

# 基于 Dlib 的面部疲劳检测模型

宋佳尧, 尉 斌, 安姝洁, 张涔嘉, 杨 玮

(天津商业大学信息工程学院, 天津 300134)

✉ 1429633682@qq.com; yubin@tjcu.edu.cn; 2741902510@qq.com; 1249439690@qq.com; 1642087781@qq.com



**摘要:**为了能够准确、实时地对稳定视频情境下人物的打哈欠等疲劳状态进行检测,文章基于 Dlib(机器学习的开源库)设计了针对人物面部特征的疲劳状态检测模型。模型包括人脸检测器、关键点定位器、特征计算器和状态预测器四个模块。首先,通过 Dlib 的相关函数进行人脸检测,然后跟踪定位面部特征点的实时坐标。其次,通过坐标值实时计算与疲劳状态相关度高的 PERCLOS(眼睛闭合时间占单位时间的百分率)参数、PMOT(张嘴时间占单位时间的百分率)参数,经对模型训练确定眼部纵横比和嘴部纵横比的阈值,对疲劳状态发出预警提示。经 YawDD 数据集验证,该模型的平均查全率约为 94.2%、平均准确率约为 93.3%,能够满足对面部疲劳检测的实时性和准确率的要求。

**关键词:**疲劳检测;Dlib;特征点定位

**中图分类号:**TP181 **文献标志码:**A

## Facial Fatigue Detection Model Based on Dlib

SONG Jiayao, YU Bin, AN Shujie, ZHANG Cenjia, YANG Wei

(College of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China)

✉ 1429633682@qq.com; yubin@tjcu.edu.cn; 2741902510@qq.com; 1249439690@qq.com; 1642087781@qq.com

**Abstract:** In order to accurately and real-time detect fatigue states of characters such as yawning in stable video scenarios, this paper proposes to design a fatigue state detection model based on Dlib (an open-source library for machine learning) for facial features of characters. The model includes four modules: face detector, key point locator, feature calculator, and state predictor. Firstly, facial detection is performed through the correlation function of the Dlib, and then real-time coordinates of facial feature points are tracked and located. Secondly, the PERCLOS (percentage of eye closure time to a specific time) parameter and PMOT (percentage of mouth opening time to a specific time) parameter, which are highly correlated with the fatigue state, are calculated in real time through the coordinate values. After training the model, the threshold values of the eye aspect ratio and mouth aspect ratio are determined, so as to issue an early warning prompt for the fatigue state. Validated by the YawDD dataset, the model achieves an average check rate of about 94.2% and an average accuracy of about 93.3%, which can meet the requirements of real-time and accuracy for facial fatigue detection.

**Key words:** fatigue detection; Dlib; feature point localization

## 0 引言(Introduction)

目前,对于疲劳状态的主要测量方法分为生理特征、视觉特征和语音特征等类型<sup>[1]</sup>。其中,基于生理特征<sup>[2]</sup>的疲劳检测大多情况下会使用侵入式的传感器,高昂的价格及侵入式设备可能引起佩戴者的不适,是影响其普及推广的两大因素;基于语音特征<sup>[3]</sup>的疲劳检测只适用于采用标准呼叫应答的场景且

语音标记样本数据较为稀缺;基于视觉特征<sup>[4]</sup>的疲劳检测具有非接触性、可直接根据人的面部关键点坐标的变化(如头部位置的偏移、眼睛闭合时长、打哈欠等)反映其疲劳状态等优势,成为目前研究的主流方法。在驾驶环境、云自习室等情境下,基于面部视觉特征的疲劳状态检测,可以帮助人们对驾驶员的疲劳状态、学生注意力不集中的状态发出预警。

## 1 基于面部特征的疲劳检测 (Fatigue detection based on facial features)

目前,应用较为广泛的基于面部特征的疲劳检测度量参数有 PERCLOS(眼睛闭合时间占单位时间的百分率)参数和 PMOT(张嘴时间占单位时间的百分率)参数。PERCLOS 是卡内基梅隆研究所提出的度量疲劳的物理量<sup>[5]</sup>。美国联邦公路管理局开展了在驾驶情境下,对 PERCLOS、眨眼频率、打哈欠等 9 种疲劳参数进行相关性分析研究,得出 PERCLOS 参数与疲劳驾驶状态相关性最高的结论<sup>[6]</sup>。

PERCLOS 参数表示闭眼时长在单位时间内所占的百分比,使用闭眼帧数与单位时间内视频总帧数  $n$  的比值表示,其计算公式如下:

$$p_{\text{PERCLOS}} = \frac{\sum_{i=1}^n f_i}{n} \times 100\% \quad (1)$$

根据参考文献<sup>[7]</sup>,将眼睑遮住瞳孔的面积达到 80%,判定为闭眼的 P80 标准对疲劳状态的检测是最准确的。

PMOT 参数的计算公式如公式(2)所示,它用来表示张嘴时长在单位时间内所占的百分比,使用张嘴帧数与单位时间内视频总帧数  $m$  的比值表示。

$$p_{\text{PMOT}} = \frac{\sum_{i=1}^n f_i}{m} \times 100\% \quad (2)$$

值得注意的是,单独使用 PMOT 参数,有时会造成打哈欠的张嘴和说话张嘴的误报,实际进行疲劳检测时,可以结合眨眼频率和头部姿态参数进行阈值的确定。

## 2 模型设计 (Design of the model)

Dlib 是一种开源的、跨平台的工具包,包含诸如图像处理、机器学习和深度学习等众多模块。对比 TensorFlow 和 PyTorch 两大框架,Dlib 模型在图像处理和特征点定位方面的通用性更强,具有更大的优越性。因此,Dlib 模型在人脸检测领域的应用也非常广泛。本文设计的 Dlib 面部疲劳状态检测模型,由人脸检测器、关键点定位器、特征计算器和状态预测器四个模块构成,面部疲劳状态检测模型框架如图 1 所示。其中,人脸检测器的任务是读取视频,并通过 `dlib.get_forntal_face_detector` 函数捕获视频流中的人脸,将人脸框发送给关键点定位器。关键点定位器的任务是通过 `dlib.shape_predictor` 函数获取人脸的 68 个关键点的坐标,并将坐标向量传递给特征计算器。特征计算器的任务是分别对眼部特征、嘴部特征和头部姿态进行计算,并将结果实时发送给状态预测器。状态预测器的任务是对疲劳状态进行判定,不依赖于单一参数,而是结合眼部、嘴部和头部姿态进行综合度量。例如,人在打瞌睡时,除了会闭眼,头部也会下垂;打哈欠时,除了嘴部的张开角度比说话时大,还会出现闭眼的情况。



图 1 面部疲劳状态检测模型框架

Fig. 1 Framework of facial fatigue state detection model

Dlib 库提供了 `dlib.get_forntal_face_detector` 函数,基于这个函数可以构造人脸检测器,人脸检测器采用 HOG 算法(Histogram of Oriented Gradient,方向梯度直方图)、线性分类器、金字塔图像结构和滑动窗口检测等技术,Dlib 模型中的 HOG 算法主要用于捕获轮廓信息,首先对图像进行灰度化处理,其次进行 Gamma 压缩和归一化图像,以减少户外光照对图像的影响;其检测效果如图 2 所示。在本文的模型设计中,人脸检测器可以针对摄像头进行实时的人脸捕获。

面部关键点定位器是基于 Dlib 库中的 `dlib.shape_predictor` 函数构造的,该函数对面部的 68 个关键点(编号从 1~68)进行标注。从图 3 可以看出,左眼的关键点编号是 37~42,右眼的关键点编号是 43~48,嘴部的关键点编号是 49~68。图 2 和图 3 中使用的测试图片来自于 WIDER FACE 公开数据集。



图 2 人脸检测器效果  
Fig. 2 Face detector effect

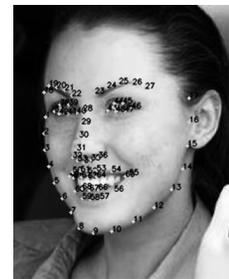


图 3 关键点定位器效果  
Fig. 3 Key locator effect

## 3 实现原理 (Implementation principle)

### 3.1 对瞌睡状态的判定

是否发生了疲劳状态下的瞌睡行为主要通过“闭眼+头部姿态变化”综合判定。

在 Dlib 库中,眼睛的位置是通过六个关键点进行标注的(图 4)。因此,眼睛纵横比的计算方法如下<sup>[8]</sup>:

$$\frac{|p_1 - p_5| + |p_2 - p_4|}{2 \times |p_0 - p_3|} \quad (3)$$

头部姿态估计主要是获得脸部朝向的角度信息,一般可以用旋转矩阵、旋转向量、四元数或欧拉角表示<sup>[9]</sup>。欧拉角的可读性更好,使用较为广泛,常用的三个欧拉角度是俯仰角 Pitch、摇头 Yaw 和滚转角 Roll。

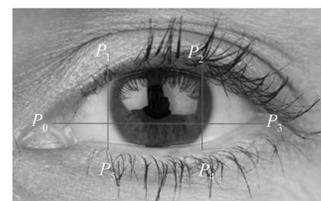


图 4 计算纵横比的眼部关键点

Fig. 4 Key points for calculating eye aspect ratios

为减少误报,提高模型判定的准确率,要对正常状态下的眨眼和疲劳时的闭眼进行区分,正常状态下眨眼的闭眼时间极短,而疲劳时闭眼时间相对较长。正常状态下眨眼时,眼睛纵横比只在瞬时(1~3 帧)处于 0 值,因此在判断眼睛纵横比的基础上,还要检测闭眼的持续时长<sup>[10]</sup>。

持续时长可通过计数器统计。在眼睛纵横比小于 0.2 时,认为当前帧处于闭眼状态,计数器加 1。一旦计数器的值大于阈值,则结合头部当前的姿态进行判定。例如,计数器的值大于 48 且头部俯仰角 Pitch 大于 15°,则判定为瞌睡行为。

### 3.2 对打哈欠状态的判定

是否发生了疲劳状态下的打哈欠行为,主要通过嘴部的张开角度和持续时间判定。在 Dlib 库中,嘴部的外部轮廓对应 12 个关键点,如图 3 中的编号 49~60,内部轮廓对应 8 个关键点,如图 3 中的编号 61~68。

类似眼睛纵横比的计算方法,选取嘴部外部轮廓的 6 个关键点(关键点的位置如图 5 所示),进行嘴部纵横比的计算,具体计算方法如公式(4)所示:

$$\frac{|p_7 - p_{11}| + |p_8 - p_{10}|}{2 \times |p_6 - p_9|} \quad (4)$$

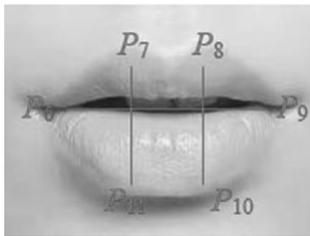


图 5 计算纵横比的嘴部关键点

Fig. 5 Key points for calculating mouth aspect ratios

一般情况下,人们在安静、说话、唱歌等不同状态时,嘴部的纵横比对应不同的数值。例如,安静时的嘴部是闭合的,正常说话和交谈时嘴部是半张开状态的,而在打哈欠时,会把嘴张大且持续一定的时间。

为减少误报,提高模型判定的准确率,要对正常状态下的说话和疲劳时的张嘴打哈欠行为进行区分,人在正常说话时,他的嘴巴张开和闭合是非常频繁的,而打哈欠时,嘴部纵横比数值大且持续时间长(大于 4 s)。持续时间的计算与“3.1”小节的计算方法类似,可通过计数器统计。在嘴部纵横比大于 0.3 时,认为当前帧处于张嘴状态,计数器加 1。当计数器的值大于 96,则判定为打哈欠。

## 4 模型测试(Model testing)

实验环境为 OpenCV3.4.9.3+Dlib-19.17.99,开发环境为 Python3.7.3,Harr 分类器选用 haarcascade\_frontalface\_alt.xml。眼睛纵横比的阈值为 0.2,嘴部纵横比的阈值为 0.3。

本文使用 YawDD 数据集作为测试数据集,YawDD 数据集是一个公开的视频数据集,共收集了 90 名汽车驾驶员的驾驶视频,其中女性驾驶员为 43 名,男性驾驶员为 47 名。视频总数有 351 条,其中内后视镜角度的视频合计 320 条(女性驾驶员为 156 条,男性驾驶员为 164 条);仪表盘角度的视频合计 29 条(女性驾驶员为 13 条,男性驾驶员为 16 条)。部分视频中,驾驶员佩戴了普通眼镜或太阳眼镜。图 6 展示了 YawDD 数据集中有代表性的视频数据,分为男性戴普通眼镜、男性戴太阳眼镜、男性未戴眼镜、女性戴普通眼镜、女性戴太阳眼镜、女性未戴眼镜六种类型。



图 6 YawDD 数据集的六类代表性数据

Fig. 6 Six types of data in the YawDD dataset

针对上述六种类型的视频数据,模型在测试时均成功捕获到驾驶员打哈欠的状态,并实时预警提示信息,如图 7 所示。



图 7 对打哈欠的捕获效果

Fig. 7 Capture effect on yawning

观察模型测试过程,发现佩戴普通眼镜或太阳眼镜,均未影响模型对眼部关键点的定位,但在佩戴太阳眼镜时,有时会因光线照射角度的不同影响对眼部关键点坐标的实时跟踪,进而产生计算延迟,对疲劳状态的检测造成了一定影响。模型在 YawDD 数据集中测试统计结果如表 1 所示,平均查全率约为 94.2%,平均准确率约为 93.3%。

表 1 测试统计结果

Tab.1 Test result statistics

视频类别	实际疲劳 次数/次	检测疲劳 次数/次	查全 率/%	误报 率/%	准确 率/%
男性戴普通眼镜	18	19	100	5	95
男性戴太阳眼镜	6	5	83	0	83
男性不戴眼镜	66	66	100	0	100
女性戴普通眼镜	16	15	94	0	94
女性戴太阳眼镜	8	7	88	0	88
女性不戴眼镜	62	62	100	0	100

以下针对查全率低于 100%的情况进行分析。从表 1 的数据来看,存在漏报导致查全率低的情况主要集中在男性戴太阳眼镜、女性戴普通眼镜和女性戴太阳眼镜三种情况。但通过对模型的检测发现,是否佩戴眼镜不是影响测试的重要因素,驾驶员的肢体移动、手势变化才是影响测试的主要因素。如图 8 所示,女性驾驶员在打哈欠的同时,用右手捂住了嘴巴,导致模型无法跟踪嘴部关键点坐标而产生了漏报。这些情况也为下一步对模型进行优化提供了方向。

(下转第 58 页)