

DOI: 10.11830/ISSN.1000-5013.202212028



复杂交通环境下二轮机动车 乘员头盔检测算法

钟铭恩¹, 谭佳威¹, 袁彬淦², 吴志华¹, 冯妍¹, 朱程林¹

(1. 厦门理工学院 机械与汽车工程学院, 福建 厦门 361024;

2. 厦门大学 航空航天学院, 福建 厦门 361104)

摘要: 针对现有二轮机动车乘员头盔检测算法在目标密集分布、随机遮挡等情况下效果较差且难以在边缘设备上应用的问题, 制作了具有针对性的数据集, 对比现有模型后, 以 YOLOv7 为参考提出一种复杂交通环境下二轮机动车乘员头盔检测算法。首先, 采用 EfficientNet-B3 作为主干网络, 可提高特征提取能力且更为轻量化; 其次, 将增大感受野模块(RFB)引入特征融合结构中, 以增大模型感受野, 提升小目标头盔检测能力; 最后, 在检测头嵌入 SimAM 机制, 在不增加参数的前提下提高算法精度。结果表明: 相较于 YOLOv7, 文中算法的准确率、召回率和平均准确率分别提高了 2.84%, 2.26% 和 3.26%, 参数量和运算量分别为 YOLOv7 的 33.1%, 23.5%, 可实现当前主流模型算法的最佳检测性能和效率; 在 NVIDIA Jetson Nano 开发板上的处理速度达到 $47.58 \text{ F} \cdot \text{s}^{-1}$, 可满足边缘设备部署需求。

关键词: 二轮机动车; 头盔检测; YOLOv7; 轻量级网络; 感受野; 注意力机制

中图分类号: TP 391.4; U 483

文献标志码: A

文章编号: 1000-5013(2023)03-0301-08

Helmet Detection Algorithm of Two-Wheeled Motor Vehicle Occupant in Complex Traffic Environment

ZHONG Ming'en¹, TAN Jiawei¹, YUAN Bin'gan²,
WU Zhihua¹, FENG Yan¹, ZHU Chenglin¹

(1. School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024, China;

2. School of Aeronautics and Astronautics, Xiamen University, Xiamen 361104, China)

Abstract: The existing helmet detection algorithms for two-wheeled motor vehicle occupant are less effective in the case of dense object distribution and random occlusion, and difficult to apply on edge devices. To address this problem, a targeted dataset is created and a helmet detection algorithm of two-wheeled motor vehicle occupant in complex traffic environment is proposed using YOLOv7 as a reference after comparing existing models. Firstly, EfficientNet-B3 is used as the backbone network to improve the feature extraction capability and make it more lightweight. Secondly, receptive field block (RFB) is introduced into the feature fusion structure to increase the receptive field of the model, and improve the detection capability of small target helmets. Finally, SimAM mechanism is embedded in the detection head to improve the accuracy of the algorithm without increasing the number of parameters. The results show that compared to YOLOv7, the accuracy, recall rate and

收稿日期: 2022-12-23

通信作者: 钟铭恩(1980-), 男, 教授, 博士, 主要从事机器视觉、人工智能和智能交通的研究. E-mail: zhongmingen@xmut.edu.cn.

基金项目: 国家自然科学基金资助项目(51978592); 福建省自然科学基金资助项目(2019J01859)

average accuracy of the proposed algorithm have been improved by 2.84%, 2.26% and 3.26% respectively, and the number of parameters and operations are 33.1% and 23.5% of YOLOv7 respectively, achieving the best detection performance and efficiency of current mainstream model algorithms. The processing speed on the NVIDIA Jetson Nano development board reaches 47.58 frames per second, which can meet the requirements of edge device deployment.

Keywords: two-wheeled motor vehicle; helmet detection; YOLOv7; lightweight network; receptive field; attention mechanism

路口等复杂交通环境下的二轮机动车违法行为是政府部门提升通行安全的一大障碍. 数据表明, 一半以上的交通死亡发生在弱势道路使用者中, 包括摩托车、电动车载乘人员等, 而未规范佩戴头盔正是二轮机动车载乘人员发生事故后死亡的重要原因^[1]. 相较于宣传教育, 行政治理由于其威慑作用而更加直接有效. 目前, 交管部门主要采用人工拦检的方式在路口查处此类行为, 需耗用大量的人力资源, 且易因驾驶员的危险逃窜行为及执法人员可能实施的不文明执法而阻碍交通, 甚至诱发严重事故. 因此, 如何在目标密集和随机遮挡频发的复杂路口环境下更科学、高效地检测和治理二轮机动车乘员不规范佩戴头盔的行为, 一直是社会各界积极探索的课题.

得益于数字图像处理、模式识别等技术的不断发展, 机器视觉在交通管理和智能交通系统中扮演着越来越重要的角色. 近年来, 相关研究主要集中于数据集制备^[2-3]和头盔佩戴检测算法^[4-8]等方面. 然而, 现有数据集的场景较为单一且对应算法的泛化能力不足. 因此, 有必要进一步提出更为复杂、多样的二轮机动车交通图像数据集, 以充分体现我国的交通现状和特点, 并探讨在该交通条件下提升算法的头盔佩戴综合检测性能的方法. 基于此, 本文提出一种复杂交通环境下二轮机动车乘员头盔检测算法.

1 数据集的制备

数据集是研究的前提, 鉴于当前公开的二轮机动车交通数据集的场景相对简单、交通密度较低、摄像头视野角度不够多样、乘员头盔较少出现重叠和遮挡等不足, 拟自行构建数据集以增强检测算法的适应能力. 通过运动相机分别采用固定和移动拍摄模式(图 1)在福建省厦门市各路口实景多角度拍摄二轮机动车的交通视频图像. 通过调整三脚架和自行车上金属立杆的高度和拍摄角度, 充分利用运动相机的防抖和高帧率摄像功能, 较好地模拟了交管部门部署于道路上方的监控探头固定视角和警用车辆(警用汽车、警用摩托车)行车记录仪的移动视角, 采集的数据有利于算法在实际交通场景的大规模应用.

重点考虑在劳动力集中型工业区路口进行数据采集, 以提高密集、混合交通图像数据的占比, 并在不同时段(白天、早高峰、晚高峰和夜间)、不同天气(晴天、阴天、雨天和雾天)、不同光照(逆光、平光和顺光)条件下进行采集, 以提高图像的复杂性和多样性. 拍摄后将图像分辨率统一调整为 800 px×600 px, 并使用开源工具 Labelimg 进行手动标注, 标签为 helmet 和 helmetless. 最终构建 1 个包含 5 381 张图像的、能够较好体现交通目标密集分布和随机遮挡特点的二轮机动车乘员头盔佩戴数据集, 并将其命名为 XMUT-HELMET. XMUT-HELMET 数据示例, 如图 2 所示.



(a) 固定 (b) 移动
图 1 固定和移动拍摄模式

Fig. 1 Fixed and moving shooting modes



图 2 XMUT-HELMET 数据示例

Fig. 2 Examples of XMUT-HELMET data

该数据集上的头盔多为小目标,且像素尺寸变化范围较大.采用深度学习技术进行目标检测时,为使模型先验框尺寸更加符合数据特点,加快网络回归,降低目标漏检率,使用 K-means++ 算法对头盔目标的真实标注框进行聚类.聚类后的先验框尺寸,如表 1 所示.

2 算法模型的设计

目前,学术界和工业界已涌现诸多经典的算法模型,例如,基于锚框的 Faster R-CNN 系列、SSD 系列和 YOLO 系列和无需锚框的 CenterNet,RetinaNet,以及从属于不同架构的 Vision transformer 等^[9-16].对比评价这些模型的性能时,更多的是在 COCO,PASCAL VOC 等大型开源数据集上进行.然而,针对不同的检测任务时,不同算法模型的检测性能很可能发生变化.

主流算法模型在 XMUT-HELMET 数据集上的检测指标,如图 3 所示.图 3 中: P_{ma} 为平均准确率; Q 为模型的参数量^[17].由图 3 可知:YOLOv4,YOLOv7 的平均准确率明显优于其他算法,对图片中密集分布和随机遮挡的小尺度头盔目标的检测效果较好;YOLOv7 的平均准确率比 YOLOv4 略低,但参数量却减少了约 50%,故选取 YOLOv7 作为基线参考模型,进一步对特征提取主干网络、特征融合结构和检测头等 3 个模型组成部分进行替换或改进,以提高平均准确率,使其在 XMUT-HELMET 数据集上达到最高水平(SOTA).

2.1 主干网络的替换

YOLOv7 主干网络由 CBS,ELANB,MPCConv 和 SPPCSPC 等模块堆叠而成,这一复杂的结构对于多个类别的公共数据集目标检测任务的适应能力较强,但对仅有两个类别的二轮机动车头盔检测任务则存在大量冗余参数和运算量.为了便于在边缘设备上部署,使用轻量级网络进行替换(改进 1),以压缩整体资源占用量.

EfficientNet^[17] 是一种轻量级网络,具有特征提取能力强、浮点运算量(FLOPs)和参数量较小的特点^[18].该网络共有 EfficientNet-B0~EfficientNet-B7 等 8 种结构,这些结构是类似的,但参数量和浮点运算量存在较大差异.EfficientNet 的参数量和浮点运算量^[19],如表 2 所示.由表 2 可知:EfficientNet 系列模型的参数量和浮点运算量从 EfficientNet-B4 开始大幅增加,导致网络的训练和调用难度也随之增加.因此,选择具有优良特征抽取能力的 EfficientNet-B3 作为主干网络,可有效降低模型参数量,满足交管部门在警用车辆、监控探头等设备上的部署需求.

2.2 特征融合结构的改进

YOLOv7 的特征融合结构通过拼接等多种操作,可充分利用二轮机动车附近区域的高低维语义信息,但也易使感受野过小,从而忽略头盔底层边缘信息,造成模型在密集、遮挡等不利条件下漏检小目标头盔.为了提高模型对复杂交通场景小目标头盔的检测能力,降低漏检率,在特征融合结构中引入增大感受野模块(RFB)^[20](改进 2).不同层数的感受野 RF_{n+1} 为

$$RF_{n+1} = RF_n + (k_{n+1} - 1) \times \prod_{j=1}^n s_j, \quad j = 1, \cdots, n.$$

(1)

式(1)中: n 为层数; k 为卷积核尺寸; s 为步长.

表 1 聚类后的先验框尺寸

Tab. 1 Prior box size after clustering

先验框尺寸	感受野		
	大	中	小
尺寸 1	(27,37)	(11,16)	(4,6)
尺寸 2	(37,53)	(15,21)	(6,9)
尺寸 3	(61,83)	(20,28)	(8,12)

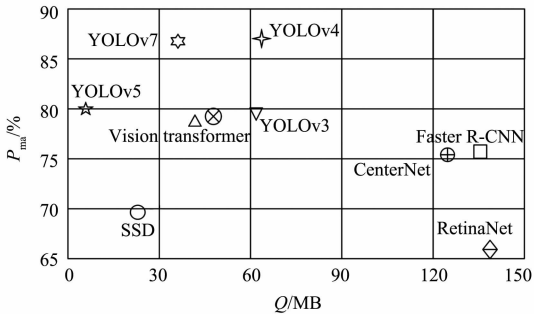


图 3 主流算法模型在 XMUT-HELMET 数据集上的检测指标

Fig. 3 Detection indicators of mainstream algorithm models on XMUT-HELMET data sets

表 2 EfficientNet 的参数量和浮点运算量^[19]

Tab. 2 Number of parameters and floating point operations of EfficientNet^[19]

网络结构	Q/MB	FLOPs/ $\times 10^9$
EfficientNet-B0	5.3	0.39
EfficientNet-B1	7.8	0.70
EfficientNet-B2	9.2	1.00
EfficientNet-B3	12.0	1.80
EfficientNet-B4	19.0	4.20
EfficientNet-B5	30.0	9.90
EfficientNet-B6	43.0	19.00
EfficientNet-B7	66.0	37.00

RFB 从卷积核尺寸着手,使用不同膨胀率的多分支空洞卷积,有效地增大卷积核,从而达到模拟人类视觉,增大感受野的目的. RFB 结构,如图 4 所示. 将 RFB 引入模型第 1 个负责小尺寸头盔检测的低维特征分支,防止细粒度特征(如头盔颜色、椭圆形边缘等)语义信息的丢失.

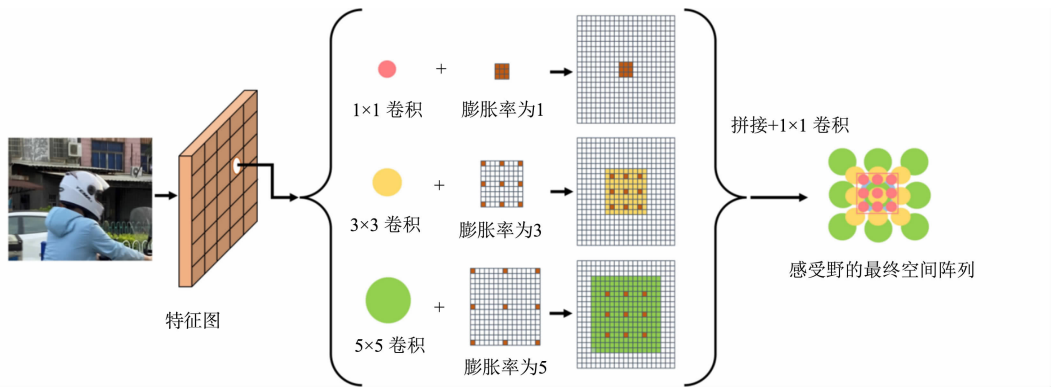


图 4 RFB 结构

Fig. 4 RFB structure

2.3 检测头的改进

YOLOv7 中的检测头先使用 RepConv 对特征融合结构的输出进行提炼,再进行头盔目标框的特征解码. 鉴于二轮车头盔检测场景的复杂性和强干扰性,若能在特征解码前使用注意力机制纠正网络感兴趣区域偏差,从大量全局信息中获取头盔重点关注区域,则能进一步降低模型漏检率和误检率.

常用的注意力机制有高效的通道注意力(ECA)^[21]等,它们大多是单一对通道或空间的神经元进行权重学习,进行权值分配时往往会缺失某些必要特征(如头盔轮廓信息等),而 SimAM^[22] 机制能更好地模拟视觉信号处理过程,使通道和空间注意力协同工作,从而对乘员头盔主要特征区域进行强调,对马路、红绿灯和公交站台等背景次要干扰特征进行抑制,且 SimAM 机制不会引入额外的参数,有利于实际部署. SimAM 机制的操作主要由能量值计算(CEV)和特征加权(WF)两部分组成,首先,通过 CEV 操作获取单个通道中神经元的能量值,再基于该能量值使用 WF 操作,对输入的图像特征进行加权,CEV 操作的描述为

$$e(T_i)=\frac{4(\hat{\sigma}^2+\lambda)}{(T_i-\hat{u})^2+2\hat{\sigma}^2+2\lambda}, \quad i=1,2,3,\cdots,W\times H, \tag{2}$$

式(2)中: T_i 为第 i 个通道中的神经元; W, H 分别为特征图的宽和高; $\hat{u}, \hat{\sigma}^2$ 分别为该通道内所有神经元的均值和方差; λ 为超参数,根据工程经验设置 $\lambda=0.0001$; $e(T_i)$ 为神经元 T_i 的能量值.

信息丰富的神经元通常表现出与周围神经元不同的放电模式,而且会抑制周围神经元,某个神经元与通道内其他神经元之间的差异越大,则对模型的推理过程越重要^[22]. $(T_i-\hat{u})^2$ 越大,神经元 T_i 越重要,而计算所得的能量值 $e(T_i)$ 越小. 因此,单个神经元 T_i 的能量值与其重要程度呈负相关.

WF 操作的描述为

$$\tilde{\mathbf{X}}=\text{Sigmoid}(1/\mathbf{E})\odot\mathbf{X}. \tag{3}$$

式(3)中: $\tilde{\mathbf{X}}$ 为头盔输出特征; \mathbf{E} 为输入特征 \mathbf{X} 在所有通道和空间维度上计算 $e(T_i)$ 后形成的张量,根据前述理论,张量 $1/\mathbf{E}$ 中单个通道的值越大,该通道对应的神经元越重要;Sigmoid 函数是为了防止张量 $1/\mathbf{E}$ 中的元素值过大,由于它是单调递增函数,故这种函数变换不会影响每个神经元的相对重要性;运算符 \odot 表示张量各对应位置元素相乘.

经过 WF 操作后,可得到加权后的头盔输出特征 $\tilde{\mathbf{X}}$. 在模型的检测头部分引入 SimAM 机制(改进 3),可提高模型在高语义特征中对头盔周围感兴趣区域的关注程度,防止颜色相近的广告牌、路灯等强干扰因素对检测过程造成负面影响.

2.4 算法模型的整体结构

综合上述参考模型和改进内容,提出二轮机动车乘员头盔检测算法(文中算法),用于实现密集目标

和随机遮挡情况下的二轮机动车乘员头盔佩戴检测。二轮机动车头盔检测算法模型的整体结构,如图 5 所示。该模型分为主干网络、特征融合结构和检测头 3 个部分。主干网络 EfficientNet-B3 通过多个模块的堆叠不断提取输入图像特征,首先,是一个包含 BN 和 Swish 激活函数的 3×3 卷积层,然后,是 7 个重复堆叠的 MBConv 结构,最后,是 1 个由 1×1 卷积层和 1 个平均池化层组成的 CP 模块。特征融合结构中沿用了 YOLO 系列模型的 FPN 和 PANet 结构思想,先进行上采样扩大特征图尺寸,再使用池化的方式提炼上、下文特征,并通过跨层连接将它们合并,最后得到用于预测的有效特征。检测头基于预先设定好的交并比(IoU)阈值和置信度阈值进行非极大值抑制,以得到较好的头盔检测结果。

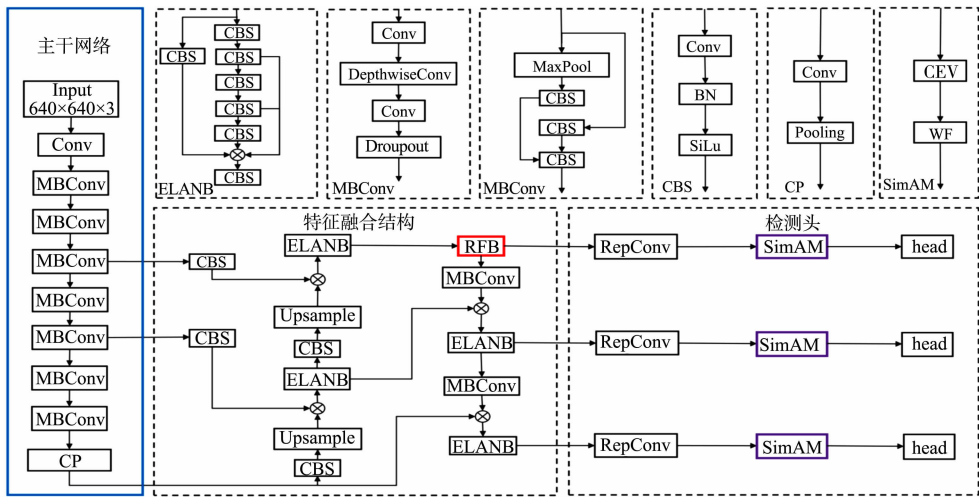


图 5 二轮机动车乘员头盔检测算法的整体结构

Fig. 5 Overall structure of helmet detection algorithm of two-wheeled motor vehicle occupant

3 实验结果与分析

3.1 评价指标

考虑到交管部门的应用需求,选择准确率(P)、召回率(R)、 P_{ma} ($IoU=0.5$)、参数量、浮点运算量和帧率作为模型性能的评估指标。准确率表示预测样本中实际正样本数占有所有正样本数的比例;召回率表示预测样本中实际正样本数占有所有预测的样本数的比例;参数量和浮点运算量反映模型占用资源的大小;帧率反映模型推理速度。 P_{ma} 的计算原理为

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad (4)$$

$$P_a = \int_0^1 P(R) d(R), \quad (5)$$

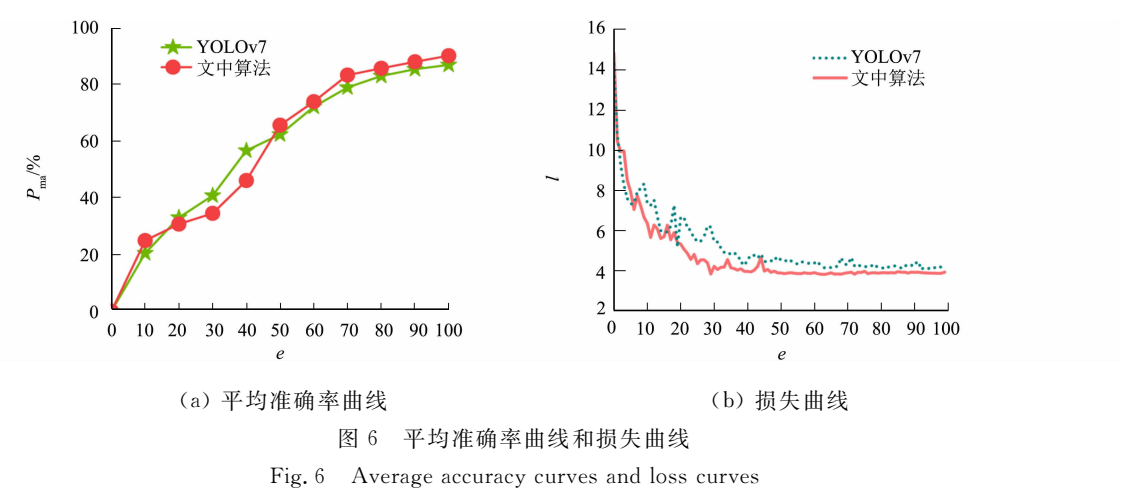
$$P_{ma} = \frac{\sum_k P_{a,r}}{k}, \quad r = 1, 2, 3, \dots, k. \quad (6)$$

式(4)~(6)中: P_a 为单个类别的平均精度;TP为真正例,表示实际为正例且被分类器划分为正例的样本数;FP为假正例,表示实际为负例,但被分类器划分为正例的样本数;FN为假反例,表示实际为正例,但被分类器划分为负例的实例数。

3.2 模型训练

实验基于 Windows 10 \times 64 平台,且通过 Python 语言编程实现,使用深度学习框架 Pytorch 1.8.0 搭建网络模型。CPU 为 Intel(R) Core(TM) i7-10700K CPU@3.80 GHz,内存为 32 GB,GPU 为 NVIDIA RTX3090,显存为 24 GB。

模型训练时,设置批尺寸为 12,初始学习率为 0.001,使用 EfficientNet-B3 预训练权重进行迁移学习。训练的平均准确率和损失(l)曲线,如图 6 所示。经过 100 个轮次(e)的训练后,YOLOv7 的平均准确率为 87.10%,而文中算法的平均准确率达到 90.36%,提升了 3.26%,损失降至 3.92。



3.3 实验结果分析

在相同的训练设备和参数配置下,使用 XMUT-HELMET 数据集分别对 YOLOv7 和文中算法进行训练,训练后使用测试集中的图片进行推理. YOLOv7 和文中算法检测结果对比,如图 7 所示.



由图 7 可知:在交通密集和随机遮挡场景下,通过 YOLOv7 进行头盔检测的效果不甚理想,鲁棒性较差,如图 7(a)中间有一个佩戴黄色头盔的骑行人员被部分遮挡,导致漏检,图 7(b)左侧有一个正常佩戴头盔的骑行人员被座位后方人员遮挡,导致其被误检,图 7(c)的左侧有一个未佩戴头盔的骑行人员被立柱式红绿灯部分遮挡,右侧有一个正常佩戴头盔的人员被密集分布的其他人员遮挡,这均被 YOLOv7 忽略,而文中算法在这些情况下依然具备优良的检测能力.

为了分析改进 1~改进 3 对模型检测效果的影响,进行消融实验(实验 A~实验 C),结果如表 3 所示.由表 3 可知以下 3 个结论:1) 采用 EfficientNet-B3 作为主干特征提取网络后,准确率、召回率、平均

表 3 消融实验结果									
Tab. 3 Results of ablation experiment									
算法	改进 1	改进 2	改进 3	$P/\%$	$R/\%$	$P_{ma}/\%$	Q/MB	FLOPs/ $\times 10^9$	帧率/ $F \cdot s^{-1}$
YOLOv7	×	×	×	89.62	81.94	87.10	37.20	43.780	50.74
实验 A	✓	×	×	91.31	82.23	89.08	10.65	9.331	68.28
实验 B	✓	✓	×	91.59	83.90	89.75	12.33	9.856	65.39
实验 C	✓	✓	✓	92.46	84.20	90.36	12.33	10.280	64.57

准确率分别提升了 1.69%,0.29%,1.98%,最主要的是模型的参数量和浮点运算量分别减少了 26.55 MB 和 34.45×10^9 ,帧率提升了 17.54 $F\cdot s^{-1}$;2) 在特征融合结构中引入 RFB 后,准确率提升较少,仅为 0.28%,但召回率增加了 1.67%,说明模块对小尺度、部分被遮挡的头盔的检测效果较好,降低了漏检率;3) 在检测头部分引入 SimAM 机制后,准确率、召回率和平均准确率分别增加了 0.87%,0.30% 和 0.61%,算法排除无关干扰的能力有所提升。

综合这 3 项改进,文中算法实现了提升目标检测效果、降低参数量和运算量的目的,这有利于将该算法部署于智能交通设施等边缘设备中。

为进一步验证文中算法的先进性,选用 Faster R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5, YOLOv7, CenterNet, RetinaNet, Vision transformer-DETR, Swin Transformer-Mask RCNN^[24] 等算法与文中算法进行对比实验,结果如表 4 所示。由表 4 可知:文中算法在准确率、召回率、平均准确率、浮点运算量和帧率上都达到了当前最佳值,仅在模型参数量上为次优(12.330 MB),比其他算法更具有优越性;文中算法的参数量和浮点运算量分别为 YOLOv7 的 33.1%,23.5%,相较于 YOLOv7,准确率、召回率、平均准确率分别提高了 2.84%,2.26%,3.26%。

表 4 各模型检测性能对比
Tab. 4 Comparison of detection performances of each model

算法	$P/\%$	$R/\%$	$P_{ma}/\%$	Q/MB	$FLOPs/\times10^9$	帧率/ $F\cdot s^{-1}$
Faster R-CNN	82.17	71.36	76.690	137.100	402.00	26.28
SSD	73.85	61.21	69.380	24.010	274.70	52.07
YOLOv3	82.16	76.69	79.550	61.580	155.10	54.21
YOLOv4	90.47	82.09	88.590	63.960	59.98	36.36
YOLOv5	84.46	78.39	80.250	7.080	16.40	58.77
YOLOv7	89.62	81.94	87.100	37.200	43.78	50.74
CenterNet	90.32	66.51	76.030	124.600	46.20	56.38
RetinaNet	89.07	51.92	66.810	140.000	69.73	32.83
Vision transformer-DETR	74.36	83.61	78.600	41.000	86.00	53.05
Swin Transformer-Mask RCNN	75.20	82.97	79.010	48.230	267.10	43.82
文中算法	92.46	84.20	90.360	12.330	10.28	64.57

此外,为验证文中算法在边缘设备上的部署能力,在 NVIDIA Jetson Nano 开发板上进行实验。该开发板采用 4 核 64 位 ARM CPU 和 128 核集成 NVIDIA GPU,提供每秒浮点运算量为 472×10^9 的计算性能。结果显示,文中模型的平均准确率为 90.24%,帧率达到 47.58 $F\cdot s^{-1}$,而同等情况下 YOLOv7 的平均准确率为 86.98%,帧率为 18.05 $F\cdot s^{-1}$,说明文中算法能在计算资源有限的情况下实现较为准确、快速的推理,具备优良的轻量化部署能力。

4 结束语

为了实现密集目标和随机遮挡下头盔佩戴情况快速、准确的检测,制作二轮机动车在复杂交通环境下的图像数据集,并以此为基础参考 YOLOv7 提出一种轻量级的二轮机动车乘员头盔检测算法。实验结果表明,在该数据集上的头盔佩戴检测任务中,文中算法在准确率、召回率、平均准确率、浮点运算量和帧率上都达到了当前最佳,并在边缘设备上有更好的部署能力,有利于在现有智能交通设施或警用车辆上应用推广。然而,从实验结果来看,文中模型的参数量仍高于 YOLOv5,存在进一步优化的空间。未来将采取模型剪枝、量化和蒸馏等措施,进一步缩减参数量。此外,文中模型对检出目标的置信度不高,这可能是由于稠密交通环境下二轮机动车背景对模型推理过程仍存在较大干扰,今后将借助 CAM 热力图等手段,探究模型推理过程的可解释性,并使用其他有效策略对模型进行进一步优化。

参考文献:

[1] World Health Organization. Global status report on road safety 2018: Summary [R]. Geneva:WHO,2018.
[2] LI Xiaofei,FABIAN F,YANG Yue,*et al.* A new benchmark for vison-based cyclist detection[C]//Proceedings of

- the IEEE Intelligent Vehicles Symposium (IV). Gothenburg: IEEE Press, 2016: 1109-1114.
- [3] LIN Hanhe, FELIX W S. HELMET dataset[EB/OL]. (2020-03-06)[2022-10-23]. <https://osf.io/4pwj8/>.
 - [4] 陈闯闯, 胡绍方. 密集场景下头盔佩戴智能检测研究[J]. 智能计算机与应用, 2020, 10(9): 223-224.
 - [5] WANG Wei, GAO Song, SONG Renjie. A safety helmet detection method based on the combination of SSD and HSV color space[M]//KIM H. IT Convergence and Security. Washington D C: Mineralogical Society of America, 2021: 117-211. DOI:10.1007/978-981-15-9354-3_12.
 - [6] 刘琛, 王江涛, 王明阳. 引入视觉机制的 SSD 网络在摩托车头盔佩戴检测中的应用[J]. 电子测量与仪器学报, 2021, 35(3): 144-151. DOI:10.13382/j.jemi.B2003332.
 - [7] 冉险生, 陈卓, 张禾. 改进 YOLOv2 算法的道路摩托车头盔检测[J]. 电子测量技术, 2021, 44(24): 105-115. DOI: 10.19651/j.cnki.emt.2107718.
 - [8] 冉险生, 张之云, 陈卓, 等. 基于改进 DeepSORT 算法的摩托车头盔佩戴检测[EB/OL]. (2022-07-27)[2022-10-27]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20220726.1653.016.html>.
 - [9] 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, 2016, 39(6): 1137-1149. DOI:10.1109/tpami.2016.2577031.
 - [10] LIU Wei, ANGUELOV D, ERHAN D, *et al.* Ssd: Single shot multibox detector[C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37. DOI:10.1007/978-3-319-46448-0_2.
 - [11] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2022-10-23]. <https://doi.org/10.48550/arXiv.1804.02767>.
 - [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2022-10-23]. <https://doi.org/10.48550/arXiv.2004.10934>.
 - [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL]. (2022-07-06)[2022-10-23]. <https://doi.org/10.48550/arXiv.2207.02696>.
 - [14] ZHOU Xingyi, WANG Dequan, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-16)[2022-10-23]. <https://doi.org/10.48550/arXiv.1904.07850>.
 - [15] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE Press, 2017: 2980-2988. DOI:10.1109/iccv.2017.324.
 - [16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16×16 words: Transformers for image recognition at scale[ED/OL]. (2020-10-22)[2022-10-25]. <http://www.arxiv-vanity.com/papers/2010.11929/>.
 - [17] TAN Mingxing, LE Q V. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. Long Beach: [s. n.], 2019: 6105-6114. DOI:10.48550/arXiv.1905.11946.
 - [18] 柳长源, 何先平, 毕晓君. 融合注意力机制的高效率网络车型识别[J]. 浙江大学学报(工学版), 2022, 56(4): 775-782. DOI:10.3785/j.issn.1008-973X.2022.04.017.
 - [19] 陶英杰, 张维伟, 马昕, 等. 面向无人机视频分析的车辆目标检测方法[J]. 华侨大学学报(自然科学版), 2022, 43(1): 111-118. DOI:10.11830/ISSN.1000-5013.202011014.
 - [20] LIU Songtao, HUANG Di. Receptive field block net for accurate and fast object detection[C]//Proceedings of the European Conference on Computer Vision. Munich: [s. n.], 2018: 385-400. DOI:10.48550/arXiv.1711.07767.
 - [21] WANG Qilong, WU Banggu, ZHU Pengfei, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Press, 2020: 13-19. DOI:10.48550/arXiv.1910.03151.
 - [22] YANG Lingxiao, ZHANG Ruyuan, LI Lida, *et al.* Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine learning. [S. l.]: IEEE Press, 2021: 11863-11874.
 - [23] 王年涛, 王淑青, 黄剑锋, 等. 基于改进 YOLOv5 神经网络的绝缘子缺陷检测方法[J]. 激光杂志, 2022, 43(8): 60-65. DOI:10.14016/j.cnki.jgzz.2022.08.060.
 - [24] LIU Ze, LIN Yutong, CAO Yue, *et al.* Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE Press, 2021: 10012-10022. DOI:10.48550/arXiv.2103.14030.

(责任编辑: 钱筠 英文审校: 吴逢铁)