

基于自适应阈值的多特征经验融合疲劳检测^①



王寰希, 张德平

(南京航空航天大学 计算机科学与技术学院, 南京 210016)

通信作者: 张德平, E-mail: depingzhang@nuaa.edu.cn

摘要: 在安防领域, 疲劳是导致安防工作人员注意力下降, 诱发各类问题的重要原因. 现有的疲劳检测方法存在各种问题, 比如生理指标检测侵入性强且价格昂贵, 人脸疲劳检测结果受个体差异和头部姿态的影响以及疲劳预警时长较短等, 因此本文提出一种基于自适应阈值和面部多特征经验融合的疲劳早后期检测算法, 使用轻量级的 SCRFD 模型进行人脸检测, 使用 MobileNetV2 模型进行人脸关键点定位, 使用梯度提升树学习头部姿态信息与眼睛纵横比 (*EAR*) 阈值的映射关系, 通过眼睑闭合时间百分比 (*PERCLOS*)、嘴巴张开时间百分比 (*FOM*) 和头部姿态 6 个自由度分别实现眨眼、哈欠、点头前后倾动作的识别. 在疲劳估计阶段, 为了将多种疲劳行为融合映射成与疲劳相关的 KSS 值, 先根据专家经验预先构建好多种人脸行为的疲劳因果图, 接着使用自定义的 singleton, mutual 和 activate/inhibit 特征算子, 结合因果图从人脸行为检测序列中计算疲劳早期和疲劳后期 KSS 值, 最后使用双尺度 KNN 实现疲劳早后期估计. 实验结果表明所提算法在 YawDD 数据集上哈欠检测准确率达到 93.81%, 在 UTA-RLDD 和 Drozy 数据集上疲劳识别准确率分别达到 67.72%, 87.88%, 仅通过 CPU, 推理实时性可达到 17.96 每秒传输帧数 (FPS).

关键词: 自适应阈值; 眨眼检测; 经验融合; 疲劳因果图; 疲劳早后期估计; 多特征融合

引用格式: 王寰希, 张德平. 基于自适应阈值的多特征经验融合疲劳检测. 计算机系统应用, 2023, 32(4): 197-205. <http://www.c-s-a.org.cn/1003-3254/9067.html>

Multi-feature Empirical Fusion Fatigue Detection Based on Adaptive Threshold

WANG Huan-Xi, ZHANG De-Ping

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: In the field of security, fatigue is an important reason for the decrease in the attention of security staff and can cause various problems. The existing fatigue detection methods, however, have many problems, such as the strong invasiveness and high costs of physiological indexes for detection, the impact of individual differences and head poses on the facial fatigue detection results, and the short early-warning time of fatigue. Therefore, this study puts forward an early and late fatigue detection algorithm based on the adaptive threshold and the empirical fusion of multiple facial features. In this algorithm, the lightweight SCRFD model is used for face detection, and the MobileNetV2 model is used to locate the key points of the face. The gradient boosting decision tree (GBDT) is applied to learn the mapping relationship between the head pose information and the eye aspect ratio (*EAR*) threshold, and the recognition of movements such as blinking, yawning, and leaning forward and backward is achieved through the six degrees of freedom of the percentage of eyelid closure over the pupil time (*PERCLOS*), the frequency of opening mouth (*FOM*), and head pose, respectively. In the fatigue estimation stage, various fatigue behaviors should be fused and mapped into fatigue-related KSS values. Hence, the fatigue causality maps of various facial behaviors are constructed in advance according to expert experience. Then, the

① 基金项目: 国防基础科研基金 (JCKY2020605C003)

收稿时间: 2022-09-25; 修改时间: 2022-10-19, 2022-11-15; 采用时间: 2022-11-18; csa 在线出版时间: 2023-03-01

CNKI 网络首发时间: 2023-03-02

self-defined singleton, mutual, and activate/inhibit feature operators are used to calculate the KSS values at the early and late stages of fatigue from the facial behavior detection sequences in combination with the causality maps. Finally, the dual-scale KNN is employed to realize the early and late fatigue estimation. The experimental results show that the yawn detection accuracy of the proposed algorithm reaches 93.81% on the YawDD dataset, and the fatigue detection accuracy of the algorithm is 67.72% and 87.88% on UTA-RLDD and Drozy datasets, respectively. The real-time inference performance can reach 17.96 frames per second (FPS) only through the CPU.

Key words: adaptive threshold; blink detection; experience fusion; fatigue causality diagram; early and late fatigue estimation; multi-feature fusion

在安防领域,特别是在航空航天、长途运输、大型服务器监管等行业,一旦操作员在精神疲劳状态下工作,将会使整个系统处于极其危险的境地.疲劳对于工作生产的安全性,尤其是对于汽车驾驶,高危行业的监督和调度等工作具有很大的影响.

对于疲劳识别方法,主要有接触式和非接触式两类.对于接触式方法,基于脑电信号,心电信号,眼电信号等生理信号分析方法虽然精度大,但成本高应用面窄,目前主要应用在卫生医疗领域.非接触式方法主要通过摄像头采集的人脸图像进行分析,该方法成本低应用范围广,已经成为该领域的研究热点. Sigari 等人^[1]使用 Haar-like 特征检测图像中的人脸,将 PERCLOS (眼睑闭合时间百分比), EDLC (上下眼睑距离), CLOSNO (闭眼频率), ROT (头部旋转) 特征作为模糊逻辑系统的输入,进行疲劳状态分析. Mandal 等人^[2]专门为公交车司机设计了一套疲劳检测系统,该系统使用睁眼次数和 PERCLOS 作为疲劳检测特征,通过梯度直方图 (HoG) 检测头部和肩部,使用支持向量机 (SVM) 检测人脸疲劳状态. Fei 等人^[3]在 YawDD 数据集^[4]上,将全局人脸图片,人脸关键点提取的人眼嘴巴区域以及该区域的局部图片作为 CNN 模型的输入,使其学习人脸表征,接着使用长短时记忆网络 (LSTM) 对人脸表征进行疲劳分析.

这些疲劳检测算法往往都存在对疲劳早期状态检测难度大,疲劳检测结果受个体差异影响大的缺点.因此本算法为了解决以上存在的问题,使用人脸校准和头部姿态信息自适应调整 EAR, MAR 阈值,分别进行眨眼和哈欠检测,使用多特征经验融合模型中自定义的 singleton, mutual 和 activate/inhibit 特征算子融合疲劳早后期 KSS 值,使用双尺度 K 近邻模型 (KNN) 实现疲劳状态的层级推理.

1 相关概念与技术

1.1 人脸检测和关键点定位

SCRFD 模型^[5]是轻量级的人脸检测模型,其通过样本再分配以及单阶段检测器中各组件的算力重分配策略,利用自动化网络搜索算法构建效能最优的人脸检测模型.

MobileNetV2 模型^[6]保留了 MobileNetV1 深度可分离卷积的优点,借鉴了 ResNet 残差连接的思想,利用翻转残差模块和线性瓶颈结构,有效避免模型训练存在梯度消失的问题.

本算法使用开源的 SCRFD-0.5GF 和 MobileNetV2-56 模型文件,分别进行人脸检测和人脸关键点定位,其中 68 个人脸检测点如图 1 所示.

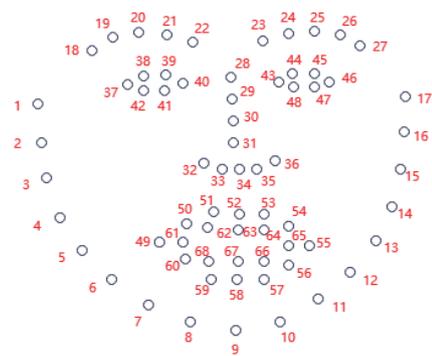


图 1 人脸 68 个关键点

1.2 梯度提升树

梯度提升树 (gradient boosting decision tree, GBDT) 是一种集成学习算法^[7],由多个决策树构成的弱分类器进行联合决策,其模型如式 (1) 所示.其中 K 为树的总个数、 f_k 为第 k 棵树、 y_i 为样本 x_i 的预测值:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (1)$$

GBDT的关键是采用最速下降法的近似方法,利用损失函数的负梯度作为提升树算法中的残差近似值,用来每次迭代拟合一个理想的回归函数. GBDT常用参数包括弱学习器的个数($n_estimators$)和每个弱学习器的学习率,通过调参可以在回归任务上取得不错的泛化性能.

1.3 疲劳行为特征

常见的疲劳行为特征有3类:基于眨眼行为的EAR和PERCLOS特征,基于哈欠行为的MAR和FOM特征,基于点头行为的6个头部姿态自由度特征.对于眨眼和哈欠检测,使用68个人脸关键点计算相应疲劳特征是有效的手段.

常见的眨眼检测方法一般是通过比较EAR(式(2))和EAR阈值得到^[8].对于快慢眨眼的识别,通过PERCLOS(式(3))和阈值得到^[9].在图1所示的人脸关键点中,左眼EAR计算过程如式(2)所示,在通过EAR特征和EAR阈值检测出眨眼行为之后,计算滑动时间窗口内眨眼帧数的占比,得到PERCLOS特征,计算过程如式(3)所示:

$$EAR = \frac{\|P_{38} - P_{42}\| + \|P_{39} - P_{41}\|}{2 \times \|P_{40} - P_{37}\|} \quad (2)$$

$$PERCLOS = \frac{\text{num_of_frames_with_blinks}}{\text{total_frames_of_windows}} \quad (3)$$

常见的哈欠检测方法是通过对比较MAR和MAR阈值得到^[10],在图1所示的人脸关键点中,MAR计算过程如式(4)所示.但是研究表明,仅通过MAR阈值的哈欠检测容易把唱歌和正常说话识别为打哈欠,因此引入了FOM特征^[11],用于计算滑动时间窗口内哈欠帧数的占比,计算过程如式(5)所示:

$$MAR = \frac{\|P_{51} - P_{59}\| + \|P_{53} - P_{57}\|}{2 \times \|P_{55} - P_{49}\|} \quad (4)$$

$$FOM = \frac{\text{num_of_frames_with_yawn}}{\text{total_frames_of_windows}} \quad (5)$$

头部姿态可用3个旋转自由度pitch, yaw, roll和3个平移自由度来表示,本文采用EPNP方法^[12]求解头部姿态问题.在式(6)中, R_x 、 R_y 、 R_z 是3个基本旋转矩阵,表示在三维空间中用右手法则将向量分别绕x、y、z轴旋转一个 θ 角,为了得到具有一般性的旋转矩阵,可以通过 R_x 、 R_y 、 R_z 做矩阵乘法得到R; t 表示关于x、y、z轴的平移向量,平移分量分别用 t_x 、 t_y 和 t_z 表示(式(7)); M 是相机的内参矩阵(式(8)), f_x 和 f_y 分别表示x和y方向上的焦距,(c_x , c_y)表示光学中

心.假设点 $P(U, V, W)$ 为世界坐标下的点, $p(x, y)$ 为图像坐标下的点,它们之间的转换关系如式(9)所示,其中 s 是一个未知比例因子,用于调节图像深度.PNP问题就是在已知世界坐标下的 P 点和对应平面坐标系中的 p 点的情况下,求解出式(9)中的 R 和 t 的过程.相机的内参矩阵 M 可使用棋盘标定算法计算得到^[13].

通过对头部3个旋转自由度的一阶差分设置阈值,可以进行点头和摇头检测.关于Z轴平移自由度的一阶差分设置阈值,可以进行前后倾运动检测.

$$\begin{cases} R = R_x(\theta)R_y(\theta)R_z(\theta) \\ R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \\ R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \\ R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{cases} \quad (6)$$

$$t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (7)$$

$$M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

$$\begin{cases} [R|t] = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \\ \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = sM[R|t] \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \end{cases} \quad (9)$$

2 多特征经验融合的疲劳识别模型

2.1 自适应阈值方法

EAR, MAR的优点是该特征受人脸距离影响较小,缺点是其特征受头部姿态3个旋转自由度影响较大,以及个体差异的影响.如果仅使用单一阈值^[7]或者波峰波谷算法^[14]对行为进行判定,容易出现漏检问题.因此本文利用头部姿态信息自适应调整EAR, MAR阈值,提高眨眼和哈欠行为检测.

2.1.1 自适应EAR阈值的二次校准方法

为了解决头部姿态对眨眼检测效果的影响,本文

先是采集了 25 个不同头部姿态下的人脸视频, 并通过这些视频采集每个视频帧中人脸头部姿态 3 个自由度以及眨眼 EAR 值, 构建 BlinkHeadpose 眨眼数据集. 接着使用梯度提升树 (GBDT) 模型, 学习头部姿态与自动标注好的 EAR 阈值之间的映射关系. 由于不同人在眼睛特征、眨眼速度上存在差异, GBDT 根据头部姿态预测出的 EAR 阈值 (记为 EAR_thres_{gbdt}) 需要进行二次校准, 在给定好数据集中受试者 A 的正脸基准 EAR 值 (记为 EAR_basis) 之后, 关于受试者 B 的 EAR 阈值校准方法如算法 1.

算法1. 自适应 EAR 阈值的二次校准方法

- 1) 采集时间窗口内关于受试者 B 的人脸区域的多个视频帧.
- 2) B 与正常人脸在左眼水平参考线、右眼水平参考线、鼻梁垂直参考线和嘴巴水平参考线相交情况, 判断待校准图片序列中的人脸是否处于正脸状态, 如果没有一帧符合正脸要求的, 则会提醒受试者摆正头部姿态, 直至采集到符合要求的人脸图片为止.
- 3) 在符合要求的正脸帧序列中, 找到最大的 EAR 并将其除以 EAR_basis , 得到关于受试者 B 的相较于受试者 A 正脸的初始 EAR 阈值比例 δ .
- 4) 最后通过 $\delta \times EAR_thres_{gbdt}$ 得到受试者 B 此时头部姿态下的 EAR 阈值.

2.1.2 自适应 MAR 阈值方法

由于头部旋转会影响 MAR 值变化, 因此利用头部姿态中的 yaw 定义 MAR 自适应阈值函数如式 (10) 所示: 初始阈值 MAR_thres 设置为 0.6, 该函数适用于 YawDD 数据集^[4] 的哈欠检测, 结合 FOM 特征, 哈欠检测准确率可达到 93.81%. 对于其他机位情况, 需要重新设置式 (7) 中 yaw 区间及对应的缩放因子.

$$MAR_thres = \begin{cases} MAR_{init} \times 1.3, & yaw < -6 \\ MAR_{init} \times 1.2, & -6 \leq yaw \leq -3 \\ MAR_{init} \times 1.0, & yaw > -3 \end{cases} \quad (10)$$

2.2 多特征经验融合算法

疲劳程度最常使用 KSS 量表进行定义. KSS 量表由 Åkerstedt 等人提出^[15], 是一种基于困倦感的测量量表, 采用 1-9 级评分, 本算法将疲劳早期和疲劳后期按 KSS 值分类如表 1 所示.

表 1 根据 KSS 值划分疲劳类别

疲劳程度	警觉	疲劳早期	疲劳后期
KSS值	1, 2, 3, 4, 5	6, 7	8, 9

对于综上检测到的人脸行为, 包括眨眼, 哈欠, 点头和前后倾, 为了将多种疲劳行为的检测结果进行融合, 本算法为每个人脸行为进行编码 (如表 2 所示), 并根据经验为每种编码赋予一个先验 KSS 值 (如图 2 所示), 比如慢眨眼行为是疲劳后期的主要检测标准, 因

而 KSS 值设置为 8; 头部前后倾运动意味着坐立不安, 此时大概率存在困意倾向, KSS 值设置为 6. 为了提高对疲劳早期状态的判断, 本文在对非明显疲劳行为设置 KSS 时保证其 KSS 不小于 4.

表 2 疲劳检测迹象编码

面部特征	面部行为编码	
	面部行为	编码
眼睛	快眨眼	e1
	慢眨眼	e2
	正常	e3
嘴巴	哈欠	m1
	正常	m2
头部姿态	点头	h1
	前后倾	h2
	正常运动	h3
	静止	h4

为了对疲劳早后期状态进行识别, 本算法使用短时间窗口 (默认 30 帧) 进行疲劳后期检测, 使用长时间窗口 (默认 240 帧) 进行疲劳早期检测.

本算法利用不同人脸行为组合来识别疲劳后期和疲劳早期. 为了提高识别准确度, 本算法提出了多特征经验融合模型, 根据经验预先构建好关于不同人脸行为组合的因果图, 通过引入 3 个 KSS 特征算子: singleton, mutual 和 activate/inhibit 算子, 计算这些人脸行为组合的 KSS 值.

根据经验, 我们可知慢眨眼、点头和打哈欠是明显疲劳迹象, 适合于疲劳后期检测; 而对于头部前后倾行为, 虽然不能直接反映真实疲劳状态, 但可用于疲劳早期的检测.

在对多特征融合模型进行经验建模之前, 需要提出两点要求: 1) 随着人脸疲劳行为编码 h1, h2, e2, e3, 统计次数的增加, KSS 值提高; 2) 随着人脸疲劳行为编码 h3, e1 统计次数的增加, KSS 值降低. 因此对每个算子的疲劳组合设计如算法 2 所示.

算法2. 对每个算子的疲劳组合设计

- 1) singleton算子只针对明显疲劳迹象: h1, m1, e2;
- 2) mutual算子只针对人脸微行为: [h2, e3], [h4, e3];
- 3) activate算子只对singleton算子起激活作用, 起到激活作用的疲劳行为编码为: h2, e3
- 4) inhibit算子只对mutual算子起抑制作用, 起到抑制作用的疲劳行为编码为: h3, e1

考虑到人的嘴巴张合习惯不同, 而且嘴巴正常行为对疲劳早期判断影响较小, 因此并没有将 m2 进行疲劳组合设计. 当进行多特征经验融合时, 先分别对短时

窗口内明显疲劳迹象和长时窗口内疲劳早期迹象进行频数统计,再使用 Softmax 函数对 KSS 值和频数进行归一化.为了解释方便,这里分别将 singleton、mutual 和 activate/inhibit 算子重命名为 f_1 、 f_2 、 f_3 和 $-f_3$ 算子,如图 3 所示,各函数公式如式(11)–式(13)所示.其中 Kn_{code_i} 和 Cn_{code_i} 分别表示某个疲劳行为编码 $code_i$ 的 KSS 归一化值和频数归一化值, α 因子可调节 KSS 值增加的加速度,控制 KSS 值的取值范围; tanh 激活函数通过 β 因子调节疲劳行为组合的激活值. f_3 算子对 f_1 算子计算的 KSS 值起激活作用, $-f_3$ 算子对 f_2 算子计算的 KSS 值起抑制作用.

完成多特征经验融合之后,使用两个 KNN 模型对 UTA-RLDD 数据集中提取的疲劳早后期分组统计数据训练,实现疲劳早期和疲劳后期的分类.在疲劳推理阶段,则使用短时 KSS 窗口和长时 KSS 窗口作为输入,使用之前训练好的两个 KNN 模型进行层级疲劳估计.

基于面部多特征经验融合的 K 近邻疲劳检测算法具体步骤如算法 3 所示.

算法3. 多特征经验融合方法

- 1) 使用轻量级SCRFD-0.5GF模型,对采集的视频流进行人脸检测.
- 2) 使用MobileNetV2模型提取68个关键点.
- 3) 利用EPNP方法计算头部姿态6个自由度,利用一阶差分 and 阈值判断,识别点头,头部前后倾,摇头和静止行为.
- 4) 通过68个人脸关键点以及校准图片,计算EAR和MAR,通过自适应阈值的判断,识别快眨眼,正常眨眼,慢眨眼,打哈欠等动作.
- 5) 对第3步和第4步所提取的疲劳行为特征进行编码,并为每个编码设置KSS值.
- 6) 根据人为经验预先构建好关于人脸疲劳行为的因果图,使用自定义的singleton, mutual和activate/inhibit算子,融合此刻所检测到的各种疲劳行为的KSS值,得到早后期疲劳KSS序列.
- 7) 使用长短时KNN对早后期疲劳KSS序列进行学习,实现疲劳早后期估计.

3 实验

3.1 实验平台和数据采集

实验平台使用 Windows 10 操作系统,处理器为 Intel(R) Core(TM) i7-6820HQ CPU,主频 2.70 GHz,内存 16 GB, GPU 为 NVIDIA Quadro M1000M.

本文所使用的疲劳检测数据集主要分为两种:一种是用于人脸疲劳行为检测的数据集,如用于眨眼检测的自建数据集 BlinkHeadpose、MINE_Blink 和 MINE_Blink2,用于哈欠检测的 YawDD;第二种是用于人脸疲劳状态检测的数据集,比如 Drozy 和 UTA-RLDD 数据集.

$$f_1 = \alpha \times Kn_{code_i} \times Cn_{code_i} \tag{11}$$

$$f_2 = \tanh \left(\beta \sum_j (Kn_{code_j} \times Cn_{code_j}) \right) + \alpha \text{Max}_j (Kn_{code_j} \times Cn_{code_j}) \tag{12}$$

$$f_3 = \tanh \left(\beta \sum_k (Kn_{code_k} \times Cn_{code_k}) \right) \tag{13}$$

BlinkHeadpose 数据集包含 1 个受试者多机位、多个头部姿态的视频,人脸主要位于摄像头的左上角、右上角、中央、左下角和右下角 5 个区域,而头部姿态类型包括正对摄像头、抬头、点头、向左旋转和向右旋转,总共 25 个视频.在保存每个视频时,系统会自动保存每一帧的头部姿态 3 个旋转自由度以及 EAR 值,眨眼行为作为标签需要结合视频进行手动标注,其中 0 表示未眨眼,1 表示眨眼,部分样本如表 3 所示.经过相应的数据处理之后 BlinkHeadpose 可用于 GBDT 模型的训练.

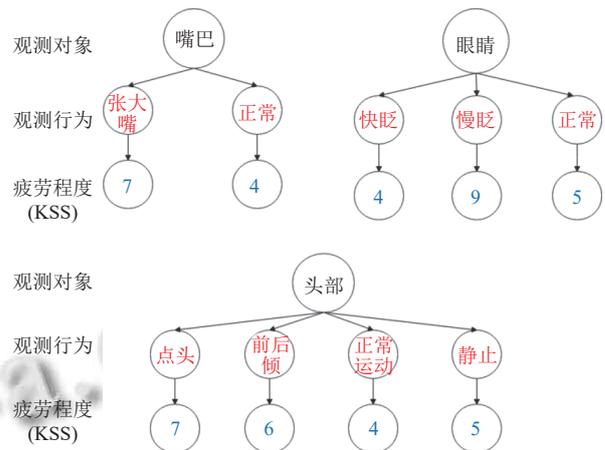


图 2 关于各疲劳迹象的先验 KSS 值

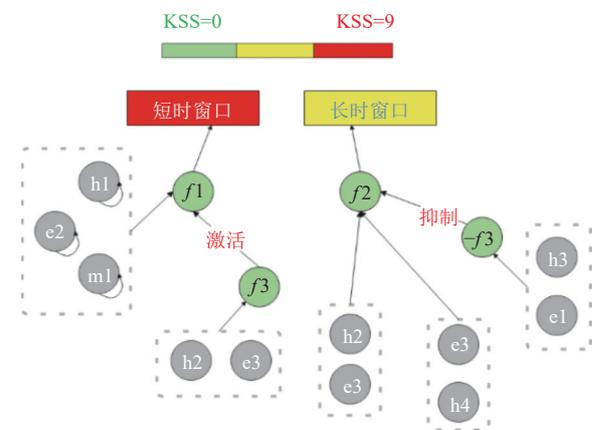


图 3 关于特征经验融合模型的 3 个算子

MINE_Blink 数据集包含 3 个受试者, 利用本地摄像头采集 120 个视频, 每个视频约 10 s, 眨眼次数按情况划分如表 4 所示。

表 3 BlinkHeadpose 部分样本数据

yaw	roll	EAR	blink
24.796211	4.200516	0.215333851	0
24.71404	4.052315	0.214905644	0
24.861328	3.852618	0.22275138	0
24.493619	4.531789	0.06571097	1
24.474487	4.370588	0.087326656	1
24.76079	3.8162	0.194849718	0
24.602299	4.476883	0.214165052	0
25.222297	3.664019	0.235837874	0

表 4 MINE_Blink 眨眼检测数据集

分类	眨眼次数
戴眼镜/不戴眼镜	376/352
快眨眼/慢眨眼	504/224
头部正/上/下/左/右	147/151/151/134/145

由于 MINE_Blink 只采集了 3 个受试者的人脸视频, 为了更好验证本文所提方法在多个受试者上的眨眼检测有效性, 本文从 UTA-RLDD, YawDD 中共抽取 10 个受试者构建了 MINE_Blink2 数据集, 并对每个视频进行眨眼次数统计。

YawDD 数据集^[4]包含 34 个参与者, 每个参与者均完成正常闭嘴、说话和打哈欠行为, 共 322 个侧脸视频。但有些视频存在两种行为标签, 因此本实验的哈欠检测并没有使用这些视频进行验证, 实际参与评估的视频个数为 307 个。

Drozy 数据集^[16]取自 14 名年轻的健康参与者, 每个参与者需要在 3 个不同时间段 (起床后 3 h, 起床后 21 h, 起床后 29 h) 分别进行长达 10 min 的精神运动警觉测试 (PVTs)。在该数据集中, 我们仅抽取 14 个受试者, 共 33 个视频进行模型验证, 其中非疲劳视频个数为 13 个, 疲劳视频个数为 20 个。

UTA-RLDD 数据集^[9]包含 60 名参与者, 每个参与者在不同实验环境中用手机摄像头记录自己 3 种不同疲劳状态, 状态分别为警醒 (alert), 欠警醒 (low vigilant) 和嗜睡 (drowsy)。我们仅抽取 24 个受试者, 共 72 个视频进行模型验证。

3.2 实验分析

3.2.1 眨眼行为模型的训练和验证

为了让 GBDT 学习到 BlinkHeadpose 数据集中 3 个旋转自由度与 EAR 阈值之间的关系, 通过大量实

验, 选择将视频眨眼时 (前一帧标签为 0, 后一帧标签为 1) 的 EAR 序列的 85% 分位数作为该视频的 EAR 阈值。从 25 个视频中可以提取出 2085 个眨眼样本, 样本可表示为 (pitch, yaw, roll, EAR_thres), 其中 EAR_thres 为对应视频中自动标注的 EAR 阈值。

由于训练样本数较少, 实验通过 10 折交叉验证对 GBDT 的 $n_estimators$ 参数进行优化, 在参数的不同取值下, 均方误差和平均耗时归一化结果如图 4 所示, 其中横坐标表示 $n_estimators$ 的取值, 纵坐标表示不同参数下的归一化结果。为了平衡模型的准确率和时间开销, 最终将该参数确定为 125。树的深度和 GBDT 的学习率分别设置为 4 和 0.01。在推理阶段则使用 EAR 阈值二次校准方法校准当前受试者的 EAR 阈值, 实现眨眼检测。

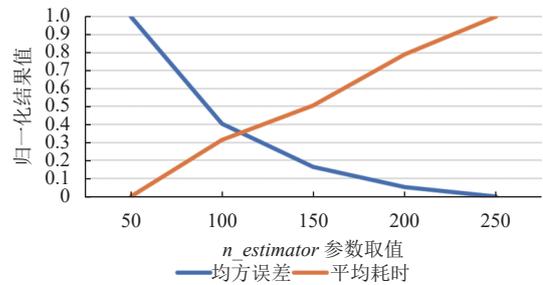


图 4 设置不同 $n_estimators$ 参数的效果比较

本实验通过 MINE_Blink, MINE_Blink2 对自适应 EAR 阈值二次校准方法进行评估, 分析眨眼的检测准确度。关于眨眼检测准确率如式 (14) 所示, 其中 $bCount$ 为该视频的实际眨眼次数, $pCount$ 为眨眼预测次数。

MINE_Blink1 眨眼检测结果如表 5 所示, 其中方法 1、方法 2、方法 3 分别表示为: 设置 EAR 阈值为 0.25^[8]、波谷计算算法^[14] (骤减阈值为 0.04) 和本文提出的自适应 EAR 阈值二次校准算法。实验结果表明, 本方法预测的 EAR 阈值在 MINE Blink 眨眼检测上, 相较于固定阈值方法和波谷计算算法, 具有更好的检测准确度; 在快眨眼和慢眨眼检测上, 虽然快眨眼检测准确率略低于波谷计算算法, 但是在慢眨眼检测上却高出了约 20%。

$$\begin{cases} acc_{mean} = \frac{\sum_{i=1}^M acc_i}{M} \\ acc_i = \begin{cases} \frac{bCount}{pCount}, & bCount < pCount \\ \frac{pCount}{bCount}, & bCount \geq pCount \end{cases} \end{cases} \quad (14)$$

MINE_Blink2 眨眼检测结果如图 5 所示, 实验结果表明: 1) 基于二次校准的自适应 EAR 阈值方法在进行眨眼检测上均优于波谷算法; 2) EAR 阈值二次校准算法在受试者 3 和 5 中表现得最好, 在其他受试者中均和其他算法一样, 存在不同程度的眨眼漏检和误检的情况。

表 5 MINE_Blink1 眨眼检测算法评估 (%)

方法	戴眼镜/不戴眼镜	快眨眼/慢眨眼	头部正/上/下/左/右
方法1	57.43/58.58	54.09/61.84	63.87/22.23/54.82/ 87.47/62.84
方法2	45.44/70.05	66.84/48.60	52.81/75.34/58.47/ 52.17/49.21
方法3	56.84/78.58	68.65/66.60	43.18/77.92/79.80/ 63.84/73.19

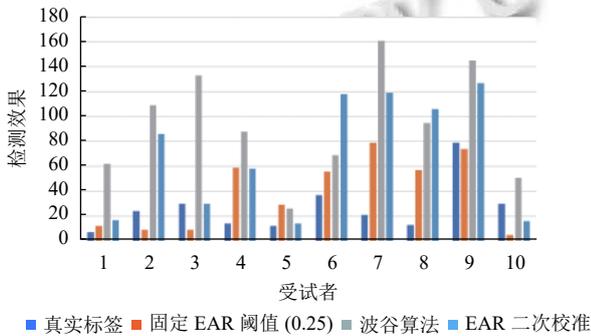


图 5 MINE_Blink2 在不同眨眼方法中的检测效果对比

在进行快眨眼和慢眨眼识别中, 仅通过 PERCLOS 就可以得到很好的效果. 本算法通过设置 PERCLOS 上下两阈值进行快眨眼, 正常眨眼和慢眨眼的识别. 当 $PERCLOS < 0.3$ 时, 表示快眨眼; 当 $0.3 \leq PERCLOS \leq 0.5$ 时表示正常眨眼; 当 $PERCLOS > 0.5$ 时表示慢眨眼。

3.2.2 哈欠行为模型的验证

对于哈欠行为检测, 本算法分成两个阶段: 第 1 阶段是使用 MAR 和 MAR 阈值进行初步的哈欠检测; 第 2 阶段使用 FOM 和 FOM 阈值对第 1 阶段哈欠检测结果进行校验. 本实验通过 YawDD 数据集对自适应 MAR 阈值算法进行评估, 其中正常视频和哈欠视频个数占比为 2:1。

在实验过程中, 使用自适应 MAR 阈值进行哈欠检测, 效果如图 6、图 7 所示. 由图 6 可知, MAR 值在区间 [100, 200] 上增加的主要原因是 yaw 值的变化, 而非嘴巴哈欠引起的, 本算法利用 yaw 值提高 MAR 阈值, 进而避免哈欠行为的误检; 对于图 7, MAR 值变化的主

要原因是嘴巴哈欠引起的, 同时又伴随着 yaw 值的缓慢变化, 该图中 MAR 值大于 MAR 阈值, 因此被初步检测为打哈欠。

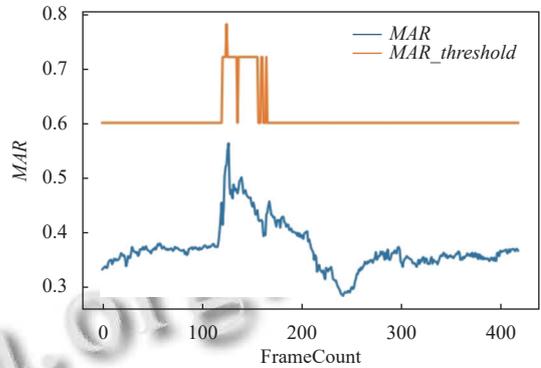


图 6 yaw 值引起 MAR 值变化的曲线图

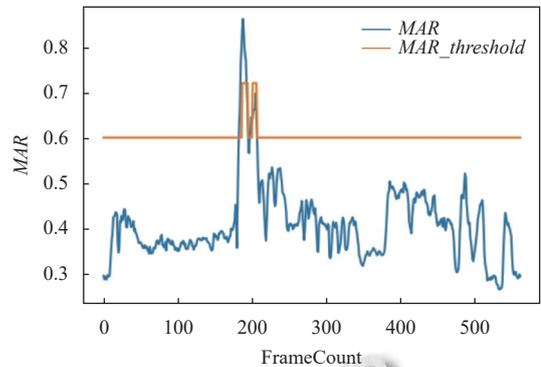


图 7 哈欠引起 MAR 值变化的曲线图

为了确定 FOM 阈值参数, 先采集每个视频中最大的 FOM 值, 接着按嘴巴行为类别绘制箱线图. 由图 8 可知, 当 FOM 阈值设置为 0.6 时, 可以有效划分正常闭嘴和哈欠行为。

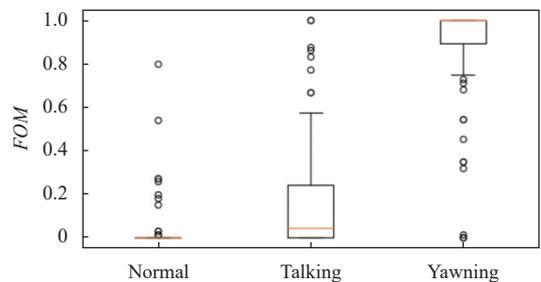


图 8 YawDD 在正常, 说话和哈欠下的 FOM 箱线图

不同检测结果如表 6 所示, 其中方法 1 表示 MAR 阈值和 FOM 阈值均设置为 0.6^[17]; 方法 2 表示使用 ResNet 和 LSTM 算法进行哈欠检测^[3]; 方法 3 表示 MAR 阈值

和 FOM 阈值均设置为 0.6, 并使用头部姿态函数修正 MAR 阈值. 实验发现使用头部姿态函数自适应调整 MAR 阈值可以提高 YawDD 哈欠检测的精确度.

表 6 YawDD 哈欠检测 (%)

方法	正常 P/R	哈欠 P/R	平均准确率
方法1	94.87/90.24	82.14/90.19	90.23
方法2 ^[3]	—	—	94.9
方法3	95.14/95.60	91.09/90.19	93.81

在 YawDD 数据集上, 对哈欠的平均检测准确度能达到 93.81%, 虽然低于文献 [3] 算法, 但在计算开销上, 仅通 MAR 阈值和 FOM 特征更适合于实时的哈欠检测.

表 7 疲劳检测各环节平均耗时

序号	疲劳检测环节	平均耗时 (ms)
1	人脸定位	24.9414
2	人脸关键点检测	8.2230
3	眨眼特征提取	1.1333
4	头部姿态估计	0.9739
5	哈欠检测	0.0189
6	多特征经验融合	16.8461

3.2.3 疲劳检测模型的比较

由表 7 可发现, 疲劳状态检测系统的耗时主要集中在人脸定位, 人脸关键点检测以及多特征经验融合上. 仅通过 CPU 进行推理, 对单个视频帧的疲劳状态分析大约耗时为 0.0556 s, 约 17.96 FPS.

本算法在 UTA-RLDD、DROZY 数据集上进行了验证, 模型性能比较如表 8 所示. 其中对于 UTA-RLDD 数据集, 该模型的检测准确率能达到 67.27%, 相较于 HM-LSTM^[9] 高出 2.7%; 仅通过 UTA-RLDD 数据集训练好的短时 KNN 和长时 KNN 对 DROZY 数据集进行验证, 准确率分别能达到 87.88%, 比同类算法高出 2.2%.

表 8 算法性能对比 (%)

算法	数据集	VA
HM-LSTM ^[9]	UTA-RLDD	65.2
MobileNetV2 ^[18]	DROZY	85.68
本算法	UTA-RLDD	67.27
	Drozy	87.88

3.2.4 多特征经验融合的消融分析

为了更好地研究不同疲劳行为特征对疲劳检测结果的影响, 多特征经验融合算法使用 RLDD 进行实验, 其中表 9–表 12 分别表示 singleton, mutual 和 activate/

inhibit 算子的消融实验结果, 在对每一个算子进行消融实验时, 其他算子的作用区间和模型参数不变.

表 9 singleton 算子消融实验 (%)

e2	m1	h1	VA
√	—	—	65.45
—	√	—	35.29
—	—	√	40
√	√	√	67.27

表 10 mutual 算子消融实验 (%)

[h2, e3]	[h4, e3]	VA
√	—	61.81
—	√	61.81
√	√	67.27

表 11 activate 算子消融实验 (%)

h2	e3	VA
√	—	50.90
—	√	52.72
√	√	67.27

表 12 inhibit 算子消融实验 (%)

h3	e1	VA
√	—	61.81
—	√	61.81
√	√	67.27

实验表明, 在 UTA-RLDD 数据集上慢眨眼行为 (e2) 的 singleton 算子起到了主要作用, 而 mutual 算子, activate 算子, inhibit 算子对疲劳检测准确度起到促进作用.

由表 9–表 12 可知, 当同时使用 3 种算子对 RLDD 进行疲劳早后期检测时效果最好, 疲劳早后期分类结果如表 13 混淆矩阵所示.

表 13 UTA-RLDD 检测结果的归一化混淆矩阵

real/pred	0	1	2
0	0.56	0.22	0.22
1	0	0.87	0.13
2	0	0.22	0.78

4 结束语

本文通过轻量级的 SCRFD 和 MobileNetV2 进行人脸检测和人脸关键点定位; 通过人脸校准和头部姿态信息提出自适应阈值方法, 可减少头部旋转对疲劳行为识别的影响; 通过自定义的 3 种特征算子和疲劳

因果图,设计了人脸多特征的经验融合方法,最后使用双尺度的KNN进行疲劳判定.实验验证,该模型对UTA-RLDD数据集的检测准确率达到67.27%,相较于HM-LSTM高出2.27%,并在Drozy数据集上高出同类方法的2.2%.该方法仅通过CPU推理,实时性可达到17.96 FPS.

本文所提出的疲劳检测方法无论是在疲劳检测精度还是在推理耗时上都取得了不错的效果,但目前仍存在以下几点问题:1)本文所构建的BlinkHeadpose数据集规模较小,导致GBDT学习到的眨眼阈值与真实阈值之间存在较大误差;2)利用头部姿态信息的自适应阈值的哈欠检测方法仍需要根据摄像头机位适当调节参数;3)多特征经验融合方法对疲劳的检测效果取决于各种疲劳行为的检测精度,存在误差累加的问题;4)本方法不支持多人同时疲劳检测.未来笔者会尝试根据以上问题继续开展研究工作.

参考文献

- 1 Sigari MH, Fathy M, Soryani M. A driver face monitoring system for fatigue and distraction detection. *International Journal of Vehicular Technology*, 2013, 2013: 263983.
- 2 Mandal B, Li LY, Wang GS, *et al.* Towards detection of bus driver fatigue based on robust visual analysis of eye state. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(3): 545–557. [doi: 10.1109/TITS.2016.2582900]
- 3 Fei YL, Li B, Wang H, *et al.* Long short-term memory network based fatigue detection with sequential mouth feature. *Proceedings of 2020 International Symposium on Autonomous Systems*. Guangzhou: IEEE, 2020. 17–22.
- 4 Abtahi S, Omidyeganeh M, Shirmohammadi S, *et al.* YawDD: A yawning detection dataset. *Proceedings of the 5th ACM Multimedia Systems Conference*. Singapore: ACM, 2014. 24–28.
- 5 Guo J, Deng JK, Lattas A, *et al.* Sample and computation redistribution for efficient face detection. *Proceedings of the Tenth International Conference on Learning Representations*. OpenReview.net, 2022. 1–17.
- 6 Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 4510–4520.
- 7 Guelman L. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications*, 2012, 39(3): 3659–3667. [doi: 10.1016/j.eswa.2011.09.058]
- 8 罗元, 云明静, 王艺, 等. 基于人眼信息特征的人体疲劳检测. *计算机应用*, 2019, 39(7): 2098–2102. [doi: 10.11772/j.issn.1001-9081.2018122441]
- 9 Ghoddoosian R, Galib M, Athitsos V. A realistic dataset and baseline temporal model for early drowsiness detection. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Long Beach: IEEE, 2019. 178–198.
- 10 Wang TS, Shi PF. Yawning detection for determining driver drowsiness. *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology*. Suzhou: IEEE, 2005. 373–376.
- 11 Alioua N, Amine A, Rziza M. Driver's fatigue detection based on yawning extraction. *International Journal of Vehicular Technology*, 2014, 2014: 678786.
- 12 Lepetit V, Moreno-Noguer F, Fua P. EPnP: An accurate $O(n)$ solution to the PnP problem. *International Journal of Computer Vision*, 2009, 81(2): 155–166. [doi: 10.1007/s11263-008-0152-6]
- 13 Zhang ZY. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330–1334. [doi: 10.1109/34.888718]
- 14 景芙蓉. 基于深度学习的疲劳驾驶检测研究与实现 [硕士学位论文]. 成都: 电子科技大学, 2021. [doi: 10.27005/d.cnki.gdzku.2021.002996]
- 15 Åkerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 1990, 52(1–2): 29–37.
- 16 Massoz Q, Langohr T, François C, *et al.* The ULg multimodality drowsiness database (called DROZY) and examples of use. *Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision*. Lake Placid: IEEE, 2016. 1–7.
- 17 刘佰强. 基于人脸视频智能分析的日间驾驶员不安全行为检测分类与预警系统 [硕士学位论文]. 西安: 西安理工大学, 2020. [doi: 10.27398/d.cnki.gxalu.2020.000333]
- 18 García-García M, Caplier A, Rombaut M. Sleep deprivation detection for real-time driver monitoring using deep learning. *Proceedings of the 15th International Conference on Image Analysis and Recognition*. Póvoa de Varzim: Springer, 2018. 435–442.

(校对责编: 牛欣悦)