

DOI:10.19651/j.cnki.emt.2209205

基于 SSD 的多因素融合的驾驶疲劳检测研究*

吕秀丽 刘希凤 白永强

(东北石油大学物理与电子工程学院 大庆 163318)

摘要: 为了降低因疲劳驾驶而导致事故发生率,提出一种利用卷积神经网络与人脸特征点、疲劳判定指标相融合的方法,共同构建疲劳驾驶检测模型。首先利用 SSD 网络定位驾驶员的眼睛与嘴巴区域,VGG16 网络学习这两个区域所包含的疲劳特征;同时再结合人脸 68 特征点、眼睛纵横比(EAR)和嘴巴纵横比(MAR)共同判定驾驶疲劳状态。最后,在相同测试集下分别计算 SSD 算法和 Faster-RCNN 算法的平均精度均值 mAP;在 YawDD 数据集上应用此模型;并通过模拟驾车环境来验证此模型的可行性。实验结果表明,SSD 算法要优于 Faster-RCNN 算法,并且此模型在 YawDD 数据集上的检测准确率约达 97.2%,摄像头也能对驾驶员的状态进行实时检测。此模型对疲劳状态的检测十分有效,可在一定程度上降低因疲劳驾驶而导致事故发生率。

关键词: SSD 网络;疲劳驾驶检测;人脸 68 特征点;眼睛纵横比;嘴巴纵横比

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Research on driving fatigue detection based on SSD multi-factor fusion

Lyu Xiuli Liu Xifeng Bai Yongqiang

(College of Physics and Electronic Engineering, Northeast Petroleum University, Daqing 163318, China)

Abstract: In order to reduce the incidence of accidents caused by fatigue driving, a method is proposed to build a fatigue driving detection model by integrating convolutional neural network with face feature points and fatigue indicators. Firstly, the driver's eyes and mouth areas are located by the SSD network, and the VGG16 network learns the fatigue features contained in the eye and mouth areas. At the same time, 68 feature points of face, eye aspect ratio and mouth aspect ratio are combined to determine the driving fatigue state. Finally, the mean average precision of SSD algorithm and Faster-RCNN algorithm is calculated under the same test set. The model is applied to YawDD dataset. And the feasibility of this model is verified by simulating driving environment. The experimental results show that SSD algorithm is better than Faster-RCNN algorithm, the detection accuracy of this model on YawDD dataset is about 97.2%, and the camera can also detect the driver's state in real-time. The model is effective in detecting fatigue state and reducing the accident rate caused by fatigue driving to a certain extent.

Keywords: SSD network; fatigue driving detection; face 68 feature points; EAR; MAR

0 引言

随着国家综合实力的增强,人民的生活水平大幅度提高,私家车的数量也日益增加,汽车保有量从 1.61 亿辆增加至 2.61 亿辆^[1],而机动车数量的增加,在方便了人们出行的同时,也导致了交通事故发生率的增加。根据世界卫生组织的报告,交通事故是损害人生命的十大原因之一^[2]。文献[3]中表示全球因交通事故导致的死亡人数已经快要突破 122 万人,受伤人数为 2 000~5 000 万人。其中,因疲劳驾驶而造成的交通事故约占 20%~30%^[4]。交通事故引起的道路安全问题已经成为全世界急需解决的重要难题

之一。因而,为避免事故的发生,对驾驶员进行疲劳检测是十分有必要的。

目前,疲劳驾车检测方式主要包括 3 大类:基于驾驶者的生理特征^[5-7]、基于驾驶车辆状况^[8-10]、以及基于驾驶者面部特征^[11-13]。基于驾驶者生理特征的检测方式要求驾驶者密切接触检测装置,从而影响驾车操作,以致造成不便;基于驾驶车辆状况的检测则受道路情况以及个人的开车习惯等影响,容易造成检测错误;基于驾驶者面部特征的检测,无需检测装备与驾驶者直接接触,在保证一定的准确性和实时性条件下,此方法更容易推广。所以,本文选择根据

收稿日期:2022-03-08

* 基金项目:黑龙江省自然科学基金(LH2019D006)项目资助

驾驶者面部特征的检测方式对驾驶者是否疲劳做出检测。

阮志强等^[11]利用深度学习,提出一种头部姿态参数的估计方法。使用级联 CNN 检测人脸后,再通过 CNN 的校准网络校准人脸区域。赵雪鹏等^[12]提出了一种基于级联卷积 CNN 的疲劳检测算法,第 1 级提取人眼,第 2 级检测人眼特征点,根据眼睛状态参数构建疲劳特征模型,进而判定疲劳状态。

目前,深度学习被广泛应用于许多领域,并取得了不错的成效,特别是在目标检测领域。针对实际问题,出现了许多基于深度学习的目标检测算法,如缺陷检测^[14]、人脸识别^[15]、交通标志识别^[16]等,疲劳检测方面也引入了许多行之有效的图像处理算法。由于基于单一特征的疲劳检测,无法准确识别驾驶员状态,因此本文在基于卷积神经网络的基础上,融合多项疲劳判定指标,共同对疲劳状态进行判定,以达到能实时检测驾驶员是否疲劳的目的,从而减少因疲劳驾驶导致的交通事故。

1 卷积神经网络

1.1 Faster-RCNN 网络

Faster-RCNN^[17]是由 RCNN 和 Fast-RCNN 不断改进和发展而来的,通过构建区域建议网络(region proposal network,RPN)选取候选框,取代了之前的选择性搜索算法,不仅将 RPN 与 Fast-RCNN 网络整合到同一个网络中实现了端到端的训练,而且大大提高了检测速度。图 1 为 Faster-RCNN 检测算法流程。

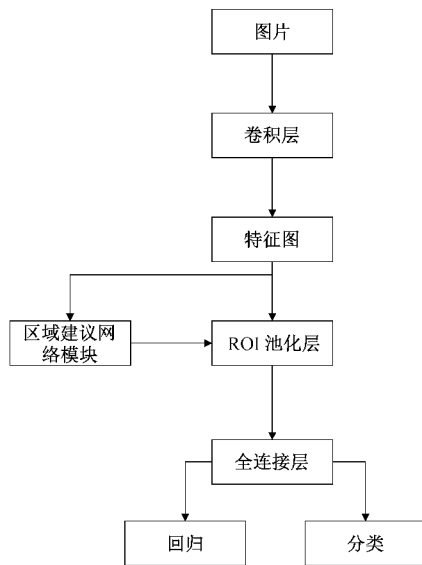


图 1 Faster-RCNN 检测算法流程

1.2 SSD 网络

SSD(single shot multibox detector)^[18]是一种端到端的基于回归的 One-stage 目标检测算法。在数据输入之后,卷积神经网络可针对不同尺度的目标,直接预测目标的位置、确定其类别、计算置信度。最后使用非极大值抑制算

法(non-maximum suppression,NMS)综合多层特征图的信息来给出检测结果。

1) 主干网络——VGG16

VGG 具有优秀的特征提取能力,本文选取 VGG16 作为主干网络。VGG16 有 13 个卷积层(convolution),5 个最大池化层(max pooling)以及 3 个全连接层(fully connected)。卷积层卷积核的大小为 3×3 ,用来增加通道数量;池化层大小为 2×2 ,用来减小特征图尺寸,增强网络的抗干扰能力;3 个全连接层中的前两层通道数为 4 096,第 3 层通道数为 1 000。VGG16 的网络框架图如图 2 所示。

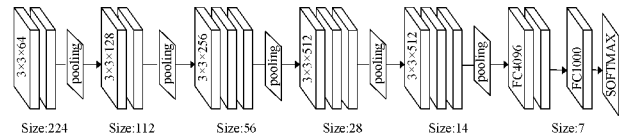


图 2 VGG16 网络结构图

2) SSD 网络结构

SSD 在原 VGG16 基础上,作了如下修改:

(1)为了在网络末端增加特征层,将原 VGG16 的两个全连接层 FC6 和 FC7 修改为卷积层 Conv6 和 Conv7,并去除 FC8 层;

(2)在网络的末端增加额外的特征层(extra feature layers):Conv8_2、Conv9_2、Conv10_2、Conv11_2。

为了适应不同尺度的检测目标,这 6 个特征图有着不同的大小。SSD 通过提取这 6 个特征图的输出来进行分类和位置回归。SSD 网络结构如图 3 所示。

2 SSD 网络训练处理

2.1 数据集的制作

本文采用自制数据集(共 2 914 张图片)对 SSD 网络进行训练,数据集的格式制作作为标准的 VOC 数据集格式。首先利用标注工具 LabelImg 对图像中的眼睛和嘴巴状态分别进行标注^[19]。标注类别分为四种,open_eye、closed_eye 表示眼睛的睁开与闭合,open_mouth、closed_mouth 表示嘴巴的张开与关闭。其中,open_eye、closed_mouth 代表正常状态;closed_eye、open_mouth 代表驾驶员处于疲劳状态。标注完成并保存后,与图片对应的 XML 格式的标注文件也被生成。

2.2 先验框的生成

依据 Faster-RCNN 中的 Anchor 概念,SSD 在 6 个不同尺度的特征图生成预选框,其中特征图 Conv4_3(38×38)、Conv10_2(3×3)、Conv11_2(1×1)都设置匹配了 4 个先验框,Conv7(19×19)、Conv8_2(10×10)、Conv9_2(5×5)都设置匹配了 6 个先验框,总共生成 8 732 个先验框,进而估计先验框的偏移量和类别置信度。

2.3 先验框匹配

在训练过程中,首先要明确与真实框进行匹配的是哪

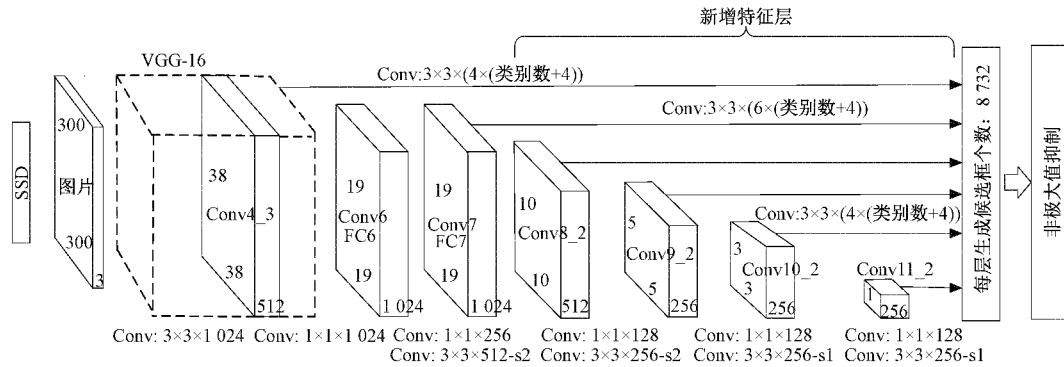


图 3 SSD 网络结构

一个先验框,然后由这个先验框所对应的边界框来对训练图片做出预测。

先验框匹配原则:对训练图片中的每一个真实框,先寻找那个和它的交并比(intersection over union, IOU)值最高的先验框加以匹配;对其他未匹配上的先验框,若某一个真实框和它的 IOU 值超过了某一阈值 T_1 , 那么它俩也进行匹配。

候选正样本是与真实框匹配的先验框,候选负样本则相反。一般来说,负样本数量远超正样本数量,为了保证平衡,一般将 T_1 设为 0.5。

2.4 损失函数

总体损失函数定义为置信度误差(conf)和位置误差(loc)的加权和^[20]。

$$L(x, c, l, g) = \frac{1}{N} [L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)] \quad (1)$$

其中, x 为先验框与真实框的匹配情况; c 为类别置信度预测值; l 为位置预测值; g 为真实框的位置参数; N 为先验框的正样本数量; α 用于调整置信度误差和位置误差之间的比例, $\alpha = 1$ 。

置信度误差采用 Softmax 损失^[20]:

$$L_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0) \quad (2)$$

其中, $\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$ i 为预测框序号, j 为真实框序

号, p 为类别序号。 $x_{ij}^p \log(\hat{c}_i^p)$ 表示预测框 i 与真实框 j 关于类别 p 匹配, 则 p 的预测概率越高, 损失越小。

位置误差采用 smooth_{L1} 损失^[20]:

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{Pos}} \sum_{m \in \{x, y, w, h\}} x_{ij}^k \text{smooth}_{L1}(l_i^m - \hat{g}_j^m) \quad (3)$$

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases} \quad (4)$$

其中, x_{ij}^k 表示第 i 个预测框与第 j 个真实框关于类别 k 是否匹配; l_i^m 表示预测框, \hat{g}_j^m 表示真实框。

2.5 非极大值抑制

非极大值抑制, 简称为 NMS 算法, 其思想是搜索局部最大值, 抑制极大值。在目标检测的过程中, 大量的候选框将会在同一目标的位置上生成, 这些候选框之间彼此交叠, 此时就需要利用 NMS 算法找到最佳的目标边界框, 去除冗余的边界框。

2.6 训练参数

本文实验在 Python 环境下, 使用 Pytorch 深度学习框架。模型训练的参数设定如下: 初始学习率设为 0.000 5, $\text{gamma} = 0.1$, $\text{momentum} = 0.9$, $\text{weight_decay} = 0.000 5$, 批量大小为 8, 共迭代 120 000 次。

2.7 SSD 训练及测试结果

将自制数据集按照 8 : 2 的比例分为训练集和测试集, 首先将训练集输入到网络中, 并按照训练参数进行训练; 然后将测试集图片输入到训练好的模型中进行测试。得出的训练和测试结果如表 1 所示, 检测准确率均超过了 97%。

表 1 训练和测试结果

数据集	数量/张	用时/s	FPS	准确率
训练集	2 331	47.556 978	49.014 889	0.973 116
测试集	583	14.114 160	41.306 036	0.970 456

3 疲劳判定指标

3.1 人脸特征点

人脸特征点定位也被称为人脸对齐, 是人脸识别过程中最为关键的一步。为提高模型的准确性, 本文利用 Dlib 官方训练好的模型进行人脸 68 点位置标定, 位置标识如图 4 所示, 43~48 点标志右眼睛, 49~68 点标志嘴巴。

3.2 眼睛纵横比

眼睛纵横比(eye aspect ratio, EAR)是指眼睛的宽度和长度之比, 用来衡量估计眼睛的睁开与闭合状态。当眼睛睁开时, EAR 值较大, 尽管 EAR 的值会随眼睛大小的不同而有所变化, 但当眼睛闭上时, EAR 值都将会趋近于 0。

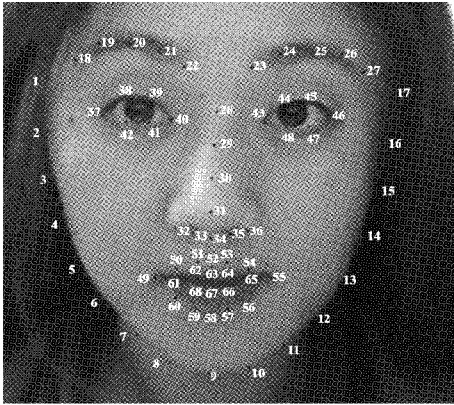


图 4 人脸 68 点位置标识

首先利用 SSD 网络定位眼部轮廓,其次利用几何方法计算 EAR 的值,以此来评估驾驶员的眼睛状态。眼部特征点如图 5 所示,利用眼部特征点 p_1 、 p_5 、 p_2 、 p_4 的纵坐标, p_0 、 p_3 的横坐标进行计算。

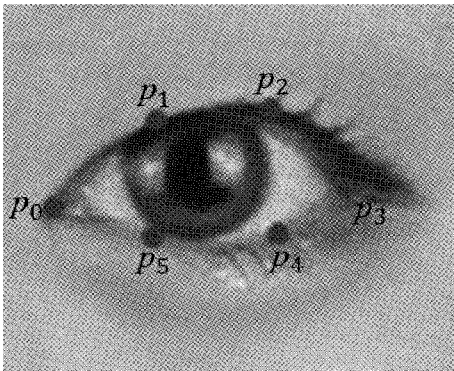


图 5 眼部特征点模型

EAR 的计算公式如下:

$$EAR = \frac{\|p_1 - p_5\| + \|p_2 - p_4\|}{2 \|p_0 - p_3\|} \quad (5)$$

经过多次试验发现,当 EAR 的值设为 0.25 时,检测准确率最高,可准确判定眼睛的张开与闭合状态。因此,当 EAR 小于最佳阈值(即 $T_2 = 0.25$) 时,判定眼睛为闭合状态。

3.3 嘴巴纵横比

借鉴眼部的 EAR 值法,通过嘴部纵横比(mouth aspect ratio, MAR)来衡量驾驶员的嘴部状态,主要检测驾驶员是否打哈欠,进而判断疲劳程度。

首先利用 SSD 网络定位嘴部轮廓,其次利用几何方法计算 MAR 的值,以此来评估驾驶员的嘴部状态。嘴部特征点如图 6 所示。利用特征点 m_{13} 、 m_{19} 、 m_{14} 、 m_{18} 、 m_{15} 、 m_{17} 的纵坐标进行计算。

MAR 的计算公式如下:

$$MAR = \frac{\|m_{13} - m_{19}\| + \|m_{14} - m_{18}\| + \|m_{15} - m_{17}\|}{3} \quad (6)$$

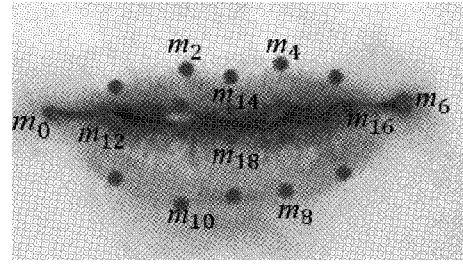


图 6 嘴部特征点模型

同样经过多次试验发现,当 MAR 的值设为 17 时,能有效区分驾驶员的说话、打哈欠状态,检测准确率最高。因此,当 MAR 大于最佳阈值(即 $T_3 = 17$) 时,判定嘴巴为张开状态,即存在打哈欠行为。

4 实验结果与分析

4.1 评价指标

精确率(Precision)衡量的是分类器预测出来的所有正样本中,实际正样本所占的概率。

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

其中, TP 表示“分类器预测为正样本,实际它确实是正样本”; FP 表示“分类器预测为正样本,但实际它是负样本”。

召回率(Recall)衡量的是分类器找出所有正样本的概率。

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

其中, FN 表示“分类器预测为负样本,但实际它是正样本”,即是一个未被检测出来的真实框。

以 Precision 为 y 轴, Recall 为 x 轴所画的曲线,为 PR 曲线;对 PR 曲线上的 Precision 值进行积分,就可以得到平均精度值(average precision, AP);最后计算出 open_eye、closed_eye、open_mouth、closed_mouth 这 4 个类别的 AP 值,再取其平均值,就得到平均精度均值(mean average precision, mAP)。

图 7、8 分别为基于 Faster-RCNN 模型和本文 SSD 模型在测试集上的 mAP 值。

实验结果表明,在使用相同测试集进行实验的情况下,SSD 模型的 mAP 值为 95.10%, Faster-RCNN 模型的 mAP 值为 77.24%。与 Faster-RCNN 模型相比,SSD 模型在 mAP 值上提升了 17.86%。其中,在 open_mouth 类别识别上 AP 值提升了 1%,在 closed_mouth 类别识别上 AP 值提升了 15%,在 closed_eye 类别识别上 AP 值提升了 26%,在 open_eye 类别识别上 AP 值提升了 29%。整体来看,SSD 模型要优于 Faster-RCNN 模型。

4.2 疲劳检测模型实验结果

疲劳检测模型是将多项疲劳判定指标结合到训练好

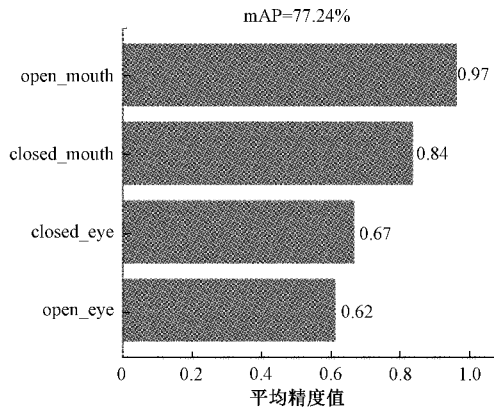


图 7 Faster-RCNN 模型

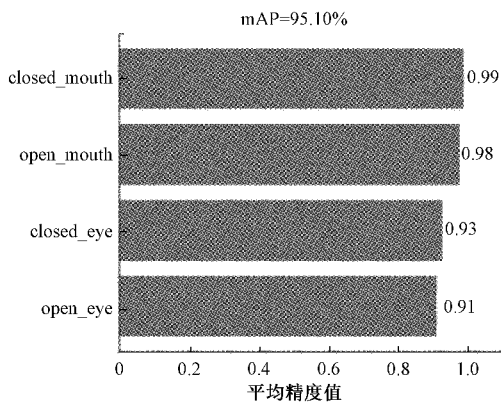


图 8 SSD 模型

的 SSD 模型中,以达到多指标融合共同检测的目的。每次检测到眼睛闭合时,显示屏显示“Eyes Closed!”;嘴巴张开时,显示屏显示“Yawning!”;以此来提示驾驶员要进行适当的休息,以避免事故的发生。

YawDD 数据集是由安装在汽车后视镜和仪表盘上的摄像头拍摄所得^[21]。该数据集是由来自不同国家、不同年龄的志愿者(63 名男性、56 名女性),在不同的光照条件下拍摄下来的。为了评估该模型的准确性,本文选取由安装在仪表盘上的摄像头所拍摄的 29 条视频,包括 13 名女性志愿者与 16 名男性志愿者,每段视频选取 10 张视频帧进行验证,共计 290 张视频帧。表 2 为疲劳检测模型在 YawDD 数据集上的检测结果,此模型在 YawDD 数据集上的准确率约达 97.2%。

表 2 疲劳检测模型在 YawDD 数据集上的检测结果

分类	数量/张	正确预测/张	准确率/%
女性无眼镜	80	79	98.8
女性戴眼镜	50	48	96.0
男性无眼镜	60	59	98.3
男性戴眼镜	100	96	96.0
总计	290	282	97.2

从表 2 所测得的数据中,可得出如下两点结论:

1) 男性戴眼镜和女性戴眼镜情况下的检测准确率相似;同样的,男性不戴眼镜和女性不戴眼镜情况下的检测准确率也相似。由此可知,疲劳检测模型的检测结果不受性别的影响。

2) 在驾驶员不戴眼镜的情况下,检测准确率要略优于戴眼镜的情况。由此可知,眼镜在一定程度上虽然会影响到检测的结果,但此模型依然可以满足实际需求。

从 YawDD 数据集中选取部分检测结果图,其中图 9、10 分别为疲劳检测模型在未戴眼镜和佩戴眼镜情况下的检测结果截图。由此可看出,该模型在戴眼镜和不戴眼镜情况下均可准确识别出眼睛闭合与打哈欠行为。



图 9 未戴眼镜检测结果



图 10 戴眼镜检测结果

图 11 为摄像头实时检测的结果图,分别截取不戴眼镜与戴眼镜情况下的正常、闭眼及打哈欠状态下的图片。

当驾驶员闭眼时,表明此时 EAR 的值小于阈值 T_2 (即 0.25),显示屏按预期设定,显示“Eyes Closed!”字样;当驾驶员打哈欠时,表明此时 MAR 的值大于阈值 T_3 (即 17),显示屏按预期设定,显示“Yawning!”字样。该模型可准确识别驾驶员的疲劳状态,符合预期设想。

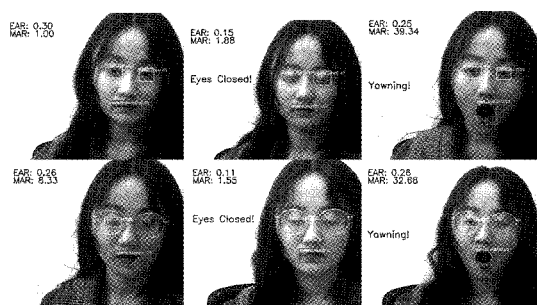


图11 摄像头实时检测结果图

5 结 论

本文基于驾驶员面部特征来进行驾驶疲劳检测,提出了一种基于卷积神经网络的多特征融合的检测方法。结果表明,本文提出的SSD算法在一定程度上克服了一般算法难以检测较小物体,并且定位不精准的缺点;同时本文结合多项疲劳判定指标,能更准确的预测眼睛与嘴巴的位置,可有效判定驾驶员的疲劳状态,有助于降低因疲劳而引起的交通事故的发生率,具有较好的发展前景与研究价值,可应用于实际。

参考文献

- [1] 李伟芯. 以人为本 夯实基础 推进“十四五”道路管理科技发展——访公安部道路交通安全研究中心主任、中国道路交通安全协会会长王长君研究员[J]. 道路交通管理, 2021(3): 28-31.
- [2] AMODIO A, ERMIDORO M, MAGGI D, et al. Automatic detection of driver impairment based on pupillary light reflex [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(8): 3038-3048.
- [3] KHAN M Q, LEE S. A comprehensive survey of driving monitoring and assistance systems [J]. Sensors, 2019, 19(11), DOI:10.3390/s19112574.
- [4] ALVARO P K, BURNETT N M, KENNEDY G A, et al. Driver education: Enhancing knowledge of sleep, fatigue and risky behaviour to improve decision making in young drivers[J]. Accident, Analysis and Prevention, 2018, 112: 77-83.
- [5] 王斐, 吴仕超, 刘少林, 等. 基于脑电信号深度迁移学习的驾驶疲劳检测[J]. 电子与信息学报, 2019, 41(9): 2264-2272.
- [6] TUNCER T, DOGAN S, SUBASI A. EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection [J]. Biomedical Signal Processing and Control, 2021, 68(2), DOI: 10.1016/j.bspc.2021.102591.
- [7] 徐礼胜, 张闻勋, 庞宇轩, 等. 基于短时心电信号的疲劳驾驶检测算法[J]. 东北大学学报(自然科学版), 2019,

40(7): 937-941.

- [8] 蔡素贤, 杜超坎, 周思毅, 等. 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(4): 77-82.
- [9] 石晓玲, 杨英仓. 基于车辆行驶轨迹特征的驾驶人疲劳状态识别技术研究[J]. 中国新通信, 2017, 19(2): 158-160.
- [10] MCDONALD A D, LEE J D, SCHWARZ C, et al. A contextual and temporal algorithm for driver drowsiness detection [J]. Accident Analysis & Prevention, 2018, 113: 25-37.
- [11] 阮志强, 方向忠. 基于深度学习的驾驶员头部姿态参数估计[J]. 信息技术, 2017(2): 97-101, 105.
- [12] 赵雪鹏, 孟春宁, 冯明奎, 等. 基于级联卷积神经网络的疲劳检测[J]. 光电子·激光, 2017, 28(5): 497-502.
- [13] 李昭慧, 张玮良. 基于改进YOLOv4算法的疲劳驾驶检测[J]. 电子测量技术, 2021, 44(13): 73-78.
- [14] 杨珂, 方诚, 段黎明. 基于深度学习模型融合的铸件缺陷自动检测[J]. 仪器仪表学报, 2021, 42(11): 150-159.
- [15] 赵梓涵, 李东新. 基于改进SRCNN算法的人脸识别技术研究[J]. 国外电子测量技术, 2020, 39(12): 74-79.
- [16] 吕禾丰, 陆华才. 基于YOLOv5算法的交通标志识别技术研究[J]. 电子测量与仪器学报, 2021, 35(10): 137-144.
- [17] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [18] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [19] 黄生鹏, 范平清. 基于深度学习的实时车辆检测研究[J]. 软件工程, 2021, 24(1): 13-16, 5.
- [20] 张雪芹, 魏一凡. 基于深度学习的驾驶场景关键目标检测与提取[J]. 华东理工大学学报(自然科学版), 2019, 45(6): 980-988.
- [21] ABTAHI S, OMIDYEGANEH M, SHIRMOHAMMADI S, et al. YawDD: A yawning detection dataset[C]. ACM Multimedia Systems(MMSys), ACM, 2014.

作者简介

吕秀丽, 教授, 主要研究方向为数字图像处理、生物特征识别技术、数字水印与信息隐藏。

E-mail: lxl_20041231@163.com

刘希凤(通信作者), 工学硕士, 主要研究方向为图像处理与模式识别。

E-mail: liuxifeng97@163.com

白永强, 教授, 主要研究方向为介观表征。

E-mail: yqbai@nepu.edu.cn