

中文引用格式:周宇,吴鑫,陈洁. 远距监控下重工业车间人员不安全行为检测算法[J]. 中国安全科学学报, 2026, 36(5): 243-250.

英文引用格式:Zhou Yu, Wu Xin, Chen Jie. Detection algorithms for unsafe behaviors of personnel in heavy industrial workshops under remote monitoring[J]. China Safety Science Journal, 2026, 36(5): 243-250.

远距监控下重工业车间人员不安全行为检测算法^{*}

周宇¹, 吴鑫^{1,2}副教授, 陈洁^{**1}副教授

(1 湖南工商大学 计算机学院, 湖南 长沙 410205; 2 湘江实验室, 湖南 长沙 410205)

中图分类号: X924.2

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2026.05.08

基金项目: 湖南省自然科学基金资助(2023JJ40238); 湖南省教育厅优秀青年基金资助(23B0596); 湖南省普通高等学校教学改革研究项目(HNJG-20230793); 湘江实验室重大项目(24XJJCYJ01004)。

【摘要】 为解决重工业车间远距监控视角下小目标检测精度不足的问题, 提出一种基于改进YOLOv7的不安全行为检测算法。首先, 利用轻量级内容感知特征重组(CARAFE)模块替代传统上采样, 通过自适应特征重组有效保留小目标的语义信息; 其次, 提出一种改进的双层高效层聚合网络(Bi-ELAN)模块, 在头部网络中融合BiFormer动态稀疏注意力机制, 强化多尺度特征融合能力, 并建立目标-背景上下文关联; 然后, 改进损失函数, 引入形状交并比(ShapeIoU)损失函数, 通过几何形状约束提升边界框回归精度; 最后, 在构建的远距监控视角数据集上, 对改进后的YOLOv7模型开展消融试验与对比试验。结果表明: 改进YOLOv7算法在保持模型轻量化的同时, 显著提升远距监控场景下小目标的检测精度, 精确率为84.2%, 召回率为78.6%, 平均精度均值(mAP)@0.5为78.8%, 相比原YOLOv7算法, 改进后的算法精确率、召回率、mAP@0.5分别提高5%、0.3%、2.6%。

【关键词】 远距监控; 重工业车间; 不安全行为; 目标检测; YOLOv7

Detection algorithms for unsafe behaviors of personnel in heavy industrial workshops under remote monitoring

Zhou Yu¹, Wu Xin^{1,2}, Chen Jie¹

(1 School of Computer Science, Hunan University of Technology and Business, Changsha Hunan 410205, China; 2 Xiangjiang Laboratory, Changsha Hunan 410205, China)

Abstract: To address the issue of insufficient small-object detection accuracy in remote monitoring of heavy industrial workshops, an unsafe behavior detection algorithm based on improved YOLOv7 was proposed. First, the traditional upsampling was replaced with a lightweight content-aware reassembly of features (CARAFE) module, which effectively preserved the semantic information of small objects through adaptive feature reassembly. Second, an improved Bi-level routing efficient layer aggregation network (Bi-ELAN) module was proposed by integrating the BiFormer dynamic sparse attention mechanism into the head network, which strengthened the multi-scale feature fusion capabilities and established target-background contextual relationships. Third, the loss function was refined by introducing the shape intersection over union (ShapeIoU) loss function, which enhanced bounding box regression accuracy

^{*} 文章编号: 1003-3033(2026)05-0243-08; 收稿日期: 2026-01-12; 修稿日期: 2026-03-24

^{**} 通信作者: 陈洁(1981—), 女, 湖南衡阳人, 博士, 副教授, 主要从事复杂系统建模、机器视觉检测、机器人协同控制等方面的研究。E-mail: chenjie@hutb.edu.cn。

through geometric shape constraints. Finally, ablation experiments and comparative experiments were conducted on the improved YOLOv7 model based on constructed remote monitoring perspective dataset. The results show that, while maintaining model lightweight characteristics, the proposed algorithm significantly improves small-object detection accuracy in remote monitoring scenarios. The improved model achieves a precision of 84.2%, a recall of 78.6%, and a mean average precision (mAP@0.5) of 78.8%. Compared to the original YOLOv7 algorithm, the improved algorithm increases precision, recall, and mAP@0.5 by 5%, 0.3%, and 2.6%, respectively.

Keywords: remote monitoring; heavy industrial workshop; unsafe behavior; object detection; YOLOv7

0 引言

随着《“十四五”国家安全生产规划》的深入推进,重工业安全智能监控已成为工业数字化转型的关键方向。据统计,我国制造业安全事故中约 32% 源于人员不安全行为,传统人工巡查方法效率低且易受主观因素影响^[1],难以满足工业安全管理需求。尽管基于深度学习的目标检测技术为安全监控提供了新思路,但在占地面积超过 5 000 m² 的大型重工业车间中,高空广角监控摄像机因拍摄距离远(通常 15~30 m)、目标成像尺寸小(平均像素占比<0.1%),给现有算法带来显著挑战。因此,开展精准稳定的远距监控视角下的重工业车间不安全行为检测算法研究具有重要的理论与实践意义^[2-3]。

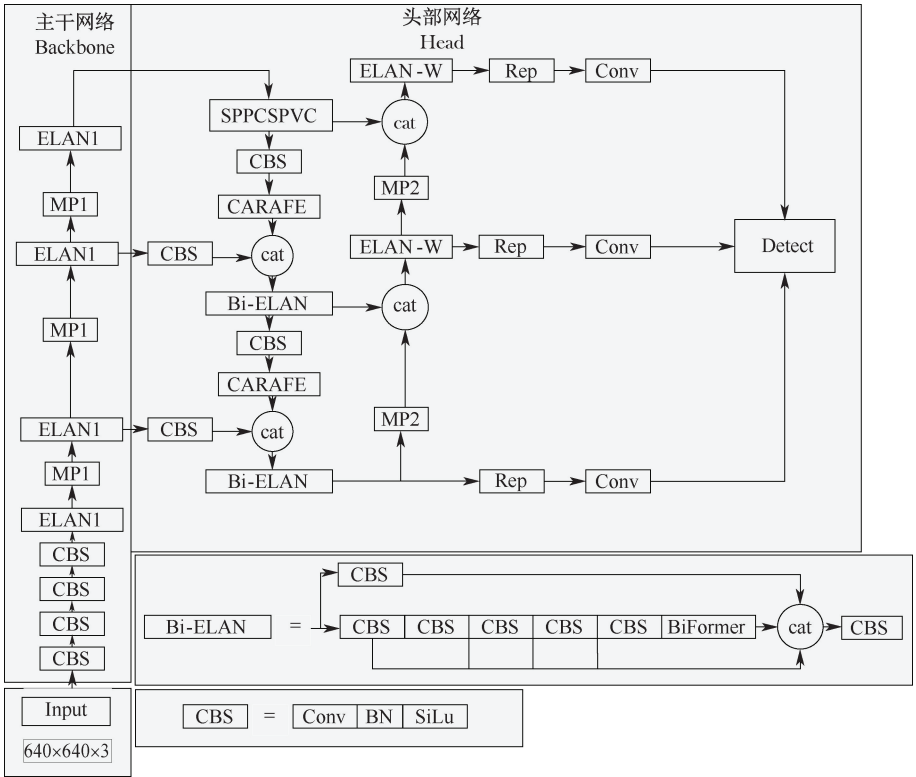
现有基于深度学习的目标检测算法主要分为双阶段网络和单阶段网络 2 类。双阶段网络以快区域卷积神经网络(Faster Region-based Convolutional Network, Faster R-CNN)^[4]为代表,通过区域生成与分类 2 个步骤实现检测,精度较高但计算复杂度大,难以满足实时及边缘部署需求。单阶段网络则以 YOLO^[5] 和单发多框检测器(Single Shot MultiBox Detector, SSD)^[6]为代表,直接在单一网络中完成目标的定位和分类,具有更快的检测速度,更适合工业检测、安全监控^[7]等实时应用场景^[8]。冯勇等^[9]设计了 YOLOv8-s-LE 算法,有效减少了模型参数。张磊等^[10]使用深度可分离卷积(Depthwise Separable Convolution, DW Conv)、改进的路径聚合网络(Path Aggregation Network, PANet)与高效交并比(Efficient Intersection over Union, EIou)损失函数优化 YOLOv5s,从而在提高检测精度的同时,降低参数量和计算量。徐壮等^[11]提出一种基于 YOLOv8n 的改进安全帽佩戴检测算法,通过引入分组空间特征融合、通道与空间注意力机制及动态交并比(Intersection over Union, IoU)优化,提升了目标检测

精度。高立鹏等^[12]提出基于 YOLOv10n 的检测算法,通过优化多尺度特征提取,提高了检测性能。郑海洋等^[13]通过数据增广、特征金字塔改进和初始候选框优化等方式提升小目标识别精度。Tang Chaoli 等^[14]通过将 YOLOX 骨干网络替换为 Swin Transformer V2,有效提高了小目标检测性能,但同时也增加了计算开销和模型复杂度。Wang Shanshan 等^[15]使用 K-means 算法对先验框进行聚类优化以提升先验框与真实框之间的匹配度;并提出针对性方法增强小目标识别能力,从而提高了检测精度。但是在复杂场景下,K-means 可能无法完美地聚类出最优的先验框^[16]。综上所述,现有目标检测方法虽然不同程度改进了单阶段算法,但仍存在不足,尤其在远距监控下的重工业车间小目标检测中,易受复杂背景及目标尺度差异等因素影响,导致检测精度偏低,难以及时预警不安全行为。

鉴于此,笔者拟提出一种基于改进 YOLOv7^[17]的远距监控视角下的重工业车间不安全行为检测算法,通过安全帽与手套小目标检测验证其在复杂工业环境中的有效性,以期提升不安全行为检测的精度与可靠性。

1 远距不安全行为检测算法模型构建

以 YOLOv7 为主干网络,首先,引入轻量级内容感知特征重组(Content-Aware ReAssembly of Features, CARAFE)模块保留更多语义信息以减少远距监控视角下小目标检测中的信息损失;然后,构建改进的双层高效层聚合网络(Bi-Level Routing Efficient Layer Aggregation Network, Bi-ELAN)模块,即在头部网络中融合 BiFormer 注意力模块提升多尺度特征提取能力,在复杂重工业车间背景下可以更精准地捕捉小目标细节;最后,引入形状 IoU(Shape IoU, ShapeIoU)损失函数增强预测框精度,提高目标检测的准确率。算法网络结构如图 1 所示。



注: 卷积-批归一化-SiLU 激活模块 (Conv + Batch Normalization + SiLU, CBS); 最大池化模块 (Max Pooling, MP); 高效层聚合网络模块 (Efficient Layer Aggregation Network, ELAN); 空间金字塔池化跨阶段部分连接模块 (Spatial Pyramid Pooling Cross Stage Partial Connections, SPPCSPC)。

图 1 改进后网络结构
Fig. 1 Improved network structure

1.1 上采样算子改进

上采样操作用于恢复特征图分辨率以进行多尺度特征融合。YOLOv7 使用的最近邻插值法虽然计算高效,但在远距离工业监控场景下,该方法对安全帽、手套等小目标的边缘平滑和细节恢复能力不足,易产生锯齿现象和特征信息丢失。

针对上述问题,引入一种高效上采样算子 CARAFE^[18]进行改进, CARAFE 能够依据输入特征

图内容动态生成卷积核权重,实现自适应的特征图重组与上采样。与传统的插值方法相比, CARAFE 能够更好地保留细节和边界信息,使得它在处理复杂场景和小目标时具有更好的表现。

CARAFE 结构如图 2 所示, CARAFE 算子由核预测模块和特征重组模块 2 个模块组成。

假设上采样倍率为 σ , 输入形状为 $H \times W \times C$ 的特征图 X , CARAFE 先通过核预测模块预测上采样

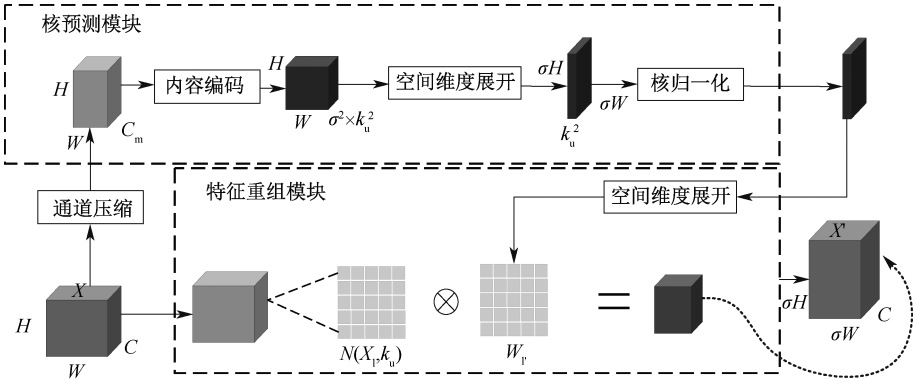


图 2 CARAFE 结构
Fig. 2 CARAFE structure

核,再利用特征重组模块完成上采样,得到形状为 $\sigma H \times \sigma W \times \sigma C$ 的输出特征图 X' 。特征图 X' 的任意目标位置 $l' = (i', j')$ 都与输入特征图 X 的初始位置 $l = (i, j)$ 对应,其中, $i = \lfloor i'/\sigma \rfloor, j = \lfloor j'/\sigma \rfloor$ 。将 $N(X_l, k)$ 定义为 X 以位置 l 为中心的 $k \times k$ 子区域。核预测模块 ϑ 利用 X_l 的子区域预测每个目标位置 l' 的上采样核 $W_{l'}$,再结合特征重组模块 ϑ 通过对 X_l 的邻域和内核 $W_{l'}$ 进行加权组合得到位置 l' 的上采样特征,其公式如下:

$$W_{l'} = \varphi(N(X_l, k_e)) \tag{1}$$

$$X'_{l'} = \vartheta(N(X_l, k_u), W_{l'}) \tag{2}$$

式中 k_e 为核预测模块中的卷积核尺寸。在核预测模块 φ 中为减小后续步骤的计算量,对形状为 $H \times W \times C$ 的输入特征图先使用 1×1 的卷积进行通道压缩。将压缩后为 $H \times W \times C_m$ 的特征图进行内容编码,使用一个 $k_e \times k_e$ 的卷积层来预测上采样核,输入的通道数为 C_m ,输出的通道数为 $\sigma^2 k_u^2$,再通过空间维度展开得到 $\sigma H \times \sigma W \times k_u^2$ 的上采样核,通过 Softmax 函数归一化处理上采样核。

在特征重组模块 ϑ 中,对于输出特征图的每个位置,先在输入特征图中找到对应的映射区域。再将这一区域的特征与该位置预测得到的上采样核 $W_{l'}$ 进行点积运算,得到形状为 $\sigma H \times \sigma W \times C$ 的输出特征图,对于目标位置为 l' 和以 $l = (i, j)$ 为中心的相对应的方形区域 $N(X_l, k_u)$ 重组公式如下:

$$X'_{l'} = \sum_{n=-z}^z \sum_{m=-z}^z W_{l'}(n, m) \cdot X_{(i+n, j+m)} \tag{3}$$

式中: z 为采样核感受野半径, $z = \lfloor k_u/2 \rfloor$; $X'_{l'}$ 为重组后的输出。

1.2 融合 BiFormer 注意力机制的特征提取

注意力机制通过模拟人类视觉的聚焦行为,能

够动态调整对输入不同区域的关注强度,从而提升深度学习模型的信息处理能力。在远距监控下的重工业车间环境中,安全帽、手套等目标成像尺寸小,背景包含大量设备和动态干扰,常规注意力机制易受噪声影响且计算成本较高,难以有效提取小目标特征。因此,在 ELAN-W 模块中引入动态稀疏注意力机制 BiFormer^[19],构建改进的 Bi-ELAN 模块,该结构通过自适应特征筛选与高效注意力分配,在复杂背景下强化小目标的特征表达,实现轻量化计算与高效特征提取的平衡。Bi-ELAN 模块结构如图 3 所示。

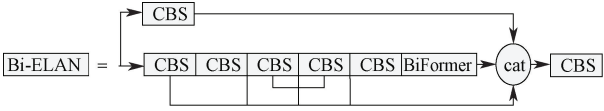


图 3 Bi-ELAN 结构
Fig. 3 Bi-ELAN structure

在 Transformer 模型中通常采用多头注意力机制 (Multi-Head Self-Attention, MHSA),将通道划分为多个子空间并行计算注意力,能有效建模长距离依赖,然而,其全连接的计算结构会带来巨大的计算开销。BiFormer 在 Transformer 的基础上引入双路由注意力 (Bi-level Routing Attention, BRA) 机制,其结构如图 4 所示。

将输入的特征图 $X \in R^{H \times W \times C}$ 划分为 $S \times S$ 个不重叠区域,其分割大小为 $\frac{H}{S} \times \frac{W}{S}$,每一个划分区域都包括 $\frac{HW}{S^2}$ 个特征向量,并通过线性映射生成查询向量 Q 、键向量 K 和值向量 V ,其表达式如下:

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v \tag{4}$$

式中 W^q, W^k, W^v 分别为查询向量 Q 、键向量 K 和值向量 V 的投影权值。

通过计算每个划分区域的查询向量 Q 和键向

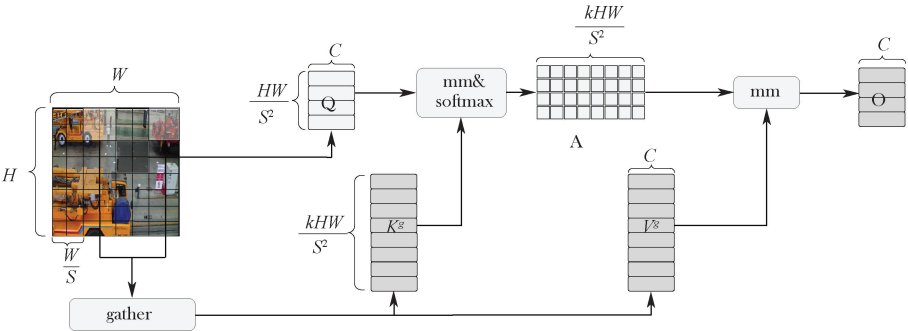


图 4 BRA 工作流程
Fig. 4 BRA workflow

量 $\mathbf{K}$ 的平均值 $\mathbf{Q}^r, \mathbf{K}^r \in R^{S^2 \times C}$, 获得区域间的邻接矩阵 $\mathbf{A}^r \in R^{S^2 \times S^2}$, 其表达式如下:

$$\mathbf{A}^r = \mathbf{Q}^r (\mathbf{K}^r)^T \quad (5)$$

通过对每个区域进行 $\text{topkIndex}(\cdot)$ 运算, 获取每个区域与其他区域之间的前 k 个连接, 形成稀疏连接路由索引矩阵 $\mathbf{I}^r \in N^{S^2 \times k}$, 其表达式如下:

$$\mathbf{I}^r = \text{topkIndex}(\mathbf{A}^r) \quad (6)$$

式中 $\mathbf{I}^r$ 的第 i 行参数 $\{\mathbf{I}_{(i,1)}^r, \mathbf{I}_{(i,2)}^r, \dots, \mathbf{I}_{(i,k)}^r\}$ 为第 i 个区域和其最具有相关性的 k 个路由区域的索引值。利用路由索引矩阵 $\mathbf{I}^r$ 来执行细粒度的令牌到令牌注意力计算。对于每个区域 i 中的查询令牌, 其仅关注由 $\{\mathbf{I}_{(i,1)}^r, \mathbf{I}_{(i,2)}^r, \dots, \mathbf{I}_{(i,k)}^r\}$ 所索引的 k 个目标路由区域中的所有 $\mathbf{K}, \mathbf{V}$ 。对所关注的键值对进行聚合操作, 其计算式如下:

$$\mathbf{K}^g = \text{gather}(\mathbf{K}, \mathbf{I}^r) \quad (7)$$

$$\mathbf{V}^g = \text{gather}(\mathbf{V}, \mathbf{I}^r) \quad (8)$$

式中 $\mathbf{K}^g, \mathbf{V}^g \in R^{S^2 \times (kHW/S^2) \times C}$, 由下式计算得到双层路由注意力:

$$\mathbf{O} = \text{softmax}\left(\frac{\mathbf{Q}(\mathbf{K}^g)^T}{\sqrt{C}}\right) \mathbf{V}^g + L(\mathbf{V}) \quad (9)$$

式中: L 为局部上下文增强模块, 通过对特征施加 5×5 深度可分离卷积操作实现进一步增强局部上下文表达能力。

1.3 基于 ShapeIoU 的损失函数改进

YOLOv7 的损失函数采用完全 IoU (Complete IoU, CIoU) 损失, 传统的 CIoU 损失函数通过同时优化目标框的重叠程度、中心点偏移以及长宽比差异, 来提升预测框与真实框之间的匹配度。CIoU 的公式如下:

$$L_{\text{IoU}} = \frac{|\mathbf{b} \cap \mathbf{b}^{\text{gt}}|}{|\mathbf{b} \cup \mathbf{b}^{\text{gt}}|} \quad (10)$$

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (11)$$

$$\mu = \frac{\nu}{\nu + (1 - L_{\text{IoU}})} \quad (12)$$

$$L_{\text{CIoU}} = 1 - L_{\text{IoU}} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})}{c^2} + \mu\nu \quad (13)$$

式中: L_{IoU} 为预测框和真实框的 IoU; L_{CIoU} 为 CIoU 损失; $\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})$ 为预测框和真实框的中心点欧氏距离; $\mathbf{b}$ 为预测框的中心点; $\mathbf{b}^{\text{gt}}$ 为真实框的中心点; c 为包含预测框和真实框的最小闭包矩形对角线距离; μ 为一个权重平衡因子; ν 用于衡量预测框和真实框的长宽比一致性; w^{gt} 和 h^{gt} 分别为真实框的宽

度与高度; w 和 h 分别为预测框的宽度和高度。但 CIoU 并未充分考虑目标框的形状特征和尺度信息。为此, 改进损失函数, 利用 ShapeIoU^[20] 损失函数替换掉 YOLOv7 的损失函数 CIoU, 如图 5 所示, 旨在进一步提升目标框与真实框之间的形状匹配精度, ShapeIoU 的相关表达公式如下:

$$\alpha = \frac{2 \times (w^{\text{gt}})^s}{(w^{\text{gt}})^s + (h^{\text{gt}})^s} \quad (14)$$

$$\beta = \frac{2 \times (h^{\text{gt}})^s}{(w^{\text{gt}})^s + (h^{\text{gt}})^s} \quad (15)$$

$$D = \beta \times \frac{(x_c - x_c^{\text{gt}})^2}{c^2} + \alpha \times \frac{(y_c - y_c^{\text{gt}})^2}{c^2} \quad (16)$$

$$\Omega = \sum_{t=w,h} (1 - \exp(-w_t))^{\theta} \quad (17)$$

$$L_{\text{Shape}} = 1 - L_{\text{IoU}} + D + 0.5\Omega \quad (18)$$

$$\begin{cases} \omega_w = \beta \times \frac{|w - w^{\text{gt}}|}{\max(w, w^{\text{gt}})} \\ \omega_h = \alpha \times \frac{|h - h^{\text{gt}}|}{\max(h, h^{\text{gt}})} \end{cases} \quad (19)$$

式中: x_c 和 y_c 分别为预测框中心点的横坐标与纵坐标; x_c^{gt} 和 y_c^{gt} 分别为真实框中心点的横坐标与纵坐标; L_{Shape} 为 ShapeIoU 损失; α 和 β 分别为水平方向和垂直方向上的权重系数; s 为用来调节宽高的重要性权重; θ 为 Ω 的归一化因子, 主要控制形状损失对最终优化目标的影响权重; D 为用于衡量预测边界框与真实边界框之间中心位置偏移的加权距离度量; ω_w 为宽度方向的差异权重; ω_h 为高度方向的差异权重。

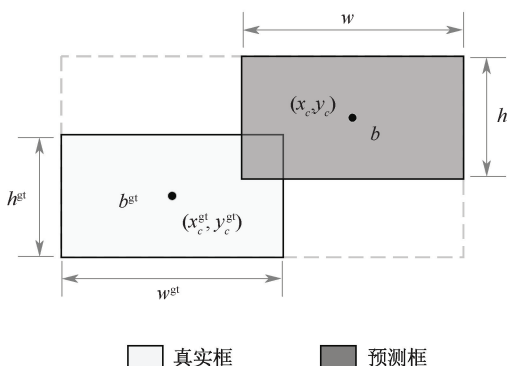


图5 ShapeIoU 损失结构

Fig. 5 ShapeIoU loss structure

基于自制数据集对 s 的取值进行对比试验, 探究不同 s 值型性能的影响。试验采用精确率 (Precision, P)、召回率 (Recall, R) 和平均精度均值 (mean Average Precision, mAP) 作为评估模型性能的指标, 其中, mAP@0.5 表示在 IoU 阈值为 0.5 时

的均值。

试验结果见表 1,当 s 值为 1.5 时,模型整体性能上最为均衡。因此,最终选择 1.5 作为 ShapeIoU 的最佳 s 值。

表 1 不同 scale 对模型的影响

Table 1 Impact of different scales on model

模型	$P/\%$	$R/\%$	mAP@ 0.5/%
YOLOv7(CIoU)	79.2	78.3	76.2
YOLOv7+ShapeloU($s=0.8$)	80.7	75.2	74.7
YOLOv7+ShapeloU($s=1.2$)	81.7	74.5	74.9
YOLOv7+ShapeloU($s=1.5$)	80.2	78.8	77.3

2 远距不安全行为检测试验与分析

2.1 数据集、试验环境及流程

基于远距监控实拍视频构建重工业车间专用数

据集,共计 1 013 张图像,涵盖不同时段、光照条件、设备布局及人员密度等典型工业场景特征。数据集中安全帽与手套 2 类目标分别标注 2 742 个和 2 809 个实例,并按 8:1:1 比例划分为训练集、验证集与测试集。在模型训练过程中采用 Mosaic 拼图数据增强方法,通过随机拼接 4 张训练图像生成复合样本,有效提升模型对小目标的检测能力与场景鲁棒性,同时缓解过拟合风险。

改进的 YOLOv7 模型基于 PyTorch 框架构建,训练采用输入尺寸 640×640,批处理大小 16,学习率 0.001,学习率动量 0.937,共进行 100 个训练轮次。

2.2 远距不安全行为检测对比试验

为进一步分析模型的性能,对比改进后的 YOLOv7 模型与 YOLOv7、YOLOv7-Tiny 和 YOLOv8,结果见表 2。

表 2 不同模型测试性能对比

Table 2 Comparison of test performance of different models

模型	$P/\%$			$R/\%$			mAP@ 0.5/%		
	总体	安全帽	手套	总体	安全帽	手套	总体	安全帽	手套
YOLOv7	79.2	86.7	71.8	78.3	97.4	59.1	76.2	92.9	59.5
YOLOv7-Tiny	72.9	85.1	60.6	56.0	90.2	21.8	56.0	89.3	22.6
YOLOv8	74.6	89.7	59.6	61.3	91.3	31.3	63.7	90.7	36.6
本文模型	84.2	89.5	78.8	78.6	97.4	59.9	78.8	94.3	63.2

由表 2 可知:相较于 YOLOv7,整体精度提高了 5%,召回率提高了 0.3%,mAP@0.5 提高了 2.6%;与 YOLOv7-Tiny 相比,3 项指标分别提高 11.3%、22.6% 和 22.8%;相较于 YOLOv8,提高 9.6%、17.3% 和 15.1%。改进后的模型在各项指标上都优于其他模型,尤其在尺寸较小、纹理细节有限且易受背景复杂度干扰“手套”类别中,改进后的 YOLOv7 模型表现突出,精确度、召回率和 mAP@0.5 都显著高于其他模型。结果表明:该模型能有效提升小目标检测的准确性与召回率,显著改善整体性能,验证了其在远距监控复工业场景中具有更优的小目标识别能力。

2.3 远距不安全行为检测结果比较

为更好地验证改进后的 YOLOv7 模型对安全帽和手套的检测性能,从测试集上选取部分图片进行测试,并局部裁剪关键区域以清晰展示检测效果,裁剪后的检测结果对比如图 6 所示。图 6 中的置信度大小可评估出模型识别的准确程度,相较于 YOLOv7 和 YOLOv7-Tiny 以及 YOLOv8 的检测结果,改进后的 YOLOv7 模型的目标框具有更高的置信度,在远距监控视角下,YOLOv7 和 YOLOv7-Tiny

以及 YOLOv8 对于小目标的检测均存在漏检的情况,而文中模型对于远景小目标检测并未出现漏检情况,说明该模型对小目标的检测效果相较于其他的对比模型表现更好。

2.4 热力图分析

利用 Grad-CAM 生成目标检测热力图,通过分析模型对图像关键区域的响应,评估其在安全帽与手套检测中的特征提取与目标敏感度。为增强分析的客观性,在测试集上引入响应区域与真实框的 IoU 作为定量指标,取所有图像中高响应区域与对应真实框 IoU 的平均值,以评估模型对目标区域的聚焦准确性。平均 IoU 试验结果见表 3。

表 3 不同模型热力图与真实框的平均 IoU

Table 3 Average IoU between heatmaps and ground truth boxes for different models

模型	平均 IoU/%
YOLOv7	78
YOLOv7-Tiny	72
YOLOv8	75
文中模型	86

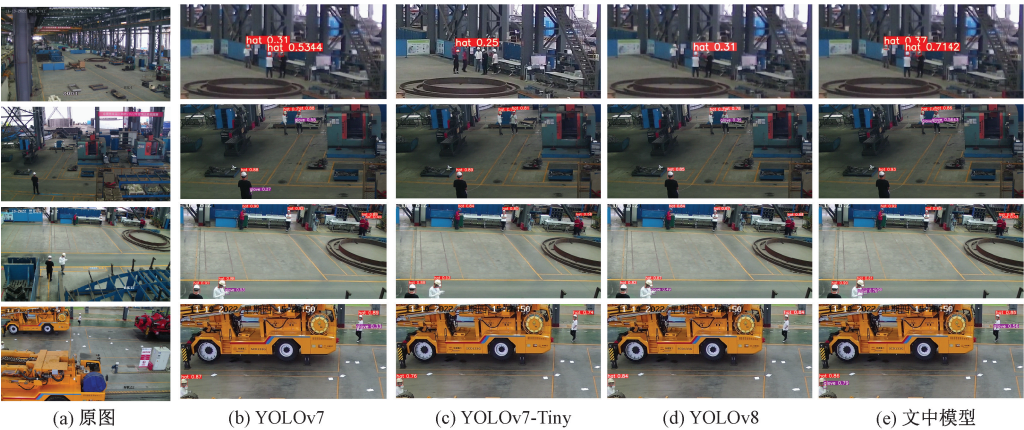


图 6 不同模型检测结果

Fig. 6 Detection results of different models

由表 3 可知:文中模型的平均 IoU 达到 86%,明显高于其他对比模型。表明其响应区域更贴近真实目标,目标定位能力更强,能更有效地反映模型对关键区域的关注程度。

2.5 模型消融试验

为评估各模块贡献,以 YOLOv7 为基准模型进行消融试验,通过逐步加入 ShapeIoU、CARAFE 和 BiFormer 模块,评估其对 P 、 R 和 mAP 的影响。消融试验结果见表 4。YOLOv7-S 表示单独加入 ShapeIoU 模块,YOLOv7-SC 表示加入 ShapeIoU 模块和 CARAFE 模块,YOLOv7-SCB 表示同时加入 3 个模块。与原始 YOLOv7 模型相比,加入 ShapeIoU 模

块后,通过增强模型对物体形状信息的感知,模型的精确率、召回率及 mAP@0.5 分别提升 1%,0.5%和 1.1%。在此基础上加入 CARAFE 模块,通过优化特征重组过程,改善了特征的细粒度表达,mAP@0.5 提升了 1.1%。在引入 ShapeIoU 和 CARAFE 的基础上加入 BiFormer 模块,通过双层路由注意力机制强化了上下文信息的捕捉能力,与原始 YOLOv7 模型相比,模型的精确率、召回率及 mAP@0.5 分别提升 5%、0.3%和 2.6%。试验结果表明:逐步引入 ShapeIoU、CARAFE 和 BiFormer 模块,能够提升模型在小目标检测任务中的精度和性能。

表 4 消融试验结果

Table 4 Ablation experiment results

模型	shapeIoU	CARAFE	Biformer	$P/\%$			$R/\%$			mAP@0.5/%		
				总体	安全帽	手套	总体	安全帽	手套	总体	安全帽	手套
—	—	—	—	79.2	86.7	71.8	78.3	97.4	59.1	76.2	92.9	59.5
YOLOv7	—	—	—	79.2	86.7	71.8	78.3	97.4	59.1	76.2	92.9	59.5
YOLOv7-S	✓	—	—	80.2	85.9	74.5	78.8	97.4	60.3	77.3	93.0	61.6
YOLOv7-SC	✓	✓	—	79.3	84.8	73.8	75.4	97.0	53.8	78.4	94.3	62.5
YOLOv7-SCB	✓	✓	✓	84.2	89.5	78.8	78.6	97.4	59.9	78.8	94.3	63.2

3 结 论

1) 通过引入轻量级 CARAFE 模块减少远距场景下小目标的信息损失,结合 BiFormer 动态稀疏注意力机制增强多尺度特征融合与背景关联,并采用 ShapeIoU 提升边界框匹配精度,有效提升了检测性能。

2) 改进后的 YOLOv7 模型有效提升了模型对小目标的识别能力。试验结果显示,改进后的模型,精确率为 84.2%,召回率为 78.6%,平均精度均值 mAP@0.5 为 78.8%。

3) 后续研究可进一步扩展模型对复杂不安全行为的识别能力,包括违规吸烟、越界操作和接打电话等行为。

参 考 文 献

[1] 郁润,张知田,郭红领. 数据驱动的施工现场工人不安全行为管理方法[J]. 中国安全科学学报,2019,29(8): 130-135.

- Yu Run, Zhang Zhitian, Guo Hongling. A data-driven management approach of construction workers' unsafe behaviors[J]. China Safety Science Journal, 2019, 29(8): 130-135.
- [2] Fang Qi, Castro-Lacouture D, Li Chengqian. Smart safety: big data-enabled system for analysis and management of unsafe behavior by construction workers[J]. Journal of Management in Engineering, 2024, 40(1): DOI: 10. 1061/JMENA. MEENG-5498.
- [3] Jiang Weiguang, Ding Lieyun. Unsafe hoisting behavior recognition for tower crane based on transfer learning[J]. Automation in Construction, 2024, 160: DOI:10. 1016/j. autcon. 2 024. 105299.
- [4] Ren Shaoqing, He Kaiming, 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, 2016, 39(6): 1 137-1 149.
- [5] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [6] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]. Computer Vision-ECCV 2016. Springer International Publishing, 2016: 21-37.
- [7] 王雪秋,高焕兵,郝泽萌.改进YOLOv8的道路缺陷检测算法[J].计算机工程与应用,2024,60(17):179-190.
Wang Xueqiu, Gao Huanbing, Jia Zemeng. Improved road defect detection algorithm based on YOLOv8[J]. Computer Engineering and Applications, 2024, 60(17): 179-190.
- [8] 纪执安,周云奕,张玉媛,等.基于改进YOLOv5的工业现场不安全行为检测[J].中国安全科学学报,2024,34(7):38-43.
Ji Zhian, Zhou Yunyi, Zhang Yuyuan, et al. Industrial site unsafe behavior detection based on improved YOLOv5[J]. China Safety Science Journal, 2024, 34(7): 38-43.
- [9] 冯勇,杨思卓,徐红艳.基于YOLO v8的轻量化安全帽佩戴检测算法[J].计算机应用,2024,44(增刊2):251-256.
Feng Yong, Yang Sizhuo, Xu Hongyan. Lightweight safety helmet wearing detection algorithm based on YOLO v8[J]. Journal of Computer Applications, 2024, 44(S2): 251-256.
- [10] 张磊,李熙尉,燕倩如,等.基于改进YOLOv5s的综采工作面人员检测算法[J].中国安全科学学报,2023,33(7):82-89.
Zhang Lei, Li Xiwei, Yan Qianru, et al. Personnel detection algorithm in fully mechanized coal face based on improved YOLOv5s[J]. China Safety Science Journal, 2023, 33(7): 82-89.
- [11] 徐壮,钱育蓉,颜丰.GCW-YOLOv8n:轻量级安全帽佩戴检测算法[J].计算机工程与应用,2025,61(3):144-154.
Xu Zhuang, Qian Yurong, Yan Feng. GCW-YOLOv8n: lightweight safety helmet wearing detection algorithm[J]. Computer Engineering and Applications, 2025, 61(3): 144-154.
- [12] 高立鹏,周孟然,胡锋,等.基于REIW-YOLOv10n的井下安全帽小目标检测算法[J].煤炭科学技术,2025,53(增刊2):417-429.
Gao Lipeng, Zhou Mengran, Hu Feng, et al. Small target detection algorithm for underground helmet based on REIW-YOLOv10n[J]. Coal Science and Technology, 2024, 53(S2): 417-429.
- [13] 郑海洋,宋纯贺,武婷婷,等.面向绝缘手套佩戴状况检测的小目标检测与匹配算法[J].小型微型计算机系统,2023,44(9):1 989-1 995.
Zheng Haiyang, Song Chunhe, Wu Tingting, et al. Small target detection and matching algorithm for wearing condition detection of insulating gloves[J]. Journal of Chinese Computer Systems, 2023, 44(9): 1 989-1 995.
- [14] Tang Chaoli, Dong Huiyuan, Huang Yourui, et al. Foreign object detection for transmission lines based on Swin Transformer V2 and YOLOX[J]. The Visual Computer, 2024, 40(5): 3 003-3 021.
- [15] Wang Shanshan, Tan Weiwei, Yang Tengfei, et al. High-Voltage transmission line foreign object and power component defect detection based on improved YOLOv5[J]. Journal of Electrical Engineering & Technology, 2024, 19(1): 851-866.
- [16] Chi T S, Ab Wahab M N, Mohamed A S A, et al. Enhancing EfficientNet-YOLOv4 for integrated circuit detection on printed circuit board (PCB)[J]. IEEE Access, 2024, 12: 25 066-25 078.
- [17] Wang Chienyao, Bochkovskiy A, Liao Hongyuan Mark. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7 464-7 475.
- [18] Wang Jiaqi, Chen Kai, Xu Rui, et al. Carafe: content-aware reassembly of features[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3 007-3 016.
- [19] Zhu Lei, Wang Xinjiang, Ke Zhanghan, et al. Biformer: vision transformer with bi-level routing attention[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10 323-10 333.
- [20] Zhang Hao, Zhang Shuaijie. Shape-iou: more accurate metric considering bounding box shape and scale[J]. arXiv Preprint arXiv, 2023: DOI:10. 48550/arXiv. 2312. 17663.

作者简介: 周宇 (2000—),男,湖南怀化人,硕士研究生,主要研究方向为计算机视觉、机器视觉检测技术等。E-mail:zhouyu4331@163.com。

