

学校代号 10532

学 号 S171000840

分 类 号 TP391

密 级



湖南大学  
HUNAN UNIVERSITY

## 硕士学位论文

# 基于深度学习多特征融合的疲劳检测算法研究

学位申请人姓名 李皓星

培 养 单 位 信息科学与工程学院

导师姓名及职称 张汗灵 教授

学 科 专 业 信息与通信工程

研 究 方 向 行为识别

论文提交日期 2020 年 4 月 29 日

学校代号：10532

学 号：S171000840

密 级：

## 湖南大学硕士学位论文

# 基于深度学习多特征融合的疲劳检测算法研究

学位申请人姓名：李皓星

导师姓名及职称：张汗灵 教授

培 养 单 位：信息科学与工程学院

专 业 名 称：信息与通信工程

论文提交日期：2020 年 4 月 29 日

论文答辩日期：2020 年 5 月 11 日

答辩委员会主席：王春华 教授

Research on Fatigue Detection Algorithm Based on Deep Learning  
with Multi-feature Fusion

by

LI Haoxing

B.E. (Xiangtan University) 2017

A thesis submitted in partial satisfaction of the

Requirements for the degree of

Master of Engineering

in

Information and Communication Engineering

in the

Graduate School

of

Hunan University

Supervisor

Professor ZHANG Hanling

April, 2020



## 摘 要

疲劳驾驶引起的交通事故给人们带来了惨痛的损失。如何快速获得驾驶员的疲劳特征，如何检测其疲劳等级，已成为疲劳检测研究领域的热点。我国对其研究起步较晚，预防疲劳驾驶的检测技术尚不完善，因此研究一种快速高效的疲劳驾驶检测方法对于预防疲劳驾驶有着重要的意义和实用价值。在这项研究中，作者提出了一种基于神经网络多特征融合的疲劳检测方法。论文主要研究工作有：

1. 图像预处理和人脸检测与面部特征提取。因为图像在采集和存储过程中会受到噪声和光照等影响导致图像质量下降。因此必需要对图像去噪和光照均衡处理。然后使用 MTCNN (Multi-task convolutional neural network) 进行人脸检测，可以获得面部 5 个关键点。在获得面部关键点后，提取面部嘴巴和眼睛区域。

2. 提出了一个基于卷积神经网络的 SC-Net (small state classification network) 和一个基于卷积神经网络和长短期记忆网络的 HP-LSTM (hope long short-term memory model)。通过 SC-Net 识别提取的眼睛和嘴巴区域，可以检测到眼睛和嘴巴的状态。在进行人脸识别的同时，通过 HP-LSTM 对图像序列进行头部姿态估计，可以获得当前帧图像是否处于点头动作区间的概率值。经对比发现使用基于卷积神经网络的方法识别面部特征状态和估计头部姿态的准确率高。

3. 根据眼、嘴、头的面部特征状态，提出了一种多特征加权求和的疲劳检测方法，并构建了疲劳检测评估基准。将获得的点头状态参数、眨眼频率参数、PERCLOS (Percent Eye Closure) 参数、持续闭眼时间参数、打哈欠参数加权求和，获得最终的疲劳特征值，评估驾驶员的状态等级。实验结果表明，我们的方法对复杂环境具有高鲁棒性，可以有效地检测疲劳。

关键词：深度学习；疲劳检测；神经网络；人脸检测；头部姿态估计

## Abstract

Traffic accidents caused by fatigue driving have brought painful losses. How to quickly obtain the driver's fatigue characteristics and detect fatigue level has become a hot spot in the research field of fatigue detection. This research started late in China, and the detection technology for preventing fatigue driving is not perfect. Therefore, it is of great significance and practical value to propose a fast and efficient fatigue driving detection method for preventing fatigue driving. In this study, the author proposes a fatigue detection method based on neural networks with multi-feature fusion. The main research work of this study are:

1. Image preprocessing, face detection and facial feature extraction. The image may be damaged by noise and light during the process of acquisition and storage, which will reduce the image quality. Therefore image denoising and lighting equalization are necessary. Then this paper uses the Multi-task convolutional neural network (MTCNN) for face detection, and can obtain five key points of the face at the same time. After obtaining facial key points, extract the mouth and eye area.

2. The author proposes a small state classification convolutional neural network (SC-Net) and a hope long short-term memory model (HP-LSTM) based on convolutional neural networks and long short-term memory (LSTM) unit. By detecting the eye and mouth areas through SC-Net, the states of eyes and the states of mouth can be detected. While performing face detection, the HP-LSTM is used to estimate the head pose of image sequence to obtain the probability value whether the current frame image is in the nodding states. Through comparison, it is found that the method based on convolutional neural network has high accuracy in detecting facial feature states and estimating head poses.

3. According to the facial features of eyes, mouth, and head, a multi-feature weighted sum fatigue detection method is proposed, and a fatigue detection evaluation benchmark is constructed. The obtained nod state parameters, blink frequency parameters, PERCLOS parameters, eye closure duration parameters, and yawn parameters are weighted and summed to obtain the final fatigue characteristic value and evaluate the driver's state level. Experimental results show that our method is robust to complex environments and can effectively detect fatigue.

**Keywords:** Deep Learning; Fatigue Detection; Convolutional Neural Network; Face Detection; Head Pose Estimation

# 目 录

学位论文原创性声明 .....	I
摘 要 .....	II
Abstract .....	III
目 录 .....	IV
插图索引 .....	VI
附表索引 .....	VII
第 1 章 绪论 .....	1
1.1 研究的背景与意义 .....	1
1.2 疲劳检测方法概述 .....	2
1.2.1 基于车辆状态的疲劳检测方法 .....	2
1.2.2 基于生理信号的驾驶员疲劳检测方法 .....	2
1.2.3 基于面部行为特征的检测方法 .....	3
1.3 驾驶员疲劳检测国内外研究现状 .....	4
1.4 本文组织结构与主要内容 .....	6
第 2 章 深度学习相关理论知识 .....	9
2.1 卷积神经网络相关介绍 .....	9
2.2 循环神经网络相关介绍 .....	10
第 3 章 图像预处理及人脸检测算法 .....	12
3.1 图像预处理 .....	12
3.1.1 图像去噪 .....	12
3.1.2 光照补偿处理 .....	14
3.1.3 实验及分析 .....	15
3.2 基于 MTCNN 的人脸检测算法 .....	16
3.2.1 人脸检测算法介绍 .....	16
3.2.2 基于 MTCNN 的人脸检测算法 .....	18
3.2.3 实验结果及分析 .....	20
3.3 眼嘴特征提取 .....	22
3.4 本章小结 .....	24
第 4 章 面部特征状态检测及头部姿态估计 .....	25
4.1 眼睛状态检测 .....	25

4.1.1 眼睛状态检测常用方法 .....	25
4.1.2 基于卷积神经网络的眼睛状态识别 .....	25
4.1.3 眼部数据集 .....	26
4.1.4 实验与结果及分析 .....	27
4.2 嘴巴状态 .....	29
4.2.1 嘴巴状态检测 .....	29
4.2.2 嘴巴数据集 .....	29
4.2.3 实验及结果分析 .....	30
4.3 头部状态估计 .....	31
4.3.1 头部状态估计定义 .....	31
4.3.2 基于 HP-LSTM 的点头状态估计 .....	32
4.3.3 实验结果分析 .....	33
4.4 本章小结 .....	34
第 5 章 基于特征融合的疲劳检测 .....	36
5.1 特征融合方法及疲劳评价基准 .....	36
5.2 面部特征参数提取 .....	37
5.2.1 眼睛信息提取 .....	37
5.2.2 嘴巴信息提取 .....	39
5.2.3 头部信息提取 .....	40
5.3 基于加权平均的疲劳检测 .....	41
5.4 实验结果及分析 .....	42
5.5 本章小结 .....	44
结论 .....	45
参考文献 .....	47
附录 A 攻读硕士学位期间发表论文目录 .....	52
附录 B 攻读硕士学位期间参与的科研活动 .....	53
致 谢 .....	54



# 插图索引

图 3.1 滤波结果对比图.....	15
图 3.2 有无光照补偿处理的差别，左图为原始图像，右图为处理后的图像.....	16
图 3.3 垂直和旋转形式的人脸形状模板.....	17
图 3.4 基于 MTCNN 的人脸检测算法示意图.....	19
图 3.5 P-Net, R-Net, O-Net 的结构示意图.....	19
图 3.6 基于 MTCNN 人脸检测算法的检测结果.....	21
图 3.7 提取面部特征结果.....	23
图 3.8 MTCNN 网络人脸识别及提取眼嘴区域示意图.....	23
图 4.1 SC-Net 网络结构.....	26
图 4.2 CEW 数据和 ZJU 数据的示例，a 是 ZJU 数据集，b 是 CEW 数据集.....	27
图 4.3 部分自收集眼睛数据示例.....	27
图 4.4 用 SC-Net 检测眼睛状态的训练过程.....	28
图 4.5 打哈欠嘴巴过程状态图.....	29
图 4.6 部分自收集嘴巴数据示例.....	30
图 4.7 用 SC-Net 检测嘴巴状态的训练过程.....	30
图 4.8 欧拉角表示头部姿态的示意结构.....	32
图 4.9 训练 hopenet: 合并交叉损失和均方误差损失的 ResNet50 结构.....	33
图 4.10 改进的 HP-LSTM 网络结构.....	33
图 5.1 疲劳检测评价基准.....	37
图 5.2 $P_{80}$ 基准原理图.....	38
图 5.3 一次睁闭眼过程示意图.....	39
图 5.4 瞌睡状态下点头区间概率图.....	40
图 5.5 基于深度学习多特征融合的驾驶员疲劳检测方法的检测流程图.....	42
图 5.6 模拟瞌睡视频测试示例.....	43

## 附表索引

表 3.1 各算法对添加高斯噪声图片的性能 .....	16
表 3.2 各算法对添加椒盐噪声图片的性能 .....	16
表 3.3 MTCNN 算法在各数据集上的准确率及效率对比 .....	22
表 4.1 在自收集数据集上使用不同网络进行眼睛状态检测的性能比较 .....	28
表 4.2 CEW 数据集和 ZJU 数据集上各种分类器的比较 .....	28
表 4.3 在自收集数据集上使用不同网络进行嘴爸状态检测的性能比较 .....	30
表 4.4 具有不同时间步长 $T$ 在验证集上的比较 .....	34
表 4.5 与一些先进方法在验证集上的精确度比较 .....	34
表 5.1 疲劳状态特征参数的权重表 .....	41
表 5.2 加权疲劳参数与疲劳状态对应表 .....	41
表 5.3 疲劳检测最终检测结果 .....	43
表 5.4 疲劳检测准确率对比 .....	43

# 第 1 章 绪论

## 1.1 研究的背景与意义

近期,随着人们生活水平的不断提高,拥有汽车的人数逐步增加。随着汽车的增加,它逐渐成为出行的主要交通工具。根据新华网近期的统计数据显示,截至 2019 年末,全国民用汽车保有量达到了 26150 万辆,其中私人汽车占比 85.5%,数量为 22635 万辆,比 2018 年末多了 1905 万辆。而民用轿车数量为 14644 万辆,比上一年增长 1193 万辆<sup>[1]</sup>。随着汽车的增加,其给人们生活带来快捷与方便的同时,汽车所引发频繁的交通事故也带来了惨重的经济损失,人民的生命也受到了威胁,道路交通事故已成为严重威胁人类生命财产安全的社会问题之一。根据 2018 年全球道路安全状况报告可以知,每年道路交通事故中的死亡人数已达到 135 万人,主要受害者是 5 至 29 岁的人群<sup>[2]</sup>,道路交通伤害现已成为全球范围内造成人员伤亡的最大原因之一。引发交通事故的因素有注意力下降、判断力丧失、目光分散、警觉性消失等,其主要原因为酒后驾驶、疲劳驾驶、超速和分心等危险行为。

驾驶疲劳,指的是驾驶员在持续的长时间行车过程中,产生了心理和生理机能的失和,相对正常情况的驾驶能力下降的状态。交通安全基金会的一项调查发现大 16%-21% 的交通事故是由于疲劳驾驶引起的<sup>[3]</sup>。根据国家统计局的统计,疲劳驾驶引起的交通事故占交通事故总数的 20% 以上,严重交通事故占总数的 40% 以上,说明疲劳驾驶是交通事故主要的原因之一。

疲劳驾驶的原因太多,比如睡眠不好、驾驶时长过长、食用能引发瞌睡的食物、心情差异,这些都将会影响司机在行进过程中的注意力、判断力、警觉能力和目光方向等。在疲劳驾驶的情况下,驾驶员会感觉瞌睡疲倦,四肢无力,注意力和判断力下降,甚至发生瞬间记忆模糊或精神恍惚,出现反应迟缓、动作变慢,对周围的场景没有正确的判断,从而不能正确的控制车辆,造成交通事故的发生,形成极其恶劣的后果,因此疲劳时应该禁止驾驶车辆。与酒后驾驶,超速驾驶和其他危险驾驶行为相比,疲劳驾驶的检测和治理难以解决,防止驾驶员困倦极为困难,同时由于驾驶员的个体差异及交通场景的差异,任何人任何时候都有可能发生疲劳驾驶事件,从而严重损害了驾驶员、乘客和行人的人身安全,因此探究高效快速的疲劳检测方法在防止交通事故的发生上有关键的作用和实用价值。

近期,各国已经意识到预防疲劳驾驶的重要性,开始对该领域进行研究,并开发了一些疲劳检测系统,在各大汽车厂商所生产汽车的高级辅助系统中得到了应用。我国对疲劳识别系统的研发相比国外开始偏晚,预防疲劳驾驶的检测技术尚不完善。根据我国国情,开发实时自动检测系统也刻不容缓。

## 1.2 疲劳检测方法概述

全球范围内疲劳驾驶所造成的安全事故，会使人们的经济损失严重、生命受到威胁。为防止危害的不可控，需要对疲劳检测技术进行研究与调查。经过相关调查与研究，发现随着科学技术的创新与发展，疲劳检测技术领域已受到国内外较多研究人员的关注。目前的疲劳检测方法是有多种形式的，但主要的检测驾驶员疲劳方法是从下面三个方面来进行检测与预警的。基于车辆状态的疲劳检测，基于生理信号的疲劳检测和基于驾驶员面部生理特征的疲劳检测。

### 1.2.1 基于车辆状态的疲劳检测方法

基于车辆状态的疲劳检测方法通过采集车辆的各项驾驶参数，与正常情况的驾驶参数作对比，分析整合多种参数的异常情况来检测驾驶员是否疲劳。该疲劳检测方法主要采用各种驾驶员行为，例如车辆的加速度，行驶速度，方向盘转动程度，方向盘的转动握力，制动踏板用力，换挡和车辆的偏离等<sup>[4-8]</sup>，需要车辆装置各种类型的传感器，用来收集车辆速度、方向盘转角、各种用力大小和发动机转速等各项传感器参数。通过综合分析上述得到的各项传感器参数可以间接性地判断驾驶员是否处于疲劳状态。

**车辆偏移检测：**通过雷达导航和计算机视觉等方法获得车辆轨迹、方向和行驶速度，判断驾驶员是否处于疲劳状态。当驾驶员疲劳时，车辆会偏离车道中线。

**行车速度检测：**当疲劳时，驾驶员无法在行车过程中正确控制恒定的行车速度、需要测量车辆的加速度和行车速度等参数来检测驾驶员是否疲劳。

**方向盘转角及用力：**当疲劳时，驾驶员在行车过程中对方向盘、制动踏板和油门的使用要比正常使用时有所差异。

基于车辆状态的驾驶员疲劳检测方法主要依赖于驾驶过程中实时采集的传感器数据，该方法比其他方法相对算法简单且技术成熟，所以该技术在疲劳检测领域广受关注且富有投资。但是，其分析结果容易受到个人驾驶习惯、天气、车辆特性及道路状况等外界环境因素的影响<sup>[4]</sup>。驾驶员的驾驶经验和道路环境会降低驾驶员疲劳识别的准确性。而且其只能在驾驶员将要引发安全事故时检测有异常情况，鲁棒性不强，不能提前警报。所以该技术的算法结果可以用于辅助检测疲劳而不能直接用于检测疲劳。

### 1.2.2 基于生理信号的驾驶员疲劳检测方法

经过相关调查与研究，当驾驶员感觉疲劳时，其身体机能会发生变化，驾驶员对环境变化的反应会有一定的延迟，体内的生理信号会产生非正常的波动。在早期的疲劳检测研究中，研究人员主要利用生物学知识来检测疲劳，采用了几种生理信号检测方法来识别驾驶员的疲劳。主要的生理信号测试包括脑电信号（EEG）、心电信号（ECG）和肌电信号（EMG）等。

**脑电信号（EEG）检测：**EEG 信号已被证明是疲劳驾驶检测最具预测性和可靠性的

信号之一，因为它是大脑活动的直接量度<sup>[9]</sup>。通过 EEG 信号可以直接获得驾驶员在行车过程中大脑神经细胞的电活动信号。经过相关调查与研究，EEG 可分解为四个基本信号波： $\alpha$ 、 $\beta$ 、 $\delta$  和  $\theta$  波，通过分析该四种波的变化规律能够精确判断驾驶员是否疲劳。

心电信号（ECG）检测：相关研究表明，在驾驶员逐渐感觉疲劳的过程中，其心率会有明显的规律性下降趋势，ECG 可以被用于辅助判断驾驶员是否疲劳。

肌电信号（EMG）检测：当驾驶员从清醒状态进入到疲劳状态时，EMG 信号的频率将会持续下降，其振幅会逐渐增大<sup>[10]</sup>。因此，通过分析 EMG 信号也能判断司机是否位于疲劳状况，并对其进行疲劳状态识别。

基于生理信号的驾驶员疲劳检测方法主要通过分析疲劳状态和正常状态之间的各种生理参数差异，来判断驾驶员是否疲倦。通常，该方法的识别准确率很高，能够客观、高效地识别驾驶员所处于的疲劳情况。然而，这种方法需要专业，昂贵的信号采集装置的支持。这种方法甚至必须附着在人体皮肤表面，这会引起驾驶员的不适，并且会因此影响驾驶员的正常操作而引发交通事故。因此，该方法很难在实际场景中得到应用，并且生理信号的个体差异大，容易受到驾驶员年龄、驾驶员性别、气候等影响，难以形成统一的标准来对驾驶员检测是否疲劳。

### 1.2.3 基于面部行为特征的检测方法

当驾驶员感觉疲劳时，其面部生理特征将不同于正常状态的生理特征。经过研究发现，使用模式识别及图像处理相关技术分析驾驶员的面部生理特征，可以有效地监测驾驶员的状态。比如眼睛状态信息（睁开/关闭）、瞳孔运动、睁眼程度、嘴巴张开状态、头部旋转和瞌睡表情等面部行为的变化可以显示驾驶员的疲劳状态。因为驾驶员的面部生理特征变化明显，易于检测，该方法是疲劳检测领域的主流和焦点。

眼睛状态信息检测：经过调查研究证明：疲劳程度与眼睛闭合时长有着密切关系，眼睛闭合时间越久，说明其疲劳程度越重。人在正常情况下眨眼一次眼睛闭合持续的时间约为 0.2s，每分钟眨眼 10-25 次，若驾驶过程中驾驶员眼睛持续闭合时间达到 0.5s 以上或者每分钟眨眼次数低于 10 次则容易发生安全事故<sup>[11]</sup>。

嘴巴张开状态检测：正常驾驶下嘴部是完全闭合的，在说话嬉笑时嘴巴的张开成度也不大，并且一次打哈欠时间大约有几秒钟。在这期间驾驶员全身神经、肌肉都会得到完全放松，并且闭目塞听，这是人在疲劳状态下自身的生理保护作用，表示大脑已经非常疲劳，需要进行休息睡眠。可根据驾驶员是否打哈欠判断驾驶员是否疲劳。

头部状态检测：在正常行车时，驾驶员头部状态不会出现点头或歪头的现象。当驾驶员感到疲劳的时候，全身的肌肉放松，头会下垂，下垂到一定程度后，有翻倒的趋势。所以，可以根据头部状态来判断驾驶员是否感觉瞌睡。

基于图像技术的优势在于提取的面部特征是驾驶员的非侵入性视觉信息，并且不受其他外部因素（例如道路环境和驾驶经验）的影响。由于基于计算机视觉的深度学习应

用发展快速,基于图像的驾驶员疲劳检测可以达到良好的性能。这种类型的方法因其非侵入性,低成本和友好的驾驶状态监视特性而受到广泛好评。尤其是基于面部生理特征方法的多特征融合技术,可以保证检测的准确性和可靠性。

### 1.3 驾驶员疲劳检测国内外研究现状

研究驾驶员疲劳检测技术可以有效地预防交通事故的发生,所以疲劳检测及识别瞌睡技术吸引了国内外研究学者的注意力。该技术不仅在学术领域蓬勃发展,且应用于现实领域,现有一些研究成果形成产品并投入应用。

欧盟等汽车公司资助的 AWAKE 项目是一个基于驾驶员行为特征的综合检测系统<sup>[12]</sup>。该系统参考了多种特征参数,他使用不同类型的传感器收集方向盘转角、轨道信息、方向盘转向握力、踏板踏动力和车辆加速度等车辆相关信息来对驾驶员的疲劳程度等级进行评估。当驾驶员出现疲劳状态时,该系统使用光学刺激及安全带颤动等方法对驾驶员进行提醒。Friedrichs F 等<sup>[13]</sup>选择 54 位驾驶员驾驶 10 辆车来开展驾驶实验,获取车道位置相关的 11 类变量,选择 LDA、KNN、高斯混合模型、贝叶斯分类器和卷积神经网络五类分类器模型构建疲劳检测模型,试验结果说明疲劳检测效果颇好。M.I. Chacon Murgui 等<sup>[14]</sup>开发了一款名为 AutoVue 的系统,在车内安装 CCD 摄像头获取驾驶数据,通过获取驾驶员的行车轨迹,若驾驶员的试车轨迹偏离原定车道时,该系统将对驾驶员进行预警。

Mao 等<sup>[15]</sup>首先通过车辆传感器获得车辆的车速、加速度等参数,然后进行去噪,接着,引入小波变换以从时域和小波系数中提取疲劳相关特征,最后构建了最小距离分类器来检测驾驶员是否疲劳。Z.L.Li 等<sup>[16]</sup>通过样本熵来分解车辆方向盘的旋转特征,提出样本熵在正常情况下比较大,说明可以通过样本熵来检测驾驶员是否疲劳。Li F 等<sup>[17]</sup>通过获得驾驶员对方向盘的握力,通过小波变换在频域中计算相应参数,后面使用小波变换分解提取特征向量,通过比较 SVM、KNN 和 LDA 这三种分类器的实验,得出方向盘握力检测可以有效地检测驾驶员是否疲劳的结论。Mao z 等<sup>[18]</sup>在频域和时域上分析了方向盘转动加速度、汽车加速度、速度和转向角四个参数,获得了七个指标为驾驶员疲劳参数;然后通过概率神经网络对参数处理,实验表明检测精度有 92%。

Zhang<sup>[19]</sup>选择 35 位驾驶员,收集他们行车时的 EEG 信号发现,驾驶员疲劳时  $\delta$  波相对频带比增加而相对熵明显减少,同时  $\theta$  波相对频带比也减少,分析驾驶员的 EEG 信号可以很好的推测他们的疲劳程度。罗希特等人<sup>[20]</sup>通过线性判别分析和支撑向量机分析 EEG 的特征,检测驾驶员的困倦状态。Kalpanadevi 和 Sangeetha<sup>[21]</sup>开发了一种用 ECG 电极和 PIC 微控制器实现疲劳检测的硬件系统,通过比较驾驶员的 ECG 信号与疲劳驾驶的 ECG 信号,推断其是否疲劳。Balasubramanian 等<sup>[22]</sup>通过模拟实验发现,当驾驶员从清醒状态进入到疲劳状态时,EMG 信号的频率将会持续下降,其振幅会逐渐增大。

Zhang 等<sup>[23]</sup>对 15 名 20-30 岁的健康男性采集 EEG 信号,提取其样本熵,发现样本熵在疲劳状态下明显下降,使用这一特性可以识别驾驶员瞌睡状况。Li 等<sup>[24]</sup>提出了心脏变异率(HRV)的概念,HRV 可在 ECG 信号中提取,并使用小波变换获取特征向量,然后使用 SVM 分类器得到驾驶员的疲劳结果。

美国高速公路管理局于 1998 年使用两类红外线(950 和 850nm)采集间隔帧的驾驶员脸部区域,接着将获得的连续图像相减得到瞳孔区域,分析瞳孔状态来判断眼睛区域的状态。Tian 等<sup>[25]</sup>采用人脸行为编码结构获得了三种眼睛状态:完全闭眼、半闭眼状态和正常状态,接着使用 Lucas-Kanade 跟踪方法来获得眼角位置和中心位置,并得到点周围部分的 Gabor 特征参数来判断眼睛状态,然后使用卷积神经网络对眼部特征检测。日本丰田汽车公司在 2008 年发布了一款疲劳驾驶检测预警系统<sup>[27]</sup>,该系统在驾驶员座位的正前方安装了相机,用于监控驾驶员的眼部,通过计算眼睛边缘的纵横比是否超出系统预定的阈值,若超过预定的阈值,则判定驾驶员处于疲劳状态,该系统就会发出预警提示,自发地将车辆减速来减少车祸的发生。ASCI(Advanced Safety Concepts)基于点头检测提出了一个头部姿态传感器系统 MINDS(Micro-Nod Detection System)<sup>[28]</sup>检测驾驶员是否处于疲劳状态。

Wu<sup>[26]</sup>等使用 Adaboost 和 Haar 特征获得面部区域,然后使用 SVM 分类器得到眼睛区域,接着获得眼睛区域的局部二值图像(LBP)作为特征参数,最后再输入一个 SVM 分类器判断眼睛特征状态。张伟等<sup>[29]</sup>获取驾驶员 11 个眼睛状态参数,构造了一种眼睛的特征空间,模仿人对环境的认知过程,在驾驶开始时先获得正常状态下驾驶员的特征空间,通过正常状态的特征空间训练得到正常驾驶的状态模型,在此基础上继续强化学习,运用贝叶斯网络来判断驾驶员是否处于疲劳状态。张万枝等<sup>[30]</sup>使用鼻子、嘴巴和眼睛构造了一个面部三角形特征,并通过提取面部特征参数对驾驶员进行头部状态估计,然后通过点头频率和点头时间来分析判断驾驶员是否处于疲劳状态。

这篇论文致力于研究基于计算机视觉的疲劳检测方法。该疲劳检测算法包含了 4 个任务:1)数据预处理;2)人脸识别及特征提取;3)并行检测;4)疲劳判定。所以介绍一下本文待使用算法的研究现状。

近期人脸检测的方法有很多。Viola 和 Jones<sup>[31]</sup>在 AdaBoost 算法的基础上,对 AdaBoost 使用 Haar-like 特征和积分图法训练出的强分类器进行级联,然后使用级联分类器进行人脸检测,该方法被广泛应用。Renliang 和 Weng 等人<sup>[32]</sup>利用 SIFT 关键点位置信息和纹理信息来对人脸进行识别。2014 年以来,由于硬件设备的高速发展,深度学习成为音视频领域中最热的方法,比如人体行为识别<sup>[33]</sup>,文本分类<sup>[34]</sup>,时序动作定位<sup>[35]</sup>等。在人脸检测方法中,神经网络也是效果显著。2015 年,Shroff F 等人提出 FaceNet<sup>[36]</sup>,成功将深度学习应用于人脸检测。2018 年, Jian Li 等人提出 DSFace 网络<sup>[37]</sup>,其继承 SSD 框架检测且效果最佳,但实时性较差。Zhang 等提出的一种多任务神经网络(MTCNN)

<sup>[38]</sup>可快速准确检测多个人脸。

下来是并行检测，包括眼睛嘴巴状态的检测和头部姿态检测。针对特征检测，YAO等<sup>[39]</sup>选择局部二值模式（LBP）来提取眼睛特征纹理并运用支撑向量机（SVM）来判断眼睛开闭状态。Dong等<sup>[40]</sup>比较了不同的眼睛状态估计方法，例如随机森林，随机森林和SVM。Jun W等<sup>[42]</sup>运用肤色建模的人脸定位，提出面部和面部特征点的位置以得到疲劳检测的分类。对于头部姿态估计，J.Ng<sup>[43]</sup>等人运用基于外观模型的方法，使用SVM方法来识别面部和检测头部状态。Qiao<sup>[44]</sup>等人在非线性回归的基础上，基于回归森林，选择深度数据建立面部方向的计算框架来进行头部状态检测，并且可以检测诸如头部大幅度的、快速变化的、暂时遮挡的数据。相比神经网络的方法效果更佳，Ruiz.N<sup>[45]</sup>等使用大量数据训练神经网络，分别预测得到人脸的欧拉角，使得头部姿态估计的准确率和实时性都得到了极大的提高。循环神经网络应用也较为广泛，主要集中在行为识别和自然语言处理<sup>[46]</sup>，致力于处理时序相关信息。本文引入循环神经网络作点头状态判断。

然后根据眼嘴状态和点头状态对疲劳的检测，大多数基于视觉的方法都是根据从眼睛区域提取的症状来推断疲劳和嗜睡状态<sup>[47]</sup>，其中Wierwille等<sup>[48]</sup>提出的PERCLOS方法应用最广泛。李娟等<sup>[49]</sup>使用PERCLOS的P80标准和眨眼频率参数为面部特征信息、车辆是否越车道线为车辆特征信息，运用SVM算法将数据融合，融合了基于面部特征信息和基于车辆特征信息的疲劳检测参数，构建了驾驶员疲劳检测模型。由清华大学和东南大学几位博士共同组建的南京研发中心联合南京远驱科技有限公司研发的疲劳预警系统gogo850平安行<sup>[50]</sup>，使用PERCLOS方法检测眼睛的闭合大小，用以检测驾驶员的疲劳程度，若驾驶员出现疲劳情况就对驾驶员预警提醒。系统是用红外相机获得人脸图像，红外相机可以预防光线对眼睛提取和识别的影响，且可以防止墨镜对检测效果的影响，提高检测的精确性。并且该系统附带额外的瞳孔识别算法，换言之，该系统可以对习惯于睁眼睡觉的人开展疲劳检测。该系统具有很好的鲁棒性，实用性很强。Wang H G等<sup>[51]</sup>提取嘴巴和眼睛特征后运用PERCLOS和张口频率（FOM）对疲劳进行判断。也有一些学者对头部姿态估计进行了研究<sup>[52],[53]</sup>，Yang X等<sup>[54]</sup>使用深度相机获得驾驶员的骨骼节点以判断头部姿态。

## 1.4 本文组织结构与主要内容

基于驾驶员面部行为特征的疲劳检测方法还存在很多问题，比如：1）采用计算机视觉的方式获得面部生理特征时，会受到复杂的光照明暗、驾驶员个体差异等影响，降低准确率。2）不同的车辆内摄像头的位置可能会存在差异，有些车辆中的识别摄像头无法获取正面清晰的驾驶员面部特征，从而影响疲劳检测的可靠性。3）基于面部行为特征的驾驶员疲劳检测方法主要通过获取眼睛状态、嘴巴状态和头部状态等单一特征进行识别，误判情况严重。4）大部分基于驾驶员面部生理特征驾驶员疲劳检测方法的识别和提取



均使用传统机器学习方法（如 SVM、随机森林等）来分类，选择相关特征需要大量的经验和一定的运气，使得分类的精度受限于特征模型的选取。针对以上问题，本文研究相关的解决方法，根据驾驶员面部生理特征的多特征融合来检测驾驶员疲劳状态，构建一种高准确率和高效率的疲劳检测方法。

本文的主要内容以及各章节的组织结构内容如下：

### 1、绪论（第 1 章）

主要介绍疲劳检测方法的研究背景与意义，同时对疲劳检测方法进行详细的介绍并分析常用疲劳检测方法的优缺点，然后对国内外疲劳检测方法的研究现状进行调查与研究，通过分析优缺点后确定本文的研究内容，最后对本文待使用技术的研究现状进行相关介绍。

### 2、深度学习相关理论知识（第 2 章）

简要介绍深度学习相关理论知识。主要介绍卷积神经网络及它的组件：卷积层、激活层、池化层、归一化层以及全连接层。然后介绍了循环神经网络相关概念，重点介绍了后文将要用到的长短期记忆网络（LSTM）。

### 3、驾驶员图像预处理及人脸识别算法（第 3 章）

主要介绍图像预处理算法、人脸识别算法和特征提取方法。为了提升人脸识别、特征提取及后续状态检测的准确率，需要对获取的相机图像进行图像预处理操作。为了减少噪声点对后续图像处理的影响，通过分析均值滤波、中值滤波、高斯滤波和动态自适应滤波方法后，选择中值滤波算法。为了使图像光照均衡，后续采用直方图均衡化算法对图像进行光照补偿。后面大致介绍了现有人脸识别算法，并将多任务卷积神经网络（MTCNN）算法与现有的一些算法进行比较，实验表明 MTCNN 算法对人脸识别的效果最佳，且耗时最短。最后介绍了驾驶员眼嘴两个面部特征的提取算法，为下一步面部特征的状态检测做好准备。

### 4、面部特征状态检测及头部姿态估计（第 4 章）

主要实现眼部嘴部的状态检测及驾驶员头部姿态估计。首先通过上述 MTCNN 算法获得了眼部和嘴部的特征，并将眼部和嘴部区域提取出来。在提取眼嘴区域后，可以将状态识别分为三个部分。（1）人眼状态检测。在获得人眼区域后，我们提出了使用基于深度学习的卷积神经网络对人眼状态进行检测，对比基于计算人眼横纵比的方法，本文所使用方法效果更佳。为了显示本文方法高精确度的优势，本文对比了 SVM、常用卷积神经网络。实验结果显示，本文所选方法可以精准的检测眼睛特征状态。（2）嘴部状态检测。在获得嘴部区域后，类似于人眼状态检测，我们使用基于深度学习的卷积神经网络对每一帧嘴部状态检测，可以准确获得每一帧图像里嘴部状态，将嘴部状态分为打哈欠、说话和正常等状态。（3）头部姿态估计。为了防止无法提取眼嘴区域，本文引入头部姿态估计来辅助判断面部特征信息。本文提出了使用 Hopenet 和 LSTM 算法对头部姿

态进行估计。为了验证该方法的准确性，我们在 NTHU-DDD 数据集上实验训练测试最佳 LSTM 参数，实验表明该方法可以很好的识别驾驶员头部姿态，准确判断驾驶员是否处于瞌睡点头区间。

### 5、疲劳状态检测实现（第 5 章）

本章主要是疲劳状态检测的实现。为了获得准确的疲劳检测模型，本文选择计算眼睛眨眼频率、闭眼状态占比和最长闭眼持续时间，嘴部计算打哈欠时间、打哈欠次数和点头状态占比，建立多特征融合加权的疲劳检测模型。对比单位时间内疲劳的累加值与对应的疲劳模型来检测驾驶员是否处于疲劳状态，若检测到疲劳，则对其进行提醒预警。最后为了验证本文算法的效果，在 NTHU-DDD 数据集和自采集疲劳检测数据集上进行实验验证。

### 6、总结与展望

最后对本文的研究方法及工作内容进行总结，概括本文的创新点和不足之处，并对将来可改进内容进行展望。

## 第 2 章 深度学习相关理论知识

深度学习是机器学习与统计学习的一个新的研究方向，可以看作为神经网络之一。其优势在于可以学习、归纳和总结，类似于人脑的学习过程。深度学习可以添加多个隐藏层，类比脑部神经元信息传递，经过逐级学习后形成深层次特征表示原始数据信息，在学习过程中优化神经元参数，便于后续分类及回归，可以减少特征提取及架构构造的不便。深度学习中最具代表的结构有卷积神经网络（CNN）以及循环神经网络（RNN），CNN 在图像分类及视频分析中具有广阔的前景，而 RNN 在时序信息处理和语音分类方面应用广泛。

### 2.1 卷积神经网络相关介绍

卷积神经网络（CNN）是以卷积运算为主的深度神经网络，它具有局部连接、参数共享、平移不变的性质使其在计算机视觉领域应用广泛。通常由输入、输出层以及隐藏层组成，而隐藏层又包括卷积层、激活层、池化层、归一化层以及全连接层等构成。

（1）卷积层：卷积层是构成 CNN 的核心组件，通常是一个矩阵，一般使用大小不同的卷积核，对前面的输入运行卷积操作获得数据特征。卷积层的操作包括两项功能，局部关联和滑动窗口。局部关联的意思是数据通过所有神经元时都要滤波，每个神经元对数据作卷积运算。滑动窗口是每一次卷积时，仅在滑动窗口内运行卷积操作，当卷积完成一次后，滑动到下一个窗口进行卷积运算。卷积层的种类相对较多，按卷积核的维度可划分为 1D 卷积、2D 卷积、3D 卷积。1D 卷积是对一维数据的处理。由于图像信号是 2D 信号并且 2D 卷积能够很好的提取边缘、颜色分布等空间特征，因此现代卷积神经网络的应用场景主要集中在图像领域，在卷积神经网络中 2D 卷积的应用也最为常见。3D 卷积在三维平面上移动，常用于处理三维图像，常见的应用领域有图像处理、视频分析等，例如 3D 医学图像中目标分割和视频中人的行为识别。

（2）池化层：池化层也叫下采样层，它实际上是一个统计函数，它通过减少特征图的空间大小来减少网络中的参数和计算量，达到简化基础计算和降低特征图维度的目的。

（3）激励层：激励层也被称为激活函数，它的常见作用是将线性映射转换为非线性映射，可以更好的解释。对于神经网络学习和理解输入与响应之间复杂的非线性的函数映射关系，激活函数非常重要。Relu 是最常用的激活函数之一，各种类型的深度卷积神经网络都经常使用它。

（4）归一化：归一化操作是对每个卷积后的数据进行归一化，可以使卷积神经网络的训练更快速高效，同时降低网络对不同初始化结果的差异性。规范的归一化是深度学习研究领域的热点，学术界和工业界已经提出了许多种类的归一化层，但是并非所有情

况都适用，例如对最大值或最小值点处理不好。

(5) 全连接层 (FC): FC 的特点是其节点与上一层的每个输出都相连，即一层中的每个神经元都与另一层中的每个神经元相连。网络架构者一般将从基础网络获得的图像特征传递到全连接层，用于驱使神经网络的做出判断，它在基于计算机视觉的检测和识别中的作用已被验证非常成功。相比于全连接层，卷积层的神经元参数都是共享的，并且其神经元只与部分输入数据相连。因为全连接层的参数量巨大，一些神经网络模型选择池化层替换全连接层，用于降低过拟合现象的产生和减少参数量。

## 2.2 循环神经网络相关介绍

循环神经网络 (RNN) 被广泛应用于深度学习中。通常使用时序数据输入 RNN，循环节点沿着时间顺序相连成为一个有向图，从而使它们能够显示顺序的行为动态。另一方面，由于循环神经网络允许前面的输出作为输入，并具有隐藏状态，也就是说循环神经网络可以拥有存储性能，所以在计算中需要考虑前置信息和参数的时域共享性质。上面提到的固有特性使其广泛用于自然语言处理和语言识别中。从原理上讲，循环神经网络还可以处理任意长度的输入，并且模型的大小不会随输入大小的增加而增加。然而循环神经网络也有很多问题，例如并行处理的速度缓慢。其较常见的问题之一是长期依赖性问题：循环神经网络在训练过程中较容易出现梯度消失或爆炸，从而难以建立长期依赖性。梯度爆炸可通过梯度裁剪轻松处理，但是梯度消失则难以控制。为了解决这个问题，研究学者提出了许多其他的循环神经网络，门控算法就是一种常用的算法。有选择地让信息通过的门通常由 Sigmoid 激活层和逐点乘法组成，特定的“门”通常具有特定的目标，并且通过组合不同的门控单元，可以提供循环神经网络控制信息积累的能力，以解决长期依赖的问题。最受欢迎的门控制算法是长短期记忆网络 (LSTM) 和门控循环单元网络 (GRU)。

长短期记忆网络是对循环神经网络基本结构的改进，它使用三种不同类型的门：输入门，遗忘门和输出门来控制内在状态（元点特征）和循环输出单元。具体地说，输入门可以通过当前时间步长的输入和前一时间步长的输出确定是否更新当前内部状态，即是否需要向记忆单元写入信息和写入什么信息。遗忘门的作用是通过上一个时间步长的内在状态影响当前时间步长的内部状态，即是否要删除历史记录。输出门根据当前时间步长的输入和内部状态确定输出。

由于不同类型的门有助于提高长短期记忆网络中的学习能力，因此可以在保持学习效果的同时，通过减少 LSTM 门的数量来简化单元结构。基于此设计概念，简化门控单元网络，通过舍弃输出门并保留遗忘门，将其分为两个门：更新门和重置门。详细地说，更新门确定是否需要更新隐藏状态，即其保存从前置的隐藏状态更新到当前隐藏状态的信息量；重置门确定是否忽略之前的内部状态。最后，使用两个门控单元可以对有用信

息保留很久，同时迅速丢弃无关信息。

## 第3章 图像预处理及人脸检测算法

为了提升人脸识别、特征提取及后续状态检测的准确率、减少后续相关图像处理的计算时间，需要对相机获取的图像进行图像预处理操作。为了准确截取驾驶员面部生理特征、缩小人眼定位及嘴部定位的区域、精确地判断驾驶员疲劳状态等级，人脸检测算法是必要的基础和核心。将现有的一些人脸检测算法进行比较，实验表明多任务卷积神经网络（MTCNN）对人脸识别的效果最佳，且耗时最短。为了进行面部特征的状态检测，必须准确提取驾驶员眼嘴两个面部特征，本文选择根据 MTCNN 检测的五个关键点提取相应的眼睛和嘴巴特征序列。

### 3.1 图像预处理

因为图像在采集和存储过程中会受到偶尔存在的各种噪声污染和不同程度的破坏，从而导致图像质量下降。在实际驾驶过程中，由于多种因素可能会造成采集图像的光照分布不均，对人脸检测、面部特征提取和疲劳检测造成较大影响。因此，在检测和识别视频图像和提取面部特征参数之前，需要对相机拍摄的图像视频序列进行预处理操作，以消除图像中的无用信息，提高有用信息的可检测性和保证原始信息的完整性，并能准确地检测和识别面部特征区域。本节介绍图像降噪和光照均衡技术，主要是为了提高人脸检测和定位的准确性。

#### 3.1.1 图像去噪

形成图像时，降低图像质量的主要因素是噪声，其产生过程是由于各种因素（例如光谱，光源，强度和相机特性（传感器响应，镜头））。噪声点隐藏了图像的重要细节，并在关键位置更改了图像像素的值，从而导致模糊和各种其他变形。我们必须去除图像中的噪音，而不丢失任何图像信息。有很多类型的噪点会破坏图像。在采集时，由于传感器嘈杂、扫描仪故障或由于数码相机故障、传输通道错误和存储介质损坏会引起噪声出现，这些噪声以不同的方式出现在图像上。

图像去噪已成为处理图像和从图像中去除不需要的噪点数据的关键步骤。图像去噪算法必须去除不必要的噪声元素，并保留图像的所有有用特征。图像去噪算法必须在两个参数之间进行权衡，即有效的噪声去除和图像细节的保留。图像在诸如天文学，医学成像和法医实验室图像等许多领域中发挥着非常重要的作用。本文用于驾驶员疲劳检测的图像也需要尽可能减少噪声，才能从图像检测中获得准确的结果。

图像去噪算法主要有两种<sup>[55]</sup>：空间域滤波去噪和频域滤波去噪。空间域滤波去噪方法有两种：1.非线性滤波：假定噪声位于高频区域，采用低通滤波器将图像与噪声分离，

空间滤波器可在很大程度上消除噪点，但会导致图像模糊，从而使图片的边缘不可见；

2.均值相关滤波：因其针对均方误差，均值滤波器是针对高斯噪声的最佳线性滤波器，若存在信号相似噪声，它们的性能较差，该方法也容易模糊锐利的边缘、线条和其他有价值的图像细节。在频域滤波去噪算法中，通过相应的变换将图像转移到频域中，然后在频域中对有用信号和噪声成分进行相关计算分析频响函数和截止频率，通过获得的频域滤波器和截止频率可以实现噪声的去除，最后将处理后的图片逆转回空间域。频域滤波去噪方法需要花费大量时间，并取决于截止频率和滤波器功能，可能会在处理后的图像中产生人工频率。考虑本文算法需要实时处理图像数据，这里使用速度较快的空间域滤波去噪方法，着重介绍均值滤波算法、中值滤波算法和自适应中值滤波方法，并将它们进行对比分析。

(1) 均值滤波：这是一个简单的滑动窗口空间滤波器。它将窗口的中心值替换为窗口中所有像素的平均值。卷积掩码提供像素及其相邻像素值的加权和。卷积掩码是方形的，并且按照移位乘和原理滤波，如下所示：

$$f(x, y) = \frac{1}{9} \times H * S(x, y) \quad (3.1)$$

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \quad (3.2)$$

其中， $S(x, y)$ 为点 $(x, y)$ 处  $3 \times 3$  滑动窗口内的像素点集， $f(x, y)$ 为点 $(x, y)$ 处的灰度值， $H$ (一般地， $h_1=h_2=\dots=h_9=1$ )为均值滤波的模板系数矩阵。

当噪声处于冲激状态时，这种均值滤波是有效的。均值滤波的作用类似于低通滤波器，并且不允许高频噪声通过。

(2) 中值滤波器：它是一个非线性滤波器。具有相同的滑动窗口概念，但中心像素值与相邻像素的中位数交换。所有像素均按数值排序，然后将中心值与窗口的中位数交换。中值滤波对于突变像素值具有非常高的抵抗力，因为异常值与像素矩阵的中位像素值不匹配。运算公式如下：

$$F(x, y) = \text{Med}(S(x, y)) \quad (3.3)$$

其中 $S(x, y)$ 为点 $(x, y)$ 处邻域内的像素点集， $\text{Med}()$ 表示中值函数，将以像素点 $(x, y)$ 为中心的邻域中的各个像素点的灰度值的大小排列，输出处于中间位置的像素灰度值。 $F(x, y)$ 表示 $(x, y)$ 处中值滤波后的新像素值。

(3) 自适应滤波器：均值滤波器与自适应滤波器之间的主要区别在于，在自适应滤波器中，加权矩阵在每次迭代后都会发生变化。它对于噪声可变的图像很有用，并且可以应用于未知图像类型，而无需知道其中存在的噪声类型，它在保留图像的细节信息的同时兼顾去噪。它是不变的低通滤波器和变化的高通滤波器的组合。加权总和可通过以

下方式获得：

在图像上选择大小为  $a * b$  的窗口  $W$ ， $u$  是窗口的均值，它是从窗口的每个元素中扣除的，结果矩阵为  $W_r$ 。

$$W_r = W - u \quad (3.4)$$

$$z = \sum_{(i, j) \in W} h(i, j) W_r \quad (3.5)$$

$h(i, j)$  为加权矩阵的  $H(i, j)$  元素。 $z$  的值和窗口平均值的和随窗口中心值变化。现在的像素值为：

$$z = \tilde{z} + \mu \quad (3.6)$$

在下一步中，将窗口更改为行主要顺序的一个像素，并如下修改权重矩阵：

$$e = W_r - \tilde{z} \quad (3.7)$$

$e$  是偏差。计算原始窗口的最大特征值。

$$e = W_r - \tilde{z} \quad (3.8)$$

新的加权矩阵为：

$$h_{k+1} = h_k + * e * W_r \quad (3.9)$$

在下一次迭代中使用。该过程一直进行到覆盖整个图像为止。

### 3.1.2 光照补偿处理

由于摄像机采集的图像的亮度很容易收到光照强度、光源位置等影响。尤其是在夜晚行车时，光照强度很低，收集到的驾驶员视频图像光照无法保持一致性。如果图像被直接训练，训练效果会很差。因此，为了保证视频图像中的人脸信息及头像信息清晰可见，并能准确提取面部特征，需要对图像进行光照补偿处理。比较常用的方法有以直方图均衡化方法为代表的灰度变换法，基于照明-反射的同态滤波法，Retinex 增强方法以及梯度域增强方法<sup>[56]</sup>，这里使用直方图均衡（HE）算法。直方图均衡化算法将调整每个像素的像素值，尽可能均衡每个灰度等级的像素点数量，改变图像灰度分布使其趋于平衡。这可以使图像的亮度在一定程度上得到补偿，同时提高正确分类图像的概率。使用以下公式对图像直方图均衡：

$$P_a(n) = \frac{1}{M} \sum_{u=1}^n h_a(u), \quad n = 1, 2, \dots, N \quad (3.10)$$



$$b(x, y) = N \times P_a[a(x, y)] \quad (3.11)$$

其中  $N$  表示图像灰度级,  $h_a(u)$  为输入图片  $a(x, y)$  的直方图,  $M$  表示像素点个数, 图片  $b(x, y)$  是前置数据直方图均衡化的结果。

### 3.1.3 实验及分析

此实验的主要目的是实现各种降噪算法, 并对其效率进行比较研究。为了能够给出良好的比较结果, 选择了高质量的图像并将噪声添加到图像中。然后, 将各种降噪算法应用于有噪声的图像, 并对三种滤波算法进行对比分析。对比图如图 3.1 所示。



图 3.1 滤波结果对比图

经过实验结果可以看出, 中值滤波和自适应滤波算法对椒盐噪声的处理效果比均值滤波的方法明显要好, 而均值滤波方法针对高斯噪声的效果更佳。

为了客观地显示每种算法的去噪性能, 计算去噪滤波算法对图像处理前后的计算时间  $T$ 、峰值信噪比  $PSNR$  及均方差  $MSE$ 。其中它们的计算公式如下所示:

$$MSE = \frac{1}{M * N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [f(x, y) - f'(x, y)]^2 \quad (3.12)$$

$$PSNR = 20 * \log_{10} \left( \frac{TOP}{\sqrt{MSE}} \right) \quad (3.13)$$

其中  $TOP$  表示图像的最大像素值, 表 3.1 和 3.2 显示实验结果。

表 3.1 各算法对添加高斯噪声图片的性能

算法	MSE	PSNR	T(s)
均值滤波	0.05721	23.5153	0.0251
中值滤波	0.05711	22.7210	0.0139
自适应中值滤波	0.07643	22.1808	0.6991

表 3.2 各算法对添加椒盐噪声图片的性能

算法	MSE	PSNR	T(s)
均值滤波	0.01829	17.4898	0.0244
中值滤波	0.01698	19.0248	0.0141
自适应中值滤波	0.00881	20.0035	0.7543

通过对实验结果分析，三种方法均能达到一定量的图像去噪性能，但是各有其优缺点。在使用中值滤波算法的时候，其耗费时间最短，但是对高斯噪声的去除能力稍逊一筹。均值滤波算法各方面比较均衡，对椒盐噪声的去除较差。自适应滤波算法性能最佳，但是耗时太长。综合考虑，本文待检测视频序列工程量较大，且高斯噪声出现概率较低，本文使用中值滤波进行图像去噪来提升图像质量。



图 3.2 有无光照补偿处理的差别，左图为原始图像，右图为处理后的图像

根据上述基于直方图均衡化的光照补偿算法进行实验，实验结果如图 3.2 所示，图 3.2 表明了图像没有光照均衡处理和有光照均衡处理之间的差别。根据实验结果可以看出，当光照强度较低时，经过光照均衡处理后，图像的亮度和观看性会得到了显著的改进。

## 3.2 基于 MTCNN 的人脸检测算法

### 3.2.1 人脸检测算法介绍

在过去的十年中，图像处理领域取得了指数级的发展，新一代计算机正变得越来越智能化，可以更快地处理 TB 级的数据。伴随着计算机硬件和图像处理的快速发展，人脸检测技术也成为了热门的研究方向。近年来，人脸检测成为了安全性检测的基础和核心，人脸检测也成为了比如生物识别系统，智能手机的前置摄像头，人体存在检测等的

关键步骤。人脸检测算法可大致分为四类：基于特征的检测方法、基于先验知识的检测方法、基于统计学习及机器学习的检测方法和基于模板匹配的检测方法<sup>[57]</sup>。

1) 基于特征：基于特征的方法利用面部的独特元素。眼睛，鼻子和嘴唇等元素可用于检测脸部。这些元素的结构和大小与非面部对象不同。因此，基于特征的方法是可靠的。同样，这些元素总是在对象的脸上，它们的位置彼此相关。如今，由于人体皮肤比非面部对象具有独特的纹理和颜色，因此皮肤被用于检测面部。此外，边缘信息在突出显示所需区域中也起着重要作用。基于特征的方法需要在图像上定位人脸。该方法需要找到皮肤区域并通过在框架中搜索至少一只眼睛进行验证。下一步是识别面部特征（基准点）。HSV 色彩空间的 H 平面用于从眼睛检测区域检测眼睛瞳孔。通过假设眼睛的预估位置在脸部的上部，可以最大程度地减少眼睛的寻找区域，定位包含眼睛的矩形区域。色调图像是阈值，并且在阈值图像中检测到轮廓。轮廓的质心被检测为眼睛瞳孔。接下来检测鼻子，在知道了眼睛的中心和鼻子的位置之后，使用一种基于面部几何结构的位置估计方法。基于眼睛和鼻尖之间保持的相对距离，可以检测到面部其他元素的相对位置。但是基于特征的检测方法会由于异常（照明，噪音等），很难定位面部特征，也会由于背景复杂，难以识别特征。

2) 基于先验知识的检测方法：基于先验知识的方法也称为自上而下的方法。此方法使用了脸部的几何特性。人脸有两只彼此对称的眼睛和一只鼻子在脸的中心。该方法分为三步：a) 脸部的中间部分具有均匀的强度值，搜索候选对象。b) 脸部中间部分和上部的平均强度值之间的差异很大，使用局部直方图均衡，然后进行边缘均衡，然后进行边缘检测。c) 经常出现两只彼此对称的眼睛，即鼻子和嘴巴，搜索眼睛和嘴巴特征以进行验证。该方法的一个有趣的特点是，使用了从粗糙到精细或者说是注意力集中策略来减少所需的计算量。

3) 基于模板匹配的检测方法：它通常用于可能会出现人脸的系统中。模板是具有统一大小和形状的预定义结构，仅通过将模板与对象进行比较即可轻松检测所需对象。在人脸检测的情况下，模板通过匹配找到输入图像或视频与人脸图像或特征之间的关系。图 3.3 显示了用于面部检测的模板。



图 3.3 垂直和旋转形式的人脸形状模板

模板匹配方法是可变的，并且基于面部轮廓。与使用神经网络的基于外观的方法不同，模板是手动编码的（未进行学习的），并使用相关性来定位人脸。但是该方法对于正面视图，面部必须没有遮挡；面部必须与模板具有相同的大小，否则匹配与否还需要改

变大小,改变比例和旋转;为了覆盖更多的面部视图,需要使用更多的模板匹配,因此需要更多时间来检测面部。

4) 基于统计学习及机器学习的检测方法:它依赖统计分析和机器学习中的技术来找到面部和非面部的相关特征。采用分布式模型或判别函数形式学习到的特征来进行面部检测。此方法降低图像的维数可以提高计算效率,并使用分类和特征来搜索候选窗口。通常使用统计变换和 Haar-like 特征,统计方法需要在大量数据中学习,但相对的,它可以用于检测处于不同姿势和方向的面部。基于 Haar-like 特征的 AdaBoost 人脸识别分类器有效的提取图像的纹理特征,从而提高了速度和准确性。但其在现实应用中因环境易变而表现欠佳,而且不能定位眼睛和嘴巴特征。其他还有多层感知器神经网络、支持向量机、隐马尔可夫模型和主成分分析等人脸检测算法也是相似的原理。

在人脸检测方法中,基于统计分析及机器学习检测方法的神经网络效果显著。Zhang 等提出的一种多任务神经网络(MTCNN)<sup>[38]</sup>可快速准确检测多个人脸,它也是一个高效和精准的人脸识别模型。下面主要对 MTCNN 算法进行详细介绍。

### 3.2.2 基于 MTCNN 的人脸检测算法

#### (1) 整体结构

采用基于级联回归的 MTCNN 网络可以快速识别对齐面部区域。通过基于 MTCNN 的人脸识别及定位,可以得到面部的回归区域和五个关键点,包括左眼中心、右眼中心、左唇端、右唇端和鼻尖的位置。这里选择 MTCNN 实现人脸识别和特征点定位。MTCNN 由 P-Net、R-Net 和 O-Net 三个小网络构成。当输入一帧图像时,对其进行不同尺度的变换,构建图像金字塔,以适应不同大小的人脸。第一步,使用全卷积网络快速形成候选框;第二步,使用复杂的神经网络拒绝错误的区域,去除重复的区域;第三步,通过最后一个神经网络输出最终的人脸框,得到五个面部特征点坐标。

建议网络(P-Net):该网络是一个全卷积网络,用来生成候选框和边框回归向量,使用边界回归的方法对候选框校正,然后使用非极大值抑制合并重叠的候选窗口。

优化网络(R-Net):该网络继续使用边界回归对候选框校正,继续采用非极大值抑制删除重叠的候选区域。然而因为 R-Net 的结构与 P-Net 的结构不同,其在输出前使用全连接层,可以去除大量的假阳性区域,将获得极佳的非极大值抑制效果。

输出网络(O-Net):O-Net 结构和 R-Net 结构相比增加了一个卷积层,会使处理的过程和结果更精确。其效果和 R-Net 结构的效果类似,不同的是该网络对人脸候选框有更严格的筛选,同时输出最终的 5 个坐标,分别为左眼中心、右眼中心、左唇端、右唇端和鼻尖。整体算法示意图如图 3.4 所示。

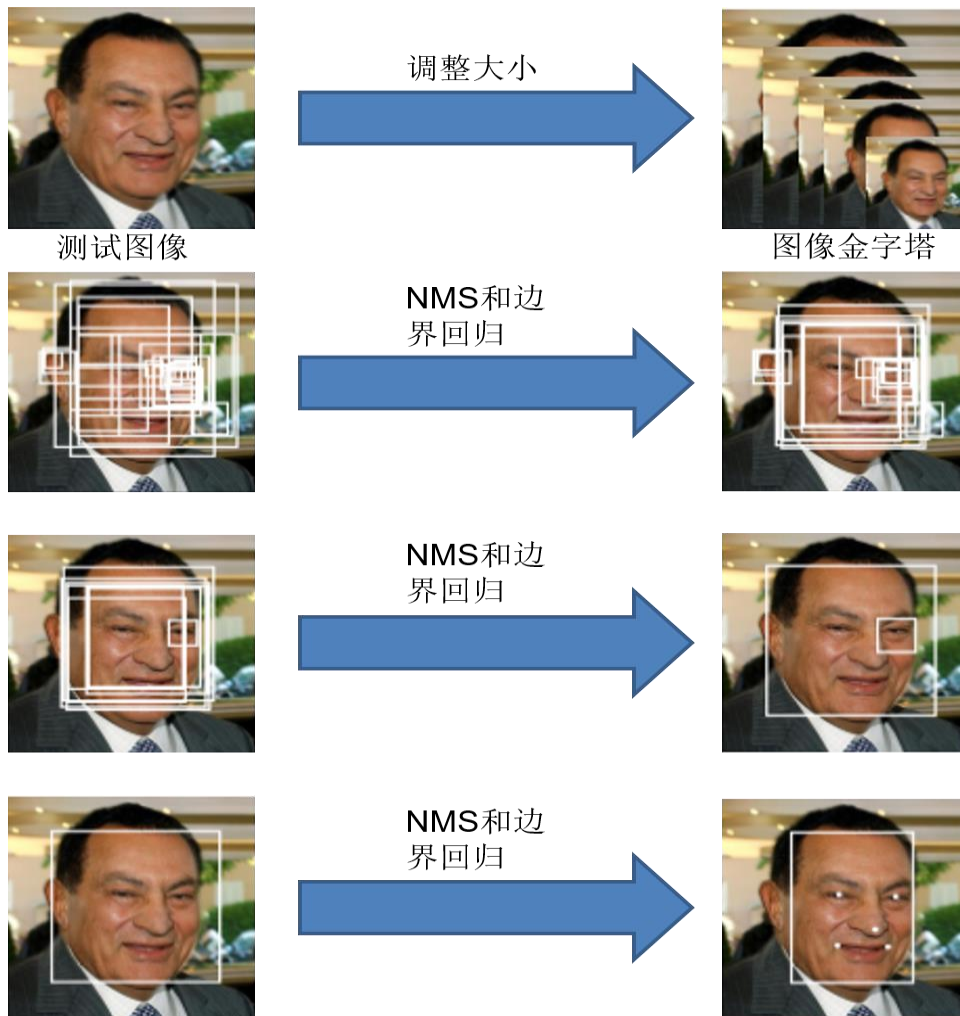


图 3.4 基于 MTCNN 的人脸检测算法示意图

## (2) MTCNN 的网络结构

MTCNN 的网络结构如图 3.5 所示。P-Net 输入为  $12 \times 12$  的 RGB 三通道图像，第一层的卷积核是一个  $3 \times 3$ ，同样的第一层池化层也是  $3 \times 3$  的最大值池化层，后面紧跟着 16 个  $3 \times 3$  的卷积核，再后面接了 32 个  $3 \times 3$  的卷积核，接着连接 32 个  $1 \times 1$  的卷积核，最后用全连接层进行面部候选框定位，边界回归和面部点定位。

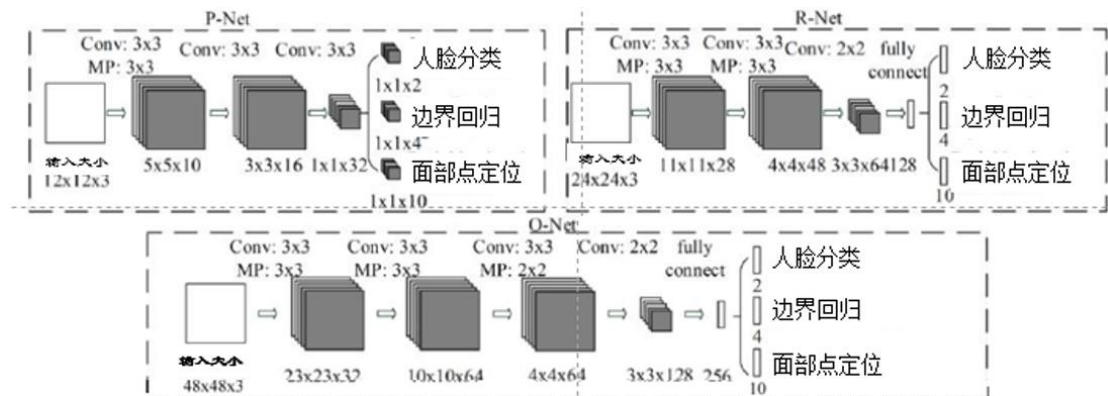


图 3.5 P-Net, R-Net, O-Net 的结构示意图

R-Net 的输入为  $24 \times 24$  的 RGB 三通道图像。第一层先用 28 个  $3 \times 3$  卷积核对图像进行卷积，同样的第一层池化层也是  $3 \times 3$  的最大值池化层，后面用 48 个  $3 \times 3$  的卷积核进



行处理，再后面接了 64 个  $2 \times 2$  的卷积核，最后用全连接层将第一层和第三层的所有神经元连接，得到一维向量，通过一个向量来进行面部候选框定位，边界回归和面部点定位。

O-Net 的输入为  $48 \times 48$  的 RGB 三通道图像，第一层先用 32 个  $3 \times 3$  卷积核对图像进行卷积，同样的第一层池化层也是  $3 \times 3$  的最大值池化层，后面用 64 个  $3 \times 3$  的卷积核进行处理，接着  $3 \times 3$  的最大池化层，再后面接了 64 个  $3 \times 3$  的卷积核，这里最大池化层为  $2 \times 2$  的最大值池化，第四层采用 128 个  $2 \times 2$  的卷积核，最后用全连接层将第一个全连接层和第四层卷积后的结果连接，得到 256 维的向量，通过该向量来进行面部候选框定位，边界回归和面部点定位。

上述网络中，所有卷积的步长是 1，池化的步长是 2，所有卷积和全连接后面会跟着一个激活函数，这里使用 PRelu 作为激活函数。

### (3) 训练及损失函数

训练时，本文使用三个任务来训练 CNN 分类器：第一个任务进行面部/非面部分类，第二个任务进行边界框回归的获得，第三个任务输出面部关键点坐标的向量。

整体目标是多源训练，因此整体的目标函数为：

$$\min \sum_{I=1}^N \sum_{J \in \{det, box, landmark\}} \alpha_j \beta_i^j L_i^j \quad (3.14)$$

其中  $N$  表示训练样本数量， $\alpha_j$  表示不同任务的重要性， $\beta_i^j \in \{0,1\}$  是样本类型指标， $L_i^j$  是不同任务的损失函数， $j$  表示不同的任务。在 P-Net 和 R-Net 中使 ( $\alpha_{det}=1$ ,  $\alpha_{box}=0.5$ ,  $\alpha_{landmark}=0.5$ )，而在 O-Net 中使 ( $\alpha_{det}=1$ ,  $\alpha_{box}=0.5$ ,  $\alpha_{landmark}=1$ ) 以获得更准确的面部标识定位。下面介绍不同任务的损失函数。

分类任务损失函数采用交叉损失熵函数，如下式(3.15)所示：

$$L_i^{det} = -(y_i^{det} \log(p_i) + (1 - y_i^{det})(1 - \log(p_i))) \quad (3.15)$$

其中  $p_i$  表示网络将样本判定为人脸的概率，每一个样本表示一张脸， $y_i^{det}$  表示真实标签。

边界框回归任务损失函数如下式(3.16)所示：

$$L_i^{box} = \|\hat{y}_i^{box} - y_i^{box}\|_2^2 \quad (3.16)$$

其中  $\hat{y}_i^{box}$  表示网络预测回归边界框的位置坐标， $y_i^{box}$  表示真实标签中边界框的位置坐标。

面部关键点坐标回归任务损失函数如下式(3.17)所示：

$$L_i^{landmark} = \|\hat{y}_i^{landmark} - y_i^{landmark}\|_2^2 \quad (3.17)$$

其中  $\hat{y}_i^{landmark}$  表示网络预测回归边界框的位置坐标， $y_i^{landmark}$  表示真实标签中边界框的位置坐标。

### 3.2.3 实验结果及分析

数据集：选择 MTCNN 网络模型识别人脸并检测眼睛嘴巴关键点，此处使用了三个

不同的数据集：WIDER FACE 数据集<sup>[58]</sup>由 32,33 个图像中的 393,703 个标记的面部边界框组成，AFLW 数据集<sup>[59]</sup>包含 24,386 个面部的面部标志注释，FDDB 数据集<sup>[60]</sup>包含一组 2,845 个图像中 5,171 个面的注释，用这些数据集训练测试 MTCNN 网络，总训练数据由 3：1：1：2（负/正/部分脸/标识脸）四个部分数据组成。

实验运行硬件条件：Intel i7-7700k 2.60GHz 的处理器、两块 NVIDIA TITAN X 12G 显卡、16G 内存；实验运行的系统环境为：Ubuntu16.04、PyCharm、PyTorch。后面的实验均在此环境和硬件条件下运行。

根据前面的介绍可知，MTCNN 网络模型的主要任务是人脸检测和面部关键点定位，下图 3.6 为不同状态和方向的人脸检测结果。表 3.3 为 MTCNN 在不同数据集下的准确率结果，并与 Haar-like 和 DSface 算法作比较。

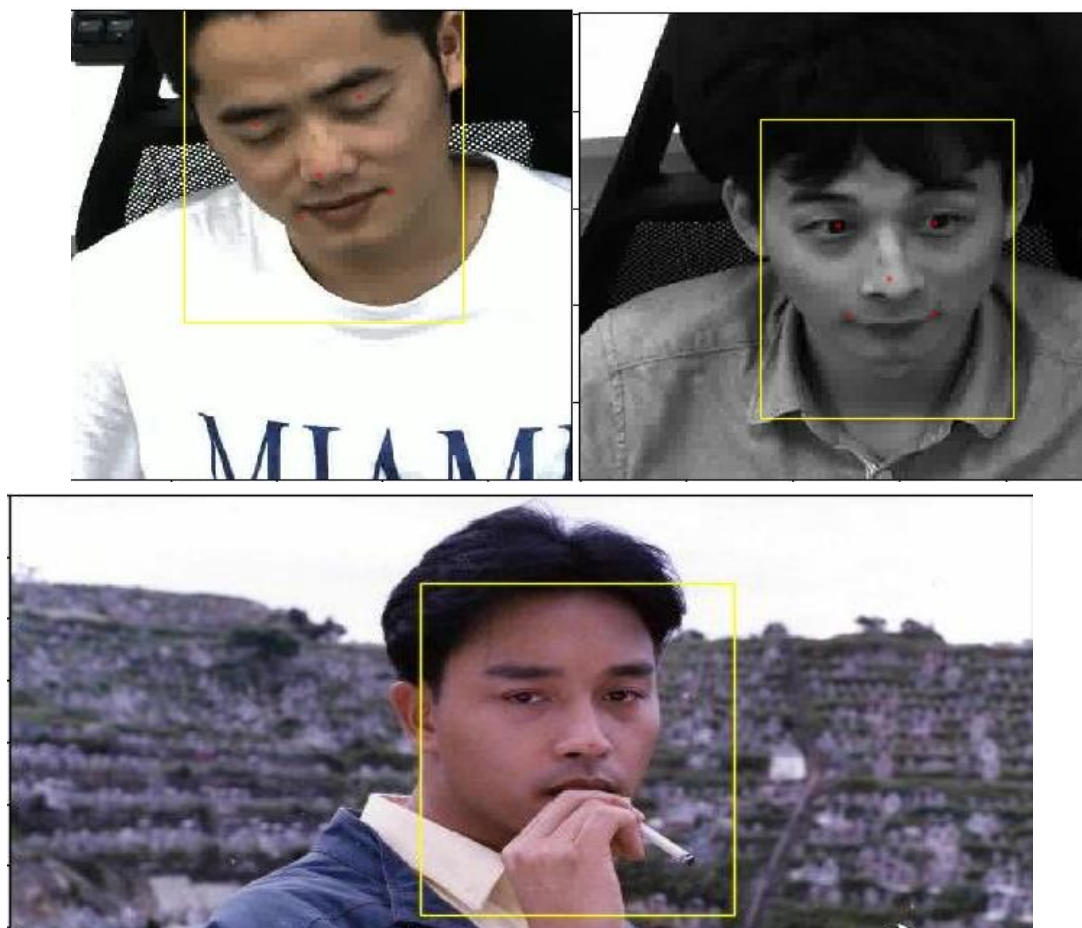


图 3.6 基于 MTCNN 人脸检测算法的检测结果

表 3.3 MTCNN 算法在各数据集上的准确率及效率对比

人脸识别算法	FDDB	WIDER FACE(easy)	速度(GPU:Nvidia Titan X)
Haar-like[31]	85.7	89.1	/
DSFace[37]	95.4	95.5	99fps
MTCNN	95.7	96.6	22fps

根据图表分析可知,虽然基于 Haar-like 的算法应用最广,但因其人脸检测在复杂场景下的准确率较低而失去竞争力;DSFace 的人脸检测算法准确率最高,但其耗时较长,无法实时处理图像数据,可能出现延迟等情况会导致疲劳检测系统出现各种问题;基于 MTCNN 的人脸检测算法不仅可以识别人脸,而且可以获得面部关键点位置坐标,其准确率和计算效率都很高,因此本本选择 MTCNN 算法进行人脸检测。

### 3.3 眼嘴特征提取

相比于人脸检测算法,想要对面部特征检测就显得更加困难,虽然人眼和嘴巴也有轮廓,但是容易受到个人差异的影响。目前常用的面部特征检测算法比较多,如: Hough 变换检测、模板匹配、基于统计特征、灰度投影算法等。

基于 Hough 变换检测的算法:从图像到参数,以参数方程的状态表示人眼的边界区域,常见的有椭圆和圆形两种形状特征。

模板匹配的算法:类似于人脸识别算法,要建立双眼灰度匹配模板,在获得的面部图形中遍历搜索,与已搭建模板匹配,匹配度最好的部分为面部特征的区域。

基于统计特征的方法:它依赖统计分析和机器学习中的技术来找到面部的相关特征,通过大量数据训练分类器获得机器学习的特征参数,得到训练后的分类器来判断是否为面部特征。

灰度投影算法:根据面部特征区域会比周围皮肤区域灰度值较低和变化明显的先验知识,通过将图像的积分投影曲线的极点值计算出来确定面部特征区域。

以上面部特征检测方法都是常见算法,但是都有一定的局限性,无法高效简洁地定位面部特征区域,不适合本文疲劳检测系统。

前面使用 MTCNN 网络后,已经获得面部 5 个关键点的坐标值:左眼、右眼、鼻子、左唇端和右唇端。但只确定出关键点坐标是无法确定正确的面部区域的,需要确定眼部和嘴部区域。我们需要通过这些关键点的坐标值,运用相应的提取算法获得脸部嘴巴眼睛区域,提取结果如图 3.7 所示。



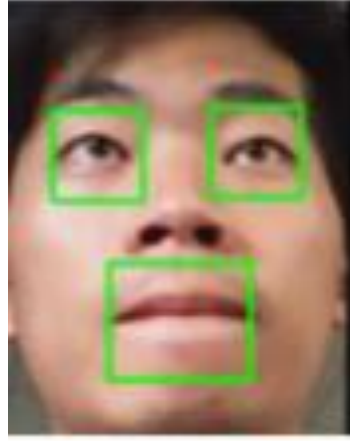


图 3.7 提取面部特征结果

人面部特征区域的分布呈现一定的规则，经过大量的对比发现，人脸符合“三庭五眼”规则，根据该规则，本文提出一个简单快速的提取方法，提取算法的规则如下：

$$\left\{ \begin{array}{l} top_{eye} = y_e - 2|y_e - y_n|/3 \\ height_{eye} = 4|y_e - y_n|/3 \\ left_{eye} = x_e - |x_e - x_n|/2 \\ width_{eye} = |x_e - x_n| \\ left_{mouth} = x_{ml} \\ right_{mouth} = x_{mr} \\ top_{mouth} = (y_{ml} + y_{mr})/2 - |(y_{ml} + y_{mr})/2 - y_n|/2 \\ height_{mouth} = |(y_{ml} + y_{mr})/2 - y_n| \end{array} \right. \quad (3.18)$$

其中， $y_i$ 表示区域的纵坐标， $x_i$ 表示区域的横坐标， $i=e, n, ml, mr$ 分别表示眼睛，鼻子，左唇端和右唇端， $height_j$ 表示区域的高度， $width_j$ 表示区域的长度， $top_i$ 表示眼睛区域的顶点坐标， $left_j$ 表示区域的左起点坐标， $right_j$ 表示区域的右起点坐标。

将 MTCNN 算法与提取算法级联就得到一个完整的面部特征提取算法，该级联结构的检测结果如下图 3.8 所示。



图 3.8 MTCNN 网络人脸识别及提取眼嘴区域示意图

### 3.4 本章小结

本章主要对图像预处理过程、人脸检测算法和面部特征提取算法进行了研究。为保证图像高质量和强可检测性，本文选择中值滤波方法对图像进行滤波去噪，不仅高效地去除了噪声而且保证了图像的完整性，接着使用直方图均衡化来对图像光照补偿，解决了低亮度条件下图像模糊不清、光强失衡等问题。在图像预处理后，使用了基于 MTCNN 的人脸检测算法，快速准确地获得面部区域，并同时获得了面部五个关键点的位置坐标。最后使用获得的五个关键点的位置坐标提取了面部特征区域（眼睛和嘴巴），将基于 MTCNN 的人脸检测算法和提取算法级联获得了一个高效简洁的面部特征提取算法。

## 第 4 章 面部特征状态检测及头部姿态估计

为了构建完整的疲劳检测系统，需要对上章提取的面部特征进行处理，本章将其分为三个子任务：头部状态估计，眼睛状态判断和嘴巴状态判断，对于眼睛和嘴巴，使用本文提出的小型分类卷积神经网络（SC-Net）模型。对于头部状态估计使用本文提出的 HP-LSTM 网络模型。

### 4.1 眼睛状态检测

#### 4.1.1 眼睛状态检测常用方法

人在疲劳状态下，闭眼时间及眨眼频率均会异常于正常状态，所以检测眼睛状态是驾驶员疲劳判定的重要特征和依据。所以对于整个系统的准确性和鲁棒性，需要精确检测驾驶员的眼睛状态。目前对于眼睛状态判定的方法较多，比如投影法、黑色素累计差值法、基于特征点的横纵比的方法和基于统计学习和机器学习的方法等。

投影法：根据眼睛在水平方向上的投影获取眼睛的开闭大小，但是该方法对于光照有极高的要求，如果驾驶员与传感器或相机的距离发生改变，则预先设定的阈值就会不实用，所以该方法无法运用在本文及实际环境中。

黑色素累计差值法：眼睛图像二值化后，经过形态学操作和中值滤波后，可以呈现眼睛轮廓及细节，根据睁闭眼二值图像黑色素累计值的差异，可以识别闭眼状态。但是该方法会随着人脸的转动及光照强度的改变而无法精确判定，所以该方法只能用于理想情况。

基于特征点的横纵比的方法：在得到眼睛周围 6 个关键点后，可以计算眼睛的横纵比。因个体之间的差异较小，闭眼时眼睛横纵比将远小于睁眼时的眼睛横纵比，运用这种差异可以获得精确的眼睛状态。本文在初始时采用的就是此方法，但是根据实验验证得知此方法实用性较差：（1）基于特征点的横纵比的方法需要获得人眼的周围 68 个关键点，此步骤需要消耗大量时间，大概 500ms/图，不符合实时性要求（2）此方法需要保证驾驶员时刻面朝正前方且不能随意摆动头部。所以该方法无法应用到本文所模拟的实际场景中。

#### 4.1.2 基于卷积神经网络的眼睛状态识别

本文将采用高准确率及高效率的基于统计学习及机器学习的方法来检测眼睛状态。人眼状态在闭合和张开之间有无数状态，但实际应用中，采集的图像是有限的，分类也应处于有限状态，所以将眼睛分为“闭合”和“睁开”两种状态，利用一定量的数据及深度学习模型训练一个状态检测分类器判断眼睛状态。本文提出小型分类卷积神经网络（SC-Net），模型其结构如图 4.1 所示。

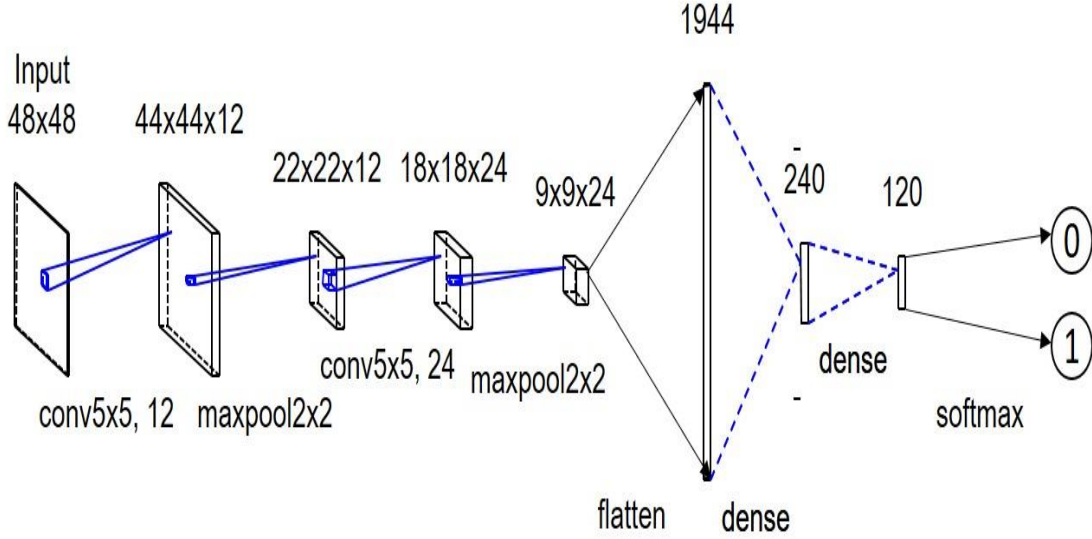


图 4.1 SC-Net 网络结构

SC-Net 直接以眼睛或嘴巴的二维图像作为输入,自动学习图像特征及数据内部的隐含关系,具有位移、缩放和扭曲不变性的优点。与 VGG-Net 的结构相似,其结构包括卷积层、池化层、全连接层及 Softmax 分类。

基于 SC-Net 的眼睛状态训练与嘴巴状态相同。所以,在这里我们以眼睛状态训练为例。在 SC-Net 里,输入图像尺度不确定,并且将其重整为 48\*48,在通过 12 个 5\*5 的卷积核和 Relu 层和最大值池化层之后,输出大小为 22\*22\*12。之后,继续通过 5\*5 的卷积核(24 个),此层的输出大小为 18\*18\*24。然后,在通过 Relu 层和最大池化之后,输出大小变为 9\*9\*24。之后,将特征图转换为一维向量并馈送到完全连接层。在全连接层中,输入层的长度为 1944,隐藏层为 3 层。最后,通过全连接和 Softmax 层,输出分为两类: 0 和 1 分别表示睁开眼睛和闭合眼睛。

这里使用交叉损失作为本文的损失函数:

$$H_{y^t}(y) = - \sum_i y'_i \log(y_i) \quad (4.1)$$

其中  $y_i$  是预测结果,  $y'_i$  是标识,标识是正确的标签类记为 1,其他情况记为 0,因为只有两个结果,所以  $i=0$  或 1。

这里选择随机梯度下降算法 (SGD) 作为本文的优化方法,更新函数如下:

$$\theta_j = \theta_j + \alpha(y - h_\theta(x))x_j \text{ (for every } j) \quad (4.2)$$

这里  $j$  表示神经网络第  $j$  层,  $\theta_j$  表示第  $j$  层的参数值,  $\alpha$  代表学习率,  $y$  表示真实值,  $h_\theta(x)$  表示预测值,  $x_j$  表示反向传播到第  $j$  层的值。

#### 4.1.3 眼部数据集

根据前面的眼部特征提取,可以将眼睛区域分割出来,从而获得眼部特征区域。这里三个数据集来训练测试 SC-Net。ZJU 数据集<sup>[61]</sup>、CEW 数据集<sup>[62]</sup>和作者自收集模

拟数据集。其中一些眼睛和嘴巴图像是通过从开放数据库（ZJU Eyeblink Database）中的 80 个视频剪辑中提取眼睛特征而获得的图像组成。原图像的分辨率为  $320 \times 240$  像素。这里我们从视频中提取 13042 张眼睛图片，其中 10804 张睁眼图片，2238 张闭眼图片。CEW 数据集是作者从互联网上收集了 2384 张眼的图像和从 Labeled Faces in the Wild 收集了 2462 张图像组成的（见图 4.2）。作者自收集模拟数据集包括来自 18 个志愿者的 30877 眼样本和 11357 嘴巴样本（见图 4.3）。这些眼睛样本分为带眼镜和不带眼镜的情况。这里将上述所有数据集中的图片缩放到  $48 \times 48$ 。

众所周知，神经网络需要大量数据才能对未看到的数据进行准确分类，并避免深度神经网络的过拟合。构建的数据集仅包含数万张图片，由于深度网络包含大量参数，这可能导致 SC-Net 容易过拟合。我们的实验中采用了数据增强技术。将每个图片在  $-15 \sim 15$  度范围内进行随机角度旋转；将图片在 0.8-1.2 的尺寸内随机缩放后再调整到  $48 \times 48$ ；考虑到因为肤色差异可能会影响到测试效果，对训练图片的色调进行小幅度微调（虽然这对数据集的测试没有多大帮助，对实用性和鲁棒性有了很大的提升）。这些图片用于微调 SC-Net，在测试时我们仅仅使用原始图像。

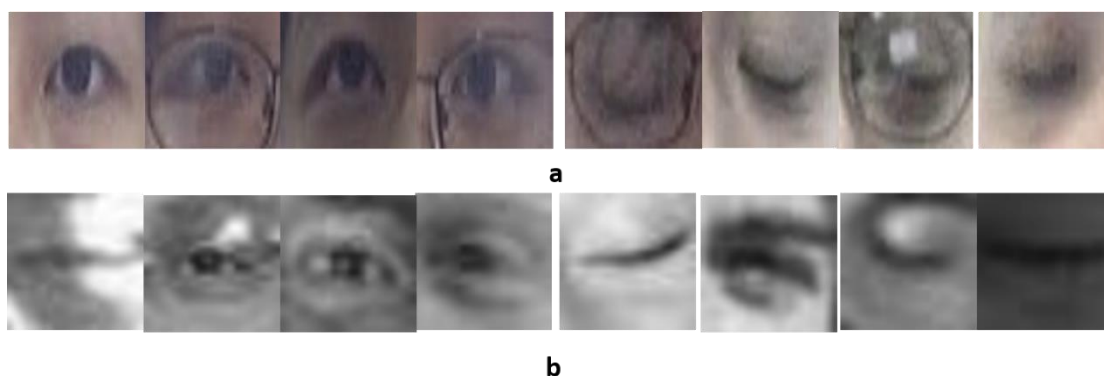


图 4.2 CEW 数据和 ZJU 数据的示例，a 是 ZJU 数据集，b 是 CEW 数据集



图 4.3 部分自收集眼睛数据示例

#### 4.1.4 实验与结果及分析

每次在训练集中随机选取 10% 的图像训练 SC-Net 模型，从实验分析可以得出，当训练次数越高时，训练精度会越高，但其测试集精度会呈现一种先升高后降低的趋势，由欠拟合到收敛状态再到过拟合状态，因此需要取得中间合适的训练次数，这里选择训练 3000 次使效果最佳。训练过程如下图 4.4 所示，可以看出，训练 3000 次后模型得到了收敛。

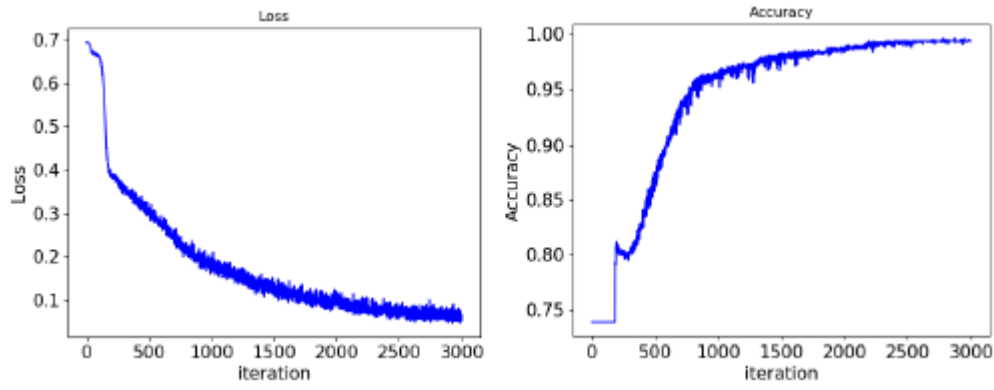


图 4.4 用 SC-Net 检测眼睛状态的训练过程

为了证明我们模型的有效性，我们首先使用自收集模拟数据集训练和测试。自收集模拟数据集中 70% 用于训练，30% 用于测试。为了验证 SC-Net 在眼睛检测方面的效果，我们将其与 VGG-16、ResNet18 和 Alexnet 进行了比较。此外，我们测试了图像预处理的效果（见表 4.1）。从实验结果可以看出，SC-Net 具有较高的检测精度，不进行图片预处理（IP）时，精度会略微下降。因此，检测前的图像预处理是非常必要的。

同时，为了证明本 SC-Net 的通用性，本文对 ZJU 数据集和 CEW 数据集进行了训练和测试（参见表 4.2）。表 4.2 给出了 SC-Net 和几个分类器在 ZJU 数据集和 CEW 数据集上的比较性能。结果表明，所有现有方法在 CEW 和 ZJU 数据集上准确率都超过了 90%；由于 CEW 数据本身比较模糊，没有任何方法可以超过 96%，相比之下，我们的 SC-Net 方法在 ZJU 数据集和 CEW 数据集上均表现优良。

表 4.1 在自收集数据集上使用不同网络进行眼睛状态检测的性能比较

网络	数据集	准确率,%
VGG-Net with IP	自收集数据集	83.35
AlexNet with IP		98.18
ResNet18 with IP		98.10
SC-Net without IP		97.04
<b>SC-Net with IP</b>		<b>98.64</b>

表 4.2 CEW 数据集和 ZJU 数据集上各种分类器的比较

方法	CEW		ZJU	
	ACC,%	AUC,%	ACC,%	AUC,%
LBP + SVM [26]	-	-	90.37	-
HOG + random forest [40]	94.57	98.17	94.70	98.37
HOG + random ferns [40]	94.15	98.78	94.76	98.89
MSP-Net[51]	95.36	99.05	96.89	98.85
<b>SC-Net(ours)</b>	<b>95.45</b>	<b>99.15</b>	<b>98.18</b>	<b>98.79</b>

## 4.2 嘴巴状态

### 4.2.1 嘴巴状态检测

嘴巴状态在闭合和张开之间有无数状态，主要处于正常情况下的闭合状态、说话和唱歌时的微张开状态和处于疲劳情况的大幅度张开状态（打哈欠状态），但实际应用中，采集的图像是有限的，分类也应处于有限状态，所以将嘴巴分为“闭合”和“张开”两种状态。类似于眼睛状态检测，利用一定量的数据及深度学习模型训练一个状态检测分类器（SC-Net）判断嘴巴状态。可以绘制出驾驶员在打哈欠时的过程状态图如下图 4.5 所示。

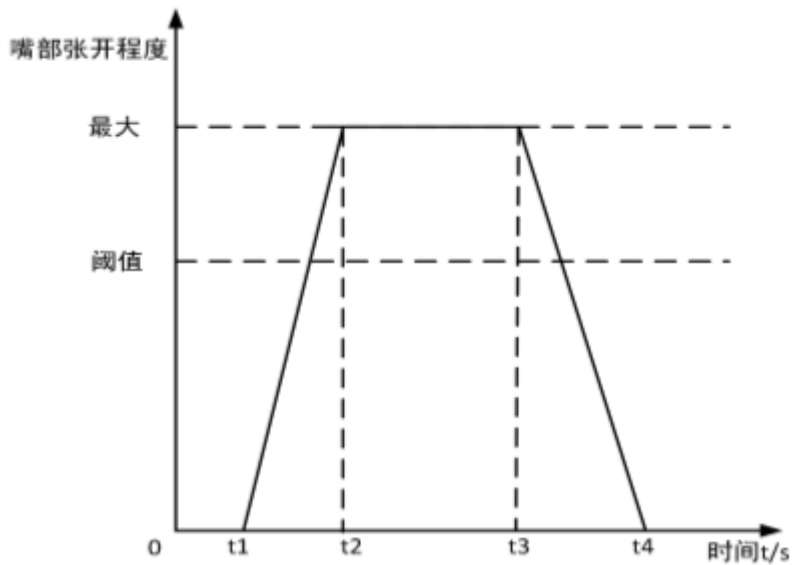


图 4.5 打哈欠嘴巴过程状态图

打哈欠是异于日常说话的，最明显的区别就是说话的时候嘴巴状态是时张时闭的，并且张嘴时间也远小于打哈欠。根据持续张开嘴部的时间，可以看出驾驶员的疲劳程度，即打哈欠的持续时间越久，驾驶员疲劳状态越严重，可以计算打哈欠的持续时间如下式：

$$T_{yawn} = T_3 - T_2 \quad (4.3)$$

经过医学实验数据表明，人们疲劳时往往伴随着频繁地打哈欠，而一次打哈欠的持续时间大概在 4-5 秒，当检测到驾驶员的嘴巴持续张开且维持的时间大于预设值时，则可以判断驾驶员可能处于疲劳状态，若打哈欠的持续时间过长或打哈欠的次数过多，则可以判定驾驶员处于重度疲劳状态。当打哈欠的参数值越大，疲劳程度越大。

### 4.2.2 嘴巴数据集

类似于眼睛特征提取，根据前面的面部特征提取，可以将嘴巴区域分割出来，从而获得嘴巴特征区域。由于网上的嘴巴数据集相对较少，本文自收集了一些嘴巴图片作为模拟数据集。自收集模拟数据集包括来自 18 个志愿者的 11357 嘴巴样本（见图 4.6）。这些嘴巴样本分为张开和闭合情况。这里将上述数据集中的图片缩放到 48\*48。





图 4.6 部分自收集嘴巴数据示例

### 4.2.3 实验及结果分析

每次在训练集中随机选取 10% 的图像训练 SC-Net 模型，从实验分析可以得出，当训练次数越高时，训练精度会越高，但其测试集精度会呈现一种先升高后降低的趋势，由欠拟合到收敛状态再到过拟合状态。当训练次数小于 500 时，准确率不到 90%，但当训练次数超过 5000 次时，训练准确率达到 99.9% 的同时测试结果只有 88.1%，因此需要取得中间合适的训练次数，这里选择训练 3000 次使效果最佳。训练过程如下图 4.7 所示，可以看出，训练 3000 次后模型得到了收敛。

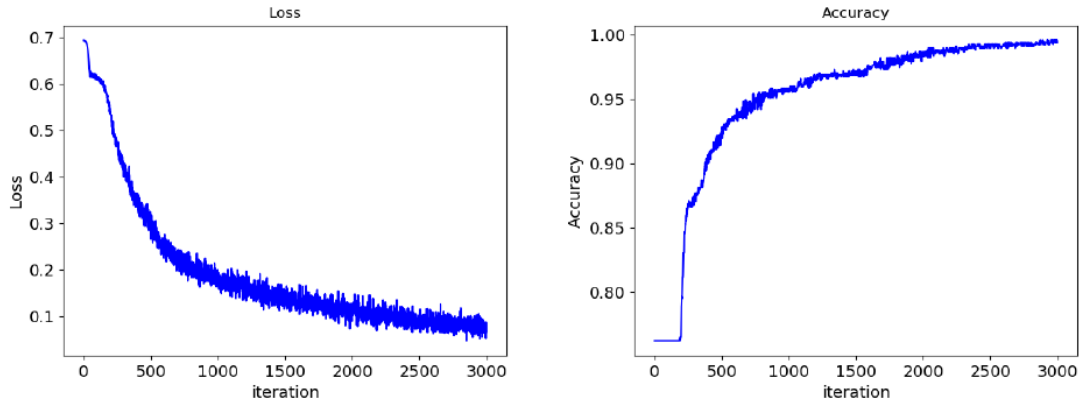


图 4.7 用 SC-Net 检测嘴巴状态的训练过程

为了证明我们模型的有效性，我们使用自收集模拟数据集训练和测试。自收集模拟数据集中 90% 用于训练，10% 用于测试。为了验证 SC-Net 在嘴巴检测方面的效果，我们将其与 VGG-16、ResNet18 和 Alexnet 进行了比较。此外，我们测试了图像预处理的效果（见表 4.3）。从实验结果可以看出，SC-Net 具有较高的检测精度，不进行图片预处理（IP）时，精度会略微下降。因此，检测前的图像预处理是非常必要的。

表 4.3 在自收集数据集上使用不同网络进行嘴巴状态检测的性能比较

网络	数据集	准确率, %
VGG-Net with IP	自收集数据集	85.93
AlexNet with IP		99.07
ResNet18 with IP		98.78
SC-Net without IP		94.87
<b>SC-Net with IP</b>		<b>99.24</b>



## 4.3 头部状态估计

在驾驶过程中，驾驶员会频繁地转动头部来观察周围环境及实时地规划驾驶路线，有时会出现比如低头、左顾右盼或背对相机，造成面部特征区域被遮挡，无法得到嘴巴或眼睛区域，影响面部特征的状态检测，从而造成疲劳检测结果不可靠。相对而言，头部姿态估计能很好的显示驾驶员的行为状态。综上所述，行驶过程中的头部姿态估计也是重要的组成部分，可以直接反映驾驶员的疲劳状态，同时可以弥补面部特征状态检测方法的缺陷，所以本文提出将头部姿态估计方法加入疲劳检测系统。

### 4.3.1 头部状态估计定义

头部状态是指某一瞬间人体头部在三维平面中相对坐标轴的偏转估计。在图像视觉领域中，头部状态估计指的是在视频或图像中对人体头部的偏转角度的计算。在实际场景中，头部状态估计是计算相对相机的偏转角度。理论上，人体头部偏转可以在三个方向上进行，且相对三位坐标轴的偏转范围为-90 度到 90 度。在实际情况中，头部偏转角度达不到这么大的范围，成年人的头部偏转范围应为-79 度到 75 度的偏航角运动、-60 度到 69 度的俯仰角运动和-41 度到 36 度的旋转角运动。

引入欧拉角的概念来分析头部状态，如图 4.8，三个不同方法的欧拉角可以表示头部的姿态，可以看作用向量表示。相应的头部状态估计可以根据计算俯仰角、偏航角和旋转角，投影到三位坐标上，x 轴表示俯仰角（pitch），y 轴表示偏航角（yaw），z 轴表示旋转角（roll）。

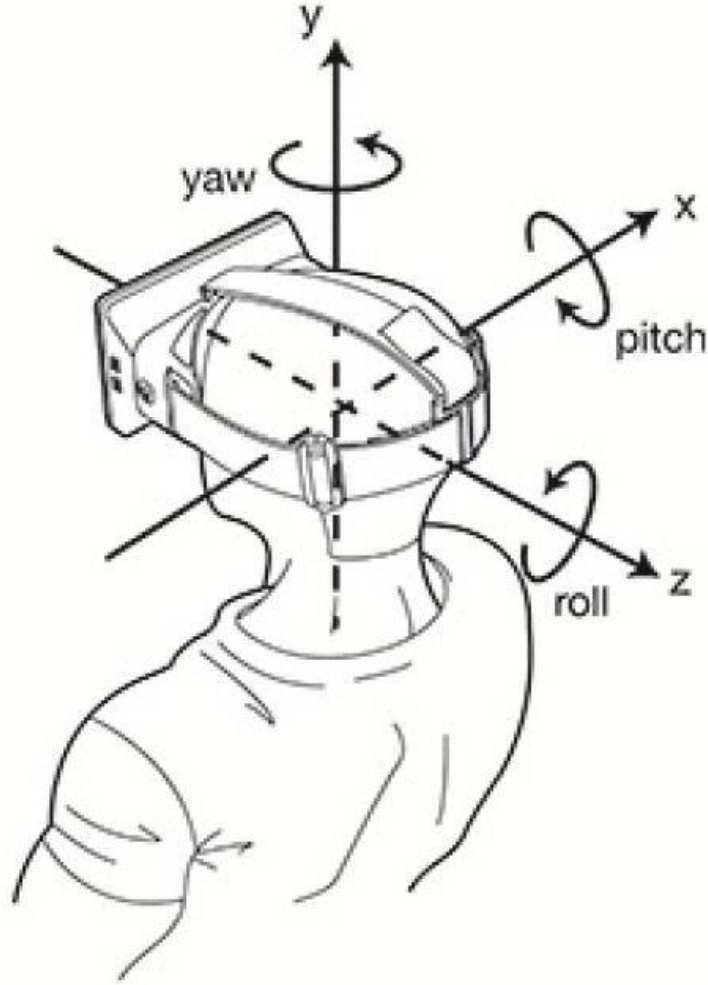


图 4.8 欧拉角表示头部姿态的示意结构

#### 4.3.2 基于 HP-LSTM 的点头状态估计

Ruiz 等<sup>[45]</sup>提供了一种在帧上估计头部姿态的神经网络 Hopenet，可快速的估计头部朝向，以供后续判断驾驶员是否点头。同时，估计头部姿态可以有效的提取对齐特征，这可有效地表示面部昏昏欲睡的状态。本文选择借鉴 Hopenet 网络和长短期记忆 (LSTM) 网络进行头部姿态估计。

借鉴 Hopenet 的思路，我们使用 resnet50 为骨架，后面接上三个全连接层 (FC)，每个层单独预测，分别预测人脸的欧拉角 (Yaw, Pitch, Roll)。其中，FC 层的全连接数是 bin 个，也就是将全部-99 到+99 一共 198 个数值每三个数分为一组，fc 连接数就是 66。先做分类，然后将分类的结果映射到一个范围，这样精度会有大的提升。而且是多个损失，分类的损失占比会影响梯度方向，从而会起到一个导向作用，引导回归往一个合适的方向，这是梯度方向上的引导。

对全连接层的结果做 Softmax，就把 fc 的值映射成概率值，所有映射结果相加为 1，映射成概率就能很方便求出期望，即网络的输出又被映射到[0,60]这个区间范围内，然后乘以 3 减去 90，这个区间范围就被映射到了[-90,+90]这个区间范围，也就是我们需要的

回归。然后就是计算回归的损失，用的是均方误差损失。

与前面分类的交叉损失熵按照一定权重加权求和，然后对最终的损失梯度反向，就完成了整个训练过程。训练的结构示意图如图 4.9 所示。

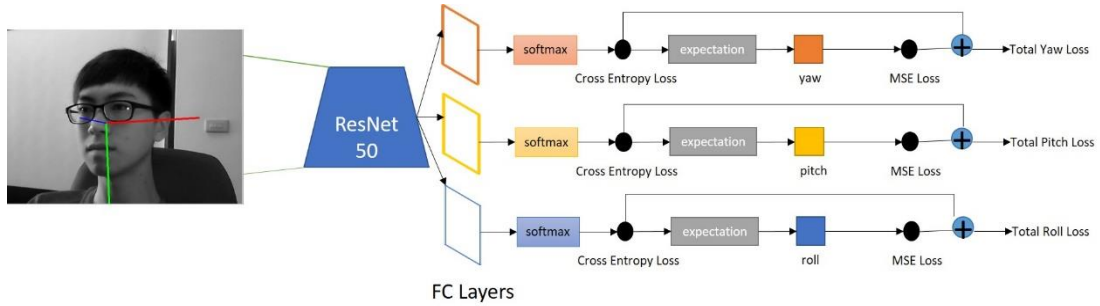


图 4.9 训练 Hopenet: 合并交叉损失和均方误差损失的 ResNet50 结构

将人脸图片输入上述训练后的 Hopenet 后，就可以得到人脸姿态方向的三个角度和 198 个中间特征值。

尽管我们训练了一个模型来预测每帧图片中驾驶员的头部姿态，但预测仅仅基于单帧，仅仅看一帧图片的信息无法检测驾驶员是否点头。为了解决这个问题，我们引入了时序的 LSTM 网络来抉择是否产生点头行为。我们通过将上述 Hopenet 网络中插入 LSTM 模块组成 HP-LSTM 网络，网络通过当前帧和当前帧之前的一系列视频帧序列（这里我们选择 50 帧）的特征值判定当前帧是否在点头过程中。Hopenet 可以得到 198 个特征值，输入 LSTM 层后，接着是新的预测层（跨时间的权重共享）以预测每个帧的点头分数  $Y_{td}$  ( $t = 0, 1$ )，即判定当前帧是否属于瞌睡点头中的一帧。我们仍然使用交叉损失熵来训练 LSTM。改进后的 HP-LSTM 网络结构如图 4.10 所示。

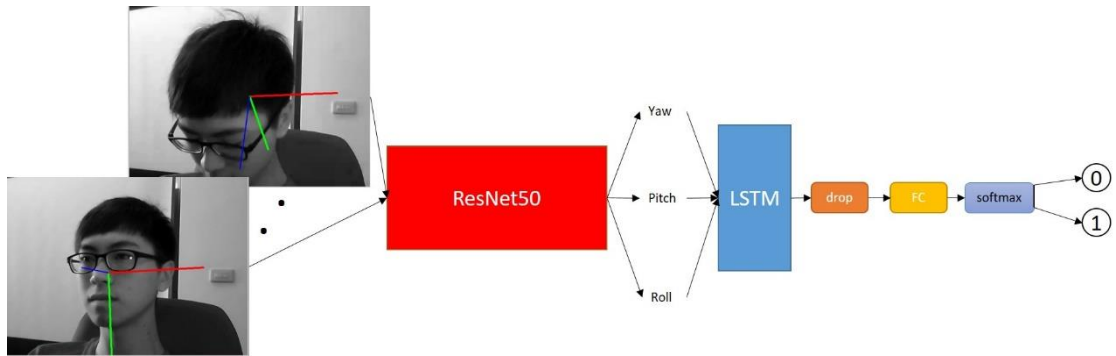


图 4.10 改进的 HP-LSTM 网络结构

### 4.3.3 实验结果分析

数据集:NTHU-DDD<sup>[63]</sup>数据集是由台湾国立清华大学开发的数据集，在 2016 年亚洲计算机视觉大会的疲劳检测会议上使用，旨在检测视频中的瞌睡。这些视频是在真实和不同的照明条件下拍摄的。场景包括不戴眼镜（白天）、戴眼镜（白天）、戴太阳镜（白天）、不戴眼镜（夜晚）和戴眼镜（夜晚）。数据集可以分为训练集、交叉验证集和测试集，每个视频疲劳状态包含两种状态:疲劳和不疲劳。每个视频都有多种状态标注，眼睛正常或闭合，嘴部讲话、正常或打哈欠，头部点头状态。这里将数据一分为二：一部分

数据有正常状态和疲劳状态比如打哈欠和闭眼，进行眼嘴状态检测；一部分数据有包括点头的疲劳状态视频和一些正常状态视频，用来点头状态估计。

数据处理：NTHU-DDD 数据集的注释中，头部姿态有三种状态（0,1,2）分别代表静止（0），点头（1），左顾右盼（2），有且仅有点头（1）与瞌睡最相关，因此我们将状态 0 和状态 2 合并为状态 0，状态 1 保持不变。

为了验证头部姿态估计在实际情景中的作用，我们使用 NTHU-DDD 数据集。将 NTHU-DDD 数据集的视频输入 HP-LSTM 网络。不同的时间步长会有不同的结果，为了得到最佳时间步长  $T$ ，选择不同的连续的  $T$  帧作为滑动窗口进行实验。实验结果如表 4.4 所示。可以看出，相比之下  $T$  为 50 时，效果最佳。同时，为了验证此模型在瞌睡点头检测的有效性，我们将其与一些先进的办法进行了比较（见表 4.5）。结果表明，其他方法因其主要目的非头部姿态预测，效果不佳，相比之下，可以看出 HP-LSTM 网络在点头检测方面表现最佳。这里选择单一 LSTM 单元实验，因为 LSTM 网络的输入参数较少，多个 LSTM 单元会因训练结果不收敛造成网络精度下降。

表 4.4 具有不同时间步长  $T$  在验证集上的比较

$T$	准确率, %
20	93.1
40	95.04
<b>50</b>	<b>96.1</b>
60	95.82
80	95.98
100	96.01

表 4.5 与一些先进方法在验证集上的精确度比较

方法	准确率, %
3D-DCNN[65]	93.8
CNN-LSTM [64]	92.5
MSTN[41]	93.3
<b>HP-LSTM(本文)</b>	<b>96.1</b>

## 4.4 本章小结

本章主要对面部特征状态检测算法、及头部姿态估计算法进行了研究。为保证眼睛及嘴巴状态检测的高精度和高效率，提出了 SC-Net 用于眼睛和嘴巴的状态检测，通过实验表明了本方法的优越性，解决了面部偏转下的眼睛状态识别不准确等问题。然后介绍了面部状态估计相关概念，在眼睛嘴巴状态检测的同时，使用了本文提出的 HP-LSTM

网络来进行驾驶员面部姿态估计，快速准确地获得头部姿态的三个欧拉角方向，并判断出驾驶员的头部是否处于点头区间。最后通过实验验证了该方法的有效性，同时通过与其他基于深度学习的头部点头估计方法进行比较，展示了本方法的优越性。

## 第5章 基于特征融合的疲劳检测

在驾驶员实际行车过程中，可以通过相机实时记录他的面部特征状态，通过图像信息去判断驾驶员的疲劳状态等级。但又因疲劳驾驶是一个复杂的行为状态，仅通过单面部特征判断驾驶员疲劳状态是非常不全面的，容易受到复杂情况的影响使得准确度不高，例如驾驶员遮挡眼睛区域、嘴部区域有异物遮蔽或者是空间电磁干扰等，因此需要使用特征融合的方法判断是否疲劳。眼睛特征、嘴巴特征和头部姿态特征与疲劳检测算法关系最为密切，本文提出一种基于眼部特征、嘴巴特征和头部姿态特征的融合算法来检测疲劳，可以使疲劳检测达到高鲁棒性、高准确性和高稳定性。

### 5.1 特征融合方法及疲劳评价基准

本文对常用特征融合算法进行了分析与研究，将其大致介绍一下。模糊理论算法：适用于不确定问题的信息融合；贝叶斯推理：适用于有先验知识的场景；投票法：适用于投票特征足够多的场景；卡尔曼滤波：适用于有一定数据规模的同构信息融合；加权平均法：适用于小规模数据的同构信息融合；主成分分析(PCA)：适用于高斯特征融合；典型相关分析(CCA)适用于特征之间具有内在联系的场景；神经网络：适用于输入输出关系不清晰的信息融合。根据以上的分析及对比后，本文选择使用简单实用的加权平均方法进行疲劳检测。

在对特征信息进行评价分析之前，需要建立一个多信息源融合的疲劳检测评价基准，构造一个具有科学性、可靠性和稳定性的评价基准是至关重要的。针对基于面部特征的疲劳检测等级划分，可以将其分为正常状态、疲劳状态和瞌睡状态三个等级。疲劳评价基准如图 5.1 所示，当驾驶员处于清醒状态时，精神力保持集中且面部特征均保持正常状态；当驾驶员处于疲劳状态时，注意力不再集中，头部会偶尔向侧面倾斜或处于点头动作区间，眼睛闭合时间会大于正常眨眼时间，且伴随着打哈欠动作；当驾驶员处于瞌睡状态时，精神力基本涣散，头部无法长时间保持正常姿态或长时间偏向一侧，嘴巴打哈欠变少，眼睛完全闭合一定的时间。

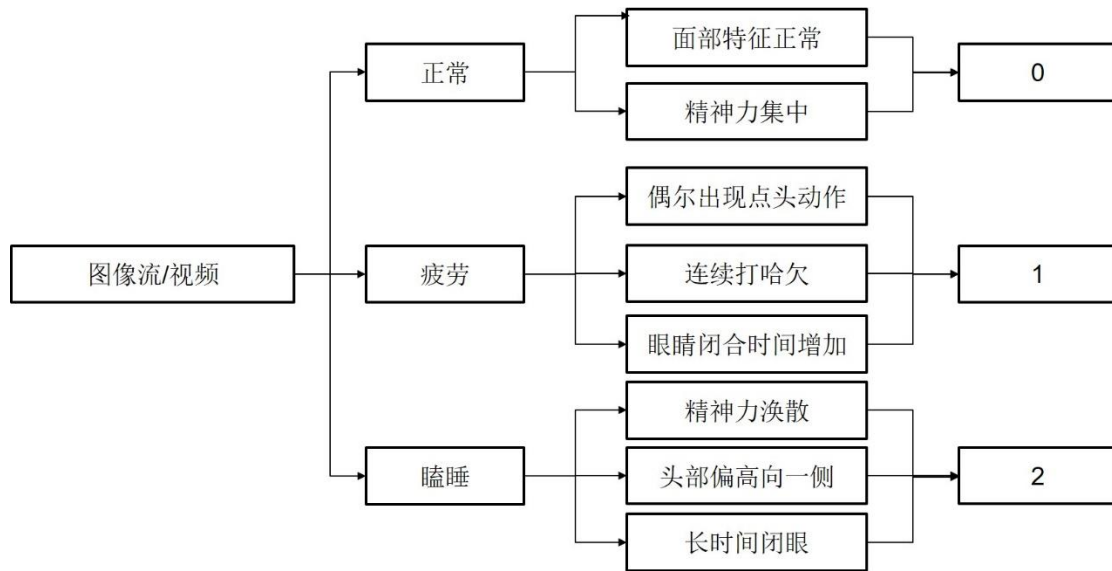


图 5.1 疲劳检测评价基准

在建立疲劳评价的初步标准后，可以构建一个模范化的疲劳驾驶基准。但不同状态下的面部特征和头部状态都会有无限的变化，所以需要更加细致的量化特征状态评价标准来检测疲劳。

## 5.2 面部特征参数提取

基于明显的驾驶员面部特征，本文建立了三级疲劳评估基准，为了对细微的特征状态有更深入地评估，本文对包括面部特征状态及头部状态信息的多特征信息进行量化提取及分析。当驾驶员疲劳时，他们通常会出现一些生理反应，比如长时间闭眼，点头，打哈欠等现象。仅依据单一的状态判断疲劳存在一定的局限性，因此本文在基于卷积神经网络的状态识别结果的基础上，以 60 秒为一个时间周期对各种特征统计分析，主要提取的信息有：PERCLOS 参数、眨眼频率、头部点头频率、打哈欠频率及最长闭眼时间。

### 5.2.1 眼睛信息提取

#### (1) PERCLOS 参数

PERCLOS 是 Percent Eye Closure 的简称，是美国联邦公路管理局公布的最可靠的疲劳检测算法，也是公认的用于疲劳检测的最有效的准则。根据对眼睛状态的不同定义，该准则定义了包含 3 个评估基准： $P_{70}$ 、 $P_{80}$  和 EM，其中  $P_{80}$  被认为是检测疲劳最好的标准。 $P_{80}$  的计算方法为：当眼睑下垂挡住瞳孔面积超过 80% 时，眼睛处于闭合状态，计算闭合时间占统计时间的比例，单次眨眼过程如图 5.2 所示。

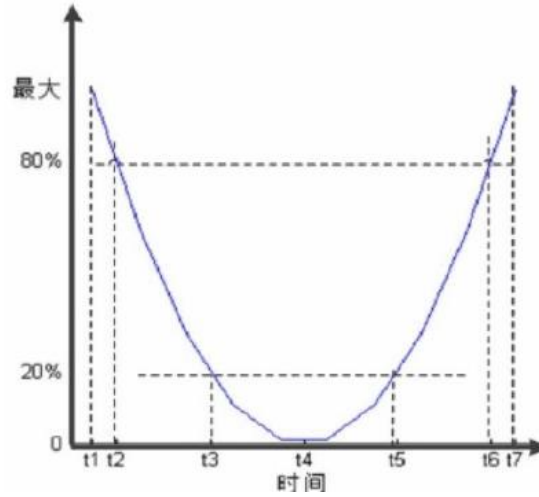


图 5.2  $P_{80}$  基准原理图

由于实际检测过程中，计算眼睑遮挡面积比较困难，而在前面章节已经实现了眼睛状态判断，因此本文基于 PERCLOSS 的定义，通过计算单位周期内眼睛闭合状态的总帧数与单位时间内总帧数的比值  $f_{per}$ 。计算公式如下：

$$f_{per} = n_{per}/N \times 100\% \quad (5.1)$$

其中  $n_{per}$  是单位周期内闭眼的帧数， $N$  是单位周期内帧的总数。PERCLOS 参数可用于量化驾驶员闭眼的程度。若  $f_{per}$  大于阈值，则将 PERCLOS 参数  $F_{per}$  置为 1，阈值设为 0.2。

## (2) 眨眼频率

眨眼频率是指人在单位时间内的眨眼次数，据统计显示，人在正常状态下每分钟眨眼约为 10-20 次，一般眨眼时间约为 2-6 秒，如下图 5.3 中从  $t_1$  至  $t_4$  所用时间。当驾驶员处于疲劳状态时，眨眼频率会增加，但当驾驶员严重疲劳或瞌睡时，眨眼次数会降低。因此，本文选择统计单位时间内眨眼次数计算眨眼频率，若眨眼频率不处于正常区间则认为驾驶员可能处于疲劳状态。



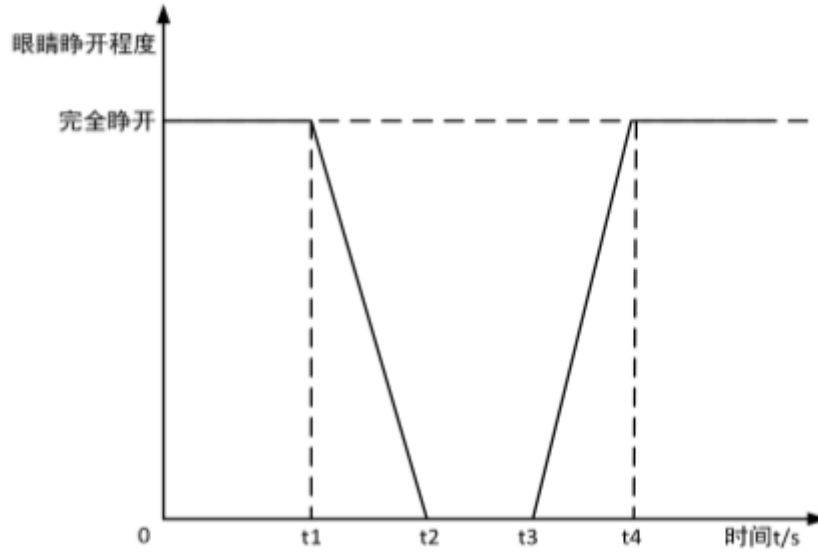


图 5.3 一次睁闭眼过程示意图

本文基于一次睁闭眼过程，通过前文获得的眼睛状态结果，判断是否经历一次眨眼过程，即眼睛从连续多帧的睁开状态到连续多帧的闭合状态再到连续多帧的睁开状态，记为一次眨眼过程。然后计算单位周期内眨眼的总次数为眨眼频率  $n_{\text{blink}}$ ，若时间周期内眨眼次数大于 25 次或小于 10 次，则将眨眼参数  $F_b$  置为 1。

### （3）最长闭眼时间

最长闭眼时间指的是单位周期内眼睛从闭合到睁开所持续的最长时间，如上图 5.3 所示的从  $t_1$  到  $t_4$  所经历的时间，本文为了计算方便，选择使用  $t_2$  到  $t_3$  所经历的时间作为一次统计的最长闭眼时间。人在疲劳状态下，闭眼时间会超过 1.5 秒，若处于瞌睡状态，则闭眼时间会超过 5 秒甚至 10 秒。运用系统内时钟计时统计最长闭眼时间  $T_{\text{mc}}$ ，若时间周期内最长闭眼时间超过阈值，则将该特征参数  $F_{\text{mc}}$  置为疲劳或瞌睡状态。

## 5.2.2 嘴巴信息提取

驾驶员在疲劳情况下会频繁打哈欠。打哈欠是一个较长的过程且嘴巴张开幅度较大。人感觉疲劳时，每次哈欠持续的时间大约是 6 秒及以上，当驾驶员张嘴超过一定时间时，认为其在打哈欠，此时需要暂停行驶并休息，我们可以通过周期内驾驶员打哈欠的次数判定其是否处于疲劳状态。由前文可知，我们获得了每一帧图像的嘴巴状态信息，当连续 3 秒嘴巴一直处于张开状态时，我们记为一次打哈欠。选择计算时间周期内打哈欠的次数作为嘴巴状态信息，若时间周期内打哈欠的次数超过阈值，则提醒驾驶员处于疲劳状态，并将嘴巴疲劳状态参数置为 1，其他情况置为 0，由于瞌睡状态下驾驶员罕有打哈欠情况，嘴巴状态信息不参与计算瞌睡判定。

$$F_Y = \begin{cases} 0, & n_y < N \text{ and } t_y < t \\ 1, & n_y \geq N \text{ or } t_y \geq t \end{cases} \quad (5.2)$$

其中  $F_Y$  为嘴巴疲劳状态信息， $n_Y$  为统计时间内打哈欠次数， $t_y$  表示一次打哈欠的

持续时间， $N=3$ ， $t=6$  秒。

### 5.2.3 头部信息提取

在正常情况下，头部姿态不会大幅度改变，即代表头部姿态的三个角度不会大幅度地改变。但当人们疲劳时，其注意力会下降，且他们对头部的控制力会大幅度下降，从而使头部下垂，为了保持驾驶状态，又要不停地抬头，因此会发生抬头和低头的周而复始的现象，即有可能会发生点头现象。当出现点头现象时，表明驾驶员已经处于疲劳状态，对车辆控制力会下降，从而可能引发交通事故，造成生命财产损失，因此需要对驾驶员进行点头频率的计算，这是对驾驶员姿态的分析，也是疲劳检测的重要因素。

根据前文提出的基于 HP-LSTM 头部姿态估计算法，因其良好的实时性和高准确率可以得到每帧图像中驾驶员的头部姿态及其变化，根据头部姿态参数可以判断出其是否处于点头区间的概率值，图 5.4 为打瞌睡时每一帧图像为点头区间的概率值。

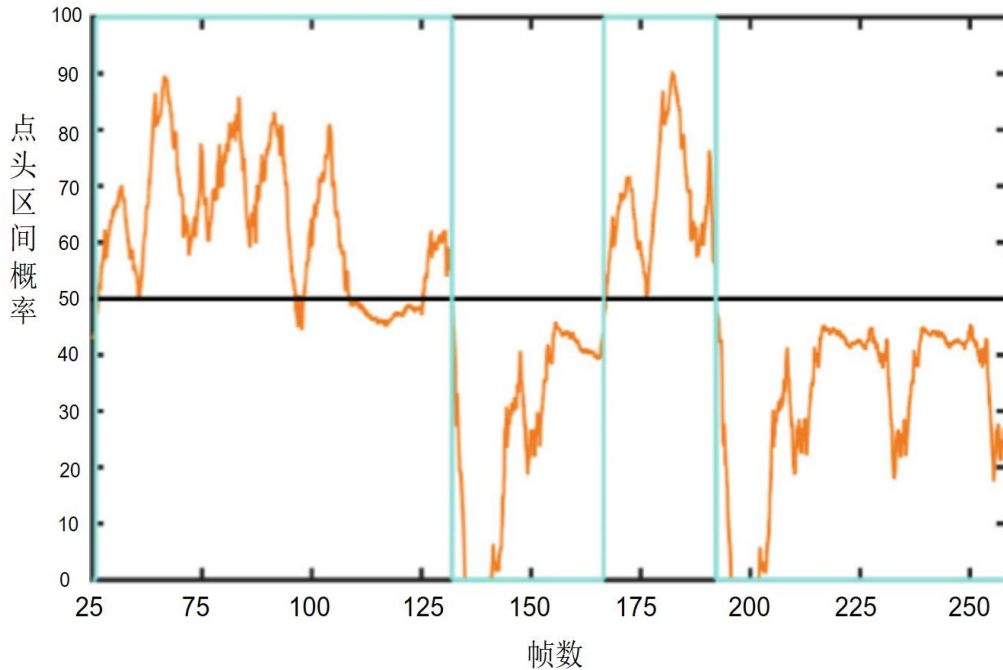


图 5.4 瞌睡状态下点头区间概率图

计算方法如下：阈值设置为 0.5，当视频帧数较多时，计算每一个时间周期内点头次数。若连续 15 帧图像的点头概率值大于阈值，则该图像序列处于点头区间，直到出现连续 15 帧图像小于阈值，结束一个点头区间，点头次数加一。得到时间周期内点头次数  $n_n$  后，将其与点头次数阈值比较，若大于阈值，则点头状态参数 ( $f_n$ ) 置为 1，其他情况置为 0。头部信息提取的方法类似嘴巴信息提取，点头状态参数的定义如下：

$$F_n = \begin{cases} 0, & n_n < N \\ 1, & n_n \geq N \end{cases} \quad (5.3)$$

其中， $n_n$  为统计时间内点头的总次数， $N$  为点头次数阈值，经过实验测试后发现其值设为 8 时准确率最佳。

### 5.3 基于加权平均的疲劳检测

根据上述眼睛、嘴巴和头部的特征信息，本文使用多特征加权平均方法进行疲劳检测，具体算法如下：

(1) 通过对眼睛、嘴巴和头部的疲劳特征运算得到疲劳状态值，通过疲劳状态值对参数加权，再相加，其运算公式如下：

$$RES = \sum W_i * F_i, \quad \sum W_i = 1 \quad (5.4)$$

其中  $F_i$  为特征的属性值， $W_i$  为对应的权重值， $i$  为眨眼频率、PERCLOS、最长闭眼时间、点头参数和打哈欠参数，状态值乘以对应的属性值，求和后得到加权疲劳参数 RES。

(2) 使用自采集视频寻找权重参数的最优值，数据集在 5.4 中介绍。使每个参数在 0.1-0.4 范围内变化，通过观察不同参数值对疲劳检测准确率的影响，可以确定眼睛、嘴巴及头部的各特征参数权重的最优值，最终的权重取值如下表 5.1 所示：

表 5.1 疲劳状态特征参数的权重表

特征参数	眼睛				嘴巴				头部			
	$F_{per}$		$F_b$		$F_{mc}$		$F_y$		$F_n$			
$F_i$	0	1	0	1	0	1	0	1	0	1		
$W_i$	0.2		0.1		0.2		0.2		0.3			

(3) 根据前文的疲劳评估基准，本文将疲劳分为三个状态：正常、疲劳和瞌睡。通过计算获得加权疲劳参数 RES 后，将该值与各疲劳状态对应，可以获得当前驾驶员的疲劳状态情况。其对应的关系表如下表 5.2 所示。

表 5.2 加权疲劳参数与疲劳状态对应表

疲劳等级	正常	疲劳状态	
		疲劳	瞌睡
$W_i$	RES<0.3	0.3≤RES<0.8	RES≥0.8

整体实现流程图如下图 5.5 所示，总结步骤如下：首先，相机连续拍摄司机，并将得到的图片进行预处理。其次，使用 MTCNN 网络获得预处理后的每帧图片的脸部和关键点来提取眼睛嘴巴区域。然后使用 SC-Net 检测每帧眼睛状态和嘴巴状态。同时将图片输入 HP-LSTM 时序神经网络结构，判断当前帧是否处于瞌睡点头动作区间得到点头状态。最后，根据眼睛嘴巴状态和点头状态的时间序列值，计算加权疲劳参数值，如果计算结果超过阈值，系统将提示并警告驾驶员已处于疲劳或瞌睡状态。

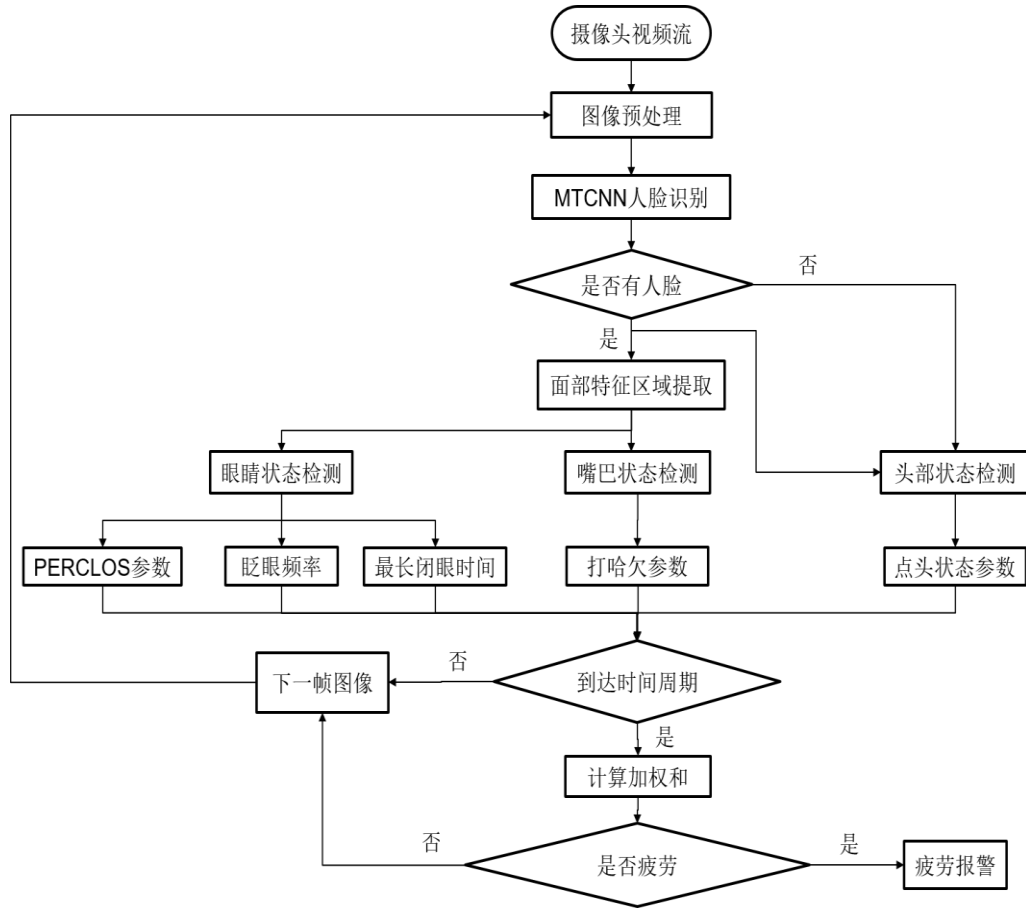


图 5.5 基于深度学习多特征融合的驾驶员疲劳检测方法的检测流程图

## 5.4 实验结果及分析

为了验证参数选择的有效性，将其用于疲劳驾驶检测实验。因为疲劳驾驶是违法行为，选择使用模拟瞌睡视频实验，采集 18 位志愿者模拟瞌睡和 ZJU 数据集的视频，如图 5.6 所示，每位志愿者 8 个视频，共计 144 个视频，每个视频 60 秒，每秒 30 帧。因为本文需要解决可能无法检测到人脸的情况，选择拍摄的视频中一半为仰角视频，一半为平角视频。实验结果如表 5.3 所示。



图 5.6 模拟瞌睡视频测试示例

表 5.3 疲劳检测最终检测结果

状态	总计	错误数量	准确率，%
仰角	72	2	97.2
平角	72	1	98.6

实验结果表明，该方法的准确率可以达到 97% 以上，且本方法的效果显著，准确率很高，可以很好的检测驾驶员是否处于疲劳状态。

最后运用模拟瞌睡，将前文基于单特征检测疲劳方法的实验结果与基于加权融合的疲劳检测方法的实验进行对比，其试验结果如下表 5.4 所示。

表 5.4 疲劳检测准确率对比

疲劳检测方法	准确率，%
MTCNN+眼睛识别	58.3
MTCNN+嘴巴检测	44.4
MTCNN+点头识别	55.6
MTCNN+多特征加权和	97.9

由表 5.4 可知，基于深度学习多特征融合疲劳检测的准确率要远远高于单面部特征疲劳检测的准确率。因此，本文选用基于深度学习多特征融合的疲劳检测方法来构筑疲劳检测系统。

## 5.5 本章小结

本章主要研究了一种基于眼睛、嘴巴和头部状态融合的疲劳检测方法。首先通过前文人脸检测及面部关键点定位提取获得每一帧的眼睛嘴巴及头部区域图像，然后利用卷积神经网络将眼睛嘴巴特征状态分类，利用循环神经网络获得头部点头状态，然后通过获得的特征状态得到眼睛的眨眼频率参数、PERCLOS 参数、最长闭眼时间参数，嘴部的打哈欠参数和点头状态参数。根据获得的各项疲劳状态参数，构建了一个疲劳检测评估基准，通过计算相应的权重加权和来检测驾驶员是否处于疲劳状态，实验表明本文基于深度学习多特征融合的疲劳检测方法准确率高且实用性好。

## 结论

近期,随着人们生活水平的不断提高,拥有汽车的人数逐步增加。随着汽车的增加,其给人们的生活带来快捷与方便的同时,汽车所引发频繁的交通事故也带来了惨重的经济损失,人民的生命也受到了威胁,道路交通事故已成为严重威胁人类生命财产安全的社会问题之一。虽然各国已经意识到预防疲劳驾驶的重要性,开始对该领域进行研究,并开发了一些疲劳检测系统,但我国相对于国外起步偏晚,预防疲劳驾驶的检测技术尚不完善,因此研究一种高效准确实时的疲劳驾驶检测方法对于预防疲劳驾驶有着巨大的意义。本文实现了一种基于神经网络多特征融合的疲劳检测方法,在相关数据集上进行测试,实验结果均表明本方法效果极佳。

### 1. 本文工作总结

本文提出了一种基于深度学习多特征加权融合的疲劳检测方法对驾驶员疲劳等级进行判定,主要研究工作与创新点如下:

(1) 总结和创新。通过了解疲劳检测领域的国内外研究现状,对当前存在的算法进行归纳与总结,提出问题与解决方案,最终选择使用基于深度学习进行疲劳检测算法研究的方案。

(2) 图像预处理。对图像进行平滑滤波及光照均衡处理,通过实验验证了各滤波算法的优缺点,选择中值滤波和直方图均衡化的光照均衡算法对图像进行滤波去噪和均衡光照强度。

(3) 基于 MTCNN 的人脸检测与特征区域提取。简要介绍了现有人脸识别算法和它们的优缺点,最后选择使用多任务级联卷积神经网络进行人脸识别及关键点定位。经过实验验证发现,基于多任务级联卷积神经网络的人脸识别算法能够精确地获得驾驶员面部区域,平均检测准确率可以达到 96.2%,并且检测速度快,实时性高。在得到人脸区域及面部五个关键点后,根据它们之间的几何关系提取眼睛和嘴巴区域,该提取方法方便快捷。

(4) 面部特征状态检测。介绍了面部特征识别的常用方法,面部特征状态识别是一个比较难以解决的问题。本文提出了两种面部状态检测方法:基于面部特征点坐标,计算关键点横纵比的特征状态检测方法;基于神经网络的面部特征状态检测方法。经对比发现神经网络识别面部特征状态准确率高、鲁棒性强,因此本文选择基于卷积神经网络的 SC-Net 识别眼睛和嘴巴状态。在获得眼睛和嘴巴区域后,该网络可以检测眼睛的睁眼、闭眼状态,还可以检测嘴巴的张开、闭合状态。不仅将 SC-Net 与其他算法进行对比,而且将其与其他基于神经网络的算法进行了对比。实验结果表明,基于卷积神经网络的 SC-Net 的识别准确率高,并且高于其他基于神经网络的算法。

(5) 头部姿态估计。研究了头部姿态估计相关方法,提出了基于卷积神经网络和长

短期记忆网络 (HP-LSTM) 的点头状态估计模型。图像序列在进行人脸识别的同时, 也进入卷积神经网络 ResNet50 模型, 将全连接层和 Softmax 层置于 ResNet50 之后, 经训练实现头部姿态估计, 可以获得头部姿态的三个欧拉角。然后级联长短期记忆网络 LSTM, 将获得的每一帧的三个欧拉角输入 LSTM 网络, 获得当前帧图像处于点头动作区间的概率值。本文在相关数据集上进行了验证, 实验结果显示, 基于 HP-LSTM 的点头状态估计方法效果最佳, 且远远领先于其他基于神经网络和循环神经网络的方法。

(6) 疲劳状态检测。根据面部特征状态识别结果, 获得眼睛的眨眼频率参数、PERCLOS 参数、最长闭眼时间参数, 嘴部的打哈欠参数; 根据点头状态估计结果, 获得点头状态参数。将眨眼频率参数、PERCLOS 参数、最长闭眼时间参数、打哈欠参数和点头状态参数进行加权信息融合, 获得最终的疲劳特征值。将疲劳特征值与疲劳等级评估基准作对比, 得到相应的疲劳等级并采取相应的措施。实验结果表明, 基于多特征融合的疲劳检测方法比基于单特征的疲劳检测方法实验效果更好、准确率更高。

## 2. 下一步工作展望

驾驶员在行驶过程中存在很多干扰因素, 由于我国相对于国外起步偏晚、预防疲劳驾驶的检测技术尚不完善, 且该领域是多个学科的交叉, 设计一个综合性能好、实时性高和稳定性强的系统仍比较困难。本文虽然在一定程度上取得了一定的进步, 但仍然存在很多缺点, 需要在以下几个方面进行进一步的改善。

(1) 融合车辆特征。本文研究的目标是基于驾驶员面部特征的疲劳检测, 重点放在计算机视觉领域, 可以添加车辆行驶特征信息如车速、车辆轨迹、方向盘转角、踏板制动力等融合检测, 可能会使疲劳检测的准确率更高、鲁棒性更强, 这些需要进一步实验。

(2) 系统在各开发平台。由于本系统整个实验是在 Linux 系统上进行实验, 与真实的驾驶环境有一定的差别, 目前只应用在办公室作业人员瞌睡检测, 且用到了造价比较昂贵的 GPU, 不利于推广与使用, 需要对硬件系统有更进一步的了解与开发, 若能够构建一个车载的实时疲劳检测系统, 更有利于疲劳检测系统的实用性。

综上所述, 如何进一步提高疲劳检测系统的准确性, 如何提高疲劳检测系统的可移植性依旧是非常艰难的挑战, 这些都是下一步工作的重点与核心。



## 参考文献

- [1] 《中华人民共和国 2019 年国民经济和社会发展统计公报》. [http://www.xinhuanet.com/finance/2020-02/28/c\\_1125637788.html](http://www.xinhuanet.com/finance/2020-02/28/c_1125637788.html), 2020-2-28
- [2] World Health Organization. Global Status Report On Road Safety 2018. 2018, 15(4): 286
- [3] Brian C, Tefft. AAA foundation for traffic safety, 2014
- [4] Shi S Y, Tang W Z, Wang Y Y. A Review on Fatigue Driving Detection. ITM Web of Conferences 12, 2017, 14-21
- [5] Zhang J Y, Qiu W W, Fu H J, et al. Review of Techniques for Driver Fatigue Detection. Applied Mechanics and Materials, 2013, 433(4350) : 928-931
- [6] Mittal A, Kumar K, Dhamija S, et al. Head movement-based driver drowsiness detection: A review of state-of-art techniques. IEEE International Conference on Engineering and Technology, Coimbatore: IEEE, 2016, 903-908
- [7] Mashko A. Review of approaches to the problem of driver fatigue and drowsiness. Smart Cities Symposium Prague, Prague: IEEE, 2015, 1-5
- [8] Kumari B M K, Kumar P R. A survey on drowsy driver detection system. 2017 International Conference on Big Data Analytics and Computational Intelligence, Chirala: IEEE, 2017, 272-279
- [9] Jap B T, Lal S, Fischer P. Comparing combinations of EEG activity in train drivers during monotonous driving. Expert Systems with Applications, 2011, 38(1):996-1003
- [10] Balasubramanian V, Adalarasu K. EMG-based analysis of change in muscle activity during simulated driving. Journal of Bodywork and Movement Therapies, 2007, 11(2):151-158
- [11] 耿磊, 袁菲, 肖志涛等. 基于面部行为分析的驾驶员疲劳检测方法. 计算机工程, 2018, 44(1): 274-279
- [12] Brandt T, Stemmer R, Rakotonirainy A. Affordable visual driver monitoring system for fatigue and monotony. 2004 IEEE International Conference on Systems, Man and Cybernetics, The Hague: IEEE, 2004, 7: 6451-6456
- [13] Friedrichs F, Yang B. Drowsiness monitoring by steering and lane data based features under real driving conditions. Signal Processing Conference. European: IEEE, 2010, 209-213
- [14] Murguia M I C, Resendiz C P. Detecting Driver Drowsiness: A survey of system designs and technology. IEEE Consumer Electronics Magazine, 2015, 4: 107-119

- [15] Mao M, Du L. Research on drive fatigue detection using wavelet transform. 2007 IEEE International Conference on Vehicular Electronics and Safety, Beijing: IEEE, 2007, 1-4
- [16] Li Z L, Jin X, Wang B J, et al. Study of Steering Wheel Movement under Fatigue Driving and Drunk Driving Based on Sample Entropy. Applied Mechanics and Materials, 2015
- [17] Li F, Wang X W, Lu B L. Detection of Driving Fatigue Based on Grip Force on Steering Wheel with Wavelet Transformation and Support Vector Machine. Neural Information Processing, 2013, 8228: 141-148
- [18] 毛喆.机动车疲劳驾驶行为识别方法研究.武汉理工大学博士学位论文, 2009
- [19] 张波.基于计算机视觉的复杂工况下驾驶人疲劳状态检测方法研究.清华大学博士学位论文, 2015
- [20] Rohit F, Kulathumani V, Kavi R, et al. Real-time drowsiness detection using wearable, lightweight brain sensing headbands. IET Intell. Transp. Syst, 2017, 11 (5): 255–263
- [21] Sangeetha M, Kalpanadevi S. Driver Fatigue Management System using Embedded ECG Sensor. International Journal for Scientific Research and Development, 2015,3(4): 1220-1224
- [22] Balasubramanian V, Adalarasu K. EMG-based analysis of change in muscle activity during simulated driving. Journal of Bodywork and Movement Therapies, 2007,11(2):151-158
- [23] Zhang A, Chen Y. EEG feature extraction and analysis under drowsy state based on energy and sample entropy. International Conference on Biomedical Engineering & Informatics, Chongqing, 2013, 501-505
- [24] Li G, Wan Y C. Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier. Sensors,2013,13(12):16494-16511
- [25] Tian Y L, Kanade T, Cohn J F. Eye-State Action Unit Detection by Gabor Wavelets. Lecture Notes in Computer Science,2000,1948:143-150
- [26] Wu Y S, Lee T W, Wu Q Z, et al. An Eye State Recognition Method for Drowsiness Detection. Vehicular Technology Conference,2010:1-5
- [27] 李延枫. 基于眼部识别的疲劳驾驶检测系统设计. 成都理工大学硕士学位论文, 2017
- [28] Kithil P W. Development of Driver Alertness Detection System Using Overhead Capacitive Sensor Array. SAE Technical Paper Series, 1998, 982292:35-56
- [29] 张伟. 基于机器视觉的驾驶人疲劳状态识别关键问题研究. 清华大学博士学位论文, 2011
- [30] 张万枝. 机器视觉感知下的车辆主动安全技术若干问题研究. 山东大学博士学位论文, 2015

- [31] Viola P, Jones M J. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 2004, 57(2): 137-154
- [32] Weng R L. Robust Feature Set Matching for Partial Face Recognition. *Proceedings of the 2013 IEEE International Conference on Computer Vision IEEE*, 2013, 601-608
- [33] Ji L, Gan C, Han S. TSM: Temporal Shift Module for Efficient Video Understanding. *ICCV*, 2018
- [34] Zhang X, Zhao J B, Cun Y L. Character-level convolutional networks for text classification. *Adv. NIPS*, 2015
- [35] Zeng R H, Huang W B, Tan M K, et al. The IEEE International Conference on Computer Vision (ICCV), 2019,7094-7103
- [36] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015:815-823
- [37] Li J, Wang Y, Wang C, et al. DSFD: Dual Shot Face Detector. *CVPR*, 2019
- [38] Zhang K P, Zhang Z P, Li Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 2016
- [39] Sheng Y, Xiao H L, Wei H Z, et al. Eyes state detection method based on LBP. *Application Research of Computers*, 2015
- [40] Dong Y, Zhang Y, Yue J, et al. Comparison of random forest, random ferns and support vector machine for eye state classification. *Multimedia Tools Appl.*, 2016, 75 (19):11763–11783
- [41] Shih T H, Hsu C T. MSTN: Multistage Spatial-Temporal Network for Driver Drowsiness Detection. *Computer Vision–ACCV 2016 Workshops*, 2017
- [42] Wang J, Yu X P, Liu Q, et al. Research on key technologies of intelligent transportation based on image recognition and anti-fatigue driving. *EURASIP Journal on Image and Video Processing*, 2019
- [43] Ng J, Gong S. Composite support vector machines for detection of faces across views and pose estimation. *Image and Vision Computing*, 2002, 20(5):359-368
- [44] Qiao T Z, Dai S L. Regression Forests for Head Pose Estimation Analysis. *Journal of Computer-Aided Design & Computer Graphics*.2014.07:1151-1158
- [45] Ruiz N, Chong E, Rehg J M. Fine-Grained Head Pose Estimation Without Keypoints. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018
- [46] Du W, Wang Y, Qiao Y. RPAN: An End-to-End Recurrent Pose-Attention Network for Action Recognition in Videos. *2017 IEEE International Conference on Computer Vision*

- (ICCV), IEEE, 2017
- [47] Sigari M H, Pourshahabi M R, Soryani M, et al. A review on driver face monitoring systems for fatigue and distraction detection. *Int. J. Adv. Sci. Technol*, 2014, 64:73–100
  - [48] Wierwille, Ellsworth W W. Evaluation of driver drowsiness by trained raters. *Accid. Anal. Prev.*, 1994, 26, (5), 571–581
  - [49] 李娟, 王富, 王维锋等. 基于数据融合的疲劳驾驶检测算法. *武汉工程大学学报*, 2016, 38(5): 505-510
  - [50] 黎亚平, 周杰, 黄磊等. 国内外驾驶疲劳状态检测技术的现状与发展. *上海学报*, 2010, 05:1-9
  - [51] Zhu Y, Gu W, He L, et al. Hierarchical CNN based Real-time Fatigue Detection System by Visual-based Technologies Using Multi-Scale Pooling Model. *IET Image Processing*, 2018
  - [52] Robertson N, Reid I. Estimating gaze direction from low-resolution faces in video. 9th IEEE ECCV, Graz, Austria, 2006, 402–415
  - [53] Gourier N, Maisonnasse J, Hall D, et al. Head pose estimation on low resolution images. *Multimodal Technologies for Perception of Humans*, New York, USA: Springer-Verlag, 2007, 270–280
  - [54] Xing Y, Lv C, Zhang Z Z, et al. Identification and Analysis of Driver Postures for In-Vehicle Driving Activities and Secondary Tasks Recognition. *IEEE Transactions on Computational Social Systems*, 2017
  - [55] 张铮, 倪红霞, 苑春红等. 精通 Matlab 数字图像处理与识别. 北京: 人民邮电出版社, 2013, 85-143
  - [56] 梁琳, 何卫平, 雷蕾等. 光照不均图像增强方法综述. *计算机应用研究*, 2010, 27(5):1625-1628
  - [57] Chaudhari M, Sondur S, Vanjare G. A review on Face Detection and study of Viola Jones method. 2015
  - [58] Yang S, Luo P, Chen C L, et al. WIDER FACE: a face detection benchmark. *Computer Vision Pattern Recognition*, Las Vegas, USA, 2016, 5525–5533
  - [59] Kostinger M, Wohlhart P, Roth P M, et al. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2011, 2144-2151
  - [60] Jain V, Learned-Miller E G. FDDB: A benchmark for face detection in unconstrained settings. Technical Report UMCS-2010-009, University of Massachusetts, Amherst, 2010
  - [61] Pan G, Sun L, Wu Z, et al. Eyeblink-based anti-spoofing in face recognition from a generic webcam. *IEEE Int. Conf. Computer Vision*, Rio de Janeiro, Brazil, 2007, 1–8

- [62] Song F, Tan X, Liu X, et al. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognition*, 2014, 47, (9):2825–2838
- [63] Weng C H, Lai Y H, Lai S H. Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network. *Asian Conference on Computer Vision*. Springer, 2016, 117-133
- [64] Wang C Y, Yan T, Jia H B. Spatial-Temporal Feature Representation Learning for Facial Fatigue Detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 2018
- [65] Jongmin Y, Sangwoo P, Sangwook L, et al. Representation Learning Scene Understanding and Feature Fusion for Drowsiness Detection. *Computer Vision–ACCV 2016 Workshops*, 2017, 165-177

## 附录 A 攻读硕士学位期间发表论文目录

- [1] 一种基于深度学习多特征融合的驾驶员疲劳检测方法，湖南大学，发明专利申请公布及进入实质审查通知书：201910974764.3，2020 年 02 月 06 号

## 附录 B 攻读硕士学位期间参与的科研活动

- [1] 太阳村室内监控系统研发，广西铁路局项目，2018-2019
- [2] 产品质量自动检测机器人，互联网+和挑战杯创新项目，2018

## 致 谢

白驹过隙，日月穿梭，转眼间在湖南大学的三年研究生生活即将结束，三年的研究生生涯经历了许多，也学会了许多，这段经历是我人生中宝贵的财富。从刚入校园时对研究生生活的憧憬与向往，到后来选择导师及研究方向的迷茫，再到论文开题报告、中期检查以及最后的论文撰写与答辩。这期间，有太多的事、太多的人值得感激。三年的研究生生活我要感谢的人很多，感谢给予我指导的导师，感谢陪伴我成长的同学朋友，感谢支持我的父母。

首先，我要对我的导师张汗灵教授表达我最真挚的感谢。研究生三年间，给予了我很多的帮助与关怀。从研究方法的选择，到论文等资料的阅读，再到实现项目的经验累积，到后面的发明专利的设计与实现，老师都给了我许多帮助。老师凭借他在机算计视觉及图像处理领域的专业能力，一步步引导着我，使我从一个图像处理领域的新手，变成了能独立完成图像处理相关科研的研究者。在平时的学习中，老师也会定期开展组会，对我们在研究当中遇到的各种问题一一地解答。导师对项目科研事业的热爱和对知识探索的积极态度潜移默化地影响着我。再次衷心感谢我的导师。

其次，我要感谢实验室王庭扬、尹兆远、夏晨星、徐昆昆和郑熠几位师兄对我的帮助与指导，认真地回答我的每一个提问，帮助我解决我所遇到的困难。还要感谢张明伟、龙亚艺和谢悦三位同门，不仅是学习上，还有在生活上，我们互相帮助、互相作伴。还有周丽娟和许馨文两位师妹，感谢你们对师兄的帮助。

另外，我最要感谢我的父母，他们一直给予我各种形式上的帮助，鼓励我不断的前进，支持我顺利度过三年的研究生学习。感谢你们为我默默的付出，让我逐渐能够独当一面。

最后，祝愿母校越来越好，祝愿实验室的成果一年比一年多！我会带着三年宝贵的经历继续前行。祝大家诸事顺利！

李皓星

2020年4月