小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems
DOI:10.20009/j.cnki.21-1106/TP.2024-0621
2026 年 2 月 第 2 期
Vol.47 No.2 2026

时序向量引导的多特征融合驾驶员疲劳检测方法

裴颂文[1],谢 影[1],张慧辰[2]

[1](上海理工大学 光电信息与计算机工程学院,上海 200093)
[2](公安部交通管理科学研究所 道路交通安全公安部重点实验室,江苏 无锡 214151)
E-mail:swpei@usst.edu.cn

摘 要: 疲劳驾驶检测是降低交通事故发生率,提高驾驶安全的关键技术.现有技术主要依赖提取驾驶员面部特征判断疲劳状态.由于光照条件、拍摄角度等复杂的交通场景带来的各类干扰信息,使得驾驶疲劳状态难以准确检测.因此,本文提出了一种时序向量引导的多特征融合驾驶员疲劳视频检测方法.首先,通过局部图像及节点图特征提取模块分别提取对应分支的疲劳特征块,减少低质图像区域对整体检测的负面影响;其次,通过关键点间的变化向量构建时序向量邻接矩阵,得到包含疲劳状态信息的时序向量;最后,通过 Transformer 多特征融合模块,使用时序向量引导双分支增强对疲劳状态动态变化的检测,提高疲劳检测的精度.在公共数据集 YAWDD 和 NTHU-DDD 的实验结果表明了本文方法的准确率分别至少提高了 2.8% 和 3.2%.

关 键 词: 疲劳驾驶检测;时序向量;Transformer 融合;交叉注意力
中图分类号: TP391 **文献标识码:** A **文 章 编 号:** 1000-1220(2026)02-0413-08

Fatigue Detection Method for Multi-feature Fusion Guided by Temporal Vectors

PEI Songwen[1], XIE Ying[1], ZHANG Huichen[2]

[1](School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)
[2](Key Laboratory of Ministry of Public Security for Road Trafic Safety, Trafic Management Research Institute of the Ministry of Public Security, Wuxi 214151, China)

Abstract: Fatigue driving detection is vital for reducing traffic accidents and improving road safety. Existing methods primarily rely on facial feature extraction, but complex traffic conditions, such as lighting and angles, hinder accurate detection. Therefore, this paper proposes a Temporal Vector-Guided Multi-Feature Fusion Method for Driver Fatigue Detection in Video Streams. First, the method extracts fatigue feature patches via local image and graph feature modules to mitigate the impact of low-quality regions; Second, a temporal vector adjacency matrix is then constructed using landmark variation vectors, encoding fatigue state information; Finally, a Transformer-based fusion module uses the temporal vector to guide a dual-branch model, enhancing dynamic fatigue detection. The experimental results on the public datasets YAWDD and NTHU-DDD demonstrate that the accuracy of the proposed method is improved by 2.8% and 3.2%, respectively.

Keywords: driver fatigue detection; temporal vector; Transformer fusion; cross-attention

0 引 言

疲劳驾驶是道路安全中最常见的危险因素之一.研究表明,驾驶员疲劳极易导致交通事故,驾驶员因其困倦发生车祸的损伤程度是清醒情况下的 3.3 倍[1].疲劳驾驶检测系统能够对驾驶员进行实时状态监测,判断驾驶员是否存在疲劳状态并及时提醒,从而最大程度上避免发生疲劳驾驶事故.随着人工智能技术的发展,疲劳检测主要分为基于生理信号检测[2]、基于车辆行为检测[3]、基于视觉特征检测[4].目前,基于视觉的检测方法有着广泛的研究,该类方法主要依靠检测人脸视频图像中是否存在闭眼,打哈欠等动作来判定疲劳.然而,因驾驶环境的不同,图像中的疲劳特征极易受光照、人脸角度、监测距离等因素的影响从而导致检测的准确度低.

基于视觉特征检测的方法主要是通过计算机视觉技术来处理驾驶员的相关图像、视频.根据驾驶员的面部特征信息、头部姿势等信息判断驾驶员的疲劳状态.由于驾驶场景复杂,在视频中检测驾驶员的疲劳状态仍然具有挑战性.驾驶室内的照明变化、驾驶员动作遮挡、车辆颠簸等都会干扰输入图像的质量,导致无法准确检测出驾驶员的疲劳状态.为了处理有干扰的视频数据,Huang 等人[5]提出了一种用于驾驶员疲劳检测的自监督多粒度图注意力网络.该网络通过图像恢复的自监督学习方法,解决了干扰图像鲁棒性弱、忽略关键帧信息的问题.然而,忽略关键帧的方法可以处理遮挡问题,但是会破坏视频中动作的连续性,容易丢失其他有效信息.Sun 等人[6]提出了一个全局人脸特征流来强调全局面部流中的重要特征,并构建了两个局部眼睛特征流用于处理关键局部特

收稿日期:2024-12-31 收修改稿日期:2025-02-12 基金项目:国家自然科学基金项目(61975124)资助;道路交通安全公安部重点实验室开放课题基金项目(2023ZDSYSKFKT04)资助;四川省重点研发项目(2024YFFK0443)资助;上海市自然科学基金项目(25ZR1401266)资助. 作者简介:**裴颂文**,男,1981 年生,博士,教授,CCF 杰出会员,研究方向为智能计算、计算机系统结构;**谢 影**,女,1999 年生,硕士研究生,研究方向为智能计算、驾驶疲劳检测;**张慧辰**,男,1988 年生,硕士,助理研究员,研究方向为智能交通.

征信息,有效解决了遮挡、头部转动等导致眼睛状态不一致的低质量图像数据.但是,该方法仅依靠眼睛区域,未采用嘴部重要区域数据,无法进一步提高检测精度.Bai 等人[7]在空间域图卷积的基础上加入了时域图卷积,其中空间图卷积负责捕获当前帧的驾驶员状态信息,而时域图卷积则检测连续帧的驾驶员疲劳特征.最后,将两个域的 softmax 值相加,得到融合后的结果以判断驾驶员是否存在疲劳状态.上述方法仅依赖驾驶员面部的关键点信息,可以解决图像遮挡条件下的检测问题,但是高度依赖人脸图像的质量,因此真实场景检测的精度受驾驶场景的影响较大.文献[8]构建了一种新的基于深度学习的驾驶员疲劳检测算法,利用反向残差方法设计了单次多边框检测的附加层网络结构,实现了对驾驶员面部的实时检测.网络模型的尺寸较小,便于应用到车载嵌入式设备中,但该方法检测的关键点少,需进一步优化后才能适应真实的复杂驾驶场景.

因此,本文提出了一种时序向量引导的多特征融合模型进行疲劳驾驶的检测,主要贡献如下:

1)针对驾驶员面部图像中左眼、右眼和嘴巴形态的变化,分别构建了局部图像特征和节点图特征提取模块,以提取局部图像数据和节点图数据中的疲劳特征.

2)根据相邻帧中对应关键点之间的位置变化向量构建时序向量邻接矩阵,引导模型对疲劳动作连续性的关注,使模型具有更强的时间表征能力.

3)提出了一个基于 Transformer 多特征融合模块,引入多头注意力机制强化图像和节点特征,以及利用时序向量通过交叉注意力机制引导特征信息融合.

1 相关工作

1.1 基于生理信号检测

第一类检测方法是依靠生理特征参数,通常包括脑电信号、心电信号、肌电信号.脑电图测量的是大脑皮层外层的电活动,在检测大脑活动方面表现出优异的准确性和灵敏度,使其成为确定疲劳程度的首选方法[9].为了识别驾驶疲劳,Hu 等人[2]提出了一种基于多通道脑电图信号时空结构的卷积神经网络(Convolutional Neural Network,CNN).张等人[10]在自监督学习与扩散模型结合的基础上,提出了 SSL-DDPM 的疲劳状态检测方法来提取脑电信号对驾驶员疲劳或警觉进行分类.X. Ding 等[11]提出了一种基于深度学习的 ResNet3D 模型,利用 3 个前额叶脑电信号通道进行驾驶疲劳检测.K. Fujiwara 等人[12]利用驾驶员的心电图信号结合自编码器开发了驾驶员疲劳检测系统来检测疲劳引起的异常变化的肾脏动脉阻力指数(Renal Artery Resistance Index,RRI)数据.基于生理参数的系统具有较高的检测准确率,但是考虑到生理信号采集设备十分复杂,其实用性、方便性和舒适性不利于驾驶疲劳检测.研究[13]采用了基于小波散射网络的驾驶疲劳分类方法,提取脑电信号的小波散射系数,作为特征向量输入到支持向量机中进行分类.

1.2 基于车辆行为检测

第二类检测方法是建立在车辆状态检测的基础上,通过对车辆信息、行驶轨迹和方向盘位置的估计来获取驾驶员的状态.研究[3]中引入了一种新的 2D 智能驾驶员模型,该模型包含了实际的车辆轨迹,重点关注车辆的横向运动,旨在提供一个清晰的实际车辆轨迹特征来评估驾驶员疲劳状态.Yang 等人[14]基于车辆轨迹数据提出了一种新型的驾驶行为安全检测方法.在分析驾驶行为内在特征的基础上,通过驾驶行为速度及车辆启动和停止状态确定危险驾驶行为特征参数.蔡素贤等人[15]采用控制器局域网(Controller Area Network,CAN)总线采集的车辆运行状态数据,提取了驾驶行为相关的特征,采用随机森林算法对疲劳驾驶进行识别,其整体的识别准确率达到 78.5%,该方法具有良好的实用性,便于嵌入到实际应用设备中.然而,不同的道路条件、驾驶技术和车辆特性对检测精度还有待提升,这是基于车辆行为检测的一大重要挑战.

1.3 基于视觉特征检测

目前基于单帧图像研究的驾驶员疲劳检测方法在正常情况下具有良好的性能,但疲劳是一种连续状态,在单帧图像条件下的疲劳检测容易忽视了驾驶员在时间特征上的疲劳信息,因此基于视频的驾驶员疲劳检测方法成为该领域的热点之一.在研究[16]中,提出了一种多级疲劳检测系统,通过 Haar 级联分类器对视频进行特征提取,设计了 2D 和 3D CNN 模型来同时检测驾驶员的打哈欠和闭眼.S. Fa 等人[17]提出了一种轻量级的多尺度时空注意图卷积网络,对驾驶员面部变化的时空特征进行提取,使其能在实际驾驶场景中完成实时疲劳检测任务.X. Lv 等人[18]首先对输入视频进行采样增强,然后提出时空自适应算法学习视频帧序列中各帧的疲劳时空特征,并自适应融合各帧的疲劳分类分数得到疲劳分类结果.研究[19]中,利用 YOLOv3 和长短期记忆网络(Long Short-Term Memory,LSTM)构建时间特征模型检索时间特征,采用 CNN 和 LSTM 构建时空特征模型提取空间特征.Yang 等人[20]提出了一种基于关键面部特征的视频驾驶员困倦检测方法,充分利用面部与疲劳相关的时空特征,并基于 transformer[21]引入多头注意块强化对时空特征的融合.

2 TV-MFFN 模型

本文提出时序向量引导的多特征融合网络模型(Temporal Vector-Guided Multi-Feature Fusion Networks,TV-MFFN),其包含以下 4 个模块:局部图像特征提取模块、节点图特征提取模块、时序向量构建模块、Transformer 多特征融合模块.模型整体检测框架如图 1 所示.首先,输入视频逐帧通过 Dlib 检测算法[22]进行人脸关键点检测,得到每帧图像中的驾驶员双眼与嘴巴关键点位置信息.然后,根据这些关键点位置确定包含双眼和嘴巴的矩形区域,从而实现对源视频的定位和划分.同时,设计了局部图像特征提取模块,对通过双眼和嘴巴关键点定位获得的局部区域图像进行疲劳特征提取.其次,关键点信息通过坐标矩阵及邻接矩阵构建图节点信息,并设计了节点图特征提取模块对其进行特征提取.在时序向量构建模块中,依据关键点在每帧图像间的位置变化,构建相应的时序向量邻接矩阵,并将其划分为对应的时序向量 patch.在 Transformer 多特征融合模块中,利用时序向量 patch 引导局

部图像及节点图特征模块强化对疲劳动作变化的关注. 最终通过多层感知机(Multi-Layer Perceptron,MLP)回归输出最终疲劳概率结果,结果越接近 1 时表示疲劳,越接近 0 时表示正常.

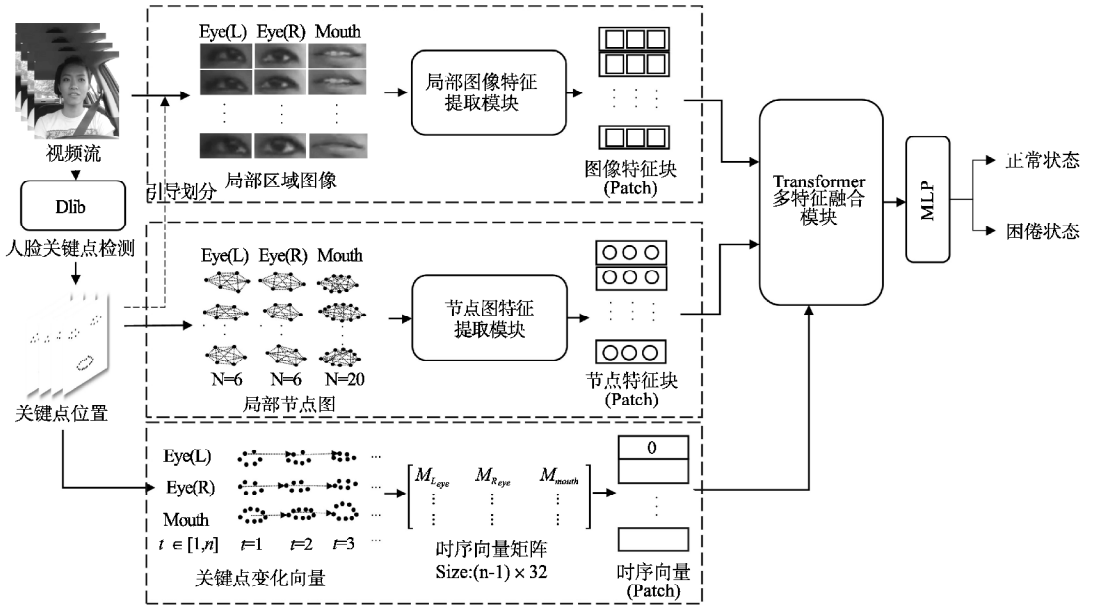

图 1　TV-MFFN 模型架构图

Fig. 1　Architecture of TV-MFFN model

2.1　局部图像特征提取模块

本文构建了局部图像特征提取模块用以充分提取主要面部图像区域(左眼、右眼,嘴巴)的疲劳特征表示. 如图 2(a)所示. 每帧的 3 个区域图像(左眼,右眼,嘴巴)分别通过该模块提取特征,其中左眼和右眼共享模型权重参数. 数据输入模型前全部进行随机数据增强操作,包括对比度调整,亮度调整以及添加随机噪声等,用以模拟实际驾驶环境中可能出现图像质量不佳的情况. 该模块主要由 3 个卷积层,3 个残差快(ResBlock)以及一个空间金字塔池化组成,其中 3 个卷积层的通道数分别为 32,64,128. ResBlock 结构如图 2(b)所示,输入首先通过一个 1×1 的卷积进行降维,通道数 C 减少一半,以减少计算参数量,再通过一个 3×3 的卷积提取特征,随后再通过 1×1 的卷积将通道数还原,最后原输入相加后形成输出. 同时接入一个 Dropout 层对特征向量部分元素随机清零,提高模型鲁棒性. 考虑局部区域图像截取因不同角度、不同位置等原因从而尺寸大小不一,本文引入空间金字塔池化对其进行统一大小. 主要包括 3 个最大池化层(1×1,2×2,4×4),拼接 3 个池化层的输出形成一个一维特征向量. 最后,同样拼接每帧图像中左眼、右眼以及嘴部的 3 个一维特征向量形成一个图像特征块(Patch).

为提高前者模型特征提取的精确性,以及对图像质量的抗干扰性,参考增强自动编码器(Adaptive Arithmetic Encoder,AAE)的核心思想[23],该模型添加一个解码器模块(如图 2(c)所示). 将每个区域的一维向量视为潜在空间向量,通过四层卷积结构从潜在空间中重建出原始图像,并通过重建损失进行约束,以增强前者模块最大程度地学习到区域特征信息,同时忽略光照等噪声影响. 其中图像重建损失采用均方误差(Mean Squared Error,MSE),如公式(1)所示:

$$L_{Recon}=\frac{1}{N}\sum_{i=1}^{i=N}(y_i-\hat{y}_i)^2 \tag{1}$$

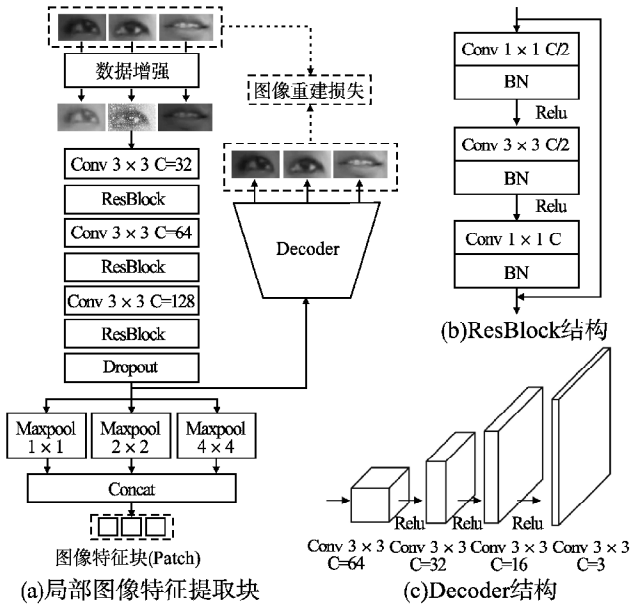

图 2　局部图像特征提取模块

Fig. 2　Local image feature extraction block

其中,N 表示图像中像素数量,y_i 表示原始图像中第 i 个像素,$\hat{y}_i$ 表示重建图像中第 i 个像素.

2.2　节点图特征提取模块

本文将 Dlib 检测到的关键点信息全部转化为图节点信息来构建左眼、右眼及嘴部区域节点图. 本文将区域关键点在

图像上的横纵位置坐标$N_i=(x_i,y_i)$作为每个节点的输入信息. 考虑到 Dlib 检测过程中,关键点位置会因为人脸角度或光照原因产生检测误差,本文对每张图中 1/3 数量的节点,在位置(x_i,y_i)进行 ±5 个像素的偏移,以提高模型鲁棒性.

眼睛区域节点数据V_{eye}用其 6 个关键点位置坐标所搭建的 2×6 的坐标矩阵表示$V_{\text{eye}}=(N_1,N_2,\ldots,N_6)$,嘴巴区域节点$V_{\text{mouth}}$用其 20 个节点位置坐标所搭建的 2×20 的坐标矩阵表示$V_{\text{mouth}}=(N_1,N_2,\ldots,N_{20})$(如图 3 左侧所示).

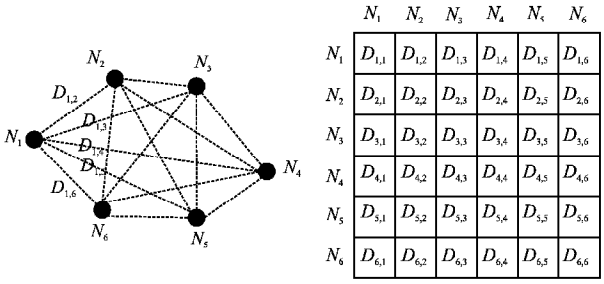

图 3 构建局部节点图

Fig. 3 Build local node graph

为捕捉各区域图节点的整体结构和特征,本文通过节点坐标计算各节点之间欧氏距离$D_{i,j}$来搭建对应的邻接矩阵 A(如图 3 右侧所示). 其中每个图为完全无向图,每个节点之间共用相同的关系距离(即$D_{i,j}=D_{j,i}$),且$D_{i,i}=0$.

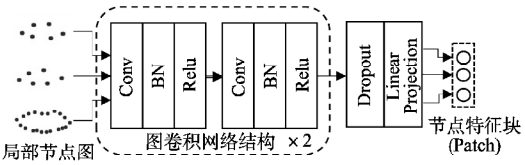

图 4 节点图特征提取模块

Fig. 4 Node graph feature extraction block

节点图特征提取块如图 4 所示,主要结构由两个图卷积神经网络[24]组成,其中包括两个卷积层,两个 BN 层以及两个 Relu 层,卷积核大小为 1×1. 输出同样经过一个 Dropout 层,并借鉴 VIT 输入的过程,对每个输出的节点特征进行一个线性的投影转换成一维向量. 最后拼接 3 个一维向量为一个节点特征块 Patch. 训练过程中,左眼和右眼仍共享模型权重参数.

2.3 时序向量构建模块

图像特征与节点特征主要依赖单帧图像中双眼与嘴巴的状态来表示疲劳,如闭眼,打哈欠等. 正常人的眨眼或者张嘴说话,都会与上述状态有类似之处,所以仅从瞬时的状态无法准确判断疲劳与否. 在视频数据中,驾驶员疲劳状态会在阶段时间的动作中正确体现. 如打哈欠,驾驶员会从初始的嘴部闭合到持续张开一段时间,最后再闭合,从而形成一个动作闭环. 因此视频中的疲劳特征除了在单帧的空间表示之外,其每帧之间的动作变化即时序变化也是重要参考特征.

如图 5 所示,本文同样以每帧图像的关键点位置坐标构建时序动作表示,首先以 2.2 节构建节点图的方式,以$N_{t,i}=(x_{t,i},y_{t,i})$的方式表示双眼或嘴部区域在第 t 帧的第 i 个关键点的坐标. 然后以$\overrightarrow{N_{t-1,i}N_{t,i}}=(x_{t,i}-x_{t-1,i},y_{t,i}-y_{t-1,i})$的方式构建双眼与嘴部在连续的两个视频帧中对应关键点的向量数据,最终构建的时序向量邻接矩阵如公式(2)所示:

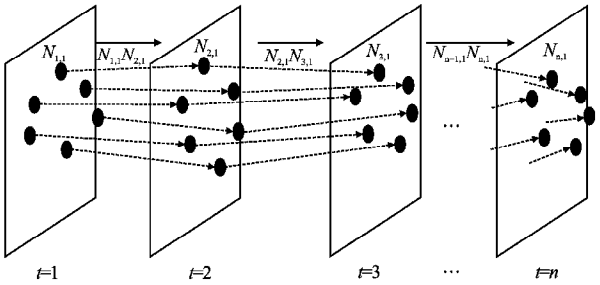

图 5 时序向量构建

Fig. 5 Build temporal vector

$$M=\begin{bmatrix}\overrightarrow{N_{1,1}N_{2,1}} & \overrightarrow{N_{1,2}N_{2,2}} & \cdots & \overrightarrow{N_{1,k}N_{2,k}}\\ \overrightarrow{N_{2,6}N_{3,6}} & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \overrightarrow{N_{n-1,1}N_{n,1}} & \cdots & & \overrightarrow{N_{n-1,k}N_{n,k}}\end{bmatrix} \tag{2}$$

其中,M 表示构建的时序向量邻接矩阵,n 表示关键点所在帧数,k 表示每个区域对应的关键点数量. 因此左眼和右眼时序向量邻接矩阵M_{Leye},M_{Reye}中的 $k=6$,嘴巴区域时序向量邻接矩阵M_{mouth}中 $k=20$,则输入后续模型的总时序向量邻接矩阵如公式(3)所示:

$$M_{\text{all}}=\text{Concat}(M_{\text{Leye}},M_{\text{Reye}},M_{\text{mouth}}) \tag{3}$$

其中,Concat( )将 3 个矩阵在列的维度进行拼接,总时序向量邻接矩阵M_{all}大小为$(n-1)\times 32$. 为了与空间帧数保持一致,本文对M_{all}最上面增补一行 0 数据,则最终输入矩阵为大小 $n\times 32$,最后将矩阵的每一行时序向量划分成为一个 Patch 块.

2.4 Transformer 多特征融合模块

本文设计了一个基于 Transformer 多特征融合模块,如图 6 所示. 首先,在上述方法中所得到的图像特征块 PatchP_{I} 与节点特征块 PatchP_{N} 依据 Transformer 流程[25]分别添加一个 ClassToken,便于后续分类回归结果. 其次,并为两者进行位置编码的一维位置嵌入,位置编码采用固定的每个 Patch 在视频中的帧数位置来编码表示. 考虑到两个输入每个对应 Patch 的帧数位置相同,因此采用相同编码,如公式(4)和公式(5)所示:

$$\text{PE}(\text{frame},2i)=\sin\left(\frac{\text{frame}}{10000^{(2i/d)}}\right) \tag{4}$$

$$\text{PE}(\text{frame},2i+1)=\cos\left(\frac{\text{frame}}{10000^{(2i/d)}}\right) \tag{5}$$

PE 为嵌入编码值,其中偶数位置(2i)用正弦,奇数位置(2i+1)用余弦. "frame" 表示每个 Patch 所在帧数位置,d 表示嵌入维度.

与其他现有的模型相比,多头注意力(Multi-Head Attention,MHA)[26]在序列特征提取方面具有明显的优势,同时具有较强并行计算能力. 因此分别采用 4 个 Transformer 模型中

的 Encoder 模块进一步提取序列特征,得到 $P_I^{'}$ 与 $P_N^{'}$. 每个 Encoder 包括一个多头注意力模块,一个多层感知机(Multi-Layer Perceptron,MLP)以及两个层归一化(Layer Normalization,LN).

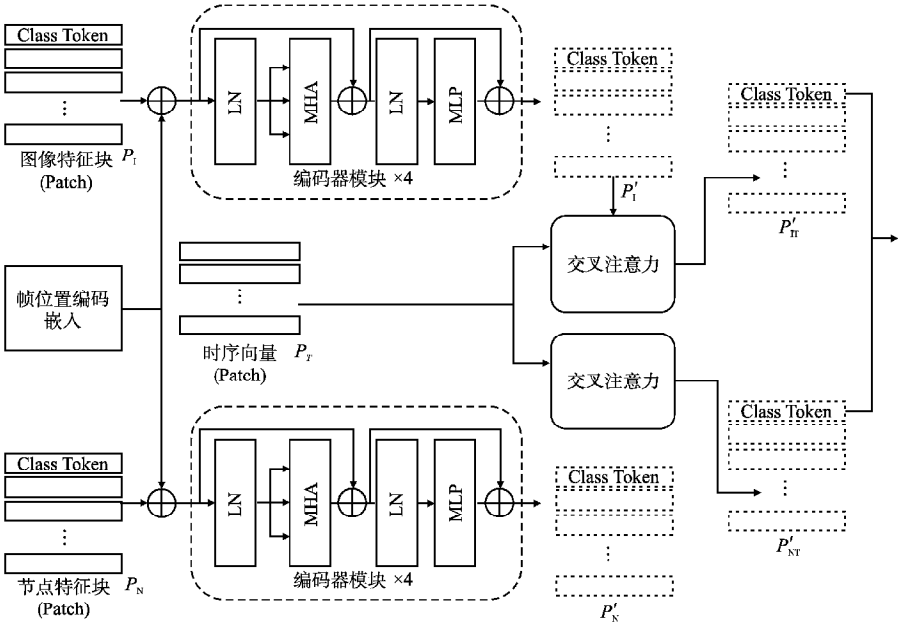

图 6　Transformer 多特征融合模块

Fig. 6　Transformer multi-feature fusion block

随后,将 $P_I^{'}$,$P_N^{'}$ 分别与时序向量 PatchP_T 中的时序特征通过交叉注意力[27]捕获两者相关性并进行信息融合,通过加强对疲劳动作的关注. 首先从 $P_I^{'}$,$P_N^{'}$ 与 P_T 中分别生成查询(Query)、键(Key)和值(Value)矩阵,如公式(6)和公式(7)所示:

$$Q_I = P_I^{'} W_{Q_I}, Q_N = P_N^{'} W_{Q_N} \tag{6}$$

$$K_T = P_T W_K, V = P_T W_V \tag{7}$$

$$I_{\text{Attention}} = softmax(\frac{Q_I K^T}{\sqrt{d}}) V \tag{8}$$

$$N_{\text{Attention}} = softmax(\frac{Q_N K^T}{\sqrt{d}}) V \tag{9}$$

其中,W_{Q_I},W_{Q_N},W_K,W_V为可训练学习的线性变换矩阵. 使用查询矩阵和键矩阵的点积来计算注意力权重$I_{\text{Attention}}$以及$N_{\text{Attention}}$,如公式(8)和公式(9)所示,d 为查询以及键的维度.

$$P_{\text{IT}}^{'} = \alpha P_I^{'} + (1-\alpha) I_{\text{Attention}} \tag{10}$$

$$P_{\text{NT}}^{'} = \beta P_N^{'} + (1-\beta) N_{\text{Attention}} \tag{11}$$

使用加权求和分别将图像特征 Patch 与节点特征 Patch 与对应的注意力权重进行融合输出,如公式(10)和公式(11)所示,其中 α,β 是一个可学习的融合权重,取值范围为[0,1].

最后,本文提取了$P_{\text{IT}}^{'}$与$P_{\text{NT}}^{'}$中的 ClassToken,并将其拼接后输出,作为最后分类回归的依据.

3　实验与结果分析

3.1　数据集准备

数据集采用了两种公开的驾驶疲劳数据集:YAWDD[28]以及 NTHU-DDD 数据集[5]. YAWDD 数据集中包含了 350 个视频对象,用于检测正常驾驶、开车时说话或打哈欠等场景. 其包括 57 名男性和 50 名女性,年龄和面部特征各不相同. NTHU-DDD 数据集由 22 个受试者模拟不同场景驾驶的视频组成,该数据集包含白天和夜晚两种场景下的正常驾驶和疲劳驾驶视频,每段视频平均 60 秒左右,每秒取 15 帧夜晚或 30 帧白天. 该数据集包含了戴眼镜、不戴眼镜、打哈欠、低头等各种动作. 这些视频是模拟真实的驾驶环境下录制,含有疲劳和清醒两种状态,共 90 个视频.

为进一步增加数据集数量,本文在从每个视频抽取并保留 5 个随机时间起点开始的 512 帧的视频. 因此数据集视频数量共 2200 个,将其按照 8∶1∶1 的方式划分数据集,其中 1760 个作为训练集,220 个作为验证集,220 个作为测试集.

3.2　实验环境与评价指标

本文方法在 Linux 操作系统上实现,GPU 为 NVDIA GeForce RTX 3060,CUDA 版本为 11.7,采用 Pytorch 深度学习框架. BatchSize 大小为 1,模型总训练 epoch 为 100,采用 Adam 优化器,初始学习率为 0.01,在 50 个 epoch 后,学习率按照线性衰减至 0.

在本研究中,为了评估所提方法的性能,采用分类任务中的 4 个评价指标. 包括准确率(Accuracy)、精确率(Precision)、召回率(Recall)及 F1 值(F1score). 本文将驾驶员疲劳状态样本视为正样本、正常状态样本视为负样本. 准确率表示正确预测结果的总体比率,精确率表示正确预测的正样本与所有预测的正样本的比率. 召回率表示正确预测的正样本在所有正样本中的比例. F1 值是通过平衡精度和召回率来评估模型整体性能.

3.3　模型性能对比实验

将本文方法与皮尔逊指数[29]、Yolov8[30]、2s-STGCN[7]、

VBFLLFA[20]、SMGA-Net[5]、FF-CNN[6]、JHPFA-Net[31]等主流的疲劳驾驶视频检测方法在 YAWDD 和 NTHU-DDD 数据集上进行实验,并在模型分类的性能指标上进行了对比.

表 1 在 YAWDD 数据集上的对比实验结果

Table 1 Comparison results on YAWDD dataset

方法	准确率(%)	精确率(%)	召回率(%)	F1 值
皮尔逊指数	72.6	71.8	73.1	0.715
Yolov8	70.4	69.5	70.1	0.716
2s-STGCN	92.3	92.5	94.0	0.895
VBFLLFA	91.6	90.2	91.8	0.909
SMGA-Net	83.6	83.0	84.0	0.821
FF-CNN	86.2	85.6	85.3	0.873
JHPFA-Net	86.7	91.0	87.5	0.888
TV-MFFN	**95.1**	**94.5**	**95.5**	**0.948**

表 2 在 NTHU-DDD 数据集上的对比实验结果

Table 2 Comparison results on the NTHU-DDD dataset

方法	准确率(%)	精确率(%)	召回率(%)	F1 值
皮尔逊指数	68.3	70.6	69.2	0.687
Yolov8	71.6	70.2	70.8	0.725
2s-STGCN	91.0	90.1	90.9	0.911
VBFLLFA	90.1	89.3	90.6	0.895
SMGA-Net	78.4	77.1	78.0	0.798
FF-CNN	82.8	83.9	83.4	0.823
JHPFA-Net	81.4	82.2	80.6	0.815
TV-MFFN	**94.2**	**93.1**	**93.4**	**0.927**

如表 1 和表 2 所示,首先传统的基于皮尔逊指数的检测方法主要依赖计算眼睛及嘴巴关键点的纵横比的变化来判断疲劳状况.该方法极其依赖关键点的准确性,且无法适应部分人群的五官比例,泛化性较低,在两个数据集上的平均准确率仅 70.4%.基于 Yolov8 的方法主要依赖 Yolo 目标检测模型,根据驾驶图像检测驾驶员面部关键部位比如眼睛、嘴巴、头部等,通过闭眼,打哈欠次数确定疲劳状.YOLOv8 对输入图像质量要求较高,光线不足或过强,会导致检测精度下降,因此准确率同样仅有 71.0%,鲁棒性较差.2s-STGCN 方法同样也考虑到时序信息与空间信息,检测精度接近本文模型.由于此方法仅仅依靠关键点的分析来进行疲劳检测,因此在两个数据集上实验的 F1 值都低于 TV-MFFN 模型,表现出的总体性能相对较差.VBFLLFA 方法将图像分为眼睛与嘴巴区域分别提取特征并综合后分类出结果,该方法一定程度能避免部分区域遮挡下对其他区域的影响,因此在两个数据集上的准确率结果高于 90.1%.SMGA-Net 与 JHPFA-Net 均以整张图像作为网络输入来提取疲劳特征,因此会受到较多冗余信息的干扰.FF-CNN 方法主要将图像中人眼作为主要参考信息,忽略了嘴巴部位的重要信息,导致最终的检测精度远远低于 TV-MFFN 方法.

在 NTHU-DDD 数据集中,其视频数据均是是在光线交叉的环境中录制,分辨率较低且为黑白图像.因此相较于在 YAWDD 数据集上的对比实验结果,SMGA-Net 与 JHPFA-Net 以整张图像输入,因此精确率对比下降较多,2s-STGCN,VBFLLFA 以及 FF-CNN 模型上具有一定的鲁棒性,因此下降幅度较小.TV-MFFN 方法虽然同样有所降低,但仍在可接受范围,同时优于其他对比方法,证明了 TV-MFFN 方法的有效性.

3.4 在复杂驾驶条件数据下的对比结果

为了验证 TV-MFFN 模型在包含驾驶室光照变化、驾驶员面部角度变化、驾驶员面部遮挡、视频低分辨率等复杂驾驶条件数据(如图 7 所示)下的鲁棒性.本文挑选了 YAWDD 和 NTHU-DDD 数据集中符合上述条件的 500 个 512 帧的视频,并进行随机数据增强(亮度对比度调整,添加图像噪声).在这批数据集上,同样将 TV-MFFN 与其他对比方法在准确率与召回率指标上进行对比.

图 7 包含复杂驾驶条件的数据集示例

Fig. 7 Dataset examples containing complex driving conditions

如表 3 所示,2s-STGCN 方法仅考虑关键点表示的特征.在驾驶室光照不足,驾驶员手部动作遮挡其眼睛嘴巴等关键特征的条件下,关键点检测效果较差,导致最终疲劳检测准确率低于 75.0%.VBFLLFA 统计结果与较为接近,较 YAWDD 数据集的结果均降低了 10.0% 以上.SMGA-Net,FF-CNN 以及 JHPFA-Net 也同样在两个指标上均低于 70.0%.TV-MFFN 模型在准确率上达到 91.9%,召回率为 91.5%,较 YAWDD 数据集结果仅下降了 32.0% 与 40.0%,在可接受范围之内,验证了本文 TV-MFFN 模型在复杂驾驶条件的情况下的鲁棒性.

表 3 在包含复杂驾驶条件数据集上的对比结果

Table 3 Comparison results on datasets containing complex driving conditions

方法	准确率(%)	召回率(%)
2s-STGCN[7]	78.6	79.3
VBFLLFA[20]	80.3	81.2
SMGA-Net[5]	64.3	65.0
FF-CNN[6]	66.5	65.4
JHPFA-Net[31]	65.7	67.6
TV-MFFN	**91.9**	**91.5**

3.5 模型结构实验分析

3.5.1 模块消融实验分析

本文将 TV-MFFN 模型中的 3 个主要模块通过消融实验进行分析,以验证嵌入的各个模块在整个方法中各自的有效性.其中,Without Image 表示去除局部图像特征提取模块,Without Graph 表示去除节点图特征提取模块,Without Tem-

poral 表示去除时序向量模块.

如图 8 所示,图中每条曲线为不同模块去除情况下的测试结果得出的 ROC 曲线图. 每条曲线下方的面积(Area Under the Curve,AUC)值即代表模型整体预测能力,面积越接近于 1,则模型性能越高. 其中,去除节点图特征提取模块后,模型检测仅依靠图像中的疲劳特征,图像中冗余信息的干扰加强,因此曲线面积区域仅 0. 73. 在去除时序向量的引导后,缺少了从连续帧中提取关键动作变化的参考,模型检测难度增大,AUC 仅 0. 78. 没有图像信息模块的情况下,虽然结果优于上述两种情况,但 Dlib 关键点检测易受到图像质量的影响,关键点位置的可靠性降低,该情况下模型的鲁棒性较低. TV-MFFN 模型融合上述 3 个模块特征,弥补了每个情况下的缺陷,AUC 值达到 0. 94.

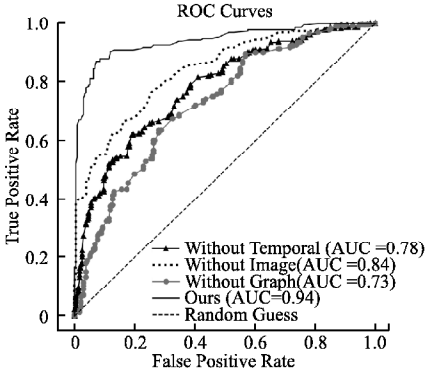

图 8　消融实验 ROC 曲线图

Fig. 8　ROC curve of ablation

3. 5. 2　融合网络结构消融实验分析

本文基于 Transformer 进行多特征融合,该模块主要包括多头注意力机制(MHA)及交叉注意力机制(CA),为了充分证明其有效性,选择了不同的模型和融合方式进行对比,包括长短期记忆网络(LSTM),卷积神经网络(CNN),以及仅使用 MHA 的方式.

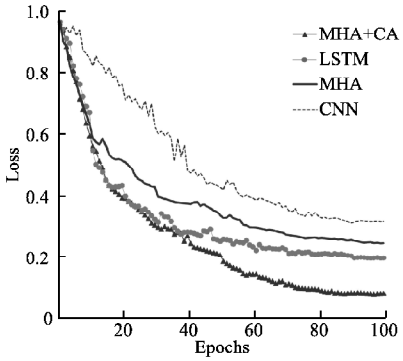

图 9　融合网络训练损失对比图

Fig. 9　Comparison of training loss of fusion network

如图 9 所示,图中的每条曲线为各种情况下的训练损失收敛的过程. 其中,使用 CNN 的训练过程中,前期震荡且下降较慢,从约第 60 个 epoch 开始进入收敛缓慢阶段,无法进一步下降,最终损失值停在 0. 38. 使用 LSTM 以及仅 MHA 的方式前期训练过程类似,但在同样的训练时间内,最终的损失值分别停在了 0. 29 与 0. 24 左右. 本文所采取 MHA + CA 的模型,在初始学习率的情况下直到约第 50 个 epoch 时收敛缓慢,因此本文将 epoch = 50 作为学习率开始线性衰减的节点. 在学习率衰减的过程中,本文模型进一步细化模型学习权重参数,最终的损失值约为 0. 06.

4　总　结

针对驾驶员视频中因光照条件、拍摄角度、遮挡等问题导致疲劳检测精度较低的问题,本文提出了一种基于时序向量引导的多特征融合模型用于驾驶员疲劳视频检测,通过局部图像模块以及节点图特征模块对划分的区域图像以及局部图提取对应的疲劳特征块(Patch),降低冗余数据信息的干扰. 同时根据相邻帧中对应关键点的位置变化构建时序向量. 最后通过设计的基于 Transformer 多特征融合模块,利用多头注意力机制进一步加强图像特征与节点特征表示,随后通过交叉注意力块融合将时序特征融入两个分支,从而引导疲劳特征对关键疲劳动作变化的关注,提高了视频检测的精度. 本文在公开数据集上将 TV-MFFN 与主流的疲劳驾驶视频检测方法进行了性能实验对比,根据实验结果分析,TV-MFFN 模型在检测的准确率及相关指标上显著优于对比方法,并且能够在复杂驾驶条件情况下保证较高的精度,具有较强的鲁棒性.

References:

[1] Liu Z, Gu C, Xie Y, et al. Realistic sketch face generation via sketch-guided incomplete restoration[C]//Proceedings of the 29th International Conference on Parallel and Distributed Systems (ICPADS),2023:32-37.

[2] Hu F,Zhang L,Yang X,et al. Eeg-based driver fatigue detection using spatio-temporal fusion network with brain region partitioning strategy[J]. IEEE Transactions on Intelligent Transportation Systems,2024,25(8):9618-9630.

[3] Pan Y,Dong Y,Wang D,et al. Comparative study on fatigue evaluation of suspenders by introducing actual vehicle trajectory data [J]. Scientific Reports,2024,14(1):5165-5179.

[4] Sedik A,Marey M,Mostafa H. An adaptive fatigue detection system based on 3D CNNs and ensemble models[J]. Symmetry, 2023,15(6):1274-1294.

[5] Huang Y,Liu C,Chang F,et al. Self-supervised multi-granularity graph attention network for vision-based driver fatigue detection [J]. IEEE Transactions on Emerging Topics in Computational Intelligence,2024,8(4):3067-3080.

[6] Sun Z,Miao Y,Jeon J Y,et al. Facial feature fusion convolutional neural network for driver fatigue detection[J]. Engineering Applications of Artificial Intelligence,2023,126:106981,doi:10.1016/j.engappai.2023.106981.

[7] Bai J,Yu W,Xiao Z,et al. Two-stream spatial-temporal graph convolutional networks for driver drowsiness detection[J]. IEEE Transactions on Cybernetics,2021,52(12):13821-13833.

[8] Ma Y,Tao Y,Gong Y,et al. Driver identification and fatigue detection algorithm based on deep learning[J]. Mathematical Biosciences and Engineering,2023,20(5):8162-8189.

[9] Alghanim M,Attar H,Rezaee K,et al. A hybrid deep neural network approach to recognize driving fatigue based on EEG signals

[J]. International Journal of Intelligent Systems, 2024, 2024(1): 1098-1110.

[10] ZHANG L H, GUO C P, XU X Z, et al. EEG fatigue state detection method based on SSL-DDPM [J] Modern Electronic Technology, 2024, 47(21): 40-45.

[11] Ding X, Chen G, Wang J, et al. Driving fatigue detection with three prefrontal EEG channels and deep learning model[C]//Proceedings of 15th International Conference on Advanced Computational Intelligence (ICACI), IEEE, 2023: 1-5.

[12] Fujiwara K, Iwamoto H, Hori K, et al. Driver drowsiness detection using RR Interval of electrocardiogram and self-attention autoencoder[J]. IEEE Transactions on Intelligent Vehicles, 2023, 9(1): 2956-2965.

[13] Wang F, Chen D, Yao W, et al. Real driving environment EEG-based detection of driving fatigue using the wavelet scattering network[J]. Journal of Neuroscience Methods, 2023, 400: 109983, doi: 10.1016/j.jneumeth.2023.109983.

[14] Yang Y, Sun B, Pan J R. A method for detecting driving behavior safety based on vehicle trajectory data[J]. Advances in Transportation Studies, 2023, 87(1): 1824-1832.

[15] CAI S X, DU C K, ZHOU S Y, et al. Fatigue driving state detection based on vehicle operation data [J]. Transportation System Engineering and Information, 2020, 20(4): 77-82.

[16] Sedik A, Marey M, Mostafa H. An adaptive fatigue detection system based on 3D CNNs and ensemble models [J]. Symmetry, 2023, 15(6): 1274-1294.

[17] Fa S, Yang X, Han S, et al. Multi-scale spatial-temporal attention graph convolutional networks for driver fatigue detection[J]. Journal of Visual Communication and Image Representation, 2023, 93: 103826-103832, doi: 10.1016/j.jvcir.2023.103826.

[18] Lv X, Zheng G, Zhai H, et al. Driver fatigue detection method based on temporal-spatial adaptive networks and adaptive temporal fusion module [J]. Computers and Electrical Engineering, 2024, 119: 109540, doi: 10.1016/j.compeleceng.2024.109540.

[19] Pandey N N, Muppalaneni N B. Dumodds: dual modeling approach for drowsiness detection based on spatial and spatio-temporal features[J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105759-105774, doi: 10.1016/j.engappai.2022.105759.

[20] Yang L, Yang H, Wei H, et al. Video-based driver drowsiness detection with optimised utilization of key facial features[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(7): 6938-6950.

[21] JING D D, LI B B, WANG S Y, et al. Bidirectional multi-level edge detection network based on transformer[J]. Journal of Chinese Computer Systems, 2024, 45(12): 3042-3049.

[22] Singhal V, Soni N, Khatri K, et al. Drowsiness detection and alert system using DLib[C]//Proceedings of International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), 2023: 242-246.

[23] Sundermeyer M, Marton Z C, Durner M, et al. Implicit 3d orientation learning for 6d object detection from rgb images[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 699-715.

[24] TAN X Y, PEI S W. Research on heterogeneous graph neural network based on higher-order neighbors[J]. Journal of Chinese Computer Systems, 2023, 44(9): 1954-1960.

[25] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]//International Conference on Learning Representations (ICLR), 2021: 11929-11950.

[26] Vaswani A. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.

[27] Chen C F R, Fan Q, Panda R. Crossvit: cross-attention multi-scale vision transformer for image classification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 357-366.

[28] Abtahi S, Omidyeganeh M, Shirmohammadi S, et al. YawDD: a yawning detection dataset[C]//Proceedings of the 5th ACM Multimedia Systems Conference, 2014: 24-28.

[29] Chen W, Zhang X, Chen S. Fatigue detection system for extracting driver's eye features[C]//Proceedings of 7th International Conference on Advanced Algorithms and Control Engineering, 2024: 891-894.

[30] Ma L, Qi W M, Chen Y, et al. Lightweight fatigue driving detection method based on improved Yolov8[C]//Proceedings of 43rd Chinese Control Conference, 2024: 7492-7497.

[31] Lu Y, Liu C, Chang F, et al. JHPFA-Net: joint head pose and facial action network for driver yawning detection across arbitrary poses in videos[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(11): 11850-11863.

附中文参考文献:

[10] 张麟华, 郭彩萍, 许骁哲, 等. 基于 SSL-DDPM 的脑电疲劳状态检测方法[J]. 现代电子技术, 2024, 47(21): 40-45.

[15] 蔡素贤, 杜超坎, 周思毅, 等. 基于车辆运行数据的疲劳驾驶状态检测[J]. 交通运输系统工程与信息, 2020, 20(4): 77-82.

[21] 荆东东, 李备备, 王诗宇, 等. BMED: 基于 Transformer 的双向多级边缘检测网络[J]. 小型微型计算机系统, 2024, 45(12): 3042-3049.

[24] 谭鑫媛, 裴颂文. 聚合高阶邻居节点的异构图神经网络模型研究[J]. 小型微型计算机系统, 2023, 44(9): 1954-1960.