

DOI:10.19651/j.cnki.emt.2211253

基于注意力机制和跨尺度特征融合的 摩托车头盔检测算法*

张鑫^{1,2} 周顺勇^{1,2} 李思诚^{1,2} 曾雅兰^{1,2}

(1, 四川轻化工大学自动化与信息工程学院 宜宾 644000; 2, 四川轻化工大学人工智能四川省重点实验室 宜宾 644000)

摘要: 在道路交通摩托车事故中,未佩戴头盔是导致骑乘人员受到致命伤害的主要原因。针对目前头盔检测中由于黑色头发、帽子和头盔的颜色和形状相似存在误检和漏检问题,提出了一种具有三重注意力机制和双向跨尺度特征融合的摩托车头盔检测算法。首先,在YOLOV5s的主干网络中引入三重注意力机制,提取了不同维度之间的语义依赖,消除了通道和权重的间接对应关系,通过关注相似样本的差异从而提升检测精度。其次,采用EIOU边框损失函数优化对遮挡和重叠目标的检测效果。最后,在特征金字塔中采用加权双向特征金字塔网络结构,实现高效的双向跨尺度连接和加权特征融合,增强了网络特征提取能力。实验结果表明,改进算法实现了98.7%的mAP@0.5、94.0%的mAP@0.5:0.95,与原算法相比,改进算法的mAP@0.5提升了3.9%以及mAP@0.5:0.95提升了7.6%,具有更高的精度和更强的泛化能力。

关键词: 头盔检测;YOLOV5s;三重注意力机制;EIOU;加权双向特征金字塔

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.2

Motorcycle helmet detection algorithm based on attention mechanism and cross-scale feature fusion

Zhang Xin^{1,2} Zhou Shunyong^{1,2} Li Sicheng^{1,2} Zeng Yalan^{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, Sichuan University of Science & Engineering, Yibin 644000, China)

Abstract: In road traffic motorcycle accidents, failure to wear a helmet is the leading cause of fatal injuries to riders. Aiming at the problems of false detection and missed detection in the current helmet detection due to the similarity in color and shape of black hair, hat and helmet, a motorcycle helmet detection algorithm with triplet attention mechanism and bidirectional cross-scale feature fusion is proposed. First, a triplet attention mechanism is introduced into the backbone network of YOLOV5s, which extracts semantic dependencies between different dimensions, eliminates the indirect correspondence between channels and weights, and improves detection accuracy by paying attention to the differences between similar samples. Second, the EIOU bounding loss function is used to optimize the detection effect of occluded and overlapping objects. Finally, the weighted bidirectional feature pyramid network structure is adopted in the feature pyramid to achieve efficient bidirectional cross-scale connection and weighted feature fusion, which enhances the network feature extraction capability. The experimental results show that the improved algorithm achieves 98.7% mAP@0.5 and 94.0% mAP@0.5:0.95. Compared with the original algorithm, the improved algorithm's mAP@0.5 increases by 3.9% and mAP@0.5:0.95 increases by 7.6%, with higher accuracy and stronger generalization ability.

Keywords: helmet detection;YOLOV5s;triplet attention mechanism;EIOU;weighted bidirectional feature pyramid

0 引言

根据世界卫生组织(World Health Organization,

WHO)发布的全球道路安全状况报告显示^[1],每年在道路交通事故中死亡人数达到了135万人,其中低收入国家的死亡率是高收入国家的3倍。在全球范围内,摩托车事故

收稿日期:2022-09-02

*基金项目:国家自然科学基金(61801319)、四川省科技厅项目(2020YFG0178)、四川省科技厅省院校合作项目(2020YFSY0027)资助

导致死亡的人数占道路交通总死亡人数的28%。尤其是在东南亚和西太平洋地区,摩托车事故的死亡率达到43%。泰国是世界道路交通死亡率最高的10个国家之一,其中与摩托车相关的死亡率世界最高,每天约60人死亡。引起摩托车事故导致死亡的主要原因是骑乘人员的头部受伤,正确规范的佩戴摩托车头盔可以使头部受伤的风险降低69%以及受到致命伤害的危险降低42%。报告表明规范佩戴头盔是防止道路交通死亡的重要方法,头盔佩戴的检测显得格外重要,但是传统的头盔检测手段费时费力,而基于深度学习的检测方法更加智能化。

目前基于深度学习的目标检测算法通常分两类。Two-stage算法有Fast R-CNN^[21]、Faster R-CNN^[31]、Mask R-CNN^[41]等,而One-stage算法为代表的是YOLO^[5-8]系列。Silva等^[9]提出了一种基于局部二值模式、定向梯度直方图和霍夫变换描述子的混合描述子,进行摩托车头盔的检测和分类,但是该算法只能应用于光线明亮、分辨率高的图像,鲁棒性差。Silva等^[10]使用圆形霍夫变换和定向梯度直方图描述来提取图像属性,通过多层感知机分类来完成检测,但是当摩托车上不止一个人时,该算法只能识别出一个人的佩戴情况,对于多人检测会失效。Vishnu等^[11]对视频帧使用自适应背景减法来获取移动对象,然后利用卷积神经网络(convolutional neural network, CNN)来识别摩托车骑乘人员,但是在人员密集的情况下算法精度不高。Kumar等^[12]提出一种基于计算机视觉的模型,该模型利用方向梯度直方图(histogram of oriented gradient, HOG)描述符和基于特征提取模块(receptive field block, RFB)的支持向量机(support vector machine, SVM)分类,第1步确定骑乘人员的驾驶工具是否为摩托车,第2步识别摩托车的骑乘人员是否佩戴头盔,但是对形状与头盔相似的帽子识别率较低。Silva等^[13]提出摩托车检测和头盔检测结合的检测系统,利用小波变换(wavelet transform, WT)作为描述符,随机森林作为分类器检测摩托车,利用圆霍夫变换(circle Hough transform, CHT)和定向梯度直方图描述符提取图像属性,最后用多层感知机(multilayer perceptron, MLP)分类器进行识别,对头盔的检测精度达到了91.37%,但是对重叠的骑手和乘客检测效果差。Dasgupta等^[14]提出了一个框架,第1阶段使用YOLOv3模型检测摩托车骑乘人员,第2阶段使用CNN对一阶段的人员进行头盔检测,但是该框架对于头部颜色的识别不敏感,很容易错误识别。Yogameena等^[15]提出使用高斯混合模型(Gaussian mixture model, GMM)对前景对象进行分割,确保摩托车的存在,然后采用Faster R-CNN来检测骑乘人员是否佩戴头盔,虽然算法在低分辨率、不同头发颜色和部分遮挡的骑乘人员的环境中表现出色,但是并未解决骑乘人员的头盔与黑色头巾以及帽子相似等问题。

冉险生等^[16]提出的MNxT-ECA-D-YOLOv2算法引入

密集连接结构和通道注意力机制,在输出卷积层增加Drop Block模块,使在原算法的基础上提高了检测精度并极大的缩减了推理时间。薛瑞晨等^[17]提出了一种基于改进YOLOv3的头盔检测算法,通过引入通道和空间注意力模块和集合密集连接网络提高特征提取效果,改进算法检测精度远高于原算法,达到了93.29%。Ubaid等^[18]所提出的EfficientDet方法来检测摩托车头盔且精度达到了95.23%,刘琛等^[19]提出了一种改进一阶多层模型(single shot multibox detector, SSD^[20])的检测算法,引入了类似人类视觉的RFB模块,降低了模型的误检率和漏检率,鲁棒性更强。Goyal等^[21]针对复杂环境和遮挡场景等问题,引入了Amodal回归器,能够有效处理拥挤场景中摩托车之间的遮挡和覆盖,这些研究成果在一定程度上能降低头盔检测中由于黑色头发、帽子和头盔颜色和形状相似导致的误检和漏检率。

本文算法针对头盔检测中由于黑色头发、帽子和头盔颜色和形状相似导致误检和漏检问题,在YOLOV5s中引入了三重注意力^[22]机制来聚焦头部区域的差异。针对摩托车骑乘人员的遮挡和重叠问题,采用高效交并比(efficient intersection over union, EIOU^[23])边界框损失函数优化筛选预测框,极大提高了摩托车头盔佩戴检测的精度。

1 YOLOV5s 算法原理

YOLOv5是一种经典的单阶段目标检测算法,它使用了depth_multiple和width_multiple来调整网络的深度和宽度,将其按照一定的数值比例能够得到不同的结构模型,其中YOLOV5s是最受欢迎的模型,它的模型参数量小检测精度也高。该模型由3个部分组成:主干特征网络、特征金字塔以及预测输出。主干特征网络负责提取输入图像特征,利用Focus操作将原始特征图每隔一个像素值取一个值,形成了4个特征图,其操作方法类似于邻近值采样,把depth和width的信息集中到了通道中。主干中最基本的结构的是C3模块,池化层中使用快速空间金字塔池化(spatial Pyramid pooling-fast, SPPF)来扩大感受野,在Neck中组成是特征金字塔网络(feature Pyramid networks, FPN)加上路径聚合网络(path aggregation network, PANet)的结构,FPN自顶向下传递强语义信息,而PANet自底向上传递准确的定位信息,它们共同构成特征金字塔,融合不同特征层之间的特征信息。输出预测部分由3个预测层构成,预测信息包括了边框位置信息、检测类别以及对应概率。

2 改进后的 YOLOV5s 算法

2.1 三重注意力机制

注意力模型在最近的几年被深度学习的很多领域广泛使用,它借鉴了人类独有的大脑信号处理机制。人类通过

扫描图像找到需要重点关注的目标区域,也可以称作注意力的焦点。

在摩托车头盔检测过程中,因为骑乘人员的黑色头发、帽子和头盔的颜色和形状相似存在误检和漏检问题,所以需要提高对头部区域的关注度,聚焦的范围也应该包括头部。这里采用在主干中引入三重注意力机制来加强对头部相似样本的聚焦。首先,给定输入张量,三重注意力机制通过选择输入张量然后进行残差变换来捕获维度间的依赖关系。注意力机制如挤压激励网络(squeeze-and-excitation networks, SENet)和卷积块注意力模块(convolutional block attention module, CBAM)需要大量额外的可学习参数,SENet 通过简单的将输入压缩到单通道获得空间注意力,CBAM 虽然有显著的性能提升,但是没有跨纬度的交互,而且在计算信道注意力时使得维度降低,对于非线性局部相关性没有帮助。

三重注意力强调了跨纬度交互的重要性,因为没有让维度降低,所以消除了通道和权重两者之间的间接映射关系,模块如图 1 所示。三重注意力机制有 3 个分支,第 1 个分支负责跨通道维度 C 和空间维度 W 计算注意力权重,第 2 个分支负责跨通道维度 C 和空间维度 H 计算注意力权重,第 3 个分支负责用于计算空间维度 W 和 H 的相关性。前两个分支通过围绕空间维度 W 建立通道维度 C 和空间维度 H 的关系,围绕空间维度 H 建立通道维度 C 和空间维度 W 的关系,最后使用对权重进行平均加权。在三重注意力机制的分支计算中, $Z-Pool$ 将 $C \times H \times W$ 维度的张量通过最大池化和平均池化形成两个 $1 \times H \times W$ 维度的张量,然后将这两个张量拼接,其计算公式为:

$$Z-Pool(X) = [MaxPool_{od}(X), AvgPool_{od}(X)] \quad (1)$$

其中, $MaxPool_{od}$ 代表对 X 进行最大池化操作, $AvgPool_{od}$ 代表对 X 进行平均池化操作。输入张量 X 通过三重注意力机制模块后得到输出张量 Y , 其计算公式为:

$$Y = \frac{1}{3}((\widehat{X}_1 \sigma(\psi_1(\widehat{X}_1^*)) + \widehat{X}_2 \sigma(\psi_2(\widehat{X}_2^*)) + X \sigma(\psi_3(\widehat{X}_3))) \quad (2)$$

其中,输入是维度为 $C \times H \times W$ 的 X , \widehat{X}_1 表示将 X 沿 H 轴逆时针旋转 90° ,其维度变为 $W \times H \times C$; \widehat{X}_2 表示将 X 沿 W 轴逆时针旋转 90° ,其维度变为 $H \times C \times W$, \widehat{X}_3 表示对 X 进行 $Z-Pool$ 操作, \widehat{X}_1^* 表示将 \widehat{X}_1 进行 $Z-Pool$ 操作, \widehat{X}_2^* 表示将 \widehat{X}_2 进行 $Z-Pool$ 操作, σ 表示 sigmoid 激活, ψ_1, ψ_2, ψ_3 分别卷积核大小为 $k \times k$ 的卷积操作。上划线表示将该张量按照对应的轴顺时针旋转 90° 。

2.2 具有双向跨尺度特征融合的特征金字塔

YOLOV5s 的 FPN 网络结构通过多次卷积下采样得到不同大小的特征图,不断获得更高级语义信息,然后不断上采样,将上下采样过程中相同尺度的特征图进行叠加。PANet 网络如图 2(a)所示,其结构在 FPN 的基础上新增

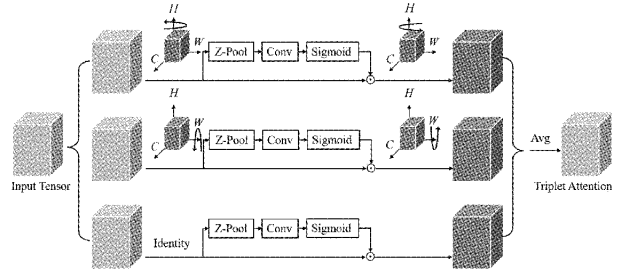


图 1 三重注意力机制模块

了二次下采样,该过程也将同尺度的特征图进行叠加,加强了定位信息。然而, PANet 只是一个平等的简单的双向融合结构,它对输出特征的信息传递是平衡的。由于不同的输入特征具有不同的尺度大小,它们通过卷积后对输出特征图的信息传递是不相等的。为了更好的进行特征融合,选择删除了图 2(a)中的 A' 和 C' 节点,达到了对不同特征层的信息传递不平衡的目的。对于同尺度的输入 B 节点和输出 P_2 节点,增加了一条额外的传递路线,只花费了少量的计算成本的代价就融入了更多的信息。

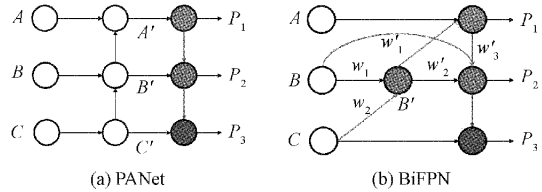


图 2 PANet 和 BiFPN 网络结构

双向特征金字塔网络 (bidirectional feature Pyramid network, BiFPN^[24]) 采用了双向交叉尺度连接和快速归一化融合,其网络如图 2(b)所示,图 2(b)中与 B 节点同尺度的 B' 节点和 P_2 节点的特征融合公式分别为:

$$B' = Conv\left(\frac{w_1 B + w_2 Resize(C)}{w_1 + w_2 + \epsilon}\right) \quad (3)$$

$$P_2 = Conv\left(\frac{w'_1 B + w'_2 B' + w'_3 Resize(P_1)}{w'_1 + w'_2 + w'_3 + \epsilon}\right) \quad (4)$$

其中, B' 是上采样过程中与 B 节点同尺度的中间特征, P_2 是下采样过程中与 B 节点同尺度的输出特征, w_1 是 B 节点和 B' 节点之间的权重, w_2 是 C 节点和 B' 节点之间的权重, w'_1 是 B' 节点和 P_2 节点之间的权重, w'_3 是 P_1 节点和 P_2 节点之间的权重, ϵ 是一个非常小的值,其大小为 0.0001 , $Resize$ 是用于尺度匹配的上采样或者下采样运算, $Conv$ 是用于特征图处理的卷积运算。

2.3 EIIOU 边框损失函数

在目标检测过程中,边界框的回归是极其关键的步骤,它能够决定预测目标的定位性能的好坏。但是目前的边界框损失函数都有一些缺点,例如在交并比(intersection over union, IOU)损失函数中,如果预测框和真实框之间没有任何交集,那么 IOU 损失始终为 0,这不能正确反映预测框和真实框之间的真实距离差距,而且其收敛速度极慢。在广

义交并比(generalized intersection over union, GIOU)损失函数中,当预测框和真实框重合或者存在包含关系时候, GIOU 等同于 IOU,其收敛速度依旧很慢。完全交并比(complete intersection over union, CIOU)在距离交并比(distance intersection over union, DIOU)的基础上加入了纵横比部分,虽然 CIOU 损失依赖于边界框的重叠面积、中心点距离和纵横比,但公式不能反映预测框的宽高和真实框之间的差距,有时反而会影响其正确优化。所以在 CIOU 的基础上,将纵横比部分拆分,提出了描述宽高的 EIOU 边界框损失函数。这种算法能够有效的加强对摩托车骑乘人员的重叠以及遮挡情况的检测,其计算公式如下:

$$L_{EIOU} = L_{IOU} + L_{Dis} + L_{Asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{(\omega^c)^2 + (h^c)^2} +$$

$$\frac{\rho^2(\omega, \omega^{gt})}{(\omega^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \quad (5)$$

其中, L_{IOU} 表示 IOU 损失, L_{Dis} 表示距离损失, L_{Asp} 表示纵横比损失, b 表示预测框的中心坐标, b^{gt} 表示真实框的中心坐标, $\rho(b, b^{gt})$ 表示 b 和 b^{gt} 之间的欧氏距离, ω^c 和 h^c 分别表示预测框和真实框最小外接矩形的宽度和高度, ω 表示预测框的宽度, ω^{gt} 表示真实框的宽度, $\rho(\omega, \omega^{gt})$ 表示 ω 和 ω^{gt} 宽度差值, h 表示预测框的高度, h^{gt} 表示真实框的高度, $\rho(h, h^{gt})$ 表示 h 和 h^{gt} 高度差值。EIOU 损失函数优化了预测框和真实框之间的高度和宽度差异最小化,使得收敛速度更快和定位更准确,有助于解决骑乘人员的互相遮挡以及头部的重叠问题。总体改进算法如图 3 所示。

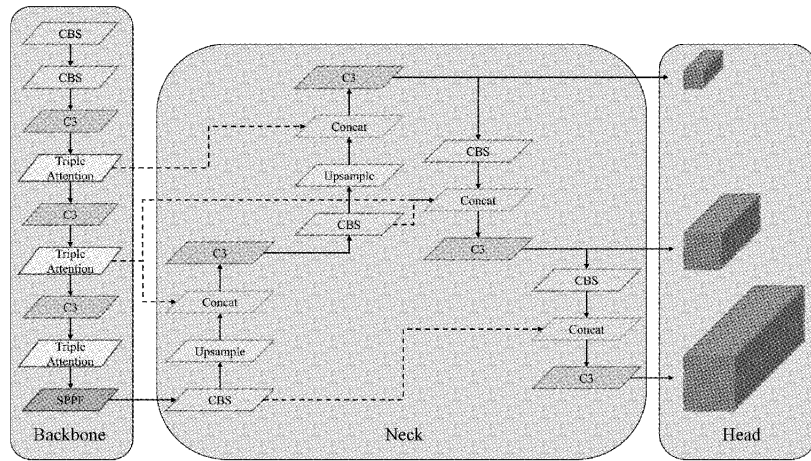


图3 具有注意力机制和加权双向特征金字塔的 YOLOV5s

3 实验结果与分析

3.1 实验环境

实验使用 NVIDIA GeForce RTX 3080, 显示内存 10.5 GB。使用深度学习框架是 Pytorch1.10.0 以及 GPU 计算框架 CUDA11.2。实验参数设置如下:学习率设置为 0.01,训练轮数为 200 轮, Batch Size 为 16, 动量为 0.937, 权值衰减为 0.005, 边界框 box 损失系数 0.05, 分类 cls 损失系数 0.5, 置信度 obj 损失系数 1.0, hsv_h 为 0.015, hsv_v 为 0.7, hsv_v 为 0.4, 并开启 mosaic 数据增强。

3.2 数据集

目前有许多学者提出的数据集大都比较简单,缺少骑乘人员重叠或者遮挡的情形,另外包括的类别少,缺乏对头部相似样本的补充,为了更加贴切真实的场景,实验使用了来自 Kaggle 以及数据集^[25]上印度新德里的道路拍摄图像,为了能准确识别出与头盔颜色和形状相似的黑色头发、帽子以及头巾,通过筛选出具有此类特征的图像,然后对图片进行手工标注,使用的标注工具是 labelimg,文件格式为 YOLO 格式。通过筛选删除掉了一些不合理的图片,最终保留了 14 160 张图片, P0 表示骑手, P1 表示乘员,按

照图 4 进行编码,形成了 5 个类别。

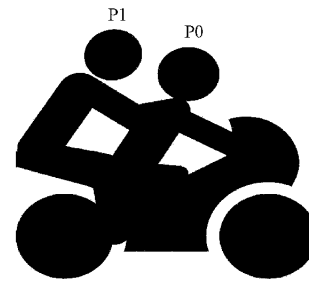


图4 骑乘人员座位情况

其按照 7:3 的比例分配为训练集和测试集,其中各个类别的标注数量如图 5 所示。

在图 5 中,类别 P0HelmetMan 表示骑手戴头盔、P0HelmetManP1HelmetMan 表示骑手和乘客都戴头盔、P0NoHelmetMan 表示骑手未戴头盔、P0HelmetManP1NoHelmetMan 表示骑手戴头盔乘客未戴头盔、P0NoHelmetManP1NoHelmetMan 表示骑手和乘客都未戴头盔。

3.3 模型评估指标

为了评估模型的检测效果,使用 Precision、Recall、

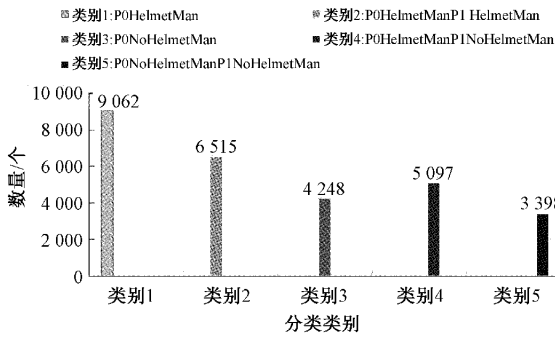


图 5 不同分类类别的数量分布柱状图

F1、mAP 来对算法进行性能评估。其计算公式如下：

$$Precision = \frac{True_positive}{True_positive + False_positive} \quad (6)$$

$$Recall = \frac{True_positive}{True_positive + False_negative} \quad (7)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

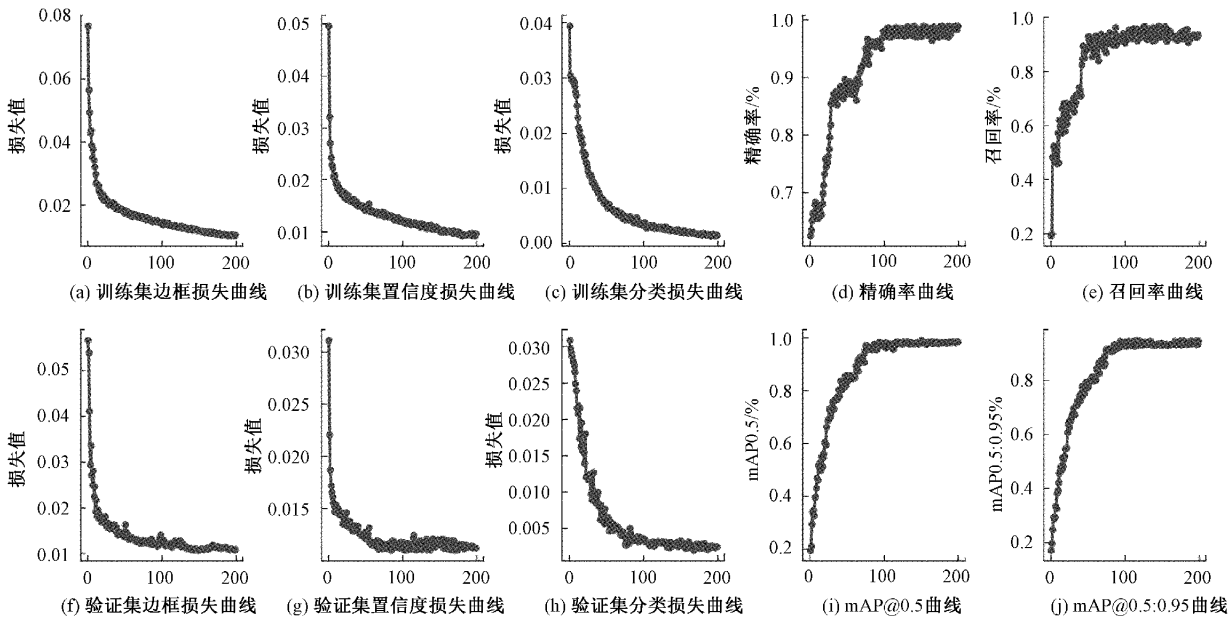


图 6 训练和验证过程中性能指标随训练轮次的变化曲线

2 排最后两个仿真结果表示 $mAP@0.5$ 和 $mAP@0.5:0.95$ 随训练轮次的变化曲线图。

图 6 中,横坐标代表训练轮次,纵坐标表示其对应的值。模型在接近 100 个 epoch 之后,该模型在 $Precision$ 、 $Recall$ 、 $mAP@0.5$ 和 $mAP@0.5:0.95$ 方面都达到了一定的稳定,并且在前 100 个 epoch 中, box_loss 、 obj_loss 、 cls_loss 急剧下降,然后呈现平稳趋势。因此,选择第 150 个 epoch 后获得的最佳权重作为摩托车头盔佩戴检测的权重。

3.5 消融实验

为了验证改进模块对摩托车头盔佩戴检测的影响,对

$$mAP = \frac{\sum_{i=1}^3 Ap(i)}{5} \quad (9)$$

其中, $True_positive$ 被正确检测为正样本, $False_positive$ 表示被错误检测为正样本, $False_negative$ 表示被错误检测为负样本。 Ap 是 $Precision$ 和 $Recall$ 包围的面积, mAP 是指每个类别的 Ap 值相加再除以所有类别个数,即所有类别 Ap 的平均值。

3.4 训练结果

对模型进行 200 个 epoch 的训练,得到了改进模型的训练结果,包括了训练集(Train Set)和验证集(Val Set)的边界框损失(box_loss)、物体置信度损失(obj_loss)和分类损失(cls_loss),还有各种评价指标例如准确率 $Precision$ 、召回率 $Recall$ 、 $mAP@0.5$ 和 $mAP@0.5:0.95$,通过训练后可以得到可视化仿真的结果如图 6 所示,第 1 排和第 2 排的前 3 个仿真结果分别显示了训练过程和验证过程的损失随训练轮次的变化曲线图,然后第 1 排最后两个仿真结果表示准确率和召回率随训练轮次的变化曲线,接着第

各个模块进行性能评估,结果如表 1 所示。其中 Base 表示基础的算法模型 YOLOV5s, A 表示引入了三重注意力机制, B 表示使用了 BiFPN, C 表示使用了 EIOU 边框损失函数。

从表 1 可以看出, A 模块三重注意力机制对摩托车头盔检测 mAP 改善较为明显,使得模型 Base 的 $mAP@0.5$ 提升了 1.7%, B 模块 BiFPN 使得模型 Base + A 的 $mAP@0.5$ 提升了 0.7%, C 模块 EIOU 边框损失函数使得模型 Base + A 的 $mAP@0.5$ 提升了 1.1%, 最终模块 A + B + C 使得模型 Base 的 $mAP@0.5$ 提升了 3.9%。 Base、Base + A、Base + A + B、Base + A + C、Base + A + B + C 的

表1 各模块对模型的性能评估

模型/评价指标	准确率/%	召回率/%	F1/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%
Base	92.7	90.9	91.8	94.8	86.4
Base+A	93.6	93.7	93.6	96.5	89.1
Base+A+B	94.3	94.6	94.4	97.2	90.0
Base+A+C	95.3	94.9	95.1	97.6	91.6
Base+A+B+C	98.9	94.2	96.5	98.7	94.0

准确率-召回率(Precision-Recall)曲线依次如图7的(a)、(b)、(c)、(d)、(e)所示,每一个模型都包括了5条关于类

别的 Precision-Recall 曲线和 1 条总的 Precision-Recall 曲线。

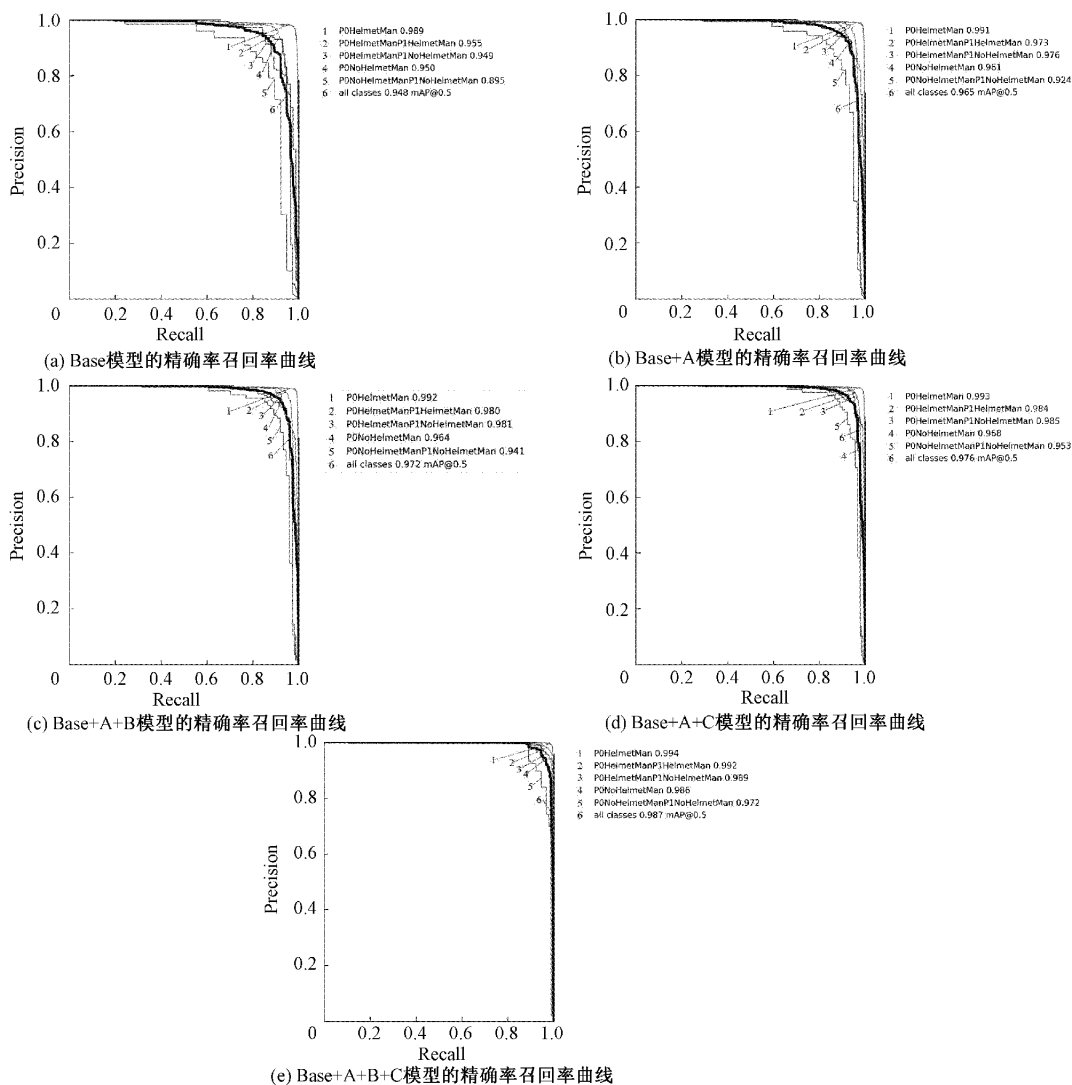


图7 不同消融模型的 Precision-Recall 曲线

3.6 改进算法与其他算法的对比

横向对比实验的设置能够体现改进模型的优势,将本文算法与 Faster RCNN、SSD、YOLOv3、YOLOv4、YOLOV5s 算法进行对比分析,所有的环境和参数变量设置都保持一致,使用相同的数据集进行训练,得出各个算法模型在 Precision、Recall、F1、 $mAP@0.5$ 、

$mAP@0.5:0.95$ 以及推理速度方面的对比结果,如表2所示。

根据表2数据对比可知,模型 YOLOV5s 比 Faster RCNN、SSD、YOLOv3 和 YOLOv4 具有更高的 $mAP@0.5$, 其值分别提升了 9.6%、12.4%、3.2% 和 2.6%。本文算法与 YOLOV5s 相比, $mAP@0.5$ 提升了 3.9%, 达到了 98.7%。

表 2 不同算法性能评估对比

模型/评价指标	准确率/%	召回率/%	F1/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%	推理时间/s
Faster R-CNN	86.2	81.4	83.7	85.2	80.4	0.256
SSD	84.4	80.8	82.6	82.4	78.9	0.132
YOLOv3	90.8	88.6	89.7	91.6	83.2	0.073
YOLOv4	91.8	89.3	90.5	92.2	84.9	0.055
YOLOV5s	92.7	90.9	91.8	94.8	86.4	0.031
本文算法	98.9	94.2	96.5	98.7	94.0	0.034

本文算法与 Faster RCNN、SSD、YOLOv3、YOLOv4 算法相比, $mAP@0.5$ 分别提升了 13.5%、16.3%、7.1% 和 6.5%, $mAP@0.5:0.95$ 分别提升了 13.6%、15.1%、10.8% 和 9.1%。通过实验数据表明, 本文算法在推理时间上与 YOLOV5s 基本无差别, 从评价指标的表现看, 优

于列表中其他目标检测算法。

3.7 可视化结果分析

为了更直观的观察改进算法对摩托车头盔的检测效果, 选择了测试集中 12 张比较具有代表性的图片进行可视化分析, 其结果如图 8 所示。



图 8 改进算法的可视化检测结果

通过对图 8 的结果分析, 本文算法能够准确识别出与头盔颜色和形状相似的黑色头发、帽子以及头巾, 也能够准确的识别出存在遮挡和覆盖的骑乘人员, 具有极好的检测效果和鲁棒性。

为了进一步观察三重注意力机制对于改进模型的作用, 选择测试集中 15 个检测结果进行 Grad-Cam 可视化分析, 模型 YOLOV5s、改进算法的可视化结果分别如图 9、10 所示。

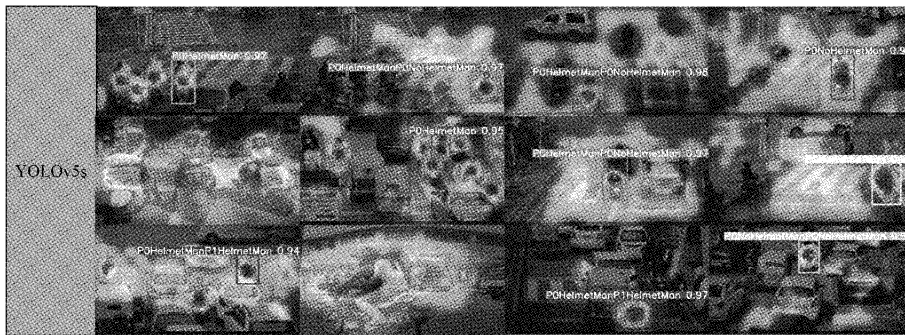


图 9 YOLOV5s 热力图可视化结果

通过对图 9 与图 10 的对比分析, 可以看出 YOLOV5s 算法重点提取的特征主要集中在骑乘人员腰部与车身接壤的位置, 没有关注需要重点区分的头部相似信息。而改进

算法将关注的焦点扩大并包括了骑乘人员的头部, 通过关注头部区域的重点信息对检测结果起了至关重要的作用, 所以改进算法具有更好的鲁棒性和更出色的检测效果。



图10 改进算法热力图可视化结果

4 结 论

本文提出了一种基于注意力机制和双向跨尺度特征融合的摩托车头盔检测算法。通过引入注意力机制模块,使得模型能够重点关注头部区域形状和颜色的关键信息。其次,使用了BiFPN,实现了简单快速的多尺度特征融合。最后,使用了EIOU边界框损失函数,优化了模型的收敛速度以及加强了骑乘人员遮挡和覆盖情况的检测。实验结果表明,改进算法在摩托车头盔佩戴中的 $mAP@0.5$ 和 $mAP@0.5:0.95$ 均高于YOLOV5s算法,且优于其他主流算法,对黑色头发、帽子和头盔的颜色和形状相似存的检测具有较强的泛化能力,且具有极好的实际应用价值。

本文算法在摩托车有一个人或两个人的情况下对头盔的检测精度非常高,但是对于超过两人的情况,检测效果还需要进一步加强。在实际应用场景中,为了达到实时的检测效果,该算法在Jetson nano嵌入式设备部署后的帧率只能达到3 fps,后续工作考虑进行模型剪枝,但同时也尽可能保持精度,通过减少参数量来达到实时检测的目的。

参考文献

- [1] ROSEN H, BARI I, PAICHADEZ N, et al. Global road safety 2010-18: An analysis of global status reports [J]. Injury, 2022, DOI: 10.1016/j.injury.2022.07.030.
- [2] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448.
- [3] REN S, HE K, GIRSHICK R. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [4] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(2): 386-397.
- [5] 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(CVPR), 2016:779-788.

- [6] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.
- [7] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv Preprint, 2018, ArXiv:1804.02767.
- [8] BOCHKOVSKIY A, WANG C, LIAO H. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint, 2020, ArXiv:2004.10934.
- [9] SILVA R, AIRES K, SANTOS T, et al. Automatic detection of motorcyclists without helmet [C]. 2013 XXXIX Latin American Computing Conference (CLEI), 2013, DOI: 10.1109/CLEI.2013.6670613.
- [10] SILVA R, AIRES K, VERAS R. Helmet detection on motorcyclists using image descriptors and classifiers[C]. 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, 2014: 141-148.
- [11] VISHNU C, SINGH D, MOHAN C, et al. Detection of motorcyclists without helmet in videos using convolutional neural network[C]. 2017 International Joint Conference on Neural Networks(IJCNN), 2017: 3036-3041.
- [12] KUMAR N, SUKAVANAM N. Detecting helmet of bike riders in outdoor video sequences for road traffic accidental avoidance[C]. International Conference on Intelligent Systems Design and Applications. Springer, 2018: 24-33.
- [13] SILVA R, AIRES K, VERAS R. Detection of helmets on motorcyclists[J]. Multimedia Tools and Applications, 2018, 77(5): 5659-5683.
- [14] DASGUPTA M, BANDYOPADHYAY O, CHATTERJI S. Automated helmet detection for multiple motorcycle riders using CNN [C]. IEEE

- Conference on Information and Communication Technology, 2019, DOI: 10.1109/CICT48419.2019.9066191.
- [15] YOGAMEENA B, MENAKA K, PERUMAAL S. Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system[J]. IET Intelligent Transport Systems, 2019, 13(7): 1190-1198.
- [16] 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.
- [17] 薛瑞晨,郝媛媛,张振,等.基于改进 YOLOv3 的头盔佩戴检测算法[J].电子测量技术,2021,44(12):115-120.
- [18] UBAID M, KIRAN A, RAJA M, et al. Automatic helmet detection using efficientDet[C]. 2021 International Conference on Innovative Computing(ICIC), 2021, DOI: 10.1109/ICIC53490.2021.9693093.
- [19] 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.
- [20] 刘寒迪,赵德群,陈星辉.基于改进 SSD 的航拍施工车辆检测识别系统设计[J].国外电子测量技术,2020, 39(7):127-132.
- [21] GOYAL A, AGARWAL D, SUBRAMANIAN A, et al. Detecting, tracking and counting motorcycle rider traffic violations on unconstrained roads [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022: 4302-4311.
- [22] MA W, CUI Y, SHAO N, et al. TripleNet: Triple attention network for multi-turn response selection in retrieval-based chatbots [J]. ArXiv Preprint, 2019, ArXiv:1909.10666.
- [23] ZHANG Y, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [24] TAN M, PANG R, LE Q. EfficientDet: Scalable and efficient object detection [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2020: 10778-10787.
- [25] SIEBERT F, LIN H. Detecting motorcycle helmet use with deep learning [J]. Accident Analysis and Prevention, 2020, DOI: 10.1016/j.aap.2019.105319.

作者简介

张鑫,硕士研究生,主要研究方向为图像处理、目标检测。

周顺勇(通信作者),硕士,副教授,主要研究方向为图像处理、信息与信号处理。

曾雅兰,硕士研究生,主要研究方向为目标检测与分割。

李思诚,硕士研究生,主要研究方向为图像处理、目标检测。

E-mail:1251897659@qq.com