

# 基于改进 YOLOv10 的驾驶疲劳检测算法 \*

殷旭鹏<sup>1,2</sup> 赵兴强<sup>1,2</sup> 王雄飞<sup>1</sup> 阮琪<sup>1</sup> 张莞<sup>1,2</sup>

(1. 南京信息工程大学自动化学院 南京 210044; 2. 南京信息工程大学江苏省大数据分析技术和智能系统省高校重点实验室 南京 210044)

**摘要:** 疲劳驾驶的检测对于确保交通安全极为重要, 实时监测并识别驾驶员的疲劳程度, 配合预警机制, 可有效降低因疲劳驾驶导致的事故风险。针对目前疲劳驾驶检测过程中驾驶员表情特征目标小、背景环境复杂的问题, 提出了一种基于 YOLOv10 改进的疲劳驾驶检测模型—YOLOv10-GMF。该模型引入全局分组坐标注意力模块(global grouped coordinate attention, GGCA), 通过分组后的全局信息与局部特征处理, 生成加权注意力图, 实现信息压缩与特征表达, 提升模型对疲劳状态下微小神态特征的捕捉能力。同时添加多尺度空洞融合模块(multi-dimension fusion attention, MDFA), 利用多尺度空洞卷积, 并行融合空间和通道注意力机制, 有效加强模型在复杂驾驶环境中对图像特征的识别能力。此外, 为进一步优化训练过程, 还设计了反馈信息驱动损失函数(feedback-driven loss, FDL), 有效加速模型的收敛过程, 提高模型的检测效率。经过对比与消融实验, 改进后的 YOLOv10-GMF 模型的检测平均精度均值(mAP)可达到 98.1%, 较 YOLOv10 提升了 14.5%, 且检测速度为 64.3 fps。通过实际车载嵌入式部署测试, 整个疲劳检测过程总耗时 19.0 ms, 完全满足驾驶过程中对疲劳状态进行实时监测的需求。

**关键词:** 小目标检测; 驾驶疲劳检测; YOLOv10; 注意力机制; 损失函数

中图分类号: TP391.4; TN911.7 文献标识码: A 国家标准学科分类代码: 520.2

## Improved YOLOv10 algorithm for driver fatigue detection

Yin Xupeng<sup>1,2</sup> Zhao Xingqiang<sup>1,2</sup> Wang Xiongfei<sup>1</sup> Ruan Qi<sup>1</sup> Zhang Wan<sup>1,2</sup>

(1. School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China;  
2. Jiangsu Province Key Laboratory of Big Data Analysis Technology and Intelligent Systems for Universities,  
Nanjing University of Information Science & Technology, Nanjing 210044, China)

**Abstract:** Fatigue driving detection is critical for traffic safety. Real-time monitoring and accurate identification of a driver's fatigue level, coupled with an early warning system, can significantly reduce the risk of accidents caused by fatigue. Addressing the challenges of small micro-expression targets and complex background environments in current driving fatigue detection, this paper proposes an improved driving fatigue detection model—YOLOv10-GMF. The model incorporates an enhanced global grouped coordinate attention (GGCA) module, which improves feature representation by weighting feature maps with global information and generating attention maps, thereby enhancing the model's ability to capture micro-expression features under fatigue conditions. Additionally, a multi-dimension fusion attention (MDFA) module is integrated, which combines multi-scale dilated convolutions with spatial and channel attention mechanisms in parallel to strengthen the model's recognition ability for image features in complex driving environments. To further optimize the training process, a feedback-driven loss function (FDL) is introduced, effectively accelerating model convergence and improving prediction accuracy. Ablation experiments demonstrate that the YOLOv10-GMF model achieves a detection accuracy of 98.1%, a 14.5% improvement over YOLOv10, with a detection speed of 64.3 fps. Through real vehicle embedded deployment tests, the average fatigue detection process takes 19.0 ms, and the model fully meets the real-time monitoring needs for fatigue driving.

**Keywords:** dim targets detection; driver fatigue detection; YOLOv10; attention mechanism; loss function

## 0 引言

2021 年世界卫生组织调查发现,每年约 130 万人死于道路交通事故。到 2024 年,交通事故已成为全球第二大伤亡原因。而疲劳驾驶是其主要诱因之一。预防驾驶疲劳对于保障道路安全至关重要。如何有效识别和评估疲劳驾驶状态从而及时预警,已成为重要课题。

早期研究者尝试借助行车数据来判断驾驶员的疲劳状态。Alkinani 等<sup>[1]</sup>通过监测车辆行驶轨迹和方向盘压力数据来检测驾驶员的疲劳情况,但这类方法易受驾驶员习惯和环境因素干扰,准确率较低。

另一类方法基于生理参数,通过专用的信号采集仪接收前额、心脏或肌肉部位电极贴片发出的信号<sup>[2]</sup>,并传输至检测系统进行实时分析。Chui 等<sup>[3]</sup>提出的一种基于支持向量机(SVM)算法的脑电信号处理方法,可以将检测延迟降低到 0.55 ms 以内。但该类方法在实际操作时需要借助外部昂贵的信号采集设备,设备的一端必须与人体皮肤相连,会制约驾驶员的驾驶动作,且会占用部分车内空间,无形中加剧了事故的发生。

随着计算机视觉技术的发展,研究人员开始利用视觉方法检测疲劳驾驶。较之前两种方法,视觉方法具有成本低、非侵入的特点,对驾驶员的正常驾驶过程没有干扰,具有极大优势。根据处理流程的不同,视觉检测方法可分为一阶段和二阶段两类。一阶段算法如 SSD<sup>[4]</sup>、YOLO<sup>[5]</sup>,而二阶段算法则主要包括卷积神经网络(CNN)<sup>[6-8]</sup>等方法。胡习之等<sup>[9]</sup>通过优化 SSD 算法和连续自适应均值漂移跟踪算法来检测人脸区域,但其算法计算量庞大,无法进行嵌入式部署。秦康等<sup>[10]</sup>针对驾驶员相似背景下的细微动作的识别过程,结合知识蒸馏方法,提出了一种基于 X3D 卷积神经网络 X3D-M-GCAE,可以适配车载硬件,但其检测精度仍存在较大提升空间。闫保中等<sup>[11]</sup>基于改进的局部二值特征法计算人眼闭合程度和视线方向判断驾驶员注意力是否分散,平均检测精度达到了 93.9%,但需要大量的数据进行训练,耗时较大且成本高昂。Ansari 等<sup>[12]</sup>提出一种基于线性修正单元层的双向长短期记忆(LSTM)疲劳检测神经网络,平均检测精度较高,但处理过程不够自动化,过程较为繁杂,不利于移动终端的移植。

近几年,YOLO 系列算法<sup>[13-16]</sup>凭借其高效的实时检测能力,已成为视觉检测领域的研究热点。Zhou 等<sup>[17]</sup>利用改进的 YOLOv8 网络将疲劳检测精度提高到 92.3%,检测帧率提高到 39.4 fps,但仍存在参数量较大、面对复杂样本时泛化能力不足的问题,检测效率还可以进一步提升。YOLOv10 作为该系列中的新版本,在驾驶疲劳检测任务中,能够兼顾检测精度与速度,精准识别并锁定驾

驶员的眼部、嘴部等关键面部特征,以便依据疲劳驾驶判定指标进一步做出准确判断。

在如何选择更加合理的疲劳驾驶判定指标方面,国内外研究者开展了大量相关研究。1994 年,美国联邦公路管理局和美国国家公路交通安全管理局通过模拟真实驾驶环境,对 9 种疲劳检测指标进行了比较,发现驾驶员的眼睛闭合程度与疲劳状态的关系最为密切。基于此,卡内基梅隆大学提出了“PERCLOS”(percentage of eyelid closure over time)指标,即单位时间眼睛闭合所占比例,该指标被广泛应用于衡量驾驶者是否疲劳。近年来,我国也推出了多个疲劳驾驶检测系统的商用产品。例如,比亚迪的 BAWS 视觉系统通过关注驾驶员的眼球运动,形成了眼球运动重复率指标;虹软科技则将目光方向与聚焦点位置量化为目光移动指标;清研微视研发的疲劳驾驶视觉预警系统通过判断驾驶员的头部摆动角度与频率指标,实现对疲劳驾驶的实时预警。这些指标在疲劳驾驶状态的识别和判定方面已显示出较高的准确性,为后续研究提供了有益的参考。然而,需要指出的是,单一指标用于判断疲劳状态可能存在偶然性问题。因此,本文尝试引入了眼部与嘴部多指标综合分析的方法,以更全面、准确地判断驾驶员的疲劳状态。

针对实际行车过程中驾驶员面部微表情特征目标小且背景环境复杂的问题,本文提出了一种基于 YOLOv10 改进的疲劳检测算法。引入全局分组坐标注意力模块(global grouped coordinate attention, GGCA),利用分组特征图在空间维度上的全局信息经压缩后生成注意力图。嵌入多尺度空洞融合模块(multi-dimension fusion attention, MDFA),提高模型对各种小尺度特征的适应性和识别能力。最后,优化原始模型的损失函数为反馈信息驱动损失函数(feedback-driven loss, FDL),并进行嵌入式部署和测试,通过实验测试结果证明改进算法成功提升模型精度及检测速度,适合应用于实时疲劳驾驶检测场景。

## 1 改进 YOLOv10 算法

### 1.1 YOLOv10 算法

YOLOv10 采用轻量级分类头、空间通道解耦下采样和排序引导块设计,以减少明显的计算冗余。传统的 YOLO 模型使用非极大值抑制(NMS)来过滤重叠的预测,这增加了推理延迟。而 YOLOv10 引入了一种双重分配策略,消除了 NMS 的需求,从而实现更快、更高效的目标检测。其网络结构由 4 个主要模块组成,即输入预处理层、主干网络、颈部网络和输出端。输入预处理层负责处理输入的图片数据,包括数据增强与归一化等预处理操作。主干网络部分采用改进的 CSPDarknet 网络结构,

以进一步增强特征提取能力。颈部网络部分使用改进的 FPN+PAN 架构,以提升多尺度目标检测的效果。最后,输出端集成了多个检测头,采用 Anchor-Free 设计,实现更精准的目标定位和分类。

## 1.2 改进的 YOLOv10 算法

本文在 YOLOv10 算法基础上作了 3 个方面改进:首先,在 YOLOv10 的 Backbone 部分嵌入创新设计的全局分组坐标注意力机制模块 GGCA,解决单一的注意力机制<sup>[18]</sup>在捕捉特征信息时有局限性,不能同时利用多维度的全局信息的问题;然后,针对当前广泛使用并行通道空

间注意力模块<sup>[19]</sup>(convolutional block attention module, CBAM),在疲劳驾驶检测的应用中存在部分特征的过拟合问题且在不同尺度图像特征捕捉方面存在局限性<sup>[20]</sup>,本文在 CBAM 的基础上进行改进,提出了多尺度空洞融合模块 MDFA;最后,本文将 YOLOv10 的损失函数替换为 FDL 损失,在训练过程中动态调节模型的优化方向,帮助模型在不同阶段集中关注最关键特征,从而优化模型训练过程并最终提高检测精度。改进后的模型整体的算法框架如图 1 所示。

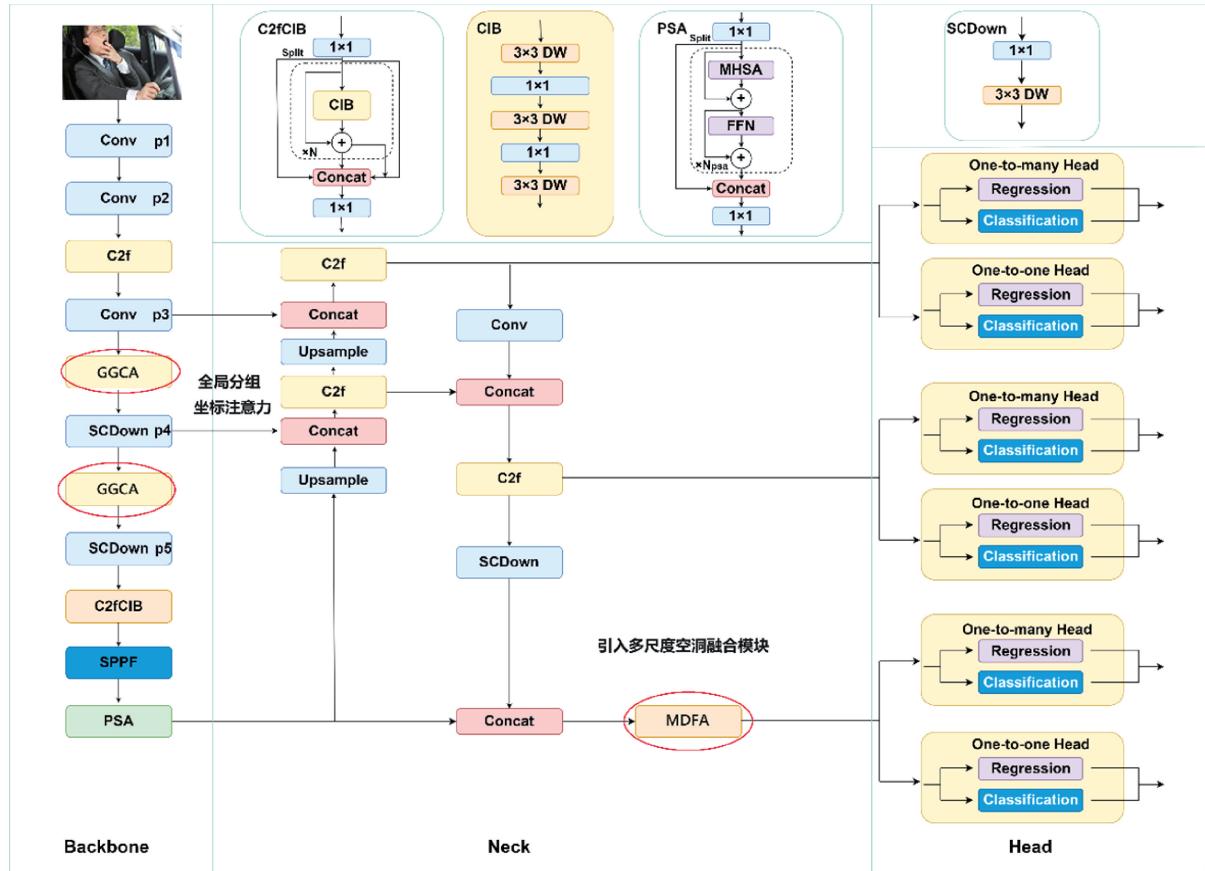


图 1 改进 YOLOv10-GMF 算法结构

Fig. 1 Improved YOLOv10-GMF algorithm structure

### 1) GGCA 模块结构与机理分析

GGCA 模块旨在利用特征图在空间维度上的全局信息来生成注意力图,并通过这些注意力图对输入特征图进行加权,增强特征表达能力。该模块的具体结构如图 2 所示。

设置训练过程中每次迭代中用于计算梯度和更新模型参数的样本特征图数量为  $B$ 。特征图的高度和宽度分别为  $H$  和  $W$ ,最初通过  $C$  个通道输入网络。即对于输入特征图  $X$ ,可表示为:

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

首先,按照通道数将输入特征图分为  $G$  组,每组包含  $C/G$  个通道。旨在将样本图中的高维特征划分为多个小维度组,以保留局部通道信息。分组后的特征图表示为:

$$X \in R^{B \times G \times \frac{C}{G} \times H \times W} \quad (2)$$

这一预先分组操作使得每组特征可以在后续分别进行空间维度的注意力计算,从而提高注意力的计算效率和泛化能力。分组后,为兼顾特征图的整体强度分布与局部显著性,在高度方向和宽度方向对其进行全局平均池化(AvgPool)和全局最大池化(MaxPool)操作,

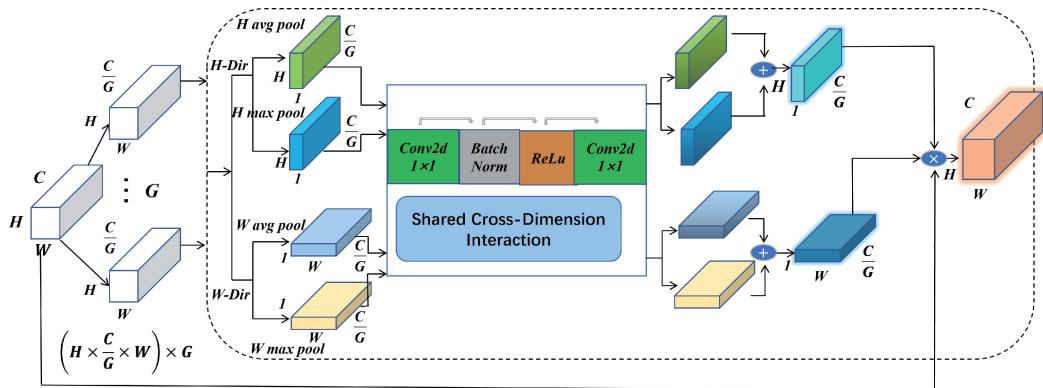


图 2 GGCA 网络结构

Fig. 2 GGCA network architecture

得到:

$$X_{h,\text{avg}} = \text{AvgPool}(X) \in R^{B \times G \times \frac{C}{G} \times H \times 1} \quad (3)$$

$$X_{h,\text{max}} = \text{MaxPool}(X) \in R^{B \times G \times \frac{C}{G} \times H \times 1} \quad (4)$$

$$X_{w,\text{avg}} = \text{AvgPool}(X) \in R^{B \times G \times \frac{C}{G} \times 1 \times W} \quad (5)$$

$$X_{w,\text{max}} = \text{MaxPool}(X) \in R^{B \times G \times \frac{C}{G} \times 1 \times W} \quad (6)$$

式中:  $X_{h,\text{avg}}$  表示在高度方向上对分组后的特征图进行全局平均池化操作得到的特征图, 该操作会将每个通道的高度维度压缩为一个特征值, 保留整体强度分布信息;  $X_{h,\text{max}}$  表示在高度方向上对分组后的特征图进行全局最大池化操作得到的特征图, 该操作会提取每个通道在高度方向上的最大值, 突出局部显著性特征。在宽度方向上得到的  $X_{w,\text{avg}}$  与  $X_{w,\text{max}}$  同理。

全局分组池化后, 将上述 4 个池化结果输入至共享卷积层中, 进行信息压缩与特征提取操作。共享卷积层由两个  $1 \times 1$  卷积层、批量归一化层和 ReLU 激活函数组成, 两个卷积层分别用于降低和恢复通道维度。通过共享卷积操作, 在高度和宽度方向上得到的特征结果可表示为:

$$Y_{h,\text{avg}} = \text{Conv}(X_{h,\text{avg}}) \quad (7)$$

$$Y_{h,\text{max}} = \text{Conv}(X_{h,\text{max}}) \quad (8)$$

$$Y_{w,\text{avg}} = \text{Conv}(X_{w,\text{avg}}) \quad (9)$$

$$Y_{w,\text{max}} = \text{Conv}(X_{w,\text{max}}) \quad (10)$$

进而将共享卷积层的输出相加, 应用 Sigmoid 激活函数  $\sigma$ , 生成高度方向和宽度方向的注意力权重图:

$$A_h = \sigma(Y_{h,\text{avg}} + Y_{h,\text{max}}) \in R^{B \times G \times \frac{C}{G} \times H \times 1} \quad (11)$$

$$A_w = \sigma(Y_{w,\text{avg}} + Y_{w,\text{max}}) \in R^{B \times G \times \frac{C}{G} \times 1 \times W} \quad (12)$$

最后, 将输入特征图按注意力权重进行加权, 得到特征提取后的输出特征图  $O$ :

$$O = X \times A_h \times A_w \in R^{B \times C \times H \times W} \quad (13)$$

GGCA 模块利用分组注意力机制和共享卷积层, 生

成高度和宽度方向的注意力图的同时, 实现信息压缩与特征加权。使得模型能更好地关注局部特征, 增强疲劳驾驶图像中眼部、嘴部等重要特征的表达, 并抑制驾驶室内环境背景等噪声的影响。针对性解决微小神态特征目标占总图比例小, 容易被环境背景、光线噪声等干扰的问题。通过增强感受野提升了模型的提取能力与融合效果。

## 2) MDFA 模块结构与机理分析

MDFA 模块设计的核心在于结合多尺度特征提取和注意力机制, 针对性解决复杂视觉任务如实时驾驶特征检测中局部细节和全局上下文信息的综合问题。

模块分为两部分, 第 1 部分是多尺度空洞卷积部分, 第 2 部分是通道和空间注意力的融合。这种结构设计既保留了不同尺度的信息, 又通过注意力机制显著提高了特征表达能力。

多尺度空洞卷积结构通过 5 个并行的卷积分支来实现不同尺度的特征提取, 每个分支配置有不同的空洞率。第 1 分支, 使用  $1 \times 1$  卷积核, 主要作用是线性投影与特征压缩; 第 2~4 分支, 使用  $3 \times 3$  卷积核, 空洞率分别为 6、12、18, 空洞率的渐增设置保证了感受野从局部到全局的递进式扩展, 捕获不同范围的空间关系; 第 5 分支, 使用一个额外的全局平均池化分支在多尺度特征中引入全局背景, 有效提高对整体语义的理解能力。

通道和空间特征合并与校准通过通道维度拼接, 形成一个综合的特征图, 整合各个分支的特征, 提升对目标的全方位理解能力。在处理复杂的驾驶图像时, 某些通道可能包含比其他通道更有用的信息。传统的卷积操作不能动态地调整通道的重要性。此外, 不同的空间位置包含的信息重要性可能不同。例如, 目标对象的区域比背景更重要等。本文引入的 MDFA 模块通过多尺度空洞卷积的设计捕捉驾驶场景中的细微特征(如眼部、嘴部的小幅变化)。再调整通道注意力, 突出关键信息通道; 同时应用空间注意力, 进一步聚焦目标区域。通过动态权

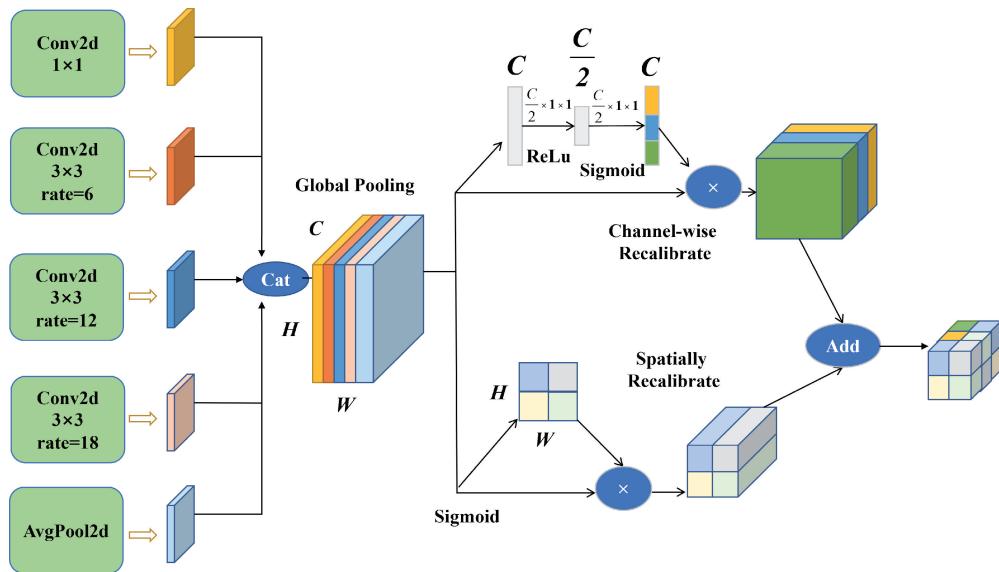


图 3 MDFA 网络结构

Fig. 3 MDFA network architecture

重实时校准通道与空间注意力机制的使用,突出驾驶员的脸部特征等关键区域,抑制背景噪声(如驾驶室内环境、窗外光线干扰等),帮助网络更好地理解图像中的每一个部分与整体的关系。从而确保即使在特征稀疏或目标特征微弱的情况下,MDFA 模块也能表现出较好的鲁棒性和特征表达能力。

### 3) 损失函数设计

在疲劳驾驶检测中,驾驶员的微表情特征具有较小的图像面积,同时背景环境中包含大量的动态变化和干扰信息(如车内景物、光线变化等),这给模型的训练优化过程带来了不确定性。原 YOLOv10 算法的损失函数以及目前主流的损失函数无法针对驾驶员的微小表情特征,动态优化模型收敛过程,训练过程中会出现过度关注背景噪声而出现误检和漏检的问题。

为此,本文提出使用反馈信息驱动的 FDL 损失函数优化模型训练过程,该损失函数可以动态调控检测过程小目标损失在总体损失中的权重,根据反馈信息自适应调整模型的学习策略,引导模型更加专注于驾驶员面部的微表情特征,并减小背景噪声的干扰。从而帮助模型最终更准确、快速地完成疲劳驾驶检测。

设总体损失为  $Loss_{(total)}$ , 小目标损失为  $Loss_{(small)}$ , 并定义反馈系数  $\tau$ , 并将其引入高斯函数  $f(\tau)$  中, 用以实时计算损失函数中的时间依赖性, 得到最终的反馈驱动损失函数:

$$\tau = \frac{Loss_{(small)}}{Loss_{(total)}} \quad (14)$$

$$Loss(location) = \lambda \sum_{i=0}^{s^2} \sum_{j=0}^B I_{ij}^{obj} f(\tau) [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] +$$

$$\lambda \sum_{i=0}^{s^2} \sum_{j=0}^B I_{ij}^{obj} f(\tau) [(w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2] \quad (15)$$

其中,  $\lambda$  为收敛比例系数, 用于在宏观上实时调整损失函数中不同部分的权重, 以平衡模型训练过程中的损失分布。如图 4 所示,  $(x_i, y_i, w_i, h_i)$  用以描述真值框, 而  $(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$  用以描述预测框。 $x_i$  和  $y_i$  表示真值框的中心点坐标,  $w_i$  与  $h_i$  分别表示真值框的宽度和高度。 $\hat{x}_i$  和  $\hat{y}_i$  表示预测框的中心点坐标,  $\hat{w}_i$  与  $\hat{h}_i$  分别表示预测框的宽度和高度。 $s$  表示网格的大小, 计算中将待处理的图像划分为  $s \times s$  的网格。 $B$  表示网格预测的边界框的数量,  $I_{ij}^{obj}$  是指示函数, 如果单元  $i$  中的第  $j$  个边界框包含对象, 则为 1, 否则为 0。

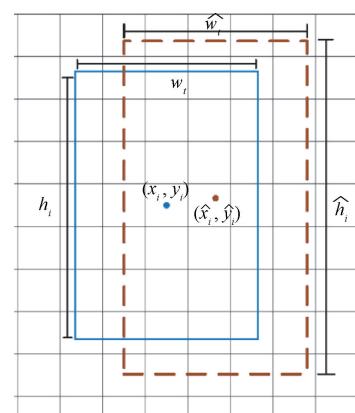


图 4 检测边界框示意图

Fig. 4 Schematic diagram of detecting bounding boxes

FDL 损失函数可以综合考虑训练过程小目标特征真

值框和预测框的重叠程度、中心点距离和形状等因素,通过引入反馈参数,利用损失分布信息作为反馈信号,实时指导模型进行下一步更新,增强模型对复杂环境中小目标检测的泛化能力,提高检测效率。

### 1.3 疲劳驾驶判定指标设计

针对疲劳检测中因单一评价指标导致检测结果不准确的问题,本文引入眼宽高比(eye aspect ratio, EAR)与嘴宽高比(mouth aspect ratio, MAR)双指标,通过监测驾驶员在一定时间内的眨眼频率和打哈欠次数来判断其是否疲劳。

首先利用改进的 YOLOv10-GMF 网络模型定位脸部并标记眼睛和嘴部特征,以匹配关键点。常规人脸关键点有 68 个,但鉴于实际检测中过多关键点不利于模型轻量化,且仅需眼睛和嘴巴信息即可判断疲劳状态,本文选其中 18 个关键点评估疲劳状态:左右眼共 12 个(编号 37~48),嘴部 6 个(编号 49、51、53、55、57、59),如图 5 所示。

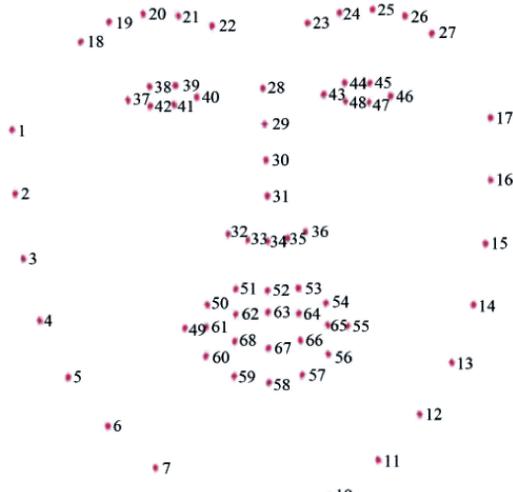


图 5 人脸关键点

Fig. 5 Face key points

根据眼部关键点的  $x$ 、 $y$  坐标得出眼睛的长宽比 EAR。参考 PERCLOS 的 P80 标准,当眼睑覆盖瞳孔面积超 80% 时,视为眼睛闭合,据此将 EAR 阈值设为 0.2,即  $EAR < 0.2$  时,判定眼睛闭合,公式如下:

$$EAR = \frac{\|y_{38} - y_{42}\| + \|y_{39} - y_{41}\|}{4\|x_{40} - x_{37}\|} + \frac{\|y_{44} - y_{48}\| + \|y_{45} - y_{47}\|}{4\|x_{46} - x_{43}\|} \quad (16)$$

$$f_e = \frac{t_e}{T_e} \times 100\% \quad (17)$$

通过界定  $f_e$  来评估驾驶员的眼部疲劳状况,其中  $t_e$  是检测时段内的闭眼帧数,  $T_e$  是检测时段的总帧数。通

常情况下,单次眼睑闭合时长在 0.1~0.2 s;疲劳状态下,该时长会超过 0.45 s。所以,当  $f_e$  的阈值被设定为 0.45 时,即单位时间内闭眼频率超过 0.45,便可判断驾驶员处于疲劳状态。

类比眼部状态指标,嘴部状态指标通过嘴部关键点的  $x$ 、 $y$  坐标计算 MAR。定义  $f_m$  来判断驾驶员嘴部疲劳状态,该指标通过计算嘴部关键点的欧氏距离得出,公式如下:

$$MAR = \frac{\|y_{51} - y_{59}\| + \|y_{53} - y_{57}\|}{2\|x_{55} - x_{49}\|} \quad (18)$$

$$f_m = \frac{t_m}{T_m} \times 100\% \quad (19)$$

式中:  $t_m$  表示检测时间内张嘴的帧数;  $T_m$  表示检测时间的总帧数。一般情况下,人打哈欠的持续时间为 4 s。设定单位检测时间为 30 s,那么在 30 s 内打哈欠超过两次(即累计 8 s)则视为疲劳驾驶。因此,本文将  $f_m$  的参数阈值设置为 0.27,即单位时间内打哈欠频率  $> 0.27$  时,可判断驾驶员处于疲劳状态。

## 2 实验与分析

### 2.1 数据集

本文采用公开数据集 Drowsy-Driving-Det1-AFLW 和 CEW。前者来源于国外开源数据集,约 25 000 张;后者来源于国内开源数据集,约 2 000 张。这两个数据集合并后约共 27 000 张,包含不同姿态、表情、光照和种族等因素影响的驾驶员行车照片,十分符合疲劳驾驶实际的应用场景。

为了使模型更好地适应实际行车中的复杂环境,本文在原始数据集的基础上,随机选取部分图像进行了包含亮度调整、噪声添加、旋转变换、镜像处理的一系列预处理操作。具体而言,亮度调整是通过将图像亮度随机设置为原始亮度的 0.8~1.2 倍来模拟不同的光照条件;噪声添加方面,随机向图像中添加均值为 0、标准差在 0.05~0.15 的高斯噪声,或者密度为 0.05~0.2 的椒盐噪声;旋转变换则是  $-20^\circ$ ~ $+20^\circ$  的范围内,以  $2^\circ$  为步长对图像进行旋转;镜像处理通过随机对图像进行水平或垂直方向的镜像翻转来增加数据多样性。

经过上述操作,最终生成了 40 000 张驾驶疲劳检测图像。随后,使用 LabelImg 工具对这些图像进行标注,依照前文所述 EAR 和 MAR 的判定标准将图像分为清醒(Awake)和疲倦(Fatigue)两类,并按照 3:1 的比例将数据集划分为训练集和测试集,分别用于模型的训练和评估。

### 2.2 实验环境和算法评价指标

实验模型的训练与参数优化过程在配置有 4 块



图 6 疲劳驾驶数据集

Fig. 6 Drowsy driving datasets

GeForce RTX 3050Ti 显卡的专用服务器上进行, 操作系统为 Ubuntu, 使用 PyTorch 框架。模型训练的超参数设置为学习率 0.01, 训练轮次 300, 批量大小 16, 优化器选择 SGD。

学习率影响模型收敛与性能。学习率过高导致模型梯度发散, 过低则延长训练时间。本文通过预实验得出, 0.01 的学习率在本文的疲劳检测任务中可平衡收敛速度与稳定性, 且保证较高的检测精度。

训练迭代次数取决于数据集规模和模型复杂度。本文数据集含约 40 000 张图像, 300 轮迭代可充分学习特征并防止过拟合。

批次大小受显存和泛化能力限制。在本研究的实验环境下, 每批次 16 张图像在显存使用和梯度稳定性上最佳。优化器选择对收敛性和泛化能力关键。

SGD 的动量项 ( $Momentum = 0.9$ ) 可加速模型收敛, 其简单的更新规则可以增强模型稳定性。

表 1 不同模型检测性能

Table 1 Different model detection performance

模型	mAP/%	Precision/%	Recall/%	Parameters/( $\times 10^6$ )	计算量/GFLOPs	检测速度/fps
SSD	76.0	68.7	63.3	4.9	14.1	57.1
Faster RCNN	79.4	73.4	71.6	16.6	48.0	10.3
Mask RCNN	82.3	79.0	76.5	13.2	36.6	9.9
YOLOv7	81.6	78.5	74.8	7.0	18.8	51.3
YOLOv8	83.1	80.8	79.9	5.3	17.2	65.3
YOLOv10	83.6	81.2	80.8	6.7	20.0	66.7

根据表 1 可知, YOLOv10 的 mAP 可达 83.6%, 相比 SSD、Faster RCNN、Mask RCNN 有显著提升, 较 YOLOv7 和 YOLOv8 也有明显优化。此外, YOLOv10 不仅在准确率和召回率方面优于其他模型, 在参数量、计算量上也具有优势, 检测帧率达 66.7 fps, 在上述算法中是最快的。

通过上述分析, YOLOv10 凭借 YOLO 系列算法单阶段的处理方式以及 v10 版本精简高效的架构, 在检测精

为贴近疲劳驾驶检测实际应用的硬件条件, 更准确地进行模型性能评估, 本研究将训练后的模型移植到 Jetson Xavier NX 嵌入式平台进行应用测试。测试使用的硬件摄像头选取三合一摄像头, 输入的实时视频流图像在 Xavier 嵌入式平台上直接进行分析和计算。实际部署平台及软硬件如图 7 所示。

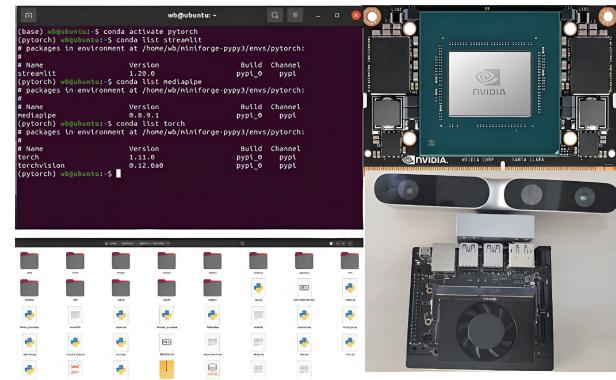


图 7 实际部署平台及软硬件设备

Fig. 7 Embedded deployment

实验中重点关注的算法指标包括平均精度均值 (mAP)、准确率 (precision)、召回率 (recall)、模型参数量 (parameters)、计算量、检测速度。

### 2.3 实验过程

#### 1) 各类原始模型或算法的检测性能对比

在统一的实验设置下, 首先对当下较经典的和较流行的目标检测算法进行对比分析。使用 SSD、Faster RCNN、Mask RCNN、YOLOv7、YOLOv8 和 YOLOv10 算法对相同的数据集进行测试。最终各算法和模型疲劳检测的各项性能指标结果如表 1 所示。

度、速度和资源利用率上表现出色, 对于疲劳驾驶检测领域优势明显, 因此后续研究选定 YOLOv10 作为基础模型进行改进。

#### 2) GGCA 模块的性能评估对比实验

在 YOLOv10 模型的颈部网络后、预测网络前添加 GGCA 模块, 与原算法以及添加经典坐标注意力 CA 模块算法的 mAP 和检测速度进行比较, 实验结果如表 2

所示。

表 2 GGCA 模块的性能评估对比

Table 2 Impact of GGCA module on performance

算法	mAP/%	检测速度/fps
YOLOv10	83.6	66.7
YOLOv10+CA	91.2	65.8
YOLOv10+GGCA	94.0	65.2

相比于原 YOLOv10 网络,添加 GGCA 模块后,模型的 mAP 提高 10.4%。相比于 YOLOv10+CA 算法,mAP 提高 2.8%。实验结果证明,基于 GGCA 全局坐标注意力模块的改进,在保证检测帧率的条件下,使网络的平均检测精度有了一定提高。

### 3) MDFA 模块的性能评估对比实验

在 YOLOv10 算法中引入多尺度空洞融合模块 MDFA,并与原算法及添加 CBAM 模块的算法在 mAP 和检测帧率方面进行对比,实验结果如表 3 所示。

表 3 MDFA 模块性能评估对比

Table 3 Impact of MDFA module on performance

算法	mAP/%	检测速度/fps
YOLOv10	83.6	66.7
YOLOv10+CBAM	90.2	65.2
YOLOv10+MDFA	94.4	65.2

相比于原 YOLOv10 网络,添加多尺度空洞融合模块 MDFA 后,mAP 提高 10.8%。相比于添加 CBAM 的算法,mAP 提升 4.2%。实验证明通过基于 MDFA 的特征融合改进,综合性能更优,在检测精度和速度上位于一个平衡的位置,更适应于实时性的高精度疲劳检测。

### 4) 各损失函数对训练过程的影响对比

为了评估 FDL 损失函数对模型训练过程的影响,分别将损失函数 GIoU、DIoU、CIoU 和 FDL 应用于 YOLOv10 的模型训练。上述 4 种损失函数的训练损失曲线如图 8 所示,表 4 是采用不同损失函数训练模型后进行检测的性能结果对比。

表 5 改进前后模型检测性能

Table 5 Model detection performance before and after improvement

模型	mAP/%	Precision/%	Recall/%	检测速度/fps
YOLOv10	83.6	80.9	80.0	66.7
YOLOv10+GGCA	94.0	84.6	85.1	65.2
YOLOv10+MDFA	94.4	85.2	86.3	65.2
YOLOv10+GGCA+MDFA	97.9	89.9	88.7	63.8
YOLOv10+GGCA+MDFA+FDL(GMF)	98.1	92.6	90.4	64.3

分析表 5 发现,同时添加 GGCA 和 MDFA 模块,较两模块单独添加,在模型检测精度、精确率、召回率方面均有提高,证明两模块同时使用性能更佳。

此外,较之实验中其他模型,融合 GGCA 与 MDFA 模块且更换损失函数的 YOLOv10-GMF 模型在疲劳驾驶实时特征检测过程中的算法指标表现最好。模型检测精

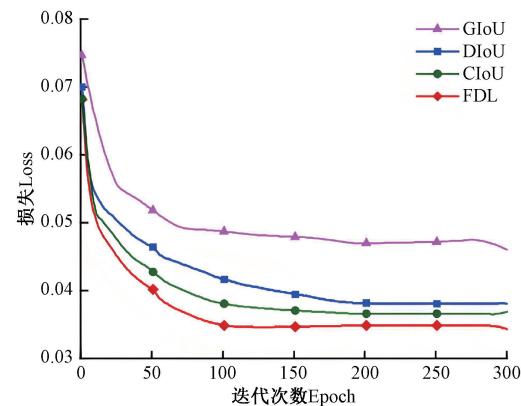


图 8 损失曲线

Fig. 8 Loss curve

表 4 使用不同损失函数训练模型的检测性能对比

Table 4 The performance comparison of models

损失函数	mAP	Precision	Recall	(%)
GIoU	76.2	79.7	73.1	
DIoU	81.6	80.2	79.5	
CIoU	83.4	80.1	80.1	
FDL	84.3	83.8	80.9	

由图 8 可以看出,损失函数在训练中最终都实现了收敛,但 FDL 收敛速度更快,并且在收敛后损失值最低。因此,该损失函数对图像特征的利用率较高。同时由表 4 可得,与其他损失函数相比,采用 FDL 损失函数训练的模型在平均检测精度、精确率和召回率方面更具优势。综上所述,将损失函数 FDL 应用于 YOLOv10 中训练优化过程,能够有效提升疲劳驾驶的检测效果。

### 5) 综合消融实验分析

为进一步验证改进后算法检测效果,在 YOLOv10 原模型基础上,分别添加 GGCA、MDFA 模块,并引入 FDL 损失函数,进行综合消融实验,分析改进效果。结果如表 5 所示。

度、精确率和召回率进一步提升,较 YOLOv10 原模型算法分别提高了 14.5%、11.7% 和 10.4%。

将本文提出的 YOLOv10-GMF 和原 YOLOv10 模型进行实际疲劳驾驶面部特征检测对比,并进行检测结果可视化,如图 9(a)~(c) 所示。



图 9 改进前后模型检测效果

Fig. 9 Visualization results of model before and after improvement

可以发现,原始的 YOLOv10 模型在疲劳驾驶检测的眼部、嘴部的特征目标较小、模糊或者有遮挡的情况下出现了漏检,而 YOLOv10-GMF 模型在同样情况下成功完成检测任务。在清醒疲倦的两种驾驶状态检测中置信度均高达 0.95,且与驾驶员实际状态完全相符。

这种提升不仅体现在视觉效果上,还反映在算法指标的量化结果中。将 YOLOv10-GMF 与其他先进检测算法进行综合性能对比,如表 6 所示。可以发现,较文献[3]、文献[9]、文献[10]和文献[11]的检测算法和模型,YOLO 系列算法在平均检测精度和帧率上展现了极大的优势。同时,相比于改进 YOLOv8 算法<sup>[17]</sup>,本文的模型 YOLOv10-GMF 在 mAP 和检测帧率上进一步提升。可以满足驾驶过程中的实时检测需求。

表 6 各模型及算法综合检测性能比较

Table 6 Comparison of comprehensive detection performance of various models and algorithms

来源	模型	mAP/%	检测速度/fps
文献[3]	ECG GA-SVM	87.2	18.2
文献[9]	SSD	82.2	23.6
文献[10]	3DCNN	75.6	30.2
文献[11]	LBF	93.9	15.2
文献[17]	改进 YOLOv8	92.3	39.4
本文	YOLOv10-GMF	98.1	64.3

### 3 测试与应用

本文选取晴朗、雾霾、雷雨的多种天气条件下的日间

和夜间环境进行实际嵌入式部署测试。在不同天气条件下,选取上午 7:00~9:00,中午 12:00~14:00,下午 17:00~19:00,夜间 22:00~24:00 的 1 d 的 4 个时段构建实际视频流样本,每个时段持续 2 h,验证模型在长时间运行中的稳定性和可靠性。黑暗的夜间环境或雾霾、雷雨天气将导致驾驶室内为低照度环境;路灯以及后车灯光干扰将导致驾驶室内为高照度环境。

设置摄像头在行车过程中以 50 fps 的速度获取视频流图像,以测试模型对连续且快速图像输入的处理能力。将获取的视频流图像按照驾驶室内环境的光照强度分为 3 类(中照度环境、低照度环境、高照度环境),为保证实验样本的多样性与普适性,每类随机选取 1 200 张图像作为实验样本,测试效果如表 7 所示。

表 7 不同环境光照条件下的样本测试效果

Table 7 Sample testing effects under various ambient lighting conditions

环境光照条件	实验样本总数	真阳性样本数	完成率/%	准确率/%
中照度环境	1 200	1 200	100	100
低照度环境	1 200	1 200	100	100
高照度环境	1 200	1 159	100	96.5

分析表 7 可以发现,测试中没有出现漏检问题,完成率达 100%。中照度环境和低照度环境下,模型表现完美,所有样本均被正确识别,准确率为 100%。高照度环境下,模型准确率略降至 96.5%,可能由于强光导致的图

像过曝或细节丢失,影响了部分样本的识别。但由于驾驶员动作的连贯性,若某一帧图像显示驾驶员处于疲劳状态,则其前后几帧图像通常也会呈现相同的疲劳动作或神态特征。通过分析连续多帧图像的检测结果,可以准确地判定驾驶员的真实状态。且在后续工作中可以通过在镜头添加偏振片等方法减弱强照度复杂光源的影响,从而实现精准检测。

模型在不同光照条件下均表现出色,尤其在中和低照度环境下达到满分表现,表明模型具有良好的鲁棒性和适应性。将检测结果进行可视化,如图 10 所示,图 10(a)为中照度环境、图 10(b)为低照度环境、图 10(c)为高照度环境。



(a) 中照度环境检测效果可视化



(b) 低照度环境检测效果可视化



(c) 高照度环境检测效果可视化

(c) Visualization of high-illumination environment detection results

图 10 实际驾驶测试模型检测效果可视化

Fig. 10 Practical driving experiment

本文在将模型移植到嵌入式系统后,统计了各帧视频流图像在 Xavier 平台上进行疲劳驾驶检测处理全过程的各阶段平均耗时。结果如表 8 所示。其中,第 1 阶段为基于 YOLOv10-GMF 模型的人脸定位与关键点获取,其平均处理速度达到 64.3 fps,相当于每帧图像的平均处理耗时为 15.5 ms。第 2 和第 3 阶段在 Xavier 平台上的总耗时为 3.5 ms。综上,3 个阶段的处理总耗时为 19 ms。而人体正常眨眼、打哈欠等动作的通常耗时在 30 ms 以上。由此可见,该模型的检测效率能够满足日常驾驶的实际需求。

表 8 疲劳检测各阶段耗时

Table 8 Time consuming for each stage

序号	阶段	平均耗时/ms
1	人脸定位与关键点获取	15.5
2	疲劳特征置信度计算	3.0
3	综合结果输出	0.5

## 4 结 论

针对当前疲劳驾驶检测中驾驶者微表情目标小、环境背景复杂,导致检测效果不理想的问题,本文提出了一种基于 YOLOv10 改进的检测算法 YOLOv10-GMF。首先,引入全局分组坐标注意力模块 GCGA,进行信息压缩的同时,分组提取全局特征信息。解决局部注意力机制在捕捉特征信息时不能同时利用多维度全局信息的问题,使检测精度比原始模型提高 10.4%。进而,针对小目标严重降采样的问题,在原始模型中引入多尺度空洞融合 MDFA 模块,针对多尺度目标特征,进行空洞融合并行处理,重点关注小目标特征,使检测精度较原模型提高了 10.8%。最后针对原始算法中损失函数仅依据固定参数优化训练过程的不合理之处,引入 FDL 损失函数,借助反馈信息实现动态优化模型训练过程的目标。

本文还将基于 YOLOv10-GMF 的疲劳驾驶检测系统部署到 Jetson Xavier NX 嵌入式平台。实验测试表明,该模型的平均精度均值达到了 98.1%,相比 YOLOv10 提升了 14.5%,同时其检测速度为 64.3 fps。通过计算各阶段耗时评估发现,整个疲劳检测过程平均耗时 19.0 ms,能够满足实时驾驶监管系统的检测需求。后续研究将着重于进一步轻量化模型架构。初步计划是运用模型剪枝技术,去除网络中的冗余连接和非关键特征通道,降低模型的复杂度。同时,通过量化感知训练,将模型参数和激活值转换为低精度格式,以减少存储需求和计算资源占用。

## 参考文献

- [1] ALKINANI M H, KHAN W Z, ARSHAD Q. Detecting human driver inattentive and aggressive driving behavior using deep learning: Recent advances, requirements and open challenges [C]. Special Section on Artificial Intelligence Empowered Intelligent Transportation Systems, 2020, 99(8): 105008-105030.
- [2] YANG Y Q, CHEN Y B, WU C X, et al. Effect of highway directional signs on driver mental workload and behavior using eye movement and brain wave [J]. Accident Analysis and Prevention, 2020, 146 (10): 105705-105721.
- [3] CHUI K T, TSANG K F, CHI H R, et al. An accurate ECG-based transportation safety drowsiness detection scheme [C]. Transactions on Industrial Informatics, 2016, 12(4): 1438-1452.

- [ 4 ] LIU W, ANGUELOV D, EEHAN D, et al. SSD: single shot multibox detector [ C ]. European Conference on Computer Vision, 2016, 8(1) : 21-37.
- [ 5 ] REDMON J, DIVVALA S K, GIRSHICK R B, et al. You only look once: Unified, real-time object detection [ C ]. IEEE Conference on Computer Vision and Pattern Recognition, 2016, 12(5) : 779-788.
- [ 6 ] REN SH Q, HE K M, GIRSHICK R B, et al. Faster R-CNN: towards real-time object detection with region proposal networks [ J ]. Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6) : 1137-1149.
- [ 7 ] ZHOU L, LI S, WANG Y. Fatigue detection and early warning system for drivers based on deep learning [ C ]. International Conference on Data Science and Computer Application, 2023: 1348-1351.
- [ 8 ] ALAMEEN S A, ALHOTHALI A M. A lightweight driver drowsiness detection system using 3DCNN with LSTM [ J ]. Computer Systems Science and Engineering, 2023, 44(1) : 895-912.
- [ 9 ] 胡习之, 黄冰瑜. 基于面部特征分析的疲劳驾驶检测方法 [ J ]. 科学技术与工程, 2021, 21 ( 4 ) : 1629-1636.
- HU X ZH, HUANG B Y. Fatigue driving detection method based on facial feature analysis [ J ]. Science, Technology and Engineering, 2021, 21(4) : 1629-1636.
- [ 10 ] 秦康, 张小俊, 张明路, 等. 基于 3DCNN 的驾驶员细微动作识别 [ J ]. 电子测量技术, 2023, 46 ( 8 ) : 51-58.
- QIN K, ZHANG X J, ZHANG M L, et al. Recognition of driver's subtle movements based on 3DCNN [ J ]. Electronic Measurement Technology, 2023, 46(8) : 51-58.
- [ 11 ] 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.
- YAN B ZH, WANG CH Y, WANG SH SH. Research on fatigue driving detection technology based on human eye features [ J ]. Applied Science and Technology, 2020, 47(1) : 47-54.
- [ 12 ] ANSARI S, NAGHDY F, DU H, et al. Driver mental fatigue detection based on head posture using new modified reLU-BiLSTM deep neural network [ C ]. Transactions on Intelligent Transportation Systems, 2022, 23(8) : 10957-10969.
- [ 13 ] 牛洪超, 胡晓兵, 罗耀俊. 基于 M-YOLO 的自动驾驶下目标识别算法 [ J ]. 计算机工程与设计, 2022, 43(8) : 2213-2220.
- NIU H CH, HU X B, LUO Y J. Target recognition algorithm for autonomous driving based on M-YOLO [ J ]. Computer Engineering and Design, 2022, 43 ( 8 ) : 2213-2220.
- [ 14 ] 邝先验, 程福军, 吴翠琴, 等. 基于改进 YOLOv7-tiny 的高效轻量遥感图像目标检测方法 [ J ]. 电子测量与仪器学报, 2024, 38(7) : 22-33.
- KUANG X Y, CHENG F J, WU C Q, et al. Efficient and lightweight target detection method for remote sensing images based on improved YOLOv7-tiny [ J ]. Journal of Electronic Measurement and Instrumentation, 2024, 38(7) : 22-33.
- [ 15 ] 曾勇杰, 范必双, 杨涯文, 等. 改进 YOLOv8 算法在风机叶片缺陷检测上的应用 [ J ]. 电子测量与仪器学报, 2024, 38(8) : 26-35.
- ZENG Y J, FAN B SH, YANG Y W, et al. Application of improved YOLOv8 algorithm in wind turbine blade defect detection [ J ]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8) : 26-35.
- [ 16 ] 胡久松, 刘张驰, 余谦, 等. 融入 GhostNet 和 CBAM 的 YOLOv8 烟雾识别算法 [ J ]. 电子测量与仪器学报, 2024, 38(8) : 201-207.
- HU J S, LIU ZH CH, YU Q, et al. Integration of GhostNet and CBAM in YOLOv8 smoke recognition algorithm [ J ]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8) : 201-207.
- [ 17 ] ZHOU C, YANG Z, LIU S B. Research on driver facial fatigue detection based on Yolov8 model [ C ]. 5th International Conference on Information Science, Parallel and Distributed Systems, 2024.
- [ 18 ] HOU Q, ZHOU D, FENG J, et al. Coordinate attention for efficient mobile network design [ C ]. Conference on Computer Vision and Pattern Recognition, 2021.
- [ 19 ] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [ J ]. Proceedings of the European Conference on Computer Vision, 2018, 13(6) : 3-19.
- [ 20 ] BOWEN Z, HUACAI L, SHENGBO Z, et al. Night target detection algorithm based on improved YOLOv7 [ J ]. Scientific Reports, 2024, 14(1) : 9-14.

### 作者简介



**殷旭鹏**, 现为南京信息工程大学本科生, 主要研究方向为目标检测与模式识别。  
E-mail: yxp12061@163.com

**Yin Xupeng** is currently a B. Sc.

candidate in Nanjing University of Information

Science & Technology. His main research

interests include object detection and pattern recognition.



**赵兴强**(通信作者), 2005 年于重庆大

学获得学士学位, 2008 年于重庆大学获得

硕士学位, 2014 年于重庆大学获得博士学

位, 现为南京信息工程大学副教授, 主要研

究方向为智能感知与检测。

E-mail: zxq8562@163.com

**Zhao Xingqiang** (Corresponding author) received his B. Sc. degree, M. Sc. degree, and Ph. D. degree from Chongqing University in 2005, 2008 and 2014 respectively. He is currently an associate professor in Nanjing University of Information Science & Technology. His main research interests include deep learning and intelligence perception.