

改进YOLOv5s的摩托车头盔佩戴检测算法

张鑫^{1,2}, 周顺勇^{1,2}

(1. 四川轻化工大学自动化与信息工程学院, 四川 宜宾 644000;

2. 人工智能四川省重点实验室, 四川 宜宾 644000)

摘要:针对摩托车头盔佩戴检测准确率低和检测速率慢的问题,提出一种基于YOLOv5s的改进摩托车头盔检测算法。首先,在YOLOv5s的多尺度特征检测中增加浅层检测尺度和4倍上采样特征融合结构,以提升检测准确率。其次,引入卷积注意力模块(Convolutional Block Attention Module, CBAM),以提升对聚集目标的关注,有效解决因遮挡、重叠导致的漏检和误检问题。最后,使用MobilenetV3的Block结构替换主干网络及颈部中的瓶颈结构,实现了降低网络参数量的目的。实验结果表明,相较于YOLOv5s算法,改进算法的*mAP*提高了2.91%,检测速率达到了36 frame/s,在保证较高检测速率的同时检测精度更高,具有一定的应用价值。

关键词:头盔检测;YOLOv5s模型;CBAM注意力机制;MobilenetV3网络

中图分类号:TP391

文献标志码:A

引言

2018年世界卫生组织发布的全球道路安全报告中^[1],摩托车交通事故导致的死亡人数占道路交通事故总死亡人数的28%。特别是在东南亚、西太平洋区域,道路交通工具以摩托车为主,摩托车交通事故死亡人数分别占总交通事故死亡人数的43%和36%。而头部受伤是导致死亡的最主要因素,如果规范佩戴摩托车头盔,死亡风险可降低42%,头部受伤风险可降低69%,从而降低摩托车事故的死亡人数。目前,对摩托车骑手是否佩戴头盔的检查仍以交警现场检查为主。随着计算机视觉技术的快速进步,智能设备在目标检测中的使用越

来越普遍,深度学习是目前的主流的方法之一。

基于深度学习的目标检测算法分为二阶段算法和一阶段算法两类。二阶段算法有Fast R-CNN^[2]、Faster R-CNN^[3]、Mask R-CNN^[4]等,虽然这类算法精度高,但是检测速度慢,不能满足实时检测的需求。一阶段算法包括YOLO系列算法^[5-7]、SSD^[8]、FCOS^[9]等,兼备高精度和小参数量模型等特点。Wu等^[10]使用混合描述符提取图像特征,并使用支持向量机(SVM)分类器将图像分为头盔和非头盔。Silva等^[11]应用了背景减法技术来识别移动车辆和主成分分析。Shine^[12]将背景减法与序列最小优化(Sequential Minimal Optimization, SMO)分类器用于检测视频中的摩托车。

收稿日期:2022-03-27

基金项目:四川省科技厅项目(2020YFG0178);四川省科技厅省院校企合作项目(2020YFSY0027)

通信作者:周顺勇(1975-),男,副教授,硕士,研究方向为图像处理,(E-mail)14409186@qq.com

主干特征网络用于提取图像特征:首先, FOCUS将原特征图每隔一个像素取一个值,类似邻近值采样,从而得到4张特征图,把W、H的信息就集中到了通道空间,减少计算量来加速模型。在主干特征网络的最后一层用了空间金字塔池化(Spatial Pyramid Pooling, SPP)来扩大感受野。根据YOLOv4中CSPnet的基本理论,设计了CSP1_X和CSP2_X,前者用于主干特征网络中,后者用于颈部结构中。在颈部中采用特征金字塔(Feature Pyramid Network, FPN)加路径聚合网络(Path Aggregation Network, PAN)的结构,FPN结构可以从上往下传达语义信息,然后PAN从下往上传达准确的定位信息,它们一起构成特征金字塔,融合不同特征层之间的特征信息。头部预测部分与YOLOv4一致,只有3个特征预测输出,预测信息包括边框信息、类别和概率。

2 改进的YOLOv5网络

2.1 多尺度特征的改进

YOLOv5s对常规数据集有很好的检测效果,但这些待检测目标通常较大,而摩托车头盔的检测属于小目标,在提取图像的信息特征时,经过多次卷积激活运算后,会影响小目标的语义信息,导致检测精度降低。为了解决这个问题,在颈部结构中增加一个尺度窗口,添加高层特征层来改进YOLOv5的多尺度检测^[19]。在保证输入大小为 640×640 的前提下,增加浅层特征大小为 160×160 的检测层,将原本三尺度特征改进为四尺度,极大增强了对小目标的检测能力。另外,为了最大程度实现不同特征层之间的特征融合,进一步减少多次卷积特征信息的损失,新增四倍上采样特征融合结构,提高对头盔的检测精度。

2.2 CBAM注意力机制

在检测摩托车头盔佩戴时,由于车身间的遮挡和车上人物重叠导致误检和漏检,降低了检测精

度。为解决这些问题,特引入CBAM^[20],该模块包括通道注意力模块(CAM)和空间注意力模块(SAM)两个独立的部分,优化了网络的特征提取能力,有效地提高了检测的精度。给定一个输入特征图为 $F \in R^{C \times H \times W}$,依次计算出一维特征图 $M_c \in R^{C \times 1 \times 1}$ 和二维特征图 $M_s \in R^{1 \times H \times W}$,计算过程如式(1)~(2)所示:

$$F' = M_c(F) \otimes F \quad (1)$$

$$F'' = M_s(F') \otimes F' \quad (2)$$

其中, \otimes 表示对应元素相乘, F' 为通道特征的输出, F'' 为空间特征的输出。将二者通过多层感知机(MLP)和隐藏层组成共享网络,得到通道注意力特征图 $M_c(F) \in R^{C \times 1 \times 1}$,其计算过程如式(3)所示:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (3)$$

其中, σ 表示激活函数sigmoid, $AvgPool(F)$ 表示平均池化, $MaxPool(F)$ 表示最大值池化, MLP 表示多层感知机的运算, $W_0 \in R^{C \times C}$ 和 $W_1 \in R^{C \times C}$ 是 MLP 的两次权值。 F_{avg}^c 和 F_{max}^c 分别表示平均池化特征和最大池化特征。

为计算空间注意力特征,首先沿通道轴计算平均池化和最大池化,将特征图压缩到一层,生成两个二维的特征: $F_{avg}^s \in R^{1 \times H \times W}$ 和 $F_{max}^s \in R^{1 \times H \times W}$,表示将 C 维通道压缩后得到一维通道,该通道具有平均池化和最大池化。卷积后得注意力特征图 $M_s(F) \in R^{1 \times H \times W}$,其计算过程如式(4)所示:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}(F_{avg}^s; F_{max}^s)) \quad (4)$$

其中, σ 表示激活函数, $f^{7 \times 7}$ 表示卷积核大小为 7×7 的标准化卷积。

CBAM注意力模块可以增强特征图中的重要通道和空间特征,有效提高对目标检测的定位精度,

增强对目标聚集的关注,解决由遮挡重叠而引起的误检、漏检问题。所以在改进YOLOv5s中的颈部结构中,每一次瓶颈结构后面都添加了CBAM注意力模块,另外在最后的四尺度预测层输出之前,也融入了CBAM注意力模块,其结构如图2所示。

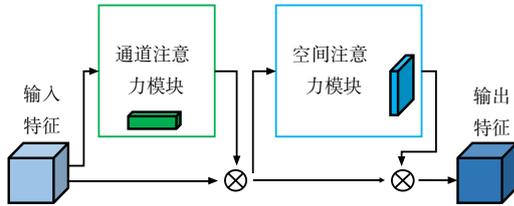


图2 CBAM注意力机制模块

2.3 MobilenetV3轻量化模块

MobilenetV3网络^[21]不但融合了MobilenetV1^[22]的深度可分离卷积、MobilenetV2^[23]的具有线性瓶颈的逆残差结构与MnasNet^[24]的轻量级注意力模型,它还融入了SE-net^[25]神经网络,即引入一种非线性swish函数,其计算如式(5)所示:

$$swish[x] = x \times \sigma(x) \tag{5}$$

其中,swish[x]表示非线性激活函数,x表示输入特征,σ(x)表示sigmoid激活函数。

虽然这种非线性计算提高了检测精度,但是在移动设备上计算sigmoid成本太高,故采用一种新的激活函数h-swish[x]来代替原来的swish[x]函数,这能极大提升检测速度,尤其是在深层网络中较为明显。其计算过程如式(6)所示:

$$h-swish[x] = x \times \frac{Relu6(x + 3)}{6} \tag{6}$$

其中Relu(x)表示Relu激活函数。

YOLOv5s主干特征网络卷积运算很多,为了极大地降低计算开销且保持较好精度,将MobilenetV3网络的Block结构替换为YOLOv5s主干的瓶颈结构,将CSP1_X中的残差卷积网络改为参数量更少的Block结构来完成特征提取。另外,在颈部中也采用该Block结构替换CSP2_X中的残差卷积,使模

型体积降低。MobilenetV3网络的Block结构如图3所示。

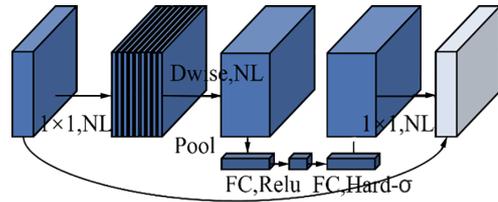


图3 MobilenetV3的Block结构

为了提高YOLOv5s模型的检测精度,本文采用了一系列改进措施,如图4所示。首先,为了避免多次卷积后丢失较小目标区域信息,增加了一个浅层来改进多尺度检测,将原来的三尺度特征扩展为四尺度特征(图4(a))。针对输入尺寸为640×640×3的图片,采用了FOCUS模块(图4(b))对其进行切片,将宽、高都裁剪为原来的一半且通道数扩张为四倍,得到320×320×12的特征图。接着,通过卷积操作将其转换为尺寸为160×160×128的特征图。在主干特征网络中,将MobilenetV3的Block结构替换残差结构CSP1_X,以此得到模块的CV(图4(c)),该模块从上至下依次输出另外4个特征图,尺寸大小分别为160×160×128、80×80×256、40×40×512和20×20×1024。为了更好地利用低层和高层的信息,并减少由于多次卷积造成的信息丢失,在原本两倍上采样的基础上新增四倍上采样,并在特征金字塔中实现不同层之间的特征融合。通过在CV的输出末端前引入CBAM模块,得到了CVB模块(图4(d)),从而减少了目标聚集问题,有效提高了遮挡和重叠目标检测中的目标定位精度。主干特征网络中卷积层CBS和池化层SPP模块结构如图4(e)和图4(f)所示,在颈部结构的输出后以及头部预测末端输出前嵌入CBAM结构,增强了各个尺度的特征信息提取,最终输出4个尺度的特征图,改进后的YOLOv5s模型如图4(a)所示。

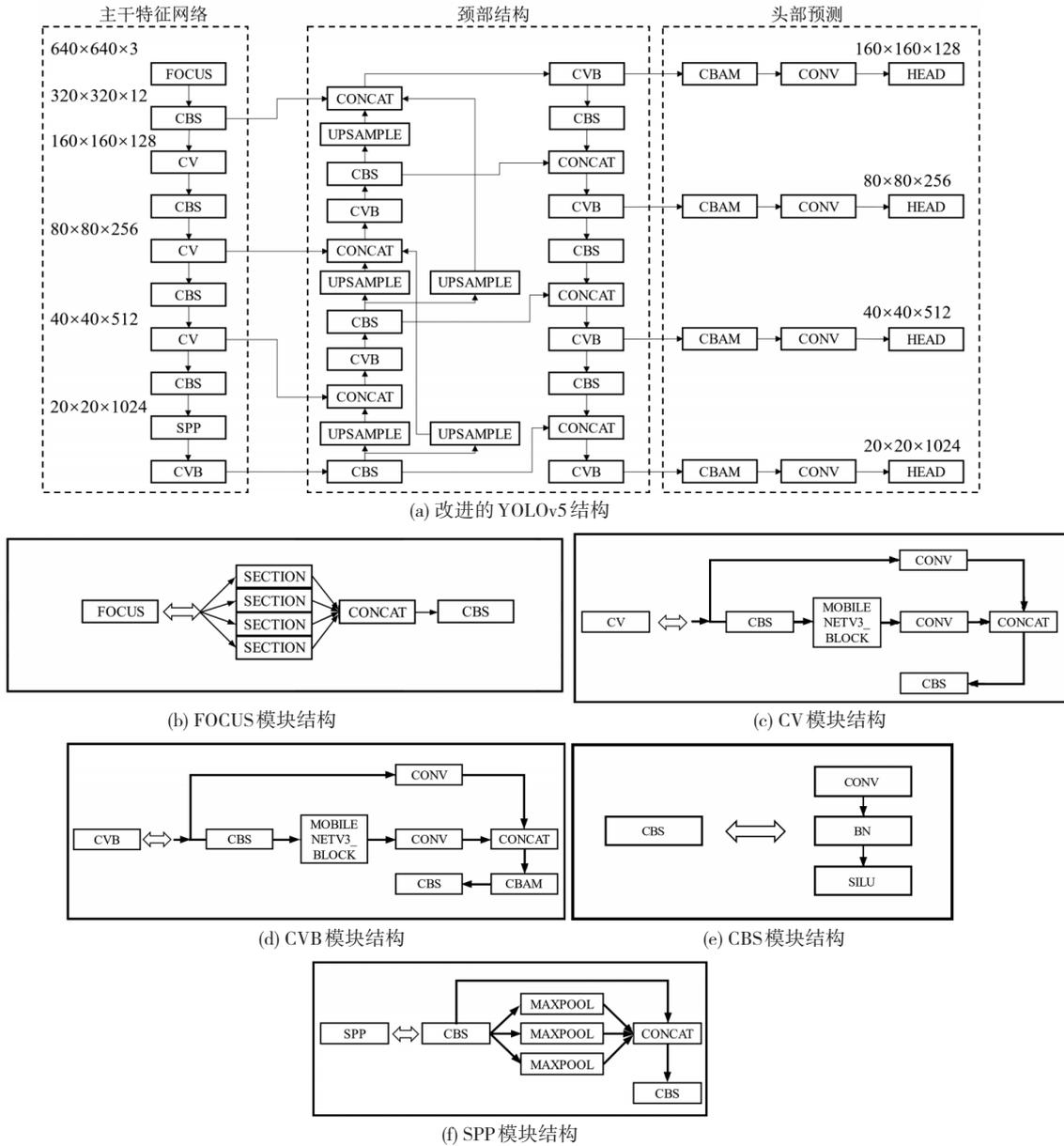


图4 改进YOLOv5s网络整体模型

3 实验结果与分析

3.1 实验环境

本实验采用 AMD Ryzen 5800H with Radeon Graphics 处理器, 内存为 16 G, 显卡为 NVIDIA GeForce RTX 3060 Laptop GPU, 其显示内存大小为 6 G。程序编译采用 Python 语言和 Pytorch 深度学习框架, 并使用 GPU 加速软件 CUDA11.1 和 CUDNN8.1.1。实验参数设置见表 1。在该实验中, 采用的数据增强方法有旋转、缩放、平移以及 Mosaic 数据增强, 它可以改善图像的对比度、噪声抑制、色

彩增强等, 此外还使用了自适应锚框计算方法 K-means 聚类, 对先验框进行聚类分析, 得到最优的初始锚框大小。

表 1 实验参数设置

| 参数名称 | 值 |
|---------------|------|
| Learning rate | 0.01 |
| Train epoch | 200 |
| Batch size | 16 |
| Momentum | 0.92 |
| Weight decay | 0.05 |
| Confidence | 0.5 |
| Nms threshold | 0.3 |

3.2 数据集

本实验数据集来自3个部分,分别为Kaggle^[26]、ImageNet大型数据集的子集和网络图像,将其组合后进行自定义标注,标注后得数据集包括2200张具有良好场景光照且分辨率较高的图像,1100张场景光照较暗的图像以及500张未佩戴安全头盔的摩托车驾驶场景。图像数据标注使用的是Labelimg软件,通过对数据集标注了两个类别,分别为佩戴安全头盔(helmet_man)与未佩戴安全头盔(no_helmet_man),并生成包含物体坐标和类别信息的标注文件,文件格式为XML,后经YOLOv5转换至txt文件即Yolo格式,以供训练使用。

此外,为了反映实际的检测场景,将训练集、验证集和测试集的比例设定为8:1:1,数据集的详细划分情况见表2。

表2 数据集

| 数据类型 | 数量 |
|-------|------|
| 训练集 | 3040 |
| 验证集 | 380 |
| 测试集 | 380 |
| 全部数据集 | 3800 |

3.3 训练与分析

为了定量评估模型的检测效果,使用准确率(Precision)、召回率(Recall)、平均精度(AP)、F1、mAP以及FPS来衡量算法的性能。

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

其中,TP表示判断为正样本实际为正样本的目标,FP表示判断为正样本实际为负样本的目标,FN表示判断为负样本实际为正样本的目标,AP表示Precision-Recall曲线包围的面积。为验证不同算法的有效性和性能,对改进YOLOv5s算法与YOLOv3、YOLOv4、YOLOv4-Tiny、SSD、Autoassign、Deformable DETR、YOLOv5s、YOLOv5m、YOLOv5l进行对比分析。首先,YOLOv5s模型以及改进YOLOv5s模型的边界框损失函数值随训练轮次曲线变化如图5所示。其次,验证了各种算法的检测性能对比,其数据对比见表3。最后,为了更直观地观察改进前、后模型的效果,对测试集的部分图像进行可视化分析,其检测效果如图6所示。

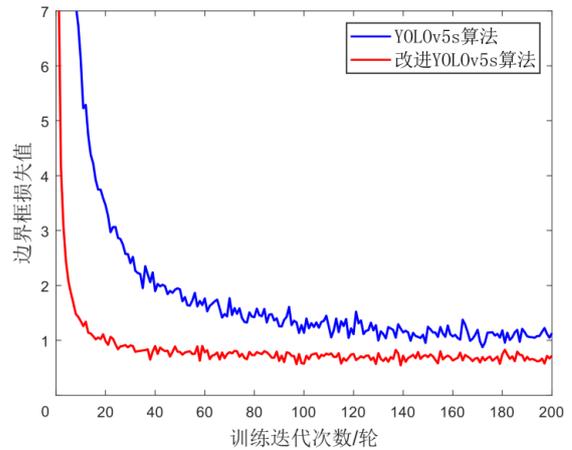


图5 改进前、后边界框损失值随迭代次数变化曲线对比

表3 各种算法检测性能对比

| 模型 | 模型大小/MB | Precision/% | Recall/% | F1 | mAP/% | FPS/(frame/s) |
|-----------------|---------|-------------|----------|------|-------|---------------|
| YOLOv5s | 14.48 | 86.62 | 88.53 | 0.88 | 92.83 | 38 |
| YOLOv5m | 41.92 | 87.09 | 89.92 | 0.89 | 94.35 | 27 |
| YOLOv5l | 91.65 | 88.90 | 92.82 | 0.91 | 95.48 | 18 |
| YOLOv3 | 236.32 | 84.22 | 85.36 | 0.85 | 90.17 | 19 |
| YOLOv4 | 244.29 | 90.66 | 93.25 | 0.91 | 95.38 | 20 |
| YOLOv4-Tiny | 22.85 | 83.57 | 82.22 | 0.83 | 88.06 | 31 |
| SSD | 102.17 | 82.13 | 81.26 | 0.81 | 86.79 | 14 |
| Autoassign | 277.06 | 89.25 | 92.07 | 0.90 | 93.74 | 17 |
| Deformable DETR | 41.64 | 88.65 | 91.22 | 0.90 | 92.12 | 22 |
| 改进YOLOv5s | 23.26 | 90.76 | 93.58 | 0.92 | 95.74 | 36 |



图6 检测效果对比图

通过对图5边界框损失值曲线对比分析,可以直观地看出,改进后的YOLOv5s算法收敛速度相比YOLOv5s算法更快,在40轮后就逐渐趋于稳定。另外改进算法的损失值也更小,稳定后的损失值由1.1降到了0.7左右,体现更优的算法性能。通过表3可知,改进YOLOv5s模型的 $F1$ 值达到了0.92, mAP 达到了95.74%,在原算法上提升了2.91%。根据图6的检测可视化效果对比,可以看出改进YOLOv5s模型能够正确地检测出遮挡和重叠的目标。在检测速度方面,改进算法比YOLOv4-Tiny提升了5 frame/s,检测速度达到了36 frame/s,在嵌入式设备Jetson nano b01中通过TensorRT加速后FPS能从3 frame/s提升到13~15 frame/s,能够满足实际场景的实时检测需求。

3.4 消融实验

为了更加直观而严谨地评估各个改进点对YOLOv5s模型的增益效果,对3种改进及其组合改进的模型编号:改进a表示采用多尺度窗口,改进b表示嵌入CBAM注意力机制,改进c表示用轻量化模型MobilenetV3替换主干特征网络。在同一环境

下,按照相同的数据集的分配比例和参数进行消融实验,实验结果见表4。

表4 消融实验对比

| 模型 | a | b | c | $mAP/\%$ | $FPS/(frame/s)$ |
|----|---|---|---|----------|-----------------|
| A | √ | — | — | 94.66 | 34 |
| B | — | √ | — | 93.57 | 36 |
| C | — | — | √ | 92.94 | 39 |
| D | √ | √ | — | 95.63 | 33 |
| E | √ | — | √ | 94.37 | 35 |
| F | — | √ | √ | 93.26 | 36 |
| G | — | — | — | 92.83 | 38 |
| H | √ | √ | √ | 95.74 | 36 |

注:“√”表示使用此种改进;“—”表示未使用此种改进

由表4可知,YOLOv5s模型经a、b、c 3种改进后在检测性能方面均有小幅度提升,综合所有模型进行对比分析,检测精度和速度均衡效果最佳的是H模型,其 mAP 最优,且检测速度达到了36 frame/s,较原模型G有较大提升,具有一定的泛化性。

4 结束语

本文提出了一种改进YOLOv5s的摩托车头盔佩戴检测算法,实现了对车身遮挡、骑乘人员重叠

情况下的头盔检测精度的提升。实验结果表明,在各种算法的检测性能对比中,改进算法的检测精度有较大提升,并且能够很好地满足实时性要求,可以植入到计算能力和内存有限的嵌入式设备和移动设备中,具有很好的实用性和应用前景。

参考文献:

- [1] WHO.Global status report on road safety 2018[R].World Health Organization,2018.
- [2] DING X,LI Q,CHENG Y,et al.Local keypoint-based faster R-CNN[J].Applied Intelligence,2020,50(10):3007-3022.
- [3] CHEN Y H,WANG H R,LI W,et al.Scale-aware domain adaptive faster R-CNN[J].International Journal of Computer Vision,2021,129(7): 2223-2243.
- [4] TIAN Y,YANG G,WANG Z,et al.Instance segmentation of apple flowers using the improved mask R-CNN model[J].Biosystems Engineering, 2020,193:264-278.
- [5] HAN X,CHANG J,WANG K.Real-time object detection based on YOLO-v2 for tiny vehicle object[J].Procedia Computer Science,2021, 183:61-72.
- [6] ZHAO L Q,LI S Y.Object detection algorithm based on improved YOLOv3[J].Electronics,2020,9(3):537.
- [7] YU J M,ZHANG W.Face mask wearing detection algorithm based on improved YOLO-v4[J].Sensors,2021,21(9):3263.
- [8] CAO J,SONG C,SONG S,et al.Front vehicle detection algorithm for smart car based on improved SSD model[J].Sensors,2020,20(16):4646.
- [9] TAN Y D,YU D,HU Y.An application of an improved FCOS algorithm in detection and recognition of industrial instruments[J].Procedia Computer Science,2021,183:237-244.
- [10] WU H,ZHAO J.An intelligent vision-based approach for helmet identification for work safety[J].Computers in Industry,2018,100:267-277.
- [11] SILVA R R,AIRES K R T,VERAS R de M S.Detection of helmets on motorcyclists[J].Multimedia Tools and Applications,2018,77(5): 5659-5683.
- [12] SHINE L,JIJI C V.Automated detection of helmet on motorcyclists from traffic surveillance videos:a comparative analysis using hand-crafted features and CNN[J].Multimedia Tools and Applications,2020,79(19):14179-14199.
- [13] ESPINOSA-OVIEDO J E,VELASTÍN S A,BRANCH-BEDOYA J W.EspiNet V2:a region based deep learning model for detecting motorcycles in urban scenarios[J].Dyna,2019,86(211):317-326.
- [14] BARBOSA R C,AYUB M S,ROSA R L,et al.Lightweight PVIDNet:a priority vehicles detection network model based on deep learning for intelligent traffic lights[J].Sensors,2020,20(21):6218.
- [15] NAGORIYA H.Live helmet detection system for detecting bikers without helmet[J].International Journal of Knowledge Based Computer System,2019,7(2):14-17.
- [16] 郭师虹,井锦瑞,张潇丹,等.基于改进的YOLOv4安全帽佩戴检测研究[J].中国安全生产科学技术,2021,17(12):135-141.
- [17] 蒋润熙,阿里甫·库尔班,耿丽婷.面向轻量化网络的安全帽检测算法[J].计算机工程与应用,2021,57(20):263-270.
- [18] 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.
- [19] 周薇娜,孙丽华,徐志京.复杂环境下多尺度行人实时检测方法[J].电子与信息学报,2021,43(7):2063-2070.
- [20] FU H X,SONG G Q,WANG Y C.Improved YOLOv4 marine target detection combined with CBAM[J].Symmetry,2021,13(4):623.
- [21] ABD E M,DAHOU A,ALSALAH N A,et al.Boosting COVID-19 image classification using MobileNetV3 and aquila optimizer algorithm [J].Entropy,2021,23(11):1383.
- [22] PATEL S.Diabetic retinopathy detection and classification using pre-trained convolutional neural networks[J].International Journal on Emerging

Technologies,2020,11(3):1082-1087.

- [23] LIU J,WANG X.Early recognition of tomato gray leaf spot disease based on MobileNetV2-YOLOv3 model[J].Plant Methods,2020,16(1):1-16.
- [24] TAN M,CHEN B,PANG R,et al.Mnasnet:platform-aware neural architecture search for mobile[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,Los Angeles,USA,June 15-20,2019:2820-2828.
- [25] HU J,SHEN L,SUN G.Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake,USA,June 18-21,2018:7132-7141.
- [26] BOJER C,MELDGAARD J.Kaggle forecasting competitions:an overlooked learning opportunity[J].International Journal of Forecasting, 2021,37(2):587-603.

引用格式:

中文:张鑫,周顺勇.改进YOLOv5s的摩托车头盔佩戴检测算法[J].四川轻化工大学学报(自然科学版),2023,36(3):50-58.

英文:ZHANG X,ZHOU S Y.Improved motorcycle helmet wearing detection algorithm of YOLOv5s[J].Journal of Sichuan University of Science & Engineering (Natural Science Edition),2023,36(3):50-58.

Improved Motorcycle Helmet Wearing Detection Algorithm of YOLOv5s

ZHANG Xin^{1,2}, ZHOU Shunrong^{1,2}

(1. School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China; 2. Artificial Intelligence Key Laboratory of Sichuan Province, Yibin 644000, China)

Abstract: Aiming at the low accuracy and slow detection rate of motorcycle helmet wearing detection, an improved motorcycle helmet detection algorithm based on YOLOv5s has been proposed. Firstly, a shallow detection scale and a 4-fold upsampling feature fusion structure are added in the multi-scale feature detection of YOLOv5s, to improve the detection accuracy. Secondly, the Convolutional Block Attention Module (CBAM) is introduced to improve the focus on the aggregation target, which effectively solves the problem of missed detection and false detection caused by occlusion and overlap. Finally, the block structure of MobilenetV3 is used to replace the bottleneck structure in the backbone network and neck, achieving the purpose of reducing the amount of network parameters. The experimental results show that, compared with the YOLOv5s algorithm, the *mAP* of the improved algorithm is increased by 2.91%, and the detection rate reaches 36 frame/s. The detection accuracy is higher with ensuring a higher detection rate, which exhibits certain application value.

Key words: helmet detection; YOLOv5s model; CBAM attention mechanism; MobilenetV3 network