛物物体，生成的标注信息保存为 XML 文件。同时，根据不同的天气条件，对视频帧进行特定的标注处理。第二阶段，基于第一阶段的标注结果，创建相应的掩模文件，保存为 PNG 格式。

对于高层危险行为的数据，我们同样使用矩形框来标识每一个危险行为的参与者，无论是人还是物体，并且详细记录下每一个动作的类型（如跳跃、跑、爬等），以及该动作发生时的环境条件（如天气、光照）。

对应的抛物标注文件举例如图 2.10 所示。



图 2.10: 某抛物视频第 10 帧标记图片

# 第3章 技术方案

## 内容提要

- 技术路线总览
- 高空抛/坠物检测核心算法
- 高层危险行为检测与识别核心算法
- 模型训练与测试

本章主要对本系统涉及到的核心技术路线与具体算法设计进行详细阐述，以提供对系统功能和操作流程的全面理解。

### 3.1 技术路线总览

本系统从保障人们的生命财产安全出发，着眼于城市高层场景，旨在解决两大社会问题：高空抛物和人员坠楼。本系统提出了一体化的高层智能检测系统，主要结合超清监控摄像头，基于事件中人和人、人和物之间的关系，对高层场景中发生的意外事件，进行有针对性、高精度、低误报的检测。

该系统主要针对各类型人群聚集区域，如景区、动物园、建筑楼层等，当镜头检测到抛物或坠楼时，该系统能以高速度、高灵敏度和高准确度定位意外发生位置，第一时间报警，有效保障人身和财产安全。此外，该系统包含的算法具备一定的鲁棒性，能对光照变化、背景干扰、空中遮挡等恶劣状况具备较强的抗干扰能力，不容易受环境变化影响发生误报。

本系统的核心技术囊括两大核心算法：**高空抛/坠物检测算法**与**高层危险行为检测与识别算法**，核心技术如图 3.1 所示：

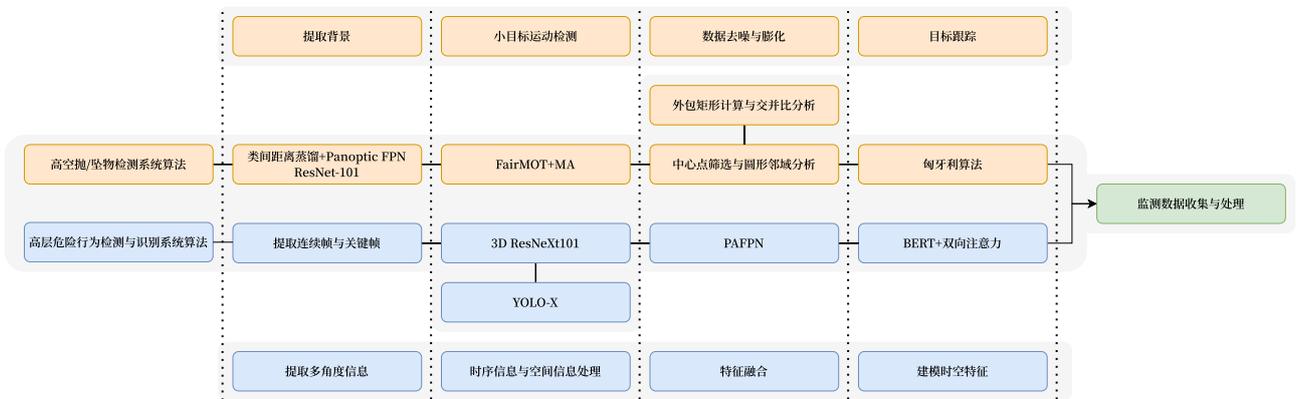


图 3.1: HaMonitorSentry 核心技术总览

其中高空抛/坠物检测算法主要包括先进的图像预处理和语义分割技术，用于精确提取和优化复杂场景中的背景，辅助以去噪，膨胀等操作。接着，利用 FairMOT 模型进行精细的小目标运动跟踪，并实时修正目标位置，同时确保信息迅速反馈到远程监控界面。

此外，高层危险行为检测与识别算法通过 2D 和 3D 网络结合人体空间定位和动作时序分析，融合使用 BERT 结构的双向注意力机制进一步分析行为模式，从而准确评估行为的危险系数并触发相应等级的警报。

通过两个核心算法，我们的系统实现了高精度的实时监控和预警。通过这种集成化和多层次的技术应用，系统能够在多变的环境中稳定运行，具有较低的漏检率和强大的环境适应性，满足高安全标准的监控需求。

## 3.2 高空抛/坠物检测核心算法

目前阶段，本系统算法流程如图 3.2 所示。首先通过语义分割算法在多种复杂场景中提取建筑物背景，辅助以去噪，膨胀等操作；然后在该背景下利用 FairMOT 模型进行小目标运动检测，并修正抛物位置，将信息立刻返回远程界面，方便管理人员进行督察。如发现高空抛物能够及时发出警报，模型整体的可用性比较强，漏检概率低。该算法在轻量化原始网络的同时，提高了语义分割模型的准确性，进一步增加了本系统的技术优势。

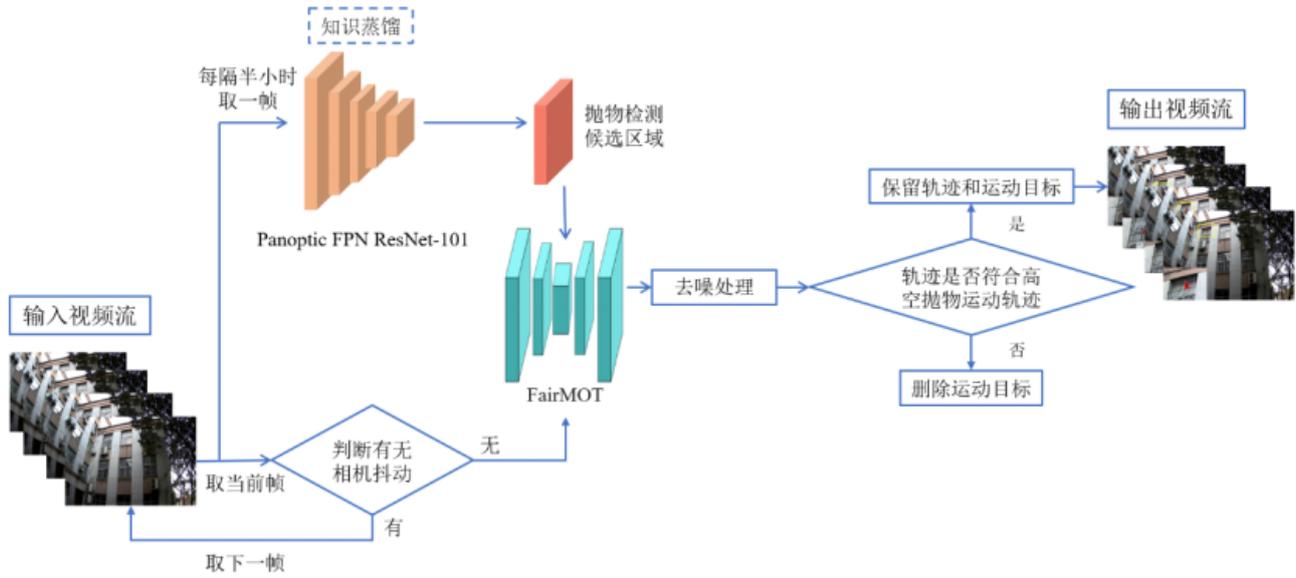


图 3.2: 高空抛/坠物检测算法结构

### 3.2.1 背景提取

首先我们提前进行知识蒸馏，得到轻量化语义分割网络 Panoptic FPN ResNet-101。在实际应用中，系统被配置为每半小时从监控视频中抽取一帧，并在此轻量化网络中进行处理，通过一系列复杂的卷积和池化操作提取出丰富的特征图。

进一步地，我们采用反卷积层对最后一个卷积层产生的特征图进行上采样，确保上采样后的特征图与原始图像尺寸相同。这一步骤关键在于它能够在进行像素级预测的同时，保留这些像素在原始图像中的准确空间位置信息。最终，通过对上采样特征图进行逐像素的分类，我们能够准确地识别出建筑物背景，这一区域是高空抛物检测的关键。

在知识蒸馏的过程中，我们特别采用了团队独创的类间距离蒸馏（IDD）方法，如图 3.3 所示，这一策略显著地提高了学生网络从教师网络学习的效率，尤其是在特征空间中的类间距离的迁移上。此外，鉴于语义分割任务对位置信息的高依赖性，我们还引入了一个专门的位置信息蒸馏模块，它特别处理水平和垂直方向的信息，以进一步增强学生网络对位置细节的捕捉能力。

这一系列技术的融合不仅提高了语义分割模型的准确性，还极大地增强了整个系统的技术优势，使其在处理高空抛物监测任务时更为高效和可靠。通过这样的技术创新，我们能够在保持高性能的同时，也保证了系统的实时响应能力。

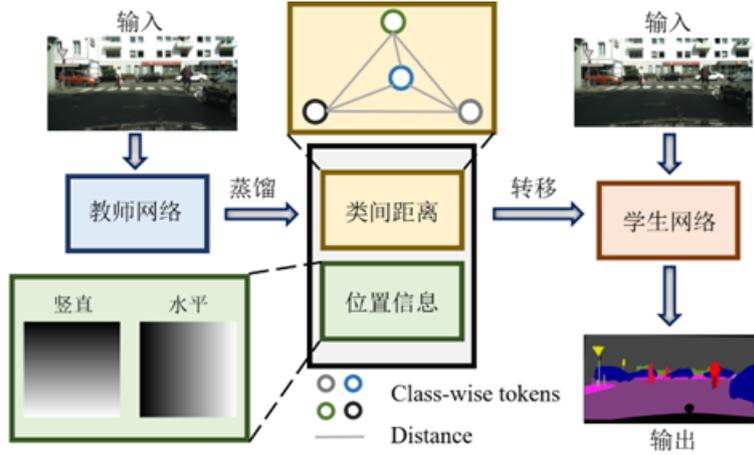


图 3.3: 类间距离知识蒸馏示意图

### 3.2.2 小目标运动检测

为了提高小目标运动检测的准确性，本系统对输入的监控视频流中的每一帧进行详尽分析，以确定是否存在由相机抖动引起的误报。具体步骤如下所述：

- 对于每次处理，首先取当前时间点  $t_N$  的帧以及前一时间点  $t_{N-1}$  的帧。运用背景差分法对两帧之间的差异进行分析，从而计算出当前帧中所有运动目标的总数  $N$ ，并收集每个运动目标的外包矩形尺寸，形成集合：

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}. \quad (3.1)$$

此集合包含每个外包矩形的长和宽，进而用以计算所有运动目标的最小外包矩形总面积：

$$S = \sum_{i=1}^N x_i \times y_i. \quad (3.2)$$

- 分析计算得到的运动区域总数  $N$  和最小外包矩形总面积  $S$ 。如果  $N$  超过预设阈值  $T_1$  或  $S$  超过阈值  $T_2$ ，则判定当前帧受到相机抖动影响，这时系统将不对该帧进行处理，而是跳过并读取下一帧图像进行分析。如果  $N$  小于  $T_1$  且  $S$  小于  $T_2$ ，则认为当前帧稳定，无相机抖动，适合继续执行后续的目标检测和分析步骤。

此处， $T_1$  和  $T_2$  是根据系统的实际监控环境和历史数据优化设定的，分别代表了在无抖动情况下的运动区域总数和外包矩形总面积的阈值。通过这种方法，我们能够有效地筛选出因相机稳定性影响而产生的误检，确保系统输出的准确性和可靠性。

利用之前步骤中得到的静态背景和无相机抖动的当前帧图像，我们采用 FairMOT 进行建筑物区域内的运动目标检测。FairMOT 是一个融合目标检测和 ReID（重新识别）的多目标追踪网络。该网络特别设计为含有两个并行分支，一个分支负责预测像素级的目标检测分数，另一个分支生成 ReID 特征，这两者共同工作以实现高效的目标跟踪。

具体来说，目标检测分支预测目标的存在概率，其数学表示为：

$$P_{\text{det}}(x, y) = \sigma(f_{\text{det}}(I(x, y))) \quad (3.3)$$

其中， $\sigma$  是激活函数， $f_{\text{det}}$  是目标检测分支的输出， $I(x, y)$  表示输入图像在像素点  $(x, y)$  的特征。

对于 ReID 分支，其生成的特征向量用于识别不同帧中的相同个体，表示为：

$$F_{\text{reid}}(x, y) = g_{\text{reid}}(I(x, y)) \quad (3.4)$$

这里， $g_{\text{reid}}$  是 ReID 分支的特征提取函数。

FairMOT 通过这种双分支结构确保了任务之间的公平性，即既不偏向于仅检测也不偏向于仅识别，而是在两者间取得平衡。这种设计显著提高了网络的综合性能，使得 FairMOT 在多目标追踪任务中达到了较高的检测和跟踪精度。

在此基础上，由于高空抛物的监测多针对高层建筑物，其背景变化不是很大，且抛物在后期速度非常快。为了有效提高高空抛物事件的监测精度并确保监测网络能专注于视频帧中的关键前景运动特征，我们在 FairMOT

网络中集成了基于前景运动信息的 Mask-Attention (MA) 模块, 如图 3.4所示。此模块优化了网络对于高层建筑背景下抛物检测的性能, 通过调整每个像素的权重来强调前景运动对象并抑制静态背景。



图 3.4: Mask-Attention

Mask-Attention 模块通过调整卷积层输出的特征图中每个像素的权重, 以加强对前景动态对象的关注并抑制背景噪声。该模块的主要步骤包括:

1. **特征提取**: 利用 FairMOT 网络的卷积层, 为输入视频序列的每一帧提取特征图:

$$F = \text{FairMOT-CNN}(I_t) \quad (3.5)$$

其中,  $I_t$  是在时间  $t$  的视频帧,  $F$  是对应的特征图。

2. **全局池化与权重计算**: 对每个特征图使用全局池化操作以计算出全局平均值, 这些全局平均值将作为注意力权重, 以此突出重要的特征并用于加权合成最终的特征表示。全局平均池化可以定义为:

$$\text{weights} = \text{GlobalPooling}(F) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F(i, j) \quad (3.6)$$

其中,  $F$  代表输入的特征图,  $F(i, j)$  是特征图在位置  $(i, j)$  的特征值,  $H$  和  $W$  是特征图的高度和宽度。

3. **前景运动信息集成**: 在视频处理中, 前景运动信息的集成通过分析连续两帧之间的光流来实现。光流计算每个像素点的运动速度和方向, 从而增强前景运动的检测精度。光流可以通过以下公式计算:

$$V = \text{OpticalFlow}(I_{t-1}, I_t) \quad (3.7)$$

其中,  $I_{t-1}$  和  $I_t$  分别表示时间  $t-1$  和  $t$  时刻的图像帧。光流  $V$  结果包括两个分量, 分别代表水平和垂直方向的像素运动速度, 通常表达为  $V(x, y) = (u, v)$ , 其中  $u$  和  $v$  分别是水平和垂直方向的速度分量。这些运动信息随后被用于更新注意力权重, 使模型能更加关注那些具有显著运动的前景对象, 从而提高处理的整体效果和精度。

4. **注意力加权**: 最后, 利用计算得到的注意力权重和前景运动信息对特征图进行加权, 以优化前景运动区域的表达:

$$F' = F \otimes (\text{weights} + \lambda V) \quad (3.8)$$

其中,  $(\lambda)$  是用于调整前景运动信息贡献的系数,  $(\otimes)$  表示元素乘法操作。

通过这一改进, MA 模块不仅加速了 FairMOT 网络的学习过程, 还提高了其对高空抛物轨迹、移动阴影和其他动态前景元素的识别能力。这种集成方法使得 FairMOT 网络能更准确地预测高空抛物事件, 同时在提高预测精度的同时, 还显著降低了误报率。

### 3.2.3 数据去噪与膨化

为了提升监控系统中小目标检测的精确度并减少误报率，我们对检测到的运动对象执行了一系列系统化的去噪步骤。该流程通过分析和筛选每个运动对象的最小外接矩形（MBR），有效提高了检测精度。

首先，我们从监控视频的当前帧中提取所有运动区域的最小外接矩形，并记录它们左上角和右下角的坐标，生成如下集合：

$$\text{REC} = \{((x_{a1}, y_{a1}), (x_{a2}, y_{a2})), \dots, ((x_{n1}, y_{n1}), (x_{n2}, y_{n2}))\} \quad (3.9)$$

此集合存储了每个识别运动区域的位置信息。接下来，我们对集合中每个矩形进行交并比（IOU）计算，这是一个衡量两个矩形重叠度的关键指标。IOU 的计算公式为：

$$\text{IOU} = \frac{\text{inter}}{\text{union}} \quad (3.10)$$

其中交集 inter 和并集 union 的计算分别为：

$$\text{inter} = (\min(x_{a2}, x_{b2}) - \max(x_{a1}, x_{b1})) \times (\min(y_{a2}, y_{b2}) - \max(y_{a1}, y_{b1})) \quad (3.11)$$

$$\text{union} = s_1 + s_2 - \text{inter} \quad (3.12)$$

$$s_1 = (x_{a2} - x_{a1}) \times (y_{a2} - y_{a1}), \quad s_2 = (x_{b2} - x_{b1}) \times (y_{b2} - y_{b1}) \quad (3.13)$$

同时，我们计算运动像素占比  $V$ ：

$$V = \frac{S_{\text{move}}}{S} \quad (3.14)$$

这里， $S_{\text{move}}$  表示外接矩形内运动区域的面积， $S$  表示外接矩形的总面积。我们还计算每个外接矩形的面积中心点坐标集合 POINT：

$$\text{POINT} = \{(x_a, y_a), (x_b, y_b) \dots, (x_n, y_n)\} \quad (3.15)$$

对于 POINT 中的每个点，以  $(x_a, y_a)$  为例，将其作为圆心，设定一定半径  $r$ ，计算圆内包含的点数  $n$  与到最近点的距离  $d$ 。如果  $n < N$  且  $d < D$ ，则认为该点代表的运动区域符合高空抛物的特征，因此保留该点。点间距离  $d$  的计算方式如下：

$$d = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (3.16)$$

对于 POINT 中的每一个点，以点  $(x_a, y_a)$  为例，将其作为圆心，设置一定的半径  $r$ ，计算圆内包含的点数  $n$  和到最近点的距离  $d$ ，如果  $n < N$  且  $d < D$ ，则认为该点所代表的运动区域符合高空抛物区域的特征，保留该点。点间距离  $d$  的计算方式如下：

$$d = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (3.17)$$

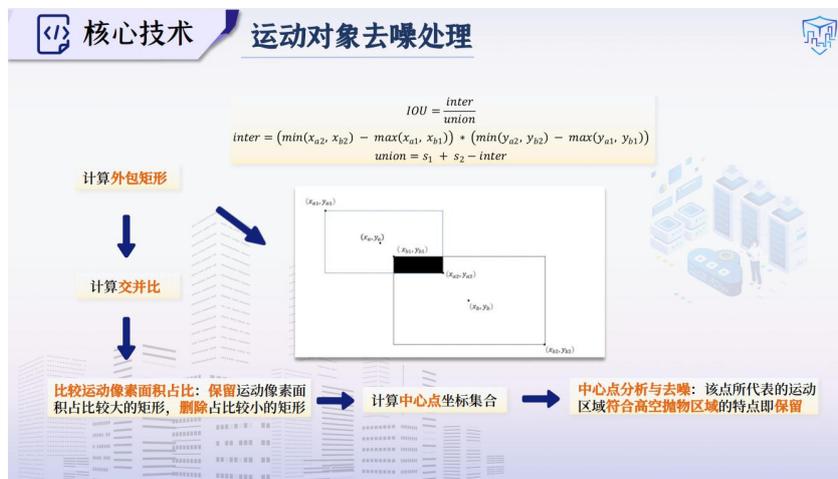


图 3.5: 运动对象去噪处理

### 3.2.4 目标跟踪

为了准确追踪并验证高空抛物事件，我们对去噪后的候选对象应用匈牙利算法进行目标追踪。这一算法通过成本最小化策略确保了目标之间的最佳匹配，从而实现连续帧中目标的有效跟踪。成功跟踪的对象将进一步分析其移动轨迹，以判断其是否符合高空抛物的典型行为。

对于轨迹分析，跟踪得到的每个对象的移动轨迹由一系列坐标点组成，表示为：

$$\text{point} = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (3.18)$$

对于这些点，我们首先计算相邻帧之间的垂直距离 ( $d$ )，该序列如下所示：

$$d = d_1, d_2, \dots, d_n, \quad \text{where } d_i = y_{i+1} - y_i \quad (3.19)$$

此距离序列的增加趋势是识别高空抛物轨迹的关键特征。

接着，为了确定一个对象的轨迹是否符合高空抛物的特征，我们设定以下条件：

1. 垂直距离  $d$  应呈现递增趋势，即  $d_1 < d_2 < \dots < d_n$ 。
2. 连续三个跟踪点的水平坐标变化应保持同一方向，具体条件为： $(x_2 - x_1) \times (x_3 - x_2) > 0$ 。

若一个轨迹满足以上所有条件，则认为该轨迹符合高空抛物轨迹，进一步确认其抛出位置和落点位置。反之，若不满足，则将该候选对象从跟踪列表中删除。

这种方法通过分析垂直距离的增长和水平移动的一致性来识别特定的运动模式，是一种高效的方式来区分高空抛物与其他类型的运动。通过这样的技术手段，我们不仅能提高系统对高空抛物事件的识别率，还能大幅减少误报，优化监控系统的整体性能。

## 3.3 高层危险行为检测与识别核心算法

目前阶段，本系统算法流程如图 3.6 所示。该算法参考双流网络的思想，在同一个网络中融合时序信息和空间信息，实现单阶段的检测。运行流程大致是：首先，把连续帧输入到 3D 网络中，得到提炼后的时序信息；同时，把连续帧中的关键帧输入到 2D 的网络中，提取空间特征，用于定位人体目标在画面中的位置；然后拼接以上两类特征，适当缩减通道数后，输入到 BERT 结构中，利用双向注意力机制进行各通道信息的融合；最后根据聚合结果，预测类别分数并拟合边界框参数，输出检测结果；最后根据检测结果进行自动预警。

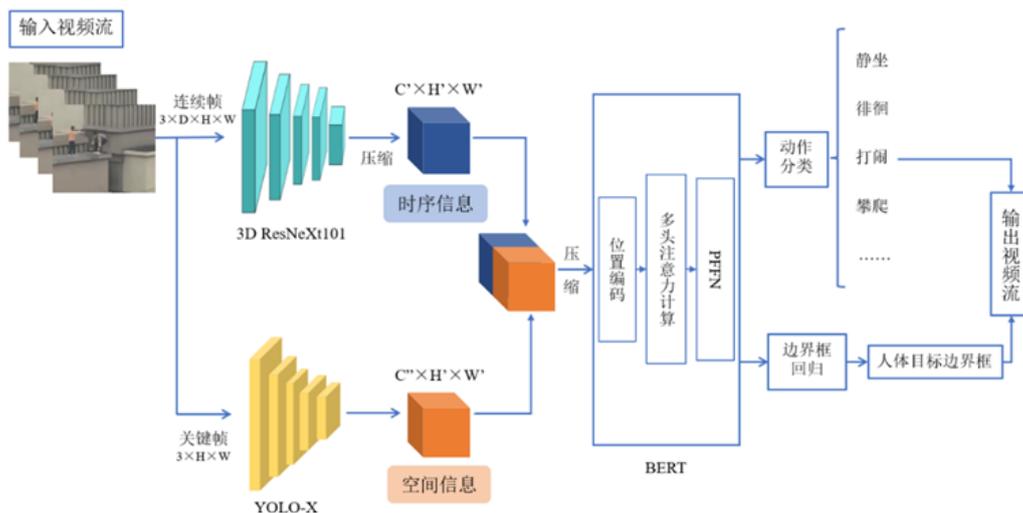


图 3.6: 高层危险行为检测与识别算法流程

### 3.3.1 提取时序特征和空间特征

在我们的核心算法中，有效地结合时序信息和空间信息是提高检测和识别精度的关键。

视频数据的预处理是确保输入数据质量的关键步骤。首先，从高分辨率的长视频中按预定的时间间隔截取视频帧。这些视频帧随后经过裁剪、旋转和平移等操作，以适应不同的视角和尺寸，增强模型的泛化能力。

为了从视频帧中提取动作的**时序特征**，我们采用了改良的 3D ResNeXt101 网络。这个网络模型通过替换二维卷积为三维卷积，能够捕捉时间上的连续性和空间上的特征。

- **3D ResNeXt101 结构概述**: 3D ResNeXt101 包含五大卷积阶段。第一阶段对特征图进行空间维度下采样，统一特征图通道数为 64。随后的四个阶段中，特征图通道数从 128 逐渐增加到 1024。残差块的配置是 3、24、36、3，特别是在第三和第四阶段，残差块较多，以实现更好的性能与参数量的平衡。

- **时间维度的处理**: 最终输出前，网络需要将时间维度压缩为 1，以便于与 2D 卷积分支的输出相适配。

同时，关键帧还需输入到基于 CSPDarkNet-53 的 YOLO-X 目标检测网络中进行**空间特征**的提取。CSPDarkNet-53 利用 CSPNet 结构优化特征传递和参数效率：

- **CSP 模块描述**: 在 CSP 模块中，输入的特征分为两路，一路通过多个瓶颈层进行深度特征提炼，另一路则进行轻量级的降维处理。两路输出最终拼接在一起，这种设计有效增加了特征的重用性，同时避免了过度的参数增加。

### 3.3.2 特征融合

1. **上采样融合**: 在骨干网络提取的高层特征图后，PAFPN 首先通过上采样（例如使用双线性插值或转置卷积）逐步将高层的特征图放大至更高分辨率的特征图尺寸。这一步的目的是捕获和传递上下文信息，使得高层的抽象特征能够与底层的细节特征进行初步的结合。
2. **注意力机制增强**: 在上采样的每一步，集成注意力机制 SE (Squeeze-and-Excitation) 模块，以增强特征的特征能力。注意力模块通过重新校准特征通道的重要性，进一步强化了有用的特征并抑制不相关的信息，从而提升特征融合的质量和精确性。
3. **下采样融合**: 经过上采样融合后，再通过下采样（最大池化）将融合后的特征图逐级降低分辨率，形成最终的特征金字塔。这一过程不仅减小了特征图的尺寸，而且进一步加强了不同尺度特征之间的交互，使得网络能够更好地处理不同尺寸的目标。
4. **输出特征层组成的元组**: 最终，PAFPN 输出三个不同尺寸的特征层，这些特征层通过二维卷积进一步处理，以确保它们与其他分支输出的尺寸一致。这样的设计使得每个特征层都能够专注于不同尺度的目标检测，增强了模型的多尺度处理能力。

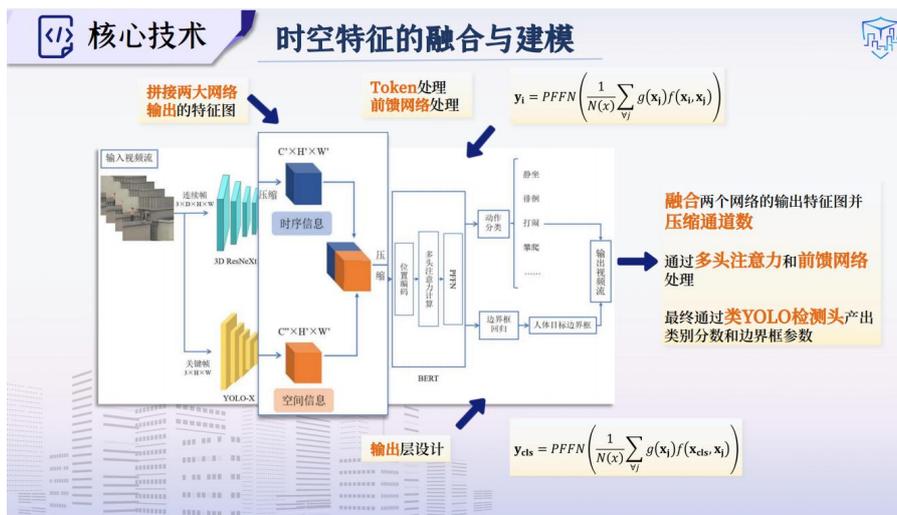


图 3.7: 同一网络融合时序信息和空间信息

通过上述的特征融合过程，PAFPN 能够有效地整合多尺度的信息，优化目标检测任务中的特征表达，从而提高检测的准确性和鲁棒性。

### 3.3.3 建模时空特征

为了有效地捕捉和建模视频数据中的时空特征，本技术采取了一种创新的多网络融合方法：

**特征图的融合和处理:** 首先，我们将来自 3D ResNeXt101 网络和 YOLO-X 网络的输出特征图进行拼接，以集成时序和空间信息。为了减少计算量和提高处理速度，接下来使用  $1 \times 1$  卷积将拼接后的特征图的通道数压缩至 512，这一处理步骤简化了后续操作，同时保持了特征的有效性。

**位置信息编码与 BERT 结构应用:** 压缩后的特征图接着被输入到 BERT 结构中，BERT 的优势在于其能够通过自注意力机制处理序列数据的依赖关系。在此过程中，网络首先学习位置编码，并将这些编码叠加到每个特征向量上，转换为  $N$  个 token ( $\mathbf{x}_1$  到  $\mathbf{x}_N$ )。此外，我们初始化一个可训练的向量——class token，它用于聚合各 token 的信息。

**多头注意力机制与信息整合:** 将  $N+1$  个 token (包括  $N$  个特征 token 和一个 class token) 输入到多头注意力结构中，执行双向自注意力计算。这允许模型从全局视角整合信息，提高特征解读的准确性。

$$y_i = PFFN \left( \frac{1}{N(x)} \sum_j g(\mathbf{x}_j) f(\mathbf{x}_i, \mathbf{x}_j) \right) \quad (3.20)$$

其中， $f(\mathbf{x}_i, \mathbf{x}_j) = \text{softmax}_j (\theta(\mathbf{x}_i)^T \phi(\mathbf{x}_j))$ ，表达了当前 token 与其他 token 的相似度， $\theta(\mathbf{x}_i)$  和  $\phi(\mathbf{x}_j)$  为线性映射，相当于 Transformer 中的 query 和 key。

**最终输出的生成:** 最终需要的结果是 class token 的输出  $y_{\text{cls}}$ ，其计算公式如下：

$$y_{\text{cls}} = PFFN \left( \frac{1}{N(x)} \sum_j g(\mathbf{x}_j) f(\mathbf{x}_{\text{cls}}, \mathbf{x}_j) \right) \quad (3.21)$$

此输出  $y_{\text{cls}}$  接着通过类似于 YOLO 检测头的卷积操作处理，以输出类别分数和边界框回归参数，完成目标检测任务。

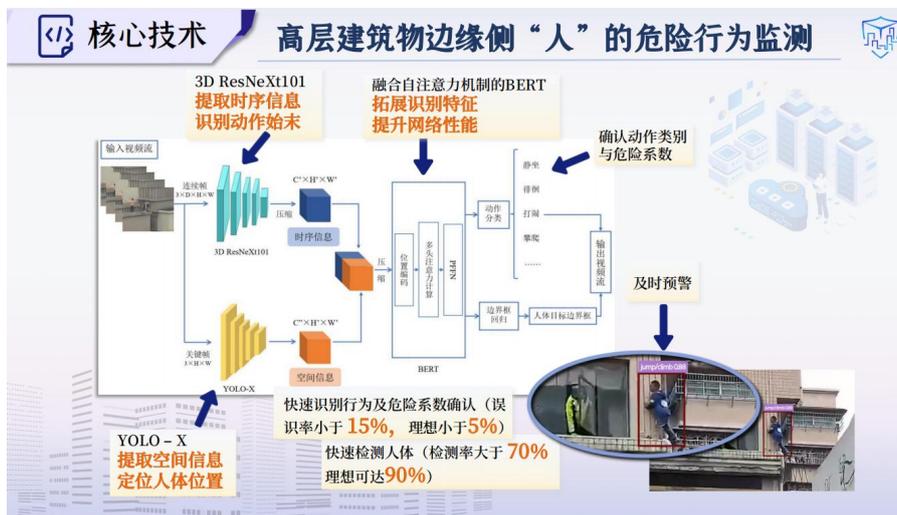


图 3.8: 时空特征的融合与建模

## 3.4 监测数据收集与处理

在部署高空抛/坠物检测与高层危险行为识别核心算法后，我们的系统将对监控数据进行全面处理，以达成 HaMonitorSentry 高层智能监测系统的功能闭环：

### 1. 实时警报机制与事件响应

- **警报触发机制：**系统利用前述的高空抛/坠物检测与高层危险行为识别核心算法实时监控并识别高空抛物和高层危险行为，一旦检测到潜在危险事件，即自动激活警报。这包括向控制中心的大屏幕发送可视警告和向管理员手机发送紧急短信通知。对于高空抛/坠物事件，会及时通知附近民众，同时启动其他的应急措施，以便事务的后续处理；对于检测出的高空危险行为，则根据其类别和行为的具体情况，对其危险行为种类，危险系数进行显示。
- **响应措施与联动：**警报响应措施设计为多层次，初级阶段包括声音警示，进一步则触发与当地应急管理部的联动，确保事件得到迅速处理。系统的视频监控与报警系统联动也使得实时数据传输至紧急响应队伍，优化了应急过程。如对于检测出的高空危险行为，进行危险系数的判定，以便后续模块触发相应的响应措施。

### 2. 数据管理与分析

- **数据采集与安全存储：**所有监控数据，包括视频文件和事件日志，都按照严格的安全协议进行加密存储，以防数据泄露。这些数据按事件类型、日期和地点分类，以便快速检索和分析。使用高标准的对称加密算法 (AES-256 [11]) 对所有传输和存储的视频文件和事件日志进行加密，确保数据在传输和存储过程中的安全性。而对于关键信息 (如管理员操作日志)，采用非对称加密技术 (RSA 算法 [13])，以保护数据访问过程中的身份验证和授权安全。
- **深度数据分析：**系统对历史监控数据进行深度学习分析，利用统计模型和机器学习技术，识别出高风险模式和异常行为趋势。通过数据挖掘技术，我们能够构建行为预测模型，预测并减少未来的危险事件。我们选择使用 ARIMA [15] (自回归积分滑动平均模型) 和季节性分解的时间序列分析技术，分析事件发生的时间规律和周期性，预测未来可能的高发时段。

### 3. 可视化反馈与操作界面

- **动态可视化展示：**分析结果通过一个交互式的图形用户界面展示，包括热力图、时间序列分析图和事件散点图，这些都是用于展示事件密集区域和高发时段的有力工具。如图 3.9 所示。我们会在收集数据的基础上运用 Apriori [6] 算法关联规则学习方法，探索不同数据项之间的关系，如特定时间和位置的高危行为关联性，从而揭示潜在的危险因素。
- **事件回放和细节检查：**系统提供高清晰度的视频回放功能，可以精确到具体事件发生的时间和地点，支持多角度查看和变速回放。这对于事后分析、追责和法律证据收集至关重要。



图 3.9: 高层危险事件可视化统计分析

通过这些综合性的技术升级，我们的系统实现了从事件的即时识别到深度分析和反馈的无缝对接，显著提升了预警准确性和响应速度。这一全面的策略不仅优化了监控系统的日常运作，也极大提升了整体的安全管理效率和效果。

## 3.5 训练与测试

面对公开可用的高层危险事件相关数据集的缺乏，模型训练和验证阶段面临了显著的挑战。为了有效克服这一障碍并提升我们网络的性能以及算法的实际效果，我们在获得必要的审批和合规之后，投入了大量资源去构建一个专门的数据集。整个过程包括了为期两个月的数据采集、预处理和分析，详细的工作流程和技术细节已在文档的“数据集的采集与预处理”部分（见 4.1）进行了全面阐述。

具体来说，我们自主构建的数据集包含了两大类：高空抛/坠物和高层危险行为。**高层人体行为数据集**包含 1,160 个视频，覆盖了 16 个不同的动作类别，这些动作类别进一步分为三大类：单人动作，人对物的动作，和人对人的动作。每个视频均在四种不同的光照强度和分辨率下拍摄，并涵盖了 16 个场景，共计 54,045 个精细标记的帧，以确保数据的多样性和丰富性。

同时，**高空抛物数据集**涵盖了 1,157 个不同的场景，包括 125,616 个视频帧。这些视频记录了四种天气状况、两种光线情况下的事件，覆盖了 17 种场景，并涉及 10 种不同抛物类别。此外，这些数据从三种摄像机拍摄角度和四种视频分辨率中捕获，确保了高度的实用性和适用性。

对于高空抛物的检测测试，这一过程包括不少于 50 次的抛物事件模拟，以及连续超过 24 小时的检测。这使我们能够全面评估模型对各类动作的识别能力，包括检出率和误检率，并据此计算出系统的整体平均精度。对于高层危险行为的检测测试，由于高层危险行为涉及显著的安全和伦理风险，实地测试无法直接进行。因此，我们采用了从互联网收集视频及现场拍摄的方式积累相关视频资料，参考抛物的测试方法进行测试和统计。具体测试结果已在文档的“作品测试与分析”部分（见 5）进行了全面阐述。

此外，通过在自主搭建的测试数据集上进行一系列的测试，我们不仅验证了模型在处理实际场景时的有效性，也为模型的进一步迭代和优化提供了宝贵的数据支持。这些工作的成功执行极大地提高了我们系统的整体性能和可靠性，为处理高风险环境中的复杂任务提供了坚实的技术基础。

## 第4章 系统实现

### 内容提要

- ❑ 数据集搭建
- ❑ 系统部署方法
- ❑ 联邦合作学习机制建设
- ❑ 多功能集成、高度迁移性的平台综合管理

### 4.1 数据集搭建

#### 4.1.1 概述

随着计算机视觉技术的迅猛发展，多种技术已被广泛应用于城市安全防范系统，包括面部识别、活动识别、车辆检测和目标跟踪等。受到这些技术的启发，我们的团队着手研究高空抛物和高层危险行为的检测技术。对于高空抛物的有效检测，构建一个包含多种天气条件、不同场景和多类抛物行为的数据集是至关重要的。然而，目前市场上缺乏专门针对高空抛物检测的标准数据集。现有的运动目标检测数据集，如 SABS[3]、CDnet 2014[19]、GTFD[9] 和 LASIESTA[4]，均无法满足高空抛物检测算法的评估需求。

为了填补这一空白，我们开发了一套全新的视频数据集，这对高空抛物检测领域具有划时代的意义。该数据集包括 1157 个视频，均在不同的城市场景下拍摄，每个建筑物的角度定义了一个独特的场景。数据集中包含总计 125616 个视频帧，每一帧都进行了详细的标注，标注示例见图4.1。

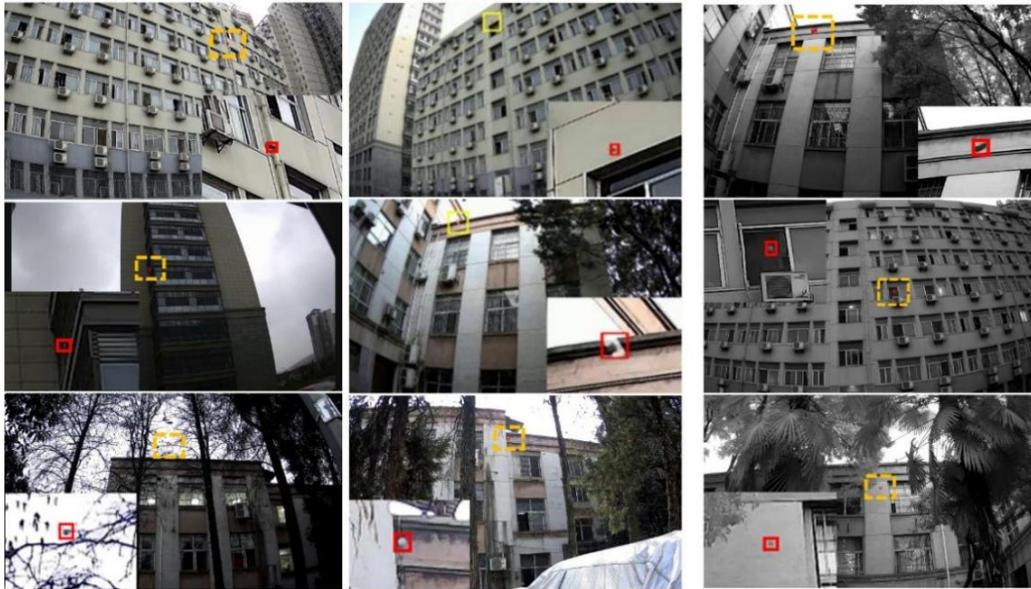


图 4.1: 高空抛物数据集的样本视频帧

高层危险行为检测与识别是基于“高层”这一特定场景的任务，一般的动作识别数据集（如 UCF101、Kinetics-400、Kinetics-700）和时空动作定位数据集（如 AVA）包含此场景的数据较少 [11]，若直接投入到本产品的算法训练中将不利于提高实际的检测效果。并且在进行本项目的数据集训练过程中，由于现有的数据集大部分只能满足单视角、无遮挡动态目标的检测，且无法满足设计的模型中对其他特定条件的要求，因此在构建数据集时采用了网上搜集和自己拍摄两种方法，对数据集进行扩展，创建了一个包含多视角、包含光照变化和着装对比度变化的高层危险行为数据库。

## 4.1.2 高空抛物数据集

在本小结，我们将从数据集的构建、数据集划分和统计、评价指标来详细介绍我们的数据集。

### 4.1.2.1 数据集的属性

本数据集广泛覆盖多种环境因素，具体属性定义如下。

#### • 物体类别

为了全面覆盖建筑物周围可能坠落的物体种类，我们收集了以下十种物品：衣服、脸盆、鞋子、厨余垃圾、水球、书籍、湿纸团、瓶子、包装袋和包装盒。

#### • 天气条件

因天气不同引起光照变化，影响高处坠落物体与背景的对比度，从而对高空抛物检测产生影响。本数据集设计了四种天气条件：晴天、阴天、多云和雨天。

#### • 光照条件

一般情况下，光照强度低于 0.04Lux 时，监控视频从 RGB 模式转为灰度模式。因此，我们提供了两种视频模式：光照强度高于 0.04Lux 的 RGB 模式和低于 0.04Lux 的灰度模式。

#### • 场景

我们的数据集包含了高空抛物可能发生的各种场景，共 17 种，包括教室、办公楼、宿舍、公寓以及在建建筑等。不同的建筑物角度代表了不同的场景。

#### • 摄像头角度

高空抛物检测的监控摄像头通常安装在建筑物外约 30 米的地方，从不同角度捕捉各楼层的情况。本数据集涵盖了三种摄像头角度：30°、45° 和 90°。

#### • 视频分辨率

监控视频具有多种分辨率，本数据集提供了四种分辨率，分别为 1280×720、1920×1080、2560×1440 和 2592×1520，以增强数据的多样性。



图 4.2: 高空抛物检测原创数据集

### 4.1.2.2 数据收集与设备配置

在获得校方的正式实验许可后，我们在校园内的高层建筑上进行了物体投掷实验。实验中使用了宽动态范围（WDR）摄像头，该摄像头配备了逐行扫描的 CMOS 传感器和点阵 LED 红外灯，专为记录高空抛物事件设计。摄像头的最高输出分辨率为 2592×1520，帧率设定为 20FPS，通过采用全局快门和高帧率的 CMOS 传感器，有效减少了由于物体高速运动所引起的图像模糊。

此外，除了我们自主收集的数据外，全球各地已有许多安装用于监控高空抛物的摄像头，这些摄像头多数采用成本效益较高的传感器。我们的数据集也广泛整合了这些监控视频，以丰富数据的多样性和实用性。部分数据收集过程如图4.3所展示。



1001.mp4	2.1 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1002.mp4	2.25 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1003.mp4	1.57 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1004.mp4	3.16 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1134.mp4	1.73 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1135.mp4	2.79 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1144.mp4	1.81 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1145.mp4	1.88 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1146.mp4	1.46 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1147.mp4	1.59 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1149.mp4	2.78 MB	Oct 2, 2022, 10:21:24 AM	rw-----
1151.mp4	2.33 MB	Oct 2, 2022, 10:21:24 AM	rw-----

图 4.3: 部分收集数据展示

### 4.1.2.3 数据预处理

在收集到的抛物视频数据上，我们的标注工作分为两个阶段完成。初步阶段包括视频分帧和使用 labeling 软件对帧进行精确标注，第二阶段则涉及生成遮罩（mask）并输出为 png 格式。

具体预处理步骤如下：

1. 开发一个分帧程序，该程序负责读取预先筛选过的抛物视频（已移除大量无抛物时间段），对视频进行分帧处理，并保存结果为图片。

frame1	类型: JPG 图片文件 分辨率: 1920 x 1080	大小: 326 KB
frame2	类型: JPG 图片文件 分辨率: 1920 x 1080	大小: 328 KB
frame3	类型: JPG 图片文件 分辨率: 1920 x 1080	大小: 329 KB
frame4	类型: JPG 图片文件 分辨率: 1920 x 1080	大小: 330 KB
frame5	类型: JPG 图片文件 分辨率: 1920 x 1080	大小: 331 KB
frame6	类型: JPG 图片文件 分辨率: 1920 x 1080	大小: 332 KB
frame7	类型: JPG 图片文件 分辨率: 1920 x 1080	大小: 332 KB

图 4.4: 视频分帧示例

2. 使用已安装的 labeling 软件对保存的帧图片进行标注。标注过程中，使用矩形边框完整包裹抛物对象，并从物体开始下落时起至消失时止进行标注。考虑到抛物体一般体积小而背景复杂，标注时常将图像放大至 300%，以提高标注精度。

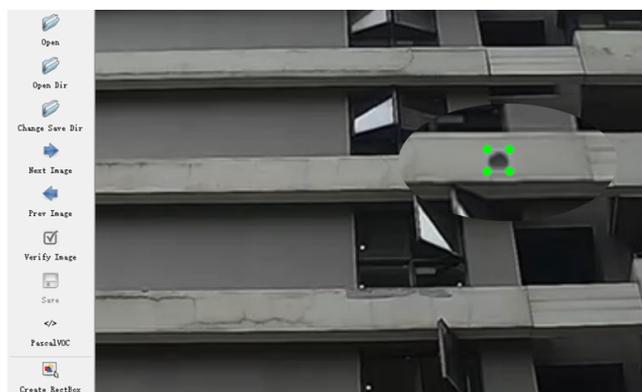


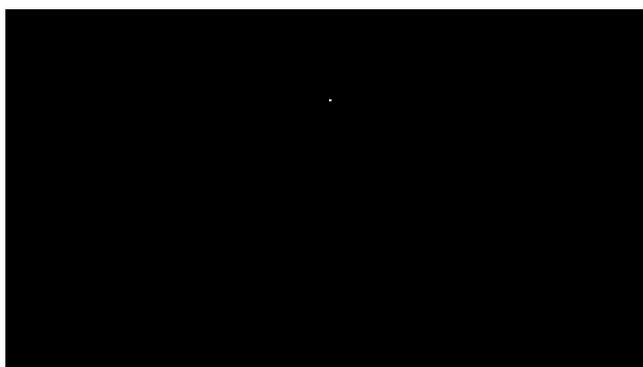
图 4.5: 实例标注过程

3. 根据不同的天气和光照条件，进行特殊的标注处理，详情如下表所示：

**表 4.1:** 天气与光照条件下的标注规范

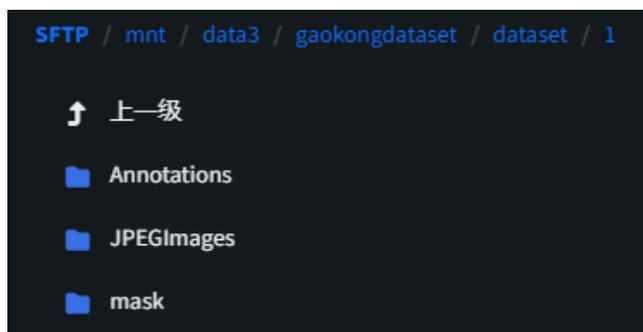
天气类型	光照强度	标注类型
晴天	强	$S_1$
晴天	强	$S_2$
阴天	弱	$C$
夜晚	无光	$N$

4. 使用一个专门编写的程序，读取标注生成的 xml 文件，并从中提取出 mask 文件，保存为 png 格式。



**图 4.6:** 从 XML 文件提取的 Mask

5. 将相关文件组织归档，确保每种文件类型存储在指定的目录下。具体地，将标注文件 (xml) 存储于 Annotations 目录，帧图片 (JPEG) 存储于 JPEGImages 目录，而 mask 文件 (png) 则存放在 mask 目录下。



**图 4.7:** 文件存储结构示意图

#### 4.1.2.4 数据集的划分和统计

##### 数据集的划分

为确保算法的泛化能力，我们根据视频分辨率、场景、光线条件、天气状况和抛物类别五种标签细致地划分了训练集、验证集和测试集。每个数据集均涵盖了所有类别的视频，并保证了数据之间的互斥性，确保不同数据集之间没有重叠。

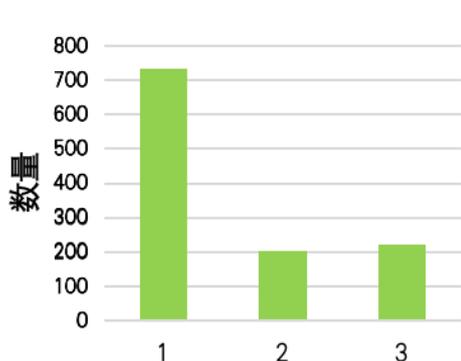
##### 数据集的统计

本数据集综合收录了多种分辨率、场景、光照条件、天气状态以及抛物类别的视频。为全面评估高空抛物检测系统，我们提供了详尽的数据统计信息。如表4.2所示，数据集中不同面积范围的抛物数量及其在数据集中的比例明确列出，其中小尺寸抛物占主导，单个抛物的平均面积占比约为  $8.45 \times 10^{-5}$ 。

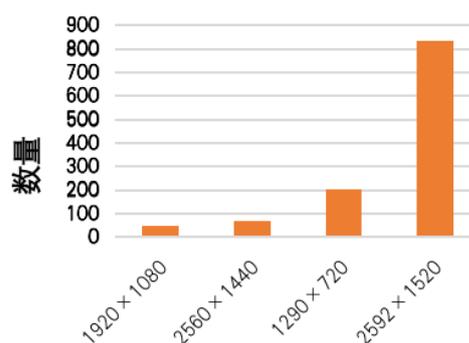
表 4.2: 高空抛物数据集各面积范围数量与占比

面积范围	数量	占比 (%)
$0^2 \leq \text{面积} \leq 5^2$	8,209	14.30
$5^2 \leq \text{面积} \leq 10^2$	16,576	29.06
$10^2 \leq \text{面积} \leq 30^2$	30,030	51.96
$30^2 \leq \text{面积} \leq 50^2$	2,601	4.25
$\text{面积} \geq 50^2$	250	0.43

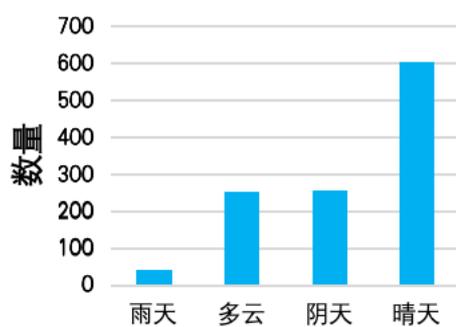
数据集的详细划分过程以及属性分类在图4.8中有进一步的阐述，以支持后续模型训练、验证和测试。



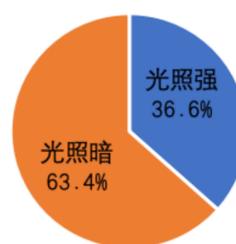
(a) 数据集划分



(b) 分辨率



(e) 天气



(f) 光照强度

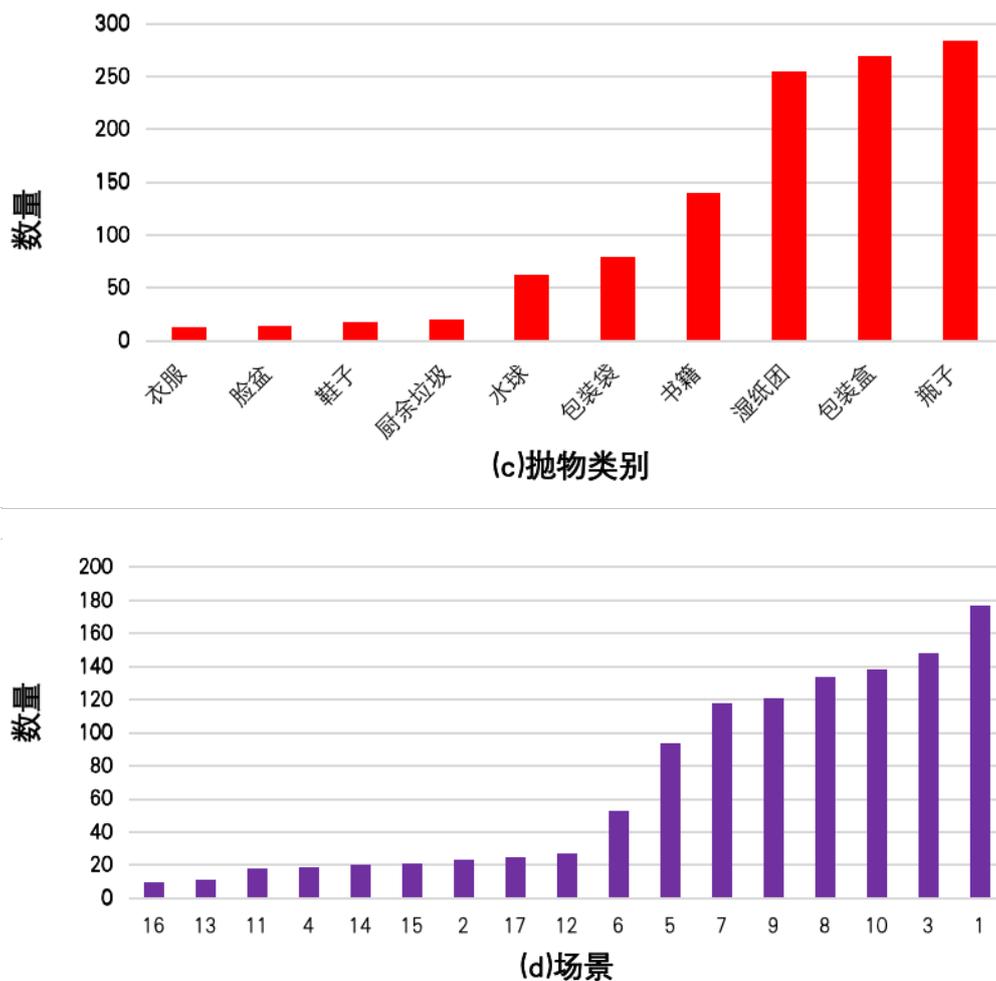


图 4.8: 数据集的统计数据

### 4.1.3 高层危险行为数据集

我们通过在网络上搜寻查找扩充数据集的过程，我们对智能识别检测目标系统有了一定的了解，这样能够发现初期系统的不足并进行改善，优化自己的算法。同时，我们可以通过对现有智能识别检测系统采取的数据集进行分析，采用部分可以兼容自己设计的新型模式识别系统的数据，总结优化旧数据集并在此基础上进行扩展改良，建立更普适、更具有通用性的开源数据集。因此，本团队基于自行收集的视频，构建了面向高层场景的时空动作定位数据集。

#### 4.1.3.1 数据集的基本情况

我们在初期收集的数据对于模型的训练相对来说不够普适，主要还是计算了特殊数据集的参数并进行识别，对于日常生活中需要实际应用到的监测场景来说泛用性不够。因此，进行数据采集时，我们有意识地从身边调取了一些可以公开使用的监控视频，来达到补充数据集的目的。考虑到实际应用场景中情况的复杂性，扩展数据集时主要关注到了三个方面，即着装变化、光照变化和视角变化，并通过在实验室场景下进行自主拍摄完善了数据集的内容。最终搜集到的数据包含多个多视角监控视频，合理训练后应能较为有效解决多视角的问题，且成功构建了针对性强，实用性强的数据集，并按照专业的评估准则对创建的数据集进行评估，既满足本项目的研究需要，也服务于其他研究团队的相关研究。

高层人体行为一共有 1160 个视频，包含 16 个不同的动作类别，16 类动作分为三大类（单人动作，人对物

的动作，人对人的动作)，同时包括 4 种光照强度，4 种不同分辨率的视频，16 个场景，54045 个标记帧，具体分类如下：

- 单人动作：弯腰、跳跃、坐、站、跑、走、爬。
- 人对物：打电话、拿/提着某物、看/玩手机、抛物、开（窗/门）。
- 人对人：打某人、抓住某人、挥手、和某人说话。



图 4.9: 高层危险行为原创数据集

### 4.1.3.2 数据集获取方式

该数据集的视频数据主要有两个来源，分别是监控视频收集与实地采集。

对于监控视频收集，先把视频大致分为新闻报道类、电影剧集类、短视频类等，分别布置相关人员去特定平台搜索和下载，同时在允许的范围内调取身边的监控。在搜集过程中把控了视频的质量，清晰度、光照情况、镜头晃动程度等满足了基本的机器识别要求。同时，搜集的视频里，动作类别、人体类型、拍摄角度等尽量多样，增强数据的泛化性。

而对于实地采集，我们团队借用了不同的楼层实施高层动作的拍摄实验，分派不同的实验人员在阳台/窗边施行各类动作，相关人员在地面架设英飞拓电子摄像头进行录像采集。实地采集的标准与网上收集相同，主要是对网上数据较缺乏的动作类别进行补充。

网上搜集和实地采集的视频数据分别如图所示：

frame1.jpg	🔄	2022/9/23 16:54	JPG 图片文件	365 KB
frame2.jpg	🔄	2022/9/23 16:54	JPG 图片文件	365 KB
frame3.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame4.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame5.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame6.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame7.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame8.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame9.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame10.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame11.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame12.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame13.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame14.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame15.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame16.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame17.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame18.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame19.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB
frame19.xml	🔄	2022/9/23 16:56	XML 文档	1 KB
frame20.jpg	🔄	2022/9/23 16:54	JPG 图片文件	366 KB

图 4.10: 部分拍摄数据展示



图 4.11: 高层危险行为数据集展示

### 4.1.3.3 数据预处理

为了使复杂场景的智能识别检测模型能够先确定一些参数，对检测数据集进行运算处理并识别，我们先挑选了一部分比较符合真实场景，但不存在着装、光照、视角等复杂变化的视频作为参考数据集，并预先进行标注处理，为其中的人员活动状态打上标签，以便模型进行学习分类运算。

在标注处理的过程中，我们选用 labeling 和 vott 软件对视频进行逐帧标注处理。一个视频对应一个文件夹，里面包含其关键帧、原视频 (.mp4) 以及预生成标签 (.json)。标注过程如下：

1. 打开所发的 via3 标注工具文件夹，用浏览器打开其中 via\_image\_annotator.html。
2. 加载待标注视频的关键帧，点击红色框。
3. 加载预生成的标签文件，点击绿色框。



图 4.12: 标注文件加载操作

4. 加载完后，删除不需要框，调整需要调整的框，并打标签。如图4.15所示。



图 4.13: 标注过程

5. 保存打好标签后的 json 文件。把所有标注好的视频的标签文件打包在一个文件夹下发送，如 6 所示。

5_dance.json	2022/5/5 20:30	JSON File	7 KB
4_work.json	2022/5/5 20:30	JSON File	7 KB
3_work.json	2022/5/5 20:30	JSON File	7 KB
2_work.json	2022/5/5 20:30	JSON File	7 KB
1_stand.json	2022/5/5 20:16	JSON File	7 KB

图 4.14: 数据集结果展示

 video1-demo 时长: 00:00:04	帧高度: 1920 帧宽度: 1080
 video1-original.mp4 时长: 00:00:04	帧高度: 1080 帧宽度: 1920
 video2-demo 时长: 00:00:12	帧高度: 360 帧宽度: 640
 video2-original 时长: 00:00:13	帧高度: 360 帧宽度: 640
 video3-demo 时长: 00:00:09	帧高度: 1280 帧宽度: 720
 video3-original 时长: 00:00:10	帧高度: 1280 帧宽度: 720
 video4-demo 时长: 00:00:10	帧高度: 636 帧宽度: 1332
 video4-original 时长: 00:00:10	帧高度: 636 帧宽度: 1332

图 4.15: 数据集处理后的部分 demo

## 4.2 系统部署方法

### 4.2.1 系统总体框架

HaMonitorSentry 高层智能监测系统框架主要分为 5 个主要模块：

#### 1. 监控输入端：

- 使用分辨率可达 70FPS 的超清摄像头。
- 采用多台摄像机接力监控的方式。
- 根据应用场景的实际情况调整安装方式，确保高层监测的精确度。

#### 2. 后台处理端：

- 采用双算法监测方式，结合团队核心技术监测高空抛物与高层危险行为。
- 使用 Deepstream 技术实现媒体流的推拉，支持硬件加速与并行处理，确保实时监测与结果分析。

#### 3. 用户操作端：

- 处理后的视频数据传至用户操作端。
- 实现及时警报，回溯取证等功能。
- 实时视频监控与报警联动采用弹窗提示、短信推送和报警联动三种预警方式，确保在事故发生后 5 秒内发出预警。

#### 4. 数据分析端：

- 视频数据用于提取重要信息，如高层事故的时间段、地点、事件类型等。
- 对历史高空坠物和高层危险行为事件进行可视化统计分析。
- 提出治理建议，助力智能安防。

系统框架图如图 4.16 所示

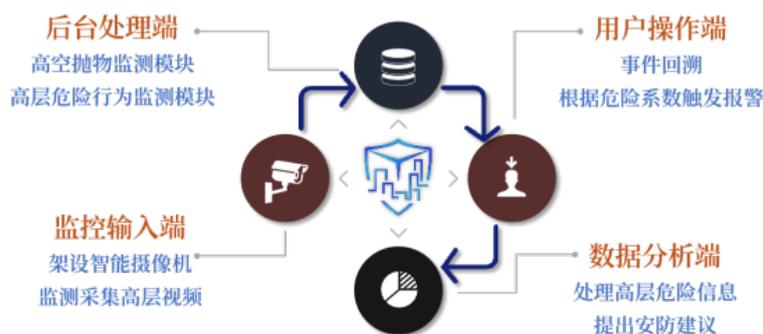


图 4.16: 系统框架图

## 4.2.2 软件开发环境

表 4.3: HaMonitorSentry 高层智能监测系统平台运行环境表

软件类别	软件
开发环境	Python: 3.8.10、PyTorch: 1.10.0+cu113、PyQt5: 5.15.10、OpenCV: 4.9.0.80
运行环境	Linux
开发语言	Python

## 4.2.3 硬件部署方案

HaMonitorSentry 高层智能监测系统的硬件布局主要落实在视频输入端，通过部署高清监控摄像头收集实时和非实时视频数据，用于探测楼层间的高层危险事件。

系统各模块之间采用 Deepstream 技术进行视频流的推拉操作，从视频输入端接收的视频流被推送至远程服务器以进行高空抛物和高层危险行为的分析处理，处理后的结果再推送回本地显示。同时，检测结果被发送到消息队列中，便于在需要展示历史数据时快速访问。

关于摄像机的安装策略，考虑到高层建筑监控的特殊性，摄像机应采用仰视方式安装，并保持至少 20 米的距离以避免仰角过大，这可能导致楼层高度的识别困难。每台摄像机负责监控 10 到 12 层，对于更高的建筑，可使用多台摄像机进行分层监控。此外，根据安装位置和楼层高度的具体情况，可选择合适的摄像机焦距，并在现场调整仰角和焦距以达到最佳监控效果。



图 4.17: 系统监控输入端配置

## 4.3 多功能集成、高度迁移性的平台综合管理

### 4.3.1 软件界面设计

HaMonitorSentry 高层智能监测系统的软件界面由三个关键模块组成：实时视频监测、历史数据统计和视频智能分析。这些模块分别负责监控摄像头的实时视频流分析、本地视频上传与检测处理、以及高空抛物与高层危险行为的历史数据统计与可视化展示。

系统界面基于 Ubuntu 20.04.3 LTS (Focal Fossa) 操作系统开发，使用 Python 语言编程，主要 GUI 框架为 PyQt5。视频数据的处理和可视化则依赖于 opencv 库，而数据存储管理则使用 OBS 服务。

界面设计强调模块化结构，每个功能模块都可以单独开发和更新，集成在统一的用户界面中。这种设计不仅便于多人团队独立工作，提升了系统的拓展性和维护性，还使得用户操作更流畅连贯。

### 4.3.2 实时视频检测模块

#### 模块概述

实时视频监测模块，如图4.18所示，连接实地摄像头以实现实时监测功能，能智能识别高空坠物和高层危险行为。此模块支持在事故发生后5秒内通过平台弹窗警告、短信通知管理员以及视频监测与报警系统的联动，从而快速响应高层安全威胁。

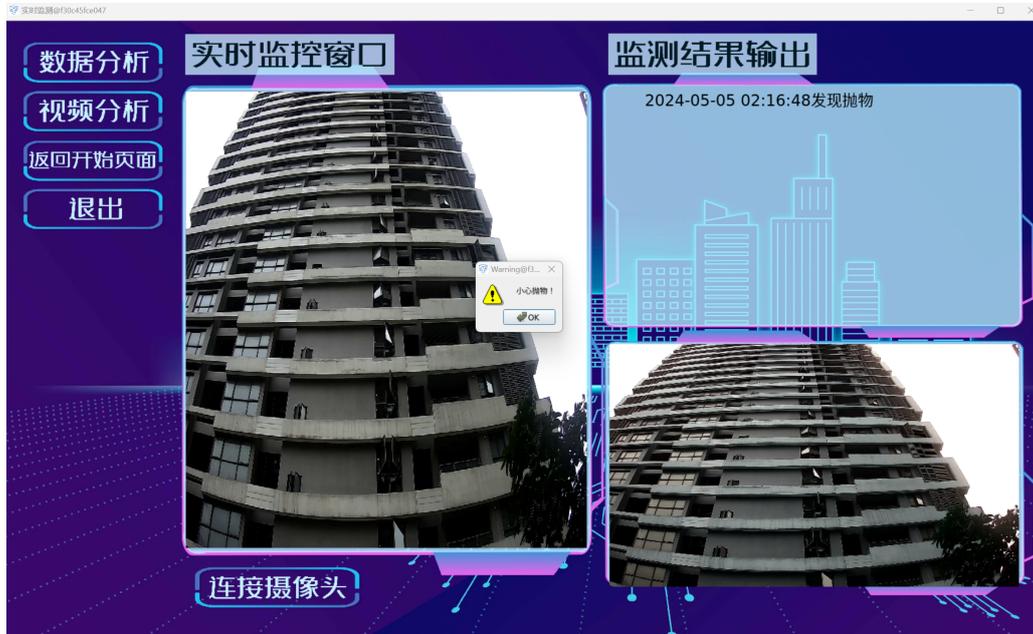


图 4.18: HaMonitorSentry 监测平台实时检测模块

#### 模块功能

通过连接实地摄像头，HaMonitorSentry 监测平台的实时视频检测模块能传输实时视频数据。当监测到高空抛物或危险行为时，模块将自动弹出警告并在界面右上方显示监测结果。

#### 界面元素

表 4.4: 实时视频监测模块功能描述

界面元素	界面事件	功能名称	限制	说明
数据分析	单击	跳转到数据分析界面		
视频分析	单击	跳转到视频分析界面		
退出	单击	退出应用平台	无	无
连接摄像头	单击	连接实地摄像头，显示实时画面		

### 4.3.3 历史数据统计模块

#### 模块概述

历史数据统计模块如图4.19所示，旨在记录并处理通过实时视频监控获得的数据。此模块不仅加工和分析这些数据，还能通过图表展示，提供居民行为的总体分析，并便于事后追溯和源头分析。

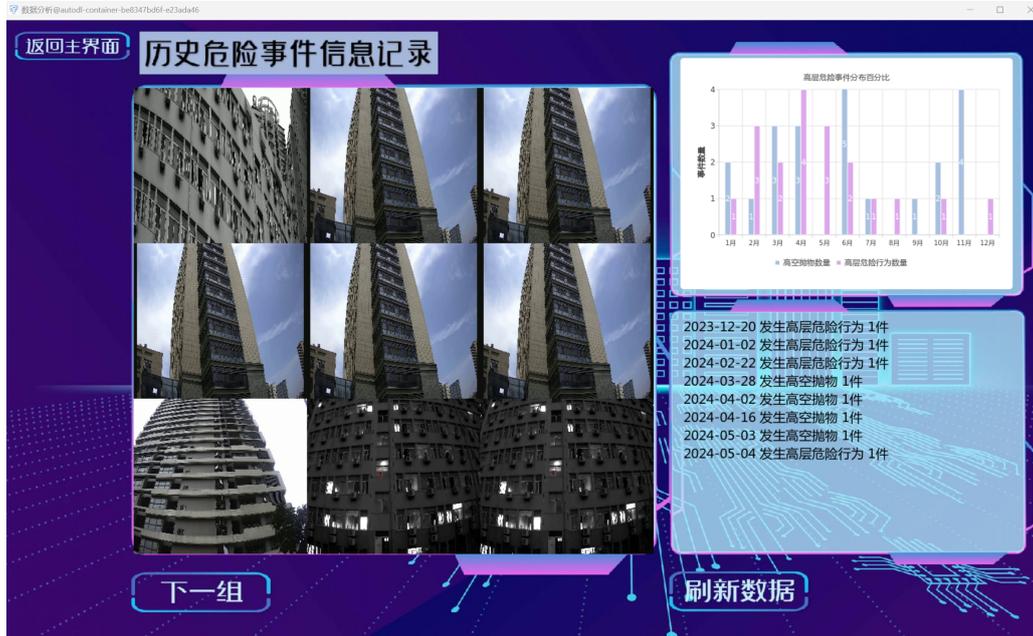


图 4.19: HaMonitorSentry 监测平台历史数据统计模块

#### 模块功能

历史数据统计模块能显示最近几次的抛物事件监测画面，并记录每次事件的具体时间。通过柱状图和文字描述，该模块详细展示了抛物和危险行为的总体趋势，使信息呈现清晰、易于理解，有助于物业和管理人员进行有效管理和追责。

#### 界面元素

表 4.5: 历史数据分析模块功能描述

界面元素	界面事件	功能名称	限制	说明
下一组	单击	更新 9 组监测图片		
刷新数据	单击	刷新高层危险事件信息及柱状图		
返回主界面	单击	返回到平台主界面	无	无
历史数据统计	无	统计并绘制高层危险事件柱状图		

### 4.3.4 视频智能分析模块

#### 模块概述

视频智能分析模块，展示于图4.20，支持用户导入特定的坠物或高层行为视频文件，对其进行深入的智能分析。该模块不仅能进行实时监测分析，还能将检测结果进行可视化展示，并提供视频回放功能，以便进行事后分析和责任追溯。



图 4.20: HaMonitorSentry 监测平台视频智能分析模块

#### 模块功能

该模块允许用户上传非实时视频，随后利用先进的高空抛物和高层危险行为检测算法对视频内容进行详细分析。分析过程包括信息捕捉、位置标注和画面呈现。模块还配备了信息回放功能，为事后分析提供支持。

#### 界面元素

表 4.6: 视频智能分析模块功能描述

界面元素	界面事件	功能名称	限制	说明
视频上传	单击	选择并上传视频文件		
开始检测	单击	开启视频内容的算法检测		
停止检测	单击	终止算法检测过程	无	无
返回主界面	单击	返回到平台主界面		
回放信息	单击	回放已检测的视频内容		

# 第5章 作品测试与分析

## 内容提要

- 性能评估指标
- 高空抛/坠“物”检测
- 高层建筑物边缘侧“人”的危险行为检测

## 5.1 性能评估指标

### 5.1.1 检出率 (Detection Rate)

检出率衡量的是系统在所有正事件中正确识别的比例。在高层建筑安全监测中，这意味着系统能够准确地识别出如抛物、人的危险行为等潜在威胁。一个高的检出率确保了所有可能造成伤害或安全隐患的行为都被系统有效捕捉，从而及时采取必要的预防或应对措施，减少事故的发生。因此，这是衡量系统有效性的关键指标。数学表达式为：

$$\text{检出率} = \frac{\text{正确检测到的正事件数量}}{\text{实际正事件总数}} \quad (5.1)$$

### 5.1.2 误报率 (False Alarm Rate)

误报率衡量的是系统在所有检测中错误标记为正事件的比例。对于高层建筑安全监测系统来说，一个低的误报率意味着减少了因错误警报引起的不必要的干扰和资源浪费。例如，在监测高空抛物或边缘侧“人”的危险行为时，频繁的误报可能导致监控人员对真正的警报不予以重视，降低系统的实用性和信赖度。控制误报率能够提升系统的运营效率和接受度。数学表达式为：

$$\text{误报率} = \frac{\text{错误检测的事件数量}}{\text{检测次数总数}} \quad (5.2)$$

综上所述，这两个指标共同构成了评估高层安全监测系统性能的基础。它们不仅影响系统的直接操作效果，也影响最终用户的满意度和系统的广泛采用。因此，精确测量和优化这些指标是设计和实施高层安全监测方案的重要组成部分。

## 5.2 高空抛/坠“物”检测

### • 测试方法：

测试协议要求在相同的实验条件下（包括统一的检测系统、相同的参数配置、一致的测试场景）进行至少 50 次的抛物事件检测，连续监测时间不少于 24 小时。为了操作方便，可将抛物事件分为两部分录像：第一部分包含集中的抛物事件视频，第二部分则为连续数小时内未发生抛物事件的视频。将第一部分视频中正确识别的抛物次数（连续多次检测同一事件视为一次）与实际抛物事件的次数比较，计算出检测的正确率。同时，将第一部分和第二部分视频中错误检测的次数与总检测次数的比值计算出误检率。此外，抛物次数与监测时长应保持适当比例，即 50 次抛物对应 24 小时连续监测，增加抛物事件时，相应增加监测时长。

### • 拟定的精度指标：

- 在良好光照、背景干扰小、空中无遮挡条件下，能够快速、高精度地检测和追踪 10 种以上的抛物和坠物，标准参数下的检出率应高于 85%，误报率控制在 5% 以下；
- 在光照不足、背景干扰小、空中无遮挡条件下，同样能够快速、高精度地完成抛物和坠物的检测和追踪，检出率应高于 75%，误报率控制在 5% 以下；

- 在良好光照、背景干扰大、空中无遮挡条件下，检测和追踪 10 种以上的抛物和坠物，检出率应高于 70%，误报率保持在 5% 以下；
- 在光照不足、背景干扰大、空中无遮挡条件下，检测和追踪 10 种以上的抛物和坠物，检出率应高于 65%，误报率不超过 10%；
- 总体性能不低于行业领先企业，符合所有公开的检测标准。对于安装角度的适应性，向上安装时应完全满足上述标准，向下安装时可适当降低精度要求，指标可比向上安装低 5 至 10 个百分点。

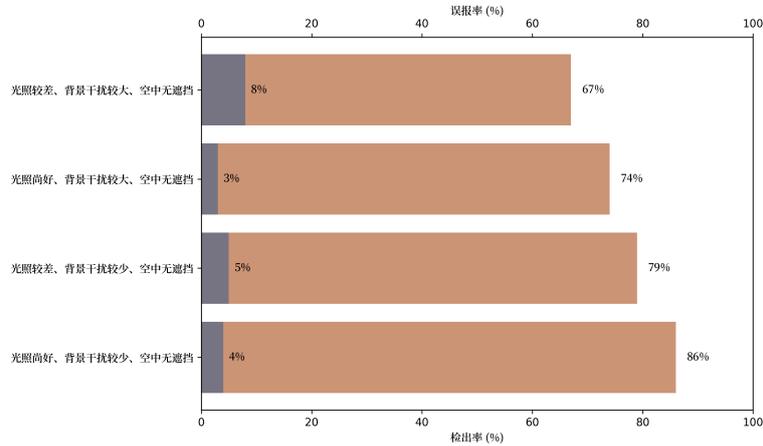


图 5.1: 各复杂状态下高空抛/坠“物”检测系统精度

### 5.3 高层建筑物边缘侧“人”的危险行为检测

#### • 测试方法:

鉴于高层危险行为难以在现场直接测试，我们通过网络资源和现场录制的方式，收集了多个高层建筑边缘行为的视频样本。这些视频资料用以模拟实际情况，进行危险行为的检测和统计分析，类似于高空抛物事件的测试方法。此过程中，我们计算了不同动作类别的检出率和误检率，及其平均精度。

#### • 拟定的精度指标:

- 能够迅速检测出处于建筑物边缘的人，并准确判断是否存在危险行为；默认设置下的检测率应超过 90%，误报率应控制在 10% 以下；
- 快速准确地识别多种危险行为，包括但不限于天台边缘静坐、身体大部分伸出窗外、边缘侧徘徊、攀爬、抛物倾向、跳楼倾向等；识别率在默认参数下应达到 90% 以上，误识率应低于 10%；
- 场景设置参考高空抛物的测试环境，最终交付的系统性能应至少匹配或超越行业领先企业的标准，满足所有公开的检测标准。

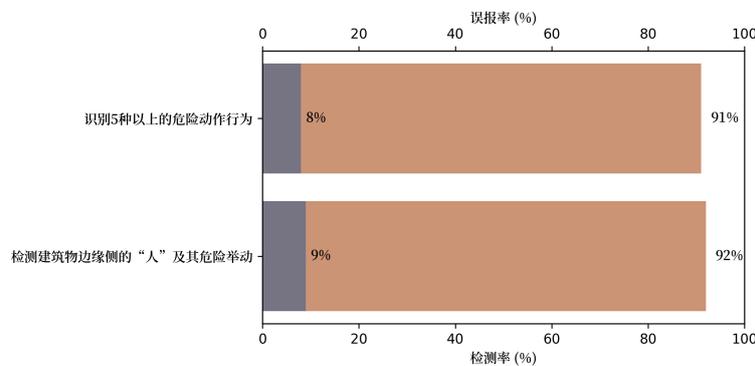


图 5.2: 高层建筑物边缘侧“人”的危险行为检测与识别系统精度指标

## 第6章 作品总结与展望

### 内容提要

- 作品特色与创新点
- 应用推广
- 成果支撑
- 作品展望

### 6.1 作品特色与创新点

本系统是一个先进的综合性高层智能监测解决方案，整合了最新的监控技术和尖端人工智能算法，专为高层建筑中的安全监控与意外预防设计。系统的核心由两个主要的创新算法模块组成：**复杂环境下的高空抛坠物检测系统**和**高层建筑物边缘侧人员危险行为检测与识别系统**，这两个系统都针对具体的安全风险设计，能够有效地识别并预警潜在的安全威胁。

#### 6.1.1 高度集成的算法与功能

系统的设计融合了多种传感技术和复杂的数据处理方法，不仅显著提升了监测结果的准确性，还增强了系统整体的稳定性和可靠性。这种高度集成的算法架构使得系统能够在多变的高层环境中提供全面的安全保护，有效降低事故发生的概率。

此外，系统还包括一个先进的数据处理和建模模块，这一模块能够深入分析收集到的监测数据，如通过挖掘高危行为发生的时间和地点，系统可以预测未来的风险并自动调整监控重点，从而实现事前的风险防控。

通过整合各类事件的相关数据（如事件发生的时间、地点及周围环境条件），本系统不仅提升了高层建筑的安全管理能力，还强化了对应急响应和责任追踪的支持，大大提高了安全监管的效率和效果。

#### 6.1.2 高层安全双重监测技术

系统采用的高空抛物和高层危险行为检测技术集成了最新的语义分割和知识蒸馏方法，这些技术能够在复杂的城市高层环境中快速准确地识别潜在风险。背景差分法和帧差法的结合使得系统能够适应快速变化的环境条件，提高目标检测的准确度。此外，引入的注意力机制进一步优化了系统对动态信息的处理能力，显著降低误检率，同时加快了特征学习的速度和网络的收敛。

在高层危险行为检测方面，系统采用了创新的双流网络结构，结合二维与三维数据处理，通过自注意力机制精确识别和定位危险行为，确保了行为动作的时空一致性，并显著提升了检测精度和系统响应速度。

#### 6.1.3 高质量的原創数据集

为了增强系统的实用性和市场竞争力，我们专门构建了两个高度适配算法模型的原創数据集。该数据集从多角度收集了涵盖光照变化和不同气象条件下的高空抛物和高层危险行为数据，提供了丰富的场景覆盖和数据多样性。

通过结合互联网搜集和现场拍摄的方法，我们不仅扩展了数据集，还提高了数据的质量和适用性。这一过程使我们能够在早期发现系统的潜在缺陷，并据此优化算法的设计和性能。

我们还对比分析了现有的智能识别系统所使用的数据集，并整合了兼容我们系统的优质数据资源。通过优化和改进旧的数据集，我们创建了一个更广泛适用且全面的开源数据集，旨在为智能安全监控领域提供更优质的研究和应用基础。

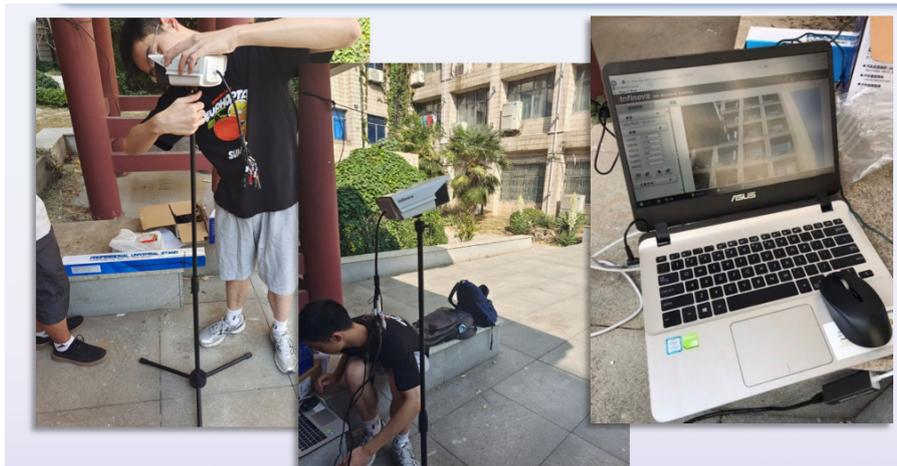


图 6.1: HaMonitorSentry 高层智能监测系统初步部署测试

## 6.2 成果支撑

目前本团队已基于本作品取得了丰富的成果，具有显著的学术及应用价值。首先，团队成员以第一作者身份成功申请了 1 项国家发明专利，名为《基于增量式背景建模与多目标跟踪的高空抛物物检测方法》。该专利目前已进入实质审查阶段，该方法通过创新的增量式背景建模和精确的多目标跟踪技术，有效提高了高空抛物行为的检测准确性与效率。

除此之外，团队还申请了 3 项软件著作权，涵盖了《高空抛物智能监测和分析平台》、《多场景高空抛物分析平台 V1.0》以及《基于运动目标检测和语义分割的高空抛物物检测系统》。这些软件工具的开发，极大地丰富了高层安全监测技术的应用场景和实用性，为我们的监测系统开发提供了重要的技术支撑。

在学术领域，团队正在投稿一篇学术论文《Class Distance Distillation for Improved Semantic Segmentation》至 CCF-A 类会议 ACMMM。该论文提出了一种新的语义分割改进方法，通过类距离蒸馏技术显著提高了模型的泛化能力和精度，此方法已集成到我们的高层智能监测系统。

更进一步，团队开源了两个原创数据集 FaDE 与 EBPersons，专注于高空抛物和高层危险行为两大角度，这些数据集在高层危险事故监测领域具有开拓性的意义。它们囊括了多样的高层场景，包括多视角、不同光照变化以及对比度变化等，为高层安全监测的研究人员提供了宝贵的资源，以进一步探索和解决高层安全监测问题。

总体而言，这些成果不仅展示了本团队在高空抛物检测技术领域的深厚研究基础，也为相关学术研究和实际应用提供了坚实的支撑，推动了高空安全监控技术的发展和革新。此外，本团队的成果在推动社会安全管理、降低公众风险方面 also 具有重要的实际意义，为智能安防建设提供了有效的科技支持。



图 6.2: HaMonitorSentry 相关成果支撑

## 6.3 应用推广

### 6.3.1 深圳英飞拓智能技术有限公司合作落地试点

目前，本作品在武汉大学测绘遥感信息工程国家重点实验室的支持下，已与深圳英飞拓智能技术有限公司建立合作关系。双方共同参与“英飞拓-武汉大学智慧城市时空信息技术研究中心”的联合研发工作，旨在开发先进的算法模型，以提升对高层危险事件的实时监测和精准预警能力。

本产品填补了国内多项技术空白。目前，我们计划将该高层智能监测系统应用于深圳湾超级总部基地以及其他数个园区，经测试可检出超过 90% 的危险事故，以提升这些区域的安全性。预期将实现累计新增销售额达 160 万元，并预计新增利润 32 万元，并带来广泛的社会效益和显著的科技示范效益。

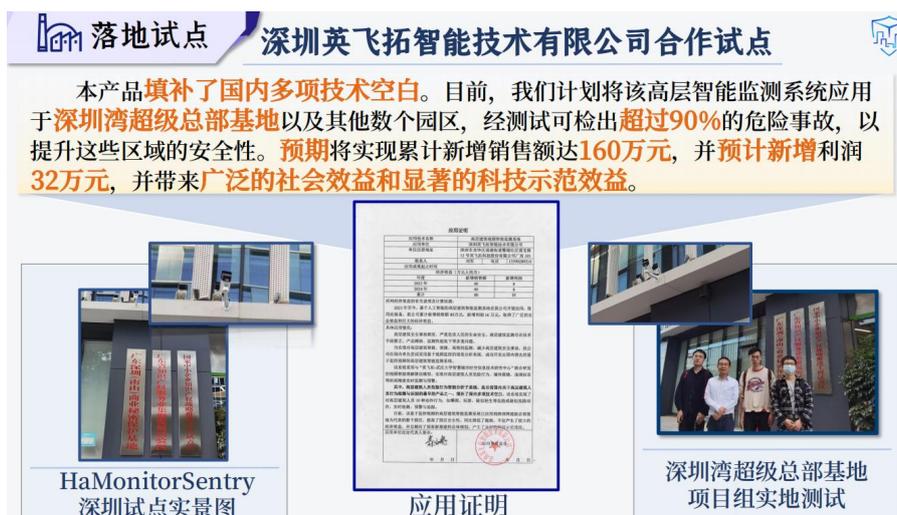


图 6.3: 深圳英飞拓智能技术有限公司合作落地试点

### 6.3.2 应用方向

本团体作品市场应用方向如下：

1. 社会责任驱动的技术作品：在“十四五”期间，国家对智能化安防技术的需求不断提高，并在政策指导文件中对智能家电、智能照明、智能安防和智能视频监控系统等行业的发展方向进行了具体阐述。我们的系统应运而生，响应了国家对智能安防技术的战略部署。此外，随着国家对建设更高标准的“平安中国”的目标的提出，本系统的推广和应用将直接支持智慧城市和新型平安城市项目，从而促进智慧化社会治理的实现。
2. 科技创新推动的技术作品：本系统采用了最新的深度学习和计算机视觉技术，这些技术的运用不仅提升了视频监控的效能，也扩展了其在安防领域的应用范围。系统通过先进的算法优化和技术创新，有效填补了市场上的技术空白，为行业提供了高效率和高准确度的智能解决方案。

### 6.3.3 应用效益分析

1. 促进社会治理和法律实施：本系统通过其先进的监控功能，不只是抑制了高空抛物等危险行为，增强了公共安全，还通过精确的事件记录和分析，提供了重要的法律执行参考。系统可以追踪事件源头并预测行为轨迹，这对于法律取证和责任归属判定至关重要，从而帮助解决法律执行中的困难，确保法律的有效实施。此外，系统的技术解决方案还能针对社区的其他关键安全问题如入侵监测、公共聚集警告、交通违规等提供支持，全面提升社区管理水平。
2. 支持新基建与智慧城市发展：随着智慧城市和新型基础设施建设的推进，本系统的开发与应用顺应了数字化转型的潮流。特别是在城市老旧小区改造中，系统的部署不仅提升了居民的安全感，还为安防设施的

现代化升级提供了技术保障。通过技术赋能，系统助力实现社区安全的智能化管理，推动了安防产业的创新升级。

- 提升公共诚信与社会责任感:** 通过实时监控和精确预警，本系统不仅预防了高空抛物和其他高危行为，还促进了社会诚信体系的建设。系统的实施鼓励公民遵守社会规范，承担相应的法律和道德责任，从而减少违法行为并增强公众的安全意识。同时，这种监控系统的存在也为公共安全教育提供了实践平台，增强了居民的社会责任感。
- 增进居民幸福感和社会和谐:** 通过减少高层建筑中的安全事故，本系统直接提高了居民的生活质量和安全感。有效的安全监控和事故预防措施减轻了居民对高处坠落事故的担忧，为儿童和工作人员提供了更安全的生活和工作环境。此外，系统的实施还促进了社区内的和谐，为构建更加稳定和安全的居住环境提供了支持，从而直接反映了以人为本的发展理念，增强了人民对未来生活的期待和满意度。

## 6.4 作品展望

在综合了实地测试反馈及内部讨论后，本团队已制定出后续升级和发展的详细规划，以进一步提升系统的性能和功能：

- 实时性能优化:** 鉴于高层危险事件对实时监测的迫切需求，我们的系统已能处理高达 25FPS 的视频帧速率。为了更快地响应并处理这类突发事件，未来我们计划在不牺牲检测精度的前提下，通过优化网络结构和引入多线程并发处理技术，显著提升系统的响应速度和处理效率。此外，我们将与部分小区合作，以低价或免费的方式提供服务，不仅为了测试和完善产品功能，也为了根据用户反馈进一步优化性能，确保监控的连续性和实效性。
- 加强数据处理和分析:** 随着团队技术的成熟及市场拓展，我们将开发具有高性价比的主要产品线。计划引入更加先进的机器学习算法来深入分析监控数据，提高事件检测的准确性，并能基于历史数据识别潜在的高风险区域，即能够从大量的视频数据中识别并预测那些频繁发生高空抛物或其他危险行为的具体区域。进一步的应用可能包括在这些高风险区域部署更多的摄像头或采用更高的监控频率，同时根据风险评估自动调整摄像头的焦距和角度，以优化监视覆盖面。这些措施将增强我们的智能基建能力，通过先进的数据分析和预警系统，显著提升高层安全。
- 功能扩展:** 展望未来，我们将利用积累的经验 and 资源大力投入技术研发，推动持续的技术创新。我们计划将监控技术的应用扩展至更广泛的智能建筑管理领域，包括但不限于自动化门禁和电梯控制系统。这些增强功能不仅提升了建筑的安全性和管理效率，而且能在紧急情况下迅速采取应急措施，如自动锁定门禁或控制电梯直达安全楼层，从而有效提升建筑的智能防护能力。此外，我们将强化监控系统的隐私保护措施，确保所有监控数据处理严格遵守隐私法规，并通过加密和访问控制增强数据安全，保障用户的隐私权益。



图 6.4: 本作品未来战略规划：发展初创到成熟

## 参考文献

- [1] Alex Biryukov and Dmitry Khovratovich. “Related-key cryptanalysis of the full AES-192 and AES-256”. In: *Advances in Cryptology—ASIACRYPT 2009: 15th International Conference on the Theory and Application of Cryptology and Information Security, Tokyo, Japan, December 6-10, 2009. Proceedings 15*. Springer. 2009, pp. 1–18.
- [2] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. “Yolov4: Optimal speed and accuracy of object detection”. In: *arXiv preprint arXiv:2004.10934* (2020).
- [3] Sebastian Brutzer, Benjamin Höferlin, and Gunther Heidemann. “Evaluation of background subtraction techniques for video surveillance”. In: *CVPR 2011*. IEEE. 2011, pp. 1937–1944.
- [4] Carlos Cuevas, Eva María Yáñez, and Narciso García. “Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA”. In: *Computer Vision and Image Understanding* 152 (2016), pp. 103–117.
- [5] Kaige Cui. “High-altitude parabolic detection system based on machine vision”. In: *2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT)*. IEEE. 2023, pp. 658–662.
- [6] Markus Hegland. “The apriori algorithm—a tutorial”. In: *Mathematics and computation in imaging science and information processing* (2007), pp. 209–262.
- [7] Wenke Huang et al. “Federated graph semantic and structural learning”. In: *Proc. Int. Joint Conf. Artif. Intell.* 2023, pp. 3830–3838.
- [8] Yukun Huang, Yongcai Guo, and Chao Gao. “Efficient parallel inflated 3D convolution architecture for action recognition”. In: *IEEE Access* 8 (2020), pp. 45753–45765.
- [9] Chenglong Li et al. “Weighted low-rank decomposition for robust grayscale-thermal foreground detection”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 27.4 (2016), pp. 725–738.
- [10] Haofeng Li et al. “Motion guided attention for video salient object detection”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 7274–7283.
- [11] Yixuan Li et al. “Multisports: A multi-person video dataset of spatio-temporally localized sports actions”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 13536–13545.
- [12] Jiang Nan et al. “Automatic Detection Method of High Altitude Falling Object Based on Interframe Difference Method”. In: *2022 7th International Conference on Intelligent Informatics and Biomedical Science (ICIIBMS)*. Vol. 7. IEEE. 2022, pp. 22–26.
- [13] Mark Shand and Jean Vuillemin. “Fast implementations of RSA cryptography”. In: *Proceedings of IEEE 11th Symposium on Computer Arithmetic*. IEEE. 1993, pp. 252–259.
- [14] Yun-Tao Shi et al. “High-altitude parabolic detection method based on GMM model and SORT algorithm”. In: *2022 4th International Conference on Intelligent Information Processing (IIP)*. IEEE. 2022, pp. 254–257.
- [15] Robert H Shumway et al. “ARIMA models”. In: *Time series analysis and its applications: with R examples* (2017), pp. 75–163.
- [16] Maryam Sultana et al. “Complete Moving Object Detection in the Context of Robust Subspace Learning”. en-US. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. Oct. 2019. DOI: [10.1109/iccvw.2019.00080](https://doi.org/10.1109/iccvw.2019.00080). URL: <http://dx.doi.org/10.1109/iccvw.2019.00080>.
- [17] Guancheng Wan, Wenke Huang, and Mang Ye. “Federated Graph Learning under Domain Shift with Generalizable Prototypes”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. 14. 2024, pp. 15429–15437.

- 
- [18] Limin Wang et al. “Temporal segment networks for action recognition in videos”. In: *IEEE transactions on pattern analysis and machine intelligence* 41.11 (2018), pp. 2740–2755.
  - [19] Yi Wang et al. “CDnet 2014: An expanded change detection benchmark dataset”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2014, pp. 387–394.
  - [20] Hanbo Wu, Xin Ma, and Yibin Li. “Spatiotemporal multimodal learning with 3D CNNs for video action recognition”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 32.3 (2021), pp. 1250–1261.
  - [21] Jie Ying et al. “GPK-YOLOv5s: Content-Aware Reassembly of Features and Self Attention for High Altitude Parabolic Detection and Tracking”. In: *MATEC Web of Conferences*. Vol. 363. EDP Sciences. 2022, p. 01012.