

系统仿真学报
Journal of System Simulation
ISSN 1004-731X, CN 11-3092/V

《系统仿真学报》网络首发论文

题目: 基于级联神经网络疲劳驾驶检测系统设计
作者: 敖邦乾, 杨莎, 令狐金卿, 叶振环
DOI: 10.16182/j.issn1004731x.joss.20-0703
收稿日期: 2020-09-15
网络首发日期: 2021-03-08
引用格式: 敖邦乾, 杨莎, 令狐金卿, 叶振环. 基于级联神经网络疲劳驾驶检测系统设计. 系统仿真学报. <https://doi.org/10.16182/j.issn1004731x.joss.20-0703>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于级联神经网络疲劳驾驶检测系统设计

敖邦乾¹ 杨莎² 令狐金卿¹ 叶振环¹

(1 遵义师范学院工学院 贵州遵义 563006; 2 遵义师范学院物理与电子科学学院 贵州遵义 563006)

摘要：算法通过调整输入图片的大小、扩大最小人脸尺寸以及减小检测窗口层间放缩比例等方式，在保证准确率的同时极大的提升了人脸检测速率，其检测效率为原始 MTCNN 的 18 倍；其次，构建新的卷积神经网络结构模型检测眼、嘴，网络检测的准确率可以达到 95.6%。将设计的检测网络与经过改进后的 MTCNN 进行级联，在已经检测出来的脸部区域继续对眼、嘴进行分类及定位；最后，设置综合性的疲劳检测函数，加强了由于单一特征而出现的虚检及误检等错误检测，准确率可以达到 95.7%。

关键词：卷积神经网络；人脸检测；疲劳判定；级联网络；边界框

中图分类号：TP277

文献标识码：A

DOI: 10.16182/j.issn1004731x.joss.20-0703

Design of Fatigue Driving Detection System Based on Cascaded Neural Network

Ao Bangqian¹ Yang Sha² Linghu Jinqing¹ Ye Zhenhuan¹

(1.College of Engineering, Zunyi Normal University, Zunyi, Guizhou, 563006, China ;

(2.College of Physics and Electronics, Zunyi Normal University, Zunyi, Guizhou, 563006, China ;)

Abstract: Methods are proposed such as *adjust the size of input image, expand minimum face size, and reduce scaling ratio between layers of the detection window*, which greatly improves face detection rate while ensuring the accuracy. The detection efficiency is 18 times of original network MTCNN. Secondly, by *building a new CNN structure model to detect eyes and mouths*, the accuracy of network detection can reach 95.6%. *The proposed network is cascaded with original MTCNN to continue locate the eyes and mouth in former face area that has been detected.* Finally, the false detection and false detection errors due to a single feature are strengthened through *setting a comprehensive fatigue detection function*, and the accuracy rate can reach 95.7%.

Key words: CNN; face detection; fatigue determination; cascaded network; bounding box

引言

近年来，随着中国经济的飞速发展及汽车制造业技术的成熟，中国的车辆保有量逐年上升，与此同时，我国由于交通事故造成的伤残死亡人数一直是世界上最多的国家之一，每年因交通事故死亡的至少有 50 万以上，交通事故已经被公认为是危害当今人类生命安全的第一大公害。根据交通事故统计分析表明，交通事故中 80%~90% 是人为因素造成的，驾驶员疲劳驾驶更会使车祸发生的可能性成倍的增加，在我国，大约 48% 的车祸是由驾驶员疲劳驾驶造成的，交通事故通常就发生在几秒钟的时间内，如果能提前 2 秒钟预警将

能减少 92% 的交通事故，提前 0.5 秒钟预警，将会避免 73% 的交通事故，因此，一种高效快速的预警系统，可以拯救成千上万人的生命，避免财产及经济的损失。

目前的疲劳检测技术及预警系统，主要分为三类：基于车辆的驾驶模式^[1-4]、基于驾驶员的生理心理特征^[5-8]、基于计算机视觉技术^[9-12]。其中基于车辆的驾驶模式容易受驾驶员个人习惯的影响出现误判；基于驾驶员生理心理特征需要相应的传感器直接接触驾驶员的身体，容易造成不适应感，并影响操作，同时其信号也容易受到噪音的干扰；基于计算机视觉技术方法，主要是通过图像分析手段对驾驶员的脸部与眼部特征进行疲劳估计，由于其具有非接触、无干扰及精度高

收稿日期：2020-09-15 修回日期：2020-12-03

基金项目：遵义市校联合科技研发资金项目(遵市科合 HZ 字[2020]10 号)、(遵市科合 HZ 字[2020]16 号)。

作者简介：敖邦乾(1984-)，男，博士，副教授，研究方向为人工智能及智能系统设计，E-mail: aobangqian@163.com；杨莎(1982-)，女，土家族，博士生，副教授，研究方向为电子线路及系统设计，E-mail: 66009731@qq.com；令狐金卿(1985-)，男，博士，副教授，研究方向为智能系统设计及优化，E-mail: 252225578@qq.com；叶振环(1984-)，男，博士，教授，研究方向为机械设计及性能优化，E-mail: 253703978@qq.com；

等优点，目前得到了广泛的研究。随着深度学习在目标检测及跟踪领域的巨大成功，其结合视频图像的卷积神经网络(Convolutional Neural Network, CNN)也推动了对于疲劳驾驶检测的研究，文献[13]提出一种基于 CNN 的计算眨眼频率的眼睛状态识别方法，同时采用红外视频检测提升了准确率，文献[14]通过利用 CNN 提取手部位置的判别信息，学习和预测安全与不安全的驾驶姿势，文献[15]提出基于 EOG 的疲劳回归模型，在有效判断状态的同时能进行智能识别，减少不利干扰。但是这些方法使用单一的检测因子，在脸部特征检测的多因素表征方面的综合性较差，也没有考虑网络检测的速度，在实时性方面较差，基于以上原因，本文设计改进 MTCNN(Multitask Casaded Convolutional Networks)[16]以及多因子共同检测的疲劳检测算法，有效的改善了目前算法的不足。

1 检测网络设计

1.1 MTCNN 人脸检测

传统的人脸检测算法主要使用特征点定位，并返回其位置数值，但其在人脸频繁动作、目标遮挡及不同光照条件下的鲁棒性较差，甚至不能精确定位。MTCNN 利用级联结构，可以实行人脸检测和对齐，是目前最快速精确的人脸检测器，其三大子网络（如图 1）功能分别为：

P-Net(Proposal Network)子网络是全卷积神经网络(fully convolutional network, FCN)，前向传播得到的特征图在每个位置是 32 维的特征向量，用于判断每个位置处约 12×12 大小的区域是否包含人脸，主要获得人脸区域的候选窗口(candidate facial windows)和边界框的回归向量(bounding box regression vectors)，同时用该边界框做回归，对候选窗口进行校准，然后通过非极大值抑制(non-maximum suppression, NMS)来合并高度重叠的候选框，通过选取合适大小的输入图片，可以减少 resize 图片金字塔的数量，从而减少冗余框的数量，极大的减少计算量，尤其是对实时性要求比较高的系统更为重要。

R-Net(Refine Network)子网络是 CNN，其功能与 P-Net 子网络的功能有点类似，也是通过边界框回归和

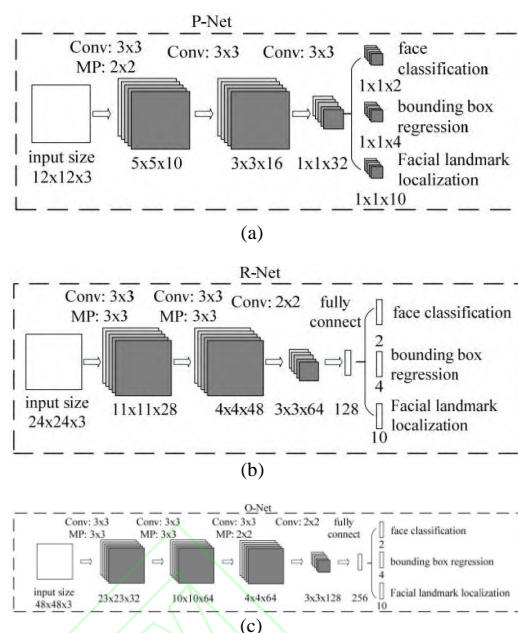


图 1 MTCNN 各子网络结构
Fig1. MTCNN Sub-network Structure

NMS 来丢弃大量的重叠窗体，此网络多了一个全连接层，更利于特征值的提取，得到的人脸检测框更加精准且冗余框更少。

O-Net(Output Network)子网络也是一个 CNN，其比 R-Net 子网络多了一个卷积层，对人脸区域进行更多更严格的监督，对人脸检测框坐标进行进一步的细化，处理的结果更加精确细致，同时会反馈回五个关键点坐标：左眼、右眼、鼻子、左唇端以及右唇端。MTCNN 中，由于每个子网络都进行了人脸与否的二维分类、边界框的大小回归和人脸五个点位置的回归，层层递进，因此具有较强的鲁棒性，同时其还采用人脸对其联合学习进行训练，因此对于旋转了一定角度的驾驶员依然能准确的进行检测，而没有必要进行额外的旋转对齐等操作。使用 MTCNN 检测到人脸信息以后，由于眼、嘴在疲劳驾驶时的表征现象最为明显，通过进一步的算法设计定位眼、嘴区域，通过其不同表象，综合判断驾驶员是否属于疲劳驾驶。

1.2 MTCNN 算法改进

MTCNN 人脸检测算法采用级联结构，网络对人脸的检测经过三个网络，层层过滤，最终定位人脸位置，并返回其中的五个特征点，其通过基于图像滑动窗口的检测方法来实现不同人脸的检测，为了保证对任意大小人脸的检测，通常是六层金字塔结构。虽然金字塔层数越多，越能保证其检测的全面覆盖性，但是其滑动检测窗口的数量也相应的增加了，加大了后

续网络的计算负担。对于实时性要求较高的疲劳驾驶检测,检测速度越快,越能保证安全性,设计合适的原始输入图像,可以减少金字塔层数来加快检测速率。对于大部分车辆而言,其驾驶位置比较固定,且在图片中占据较大的比例,通过适当的调整图像采集设备与驾驶员的位置,可以快速得到人脸图像。通过对原始网络进行分析及多次实验测试,根据等比例放缩原理,当摄像头安装在离驾驶员约 60-80cm 距离,输入图片大小为 320×240 ,同时设置最小人脸尺寸为 60,并适当的缩小检测窗口相邻层间放缩比例,此处设置为 0.637。则只需要两层金字塔,就可以对视频设备采集

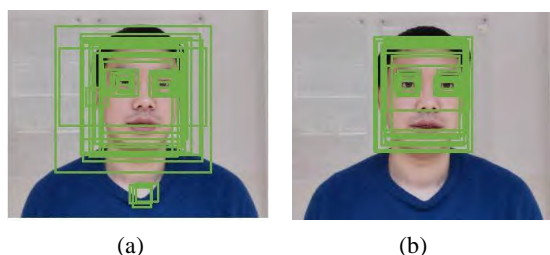


图 2 优化前后检测框对比图
Fig2. The comparison of detection results

到的原始图像的任意位置的人脸进行检测,极大的减少了后面检测框的数量。同时,在行驶过程中,驾驶员位置相对固定,不会存在幅度较大的晃动,因此下一步的检测,对前几帧中有边界框的帧图像进行采样,作为当前帧滑动窗口的搜索区域,避免从原始图像中重新采样,简化了搜索背景,特征值的进一步提取范围也得到了缩小,大大的加快了检测的速率,其优化前后对比效果如图 2 所示,从图中可以看出,经过算法优化处理以后,其返回的检测框区域更加接近真实人脸区域,同时也降低了返回框的误检率。

2 眼、嘴检测网络设计

MTCNN 在提取到人脸信息并返回相关边界框及特征点以后,对于是否属于疲劳驾驶,还需要进一步状态的判定,本文仍然采用级联结构,在 MTCNN 的最后一个子网络 O-Net 的输出端,级联一个卷积神经网络,对眼、鼻进行分类定位并返回边界框,本文设计 EMs-Net(Eyes and Mouth detection Net)来实现这个功能,所设计的眼、嘴检测网络由两个卷积层、两个池化层和两个全连接层构成,其具体结构如图 3 所示,网络的输入图片大小为 48×48 。所设计的 EMs-Net 子网

络在卷积层和池化层之后,有两个全连接层,这样设计的优点在于可以对目标数据集的图片进行准确的分类,所输出眼、嘴分类结果为 $1 \times 1 \times 4$ (是否眼睛、是否鼻子)的向量,边界框返回的是 $1 \times 1 \times 8$ (眼睛边界框左上角及右下角两个坐标、鼻子边界框左上角及右下角两个坐标)的向量。

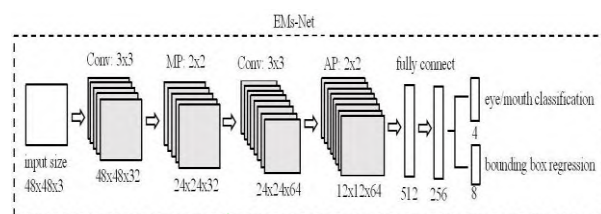


图 3 眼嘴检测网络设计

Fig. 3 The design of eye and mouth detection network

EMs-Net 子网络设计如下:

Step 1: 输入的眼、鼻等训练集图片分 3 个通道进入第 1 个卷积层(Conv)C1,使用 32 个 3×3 卷积核,生成 32 个大小为 48×48 的 feature map;

Step 2: step 1 中的 feature map 经过池化层 S2 进行降维处理,选用 2×2 的 max-pooling(MP),获得 32 个大小为 24×24 的 feature map;

Step 3: step 2 中的 feature map 进入第二个卷积层 C3 进行进一步的特征提取,使用 64 个卷积核,尺寸大小仍然选择 3×3 ,生成 64 个大小为 24×24 的 feature map;

Step 4: step 3 中的 feature map 经过池化层 S4 进行降维处理,选用大小为 2×2 的 ave-pooling(AP),获得 64 个大小为 12×12 的 feature map;

Step 5: step 4 中的 feature map 分别经过神经元为 512、256 的全连接层 F5、F6,最后一层为 Softmax 层,对图片的输出进行 4 分类。

子网络设计完成以后,准备合适的数据集对其进行训练,数据的多样性会极大的提高特征的表征能力,因此需要采集不同光照条件下、不同性别及年龄的数据集,本文所使用的数据据主要来自人脸数据库 Faces in the Wild、MTFL、Face Warehouse,眨眼视频数据集 ZJU 和疲劳驾驶视频数据集 YawDD 等,还有一部分来自于实验室的各位成员,同时为了增加样本的多样性及降低数据采集工作量,将部分图片数据进行水平、倾斜以及垂直翻转等处理,最后收集到眼部数据共计 11234 张,嘴部数据集 12365 张,将数据集命名为

EMs(Eyes and Mouses dataset), 同时从该数据集中各选出 2000 张(正、负样本各 1000 张)用于模型的测试, 其余样本则用于模型训练。

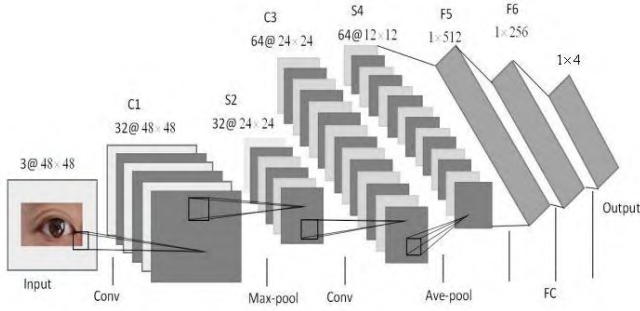


图 4 眼嘴网络训练模型

Fig. 4 The CNN training model of eyes and mouth

本文所收集的用于训练的眼、嘴数据集是人体正常状态下的数据集, 也即只有睁眼状态下和嘴巴闭合状态下的数据集, 这极大的减轻了网络的计算量, 这样选择训练数据集是因为部分驾驶员在打哈欠时, 喜欢用手捂住嘴部, 或者眼部疲劳时, 用手进行揉搓, 如果同时训练睁眼闭眼数据及张嘴闭嘴时, 会因为个人习惯, 检测不到目标而导致误警及虚警, 同时由于返回的类别较少, 在实时性方面也比较有优势。上述设计的网络在经过训练以后, 能成功的检测到相关的器官, 然后再进行状态的判断, 其训练模型如 4 所示。

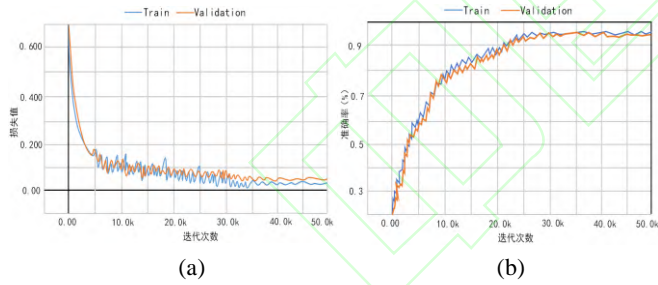


图 5 数据集训练及验证损失值及精度

Fig. 5 Training/Validation Loss and accuracy

EMs-Net 子网络中, 输出有两个, 一个为有眼睛-无眼睛及有嘴巴-无嘴巴的分类, 主要是对眼、嘴进行四分类, 是一个分类问题, 其分类的 loss 函数使用交叉熵 loss 函数:

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

上式中, p_i 表示眼、鼻出现的概率, y_i^{cls} 表示该区域的真实标签; 网络另外一个输出为分类目标的边界框, 其值越精确越好, 是一个回归问题, 采用欧氏距离作为距离度量的 loss 函数:

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

其中 \hat{y}_i^{box} 为子网络预测得到的边界框坐标, y_i^{box} 为目标的实际边界框坐标, 子网络的总 loss 函数为上述两个 loss 的加权求和:

$$\min \sum_{i=1}^N \sum_{j \in \{cls, box\}} \omega_j \beta_i^j L_i^j \quad (3)$$

式(3)中 ω 表示权重, $\beta = (0,1)$ 表示数据集样本类型, 本文所设计的网络在经过 50000 次的迭代后, 其损失值和准确率曲线如图 5 所示, 从图中可以看出, 其损失值在不断的下降, 最后收敛到一个较小的值, 同时其准确率也在不断的提高, 最终稳定在一个较高的数值。为了验证本文算法的可行性及优越性, 将本文提出的 EMs-Net 与 SSD^[17] 进行对比, 并在准确率及单一图片的时间开销两方面进行了对比, 其对比结果如表 1、表 2 所示。

表 1 网络分类准确率对比(%)

Tab.1 Comparison of classification accuracy (%)

分类	测试样本量	本文样本准确率(%)	SSD 样本准确率(%)
左眼	2000	95.6	98.6
右眼	2000	96.4	98.8
嘴部	2000	96.8	98.9

表 2 单张图片时间开销(ms)

Tab.2 Time cost of single picture (ms)

测试任务	本文时间开销	SSD 样本时间开销
左眼检测分类	2.685	6.453
右眼检测分类	2.668	6.386
嘴部检测分类	2.488	6.568

从测试结果表 1 可知, 对于给定的测试样本, 两种算法都能达到 95.6% 以上的准确率, 从对比测试的结果可知, 本文提出的算法在分类方面比 SSD 的准确率较低, 这是因为 SSD 使用了足够多的卷积层进行了特征信息的提取, 同时其分类框也进行了更为细致的划分, 最后使用的非最大值抑制算法也加大了其准确率, 相比较而言, 本文只使用了少量卷积层和池化层进行特征信息的提取, 紧随其后的全连接层也只有两层, 因此分类准确率较低, 但是相应的, 从表 2 可知, 由于计算量少及结构相对简单, 其对于单张图片的时间开销也大大的降低了, 这对于实时性要求很高的疲劳检测系统, 在牺牲一定程度的准确率的代价基础上, 大大提高其检测速度, 是完全可行的, 同时, 由于本文

所设计的网络是级联在原网络的后面，在已经检测到人脸的区域进行眼、嘴的分类及定位，其搜索区域大大的缩小了，极大的减少了计算量。

经过上述设计及训练以后，网络在检测到人脸的基础上，当检测到睁开的眼睛及闭合状态下的嘴巴数据时，返回其边界框，当检测闭合状态下的眼睛以及嘴部形状明显变化的嘴部时，则只返回脸部的边界框，



图 6 正常情况下眼嘴检测结果返回框
Fig. 6 The detection with bound box under normal conditions



图 7 眼嘴各种状态下检测结果返回框
Fig. 7 The detection with bound box in various states

而没有眼、嘴的边界框，图 6 为部分人脸检测结果。从图 6 可以看出，本算法网络对不同性别以及不同背景环境下的人脸及眼、嘴能进行精确的检测，其对于脸部侧移、背景转换及光照变化也具有鲁棒性，特定位置的返回框对也比较准确，为后面的疲劳检测算法设计提供了很好的数据依据，同时，从图 7 中也可以看出，由于未训练闭眼状态及张嘴状态下数据，而由于其形态发生了较大的变化，因此检测不到眼、嘴数据，也就没有返回框。

3 疲劳判定算法设计

疲劳驾驶时，会出现眨眼频率降低、眨眼次数减少以及打哈欠等特征，但是单一的眼部及嘴部特征的变化有局限性，容易受外界干扰以及驾驶员的个人习惯影响，出现误警及虚警等，疲劳驾驶判定需要综合多方面的因素才能做出比较精确的判定。本文在精确定位到眼睛和嘴巴的位置并实时跟踪这些器官的同时，通过计算疲劳指数，来进行相应的后续操作，其主要参数有如下两个方面：

(1) 眨眼频率(BlinkFreq, BF)参数 f_{BF} :

PERCLOS 参数 f_p 是目前国际上公认的疲劳判定准则，表示在一定时间内，眼睛处于闭合状态的时间

占据的比例，视频流信息的帧序列具有连续性，因此使用在一段时间内眼睛闭合状态下的总帧数 N_{close} 与该段时间内总的帧数 N_{total} 的比值来求取 PERCLOS 的值 f_p 如下：

$$f_p = N_{close} / N_{total} \times 100\% \quad (4)$$

使用 POERCLOS 参数，就必须检测睁眼及闭眼两种状态，数据集必须包含闭眼数据集及睁眼数据集两种数据，这极大的增加了神经网络训练时的数据量及计算量，同时在实时检测过程中，由于本文在原 MTCNN 三级级联结构的基础上，又级联了一层，多一种状态的测试，会使检测结果有比较明显的延迟，因此 POERCLOS 参数不适合本文的神经网络结构，本文使用 BF 参数 f_{BF} 来进行眼部状态的判断。

正常情况下人的 BF 为每分钟 20 ~ 30 次，每次眨眼时间为 0.25 ~ 0.3 s，如果在一段时间内 BF 过高，基于本文算法，其在较长时间内没有眼部检测返回框，则有可能处于疲劳状态，其更能表征驾驶员的驾驶状态。本文对某一段视频中的部分时段的眼部状态进行了测

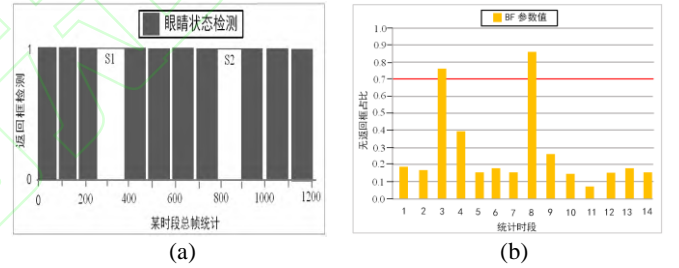


图 8 某时段眨眼频率参数

Fig. 8 Blink frequency parameter in a certain period of time

试分析，图 8 显示其眼部状态返回框的变化情况及 BF 的数值，其中 0 表示没有眼部返回框，1 表示有眼部返回框，分别对应于眼部闭眼及睁眼状态，从图中可以看出，其能对眼部正常的眨眼行为和瞌睡行为 S1、S2 进行精确的帧统计及行为区分，假设一段时间内的眨眼次数 n ，每次眨眼时眼睛闭合状态下无返回检测框的帧数为 M ，这段时间内的总帧数 N ，则 BF 按照如下公式进行计算：

$$f_{BF} = \frac{nM}{N} \times 100\% \quad (5)$$

(2) 打哈欠参数 f_{yawn} :

通常嘴部动作有闭合、说话以及打哈欠三种状态，与眼部数据集的收集一样，本文在搜集嘴部数据集时，主要收集闭合状态下的嘴部数据，同样可以减少训练时的工作量，同时在网络检测状态时，也只需要较少

的时间损耗。由于个人习惯不同,有些人习惯嘴部微张,或者在驾驶过程中会进行说话交流,由于其幅度不是太大,因此数据集中也需要包含一部分嘴部微张的数据集,以增加数据的多样性,这样在检测时,就能检测到闭合及说话等正常状态下的嘴部位置,并返回检测框,而打哈欠时,嘴部动作状态变化明显,网络认为不是嘴部部位,因此不会返回检测框。图 9(a)

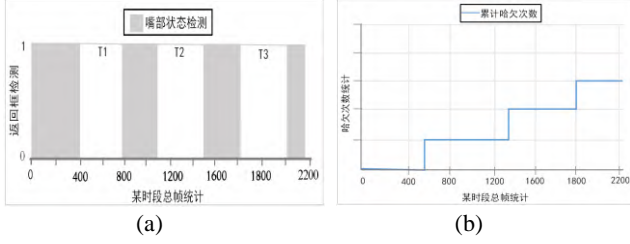


图 9 某时段嘴部状态返回框数量统计
Fig. 9 Mouth states in a certain period of time

中为对某一具有哈欠动作的视频进行了测试,同时显示了其嘴部状态返回框的变化情况及打哈欠的统计数值,其中 0 表示没有返回框,1 表示有返回框,分别对应于嘴部打哈欠的大幅度张开及正常闭合或说话

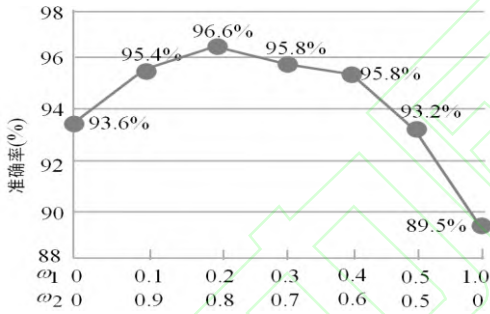


图 10 权值 ω_1 、 ω_2 寻优
Fig. 10 The weight optimization of ω_1 、 ω_2

状态,从图中可以看出,其能对嘴部正常行为和打哈欠行为 T1、T2、T3 也都进行了精确的帧统计及行为区分,同时疲劳驾驶时,打哈欠的频率也会增大,可以计算其比值来统计疲劳程度,图 9(b)对这段视频里面的哈欠次数进行了统计。通过统计在一定时间内没有嘴部返回框帧数量 s ,与这段时间内的帧数 S ,确定其打哈欠参数:

$$f_{yawn} = \frac{s}{S} \times 100\% \quad (6)$$

从图 8 中可以看出,眼部状态正常情况下,其 f_{BF} 值通常会小于 0.3,疲劳情况下一般则会超过 0.7. 同时从图 9 中可以看出,正常情况下,由于一直可以检测出嘴部位置,因此其 f_{yawn} 通常为零,当有打哈欠行为

时,尤其是疲劳状态下的深哈欠,其无返回框的帧数会持续较长一段时间,其帧占比相应的增大,在设置适当的阈值时,可以统计为一次哈欠行为,只有综合

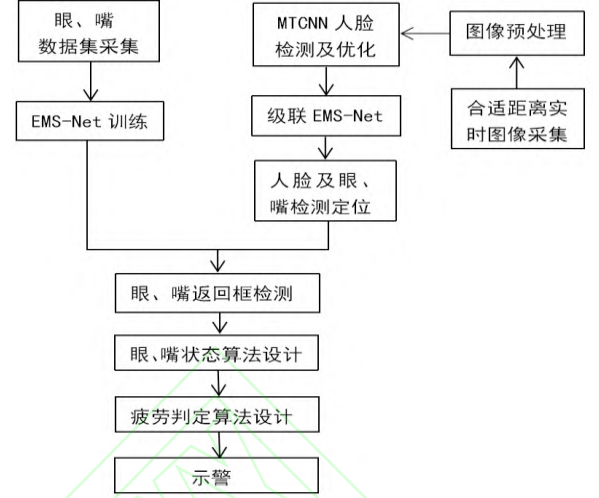


图 11 本文眼、嘴检测网络流程图

Figure. 11 The eyes/mouth detection network flow chart

考虑两种参数,才能对疲劳行为进行准确的检测。单一睁眼频率 f_{BF} 、打哈欠参数 f_{yawn} 无法准确的反应驾驶员的疲劳与否以及疲劳程度,只有综合两个参数的疲劳特征融合,才能较准确的检测出驾驶员的疲劳与否,本文为两个参数设定不同的权重 ω_1, ω_2 ($\omega_1 + \omega_2 = 1$), 其中 $\omega_1 < \omega_2$, 其疲劳指数计算公式为:

$$f = \omega_1 f_{BF} + \omega_2 f_{yawn} \quad (7)$$

根据 f 值分别设定为极度疲劳、一般疲劳及正常等三种情况,对于前两种情况,需要进行相应的示警及提醒操作。不同的权值对于疲劳驾驶的准确度有较大的影响,图 10 为本文算法中测试时的权值寻优图,从图中可以看出,在 $\omega_1 = 0.2, \omega_2 = 0.8$ 时精度可以达到 96.6%。经过上述设计以后,对本文的设计及算法进行相关的测试,图 11 为本文的设计流程图。

4 实验测试及分析

本文所设计的神经网络、所有样本的采集、网络训练和测试都是基于 TensorFlow 2.0 平台,使用 Python 3.6 编程实现的,训练及测试是在处理器为 i7-6700 以及 NVIDIA GeForce GTX 1070 GPU 环境下进行的, GPU 极大的加快了训练速度,同时 CPU 也具有较快的检测速度,为了验证其普适性,本文采集 8 段视频图像对本文算法进行测试,视频图像包含不同年龄以

及不同的光照环境，同时为了测试本算法的有效性，在视频图像采集中，尽量包含各种行为如正常眨眼、说话、闭眼及打哈欠行为等。其测试结果如表 3 所示。

表 3 眼、嘴异常及疲劳行为判定检测
Table 3 The fatigue behavior determination test

视频序号	异常眨眼行为	算法判定疲劳行为	哈欠行为	算法判定疲劳行为
1	5	5	7	7
2	12	11	5	5
3	15	14	9	9
4	8	8	12	12
5	15	14	10	10
6	6	6	8	8
7	13	12	6	6
8	20	20	10	10
有效检出率	95.7%		100%	

表 4 各种算法疲劳行为检出准确率对比
Table 4 Comparison of various algorithms

算法	检出准确率(%)
文献[6]算法	95.0%
文献[7]算法	91.6%
文献[8]算法	92.5%
本文算法	95.7%

从实验测试结果可知，本文算法对哈欠行为的判定准确率很高，几乎没有误判行为，但是在眼睛异常行为的判定时，存在漏检问题，通过返回对所采集的视频进行分析可知，几个漏检的视频中主要包括有带眼镜驾驶员的异常眨眼行为以及在有强光照射下的眼部异常行为，这种漏检情况在后续的研究学习中，尝试使用一些对光照不敏感的视频采集器以及更多的戴眼镜的数据集，以提高网络的准确度。与几种算法的检测准确率对比中，文献[6]所提的基于卷积神经网络分类和三元组定位特征点的算法检出准确率较高，但是其计算量较大，在实时性方面较差，而且算法不够稳定，同时其提出的嘴部宽高比的算法，加大了算法的复杂度，文献[7]使用多尺度卷积神经网络对输入尺寸进行分层输入，同时使用普通摄像头及红外摄像头在不同环境下进行了测试，但是其对眼、嘴两个参数的选择没有设定相应的权重比值，疲劳判定存在很大的随机性，文献[8]提出的基于面部行为的检测算法存在特征不易区分等问题。本文算法则综合了多种疲劳

状态下的特征变化，同时训练比较稳定的卷积神经网络，因此其效果更佳。

表 5 单一图片处理时间开销
Table 5 Single image processing time

任务类别	时间开销(ms)
原始图像获取及优化	3.034
MTCNN 人脸检测	5.686
EMs-Net 眼、嘴提取	2.563
眼部状态返回框	4.569
嘴部状态返回框	3.326
总时间开销	19.177

在高效率检测并分析出疲劳行为的同时，也需要综合考虑实时性，也即单张图片的处理时长，通过对其完成级联任务及处理时间进行统计分析，如表 5 所示，其处理单张图片的时间为 19.177ms，每秒可以处理大约 52 帧图像，并能达到较高的准确率，是完全能够满足实时性要求的。以下测试结果为在实验室外所测试，在使用于车辆时，具有同样的测试效果，其部分测试状态结果如下：

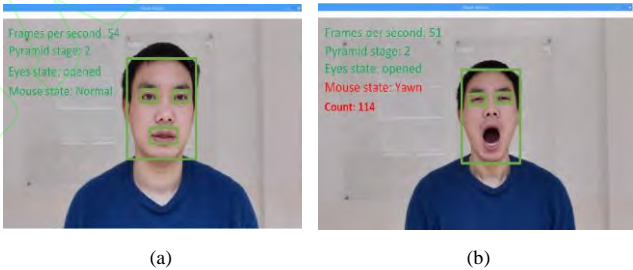


图 12 本文眼嘴检测图

Fig. 12 Eye and mouth detection diagram

在卷积神经网络检测框中，最开始使用的是 MTCNN 人脸检测网络，因此总是会返回人脸框，同时本文级联了 EMs-Net 眼、嘴检测网络，并返回检测框，如图 12 所示，本文算法测试中，同时返回当前检测速率以及人脸金字塔层数，其正常情况下的检测效果如图 12(a)所示，图 12(b)显示其打哈欠状态，并返回网络检测到的帧数，由于可能是才开始的哈欠行为，因此还不算疲劳行为，算法需要继续统计哈欠的次数。图 13 显示了当眼睛闭合一段时间以后，网络的检测结果，当网络检测到闭眼行为超过一定时间以后，本文算法中，当网络检测到 48 帧闭眼数据时，根据本文的算法设计，大概是一秒钟左右，这个属于疲劳行为，系统

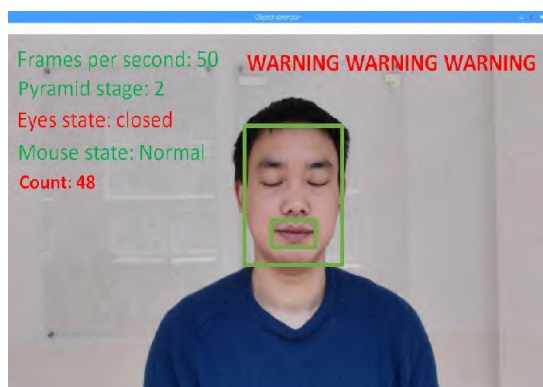


图 13 疲劳行为检测结果
Fig. 13 Fatigue behavior test results

会进行警告提醒。综合疲劳程度，也就是深度疲劳、一般疲劳和正常状态下的检测，本文算法也能进行精确的检测。

5 结论

本文使用卷积神经网络检测算法对疲劳驾驶行为进行了分析研究，首先，使用 MTCNN 对车内驾驶员的人脸进行检测，通过设置合适的视频图像采集距离、增大最小人脸以及缩小层间放缩比例等方式改进优化了其检测速率，同时也不影响其准确率，网络检测速度提升了 18 倍；其次，在原检测网络后面级联 EMs-Net 眼、嘴检测网络，只检测正常状态下的眼睛、嘴部数据，这不仅大大的缩小了数据集采集及训练时的计算量，同时也减少了网络实时检测的计算量，加快了检测速度；最后，通过综合统计眼睛、嘴部图像帧数据，并优化其权值，得到综合的疲劳驾驶行为判断准则。其对深度疲劳、一般疲劳以及正常驾驶行为都有比较精确的测量，准确率可以达到 95.7%，通过分析其检测速率及准确率，并和其它几种算法进行了对比，仿真测试结果也证明了本文算法的实时性及优越性，可以实现真正意义上的车载实时检测，具有较大的应用价值。

参考文献

- [1] Krajewski J, Sommer D, Trutschel U, et al. Steering wheel behavior based estimation of fatigue. Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment Training and Vehicle Design. 2009. 118-124.
- [2] 刘炜煌,钱锦浩,姚增伟,等.基于多面部特征融合的驾驶员疲劳检测算法[J]. 计算机系统应用, 2018, 27 (10): 181-186.(Liu Weihuang, Qian Jinhao, Yao Zengwei, et al. Driver Fatigue Detection Algorithm Based on Multi-Facial Feature Fusion [J]. Computer Systems & Applications, 2018, 27 (10): 181-186.)
- [3] 陈志勇,杨佩,彭力,等. 基于 BP 神经网络的驾驶员疲劳监测研究[J]. 计算机科学, 2015, 42 (Z6): 67-69. (Chen Zhiyong, Yang Pei, Peng Li, et al. Fatigue driving monitoring based on BP neural network [J]. Computer Science, 2015, 42(Z6): 67-69.)
- [4] Mott G E, Forsman P, Short K R, et al. Efficient driver drowsiness detection at moderate levels of drowsiness[J]. Accid Anal Prev, 2013, 50 (1): 341-350.
- [5] Sun Chao, Li Jianhua, Song Yang, et al. Real-time driver fatigue detection based on eye state recognition [C]// Applied Mechanics and Materials. Trans Tech Publications, 2014, 457: 944-952.
- [6] 戴诗琪,曾智勇.基于深度学习的疲劳驾驶检测算法[J]. 计算机系统应用, 2018, 27 (7): 113-120. (Dai Shiqi, Zeng Zhiyong. Fatigue driving monitoring based on BP neural network [J]. Computer Systems & Applications, 2018, 27 (7): 113-120.)
- [7] 顾王欢,朱煜,陈旭东,等. 基于多尺度池化卷积神经网络的疲劳检测方法研究[J]. 计算机应用研究, 2019, 36(11): 3471-3475. (Gu Wanghuan, Zhu Yu, Chen Xudong, et al. Driver's fatigue detection system based on multi-scale pooling convolutional neural networks [J]. Application Research of Computers, 2019, 36(11): 3471-3475.)
- [8] 耿磊,袁菲,肖志涛,等. 基于面部行为分析的驾驶员疲劳检测方法[J]. 计算机工程, 2018, 44 (1): 274-279. (Geng Lei, Yuan Fei, Xiao Zhitao, et al. Driver Fatigue Detection Method Based on Facial Behavior Analysis [J]. Computer Engineering, 2018, 44 (1): 274-279.)
- [9] Chuang C H, Huang C S, Ko L W, et al. An EEG-based perceptual function integration network for application to drowsy driving[J]. Knowledge-Based Systems, 2015, 80: 143-152.
- [10] W. Zheng et al., "Vigilance Estimation Using a Wearable EOG Device in Real Driving Environment," in IEEE Transactions on Intelligent Transportation Systems, Jan. 2020, vol. 21, no. 1, pp. 170-184.
- [11] Sun Chao, Li Jianhua, Song Yang, et al. Real-time driver fatigue detection based on eye state recognition[C]// Applied Mechanics and Materials. Trans Tech Publications, 2014, 457: 944-952.
- [12] Tang Xinxing, Zhou Pengfei, Wang Ping. Real-time image-based driver fatigue detection and monitoring system

-
- for monitoring driver vigilance[C]//Proc of the 35th Chinese Control Conference (CCC). 2016: 4188-4193.
- [13] Zhang F, Su JJ, Geng L, et al. Driver fatigue detection based on eye state recognition. International Conference on Machine Vision and Information Technology.Singapore. 2017. 105-110.
- [14] Yan C, Zhang BL, Coenen F. Driving posture recognition by convolutional neural networks. Proceedings of the 11th International Conference on Natural Computation. Zhangjiajie, China. 2015. 680-685.
- [15] Zhu XM, Zheng WL, Lu BL, et al. EOG-based drowsiness detection using convolutional neural networks. International Joint Conference on Neural Networks. Beijing, China. 2014. 128-134.
- [16] Zhang Kaipeng, Zhang Zhanpeng, Li Zhifeng, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23 (10): 1499-1503.
- [17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot Multi-Box detector," in Proc. ECCV, pp. 21-37, 2016.