

中图分类号： TP391.41

论文编号： _____

学科分类号： 520.6040

密 级： 公 开

安徽理工大学

专业学位硕士学位论文

(全日制)

多特征融合车载疲劳检测方法研究

作者姓名： 叶华洲

专业名称： 计算机技术

研究方向： 图像处理

导师姓名： 王军号 高级实验师

导师单位： 安徽理工大学

答辩委员会主席： 朱家兵

论文答辩日期： 2022 年 12 月 04 日

安徽理工大学研究生院

2022 年 12 月 10 日

A Dissertation in computer technology

**Research on multi-feature fusion vehicle fatigue
detection method**

Candidate: Ye Huazhou

Supervisor: Wang Junhao

Computer Science and Engineering School

Anhui University of Science and Technology

No. 168, Taifeng Street, Huainan, 232001, Anhui, P. R. China

摘 要

随着经济的发展，私家车的数量不断上升，我国每年的交通事故数量也随之攀升。相关报道显示，以疲劳驾驶为主要原因的交通事故数量占总体事故数量的20%左右。若可以在驾驶途中及时对驾驶员进行疲劳警示，就能降低由疲劳驾驶造成的交通事故数量。在目前的疲劳检测算法中，基于驾驶员生理信号特征的疲劳检测方法会干扰驾驶员，使其本身成为危害驾驶安全的因素。基于车辆行为特征的疲劳检测受路况、天气影响，且不同车辆的参数不同，其鲁棒性较低。基于驾驶员行为特征的疲劳检测，结合一定的图像处理技术，可以在不干扰驾驶员的前提下进行疲劳检测，且不受环境、车辆参数的影响，具有高精度、高鲁棒性的优点。本文研究了一种分析驾驶员面部行为特征判断驾驶员疲劳状态的方法，研究内容主要有以下几方面：

1.疲劳特征区域定位。优化了多特征提取流程，提出了多任务卷积网络从图像中检测人脸区域并获得头部姿态，使用 ERT 算法获得人脸特征点定位眼部区域与嘴部区域。多任务卷积神经网络将人脸分类、边界框回归与头部姿态估计任务融合到级联网络中，人脸分类任务准确率达到 99.2%，边界框回归任务在 WIDER FACE 官方评价中取得 0.665 的分数，头部姿态估计任务平均绝对误差为 8.021。相较于使用 MTCNN+FAS-Net 的方式，每秒检测帧数由 12FPS 提升至 25FPS。

2.疲劳参数提取。获取特征区域后，在眼部区域使用眼部纵横比获取眨眼信息，使用 PERCLOS 均值与眨眼次数衡量眼部疲劳状态。在嘴部区域使用嘴部纵横比获取哈欠信息，使用哈欠次数衡量嘴部疲劳状态。在头部使用头部姿态（俯仰角、偏航角和滚转角）判断点头与歪头动作，使用点头次数与异常姿态时长衡量头部疲劳状态。

3.多特征融合疲劳检测。因为单种疲劳指标存在准确率低，容易误判的问题，本文采取多特征融合的方法进行疲劳状态判断。使用以 CART 树为基决策树的随机森林作为多特征融合方法，将 5 种疲劳参数作为特征。对 UTA 数据集进行处理，以 30s 的短视频为样本，选取其中部分作为训练集与测试集。实验结果表明，本文采取的疲劳检测方法可以达到 96.2%的准确率，对于疲劳驾驶检测具备较好的可行性。

图 36 表 13 参 70

关键词：疲劳驾驶检测；CART 树；特征融合；PERCLOS；ERT

分类号：TP39；

Abstract

With the development of economy, the number of private cars is rising, and the number of traffic accidents per year in China is also rising. Relevant reports show that the number of traffic accidents with fatigue driving as the main reason accounts for about 20% of the total number of accidents. If the driver can be alerted to fatigue in time while driving, the number of traffic accidents caused by fatigued driving can be reduced. In the current fatigue detection algorithm, the fatigue detection method based on the characteristics of the driver's physiological signal will interfere with the driver, making it a factor endangering driving safety. Fatigue detection based on vehicle behavior characteristics is affected by road conditions and weather, and the parameters of different vehicles are different, so its robustness is low. Fatigue detection based on driver behavior characteristics, combined with certain image processing technology, can perform fatigue detection without disturbing the driver, and is not affected by the environment and vehicle parameters, so it has the advantages of high precision and high robustness. This paper studies a method to analyze the driver's facial behavior characteristics to judge the driver's fatigue state. The research content mainly includes the following aspects:

1. Fatigue feature area localization. Optimized multi-feature extraction process, proposed multi-task convolutional network to detect face regions from images and obtain head pose, use the ERT algorithm to obtain the face feature points to locate the eye area and the mouth area. The multi-task convolutional neural network integrates face classification, bounding box regression and head pose estimation tasks. The multi-task convolutional neural network has an accuracy rate of 99.2% for face tasks, the bounding box regression task achieves a score of 0.665 in the official WIDER FACE evaluation, and the average absolute error of the head pose estimation task is 8.021. Compared with using MTCNN+FAS-Net, the frames per second detection is increased from 12FPS to 25FPS.

2. Fatigue parameter extraction. After obtaining the feature area, use the eye aspect ratio to obtain the blink information in the eye area, and use the PERCLOS mean and the number of blinks to measure the eye fatigue state. Use the mouth aspect ratio to obtain yawn information in the mouth area, and use the number of yawns to measure

mouth fatigue status. The head posture (pitch angle, yaw angle, and roll angle) was used to judge the nodding and tilting of the head, and the number of nods and the duration of abnormal postures were used to measure the fatigue state of the head.

3. Multi-feature fusion fatigue detection. Because a single fatigue index has the problem of low accuracy and easy misjudgment, this paper adopts the method of multi-feature fusion to judge the fatigue state. Random forest based on CART tree is used as a multi-feature fusion method, and 5 kinds of fatigue parameters are used as features. Process the UTA data set, take 30s short videos as samples, and select some of them as training sets and test sets. The experimental results show that the fatigue detection method adopted in this paper can achieve an accuracy of 96.2%, which has good feasibility for fatigue driving detection.

Figure 36 Table 13 Reference 70

Keywords: fatigue driving detection; CART tree; feature fusion; PERCLOS; ERT
Chinese Books Catalog:TP39

目 录

1 绪论.....	1
1.1 课题研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 基于驾驶员生理信号特征的疲劳检测	4
1.2.2 基于车辆行为特征的疲劳检测.....	5
1.2.3 基于驾驶员行为特征的疲劳检测.....	5
1.3 本文研究内容.....	6
1.4 论文的组织结构.....	7
2 相关理论与技术分析	8
2.1 图像增强.....	8
2.2 卷积神经网络在人脸检测中的应用	9
2.3 人脸特征点检测算法.....	10
2.4 本章小结.....	11
3 疲劳特征区域定位	12
3.1 疲劳特征区域定位流程优化	12
3.2 图像预处理.....	14
3.2.1 图像去噪.....	14
3.2.2 暗光图像处理.....	16
3.3 多任务卷积神经网络.....	17
3.3.1 网络任务与检测流程.....	17
3.3.2 网络结构.....	18
3.3.3 损失函数.....	23
3.3.4 网络训练.....	25
3.3.5 训练结果与网络性能分析.....	30
3.4 人脸特征点检测.....	32
3.4.1 建立模型	32
3.4.2 模型训练.....	34
3.5 本章小结.....	35
4 疲劳参数提取	37
4.1 眼部疲劳参数.....	37
4.1.1 眨眼判断.....	37
4.1.2 PERCLOS 参数获取	39
4.1.3 眨眼次数获取.....	40
4.2 嘴部疲劳参数.....	41

4.3 头部姿态疲劳参数.....	43
4.3.1 点头次数获取.....	43
4.3.2 头部异常姿态时长获取.....	44
4.4 本章小结.....	45
5 多特征融合疲劳检测	46
5.1 疲劳状态评价.....	46
5.2 多特征融合疲劳检测算法	47
5.1.1 CART 树	47
5.1.2 随机森林.....	49
5.3 实验结果与分析.....	50
6 总结与展望.....	53
6.1 总结.....	53
6.2 展望.....	54
参考文献.....	55

1 绪论

1.1 课题研究背景与意义

随着我国经济发展水平的提高,人民越来越富裕,对生活的品质愈加关注。私家车作为一种可以随时出行、体验舒适的交通工具成为每个家庭的必备品。国家统计局数据显示,2020 年我国每百户保有家用汽车 37.1 辆^[1]。从 2014 到 2020 年我国民用汽车和机动车驾驶员数量稳步增长,详细情况如表 1 所示。

表 1 2014-2020 年全国车辆和驾驶员数量统计表

Table1 The number of vehicles and drivers in China from 2014 to 2020

年份	民用汽车拥有量/万辆	机动车驾驶员/万人
2014	14 598.11	29 892.32
2015	16 284.45	32 853.05
2016	18 574.54	35 876.98
2017	20 906.67	36 016.94
2018	23 231.23	41 030.16
2019	25 376.38	43 636.74
2020	27 340.92	45 702.49

汽车数量以及机动车驾驶员数量不断增加,在如此庞大基数的情况下,道路交通事故的数量也不断增加。2020 年共发生 244671 起交通事故,其中机动车事故 211074 起,造成 55950 人死亡^[1]。复杂交通环境与极端天气是引发交通事故的客观因素,驾驶员的危险驾驶行为是引发交通事故的主观因素,疲劳驾驶就属于危险驾驶行为的一种。Gander^[2]和 Forsman^[3]等人研究发现,疲劳驾驶是诱发交通事故的重要因素,相比于正常驾驶状态,疲劳状态下发生交通事故的几率高出 4 到 6 倍。每年以疲劳驾驶为主要原因的交通事故数量占全部交通事故数量的 10%~20%,占特大交通事故数量的 40%以上^[4]。

驾驶员在驾驶中长期保持固定姿势会引起血液循环不畅,进而引起肢体疲劳,长时间观察路况,精神缺少放松时间进而引起注意力涣散。生理与心理状态的恶化导致驾驶员在遭遇紧急情况时反应能力下降,极大增加了交通事故发生的概率。迫于生计或存有侥幸心理,部分驾驶员在行驶过程中即使意识到自己已经陷入疲劳状态,对车辆的操控能力下降,仍会强忍疲劳继续行驶。更多的驾驶员则是在车辆行驶过程中,没有意识到自己正在疲劳驾驶,甚至已经进入深度疲劳状态仍

不自知，在无意识情况下操纵车辆，引发单方事故的风险急剧升高，也更容易引发重大交通事故。当驾驶员处于疲劳时，会进入以下危险状态：

- 1.打哈欠，表情僵硬；
- 2.视线模糊，漏看交通指示灯、指示牌等信息；
- 3.感觉眼睛难以睁开，睁眼时眼睛睁开程度变小；
- 4.眼睛发红、发干，频繁揉眼；
- 5.感觉头部沉重，无法保持头部水平，甚至频繁点头；
- 6.注意力难以集中，接受信息并判断的速度下降；
- 7.对肢体的操控能力下降，无法保持方向盘稳定，车辆频繁发生小幅转向；
- 8.方向感与距离感下降，对车身转向幅度、车间距的把控能力降低；
- 9.对车辆速度感知能力下降，无意识的加速或减速，速度不稳定。

状态 1、4 与状态 5 使司机在短时间内失去车辆前方视野，在回复视野的窗口期遭遇突发状况时司机无法及时反应，从而酿成大祸；司机处于状态 2 下可能进行违规变道、闯红灯等极易引发交通事故的行为；状态 7、8 与状态 9 令司机的驾驶轨迹飘忽不定，对其他驾驶者造成严重威胁。疲劳状态会降低司机对车辆的操作能力，大大提高交通事故的发生率。大型交通与铁路机构会制定科学的排班制度并对司机定期培训与警示，这些受雇的司机知悉疲劳驾驶的危害，受制于严厉的惩罚机制，依赖科学严谨的排版与作息，能够尽量避免在行驶过程中进入疲劳状态。但是更多的私家车驾驶员、个人经营的货车驾驶员不够重视自己的驾驶状态，往往也不能发现自己已经疲惫不堪，因此开发一种可靠的、实时的车载疲劳驾驶检测系统是必要的。该系统能够准确及时的对驾驶员的疲劳程度进行判定并提醒驾驶员，从而减少因疲劳驾驶发生的交通事故，既能保障驾驶员自身安全，也提高了交通安全系数。

1.2 国内外研究现状

针对疲劳状态的检测，尤其是针对驾驶员的疲劳检测一直是研究的重点。疲劳驾驶检测涉及到生理学、心理学、图像处理、运动跟踪、模式识别等多个领域，是一个复杂而又同时兼具理论与现实价值的研究课题^[5]。在长时间驾驶中，驾驶员的疲劳随驾驶时间逐渐加深，若能实时检测驾驶员疲劳状态并采取警示、降速和人工智能接管等操作，可以大大减低交通事故发生率。

疲劳检测根据检测方式可以分为俩类：主观评测法和客观评测法。

1.主观评测法

主观评测法指通过他人问询、观测或被测人自述的感受、填写的问卷等方式获得的主观评价来判断被测人疲劳状态的方法。主观测评法的具体操作方式主要有以下三种：

1) 主观问卷方式：精心设计好调查问卷与相应的评估方法，由被测人填写调查问卷并通过评估方法得出结论。主观问卷方式也被用于精神状态评估、世界观评估等领域。

2) 自我评价方式：通过被测人叙述自身的感受评估被测人疲劳状态。自我评价方式由受测人凭感官口述，受测人主观意愿会极大影响评价结果。

3) 反馈应答方式：通过对被测人进行人工问询，观察其反应与回答判断被测人是否处于疲劳状态。

主观评测法简单易用，但其准确性不高，不同个体的差异很大，每个人对疲劳的认识不同，且存在受测人主观谎报的情况。主观测评法一般用于事后分析，在事故发生后划分事故责任。

2.客观评测法

客观评测法指通过传感器、图像采集处理设备 etc 器械获取受测人的生理信息或行为特征推断受测人疲劳程度的方法。在针对驾驶员的疲劳检测方法中，还可以通过车辆信息推断驾驶员疲劳状态。针对驾驶员的疲劳检测难题，国内外研究人员提出了四种基于不同技术的解决方案：

1) 基于驾驶员生理信号特征的疲劳检测。医学上研究人体生理信号如肌电信号（EMG）、脑电信号（EEG）和心电信号（ECG）等发现，人体生理信号在清醒状态与疲劳状态之间有较明显区别，在此基础上通过研究疲劳时生理信号的特征来判断受测人的疲劳状态；

2) 基于车辆行为特征的疲劳检测。驾驶员的疲劳会反映在其对车辆的操控上，通过检测诸如方向盘偏转幅度、车辆加速度、车身与车道的夹角等参数来判断驾驶员的疲劳状态；

3) 基于驾驶员行为特征的疲劳检测。当驾驶员生理发生变化时，其行为往往也会发生变化，表现出一些明显的特征，如打哈欠、揉眼睛、点头等。可以通过计算机视觉的方法提取出这些行为特征用来判断受测人的疲劳状态。

4) 多源信息融合的疲劳检测。单一的检测指标准确率不高，且有可能出现指标中断导致检测失败的情况。为提高疲劳检测的准确度与鲁棒性，常常运用多种

检测方式进行检测，运用多种指标综合判断驾驶员的疲劳状态。

以下小节将介绍前三种客观评测法的研究现状。

1.2.1 基于驾驶员生理信号特征的疲劳检测

基于驾驶员生理信号的疲劳检测中使用的生理信号包含脑电波信号（EEG）、心电波信号（ECG）、眼电图（EOG）、肌电波信号（EMG）、脉搏和血氧饱和度（SpO2）等。

南姣芬^[6]分别使用了 S 变换、希尔伯特-黄变换以及高阶谱分析方法提取驾驶疲劳时的脑电特征，判断驾驶员的疲劳程度。实验表明当驾驶员连续驾驶超过 40 分钟后，驾驶员的精神活动开始减少，驾驶 40-60 分钟时间段内驾驶员视觉疲劳先于精神疲劳出现。王海玉^[7]研究多熵值特征下的 EEG 疲劳检测模型，通过对疲劳检测过程中的去噪、特征提取和分类等三个主要步骤进行分析研究，建立起三个准确率较高、特征稳定的疲劳检测分析模型。采用 ICA 对不同的主体进行处理，然后对样本熵、信息熵、模糊熵和 AR 系数进行特征提取，最终使用 AdaBoost 将最小二乘矢量积集成到基于三种内核分类器的强分类器中进行疲劳判断。王斐^[8]提出了基于电极-频率分布图和迁移学习策略的疲劳状态检测模型，相较于传统的方法，具有识别率高，特征提取快速的优点。EEG 信号检测的可靠性和灵敏度很好，其抗干扰性也很强，但是其检测成本高、装置结构复杂、可扩展性较差^[9]。

郭玮珍^[10]从心电信号中提取心率（HR）和心率变异性（HRV），并对其 5 项时频域指标进行分析。实验结果发现采用 5 项心电图时频域指标，可以对疲劳程度进行量化的反映和评价。郝贺^[11]提出一种 QRS 波群检测算法处理心电信号。QRS 波检测是心电信号检测的基础，通过检测心电信号中的 QRS 综合波和 R 峰值，可以得到有关心率、传导速度、心脏内各种组织状态和各种不正常状况的信息，对于心电信号的分析具有十分重要的意义。

眼电图是通过在眼睛上下方放置的电极，获得 EOG（electrooculogram）在垂直方向上的变化。Ohsuga M^[12]通过研究眨眼过程的 EOG 变化情况得出，在闭眼时 EOG 达到峰尖幅值，EOG 上升时间长度与睁眼持久度相关，EOG 下降时间长度与闭眼持久度相关。在此基础上，Ohsuga M 通过对 EOG 波形中峰尖时间，上升时间与下降时间进行聚类，得到不同精神状态下的 EOG 波形图，以此来判断受测人疲劳状态。

1.2.2 基于车辆行为特征的疲劳检测

基于车辆行为特征的疲劳检测属于间接疲劳检测方法。基于车辆行为特征的疲劳检测方法通过采集车辆信息，使用统计学分析与机器学习的方法推测驾驶员的疲劳状态。间接疲劳检测的精准度不高，但不需要对车辆进行大幅度改造，不会提高车辆制造成本，且该方法对驾驶员透明，不会对驾驶员造成任何影响，因此各大汽车生产商纷纷投入研发，并已产品化。

梅赛德斯-奔驰公司开发了一套名为 **Attention Assist** 的车载疲劳检测系统。该系统基于车辆状态参数来推断驾驶员驾驶行为与状态，例如车速、发动机转速、横摆角速度等及各信号的后处理参数，综合考虑以上因素进行分析计算得到驾驶员状态监测结果^[13]。大众则通过转向角传感器和电子助力转向系统收集驾驶员转向动作信息来确定其是否处于疲劳状态。在行驶过程中驾驶员疲劳探测系统会不断更新参数，计算出疲劳指数。日产公司利用电动助力转向(**Electric Power Steering, EPS**)系统的转向盘转角与角速度、转向盘力矩信号，结合车辆状态信号和车内环境信息判断驾驶员是否疲劳。

澳大利亚国际大学开发的 **DAS (driver assistance systems)**^[14]系统，可以检测车辆盲点、检测车辆间的距离辅助会车、检测驾驶员对车道保持的状态以及方向盘握力来判断疲劳。**Pilutti T**^[15]等人使用道路偏离警告和干预系统的背景下进行驾驶员状态评估，采用车辆侧边与道路边沿的距离作为评估算法的输入，方向盘方向作为输出。在行驶中算法不断在线更新参数，模拟驾驶的结果表明使用车辆侧边与道路边沿的距离作为评价指标是有效的。**Bertozzi M**^[16]等人开发了一种基于立体视觉的硬件和软件系统，可以检测障碍物和道路航线来判断车辆是否偏航，以此判断驾驶员状态。**Mao M**^[17]等人通过收集车辆驾驶时的车速、方向盘角度、状态加速器以及转向灯的状态，使用离散小波变换处理数据，得出疲劳驾驶状态与正常驾驶状态的参数区别，借此判断驾驶员状态。

1.2.3 基于驾驶员行为特征的疲劳检测

基于驾驶员行为特征的疲劳检测方法通过检测驾驶员疲劳时的外在行为表现来判断其疲劳程度。

早期的基于驾驶员行为特征的疲劳检测系统需要对车辆进行较大改装。**ASCI (Advanced Safety Concepts Inc.)**^[18]研制了一种定位驾驶员头部的装置，该装置安置在座椅正上方，通过电容传感器阵列定位驾驶员头部，并通过头部位置变化

信息，即点头摇头等动作判断其是否处于疲劳状态。该系统需提前在驾驶员头部贴上标志才可使仪器定位头部，影响驾驶员状态，且在座椅上方安装定位装置也需要对车辆进行大幅改造。

基于计算机视觉的疲劳检测方法随着深度学习的兴起得以盛行，只需在车内安装摄像头便可获得驾驶员的多种行为特征。邹昕彤^[19]通过 Adaboost 算法在视频中检测出人脸区域，之后使用灰度直方图统计特征值的方法判断眼部状态，使用似圆度判断嘴部打哈欠情况，通过眼部与嘴部疲劳状态综合判断受测人疲劳状态。刘忠旭^[20]使用 MTCNN 网络获得人脸区域，之后通过 Shi-Tomasi 算法进行眼睛角点定位。在获得人脸区域与眼睛角点后，截取并校正眼部与嘴部图像并将其分别输入眼部状态识别模型与嘴部状态识别模型。获得眼部与嘴部状态后，融合打哈欠参数、PERCLOS 值与闭眼时长三个参数判断驾驶员疲劳状态。邬敏杰^[21]通过 harr 特征得到面部区域，使用人眼模板匹配眼睛位置，得到人眼位置后，通过俗称“三庭五眼”人脸的几何特征对嘴部区域进行搜寻。获得人眼 PERCLOS 值与嘴部动作频率后，使用加权平均法求得最终状态参数判断疲劳状态。

1.3 本文研究内容

结合国内外研究现状，在三种检测方法中，基于驾驶员生理信号的疲劳检测方法需要驾驶员穿戴检测设备，接触式的生理信号获得方式会干扰驾驶员正常驾驶；基于车辆行为特征的疲劳检测受到道路状况的干扰，且不同车辆正常驾驶时参数范围不同，泛用性差；基于驾驶员行为特征的疲劳检测方式与计算机视觉结合后，使用摄像头这种非接触式设备获得驾驶员状态信息，不受路况限制，不同车辆进行简单改造即可使用，泛用性强。本文研究内容主要有以下几个方面：

1. 优化疲劳检测流程，采取合并任务的方式缩短疲劳特征区域定位流程。
2. 研究一种可以同时检测出人脸区域和头部偏转角的多任务卷积神经网络。使用一种人脸特征点定位算法定位人眼区域与嘴部区域。
3. 研究如何通过人脸特征点获得眼部与嘴部状态，通过头部偏转角获得头部状态。研究如何获得合适的前述状态的参数化表示用于疲劳检测。
4. 研究一种多特征融合的疲劳检测方法。使用以 CART 决策树为基决策器的随机森林综合单位时间内 PERCLOS 均值、眨眼次数、哈欠次数、点头次数和头部异常姿态时长来判断驾驶员疲劳程度。

1.4 论文的组织结构

本文共分六个章节，章节内容如下：

第 1 章绪论，介绍论文的研究背景及意义。阐述当前疲劳检测技术的分类，分析基于驾驶员生理信号的疲劳检测、基于机动车行为特征的疲劳检测和基于驾驶员行为特征的疲劳检测的研究现状。明确研究内容，结合论文完成的任务确定论文主要的工作。

第 2 章相关理论和技术分析，介绍常见的图像增强算法及其原理；介绍卷积神经网络在人脸检测任务中的应用；介绍人脸特征点检测的几种方式。

第 3 章疲劳特征区域定位，介绍优化后的疲劳检测流程，疲劳特征区域定位方法；介绍适用于本文的图像增强算法；介绍多任务卷积神经网络的构成，网络训练结果及其在数据集上的表现；介绍人脸特征点检测方法。

第 4 章疲劳参数提取，介绍在第三章获取疲劳特征区域的基础上，从图像中提取疲劳参数并进行合适的数值化表示。

第 5 章多特征融合疲劳检测，介绍使用以 CART 决策树为基决策器的随机森林综合单位时间内 PERCLOS 均值、眨眼次数、哈欠次数、点头次数和头部异常姿态时长来判断驾驶员疲劳程度。

第 6 章总结与展望，对本课题中所完成的工作进行总结，并对下一步的研究工作做出展望。

2 相关理论与技术分析

2.1 图像增强

图像增强伴随数字图像的诞生而出现，用于提高成像不清晰的图像质量或突出图像的某些特征，使得图像更易被人眼或是检测算法识别，提高可读性。图像增强算法在软件方向可以分为两大类：空域增强处理法(Spatial-based Domain Enhancement)和频域增强处理法(Frequency-based Domain Enhancement)。多数空域增强算法的原理是对图像本身进行处理，包括直方图变换、灰度变换、滤波器处理、模糊逻辑增强、基于遗传算法优化等。基于频域的图像增强处理方法先使用某种变换将图像转换成其他形式，再对变换后的形式进行某种处理，最后进行反变换将处理后的变换形式转换成图像来达到去除图像噪声的目的^[22]。常见的频域增强处理方法有傅里叶变换、小波变换、余弦变换等。

图像在生成过程中由于光线问题致使图像特征不明显，通常采用空域增强处理法来突显图像特征。

灰度变换是空域增强处理法中最基本的方法，直接对单个像素点进行操作。常用的灰度变换有：图像反转、Gamma 矫正（矫正图像数据来减轻显示器失真）、对比度拉伸（将灰度密集区域变换到较大范围）、灰度级分层（增强需要层级的灰度区域）等。直方图变换方法基于图像的灰度直方图对像素进行操作。直方图变换方法统计图像中每一级灰度出现的次数，也可以标准化为概率显示，在此基础上通过不同策略对像素进行操作获得不同的变换效果。直方图拉伸指对灰度集中的图像，将其灰度的分布区间由 $[1,255]$ 的子集拉伸至 $[1,255]$ ，从而加强图像对比度，提高可读性。直方图均衡化指将原始图像灰度直方图从某一比较集中的区间转化为在全区间上的均匀分布，相较于直方图拉伸，直方图均衡化不仅扩大了灰度范围，也对每一灰度上的像素点数量做出了限制。

图像在生成、传输和记录的过程中，由于电磁波的干扰、传感器自身的局限、通信系统的故障等因素，产生冗余和不必要的干扰信息，这就是图像噪声。图像噪声严重影响图像的视觉效果，干扰图像信息的提取，因此在进行边缘检测、图像分割、特征提取和模式识别等工作之前，采用适当的方法减少噪声是一项非常重要的图像预处理步骤^[23]。噪声根据其性质可以分为高斯噪声（白噪声）与脉冲噪声（椒盐噪声），高斯噪声的概率密度函数符合正态分布，脉冲噪声是离散型噪声的统称。

滤波器变换将图像看作一个二维信号，将像素值的灰度看作信号的强弱，使用信号增强的方式处理图像。常见的滤波器分为俩类：线性滤波器与非线性滤波器。线性滤波器对邻域中的像素的计算为线性运算，如利用窗口函数进行平滑加权求和的运算，或者某种卷积运算，都可以称为线性滤波，其代表有均值滤波、高斯滤波等。非线性滤波器产生的新图像与原图像具有某种逻辑关系。均值滤波去噪将数字图像中某一像素点的值用该点的一个邻域中各像素点的均值替换，在实践中可以很好的去除高斯噪声，对脉冲噪声效果不佳，同时会模糊图像的边缘信息，丢失图像细节；中值滤波去噪将数字图像中某一像素点的值用该点的一个邻域中各点像素点的中值替换，可以有效的消除脉冲噪声，但随着脉冲噪声密度的增大，中值滤波去噪也会丢失图像细节^[24]。

2.2 卷积神经网络在人脸检测中的应用

人脸检测是计算机视觉领域的一个重要分支，大量任务都有涉及人脸检测或需要人脸检测作为前置条件。身份认证任务需要首先获得人脸区域才能进行人脸特征匹配，人机交互任务需要获得人脸区域以识别人脸动作、表情作为交互的信息，头部姿态识别、人脸特征点检测等任务也都基于人脸检测的基础上。随着人脸检测的广泛应用，产品对人脸检测算法的性能要求也越来越高，实际使用场景中难以实现将人脸约束在某一姿势下，故无约束条件下人脸检测任务的准确率与检测速度受到科研人员的重视并作为独立课题进行研究。

基于卷积神经网络的人脸检测方法最早可以追溯到 1994 年^[25]，随着训练数据的大量增加，卷积神经网络在人脸检测任务上的表现愈发优秀。在卷积神经网络的基础上，不断有新的网络被发表。级联神经网络（Cascaded Neural Network）^[26]使用多级网络模型以期将多种网络的优点结合起来，采取多种不同的特征提取策略避免对特征学习的遗漏。在人脸检测任务中级联神经网络通过前几级网络筛选大量非人脸区域，极大的提高人脸识别速度。多任务卷积神经网络（multi-task convolutional neural networks）^[27]将人脸检测任务与人脸对齐任务同时进行，实验结果表明多任务同时进行相比单任务提高了准确率，说明同时进行有关联的任务时可以获得比单任务更好的效果。除了对网络本身的改进，UnitBox^[28]使用 IOU（重叠度）损失替代 L2 损失函数来进行网络训练，将表示人脸位置的四个点作为一个整体，使训练结果更加准确。由于 IOU 表示预测区域与准确区域的重叠度，其值介于[0,1]，这天然的归一化属性使其在多尺度任务上有更强的处理能力。Grid Loss^[29]针对人脸遮挡问题，提出将整个人脸部分划分为多个子区域，分别为这些

子区域训练独立的检测器。当人脸的某一部分被遮挡时，网络仍可以通过对未遮挡区域的检测定位到人脸区域。

人脸检测任务是单目标检测任务，因此有些更具普适性的目标检测算法也可以被用于人脸检测。区域卷积神经网络(Region-CNN)^[30]是第一个成功在目标检测任务上应用深度学习的算法。R-CNN 在图片上选出所有目标可能出现的区域框，对每个框提取特征、图像分类、非极大值抑制进行目标检测。由于候选区域过多，R-CNN 的识别速度极慢，为克服这一缺点，出现了很多基于 R-CNN 的改良网络：SPP-Net^[31]、Fast R-CNN^[32]、Faster R-CNN^[33]、YOLO^[34]等，这些网络均可以用于人脸检测。

部分人脸检测算法在 Fddb 数据集上的表现如表 2 所示。

表 2 部分人脸检测算法在 Fddb 数据集上的检测表现

Table2 Detection performance of some face detection algorithms on Fddb

算法	True positive rate	False positive
CascadeCNN ^[26]	0.8567	2000
Joint Cascade ^[35]	0.8667	2000
HeadHunter ^[36]	0.8808	2000
Grid Loss-big ^[29]	0.8940	1000
Faceness ^[37]	0.9098	2000
UnitBox ^[28]	0.9442	500
MTCNN ^[27]	0.9504	2000

2.3 人脸特征点检测算法

人脸特征点检测任务可以看作在给定人脸图像前提下，从人脸图像上定位一系列事先定义的人脸特征点。这些特征点通常用来描述人眼、鼻子、嘴唇和面部的轮廓。通过这些特征点可以对人脸属性进行分析，在性别识别、表情识别、年龄识别、面部动作捕捉等方面有广泛的应用。人脸特征点检测算法根据其使用的模型种类可以被分为两大类：基于参数模型的方法和基于非参数模型的方法。基于参数模型的方法假定人脸特征点的位置符合某种概率分布，如混合高斯分布、多元高斯分布等，事先人为设定参数数量。基于非参数模型的方法无须假定一个概率分布，这种方法并非没有参数，而是在训练过程中，参数量随着训练数据量的增加而增加^[38]。

主动形状模型（Active Shape Model）是经典的基于参数模型的人脸特征点检测算法，通常用于预测人脸 68 个特征点。Coote 等人^[39]使用每个特征点的 x, y 坐标构成一个 136 维的向量来描述人脸形状，在假定人脸形状满足各向同性高斯分布的前提下使用主成分分析（principal component analysis, PCA）对形状向量进行降维，在训练中不断使用新的人脸形状来更新形状向量。由于 ASM 迭代时仅考虑图像的局部信息，Coote 随后又提出主动外观模型（active appearance model, AAM）^[40]，使用图像金字塔从训练集中获得不同分辨率下的平均模型。AAM 不仅包括形状轮廓模型，还包括纹理轮廓模型，使其在图像分割方向取得较 ASM 更好的效果。

随着卷积神经网络被用于人脸检测并取得较好的效果，在人脸特征点检测领域也不断有基于卷积神经网络的方法被提出。Yi Sun 等^[41]首次将级联卷积神经网络用于人脸特征点定位，并在当年取得 state-of-art 的结果，这种级联的多层卷积神经网络又被称为 DCNN（Dynamic Convolution Neural Network）。DCNN 的具体结构如图 1 所示。

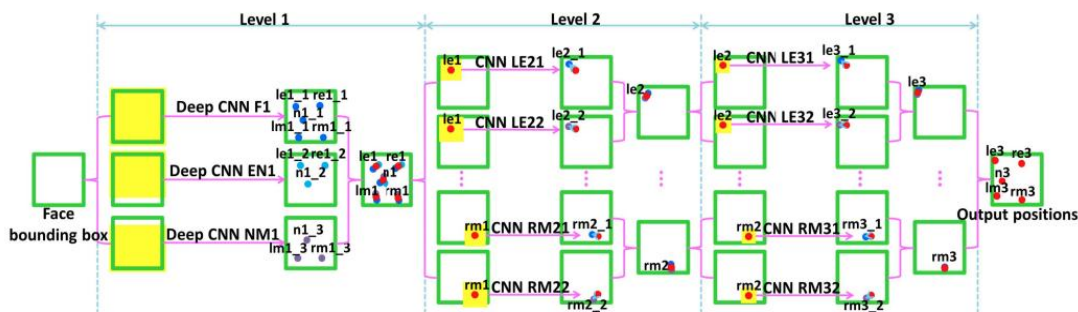


图 1 DCNN 结构图

Fig1 Architectures of DCNN

DCNN 共使用三个等级的网络完成不同的任务，最后输出精确的人脸关键点。人脸检测器得到人脸位置输入 Level1 网络，Level1 网络得到包含特征点的最小图像区域输入 Level2 网络。在 Level2 网络中，特征点被粗略的定位，通过这些被粗略定位的特征点裁剪出包含单个特征点的小区域图像，将这些小区域图像送至 Level3 网络中得到最终的特征点位置。

2.4 本章小结

本章包含三部分内容，一是图像增强的常用方法，介绍了图像增强的分类。二是卷积神经网络在人脸检测中的应用，介绍了卷积神经网络的发展与历代网络创新之处。三是人脸特征点检测算法，介绍了人脸特征点检测算法的发展与分类。

3 疲劳特征区域定位

3.1 疲劳特征区域定位流程优化

本文所需检测的疲劳特征包括眼部疲劳特征、嘴部疲劳特征与头部姿态疲劳特征，即需要定位驾驶员眼部区域图像，嘴部区域图像与头部图像并分析。在传统的疲劳检测流程中，都是在获得人脸区域的基础上，使用不同的方法定位眼部与嘴部区域并获得头部姿态，之后由这些原始图像信息获得各部分疲劳状态并进行综合判断。传统的疲劳检测流程如图 2 所示。

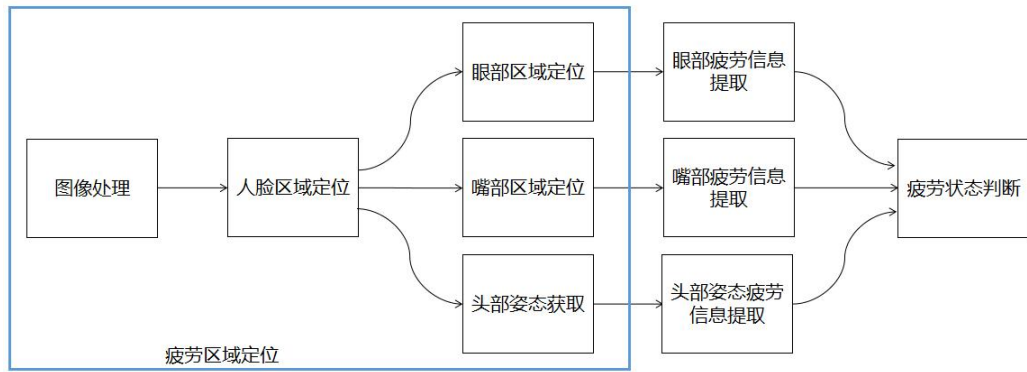


图 2 传统的疲劳检测流程图

Fig2 Traditional fatigue testing flow chart

图 2 中有蓝色方框包围的流程即为疲劳区域定位流程。文献^[42]使用 Haar-like 特征与 Adaboost 分类器进行人脸检测，使用标准人脸模板粗定位人眼位置。在对粗定位的图像进行处理后，使用边缘提取的方式获得眼睛部位的精确像素图，最后使用模糊综合评价判断眼睛睁闭情况。文献^[43]获得眼部区域精确像素图后，判断其最大外接矩形的大小来判断眼睛睁闭情况，监测人脸区域中心坐标判断驾驶员是否点头。文献^[44]使用分段高斯模型在 YCbCr 空间下进行肤色分割，通过唇色样本统计和聚类分析得到唇色的三重闭值，在可能的肤色区域中搜索嘴唇区域。文献^[45]使用 PFLD 算法同时检出人脸特征点与头部姿态，通过人脸特征点定位眼部与嘴部区域。

为了缩短疲劳检测流程，受 MTCNN 网络的构建思想启发，本文将人脸检测任务与头部姿态检测任务统一，使用级联的多任务卷积神经网络同时获得人脸区域与头部姿态。使用 ERT 人脸特征点检测方法获得 68 人脸特征点并以此定位眼部区域与嘴部区域。优化后的疲劳检测流程如图 3 所示。改进后，检测流程中的人脸检测与头部姿态检测合并为一个任务，人眼区域检测与嘴唇区域检测合并为

一个任务，总任务由 4 个变为 2 个，简化了检测流程，提高了检测效率。

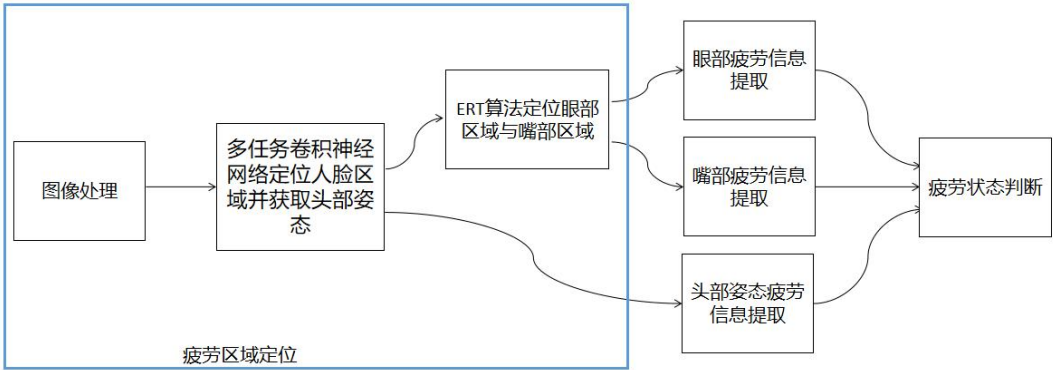


图 3 本文的疲劳检测流程图

Fig3 My fatigue testing flow chart

优化后的疲劳区域定位流程如图 4 所示。

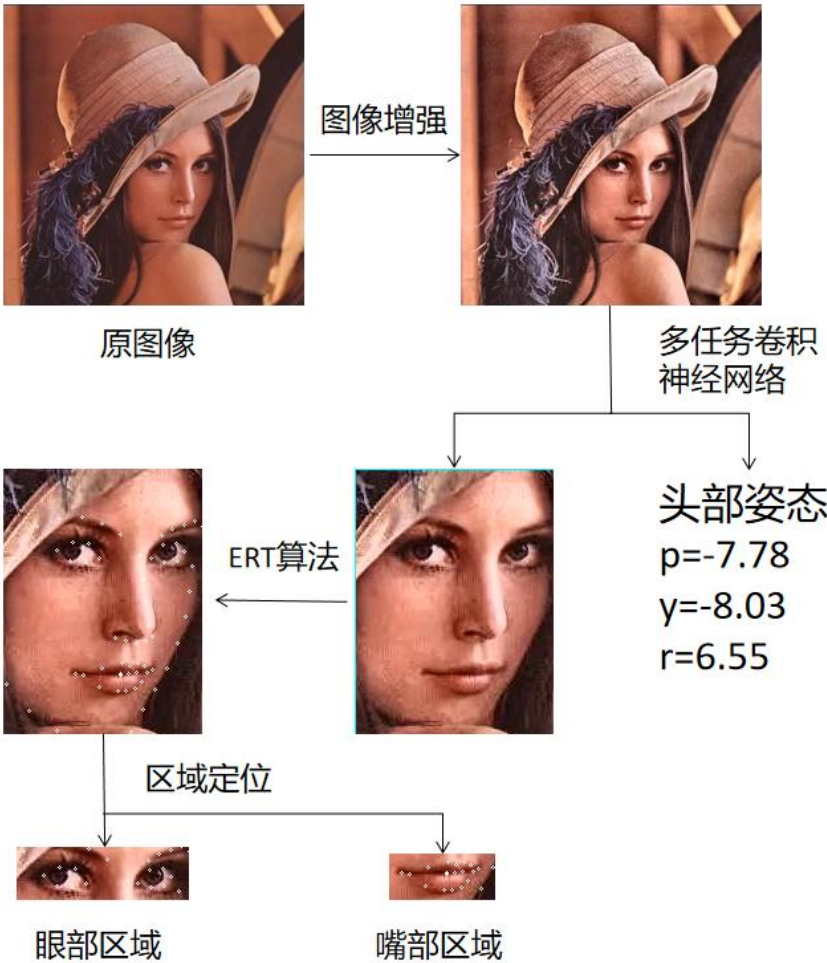


图 4 优化后的疲劳区域定位流程

Fig4 Optimized fatigue area positioning process

3.2 图像预处理

3.2.1 图像去噪

小波变换 (wavelet transform) 作为一种变换域信号处理方法, 在信号分析、图像处理、地震勘探和非线性科学等诸多领域得到了广泛应用^[46]。小波变换在图像处理领域具有低熵性、多尺度特性、去相关性与选基灵活性等优点^[47]。小波变换去噪的基本依据是信号点的频率与噪声点的频率不同, 一般而言信号点的频率较低而噪声点的频率较高, 图像的高频信号几乎由噪声组成, 将源信号进行小波变换后, 信号点的小波系数幅值较大, 噪声点的小波系数幅值较小, 此时将幅值小的小波系数置 0 即可去除噪声点。基于此原理, 小波变换去噪的步骤如下:

1. 二维信号的小波分解。对含噪信号 S 进行小波变换, 得到一组小波分解系数 w_{jk} ;

2. 设定一个阈值, 使用阈值函数对小波分解后的系数进行处理, 得到处理后的系数 \hat{w}_{jk} ;

3. 二维信号的小波重构。利用处理后的系数 \hat{w}_{jk} 计算二维信号的小波重构。

小波变换去噪的常用阈值函数可以分为硬阈值函数、软阈值函数和半软阈值函数。

硬阈值函数

$$\hat{w}_{jk} = \begin{cases} w_{jk} & |w_{jk}| > \lambda \\ 0 & |w_{jk}| \leq \lambda \end{cases} \quad (3-1)$$

软阈值函数

$$\hat{w}_{jk} = \begin{cases} \text{sign}(w_{jk})(|w_{jk}| - \lambda) & |w_{jk}| > \lambda \\ 0 & |w_{jk}| \leq \lambda \end{cases} \quad (3-2)$$

式中: λ 表示小波系数的阈值, $\text{sign}()$ 表示符号函数, w_{jk} 表示处理前的小波分解系数, \hat{w}_{jk} 表示处理后的小波分解系数。

半软阈值函数

$$\hat{w}_{jk} = \begin{cases} w_{jk} & |w_{jk}| > \lambda_1 \\ \text{sign}(w_{jk}) \frac{\lambda_2(|w_{jk}| - \lambda_2)}{\lambda_1 - \lambda_2} & \lambda_2 \leq |w_{jk}| \leq \lambda_1 \\ 0 & |w_{jk}| < \lambda_2 \end{cases} \quad (3-3)$$

式中: λ_1 表示小波系数的阈值, λ_2 表示小波系数的阈值, $\text{sign}()$ 表示符号函数, w_{jk}

表示处理前的小波分解系数， \hat{w}_{jk} 表示处理后的小波分解系数。

使用硬阈值函数可以较好的保留图像边缘等局部特征，但是会产生震铃，伪 Gibbs 效应等视觉失真，且小波系数在 $\pm\lambda$ 附近不连续。使用软阈值函数处理结果相对平滑，但由于改变了对应图像信号的小波系数，在重构后会丢失图像局部细节，造成图像边缘模糊。半软阈值函数在选择合适的阈值的前提下，可以在硬阈值函数与软阈值函数之间很好的折中。

Donoho^[48]和 Johnstone^[49]提出了全局阈值选取模型，其计算公式为：

$$T = \sigma\sqrt{2\ln N} \quad (3-4)$$

式中：T 表示最终阈值， σ 表示噪声图像的标准方差，N 表示图像的长度。

在正态高斯模型下，对于多维独立正态变量联合分布，当维数趋于无穷时，大于该阈值的系数含有噪声信号的概率趋近于 0。本文使用半软阈值函数作为小波去噪的阈值函数，并令 $\lambda_1 = 2T, \lambda_2 = T$ 。作为对比的硬阈值函数与软阈值函数也使用 T 作为阈值。对原图像加入均值为 0，方差为 25 的高斯噪声，原图、噪声图及使用三种阈值函数去噪后的结果如图 5 所示。



图 5 三种去噪方式效果对比图

Fig5 Comparison of three denoising methods

主观上很难判断使用哪一种阈值函数的去噪效果更好，为了更好的衡量图像去噪的效果，使用 PSNR (Peak Signal to Noise Ratio) 判断去噪效果的好坏。PSNR 意为峰值信噪比，单位为 dB，越大代表去噪效果越好,其计算公式如下所示：

$$PSNR = 10 \times \log_{10} \frac{MAXI^2}{MSE} \quad (3-5)$$

式中：MAXI 表示图像颜色的最大数值，在灰度图像中为 255。MSE 表示噪声图像与原图像之间的均方误差。

MSE 的计算公式如下所示：

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i,j) - K(i,j)\|^2 \quad (3-6)$$

式中：m 表示图像的长度,n 表示图像的宽度，I 表示噪声图像，K 表示原图像。

噪声图像与去噪后的图像相对于原图的 PSNR 如表 3 所示。

表 3 不同去噪方式的 PSNR 值

Table3 PSNR values of different denoising methods

图像	噪声图像	软阈值函数去噪	硬阈值函数去噪	半软阈值函数去噪
PSNR	20.2886	21.1122	20.2910	20.4125

由表 3 可知，半软阈值函数去噪效果处于软阈值函数与硬阈值函数之间，得到了与理论相符的折中效果。

3.2.2 暗光图像处理

驾驶员在夜间行驶更易陷入疲劳，然而夜间光线不足，通过摄像头得到的画面存在轮廓不明显，人脸特征难以识别的问题。使用一种快速有效增强暗光条件下人脸对比度的方法：直方图均衡化（Histogram Equalization）。灰度直方图（histogram）是灰度级分布的函数，它表示图像中具有每种灰度级的像素的个数，反映图像中每种灰度出现的频率。直方图均衡化通过图像直方图均衡的操作将原图像密集的灰度映射到更大的灰度范围，均匀化各灰度范围的像素量，以此达到增强对比度的效果。如图 6 所示，蓝色部分为原始灰度图的直方图，橙色部分为算法增强后得到的均衡图像的直方图。

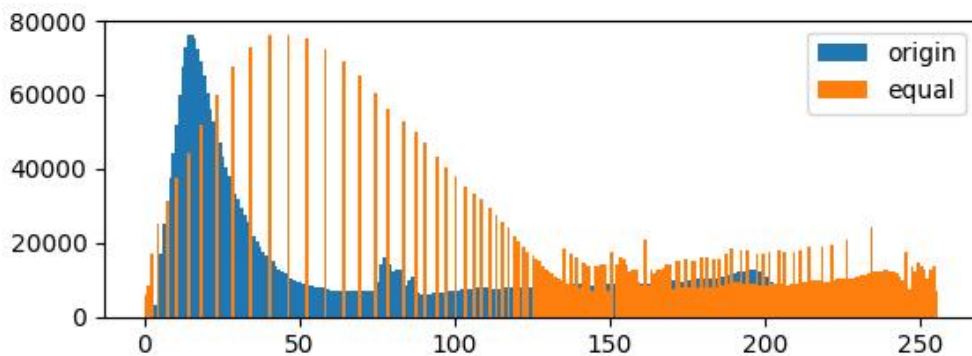


图 6 均衡化前后图像直方图对比

Fig6 Comparison of image histogram before and after equalization

直方图均衡化的原理是：使用累计分布函数对原图像的像素进行线性映射，映射后的像素保持原有的像素值大小顺序。映射公式如下所示：

$$S_k = \sum_{i=0}^k \frac{n_i}{N}, k = 0, 1, 2, \dots, L - 1 \quad (3-7)$$

式中：N 表示图像中像素的总数，L 表示图像的灰度级数， n_k 表示第 k 级灰度在图像内的像素点个数， S_k 表示第 k 级像素映射后的像素值。

由于待处理图像是彩色图像，所以分别在原图的 RGB 三色通道上进行直方图均衡化，再填充各像素点新的灰度值，最后联结这三个数组形成三维矩阵，恢复直方图处理后的图像。处理结果如图 7 所示，可以看到人脸的轮廓特征被加强，画面也更加明亮。



图 7 均衡化前后对比图

Fig7 Comparison before and after equalization

3.3 多任务卷积神经网络

3.3.1 网络任务与检测流程

本文所述的多任务卷积网络，借鉴 MTCNN 的多任务思想，同时完成人脸区域定位任务与头部姿态估计任务以缩短检测流程。

MTCNN 网络使用了多任务级联的网络架构，在完成面部检测功能的同时加入了同样是回归任务的人脸关键点检测，在同一网络完成了人脸定位与人脸五个关键点的输出，实验结果表明该举措同时提高了两个任务的准确率。考虑到当人脸偏转幅度过大导致脸部被遮挡时，人脸关键点必然产生缺失，而人脸检测与头部姿态估计对头部偏转的鲁棒性更好，因此将这两个任务同时进行。此外，首先进行头部姿态估计，可以在头部偏转角过大时，不再进行人脸关键点检测，避免

因遮挡或大姿态下人脸关键点算法所得关键点位置偏移过大造成疲劳检测失误。

多任务卷积神经网络运行流程如图 8 所示，通过三个层级的网络不断提升检测精度，最终由 O-Net 网络输出人脸位置与头部偏转角。

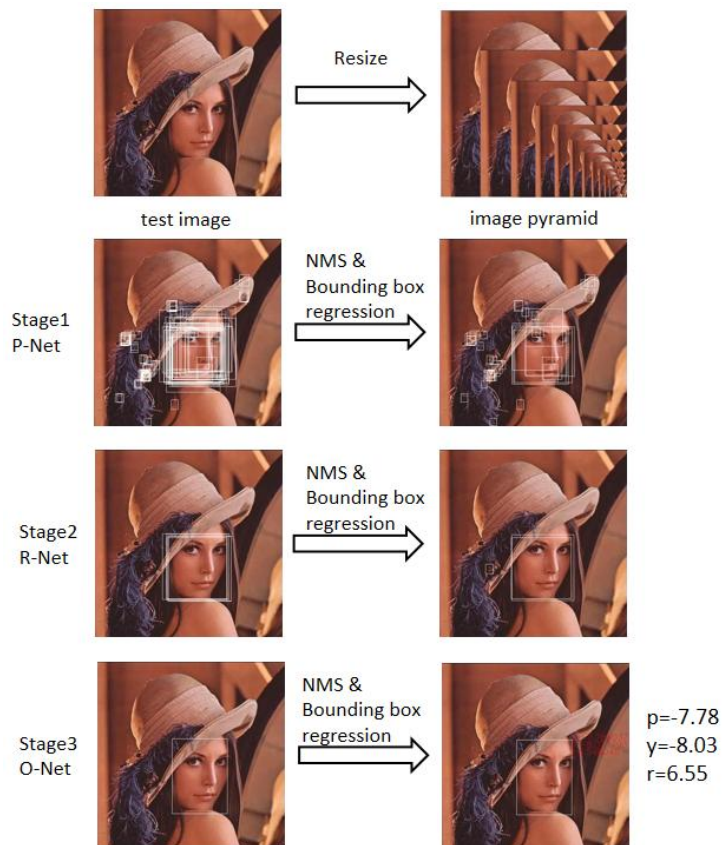


图 8 网络检测流程图

Fig8 Network detection flow chart

3.3.2 网络结构

针对待解决的任务，选择合适的网络模型，并在此基础上加以改进及优化是深度学习研究的重要方法。多任务卷积网络就是在借鉴 MTCNN 的网络结构基础上，针对实际问题进行的改进与优化。

1. 建议网络（Proposal Network, P-Net）

P-Net 是三层网络中的第一个网络，网络结构是 FCN 类型的全连接网络，其作用是对输入的图片进行简单预测，去除大量不是人脸的区域，输出认为有可能是人脸的区域坐标。原始图像进行不同尺度变换生成图像金字塔后输入 P-Net 完成人脸区域建议，同时输出各区域的人脸概率，最终将不同大小图片的检测结果映射回原图。如图 9 所示，P-Net 网络的输入是尺寸为 12×12 的三通道 RGB 图

像。第 1 个卷积层使用 10 个 3×3 的卷积核，并进行尺寸为 2×2 的最大池化，生成了 10 个 5×5 的特征图；第 2 个卷积层使用 16 个 3×3 的卷积核，生成了 16 个 3×3 的特征图；第 3 层采用 32 个大小为 3×3 的卷积核，生成了 32 个 1×1 的特征图；最后一层使用卷积核为 1×1 的卷积操作代替全连接，获得 $1 \times 1 \times 2$ 的人脸置信度， $1 \times 1 \times 4$ 的人脸边框和 $1 \times 1 \times 3$ 的头部姿态。

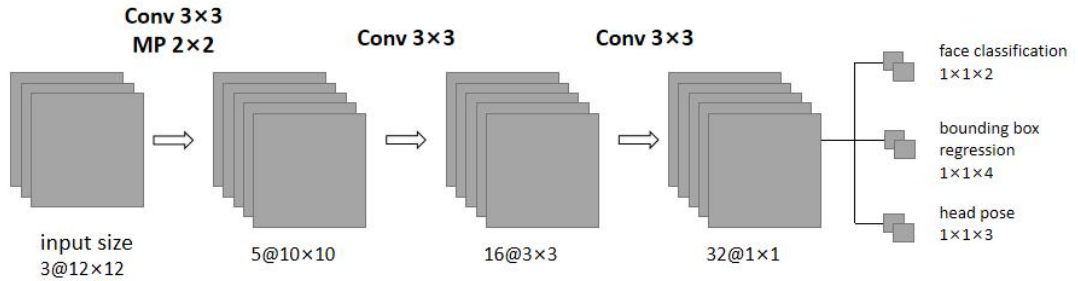


图 9 P-Net 网络结构图

Fig9 The architectures of P-Net

2.提纯网络（Refine Network, R-Net）

第二层 R-Net 对 P-Net 输出的人脸候选区域的进一步筛选,相对于 P-Net 拥有更复杂的网络结构。R-Net 的输入是 P-Net 的输出,即可能是人脸区域的原图切片,R-Net 将对这些切片进行进一步检测,删除置信度低的区域并对边界框进行调整,从而达到大量过滤低质量预测和人脸区域优化的效果。如图 10 所示,R-Net 的输入图像是一个 24×24 的三通道的 RGB 图像,该图像由 P-Net 生成的候选区域缩放得来。第一个卷积层使用 28 个 3×3 的卷积核,并进行尺寸为 3×3 的最大池化,生成了 28 个 11×11 的特征图;第二个卷积层使用 48 个 4×4 的卷积核,并进行尺寸为 3×3 的最大池化,生成了 48 个 4×4 的特征图;第三个卷积层使用 64 个 2×2 大小的卷积核,生成了 64 个 3×3 的特征图。将这 64 个特征图压缩至 1×576 后,使用全连接操作得到一个 128 维向量,从该向量中获得三个任务的输出。由此完成人脸分类、人脸候选框回归以及头部姿态估计这三大任务。

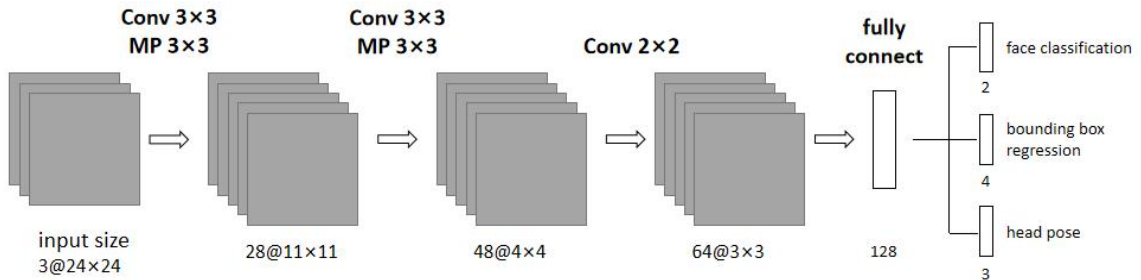


图 10 R-Net 网络结构图

Fig10 The architectures of R-Net

3. 输出网络（Out Network, O-Net）

第三层网络 O-Net 输出输出人脸置信度、对人脸候选框筛选后人脸的最终回归边界框、三个头部偏转角。如图 11 所示，O-Net 的输入图像尺寸更大，为 48×48 的三通道 RGB 图像，能够包含更多图像细节。第一个卷积层使用 32 个 3×3 的卷积核，并进行尺寸为 3×3 的最大池化，生成了 32 个 23×23 的特征图；第二个卷积层使用 64 个 3×3 的卷积核，并进行尺寸为 3×3 的最大池化，生成了 64 个 10×10 的特征图；在此之后，使用 CBAM（Convolutional Block Attention Module）^[50] 对得到的特征图施加通道注意力与空间注意力，使得网络能够抑制无相关噪声信息；第三个卷积层同样使用 64 个 3×3 的卷积核，并进行尺寸为 2×2 的最大池化，生成了 64 个 4×4 的特征图；第四个卷积层仅使用 128 个 2×2 的卷积核，生成了 128 个 3×3 的特征图；将这 128 个特征图压缩为 1×1152 并使用得到一个 256 维向量，并由此向量获得最终输出的人脸框与对应的头部偏转角。

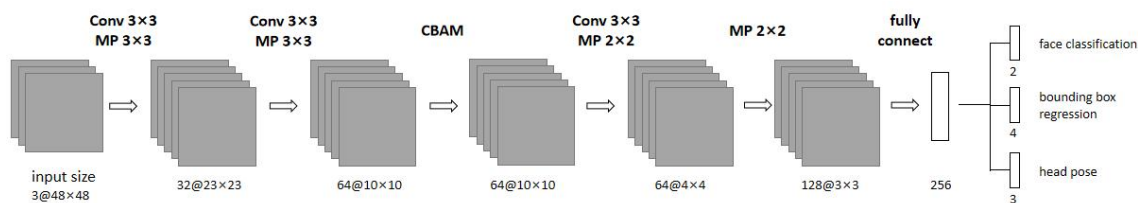


图 11 O-Net 网络结构图

Fig11 The architectures of O-Net

卷积神经网络注意力模块（CBAM）由通道注意力机制模块与空间注意力机制模块串联而成，其对输入特征层的处理过程即主要结构如图 12 所示。

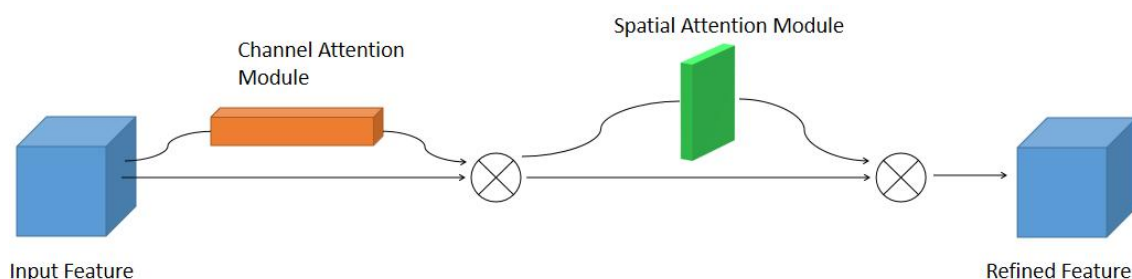


图 12 CBAM 主要结构示意图

Fig12 Main structure diagram of CBAM

CBAM 的前半部分为通道注意力机制，该部分首先对输入特征分别进行一次全局平均池化和全局最大池化得到俩个 $\text{Channel} \times 1 \times 1$ 的结果，对池化后的结果使用共享的全连接层进行缩小后再放大，将得到的俩个结果相加并使用 sigmoid 函数将参数映射到 0~1 之间，就得到了每一个通道的权重。其计算如公式（3-8）所示，Channel Attention 系数计算流程如图 13 所示。

$$\begin{cases} F_{out} = F_{input} \times \theta_{attention} \\ \theta_{attention} = \text{sigmoid}\left(MLP(\text{Maxpool}(F_{input})) + MLP(\text{Avgpool}(F_{input}))\right) \end{cases} \quad (3-8)$$

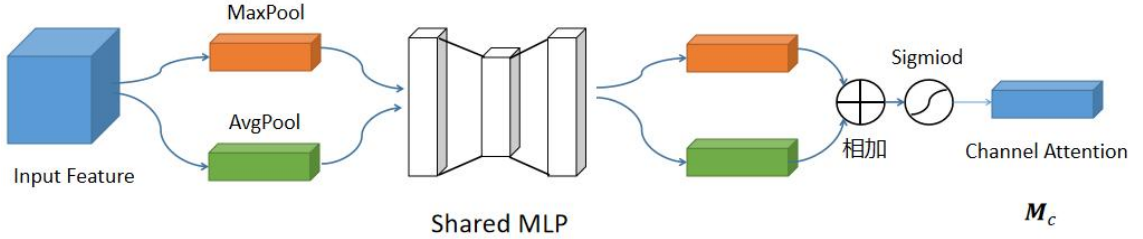


图 13 通道注意力系数计算流程图

Fig13 Flow chart of channel attention coefficient calculation

CBAM 的后半部分为空间注意力机制,特征层经过空间注意力机制计算过后,对每个通道施加空间注意力机制。首先对特征层的每一个通道分别取最大值与平均值得到俩个 $\text{Weight} \times \text{Height}$ 的矩阵,将这两个矩阵堆叠,使用 1×1 的卷积核进行卷积,对卷积得到的结果使用 sigmoid 函数映射到 0~1 之间,得到特征层每一个点的权重。Spatial Attention 系数计算流程如图 14 所示。

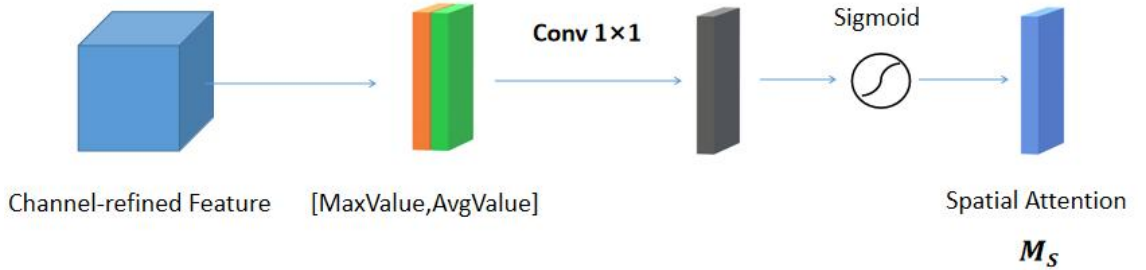


图 14 空间注意力系数计算流程图

Fig14 Flow chart of spatial attention coefficient calculation

4. 激活函数

在回归任务中,为了提高网络在零值附近的拟合能力,采用 P-ReLU 激活函数替换常用的 ReLU 激活函数。P-ReLU 激活函数在 ReLU 激活函数的基础上,在负值方向增加了梯度,避免了使用 ReLU 激活函数在正向传播时,若 $x < 0$ 会导致神经元未被激活的问题。P-ReLU 激活函数的解析式为: $\text{P-ReLU} = \max(x\alpha, x)$,其中 α 为初始随机的超参数,可以在反向传播中被学习,从而在当前神经元获得最好的负值梯度。P-ReLU 函数的图像如图 15 所示。

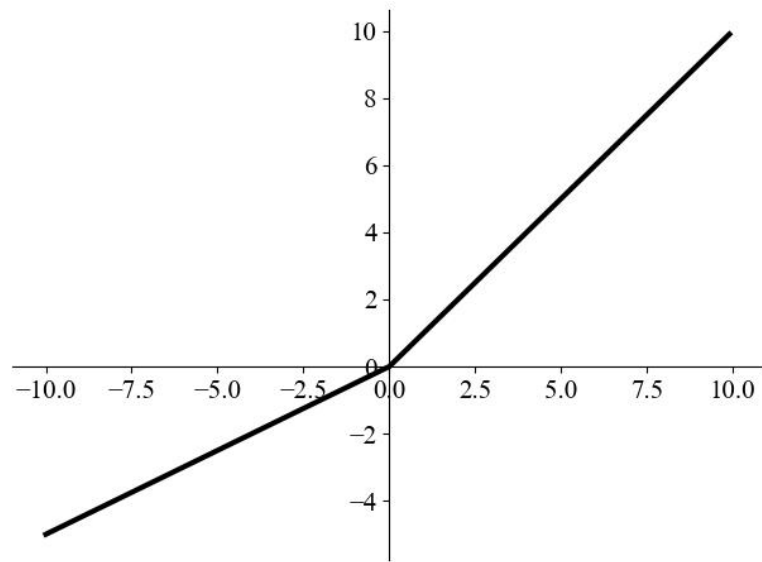


图 15 P-ReLU 函数示意图

Fig15 Schematic diagram of P-ReLU function

在分类任务，即分辨当前区域是否为人脸时，使用 Tanh 激活函数替代 sigmoid 激活函数。Tanh 激活函数由 sigmoid 激活函数变体而来，解决了 sigmoid 函数输出不以零为中心的问题，二者的关系为： $\text{Tanh}(x) = 2\text{sigmoid}(2x) - 1$ 。Tanh 激活函数的解析式为：

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3-9)$$

Tanh 激活函数的图像如图 16 所示。

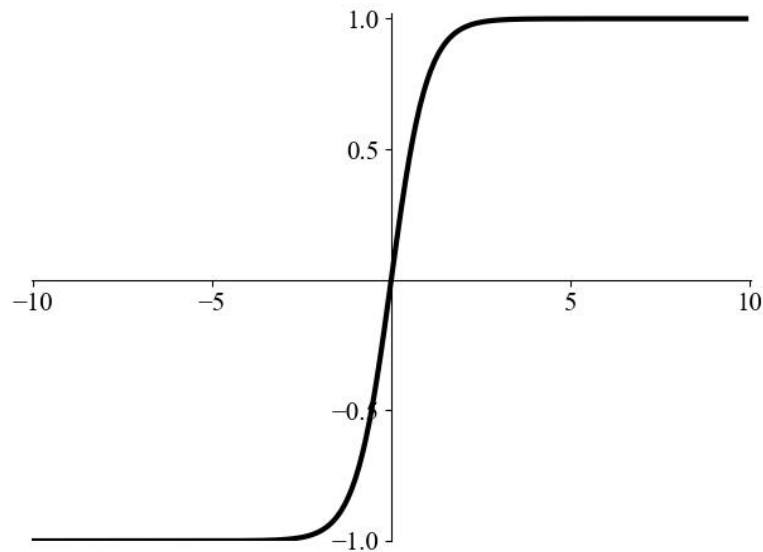


图 16 Tanh 函数示意图

Fig16 Schematic diagram of Tanh function

3.3.3 损失函数

多任务卷积神经网络共有三个任务：人脸与非人脸的二分类任务、人脸边界框的回归任务和头部姿态估计的回归任务。因此本网络需要对这三个任务与总任务分别设计损失函数，接下来分别介绍这些损失函数。

1.人脸分类任务的损失函数

人脸分类任务是一个二分类任务，人脸分类器输出的是输入图片包含人脸的概率，损失函数越小则代表模型越稳定、抗干扰能力越强。对于每一个训练样本 x ，定义其交叉熵损失函数为：

$$L_i^{det} = -((1 - y_i^{det})(1 - \log(p_i)) + y_i^{det} \log(p_i)) \quad (3-10)$$

式中： p_i 表示网络输出的样本 x 中包含人脸的概率， $y_i^{det} \in \{0,1\}$ 表示由标签提供的样本 x 中包含人脸的真实概率， L_i^{det} 表示分类的损失函数。

2.边界框（bounding box）回归任务的损失函数

人脸候选框的坐标位置被表示为区域左上角点的横纵坐标及区域的长宽。该任务的学习目标是一个回归问题，输出为连续量，对于每一个预测的候选边界框，计算它与最近的真实人脸区域的偏移量，使用均方误差损失函数来度量其误差，其公式如下：

$$\begin{cases} L_i^{box} = ||\hat{y}_i^{box} - y_i^{box}||_2^2 \\ \hat{y}_i^{box}, y_i^{box} \in R^4 \end{cases} \quad (3-11)$$

式中： L_i^{box} 表示边界框回归的损失函数， $\hat{y}_i^{box}, y_i^{box}$ 为四元组，包含区域左上角点的横坐标、纵坐标、区域右下角点的横坐标、纵坐标， \hat{y}_i^{box} 表示包含人脸区域的真实坐标， y_i^{box} 表示网络预测的人脸区域坐标。

3.头部姿态估计任务的损失函数

人头部的旋转可以由三个欧拉角(Euler Angle)来表示： $pitch$ (围绕 X 轴旋转)， yaw (围绕 Y 轴旋转)和 $roll$ (围绕 Z 轴旋转)，学名为俯仰角、偏航角和滚转角。其图像表示如图 17 所示。

头部姿态检测器的输出为三个连续的角度数值，因此该任务为回归任务。若同边界框回归任务一样，使用均方误差损失函数来度量其误差，在网络训练中发现头部姿态估计任务的损失难以减小，致使在收敛时整体任务的损失过大，严重影响训练效果。使用文献^[51]中所提出的多重损失函数作为头部姿态检测任务损失函数。该损失函数预先将人脸偏转角 $[-90^\circ, 90^\circ]$ 以间隔 3° 分为 60 个类别，对网络得到的偏转角，既计算其与真实值的误差，也计算其所在类别与真实值所在类

别的分类误差。头部姿态估计的损失函数公式（3-11）所示，计算其交叉熵损失函数与均方差损失函数的加权和。

$$\begin{cases} L_i^{angle} = H(y, \hat{y}) + \alpha \cdot \|\hat{y}_i^{angle} - y_i^{angle}\|_2^2 \\ \hat{y}_i^{angle}, y_i^{angle} \in R^3 \end{cases} \quad (3-12)$$

式中： L_i^{angle} 表示边界框回归的损失函数， $H(y, \hat{y})$ 表示输出偏转角所在类别与真实偏转角所在类别的多分类损失函数， $\hat{y}_i^{angle}, y_i^{angle}$ 为三元组，包含俯仰角、偏航角和滚转角， \hat{y}_i^{box} 表示头部的真实旋转度数， y_i^{box} 表示网络预测的头部偏转度数。本文中加权系数 α 取1。

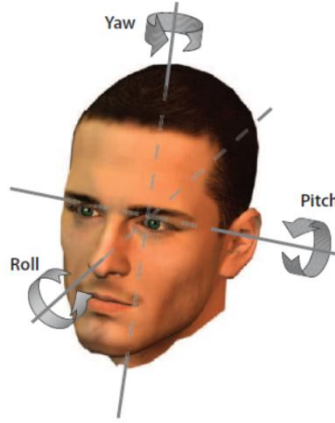


图 17 头部姿态示意图

Fig17 Schematic diagram of head posture

4.多任务训练损失函数

每一层网络的训练任务不同：P-Net 侧重于人脸识别，旨在找出包含人脸的区域；R-Net 侧重于精确的人脸识别与粗略的边界框回归；因为 R-Net 在人脸识别任务上已经足够精确，O-Net 对边界框回归与头部姿态估计更加关注。在每个 CNN 中，训练时都有不同类型的训练图像来满足不同的训练任务，因此并不是每个样本都需要用到上述三个损失函数。为了表示某个样本需要用到怎样的损失函数，使用一个指示值来建立所有网络整体的目标函数即多任务损失函数，其公式如下所示：

$$Loss = \min \sum_{i=1}^N \sum_{j \in \{det, box, engle\}} \alpha_j \beta_i^j L_i^j \quad (3-13)$$

式中： N 表示训练的样本总数， α_j 表示当前网络训练时三种损失函数的权重， β_i^j 表示当前样本的类型， L_i^j 表示上述三种类型的损失函数。

对于 P-Net 与 R-Net， α_j 被赋值为 $\alpha_{det} = 1$ 、 $\alpha_{box} = 0.5$ 、 $\alpha_{engle} = 0.5$ ，在 O-Net 中， α_j 被赋值为 $\alpha_{det} = 0.5$ 、 $\alpha_{box} = 1$ 、 $\alpha_{engle} = 1$ 。

3.3.4 网络训练

1.数据集

本章所提出的多任务卷积神经网络训练需要两个以不同任务为目标的数据集共同训练：对于人脸分类与边界框回归任务，采取 WIDER FACE^[52]数据集进行训练；对于头部姿态估计任务，采取 Biwi Kinect Head Pose Database (BIWI)^[53]数据集进行训练。

WIDER FACE 是一个人脸检测的 benchmark 数据集，其中的图像都是从公开的数据集中获取。WIDER FACE 共包含 32203 图像，这些图像中共有 393703 张在比例、姿势和遮挡方面具有高度变化范围的人脸。其部分样本如图 18 所示。

BIWI 包含 20 个人（14 位男性与 6 位女性）的超过 15000 张含标注的人脸姿态图像，头部姿态范围包括大约 $\pm 75^\circ$ 的偏航角、 $\pm 60^\circ$ 的俯仰角和 $\pm 50^\circ$ 的滚转角。其部分样本如图 19 所示。



图 18 WIDER FACE 部分样本

Fig18 Partial samples of WIDER FACE



图 19 BIWI 部分样本

Fig19 Partial samples of BIWI

2. 训练数据集生成

模型训练时需要生成不同类型的训练数据，为了更好的对训练数据做出分类，引入 Intersection over Union（IOU，交并比）作为数据分类的依据。

假设区域 A 左上角坐标为 (x_1, y_1) ，右下角坐标为 (x_2, y_2) ，区域 B 左上角坐标为 (p_1, q_1) ，右下角坐标为 (p_2, q_2) 。区域 A 与区域 B 按图 20 所示重合。

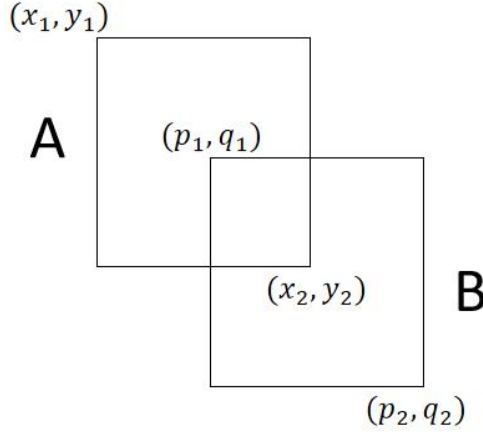


图 20 IOU 计算示意图

Fig20 Schematic diagram of IOU calculation

区域 A 与区域 B 的相交区域面积为：

$$Area_{A \cap B} = |p_1 - x_2| \times |q_1 - y_2| \quad (3-14)$$

区域 A 与区域 B 并集的面积表示为：

$$Area_{A \cup B} = |x_1 - x_2| \times |y_1 - y_2| + |p_1 - p_2| \times |q_1 - q_2| - Area_{A \cap B} \quad (3-15)$$

区域 A 与区域 B 的交并比表示为：

$$IOU = \frac{Area_{A \cap B}}{Area_{A \cup B}} \quad (3-16)$$

级联网络的每一级网络（P-Net、R-Net 和 O-Net）都需要通过上述两个数据集生成训练集，每一个网络的训练集各不相同但都包含四类训练数据：

1) **positive** 训练集：训练图片中包含的人脸部分与完整人脸的 IOU 大于 0.65，有人脸坐标标注。

2) **part** 训练集：训练图片中包含的人脸部分与完整人脸的 IOU 介于 0.4 到 0.65 之间，有人脸坐标标注。

3) **negative** 训练集：训练图片中包含的人脸部分与完整人脸的 IOU 低于 0.4 或图片中不含人脸，没有人脸坐标标注。

4) **pose** 训练集：训练图片中包含完整人脸并有头部姿态标注。

生成训练集数据时这四类数据的比例为 3: 1: 1: 2。对于三个级联网络，后一个网络的训练集生成需要依赖前一个网络的输出，即 R-Net 的训练集需要在 P-Net 训练完成后生成，O-Net 的训练集需要 R-Net 训练完成后生成。

第一个网络 P-Net 的训练集需要用到图像金字塔算法，通过对一张原始图像的缩放来检测不同大小的人脸。图像金字塔算法如下所示：

算法 3.1 图像金字塔

输入： 原始图片 source

输出： 图片集合 results

过程：

- 1) 确定图片缩放系数 `resize_factor`。确定图片缩放的最小尺寸 `minsize`。
- 2) `result = source × resize_factor`
- 3) `results.append(result)`
- 4) `while (result.size < minsize) :`
 `result = result × resize_factor`
 `results.append(result)`

`resize_factor` 系数其大小一般在 0.7 与 0.8 之间，设定较大缩放生成的图片较多，使检测时间延长，系数设定较小则会导致漏检中小型人脸。结合检测疲劳驾驶的具体要求，摄像头获得的图像中司机人脸区域偏大，因此设定 `resize_factor=0.708`。`minsize` 即网络所需识别的最小区域的大小，其作用为过滤人脸得分过低的识别框，在本文所述网络中 `minsize=40`。

对 WIDER FACE 数据集中的图片，生成其图像金字塔，随机从中截取一些 12×12 大小的图像，将不包含人脸的图像标记为 `negative` 数据。在人脸坐标周围随机截取 12×12 大小的图像，将与完整人脸的 IOU 小于 0.3 的作为 `negative` 数据，介于 0.4 到 0.65 之间的作为 `part` 数据，高于 0.65 的作为 `positive` 数据。对 BIWI 数据集中的图片，截取其中的人脸并缩放至 12×12 作为 `pose` 数据。以上数据按照 3: 1: 1: 2 的比例构成 P-Net 的训练集。

在 P-Net 训练完成后，使用 P-Net 对 WIDER FACE 数据集中的图片进行人脸识别，对识别结果使用非极大值抑制（Non-maximum suppression, NMS）算法去除重复度高的区域。NMS 算法如下所示：

算法 3.2 非极大值抑制

输入: 面部候选区域集合 $R = \{r_1, r_2, \dots, r_n\}$ 与对应的分数集合 $S = \{s_1, s_2, \dots, s_n\}$

输出: 筛选后的候选区域集合 $D = \{d_1, d_2, \dots, d_k\}$ 与对应的分数集合 $S_k = \{s_1, s_2, \dots, s_k\}$

过程:

- 1) while $R \neq \emptyset$ do:
 - 2) $r_{\max} = \text{findmaxS}(R)$
 - 3) 将 r_{\max} 及其分数加入 D 与 S_k , 从 R 与 S 中去除 r_{\max} 及其分数
 - 4) for r_i in R do:
 - 5) if $\text{IOU}(r_{\max}, r_i) > 0.5$:
 - 6) 将 r_i 及其分数从 R 与 S 中去除
-

对 P-Net 的预测结果进行 NMS 后将人脸框缩放至 24×24 , 并将其按照与完整人脸的 IOU 分类为 positive、part 与 negative 数据。对 BIWI 数据集中的图片, 截取其中的人脸并缩放至 24×24 作为 pose 数据。以上数据按照 3: 1: 1: 2 的比例构成 R-Net 的训练集。

当 R-Net 训练完成后, 使用 P-Net, R-Net 对 WIDER FACE 数据集中的图片进行级联预测并输出人脸位置, 使用 NMS 算法精简后, 将预测的人脸位置缩放至 48×48 , 并将其按照与完整人脸的 IOU 分类为 positive、part 与 negative 数据。对 BIWI 数据集中的图片, 截取其中的人脸并缩放至 48×48 作为 pose 数据。以上数据按照 3: 1: 1: 2 的比例构成 O-Net 的训练集。

3. 模型训练结果

图 21 所示为第一级 P-Net 网络训练准确率曲线 (a) 与 loss 曲线 (b)。

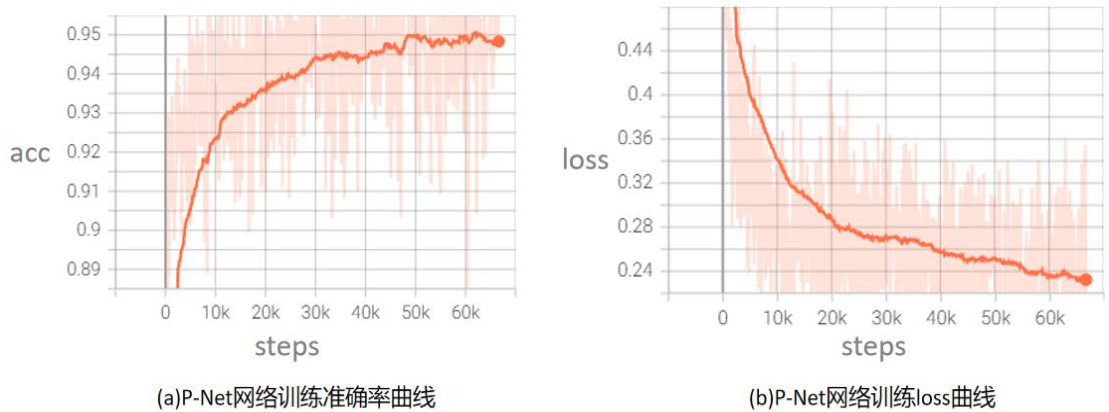


图 21 P-Net 网络训练准确率与 loss 曲线

Fig21 P-Net training accuracy and loss curve

图 22 所示为第二级 R-Net 网络训练准确率曲线 (a) 与 loss 曲线 (b)。

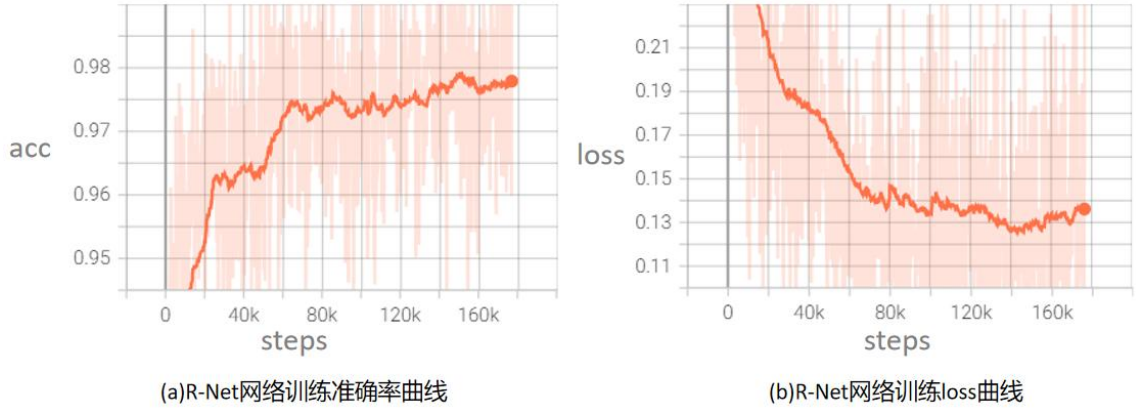


图 22 R-Net 网络训练准确率与 loss 曲线

Fig22 R-Net training accuracy and loss curve

图 23 所示为第三级 O-Net 网络训练准确率曲线 (a) 与 loss 曲线 (b)。

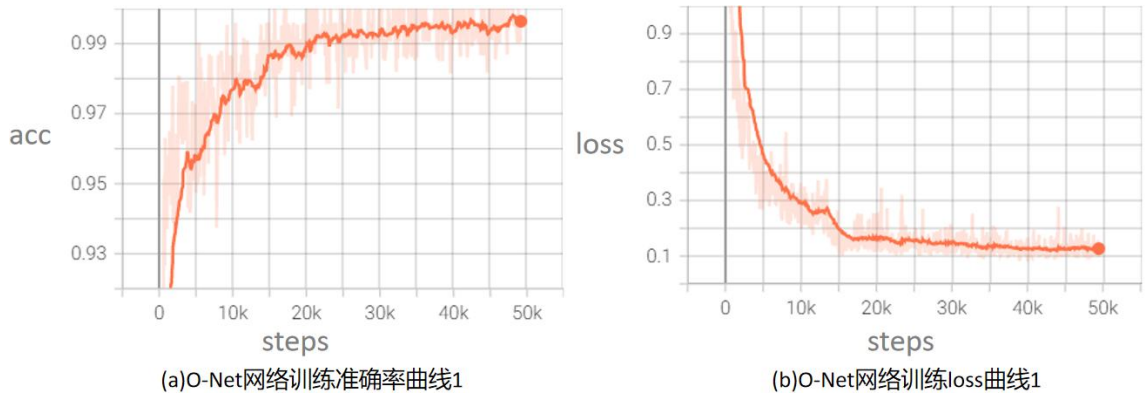


图 23 O-Net 网络训练准确率与 loss 曲线

Fig23 O-Net training accuracy and loss curve

图 24 所示为去除 CBAM 模块后第三级网络 O-Net 网络训练准确率曲线 (a) 与 loss 曲线 (b)。

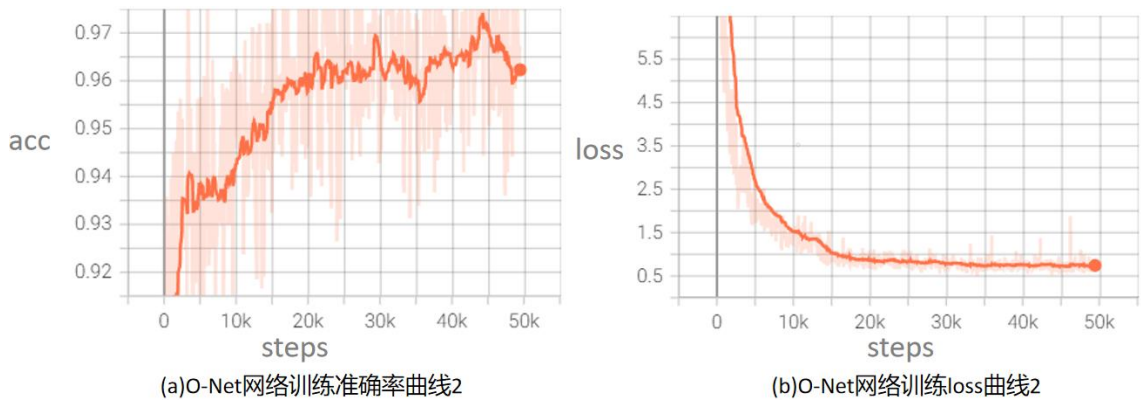


图 24 去除 CBAM 模块 O-Net 网络训练准确率与 loss 曲线

Fig24 O-Net training accuracy and loss curve without CBAM

3.3.5 训练结果与网络性能分析

1. CBAM 模块的效果

由图 23、24 可知，O-Net 在加入 CBAM 模块后多任务损失降低。O-Net 的多任务损失函数中各任务损失函数占比为 0.5: 1: 1，需要网络更多的关注边界框回归任务与头部姿态回归任务。CBAM 模块使得网络能从人脸区域中自适应的寻找关键特征，在参数量不变的前提下提升网络性能。

2. 人脸分类任务

由图 21、22 与 23 可知，随着网络级数的增加，人脸分类任务的准确率由 P-Net 的 95% 左右增加到 R-Net 的 98% 附近，在 O-Net 上准确率增加至 99.2% 左右。三个网络中 O-Net 收敛最快，R-Net 的训练数据最多。

3. 边界框回归任务

使用 WIDER FACE 数据集提供的模型性能评估方法对多任务卷积神经网络（MyNet）的性能进行评估。对比网络的算法数据由 WIDER FACE 官网得到，对比结果如图 25 所示。

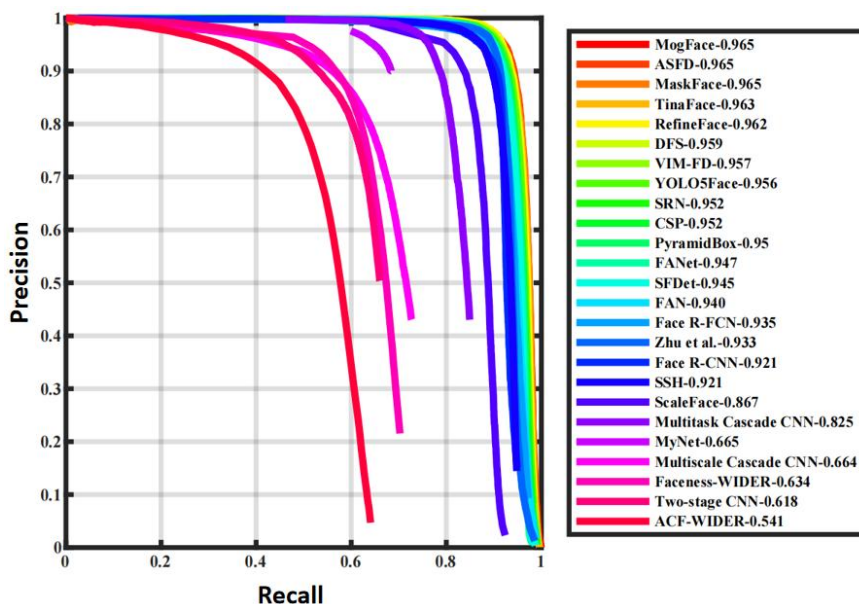


图 25 WIDER FACE 官网对比结果

Fig25 Comparison results of WIDER FACE official website

由在 WIDER FACE 上算法测评的结果来看，本章提出的多任务卷积神经网络模型在边界框回归任务上优于部分人脸检测算法，同时与一些优秀的人脸检测算法如 PyraminBox^[54]、Face R-CNN^[55]等仍然存在较大差距，略弱与参考模型 MTCNN。这是由于本章所提出的为轻量化网络模型，相较于深度网络精度存在

劣势，头部姿态估计任务与边界框回归任务的关联度没有人脸关键点检测任务与边界框回归任务的关联度高，因此在最终性能提升上相比 MTCNN 较少。该模型的主要任务是给出可用的人脸区域供人脸特征点检测算法使用并给出较精确的头部姿态估计，对检测出的人脸区域要求并不严苛。

4. 头部姿态估计任务

头部姿态估计任务性能的评估使用 AFLW2000 数据集进行，使用平均绝对误差（Mean Absolute Error, MAE）来评估网络在头部姿态估计任务上的性能，其公式如下：

$$MAE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i) \quad (3-17)$$

面部姿态估计性能测试中，选择俩个人脸关键点检测算法 FAN^[56]和 ERT^[57]，俩个头部姿态估计算法 Multi-Loss ResNet50^[51]和 3DDFA^[58]作为对比。对比结果如表 4 所示。

表 4 头部姿态估计任务 MAE 评估结果

Table4 MAE evaluation results of head pose estimation task

方法名称	Yaw	Pitch	Roll	MAE
Multi-Loss ResNet50(α=2)	6.470	6.559	5.436	6.155
3DDFA	5.400	8.530	8.250	7.393
MyNet	6.768	8.653	8.642	8.021
FAN(12 points)	6.358	12.277	8.714	9.119
Dlib(68 points)	23.153	13.633	10.545	15.777

从表 4 可知，本文所述模型在头部姿态估计任务上优于使用人脸特征点对头部姿态进行估计，劣于深度网络在头部姿态估计任务上的性能。同时模型在 Pitch 角和 Roll 角的检测上表现较好，接近 3DDFA 的性能，在 Yaw 角检测上表现较差，略逊于 Multi-Loss ResNet50。

5. 网络运行效率

本文所述多任务卷积神经网络同时完成了人脸区域定位与头部姿态估计任务，相较于先使用 MTCNN 进行人脸区域定位任务再使用 FAS-Net 进行头部姿态估计任务的方法降低了检测时间，每秒检测帧数从 12FPS 提升至 25FPS。相较于先进行人脸区域定位任务再使用 ERT 算法进行人脸特征点检测任务，之后使用人脸特征点估计头部姿态的方法，平均绝对误差从 15.777 降低至 8.021。

3.4 人脸特征点检测

在 3.3 节定位人脸区域的基础上,需要定位驾驶员眼部区域与嘴部区域来获得眼部与嘴部状态。本节基于级联回归树 (Ensemble of Regression Trees, ERT) 算法建立人脸特征点检测模型, 进而对眼部与嘴部区域进行定位并获得状态信息。ERT^[57]是一种基于回归树的人脸对齐算法, 算法通过建立的一个级联残差回归树 (GBDT) 来预测当前输入图像中人脸特征点的位置。GBDT 由很多棵树构成, 这些树的每一个叶子节点上都存储着一个残差回归量, 输入的图片经过树的分叉最终会落到一个叶子节点上, 将该叶子节点存储的残差加到输入上, 使输入向真实值回归, 当所有树遍历完时, 所有残差叠加在一起就获得了最终的人脸特征点输出。该算法的核心是俩层级联回归树建立的模型, 主要包括建立训练模型与模型拟合两个部分。ERT 算法在保证高准确率的前提下, 能够快速获得 68 个人脸特征点, 包含人眼、嘴唇、鼻子和脸部轮廓信息。

3.4.1 建立模型

在建立 ERT 模型之前, 先要确定人脸关键点的模型。设 $X_i \in R^2$ 表示输入图像 I 的第 i 个特征点坐标, 人脸形状向量 S 被表述为 $S = \{X_1^T, X_2^T, \dots, X_n^T\}$, 其中 n 表示人脸特征点的数量。在本文所述 ERT 算法中, 检测的人脸特征点数量为 68 个。标准人脸 68 个特征点标注如图 26 所示。

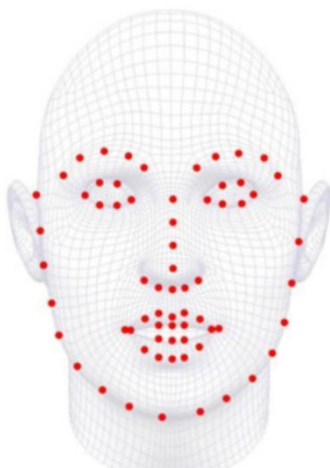


图 26 标准人脸 68 个特征点标注

Fig26 Standard 68 feature point annotations

1. 回归模型的建立

ERT 算法采用级联回归的方式建立多个级联的回归器。

记训练数据集为 $(I_1, S_1), (I_2, S_2), \dots, (I_n, S_n)$ ，其中 (I_n, S_n) 表示第 n 张训练图及其人脸形状。初始迭代公式为：

$$\hat{S}^{t+1} = \hat{S}^t + \gamma_t(I, \hat{S}^t) \quad (3-18)$$

式中： \hat{S}^t 表示当前对人脸形状 S 的估计，由人脸特征点的坐标组成； γ_t 表示回归器，其输入为当前人脸图片 I 与当前估计的形状 \hat{S}^t ，输出为当前人脸形状的更新。

每经过一个迭代器，人脸形状估计就会被更新，使之与真实人脸形状愈发接近。为了使迭代能顺利进行，首先要得到每张训练图像的初始人脸形状。记人脸图像三元数组为 $(I_{\pi_i}, \hat{S}_i^{(0)}, \Delta S_i^{(0)})$ ，依次代表人脸图像，初始人脸形状和目标更新步长。每张训练图像三元数组的计算如下：

$$\begin{cases} \pi_i \in \{1, 2, \dots, n\} \\ \hat{S}_i^{(0)} \in \{S_1, S_2, \dots, S_n\} \setminus S_{\pi_i} \\ \Delta S_i^{(0)} = S_{\pi_i} - \hat{S}_i^{(0)} \end{cases} \quad (3-19)$$

获得三元数组后，对于每一个训练图像，使用其人脸图像三元数组迭代训练回归器 $(\gamma_0, \gamma_1, \dots, \gamma_{k-1})$ ，其中 k 为设置的回归器数量。回归器的最终迭代训练公式为：

$$\begin{cases} \hat{S}_i^{(t+1)} = \hat{S}_i^{(t)} + \gamma_t(I_{\pi_i}, \hat{S}_i^{(t)}) \\ \Delta S_i^{(t+1)} = S_{\pi_i} - \hat{S}_i^{(t+1)} \end{cases} \quad (3-20)$$

当 $t=0$ 时，使用初始人脸图像三元组进行训练，之后在训练过程中不断动态调整人脸图像三元组中的人脸形状与目标更新步长。不断重复上述迭代公式，经过 k 次迭代可得到 k 个回归器。

2. 回归器的训练

ERT 算法使用基于平方误差总和的梯度提升树（gradient tree boosting）来训练回归器 γ_t 。其训练算法如下：

算法 3.3 回归器训练算法

输入：数据集 $\{I_{\pi_i}, \hat{S}_i^{(t)}, \Delta S_i^{(t)}\}_{i=1}^N$

输出：回归函数 $\gamma_t(I, \hat{S}^{(t)})$

- 1) 设置学习速率 $0 < v < 1$
- 2) 回归器的初始化：

$$f_0(I, \hat{S}^{(t)}) = \underset{\gamma \in R^{2p}}{\operatorname{argmin}} \sum_{i=1}^N \|\Delta S_i^{(t)} - \gamma\|$$

3) 对于 $k = 1, 2, \dots, K; i = 1, 2, \dots, N$, 每次迭代更新 γ_{ik} 的值, 更新公式如下:

$$\gamma_{ik} = \Delta S_i^{(t)} - f_{k-1}(I_{\pi_i}, \hat{S}_i^t)$$

4) 更新 $f_k(I, \hat{S}^t)$:

$$f_k(I, \hat{S}^{(t)}) = f_{k-1}(I, \hat{S}^{(t)}) + v g_k(I, \hat{S}^{(t)})$$

其中 $g_k(I, \hat{S}^{(t)})$ 是由 γ_{ik} 得到的弱回归方程。

5) 不断重复步骤 3)、4), 最终达到迭代次数或迭代收敛, 收敛条件为:

$$\gamma_t(I, \hat{S}^{(t)}) = f_k(I, \hat{S}^{(t)}), \text{ 最终得到回归函数 } \gamma_t(I, \hat{S}^{(t)}).$$

ERT 算法采用的损失函数为平方误差:

$$\varepsilon = \frac{1}{2} [y_i - f(x_i)]^2 \quad (3-21)$$

在每一步迭代中, 使用平方误差的导数作为拟合对象构建模型。每次迭代时, 随机选取检测到的人脸形状中的 500 对点对进行实验, 以最大化降低方差。将选取的点对的像素差值表征为图像的人脸特征点, 选取差值可以有效降低光照不均对特征点检测的影响。使用这些像素差值特征构成回归树, 其中非叶子节点保存分类阈值, 叶子节点保存特征点的残差。

3. 模型拟合

在获得回归模型后, 本段介绍得到最优模型后的搜索拟合过程, 过程如下:

1) 初始化人脸形状。从输入图像中检测出人脸特征点, 并得到初始人脸形状向量。

2) 基于人脸检测的结果, 计算人脸特征点对的像素差值, 使用像素差值进行树的分裂。

3) 使用所有的决策树求残差。将像素差值与非叶子节点中保存的阈值进行比较, 计算出差值, 得到的残差保存在叶子节点上。将所有的残差求和得到全局残差。

4) 将本次迭代获得的全局残差与上一次迭代的结果求和。

重复 2)、3)、4) 的操作直至达到迭代次数。

3.4.2 模型训练

人脸特征点是很多其他任务的前置任务, 因此有大量公开的人脸特征点数据

集供算法训练使用。本文选用 300-W 数据集训练 ERT 算法，300-W 数据集包含 afw (337)、helen (train 2000+test 330)、ibug (135) 和 lfpw (train 811+test 224)，共计 3148 张训练图片与 689 张测试图片。该数据集中包含的图像可能不止一个人脸但是只有一个人脸会被标注，标注方式为标准 68 个人脸特征点。模型训练采用第三方库 Dlib 提供的训练方法，训练参数如表 5 所示。

ERT 算法在不同级联级数下的效果如图 27 所示。其中，(f) 为真实人脸坐标标注，T 代表级联级数，当级联数为 0 时，人脸关键点向量是初始值，通过不断地级联回归，可以看出随着 T 的增大，人脸关键点的坐标位置与实际位置越来越接近，当级联级数为 T 为 10 时，基本接近真实坐标。

表 5 ERT 模型训练参数

Table5 Parameter of ERT training

参数名	设置
级联级数	10
每层级联树的数量	500
树的深度	5
正则项	0.1
特征池的大小	400
回归树节点分裂个数	20
样本扩大倍数	20

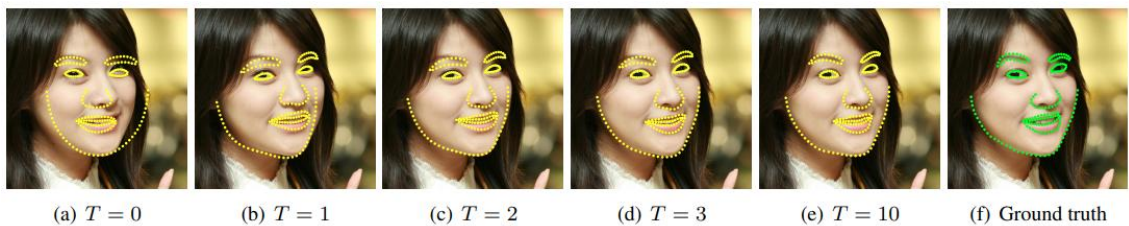


图 27 ERT 算法流程图

Fig27 Flow chart of ERT

3.5 本章小结

本章介绍了获取包含疲劳特征的头部姿态及面部区域（眼部区域与嘴部区域）的方法：多任务卷积神经网络和基于计算回归树的人脸特征点检测。本章详细介绍了多任务卷积神经网络的网络构成、训练策略以及性能评估，多任务卷积神经网络在人脸分类任务上取得了 98% 的准确率，在边界框回归任务上性能优于部分

网络，在头部姿态估计任务上优于使用人脸特征点计算头部姿态的方法。多任务卷积神经网络为 ERT 的使用奠定了基础，为第四章疲劳参数的提取建立了前提与保障。

4 疲劳参数提取

4.1 眼部疲劳参数

4.1.1 眨眼判断

疲劳状态下，驾驶员会感到眼睛酸涩、眼皮沉重，眨眼频率与眨眼动作特征都与清醒状态下不同。眨眼频率使用单位时长的眨眼次数来统计，单次眨眼动作疲劳判断使用 PERCLOS 准则。

眼部疲劳特征的提取基于对眨眼动作的正确判断。当前图像驾驶员是否闭眼可以使用 EAR（Eye Aspect Ration，眼睛纵横比）的大小来判断。EAR 的计算公式如下：

$$EAR = \frac{width}{length} \quad (4-1)$$

EAR 能够直观的反应人眼的睁闭状态，值越大表明眼睛睁开幅度越大。使用比例而非眼睛睁开的宽度可以避免人眼大小对判断结果的影响，结合人脸特征点，可以快速、精准的反应人眼状态。ERT 算法在人眼处的标注如图 28 所示，结合后的 EAR 算法如下：

$$EAR_L = \frac{|p_{44} - p_{48}| + |p_{45} - p_{47}|}{2|p_{43} - p_{46}|} \quad (4-2)$$

$$EAR_R = \frac{|p_{38} - p_{42}| + |p_{39} - p_{41}|}{2|p_{37} - p_{40}|} \quad (4-3)$$

$$EAR = \frac{EAR_L + EAR_R}{2} \quad (4-4)$$



图 28 ERT 算法在人眼处的标注

Fig28 Annotation of ERT algorithm at human eyes

选取一名测试人员记录其眨眼过程的 EAR 值，其变化如图 29 所示。

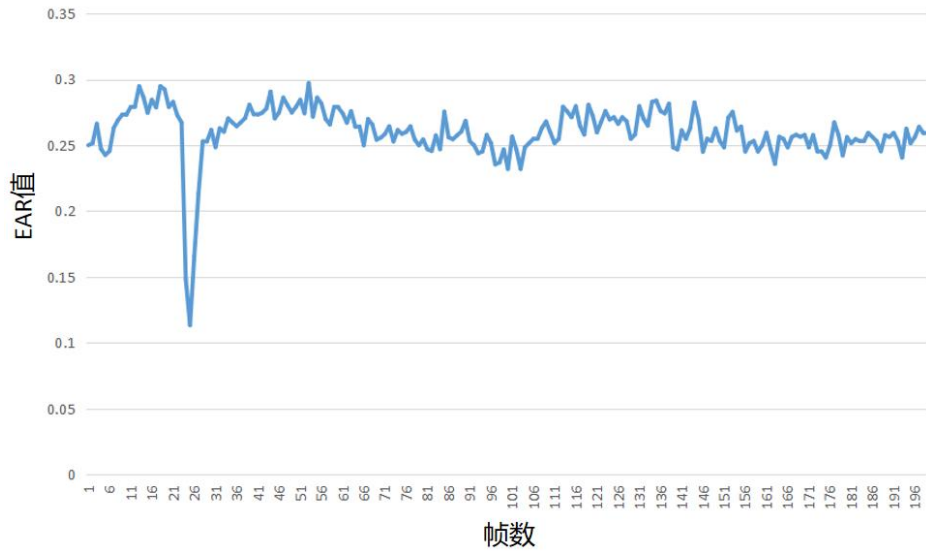


图 29 眨眼时 EAR 值的变化曲线

Fig29 Curve of EAR value during blinking

由图 29 可知：在非眨眼的情况下，EAR 值基本维持在一个范围内，当眨眼行为发生时 EAR 值急速减少之后快速恢复。为了准确判断眨眼动作的发生，本文使用 UTA-RLDD^[59]（UTA Real-Life Drowsiness Dataset）中的数据进行 EAR 阈值测定实验。UTA-RLDD 共包含 11 名受测人员拍摄的包含正脸的视频，选取每名受测人员各一段视频，总计 270s，眨眼 63 次。为更好的描述检测效果，引入精确率(precision)、召回率(recall)和 F_1 来对检测结果进行评估，其计算公式如下。

$$P = \frac{TP}{TP + FP} \quad (4-5)$$

$$R = \frac{TP}{TP + FN} \quad (4-6)$$

$$F_1 = \frac{2PR}{P + R} \quad (4-7)$$

式中： P 表示精确率， R 表示召回率， F_1 是对精确率和召回率的加权平均，能更好的反应整体效果， F_1 越高，说明整体效果越好。 TP 表示眨眼行为被检测为眨眼的次数； FP 表示非眨眼行为被检测为眨眼的次数，即误检次数； FN 表示眨眼行为被检测为非眨眼行为的次数，即漏检次数。

实验结果如表 6 所示。由表 6 可知，取 EAR 阈值为 0.205 的眨眼检测效果较好。

表 6 眨眼判断 EAR 阈值实验

Table6 EAR threshold experiment of blink judgment

EAR 阈值设定	实际眨眼次数	TP	FP	FN	P	R	F_1
0.25	63	58	120	5	0.326	0.921	0.4972
0.23	63	56	94	7	0.373	0.889	0.5255
0.21	63	53	22	10	0.707	0.841	0.7682
0.205	63	52	11	11	0.825	0.825	0.8250
0.20	63	45	8	18	0.849	0.714	0.7757
0.19	63	34	6	29	0.850	0.540	0.6604

4.1.2 PERCLOS 参数获取

当驾驶员处于疲劳状态时，单次眨眼动作时间延长，眨眼动作各部分所占比重也不同。卡内基梅隆研究所经过大量实验论证，提出 PERCLOS（percentage of eyelid closure over the pupil over time）准则用于判断驾驶员疲劳程度。PERCLOS 指在单位时间内眼睛闭合时间占总时间的比值，是在疲劳检测中被广泛使用的判断指标。该准则通常有 P_{70} ， P_{80} 和 EM 三种测量方式：

P_{70} ：眼皮盖过眼球的面积超过 70% 所占的时间比例。

P_{80} ：眼皮盖过眼球的面积超过 80% 所占的时间比例。

EM ：眼皮盖过眼球的面积超过 50% 所占的时间比例。

1999 年美国联邦高速公路管理局召集专家学者，研究讨论九种不同疲劳指标的有效性。通过对实验数据的对比，证明了 PERCLOS 相对于驾驶员其他特征，更能直接反映驾驶员的疲劳程度，其中使用 P_{80} 测量方式判断结果最佳^[60]。

在 P_{80} 测量方式下，眨眼过程可以被划分为图 30 所示。其中 t_1 为开始时刻， t_2 时眼睛睁度下降至 20%，即眼皮盖过眼球 80%， t_3 时眼睛睁度上升至 20%， t_4 为结束时刻。

在如图 30 所示标注下 PERCLOS 的计算公式为：

$$PERCLOS = \frac{t_3 - t_2}{t_4 - t_1} \quad (4-8)$$

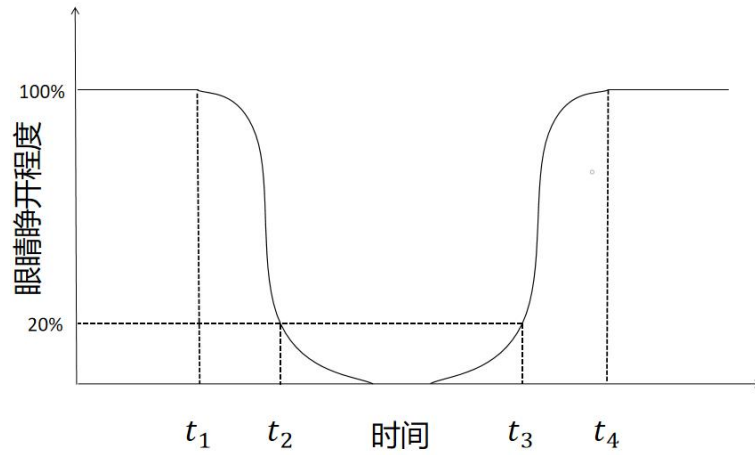


图 30 眨眼过程划分

Fig30 Division of blink process

为了计算 PERCLOS 值，需要得到眼球被覆盖 80% 时 EAR 的值。由于眨眼时眼睛的宽不变，因此眼球覆盖 80% 时即 EAR 值域的前 20% 处，由此可得：

$$EAR_p = (EAR_{max} - EAR_{min}) \times 0.2 + EAR_{min} \quad (4-9)$$

由 4.1.1 节可知， $EAR_{max} = 0.205$ ，对多个闭眼对象计算其 EAR 值，可以得到 $EAR_{min} = 0.15$ ，可计算 $EAR_p = 0.161$ 。将 PERCLOS 准则与 EAR 值结合，PERCLOS 计算公式变形为：

$$PERCLOS = \frac{F_p}{F_{close}} \quad (4-10)$$

式中： F_{close} 表示一次眨眼动作的总帧数， F_p 表示一次眨眼动作中 EAR 值低于 EAR_p 的总帧数。

由国内外实验人员的分析可知，当 PERCLOS 值小于 0.1 时，推断驾驶员处于清醒状态；当 PERCLOS 值处于 0.1 到 0.3 时，推断驾驶员处于轻度疲劳状态；当 PERCLOS 值处于 0.3 到 0.5 时，推断驾驶员处于中度疲劳状态；当 PERCLOS 值大于 0.5 时，推断驾驶员处于重度疲劳状态^[61]。

4.1.3 眨眼次数获取

当驾驶员疲劳时，还有几种特征可能表现在眼部：长时间睁眼、长时间闭眼和频繁眨眼。一般情况下，一分钟时间内眨眼次数约为 15~20 次，大约每隔 3~6 秒眨一次眼睛，每次眨眼时间约为 0.2~0.4 秒^[62]。处于疲劳状态时，驾驶员会不自觉进入走神状态，此时驾驶员目光呆滞，眼睛长时间睁开，眨眼频率下降；驾驶员也可能感到眼睛酸涩，频繁眨眼，眨眼频率上升。当处于重度疲劳时，驾驶

员可能发生长时间闭眼的情况，此时 PERCLOS 准则失效，极易发生交通事故。为了判断出上述三种情况下驾驶员的疲劳状态，使用眨眼频率作为疲劳特征之一。当驾驶员的眨眼频率超出或低于正常范围时，推断其处于疲劳状态。

一次眨眼动作指眼睛从开始闭合到完全睁开，在本文中指 EAR 值从降低到阈值以下到上升至阈值。如图 31 所示，通过监测 EAR 值与阈值的关系即可判断眨眼次数。图 31 共检测到 2 次眨眼动作。

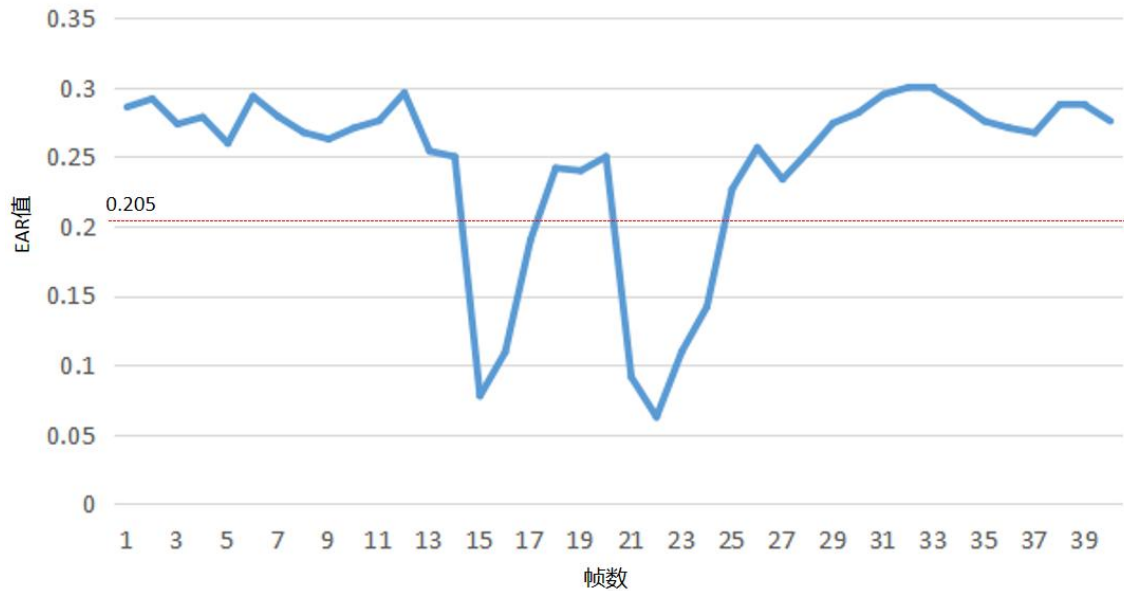


图 31 眨眼次数统计方法示意图

Fig31Schematic Diagram of Blinking Times Statistics

4.2 嘴部疲劳参数

除眼部疲劳特征外，嘴部在疲劳时也会表现出明显特征：打哈欠。当人处于疲劳、困倦状态时，呼吸节奏放缓，血液中二氧化碳含量增加，氧气含量变少。此时大脑呼吸中枢感知到血液变化，令人不受控制的进行深呼吸动作吸入氧气，同时令面部肌肉紧张，加速头部血液流通，其外在表现便是打哈欠。由此可以知道哈欠动作是很好的疲劳表现，单位时间内哈欠的次数与疲劳程度密切相关。

仿照眨眼判断的思想，通过对嘴部纵横比的检测来判断嘴张开的大小，即使用 MAR (Mouth Aspect Ration, 嘴部纵横比) 检测哈欠行为。在 ERT 算法中，嘴部的标注如图 32 所示：

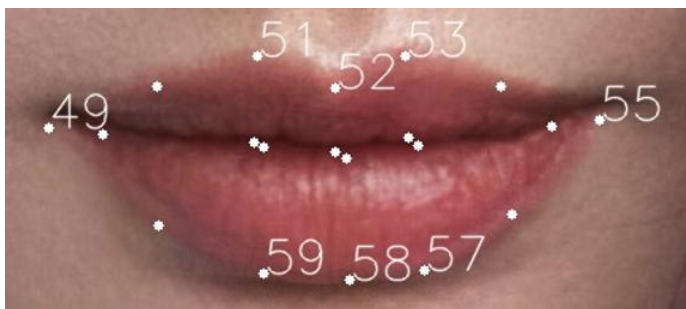


图 32 ERT 算法在嘴唇处的标注

Fig32 Annotation of ERT algorithm at human mouth

通过人脸特征点检测算法得到嘴部外侧的 8 个二维坐标位置，在 ERT 中被标记为 49、51、52、53、55、57、58、59，由此可得到 MAR 的计算公式：

$$MAR = \frac{|p_{59} - p_{51}| + |p_{58} - p_{52}| + |p_{57} - p_{53}|}{3|p_{55} - p_{49}|} \quad (4-11)$$

选取嘴部外侧的特征点可以避免当张嘴时内部特征点标注受牙齿、舌头的影响出现偏移，而外部特征点的干扰较少。

邀请一名测试人员录制一段视频，测试人员进行说话、唱歌和打哈欠行为，其 MAR 变化如图 33 所示。由图 33 可知，当测试人员说话或唱歌时 MAR 快速上下波动，打哈欠时 MAR 峰值明显高于前两者，且会维持一段时间。

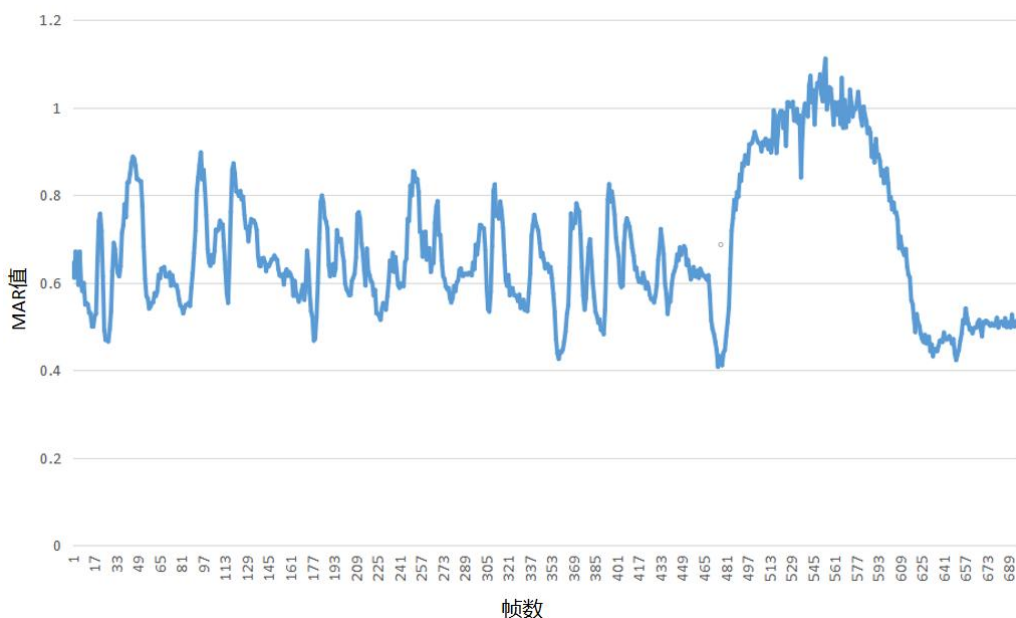


图 33MAR 值变化曲线

Fig33MAR value change curve

哈欠行为的 MAR 均值与说话、唱歌行为的 MAR 均值区别较大，可以确定一个阈值来区分哈欠行为与其他行为。此外，哈欠行为通常持续 2~3s，因此低于

1s 的高 MAR 值区间应被视为检测失误或说话、唱歌行为的特异点。本文使用 YawDD^[63]来测定哈欠行为的 MAR 阈值, YawDD 包含不同性别、不同人种以及佩戴眼镜的驾驶员在不同光照下模拟驾驶的视频。选取其中 8 段视频, 共计哈欠 26 次。实验结果如表 7 所示。

表 7 哈欠判断 MAR 阈值实验

Table7 MAR threshold experiment of yawn judgment

MAR 阈值	实际哈欠次数	<i>TP</i>	<i>FN</i>	<i>FP</i>	<i>P</i>	<i>R</i>	<i>F₁</i>
0.95	26	12	14	0	1	0.4615	0.6315
0.9	26	18	8	0	1	0.6923	0.8182
0.85	26	26	8	8	0.6923	0.6923	0.6923
0.8	26	18	4	6	0.7857	0.8461	0.8148
0.75	26	26	2	2	0.9231	0.9231	0.9231
0.7	26	26	0	5	0.8387	1	0.9123

实验数据分析: 阈值设定过高会导致部分哈欠无法被检测到, 使得漏检次数过多; 设置在 0.8 左右会在一次哈欠行为进行中出现 MAR 低于阈值的情况, 从而导致系统将一次哈欠判定为两次, 误检次数明显上升; 设置在 0.7 会将非哈欠行为判断为哈欠行为, 致使误检次数上升; 设置在 0.75 可以获得较高的 F_1 值, 即获得较好的效果。

4.3 头部姿态疲劳参数

驾驶员陷入疲劳时, 很难维持正常的头部姿态, 出现不断点头、歪头乃至长时间低头不动的情况, 所以驾驶员头部姿态可以有效反应驾驶员疲劳状态。在第三章获得头部姿态的前提下, 使用欧拉角描述驾驶员头部姿态, 使用单位时间内点头次数和单位时间内头部异常姿态时长来描述疲劳特征。

4.3.1 点头次数获取

驾驶员在疲劳时, 会陷入短暂失神导致无力控制头部肌肉, 点头后惊醒并快速调整头部姿态, 如此反复。摄像机设定在驾驶员前方方向盘处, 在描述头部姿态的欧拉角中, 俯仰角 *pitch* 与点头行为的关系最为密切, 当驾驶员点头时, *pitch* 值上下波动。点头前后的 *pitch* 值变化如图 34 所示。

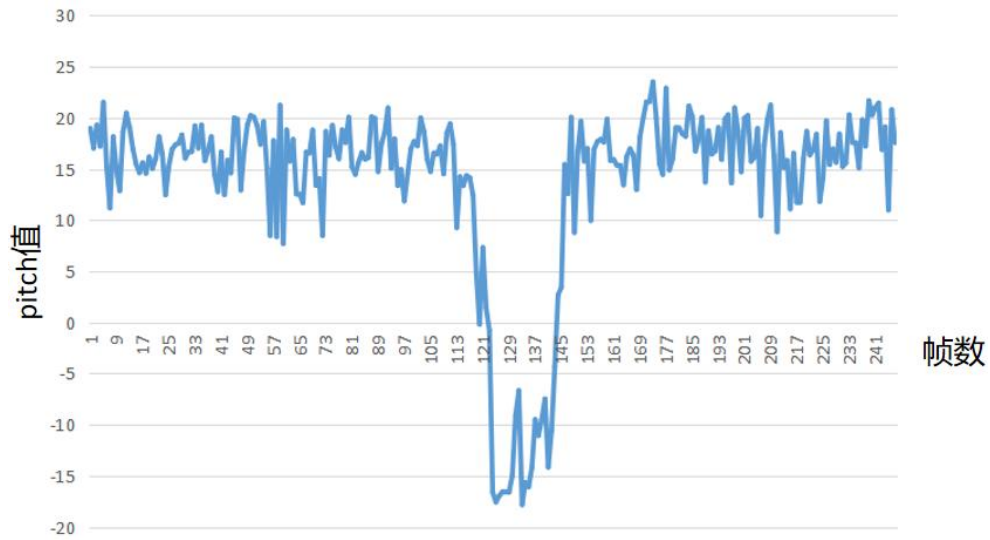


图 34 点头时 pitch 值的变化

Fig34 Change of pitch value when nodding

从图 34 中可以看出，在点头动作发生时 pitch 值有了大幅度的下降，而在非点头时 pitch 值也是不断波动的，这是由于驾驶员不可能保持头部静止不动，而且车辆在行驶过程中难免有震动。考虑到不同驾驶员的身高与开车习惯的不同，点头动作的检测不使用 pitch 阈值比较的方式进行判断，算法检测连续 10 帧的 pitch 值，若 10 帧内 pitch 值减少 25° ，认为进行了点头动作。

4.3.2 头部异常姿态时长获取

当驾驶员由于疲劳进入失神状态时，对头部的控制力减弱，致使头部姿态与清醒状态下不同，出现歪头、偏头与低头的姿态。由于 pitch 角已经被用于描述点头次数，本小节使用 yaw 角与 roll 角描述头部异常姿态。yaw 角与 roll 角的理论范围为 -90° 至 90° ，然而由于人的生理限制，成年人的头部姿态角事实范围无法达到理论值，其中 yaw 角范围为 -40.9° 至 36.3° ，roll 角范围为 -79.8° 至 75.37° [64]。仿照 P_{80} 原则，以 0° 为中心，将偏离中心超过值域 20% 的头部姿态角定义为异常姿态，偏离中心 20% 以内的头部姿态角定义为正常姿态，即当 yaw 角小于 -8.18° 或大于 7.26° ，roll 角小于 -15.96° 或大于 15.07° 时，认为驾驶员处于头部异常姿态下。头部异常姿态时长的获取公式如下：

$$time = \frac{F}{FR} \quad (4-12)$$

式中：time 为头部异常姿态时长，F 为头部异常姿态帧数，FR 为视频帧率。

4.4 本章小结

本章介绍了如何在第 3 章获取疲劳特征区域与头部姿态角的基础上提取疲劳参数。在眼部区域，可以提取单位时间内 PERCLOS 均值与单位时间内眨眼次数作为疲劳判断依据；在嘴部区域，可以提取哈欠次数作为疲劳判断依据；通过头部姿态角，可以提取单位时间内点头次数与单位时间内头部异常姿态时长作为疲劳判断依据。

5 多特征融合疲劳检测

5.1 疲劳状态评价

目前疲劳判断没有统一的标准,本文采用 Ghoddosian R^[59]在文献中提出的疲劳等级判断方法,并使用其发布的开源数据集 UTA Real-Life Drowsiness Dataset(UTA-RLDD)。UTA-RLDD 对疲劳等级的定义如表 8 所示。

表 8 疲劳状态定义

Table8 Definition of fatigue state

疲劳分类	疲劳等级
警惕 (Alert)	1.极度警觉
	2.非常警觉
	3.警觉
	4.相当警觉
警惕性低 (Low Vigilant)	5.既不警觉也不困倦
	6.有一些困倦的迹象
	7.困倦, 但容易保持清醒
昏昏欲睡 (Drowsy)	8.困倦, 努力保持清醒
	9.极度困倦, 竭力抵抗入睡

Ghoddosian R 将 1 到 3 级疲劳定义为 Alert, 即受试者能保持警惕, 处于清醒状态, 可以轻松长时间驾驶; 4 到 6 级被定义为 Low Vigilant, 即受试者出现困倦的迹象或是不需要警惕的细微状况, 此时受试者可能可以驾驶但不鼓励这么做; 7 到 9 级被定义为 Drowsy, 即受试者需要努力抵抗睡眠, 此时不可进行驾驶。

UTA-RLDD 数据集由 60 位志愿者拍摄的视频组成。志愿者被要求在三种状态下, 使用任意摄像设备拍摄自己当前状态, 每段视频时长 10 分钟。所有视频都是以双目可见的角度拍摄, 并且要求相机放置在距离拍摄对象一臂的距离内, 志愿者还被要求进行阅读、观看视频和发呆动作, 这些操作可以模拟在汽车中获取的视频的状态, 例如将摄像头设置在仪表盘上拍摄, 驾驶员专注路况或是候车等情况。数据集中部分帧图像如图 35 所示。



图 35 UTA-RLDD 图例

Fig35 Sample of UTA-RLDD

5.2 多特征融合疲劳检测算法

5.1.1 CART 树

CART (Classification And Regression Trees) 意为决策与回归树，即该算法既可以用来进行分类任务，也可以进行回归任务。与传统的 IDE3 算法相比，CART 算法可以同时处理离散型特征与连续型特征，适用于本文所获得的疲劳特征；与 C4.5 算法相比，CART 算法生成的决策树为二叉树，不易产生数据碎片，精度往往比多叉树更高^[65]。故本文选用 CART 算法作为基分类器构造多特征融合疲劳检测方法。

1. 特征选择

决策树每一次对数据集的划分都依赖于特征选择，而找到最优特征的前提是找到最优切分点。CART 算法使用基尼指数 (Gini index) 来衡量每一个可能的切分点划分后数据集的纯度，其公式如下所示：

$$Gini(D) = \sum_{k=1}^{|y|} \sum_{n \neq k} p_k p_n = 1 - \sum_{k=1}^{|y|} p_k^2 \quad (5-1)$$

式中： $|y|$ 表示数据集 D 中的样本总类别数， p_k 表示第 k 类样本所占的比例， $Gini(D)$ 数值越小，数据集 D 的纯度越高。 $Gini(D)$ 反映了从数据集 D 中随机抽取两个样本，其类别不一致的概率。在属性 a 上划分后，得到的 D_1 与 D_2 两个子集的基尼指数计算公式如下：

$$Gini(D|a) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (5-2)$$

算法选择划分后基尼指数最小的划分点作为最优划分进行树的分支。针对不同的特征类型，算法采用不同的划分方法。

1.连续型特征划分方法

假设 $\{a_1, a_2, \dots, a_m\}$ 为数据集 D 中某个连续型特征 A 的取值并按升序排列。取这些连续值相邻两个值的均值作为划分点，共得到 $m-1$ 种划分方式，计算这 $m-1$ 种划分方式对数据集 D 进行划分后得到的子集的基尼指数，取使基尼指数最小值的划分点作为该特征 A 上的最佳划分。

2.离散型特征划分方法

假设 $\{b_1, b_2, \dots, b_n\}$ 为数据集 D 中某个离散型特征 B 的取值。对于任意离散值 b_k ，将数据集 D 划分为包含特征值 b_k 与不包含特征值 b_k 的俩部分，共得到 n 种划分方式，计算这 n 种划分方式对数据集 D 进行划分后得到的子集的基尼指数，取使基尼指数最小值的划分点作为该特征 B 上的最佳划分。

在得到全部特征上的最佳划分后，比较其基尼指数大小，取基尼指数最小的特征与其最佳划分点进行数据集划分，即 CART 树的分叉。

2.树的生成

树的生成算法如下所示：

算法 5.1 CART 树的生成

输入：数据集 D ，停止计算的条件

输出：二叉决策树

1.判断当前数据集是否满足停止条件，若满足则算法结束，不满足则以当前数据集为根结点。

2.遍历当前数据集上的特征，判断其性质（离散型或连续型），使用对应方法

找到其最佳划分。比较每个特征上最佳划分后的基尼指数，选择基尼指数最小的特征为最优特征。

3.将当前数据集以最优特征进行最佳划分得到俩个子集 D_1, D_2 ，建立当前结点的子结点，左结点为 D_1 ，右结点为 D_2 。将此最优特征的最佳划分点去除。

4.对 D_1, D_2 重复进行操作 1~4。

算法的停止条件有多种：1) 当前数据集基尼指数小于某一指定阈值，即认为当前数据集不需要再划分；2) 当前数据集每个特征上仅有一个值，无法划分；3) 当前数据集样本数小于某一指定阈值，即认为当前数据集没有划分意义；4) 当前 CART 树深度达到预设值，停止分裂预防过拟合。

5.1.2 随机森林

过拟合是决策树学习算法面对的重要问题，在决策树划分过程中，有时会出现分支过多，过度的学习训练集，将一些训练集自身的特点或偏好作为普适的特征学习，造成决策树过拟合，泛化性能下降。为了解决过拟合问题，国内外学者提出了诸多办法：预剪枝、后剪枝和分类器组合等办法。随机森林是分类器组合方法中的一种，其理论来自于统计学习。

随机森林利用 bootstrap 方法从原始样本中进行随机放回抽样，最终得到 k 个与原始样本相同大小的数据集，使用这些数据集得到 k 个决策树。在面对新的输入时，使用 k 个决策树对输入进行预测，使用给定的结合策略对 k 个决策树的结果进行结合，输出最终结果。在树的生成过程中，对决策树的每一个结点，首先从该结点的属性集合中随机抽取包含 m 个属性的子集，然后从属性子集中选取最优划分进行树的分裂，在一般情况下，推荐 $m = \log_2 d^{[66]}$ 。

随机森林生成算法如下所示：

算法 5.2 随机森林的生成

输入： 训练集 $D = \{\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_n, y_n\}\}$ ；生成决策树数量 k ；

输出： k 个 CART 决策树

过程：

1.随机放回抽样生成 k 个大小为 n 的数据集，记为 $D_r = \{D_1, D_2, \dots, D_k\}$

2.for D_i in D_r :

3. 计算每次分裂时属性数量 d ，随机选择 $\log_2 d$ 个属性进行分裂

5.3 实验结果与分析

第4章所提取的疲劳参数大多为过程量，需要一定检测时间才可得出，同时为了获得更多样例，将 RLDD 数据集中的视频裁切为 30s 时长的短视频，共计获得 633 个样本，三类样本各 211 个。选取其中 501 个样本进行训练，132 个样本进行测试。测试数据集中五个疲劳参数的取值方法如下：

1. PERCLOS 值疲劳参数取 30s 内的 PERCLOS 平均值；2. 眨眼次数疲劳参数取 30s 内眨眼次数；3. 哈欠次数疲劳参数取 30s 内哈欠次数；4. 点头次数疲劳参数由 30s 内点头次数；5. 头部异常姿态时长疲劳参数取 30s 内头部异常姿态时间。

部分样本的五个疲劳参数如下所示：

表 9 部分样本数据

Table9 Partial sample data

疲劳类型	PERCLOS 均值	眨眼次数	哈欠次数	点头次数	头部异常姿态时长
Alert	0.072	3	0	0	6.367
Alert	0.036	7	0	0	4.333
Low Vigilant	0.143	7	0	3	17.833
Low Vigilant	0.194	8	0	2	16.2
Drowsy	0.896	3	2	4	19.833
Drowsy	0.452	12	0	2	25.433

使用训练集训练一颗 CART 树，训练结果如图 36 所示。

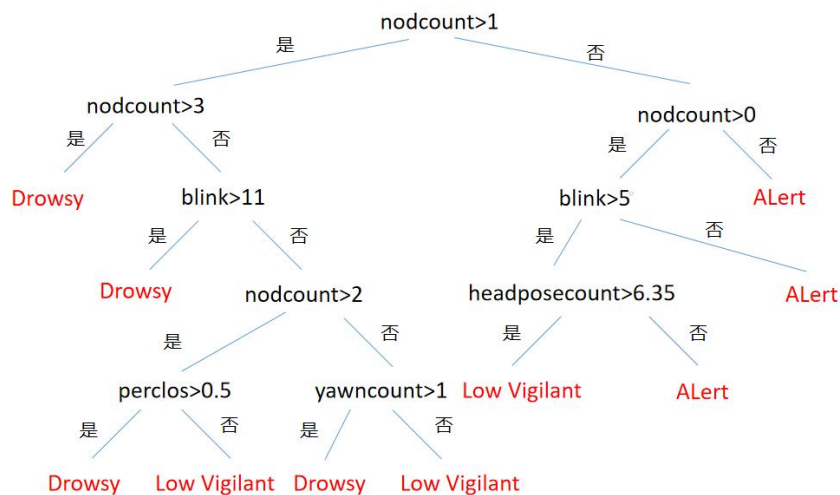


图 36 单棵 CART 决策树

Fig36 Single CART decision tree

在训练集中，决策树的分类结果如表 10 所示。

表 10 单棵决策树在训练集上的分类结果

Table10 Classification result of single decision tree on training set

样本类别	样本个数	正确识别样本数	错误识别样本数	准确率
Alert	167	162	5	97.0%
Low Vigilant	167	156	11	93.4%
Drowsy	167	167	0	100%

由表 10 可以发现，单棵 CART 树对 Alert 与 Low Vigilant 的分类失误较多。这是由于这俩类数据特征相近，单棵 CART 树无法很好的区分这俩者，在叶结点出现样本数量低于停止条件但是当前集合仍不纯净的情况。随机森林可以很好的解决单棵树分类能力不足的缺点，本文所提随机森林采用投票法得出结果，共由 50 棵 CART 决策树构成。在训练集中，随机森林的分类结果如表 11 所示。

表 11 随机森林在训练集上的分类结果

Table11 Classification results of random forests on training sets

样本类别	样本个数	正确识别样本数	错误识别样本数	准确率
Alert	167	164	3	98.2%
Low Vigilant	167	161	6	96.4%
Drowsy	167	167	0	100%

由表 11 可以看出，随机森林在训练集上的分类能力相较单棵 CART 树更强。

使用单特征 *perclos* 值与点头次数分别在训练集上建立 CART 决策树，四种决策树在测试集上的表现如表 12 所示。

表 12 不同决策树在测试集上的分类结果

Table Classification results of different decision trees on test sets

决策树类别	样本个数	正确识别样本数	错误识别样本数	准确率
基于 <i>perclos</i> 的决策树	132	70	62	53.0%
基于点头次数的决策树	132	86	46	65.2%
单棵多特征决策树	132	104	28	78.8%
随机森林	132	127	5	96.2%

由表 12 可以得出，多特征决策树性能优于单特征决策树，单特征决策树无法将样本正确分类。随机森林性能优于单棵决策树，对特征相近的样本分类能力更强。

本文对不同文献的疲劳检测方法进行对比, 实验结果如表 13 所示。

表 13 不同疲劳检测方法对比

Table13 Comparison of different fatigue testing methods

文献	疲劳检测方法	准确率
文献[67]	无监督学习+面部特征提取+K 均值聚类	87.5%
文献[68]	PERCLOS+哈欠次数+车辆运行参数	90%
文献[69]	DINN+LBF+MSPM+ASPM+PERCLOS	90.15%
文献[70]	EnlightenGAN+MTCNN+ESC-Net	93.79%
文献[42]	人眼状态+SSD	96%
本文	多特征随机森林	96.2%

文献[67]使用无监督学习的方法直接从面部提取特征, 对得到的特征使用 Softmax 预测的类别概率来改进 K 均值聚类算法中的均值向量初始化方法, 进行驾驶员疲劳状态类别划分。使用混淆矩阵进行检测结果的评价, 其中疲劳状态的检测正确率达到 87.5%。

文献[68]使用 HOG 与 SVM 结合完成人脸检测, 使用级联回归树得到人脸特征点定位眼部区域与嘴部区域。结合 PERCLOS、哈欠次数与车辆运行时长综合判断驾驶员疲劳状态, 其检测准确率达到 90%。

文献[69]使用深度混合神经网络 (DINN) 识别驾驶员眼睛开闭状态, 提出了基于时空局部二值模式 (时空 LBF) 和 Gabor 融合特征的驾驶员动态疲劳表情识别模型。使用 9 个特征建立特征空间, 将通用疲劳特征空间和自适应疲劳特征空间结合构建一种驾驶员自适应疲劳判别模型, 检测驾驶员疲劳状态, 其准确率达到 90.15%。

文献[70]针对低光条件下识别率地下的问题, 采用 EnlightenGAN 对夜间图像进行增强, 使用 MTCNN 定位眼部区域并判断眼睛睁闭, 采取 ESC-Net 进行疲劳分类, 最终检测准确率达到 93.19%

文献[42]采用深度卷积神经网络 (SSD) 进行目标检测, 直接得出人脸状态, 精确率达 99.7%。使用模糊评价思想与单指标批判方法进行疲劳程度判定, 在测试集上的总体准确率为 96%。

6 总结与展望

6.1 总结

随着经济水平的快速发展，私家车的数量不断增多。大量新增驾驶员缺乏驾驶经验，对驾驶技能不够重视，同时工作时间的普遍延长致使人们经常无法获得足够的休息，这些因素导致疲劳驾驶风险的升高，进而引发交通事故。在前人工作的基础上，本文提出了一种多特征融合车载疲劳检测的方法，旨在当驾驶员陷入疲劳状态时及时提醒驾驶员，避免因此引发的交通事故。本文的研究内容包括以下几点：

1.介绍了本文的研究背景。总结了以往的疲劳检测方法，包括主观疲劳检测法与客观疲劳检测法，客观检测法包含以下几类：基于驾驶员生理信号特征的疲劳检测方法；基于车辆行为特征的疲劳检测方法；基于驾驶员行为特征的疲劳检测方法。

2.介绍了本文所使用的相关理论与技术。图像增强算法历史悠久，是大部分图像处理工作的前置工序；卷积神经网络在人脸识别领域有大量应用，是很好的人脸定位工具；人脸特征点检测算法形式多样，主动形状模型与神经网络都可以很好的完成人脸特征点检测。

3.优化了特征区域的获取方法。本文优化了特征区域获取流程，提出了多任务卷积神经网络获取人脸区域与头部姿态信息，使用 ERT 人脸特征点检测算法获取眼部区域与嘴部区域。

4.提出了从特征区域提取疲劳参数的方法。使用 EAR 参数从眼部区域提取 PERCLOS 值与单位时间内眨眼次数；使用 MAR 参数从嘴部区域提取哈欠次数；使用 pitch 角、roll 角与 yaw 角从头部姿态信息中提取单位时间内点头次数与头部异常姿态持续时间。

5.提出了多特征融合疲劳检测方法。使用 UTA-RLDD 定义的疲劳状态等级与数据，使用以 CART 树为基的随机森林作为多特征特征融合疲劳检测方法。在自制测试集上取得了 96.2%的准确率。

6.2 展望

本文研究了基于驾驶员行为特征的疲劳检测方法，综合了单位时间内眨眼次数、哈欠次数、点头次数、PERCLOS 均值与头部异常姿态持续时间五个疲劳参数判断驾驶员疲劳状态。回顾实验进行中遇到的问题，发现仍有如下几点缺点与改进方法：

1.疲劳数据集中的视频帧率均为 30FPS。在进行 PERCLOS 值计算时，由于视频帧率的限制，包含眨眼动作的图像中存在闭眼图片缺失的情况，由此计算的部分 PERCLOS 值偏小。可以在录制数据集时提高视频帧率来避免这种现象。

2.本文的 EAR 阈值测定仅采用了疲劳数据集中的 11 名受测人员的视频测定，样本数据量少，在实际场景中使用预计效果不佳。解决方法有俩种：扩大样本量确保 EAR 阈值具有普适性；提出一种自适应 EAR 阈值方法，在车辆启动后对当前驾驶员进行采样并确定其个人的 EAR 阈值。

3.本文提出的疲劳检测算法仅融合了驾驶员行为特征的疲劳参数，在实践中可以针对不同车型在模型中加入车辆运行时的参数提高疲劳检测算法的检测能力。

4.本文所采取的特征区域定位方法为多任务卷积神经网络+ERT 人脸特征点检测算法，检测流程虽有优化，但仍旧相对复杂缓慢，在未来的研究中可以将一次性检出特征区域作为研究方向。

参考文献

- [1] 中华人民共和国国家统计局.中国统计年鉴[J].北京:中国统计出版社,2021.
- [2] Gander P H, Marshall N S, James I, et al. Investigating driver fatigue in truck crashes: Trial of a systematic methodology[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2006, 9(1): 65-76.
- [3] Forsman P M, Vila B J, Short R A, et al. Efficient driver drowsiness detection at moderate levels of drowsiness[J]. Accident Analysis & Prevention, 2013, 50: 341-350.
- [4] 牛清宁,周志强,于鹏程,王秋鸿.国内外货运车辆驾驶人疲劳驾驶管理政策研究[J].交通工程,2017,17(06):45-48.DOI:10.13986/j.cnki.jote.2017.06.009.
- [5] 薛雷. 考虑驾驶员生物电信号的疲劳驾驶检测方法研究[D].吉林大学,2015.
- [6] 南姣芬. 基于脑电信号的驾驶疲劳检测方法研究[D].陕西师范大学,2011.
- [7] 王海玉. 基于脑电信号多熵特征的驾驶疲劳检测模型研究[D].江西农业大学,2019.DOI:10.27177/d.cnki.gjxnu.2019.000242.
- [8] 王斐,吴仕超,刘少林,张亚徽,魏颖.基于脑电信号深度迁移学习的驾驶疲劳检测[J].电子与信息学报,2019,41(09):2264-2272.
- [9] 吴雅萱,李文书,施国生,周涛.疲劳驾驶检测技术研究综述[J].工业控制计算机,2011,24(08):44-46+49.
- [10] 郭玮珍,郭兴明,万小萍.以心率和心率变异性为指标的疲劳分析系统[J].医疗卫生装备,2005(08):1-2.
- [11] 郝贺. 基于心率变异性的人体精神疲劳检测系统研究[D].东北大学,2014.
- [12] Ohsuga M , Kamakura Y , Inoue Y , et al. Classification of Blink Waveforms Toward the Assessment of Driver's Arousal Levels - An EOG Approach and the Correlation with Physiological Measures. 2007.
- [13] 于立娇,吴振昕,王文彬,高洪伟.驾驶员疲劳状态监测系统综述[J].汽车文摘,2019(03):24-30.
- [14] Fletcher, L, Apostoloff, et al. Vision in and out of vehicles[J]. Intelligent Systems, IEEE, 2003.
- [15] Pilutti T , Ulsoy G . On-line identification of driver state for lane-keeping tasks[C]// Proceedings of 1995 American Control Conference - ACC'95. IEEE, 1995.
- [16] Bertozzi M , Broggi A . Real-Time Lane and Obstacle Detection on the GOLD System[C]// Intelligent Vehicles Symposium, 1996. Proceedings of the 1996 IEEE.

IEEE, 1996.

[17] Mao M , Du L . Research on drive fatigue detection using wavelet transform[C]//IEEE International Conference on Vehicular Electronics & Safety. IEEE, 2008.

[18] Kithil P W. Driver alertness detection research using capacitive sensor array[C]//Driving Assesment Conference. University of Iowa, 2001, 1(2001).

[19] 邹昕彤,王世刚,赵文婷,赵晓琳,李天舒.基于眼睛与嘴部状态识别的疲劳驾驶检测[J].吉林大学学报(信息科学版),2017,35(02):204-211.DOI:10.19292/j.cnki.jdxxp.2017.02.015.

[20] 刘忠旭. 基于面部特征的驾驶员疲劳状态检测[D].长春工业大学,2022.DOI:10.27805/d.cnki.gccgy.2022.000640.

[21] 邬敏杰,穆平安,张彩艳.基于眼睛和嘴巴状态的驾驶员疲劳检测算法[J].计算机应用与软件,2013,30(03):25-27+54.

[22] 刘祝华. 图像去噪方法的研究[D].江西师范大学,2005.

[23] 关新平,赵立兴,唐英干.图像去噪混合滤波方法[J].中国图象图形学报,2005(03):332-337.

[24] 赵高长,张磊,武风波.改进的中值滤波算法在图像去噪中的应用[J].应用光学,2011,32(04):678-682.

[25] Vaillant R, Monrocq C, Le Cun Y. Original approach for the localisation of objects in images[J]. IEE Proceedings-Vision, Image and Signal Processing, 1994, 141(4): 245-250.

[26] Li H, Lin Z, Shen X, et al. A convolutional neural network cascade for face detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 5325-5334.

[27] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE signal processing letters, 2016, 23(10): 1499-1503.

[28] Yu J, Jiang Y, Wang Z, et al. Unitbox: An advanced object detection network[C]//Proceedings of the 24th ACM international conference on Multimedia. 2016: 516-520.

[29] Opitz M, Waltner G, Poier G, et al. Grid loss: Detecting occluded faces[C]//European conference on computer vision. Springer, Cham, 2016: 386-402.

- [30] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [31] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.
- [32] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [33] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
- [34] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [35] Chen D, Ren S, Wei Y, et al. Joint cascade face detection and alignment[C]//European conference on computer vision. Springer, Cham, 2014: 109-122.
- [36] Mathias M, Benenson R, Pedersoli M, et al. Face detection without bells and whistles[C]//European conference on computer vision. Springer, Cham, 2014: 720-735.
- [37] Yang S, Luo P, Loy C C, et al. From facial parts responses to face detection: A deep learning approach[C]//Proceedings of the IEEE international conference on computer vision. 2015: 3676-3684.
- [38] Jin X, Tan X. Face alignment in-the-wild: A survey[J]. Computer Vision and Image Understanding, 2017, 162: 1-22.
- [39] Cootes T F, Taylor C J. Active shape models— ‘smart snakes’ [M]//BMVC92. Springer, London, 1992: 266-275.
- [40] Cootes T F, Edwards G J, Taylor C J. Active appearance models[J]. IEEE Transactions on pattern analysis and machine intelligence, 2001, 23(6): 681-685.
- [41] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[C]//Proceedings of the IEEE conference on computer vision and pattern

recognition. 2013: 3476-3483.

[42] 王迪. 基于人眼状态的疲劳检测算法研究与应用[D].电子科技大学,2020.DOI:10.27005/d.cnki.gdzku.2020.001881.

[43] 牛正存. 驾驶员疲劳特征提取方法的研究及检测系统的设计[D].电子科技大学,2017.

[44] 马艳. 基于颜色与模板匹配的人脸检测方法[D].大连理工大学,2006.

[45] 郑伟成,李学伟,刘宏哲,代松银.基于深度学习的疲劳驾驶检测算法[J].计算机工程,2020,46(07):21-29.DOI:10.19678/j.issn.1000-3428.0055912.

[46] 王剑平,张捷.小波变换在数字图像处理中的应用[J].现代电子技术,2011,34(01):91-94.DOI:10.16652/j.issn.1004-373x.2011.01.049.

[47] B. Vidakovic and C. B. Lozoya, "On time-dependent wavelet denoising," in IEEE Transactions on Signal Processing, vol. 46, no. 9, pp. 2549-2554, Sept. 1998, doi: 10.1109/78.709544.

[48] Donoho, D. L. De-noising by soft-thresholding[J]. IEEE Transactions on Information Theory, 2002, 41(3):613-627.

[49] Donoho D L, Johnstone I M. Adapting to Unknown Smoothness via Wavelet Shrinkage[J].

[50] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[51] Ruiz N, Chong E, Rehg J M. Fine-Grained Head Pose Estimation Without Keypoints[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018.

[52] Yang S, Luo P, Loy C C, et al. Wider face: A face detection benchmark[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 5525-5533.

[53] Fanelli G, Dantone M, Gall J, et al. Random forests for real time 3d face analysis[J]. International journal of computer vision, 2013, 101(3): 437-458.

[54] Tang X, Du D K, He Z, et al. Pyramidbox: A context-assisted single shot face detector[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 797-813.

- [55] Wang H, Li Z, Ji X, et al. Face r-cnn[J]. arXiv preprint arXiv:1706.01061, 2017.
- [56] Bulat A, Tzimiropoulos G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1021-1030.
- [57] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1867-1874.
- [58] Zhu X , Zhen L , Liu X , et al. Face Alignment Across Large Poses: A 3D Solution[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
- [59] Ghoddoosian R, Galib M, Athitsos V. A realistic dataset and baseline temporal model for early drowsiness detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 0-0.
- [60] 薛丽. 基于改进 MTCNN 网络的多特征疲劳驾驶检测研究[D].长安大学,2021.DOI:10.26976/d.cnki.gchau.2021.001197.
- [61] 朱名流,李顶根.基于人脸特征点的疲劳检测方法研究[J].计算机应用研究,2020,37(S2):305-307.
- [62] 徐忠帅. 基于面部特征的驾驶员疲劳检测研究[D].辽宁工业大学,2019.
- [63] Abtahi S , Omidyeganeh M , Shirmohammadi S , et al. YawDD: A yawning detection dataset[C]// ACM Multimedia Systems (MMSys). ACM, 2014.
- [64] 陈书明,陈美玲.头部姿态估计技术研究综述[J].泉州师范学院学报,2015,33(06):78-85.DOI:10.16125/j.cnki.1009-8224.2015.06.015.
- [65] 栾丽华,吉根林.决策树分类技术研究[J].计算机工程,2004(09):94-96+105.
- [66] Breiman, L. Random Forests. Machine Learning 45, 5 - 32 (2001).
- [67] 王静静. 基于深度学习的驾驶员疲劳检测 [D]. 合肥工业大学,2020.DOI:10.27101/d.cnki.ghfgu.2020.000088.
- [68] 王仕卿. 货运车辆疲劳驾驶检测方法研究 [D]. 吉林大学,2021.DOI:10.27162/d.cnki.gjlin.2021.006918.
- [69] 赵磊. 基于深度学习和面部多源动态行为融合的驾驶员疲劳检测方法研究 [D].山东大学,2018.
- [70] 李晓星. 基于深度学习的疲劳驾驶检测方法研究[D].中国科学技术大

学,2020.DOI:10.27517/d.cnki.gzkju.2020.000652.