



## 视觉协同的违规驾驶行为分析方法

高尚兵, 黄子赫, 耿璇, 臧晨, 沈晓坤

引用本文:

高尚兵, 黄子赫, 耿璇, 等. 视觉协同的违规驾驶行为分析方法[J]. *智能系统学报*, 2021, 16(6): 1158–1165.

GAO Shangbing, HUANG Zihe, GENG Xuan, et al. A visual collaborative analysis method for detecting illegal driving behavior[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(6): 1158–1165.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202101024>

## 您可能感兴趣的其他文章

### 面向自动驾驶目标检测的深度多模态融合技术

Deep multi-modal fusion in object detection for autonomous driving

*智能系统学报*. 2020, 15(4): 758–771 <https://dx.doi.org/10.11992/tis.202002010>

### 深度学习的双人交互行为识别与预测算法研究

Human interaction recognition and prediction algorithm based on deep learning

*智能系统学报*. 2020, 15(3): 484–490 <https://dx.doi.org/10.11992/tis.201812029>

### 基于车内外视觉信息的行人碰撞预警方法

Pedestrian collision warning system based on looking-in and looking-out visual information analysis

*智能系统学报*. 2019, 14(4): 752–760 <https://dx.doi.org/10.11992/tis.201801016>

### 多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene

*智能系统学报*. 2019, 14(2): 306–315 <https://dx.doi.org/10.11992/tis.201710019>

### 自动驾驶场景下小且密集的交通标志检测

Detection of small and dense traffic signs in self-driving scenarios

*智能系统学报*. 2018, 13(3): 366–372 <https://dx.doi.org/10.11992/tis.201706040>

### 智能手机车辆异常驾驶行为检测方法

Abnormal driving behavior detection based on the smart phone

*智能系统学报*. 2016, 11(3): 410–417 <https://dx.doi.org/10.11992/tis.201504022>

 微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202101024

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210901.1405.006.html>

## 视觉协同的违规驾驶行为分析方法

高尚兵<sup>1,2</sup>, 黄子赫<sup>1</sup>, 耿璇<sup>1</sup>, 臧晨<sup>1</sup>, 沈晓坤<sup>1</sup>

(1. 淮阴工学院 计算机与软件工程学院, 江苏 淮安 223001; 2. 淮阴工学院 江苏省物联网移动互联网技术工程实验室, 江苏 淮安 223001)

**摘要:** 本文针对危险驾驶识别中主流行为检测算法可靠性差的问题, 提出了一种快速、可靠的视觉协同分析方法。对手机、水杯、香烟等敏感物体进行目标检测, 提出的 LW(low weight)-Yolov4(You only look once v4) 通过去除 CSPDarknet53(cross stage partial Darknet53) 卷积层中不重要的要素通道提升了检测速度, 并  $L_1$  正则化产生稀疏权值矩阵, 添加到 BN(batch normalization) 层的梯度中, 实现优化网络模型的目的; 提出姿态检测算法对驾驶员指关节关键点进行检测, 经过仿射逆变换得到原始帧中的坐标; 通过视觉协同分析对比敏感物品的检测框位置与驾驶员手部坐标是否重合, 判定驾驶员是否出现违规驾驶行为及类别。实验结果表明, 该方法在识别精度与检测速度方面均优于主流的算法, 能够满足实时性和可靠性的检测要求。

**关键词:** 驾驶行为识别; 模型剪枝; 目标检测; 姿态估计; 协同检测; 模型优化; 深度学习; 卷积神经网络

**中图分类号:** TP391.4   **文献标志码:** A   **文章编号:** 1673-4785(2021)06-1158-08

中文引用格式: 高尚兵, 黄子赫, 耿璇, 等. 视觉协同的违规驾驶行为分析方法 [J]. 智能系统学报, 2021, 16(6): 1158-1165.

英文引用格式: GAO Shangbing, HUANG Zihé, GENG Xuan, et al. A visual collaborative analysis method for detecting illegal driving behavior[J]. CAAI transactions on intelligent systems, 2021, 16(6): 1158-1165.

## A visual collaborative analysis method for detecting illegal driving behavior

GAO Shangbing<sup>1,2</sup>, HUANG Zihé<sup>1</sup>, GENG Xuan<sup>1</sup>, ZANG Chen<sup>1</sup>, Shen Xiaokun<sup>1</sup>

(1. College of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223001, China; 2. Laboratory for Internet of Things and Mobile Internet Technology of Jiangsu Province, Huaiyin Institute of Technology, Huaian 223001, China)

**Abstract:** This study proposes a fast and reliable visual collaborative analysis method to improve the reliability of mainstream behavior detection algorithms in dangerous driving recognition. First, the algorithm performs target detection on sensitive objects such as mobile phones, water cups, and cigarettes. The proposed low weight-Yolov4 algorithm improves the detection speed by removing unimportant element channels in the cross-stage partial Darknet53 convolutional layer and regularizes  $L_1$  to generate a sparse weight matrix. Besides, the obtained matrix is added to the gradient of the batch normalization layer to optimize the network model. Then, an attitude detection algorithm is used to detect key points of the driver's knuckles, and the coordinates in the original frame are obtained through the affine inverse transformation. Finally, the driver's illegal driving behavior and its category are determined through visual collaborative analysis and comparison of the position of the detection frame of sensitive objects and coordinates of the driver's hands. Experimental results show that the recognition accuracy and detection speed of the proposed method are better than those of mainstream algorithms, which can meet the detection requirements of real-time and reliability.

**Keywords:** driving behavior recognition; model pruning; target detection; attitude estimation; collaborative detection; model optimization; deep learning; convolutional neural network

收稿日期: 2021-01-18. 网络出版日期: 2021-09-01.

基金项目: 国家重点研发计划项目 (2018YFB1004904); 江苏高校“青蓝工程”项目; 江苏省高校自然科学研究重大项目 (18KJA520001).

通信作者: 高尚兵. E-mail: [luxiaofen\\_2002@126.com](mailto:luxiaofen_2002@126.com).

随着机动车数量的增多, 频发的交通事故成为社会关注的难题, 而司机的不规范驾驶行为 (如抽烟、玩手机等) 是造成事故发生的重要原

因。通过约束驾驶员的行为可以在一定程度上减少此类交通事故的发生, 众多研究者开始致力于研究如何快速有效地检测驾驶员的不规范驾驶行为。近年来, 图像识别成为了检测驾驶员行为的趋势, 深度学习在计算机视觉领域被广泛用于图像分类和目标检测, 卷积神经网络 (convolutional neural network, CNN) 在图像识别和物体检测方面功不可没<sup>[1]</sup>。

目前基于 CNN 的检测驾驶员行为的方法包括: Le 等<sup>[2]</sup>利用 R-CNN 目标检测算法, 对驾驶员头部、手部和方向盘进行检测, 通过检测头部和手部的的位置, 判断驾驶员是否在打电话, 检测手部和方向盘的位置, 判断驾驶员的手是否脱离方向盘; 李俊俊等<sup>[3]</sup>对经典卷积神经网络进行改进, 提出了结合 FCNN 与三级级联神经网络融合的模式, 实验结果表明该模型能有效识别违规行为, 但仍需提高训练的准确率和速度; 魏泽发<sup>[4]</sup>构建了一种基于 SSD<sup>[5]</sup> 目标检测算法的出租车司机违规行为检测方法, 他选用 VGG16<sup>[6]</sup> 作为 SSD 算法的基础网络, 结合自建数据集的特点, 采用迁移学习的方法训练数据集, 准确率高达 94.22%, 较原始 SSD 算法提升了 2.42%, 同时能保证模型处理速度为 33 f/s。Jin 等<sup>[7]</sup>采用了一种基于双流卷积神经对驾驶员使用手机这种行为进行实时检测, 并在自建数据集上取得了 95.7% 的准确率; Huang 等<sup>[8]</sup> 建立了一个混合 CNN 的框架, 先采用一个联合的预训练的模型对驾驶员的行为特征进行提取, 再将这些提取出来的特征送入到全连接层进行分类, 实验结果表明, 检测准确率达到 96.74%。He 等<sup>[9]</sup> 基于 CornerNet-Lite<sup>[10]</sup> 神经网络对驾驶员打电话这种行为进行检测, 在取得 86.2% 的准确率的同时还有着 30 f/s 的实时检测速度, 并且即使在具有噪声干扰的环境下, 仍然能够保持较为稳定的鲁棒性。Masood 等<sup>[11]</sup> 在内安装一个摄像头, 利用 VGG16 对驾驶员的违规行为进行识别, 实验结果表明, 平均准确率达到 99%。

实时性和有效性是驾驶员行为检测的重要指标。将基于 CNN 的目标检测方法如 YOLOv4<sup>[12]</sup>、YOLOv3<sup>[13]</sup>、EfficientDet<sup>[14]</sup>、Faster-RCNN<sup>[15]</sup>、RetinaNet<sup>[16]</sup>、R-FCN<sup>[17]</sup> 等算法进行对比, YOLOv4 算法更加高效且适合单次 GPU 训练。另外, 姿态检测算法 Open Pose<sup>[18]</sup> 能够实现多人姿态检测, 优点是能在人物数量多的情况下, 既能保持精度, 又可以提升检测速度。

本文提出了一种新颖算法 LW-Yolov4, 通过去除网络模型中权重较低的模型通道数来简化

Yolov4 网络模型大小, 从而提高了检测速度。驾驶员违规行为的识别需要通过提出的 LW-Yolov4 网络和姿态检测算法 OpenPose 对驾驶员的监控数据进行协同检测, 将视频帧图像送入 LW-Yolov4 网络获取置信度较高的敏感物品检测框位置坐标, 并同时对人体姿态关键点的检测, 获取手指部分关键点坐标。通过判断双手手指部分关键点与敏感物品识别区域是否有重叠, 以检测行车过程中玩手机、抽烟、喝水等违规驾驶行为。

## 1 Yolov4 目标检测算法

Yolov4 目标检测算法平衡了精度与速度, 相较于 Yolov3, mAP 提升了 10%, 速度提升了 12%。Yolov4 的主干网络采用了 CSPDarknet53 网络, 并将 SPP (spatial pyramid pooling) 模型<sup>[19]</sup> 加入到 CSPDarknet53 中, 使用 PANet<sup>[20]</sup> (path aggregation network) 代替了 Yolov3 的 FPN<sup>[21]</sup>。用 Focal Loss 来解决数据不平衡的问题。回归框的损失则采用 CIOU Loss<sup>[22]</sup>。基于 CutMix<sup>[23]</sup> 数据增强方法, Yolov4 提出了一种新的数据增强方法 Mosaic, 将 4 张图片拼接成一张, 能够丰富检测物体的背景, 降低批处理数量, 使得图片在单 GPU 上训练更为轻松。Mosaic 拼接效果如图 1 所示。



图 1 Mosaic 拼接效果  
Fig. 1 Mosaic effect

### 1.1 Yolov4 基础网络结构

Yolov4 采用的 CSPDarknet53 网络包含了 29 个卷积层, 感知野大小为 725×725。Darknet53 结构如图 2 所示, 由 5 个残差块组成。

CSPDarknet53 为 Darknet 改进而来, 在 Darknet 基础上添加了 CSPNet<sup>[24]</sup> (cross stage partial network), 能在轻量化的同时保持准确性, 降低计算的成本。CSPDark 网络是在每一个残差块上加上 CSP, 图 3 为残差块 1 的 CSPDarknet53 模型图, 其中, 层 [0,1,5,6,7] 与 Darknet53 的原网络一致, 而层 [2,4,8,9,10] 为新添加的 CSPNet。

类型	卷积核数量	大小	输出
卷积	32	3×3	256×256
卷积	64	3×3 /2	128×128
1×	卷积	32	1×1
	卷积	64	3×3
	残差		128×128
2×	卷积	128	3×3 /2
	卷积	64	1×1
	卷积	128	3×3
8×	残差		64×64
	卷积	256	3×3 /2
	卷积	32×32	
8×	卷积	128	1×1
	卷积	256	3×3
	残差		32×32
8×	卷积	512	3×3 /2
	卷积	16×16	
	卷积	256	1×1
4×	卷积	512	3×3
	卷积	1024	3×3
	残差		8×8
平均池化		Global	
全连接		1000	
分类			

图 2 Darknet53 网络结构  
Fig. 2 Darknet53 network structure

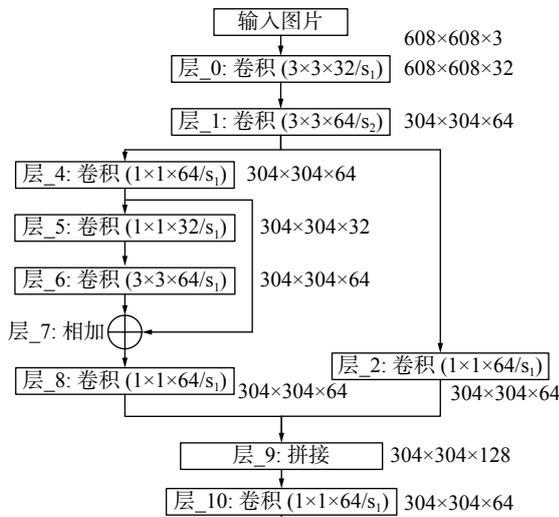


图 3 残差块 1 的 CSPDarknet53  
Fig. 3 CSPDarknet53 of residual block 1

1.2 LW-Yolov4 算法

Yolov4 算法计算精度高, 检测速度快, 通过加入 CSP(cross stage partial)、SPP(spatial pyramid pooling) 模块和 PANet(path aggregation network) 进行多尺度融合的同时利用路径聚合网络将底层特征信息与高层特征进行融合, 从而有效增强模型的表达能力, 仅牺牲少量训练速度的代价来换得精度的提升。由于车载系统属于低功耗场景, 实时检测需要模型的计算量和大小尽可能的小, 为了让驾驶员行为检测平台对数据能够实时地处理

反馈, 针对驾驶员协同检测算法中的目标检测网络模型结构, 本文提出了一种 Yolov4 的网络模型简化算法——LW-Yolov4, LW(light weight) 代表了轻量化, 即轻量化的 Yolov4 网络模型。

LW-Yolov4 算法简化了神经网络模型, 通过去除卷积层中不重要的要素通道, 达到提升检测速度的目的。首先通过  $L_1$  正则化<sup>[12]</sup> 产生稀疏权重矩阵<sup>[25]</sup>, 得到的梯度添加到 BN (batch normalization) 层的梯度中。 $L_1$  正则化表示为

$$\Omega(\omega) = \|\omega\|_1 = \sum_i |\omega_i| \quad (1)$$

式中:  $\omega_i$  为模型权重系数;  $\Omega(\omega)$  为惩罚项。使用 BN 层中的缩放因子  $\gamma$  评估参数, 根据每个神经元的  $L_1$  绝对值的权重参数  $\gamma$  从小到大进行排序。设立合适  $\gamma$  的阈值, 该阈值决定了网络模型最终的大小。本实验将  $\gamma$  阈值设置为 0.8, 即优化后的神经网络为原网络的模型的 0.2 倍。将  $\gamma$  低于阈值 0.8 的权重参数置 0。然后将由 BN 层得到的卷积计算结果与合并后的权值参数  $\beta$  (偏置量) 进行计算。通过去除这些不重要的像素通道, 网络模型在进行模型推理的时候, 可以跳过这些通道, 仅仅去加载那些权重大的通道, 从而减少模型的计算量, 提高模型的检测速度。LW-Yolov4 算法流程如图 4 所示。

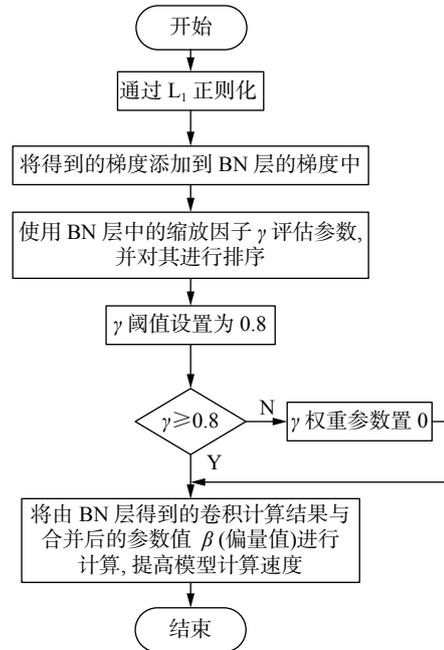


图 4 LW-Yolov4 算法流程  
Fig. 4 Flow chart of LW-Yolov4 algorithm

模型通道数量的改进前后对比效果见表 1。通过对比改进前后的模型通道数可知 Yolov4 网络模型得到了简化。

表1 模型通道数对比  
Table 1 Model channel number comparison

卷积层下标	总通道数	剩余通道数
0	32	19
3	32	24
6	64	25
9	64	36
13	128	45
22	128	48
25	128	39
32	128	47
34	128	51
38	256	42
42	256	38
46	256	52
49	256	36
51	256	46
56	256	35

### 1.3 基于LW-Yolov4的敏感物品检测

本文通过LW-Yolov4算法进行车内敏感物品的检测,训练前需对数据集图像进行归一化<sup>[26]</sup>处理,归一化采用了z-score标准化方法,基于原始数据的均值和标准差,经过处理的数据符合标准正态分布,转换函数为

$$x^* = \frac{x - \mu}{\sigma} \quad (2)$$

式中: $\mu$ 为所有样本数据的均值; $x^*$ 为样本数据的标准差。

训练函数采用了CIoU Loss, CIoU Loss加入了一个影响因子,这个因子把预测框长宽比、拟合目标框的长宽比考虑进去,公式为

$$\text{CIoU} = \text{IOU} - \frac{\rho^2(b, b^{\text{gt}})}{c} - av \quad (3)$$

式中:CIoU为交并比; $v$ 为权重函数,用来度量长宽比; $b, b^{\text{gt}}$ 分别代表了预测框和真实框的中心点; $c$ 代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离; $\alpha$ 为正权衡参数。

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (4)$$

式中: $w, w^{\text{gt}}$ 分别代表了预测框宽和真实框宽; $h, h^{\text{gt}}$ 分别代表了预测框高和真实框高。损失函数公式为

$$\text{LOSS}_{\text{CIoU}} = 1 - \text{IOU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + av \quad (5)$$

## 2 视觉协同检测算法融合

将敏感物品的识别与人体的姿态进行协同融合检测,从而达到驾驶员违规驾驶行为的识别的目的。

### 1) 人体姿态算法 OpenPose

OpenPose人体姿态识别算法<sup>[10]</sup>由美国卡耐基梅隆大学提出,可以实现人体动作、面部表情、手指运动等姿态估计。OpenPose使用的是联合的多branch分支,一个分支负责关键点的检测,另一个分支负责将关键点连接成骨架,再通过二分图匹配匈牙利算法。

### 2) 算法融合策略

将待检测视频的每一帧同时送入LW-Yolov4网络模型和OpenPose模型,LW-Yolov4算法检测后得到敏感物品的位置信息以及置信度,OpenPose检测后得到人体关键点信息及手部的的位置坐标。设定一个阈值,若连续1s内检测到驾驶员的手指位置与敏感物品重合则判定驾驶员出现了违规驾驶行为,保留视频节点并进行语音提醒。算法融合流程如下:

1) 将采集到的动作图片序列输入检测系统,对图片进行裁剪、缩放、中值滤波等图像预处理操作,获得416像素×416像素的图片,作为模型的输入;

2) 将图片输入到LW-Yolov4模型中进行敏感检测,并得到敏感物的坐标;

3) 将图片输入到OpenPose中获得手部关键点的坐标;

4) 进行图片渲染,整合到一张图片上;

5) 判断敏感物坐标是否与手部坐标重叠,若重叠则进行相应的预警。

## 3 实验分析

本文的驾驶员行为识别分析系统是通过对比车内敏感物品的检测框位置与驾驶员手部位置坐标是否出现重合来判定驾驶员是否出现了违规驾驶行为的。因此,实验分为3个部分,分别为车内敏感物品的检测、人体关键点的检测以及视频融合与行为判定。

### 3.1 实验平台及数据

实验数据来自淮安公共交运平台的驾驶员行车过程中的驾驶员视角的监控视频,视频数据大小为430GB。视频像素大小为1280×720,包含了同一场景下的红外摄像头拍摄的灰度图像以及高清摄像头拍摄的日间行车图像。使用标注工具

Labellmg 对视频中截取的敏感物品进行标注, 主体标注对象为手机、水杯和香烟 3 个类别。将标注完成的图片制作成 VOC 格式的数据集。数据集包含了 15 000 张训练集以及 6 000 张验证集, 实验平台的环境配置如表 2 所示。

表 2 实验环境配置  
Table 2 Experimental environment configuration

名称	相关配置
操作系统	Windows10 1805
CPU/GHz	AMD Ryzen 2700X CPU
RAM/GB	16,3200
GPU	NVIDIA GeForce GTX 2060,6
硬盘	Samsung 512 GB 970 PRO SSD
框架	Pytorch
GPU加速库	CUDA10.0, CUDNN7.4

### 3.2 协同检测算法融合训练过程

将制作完成的 VOC 格式的数据集图片采用 Mosaic 数据增强方法进行裁剪、旋转以及缩放操作。LW-Yolov4 训练参数的初始学习率设置为 0.001, 采用阶跃衰减学习率<sup>[27]</sup> 调度策略, 预测分类类别数设置为 3(检测手机、水杯、香烟这三类敏感物品), 经过 2 000 次的迭代训练。读取待检测视频的每一帧图像, 通过训练好的网络模型进行敏感物品的检测。图 5 为日间行车敏感物品检测效果图, 图 6 为灰度图像敏感物品检测效果图。



图 5 日间行车检测效果  
Fig. 5 Renderings of daytime driving detection

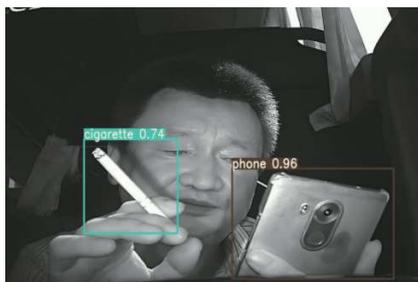


图 6 灰度图像检测效果  
Fig. 6 Grayscale image detection renderings

将送入 LW-Yolov4 网络模型的视频帧同时送入 OpenPose 进行人体关键点检测, 得到人体关键点信息。结合敏感物品的检测窗口及姿态检测的信息对原视频每一帧进行渲染, 并进行文本可视化, 根据检测到的敏感物品以及手部信息在左上角给出文本进行提示, 渲染后的效果图如图 7、图 8 所示。



图 7 日间行车融合效果  
Fig. 7 Renderings of daytime driving fusion

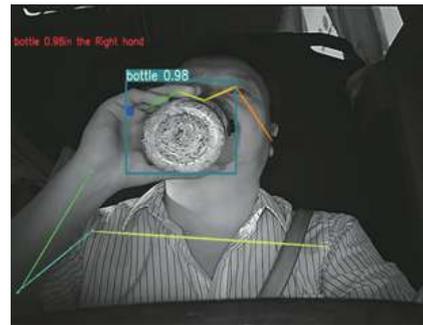


图 8 灰度图像融合效果  
Fig. 8 Gray-scale image fusion effect

### 3.3 实验结果

本文选取了 3 段驾驶员行车过程监控视频 (10 800 f) 进行检测, 检测对象喝水、打电话、玩手机等违规驾驶行为。通过统计含有违规驾驶行为的帧数与检出违规驾驶行为的帧数计算检出率, 检测结果如表 3 所示。

表 3 检测结果  
Table 3 Test results

视频编号	行为类型	检出率/%	FPS
01	喝水	95.65	20.95
	玩手机	94.32	21.04
	抽烟	93.48	21.53
02	喝水	96.68	21.27
	玩手机	95.46	22.04
	抽烟	93.41	20.98
03	喝水	95.36	21.96
	玩手机	94.17	20.87
	抽烟	93.29	21.63

通过表 3 的数据可以看出, 驾驶员的违规驾驶行为具有较高的检出率, 3 个视频的违规行为总体检出率为 94.76%, 其中喝水的平均检出率达到了 95.90%, 玩手机的平均检出率为 94.65%, 抽烟检出率为 93.40%。整体来看, 喝水的检出率略高于玩手机及抽烟的检出率。检测过程的 FPS 为

21.36, 能够满足驾驶员违规行为检测的准确性和实时性要求。

为了进一步验证 LW-Yolov4 算法的性能, 本文对面前应用的主流算法进行训练, 对比实验结果。评价指标为平均精度召回率 mAP、精确率 Precision 以及召回率 Recall。实验结果如表 4 所示。

表 4 LW-Yolov4 算法性能测试结果  
Table 4 Test results of LW-Yolov4 algorithm

检测方法	基础网络	mAP	精准率	召回率	速度/(毫秒/张)
Yolov3	Darknet-53	0.874	0.851	0.866	32
Yolov4	CSPDarknet	0.984	0.967	0.981	33
LW-Yolov4	CSPDarknet+	0.979	0.962	0.975	30
SSD	Residual-101	0.802	0.767	0.793	60
Faster-R-CNN	VGG-16	0.793	0.805	0.774	64
R-FCN	Residual-101	0.813	0.816	0.814	88

通过表 4 中的数据可以看出, LW-Yolov4 算法的运行速度比 Yolov4 高出了 9.1 个百分点, 比 Yolov3 高出了 6.7 个百分点。但是在精度上, LW-Yolov4 下降的原因在于去除了部分权重较低的网络通道。与 SSD、Faster-RCNN 和 R-FCN 相比, 本文提出的 LW-Yolov4 在速度和精度上均有了较大幅度的提升。根据实验结果可知, LW-Yolov4 的算法优势在于, 精度小幅下降的情况下, 大幅提升了检测速度, 能够满足实时检测的要求。为了进一步验证融合算法对于危险驾驶行为的检测效果, 本文进行了协同检测算法的性能和精度对比。实验对比结果如表 5 所示。

10% 以上。

由对比实验可以看出, 本文提出的协同检测算法在检测精度、检测速度方面均有优势。

#### 4 结束语

本文提出的 LW-Yolov4 算法是在 Yolov4 的基础上通过精简网络模型改进得到的。通过 LW-Yolov4 算法与人体姿态算法进行融合从而达到检测驾驶员行车过程中违规行为的目的。目前众多深度学习算法模型在精度上表现都较为优越, 但很多行业对数据的实时性处理要求较高, 在检测速度上还未能完全得到满足。本文提出的 LW-Yolov4 算法在精度能够达到实际应用的同时大幅提高了检测的速度, 能够满足基础的应用, 但距离工业化的检测速度要求还有一定的差距。在未来目标检测的前进方向上, 目标检测速度算法的提升仍会成为重点的研究方向。

表 5 各目标检测算法性能测试结果

Table 5 Test results of the accuracy of each target detection algorithm

检测方法	类别				FPS
	mAP	手机	水杯	香烟	
YOLOv3	0.892	0.921	0.903	0.847	34
YOLOv4	0.932	0.941	0.927	0.836	36
协同检测算法	0.985	0.987	0.982	0.991	40
SSD	0.795	0.834	0.843	0.709	15
Faster-RCNN	0.763	0.799	0.817	0.675	12
R-FCN	0.804	0.875	0.834	0.703	17

在 3 类违规驾驶行为的综合检出率中, 协同检测算法相较于 Yolov4, 检测精度提升了 5.3%, 检测速度提升了 10%。相较于其他主流算法, 融合检测算法的检测速度和精度也都具有明显的检测优势, 检测精度提升 5% 以上, 检测速度提升了

#### 参考文献:

- [1] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望 [J]. 自动化学报, 2017, 43(8): 1289-1305.  
ZHANG Hui, WANG Kunfeng, WANG Feiyue. Advances and perspectives on applications of deep learning in visual object detection[J]. Acta automatica sinica, 2017, 43(8): 1289-1305.
- [2] LE T H N, ZHENG Yutong, ZHU Chenchen, et al. Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection[C]//2016

- IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas, USA, 2016: 46–53.
- [3] 李俊俊, 杨华民, 张澍裕, 等. 基于神经网络融合的司机违规行为识别 [J]. 计算机应用与软件, 2018, 35(12): 222–227, 319.
- LI Junjun, YANG Huamin, ZHANG Shuyu, et al. Driver's illegal behavior recognition based on neural network fusion[J]. Computer applications and software, 2018, 35(12): 222–227, 319.
- [4] 魏泽发. 基于深度学习的出租车司机违规行为检测 [D]. 西安: 长安大学, 2019.
- WEI Zefa. Taxi driver violation detection based on deep learning[D]. Xi'an: Chang'an University, 2019.
- [5] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 21–37.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10)[2020-12-03] <https://arxiv.org/abs/1409.1556>.
- [7] JIN Chongchong, ZHU Zhongjie, BAI Yongqiang, et al. A deep-learning-based scheme for detecting driver cell-phone use[J]. *IEEE access*, 2020, 8: 18580–18589.
- [8] HUANG Chen, WANG Xiaochen, CAO Jiannong, et al. HCF: a hybrid CNN framework for behavior detection of distracted drivers[J]. *IEEE access*, 2020, 8: 109335–109349.
- [9] HE Anqing, CHEN Guohua, ZHENG Wei, et al. Driver cell-phone use detection based on CornerNet-Lite network[C]//OP Conference Series: Earth and Environmental Science. Smolensk, Russian, 2021: 042004.
- [10] LAW H, TENG Yun, RUSSAKOVSKY O, et al. Cornernet-lite: efficient keypoint based object detection [EB/OL]. (2019-04-18)[2020-12-03] <https://arxiv:1904.08900.2019>.
- [11] MASOOD S, RAI A, AGGARWAL A, et al. Detecting distraction of drivers using convolutional neural network[J]. *Pattern recognition letters*, 2020, 139: 79–85.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL].(2020-04-23)[2020-12-03] <https://arxiv.org/abs/2004.10934>.
- [13] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL].(2018-04-08)[2020-12-03] <https://arxiv.org/abs/1804.02077>.
- [14] TAN Mingxing, PANG Ruoming, LE O V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA, 2020: 10778–10787.
- [15] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(6): 1137–1149.
- [16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2999–3007.
- [17] DAI Jifeng, LI Yi, HE Kaiming, et al. R-FCN: object detection via region-based fully convolutional networks [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain, 2016: 379–387.
- [18] CAO Zhe, HIDALGO G, SIMON T, et al. OpenPose: realtime multi-person 2D pose estimation using part affinity fields[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 43(1): 172–186.
- [19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015, 37(9): 1904–1916.
- [20] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8759–8768.
- [21] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 936–944.
- [22] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 658–666.
- [23] YUN S, HAN D, CHUN S, et al. Cutmix: regularization strategy to train strong classifiers with localizable features[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea

(South), 2019: 6022–6031.

- [24] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, USA, 2020: 1571–1580.
- [25] 赵传君, 王素格, 李德玉. 跨领域文本情感分类研究进展 [J]. 软件学报, 2020, 31(6): 1723–1746.  
ZHAO Chuanjun, WANG Suge, LI Deyu. Research progress on cross-domain text sentiment classification[J]. Journal of software, 2020, 31(6): 1723–1746.
- [26] 王岩. 深度神经网络的归一化技术研究 [D]. 南京: 南京邮电大学, 2019: 1179–1185.  
WANG Yan. Analysis of normalization for deep neural networks[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2019: 1179–1185.
- [27] ZHANG Shuang, SONG Zongxi. An ethnic costumes classification model with optimized learning rate[C]//The 11th International Conference on Digital Image Processing. Guangzhou, China, 2019: 1179–1185.

### 作者简介:



高尚兵, 教授, 博士, 主要研究方向为机器学习、计算机视觉、模式识别和数据挖掘。获中国仿真学会科技进步二等奖、吴文俊人工智能科技进步三等奖。发表学术论文 100 余篇。



黄子赫, 硕士研究生, 主要研究方向为计算机视觉、模式识别和数据挖掘。



耿璇, 本科生, 主要研究方向为图像识别。

## 自然语言处理技术创新大赛——中文文本纠错比赛圆满落幕

近日, 由中国人工智能学会主办, 科大讯飞股份有限公司组织, 认知智能国家重点实验室(科大讯飞)、哈尔滨工业大学社会计算与信息检索研究中心承办的“第三届中国‘AI+’创新创业大赛——自然语言处理技术创新大赛——中文文本纠错比赛”(CIEC-CTC 2021)圆满落幕。

自 2021 年 6 月 10 日开赛以来, 包括阿里、网易、京东等企业以及清华大学、北京大学等高校的 124 支队伍报名参赛, 其中 42 队通过资格赛进入决赛, 经过决赛的激烈角逐, 最终评选出冠军 1 队、亚军 2 队、季军 3 队。

CIEC-CTC 2021 赛程历时 4 个月, 旨在促进中文自然语言技术的发展。比赛选择互联网上中文母语写作者撰写的网络文本作为校对评测数据, 从拼写错误、语法错误、语病错误等多个方面考察机器的认知智能能力, 为从事中文自然语言处理研究的研究人员、产业界从业人员以及 AI 技术爱好者提供了一个良好的交流平台。