

基于 CRS-YOLO 算法的智慧楼宇无权限人员入侵检测

宋 超¹ 刘剑磊¹ 王志乾¹
SONG Chao LIU Jianlei WANG Zhiqian

摘 要

采用固定卷积核, 仅能处理特定尺度范围的目标, 降低了入侵检测的精准度。为了解决上述问题, 文章提出了基于 CRS-YOLO 算法的智慧楼宇无权限人员入侵检测。通过科学布局摄像头构建全方位监控网络, 结合人脸图像预处理、检测识别与比对搜索技术, 精准甄别楼宇人员权限并标注无权限人员。针对智慧楼宇场景特点, 改进 YOLO 基础架构, 引入通道-空间混合注意力机制并设计新损失函数, 构建基于 CRS-YOLO 算法的无权限人员入侵检测模型, 使得模型可以更好地捕捉不同尺度目标的关键特征。对标注数据集进行格式转换与超参数配置, 依托 CRS-YOLO 模型进行训练并采用早停策略防止过拟合, 最终部署模型实现无权限人员入侵的实时检测与追踪。实验结果表明, 研究方法能够通过更优的 ROC 曲线与混淆矩阵表现, 显著提升无权限人员入侵检测的精准度。

关键词

CRS-YOLO 算法; 智慧楼宇; 无权限人员; 入侵检测; 损失函数

doi: 10.3969/j.issn.1672-9528.2025.11.006

0 引言

当前社会对安全问题高度重视, 无权限人员入侵检测技术在智慧楼宇、工业场地等关键区域的安全保障中至关重要。然而, 现有入侵检测方法在实际应用中面临诸多挑战。基于红外光栅技术的方法依赖设备部署, 且易受光线变化影响导致误报^[1]; 基于深度学习的方法虽能提取视频特征, 但在处理快速或小目标时, 速度和精度难以兼顾^[2]; 基于 CSA-ResNet 的方法引入通道注意力机制, 但对非结构化入侵行为泛化能力不足^[3]; 改进 YOLOv5 的方法虽优化网络结构, 但在处理小目标和遮挡目标时精度受限^[4]。

针对上述方法的不足, 本文提出基于 CRS-YOLO 算法的智慧楼宇无权限人员入侵检测方法。

1 标注无权限人员

在智慧楼宇无权限人员入侵检测研究中, 视频数据无标签导致模型学习困难。本文以科学布局摄像头为基础, 结合人脸图像处理技术实现人员权限甄别与标注。依据楼宇布局, 在入口、走廊、电梯间、停车场及敏感区域等关键节点安装高清摄像头, 构建全方位监控网络。不同区域摄像头安装位置各有侧重, 以全面覆盖人员行为。由于采集的人脸图像易受干扰, 进行预处理, 包括几何规范化和灰度规范化^[5]。几何规范化调整瞳孔间距使图像符合标准, 灰度规范化采用技术改善质量、统一灰度范围。

预处理后, 利用人脸检测算法定位人脸, 通过特征提取技术建模并提取特征。将人像图片保存到 FastDFS 进行比对搜索^[6], 对比检测到的人脸特征与授权人员特征库, 确定人员权限。最后, 用 Labellmg 工具标记无权限人员位置, 按比例划分数据集, 为模型训练提供数据支撑。此方法能有效应对视频数据无标签问题, 为智慧楼宇无权限人员入侵检测提供可行方案。

2 基于 CRS-YOLO 算法建立入侵检测模型

为了适应智慧楼宇场景人员活动复杂、目标尺度差异大、背景干扰多的情况, 对 YOLO 基础架构进行改进。增加卷积层深度和宽度, 强化其特征提取能力, 并动态调整卷积核尺寸与步长, 让模型能同时捕捉楼宇内人员的全局运动模式和局部细节特征。CRS-YOLO 算法是在经典 YOLO 算法的基础上进行改进而来, 融入了通道注意力机制(channel-wise spatial attention, CSA)。这些技术的引入旨在增强模型对多尺度目标的检测能力, 以及提升模型在复杂环境下的适应性。CRS-YOLO 模型由输入层、多个卷积层、池化层、残差块、上采样层和检测头等部分组成。其中, 检测头负责输出不同尺度的特征图, 每个特征图上包含多个锚框(anchor boxes), 用于预测不同大小和形状的目标, 从而实现对多尺度入侵人员的有效检测。

在特征提取阶段, 生成不同分辨率的智慧楼宇无权限人员入侵特征图集合 $\{F_1, F_2, \dots, F_n\}$, 再用自适应融合策略整合成统一特征 F_f 表示:

1. 甘肃同兴智能科技发展有限公司 甘肃兰州 730050

$$F_f = \sum_{m=1}^M w_m \otimes F_m \otimes \phi_m \quad (1)$$

式中： w_m 表示第 m 个无权限人员入侵特征图的动态融合权重； ϕ_m 表示用于特征图尺度对齐的变换矩阵。

本文在模型中引入了通道-空间混合注意力机制，生成通道权重 s_k ，从而强化人员轮廓、姿态等关键特征通道：

$$s_k = S(W_2 \delta(W_1 z_k)) \quad (2)$$

式中： S 表示Sigmoid激活函数； W_1 、 W_2 分别表示第一、二层全连接网络的权重矩阵； δ 表示ReLU激活函数； z_k 表示智慧楼宇无权限人员入侵特征图上所有位置第 k 通道的特征值的平均。

在智慧楼宇无权限人员入侵检测中，通道注意力机制能强化捕捉无权限人员关键特征，提升检测准确性，无权限人员入侵特征图 F_c 可以通过公式表示为：

$$F_c = s_k \otimes F_f \quad (3)$$

在训练过程中，通过损失函数对模型进行持续优化，引导模型朝着更准确的方向学习。针对传统IoU损失在预测框定位精度上的不足，本文的智慧楼宇无权限人员入侵检测模型损失函数 L_S ：

$$L_S = 1 - \text{IoU} + \lambda_1 L_a + \lambda_2 L_d + \lambda_3 L_s \quad (4)$$

式中：IoU表示交并比； λ_1 、 λ_2 、 λ_3 表示权重系数； L_a 、 L_d 、 L_s 分别表示角度成本、距离成本、形状成本^[7]。

在楼宇监控中，该损失函数能够有效减少对远处小目标的定位偏差，避免将正常工作人员误检为入侵者，提高检测的稳定性和可靠性。

3 实现智慧楼宇无权限人员入侵实时检测

为了实时检测智慧楼宇无权限人员入侵，本文把已完成标注和数据集划分后的训练集等，生成符合YOLO格式的标签文件，每个目标都有类别索引、中心点坐标等信息。

在完成人脸图像采集后，利用LabelImg工具对无权限人员进行位置标记。标注人员需在图像中框选出无权限人员的面部区域，并为其分配相应的类别标签，如“无权限人员”。随后，按照70%训练集、15%验证集、15%测试集的比例，将标注好的数据集进行划分。这种划分方式确保了模型在训练、验证和测试过程中数据的独立性与公正性，有助于全面评估模型的性能。完成标注与划分后，需生成符合YOLO格式的标签文件。YOLO格式的标签文件包含每个目标的类别索引、中心点坐标（相对于图像宽度和高度的归一化值）、宽度和高度（同样归一化）等关键信息。为实现这一转换，编写专门的脚本至关重要。该脚本需读取标注文件，精准提取目标框的坐标信息，并依据图像宽度和高度进行归一化计算，最终按照YOLO格式要求将数据写入标签文件，为后续模型训练提供规范的数据输入。

CRS-YOLO模型的构建与训练是整个检测技术的核心。CRS-YOLO算法是在经典YOLO算法的基础上进行改进而来，可能融入了通道注意力机制（channel-wise spatial attention, CSA）等先进技术。这些技术的引入旨在增强模型对多尺度目标的检测能力，以及提升模型在复杂环境下的适应性。从网络结构来看，CRS-YOLO模型通常由输入层、多个卷积层、池化层、残差块、上采样层和检测头等部分组成。其中，检测头负责输出不同尺度的特征图，每个特征图上包含多个锚框（anchor boxes），用于预测不同大小和形状的目标，从而实现了对多尺度入侵人员的有效检测。在模型配置方面，本文为CRS-YOLO模型设置了初始学习率为0.01，这一参数用于控制模型参数更新的步长，合适的初始学习率有助于模型快速且稳定地收敛。批次大小设置为32，即每次训练时使用32个样本进行参数更新，这种批量处理方式有助于提高训练效率。训练轮次设定为1000轮，以确保模型能够充分学习数据集中的特征，捕捉到各种无权限人员入侵的模式。在模型训练过程中，优化器的选择至关重要。

本文采用Adam优化器，结合动量法和自适应学习率的思想，能够根据训练过程中的梯度信息自动调整学习率，加速模型收敛并提高训练的稳定性。同时，针对检测任务的特点，设计合适的损失函数是优化模型性能的关键。该损失函数通常包含分类损失和定位损失两部分。分类损失用于衡量模型对目标类别的预测准确性，定位损失则用于优化模型对目标位置的预测精度。在训练过程中，通过式（4）设计的损失函数对模型进行持续优化，引导模型朝着更准确的方向学习。为防止模型过拟合，本文采用早停策略。在每个训练周期结束后，计算验证集的损失值，并与历史最优损失值进行对比。如果最近 N 个周期损失值都没有明显下降，则认为模型已经达到收敛状态，提前结束训练，避免模型在训练集上过度拟合而降低在测试集和实际应用中的泛化能力。

模型训练完成后，需将其部署到智慧楼宇监控系统以实现实时检测。在模型部署阶段，首先要进行系统集成。这涉及将训练好的CRS-YOLO模型文件加载到监控系统中，初始化推理引擎，并确保模型能够与监控视频流实现无缝对接。为确保实时检测的性能，可能还需要对模型进行一系列优化操作，如量化、剪枝等。量化操作可以将模型中的浮点数参数转换为定点数，减少模型的大小和计算量；剪枝操作则可以去除模型中不重要的神经元或连接，进一步提高模型的运行效率。进入实时检测阶段，使用训练好的CRS-YOLO模型对实时监控视频帧进行处理。具体而言，将每一帧视频图像输入到模型中，模型经过计算输出每个目标的类别、位置等信息。根据模型输出的结果，系统能够快速判断是否存在无权限人员入侵。一旦检测到入侵情况，系统将立即发出

警报,并通过视频跟踪算法锁定入侵者的位置,实现对入侵人员的实时追踪,为后续的安全处置提供有力支持。

基于 CRS-YOLO 算法的智慧楼宇无权限人员入侵检测流程如图 1 所示。

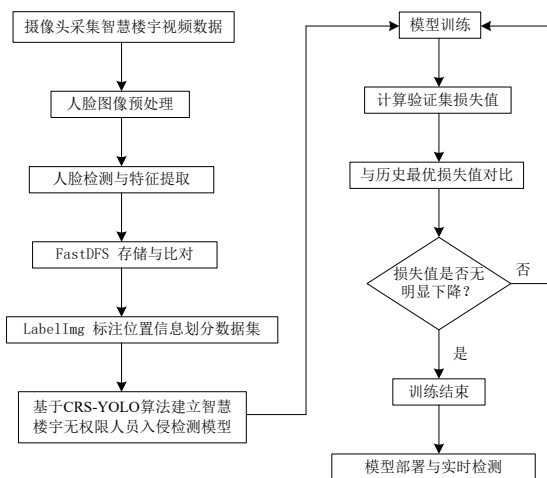


图 1 智慧楼宇无权限人员入侵检测流程图

4 实验

4.1 实验准备

在实验准备阶段,本文配置了实验所需的硬件与软件环境,具体配置情况如表 1 所示。

表 1 实验环境配置表

类别	设备/软件名称	型号/版本	参数/配置
硬件环境	摄像头	海康威视 DS-2CD2T45G0P-I8	400 万像素,2 560×1 440 分辨率,水平 80° 视角,垂直 45° 视角,25 帧/s
	服务器	Dell PowerEdge R740	2×Intel Xeon Gold 6248R 处理器(2.4 GHz, 24 核/48 线程),256 GB DDR4 ECC 内存,2×960 GB SSD 硬盘
	网络设备	Cisco Catalyst 9300 系列	48 个千兆以太网端口,4 个万兆 SFP+ 端口,支持 PoE+ 供电
	存储设备	NetApp AFF A220 存储阵列	2×960 GB SSD 硬盘,2×4 TB HDD 硬盘,支持 RAID 5 配置
软件环境	操作系统	Ubuntu 20.04 LTS	64 位操作系统,支持多线程处理,配备最新的安全补丁
	数据库管理系统	MySQL 8.0	提供数据存储和查询能力

根据表 1 列出的硬件设备与软件系统,对本文的智慧楼宇无权限人员入侵检测实验环境进行了搭建,为本文的实验提供了良好的环境基础。

本文为 CRS-YOLO 模型设置了初始学习率为 0.01,这一参数用于控制模型参数更新的步长,合适的初始学习率有

助于模型快速且稳定地收敛。批次大小设置为 32,即每次训练时使用 32 个样本进行参数更新,这种批量处理方式有助于提高训练效率。训练轮次设定为 1 000 轮,以确保模型能够充分学习数据集中的特征,捕捉到各种无权限人员入侵的模式。在模型训练过程中,优化器的选择至关重要。本文采用 Adam 优化器,它结合了动量法和自适应学习率的思想,能够根据训练过程中的梯度信息自动调整学习率,加速模型收敛并提高训练的稳定性^[8]。

在完成人脸图像采集后,利用 LabelImg 工具对无权限人员进行位置标记。标注人员需在图像中框选出无权限人员的面部区域,并为其分配相应的类别标签,如“无权限人员”。随后,按照 70% 训练集、15% 验证集、15% 测试集的比例,将标注好的数据集进行划分。

4.2 实验方案设计

为了验证研究方法的优越性,本文挑选了另外几种入侵检测方法作为参照,与本文提出的基于 CRS-YOLO 算法的入侵检测方法进行对比,以此来设计本文的对比实验方案。具体方案如下:将基于红外光栅技术的高压场地作业区域人员入侵检测方法作为对比方法 1,基于深度学习的动态危险区域入侵检测方法作为对比方法 2,基于改进 YOLOv5 的高铁周界入侵人员检测方法研究作为对比方法 3。

在实验开展过程中,首先按照表 1 所列配置清单完成实验环境的搭建工作,确保所有设备均处于正常运行状态且网络连接稳定可靠。在实验区域内模拟不同场景,使用摄像头采集视频数据。将上述 4 种入侵检测方法部署在服务器上,并根据实验需求对各方法的相关参数进行配置。将采集到的视频数据依次输入到这 4 种入侵检测方法中,记录各方法的 ROC 曲线与混淆矩阵,评估 4 种方法在楼宇无权限人员入侵检测任务中的性能。

4.3 实验结果及分析

按照上述设计的实验方案,本文进行了实验,并得到了 4 种入侵检测方法的 ROC 曲线结果,具体如图 2 所示。

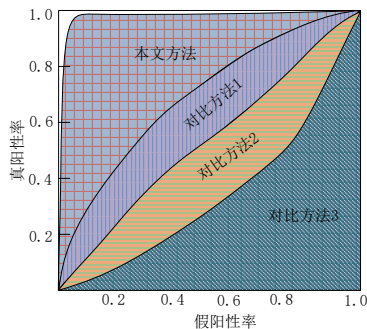


图 2 4 种入侵检测方法 ROC 曲线对比结果图

从图 2 中可以看出,本文基于 CRS-YOLO 算法的入侵

检测方法在 ROC 曲线上展现出了显著的优势。在相同阈值设定下，该方法对应的 ROC 曲线明显更靠近左上角。ROC 曲线越靠近左上角，意味着模型在保证较低假正例率的同时，能够获得较高的真正例率，这表明模型在平衡误报率和召回率方面具有更出色的性能。此外，曲线下面积值是衡量 ROC 曲线性能的关键指标，AUC 值越大，模型的分类性能越好。经过精确计算，本文方法的 AUC 值显著高于其他 3 种方法。这一结果充分说明，在面对智慧楼宇中复杂多变的入侵检测场景时，CRS-YOLO 算法能够更有效地区分无权限入侵人员和正常人员，为楼宇安全提供更可靠的保障。这种优异的性能得益于 CRS-YOLO 算法对传统 YOLO 框架的深度优化。其中，引入的通道注意力机制起到了至关重要的作用。通道注意力机制能够自适应地聚焦于与入侵检测任务高度相关的特征通道，同时过滤掉背景噪声等无关信息。在智慧楼宇的实际场景中，存在着各种复杂的干扰因素，如人员的正常走动、光照变化、背景物体的遮挡等。通道注意力机制可以使模型更加专注于与入侵行为相关的特征，如人员的面部特征、异常的移动轨迹等，从而在复杂环境下实现高精度的入侵检测。

为了进一步分析几种方法在分类性能上的差异，本文还对 4 种入侵检测方法的混淆矩阵进行了对比分析，具体结果如图 3 所示。

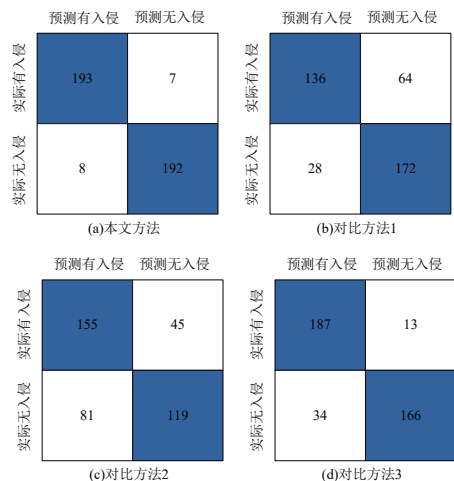


图 3 4 种入侵检测方法混淆矩阵对比结果图

从图 3 中的混淆矩阵可以看出，本文方法在识别真正例和真负例方面表现更为精准，同时假正例和假负例的数量明显更少。真正例表示模型正确识别出的无权限入侵人员数量，真负例表示模型正确判断为正常人员的数量。假正例则是将正常人员误判为无权限入侵人员，这会导致不必要的警报和资源浪费；假负例则是将无权限入侵人员误判为正常人员，这会给楼宇安全带来严重威胁。本文方法的优异表现主要归因于其通过引入通道注意力机制，能够动态调整不同特征通道的权重。在智慧楼宇环境中，不同的区域和场景具有各自

独有的特征，通道注意力机制可以使模型根据不同的场景动态分配注意力，捕捉到智慧楼宇中与入侵检测相关的关键特征，从而实现精准的入侵检测。

5 结语

本文首先对智慧楼宇的摄像头网络进行了布局，结合人脸识别算法，标注了无权限人员数据集。改进 YOLO 算法架构，引入多尺度特征融合与混合注意力机制，设计新损失函数减少定位偏差，提升了模型检测精度。对构建的 CRS-YOLO 模型进行训练，实现了智慧楼宇无权限人员入侵的精准检测。

参考文献：

[1] 肖耀辉,李为明,丁伟锋,等.基于红外光栅技术的高压场地作业区域人员入侵检测方法[J].激光杂志,2024,45(8):87-91.

[2] 朱鹏浩,李军,张世义,等.基于深度学习的动态危险区域入侵检测方法[J].中国安全生产科学技术,2024,20(7):170-178.

[3] 张雷,鲍蓉,朱永红,等.基于CSA-ResNet的人员入侵检测方法[J].计算机工程,2023,49(4):297-302.

[4] 傅荟瑾,史天运,王瑞,等.基于改进YOLOv5的高铁周界入侵人员检测方法研究[J].铁道标准设计,2023,67(9):162-169.

[5] 毛清华,苏毅楠,贺高峰,等.基于改进YOLOv8模型的井下人员入侵带式输送机危险区域智能识别[J].工矿自动化,2025,51(1):11-20.

[6] 吴立舟,李华,李典斌,等.基于计算机视觉的塔吊危险区域入侵预警[J].中国安全科学学报,2024,34(7):139-145.

[7] 张永强,李胜男,张子强,等.R-YOLO轨道人员目标检测模型[J].河北科技大学学报,2023,44(6):580-588.

[8] 傅荟瑾,郭鹏跃,徐成伟,等.基于多传感技术融合的高速铁路周界入侵监测技术方案研究[J].铁道运输与经济,2022,44(9):122-129.

【作者简介】

宋超(1978—)，男，辽宁开原人，本科，助理工程师，研究方向：信息技术。

刘剑磊(1982—)，男，甘肃武威人，本科，助理工程师，研究方向：信息数字化技术。

王志乾(1992—)，男，甘肃兰州人，本科，研究方向：信息数字化技术。

(收稿日期：2025-06-04 修回日期：2025-11-14)