文章编号:1671-6833(2026)02-0001-08

改进 YOLOv7-tiny 的施工现场安全衣帽穿戴检测算法

张 震[1], 张晨稳[2], 张俊杰[3], 裴胜利[3], 王文娟[4]

(1.郑州大学 电气与信息工程学院,河南 郑州 450001;2.郑州大学 河南先进技术研究院,河南 郑州 450001;3.河南汇融油气技术有限公司,河南 郑州 450001;4.广东省轻工业技师学院 机电工程学院,广东 广州 511330)

摘 要:针对当前施工现场安全衣帽穿戴检测算法在复杂背景、弱光环境及目标遮挡情况下的抗干扰能力不足,导致检测精度低、漏检率高及误检现象频繁等问题,提出了一种改进 YOLOv7-tiny 的施工现场安全衣帽穿戴检测算法。首先,在特征提取区域引入 EMA 注意力机制增强网络特征提取能力,弱化复杂背景干扰;其次,在特征融合部分插入 RFEM 模块提升网络感受野,获取更广泛的上下文信息,增强对小目标的感知能力;最后,采用 Shape-IoU 替换 IoU 边界回归损失函数,提升检测准确性。实验结果表明:改进模型在自制数据集上的 $mAP@0.5$ 达到 90.4%,相比原模型提高 3.0 百分点;帧率达到了 93 帧/s,模型参数量仅为 6.1×10^6。相比 YOLOv8s、YOLOv9s 等模型,所提算法在检测精度、速度和模型轻量化方面更具优势,适合施工现场的实时检测应用。

关键词:YOLOv7-tiny; 注意力机制; RFEM; Shape-IoU; 安全衣帽检测

中图分类号:TP391 **文献标志码**:A **doi**:10.13705/j.issn.1671-6833.2025.05.001

建筑业作为中国先导性的基础产业在国民经济中占有重要地位[1],但建筑施工现场经常存在各种各样的施工安全风险,如高处坠落、物体打击、坍塌、起重伤害及其他伤害等,这些安全风险时刻威胁着施工人员的人身安全。穿戴反光衣可以提高工人的可见性,减少事故发生的风险;佩戴安全帽可以最大限度地减少创伤性脑损伤风险进而保护施工人员的安全。人工巡检施工现场人员安全衣帽的穿戴问题消耗大量人力,且无法做到持续监管,针对当前施工现场安全生产要求,基于深度学习的目标检测算法更适合施工现场的实时监控。

当前的目标检测算法可分为两阶段目标检测和单阶段目标检测算法[2]。两阶段目标检测算法包括 Fast R-CNN[3]、R-FCN[4] 和 Mask R-CNN[5] 等。这些算法在图像中生成一系列候选区域,然后对这些候选区域进行分类和回归。单阶段目标检测算法主要包括 YOLO[6-10]、SSD[11]、CornerNet[12] 和 M2Det[13] 等。这些算法无须生成候选区域,在一个步骤中直接预测对象的类别和位置。对比两阶段目标检测算法,单阶段目标检测算法虽然会损失少量检测精度,但能够快速完成目标检测,在工程应用上更具优势。

单阶段目标检测算法 YOLOv7-tiny[8] 作为基础网络,在保持轻量化设计的同时,能够在计算资源受限的环境中实现高效的目标检测,适用于施工现场实时检测场景。本文以 YOLOv7-tiny 为基础网络,提出了一种改进 YOLOv7-tiny 网络。改进 YOLOv7-tiny 网络可以避免施工现场安全衣帽穿戴检测中因复杂背景干扰、目标尺寸小、目标遮挡等问题导致的检测精度低、漏检和误检等现象的发生。为增强检测鲁棒性,首先,在特征提取网络中引入基于跨空间学习的高效多尺度注意力机制(efficient multi-scale attention,EMA)[14],用以捕捉不同尺度特征;其次,在特征融合网络插入感受野增强模块(receptive field enhancement module,RFEM)[15]来扩大特征图的感受野;最后,采用 Shape-IoU[16] 边界回归损失函数替换 IoU[17] 边界损失函数,从而提升检测的准确性。

1 相关工作

在基于 YOLO 的安全帽和反光衣的检测中,反

收稿日期:2025-09-23;**修订日期**:2025-11-20

基金项目:河南省重点研发专项(231111211600);河南省重大公益专项(201300311200)

作者简介:张震(1966—),男,河南郑州人,郑州大学教授,博士,博士生导师,主要从事计算机视觉研究,E-mail:zhangzhen66@126.com。

引用本文:张震,张晨稳,张俊杰,等.改进 YOLOv7-tiny 的施工现场安全衣帽穿戴检测算法[J].郑州大学学报(工学版),2026,47(2):1-8.(ZHANG Z, ZHANG C W, ZHANG J J, et al. Improved YOLOv7-tiny for safety clothing and helmet wearing detection algorithm on construction site[J]. Journal of Zhengzhou University (Engineering Science), 2026,47(2):1-8.)

光服检测算法 YOLO-FL[18] 在 YOLOv4-tiny 的颈部网络引入 SPPF 模块,并提出了一种新的交并比计算方法 KIoU,将其引入到原 *K*-means++聚类方法中,使得模型平均检测精度比原算法提高了 8.29%;YOLO-S[19] 是一款轻型头盔佩戴检测模型,通过数据增强来平衡数据类别,并用轻量级网络 MobileNetV2 取代 YOLOv5s 原来的骨干网络,最后对模型通道进行剪枝和知识蒸馏,减少了网络计算量;YOLO-M[20] 对 YOLOv5s 进行改进,选择轻量级网络 MobileNetV3 作为骨干网络,设计 BiCAM 的注意模块并插入到网络中,特征融合模块引入残差结构,在汇聚原始图像信息的同时,降低模型参数量,进一步提高了安全帽检测的精度;Peng 等[21] 在 YOLOv5 头部网络中添加了有效的通道注意力(ECA)机制,并将损失函数替换为 NWD,在不增加太多计算复杂度的情况下有效提升 YOLOv5 中小型安全帽的检测精度;Ma 等[22] 使用基于 YOLOv4 的稀疏训练和修剪策略,改进后的 Light-YOLOv4 模型大小压缩至原模型的 10%,虽精度略有损失,但检测速度却大大提高。

虽然上述方法都提高了检测性能,但对于施工现场安全衣帽穿戴的检测仍存在一些挑战。目前对安全衣帽穿戴检测的研究主要集中在安全帽上,忽视了对反光服装的检测;缺乏包含安全帽、反光衣的图像数据集;施工现场环境复杂,拍摄得到的检测目标小、存在遮挡、背景复杂的图像存在目标特征信息丢失的情况,从而导致检测结果出现漏检和误检等现象。

2 改进 YOLOv7-tiny

2.1 改进 YOLOv7-tiny 网络结构

YOLOv7-tiny[8] 是 YOLOv7 系列的轻量化版本,专为计算资源受限环境设计,旨在实现快速高效的目标检测。改进后的 YOLOv7-tiny 网络在保持轻量化的同时,针对复杂背景、小目标检测和精度提升进行了优化。其结构包括特征提取网络(Backbone)、特征融合网络(Neck)和检测头(Head),改进 YOLOv7-tiny 网络结构如图 1 所示。在特征提取网络(Backbone)中,本文引入了高效多尺度注意力机制 EMA[14],通过多尺度学习增强特征捕捉能力,减少复杂背景干扰,提升精度和鲁棒性,同时保留 YOLOv7-tiny 的核心卷积层、ELAN 模块和 MP 模块。在特征融合网络(Neck)中,添加感受野增强模块 RFEM[15],利用不同扩张率的卷积层扩大感受野,增强多尺度特征理解,提升复杂背景下小目标检测能力。改进后的 Neck 网络不仅保留了 YOLOv7-tiny 的轻量化特点,还在特征融合过程中强化了对多尺度特征的理解。在检测头(Head)部分,采用 Shape-IoU[16] 损失函数代替传统的 IoU[17] 损失函数,优化边界框回归精度,特别在复杂场景中提升不规则目标回归的准确性。检测头通过 CBL 模块进行边界框定位、分类和精度估计,生成最终结果。

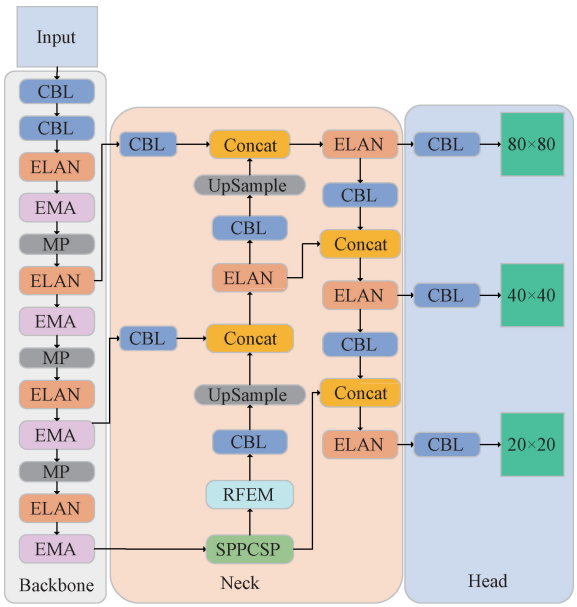


图 1 改进 YOLOv7-tiny 网络结构

Figure 1 Improved YOLOv7-tiny structure

2.2 EMA 模块

高效多尺度注意力机制 EMA[14],通过多尺度并行网络建立短期和长期依赖关系,实现局部跨通道交互并整合不同尺度的上下文特征。图 2 为 EMA 模块结构。如图 2 所示,EMA 模块重构部分通道,将特征划分为不同分支以学习多层语义信息。EMA 模块包含 3 条分支:两条 1×1 分支和 1 条 3×3 分支。两条 1×1 分支分别沿水平方向和垂直方向进行编码,获取注意力信息,池化后的特征通过 1×1 卷积和 Sigmoid 生成注意力图。在 3×3 分支中,通过堆叠卷积捕获多尺度特征,以增强表达。1×1 和 3×3 分支输出的特征图通过矩阵点积生成空间注意力图,对目标精确定位,最终融合生成像素级的注意关系,提取全局上下文信息,突出重要像素。在施工现场安全检测中,复杂背景常导致目标识别精度下降,引入 EMA 模块可以在不同尺度下提取和整合特征,捕捉多层次的上下文信息,提升模型在复杂背景下的特征表达能力,增强网络对目标的鲁棒性。EMA 模块通过多尺度特征融合,在复杂场景中可有效区分目标和背景,从而提高目标识别的精度。

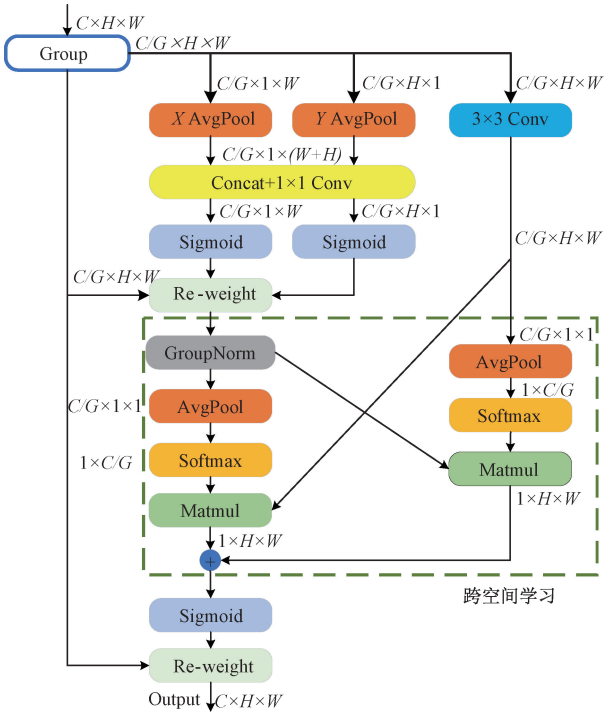


图 2　EMA 模块结构

Figure 2　Structure of EMA module

2.3　RFEM 模块

感受野增强模块 RFEM[15] 用于增强感受野,提升多尺度目标检测与识别的准确性。该模块由多个分支和聚合加权层组成。多分支部分通过不同扩张率的卷积捕捉多尺度信息,各分支共享权重,唯一的区别在于感受野的大小。每个分支采用 3×3 卷积核,并通过残差连接防止梯度消失和梯度爆炸。聚合加权层则整合各分支信息并对其进行加权,以平衡不同分支的表示能力。将 RFEM 模块与 SPPCSP 结合后,能够有效扩大特征图的感受野,改善上下文理解与特征提取,进而提升小目标检测和复杂背景下的安全衣帽穿戴检测性能。改进后的 RFEM 模块结构如图 3 所示。

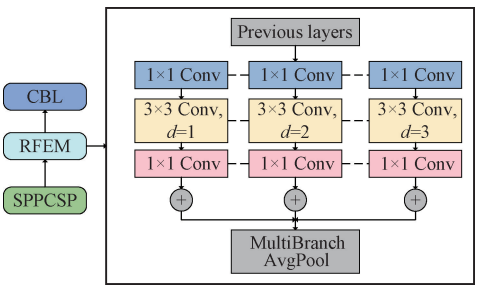


图 3　改进后的 RFEM 模块结构

Figure 3　Structure of improved RFEM module

2.4　Shape-IoU

Shape-IoU[16] 方法主要针对目标检测任务中的边界框回归准确性问题。传统的边界框回归方法通常仅考虑预测框与真实框之间的几何关系,忽略了边界框自身的形状和尺度对回归精度的影响。Shape-IoU 通过关注边界框的形状和尺寸信息,进一步提高了回归的准确性,从而提高目标检测的整体效果,其计算示意图如图 4 所示,相关公式及计算如下。

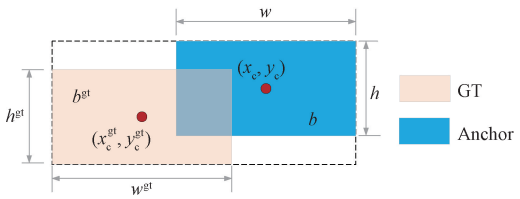


图 4　Shape-IoU 计算示意图

Figure 4　Diagram of Shape-IoU calculation

$$\mathrm{IoU} = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|}。\tag{1}$$

式中:IoU(intersection over union)为交并比,用于评估预测框与真实框的重叠情况,其值为 0~1,值越大表示重叠度越高,IoU 帮助网络判断检测准确性,提高检测精度;b 表示预测边界框;b^{gt} 表示真实边界框(GT 框)。

$$ww = \frac{2(w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}}。\tag{2}$$

式中:ww 表示真实框在水平方向上的相关权重系数,通过引入该系数考虑预测边界框与真实边界框的宽度差异,它对边界框的水平方向尺度进行精细调整,从而提升定位准确性;w^{gt} 表示真实边界框的宽度;h^{gt} 表示真实边界框的高度;$scale$ 为比例因子,与数据集中目标的比例有关。

$$hh = \frac{2(h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}}。\tag{3}$$

式中:hh 为引入垂直方向的相关权重系数,以修正预测框与真实框在高度上的差异,使检测框在垂直方向上的误差减小,增强上下边界的拟合能力。

$$distance^{shape} = hh \cdot (x_c - x_c^{gt})^2/c^2 + ww \cdot (y_c - y_c^{gt})^2/d^2。\tag{4}$$

式中:$distance^{shape}$ 表示预测框和真实框中心点之间的形状调整距离,相比传统 IoU,该公式结合了边界框的形状差异,使 Shape-IoU 能更好地拟合不规则目标;x_c 表示预测边界框的中心点的 x 坐标;y_c 表示预测边界框的中心点的 y 坐标;x_c^{gt} 表示真实边界框的中心点的 x 坐标;y_c^{gt} 表示真实边界框的中心点的 y 坐标;d 表示预测边界框中心点与真实边界框中心点的距离。

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-\omega_t})^{\theta}, \theta = 4。\tag{5}$$

$$\begin{cases}\omega_w = hh \cdot \dfrac{|w - w^{gt}|}{\max(w, w^{gt})}; \\ \omega_h = ww \cdot \dfrac{|h - h^{gt}|}{\max(h, h^{gt})}。\end{cases} \tag{6}$$

式中:Ω^{shape} 为 Shape-IoU 度量中的形状相似性项,用于表示预测边界框与真实边界框之间的形状相似性,引入形状相似性可优化边界框回归,确保预测边界框在覆盖目标的同时也与目标形状匹配;θ 为权重因子;w 和 h 分别为预测边界框的宽度和高度;ω_t 表示 Shape-IoU 计算中的权重因子,t 的取值为 w 和 h,其计算公式为式(6),式(6)通过引入边界框的宽度与高度的权重因子,进一步平衡边界框形状和尺寸的影响。

Shape-IoU 的边界框回归损失函数如式(7)所示。区别于传统 IoU 方法,Shape-IoU 在回归过程中考虑形状和尺寸因素。通过宽度和高度的比例权重,Shape-IoU 能更精确地处理不同尺度和形状的目标,提升匹配能力,并在处理形状复杂对象时展现出更强的鲁棒性和回归性能。

$$L_{\text{Shape-IoU}} = 1 - \text{IoU} + distance^{shape} + 0.5\Omega^{shape}。 \tag{7}$$

3 实验结果与分析

3.1 数据集介绍

目前尚无同时包含反光衣和安全帽检测的开源数据集,因此本文构建了一个施工现场安全衣帽穿戴检测数据集。数据集部分采用了开源安全帽数据集 SHWD(safety helmet wearing-dataset),该数据集包含佩戴安全帽和未佩戴安全帽两个类别,但不包含反光衣的检测类别。为此,本文通过网络爬虫收集施工现场工人的图像,并将其与 SHWD 数据集一起上传至 Robflow 数据标注平台,对图像数据进行清洗、标注、划分。整理后的数据集包含 7 017 张图像,标注了 3 个类别:人(person)、佩戴安全帽(helmet)和穿戴反光衣(vest),标注的数量分别为 17 148,9 465 和 7 866 个。数据集按照 8∶1∶1的比例划分为训练集、验证集和测试集,数据集示例如图 5 所示。本文所有实验均基于该自制数据集进行,以确保评估结果的统一性和可靠性。

3.2 实验平台

本实验在 Windows10 操作系统上进行,使用深度学习框架 PyTorch1.7.1 搭建网络模型,编程语言为 Python3.8。GPU 加速采用 CUDA 11.0,硬件配置为 NVIDIA GRID T4-8Q 显卡(8 GB 显存)、QEMU Virtual CPU 和 16 GB 内存。实验设置 *batch size* 为 8,训练周期为 150 个轮次,初始学习率为 0.01,动量为 0.8,权重衰减系数为 0.000 5。

图 5 施工现场安全衣帽穿戴检测数据集示例

Figure 5 Example of safety clothing and helmet wearing detection data set on construction site

3.3 评价指标

实验选取精确率 P、召回率 R、平均精度均值 mAP、帧率、模型参数量和计算量作为评价指标。平均精度均值 mAP 是所有类别的平均精度,每个类别的平均精度 AP 是该类别 P-R 曲线的面积;$mAP@0.5$ 表示在 IoU 阈值为 0.5 时的 mAP;参数量用于评估模型大小,参数量和计算量越小,说明模型越轻量;帧率表示每秒钟刷新的图像数量,可用于评估模型在检测时的效率。

3.4 改进前后模型实验结果对比

在自制的施工现场安全衣帽穿戴检测数据集上,分别使用 YOLOv7-tiny 和改进 YOLOv7-tiny 模型进行实验,对比结果如表 1 所示。从表 1 可知,改进 YOLOv7-tiny 模型的 $mAP@0.5$ 指标提升了 3.0 百分点,其中 person、helmet、vest 3 类目标的 AP 值分别提高了 5.8 百分点、1.9 百分点和 1.6 百分点,改进模型在 person 类和 helmet 类的检测中效果较为显著,因这两类目标较小,容易受到施工现场光线和遮挡的影响。结果验证了改进 YOLOv7-tiny 模型在复杂施工环境中对安全衣帽检测精度的有效提升。

表 1 YOLOv7-tiny 改进前后检测效果对比

Table 1 Comparison of detection effects before and after the improvement of YOLOv7-tiny

模型	AP/%			$mAP@0.5$/%
	person	helmet	vest	
YOLOv7-tiny	86.8	81.2	93.9	87.4
改进 YOLOv7-tiny	92.6	83.1	95.5	90.4

3.5 注意力机制对比实验

为了验证 EMA 注意力机制的优势,本文在 YOLOv7-tiny 算法基础上分别引入了空间注意力机制 GAM[23]、通道注意力机制 SE[24]、ECA[25]、混合注

意力机制 CBAM[26] 及 EMA[14] 注意力机制,并对各注意力机制进行了实验对比,实验结果如表 2 所示。由表 2 可知,除了 CBAM 导致 *mAP*@0.5 略微下降 0.1 百分点外,添加 SE、ECA、GAM 和 EMA 注意力机制分别提升了 0.1 百分点、0.4 百分点、0.6 百分点和 0.9 百分点,其中 EMA 的提升幅度最大。此外,EMA 在召回率 *R* 上比 GAM 高了 0.4 百分点,并且计算量远小于 GAM。实验结果表明,EMA 的多尺度特征融合机制能够有效提取复杂背景中的关键特征,突出目标信息,从而在施工现场等复杂场景中表现出更高的鲁棒性和检测精度。

表 2　注意力机制对比实验

Table 2　Comparative experiment of attention mechanism

模型	*R*/%	*mAP*@0.5/%	计算量/GFLOPs
YOLOv7-tiny	86.4	87.4	13.2
YOLOv7-tiny+SE	86.5	87.5	14.2
YOLOv7-tiny+CBAM	86.7	87.3	14.3
YOLOv7-tiny+ECA	86.2	87.8	14.2
YOLOv7-tiny+GAM	87.9	88.0	34.2
YOLOv7-tiny+EMA	88.3	88.3	14.2

3.6　损失函数对比实验

为了验证不同损失函数对网络性能的影响,本文在自制的安全衣帽检测数据集上,对 YOLOv7-tiny 网络分别采用了 IoU[17]、WIoU[27]、EIoU[28]、SIoU[29] 和 Shape-IoU[16] 5 种损失函数进行训练,并对实验结果进行对比分析,结果如表 3 所示。由表 3 可知,不同损失函数对网络训练结果的影响各不相同。在平均精度均值 *mAP*@0.5 指标上,相较于 IoU,WIoU、EIoU、SIoU 和 Shape-IoU 分别提高了 0.1 百分点、0.3 百分点、0.2 百分点和 0.6 百分点,其中 Shape-IoU 的提升幅度最大。在精确率 *P* 方面,WIoU 通过动态聚焦机制,优化了预测框与真实框的重叠区域,因而在边界对齐上表现更佳。

表 3　损失函数对比实验

Table 3　Comparative experiment of loss function　单位:%

损失函数	*P*	*R*	*mAP*@0.5
IoU	85.0	86.4	87.4
WIoU	86.1	85.2	87.5
EIoU	85.0	86.6	87.7
SIoU	84.3	87.0	87.6
Shape-IoU	85.3	87.1	88.0

Shape-IoU 不仅关注重叠面积,还综合考虑了边界框的形状和尺寸特征,重点提升了整体的形状匹配与回归精度,这使其在召回率 *R* 和 *mAP*@0.5 上均取得了最高值。*mAP*@0.5 作为衡量模型综合检测性能的核心指标,进一步验证了 Shape-IoU 在检测精度上的优势。图 6 为 *mAP*@0.5 指标下不同损失函数训练示意图。由图 6 可知,Shape-IoU 在训练过程中收敛更快,并表现出较高的稳定性。因此,Shape-IoU 作为改进 YOLOv7-tiny 的边界回归损失函数在施工现场安全衣帽检测任务中表现出更优的综合性能。

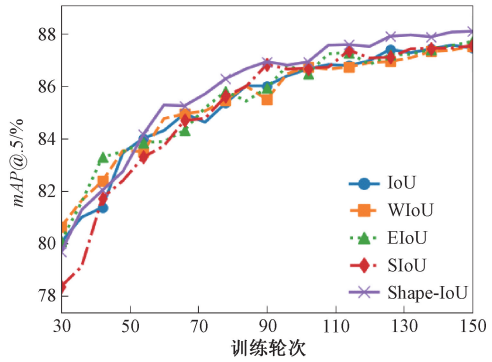


图 6　*mAP*@0.5 指标下不同损失函数训练示意图

Figure 6　Training comparison of loss functions with *mAP*@0.5

3.7　消融实验

为验证各改进点对网络的增益,本文进行了以下 7 组消融实验,结果如表 4 所示。由表 4 可知,各个改进点单独使用时均能提升各项指标。其中,RFEM 通过提高特征图感受野,在目标检测性能上提升最显著,*mAP*@0.5 提升了 1.0 百分点;EMA 则通过捕捉不同尺寸特征,使 *mAP*@0.5 提升了 0.9 百分点;Shape-IoU 损失函数优化边界框回归的精度,*mAP*@0.5 也有 0.6 百分点的提升。在组合实验中,各改进点的组合使用均显示出明显的提升效果,表明了不同改进策略间具有良好的互补性,当 3 个改进点同时应用时提升最大,精确率 *P*、召回率 *R* 和 *mAP*@0.5 分别提升了 3.7 百分点、2.3 百分点和 3.0 百分点。

表 4　消融实验

Table 4　Ablation experiments　单位:%

EMA	RFEM	Shape-IoU	*P*	*R*	*mAP*@0.5
			85.0	86.4	87.4
√			87.5	88.3	88.3
	√		85.7	87.2	88.4
		√	85.3	87.1	88.0
√	√		87.7	84.5	89.3
√		√	87.4	87.9	89.6
	√	√	86.6	87.6	89.1
√	√	√	88.7	88.7	90.4

3.8　改进 YOLOv7-tiny 与其他网络模型对比

为了验证改进 YOLOv7-tiny 的创新性和优势,本

文与多种主流目标检测算法进行了对比实验。结果如表5所示。表5展示了包括YOLOv5s、YOLOv6s、YOLOv8s、YOLOv9s、YOLOv7与本文改进YOLOv7-tiny模型在多个关键性能指标上的对比。改进YOLOv7-tiny模型在*mAP*@0.5上达到了90.4%,略低于YOLOv7的90.6%,但在实时性、模型大小和计算复杂度等方面表现更为优越。改进模型帧率为93帧/s,显著高于YOLOv5s的72帧/s和YOLOv7的42帧/s,适合动态环境下的安全衣帽穿戴的监测。在模型大小方面,改进YOLOv7-tiny参数量为6.1×10^6,低于其他模型,更适合资源受限的应用场景。在模型计算量上,改进后的YOLOv7-tiny计算量远低于YOLOv6s和YOLOv7,表明该模型在保持较高检测精度的同时,具备更低的计算复杂度。综上所述,改进YOLOv7-tiny模型在目标检测精度、实时性、模型大小及计算复杂度等多个方面均展现出优异的性能,为施工现场安全衣帽穿戴检测提供了理想的解决方案。

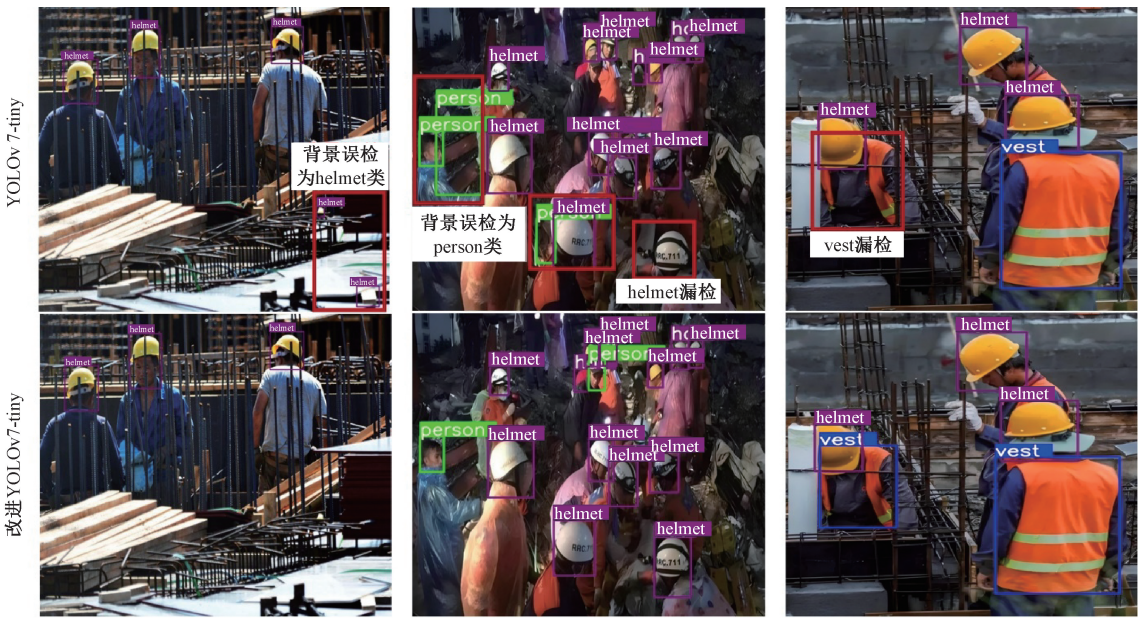
表5 不同算法性能对比

Table 5 Performance of different algorithms

算法	*mAP*@0.5/%	帧率/(帧·s^{-1})	参数量/10^6	计算量/GFLOPs
YOLOv5s	87.1	72	7.2	16.0
YOLOv6s	87.9	88	18.5	45.3
YOLOv8s	87.6	80	11.2	28.6
YOLOv9s	87.8	90	7.1	26.4
YOLOv7	90.6	42	37.2	105.2
改进YOLOv7-tiny	90.4	93	6.1	14.2

3.9 检测效果分析

图7展示了在检测框置信度阈值设定为0.25时YOLOv7-tiny与改进YOLOv7-tiny在不同施工场景下的施工人员安全衣帽穿戴检测效果对比。在图7(a)和7(b)中,由于复杂背景和弱光环境下的背景干扰强烈、目标特征信息不明显,YOLOv7-tiny检测容易产生误检和漏检(红框标注)。在图7(c)中,目标被遮挡导致特征信息不完整,进而使YOLOv7-tiny出现漏检问题(红框标注)。改进YOLOv7-tiny通过EMA注意力机制学习并融合不同层次的特征信息,增强了对目标的语义建模能力;通过RFEM模块得到更大感受野,提升了网络的鲁棒性和定位精度,有效减少了误检和漏检的问题。



(a) 复杂背景 (b) 弱光环境 (c) 目标遮挡

图7 不同环境下检测对比

Figure 7 Comparison of detection in different environments

4 结论

本文提出了一种改进YOLOv7-tiny的施工现场安全衣帽穿戴检测算法,针对当前算法在复杂背景、小目标检测抗干扰能力不足及精度差等问题进行优化。研究构建了包含person、helmet和vest 3类目标的自制数据集,结合EMA注意力机制提升特征提取能力,RFEM模块扩大网络感受野,并采用Shape-IoU边界回归损失函数优化检测精度。实验结果表明,改进模型的*mAP*@0.5提高3.0百分点,person、helmet和vest类别的*AP*值分别提高5.8百分点、1.9百分点和1.6百分点。对比实验显示,该算法

在检测精度、实时性和模型轻量化方面均优于 YOLOv5s 等主流算法,尤其适用于复杂施工环境的实时监测。未来工作将聚焦于实际施工场景及边缘设备上的模型性能评估。

参考文献:

[1] 宋健, 黄建军. 建筑工人安全参与行为演化博弈分析[J]. 安全, 2021, 42(10): 62-67.
SONG J, HUANG J J. Evolutionary game analysis on safety participation behavior of construction workers[J]. Safety & Security, 2021, 42(10): 62-67.

[2] 张震, 王晓杰, 晋志华, 等. 基于轻量化 YOLOv5 的交通标志检测[J]. 郑州大学学报(工学版), 2024, 45(2): 12-19.
ZHANG Z, WANG X J, JIN Z H, et al. Traffic sign detection based on lightweight YOLOv5[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(2): 12-19.

[3] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2015: 1440-1448.

[4] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks. [EB/OL]. (2016-05-20)[2025-06-22]. https://doi.org/10.48550/arXiv.1605.06409.

[5] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.

[6] Ultralytics. YOLOv5[EB/OL]. (2020-05-18)[2025-06-22]. https://github.com/ultralytics/yolov5.

[7] LI C Y, LI L L, GENG Y F, et al. Yolov6 v3.0: a full-scale reloading[EB/OL]. (2023-01-13)[2025-06-22]. https://arxiv.org/abs/2301.05586.

[8] 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[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway:IEEE, 2023: 7464-7475.

[9] Ultralytics. YOLOv8[EB/OL]. (2023-01-10)[2025-06-22]. https://github.com/ultralytics/ultralytics.

[10] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024-02-21)[2025-06-22]. https://arxiv.org/abs/2402.13616v2.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multiBox detector[C]// European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[12] LAW H, DENG J. CornerNet: detecting objects as paired keypoints[EB/OL]. (2024-02-21)[2025-06-22]. https://doi.org/10.48550/arXiv.1808.01244.

[13] ZHAO Q J, SHENG T, WANG Y T, et al. M2Det: a single-shot object detector based on multi-level feature pyramid network[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 9259-9266.

[14] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2023: 1-5.

[15] YU Z P, HUANG H B, CHEN W J, et al. YOLO-FaceV2: a scale and occlusion aware face detector[J]. Pattern Recognit, 2022, 155: 110714.

[16] ZHANG H, ZHANG S J. Shape-IoU: more accurate metric considering bounding box shape and scale[EB/OL]. (2023-12-29)[2025-06-22]. https://arxiv.org/abs/2312.17663.

[17] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[C]//European Conference on Computer Vision. Cham: Springer, 2018: 816-832.

[18] MA W B, GUAN Z, WANG X, et al. YOLO-FL: a target detection algorithm for reflective clothing wearing inspection[J]. Displays, 2023, 80: 102561.

[19] 赵红成, 田秀霞, 杨泽森, 等. YOLO-S: 一种新型轻量的安全帽佩戴检测模型[J]. 华东师范大学学报(自然科学版), 2021(5): 134-145.
ZHAO H C, TIAN X X, YANG Z S, et al. YOLO-S: a new lightweight helmet wearing detection model[J]. Journal of East China Normal University (Natural Science), 2021(5): 134-145.

[20] WANG L L, ZHANG X J, YANG H L. Safety helmet wearing detection model based on improved YOLO-M[J]. IEEE Access, 2023, 11: 26247-26257.

[21] PENG J, PENG F, JIN S Z, et al. Research on safety helmet wearing detection based on YOLO[C]//49th Annual Conference of the IEEE Industrial Electronics Society. Piscataway: IEEE, 2023: 1-6.

[22] MA X J, JI K F, XIONG B L, et al. Light-YOLOv4: an edge-device oriented target detection method for remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 10808-10820.

[23] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2021-12-10)[2025-06-22]. 2021: 2112.05561. https://arxiv.org/abs/2112.05561v1.

[24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vi-

sion and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.

[25] WANG Q L, WU B G, ZHU P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11531-11539.

[26] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[EB/OL]. (2018-07-11)[2025-06-22]. https://doi.org/10.48550/arXiv.1807.06521.

[27] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB/OL](2023-01-24)[2025-06-22]. https://arxiv.org/abs/2301.10051.

[28] 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.

[29] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[EB/OL]. (2022-05-25)[2025-06-22]. https://doi.org/10.48550/arXiv.2205.12740.

Improved YOLOv7-tiny for Safety Clothing and Helmet Wearing Detection Algorithm on Construction Site

ZHANG Zhen[1], ZHANG Chenwen[2], ZHANG Junjie[3], PEI Shengli[3], WANG Wenjuan[4]

(1. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. School of Henan Institute of Advanced Technology, Zhengzhou University, Zhengzhou 450001, China; 3. Henan Huirong Oil & Gas Technology Co., Ltd., Zhengzhou 450001, China; 4. School of Mechanical and Electrical Engineering, Guangdong Province Technician College of Light Industry, Guangzhou 511330, China)

Abstract: In response to the insufficient robustness of current safety clothing and helmet wearing detection algorithm in complex backgrounds, and frequent false positive rates, an improved YOLOv7-tiny detection algorithm for safety clothing and helmet wearing on construction site was proposed. Firstly, the EMA attention mechanism was introduced into the feature extraction module to enhance the networks feature extraction capability and mitigate complex background interference. Secondly, the RFEM module was integrated into the feature fusion stage to improve the networks receptive field, acquiring broader contextual information, and enhance perception for small targets. Finally, Shape-IoU was employed to replace the IoU boundary regression loss function, to improve detection accuracy. Experimental results showed that the improved model achieved a *mAP*@0.5 of 90.4% on the proprietary dataset, 3.0 percentage points higher than the original model. The frame rate speed reached 93 frames/s, and the model size comprised only 6.1 million parameters. Compared to YOLOv8s, YOLOv9s, and other models, the proposed algorithm demonstrated superior performance in detection accuracy, speed, and model efficiency, making it suitable for real-time detection applications on construction site.

Keywords: YOLOv7-tiny; attention mechanism; RFEM; Shape-IoU; safety clothing and helmet delection