



基于ODEMYOLO的水泥骨料生产车间工人安全穿戴检测模型

李鑫 胡慢谷 佟瑞鹏

A safety wear detection model for workers in cement aggregate production workshop based on ODEM-YOLO

LI Xin, HU Mangu, TONG Ruipeng

引用本文:

李鑫, 胡慢谷, 佟瑞鹏. 基于ODEM - YOLO的水泥骨料生产车间工人安全穿戴检测模型[J]. 北科大: 工程科学学报, 2025, 47(12): 2578-2588. doi: 10.13374/j.issn2095-9389.2025.03.31.003

LI Xin, HU Mangu, TONG Ruipeng. A safety wear detection model for workers in cement aggregate production workshop based on ODEM - YOLO[J]. *Chinese Journal of Engineering*, 2025, 47(12): 2578-2588. doi: 10.13374/j.issn2095-9389.2025.03.31.003

在线阅读 View online: <https://doi.org/10.13374/j.issn2095-9389.2025.03.31.003>

您可能感兴趣的其他文章

Articles you may be interested in

基于 ODEM–YOLO 的水泥骨料生产车间工人安全穿戴检测模型

李 鑫, 胡慢谷, 佟瑞鹏[✉]

中国矿业大学(北京)应急管理与安全工程学院, 北京 100083

✉通信作者, E-mail: tongrp@cumtb.edu.cn

摘 要 针对水泥骨料生产车间监控视频中工人安全穿戴检测中多尺度小目标识别困难、漏检误检频繁以及实时检测效率不高等问题, 本文提出了一种基于改进 YOLOv8 的轻量化目标检测模型 ODEM–YOLO(Omni-dimensional efficient attention and multiscale enhancement YOLO). 首先, 在 YOLOv8 基础上引入全维动态卷积 (ODConv) 模块, 增强浅层特征提取能力, 有效捕获小目标的关键特征; 其次, 结合改进的高效多尺度注意力机制 (iEMA) 优化 Neck 网络, 有效提高多尺度目标的特征表达能力; 同时, 提出 C2f 多尺度边缘信息增强 (C2f_MSEIE) 模块显式增强目标边缘信息, 提高对安全装备边界特征的识别精度. 实验基于实际水泥骨料车间监控数据, 构建了包含 9877 个多尺度小目标样本的数据集, 开展模型性能评估. 实验结果表明, ODEM–YOLO 模型在保持结构轻量化 (6.9 MB) 的同时, 整体检测精度 (mAP@0.5) 达到 0.896, 小目标 (口罩) 检测精度 (AP@0.5mask) 达到 0.746, 单图推理时间达 8.2 ms, 优于 YOLOv5n、YOLOv10n 等主流模型. 并且在 NVIDIA Jetson Nano B01 嵌入式设备上实际部署测试达到 25 frame·s⁻¹ 的实时检测效果, 充分满足工业现场实时安全监控需求.

关键词 安全穿戴检测; 小目标检测; 多尺度检测; YOLO; 高效多尺度注意力 (EMA); 全维动态卷积 (ODConv)

分类号 TP183;TD76;TP391.41

A safety wear detection model for workers in cement aggregate production workshop based on ODEM–YOLO

LI Xin, HU Mang, TONG Ruipeng[✉]

School of Emergency Management and Safety Engineering, China university of mining and technology-Beijing, Beijing 100083, China

✉Corresponding author, E-mail: tongrp@cumtb.edu.cn

ABSTRACT Safety compliance, particularly the correct usage of personal protective equipment (PPE), is critical in high-risk industrial settings such as cement aggregate production workshops, where traditional manual supervision is often insufficient owing to harsh conditions and operational dynamics. While artificial intelligence-driven video surveillance offers a promising solution, existing object detection models frequently struggle with accurately identifying small and multiscale targets, leading to high error rates and limited practical effectiveness. To address these limitations, this paper introduces ODEM–YOLO, a novel, lightweight yet highly accurate object detection model based on an enhanced YOLOv8 architecture, specifically engineered for robust safety wear detection. Methodologically, ODEM–YOLO incorporates several key innovations. First, the omni-dimensional dynamic convolution (ODConv) module is integrated into the early backbone stages. Unlike standard convolutions with fixed kernels, ODConv employs a multidimensional attention mechanism to dynamically learn kernel weights across spatial, input channel, output channel, and kernel number dimensions, enabling adaptive focus on salient features of small targets in complex scenes and enhancing shallow-level feature map discrimination. Second,

收稿日期: 2025–03–31

基金项目: 国家自然科学基金资助项目 (52074302); 北京市自然科学基金资助项目 (8212015)

the Neck network is optimized with an improved efficient multiscale attention (iEMA) mechanism, centered around an inverted residual mobile block core. This module strategically uses 1×1 pointwise convolutions for channel manipulation and 3×3 depth-wise separable convolutions for efficient spatial feature learning, allowing effective capture and fusion of multiscale contextual information with significantly reduced computational complexity to improve the representation of diverse PPE sizes. Third, a novel C2f multi-scale edge information enhancement (C2f_MSEIE) module replaces original C2f blocks, explicitly enhancing target edge information for clearer boundary definition. It comprises a local convolution branch for preserving fine-grained details and a multiscale edge modeling branch that utilizes AdaptiveAvgPool2d with multiple bin sizes and an innovative Edge Enhancer submodule to extract and reinforce high-frequency edge features, providing a more robust understanding of object contours for precise localization. The efficacy of ODEM-YOLO was rigorously validated on a custom dataset of 9877 images from actual cement aggregate workshops, featuring diverse small and multiscale targets under realistic and challenging conditions. The experimental results demonstrate ODEM-YOLO's superior performance, achieving an overall mean average precision (mAP@0.5) of 0.896 and an AP@0.5mask (for the challenging small "mask" objects) of 0.746. Despite these significant accuracy gains, the model maintains a compact size of only 6.9 MB and achieves a rapid single-image processing time of 8.2 ms (utilizing 9.5 GFLOPs), outperforming other mainstream lightweight models such as YOLOv5n and YOLOv10n. Ablation studies systematically confirmed the individual and synergistic contributions of the ODConv, iEMA, and C2f_MSEIE modules to the overall performance improvement. Furthermore, practical deployment on an NVIDIA Jetson Nano B01 embedded device demonstrated ODEM-YOLO's capability of real-time detection at 25 frames per second, fully satisfying the demanding requirements of industrial on-site safety monitoring. In conclusion, ODEM-YOLO presents a highly effective and efficient solution for real-time safety wear detection in challenging industrial environments. Its architectural innovations specifically target the difficulties of small and multiscale object detection, leading to substantial improvements in accuracy and reliability while preserving a lightweight structure crucial for edge deployment. ODEM-YOLO is a valuable and practical tool for enhancing occupational safety and potentially reducing accident rates.

KEY WORDS safety wearable detection; small object detection; multiscale object detection; YOLO; efficient multi-scale attention (EMA); omni-dimensional dynamic convolution (ODConv)

近年来, 矿山及水泥骨料生产等高危作业环境中, 不正确佩戴可穿戴防护设备等不安全行为仍然普遍存在, 导致事故发生率居高不下. 究其原因, 一方面是部分工人安全意识淡薄或图一时方便而缺乏主动佩戴意愿^[1]; 另一方面, 作业现场环境恶劣, 粉尘和碎石飞溅等因素也使得管理人员难以及时、有效地进行监督. 为减少安全隐患, 业界普遍采用基于 AI 视频智能分析的技术手段, 对工人安全帽、口罩、反光衣等防护用品的佩戴情况进行检测与识别^[2-3]. 然而, 现有检测算法仍面临检测准确率偏低、误报和漏报率较高的问题. 其原因主要包括: 监控摄像头取景范围较大、拍摄距离远, 小目标在图像中所占像素较少、特征表述不明显^[4]; 同时, 安全帽、反光衣和口罩等目标在图像中的尺度差异较大, 算法更容易捕捉特征明显的较大目标, 导致小目标检测精度不足^[5]. 因此, 亟需研究更有效的工人安全穿戴检测算法, 提高检测准确率, 以便在现场作业中及时发现并干预工人的不安全行为, 从而有效降低事故发生率并切实保障工人的人身与生产安全.

上述问题可归类为小目标检测 (Small object

detection, SOD) 与跨尺度检测问题, 是目标检测领域的热点和难点. MS COCO (Microsoft common objects in context) 数据集将小目标定义为小于 32×32 像素的目标, 跨尺度问题定义为各对象的大小相对于图像的比例几十倍以上的跨度^[6]. 主流的目标检测器, 如 YOLO (You only look once)^[7]、SSD (Single shot multi box detector)^[8] 等普遍以特征金字塔网络 (Feature pyramid network, FPN)^[9] 的多尺度特征融合思想设计, 利用自顶向下的架构, 从高分辨率的浅层特征开始, 逐渐向下到低分辨率的深层特征, 使网络能够捕获到图像的多尺度信息, 但特征也在网络逐步加深的过程中越来越抽象, 丢失了较小目标的特征, 使本就难以检测的小目标更难检测. 不同尺度的目标对检测算法中的交并比计算等操作敏感性不同, 小目标由于尺寸小, 更易受到检测阈值的影响, 从而难以被准确识别, 最终导致精度不均衡的现象^[10].

国内外学者对多尺度目标中混合小目标的检测任务提出许多解决办法. Cheng 等^[11]总结了现有基于深度学习的小目标检测方法, 即面向样本方法、规模感知方法、基于注意力方法、特征模仿

方法、上下文建模方法、聚焦检测方法, 这些方法的应用可以解决大部分场景中的小目标检测问题. Yuan 等^[12]提出了两段式检测框架 CFINet, 包括 FI 特征模仿分支和 CRPN 动态锚点策略, 并集成到 Faster R-CNN 算法中, 得到专为小目标检测设计的检测网络. 王兵等^[13]针对 YOLO 轻量网络在口罩佩戴检测任务中特征提取不足且利用率低的问题, 增加了 Max Module 结构, 改进了 YOLOv4-tiny 轻量化网络算法, 提高了特征利用率. 田枫等^[14]构建了油田场景下 YOLO-people 与 YOLO-dress 级联的小目标检测模型, 先定位行人目标, 然后裁剪出行人区域并进行尺度变换, 最后对行人进行安全穿戴检测. 针对交通标志检测中存在的误检、漏检问题, Zhang 等^[15]基于 YOLOv8 算法展开了针对性改进, 通过引入高分辨率小目标检测层, 并构建双向跨层级特征交互机制, 优化了特征金字塔网络 (BiFPN) 的多尺度融合能力, 显著提升了对微小交通标志的定位精度; Wei 等^[16]则在 YOLOv8 主干网络中嵌入改进型卷积块注意力模块 (FCBAM), 通过融合空间位置感知增强了复杂光照环境下小目标的特征判别性. Shen 等^[17]通过多尺度注意力混合头、多通道增强颈和星形融合模块等改进, 提出了一种基于改进 YOLOv8 的轻量级模型 MSA-YOLO (Multi-scale adaptive YOLO), 显著提升了小目标检测和分割的准确性和效率. 然而, 现有方法对密集小目标及多尺度特征的适应性仍有待提升.

综上, 水泥骨料生产车间工人安全穿戴检测面

临着小目标和多尺度检测等难题, 现有的检测算法存在检测精度低、漏检误检率高和实时性差等问题. 为此, 本文针对车间的实际工况, 提出了一种基于改进 YOLOv8 基线模型的工人安全穿戴检测算法 ODEM-YOLO (Omni-dimensional efficient attention and multiscale enhancement YOLO), 引入 ODConv (Omni-dimensional dynamic convolution)^[18]提升网络的特征提取能力, 引入 EMA (Efficient multi-scale attention)^[19]提升特征表示能力, 并利用倒置残差移动块^[20-21]改善 EMA 带来的后处理效率问题, 同时, 提出 C2f_MSEIE (C2f multi-scale edge information enhancement) 模块以增强边缘结构感知能力. 使用自制数据集验证工人安全穿戴检测的精度、推理速度、后处理速度等指标.

1 安全穿戴检测算法

本文提出的 ODEM-YOLO 检测算法以 YOLOv8 模型为基线, 整体结构由主干网络 (Backbone)、颈部网络 (Neck) 和头部网络 (Head) 三部分组成, 如图 1 所示. 其中, 主干网络借鉴了 YOLOv8 的 CSP (Cross stage partial) 设计思想, 主要由 OD-CBS (ODConv + BN + SiLU)、CBS 和 SPPF 模块构成; 颈部网络采用融合特征金字塔的路径聚合网络 Path aggregation network (PAN) 结构, 并在其中引入注意力模块 iEMA (Improved efficient multiscale attention); 检测头部沿用 YOLOv8 的设计, 用于输出各目标类别的定位和置信度. 此外, 本文还将所

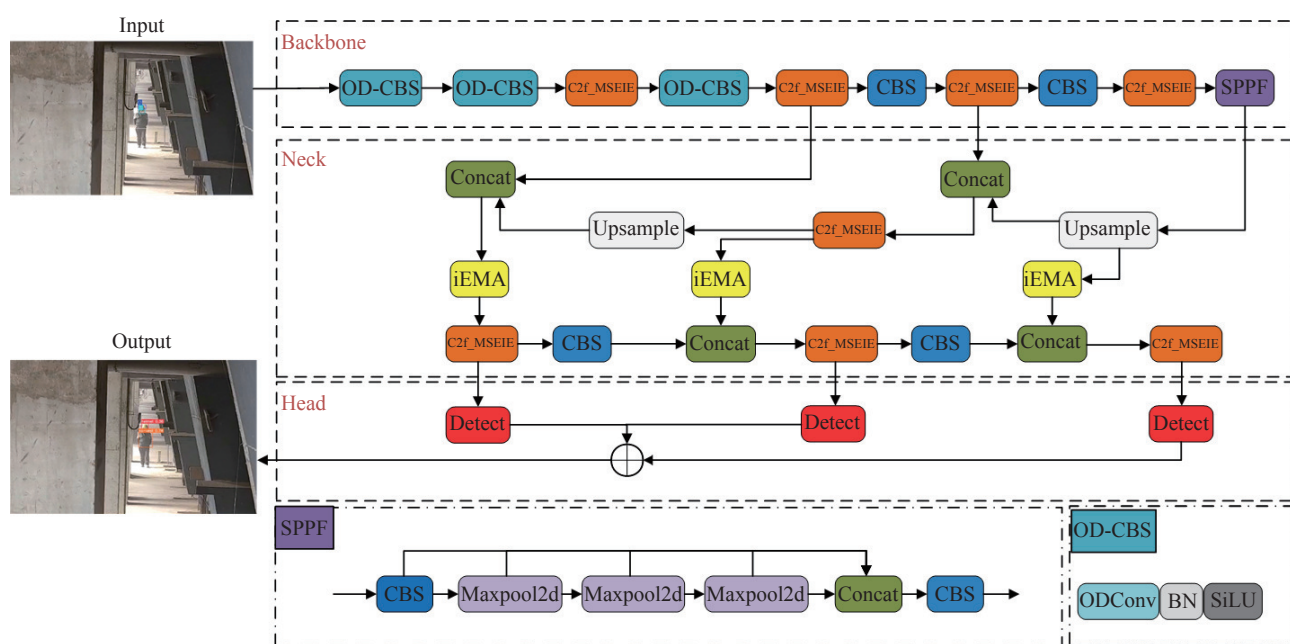


图 1 ODEM-YOLO 结构图

Fig.1 Structure of ODEM-YOLO

有 C2f 模块替换为设计的 C2f_MSEIE 模块, 以进一步增强网络对多尺度目标边缘信息的捕获能力. 以下分别介绍本模型中三个关键改进模块: ODConv 卷积单元、iEMA 注意力机制和 C2f_MSEIE 模块.

1.1 ODConv 特征提取卷积核

传统卷积在提取特征时通常采用固定的卷积核参数, 无法根据输入图像特征的变化自适应地调整卷积核权重, 导致网络难以有效应对工业场景中复杂背景、多尺度以及小目标检测任务的特征多样性问题. 本文引入了 ODConv 模块, 以增强模型在浅层特征提取阶段对不同尺度尤其是小目标的特征捕获能力. ODConv 以注意力机制动态调整多个并行卷积核的权重, 各卷积输出与对应的注意力权重相乘后, 再将所有加权输出求和形成最终的卷积结果. 其中, 并行的卷积核共享内核尺寸和输入/输出维度. 注意力机制则根据输入动态计算每个卷积核的注意力权重, 以表示各卷积核对输出的贡献程度, 如公式 (1) 所示.

$$y = (\alpha_{wi} \odot \alpha_{fi} \odot \alpha_{ci} \odot \alpha_{si} \odot W_i + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x \quad (1)$$

式中, * 代表的是标准的卷积操作, W_i 表示第 i 个卷积核; α_{wi} 是 W_i 核数量维度的注意力权重, 它决定了每个卷积核在聚合过程中的贡献程度; α_{fi} 是输出通道维度的注意力权重, 使模型能够调节不同输出特征图的重要性; α_{ci} 是输入通道维度的注意力权重, 强调对最终输出更为重要的通道; α_{si} 是空间维度的注意力权重, 对卷积核的每个空间位置学习一个不同的注意力权重, 即为不同位置的卷积操作分配不同的重要性.

ODConv 的多维注意力权重由多头注意力模块 $\pi(x)$ 生成, 包括全局平均池化 (GAP)、全连接层 (FC) 和激活函数过程, 以精细化卷积操作的特征选择. 每个卷积核生成的特征图与注意力权重相乘后有效增强或弱化不同区域, 突出任务相关特征. 最后, 将所有加权特征图求和, 得到最终输出特征图.

1.2 iEMA 特征增强模块

在水泥骨料生产车间的安全穿戴检测任务中, 由于目标尺度差异较大, 尤其是口罩等小目标的特征难以捕获, 传统卷积网络结构在应对此类多尺度检测问题时效果欠佳. 为更好地突出不同尺度目标的关键信息, 本文引入了一种高效的改进注意力模块 iEMA. 该模块在 EMA 模块的基础上, 以倒置残差移动块 (Inverted residual mobile block, iRMB) 为核心进行结构优化. iEMA 包括

1×1 逐点卷积升维、 3×3 深度可分离卷积 (DWConv) 以及 1×1 逐点卷积降维过程, 能够有效实现特征的融合与重用. 与传统 EMA 结构不同, 改进后的 iEMA 模块去除了复杂的特征映射和矩阵相乘等操作, 使网络的参数量与计算复杂度显著降低. iEMA 模块被嵌入到 YOLOv8 网络的 Neck 部分 (C2f 模块之后), 用于特征融合阶段对不同尺度特征进行有效的注意力加权. 通过这种结构优化, 网络能够更高效地捕获工业场景中小尺度目标 (如未佩戴口罩的人脸区域) 的关键细节信息, 显著提升了整体检测精度与实时性能.

1.3 C2f_MSEIE 模块

为提升 YOLOv8 模型在水泥骨料生产车间工人安全穿戴检测任务中的边缘结构感知能力, 本文设计并提出了一种新型的特征增强结构 MSEIE 模块 (Multi-scale edge information enhancement, 多尺度边缘信息增强). 该模块旨在通过多尺度上下文建模与边缘显式增强机制, 提升模型对细粒度小目标 (如口罩、安全帽、反光背心等) 的判别能力. 将 MSEIE 模块作为 C2f 模块中的核心单元, 替代原始由多个 Bottleneck 模块组成的堆叠结构, 构成改进后的 C2f_MSEIE 模块, 整体结构如图 2 所示.

原始 C2f 模块中的 Bottleneck 主要通过浅层残差与卷积叠加方式进行局部特征压缩与传播, 但其对结构边缘与高频细节的建模能力不足. 考虑到工人佩戴状态常伴随轮廓边界的断裂 (如“未佩戴口罩”、“未穿反光背心”), 提出的 MSEIE 模块从设计上引入多尺度池化路径, 结合高频边缘残差信号建模机制, 使网络能够对小尺度区域中变化明显的边界区域进行显式增强.

在结构上, MSEIE 模块由两部分组成: 局部卷积分支和多尺度边缘建模分支. 局部卷积分支通过一个 3×3 卷积提取输入特征的局部细节信息, 保留了精细的纹理和边缘细节. 多尺度边缘建模分支则侧重引入不同空间尺度的上下文信息和边缘特征增强. 设计的 C2f_MSEIE 模块中, MSEIE 子模块首先采用多尺度的自适应平均池化操作 (AdaptiveAvgPool2d) 对输入特征图进行分块降维, 以获取不同空间尺度下的上下文特征信息. 该多尺度分支设置三个 bin 尺寸: bin1、bin2、bin3, 分别对应池化输出尺寸为 1×1 、 2×2 和 4×4 , 从而分别提取全局、区域和局部层级的语义特征. 其中, bin1 的 1×1 输出可视为对整幅特征图的全局压缩, 相当于感受原图约 16×24 的区域信息; bin2 (2×2) 特征图捕捉较粗粒度的区域上下文, 其每个单

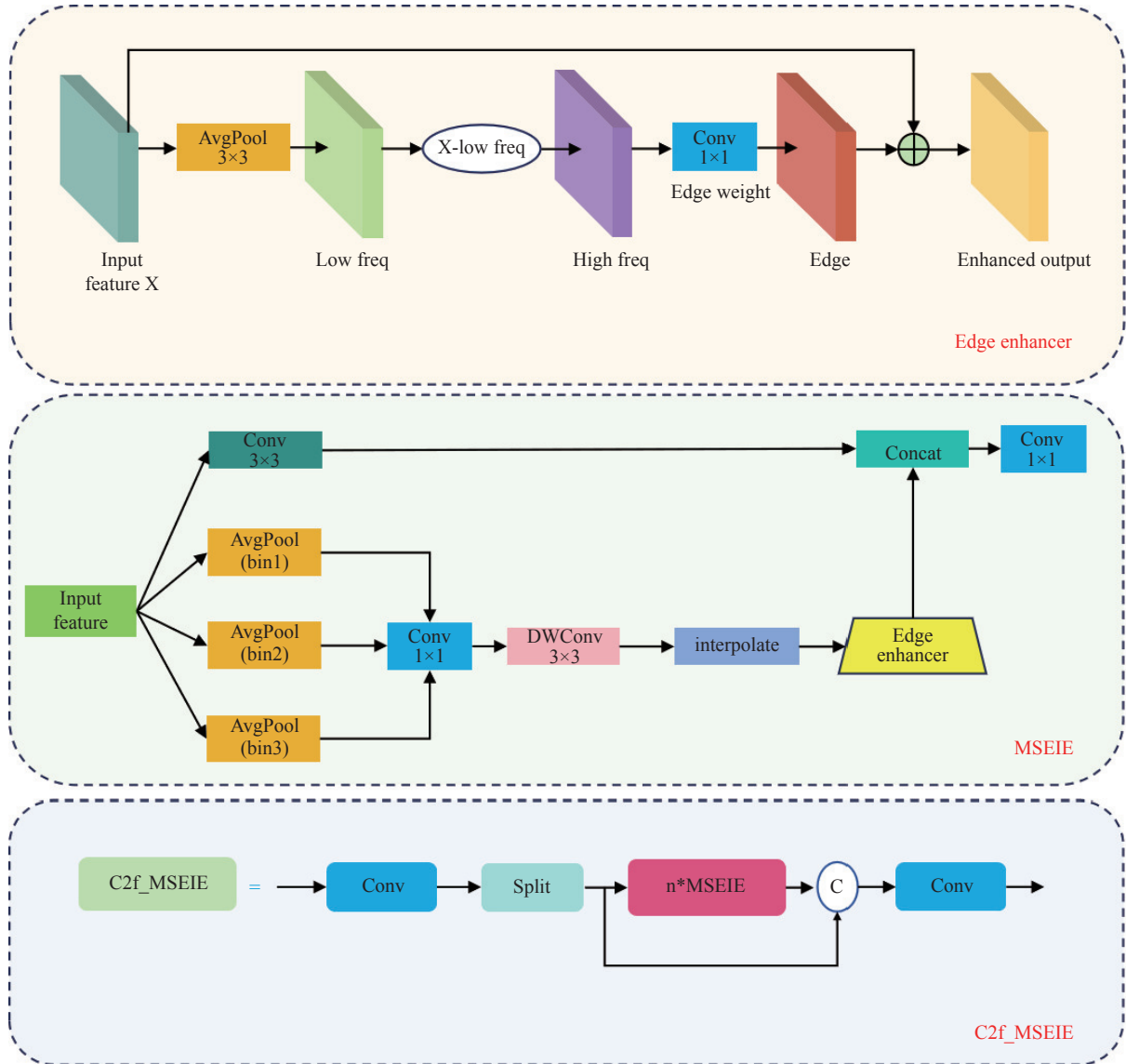


图 2 C2f_MSEIE 模块结构

Fig.2 Structure of C2f_MSEIE

元感受原图约 8×12 区域; bin3(4×4) 提供更高分辨率的空间特征, 每个单元对应原图约 4×6 的区域. 随后, 每一路特征先经过一个 1×1 卷积进行通道降维, 减少冗余并降低计算量; 再通过一个 3×3 深度可分离卷积(DWConv 3×3) 在低分辨率特征图上进行空间建模, 提取局部结构信息.

完成卷积建模后, 将每一路特征图通过双线性插值上采样至与原始特征图相同的空间尺寸, 以便与其他分支进行特征融合. 接下来, 引入专门设计的 Edge enhancer 子模块对各尺度路径进行边缘提取与增强处理. Edge enhancer 模块首先对输入特征图进行 3×3 的平均池化, 提取低频特征(Low

freq), 然后将原始输入特征图减去低频图, 获取包含边缘与高频细节的残差图(High freq). 再通过 1×1 卷积、批归一化(BN)以及 Sigmoid 激活, 将高频图转化为边缘权重图(Edge weight), 最后以残差相加的方式与原输入特征融合, 生成显式增强边缘细节的输出特征(Enhanced output). 这一机制使网络具备对细微边界区域更为敏感的判别能力.

最后, MSEIE 模块将局部卷积分支的输出特征与三条多尺度边缘增强分支结果在通道维度进行拼接(Concat)融合, 通过 1×1 卷积进行通道维度的压缩与特征融合, 形成最终融合了边缘信息的输出特征图. 整体计算过程可表达如公式 (2):

$$\hat{X} = \text{Conv}_{1 \times 1} \left(\text{Concat} \left(\text{Conv}_{3 \times 3}(X), \sum_{i=1}^N \text{EE}_i(\text{Upsample}(F_i(X))) \right) \right) \quad (2)$$

式中, Conv 表示卷积操作, 是一种通过特定大小的滑动滤波器(如 3×3 或 1×1)对输入特征图进行加权求和的数学运算; Upsample 表示上采样操作, 是一种将输入特征图的尺寸(高度和宽度)进行放大的过程. X 表示输入特征图, $F_i(X)$ 表示第 i 个尺度池化路径处理结果, EE_i 为对应的 EdgeEnhancer 模块, Concat 表示通道拼接操作, 最终通过 1×1 卷积进行维度融合.

2 数据集制备与实验设置

2.1 实验设置

实验基于 PyTorch2.0.0 框架, 软硬件环境配置如表 1 所示. 输入图像输入尺寸为 720×720 像素, 训练迭代次数为 100, 每个批次大小为 16, 采用随机梯度下降算法优化模型, 学习率初值设置为 0.01, 动量设置为 0.937, 权重衰减因子设置为 0.0005.

表 1 软硬件平台配置

Table 1 Software and hardware platform configurations

Setup	Version parameter
System	Ubuntu20.04
GPU	NVIDIA RTX 3090 (24 GB)
CPU	14 vCPU Intel(R) Xeon(R) Gold 6330 CPU @ 2.00GHz
CUDA	11.8
Python	3.8.10

2.2 实验数据集

为了训练和检验 ODEM-YOLO 算法, 提取了水泥骨料各生产车间不同时段和不同机位的监控视频中的 3500 余张图片, 制备了口罩、安全帽和反光衣等类别的数据样本 9877 个, 如表 2 所示. 其中, Vest、No-vest、Helmet、Head、Mask、No-mask 分别为有无穿戴反光衣、有无穿戴安全帽和有无佩戴口罩的样本标签. 并以 7:3 的比例随机划分训练集和测试集.

表 2 实验数据集构成

Table 2 The composition of the experimental dataset

Category	No-vest	Vest	Helmet	Head	Mask	No-mask	Total
Quantity	2407	1100	2512	1022	1806	1030	9877

2.3 评价指标

为评价所提出的 ODEM-YOLO 算法性能, 本实验采用平均精度 $mAP@0.5$ 来评估算法对安全帽、反光衣和口罩等多目标的整体检测精度, 采用 $AP@0.5mask$ 评估算法模型对多尺度目标检测

中口罩小目标的检测能力, 采用权重大小、推理速度、后处理速度和 GFLOPs 等评价算法的实时性. $mAP@0.5$ 是指模型在 50% 交并比阈值下对所有类别的平均精度 (Average precision, AP) 取平均值得到的指标. $AP@0.5mask$ 是指模型在 50% 交并比阈值下口罩目标的平均检测精度, 即模型对口罩目标的 Precision-recall 曲线下的面积. 所用计算公式为:

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{\sum_{i=1}^K AP_i}{K} \quad (4)$$

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

式中, TP 为模型预测为正类的正样本数, TN 为模型预测为负类的负样本数, FP 为模型预测为正类的负样本数, FN 为模型预测为负类的正样本数.

3 实验结果与分析

3.1 不同注意力机制对比实验

为验证 iEMA 在多尺度特征提取和后处理改进的有效性, 以 YOLOv8 为基线模型, 在相同训练参数、相同添加位置的情况下开展不同注意力机制的对比实验. 选取了 SE(Squeeze-and-excitation)^[22]、CBAM(Convolutional block attention module)^[23]、CA(Coordinate attention)^[24] 等主流注意力机制对基线模型进行了改进, 实验结果如表 3 所示.

表 3 不同注意力机制的对比实验结果

Table 3 Comparative experimental results of different attentional mechanisms

Attention	$mAP@0.5$	$AP@0.5mask$	Post-processing speed/ms	Weight size/MB
Baseline	0.809	0.568	2.3	8.99
+SE	0.837	0.505	8.2	8.50
+CBAM	0.836	0.620	6.2	11.4
+CA	0.824	0.656	6.8	11.5
+EMA	0.827	0.716	5.4	8.6
+iEMA	0.828	0.713	2.8	8.8

结果表明, SE 的引入优化了模型结构, 在降低了权重大小的同时提升了模型的检测精度, 但拖慢了后处理速度, 实时性变差, $AP@0.5mask$ 表现出降低, 小目标检测能力变差. CBAM 因其在空间

维度上对特征图加权, 强调重要的空间位置, 使得 AP@0.5mask 提升 9.2%, 说明空间位置的加强利于小目标的定位. CA 将特征的空间坐标信息集成到通道注意力中, 进一步提升了网络对空间特征的感知能力, 虽然 mAP@0.5 表现不如 CBAM, 但 AP@0.5mask 有 15% 的提升, 表明 CA 具有良好的小目标检测能力. EMA 在提升多尺度目标特征表示上效果显著, AP@0.5mask 达到了 0.716, 且权重大小有少量缩减, 但同样拖慢了模型的后处理速度, 实时性差. iEMA 在保持 EMA 检测性能的同时, 后处理速度提升了近 1 倍, 表明引入 iRMB 结构增强了模型的特征重用能力, 减少了后处理阶段的计算量, 大大提升了算法的实时性.

3.2 不同卷积对比试验

为验证 ODConv 在生产车间工人安全穿戴检测中的有效性, 将骨干网络首个标准卷积分别替换为 DWConv、GhostConv、GSConv、SPDConv 和 ODConv, 并在相同的训练与推理条件下进行对比. 表 4 中的实验结果表明, 在参数量几乎不变(3.162M 对 3.157M)的情况下, ODConv 将 AP@0.5mask 从 0.568 提升到 0.631, 将 mAP@0.5 从 0.809 提升至 0.818, 性能优于所有对比方法. DWConv 的计算量最低, 但因为深度可分卷积削弱跨通道信息耦合, 精度反而下降. GhostConv 和 GSConv 依靠冗余特征重用或空间通道重排提高效率, 但难以针对口罩等小目标自适应捕获关键信息, 性能仅与基线持平或略高. SPDConv 通过扩展空间维度将精度提升, 相较于 Baseline 却增加 1.6 GFLOPs 的算力代价, 性价比较低. ODConv 在空间、输入输出通道和卷积核维度引入动态注意力机制, 可对浅层特征进行精细重加权; 对口罩这类小目标能增强局部纹理, 对尺度变化大的区域能自适应调整感受野, 因此在精度与速度之间取得最佳平衡, 展示出在工业安全监测中的应用潜力.

3.3 消融实验

根据表 5 的 ODEM-YOLO 消融实验结果可

表 4 不同卷积的对比实验结果

Method	Parameters	GFLOPs	AP@0.5mask	mAP@0.5
Baseline	3157184	8.9	0.568	0.809
+DWConv	3135008	8.4	0.560	0.801
+GhostConv	3146848	8.6	0.562	0.806
+GSConv	3146848	8.6	0.580	0.811
SPDConv	3227600	10.5	0.600	0.815
+ODConv	3162353	9.2	0.631	0.818

知, 其中①、②、③分别代表 ODConv、iEMA、C2f_MSEIE 模块, 各模块单独应用时均有效改善了基线模型的检测能力. 具体而言, 在 Backbone 引入 ODConv 模块后, 通过动态卷积机制增强了网络对复杂背景下目标特征的捕获能力, 使整体检测精度 mAP@0.5 提升至 0.818. 在 Neck 部分使用 iEMA 模块, 通过迭代特征融合减少了特征信息的损失, 进一步将 mAP@0.5 提升至 0.828, 召回率也显著提高至 0.799, 在小目标检测方面, 口罩类别的平均精度 (AP@0.5mask) 提高至 0.713. C2f_MSEIE 模块通过有效融合多尺度特征并增强细节表达能力, 使整体 mAP@0.5 达到 0.854, 召回率与精确率均得到明显提升, 小目标定位表现显著改善, AP@0.5mask 提升至 0.729, 进一步体现出该模块对安全穿戴检测任务的有效性. 当同时应用 ODConv、iEMA 和 C2f_MSEIE 三个模块时, 模型的整体检测性能达到最佳, mAP@0.5 提高至 0.896, 召回率达到 0.861, 精确率达到 0.927, 且针对口罩类小目标的检测精度 (AP@0.5mask) 显著提高至 0.746. 这体现了三个模块在特征捕获、特征融合及特征重用方面的优势互补, 全面提升了模型在多尺度目标检测任务中的表现.

表 5 ODEM-YOLO 消融实验结果

Table 5 Results of ODEM-YOLO ablation experiment

Model structure			Precision	Recall	AP@0.5mask	mAP@0.5
①	②	③				
—	—	—	0.869	0.766	0.568	0.809
√	—	—	0.886	0.782	0.631	0.818
—	√	—	0.875	0.799	0.713	0.828
—	—	√	0.917	0.835	0.729	0.854
√	√	—	0.925	0.832	0.722	0.868
√	√	√	0.927	0.861	0.746	0.896

3.4 不同检测模型对比实验

为体现本文算法优势, 在相同的数据集和训练参数下, 对比分析了 R-CNN^[25-26] 系列与 YOLO^[27-29] 系列主流目标检测模型在小目标多尺度检测任务中的性能表现, 实验结果如表 6 所示. 由表 6 可知, RT-DETR-L^[30] 在整体检测精度 (mAP@0.5) 和小目标检测精度 (AP@0.5mask) 方面表现突出, 分别达到 0.884 和 0.750, 但模型的权重 (112.7 MB) 和计算复杂度 (126.6 GFLOPs) 最高, 推理速度也较慢 (单张图像 16.9 ms), 因此更适用于对检测精度要求较高但对实时性要求不严格且计算资源充裕的场景. R-CNN 系列的代表模型 Faster R-CNN 整体精度较

高 ($mAP@0.5$ 为 0.857, $AP@0.5mask$ 为 0.707), 但其权重 (166.0 MB) 和计算量 (215.4 GFLOPs) 明显偏大, 推理速度最慢 (单张图像 21.5 ms), 限制了在实际生产环境中的实时应用. YOLO 系列随着版本更新呈现出明显的轻量化与高效化趋势, 性能逐步提高. 其中 YOLOv10n 表现尤为出色, 得益于无需非极大值抑制 (NMS) 的端到端预测机制, 其整体精度 ($mAP@0.5$) 达到 0.836, 小目标检测精度

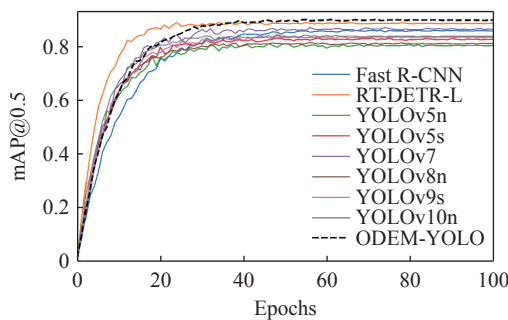
($AP@0.5mask$) 为 0.620, 推理速度极快 (单张图像 7.1 ms), 模型体积仅为 6.3 MB, 适合实时部署. 本文提出的 ODEM-YOLO 模型整体性能最佳, 其 $mAP@0.5$ 达到 0.896, 小目标检测 ($AP@0.5mask$) 也提升至 0.746, 显著优于所有 YOLO 系列模型. 同时, 该模型维持了较小的权重 (6.9 MB)、适中的计算量 (9.5 GFLOPs) 和较快的推理速度 (单张图像 8.2 ms), 在性能与实时性之间实现了理想的平衡.

表 6 不同目标检测模型实验结果对比

Table 6 Comparison of experimental results of different target detection models

Model	$mAP@0.5$	$AP@0.5mask$	Weight size/MB	GFLOPs	Inference speed/ms
YOLOv5n	0.802	0.545	3.9	4.5	6.1
YOLOv5s	0.825	0.580	13.8	15.8	8.8
YOLOv7	0.863	0.690	37.5	105	12.5
YOLOv8n	0.809	0.568	6.2	8.9	7.9
YOLOv9s	0.833	0.602	20.1	26.4	11.7
YOLOv10n	0.836	0.620	6.3	7.9	7.1
RT-DETR-L	0.884	0.750	112.7	126.6	16.9
Faster R-CNN	0.857	0.707	166.0	215.4	21.5
ODEM-YOLO(Ours)	0.896	0.746	6.9	9.5	8.2

为对比不同模型的 $mAP@0.5$ 随训练的收敛情况, 绘制了其检测精度 $mAP@0.5$ 随训练轮数 (Epoch) 的变化曲线, 如图 3 所示. 对比发现, ODEM-YOLO 在特征学习速度方面显著增强, $mAP@0.5$ 在 30 轮时已与其他 YOLO 模型有明显差异, 表明其在训练过程中以更高的效率学习到特征表示.

图 3 不同模型 $mAP@0.5$ 随迭代轮次的变化曲线Fig.3 Curve of $mAP@0.5$ with iteration rounds for different models

为验证 ODEM-YOLO 算法在实际场景中的性能, 从测试集中抽取典型图像开展安全穿戴检测实验, 结果如图 4 所示. 图 4 直观地展示了各车间典型场景下对工人实际安全穿戴状态的识别效果. 从识别结果来看, 该算法在实际应用场景中的预测效果与训练阶段的表现基本一致. 除口罩 (Mask)

等小尺寸目标的检测置信度稍低外, ODEM-YOLO 算法对各尺度目标在 50% 交并比阈值下的置信度均达到了 0.5 以上, 且对不同尺度目标的检测精度保持较好的均衡性. 这表明 ODEM-YOLO 算法在真实水泥骨料生产车间环境中具备较高的准确性和稳定性, 满足实际现场部署需求.

3.5 检测结果及不足情形分析

ODEM-YOLO 对各检测类别的结果如表 7 所示, 安全帽 (Helmet) 和反光衣 (Vest) 类别的精确率和召回率较高, 表现较为理想, 其中安全帽的精确率为 95.1%、召回率为 96.2%, 反光衣的精确率为 98.6%、召回率为 99.0%. 然而, 口罩 (Mask) 的精确率为 85.8%、召回率仅为 60.4%, 低于其他类别, 表明该类别的检测准确性和召回能力存在明显不足, 需针对性优化, 以提升模型整体性能.

尽管 ODEM-YOLO 模型在大多数工业场景下表现出良好的检测效果, 但在实际应用中仍存在部分复杂情形会导致目标漏检或误检. 为系统识别这些潜在问题并推动模型进一步优化, 本文选取若干具有代表性的检测失败案例进行深入分析. 如图 5 所示, 图 5(a) ~ 5(c) 三张图片为真实标注, 其中数字 0、1、2、5 分别表示安全帽 (Helmet)、口罩 (Mask)、未穿反光衣 (No-vest) 和穿反光衣

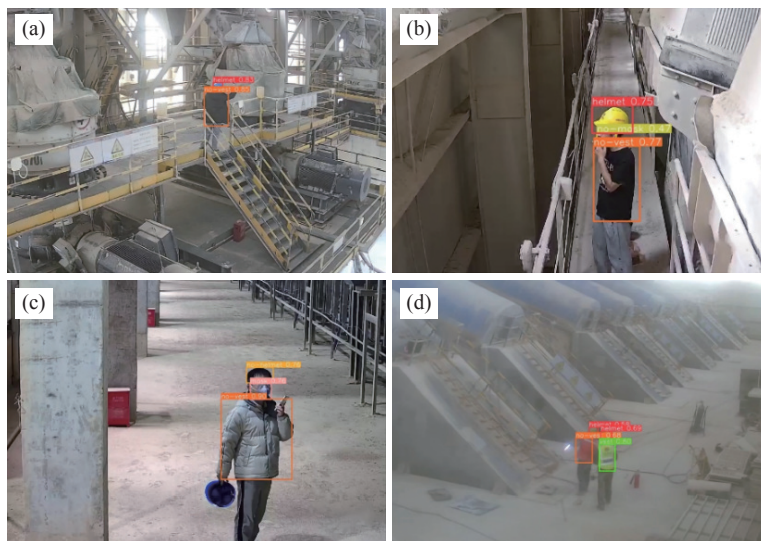


图 4 典型检测场景中工人实际穿戴状态的识别效果. (a) 检测场景 1; (b) 检测场景 2; (c) 检测场景 3; (d) 检测场景 4

Fig.4 Detection results of workers' actual safety gear status in typical monitoring scenarios: (a) detection scenario 1; (b) detection scenario 2; (c) detection scenario 3; (d) detection scenario 4

表 7 ODEM-YOLO 对不同类别的检测性能

Table 7 Detection performance of ODEM-YOLO for different class

Class	Precision	Recall	AP@0.5
Helmet	0.951	0.962	0.968
Mask	0.858	0.604	0.746
No-vest	0.979	0.993	0.983
No-helmet	0.906	0.893	0.924
No-mask	0.884	0.724	0.771
Vest	0.986	0.990	0.985
Total	0.927	0.861	0.896

态或视角变化造成的遮挡或形变. 例如, 当工人以背向摄像头且低头作业的姿势时, 安全帽的完整轮廓难以呈现, 仅局部区域暴露, 尤其当安全帽颜色与背景环境接近时, 模型可能会错误地判定为未佩戴安全帽, 导致漏检. 口罩 (Mask) 目标在所有检测类别中尺度最小, 因而检测难度最大. 尤其在远距离拍摄情况下, 工人的口鼻区域特征不明显, 面部细节丢失, 进一步增加了识别难度, 使模型难以准确检测出口罩目标, 出现漏检的情况. 此外, 反光衣 (Vest) 检测虽相对较容易, 但在复杂场景中也可能存在不足, 例如工人彼此站位密集或身体遮挡明显, 且现场光照条件较差时, 反光衣的反光特征弱化, 模型便可能因特征不明显而产生漏检现象.

(Vest); 5(d) ~ 5(f) 三张为本文模型的检测结果图. 分析发现, 安全帽 (Helmet) 的漏检主要由于工人姿

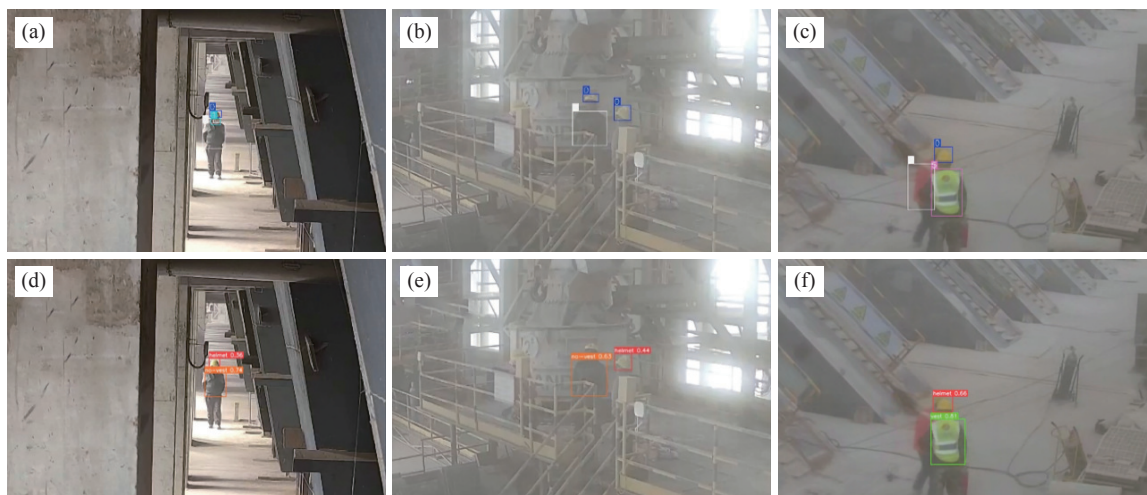


图 5 检测不成功的案例. (a) 安全帽和口罩标注; (b) 安全帽和未穿反光衣标注; (c) 安全帽、未穿反光衣和穿反光衣标注; (d) 安全帽和口罩输出; (e) 安全帽和未穿反光衣输出; (f) 安全帽、未穿反光衣和穿反光衣输出

Fig.5 Examples of unsuccessful detections: (a) label of helmet and mask; (b) label of helmet and no-vest; (c) label of helmet, no-vest, and vest; (d) output of helmet and mask; (e) output of helmet and no-vest; (f) output of helmet, no-vest, and vest

这些分析结果说明,在工业现场安全穿戴检测任务中,模型对视角变化、尺度过小、遮挡严重和光照条件差等复杂情境仍需进一步优化和提高鲁棒性。

4 应用讨论与展望

在工业安全监管实际部署中,为避免因漏检未佩戴安全装备而导致严重安全事故,将 ODEM-YOLO 模型的置信度阈值从默认的 0.25 降低至 0.10,以提升召回率。虽然此举略微增加了误报,但可通过人工复核及多步筛查机制有效控制。为验证模型的实时性能及资源利用效率,将其部署至 NVIDIA Jetson Nano B01 嵌入式设备上,该设备采

用 128 核 Maxwell GPU、四核 ARM Cortex-A57 CPU 和 4GB LPDDR4 内存(带宽 $25.6 \text{ GB}\cdot\text{s}^{-1}$),适合边缘端推理需求。部署实例效果如图 6 所示:模型在工业现场画面中自动检测出安全帽(Helmet)和未佩戴反光背心(No-vest),并以绿色检测框实时呈现其位置,直观体现了 ODEM-YOLO 在边缘端的高效检测能力。通过 TensorRT 的 FP16 精度推理及 INT8 量化优化后,模型大小仅为 6.9 MB,计算量约 9.5 GFLOPs,在 Jetson Nano 上实现了 $25 \text{ frame}\cdot\text{s}^{-1}$ 的实时推理,满足工业现场的实时监控需求。未来进一步开展模型剪枝、知识蒸馏及内存管理优化等工作,以持续提升模型在边缘设备上的运行效率。

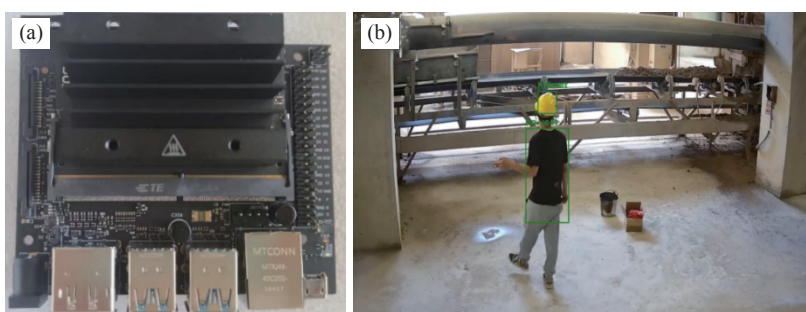


图 6 Jetson Nano B01 上 ODEM-YOLO 的检测效果示例。(a) NVIDIA Jetson Nano B01 嵌入式设备;(b) ODEM-YOLO 的检测效果示例

Fig.6 Example detection results of ODEM-YOLO on Jetson Nano B01: (a) NVIDIA Jetson Nano B01 embedded device; (b) sample detection results of ODEM-YOLO

5 结论

本文针对水泥骨料生产车间的安全穿戴检测任务,提出了基于 YOLOv8 模型改进的 ODEM-YOLO 算法。研究中,通过引入 ODConv 模块增强了浅层网络对多尺度小目标的特征提取能力,引入的 iEMA 注意力机制有效提升了 Neck 网络对多尺度目标特征的表达能,并设计 C2f_MSEIE 模块强化了对目标边缘细节信息的捕获。这些结构上的创新使模型能够显著提升对安全帽、口罩、反光衣等小目标的检测精度,尤其是提升了难度较大的口罩类目标的检测表现。

实验构建了基于真实工业环境的多尺度小目标数据集,验证表明 ODEM-YOLO 在保持轻量化模型和低计算复杂度的同时,实现了优于当前主流模型的检测精度,mAP@0.5 达到 0.896,口罩类目标 AP@0.5mask 提高至 0.746,单张图像推理时间达到 8.2 ms。通过在 NVIDIA Jetson Nano B01 边缘设备上的实际部署,模型的实时推理帧率稳定达到 $25 \text{ frame}\cdot\text{s}^{-1}$,满足了工业现场实时安全监控的实际需求。

尽管如此,模型在实际应用场景中仍存在因视角变化、遮挡严重、光照条件差等因素导致的漏检问题,这提示未来工作需要进一步提升模型的鲁棒性。下一步研究将着重探索模型剪枝、知识蒸馏及内存管理优化等方法,进一步提升模型在复杂工业场景下的适应性和边缘设备部署效率,以实现更加高效、稳定和精确的工业安全监控方案。

参 考 文 献

- [1] Yang L, Guo Y M, Huo Y B, et al. Application of improved YOLOv5 for illegal wearing detection in electric power construction. *Power Syst Prot Control*, 2023, 51(14): 160
(杨乐,郭一鸣,霍勇博,等.改进 YOLOv5 在电力生产违规穿戴检测中的应用.电力系统保护与控制,2023,51(14):160)
- [2] Cui Y Z, He J R, Ren Y Y. Exploration of artificial intelligence platform technology suitable for coal mine underground. *Coal Sci Technol*, 2025, 53(S1): 275
(崔亚仲,贺建荣,任艳艳.神东矿区人工智能安全生产管控平台应用研究.煤炭科学技术,2025,53(S1):275)
- [3] Hou G Y, Chen Q H, Yang Z H, et al. Helmet detection method based on improved YOLOv5. *Chin J Eng*, 2024, 46(2): 329
(侯公羽,陈钦煌,杨振华,等.基于改进 YOLOv5 的安全帽检测

- 算法. 工程科学学报, 2024, 46(2): 329)
- [4] Zhang S, Wang B T, Tu J Y, et al. SCE-YOLO: Improved YOLOv8 vision detection algorithm for lightweight UAV. *Comput Eng Appl*, 2025, 61(13): 100
(张帅, 王波涛, 涂嘉怡, 等. SCE-YOLO: 改进 YOLOv8 的轻量化无人机视觉检测算法. 计算机工程与应用, 2025, 61(13): 100)
- [5] Bai P R, Wang R, Liu Q Y, et al. DS-YOLOv5: A real-time detection and recognition model for helmet wearing. *Chin J Eng*, 2023, 45(12): 2108
(白培瑞, 王瑞, 刘庆一, 等. DS-YOLOv5: 一种实时的安全帽佩戴检测与识别模型. 工程科学学报, 2023, 45(12): 2108)
- [6] Krishna H, Jawahar C V. Improving small object detection // 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR). Nanjing, 2017: 340
- [7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 779
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector. *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016: 21
- [9] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 936
- [10] Tong K, Wu Y Q. Research advances on deep learning based small object detection benchmarks. *Acta Electron Sin*, 2024, 52(3): 1016
(童康, 吴一全. 基于深度学习的小目标检测基准研究进展. 电子学报, 2024, 52(3): 1016)
- [11] Cheng G, Yuan X, Yao X W, et al. Towards large-scale small object detection: Survey and benchmarks. *IEEE Trans Pattern Anal Mach Intell*, 2023, 45(11): 13467
- [12] Yuan X, Cheng G, Yan K B, et al. Small object detection via coarse-to-fine proposal generation and imitation learning // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, 2023: 6294
- [13] Wang B, Le H X, Li W J, et al. Mask detection algorithm based on improved YOLO lightweight network. *Comput Eng Appl*, 2021, 57(8): 62
(王兵, 乐红霞, 李文璟, 等. 改进 YOLO 轻量化网络的口罩检测算法. 计算机工程与应用, 2021, 57(8): 62)
- [14] Tian F, Jia H P, Liu F. Small target detection in oilfield operation field based on improved YOLOv5. *Comput Syst Appl*, 2022, 31(3): 159
(田枫, 贾昊鹏, 刘芳. 改进 YOLOv5 的油田作业现场安全着装小目标检测. 计算机系统应用, 2022, 31(3): 159)
- [15] Zhang H, Liang M Y, Wang Y F. YOLO-BS: A traffic sign detection algorithm based on YOLOv8. *Sci Rep*, 2025, 15(1): 7558
- [16] Wei L, Tong Y F. Enhanced-YOLOv8: A new small target detection model. *Digit Signal Process*, 2024, 153: 104611
- [17] Shen Q, Zhang X Y, Shen M C, et al. Multi-scale adaptive YOLO for instance segmentation of grape pedicels. *Comput Electron Agric*, 2025, 229: 109712
- [18] Li C, Zhou A J, Yao A B. Omni-dimensional dynamic convolution [J/OL]. arXiv preprint (2022–09–16) [2025–06–01]. <https://arxiv.org/abs/2209.07947>
- [19] Ouyang D L, He S, Zhang G Z, et al. Efficient multi-scale attention module with cross-spatial learning // ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island, 2023: 1
- [20] Zhang J N, Li X T, Li J, et al. Rethinking mobile block for efficient attention-based models // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, 2023: 1389
- [21] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4510
- [22] Hu J, Shen L, Sun G. Squeeze-and-excitation networks // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. IEEE, 2018: 7132
- [23] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block attention module // *Computer Vision—ECCV 2018*. Cham: Springer International Publishing, 2018: 3
- [24] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, 2021: 13708
- [25] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(6): 1137
- [26] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell*, 2020, 42(2): 386
- [27] Wang J F, Chen Y, Dong Z K, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection. *Neural Comput Appl*, 2023, 35(10): 7853
- [28] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: Learning what you want to learn using programmable gradient information // *Computer Vision—ECCV 2024*. Cham: Springer Nature Switzerland, 2024: 1
- [29] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection. *Adv Neural Inf Process Syst*, 2024, 37: 107984
- [30] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2024: 16965