

DOI:10.19651/j.cnki.emt.2416177

基于改进 YOLOv8-pose 的分心驾驶检测与识别^{*}

朱周华 侯智杰 田成源 周怡纳

(西安科技大学通信与信息工程学院 西安 710000)

摘要: 针对现有的分心驾驶检测算法存在检测率低、检测速率慢等问题,本文构建了一种基于改进 YOLOv8-pose 的分心驾驶检测识别模型 YOLOv8-EFM。首先,通过使用 EfficientViT 更换 YOLOv8-pose 的主干网络,结合 CNN 和 ViT 之间的互补性,提升了检测的准确率;其次,使用 FasterBlock 模块替换 C2f 中的 Bottleneck 模块,增加了检测速率并减小模型参数;最后在 SPPF 后加入了轻量级的 MLCA 注意力模块,在模型大小和准确性之间取得了良好的平衡。实验结果表明,本文所构建的 YOLOv8-EFM 模型, mAP50 可以达到 98.9%,模型大小只有 9.7 M,该方法不仅可以识别出具体分心行为,还可以检测上半身的人体骨架,可以有效应用在驾驶员分心驾驶的检测场景中。

关键词: 分心检测;人体姿态估计;YOLOv8-pose;EfficientViT;FasterNet;MLCA

中图分类号: TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.50

Distracting driving detection and identification based on an improved YOLOv8-pose

Zhu Zhouhua Hou Zhijie Tian Chengyuan Zhou Yina

(School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710000, China)

Abstract: Aiming at the existing distracted driving detection algorithms, this paper constructs a YOLOv8-EFM distracted driving detection and recognition model based on improved YOLOv8-pose. Firstly, by replacing the backbone network of YOLOv8-pose with EfficientViT, combined with the complementarity between CNN and ViT, the detection accuracy is improved; secondly, replacing the Bottleneck module in C2f with FasterBlock module, increasing the detection rate and reducing the model parameters; finally, the lightweight MLCA attention module is added after SPPF, achieving a good balance between model size and accuracy. The experimental results show that the YOLOv8-EFM model constructed in this paper can detect mAP 50 with 98.9%, and the model size is only 9.7 M. This method can not only detect the specific distraction behavior, but also detect the human skeleton of the upper body, which can be effectively applied in the detection scene of distracted driving.

Keywords: distraction detection; human pose estimation; YOLOv8-pose; EfficientViT; FasterNet; MLCA

0 引言

自 1885 年世界上第一辆汽车产生至今,随着科技的进步和人们对于汽车需求量的增加,汽车正在作为重要的交通工具,逐步推动着人类各行业的蓬勃发展。随着汽车数量的不断增加,交通安全问题呼之欲出,而人-车-路是复杂道路交通系统的重要组成部分,每个因素都起到至关重要的作用,任何因素的不稳定问题,都是导致交通事故发生的潜在风险,其中,驾驶人是造成交通事故的主要因素,而分心驾驶行为则是最大的原因。分心驾驶行为是指在驾驶过程中,驾驶员因为各种因素导致注意力不集中或转移

注意力,而影响到驾驶行为的安全性,包括但不限于打电话、发短信、抽烟^[1]、吃东西、喝水、听音乐等行为。因此对分心驾驶行为的检测,并对驾驶危险行为进行警报^[2]相当重要,这样可以在一定程度上减少甚至避免交通事故的发生,更有利于保障驾驶员与车上乘客的人身安全,对于构建平安中国有着相当重要的社会意义和应用价值。

早期,研究者只是使用心电图、脑电图^[3-4]等对驾驶员的疲劳程度^[5]进行测量与判断,但是这种方法需要将相关仪器佩戴在驾驶员身上,这反而会阻碍驾驶员进行正常的驾驶操作。后来则是对车辆行驶数据的研究,通过检测行驶速度^[6]、方向盘运动^[7]等进行驾驶员分心行为检测,但是

收稿日期:2024-06-04

^{*} 基金项目:国家自然科学基金(U19B2015)项目资助

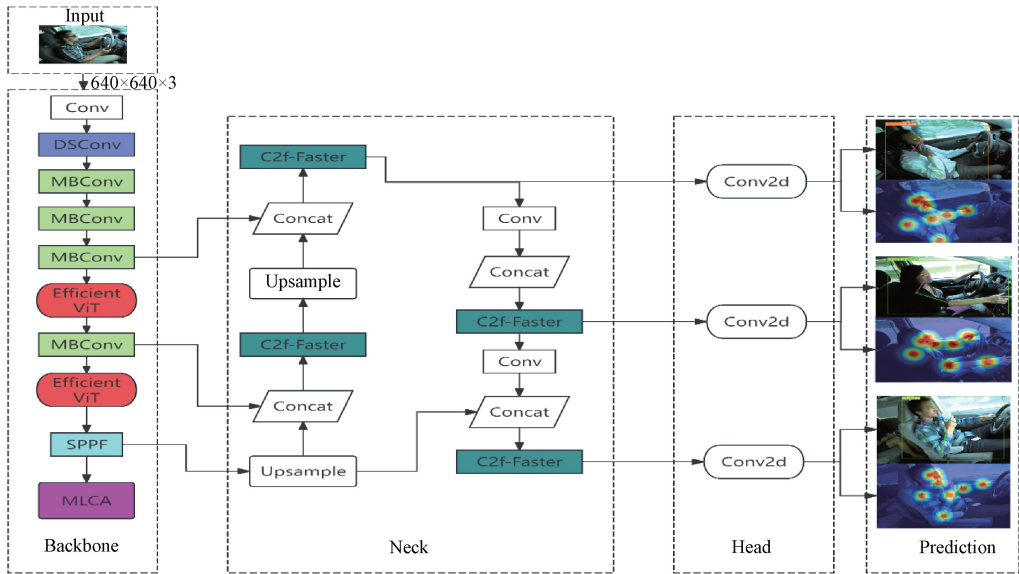


图2 YOLOv8-EFM网络结构

Fig. 2 Structure of YOLOv8-EFM network

网络架构如图3所示。首先,输入维度为 $640 \times 640 \times 3$ 的三通道RGB图像,之后特征图依次通过DSConv模块,MBConv模块和EfficientViT模块,最终将输出输入到SPPF模块中。

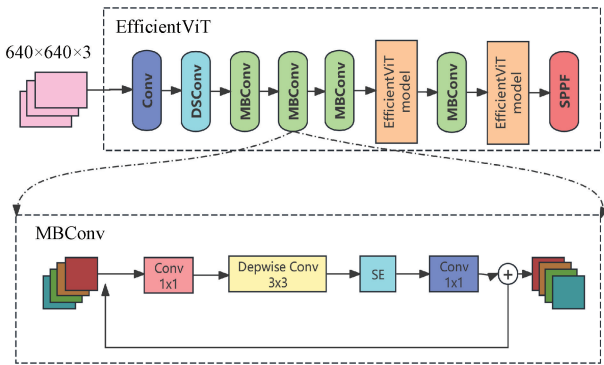


图3 EfficientViT网络整体架构

Fig. 3 The overall architecture of the EfficientViT network

在特征提取过程中,主要使用MBConv (mobilenetV3 block convolution, MBConv)模块进行图像特征提取,MBConv模块继承了Mobilenet v1的深度可分离卷积和Mobilenet v2的具有线性瓶颈到底倒残差结构,并引入SE注意力机制。如图3所示,MBConv首先通过一个 1×1 的轻量级卷积。可以降低输入特征图的通道数,接着使用Depthwise Conv深度可分离卷积对输入特征图像进行处理,然后输送到SE模块,SE模块通过一个全局平均池化将特征图降维,然后通过两个全连接层生成每个通道的注意力权重,最后这些权重与输入特征相乘,得到新的特征图。SE模块的输入是之前模块得到的特征图 (C, W, H) ,如式(1)所示。

$$F \in R^{H \times W \times C} \quad (1)$$

输出如式(2)所示。

$$F' = F \otimes \text{Sigmoid}(W_2(\text{RELU}(W_1 \text{Pool}(F)))) \quad (2)$$

式中: \otimes 表示基于元素的乘法,两个全连接层的激活函数分别为ReLU函数和Sigmoid函数,最后通过一个 1×1 的卷积将通道数恢复。

2) FasterNet 网络

针对目前在驾驶员分心驾驶检测任务中参数量过大导致在边缘设备上推理速度较慢的问题,因此使用FasterNet^[21]网络对C2f模块进行改进,使用FasterBlock模块替换C2f中的Bottleneck模块,可以有效降低检测任务中的计算量和参数量,增加检测速率,并且具备一定的轻量化水平。

FasterNet是一种高效的神经网络,通常用于完成目标检测任务,并在速度和精度方面进行了优化,核心思想是在保持轻量级和高速度的基础上,提高特征表达能力和感受野的覆盖范围。FasterNet网络结构由4个阶段组成,如图4所示。每个阶段都是对特征进行提取,仅在卷积核大小上有区别,Embedding模块由一个步长为4的正则卷积组成,Merging层是由一个步长为2的卷积组成,用于空间下采样和通道数扩展。

FasterNet Block模块是FasterNet的核心模块,该模块设计思想来源于GhostNet,在一定程度上解决了特征卷积通道上具有一定冗余的问题,但与GhostNet相比,并没有采用DWConv(深度可分离卷积),而是提出了一种新的部分卷积(PCConv)。PCConv仅在输入通道的一部分上应用常规卷积进行空间特征提取,其余通道则是保持大小不变。对于连续后者规则的内存访问,将第1个或最后1个连续的通道视为整个特征图的代表进行计算,在不丧失一般性的情况下认为输入和输出特征图具有相同数量的通

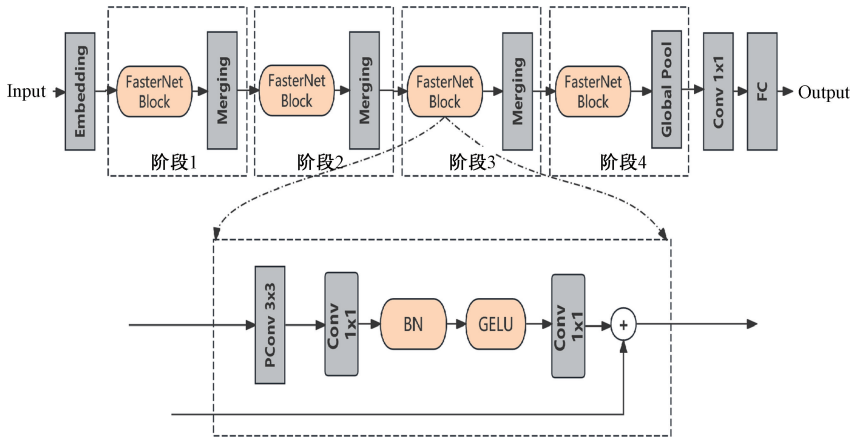


图 4 FasterNet 网络结构图

Fig. 4 FasterNet network structure

道, PConv 的计算量如式(3)所示。

$$F = h \times w \times k^2 \times cp^2 \quad (3)$$

PConv 的内存访问量如式(4)所示。

$$M = h \times w \times 2cp + k^2 \times cp^2 \quad (4)$$

式中: h 为通道的高, w 为通道的宽, cp 为连续的网络通道数, k 为卷积核尺度。因此, FasterNet Block 采用了部分卷积(PConv), 可以通过减少冗余计算和内存访问来更有效

地提取空间特征。

3) MLCA 注意力模块

针对增加检测速率所带来的检测精度下降问题, 本文在 SPPF 后加入了 MLCA 注意力模块。MLCA 是一种轻量级的注意力模块, 该模块同时考虑通道信息和空间信息, 并结合局部信息和全局信息以提高网络的表达效果, MLCA 的结构和工作原理如图 5 所示。

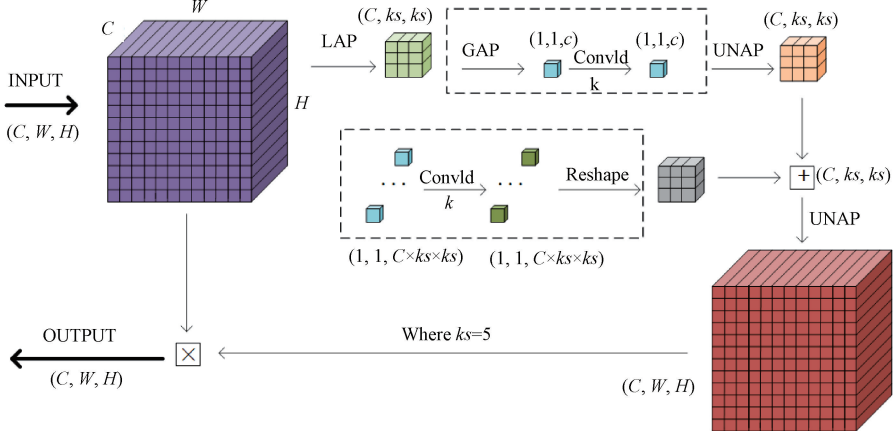


图 5 MLCA 网络结构和工作原理图

Fig. 5 The structure and working principle of the MLCA network

首先输入特征图 (C, W, H) 进行局部平均池化(LAP)和全局平均池化(GAP)处理。其中局部池化关注局部区域的特征, 而全局池化捕捉整个特征图的统计信息。之后局部池化后的特征和全局池化后的特征都经过一个 1D 卷积(Conv1d)来进行特征转换。此处的 1D 卷积用于压缩特征通道, 同时保持空间维度不变。经过 1D 卷积后, 特征被重新排列, 使其适应后续操作。对局部池化后的特征, 使用 1D 卷积后进行重新排列, 然后通过“乘法”操作与原始输入特征相结合。这个过程相当于一种特征选择, 强化了对有用特征的关注的。再对全局池化后的特征, 经过 1D 卷积和重新排列后, 通过“加法”操作与局部池化特征相结

合。这个步骤在特征图中融合了全局上下文信息。最后, 经过局部和全局注意力处理的特征图再次通过反池化(UNAP)操作, 恢复到原始的空间维度。

2 实验结果与分析

2.1 数据集

本文实验所使用的数据集为 Kaggle 机构推出了 StateFarm 分心数据集, 总共 10 个分心类型, 分别为安全驾驶, 右手发短信, 右手打电话, 左手发短信, 左手打电话, 调中控台, 喝水, 拿后座东西, 整理头发或化妆, 与副驾驶交谈, 如图 6 所示。



图 6 分心驾驶类型
Fig. 6 Distracted driving type

由于驾驶人在开车时主要的分心驾驶行为体现在上半身,所以在检测时主要对眼睛,鼻子,肩膀以及双手臂进行检测。首先对数据集进行关键点标注,借助 labelme 标记工具,对驾驶员进行框选关键点标注,关键点主要包括:鼻子,左右眼,左右肩膀,左右手肘,左右手腕,标记示意图如图 7 所示。

2.2 实验环境与模型参数设置

本文所有实验都是基于表 1 环境运行,在预训练阶段,所有模型训练周期 (Epoch) 为 100,批量大小 (Batch size) 为 16,初始学习率为 0.01。

2.3 消融实验

为了验证改进算法的有效性,首先使用预训练模型,保持相同的软硬件实验配置,分别在原始 YOLOv8-pose



图 7 驾驶员关键点标注示意图
Fig. 7 Schematic diagram of driver key point annotation

表 1 实验环境

Table 1 Experimental environment

名称	相关配置
操作系统	Ubuntu18.04.5
CPU	Intel i9-10900X CPU@3.7 GHz×20
内存	200 GB
GPU	NVIDIA RTX 2080Ti
CUDA	11.3
PyTorch	1.12.1
Python	3.8.18

的基础上依次加入 EfficientViT、C2f-Faster 和 MLCA 这 3 个改进点,以验证每个改进点对模型的提升效果。精度曲线图和损失曲线图如图 8、9 所示。可以看出 YOLOv8-EFM 表现出了最佳性能,精度达到了 98.6%,相比最初的 YOLOv8-pose 提升了 2.3%,损失则是 0.10,比 YOLOv8-pose 下降了 0.04。

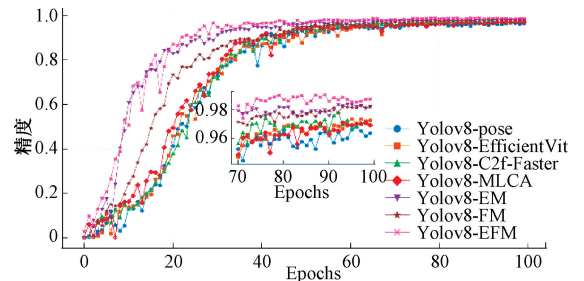


图 8 消融实验精度曲线图

Fig. 8 Accuracy curve of ablation experiment

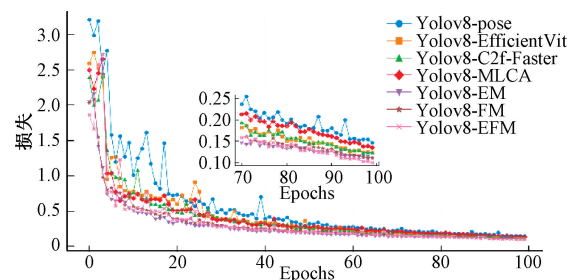


图 9 消融实验损失曲线图

Fig. 9 Loss plots of the ablation experiments

其次将精确率 (Precision)、召回率 (Recall)、mAP50、模型大小作为消融实验验证标准,精确率和 mAP50 可以

表示模型检测的精度效果,模型大小可以表示网络的轻量化程度。消融实验分为 7 个实验序列,结果如表 2 所示。

表 2 消融实验结果对比

Table 2 Comparison of the results of the ablation experiments

实验序列	EfficientViT	C2f-Faster	MLCA	Precision/%	Recall/%	mAP50/%	模型大小/M
1				93.3	93.4	96.8	10.5
2	✓			95.6	92.0	98.1	12.5
3		✓		95.8	93.4	97.8	5.9
4			✓	96.3	92.4	97.3	10.6
5	✓		✓	95.2	94.0	98.4	15.3
6		✓	✓	94.8	96.2	98.2	8.1
7	✓	✓	✓	95.3	97.7	98.9	9.7

由实验 1 和实验 2 对比结果可知,采用 EfficientViT 更换主干网络之后,模型的 mAP50 得到 1.3% 的提升,模型大小增加了 2 M。由实验 1 和实验 3 对比结果可知,使用 FasterBlock 模块替换 C2f 中的 Bottleneck 模块,mAP50 提升了 1%,模型大小下降 4.6 M。由实验 1 和实验 4 对比结果可知,单独增加 MLCA 注意力模块,mAP50 仅提升了 0.5%,但是模型的 P 却增加了 3.3%,模型大小增加了 0.1 M。由实验 2 和实验 5 对比可知,更换主干网络并且添加 MLCA 注意力模块相比单独更换主干网络,模型检测精度更高,但模型却更大。由实验 3 和实验 6 对比结果可知,单独使用 FasterBlock 模块替换 C2f 中的 Bottleneck 模块相比较基础上再增加 MLCA 注意力模块,模型更小,但是 mAP50 也更低。由实验 1 和实验 7 对比结果可知,经过更换主干网络,改进 C2f 模块,增加注意力模块后,mAP50 提升了 2.1%,模型的 Recall 也达到了 97.7%,提升了 4.3%,模型大小下降了 0.8 M。由此可以认为,改进之后的模型在检测精度有所提升,并且有更小的参数量,可以高效准确完成驾驶员的分心驾驶检测。

2.4 对比实验

为了验证改进模型的性能,将本文的 YOLOv8-EFM 与 VGG、EfficientNet、MobileNetv2、ResNet50、SwinTransformer 进行了对比,不同网络的精度曲线图如图 10 所示,不同网络的损失曲线图如图 11 所示。可以看出,YOLOv8-EFM 的性能最强,超过了 MobilNetv2 和 SwinTransformer,精度和损失分别为 98.6% 和 0.10。表 3 总结了不同网络的 mAP50 和模型参数量,其中 YOLOv8-EFM 最高为 98.9%,之后依次为 SwinTransformer、MobileNetv2、ResNet50、EfficientNet,而 VGG 最低,仅有 87.33%,且 YOLOv8-EFM 和 MobileNetv2 模型都相对较小一些,分别是 9.7 M 和 7.8 M。所以可以得到结论,本文所改进模型不仅准确率高,并且模型大小也合适,方便后期进行嵌入式边缘部署。

2.5 实验结果可视化

为了进一步展示改进模型的性能,对于各类驾驶行为

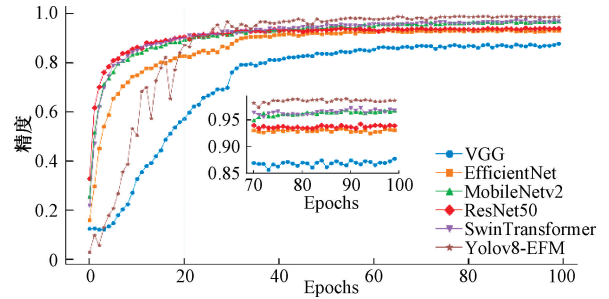


图 10 不同网络的精度曲线图

Fig. 10 Precision plots of the different networks

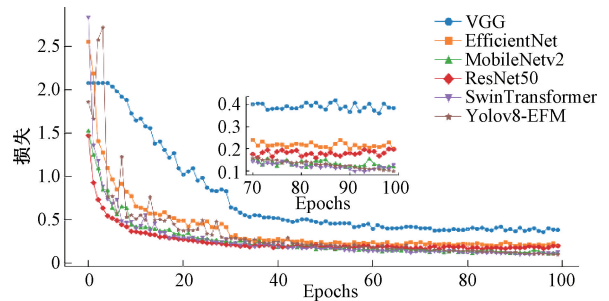


图 11 不同网络的损失曲线图

Fig. 11 Loss plots of the different networks

表 3 不同网络的 mAP50 平均精确率和模型参数量

Table 3 Average mAP 50 accuracy and number of model parameters for the different networks

模型	mAP50/%	模型大小/M
VGG	87.33	233.1
EfficientNet	92.05	32.7
MobileNetv2	97.15	7.8
ResNet50	93.39	188.8
SwinTransformer	97.46	40.0
YOLOv8-EFM	98.90	9.7

的预测结果进行了可视化,如图 12 所示,左半部分为具体分心行为预测结果以及关键点连线图,右半部分则是热图

展示。从图中可以看到,9个关键点都可以预测出来,基本准确,置信度均在0.9以上,即使图12(i)图中拿后座东西

导致看不到右肘以及右手臂,但是其他7个关键点依旧可以准确预测。

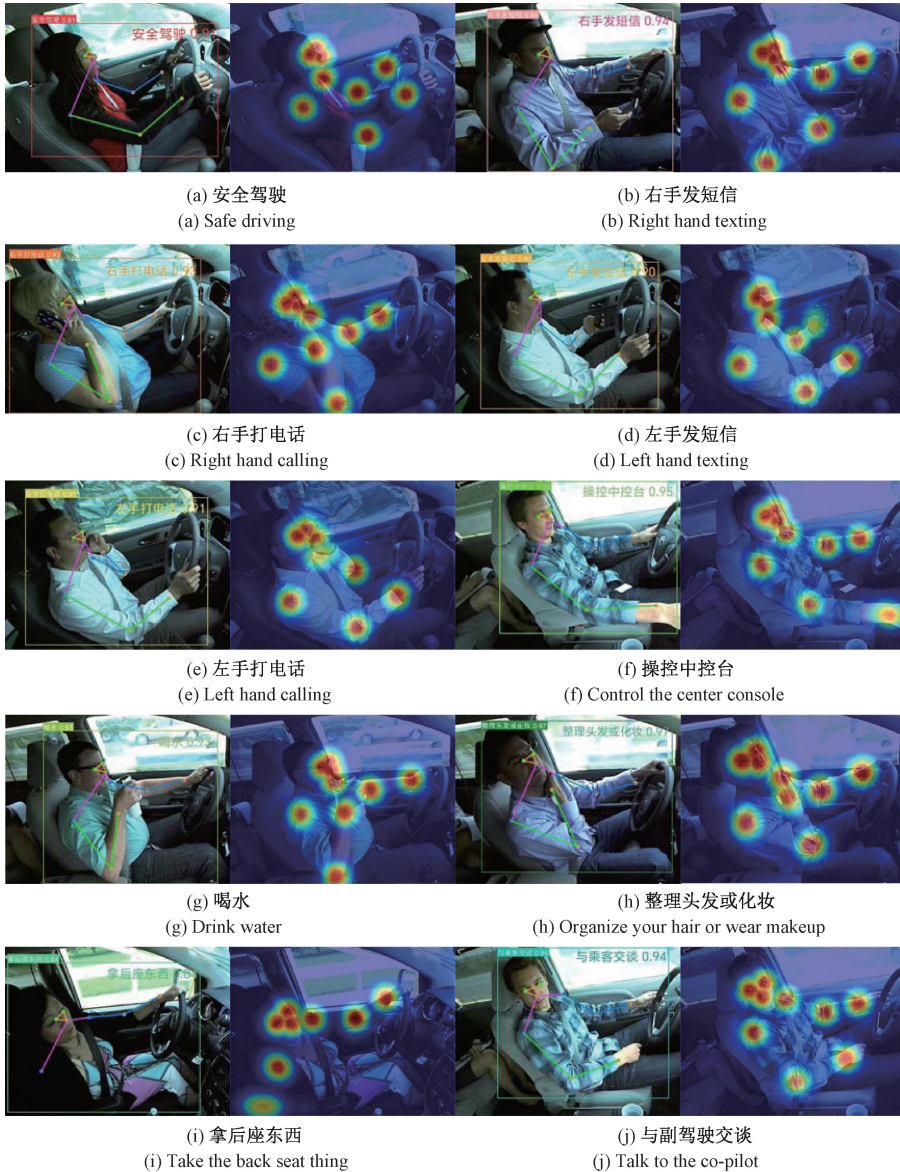


图12 模型预测可视化

Fig. 12 Model prediction visualization

3 结 论

由于分心驾驶的危害性较大,所以对于驾驶员分心驾驶的正确检测需求迫切。因此针对这一问题本文构建了YOLOv8-EFM分心驾驶检测与识别模型,不仅能够标记驾驶员的上半身关键点,同时可以正确检测具体分心行为,实验结果也表明经过改进之后的模型在保证准确率提高的同时不仅提升了检测速率,并且模型参数量也有所下降,满足对分心驾驶的检测要求。但还存在一些问题亟待进一步的研究,本次只研究了驾驶员的9种分心行为,在实际生活场景中并不止于此,因此需要构建更全面的数据

集;本文对驾驶员分心驾驶行为的检测基本都是正常光照条件下,但在夜间驾驶时车内光照不足,所以还需研究夜间行车时的检测方法;从实用性出发,下一步应将模型部署到移动设备或者嵌入式设备上完成实时检测。

参考文献

- [1] 娄文,郭杜柱,张杰,等. 基于YOLOv7的驾驶人使用手机与抽烟行为识别方法[J]. 电子测量技术, 2023, 46(21):123-131.
- LOU W, GUO D D, ZHANG J, et al. Identification method of mobile phone use and smoking behavior of

- drivers based on YOLOv7 [J]. *Electronic Measurement Technology*, 2023, 46(21):123-131.
- [2] 禹江林,张云.一种基于多因素融合的驾驶预警方法[J].*电子测量技术*,2021,44(11):103-108.
YU J L, ZHANG Y. A driving early warning approach based on multi-factor fusion[J]. *Electronic Measurement Technology*, 2021, 44(11): 103-108.
- [3] 陈骥驰,王宏,王翘秀,等.基于脑电信号的疲劳驾驶状态研究[J].*汽车工程*,2018,40(5):515-520.
CHEN J CH, WANG H, WANG Q X, et al. Study on fatigue driving status based on EEG signals[J]. *Automotive Engineering*, 2018, 40(5): 515-520.
- [4] SIGARI M H, POURSHAHABI M R, SORYANI M, et al. A review on driver face monitoring systems for fatigue and distraction detection[J]. *International Journal of Advanced Science and Technology*, 2014, 64: 73-100.
- [5] 高绮焯,谢凯,贺正方,等.复杂环境下多模态特征融合的疲劳驾驶检测[J].*电子测量技术*,2023,46(6):106-115.
GAO Q H, XIE K, HE ZH F, et al. Fatigue driving detection with multi-modal feature fusion in complex environments [J]. *Electronic Measurement Technology*, 2023, 46(6): 106-115.
- [6] ZHANG W, ZHANG H. Research on distracted driving identification of truck drivers based on simulated driving experiment [C]. *IOP Conference Series: Earth and Environmental Science*, 2021, 638(1): 012039.
- [7] 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.
- [8] HSSAYENI M D, SAXENA S, PTUCHA R, et al. Distracted driver detection: Deep learning vs handcrafted features [J]. *Electronic Imaging*, 2017, 29: 20-26.
- [9] ALMAZROI A A, ALQARNI M A, ASLAM N, et al. Real-time CNN-base driver distraction & drowsiness detection system [J]. *Intelligent Automation & Soft Computing*, 2023, 37(2): 2153-2174.
- [10] CHEN J, JIANG Y N, HUANG ZH X, et al. Fine-grained detection of driver distraction based on neural architecture search [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(9): 5783-5801.
- [11] 李富,徐凯,朱灵龙,等.基于上下文语义联合YOLOv7的分心驾驶检测算法[J].*国外电子测量技术*,2024,43(3):121-128.
LI F, XU K, ZHU L L, et al. Distracted driving detection algorithm based on context-semantic joint YOLOv7 [J]. *Foreign Electronic Measurement Technology*, 2024, 43(3): 121-128.
- [12] 钟可佳.基于深度学习的分心驾驶行为检测方法研究[D].南昌:南昌大学,2023.
ZHONG K J. Research on the distracted driving behavior detection method based on deep learning[D]. Nanchang: Nanchang University, 2023.
- [13] SONG L CH, YU G, YUAN J S, et al. Human pose estimation and its application to action recognition: A survey [J]. *Journal of Visual Communication and Image Representation*, 2021, 76: 103055.
- [14] SHAHROUDY A, LIU J, NG T T, et al. NTU RGB+D:A large scale dataset for 3D human activity analysis[C]. *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2016: 1010-1019.
- [15] 夏瀚笙.基于Stacked Hourglass网络的二维人体骨骼关键点检测[D].南京:南京航空航天大学,2019.
XIA H SH. Two-dimensional human bone critical point detection based on the Stacked Hourglass network [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019.
- [16] JIAO SH J, LIU L Y, LIU Q. A hybrid deep learning model for recognizing actions of distracted drivers[J]. *Sensors*, 2021, 21(21):7424.
- [17] 张志威.基于机器视觉的异常驾驶行为检测方法研究[D].长沙:湖南大学,2020.
ZHANG ZH W. Research on the detection method of abnormal driving behavior based on machine vision[D]. Changsha: Hunan University, 2020.
- [18] 陈帅超.基于人体关键点检测的分心驾驶行为识别算法研究[D].成都:电子科技大学,2023.
CHEN SH CH. Research on the distracted driving behavior recognition algorithm based on human key point detection[D]. Chengdu: University of Electronic Science and Technology, 2023.

- [19] 李一凡,袁龙健,王瑞. 基于 OpenPose 改进的轻量化人体动作识别模型[J]. 电子测量技术, 2022, 45(1): 89-95.
- LI Y F, YUAN L J, WANG R. A improved lightweight human action recognition model based on OpenPose[J]. Electronic Measurement Technology, 2022, 45(1): 89-95.
- [20] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [21] CHEN J R, KAO SH H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural network[C]. 2023 IEEE/CVF Conference on Computer Vision and

Pattern Recognition(CVPR), 2023: 12021-12031.

作者简介

朱周华, 副教授, 博士研究生, 主要研究方向为图像处理、模式识别等。

E-mail: 514210183@qq.com

侯智杰(通信作者), 硕士研究生, 主要研究方向为深度学习、目标检测等。

E-mail: sy913589711@163.com

田成源, 硕士研究生, 主要研究方向为深度学习、图像处理等。

E-mail: 614489450@qq.com

周怡纳, 硕士研究生, 主要研究方向为机器学习、目标检测等。

E-mail: 875321730@qq.com