

文章编号: 1003-1421(2026)10-0188-10

引用格式: 李小平, 王建功, 钟岱兵. 基于面部特征的列车司机注意力分散检测模型[J]. 铁路运输与经济, 2026, 48(10): 188-197.
LI Xiaoping, WANG Jiangong, ZHONG Daibin. Train Driver Distraction Detection Model Based on Facial Features[J]. Railway Transport and Economy, 2026, 48(10): 188-197.

基于面部特征的列车司机注意力分散检测模型

李小平¹, 王建功², 钟岱兵³

(1. 兰州交通大学 自动化与电气工程学院, 甘肃 兰州 730070; 2. 中国铁路兰州局集团有限公司, 甘肃 兰州 730000; 3. 兰州交通大学 机电工程学院, 甘肃 兰州 730070)

摘 要: 针对现有列车司机注意力分散状态检测方法因过度依赖头部姿态角或单一眼动数据而导致的误判率高、检测精度不足的问题, 提出了一种基于面部特征的列车司机注意力分散检测模型。首先, 采用改进的 MobileNetV4 网络作为主干特征提取网络, 通过调整通用倒置瓶颈搜索块配置, 并以 CA 注意力模块替换部分 Mobile MQA 模块, 以增强模型对面部关键区域的关注度, 有效抑制无关背景信息的干扰; 其次, 引入双向特征金字塔网络 BiFPN 特征融合方法, 以优化多尺度面部特征融合; 再次, 利用多头注意力机制和加权融合对人脸关键点特征、眼部区域特征和嘴部区域特征进行多特征融合; 最后, 将融合得到的高维特征向量输入门控循环单元 GRU 网络, 以实现对注意力分散状态的序列建模与检测。在自建数据集上的实验结果表明, 本模型检测精确率达到 94.5%, 检测帧率为 60 FPS, 验证了所提方法在保证高精度的同时, 具有良好的实时性。

关键词: 铁路运输; 注意力分散; 人脸关键点; MobileNetV4; 双向特征金字塔网络

中图分类号: U298.1+2 **文献标识码:** A **DOI:** 10.16668/j.cnki.issn.1003-1421.20250604004

Train Driver Distraction Detection Model Based on Facial Features

LI Xiaoping¹, WANG Jiangong², ZHONG Daibin³

(1. School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China; 2. China Railway Lanzhou Group Co., Ltd., Lanzhou 730000, Gansu, China; 3. School of Mechanical and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China)

Abstract: To address the high misjudgment rate and low detection accuracy of existing train driver distraction detection methods, which overly depend on head posture angle features or single eye-movement data features, a facial-feature-based train driver distraction detection model was proposed. First, an improved MobileNetV4 was adopted as the backbone feature extraction network. The configuration of the general inverted bottleneck search block was adjusted, and the coordinate attention (CA) module was used to replace the Mobile MQA module, enhancing the model's focus on key facial regions and effectively suppressing interference from irrelevant background information. Second, the bidirectional feature pyramid network (BiFPN) was introduced to optimize multi-scale facial feature fusion. Third, multi-head attention combined with weighted fusion is utilized to integrate facial keypoint, eye-region, and mouth-region features. Finally, the resulting high-dimensional fused feature vector was fed into a gated recurrent unit (GRU) network to achieve sequence modeling and detection of distraction states. Experiments on a self-built dataset demonstrate that the proposed

收稿日期: 2025-06-04

基金项目: 国家自然科学基金资助项目(52462051); 甘肃省科技计划项目联合科研基金重点项目(25JRRA107)

model achieves a detection precision of 94.5% at 60 FPS, verifying that the proposed method maintains high accuracy with favorable real-time performance.

Keywords: Railway Transportation; Distraction; Facial Key Points; MobileNetV4; Bidirectional Feature Pyramid Network (BiFPN)

0 引言

注意力分散驾驶是目前交通安全领域面临的一大挑战,尤其在轨道交通领域,其危害性更是不容忽视。相比于汽车驾驶,列车驾驶的特殊性在于列车的高速运行和长制动距离,一旦发生意外,后果往往更为严重^[1]。因此,对列车司机的注意力分散进行检测,对于确保列车运行安全具有重要意义。

目前,司机注意力分散检测主要可分为3类:基于生理心理数据、基于外部环境数据、基于视觉特征的方法^[2]。基于生理心理数据的检测方法通常需要佩戴或安装特殊传感器等检测司机是否处于注意力分散状态,成本较高,且佩戴传感器也会影响司机正常驾驶^[3]。基于外部环境数据的注意力分散检测方法从方向盘、GPS、车载诊断和摄像头等收集数据,实时性较差,且容易受道路、车型等条件限制和司机个人习惯的影响,导致误判率较高^[4]。基于视觉特征的注意力分散检测方法,根据摄像头采集司机驾驶视频图像,分析司机头部姿态、眼动特征、面部表情及嘴部变化以及身体动作等特征判断其注意力分散的状态,具有非接触性、实时性等优点,逐渐成为注意力分散检测领域研究的热点方向^[5]。

Sun等^[6]提出一种新型的轻量级混合模型MAViT,利用多尺度逆残差(MSIR)块和LGMA注意力模块提取不同尺寸的关键特征、建立局部特征与全局特征的相互影响关系,增强模型在驾驶场景中的分类能力;Zhao等^[7]提出了一种基于头部姿势的注意力分散检测方法,以头部姿态作为驾驶注意力分散特征参数并建立深度学习模型进行检测;贺宜等^[8]依托实车试验,构建注意力分散行为数据集,通过改进优化MobileViT网络对注意力分散行为识别,提高检测准确性和泛化性,有效实现了对注意力分散行为的检测。陈仁祥等^[9]利用YOLOv5模型提取面部关键点参数并结合头部姿态检测算法实现了司机注意力分散识别;Li等^[10]提出一种具有倍频程状卷积混合块的轻量级网络OLCMNet,引入SE模块和PC算子对低频和低频分支的特征串联进行顺序编码,实

现特征全局信息融合;Tang等^[11]引入CaTNet模型,简化ConvNext架构并结合自注意力机制来增加表示能力;李泰国等^[12]采用PFLD检测司机面部关键点坐标,建立面部倒立摆模型,通过CNN处理注意力分散特征信息熵值,建立信息熵和司机注意力分散状态的联系,实现分类预测;Shi等^[13]基于眼动追踪提出VFOD-Net目标检测网络,结合了自下而上和自上而下的关注,实现了接近全监督模型检测性能。

综上所述,现有研究大多基于公路车辆驾驶场景构建模型,而基于列车驾驶场景的建模研究相对较少,且列车与公路车辆在驾驶环境、操作方式等方面存在显著差异,行驶速度远高于大部分公路车辆,这要求检测方法具有更快的检测效率和更高的准确性,因此上述方法不能够完全适应于列车司机注意力分散检测,故本研究提出了一种基于面部特征的列车司机注意力分散状态检测方法,使用改进后的轻量化MobileNetV4网络来进行特征提取,提高检测速度,同时调整UIB搜索块配置,使用CA注意力机制模块替换Mobile MQA模块,增强面部关键区域关注度,抑制无关背景信息干扰的同时不影响模型的实时检测速度;其次,结合BiFPN特征融合方法,优化多尺度面部特征提取;利用多头注意力机制和加权融合进行多特征融合;最后,将融合得到的特征向量输入GRU网络,实现列车司机注意力分散状态的精确检测,有效提升模型的检测精度。

1 基于面部特征的注意力分散状态检测模型

基于面部特征的注意力分散状态检测模型如图1所示,主要由视频帧采集、特征提取、多尺度特征融合、多特征融合和输出5个部分组成。将从列车司机实时驾驶视频中抽取的视频帧图像按时间序列输入模型,特征提取部分在MobileNetV4网络结构基础上引入CA注意力模块替换原始网络中的Mobile MQA模块,增强面部关键区域的特征提取。首先UIB模块的多尺度特性可以融合不同尺度的特征同时捕捉面部关键区域的细节信息和全局上下文。其次CA注意力模块能够更精准地捕捉面部

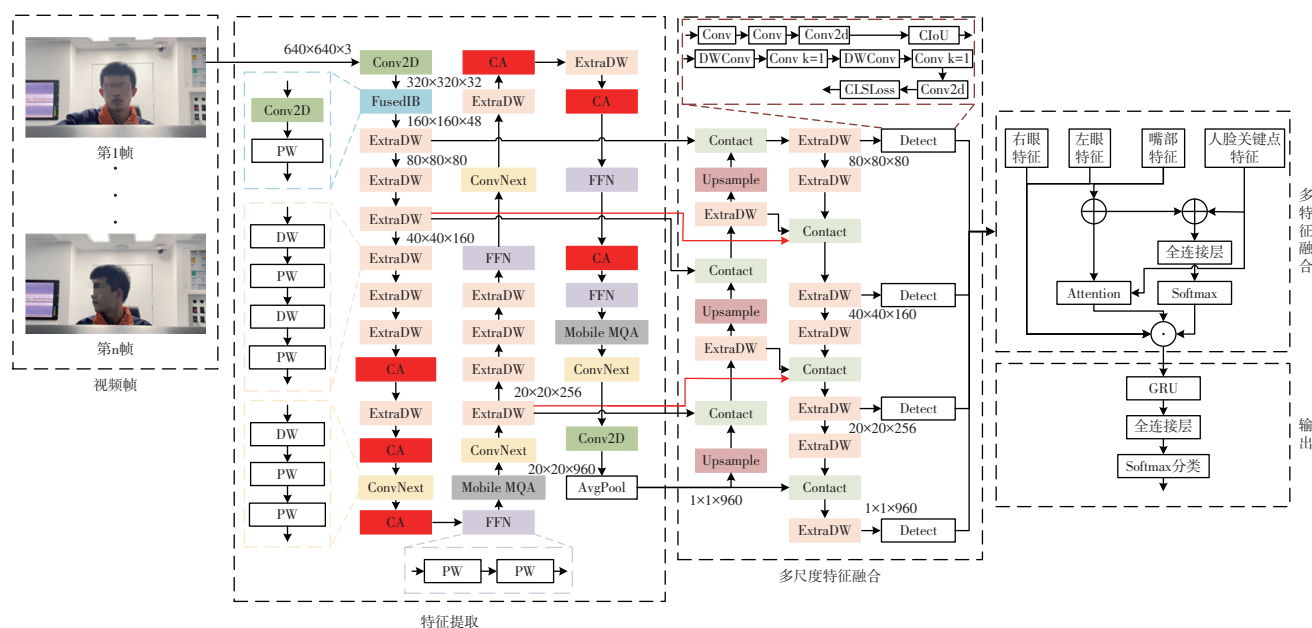


图1 基于面部特征的注意力分散状态检测模型
Fig.1 Distraction state detection model based on facial features

关键区域的空间位置信息,通过分解通道注意力为水平方向聚焦于面部中轴线附近的区域(如鼻子特征、嘴部特征);其垂直方向聚焦于左右眼特征、眉毛等纵向特征。提取得到4种不同尺度的特征图送至BiFPN特征融合网络,利用多尺度特征融合提高注意力分散状态检测能力;多特征融合部分融合左右眼特征、嘴部特征和人脸关键点特征,强化关键点对局部特征的引导作用;最后将融合得到的连续帧特征序列输入GRU中进行图像序列时间特征编码,最终得到列车司机驾驶中注意力分散状态检测结果。

1.1 基于MobileNetV4的特征提取

列车司机注意力分散状态检测模型的特征提取模块主要由改进MobileNetV4组成。

MobileNetV4是为移动设备提供的普遍高效的架构设计^[14],引入了通用倒置瓶颈(UIB)搜索块,融合了倒置瓶颈(IB)、下一代卷积网络(ConvNext)、前馈网络(FFN)以及额外深度卷积(ExtraDW),并且拥有专为移动加速器设计的注意力块Mobile MQA。因此MobileNetV4在多种硬件上检测精度和速度均优于其他模型。

UIB的4种变体如图2所示。本研究利用改进MobileNetV4网络进行特征提取,选用新变体ExtraDW模块,综合了IB与ConvNext的优势,以较低的成本

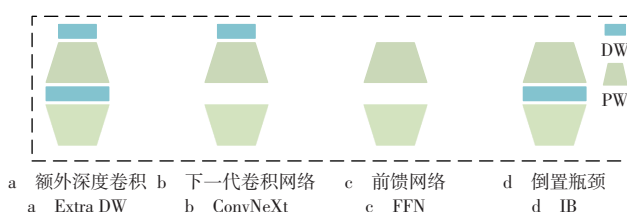


图2 UIB的4种变体
Fig.2 Four variants of UIB

本增加网络深度和感受野,适合复杂环境下的细粒度特征学习,同时添加了4个FFN和ConvNext模块,减少计算量,加速检测效率;为增强模型对面部关键区域的定位,选用CA注意力模块替换原网络的部分Mobile MQA注意力模块,保留2个Mobile MQA注意力模块为模型加速,如图1所示,在不影响实时检测的情况下增强网络对眼部区域、嘴部区域以及人脸关键点的定位,提高检测精度。

CA注意力机制同时建模通道关系和空间位置信息,通过分解坐标方向捕获长距离依赖,更适用于图像的检测、分割、定位任务^[15]。SE通道注意力机制也能为图像的检测、分类增强通道特征表达,但仅关注通道维度无法捕捉空间位置关系,且通道数较高时全连接层参数数量会显著增加;ECA注意力机制在SE的基础上通过一维卷积来捕捉通道间的依赖关系,减少了参数数量和计算量,计算速度显著提升,但可能无法捕捉长距离通道交互范围;CA相

比SE和ECA需额外分解坐标方向,但参数量与SE相近,因此本研究在拥有FFN, ConvNeXt, MobileMQA等模块为模型加速的前提下,选用CA注意力机制模块,增强人脸关键区域检测定位。

CA注意力机制将通道注意力分解为2个并行的1D特征编码过程,利用2个1D全局池化操作,分别沿垂直和水平方向聚合输入特征,生成2个包含方向特定信息的特征图后编码为2个注意力图,每个图都能够捕捉输入特征图沿一个空间方向的长距离依赖性,CA注意力机制结构图如图3所示。

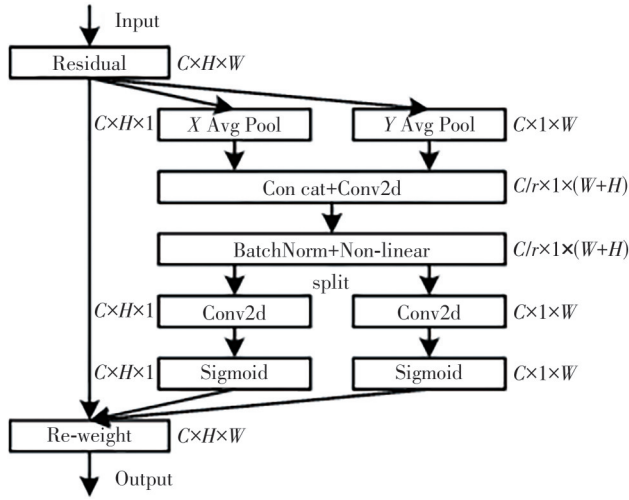


图3 CA注意力机制结构图
Fig.3 CA mechanism structure

由图3可知,以 x 作为输入,则输出为

$$CA(x) = \text{Sigmoid}\left(f_c\left(\text{Concat}\left(\text{GAP}(x_h), \text{GAP}(x_w)\right)\right) \otimes x\right) \quad (1)$$

式中: x_h 和 x_w 分别为水平与垂直方向的特征图; f_c 为卷积操作。

1.2 基于BiFPN网络的多尺度特征融合

受YOLOv11启发本研究选用BiFPN多尺度特征融合方法^[16],通过双向路径和加权融合机制,优化不同分辨率特征图的融合效果,同时减少参数量和计算量,以便实时应用。

BiFPN在传统FPN和PANet的基础上,引入了双向跨尺度连接,允许特征在不同层级间多次融合,4种常见的特征融合结构图如图4所示,删除只有一条输入边的节点,在不影响特征融合的同时简化双向网络。

每个节点接收来自多个层级的输入(如相邻层和同层级跳跃连接),并通过加权求和整合信息,对每个输入特征图分配可学习权重,动态调整不同分辨率特征的重要性,以 P_6 为例,其计算公式如下。

$$P_6^{\text{id}} = \text{Conv}\left(\frac{\omega_1 P_6 + \omega_2 \text{Resize} P_7}{\omega_1 + \omega_2 + \varepsilon}\right) \quad (2)$$

$$P_6^{\text{out}} = \text{Conv}\left(\frac{\omega_1' P_6 + \omega_2' P_6^{\text{id}} + \omega_3' \text{Resize} P_5^{\text{out}}}{\omega_1' + \omega_2' + \omega_3' + \varepsilon}\right) \quad (3)$$

式中: P_6, P_7 是BiFPN双向特征金字塔中第6、第7中间层级; P_6^{id} 为第6层的自顶向下路径中间融合特征; P_6^{out} 为第6层经过BiFPN融合后的输出特征图; Resize 为上采样或下采样; ω_i, ω_i' 为第 i 个输入特征对应的可学习权重; ε 为快速归一化融合的防震荡常数。

通过BiFPN多尺度特征融合,同时融合低层高分辨率特征(捕捉细节,如眼睛、瞳孔、嘴巴)和高层低分辨率特征(捕捉语义信息,如头部姿态),对不同层级的特征赋予可学习的权重,如在检测眼部时更关注低层细节特征;在判断头部姿态时,则侧

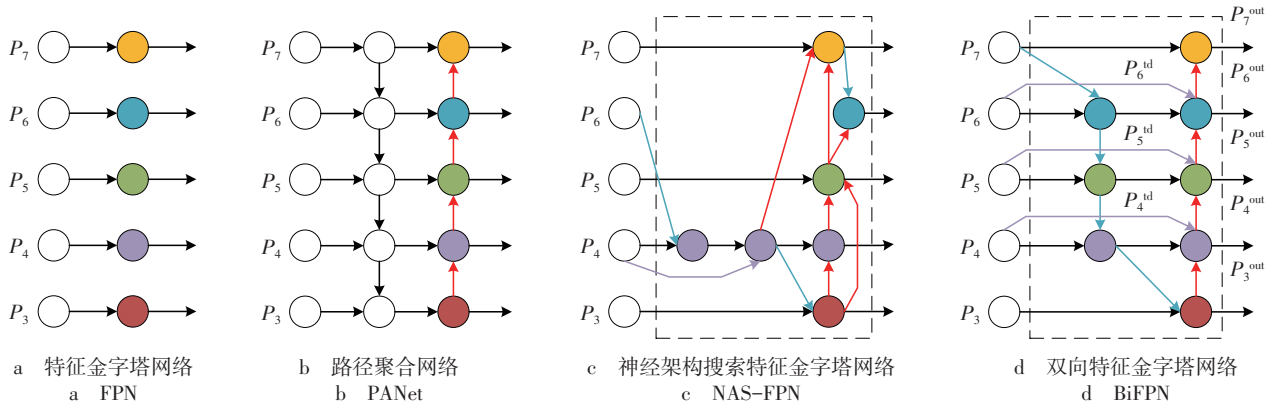


图4 4种常见的特征融合结构图
Fig.4 Four common feature fusion structures

重高层语义特征,解决单一尺度特征表达能力不足的问题。利用深度可分离卷积和2维卷积层提取得到左眼特征向量 F_{eye_l} 、右眼特征向量 F_{eye_r} 、嘴部特征向量 F_{mouth} 和人脸关键点特征向量 $F_{keypoints}$ 。

1.3 多特征融合

将多尺度特征融合输出的左眼特征 eye_l 、右眼特征 eye_r 、嘴部特征 $mouth$ 和人脸关键点特征 $keypoints$ 使用多头注意力机制动态分配权重,其矩阵分别表示为 F_{eye_l} , F_{eye_r} , F_{mouth} 和 $F_{keypoints}$,以 $F_{keypoints}$ 作为查询矩阵(Q), F_{eye_l} , F_{eye_r} 和 F_{mouth} 特征的拼接作为键矩阵(K)和值矩阵(V)。

$$Q = F_{keypoints} \quad (4)$$

$$K = V = [F_{eye_l}; F_{eye_r}; F_{mouth}] \quad (5)$$

多头注意力机制会计算查询与键之间的注意力权重,并基于这些权重对值进行加权求和,得到注意力头 A 。

$$A = \text{Attention}(Q, K, V) \quad (6)$$

将左眼、右眼、嘴巴和面部关键点的特征在维度1上进行拼接,并通过一个全连接层计算每个特征的权重,然后使用Softmax函数将这些权重归一化到0和1之间。

$$\theta = \text{Softmax}(\theta'(\text{concat}(eye_l, eye_r, mouth, keypoints))) \quad (7)$$

式中: θ' 为全连接层的权重。根据计算出的权重 θ ,对左眼、右眼、嘴巴和注意力输出进行加权求和,得到最终的融合特征 F 。

$$F = \theta_1 F_{eye_l} + \theta_2 F_{eye_r} + \theta_3 F_{mouth} + \theta_4 A \quad (8)$$

式中: $\theta_1 \sim \theta_4$ 分别为 F_{eye_l} , F_{eye_r} , F_{mouth} 和 A 的动态权重。将融合得到的独立帧图像特征向量 F 按照时间序列输入到GRU网络,以编码列车司机注意力分散状态的时间依赖性。通过全连接层和softmax层进行分类,输出列车司机注意力分散状态检测结果。

2 实验方案设计

2.1 测试人员的选取

本研究实验共招募23名测试人员,其中铁道技术学院铁路机车运用与维护专业列车驾驶实训学员15名,兰州西机务段铁路列车司机8名,男性20名,女性3名。实训学员要求具备20 h以上模拟器驾驶经验,铁路列车司机需满足10 000 km以上实际驾

驶里程。所有测试人员均身体健康、裸眼视力或矫正视力不低于4.8、无辨色能力缺陷(色盲/色弱)、无听觉功能障碍,测试前48 h内未摄入酒精以及精神类药物,精神状态评估良好。

2.2 实验设备

本实验在以下2个平台上进行:①兰州交通大学“甘肃省轨道交通服役环境与智能运维重点实验室”城市轨道交通列车司机模拟驾驶台;②兰州交通大学铁道技术学院列车司机实训室韶山7E型机车模拟仿真驾驶平台。列车模拟仿真驾驶平台如图5所示。



a 城市轨道交通列车模拟驾驶台
a Urban rail transit train simulation driving console



b 城市轨道交通列车模拟驾驶台操作场景
b Simulated driving console operation scenario



c 韶山7E型机车模拟仿真驾驶平台
c Shaoshan 7E locomotive simulation driving platform



d 韶山7E型机车模拟仿真驾驶平台操作场景
d Shaoshan 7E locomotive driver's platform operating scene

图5 列车模拟仿真驾驶平台

Fig.5 Train simulation driving platform

城市轨道交通列车司机模拟驾驶台(如图5a和图5b所示)主要由列车实时控制器、双27英寸显示屏与单68英寸显示屏构成显示系统、虚拟列车管理工作站以及双海康威视DS-U68高清摄像头组成。该驾驶台集成软件系统,可模拟城郊轨道、隧道环境等3D动画场景,支持交通参数设置、模型库调用(含有50余组模型和材质)以及动态环境模拟功能,可复现日照变化、灯光效果以及恶劣天气等复杂工况。

韶山7E型机车模拟仿真驾驶平台整体外观与内部驾驶舱分别如图5c和图5d所示,1:1还原真实机车驾驶舱,通过3D建模技术复现部分路段场景,采用曲面屏投影构建驾驶环境,集成手指口呼规范

提示、制动操作等真实作业辅助系统,同时配备双海康威视DS-U68高清摄像头完成图像数据采集。

2.3 实验设置

2.3.1 实验场景

本研究实验在2个模拟驾驶平台上进行,分别构建了3D模拟驾驶场景,模拟驾驶部分场景示例如图6所示。

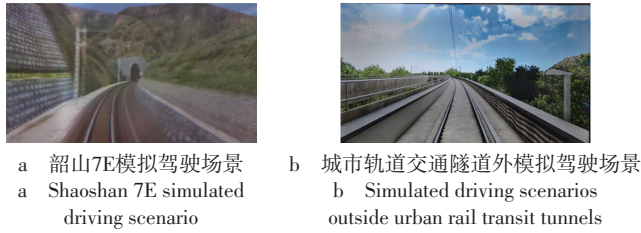


图6 模拟驾驶部分场景示例

Fig.6 Example of a simulated driving scenario

在韶山7E型机车仿真驾驶平台中,选取某铁路局集团公司阳平关—朝天南区间的模拟路线,韶山7E模拟驾驶场景如图6a所示。该路段包括站台停靠、山谷穿行以及隧道驾驶等场景模拟,驾驶时长超过1 h,环境要素变化相对单一,这种长时间、重复性的驾驶场景容易诱发列车司机注意力分散或做出与驾驶无关的动作行为,适用于列车司机注意力分散驾驶检测实验。

城市轨道交通列车模拟驾驶台设置了城郊隧道外环境测试场景,城市轨道交通隧道外模拟驾驶场景如图6b所示。该路段以直线轨道为主,局部包含缓曲线,驾驶速度维持在较低水平,持续时间约20 min。司机持续注视轨道流动视野,相对单调的视觉环境易引发注意力分散。

2个实验场景经过精心设计,均能有效模拟真实作业中注意力分散驾驶的高发情境,为列车司机注意力分散研究提供了有效的数据采集环境。

2.3.2 实验流程

本次列车司机注意力分散驾驶检测实验的开展,已通过了相关伦理道德委员会的批准。实验开始前,向测试司机详细说明研究目的、任务要求以及操作流程。再由经验丰富的列车司机指导实训学员进行30 min模拟驾驶适应性训练,熟悉平台操作界面及虚拟驾驶环境特征。实验开始后,分别进行2个驾驶阶段实验:首先进行基准驾驶:要求司机在无干扰条件下,以常规速度(韶山7E平台:120 km/h;城

轨平台:60 km/h)完成自由驾驶任务。随后进入注意力分散驾驶阶段,在此阶段,测试人员需执行一系列根据文献^[17-19]改进设计的标准化分心任务。实验全程通过高清摄像头及平台监测系统采集视频数据。

2.3.3 数据采集

为获取高质量且适用的数据,本研究采用海康威视DS-U68高清摄像头(分辨率3 480×2 160)来采集原始视频,并对所采集的视频数据进行预处理:首先,从所有视频中提取帧图像,从中筛选出9 842张能够清晰表征注意力分散驾驶行为的图像,然后利用Python脚本将图像的分辨率统一调整为640×640像素,以适配模型输入要求。同时,针对司机正面驾驶视角的典型样本IMG_5328.mp4,提取出该视频的全部帧序列,并专门将分辨率调整至480×320像素,为后续构建多尺度检测模型提供多样化数据支持。

2.4 实验验证

实验平台的硬件配置为Intel(R)Core(TM)i5-10300H, 2.50 GHz处理器,显卡为NVIDIA GeForce RTX1650Ti 4 GB,配置较低,适合轻量化部署。在Win11环境下基于Pytorch深度学习框架,Python版本为3.8,Pytorch版本为1.10.1。

由于目前没有列车司机人脸公开数据集,因此本研究采用对AFLW公开数据集和LFW数据集各随机选取5 000张图片作为训练集,选取UMDFaces数据集和DGW数据集作为测试集,自建数据集作为验证集。自建数据集数据来源于本实验设置的2个机车模拟驾驶台拍摄数据,共328个视频数据,由专业技术分析人员对视频数据中的司机状态进行了标记(注意力分散、非注意力分散),在实验过程中采用从列车司机实时驾驶视频中抽取的视频帧图像作为实验数据按时间序列输入模型中。

为了评估本研究对列车司机注意力分散检测的定量评价,实验在检测性能和模型参数2个方面进行评价。实验指标包括精确率(Precision, P),召回率(Recall, R),参数量(Parameter quantity, $Paras$)和浮点运算次数(Floating Point Operations Per Second, $GLOPs$),帧率(Frames Per Second, FPS)。计算公式如下。

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

$$R = \frac{TP}{TP + FN} \tag{10}$$

式中： TP 表示正样本被识别正确的数量； FP 与 FN 分别表示正样本与负样本被识别错误的数量。 P 和 R 值越高证明模型的检测性能越好。

$$GLOPS = 2 \times 10^{-9} HW (C_{in} K^2 + 1) C_{out} \tag{11}$$

$$FPS = \frac{1}{t} \tag{12}$$

式中： HW 表示输出特征图的维度； K 表示卷积核的大小； C_{in} 和 C_{out} 分别表示输入和输出通道的数量。

2.5 实验结果分析

2.5.1 司机注意力分散检测结果

将本研究所提模型与 VGG，YOLOv11n，MobileNetV4 和 ResNet 等主流网络模型在相同训练集上进行对比，初始学习率、终止学习率、权重衰减系数、动量因子、输入批量大小等训练参数均设置相同参数，均使用 Adam 优化器训练 50 个 epochs，本研究模型与其他模型精确率对比图如图 7 所示。

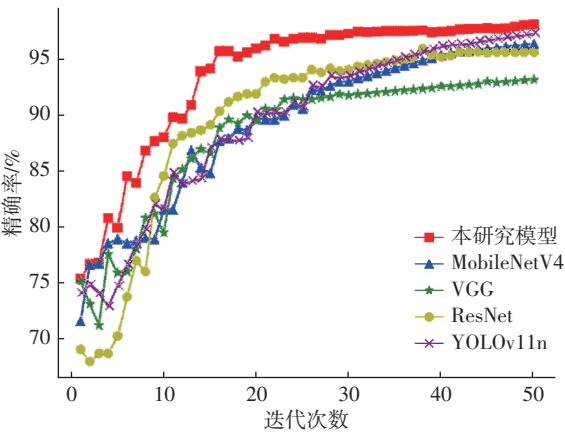


图 7 本研究模型与其他模型精确率对比图

Fig.7 Comparison of precision between this model and other models

由图 7 可知，在训练中，本研究模型迭代 25 轮后精确率趋于稳定，可达 96.88%，MobileNetV4 和 YOLOv11 模型均在 45 轮迭代后精确率开始趋于稳定，分别达到 95.8% 和 96.6%，本研究模型优势明显。将训练好的模型用验证集数据进行模型过拟合验证，然后用测试数据集进行测试。训练集与验证集的总损失曲线均在 20 个 epochs 后收敛于 0.3 以下，且两者数值接近、波动较小，表明模型未出现过拟合，具有良好的列车司机注意力分散检测能力。

部分检测结果图如图 8 所示，其中图 8a 为测试人员正常驾驶，图 8b 和图 8c 为测试人员出现注意力

分散行为但持续时间短暂不影响正常驾驶，图 8d、图 8e 和图 8f 测试人员均处于注意力分散状态。



图 8 部分检测结果图

Fig.8 Part of test results

2.5.2 注意力模块对比

为了验证 CA 注意力模块在本研究模型的有效性，采用与主流的注意力模块在同一模型下使用同一数据集和实验参数进行对比实验，注意力模块对比如表 1 所示。

表 1 注意力模块对比

Tab.1 Comparison of attention modules

注意力模块	P/%	R/%	Params/M	FLOPs/G	FPS
Mobile MQA	85.6	80.9	5.0	0.70	163
SE	86.7	83.4	5.6	0.80	115
ECA	86.4	78.9	5.1	0.70	150
CBAM	87.9	85.5	6.2	0.90	98
BAM	88.5	86.3	6.5	0.95	84
CA	89.4	87.8	5.8	0.80	110
CA+Mobile MQA	89.2	87.9	5.4	0.75	142

由表 1 可知 Mobile MQA 注意力模块作为移动加速器，其参数量($Params$)和计算量($FLOPs$)均最低，检测速度最快，因此将其替换为其他注意力机制模

块,精确率(P)、召回率(R)、 $Params$ 、 $FLOPs$ 均有所增加,其中CA注意力模块在精度上的提升效果最为明显,召回率(R)的指标也相比于其他的模型有所增加,但帧率(FPS)下降了30%左右,因此为了避免引入多特征融合影响检测的实时性,研究在替换CA注意力模块的情况下保留了2个Mobile MQA模块用以加速,在提升精确率的同时确保检测效率。

2.5.3 特征融合网络对比

为验证BiFPN多尺度特征融合网络对本研究模型的有效性,保留模型的特征提取和多尺度特征融合部分,由Detect模块的输出作为判断标准,将主流的特征融合网络FPN, PANet, NAS-FPN与BiFPN进行替换并使用同一数据集测试,对比分析,特征融合网络对比如表2所示。

特征融合网络	$P/\%$	$R/\%$	$Params/M$	PFS
FPN	84.3	78.5	4.2	108
PANet	87.5	85.3	5.8	84
NAS-FPN	89.7	86.8	7.1	69
BiFPN	89.5	87.1	4.5	98

由表2可知,NAS-FPN的精确率最高,其通过神经架构搜索优化了跨尺度连接,但参数量与计算成本过高;BiFPN精确率略低于NAS-FPN,但召回率(R)相比于NAS-FPN提升了0.3%,检测帧率提升了29,且BiFPN的精确率明显高于FPN和PANet,因此本研究选用了BiFPN特征融合网络作为多尺度特征融合的主要架构。

2.5.4 模型对比

将本研究模型与现有主流注意力分散检测模型在相同实验环境下进行对比,对比实验如表3所示。

由表3可知,本研究模型检测精确率达到94.5%,召回率达到89.8%,均优于文献[6]—[13]的模型。检测帧率达到60FPS,仅低于文献[9]的改进YOLOv5模型,能够满足实时检测需求。综合对比各项指标,本研究模型更适合在低算力设备上实现注意力分散状态的实时检测任务。

2.5.5 消融实验

为了验证所提模型中的改进模块对原网络模型在列车司机注意力分散检测的有效性,从而确定哪些模块对提升整体模型的性能最为关键,依次

表3 对比实验

文献	模型	$P/\%$	$R/\%$	$Params/M$	FPS
[6]	MAViT	88.1	75.1	12.5	25
[7]	HPE_Resnet	90.9	80.3	25.6	15
[8]	MobileViT	85.4	88.3	6.2	30
[9]	改进YOLOv5	82.3	80.4	4.7	80
[10]	OLCMNet	84.8	82.3	7.1	25
[11]	CaTNet	86.1	85.7	9.3	22
[12]	PFLD-CNN-LSTM	89.9	85.9	8.5	18
[13]	VFOD-Net	88.4	86.1	28.9	10
本研究	改进MobileNetV4	94.5	89.8	5.8	60

在基线网络模型的基础上进行了消融实验,选择以MobileNetV4模块作为基线模型,设5组消融算法,分别是:①MobileNetV4;②MobileNetV4+新变体UIB;③MobileNetV4+新变体UIB+CA;④MobileNetV4+新变体UIB+CA+BiFPN;⑤MobileNetV4+新变体UIB+CA+BiFPN+Multi-attention。消融实验如表4所示。

表4 消融实验

基线模型	新变体 UIB	CA	BiFPN	Multi- attention	$P/\%$	$R/\%$	$Params/M$	FPS
					85.4	85.3	6.2	30
	√				86.4	80.9	5.9	35
MobileNetV4	√	√			89.8	87.5	6.1	44
	√	√	√		91.9	88.4	6.0	50
	√	√	√	√	94.5	89.8	5.8	60

注:“√”表示模型添加模块。

从表4可以看出,单纯改变新变体UIB模块后,MobileNetV4网络的精确率提高了1个百分点,Params降低了0.3个百分点,检测帧率提高了5,这是因为修改后的新变体ExtraDW模块综合了IB与ConvNext的优势,以较低的成本增加网络深度和感受野,能够增强模型对面部关键区域的定位。再引入CA注意力机制的MobileNetV4网络的精确率提高了3.4个百分点,召回率提高了6.6个百分点,检测帧率提高了9。可以看出在列车司机注意力检测时能够通过给模型增加CA注意力机制,增强人脸关键区域检测定位。将BiFPN模块加入模型中,精确率提高了2.1个百分点,召回率提高了0.9个百分点,Params降低了0.1,检测帧率提高了6。因此采用BiFPN模块实现了双向跨尺度连接,允许特征在不同层级间多次融合。将新变体UIB模块、CA模块和BiFPN

模块以及 Multi-attention 模块一起引入 MobileNetV4 模型中,其精确率提高了2.6个百分点,召回率提高了1.4个百分点,Params降低了0.2,检测帧率提高了10。消融试验结果表明,提出的改进 MobileNetV4 模型在进行列车司机注意力检测时所做的4项改进均发挥了作用,虽然在效率上有所降低,但在检测精度上得到了显著提升。

3 结论

本研究针对当前列车司机注意力分散检测方法误判率较高、精确率较低的问题,提出一种基于面部特征的列车司机注意力分散状态检测模型,实现了融合头部姿态特征和眼动特征的列车司机注意力分散状态检测。

(1) 在特征融合部分利用改进 MobileNetV4 网络,选用 ExtraDW 模块增加网络深度和感受野,将部分 Mobile MQA 模块替换为 CA 注意力模块,增强人脸关键区域检测定位。

(2) 选用 BiFPN 多尺度特征融合网络,融合低层高分辨率特征和高层低分辨率特征,输出得到眼部特征向量、嘴部特征向量和人脸关键点特征向量。

(3) 使用多头注意力机制动态分配权重,融合得到独立帧图像特征向量;按照时间序列输入到 GRU 网络,以编码列车司机注意力分散状态的时间依赖性。通过全连接层和 softmax 层进行分类,输出列车司机注意力分散状态检测结果。

(4) 经实验验证,检测精确率达到94.5%,检测帧率60FPS,比现阶段主流的列车司机注意力分散状态检测算法均具有一定的优势。

参考文献:

[1] 郭凤香,曲思柔,万华森,等. 分心驾驶研究文献计量学分析[J]. 交通运输系统工程与信息, 2022, 22(6): 26-39.
GUO Fengxiang, QU Sirou, WAN Huasen, et al. A Bibliometric Analysis of Distracted Driving Research[J]. Journal of Transportation Systems Engineering and Information Technology, 2022, 22(6): 26-39.

[2] KOAY H V, CHUAH J H, CHOW C O, et al. Detecting and Recognizing Driver Distraction through Various Data Modality Using Machine Learning: A Review, Recent Advances, Simplified Framework and Open Challenges (2014-2021)[J].

Engineering Applications of Artificial Intelligence, 2022, 115: 105309.

[3] ZHAO L, YANG F, BU L G, et al. Driver Behavior Detection via Adaptive Spatial Attention Mechanism[J]. Advanced Engineering Informatics, 2021, 48: 101280.

[4] 高 珍, 陈 超, 许靖宁, 等. 基于车载视觉的端到端驾驶员疲劳检测模型[J]. 同济大学学报(自然科学版), 2024, 52(2): 284-292.
GAO Zhen, CHEN Chao, XU Jingning, et al. End-to-End Driver Fatigue Detection Model Based on In-Vehicle Vision[J]. Journal of Tongji University (Natural Science), 2024, 52(2): 284-292.

[5] 李小平, 代旭鹏, 孙守庆, 等. 列车司机手势识别方法研究[J]. 铁道科学与工程学报, 2024, 21(2): 533-544.
LI Xiaoping, DAI Xupeng, SUN Shouqing, et al. Research on Gesture Recognition Method for Train Driver[J]. Journal of Railway Science and Engineering, 2024, 21(2): 533-544.

[6] SUN H B, MA Y J. MAViT: A Lightweight Hybrid Model with Mutual Attention Mechanism for Driver Behavior Recognition[J]. Engineering Applications of Artificial Intelligence, 2025, 143: 109921.

[7] ZHAO Z P, XIA S L, XU X Z, et al. Driver Distraction Detection Method Based on Continuous Head Pose Estimation[J]. Computational Intelligence and Neuroscience, 2020, 2020(1): 9606908.

[8] 贺 宜, 鲁曼可, 高 嵩, 等. 基于 MobileViT-CA 模型的营运车辆驾驶人分心行为检测[J]. 中国公路学报, 2024, 37(1): 194-204.
HE Yi, LU Manke, GAO Song, et al. Distracted Behavior Detection of Commercial Vehicle Drivers Based on the MobileViT-CA Model[J]. China Journal of Highway and Transport, 2024, 37(1): 194-204.

[9] 陈仁祥, 胡超超, 胡小林, 等. 基于改进 YOLOv5 的驾驶员分心驾驶检测[J]. 吉林大学学报(工学版), 2024, 54(4): 959-968.
CHEN Renxiang, HU Chaochao, HU Xiaolin, et al. Driver Distracted Driving Detection Based on Improved YOLOv5[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(4): 959-968.

[10] LI P H, YANG Y F, GROSU R, et al. Driver Distraction Detection Using Octave-Like Convolutional Neural Network[J].

- IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 8823–8833.
- [11] TANG X X, CHEN Y, MA Y F, et al. A Lightweight Model Combining Convolutional Neural Network and Transformer for Driver Distraction Recognition[J]. Engineering Applications of Artificial Intelligence, 2024, 132: 107910.
- [12] 李泰国, 张天策, 李超, 等. 基于面部倒立摆模型与信息熵的驾驶员疲劳检测[J]. 交通运输系统工程与信息, 2023, 23(5): 24–32.
- LI Taiguo, ZHANG Tiance, LI Chao, et al. Driver Fatigue Detection Based on Facial Inverted Pendulum Model and Information Entropy[J]. Journal of Transportation Systems Engineering and Information Technology, 2023, 23(5): 24–32.
- [13] SHI Y, QIN L, ZHAO S X, et al. Weakly Supervised Fixated Object Detection in Traffic Videos Based on Driver's Selective Attention Mechanism[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 11478–11492.
- [14] QIN D F, LEICHNER C, DELAKIS M, et al. MobileNetV4: Universal Models for the Mobile Ecosystem[C]//Computer Vision–ECCV 2024. Cham: Springer, 2025: 78–96.
- [15] HOU Q B, ZHOU D Q, FENG J S. Coordinate Attention for Efficient Mobile Network Design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20–25, 2021, Nashville, TN, USA. IEEE, 2021: 13708–13717.
- [16] KHANAM R, HUSSAIN M. YOLOv11: An Overview of the Key Architectural Enhancements[EB/OL]. 2024: arXiv: 2410.17725. <https://arxiv.org/abs/2410.17725>.
- [17] 郑明强. 基于Stacking集成学习的驾驶分心识别模型研究[D]. 镇江: 江苏大学, 2022.
- [18] 马勇, 付锐, 袁钰, 等. 基于驾驶人分心影响程度的驾驶分心任务评级[J]. 交通信息与安全, 2014, 32(5): 47–51, 57.
- MA Yong, FU Rui, YUAN Yu, et al. Rating Driving Distraction Tasks Based on Distraction Impact on Drivers[J]. Journal of Transport Information and Safety, 2014, 32(5): 47–51, 57.
- [19] 代旭鹏. 基于深度学习的列车司机分心驾驶检测方法研究[D]. 兰州: 兰州交通大学, 2024.
- (责任编辑 吴文娟)

(上接第187页)

- [12] 王宇嘉, 贾永刚, 孙玉明, 等. 基于粒度分析理论的铁路应急处置技术规章制修订优化研究[J]. 铁道运输与经济, 2025, 47(6): 154–161.
- WANG Yujia, JIA Yonggang, SUN Yuming, et al. Formulation, Revision, and Optimization of Technical Regulations for Railway Emergency Response Based on Granularity Analysis Theory[J]. Railway Transport and Economy, 2025, 47(6): 154–161.
- [13] 冯姗姗. 国外铁路技术规章管理经验分析及启示[J]. 铁道运输与经济, 2016, 38(4): 83–88.
- FENG Shanshan. Analysis and Revelation of Experience of Railway Technical Regulation Management in Foreign Countries[J]. Railway Transport and Economy, 2016, 38(4): 83–88.
- [14] 晁阳. 高速铁路跨局作业中的设备故障应急处置规范化研究[J]. 中国铁路, 2020(2): 46–51.
- CHAO Yang. Normalization Research on the Emergency Disposal of Equipment Failure in the Cross-CR Branches Operation of High Speed Railway[J]. China Railway, 2020(2): 46–51.
- [15] 孙玉明, 贾永刚, 田雷. 我国铁路运输技术规章管理机制优化研究[J]. 铁道运输与经济, 2020, 42(11): 44–50.
- SUN Yuming, JIA Yonggang, TIAN Lei. A Study on Optimizing Technical Regulation Management Mechanism of Railway Transportation[J]. Railway Transport and Economy, 2020, 42(11): 44–50.
- [16] 孙玉明. 关于加强铁路技术规章管理体系建设的思考[J]. 铁道运输与经济, 2016, 38(2): 52–55, 66.
- SUN Yuming. Thoughts on Strengthening Establishment of Management System of Railway Technical Regulation[J]. Railway Transport and Economy, 2016, 38(2): 52–55, 66.
- (责任编辑 魏然)