分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2210330152

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于深度学习的疲劳驾驶检测技术研究

论文作者　张兴旺
校内导师　王凤随（教授）
校外导师　郭振

专业学位类别　电子信息
研究方向（领域）　智能信息处理及应用

论文提交日期：2024 年 5 月 31 日

分类号：TP391　　单位代码：10363

密　级：公开　　学　　号：2210330152

题　目　基于深度学习的疲劳驾驶检测技术研究

题　目　Research on Fatigue Driving Detection Technology Based on Deep Learnin

学　生　姓　名：张兴旺

校　内　导　师：王凤随（教授）

校　外　导　师：郭振

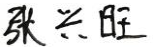
专　　业：电子信息

研　究　方　向：智能信息处理及应用

论文答辩日期：2024 年 5 月 23 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名： 

日期：　　2024 年 5 月 28 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 √。

学位论文作者签名：张兴旺 指导教师签名：王凤随

日期： 2024 年 5 月 28 日 日期： 2024 年 5 月 28 日

基于深度学习的疲劳驾驶检测技术研究

摘 要

随着经济的不断发展，汽车已成为人们出行中不可或缺的交通工具之一，但随着汽车数量的快速增加，交通事故、交通拥堵等问题也频繁发生。在导致交通事故的原因中，疲劳驾驶是其中的主要因素之一。然而，当前疲劳驾驶检测存在实时性不足和准确度低等问题，为此，本文提出了基于深度学习的驾驶员疲劳检测算法。算法采用了检测速度较快的单阶段检测算法 YOLOv5s 和 YOLOv7 作为基准模型，并引入了注意力机制和下采样卷积等技术来实现多特征信息融合，提高算法的准确率，为快速、准确地检测驾驶员的疲劳状态提供了有效的帮助。具体研究内容和工作如下：

（1）针对目前驾驶员疲劳驾驶检测不能兼顾检测速度与检测准确率的问题，提出了一种基于 YOLOv5s 改进的疲劳驾驶检测算法。首先，为提高模型的收敛速度，增强模型的检测性能，算法将深度重参数化卷积层替换传统的卷积层，通过增加可学习的参数加快拟合过程。其次，为了提高模型捕获不同尺度特征之间的空间和语义信息和对物体结构和上下文的理解能力，在模型特征表达性能较弱的深层网络中加入高效多尺度注意力机制，使得模型可以自适应地调整对不同尺度信息的关注程度，从而提升模型的性能和鲁棒性。

实验结果表明，算法在 WIDER FACE 数据集上达到了 76.8%，较基线算法提升了 2.2%，显著提升了对人脸目标检测的精度。

（2）针对复杂环境下对人脸检测精度低的问题，提出了一种基于 YOLOv7 改进的疲劳驾驶检测算法。首先，为解决在有遮挡目标的场景下传统的下采样过程容易导致特征丢失严重的问题，以及提高对遮挡情况下目标检测的准确度，算法引入了基于 SE 注意力改进的 DS-Conv 模块，通过对每个通道进行自适应的特征加权，使得模型能够更有针对性地保留重要的特征，将更多的注意力集中在对当前任务有用的通道上，从而增强模型的表达能力和性能。其次，为了提高模型对小目标的检测能力，算法在特征提取层中引入了多尺度特征提取 MSS 注意力模块，能够在不同尺度上捕捉目标的细节和上下文信息，使其在各种场景中更加有效和可靠。实验结果表明，改进算法在 WIDER FACE 数据集的 Easy，Medium，Hard 子集上分别达到了 96.0%，94.6%，88.1%，大幅度提高了复杂环境下的人脸检测精度。

（3）针对驾驶员处于疲劳状态时，通常会表现出的一些特征，例如闭眼和打哈欠等，设计了一套驾驶员疲劳检测系统。首先，通过安装在方向盘管柱上的摄像头输入视频数据，之后根据输入的图像检测人脸关键点点位，并计算得出眼嘴部张开大小和眼嘴部纵横比，判定眼嘴部状态。最后对眼部和嘴部的不同状态进行识别，并实时计算单位时间内的 PERCLOS（闭眼时间占比）和打哈欠次数等

疲劳参数。通过这些参数的计算结果，综合对驾驶员的疲劳状态进行判定，这种综合判定方法能够更准确地评估驾驶员的疲劳程度。通过实验验证，提出的驾驶员疲劳检测系统能有效识别出驾驶员的疲劳状态，并完成疲劳检测任务，检测准确率达到了 94%，保障了驾驶员的生命财产安全。

关键词：疲劳驾驶检测，深度学习，目标检测，多尺度注意力，多特征信息融合

RESEARCH ON FATIGUE DRIVING DETECTION TECHNOLOGY BASED ON DEEP LEARNING

ABSTRACT

With the continuous development of economy, vehicles have become one of the indispensable means of transportation for people to travel. However, with the rapid increase in the number of cars, problems such as traffic accidents are also rising. Among the causes of traffic accidents, fatigue driving is one of the main factors. However, the current fatigue driving detection system has problems such as insufficient real-time performance and low accuracy. Therefore, this paper proposes a driver fatigue detection algorithm based on deep learning. The algorithm adopts the fast single-stage detection algorithms YOLOv5s and YOLOv7 as benchmark models, and introduces the attention mechanism and downsampling convolution to realize the fusion of multiple feature information, improve the accuracy of the algorithm, and provide effective help for quickly and accurately monitoring the fatigue state of drivers. The specific research content and work are as follows:

(1) Aiming to solve the problem that the current driver fatigue driving detection cannot take into account the detection speed and accuracy, an improved fatigue driving detection algorithm based on YOLOv5s is proposed. Firstly, in order to improve the convergence speed and enhance the detection performance of the model, the algorithm replaces the traditional convolutional layer with a deep over-parameterized convolutional layer, which accelerates the fitting process by adding

learnable parameters. Secondly, in order to improve the ability of the model to capture the spatial and semantic relationships between features of different scales and improve the ability to understand the structure and context of the objects, an efficient multi-scale attention mechanism is added to the deep network layer, which has a weaker performance in the expression of the model's features, so that the model can adaptively adjust the degree of attention to the information of different scales, so as to improve the performance and robustness of the model. The experimental results show that the algorithm achieves 76.8% on the WIDER FACE dataset, which is 2.2% higher than that of the baseline algorithm, significantly improving the accuracy of face target detection and ensuring the driver's safety.

(2) Aiming to solve the problem of low accuracy of face detection in complex environments, an improved fatigue driving detection algorithm based on YOLOv7 is proposed. Firstly, to solve the problem that the traditional downsampling process tends to lead to serious feature loss in the scene with occluded targets, and to improve the accuracy of target detection in the case of occlusion, the DS-Conv module based on SE attention improvement is introduced, which enables the model to retain the important features in a more targeted way by adaptively weighting the features of each channel, and focus more attention on the channels useful for the current task, thus enhancing the model's expressiveness and performance. Secondly, to improve the model's ability to detect small targets, the algorithm introduces a multi-scale feature extraction MSS attention module to the feature extraction layer, which is able to capture the details and contextual information of the target at different scales, making it more effective and reliable in various scenarios. The experimental results show that the improved algorithm achieves 96.0%,

94.6%, and 88.1% on the Easy, Medium, and Hard subsets of the WIDER FACE dataset, respectively.

(3) A driver fatigue detection system was designed to focus on some characteristics that drivers usually show when they are in a state of fatigue, such as eyes closing and yawning. Firstly, video data is input through the camera installed on the steering wheel tube. Then, the key points of the face are detected based on the input image, and the size of the eye-mouth opening and the eye-mouth aspect ratio are calculated to determine the state. Finally, the different states of the eyes and mouth are identified, and the fatigue parameters such as PERCLOS (proportion of eye closing time) and the number of yawning are calculated in real-time. By integrating the calculation results of these parameters, the driver's fatigue state can be determined, and this comprehensive judgment method can evaluate the driver's fatigue degree more accurately. Through experimental verification, the driver fatigue detection system can effectively identify the driver's fatigue state and complete the fatigue detection task, with the detection accuracy of 94%, ensuring the safety of the driver's life and property.

KEY WORDS: fatigue driving detection, deep learning, target detection, multiscale attention, multi-feature information fusion

目录

第 1 章 绪论

1.1 研究背景及意义

疲劳驾驶指在长时间驾驶后，由于身体和大脑的疲劳，导致驾驶员的反应能力和判断能力下降，从而增加了交通事故的风险。疲劳驾驶已经成为世界范围内的交通安全问题，每年造成大量的人员伤亡和财产损失。因此，研究疲劳驾驶的检测方法和防止措施具有重要的意义。

疲劳驾驶是一个多因素的问题，受到驾驶者的身体状况、驾驶环境、行车时间等多种因素的影响。Beirness 等人调查并统计了经历过疲劳驾驶的驾驶员的相关原因。研究结果显示，驾驶时产生疲劳的驾驶员通常存在以下情况：没有达到充足的 8 小时睡眠时间，或者他们的睡眠质量较差[1]。在长时间驾驶后，驾驶者的身体和大脑会出现疲劳的症状，包括注意力不集中、反应迟钝、眼睛干涩等[2]。这些症状会严重影响驾驶者的驾驶能力，导致他们无法及时判断和应对道路情况，从而增加了交通事故的发生概率。

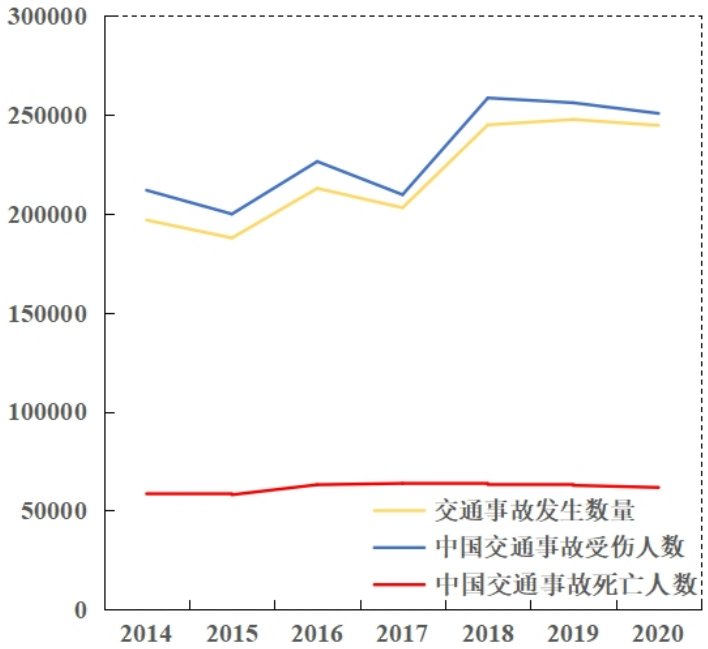


图 1-1 中国交通事故发生数量以及受伤和死亡人数

根据世界卫生组织的数据[3]，疲劳驾驶是导致交通事故的最主要原因之一，仅次于酒驾和超速驾驶[4]。其中，驾驶员疲劳驾驶占交通事故总数的 20%，占大型卡车和公路交通事故总数的 37%，占重大交通事故总数的 43%。根据社会

调查，疲劳驾驶在司机的行车过程中已经成为常态。根据国家统计局数据[5]，如图 1-1 所示为我国在 2014 年至 2020 年交通事故发生数量和交通事故中死亡人数及受伤人数，平均每年发生 22 万起交通事故，死亡人数和受伤人数分别达到了每年 6 万人和 23 万人。在美国，疲劳驾驶每年导致约 4 万起交通事故，其中包括约 1.5 万起严重事故，造成约 7 万人受伤和 1500 人死亡。在欧洲，疲劳驾驶每年导致约 5 万起交通事故，其中包括约 2 万起严重事故，造成约 3 万人受伤和 1000 人死亡。可见，疲劳驾驶已经成为全球性的交通安全问题[6]，为了保护人们的生命财产安全，找到一种实时、高效和简易的疲劳驾驶检测方法，检测和提醒驾驶员疲劳驾驶行为，从源头解决由于疲劳驾驶导致的交通事故是非常有必要的。总的来说，研究疲劳驾驶的检测方法和防止措施具有重要的意义，可以有效地保障人们的交通安全，减少交通事故的发生。

1.2 国内外研究现状

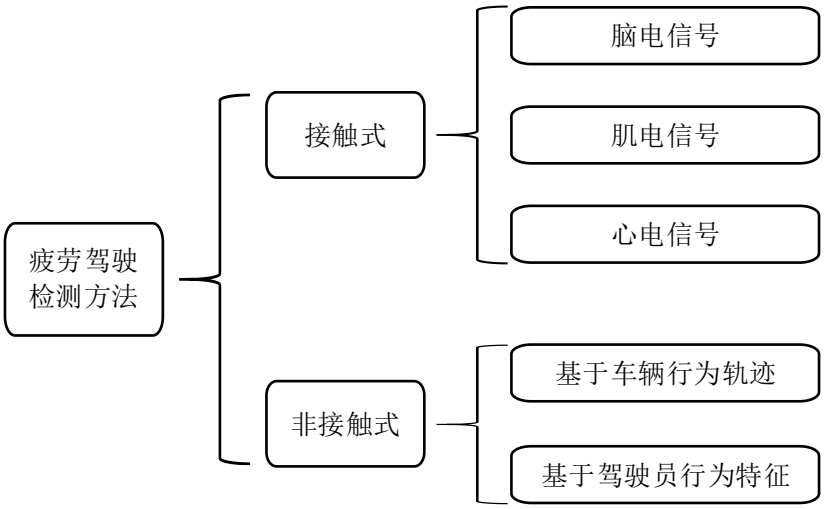


图 1-2 疲劳驾驶检测方法

如图 1-2 所示为目前疲劳驾驶检测的相关方法[7]。疲劳驾驶行为检测方法主要分为两种，区分的标准在于是否需要和驾驶员进行接触：第一种是需要进行接触的方法，统称为接触式检测[8]，另外一种是不需要和驾驶员进行接触的检测方法，统称为非接触式检测[9]。接触式检测方法主要是设计一款检测设备通过直接接触人体来采集相关数据进行疲劳状态的检测，比如脑电信号、肌电信号和心电信号等，这类方法称为基于驾驶员生理信号的疲劳检测方法。对于非接触式检测方法来说，又可划分为两类，一类是以车辆为载体，称为基于车辆

行为轨迹的疲劳检测方法；另一类是以驾驶员行为为载体，称其为基于驾驶员行为特征的疲劳检测方法。

疲劳驾驶检测可以采用接触式检测和非接触式检测两种方法。接触式检测通常需要驾驶员与检测设备直接接触。如通过眼动仪追踪驾驶员的眼球运动来检测疲劳迹象，使用头戴式设备来监测头部姿势和头部运动以判断驾驶员是否出现头部下垂等疲劳姿势，通过监测生理信号，如心率、皮肤电阻等，来评估驾驶员的疲劳状态。非接触式检测方法可以在不需要驾驶员与设备直接接触的情况下进行监测。如利用车辆内部或道路上的摄像头来捕捉驾驶员的面部特征，如眼睛闭合、头部姿势等，以判断疲劳驾驶行为。其次，还可以通过分析车辆操作数据，如方向盘的微小变化、车道偏离等，来识别可能的疲劳驾驶迹象。无论是接触式检测还是非接触式检测，都有其优缺点。接触式检测通常可以提供更准确的结果，但需要特定的设备和驾驶员的配合。非接触式检测更为方便，但可能受到环境条件和摄像头角度等因素的影响，可能存在一定的误判率。因此综合考虑实际应用的需求和条件，选择适合的疲劳驾驶检测方法是很重要的。

1.2.1 基于驾驶员生理特征信号的疲劳驾驶检测方法

研究表明，当驾驶员处于疲劳状态时，他们的脑电信号（EEG）、心电信号（ECG）、肌电信号（EMG）等生理信号与清醒状态下存在着明显的差异。这些差异可以用来判断驾驶员的疲劳程度。通过分析和比较这些生理信号的变化，可以获得与疲劳相关的特征，进而进行疲劳驾驶的检测和预警，以帮助预防和减少因驾驶员疲劳而引起的交通事故。

Jap 等[10]人的研究发现，调查的四种 EEG 活动与疲劳驾驶之间存在着密切的关联。另外，Min 等[11]人提出一种基于前额脑电图的快速有效的多熵特征驱动疲劳识别框架，该框架采用了混合模型，利用了几种常见的熵来提高脑电信号的特征质量，实验结果表明，所提出的方法既方便又有效。同时，由于脑电信号能够直接反映大脑的状态变化，因此 EEG 被认为是最有效、最可靠的人体生理指标之一[12]。这些研究结果表明，脑电信号的分析可以用于检测和评估驾驶员的疲劳状态。

当驾驶员处于疲劳状态时，可以通过分析心率的变化来评估其疲劳程度。Halomoan 等[13]人提出了驾驶疲劳检测框架中的特征提取和特征选择的改进。在特征提取阶段，利用心率片段提取非线性特征来分析驾驶员的认知状态，这些特征与从心率变异性分析中获得的时域、频域和非线性域特征相结合。在特征选择阶段，利用互信息过滤冗余特征。实验结果表明，具有 44 个选定特征的随机森林算法的模型性能最佳。

通过利用检测驾驶员肌肉电活动来评估其疲劳程度是检测疲劳驾驶的一种方法。肌电信号是由肌肉收缩和放松过程中产生的电活动产生的生物电信号。通过监测和分析肌电信号的特征和变化，可以推断驾驶员的肌肉疲劳程度和身体疲劳状态。Balasubramanian 等[14]人通过实验发现，当 EMG 信号变弱，肌电幅值增大时，疲劳程度越深。通过肌电信号检测疲劳驾驶有以下优点：肌电信号检测是通过佩戴肌电传感器获取驾驶员的肌肉电活动，不需要进行任何侵入性操作，对驾驶员的身体没有任何伤害；肌电信号可以实时获取和分析，能够提供及时的疲劳状态评估结果，有助于驾驶员及早采取相应的措施。然而，通过肌电信号检测疲劳驾驶也存在一些缺点：肌电信号的获取和处理需要专门的传感器和设备，并且需要进行信号处理和分析，这对于一般的驾驶情境可能不太方便；肌电信号受到多种因素的影响，如肌肉疲劳、情绪状态、肢体运动等，这些因素的干扰可能会导致疲劳评估的准确性有所降低；不同驾驶员的肌电信号可能存在差异，因此需要建立个体化的模型和基准来进行准确的疲劳评估。综上所述，通过肌电信号检测疲劳驾驶虽然具有非侵入性和实时性等优点，但需要专门的设备和信号处理，并且受到多因素影响和个体差异的限制，因此在应用肌电信号检测疲劳驾驶时需要考虑这些因素并综合其他指标进行综合评估。

通过生理特征信号检测疲劳驾驶是一种利用驾驶员的生理信号来评估其疲劳状态的方法。生理特征信号可以提供细微的变化，具有较高的灵敏度和准确性，通过对驾驶员的生理特征信号进行实时监测和分析，可以及时发现疲劳迹象并提供预警，帮助驾驶员采取相应的措施，从而减少疲劳驾驶带来的安全风险。不过检测生理特征信号也有一下缺点：第一，生理特征信号的获取需要专门的传感器和设备，并且可能需要进行信号处理和分析，增加了系统的复杂性和成本。第二，不同驾驶员的生理特征信号可能存在差异，因此需要建立个体

化的模型和基准来进行准确的疲劳评估。第三，生理特征信号受到多种因素的影响，如情绪状态、环境条件等，这些因素的干扰可能会降低对疲劳评估的准确性。第四，生理特征信号由于需要接触式采集，因此可能对驾驶员在驾驶过程中造成影响。

1.2.2 基于车辆行为特征的疲劳检测

基于车辆行为特征的疲劳检测是一种通过分析驾驶行为和车辆运动的特征来评估驾驶员疲劳程度的方法。下面是一些常用的基于车辆行为特征的疲劳检测方法：方向盘运动特征、车辆位置和偏离特征、加速度和减速度特征、车辆稳定性特征、加速踏板和制动踏板操作特征等。

疲劳驾驶时，驾驶员的方向盘操作可能变得较为模糊或缓慢。通过监测方向盘的转动角度、转动速度、转动力度等特征，可以评估驾驶员的疲劳程度。Lawoyin 等[15]人提出采用三轴加速度计，提供一种经济、简便、有效的方法来监测方向盘运动，而无需对现有车辆装置进行任何改动。模拟和实验结果表明，加速度计测量到的方向盘旋转角度与电位计记录到的方向盘实际旋转角度呈线性相关，因此加速度计是监测方向盘运动以检测瞌睡驾驶的有用工具。Devi 等[16]人提出从实时方向盘角度时间序列的固定滑动窗口中提取近似熵特征，计算样本数据线性特征序列之间的翘曲距离，利用翘曲距离判断驾驶员是否处于疲劳状态。Tian[17]提出了一种基于方向盘角速度的疲劳检测方法。通过对驾驶员疲劳状态下的转向行为进行分析，在时间检测窗口中，应用可变性和程度约束，对检测特征进行判断，识别驾驶员的疲劳状态。

其次，在疲劳驾驶中，驾驶员可能更容易让车辆偏离车道中心线或频繁变道。通过监测车辆的位置偏离、车道变更次数、车道保持能力等特征，可以推断驾驶员的疲劳情况。Zhang 等[18]人提出了基于 Hough 改变的车道检测，以此判断驾驶员是否处于疲劳状态。除此之外，当出现疲劳驾驶，驾驶员还可能会出现加速度和减速度的变化。通过分析车辆的加速度和减速度模式、频率、幅度等特征，可以间接反映驾驶员的疲劳状态。或者由于疲劳驾驶可能会导致驾驶员对车辆的控制能力下降，车辆的稳定性可能受到影响，因此通过分析车辆的姿态、横摆角速度、侧倾角等特征，可以评估驾驶员的疲劳水平。

然而，基于车辆行为特征的疲劳检测方法也存在一些限制：车辆行为特征可能受到其他因素的干扰，如道路条件、交通情况等，这可能导致误判或不准确的结果；不同驾驶员的行为习惯和驾驶风格可能存在差异，因此需要建立个体化的模型和基准来进行准确的疲劳评估。因此，基于车辆行为特征的疲劳驾驶检测技术准确性和可靠性不高[19]。

1.2.3 基于人脸特征的疲劳检测

驾驶员开始进入疲劳状态后，将出现不同表现的身体特征[20]，基于人脸特征的疲劳检测利用计算机视觉技术来分析驾驶员的面部表情、眼睛状态和头部姿态等特征，从而判断其疲劳程度。Li 等[21]人提出基于面部行为分析的驾驶员疲劳检测方法，通过检测眼睛和嘴巴的状态，判断驾驶员是否处于疲劳状态。

基于眼部的疲劳特征检测。眼部特征检测是一种常用的方法，通过分析驾驶员的眼睛状态和眼部行为来判断其疲劳程度。以下是一些常用的眼部特征检测方法：通过分析驾驶员眼睛的开闭状态，可以判断是否存在疲劳。长时间的眼睛闭合或频繁的眨眼间隔可能暗示驾驶员处于疲劳状态；通过检测和跟踪眼球的运动，可以获取关于驾驶员注意力和疲劳状态的信息，例如，当驾驶员眼球的运动范围较小或缓慢时，可能表明注意力不集中或疲劳；分析驾驶员眼睑的活动情况，如眨眼频率、眼睑的开合速度等，较低的眨眼频率和缓慢的眼睑运动可能与疲劳相关；瞳孔的大小可以反映驾驶员的注意力和兴奋程度，当驾驶员疲劳时，瞳孔通常会变小，因为疲劳会导致注意力下降；疲劳状态下，眼部周围的肌肉可能出现松弛或下垂的情况，通过分析驾驶员眼部的皱纹、褶皱、眼袋等特征，可以提供关于疲劳状态的线索。值得注意的是，眼部特征检测方法仅作为一种指示疲劳状态的特征，综合考虑其他因素会提高疲劳检测的准确性和可靠性。此外，个体差异和环境条件（如光照）也可能对眼部特征检测的结果产生一定的影响。

基于嘴部的疲劳特征检测。嘴部特征检测是用于判断驾驶员疲劳状态的另一种常用方法。通过检测分析驾驶员嘴唇的张合程度，可以判断驾驶员的疲劳状态，如果驾驶员的嘴唇保持松弛或张嘴的状态时间较长，可能表明存在疲劳。嘴部特征检测还可以通过检测和跟踪嘴部的运动，获取关于驾驶员疲劳状态的

信息。例如，当驾驶员的嘴部运动缓慢或范围较小时，可能表明注意力不集中或疲劳。同时，分析驾驶员嘴部周围的肌肉活动和表情变化，如嘴角的上扬或下垂等，也可以反映驾驶员的疲劳程度。这些嘴部特征可以通过计算机视觉技术和图像处理算法进行提取和分析。通常，人工智能和机器学习方法可以应用于嘴部特征检测，以训练分类器来自动判断驾驶员的疲劳状态。

综上所述，疲劳驾驶检测方法有多种，每种方法都有其优点和缺点。如脑电图是一种直接测量脑电活动的方法，可以提供关于驾驶员疲劳状态的客观指标。通过分析脑电图信号的频谱特征和时域特征，可以准确判断疲劳程度。但脑电图检测需要使用专门的设备，如脑电图电极，对驾驶员来说比较麻烦和不便。此外，脑电图信号的分析和解读需要专业知识和经验。基于人脸特征的疲劳检测通常采用 Adaboost 分类器算法进行人脸定位[22]，可以通过分析眼部和嘴部的状态以及行为来判断疲劳状态。眼嘴部特征在疲劳驾驶中具有较高的相关性，因为疲劳会导致眼睛闭合程度、眼球运动、瞳孔大小、嘴唇的张合程度和嘴部表情变化等方面的变化。但人脸特征在不同人之间存在较大的个体差异，包括面部结构、肌肉运动模式等。这些个体差异可能导致疲劳检测的准确性受到影响。人脸特征的检测还可能受到环境条件的限制，如光照、摄像头质量等。不良的环境条件可能导致人脸图像的质量下降，影响疲劳检测的准确性。因此综合考虑多种方法，并结合其他因素，可以提高疲劳驾驶检测的准确性和可靠性。

1.3 论文的主要研究内容

本文主要以研究疲劳驾驶检测开展相关工作，并对该选题的背景和意义进行了叙述。同时，对国内外对疲劳驾驶的研究现状进行了调查和分析。最后在对目前存在的三种疲劳检测方法进行了比较后，选择了基于人脸特征的疲劳检测方法。这种方法可以通过分析驾驶员的面部特征来进行疲劳检测，无需使用额外的传感器或设备，对驾驶员来说非常便利和舒适。基于人脸特征的疲劳检测方法还可以根据个体差异进行自适应，通过建立模型对每个驾驶员的面部特征进行训练和学习，可以针对个体的特点进行疲劳检测，提高检测的准确性和适应性。主要研究内容有以下几个方面：

（1）基于深度学习的轻量级疲劳驾驶检测研究

为弥补小模型学习能力不足的问题，将模型中的传统卷积（卷积核大小不为 1×1）更换为深度重参数化卷积层，通过增加模型可学习的参数，增强模型的表达能力，捕捉输入数据中更复杂的特征，帮助模型更好地理解数据，从而提高模型在训练集上的拟合度。同时其学习的函数也更加复杂，有助于模型在未见过的数据上（测试集）表现更好，提高模型的泛化能力。接着为了增强模型深层网络层的特征提取能力，增加了高效多尺度注意力机制，在将一个 1×1 分支分解为两个一维特征编码的同时，将另外一个 3×3 分支与其并行放置。高效多尺度注意力机制不仅通过 1×1 分支沿一个空间方向捕获远程依赖关系，沿另一空间方向保留精确的位置信息。还可以从 3×3 分支中学习到不同尺度的特征信息，通过不同尺度对数据进行分析和学习，模型可以从多种视角理解数据，从而捕捉到更丰富的特征。同时多尺度学习能使模型对物体的尺度变化更为鲁棒，因为模型已经学会了在不同尺度下提取有效信息，能够更好地识别和定位物体。

（2）基于深度学习的多特征信息融合疲劳驾驶检测研究

算法在特征提取层中添加了多尺度特征提取 MSS 注意力模块，即使目标在图像中可能以不同的尺度出现，也可以帮助模型对不同尺度的目标进行感知和定位。同时帮助模型适应这种尺度变化，捕捉目标在不同尺度下的视觉特征。多尺度特征提取还可以提高目标定位的准确性，较小的尺度能够更好地捕捉目标的局部细节，而较大的尺度能够提供目标的全局上下文信息，通过综合利用不同尺度的特征，模型可以更准确地定位目标并减少定位误差。最后 MSS 通过融合不同尺度的特征来提高目标检测的性能，提供更全面和丰富的信息，有助于提高模型的准确性、鲁棒性和泛化能力，使其在各种场景中更加有效和可靠。接着引入了基于 SE 注意力改进的通道注意力 DS-Conv，替换原有模型的下采样卷积层。通过对每个通道进行自适应的特征加权，使得模型能够更有针对性地保留重要的特征，将更多的注意力集中在对当前任务有用的通道上，从而增强模型的表达能力和性能。通道注意力还可以通过对通道进行加权乘法运算，减少对冗余信息的关注，这有助于降低模型的计算和存储成本，提高模型的计算效率。并且通过引入通道注意力，模型可以更好地适应不同的输入情况和场景

变化，自适应地调整通道权重，从而提高模型的鲁棒性，使其在各种输入条件下都能有更好的性能表现。

（3）基于深度学习的驾驶员疲劳检测系统设计与应用

为解决疲劳驾驶带来的安全问题，并将本文研究算法应用于实际当中，设计了一套驾驶员疲劳检测系统。通过实时监测和分析驾驶员的 EAR 和 MAR 值，系统可以识别出眼睛闭合或打哈欠等行为模式，从而提示或警告驾驶员其处在疲劳状态。首先利用安装在方向盘管柱上的摄像头输入驾驶员视频数据，为了减少由于脸部遮挡现象给检测系统带来的影响，采用了红外摄像头作为采集驾驶员视频的设备。同时为了希望可以更直接、清晰地捕捉到驾驶员的眼睛和面部表情，也便于专注于检测瞳孔大小和眨眼频率等细节，红外摄像头利用支架外壳安装在方向盘管柱上。系统检测输入的驾驶员视频数据，根据输入的 Frame，描绘出人脸关键点检测效果图，驾驶员疲劳检测系统可以实时检测并计算得出眼嘴部张开大小和眼嘴部纵横比，判定眼嘴部和驾驶员疲劳状态。

1.4 论文的结构安排

本文通过围绕疲劳驾驶检测技术展开相关研究，总共分为六个章节。

第一章绪论。本章的主要内容为叙述了疲劳驾驶检测的研究背景及意义，分析了国内外疲劳驾驶检测的研究现状。

第二章基于深度学习的疲劳驾驶检测相关技术分析。本章作为本文研究的理论基础，简述了神经网络和深度学习，并说明了卷积神经网络的发展历程以及卷积神经网络的概念和组成，同时还介绍了本文使用的网络模型算法 YOLOv5 以及 YOLOv7，并对其网络结构以及创新点进行说明。

第三章基于深度学习的轻量级疲劳驾驶检测研究。本章主要讲述了基于 YOLOv5s 算法改进的疲劳驾驶检测，为了保证检测模型的实时性，选取 YOLOv5s 作为检测算法模型。接着说明了对 YOLOv5s 的改进方法。通过将深度重参数化卷积层替换 YOLOv5s 的传统卷积层，在深层网络层中增加了高效多尺度注意力机制，通过消融实验和与其他模型的对比试验，证明了模型在基本不增加参数的情况下完成了精度的提高。最后对本章做了总结。

第四章基于深度学习的多特征信息融合疲劳驾驶检测研究。本章主要针对

在复杂环境下，对人脸检测的检测精度。本章算法基于 YOLOv7 进行改进，通过将 DS-Conv 替换传统的下采样卷积，并在特征提取层中添加了多尺度特征提取 MSS 注意力，显著提高了在干扰情况较多检测环境下的检测精度。

第五章基于深度学习的驾驶员疲劳检测系统设计与应用。本章阐述了判定疲劳状态的判定方法，并展示了车载驾驶员疲劳检测系统的应用。

第六章总结与展望，简述本文所做的主要工作，并指出当前算法存在的不足以及今后的工作。

第 2 章 基于深度学习的疲劳驾驶检测相关技术分析

2.1 神经网络与深度学习简介

神经网络是一种计算模型，受到人类神经系统结构和功能的启发而设计。它由大量相互连接的简单处理单元（神经元）组成，通过这些连接进行信息传递和处理。在神经网络中，神经元接收输入信号，对其进行加权求和，并通过激活函数进行非线性变换，产生输出信号。这些输出信号又被传递到下一层神经元，形成层级结构。通常，神经网络由输入层、隐藏层和输出层组成，其中隐藏层可以有多层。每个神经元的权重（连接的强度）可以通过训练过程进行自动调整，以使网络能够学习并适应输入数据的模式和关系。神经网络的训练通常采用反向传播算法。该算法基于损失函数（衡量网络输出与期望输出之间的差异）和梯度下降优化方法，通过迭代地调整神经元的权重，以最小化损失函数。这样，网络可以学习从输入到输出之间的复杂映射关系，并具备对新数据进行预测或分类的能力。神经网络在机器学习和人工智能领域有广泛应用。它们可以用于图像识别、语音识别、自然语言处理、推荐系统、预测分析等任务。随着深度学习的兴起，深度神经网络（拥有多个隐藏层）如卷积神经网络和循环神经网络等被广泛应用，并在许多领域取得了显著的成果。

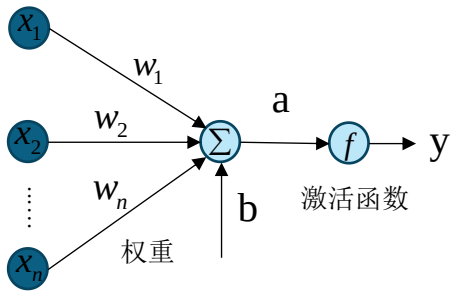


图 2-1 神经元结构图

神经元结构图如图 2-1 所示，用数学方式展示如公式（2-1）所示：

$$h_{w,b}(x) = f(W^T x + b) = f(\sum_{i=1}^{3} W_i x_i + b) \qquad (2\text{-}1)$$

其中 $x_i(i = 1, 2, \cdots, n)$ 作为输入参数，W 为输入信号对应的权重，b 为一个偏置，$f()$为非线性函数。

深度学习旨在模仿人类大脑的工作原理。它是人工智能领域的一个分支，通过构建和训练深度神经网络来实现模式识别、分类、预测和决策等任务。深度学习的核心思想是通过多层次的神经网络模型来学习数据中的抽象表示。这些神经网络由大量的人工神经元组成，每个神经元都接收来自前一层神经元的输入，并将其传递给下一层神经元。通过逐层传递和处理信息，深度学习模型可以学习到更加复杂和抽象的特征表示。深度学习的一个重要特点是端到端学习。传统的机器学习方法通常需要手工提取和选择特征，而深度学习则可以直接从原始数据中学习到高层次的特征表示，无需手动设计特征提取器。深度学习在多个领域取得了显著的突破，如计算机视觉、自然语言处理、语音识别等。例如，在计算机视觉领域，深度学习模型可以自动识别和分类图像中的对象和场景，甚至可以生成逼真的图像。在自然语言处理领域，深度学习模型可以理解和生成自然语言，实现机器翻译、文本摘要等任务。深度学习的成功得益于大数据和强大的计算资源的可用性。训练一个复杂的深度学习模型通常需要大量的标记数据和高性能的图形处理器（GPU）等硬件加速器。总之，深度学习是一种强大的机器学习技术，通过构建和训练深层神经网络来实现模式识别和决策任务。它在多个领域展现了巨大的潜力，并对人工智能的发展产生了深远影响。

2.2 卷积神经网络

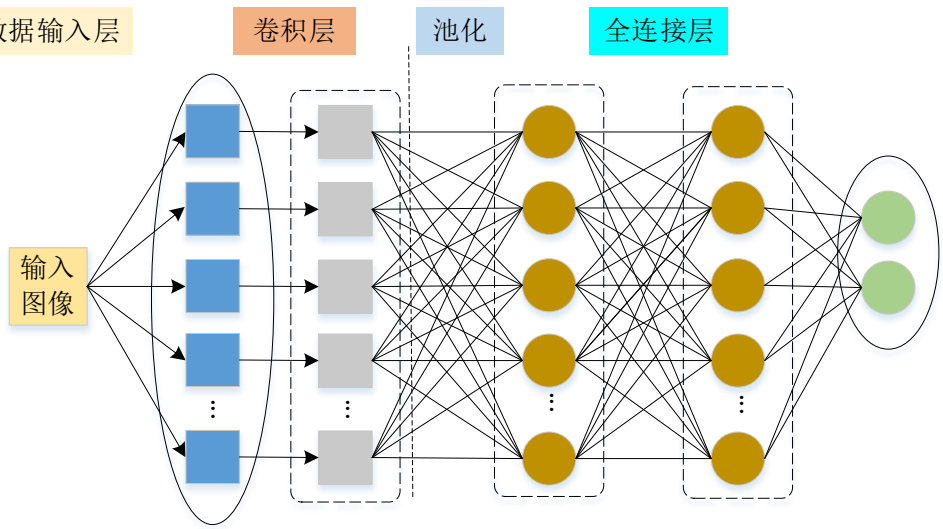


图 2-2 卷积神经网络结构图

卷积神经网络[23]（Convolutional Neural Network，CNN）的原理建立在卷积运算和神经网络的基础上，如图2-2所示。它通过多个卷积层、激活函数、池化层和全连接层构成，以实现对图像和其他网格结构数据的高效处理和特征提取。卷积操作是 CNN 的核心。它通过将一个可学习的过滤器（卷积核）应用于输入数据的局部区域，计算得到特征映射。卷积运算在整个输入数据上滑动，并在每个位置上执行卷积操作。这样可以有效地捕捉输入数据的局部特征，例如图像中的边缘和纹理。CNN 中的卷积核在整个输入数据上共享权值。这意味着在卷积层中，无论特征在图像中的位置如何，相同的卷积核被应用于所有位置。这种权值共享的机制大大减少了需要学习的参数数量，使得 CNN 能够处理大规模数据并保持模型的轻量化。池化层用于减小特征映射的空间尺寸并提取主要特征。最常见的池化操作是最大池化，它在每个池化窗口中选择最大值作为池化后的值。池化操作帮助减少数据的维度，并增强模型对平移不变性的学习能力。CNN 通常由多个卷积层和全连接层组成，形成层次结构。低层的卷积层可以学习到边缘和纹理等低级特征，而高层的卷积层可以学习到更抽象和复杂的特征。最后的全连接层通过学习权重来进行分类或回归。

随着时间的推移，CNN 在计算机视觉领域取得了巨大的发展和突破。在 1998 年，LeNet-5 是由 Yann 等人提出的第一个卷积神经网络模型，用于手写数字的识别，它被认为是现代 CNN 的鼻祖。AlexNet 是由 Alex 等人于 2012 年在 ImageNet 图像识别竞赛中获得冠军的 CNN 模型。AlexNet 的成功引发了对深度学习和 CNN 的广泛关注。在 2014 年，VGGNet 是由 Karen 和 Andrew 提出的 CNN 模型。它以其深度和简单的结构而著名，证明了增加网络深度可以提高性能。同年 GoogLeNet 是由 Google 团队提出的 CNN 模型，其主要特点是引入了 Inception 模块，有效地减少了参数数量。ResNet 是由 Kaiming 等人于 2015 年提出的 CNN 模型，通过使用残差连接解决了深层网络训练中的梯度消失问题，使得更深的网络可以被训练。

迁移学习是指将在一个任务上训练好的 CNN 模型应用于另一个相关任务上。迁移学习使得在小数据集上训练的模型也能取得良好的性能，并加速了模型的训练过程。GAN 是一种由生成器和判别器组成的神经网络结构，其中生成器试图生成逼真的图像，而判别器则尝试区分真实图像和生成图像。CNN在GAN中

广泛应用，如 DCGAN（Deep Convolutional GAN）和 PGGAN（Progressive Growing of GANs）。

总的来说，CNN 的发展使得计算机视觉领域取得了显著的进展。它在图像分类、目标检测、人脸识别、图像生成等任务上取得了令人瞩目的成果，并成为深度学习领域的基础模型之一。随着硬件的进步和研究的不断深入，CNN 仍然在不断演进和应用于更广泛的领域。

2.2.1 卷积层

卷积层是卷积神经网络的核心组件之一，用于提取输入数据（如图像）中的特征。卷积层的原理基于卷积运算和权值共享的概念。

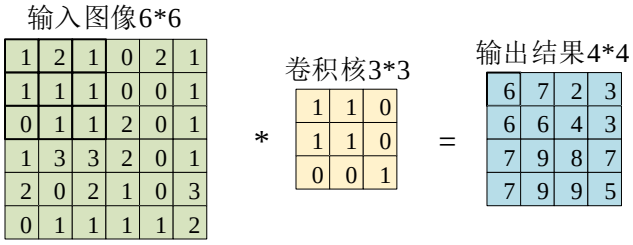


图 2-3 卷积计算过程

卷积运算是卷积层的核心操作，如图2-3所示。在卷积运算中，卷积层通过将一个可学习的过滤器（也称为卷积核或滤波器）应用于输入数据的局部区域来提取特征。卷积核是一个小的二维矩阵，它在输入数据上滑动，并与输入数据的每个局部区域进行逐元素相乘，并将结果相加得到输出特征映射的对应位置。

卷积运算的步骤如下：第一步定义卷积核的大小和数量。每个卷积核都可以提取不同的特征。之后将每个卷积核与输入数据的局部区域进行逐元素相乘。将逐元素相乘的结果进行求和，得到输出特征映射的对应位置的值。最后通过滑动卷积核，重复上述步骤，直到覆盖整个输入数据。

卷积层的主要特点是权值共享。在卷积层中，每个卷积核的权重在整个输入数据上共享。这意味着，无论特征在输入数据中的位置如何，相同的卷积核将被应用于所有位置。这种权值共享的机制大大减少了需要学习的参数数量，使得卷积层具有更好的泛化能力和更高的效率。卷积层的输出称为特征映射（Feature Map），它是对输入数据的特征提取。每个卷积核在输入数据上滑动，计算得到一个特征映射。多个卷积核可以同时应用于输入数据，生成多个特征

映射。这些特征映射捕捉了输入数据中的不同特征，如边缘、纹理和形状等。卷积层通常还包括激活函数（Activation Function），它引入非线性，增加网络的表达能力。常用的激活函数包括 ReLU（Rectified Linear Unit）、Sigmoid 和 Tanh 等。激活函数在卷积层的输出上逐元素地应用，将负值设为零，保留正值。

卷积核（Convolutional Kernel）是卷积层的组成部分之一。它是一个小的二维矩阵，用于在卷积层中进行卷积运算。卷积核定义了卷积层的特征提取方式和学习的参数。每个卷积核可以捕捉输入数据的不同特征，如边缘、纹理或形状等。在卷积层中，多个卷积核可以并行地应用于输入数据，从而生成多个特征映射。在应用卷积核之前，可以在输入数据的边界周围添加额外的像素，称为填充。填充可以有助于保持输出特征映射的尺寸与输入数据的尺寸相同，减少信息损失，并提高模型对边缘和边界的感知能力。步幅定义了卷积核在输入数据上滑动的距离。通常，默认情况下，步幅为 1，表示卷积核逐个元素地滑动，但可以通过调整步幅的大小来控制输出特征映射的尺寸。较大的步幅可以减小输出特征映射的尺寸，同时可能损失一些空间信息。输入数据可以具有多个通道，例如彩色图像具有 RGB 三个通道。卷积层中的卷积核也具有与输入数据相同数量的通道。卷积核在每个通道上进行卷积运算，并将结果相加以生成单个特征映射。

卷积层可以通过堆叠多个卷积核来增加输出特征映射的深度。深度决定了卷积层的特征提取能力和表示能力。每个卷积核都可以学习不同的特征，因此更深的卷积层可以学习到更丰富和抽象的特征。在卷积层中，卷积核的权重在整个输入数据上共享。这意味着，无论特征在输入数据中的位置如何，相同的卷积核将被应用于所有位置。参数共享大大减少了需要学习的参数数量，减少了过拟合的风险，并使卷积层具有更好的泛化能力。

通过多个卷积层的堆叠和其他组件（如池化层和全连接层），CNN 能够从原始输入数据中提取出层次化的特征表示，进而用于分类、目标检测、图像生成等任务。总结起来，卷积层是 CNN 中用于特征提取的核心组件，通过卷积运算和权值共享来捕捉输入数据的局部特征。它的原理和设计灵感来自于生物学中视觉皮层的工作原理，并在计算机视觉领域取得了巨大的成功。

2.2.2 激活函数

激活函数（Activation Function）是神经网络中的一种非线性函数，它被用于在神经元中引入非线性变换，从而增加神经网络的表达能力和拟合复杂函数的能力，使得神经网络可以对非线性模式和复杂关系进行建模。如果没有激活函数，神经网络只能表示线性变换，无法进行非线性建模。激活函数通常被应用于神经网络的隐藏层和输出层。

以下是几种常见的激活函数及其简介[24]：

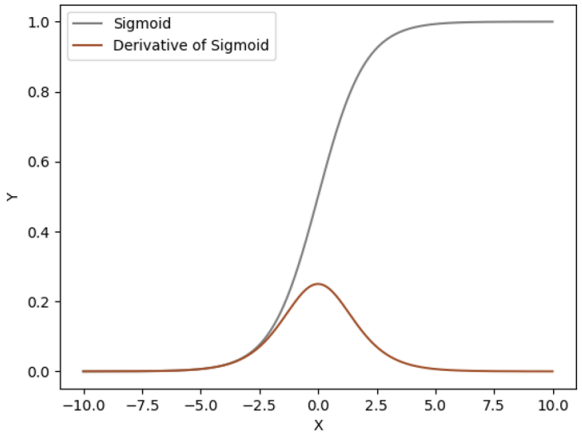


图 2-4 Sigmoid 函数及其导数图像

Sigmoid 函数是一种常用的激活函数，Sigmoid 函数及其导数图像如图 2-4 所示，它将输入映射到一个介于 0 和 1 之间的连续输出。Sigmoid 函数的形状呈现 S 型曲线，因此得名。Sigmoid 函数的计算公式如公式（2-2）所示：

$$S(x)=\frac{1}{1+e^{-x}} \tag{2-2}$$

Sigmoid 函数引入了非线性变换，使得神经网络能够对非线性模式和数据分布进行建模。Sigmoid 函数的输出范围介于 0 和 1 之间，可以被解释为概率或激活的程度，因此在二分类任务中，可以将输出阈值设置为 0.5，将大于阈值的预测视为正类，小于阈值的预测视为负类，同时还可以用作输出层的激活函数，并用于计算交叉熵损失函数。Sigmoid 函数在整个输入范围内都是可微分的，这对于基于梯度的优化算法（如反向传播）是很重要的。Sigmoid 函数常用于逻辑回归模型中，用于将线性模型的输出转换成概率。

然而，Sigmoid 函数也存在一些限制和问题。在 Sigmoid 函数的两个极端（接近 0 和 1）附近，梯度接近于零，这可能导致梯度消失问题，在深层网络中限制了梯度的传播和模型的训练能力。因此在某些情况下，其他激活函数（如

ReLU、Leaky ReLU、ELU 等）可能更常用，特别是在深度神经网络中。然而，在某些特定的任务和模型中，Sigmoid 函数仍然具有一定的应用价值。

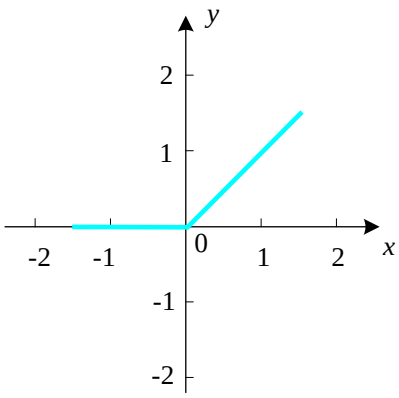


图 2-5 ReLU 激活函数

ReLU（Rectified Linear Unit）函数是一种简单而常用的激活函数，如图 2-5 所示，它在深度学习中广泛使用。ReLU 函数在输入大于零时直接输出输入值，而在输入小于等于零时输出零。ReLU 函数的计算公式如公式（2-3）所示：

$$\mathrm{ReLU}(x)=\max(0,x) \tag{2-3}$$

ReLU 函数引入了非线性变换，使得神经网络能够拟合更复杂的非线性模式和数据分布，在输入小于等于零的情况下输出零，这意味着它可以使得网络中的一部分神经元变得不活跃，从而实现稀疏性表示，减少参数的冗余性。ReLU 函数在输入大于零时具有恒定的梯度，这有助于解决梯度消失的问题，并促进梯度在深层网络中的有效传播。ReLU 函数的计算非常简单，只需比较输入值和零，适合在大规模神经网络中使用。

但 ReLU 函数也存在一些限制和问题，当输入小于等于零时，ReLU 函数的输出为零，这可能导致神经元的“死亡”，即对应的权重无法更新，无法学习新的特征。ReLU 函数在输入大于零时输出恒定的梯度，这可能导致某些神经元在训练过程中变得过于饱和，使得梯度变得非常小，减缓学习速度。ReLU 函数在输入小于零时输出为零，而在输入大于零时输出与输入相等，这种非对称性可能导致一些问题，如梯度爆炸。为了克服 ReLU 函数的一些限制，还出现了一些改进的变体，如 Leaky ReLU、ELU 和 PReLU 等，它们在负输入值范围内引入了一定的斜率或曲线，以解决神经元死亡和饱和问题。这些改进的 ReLU 变体在实践中也得到了广泛应用。

Softmax 函数是一种常用的激活函数，通常用于多类别分类任务中的输出层。它将一组实数映射到一个概率分布上，使得所有输出值的总和等于 1。Softmax 函数可以将神经网络的输出解释为各个类别的概率。Softmax 函数的计算公式如公式（2-4）所示：

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{2-4}$$

其中 e 是自然对数的底数，$z=[z_1, z_2, ..., z_j]$为实数向量。

Softmax 函数对输入向量中的每个元素进行指数运算，然后将所有指数值相加，并将每个元素除以总和，以确保结果是介于 0 和 1 之间，并且所有结果的总和等于 1。这样，Softmax 函数将输入向量转换为一个概率分布，其中每个元素表示对应类别的概率。需要注意的是，Softmax 函数在处理具有较大差异的输入值时可能存在数值稳定性问题，因为指数函数的值可能非常大。为了处理这个问题，可以使用数值稳定的实现方法，例如通过减去输入向量中的最大值来避免指数溢出。Softmax 函数广泛应用于多类别分类任务中的神经网络输出层，例如图像分类、语音识别和自然语言处理等。

除了上述激活函数，还有许多其他的激活函数变体和改进方法，每个都有其自身的特点和适用性。选择适当的激活函数需要根据具体问题和网络架构进行实验和评估。此外，激活函数的选择也可能取决于梯度消失和梯度爆炸等训练中的问题，以及对计算效率和模型复杂性的要求。

2.2.3 池化层

池化层是深度学习中常用的一种层类型，用于对输入数据在空间维度上进行下采样（降维），以减少参数数量、提取主要特征和减轻模型对输入数据的位置敏感性。池化层通过对输入数据的局部区域进行汇聚操作，将该区域的信息聚合到一个单一输出值或特征图中，常用的池化操作包括最大池化和平均池化，如图 2-6 所示。

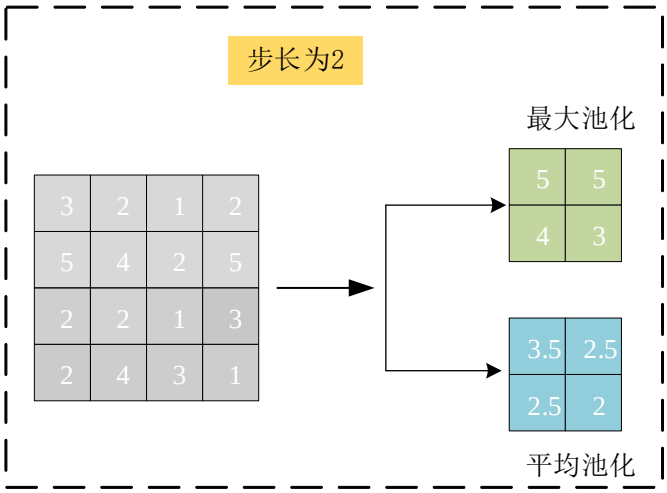


图 2-6 最大池化和平均池化

最大池化是一种常用的池化操作，用于在输入数据的局部区域中选择最大的元素作为输出值。最大池化通常应用于卷积神经网络中，作为特征提取的一部分。计算过程如下：首先定义池化窗口大小和步幅（通常是正方形），将输入数据划分为不重叠的窗口，窗口的大小与步幅相匹配；然后在每个窗口中，选择窗口内的最大值作为输出值；最后将所有窗口的最大值组成输出特征图。

最大池化通过选择局部区域内的最大值，提取出输入数据的主要特征。它可以帮助减少冗余信息，保留图像或特征图中最显著的特征。最大池化还具有平移不变性，即无论特征出现在图像的哪个位置，其在池化操作后得到的特征值是相同的。这种性质使得模型对于目标在图像中的具体位置变化具有鲁棒性。最大池化对于输入数据的尺度变化相对不敏感，因为它只关注局部区域内的最大值，这使得模型能够在不同尺度的目标上进行特征提取。同时最大池化可以减少输入数据的空间维度，从而降低计算负担和模型的参数数量，避免过拟合问题。需要注意的是，最大池化没有需要学习的参数，它是一种无参数的操作。

平均池化作为另外一种常用的池化操作，用于在输入数据的局部区域中计算所有元素的平均值作为输出值。平均池化通常用于卷积神经网络中，作为特征提取的一部分。平均池化的计算过程如下：首先定义池化窗口大小和步幅（通常是正方形），将输入数据划分为不重叠的窗口，窗口的大小与步幅相匹配；然后在每个窗口中，计算窗口内所有元素的平均值作为输出值；最后将所有窗口的平均值组成输出特征图。

平均池化通过计算局部区域内的平均值，提取出输入数据的主要特征。与最大池化不同，平均池化能够平滑图像或特征图中的细节，对整体特征进行描

述。不过与最大池化类似的，平均池化也具有平移不变性，即无论特征出现在图像的哪个位置，其在池化操作后得到的特征值是相同的。这种性质使得模型对于目标在图像中的具体位置变化具有鲁棒性。平均池化对于输入数据的尺度变化相对不敏感，因为它考虑了局部区域内所有元素的平均值。这使得模型能够在不同尺度的目标上进行特征提取。平均池化还可以减少输入数据的空间维度，从而降低计算负担和模型的参数数量，避免过拟合问题。与最大池化类似，平均池化也是一种无参数的操作，没有需要学习的参数。在卷积神经网络的架构中，平均池化通常在卷积层之后或者多个卷积层之间使用，以逐步减小特征图的空间维度，并提取出更具有代表性的特征。

池化层的使用通常和卷积层一起构成卷积神经网络，其中卷积层用于提取特征，而池化层则用于减少特征的空间维度。这样的设计可以有效地降低计算量、减少模型的参数量，并在图像和其他二维数据的处理中取得良好的效果。

2.2.4 全连接层

全连接层，也称为密集连接层，是深度学习中常用的一种层类型。全连接层的每个神经元都与前一层的所有神经元相连，通过权重将输入特征与输出特征进行线性组合，并通过激活函数进行非线性变换。

全连接层的计算过程如下：首先将输入特征展平为一个向量，然后每个神经元计算输入特征向量与权重向量的点积，并加上偏置项；最后对每个神经元的计算结果应用激活函数，产生输出特征向量。

全连接层通过权重的线性组合和非线性变换，能够将输入特征进行复杂的组合，从而得到更高级别的特征表示，提高模型的表达能力。全连接层中的权重和偏置项是需要通过训练来学习的参数。模型可以根据训练数据调整这些参数，使得网络能够学习到适合任务的特征表示。全连接层通常在线性组合后应用激活函数，引入非线性变换，使得模型能够学习到更复杂的非线性关系。这对于解决许多实际问题非常重要。全连接层通常作为神经网络的最后一层，用于对输入进行分类或回归预测。输出层的神经元数量与任务的类别数或预测目标的维度相匹配。需要注意的是，全连接层的参数数量随着输入和输出维度的

增加而增加。这可能导致模型过于复杂，容易过拟合。因此，在实践中，通常会结合其他层类型，如卷积层和池化层，来构建更有效的深度学习模型。

2.3 网络模型简介

目前卷积神经网络目标检测算法分为单阶段检测和两阶段检测。两阶段检测算法其基本思想是先生成一些候选框，再对这些候选框进行分类和回归，以得到最终的检测结果。常见的两阶段目标检测算法有 MaskR-CNN[25]、R-CNN[26]、FastR-CNN[27]、FasterR-CNN[28]等。两阶段目标检测算法虽然准确率较高，但计算量较大，检测速度较慢，不能满足驾驶员疲劳驾驶检测实时性的要求。单阶段检测算法则可以直接预测边界框和类别，而不需要通过两个阶段（即候选区域生成和分类）来完成。由于没有额外的候选区域生成和筛选过程，单阶段检测算法能够在减少计算资源需求的同时，实现较快的检测速度，常见的单阶段目标检测算法有 YOLO 系列[29-31]、VGG[32]、MobileNet[33]、GoogleNet[34]、SSD[35]、EfficientDet[36]、RetinaNet[37]等。但相较于两阶段检测算法，单阶段检测算法在准确率上可能略有不足。

针对单阶段检测算法在疲劳驾驶检测任务中准确率不足的问题，仲鹏宇[38]用 MobileNet 替换 SSD 的主干特征提取网络 VGG16，并在模型的浅层特征上加入轻量级注意力机制网络 SENet 提升模型对小目标的检测精度。李京徽[39]将 MobileNet 作为 YOLOv3-Tiny 的主干网络，对输出的特征图的后续处理没有完全照搬 YOLOv3-Tiny 中的过程，而是保留了一个包含 3 个卷积层的小卷积集，另外也适当增加了卷积核的数量，以提取更多更高层次的抽象特征来增加结果预测的准确率。臧露奇[40]在 YOLOv7 中进行了一些改进，他增加了感受野增强模块，通过捕获多尺度信息和不同距离的依赖关系，以提升模型的检测性能。此外他还引入了排斥损失和注意力模块，以增强对于复杂场景下的目标检测能力。然而上述方法在提升准确率的同时，检测速度明显下降。许卓凡[41]提出了一种不同的思路来解决准确率问题，他引入了浅层特征和深层特征的融合机制，通过使用两种类型的金字塔结构，将上下文信息和背景信息引入金字塔结构中。这种融合机制避免了特征融合过程中的不适应现象，并提高了对于小尺度和遮挡人脸的检测能力。另外 Jia[42]采用 GhostNet 为 backbone 来提取特征，并通过

特征融合进一步增强模型的性能，同时设计了一个新的损失函数，以平衡模型的检测速度和准确率，然而最终表现的准确率并不优异。

YOLO 是一种实时目标检测算法，由 Joseph Redmon 等人提出。它的特点是速度快，可以在一张图像上一次性检测出多个目标，并且准确率也比较高。YOLO 系列包括 YOLO、YOLOv5、YOLOv7 等不同版本，每个版本都有不同的改进和优化，在目标检测领域都取得了不错的成绩。

2.3.1 YOLOv5 模型简介

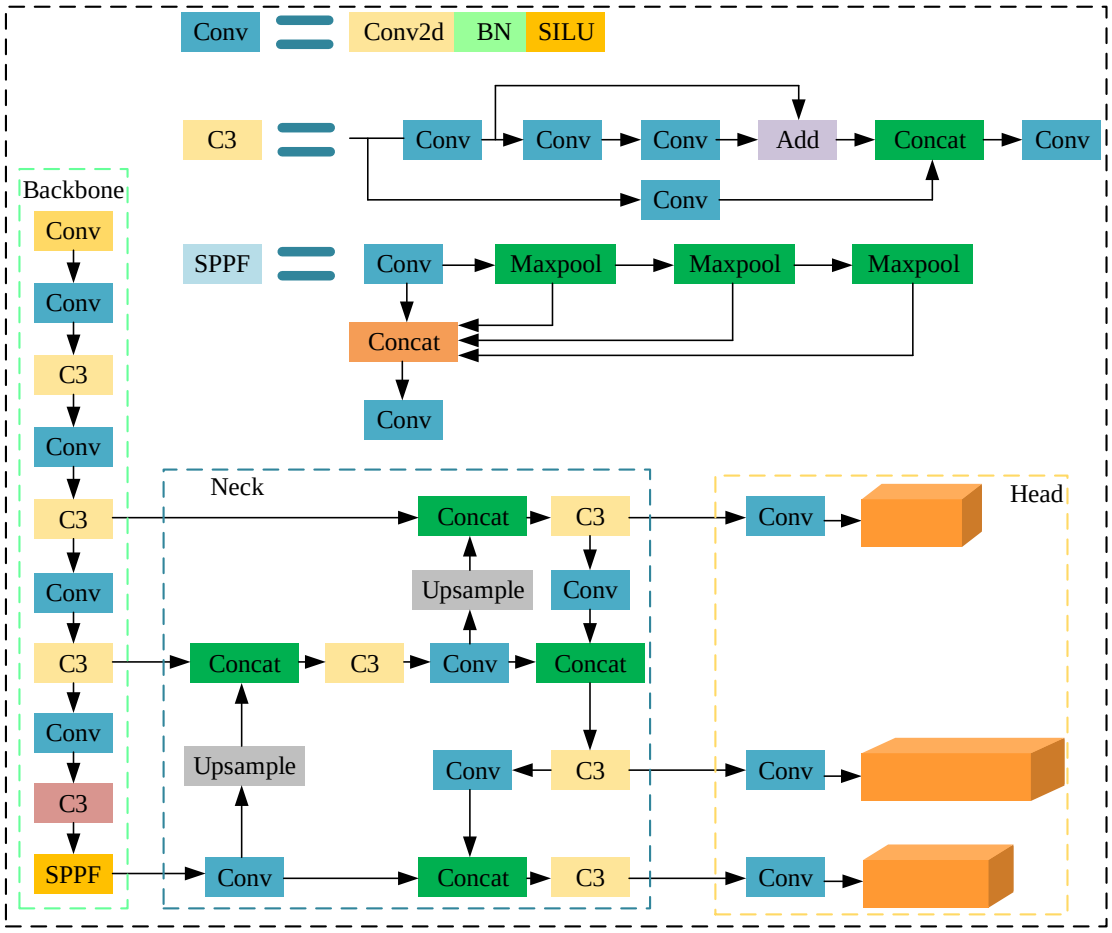


图 2-7 YOLOv5s 网络结构图

YOLOv5 是一个流行的深度学习目标检测算法，属于 YOLO 系列算法的版本之一，YOLOv5s 的网络结构图如图 2-7 所示。YOLO 算法是一种端到端的目标检测方法，能够在单个网络前向传播过程中同时预测出图像中的目标位置和类别。这种方法的主要优势在于其速度快、效率高，适用于实时目标检测任务。YOLOv5 提供了多个模型版本，包括 YOLOv5s、YOLOv5m、YOLOv5l 和

YOLOv5x，不同的版本在模型大小、速度和检测精度上有所不同，用户可以根据实际需求和资源限制选择合适的版本。

YOLOv5 采用一系列网络轻量化技术，比如模块化设计、路径聚合网络增强特征提取能力等，使得模型即便在资源受限的设备上也能运行。YOLOv5 在使用和部署上较为简单，提供了丰富的文档和教程，支持多种编程环境和平台，包括 Python、Docker 等，并且可以方便地导出到 ONNX、TensorRT、CoreML 等格式，便于在不同的应用场景中部署。YOLOv5 还引入了多种数据增强技术、优化后的损失函数和更高效的训练策略，进一步提高模型的泛化能力和准确率。作为一个开源项目，YOLOv5 拥有活跃的社区支持，不断有来自社区的贡献促使模型和工具得到更新和优化，凭借其在多个场景下的出色表现，成为了目标检测领域内的一个重要工具。

2.3.2 YOLOv7 简介

Yolov7 是目标检测算法 YOLO 系列的一种改进版本，它以快速且准确的目标检测能力而受到广泛关注。

Yolov7 主干网络采用了 Darknet，不仅具有较少参数，还拥有着高效的计算性能，同时采用了更深的网络层次结构，使得模型能够学习更丰富的特征。Yolov7 通过引入特征金字塔，可以在不同层次的特征图上进行目标检测，可以更好地处理不同尺度的目标，并使用了 PANet 来进行特征融合。PANet 通过自上而下和自下而上的路径，将来自不同尺度特征图的信息进行聚合，以提高目标检测的准确性。Yolov7 在训练和推理过程中采用了一些优化策略，如使用 Mish 激活函数、使用 Group Normalization（组归一化）等，以提高模型的效率和准确性。Mish 激活函数是一种非线性激活函数，可以提供更广泛的激活范围和更好的梯度性质，进而改善模型的准确性。组归一化用来规范化特征图。相比于传统的批量归一化，组归一化对小批量数据更稳定，并且在训练过程中不依赖批量大小，有助于提高模型的泛化性能。Yolov7 在训练过程中还应用了多种数据增强技术，如随机缩放、随机裁剪、图像翻转等，以增加数据样本的多样性，提高模型的鲁棒性。

2.4 深度学习在疲劳驾驶检测的应用

现阶段卷积神经网络已广泛使用在图像分类、目标检测、人脸识别以及自动驾驶等任务中[43-44]，并且在各任务中都卓有成效。深度学习在疲劳驾驶检测方面具有广泛的应用[45-46]，通过深度学习算法，可以从输入的图像或视频数据中学习和提取特征，以准确地检测和识别驾驶员的疲劳状态。以下是深度学习在疲劳驾驶检测中的常见应用：

驾驶员脸部关键点识别：深度学习模型可以训练用于检测人脸关键点的神经网络，从而准确地定位驾驶员的眼睛、嘴巴等部位。通过监测这些关键点的位置和变化，可以判断闭眼和打哈欠等疲劳迹象。

表情识别：深度学习模型可以训练用于表情识别的神经网络，通过分析驾驶员面部表情的变化，可以判断是否存在疲劳的迹象，如困倦、无表情等。

眼部状态检测：深度学习可以应用于眼部状态检测，例如通过训练神经网络来检测闭眼、眨眼或眼睛睁开程度的变化，从而判断驾驶员的疲劳程度。

多模态数据融合：深度学习可以用于融合多种传感器数据，如图像、声音和车辆传感器数据。通过将这些数据结合起来，可以更全面地评估驾驶员的疲劳状态。

实时疲劳检测：深度学习模型可以实时处理和分析视频流数据，以实时检测驾驶员的疲劳状态。这种实时检测能够及时发现驾驶员的疲劳迹象，并采取相应的预警或提醒措施。

通过深度学习的应用，驾驶员疲劳检测系统可以更准确地判断驾驶员的疲劳程度，提高道路安全性，并有效预防因疲劳驾驶导致的交通事故。

2.5 本章总结

本章首先简要介绍了神经网络和深度学习的基本概念，然后概述了卷积神经网络的发展历程和现状，并详细介绍了卷积神经网络的主要组成结构，包括卷积层、激活函数、池化层和全连接层的原理和概念。此外，本章还介绍了后续实验中使用的模型 YOLOv5s 和 YOLOv7。总体来说，本章在全文中属于理论基础部分，为后续的工作内容提供了基础知识。

第 3 章 基于深度学习的轻量级疲劳驾驶检测研究

3.1 引言

由于汽车车机的计算资源有限，当疲劳驾驶检测模型参数较大时，可能导致车机系统卡顿等情况的发生。同时，为了驾驶人员的生命财产安全，疲劳驾驶检测必须保证人脸检测的准确率。针对现有算法无法兼顾疲劳驾驶检测准确率和检测速度，提出了基于改进 YOLOv5s 的疲劳驾驶检测算法。YOLOv5s 是一种实时目标检测算法，具有参数少和准确率高的优点。对于人脸检测任务，它能够在实时性要求较高的场景中快速准确地检测人脸。同时 YOLOv5s 在目标检测任务中取得了卓越的准确性。它采用了一系列的改进策略，包括特征金字塔网络、PANet 和框回归等技术，能够提高对小目标的检测精度，从而提升了人脸检测的准确性。其次，YOLOv5s 采用了多尺度检测的策略，可以在不同尺度下检测人脸。这对于应对人脸在图像中大小变化的情况非常有效，使得算法具有更好的适应性和鲁棒性。最后，相比于一些复杂的目标检测算法，YOLOv5s 具有简单轻量的特点。它的网络结构相对简洁，参数量较小，方便在嵌入式设备或者边缘计算平台上部署和应用。

3.2 算法流程

通过在网络模型中添加了深度重参数化卷积层[47]（Depthwise Overparameterized Convolutional, DO-Conv）模块和高效多尺度注意力[48]（Efficient Multi-Scale Attention, EMA）模块提高网络模型的检测精度。首先将 DO-Conv 替换传统的卷积层，它通过额外的深度卷积来增强卷积层，输入与传统卷积和深度卷积组成的组合核进行卷积。这两个卷积的组合构成了一个过度参数化，增加了可学习的参数，此外，在推理阶段转换成传统卷积，将计算量减少到完全等同于传统卷积层的计算量，因此 DO-Conv 可以在不增加推理计算复杂度的情况下实现性能提升。同时，由于 YOLOv5s 的浅层特征层的特征表达性能已经足够强，所以将 EMA 模块添加在 Backbone 的深层网络层中。EMA 模块为了保留每个通

道的信息并减少计算，将通道维度分组为多个子特征，除了对全局信息进行编码以重新校准每个平行分支的通道权重外，两个平行分支的输出特征还通过跨维度的交互作用进一步聚合，使空间语义特征在每个特征组中得到良好的分布。

3.3 算法原理

3.3.1 深度重参数化卷积层 DO-Conv

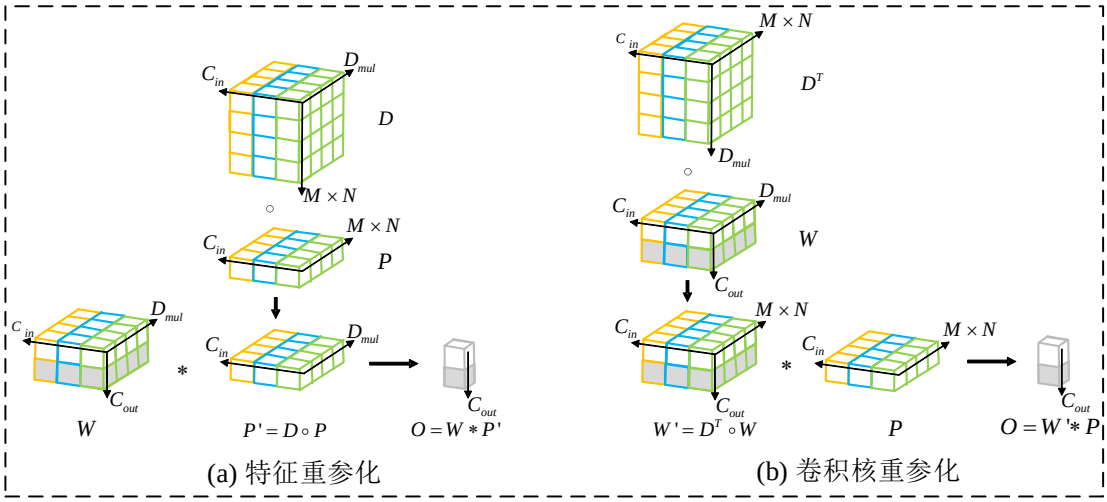


图 3-1 特征重参化和卷积核重参化

深度重参数化卷积层是由深度卷积 D 和传统卷积 W 两个可训练卷积组成，并且有两种重参化的方法，分别为特征重参化和卷积核重参化，如图 3-1 所示。特征重参化时，如图 3-1（a）所示，特征 P 首先应用深度卷积算子 $\circ$，通过深度卷积 D 得到一个转换后的特征 $P'=D\circ P$，然后应用常规卷积算子 $*$，通过传统卷积 W 转换 P' 得到输出 $O=W*P'$。卷积核重参化时，如图 3-1（b）所示，$\circ$ 和 $*$ 的组合是通过一个复合核 W' 实现的，首先通过 D^T 来转换传统卷积 W，D^T 是 D 在第一维度和第二维度上的转置，得到 $W'=D^T\circ W$，然后在 W' 和特征 P 之间应用传统的卷积算子 $*$，得到 $O=W*P'$。给定一个输入特征 $P\in R^{(M\times N)\times C_{in}}$，特征重参化和卷积核重参化可用公式（3-1）表示：

$$\begin{aligned} O &= (D,W)\circledast P \\ &= W*(D\circ P) \\ &= (D^T\circ W)*P \end{aligned} \tag{3-1}$$

其中 $D\in R^{(M\times N)\times D_{mul}\times C_{in}}$，$W\in R^{C_{out}\times D_{mul}\times C_{in}}$，$D^T\in R^{D_{mul}\times(M\times N)\times C_{in}}$，$D_{mul}\geq M\times N$，

$\circledast$ 为深度重参数化卷积算子，$O=(D,W)\circledast P$ 是 DO-Conv 的输出。

虽然特征重参化和卷积核重参化两种方法在数学上是等价的，但训练效率并不相同。当使用乘法和累加运算次数（MACC）衡量特征重参化和卷积核重参化计算量，假设 stride=1，特征重参化和卷积核重参化分别作用于输入 $R^{H\times W\times C_{in}}$ 上时，MACC 分别可以计算如公式（3-2）、（3-3）所示：

特征重参化：

$$\begin{aligned} P' &= D\circ P : D_{mul}\times(M\times N)\times C_{in}\times(H\times W), \\ O &= W * P' : C_{out}\times C_{in}\times H\times W\times D_{mul}. \end{aligned} \tag{3-2}$$

卷积核重参化：

$$\begin{aligned} W' &= D^{T}\circ W : D_{mul}\times(M\times N)\times C_{in}\times C_{out}, \\ O &= W' * P : C_{out}\times C_{in}\times H\times W\times(M\times N). \end{aligned} \tag{3-3}$$

其中 H 和 W 分别为特征映射的高度和宽度。在大多数情况下，由于 $H\times W\gg C_{out}$，$D_{mul}\geq(M\times N)$，卷积核重参化会比特征重参化需要更少的 MACC 操作。同时卷积核重参化中 W' 消耗的内存也小于特征重参化中 P' 消耗的内存，因此卷积核重参化在训练阶段更可取。

DO-Conv 是卷积层的一个过度参数化。传统卷积层可训练权重的参数化 A 可表示为公式（3-4）：

$$A: C_{out}\times(M\times N)\times C_{in} \tag{3-4}$$

DO-Conv 可训练权重的参数化 A 可表示为公式（3-5）：

$$A': (M\times N)\times D_{mul}\times C_{in}+C_{out}\times D_{mul}\times C_{in} \tag{3-5}$$

其中 $D_{mul}\geq(M\times N)$。即使在 $D_{mul}=(M\times N)$ 的情况下，参数的数量也增加了 $(M\times N)\times D_{mul}\times C_{in}$。

在用 DO-Conv 进行 YOLOv5s 的训练和推理时，DO-Conv 的感受野与传统卷积层的感受野相同，因此DO-Conv可以很容易地取代YOLOv5s中的卷积层。同时 YOLOv5s 使用的是 Adam 优化器的算法，它是一种自适应学习率的梯度下降算法。由于公式（3-1）中定义的$\circledast$ 算子是可微的，所以 DO-Conv 的 D 和 W' 都可以用基于梯度下降的优化器进行优化。DO-Conv 通过增加可学习参数可以

使模型具备更强的表达能力，能够更好地拟合更复杂的数据，同时学习到更细致和精确的特征信息，提高模型在检测任务上的性能。DO-Conv 还可以提高模型的泛化能力，通过增加可学习参数，模型可以更好地捕捉训练数据中的变化和复杂性，从而使模型在未见过的数据上表现更好。

3.3.2 高效多尺度注意力机制 EMA

全局池化方法通常用于通道注意编码空间信息的全局编码，但由于它将全局空间信息压缩到通道描述符中，导致难以保存位置信息。Coordinate Attention[49]（CA）注意力机制将通道注意力分解为两个一维特征编码，在沿一个空间方向捕获远程依赖关系的同时，还可以沿另一空间方向保留精确的位置信息。然后将生成的特征图分别编码，互补地应用于输入特征图，以增强特征学习。CA 按照公式公式（3-6）分解了全局池化，转化为一对一维特征编码操作：

$$Z_c = \frac{1}{H \times W} \sum_{j}^{H} \sum_{i}^{W} x_c(i, j) \tag{3-6}$$

其中 x_c 表示第 C 通道的输入特征。当给定输入 $X \in R^{C \times H \times W}$，CA 分别沿着水平坐标和垂直坐标对每个通道进行编码。因此，高度为 H 和宽度为 W 的第 C 通道的输出分别为公式（3-7）所示：

$$\begin{aligned} Z_c^H(H) &= \frac{1}{W} \sum_{0 \le i \le W} x_c(H, i) \\ Z_c^W(W) &= \frac{1}{H} \sum_{0 \le j \le H} x_c(j, W) \end{aligned} \tag{3-7}$$

其中 x_c 表示第 C 通道的输入特征。上述两种变换在两个空间方向上分别聚合特征，生成了两个一维特征编码向量，然后将两个一维特征编码向量通过卷积层连接起来。这种变换方式允许注意力模块捕捉到沿着一个空间方向的长期依赖关系，并保存沿着另一个空间方向的精确位置信息，这有助于网络更准确地定位感兴趣的目标。不过两个一维全局平均池化的设计虽然是为了沿两个空间维度方向对全局信息进行编码，并沿不同维度方向在空间上捕捉长距离的相互作用，但同时也忽略了完全空间位置之间相互作用的重要性。此外 1×1 卷积

核的有限接受域不利于对局部跨通道交互的建模和对背景信息的利用。

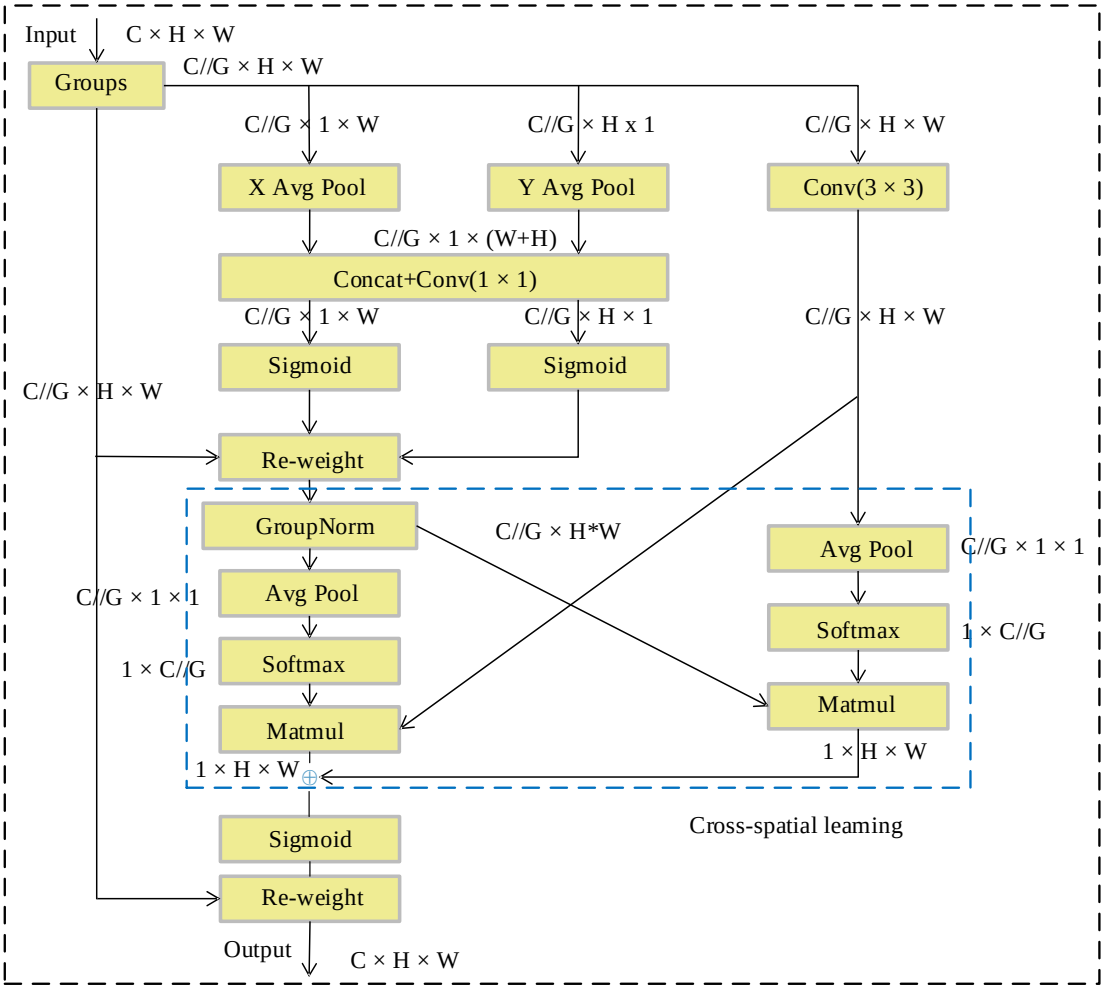


图 3-2 EMA 模块结构图

EMA模块结构图如图3-2所示。EMA选取CA模块中1×1卷积的共享分量，在 EMA 中命名为 1×1 分支，同时为了获得多尺度空间结构信息，将一个 3×3 卷积与该 1×1 分支并行放置，命名为 3×3 分支。EMA 利用三条并行路线来分组提取描述特征图的注意力权重，其中 1×1 分支中有两条并行路线，3×3 分支作为第三条路线。为了捕捉通道之间的依赖关系并降低计算成本，EMA 在通道方向上建立了跨通道信息交互模型。在 1×1 分支中，有两个一维全局平均池化运算操作，分别沿两个空间方对通道进行编码，而在 3×3 分支中，一个 3×3 卷积用于捕捉多尺度特征信息。

EMA 将两个一维特征编码并联起来，使其在 1×1 分支中共享 1×1 卷积且不进行降维，同时 3×3 分支通过 3×3 卷积捕捉到不同尺度上的空间特征，获得更大范围的上下文信息。特征分组和多尺度结构的形式，可以有效地建立短距离和长距离的依赖关系，保留精确的空间结构信息，获得更好的性能，还可以帮

助模型对不同尺度的目标进行感知和定位。最后通过融合不同尺度的特征来提高目标检测的性能，提供更全面和丰富的信息，提高了模型的准确性、鲁棒性和泛化能力，使其可以完成在各种场景下的检测任务。不过值得注意的是，由于 DO-Conv 并不适用于尺寸为 1×1 的卷积核，因此在实验中 EMA 还是调用传统卷积

3.4 实验结果与分析

3.4.1 实验数据集及环境配置

实验采用的数据集是 WIDER FACE[50]数据集。WIDER FACE 数据集是一个大规模的人脸检测数据集，由伯克利大学的研究人员于 2016 年发布。该数据集旨在促进人脸检测方法的研究和发展，特别是针对在实际应用中常遇到的问题的各种挑战，例如复杂度、遮挡、姿势变化等。数据集共包含 32203 张训练图像和 393703 张个人脸部标注，覆盖大量的人脸大小、姿势、表情和遮挡情况。同时数据集基于 61 个事件类进行组织，对于每个事件类别，随机选择 40%/10%/50%数据作为训练、验证和测试集。这使得数据集非常适合和评估人脸检测算法。除了人脸边界框之外，WIDER FACE 还提供了五个面部关键点（左右眼、鼻子、左右嘴角）的标注，这有助于开展面部关键点检测和姿势估计等研究。数据集还包含了从网络收集的各种场景，如室外、室内、群体照片等。这使得 WIDER FACE 具有多样性，有助于开发出更加鲁棒性的人脸检测算法。WIDER FACE 数据集已成为人脸检测领域的一个重要基准，许多研究人员和工程师利用它来开发、和评估新的人脸检测方法。

WIDER FACE 数据集的训练集包含了 12876 张图片，验证集包含了 3226 张图片。实验均在 Linux 操作系统的环境下对数据集进行训练及验证，具体运行环境配置如表 3-1 所示。

表 3-1 实验运行环境

类型	具体参数
操作系统	Linux
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz
GPU	GeForce RTX 3090
PyTorch	1.90
CUDA	11.4
python	3.8

3.4.2 参数设置和评价指标

实验共训练 300 个 epoch，具体的网络模型参数设置如表 3-2 所示。

表 3-2 网络模型参数

参数	具体数值
Batchsize	64
epoch	300
device	0
conf_thres	0.001
iou_thres	0.65

实验通过准确率 P（Precision）、召回率 R（Recall）、mAP（mean Average Precision）值作为评价网络模型优劣的指标。精确率P表示分类器预测为正样本且实际为正样本的目标在所有被分类器预测为正样本的目标中所占的比例。精确率 P 计算公式如公式（3-8）所示：

$$P=\frac{T_P}{T_P+F_P} \tag{3-8}$$

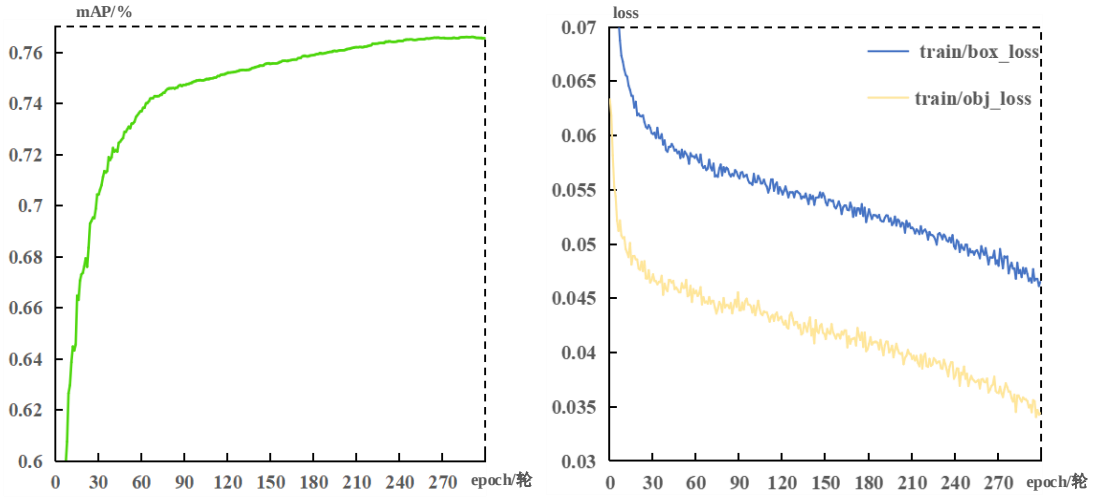
其中，T_P 表示实际为正样本且预测结果为正样本的目标，即正确分类的正样本；F_P 表示实际为负样本但预测结果为正样本的目标，即错误分类的负样本。

召回率 R 表示分类器预测为正样本且实际为正样本的目标在所有实际为正样本的目标中所占的比例。召回率 R 计算公式如公式（3-9）所示：

$$R=\frac{T_P}{T_P+F_N} \tag{3-9}$$

其中，T_P 与精确率计算中的含义相同；F_N 表示实际为正样本但预测结果为负样本的目标，即错误分类的正样本。以召回率 R 为横坐标，精确率 P 为纵坐标，绘制 P-R 曲线，平均精度 AP 的值可以表示为该曲线下的面积大小。

3.4.3 实验结果



(a) mAP变化图　　(b) 损失函数变化图

图 3-3 改进算法后的 mAP 与损失函数的变化图

图 3-3 为改进算法的 mAP 与损失函数的变化图。在完成 300 个 epoch 的训练后，从实验结果可以看到，在网络模型中加入了 DO-Conv 模块加快收敛速度后，算法模型在完成 250 个左右 epoch 时，可以完成收敛精度达到最高，并由于通过 EMA 模块来增强深层网络层的特征表达，网络模型 mAP 可以达到 76.8%。

表 3-3 不同算法的检测结果对比

模型	Backbone	参数	GFLOPs	mAP
YOLOv5s	CSPDarNet53	7.02M	15.8	74.6
YOLOv3-tiny	DarNet53-tiny	33.25M	5.56	55.1
YOLOv6s	EfficientRep	17.2M	44.2	62.9
YOLOv7-tiny	ELAN	6.01M	13.1	72.6
Ours	**CSPDarNet53**	**7.41M**	**16.8**	**76.8**

由表 3-3 可知，改进的 YOLOv5s 算法 mAP 达到了 76.8%，与同级别参数下的其它模型精度相比，改进后的算法比 YOLOv7-tiny 高出 4.2%，比 YOLOv6s 高出 13.9%，比 YOLOv3-tiny 高出 21.7%。相较于基线网络 YOLOv5s，检测精度提升了 2.2%，并且采用单阶段检测，适合在车载系统等资源有限的环境下部署。

3.4.4 消融实验

为了证明实验算法改进的有效性，通过在基线网络模型中添加深度重参数化卷积层 DO-Conv 模块和 EMA 模块来比较添加后的人脸检测的性能。在实验中，用 WIDER FACE 数据集来训练和验证，并在一致的配置环境下进行了实验。

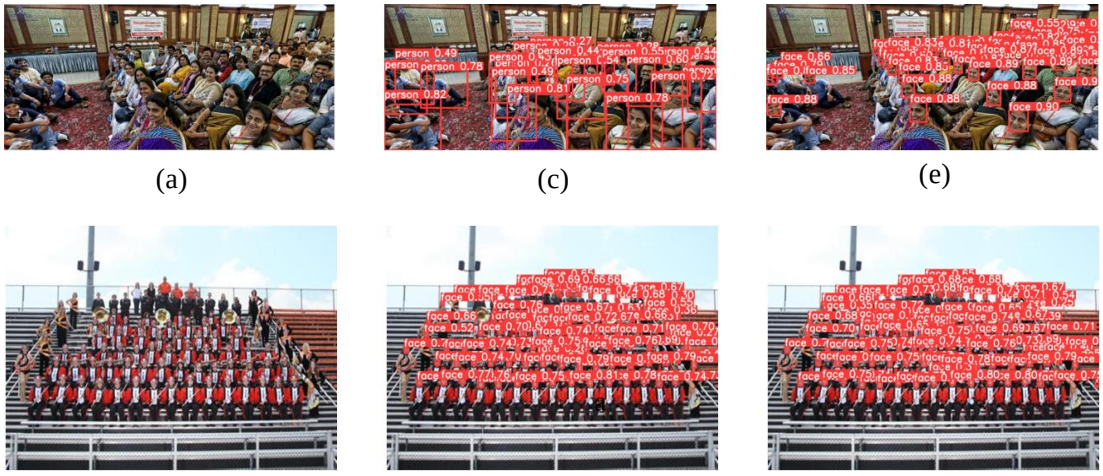
实验结果如表 3-4 所示。

表 3-4 消融实验检测结果对比

模型	参数	mAP_0.5/%	mAP_0.5:0.95/%
YOLOv5s	7.02M	74.6	40.9
YOLOv5s+DO-Conv	7.40M	75.2	41.2
YOLOv5s+DO-Conv+EMA	7.41M	76.8	42.4

首先，在添加了 DO-Conv 模块到网络模型后，通过增加可学习的参数，提高模型特征表达能力，增强了模型的检测性能，mAP_0.5/%检测精度提升 0.6%，达到 75.2%。其次，添加 EMA 模块来提高网络的多尺度感知能力和注意力机制。EMA 模块通过并行子网络的结构，有助于有效减少深度同时保持了高性能，改进算法在参数几乎不变的情况下精确度得到 1.6%的提升。最后将 DO-Conv 模块和 EMA 模块同时添加到网络层中，检测精度达到了 76.8%，较原基线网络提升了 2.2%。因此在基线网络中添加 DO-Conv 模块和 EMA 模块可以显著提高人脸检测的性能，证明了改进算法的有效性。

3.4.5 可视化结果



(a) (c) (e)

(b) (d) (f)

图 3-4 图像识别结果对比图

图 3-4 展示了使用改进算法与基线算法得到的检测结果对比图。第一列图 3-4（a）和图 3-4（b）为待检测原图，第二列图片 3-4（c）和 3-4（d）为原基线算法的检测结果，第三列图片 3-4（e）和 3-4（f）为改进算法的检测结果。在图 3-4（c）和图 3-4（e）中可以明显看到，在增加 EMA 模块后，对小目标和遮挡目标的检测效果有了极大的提升，改进算法的漏检率大幅度降低。在图 3-4

（d）和图 3-4（f）中可以看到，在增加了 DO-Conv 模块和 EMA 模块后，识别准确度和置信度都有着明显提升。

3.5 本章小结

提出基于改进的 YOLOv5s 疲劳驾驶检测算法，首先将 DO-Conv 模块替换传统的卷积层，采用深度卷积和传统卷积组合的复合核的方式，在不增加网络模型推理计算复杂度的情况下加快了模型的收敛，提升了准确率。其次，在深层网络层中添加了 EMA 模块。EMA 通过两个一维特征编码，在捕获远程依赖关系的同时，还保留着精确的位置信息。同时通过并行增加的 3×3 卷积分支多尺度提取特征，能够从不同尺度上捕捉目标的细节和上下文信息，增强对小目标和遮挡场景下目标的检测能力。最后在 WIDER FACE 数据集上，实验算法的检测准确率较基线提升了 2.2%。综上所述，提出的算法参数量较少、准确率较高，适用于计算资源受限的汽车车机系统，因此具有实际应用的潜力。

第 4 章 基于深度学习的多特征信息融合疲劳驾驶检测研究

4.1 引言

汽车在行驶过程以及驾驶员在疲劳状态时，都有可能进行汽车的转向操作，因此会发生转动方向盘导致对人脸的部分遮挡，造成人脸目标检测准确度降低。同时因个体差异，如果驾驶员眼睛过小可能会出现眼部关键点定位不准确的情况。　　提出了一种基于改进 YOLOv7 的疲劳驾驶检测算法。Yolov7 具有高效的检测性能，采用更有效的特征提取和聚合机制，从而确保即使在复杂环境中也能准确提取目标的关键信息。虽然 Yolov7 主要用于目标检测，但它可以提取人脸图像中的丰富特征。这些特征可以用于进一步的人脸识别任务，例如提取人脸特征向量并进行比对。Yolov7 的网络结构可以根据需要进行修改和扩展，以适应特定的人脸识别任务，同时可以在 Yolov7 的基础上添加额外的网络层或模块，以提高人脸识别的准确性。

4.2 算法流程

首先，通过将传统卷积层替换为深度重参数化卷积层，增加可学习的参数，使模型能够学习到更复杂和精细的特征信息。其次，在特征提取层中添加了多尺度特征提取 MSS[51]注意力模块，该模块能够在不同尺度上捕捉目标的细节和上下文信息，从而提高对小目标的检测能力。最后，引入了基于 SE[52]注意力改进的 DS-Conv 模块，用于替换原有模型的下采样卷积层。DS-Conv 模块能够帮助模型关注与目标相关的通道信息，减少对于无关信息的依赖，提高目标检测的准确性，它还可以学习到对于遮挡目标更具区分度的通道信息，从而提高对于遮挡目标的检测能力。通过以上改进，提出的基于改进 YOLOv7 的疲劳驾驶检测算法在提高检测准确性的同时，也保持了较高的检测速度，从而有效解决现有算法在疲劳驾驶检测任务中的问题。

4.3 算法原理

4.3.1 下采样卷积层 DS-Conv

下采样操作通常会减小特征图的尺寸，从而丢失一些重要的特征信息，通过添加通道注意力，可以对每个通道进行自适应的特征加权，使得模型能够更有针对性地保留重要的特征，将更多的注意力集中在对当前任务有用的通道上，从而增强模型的表达能力和性能。通道注意力还可以通过对通道进行加权乘法运算，减少对冗余信息的关注，这有助于降低模型的计算和存储成本，提高模型的计算效率。并且通过引入通道注意力，模型可以更好地适应不同的输入情况和场景变化，自适应地调整通道权重，从而提高模型的鲁棒性，使其在各种输入条件下都能有更好的性能表现。

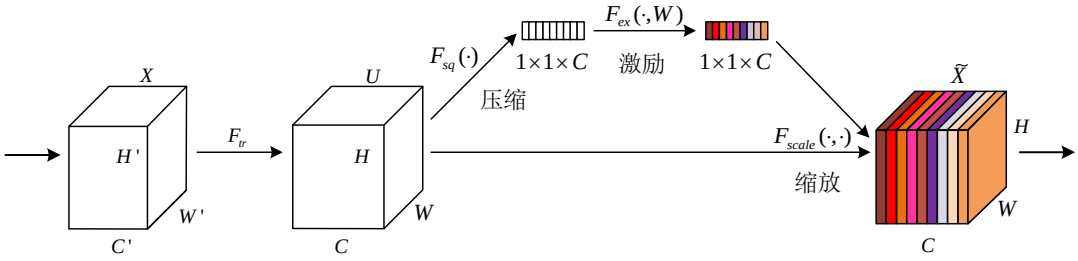


图 4-1 SE 注意力模块结构图

SE（Squeeze-and-Excitation）注意力模块结构图如图 4-1 所示。它通过学习通道间的关系，动态地调整每个通道的权重，以便更好地捕捉重要的特征，增强模型的表达能力。SE注意力机制由三个模块组成，分别为压缩（Squeeze）模块、激励（Excitation）模块和缩放（Scale）模块。压缩模块将卷积神经网络的输出特征图通过全局平均池化操作进行压缩，将通道维度压缩为一个特征向量。这个特征向量捕获了整个特征图的全局统计信息，反映了每个通道的重要性。激励模块通过使用两个全连接层（或卷积层）对压缩后的特征向量进行处理，以产生一个激励向量。第一个全连接层用于降低维度，第二个全连接层用于增强特征。最后使用激活函数将激励向量的值限制在 0 到 1 之间。缩放模块将激励向量乘以输入特征图，对每个通道进行缩放。这个缩放操作根据通道的重要性调整特征图中每个通道的权重，从而使更重要的特征得到增强，不重要的特征得到抑制。

SE 注意力机制的作用是通过自适应地学习通道权重，增强模型对重要特征

的关注，抑制对于任务无关或不重要的特征。它可以帮助模型更好地适应不同样本和场景，提高模型的表达能力和准确性。同时 SE 注意力机制可以嵌入到现有的卷积神经网络中，而不需要额外的复杂结构。

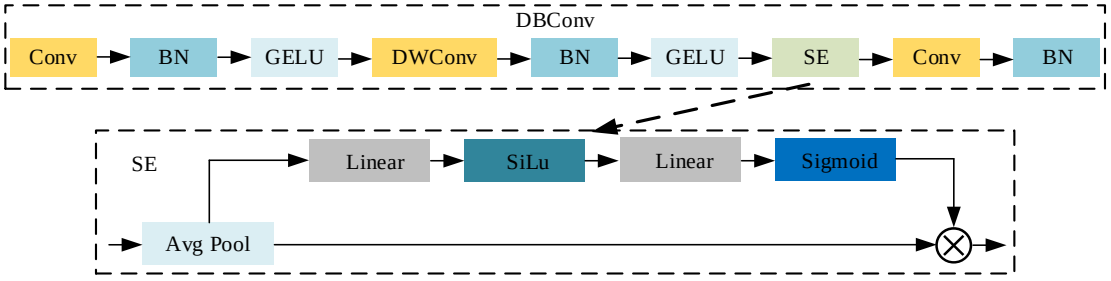


图 4-2 DS-Conv 模块结构图

DS-Conv 模块结构图如图 4-2 所示。在除了添加 SE 通道注意力以外，DS-Conv 还增加了几个卷积层。首先通过添加第一个普通卷积层将输入的通道数增加至原来的四倍，增加网络的表达能力和特征提取能力。然后添加一个深度可分离卷积层，减少参数量和降低计算量并获取各个通道的信息，目的是通过深度可分离卷积层来提取更加丰富和有表达力的特征。之后将特征输入到 SE 模块中，SE 模块对每个通道的特征进行降维，从而减少计算量，并在学习各个通道的权重后，融合各个通道的特征信息。最后增加一个普通卷积层恢复增加的通道数，为后续的操作提供适当的输入。最终通过整合和融合各个通道的特征信息，DS-Conv 可以提高网络的性能和表达能力，从而更好地处理输入数据并提取有用的特征。

4.3.2 MSS 注意力

MSS 注意力主要由多头混合卷积（Multi-Head Mixed Convolution，MHMC）、多尺度感知聚合（Scale-Aware Aggregation，SAA）和尺度感知调制器（Scale-Aware Modulation，SAM）三个模块组成，结构图如图 4-3 所示。

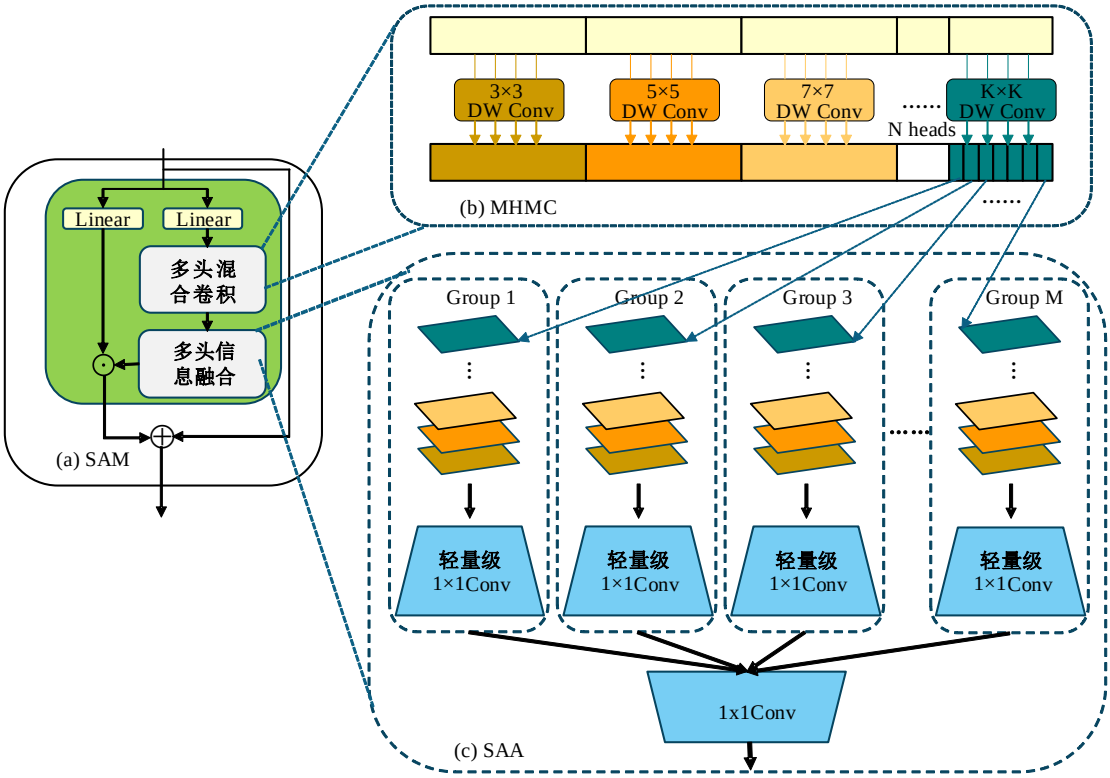


图 4-3 MSS 注意力结构图

MHMC 设置了 N 个卷积核，并将卷积核的大小初始化为 3×3，以 2 为步幅逐头递增，对每个检测头应用独立的深度可分离卷积操作。通过增加检测头的数量，可以使用更大的卷积核来扩大感受野，从而增强模型建模长距离依赖关系的能力。由于逐渐增加卷积核的大小，从而可以捕捉到不同尺度上的空间特征，较小的卷积核关注局部细节，较大的卷积核则能够捕捉更大范围的上下文信息。通过多个检测头的并行操作，模型能够同时处理多个尺度上的特征，从而提高目标检测的性能。因此 MHMC 具有通过调整头的数量来调节感受野的范围和多粒度信息的能力，MHMC 结构图如图 4-3（b）所示，过程可表述为公式（4-1）：

$$O_{MHMC(X)} = Concat(DW_{k_1\times k_1}(x_1),...,DW_{k_n\times k_n}(x_n)) \quad (4\text{-}1)$$

其中 $O_{MHMC(X)}$ 为 MHMC 的输出，$DW_{k_n\times k_n}(x_n)$ 为第 n 个深度可分离卷积检测头的输出，$x=[x_1,x_2,...,x_n]$ 表示分割成的 N 个检测头，$k_i\in\{3,5,...,K\}$ 表示卷积核大小单调增加 2。

由于 MHMC 检测头之间存在信息交互不足的问题，因此引入了 SAA 模块，设计如图 4-3（c）所示。SAA 模块从每个检测头中选择一个通道来共同构建一个组，这样每个组包含了来自不同检测头的特征通道。通过这种方式，SAA 模

块对 MHMC 生成的不同粒度的特征进行重新组合和分组，以增强多尺度特征的多样性。然后 SAA 模块使用一个 1×1 卷积操作来进行组内和组间的信息融合，这种跨组信息融合的方式可以实现轻量化和高效的聚合效果。通过将不同组内的特征进行融合，SAA 模块促进了多尺度特征之间的交互，从而提高了目标检测模型的性能。当给定输入 $X \in R^{H\times W\times C}$，组的数量 *Groups=C/Heads*，每个组中包含了 N 个特征通道，其中 C 为通道数，N 为检测头的数量，SAA 的过程可表述为公式（4-2）：

$$\begin{aligned} M &= W_{inter}([G_1, G_2, ..., G_M]), \\ G_i &= W_{intra}([H_1^i, H_2^i, ..., H_N^i]), \\ H_j^i &= DWConv_{k_j\times k_j}(x_j^i) \in R^{H\times W\times 1}. \end{aligned} \tag{4-2}$$

其中 W_{inter} 和 W_{intra} 是点卷积的权矩阵，$j \in \{1, 2, ..., N\}$，$i \in \{1, 2, ..., M\}$。$H_j \in R^{H\times W\times M}$ 表示深度可分离卷积的第 j 个头部，H_j^i 表示第 j 个头部的第 i 个通道。

如图 4-3（a）中，在使用 MHMC 捕捉多尺度空间特征并通过 SAA 进行信息融合后，得到一个输出特征图，然后通过 SAM 进行调制。对于输入 $X \in R^{H\times W\times C}$，计算输出 Z 如公式（4-3）所示：

$$\begin{aligned} Z &= M \odot V, \\ V &= W_v X, \\ M &= SAA(MHMC(W_S X)). \end{aligned} \tag{4-3}$$

其中⊙是元素相乘，W_v 和 W_s 是线性层的权矩阵。SAM 保留了信道维度，可以增强模型对不同通道特征的建模能力和灵活性。同时还可以在不同的输入下动态变化，实现自适应自调制，使模型能够独立调节每个通道的特征权重，实现特征选择和多样性的特征融合，从而提高目标检测模型的性能。

通过 MSS 多尺度特征的提取，即使目标在图像中可能以不同的尺度出现，也可以帮助模型对不同尺度的目标进行感知和定位。同时帮助模型适应这种尺度变化，捕捉目标在不同尺度下的视觉特征。多尺度特征提取还可以提高目标定位的准确性，较小的尺度能够更好地捕捉目标的局部细节，而较大的尺度能够提供目标的全局上下文信息，通过综合利用不同尺度的特征，模型可以更准确地定位目标并减少定位误差。最后 MSS 通过融合不同尺度的特征来提高目标检测的性能，提供更全面和丰富的信息，有助于提高模型的准确性、鲁棒性和

泛化能力，使其在各种场景中更加有效和可靠。

4.3.3 卷积核重参化

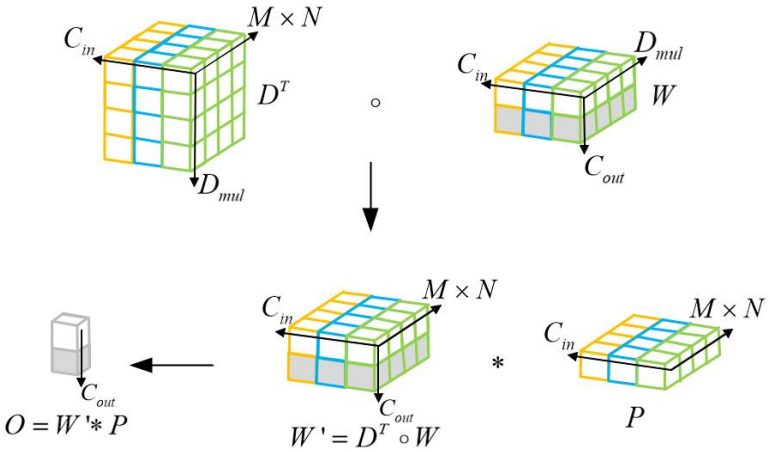


图 4-4 卷积核重参化

如图 4-4 为卷积核重参化。结合上文研究，在 YOLOv7 模型中并未将 DO-Conv 代替所有的传统卷积，而仅仅替换了 ELAN 结构中的部分卷积，修改后的 ELAN 结构如下图 4-5 所示。

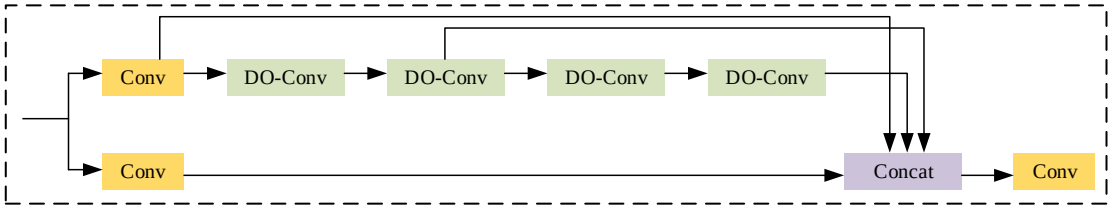


图 4-5 改进 YOLOv7ELAN 结构图

通过增加可学习参数，模型可以更好地拟合训练数据，提高模型的表达能力，从而更好地适应复杂的数据分布和任务要求。同时也增加了模型的灵活性，降低欠拟合的风险，提高了模型的拟合能力。在一些复杂的任务和大规模数据集上通常需要更强大的模型来提取更复杂的特征信息，通过增加可学习参数，模型可以更好地处理这些挑战，从而提高任务的准确性和性能。

4.4 实验结果与分析

4.4.1 实验数据集及环境配置

由于尺度、遮挡、姿势和背景杂乱等因素的巨大变化，WIDER FACE 数据集极具挑战性。为了量化这些特性，采用了通用对象建议方法[53-54]，这种方法

专门用于发现图像中的潜在对象（人脸可视为一个对象）。通过测量提议的数量与人脸检测率，可以初步评估数据集的难度。在评估中，采用 EdgeBox[55]作为对象建议，它在准确性和效率方面都有很好的表现[56]。最终 WIDER FACE 数据集根据检测难度分为了 Easy，Medium，Hard 三个子集，实验均在 Linux 操作系统的环境下对数据集进行训练及验证，具体运行环境配置如表 4-1 所示。

表 4-1 实验运行环境

类型	具体参数
操作系统	Linux
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz
GPU	GeForce RTX 3090
PyTorch	1.90
CUDA	11.4
python	3.8

实验训练 100 个 epoch，具体的网络模型参数设置如表 4-2 所示。

表 4-2 网络模型参数

参数	具体数值
Batchsize	16
epoch	100
device	0
workers	8

4.4.2 实验结果

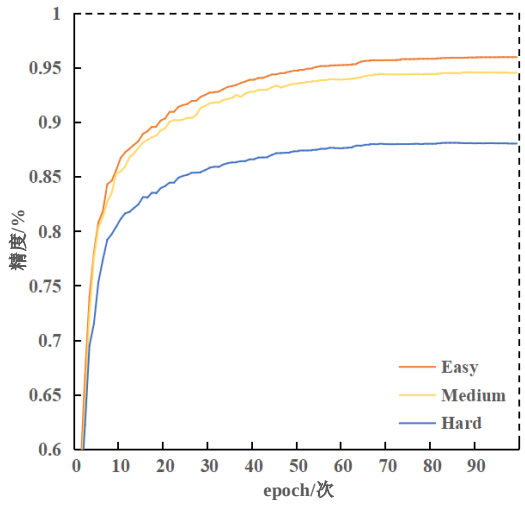


图 4-6 算法在 WIDER FACE 验证集 Easy，Medium，Hard 子集上的 mAP

模型以精度(mean average precision, mAP)作为评价指标，mAP 是所有类别 AP 的均值。在 WIDER FACE 验证集的三个子集上的 mAP 曲线如图 4-6 所示，提出的算法最终在 WIDER FACE 验证集的 Easy，Medium，Hard 子集上分别达

到 96.0%，94.6%，88.1%。为了进一步证明使用模块的有效性，将所提出的算法与其它算法进行对比，结果如表 4-3 所示。

表 4-3 不同算法在 WIDER FACE 验证集上的结果对比

模型	Easy	Medium	Hard
DSFD[57]	96.6	95.7	90.4
SFDet[58]	95.4	94.5	88.8
FANet[59]	95.2	94.0	90.0
FA-RPN[60]	94.9	94.1	89.4
文献[61]	94.9	93.0	83.6
SFD[62]	93.7	92.5	85.9
HR[63]	93.7	92.1	83.1
SSH[64]	93.1	92.1	84.5
CMS-RCNN[65]	89.9	87.4	62.4
Ours	**96.0**	**94.6**	**88.1**

通过表 4-3 可以看出，提出的算法在准确率上仍有提升的空间，不过参数量较少，仅有 49.5M，并且采用单阶段检测。相比之下，DSFD 算法的参数量为 459M，需要更多的计算资源。因此提出的算法在性能和计算资源之间取得了一定的平衡。此外与一些多阶段的人脸检测算法，如 CMS-RCNN 相比，提出的算法在准确率上取得了显著的提升。

4.4.3 消融实验

通过对 YOLOv7 模型进行了一系列的改进，包括将传统卷积层替换为 DO-Conv，把多尺度特征提取 MSS 注意力模块添加在特征提取层中，引入基于通道注意力 SE 改进的 DS-Conv 模块，用于替换原有模型的下采样卷积层。实验的结果如表 4-4 所示。

表 4-4 消融实验检测结果对比

模型	DO-Conv	DS-Conv	MSS	参数	Easy	Medium	Hard
基线算法	×	×	×	37196556	94.7	93.2	86.1
消融实验 1	✓	×	×	38518962	95.0	93.6	86.3
消融实验 2	✓	✓	×	41562290	95.3	94.0	87.0
消融实验 3	✓	✓	✓	49545394	96.0	94.6	88.1

根据表 4-4 中的数据可以看出，在增加多尺度特征提取 MSS 注意力模块后，模型可以更好地捕捉目标在不同大小和比例下的特征信息，Easy、Medium 和 Hard 子集上的精度均有较高提升，分别提升了 0.7%、0.6%和 1.1%。另外在添加 DS-Conv 模块后，Hard 子集提升较多，精度提高了 0.7%，这是因为在 Hard 子集中小脸所占的像素太少，导致在下采样过程中特征丢失严重，通过引入通道注意力，可以更好地关注细节信息，从而提升在 Hard 子集上的准确率。最后

引入 DO-Conv 增加了模型可学习的参数，Easy、Medium 和 Hard 子集上的准确率也有所提升，分别提高了 0.3%、0.4%和 0.2%。综上所述，通过增加 MSS 注意力模块、DS-Conv 模块和 DO-Conv 模块，可以改善模型在不同环境下的特征提取能力，提升检测准确性，为模型提供了改进和优化。

4.4.4 可视化结果

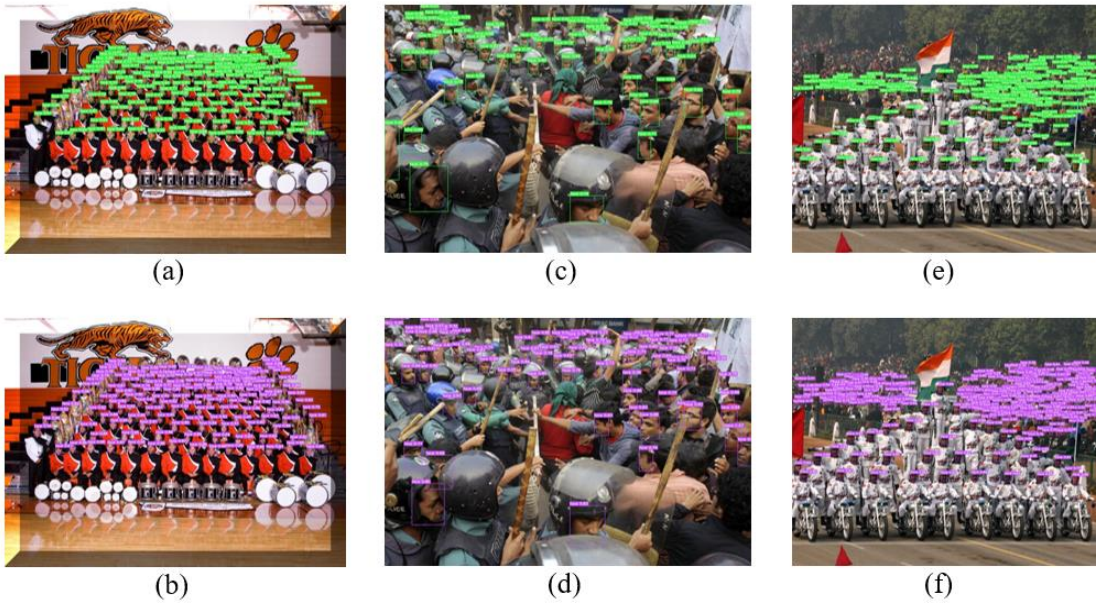


图 4-7 图像识别对比结果

图 4-7 比较了使用改进算法与基线算法进行人脸检测任务的结果。图 4-7（a）、图 4-7（c）、图 4-7（e）展示了基线算法的检测结果，图 4-7（b）、图 4-7（d）、图 4-7（f）展示了改进算法的检测结果。在图 4-7（a）、图 4-7（b）中可以看到经过训练的 YOLOv7 模型在人脸识别任务中取得了出色的表现，但与改进算法结果相比，基线算法出现了一例误检，证明了改进算法在检测任务中的稳定性。在图 4-7（c）、图 4-7（d）中，当人脸存在遮挡情况时，基线算法的漏检率和误检率较高。相比之下，改进算法基本完成了检测任务，极大降低了漏检率和误检率，这表明改进算法具备较高的准确性，能够应对遮挡情况下的人脸检测挑战。在图 4-7（e）、图 4-7（f）中，在人脸较为密集且包含大量小脸的场景下，基线算法对于图片中远处小脸的检测效果较差，然而改进算法检测出的小脸数量大幅提升，对人脸的定位也更加精准。基线算法和改进算法的漏检率和误检率如表 4-5 所示，在 Hard 子集中，改进算法漏检率下降了 2.7%，对复

杂环境下的目标检测的检出率大幅度提高。在误检率上，改进算法较基线算法在 Easy、Medium 和 Hard 上分别下降 0.6%、0.7%和 0.9%，表明改进算法在不同难度级别的场景中能够减少错误检测的情况。实验结果表明改进算法在人脸检测任务中相较于基线算法展现了更好的稳定性、准确性，以及在遮挡和小脸等复杂场景情况下也表现了更出色的检测能力。

表 4-5 漏检率和误检率检测结果对比

模型	数据集	漏检率%	误检率%
基线算法	Easy	12.5	7.6
	Medium	14.2	8.1
	Hard	21.1	10.2
改进算法	Easy	11.2	7
	Medium	12.7	7.4
	Hard	18.4	9.3

4.5 本章小结

提出了一种基于 YOLOv7 改进的疲劳驾驶检测算法，显著提高了复杂环境下人脸目标的准确率。首先，通过将深度重参数化卷积层 DO-Conv 替换传统的卷积层，在不增加推理计算复杂度的情况下加快拟合过程，增强模型的检测性能。其次，引入了基于通道注意力 SE 改进的 DS-Conv 模块，用于替换原有模型的下采样卷积层，极大改善了遮挡目标场景下下采样过程特征丢失严重的问题，提高了目标检测的准确性。再次，在特征提取层中添加多尺度特征提取 MSS 注意力模块，能够从不同尺度上捕捉目标的细节和上下文信息，增强了对小目标的检测能力。最后实验结果表明，提出的算法在 WIDER FACE 数据集的 Easy，Medium，Hard 子集上的检测精度分别达到了 96.0%，94.6%，88.1%。综上所述，提出的算法结构简单、准确率较高，适应因个体差异和转动方向盘导致人脸遮挡等问题对人脸目标检测的影响，这使得提出的算法成为一种值得考虑的疲劳驾驶检测模型解决方案。

第 5 章 基于深度学习的驾驶员疲劳检测系统设计与应用

5.1 引言

疲劳是一种常见的生理和心理状况，通常由过度劳累、压力、缺乏睡眠或其他健康问题引起。它既可以是身体的也可以是心理的表现，影响个人的能力来执行日常活动，减少工作效率并影响生活质量。疲劳的产生机制复杂，可以从物理、心理和行为上表现出来，包括以下几个方面。物理层面：长时间或高强度的体力活动导致肌肉力量下降；长期的劳累或没有得到足够的恢复，身体的能量储备耗尽，导致机体功能下降。心理层面：长期的精神压力、紧张和焦虑可以导致情绪耗竭，表现为兴趣丧失、情绪低落等。行为层面：身体和心理的疲惫会直接影响到个人的行为表现，如工作效率降低、易出错等，疲劳还可能导致睡眠质量变差，形成一个恶性循环。

疲劳对驾驶员的影响是深远和危险的，它能显著降低驾驶的安全性。疲劳会减慢驾驶员的反应速度，在紧急情况下，这种延迟可能意味着无法及时做出必要的操作来避免事故，同时导致驾驶员注意力分散，使得驾驶员更难持续关注道路状况和交通标志，增加出错或遇险的风险。疲劳还会影响到驾驶员的决策能力和判断力，使他们可能错误估计速度、距离和自己的能力，从而做出危险的驾驶决策，并且由于视觉模糊和对动态视觉信息处理能力降低，对于识别和理解路况变更，处理夜间或复杂光线条件的能力也会受到影响。疲劳驾驶最严重的后果之一是可能引发微睡或甚至在驾驶时睡着。微睡是指不自主地短暂入睡，这种状态下的驾驶员对周围环境几乎没有任何反应，极度危险，严重危害人们的生命财产安全。

为了验证本文提出的驾驶员疲劳检测系统，本章以家用汽车为实验对象，通过红外摄像头作为视频采集设备，对驾驶员疲劳检测系统进行了测试。测试结果证明了本文提出的驾驶员疲劳检测系统的有效性，并在 YAWDD 数据集上准确率达到了 94%，因此具有实际应用的潜力。

5.2 疲劳状态检测系统整体框架

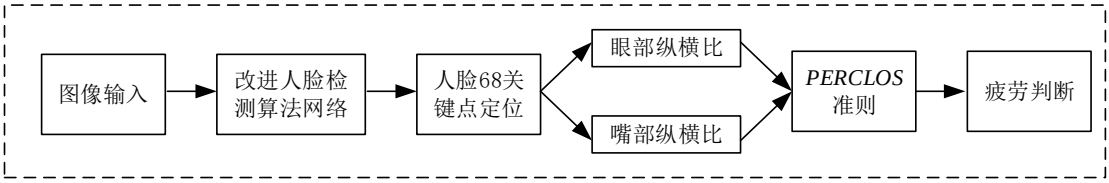


图 5-1 疲劳驾驶检测整体流程图

如图5-1所示为疲劳驾驶检测流程图，通过检测驾驶员眼睛是否处于闭合状态以及嘴巴是否打哈欠的方法，来判定驾驶员是否处于疲劳状态。在将采集到的图像输入后，经过改进算法模型检测人脸[66-67]，后续对人脸检测框采取人脸68 关键点定位处理。人脸关键点检测是指识别和定位人脸中重要的关键点，如眼睛、鼻子、嘴巴等部位的位置，是在人脸检测识别图像中是否存在人脸，并将其框出来后，进一步对检测框的人脸提取关键点的操作。人脸关键点检测器部分代码如下：

```
# 遍历检测到的人脸
for face in faces:
    # 使用 dlib 人脸关键点检测器检测人脸关键点
    landmarks = predictor(gray, face)
    # 遍历关键点，并在图像上绘制出来
    for n in range(0, 68):
        x = landmarks.part(n).x
        y = landmarks.part(n).y
        cv2.circle(image, (x, y), 1, (0, 255, 0), -1)
# 显示结果图像
cv2.imshow("Facial Landmarks", image)
cv2.waitKey(0)
cv2.destroyAllWindows()
```

如图 5-2 为判定驾驶员疲劳状态的流程图。根据检测到的人脸 68 关键点，计算眼部、嘴部纵横比，判断驾驶员眼部和嘴部的状态，并实时计算闭眼时间占比（PERCLOS）和打哈欠等疲劳参数。通过综合计算和分析这些参数的结果，评估驾驶员的疲劳状态。

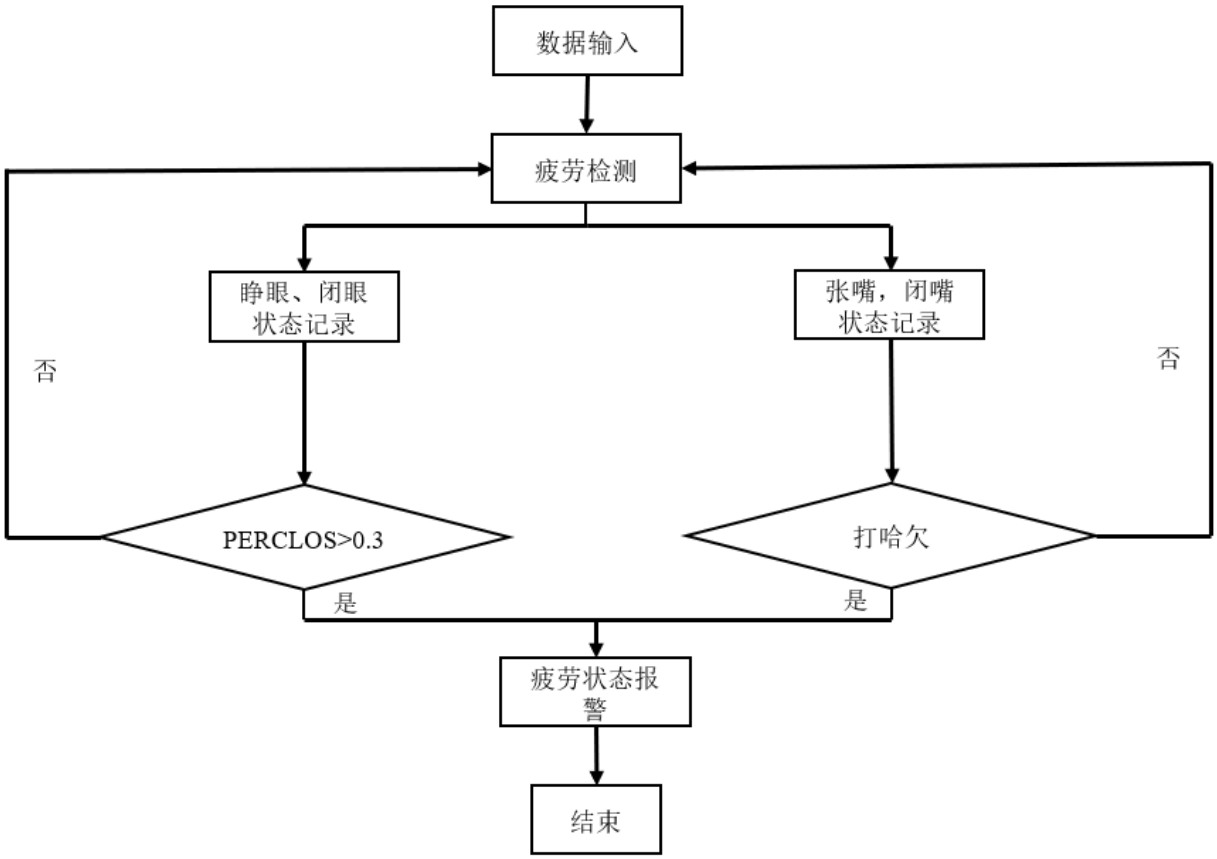
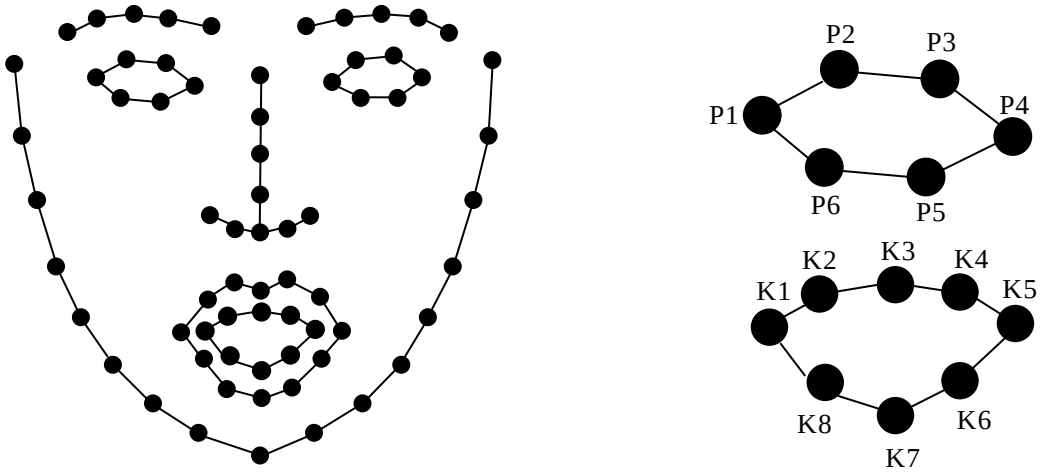


图 5-2 驾驶员疲劳检测流程图

5.3 驾驶员疲劳状态判定



(a) Dlib人脸关键点坐标　　(b) 右眼及嘴巴关键点坐标

图 5-3 Dlib 人脸关键点检测器的 68 个关键点坐标和右眼及嘴巴关键点坐标

如图 5-3 为 Dlib 人脸 68 关键点坐标图，人脸 68 关键点检测是一种常见的计算机视觉技术，用于识别和定位人脸上的特定部位，如眼睛、眉毛、鼻子、嘴巴、下巴等。这项技术广泛应用于人脸识别、情绪分析、虚拟现实、增强现

实等领域。人脸 68 关键点检测一般基于机器学习或深度学习模型完成。这些模型通过大量的标记数据进行训练，学习人脸的各种特征和模式。在推理阶段，给定一张人脸图片，模型能够预测出 68 个关键点的精确位置。常见的 68 个关键点模型为 Dlib 的人脸关键点检测模型。Dlib 是一个包含各种机器学习算法的 C++库，其中包括用于人脸关键点检测的预训练模型。Dlib 的人脸检测基于 HOG（直方图梯度）特征和线性分类器，其关键点检测部分利用回归树集成学习方法。人脸 68 关键点检测技术的进步为人脸识别、表情识别、人机交互等多种应用提供了强大的技术支撑，促进了相关领域的发展。

在疲劳检测的应用中，通过实时监测和分析驾驶员的 *EAR* 和 *MAR* 值，系统可以识别出频繁的眨眼、眼睛长时间保持闭合（比如微睡）或频繁打哈欠等行为模式，从而提示或警告驾驶员其正在经历疲劳状态，需要休息或采取其他措施以确保驾驶安全。这种技术在提高道路行车安全、减少因疲劳驾驶引起的事故方面具有重要意义。*EAR*（眼睛纵横比，Eye Aspect Ratio）和 *MAR*（嘴巴纵横比，Mouth Aspect Ratio）是两种常用于计算机视觉和人脸分析中的度量，特别用于疲劳检测。它们通过分析驾驶员的面部特征变化来判断是否存在疲劳状况。

5.3.1 基于眼部状态的疲劳判定

首先通过改进的人脸检测网络模型检测输入图像，从图像中检测到脸部区域后，将脸部区域作为输入图像输入，之后通过基于 Dlib 库的人脸关键点检测器获取关键点，并在图像中绘制出关键点位置。如图 5-3（a）为人脸 68 个关键点描绘图，图 5-3（b）为右眼及嘴巴关键点坐标。通过提取眼部特征并计算 *EAR* 判断眼睛是否闭合，以左眼为例，*EAR* 计算公式如（5-1）所示：

$$EAR=\frac{\|P_2-P_6\|+\|P_3-P_5\|}{2\|P_1-P_4\|} \tag{5-1}$$

P_2 和 P_6 之间的距离 ρ 利用欧几里得距离公式计算得到，计算公式如（5-2）所示：

$$\rho=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \tag{5-2}$$

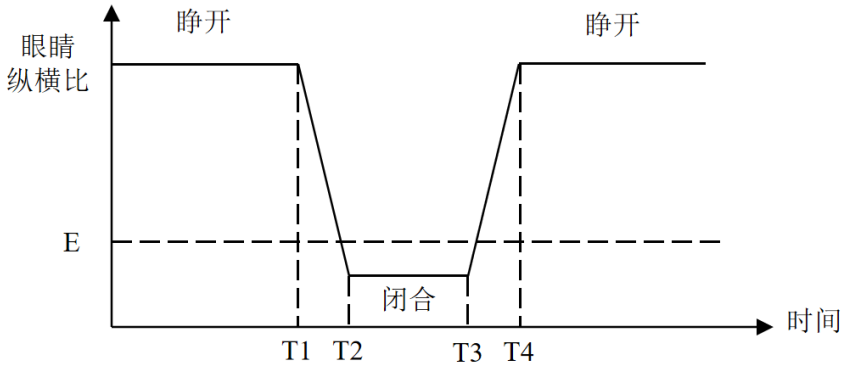


图 5-4 *EAR* 在眨眼过程中变化示意图

同理，计算出右眼的 *EAR*，然后将计算得出的两个 *EAR* 进行平均后获得最终 *EAR* 结果，得到的 *EAR* 结果结合 PERCLOS 准则，判断驾驶员是否处于疲劳状态。图 5-4 为一次正常情况下眨眼过程的 *EAR* 变化示意图，虚线 E 代表 *EAR* 的阈值。PERCLOS 是一种用于疲劳检测的指标，全称为“闭眼时间百分比”。目前 PERCLOS 有 EM 准则、P70 准则和 P80 准则三种准则判断被检测者是否处于疲劳状态。研究表明当处于驾驶状态时，P80 准则与驾驶员疲劳状态的相关性最高，即瞳孔被眼睑覆盖超 80%的面积时判定眼睛处于闭合状态，并计算检测周期时间内闭合状态所占的时间比例，因此实验中的 PERCLOS 采用 P80 准则。一般来说，闭眼时间百分比在 0%到 15%之间的驾驶员为正常状态，闭眼时间百分比在 15%到 30%之间的驾驶员可能处于疲劳状态，闭眼时间百分比超过 30%的驾驶员则极有可能处于严重的疲劳状态。闭眼时间百分比 P_{P80} 计算方式如公式（5-3）所示：

$$P_{P80}=\frac{F_{close}}{F_a} \tag{5-3}$$

其中，F_a 为检测周期时间内的总帧数，F_{close} 为检测周期内处于闭眼状态时的总帧数。

5.3.2 基于嘴部状态的疲劳判定

在疲劳检测时，可以通过嘴部张开的程度来判断人的疲劳程度。以下是几种常见的嘴部张开状态及其区别：自然状态下嘴部处于正常闭合状态，没有明显的张开。这表示人处于清醒和正常状态下。当嘴部略微张开时，这可能表明人处在交谈或自己唱歌状态下。当嘴部张开得很大，嘴唇之间有较大的间隙，

这表示人可能在打哈欠，这时通常表示人处于疲劳状态。

通过提取嘴部特征并计算嘴部纵横比（*MAR*）判断嘴巴是否打哈欠，*MAR* 计算公式为（5-4）所示：

$$MAR = \frac{\|K_2 - K_8\| + \|K_3 - K_7\| + \|K_4 - K_6\|}{3\|K_1 - K_5\|} \tag{5-4}$$

驾驶员在说话时嘴巴也处于张开状态，但当打哈欠时，*MAR* 值会持续上升，并会持续一定的时间。刘佰强在论文中通过计算实验者模拟说话和打哈欠时的 *MAR* 值，并绘制了具有两条 *MAR* 曲线的图。通过观察这个图，发现当 *MAR* 值等于 0.3 时，可以将打哈欠和说话时的 *MAR* 区分开，因此将 0.3 作为判定嘴部疲劳的阈值。相比于使用神经网络分类，计算眼部状态和嘴部状态来判断驾驶员是否疲劳的方法更加简便，同时可以提高判断的准确性。

5.3.3 眼部及嘴部检测实验结果

在完成驾驶员疲劳检测系统的设计后，对其进行了测试，并获得了图5-5所示的测试结果。图 5-5（a）和图 5-5（b）展示了正常状态下的检测结果，而图 5-5（c）和图 5-5（d）展示了闭眼状态下的检测结果。此外，图 5-5（e）和图 5-5（f）展示了打哈欠状态下的检测结果。从这些结果可以看出，检测模型基本完成了检测任务，并且相对准确地判断出了驾驶员眼部和嘴部的状态。这表明该系统在识别疲劳驾驶方面具有较高的准确性和可靠性。

图 5-5 疲劳检测结果

5.4 车载驾驶员疲劳检测系统的应用

通过在车上安装红外摄像头来获取驾驶员的驾驶视频数据。红外摄像头可以在黑暗或低光条件下工作，因为它们可以依靠红外辐射来捕捉热量分布。这使得红外摄像头在夜间驾驶或暗光环境中进行疲劳检测非常有效，无需额外的照明设备。此外，红外摄像头还可以提供详细的热量分布图像，能够捕捉到微小的变化，通过分析驾驶员的眼部区域，红外摄像头可以检测到眼睛的瞬目频率、瞳孔直径等特征，从而准确地判断疲劳状态。红外摄像头的示意图如图 5-6

（a）所示。同时为了便于摄像头固定，安装摄像头的结构如图 5-6（b）所示。

（a）红外摄像头

（b）安装摄像头的支架结构

图 5-6 红外摄像头及其安装结构

驾驶员疲劳检测摄像头在车上的安装位置如图 5-7 所示。驾驶员疲劳检测摄像头安装的位置有多种，如安装在车辆的 A 柱或方向盘管柱，这里将摄像头安装在了方向盘管柱上。摄像头安装在方向盘管柱可以靠近驾驶员，更直接、清晰地捕捉到驾驶员的眼睛和面部表情，也便于专注于检测瞳孔大小和眨眼频率等细节。相比 A 柱，方向盘管柱的位置更容易进行布线和安装，对车辆结构的改动较小。

图 5-7 摄像头在车上的安装位置

基于安装的驾驶员疲劳检测红外摄像头，采集到的驾驶员的驾驶状态视频数据如图 5-8 所示。

图 5-8 摄像头采集视频图像

基于输入的驾驶员视频数据，输出的人脸关键点点位描绘图如图 5-9 所示。

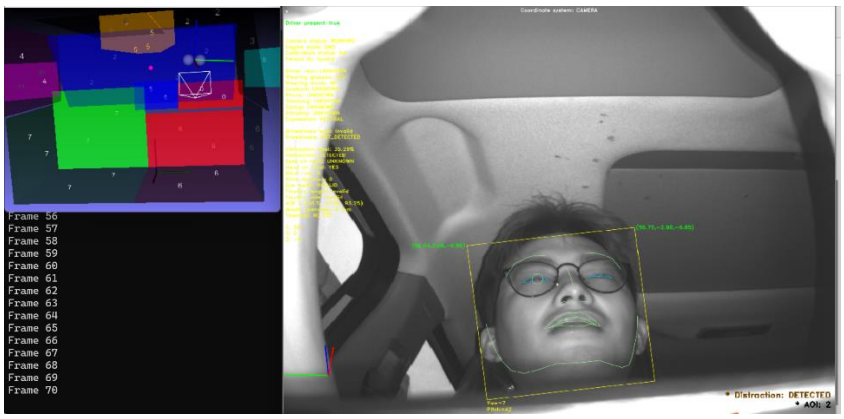
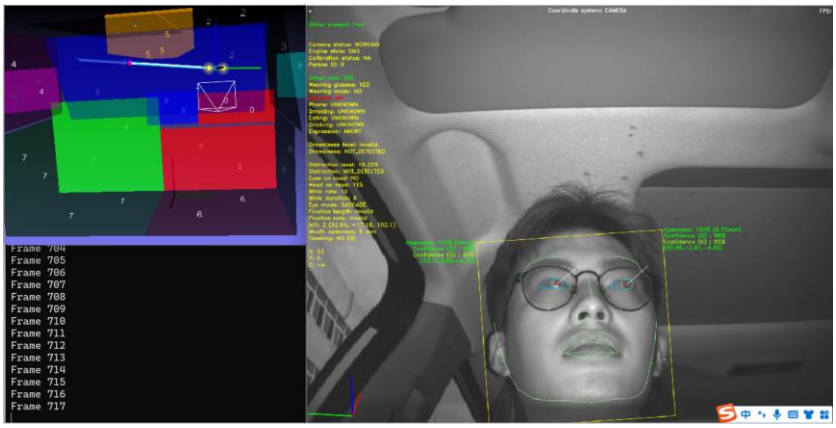
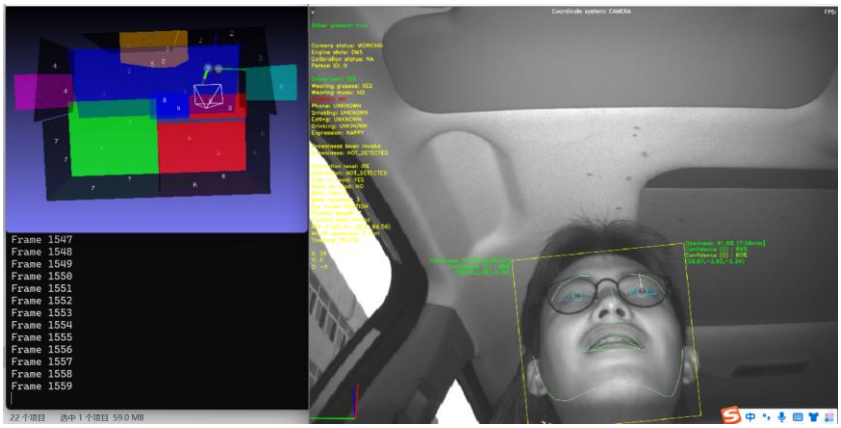


图 5-9 人脸关键点点位描绘图

以下对驾驶员疲劳检测系统功能进行测试。疲劳检测测试结果图如图 5-10 所示。通过系统在实际应用中的人脸关键点检测效果图，模型开始计算并判定疲劳检测的工作，系统对驾驶员驾驶状态的判断结果如图所示。在图中可以看到，根据输入的 Frame，驾驶员疲劳检测系统可以实时检测并计算得出眼嘴部张开大小和眼嘴部纵横比，判定眼嘴部状态。



（a）



（b）

图 5-10 疲劳检测测试结果图

5.5 实验结果与分析

YAWDD[68]是一个专门用于研究和开发驾驶员疲劳检测系统的公开数据集。YAWDD 包含了多个年龄段、性别以及种族的参与者，这样可以确保检测算法的普遍适用性。同时数据集中包含不同的光照条件，包括日间和夜间的情景，这对于开发在实际驾驶条件下都能稳定工作的疲劳检测系统至关重要。因此，YAWDD 数据集是一个极具价值的资源，能够帮助研究人员和开发者在驾驶员疲劳检测领域开展工作，进而提升驾驶安全。

由于 YAWDD 数据集包含的驾驶员特征种类较多，为了验证本文驾驶员疲劳状态检测系统的准确性，从 YAWDD 数据集中抽取只包含打哈欠和闭眼的疲劳状态视频以及正常状态的视频作为实验检测数据。最终获得了 50 条检测视频，分为 10 组，其中 1-4 组为驾驶员清醒的视频，5-10 组为驾驶员疲劳的视频，每组 5 条视频。驾驶员疲劳状态检测系统的检测结果如下表 5-1 所示。

表 5-1 驾驶员疲劳状态检测系统的检测结果

组号	PERCLOS	眨眼频率（次/30s）	打哈欠次数（次/30s）	实际疲劳数	检测疲劳数	准确率%
1	6.9%	9	0	0	0	100
2	12.5%	7	0	0	1	80
3	9.5%	8	0	0	0	100
4	7.9%	7	0	0	0	100
5	35.6%	5	2	5	5	100
6	19.7%	7	3	5	4	80
7	43.8%	4	2	5	5	100
8	31.2%	6	2	5	5	100
9	22.5%	6	2	5	5	100
10	29.8%	5	3	5	4	80

通过表 5-1 可以看出，本文驾驶员疲劳状态检测系统基本完成了检测任务，驾驶员清醒的检测准确率为 95%，驾驶员疲劳的检测准确率为 93.3%，平均检测准确率达到了 94%。在第 2 组中，将清醒误判为疲劳，主要是因为被检测者的眼睛过小，造成系统误检为眼睛闭合，但从整体的实验结果来看，本文设定的眼睛闭合阈值是具有一定准确性的。在第 6 和第 10 组中，分别出现一例漏检，主要是因为由于驾驶环境（如拍摄角度遮挡等）和人在疲劳状态时的生理和心理状态差异（如有些人疲劳状态表现为目光散漫）等问题造成的误差，针对这些潜在的漏检原因，疲劳检测系统的开发需要不断地优化算法，增加和多样化训练数据集，提高系统对不同环境和个体的适应能力，以减少漏检的发生。

5.6 本章小结

为了解决由于疲劳驾驶带来的安全问题，并将本文研究算法应用于实际当中，本章根据前几章研究的算法设计了一套驾驶员疲劳检测系统。首先，介绍了疲劳状态检测系统整体框架和驾驶员疲劳状态判定的依据，并以此依据展示了检测的实验结果图。然后，展示了驾驶员疲劳检测系统在家用汽车上的应用，通过红外摄像头输入视频数据，系统描绘出人脸关键点点位并计算眼嘴部张开大小和纵横比，判定眼嘴部状态。最后在 YAWDD 数据集上准确率达到了 94%，证明了本文提出的驾驶员疲劳检测系统的有效性。

第 6 章 总结与展望

6.1 工作总结

针对疲劳驾驶这一问题做了相关的研究和实验证明，根据应用场景对目标检测算法的需求，提出了实时性较高的基于 YOLOv5s 的算法模型和应对复杂场景下的基于 YOLOv7 的算法模型，并基于 Dlib 库和结合 PERCLOS 识别判定疲劳状态。详细工作内容如下：

（1）详细探讨了疲劳驾驶检测领域以及深度学习的相关研究方法和成果，分析并总结了疲劳驾驶检测的研究现状和相关理论基础。

（2）在轻量级疲劳驾驶检测任务中，首先，采用深度重参数化卷积层替换传统的卷积层，通过深度卷积和传统卷积组合的复合核增加可学习的参数，增强了模型的表达能力，提高了模型的收敛速度和检测性能。其次，在模型的深层网络层中加入了高效多尺度注意力机制，通过在不同的尺度上提取特征，获取更加全面的信息，可以对不同大小的检测目标都具有很好的适应性，提高检测的准确率和鲁棒性。

（3）在多特征信息融合疲劳驾驶检测任务中，首先，本文设计了基于 SE 注意力改进的 DS-Conv 模块。在 DS-Conv 模块中，通过添加的几个额外卷积层增强了模块的性能。第一个卷积层将输入的通道数增加至原来的四倍，第二个深度可分离卷积层减少参数量和降低计算量的同时，获取各个通道的信息，最后一个卷积层在学习各个通道的权重后，融合各个通道的特征信息。通过多通道信息的提取并加以融合，可以更好地适应不同的输入情况和场景变化，提高了模型的鲁棒性和检测性能。其次，在特征提取层中添加了多尺度特征提取 MSS 注意力模块，捕捉目标的细节和上下文信息。通过综合不同尺度的特征信息，模型可以更准确地定位目标，减少定位造成的误差。之后融合不同尺度的特征，提供更全面和丰富的信息，有助于提高模型的准确性、鲁棒性和泛化能力，使其在各种场景中更加有效和可靠。

（4）提出了一套驾驶员疲劳检测系统，该系统能够实时检测驾驶员在疲劳状态下表现出的特征。首先，安装在方向盘管柱上的摄像头用于输入驾驶员的

视频数据。系统通过对输入图像进行处理，检测人脸关键点的位置，并计算眼嘴部的张开大小和纵横比，以判断驾驶员的眼嘴部状态，同时计算闭眼时间占比和打哈欠次数等疲劳参数。通过综合计算这些参数的结果，系统能够对驾驶员的疲劳状态进行判定，这种综合判定方法能够更准确地评估驾驶员的疲劳程度。

根据以上分析，本文提出的基于深度学习的疲劳驾驶检测技术研究具有以下优点：

（1）提出的轻量级疲劳驾驶检测算法在 WIDER FACE 数据集上平均精度 mAP 达到了 76.8%，较基线算法提升了 2.2%。与同级别参数下的其它模型精度相比，比 YOLOv7-tiny 高出 4.2%，比 YOLOv6s 高出 13.9%，比 YOLOv3-tiny 高出 21.7%，具有参数量少，准确率高的优点。

（2）提出的多特征信息融合疲劳驾驶检测算法在 WIDER FACE 数据集的 Easy、Medium 和 Hard 子集上平均精度 mAP 分别达到了 96.0%、94.6%和 88.1%，大幅度提高了在复杂环境下的人脸检测精度。相较于目前同级别各种先进算法，在检测精度上得到显著提升。

（3）提出的驾驶员疲劳状态检测系统应用，在实际测试中基本完成了检测任务，并在 YAWDD 数据集上，检测准确率达到了 94%。同时本文提出的驾驶员疲劳状态检测系统成本低，不妨碍驾驶员的驾驶活动，达到做为车载驾驶员疲劳检测的要求。

6.2 工作展望

本文的疲劳驾驶检测研究工作仍然存在一定的不足性。

第一，虽然基于 YOLOv5s 的疲劳检测算法具有较少的参数和实时性高的特点，但其检测精度在对于疲劳驾驶等任务的要求上仍存在一定不足。相比之下，基于 YOLOv7 改进的疲劳检测算法在 WIDER FACE 数据集的 Easy、Medium 和 Hard 子集上表现出较高的精度，但其参数量达到了 49.5MB，对于实时性的要求仍有待改进。

第二，随着深度学习的不断发展，计算机视觉人脸检测领域涌现出了性能更好的人脸检测算法，例如 YOLOv8 等。然而，本文未及时进行对新算法的对比实验，以评估其在疲劳检测任务中的性能表现。

第三，本文的疲劳状态测试数据量较少，因此在特殊情况下，例如眼睛过小的情况下，可能存在误判的可能性。为了改进这一问题，后续可以考虑增加驾驶员眼部和嘴部信息的存储系统，以提高对不同情况的准确性和鲁棒性。

参考文献

[1] Beirness D J, Simpson H M, Desmond K. The road safety monitor 2004: Drowsy driving[R]. Ottawa: TrafficInjury Research Foundation, 2005.

[2] Watling C N, Armstrong K A, Radun I. Examining signs of driver sleepiness, usage of sleepiness countermeasures and the associations with sleepy driving behaviours and individual factors[J]. Accident Analysis & Prevention, 2015, 85: 22-29.

[3] 道路安全行动十年. 2021-2030 全球计划[DB/OL].https://www.who.int/zh/publications/m/item/global-plan-for-the-decade-of-action-for-road-safety-2021-2030: 世界卫生组织, October 20, 2021

[4] Maclean A W, Davies D R T, Thiele K. The hazards and prevention of driving while sleepy[J]. Sleep Medicine Reviews, 2003, 7(6): 507-521.

[5] 中华人民共和国国家统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2011-2021.

[6] Davidović J, Pešić D, Antić B. Professional drivers' fatigue as a problem of the modern era[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2018, 55: 199-209.

[7] 陈见哲. 疲劳驾驶检测方法研究进展[J]. 汽车实用技术, 2023, 48(21): 179-186.

[8] Prabhakar S K, Won D O. Multiple robust approaches for EEG-based driving fatigue detection and classification[J]. Array, 2023, 19: 100320.

[9] Beles H, Vesselenyi T, Rus A, et al. Driver drowsiness multi-method detection for vehicles with autonomous driving functions[J]. Sensors, 2024, 24(5): 1541.

[10] Jap B T, Lal S, Fischer P, et al. Using EEG spectral components to assess algorithms for detecting fatigue[J]. Expert Systems with Applications, 2009, 36(2): 2352-2359.

[11] Min J, Xiong C, Zhang Y, et al. Driver fatigue detection based on prefrontal EEG using multi-entropy measures and hybrid model[J]. Biomedical Signal Processing and Control, 2021, 69: 102857.

[12] Kecklund G, Akerstedt T. Sleepiness in long distance truck driving: An ambulatory EEG study of night driving[J]. Ergonomics, 1998, 36(9): 1007.

[13] Halomoan J, Ramli K, Sudiana D, et al. ECG-based driving fatigue detection using heart rate variability analysis with mutual information[J]. Information, 2023, 14(10): 539.

[14] Balasubramanian V, Adalarasu K. EMG-based analysis of change in muscle activity during simulated driving[J]. Journal of Bodywork and Movement Therapies, 2007, 11(2): 151-158.

[15] Lawoyin S, Fei D Y, Bai O. Accelerometer-based steering-wheel movement monitoring for drowsy-driving detection[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of automobile engineering, 2015, 229(2): 163-173.

[16] Devi M S, Bajaj P R. Fuzzy based driver fatigue detection[C]. Proceedings of the International Conference on Systems, Man and Cybernetics. 2010: 3139-3144.

[17] Tian Y, Zhang Q, Ren Z, et al. Multi-scale dilated convolution network based depth estimation in intelligent transportation systems[J]. IEEE Access, 2019, 7: 185179-185188.

[18] Zhang R H, He Z C, Wang H W, et al. Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system[J]. IEEE Access, 2017, 5: 6649-6660.

[19] Shi S Y, Tang W Z, Wang Y Y. A review on fatigue driving detection[J]. ITM Web of Conferences, 2017, 12: 14-21.

[20] Singh H, Bhatia J S, Kaur J. Eye tracking based driver fatigue monitoring and warning system[C]. Proceedings of the India International Conference on Power Electronics. 2011: 1-6.

[21] Li D, Zhang X, Liu X, et al. Driver fatigue detection based on comprehensive facial features and gated recurrent unit[J]. Journal of Real-Time Image Processing, 2023, 20(2): 19.

[22] Chen X, Luo X, Liu X, et al. Eyes localization algorithm based on prior MTCNN face detection[C]. Proceedings of International Information Technology and Artificial Intelligence Conference. 2019: 1763-1767.

[23] Kurniawan A, Erlangga E, Tanjung T, et al. Review of deep learning using

convolutional neural network model[J]. Engineering Headway, 2024, 3: 49-55.

[24] Apicella A, Donnarumma F, Isgrò F, et al. A survey on modern trainable activation functions[J]. Neural Networks, 2021, 138: 14-32.

[25] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]. Proceedings of the International Conference on Computer Vision. 2017: 2961-2969.

[26] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2014: 580-587.

[27] Girshick R. Fast R-CNN[C]. Proceedings of the International Conference on Computer Vision. 2015: 1440-1448.

[28] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017, 39(6): 1137-1149.

[29] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2016: 779-788.

[30] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2017: 7263-7271.

[31] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[C]. Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition. 2020:321-333.

[32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of the International Conference on Learning Representations. 2015: 1446-1466.

[33] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[34] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[35] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. Proceedings of the European Conference on Computer Vision. 2016: 21-37.

[36] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2020: 10781-10790.

[37] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. Proceedings of the International Conference on Computer Vision. 2017: 2980-2988.

[38] 仲鹏宇, 杨娟. 基于轻量级双模态 SSD 算法的疲劳驾驶检测[J]. 电子技术与软件工程, 2022, No. 221(3): 145-149.

[39] 李京徽, 郭昕刚. 一种疲劳驾驶检测方法[J]. 长春工业大学学报, 2021, 42(04）：371-377.

[40] 臧露奇. 基于 YOLOv7 改进的人脸检测算法[D]. 南昌: 南昌大学, 2023.

[41] Xu Z, Bai H, Xiao J, et al. Occluded and tiny face detection with deep and shallow features fusion and compensation[J]. Intelligent Information Hiding and Multimedia Signal Processing. 2022, 277: 181-190.

[42] Jia H, Xiao Z, Ji P. Real-time fatigue driving detection system based on multi-module fusion[J]. Computers & Graphics, 2022, 108: 22-33.

[43] Florez R, Palomino-Quispe F, Coaquira-Castillo R J, et al. A CNN-based approach for driver drowsiness detection by real-time eye state identification[J]. Applied Sciences, 2023, 13(13): 7849.

[44] Kim M, Jain A K, Liu X. Adaface: Quality adaptive margin for face recognition[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2022: 18750-18759.

[45] Qu J, Wei Z, Han Y. An embedded device-oriented fatigue driving detection method based on a YOLOv5s[J]. Neural Computing and Applications, 2024, 36(7): 3711-3723.

[46] Deshpande A V, Subashini M M. Transfer learning-based drowsiness detection system for driver assistance and classification of traffic signs employing a deep convolutional neural network[J]. International Journal on Artificial Intelligence Tools, 2023, 32(08): 2350035.

[47] Cao J, Li Y, Sun M, et al. DO-Conv: Depthwise over-parameterized convolutional layer[J]. IEEE Transactions on Image Processing, 2022, 31: 3726-3736.

[48] Ouyang D, He S, Zhang G, et al. Efficient multi-Scale attention module with cross-spatial learning[C]. Proceedings of the International Conference on Acoustics, Speech and Signal Processing. 2023: 1-5.

[49] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2021: 13713-13722.

[50] Yang S, Luo P, Loy C C, et al. Wider face: A face detection benchmark[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2016: 5525-5533.

[51] Lin W, Wu Z, Chen J, et al. Scale-aware modulation meet transformer[C]. Proceedings of the International Conference on Computer Vision. 2023: 6015-6026.

[52] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[53] Arbeláez P, Pont-Tuset J, Barron J T, et al. Multiscale combinatorial grouping[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2014: 328-335.

[54] Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104: 154-171.

[55] Zitnick C L, Dollár P. Edge boxes: Locating object proposals from edges[C]. Proceedings of the European Conference on Computer Vision. 2014: 391-405.

[56] Hosang J, Benenson R, Schiele B. How good are detection proposals, really?[C]. Proceedings of the British Machine Vision Conference. 2014: 5432-5444

[57] Li J, Wang Y, Wang C, et al. DSFD: Dual shot face detector[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 5060-5069.

[58] Zhang S, Wen L, Shi H, et al. Single-shot scale-aware network for real-time face detection[J]. International Journal of Computer Vision, 2019, 127: 537-559.

[59] Zhang J, Wu X, Hoi S C H, et al. Feature agglomeration networks for single stage face detection[J]. Neurocomputing, 2020, 380: 180-189.

[60] Najibi M, Singh B, Davis L S. FA-RPN: Floating region proposals for face

detection[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2019: 7723-7732.

[61]董子平, 陈世国, 廖国清. 基于 YOLOv5s 的密集多人脸检测算法[J]. 计算机工程与科学, 2023, 45(10): 1838-1846.

[62]Zhang S, Zhu X, Lei Z, et al. S3FD: Single shot scale-invariant face detector[C]. Proceedings of the International Conference on Computer Vision. 2017: 192-201.

[63]Hu P, Ramanan D. Finding tiny faces[C]. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2017: 951-959.

[64]Najibi M, Samangouei P, Chellappa R, et al. SSH: Single stage headless face detector[C]. Proceedings of the International Conference on Computer Vision. 2017: 4875-4884.

[65]Zhu C, Zheng Y, Luu K, et al. CMS-RCNN: Contextual multi-scale region-based CNN for unconstrained face detection[J]. Deep Learning for Biometrics, 2017: 57-79.

[66]Hasan M K, Ahsan M S, Newaz S H S, et al. Human face detection techniques: A comprehensive review and future research directions[J]. Electronics, 2021, 10(19): 2354.

[67]杨艳艳, 李雷孝, 林浩. 提取驾驶员面部特征的疲劳驾驶检测研究综述[J]. 计算机科学与探索, 2023, 17(06): 1249-1267.

[68]Abtahi S, Omidyeganeh M, Shirmohammadi S, et al. YawDD: A yawning detection dataset[C]. Proceedings of the Multimedia Systems Conference. 2014: 24-28.

致谢

时光荏苒，三年飞快地过完了。

首先，感谢恩师王凤随老师辛苦的教导，对于我在学术研究方面，提出了很多关键性的建议，帮助我克服了一个个难关。在一年四季，老师总是来得最早，走的最迟，有一次半夜一点因事回实验室，看到老师还在实验室工作。

其次，感谢我的对象董舒婷对我英文摘要的帮助，让一名刚过四级的苦瓜英语学习者顺利完成复杂的英文摘要。

然后，我要感谢我的同门，对我学术上的颇多指导，一起学习，一起进步，让学习变得快乐。

同时，我要感谢我的家人，他们永远是我坚实的后盾，在我低沉时鼓励我，消极时开导我，因为有了他们，让我学习有了更大的动力。

最后对评阅本论文的专家们报以真诚的感谢！

张兴旺

2024.5.24

攻读硕士学位期间的研究成果

1 发表学术论文情况

[1] 张兴旺, 王凤随, 杨海燕. 基于多特征信息融合的疲劳驾驶检测算法[J/OL]. 重庆工商大学学报(自然科学版): 1-11.

[2] 杨海燕, 王凤随, 张兴旺. 基于上下文增强和特征融合的目标检测算法[J/OL]. 重庆工商大学学报(自然科学版): 1-8.

2 参与科研项目情况

（1）安徽省自然科学基金(2108085MF197)

（2）安徽高校省级自然科学研究重点项目(KJ2019A0162)

（3）安徽工程大学国家自然科学基金预研项目(Xjky2022040)