DOI:10.19651/j.cnki.emt.2106365

一种基于多因素融合的驾驶预警方法*

禹江林[1,2] 张 云[1,2]
(1.昆明理工大学 信息工程与自动化学院 昆明 650500;2.云南省计算机技术应用重点实验室 昆明 650500)

摘 要:本文提出了一种新的基于多因素融合的驾驶预警方法。首先,结合现有的疲劳评判因素,提出了一种基于多因素的危险评判标准,克服了传统单因素方法的应用的局限及易受外界干扰的缺点。其次,提出了一个基于 SSD 的检测网络,其中,用先进的 MobileNetV3 替换了主干网络 VGG,用修改的 NMS 层实现了快速目标检测,最后用新设计的多任务检测器及损失函数实现了多任务检测。在预训练权重的迁移学习后,实测的检测准确率为 95.7%,速度为 41 fps,实现了准确及实时性。

关键词:驾驶预警;目标分类;头部姿态检测;PERCLOS 值
中图分类号:TP391 **文献标识码**:A **国家标准学科分类代码**:510.4050

A driving early warning method based on multi-factor fusion

Yu Jianglin[1,2] Zhang Yun[1,2]
(1. College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;
2. Yunnan Key Lab for Computer Technology Applications, Kunming 650500, China)

Abstract: Based on multi-factor fusion, this paper proposes a new driving early warning method. First, combining the existing fatigue factors, a risk assessment criterion is proposed based on multiple factors, which overcomes traditional single-factor methods' limitations in application and shortcomings of being vulnerable to external interference. Secondly, a detection network is proposed based on SSD, in which the backbone network VGG is replaced with the advanced MobileNetV3, fast target detection is achieved with a modified NMS layer, and finally multi-task detection is realized with newly designed multi-task detectors and loss functions. After transfer-learning with the pre-training weights, the tested detection accuracy rate is 95.7%, and the speed is 41 fps, which is accurate and real-time.

Keywords: driving early warning; object classification; head pose detection; PERCLOS

0 引 言

驾驶预警技术是计算机视觉领域的一个重要的研究课题,它在自动驾驶、智慧城市等当今社会生活的多个方面有着广泛的应用价值。

目前的危险驾驶行为预警方法主要有如下 3 种:1)根据检测驾驶员生理参数进行预警,如 Jose 等[1]和 Luo 等[2]通过检测脑电信息(EEG),Markus 等[3]通过检驾驶员的心电信号(ECG),该类方法通过接触式检测结果准确可靠但却容易影响驾驶员正常操作,因检测设备较为专业且过于昂贵无法大规模普及。2)根据驾驶行为进行预警,如 Chai 等[4]根据驾驶员操作方向盘的方向角变化,张明明[5]通过检测操作方向盘的力度大小,该类方法具有所需成本低、检测速度快的优点,但容易因驾驶员个体的不同,车辆路况的不同使准确率受到干扰。3)根据驾驶员的面部特征进行预警,如 Xu 等[6]通过追踪眼球运动;田璐萍等[7]通过眼部信息进行检测;Knapik 等[8]通过检测嘴部位置然后跟踪嘴部是否存在打哈欠状况,该类方法通常由多个模型串联组成,容易造成误差的累积,或因单因素判断不够全面,易受到外界因素的干扰。

针对上述问题,本文提出了一个更加全面的用于评判危险驾驶的标准,同时基于 SSD 网络设计了一个端到端的多任务融合的结构模型,实测也表明该方法的检测准确率为 95.7%,速度为 41 fps,具有准确及实时性。

1 多任务融合框架

卷积网络在目标检测[9-10]与目标分类[11-12]任务中有广泛的应用,本文通过使用更先进的检测网络 MobileNetV3[13]

收稿日期:2021-04-12
* 基金项目:国家自然科学基金项目(61262043)、云南省科技计划项目(2011FZ029)资助

替换 SSD[14] 原主干网络 VGG[15]，使速度提高了 21.7%，准确率提高 5.9%；对多个模块串联的传统方法进行了改进，如图 1 所示多模块的串联不仅需要对图片进行重复的特征提取耗时较长，且多个模块容易形成误差累积，更不便于跨平台的移植。修改了末端检测器的输出和默认框(default box)的比例，使其能够在一个网络下实现对头部姿态偏转角和眼睛嘴巴张开状态与闭合状态的同步检测，框架如图 2 所示，鉴于现存的面部姿态检测算法均先对面部进行精确定位再对面部图像提取特征两个步骤构成，因此本文把面部置信度最高的默认框的特征作为输出头部姿态检测卷积的输入，即默认面部检测得分最高的检测框对应的姿态角可信度最高，并把其作为模型的姿态角输出。由于应用场景为针对驾驶员(单人)的目标检测，且两只眼睛的动作和频率具有同步性，因此在 NMS 层省去了置信度阈值设定与计算重叠度的过程，直接输出每个类置信度得分最大的检测框作为每个类的检测输出，与置信度阈值 0.1 的检测速度相比，本文速度上提高了 40%，检测速度可达 41 fps。同时，把对眼睛和嘴巴状态的检测问题转化为目标分类问题，即把睁眼图像和闭眼图像作为两个不同的分类，嘴巴类似，在输出端根据其置信度得分把得分较高的输出作为眼睛和嘴巴的状态，省去了传统方法如根据 EAR 值判断张合度等方法后期计算的过程，在一定程度上避免了误差的累积。通过统计每分钟内眼睛和嘴巴的张开闭合情况得到眼睛和嘴巴的 PERCLOS 值，结合头部姿态的偏转角综合地判断驾驶员是否存在危险驾驶，具有实时、准确、便于移植的优点。

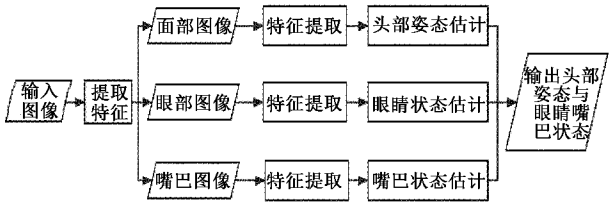

图 1　常用检测框架

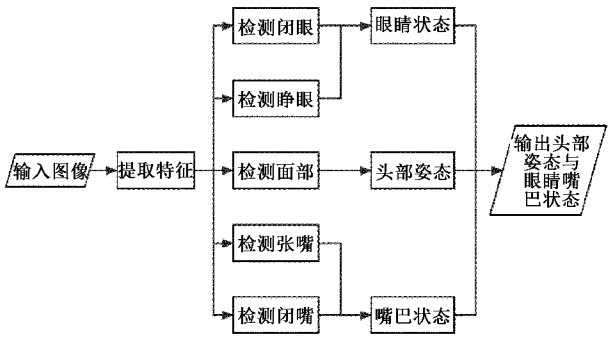

图 2　本文检测框架

2　多因素评判

在立体空间内，物体的姿态可以用如图 3 所示的 3 个欧拉角来表示：pitch(围绕 X 轴旋转)、yaw(围绕 Y 轴旋转)和 roll(围绕 Z 轴旋转)，分别称为俯仰角、偏航角和滚转角。

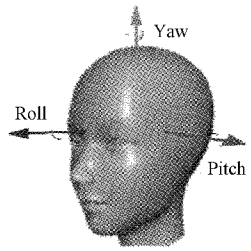

图 3　头部姿态角

PERCLOS[16] 参数表示了单位时间闭眼程度超过一定值(70%，80%或 50%)的时间相对单位总时间的比例关系如图 4 所示，可得 PERCLOS 值计算如式(1)所示。

$$\mathrm{PERCLOS}=(t_4-t_3)/(t_6-t_1)\times 100\% \tag{1}$$

PERCLOS 与疲劳程度的相关性较高，故在疲劳评判中具有广泛的应用。本文借鉴 PERCLOS 计算公式，将单位时间内闭眼时间所占总时间的比值转换为单位时间内闭眼帧数所占总帧数的比值，故闭眼的 PERCLOS 值 $perclos_{eye}$ 和张嘴的 PERCLOS 值 $perclos_{mouth}$ 可由式(2)、(3)计算得出：

$$perclos_{eye}=t_{eye}/T_{eye}\times 100\% \tag{2}$$

$$perclos_{mouth}=t_{mouth}/T_{mouth}\times 100\% \tag{3}$$

式中：t_{eye} 为单位时间内闭眼的帧数；T_{eye} 为单位时间内的总帧数。同理，t_{mouth} 为单位时间张嘴帧数；T_{mouth} 为单位时间总帧数。

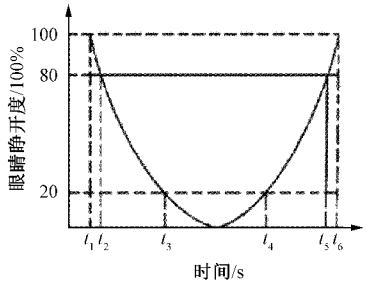

图 4　PERCLOS 示意图

国家标准[17] 中对头部姿态异常行为的定义为左右偏转(Yaw)超过 45°或上下偏转(Pitch)超过 30°且时间大于 3 s；对闭眼的定义为眼睑完全闭合超过 2 s。为了使系统更加敏感，故本文把头部偏转持续时间超过 2 s 或闭眼持续时间超过 1 s 即视为危险驾驶。文献[18]中把眼睛 PERCLOS 值大于 0.3 的情况视为轻度疲劳，大于 0.5 的视为重度疲劳，综上所述，本文的疲劳评判标准如下：

$$\begin{cases}\begin{cases}|yaw|>45^\circ\\ |roll|>30^\circ \text{ 和 } time>2\text{ s}\\ |pitch|>30^\circ\end{cases}\\ \begin{cases}|yaw|<45^\circ\\ |roll|<30^\circ \text{ 和}\begin{cases}perclos_{eye}>0.3\\ perclos_{mouth}>0.3\\ time_{close\ eye}>1\text{ s}\end{cases}\\ |pitch|<30^\circ\end{cases}\end{cases} \tag{4}$$

式中:yaw,$roll$,$pitch$ 为头部姿态的 3 个角度;$perclos_{eye}$ 为每分钟内闭眼的 PERCLOS 值;$perclos_{mouth}$ 为每分钟内张嘴的 PERCLOS 值;$time_{close\_eye}$ 表示持续闭眼的时间。

3 模型改进

模型结构方面,本文主要对 SSD 的主干网络和检测器做了改进,优化了默认框的比例和在各特征层上的分布。使用了 MobileNetV3 替换了 SSD 原主干网络 VGG,分别抽取如图 5 所示的 6 个特征层作为面部器官检测器和脸部检测器的输入。

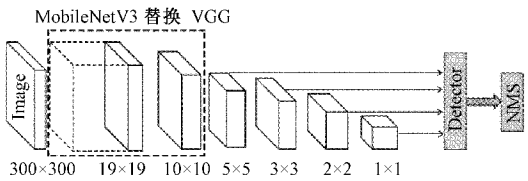

图 5 模型网络结构

对于默认框的比例设定,因前 3 层模型的尺寸较大,每个特征单元映射到输入图片的面积较小,故感受野较小的检测框适合检测眼睛嘴巴等图像面积较小的目标,后 3 层感受野较大,适合检测面部等图像面积较大的目标,故对于前 3 层特征单元生成的默认框,如图 6 所示把宽高比设为(1∶1,1∶2,2∶1,3∶1,4∶1),使检测框与眼睛和嘴巴的轮廓更加贴合,对于后 3 层的特征层,如图 7 所示把默认框的比例设为(1∶1,1∶2)主要针对面部检测。对于 1∶1 的检测框,除了默认尺寸比例的默认框之外,还有个默认尺寸开根号为尺寸比例的正方形,故倘若存在宽高比例为 1∶1 的正方形,则会具有两个纵横比为 1 但大小不同的默认框。

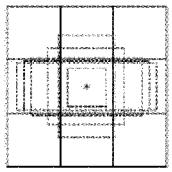

图 6 前 3 层默认框

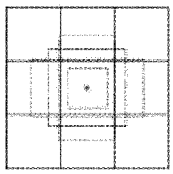

图 7 后 3 层默认框

综上所述,前 3 层特征层上每个特征单元生成 6 个默认框,后 3 层特征层上每个特征单元生成 3 个默认框,每个默认框对应一个维度的向量,模型共生成 $6\times(19\times19+10\times10+5\times5)+3\times(3\times3+2\times2+1\times1)=2\ 958$ 个向量,每个向量用来预测 6+4+3 个值,分别对应分类卷积生成的头部,睁眼、闭眼、张嘴、闭嘴和背景的 6 个值;位置回归卷积生成 cx,cy,Δw 和 Δh 的 4 个值;头部姿态回归卷积生成 yaw,$pitch$,$roll$ 的 3 个值。为减小训练时的波动,本文不直接回归目标框位置,而是回归目标框相对于默认框的位置偏差,即 cx,cy,Δw 和 Δh。鉴于以往的头部姿态检测网络都需要对面部进行精确定位后在面部图像上进行特征提取,故本文默认头部检测得分最高的检测框上得出的头部姿态值可信度最高,并把其作为模型头部姿态估计的输出。对于眼睛和嘴巴的检测框得出的姿态估计值对其置为 0。对每个特征单元生成的检测框进行预测的检测器结构如图 8 所示。

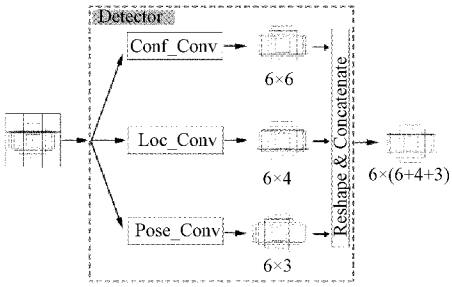

图 8 检测器结构

模型输出端的非极大抑制层(NMS)用来筛选出默认框的预测结果作为模型的输出,其本质是挑选出每个类达到阈值的目标框,并对这些框中与置信度得分最大框的重叠度超过一定值的框进行剔除,保留剩下的框。若置信度阈值设定过大或过小,会造成目标得漏检或误检,由于人的两只眼睛睁眼闭眼等生理反应的频率相同,故本模型直接输出每个类的置信度最高的检测框作为检测输出,省去了各检测框之间重叠度的计算与置信度阈值的设定,避免了因头部姿态和光线等外界因素对检测结果造成的干扰,与置信度阈值为 0.1 时的运行速度对比有了 40% 的提升。在输出前把睁眼置信度和闭眼置信度中的较大值作为眼睛状态的输出;同理把张嘴置信度和闭嘴置信度中的较大值作为嘴巴状态的输出。

对原 SSD 网络的损失函数进行扩展,针对任务的不同设置了不同的损失函数,最后进行加权求和,使其能够进行头部姿态回归的训练,本模型总体损失函数公式如下:

$$L=\frac{1}{N}(L_{conf}(c)+\alpha L_{loc}(pl,gl)+\beta L_{pose}(pa,ga)) \tag{5}$$

式中:$L_{conf}(c)$ 是置信度的损失;$L_{loc}(pl,gl)$ 为目标框的回归损失;$L_{pose}(pa,ga)$是头部姿态角的损失。α、β 用于在训练的不同阶段调整权重的参数,N 为正样本的个数。其中,$L_{conf}(c)$ 采用 softmax 损失函数,$L_{loc}(pl,gl)$ 和 $L_{pose}(pa,ga)$采用 Smooth L1 损失函数。

$L_{conf}(c)$损失用来进行面部、睁眼闭眼、张嘴闭嘴和背景分类的训练。其公式如下:

$$L_{conf}(c)=-\sum_{i\in pos}^{N}\log\left(\frac{\exp(c_i^p)}{\sum_p\exp(c_i^p)}\right)-\sum_{i\in neg}^{N}\log\left(\frac{\exp(c_i^0)}{\sum_p\exp(c_i^p)}\right) \tag{6}$$

式中：p 的值取 0～5，c_i^1 对应表示第 i 个预设框为面部的置信度，c_i^2 对应表示第 i 个预设框为睁眼的置信度，c_i^3 对应表示第 i 个预设框为闭眼的置信度，c_i^4 对应表示第 i 个预设框为张嘴的置信度，c_i^5 对应表示第 i 个预设框为闭嘴的置信度，c_i^0 表示第 i 个预设框为背景的置信度。

$L_{loc}(pl,gl)$ 用来对位置框回归进行训练，其公式如下：

$$L_{loc}(pl,gl)=\sum_{i\in pos}^{N}\sum_{m\in\{cx,cy,w,h\}}smooth_{L1}(pl^m-\hat{g}l^m) \tag{7}$$

$$\hat{g}l^{cx}=\frac{g^{cx}-d^{cx}}{d^w} \tag{8}$$

$$\hat{g}l^{cy}=\frac{g^{cy}-d^{cy}}{d^h} \tag{9}$$

$$\hat{g}l^{w}=\log\left(\frac{g^{w}}{d^{w}}\right) \tag{10}$$

$$\hat{g}l^{h}=\log\left(\frac{g^{h}}{d^{h}}\right) \tag{11}$$

式中：相比 cx 和 cy，h 和 w 增加了对数操作，确保了每个默认框是以中心点为基准，避免了检测框的大幅度漂移，使训练过程更加平稳。

$L_{pose}(pa,ga)$ 用来训练模型姿态角的回归，为解决训练数据集中的大幅度姿态偏转数据较少的问题，本文模型引入了训练集的角度权重，对大角度偏转数据产生的 loss 进行加权，公式如下：

$$L_{pose}(pa,ga)=\sum_{i\in pos}^{N}f(ga^m)\sum_{m\in pitch,yaw,roll}smooth_{L1}(pa^m,ga^m) \tag{12}$$

$$f(ga^m)=\frac{1}{3}\sum_{m\in pitch,yaw,roll}(1+\sin(ga^m)) \tag{13}$$

$f(ga^m)$ 为权重函数，训练时输入图片的姿态角 ga^m 越大，sin 值也就越大，该图片的权重值越大，加强了对大姿态角偏转图片的训练力度。

本文模型中大量运用 Smooth L1 损失函数，在训练差值较小时，其梯度也会变小，而当训练差值很大时，其梯度的绝对值达到上限值 1，避免了整个模型因为某个损失差值过大导致梯度爆炸，使训练更加平滑。

4　训练与实验

本文选取了 300W 数据集和 AFLW 人脸数据集，根据其关键点生成目标框坐标，并对其眼睛和嘴巴状态进行标注，读取数据集标注文件中的“pt2d”信息作为训练输入的头部姿态值，并把眼睛和嘴巴的姿态值置为 0，此外录制了一些实际应用场景的视频进行标注。由于 SSD 的输入为固定大小，输入的图像被统一缩放为 300×300 的正方形，为了避免非正方形图像缩放后的面部几何特征发生改变，本文统一对非正方形图像以图像中心为基准做最大内切正方形裁剪，并对坐标信息进行相应的转换，生成的训练集和验证集图像如图 9 所示。

图 9　数据集图片标注

对于眼睛嘴巴的分类标准，国标中对打哈欠嘴巴张开的定义为高宽比大于 0.6，对闭眼的定义为上下眼皮完全闭合，本文参照其对数据集进行标注，把嘴巴高宽比近似大于 0.6 且能看到口腔内的图像标记为张嘴，其他的标记为闭嘴；对眼睛完全闭合或近似闭合且看不到眼仁的图像标记为闭眼，其他的情况标记为睁眼。眼睛嘴巴状态的分类示例图如图 10 所示。

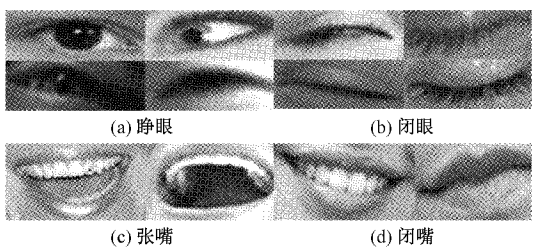

图 10　眼睛嘴巴分类示意

通过使用训练好的 MobileNetV3 权重进行迁移学习，初始学习率为 0.001，batch size 设为 64，正负样本比例设为 1∶3，采用指数衰减法逐渐减小学习率。训练分为两个阶段，先在 300W 和 AFLW 组成的训练集上对检测、分类、头部姿态回归等任务损失进行全面训练，120 epoch 后在自己标注的数据集上对分类与检测损失进行微调，初期训练曲线如图 11 所示。

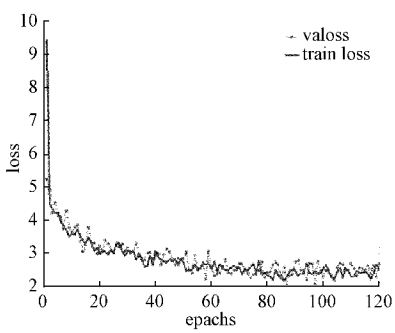

图 11　训练初期 loss 曲线

由图可以看出，当训练到 60 个 epochs 后，loss 基本趋于平稳，可见在更改了输出端检测器之后，模型的训练依然可以稳定收敛，且抽取部分数据集做消融实验，更改检测器前的检测正确率为 0.922，添加姿态角回归卷积后在相同的数据集上检测的正确率为 0.929，证明了添加头部姿态的卷积模块输出不会降低模型检测的正确率，其在验证集上的角度误差值如表 1 所示。

表 1 验证集各角度误差均值

数据集	Yaw	Pitch	Roll	Mean
AFLW	5.03	5.75	4.87	5.21
300W	4.87	5.29	4.53	4.89

实验发现,在 300W 验证集的预测结果要优于 AFLW,这是由于数据集 AFLW 图像包含了更多的头部姿态变化和光线的变化,而 300W 数据集图片所包含的姿态偏转角度较小,故在数据集 300W 上的结果要好。

对于小汽车和公交车等不同的应用场景下,当驾驶员偏转头部或佩戴眼镜时,本文模型均能实现较好的检测效果,即使是小部分的面部遮挡和光线亮暗发生变化,依然能够对目标实现准确的定位,具有较好的鲁棒性,检测效果如图 12 所示。随后本文抽取 YawDD 数据集中若干真实驾驶场景的视频片段,结合作者录制的视频,以眼睛持续两帧以上的闭眼状态作为一次眨眼次数,嘴巴持续两帧以上的张嘴状态作为一次打哈欠次数,人工统计出视频人物的眨眼次数与打哈欠次数,以及头部姿态异常的次数,分别与模型检测的结果作比较,结果如表 2 所示。

图 12 实际应用场景检测效果

表 2 视频片段检测结果

片段 ID	闭眼次数	检测次数	眨眼测全率/%	打哈欠次数	检测次数	打哈欠测全率/%	缓慢眨眼次数	姿态异常次数	报警次数
1	59	57	96.61	5	5	100	2	3	5
2	63	60	95.23	2	2	100	2	0	2
3	71	68	95.77	3	3	100	2	2	4
4	51	49	96.07	1	1	100	0	3	3
5	39	37	94.87	1	1	100	0	5	5
6	26	25	96.15	4	4	100	0	5	5

由表 2 可以看出,本文模型对打哈欠时的测全率为 100%,且对于头部姿态异常时的情况均能检测出来并报警予以提示,眨眼的漏检率低于 5%,漏检原因主要为头部姿态大幅偏转造成目标的丢失,通过后期增加大姿态图像训练可以进一步提高准确率。在单卡 2080Ti 上的检测速度为 41 fps,满足了实时检测的需求。

5 结 论

本文模型结合了现有的不同的疲劳评判因素提出了一种新的多因素融合的疲劳评判方法。改进了原有的 SSD 网络结构,并且修改了输出端检测器和 NMS 层,实现了目标检测、目标分类、头部姿态回归等多任务的同步进行。另外将眼睛和嘴巴的状态检测转化为分类问题,避免了传统多任务模块串联方法的误差累积和特征重复提取。在实际应用场景下的检测准确率达 95.1%,单卡检测速度可达 41 fps,具有较好的应用价值。

参考文献

[1] JOSE M M, CAROLINA D P, HECTOR R, et al. Monitoring driver fatigue using a single-channel electroencephalographic device: A validation study by gaze-based, driving performance, and subjective data[J]. Accident Analysis and Prevention, 2017, 109: 62-69.

[2] LUO H W, QIU T R, LIU C, et al. Research on fatigue driving detection using forehead EEG based on adaptive multi-scale entropy[J]. Biomedical Signal Processing and Control, 2019, 51: 50-58.

[3] MARKUS G, DAVID S, NATIVIDAD M M, et al. ECG sensor for detection of driver's drowsiness[J]. Procedia Computer Science, 2019, 159: 1938-1946.

[4] CHAI M, LI S W, SUN W C, et al. Drowsiness monitoring based on steering wheel status [J]. Transportation Research Part D: Transport and Environment, 2019, 66: 95-103.

[5] 张明明. 基于方向盘握力的疲劳驾驶检测研究[D]. 镇江:江苏大学, 2016.

[6] XU J L, MIN J L, HU J F. Real-time eye tracking for the assessment of driver fatigue[J]. Healthcare Technology Letters, 2018, 5(2): 54-58.

[7] 田璐萍, 稀启春. 基于眼部信息融合的疲劳驾驶检测的研究[J]. 国外电子测量技术, 2019, 38(10): 26-29.

[8] KNAPIK M, CYGANEK B. Driver's fatigue recognition based on yawn detection in thermal images [J].

Neurocomputing, 2019, 338: 274-292.
[9] 白中浩，朱磊，李智强. 多模型融合和重新检测的高精度鲁棒目标跟踪[J]. 仪器仪表学报，2019,40(9): 132-141.
[10] 李智伟，杨亚莉，钟卫军,等. 基于改进的 SSD 模型手机违规使用目标检测[J]. 电子测量与仪器学报，2021，35(1):120-127.
[11] 李钊光. 基于深度学习和迁移学习的体育视频分类研究[J]. 电子测量技术，2020，43(18):21-25.
[12] 王永浩. 基于加权结构 SVM 的钢板表面缺陷分类[J]. 电子测量技术，2020，43(11):69-73.
[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. European Conference on Computer Vision, (ECCV), 2016: 21-37.
[14] CHU X X, ZHANG B, XU R J. MoGA: Searching beyond MobileNetV3[C]. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP) Barcelona, Spain, 2020: 4042-4046.
[15] KAREN S Y, ANDREW Z. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014, 9: 1-14.
[16] WANG J J, XU W, GONG Y H. Real-time driving danger-level prediction[J]. Engineering Applications of Artificial Intelligence, 2010, 23(8): 1247-1254.
[17] 汽标委智能网联汽车分标委. 驾驶员注意力监测系统性能要求及试验方法[EB/OL]. (2020-12-01)[2021-06-01]. http://www.doc88.com/p-90329250373419.html.
[18] 闫保中，王晨宇，王帅帅. 基于人眼特征的疲劳驾驶检测技术研究[J]. 应用科技，2020，47(1): 47-54.

作者简介

禹江林，硕士研究生，主要研究方向为图像处理。
E-mail:184333195@qq.com

张云(通信作者)，博士，教授，硕导，主要研究方向为图像处理。
E-mail:zhangyun92x@163.com