DOI: 10. 13382/j. jemi. B2407813

改进 RT-DETR 的双轮车头盔检测算法*

孙光灵1,2 王薪博1 李艳秋1

(1.安徽建筑大学电子与信息工程学院 合肥 230601;2.安徽建筑大学安徽省古建筑智能感知与高维建模国际 联合研究中心 合肥 230601)

摘 要:针对双轮车头盔检测算法中出现的密集目标,远景小目标等复杂场景下的漏检、误检和检测准确率低的现象,以RT-DETR-r18为基础,提出一种改进的RT-DETR 双轮车头盔检测算法。设计了二重跨阶段的多尺度特征融合模块(DespBlock),将多核初始化模块(PKIBlock)融入到跨阶段模块中,在降低模型参数量的同时,有效增强了网络对远近场景中不同尺度目标的捕获能力;引入了小目标检测模块 Decoderhead-p2,有效增强了模型对小型目标的检测准确性;为了缓解复杂检测场景出现的正负样本不平衡以及边界框定位不准确问题,使用改进的损失函数 MPD_Focaler-IOU 替换原模型的 GIOU,通过设置阈值参数来改进 IOU 的计算方式,从而减少正负样本不平衡对模型性能的影响,引入最小垂直距离的计算方式,使得边界框在精细定位上有着更好的表现。实验表明,在 TSHW 数据集上,改进的 RT-DETR 相较于原模型,平均精度均值(mAP)mAP@ 0.5 提升了3.6%,参数量降低了 17.6%,同时保持较小的计算量,说明改进的模型可以有效提升复杂场景中对双轮车头盔检测的性能。 关键词: RT-DETR;目标检测;双轮车头盔;损失函数

中图分类号: TN957.52; TP391.4 文献标识码: A 国家标准学科分类代码: 520.60

Improved helmet detection algorithm for two-wheeled vehicles of RT-DETR

Sun Guangling^{1,2} Wang Xinbo¹ Li Yanqiu¹

(1. School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230601, China; 2. Anhui International Joint Research Center for Intelligent Perception and High-dimensional Modeling of Ancient Buildings,

Anhui Jianzhu University, Hefei 230601, China)

Abstract: Aiming at the phenomena of leakage, false detection and low detection accuracy in complex scenes such as dense targets, small targets in distant view, etc., which occur in the helmet detection algorithm for two-wheeled vehicles, an improved RT-DETR helmet detection algorithm for two-wheeled vehicles is proposed on the basis of RT-DETR-r18. A dual cross-stage multi-scale feature fusion module (DcspBlock) is designed, and a multi-core initialization module (PKIBlock) is incorporated into the cross-stage module, which reduces the number of model parameters while effectively enhancing the network's ability to capture targets of different scales in the near and far scenes; a small target detection module Decoderhead-p2 is introduced, which effectively enhances the model's ability to detect small target detection accuracy; in order to alleviate the positive and negative sample imbalance and the inaccurate positioning of the bounding box in complex detection scenarios, the original model's GIOU is replaced by the improved loss function MPD_Focaler-IOU, and the computation of the IOU is improved by setting the threshold parameter, so as to minimize the impact of the positive and negative sample imbalance on the model's performance, and the computation of the minimum vertical distance is introduced to enable the bounding box to be finely localized. which makes the bounding box have better performance in fine localization. The experiments show that on the TSHW dataset, the improved RT-DETR improves the mAP@ 0.5 by 3. 6% and reduces the number of parameters by 17. 6% compared with the original model, while keeping a smaller computational volume, indicating that the improved model can

收稿日期: 2024-09-07 Received Date: 2024-09-07

*基金项目:中国建设教育协会教育教学科研立项课题(2023069)、2023 年安徽省住房城乡建设科学技术计划(2023-YF058,2023-YF113)、安徽 省高等学校科学研究重点项目(2023AH050164)、安徽省高校杰出青年科研项目(2023AH020022)资助 effectively enhance the performance of two-wheeled vehicle helmet detection in complex scenes. Keywords:RT-DETR; object detection; two-wheeler helmet; loss function

0 引 言

随着双轮车(电动车,摩托车,自行车)的不断普及, 双轮车已经成为我国市民主要的交通工具。近年来,公 民乱闯红的,不戴头盔等交通违章行为屡见不鲜,因此检 测并督促驾驶人员佩戴头盔具有重要意义。现阶段在智 能交通方面的目标检测仍存在着一些挑战,比如目标尺 度的变化多样性,复杂的上下文信息等环境因素。因此, 需要进一步展开研究,以提升在双轮车头盔检测技术上 的性能。

随着近些年深度学习的兴起,基于卷积神经网 络(convolutional neural networks, CNN)的目标检测算法, 比如双阶段的 Faster R-CNN (fast region-convolutional neural network)^[1] 和单阶段的 YOLO (you only look once)^[2]等算法,逐步提升了双轮车头盔的检测的性能, 对于 Faster R-CNN,在检测时会生成候选区域,对每个候 选区域内进行细致的分类和边界框回归,具有很高的准 确率。对于 YOLO 算法,将目标检测问题转化为一个回 归问题,不需要生成候选区域,并且结构比较简单,训练 的速度非常快。但是基于卷积神经网络的检测算法还存 在一些缺陷,虽然对局部的特征的获取比较突出,但对远 距离以及多尺度信息的捕获相对有限,对处理复杂环境 的场景的能力也略显不足。与此相反, DETR (detection transformer)^[3]是基于 Transformer^[4] 框架的目标检测网 络,对图像中的全局特征的捕获表现突出,适合处理具有 远距离依赖关系的复杂场景。并且 DETR 避免了手工设 计锚框和非极大值抑制 (non-maximum suppression, NMS)^[5],简化了检测流程。尽管 DETR 在某些方面相较 于卷积神经网络算法具有一些优势,但是 Transformer 的 自注意力机制的计算量较高,对于高分辨率图像,计算资 源需求较大,检测方法通常需要较长的训练时间,数据和 计算资源需求较高,并且对小目标的检测效果不如基于 卷积神经网络的检测算法。为克服这些缺点, Deformable-DETR^[6]引入了可变性卷积和多尺度的特征 融合[7]来减少计算复杂度,使得模型在处理高分辨率图 像时更加高效,提高模型对不同大小目标的检测能力。 Anchor DETR^[8]结合锚框思想和 DETR 的 Transformer 架 构,引入了锚点机制,设计锚点和查询一一对应,提高了 定位精度。在检测后处理过程中,结合 NMS 技术,提高 了检测结果的准确性。虽然先前的工作已经对 DETR 算 法做出了不断优化,但是基于 DETR 的检测算法的计算 成本仍然很高,这限制了 DETR 算法的实际应用。近期,

百度飞桨团队提出了实时目标检测算法 RT-DETR(realtime DETR)^[9],有效解决了 DETR 在训练效率、计算复杂 度和实时性方面存在的不足,在保持高检测精度的同时, 实现更高的检测速度和更低的计算资源需求。

针对头盔检测中存在的问题,刘琛等^[10]基于 SSD 网络,通过添加通道注意力、空间注意力模块和群体感受野模块,并通过衰减学习率的方法优化网络,有效提升了检测精度,但边界框定位不够准确。韩东辰等^[11]提出了ML-CSPDarknet53 网络,基于 YOLOv5,在主干网络的基础上添加 MLCA 注意力机制,可以同时关注通道信息、空间信息、局部信息和全局信息,增强特征提取能力,但检测场景较为简单,对小目标的检测效率不高。储开斌等^[12]针对 Faster R-CNN 算法在小目标上检测精度低的现象,采用主要样本注意力机制方法对样本按权值重新进行排序,优化样本不均衡问题,用可变形卷积模块代替模型中的标准卷积,进一步改善识别精度,但是模型的参数量并未得到优化,需要进一步的轻量化处理。

RT-DETR 主要基于 Transformer 架构,虽然捕获全局 信息方面比较突出,但是在多尺度的特征提取和上下文 信息的利用率不如传统的 CNN 架构,导致复杂场景下的 双轮车头盔检测中容易出现漏检误检现象。本文以 RT-DETR 算法为基础,提出了一种针对于复杂场景双轮车 头盔检测的多尺度特征融合的改进算法,有效地提高了 检测性能。

本文提出了一种改进二重跨阶段特征融合模 块(double cross stage partial block, DcspBlock)。参照 CSP^[13]结构以及 PKINet (Poly kernel inception network)^[14]中的 PKIBlock(Poly kernel inception block)进 行改进,保证轻量化以及跨多尺度捕获上下文信息的同 时进一步提高了特征提取的性能。在 Encoder 模块中引 入深层次的 Decoderhead-p2 检测模块,以此来增强对特 征中小目标的检测能力。结合 Focaler-IOU^[15]和 MPDIOU^[16]设计损失函数 MPD_Focaler-IOU,用其替换原 始模型中的 GIOU^[17]。更准确地评估正负样本的不平衡 和边界框之间的位置关系。

1 RT-DETR 模型

RT-DETR 按照参数可分为 RT-DETR-r18、RT-DETRr50、RT-DETR-r101、RT-DETR-l、RT-DETR-x。其中 RT-DETE-r18 涉及的参数规模相对较小,基于实验配置以及 需求,选择 RT-DETR-r18 作为基线模型。RT-DETR-r18 的成分包括主干(backbone)、编码器(encoder)和解码 器(decoder)等部分组成,如图1所示。

主干网络主要包括 ConvNormLayer 结构、MaxPool 结构和 BasicBlocks 结构。ConvNormLayer 结构是由卷积层 归一化层组成,用于提取图像的基本特征并加速模型收敛,这种多层次的 ConvNormLayer 组合能够更好地从输入数据中提取特征,并在处理过程中逐步提高特征的抽象度。MaxPool 结构是最大池化层,通过取每个池化窗口内的最大值来减少特征图的空间尺寸来降低参数量和计算复杂度,同时保留最显著的特征,由于减少了后续层的参数数量,因此在一定程度上可以减轻模型的过拟合风险。BasicBlocks 结构引入残差连接的方式,可以有效缓解深层网络中的梯度消失问题,增强模型的特征表示以及模型拟合能力。

编码器部分主要包括 AIFI (attention-based image feature integration)结构和 RepC3 (repetitive convolutional

contextualization with 3×3 kernels)结构。其中 AIFI 是一种注意力机制的特征融合模块,作用在目标检测模型中的特征融合和特征提取阶段。AIFI 的目的在于提高模型的特征表达能力和检测性能,虽然 AIFI 模块引入了复杂的注意力机制,但它在设计时考虑了计算效率,保证模型在保持高检测精度的同时,能够实现实时检测。为了增强模块的特征提取能力,RepC3 通过设计重复的 3×3 卷积核来增加网络的深度,这种重复卷积的设计使得网络深度增加,有助于提取图像中的粒度信息。

解码器采用自注意力机制和交叉注意力机制结合的 编码器生成的图像特征,以此来生成目标检测结果。在 解码器的最后部分,检测头将多层解码处理后的特征映 射到类别标签和目标框的坐标。这些检测结果会经过后 序处理(如非极大值抑制)去除重复的检测目标框,并得 到最终的检测结果。



图 1 RT-DETR 结构 Fig. 1 RT-DETR structure

2 改进 RT-DETR 模块

本文从以下方面对 RT-DETR-r18 模型进行如下改进:在主干网络中引入量级的 DcspBlock 模块,有效提升

了模型对多尺度特征目标的捕获和检测能力;在 Encoder 中添加 Decoder-p2 检测头,提升对小目标的检测效果;在 损失函数方面结合了 Focaler-IOU 和 MPDIOU,以增强框 定位和匹配的准确性。改进模型结构如图 2 所示。



图 2 改进 RT-DETR 结构 Fig. 2 Improved RT-DETR structure

2.1 主干网络改进

由于道路上交通的复杂场景,使得双轮车头盔的检测中出现多种不同尺度的特征目标,同时多样化的上下 文信息,会出现精度较低、误检漏检等问题。为了捕获多 尺度的特征信息,以及应对复杂的检测场景,轻量级网络 PKINet采用无扩展的多尺度卷积核来提取不同尺度的 目标特征并捕获局部上下文。其中 PKIBlock 模块设计 的并行多尺度卷积对于捕获复杂环境的多尺度特征展现 出了不错的效果。PKINet 结构如图 3 所示。



图 3 PKINet 结构 Fig. 3 PKINet structure diagram

partial)^[18]结构按照通道数分为两个输入特征,让两个特征分别进入前馈神经网络(FFN)和 PKIBlock 得到两个输出,并将得到的输出进行特征融合^[19],使得模型既有精细的局部表示,又有全局的上下文信息,以此来增强模型的细节表征能力和的全局表征能力。PKINet 单阶段模块的输出结果如式(1)所示。

 $F = Conv(Concat(FFN(split_1(X))),$ PKI(split_1(X))))

(1)

其中, *split*₁和 *split*₂表示的是将输入特征沿着通道 分为两部分,假设输入的通道维度为 X(C,H,W), *split*₁ 表示提取前 C/2 个通道,得到特征图 *split*₁(X) 的维度 为(C/2,H,W),同理 *split*₂(X)表示提取后 C/2 个通道 数。得到的两个通道的特征分别流入到 FFN 和 PKI 模 块进行处理(式(2))。FFN 中的 Liner 表示非线性变换 的全连接层,通过堆叠多个线形层和激活函数将特征映 射到更高维度的空间。PKIBlock 中设计了并行处理的深 度卷积(depthwise convolution, DWConv)^[20]进行特征融 合,这种并行的 DW 卷积可以在不同尺度上进行卷积操 作,能够从多个角度提取特征,有效提高了在复杂场景中 的特征提取能力,同时深度可分离卷积减少了计算量和 参数量,进一步提高了模型的计算效率。

$$FFN(split_{1}(X)) = Liner_{2}(Liner_{1}(X))$$

$$PKI(split_{2}(X)) = Conv(split_{2}(X) + \sum_{i=1}^{4} DWConv_{(2k+1)(2k+1)}(split_{2}(X)))$$

$$k = 1, 2, 3, 4$$
(2)

在此基础上为了进一步增强模型对复杂场景数据任务的处理,提出了 DespBlock。参照 CSP 结构并结合 PKINet 中的 PKIBlock 进行设计。整个 DespBlock 模块采 用二重 CSP 结构,如图 4 所示,将输入特征按照通道数划 分流入两条分支,一条分支流入到 PKIBlock 模块,用来获取丰富的多尺度的特征信息。另一条分支则经过第二 重 CSP 结构进行推理和特征融合,其中瓶颈

模(bootleneck)^[21]用来提取丰富的局部特征,同时保留 较低的复杂度; silu 激活函数用来结合线性和非线性的 特征,提高模型的表达能力。第1个 Concat 操作将 C/2 的原始通道和经过 Bootleneck 提取后的 C/4 个原始通道 特征进行拼接,第2个 Concat 把第1次 Concat 处理后的 C/2 通道和经过 PKIBlock 处理后的 C/2 个通道信息进行 特征融合。这种设计有助于模型处理复杂的决策边界, 使其能够更好地应对多样化的输入数据。DcspBlock 的 多尺度特征融合机制使得模型能够同时捕捉到局部细节 和全局信息,增强了特征表示的丰富性,同时在保证提取 丰富特征的前提下减少计算和参数量。



Fig. 4 DcspBlock structure diagram

2.2 Encoder 部分改进

RT-DETR 的 Encoder 部分使用了 3 个检测模块,分 别为 Decoderhead-p3、Decoderhead-p4 和 Decoderhead-p5。 但是对于双轮车头盔检测而言表现不佳,比如一些复杂 场景会存在远距离的小目标难以捕获的现象。考虑到上 述情况,通过增加更浅层的检测头 Decoderhead-p2 来缓 解小目标的检测效果,如图 5 所示,为 Encoder 模块增加 更浅层的粒度信息,使深层的网络能够增强对特征图中 局部小型目标的捕获。更好地提升模型对小型目标的捕 获能力以及检测的准确性。

2.3 损失函数改进

目标检测的性能与边界框回归的准确性密切相关, RT-DETR 中采用了 GIOU 损失函数。假设(x_1^{st} , y_1^{st})、 (x_2^{st} , y_2^{st})和(x_1^{prd} , y_1^{prd})、(x_2^{prd} , y_2^{prd})分别代表真实框和预 测框的左上角和右下角坐标,定义如图 6 所示。A 和 B 分 别代表真实框和预测框的面积, C 是最小闭包框,指的是 能同时框住真实框和预测框的最小面积。GIOU 公式如



图 5 添加 Decoderhead-p2 检测头

Fig. 5 Adding the Decoderhead-p2 detection header $式(3) \sim (6)$ 所示。

$$L_{GIOU} = 1 - IOU + \frac{|C - A \cup B|}{|C|}$$
(3)

$$A = | x_2^{g_1} - x_1^{g_1} | \times | y_2^{g_1} - y_1^{g_1} |$$
(4)

$$B = |x_2^{prd} - x_1^{prd}| \times |y_2^{prd} - y_1^{prd}|$$
(5)

$$C = (\max(x_2^{gt}, x_2^{prd}) - \min(x_1^{gt}, x_1^{prd})) \times (\max(y_2^{gt}, y_2^{prd}) - \min(y_1^{gt}, y_1^{prd}))$$
(6)



Fig. 6 Bounding box definition

在目标检测任务中,正样本(目标物体)一般小于负 样本(背景),但 GIOU 将所有样本一视同仁,可能会导致 在训练过程中过于关注容易区分的负样本而忽略了复杂 的正样本进而会导致一些漏检现象。虽然 GIOU 考虑了 最小闭包框,但对框的相对位置和长宽比不够敏感。当 两个框的长宽比相差较大时,GIOU 并没能突出这种差 异。为了应对这些问题,使用改进的损失函数 MPD_ Focaler-IOU 替换原模型的 GIOU。损失函数的定义如式 (7)所示。

$$L_{MPD_Focaler-IOU} = 1 - \frac{IOU - m}{u - m} + \frac{d_1}{hw} + \frac{d_2}{hw}$$
(7)

式中: m 和 u 分别代表低阈值和高阈值参数,用来调节 MPD_Focaler-IOU 的敏感度。通过调节这两个参数对正 样本赋予更高的权重来改善模型对复杂场景中正样本的 识别能力。hw 为对角线差异归一化参数,用于归一化两 个边界框对角线的平方差,通过调节 hw 的值使模型适应 不同的检测场景。考虑到使用最小垂直距离来表示两个 框之间的位置关系,用 d₁ 和 d₂ 代表真实框和预测框对角 线的平方差。用公式可表示为:

$$d_{1} = \sqrt{(x_{1}^{prd} - x_{1}^{gt})^{2} + (y_{1}^{prd} - y_{1}^{gt})^{2}}$$
(8)

$$d_2 = \sqrt{(x_2^{prd} - x_2^{gt})^2 + (y_2^{prd} - y_2^{gt})^2}$$
(9)

与原模型的 GIOU 相比, MPD_Focaler-IOU 设置的调 节参数有效减少了正负样本不平衡对模型性能的影响, 同时进一步考虑了真实框和预测框的最小垂直距离, 使 模型在框的相对位置和长宽比上有更好的表现, 可以更 准确地反映真实框和预测框之间的位置关系, 从而在边 界框的精细定位上有着更好的表现。

3 实验与结果分析

3.1 数据集介绍

数据集采用的是开源双轮车头盔检测数据集 TSHW(traffic safety helmet wear)^[22],该数据集包含不同 角度和不同时间的检测图像,在不同类别目标的数量上 保持相对平衡。按照双轮车与驾乘人员整体(two_ wheeler)、未戴头盔的人头(without_helmet)、戴头盔的人 头(helmet)进行定位与分类标注。图片划分训练集、验 证集、测试集的比例为8:1:1。

3.2 实验环境和参数设置

实验使用的是 Windows10 操作系统, CPU 型号为 i5-12400F, GPU 型号为 NVIDIA GeForceRTX 3060(12 GB), 训练框架采用了 Python3.9、Pytorch2.0.0、CUDA11.8。 训练参数如表1所示。实验设置了对比试验和消融实验 用来验证改进模型的有效性。

表1 实验参数配置

Table 1 Configuration of experimental parameters

参数名	参数值
迭代次数(epochs)	100
学习率(h0)	0.000 1
图片尺寸(imgsz)	640
批处理大小(batch)	4
梯度动量(momentum)	0.9

3.3 评价指标

本文实验采用平均精度均值(mAP)、召回率(recall)、浮点计算量和参数量(parameters)作为衡量指标。

召回率是衡量模型在识别正例方面表现如何的指标,表示模型成功找到的正例数占所有真实正例数的比例,如式(10)所示。

$$Recall = \frac{TP}{TP + FN}$$
(10)

式中: TP 表示模型正确检测为正例的样本数; FN 表示 模型错误检测为负例,但实际为正例的样本数。

平均精度均值反应了整个模型的全面性和准确率。 对于多类别的目标检测任务,首先计算每个类别的平均 精度(AP),然后对所有类别的 AP 取平均值,得到 mAP 为:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i$$
(11)

式中:N是类别数;AP;是第 i个类别的平均精度。

3.4 消融实验

1) 模块消融实验

为验证改进模块的有效性,选取 RT-DETR-r18 为基 准模型进行消融实验,且每次实验的实验环境和参数设 置均保持一致,得到 8 组实验数据,如表 2 所示。表 2 中 A 代表使用改进模块 DcspBlock 替换原始模型的 BasicBlocks,B 代表在 Encoder 部分引入小目标检测头 p2,C 代表使用改进损失函数 MPD_Focaler-IOU 替换原 模型的 GIOU。从消融数据可以得出,由于特征提取方式 的改变和网络结构的不断优化,模型的计算量有所增加, 但模型的检验精度值提升较明显,且参数量也得到很大 程度的优化。加入改进点 A 后 mAP@0.5提升了1.2%, 参数量降低了11.1%;引入改进点 B,mAP@0.5提升了 1.6%,参数量进一步降低了6.5%;引入改进点 C, mAP@0.5提升了0.8%。实验表明基于的改进与原始 的 RT-DETR 模型相比 mAP@0.5提升了3.6%,参数量 降低了17.6%。表明改进后的模型在保证精度上涨的同 时,又进一步实现了模型的轻量化。改进前后训练过程 中精度以及损失曲线的变化情况如图7所示,改进后模 型在数据拟合效果上有显著的提升。

2) 损失函数消融实验

为了评价改进后损失函数 MPD_Focaler-IOU 的优势,将 MPDIOU、Focaler-IOU、MPD_Focaler-IOU 和原始 RT-DETR 模型使用的 GIOU 结合改进后的模型进行定量 对比,结果如表 3 所示,结果表明 MPD_Focaler-IOU 相较 于其他 IOU 取得了最高的精度,验证改进模型中损失函 数的优势。



Fig. 7 Comparison of training process

Table 2 Nodule ablation experiments									
٨	D	С	D II.(c)	mAP@ 0. 5/%	mAP@ 0. 5/%	mAP@ 0. 5/%	mAP@ 0. 5/%	Parameters/	浮点数/
А	D		Recall/ %	(all)	(two_wheeler)	(helmet)	(without_helmet)	$(\times 10^{6})$	GFLOPs
			74.6	77.4	92.5	84.9	55.0	19.9	56.9
\checkmark			77.1	78.6	93.3	84.9	57.6	17.7	57.7
	\checkmark		76.2	78.6	92.9	86.1	56.8	18.6	78.1
			74.8	77.6	92.1	83.8	57.1	19.9	56.9
	\checkmark		77.6	80. 2	92.8	86.4	61.3	16.4	79.7
			77.5	79.4	92.7	85.7	60.0	17.7	57.7
	\checkmark	\checkmark	77.4	78.8	92.7	86.4	57.3	18.6	78.1
	\checkmark	\sim	77.9	81.0	92. 7	87.3	63. 2	16.4	79.7

表 2 模块消融实验

表 3 损失函数消融实验

Table 3	Loss	function	ablation	experiments	

IOU	GIOU	MPDIOU	Focaler-IOU	MPD_Focaler-IOU
mAP@ 0. 5/%	80.2	80.8	80.6	81.0

3.5 对比实验

1) 检测头对比

为了验证在 Encoder 模块中引入浅层检测头 Decoderhead-p2的有效性,将p2检测头结合p3、p4、p5检 测头做了删减实验,如表4所示,实验数据表明,添加更 浅层的 p2 检测头可以有效地提高模型的检测精度。

2) 损失函数对比

为了验证改进损失函数的优势,将 MPD_Focalar-IOU 分别与 CIOU^[23]、WIOU^[24]、EIOU^[25]、SIOU^[26]、DIOU^[27] 以及原模型的 GIOU 进行定量对比,如表 5 所示。结果 表明 MPD_Focaler-IOU 取得了最高的 mAP,验证了损失 函数的性能优势。

表 4 检测头对比 Table 4 Comparison of detection heads

Tuble T Comparison of detection routes						
Decoder head	p(3+4+5)	p(2+3+4)	p(2+4+5)	p(2+3+5)	p(2+3+4+5)	
mAP@ 0. 5/%	79.4	80.0	79.5	80. 4	81.0	

表 5 损失函数对比 Table 5 Comparison of loss functions

IOU 类型	CIOU	WIOU	EIOU	SIOU	DIOU	GIOU	MPD_Focaler-IOU
mAP@ 0. 5/%	80. 2	80.6	80.1	80. 2	80.4	80.2	81.0

3) 不同模型对比

为了进一步验证改进模型的性能,将改进后的模型 与其他目标检测模型 Faster R-CNN、YOLOv5s^[28]、 YOLOv8s^[29]、YOLOv8m、YOLOv9c^[30]、YOLOv10b^[31]、 YOLOv8-ghoat^[32]、Deformable DETR 等模型进行了对比, 如表 6 所示。实验结果表明,改进后模型在 mAP@0.5 指标上比上述模型分别高出 18.3% \ 8.6% \ 7.2% \ 5.3% \ 2% \ 8.1%和 5%。经对比,模型在保证精度高的同时又 保持着较小的参数量。因此改进的模型在检验精度以及 参数量之间实现了良好的平衡。

Table 6 **Comparison of models** 模型 Recall/% mAP@ 0. 5/% mAP@ 0.5 : 0.95/% 浮点数/GFLOPs Parameters/($\times 10^6$) Faster R-CNN(r50) 60.6 62.7 48 4 41 56 206.4 YOLOv5s 54.6 9.1 24.1 69.8 72.4 YOLOv8s 70.8 73.8 55.3 11.1 28.4 YOLOv8m 73.7 75.7 57.0 25.8 78.7 YOLOv9c 78.2 25.3 75.4 58.6 103.7 YOLOv10b 74.8 79.0 59.3 20.5 98.7 YOLOv8s-ghoat 70.6 72.9 55.1 5.9 16.3 Deformable DETR 72.9 76.0 56.8 40 179.3 **RT-DETR-L** 75.4 77.4 57.8 32.0 103.4 本文 77.9 81.0 61.0 16.4 79.7

表 6 各模型对比 Table 6 Comparison of mode

3.6 泛化性检验

为了验证改进模型的泛化性,分别将改进模块作用 在 RT-DETR-r18 和 RTDETR-r50 上进行实验,验证了改 进模块在不同网络上的鲁棒性和泛化性。如表 7 所示。 相比之下,改进后 RTDETR-r50 模型在 mAP@0.5 上提 升了 1.5%,在 mAP@0.5:0.95 上提升了 2.0%。

表 7 不同网络对比 Table 7 Comparison of different networks (%)

	- F · · · · ·			.,
模型	指标	改进前	改进后	
DTDETD "19	mAP@ 0. 5	77.4	81.0	
KIDEIK-II8	mAP@ 0. 5 : 0. 95	57.3	61.0	
DTDETD	mAP@ 0. 5	79.7	81.2	
RIDEIR-r50	mAP@ 0. 5 : 0. 95	59.3	61.3	

此外,实验选取 Kaggle^[33] 网址中的头盔检测数据集 Helmet、夜间行人检测数据集 CTIR、红外检测数据集 infrared 分别进行实验,进一步验证改进模型在不同数据 集上的泛化能力。如表 8 所示。Helmet 数据集的 mAP@ 0.5 和 mAP@ 0.5 : 0.95 相较于原始模型分别提升了 3.7%和4.7%,在 CTIR 数据集上 mAP@ 0.5 和 mAP@ 0.5 : 0.95 分别提升了 1.0%和1.2%,在 infrared 数据集 上分别提升了 2.3%和1.5%。改进后的模型在多个数据 集上均有效提升了 mAP,证明模型在不同数据集上具有 较好的泛化能力。

表 8 不同数据集对比

 Table 8 Comparison of different datasets
 (%)

粉捉隹	改	双进前	改进后		
<u>奴</u> '师朱	mAP@ 0. 5	mAP@ 0. 5 : 0. 95	mAP50	mAP@ 0. 5 : 0. 95	
Helmet	80. 2	65.1	83.9	69.8	
CTIR	82.3	56.5	83.3	57.7	
infrared	49.8	29.3	52.1	30. 8	

3.7 定性分析

为了直观验证模型的准确性,使用 TSHW 数据集中 的测试集进行了检测效果对比。将改进后的 RT-DETR 模型与 RT-DETR 基准模型进行比较,分析检测效果。检 测出的类别包括双轮车和骑行人员整体(two_wheeler)、 佩带头盔(helmet)和未佩带头盔(without_helmet)。通常 情况下需要检测出双轮车整体和骑行人员整体,同时要 检测出骑行人员有无佩带头盔。改进前后模型效果的对 比如图 8 所示。从检测效果可以看出,对于远景目标, RT-DETR 原模型虽然可以清晰检测出骑行人员整体,但 并未检测出骑行人员的头盔佩带情况,出现了漏检现象, 而改进后的 RT-DETR 模型可以清晰地检测出远景目标



(a) 原图 (a) Original drawing



(b) RT-DETR检测效果 (b) RT-DETR detection effect



(c) 改进后RT-DETR检测效果 (c) Improved RT-DETR detection effect effect

图 8 与 RT-DETR 原模型对比 Fig. 8 Comparison with the Orange RT-DETR model

的头盔佩带情况。

为进一步检验模型的检测性能,用改进的 RT-DETR 模型和现阶段流行的检测器 YOLOv8m 作对比实验,如图 9 所示。结果表明 YOLOv8m 检测器对遮挡目标和部分 远景和近景存在着明显漏检和误检;而改进的 RT-DETR 检测器很好地避免了 YOLOv8 检测器出现的问题,体现 出改进模型的优势。

实验结果表明,改进后的模型有效地缓解了检测器 误检和漏检现象,尤其针对于遮挡物体以及对远景部分 的小目标的捕获展现出了比较优异的性能。

4 结 论

本文提出的基于 RT-DETR 的改进模型在解决双轮 车头盔检测中复杂场景下的误检、漏检等问题上取得了 较好的效果,得到如下结论。通过优化网络结构、引入新 型损失函数以及增强模型对多尺度目标和小型目标的特 征提取能力,显著提升了模型在复杂场景下的检测性能; 尽管改进后的模型在精度和参数量方面得到了优化,但 模型的计算量方面的优化不够明显;未来的研究将集中 在降低计算资源的消耗,比如通过蒸馏和剪枝等技术进 一步优化模型的性能。

参考文献

- REN SH Q, HE K M, GIRSHICK R, et al. Faster RCNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.
- [2] LI Z, XIE W, ZHANG L, et al. Toward efficient safety helmet detection based on YOLOv5 with hierarchical positive sample selection and box density filtering [J].
 IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-14.
- [3] CARIONN, MASSAF, SYNNAEVEG, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer, 2020:



(a) 原图 (a) Original drawing



(b) YOLOv8m检测效果 (b) YOLOv8m detection effect



(c) 改进后RT-DETR检测效果 (c) Improved RT-DETR detection effect effect

图 9 与 YOLOv8m 进行检测对比 Fig. 9 Comparison of detection with YOLOv8m

213-229.

- [4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:5998-6008.
- [5] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMSimproving object detection with one line of code [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 5561-5569.
- [6] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable transformers for end-to-end object detection [C]. International Conference on Learning, 2020.
- [7] 黄志海,邓耀华,吴光栋. IC 器件表面缺陷多光谱图像特征融合检测方法[J]. 仪器仪表学报, 2024, 45(9):24-33.
 HUANG ZH H, DENG Y H, WU G D. Multispectral image feature feature feature feature feature for the starting performance of the st

image feature fusion method for detecting surface defects in IC devices [J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 24-33.

- [8] WANG Y M, ZHANG X Y, YANG T, et al. Anchor DETR: Query design for transformer based detection [C]. Proceedings of the 36th AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2021: 2567-2575.
- [9] ZHAO Y AN, LYU W Y, XU SH L, et al. DETRS beat YOLOs on real-time object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [10] 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.
 LIU CH, WANG JT, WANG MY. Application of SSD network with visual mechanism in motorcycle helmet wearing detection[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(3): 144-151.
- [11] 韩东辰,张方晖,王诗洋,等. 基于改进 YOLOv5 的复 杂场景电动车头盔检测方法[J].现代电子技术,
 2025,48(1):123-129.
 HAN D CH, ZHANG F H, WANG SH Y, et al.

Improved YOLOv5 based electric bicycle helmet detection method in complex scenes [J]. Modern Electronics Technique, 2025,48(1):123-129.

[12] 储开斌,叶托,张继. 基于改进 Faster R-CNN 的头盔检 测算法研究[J]. 国外电子测量技术, 2022, 41(6): 86-92.

CHU K B, YE T, ZHANG J. Research on helmet detection algorithm based on improved Faster R-CNN [J]. Foreign Electronic Measurement Technology, 2022, 41 (6): 86-92.

[13] 韩强. 面向小目标检测的改进 YOLOv8 算法研究[D]. 长春:吉林大学, 2023.

HAN Q. Research on improved YOLOv8 algorithm for small target detection[D]. Changchun: Jilin University, 2023.

- [14] CAI X, LAI Q, WANG Y, et al. Poly kernel inception network for remote sensing detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27706-27716.
- [15] ZHANG H, ZHANG S. Focaler-IoU: More focused intersection over union loss [J]. ArXiv preprint arXiv: 2401.10525, 2024.
- [16] MA S, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression [J]. ArXiv preprint arXiv:2307. 07662, 2023.
- [17] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [18] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]. 2020 IEEE/CVF Conference on Computer Vision and 597 Pattern Recognition Workshops (CVPRW). IEEE,2020:1571-1580.
- [19] CHOLLET F. Xception: Deep learning with depth wise separable convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [20] SUN Q S, ZENG S G, LIU Y, et al. A new method of feature fusion and its application in image recognition [J].
 Pattern Recognition, 2005, 38(12): 2437-2448.
- [21] SRINIVAS A, LIN Y L, PARMAR N, et al. Bottleneck Transformers for visual recognition [C]. Conference on Computer Vision and Pattern Recognition, 2021.
- [22] DONGDONG G, BO S. Traffic safety helmet wear detection based on improved YOLOv5 network [J]. Optoelectronics Letters, 2025, 21(1):35-42.

- [23] ZHENG Z H, WANG P, REN D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2022, 52(8):8574-8586.
- [24] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [J]. ArXiv preprint arXiv:2301.10051, 2023.
- [25] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022,506: 146-157.
- [26] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression [J]. ArXiv preprint arXiv: 2205. 12740, 2022.
- ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C].
 Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [28] 米增,连哲.面向通用目标检测的 YOLO 方法研究综述[J].计算机工程与应用,2024,60(21):38-54.
 MI Z, LIAN ZH. A review of YOLO methods for universal object detection[J]. Computer Engineering and Applications, 2024, 60(21): 38-54.
- [29] WEN L H, JO K H. Three-attention mechanisms for one stage 3-D object detection based on LiDAR and camera[J].
 IEEE Transactions on Industrial Informatics, 2021, 17(10):6655-6663.
- [30] WANG C Y, YEH I H, MARK L H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision. Springer, 2025: 1-21.
- [31] WANG A, CHEN H, LIU L, et al. YOLOv10: Realtime end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.
- XIONG E J, ZHANG R F, LIU Y H, et al. Ghost-YOLOv8 detection algorithm for traffic signs [J]. Computer Engineering and Applications, 2023, 59(20): 200-207.
- [33] BOJER C, MELDGAARD J. Kaggle forecasting competitions: An overlooked learning opportunity [J]. International Journal of Forecasting, 2021, 37 (2): 587-603.

作者简介



孙光灵(通信作者),2002 年于合肥工 业大学大学获得学士学位,2005 年于合肥 工业大学获得硕士学位,现为安徽建筑大学 副教授,主要研究方向为机器学习、图像 处理。

E-mail: sunguangling@163.com

Sun Guangling (Corresponding author) received the B. Sc. degree from Hefei University of Technology in 2002, and the M. Sc. degree from Hefei University of Technology in 2005. Now he is an associate professor at Anhui Jianzhu University. His main research interests include machine learning and image processing.



王薪博,2023年于安徽大学江淮学院 获得学士学位,现为安徽建筑大学硕士研究 生,主要研究方向为图像处理、目标检测。 E-mail: 949944196@163.com

Wang Xinbo received his B. S. degree from Jianghuai College of Anhui University in

2023. He is now a M. Sc. candidate at Anhui Jianzhu University. His main research interests include image processing

and object detection.



李艳秋,2010年于巢湖学院获得学士 学位,2014年于合肥工业大学获得硕士学 位,2018年于合肥工业大学获得博士学位, 现为安徽建筑大学讲师,主要研究方向为情 感计算、图像处理、集成学习。

E-mail: liyanqiu2012@163.com

Li Yanqiu received the B. Sc. degree from Chaohu College in 2010, the M. Sc. degree from Hefei University of Technology in 2014, and the Ph. D. degree from Hefei University of Technology in 2018. She is now a lecturer at Anhui Jianzhu University. Her main research interests include affective computing, image processing, integrated learning.