

DOI:10.19651/j.cnki.emt.2210750

复杂环境下多模态特征融合的疲劳驾驶检测*

高绮煌¹ 谢凯¹ 贺正方² 文畅³ 贺建彪⁴ 张伟⁴

(1. 长江大学电子信息学院 荆州 434023; 2. 浙江宇视科技有限公司 杭州 310051;
3. 长江大学计算机科学学院 荆州 434023; 4. 中南大学计算机学院 长沙 410083)

摘要: 为了避免因疲劳驾驶而导致交通事故的发生,维护城市道路交通和驾乘人员生命安全,该项目针对传统疲劳驾驶检测方法存在着精度低、参数复杂、泛化能力差等核心问题,采用MTCNN模型和基于红外的rPPG等理论,在光照变化、部分遮挡和头部偏转等复杂行车环境下精确提取驾驶员面部与生理信息;同时在深层挖掘多模态的特定疲劳信息后,结合多损失重构(MLR)的特征融合模块利用各模态间的互补信息,避免了单模态检测方法存在的局限性,进一步构建了多模态特征融合模型,增强模型的准确性与鲁棒性;最后考虑到疲劳的时序性,基于Bi-LSTM模型建立了疲劳驾驶检测模块。在自制数据集FAHD上展开实验,证明了红外生理特征提取模型的可靠性,多模态特征输入的有效性,同时与现有融合方法相比,本文方法融合后的预测结果与疲劳标定值间的相关系数提高了5.6%,均方根误差减少25%,疲劳检测系统准确率达到了96.7%,在推动智慧交通发展的同时对维护交通安全也有较好的积极意义。

关键词: 多模态;多损失重构特征融合;疲劳驾驶;远程光电容积描记术;深度学习

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Fatigue driving detection with multi-modal feature fusion in complex environments

Gao Qihuang¹ Xie Kai¹ He Zhengfang² Wen Chang³ He Jianbiao⁴ Zhang Wei⁴

(1. School of Electronic Information, Yangtze University, Jingzhou 434023, China; 2. Zhejiang Uniview Technologies Co., Ltd., Hangzhou 310051, China; 3. School of Computer Science, Yangtze University, Jingzhou 434023, China; 4. School of Computer Science, Central South University, Changsha 410083, China)

Abstract: In order to avoid the occurrence of traffic accidents caused by fatigue driving and to safeguard urban road traffic as well as occupant safety, this project addresses the core problems in traditional fatigue driving detection methods, such as low accuracy, elaborate parameters and poor generalization, by using the MTCNN and infrared-based rPPG to accurately extract driver's facial and physiological information in complex driving environments with changing light, partial occlusion and head deflection. At the same time, after deep mining the specific fatigue information of multi-modal modes, combined with the multi-loss reconstruction (MLR) feature fusion module to use the complementary information between each mode are employed to further construct the multimodality feature integration model, which breaks the limitation of single-mode detection methods and improves its the accuracy and robustness in complicated driving environments. Finally, by using the time-series nature of fatigue, a fatigue driving detection system based on the Bi-LSTM model is established. Experiments were conducted on a home-made dataset FAHD, which demonstrated the reliability of the infrared physiological feature extraction model. In addition, the accuracy of multimodal input increased by at least 5.6% compared to the single-modal input, while the correlation coefficient improved by 5.6% and the root mean square error was reduced by 25% compared to existing fusion methods, achieving an accuracy of 96.7%. While promoting the development of intelligent transportation, it also has a good positive significance for the maintenance of traffic safety.

Keywords: multimodal; multi-loss reconstruction feature fusion; fatigue driving; remote photoplethysmography; deep learning

0 引言

疲劳驾驶是指人在驾驶车辆时持续重复着简单、机械

的动作而出现驾驶机能下降的情况^[1],主要表现为注意力分散、打瞌睡、判断能力下降、操作迟误或过早以至发生追尾、碰撞、偏离道路等交通事故,严重威胁到道路安全。因

收稿日期:2022-07-20

* 基金项目:国家自然科学基金(62272485)、新疆维吾尔自治区自然科学基金(2020D01A131)项目资助

此探究驾驶员疲劳的机理,研制疲劳驾驶检测系统,准确地检测出驾驶员的实时状态并做出提醒,对促进城市交通的良好发展和保障人们出行安全有着重大意义。

现有的疲劳驾驶检测方法大致由主观检测、单模态特征检测和多模态特征检测^[2-4]组成。主观检测主要通过各类调查问卷和主观量表进行评估,但这类方法个体差异较大,且花费时间较长,无法满足实时检测和预防需求。

单模态特征检测是基于单一模态的特征来判断疲劳,包括车辆特性特征^[5-6]、生理特征^[7-10]、视觉特征^[11-16],同时为了提高驾驶员疲劳检测的准确性,利用多种同模态特征结合^[17-19]来检测疲劳。例如 Yu 等^[18]提取融合不同场景下驾驶员头部、眼睛和嘴巴的时空特征表示; Qiu 等^[19]采集 EEG、ECG 和近端皮肤温度信号对受试者进行了疲劳评估实验。结合同模态多特征相比于单一特征,该类方法精度都得到了明显的提高,但人们对于疲劳驾驶检测的要求不断提高,因此急需开展多模态特征融合理论模型的研究,设计一个高精度、鲁棒性强,且兼备便携性和实时性的检测系统。

现有的多模态融合大多基于决策融合和特征融合的方法来实现,例如获取司机的驾驶行为、车辆运行参数、面部信息及生理参数等多种模态信息,对这些信息进行融合来分析疲劳状态。Sun 等^[20]将上下文特征、驾驶员面部特征和车辆行为特征输入三个独立的多类支持向量机(MCSVM)以融合多模态信息,通过双层融合结构先后进行特征融合和决策融合; Du 等^[21]提出了一种新型的多模

态融合递归神经网络(MFRNN)融合了心率、眼睛开合度和嘴巴开合度,将3个特征的时间信息输入到包含两级RNN的新关系层来识别其间关系。Huang 等^[22]设计了多粒度深度卷积模型(RF-DCM),通过特征校正子网络和特征融合子网络,自适应重校准和融合多粒度特征,并利用长短期记忆网络(long short-term memory, LSTM)来探寻序列帧之间的关系。该类方法的准确率和鲁棒性都得到了一定的提升,但大部分并没有很好地利用到多模态特征之间的高阶交互表达信息,同时在实际行车环境中,还存在光照变化、人脸较大幅度偏转、俯仰以及部分遮挡等复杂情况影响检测精度。

为了应对以上种种挑战,本文构建了一个复杂环境下的特征提取模型,并利用各模态间的深层疲劳信息设计了多模态融合框架,通过多损失重构(multi-loss reconstruction, MLR)的特征融合模块充分利用了多种模态数据间的特定性与互补性,考虑到疲劳的时序性,建立基于双向长短期记忆网络(Bi-LSTM)模型的疲劳驾驶检测模型,最后通过实验验证本文方法的有效性。相比于 MCSVM 和 RF-DCM 方法,本文构建的检测模型精度较高,鲁棒性强,且在实际驾驶场景中实用性较好,对于实现道路的智能交通,维护城市交通安全具有一定的参考价值。

1 原理与方法

本文构建的多模态融合疲劳驾驶检测模型框架如图1所示。

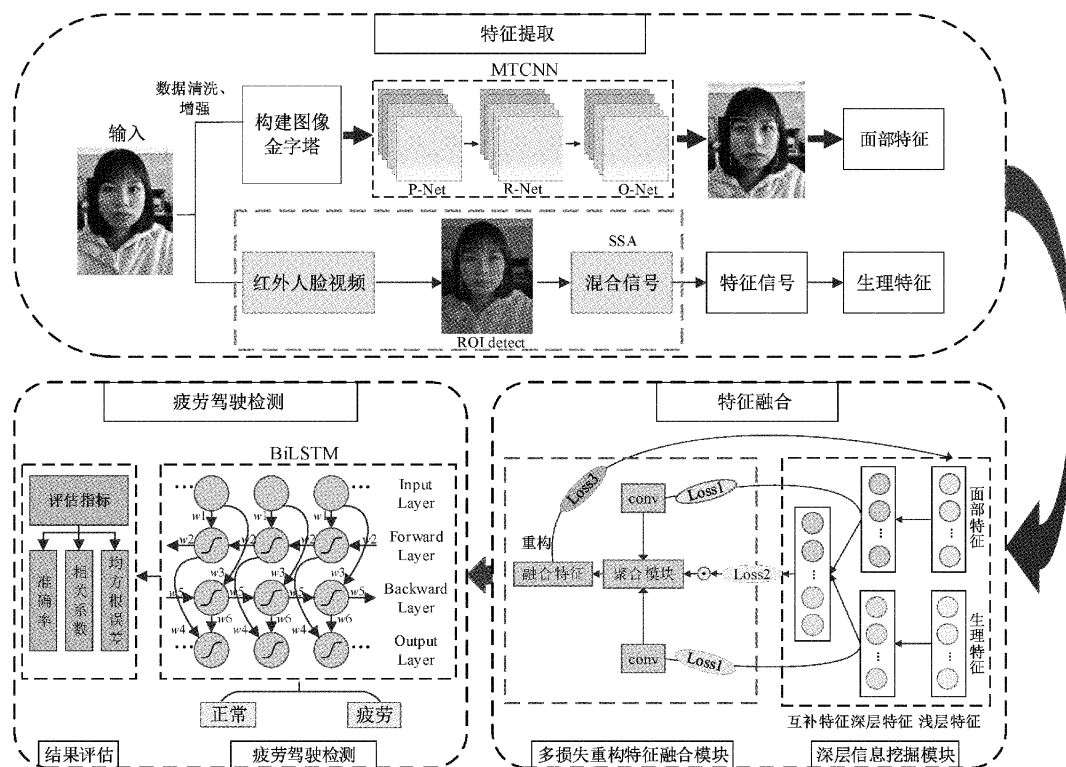


图1 模型总体框架

该模型主要包括以下 3 个部分:

1) 多模态特征提取。该部分主要将采集到的 RGB 和红外视频数据通过清洗增强后,分别输入面部特征提取和生理特征提取模型中,得到各模态间关于疲劳的隐藏表征。

2) 多模态特征融合。通过深层信息挖掘模块分别提取出模态间的深层信息和交互信息,得到更具表达的疲劳特征。同时设计 MLR 特征融合模块,在避免冗余信息、加强信息表达的基础上将多种深层模态信息进行融合。

3) 疲劳驾驶识别模型。将融合后的多模态特征输入到 Bi-LSTM 网络中,进一步挖掘特征与疲劳间的时序关系,达到准确、高效检测疲劳驾驶的目的。

1.1 多模态特征提取

1) 面部特征提取

为了提取到行车途中驾驶员疲劳时的面部特征,本文采用多任务卷积神经网络(multi-task convolutional neural network, MTCNN)作为面部特征提取模块的核心。MTCNN 算法由 3 个级联网络(P-Net、R-Net、O-Net)构成,用来检测人脸及关键点定位,比传统算法更加快速准确,同时可以减少不同光照带来的影响,在人脸较大角度偏转、俯仰以及部分遮挡等情况下依然能够对面脸和关键点进行准确地检测与定位,因此非常适用于在行车途中对驾驶员的检测。

对输入的 N 段视频进行 20 帧/s 的帧采样,得到 n 组帧图像,将这些图像尺寸缩小到原图像的 0.5 倍并组成 n 个 X_i 样本集。 X_i 样本集作为 MTCNN 网络的输入,计算如下:

$$L_i(\det, box, landmark) = \overline{MTCNN}(X_i) \quad (1)$$

其中, $i \in [1, n]$, $L_i(\det, box, landmark)$ 表示 X_i 通过网络输出得到的人脸候选框和关键点。其中 $L_i(\det, box, landmark)$ 包括了 5 个人脸关键点(左眼、右眼、鼻子和嘴唇的左右两角)的坐标值。

得到了关键点精确的坐标值后,利用关键点坐标值提取出眼部和嘴部区域的图像,并对该区域图像的眼部和嘴部状态进行分类。首先连接左、右眼中心点得到 a 线,并将该连线 a 与水平线形成的夹角定义为 α , 该眼部区域框的宽、高分别定义为 w 和 $h = w/2$ 。然后连接嘴唇左右两角得到 b 线,并将鼻子关键点对该连线 b 作垂线 c , 定义垂距为 d , 则嘴部区域框的上边沿为 $d/2$, 下边沿为该垂线 c 的延长线 $5d/3$ 处,在得到眼部和嘴部区域框后对其进行二分类。以 60 s 为间隔,共有 1 200 帧图像,通过眼睛闭合时间比例(PERCLOS)准则和先验知识,可知:

$$P = \frac{\text{眼睛闭合帧数}}{\text{检测时间段总帧数}} \times 100\% \quad (2)$$

$$L = \frac{\text{打哈欠的帧数}}{\text{检测时间段总帧数}} \times 100\% \quad (3)$$

PERCOLS 算法被证明确实能做到实时地、准确地对

驾驶员进行疲劳判定。同时根据先验知识,通过检测受试者每分钟打哈欠的次数可判定驾驶员是否处于疲劳状态。因此选取 P 值与 L 值作为面部特征。

2) 生理特征提取

研究表明:人的精神状态可由自主神经系统的活动状态体现,而心率变异性(heart rate variability, HRV)是人体内每次心跳时 R-R 间期的细微涨落变化,最能反映自主神经系统的活动状态。Chang 等^[23]证实了 HRV 信号与疲劳驾驶之间的联系,指出其频段指标中低、高频的比值(LF/HF)在清醒和疲劳状态间存在明显差异,反映出疲劳驾驶时自主神经系统发生的变化。

为了精确地得到心电信号中的心率(heart rate, HR)及 HRV 信号,并且考虑到司机开车的便捷性,我们采用了远程光电容积描记术(remote photoplethysmography, rPPG)实现无接触监测心率,很好地避免了传统方法带来的缺陷,在保证准确性的同时也非常便携稳定。但传统 rPPG 方法容易受到周围环境光亮暗的影响,而疲劳驾驶等情况多发于司机夜间行车途中,且汽车的运动会使得相机在渐进扫描帧的不同部分时会产生相同 PPG 信号的微小相移,从而导致 rPPG 的错误估计。Magdalena 等^[24]和 Mironenko 等^[25]都开展了深入研究。

因此提出了基于红外的远程光电容积描记术等理论,建立生理特征提取模型来满足夜间疲劳驾驶检测需求,基于稀疏傅里叶的生理特征提取算法如下:

算法:红外光电容积描记术

- 1) 对输入的红外视频进行人脸 ROI 区域的分割;
- 2) 将分割好的 ROI 区域里所有像素进行平均;
- 3) 将时刻 t 作为起点,20 帧/s 的帧率来采集 1 min 的视频(共 1 200 帧)。用 x_t 序列代表连续帧中的 ROI 区域, X_t 代表初始时刻 t 输入的时间序列,即长度为 t 的时间窗口内的视频帧数;
- 4) 对 X_t 进行零均值预处理得到,构建 \tilde{X}_t 的延迟矩阵 X , 并通过奇异值分解得到 X 的特征矩阵,并分解成 M 个向量;
- 5) 将上面分解得到的 M 个向量再重构,得到一个独立的延迟矩阵 R_i , 并从 R_i 中复原输出信号 $y_i(t)$;
- 6) 该测量信号的建模如下: $p_i = h_i \times F^{-1} y_i + n_j$, F 为大小为 T 的一维离散傅里叶变换, y_i 为心跳信号的稀疏频谱, n_j 为信道噪声。
- 7) 来自所有区域 t 的信号 y_i 的无噪声频谱 x_j 具有相同的支持度,因此将鲁棒主成分分析法(RPCA)的去噪输出建模,并以矩阵形式表示如下:

$$Z = F^{-1} X + E = [F^{-1} I] \begin{bmatrix} X \\ E \end{bmatrix} \quad (4)$$

- 8) 对复原得到的输出信号进行上述操作后,得到频谱 $Y_i(k)$;

9)选择 SNR 最大的 $Y_i(k)$, 其频谱的峰值就是所求的瞬时心率 $HR_{(n)}$;

10)取 N 个心动周期, 记为 $t_{(n)}, n \in \{1, N\}$;

11)对 $t_{(n)}$ 做频谱变换, 得到频谱为 $F_{(w)}$, 功率谱密度为 $P_{(w)}$;

12)最后通过下式计算 LF 、 HF 和 LF/HF :

$$LFP = \sum P_{(w)}, 0.04 < W < 0.15 \quad (5)$$

$$HFP = \sum P_{(w)}, 0.15 < W < 0.4 \quad (6)$$

$$LF/HF = \frac{LFP}{HFP} \quad (7)$$

1.2 多模态特征融合

1) 深层信息挖掘

大量研究表明, 驾驶员面部与生理的两大因素直接影响着疲劳驾驶检测系统的性能。但是早期疲劳时面部特征变化不大导致检测困难, 而从医学角度入手, 结合生理特征检测可以改善精度低和误检等缺陷, 因此有必要建立多模态特征融合的疲劳驾驶检测理论模型。

受限玻尔兹曼机 (restricted Boltzmann machine, RBM) 具有自动学习特征间组合信息的功能, 因此本文选取 RBM 作为特征融合网络的基础。单个受限玻尔兹曼机是一个两层的神经网络, 包括了可见层 v 和隐藏层 h , 受限来源于其同层神经元间是互不关联的, 但可见层 v 和与隐藏层 h 之间则由对称的权值矩阵相互连接, 结构如图 2 所示。

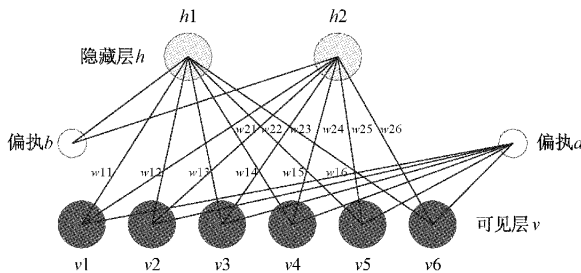


图 2 RBM 网络结构示意图

其能量公式定义如式(8):

$$E(v, h) = - \sum_{i \in v} a_i v_i - \sum_{j \in h} b_j h_j - \sum_{i, j} v_i h_j \omega_{ij} \quad (8)$$

其中, v_i 指原始特征, 作为可见层 v 的输入, h_j 是隐藏层神经元, a_i 和 b_j 分别为可见层和隐藏层的偏置, ω_{ij} 为两层神经元间连接的权值, 此时两层网络之间神经元 i, j 可能存在连接的概率可以计算:

$$p(v, h) = \frac{1}{Z} e^{-E(v, h)} \quad (9)$$

其中, Z 代表可见层和隐藏层之间所有可能存在的连接之和, 也被叫做归一化因子。

因为 RBM 在同层神经元间相互无连接的性质, 可以认为同层神经元都是相互独立的, 因此可见神经元 v_i 和隐藏神经元 h_j 出现的激活概率为:

$$p(v_i = 1 | h) = \sigma(a_i + \sum_j h_j \omega_{ij}) \quad (10)$$

$$p(h_j = 1 | v) = \sigma(b_j + \sum_i v_i \omega_{ij}) \quad (11)$$

同时用对比散度算法来优化 RBM, 这样只需要较少的抽样次数就可以对可见层的样本分布有足够好的近似估计, 加快学习速度。

考虑到面部特征和生理特征之间是高度非线性的, 如果只是简单的通过一层单独的 RBM, 只能得到单模态信号之间的连接组合, 很难学习到两者间跨模态的非线性特征。为了能够更好地提取出面部特征和生理特征之间的跨模态表达, 首先使用一层单独的 RBM 提取各模态的隐藏神经元表达。将前面提取到的特征输入到可见层 v 中, v^{Face} 表示面部特征, v^{Heart} 表示生理特征, 两种特征分别通过一层 RBM 提取中间隐单元特征, 此时第 1 层模型中学习到隐藏层后验分布概率 (即提取到的面部深层特征 h^{Face} 和生理深层特征 h^{Heart}) 为:

$$p(h_j^{Face} | v^{Face}) = \text{sigmoid} \left(\frac{1}{\sigma^2} (b_j^{Face} + \omega_j^{Face \top} v^{Face}) \right) \quad (12)$$

$$p(h_j^{Heart} | v^{Heart}) = \text{sigmoid} \left(\frac{1}{\sigma^2} (b_j^{Heart} + \omega_j^{Heart \top} v^{Heart}) \right) \quad (13)$$

将 RBM⁽¹⁾ 学习到的隐单元特征输入到 RBM⁽²⁾ 中, 使得第 2 层学习得到的隐单元能够拥有面部特征和生理特征之间更高维度的相互关系, 这种共享特征能够更有效的表达疲劳驾驶状态。

第 2 层模型中学习到隐藏层后验分布概率 (即提取到的互补特征 H) 为:

$$p(H_j^{(2)} | h_j^{Face}, h_j^{Heart}) = \text{sigmoid} \left(\frac{1}{\sigma^2} (b_j^{(2)} + \omega_j^{(2) \top} h_j^{Face, Heart}) \right) \quad (14)$$

2) 多损失重构特征融合

第 1 层 RBM 提取到的各模态隐藏表征除了输入到第 2 层 RBM 中得到模态间的互补特征外, 再分别输入到 3 个 1×1 的卷积层中, 并通过一个 ReLU 层来保证网络的非线性能力。为了简洁起见, 把这些卷积层和 ReLU 层组合表示为一个 Conv 模块。接着引入一个多模态特征聚合模块, 将这两级特征有效聚合以得到更具判别性和鲁棒性的聚合特征。

将第 1 层中提取到的各模态深层特征输入到 conv 模块中, 再输入到聚合模块得到聚合特征:

$$h_{1,2} = \text{conv}(h_j^{Face, Heart}) \quad (15)$$

$$f = \text{ReLU}[\text{conv}(\text{Concat}(h_1, h_2))] \quad (16)$$

由于共享和特定特性之间的信息重叠, 整体性能可能

会受到影响。首先,如果共享特征包含较多的特定于模态的信息,则会导致特征传递不准确。其次,如果特定特征与共享特征高度相关,那么特定特征只能对共享特征提供很少的补充。特定特征中的冗余信息也会影响模型灵敏度。

为了缓解这两个问题,将第 2 层模型中输出的互补特征 H 与聚合特征 f 各自通过损失函数^[26]进行约束,在互补模态损失 $L_2(H_i, H_k)$ 的基础上,引入了聚合特征损失 $L_1(f_i, f_k)$, 计算了交叉模态的相似性约束 L , 这样在保证共享特征和特定特征之间相关性较低的同时,增强这两个特征的互补性。

$$L = \sum \max[\rho_1 - L_1(f_i, f_k), 0] + \sum \max[\rho_2 - H_2(H_i, H_k), 0] \quad (17)$$

其中, ρ_1 和 ρ_2 是一个预定义的边界, i, k 表示与每个锚点 j 在小批次内的交叉模态间相同的指数。这种交叉模态约束确保了最难的交叉模态负样本也应该远离其对应的交叉模态正样本。

接着将两者进行点乘,得到所需的融合特征:

$$F = H \odot f \quad (18)$$

同时,最后通过重构任务对总体信息损失进行约束,结合全局损失函数^[27]引导共享和聚合特征自我区别和互补。

众所周知,各模态间的深层特征更加具有表达能力,面部深层特征有助于简单明了的检测疲劳,但存在早期疲劳容易误检且易受光线影响。而生理深层特征可以提供疲劳时人体内更细微的变化。这两种模态特征是互补的。此外,结合两种模态特征有助于从多角度来更完整、更准确的检测疲劳驾驶。

1.3 疲劳驾驶检测模型

1) 疲劳驾驶检测

驾驶员在持续不间断的行车过程中容易产生疲劳,随着时间进行,驾驶员逐渐会出现明显的疲劳特征,同时驾驶疲劳具有持续性,驾驶疲劳状态具有明显的时序性特点,其各种特征反应也具有不同的时序性信息。

LSTM 是一种特殊递归神经网络,由于其网络结构的独特性,能够很好地学习特征中的时序性信息,从而显著改善其他递归神经网络中的长期依赖(long term dependencies)问题,其单元结构如图 3 所示。

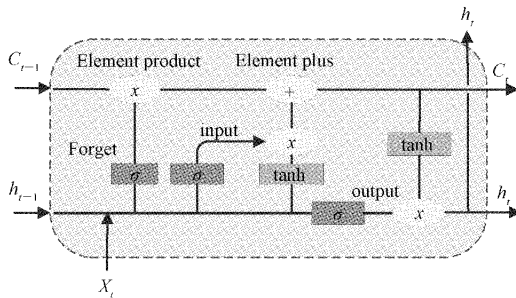


图 3 LSTM 单元结构示意图

LSTM 的原理许多学者已阐述得十分清楚,本文就不再过多赘述,简而言之,在 LSTM 中,可以通过遗忘门将无关信息遗忘,内部的当前状态会不断被更新,这样的结构能够让 LSTM 在处理序列数据的关系时更加得心应手。

而对于疲劳驾驶的预测,结合驾驶员前后状态的若干输入来共同决定会得到更为准确的结果。同时考虑到可能存在将正常眨眼误判成疲劳闭眼特征输入网络等问题,因此选取了 Bi-LSTM 网络作为检测模型,该网络包括了前向和后向的 LSTM。在隐含层中每个时刻都保存了正、反向计算的输出值,最后将双向隐含层中每个相同时刻的输出相结合得到最终结果。这样我们可以充分利用驾驶疲劳的时序性信息来提高识别的准确性。

2) 检测评估指标

本文选取准确率(accuracy)、相关系数(CORR)和均方根误差(RMSE)来评估模型性能。准确率指预测正确的占总比例的多少,相关系数能够有效的衡量疲劳驾驶检测算法预测结果和疲劳标定值之间的线性相关性,对两个序列之间的一致性进行标注;均方根误差这一指标可以对疲劳预测值与标定的疲劳状态真实值之间的误差进行具体的描述。

$$accuracy = \frac{TP + TN}{P + N} \quad (19)$$

P 指被分类为正例, N 指被分类为负例, TP 为被正确分类成 P 的个数, TN 为被正确分类成 N 的个数。

$$CORR(Y, \hat{Y}) = \frac{\sum_{i=1}^m (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^m (y_i - \bar{y})^2 \sum_{i=1}^m (\hat{y}_i - \bar{\hat{y}})^2}} \quad (20)$$

$$RMSE(Y, \hat{Y}) = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (21)$$

y_i 指标定的疲劳状态真实值, \hat{y}_i 指疲劳驾驶检测算法的预测值, y 和 \bar{y} 分别为 y_i 和 \hat{y}_i 的平均值。

2 实验与结果分析

本节中包括了实验环境、采集的车内驾驶数据集、评价指标和目前常用的模型的介绍,以及最终的实验结果分析。实验流程如图 4 所示。

2.1 实验平台、数据集

实验设备和环境: Windows11 64 位操作系统, Intel i7 2.30 GHz 处理器, 16 GB 内存, NVIDIA RTX 2060 (GPU), Python(3.7), Keras (Tensorflow2.1) 框架。

在数据采集的过程中,由于目前很少有公开的 RGB 与红外多模态疲劳驾驶数据集,因此使用了自制的疲劳驾驶数据集 FAHD (face and heart data)。为了获取驾驶数据集,使用了奥尼 S500 双目红外摄像头,可灵活安装在车内的后视镜或仪表板上,该双目摄像头由彩色 RGB 和红外 IR 传感器组成,采集速率为 20 Hz,同时捕获 RGB、IR



图4 实验流程

图像,分辨率为 640×480 ,虽然RGB和IR摄像头的位置有所偏移,但差异并不明显。

数据集包含私家车、出租车和货车等各类车辆在白天和夜间的真实驾驶场景,参与者共有56名(30名男性,26名女性),年龄在22~50岁之间,每位参与者被要求在不同的路段执行一次驾驶以及常见的行为活动,整个行车时长约在2h左右,行车中存在光照变化、驾驶员头部偏转以及部分遮挡等多种真实存在的复杂情形,保证了样本的多样性与随机性。自制数据集部分样本图像如图5所示。



图5 FAHD数据集样本图像

视频采集完成后整理切分成60s的短视频,按照KSS疲劳量表并根据时间顺序精确标记为清醒状态或疲劳状态,包含5000段 1200×650 的RGB-IR驾驶视频,其中清醒状态3720段,疲劳状态1280段。

2.2 红外rPPG-HR检测方法评估

1) 疲劳与HRV信号的关联实验

为了测试出疲劳与HRV信号的关系,验证清醒和疲劳状态的LF/HF比值变化。使用了目前较为先进且被广泛使用的心电图监测系统,让10名受试者佩戴整整一个小时的心电监测器,并分别收集了他们清醒与疲劳时的心电数据。结果如图6所示。

图6证实了Chang等^[23]关于清醒状态和疲劳状态之间HRV变化的发现。可以观察到,当受试者处于清醒或者疲劳时,LF/HF比值差异明显,前者的LF/HF比值高于后者。在整体的10个实验中,平均清醒时比平均疲劳

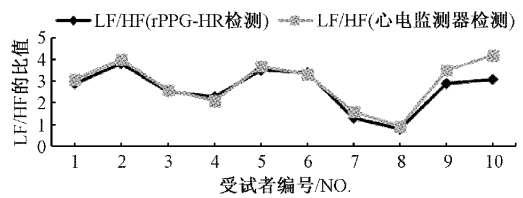


图6 清醒与疲劳时LF/HF的比值变化

时比值大约高15%~35%。当驾驶员疲劳程度加深时LF下降,HF上升,LF/HF的整体数值降低符合疲劳状态下自主神经系统的生理反应。

2) 与相关HR检测的对比实验

在这一部分中,测试了提出的红外rPPG-HR检测系统的准确性和稳定性。随机从56名受试者中选取了10名,要求他们佩戴心电监测器,其中1~5号受试者在中午12:00,6~10号受试者在晚上23:00进行实验,以便能够准确提取他们的HR和HRV,包括LF和HF数据,与本文心率提取方法的检测结果对比如图7所示。

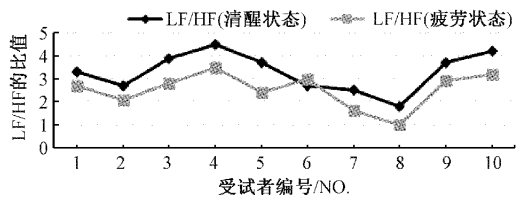


图7 不同的HRV信号提取方法对比结果

图7通过比较红外rPPG-HR方法检测结果与权威的心电检测器结果,可以看出两者在白天与夜间时基本保持一致,表明本文方法能够准确提取出有效的反映疲劳的生理特征。经过八轮实验,准确率达到了90%以上。

2.3 多特征融合方法评估

1) 不同模态融合方法的对比实验

不同的模态特征融合方法会影响到疲劳驾驶检测的结果,这里选取了目前常用的决策融合和特征融合,并将融合后的特征输入到支持向量机(support vector machines, SVM)中,通过相关系数和均方根误差两个指标来对本文提出的多模态融合算法进行对比评估。实验结果如表1所示。

表1 不同多模态融合方法的相关系数与均方根误差对比

融合方法	CORR		RMSE	
	平均值	标准差	平均值	标准差
决策融合	0.863	0.095	0.104	0.027
特征融合	0.886	0.074	0.097	0.020
本文方法	0.914	0.051	0.078	0.014

实验表明本文方法的相关系数最高,且均方根误差均小于其他两种方法,融合性能最好。原因是决策融合和特征融合算法虽然都利用了多模态信息,却也容易导致信息

冗余,且并未考虑到多模态信息之间的交互表达,因此效果一般,有所欠缺。而本文方法通过充分融合两种模态间的深层特征与互补特征,并采用 MLR 模块使得模型在训练过程中剔除冗余信息,学习到更加有效的多模态特征表达。

2) 不同模态特征输入不同算法的对比实验

为了能更有效的对多模态数据融合后的分类效果与单模态分类效果进行评判,确保实验环境和受试者不变的前提下,本研究将不同模态特征输入不同算法中进行对比实验。特征类别一共分为生理特征,视觉特征,混合视觉特征和混合多种特征 4 组,将 4 组特征类型通过给定的标签分为疲劳状态和清醒状态,同时输入到传统的 SVM 与 DBN 网络中,将得到的分类准确率与本文方法进行比较。实验结果如表 2 所示。

据表 2 可知,本文方法对于生理特征组(EGC, rPPG)的分类准确率分别为 0.874 和 0.861,可以证明 ECG 确实更有效但差别不大只有 0.013,考虑到检测的便捷性,后续生理特征采用 rPPG,对于视觉特征组(Eyes, Mouth)的分类准确率分别为 0.868 和 0.855,而混合视觉特征组(Eyes+Mouth)的准确率为 0.892,高于其他单一特征组的准确率,说明多特征能够比单一特征更有效地描述疲劳状态,对于混合多类特征组(Eyes+Mouth+rPPG)的准确率高达 0.929,据实验结果可知基于多模态数据的检测结果均优于仅输入单模态数据的方法,充分验证了多模态信

息在疲劳驾驶检测算法中的有效性,且在多种融合方法中,本文融合方法结果最优。

表 2 不同特征类型输入不同算法的准确率对比

类别	特征	指标	SVM	DBN	本文方法
生理特征	ECG	HR, HRV	0.792	0.853	0.874
	rPPG	HRV	0.778	0.827	0.861
视觉特征	Eyes	PERCLOS	0.785	0.849	0.868
	Mouth	Yawn frequency	0.757	0.806	0.855
混合视觉特征	Eyes+Mouth	PERCLOS, Yawn frequency	0.814	0.869	0.892
混合多类特征	Eyes+Mouth+rPPG	PERCLOS, Yawn frequency, HRV	0.774	0.871	0.929

2.4 疲劳驾驶检测系统实验评估

1) 疲劳驾驶检测系统的性能实验

为了更清晰直观地展现疲劳驾驶检测系统运行效果,我们在不同情况下分别截取了 1 段视频来进行训练和测试。如图 8 所示,展示了该系统在多位驾驶员头部偏转、俯仰,部分遮挡和光线变化等多种复杂环境下对于清醒状态和疲劳状态的检测和判定效果。





图8 复杂情况下疲劳驾驶检测结果展示图

图8(a)~(c)为正常情况,无遮挡且光照条件良好;图8(d)~(f)中驾驶员正在进行转头和仰头的头部动作;图8(g)~(i)为驾驶员佩戴墨镜和口罩,遮挡了部分关键检测区域;图8(j)~(l)为强光情况,来源于夜晚的灯光和白天的太阳光;图8(m)~(o)为夜晚行车时的弱光情况。可以看到该系统在正常情况和四种复杂情况下都能够较为准确的划分出感兴趣区域,且对于驾驶员的疲劳状态都能够得到准确的判断结果。

2) 不同疲劳驾驶检测方法的对比实验

为了使实验结果更加普遍和可靠,将本文算法与当前主流疲劳驾驶检测算法进行了比较,包括 Sun 等^[20]提出的多类支持向量机(MCSVM)和 Huang 等^[22]提出的多粒度深度卷积模型(RF-DCM)。实验数据为自制数据集 FAHD,其包括驾驶员在各类复杂环境中连续驾驶时的脸部视频。经过相应算法的处理后,比较了每个算法检测疲劳的准确性,实验结果如表3。

表3 疲劳驾驶检测方法在复杂环境中的准确率比较

方法	准确率/%
MCSVM ^[20]	87.3
RF-DCM ^[22]	94.6
本文方法	96.7

表3中的结果表明,在连续的真实环境中,我们算法的可靠性和准确性是最好的。主要原因是我们在设计面部特征与生理特征融合网络时,通过深层信息挖掘模块获得了更丰富的深层疲劳信息,并结合 MLR 特征融合模块有效利用了多模态特征的深层信息与互补信息,且采用红外 rPPG-HR 方法来提取夜间的生理信息,提高了检测精度的同时也增强了鲁棒性。

3 结论

为了解决传统疲劳检测方法在复杂环境下存在的精度低、鲁棒性差等关键问题,本文提出了基于多模态特征融合模块的疲劳驾驶检测模型,首先通过 MTCNN 网络和红外的 rPPG 方法来提取多种疲劳特征,提高了模型在低光和头部偏转等多种外部环境影响下的检测精度;其次利用 RBM 挖掘两种模态间的深层特征,再结合 MLR 模块加强深层特征与互补特征表达,去除冗余信息并充分融合特

征从多角度来更完整、更准确的反映疲劳驾驶;最后利用疲劳的时序性,输入到 Bi-LSTM 网络进行检测。我们在自制数据集上与几个目前常用的方法进行了充分实验与比较,表明本文模型具有良好的检测效果,并且具有较强的鲁棒性,实用性好。进一步的研究可以着眼于如何将疲劳驾驶检测和人工智能技术的深度融合,致力于推动道路智能化的进程,进而促进智慧交通革命,带动产业升级。

致谢

在这里感谢每一位同意参与数据集采集的司机的支持与鼓励。

参考文献

- [1] 李都厚,刘群,袁伟,等. 疲劳驾驶与交通事故关系[J]. 交通运输工程学报,2010,10(2):104-109.
- [2] SIKANDER G, ANWAR S. Driver fatigue detection systems: A review [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(6): 2339-2352.
- [3] 秦斌斌,彭良康,卢向明,等. 司机分心驾驶检测研究进展[J]. 计算机应用,2021,41(8):2330-2337.
- [4] NEMCOVA A, SVOZILOVA V, BUCSUHÁZY K, et al. Multimodal features for detection of driver stress and fatigue [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3214-3233.
- [5] LI Z, LI S E, LI R, et al. Online detection of driver fatigue using steering wheel angles for real driving conditions[J]. Sensors, 2017, 17(3): 495.
- [6] MCDONALD A D, SCHWARZ C, LEE J D, et al. Real-time detection of drowsiness related lane departures using steering wheel angle[C]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Los Angeles, USA: SAGE Press, 2012, 56(1): 2201-2205.
- [7] 张亚龙,张世武,孙帅帅,等. 融合肌电信号与 A 型超声的新型肌肉疲劳检测方法[J]. 电子测量与仪器学报,2022,36(6):13-21.
- [8] CHAUDHURI A, ROUTHAY A. Driver fatigue detection through chaotic entropy analysis of cortical

- sources obtained from scalp EEG signals[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(1): 185-198.
- [9] DU G, WANG Z, LI C, et al. A TSK-type convolutional recurrent fuzzy network for predicting driving fatigue [J]. IEEE Transactions on Fuzzy Systems, 2020, 29(8): 2100-2111.
- [10] ZHENG W L, GAO K, LI G, et al. Vigilance estimation using a wearable EOG device in real driving environment [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(1): 170-184.
- [11] 李昭慧,张玮良. 基于改进 YOLOv4 算法的疲劳驾驶检测[J]. 电子测量技术, 2021, 44(13): 73-78.
- [12] 李响,李国正,邓明君,等. 基于语音频谱图像特征的人体疲劳检测方法[J]. 仪器仪表学报, 2021, 42(2): 123-132.
- [13] 禹江林,张云. 一种基于多因素融合的驾驶预警方法[J]. 电子测量技术, 2021, 44(11): 103-108.
- [14] MANDAL B, LI L, WANG G S, et al. Towards detection of bus driver fatigue based on robust visual analysis of eye state [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 18(3): 545-557.
- [15] YANG H, LIU L, MIN W, et al. Driver yawning detection based on subtle facial action recognition[J]. IEEE Transactions on Multimedia, 2020, 23: 572-583.
- [16] REDDY B, KIM Y H, YUN S, et al. Real-time driver drowsiness detection for embedded system using model compression of deep neural networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA: CVPR Press, 2017: 121-128.
- [17] 许智熠,狄长安,潘绪仲,等. 穿戴式火炮人机工效测试评估方法研究[J]. 国外电子测量技术, 2022, 41(1): 106-112.
- [18] YU J, PARK S, LEE S, et al. Driver drowsiness detection using condition-adaptive representation learning framework [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(11): 4206-4218.
- [19] QIU X, TIAN F, SHI Q, et al. Designing and application of wearable fatigue detection system based on multimodal physiological signals[C]. Proceedings of IEEE International Conference on Bioinformatics and Biomedicine, Seoul, Korea(South): BIBM Press, 2020: 716-722.
- [20] SUN W, ZHANG X, PEETA S, et al. A real-time fatigue driving recognition method incorporating contextual features and two fusion levels[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(12): 3408-3420.
- [21] DU G, LI T, LI C, et al. Vision-based fatigue driving recognition method integrating heart rate and facial features [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(5): 3089-3100.
- [22] HUANG R, WANG Y, LI Z, et al. RF-DCM: Multi-granularity deep convolutional model based on feature recalibration and fusion for driver fatigue detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(1): 630-640.
- [23] CHANG Y L, FENG Y C, CHEN O T C. Real-time physiological and facial monitoring for safe driving[C]. Proceedings of the IEEE Engineering in Medicine and Biology Society, Orlando, FL, USA: EMBC Press, 2016: 4849-4852.
- [24] MAGDALENA NOWARA E, MARKS T K, et al. SparsePPG: Towards driver monitoring using camera-based vital signs estimation in near-infrared [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, Utah: CVPR Press, 2018: 1272-1281.
- [25] MIRONENKO Y, KALININ K, KOPELIOVICH M, et al. Remote photoplethysmography: Rarely considered factors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, WA, USA: CVPRW Press, 2020: 296-297.
- [26] YE M, WANG Z, LAN X, et al. Visible thermal person re-identification via dual-constrained top-ranking [C]. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: IJCAI-18 Press, 2018: 1092-1099.
- [27] KUMAR B V, CARNEIRO G, REID I. Learning local image descriptors with deep siamese and triplet convolutional networks by minimizing global loss functions [C]. Proceedings of the Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA: CVPR Press, 2016: 5385-5394.

作者简介

高绮煌, 硕士研究生, 主要研究方向为深度学习、图像处理、人工智能。

E-mail: 2878834552@qq.com

谢凯(通信作者), 博士, 教授, 主要研究方向为人工智能、图像处理、信号处理、数据挖掘与三维建模等。

E-mail: pami2009@163.com

贺正方, 硕士, 中级工程师, 主要研究方向为物联网、数字图像处理和机器视觉等。

E-mail: hezhengfang1972@163.com

刘畅, 硕士, 讲师, 主要研究方向为人工智能、地震信号处理等。

E-mail: wenchang2016paper@163.com

贺建飏, 硕士, 副教授, 主要研究方向为信号与信息处

理等。

E-mail: hejianbiaopaper@163.com

张伟, 讲师, 博士, 主要研究方向为图像处理和计算机视觉等。

E-mail: zhangwei2022paper@163.com