doi: 10.3969/j.issn.1672-6073.2025.06.015

基于面部特征与 Bi-LSTM 的智轨驾驶员疲劳检测算法

闵思远[1,2]，孟新宇[1,2]，张锐骥[1,2]，辛开元[1,2]，殷　勇[1,2]
（1. 西南交通大学交通运输与物流学院，成都 611756；2. 西南交通大学宜宾研究院，成都 644000）

摘　要：针对现有视觉疲劳检测方法普遍存在场景适应性不足以及时间特征依赖关系建模不充分的问题，提出一种基于面部多特征与双向长短期记忆网络的智轨驾驶员疲劳检测算法。首先，算法通过开源机器学习与计算机视觉库(Dlib)定位驾驶员面部关键点，计算出眼睛纵横比(eye aspect radio，EAR)、嘴部纵横比(mouth aspect ratio，MAR)、头部姿态角等 5 个疲劳特征参数，并提出纵向头部巡视行为频率用于分心检测；其次，通过双向长短期记忆网络(bidirectional long short-term memory，Bi-LSTM)网络建立时序依赖关系，完成驾驶状态的判定；再次，引入时间注意力来增强关键特征的表征能力，提高算法整体性能；最后，与基于眼部状态检测的模型相比，本文模型的准确率为 94.24%，F1-Score 值由 0.8324 提升至 0.9027，验证了本文方法的有效性与可行性。研究结果可为公共交通安全保障提供参考。

关键词：深度学习；面部多特征；疲劳检测；Bi-LSTM；人脸关键点
中图分类号：U231.1　　**文献标志码：**A　　**文章编号：**1672-6073(2025)06-0116-09

Fatigue Detection Algorithm for Intelligent Rail Drivers Based on Facial Features and Bi-LSTM

MIN Siyuan[1,2], MENG Xinyu[1,2], ZHANG Ruiji[1,2], XIN Kaiyuan[1,2], YIN Yong[1,2]
(1. School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756;
2. Yibin Research Institute, Southwest Jiaotong University, Yibin 644000)

Abstract: Aiming at the common issues of insufficient scene adaptability and inadequate modeling of temporal feature dependencies in existing visual fatigue detection methods, this paper proposes an intelligent rail driver fatigue detection algorithm based on facial multi-features and Bi-LSTM. First, the algorithm uses Dlib to locate the driver's facial keypoints and calculates five fatigue feature parameters, including the Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and head pose angles. It also proposes the frequency of longitudinal head scanning behavior for distraction detection. Second, the Bi-LSTM network is employed to establish temporal dependencies and complete the determination of the driving state. Third, temporal attention is introduced to enhance the representation capability of key features, improving the overall performance of the algorithm. Finally, the research results show that, compared to traditional models, the proposed method achieves an accuracy of 94.24% and improves the F1-Score from 0.8324 to 0.9027, validating the effectiveness and feasibility of the proposed approach. The findings can provide a reference for public transportation safety assurance.

Keywords: deep learning; multiple facial features; fatigue detection; Bi-LSTM; facial landmarks

收稿日期：2025-05-14　修回日期：2025-07-10

第一作者：闵思远，男，硕士研究生，从事驾驶疲劳检测研究，minsiyuan@my.swjtu.edu.cn
通信作者：殷勇，男，博士，副教授，从事交通运输组织优化研究，yinyong@home.swjtu.edu.cn

基金项目：宜宾双城市校协议专项科研经费科技项目管理委员会(YBSCXY2023020011)；国家重点研发计划(2022YFB4300503)
引用格式：闵思远，孟新宇，张锐骥，等. 基于面部特征与 Bi-LSTM 的智轨驾驶员疲劳检测算法[J]. 都市快轨交通，2025，38(6)：116-124.
MIN Siyuan, MENG Xinyu, ZHANG Ruiji, et al. Fatigue detection algorithm for intelligent rail drivers based on facial features and Bi-LSTM[J]. Urban rapid rail transit, 2025, 38(6): 116-124.

随着城市智能轨道快运系统(简称智轨)出行需求和运营规模的持续扩大，其在公共交通领域的大规模应用正面临着新的安全挑战。据统计数据显示[1]，在智轨系统运营过程中，由人为因素引发的安全事故占比高达运输事故总量的 82.3%，其中由驾驶人员疲劳状态所致的安全隐患尤为突出。疲劳状态会对驾驶员警觉度、感知能力、决策力等多个方面产生负面影响，极易导致交通事故，研究表明驾驶员因其疲劳状态发生的事故损伤程度是清醒情况下的 3.3 倍[1]。因此，研究疲劳驾驶检测非常重要。

围绕疲劳检测的实现路径，现有疲劳驾驶检测系统按传感器类型可分为基于车辆特征、生理特征和面部特征 3 类。然而，复杂的驾驶环境可能会降低基于车辆特征的疲劳检测准确性。脑电图(electroencephalography，EEG)、心电图(electrocardiogram，ECG)等生理信号是直接衡量疲劳的标准，但该类生理特征的收集需要将传感器贴附在人体皮肤上，这可能会引起驾驶员不适，进而影响驾驶员的正常驾驶行为。基于面部特征的疲劳检测以其直观性强和非侵入式的优势得到广泛应用。因此，利用驾驶员面部特征进行疲劳检测成为近年来较为热门的研究方法。CHANG 等[2]提出一种基于驾驶员视频分析非侵入式的疲劳检测系统，以闭眼持续时间作为参数，再使用眼睑闭合百分比(PERCLOS)[3]对驾驶状态进行分类。这种方法使用传统的机器学习库提取眼睛和嘴巴的特征，使用固定阈值计算疲劳状态，其准确率有待提高。为提升预测准确率，FRIEDRICHS 等[4]在 18 种不同眼部特征中筛选，最终发现以 PERCLOS 和睁眼速度作为特征参数的准确率最高。为进一步提升检测速度和精度，LI 等[5]引入改进的 YOLOv3 模型来捕捉人脸面部区域，并利用眼部特征向量和嘴部特征向量计算闭眼时间、眨眼频率和哈欠频率，从而评估驾驶员的疲劳状态。ZHANG 等[6]利用红外视频检测，提出一种基于卷积神经网络(convolutional neural network，CNN)的眼部状态识别方法，最终通过计算 PERCLOS 和眨眼频率来检测疲劳程度，但此时疲劳检测的判断仅依赖于眼部状态。为增加特征维度，DWIVED 等[7]通过深度学习对用于疲劳检测的特征进行了筛选，并整理出可用于疲劳检测的特征。LIU 等[8]提出一种基于卷积神经网络和长短期记忆(long short-term memory，LSTM)的驾驶员疲劳检测方法，将 PERCLOS、嘴部特征和脸部方向作为参数，并将连续时间序列上的这些特征参数输入到 LSTM 中，完善了 CNN 不能提取时序信息的缺点。

虽然基于视觉的疲劳检测技术已经取得较好进展，但仍存在一些待解决的问题。本文提出一种新的智轨驾驶员疲劳检测模型。首先，提出基于眼睛纵横比、嘴部纵横比及头部姿态的多面部特征融合方案，并引入纵向头部巡视行为频率作为智轨场景下的分心检测专属特征；然后，在双向长短期记忆网络(bidirectional long short-term memory，Bi-LSTM)上融合时间注意力机制，对序列帧按重要性加权，并将特征编码层与注意力化的 Bi-LSTM 结合，实现具备时序上下文感知的疲劳判断，满足实时推理要求；最后，基于混合数据集(公开数据+自建智轨匹配样本)进行训练与评估，以场景适配性为导向的数据预处理与标注策略。

1 模型方法

本文提出了一种基于深度学习的智轨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，结合使用了开源机器学习与计算机视觉库(Dlib)、多尺度特征编码(heteroscale-temporal encoder，HSTE)和 Bi-LSTM，并在 Bi-LSTM 上融合了时间注意力机制(time atention mechanism，TAM)。

疲劳检测算法主要由 3 个阶段组成，包括人脸关键点检测、面部特征提取和驾驶状态分类。

在人脸关键点检测阶段，Dlib 基于稀疏像素强度数据子集得到视频流中每帧的关键点位置信息，其检测结果直接决定了后续面部特征的提取及疲劳驾驶检测的准确性[9]。

在特征提取阶段，依据关键点的坐标信息计算得到驾驶员的眼部、嘴部及头部姿态的 5 个疲劳特征参数，并通过 HSTE 为后续的 Bi-LSTM 网络提供高阶时序编码。同时，基于头部姿态提出了纵向头部巡视行为频率智轨场景下的专属特征，用以实现驾驶员分心行为的判定。

在分类阶段，使用基于时间注意力的 Bi-LSTM 捕捉不同疲劳行为与时间序列之间的依赖相关性，最终基于 Softmax 模块输出每个驾驶片段是否疲劳的二分类结果。智轨驾驶员疲劳检测算法的总体架构如图 1 所示。

1.1 面部特征提取

1.1.1 基于 Dlib 的人脸关键点检测

本文选用 Dlib 提取驾驶员人脸细粒度特征，Dlib 是一个基于现代 C++开发的工具库，集成了机器算法

和工具组件，包含了 68 个面部关键点检测模型，该模型由数百万张人脸训练而成。该方法检测人脸时，通过级联形状回归，直接从稀疏的像素强度数据子集中估计面部关键点的位置，兼具高精度与耗时低的特性，适合车载部署。检测过程如图 2 所示，当 Dlib 检测智轨驾驶员面部时，可实时捕捉驾驶员面部 68 个关键点，将其编码为表征人脸信息的 128 维特征向量，用于后续计算面部特征。

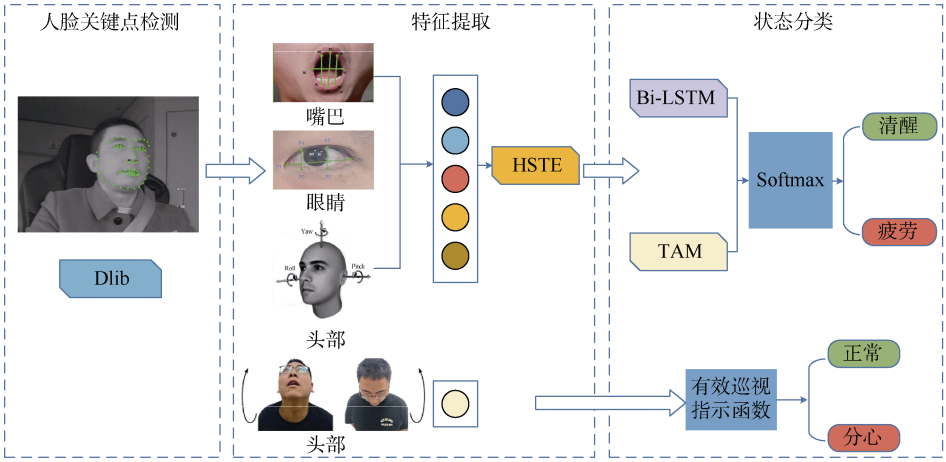


图 1　智轨驾驶员疲劳检测算法框架

Figure 1　Driver fatigue detection algorithm framework

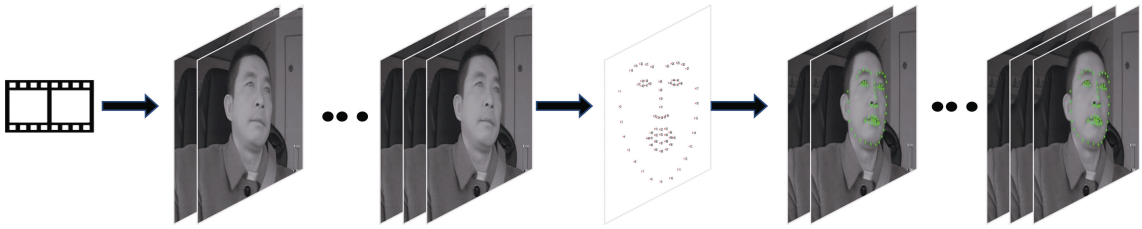


图 2　人脸关键点检测流程

Figure 2　Facial landmark detection process

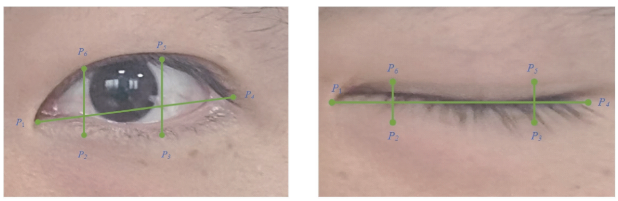
1.1.2　基于几何的面部特征

1) 眼睛纵横比。眼睛纵横比(eye aspect radio，EAR)，即眼睛的长宽之比。当驾驶员眼部为睁开状态时，EAR 值较大，当为闭合状态时，EAR 值较小。所以本文选用眼睛纵横比值作为衡量驾驶员眼部开合大小的特征，用于检测驾驶员眼睛是否闭合。

首先利用 Dlib 对视频流中的驾驶员面部进行标记，得到眼部关键点的定位如图 3 所示。再利用眼部关键点坐标信息，得到点之间的欧氏距离来计算 EAR，最终 EAR 的值由左眼和右眼同权重计算。EAR 的计算式为：

$$\mathrm{EAR}=\frac{\|p_2-p_6\|+\|p_3-p_5\|}{2\|p_1-p_4\|} \tag{1}$$

式中，p_1～p_6 为人眼周围的 6 个关键点的二维坐标。



(a) 睁眼点位示意图　　(b) 闭眼点位示意图

图 3　眼睛纵横比

Figure 3　Eye aspect ratio

2) 嘴部纵横比。嘴部纵横比(mouth aspect ratio，MAR)，即嘴部的长宽之比。当驾驶员打哈欠时，MAR 值较大，当为闭合状态时，MAR 值较小，所以本文选用 MAR 值作为衡量驾驶员嘴部开合大小的特征，利用嘴部状态判断驾驶员是否进行打哈欠行为。

同 EAR 值相似，使用 Dlib 得到的嘴部关键点的定位如图 4 所示。再利用点间欧氏距离计算 MAR，MAR 的表达式为

$$\mathrm{MAR}=\frac{\|M_2-M_8\|+\|M_3-M_7\|+\|M_4-M_6\|}{3\|M_1-M_5\|} \tag{2}$$

式中，M_1～M_8 为人脸的 8 个关键点的二维坐标。

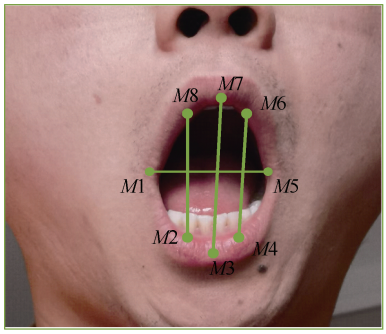


图 4　嘴巴纵横比

Figure 4　Mouth aspect ratio

3) 头部姿态。当驾驶员产生瞌睡动作时，即点头或歪头，其俯仰角或翻滚角会瞬间变大，因此本文选用 θ_{Pitch}(头部俯仰角，绕 X 轴旋转角度)、θ_{Yaw}(头部偏航角，绕 Y 轴旋转角度)、θ_{Roll}(头部翻滚角，绕 Z 轴旋转角度)作为衡量驾驶员疲劳状态的特征。

头部特征的计算由以下步骤实现[10]：首先，使用 Dlib 从图像中提取特定 2D 点，将其与 3D 人脸模型之间建立 2D-3D 严格匹配，模型如图 5 所示。其次，构建相机矩阵，假设焦距近似为图像宽度，形成投影模型。再次，使用 solvePnP 函数求解旋转向量和平移向量，再通过 Rodrigues 变换将旋转向量转换为旋转矩阵 $\boldsymbol{R}$，最终 RQ 分解求得 Pitch、Yaw，Roll，RQ 分解式为

$$\begin{cases} \theta_{\text{Pitch}} = \arctan 2\left(R_{2,1}, R_{2,2}\right) \\ \theta_{\text{Yaw}} = \arctan 2\left(-R_{2,0}, \sqrt{R_{2,1}^2 + R_{2,2}^2}\right) \\ \theta_{\text{Roll}} = \arctan 2\left(R_{1,0}, R_{0,0}\right) \end{cases} \tag{3}$$

式中：θ_{Pitch}、θ_{Yaw}、θ_{Roll} 分别为头部俯仰角、头部偏航角、头部翻滚角的值；$R_{i,j}$ 表示旋转矩阵 $\boldsymbol{R}$ 的第 i 行，第 j 列元素。

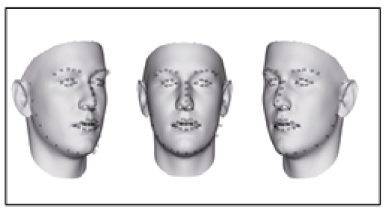


图 5　头部姿态三维模型

Figure 5　3D head pose model

基于式(1)、(2)、(3)，得到特征 EAR、MAR、θ_{Pitch}、θ_{Yaw}，θ_{Roll} 的值，但由于不同指标量纲不同，需对所有得到的特征值进行标准化处理，消除量纲差异。本文使用 Z-Score 标准化将不同尺度的数据转换为均值为 0、标准差 1 的标准正态分布，使得模型可以均衡学习不同量纲特征，保证梯度平稳，进而提升模型收敛速度，转换式为

$$\begin{cases} \mu = \dfrac{1}{N}\sum_{i=1}^{N} x_i \\ \sigma = \sqrt{\dfrac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2} \\ z_i = \dfrac{x_i - \mu}{\sigma} \end{cases} \tag{4}$$

式中，μ 为每个特征值所有数据点的平均值；σ 为标准差；z_i 为最后标准化后的数据；N 为样本总数；x_i 为对应特征值(5 种特征值处理方式一致)。

4) 纵向头部巡视行为频率。自动循迹模式下，驾驶员需周期性检查仪表盘信息与前方路况，以确保能够随时接管车辆，这一操作行为通常伴随着规律性的低头与仰头动作。为此，本文提出纵向头部巡视行为频率作为新特征量化操作频率，用于建模该类周期性头部姿态变化，以检测驾驶员因分心导致的操作不规范。

在每一个独立的 1 min 时间窗口内(不跨分钟计算)，在任意连续 15 帧内，若驾驶员头部俯仰角从 5°～15° 下降至–15°～5°[11]，并在随后 15 帧内再次回到 5°～15°，则视为一次有效的纵向头部巡视行为。当 $N_{\text{scan}}^{(m)} \leqslant 1$ 时，表明第 m min 驾驶员未能按规定周期性检查仪表盘与路况，判定为分心状态，反之则为正常。

有效巡视指示函数定义为

$$I^{(\text{m})}(s) = \begin{cases} 1, & \\ & s_1 \in [s, s+14], \quad s_2 \in [s_1, s_1+14] \\ & \theta_{\text{pitch}}(s) \in [5^\circ, 15^\circ] \\ & \theta_{\text{pitch}}(s_1) \in [-15^\circ, 5^\circ] \\ & \theta_{\text{pitch}}(s_2) \in [5^\circ, 15^\circ] \\ 0, & \text{否则} \end{cases} \tag{5}$$

式中：s、s_1、s_2 为视频帧序号，m 为分钟序号。

第 m min 巡视次数的定义为

$$\text{N}_{\text{scan}}^{(m)} = \sum_{t=(m-1)\times 1800+1}^{m\cdot 1800-29} I^{(m)}(\text{s}) \tag{6}$$

1.1.3　多尺度特征提取及时序编码

在进行驾驶疲劳状态分类前，提出了 HSTE 模块，用于承担多尺度时间特征提取和时序信号压缩编码任

务，为后续的 Bi-LSTM 网络模型提供更有效的高层次特征，提高驾驶员疲劳检测模型的整体性能。网络架构细节如表 1 所示。

表 1 特征编码层的网络架构

Table 1 Feature encoding layer network architecture

层级类别	层类型	卷积核尺寸	输出形状	参数数量
多分支卷积	1D Conv	1×3	(32，64，90)	1 024
	ReLU	-	-	-
	1D Conv	1×5	(32，128，90)	3 328
	ReLU	-	-	-
	1D Conv	1×7	(32，256，90)	9 216
	ReLU	-	-	-
	通道拼接	-	(32，448，90)	-
特征融合	Conv1d	1×1	(32，256，90)	114 944
	ReLU	-	-	-
空洞卷积层 1	Conv1d	1×3	(32，256，90)	196 864
	BatchNorm1d	-	-	-
	ReLU	-	-	-
空洞卷积层 2	Conv1d	1×3	(32，256，90)	196 864
	BatchNorm1d	-	-	-
	ReLU	-	-	-
	残差连接	-	(32，256，90)	
一维最大池化	Maxpol1d	1×3	(32，256，30)	-
	ReLU	-	-	-

该网络首先将包含 5 个特征的时间序列作为输入，经 3 个并行的卷积分支提取多尺度特征，卷积核大小分别为 3、5、7，小核可聚焦局部细节，如眼部瞬时闭合，大核可提取长段上下文特征，如头部姿态连续变化。其次，通过通道拼接和特征融合，保留各分支原始特征信息，避免因池化导致信息丢失。再次，经过残差空洞卷积扩展时序感受野，结合残差连接稳定梯度传播。最后，通过最大池化压缩序列长度，得到特征维度 256 的时间序列。

1.2 基于 Bi-LSTM 与 TAM 的疲劳状态分类

正常人的眨眼、打哈欠或者低头等行为都是一个动态的连续过程，持续时间约为 1～2 s，仅依赖单帧图像中的面部状态无法准确判断是否疲劳。因此，在视频数据中，驾驶员疲劳状态会在阶段时间内的动作中得到体现。为了通过连续帧特征之间的关系来确定驾驶员的当前状态，本文使用了特殊类型的循环神经网络(recurrent neural network，RNN)，即长短期记忆网络(LSTM)[12]，并引入了注意力机制，动态调整时序数据上不同时间步的权重，以加权聚焦于更重要的疲劳信号。

本文提出的基于时间注意力机制的 Bi-LSTM 网络架构如表 2 所示。

表 2 基于时间注意力机制的 Bi-LSTM 的网络架构

Table 2 Network architecture of Bi-LSTM with temporal attention mechanism

层级类别	操作	输出形状	参数数量
BiLSTM	-	(32，30，256)	395264
时间注意力机制	线性变换	(32，30，1)	257
	归一化权重	(32，30，1)	0
	加权求和	(32，256)	0
分类器	全连接层	(32，128)	32896
	ReLU	-	-
	Dropout	(32，128)	-
	全连接层	(32，2)	258

双向长短期记忆网络作为一种特殊的 RNN，由两个独立的 LSTM 网络组成[13]：一部分按照序列的正向顺序处理数据，另一部分按照反向顺序处理数据，并共同决定最终的输出值，输入数据序列经 Bi-LSTM 得到一系列正向和反向的隐藏状态，函数式为

$$\begin{cases} \boldsymbol{h}_{ft} = \overrightarrow{\mathrm{LSTM}}\left(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}\right) \\ \boldsymbol{h}_{bt} = \overleftarrow{\mathrm{LSTM}}\left(\boldsymbol{x}_t, \boldsymbol{h}_{t+1}\right) \\ H_f = \left(\boldsymbol{h}_{f1}, \boldsymbol{h}_{f1}, \cdots, \boldsymbol{h}_{fT}\right) \\ H_b = \left(\boldsymbol{h}_{bT}, \boldsymbol{h}_{f(T-1)}, \cdots, \boldsymbol{h}_{b1}\right) \\ Y_t = H_f + H_b \end{cases} \tag{7}$$

式中，$\boldsymbol{x}_t$ 为时间步 t 的输入向量；h_{t-1} 为上一个时刻的隐藏状态；h_{ft} 为前向 LSTM 的隐藏状态；h_{bt} 为后向 LSTM 的隐藏状态；$\overrightarrow{\mathrm{LSTM}}$ 为前向传递网络；$\overleftarrow{\mathrm{LSTM}}$ 为后向传递网络；Y_t 为前后向合并的隐藏状态；T 表示时间步的总长度，即序列的最后一个时间点的索引。

LSTM 网络的循环单元结构如图 6 所示，h_t 为当前时刻的隐藏状态；c_t 为当前时刻的细胞状态；c_{t-1} 为上一个时刻的细胞状态；f_t 是遗忘门，决定丢弃多少信息；i_t 是输入门，决定写入多少新的信息到细胞状态；$\tilde{c}_t$ 是候选细胞状态，表示要加入细胞状态的新信息；o_t 是输出门，决定从细胞状态输出多少到 h_t；σ 是 sigmoid 激活函数；tanh 是双曲正切激活函数。LSTM 通过遗忘门、输入门和输出门协同工作，对细胞状态进行选择性遗忘、更新与输出，从而能够存储并动态

调整历史信息。然而，传统 LSTM 仅依赖过去观测，缺乏对未来时刻信号的利用，容易在序列开头或末尾对关键疲劳动作(如闭眼持续、频繁点头)做出错误或迟滞判断。疲劳状态具有持续性，比如闭眼持续时间变长、频繁点头等，体现这些疲劳行为的特征在时间序列中具有一定的前后关联，而双向长短期记忆网络可以捕获前向和后向的信息，结合未来几帧的数据信息可以辅助验证当前状态的决策，减少决策误判，显著提升模型的预测精度。

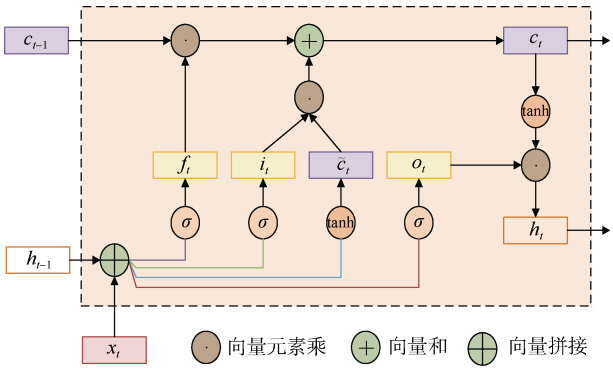


图 6　LSTM 网络的循环单元结构

Figure 6　The recurrent unit structure of an LSTM network

尽管 Bi-LSTM 在时序依赖建模上较 LSTM 有明显优势，但序列中各时间步权重相同，无法主动区分高价值关键帧与普通帧，导致模型在多场景下的检测性能波动较大。基于此，本文在 Bi-LSTM 基础上引入时间注意力机制，通过对每个时间步的隐藏状态计算注意力得分并归一化为权重，使模型自动聚焦于关键疲劳信号所在帧。这样既保留了双向上下文的优势，又能在序列中突出闭眼、点头等动作的重要性，显著提升了模型的鲁棒性与精度。

2　实验结果与分析

为验证所提模型在精度与实时性上的优越性，量化各模块对序建模能力的实际增益，本文先基于公开数据集 YAWDD 与自建男性驾驶员数据，经过筛选、切片、Dlib 特征提取与 Z-Score 标准化，构建数据集并设置训练超参；再通过模型对比实验(与 RetinaFace+ERT、HOG+CNN 比较)评估所提算法的整体检测性能，并用逐步递进的消融实验考察各关键组件对准确率、推理时间和 AUC(ROC 曲线下面积)的独立贡献。

2.1　数据集收集与预处理

本文数据集采用了公开数据集 YAWDD[14]与自建数据集：YAWDD 包含 57 名男性及 50 名女性志愿者的模拟驾驶视频，覆盖多年龄、种族和面部特征，记录不同光照条件下的正常驾驶、闭眼、打哈欠等行为；鉴于宜宾智轨驾驶员群体中 36～50 岁男性占比达 88%，自建数据集通过模拟驾驶平台，采集 10 名男性驾驶员(含 28 岁 1 名、52 岁 1 名、36～50 岁 8 名)的 30 段视频(总时长 50 min)，重点捕获正常驾驶、长时间闭眼、打哈欠、头部大角度转动、后仰及瞌睡状态等行为，以适配实际应用场景。

因当前智轨驾驶员都为男性，为与智轨工作场景匹配，本实验对 YAWDD 数据集进行以下处理：去除女性驾驶员视频及佩戴墨镜的男性驾驶员视频，剩余共计 243 段。两种数据集视频原始帧率都为 30，为提升样本数量，本文将驾驶视频切割为共计 90 帧的视频片段，共计获得样本 4 072 份。依据表 3 所列的疲劳特征标准，对所有视频样本进行人工标注：完全符合判定标准的样本标记为疲劳状态，其他样本标记为清醒状态。经 4 人复核验证，最终获得疲劳样本 3 563 份，清醒样本 509 份。将混合后的疲劳驾驶数据集按 7∶2∶1 的比例划分为训练集、验证集和测试集，数量分别为训练集 2 851 份、验证集 814 份、测试集 407 份。然后，使用 Dlib 提取视频数据集中的特征数据，针对时序特征中的异常值，剔除超出合理范围的数值，使用前后两帧的均值进行填充，再对得到的数据进行 Z-Score 标准化。

表 3　疲劳特征标准

Table 3　Fatigue feature criteria

疲劳特征	标准描述
闭眼持续	眼部完全闭合连续超过 24 帧
打哈欠	样本视频中连续出现打哈欠动作≥2 次
头部低垂	面部明显向下偏离水平视线，持续超过 30 帧
微睡	出现连续≥2 次的“慢速眨眼”，或眼睑明显下垂且伴随头部轻微前倾动作，任意一种情况持续≥30 帧
深睡	大角度仰头靠座位睡眠及低头睡眠

在训练过程中，使用训练集数据进行迭代训练，得到本文所提出模型的有效参数，然后利用验证集确定模型的最优参数，最后，使用测试集验证模型的整体准确性。

2.2　实验环境与评价指标

在本文实验中，批量大小设定为 32，训练 100 个周期，初始学习率设定为 0.0001，L2 正则化系数设定为 0.001，防止模型过拟合，提升泛化能力。实验环境配置如表 4 所示。

表 4 实验环境配置

Table 4 Experimental environment configuration

设备名称	设备信息
操作系统	Windows 11
CPU	I9-10900K
GPU	RTX3090
Python 版本	3.8
CUDA 版本	11.6
Pytorch 版本	2.6.0

针对二分类问题，预测结果通常存在着 4 种互斥的预测情况，一般情况下用 TP、TN、FP、FN 代表 4 种预测结果。TP 为样本是正类，模型预测结果也为正类；TN 为样本是负类，样本预测结果也为负类；FP 是样本实际为负类，但模型误判为正类；FN 是样本实际为负类，但模型误判为正类。本文使用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 值(F1-Score)，共 4 种评价指标评估模型优劣，其相关计算式为

$$\begin{cases} \text{Accuracy} = \dfrac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \\ \text{Precision} = \dfrac{\text{TP}}{\text{TP+FP}} \\ \text{Recall} = \dfrac{\text{TP}}{\text{TP+FN}} \\ \text{F1} = 2 \times \dfrac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{cases} \tag{8}$$

2.3 模型对比实验

为验证本文算法模型检测性能，将不同模型基于同一数据集训练测试，最终得到不同模型的 4 种指标评价结果，如表 5 所示，基于 RetinaFace+ERT 模型的疲劳检测仅对人眼特征进行提取，疲劳状态的判断完全依赖于眼睛状态，忽略其他部位的重要信息，造成整体算法漏判可能性提高，准确率仅为 0.6812；在基于 HOG+CNN 的模型中，将图像分为眼睛和嘴巴区域，再输入到网络中进行卷积，提取特征进行疲劳判断，但由于受到多余冗杂信息干扰，且未考虑前后状态之间的联系，准确率为仅为 82.86；本文提出的疲劳检测算法相比于以上算法具有更高的准确率，准确率达到 94.24%。因此，基于面部多特征使用与前后驾驶状态有关的算法检测可以实现较高的准确率。

表 5 模型对比结果

Table 5 Model comparison results

模型	准确率/%	精确率/%	召回率/%	*F*1
RetinaFace+ERT[15]	69.54	66.89	69.12	0.6835
HOG+CNN[16]	82.86	83.91	82.97	0.8324
本文算法	94.24	92.15	92.61	0.9027

2.4 网络结构消融实验分析

为系统评估本文所提出模型结构中各关键模块对疲劳驾驶检测性能的实际贡献，本文设计了 4 个递进式的消融对比实验，模型结构详见表 6，A 模型为本文提出模型；B 模型在 A 模型基础上移除 Bi-LSTM，使用 LSTM 替代；C 模型在 B 模型基础上移除注意力机制；D 模型在 C 模型基础上移除 HSTE。每个实验模型均在统一数据集和训练策略下进行训练与测试，确保对比的公平性和一致性。实验选取准确率、推理耗时以及 ROC 曲线面积(AUC)作为主要评价指标，综合反映模型在识别精度、响应效率及整体判别能力方面的表现。通过逐步剔除或保留 HSTE 模块、注意力机制和 Bi-LSTM 结构，可以清晰识别出各组件在模型中的独立作用及其协同效果，如表 6 和图 7 所示。

表 6 消融实验模型结构及结果

Table 6 Ablation study model architecture and results

模型编号	HSTE	注意力机制	Bi-LSTM	准确率%	推理时间/(ms/*f*)
A	√	√	√	94.24	25.87
B	√	√	×	87.07	21.15
C	√	×	×	81.35	19.12
D	×	×	×	77.12	16.94

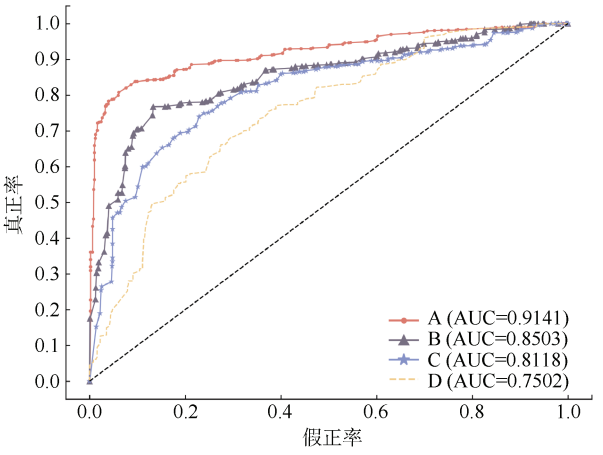


图 7 ROC 曲线

Figure 7 ROC curves

由表 6 和图 8 可知：当去除 HSTE、注意力机制与 Bi-LSTM 后，模型 D 由 LSTM 和 Softmax 分类器构成，其仅依赖单层时序建模结构进行疲劳识别，缺乏对多尺度特征与关键帧的建模能力，冗余信息干扰增强，导致准确率降至 77.12%，ROC 曲线面积(AUC)

仅 0.7502，整体识别性能较弱，推理时间为 16.94 ms/帧；在仅保留 HSTE 模块的模型 C 中，由于缺乏注意力机制的时序加权，模型在连续帧中难以聚焦关键疲劳动作，AUC 值仅为 0.8118，准确率为 81.35%，推理时间为 19.12 ms/*f*，虽然速度较快，但鲁棒性依然有限；模型 B 在模型 C 基础上引入了注意力机制，增强了对重要帧段(如连续闭眼、打哈欠等)的响应能力，使 AUC 值提升至 0.8503，准确率也提升至 87.07%，但由于仍为单向 LSTM，无法感知前后时序间的完整依赖关系，推理延时为 21.15 ms/*f*；完整模型 A 融合了 HSTE、注意力机制与 Bi-LSTM 3 个模块，既能实现多尺度疲劳信号建模，又可动态分配帧权重并引入前后文依赖，使模型在时序上下文中实现更充分的疲劳动作理解，最终准确率达到 94.24%，AUC 值提升至 0.9141，同时推理时间控制在 25.87 ms/*f*，满足 30 FPS 实时检测需求，在精度与效率之间取得了良好平衡。该实验充分验证了各模块设计的合理性及其在疲劳检测中的增益作用，检测实际效果如图 8 所示。



(a) 闭眼检测示例　(b) 低头检测示例



(c) 哈欠检测示例　(d) 瞌睡检测示例

图 8　实际检测效果

Figure 8　Practical detection effectiveness

3　结论

1) 针对现有视觉疲劳检测方法在智轨场景下适配性不足的问题，本文提出将眼动、嘴部和头部姿态面部特征用于疲劳检测，并针对智轨场景提出了纵向头部巡视行为频率指标用于分心检测。实验结果表明，在模拟智轨驾驶场景下，本文算法均能保持 90%以上的准确率，较单一眼动特征方法提升了 24.7%。这说明多特征协同提取与多尺度融合有效提高了模型在智轨场景的适应性。

2) 为克服传统时序分类模型对关键疲劳动作响应滞后、时间依赖建模不充分的缺陷，本文在 Bi-LSTM 基础上引入时间注意力机制，实现了对闭眼、打哈欠等高价值帧的自动加权。与基于眼部状态检测的模型相比，本文模型的 F1-Score 分别由 0.683 5 和 0.832 4 提升至 0.902 7，显著降低了误检与漏检率。进一步的消融实验表明，TAM 与 Bi-LSTM 联合使用可在保持实时性的同时增强长短时依赖捕捉能力，改善了现有方法时间特征依赖建模不充分的问题。

本文仅就面部特征信息进行了研究，未来可融合心率、皮电、脑电等生理信号与面部动作建模，实现多场景疲劳预警，并可推广至无人驾驶与智能交通系统，提升整体行车安全与效率。

参考文献

[1] LIU Z J, GU C H, XIE Y, et al. Realistic sketch face generation *via* sketch-guided incomplete restoration[C]//2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS). Ocean Flower Island, 2023: 32-37.

[2] CHANG R C, WANG C Y, CHEN W T, et al. Drowsiness detection system based on PERCLOS and facial physiological signal[J]. Sensors, 2022, 22(14): 5380.

[3] WIERWILLE W, WREGGIT S, KIRN C, et al. Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorithms for detection of driver drowsiness. final report[R]. 2019.

[4] FRIEDRICHS F, YANG B. Camera-based drowsiness reference for driver state classification under real driving conditions[C]//2010 IEEE Intelligent Vehicles Symposium. La Jolla, 2010: 101-106.

[5] LI K N, GONG Y B, REN Z L. A fatigue driving detection algorithm based on facial multi-feature fusion[J]. IEEE Access, 2020, 8: 101244-101259.

[6] ZHANG F, SU J J, GENG L, et al. Driver fatigue detection based on eye state recognition[C]//2017 International Conference on Machine Vision and Information Technology (CMVIT). Singapore. 2017: 105-110.

[7] DWIVEDI K, BISWARANJAN K, SETHI A. Drowsy driver detection using representation learning[C]//2014 IEEE International Advance Computing Conference (IACC). Gurgaon, 2014: 995-999.

[8] 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): 576-595.

[9] 杨艳艳, 李雷孝, 林浩. 提取驾驶员面部特征的疲劳驾驶检测研究综述[J]. 计算机科学与探索, 2023, 17(6): 1249-1267.
YANG Yanyan, LI Leixiao, LIN Hao. Review of research on fatigue driving detection based on driver facial features[J]. Journal of frontiers of computer science and technology, 2023, 17(6): 1249-1267.

[10] CAO Z W, CHU Z C, LIU D F, et al. A vector-based representation to enhance head pose estimation[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, 2021: 1187-1196.

[11] YE M, ZHANG W W, CAO P C, et al. Driver fatigue detection based on residual channel attention network and head pose estimation[J]. Applied sciences, 2021, 11(19): 9195.

[12] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. Neural computation, 2000, 12(10): 2451-2471.

[13] HU X M, YUAN S S, XU F Z, et al. Scalp EEG classification using deep Bi-LSTM network for seizure detection[J]. Computers in biology and medicine, 2020, 124: 103919.

[14] ABTAHI S, OMIDYEGANEH M, SHIRMOHAMMADI S, et al. YawDD: a yawning detection dataset[C]//Proceedings of the 5th ACM Multimedia Systems Conference. Singapore, 2014: 24-28.

[15] 雷富强, 张博雅, 张一帆, 等. 基于 RetinaFace 和 ERT 的眼部疲劳检测方法[J]. 计算机应用与软件, 2024, 41(10): 227-232.
LEI Fuqiang, ZHANG Boya, ZHANG Yifan, et al. Eye fatigue detection based on retinaface and ert[J]. Computer applications and software, 2024, 41(10): 227-232.

[16] 周先春, 邹清宇, 陆滇. 基于卷积神经网络的疲劳检测改进算法[J]. 计算机应用与软件, 2024, 41(6): 156-160, 168.
ZHOU Xianchun, ZOU Qingyu, LU Dian. An improved algorithm for fatigue detection based on cnn[J]. Computer applications and software, 2024, 41(6): 156-160.

（编辑：王艳菊）

（上接第 97 页）

[4] 林晓玲. 基于多约束三维场景的漫游导航算法研究[D]. 深圳: 深圳大学, 2022.
LIN Xiaoling. Research on roaming navigation algorithm based on multi-constrained three-dimensional scene[D]. Shenzhen: Shenzhen University, 2022.

[5] 范志伟. 基于数字孪生的三维场景建模及可视化研究[D]. 桂林: 桂林电子科技大学, 2024.
FAN Zhiwei. Research on 3D scene modeling and visualization based on digital twin[D]. Guilin: Guilin University of Electronic Technology, 2024.

[6] 张正, 李鑫. CIM 平台三维漫游功能的实现与优化技术研究[J]. 电子设计工程, 2023, 31(22): 16-20.
ZHANG Zheng, LI Xin. Research on realization and optimization technology of 3D roaming function of CIM platform[J]. Electronic design engineering, 2023, 31(22): 16-20.

[7] 袁占全, 李明益. 基于 BIM+GIS 的航道三维漫游系统设计与实现[J]. 水运工程, 2022(2): 153-157.
YUAN Zhanquan, LI Mingyi. Design and implementation of three-dimensional waterway roaming system based on BIM+GIS[J]. Port & waterway engineering, 2022(2): 153-157.

[8] 樊文平, 王建华, 张静, 等. 基于无人机倾斜摄影测量的虚拟校园漫游系统设计与实现[J]. 测绘工程, 2023, 32(4): 21-28.
FAN Wenping, WANG Jianhua, ZHANG Jing, et al. Design and implementation of virtual campus roaming system based on UAV tilt photogrammetry[J]. Engineering of surveying and mapping, 2023, 32(4): 21-28.

[9] 胡娟. 校园三维可视化信息系统的设计与实现[J]. 信息记录材料, 2024, 25(1): 112-114.
HU Juan. Design and implementation of campus 3D visualization information system[J]. Information recording materials, 2024, 25(1): 112-114.

[10] 冯振华, 王博, 蔡文文. BIM 和 SuperMap 三维 GIS 融合的技术探索[C]//第三届全国 BIM 学术会议论文集. 上海, 2017: 153-156.

（编辑：王艳菊）