

# 基于轻量化改进 Yolov5s 的驾驶员人脸检测算法

贺 帅, 李 锐

(重庆交通大学机电与车辆工程学院, 重庆 400047)

**摘要:**人脸检测是驾驶员驾驶行为检测和驾驶疲劳检测的首要前提,而实时检测效率和检测精度则直接影响最终的判别效果,同时轻量化设计也更有助于后期在嵌入式设备上的计划部署;为此,基于 Yolov5s 设计了一种轻量化的驾驶员人脸检测算法。首先采用 Shufflenetv2 作为特征提取主干,降低网络的参数数量和计算复杂度;然后引入 ASFF 实现不同层间的人脸信息融合,并结合 SIOU Loss 提升模型训练速度和检测精度。结果表明,该检测算法的性能和轻量化程度均得到了有效的提升,节省了硬件成本,满足设计要求。

**关键词:** Yolov5s; 人脸检测; 轻量化; ASFF; SIOU Loss

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 2096-9759(2023)08-0043-04

## Driver face detection algorithm based on lightweight improved Yolov5s

HE Shuai, LI Rui

(School of Mechanical and Electrical and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400047)

**Abstract:** Face detection is a prerequisite for driver behavior detection and driving fatigue detection. Real-time detection efficiency and detection accuracy directly affect the final discrimination effect. At the same time, lightweight design is more conducive to the planned deployment on embedded devices in the later stage. Therefore, a lightweight driver face detection algorithm based on Yolov5s is designed. Firstly, Shufflenetv2 is used as the feature extraction backbone to reduce the network's parameter volume and computational complexity. Then, ASFF is introduced to achieve facial information fusion between different layers and combined with SIOU Loss to improve the model's training speed and detection accuracy. The results show that the performance and lightweight level of this detection algorithm have been effectively improved, which saves hardware costs and meets design requirements.

**Key words:** yolov5s; face detection; lightweighting; adaptive feature fusion; SIOU Loss

## 0 引言

随着目标检测算法的不断迭代更新,快速部署以及精准检测的优势已经让其在工业领域得到了广泛的应用。而驾驶员人脸检测作为在驾驶辅助系统中提取面部特征信息的关键步骤,已经成为了机器视觉领域的研究热点。并且由于面部特征信息采集的自然性、非强制性和非接触性,在驾驶过程中不会对驾驶员操作造成任何干扰,因此,它越来越广泛地应用于驾驶疲劳检测和驾驶行为检测。

张文霞等人<sup>[1]</sup>基于 Adaboost 算法,从车辆主动安全角度出发提出人脸分类检测器,并通过大量的驾驶室样本对其算法的快速和准确性进行了有效验证,为驾驶辅助系统的开发提供了参考依据。冯辉宗等人<sup>[2]</sup>为了解决驾驶室内由于光线不足引起的人脸检测定位准确率和实时性低的问题,基于多尺度 Retinex 算法,对人脸皮肤区和非皮肤区图像进行光照补偿处理。设计了一种肤色聚类的人脸检测算法来实现对不同光照变化和旋转角度的人脸进行检测,但是当检测背景中出现类肤色的部分时容易引起误检。李永杰等人<sup>[3]</sup>针对头部姿态估计中人脸检测框的尺寸难以学习的问题,提出了一种基于 YOLOv3 算法的人脸检测与头部姿态估计融合的算法,充分利用特征层间的信息,实现对特征类别的多尺度检测。但是 YOLOv3 模型整体不够轻量,不利于后期边缘设备的部署。卢嫚等人<sup>[4]</sup>基于 MTCNN 和 MobileFaceNet 算法设计了一种人脸识别系统,通过人脸候选框和人脸关键点坐标输出,实现对移动人脸的活体检测识别。综上,针对目前人脸检测算法体积臃肿和识别率较低的问题,本文将着重对网络的轻量化和检测效率进行研究。

目前主流的目标检测算法可以分为以 Yolo 系列<sup>[5-8]</sup>为代表的一阶段算法和以 R-CNN<sup>[9]</sup>、Fast R-CNN<sup>[10]</sup>等为代表的二阶段算法,其中一阶段目标检测算法具有更快、更高的检测速度和检测实时性,并且经过不断的优化改进,Yolo 系列算法相较于二阶段算法的检测精度已经得到了显著的提升。因此,针对车辆驾驶员的人脸检测,本文在目标检测算法 Yolov5s 的基础上,引入轻量级主干网络 Shufflenetv2<sup>[11]</sup>,降低模型的参数数量和计算复杂度。然后通过自适应空间特征融合模块 ASFF (Adaptively Spatial Feature Fusion)<sup>[12]</sup>,增强人脸特征之间的信息传递。最后引入边框回归损失 SIOU Loss<sup>[13]</sup>,避免预测框在回归时的随意匹配,减少特征层间的信息损失,提升模型的训练速度和人脸检测精度。最终构建轻量化且高效的人脸检测方法。

## 1 人脸检测方法

### 1.1 人脸检测模型网络结构

在移动端的检测任务中,检测精度和检测速度以及模型体积的大小是衡量模型是否能应用在工程上的重要参考指标,并且考虑到移动端边缘计算设备的计算能力和内存空间的大小,本文研究基于 Yolov5s 做了进一步的适应性改进,优化了模型的检测识别性能,如图 1 所示。本文针对 Yolov5s 主要做了三方面改进:(1)为了降低模型参数数量和计算复杂度,提高运行速度,引入轻量级网络 Shufflenetv2 替换 Yolov5s 的特征提取主干。(2)通过自适应空间特征融合 ASFF,将高层与底层的特征重新融合,充分利用不同特征尺度的信息,进一步提高人脸检测的精度。(3)采用 SIOU Loss 替换 CIOU Loss<sup>[14]</sup>,提升模型训练时的鲁棒和速度。

收稿日期:2023-04-19

通讯作者:贺帅(1997-),男,四川巴中人,硕士研究生,重庆交通大学,主要研究方向为深度学习、图像处理;李锐(1998-),男,重庆人,硕士研究生,重庆交通大学,主要研究方向为深度学习、图像处理。

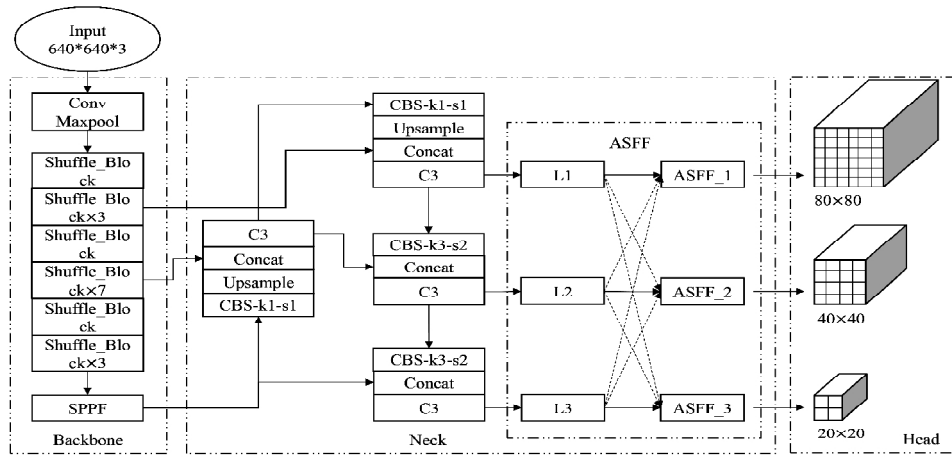
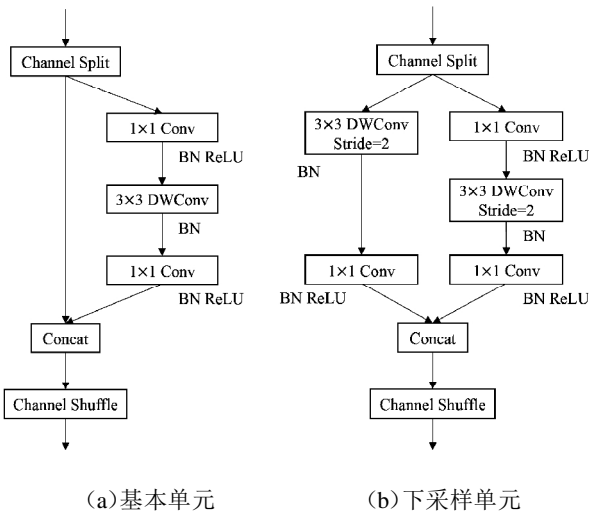


图1 改进后 Yolov5s 网络结构

## 1.2 模型结构的改进

### 1.2.1 Shufflenetv2 网络

为满足将模型搭载至车载终端的轻量化需求,在 Yolov5s 的基础上,引入参数数量和计算量更小的 Shufflenetv2 网络作为初始结构的特征提取主干,以提高检测速度。如图 2 所示,Shufflenetv2 主要由基本单元和下采样单元构成,它沿用了 Shufflenet<sup>[15]</sup> 的 Channel Shuffle 和深度可分离卷积。其中基本单元的输入特征经过 ChannelSplit 实现通道分流,形成双分支结构,两个分支在卷积操作之后将进行通道相加和 Channel shuffle,以充分利用各分支间传递的信息,使其在实现分组卷积运算量降低的同时,避免各卷积组相互学习组内特征信息的局限性。下采样单元与基本单元结构相似,不同的是其下采样操作是通过在两条通道路径中引入步长为 2 的深度可分离卷积来实现的。



(a)基本单元

(b)下采样单元

图2 Shufflenetv2 基本构建单元

### 1.2.2 自适应空间特征融合 ASFF

考虑到 Yolov5s 网络模型中 FPN+PAN 结构自上而下的人脸特征信息传递效率低且不能充分利用不同尺度特征的问题,本文在 PANet 的基础上引入 ASFF。该模块能够通过自适应学习获取特征层上人脸信息的位置权重,然后学习如何对冲冲突信息进行空间过滤以抑制不一致性,使其在特征融合时占据主导地位,从而提高面部特征的尺度不变性和人脸特征检测过程中的精度。

如图 3 所示,该模块的实现主要分为特征尺度缩放和自适应特征融合两个步骤,其中 F1、F2、F3 分别表示来自主干特征层的特征输出。对于特征缩放,由于各特征层通道数和分辨率的差异,需要修改每个特征尺度的上采样和下采样策略,使特征尺寸达到一致。对于特征层的自适应融合,本文以 ASFF-3 为例,为实现人脸特征信息的融合操作,只需将 L1、L2、L3 层输出的特征信息与权重参数  $\alpha_{ij}^3$ 、 $\beta_{ij}^3$ 、 $\gamma_{ij}^3$  的乘积求和即可,计算如式(1)所示:

$$y_{ij}^3 = \alpha_{ij}^3 \cdot x_{ij}^{1 \rightarrow 3} + \beta_{ij}^3 \cdot x_{ij}^{2 \rightarrow 3} + \gamma_{ij}^3 \cdot x_{ij}^{3 \rightarrow 3} \quad (1)$$

式中,  $y_{ij}^3$  表示经过 ASFF-3 获得的新特征图,  $\alpha_{ij}^3$ 、 $\beta_{ij}^3$ 、 $\gamma_{ij}^3$  表示来自不同特征层的权重值,  $x_{ij}^{1 \rightarrow 3}$ 、 $x_{ij}^{2 \rightarrow 3}$  表示 L1、L2 层的特征经过特征缩放生成的与 L3 层的  $x_{ij}^{3 \rightarrow 3}$  尺寸相同的特征。

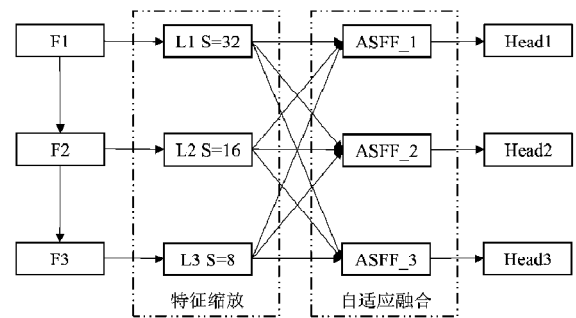


图3 ASFF 基本结构

### 1.2.3 损失函数优化

目标检测任务的有效性在很大程度上取决于损失函数的优化,传统的损失函数考虑到了预测框与真实框的距离、重叠面积以及纵横比等因素,但是忽略了回归方向不匹配的问题,进而导致收敛速度较慢。因此,本文采用边框回归损失 Siou Loss 替换 Yolov5s 的 Ciou Loss,重新定义惩罚机制,并且充分考虑回归向量的角度,增加角度惩罚项,减少自由度的数量,从而提高人脸检测性能。Siou Loss 的计算主要有向量角度成本 (Angle cost)、距离成本 (Distance cost)、形状成本 (Shape cost) 和面积重叠成本 (IoU cost) 四个部分。

其中, Angle cost 能够最大限度减少与距离相关变量的数量,避免回归随意匹配的问题。根据 Angle cost,重新定义其

余三部分的成本计算,最终得到 SIOU Loss,如式(2)所示:

$$SIOU\ Loss=1-IoU+(\Delta+\Omega)/2 \quad (2)$$

式中,“ $\Delta$ ”表示 Distance cost,“ $\Omega$ ”表示 Shape cost,“IoU”表示 IoU cost。

## 2 实验与结果分析

### 2.1 数据集与环境平台

实验所需的人脸数据集分别来自模拟驾驶视频数据 YawDD<sup>[16]</sup>和闭眼数据集 CEW<sup>[17]</sup>。首先对 YawDD 视频数据集按每 20 帧截取一张图片的方式进行镜像、旋转和裁切增强处理,然后添加闭眼数据集 CEW 扩增数据集的类别特征,提高模型训练的稳定性。最终将数据集标注为 face、yawn、o\_eyes、n\_yawn 和 c\_eyes 五个类别,共 3604 张数据图片,并按照 8:2 比例划分为训练集和验证集。

实验平台为一台机带 11th Gen Intel(R)Core(TM)i7-11800H @ 2.30GHz 处理器,16GB 内存,RTX3060(6GB)显卡的笔记本电脑,其中 CUDA 版本为 11.7,深度学习框架为 Pytorch1.12.1。模型训练轮数为 200,批大小为 8。

### 2.2 评价指标

本文实验采用的模型性能评价指标为平均精度均值 mAP、召回率 R 以及每秒处理图像的帧数 FPS;轻量化评价指标为浮点运算量 FLOPs、权重文件体积 Vol 和参数量 Paras。

其中 R 和 mAP 的计算方法如式(3)-(5)所示:

$$R=TP/(TP+FN) \quad (3)$$

$$AP=\int_0^1 P(R)dR \quad (4)$$

$$mAP=(\sum_{c=1}^C AP_c)/n \quad (5)$$

式中,TP 表示正样本的正确预测量,FN 表示负样本的错误预测量,AP 表示衡量模型在每个类别上的性能好坏,P 为精确率,c 为预测类别,n 为预测类别的数量。

### 2.3 实验结果分析

#### 2.3.1 消融实验

为了验证分析轻量化算法的有效性和泛化性,对各部分改进优化进行消融实验对比,如表 1 所示。

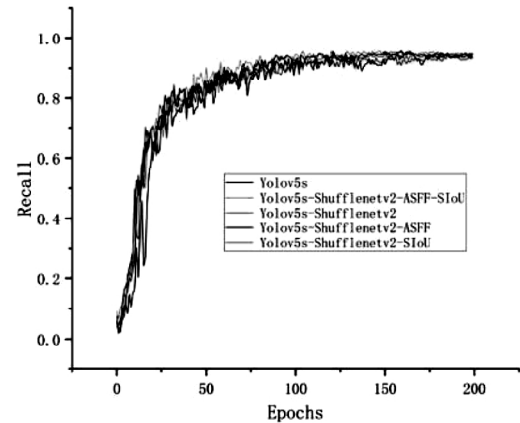
表 1 消融实验结果

Yolov5s	Shufflenetv2	ASFF	SIOU	R/%	mAP/%	Paras/M	FLOPs/G	Vol/M	FPS
✓				90.9	97.1	7.02	15.8	13.7	102
✓	✓			93	96.9	1.05	2.0	2.4	105
✓	✓	✓		93.9	98.1	2.42	4.2	5.2	85
✓	✓		✓	94.4	97.5	1.05	2.0	2.4	106
✓	✓	✓	✓	94.9	98.2	2.42	4.2	5.2	79

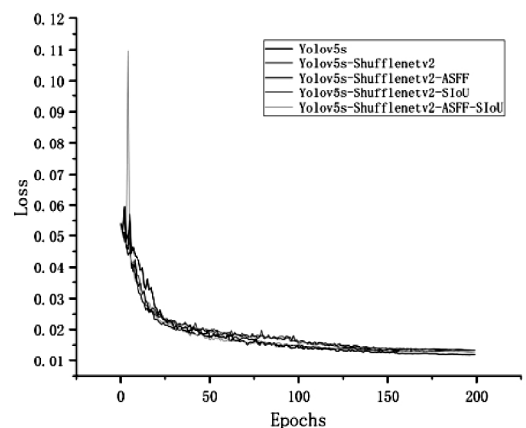
从表 1 的实验结果可知,本文设计的人脸检测算法虽然在 FPS 上稍低于 Yolov5s,检测速度相对较慢,但是并不影响模型最终的检测速度,符合快速实时检测的预期。并且从网络的检测性能和模型的轻量化程度两方面表明,改进后的召回率 R 和平均检测精度 mAP 分别提升了 4% 和 1.1%;轻量化程度虽然在 Shufflnetv2 的基础上有所降低,但是相比于 Yolov5s 算法,浮点运算量 FLOPs 和参数量 Paras 分别降低了 73.4%、65.5%,并且权重文件体积 Vol 也仅为 5.2M,获得了较为优秀的检测和轻量化效果。

为了更加直观的比较 Yolov5s 算法改进后各个模块对其

性能的影响,在自建数据集上的训练结果如图 4 所示。图 4(a)为基于 Yolov5s 各改进点的召回率 R 随训练轮数(Epochs)的变化情况,可以直观的看出模型在训练过程中正样本的实际预测正确率随着各改进模块的融入逐渐提升,且最终模型的整体训练鲁棒性得到明显改善,与上文统计数据相互对应。从图 4(b)中可以看出,改进 Yolov5s 的 SIOULoss 相比于 Yolov5s 有所增大,但是差距很小,并且前期的收敛速度更快,损失更低。



(a) Recall



(b) Loss

图 4 不同模型训练的召回率和损失曲线

综上所述,本文针对 Yolov5s 提出的改进方案能够在一定程度上提升模型的整体性能,验证了改进方法的有效性。

#### 2.3.2 模型对比实验

为了进一步验证改进后的轻量化人脸检测算法的性能和优势,在自建数据集的基础上,将其他主流的目标检测算法与本文改进算法进行对比,结果如表 2 所示。

表 2 不同模型对比试验

Algorithm	mAP/%	FLOPs/G	Paras/M	Vol/M	FPS
Yolov3	99	154.6	61.52	123.5	84
Yolov3-Tiny	97.9	12.9	8.68	66.7	90
ResNet50	78.9	15	25.56	102	65
Yolov5s	97.1	15.8	7.02	13.7	102
Yolov5n	94.3	4.1	1.77	3.9	104
本文	98.2	4.2	2.42	5.2	79

从上表可知,本文算法与主流 Yolov3、Yolov3-Tiny、Res-Net50 和 Yolov5n 目标检测相比,虽然 mAP 略低于 Yolov3,但是在检测速度和轻量化程度方面的提升较为明显;其中模型的体积 Vol 和浮点运算量 FLOPs 仅为它的 4.2%和 2.7%。与 Yolov5 系列最轻量的 Yolov5n 相比,整体的轻量化程度在非

常接近的同时,检测效率也得到了明显提升,验证了改进算法的高性能和优越性。

为了更加直观的反映本文设计的人脸检测算法的性能优势,针对 YawDD 模拟驾驶视频数据集进行算法改进前后的可视化对比分析,检测结果如图 5 所示。



图 5 Yolov5s 改进前后可视化对比

从图 5 中可知,改进后算法的检测识别精度具有较为明显的优势,即使驾驶员头部转动幅度较大或是在受到外界环境影响下,仍具有较高的检测效率,且对于眼睛这类相对较小目标的识别表现更为明显。

### 3 结语

针对驾驶员人脸检测任务,基于 Yolov5s 设计了一种轻量化的驾驶员人脸检测算法:使用轻量级网络 Shufflenetv2 降低模型参数和计算量;然后在 PANet 的基础上引入 ASFF 加强网络层间的特征信息融合,提升模型的人脸特征提取能力,并结合 SIOU Loss 改善算法轻量化后人脸检测性能不佳的问题。在自建数据集上的实验结果表明, mAP 和 R 分别提升 1.1%和 4%的同时,有效的压缩了网络参数数量和计算复杂度,且能够满足快速实时检测的设计要求。对于后期驾驶员的驾驶行为检测和驾驶疲劳检测以及嵌入式设备的轻量化部署都具有一定的参考价值。

#### 参考文献:

- [1] 张文霞,王旭.驾驶辅助系统中人脸检测方法研究[J].汽车实用技术,2017,No.249(18):83-84.
- [2] 冯辉宗,谢静,蒋建春.针对机动车驾驶员的人脸检测方法[J].传感器与微系统,2012,31(09):112-115.
- [3] 李永杰,周桂红,刘博.基于 YOLOv3 模型的人脸检测与头部姿态估计融合算法[J].广西师范大学学报(自然科学版),2022,40(03):95-103.
- [4] 卢嫚,邓浩敏.一种基于 MTCNN 和 MobileFaceNet 人脸检测及识别方法[J].自动化与仪表,2023,38(02):76-80+97.
- [5] Redmon J,Divvala S,Girshick R,et al. Youonly look once: Unified, real time object detection[C]. Computer Vision and Pattern Recognition,2017:6517-6525.
- [6] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C]. IEEE Conference on Computer Vision and Pattern Recognition,2017: 6517-6525.

- [7] Redmon J, Farhadi A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
- [8] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [9] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. IEEE Conference on Computer Vision & Pattern Recognition, IEEE Computer Society, 2014: 580-587.
- [10] Girshick R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.
- [11] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [C] // Computer Vision-ECCV 2018-15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XIV, 2018: 122-138.
- [12] Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection [J]. arXiv preprint arXiv: 1911.09516, 2019.
- [13] Gevorgyan Z. SIOU loss: More powerful learning for bounding box regression [J]. arXiv preprint arXiv: 2205.12740, 2022.
- [14] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021.
- [15] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6848-6856.
- [16] Abtahi S, Omidyeganeh M, Shirmohammadi S, et al. YawDD: A yawning detection dataset [C] // Proceedings of the 5th ACM multimedia systems conference. 2014: 24-28.
- [17] Weng C H, Lai Y H, Lai S H. Driver drowsiness detection via a hierarchical temporal deep belief network [C] // Asian Conference on Computer Vision. Springer, Cham, 2017: 117-133.