

分类号_____

密级_____

U D C _____

编号_____

華東交通大學

全日制专业硕士学位论文

基于多特征的驾驶员疲劳检测方法

学位申请人： 张笔豪

专业领域： 交通运输工程

校内导师： 陆荣秀 教授

答辩日期：2022 年 6 月 1 日

基于多特征的驾驶员疲劳检测方法

摘要

随着我国汽车保有量不断上升，频发的交通事故也带来了巨大的伤亡和损失。根据研究，在造成事故的众多因素中，疲劳驾驶尤为严重。所幸的是，在疲劳导致交通事故之前，驾驶员会表现出明显的疲劳特征，这使得在疲劳导致交通事故之前就可以提醒驾驶员注意疲劳。因此，研究快速、可靠、高精度的驾驶员疲劳检测方法对道路交通安全有着深远意义。

道路交通安全的问题一直备受国内外众多研究学者的关注，在疲劳驾驶检测方面也出现了较多优秀的方法和成果。其中利用机器视觉通过驾驶员行为特征进行疲劳辨识的方法体现出无接触、低成本、高精度等优越性，但同时也存在着疲劳特征参数单一、特征提取网络层数浅和不能针对不同驾驶员定制疲劳阈值等问题。针对以上问题，本文提出了一种基于多特征的疲劳检测方法，具体描述如下：

首先，采用 HOG 和 ERT 算法进行人脸检测和人脸关键点定位，利用驾驶员二维面部关键点结合坐标系变换与相机标定进行头部姿态估计，解决人脸检测、关键点定位与头部欧拉角估计的问题。然后，针对驾驶员疲劳特征提取，利用 EAR、MAR 算法和头部姿态欧拉角提取驾驶员眼睛、嘴巴、头部姿态多重疲劳特征，同时将深度残差神经网络与 EAR 眼部疲劳特征提取算法相结合，并采用迁移学习算法使模型针对性和准确度更高。最后，利用眼睛、嘴巴、头部姿态的疲劳特征，建立了针对不同驾驶员的支持向量机个性化疲劳阈值模型，并综合 PERCLOS、眨眼频率、闭眼时长、哈欠次数、低头频率和头部异常角度等疲劳特征参数判断驾驶员的疲劳状态，并对其进行预警。实验表明，本文提出的基于多特征的驾驶员疲劳检测方法在 Yaw DD 和自建疲劳模拟数据集上具有较高的准确率和鲁棒性，在单一疲劳特征检测受阻时仍能提供良好的疲劳预警。

关键词：疲劳驾驶；人脸检测；头部姿态；深度学习；支持向量机

DRIVER FATIGUE DETECTION METHOD BASED ON MULTIPLE FEATURES

ABSTRACT

With the continuous increase of car ownership in our country, frequent traffic accidents have also brought huge casualties and losses. According to research, among the many factors that contribute to accidents, drowsy driving is particularly serious. Fortunately, drivers exhibit obvious fatigue signatures before fatigue leads to a traffic accident, which makes it possible to alert drivers to fatigue before it leads to a traffic accident. Therefore, the research on fast, reliable and high-precision driver fatigue detection methods has far-reaching significance for road traffic safety.

The issue of road traffic safety has always attracted the attention of many researchers at home and abroad, and there have been many excellent methods and achievements in fatigue driving detection. Among them, the method of using machine vision for fatigue identification through driver behavior characteristics reflects the advantages of non-contact, low cost, high precision, etc., but at the same time, there are also single fatigue characteristic parameters, shallow feature extraction network layers, and inability to target different drivers. Issues such as custom fatigue thresholds. In view of the above problems, this paper proposes a multi-feature-based fatigue detection method, which is described as follows:

First, the HOG and ERT algorithms are used for face detection and face key point positioning, and the driver's two-dimensional face key points are combined with coordinate system transformation and camera calibration for head pose estimation to solve face detection, key point positioning and head pose estimation. Euler angle estimation problem. Then, for the extraction of driver fatigue features, the EAR, MAR algorithm and the Euler angle of the head posture are used to extract the multiple fatigue features of the driver's eyes, mouth and head posture. At the same time, the deep residual neural network and the EAR eye fatigue feature extraction algorithm are combined, and the transfer learning algorithm is used to make the model more targeted and accurate. Finally, using the fatigue characteristics of eyes, mouth, and head posture, a personalized support vector machine fatigue threshold model for different drivers is established, and PERCLOS, blink frequency, eye closure duration, yawn times, head bowing frequency and head abnormality are integrated. Fatigue characteristic parameters such as angle determine the driver's fatigue state and give an early warning. Experiments show that the multi-feature-based driver fatigue detection method proposed in this paper has high accuracy and robustness on Yaw DD and self-built fatigue simulation datasets, and can still provide good fatigue warning when single fatigue feature detection is blocked.

Key Word: Fatigue driving, face detection, head posture, deep learning, SVM

目录

摘要.....	I
ABSTRACT.....	II
目录.....	III
第一章 绪论.....	1
1.1 研究背景与意义.....	1
1.2 疲劳驾驶检测方法国内外现状.....	3
1.2.1 基于主观评定的疲劳检测方法.....	4
1.2.2 基于生理参数的疲劳检测方法.....	5
1.2.3 基于车辆信息的疲劳检测方法.....	6
1.2.4 基于驾驶员行为特征的疲劳检测方法.....	7
1.3 研究的主要内容.....	9
1.4 论文组织结构.....	10
第二章 相关技术理论.....	12
2.1 深度学习理论.....	12
2.1.1 深度学习理论基础.....	12
2.1.2 卷积神经网络介绍与特点.....	12
2.1.3 卷积神经网络的结构.....	14
2.1.4 卷积神经网络训练过程.....	16
2.1.5 常见卷积神经网络.....	17
2.2 迁移学习.....	19
2.2.1 迁移学习定义.....	19
2.2.2 迁移学习分类.....	19
2.3 本章小结.....	20
第三章 驾驶员人脸检测和头部姿态估计.....	21
3.1 基于 HOG 的驾驶员人脸检测算法.....	21
3.1.1 HOG 算法原理.....	21
3.1.2 基于 HOG 算法的人脸检测实现流程.....	22
3.2 基于级联回归树的驾驶员人脸关键点检测算法.....	24
3.2.1 级联回归树人脸关键点检测算法原理.....	25
3.2.2 建立级联回归器.....	26
3.2.3 基于树的回归器.....	26
3.2.4 实验结果与分析.....	27

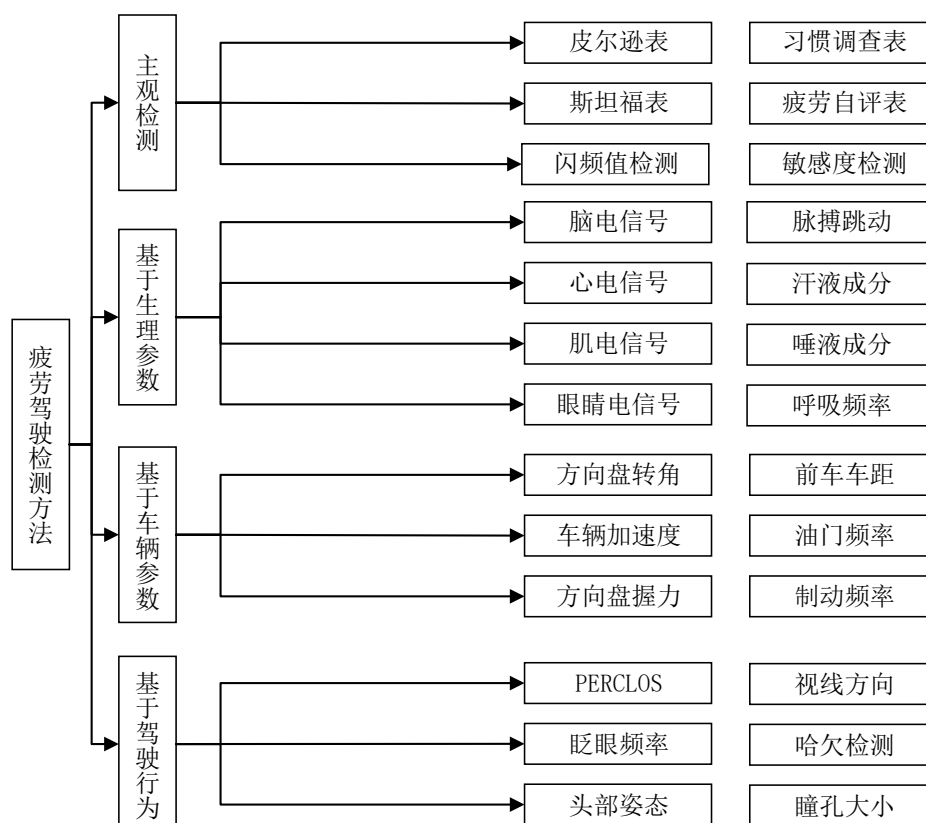


图 1-2 常见疲劳检测方法

Fig.1-2 Common fatigue detection methods

1.2.1 基于主观评定的疲劳检测方法

主观评定法是将驾驶员当下所反馈的主观感受、反应状态和心理状态等指标量化为驾驶员主观疲劳状态的一种方法。主观评定法可以划分为主观自评法和主观他评法，主观自评是驾驶员通过调查问卷、睡眠记录等方式对自我状态进行评定，主观他评是通过专家对驾驶员进行问答或通过观察驾驶员行为视频进行疲劳状态的评定。常用的主观自评的方法有：斯坦福嗜睡表 SSS（Stanford Sleepiness Scale）、卡洛琳斯卡睡眠尺度表 KKS（Karolinska Sleepiness Scale）、视觉类比量化表 VAS（Visual Analogue Scales）、Rhoten 疲劳量表和皮尔逊疲劳量表等。常用的主观他评方法是通过一组受训练的专家对实际场景驾驶下驾驶员面部特征和行为姿态进行疲劳评定。

基于主观评定的疲劳检测方法的优点是：操作简单、成本低、容易实施。其缺点首先是对于具有问答性质的疲劳量化表，驾驶员被专家询问时具有刺激性，会对当下的疲劳状态产生影响，准确度不高。其次，该方法只能离线检测，实时性差，不能在实际驾驶环境中驾驶员出现疲劳状态时及时评定并给予预警，只能用作实验室场景下事故的事后分析和复盘。

1.2.2 基于生理参数的疲劳检测方法

基于生理参数的疲劳检测方法就是通过接入人体的仪器来检测分析相关生理参数指标的变化进而判定人体是否疲劳，早期的疲劳研究便是在医学领域基于人体生理参数进行的，所以该类方法的发展时间较早。人体常见的可以反映疲劳状态的生理指标有脑电信号 EEG（Electroencephalogram）、心电信号 ECG（Electrocardiogram）、肌电信号 EMG（Electromyogram）和眼电信号 EOG（Electrooculography）等^[7]。在交通安全领域，常用于检测驾驶员疲劳的生理指标主要有以下三种：

（1）脑电信号检测（EEG）：由于人体的大脑信号反映着大脑此刻的状态，在人体处于疲劳时，大脑电信号会发生显著的变化，记录并分析这一信号的变化便可以判断人体是否处于疲劳状态，脑电信号被誉为疲劳检测上的“金标准”^[8]早在二十多年前，文献[9]就连续记录了 18 个在瑞典南部和斯德哥尔摩（500 公里）之间夜间行驶的卡车司机的脑电图，并得出驾驶员的嗜睡度与 EEG 中的 alpha 和 theta 波呈相关性，为基于 EEG 检测驾驶员疲劳状态奠定了基础。Tuncer T 等人基于一种动态中心和多阈值点的稳定特征提取网络，设计了一种基于 EEG 的驾驶员疲劳检测智能系统，实现了 92% 的驾驶员疲劳识别精度^[10]。陈骥驰等人对采集到的驾驶员脑电信息进行小波包分解与重构提取节律信号，通过计算和构建矩阵提取脑网络特征，最后以较少的电极数量实现了驾驶员的疲劳检测^[11]。Fnu 和 Rohit 等人使用具有径向基核的支持向量机(SVM)对来自驾驶员身上轻量化的脑电传感器 EEG 信号进行的频谱分析，对驾驶员的疲劳状态实现了高精度的分类^[12]。王斐等人在脑电信号的电极-频率分布图的基础上，用 SEED 脑电情绪数据集对搭建好的深度学习神经网络进行预训练后，再用迁移学习的算法将其运用于驾驶员疲劳检测上^[13]。

（2）心电信号检测（ECG）：当驾驶员出现疲劳时，ECG 信号（如：心率和心率变异性）等指标会随之规律变化。文献[14]通过将驾驶员的心率变异性使用小波变换后进行支持向量机建模，实验结果优于使用傅立叶变换的心率变异性建立的模型。徐礼胜等人利用差分阈值的方法确定 R 波的位置计算 R-R 序列的时域和频域特性后利用深度学习网络特征提取网络和随机森林分类器将驾驶员疲劳检测的准确率达到 91%^[15]。

（3）肌电信号检测（EMG）：肌电信号是人体中众多肌肉单元中的动作电位在时间和空间上叠加而来，驾驶员从长时间的高度精神集中驾驶转变为疲劳驾驶时，肌肉从紧绷转换到松弛僵硬，此时可以通过检测驾驶员 EMG 信号的变化来判断驾驶员是否疲劳。王琳等人采用颈 6 左右两侧上斜方肌和腰 4 左右两侧竖脊肌的表面肌电信号并通过降维构建基于 EMG 的驾驶员疲劳监测模型，最终实现了 90% 以上的准确率^[16]。

基于脑电、心电和肌电信号的疲劳检测方法能最直接反映出人体中因疲劳导致的生理参数变化，其优点是可靠性高、低误检率、实时性高，可以在驾驶员出现疲劳导致事故之前对驾驶员进行疲劳预警。但是这类方法的缺点是需要对驾驶员身体接入不同程度的电极等测量生理参数的设备仪器，对驾驶员的正常驾驶行为产生入侵性，影响驾驶员

的舒适性,甚至可能由于不适而导致交通事故。此外,这类方法需要价格较为高昂的仪器,系统成本高,不利于普及。

1.2.3 基于车辆信息的疲劳检测方法

当驾驶员处于疲劳状态时,在做出例如避让行人、转向、保持车道、控制车速等驾驶行为时会出现异常。基于车辆信息的疲劳检测方法就是在机动车上安装譬如油门踏板传感器、制动踏板传感器、方向盘传感器和车道图像传感器等设备来记录、检测车辆行驶数据进而判断驾驶员的精神状态。常用的检测疲劳驾驶的车辆参数信息有车辆加速度、方向盘压力、转角和车道偏移率。

通过在机动车加速踏板和制动踏板处安装传感器记录车辆的加速度曲线和踩下加速制动踏板的频率,并建立数据模型分析,便可以检测出驾驶员是否疲劳。Wang M S 等人通过实验证明使用车辆的横向加速度和纵向加速度进行组合,并使用 RF 算法进行输入数据处理,对驾驶员的疲劳辨识准确率要高于其他参数的组合,最终准确率达到 84.8%^[17]。文献[18]通过将提取到的驶员车辆的速度、横向位置和方向盘信息组成样本熵,构建 BP 神经网络对驾驶员进行疲劳检测。蔡素贤等人基于 CAN 总线提取车辆的 18 个与驾驶行为具有相关性的特征,并使用随机森林算法对提取的特征数据进行分析处理进而判断驾驶员是否疲劳^[19]。

驾驶员与车辆接触最直接也是最频繁的部件之一就是方向盘,所以可以借由方向盘信息分析驾驶员的状态,一般而言,当驾驶员处于疲劳状态时,对方向盘的操作频率和握住方向盘的压力会随之下降。通过在机动车的方向盘控制杆上安装角速度和压力传感器,便可以检测车辆在行驶过程中的方向盘转动的角度、角速度、频率和握力。Kyeheonk 等人通过在方向盘安装压力传感器,实现了对驾驶员的疲劳预警^[20]。张海兵通过柔性力传感器和 STM32 微控制器等传感器对驾驶员的手握方向盘的压力数据进行采集,再通过 RealView MDK 设计的系统将采集的数据与设定的疲劳阈值进行对比进而判断驾驶员是否疲劳^[21]。文献[22]提出了一种基于车辆方向盘转角复杂度与近似熵的驾驶员疲劳检测方法,通过输入方向盘时序 ApEn 和复杂度的非线性特征,建立疲劳等级辨识模型,实现了 84.6%准确率的疲劳辨识。柴萌运用马尔可夫模型(GM-HMM)对长途客车驾驶员疲劳参数中优选出 NMRHOLD、SWDR 等 7 个疲劳特征参数,建立长途客车驾驶员疲劳辨识模型,最终实现了 90.63%的准确率^[23]。

基于车辆信息的疲劳检测方法是一种基于车辆的异常状况间接反映驾驶员疲劳状态的方法。其优点是具有实时性、无接触等,相较于前一种方法由于不需要在身上放置传感器,所以对驾驶员无刺激性,不影响驾驶员的舒适度。但是,该类方法需要大量真实疲劳驾驶场景下的车辆特征数据进行模型训练,而研究者在真实疲劳状态时,驾驶车辆在道路上采集数据对驾驶员自身或无关的车辆行人具有很大程度的危险性。其次,该类方法会随着车辆行驶年限的增加,传感器设备老化而失效,成本高不易普及。除此之

外,该类方法在检测出车辆异常状况时,通常驾驶员已经完全失去了车辆的控制权,不能在驾驶员出现前期疲劳、车辆出现危险状况之前对驾驶员进行及时的疲劳预警。

1.2.4 基于驾驶员行为特征的疲劳检测方法

当驾驶员处于疲劳状态时,会不同程度地反应在驾驶员的面部和头部姿态,例如:眨眼、打哈欠或者低头。基于驾驶员行为特征的疲劳检测方法就是运用机器视觉算法提取驾驶员面部特征或头部姿态进行疲劳检测,通过分析驾驶员眼部特征(如:眨眼频率、眨眼速度、PERCLOS、瞳孔直径、注视方向等)、嘴部特征(如:哈欠频率等)和头部姿态特征(如:头部异常角度、低头频率等),提取能表征驾驶员疲劳的参数指标对驾驶员进行疲劳检测。

驾驶员眼部特征:眼睛不仅是人体获取外界信息最直接最重要的器官之一,还是反映人体状态的重要“窗口”。驾驶员在驾驶车辆过程中,80%-90%的信息是通过眼睛获得,眼部特征信息也是反映驾驶员疲劳状态的重要特征。在疲劳状态下,驾驶员眼睛的眨眼速度会变慢、眼睑闭合时长增加。不仅如此,进入疲劳状态时,眼睛瞳孔直径、注视焦点也会相应的变化。Catalbas M C 等人通过红外 LED 摄像机获取驾驶员眼部信息,运用眼睛瞳孔的加速度、速度和大小数据进行分析,建立了基于眼球扫视运动的驾驶员疲劳检测系统^[24]。文献[25]结合 Adaboost 和轮廓圆,通过 Adaboost 检测人脸和眼部区域,提取连通域的中心和半径作为特征向量,最后根据 PERCLOS 阈值判断眨眼。文献[26]建立了一套基于人脸检测、眼部检测、眨眼检测、PERCLOS 检测、头肩检测、疲劳程度分类等模块的公交车驾驶员疲劳检测系统,该系统采用一种基于光谱回归的持续睁眼程度估计方法,并基于 PERCLOS 对驾驶员状态进行了分类。文献[27]通过将人工先验信息集成到轻度级深度网络,使用 GP-VGG16 深度学习网络对驾驶员的眼部疲劳进行识别,提高眼部状态识别的准确性、稳定性和实时性。Song F 等人结合了多种特征集的优势来表征眼睑的丰富信息,并构建了眼睛状态识别模型,最终在 ZJU、CEW 数据集上表现出较好的识别效果^[28]。Chen P 等人基于肤色的特点,利用 YCbCr 的色彩空间变换实现人脸定位,采用二值化和二次积分投影的方法对人脸区域进行处理得到驾驶员眼睛的位置,最后分析眼睛瞳孔开度,通过双眼开闭比,检测疲劳驾驶^[29]。文献[30]通过将摄像头放置在驾驶员左侧而非正面,使用 SVM 和 Adaboost 在驾驶员单眼上完成眼睛是睁闭分类,最后根据眼睛状态确定微睡眠模式,并在需要时触发警报以警告驾驶员。Zhao Z 等人使用 MTCNN 进行人脸检测和特征点定位,再用 EM-CNN 卷积神经网络对 ROI 区域进行眼部和嘴部的状态识别从而判断驾驶员是否疲劳^[31]。高宁通过提出多通道时空信息量化驾驶员眼睛状态和基于差异化时空多尺度分析眨眼的方法,增加了提取帧间时序特征、几何特征和帧内图像特征的精度,提高了驾驶员眨眼检测的精确度^[32]。

驾驶员嘴部特征:当驾驶员在驾驶机动车时,嘴部状态通常有几种:嘴部闭合松弛、张嘴说话、唱歌、打哈欠。这几种嘴部状态嘴部的开合程度不尽相同,故可以根据嘴部几何形状的变化和持续时间将几种不同嘴部状态检测出来。其中打哈欠是驾驶员进入疲

劳状态的一种特征行为,所以检测是否疲劳驾驶可以通过检测驾驶员嘴部是否存在哈欠行为以及哈欠频率来进行。文献[33]通过改进的 Viola-Jones 算法和反投影理论检测司机嘴部变化,并于 APEX 平台搭建哈欠检测系统。Alioua N 等人提出使用支持向量机(SVM)和基于圆形霍夫变换(CHT)的驾驶员嘴部检测方法,并使用了 6 个模拟真实疲劳驾驶场景的视频样本进行测试,实验结果证明该方法的准确度达 98%^[34]。Du G 等人使用单个 RGB-D 摄像头来提取三个疲劳特征,然后把 RNN 层加入多模态融合递归神经网络获取特征的时间信息,最后整合三个特征来提高驾驶员疲劳检测的准确性^[35]。文献[36]将提取的具有低时间采样特性的 3D 深度学习网络用于细微的面部动作识别。

驾驶员头部姿态特征:驾驶员处于正常驾驶状态时,头部大部分时间朝向前方与眼睛注视方向一致,在进行转向和掉头等操作时,驾驶员眼睛看向车辆后视镜位置方向,头部随之进行一个侧向转动。在严重疲劳状态下,驾驶员头部会出现上下点头或者向左右下方歪头等异常行为姿态,这些由疲劳所导致的头部异常姿态与正常驾驶时的头部姿态大有径庭,所以对于疲劳状态的判断,可以根据检测驾驶员的头部姿态的异常或低头频率来实现。Tawari A 等人通过头部姿势进行动态注视区域估计,用以判断驾驶员是否处于正常驾驶状态^[37]。文献[38]计算列车司机头部旋转加速度和角速度,依据驾驶员头部的倾斜角度以及旋转角速度和加速度判断驾驶员是否疲劳^[38]。Shen Q 等人首先将定位检测到的驾驶员面部图像进行疲劳检测并计算相应的光流,其中在系统中使用了对比度受限的自适应直方图均衡(CLAHE)来减少不同光照条件的影响,然后针对每个特征设计了三个独立的双流网络结合 3D 注意力机制和 3D 深度可分离卷积来提取时间信息,最后三个子网络的输出将被级联并发送到全连接网络判断驾驶员状态^[39]。

基于驾驶员行为特征的疲劳检测方法相较于基于主观评定的疲劳检测方法具有实时性和前瞻性,可用作实际驾驶场景下的提前疲劳预警,因其不对驾驶员使用电极等传感器测量生理参数,相较于基于生理特征的疲劳检测方法具有无接触、无入侵性和不影响驾驶员舒适度的优点,此外由于硬件设备仅需摄像头,故相比基于车辆信息的疲劳检测方法具有硬件设备简单、成本低、易普及等优势。但是该类方法的缺点在于搭建模型系统涉及人脸识别、眼部识别、嘴部识别和头部姿态估计等复杂识别算法,开发难度大,且由于仅使用摄像头设备对驾驶员的特征信息进行提取分析,在某一特征(如:戴墨镜)收到遮挡时,该类方法就会因无法提取到特征而失效。

通过对国内外研究现状的分析,可以得出不管是国内还是国外,疲劳驾驶检测的研究从最初在驾驶员身上接入电极设备的接触式检测转向基于车辆参数信息和基于机器视觉的驾驶员行为特征的非接触式检测。随着机器学习和深度学习的快速发展,计算机算力呈指数级增长,基于机器学习和深度学习的机器视觉算法开始在人脸识别、图像处理、数据挖掘和分类预测等方面发挥重要作用。目前,越来越多的研究学者已经开始将目光聚焦在基于机器视觉的疲劳检测方法。

本文通过综合分析各类疲劳检测方法的利弊如表 1-2 所示，通过分析表中生理参数检测法、车辆参数检测法和驾驶员行为检测方法之间的优缺点，得出基于机器视觉的驾驶员行为检测方法具有其他方法难以企及的易用性、低成本、非接触的优点。随着机器学习和深度学习算法的发展，计算机硬件设备算力不断增加，软件开发难度也在下降，该方法受光照、遮挡的影响将进一步减小，准确度将进一步提高，未来更加具有实用性。

表 1-2 各类疲劳检测方法优缺点

方法	特征	优点	缺点
基于生理参数的疲劳检测方法	通过在驾驶员身上接入电极等设备采集生理参数信息	准确度高，可靠性强，可以客观反映驾驶员的疲劳程度	接触式测量，对驾驶员有入侵性，影响驾驶员正常操作
基于车辆信息的疲劳检测方法	通过在车辆安装传感器实时采集车辆运行参数	对驾驶员无接触，可以获取较多的判别信息	成本高，属于间接性检测，传感器易老化失效
基于驾驶员行为特征的疲劳检测方法	通过机器视觉，检测驾驶员面部和头部行为特征	非接触式，对驾驶员无入侵，硬件设备简单，易普及	算法复杂，特征提取容易受到光照和遮挡的影响

1.3 研究的主要内容

本文围绕如何实现对驾驶员快速、准确、稳定的疲劳检测以改善道路交通安全这一目标展开研究，明确由疲劳导致的交通事故占比高和基数庞大这一选题背景及其背后的研究意义后，对目前现有的四类疲劳检测方法进行了对比。综合分析得出，基于驾驶员行为特征的疲劳检测方法不仅精度高、无干扰，而且设备简单成本低廉。本文对该类方法进行深入探讨和研究发现其虽然有着其他方法不可比拟的优点，但是也存在着疲劳特征单一、特征提取网络不够深和对不同驾驶员个性化疲劳阈值等问题。基于此，本文主要研究内容和创新点如下：

(1) 使用基于 HOG 和级联回归树 ERT 的人脸检测和关键点定位算法。本文使用计算机视觉较为优秀的方向梯度直方图算法 HOG 对驾驶员进行人脸区域检测，该算法采用细胞单元 cell 的方式对图像进行梯度方向量化，具有几何不变性和光照不变性的优点，适合在具有不同光照情况的车辆中检测人脸。在驾驶员面部关键点定位环节，采用级联回归树 ERT 的人脸关键点检测算法对 HOG 算法输出的人脸区域进行关键点定位，为后续疲劳特征的提取提供坐标基础。

(2) 设计了基于面部和头部多特征的驾驶员疲劳辨识模型。目前现有疲劳检测方法对驾驶员进行疲劳特征提取时大多采用单一特征提取，基于机器视觉对单一特征进行疲劳特征提取时容易因这一特征受遮挡而导致整个疲劳检测系统失效。本文针对目前疲劳检测方法中疲劳特征参数单一、对疲劳状态表征不全的问题，提出基于驾驶员面部信

息和头部姿态多特征的疲劳检测方法,该方法使用 EAR、MAR 算法对驾驶员眼部和嘴部几何特征进行分析,计算实时眼睑和嘴部纵横比,并进行坐标系转换估计驾驶员头部姿态。因此,本文是通过实时输出驾驶员眼部、嘴部疲劳特征参数和头部姿态欧拉角,使用多个疲劳特征参数结合疲劳判定标准对驾驶员的疲劳状态进行判定。

(3)建立了真实驾驶环境下的疲劳模拟数据集。相较于人脸表情数据集,目前现有公开的针对疲劳驾驶的数据集较少,本文通过招募人员,对司机剥夺睡眠后在真实驾驶环境下进行模拟疲劳驾驶,并通过硬件设备采集建立模拟疲劳视频数据集,为后续模型的训练和实验验证提供数据基础。

(4)提出了一种利用深度残差神经网络提取驾驶员眼部疲劳特征的方法,并利用迁移学习算法对该网络进行训练。本文利用深度残差神经网络提取出驾驶员眼部疲劳特征,由于存在残差单元,网络不会随着深度的增加而导致梯度消失或退化,使用这一网络对驾驶员复杂的眼部疲劳特征进行深层次提取,相较于传统网络有巨大优势。深度神经网络模型最终的优劣取决于网络的训练是否具有针对性,目前的疲劳特征提取网络大多采用人脸数据集而非真实环境下的疲劳数据集进行训练,针对性低。本文提出采用迁移学习算法并使用真实场景下的疲劳数据集进行深度神经网络模型训练,经过该方法训练的网络模型更具针对性、精度更高。

(5)提出了基于支持向量机 SVM 的驾驶员个性化疲劳阈值辨识算法。通常情况下,一辆车有 1 至 3 名驾驶员,目前的疲劳检测方法对疲劳阈值的设定固定且单一,无法在实际场景中满足不同驾驶员的疲劳辨识需求。本文提出的基于 SVM 的疲劳辨识算法,具有针对不同驾驶员训练的个性化疲劳阈值,经试验证明,与单一疲劳阈值和平均疲劳阈值的疲劳辨识算法相比,该算法对不同驾驶员的疲劳辨识具有较高的精确度。

1.4 论文组织结构

第一章:绪论。本章首先通过查阅和分析了全球以及我国的道路交通统计数据,介绍了疲劳驾驶的研究背景及其意义。然后,通过阅读大量相关文献,总结并比较了目前四类常见方法的优缺点,结合国内外研究现状提出本文的研究方法。最后介绍了本文的主要研究内容、创新性和篇章结构。

第二章:相关技术理论。本章主要阐述了深度学习和迁移学习算法等理论基础。在深度学习理论中,介绍了卷积神经网络局部感知和权值共享的特点,详细探究了卷积层、池化层等网络层级的原理及其之间如何组合共同构成卷积神经网络,并且还具体描述了网络的训练过程以及 AlexNet、VGGNet 等常见的卷积神经网络。最后,本章还介绍了迁移学习算法的定义、分类以及与深度学习相结合的发展与使用场景。

第三章:驾驶员人脸检测和头部姿态估计。本章主要进行了驾驶员人脸检测、面部关键点定位和头部姿态估计这三方面的研究工作,为特征提取奠定基础。首先在人脸检测中使用了基于 HOG 的人脸检测算法,并对其原理和算法实现流程进行了详细阐述。

然后在人脸关键点定位中使用目前较为先进的 ERT 算法，精确定位了驾驶员在不同角度、不同表情下的面部关键点。最后在头部姿态估计方面，利用相机标定实验和坐标系转换搭建了二维空间与三维空间的映射关系，通过 ERT 算法输出的二维面部关键点坐标联合相机内部参数矩阵求解驾驶员的头部姿态欧拉角。

第四章：基于面部信息和头部姿态的驾驶员疲劳特征提取。本章以驾驶员多种疲劳特征提取作为主要工作，通过提取眨眼、哈欠和低头等特征为后续疲劳判定提供数据基础。在眼部疲劳特征提取方面，基于 EAR 和深度残差神经网络结合的方法对眨眼进行检测，并使用迁移学习算法对残差网络进行训练。在嘴部特征提取上，运用 MAR 算法实现了驾驶员嘴部状态的检测。在头部姿态特征方面，基于国际 P80 标准对实时估计的头部姿态进行低头和异常角度检测。

第五章：基于多特征的驾驶员疲劳状态检测方法。本章首先介绍了基于多特征的驾驶员疲劳检测方法的总体结构、各部分功能以及实现流程。然后，建立了疲劳数据集并且制定了本文疲劳检测所使用的眼部、嘴部和头部姿态的疲劳判定标准。最后，基于 SVM 算法建立了针对不同驾驶员个性化疲劳阈值模型，通过实验对比分析证明了本文所提出的疲劳检测方法具有高精度、高鲁棒性，且不因某一特征检测受阻而失效等优点。

第六章：总结与展望。总结全文的研究工作内容，提出不足以及后续的改进方向。

第二章 相关技术理论

在驾驶员疲劳特征提取中，本文使用了深度学习算法与机器学习相结合的方法，并运用了迁移学习算法对驾驶员疲劳特征提取模型进行训练。本章对相关算法以及理论进行具体阐述，为后文模型的建立奠定理论基础。

2.1 深度学习理论

2.1.1 深度学习理论基础

深度学习是人工智能领域的重要成果之一，作为由机器学习发展而来的新技术，在语音图像识别、机器视觉等领域中获得了成功的应用。相比于通常只有 1 到 2 层的隐性马尔可夫模型、逻辑回归模型和高斯混合模型等浅层学习结构模型，深度学习可以简单理解成是深层次的神经网络模型。如图 2-1 所示，深度学习和传统的神经网络模型总体结构具有相似之处，都是由输入层、隐含层和输出层构成，但不同之处在于其有更多的隐含层，可以提取到更深层次的特征。

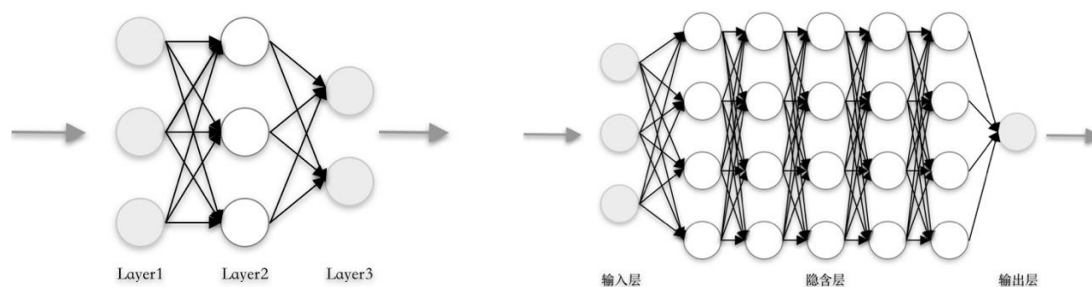


图 2-1 传统神经网络与深度学习

Fig.2-1 Traditional neural networks and deep learning

世界上第一个神经网络数学模型早在二十世纪四十年代就被提出^[40]，而深度学习却在 2006 年 Hinton 等人发表在《Neural Computation》^[41]和《Science》^[42]的两篇论文后才开始蓬勃发展，从此大量深度学习模型不胜枚举，其中包括强化学习网络（RLN）、卷积神经网络（CNN）、循环神经网络（RNN）和生成对抗网络（GAN）等。近年来，随着计算机计算能力的提高和高性能 GPU 的出现，加之有许许多多易于访问的带标签的数据集，卷积神经网络在图像识别和机器视觉中表现突出，本文将对其理论展开详细介绍。

2.1.2 卷积神经网络介绍与特点

2012 年 Hinton 教授以及他的学生凭借卷积神经网络 AlexNet^[43]在世界机器视觉领域的权威竞赛 ImageNet 远超第二名摘得桂冠引爆了卷积神经网络的热潮。如今，卷积神经网络在图片分类、姿态估计和目标检测等图像视频领域中取得了广泛且成功的应用。

传统的全连接神经网络在处理大像素图片时具有明显的缺陷，如：将图片转化为向量时丢失临近像素的空间特征信息、图片数量庞大导致参数过多进而影响训练效率、模型参数过多致使网络出现过拟合状态等。卷积神经网络相比于传统的神经网络，具有更少的权重参数和更快的计算速度，并且不会因为出现大量的参数而导致网络出现过拟合状态，其原因是该网络中具备局部感受野和权值共享的特点。

（1）局部感受野

“感受野”的概念最初由 Hubel 在对动物的视觉皮层的研究中提出^[44]。在卷积神经网络中，局部感受野就是神经元的局部连接。与传统神经网络的全连接不同，卷积神经网络中某一层的神元只与上一层的部分神经元相联系，如图 2-2 所示。局部连接的特点在网络训练中极大减少了连接数量、降低了网络的复杂度和计算量，也更加符合图像的特征规律。对于一张图像而言，图像中某一个局部的图像像素信息通常与其周围像素信息相关度高而与整个图像所有的像素信息相关性低，通过卷积神经网络局部连接的特点对图像进行局部感知最后再整体联系，即可得到图像的全局特征，相比于全局感知图像特征，提取效果更佳、计算更少。

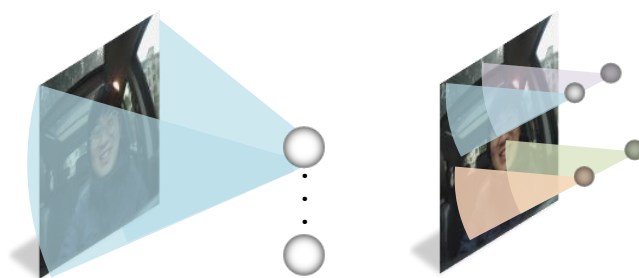


图 2-2 全连接与局部连接

Fig.2-2 Full connection and partial connection

（2）权值共享

卷积神经网络的权值共享是其具有图像平移不变性的原理体现。传统神经网络在特征提取时权重矩阵中的每个元素仅使用一次，而卷积神经网络对该图像全部区域仅使用单独的卷积核进行卷积，再使用不同的卷积核获取多种特征，如图 2-3 所示。基于此，无论图像如何进行平移变化依然可以提取到相同的特征，在赋予网络图像平移不变性的同时，大大降低了参数浪费，减小了模型复杂度和训练难度。



图 2-3 卷积神经网络权值共享

Fig.2-3 Convolutional neural network weight sharing

2.1.3 卷积神经网络的结构

标准的卷积神经网络通常由输入层、卷积层、池化层、全连接层和输出层构成，如图 2-4 所示。以下分别对不同网络层进行介绍。

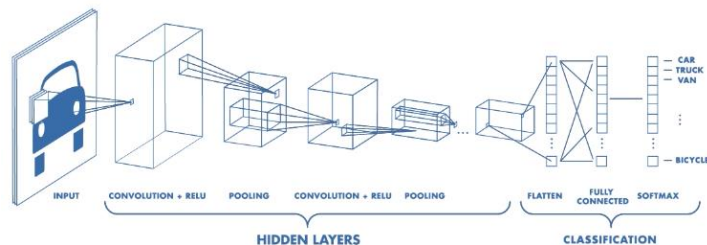


图 2-4 卷积神经网络结构图

Fig.2-4 Convolutional neural network structure diagram

（1）卷积层

卷积层是卷积神经网络中最具代表也是最核心的一个网络层级。卷积层通过卷积核对图像进行卷积计算提取图像特征，卷积的定义见式（2-1）：

$$s(i_c, j_c) = (X_c * W_c)(i_c, j_c) = \sum_{m_c} \sum_{n_c} x(i_c + m_c, j_c + n_c) w(m_c, n_c) \quad (2-1)$$

式中， W_c 为卷积核， m_c 、 n_c 为卷积核的维度。具体卷积过程如图 2-5 所示，假设样本大小为 5×5 ，卷积核大小为 3×3 ，卷积核在 5×5 样本的区域依次平移一个像素点，并将对应区域乘积之和输出为 3×3 的特征图。

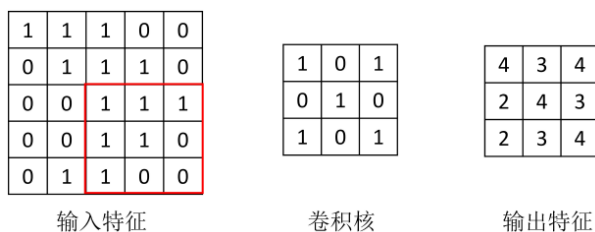


图 2-5 卷积过程示意图

Fig.2-5 Schematic diagram of the convolution process

一般情况下，卷积层中的卷积核与图像像素进行卷积计算后需要经过激活函数将特征进行非线性化，再将激活函数的输出作为卷积层的输出。

（2）激活函数

激活函数也称之为非线性处理单元。网络完成卷积层的线性运算后，需要经过激活函数增加非线性拟合能力，其中常见的激活函数有如下几种（见表达式 2-2 至 2-4），其函数图像见图 2-6：

Sigmoid 函数表达式为：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2-2)$$

Tanh 函数表达式为：

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (2-3)$$

ReLU 函数表达式为:

$$f(x) = \max(0, x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (2-4)$$

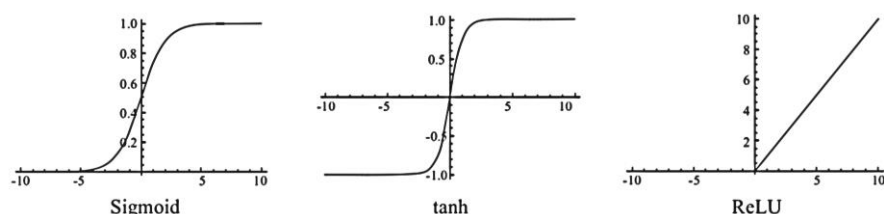


图 2-6 激活函数图

Fig.2-6 Activation function graph

由图 2-6 和表达式可知, sigmoid 函数和 tanh 函数的计算和求导都是指数运算, 存在梯度饱和问题。以 sigmoid 函数为例, 该函数可将输入映射到 $[0, 1]$ 区间内, 当输入由 0 趋近正负无穷时, 函数梯度快速消失并趋近于 0, 在进行反向传播时, 误差无法传递到前层, 导致无法完成深层次的网络训练。随着深度学习的发展, sigmoid 这类函数逐渐不能满足深层次网络的需求。基于此, 修正性单元 (Rectified Linear Unit, ReLU) 被引入到神经网络中, 由表达式可知, 该函数为分段函数, 在输入大于 0 时不会出现梯度饱和和效应, 适合作为深度神经网络中的激活函数。

(3) 池化层

卷积神经网络中, 图像经过卷积层虽然可以大幅度减少参数量、降低维度, 但仅仅依靠卷积层进行降维, 不仅降维速度缓慢, 而且会由于大量的卷积层堆叠带来过拟合问题。因此, 在卷积神经网络中引入了池化层, 池化层也称下采样层, 常伴随卷积层之后起到提高计算效率、防止训练过拟合等作用。池化的核心思想是降维, 其分为最大池化层 (Max pooling) 和平均池化层 (Average pooling), 最大池化层是通过在目标区域内, 选取其最大值作为其整个目标区域的代表值。平均池化层则是选取该目标区域的平均值作为代表值进行输出, 起到降低数据维度的目的。以最大池化为例, 其具体运算过程如图 2-7 所示, 在 4×4 的样本区域内, 使用步长为 2, 即 2×2 的目标区域进行最大池化操作, 将每个目标区域的最大值进行输出, 最终得到 2×2 的一个区域, 该区域表征了原区域的特征并降低了特征维度。



图 2-7 最大池化过程

Fig.2-7 Maximum pooling process

(4) 全连接层

全连接层可以视为传统网络中的隐含层，通常在卷积神经网络的最后部分承接输出，对前层特征进行融合并将其转变为一维特征向量。全连接层中的神经元前后之间全部相联系，目的是完整无遗漏地提取到前层卷积层和池化层的特征信息，其结构图如图 2-8 所示。

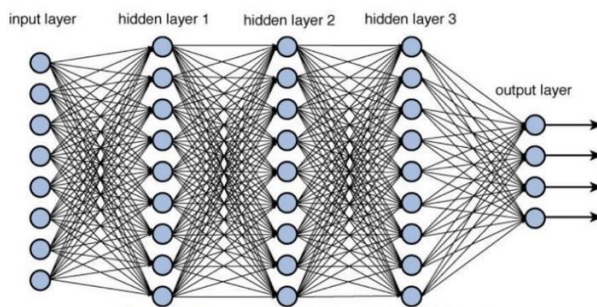


图 2-8 全连接层结构图

Fig.2-8 Connection layer structure diagram

(5) 输出层

输出层在卷积神经网络中的作用是输出分类结果。在进行实际的图像多分类任务时，通常在网络末位的全连接层后加上非线性激活函数用于最后的分类，一般由 softmax 函数实现。网络经过 softmax 分类器后输出各样本类别的预测概率，softmax 函数定义如下：

$$S_{soft} = \frac{e^{V_z}}{\sum_j e^{V_j}} \quad (2-5)$$

式中， S_{soft} 表示当前元素的指数与所有元素指数之和的比， V_z 表示 softmax 分类器输出， z 表示类别索引， C 为总类别个数。在网络训练过程中，输出层发挥着至关重要的作用，输出层将网络的分类预测结果与标签结果相比较，计算误差并进行反向传播更新权重。

2.1.4 卷积神经网络训练过程

卷积神经网络在进行特征训练时，一般要经过前向传播和反向传播两个阶段。其中前者是图像由网络的输入层到卷积层、池化层，最后由输出层的 softmax 分类器输出类别的预测概率这一过程。反向传播过程是一个有监督的学习过程，即在卷积神经网络训练过程中，带标签的训练数据经过前向传播阶段从输入层到输出层输出分类结果并与标签值进行比较，最后将误差进行反向传递到前层神经元并修正权重。反向传播算法本质上是一种递归梯度下降算法。

梯度下降算法 (GradientDescent Optimization) 是常用的模型最优化方法之一，又称最速下降算法，由 Cauchy 于 1874 年提出，其核心是用负梯度方向作为下降方向。设 $f(x)$ 在 x_k 连续且可微，令 $x = x_k + \tau d$ ，其中 $\|d\|=1$ ，若 $h_k = \nabla f(x_k) \neq 0$ ，则泰勒展开为：

$$f(x) = f(x_k) + (\nabla f(x_k))^T (x - x_k) + o(\|x - x_k\|) \quad (2-6)$$

简化为：

$$f(x_k + \tau d) = f(x_k) + \tau h_k^T d + o(\tau), \tau > 0 \quad (2-7)$$

设 ω 为 d 与 $-h_k$ 间夹角，则有 $h_k^T d = -h_k \cos \omega$ ，由上式可见，当 $\omega = 0$ 时， $\cos \omega = 1$ ， $h_k^T d$ 值最小，因此 $f(x)$ 下降最快，负梯度方向 $-h_k$ 便是 $f(x)$ 在 x_k 附近下降最快的方向。

反向传播算法（Backpropagation algorithm）是神经网络进行误差传递学习的一种重要算法，其通过梯度下降的方式由预测结果与真实结果之间的差值向前层神经元更新权重使其不断逼近真实值。设样本数据集为 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，经梯度下降法求解其整体代价函数可得：

$$J(D, b) = \left[\frac{1}{n} \sum_{i=1}^n \frac{l}{2} \|h_{D,b}(x^{(i)}) - y^{(i)}\|^2 \right] + \frac{l}{2} \sum_{l=1}^{ml-1} \sum_{i=1}^{Sl} \sum_{j=1}^{Sl+1} (D_{ji}^{(l)})^2 \quad (2-8)$$

其中参数 D 、 b 按如下公式进行迭代更新，式中 α_l 为学习率：

$$D_{ij}^{(l)} = D_{ij}^{(l)} - \alpha_l \frac{\partial}{\partial D_{ij}^{(l)}} J(D, b) \quad (2-9)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha_l \frac{\partial}{\partial b_i^{(l)}} J(D, b) \quad (2-10)$$

2.1.5 常见卷积神经网络

AlexNet: AlexNet 是在 2012 年“深度学习之父”Hinton 教授与其学生 Alex Krizhevsky 提出的卷积神经网络模型。该模型是首个在 ILSVRS 比赛中获得冠军的卷积神经网络模型，首次向世界证明由神经网络模型自身学习到的特征可以超越人工提取的特征，为后来卷积神经网络的发展奠定了基础。AlexNet 的网络结构如图 2-9 所示：

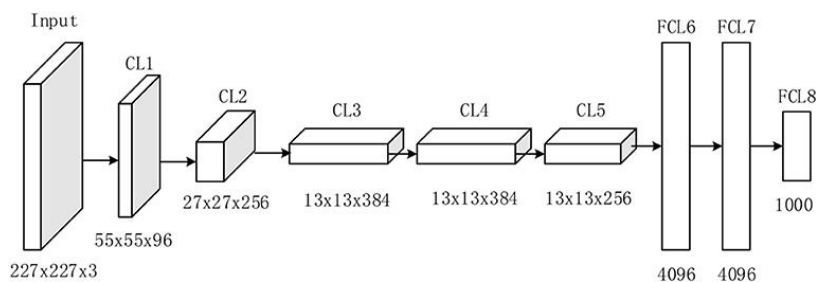


图 2-9 AlexNet 网络结构

Fig.2-9 AlexNet network structure

由图 2-9 可知，AlexNet 主要由 5 个卷积层和 3 个全连接层组成，其中还包括 3 个最大池化层对特征进行降维，网络的输入为 227×227 RGB 颜色通道的图像，输出为经过 softmax 分类器分类的 1000 个类别的图像概率。AlexNet 在卷积神经网络中取得突破性优势的原因在于：1、相较于以往具有更深层的网络结构。2、使用了 ReLU 函数作为

激活函数。3、使用多种方法防止过拟合。（如：对训练数据随机裁剪以增强数据、使用 dropout 的方式强迫网络学习出更稳定的特征。）

VGGNet: VGGNet 是由 University of Oxford 与 Google DeepMind 中的专家学者在 2014 年共同研发的卷积神经网络模型^[45]，同年该模型取得了 ILSVRS 竞赛的亚军和定位项目的冠军。VGGNet 模型包含两种结构，分别是 VGG16 和 VGG19，其中 VGG16 的网络结构如图 2-10 所示：

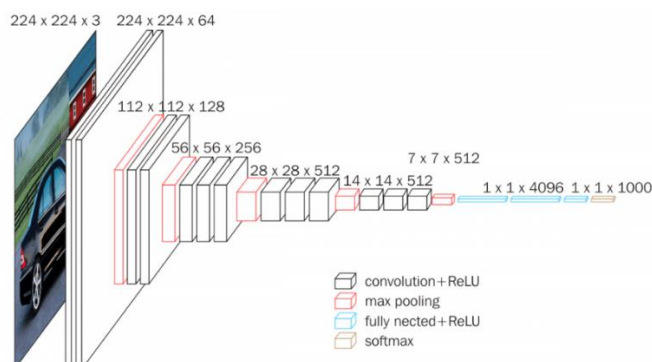


图 2-10 VGG16 网络模型

Fig.2-10 VGG16 network structure

由图 2-10 可知，VGG16 网络的拥有 12 层卷积层和 5 个最大池化层，相比于 AlexNet 拥有更深的网络结构和更多的 ReLU 函数。其创新性在于 VGG 网络重复使用较小的 3×3 卷积核堆叠来构造更深层次的卷积神经网络，而使用较小的卷积核相对于大卷积核不仅能够获取到更深层次的特征，还降低了参数计算量。

GoogLeNet: GoogLeNet^[46]是 Google 基于 inception 模块推出的深度神经网络模型，2014 年在 ILSVRS 竞赛中大放异彩取得冠军。以往的 AlexNet、VGGNet 都是通过增加网络深度来提高分类的准确率，但是复杂的网络往往会由于大量卷积核参数没有完全利用而浪费大量计算资源，而 GoogLeNet 通过创新性提出 inception 块带来了另一种提升网络精度的思想。图 2-11 为简单 inception 块结构：

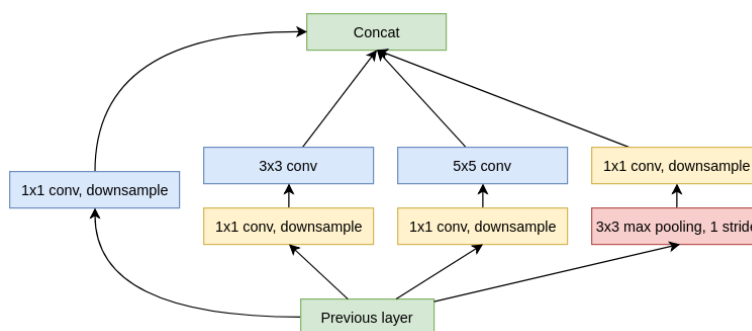


图 2-11 inception 结构图

Fig.2-11 inception structure diagram

由图 2-11 可知，inception 块通过 4 条由不同大小的卷积层和池化层组成的并行通道对图像进行特征提取。GoogLeNet 将多个卷积层和池化层通过一定的方式组成一个模块，即 inception 块，然后再通过组合各种模块构造网络。

2.2 迁移学习

2.2.1 迁移学习定义

迁移学习是机器学习算法中发展出来的一个重要分支，是将知识经验从源域向目标域进行迁移的一种方法^[47]。迁移学习就是要找到原任务与新任务之间的联系，并通过将经验知识迁移，利用模型的迁移特性以达到跨主题实现预测任务的目的。迁移学习本质是利用拥有足够标签样本的源域数据知识，来训练缺少标签或无标签样本的目标域，获得泛化性优秀的目标域模型。传统的机器学习与迁移学习的原理如图 2-12 所示。

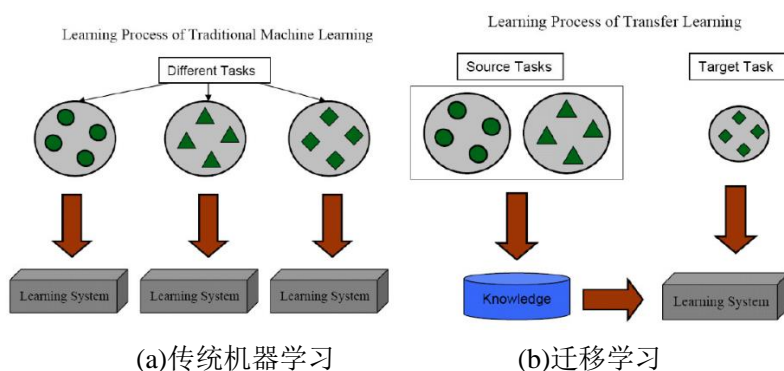


图 2-12 传统机器学习与迁移学习的原理对比

Fig.2-12 Comparison of the principles of traditional machine learning and transfer learning

由图 2-12 可知，传统的机器学习训练一个用于目标域的模型需要满足：1.足够多带标签的样本训练数据；2.训练数据域的特征空间与目标域一致；3.对每一个预测目标需要重新训练预测模型。而迁移学习可以高效地利用原域知识，并从中挑选出适合目标域的经验知识，在目标域建立泛化性好的迁移学习模型。其既不需要大量与目标域有相同特征空间的标签训练数据，也不需要针对每一个不同的预测任务从零开始构建模型，大大提高了模型的泛化性，节省了成本和计算资源。

2.2.2 迁移学习分类

按历史发展的顺序将迁移学习进行分类，可以分为基于机器学习和基于深度学习的迁移学习算法，按照迁移学习方法可分为基于样本、特征、参数和模型的迁移学习算法，下面对各类迁移学习算法进行简单介绍。

(1) 基于样本的迁移学习算法

基于样本的迁移学习算法就是将源域的样本特征空间经过一定的权重处理，对与目标域样本特征空间相近的样本赋予高权重，以此来筛选出与目标域数据特征更为相似的数据样本的算法^[48]。该类算法适用于目标域和源域数据特征较为相近的预测任务。

（2）基于特征的迁移学习算法

基于特征的迁移学习算法可以简单理解为对样本特征进行变换以满足不同域之间的特征需求。其原理是将源域和目标域的样本特征共同映射到其他特征空间，或者将源域的样本特征映射到目标域的特征空间中去，最终实现对目标域的分类预测。

（3）基于参数的迁移学习算法

基于参数的迁移学习算法通过模型共享参数和先验分布，在源域和目标域中找到共同的参数和先验知识，达到知识迁移的目的，该类方法迁移的知识是模型的参数而非数据本身，对源域数据有一定的保护^[49]。

（4）基于模型的迁移学习算法

基于模型的迁移学习算法一般运用在神经网络模型上，利用神经网络封装的特性，将源域已经学习到知识的模型迁移到目标域内，再结合目标域的样本特征进行调整学习得到新的模型^[50]。对于一个新任务而言，利用模型的迁移学习算法不需要从零开始训练搭建网络模型，运用已有的高精度模型进行迁移训练，不仅节省了大量的计算成本，更为重要的是给一些缺少标签样本的目标域提供了一种新的解决方法。

近年来，随着深度学习神经网络的不断发展，有越来越多在源域经过大量数据和深层次训练的高精度、高鲁棒性模型，得益于深度神经网络结构的高效性和封装性，使得模型不仅可以获得表征能力突出的迁移特性，还能满足实际任务中端到端的开发需求。迁移学习利用深度神经网络的结构优势和迁移学习算法赋予模型的泛化能力，不仅节省了在实际任务中的训练模型的时间、计算成本，还解决了小样本目标域难以使用高精度深度神经网络模型的难题，是未来人工智能发展的重要方向之一。

2.3 本章小结

本章对深度学习和迁移学习的相关理论进行了详细介绍，首先介绍了卷积神经网络局部感知和权值共享的特点，以及网络结构（卷积层、池化层、全连接层）和网络的前向、反向传播的训练过程，然后简单介绍了 AlexNet、VGGNet 等常见的卷积神经网络模型。最后，阐述了迁移学习算法的原理定义以及其分类和使用场景，为后文疲劳检测模型的建立奠定理论基础。

第三章 驾驶员人脸检测和头部姿态估计

基于机器视觉对驾驶员进行疲劳特征提取并判断是否疲劳，首要任务是对驾驶舱内驾驶员的人脸进行快速准确的检测，然后将检测到的人脸进行眼睛、嘴巴等面部 2D 关键点的定位，最后将驾驶员 2D 面部特征信息映射到 3D 空间上建立驾驶员头部姿态运动估计模型，为后续进行驾驶员眼部、嘴部和头部姿态的疲劳特征提取做准备。

3.1 基于 HOG 的驾驶员人脸检测算法

在对驾驶员进行疲劳检测的过程中，对于由机器视觉构建的疲劳检测系统，首先要获取带有驾驶员人脸的图像，才能进行后续的面部关键点和驾驶员疲劳特征提取。因此使用一种快速且准确的人脸检测算法对于整个疲劳检测系统而言至关重要。

3.1.1 HOG 算法原理

对由摄像头传递到疲劳检测系统的视频图像，利用人脸检测算法返回驾驶员人脸区域位置坐标信息是研究驾驶员疲劳检测方法的首要任务。现有的人脸检测方法主要分为基于特征、基于知识、基于模板匹配和基于统计学习的人脸检测算法，其优缺点对比如表 3-1 所示：

表 3-1 常用的人脸检测算法对比

算法	原理	优点	缺点
基于特征的方法	基于面部不变特征（如：肤色特征）实现人脸检测	具有旋转和尺度不变形	光照、姿态和噪声会破坏图像特征，影响检测准确率
基于知识的方法	基于面部器官的位置分布，利用先验知识制定检测规则	具有简单快速的优点	对先验知识要求高、先验知识和规则定义不满足所有人脸
基于模板匹配的方法	通过预先设立模板，将输入图像与模板进行阈值匹配	开发成本低，算法易实现	不能有效处理尺度、姿态和形态变化，计算资源需求高
基于统计学习的方法	通过将人脸检测视作模式识别问题，使用正负样本进行训练	无需先验知识、实时性高、精度高	需要大量样本进行训练，开发难度大

通过对比分析常用的人脸检测算法，本文选择使用兼备实时性与检测精度的基于统计学习的人脸检测算法作为疲劳驾驶检测的人脸检测算法。其中，在该类人脸检测算法中，Dalal N 在 CVPR 会议中提出的方向梯度直方图算法（Histogram of Oriented Gradients,HOG）是一种兼备快速性和高精度的目标检测算法^[51]。利用梯度和边缘方向密

度分布可以较为容易表征物体边缘形状这一特性，HOG 算法通过表征图像局部纹理特征计算图像局部像素点的梯度和边缘方向直方图，再将直方图组合建立特征表述器以表征待测物体的特征。相比于其他图像检测算法，HOG 算法采用细胞单元 cell 的方式对图像进行梯度方向量化，以表征物体边缘的结构特征。由于算法在位置和方向空间上进行量化，减少了物体旋转平移的影响，并且由于物体的梯度信息不易受实际环境光照的影响和算法对计算机资源消耗小，所以运用在光照变化较大且算力较小的车载驾驶员疲劳检测系统中较为合适。

3.1.2 基于 HOG 算法的人脸检测实现流程

基于 HOG 算法的人脸检测实现分为四步，具体过程如图 3-1 所示：

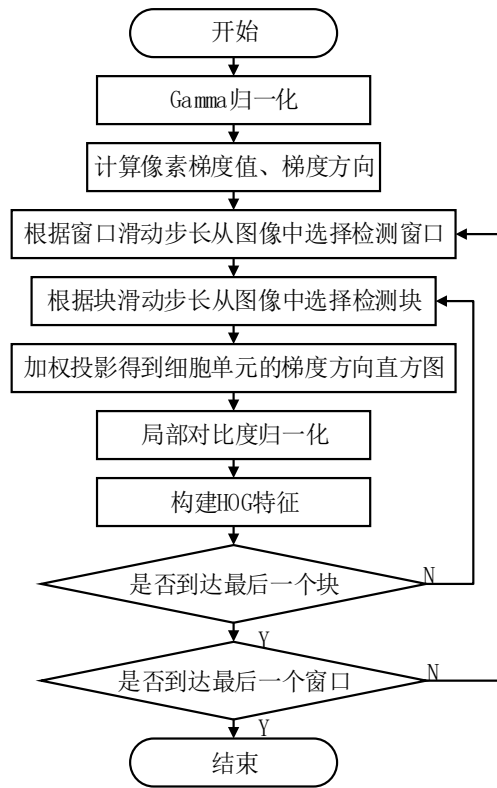


图 3-1 HOG 算法流程图

Fig.3-1 HOG algorithm flow chart

(1) 图像归一化

由于摄像头输入的图像大多为 RGB 颜色空间，特征维度高，不利于算法直接处理，首先需要将其按式 (3-1) 转化为灰度图。然后利用 Gamma 算法对图像进行归一化压缩处理以减少实际驾驶环境场景中光照、颜色和阴影的影响，其公式如式 (3-2) 所示，式中 (x, y) 为图像中的像素点。

$$Gray = 0.299R + 0.587G + 0.114B \quad (3-1)$$

$$I(x, y) = I(x, y)^{gamma} \quad (3-2)$$

(2) 计算图像像素梯度

由于物体形状边缘的梯度变化大，故想要获取物体轮廓信息，只需遍历计算像素梯度值 $G(x, y)$ ，式 (3-3) 和 (3-4) 为水平和垂直方向的梯度计算公式。

$$G_x(x, y) = H(x+1, y) - H(x-1, y) \quad (3-3)$$

$$G_y(x, y) = H(x, y+1) - H(x, y-1) \quad (3-4)$$

式中， $H(x, y)$ 为图像的像素值。遍历计算获取到像素点的 $G_x(x, y)$ 和 $G_y(x, y)$ 后，利用式 (3-5) 和 (3-6) 求解像素点梯度值 $G(x, y)$ 和梯度方向 $\alpha(x, y)$ 。

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (3-5)$$

$$\alpha(x, y) = \tan^{-1} \left[\frac{G_y(x, y)}{G_x(x, y)} \right] \quad (3-6)$$

(3) 构建梯度直方图

将输入图像平均分割成若干个相同大小的小区域也称 cells 单元，每个区域的像素为 $n \times n$ ，再统计每个 cells 的梯度直方图，具体做法是：将 $[0^\circ, -180^\circ]$ 或 $[0^\circ, -360^\circ]$ 平均为 bin 份，然后将像素值的模长放入对应角度值的 bin 区域内，遍历完 cells 的像素点后，将 bin 区域内的模长相加，便可以求得 cells 的特征向量。最后将每个 cells 在根据式 (3-5) 和 (3-6) 为其所属方向区间加权投影，生成梯度方向直方图。若 $\text{bin}=8$ ，即将 360° 分为 8 个区域，每个区域为 45° ，如图 3-2 所示，图中 A、B、C 为 cell 内三个像素点。

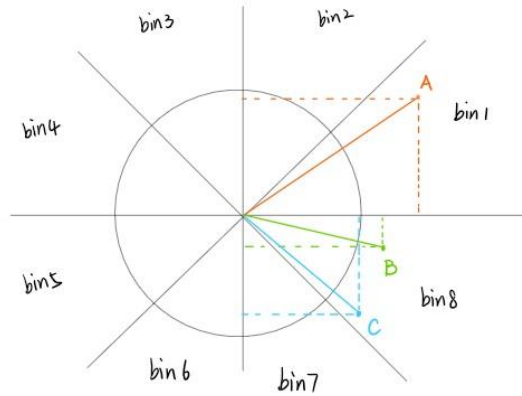


图 3-2 梯度直方图分割计算

Fig.3-2 Gradient histogram segmentation calculation

(4) 梯度直方图合并与归一化

由于驾驶环境中前、背景光照对比度大，为了降低梯度强度变化过大的问题，将若干 cells 单元合并，组成互通的大单元块 blocks，如下图所示，再对每一个 blocks 中的梯度信息作归一化处理以提升算法性能。

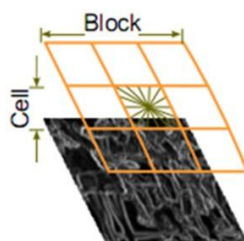


图 3-3 合并 cells 单元

Fig.3-3 Merge cells

最后，利用单元块作为滑动窗口，扫描摄像头输入的驾驶舱图像，提取 HOG 特征，并将其输入至分类器。在输出的众多人脸候选框中使用非极大值抑制（NMS）算法进行处理，最后得到人脸候选框如图 3-4 所示：

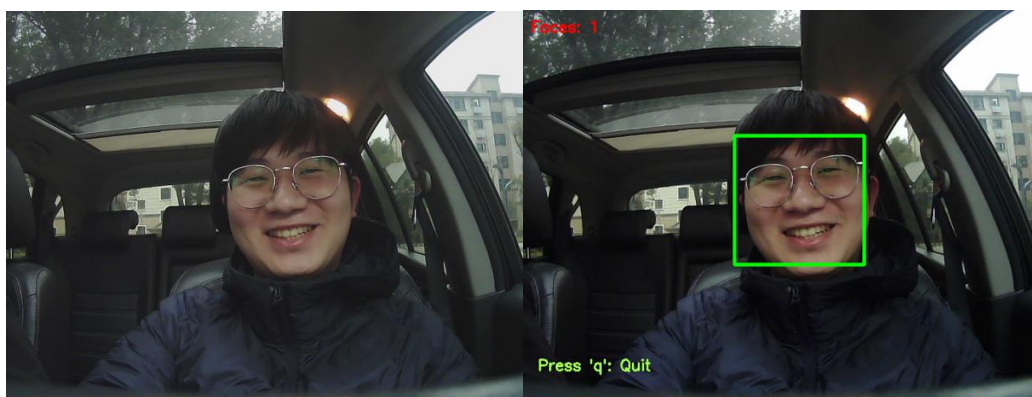


图 3-4 基于 HOG 算法人脸检测实验结果

Fig.3-4 Experimental results of face detection based on HOG algorithm

图 3-4 为测试驾驶员在实际驾驶环境中在摄像头实时拍摄的视频中截取的一帧图像。由实验结果可知，HOG 算法在实际驾驶环境对驾驶员的人脸检测拥有较好的实时性和准确度，在实际光线不足的情况下，也能对驾驶员进行准确的人脸检测。

基于 HOG 的人脸检测算法是根据车载摄像头实时传输视频帧中的局部区域的梯度统计信息来表征人脸特征的。在实际驾驶环境中，驾驶员位置相对固定，人脸处于采集图像的中间位置，HOG 算法不会因为驾驶员面部表情和驾驶姿态的变化而频繁丢失检测目标，所以在疲劳驾驶检测领域采用 HOG 特征的人脸检测算法具有较为优秀的检测精度。并且，由于 HOG 算法的主要原理是对图像的梯度进行计算，其中大多为线性运算，时间复杂度低，检测实时性好，更适合在 CPU 算力较小的车载系统作为人脸检测算法。

3.2 基于级联回归树的驾驶员人脸关键点检测算法

人脸关键点检测是机器视觉中描述人脸图像面部几何特征的一种技术，通过给定具有人脸信息的图像，人脸关键点检测算法根据目标需求，识别并定位出如眼睛、嘴巴、鼻子等面部几何特征信息。本节在 HOG 人脸检测算法输出的具有人脸范围信息图像的

基础上,使用基于级联回归树算法定位出驾驶员人脸关键点信息,为后续疲劳特征的提取奠定基础。

3.2.1 级联回归树人脸关键点检测算法原理

人脸关键点定位通常是在人脸检测的基础上,对人脸图像进行目标特征点的坐标定位,将输入的具有人脸信息的图像,快速准确地识别并定位出面部的眼睛、眉毛、鼻子、嘴巴等面部特征信息,输出坐标集合。目前,人脸关键点检测方法主要有:主动形状模型^[52]、CNN 检测算法、主动表现模型^[53]和基于回归的检测算法。表 3-2 为不同算法在相同测试图像集的检测速度和精度的横行对比^[54]。

表 3-2 不同人脸对齐算法对比

算法	ASM	AAM	STASM	EGM	SDM	ESR	ERT	LBF
速率 (fps)	≈1	≈1	15	<1	70	120	1000	3000
平均误差 (像素)	8.7	9.5	11.1	3.98	5.6	3.47	4.74	4.95

由表可知,ERT 算法满足实时检测要求且精度较高。因此,本文选择采用基于级联回归树(Ensemble of Regression Trees,ERT)的人脸关键点定位算法,ERT 算法由 V. Kazemi 和 J. Sullivan 在 CVPR 中提出^[55],其原理是通过构建梯度提升树(Gradient Boost Decision Tree,GBDT)来不断学习回归树叶子节点残差,逐渐将人脸候选框中的人脸形状拟合成真实形状,最后得到人脸关键点信息。

与传统算法相比,ERT 不仅可以解决在训练过程中出现的标签缺失或不确定等问题,还可以在执行形状不变特征选择的同时,最大范围地减少训练时相同的损失函数。ERT 算法在学习回归树的过程中,每一颗树都是建立在前一颗树上的串行关系,每颗树的叶子节点存在残差回归量,人脸形状残差将存入叶子节点中,通过不断学习更新得到人脸关键点信息,图 3-5 为人脸关键点标记信息。

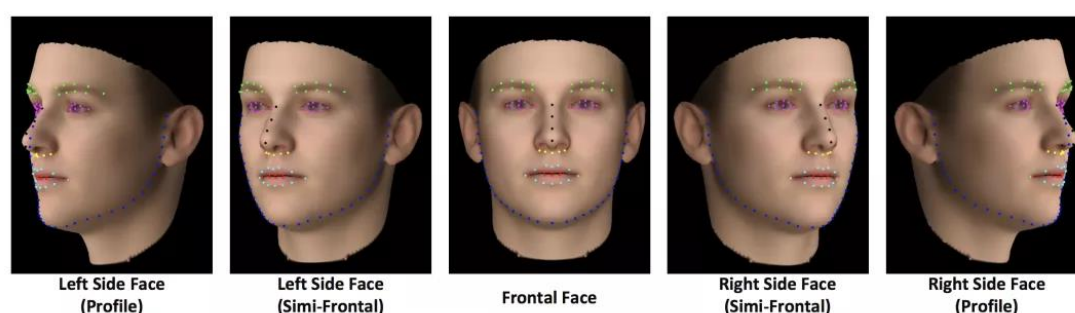


图 3-5 人脸关键点标记信息

Fig.3-5 Face key point mark information

3.2.2 建立级联回归器

基于 ERT 算法的人脸对齐步骤是先建立级联回归器，设 $x_i \in R^2$ 为人脸图像 I 上面部关键点 i 的坐标， $S = (x_1^T, x_2^T, \dots, x_p^T) \in R^{2p}$ 为图像 I 中 p 个面部关键点坐标组成的形状向量，其人脸形状回归计算如下：

$$\hat{S}^{(t+1)} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)}) \quad (3-7)$$

式 (3-7) 中， $\hat{S}^{(t)}$ 为当前人脸形状索引，即对 S 的估计， $r_t(I, \hat{S}^{(t)})$ 表示回归器， $\hat{S}^{(t+1)}$ 表示更新的形状。在对驾驶员进行人脸关键点定位时，用 $\hat{S}^{(0)}$ 平均初始人脸形状，回归器 r_t 根据图像 I 和 $\hat{S}^{(t)}$ 索引计算特征值预测级联的临界点。

3.2.3 基于树的回归器

(1) 训练级联结构的每个回归器

第 i 张人脸图像为 I_i ，图像的形状向量为 S_i ，将 $(I_1, S_1), (I_2, S_2), \dots, (I_n, S_n)$ 作为回归器的训练数据，设输入参数 $\{(I_i, \hat{S}_i^{(t)}, \Delta S_i^{(t)})\}_{i=1}^N$ ，对回归器进行下式初始化：

$$f_0(I, \hat{S}^{(t)}) = \arg \min_{\gamma \in R^{2p}} \sum_{i=1}^N \|\Delta \hat{S}^{(t)} - \gamma\|^2 \quad (3-8)$$

$$\Delta S_i^{(t)} = S_i - \hat{S}_i \quad (3-9)$$

回归器中 r_t 按照式 (3-10) 迭代，同时更新 $f_k(I_i, \hat{S}_i^{(t)})$ 如下：

$$r_{ik} = \Delta S_i^{(t)} - f_{k-1}(I_i, \hat{S}_i^{(t)}) \quad (3-10)$$

$$f_k(I_i, \hat{S}_i^{(t)}) = f_{k-1}(I_i, \hat{S}_i^{(t)}) + g_k(I, \hat{S}^{(t)}) \quad (3-11)$$

其中， k 为 r_t 的训练迭代次数， $g_k(I, \hat{S}^{(t)})$ 表示弱分类器的回归方程，最终得到回归器表达式如下：

$$r_t(I, \hat{S}^{(t)}) = f_k(I_i, \hat{S}_i^{(t)}) \quad (3-12)$$

(2) 形状不变的节点分裂

训练好基于树的回归器后，在决定叶子节点分裂前，需要设定一个分裂阈值，计算图像在特征池中的一对像素点的像素值和像素差，并与分裂阈值进行对比，经过阈值判断进行节点分裂，具体如下：

设两个像素点的位置分别为 u 和 v ， k_u 为平均形状中面部关键点索引， k_u 与 u 的偏移量计算公式如下：

$$\delta x_u = u - \bar{x}_{k_u} \quad (3-13)$$

S_i 为图像 I_i 的形状，在图像中与点 u 定性相似的点的位置可由式 (3-14) 求解得到：

$$u' = x_{i, k_u} + \frac{1}{S_i} R_i^T \delta x_u \quad (3-14)$$

式中, s_i 是 S_i 变换到平均形状 \bar{S} 的比例, R_i 为旋转矩阵, 通过式 (3-15) 可以将其最小化。

$$\sum_{j=1}^p \|\bar{x}_j - (s_i R_i x_{i,j} + t_j)\|^2 \quad (3-15)$$

形状不变的节点分裂公式如下, 式中 $\theta = (t_{sp}, u, v)$, t_{sp} 为分裂阈值。

$$h(I_{\pi i}, \hat{S}_i^{(t)}, \theta) = \begin{cases} 1 & I_{\pi i}(u') - I_{\pi i}(v') > t_{sp} \\ 0 & otherwise \end{cases} \quad (3-16)$$

(3) 节点分裂

在训练时, 对每个节点随机生成候选分裂阈值集合, 再通过最小化平方误差之和, 计算出最优的节点分裂阈值 θ^* , 其计算公式如下:

$$E(Q, \theta) = \sum_{s \in \{l, r\}} \sum_{i \in Q_{\theta, s}} \|r_i - \mu_{\theta, s}\|^2 \quad (3-17)$$

式中, Q 为节点的训练样本, r_i 为图像通过梯度增加算法计算的所有残差向量, $\mu_{\theta, s}$ 的计算如下:

$$\mu_{\theta, s} = \frac{1}{|Q_{\theta, s}|} \sum_{i \in Q_{\theta, s}} r_i \quad s \in \{l, r\} \quad (3-18)$$

忽略部分因素 θ , 将上式重新整理可得到最佳分裂阈值的计算公式:

$$\arg \min_{\theta} E(Q, \theta) = \arg \max_{\theta} \sum_{s \in \{l, r\}} |Q_{\theta, s}| \mu_{\theta, s}^T \mu_{\theta, s} \quad (3-19)$$

最后利用式 (3-20) 优先寻找距离相近的一对像素点来选择每一个节点。

$$P(u, v) \propto e^{-\lambda_p \|u - v\|} \quad (3-20)$$

3.2.4 实验结果与分析

基于 ERT 的人脸关键点检测算法, 通过建立梯度增加学习框架, 优化损失函数和误差总和来学习 Tree。最后由 HOG 人脸检测算法输出的人脸候选图像经过 ERT 人脸关键点预训练模型可以提取到关键点坐标。

预训练模型由 300W^[56]人脸基准数据集训练而成, 其中包括 AFW^[57]、HELEN^[58]、LFPW^[59]和 IBUG^[60]数据集。300W 数据集采用半监督标注的方式标注了人脸 68 个关键点坐标, 一共 3837 张人脸图像, 模型采用 3148 张作为训练, 689 张作为测试, 其中对训练集数据进行数据增强: 对人脸图像进行 -45° 到 45° 的随机旋转、对图像进行随机的中心裁剪、对图像的亮度、对比度、饱和度和色相做随机调整等。

最终在光线较暗的实际驾驶环境中, 由车载摄像头采集的人脸图像在经过 HOG 和 ERT 人脸检测和关键点定位算法输出的实时结果如图 3-6 所示:

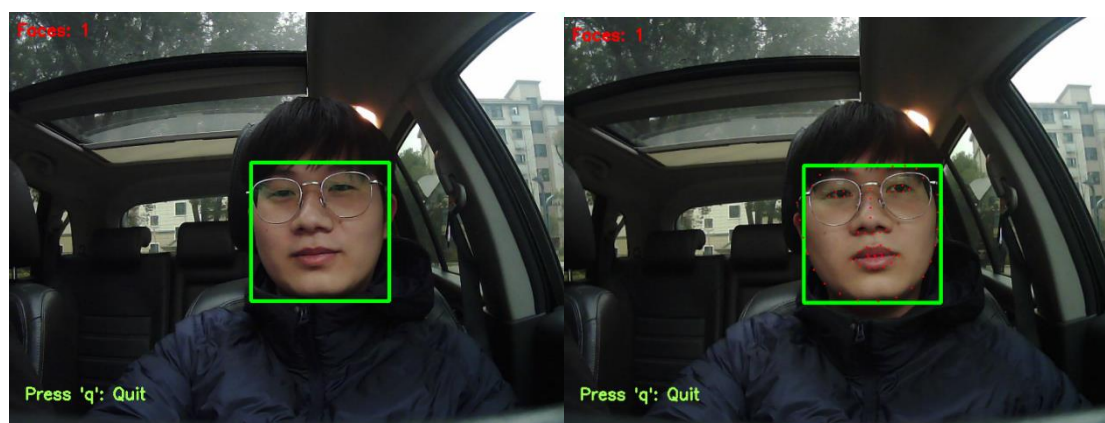


图 3-6 基于 ERT 的驾驶员人脸关键点检测实验结果图

Fig.3-6 ERT-based driver's face key point detection experiment result graph

由实验结果可知，对于由 HOG 人脸检测算法输出的人脸候选框，ERT 人脸关键点检测算法都可以展现出较高的稳定性和准确度，对不同状态的人脸图像都可以实现优秀的人脸关键点定位。

3.3 基于 PnP 问题的驾驶员头部姿态估计

人体的头部重量大约在 5.5kg，当人体进入疲劳状态时，由于大脑对肌肉的控制能力下降，支撑头部重量的头/颈夹肌、胸锁乳突肌等颈部肌肉僵硬松弛，导致头部出现区别于正常状态时的低头等异常行为。通过对驾驶员进行头部姿态估计，可以有效地检测出驾驶员是否出现严重的疲劳特征行为。

3.3.1 人体头部姿态分析

由于人体头部在旋转时，大小、和面部特征点位置不发生变化，所以可以在世界坐标空间中使用旋转轴来表征头部的旋转角度。在三维空间中，可用于表征人体头部旋转的方法有 DCM-方向余弦阵、四元数、旋转向量和欧拉角等，本文使用欧拉角来表征驾驶员头部的旋转姿态。

本文用欧拉角描述驾驶员头部姿态时，将水平和垂直方向的旋转轴分别设为 X 轴和 Y 轴， X 轴与 Y 轴组成的平面法向量为 Z 轴， Z 轴方向按右手法则确定，然后，将驾驶员头部分别绕三维空间中 X 、 Y 、 Z 轴旋转的角度定义为：俯仰角 (*pitch*)、偏航角 (*yaw*) 和滚动角 (*roll*) 如图 3-7 所示：

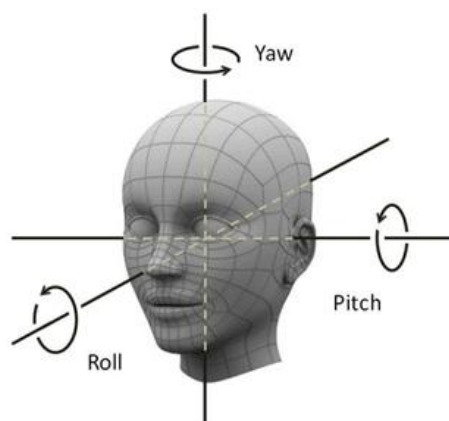


图 3-7 人体头部旋转轴

Fig.3-7 Human head rotation axis

本文对驾驶员头部姿态的估计是根据 3.3 节得到的人脸 2D 坐标通过解决 PnP 问题实时计算头部三维空间欧拉角 (*pitch*、*yaw*、*roll*) 来进行的。其中，人体头部的旋转处于以世界坐标系表征的三维空间下，而疲劳检测系统是通过摄像头采集人体头部图像，是处于图像坐标系下的二维空间。所以需要将以世界坐标系表示的三维人脸坐标与以图像坐标系表示的二维人脸坐标构建映射关系，在构建映射关系时就需要进行相机标定以获取内参矩阵。

3.3.2 相机标定

搭建世界坐标系与相机坐标系的映射，完成 2D 人脸坐标与 3D 人脸坐标的转换，需要对相机进行参数标定以获取表征投影特性的相机内部参数。通常相机标定需要经过两次坐标系的转换，即构建世界坐标系与相机坐标系之间和相机坐标系与图像坐标系之间的映射关系以完成坐标系转换，其转换如图 3-8 所示：

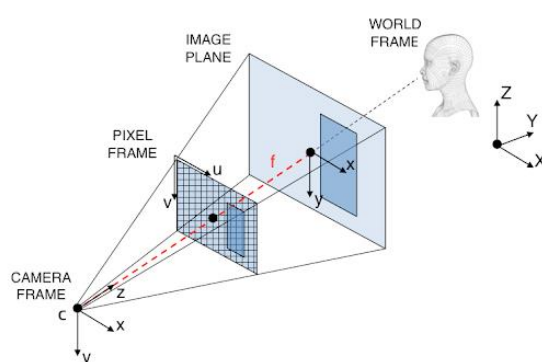


图 3-8 坐标系原理图

Fig.3-8 Coordinate system principle diagram

图 3-8 中，从左自右依次为相机坐标系 (*Camera coordinate system*)、图像坐标系 (*Pixel coordinate system*)、成像平面坐标系 (*Image plane coordinate system*) 和世界坐标系 (*World coordinate system*)。

真实驾驶环境中，驾驶员的头部处于世界坐标系表示下的三维空间，而基于机器视觉的检测系统通过摄像头向系统输入相机坐标系下的图像。因此通过将世界坐标系与相机坐标系的坐标轴对齐^[61]，进行式（3-21）的转换，构建两个坐标系的映射关系。由坐标系原理图可知，成像平面坐标系是相机坐标系和图像坐标系之间的过渡坐标系。所以构建相机坐标系与成像平面坐标系的映射关系见式（3-22）。

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3-21)$$

$$\lambda \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \quad (3-22)$$

式中， \mathbf{R} 表示旋转矩阵， \mathbf{T} 为平移矩阵 f 为焦距，最后根据式（3-23）构建成像平面坐标系与图像坐标系的映射关系：

$$\begin{bmatrix} X_p \\ Y_p \\ 1 \end{bmatrix} = \begin{bmatrix} m_x & s & x_0 \\ 0 & m_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} \quad (3-23)$$

其中， x_0, y_0 表示光轴与像素平面的交叉点， m_x, m_y 表示像素比例因子， s 是传感器倾斜因子。最终经过 DLT（Direct Linear Transform）算法将式（3-21）-（3-23）进行融合，构建出世界坐标系与图像坐标系的映射关系，即表述二维和三维的映射关系：

$$\begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \lambda \begin{bmatrix} fm_x & fs & x_0 \\ 0 & fm_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} X_p \\ Y_p \\ 1 \end{bmatrix} \quad (3-24)$$

简化为式（3-25），式中 \mathbf{K} 表示相机内参数矩阵。

$$\begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \mathbf{X}_w = \lambda \mathbf{K}^{-1} \mathbf{X}_p \quad (3-25)$$

$$\lambda \mathbf{X}_p = \mathbf{K}[\mathbf{R}, \mathbf{T}] \mathbf{X}_w \quad (3-26)$$

式（3-26）中，通过 3.3 节 ERT 算法可以提取到 2D 人脸坐标 $\mathbf{X}_p(u, v)$ ，通过在一个统计测量模型下获取标准人脸三维坐标获得 3D 人脸坐标 \mathbf{X}_w ，最后需要经过相机标定实验获取相机内部参数矩阵 \mathbf{K} 。

本文使用文献[62]所提出的一种新的相机标定算法进行标定，首先定义矩阵 $\mathbf{C} = \mathbf{K}[\mathbf{R}, \mathbf{T}]$ ，将式（3-26）化为 $\lambda \mathbf{X}_p = \mathbf{C} \mathbf{X}_w$ ，将等式写成向量形式为：

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{C} \begin{pmatrix} r_1 & r_2 & r_3 & t \end{pmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3-27)$$

式中， r_1 、 r_2 、 r_3 为矩阵 \mathbf{R} 的列向量，由于实际标定使用平面棋盘格标定板进行标定，所以 $Z=0$ ，故式 (3-27) 化为：

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{C} \begin{pmatrix} r_1 & r_2 & t \end{pmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (3-28)$$

将 c_1 c_2 c_3 表示为 \mathbf{C} 的列向量， \mathbf{C} 可以表示为：

$$\mathbf{C} = [c_1 \ c_2 \ c_3] = \frac{1}{\lambda} \mathbf{K} [r_1 \ r_2 \ t] \quad (3-29)$$

加上旋转向量点积为 0 和向量长度相等两个约束条件：

$$\begin{aligned} r_1^T r_2 &= 0 \\ \|r_1\| &= \|r_2\| \end{aligned} \quad (3-30)$$

式 (3-29) 变换为：

$$\begin{cases} c_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} c_2 = 0 \\ c_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} c_1 = c_2^T \mathbf{K}^{-T} \mathbf{K}^{-1} c_2 \end{cases} \quad (3-31)$$

至此，使用两张以上的标定照片就可以将相机内参矩阵 \mathbf{K} 求出。如图 3-9 所示，实验使用 8×6 的棋盘，通过算法获取角点特征，并标定函数生成矩阵，输出畸变系数矢量。经过实验相机标定参数结果如表 3-3 所示：

表 3-3 相机标定参数

类型	具体参数		
相机内部参数	552.9714072890398	0.00000000000000	328.3091158769222
	0.00000000000000	553.4329492987268	236.9082458910770
	0.00000000000000	0.00000000000000	1.00000000000000
畸变参数	-0.3838811126603	0.1649164927455	-0.00393849945373
	-0.0028754001586	0.00000000000000	

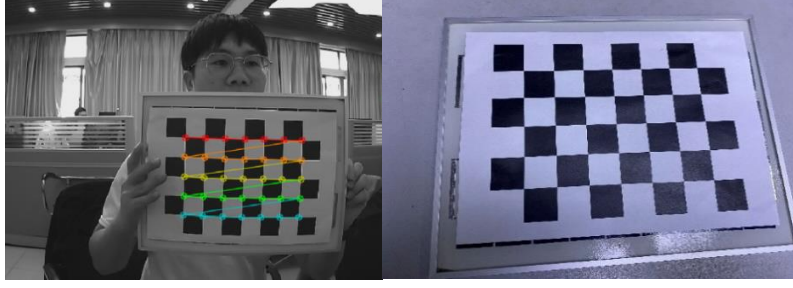


图 3-9 相机标定棋盘格

Fig.3-9 Camera calibration checkerboard

表 3-1 为相机标定实验结果，经过实验标定的相机畸变参数可用于修正镜头畸变，而内参矩阵 \mathbf{K} 则是构建坐标系映射关系，求解旋转、平移矩阵的关键参数。

3.3.3 基于 PnP 求解头部姿态

PnP (Perspective-N-Point) 问题的定义最早由 Horaud^[63]提出，就是通过世界坐标系和图像坐标系中已知的坐标信息，利用构建的映射关系求解 \mathbf{R} 和 \mathbf{T} 的问题。即是已知 \mathbf{K} 、 \mathbf{X}_p 、 \mathbf{X}_w 求解式 (3-26) 中的旋转矩阵 \mathbf{R} 和平移矩阵 \mathbf{T} 。

在对头部姿态 PnP 问题的求解中，本文通过 ERT 人脸关键点检测算法获取了驾驶员面部在图像坐标系的特征坐标 \mathbf{X}_p ，在已知统计模型标定中得到人脸特征在世界坐标系的坐标矩阵 \mathbf{X}_w ，并且利用相机标定计算出了相机内参矩阵 \mathbf{K} ，通过计算可得：

$$\mathbf{R} = \begin{bmatrix} r_{00} & r_{01} & r_{02} \\ r_{10} & r_{11} & r_{12} \\ r_{20} & r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} \cos \phi \cos \gamma \\ \cos \phi \sin \gamma + \sin \phi \sin \phi \cos \gamma \\ \sin \phi \sin \gamma - \cos \phi \sin \phi \cos \gamma \end{bmatrix} \quad (3-32)$$

$$+ \begin{bmatrix} -\cos \phi \cos \gamma \\ \cos \phi \sin \gamma + \sin \phi \sin \phi \sin \gamma \\ \sin \phi \cos \gamma + \cos \phi \sin \phi \sin \gamma \end{bmatrix} + \begin{bmatrix} \sin \phi \\ -\sin \phi \sin \phi \\ \cos \phi \sin \phi \end{bmatrix}$$

$$\begin{cases} \varphi = \text{atan}(-r_{12}, r_{22}) \\ \phi = \text{atan}(r_{02}, \sqrt{r_{12}^2 + r_{22}^2}) \\ \gamma = \text{atan}(-r_{01}, r_{00}) \end{cases} \quad (3-33)$$

式中 φ 、 ϕ 、 γ 为驾驶员头部的三个欧拉角的值，本文分别对驾驶员头部姿态的俯仰角 (*pitch*)、偏航角 (*yaw*)、滚动角 (*roll*) 进行实验检测，其结果如图 3-10 所示：

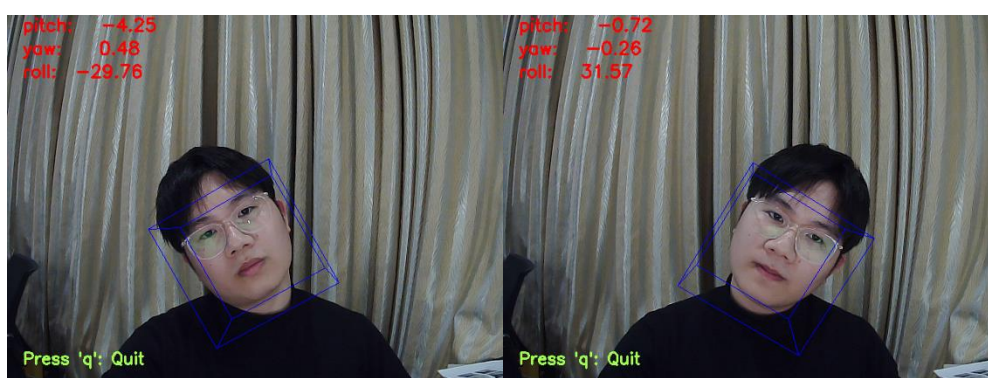
(a) 俯仰角 *pitch* 姿态检测(b) 偏航角 *yaw* 姿态检测(c) 滚动角 *roll* 姿态检测

图 3-10 驾驶员不同头部姿态检测实验结果

Fig.3-10 Driver's different head posture detection experiment results

由实验结果图可知，当驾驶员头部绕 X, Y, Z 轴转动时，模型实时计算欧拉角并估计驾驶员的头部姿态，当驾驶员出现低头动作时，反应在欧拉角上是 *pitch* 数值的变化，如 (a) 图中低头时 *pitch* 值为 13.25，抬头时为 -22.49。当驾驶员左右摇头时，体现为 *yaw* 角度的变化，如图 (b) 中驾驶员向左摇头和向右摇头的 *yaw* 角度分别为 30.71 和 -31.98。当驾驶员出现偏头动作时，体现为 *roll* 角度的变化，如图 (c) 驾驶员左右偏头的角度分别为 -29.76 和 31.57。

在驾驶环境对驾驶员头部姿态估计算法进行实验，实验结果如图 3-11 所示：

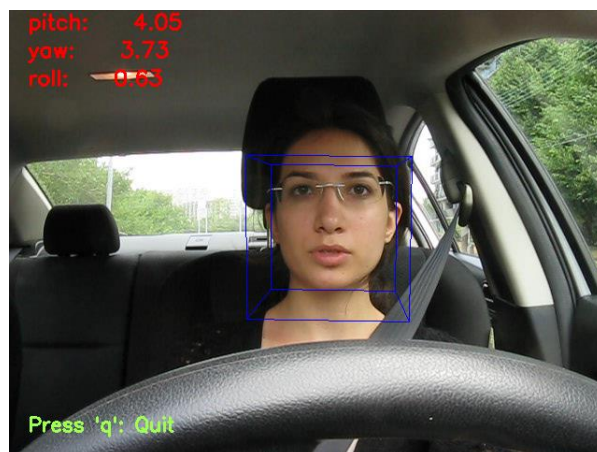


图 3-11 驾驶员头部姿态估计实验结果

Fig.3-11 Driver's head posture estimation experiment results

由图 3-11 左上角的结果显示，在自然光条件下，模型对驾驶员的头部姿态能够进行很好的估计，满足疲劳检测系统对实时性和准确度的要求。

至此，便完成了驾驶员实时的头部姿态估计，为后续驾驶员头部疲劳特征的提取奠定基础。

3.4 本章小结

本章实现了驾驶员的人脸检测和面部关键点定位，并且基于 PnP 问题进行了驾驶员实时的头部姿态估计。在人脸检测方面，在分析了 HOG 算法的优缺点和实际疲劳检测系统的需求后，使用 HOG 人脸检测算法用作疲劳检测系统的人脸检测环节，并对其原理、具体算法的实现和实验结果进行了详解阐述。在人脸关键点定位方面，本文选择目前较为先进的 ERT 人脸对齐算法，对其算法实现流程和实验结果做了进一步的展示，精确定位了驾驶员在不同角度下面部 68 个关键点坐标。在头部姿态估计方面，首先对欧拉角的定义和 4 种坐标系进行了介绍，然后通过坐标系转换和相机标定实验建立了二维空间坐标与三维空间坐标的映射关系，最后根据 ERT 人脸关键点检测输出的二维人脸关键点坐标进行 PnP 问题的求解，得出驾驶员头部姿态欧拉角，为后续疲劳特征提取奠定基础。

第四章 基于面部信息和头部姿态的驾驶员疲劳特征提取

有研究表明，驾驶员在进入疲劳状态时，其疲劳特征通常表现在面部和头部，具体体现为：眨眼状况的改变、频繁哈欠甚至低头瞌睡等。基于机器视觉对驾驶员进行疲劳检测，在 HOG 和 ERT 算法的输出以及估计出头部姿态欧拉角的基础上，提取驾驶员眼部、嘴部、头部的疲劳特征，并将特征的变化作为疲劳检测的基础。

4.1 基于驾驶员眼部疲劳特征的提取方法

4.1.1 驾驶员眼部疲劳特征概述

在疲劳驾驶过程中，驾驶员眼部的疲劳特征尤为重要，眼睛作为“心灵之窗”往往能反映出当下人体的状态信息，是最能反应驾驶员疲劳程度的疲劳特征之一。

通常情况下，驾驶员在正常驾驶状态下每次眨眼的速度较快，约为 250ms。当驾驶员疲劳时，眼睛由睁到闭再恢复睁眼这一过程的速度减缓约 0.6 秒、单次闭眼时长增加，相同时间内眨眼频率上升。故利用驾驶员眼部疲劳特征对驾驶员进行疲劳检测时，可将眨眼频率和闭眼时长用作特征参数以辨识疲劳。本文将眨眼频率定义为 f_{eye} ：

$$f_{eye} = \frac{T_{eye}}{M} \quad (4-1)$$

式 (4-1) 中，由摄像头输入的单位时间内视频连续帧中的闭眼帧数为 T_{eye} ， M 为视频总帧数。基于眨眼频率的定义，算法只需在连续的视频图像帧中检测出驾驶员闭眼帧数，计算眨眼频率，就可以获得由眨眼表征的眼部疲劳特征。

4.1.2 驾驶员眼部疲劳特征提取技术路线

对驾驶员眼部疲劳特征的检测是否准确，关键在于对驾驶员的眼部特征提取是否准确，即对驾驶员眨眼检测的精度是否够高。由于人体的眨眼速度在毫秒级，在实际驾驶状态中，驾驶员不仅眨眼次数多、眨眼速度快，还会由于转向、倒车等操作出现不同角度影响检测的情况，这对驾驶员的眨眼检测的实时性和准确度提出较大的挑战。

基于眼部疲劳特征的复杂程度，本文结合 EAR 算法和深度残差神经网络，提出了一种多模型、多方位提取驾驶员眼部疲劳特征的方法，以提高眨眼检测的精确度，降低漏检率。

4.1.3 基于 EAR 的眼部疲劳特征提取

正常情况下，人眼主要有睁眼和闭眼两种状态，无论哪一种状态，眼睛的宽度基本保持不变，当眼睛睁开时，从上眼睑到下眼睑的高度差变大，反之变小，故可以利用眼

睑的纵横比来反映眼睛睁闭的情况，本文从驾驶员面部的关键点出发，提出用眼睑比 EAR (Eye Aspect Ratio) 算法来判断驾驶员眼睛的开闭程度。

人脸图像在经过本文第三章基于 ERT 的人脸关键点定位，得到人眼 12 个关键点的坐标索引，其中左眼的 6 个关键点坐标为 $(x_1, y_1)_{left} \cdots (x_6, y_6)_{left}$ 如图 4-1 所示：

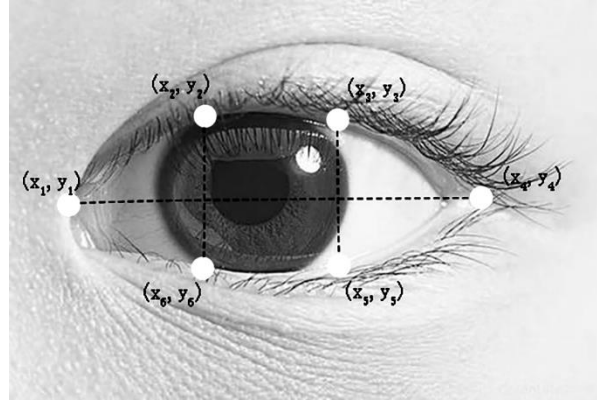


图 4-1 眼（左）部关键点

Fig.4-1 Key points of the eye (left)

基于 EAR 算法对眼部关键点进行计算，其左眼的 EAR 计算公式如下：

$$EAR_{left} = \frac{\|y_{l2} - y_{l6}\| + \|y_{l3} - y_{l5}\|}{2\|x_{l1} - x_{l4}\|} \quad (4-2)$$

式中，分子计算的是驾驶员上眼睑与下眼睑之间的欧式距离，分母计算的是眼角之间的水平距离。同理，右眼 EAR 计算如下：

$$EAR_{right} = \frac{\|y_{r2} - y_{r6}\| + \|y_{r3} - y_{r5}\|}{2\|x_{r1} - x_{r4}\|} \quad (4-3)$$

眼睛的关键点之间的距离使用欧几里德距离计算公式计算所得，其公式如式 (4-4) 所示：

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4-4)$$

由于人眼的两只眼睛眨眼是同步进行的，故将左右眼实时计算的 EAR 合并成最终的 EAR 值：

$$EAR = \frac{EAR_{left} + EAR_{right}}{2} \quad (4-5)$$

本文基于人体眨眼速度快、特征复杂，依据单一 EAR 算法进行驾驶员的眼部特征提取容易出现漏检的问题。在眼部特征提取上，提出将 EAR 算法与深度残差神经网络模型结合补充，以提升整体疲劳检测方法的适用性和准确度。

4.1.4 基于深度学习的眼部疲劳特征提取

目前在疲劳检测领域，大多采用传统的卷积神经网络对驾驶员的行为特征进行提取，而这种 CNN 网络随着深度和网络层数的增加，网络的梯度会不断消失直至退化^[64]。这

种不能深度叠加卷积层的缺点也导致了在疲劳检测中网络提取的疲劳特征层次浅，检测精度低等问题。针对这些不足，对于复杂的眼部疲劳特征，使用深度残差神经网络（Deep Residual Network, ResNet）与眼睑纵横比 EAR 相补充，以此来深层次提取驾驶员眼部疲劳特征。网络结构如图 4-2 所示：

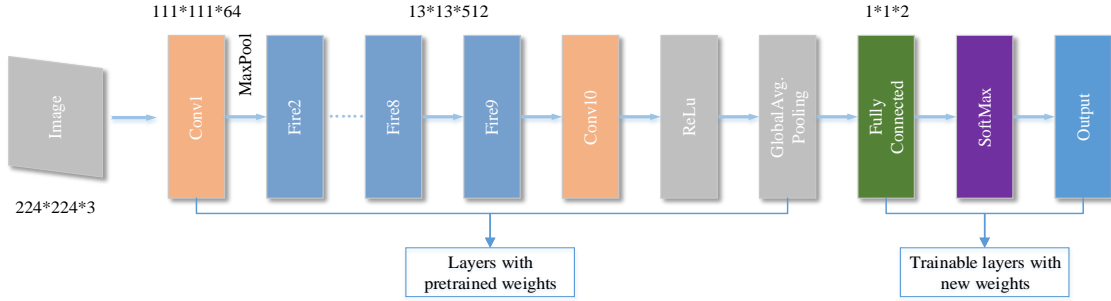


图 4-2 深度残差神经网络结构图

Fig.4-2 Deep residual neural network structure diagram

若假设传统的 CNN 损失函数为 $f(X, W)$ ，则在网络进行反向传播时公式如下：

$$\frac{\partial Loss}{\partial X} = \frac{\partial f(X, W)}{\partial X} \quad (4-6)$$

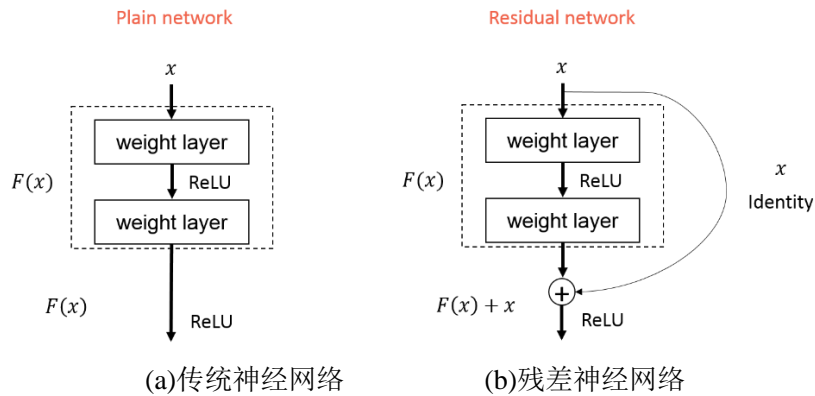
将网络叠加拓展至深层公式将变为：

$$Loss = f_n(X_n, W_n), \dots, L_2 = f_1(X_1, W_1) \quad (4-7)$$

计算式（4-7）中网络的第 n 层梯度：

$$\frac{\partial Loss}{\partial X_1} = \frac{\partial f_n(X_n, W_n)}{\partial X_n} \times \dots \times \frac{\partial f(X_{i+1}, W_{i+1})}{\partial X_i} \quad (4-8)$$

由上式易得，传统的 CNN 网络在层数较深时，浅层网络的梯度会在进行式（4-6）的训练迭代时逐渐消失。而与之不同的是，残差神经网络将残差单元加在传统神经网络的结构中，如图 4-3(b)所示：



(a)传统神经网络 (b)残差神经网络

图 4-3 传统神经网络与残差神经网络

Fig.4-3 Traditional neural network and residual neural network

由图 4-3 可知，原先的网络结构在增加映射 x 后，输出由 $H(x) = F(x)$ 转化为 $H(x) = F(x) + x$ ，这样残差神经网络便可以通过将 $F(x)$ 权重置 0，通过同等映射 x 维持梯度。基于此，式（4-7）将变为式（4-9）。

$$\frac{\partial X_{i+1}}{\partial X_i} = \frac{\partial X_i + \partial f(X_n, W_n)}{\partial X_i} = 1 + \frac{\partial f(X_{i+1}, W_{i+1})}{\partial X_i} \quad (4-9)$$

由此，浅层网络的梯度便不会随着深度的提升而消失。其中网络为了提高训练的稳定性，还添加了批规范化层（BatchNorm），批规范化层的原理就是在训练网络的过程中，对一个 batch 取图像时先对其进行式（4-10）的归一化操作，本文所使用的残差网络在每一次卷积操作以后都加入一层 BatchNorm 层，目的是加速神经网络的收敛速度并提高稳定性。

$$\hat{x}_k = \frac{x_k - E[x_k]}{\sqrt{\text{Var}(x_k)}} \quad (4-10)$$

在实际任务场景下，深度神经网络往往需要庞大的数据量用以训练，且深度越深数据量的需求越大。由于真实驾驶环境的人脸数据集较少，现有的用以疲劳检测的 CNN 网络为了满足网络训练所需数据量的需求，一般采用公开的人脸数据集来进行训练，而不是采用真实驾驶环境中的人脸图像。虽然在数量上满足了网络的需求，但是由于训练域与任务域的不同，网络的精度最终在实际场景下往往不能满足疲劳检测的要求。

如今的深度学习相较于传统的机器学习算法虽然分类预测的准确度有了明显的提高，但精确度的提高是以对计算能力的巨大需求为代价的，表 4-1 为传统机器学习算法与深度学习之间的对比：

表 4-1 传统机器学习算法与深度学习算法对比

算法	机器学习	深度学习
数据集	用小数据集得出结果	需要非常大的数据集
计算需求	快速训练模型	计算密集型
模型特征	需要尝试不同分类器和特征	自动学习特征和分类器
模型准确度	精度进去平台期	精度无限制

由表 4-1 可知，尽管深度学习具有更强的网络表征和特征提取能力，但是其高精度背后是不可估量的具有针对性的目标数据集。在传统的深度学习模型训练中，往往将来源于同域的数据拆分成训练、测试集，并假设实际运用场景的输入数据与先前模型训练数据集来自同一个域，即认为输入数据的特征空间与模型表征参数总是向契合的，实际上以该数据域的数据测试出来的模型准确率与实际场景并不符合，也不合理，是过于理想化的。（如：使用人脸数据集域训练出来的疲劳检测模型与真实场景下驾驶员在车辆环境的数据域并不相同）实际使用场景下，获取完全覆盖模型使用场景的训练集数据域

不仅需要取得大量来源于使用场景的数据源,还需要支付巨额的人力成本用以标注数据。这样便需要提出一种可迁移模型特征的算法。

迁移学习算法就是利用深度学习神经网络的封装性,将网络模型权重从源域迁移至目标任务域的一种网络训练算法。通过把网络原始特征空间领域的特征数据映射到新的特征空间中,网络模型便可以把源域中学习到的特征权重迁移至目标任务域,并进行实际分类任务。由于公开的疲劳驾驶数据集较少,现有的真实驾驶环境数据集的数量不足以训练深度神经网络模型。针对以上难题,本文提出使用基于 ResNet 的迁移学习模型^[65],迁移学习模型利用 ResNet-50 中经源域充分训练的网络权重参数,构建深度残差神经网络,以提高网络的泛化性。

迁移学习模型输入维度为 $224 \times 224 \times 3$,将视频帧输入模型后,将人脸图像中眼睛部分 ROI 裁剪为 224×224 ,并转化为 RGB 三通道图像。首先,为了提取特征信息并降低维度,将 ROI 图像输入到 $7 \times 7 \text{kernel}, \text{pad} = 3, \text{stride} = 2$ 的卷积层和 $3 \times 3, \text{stride} = 2$ 的最大池化层。随后,为了更深层次提取眼部图像特征信息,将特征图依次经过 52 个 $3 \times 3 \text{kernel}$ 卷积池化层,其中在卷积层和池化层使用 ReLU 函数进行非线性化。模型采用多个小卷积核堆叠可以大大减少模型训练的参数量并可以提取到更深层次的特征和更复杂的非线性关系。最后,模型采用平均池化层和全连接层来连接前层网络,通过使用 softmax 函数回归二维向量,预测眼睛睁闭情况。经过迁移学习算法训练的深度残差神经网络与 EAR 眼睑纵横比算法相补充,提高系统的检测精确度。

4.1.5 实验结果与分析

眼部疲劳特征是判定驾驶员是否疲劳的主要特征,本文基于眼部几何特征和深度残差神经网络提取驾驶员眼部疲劳特征,检测是否眨眼。使用两种模型相补充的方法,相比单一模型提高了检测的准确度,降低了眨眼漏检率。

(1) 基于 EAR 眼睑纵横比的眼部特征提取有效性实验

在对 EAR 眼睑纵横比算法的实验中,实际驾驶环境中采集的每帧图像经过 HOG 和 ERT 算法检测出驾驶员眼部的关键点坐标,通过 EAR 眼睑纵横比算法根据式(4-5)实时计算驾驶员 EAR 值。选取一段模拟驾驶视频,驾驶员在正常驾驶的过程中进行若干次眨眼,其中截取两次眨眼进行分析,实验结果如图 4-4 所示:

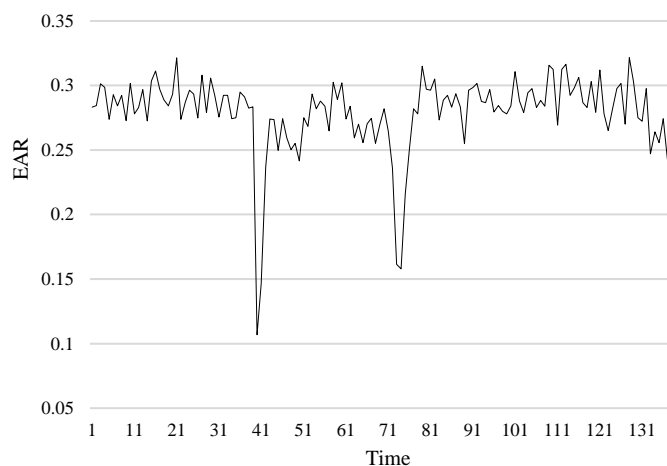


图 4-4 单位时间序列 EAR 值

Fig.4-4 Unit time series EAR value

由实验结果图可知，单位时间序列中，实验测试的驾驶员处于正常睁眼状态时，EAR 值的范围在 $[0.25, 0.32]$ 之间，在驾驶员进行眨眼动作时，由图 4-4 可知，EAR 值以极大斜率趋于 0。由此可以设置一定的阈值，根据 EAR 判断当前眼睛的状态，检测眨眼。本文经实验将阈值设为 0.25，模型检测到驾驶员实时的 EAR 值连续 2 帧小于阈值则认为驾驶员进行了一次眨眼行为，本文将 EAR 值由大于阈值到低于阈值，持续 2 帧或以上，再由小于阈值上升至阈值以上的这一完整过程，认定为一次眨眼行为。

(2) 基于深度残差神经网络模型的眼部特征有效性验证

深度残差神经网络模型采用 CEW^[66]、YawDD^[67]数据集进行迁移学习训练。将 YawDD 视频提取关键帧获得真实驾驶环境下的训练样本图像，如图 4-5 (a) 所示。然后将标签好驾驶员状态的 4538 张来源于 YawDD 的图像和来源于 CEW 的 2423 张图像 ROI 截取并降维，如图 4-5 (b) 所示：



(a)YawDD 数据集预览 (b)眼部 ROI

图 4-5 YawDD 数据集预览与眼部 ROI

Fig.4-5 YawDD dataset preview and eye ROI

获取网络的训练数据集后，实验对残差神经网络的训练如下：（1）图像调整为 224×224 以满足网络的输入。（2）为了保障迁移学习的效率和精度，对训练样本进行

数据增强，并将像素数值与 ImageNet 平均 RGB 值做差。（3）建立深度卷积神经网络 DCNN。基于 ResNet-50 的网络结构，使用维度为 2 的全连接层和 softmax 层将原先的层级加以替换。然后将网络中卷积层的权重参数迁移至 DCNN 网络模型中。（4）使用迁移学习算法进行训练。冻结卷积层的权重参数，使用真实驾驶环境的图像对全连接层和 softmax 层进行完全训练。其中，训练参数设置为：1.学习率：采用动态学习率。（0.001，每 7 个 epoch 减少 1/10）2.优化器：Adam。3.batch = 32；epoch = 100。网络的训练集和验证集在整个过程的准确率变化如图 4-6 所示：

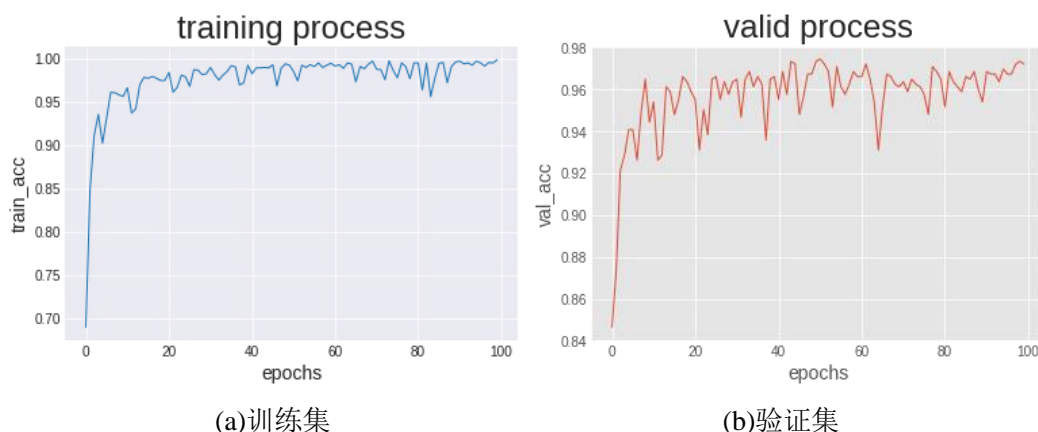


图 4-6 训练过程结果图

Fig.4-6 Training process result graph

由模型的实验结果图可知，深度残差神经网络在经过迁移学习训练迭代后，在训练集和验证集都表现出优秀的准确率，分别高达 98.9%和 97.5%。经过真实驾驶环境图像训练的深度学习网络模型与眼睑纵横比 EAR 算法互相结合补充用作驾驶员眼部疲劳特征的提取和检测，可以提高疲劳检测方法整体的检测精度，减少眼部特征的漏检情况。

4.2 基于驾驶员嘴部疲劳特征的提取方法

4.2.1 驾驶员嘴部疲劳特征概述

驾驶员在产生疲劳时，不仅存在眨眼频率、闭眼时长等眼部疲劳特征，还通常伴随着哈欠行为。因此，可以通过检测驾驶员在驾驶车辆过程中嘴部的哈欠行为来判定驾驶员是否进入疲劳状态。通常情况下，驾驶员的嘴部状态有：（1）正常状态下松弛嘴部紧闭；（2）说话时嘴部持续微张；（3）哈欠行为时嘴部由紧闭到微张，然后持续张到最大后恢复紧闭。因此根据驾驶员嘴部的开合特征可以较为准确地分辨出当前嘴部的状态。

4.2.2 驾驶员嘴部疲劳特征提取技术路线

基于驾驶员嘴部状态来提取疲劳特征时，即是在驾驶员嘴部闭合、说话和哈欠 3 种可能存在的嘴部行为中，检测出哈欠行为。相较于眼部特征，驾驶员嘴部特征复杂度低、

持续时间长，将嘴部关键点定位准确即可快速根据几何特征识别嘴巴状态，无需使用残差神经网络模型，以降低车机系统的 CPU 消耗。

由于个体的差异性，并非每位驾驶员在进入疲劳状态的时候都伴随着哈欠行为，所以本文将驾驶员嘴部哈欠检测作为一种辅助检测手段，在眼部检测受阻时（如：佩戴墨镜）提供一种检测方法，提高系统的鲁棒性。

4.2.3 基于 MAR 的嘴部疲劳特征提取

驾驶员嘴部疲劳特征的提取与基于 EAR 提取眼部疲劳特征相似，由 3.2 节输出的结果可知，嘴巴部外轮廓 12 个关键点坐标为 $(x_1, y_1), (x_2, x_2) \cdots (x_{12}, x_{12})$ ，如图 4-7 所示：

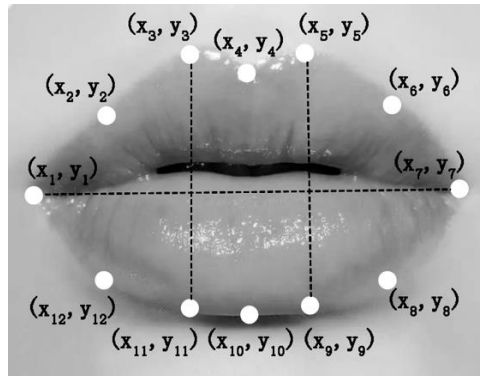


图 4-7 驾驶员嘴部外轮廓关键点

Fig.4-7 Key points of the outline of the driver's mouth

基于嘴部纵横比（Moth Aspect Ratio, MAR）来进行驾驶员嘴部状态的识别，其公式如下：

$$MAR = \frac{\|y_3 - y_{11}\| + \|y_5 - y_9\|}{2\|x_1 - x_7\|} \quad (4-11)$$

式中，关键点之间的距离与眼睑纵横比类似，采用欧几里德距离计算公式进行计算，其中分子计算的是上半唇与下半唇之间的欧式距离，分母计算的则是左右嘴角的欧式距离。由 MAR 公式可知，驾驶员嘴部张开时，上下半嘴唇距离变大，嘴部宽度变小，MAR 值变大。当驾驶员嘴部闭合时，上下半嘴唇距离最小，即 MAR 值最小。驾驶员说话时嘴部张开程度要小于哈欠时的程度，所以说话时的 MAR 值要小于哈欠时的 MAR，嘴部紧闭时 MAR 值最小。故可以设定一定的阈值，将驾驶员嘴部的三种状态判别出来。

实际驾驶环境中采集的每帧图像，经过 HOG 和 ERT 算法检测出驾驶员嘴部外轮廓的 12 个关键点坐标，基于 MAR 嘴部纵横比算法实时计算驾驶员嘴部高度与宽度的欧氏距离比，实验选取一段模拟驾驶视频，在视频的开始驾驶员按照实验指令保持嘴部处于正常闭合状态，接着与实验人员进行一小段的谈话，最后模拟两次哈欠行为，计算视频中每一帧图像的 MAR，其实验结果如图 4-8 所示：

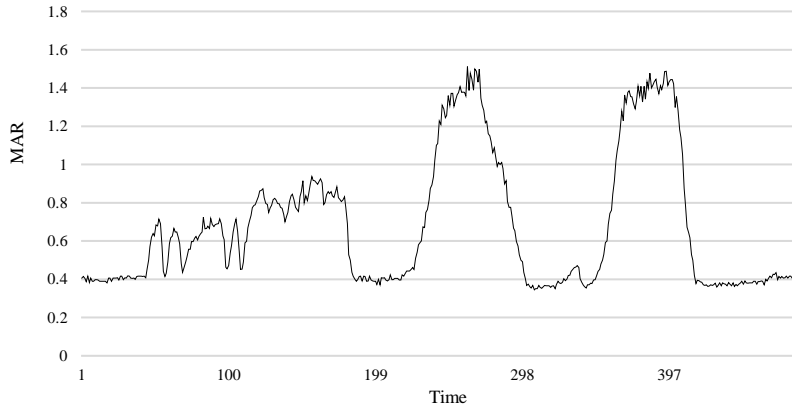


图 4-8 单位时间序列驾驶员嘴部 MAR 值

Fig.4-8 The MAR value of the driver's mouth per unit time series

由图 4-8 可知，当驾驶员嘴部处于正常闭合状态时，MAR 稳定在一个较小的值上下波动；当驾驶员嘴部处于说话状态时，MAR 的实时值在[0.6, 1.0]范围波动；当驾驶员进行哈欠行为时，MAR 的实时值由于嘴部几何形状的剧烈变化而迅速上升到[1.2, 1.5]范围内。

值得注意的是，实验中当驾驶员处于说话状态时，其嘴部 MAR 的实时值不是恒定处于某一值，MAR 的值会由于嘴部说话时一张一合的状态而变化较大，不能根据单一的阈值判断驾驶员是否处于说话状态，所以经过实验，本文采用双阈值的方法，分别使用 0.6 和 0.9 作为说话和哈欠阈值，将 10 和 5 分别作为说话和哈欠的帧数阈值，当实时计算的 MAR 值连续 10 帧大于 0.6 小于 0.9 时，判断驾驶员处于说话状态，当 MAR 连续 5 帧大于 0.9 时判断驾驶员处于哈欠状态，由此通过 MAR 值便可判断驾驶过程中的嘴部状态。

4.3 基于驾驶员头部姿态的疲劳特征提取方法

人体在疲劳时，头部姿态也会与眼部和嘴部一样出明显的疲劳特征。在正常驾驶状态下，驾驶员需要随时对车辆外界信息保持关注，会灵活转动头部以查看后视镜。而当驾驶员在进入疲劳状态时，由于注意力的下降，转动头部查看后视镜的行为将随之减少，随着大脑对神经中枢的控制能力下降，松弛的颈部肌肉很难将头部持续维持在正常位置，在疲劳程度加深后，驾驶员会出现点头、后仰、左右偏倒等异常行为。当驾驶员出现头部异常行为时，说明其已经处于随时可能发生事故的严重疲劳状态。若要描述这些头部异常行为，便可使用俯仰角(*pitch*)、偏航角(*yaw*)、滚动角(*roll*)三个维度来表征头部姿态的转动角度。

根据前文基于相机标定和坐标系转换的 PnP 求解，估算得到驾驶员实时的头部姿态欧拉角(*pitch*、*yaw*、*roll*)，通过检测计算驾驶员头部异常角度和低头频率 f_{head} 便可以检测出驾驶员是否疲劳驾驶， f_{head} 定义如下：

$$f_{head} = \frac{T_{head}}{M} \quad (4-12)$$

定义中， T_{head} 为摄像头输入的单位时间内视频连续帧中的低头帧数。据文献[68]可知，正常情况下，驾驶员的头部绕中轴的 X 轴、 Y 轴、 Z 轴，即 ($pitch$ 、 yaw 、 $roll$) 的最大旋转角度分别是 $[-60.4^\circ, 69.6^\circ]$ 、 $[-90^\circ, 75^\circ]$ 和 $[-40.9^\circ, 36.3^\circ]$ 。定义低头频率 f_{head} 后，便可通过实时估算 $pitch$ 、 yaw 、 $roll$ 来计算驾驶员头部异常角度。

由于驾驶员的头部异常状态主要反映在俯仰角 ($pitch$) 和翻滚角 ($roll$)，本文参照国际标准 PERCLOS 中的 P80 进行头部姿态的特征提取，若检测到实时计算的头部姿态欧拉角偏离正常旋转范围的 80%，即平均 $pitch > 25^\circ$ ， $roll > 15.4^\circ$ ，则对驾驶员发出头部姿态异常预警。

4.4 本章小结

本章主要对驾驶员的面部信息和头部姿态进行了疲劳特征提取。首先使用眼睑纵横比 EAR 算法提取眼部疲劳特征，然后提出深度残差神经网络与之相补充的方法以提高眨眼检测精度，减少漏检率，其中针对疲劳数据集不足的问题采用迁移学习算法进行网络的训练以提高网络的鲁棒性和泛化能力。然后，采用 MAR 算法实现了嘴部的三种状态（闭合、说话、哈欠）的判别。最后在头部姿态的特征提取中，根据实时计算的头部姿态欧拉角并参照 P80 标准定义了低头频率和头部异常角度，进而提取头部姿态的疲劳特征。通过本章对驾驶员面部信息和头部姿态的疲劳特征提取，为后文判断驾驶员是否疲劳提供特征数据基础。

第五章 基于多特征的疲劳状态检测方法

经过人脸检测、关键点定位和头部姿态估计后，利用驾驶员面部和头部姿态疲劳特征检测是否疲劳驾驶。在实际驾驶中，对不同驾驶员进行疲劳状态检测是一个复杂的任务，仅仅使用单一的疲劳特征和固定的疲劳判定阈值不仅鲁棒性低还不能满足不同驾驶员对检测系统的需求。因此本文提出多特征的疲劳检测方法并基于支持向量搭建针对不同驾驶员个性化的检测模型。

5.1 基于多特征的疲劳检测方法总体结构

本文提出的基于多特征的疲劳检测方法，通过机器视觉将机器学习和深度学习算法相结合，采取了眼部、嘴部和头部姿态多种疲劳特征，并使用深度残差神经网络来提升眼部特征提取的准确度，最后针对不同驾驶员建立了 SVM 个性化疲劳检测模型，其原理图如图 5-1 所示：

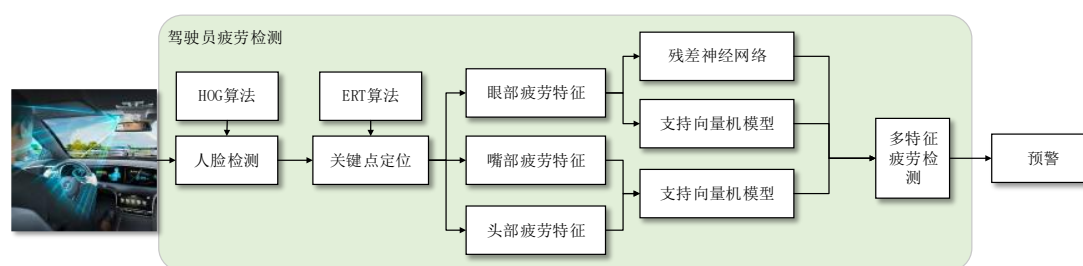


图 5-1 基于多特征的疲劳检测方法原理图

Fig.5-1 Schematic diagram of fatigue detection method based on multiple features

由原理图可知，基于多特征的疲劳检测方法主要包括 3 个部分：

（1）人脸检测定位和头部姿态估计

通过红外摄像头采集驾驶员实时驾驶视频信息，首先，HOG 人脸检测算法将实时传输的视频中一帧帧图像进行人脸检测并标出人脸候选框。其次，使用级联回归树 ERT 算法对人脸候选框进行关键点定位。最后，进行相机标定和坐标系转换，利用人脸二维关键点坐标求解 PnP 问题得到驾驶员头部欧拉角。

（2）多种疲劳特征提取

对驾驶员面部和头部多个疲劳特征进行提取。通过眼睑纵横比 EAR 和深度残差神经网络对驾驶员进行眨眼检测提取眼部疲劳特征，并且根据嘴部几何特征的 MAR 算法对驾驶员进行说话和哈欠检测。在驾驶员头部姿态方面，使用实时计算的欧拉角对驾驶员进行低头和头部异常角度检测。通过利用驾驶员眼部、嘴部和头部姿态多种疲劳特征，大大提高了系统的鲁棒性。

（3）驾驶员疲劳检测

通过设计采集和训练流程，搭建了针对不同驾驶员的 SVM 模型。SVM 模型可以对每一位驾驶员个性化定制疲劳特征阈值，在定制个性化阈值后，使用国际 PERCLOS 判定标准对提取的眨眼频率、哈欠次数、低头次数和异常角度等特征进行疲劳辨识，完成驾驶员的疲劳预警。

基于机器视觉的多特征疲劳检测方法对驾驶员进行疲劳辨识的具体流程如图 5-2 所示：

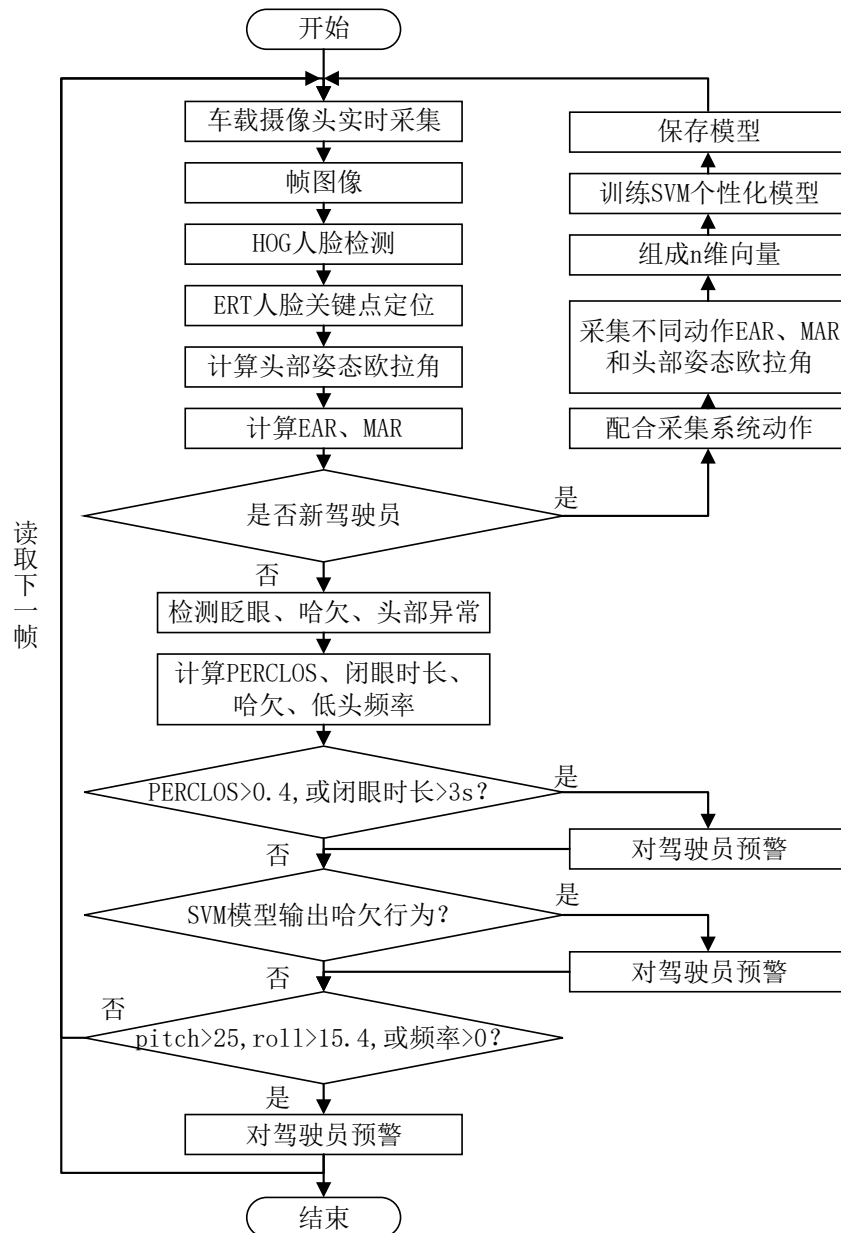


图 5-2 疲劳检测方法总流程图

Fig.5-2 General flow chart of fatigue testing methods

由流程图 5-2 可知，在车载相机开始工作时，对实时拍摄的每一帧图像中驾驶舱内的驾驶员进行人脸检测和关键点定位，并估计头部旋转欧拉角，此时，若驾驶员是新驾驶员，则对驾驶员进行动作采集并根据动作行为特征训练 SVM 个性化模型。若驾驶员

不是新驾驶员，则进行后续的眨眼、哈欠和头部异常检测，根据定义的疲劳判定标准，当实时计算检测的疲劳特征参数符合疲劳指标时，则对驾驶员进行疲劳预警。至此，便完成了一张图像帧的疲劳检测工作，依此循环检测每一帧实时的图像，直至关闭。

5.2 建立疲劳模拟数据集及疲劳判定标准

5.2.1 疲劳驾驶数据集

本文对采用 YawDD 数据集以及自建数据集进行疲劳检测实验。YawDD 数据集是加拿大渥太华大学针对疲劳驾驶建立的哈欠数据集，是目前为数不多的公开疲劳数据集，与本文所研究的方向相匹配。该数据集分别在汽车内后视镜和汽车仪表盘处安装摄像头，在自然光条件下拍摄了不同性别、不同种族和是否佩戴眼镜等不同情况的两种角度的视频，其中记录了驾驶员正常驾驶、说话唱歌、哈欠疲劳三种情况。数据集中的部分样本如图 4-5 所示：

由于 YawDD 数据集并没有拍摄驾驶员疲劳低头瞌睡等情况，本文还建立了疲劳模拟数据集进行实验。通过招募四名男驾驶员和一名女驾驶员，对驾驶员进行剥夺睡眠，并选用饭后血糖升高容易困倦的时间，在户外真实车辆中进行实验采集录制疲劳模拟视频。其中为了保障驾驶员和行人安全，实验采取模拟驾驶的方式，没有选择让处于疲劳状态的驾驶员驾驶车辆进入真实道路环境。自建数据集采集了实验驾驶员正常驾驶、交谈说话、疲劳哈欠和低头瞌睡 4 种状态，如图 5-3 所示：



图 5-3 自建模拟疲劳数据集样本

Fig.5-3 Self-built simulated fatigue data set sample

5.2.2 疲劳驾驶参数判定标准

通过疲劳特征提取，完成状态识别后，需要对疲劳状态的参数制定判定标准。将实时计算的疲劳特征参数与后文 SVM 个性化疲劳阈值模型相结合，完成疲劳预警任务。

(1) PERCLOS

PERCLOS (Percentage of Eyelid Closure Over the Pupil Time) 是国际公认的疲劳状态判定准则, 1999 年美国联邦公路管理局 (FHWA) 为了能探讨出最有效表征疲劳的判定指标, 集结了大量学者专家对多种疲劳判定指标进行实验对比分析, 得出 PERCLOS 是与驾驶员疲劳程度相关性最好的疲劳判定指标。PERCLOS 表示单位时间序列中, 眼睛闭合时间所占的比例, 当比例大于某一阈值则可以判定为疲劳状态。根据眼部状态的不同, PERCLOS 可以分为三类:

EM: 当瞳孔的 50% 以上面积被眼皮盖住时, 认为眼睛处于闭合, 统计单位时间内闭眼所占时长的比重。

P70: 当瞳孔的 70% 以上面积被眼皮盖住时, 认为眼睛处于闭合, 统计单位时间内闭眼所占时长的比重。

P80: 当瞳孔的 80% 以上面积被眼皮盖住时, 认为眼睛处于闭合, 统计单位时间内闭眼所占时长的比重。经过实验研究发现, PERCLOS 的 P80 标准在疲劳驾驶判定中相关性最高, 最适合在疲劳驾驶检测中作为疲劳判定标准, 其原理如图 5-4:

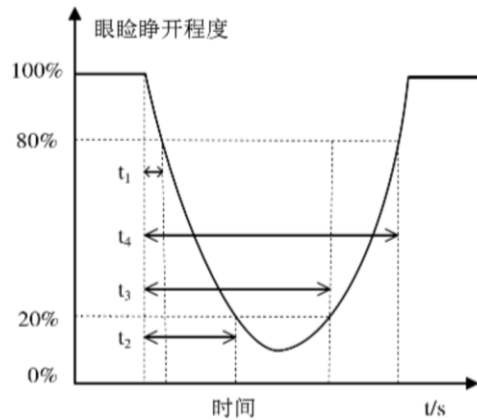


图 5-4 P80 检测原理图

Fig.5-4 P80 detection principle diagram

图 5-4 表示的是一次眨眼行为的时间统计, t_1 表示上眼睑完全睁开到遮住瞳孔 80% 时所经过的时间, t_2 表示上眼睑完全睁开到遮住瞳孔 20% 时所经过的时间, 以此类推, 得出 PERCLOS 中 P80 的计算公式如下:

$$P_{80} = \frac{t_3 - t_2}{t_4 - t_1} \times 100\% \quad (5-1)$$

(2) 眨眼频率

在实际疲劳驾驶中, 由于摄像头实时采集的视频图像帧具有不间断的特性, 因此, 通过检测算法检测驾驶员在视频帧中的闭眼帧, 并统计单位时间的帧数占比, 根据 PERCLOS 的定义, 便可以利用第四章给出的眨眼频率 f_{eye} 来近似求解 PERCLOS 值。

(3) 单次闭眼时长

在驾驶过程中，驾驶员都是通过眼睛获取 80%-90% 的外界信息。驾驶员在进入疲劳状态时，单次闭眼时长会增加，且不论驾驶员由于何种原因，如果驾驶员的单次闭眼时间增加了，就意味着驾驶员对驾驶的路况信息丢失了实时的掌控，而路况信息是瞬息万变的。因此，对驾驶员进行单次闭眼时长的检测不仅可以检测出驾驶员是否疲劳，也保障了处于其他原因长时间闭眼的驾驶员的生命安全。

(4) 哈欠次数

人在疲劳时会出现哈欠行为，哈欠次数作为一种疲劳判断的参数指标，可以很好地辅助眼部的疲劳特征参数。本文在 4.2 节对驾驶员嘴部疲劳特征进行提取，并根据嘴部几何特征使用 MAR 检测出驾驶员是否出现哈欠行为，后文将通过对不同驾驶员建立 SVM 个性化阈值模型来更加准确地检测出驾驶员的哈欠次数。

(5) 低头次数

除了前面几种可以作为疲劳判定标准的面部特征参数外，驾驶员头部姿态参数也可以作为疲劳判定的指标。当驾驶员进入疲劳状态时，头部会失去控制而出现规律性点头甚至低头睡着，根据前文对驾驶员头部姿态的实时估计，结合 SVM 驾驶员个性化疲劳阈值模型，可以检测出驾驶员是否出现低头行为，将低头次数以及低头频率作为疲劳驾驶判定的参数，在驾驶员出现低头行为时给予预警。

(6) 头部异常角度

驾驶员处于清醒状态时，其头部一般保持在一个正常的角度。根据研究，正常成年人的头部运动欧拉角范围是： $(pitch)$ 角度为 $[-60.4^\circ, 69.6^\circ]$ ， (yaw) 角度为 $[-90^\circ, 75^\circ]$ ， $(roll)$ 角度为 $[-40.9^\circ, 36.3^\circ]$ ，本文参照 PERCLOS 的 P80 标准，判断实时计算的头部角度是否超过正常范围的 80%，若超过则对驾驶员进行预警。

5.3 基于支持向量机的驾驶员个性化疲劳检测方法

在实际运用中，由于驾驶员面部特征不尽相同，不同驾驶员都拥有自己独特的面部特征，而通常情况下，一辆私家车有 1 到 3 名的驾驶员，对不同驾驶员采用单一的疲劳阈值容易造成误检，所以本章针对不同驾驶员采用基于支持向量机的个性化疲劳预警模型，以提高疲劳检测的鲁棒性。

5.3.1 支持向量机原理

支持向量机 (Support Vector Machine, SVM) 是用以解决线性回归和分类问题的一种兼备高效率和高精度的机器学习算法。

(1) 线性 SVM

SVM 算法的目的就是要找到一个最优的分类超平面。在求解线性问题中，SVM 可以在线性可分的数据中利用最优的超平面将样本数据进行分类，设 $f(x) = \bar{w} \cdot \bar{x} + b$ 为目标任务平面的决策面，当 $f(x)$ 距离直线两边的支持向量距离最大，且两边距离相等时，

$f(x)$ 为最优决策面，此时 $f(x) = \bar{w} \cdot \bar{x} + b = 0$ ，支持超平面为 $f(x) = \bar{w} \cdot \bar{x} + b = \pm 1$ ，则图像中的分类模型为：

$$f(x) = \bar{w} \cdot \bar{x} + b = |\bar{w} \cdot \bar{x} + b| \quad (5-2)$$

根据拉格朗日算子将公式最优解得到 SVM 函数：

$$y(x) = \sum_{j=1}^n \alpha_j y_j (x_j^T x) + b \quad (5-3)$$

(2) 非线性 SVM

相比于线性 SVM，非线性分类问题由于数据样本不满足函数间隔大于等于 1 的性质，在处理非线性问题时，通常将非线性样本数据转化成线性数据进行分类。若样本数据在当前维度不可分，便使用核函数将其映射到其他维度空间中，实现线性可分的状态。其中，在实验中常见的核函数为：

线性核函数：

$$Ker(x, x_i) = x^T x_i \quad (5-4)$$

多项式核函数：

$$Ker(x, x_i) = (\gamma x^T x_i + r)^p, \gamma > 0 \quad (5-5)$$

径向基函数：

$$Ker(x, x_i) = \exp(-\gamma \|x - x_i\|^2), \gamma > 0 \quad (5-6)$$

两层感知器核函数：

$$Ker(x, x_i) = \tanh(\gamma x^T x_i + r) \quad (5-7)$$

其中，径向基函数拥有较强的泛化能力、抗干扰能力，并且可较好的反应局部信息。故本文采用径向基函数作为模型的核函数。

5.3.2 基于 SVM 个性化疲劳辨识模型建立

在日常生活中，一个家庭的私家车不是只有一个驾驶员，而是由若干个家庭成员组成，因此若一个疲劳检测系统对具有不同面部特征的司机设定统一疲劳阈值，并使用这种固定的单一阈值对不同司机的哈欠行为、眨眼行为和低头瞌睡进行判定，既不可靠，也不符合实际使用场景中对多名驾驶员进行疲劳判断的要求^[69]。针对该问题，为了提高疲劳检测方法的整体鲁棒性和准确度，并满足实际应用场景中多位驾驶员的情况，本文提出了针对不同驾驶员个性化训练疲劳阈值的 SVM 疲劳辨识模型。

本文建立的 SVM 个性化疲劳辨识模型原理如图 5-5 所示：

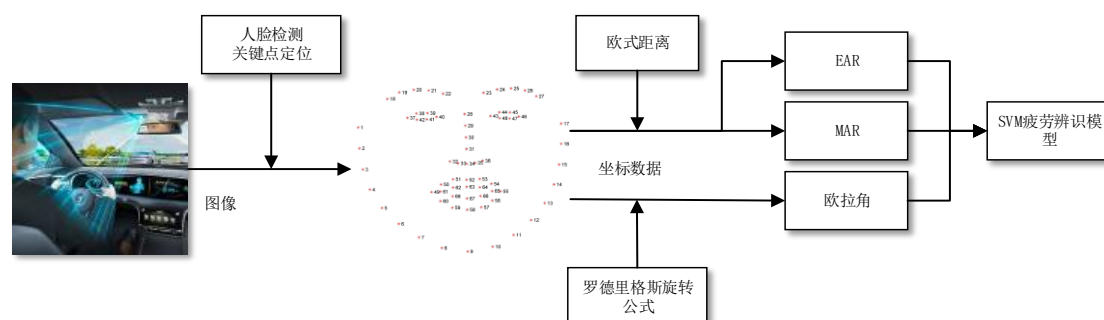


图 5-5 SVM 疲劳辨识模型原理

Fig.5-5 SVM fatigue identification model principle

由原理图 5-5 可知，在实际的疲劳检测中，要搭建基于支持向量机的驾驶员个性化疲劳辨识模型，将模式识别数据分类中优秀的 SVM 机器学习算法运用在疲劳检测问题上。首先需要将由摄像头实时采集的驾驶员人脸图像信息通过 HOG 和 ERT 的人脸检测和关键点定位算法，通过检测提取人脸关键点坐标将人脸视频帧信息变成可供计算的数据；然后，通过眼部和嘴部的几何特征计算欧式距离，将提取的面部数据变成表征几何意义的 EAR 和 MAR 值，同时，通过求解 PnP 问题将现实世界中驾驶员头部旋转的姿态量化为欧拉角 *pitch*、*yaw*、*roll*；最后将求解得到的 EAR、MAR 和欧拉角，针对不同驾驶员，通过支持向量机建立不同的个性化疲劳辨识模型。其具体步骤分为 3 步，分别是：人脸特征采集—特征提取—模型训练。

（1）人脸特征采集

为了提取不同驾驶员在各种行为动作时的面部特征信息和头部姿态欧拉角，当驾驶员初次进入车辆时，系统将对其进行一次简短的特征采集，根据系统步骤指南做出睁闭眼、张开嘴、念出一段文字和低头等动作，系统使用摄像头采集驾驶员每种动作的特征信息（眼部 EAR 数据、嘴部 MAR 数据和头部旋转欧拉角），如图 5-6 所示。



图 5-6 人脸特征采集实时图

Fig.5-6 Real-time image of face feature collection

（2）特征提取

通过上一步骤对驾驶员的人脸特征进行采集后，将采集到的驾驶员眼部纵横比、嘴部纵横比和头部姿态欧拉角数据组成 N 维向量，由于本文自建疲劳数据集由四名男司机和一名女司机拍摄采集而成，故对招募的实验驾驶员提取眼睛睁闭眼 EAR 特征数据、

嘴巴三种状态 MAR 特征数据（闭合、说话、打哈欠）和疲劳异常的 *pitch*、*yaw*、*roll*。将采集的特征参数信息针对不同驾驶员组成 N 维向量，训练个性化模型。

（3）模型训练

系统对采集的不同驾驶员的特征信息进行自动编码，将驾驶员眼部的睁闭眼分别标签为 0 和 1，将驾驶员嘴部闭合、说话、哈欠三种状态分别标签为 0、1、2，以此类推。最后将每一位驾驶员带标签的数据送入支持向量机模型进行训练，得到针对不同驾驶员生成的个性化疲劳辨识模型。

5.4 驾驶员疲劳辨识实验与分析

为了验证所提出的疲劳检测方法的有效性，本文进行了两部分的实验：第一部分是针对 SVM 个性化疲劳辨识模型的实验；第二部分是针对驾驶员多特征疲劳检测方法进行实验。

5.4.1 SVM 个性化疲劳辨识模型实验

为了完成对不同驾驶员个性化疲劳阈值这一目标，在驾驶员首次进入车辆时，通过 5.3 节设计的步骤，系统将根据不同驾驶员进行特征信息采集，将采集到的不同驾驶员在不同状态下的 EAR、MAR 和头部姿态欧拉角构建 SVM 模型，表 5-1 为系统提取的部分数据：

表 5-1 系统对不同驾驶员采集的特征数据（部分）

测试员编号	EAR		MAR			头部姿态（正常）			头部姿态（动作）		
	open	close	normal	talking	yawning	pitch	yaw	roll	pitch	yaw	roll
1	0.27405	0.08593	0.37279	0.64003	1.52164	-3.94691	6.33745	1.85180	15.5987	22.1555	6.90821
2	0.26203	0.10784	0.38015	0.75042	1.42813	-3.01379	14.8119	0.44408	17.1604	11.2211	2.17314
3	0.25032	0.13944	0.15520	0.50029	1.06314	-4.56941	4.11723	1.31819	16.9147	5.22492	-0.70400
4	0.29188	0.15934	0.29554	0.86710	1.16041	-1.92371	10.1082	0.93463	15.0416	5.43838	-1.12942
5	0.23663	0.14575	0.42768	0.79457	1.23892	-2.92260	6.29364	0.94492	22.0486	4.30100	0.31905

分析上表数据可知，不同驾驶员根据系统指令进行动作时的眼部睁闭眼、嘴部三种状态和头部姿态欧拉角 *pitch*、*yaw*、*roll* 具有一定的趋势性。实验根据驾驶员的真实动作状态对数据进行自动标签，然后分别将不同驾驶员的特征数据进行训练生成个性化疲劳阈值模型。实验中，系统按照视频帧数对驾驶员的动作特征数据进行采集：对驾驶员睁眼状态、闭眼状态、正常松弛闭嘴状态、说话状态、哈欠状态、头部四个方向欧拉角（正常姿态、正下方低头、偏向左右下方低头）分别采集 500 个特征数据。

模型训练具体参数如下：采集获取每一位驾驶员 4500 个特征数据（眼部特征数据、嘴部特征数据和头部姿态特征数据分别为：1000、1500、2000）；分类器核函数：RBF

径向基函数；核函数 cache 缓存大小：200；惩罚系数 C：1.0；probability=False；停止训练误差：1e-3。

模型测试采用自建模拟疲劳数据集（见 5.2 节）作为测试样本，并招募 5 名观察记录员对自建模拟疲劳数据集视频中驾驶员的眨眼、哈欠、低头瞌睡次数进行统计，对说话时长进行记录。

为验证本文提出的基于 SVM 的个性化疲劳阈值模型的有效性，实验将本文提出的 SVM 个性化疲劳阈值模型与前文实验效果较好的单一疲劳阈值和视频连续帧平均阈值三种方法模型对比分析。其中对眼部进行算法对比分析，单一阈值选取 0.25，当算法检测到驾驶员眼部实时的 EAR 值在视频连续帧中的两帧均小于阈值，则系统认为驾驶员进行了一次眨眼；平均阈值法取驾驶员在连续的视频图像帧中在眼睛处于睁眼状态下的 EAR 平均值作为判定阈值。其实验结果见表 5-2：

表 5-2 眼部检测算法对比结果

视频编号	实际闭眼数	检测闭眼数/准确率		
		单一阈值	平均阈值	本文方法
1	90	78/86.7%	83/92.2%	89/98.9%
2	157	148/94.3%	142/90.4%	154/98.1%
3	131	97/74.0%	120/91.6%	129/98.5%
4	87	60/69.1%	77/88.5%	85/97.7%
5	106	74/69.8%	98/92.5%	103/97.2%

对驾驶员哈欠进行算法对比，单一阈值法取阈值为 0.9，当算法检测到驾驶员嘴部实时的 MAR 值在视频连续帧中的五帧均大于阈值，则系统认为驾驶员进行了一次哈欠行为；以此类推，平均阈值法阈值选择同眼部平均阈值，其实验结果见表 5-3：

表 5-3 哈欠检测算法对比结果

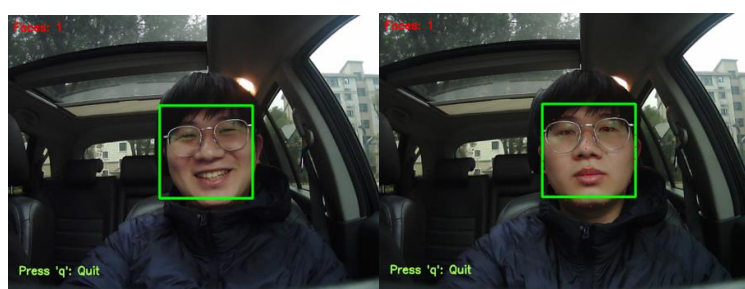
视频编号	实际哈欠数	检测哈欠数/准确率		
		单一阈值	平均阈值	本文方法
1	3	2/66.7%	3/100%	3/100%
2	0	0	0	0
3	2	2/100%	1/50%	2/100%
4	5	6/83.3%	4/80%	6/83.3%
5	4	2/50%	4/100%	4/100%

从实验结果可以看出，采用单一阈值的疲劳辨识方法对某些个体的眨眼检测精度较高（例如：对 2 号视频的驾驶员检测的准确度为 94.3%），但是对于所有的驾驶员这一检测精度不能维持，由于每位驾驶员的面部特征的不一致，导致基于单一阈值的检测方法对实验样本的眨眼检测精度在[69.1%，94.3%]之间，该方法对于不同驾驶员的可移植

性差、系统整体鲁棒性不高。平均阈值的疲劳辨识方法相较于单一阈值的疲劳辨识方法，其整体准确率和泛化能力有所提高，大约稳定在 90%，但这样的准确度依然达不到驾驶员对于疲劳检测系统的要求。而采用本文所提出的基于 SVM 的个性化疲劳阈值方法，相较于以上两种方法，对实验整体样本数据，对不同测试驾驶员的眨眼检测准确率都高于 97.2%。不仅如此，对驾驶员嘴部的哈欠特征行为的检测也拥有优秀的准确度，具有高鲁棒性、强泛化能力和优秀的检测准确度，相比于单一阈值和平均阈值方法，本文提出的方法具有更优秀的检测性能。

5.4.2 驾驶员多特征疲劳检测方法实验

对本文所提出的驾驶员多特征疲劳检测方法进行实验验证，实验选择在户外真实车辆中进行。首先，进行定制疲劳阈值环节。驾驶员首次进入车辆配合系统提示进行简短的特征采集，并针对该驾驶员建立 SVM 个性化疲劳阈值模型。然后，进行人脸检测 and 关键点定位环节。通过相机对驾驶员基于 HOG 算法进行实时的人脸检测，实验结果如图 5-7 (a) 所示，在人脸候选框中基于 ERT 进行人脸关键点定位，结果图如图 5-7 (b) 所示：



(a) 人脸检测候选框 (b) 人脸关键点定位

图 5-7 人脸检测和人脸关键点定位实验结果图

Fig.5-7 The experimental results of face recognition and key point positioning

接着，进行疲劳特征提取环节。通过上一环节算法的输出结果，提取眼睛部分、嘴巴部分和头部姿态疲劳特征，特征间相互独立、补充，当模型在进行实际任务时，出现某种特征提取失效时也拥有其他疲劳特征进行疲劳辨识，其实验结果如图 5-8 所示：



图 5-8 多疲劳特征提取实验结果图

Fig.5-8 Multiple fatigue feature extraction experiment result graph

由实验结果图 5-8 可知，图中左上角的实验结果图为驾驶员眼部眨眼检测，由实验结果可知，算法准确定位了眼睛并将实时计算的眨眼次数标在界面上方。图中右上角和左下角的实验结果图分别为驾驶员的哈欠检测和说话检测，算法准确定位嘴巴轮廓，通过算法实时判断嘴部状态。图中右下角为头部姿态估计实验结果图，算法实时估计头部姿态，并将计算的欧拉角标注在界面左上方。

最后，疲劳辨识环节。在提取了驾驶员多种疲劳特征后，使用 5.2 节的疲劳判定标准进行判定，其中，根据文献研究^[70]，驾驶员眼部 PERCLOS 参数指标与疲劳状态之间的关系如下：

$$PERCLOS \begin{cases} < 0.1, & \text{正常驾驶状态} \\ [0.1, 0.3], & \text{轻度疲劳状态} \\ [0.3, 0.5], & \text{中度疲劳状态} \\ > 0.5, & \text{重度疲劳状态} \end{cases} \quad (5-8)$$

本文对疲劳辨识参数标准的设置如下：

眼部疲劳辨识参数：经实验，本文将眼部 PERCLOS 阈值设为 0.4，time=30 秒，当模型实时计算 $f_{eye} > 0.4$ ，或者闭眼时长大于 3 秒时，系统发出疲劳警报。

嘴部疲劳辨识参数：疲劳辨识模型实时输出驾驶员的 MAR 值，将 MAR 送入针对驾驶员训练后的 SVM 个性化疲劳辨识模型，模型根据 MAR 值实时判定是否出现哈欠行为，在出现哈欠行为时对驾驶员进行疲劳预警。

头部姿态疲劳特征参数：疲劳辨识模型将实时估计的欧拉角送入 SVM 模型，依据输出结果判定实时的头部姿态，若检测到驾驶员出现低头行为，即 $f_{head} > 0$ 时，对驾驶员进行疲劳预警。同时，模型根据欧拉角判定头部姿态是否超过正常范围的 80%，若超过则判定为头部异常角度，即 $pitch > 25^\circ$ 、 $roll > 15.4^\circ$ ，对驾驶员进行预警。

基于自建模拟疲劳数据集对多特征疲劳辨识模型进行实验验证,实验测试样本数据由 4 个疲劳驾驶视频和 1 个正常驾驶视频组成,其中除了 2 号视频,其余都为疲劳驾驶视频,其实验结果如表 5-4 所示:

表 5-4 多特征疲劳辨识模型实验结果

视频 编号	眼部			嘴部				头部		预警 次数
	实际闭 眼数	检测闭 眼数	准确率	实际哈 欠数	检测哈 欠数	实际说话 时间/s	检测说话 时间/s	实际低 头数	检测低 头数	
1	90	89	98.90%	3	3	12	9	5	5	8
2	157	155	98.70%	0	0	20	18	1	0	0
3	131	130	99.20%	2	2	17	13	4	5	7
4	87	85	97.70%	5	6	6	8	8	6	13
5	106	104	98.10%	4	4	18	17	13	15	21

由实验结果可知,本文提出在驾驶员眼部疲劳特征提取上使用 EAR 算法与残差神经网络相补充,相比于只使用基于眼部几何特征的眼睑纵横比来进行眨眼检测,将最低精度再一次提升了 0.5%,在不同驾驶员身上的精度都超过了 97.7%。本文利用驾驶员面部和头部特征建立的多特征疲劳辨识模型在实验测试集中,对存在疲劳驾驶的测试视频都进行了疲劳预警,在正常驾驶的 2 号实验测试视频上没有出现疲劳预警,通过实验验证,证明了本文提出的方法对处于疲劳状态的驾驶员可以进行有效的疲劳预警。

5.5 本章小结

本章首先对基于多特征的驾驶员疲劳检测方法的总体设计结构进行了详细的描述,重点介绍了各部分的功能作用以及驾驶员疲劳检测的实现流程。然后,对所使用的公开及自建的疲劳驾驶数据集进行了介绍,并且制定了本文用于疲劳判定的眼部、嘴部和头部姿态的疲劳参数。同时,阐述了基于 SVM 算法搭建针对不同驾驶员的个性化疲劳阈值模型的具体过程,并与单一阈值、平均阈值的方法对比分析,验证了本文所提方法的优越性。最后通过实验证明了本文提出的基于多特征的疲劳检测方法可以实时有效地对驾驶员进行疲劳检测,在驾驶员出现眼部、嘴部或头部姿态疲劳特征时给予预警,同时提高了在某一特征检测受阻或失效时的普适性。

第六章 总结与展望

6.1 工作总结

据统计,疲劳驾驶是导致交通事故的主要因素之一,为了缓解这一问题,需要提出一种兼顾实时性和高精度的疲劳检测方法对驾驶员进行疲劳检测。本文通过对比分析国内外关于疲劳驾驶检测的研究成果及文献,发现现有的疲劳驾驶检测方法存在疲劳特征参数不够、网络模型不够深和疲劳阈值不能定制等问题,针对这些问题,提出了一种基于多特征的疲劳检测方法,创新性地将多种疲劳特征、深度残差神经网络和 SVM 个性化阈值模型应用于驾驶员疲劳检测中。该方法不仅不会因某种疲劳特征检测受阻而失效,具有无接触、低成本、高精度和高鲁棒性等优点,还能根据不同驾驶员个性化疲劳阈值。本文主要的工作总结如下:

(1) 对比分析现有的驾驶员疲劳检测方法。本文通过相关文献的整理阅读,总结了目前在疲劳驾驶检测领域的四类常见方法,并对其准确度、鲁棒性、接触性和成本进行对比,最后将研究重点放在基于驾驶员行为特征的疲劳检测方法上,并针对目前该方法中存在的不足提出了本文的研究方法。

(2) 驾驶员的人脸检测和关键点定位。本文使用计算机视觉较为优秀的方向梯度直方图算法 HOG 对驾驶员进行人脸区域检测,该方法具有几何不变性和光照不变性的优点,适合在不同光照场景下使用。在人脸关键点定位上,使用了一种新的毫秒级 ERT 人脸对齐算法,该算法能在毫秒级的时间单位上将驾驶员眼部和嘴部融为一体,定位出面部关键点坐标,为后续的疲劳特征提取和头部姿态估计奠定了基础。

(3) 驾驶员头部姿态估计。根据 2D 人脸关键点进行了坐标系转换,建立了图像坐标系与世界坐标系之间的映射关系,并通过相机标定解决了 PnP 问题,实时计算出了驾驶员头部姿态欧拉角。

(4) 提出了使用几何特征和深度残差网络进行疲劳特征提取的方法。通过几何特征,分别对驾驶员眼部和嘴部提取 EAR 和 MAR 特征参数。针对眼部疲劳特征复杂的问题,通过加入深度残差神经网络进行补充,该网络通过使用残差单元解决梯度消失和网络退化等问题,并使用迁移学习算法进行训练,经过迁移训练的网络模型更具针对性、精度更高。

(5) 提出一种针对不同驾驶员的个性化疲劳阈值模型。为了应对实际任务中存在不止一位驾驶员的情况,增强疲劳检测系统的定制功能,本文提出了基于支持向量机 SVM 的疲劳辨识方法,该方法针对不同驾驶员训练的个性化疲劳阈值,与单一疲劳阈值和平均疲劳阈值的疲劳辨识方法相比,具有较高的鲁棒性和检测准确度。

(6) 提出一种多疲劳特征参数判别方法。常用的基于机器视觉的疲劳检测方法大多采用单一疲劳特征进行疲劳判别,当某一疲劳特征检测受限时,系统将失效。本文综合 PERCLOS、眨眼频率、闭眼时长、哈欠次数、低头频率和头部异常角度等疲劳特征参数对驾驶员进行疲劳检测并进行预警,相比于单一疲劳特征参数方法,该方法适用性更广。

6.2 研究展望

虽然本文所提出的基于多特征的疲劳检测方法取得了一定成果,但由于实验设备和时间的限制,该方法还存在一些缺陷和不足,关于本文所提方法的不足以及未来的研究工作如下:

由于条件的受限和现有的针对疲劳驾驶的公开数据集不足,所以本文通过自行招募人员拍摄模拟疲劳视频数据集进行数据补充。但是由于条件的限制,没有封闭路段用以进行实验数据的采集,所以选择了在停止的车辆中进行拍摄采集,与在真实行驶的车辆存在一定差别。后续需要进一步优化数据集的真实性,使数据集尽可能接近真实驾驶环境。

由于本文是基于驾驶员面部图像使用人脸检测和关键点定位算法,当面部出现大范围遮挡时,算法可能会失效。后续需要继续优化驾驶员的人脸检测和关键点定位算法,以保证在出现大范围遮挡时也能检测出人脸定位出关键点。本文对驾驶员只进行了疲劳驾驶检测,在实际驾驶环境中,还存在着例如:打电话、抽烟等危险行为,后续可以将疲劳检测与驾驶员的姿态检测结合起来,检测驾驶员是否存在打电话、抽烟等危险行为,与疲劳驾驶行为一同给予驾驶员危险预警。

参考文献

- [1] 《2021 中国统计年鉴》
- [2] 世界卫生组织. 2011-2020 年道路安全行动十年[DB/OL]. <https://www.who.int/roadsafety/decade of action/plan/plan ch.pdf?ua=1>: 世界卫生组织, March 4,2021.
- [3] Abdullah, M. H. , et al. Driver fatigue detection. International Conference on Information Science and Applications, ICISA 2016, February 15, 2016 - February 18,2016, Minh City, Vietnam, Springer Verlag.
- [4] Thiffault P, Bergeron J. Monotony of road environment and driver fatigue: a simulator study[J]. Accident Analysis & Prevention, 2003, 35(3): 381-391.
- [5] Nobe S A, Wang F Y. An overview of recent developments in automated lateral and longitudinal vehicle controls[C]//2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236). IEEE, 2001, 5: 3447-3452.
- [6] 焦影影. 基于脑电和眼电信号的驾驶员睡意检测研究[D]. 上海交通大学,2019.
- [7] Kurt M B, Sezgin N, Akin M, et al. The ANN-based computing of drowsy level[J]. Expert Systems with Applications, 2009, 36(2): 2534-2542.
- [8] Göran Kecklund, Torbjörn Åkerstedt. Sleepiness in long distance truck driving: an ambulatory EEG study of night driving[J]. Ergonomics, 1993, 36(9): 1007-1017.
- [9] Kecklund G , T Åkerstedt. Sleepiness in long distance truck driving: an ambulatory EEG study of night driving[J]. Ergonomics, 1993, 36(9): 1007-1017.
- [10] Tuncer T, Dogan S, Ertam F, et al. A dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals [J]. Cognitive Neurodynamics(1871-4080), 2021, 15(7): 223-237.
- [11] 陈骥驰,王宏,王翹秀,化成城,刘冲. 基于脑电信号的疲劳驾驶状态研究[J]. 汽车工程,2018,40(05): 515-520.
- [12] Fnu, Rohit, Vinod, et al. Real-time drowsiness detection using wearable, lightweight brain sensing headbands[J]. IET Intelligent Transport Systems, 2017, 11(5): 255-263.
- [13] 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.
- [14] Li G, Chung W Y. Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier[J]. Sensors, 2013, 13(12): 16494-16511.
- [15] 徐礼胜,张闻勘,庞宇轩,吴承暘. 基于短时心电信号的疲劳驾驶检测算法[J]. 东北大学学报(自然科学版),2019,40(07): 937-941.
- [16] 王琳,罗旭,姜鑫,王宏. 基于生物力学和颈腰部 EMG 判别驾驶员疲劳状态[J]. 汽车工程,2017,39(08): 955-960+967.
- [17] Wang M S, Jeong N T, Kim K S, et al. Drowsy behavior detection based on driving information[J]. International journal of automotive technology, 2016, 17(1): 165-173.
- [18] 万蔚,王振华,王保菊. 基于驾驶行为的疲劳驾驶判别算法研究[J]. 道路交通与安全,2016,16(06): 21-24.
- [19] 蔡素贤,杜超坎,周思毅,王雅斐. 基于车辆运行数据的疲劳驾驶状态检测[J]. 交通运输系统工程与信息,2020,20(04): 77-82.
- [20] Kyehoon, LEE, SUNG-AE, et al. Detecting Driver Fatigue by Steering Wheel Grip Force [J]. International Journal of Contents(1738-6764), 2016, 12(1): 44-48.

- [21] 张海兵. 基于方向盘握力的驾驶员状态识别系统设计[D].太原理工大学,2020.
- [22] 李作进,李仁杰,李升波,王文军,成波. 基于方向盘转角近似熵与复杂度的驾驶人疲劳状态识别[J]. 汽车安全与节能学报,2016,7(03): 279-284.
- [23] 柴萌. 长途客车驾驶员疲劳状态辨识与预警[D].吉林大学,2019.
- [24] Catalbas M C, Cegovnik T, Sodnik J, et al. Driver fatigue detection based on saccadic eye movements[C]//2017 10th International Conference on Electrical and Electronics Engineering (ELECO). IEEE, 2017: 913-917.
- [25] Wang M, Guo L, Chen W. Blink detection using Adaboost and contour circle for fatigue recognition [J]. Computers&Electrical Engineering(0045-7906), 2017, 58: 502-512.
- [26] Mandal B, Li L, Wang G S, et al. Towards detection of bus driver fatigue based on robust visual analysis of eye state[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 18(3): 545-557.
- [27] 樊星,刘占文,林杉,窦瑞娟. 融合人眼特征与深度学习的疲劳驾驶检测模型[J]. 计算机工程,2021,47(08): 243-250.
- [28] Song F, Tan X, Liu X, et al. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients[J]. Pattern Recognition, 2014, 47(9): 2825-2838.
- [29] Chen P. Research on driver fatigue detection strategy based on human eye state[C]//2017 Chinese Automation Congress (CAC). IEEE, 2017: 619-623.
- [30] Fatima B, Shahid A R, Ziauddin S, et al. Driver fatigue detection using viola jones and principal component analysis[J]. Applied Artificial Intelligence, 2020, 34(6): 456-483.
- [31] Zhao Z, Zhou N, Zhang L, et al. Driver fatigue detection based on convolutional neural networks using em-cnn[J]. Computational intelligence and neuroscience, 2020, 2020.
- [32] 高宁. 面向驾驶人疲劳检测的人脸分析方法研究[D].大连理工大学,2019.
- [33] Omidyeganeh M, Shirmohammadi S, Abtahi S, et al. Yawning detection using embedded smart cameras [J]. IEEE Transactions on Instrumentation and Measurement (0018-9456), 2016, 65(3): 570-582.
- [34] Alioua N , Amine A , Rziza M . Driver's Fatigue Detection Based on Yawning Extraction[J]. International Journal of Vehicular Technology, 2014, 2014: 47-75.
- [35] Du G , Li T , Li C , et al. Vision-Based Fatigue Driving Recognition Method Integrating Heart Rate and Facial Features[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, PP(99): 1-12.
- [36] Yang H, Liu L, Min W, et al. Driver yawning detection based on subtle facial action recognition[J]. IEEE Transactions on Multimedia, 2020, 23: 572-583.
- [37] Tawari A, Trivedi M M. Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. IEEE, 2014: 344-349.
- [38] 李勇达,张超,孟令君. 基于头部姿态特征的列车司机疲劳驾驶检测系统研究[J]. 交通信息与安全,2014,32(05): 114-119.
- [39] Shen Q, Zhao S, Zhang R, et al. Robust Two-Stream Multi-Features Network for Driver Drowsiness Detection[C]//Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence. 2020: 271-277.
- [40] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.
- [41] Hinton G E , Osindero S , Teh Y W . A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2014, 18(7): 1527-1554.
- [42] Fischer A, Igel C. Training restricted Boltzmann machines: An introduction[J]. Pattern Recognition, 2014, 47(1): 25-39.

- [43] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.
- [44] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of physiology, 1962, 160(1): 106-154.
- [45] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [46] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [47] Pan S J , Qiang Y . A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [48] Tang D, Yang X, Wang X. Improving the transferability of the crash prediction model using the TrAdaBoost. R2 algorithm[J]. Accident Analysis & Prevention, 2020, 141: 105551.
- [49] 余欢欢,魏文戈. 模型参数自适应迁移的多源域适应[J].计算技术与自动化,2019,38(04): 87-90.
- [50] Oquab M, Bottou L, Laptev I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1717-1724.
- [51] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.
- [52] Cootes T F, Taylor C J, Cooper D H, et al. Active Shape Models-Their Training and Application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.
- [53] 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.
- [54] 王子彦. 基于面部表情的儿童社会情绪能力评测研究[D].东南大学,2016.
- [55] V. Kazemi and J. Sullivan, One millisecond face alignment with an ensemble of regression trees[C]// IEEE Conference on Computer Vision and Pattern Recognition, Columbus, New York: IEEE Press, 2014, 1867-1874.
- [56] Sagonas C, Antonakos E, Tzimiropoulos G, et al. 300Faces in-the-wild Challenge: Database and Results [J]. Image and vision computing(0262-8856), 2016, 47: 3-18.
- [57] Zhu X, Ramanan D. Face Detection , Pose Estimation , and Landmark Localization in the wild [C]// IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE,2012: 2879-2886.
- [58] Le V, Brandt J, Lin Z, et al. Interactive Facial Feature Localization [C]// European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012: 679-692.
- [59] Belhumeur P N, Jacobs D W, Kriegman D J, et al. Localizing Parts of Faces Using a Consensus of Exemplars [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (0162-8828), 2013, 35(12): 2930-2940.
- [60] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 Faces in-the-wild Challenge:The First Facial Landmark Localization Challenge [C]// Proceedings of the IEEE International Conference on Computer Vision Workshops.2013: 397-403.
- [61] Rafael C, Richard E. Digital image processing[M]. PHI.2017.
- [62] Zhang Z. A flexible new technique for camera calibration [J]. IEEE Transactions on pattern analysis and machine intelligence(0162-8828), 2000, 22(11): 1330-1334.
- [63] Xu D, Tan M, Jiang Z, et al. A shape constraint based visual positioning method for a humanoid robot[J]. Robotica, 2006, 24(4): 429-431.
- [64] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Las Vegas,NV,2016.770-778.

- [65] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks?[C]// International Conference on Neural Information Processing Systems. MIT Press, 2014. 3320–3328.
- [66] Song F, Tan X, Chen S, et al. A literature survey on robust and efficient eye localization in real-life scenarios [J]. Pattern Recognition(0031-3203), 2013, 46(12): 3157-3173.
- [67] Abtahi S, Omidyeganeh M, Shirmohammadi S, et al. YawDD: A yawning detection dataset[C]// ACM Multimedia Systems (MMSys). ACM, 2014.24-28.
- [68] Ferrario V F, Sforza C, Serrao G, et al. Active Range of Motion of the Head and Cervical Spine: A Three-dimensional Investigation in Healthy YoungAdults [J]. Journal of Orthopaedic Research (1554-527X), 2002, 20(1): 122-129.
- [69] 陆荣秀,张笔豪,莫振龙. 基于脸部特征和头部姿态的疲劳检测方法[J/OL].系统仿真学报:1-13.21-0583.
- [70] 汪磊,孙瑞山. 基于面部特征识别的管制员疲劳监测方法研究[J].中国安全科学学报,2012,22(07): 66-71.

致谢

圣诞节早晨，寒风瑟瑟，站在每日都要经过数十次的 19 栋回廊上，看向窗外的双港东大街的三岔路，我的视线被灰白色的屏障所阻隔，不知是雾还是霾。白驹过隙，路行至此，恍惚间我的 7 年校园生涯也进入了尾声。回想起这 7 年的求学之路，思绪万千，五味杂陈，从带着青涩脸庞，穿着褪色军训服在操场上的欢声笑语，到图书馆日日夜夜的埋头苦学，最后在实验室里敲击着键盘将这篇文章带到你们面前。2015 年，我拖着重重的行李箱从南方偏远小城乘坐近 10 小时的火车来到这里，不知不觉，我已经离开家乡独自咬牙走了一段很长的路，今当离别，临表涕零，尽管我还是很讨厌这里的潮湿和妖风，但我在这里有幸遇到的老师、朋友和我的女朋友共同编织成了我生命中最美的记忆。

我很感谢陆荣秀老师在我身处谷底而不自知时温柔地将我拉起，带我选择了交通运输工程专业，将我领入了实验室这个大家庭。这三年无疑是我人生中成长最快的三年，在这个大家庭里，我过得充实且幸福。在科研中，从研究的选题、新冠疫情在家进行文献阅读的指导、到一遍又一遍耐心的修改润色着我的论文，陆老师对待学术的一丝不苟深深的感染了我，除了专业知识的传授，陆老师同样向我传授了为人处事的积极态度，使我受用终生。在生活中，由于寒冷潮湿和久坐，我时常腰痛，陆老师不仅督促着我进行锻炼恢复健康，还在我疼痛急性期时给我拿来膏药，如同异乡的母亲一般给予了我太多的关心与爱，遗憾的是，我距离陆老师对我的期望还有一些距离，希望我能带着老师的这份期望，在今后的工作岗位中，用另一种方式弥补这个遗憾。同时还要感谢杨辉老师、李中奇老师、杨刚老师、朱建勇老师和实验室的其他老师，老师们总能发现我的问题，并为我答疑解惑。在这里衷心的对老师们说一句：谢谢！

很感谢赖路璐、何权恒、饶运春、陈明明、陈赫鹏、徐彬、黄学文、凤冰霞、陶力宏、黄伟等实验室的师兄师姐在科研道路和生活中给予的帮助，特别感谢赖路璐师姐在我一无所知时教会我如何进行科研写作、投稿论文等等，以及日常给予的处处照顾与关心，感谢同门兄弟蒋松、丁康豪以及肖孙博、刘洪亮、邓彪、李洁、方立坚、刘晓霞师弟师妹带给我的快乐与包容。还要感谢李明振、裴结安两位好兄弟对我科研上的帮助以及生活中的照顾。

这里还需要感谢从大一就一直陪伴我的女朋友程锦，从书画协会中相识，到肩并肩一起读研，到如今一步步互相支持着离开校园走向社会，你总在我遇到困难时陪着我，一起迈过一道又一道坎，希望我们能携手迈过面前的一又道坎，如你的名字般，前程似锦。最后，感谢养育我的父母，在外求学进十载，毕业依旧离家漂泊，慈母倚门情，游子行路苦，父母的养育之恩用文字难以形容，需要我用一生去守护！

3.3 基于 PnP 问题的驾驶员头部姿态估计.....	28
3.3.1 人体头部姿态分析.....	28
3.3.2 相机标定.....	29
3.3.3 基于 PnP 求解头部姿态.....	32
3.4 本章小结.....	34
第四章 基于面部信息和头部姿态的驾驶员疲劳特征提取.....	35
4.1 基于驾驶员眼部疲劳特征的提取方法.....	35
4.1.1 驾驶员眼部疲劳特征概述.....	35
4.1.2 驾驶员眼部疲劳特征提取技术路线.....	35
4.1.3 基于 EAR 的眼部疲劳特征提取.....	35
4.1.4 基于深度学习的眼部疲劳特征提取.....	36
4.1.5 实验结果与分析.....	39
4.2 基于驾驶员嘴部疲劳特征的提取方法.....	41
4.2.1 驾驶员嘴部疲劳特征概述.....	41
4.2.2 驾驶员嘴部疲劳特征提取技术路线.....	41
4.2.3 基于 MAR 的嘴部疲劳特征提取.....	42
4.3 基于驾驶员头部姿态的疲劳特征提取方法.....	43
4.4 本章小结.....	44
第五章 基于多特征的疲劳状态检测方法.....	45
5.1 基于多特征的疲劳检测方法总体结构.....	45
5.2 建立疲劳模拟数据集及疲劳判定标准.....	47
5.2.1 疲劳驾驶数据集.....	47
5.2.2 疲劳驾驶参数判定标准.....	47
5.3 基于支持向量机的驾驶员个性化疲劳检测方法.....	49
5.3.1 支持向量机原理.....	49
5.3.2 基于 SVM 个性化疲劳辨识模型建立.....	50
5.4 驾驶员疲劳辨识实验与分析.....	52
5.4.1 SVM 个性化疲劳辨识模型实验.....	52
5.4.2 驾驶员多特征疲劳检测方法实验.....	54
5.5 本章小结.....	56
第六章 总结与展望.....	57
6.1 工作总结.....	57
6.2 研究展望.....	58
参考文献.....	59
个人简历 在读期间发表的学术论文.....	63

致谢.....	64
---------	----

第一章 绪论

1.1 研究背景与意义

随着经济的飞速发展，人民的生活质量在不断提高，汽车的数量也在不断增加。根据中国汽车工业协会发布的中国汽车市场的产销数据，2021 年 1、2 季度，全国汽车销售 1036.32 万辆，相比上年同一时期增加了 31.5%，生产和销售总量以一骑绝尘之势连续 12 年蝉联世界第一。《2021 年中国统计年鉴》显示^[1]，我国民用汽车保有量从 2005 年末至 2020 年逐年上升，15 年间从 3159.66 万辆增加到 27340.92 万辆，年均增长率为 15.6%（见图 1-1）。我国民用汽车保有量不仅基数庞大，且在 2020 年又新增了近 2400 万辆民用车辆。随着我国以民用汽车为代表的各类交通车辆数量的有增无已，所带来的安全隐患和导致的道路交通事故也在与日俱增，所以公众对驾驶员人身安全的关注度也随之上升。

交通事故对驾驶员造成的人身伤害和财产损失一直是全世界所面临的问题，如今道路交通事故已经成为全球人口伤亡最大的意外事件之一，WHO 的数据显示^[2]，每年全球因交通事故死亡的人数约有 135 万人，即每小时约有 158 人丧生，3.5 千万人受伤致残。由于我国人口占世界人口的 1/5，且汽车保有量位居世界第一，所以我国在道路交通安全问题上较世界其他国家基数更为庞大。在我国，交通事故导致的人员伤亡和财产损失都颇为严重，表 1-1 为国家统计局发布的 2011 年至 2020 年的道路交通事故数据。由表 1-1 可知，在过去 10 年间，交通事故发生总数高达 2151147 起，导致伤亡的人数达 2900667 人，造成直接的经济损失大约为 118.7 亿元。

分析表 1-1 数据可知，近 10 年来我国交通事故发生数高居不下，造成的直接经济损失呈上升趋势。近年来，经过国家的积极处理和整治，情况有了些许好转，但是由于我国人口基数庞大，形势依然不容乐观。在这些交通事故数据的背后，大多数是由于驾驶员的危险驾驶行为导致，例如：酒后驾驶、疲劳驾驶等。其中，疲劳驾驶是引发交通事故的主要因素，约占总事故数量的百分之二十。美国 NTSB 和 NHTSA 的调查数据表明，每年由疲劳驾驶导致的事故约有 100000 起，共造成 70000 人受伤^[3]。在这类事故中，由于存在驾驶员在事故中死亡或者在幸存以后隐瞒疲劳驾驶的情况，且没有明确的疲劳驾驶判定标准，所以统计的事故数可能只是冰山一角，实际数据要远大于统计数据。

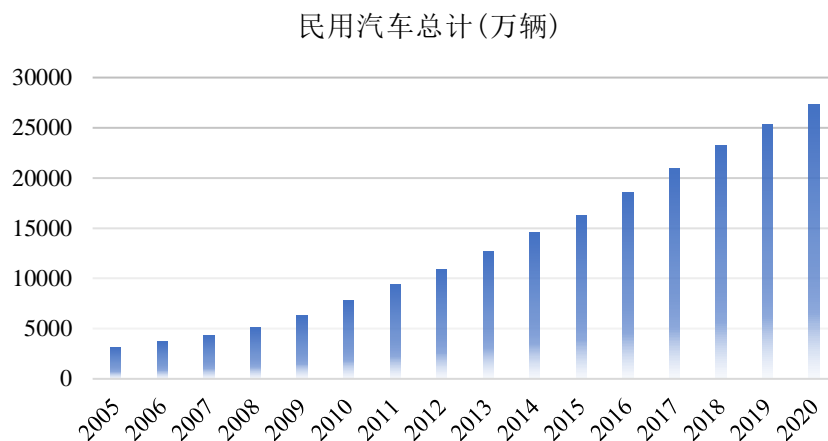


图 1-1 2005 年至 2020 年全国民用汽车保有量

Fig.1-1 National civil vehicle ownership from 2005 to 2020

表 1-1 我国 2011 年至 2020 年交通事故状况

年份	交通事故发生数 (起)	死亡人数 (人)	受伤人数 (人)	直接财产损失 (万元)
2011	210812	62387	237421	107873.0
2012	204196	59997	224327	117489.6
2013	198394	58539	213724	103896.6
2014	196812	58523	211882	107542.9
2015	187781	58022	199880	103691.7
2016	212846	63093	226430	120759.9
2017	203049	63772	209654	121311.3
2018	244937	63194	258532	138555.9
2019	247646	62730	256101	134617.9
2020	244674	61703	250723	131360.6

疲劳驾驶是驾驶员在行驶的车辆中产生疲劳,注意力下降,从而对机动车控制能力下降的一种危险驾驶行为。在实际驾驶环境中,驾驶员产生疲劳的原因可分为内部因素和外部因素^[4],内部因素可能是驾驶员患有引起大脑供血不足的疾病(如:贫血、颈椎病、低血压和脑血管病变等)或者患有内分泌和精神类疾病(如:糖尿病、甲状腺疾病和抑郁症等),此外,患有睡眠呼吸暂停综合症或者夜班熬夜都可能导致驾驶员在驾驶机动车过程中疲劳瞌睡;外部因素可能是驾驶员驾驶机动车长时间行驶在高速公路这种单一路况,环境单调,视觉敏感度降低导致疲劳,或者天气闷热等。国家相关部门和网约车平台对预防疲劳驾驶进行了相应措施,如:《道路交通安全法实施条例》中规定了司机的最大连续驾驶时间;网约车平台也对网约车司机的最大连续驾驶时间进行监控,

若移动应用端检测到司机长时间驾驶未休息，平台将强制网约车司机下线二十分钟以保证乘客和司机的安全。

疲劳状态因人而异，仅仅从驾驶时长去界定疲劳驾驶不能在根本上预防交通事故。无论驾驶员是由于内部因素还是外部因素导致的疲劳，都有可能在道路上酿出恶果，对自身家庭和国家造成不可弥补的损失。但是在疲劳驾驶引发交通事故之前，驾驶员会有明显的前期疲劳特征，此时对驾驶员进行疲劳预警可以很大程度上避免交通事故的发生。研究显示，如果驾驶员在事故发生前一秒被警告疲劳，60%与疲劳有关的交通事故是可以预防的^[5]。然而，目前对于疲劳预警的研究仍存在很大空缺，所以，研究如何在事故发生前提醒驾驶员注意疲劳驾驶，对于提高交通安全和对人民群众对美好生活的渴望具有重要意义。

1.2 疲劳驾驶检测方法国内外现状

在当今社会，由疲劳驾驶引发的交通事故率在全球范围内高居不下，已然成为一种严重的社会问题。现如今，研究如何对处于疲劳状态的驾驶员进行检测预警已成为全球专家学者研究的热点，受到了各国家政府和人民的重视。在疲劳驾驶检测方面，国内开始研究的时间要晚于国外，但是得益于国家和政府的重视，我国近几年对疲劳驾驶检测进行研究的企业和人员越来越多，成果也在不断涌现。

早期对人体的疲劳检测主要存在于医学领域的实验室，通过相关的医疗器械对人体的生理指标进行分析，确定疲劳程度。目前，交通领域的疲劳检测方法主要有主观检测和客观检测两种^[6]，通过疲劳评定表和睡眠调查表来判断疲劳程度的方法称为主观检测法。客观检测分为两类，一类是以车辆为载体，基于车辆的行驶信息对车辆的加速度、方向盘转角进行分析，间接判定驾驶员是否疲劳；一类是以驾驶员为载体，基于驾驶员的生理信号参数对驾驶员脑电、心电和肌电信号进行分析进而判断驾驶员是否疲劳；和基于驾驶员行为特征对驾驶员面部疲劳特征和头部姿态疲劳特征进行检测进而判断是否疲劳。常见的疲劳检测方法如图 1-2 所示。