

# 基于注意力机制改进的疲劳驾驶检测方法<sup>\*</sup>

徐敬一<sup>1</sup>, 田 瑾<sup>1</sup>, 刘 翔<sup>1</sup>, 龚 利<sup>2</sup>

(1. 上海工程技术大学 电子电气工程学院, 上海 201620;

2. 华东师范大学 通信与电子工程学院, 上海 200062)

**摘 要:** 由于疲劳驾驶采集过程中存在因识别角度不佳, 部分区域遮挡等原因, 在不同时间段丢失了不同特征的时间信息, 导致算法的泛用性较差。此外, 驾驶疲劳的检测需要在保证综合准确率的同时, 需要具有更低的漏检率。针对以上问题, 提出了一种基于注意力机制和长短期记忆(LSTM)神经网络的疲劳驾驶检测模型。通过对不同特征定位点计算多维特征向量, 并对特征向量的时间序列进行学习, 同时引入注意力机制, 赋予各维度隐含状态不同的概率权重, 加强重要信息对疲劳状态判定的影响和降低丢失特征信息的历史数据对参数的影响。根据实验可得, 该方法在更普遍的检测环境下有着 92.19% 的准确率和 1.9% 的漏检率, 同时在丢失部分特征的环境下漏检率仅有 3.07%。

**关键词:** 疲劳检测; 特征丢失; 注意力机制; 长短期记忆

中图分类号: TP391

文献标识码: A

文章编号: 1000-9787(2024)04-0115-04

## Improved fatigue driving detection method based on attention mechanism<sup>\*</sup>

XU Jingyi<sup>1</sup>, TIAN Jin<sup>1</sup>, LIU Xiang<sup>1</sup>, GONG Li<sup>2</sup>

(1. School of Electronic and Electrical Engineering, Shanghai University of Engineering and Technology, Shanghai 201620, China;

2. School of Communication and Electronic Engineering, East China Normal University, Shanghai 200062, China)

**Abstract:** Due to the poor recognition angle and partial area occlusion in the process of fatigue driving acquisition, the time information with different characteristics is lost in different time periods, resulting in poor universality of the algorithm. In addition, the detection of driving fatigue needs to not only ensure the comprehensive accuracy, but also have a lower missed detection rate. To solve the above problems, a fatigue driving detection model based on attention mechanism and long short-term memory(LSTM) neural network is proposed. By calculating multi-dimensional feature vectors for different feature localization points and learning the time series of feature vectors, the attention mechanism is introduced to give different probability weights to the hidden states of each dimension, so as to strengthen the influence of important information on the determination of fatigue state and reduce the influence of historical data losing feature information on parameters. According to the experiment, this method has accuracy of 92.19% and missed detection rate of 1.9% in more general detection environment, and the missed detection rate is only 3.07% in the environment where some features are lost.

**Keywords:** fatigue detection; feature loss; attention mechanism; long short-term memory(LSTM)

### 0 引 言

疲劳驾驶是威胁道路交通安全、造成严重伤害、死亡和经济损失的主要因素之一。由于疲劳状态下, 驾驶员从清醒到睡眠的无意识过渡会产生缺乏警觉, 从而使得驾驶性能下降, 将会导致严重的交通事故。目前, 基于视觉的疲劳检测算法<sup>[1]</sup>从单帧视频中提取驾驶员的面部特征, 包括眼部行为<sup>[2,3]</sup>和打呵欠频率<sup>[4]</sup>, 并通过计算时间帧在滑动窗口内的疲劳时间帧比值来判断驾驶员是否疲劳, 或是通过对多

用户特征的综合分析判断疲劳状态<sup>[5]</sup>, 提高单帧的检测精度。此类算法无法获取准确的时序特征, 或是在不同时间段丢失了不同特征的时间信息, 虽然在特定的采集环境中有着不错的准确率, 但泛用性较差, 抗干扰能力弱, 应用性不高。

针对以上问题, 本文提出了一种基于注意力机制和长短期记忆(long short-term memory, LSTM)神经网络的疲劳驾驶多注意力(multi-attention)检测模型。通过对尺度不变

特征变换(scale invariant feature transform, SIFT)算法获取的定位点计算多维特征向量,使用 LSTM 网络学习特征向量的时间序列进行建模学习,同时引入注意力机制赋予各维度隐含状态不同的概率权重,加强重要信息对疲劳状态判定的影响。此模型旨在利用不同结构更有效的分析处理驾驶员面部数据,以更准确地判断驾驶员的疲劳状态变化。实验证明,与其他方法对比,本文所提方法具有更高检测精度和更低的漏检率。

## 1 模型原理

### 1.1 LSTM 神经网络

LSTM 网络<sup>[6]</sup>的主体是一个自我不断循环的神经元结构,随着输入数据的不断增加,这个自循环神经元能将上一次的状态和下一轮的输入一起作为新的输入数据进行当次的学习训练。通过自循环结构神经元,LSTM 网络能实现对时间序列的数据的记忆学习。其网络单元是由存储单元及它们包含的栅极单元组成的特殊结构,主要由遗忘门,输入门和输出门组成。长期状态  $c_{t-1}$  从左向右在网络中传播,依次经过遗忘门时丢弃一些记忆,并通过加法操作从输入门中计算获取并更新一部分记忆。相关函数如下

$$\begin{aligned} f_t &= \sigma(W[x_t, h_{t-1}] + b_f), \tilde{c}_t = f_t \times c_{t-1}, \\ g_t &= \sigma(W[x_t, h_{t-1}] + b_g), \\ i_t &= \tanh(W[x_t, h_{t-1}] + b_c), \\ c_t &= \tilde{c}_t + i_t \times g_t, \\ o_t &= \sigma(W[x_t, h_{t-1}] + b_o), \\ h_t &= \tanh(c_t) \times o_t \end{aligned} \quad (1)$$

其中,  $x_t$  为输入数据序列,  $h_{t-1}$  为前一个神经单元的输出数据,  $c_{t-1}$  为前一个神经单元的存储单元内容,  $b_f, b_g, b_c, b_o$  为偏置向量。  $W$  表示每个输入为独立的向量,  $\sigma$  为 logistic Sigmoid 函数。通过对激活函数的输出  $f_t$  和  $c_{t-1}$  进行元素乘积运算,确定前一个存储单元的各部分内容对 LSTM 的有效程度,若存在过低的值,则该值对应的信息将被舍弃,并更新输出。  $g_t$  通过 Sigmoid 函数计算获取待更新的参数向量,  $i_t$  通过 tanh 函数创建候选向量,并与遗忘门的输出相加,决定该神经单元更新后的长期记忆,最终结果更新保存至神经单元的存储单元  $c_t$ 。  $o_t$  决定哪些特征参数将被添加至短期记忆中。  $h_t$  为输出门参数决定该神经单元输出的短时记忆值。

### 1.2 注意力机制

注意力机制<sup>[7]</sup>是一种模拟了人脑注意力的资源分配机制,其核心思想是巧妙合理的改变对信息的注意力,忽略无关信息并放大所需信息,其结构如图 1 所示。注意力机制通过概率分配的方式,对关键信息分配足够的关注,突出重要信息的影响,从而提高模型的准确率。

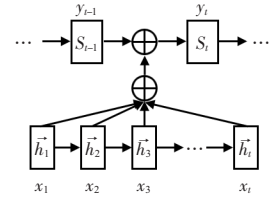


图1 注意力机制结构

## 2 基于 LSTM 网络的疲劳驾驶检测方法

### 2.1 提取各区域的特征矢量

本文首先对视频帧获取 SIFT 特征点作为不同特征的定位点,并整合出 12 个嘴部定位点集,12 个嘴部定位点集和 6 个嘴部定位点集,并计算各自的特征矢量。

1) 计算嘴部特征矢量。嘴部特征包含嘴型区域的 12 个定位点  $P_1 \sim P_{12}$  (如图 2(a)), 使用嘴型纵向长度比横向长度的比值建立嘴型的时序曲,其公式如下

$$\begin{aligned} r_{\text{EAR}} &= (\|P_3 - P_{11}\| + \|P_4 - P_{10}\| + \|P_5 - P_9\|) / 3, \\ W &= \|P_1 - P_7\|, f = H / W \end{aligned} \quad (2)$$

式中  $H$  为嘴的高度,以上下唇的中心 6 点在纵轴方向上的欧氏距离和的平均值表示;  $W$  为嘴的宽度,以嘴角  $P_1$  和  $P_7$  在横轴方向上的距离表示;  $f$  为 1 个视频帧中嘴的宽度和高度的比率,  $f$  的值越大嘴的张开程度越大,反之越小。

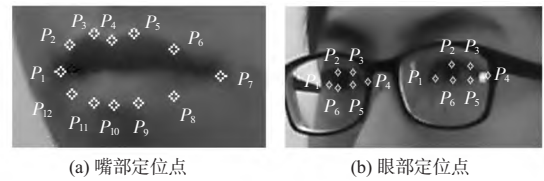


图2 嘴部和眼部定位点图

2) 计算眼部特征矢量。眼部特征由每个眼睛的 6 个定位点决定其特征 (如图 2(b) 所示), 并由纵横比来确定眼睛的状态, 即定义人眼的张合度为  $r_{\text{EAR}}$ , 通过单只眼的 6 个特征点的距离运算获得  $r_{\text{EAR}}$  的值, 其公式如下

$$r_{\text{EAR}} = (\|P_2 - P_6\| + \|P_3 - P_5\|) / 2 \|P_1 - P_4\| \quad (3)$$

通过计算  $P_2$  与  $P_6$ ,  $P_3$  与  $P_5$  的差值, 获取眼睛在纵轴方向上的欧氏距离和, 将  $P_1$  与  $P_4$  的差值作为眼睛在横轴方向上的距离, 其比值  $r_{\text{EAR}}$  为眼睛特征点的纵横比。该值越大表明眼睛越张开; 反之, 则表明眼睛越接近闭合状态。

3) 计算头部特征矢量。选用梁、鼻尖、右眼右眼角、右眼左眼角、左眼左眼角、左眼右眼角 6 个点坐标。通常头部姿态由  $\alpha_{\text{yaw}}$ ,  $\alpha_{\text{pitch}}$  和  $\alpha_{\text{roll}}$  角度来定义。其中,  $\alpha_{\text{yaw}}$  为水平偏移角度,  $\alpha_{\text{pitch}}$  为竖直偏移角度,  $\alpha_{\text{roll}}$  为平面偏移角度, 根据这 3 个参数可得出人脸所在位置的空间角, 继而可以对头部姿态进行较好的估计。因此, 本文以眼部直线的近似向量  $\vec{L}_1$ , 鼻梁的近似向量  $\vec{L}_2$ , 相机的水平向量  $\vec{L}_3$  和相机的垂直向

量  $\vec{L}_4$  作为计算 3 个角度的基准向量

$$\begin{aligned}\vec{L}_1 &= (P_1 + P_2 + P_3 + P_4)/4 - P_1, \vec{L}_2 = P_5 - P_6, \\ \vec{L}_3 &= (1, 0), \vec{L}_4 = (0, 1)\end{aligned}\quad (4)$$

基于以上 4 个基准向量,设定  $\alpha_{yaw}$  为眼部向量与相机水平向量的夹角,  $\alpha_{pitch}$  为鼻梁向量与相机垂直向量的夹角,  $\alpha_{roll}$  为眼部向量与鼻梁向量的夹角,其计算公式如下

$$\begin{aligned}\alpha_{yaw} &= \frac{\vec{L}_1 \cdot \vec{L}_3}{\|\vec{L}_1\| \times \|\vec{L}_3\|}, \alpha_{pitch} = \frac{\vec{L}_2 \cdot \vec{L}_4}{\|\vec{L}_2\| \times \|\vec{L}_4\|}, \\ \alpha_{roll} &= \frac{\vec{L}_1 \cdot \vec{L}_2}{\|\vec{L}_1\| \times \|\vec{L}_2\|}\end{aligned}\quad (5)$$

## 2.2 计算疲劳概率结果

将计算的特征矢量输入至基于注意力机制设计的 LSTM 网络中进行二次分类。第一次分类基于 encoder 模块,通过将输入序列(特征时间序列)输入 LSTM 网络中,获取当前时间步对特征状态的状态概率向量  $h_i$ ,并将该向量压缩成指定长度的向量  $c$ ,公式如下

$$y_i = \text{SoftMax}(h_i), c = q(y_1, y_2, \dots, y_i) \quad (6)$$

第二次分类基于 decoder 分类模块,将接收到的每一个隐含特征计算一个注意力权重,即将计算得到的概率值与隐含特征相乘,其公式如下

$$c_i = \sum_{i=1}^T \alpha_i h_i \quad (7)$$

将注意力权重与语义向量此时间步上的向量值相乘加权后的得到输入向量,作为此模块中当前时间步的输入,其公式如下

$$X_i = C_i \times c_i \quad (8)$$

最后将同时时间帧不同特征判断值和其准确率组成该时间帧的特征向量  $X_i$  输入至一个具有独立参数和层数的 LSTM 中,学习并获得输出序列  $H_i$ 。通过激活函数 SoftMax 对  $H_i$  进行归一化处理,获得当前时间步特征状态的概率向量,计算获得最终的疲劳状态。

## 3 实验结果与分析

### 3.1 实验内容与评估方式

实验选用 NTHU Drowsy Driver Detection (DDD) 数据集<sup>[8]</sup>,该数据集由 18~40 岁之间,具有不同面部特征的男女性成员,在不同场景下完成模拟驾驶的视频构成。本文从中筛选分割出包含嘴部,眼部,头部不同状态的 600 组视频源,每组视频源包含 500 帧的驾驶视频作为本实验的训练和测试数据集和 10 组每组包含 500 帧视频帧的驾驶视频作为验证数据集。根据五折交叉验证的思路,将 600 组视频源随机分成 5 等分,最终训练集、测试集、验证集的数据比例为 48:12:1。本文采用查准率  $P$  和查全率  $R$  作为评估模型的性能的指标,同时选用漏检率作为模型检测能力的额外指标,即

$$P = TP / (FP + TP), R = TP / (TP + FN), \quad (9)$$

漏检率 =  $FN / \text{疲劳总数}$

其中,TP 为判定正确且实际正确的数据数量,FP 为判定正确实际错误的数据数量,FN 为判定错误但实际正确的数据数量。

同时,本文采用额外以上视频源外包含 4 种不同情况合计 120 组 60 帧的视频源(包含 50 组丢失嘴部区域特征,50 组丢失眼部区域特征,10 组丢失头部区域特征和 40 组丢失多个特征区间的视频片段)作为异常片段的对照数据集,每种情况均含有疲劳和非疲劳的视频源,比例为 3:1。对照组采用计算判断正确帧数/总帧数的准确率作为不同模型在特征丢失区间检测能力的判定标准。

### 3.2 模型学习结果

图 3 为网络训练过程中随迭代次数增加,其准确率和损失值的变化。

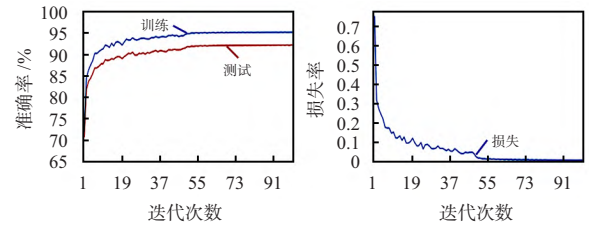


图3 准确率和损失率曲线

由图 3 可以看出,随迭代次数的增加,损失值呈下降趋势,且趋于平缓,网络收敛较快,准确率整体呈上升趋势,在迭代 55 次时进入平稳状态,最终完成训练时,其训练准确率为 95.22%,测试准确率为 92.19%。

### 3.3 模型对比分析

目前常用的 3 类模型进行性能分析,第一类为通过卷积网络获取不同特征区域图块后直接识别的方法<sup>[9]</sup>;第二类为获取关键点后通过 PERCLOS 算法计算疲劳概率的方法<sup>[10]</sup>;第三类为使用仅使用 LSTM 学习判断的方法<sup>[1]</sup>。

Multi-Attention 与上述 3 类模型进行比较分析,实验结果如表 1 所示。

表1 不同网络模型结果对比

	丢失特征区间准确率				查准率	查全率	准确率	漏检率
	眼部	嘴部	头部	多个				
Multi-Attention	94	89	98	5	93.32	98.70	92.19	1.30
CNN <sup>[9]</sup>	65	74	83	32	90.48	90.81	82.89	9.19
PERCLOS <sup>[10]</sup>	80	81	93	35	83.19	86.86	73.89	13.14
LSTM <sup>[1]</sup>	85	87	97	55	91.68	96.37	88.62	3.63

实验表明,引入注意力机制后的模型保留了 LSTM 网络在时间上的学习能力,相比于其他模型能获取更多的时间信息。同时相比于 LSTM 模型在查准率,查全率和准确率上均有提高。

表 2 为验证数据集和对照数据集所有数据判断结果分

类总和,其中对照数据集在“丢失多个特征区间”的数据在所有模型上的判断准确率均较低,不存在参考性并剔除相关结果。由表中数据可得,Multi-Attention 模型未查出的疲劳状态在所有错误中仅占 1.3%,相比于 LSTM 的 3.63%,其漏检率降低了 2.33%,丢失特征区间的疲劳漏检率降低了 7.21%,表明本文方法有更好的疲劳检出能力。

表 2 不同网络模型的错误分布

	验证数据集				对照数据集			
	TP	FP	FN	漏检率/%	TP	FP	FN	漏检率/%
Multi-Attention	276 570	19 786	3 644	1.30	6 221	227	152	3.07
CNN <sup>[9]</sup>	248 688	26 153	25 159	9.19	4 817	901	882	17.81
PERCLOS <sup>[10]</sup>	221 689	44 793	33 518	13.14	5 406	576	618	12.48
LSTM <sup>[11]</sup>	265 861	24 134	10 005	3.63	5 720	371	509	10.28

#### 4 结 论

本文提出了一种基于注意力机制和 LSTM 神经网络的疲劳驾驶检测模型。通过引入注意力机制赋予各维度隐含状态不同的概率权重,增加对有效特征的概率倾斜,加强重要信息对疲劳状态判定的影响,能降低驾驶过程中的疲劳漏检率,对提高驾驶安全有较高的现实意义,未来可以通过捕获更多的特征区间,进一步提高模型对疲劳状态检测的适用性和判断能力。

#### 参考文献:

- [1] DENG W H, WU R X. Real-time driver-drowsiness detection system using facial features [J]. IEEE Access, 2019 (99): 118727–118738.
- [2] 王从澳,黄润才,孙延标,等. 基于特征融合与选择的小样本表情识别[J]. 传感器与微系统,2022,41(10):112–116.
- [3] 冯桑,方涂杰,严楷淳,等. 面向智能驾培系统的深度学习表情识别[J]. 传感器与微系统,2022,41(2):140–143.
- [4] SARADADEVI M, BAJAJ P. Driver fatigue detection using mouth and yawning analysis [J]. International Journal of Computer Science and Network Security, 2008, 8(6): 183–188.
- [5] LAN X Y, YE M, ZHANG S P, et al. Modality-correlation-aware sparse representation for RGB-infrared object tracking [J]. Pattern Recognition Letters, 2020, 130: 12–20.
- [6] YILDIRIM O. A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification [J]. Computers in Biology & Medicine, 2018, 96(5): 189–202.
- [7] 赵兵,王增平,纪维佳,等. 基于注意力机制的 CNN-GRU 短期电力负荷预测方法[J]. 电网技术, 2019, 43(12): 4370–4376.
- [8] SHIH T H, HSU C T. MSTN: Multistage spatial-temporal network for driver drowsiness detection [C] // Asian Conference on Computer Vision, 2016: 146–153.
- [9] 刘明周,蒋倩男,扈静. 基于面部几何特征及手部运动特征的驾驶员疲劳检测[J]. 机械工程学报, 2019, 55(2): 18–26.
- [10] 耿磊,袁菲,肖志涛,等. 基于面部行为分析的驾驶员疲劳检测方法[J]. 计算机工程, 2018, 44(1): 274–279.

#### 作者简介:

徐敬一(1996—),男,硕士研究生,研究方向为疲劳检测,计算机视觉。

田 瑾(1982—),女,通讯作者,博士,副教授,研究领域为计算电磁学,异构并行计算。

刘 翔(1972—),男,博士,副教授,研究领域为医学图像分析,模式识别。

龚 利(1981—),男,博士,副教授,研究领域为能源政策。

#### (上接第 114 页)

- [2] USUNG P, RHIM J, JUEON J U, et al. A micromachined differential resonant accelerometer based on robust structural design [J]. Microelectronic Engineering, 2014, 129(5): 5–11.
- [3] LEFORT O, THOMAS I, JAUD S. To the production of a robust and highly accurate MEMS vibrating accelerometer [C] // 2017 DGON Inertial Sensors and Systems, Karlsruhe, Germany, Institute of Electrical and Electronics Engineers, 2017: 1–19.
- [4] JOHNSON B, ALBRECHT C, BRAMAN T, et al. Development of a navigation-grade MEMS IMU [C] // 2021 IEEE International Symposium on Inertial Sensors & Systems, Kailua Kona, HI, USA, IEEE, 2021: 1–4.
- [5] FANG Z X, YIN Y G, CHEN C, et al. A sensitive micromachined resonant accelerometer for moving-base gravimetry [J]. Sensors and Actuators A, 2021, 325: 112694.
- [6] HUANG J, ZHAO Y, XIA G M, et al. Systematic modeling of a MEMS resonant accelerometer based on displacement coordination [J]. IEEE Sensors Journal, 2022, 22(7): 6454–6465.
- [7] 范书聪,张基强,黄丽斌,等. 硅微谐振加速度计电路宽带测试方法研究[J]. 传感器与微系统, 2020, 39(8): 20–22.
- [8] DING H, WU C J, XIE J. A MEMS resonant accelerometer with high relative sensitivity based on sensing scheme of electrostatically induced stiffness perturbation [J]. Journal of Microelectromechanical Systems, 2021, 30(1): 32–41.
- [9] WANG C, HU Q F, WANG Y, et al. Structure design and simulated optimization of silicon resonant accelerometer [J]. Navigation and Control, 2016, 15(1): 41–46.
- [10] ZHANG J, SU Y, SHI Q, et al. Microelectromechanical resonant accelerometer designed with a high sensitivity [J]. Sensors, 2015, 15(12): 30293–30310.

#### 作者简介:

高乃坤(1991—),男,博士,高级工程师,研究领域为 MEMS 惯性器件技术。