

DOI: 10.19650/j.cnki.cjsi.J2311123

基于改进 ByteTrack 的高铁周界入侵监测方法研究*

傅荟瑾¹, 史天运², 王瑞³, 马祯³, 张万鹏³

(1. 中国铁道科学研究院研究生部 北京 100081; 2. 中国铁道科学研究院集团有限公司科信部 北京 100081;
3. 中国铁道科学研究院集团有限公司电子计算技术研究所 北京 100081)

摘要:本文旨在应对高铁周界环境复杂、小目标多等情况,研究周界入侵行为的识别与跟踪问题,并提出一种改进 ByteTrack 算法。本文融合 YOLOv7-X 与 BYTE 数据关联方法对模型进行改进,并且引入卷积块注意力机制以提升周界复杂环境下前景目标的识别效果,利用空间-深度转化模块优化跨步卷积与池化层,改善小目标识别时下采样导致的细粒度信息丢失情况。制作铁路周界入侵数据集进行实验,实验结果表明,改进后的模型平均精度达到 95.6%,提升了 9.4%,对大中小目标识别的平均精度均有提升,尤其是对小目标识别效果提升显著,提升了 22.2%。结果表明改进 ByteTrack 算法在高铁周界复杂环境下能实现入侵行为的识别与跟踪,为高铁周界防护提供技术支持。

关键词: 高速铁路;周界入侵;ByteTrack;YOLOv7

中图分类号: TH865 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Research on intrusion detection of high speed railway perimeter based on the improved ByteTrack

Fu Huijin¹, Shi Tianyun², Wang Rui³, Ma Zhen³, Zhang Wanpeng³

(1. Postgraduate Department, China Academy of Railway Science, Beijing 100081, China; 2. Science and Information Department, China Academy of Railway Sciences Group Co., Ltd., Beijing 100081, China; 3. Institute of Electronic Computing Technology, China Academy of Railway Sciences Group Co., Ltd., Beijing 100081, China)

Abstract: To address the problems of high-speed railway perimeter intrusion detection such as complex surroundings and a large number of small targets, an improved ByteTrack algorithm is proposed to realize the identification and tracking of perimeter intrusion. The model is improved by integrating YOLOv7-X and the data association method of BYTE. The convolution block attention module is introduced to improve the recognition effect of foreground targets in complex surroundings. The space-to-depth layer and the non-strided convolution layer are used to optimize the step convolution and pooling layers to improve the loss of fine-grained information caused by down-sampling in small target recognition. The railway perimeter intrusion dataset is established for experiments. The experimental results show that the AP of the improved module is 95.6%, an increase of 9.4%, and has improved the AP of target recognition for large, small, and medium-sized targets, especially for small targets, with a significant improvement of 22.2%. The improved ByteTrack algorithm can realize the identification and tracking of intrusion behavior in the complex environment of high-speed railway perimeter, and provide technical support for high-speed railway perimeter protection.

Keywords: high speed railway; perimeter intrusion; ByteTrack; YOLOv7

0 引言

在“十三五”期间,中国已建成世界最发达高速铁路

网,中国高铁以技术先进、安全正点、便捷高效等特点享誉国内外。随着高铁规模不断扩大和时速不断提高,高铁周界运行环境的安全问题已成当务之急。周界入侵可能造成运输秩序严重被扰乱,甚至发生人员伤亡,因此亟

需对高铁周界进行安全防护。相较于一般周界,高铁周界环境复杂,列车运行横跨复杂多样的地形,气候变化不定,为高铁周界防护带来极大挑战,需不断深化高铁周界入侵监测技术研究。

随着计算机视觉和神经网络的快速发展,图像识别技术被认为是高铁周界入侵监测的重要发展方向,近年来众多学者对基于图像识别的高铁周界入侵监测技术开展了研究。前期,高铁周界入侵检测多采用传统机器学习的方法,如 Teng 等^[1]结合支持向量机与超像素轨道目标检测方法,实现铁轨上异物的监测,此类方法依赖于目标的浅层特征,模型泛化能力差。由于深度学习方法的发展,基于深度学习的目标检测成为近年来周界入侵监测的主流方法,主要是对图像中感兴趣对象进行分类与定位。基于深度学习的目标检测依赖于大量的训练数据,大都使用 COCO 等公开数据集,但由于铁路场景环境干扰多、气候变化莫测,利用公共数据集训练得到的模型在铁路场景下泛化能力差,不具有针对性,易出现漏报误报等问题。因此,一些学者针对铁路周界复杂场景对目标检测模型进行了优化, Ye 等^[2]和 Li 等^[3]在单步多框目标检测(single shot multibox detector, SSD)模型基础上修改了卷积类型,用深度可分离卷积替代普通卷积以节省计算成本,实现铁路的目标检测。Guo 等^[4]提出了一种基于改进深度卷积生成对抗网络的铁路入侵目标图像生成方法。针对铁路小目标检测,王瑞等^[5]提出一种改进的基于级联掩膜区域的卷积神经网络模型,在网络中增加了特征金字塔网络模块和空洞金字塔汇聚模块以提升高铁周界入侵中小目标的检测效果。贺加贝^[6]将深层特征图的语义信息融合到浅层,生成新的特征图,在保持较多的细节信息同时增加了语义信息,加强对铁路场景下小目标行人的检测能力。

进一步地,经过前期对高铁周界入侵行为调研分析发现,人员进入周界后大多会离开入侵点在界限内其他位置进行破坏,需要实现对人员的跟踪,因此多目标跟踪(multiple object tracking, MOT)方法在周界场景下更具适应性。多目标跟踪实现跨视频帧关联检测目标,从而获得目标的整个运动轨迹。近年来,学者们对多目标跟踪方法进行了广泛的研究,利用图聚类^[7]、图神经网络^[8]等算法,实现跨帧和对象的信息聚合;利用卡尔曼滤波^[9]、循环神经网络^[10]和长短期记忆网络^[11]等算法,提高与运动线索的关联性;利用基于检测的跟踪范式(tracking-by-detection, TBD)^[12]和联合检测跟踪范式(joint detection and tracking, JDT)^[13]等,提高多帧检测性能。目前,基于 TBD 范式的多目标跟踪方法在各个场景表现出较好的鲁棒性,该方法将多目标跟踪分为两个部分,检测与数据关联,即先利用性能优秀的检测器得到目标的预测框,然后基于重识别、运动预测、相似度计算等

方法,将检测结果匹配至正确的跟踪目标并形成目标的轨迹。目前多目标跟踪在周界入侵领域并未广泛应用,农怀恩^[14]基于多目标跟踪算法深度简单在线实时跟踪(deep simple online realtime tracking, Deep_SORT),实现了周界行人跟踪,利用卡尔曼滤波器实现了轨迹预测与跟踪。石瑞娇^[15]针对铁路视频监控中行人姿态变化导致的帧间漏检问题,将 Deep_SORT 与检测网络结合,对轨迹上的漏检行人进行补充。针对夜晚或光线较差情况下的铁路周界入侵检测需求,王瑞等^[16]采用三帧差法、混合高斯背景建模法叠加 Deep_SORT 降低误报。

在高铁周界场景下,存在环境复杂、小目标识别困难等问题,为高铁周界入侵识别跟踪带来挑战。基于此,本文提出融合 YOLOv7-X 与 BYTE 数据关联方法的改进 ByteTrack 算法,实现高铁周界入侵目标的识别。

1 相关算法

1.1 基于 YOLOv7 的检测算法

YOLOv7^[17]是你只需看一次(you only look once, YOLO)系列 2022 年 7 月推出的算法,在 COCO 数据集上,比以往 YOLO 系列具有更优的检测精度和速度。

YOLOv7 网络结构主要包括输入层、骨干网络层和头部层。输入层对输入的图片进行预处理后传输到骨干网络层;骨干网络主要用于特征提取,输出 3 个尺寸不同的特征信息,由不同的卷积组合模块构成,其主要特点是采用了高效长程注意力网络 ELAN,利用扩张、混洗、聚合特征的方法不断增强网络学习能力;头部层主要用于融合骨干网络层输出的特征,生成边界框并预测目标类别。YOLOv7 网络结构如图 1 所示,网络包括 CBS 模块、池化卷积模块(maxpool and convolutional module, MPConv)、融合跨阶段局部网络的空间金字塔池化模块(spatial pyramid pooling cross stage partial network, SPPCSPC)模块、Rep 模块(re-parameterization, Rep)等。其中 CBS 模块是由卷积层(convolutional layer, Conv)、归一化层(batch normalization, BN)和 SiLU 激活函数组成,实现特征提取;MPConv 包括池化层与卷积层,可以有效缩减特征图的尺寸;SPPCSPC 模块在空间金字塔池化结合小残差结构时,引入了大的残差边,提取出更多有效特征;Rep 模块为重参数化模块,用以提取和平滑特征。

YOLOv7 虽然在公共数据集上表现出色,但将其直接应用于高铁周界场景中,一方面,YOLOv7 下采样时会导致像素点丢失,高铁周界为带状监测场景,存在远距离小目标漏检误检问题。另一方面,在周界复杂环境下存在前景目标易丢失导致的漏检问题,因此需要对 YOLOv7 模型进行进一步改进以克服以上问题。

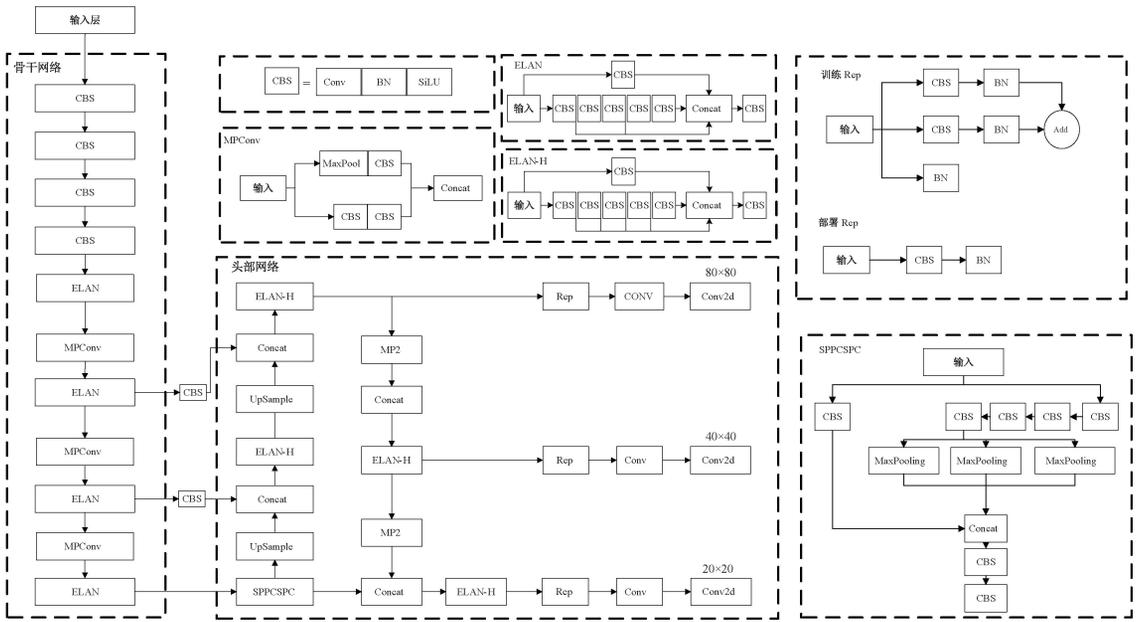


图 1 YOLOv7 结构示意图
Fig. 1 The architecture of YOLOv7

1.2 ByteTrack 多目标跟踪算法

ByteTrack^[18] 是基于 TBD 范式的目标跟踪算法,利用目标检测得到边界框进行轨迹追踪。在检测阶段,如果目标暂时被遮挡会造成检测框得分低,此前大多数 TBD 算法是将低分检测框直接丢弃,但这样可能会造成被遮挡的真实目标丢失跟踪轨迹。ByteTrack 为了改进此类问题,通过数据关联方法 BYTE 对高分检测框建立轨迹的同时,从低分检测框中滤除背景,挖掘真实目标,保持轨迹的连贯性。

BYTE 数据关联方法具体的流程如下:

- 1) 通过目标检测算法得到检测框置信度,设置高、低置信度阈值 0.45 和 0.1,将检测框分为 3 部分,其中置信度小于 0.1 的检测框直接丢弃,大于 0.45 的为高分检测框,小于 0.45 大于 0.1 的为低分检测框;
- 2) 第 1 次匹配时利用高分检测框和之前的跟踪轨迹进行匹配,并利用卡尔曼滤波器得到预测目标轨迹;
- 3) 第 2 次匹配时利用低分检测框和第 1 次没有匹配上高分检测框的跟踪轨迹进行匹配;
- 4) 对于两次都没有匹配上跟踪轨迹的检测框,与目标跟踪池里中间出现的新轨迹进行匹配,即与未激活的轨迹匹配;
- 5) 仍然没有匹配上检测框的跟踪轨迹,保留 30 帧,当其再次出现时进行匹配,如 30 帧仍未再出现则丢弃。

2 基于改进的 ByteTrack 周界入侵目标跟踪算法

2.1 改进的 YOLOv7 模型

为了提升 YOLOv7 模型在周界复杂环境下的识别性能,利用卷积块注意力机制和空间-深度转化模对模型进行改进。

1) 卷积块注意力机制

注意力机制的主要思路与人类观察周围环境的模式类似,通过对全局环境的粗略扫描观察,得到重点注意区域,并忽略抑制其他区域的信息,从而提高对环境感知的处理效率和准确性。目前,在视觉任务中常用的基本注意力机制包括通道注意力机制、空间注意力机制等。

卷积块注意力机制^[19] (convolutional block attention module, CBAM) 结构如图 2 所示,通道注意力机制对输入特征图中更重要的通道增加权重,网络选择重要的通道进行特征响应,使网络更加关注感兴趣区域。空间注意力机制通过压缩通道计算输入特征图的空间注意力权重,使网络更加关注到图片中上下文信息丰富的位置,达到对感兴趣区域信息增强和其他背景区域弱化的效果。CBAM 将通道注意力与空间注意力机制串联,突出前景人员目标的权重,使得网络在学习的过程中更关注于前景人员的特征。将其应用于高铁周界入侵识别中,可以提升算法对目标特征的提取能力,改善由于夜晚、雨雪等复杂环境引起的前景人员特征减弱引起的识别准确性降低的问题。

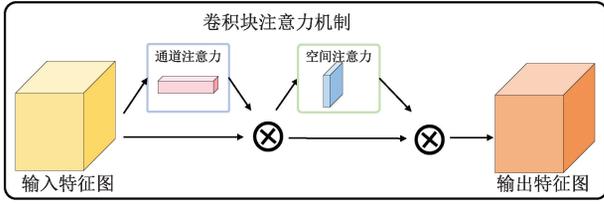


图2 卷积块注意力机制

Fig. 2 Convolutional block attention module

2) 空间-深度转化模块

小目标检测一直都是高铁周界检测中具有挑战性的任务,小目标分辨率较低,与大目标共存时,特征学习过程往往以大目标为主,造成小目标漏检。通过前期实验发现,在图像分辨率高,物体大小适中的场景中存在大量冗余像素信息,YOLOv7的跨步卷积和池化可以方便地跳过这些信息,模型学习特征能力仍然很强,但当图像模糊或者小目标存在时,并不存在大量冗余信息,跨步卷积和池化会造成细粒度信息丢失,使特征学习能力大大降低。

针对此,本文利用空间-深度转化模块^[20](space-to-depth layer and non-strided convolution layer, SPD-Conv)优化跨步卷积和池化层,实现下采样的同时不丢失可学习特征。SPD-Conv模块由一个空间-深度层和一个非跨步卷积层组成,空间-深度层将原始特征图转换为带有特征鉴别信息的中间层特征图,实现下采样。非跨步卷积层步长为1,尽可能地保留所有的特征信息的同时对通道数进行调整,方便与其他层进行特征融合。

在空间-深度层,对任意尺寸为 $S \times S \times C$ 的原特征图 X 以比例因子 $scale$ 进行切分,得到 $scale^2$ 个特征子图 $f_{x,y}$,尺寸均为 $\left(\frac{S}{scale}, \frac{S}{scale}, C\right)$,实现了原特征图 X 的 $scale$ 倍下采样。接着将特征子图沿通道维度连接,得到中间层特征图 $X' \left(\frac{S}{scale}, \frac{S}{scale}, scale^2 C\right)$,保留了通道维度中的所有信息。具体计算过程如式(1)所示。

$$\begin{aligned} f_{0,0} &= X[0:S:scale, 0:S:scale] \\ f_{1,0} &= X[1:S:scale, 0:S:scale], \dots \\ f_{scale-1,0} &= X[scale-1:S:scale, 0:S:scale] \\ f_{0,1} &= X[0:S:scale, 1:S:scale] \\ f_{1,1} &= X[1:S:scale, 1:S:scale], \dots \\ f_{scale-1,1} &= X[scale-1:S:scale, 1:S:scale]; \dots \\ f_{0,scale-1} &= X[0:S:scale, scale-1:S:scale], \dots \\ f_{1,scale-1}, \dots \\ f_{scale-1,scale-1} &= X[scale-1:S:scale, scale-1:S:scale] \end{aligned} \quad (1)$$

图3以 $scale=2$ 为例,原特征图 $X(S, S, C)$ 切分后得到4个特征子图 $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1}$,尺寸均为 $\left(\frac{S}{2}, \frac{S}{2}, C\right)$,实现了对 X 的2倍下采样。然后将特征子图连接,得到中间层特征图 $X' \left(\frac{S}{2}, \frac{S}{2}, 4C\right)$ 。 X' 长度和宽度分别减少为原来的1/2,通道维度变为原来的4倍。

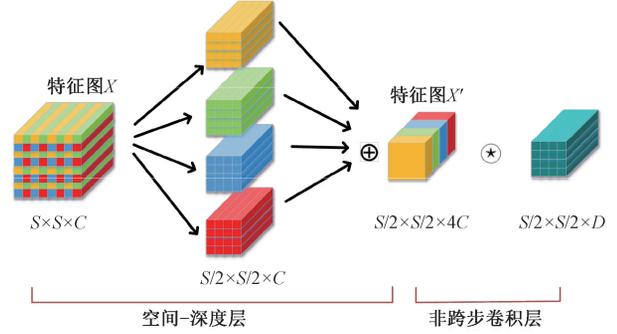


图3 空间-深度转化模块

Fig. 3 SPD-Conv module

在非跨步卷积层,利用D滤波器,进一步对中间层特征图 X' 进行转换,得到特征图 $X'' \left(\frac{S}{scale}, \frac{S}{scale}, D\right)$ 。由于非跨步卷积最大程度的保留了 X' 的特征信息,因此SPD-Conv模块实现了下采样的同时尽可能保留了特征信息。

SPD-Conv模块的输入是一个浮点数组,表示对称正定矩阵,利用输入矩阵进行下采样。而传统卷积层与池化层输入的是多维图像数据,利用权值矩阵和偏置向量进行下采样。

2.2 总体流程

基于TBD设计高铁周界入侵检测跟踪算法,先进行目标检测,再将目标链接到跟踪轨迹上。在这种方法中,跟踪目标的数量和类型完全取决于目标检测的结果,因此目标检测的效果尤为重要,本文重点对目标检测算法进行改进。接着利用优化后的YOLOv7融合BYTE数据关联方法对入侵目标实现识别与跟踪。

检测部分整体和YOLOv7保持一致,仍然是有锚框结构。首先基于训练集中真实标注框,通过k均值聚类算法,获得9个从小到大排列的锚框作为先验。将每个真实标注框与9个锚框匹配,分别计算它们与9个锚框对应的长与宽的比值,取最大的比值,若该比值小于设定的阈值,则该锚框的预测框为正样本。一个真实标注框可能与几个锚框均能匹配,因此一个真实标注框可能在不同的网络层上做预测训练,大大增加了正样本的数量,加速模型收敛。

CBAM注意力机制为即插即用模块,如果将其添加到骨干网络中,一方面改变了已设计好的骨干网络结构,

影响其特征提取能力,另一方面会降低骨干网络的一些原始权重,导致网络预测结果出现错误。本文设计将 CBAM 增加在 YOLOv7 头部层,此时包含更丰富的语义信息,让网络更好收敛并突出重点,对图像中人员信息予以聚焦与关注。随着模型网络层不断加深,卷积与池化操作增加,造成小目标特征的丢失,利用 SPD-Conv 模块进行改善。但 SPD-Conv 会大幅增加训练时间,为了平衡训练时间和检测精度,利用 SPC-Conv 模块优化网络浅层的卷积层与池化层,对小目标特征信息最大程度保留。

跟踪部分采用 BYTE 数据关联方法。选取高、低置信度阈值 x 和 n ,进行两次匹配,如两次匹配均失败,与目标跟踪池中未激活的轨迹进行匹配,检测结果与当前跟踪池没有匹配上的结果时,保存 30 帧。

改进算法的总体流程如图 4 所示。

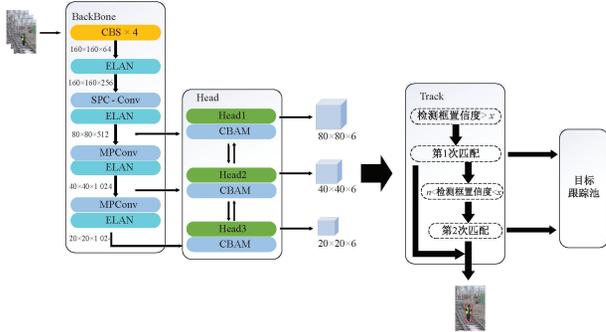


图 4 改进算法流程

Fig. 4 Algorithm improvement flowchart

损失函数部分,常见的目标检测算法的损失函数由坐标损失、目标置信度损失和目标分类损失 3 部分构成,由于高铁周界场景只针对人员一种类别,因此损失函数分为坐标损失、目标置信度损失两部分。其中目标置信度损失采用二分类交叉熵损失(binary crossentropy loss, BCELoss)^[21]的变体 BCEWithLogitsLoss,该损失结合了 BCELoss 和 sigmoid 函数,其比单独使用 BCELoss 和 sigmoid 在数值上更加稳定。坐标损失采用 Clou 损失,考虑了边界框的中心点距离和宽高比,一定程度上可以提高高铁周界入侵识别抗遮挡干扰能力。BCEWithLogitsLoss 计算方法如式(2)所示。

$$l = -w [y \log \sigma(x) + (1 - y) \log(1 - \sigma(x))] \quad (2)$$

其中, $\sigma(x) = \text{sigmoid}(x)$, x 表示预测输出, y 表示置信度标签, w 为一个固定参数,当标签不均衡时设置,正常情况下设为 1。

Clou 损失计算方法如式(3)~(6)所示。

$$L_{Clou} = 1 - IoU + \frac{\rho^2(p, p^{gt})}{c^2} + \alpha v \quad (3)$$

$$IoU = \frac{|B_p \cap B_g|}{|B_p \cup B_g|} \quad (4)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (5)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (6)$$

其中, B_p 为预测框, B_g 为标注框, ρ 为欧氏距离, p 为预测框中心点, p^{gt} 为标注框中心点, c 为框之间最小外接矩形框的对角线距离, w 和 h 为预测框的宽和高, w^{gt} 和 h^{gt} 为标注框的宽和高。

3 实验验证

3.1 数据集

目前多目标跟踪数据集主要是 MOT 数据集,包括 MOT 15、MOT 16、MOT 17 等,主要侧重于密集场景下的行人跟踪。在高铁周界入侵场景中,由于人员入侵行为姿态各异,周围环境复杂多变,使用多目标跟踪数据集进行训练会导致效果不佳。因此,本文制作了铁路周界入侵数据集,缩小了 MOT 数据集的监测范围,同时也可应用于目标检测的训练。

高速铁路横跨复杂多变的地理环境,深入调研多种场景下高铁周界人员入侵现状,分析归纳周界入侵典型场景包括路基地段、隧道口、路堑、桥头等,通过真实入侵与模拟入侵两种方式采集数据。真实入侵数据通过铁路沿线综合视频监控系统采集,模拟入侵数据通过搭建模拟场景进行采集。在国家铁道试验中心搭建周界入侵典型场景,轨旁新立监控杆,在距钢轨面 4 m 处安装海康威视 DS-2CD7047EWD 摄像机。在每个场景中开展模拟试验,结合不同时间与天气条件下的入侵行为与入侵动作,建立数据集。考虑到环境的复杂性与多样性,数据集混杂了不同难度的干扰,包括遮挡、灯光干扰、小目标等。共收集了 2 000 段视频,每条视频约为 4~6 s,对每条视频进行标注,最终形成铁路周界入侵数据集。相较于公共 MOT 数据集,自建数据集为高铁周界入侵识别提供了更加集中和具体的基准。数据集一些重要统计数据如表 1 所示,部分图像示例如图 5 所示,其中将目标按照不同尺寸进行统计,大目标指尺寸大于 96 pixel×96 pixel 的目标,小目标指尺寸小于 32 pixel×32 pixel 的目标,中目标介于两者之间。

3.2 实验环境和参数设置

1) 实验环境

实验采用 Pytorch1.11 深度学习框架,CPU 配置为 Intel® Core TMi9-12900k, GPU 为 GeForce RTX 3090 Ti×2,系统为 ubuntu18.04LTS, Python3.8, CUDA11.2 环境。

表1 铁路周界入侵数据集数据统计

Table 1 Data statistics of railway perimeter intrusion dataset

类别	子类别	数量
场景	路基地段	792
	隧道口	435
	路堑	450
	桥头	323
	白天	1 494
时间	晚上	506
	大目标	555
目标类型	中目标	577
	小目标	306
	无	562
	晴	1 605
天气	雨	290
	雪	65
	雾霾沙尘	40
入侵行为	攀爬	74
	破坏	95
	下挖	22
	横穿线路	994
	沿线路运动	815
入侵动作	站立	541
	行走	976
	奔跑	169
	蹲行	136
	弯腰	93
	爬行	51
干扰	躺卧	34
	灯光	185
	遮挡	159

2) 实验参数设置

训练时使用80%的铁路周界入侵数据集和MOT数据集混合训练,并使用20%的铁路周界入侵数据集作为验证。网络训练时采用基于动量的随机梯度下降算法对模型进行120 epoch的训练,动量因子为0.9,初始学习率为 10^{-4} ,Batch size大小设置为24,输入图像的大小调整为 640×640 ,在80 epoch时学习率下降到 10^{-5} 。采用标准数据增强技术,包括旋转、缩放和颜色抖动等方式,提升模型泛化能力。在两个RTX 3090 Ti GPU上,训练时长大约需要96 h。

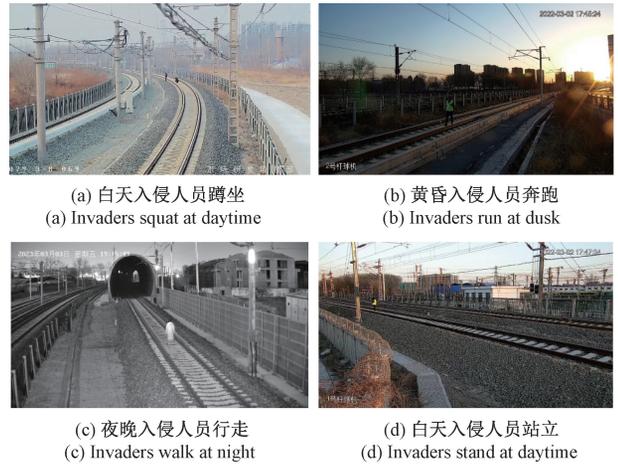


图5 数据集部分图像示例

Fig. 5 Partial image of datasets

3.3 性能评价指标

在周界入侵识别领域,常用误报率(false positive rate, FPR)、漏报率(false negative rate, FNR)、平均精度(average precision, AP)作为性能评价指标,同时在实际应用时,需要考虑检测速度(frames per second, FPS)和模型参数量。将数据划分为真正例(true positive, TP)、假正例(false positive, FP)、真负例(true negative, TN)、假负例(false negative, FN),分别代表识别出人员入侵时实际有人入侵与实际无人入侵,识别为无人入侵时实际无人入侵与实际有人入侵。精确率(precision, P)为真正例占识别为有人入侵实例的比例,召回率(recall, R)为真正例占全部真实值的比例,FPS为每秒传输帧数。FPR、FNR、AP、FPS计算公式如下:

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

$$FNR = \frac{FN}{TP + FN} \quad (8)$$

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$AP = \int_0^1 P(R) dR \quad (11)$$

$$FPS = \frac{1000}{T_{pre-process} + T_{inference} + T_{NMS}} \quad (12)$$

式中: $T_{pre-process}$ 为单张图像预处理时间, $T_{inference}$ 为推理时间, T_{NMS} 为后处理时间。

3.4 实验结果分析

1) YOLOv7与YOLO系列模型对比

为验证YOLOv7在铁路周界入侵数据集上的检测效果,选择YOLOv5系列中AP表现最好的YOLOv5-X, YOLOX^[22]系列中AP表现最好的YOLOX-X, YOLOv7,进

行复合缩放后的 YOLOv7-X,使用 E-ELAN 的 YOLOv7-E6E 进行试验,将上述 6 种网络模型在周界入侵数据集上进行测试,实验结果如表 2 所示。其中 AP_s 、 AP_M 、 AP_L 分别表示小、中、大目标的 AP 值。由表 2 可知, YOLOv5-X 到 YOLOv7-E6E 的 AP 递增, YOLOv7-E6E 虽然 AP 表现最佳,但是推理速度大大下降,模型参数量达到 151.7 MB,考虑到后续需增加新的模块对模型进行改进,模型参数量将进一步增加,推理速度也会受到影响, YOLOv7-E6E 无法满足高铁周界入侵对推理速度的要求。 YOLOv7-X 与 YOLOv5-X 相比,AP 提升了 4.1%,推理速度提高了 44 帧,减少了 7.6% 的参数量。 YOLOv7-X 与 YOLOX-X 相比,AP 提升了 2.4%,推理速度提高 63 帧,同时模型大小减少 26.1 M。 YOLOv7-X 与 YOLOv7 相比,模型进行了复合缩放,将计算块的深度加大 1.5 倍,将过渡块的宽度扩大 1.25 倍,AP 提升了 2.1%,大中小目标样本的 AP 分别提升了 1.6%、2.7%、1.2%,尤其对小目标 AP 的提升较多。虽然 YOLOv7-X 推理速度下降了 44 帧,但 90 帧/s 的推理速度已能满足高铁周界入侵监测需求。

表 2 YOLOv7 与 YOLO 系列模型检测效果对比

Table 2 Comparison of YOLOv7 and YOLO series model detection effects

算法	AP/%	AP _s /%	AP _M /%	AP _L /%	模型大小/MB	FPS/(帧·s ⁻¹)
YOLOv5-X	83.7	74.5	85.6	89.3	61	46
YOLOX-X	85.4	77.9	86.4	92.3	99.1	27
YOLOv7	85.7	78.4	86.6	92.9	85.7	134
YOLOv7-X	87.8	79.6	89.3	94.5	71.3	90
YOLOv7-E6E	90.1	83.4	91.2	95.9	151.7	16

综合考虑推理速度与精度的平衡,文本选用 YOLOv7-X 作为检测器,并在此基础上进行改进以增强其在复杂环境下对周界入侵人员的识别效果。

2) CBAM 对模型影响

为了验证不同注意力机制对模型的影响,分别将压缩-激励模块 (squeeze-and-excitation, SE)^[23]、高效通道注意力 (efficient channel attention, ECA)^[24] 和 CBAM3 种注意力机制加入到 YOLOv7-X 网络的头部层进行实验,实验结果如表 3 所示。由此可见,3 种注意力机制均降低了模型的推理速度,分别下降了 9、5 和 11 帧,CBAM 对速度影响最大。增加 SE 与 ECA 注意力后模型的 AP 反而降低了 0.2% 和 0.6%,SE 对小目标和中目标样本的 AP 降低了 1.7% 和 0.2%,对大目标样本的 AP 提升了 0.2%,ECA 对小目标样本的 AP 降低了 0.9%,对中目标和大目标样本的 AP 提升了 0.4% 和 0.2%。CBAM 对模型的 AP 提升了 2.5%,对小中大目标样本的 AP 提升了 0.4%、3.5%,

4%。说明模型中加入 CBAM 对入侵人员的识别效果有提升,并且优于其他两种注意力机制。虽然 CBAM 的加入降低了模型推理速度,但 79 帧/s 的推理速度已能满足高铁周界入侵监测需求。与 SE 和 ECA 相比,CBAM 模块不仅改进了通道注意力模块,还增加了空间注意力模块,有助于明确小目标位置,从两个维度对特征进行了强化。

表 3 注意力机制对 YOLOv7-X 模型影响

Table 3 The influence of attention mechanism on YOLOv7-X model

注意力机制			AP/%	AP _s /%	AP _M /%	AP _L /%	FPS/(帧·s ⁻¹)
SE	ECA	CBAM					
×	×	×	87.8	79.6	89.3	94.5	90
√	×	×	87.2	77.9	89.1	94.7	81
×	√	×	87.6	78.7	89.7	94.8	85
×	×	√	90.3	80.0	92.8	98.5	79

3) SPD-Conv 对模型影响

为了验证 SPD-Conv 对模型的影响,利用 SPD-Conv 对骨干网络浅层 12~16 层的卷积层与池化层进行了优化,优化前后的实验结果如表 4 所示。可以看出,优化后模型的 AP 提升了 3.5%,小中大目标样本的 AP 分别提升了 9.1%、1.7%、0.1%,对小目标的识别效果提升显著。这是因为在原本的模型中,随着网络层数的增加与感受野的变大,网络提取的高级特征反映了原始图像中更大的像素范围,更多地反映原始图像中目标的宏观轮廓。但是小目标包含的信息少,随着网络加深,卷积与池化的增加会造成细粒度信息的丢失,从而导致识别失败。因此利用 SPD-Conv 在骨干网络的浅层对小目标特征信息最大程度保留。

表 4 SPD-Conv 模块对 YOLOv7-X 影响

Table 4 Influence of SPD-Conv module on YOLOv7-X

SPD-Conv	AP/%	AP _s /%	AP _M /%	AP _L /%	FPS/(帧·s ⁻¹)
×	87.8	79.6	89.3	94.5	90
√	91.3	88.7	91	94.6	65

4) YOLOv7-X+ByteTrack 消融实验

为了验证 YOLOv7-X 融合 BYTE 数据关联方法对模型的影响,本文将 ByteTrack、YOLOv7-X+BYTE、改进后 YOLOv7-X+BYTE 等模型进行对比,实验结果如表 5 所示。由此可见,相比于原始的 ByteTrack 模型,融合了 YOLOv7-X 后 AP 提升了 3%,在 YOLOv7-X 中添加 CBAM 后模型 AP 比 ByteTrack 提升了 5%,利用 SPD-Conv 优化后模型 AP 比 ByteTrack 提升了 6%,本文

模型 AP 比 ByteTrack 提升了 9.4%, 对小中大目标样本的 AP 分别提升了 22.2%, 4.1%, 1.9%, 可见对小目标检测精度显著提升。对于小目标, YOLOv7-X 融合 BYTE 后 AP 提高了 9.3%, 添加 CBAM 后 AP 提升了 9.7%, 利用 SPD-Conv 优化后模型的 AP 提升了 18.9%, 可见 SPD-Conv 对提升小目标检测精度有突出贡献。在推理速度

方面, 相比于原始的 ByteTrack 模型, 融合了 YOLOv7-X 后速度提升了 51 帧, 但由于 CBAM 与 SPD-Conv 对模型的进一步优化, 推理速度降低了 18 帧, 但仍然比原始 ByteTrack 提高了 16 帧, 满足高铁周界入侵监测需求。综上, 改进的 YOLOv7-X 融合 BYTE 数据关联方法, 提高了复杂环境下小目标的识别效果。

表 5 YOLOv7+ByteTrack 消融实验结果
Table 5 YOLOv7+ByteTrack Ablation results

算法	$AP\%$	$AP_S\%$	$AP_M\%$	$AP_L\%$	FPS/(帧·s ⁻¹)
ByteTrack	86.2	70.5	90.8	97.3	30
YOLOv7-X + ByteTrack	89.2	79.8	91.4	96.5	81
YOLOv7-X + CBAM+ByteTrack	91.2	80.2	94.3	99.1	78
YOLOv7-X + SPD-Conv+ByteTrack	92.2	89.4	92.5	94.7	60
YOLOv7-X + CBAM +SPD-Conv+ByteTrack	95.6	92.7	94.9	99.2	46

5) 本模型与其他模型实验结果对比

为了验证本文算法的有效性, 对比本文模型与目前先进的 MOT 模型如 FairMOT 模型^[25]、ByteTrack 模型, 及目标检测算法 YOLOv7-X 模型, 采用一致的实验设置, 在铁路周界入侵数据集上进行验证, 结果如表 6 所示。可以看出, 本文模型 AP 相较于其他模型有较大的提升, 尤其是对小目标样本的识别效果提升显著。

本文模型 AP 比 FairMOT 提升了 13.9%, 对小目标样本提升了 30.3%; 比 YOLOv7-X 提升了 7.8%, 对小目标样本提升了 13.1%。本文模型误报率和漏报率分别为 4.4% 和 8.1%, 较其他方法有所降低, 对误报率改善更为显著。对于推理速度, 本模型的推理速度较于 FairMOT 提升了 27 帧, 但比 ByteTrack 与 YOLOv7-X 速度均有所降低。

表 6 本模型与其他模型实验结果对比

Table 6 Comparison of the experimental results of our model with other models

算法	$AP\%$	$AP_S\%$	$AP_M\%$	$AP_L\%$	FPR/%	FNR/%	FPS/(帧·s ⁻¹)
ByteTrack	86.2	70.5	90.8	97.3	11.2	14.5	30
FairMOT	81.7	62.4	91.2	91.7	15.3	12.9	19
YOLOv7-X	87.8	79.6	89.3	94.5	11.7	13.2	90
本文	95.6	92.7	94.9	99.2	4.4	8.1	46

实验结果可视化对比如图 6 所示, 选取了不同光线、天气等条件下小目标存在的几种场景, 为了更直观的对比不同算法预测框, 每组中第 1 张图展示所有算法预测框, 同时为了更清晰地展示每种算法预测框, 每组中第 2~5 张图分别为 4 种算法识别的结果。图 6(a) 为白天入侵人员在铁轨蹲坐的场景, 只有 ByteTrack 和本文算法识别出目标, 并且本文算法的预测框比 ByteTrack 更为精准。图 6(b) 为夜晚两名入侵人员在铁轨上行走与蹲坐的场景, 只有本文算法识别出两名不同形态的入侵人员。图 6(c) 为雨天两名入侵人员在铁轨上行走的场景,

YOLOv7 识别出一名入侵人员, 本文算法识别出两名入侵人员且预测框更为精准。图 6(d) 为夜晚雨天并且存在灯光干扰的条件下, 多名入侵人员在铁轨上行走的场景, 对于近距离目标 4 种算法均可识别, 但 ByteTrack 和 FairMOT 未识别人员脚部, 画面右端穿浅色衣服受到灯光干扰影响的人员, 只有 ByteTrack 和本算法识别出目标且本文算法预测框更加精准, 但由于光线、天气等干扰存在且遮挡严重, 4 种算法均将并列行走的两人识别为一个目标。总之由多种场景下的对比图可得, 本文算法对复杂环境下小目标的识别效果均有提升。

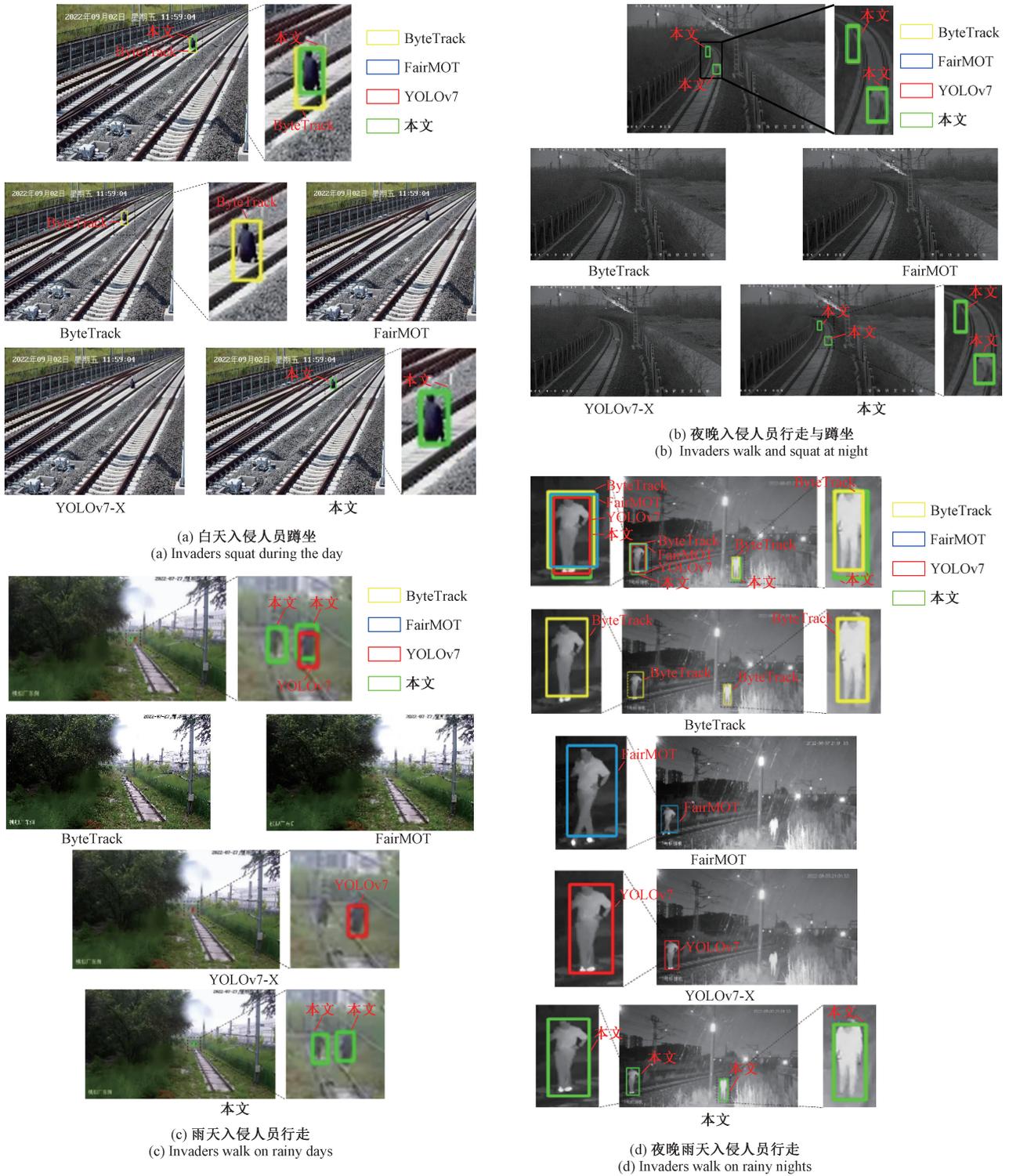


图 6 不同算法在铁路周界入侵数据集中识别结果

Fig. 6 Different algorithms identification results in the railway perimeter intrusion dataset

4 结 论

本文针对高铁周界背景复杂、小目标多造成的入侵

人员识别率低的问题,建立铁路周界入侵数据集,融合 YOLOv7-X 与 BYTE 数据关联方法,增加 CBAM 提高模型对复杂背景下入侵人员检测精准率,利用 SPD-Conv 模块优化 YOLOv7 骨干网络提升小目标识别效果,并与其

他典型模型进行对比,得出结论如下:

1)在本文铁路周界入侵数据集上,改进后的模型与原模型相比, AP 达到了95.6%,提升了9.4%,小目标样本的 AP 提升了22.2%;与YOLOv7-X相比, AP 提升了7.8%,小目标样本的 AP 提升了13.1%;与FairMOT相比,提升了13.9%,小目标样本的 AP 提升了30.3%。无论是综合识别效果还是对小目标样本的 AP ,改进后模型优势都较为明显。

2)在YOLOv7-X头部层增加CBAM可以提高复杂背景下入侵人员的识别效果,利用SPD-Conv改进YOLOv7-X浅层骨干网络的卷积与池化层,可以改善下采样时小目标特征丢失的问题,融合YOLOv7-X与BYTE数据关联方法可以进一步提升模型识别效果并实现目标跟踪。但以上改进均会降低模型的推理速度,降低后速度仍可以满足周界入侵监测的需求,在实际应用时,在保证 AP 的前提下,可以接受牺牲较小的速度指标。

本文所提方法可以有效改善网络模型在高铁周界入侵复杂环境下对小目标的识别与跟踪效果。但如在夜间且存在大量灯光干扰的情况下,视频质量严重降低,今后的研究中,将从低照度图像增强等角度进一步提升模型在夜晚的 AP 。

参考文献

- [1] TENG Z, LIU F, ZHANG B P. Visual railway detection by superpixel based intracellular decisions [J]. *Multimedia Tools and Applications*, 2016, 75 (5): 2473-2486.
- [2] YE T, WANG B, SONG P, et al. Automatic railway traffic object detection system using feature fusion refine neural network under shunting mode [J]. *Sensors*, 2018, 18 (6): 1916.
- [3] LI J, ZHOU F, YE T. Real-world railway traffic detection based on faster better network [J]. *IEEE Access*, 2018, 6: 68730-68739.
- [4] GUO B, GENG G, ZHU L, et al. High-speed railway intruding object image generating with generative adversarial networks [J]. *Sensors*, 2019, 19 (14): 3075.
- [5] 王瑞,李霄峰,史天运,等.基于视频深度学习的铁路周界入侵检测算法研究 [J]. *交通运输系统工程与信息*, 2019, 20 (2): 61-68.
WANG R, LI X F, SHI T Y, et al. Railway perimeter intrusion detection algorithms based on video deep learning [J]. *Journal of Transportation Systems*

Engineering and Information Technology, 2020, 20 (2): 61-68.

- [6] 贺加贝. 铁路场景下的行人检测算法研究 [D]. 成都:西南交通大学, 2021.
HE J B. Research on pedestrian detection algorithm for railway scene [D]. Chengdu: Southwest Jiaotong University, 2021.
- [7] TANG Z, WANG G, XUAO H, et al. Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018: 108-115.
- [8] WANG Y, KITANI K, WENG X. Joint object detection and multi-object tracking with graph neural networks [C]. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021: 13708-13715.
- [9] KIM D Y, JEON M. Data fusion of radar and image measurements for multi-object tracking via Kalman filtering [J]. *Information Sciences*, 2014, 278: 641-652.
- [10] MILAN A, REZATOFIGHI S H, DICK A, et al. Online multi-target tracking using recurrent neural networks [C]. *National Conference on Artificial Intelligence*, San Francisco: AAAI Press, 2017: 4225-4232.
- [11] LU Y, LU C, TANG C K. Online video object detection using association LSTM [C]. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2344-2352.
- [12] ZHOU Q, ZHONG B, ZHANG Y, et al. Deep alignment network based multi-person tracking with occlusion and motion reasoning [J]. *IEEE Transactions on Multimedia*, 2018, 21(5): 1183-1194.
- [13] ZHOU X, KOLTUN V, KRÁHENBÜHL P. Tracking objects as points [C]. *European Conference on Computer Vision*. Springer, Cham, 2020: 474-490.
- [14] 农怀恩. 基于计算机视觉的周界预警系统设计与实现 [D]. 武汉:华中科技大学, 2021.
NONG H EN. Design and implementation of perimeter warning system based on computer vision [D]. Wuhan: Huazhong University of Science and Technology, 2021.
- [15] 石瑞姣. 铁路视频监控中小尺度行人检测算法研究 [D]. 北京:北京交通大学, 2021.
SHI R J. Study on the small-scale pedestrian detection for video surveillance of railway [D]. Beijing: Beijing

Jiaotong University, 2021.

- [16] 王瑞, 史天运, 包云. 一种基于视频的铁路周界入侵检测智能综合识别技术研究[J]. 仪器仪表学报, 2020, 41(9): 188-195.
WANG R, SHI T Y, BAO Y. Research on an intelligent comprehensive recognition technology in railway perimeter intrusion detection based on videos [J]. Chinese Journal of Scientific Instrument, 2020, 41(9): 188-195.
- [17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [J]. Arxiv Preprint, 2022, Arxiv: 2207.02696.
- [18] ZHANG Y, SUN P, JIANG Y, et al. ByteTrack: Multi-object tracking by associating every detection box [C]. European Conference on Computer Vision. Springer, Cham, 2022: 1-21.
- [19] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [20] SUNKARA R, LUO T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects [J]. Arxiv Preprint, 2022, ArXiv: 2208.03641.
- [21] 戴琳琳, 阎志远, 景辉. Insightface 结合 Faiss 的高并发人脸识别技术研究[J]. 铁路计算机应用, 2020, 29(10): 16-20.
DAI L L, YAN ZH Y, JING H. Highly concurrent face recognition technology with Insightface and Faiss [J]. Railway Computer Application, 2020, 29(10): 16-20.
- [22] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [J]. Arxiv Preprint, 2021, ArXiv: 2107.08430.
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

- [24] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [25] ZHANG Y, WANG C, WANG X, et al. FairMOT: On the fairness of detection and re-identification in multiple object tracking [J]. International Journal of Computer Vision, 2021, 129: 3069-3087.

作者简介



E-mail: Fuhuijin@yeah.net

傅荟瑾, 2016 年于北京理工大学获得学士学位, 2019 年于北京理工大学获得硕士学位, 现为中国铁道科学研究院博士研究生。主要研究方向为铁路周界入侵检测、交通信息工程、铁路安全保障等。

Fu Huijin received her B. Sc. degree and M. Sc. degree both from Beijing Institute of Technology in 2016 and 2019. She is currently a Ph. D. candidate at China Academy of Railway Sciences Group Co., Ltd. Her main research interests include railway perimeter intrusion detection, traffic information engineering and railway safety guarantee.



E-mail: tyshi@rails.cn

史天运 (通信作者), 1990 年于北京理工大学获得学士学位, 1995 年于北京理工大学获得硕士学位, 1998 年于北京理工大学获得博士学位, 现为中国铁道科学研究院集团有限公司科信部部长、研究员, 主要研究方向为交通信息工程、铁路防灾、智能铁路等。

Shi Tianyun (Corresponding author) received his B. Sc. degree, M. Sc. degree and Ph. D. degree all from Beijing Institute of Technology in 1990, 1995 in 1998. He is currently the director and a research fellow in Science and Information Department, China Academy of Railway Sciences Group Co., Ltd. His main research interest includes traffic information engineering, railway disaster prevention, intelligent railway, etc.