基于改进 YOLO v5 的骑行头盔佩戴检测算法研究

白熠宸[1]，贾　蒙[2]，申小萌[2]

（1.哈尔滨工业大学（威海），山东 威海 264200；2.新乡学院，河南 新乡 453000）

摘要：针对电动车骑行场景中头盔佩戴检测存在的小目标识别困难、复杂遮挡及人工监管效率低下等问题，提出一种基于改进 YOLO v5 的骑行头盔佩戴检测算法。通过在 YOLO v5s 基准模型中引入双重注意力机制与优化损失函数，显著提升检测性能：在 Backbone 中嵌入 CA 坐标注意力模块，增强模型对目标位置的敏感度；在 Neck 部分集成 CBAM 注意力机制模块，通过通道-空间双域协同机制抑制背景噪声干扰；采用 DIoU 损失函数替代 GIoU，优化边界框回归轨迹，提升重叠目标的定位精度。实验基于自建数据集（2360 张图像），改进模型的精确率（P）、召回率（R）及平均精度（mAP）分别达到 88.2%、90.4% 和 90.9%，较基准模型提升 3.5%、4.3% 和 4.1%，同时优于 Faster R-CNN 等对照模型。该算法兼顾轻量化与高精度，为复杂交通场景下的头盔佩戴自动化监管提供了高效的解决方案，具有实际应用价值。

关键词：头盔佩戴检测；YOLO v5；CA 坐标注意力；CBAM 注意力机制；DIoU 损失函数

中图分类号：TP391.41　　**文献标志码**：A　　**文章编号**：2096-7772-(2025)03-0014-06

0　引言

在电动车交通事故中，驾乘人员的头部易遭受极为严重的损伤，从而使电动车头盔成为维护驾乘人员头部安全的关键装备。研究表明，佩戴安全头盔能够将道路交通事故导致的死亡率及头部重伤减少 20% 至 45%[1]。为此，我国公安交管部门自 2020 年 4 月起实施了“一盔一带”安全保护措施，通过法律手段纠正电动车驾驶者未佩戴安全头盔的非法行为，旨在促进安全驾驶习惯的形成。然而，当前对于安全头盔佩戴的监管主要依赖于人工执行，这不仅耗时耗力，同时也难以确保监管效率。

得益于深度神经网络架构的突破性进展，计算机视觉技术正在深度融入城市交通安全治理场景。对于电动车驾乘人员头盔佩戴规范性的检测，现阶段主流模型可分为两类：以 Faster R-CNN[2] 为代表的双阶段检测框架（two-stage detection）和以 SSD[3]、YOLO[4-6] 系列为核心的单阶段检测框架（single-stage detection）。常见的双阶段算法虽然精确度较高，但是具有较大的运算量和较慢的运算速度，在移动端难以部署。相比之下，单阶段算法则在保证检测准确率的前提下，具备更好的轻量化性能和更快的检测速度。

在实际交通场景中，传统目标检测模型的性能并不理想，具体问题包括监控视频中的小目标个体难以提取，复杂场景中目标间相互遮挡等。针对以上现实问题，综合考虑模型的体积及运行速度，本研究基于 YOLO v5 模型进行了改进，具体创新点有：在 Backbone 中加入坐标注意力（Coordinate Attention，CA），在 Neck 中加入卷积块注意力模块（Convolutional Block Attention Module，CBAM），使用 DIoU 改进损失函数。

1　YOLO v5 检测算法

YOLO v5 作为单目标检测算法 YOLO 算法的第五代版本，采用梯度化设计理念构建了 YOLO v5n/s/m/l/x 系列模型。该系列通过调整网络宽度（Channel）

收稿日期：2025-02-16
基金项目：国家自然科学基金（61501391）；河南省科技创新人才项目（21HASTIT021）
第一作者简介：白熠宸（2004—　），男，河南新乡人，回族，主要从事人工智能及计算机视觉研究。
通信作者简介：贾蒙（1981—　），男，河南新乡人，教授，博士，主要从事人工智能及智能制造研究。

与深度（Layer）的相应参数，形成性能渐进的模型序列[7]。考虑到嵌入式系统计算资源的约束，YOLO v5s 所具备的轻量化网络架构通过深度可分离卷积与通道剪枝技术，可有效平衡检测精度与实时性需求，为安全头盔检测等边缘计算场景提供了最优解决方案。

YOLO v5s 算法架构采用模块化设计，其工作流程可分解为四个功能层级：输入预处理模块（Input）、特征提取骨干网络（Backbone）、特征融合层（Neck）和目标检测头部（Head）。该模型的层级化架构如图 1 所示。

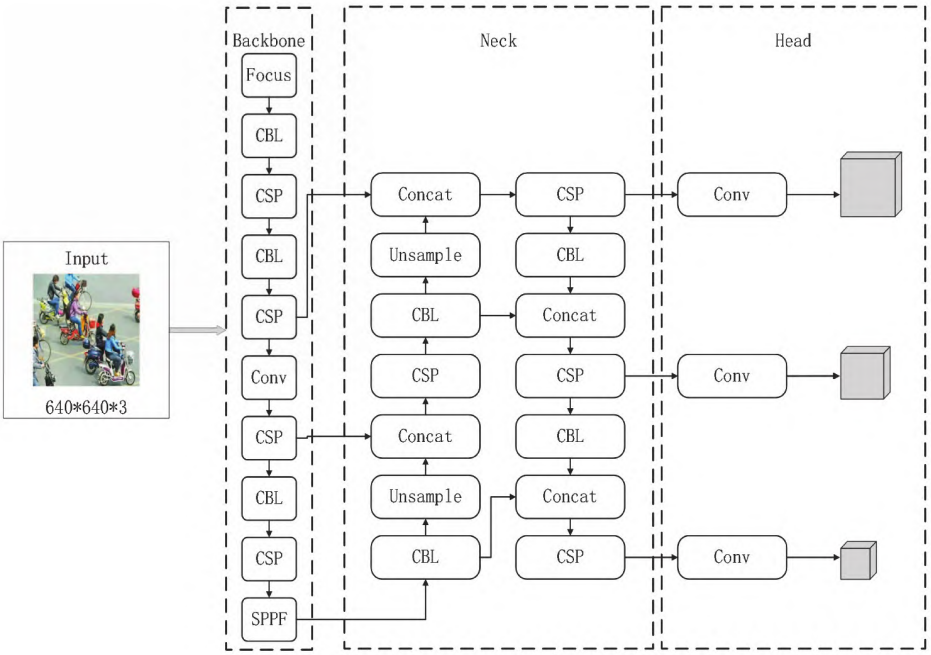

图 1　YOLO v5 层级化架构

在输入图片后，输入预处理模块（Input）进行图像预处理和归一化等操作。主要包括：通过 Mosaic 增加训练数据多样性，自动调整检测框的预设尺寸，自适应将图像缩放至固定尺寸，在保持图片不变形的同时让模型运算更高效。

特征提取骨干网络（Backbone）通过多层次卷积运算，构建图像的高维表征空间。该网络采用跨阶段局部连接机制（CSP）[8]，通过梯度分流策略将输入特征划分为两条异构处理路径：主路径保留完整特征传递，辅助路径实施深度可分离卷积与通道重标定。经过特征融合层对双路径输出实施通道维度拼接（Concat）后，有效缓解了传统架构中的梯度冗余问题，并通过跨层特征互补显著提升了多尺度特征的判别性表达能力。

特征融合层（Neck）负责整合来自不同抽象层次、分辨率尺度及语义级别的特征表示。该部分通过细粒度与粗粒度特征的交互，捕捉数据中的细微差异，促进多层级信息的有效融合，以增强模型对复杂情况的解析能力，优化分类任务的表现。整合后，模型的泛化能力和鲁棒性显著提升，减少了过拟合的风险。

目标检测头部（Head）负责处理 Backbone 的输出，预测目标信息。其通过层级化预测结构对 Backbone 输出的多维特征图进行解码，生成包含空间定位、置信评估和类别判别的复合检测结果，构建完整的物体特征识别体系。

2　YOLO v5 检测算法优化

本文对 YOLO v5s 基准模型进行如下优化：针对监控视频中小目标个体难以提取，复杂场景中目标间相互遮挡的问题，添加多层注意力机制，分别在 Backbone 中引入 CA[9]，在 Neck 中添加 CBAM[10]。针对模型在处理重叠目标和边界框回归时的效果不理想问题，使用 DIoU 损失函数[11]来替换 YOLO v5s 基准模型的 GIoU 损失函数。

2.1　引入 CA

注意力机制能够在特征图中进行信息筛选，去除无关紧要的细节，从而提取出关键特征，以提升检测性能。该机制通过学习全局信息，增强对富含信息区域的关注，同时抑制非关键特征的影响。在复杂的实际交通场景下，为了使识别系统更有效地关注电动车和头盔等目标，减少背景噪声的干扰，引入 CA。作为一种轻量级的注意力机制，CA 提高了模型对感

兴趣区域的定位精度，增强了对这些区域的识别能力。具体而言，CA 通过对输入特征图沿水平和垂直方向进行池化操作，生成两个一维特征向量。这两个向量分别捕捉了特征图在各自方向上的分布特性，从而使得模型能够更加精确地定位目标位置，并获得相关属性[12]。将编码后的坐标信息与原始特征图进行融合，通过一系列的卷积和激活操作，生成注意力权重图。该权重图对原始特征图的每个位置进行加权，从而突出重要区域，抑制不重要的区域，帮助模型更准确地进行定位和识别。CA 的网络结构图如图 2 所示。

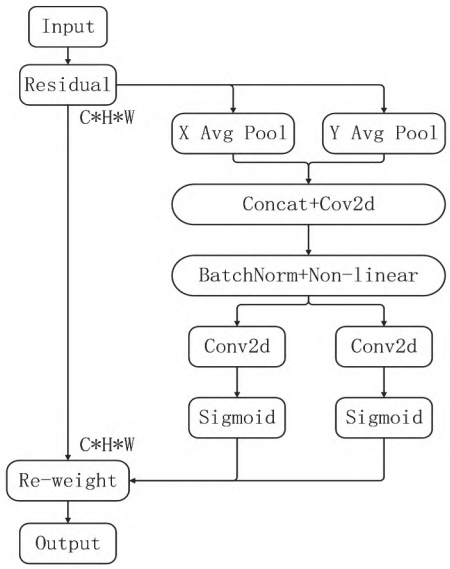

图 2 CA 的网络结构

2.2 添加 CBAM

CBAM 构建了通道 – 空间双域协同注意机制，区别于传统卷积神经网络 CNN 中单一维度的通道注意力机制，该模块采用级联结构，有机结合跨通道特征校准和空间域显著性筛选。在特征处理过程中，首先通过通道注意力模块（Channel Attention Module，CAM）[13] 动态调整各特征通道的权重系数，解决特征图通道间的信息分配问题。随后，应用空间注意力模块（Spatial Attention Module，SAM）[14] 对特征图中的关键区域进行定位，生成二维注意力掩码，强化关键区域的响应强度。这种双阶段注意力机制在保持通道域细粒度特征分析的同时，引入空间维度的自适应聚焦能力，使网络能够有效抑制冗余特征并增强判别性特征的表征能力。CBAM 注意力机制的结构如图 3 所示。

CAM 通过建立特征通道的自适应权重分配机制，实现对多维特征张量的通道域信息优化。该模块采用双路池化策略，首先通过全局平均池化（GAP）与全局最大池化（GMP），并行提取通道级统计描述量，构建联合特征表征；然后通过共享权重的多层感知机（MLP）对融合特征进行非线性变换，学习各特征通道的调制系数；最终，利用 sigmoid 函数将注意力权重归一化，并与原始特征图进行逐通道乘法运算，形成具有通道选择增强特性的优化特征空间。这种动态通道校准机制通过强化高语义关联通道的激活响应，有效提升特征映射的鉴别力。CAM 的计算式如式（1）所示，机制结构如图 4 所示。

$$M_c(F)=\sigma\{MLP[Augpool(F)]+MLP[Maxpool(F)]\}$$
$$=\sigma\{W_1[W_0(F^cAug)]+W_c[(W_0(F^cMax)]\} \quad (1)$$

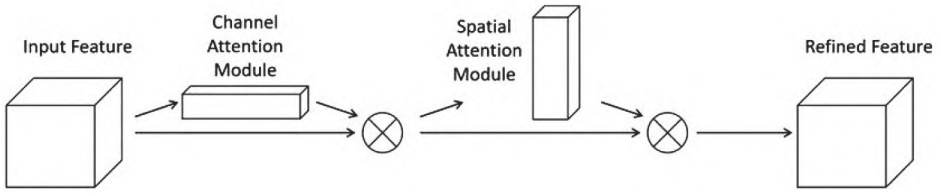

图 3 CBAM 注意力机制结构

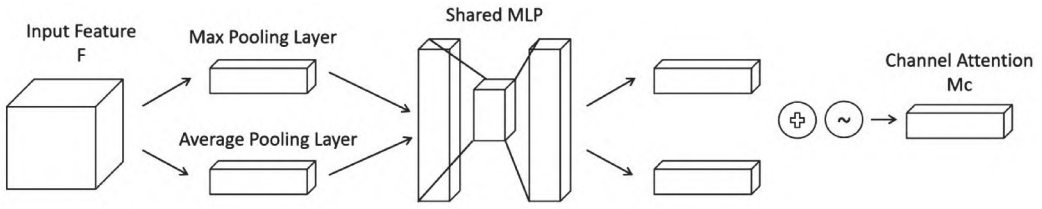

图 4 CAM 通道注意力机制结构

SAM 为输入特征图的每个位置分配一个注意力权重，这些权重可帮助网络集中注意力于感兴趣的区域。首先对输入的特征图实施全局平均池化（GAP）与全局最大池化（GMP）操作，分别提取特征图中的平均值信息和最大值信息；然后将池化后的特征图依据通道相加，得到两个一维向量，对这两个向量进行点积，

形成一个注意力权重矩阵；最后将注意力权重矩阵应用于输入特征图，得到空间注意力调整后的特征图。SAM 的计算式如式（2）所示，机制结构如图 5 所示。

$$M_s(F)=\sigma\left\{f^{7*7}\left\{[Augpool(F);Maxpool(F)]\right\}\right\}=\sigma\left\{f^{7*7}\left\{[f_{Aug}{}^{s};f_{Max}{}^{s}]\right\}\right\} \tag{2}$$

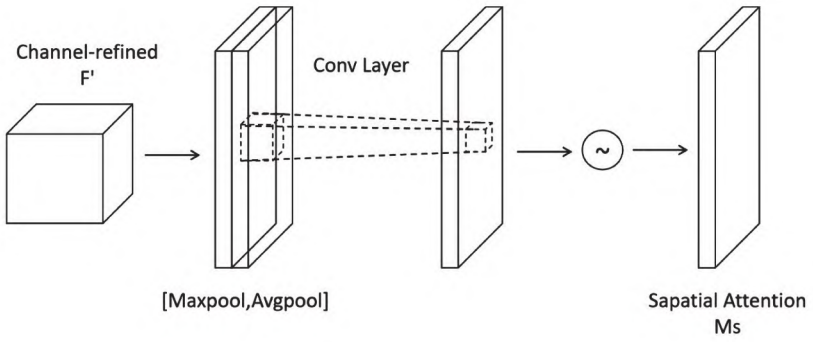

图 5　SAM 空间注意力机制结构

2.3　损失函数的改进

交并比（Intersection over Union,IoU）是目标检测任务中常用的一个评估指标，用于衡量预测框与真实框之间的重叠程度，重叠程度越高，IoU 越趋近于 1。其计算公式为：

$$IoU=\frac{Area_{\text{intersection}}}{Area_{\text{union}}} \tag{3}$$

在 YOLO v5s 基准模型中，采用广义交并比（Generalized Intersection over Union,GIoU）作为边界框回归损失函数存在固有局限性。当预测边界框完全包含于真实标注框内部时，GIoU 会退化为 IoU 进行运算，导致其因无法有效区分预测框的不同空间分布，使回归质量不稳定。为克服 GIoU 对目标框重叠形态的回归缺陷，本研究采用距离交并比（Distance Intersection over Union,DIoU）损失函数进行算法改进。DIoU 在重叠区域度量的基础上，引入边界框中心点欧氏距离惩罚项，构建具有几何解耦特性的评估准则，从而更精确地量化预测框与真实框的空间位置关系。DIoU 的计算式如下：

$$DIoU=IoU-\frac{\rho^2(b,b^{gt})}{c^2}=IoU-\frac{d^2}{c^2} \tag{4}$$

$$L_{DIoU}=1-DIoU \tag{5}$$

式中：b 和 b^{gt} 分别代表预测框和目标框的中心坐标参数，$\rho^2(b,b^{gt})$和 d 代表预测框和目标框的欧氏几何距离，c 表示预测框和目标框的最小外接矩形框的对角距离[15]。

3　实验配置与结果分析

3.1　实验环境与参数配置

本文所使用的实验环境和超参数如表 1、表 2 所示。

表 1　实验环境

环境配置	名称	参数
硬件配置	CPU	AMD R9-7945HX
	GPU	NVIDIA GeForce RTX4060
	显存	8GB
	操作系统	Windows 11
软件环境	Python	3.8.5
	Pytorch	1.8.0
	CUDA	12.0

表 2　超参数设置

名称	数值
Image_size	640 × 640 × 3
Epochs	300
Batch_size	16
Weight_decay	0.0005

3.2　数据集搜集与处理

本实验数据来源于 Open Images Dataset 开放平台，通过搜索条件“electric bicycle”与“helmet”，采集初始图像样本 2500 张，包括佩戴头盔与未佩戴头盔两类骑行场景。通过数据清洗，剔除成像模糊及存在标注歧义的样本后，最终保留有效图片 2360 张。为确保标注准确性，避免自动标注工具可能产生的误差，本实验采用 LabelImg 工具对样本进行人工标注，构建三元标签体系：‘helmet’（佩戴头盔）、‘head’（未佩戴头盔）、‘two_wheeler’（电动自行车）。最后采用分层随机抽样策略，按 8:2 比例将数据划分为训练集（1894 张）和测试集（473 张），确保两类目标在训练集和测试集中的分布一致性。

3.3　评价指标

在头盔佩戴检测实验中，通常选取精确率（P）、

召回率（R）以及平均精确度（mAP）作为模型性能的评价指标。精确度和召回率的计算公式如下：

$$P=\frac{TP}{FP+TP} \tag{6}$$

$$R=\frac{TP}{TP+FN} \tag{7}$$

式中：*TP* 代表预测结果为正样本，真实情况也为正样本；*FP* 代表预测结果为正样本，真实情况为负样本；*FN* 代表预测结果为负样本，真实情况为正样本[16]。

对 PR 曲线进行积分求出面积得到平均精度 *AP*，再对 *AP* 求平均值即可得到训练结果的 *mAP*。计算公式如下：

$$AP=\int_0^1 p(r)dr \tag{8}$$

$$mAP=\frac{\sum_{i=1}^{K} AP_i}{K} \tag{9}$$

式中: *K* 表示类别数，AP_i 代表第 *i* 个类别的精度。

3.4 数据与结果分析

在相同实验环境下，分别对 YOLO v5s 基准模型、YOLO v5s 改进模型和 SSD 算法模型和 Faster R-CNN 算法模型进行测试，对比结果如表 3 所示：

表 3 不同模型对比 %

算法模型	*P*	*R*
YOLO v5s 改进模型	88.2	90.4
YOLO v5s 基准模型	84.7	86.1
SSD	82.5	84.2
Faster R-CNN	87.8	88.1

实验数据分析表明，在电动车头盔检测任务的模型对比研究中，三类检测架构呈现显著性能差异。在 YOLO v5s 基准模型与 SSD、Faster R-CNN 构成的对照组中，Faster R-CNN 在精确率（Precision, 87.8%）、召回率（Recall, 88.1%）及平均精度均值（mAP, 88.9%）三项核心指标上均展现出相对优势。然而，经算法优化的 YOLO v5s 改进模型实现了检测性能突破，其 Precision、Recall、mAP 分别达到 88.2%、90.4% 和 90.9%，相较 Faster R-CNN 分别获得 0.4%、2.3% 和 2.0% 的性能提升。经数据对比证实，改进后的轻量化检测架构在保持实时性的同时，其综合检测效能显著优于传统两阶段检测器与基准单阶段模型，为复杂场景下的安全装备识别提供了更优解决方案。

3.5 消融实验

为了进一步显示本文改进措施的效果，在相同实验环境下进行消融实验，评估各个模块的性能，具体如表 4 所示。表中，A 代表引入 CA 坐标注意力，B 代表添加 CBAM 注意力机制，C 代表使用 DIoU 改进损失函数。

表 4 消融实验结果 %

算法模型	*P*	*R*	*mAP*
YOLO v5s	84.7	86.1	86.8
YOLO v5s+A	86.2	88.1	89.8
YOLO v5s+B	87.9	89.0	90.7
YOLO v5s+C	85.0	86.3	88.4
YOLO v5s+A+B+C	88.2	90.4	90.9

由实验数据定量分析，CBAM 注意力机制对头盔检测效果增益最为明显，使模型 mAP 提高了 3.9%，CA 坐标注意力使模型 mAP 提高了 3%，DIoU 损失函数使模型 mAP 提高了 0.6%。相较于 YOLO v5s 基准模型，引入 CA 坐标注意力、CBAM 注意力机制和 DIoU 后，改进模型的 mAP 提高了 4.1%，同时精确率和召回率也明显提高。这验证了双路径注意力协同机制的有效性——CA 通过通道 – 位置坐标编码增强小目标特征表征能力，CBAM 则通过空间 – 通道联合注意力抑制复杂遮挡场景下的背景噪声干扰，同时使用 DIoU 损失通过中心距约束优化边界框回归轨迹，使重叠目标的定位精度得到显著改善。

4 总结

为解决现阶段电动车头盔检测中小目标个体提取困难，目标间相互遮挡等问题，本文对基于 YOLO v5 的骑行头盔佩戴检测算法进行了改进。首先，针对复杂交通背景中信息提取的问题，引入双重注意力机制，通过 CA 坐标注意力和 CBAM 突出电动车头盔佩戴相关特征，弱化无用特征；然后通过引入 DIoU 替换 GIoU，对损失函数进行改进，优化边界框回归轨迹，改善重叠目标的定位精度，提高模型的收敛速度。经实验，改进后的 YOLO v5 模型的 Precision 提高了 3.5%，Recall 提高了 4.3%，mAP 提高了 4.1%。由实验结果可知，改进后的模型性能有可观的提升，有助于更高效地进行监管工作。

参考文献：

[1] 王轶宁，金文扬 . 从临床诊疗角度分析临海市整治不戴头盔骑坐电瓶车行为的促进意义 [J]. 中国乡村医药 ,2021,28(4):26.

[2] REN S Q, HE K M, 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,2017,39(6):1137-1149.

[3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on

Computer Vision, 2016: 21–37.
[4] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017: 6517–6525.
[5] REDMON J ,DIVVALA K S ,GIRSHICK B R , et al. You only look once: unified, real–time object detection.[J]. CoRR,2015,abs/1506.02640 22.
[6] ZHAO J W,TIAN G Z, QIN C, et al.Weed detection in potato fields based on improved YOLO v4: optimal speed and accuracy of weed detection in potato fields[J].Electronics,2022,11(22):3709–3709.
[7] 夏薪喆 . 基于 YOLO v5 的交通标志识别轻量化技术研究 [D]. 哈尔滨：黑龙江大学 ,2023.
[8] HU X ,DAI N ,HU X, et al. YOLO vT: CSPNet–based attention for a lightweight textile defect detection model[J].Textile Research Journal,2024,94(9–10):1021–1039.
[9] LIU S H, SONG Z D, HE L. Improving ECAPA–TDNN performance with coordinate attention[J].Journal of Shanghai Jiaotong University (Science),2024,(prepublish):1–7.
[10] BO P .Classification of images using EfficientNet CNN model with convolutional block attention module (CBAM) and spatialgroup–wise enhance module (SGE)[C]//Beijing Jiaotong Univ. (China),2022.
[11] ZHENG Z ,WANG P ,LIU W , et al. Distance–IoU Loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2020,34(7):12993–13000.
[12]张学立 , 贾新春 , 王美刚 , 等 . 安全帽与反光衣的轻量化检测：改进 YOLO v5s 的算法 [J]. 计算机工程与应用 ,2024,60(1):104–109.
[13] LIU Y ,ZHONG Y ,SHI S , et al.Scale–aware deep reinforcement learning for high resolution remote sensing imagery classification[J].ISPRS Journal of Photogrammetry and Remote Sensing,2024,209：296–311.
[14] ZHU X ,CHENG D ,ZHANG Z , et al.An empirical study of spatial attention mechanisms in deep networks.[J]. CoRR,2019,abs/1904.05873
[15] 李鸿治 , 舒远仲 , 肖靖 , 等 . 基于改进 YOLO v5s 的骑行头盔佩戴检测算法研究 [J]. 南昌航空大学学报 (自然科学版),2024,38(3):95–102.
[16] 陈宁 , 张武 , 王凯 , 等 . 基于改进 YOLO v7 的水下目标检测算法 [J]. 指挥信息系统与技术 ,2024,15(5):77–83.

Research on Helmet Wearing Detection Algorithm for Electric Bike Riders Based on Improved YOLO v5

BAI Yichen[1], JIA Meng[2], SHEN Xiaomeng[2]
(1. Harbin Institute of Technology (Weihai), Weihai 264200, China;
2. Xinxiang University, Xinxiang 453000, China)

Abstract: To address challenges in helmet–wearing detection for electric bike riders, such as difficulties in small–target recognition, complex occlusions, and inefficient manual supervision, this study proposes an improved YOLO v5–based detection algorithm. By introducing a dual attention mechanism and optimizing the loss function in the YOLO v5s baseline model, the detection performance is significantly enhanced. Specifically, a CA (Coordinate Attention) module is embedded into the Backbone to improve the model’s sensitivity to target locations; a CBAM (Convolutional Block Attention Module) is integrated into the Neck to suppress background noise through channel–spatial dual–domain collaboration; and the DIoU loss function replaces GIoU to optimize bounding box regression and improve localization accuracy for overlapping targets. Experiments on a self–built dataset (2,360 images) show that the improved model achieves precision, recall, and mean average precision (mAP) of 88.2%, 90.4%, and 90.9%, respectively, outperforming the baseline model by 3.5%, 4.3%, and 4.1%, and surpassing comparative models like Faster R–CNN. The algorithm balances lightweight design and high accuracy, offering an efficient solution for automated helmet–wearing supervision in complex traffic scenarios, with practical application value.

Key words: helmet wearing detection; YOLO v5; CA coordinate attention; CBAM attention mechanism; DIoU loss function

（责任编辑　王　磊）