

基于车载视觉的端到端驾驶员疲劳检测模型

高珍¹, 陈超¹, 许靖宁¹, 余荣杰², 宗佳琪¹

(1. 同济大学软件学院, 上海 201804; 2. 同济大学道路与交通工程教育部重点实验室, 上海 201804)

摘要: 营运驾驶员长时间疲劳驾驶是导致事故发生的重要原因, 为此, 企业在营运车辆上安装相机采集驾驶员面部视频, 基于模型和算法自动识别驾驶员的疲劳状态, 通过语音提醒甚至启用远程护航进行疲劳干预, 以此提高驾驶安全。现有的疲劳检测研究大多数都是基于面部关键点检测的算法, 该类算法对面部视频的质量要求严格。在真实的营运行车环境中, 夜晚光线过差, 相机位置安装不理想, 驾驶员面部遮挡等均会造成关键点检测失效, 从而影响模型的准确性。基于卷积神经网络(CNN)和长短时记忆神经网络(LSTM)设计了一种端到端营运驾驶员疲劳检测模型, 该模型以相机采集的驾驶员面部视频作为输入, 使用CNN网络提取视频单帧特征, 在此基础上将时序单帧特征作为LSTM网络的输入来最终识别驾驶员的疲劳状态, 实验表明, 模型的接收者操作特征曲线下面积(AUC)为0.9, 远优于现有的面部关键点模型。此外, 为了提高该模型在实际行车环境中的鲁棒性, 基于光线变化及相机变化的模拟操作在训练数据上进行了数据增强, 通过模型重训练进一步提高了模型的精度及鲁棒性。实验结果表明, 改进前, 营运车辆行车环境下模型的AUC相比实验室模型下降37.3%, 而改进后AUC仅下降9.7%, 模型的鲁棒性得到改善, 能够更好地适应复杂的营运车辆自然驾驶环境。

关键词: 车载视觉; 疲劳检测; 端到端模型; 鲁棒性

中图分类号: TP311; TP391.

文献标志码: A

End-to-End Driver Fatigue Detection Model Based on In-Vehicle Vision

GAO Zhen¹, CHEN Chao¹, XU Jingning¹, YU Rongjie², ZONG Jiaqi¹

(1. School of Software Engineering, Tongji University, Shanghai 201804, China; 2. Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

Abstract: Long-term fatigue driving is an important cause of accidents for operational drivers. To ensure driving safety, companies install cameras on operational

vehicles to collect drivers' facial videos, automatically identify the drivers' fatigue state based on a fatigue detection model, and use voice reminders or even enable remote escort to prevent fatigue. Most of the existing fatigue detection research is based on the extraction of the key points of drivers' faces, which has high requirements for video quality. However, in the real commercial vehicle environment, the detection of key points easily fails due to the poor light at night, the imperfect position of the camera and the obscured face of the drivers, thus affecting the accuracy of the model. Therefore, this paper proposes an end-to-end fatigue detection model for operational drivers with a high robustness, based on convolutional neural network (CNN) and long short-term memory neural network (LSTM). The model takes the drivers' facial videos collected by cameras as input, and uses the CNN network to extract the single-frame features of the videos. On this basis, the temporal single-frame features are used as the input of the LSTM network to finally identify the drivers' fatigue state. The experimental results show that the area under curve (AUC) of the model is 0.9, which is much superior to existing models based on facial key points. In addition, in order to improve the robustness of the model in the actual driving environment, data augmentation is applied to the training data, simulating both light and camera changes. The accuracy and robustness of the model are further improved through model retraining. Before the improvement, the AUC of the model in the actual driving environment for commercial vehicles is reduced by 37.3%, compared with the laboratory model. However, after the improvement, the AUC is only reduced by 9.7% which indicates that the robustness of the model is improved and the model can better adapt to the complex naturalistic driving environment for commercial vehicles.

Keywords: in-vehicle vision; fatigue detection; end-to-

收稿日期: 2022-05-19

基金项目: 国家自然科学基金(5217120344)

第一作者: 高珍, 副教授, 工学博士, 主要研究方向人工智能、智能交通。Email: gaozhen@tongji.edu.cn

通信作者: 许靖宁, 工学硕士, 主要研究方向为人工智能、智能交通。Email: xsimba@tongji.edu.cn



论文
拓展
介绍

end model; robustness

随着全世界交通运输业和城市道路的蓬勃发展,车辆营运数量和驾驶员人数增长迅速,道路交通事故和因车祸伤亡的人数居高不下。疲劳驾驶被认为是引发交通事故的重要原因之一,仅在美国,每年就有约1500人丧生于驾驶员疲劳驾驶引起的车祸^[1]。其中,营运车辆驾驶员疲劳驾驶引发的事故占比54%,我国有关部门对营运车辆驾驶人抽样调查结果显示:84%货运车辆驾驶人日均驾驶时间超过8h,其中40%超过12h,64%货运车辆只配备了1名驾驶人,具有巨大的交通事故隐患。因此,对营运车辆驾驶员疲劳驾驶状态进行检测、定级并做出相应的预警,对于保障交通运输安全至关重要。

疲劳的检测与分析常从驾驶员的面部特征入手,因为面部特征能够最直观、最明确地描述身体的疲倦状态,是目前研究中使用最多的一种方法。驾驶员在疲劳驾驶时,其面部特征可主要表现为抑制自发眨眼、打哈欠、头部俯仰变化频率提高等,因此能够通过眼部、嘴部、头部的疲劳状态表征和检测模型实现疲劳状态的分类。若驾驶员处于疲劳状态,各类特征值均会出现显著上升趋势,作为对驾驶员疲劳行为的定义,进而实现疲劳检测或疲劳定级。然而,在真实的营运车辆驾驶环境中,通过车载视觉系统收集的图像和视频数据极易受到车内照明条件的变化、图像或视频背景、相机位置及拍摄角度或其他因素的影响,出现图像过曝、驾驶员人脸不清晰或部分遮挡等问题,导致检测精度降低,严重影响模型性能。基于视觉特征进行驾驶员疲劳检测的方法仍存在一定的局限性,模型投入应用后往往存在召回率大幅降低等问题,急需对算法鲁棒性的提高方法进行进一步探索。

本文主要针对营运驾驶员的疲劳检测开展研究,在真实营运车辆驾驶环境下收集了901辆车、681位驾驶员的共1398条驾驶片段,其中有456条片段属于疲劳驾驶。本文的主要工作及贡献为:

(1)提出了一种融合卷积神经网络(convolutional neural network, CNN)和长短时记忆神经网络(long short-term memory neural network, LSTM)架构的端到端模型,来提取驾驶员的疲劳时序特征,在工业数据集上取得了0.83的接收者操作特征曲线下面积(area under curve, AUC)性能;

(2)在原始数据集的基础上,对实际运营环境中因天气光照的变化或者相机位置等原因可能出现的极端情况进行模拟,分别构建了光照变换、几何变换及综合变换三套增强数据集,进行了模型重训练,在工业数据集上取得了0.90的AUC性能,相较原始模型,增强模型的召回率提升了25%,也有效提升了模型的鲁棒性;

(3)使用了一种神经网络的可视化方法Grad-CAM^[18]将本文提出的模型中的CNN模块进行可视化,对模型在图像空间的关注区域进行了解读,通过可视化的方法分析了本文提出的端到端模型相比传统人脸关键点模型的优势。

1 研究综述

目前已有研究中,驾驶员的疲劳检测主要采用以下三种方法:(1)基于驾驶员生理信号的疲劳检测,该方法通常使用生理信号测量仪器,获取被测驾驶员的脑电图(EEG)、心电图(ECG)^[2]、肌电图(EMG)^[3]、眼电图(EOG)^[4]、心率变异性(HRV)^[5]等。其中脑电图(EEG)和心电图(ECG)是最常用的方法,分别用于跟踪驾驶员的大脑活动和心率^[2],如Awais等结合使用EEG和ECG两种模式,将驾驶员的疲劳程度分为了两个等级^[6]。虽然这类方法在疲劳检测问题上已被证明具有较高的准确性和客观性,但它们在现实环境中可实行度很低。在驾驶员身上连接电极或佩戴相应检测设备是一种干扰正常驾驶的入侵机制,可能导致驾驶员身体不适,干扰驾驶。(2)基于车辆状态的疲劳检测,这种方法通过分析车辆运动学数据进行疲劳检测,是一种间接的检测方法。该方法常借助车辆总线和视觉传感器等设备来监控车速、车辆在道路上的位置和方向盘转弯角度等^[7],并根据上述数据判断车辆的行驶状态以及驾驶员的疲劳状态。然而,从车辆上获取的实时数据容易受到驾驶员的驾驶习惯和外部环境的影响,因此检测的准确性与驾驶员和驾驶环境密切相关,难以应用于真实驾驶场景中驾驶员疲劳的检测或预测中。(3)基于驾驶员面部特征的疲劳检测,面部特征在当前疲劳检测研究中最广泛使用,Charlotte等的研究表明,在生理信号数据集、车辆状态数据集、面部特征数据集中,面部特征数据集的表现最佳^[8]。面部特征可包括头部姿态、打哈欠周期、眨眼频率等。Wang等^[9]提出了一种改进的人脸检测方法,并基于PERCLOS指标评估驾驶员的疲劳

状态。Ji等根据驾驶员的眼睛状态,瞳孔运动和面部位置来判断驾驶员的疲劳水平^[10]。Nawal等基于支持向量机(SVM)提取人脸,然后基于循环霍夫变换(HT)检测嘴部区域,最后基于打哈欠的特征检测驾驶员的疲劳状态^[11]。尽管人脸分类准确度已经达到了很高的水平,但精确确定驾驶员是否疲劳仍然是一个挑战,尤其是在处理视频数据时^[12]。随着深度学习技术的发展,一些研究提出了融合CNN和LSTM的LRCN算法^[13]对视频数据进行分类,能取得更好的视频分类效果。Wang等^[14]基于LRCN算法对驾驶员视频进行了疲劳检测算法,通过CNN对事先提取出的嘴部和眼部区域提取特征,然后通过DHLSTM网络进行疲劳判断。Liu等^[7]也设计了一种基于CNN-LSTM模型检测驾驶员疲劳的方法,先通过CNN网络提取驾驶员的眼部和嘴部区域,然后计算眼部闭合特征PERCLOS,统计张嘴及低头的时间占比特征,联合方向盘转角特征一起输入到LSTM网络中进行疲劳判断。这些方法^[7,14]对视频质量要求高,算法精度依赖于驾驶员眼部和嘴部的准确识别,有的还使用了视频之外的其他车载数据来源。除了算法有待提高之外,目前多数研究是在仿真实验环境下而非实车中收集驾驶员视频或图像数据。即使一些研究在实车上收集数据,也多是选择光照充足的理想驾驶环境,对相机位置及参数进行了精心设置,实验获取的原始数据清晰稳定,能够精准提取人脸关键点数据并计算疲劳特征,模型的评测性能较高。但在真实的营运车辆驾驶环境中,驾驶员图像质量极易受到天气及照明条件的变化、车厢背景、相机位置及拍摄角度等诸多因素的影响,视频画面存在色彩扰动,或人脸存在过大、过小、不能完全入画、画质不清晰等问题,导致模型的疲劳检测精度大幅降低,如何提高模型鲁棒性在现有研究中也未被深入讨论。

2 建模方法

本文提出了一种基于深度学习算法的端到端驾驶员疲劳检测模型。该模型以驾驶员面部视频作为输入,网络包含CNN及LSTM两个主要部分。其中CNN网络部分被用来从非结构化视频数据中提取驾驶员面部特征;而LSTM网络用于学习驾驶员疲劳的时间序列特征,并解决复杂场景中的长期依赖性问题。

此外,和实验室的理想测试环境相比,实际运营环境会存在由光照、天气不同而引起的车内光线变

化干扰,以及由相机安装位置及镜头角度等不同而引起的图像几何变化干扰,本文在原始训练数据集上采用图像亮度变换及仿射变换得到增强数据集,然后进行模型重训练,以进一步提升模型的泛化能力及鲁棒性。

2.1 基于驾驶员面部视频的端到端模型

2.1.1 单帧视频特征提取网络

卷积神经网络(CNN)是一种前馈神经网络。在本文使用的端到端模型中,首先从视频数据中提取单帧图片作为模型输入,经过卷积层、池化层和全连接层的处理,得到该视频帧对应的特征信息,网络结构如图1所示,图中Convolution表示卷积层,Batch Norm表示批量归一化层,ReLU表示线性整流函数即激活函数。将每一帧提取出的特征按照时间排列,即得到了原视频数据所对应的时序特征。

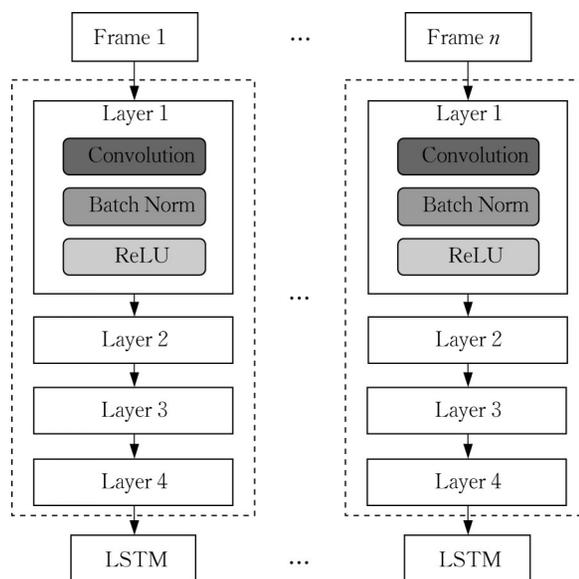


图1 CNN模型结构

Fig. 1 Structure of CNN model

在设计CNN部分的结构时,参考了经典的AlexNet^[15]图像分类网络,并在此基础上通过对网络层及卷积核参数等进行优化调整,得到了最终的轻量级CNN网络结构。具体来说,在设计CNN网络结构中,共包含4个层,其中每个层包含卷积层、批处理归一化层和激活层。且4个层对应的卷积核分别被设置为32、64、128和256,步长均设置为2,每层选用的激活层均为ReLU函数。经过连续的4次卷积操作后,一张 256×256 的RGB图像将被转化为一个 $4 \times 4 \times 256$ 的三维矢量阵列。然后对这个三维矢量矩阵进行扁平化操作,将其压缩成长度为4 096

的一维向量数组。最后将这个数组经过一层全连接层,计算得到一个长为256的一维向量,作为CNN网络对该帧图像的特征提取结果。

2.1.2 视频帧间时间序列特征提取网络

作为循环神经网络(RNN)的进化版本,长短期记忆神经网络(LSTM)能够进行长短期记忆,并有效地解决长序列训练中的梯度消失和梯度爆炸问题。在本研究中,将CNN模型提取的特征数据输入LSTM网络。此处的LSTM网络主要负责疲劳检测的分类,最终输出的是每个驾驶片段是否有疲劳现象的二分类结果。

2.1.3 端到端模型总体框架

本文提出的端到端模型整体结构如图2所示。视频中的每一帧经过前一部分的CNN提取之后,会得到对应的面部特征,之后将每一帧提取出的特征按照时间排列,转化为时序数据。本文使用的LSTM共包含两层隐藏层,每层隐藏层有256个维度的特征。在LSTM网络之后,紧跟着一个全连接层,用于输出整个模型最终的疲劳识别结果。

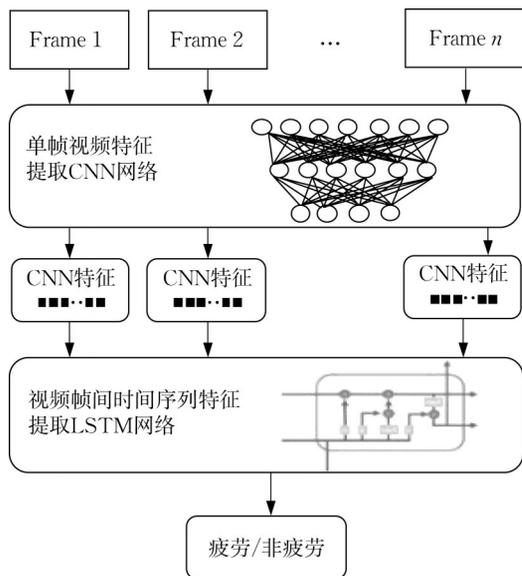


图2 端到端模型总体框架图

Fig. 2 Overall structure of end-to-end model

2.2 基于数据增强提升模型鲁棒性

2.2.1 车内光线变化模拟

实际运营环境中,车身周围环境会对车内光线情况产生较大的影响。例如,建筑物或树木的遮挡可能导致车内光线减弱;太阳光、环境车的车灯,尤其后视镜的反光等可能导致车内部分区域的光线陡然增强。此外,白天和黑夜车内光线状况也截然不

同。因此,对视频片段进行车内光线变化的模拟是十分必要且符合真实车辆运行状况的。

(1) 亮度模拟

亮度一般被定义为一个光源相对于对比光源所输出的能量大小。一般而言,亮度越高,像素越接近白色;亮度越低,像素越接近黑色。本研究中,首先设置了亮度变换随机因子 f_b ,原始图像 i_{ori} 和亮度随机变化后的图像 i_b 之间的关系表示为

$$i_b = i_{ori} \cdot f_b \quad (1)$$

(2) 饱和度变化

饱和度指颜色的纯度或强度,即颜色中的灰色量含量的高低。本研究设置的饱和度变换随机因子为 f_s ,首先通过计算得到原图 i_{ori} 对应的灰度图 i_{gray} ,则原始图像 i_{ori} 和亮度随机变化后的图像 i_s 之间的关系表示为

$$i_s = i_{ori} \cdot f_s + (1 - f_s) \cdot i_{gray} \quad (2)$$

(3) 对比度变化

对比度指数字图像中各种图像特征之间存在的颜色或灰度区分的数量。设对比度变换随机因子为 f_c ,首先通过计算得到原图 i_{mo} 对应的灰度图的像素均值 \bar{I} 。则原始图像 i_{mo} 和亮度随机变化后的图像 i_c 之间的关系表示为

$$i_c = i_{mo} \cdot f_c + (1 - f_c) \cdot \bar{I} \quad (3)$$

(4) 色相变化

色相变化指的是对原图片的色彩进行相位的变化,具体做法是:首先将原图片转换成六角锥体模型(hue saturation value, HSV)格式得到色相 H ,之后通过色相随即变换因子 f_h 进行随机变换,再将新的HSV格式的图像转换回原格式,转化公式为

$$H_{new} = H_{ori} + f_h \cdot 255 \quad (4)$$

2.2.2 相机变化模拟

由于车内环境及驾驶员本人体态特征等的差异,即使相机是按照统一的安装标准进行安装,也会造成诸如驾驶员距离镜头远近、驾驶员面部在整个视频中所处位置等不一致的情况。此外,由于车辆颠簸等原因,相机极有可能发生诸如平移、旋转之类的几何变换。对于一些模型来说,这种几何变换可能是致命的,甚至会直接导致模型失效。

在本研究中,主要对相机位置变化等引起的图像几何变换进行模拟,包括平移、旋转和缩放。各个变换对应的随机因子如表1所示。通过这三类仿射变换,能够较好地覆盖上述分析的相机引起的图像几何变换情况。

表 1 各类变换及对应的随机因子

Tab. 1 Each transformation and corresponding random factors

变换类别	随机因子	参数范围
平移	x轴平移比例/%	[-0.1, 0.1]
	y轴平移比例/%	[-0.1, 0.1]
旋转	旋转角度/(°)	[-25, 25]
缩放	缩放比例/%	[0.75, 1.25]

3 模型鲁棒性的提升

在训练端到端模型时使用 Adam 优化器进行优化,学习率被设定为 0.001,最大训练周期数约为 500,损失函数选取的是分类任务中常用的交叉熵。端到端模型的实现和训练是基于 PyTorch 平台实现,模型的训练和评估是在装有 NVIDIA Tesla K40C GPU 和 Intel Core i7 处理器的工作站上进行,模型参数量为 3.168×10^6 ,对应的每秒浮点运算次数为 17.864×10^9 ,每条 10 s 时长视频的平均推断时间为 37 ms,模型具有实时性。

3.1 数据采集

本文的数据来源于某货运公司的车载视觉系统。通过在真实的出车任务、真实的驾驶环境中录制营运驾驶员面部视频,共采集到视频数据 1 398 条,视频时长 10 s,分辨率为 352×288 。采用 3 名专家联合标定的方法将所有视频人工标定为疲劳、非疲劳两种类型,其中被标定为疲劳的视频数据共 456 条,被标定为非疲劳的视频数据共 942 条。实验中按照 7:1.5:1.5 将视频数据划分为训练集、验证集和测试集。

随机采集的视频数据共涉及 681 名驾驶员,年龄分布情况如图 3 所示。其中,90% 的驾驶员年龄集中在 35~50 岁范围内,驾驶员的平均年龄为 42.5 岁。实验包含的驾驶员人数众多,车辆型号不一且车内布置各不相同,相机的拍摄角度及位置也有明显差异。如图 4c,驾驶员着装没有统一规定,多位驾驶员存在戴帽子、佩戴眼镜、甚至墨镜,疫情期间,驾驶员戴口罩的情况也比较常见。如图 4a 和 4b,出车时间涉及白天及黑夜,覆盖了不同的光照情况。在个别视频中,如图 4d,由于相机的安装位置不当,还存在驾驶员面部采集不完整的情况,只能看到面部的一部分,比如眼睛或嘴巴。综上所述,本次研究所使用的数据集涵盖了真实运营环境中可能出现的大多数情况,具有真实性、多样性及丰富度。

3.2 基础模型的鲁棒性评估

首先,本文在某货运公司的驾驶员车载视觉数

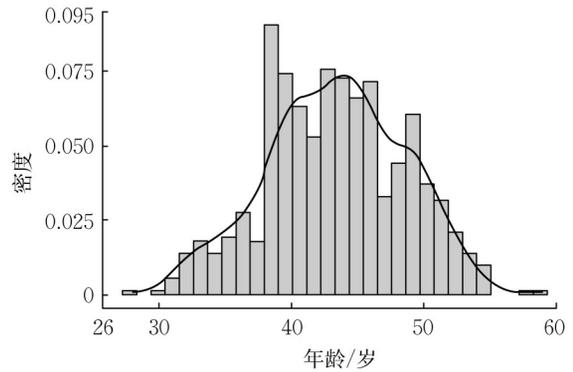


图 3 驾驶员年龄分布图

Fig. 3 Driver age distribution chart



图 4 数据集中存在问题的样本的样例图

Fig. 4 Samples with problems from the dataset

据上训练了一个端到端基础模型来识别驾驶员的疲劳状态,数据随机划分为训练集、验证集和测试集,模型在 210 条测试数据上的各项指标如表 2 所示,模型输出的分类阈值根据约登指数得到^[16],表中各项指标基于混淆矩阵进行计算。实验结果表明端到端基础模型在基础测试集上的 AUC 为 0.830,准确率达到 84.8%,如表 2 所示,能够有效识别营运驾驶员的疲劳状态。

车辆运营过程中可能会出现图像过曝、相机位置旋转过度等问题,为了充分测试模型的鲁棒性,在测试数据集上进行了车内光线变化及相机位置变化的模拟,对基础模型进行鲁棒性评估。鲁棒性强的模型在各个测试环境下理应拥有接近的性能;反之,鲁棒性弱的模型在一些特殊的测试环境下会有明显的性能指标下降。因此,使用测试环境改变时 Accuracy 和 AUC 等指标的下降程度来衡量模型的鲁棒性。

表2 基础模型鲁棒性评估

Tab. 2 Robustness evaluation of base model

模型测试环境	Accuracy/%	Recall/%	Precision/%	AUC
基础测试集	84.8	57.1	76.2	0.830
光线变化扩展测试集	83.0	57.1	69.6	0.750
相机变化扩展测试集	69.6	60.7	42.5	0.664
综合扩展测试集	68.6	39.3	37.9	0.520

从表2的实验结果来看,该模型在应对这两种变化时均有较为显著的性能下降,表中 Accuracy 表示准确率,Recall 表示召回率,Precision 表示精确率。相比光线变化,相机位置等变化引发的图像几何变换对模型性能的影响更为明显,AUC 下降达到了 20%,说明该模型对图像几何变换十分敏感。相对而言,车内光线变化对模型性能影响较小,在引入了车内光线变化模拟的测试集中,AUC 相比基础测试集下降了 9.6%,Precision 下降达到了 8.7%。该现象出现的原因可能是实验中采用的数据多数为黑白图像,因此图像亮度变化对原数据的改变较小。最后,测试了综合车内光线变化和相机相关几何变化对应的模型性能,从结果可以看出,模型在这种综合变化下基本完全失效,AUC 值已经接近 0.5,相比基础测试集,AUC 下降了 37.3%。实验结果表明端到端基础模型的鲁棒性较差,在基础测试集代表的实验室环境中其 AUC 为 0.83,但是在包含了光线变化及相机位置变化模拟的综合扩展测试集上,模型基本失效。

3.3 基于增强数据集的模型改进

在 3.2 节中,根据基础模型的鲁棒性评估结果,可以推测视车内光线变化及相机位置变化是影响基础模型鲁棒性的重要因素。为了进一步提高模型的泛化能力及鲁棒性,随机选取训练集的视频片段,进行两种变化模拟,分别构建了光照变换、几何变换及综合变换三套增强数据集,用于模型重训练。基于这三套增强数据集训练出的三个增强模型和基础模型的性能对比如表 3 所示。对应的 ROC 曲线对比图

如图 5 所示,这里的 AUC 是模型在基础测试数据集上的推断结果。从结果可知,增强模型相较于基础模型来说,模型性能均有明显提升,其中基于综合变换的增强模型泛化能力提升最为显著,AUC 达到了 0.9,相比基础模型,AUC 提升了 8.4%。

表3 三个增强模型和基础模型的性能对比

Tab. 3 Performance comparison between three augmented models and base model

模型测试环境	Accuracy/%	Recall/%	Precision/%	AUC
基础测试集	84.8	57.1	76.2	0.830
光照增强数据模型	84.8	42.9	92.3	0.846
几何增强数据模型	86.6	46.4	100.0	0.880
综合增强数据模型	88.4	71.4	80.0	0.900

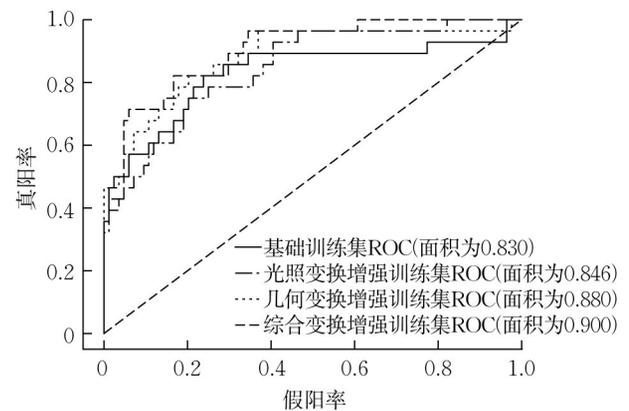


图5 增强模型的 ROC 曲线对比

Fig. 5 ROC curve comparison of enhanced models

此外,为了检验增强模型的鲁棒性,本文测试了增强模型在综合扩展测试集上的性能,如表 4 所示,模型鲁棒性对比如图 6 所示。从实验结果可以看出,改进后的三个增强模型,在应对光线及相机位置及参数变化时,均体现出了更好的鲁棒性。改进前,模型在基础测试集上 AUC 为 0.83,在综合扩展测试数据集上 AUC 为 0.52,性能下降 37.3%;改进后,模型在基础测试集上 AUC 为 0.9,在综合扩展测试数据集上 AUC 为 0.813,性能下降 9.7%,相对改进前,模型的鲁棒性有明显的改善。

表4 基础模型与增强模型的鲁棒性对比

Tab. 4 Comparison of robustness between base model and enhanced model

模型测试环境	改进前的基础模型			改进后的增强模型		
	Accuracy/%	AUC	AUC 下降/%	Accuracy/%	AUC	AUC 下降/%
基础测试集	84.8	0.830		88.4	0.900	
光线变化扩展测试集	83.0	0.750	9.6	83.0	0.828	8.0
相机变化扩展测试集	69.6	0.664	20.0	84.8	0.828	8.0
综合扩展测试集	68.6	0.520	37.3	79.4	0.813	9.7

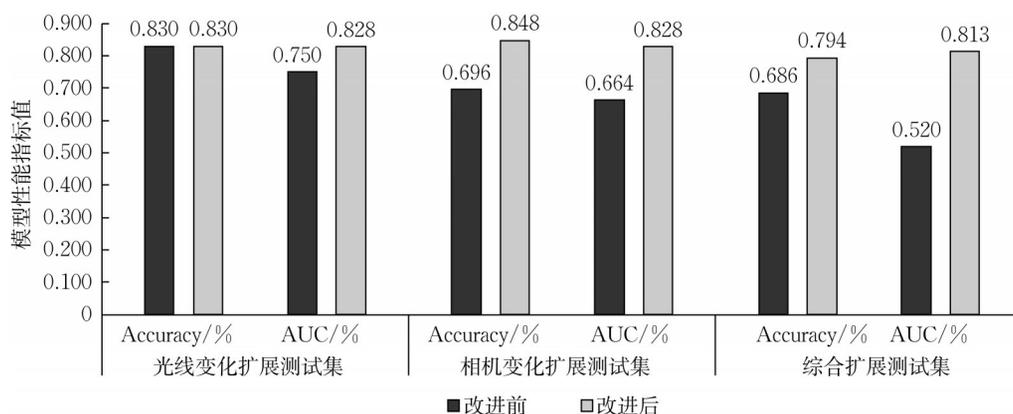


图6 基于增强数据集改进前后模型鲁棒性对比

Fig. 6 Comparison of model robustness before and after improvement based on enhanced dataset

4 模型解读及性能对比

4.1 模型解读

由于基于深度学习算法的端到端模型是黑盒模型,为了探究模型是否按照预期捕获了疲劳的相关特征,使用了一种神经网络的可视化方法 Grad-CAM^[17]将本文提出的模型中的CNN模块进行可视化,结果如图7所示,图中高亮部分即为模型的重点关注区域。对于图7a对应的视频片段,可以看到模型的关注重点主要有两块,分别是眼部和嘴部,说明模型与预期一致地捕捉到了驾驶员眼部闭眼和嘴部张开打哈欠的特征。在图7b中,虽然由于角度原因眼部特征较难抓到,但是模型成功地抓捕到驾驶员嘴部张开打哈欠的信息,因而成功地完成识别。图7c对应的情况和图7b相反,由于摄像头位置问题,没法完整捕捉驾驶员嘴部信息,因此模型主要通过驾驶员是否闭眼完成判断。在最后一张图7d中,可以看到驾驶员在打哈欠的时候做了一个捂嘴的动作,这种动作会导致人脸关键点完全捕捉不到,严重影响基于关键点的模型性能,而端到端模型则能够对驾驶员打哈欠的手部动作进行较好的捕捉,驾驶员的疲劳状态被正确识别。模型可视化结果表明,即使在光线不足、驾驶员面部视频不完整、采集角度不理想等多种复杂的自然驾驶情况下,设计的端到端模型仍然能够有效捕捉疲劳相关特征,完成驾驶员疲劳状态的正确识别。

4.2 模型性能对比

现有疲劳评估的研究大多数是基于驾驶员面部关键点的提取进行建模。主要分为两大类,基于关键点计算面部统计特征建模或基于关键点计算面部时序特征建模。在第一类方法中,研究者^[18]通常使

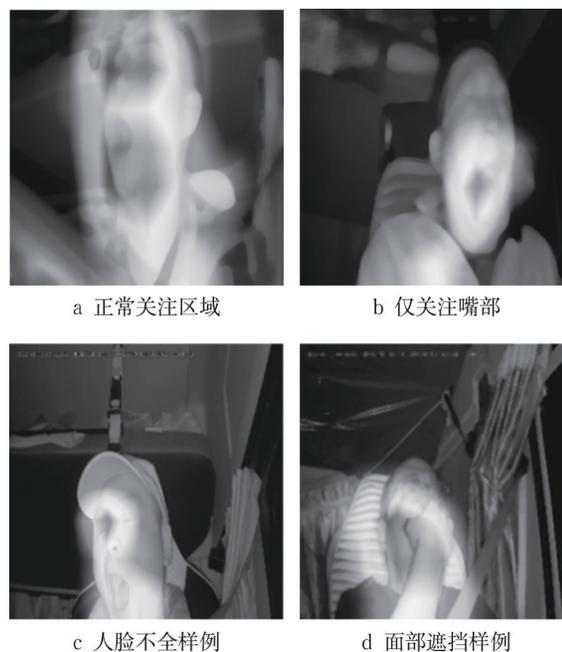


图7 端到端模型可视化结果

Fig. 7 Model visualization results in this paper

用算法检测脸部的68个或更多关键点,根据每帧的眼部及嘴部关键点坐标判断该时刻眼睛及嘴巴是否闭合,也有文献^[19]根据鼻尖及下颌等部位的坐标进行头部运动情况的判断,然后在视频的连续图像帧上计算特征统计值,最后基于规则进行疲劳状态的判断,如连续闭眼3s则视为疲劳,或者根据单位时间内眼睛闭合频率PERCLOS值^[20]进行疲劳判断,称这一类算法为面部关键点检测统计模型,他们都是先使用算法检测面部关键点,然后依据关键点坐标计算的眼部、嘴部及头部的统计特征进行疲劳判断。而第二类方法,研究者认为疲劳是一个连续的状态,通过面部时序特征进行疲劳判断会更准确,这类算法^[21]通常先检测面部关键点,在每一帧上计算

包括眼部闭合、嘴部闭合、头部倾斜角等在内的面部特征,然后将这些特征按照时间排列,通过RNN/LSTM等时序网络判断疲驾驶员的疲劳状态,称这类算法为面部关键点检测时序模型,这类算法往往比第一类算法的精度更高。但是这两类方法存在两个较严重的问题。首先它们都严重依赖于关键点检测算法(如DLIB)的精度,在复杂的营运车辆自然驾驶环境中,由于相机角度位置的不同会导致人脸画面捕捉不全、或面部被手遮挡、或佩戴墨镜、口罩等均会引起关键点检测失效,直接造成疲劳评估模型的性能下降;其次这类算法仅检测面部部分区域进行疲劳判断,如眼部、嘴部等区域,往往会将驾驶员低头看手机等分神驾驶误判为疲劳驾驶。因此要在营运车辆的自然驾驶环境中进行疲劳状态判断,必须要提取驾驶员面部甚至头部更全面和丰富的特征,如面部表情,频繁抬眼,捂嘴、转头、揉眼等肢体动作等,设计的端到端模型对包含驾驶员面部影像的整帧图像进行高层特征提取,增强了更多特征提取的可能性,能够有效避免因关键点检测失效而造成的误判。

将面部关键点检测统计模型和面部关键点检测时序模型应用于本文的实验数据,与端到端模型性能进行对比,结果见表5。对于关键点检测统计模型的构建,是在视频帧上基于DLIB算法提取面部68个关键点,在10 s视频(帧率为10 帧·s⁻¹,合计100幅视频帧)上根据嘴部及眼部关键点坐标统计计算得到打哈欠次数、单位时间内眼睛闭合频率(PERCLOS)、嘴巴张开次数、最大闭眼时间和点头次数,基于这5个特征训练随机森林模型判断驾驶员是否疲劳,得到DLIB-STATS模型,模型AUC为0.58。而对面部关键点检测时序模型,则是根据关键点坐标分别计算每帧的左、右眼纵横比值、嘴巴纵横比值作为时序特征输入到LSTM模型中判断驾驶员疲劳情况,得到DLIB-LSTM模型,模型AUC为0.61。在复杂的营运车辆行车环境下,提出的端到端的模型AUC为0.9,远优于这两种主流的基于面部关键点检测的疲劳评估算法。

表5 模型性能对比

Tab. 5 Comparison of model performance

模型	Accuracy/%	AUC
本文端到端模型	88.4	0.90
面部关键点检测统计模型(DLIB-STATS)	68.6	0.58
面部关键点检测时序模型(DLIB-LSTM)	69.9	0.61

5 结语

本文基于CNN和LSTM网络构建了端到端的疲劳检测模型,对包含驾驶员影像的视频帧进行高层时序特征提取,能够有效提高自然驾驶环境下营运驾驶员的疲劳检测精度。为了模拟营运车辆复杂行车环境下的光线变化及相机位置变化,本文对训练数据集的视频帧进行了光照和几何变换,得到了增强数据,进而对模型进行重训练。实验结果表明,增强模型的AUC提升了8.4%,达到0.9。在包括了光线变化模拟及相机位置变化模拟的综合扩展测试集上,基础模型AUC下降37.2%,而增强模型AUC仅下降9.7%,优化后的模型鲁棒性有了明显提升,有效降低了自然驾驶环境中因光线变化或相机位置移动对模型精度带来的影响。后期拟采用对抗训练的方法进一步提升模型的鲁棒性。在实验过程中也发现,车内背景会对模型的识别造成一定的干扰。在未来的工作中,也拟通过添加注意力机制等方法降低驾驶员背景的干扰。另一方面,也将尝试通过分阶段处理先对视频数据中的无关信息进行过滤,例如通过目标检测的方法先提取出驾驶员面部区域,再通过深度神经网络进行疲劳特征提取,以进一步提升营运驾驶员疲劳检测的准确性。此外,未来在算法实车部署时还需要对算法功耗、可靠性及安全性进行评估及研究。

作者贡献声明:

高 珍:实验方法设计、论文撰写、审阅与修订。

陈 超:实验数据整理、实验数据分析、论文撰写。

许靖宁:实验数据分析、模型构建、论文撰写。

余荣杰:提供研究资源、实验方法设计、论文审阅与修订。

宗佳琪:实验数据分析、模型构建、论文撰写。

参考文献:

- [1] MONAGI H A, WAZIR Z K, QURATULAIN A. Detecting human driver inattentive and aggressive driving behavior using deep learning: Recent advances, requirements and open challenges[J]. IEEE Access, 2020, 8(1): 105008.
- [2] MOUJAHID A, DORNAIKI F, ARGANDA-CARRERAS I, *et al.* Efficient and compact face descriptor for driver drowsiness detection [J]. Expert Systems with Applications, 2021, 168(12):114334.
- [3] TIBERIU V, ALEXANDRU R, TUDOR M, *et al.* Vehicle driver drowsiness monitoring and warning system [C]// CONAT 2016 International Congress of Automotive and Transport Engineering. Cham, Switzerland: Springer, 2017:

- 873-880.
- [4] CHI Z, HONG W, FU R. Automated detection of driver fatigue based on entropy and complexity measures [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(1):168.
- [5] LI G, CHUNG W Y. Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier[J]. *Sensors*, 2013, 13(12):16494.
- [6] MUHAMMAD A, NASREEN B, MICHEAL D. A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability [J]. *Sensors*, 2017, 17(9):1991.
- [7] LIU M Z, XU X, HU J, *et al.* Real time detection of driver fatigue based on CNN-LSTM [J]. *IET Image Processing*, 2022, 16(2): 576.
- [8] CHARLOTTE J D N, CHRISTOPHE B, ANCA S, *et al.* Detection and prediction of driver drowsiness using artificial neural network models[J]. *Accident Analysis and Prevention*, 2019,126:95.
- [9] WANG Q, YU F L, SONG L X. Driver fatigue detection system based on DSP [J]. *Proceedings of SPIE — The International Society for Optical Engineering*, 2012, 8335(5):55.
- [10] QIANG J, YANG X. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance [J]. *Real-Time Imaging*, 2002, 8(5):357.
- [11] ALIOUA N, AMINE A, RZIZA M. Driver's fatigue detection based on yawning extraction [J]. *International Journal of Vehicular Technology*, 2014, 2014:47.
- [12] ZHAO Y F, XIE K, ZOU Z Z, *et al.* Intelligent recognition of fatigue and sleepiness based on inceptionV3-LSTM via multi-feature fusion[J]. *IEEE Access*, 2020, 8: 144205.
- [13] DONAHUE J, HENDRICKS L A, ROHRBACH M, *et al.* Long-term recurrent convolutional networks for visual recognition and description [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4):677.
- [14] WANG Y, HE Z, WANG L. Truck driver fatigue detection based on video sequences in open-pit mines [J]. *Mathematics*, 2021, 9(22): 2908.
- [15] KRIZHEVSKY A, SUTSKEVER I, GEOFFREY E H. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6):84.
- [16] GAO Z, LIU Y J, ZHENG J Y, *et al.* Predicting hazardous driving events using multi-modal deep learning based on video motion profile and kinematics data [C]// 2018 IEEE International Conference on Intelligent Transportation Systems (ITSC). [S.l.]:IEEE, 2018: 3352-3357.
- [17] SELVARAJU R R, COGSWELL M, DAS A, *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization [J]. *International Journal of Computer Vision*, 2020, 128(2):336.
- [18] SHRUTI M, SHRUTI V H, SUPRIYA P, *et al.* Design of real-time drowsiness detection system using Dlib [C] //2019 IEEE International WIE Conference on Electrical and Computer Engineering: WIECON-ECE 2019. Bangalore, India: Institute of Electrical and Electronics Engineers, 2019: 229-232.
- [19] 苏锦瑾,霍春宝,王特特. 基于面部特征的驾驶员疲劳检测系统[J]. *信息技术与信息化*, 2021(9):114.
SU Jinjin, HUO Chunbao, WANG Tete. Driver fatigue detection system based on facial features [J]. *Journal of Information Technology and Informatization*, 2021(9):114.
- [20] JUNAEDI S, AKBAR H. Driver drowsiness detection based on face feature and PERCLOS[C]//International Conference on Computation in Science and Engineering (ICCSE 2017). Bandung, Indonesia: Institute of Physics, 2017:293-299.
- [21] CHEN L, XIN G, LIU Y, *et al.* Driver fatigue detection based on facial key points and LSTM [J]. *Security and Communication Networks*, 2021, 2021(8):1.