



浙江工业大学

硕士学位论文

基于机器视觉的驾驶员注意力检测系统设计

作者姓名	唐敏
指导教师	赵云波 教授
学科专业	控制工程
学位类型	工程硕士
培养类别	全日制专业型硕士
所在学院	信息工程学院

提交日期：2021 年 1 月

Design of Driver Attention Detection System Based on Machine Vision

Dissertation Submitted to

Zhejiang University of Technology

in partial fulfillment of the requirement

for the degree of

Master of Engineering



by

Dissertation Supervisor: Yun-Bo Zhao

Jan., 2021

浙江工业大学学位论文原创性声明

本人郑重声明：所提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研究成果。除文中已经加以标注引用的内容外，本论文不包含其他个人或集体已经发表或撰写过的研究成果，也不含为获得浙江工业大学或其它教育机构的学位证书而使用过的材料。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

作者签名：唐敏

日期：2020 年 12 月

学位论文版权使用授权书

本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，允许论文被查阅和借阅。本人授权浙江工业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。

本学位论文属于 1、保密□，在一年解密后适用本授权书。

2、保密□，在二年解密后适用本授权书。

3、保密□，在三年解密后适用本授权书。

4、不保密 ☒。

（请在以上相应方框内打“√”）

作者签名：唐敏

日期：2020 年 12 月

导师签名：赵一波

日期：2020 年 12 月

中图分类号 TH741

学校代码 10337

UDC 681.2

密级 公开

研究生类别 全日制专业型硕士研究生



浙江工业大学

工程硕士学位论文

基于机器视觉的驾驶员注意力检测系统设计

Design of Driver Attention Detection System Based on
Machine Vision

作者姓名 唐敏

第一导师 赵云波

学位类型 工程硕士

学科专业 控制工程

培养单位 信息工程学院

研究方向 人工智能

答辩日期: 2020 年 12 月 04 日

基于机器视觉的驾驶员注意力检测系统设计

摘 要

驾驶员注意力不集中是指驾驶员的注意力被一些分心驾驶行为所分散从而不集中。车辆驾驶过程中的注意力分散行为将会使得驾驶员的反应过程被拉长,进而使得驾驶员对车辆能否安全行驶的判断能力和对突发情况的反应能力降低,极易引起交通事故。近年来,驾驶员注意力检测已经越来越多地受到大家的关注。

本文针对驾驶员注意力检测场景,设计了基于机器视觉的驾驶员注意力检测系统,进行了驾驶员注意力检测系统的软硬件设计,设计的基于多粒度特征和中间层特征的 MGMN 算法在现有的公开的分心驾驶数据集中取得了更高的精确度,设计的硬件系统成本较低并且在真实场景实验中也有着较高的实用性。本文的主要工作有以下两个方面:

(1) 设计了基于多粒度特征和中间层特征的驾驶员注意力检测算法,该算法具有精度高、特征提取丰富且准确的特点。考虑到驾驶员在驾驶时头部和手部都处于图像的相对固定的位置,并且头部一般处于图片的上方,手部一般处于图片的中下方。如果可以针对驾驶图像的每一部分进行学习,提取每一部分特有的特征,本文在 ResNet50 模型中加入了多粒度特征提取部分,该部分有全局分支、二分支和三分支三部分,共同提取驾驶图像的全局和局部特征。同时又抽取了驾驶图像的中间层特征,中间层包含了丰富的图像特征,最后将多粒度特征和中间层特征一起输入到 Softmax 损失函数中,设计出了 MGMN 算法。MGMN 算法无论是在现有的分心驾驶检测公开数据集上还是在自建数据集上都有着更高的精确度和较短的运行时间。

(2) 设计了“树莓派+远程服务器”的驾驶员注意力检测系统。采用“树莓派+远程服务器”的结构,通过局域网将树莓派与远程服务器相连接,具有较高的信息传输速率,树莓派与服务器之间采用 Socket 通信方式进行通信,做到终端与服务器之间的信息的准确传输。同时采用成本低、体积小、可靠性高的树莓派作为驾驶图像实时采集终端,采用配置高、运行效率高的服务器作为状态判别的设备,降低了设计成本,实现了对驾驶员注意力状态实时的高精度检测。

关键词: 驾驶员, 注意力检测系统, 多粒度特征, 中间层特征, 卷积神经网络

DESIGN OF DRIVER ATTENTION DETECTION SYSTEM BASED ON MACHINE VISION

ABSTRACT

Driver inattention means that the driver's attention is distracted by some distracted driving behaviors and thus is not concentrated. The distracting behavior during the driving of the vehicle will lengthen the driver's reaction process, thereby reducing the driver's ability to judge whether the vehicle is safe to drive and the ability to respond to emergencies, which can easily cause traffic accidents. In recent years, driver attention detection has attracted more and more attention.

Aiming at the driver's attention detection scene, this paper designs a driver's attention detection system based on machine vision, and carries out the software and hardware design of the driver's attention detection system. The designed MGMN algorithm based on multi-granularity features and middle layer features is in the existing The public distracted driving data set has achieved higher accuracy, the designed hardware system has a lower cost and has a higher practicability in real-scene experiments. The main work of this paper has the following two aspects:

(1) A driver attention detection algorithm based on multi-granularity features and mid-level features is designed. The algorithm has the characteristics of high accuracy, rich feature extraction and accuracy. Considering that the driver's head and hands are in relatively fixed positions of the image when driving, and the head is generally above the picture, and the hands are generally below the middle of the picture, if you can learn for each part of the driving image, To extract the unique features of each part, we added a multi-granularity feature extraction part in the ResNet50 model. This part is divided into three parts: the global branch, the second branch and the third branch to jointly extract the global and local features of the driving image. At the same time, the middle layer features of the driving image are extracted. The middle layer contains rich image features. Finally, the multi-granularity features and the middle layer features are input into the Softmax loss function together, and the MGMN algorithm is designed. The MGMN algorithm has higher accuracy and shorter running time on both the existing public data of distracted driving detection and the self-built data set.

(2) This paper designs the driver's attention detection system of "Raspberry Pi + remote server". Using the structure of "Raspberry Pi + Remote Server", the Raspberry Pi is connected to the remote server through the local area network, which has a higher

information transmission rate. The Raspberry Pi and the server use Socket communication to communicate, so that the terminal can communicate with the server. Accurate transmission of information between servers. At the same time, the Raspberry Pi with low cost, small size and high reliability is used as the real-time acquisition terminal for driving images, and the server with high configuration and high operation efficiency is used as the equipment for state discrimination, which reduces the design cost and realizes the state of attention to the driver. Real-time high-precision detection.

KEY WORDS: Driver, attention detection system, multi-granularity features, middle layer features, convolutional neural network

目 录

摘 要.....	I
ABSTRACT.....	II
插图清单.....	VII
表格清单.....	IX
第一章 绪 论.....	1
1.1 研究背景和意义.....	1
1.2 国内外研究现状.....	3
1.2.1 基于驾驶员生理特征的检测方法.....	4
1.2.2 基于机动车行为特征的检测方法.....	4
1.2.3 基于驾驶员面部特征分析的检测方法.....	5
1.2.4 基于驾驶员动作的检测方法.....	7
1.3 技术路线.....	8
1.4 本文组织架构.....	9
1.5 本章小结.....	9
第二章 卷积神经网络综述.....	10
2.1 人工神经网络.....	10
2.1.1 人工神经元.....	10
2.1.2 人工神经网络.....	12
2.1.3 多层感知机.....	14
2.1.4 反向传播算法.....	15
2.2 卷积神经网络原理.....	16
2.2.1 卷积层.....	17
2.2.2 激活层.....	19
2.2.3 池化层.....	21
2.2.4 全连接层.....	21
2.3 深度学习框架 Pytorch 和计算机视觉库 Opencv 介绍	22

2.3.1 深度学习框架 Pytorch	22
2.3.2 计算机视觉库 Opencv	23
2.4 本章小结.....	24
第三章 驾驶员注意力检测算法设计.....	25
3.1 驾驶员注意力状态分类及其特征.....	25
3.1.1 驾驶员注意力状态分类.....	25
3.1.2 驾驶员注意力各状态特征.....	26
3.2 MGMN 算法设计	28
3.2.1 多粒度特征的提取.....	28
3.2.2 中间层特征的提取.....	31
3.2.3 损失函数设计.....	32
3.3 本章小结.....	34
第四章 MGMN 算法实验与结果对比分析.....	35
4.1 实验设置与环境.....	35
4.2 实验所用数据集和处理方法.....	36
4.2.1 驾驶员注意力检测数据集.....	36
4.2.2 驾驶员注意力检测数据集的处理.....	37
4.3 算法精度实验与分析.....	38
4.3.1 实验设置.....	38
4.3.2 实验结果对比与分析.....	38
4.4 算法特征提取能力实验与分析.....	42
4.4.1 实验设置.....	42
4.4.2 实验结果与分析.....	43
4.5 本章小节.....	46
第五章 驾驶员注意力检测系统开发.....	47
5.1 系统总设计.....	47
5.1.1 系统软硬件的功能需求分析.....	47
5.1.2 硬件平台介绍.....	49
5.1.3 Socket.....	51
5.2 驾驶员注意力检测系统验证.....	51

5.2.1 实验环境与设置.....	51
5.2.2 实验结果分析.....	52
5.2.3 本章小节.....	53
第六章 总结与展望.....	54
6.1 总结.....	54
6.2 展望.....	55
参考文献.....	57
致 谢.....	60
作者简介.....	62
1 作者简历.....	62
2 攻读硕士学位期间发表的学术论文.....	62
3 参与的科研项目及获奖情况.....	62
4 发明专利.....	62
学位论文数据集.....	64

插图清单

图 1-1 2013 年~2018 年全国交通事故发生情况汇总分析	2
图 2-1 神经元	10
图 2-2 M-P 神经元模型	11
图 2-3 Sigmoid 函数	12
图 2-4 神经网络结构	13
图 2-5 单隐层感知机	14
图 2-6 卷积神经网络结构	17
图 2-7 卷积操作矩阵	18
图 2-8 卷积过程	18
图 2-9 卷积结果	18
图 2-10 tanh 函数	19
图 2-11 ReLU 和 PReLU 激活函数	20
图 2-12 两种典型的池化方式	21
图 2-13 全连接层	22
图 3-1 驾驶员注意力不集中行为	26
图 3-2 驾驶员处于注意力集中状态的特征	27
图 3-3 驾驶员注意力不集中行为的特征	27
图 3-4 驾驶员使用右手收发消息时的特征热力图	28
图 3-5 MGMN 算法结构图	29
图 3-6 多粒度特征提取模块	30
图 3-7 MGMN 算法的三支结果示意图	31
图 3-8 MGMN 算法对图片的中间层特征提取过程	31
图 3-9 Softmax 函数运算过程	33
图 4-1 State Farm 的 “Districted-Driver-Detection” 数据集	36
图 4-2 “AUC-Districted-Driver-Detection” 数据集	37
图 4-3 “ZJUT- Districted-Driver-Detection” 数据集	37
图 4-4 不同的卷积神经网络在 State Farm 公司的数据集上的结果的对比	39
图 4-5 不同的卷积神经网络在 “AUC-Districted-Driver-Detection-Clean” 数据集上的结果的对比	40
图 4-6 不同卷积神经网络在 “ZJUT-Districted-Driver-Detection” 数据集上的结果的对比	42
图 4-7 MGMN 算法在 State Farm 公司的数据集上的热力图	43

图 4-8 ResNet50 算法在 State Farm 公司的数据集上的热力图	43
图 4-9 MGMN 算法在 “AUC-Districted-Driver-Detection-Clean” 数据集上的热力图	44
图 4-10 ResNet50 算法在 “AUC-Districted-Driver-Detection-Clean” 数据集上的热力图	45
图 4-11 MGMN 算法在 “ZJUT- Districted-Driver-Detection” 数据集上的热力图	45
图 4-12 ResNet50 算法在 “ZJUT- Districted-Driver-Detection” 数据集上的热力图	46
图 5-1 系统架构图	48
图 5-2 两位志愿者	52

表格清单

表 3-1 ResNet50 模型结构.....29

表 4-1 实验环境35

表 4-2 MGMN 算法在 State Farm 公司的数据集上的实验结果39

表 4-3 MGMN 算法在 “AUC-Districted-Driver-Detection-Clean” 数据集上的实验结果.....40

表 4-4 MGMN 算法在 “ZJUT- Districted-Driver-Detection” 数据集上的实验结果41

表 5-1 树莓派 4B 参数表.....49

表 5-2 摄像头参数表50

表 5-3 服务器配置50

表 5-4 男性志愿者的实验结果52

表 5-5 女性志愿者的实验结果53

第一章 绪 论

驾驶员注意力不集中在世界范围内都是引起严重道路交通事故的最常见原因之一，特别是在汽车和驾驶员的数量急剧增长的今天，与驾驶员注意力不集中相关的事故频繁发生，这引起了社会的高度关注，多年来，检测驾驶员注意力状态一直是研究的热点问题，迄今为止已经出现了很多检测方法。本章概述了用于检测驾驶员注意力状态的当前现有方法。

1.1 研究背景和意义

随着我国工业的不断发展，社会的不断进步，人们的生活水平在得到了不断的提高的同时，汽车时代也随之而来。汽车作为社会上最为普通的代步工具之一，其数量也在近几年有了大幅度的增长，汽车行业也成为了我国的一个重要的产业。据中华人民共和国公安部的数据显示，2019 年全国新注册登记汽车 2500 余万辆，全国汽车保有量达到 26000 万辆，同时全国汽车驾驶员数量达到近 4 亿人^[1]。2020 年 4 月 9 日，国务院联防联控机制在北京举行的发布会指出：中国的汽车保有量有望在 2020 年底超越美国，成为世界最大的汽车拥有国^[2]。

但就在同时，伴随着汽车和驾驶员数量的不断增长而来的是越来越频发、越来越严重的交通安全问题。据世界卫生组织报道，全球每年有 120 余万人死于交通事故。我国每年仍约有 20 万起交通事故并且有 6 万人死于其中，有些年份这个数字甚至超过 8 万，同时每年还有 20 余万人遭受交通事故带来的非致命伤害^[3]。这给个人、家庭和国家都带来了巨大的伤害，每年全国交通事故带来的直接经济损失就有 10 余亿元，多年以来，我国由此带来的人和财产损失都位居世界第一。如今，我国的汽车保有量不断增长，若不采取一些有针对性的措施，我国的交通安全情况可能会越来越糟糕。图 1-1 为 2013 年~2018 年全国交通事故发生情况汇总分析。

随着人工智能的不断发展，无人驾驶汽车已成为必然的发展趋势，2020 年 4 月 19 日，百度的 Apollo Robotaxi 服务的 Dutaxi 正式在长沙落地，虽然说长沙市民已经可以打到无人驾驶出租，但是这些无人驾驶出租依然有两名安全员，这两名安全员会在车辆遇到紧急状况时进行人为干预。2016 年，美国交通运输部 DOT 部采纳了 SAE 颁布的将智能汽车从 0 到 5 进行分级的分类标准 SAEJ016^[4]。在该分类标准中：0 级指由驾驶员独立完成所有驾驶任务，5 级是指由汽车的无人驾驶系统独立完成所有的驾驶操作。目前来看，本文所研究的车辆的自主性已经到了 3 级，也就是可以实现有条件的自动化。人类驾驶员在车辆行驶的驾

驶前进的过程中会有一些预见性，而机器只会有一定的危险距离内进行反应，因此在未来较长的时间内，人机共驾仍会成为主流，虽然现在有些车辆已经将一些先进的技术加入到汽车的辅助驾驶系统 ADAS 中去，但是距离车辆可以完全自主的无人驾驶还有很长的一段路要走^[5]。



图 1-1 2013 年~2018 年全国交通事故发生情况汇总分析

Figure 1-1. Summary of traffic accidents across the country from 2013 to 2018

人机共驾是指由驾驶员和汽车的驾驶系统共同完成对车辆的控制。人机共驾系统需要确定驾驶员的自主性边界，自主性边界是指驾驶员可以独自安全地进行车辆驾驶的一定的范围，当其接近自主性边界时，驾驶员自动驾驶汽车的安全性会下降，一旦其越过了自主性边界，系统将不允许驾驶员独自进行车辆的驾驶，此时就需要系统进行介入。在人机共驾系统中，驾驶员的注意力集中状态是自主性边界的一个重要组成部分，若其注意力不集中，驾驶的安全性将会降低，很有可能会引起驾驶事故。因此人机共驾系统需要能够对驾驶员的注意力状况进行检测和判别，随后系统根据驾驶员的注意力状态判断是否需要介入。目前，汽车市场上对于驾驶辅助系统 ADAS 的需求也越来越多，而驾驶员注意力检测也是其中一个非常重要的需求，通过对驾驶员的注意力检测，可以有效提高行车的安全性。

驾驶员注意力不集中是指驾驶员在进行车辆的驾驶过程中注意力被一些行为所分散从而不集中。车辆驾驶过程中的注意力分散行为将会使得驾驶员的反应过程被拉长，从而使得其对车辆能否安全行驶的判断能力和突发情况的反应能力降低。同时，注意力分散行为会使得驾驶员对道路环境的观察减少，极易引起交通安全事故，每年全国都会有很多因为开车玩手机打电话等行为引起的惨烈车祸。我国公安部交通管理科学研究所的分析，有超过 93% 的交通安全事故是由于驾驶员操作汽车失误的因素导致的，而认知和决策失误引起了四分之三以上的事故，在这四分之三中，驾驶的疲劳驾驶和注意力不集中占了极大的

比例^[6]。根据美国交通管理部门的统计结果，每年有三成以上的交通死亡事故是由驾驶员疲劳驾驶和注意力不集中引起的。

随着信息社会的不断发展，据中华人民共和国工信部数据显示，截至 2019 年底，我国的移动电话用户普及率达 114.4 部/百人，手机上网用户达到 16 亿人次^[7]。手机普及率的不断升高，人们与外界之间的联系变得越来越必不可少，即使是在驾车的过程中，不少人仍会忍不住掏出手机，接打电话、回复微信或者查看微博消息，现在警方已经可以通过路边的监控摄像头获取汽车行驶时的图像，并且利用图像识别技术分类识别出驾驶员是否在使用手机，如果发现驾驶员在行车过程中有使用手机等情况，将会对其进行处罚。尽管如此，驾驶员开车接打电话等行为仍然屡禁不止，每年都会因此发生大量交通事故。除此之外，还有很多引起驾驶员注意力分散的行为，诸如和车上的乘客交谈、车上饮食、使用收音机、查询地图导航和到后座取物等等，不安全的、分散注意力的驾驶行为就时刻发生驾驶的过程中，除非驾驶员全神贯注于驾驶任务，否则将无法安全驾驶。驾驶员从事的任何非驾驶活动都可能分散注意力，并增加发生车祸的风险。如果不对驾驶员的注意力进行检测并进行预警，驾驶行为的安全性就得不到有效的保障。

由此可见，研究并开发一套切实有效的驾驶员注意力检测系统，对驾驶员进行注意力检测和判别，在其注意力分散的情况下进行预警并且采取一定行之有效的措施提醒驾驶员要集中注意力。对提高驾驶员行车的安全系数和降低道路交通事故的发生率有着重要的意义。

1.2 国内外研究现状

驾驶员注意力检测技术可大致分为直接监视驾驶员的技术和监视作为驾驶员结果的车辆行为的技术这两种。直接监视驾驶员的技术包括测量生理状况（例如脑电波^[8]，心电信号和肌电信号等^[9]）以及监视眼睑的运动，头部运动和面部状况（例如经常打哈欠）。生理信号的测量在驾驶员疲劳检测方面可达到最佳精度，但由于需要将电极连接到驾驶员，因此会带来许多干扰，这通常被证明是驾驶员的烦恼，并且在大多数情况下是不切实际的。但是，在评估其他技术时，生理测量值可作为良好的基准。通过使用摄像机来监视眼睑的运动和凝视是用于驾驶员注意力检测的成熟技术，因此，致力于开发这些技术的研究工作很多。使用摄像机监视驾驶员也是非侵入性的（与 EEG 测量相反），因为不需要驾驶员进行任何交互。然而，这些技术确实倾向于对诸如驾驶员的亮度或外观（例如，驾驶员是否戴着眼镜）的外部因素敏感。监视车辆行为的技术考虑了车辆速度，横向位置，转弯角度和移动路线。这些技术也完全不会对驱

动程序造成干扰，使其适用于实际系统。然而，这些技术也具有局限性，例如受到驾驶员经验，驾驶条件和车辆类型的限制。

1.2.1 基于驾驶员生理特征的检测方法

当驾驶员处于疲劳状态时，他的注意力会明显不如轻松时集中，因此驾驶员是否疲劳是其注意力是否集中的重要标志。当人处于不同状态时，其生理信号会发生明显的变化。所以通过生理上或者医学上的方法来对驾驶员的疲劳状态进行检测研究就发展的比较早。基于驾驶员生理特征的检测方法需要驾驶员穿戴一些医疗监测设备，来获取诸如脑电信号、肌电信号、心电信号、眼电信号和驾驶员的呼吸和脉搏等信号来进行直接测量。实验表明，驾驶员在处于疲劳状态时，其脑电波、心率和脉搏等会发生明显的变化，通过检测这些变化，可以获取驾驶员的疲劳状态也就是获得驾驶员注意力分散的标志。

通过激发大脑内神经元而沿头皮记录的电活动被称为脑电图（EEG）。基于脑电波 EEG 的检测方法被称为驾驶员疲劳检测的金标准^[10]，人脑由四种脑电波，这四种脑电波在人处于清醒和疲劳状态下的出现的次数是不同的。MINDS 项目（微点检测系统^[11]）提出了一种使用电容式传感器阵列的头部位置传感器，旨在检测人的头部点头和微睡眠事件。在由^[12]开发的一种基于 EEG 的算法中，使用了人工神经网络（ANN）将驾驶员的状态分类为警报或困倦。该研究的数据来自 17 个受试者，并由学习矢量量化（LVQ）神经网络用于分类过程。

基于心电信号疲劳检测方法是指分心驾驶员的心电信号来对驾驶员的状态进行判别的方法。当驾驶员处于疲劳状态下时，其心率变异性 HRV 会产生明显变化，曾超等人^[13]将 HRV 作为驾驶员疲劳检测重要标准。王琳虹等人^[14]通过改变驾驶员的驾驶时长，发现驾驶员的 HRV 与道路景观的色彩均值有着密切联系，驾驶员的 HRV 会随着疲劳状态改变而改变。

基于肌电信号疲劳检测方法是指分心驾驶员的肌电信号来对驾驶员的状态进行判别的方法。合肥工业大学的扈静^[15]将人的心电信号、肌电信号作为评价驾驶员注意力检测的指标，开发出了一套基于 LVQ 神经网络的表征驾驶员疲劳状态的评价方法。东北农业大学的祝荣欣^[16]采集了联合收割机驾驶员的心电信号和肌电信号，从中选取出最能够表示驾驶员疲劳状态的指标，采用信息融合的方法构建出了疲劳等价评价模型。

但是，对于驾驶员来说，使用 EEG 或 ECG 等传感器都不是很实用，也不舒服。尽管生物医学传感器得到了发展，但先前提到的所有方法都具有某种程度上的侵入性和不舒适性，虽然它们已经被证明能够在受控环境中很好地工作，但是在实际驾驶条件下并不能进行大规模推广。

1.2.2 基于机动车行为特征的检测方法

在驾驶员注意力集中状态的检测方法中，有一种方法是检测机动车的行为特征。研究人员从汽车方向盘的旋转角度、车辆的横向位置和车速的变化中可

以看出驾驶员的注意力状态。与基于驾驶员生理信号的检测方法相反，这类技术不具有侵入性，驾驶员不用佩戴测量仪，也不会影响正常驾驶，但是，这类技术大多会受到车辆的类型、道路环境条件和驾驶员的驾驶经验等的限制，因此，与基于驾驶员生理特征的检测方法相比，这类检测方法的准确性可能较低。

蔡素贤、杜超坎等人^[17]使用 CAN 总线采集了车辆运行状态数据，通过采用提取的 18 项与驾驶行为相关的特征开发出了一套基于随机森林算法的驾驶员疲劳检测算法；Friedrichs.F 等人^[18]通过对驾驶员进行实车实验提取了用于判别驾驶员注意力状态的 11 个指标，这些指标来自与车辆的横向位置、方向盘转角等，在这些指标的基础上使用 BP 神经网络构建起驾驶员疲劳检测模型；上海交通大学的李伟等人^[19]对 10 名驾驶员的车辆方向盘转角以及车辆的横向位置的方差做离散化和归一化的处理，将其作为疲劳驾驶的评价指标，从而构建了基于神经网络算法的疲劳识别模型；Sandberg D 和 Wahde M^[20]在汽车驾驶模拟器上进行模拟驾驶，并对车辆的横向位置、方向盘转角、速度等信息进行采集，并将这些信息输入到神经网络中开发出了驾驶员的疲劳状态检测算法；江苏大学的刘军等人^[21]使用角度传感器采集汽车方向盘的转角数据，根据方向盘转角和静止百分比构建驾驶员注意力检测模型；基于智能手机自带的加速度传感器、陀螺仪、GPS 等传感器，Saurabh Daptardar 等人^[22]获得了汽车的加速度、速度和方向等参数，利用这些参数判断是否存在驾驶危险。

1.2.3 基于驾驶员面部特征分析的检测方法

一个人的注意力集中程度可以通过他的眼睑运动来确定。为了分析人的眼睑运动，必须进行眼睛检测和跟踪，而眼睛得检测和跟踪还可以用作提取其他视觉特征的基础，这些特征可以指示驾驶员注意力集中的状态，因此在下面进行介绍。

文章^[23]提出的计算机视觉系统结合了眼睑运动，头部运动和面部表情来检测驾驶员注意力集中状态。疲劳监测系统和注意力分散检测系统已经在市场上出售，适用于工作场所环境和车辆（或铁路）驾驶员。它们大多数基于图像捕获，面部识别和用于疲劳检测的复杂图像处理。在面部和眼睛的检测和跟踪中所做的工作可以进一步分为两类：基于被动外观的方法和基于主动红外（IR）的方法^[24]。这些检测系统将人工神经网络或支持向量机来作为面部检测算法。在检测到面部的上部区域之后，必须使用眼睛检测算法来检测眨眼率，眼睛运动和闭眼百分比（PERCLOS）。在实际使用中，PERCLOS 方法已被证明在驾驶模拟器上表现良好^[25]，PERCLOS 是指测量驾驶员的眼睑闭合程度大于等于 80% 占总时间的比例。很多文章中作者为了实现在弱光条件下的测量，使用了主动红外方法，借助于指向驾驶员面部的红外聚光灯。Brandt 等人就是一个例子，他们在文章^[26]中设计的系统由使用两个网络摄像头，一个红外聚光灯，一个带有 28 个红外发光二极管（LED）的网格和一个布置了图像处理算法的个人

计算机组成。格雷斯等人^[27]在文中使用了以下方法来检测驾驶员的瞳孔：使用不同波长的两个不同的红外照明源以，其中 850nm 的 IR 波长获得一个图像，而使用 950nm 的波长获得另一图像。除了图像中瞳孔的亮度外，这两个图像是相同的。通过计算图像之间的差异，可以获得眼睑，因此可以确定 PERCLOS。

也有一些使用不同机位摄像机的检测方法，在文章^[28]中，Chu 等人提取了驾驶员嘴的形状和位置用来检测驾驶员是否有可能打哈欠。文章^[29]中在开始阶段使用了两个摄像机，一个在驾驶员的左边，另一个在驾驶员的右边，并使用了一个与特定摄像机匹配的实时处理器来处理来自两个图像的特征，以确定每个特征的 3D 位置。即使驾驶员戴着眼镜，它也可以确定视线，睁眼和眨眼的频率。监视眼睛凝视的方法也可能是有用的，因为已经有实验证明疲劳的驾驶员会减少对外部视图的视线，更多的时间内会选择朝前看，从而减少了对仪器和后视镜的检查^[30]。但是，这类系统即使是非商业的也很昂贵，并且它们的成本太高也会导致大多数驾驶员不会去使用它们。

也有一些科研人员设计了一些使用摄像机和微控制器的可控成本原型用于驾驶员疲劳检测。Sugany 等人^[31]使用了 Raspberry Pi 微控制器和红外敏感相机。Palani 和 Kothandaraman 在实验室环境中使用常规的 CCD 相机和个人计算机组成了检测系统，但是该系统仍存在一些缺陷，因为它不适用于深色皮肤的驾驶员^[32]，如果驾驶员皮肤较深的话，系统无法进行分辨。

但是，由于担心隐私问题，一些用户并不愿意使用仪器对他们的脸部进行扫描。为了避免使用照相机和生物医学传感器，可以使用一些解决方案来检测驾驶员与分心驾驶相关的行为。Rogado 等人使用压力传感器来检测驾驶员的手和 ECG 电极是否在方向盘上^[33]。Sendra 等人在方向盘上安装了温度传感器来检测驾驶员的手是否放在方向盘上^[34]，Sendra 使用的另一种方法是使用光传感器来检测驾驶员头部位置的变化，因为尤其是在驾驶员开始点头时，传感器应该会吸收更多的光。

从上文的分析中可以看出，目前存在着多种检测驾驶员注意力状态的方法。其中，通过对驾驶员的脑电波进行脑电图测量，可以说是检测驾驶员注意力状态的最可靠技术。多年来，包括通过 EEG 测量监视人脑行为在内的直接监视驾驶员注意力行为的方法已成为公认的最佳检测方法。然而，由于需要将多个电极附着在受试者的头皮上，这使得基于脑电图的方法具有较高的侵入性并且严重限制了其作为实际疲劳检测方法的使用。迄今为止，研究人员们最大的努力主要集中在从驾驶员身上提取面部特征（尤其是从眼睛中提取），很多注意力检测系统都使用眼睑移动作为注意力集中的度量，在这类方法中通过照相机提取驾驶员的头部和面部特征是非侵入性的，使其成为用于现实系统的合适方法。

1.2.4 基于驾驶员动作的检测方法

目前也有着许多的基于驾驶员驾驶动作的注意力检测方法，这些方法主要是以卷积神经网络为工具，使用卷积神经网络对驾驶员的驾驶过程的图像中驾驶动作进行识别，判断驾驶员的注意力状态。

卷积神经网络在人体的姿态估计中的使用由来已久，目前被广泛应用在人体行为识别、虚拟现实等领域，相较于传统的姿态估计方法。卷积神经网络能够很好地提取图像或者视频中的特征，在很多场合有着不俗的表现，这些场合主要有：人体的关节被遮挡、关节活动过于灵活、人体目标尺寸变化、人体相互重叠等。目前常用于基于驾驶员动作的方法主要有以下四类：（1）基于手和脸检测相结合的算法、（2）基于多分类器的手部运动检测的算法、（3）基于前景提取的算法、（4）基于方向梯度直方图和支持向量机的算法。

（1）基于手和脸检测相结合的算法

Yehya Abouelnaga 和 Hesham M.Eraqi 等人^[35]在文章中首先使用检测器来检测驾驶员的手和脸。他们使用遗传算法来学习：“驾驶员的手”、“驾驶员的脸”、“驾驶员的整体图像”、“驾驶员的手+脸”四种图像的权重。积分后将图片输入到 Softmax 函数。但是，他们发现，在进行手部和脸部识别的时候，检测器有时会失效，无法检测到手和脸，这会导致分心驾驶行为检测的准确性的降低。同时，他们也在检测的过程中发现，手部所占的比例要高于脸部，因此，通过使用聚焦在“脸部”和“脸部+手部”上的图像来进行检测的准确性明显低于手部图像的准确性。Abouelnaga 等人^[36]还提出了一个由遗传加权卷积神经网络集合组成的系统，该系统在原始图像，皮肤分割图像，脸部图像，手部图像和“脸部+手部”图像上进行训练，效果提升并不明显，依旧无法解决因为脸部所占比例低于手部而带来的精确度降低的问题。

（2）基于多分类器的手部运动检测的算法

Ohn-bar 和 Martin^[37]在文章中提出了基于多分类器的手部运动检测的算法。他们将图像分割为三个区域：方向盘，挡位和控制面板（即收音机）。他们为每个区域开发一个分类器，并且在这些区域中检测驾驶员的手是否存在。来自这些区域的信息被传递到“动作分类器”，该“动作分类器”合并三个分类器的特征，检测驾驶员实际的动作（如调节收音机，操作方向盘等）。尽管如此，他们分析的手部动作的种类还是不够的，还有一些驾驶动作没有被分类到关注到：如（1）驾驶员两只手都在方向盘上，（2）驾驶员一只手在控制面板（即收音机）上，（3）驾驶员手部处于挂档状态。而且，只对这三个区域进行检测，检测结果容易受到干扰。于是，Ohn-bar 和 Trivedi 在后续的研究中又收集了一个新的 RGBD 数据集，他们通过研究该数据集中的驾驶员驾驶动作，将图像中驾驶员手部出现的情况分为以下几类：只有一只手，没有出现手，两只手，两只手+手机，两只手+地图，以及两只手+瓶子。

（3）基于前景提取的算法

Maitree Leekhaa 和 Mononito Goswami^[38] 等人提出了一种使用前景提取和卷积神经网络（CNN）的简单架构。他们使用 GrabCut 通过前景提取合并其他功能（例如驾驶员的姿势），可将驾驶员与嘈杂的背景区分开，提高模型的性能。而 GrabCut 是一种交互式工具，可在配置生产环境中所需的预处理模块时最大程度地减少用户的工作量。但是，由于光线和环境的影响，该模型具有较大的误差。实际的驾驶活动是 3D 透视图，而视频是 2D 的，实时检测的准确性较低。

（4）基于方向梯度直方图和支持向量机的算法

Bu 等人^[39]以驾驶员分心驾驶行为的特征为分类目标，提出了一种基于梯度直方图（HOG）和支持向量机（SVM）的分心驾驶行为的检测方法。他们首先获取图像中感兴趣的区域，其次对该部分进行增强，去噪和归一化处理，然后，提取图像 HOG 特征，然后使用交叉验证方法对 SVM 分类器中的参数进行优化，最后，对视频图像中驾驶员的不同行为进行分类和识别。

Zhao 等人^[40]收集了更为全面的分心驾驶数据集——东南大学（SEU）驾驶姿势数据集，其中包含进行以下四项活动的驾驶员侧视图图像：（1）安全驾驶，（2）操作操纵杆，（3）吃喝（4）在手机上聊天。他们使用 Contourlet 变换提取了图像特征，并评估了四个不同分类器的性能：线性感知器，K 最近邻算法（KNN），多层感知器（MLP）和随机森林算法（RF）。该技术的平均分类准确度分别为 85.09%，87.16%，37.21%和 90.63%。后来，他们结合了来自空间特征提取器的特征和定向梯度的金字塔直方图（PHOG），模型精度相对来说还是比较低。

1.3 技术路线

本文是基于机器视觉的驾驶员注意力检测系统设计。涉及到驾驶员注意力检测的软硬件的开发。

本文中在设计驾驶员注意力检测算法 MGMN 时，组合提取了驾驶员检测图像的多粒度特征和中间层特征。多粒度特征提取模块在 ResNet50 模型的基础上，在 Conv5_x 模块都形成三个分支共同提取多粒度特征；在进行多粒度特征提取的同时抽取图像的中间层特征输入到 Softmax 损失函数中去。

本文在进行驾驶员注意检测系统设计时，选取了“树莓派+远程服务器”的结构，在服务器中部署了 MGMN 算法，通过局域网进行 Socket 通信。进行了真实驾驶场景实验，获得了很高的准确精度和较好的实时性。

1.4 本文组织架构

本文是基于机器视觉的驾驶员注意力检测系统设计，重点对驾驶员注意力检测的算法精度、检测系统设计等方面进行研究。文章共分五个章节进行阐述，其组织架构如下：

第一章：绪论。本章介绍了基于机器视觉的驾驶员注意力检测系统设计的研究背景和意义、国内外研究现状。

第二章：深度学习算法 CNNs 综述。本章介绍了一些常用方法和研究工具。

第三章：驾驶员注意力检测算法设计。本章提出了一种驾驶员注意力检测算法，通过对驾驶员驾驶过程图像的多粒度特征和中间特征的提取，较高精度地获得了驾驶员的注意力集中状态。

第四章：本章介绍了实验配置和环境、实验所用到的公用数据集、自建数据集的构建和处理方法。通过在现有公开数据集和自建数据集上进行对比实验，验证算法的精度和特征提取能力

第五章：驾驶员注意力检测系统设计。本章设计了一种驾驶员注意力检测系统，通过树莓派和远程服务器连接，服务器上配置的检测算法对树莓派传输来的图片进行驾驶员的注意力状态的识别后再由树莓派的显示器显示输出验证了 MGMN 算法在真实驾驶场景中的实用性。

第六章：总结与展望。本章总结全文工作，分析目前工作中存在的优点与不足，并对未来可以发展的方向进行展望。

1.5 本章小结

本章首先介绍了驾驶员注意力检测的研究背景和意义，从基于生理特征的检测方法、基于机动车行为的检测方法、基于驾驶员面部特征的检测方法以及基于驾驶员动作的检测方法四个方面介绍了近年来的研究现状。最后介绍了本文各章节的组成架构。

第二章 卷积神经网络综述

本文是基于机器视觉的驾驶员注意力检测系统设计，需要对深度学习中的卷积神经网络进行学习，了解其基本原理和构成，也需要了解现有的驾驶员注意力检测算法、深度学习框架和图像处理技术。本章介绍了早期的人工神经网络模型和卷积神经网络原理，包括卷积神经网络的反向传播、卷积层、激活层、池化层、全连接层结构、常用深度学习框架 Pytorch 和计算机视觉库 Opencv 等。通过阅读本章内容，可以加深对本文后续内容的理解。

2.1 人工神经网络

2.1.1 人工神经元

人的神经系统负责传递有关周围环境和人本身的相关信息，它处理这些信息并会引起身体其它部位的反应，神经系统的基本组成是一种被称为神经元的专门细胞，神经元的数量巨大，每个神经元可以与多达 5000 个其他神经元连接^[41]。神经元的图片如图 2-1 所示，神经元包含主干、树突和轴突三个部分，树突是输入通道，轴突是输出通道，在每个轴突的末端，都有许多突触，与其他神经元的树突相连。神经元所传递的信号实质上是沿着它移动的电荷，突触有两个状态，分为兴奋和抑制和，处于兴奋状态时，信号将被放大，处于抑制状态时，信号将被减弱。突出通透性的水平是可以发生改变的，这就使得神经元对于新的环境具有适应性，如果突出的通透性保持一致被认为是神经元的记忆。

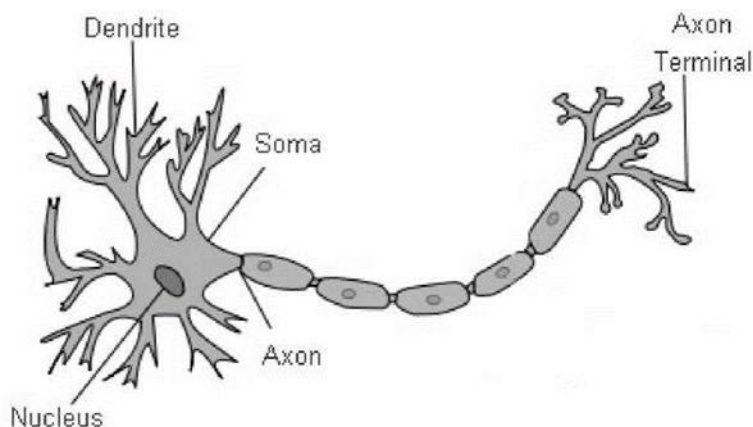


图 2-1 神经元(图片来源于维基百科词条神经元)

Figure 2-1. Natural neuron (Picture borrowed from Wikipedia page about neurons at <http://en.wikipedia.org/wiki/Neuron>)

人工神经元模型又称 M-P 模型^[42]（如图 2-2 所示）是人工神经网络模型的基本元素，它是仿照神经生理学上的神经元设计出来的。人工神经元（以下简

称神经元) 有 n 个用来模拟树突的真实输入 $x_1, x_2, \dots, x_n$, 输入 x_i 具有与之相对应的权重 w_i , 该权重模拟了突触通透性的, 当权重 w_i 为负值时表示了突触的抑制特性。神经元的内部电势 z 计算为输入的加权总和减去阈值 θ , 最后对神经元的内部电势施加一个激活函数 (Activation function) $f(z)$ 构成该神经元的输出。

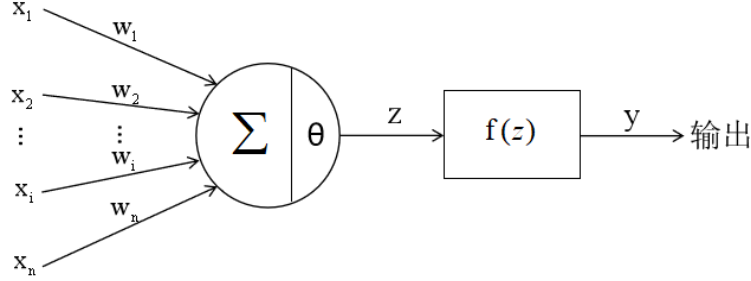


图 2-2 M-P 神经元模型

Figure 2-2. M-P neuron model

M-P 神经元模型用公式描述为:

$$y = f(z) = f\left(\sum_{i=1}^n w_i x_i - \theta\right) \quad (2-1)$$

输出值 y 对内部电势 z 的依赖是非线性的, 并且由激活函数 f 定义, 因此 $y=f(z)$ 。激活函数的种类又很多, 最常用的有以下三种:

(1) 阶跃函数 (如公式 2-2 所示)

$$f(x) = \begin{cases} 0, & x < 0; \\ 1, & x \geq 0. \end{cases} \quad (2-2)$$

由于阶跃函数的输出仅为“0”和“1”, 分别表示抑制和兴奋, 但是阶跃函数的表达能力不强, 极大地抑制了人工神经元的表达能力, 因此人们设计出了分段线性函数。

(2) 分段线性函数 (如公式 2-3 所示)

$$f(x) = \begin{cases} 0, & x < -\frac{1}{2}; \\ 0, -\frac{1}{2} \leq x < \frac{1}{2}; \\ 1, & x \geq \frac{1}{2}. \end{cases} \quad (2-3)$$

分段线性函数增强了神经元的表达能力, 但是其连接处并不光滑, 不能实现有效地过度, 因此, 人们又设计出了光滑且连续的 Sigmoid^[43]函数。

(3) Sigmoid 函数 (如公式 2-4 所示)

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2-4)$$

Sigmoid 函数图像如图 2-3 所示：

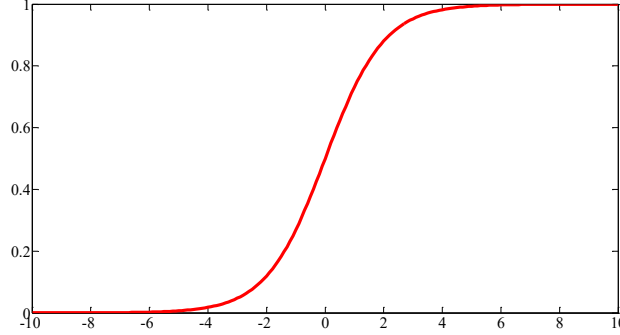


图 2-3 Sigmoid 函数
Figure 2-3. Sigmoid function

这三类激活函数在它们的输入值较小时，输出接近于“0”或者直接为“0”；当输入的值较大时，输出接近于“1”或者直接为“1”。其中 Sigmoid 函数处处可导，导数也非常简洁明显。因此，Sigmoid 函数在反向传播计算中的使用非常普遍，其公式为：

$$Sigmoid'(x) = Sigmoid(x) * (1 - Sigmoid(x)) \quad (2-5)$$

2.1.2 人工神经网络

通过一定的组合规则，将一个神经元的输出与若干个其他神经元的输入相连接，这就组成了人工神经网络（以下简称神经网络）人们将神经网络中的神经元分成三大类，分别为：输入神经元，隐藏神经元和输出神经元，分别对应着生物神经学中的受体、效应子和信号传播途径。输入神经元构成输入层，隐藏神经元构成了隐藏层，输出神经元构成输出层。隐藏层和输出层的神经元都是 M-P 模型^[44]。

整个神经网络的状态被定义为组成这个网络的所有神经元的状态，神经网络的配置被定义为所有相互连接的突触的权重，神经元的状态会随时间的变化而发生变化，因此，神经网络的状态也是时变的。人们确定了三种类型的动力学。分别是：结构动态、活动动态和自适应动态。结构动态是指网络拓扑的变化，活动动态是指网络状态的变化，而自适应动态是配置的变化。通过定义这三个动力学，就创建了一个神经网络模型^[45]。

（1）结构动态

神经网络的结构动态是指一个神经网络的体系结构和可能存在的改动。拓扑的更改通常与网络适应性一起发生，目的是在需要新的神经元的时候能够创建它们。神经网络的结构主要分为两类，一类是循环结构的神经网络（如图 2-4(a)所示），一类是非循环结构的神经网络（如图 2-4(b)所示）。

循环结构的神经网络是指一组神经元依次首尾相连，形成一个循环。循环结构是最简单的反馈情况，神经网络的输出连接到同一神经网络的输入。

非循环网络结构中没有循环的存在。这种结构的神经网络可以被划分成几层，一层在另外一层之上，这样每一层的连接都只通向更高层，这就形成了一个多层神经网络。这种神经网络的输入层仅包含输入神经元，输出层仅包含输出神经元，可以有一层或多层隐藏层。这些层从输入层开始向输出层那个依次按照神经元的个数进行编号，这样就可以简单地通过说明其每一层中神经元的数量来定义体系结构，这些层通常用短划线隔开（例如 3-4-3-2），如图 2-4(b)所示。

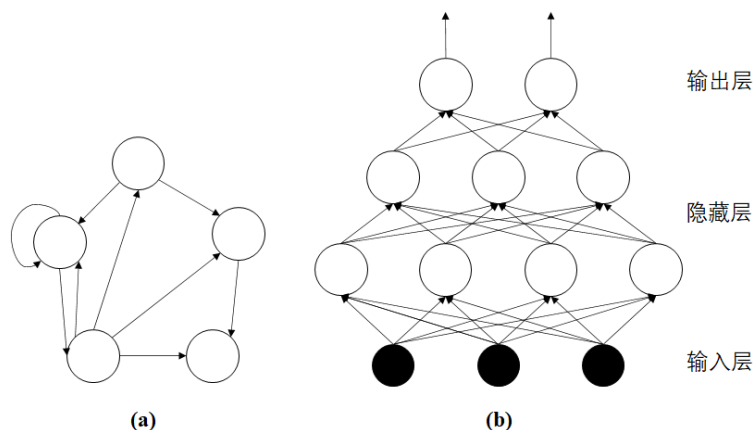


图 2-4 神经网络结构；(a) 循环结构；(b) 非循环结构

Figure 2-4. Neural network structure; (a) Cyclic structure; (b) Acyclic structure

(2) 活动动态

处于活动中的神经网络主要用于计算，主动动力学定义了网络的初始状态以及所网络的状态随时间变化的方式。

当网络刚刚进入活动状态时所有的神经元均处于初始状态，输入神经元开始接收输入信号，随后，网络可以开始进行计算。通常，神经网络的状态可以由时间的连续微分函数定义，但人们通常假设离散时间是由 0 开始的，并且状态仅在时间 1、2、3 等处发生变化。在每个步骤中，可以选择和更新一个神经元（顺序计算）或多个神经元（并行计算），以便根据输入来计算输出。同步更新意味着该过程是集中控制的，异步更新意味着神经元可以自主更新，网络（或网络功能）的输出就是所有输出神经元的输出。按照以上的假设，图 2-3(b) 中神经网络的离散计算可以如下进行：

在时间 0，为输入层中的输入神经元设置输入值，并计算它的输出；在时间 i ，第 i 层中的神经元被更新（默认输入层为第 0 层）；在每一层的每一个神经元中，执行以下步骤：应用激活函数，并将结果发送到输出。在输出层中完成计算后，整个网络的计算过程完成。

(3) 自适应动态

自适应动力学定义了神经网络的初始配置以及突触的权重随时间变化的方式。当网络进入自适应模式时，网络所有连接的权重将被设置为初始值（这些值可以是随机的）。与主动动力学类似，自适应模式下的神经网络的配置可以表示为时间的连续函数，但更多情况下使用离散方法。

活动模式下的神经网络功能取决于网络配置。自适应的目的是找到一种能够使处于活动模式的网络正确计算给定的网络功能的配置。换句话说，自适应模式表示网络的训练（或学习，编程）。如今，已经有了用于不同神经网络模型的各种训练算法。最常用的是著名的用于多层神经网络的反向传播算法。网络训练是一个复杂的非线性优化问题，即使对于一个很小的问题，其计算也会很耗时，训练所需的网络功能通常表示为一组成对的形式（网络输入，网络输出）。

监督训练是训练方法的一种，其中有人或机器能够提供关于神经网络预期输出的训练算法信息。在某些情况下，主管应该根据输出相对于输入的良好程度来打分。这称为分类训练。自组织是一种没有主管的适应方法。

2.1.3 多层感知机

多层感知机是一类特殊的神经网络。它具有神经网络的一般结构：输入层、隐含层和输出层。上文已经仔细讲过，这里不再赘述。特别地，只有一层隐藏层的感知机被称为单隐层感知机，有两层隐藏层的感知机被称为双隐层感知机。图 2-5 是一个典型的单隐层感知机模型。

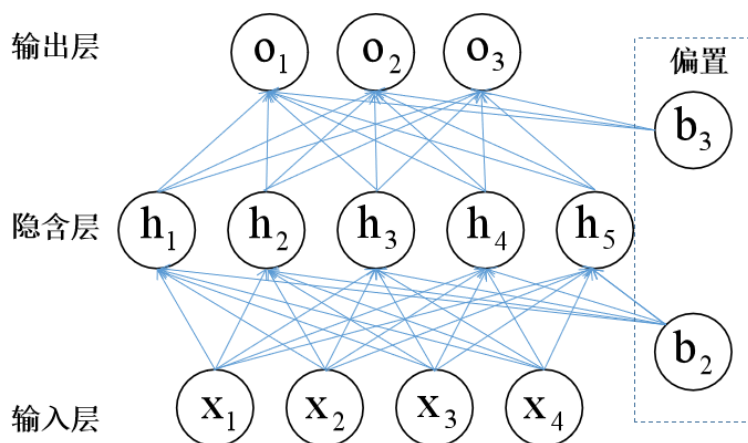


图 2-5 单隐层感知机

Figure 2-5. Single hidden layer perceptron

如上图所示，该单隐层感知机的输入层，隐藏层和输出层各包含 4, 5, 3 个神经元，输入的为四个变量 $[x_1, x_2, x_3, x_4]$ 。 x_i 、 h_i 、 o_i 分别表示输入层、隐藏层和输出层中的第 i 个神经元的输出信号。由此可以获得隐藏层和输出层神经元输出公式如下：

$$\begin{aligned} h_i &= f(w_2^{i1}x_1 + w_2^{i2}x_2 + w_2^{i3}x_3 + w_2^{i4}x_4 + b_2^i) \\ &= f\left(\sum_{j=1}^4 w_2^{ij}x_j + b_2^i\right) \end{aligned} \quad (2-6)$$

$$\begin{aligned} o_i &= f(w_3^{i1}h_1 + w_3^{i2}h_2 + w_3^{i3}h_3 + w_3^{i4}h_4 + w_3^{i5}h_5 + b_3^i) \\ &= f\left(\sum_{j=1}^5 w_3^{ij}h_j + b_3^i\right) \end{aligned} \quad (2-7)$$

式中的 w_l^{ij} 表示第 l 层的第 i 个神经元与前一层中第 j 个神经元之间的权重， b_l^i 表示第 l 层中第 i 个神经元的偏置， f 表示激活函数。上面两个公式可以变形形成以下形式：

$$\begin{aligned} \mathbf{H} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \end{bmatrix} &= f\left(\begin{bmatrix} w_2^{11} & w_2^{12} & w_2^{13} & w_2^{14} \\ w_2^{21} & w_2^{22} & w_2^{23} & w_2^{24} \\ w_2^{31} & w_2^{32} & w_2^{33} & w_2^{34} \\ w_2^{41} & w_2^{42} & w_2^{43} & w_2^{44} \\ w_2^{51} & w_2^{52} & w_2^{53} & w_2^{54} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b_2^1 \\ b_2^2 \\ b_2^3 \\ b_2^4 \\ b_2^5 \end{bmatrix}\right) \\ &= f(\mathbf{W}_2\mathbf{X} + \mathbf{B}_2) \end{aligned} \quad (2-8)$$

$$\begin{aligned} \mathbf{O} = \begin{bmatrix} o_1 \\ o_2 \\ o_3 \end{bmatrix} &= f\left(\begin{bmatrix} w_3^{11} & w_3^{12} & w_3^{13} & w_3^{14} & w_3^{15} \\ w_3^{21} & w_3^{22} & w_3^{23} & w_3^{24} & w_3^{25} \\ w_3^{31} & w_3^{32} & w_3^{33} & w_3^{34} & w_3^{35} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \end{bmatrix} + \begin{bmatrix} b_3^1 \\ b_3^2 \\ b_3^3 \\ b_3^4 \\ b_3^5 \end{bmatrix}\right) \\ &= f(\mathbf{W}_3\mathbf{H} + \mathbf{B}_3) \end{aligned} \quad (2-9)$$

2.1.4 反向传播算法

神经网络中从输入层到隐藏层再到输出层的运算过程被称为前向传播；反向传播是指神经网络中根据前向传播得到的误差来评价每一层网络中的权重对其的影响程度，通过使用链式规则和梯度下降对权重和偏置进行更新的过程。

下面将以图 2-4 中单隐层感知机为例，演算反向传播算法。假设神经网络的输出为 o_i ，真实标签为 o_i^t ， n 表示输出层中神经元的个数，则平方差损失函数 E 可以表示为：

$$E = \frac{1}{2} \sum_{i=1}^n (o_i^t - o_i)^2 \quad (2-10)$$

反向传播算法通过迭代的方式逐步修正神经元中的权重 w_l^{ij} 和偏置 b_l^i ，定义学习率为 η ，每次调整的值分别为：

$$\Delta w_l^{ij} = -\eta \frac{\partial E}{\partial w_l^{ij}} \quad (2-11)$$

$$\Delta b_l^i = -\eta \frac{\partial E}{\partial b_l^i} \quad (2-12)$$

调整后的值为：

$$v = v + \Delta v \quad (2-13)$$

分别对输出层中神经元的权重 w_3^{ij} 和偏置 b_3^i 进行求导：

$$\frac{\partial E}{\partial w_3^{ij}} = \frac{\partial E}{\partial o_i} \cdot \frac{\partial o_i}{\partial \beta_3^i} \cdot \frac{\partial \beta_3^i}{\partial w_3^{ij}} \quad (2-14)$$

$$\frac{\partial E}{\partial b_3^i} = \frac{\partial E}{\partial o_i} \cdot \frac{\partial o_i}{\partial \beta_3^i} \cdot \frac{\partial \beta_3^i}{\partial b_3^i} \quad (2-15)$$

其中 β_3^i 表示输出层中第 i 个神经元的输入，经过 Sigmoid 激活函数映射后得到输出 o_i ，即 $o_i = f(\beta_3^i)$ 。 β_3^i 可表示为：

$$\beta_3^i = \sum_{j=1}^m w_3^{ij} h_j + b_3^i \quad (2-16)$$

因此式(2-13)和式(2-15)可以进一步表示为：

$$\frac{\partial E}{\partial w_3^{ij}} = \frac{\partial E}{\partial o_i} \cdot o_i' \cdot h_j = (o_i - o_i^t) \cdot o_i \cdot (1 - o_i) \cdot h_j \quad (2-17)$$

$$\frac{\partial E}{\partial b_3^i} = \frac{\partial E}{\partial o_i} \cdot o_i' = (o_i - o_i^t) \cdot o_i \cdot (1 - o_i) \quad (2-18)$$

经过一次迭代， w_3^{ij} 和 b_3^i 的值更新为：

$$w_3^{ij} = w_3^{ij} - \eta \cdot \frac{\partial E}{\partial w_3^{ij}} \quad (2-19)$$

$$b_3^i = b_3^i - \eta \cdot \frac{\partial E}{\partial b_3^i} \quad (2-20)$$

同理，对于其他层的权重和偏置参数都可以使用链式法则求得导数，再根据式(2-13)进行更新。

2.2 卷积神经网络原理

卷积神经网络^[46]，也称为 CNN，是一种特殊的人工神经网络。它也是一种的图像识别方法，它能够识别二维图形，它能够使图像在被进行平移，缩放，倾斜或其他形式的变形的情况下特征保持不变，通过卷积运算、池化运算和反向传播等操作中学习数据特征。卷积神经网络主要包含：输入层、卷积层、激

活层、池化层、全连接层以及一个由卷积层和子采样层组成的特征抽取器。卷积神经网络被广泛应用于图像识别、自然语言处理、语音识别和其他领域。卷积神经网络结构如图 2-6 所示：

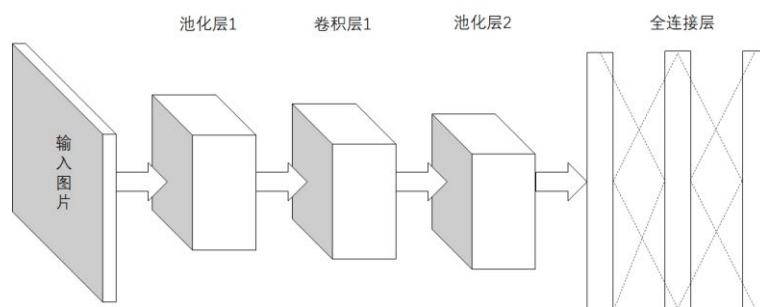


图 2-6 卷积神经网络结构

Figure 2-6. Structure of convolutional neural network

2.2.1 卷积层

卷积层 (Convolutional Layer) 可以说是卷积神经网络中最重要的部分，卷积神经网络的名称中的卷积两字也是由此得来。学习卷积层，首先需要理解什么是卷积。数学上的卷积是一种运算。

连续情况下它的定义为：

$$(f * g)(n) = \int_{-\infty}^{\infty} f(t) g(n - t) dt \quad (2-21)$$

$(f * g)(n)$ 就被称为 f 和 g 的卷积，连续情况下的卷积可以理解成存在两个函数 $f(x)$ 和 $g(y)$ ，且满足 $y = n - x$ ，它们的卷积是指对 $f(x) * g(y)$ 进行积分。

离散情况下它的定义为：

$$(f * g)(n) = \sum_{t=-\infty}^{\infty} f(t) g(n - t) \quad (2-22)$$

$(f * g)(n)$ 就被称为 f 和 g 的卷积，离散情况下的卷积可以理解成存在两个函数 $f(x)$ 和 $g(y)$ ，且满足 $y = n - x$ ，它们的卷积是指对 $f(x) * g(y)$ 进行求和。

在卷积神经网络中，卷积是指两个矩阵之间的操作。如图 2-7 所示，2-7(a) 是一个 5×5 的原始图像矩阵 A ，2-7(b) 是一个 3×3 的内核矩阵 B ，卷积就是将矩阵 A 和 B 对应的数值相乘相加之后再求和，卷积核 B 在矩阵 A 移动的距离为步长。具体卷积过程如图 2-8 所示。A 和 B 经过卷积操作最后得到的特征图如图 2-9 所示：

原始图像通过卷积层时，进行一系列的卷积操作，获得了最终的卷积特征图，该特征图中包含了大量原始图像的信息，通过逆运算也可以获得原始图像。

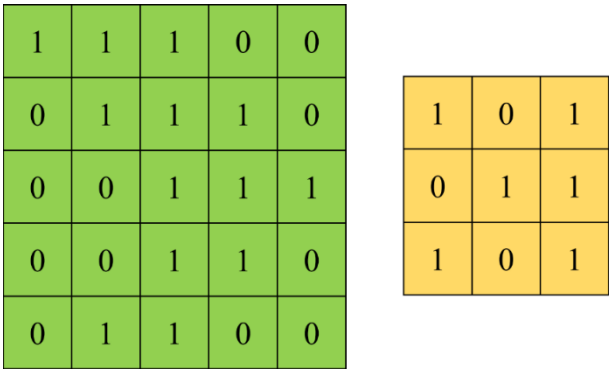


图 2-7 卷积操作矩阵；（a）原始图像矩阵 A；（b）内核矩阵 B
Figure 2-7. Convolution operation matrix; (a) Original image matrix A;
(b) Kernel matrix B

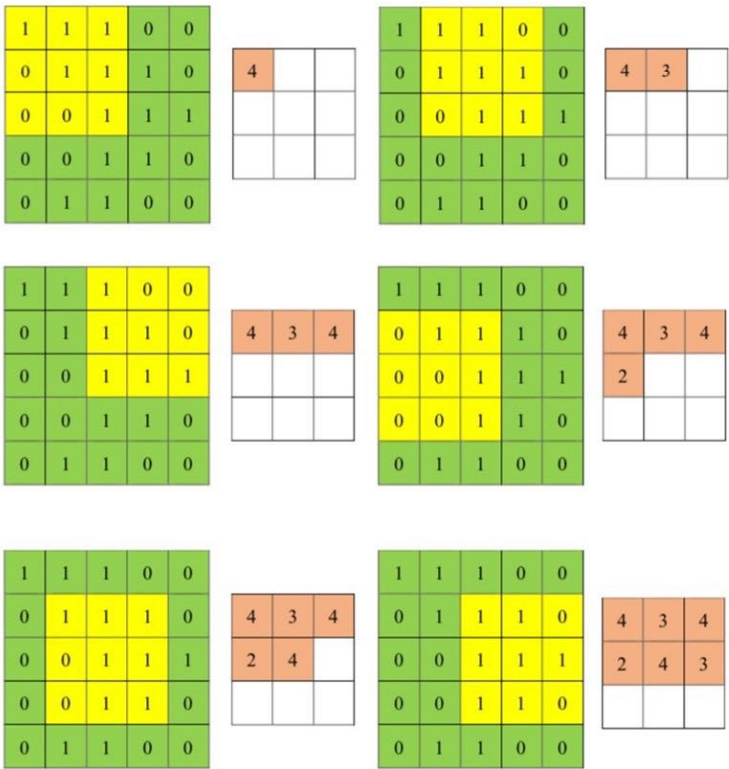


图 2-8 卷积过程
Figure 2-8. Convolution process

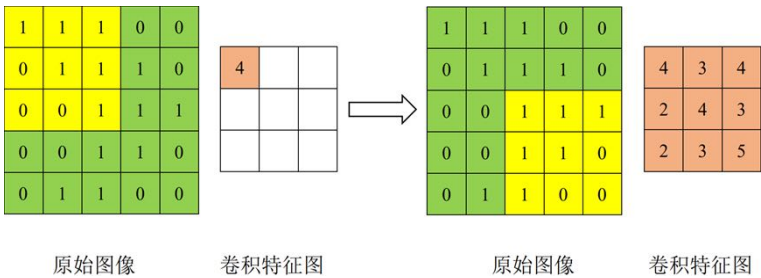


图 2-9 卷积结果
Figure 2-9. The result of convolution

从上图操作，本文可以发现，矩阵 A 中的每一小块都是被相同的卷积核 B 去扫的，因此每一次乘积求和的过程中的权重是相同的，这就是卷积层的一个特点，权值共享，通过权值共享减少了卷积层中参数的个数，使得运算更为方便简洁。由于每次运算时卷积核 B 只与原始图像 A 的一部分相连，这也使得进行运算的参数较少，加快了运算的进程。同时，同一张图片上距离较近的像素点之间的关联性较大，而距离较远的像素点之间的关联性较小，卷积核与图像的局部相结合，能够很好地提取图像的特征。

2.2.2 激活层

图像在卷积层进行卷积操作时，图像中的像素点和卷积核的乘积相加再求和，这显然是一个线性的过程，但是要怎么处理非线性的信息呢？神经网络中的激活层(Activation Layer)运行时，激活函数将激活该部分地神经元，将激活信息传递到下一层神经网络，因为激活函数中包含了非线性的信息，这就赋予了模型非线性信息的表达力。

目前使用较多的激活函数主要有两种，一种是处处可微的平滑非线性函数，如 Sigmoid、tanh 等等，还有一种是连续但不是处处可微的函数，如 Relu、Relu6 等等。

(1) Sigmoid 函数

Sigmoid 函数是传统神经网络中常用的激活函数之一，它的性质在前面已经讲过，此处不再赘述，Sigmoid 函数由于它的输出在[0,1]且是单调连续的函数，求导简单，适合做输出层。但是当 Sigmoid 函数的输入值过大或者过小时，其导数就会趋向于 0，梯度就会消失。

(2) tanh 函数

tanh 函数的如公式描述如下，其函数图像如图 2-10 所示

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (2-23)$$

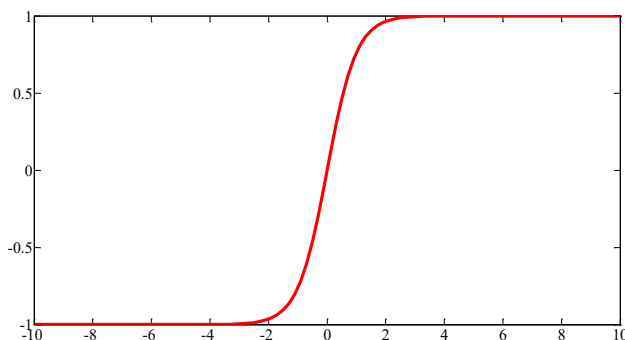


图 2-10 tanh 函数

Figure 2-10. Tanh function

(3) ReLU/PReLU 函数

ReLU 函数因为 AlexNet 算法在 2012 年的图像识别大赛中获得冠军而被大众知晓，ReLU 函数的表达式如下，其函数图像如图 2-11(a)所示。

$$f(x) = \begin{cases} x, & x \geq 0; \\ 0, & x < 0. \end{cases} \quad (2-24)$$

从图 2-11(a)中可以看出，当输入小于 0 时，ReLU 函数的输出值为 0，输入大于 0 时，输出值等于输入值，导数为 1，能够保证梯度不发生衰减，能够解决梯度消失的问题，并且提供了神经网络的稀疏表达能力。但是，在训练的过程中，部分输入会小于 0，会使得对应的权重无法更新，从而导致“神经元死亡”，神经元的输出持续为 0。鉴于此，He 等人^[47]提出了改进的 ReLU 函数——PReLU 函数，其公式如下，函数图像如图 2-11(b)所示。

$$f(x) = \begin{cases} x, & x \geq 0; \\ a_i x, & x < 0. \end{cases} \quad (2-25)$$

式中的 a_i 表示特征图的第 i 个通道的负向斜率，不为 0，这里的 a_i 不是一个固定的值，而是可学习的，它的出现使得 PReLU 函数不会发生因为输入过小而发生“神经元坏死”的情况。

在进行激活函数的选择的时候，当输入的数据的特征差异明显的时候，选择一般选择使用 tanh 函数，它会在不断循环的过程中将图像的特征扩大显示出来；当输入数据的特征差异不是很明显的时候，Sigmoid 函数的效果比较好。

但是当使用 tanh 和 Sigmoid 作为激活函数的情况下。需要对数据进行规范化的处理，不然可能会出现激活后的数据全部都在平坦区，隐藏层的输出趋同，丧失了原有的特征表达。但是在选用 ReLU 函数时，不需要进行数据规划化来避免丧失特征表达的情况。

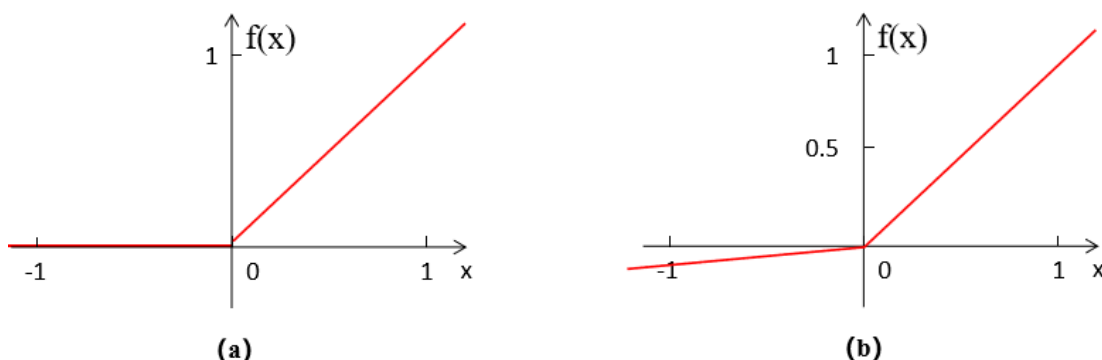


图 2-11 ReLU 和 PReLU 激活函数; (a) ReLU 函数; (b) PReLU 函数

Figure 2-11. Activation function of ReLU and PReLU, (a) ReLU function, (b) PReLU function

2.2.3 池化层

卷积神经网络中，通常会在卷积层之间周期性地加入池化层（Pooling Layer），池化操作分为最大池化和平均池化。

最大池化是指将输入的矩阵平均分成几个子矩阵，输出每个子矩阵的最大值，图 2-12(a)中所示，就是将一个 4×4 矩阵，将其平均分成四块，取每一部分的最大值，执行了这个最大池化操作获得一个 2×2 矩阵；最大池化是最常用的池化方式，对学习图像的纹理和边缘特征有着很大的帮助。

平均池化是指将输入的矩阵平均分成几个子矩阵，输出每个子矩阵的数值之和的平均值，图 2-12(b)中所示，就是将一个 4×4 矩阵，将其平均分成四块，取每一部分的和的平均值，执行了这个平均池化操作获得一个 2×2 矩阵。

池化操作的作用主要是通过采样来降低数据体的维数，从而减少运算时的参数，降低运算量，因为图片中也会带有一些冗余的信息，通过池化操作，可以实现对图像中多余的消息的过滤，这么做也可以避免过拟合的情况发生，提高了模型的泛化能力，提高了所提取特征的鲁棒性。

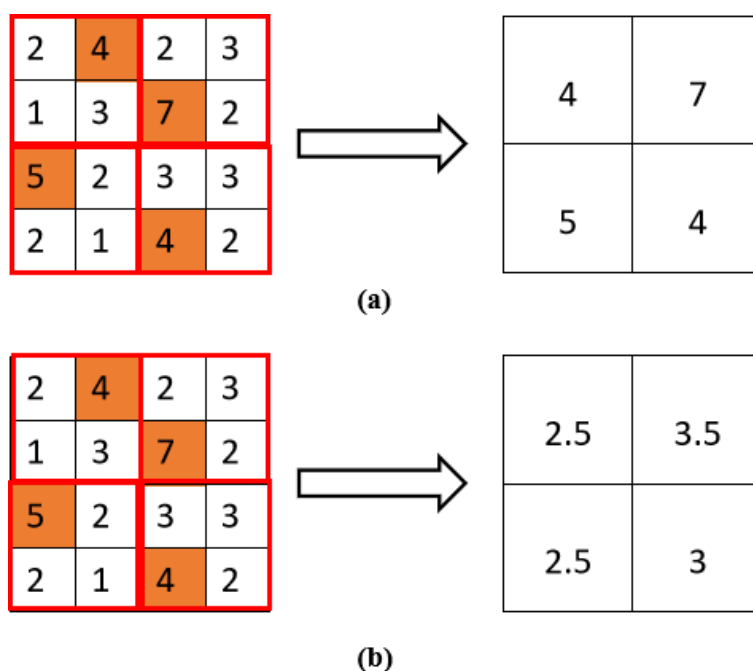


图 2-12 两种典型的池化方式；(a) 最大值池化；(b) 均值池化

Figure 2-12. Two typical pooling methods, (a) Max Pooling, (b) Average Pooling

2.2.4 全连接层

全连接层中(Full Connected Layer)的神经元都与上一层中所有的神经元相连接，如图 2-13 所示。卷积操作提取的是部分特征，这些局部特征输入到全连接层就可以被综合起来，全连接层通过融合全局信息，可以实现分类，起到了一个分类器的作用。因此全连接层的参数数量是非常大的，通常出现在卷积神经网络的最后几层且在神经网络中不宜加入太多。

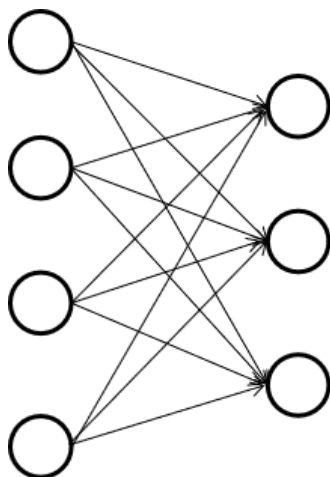


图 2-13 全连接层

Figure 2-13. Full Connected Layer

2.3 深度学习框架 Pytorch 和计算机视觉库 Opencv 介绍

2.3.1 深度学习框架 Pytorch

Pytorch 是由 Facebook 在 2017 年公布的深度学习框架，它是在 2002 年纽约大学公布的 Torch 基础上，使用 Python 语言进行封装和重构而成的^[48]。Pytorch 有着简洁高效、运行速度快、简单易用、社区活跃等特点。

(1) 简洁高效

Pytorch 设计时按照 `tensor`→`variable`(`autograd`)→`nn.Module` 这三个由低到高的抽象层次，分别代表着高维数组（张量）、自动求导（变量）和神经网络（层/模块），它们之间密切关联，可以同时进行操作。简洁的设计使得 Pytorch 的源码只有在使用时同样流行的 Tensorflow 的一成左右，更具体直观的设计理念使得 Pytorch 的源码极具可读性，很好理解。

(2) 运行速度快

Pytorch 的简介性不是在牺牲速度的基础上获得的，尽管源代码只有 Tensorflow 的十分之一，但是在众多评测中，Pytorch 的运行速度是高于 Tensorflow 的，代码的运行效率很大程度上与编码水平有关的，所以同样的算法，使用 Pytorch 完全有可能优于其他框架。

(3) 简单易用

基于 Lua 语言的 Torch 框架公布于 2002 年，虽然 Lua 语言因为及其小众导致 Torch 的使用并不是很流行，但是 Torch 的面对对象的接口设计非常的简单灵活易用，而 Pytorch 框架源于 Torch，其 API 和接口的设计都和 Torch 一脉相承，使用起来效率极高且设计上符合用户的思维模式。

(4) 社区活跃

Pytorch 为用户提供了完整而由浅入深的指导教程，它的开发者为用户们提供了供大家交流学习的论坛，同时 Pytorch 有着以上所述的种种优点，这也使得 Pytorch 受到了大家的追捧，它的社区更加活跃，活跃的社区使得用户更加积极使用 Pytorch，形成了这样的一个正向反馈，Pytorch 的社区活跃度和用户使用量都在不断上升。

Pytorch 自 2017 年公布以来，已经成为使用最广泛的深度学习框架之一，本文使用的深度学习方法中，就将使用 Pytorch0.4.0 结构。

2.3.2 计算机视觉库 Opencv

Opencv 全称 Open Source Computer Vision Library，是一个在 2000 年由 IEEE 公开的跨平台的开源计算机视觉库^[49]。

首先需要明确计算机视觉和图像处理这两个概念的区别。计算机视觉是指使用计算机技术来模拟和复制模拟人的视觉功能^[50]，主要包括：图像处理，模式识别、摄影测量等；而图像处理主要是指对图像进行处理^[51]，涉及到图像的转换，主要包括：增强、还原、去噪和分割等功能。

如上所述，计算机视觉是一门学科，被广泛应用于机器人、现代医学、工业自动化场景、安全应用、交通运输等各个领域，图像处理作为它的一个方向，也有着广泛的应用。

Opencv 作为一个模块化的计算机视觉库，它主要包含以下几个模块：

(1) 核心功能模块

核心功能模块 Core functionality (core) 定义了 opencv 最基本的数据结构，包括标量、点，范围等等。还定义了多维数组 Mat，可进行数组操作。

(2) 图像处理模块

图像处理模块 Image Processing (imgproc) 包含了各种的图像处理操作，主要包括：线性过滤，非线性过滤、几何图像转换，颜色空间转换，直方图，结构分析和形状描述，运动分析和对象跟踪，特征检测，目标检测等等。

(3) 相机校准和 3D 重建模块

相机校准和 3D 重建模块 Camera Calibration and 3D Reconstruction (calib3d) 主要包括了：基本的多视图几何算法，单个立体摄像机标定，物体姿态估计，立体相似性算法、三维信息重建等算法。

(4) 机器学习模块

机器学习模块 Machine Learning (ml) 包含了统计模型和分类算法，主要包括：统计模型，一般贝叶斯分类，K-近邻，支持向量机，决策树，梯度提升树，随机树，超随机树，神经网络等等算法。

Opencv 自 2002 年公布以来，已经成为一个使用极其广泛的计算机视觉库，本文使用的深度学习方法中，就将使用到 Opencv。

2.4 本章小结

本章主要介绍了人工神经网络的组成和卷积神经网络原理，现有的驾驶员注意力检测算法，以及深度学习框架 Pytorch 和计算机视觉库 Opencv。本章首先介绍了人工神经元的结构、人工神经网络的结构和运算过程，在多层感知机部分进行了反向传播演算；其次详细地进行了图像处理中卷积操作的演示、现有的激活函数地优缺点的分析、池化层和全连接层功能的介绍和演示；最后介绍了本文中将要用到的目前非常流行的深度学习框架 Pytorch 和计算机视觉库 Opencv。下文我将进行驾驶员注意力检测算法的设计，通过本章的阅读，会有利于对后续内容的理解。

第三章 驾驶员注意力检测算法设计

基于机器视觉的驾驶员注意力检测系统中，驾驶员注意力检测算法设计是其重要的组成部分，驾驶员注意力检测算法的检测准确率和快速性直接影响着检测的效率。图像识别技术可以通过使用卷积神经网络模型提取图片中的特征，给尚未进行学习的图片打上标签，对其进行分类。因此，本章将从驾驶员分心驾驶行为检测的角度出发，提取汽车驾驶员处于不同驾驶行为下不同部位的不同特征，设计出基于多粒度特征和中间层特征的卷积神经网络算法。

3.1 驾驶员注意力状态分类及其特征

3.1.1 驾驶员注意力状态分类

驾驶过程中驾驶员的注意力不集中是引起驾驶事故的重要因素之一，对其实注意力状态进行检测需要区分驾驶员具体的注意力状态。驾驶员的注意力状态分为集中与不集中两种状态，其中，引起驾驶员注意力不集中的状态主要有以下几种类型（如图 3-1 所示）：

- 状态 1：驾驶员开车时左手接打电话；
- 状态 2：驾驶员开车时右手接打电话；
- 状态 3：驾驶员开车时左手收发消息；
- 状态 4：驾驶员开车时右手收发消息；
- 状态 5：驾驶员开车过程中与乘客谈话；
- 状态 6：驾驶员开车过程中饮食；
- 状态 7：驾驶员操作汽车的收音机和控制屏幕；
- 状态 8：驾驶员开车过程中回头向后座取物；
- 状态 9：驾驶员在驾驶过程中整理仪容。

这些动作是现实中非常常见的分散驾驶员注意力的行为，都容易引起交通安全事故。



图 3-1 驾驶员注意力不集中行为
Figure 3-1. Driver inattention behaviors

3.1.2 驾驶员注意力各状态特征

不同的驾驶行为有着不同的行为特征，驾驶员注意力集中正常驾驶状态如图 3-2 所示，通过驾驶员注意力不集中的状态特征如下，如图 3-3 所示，

（1）当驾驶员处于状态 1 时，其左手会摆放在头部，右手会摆在方向盘上，头部竖立朝前；

（2）当驾驶员处于状态 2 时，其右手会摆放在头部；左手会摆在方向盘上，头部竖立朝前；

（3）当驾驶员处于状态 3 时，其左手会摆放在方向盘和头部之间，右手会摆在方向盘上，低头；

（4）当驾驶员处于状态 4 时，其右手会摆放在方向盘和头部之间，左手会摆在方向盘上，低头；

（5）当驾驶员处于状态 5 时，其头部会向右转（左舵车），同时双手会摆在方向盘上；

（6）当驾驶员处于状态 6 时，其会有一只手里拿着食物或水杯放在头部嘴附近，两外一只手摆在方向盘上；

（7）当驾驶员处于状态 7 时，其左手会摆在方向盘上，右手向前下方伸，头部会稍稍向前倾；

（8）当驾驶员处于状态 8 时，人会转身，右手向后伸去，左手摆放在方向盘上；

（9）当驾驶员处于状态 9 时，其会有一只手放在头部嘴附近，而另外一只手放在方向盘上。

(10) 当驾驶员处于安全驾驶情况下时，其双手一般情况下都会摆放在方向盘上，并且保持面部朝向前方。

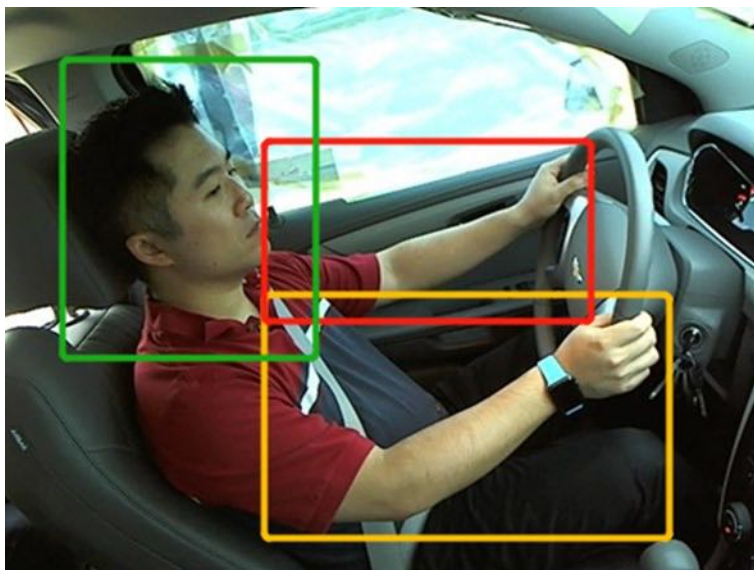


图 3-2 驾驶员处于注意力集中状态的特征

Figure 3-2. Characteristics of the driver's concentration

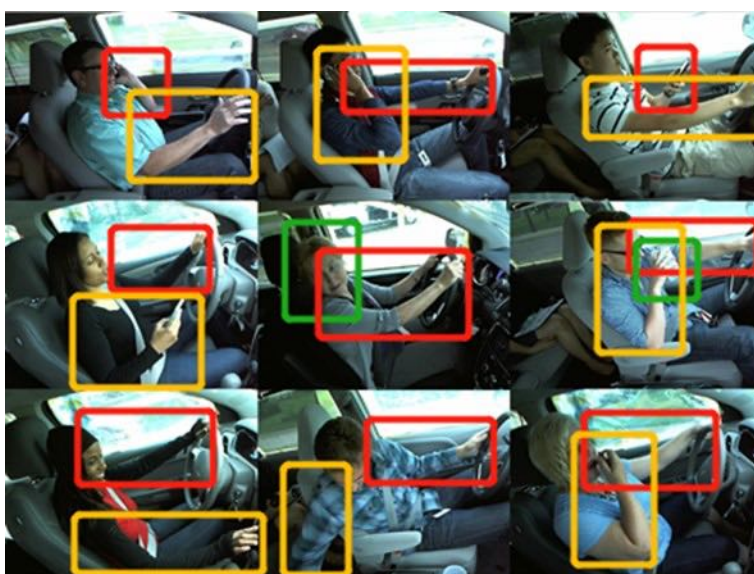


图 3-3 驾驶员注意力不集中行为的特征

Figure 3-3. Characteristics of driver inattention behavior

图 3-4 中本文使用 Grad-CAM 算法绘制了驾驶员右手收发消息时的特征热力图，该图显示了驾驶员在使用右手收发消息时的图形特征。本文可以从图中看出，算法检测到的动作特征是方向盘、左手握住方向盘的这个动作以及右手手里抓着一个手机。

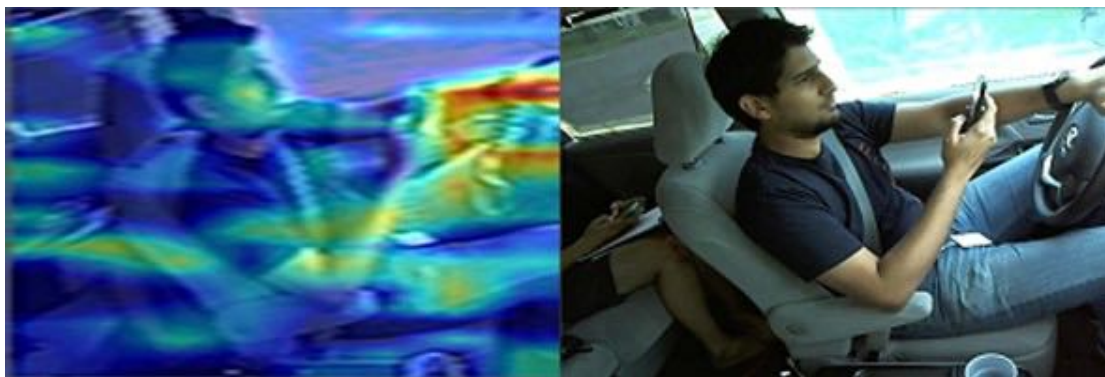


图 3-4 驾驶员使用右手收发消息时的特征热力图

Figure 3-4. The characteristic heat map when the driver uses his right hand to send and receive messages

3.2 MGMN 算法设计

针对 3.1 节中本文提到的驾驶员在不同驾驶状态下的头部与手部处在不同的特征，考虑到驾驶员在驾驶时头部和手部都处于图像的相对固定的位置，并且头部一般处于图片的上方，手部一般处于图片的中下方，如果可以针对驾驶图像的每一部分进行学习，提取每一部分特有的特征，再和整体图像的特征结合，实现驾驶图像特征的丰富提取。

基于上有的考虑，本节中本文设计了一个可以同时提取驾驶员驾驶图像中的多粒度特征与中间层特征的算法 MGMN（Multiple-Granularity-Middle Network），多粒度特征的提取使得驾驶员图像的每一部分的图像被充分使用，从每一部分去确定驾驶员的具体动作，中间层特征的提取使得在充分利用局部特征的基础上加上全局特征，丰富了提取到的特征，从整体上去确定驾驶员的动作。MGMN 算法使得图像中的整体与部分特征兼顾而结合，实现了驾驶图像特征的充分利用。MGMN 的算法结构如图 3-5 所示。

3.2.1 多粒度特征的提取

多粒度特征源于上海交通大学的 Guan-shuo Wang 等人首次在论文《Learning Discriminative Features with Multiple Granularities for Person Re-Identification》中提出的多粒度神经网络。

本文构造的 MGMN 算法中使用的多粒度特征提取模块以深度卷积残差网络 ResNet50 模型（表 3-1 ResNet50 网络结构）为基础，在它的 conv4_1 模块后划分出三个独立的分支，三个分支由上到下依次为：全局分支、二分支和三分支（如图 3-6 所示），而这三个分支与原始的 ResNet50 模型使用相似的体系结构^[52]。

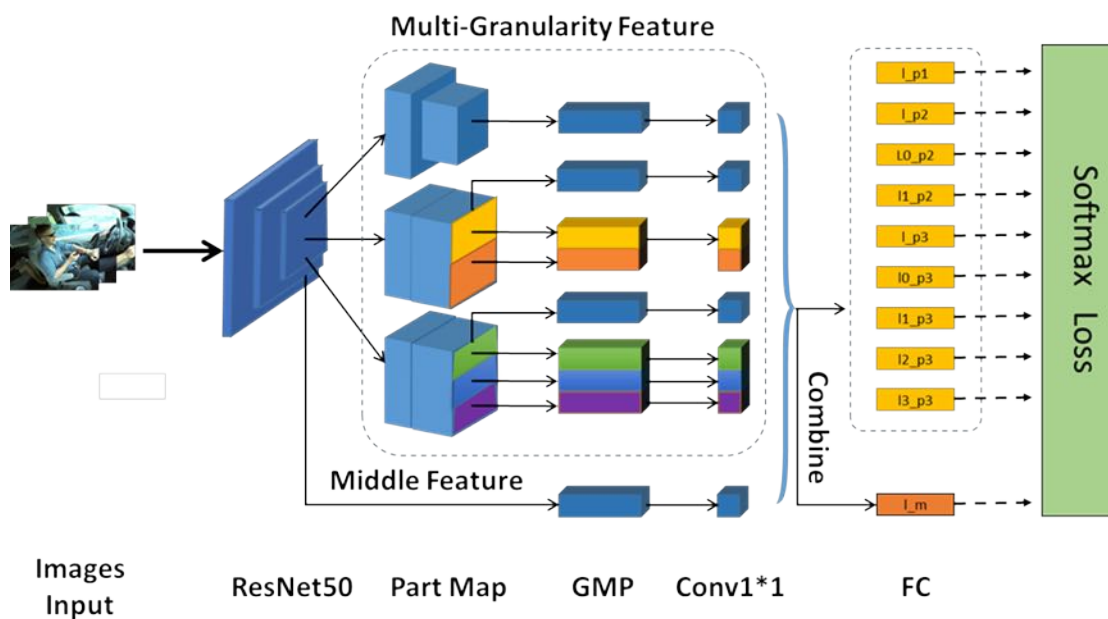


图 3- 5 MGMN 算法结构图

Figure 3-5. Structure diagram of MGMN algorithm

表 3-1 ResNet50 模型结构

Table 3-1. Structure of the ResNet50 model

Layer Name	Output Size	50-Layer
Conv1	112×112	7×7, 64, stride2
		3×3, max pool, stride2
Conv2_x	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4_x	Conv4_x	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	Average pool, 1000-d fc, Softmax
FLOPs		3.8×10^9

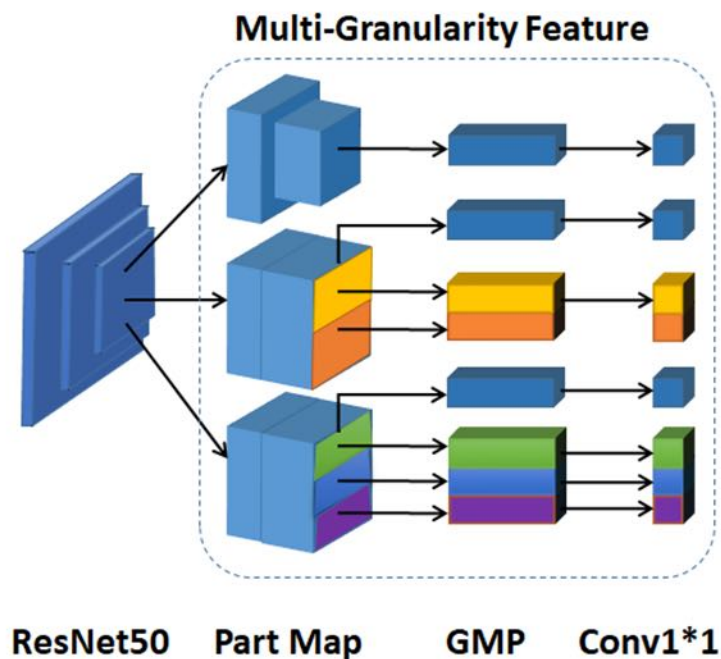


图 3-6 多粒度特征提取模块

Figure 3-6. Multi-granularity feature extraction module

全局分支学习的是不带有局部特征的全局信息。首先，它在 `conv5_1` 模块之后使用了带有 `stride-2` 卷积层的下采样操作，其次，在对输出的特征图进行了全局最大池化操作之后获得了 2048 维的特征图，随后，在该特征图使用归一化的 1×1 卷积和并使用激活函数 `ReLU` 后，实现了将特征像素减少到了只有 256 维。

二分支使用和全局分支类似的体系结构，但是它会在图像的输出特征图的水平方向上划分出两块，这两块将会被执行和全局分支一样的操作，分别提取图像中的上下两部分的信息。

三分支的结构和二分支完全相同，但是它是在图像的输出特征图的水平方向上划分成三块，这三块将会执行和全局分支一样的操作，分别提取图像中的上中下三部分的信息。对于驾驶员的驾驶动作图像，二分支和三分支具有很好的针对性，三分支结果示意图如图 3-7 所示，水平划分成三部分之后，驾驶员的头部和手部居于图像的第一第二层。

通过分成三个分支，提取全局信息、分成两部分和三部分后的全局特征相结合。



图 3-7 MGMN 算法的三支结果示意图

Figure 3-7. Schematic diagram of three-branch results of MGMN algorithm

3.2.2 中间层特征的提取

当驾驶员驾驶图像被输入到卷积神经网络中后，神经网络会对图像的各层特征进行提取，其中包括浅层特征和深层特征。浅层特征主要包括图片经过神经网络模型前级的卷积而后得到的特征图。这一部分特征图主要关注于图像中的纹理细节等特征；深层特征主要包括图片经过神经网络模型后级的卷积层而后得到的特征图信息，这一部分的信息在语境上更为抽象。

本文在构建 MGMN 算法时，在 ResNet50 的 Conv5_x 模块后抽取了图像的中间层特征，在对输出的特征图进行全局最大池化操作后，获得了 2048 维的特征图，随后，在该特征图使用归一化的 1×1 卷积和并使用激活函数 ReLU 后，实现了将特征像素减少到了只有 256 维。中间层特征提取模块的输出在和多粒度特征模块的输出合并后，共同输入到 Softmax 损失函数中去。中间层特征提取过程如图 3.8 所示。

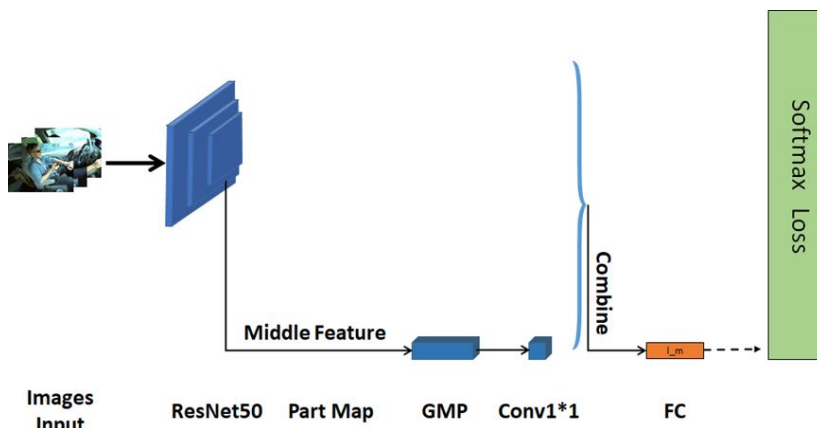


图 3-8 MGMN 算法对图片的中间层特征提取过程

Figure 3-8. Process of MGMN algorithm for image feature extraction in the middle layer

3.2.3 损失函数设计

损失函数是指将一些随机运算的结果映射成一个大于或等于 0 的实数，用这个实数的大小来进行结果的评价，以此作为依据，可以用来对神经网络的评价，从而优化神经网络的结构。

每一张图像中都含有一定量的信息，香农在信息论中提出了信息量的概念，香农认为信息量是用来消除随机不确定性的东西，该信息发生的可能性越大则信息量越小，信息发生的可能性越小，那么信息量就会比较大。而信息熵则是用来反映信息中的信息量的东西，用来表示所有信息量的期望，而数学中的期望是指事件所有可能发生的结果和它对应的概率乘积的总和，信息熵也可以被表示成如下公式：

$$H(x) = - \sum_{i=1}^n p(x_i) \log(p(x_i)), x(x_1, x_2, x_3, \dots, x_n) \quad (3-1)$$

相对熵，又称 KL 散度，是用来表示如一个随机变量 x 含有的两个独立的概率分布 $p(x)$ 和 $q(x)$ 之间的差异。KL 散度的公式如下：

$$D_{KL}(p||q) = \sum_{i=1}^n p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \quad (3-2)$$

在机器学习中常用 KL 散度来表示模型预测分布和真实分布之间的差异，如果 KL 散度越小，则表示 $p(x)$ 和 $q(x)$ 的分布更为接近，模型的效果也就越好，为了达到这个效果机器学习中会对 $q(x)$ 进行反复的训练使它的分布更加逼近 $p(x)$ 。

$$\begin{aligned} D_{KL}(p||q) &= \sum_{i=1}^n p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \\ &= \sum_{i=1}^n p(x_i) \log(p(x_i)) - \sum_{i=1}^n p(x_i) \log(q(x_i)) \\ &= -H(x) + \left[- \sum_{i=1}^n p(x_i) \log(q(x_i))\right] \\ &= \left[- \sum_{i=1}^n p(x_i) \log(q(x_i))\right] - H(x) \end{aligned} \quad (3-3)$$

在这里本文将 KL 散度的公式进行拆分，变成了前后两部分，后面的 $H(x)$ 就是信息熵，剩下的一块本文称之为交叉熵，即

$$\text{KL 散度} = \text{交叉熵} - \text{信息熵} \quad (3-4)$$

即

$$\text{交叉熵} = \text{KL 散度} + \text{信息熵} \quad (3-5)$$

在本文训练机器学习网络时，因为用来训练的数据几乎都是已有的，输入的数据和标签常常是已经确定的，其真实的概率分布 $p(x)$ 已经是确定的了，而信息熵只与真实的概率分布有关，因此，信息熵也是确定的了，KL 散度也就只和交叉熵有关，而交叉熵的公式相较于 KL 散度要简单得多，因此机器学习算法中常用交叉熵做为损失函数来判断预测结果的准确性。

驾驶员注意力检测需要对驾驶员的驾驶图像进行分类，给出其分类的确切结果，其本质是一个对驾驶员的驾驶图像的信息进行提取的过程，也就是一个多分类问题。本文使用交叉熵损失函数来解决这个问题。

Softmax^[53]是一个深度学习中的很重要的函数，它的作用是将人工神经元的输出结果映射到[0,1]空间中去，并且能够保重归一化后的和为1，这使得多分类问题的各个结果的概率之和刚好为1，并且保证输出的结果为最大概率的值。Softmax 函数的表达式如下：

$$S_i = \frac{e^i}{\sum_i e^i} \quad (3-6)$$

图 3-9 是一个 Softmax 的例子，数组[3,1,-3]在 Softmax 函数的作用下，映射成[0.88,0.12,0]，这三个值满足和为1，最后选取输出时，选择概率最大的那个值，也就是 0.88 对应的值。

在本文的 MGMN 算法中，本文将多粒度特征提取模块的输出结果单独输入到 Softmax 函数的同时，再将多粒度特征提取模块和中间层特征提取模块的输出结果进行组合后一起输入到 Softmax 函数。

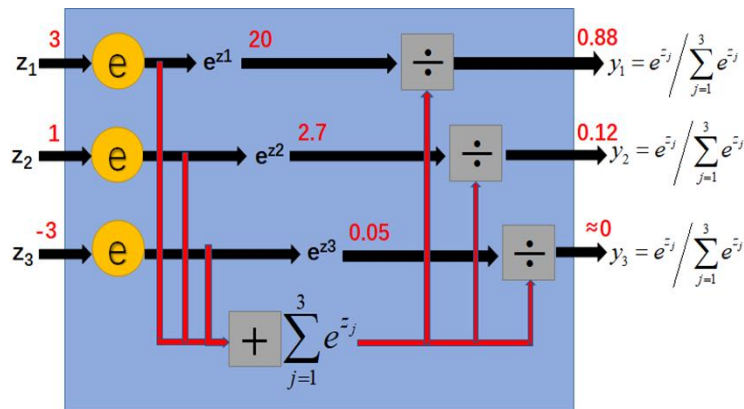


图 3-9 Softmax 函数运算过程

Figure 3-9. The calculation process of Softmax function

3.3 本章小结

在本章中，本文首先分析了驾驶员驾驶过程中注意力图像的特征，并针对这些特征提出了可以针对驾驶图像的每一部分进行学习，提取每一部分特有的特征，再和整体图像的特征结合，实现驾驶图像特征的丰富提取的想法，提出了基于多粒度特征和中间层特征的 MGMN 算法。随后对 MGMN 算法的每一部分的构成进行了详细的介绍和分析，以及选择交叉熵损失函数 Softmax 的过程进行了详细的推导。

第四章 MGMN 算法实验与结果对比分析

本文的第三章有针对性地根据驾驶员注意力检测场景的特征设计了基于多粒度特征和中间层特征的 MGMN 算法。为了验证算法设计初衷是否实现，本章设计了一系列算法对比实验，MGMN 算法将与现有的 CNNs 算法在本文的自建数据集“ZJUT-Districted Driver Detection”和现有的公开数据集：State Farm 公司的“Districted Driver Detection”数据集和“AUC-Districted-Driver-Detection”数据集上进行结果对比分析。

设计的实验主要包括算法精度对比试验和算法特征提取对比试验。使 MGMN 算法在训练速度和精度以及图像特征的与现有的算法进行比较，证明 MGMN 算法设计的合理性和有效性，证明该算法设计初衷的实现。

4.1 实验设置与环境

按照实验所要求，本文在实验室服务器上搭建了相关的深度学习软件环境，使用服务器进行训练，该服务器相关的软硬件配置如表 4-1 所示，该服务器配置已能满足本文实验的需求。

表 4-1 实验环境
Table 4-1. The experiment environment

参数名		参数值
处理器	型号	i7 7800X
	频率	3.5GHz
CPU	型号	GTX1080Ti
	容量	11G
	数量	2
	型号	DDR4 2666
内存	容量	16G
	数量	2
操作系统	型号	Ubuntu16.04
深度学习框架	类型	Pytorch0.4.0

4.2 实验所用数据集和处理方法

4.2.1 驾驶员注意力检测数据集

目前常用的驾驶员注意力检测数据集主要有 State Farm 公司公开“Districed-Driver-Detection”数据集和开罗美国大学公开“AUC-Districted-Driver-Detection”数据集^[35]。State Farm 公司的“Districed-Driver-Detection”数据集是 2016 年在分心驾驶检测比赛中公开的首个分心驾驶数据集，该数据集共包括 26 个驾驶员共 22424 张驾驶过程中的动作图片，其中训练集包含 17424 张，测试集包含 5000 张。这 22424 张图像被分成十类，依次分别是：c0 安全驾驶、c1 右手收发消息、c2 右手接打电话、c3 左手收发消息、c4 左手接打电话、c5 调整收音机、c6 饮食、c7 伸手向后座、c8 整理仪容、c9 和身边乘客讲话。如图 4-1 所示：



图 4-1 State Farm 的“Districed-Driver-Detection”数据集

Figure 4-1. State Farm's "Districed-Driver-Detection" dataset

开罗美国大学公开的数据集“AUC-Districted-Driver-Detection”是由 Yehya Abouelnaga 和 Hesham M.Eraqi 等人在论文中首次公开的数据集，该数据集。该数据集共包括 17310 张驾驶员驾驶过程中的动作图片，其中训练集 12978 张，测试集 4332 张，总体数量上略少于 State Farm 公司的“Districed-Driver-Detection”数据集。这 17310 张图像也被分成十类，依次分别是：c0 调整收音机、c1 饮食、c2 安全驾驶、c3 整理仪容、c4 伸手向后座、c5 左手接打电话、c6 和身边乘客讲话、c7 右手接打电话、c8 左手收发消息、c9 右手收发消息。该数据集如图 4-2 所示：



图 4-2 “AUC-Districted-Driver-Detection”数据集

Figure 4-2. “AUC-Districted-Driver-Detection” dataset

但是以上两个数据集都是欧洲人的针对驾驶员注意力检测的真实场景，为了开发更有效的针对亚洲人的驾驶员注意力检测算法，本文邀请到了 10 为中国志愿者，为他们拍摄了驾驶过程中的各类分心驾驶行为的照片，并制作了数据集，取名为“ZJUT- Districted-Driver-Detection”。该数据集包含 10 类驾驶行为共 5093 张照片，分为训练集和数据集两块，其中训练集有 4214 张，测试集有 879 张照片。“ZJUT- Districted-Driver-Detection”数据集的内容如图 4-3 所示。



图 4-3 “ZJUT- Districted-Driver-Detection”数据集

Figure 4-3. “ZJUT-Districted-Driver-Detection” dataset

4.2.2 驾驶员注意力检测数据集的处理

由于“AUC-Districted-Driver-Detection”数据集的训练集中的照片是由录制的驾驶员驾驶动作视频截取部分图片而成，因此混杂着大量的不属于对应分类的图片，因此本文在使用该数据集时对这个数据集进行了清洗，将数据集中的分类错误的数据集进行剔除，使得“AUC-Districted-Driver-Detection”数据集中剩下的都是分类正确的图片，清洗后的数据集本文称之为“AUC-Districted-Driver-Detection-Clean”数据集，在下面将要进行的算法性能验证中本文将使用“AUC-Districted-Driver-Detection-Clean”数据集进行实验。

本文自建的数据集取名为“ZJUT-Districted-Driver-Detection”，由于制作数据集时使用的摄像头像素过高，照片所占存储空间过大，在真实的使用场景中不一定会有这么高像素的摄像头，存储空间也是有限的，照片过大也会消耗算法读取和压缩照片的时间，因此本文对已获取的图像，编写了脚本，将其像素调制分辨率为 2304*1728，水平和垂直分辨率为 72dpi，这么做将大大减少程序运行的时间，也有利于将来的轻量化的 CNNs 算法的实现。

本文在分配训练集和测试集时，确保了训练集中的驾驶员没有出现在训练集中，其中 9 位志愿者的照片用于训练，1 位志愿者的照片用于测试，用于测试的志愿者的驾驶员了该数据集中的所有车型，这样做确保模型不会根据人物各自的特征以及车辆的特征进行分类，而是根据驾驶员的驾驶行为进行分类。这一点上，State Farm 公司的“Districted-Driver-Detection”数据集和“AUC-Districted-Driver-Detection”数据集没有做到，他们的训练集和测试集中的人物是一样的，这就存在着模型只依据人物的其他特征就可以进行分类的可能。

4.3 算法精度实验与分析

检验一个算法的有效性，需要对算法的精度进行实验，本节将对 MGMN 算法进行精度验证实验，并将实验结果与现有的算法进行对比，证明 MGMN 算法的有效性。

4.3.1 实验设置

针对这三个不同的数据集，本文设计了三组实验：第一组为 State Farm 公司的“Districted Driver detection”数据集实验，分别使用 MGMN 算法和其他算法在该数据集上进行训练和测试，并将运行精度进行对比分析。第二组为“AUC-Districted-Driver-Detection-Clean”数据集实验，分别使用 MGMN 算法和其他算法在该数据集上进行训练和测试，并将运行精度进行对比分析。第三组为自建数据集“ZJUT-Districted-Driver-Detection”实验，分别使用 MGMN 算法和其他算法在该数据集上进行训练和测试，并将运行精度进行对比分析。

4.3.2 实验结果对比与分析

(1) State Farm 公司的“Districted Driver detection”数据集实验

(i) MGMN 算法实验精度

本文使用 MGMN 算法在 State Farm 公司的“Districted Driver detection”数据集上进行了 9 组实验，其运行精度如表 4-2 所示：

表 4-2 MGMN 算法在 State Farm 公司的数据集上的实验结果

Table4-2. Experiment results of MGMN algorithm on State Farm's dataset

实验次数	精度
1	99.8%
2	99.72%
3	99.76%
4	99.8%
5	99.74%
6	99.74%
7	99.44%
8	99.72%
9	99.68%

如表 4-2 所示，MGMN 算法在 State Farm 公司的“Districted Driver detection”数据集上进行了 9 次实验，其中训练精度最高为 99.8%，精度最低为 99.44%，平均精度为 99.71%，MGMN 算法很好地提取了数据集中图像的特征

(ii) 其他算法精度对比实验

本文又将 ResNet50 模型、SqueezeNet 模型、VGG-16 模型、VGG-19 模型和 Maitree 网络模型在 State Farm 数据集上进行实验，算法的精度对比如图 4-4 所示，图表中我们给出了算法精度和训练时间，算法精度越高越好，时间越短越好。

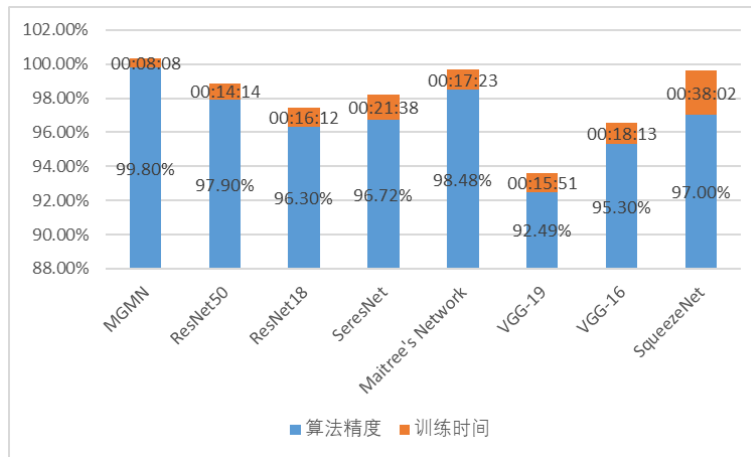


图 4-4 不同的卷积神经网络在 State Farm 公司的数据集上的结果的对比

Figure4-4: Comparison of the results of different convolutional neural networks on State Farm's dataset

由上图中本文可以看出，MGMN 算法相较于其他的几种算法在精度上有着较高的优势，有着 0.6%-3.28%的提升。

(2) “AUC-Districted-Driver-Detection-Clean”数据集

(i) MGMN 算法实验精度

本文使用 MGMN 算法在“AUC-Districted-Driver-Detection-Clean”数据集上进行了 8 组实验，其运行精度如表 4-3 所示：

表 4-3 MGMN 算法在“AUC-Districted-Driver-Detection-Clean”数据集上的实验结果
Table4-3. Experimental results of the MGMN algorithm on the "AUC-Districted-Driver-Detection-Clean" dataset

实验次数	学习率	精度
1	0.0001	98.70%
2	0.0002	98.59%
3	0.0003	98.50%
4	0.0004	98.20%
5	0.0005	97.60%
6	0.001	96.50%
7	0.001	96.80%
8	0.001	95.60%

如表 4-3 所示，MGMN 算法在“AUC-Districted-Driver-Detection-Clean”数据集上进行的 8 次实验，其中训练精度最高为 98.7%，精度最低为 95.6%，MGMN 算法很好地提取了数据集中图像的特征。

(ii) 其他算法精度对比实验

本文又将 ResNet50 模型、ResNet18 模型、SqueezeNet 模型、MoblieNet 模型、Xception 模型和 Seresnet 模型在“AUC-Districted-Driver-Detection-Clean”数据集上进行实验，算法的精度对比如图 4-5 所示：

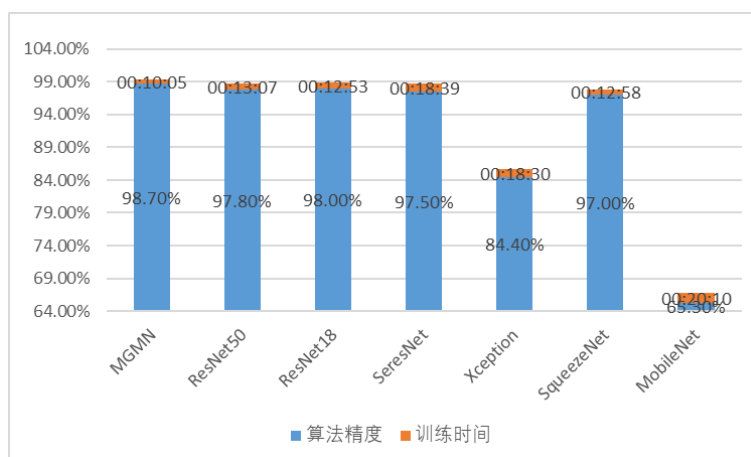


图 4-5 不同的卷积神经网络在“AUC-Districted-Driver-Detection-Clean”数据集上的结果的对比

Figure4-5. Comparison of the results of different convolutional neural networks on the "AUC-Districted-Driver-Detection-Clean" dataset

由上图中本文可以看出，MGMN 算法相较于其他的几种算法在精度上有着较高的优势，有着 0.7%-33.4%的提升。

(3) “ZJUT- Districted-Driver-Detection”数据集

(i) MGMN 算法实验精度

本文使用 MGMN 算法在 “ZJUT- Districted-Driver-Detection”数据集上进行了 8 组实验，其运行精度如表 4-4 所示：

表 4-4 MGMN 算法在 “ZJUT- Districted-Driver-Detection” 数据集上的实验结果
Table4-4. Results of MGMN algorithm on the "ZJUT- Districted-Driver-Detection" dataset

实验次数	学习率	精度
1	0.0001	96.40%
2	0.0001	97.80%
3	0.0001	96.60%
4	0.0001	96.80%
5	0.0002	96.82%
6	0.0003	95.10%
7	0.0004	95.70%
8	0.0005	96.32%

如表 4-4 所示，MGMN 算法在 “ZJUT- Districted-Driver-Detection”数据集上进行的 8 次实验。虽然数据集的量比较小，但是 MGMN 算法很好地提取了数据集中图像的特征，但是数据集的数量比较小易发生过拟合的情况，但是训练集和测试集选用了完全不同的驾驶员，较好地预防了这个问题。

(ii) 其他算法精度对比实验

接着本文又将 ResNet50 模型、ResNet18 模型、VGG-16 模型、VGG-19 模型和 Seresnet模型在 “ZJUT- Districted-Driver-Detection”数据集上进行实验，算法的精度对比如图 4-3 所示。

由图 4-6 中本文可以看出，MGMN 算法在相较于其他的几种算法在 “ZJUT- Districted-Driver-Detection”数据集的精度上有着很高的优势，有着 2.6%-5.21%的提升。

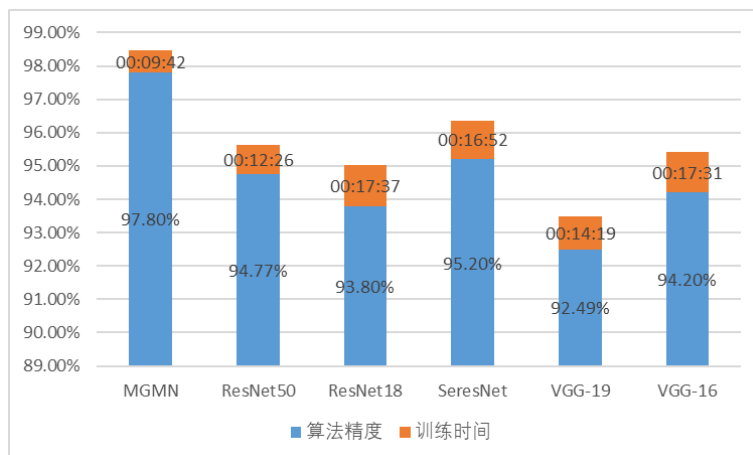


图 4-6 不同卷积神经网络在“ZJUT- Districted-Driver-Detection”数据集上的结果的对比
Figure4-6. Comparison of the results of different convolutional neural networks on the "ZJUT- Districted-Driver-Detection" dataset

4.4 算法特征提取能力实验与分析

算法的图像特征的抓取能力是评价一个深度学习算法的重要指标，如果算法不能很好地提取图像中的特征，那么算法的精确度是没有办法提高的。进行算法设计时不能仅仅把好不好用作为衡量深度学习算法模型的评价标准，更需要考虑到模型的可解释性问题。近年来，以剑桥大学的盖尔博士为代表的一群人就在研究着深度学习的可解释性问题，试着去揭开那层神秘的面纱。

为了更好地理解 MGMN 算法的特点，对算法做出更好的解释，本节中设计了 MGMN 算法的可解释性实验。可解释性实验是在算法的图像特征抓取实验的基础上，使用 Grad-CAM 算法绘制可视化热力图来查看 MGMN 算法的特征提取能力。Grad-CAM，全称 Gradient-weighted Class Activation Mapping，中文名梯度加权类激活映射，是一种通过生成视觉解释来解释 CNN 模型的方法，并不会给出图像中用于分类的具体细粒度信息。Grad-CAM 求取热力图的具体流程如下 [54]：

- (1) 求图像经过特征提取后最后一次卷积得到的所有特征图；
- (2) 利用反向传播求取每一张特征图的权重；
- (3) 每一张特征图乘以该特征图对应的权重得到带权重的特征图，在第三维求取平均值，使用 ReLU 函数进行激活，进行归一化处理；
- (4) 将处理后的特征图转换到原始图像的大小，便于与图像进行加权。

4.4.1 实验设置

在这个实验中，本文将使用上一节中实验精度最高的 ResNet50 算法和 MGMN 算法训练数据集获得的权重，来对这两个算法进行特征提取能力的对比实验。具体的，本文将共进行 3 组实验，每组实验都将 ResNet50 算法和 MGMN

算法在三个数据集上依次进行训练，利用卷积神经网络的前向传播获得最后一次卷积时得到的所有特征图，利用反向传播求的每一张特征图的权重，并进行加权求和，使用激活函数进行激活后获得的激活图与原图像进行叠加，得到的六组图像，可视化的图层是 p3.1.0.conv1 层。

4.4.2 实验结果与分析

两个算法在三个数据集上的热力图如下图所示，图像中颜色越鲜艳的地方表示算法对该特征的反映越强烈：

(1) State Farm 公司的“Districted Driver detection”数据集

MGMN 算法和 ResNet50 算法在 State Farm 公司的数据集上的特征热力图分别如图 4-7 和图 4-8 所示。

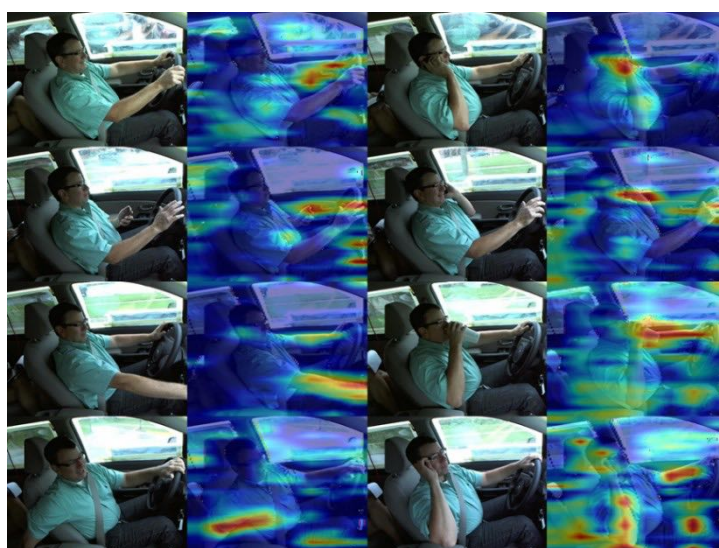


图 4-7 MGMN 算法在 State Farm 公司的数据集上的热力图

Figure 4-7. Heat map of MGMN algorithm on State Farm's dataset

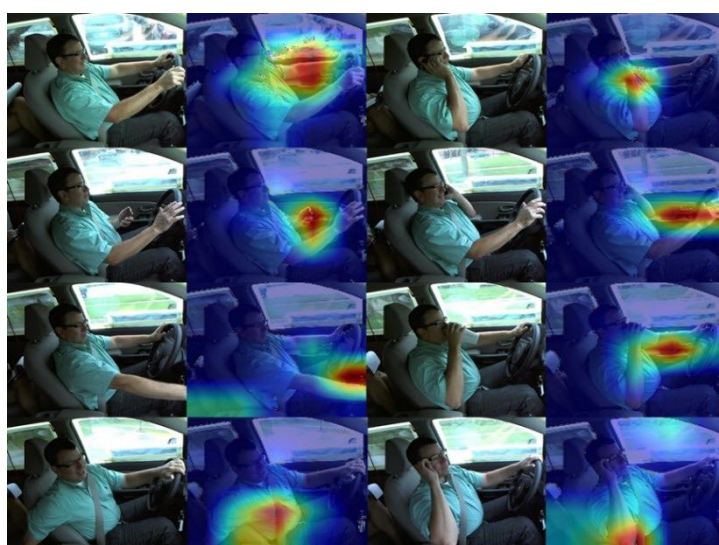


图 4-8 ResNet50 算法在 State Farm 公司的数据集上的热力图

Figure 4-8. Heat map of ResNet50 algorithm on State Farm's dataset

如上图所示，ResNet50 算法关注的点比较集中，如驾驶员右手打电话时，它关注的点仅为驾驶员的右臂，而 MGMN 算法则不一样，在 MGMN 算法的热力图中可以发现，驾驶员的右臂和右手掌部位的颜色特别鲜艳，这是算法关注的表现，同时还可以注意到，MGMN 算法也提取到了驾驶员的左臂放在方向盘上的这个特征，这个特征是 ResNet50 算法没有关注到的点，从其他分心驾驶行为的图片中本文也可以看出。

(2) “AUC-Districted-Driver-Detection-Clean”数据集

MGMN 算法和 ResNet50 算法在“AUC-Districted-Driver-Detection-Clean”数据集上的特征热力图如图 4-9 和 4-10 所示。

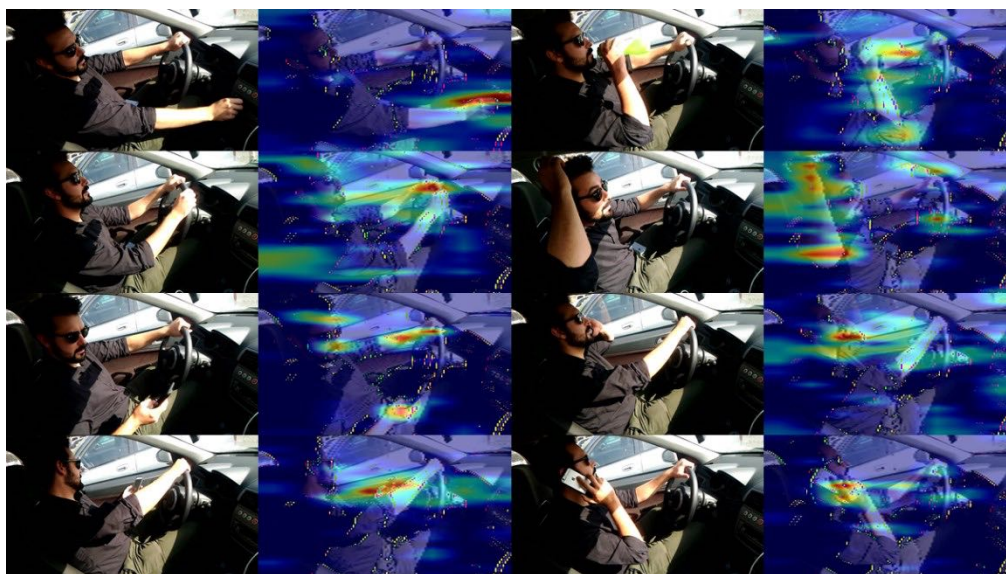


图 4-9 MGMN 算法在“AUC-Districted-Driver-Detection-Clean”数据集上的热力图

Figure 4-9. Heat map of MGMN algorithm on "AUC-Districted-Driver-Detection-Clean" dataset

ResNet50 算法关注的点比较集中，因此如果出现误判那就会直接导致错误的发生，如图 4-10 中的驾驶员右手收发消息时，算法并没有抓住右手这个特征，而是把关注的点放在了驾驶员的胸口处，图 4-9 中 MGMN 算法在进行驾驶员右手收发消息的判别时，提取了多处特征，驾驶员的右手有手机、左手在方向盘上以及提取到了人脸的特征，三种特征相结合，提高了特征提取的容错率，提高了算法的准确率。

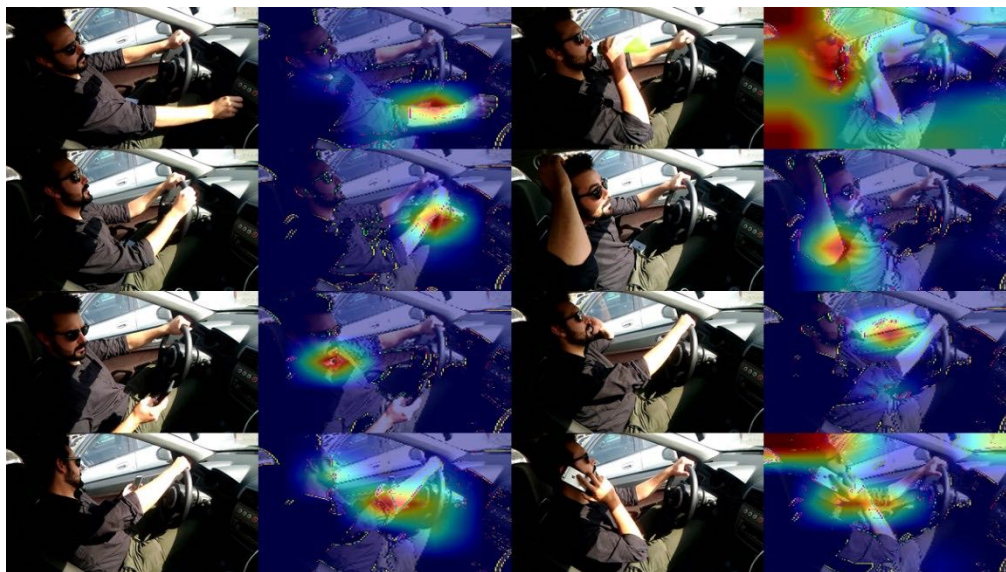


图 4-10 ResNet50 算法在“AUC-Districted-Driver-Detection-Clean”数据集上的热力图
Figure 4-10. Heat map of ResNet50 algorithm on "AUC-Districted-Driver-Detection-Clean" dataset

(3) “ZJUT- Districted-Driver-Detection” 数据集

MGMN 算法和 ResNet50 算法在 “ZJUT- Districted-Driver-Detection”数据集上的特征热力图如图 4-11 和 4-12 所示。

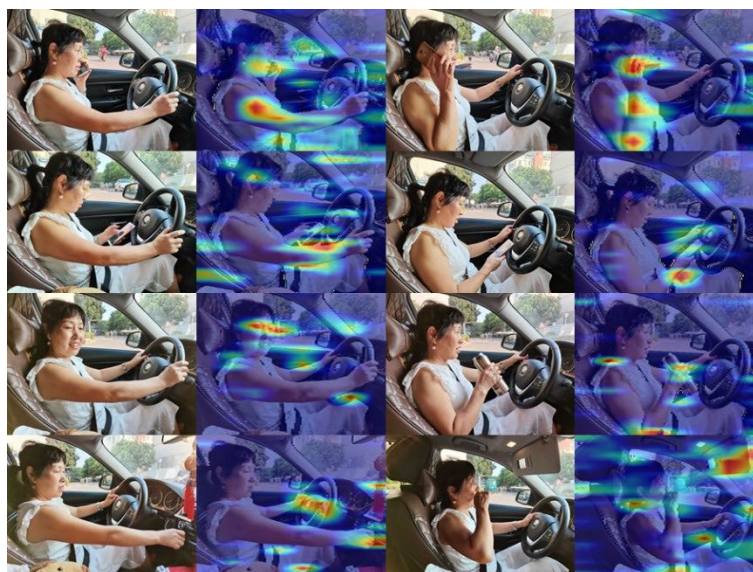


图 4-11 MGMN 算法在“ZJUT- Districted-Driver-Detection”数据集上的热力图
Figure 4-11. Heat map of MGMN algorithm on "ZJUT- Districted-Driver-Detection" dataset



图 4-12 ResNet50 算法在“ZJUT- Districted-Driver-Detection”数据集上的热力图
Figure 4-12. Heat map of ResNet50 algorithm on "ZJUT- Districted-Driver-Detection" dataset

MGMN 算法提取的特征非常丰富，但是它对背景等信息没有兴趣，只关注驾驶员的手部动作和其他的分心驾驶特有特征，因此 MGMN 算法能够准确地进行驾驶员分心驾驶动作的识别；精度也很高的 ResNet50 模型关注的点比较集中，但是它的特征提取不如 MGMN 算法丰富，因此它关注的点也被 MGMN 算法关注在内，精度上有一些差距，训练时间上 MGMN 算法也不存在比其他算法的明显弱势。

4.5 本章小节

本章对 MGMN 算法和其他现有的算法在精度上和可解释性上在三个数据集上进行了对比，实验结果表明，MGMN 算法在算法精度上相较于其他现有的算法都有着更高的精度，有着很强的特征提取能力。本文提出的自建数据集“ZJUT- Districted-Driver-Detection”在实验中也能准确地反应驾驶员分心驾驶行为的特征，是一个有效的分心驾驶数据集。

第五章 驾驶员注意力检测系统开发

本章本文将根据驾驶员注意力检测场景的特点设计出一套检测系统，并在系统中部署第三章开发的 MGMN 算法，将会在实际场景中使用验证系统的实用性。本章本文将会用到“树莓派+远程服务器”结构，在树莓派与服务器之间通过 Socket 方式实现远程通信，树莓派上连接的摄像头收集到驾驶员驾驶图像后使用传输到远程服务器，远程服务器对驾驶员的注意力状态进行分类，将检测结果传送到树莓派上并进行输出。

5.1 系统总设计

本节本文将针对驾驶员注意力检测环境的特点进行软硬件结构的分析，设计出实用可行的软硬件结构。

5.1.1 系统软硬件的功能需求分析

本文设计的基于多粒度特征和中间层特征的驾驶员注意力检测算法 MGMN，本质是一个卷积神经网络，用它来进行驾驶员注意力检测的实质是首先通过训练现有的分心驾驶行为的图像来提取驾驶员每一类注意力状态行为的特征，当新的驾驶图像来时，会根据新的图像中的特征对这张图片进行分类。因此，该系统的软硬件设计应满足如下功能：

(1) 驾驶图像的实时采集

实时采集驾驶员驾驶图像，高频率低延迟实时传输到注意力状态分析模块。

(2) 注意力状态判别

使用开发的 MGMN 算法对实时采集的驾驶员注意力状态进行判别。

(3) 注意力状态的实时输出

将判别得到的驾驶员注意力状态实时输出。

为了满足以上的三个主要功能，需要对系统进行软硬件分析：

(1) 独立的操作系统

本文需要进行算法计算功能，对 CPU 的运算能力有着一定的要求，如果条件允许，使用 GPU 进行计算的化运行效果会更好，为了方便文件的配置和管理，使用独立的操作系统很有必要

(2) 硬件尺寸不宜过大

驾驶员注意力检测是在汽车驾驶室内进行的，空间狭小，所选硬件的尺寸必须尽可能小，固定在驾驶室右侧车窗上部。

(3) 实时性

驾驶员注意力检测时一个对实时性要求较高的操作，如果图像采集及检测及结果不能实时，注意力检测也就没有意义。

(4) 算法易移植

针对驾驶员注意力检测，需要进行卷积神经网络的训练，需要配置环境，便于检测算法的植入。

(5) 系统易于扩展

本系统在本文中暂时只进行驾驶员驾驶图像的输入输出和驾驶注意力状态种类的判别，但是后期可以根据需求加入针对驾驶员分心驾驶状态的声光预警系统，因此，该硬件系统需要易于扩展，以方便后续工作的展开

本文设计的 MGMN 算法体积较大，现有的价格较低的移动终端并不具备运行该算法的能力，但是本文设计的初衷就是使用较低的成本实现高精度的驾驶员注意力检测，因此不能选用价格昂贵的设备。

价格低廉的终端不具备卷积神经网络的运算能力，本文考虑可以选择一块价格低廉的移动终端作为输入输出设备，通过局域网与远程服务器相连接，实现两者间信息的交互。

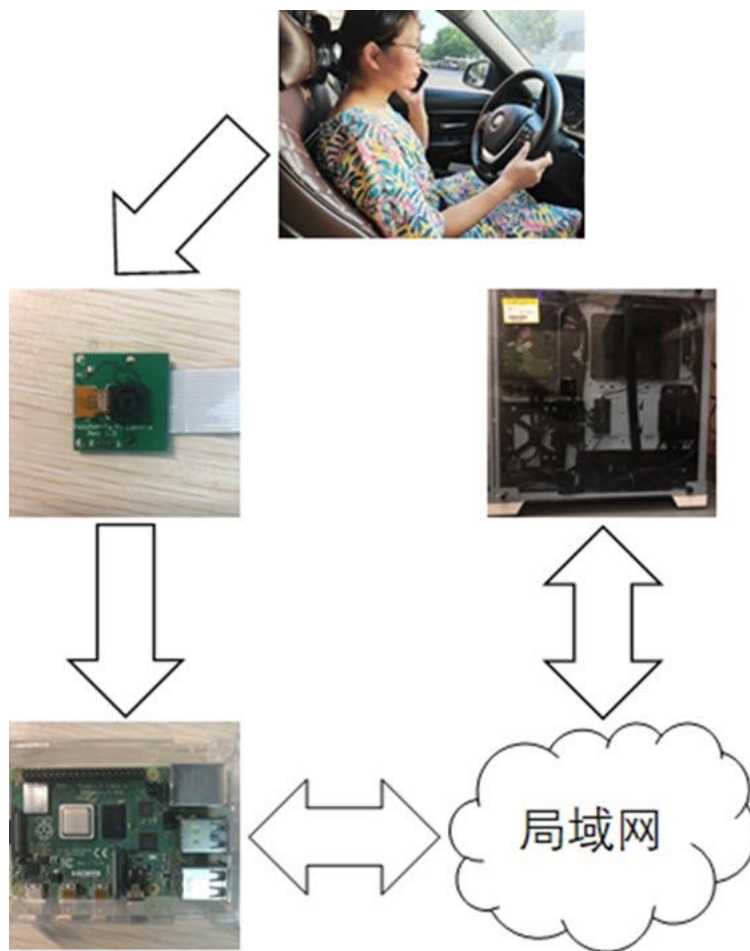


图 5-1 系统架构图

Figure5-1. System architecture diagram

基于上一节中的系统的软硬件设计需求的分析，论文决定选用“树莓派+远程服务器”的结构对驾驶员注意力检测系统进行开发，主要包括摄像头、树莓派 4B、远程服务器等等，使用局域网使得树莓派与远程服务器相连并且使用 Socket 方式使得信息在服务器和树莓派之间快速传递，由于局域网得高速性，使得传输获得了较低的延迟。检测系统的结构如图 5-1 所示。

5.1.2 硬件平台介绍

下面将详细介绍本章用于组建驾驶员注意力检测系统所使用到的各部分硬件及其配置结构。

(1) 树莓派 4B

树莓派在及其摄像头在本文的检测系统中属于终端的角色，用于实时采集驾驶员注意力检测图像以及信息的输入和输出，本文采用的是树莓派 4B。树莓派 4B 的详细参数如表 5-1 所示。

表 5-1 树莓派 4B 参数表
Table 5-1. Raspberry Pi 4B parameter table

名称	Raspberry Pi 4
Soc	Broadcom BCM2711
CPU	64-位 1.5GHz 四核处理器（28nm 工艺）
操作系统	Raspberry Pi OS(32-bit)
蓝牙	蓝牙 5.0
GPU	500MHz VideoCore VI
内存	4GB DDR4
影像输出	双 micro HDMI 端口
最大分辨率	4K 60Hz+1080p
USB 端口	2×USB3.0
有线网络	千兆以太网
无线网络	802.11ac(2.4/5GHz)
充电端口	USB Type-C
电力需求	3A,5V
多媒体	H.265(4Kp60 decode); OpenGL ES, 3.0 graphics
尺寸	85.60mm×56mm×21mm
重量	45g

由树莓派技术参数可知，使用其对 MGMN 算法进行训练与测试有一定的难度，即使算法能够顺利运行，运行时间可能还是会很长，因此并不适用做 MGMN 算法的处理器。

(2) 树莓派广角摄像头

本系统所用的摄像头是树莓派广角摄像头，其像素为 500 万像素。具体参数如表 5-2 所示。

表 5-2 摄像头参数表
Table 5-2. Camera parameter table

名称	树莓派广角摄像头
像素	500 万像素
感光芯片	OV5647
静态图片分辨率	2592*1944
尺寸	25mm×24mm×9mm
角度	65 度

(3) 服务器

该系统中所用到的服务器主要用于图像处理，因此需要较高的配置，本文采用了浙江工业大学网络化智能控制实验室中的服务器作为本文的服务器，在现实使用场景中，可以采用终端通过 5G 连接云服务器的方法去配置。本文采用的服务器的配置如表 5-3 所示：

表 5-3 服务器配置
Table 5-3. The experiment environment

参数名		参数值
处理器	型号	i7 7800X
	频率	3.5GHz
CPU	型号	GTX1080Ti
	容量	11G
	数量	2
内存	型号	DDR4 2666
	容量	16G
	数量	2
操作系统	型号	Ubuntu16.04
深度学习框架	类型	Pytorch0.4.0

5.1.3 Socket

同一网络中不同主机上的应用之间时常需要进行双向通信，而它们进行双向通信的端点称之为 **Socket**，也称套接字，主要用于完成两个程序之间的数据传输。

Socket 是从 **Unix** 中发展而来的，在 **Linux** 系统中可以使用一些函数对 **Socket** 进行操作，包括读写 **IO**、打开，关闭等等。在 **Linux** 中，**Socket()** 函数用于返回一个整型的 **Socket** 描述符，之后就可以建立连接并且进行数据传输了。**Socket** 通信的数据传输方式，有两种比较常用：

(1) SOCK_STREAM

SOCK_STREAM 是一种面向连接的数据传输方式。本文常见的 **http** 协议就是使用的 **SOCK_STREAM** 传输数据，因为它可以确保数据准确无误地在两台计算机之间进行传输，假如数据包损坏或者丢失，**SOCK_STREAM** 会将数据重新发送，因此传输效率相对较慢，但是可以保证传输数据的准确性。

(2) SOCK_DGRAM

SOCK_DGRAM 是一种无连接的数据传输方式，在 **SOCK_DGRAM** 的传输过程中，计算机只负责数据的传输，并不会对数据进行校验，假如数据在传输的过程中发生了丢失或者损坏，并不会重新发送数据，数据就会无法准确地到达计算机，因此数据传输效率高但是准确率相对较低。

在进行驾驶员注意力检测的过程中，由于采用的是局域网连接，网速较快，而所需传输图像较小，因此本文在系统之选用 **SOCK_STREAM** 是这种面向连接的数据传输方式，确保数据传输的准确率。

5.2 驾驶员注意力检测系统验证

5.2.1 实验环境与设置

为了进行驾驶员注意力检测系统验证，该节设计了驾驶员注意力的实时检测实验。

本文选取了一辆左舵汽车，在浙江工业大学的道路上进行了驾驶员注意力检测实验，驾驶员由两位志愿者担任，这辆汽车和两位驾驶员都没有在任何分心驾驶数据集中出现过。信息采集终端（树莓派+摄像头）被固定在右侧车窗上部进行驾驶动作的采集，树莓派通过连接浙江工业大学校园网和远程服务器连在了同一个局域网内。本文在树莓派上和服务器上都已配置好 **Opencv**，树莓派作为终端，使用 **Socket** 通信方式向服务器发送图像，服务器接收图像，在服务器上收到树莓派的摄像头获取的帧，使用 **MGMN** 算法进行注意力状态的判别并将结果传输到树莓派显示器上进行显示。服务器上获取的志愿者状态图像如图 5-2 所示。



图 5-2 两位志愿者

Figure5-2. Two volunteers

5.2.2 实验结果分析

本文在表 5-4 和表 5-5 中给出了两位志愿者的 10 类驾驶状态的 MGMN 算法的 10 类状态的检测结果和延迟信息。

由以下两表本文可以看出，部署了 MGMN 算法驾驶员注意力检测系统具有很高的准确性，但是在该实验中系统的延时相对较高，这是由于本文所使用的浙江工业大学无线校园网的带宽较窄引起的，上传下载速度最高也达不到 1M/s，数据传输速度较慢，因此具有一定的延迟。系统平均延迟为 1.66s，这在现实中尤其是高速上时不允许的，在高速公路上如果以 120km/h 行驶的话，延迟 1.66 秒就是已经是开过了 50 多米了，但是在低速行驶的情况下会好一点。

表 5-4 男性志愿者的实验结果

Table 5-4. Experimental results of male volunteers

志愿者	驾驶行为	输出结果	系统延迟/s
志愿者 1 (男性)	左手打电话	talk left	1.7
	右手打电话	talk right	1.6
	左手发消息	message left	1.9
	右手发消息	message right	1.7
	和乘客谈话	talk passenger	2.0
	饮食	drink&eat	1.6
	调收音机	adjust radio	1.6
	转身回头取物	turn back	1.5
	整理仪容	makeup	1.7
	安全驾驶	safe drive	1.3

表 5-5 女性志愿者的实验结果

Table 5-5. Experimental results of female volunteers

志愿者	驾驶行为	输出结果	系统延迟/s
志愿者 2 (女性)	左手打电话	talk left	1.6
	右手打电话	talk right	1.8
	左手发消息	message left	1.6
	右手发消息	message right	1.7
	和乘客谈话	talk passenger	1.9
	饮食	drink&eat	1.6
	调收音机	adjust radio	1.5
	转身回头取物	turn back	1.6
	整理仪容	makeup	1.7
	安全驾驶	safe drive	1.4

5.2.3 本章小节

本章首先针对驾驶员注意力检测场景的需求进行了软硬件需求分析，确定了基于“树莓派+远程服务器”结构的检测系统，并对选用的硬件配置和传输方式进行了详细介绍，对两个没有出现在任何一个分心驾驶检测数据集中的志愿者的驾驶图像对检测系统进行验证，系统取得了很高的精度和较好实时性，在大带宽的网络下可以满足现有的驾驶员注意力检测场景的需求，可以说是达到了本文预期的设计效果。

第六章 总结与展望

6.1 总结

随着我国工业的不断发展，汽车和驾驶员数量的不断增长，道路安全问题越来越严重，有很大一部分的驾驶事故是由驾驶员的注意力不集中引起的；随着技术的不断进步，在未来较长的一段时间内，人机共驾将会成为驾驶技术的主流，人机共驾系统需要检测驾驶员的状态是否可以满足独自控制车辆，如果不满足，系统就会强制介入，获得车辆的控制权以提高行车的安全性。其中，驾驶员的注意力集中状态就是一个重要的检测指标。驾驶员注意力不集中是指驾驶员在进行车辆的驾驶过程中注意力被一些行为所分散从而不集中。车辆驾驶过程中的注意力分散行为将会使得驾驶员的反应过程被拉长，从而使得其对车辆能否安全行驶的判断能力和突发情况的反应能力降低。同时，注意力分散行为会使得驾驶员对道路环境的观察减少，极易引起交通事故。

现有的检测驾驶员的注意力的检测方法主要有：基于生理特征的检测方法、基于车辆行驶数据的检测方法、基于驾驶员面部特征的检测方法和基于驾驶员动作的检测方法。基于生理特征的检测方法的准确度高，但是由于需要佩戴用于检测生理信号的仪器，这些仪器成本较高不易推广；基于车辆行驶数据的检测方法不具有侵入性，不会对驾驶员驾驶产生影响，但是这种方法会受到驾驶员的驾驶经验、车辆情况以及道路情况的影响；基于驾驶员面部特征的分析方法大多数是基于驾驶图像捕获，面部识别和用于疲劳检测的复杂图像处理。但是由于隐私问题，一些用户不愿意对其脸部进行扫描。

现有的基于机器视觉的驾驶员注意力检测算法主要有：基于手和脸检测相结合的算法、基于多分类器的手部动作检测的算法、基于前景提取的算法以及基于方向梯度直方图和支持向量机的算法。针对手部或者脸部的检测算法，在检测时，会出现检测失灵、人的各个部位所占比例不同带来的精度降低的问题；基于前景提取的检测算法易受光照和环境的影响，散发具有较大的误差；基于梯度直方图和支持向量机的算法总的来说算法精度还是比较低的。

本文针对驾驶员注意力检测场景，设计了基于机器视觉的驾驶员注意力检测系统，进行了驾驶员注意力检测系统的软硬件设计，设计的基于多粒度特征和中间层特征的 MGMN 算法在分心驾驶数据集中取得了更高的精确度，设计的硬件系统成本较低并且在真实场景实验中也有着很高的实用性。本文的主要工作有以下两个方面：

(1) 设计了基于多粒度特征和中间层特征的驾驶员注意力检测算法, 该算法具有精度高、特征提取丰富且准确的特点。考虑到驾驶员在驾驶时头部和手部都处于图像的相对固定的位置, 并且头部一般处于图片的上方, 手部一般处于图片的中下方, 如果可以针对驾驶图像的每一部分进行学习, 提取每一部分特有的特征, 本文在 ResNet50 模型中加入了多粒度特征提取部分, 该部分分成全局分支、二分支和三分支三部分, 共同提取驾驶图像的全局和局部特征。同时有抽取了驾驶图像的中间层特征, 中间层包含了丰富的图像特征, 最后将多粒度特征和中间层特征一起输入到 Softmax 损失函数, 设计出了 MGMN 算法。MGMN 算法无论是在现有的分心驾驶检测公开数据上还是在自建数据集上都有着更高的精确度和较短的运行时间。

(2) 本文设计了“树莓派+远程服务器”的驾驶员注意力检测系统。采用“树莓派+远程服务器”的结构, 通过局域网将树莓派与远程服务器相连接, 具有较高的信息传输速率, 树莓派与服务器之间采用 Socket 通信方式进行通信, 做到终端与服务器之间的信息的准确传输。同时采用成本低、体积小、可靠性高的树莓派作为驾驶图像实时采集终端、采用配置高、运行效率高的服务器作为状态判别的设备, 降低了设计成本, 实现了对驾驶员注意力状态实时的高精度的检测。

6.2 展望

本文提出的方法相比其他的驾驶员注意力检测算法精确度有了一定的提升, 设计的驾驶员注意力检测系统有一定的实用性, 但是仍然存在一些可以改进和提升的地方。

(1) 缺乏大规模的数据集。截止到目前为止, 公开的分心驾驶检测数据集中只有 State Farm 公司的数据集包含的驾驶图像最多, 有 20000 余张, 开罗美国大学公开的数据集只有 10000 余张, 并且其中混入了大量不可用的数据, 本文用到的自建数据集虽然质量高但是图像数量只有 5000 余张, 进行数据集的制作比较耗时耗力。驾驶员的注意力检测很有必要, 但是目前相对来说还没有大规模的驾驶图像数据集, 尤其时针对夜间行车的数据集, 现有的公开数据集都是针对的白天的驾驶行为。分心驾驶尤其是疲劳驾驶更多地会发生在夜间行车地场合, 但是截止到目前为止, 尚无针对夜间行车的分心驾驶数据集, 后续我将会制作针对夜间行车的驾驶员分心检测数据集。

(2) 本文设计的驾驶员注意力检测系统中, 采用了终端与服务器远程连接的情况, 是考虑到了目前可进行深度学习计算的终端价格较贵, 但是随着技术的不断地发展, 在不久的将来, 可进行深度学习计算的终端价格会逐渐降低,

在终端上部署轻量化的模型，可以在终端上独立地进行驾驶员注意力图像的采集、检测和结果的输出。

（3）本文设计的是“1+1”情况下的驾驶员注意力检测系统，即 1 个终端连接了 1 个远程服务器，在日后的研究中，可以设计基于“n+1”结构的检测系统，一台服务器通过高速无线网络与 n 台终端相连接，进行大范围内驾驶员注意力的检测，实现计算资源的充分使用。

参考文献

- [1] 蒋菱枫. 全国私家车保有量首次突破 2 亿辆, 66 个城市汽车保有量超过百万辆[EB/OL]. (2020). [2020-09-30]. <https://www.mps.gov.cn/n2254314/n6409334/c6852472/content.html>.
- [2] 中国新闻网. 商务部: 2020 年末中国汽车保有量将有望超越美国[EB/OL]. (2020-04-09). [2020-09-30]. <http://www.chinanews.com/auto/2020/04-09/9152030.shtml>.
- [3] World Health Organization. Global status report on road safety2018[M]. Switzerland: World Health Organization, 2018: 2-8.
- [4] Li C, Jiang X, Wang W, et.al. A Simplified Car-following Model Based on the Artificial Potential Field[J]. Procedia Engineering, 2016, 137:13-20.
- [5] 张玉峰. 驾驶员状态检测及其在人机共驾中的应用[D]. 重庆: 重庆大学汽车工程学院, 2018.
- [6] Hong M. J., Arshad M. R.. A Balance-Artificial Potential Field Method for Autonomous Surface Vessel Navigation in Unstructured Riverine Environment[J]. Procedia Computer Science, 2015, 76: 198-202.
- [7] 中华人民共和国工业和信息化部. 2019 年通信业统计公报[EB/OL]. (2020-02-27). [2020.09.30]. <http://www.miit.gov.cn/n1146312/n1146904/n1648372/c7696411/content.html>.
- [8] Lal S K, Craig A, Boord P, et al. Development of an algorithm for an EEG-based driver fatigue countermeasure[J]. Journal of Safety Research, 2003, 34(3): 321-328.
- [9] Seong K., Haet L., Jung K., et al. ECG, EOG detection from helmet based system[C]. //2007 6th International Special Topic Conference on Information Technology Applications in Biomedicine. Tokyo, Japan: ITAB, 2007: 191-193.
- [10] Kecklund G., Akerstedt T.. Sleepiness in long distance truck driving: an ambulatory EEG study of night driving[J]. Ergonomics, 1993, 36(9): 1007-1017.
- [11] Kithil P. W., Jones R. D., Mccuish J. Driver Alertness Detection Research Using Capacitive Sensor Array[C]. //Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Aspen, CO, USA, 2001: 324-329.
- [12] Bhuwan Mohan Karan, Rakesh Kumar Sinha, Prabhat Kumar Upadhyay. Predicting heat-stressed EEG spectra by self-organising feature map and learning vector quantizers——SOFM and LVQ based stress prediction[J]. Journal of Biomedical Science and Engineering, 2010, 3(5): 529-537.
- [13] Li-Wei Ko, Wei-Kai Lai, Wei-Gang Liang, et al. Ingle channel wireless EEG device for real-time fatigue level detection[C]. //2015 International Joint Conference on Neural Networks. Killarney: IJCNN, 2015: 1-5.
- [14] 王琳虹, 李世武, 周茹波, 杨志发, 冀秉魁, 姚雪萍. 高速公路路侧景观色彩对驾驶员心率指标的影响[J]. 吉林大学学报(工学版), 2013, 43(01): 74-80.
- [15] 扈静, 叶成文, 余涛, 孙蕊. 基于心电机电生理信号的驾驶疲劳评价[EB/OL]. (2018-04-11). [2020.09.30]. <http://www.paper.edu.cn/releasepaper/content/201804-107>.
- [16] 祝荣欣. 基于生理信号的联合收获机驾驶疲劳检测与评价[D]. 哈尔滨: 东北农业大学, 2016.
- [17] 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.

- [18] Friedrichs F., Yang Bin. Drowsiness monitoring by steering and lane data based features under real driving conditions[C]. //18th European Signal Processing Conference. Aalborg, Denmark: EUSIPCO, 2010: 209-213.
- [19] 李伟, 何其昌, 范秀敏. 基于汽车操纵信号的驾驶员疲劳状态检测[J]. 上海交通大学学报, 2010, 44(2): 292-296.
- [20] Sandberg D., Wahde M. Particle swarm optimization of feedforward neural networks for the detection of drowsy driving[C]. //IEEE International Joint Conference on Neural Networks. Hong Kong, China: IEEE, 2008:788-793.
- [21] 刘军, 王利明, 聂斐, 袁俊, 程伟. 基于转向盘转角的疲劳驾驶检测方法研究[J]. 汽车技术, 2016(04): 42-45.
- [22] Daptardar S., Lakshminarayanan V., Reddy S., et al. Hidden Markov Model based driving event detection and driver profiling from mobile inertial sensor data[C]. //2015 IEEE Sensors. Busan, South Korea: IEEE, 2015:1-4.
- [23] Q. Ji, Z. Zhu, and P. Lan. Real-time nonintrusive monitoring and prediction of driver fatigue[J]. IEEE transactions on vehicular technology, 2004, 53(04): 1052-1068.
- [24] Coetzer R. C., Hancke G. P. Driver fatigue detection: A survey[C]. //Africon 2009. Nairobi: IEEE, 2009: 1-6.
- [25] Dinges D. F., Grace R.. PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance[EB/OL]. (1998). [2020.09.30]. <https://permanent.fdlp.gov/gpo6571/tb98-006.pdf>.
- [26] Brandt T., Stemmer R., Rakotonirainy A. Affordable visual driver monitoring system for fatigue and monotony[C]. //IEEE International Conference on Systems, Man and Cybernetics. The Hague, The Netherlands: IEEE, 2004: 6451-6456.
- [27] Grace R., Steward S. Drowsy Driver Monitor and Warning System[C]. //Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Aspen, CO, USA: University of Iowa, 2001: 64-69.
- [28] Rong-Ben W, Ke-You G, Shu-Ming S, et al. A monitoring method of driver fatigue behavior based on machine vision[C]. //IEEE Intelligent Vehicles Symposium. Columbus, USA: IEEE, 2003: 110-113.
- [29] Gong S, Buxton H. Understanding visual behaviour[J]. Image & Vision Computing, 2002, 20(12): 825-826.
- [30] Harbluk J, Noy Y, Trbovich P, et al. An on-road assessment of cognitive distraction: impacts on drivers' visual behavior and braking performance[J]. Accid Anal Prev, 2007, 39(2): 372-379.
- [31] Suganya, G, Premalatha, M, Bharathiraja, S, Rohan, A. A low cost design to detect drowsiness of driver[J]. International Journal of Civil Engineering and Technology, 2017, 8(9):1138-1149.
- [32] Palani S, Kothandaraman S. A Low Cost Drivers Drowsiness Detection System For Autonomous Mobile Vehicle[J]. Ipsj Sig Notes, 2008, 2008(02): 61-67.
- [33] Rogado, E.; Garcia, J.L.; Barea, R.; Bergasa, L.M. Driver Fatigue Detection System[C]. //Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand: IEEE, 2009: 1105-1110.
- [34] Sendra, S., Garcia, L., Jimenez, J.M., Lloret, J. Low-Cost Vehicle Driver Assistance System for Fatigue and Distraction Detection[C]. //Proceedings of the First International Conference Future Intelligent Vehicular Technologies. Porto, Portugal: Springer International Publishing, 2016: 69-78.

- [35] Abouelnaga Yehya, Hesham M. Eraqi, Moustafa M. N. Real-time Distracted Driver Posture Classification[C]. //32nd Conference on Neural Information Processing Systems. Montréal, Canada: NIPS, 2018:1-8.
- [36] Hesham M. Eraqi, Yehya Abouelnaga, Mohamed H. Saad, Mohamed N. Moustafa. Driver Distraction Identification with an Ensemble of Convolutional Neural Networks[J]. Journal of Advanced Transportation, Machine Learning in Transportation (MLT) Issue, 2019:1-16.
- [37] E. Ohn-Bar, S. Martin, M. M. Trivedi. Driver hand activity analysis in naturalistic driving studie[J]. Journal of Electronic Imaging, 2013, 22(4): 041119-041119.
- [38] Leekha M., Goswami M., Shah R. R., et al. Are You Paying Attention? Detecting Distracted Driving in Real-Time[C]. //The 5th IEEE International Conference on Multimedia Big Data. Newark, California, USA: IEEE, 2019:171-180.
- [39] Qing-Zhi Bu, Jun Qiu, and Hu Chao Hu. Research on Driver Attention Behavior Detection Method Based on HOG Feature Extraction and SVM[J]. Integrated Technology, 2019, 8(04): 69-75
- [40] C. H. Zhao, B. L. Zhang, J. He, and J. Lian. Recognition of driving postures by contourlet transform and random forests[J]. IET Intelligent Transport Systems, 2012, 6(2):161-168.
- [41] Wikipedia. Natural neuron[EB/OL]. (2020.08.30). [2020.09.30]. <http://en.wikipedia.org/wiki/Neuron>.
- [42] 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.
- [43] Cybenko G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of control, signals and systems, 1989, 2(4): 303-314.
- [44] Wikipedia. Artificial neural network[EB/OL]. (2020-08-30). [2020-09-30]. http://en.wikipedia.org/wiki/Artificial_neural_networks.
- [45] Jürgen, Schmidhuber. Deep Learning in neural networks: An overview[J]. Neural Networks, 2015, 61: 85-117.
- [46] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [47] He K., Zhang X., Ren S.. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]. //Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 1026-1034.
- [48] FAIR. From research to production[EB/OL]. (2020). [2020-09-30]. <https://pytorch.org>.
- [49] OpenCV team. About[EB/OL]. (2020). [2020-09-30]. <https://opencv.org/about>.
- [50] Karimi A. M., Fada J. S., Parrilla N. A., et al. Generalized and Mechanistic PV Module Performance Prediction From Computer Vision and Machine Learning on Electroluminescence Images[J]. IEEE Journal of Photovoltaics, 2020, 10(3): 878-887.
- [51] Rafael C., Gonzalez, Richard E Woods. Digital image processing[J]. Prentice Hall International, 2008, 28(4): 484-486.
- [52] Wang G., Yuan Y., Chen X., et al. Learning Discriminative Features with Multiple Granularities for Person Re-Identification[C]. //2018 ACM Multimedia Conference. Seoul, Korea: ACM, 2018: 22-26.
- [53] Michel Tokic, Günther Palm. Value-Difference Based Exploration: Adaptive Control between Epsilon-Greedy and Softmax[C]. //Annual Conference on Artificial Intelligence. Berlin, Heidelberg: Springer, 2011: 1-12.
- [54] Selvaraju R. R., Cogswell M., Das A., et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

致 谢

时光飞逝，岁月荏苒，两年多的硕士研究生生涯即将结束，我依然记得那个 2018 年的夏天，我满怀着憧憬步入浙江工业大学校园，而今已是 2020 年秋天。在这美好的秋日，细细回味这人生中短暂的两年多的时光，我虽然时而担心自己的头发，时而担心自己的体重，但是我不仅收获了知识，也深切地感受到了人与人之间的美好的关怀和无私的帮助。在这里，我要向每一位帮助过我的老师、同学、朋友和家人，致以最诚挚的感谢！

首先，我要感谢我的导师赵云波教授。赵老师尊重每一位学生的研究兴趣，并没有给我硬性规定研究方向，不仅如此，他还在我的研究进展遇到瓶颈时提供无私的帮助。赵老师为我创造了良好的科研条件和环境，只要是能够对本文的科研有帮助的地方，老师都会尽力满足我，这两年中，我一直在其他老师的学生的艳羡中度过。在撰写学术论文的过程中，赵老师反复教导，教会了我如何写科技文献，如何排版措辞，如何描述清楚一个问题，老师花费了大量时间在我的论文上，没有老师的悉心指导，我的论文就不能顺利产出。当然，我最崇敬的是赵老师的人格魅力，在物欲横流的时代，赵老师始终坚守着科研的底线，对学术剽窃零容忍，老师对学术有着崇高的追求，宁缺毋滥，始终把握着论文的质量关。

同时我还要感谢我的师兄、现供职于华为杭州研究院的林建武工程师，跟着师兄学习的半年时间里，我受益匪浅。在我研究遇到障碍时，就会向师兄求助，师兄总会不厌其烦地为我提供指导，陪我一起讨论论文的创新点、算法的改进之处等。师兄刻苦钻研的精神也深深地感染着我，每天一早就能在实验室看到师兄的身影，直到深夜方才离开。衷心祝福林师兄前程似锦，早日找到心仪的女生。

特别地，我还要感谢我的辅导员严庆乾老师，是严老师在我人生迷茫的时候为我指明了方向，告诉我专业硕士虽然可以不发表论文，但是没有一篇论文的研究生生涯是不完美的，也是他时常督促我要努力上进，不能浑浑噩噩。在这里再次感谢严庆乾老师给予我的无私帮助。

其次我还要感谢我的朋友：陈鹏程、胡成鹏、潘继炜、乔洪帅、沈佳琪等等，你们经常在科研之余和我出去吃一顿，缓解科研压力，排解我心里的烦闷，为我枯燥的学习时光增添了不少乐趣，使我的性格更为开朗乐观，我也收获了真挚的友谊。

此外，我还要感谢我的同届：梁启鹏、王岭人、朱创、朱巧慧，更要感谢我的师兄师姐师弟师妹们，是你们在我的科研和生活中给了我很大的帮助。

最重要的是，感谢我的家人在这二十多年来的养育，无论经历多少的风风雨雨，都会支持着我，让我可以安心地学习，感谢我的家人为我付出的一切！由衷地感谢评阅专家、老师和学者，在百忙之中审阅我的论文！

作者简介

1 作者简历

1994 年 11 月出生于江苏省常州市。

2014 年 9 月——2018 年 6 月，宿迁学院机电工程学院自动化专业学习，获得工学学士学位。

2018 年 9 月——2021 年 1 月，浙江工业大学信息工程学院控制工程专业学习，攻读专业型硕士学位。

2 攻读硕士学位期间发表的学术论文

[1] Min Tang, Fang Wu, Li-Li Zhao, Qi-Peng Liang, Jian-Wu Lin, Yun-Bo Zhao. Detection of Distracted Driving Based on Multi-Granularity and Middle-Level Features. 中国自动化大会, 2020

[2] 李灏, 唐敏, 林建武, 赵云波. 基于改进困难三元组损失得跨模态行人重识别框架. 计算机科学, 2020

3 参与的科研项目及获奖情况

[1] 基于资源调度和预测控制的无线网络化控制系统的联合设计. 中国国家自然科学基金项目(61673350).

[2] 复杂环境下非完全信息博弈决策的智能基础模型研究. 科技创新 2030-“新一代人工智能”重大专项课题(2018AAA0100801)

4 发明专利

[1] 赵云波, 何江涛, 唐敏, 倪洪杰, 余世明. 一种基于 IEEE 802.15.4 协议的无线网络化控制系统的调度算法. 中国, CN109327818A[P]. 2019-02-12.

[2] 赵云波, 唐敏, 赵丽丽. 一种人机系统中人的状态的识别方法. 中国, CN110598616A[P]. 2019-12-20.

[3] 赵云波, 唐敏, 赵丽丽, 吴芳. 一种程序员的疲劳程度的检测方法. 中国, CN110786869A[P]. 2020-02-14.

- [4] 赵云波, 唐敏, 吴芳, 赵丽丽. 一种基于机器视觉的驾驶员注意力检测方法. 中国, CN111553190A[P]. 2020-08-18.
- [5] 赵云波, 林建武, 唐敏, 李灏. 一种基于困难五元组的跨模态行人重识别方法. 中国, CN111597876A[P]. 2020-08-28.
- [6] 赵云波, 许德衡, 唐敏. 一种基于贝叶斯推断的网络丢包率学习方法. 中国, CN111614510A[P]. 2020-09-01.
- [7] 赵云波, 许德衡, 唐敏. 一种基于贝叶斯推断的动态数据包打包方法. 中国, CN111614436A[P]. 2020-09-01.
- [8] 赵云波, 唐敏, 朱创, 孙悦铖. 一种基于图像的电脑操作员的疲劳检测方法. 中国, CN111695510A[P]. 2020-09-22.
- [9] 赵云波, 唐敏, 花婷婷, 赵丽丽. 一种基于多粒度特征与中层特征的分心驾驶检测方法. 中国, CN111695535A[P]. 2020-09-22.
- [10] 赵云波, 朱创, 唐敏, 周庆瑞. 一种基于高斯混合概率假设密度滤波器的多目标航迹提取方法. 中国, 202010636230[P]. 2020-07-03.

学位论文数据集

密 级*	中图分类号*	UDC*	论文资助
公开	TH741	681.2	
学位授予单位名称	学位授予单位代码	学位类型*	学位级别*
浙江工业大学	10337	工程硕士	全日制专业型硕士
论文题名*	基于机器视觉的驾驶员注意力检测系统设计		
关键词*	驾驶员, 注意力检测系统, 多粒度特征, 中间层特征, 卷积神经网络		论文语种*
并列题名*	无		中文
作者姓名*		学 号*	
培养单位名称*	培养单位代码*	培养单位地址	邮政编码
浙江工业大学 信息工程学院	10337	杭州市潮王路 18 号	310032
学科专业*	研究方向*	学 制*	学位授予年*
控制科学与工程	计算机视觉	2.5 年	2021 年
论文提交日期*	2021 年 01 月		
导师姓名*	赵云波	职 称*	教授
评阅人	答辩委员会主席*	答辩委员会成员	
盲评	石崇源	宣琦, 赵云波、闵勇、傅晨波	
电子版论文提交格式: 文本 (☒) 图像 (☐) 视频 (☐) 音频 (☐) 多媒体 (☐) 其他 (☐)			
电子版论文出版 (发布) 者	电子版论文出版 (发布) 地		版权声明
论文总页数*	62 页		
注: 共 33 项, 其中带*为必填数据, 为 22 项。			