



基于深度学习的疲劳驾驶检测算法

郑伟成, 李学伟, 刘宏哲, 代松银

(北京联合大学 北京市信息服务工程重点实验室, 北京 100101)

摘 要: 为实现复杂驾驶环境下驾驶人员疲劳状态识别与预警, 提出基于深度学习的疲劳驾驶检测算法。利用基于 shuffle-channel 思想的 MTCNN 模型检测常规摄像头实时采集的驾驶人员人脸图像, 使用 PFLD 深度学习模型进行人脸关键点检测以定位眼部、嘴部和头部位置, 从中提取眨眼频率、嘴巴张开程度和点头频率等特征参数, 并通过多特征融合策略获取驾驶人员疲劳状态, 从而实现疲劳驾驶的有效预警。实验结果表明, 该算法给出的疲劳驾驶预警结果均未出现误判情况, 具有较高的检测准确率和较好的鲁棒性。

关键词: 疲劳驾驶检测; 疲劳特征提取; PERCLOS 值; 人脸检测; 人脸关键点检测; 头部姿态估计

开放科学(资源服务)标志码(OSID):



中文引用格式: 郑伟成, 李学伟, 刘宏哲, 等. 基于深度学习的疲劳驾驶检测算法[J]. 计算机工程, 2020, 46(7): 21-29.

英文引用格式: ZHENG Weicheng, LI Xuewei, LIU Hongzhe, et al. Fatigue driving detection algorithm based on deep learning[J]. Computer Engineering, 2020, 46(7): 21-29.

Fatigue Driving Detection Algorithm Based on Deep Learning

ZHENG Weicheng, LI Xuewei, LIU Hongzhe, DAI Songyin

(Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China)

[Abstract] To realize identification and warning of fatigue driving detection in complex driving environment, this paper proposes an algorithm for fatigue driving detection based on deep learning. The algorithm uses the MTCNN model based on the shuffle-channel concept to detect the facial images of drivers collected in real time by normal cameras. Then the PFLD deep learning model is used for facial keypoint detection to locate the eyes, the mouth and the head, so as to extract the feature parameters including the blinking rate, the extent to which the mouth opens, and the nodding frequency. Finally, based on the multi-feature fusion strategy, the fatigue state of the driver is obtained to implement effective alarming for fatigue driving. Experimental results show that false warning do not occur in fatigue driving warning generated by the proposed algorithm, which means the proposed algorithm has a high detection accuracy and robustness.

[Key words] fatigue driving detection; fatigue feature extraction; PERCLOS value; face detection; face keypoint detection; head pose estimation

DOI: 10.19678/j.issn.1000-3428.0055912

0 概述

疲劳驾驶是指驾驶人员在长时间连续行车后, 造成生理和心理机能的失调, 而在客观上出现驾驶技能下降的现象。驾驶人员疲劳后继续驾驶车辆, 身体会出现困倦瞌睡、注意力不集中、驾驶操作停顿或对危险情况的预判能力下降等不安全因素, 极易发生道路交通事故。近年来, 世界各大汽车厂商与

科研机构将疲劳驾驶检测技术作为研究重点并取得了相关研究成果, 其主要分为主观检测和客观检测两大类。主观检测方法主要靠驾驶人员的主观因素进行判断, 可靠性不高, 很难成为评价疲劳驾驶的标准, 而客观检测方法可靠性相对较高, 已成为疲劳驾驶检测的主要方法。

客观检测方法主要分为基于生理特征的检测方法、基于车辆行驶信息的检测方法及基于驾驶人员

基金项目: 国家自然科学基金(61871039, 61802019, 61906017); 国家科技支撑计划项目(015BAH55F03); 北京市自然科学基金(4184088); 北京市属高校高水平教师队伍建设支持计划项目(IDHT20170511); 北京市教委项目(KM201911417001, KM201711417005); 北京联合大学领军人才项目(BPHR2019AZ01); 智能驾驶大数据协同创新中心项目(CYXC1902)。

作者简介: 郑伟成(1986—), 男, 硕士研究生, 主研方向为目标检测、智能驾驶; 李学伟(通信作者), 教授、博士生导师; 刘宏哲, 教授; 代松银, 博士。

收稿日期: 2019-09-04

修回日期: 2019-10-12

E-mail: andy_cheng0617@163.com

行为特征的检测方法 3 类。其中,基于驾驶人员行为特征的检测方法通过分析驾驶人员的眼睛状态(如眨眼幅度、眨眼频率等)、嘴部运动、头部姿态等,实现驾驶人员的疲劳检测。此类方法与前两者相比,有非接触性、准确性高、成本低及易实现等优点,因此疲劳驾驶预警系统具有极其广阔的市场前景。

基于驾驶人员行为特征的检测方法根据面部疲劳特征又可分为基于眼部特征的检测方法、基于嘴部特征的检测方法及基于头部姿态的检测方法 3 类。在基于眼部特征的检测方法中,当驾驶人员出现疲劳驾驶时,其眼睛闭合持续时间会明显变长,眨眼频率会加快或减少,因此眼部特征直接反映驾驶人员的精神状态。文献[1]定位出人眼区域后,在人眼定位过程中,采用 OSTU 阈值分割、非线性点运算和积分投影等预处理过程消除眉毛,并利用模糊综合评价算法对眼睛矩形区域的长宽比、拟合椭圆面积、瞳孔黑色素所占比例这 3 个影响因子进行分析,判别出眼睛的睁开闭合状态,最后根据 PERCLOS^[2]原理对驾驶人员的疲劳状态做出判断。在基于嘴部特征的检测方法中,当驾驶人员感到疲劳时,嘴部会自发地发生打哈欠的行为,嘴部张开程度大,张开时间持久。文献[3]通过分析标注剥夺睡眠的驾驶人员视频数据集,说明面部触摸、人脸遮挡及打哈欠之间的关系,提出面部触摸可作为检测打哈欠的新线索,并提出一种通过提取嘴部和眼部区域的几何和外观特征来检测打哈欠的方法。在基于头部姿态的检测方法中,当驾驶人员产生疲劳时,打瞌睡点头次数会增多,出现长时间不正视前方等姿态。文献[4]提出一种基于面部特征点定位的头部姿态估计方法。在检测到的人脸区域,使用 Hough 圆检测方法定位眼睛和鼻孔,利用人眼和鼻孔的位置信息,将眼睛、鼻子

定位结果与正脸头部姿态中的眼睛、鼻子进行对比,从而对不同的头部姿态进行粗估计。文献[5]提出一种简单且鲁棒的头部姿态确定方式,通过训练一个多损失函数的卷积神经网络,直接使用 RGB 结合分类和回归损失来预测头部姿态欧拉角。

传统基于驾驶人员行为特征的检测方法使用 HOG、SIFT、Haar 等特征描述子,人工提取特征进行模式识别,在实际复杂的驾驶环境下检测精度通常不能满足需求,且鲁棒性也有待提高。深度学习^[6-7]技术是一类能从数据中自动学习特征表示的方法,近年来在目标检测^[8-9]、目标跟踪^[10-11]、人脸识别^[12-13]、场景理解^[14]等图像处理领域取得了快速发展,其强大的特征学习能力,对部分遮挡、光照变化等外界条件表现出较强的鲁棒性。因此,本文提出基于深度学习的疲劳驾驶检测算法,选取表征疲劳驾驶常用且有效的眼睛闭合程度、嘴巴张开程度和头部姿态等参数指标,利用多特征融合策略检测驾驶人员的疲劳程度。

1 疲劳驾驶检测算法整体框架

本文提出的基于深度学习的疲劳驾驶检测算法具体步骤为:1)利用改进的 MTCNN^[15]人脸检测模型检测出摄像头实时获取画面中的驾驶人员人脸图像;2)基于人脸图像,使用具有较高效率和准确率的 PFLD^[16]模型进行人脸关键点检测,从而定位出眼部和嘴部;3)根据人脸关键点计算出人脸图像的头部姿态欧拉角,即俯仰角(Pitch)、偏航角(Yaw)及滚转角(Roll);4)采用多特征融合策略,根据眼部、嘴部和头部的疲劳特征建立疲劳驾驶预测模型,实现对疲劳驾驶的预警。算法总体流程如图 1 所示。

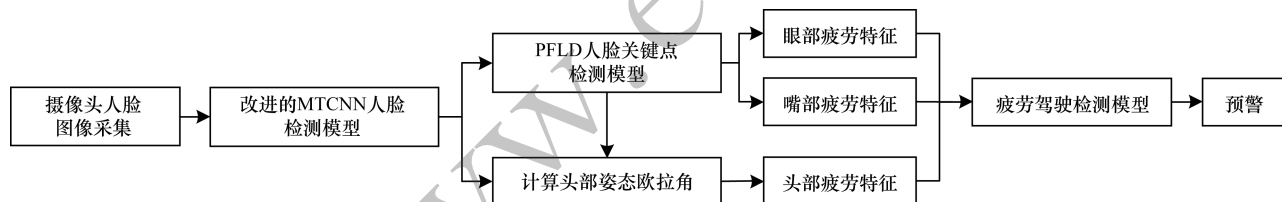


图 1 基于深度学习的疲劳驾驶检测算法流程

Fig. 1 Procedure of fatigue driving detection algorithm based on deep learning

2 改进的 MTCNN 人脸检测模型

MTCNN 是结合级联与由粗到精思想的深度学习模型(级联结构包括 PNet、RNet 及 ONet),其是目前应用十分广泛的特定目标检测器,也是少数能在传统硬件上实现进行检测器。对于人脸检测任务,在人脸数少于 5 的情况下,其具有较高的检测水准并且有极大的优化空间,通过模型训练可在移动端达到极快的检测速度。实验结果表明,原始 MTCNN 模型在速度方面存在瓶颈,具体为:1)输入

图片越大, PNet 耗时越长;2)人脸越多, ONet 和 RNet 耗时越长;3)噪点比较多的夜晚图像会导致 PNet 误检增加。

针对 PNet、ONet 和 RNet 耗时长的问题,采取优化网络结构的策略,使其在精度下降不明显情况下,尽可能减少计算量。改进方法主要为使用 ShuffleNet^[17]中的 shuffle-channel 思想,将特征的通道平均分到不同组中,使每个组在卷积时能得到其他组的信息,增加不同 channel 之间的关联性。实验为对比改进方法的优越性,采用 Xception^[18]中

depthwise 卷积思想以减少计算量和提升效率。首先对输入图的每个通道进行卷积,然后由 1×1 卷积将通道进行合并。实验结果证明,此操作基本可以等同于普通的空间卷积,并且在 I/O 效率和性能不变的情况下,计算量降低了 90%。因此,利用该方法替换 PNet、RNet 和 ONet 中的卷积操作,可使模型检测速度得到大幅提升。

由于驾驶场景下摄像机采集的一般为单张人脸图像,因此实验仅在单张人脸数据集上进行测试,使用的深度学习框架为 TensorFlow,测试硬件环境为 i7-9750H CPU,测试软件为 Pycharm2018。在某些监控场景的夜晚图片中,会有大量噪点的出现,导致 PNet 产生大量的误检候选区。为减少这些噪点的出现,实验在 PNet 检测前,使用一次中值滤波进行快速去噪,具有较好的实验效果。

3 人脸关键点检测模型

在疲劳驾驶检测应用场景中,除了根据人脸关键点定位眼部及嘴部外,还可利用关键点进行头部姿态估计,计算头部姿态的欧拉角。图 2 为人脸中的 68 个关键点。

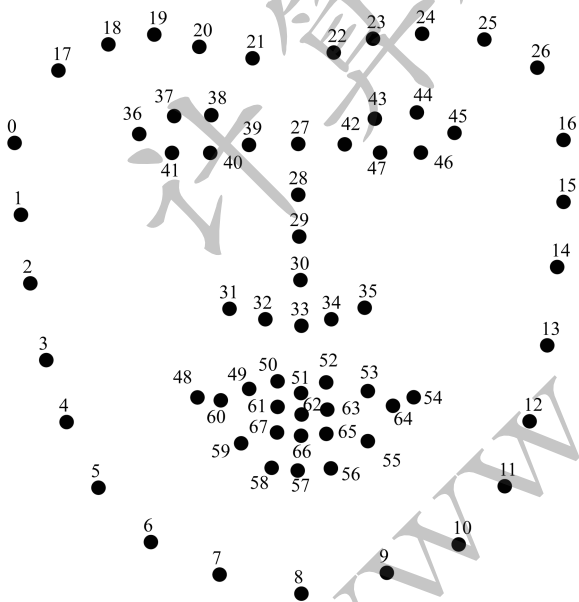


图 2 人脸中的 68 个关键点
Fig.2 68 keypoints in the face

从实用性角度来看,人脸关键点检测面临诸多问题:1)人脸表情变化很大,真实环境光照复杂,而且现实中存在大量人脸局部被遮挡的情况等;2)人脸是 3D 的,位姿变化多样,另外,因拍摄设备和环境影响,成像质量也有所不同;3)现有训练样本的各个类别存在不平衡问题;4)在计算受限的设备(如手机)终端时,必须考虑计算速度和模型规格问题。

为解决以上问题,本文提出人脸关键点检测模

型 PFLD。该模型的主干网络用于定位关键点的位置,在训练时加入预测人脸姿态的辅助网络以提高定位精度,但在测试时省略该网络。对于训练样本类别不平衡问题,PFLD 通过修改 loss 函数,在训练时关注稀有样本,如果某类别样本较少则赋予较大的权重,并使用轻量级模型提高计算速度和降低模型规格。

在模型设计上,PFLD 模型的骨干网络未采用 VGG16^[19]、ResNet^[20] 等网络,但为增加模型的表达能力,对 MobileNet^[21] 的输出特征进行结构修改,通过融合 3 个不同尺度的特征来增加模型表达能力,并且为更好地兼容嵌入式设备,PFLD 模型进行了性能优化,在骁龙 845 芯片中的效率可达 140 frame/s,模型内存为 2.1 MB,在许多关键点检测的 Benchmark 中也取得了较好的结果。

4 头部姿态估计

在实际驾驶环境中,头部姿态主要用来判断驾驶人员的注意力情况,如检测长途司机是否在目视前方,若长时间不目视前方,则给予相应警告,以保证安全,减少事故。

头部姿态欧拉角的计算分为:1)2D 人脸关键点检测(使用 PFLD 模型),该部分采用的人脸关键点有鼻尖、下巴、左眼的左角、右眼的右角、左嘴角、右嘴角;2)3D 人脸模型匹配;3)求解 3D 点和对应 2D 点的转换关系;4)根据旋转矩阵求解欧拉角。

人脸相对于相机的姿态可以使用旋转矩阵和平移矩阵进行表示,其中:平移矩阵(t)表示物体相对于相机的空间位置关系矩阵;旋转矩阵(R)表示物体相对于相机的空间姿态关系矩阵。具体的计算过程为:人脸特征的 3D 坐标是以世界坐标系进行表示,如果旋转和平移均已知,那么可以将世界坐标中的 3D 点转换到相机坐标系中的 3D 点;然后采用相机的内在参数(焦距、光学中心等),将相机坐标系中的 3D 点再投影到图像平面(如图像坐标系)。图 3 为坐标系统的工作原理。

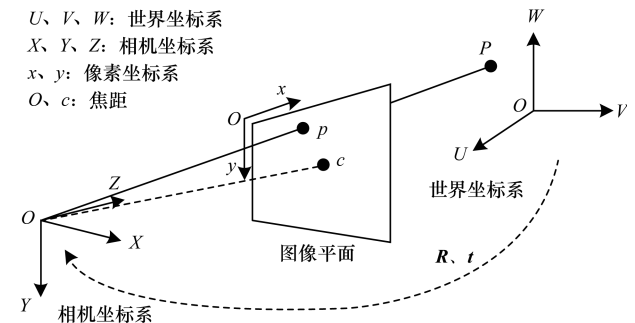


图 3 坐标系统工作原理

Fig.3 Working principle of coordinate system

世界坐标系到相机坐标系的转换、相机坐标系到像素坐标系的转换如式(1)、式(2)所示:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \mathbf{R} \begin{pmatrix} U \\ V \\ W \end{pmatrix} + \mathbf{t} = [\mathbf{R} | \mathbf{t}] \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix} \quad (1)$$

$$s \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (2)$$

像素坐标系和世界坐标系的关系为:

$$s \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{pmatrix} [\mathbf{R} | \mathbf{t}] \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix} \quad (3)$$

式(3)可用 DLT 算法结合最小二乘进行迭代求解,最小二乘的目标函数为:

$$J = (\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2 \quad (4)$$

其中,带“^”的变量为预测值,其余为测量值。若相机存在径向和切向畸变,则需先将相机坐标系转换到图像中心坐标系(如式(5)所示),然后计算畸变(如式(6)所示),最后得到图像中心坐标系和像素坐标系的关系(如式(7)所示)。

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} X/Z \\ Y/Z \end{pmatrix} \quad (5)$$

$$\begin{cases} u = u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 uv + p_2(r^2 + 2u^2) \\ v = v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_2 uv + p_1(r^2 + 2v^2) \end{cases} \quad (6)$$

$$s \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (7)$$

由式(7)可以看出,若已知世界坐标系内点的位置、像素坐标位置和相机参数,则可以使用 Opencv 中求解 PnP 问题的 solvePnP 函数计算旋转和平移矩阵,计算得到的旋转矩阵为:

$$\mathbf{R} = \begin{pmatrix} r_{00} & r_{01} & r_{02} \\ r_{10} & r_{11} & r_{12} \\ r_{20} & r_{21} & r_{22} \end{pmatrix} = \mathbf{A} + \mathbf{B} + \mathbf{C} \quad (8)$$

其中:

$$\mathbf{A} = \begin{pmatrix} \cos \phi \cos \gamma \\ \cos \phi \sin \gamma + \sin \phi \sin \phi \cos \gamma \\ \sin \phi \sin \gamma - \cos \phi \sin \phi \cos \gamma \end{pmatrix} \quad (9)$$

$$\mathbf{B} = \begin{pmatrix} -\cos \phi \cos \gamma \\ \cos \phi \cos \gamma + \sin \phi \sin \phi \sin \gamma \\ \sin \phi \cos \gamma + \cos \phi \sin \phi \sin \gamma \end{pmatrix} \quad (10)$$

$$\mathbf{C} = \begin{pmatrix} \sin \phi \\ -\sin \phi \sin \phi \\ \cos \phi \sin \phi \end{pmatrix} \quad (11)$$

最终得到的欧拉角为:

$$\begin{cases} \varphi = \text{atan}(-r_{12}, r_{22}) \\ \phi = \text{atan}(r_{02}, \sqrt{r_{12}^2 + r_{22}^2}) \\ \gamma = \text{atan}(-r_{01}, r_{00}) \end{cases} \quad (12)$$

其中,世界坐标系的中心位置由 3DMM 模型进行确定^[22]。实验中首先对 DLT 算法进行了改进,即随机轻微地改变姿态(\mathbf{R} 和 \mathbf{t}),然后检查重投影误差的衰减情况,如果是衰减,则接受为新的姿态,最后继续调整 \mathbf{R} 和 \mathbf{t} 以寻找更好的姿态。

5 疲劳特征提取

5.1 眼部疲劳特征提取

驾驶人员眼部的运动和眨眼信息被认为是反映疲劳状态的重要特征,眼部疲劳检测算法利用 PERCLOS 中的 P80 标准和眨眼频率进行眼部疲劳特征提取。

5.1.1 PERCLOS 值

PERCLOS 是卡内基梅隆研究所经过反复实验和论证提出的度量疲劳/瞌睡的物理量。研究表明,在 PERCLOS 常用的判定标准中,P80 标准与疲劳驾驶的相关性较高,P80 标准指当眼睑遮住瞳孔的面积超过 80% 时记为眼睛闭合,统计在一定时间内眼睛闭合时所占的时间比例,其原理如图 4 所示。

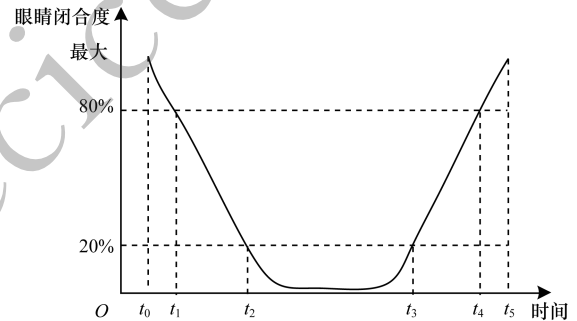


图 4 PERCLOS 中的 P80 标准原理

Fig.4 Principle of P80 standard in PERCLOS

PERCLOS 值的计算公式如下:

$$f_p = \frac{t_4 - t_2}{t_5 - t_1} \quad (13)$$

其中, f_p 是 PERCLOS 值,若 D 为眼睛完全睁开时的上下眼睑距离,则 t_1 是 D 减少 20% 的时刻, t_2 是 D 减少 80% 的时刻, t_4 是眼睛由闭合增加至 D 的 20% 的时刻, t_5 是眼睛由闭合增加至 D 的 80% 的时刻。PERCLOS 值与眼睛睁开程度密切相关,眼睛睁开程度由眼睛上下眼睑围成面积的变化进行表示。将眼睛上下眼睑的轮廓与虹膜看作两个嵌套的椭圆,如图 5 所示。

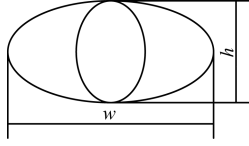


图 5 P80 标准的眼轮廓模型

Fig. 5 Eye contour model of P80 standard

眼睛面积为 $S = \pi wh$, 其中, w 表示眼睛宽度, h 表示上下眼睑的垂直距离。 w 和 h 的计算公式如下:

$$\begin{cases} w = x_{39} - x_{36} \\ h = y_{40} - y_{38} \end{cases} \quad (14)$$

其中, x_{39} 、 x_{36} 为第 39 个和第 36 个关键点的横坐标, y_{40} 、 y_{38} 为图 2 中第 40 个和第 38 个关键点的纵坐标 (索引下标从 0 开始)。 S 的最终取值为左右眼睛面积的平均值, 设眼睛睁开百分比为 R , 则可以得到:

$$R = \frac{S}{S_{\max}} = \frac{\pi wh}{\pi w_{\max} h_{\max}} = \frac{wh}{w_{\max} h_{\max}} \quad (15)$$

其中, w_{\max} 、 h_{\max} 、 S_{\max} 分别为眼睛完全睁开时的值, 终端采集视频中每帧图片的眼睛睁开程度按照 R 值进行评定, 再根据某时间段推算得到 f_p 值。在一般情况下, 正常人的闭眼速度为 0.2 s, 如果闭眼持续时间超过 0.8 s, 则为疲劳状态。

5.1.2 眨眼频率

据统计, 正常人每分钟眨眼约 20 次, 每次眨眼时间间隔约 3 s, 发生疲劳时眨眼的频率会加快, 若平均每秒眨眼一二次, 属于应预警的疲劳级别, 因此可以通过眨眼频率 (即单位时间内眨眼次数) 来判断驾驶人员是否疲劳。眨眼频率 f_{blink} 为:

$$f_{\text{blink}} = \frac{N_B}{N_p} \quad (16)$$

其中, N_B 为时间段 T 内发生眨眼动作的次数, 可由眼睛纵横比 (Eyes Aspect Ratio, EAR) 来确定, N_p 为时间段 T 内的视频总帧数。以图 2 中的左眼为例, 提取的人脸左眼关键点分布如图 6 所示。

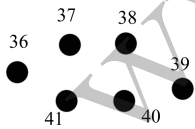


图 6 人脸左眼关键点分布

Fig. 6 Distribution of keypoints of the left eye in the face

EAR 的计算公式为:

$$\text{EAR} = \frac{|y_{37} - y_{41}| + |y_{38} - y_{40}|}{2|x_{36} - x_{39}|} \quad (17)$$

其中, y_{37} 、 y_{41} 、 y_{38} 、 y_{40} 分别为第 37 个、第 41 个、第 38 个和第 40 个关键点纵坐标的坐标值, x_{36} 、 x_{39} 分

别为第 36 个、第 39 个关键点横坐标的坐标值。EAR 的最终取值为左右眼 EAR 的平均值。EAR 在眼睛睁开时基本保持不变, 在小范围内会上下浮动; 当眼睛闭合时, EAR 会迅速下降, 即进行一次眨眼动作, 由此计算出时间段 T 内发生眨眼动作的次数 N_B 。

5.2 嘴部疲劳特征提取

嘴部状态通常有闭合、说话及打哈欠 3 种, 在疲劳状态下, 人会频繁地打哈欠。类似于眼部疲劳特征的提取, 提取的人脸嘴部区域的关键点分布如图 7 所示。

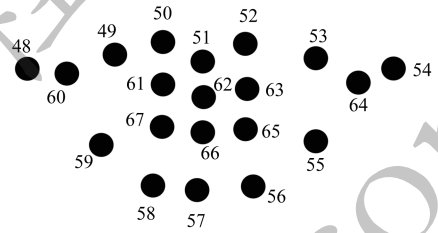


图 7 人脸嘴部关键点分布

Fig. 7 Distribution of keypoints of the mouth in the face

嘴部宽度 W 及高度 H 计算如下:

$$\begin{cases} W = |x_{\max} - x_{\min}| \\ H = |y_{\max} - y_{\min}| \end{cases} \quad (18)$$

W 与 H 的比值 K 为:

$$K = \frac{W}{H} \quad (19)$$

其中, x_{\max} 、 x_{\min} 分别为第 54 个和第 48 个人脸关键点的横坐标, y_{\max} 为第 56 个、第 58 个关键点纵坐标的平均值, y_{\min} 为第 50 个、第 52 个关键点纵坐标的平均值。打哈欠时嘴巴的张开程度最大, 此时 K 值变小, 由此可判断驾驶员是否打哈欠。打哈欠的频率 f_{yawn} 为:

$$f_{\text{yawn}} = \frac{N_Y}{N_T} \quad (20)$$

其中, N_Y 为时间段 T 内打哈欠的次数, N_T 为时间段 T 内的视频总帧数。 N_Y 的计算方法同眨眼次数 N_B , 可由嘴部纵横比 (Mouth Aspect Ratio, MAR) 来确定。

5.3 头部姿态疲劳特征提取

文献[23]研究发现, 成年人的头部运动范围包括 $-60.4^\circ \sim 69.6^\circ$ 的矢状屈曲和伸展 (即从颈部向后运动, Pitch), $-40.9^\circ \sim 36.3^\circ$ 的正面侧向弯曲 (即颈部从右向左弯曲, Roll), $-79.8^\circ \sim 75.3^\circ$ 的水平轴向旋转 (即从头部向左旋转, Yaw)。当驾驶人员处于疲劳状态时, 通常会无意识地出现身体的

前倾后仰、左右倾斜或其他疲劳姿态,导致不能进行正确的驾驶操作,最常见的表现为驾驶人员头部的点头频率和异常姿态,即 Roll、Pitch 及 Yaw 旋转角度的变化。

5.3.1 点头频率

头部姿态检测得到的欧拉角中前后方向的 Pitch 与驾驶人员的点头频率相关。当 Pitch 达到一定阈值时判定为点头动作。点头频率 f_{node} 为:

$$f_{\text{node}} = \frac{N_N}{N_T} \quad (21)$$

其中, N_N 为时间段 T 内发生点头动作的次数, N_T 为时间段 T 内的视频总帧数。

5.3.2 异常姿态

本文疲劳驾驶检测算法在头部姿态检测的基础上,结合 PERCLOS 的 P80 标准,当头部姿态欧拉角发生 20% 的角度变化时,则判定驾驶人员处于异常姿态,即 Roll 的异常范围为 $-56.3^\circ \sim -40.9^\circ$ 、 $36.3^\circ \sim 51.7^\circ$, Pitch 的异常范围为 $-86.4^\circ \sim -40.9^\circ$ 、 $69.6^\circ \sim 95.6^\circ$, Yaw 的异常范围为 $-110.8^\circ \sim -79.8^\circ$ 、 $75.3^\circ \sim 106.3^\circ$ 。异常姿态指标 f_{pose} 为:

$$f_{\text{pose}} = \frac{N_P}{N_T} \quad (22)$$

其中, N_P 为时间段 T 内检测出异常姿态的帧数。

6 实验结果与分析

6.1 生成模型

6.1.1 人脸检测模型

改进的 MTCNN 人脸检测模型使用基准人脸数据集 WIDER FACE^[24]进行训练,在下载的数据集上共有 12 880 幅图像,其中包括 159 424 张人脸。如图 8 所示,其中矩形框为人工标注边界框,人脸在尺度、姿态、遮挡、表情、装扮、光照等方面都有很大差异,人脸特征具有多样性。在训练模型时,根据图像的识别难度将数据集划分为 3 个子集,其中 50% 作为测试集,40% 作为训练集,10% 作为验证集。训练模型正负样本的处理参考原始 MTCNN 的训练方法。为验证 MTCNN 改进模型的性能,实验分别采用 ShuffleNet 中的 shuffle-channel 和 Xception 中的 depthwise 思想进行模型训练和测试,实验结果如表 1 所示,其中第 2 组比第 3 组速度慢的原因为 1×1 卷积占据了 depthwise 卷积 80%~90% 的计算量,shuffle-channel 正好解决了计算量和不同通道通信的问题,可以看出 shuffle-channel 方法具有更高的效率。在驾驶环境下,MTCNN-shuffle 人脸检测模型的检测效果如图 9 所示。



图 8 WIDER FACE 数据集样本
Fig.8 WIDER FACE dataset sample

表 1 MTCNN 及其改进模型的比较
Table 1 Comparison of MTCNN and its improved models

组号	模型	速度/(frame · s ⁻¹)
1	MTCNN	25
2	MTCNN-DW	138
3	MTCNN-shuffle	231

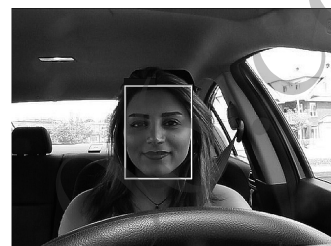


图 9 驾驶环境中的人脸检测
Fig.9 Face detection in the driving environment

6.1.2 人脸关键点检测模型

人脸关键点检测模型 PFLD 使用 300W^[25] 和 AFLW^[26] 两个基准数据集训练和测试模型。300 W 是用于人脸关键点检测的竞赛数据集,其中每幅图像的每张人脸有 68 个关键点。300W 数据集包含 LFPW^[27]、AFW^[28]、HELEN^[29] 及 IBUG^[30] 4 个数据集,共 3 837 幅人脸图像。实验使用 3 148 幅图像作为训练集,689 幅图像作为测试集。具体地,测试集分为 3 个子集:1) 具有挑战性的集合 (IBUG 中的 135 幅图像);2) 一般集合 (LFPW 测试集中的 224 幅图像,HELEN 测试集中的 330 幅图像,共 554 幅图像);3) 全集 (全部测试图像,共 689 幅图像)。AFLW 数据集提供了大规模人脸图像集合,每张人脸面部有 21 个标志,表现出各种各样的外观及一般的成像和环境条件。实验分别使用 20 000 幅图像和 4 386 幅图像作为训练集和测试集。

在训练期间,所有人脸图像被裁剪或重置大小至 112×112 。实验采用深度学习框架 TensorFlow,批量大小 (BatchSize) 设置为 512,优化器为 Adam,权重衰减设置为 $10e-6$,动量为 0.9,最大迭代次数为 70 000 次,整个训练过程的学习率固定为 10^{-4} 。训练硬件环境为单卡 Nvidia GTX 2060 GPU。对于

300W 数据集,实验采用数据增强技术,如对每个样本水平翻转并以 5° 为间隔从 -30° 至 30° 对图像样本进行旋转。考虑到真实驾驶环境,进一步增强模型泛化能力,并对每个样例中的人脸图像做了 20% 的随机遮挡。对于 AFLW 数据集,实验没有使用任何数据增强技术而是直接输入原始图像。图 10 为驾驶环境下人脸关键点检测及定位出的眼部及嘴部。

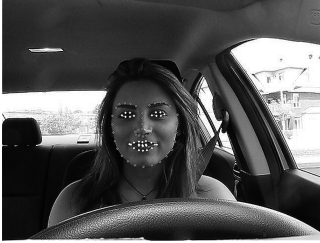


图 10 驾驶环境中的人脸关键点检测

Fig. 10 Face detection of keypoints in the driving environment

6.1.3 疲劳驾驶检测模型

疲劳驾驶检测模型使用 YawDD^[31] 数据集作为疲劳检测的标准数据集,该数据集包含不同性别、年龄的驾驶人员在不同光照环境下拍摄的模拟驾驶视频。驾驶人员的面部特征包括:1) 正常驾驶(不说话,标签为 Normal);2) 驾驶时说话或唱歌(标签为 Talking);3) 开车打哈欠(标签为 Yawning)。数据样本如图 11 所示。



(a)Normal

(b)Talking

(c)Yawning

图 11 YawDD 数据集样本

Fig. 11 YawDD dataset sample

由于数据集中的视频时间长短不一,为保证准确性,实验随机选取时间大于 30 s 的测试样本,并将视频的 70% 作为训练集,30% 作为测试集。同时,由于每一个视频中驾驶人员的眨眼、打哈欠及异常姿态次数并未标注,因此实验时寻找 10 位观察者记录每一个视频中驾驶人员的眨眼、打哈欠及异常姿态次数,最后取 10 位观察者的平均值作为真实数据,同时使用 PFLD 模型检测并记录 PERCLOS 均值,数据样本如表 2 所示。

表 2 基于 10 位观察者收集的数据样本

Table 2 Data samples collected from 10 observers

视频 ID	眼部状态		嘴部 打哈欠 次数	头部姿态		标签
	眨眼 次数	PERCLOS 均值		点头 频率	异常姿 态次数	
10	32	0.59	3	5	2	Yawning
17	41	0.39	0	0	0	Normal
20	39	0.35	0	1	1	Talking

由于疲劳特征的多样性,单一的疲劳特征很难作为疲劳驾驶的检测标准,因此疲劳驾驶检测模型采用多种面部疲劳特征相融合的策略,具体过程如下:

1) 获取各疲劳特征阈值与连续检测帧数。对单个视频而言,随机抽取样本分别得出眼部 EAR 阈值 α 、嘴部 MAR 阈值 β 、点头频率阈值 γ 、异常姿态阈值 θ 与连续检测视频帧数 F 的设定对相应检测准确率的影响。实验时所有阈值的设定范围为 0.35 ~ 0.75,间隔取值为 0.05; F 设定的值为 [10, 15, 20]。以眼部 EAR 阈值为例,准确率为检测到的眨眼次数与标注数据的比值,如图 12 所示。根据以上方法在整个训练集上进行实验,最后计算所有样本的眼部 EAR 阈值 α 、嘴部 MAR 阈值 β 、点头频率阈值 γ 、异常姿态阈值 θ 在准确率最高值时的平均值分别作为疲劳检测时的各特征阈值, F 选取准确率最高的值。

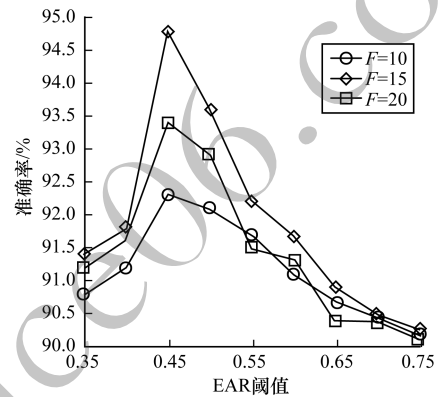


图 12 眼部 EAR 阈值与连续检测视频帧数对准确率的影响

Fig. 12 The influence of eye EAR threshold and the number of continuously video frame on accuracy

2) 基于决策树的疲劳驾驶预测与分类模型。根据以上 10 位观察者的标注数据,使用 sklearn 机器学习库构建决策树来预测疲劳驾驶检测模型,特征选择标准为 Gini 指数,输入为所有疲劳特征组成的向量,输出为连续检测给定帧数后判别驾驶人员为正常驾驶、驾驶时说话或唱歌及开车打哈欠这 3 种情况。

6.2 测试模型

在测试阶段,为验证模型准确率,实验中还检测并统计了测试样本的相关疲劳特征数据,如表 3 所示。可以看出,驾驶人员在疲劳驾驶时,PERCLOS 值、打哈欠次数及头部异常姿态都有所增加或加重。在测试集的所有视频中,本文疲劳驾驶检测模型给出的预警均未出现误判的情况。

表 3 测试样本的相关疲劳特征数据
Table 3 Related fatigue characteristic data of test samples

视频 ID	眼部状态				嘴部运动		头部姿态		预警次数
	实际眨眼次数	检测眨眼次数	准确率/%	PERCLOS 均值	实际打哈欠次数	检测打哈欠次数	实际异常姿态次数	检测异常姿态次数	
1	32	31	96.9	0.25	0	0	0	0	0
2	32	30	93.8	0.36	0	0	0	0	0
3	15	14	93.3	0.56	1	1	0	0	1
4	52	51	98.1	0.51	3	3	0	0	3
5	56	54	96.4	0.63	2	2	12	10	14
6	59	57	96.6	0.57	2	2	2	2	4

7 结束语

在复杂的驾驶环境下,鲁棒的人脸检测和人脸关键点检测是疲劳驾驶检测的前提,而疲劳特征的选取是疲劳驾驶检测的关键。为此,本文提出一种基于深度学习的驾驶疲劳检测算法,通过改进 MTCNN 模型提升人脸检测效率,使用 PFLD 模型保证人脸关键点检测鲁棒性,采用基于决策树构建的疲劳驾驶预测与分类模型提高疲劳驾驶检测准确率。实验结果表明,该算法在疲劳驾驶检测准确率和效率方面均有较大的性能提升。后续将通过解决视频样本中驾驶人员的极端与异常头部姿态估计问题,进一步提高疲劳驾驶检测算法的实时性和准确率。

参考文献

- [1] PAN Zhigeng, LIU Rongfei, ZHANG Mingmin. Research on fatigue driving detection algorithm based on fuzzy comprehensive evaluation [J]. Journal of Software, 2019, 30(10): 2954-2963. (in Chinese)
潘志庚,刘荣飞,张明敏.基于模糊综合评价的疲劳驾驶检测算法研究[J].软件学报,2019,30(10): 2954-2963.
- [2] DINGES D F, GRACE R. PERCLOS: a valid psychophysiological measure of alertness as assessed by psychomotor vigilance: FHWA-MCRT-98-006 [R]. Washington D. C., USA: Federal Highway Administration, 1998.
- [3] JIE Z, MAHMOUD M, STAFFORD-FRASER Q, et al. Analysis of yawning behavior in spontaneous expressions of drowsy drivers [C]//Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition. Washington D. C., USA: IEEE Press, 2018: 571-576.
- [4] MIN Qiusha, LIU Neng, CHEN Yating, et al. Head pose estimation based on facial feature point localization [J]. Computer Engineering, 2018, 44(6): 263-269. (in Chinese)
闵秋莎,刘能,陈雅婷,等.基于面部特征点定位的头部姿态估计[J].计算机工程,2018,44(6): 263-269.
- [5] RUIZ N, CHONG E, REHG J M. Fine-grained head pose estimation without keypoints [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Press, 2018: 2074-2083.
- [6] HUANG Baisheng, JIANG Tao, ZHOU Weiguang. Performance analysis of HRRP ship target recognition based on deep learning [J]. Modern Radar, 2020, 41(2): 54-57. (in Chinese)
黄柏圣,江涛,周伟光.基于深度学习的HRRP舰船目标识别性能分析[J].现代雷达,2020,41(2): 54-57.
- [7] LI Shiguo, ZHANG Ruiguo, SUN Jingming, et al. A study on the architecture of radar ATR based on deep learning [J]. Modern Radar, 2019, 40(11): 57-61. (in Chinese)
李士国,张瑞国,孙晶明,等.基于深度学习的雷达自动目标识别架构研究[J].现代雷达,2019,40(11): 57-61.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [EB/OL]. [2019-08-04]. <https://arxiv.org/abs/1506.01497>.
- [9] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. [2019-08-04]. <https://arxiv.org/abs/1804.02767>.
- [10] CHOI J, CHANG H J, FISCHER T, et al. Context-aware deep feature compression for high-speed visual tracking [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 479-488.
- [11] LU Huchuan, LI Peixia, WANG Dong. Visual object tracking: a survey [J]. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 61-76. (in Chinese)
卢湖川,李佩霞,王栋.目标跟踪算法综述[J].模式识别与人工智能,2018,31(1): 61-76.
- [12] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: a unified embedding for face recognition and clustering [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 815-823.
- [13] ZOU Guofeng, FU Guixia, GAO Mingliang, et al. Pose varied face recognition based on self learning deep convolutional neural network [J]. Journal of Chinese Computer Systems, 2018, 39(6): 1156-1162. (in Chinese)
邹国锋,傅桂霞,高明亮,等.基于自学习深度卷积神经网络的姿态变化人脸识别[J].小型微型计算机系统

- 统,2018,39(6):1156-1162.
- [14] ANDERSON P, HE X, BUEHLER C, et al. Bottom-up and top-down attention for image captioning and casual question answering[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 6077-6086.
- [15] ZHANG K P, ZHANG Z P, LI Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [16] GUO Xiaojie, LI Siyuan, YU Jinke, et al. PFLD: a practical facial landmark detector[EB/OL]. [2019-08-04]. <https://arxiv.org/abs/1902.10859>.
- [17] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 6848-6856.
- [18] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1251-1258.
- [19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2019-08-04]. <https://arxiv.org/abs/1409.1556>.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [21] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 4510-4520.
- [22] HU G S, YAN F, JOSEF K, et al. Efficient 3D morphable face model fitting[J]. Pattern Recognition, 67(C): 366-379.
- [23] FERRARIO V F, SFORZA C, SERRAO G, et al. Active range of motion of the head and cervical spine: a three-dimensional investigation in healthy young adults[J]. Journal of Orthopaedic Research, 2002, 20(1): 122-129.
- [24] YANG S, LUO P, LOY C C, et al. WIDER FACE: a face detection benchmark[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 5525-5533.
- [25] SAGONAS C, ANTONAKOS E, TZIMIROPOULOS G, et al. 300 faces in-the-wild challenge: database and results[J]. Image and Vision Computing, 2016, 47: 3-18.
- [26] KOESTINGER M, WOHLHART P, ROTH P M, et al. Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization[C]//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops. Washington D. C., USA: IEEE Press, 2011: 2144-2151.
- [27] BELHUMEUR P N, JACOBS D W, KRIEGMAN D J, et al. Localizing parts of faces using a consensus of exemplars[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2930-2940.
- [28] ZHU X, RAMANAN D. Face detection, pose estimation, and landmark localization in the wild[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 2879-2886.
- [29] LE V, BRANDT J, LIN Z, et al. Interactive facial feature localization[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2012: 679-692.
- [30] SAGONAS C, TZIMIROPOULOS G, ZAFEIRIOU S, et al. 300 faces in-the-wild challenge: the first facial landmark localization challenge[C]//Proceedings of IEEE International Conference on Computer Vision Workshops. Washington D. C., USA: IEEE Press, 2013: 397-403.
- [31] ABTAHI S, OMIDYEGANEH M, SHIRMOHAMMADI S, et al. YawDD: a yawning detection dataset[C]//Proceedings of the 5th ACM Multimedia Systems Conference. New York, USA: ACM Press, 2014: 24-28.