

◆特邀栏目◆

基于改进 YOLOV7 与 StrongSORT 算法的列车司机手比行为检测*

宋吉超, 黄伟**, 陈振棠, 周成才

(柳州铁道职业技术学院动力技术学院, 广西柳州 545000)

摘要:列车司机驾驶行为的规范性直接影响到列车行车安全与状态,但当前对列车司机手比行为的检测仍存在不足。为了对列车司机手比行为进行有效检测,本文利用动车组模拟驾驶系统所采集的司机乘务作业影像,结合采用融合注意力机制的 You Only Look Once Version 7 (YOLOV7)神经网络模型与 Strong Simple Online and Realtime Tracking (StrongSORT)算法,对动车组司机驾驶过程中的5种手比行为进行检测。实验结果表明:本文算法可以有效提升对列车司机乘务作业时不同类型手比行为的检测效果,其中检测精确率平均提升1.2%,检测召回率平均提升1.9%。本文提出的算法将有助于改进铁路院校、机务部门对列车司机日常训练、工作考评的效果,提高列车行车过程中的安全性。

关键词:列车司机;乘务作业;行为检测;YOLOV7;StrongSORT;注意力机制

中图分类号:O436, TP391.41 文献标识码:A 文章编号:1002-7378(2023)04-0471-08

DOI:10.13657/j.cnki.gxkxyxb.20231226.014

当下人工智能(Artificial Intelligence, AI)技术的应用场景愈发广泛^[1],轨道交通车辆的运营与维护领域也日渐成为了AI技术落地的热门领域^[2]。列车司机驾驶行为的规范性直接影响到列车的行车安全,对于列车司机驾驶行为的监测是AI技术落地轨道交通领域的重要切入点。自2012年卷积神经网络(Convolutional Neural Network, CNN)模型 Alexnet^[3]被提出以来,CNN在图像识别方面取得了重大突破。CNN同样也推动了列车司机行为监控领域的

进步与发展,李伟^[4]通过CNN对列车司机面部特征进行提取,随后将特征送入长短时记忆(Long Short Term Memory, LSTM)网络完成时序建模,增强了疲劳检测算法的适用性与有效性。姚巍巍等^[5]采用人体关键点检测算法 Pose proposal network,实现了对使用电话、站立、端坐、瞌睡、离岗等列车司机行为的识别,并将网络模型通过 TensorRT 加速优化后,在 Jetson TX2 中完成了实际部署。所达^[6]通过区域三维卷积神经网络(Region Convolutional 3D Net-

收稿日期:2022-09-04

修回日期:2023-02-09

*柳州铁道职业技术学院2022年度校级立项项目“基于人工智能技术的列车司机行为监测系统”(2022-KJB14)资助。

【第一作者简介】

宋吉超(1994-),男,硕士,实验师,主要从事机械设计与智能控制研究。

【**通信作者简介】

黄伟(1986-),男,博士,副教授,主要从事人工智能、云计算与大数据等研究,E-mail:huangwei2913@gmail.com。

【引用本文】

宋吉超,黄伟,陈振棠,等.基于改进YOLOV7与StrongSORT算法的列车司机手比行为检测[J].广西科学院学报,2023,39(4):471-478.

SONG J C, HUANG W, CHEN Z T, et al. Gesture Behavior Detection of Train Drivers Based on Improved YOLOV7 and StrongSORT [J]. Journal of Guangxi Academy of Sciences, 2023, 39(4): 471-478.

work, R-C3D) 实现了对城轨列车司机确认手势的快速准确识别与起始时间的精确预测。何理^[7]分别采用 Deep LSTM、baseline Temporal Convolutional Network (TCN)、Spatial Temporal Graph Convolutional Networks (ST-GCN) 模型实现了列车司机手势动作的识别, 并通过对比分析验证了图卷积神经网络的有效性。

手比行为是列车司机驾驶过程中用以确认驾驶室仪表信息、确认道路信号信息等必须进行的微小手部动作, 动作执行的到位率不仅影响着列车行驶过程中的乘务安全, 而且也是评价列车司机驾驶合理性的重要标准^[8]。现阶段业界主要通过人工抽检车载监控视频来分析司机手比行为执行的到位率, 但是人工分析存在效果低下、质量不高以及容易出现遗漏等问题。因此, 使用计算机视觉技术对列车司机手比行为进行有效识别能够帮助铁路机务部门进行日常的监控与管理^[9]。

现阶段对于列车司机行为识别的研究多集中于对疲劳驾驶、玩手机等不规范行为的检测, 而对列车司机手比行为的分析, 特别是对动车组司机乘务作业过程中不同类型的手比行为分析, 国内外的相关研究基本为空白。手比行为技术动作起源于日本, 在日本被称为“指差确认”^[10], 目前主要是日本与中国的铁路系统使用该技术动作, 欧美铁路系统较少应用手比行为技术。中国铁路系统中不同路局乘务作业标准对该技术动作的要求也有所区别。

为实现对列车司机乘务作业时手比行为的有效检测, 本文依据某铁路局机务系统列车司机乘务作业时手比行为的标准化作业要求, 分析不同类型手比行为的技术特征。由于手比行为是小目标, 因此本文采用目标检测算法 You Only Look Once Version 7 (YOLOV7) 进行检测^[11], 并通过融合注意力机制的方法对算法进行改进, 以增强算法的检测效果^[12]。针对司机乘务过程中手比行为不断变化的特点, 采用 Strong Simple Online and Realtime Tracking (StrongSORT) 算法进行跟踪^[13]。

1 准备工作

“手比眼看”是列车司机乘务作业时的规范准则。其中, “手比”指列车司机乘务作业时所进行的一系列手势动作。本文以某铁路局机务系统《CR400 型动车组司机一次乘务作业指导书》为依据, 将列车司机乘务作业时的手比类型概括如图 1 所示。

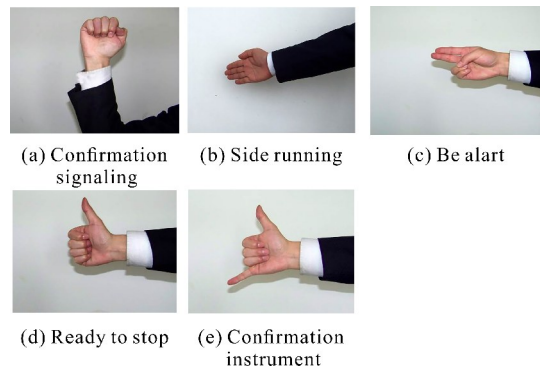


图 1 手比类型

Fig. 1 Gesture type

检测司机手比行为主要存在 2 个难点: (1) 列车司机乘务作业过程中, 不同手势的变化主要为手指状态的改变, 在监控画面中所占范围极小; (2) 列车司机乘务作业过程中, 列车运行指示信号的变化以及司机端坐、站立等不同状态都会造成手比行为在监控画面中位置的不同。图 2 为实时监控画面中列车司机做出的“准备停车”以及“确认仪表”的手比行为, 可以看出手比行为在监控画面中是小目标, 而且手势位置会随着列车的运行、信号的不同而发生变化。因此检测算法必须对小目标有着良好的跟踪检测能力。

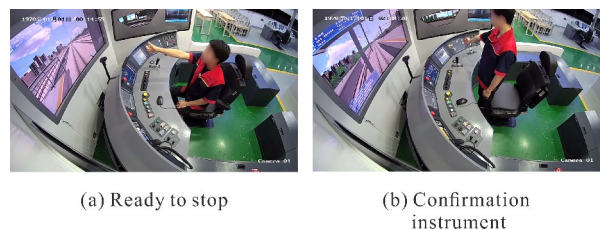


图 2 监控画面

Fig. 2 Monitoring screen

本文的列车司机监控影像数据来自于畅想高科 CR400BF 动车组驾驶仿真系统所采集的实训操作影像。CR400BF 型列车运行时, 对列车司机手比行为的监控主要来自于侧方摄像头(图 2), 监控画面覆盖整个司机室, 且画面清晰。司机室为独立、封闭的安全环境, 不存在其他恶劣外界条件的干扰。

2 检测与跟踪算法

2.1 YOLOV7 模型

YOLO 是 Redmon 等^[14]于 2016 年提出的首个单阶段目标检测算法。该算法只需提取一次特征即可实现目标检测, 目前已成为目标检测领域最具代表性的模型。2022 年 7 月, Alexey 团队发布了最新开源的 YOLOV7^[15]。在相同体量下官方版的 YOLOV7 比 YOLOV5 精度更高, 速度(单位为 fps)

快120%,比YOLOX快180%,比Dual-Swin-T快1200%,比ConvNext快550%,比Swin-L快500%。

YOLOV7整体网络结构主要由以下3部分组成:Input层、Backbone层、Head层^[16]。YOLOV7网络模型框架如图3所示。

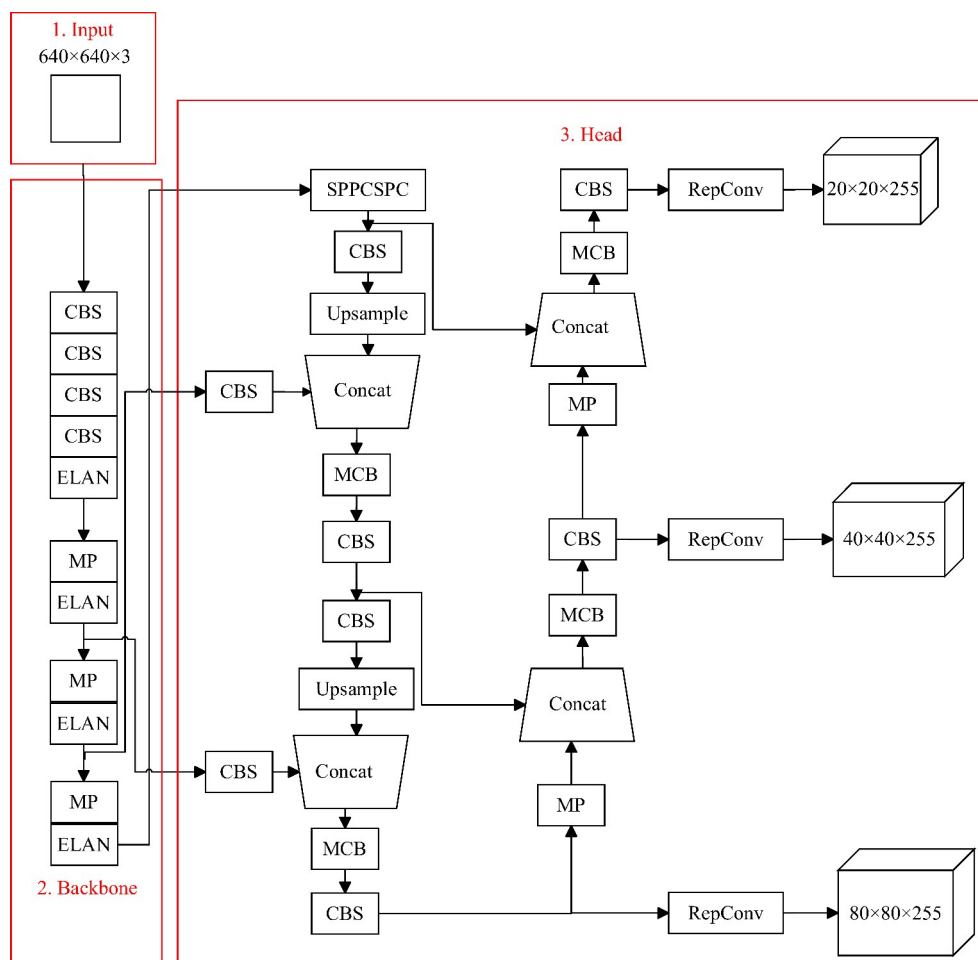


图3 YOLOV7模型框架

Fig. 3 YOLOV7 model framework

Input层采用一系列的方法对输入图片进行处理,从而达到数据增强的目的。Backbone层中CBS模块是卷积运算 Convolution、批量归一化 Batch normalization与激活模块 Silu 的组合;ELAN 模块是由多个卷积拼接而成;MP 模块由最大池化 Maxpooling 和常规卷积组合而成。Head层中的SPPCSPC模块通过CBS与Maxpooling使算法的计算量降低、计算速度加快;MCB模块由多个CBS拼接而成;最后,Head层通过输出3组包含了置信度和像素坐标等信息的特征图,再利用非极大抑制消除多余框以寻找最佳的物体检测位置。本文算法使用 Generalized Intersection over Union (GIoU)来计算定位损失。

2.2 注意力机制的融合

注意力机制在图像识别中的有效性已经得到证

明^[17]。图4为卷积注意力模块(Convolutional Block Attention Module, CBAM),注意力机制被用于获取特征图中可用的注意力信息,即经由通道注意力模块(Channel Attention Module, CAM)和空间注意力模块(Spatial Attention Module, SAM)两个子模块输出新的特征图。该模块通过强调重要特征并抑制一般特征,以提升目标检测效果。

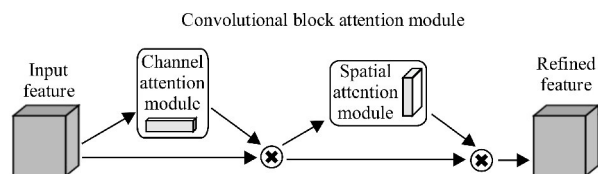


图4 CBAM

Fig. 4 CBAM

为使注意力机制尽可能突出特征图中的关键信息,根据YOLOV7的特性,本文将CBAM位置置于

Backbone 层与 Head 层之间, 添加 CBAM 后的模型 网络结构如图 5 所示。

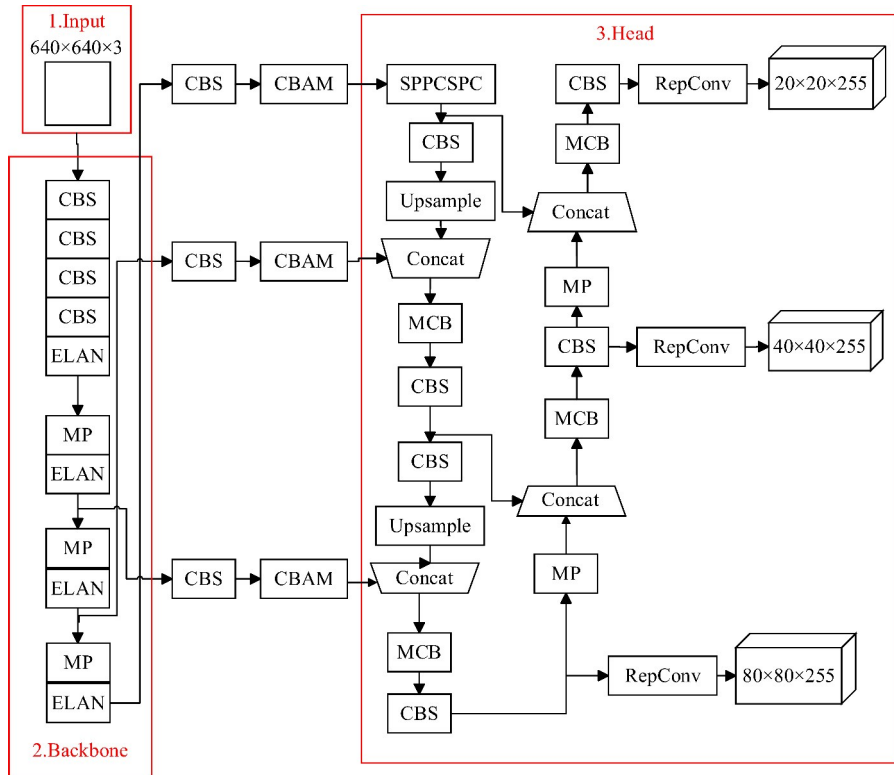


图 5 融合 CABM 后的改进 YOLOV7 网络

Fig. 5 Improved YOLOV7 network integrated with CABM

2.3 StrongSORT 算法

考虑到列车司机进行手比行为时, 手势位置不断发生变化, 为提高检测的效果, 本文采用了 Du 等^[18]提出的 StrongSORT 算法(图 6), StrongSORT 算法是经典的 DeepSORT 算法^[19]的升级。StrongSORT 算法具备通用性, 可以应用在任何可用的视频目标检测模型上。

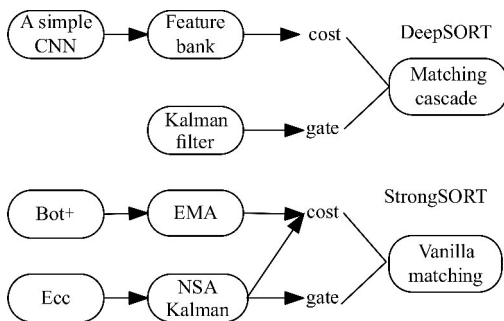


图 6 DeepSORT 和 StrongSORT 的应用原理

Fig. 6 Application principle of DeepSORT and StrongSORT

StrongSORT 算法与 DeepSORT 相比, 在外观分支采用了更强的外观特征提取器 Bot+ 取代原来简单的卷积神经网络, 以 ResNeSt50 为骨干, 可以采集更多的判别特征。此外用特征更新策略替换特征

库, 并以指数移动平均 (Exponential Moving Average, EMA) 方式更新帧 t 处的第 i 个轨迹的外观状态 e_i^t , 计算公式如下

$$e_i^t = \alpha e_i^{t-1} + (1 - \alpha) f_i^t, \quad (1)$$

其中, f_i^t 是当前匹配检测的外观嵌入, $\alpha = 0.9$ 是动量项。对于运动分支, 采用 Enhanced Correlation Coefficient Maximization (ECC) 进行相机运动补偿。

在卡尔曼滤波算法中, 噪声尺度是一个常数矩阵。由于卡尔曼滤波器会受低质量检测的影响, 从而忽略检测噪声尺度的信息, 因此本文采用 Noise Scale Adaptively (NSA) 卡尔曼算法, 通过公式(2)来自适应地计算噪声协方差^[20]:

$$\tilde{R}_k = (1 - c_k) R_k, \quad (2)$$

其中, R_k 是预设的恒定测量噪声协方差, c_k 是状态 k 时刻下的检测置信度得分。即通过预先给定 R_k , 再根据 k 时刻检测到的置信度得分, 以达到能在不同时刻自适应调整噪声斜方差, 从而使估值更为准确。

本文将 CABM 融合进 YOLOV7, 并将融合后的算法(Our algorithm)作为目标检测算法检测不同类型手比行为, 接着将检测结果作为 StrongSORT 算法的跟踪目标, 以实现对列车司机乘务作业过程中手

势的实时跟踪。

3 仿真实验与分析

3.1 实验数据

本文的数据集采集自柳州铁道职业技术学院 10 个班级 400 名不同学生在畅想高科 CR400BF 动车组模拟驾驶仿真实训设备中的某次乘务作业时的手比行为监控视频画面。每位学生在站立与端坐两种姿态下分别进行图 1 中的手比行为。将监控视频中每位同学在每种姿态下的 5 种手比行为各截取 2 张作为原始数据,即 5 种类型的手比行为各 1 600 张,共计 8 000 张图片,并按照 7 : 1 的比例划分数据,即训练集 7 份,验证集 1 份。

采用 Labelimg 将上述数据标注为 YOLO 格式。5 种手比行为的标准代号分别为确认信号 XH、侧线运行 YX、注意警惕 JT、准备停车 TC 和确认仪表 YB。标注数据时应当注意,由于列车司机仅用单手进行手比行为,此时若仅对手掌进行标注则会误检左右手掌,因此为避免左右手掌的干扰,标注时应对整个进行手比行为的手臂进行框选(图 7)。



图 7 行为标注

Fig. 7 Behavior labeling

3.2 实验环境搭建与参数配置

在进行模型训练前,需对计算机及 YOLOV7 模型的一些参数进行初始化设置。本文所进行的实验环境搭建与参数配置如表 1 所示。

3.3 评价指标

对于列车司机检测的评价既要全面又要准确,通常精确率(Precision)和召回率(Recall)两项指标可以评价这两点^[21],两者构成的混淆矩阵如表 2 所示。其中,FP 表示将负类别预测为正类别的个数,TP 表示将正类别预测为正类别的个数,FN 表示将正类别预测为负类别的个数,TN 表示将负类别预测为负类

别的个数。

表 1 实验环境与参数配置

Table 1 Experimental environment and parameter configuration

环境/参数类型 Type of environment/parameter	配置 Configuration
Operating system	Windows 10
GPU	NVIDIA GeForce GTX 1650 SUPER
Programming language	Python (Version 3.8.8)
Deep learning architecture	Pytorch (Version 1.8.2)
Accelerating environment	CUDA (Version 10.2) cuDNN (Version 10.2)
Number of training iterations	300
Initial learning rate	0.001
Batch size	16

表 2 混淆矩阵

Table 2 Confusion matrix

真实标签 True label	预测结果 Forecast result	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

精确率为预测正确的正样本个数除以真实的正样本个数[公式(3)],其可以评价检测的准确性。召回率为正确类别被模型预测正确的概率[公式(4)],其可以评价检测的全面性。

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}), \quad (3)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}). \quad (4)$$

综上,本文采用精确率与召回率作为评价指标。

3.4 检测结果

图 8 为在训练 300 个迭代过程中,本文算法的损失函数曲线。可以看出,在前 50 个训练周期损失函数快速下降,在 50 - 200 个周期中损失函数的下降速度越来越缓慢,在 200 - 300 个训练周期中变化平稳,说明训练 300 个周期内损失函数逐步达到收敛状态。

在训练 300 个迭代周期后,本文算法与 YOLOV7 算法对各手比行为的检测结果见表 3。由表 3 可知,YOLOV7 的 2 项指标的平均值分别达到了 95.3%、93.3%,对 5 种手比行为的精确率均达到 92% 以上,其中,JT 的识别精度最高,可能是因为注意警惕行为除手部行为外,还具备手臂成 90° 的明显特征。YX 与 TC 的识别精确度较低,分别有 7.6% 和 7.3% 的错误率,这主要是因为这两个动作极为接

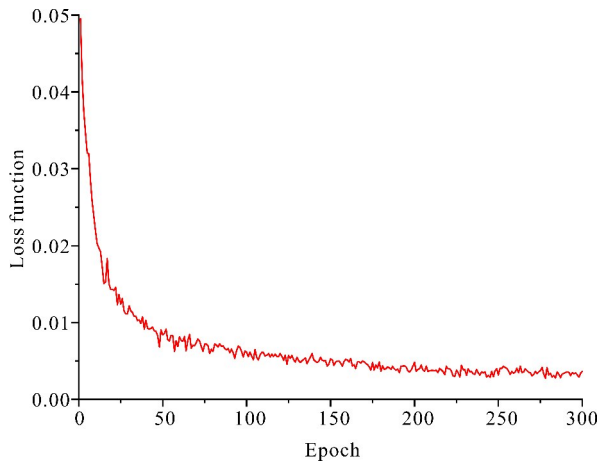


图8 损失函数

Fig. 8 Loss function

近。侧线运行为握拳后比小拇指与大拇指,准备停车为握拳后比大拇指,两者的区别仅表现在小拇指上,两者差别过小。除 YB 外的 4 种手比行为的召回率均达到 93.0% 以上,而 YB 的召回率仅为 89.2%,说明对注意仪表行为的检测可能存在漏检。这可能是由于列车司机需要根据不同的仪表信息从左至右指认不同仪表。如开左门时,需要手指最左侧,而开右侧车门时需要手指最右侧。而监控摄像头仅在左侧,则在右侧指认时,距离及身体会遮挡摄像头部分视线,从而造成漏检。

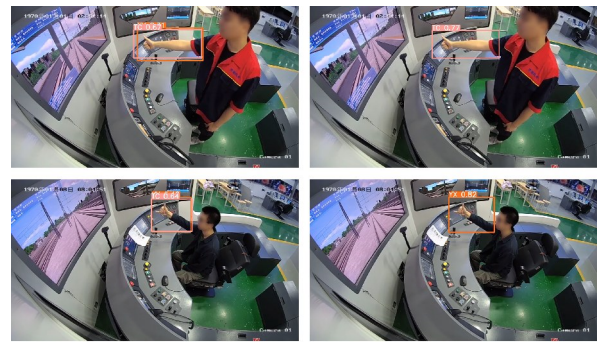
表 3 YOLOV7 和本文算法的精确率、召回率

手比类型 Gesture type	精确率 Precision				召回率 Recall			
	YOLOV7		Our algorithm		YOLOV7		Our algorithm	
	YOLOV7	Our algorithm	YOLOV7	Our algorithm	YOLOV7	Our algorithm	Our algorithm	
XH	96.2	96.9	94.2	95.9				
YX	92.4	94.6	94.4	95.5				
JT	98.9	99.5	95.8	96.9				
TC	92.7	94.8	93.0	94.1				
YB	96.3	96.9	89.2	93.8				
Average value	95.3	96.5	93.3	95.2				

相比于 YOLOV7 算法,本文算法的效果有所提升。YX 精确率从 92.4% 提高到了 94.6%,YC 精确率从 92.7% 提高到了 94.8%。YOLOV7 存在确认仪表漏检的情况,而采用 StrongSORT 算法后可以有效检出确认仪表行为,因此召回率从 89.2% 提高到了 93.8%。在 XH、JT 和 TC 的检测结果中,本文算法的精确率比 YOLOV7 分别提升了 0.7%,0.6%

和 2.1%,召回率分别提高了 1.9%、1.1% 和 1.1%。

图 9、图 10 为两个算法的检测效果。图 9 中,第 1、2 行分别为注意停车和侧线运行的检测效果,这两种手势的区别仅表现在小拇指是否伸出。YOLOV7 算法对注意停车与侧线运行存在混检,不能准确、有效地区分手指间的细小差别。经过改进后的本文算法可以准确检出被混检的图片类型。图 10 为不同算法对确认仪表的检测效果,因为图中手势确认的是右侧仪表,所以监控中手势画面大小相较于中间或左侧的更小,画质更低,这可能导致 YOLOV7 算法未能有效检测出该画面,而本文算法可以将该情况下的手比准确检测为确认仪表。

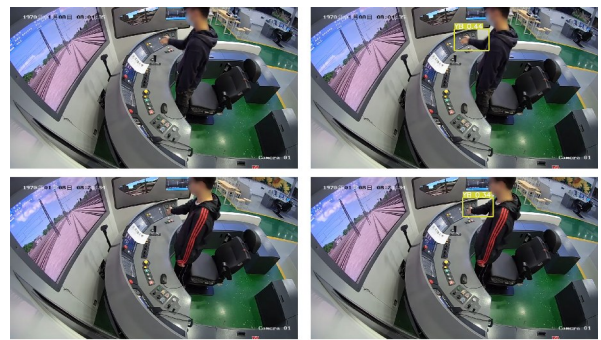


(a) YOLOV7

(b) Our algorithm

图 9 注意停车与侧线运行检测效果

Fig. 9 Ready to stop and side running detection effect



(a) YOLOV7

(b) Our algorithm

图 10 确认仪表检测效果

Fig. 10 Confirmation instrument detection effect

综上,采用融合注意力机制并添加 StrongSORT 算法可以有效地提高 YOLOV7 算法的检测效果。主要是因为本文算法具有更加强大的特征提取功能,可以更有效地注意到图像中的小细节,比如手比行为中小拇指的变化。本文算法还可以更好地注意到图像不同时刻的变化特征,比如确认仪表中不同时刻手势移动产生的变化特征。

4 结论

本文从动车组模拟驾驶实训系统采集了列车司机乘务作业时的手比行为数据集, 并使用融合注意力机制的YOLOV7卷积神经网络模型对列车司机不同类型的手比行为进行检测, 再通过StrongSORT算法增强了对手比行为的检测。本文算法可以辅助铁路院校日常驾驶培训、机务部门考核列车司机乘务作业时的驾驶行为, 是对车辆人员进行智能监测的重要一环。当然, 对列车司机乘务作业时的监测与考核远非仅有手比行为一项, 对列车司机的实时监测与有效考核的全面智能监测仍然是一项庞大且复杂的工作。

参考文献

- [1] 何新宇, 赵时璐, 张震, 等. 机器视觉的研究及应用发展趋势[J]. 机械设计与制造, 2020(10): 281-283, 287.
- [2] 郭建伟, 司军民, 赵梦露, 等. 基于全要素、全过程数据融合的城市轨道交通车辆智能运维系统[J]. 城市轨道交通研究, 2022, 25(1): 210-215.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [4] 李伟. 列车司机疲劳状态识别技术研究[D]. 成都: 西南交通大学, 2019.
- [5] 姚巍巍, 张洁. 融合目标检测和人体关键点检测的铁路司机行为识别[J]. 计算机测量与控制, 2020, 28(6): 212-216.
- [6] 所达. 城轨列车司机行车确认手势动作行为识别方法研究[D]. 北京: 北京交通大学, 2021.
- [7] 何理. 基于深度学习的铁路司机行为规范性识别系统研究[D]. 成都: 西南交通大学, 2021.
- [8] 魏秀琨, 所达, 魏德华, 等. 机器视觉在轨道交通系统状态检测中的应用综述[J]. 控制与决策, 2021, 36(2): 257-282.
- [9] 房海宽. 探索乘务司机标准化作业对乘务安全管理的重要性[J]. 时代汽车, 2021(19): 197-198.
- [10] VIOLATO E, CHAO I C I, MCCARTAN C, et al. Pointing and calling the way to patient safety: an introduction and initial use case [J]. Journal of Patient Safety and Risk Management, 2022, 27(2): 86-93.
- [11] 戚玲珑, 高建瓴. 基于改进YOLOv7的小目标检测[J]. 计算机工程, 2023, 49(1): 41-48.
- [12] 任欢, 王旭光. 注意力机制综述[J]. 计算机应用, 2021, 41(S1): 1-6.
- [13] BENAHMED H K, BENSAAD M L, CHAIB N. Detection and tracking of honeybees using YOLO and StrongSORT [C]//2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS). Piscataway, NJ, USA: IEEE, 2022: 18-23.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.
- [15] YUAN W A. Accuracy comparison of YOLOv7 and YOLOv4 regarding image annotation quality for apple flower bud classification [J]. AgriEngineering, 2023, 5(1): 413-424.
- [16] 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 [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 7464-7475.
- [17] NIU Z Y, ZHONG G Q, YU H. A review on the attention mechanism of deep learning [J]. Neurocomputing, 2021, 452: 48-62.
- [18] DU Y H, ZHAO Z C, SONG Y, et al. StrongSORT: make DeepSORT great again [J]. IEEE Transactions on Multimedia, 2023, PP(99): 1-14.
- [19] 李志华, 于杨. 基于检测的多目标跟踪算法综述[J]. 物联网技术, 2021, 11(4): 20-24.
- [20] YEH K H, HSU I C, CHOU Y Z, et al. An aerial crowd-flow analyzing system for drone under YOLOv5 and StrongSort [C]//2022 International Automatic Control Conference (CACS). Piscataway, NJ, USA: IEEE, 2022: 1-6.
- [21] 包文歧, 谢立强, 徐才华, 等. 基于YOLOv5的微型无人机实时探测方法[J]. 兵器装备工程学报, 2022, 43(5): 232-237.

Gesture Behavior Detection of Train Drivers Based on Improved YOLOV7 and StrongSORT

SONG Jichao, HUANG Wei^{**}, CHEN Zhentang, ZHOU Chengcai

(Power Technology Institute, Liuzhou Railway Vocational Technical College, Liuzhou, Guangxi, 545000, China)

Abstract: The standardization of train driver's driving behavior directly affects the running safety and state of train operation. At present, the detection of train driver's gesture behavior is insufficient. In order to effectively detect the gesture behavior of train drivers, this article uses the images of drivers' crew work collected by the EMU simulation driving system, combined with the You Only Look Once Version 7 (YOLOV7) neural network model using the fusion attention mechanism and the Strong Simple Online and Realtime Tracking (StrongSORT) algorithm, five kinds of gesture behaviors of EMU drivers during driving were tested. The experimental results show that the algorithm in this article can effectively improve the detection effect of different types of gesture behaviors when the train driver is working. The detection accuracy rate is increased by 1.2% on average, and the detection recall rate is increased by 1.9% on average. The algorithm proposed in this article will help to improve the effect of daily training and work evaluation of train drivers by railway colleges and locomotive departments, and improve the safety in the process of train operation.

Key words: train drivers; crew work; behavior detection; YOLOV7; StrongSORT; attention mechanism

责任编辑:米慧芝,陈少凡



微信公众号投稿更便捷

联系电话:0771-2503923

邮箱:gxxkxyxb@gxas.cn

投稿系统网址:<http://gxxk.ijournal.cn/gxxkxyxb/ch>