

基于YOLOv10改进的车间人员不安全行为检测算法

陈澎哲

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2026年3月27日; 录用日期: 2026年4月22日; 发布日期: 2026年4月27日

摘要

针对车间、工地等生产作业场景中作业人员因未佩戴安全帽、吸烟、使用手机、未穿工服等不安全行为易引发严重安全事故的问题, 本文提出一种基于YOLOv10改进的不安全行为检测模型。首先, 引入小波池化模块重构YOLOv10的上下采样结构, 以增强模型对细粒度特征的感知性能; 其次, 在主干网络中构建LSKA_SPPF模块, 通过融合大核注意力机制提升多尺度特征提取能力; 最后, 设计遮挡感知检测头SEAMHead, 通过显式建模通道与空间依赖关系, 增强模型对遮挡区域的特征表达与补偿能力。在自建的WorkerBehavior数据集上的实验结果表明, 改进后的模型在检测性能上优于基准YOLOv10, mAP@0.5提升了1.1个百分点, mAP@0.5:0.95提升了0.6个百分点, 验证了所提方法在复杂工业场景下对车间人员不安全行为检测的有效性。

关键词

行为检测, YOLOv10, 多尺度, 特征提取

Improved Unsafe Behavior Detection Algorithm for Workshop Personnel Based on YOLOv10

Pengzhe Chen

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: March 27, 2026; accepted: April 22, 2026; published: April 27, 2026

Abstract

Unsafe behaviors in production environments such as workshops and construction sites, including

not wearing safety helmets, smoking, using mobile phones, and failing to wear work uniforms, can easily lead to serious safety accidents. To address this issue, this paper proposes an unsafe behavior detection model based on an improved YOLOv10. Firstly, a wavelet pooling module is introduced to reconstruct the up-sampling and down-sampling structure of YOLOv10, thereby enhancing the model's ability to perceive fine-grained features. Secondly, an LSKA_SPPF module is constructed in the backbone network, which improves multi scale feature extraction capability by integrating a large kernel attention mechanism. Finally, an occlusion aware detection head named SEAMHead is designed. By explicitly modeling channel and spatial dependencies, the proposed head enhances the model's feature representation and compensation capability for occluded regions. Experimental results on the self constructed WorkerBehavior dataset show that the improved model outperforms the baseline YOLOv10 in detection performance. The mAP@0.5 increases by 1.1 percentage points and the mAP@0.5:0.95 increases by 0.6 percentage points. These results demonstrate that the proposed method is effective for detecting unsafe worker behaviors in complex industrial scenarios.

Keywords

Behavior Detection, YOLOv10, Multi-Scale, Feature Extraction

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在工业生产过程中, 车间作业人员的不安全行为(如未佩戴安全帽、未穿着工服、违规吸烟等)直接威胁生产安全, 易引发严重事故, 造成人员伤亡与财产损失。因此, 对作业人员行为进行实时监测与主动干预, 对于保障人身安全、降低事故风险、提升生产稳定性具有重要意义。当前, 工厂主要依赖线下安全教育培训与人工巡检等传统手段进行防控。然而, 人工监管方式受主观因素影响, 存在漏判、误判等问题, 且受限于人力资源与注意力持续性, 难以实现全天候、全覆盖的监测。

为突破上述人防模式的固有局限, 并满足工业场景对实时性、高精度及自动化水平的迫切需求, 基于计算机视觉的智能监测技术成为研究与应用的热点。近年来, 以卷积神经网络为核心的深度学习方法能够自动提取图像深层特征, 实现对作业人员的高效识别与定位。现有目标检测算法主要分为两阶段方法与一阶段方法: 前者如 RCNN [1]、Faster-RCNN [2]等, 通过生成候选区域再进行分类与回归, 检测精度较高; 后者如 YOLO [3]、SSD [4]等, 采用端到端预测, 推理速度更具优势。这些技术为工业场景下不安全行为的智能识别与预警提供了有力支撑。

尽管现有的检测方法已取得显著研究进展, 但仍面临两方面的主要挑战: (1) 对于尺度较小的目标, 检测过程中仍存在较高的误检与漏检率; (2) 在目标受到遮挡或背景干扰等复杂场景下, 模型的检测性能与鲁棒性往往难以保证。为突破上述局限, 本文以 YOLOv10 [5]为基础模型, 从多个角度对其进行改进, 整体架构如图 1 所示。需要说明的是, 图 1 中标注的 C2fCIB 模块属于 YOLOv10 自带的变体结构, 该模块在原始模型中用于优化特征提取与特征融合过程, 其设计延续了 YOLO 系列在计算效率与表征能力之间平衡的考量。本文的核心改进则在此基础上展开, 主要贡献如下:

1) 针对现有模型上下采样操作易导致细粒度空间信息丢失、进而影响小目标检测性能的问题, 引入小波池化模块重构 YOLOv10 的上下采样过程, 在提升特征图结构保持能力的同时, 通过多尺度细节信息提取有效抑制过拟合, 增强模型对细粒度特征的感知能力。

2) 提出 LSKA-SPPF 改进模块,旨在提升模型在多尺度特征提取方面的能力。该模块在原始 SPPF 结构基础上嵌入大可分核注意力(LSKA)机制,利用其将大核深度卷积分解为级联一维卷积的特性,显著提升了模型对多尺度上下文信息的感知能力,进而增强了对复杂场景下目标的特征表达与判别能力,同时,有效解决复杂工业场景中目标尺度变化剧烈、现有空间金字塔池化结构对多尺度上下文建模能力不足的问题。

3) 针对遮挡目标在特征图中易被抑制、导致漏检与定位偏差的问题,设计适用于复杂场景的遮挡感知检测头 SEAMHead,通过显式建模通道间依赖关系与空间上下文信息,增强模型对遮挡区域的特征表达与特征补偿能力。

4) 构建面向人员不安全行为检测的专用数据集(WorkerBehavior),并在该数据集进行了系统实验,验证了所提算法的有效性。

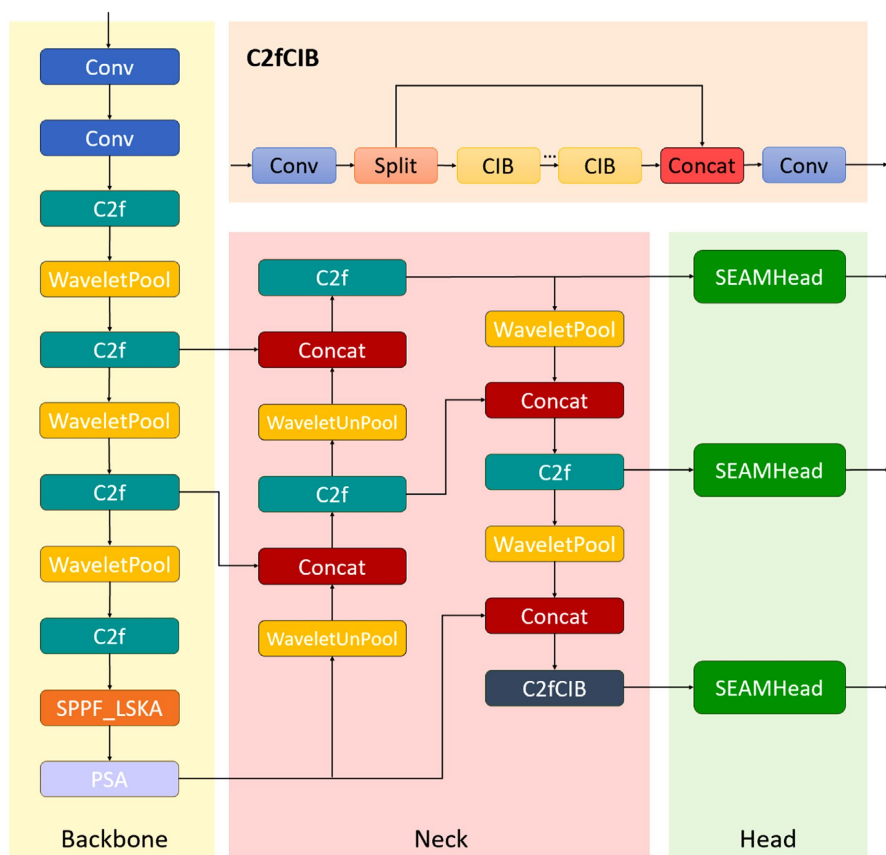


Figure 1. Network structure of the improved YOLOv10
图 1. 改进的 YOLOv10 网络结构图

2. 相关工作

在工业场景下的人员不安全行为检测研究中,现有方法虽取得一定进展,但仍存在明显局限,为本研究提供了明确的切入点。Gao [6]等人采用形态学滤波、光流法及 Mean-Shift 算法实现实验室场景下的图像去噪、目标检测与跟踪,虽在受控环境中表现有效,但在复杂工业现场中面临形态学滤波噪声适应性有限、光流法计算复杂度过高;Wu [7]等人针对电力作业人员提出基于时空视觉注意力的关键帧提取方法,通过注意力机制挖掘关键帧信息提升了识别精度,但关键帧提取可能丢失动态行为的连续性信息,

尤其对短时小幅动作的表征能力不足,对复杂背景干扰敏感,其传统行为识别框架的特征表示能力较弱; Tian [8]等人针对煤矿井下矿工提出改进 AlphaPose 姿态估计结合 ST-GCN 的行为识别方法,通过引入 RTDETR 与 Swin-Transformer 模块提升了姿态估计精度,然而 AlphaPose 作为自上而下的姿态估计方法,其性能高度依赖目标检测的准确性,在密集人群或遮挡场景下易产生漏检,且两阶段架构导致计算流程较长,难以满足工业场景的实时性要求。

综上所述,现有研究在复杂光照、遮挡、尺度变化等车间典型干扰下的鲁棒性仍有不足,且部分方法仅实现检测跟踪而未完成行为分类。因此,本文以 YOLOv10 一阶段检测模型为基础,针对车间人员不安全行为检测任务,从多尺度特征增强与遮挡感知检测头设计等角度对模型进行改进,旨在构建兼具精度与速度的端到端检测方法,以满足复杂工业场景下的实时监测需求。

3. 方法

3.1. WaveletPool 和 WaveletUnPool 模块

在目标检测任务中,特征图的下采样与上采样操作对检测精度与模型鲁棒性具有重要影响。传统的 YOLOv10 模型在特征金字塔网络中采用步幅卷积或最近邻插值等方法进行空间维度的变换,这些方法虽然高效,但容易引入信息损失或边缘伪影,限制了模型对细节特征的表达能力。受 Williams [9]等人提出的 Wavelet Pooling 启发,本研究提出一种基于小波变换的双向采样模块,用于替代 YOLOv10 中传统的下采样与上采样操作。该方法利用二维离散小波变换对特征图进行多尺度分解,在下采样过程中保留高频细节信息,

$$W_{\phi}[j+1,k]=h_{\phi}[-n]*W_{\phi}[j,n] \Big|_{n=2k,k \leq 0} \quad (1)$$

$$W_{\phi}[j,k]=h_{\phi}[-n]*W_{\phi}[j+1,n]+h_{\psi}[-n]*W_{\psi}[j+1,n] \Big|_{n=\frac{k}{2},k \leq 0} \quad (2)$$

其中, W_{ϕ} 表示近似系数, W_{ψ} 表示小波系数, j 为分辨率级别, n 为样本索引,并通过逆小波变换实现高质量的上采样重建。

$$W_{\psi}[j+1,k]=h_{\psi}[-n]*W_{\psi}[j,n] \Big|_{n=2k,k \leq 0} \quad (3)$$

具体而言,我们在 YOLOv10 的 FPN 结构中引入小波下采样模块和小波上采样模块,分别替代原有的卷积下采样和最近邻插值上采样。相较于传统方法,该方法不仅具备更优的结构保持能力,还能有效抑制过拟合,提升模型在复杂场景下的人员检测性能。

3.2. LSKA_SPPF 模块

在 YOLOv10 的架构设计中,空间金字塔池化快速(SPPF)模块主要用于聚合不同尺度的特征信息,以提升模型对多尺度目标的感知能力。然而,原始 SPPF 模块在复杂场景下面临一定的局限性:其感受野受限于固定尺寸的池化操作,难以有效捕获大范围的上下文信息;同时,在面对遮挡、模糊及小目标等干扰时,特征表示往往缺乏足够的判别性和鲁棒性。此外,传统的大卷积核虽然能够扩展感受野,但会带来显著的计算开销和参数量增长,难以在实时检测任务中取得平衡。

为克服上述不足,本文提出一种改进型模块——LSKA-SPPF,其结构如图 2 所示。该模块将大可分核注意力机制嵌入原始 SPPF 结构中。LSKA [10]通过将大核深度卷积分解为级联的一维深度卷积,在保持大感受野的同时实现对空间特征的有效建模。将 LSKA 引入 SPPF 后,模型能够借助注意力机制动态增强关键特征的响应,有效抑制无关噪声干扰,从而在多个尺度上聚合更丰富的上下文信息,显著提升

了多尺度特征的提取能力与判别性。

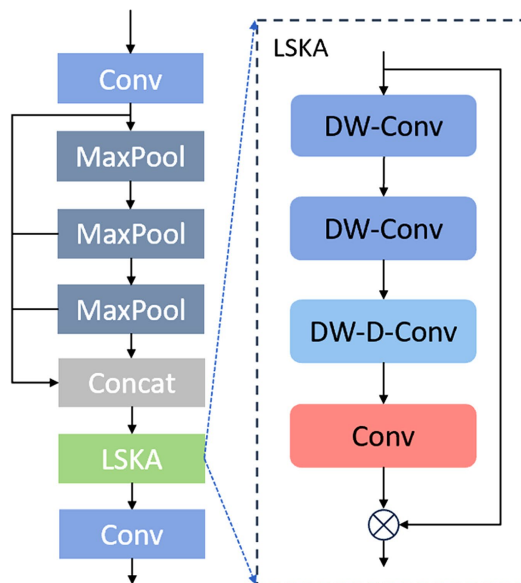


Figure 2. Structure of the LSKA_SPPF module
图 2. LSKA_SPPF 模块结构图

3.3. SEAMHead 检测头

在 YOLOv10 的原始检测头设计中，其通过多尺度特征融合与回归分类分支实现对目标的定位与识别。然而，在复杂工业场景下，人员目标常面临严重遮挡、形变及尺度变化等挑战，原始检测头存在以下固有缺陷：首先，检测头所依赖的特征图主要经由卷积核局部感受野提取，缺乏对全局上下文信息与通道间长程依赖关系的显式建模，导致被遮挡区域的特征响应被背景噪声淹没，难以有效表征局部可见的目标部分；其次，传统检测头在处理密集人群或重叠目标时，其分类与回归分支难以从混叠特征中解耦出个体目标的专属特征，从而引发漏检与误检率的上升；此外，原始结构未引入针对遮挡区域的注意力补偿机制，使得模型在部分特征缺失的情况下缺乏鲁棒的特征重建能力。

为突破上述局限，本文提出一种基于分离与增强注意力模块的遮挡感知检测头，即 SEAMHead，并将其集成至 YOLOv10 检测框架中。SEAMHead 旨在通过显式建模通道间依赖关系与空间上下文信息，增强模型对遮挡区域的特征表达与补偿能力。具体而言，SEAMHead 首先引入深度可分离卷积结合残差连接的结构，以轻量化方式提取多尺度局部特征，保留空间细节的同时降低计算开销；随后，通过通道全连接层学习特征图各通道之间的全局依赖关系，并利用指数归一化函数生成注意力权重，从而动态突出包含关键判别信息的通道；最终，将生成的注意力图与原始特征逐元素相乘，实现对遮挡区域特征的自适应增强。该设计使得被遮挡目标的局部特征能够通过全局通道关系得到补偿与强化，显著提升了检测头在部分特征缺失情况下的鲁棒性。

4. 实验设计与结果分析

4.1. 数据集

本文针对实际的检测任务以及应用场景自制 WorkerBehavior 数据集。该数据集通过网络爬取、整合公开数据集 SHWD 中的部分安全帽图片以及从多平台获取的图片，力求覆盖多样化的实际场景。数据集

中的不安全行为被划分为 6 类: helmet、head、smoke、telephone、DailyClothes 和 WorkClothes。该数据集总共包含 7388 张图片, 并按照 8:1:1 的比例划分为训练集、验证集和测试集, 具体分布为训练集 5869 张、验证集 693 张和测试集 692 张。该数据集不仅包含多来源、多场景的图片数据, 而且经过规范标注与精细化处理, 具备较高的数据质量, 为员工不安全行为检测提供了全面且可靠的数据支撑。

4.2. 评价指标

在目标检测任务中, 模型性能的评估通常从检测精度与模型复杂度两个维度展开。检测精度主要通过精确率(Precision)、召回率(Recall)及平均精度均值(mAP)三项指标进行量化。其中, 精确率衡量模型预测为正样本中真正样本的比例, 召回率则反映真正样本中被正确检测出的比例:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

平均精度(Average Precision, AP)是评估模型在某一类别上检测性能的核心指标, 其通过计算精确率-召回率(P-R)曲线下的面积获得。平均精度均值(mAP)则是对所有类别 AP 值取平均, 用以综合评价模型的整体检测性能,

$$mAP = \frac{1}{n} \sum_{i=1}^n \left(\int_0^1 Precision(Recall) d(Recall) \right)_i \quad (6)$$

4.3. 对比实验

为验证本文所提改进模型的有效性, 将本算法与 YOLOv5、YOLOv8 及 YOLOv10 基线模型在自建 WorkerBehavior 数据集上进行对比实验, 评估指标包括检测精度(mAP@0.5、mAP@0.5:0.95)、参数量(Params)及计算复杂度(GFLOPs), 实验结果如表 1 所示。

实验结果表明, 本文算法在检测精度与模型复杂度之间取得了良好平衡。在检测精度方面, 本文算法在 mAP@0.5 指标上达到 84.0%, 较 YOLOv10 基线模型提升 1.1 个百分点; 在更为严格的 mAP@0.5:0.95 指标上达到 58.9%, 较 YOLOv10 提升 0.6 个百分点, 充分验证了所提改进模块在多尺度特征提取与复杂场景建模方面的有效性。在模型复杂度方面, 本文算法的参数量较 YOLOv10 略有增加, 但仍低于 YOLOv8, 且计算量优于上述两种对比模型。这表明, 所引入的 WaveletPool、LSKA_SPPF 及 SEAMHead 模块在有效提升检测性能的同时, 并未显著增加模型的计算负担, 实现了检测精度与推理效率的良好平衡。综上, 本文算法在检测精度与模型效率两方面均表现出优越性, 能够更好地适应车间复杂场景下工人不安全行为实时检测的实际需求。

Table 1. A comparison of performance metrics for different model training

表 1. 不同模型训练的性能指标对比

Module	mAP@0.5/%	mAP@0.5-0.95/%	Params/10 ⁶	GFLOPs
YOLOv5	82.5	58.0	1.7	4.2
YOLOv8	81.6	58.6	3.0	8.1
YOLOv10	82.9	58.3	2.2	6.5
ours	84.0	58.9	2.5	5.8

为进一步直观验证本文改进算法的有效性, 我们对 YOLOv10 基线模型与改进后模型进行了热力图

可视化对比分析,结果如图3所示。从图中可以看出,YOLOv10的热力响应存在若干不足:首先,响应区域较为分散,难以形成连续且聚焦的目标轮廓,导致目标边界定位模糊;其次,在多目标共存场景下,各目标的响应强度分布不均,部分次要目标响应微弱甚至完全缺失,导致漏检问题尤为突出。相比之下,改进后的模型在热力响应方面表现出更为优越的特性。具体而言,其响应区域高度集中于真实目标所在空间范围,有效抑制了背景噪声干扰,表明模型具备更强的特征筛选能力;同时,热力分布与目标的几何结构高度吻合,尤其在行人肢体边缘等细粒度区域能够形成连续稳定的响应,体现了模型在空间感知精度上的显著提升。

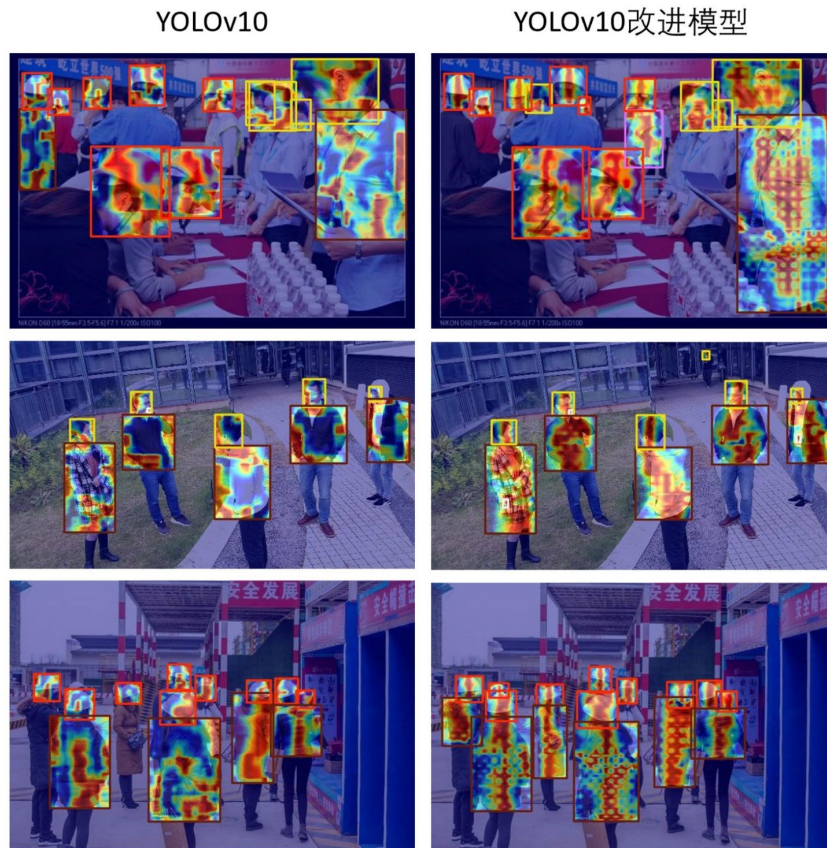


Figure 3. Heatmap comparison between the improved algorithm and the original YOLOv10 algorithm

图3. 改进算法与原YOLOv10算法热力对比图

4.4. 消融实验

为验证所提各模块的有效性,以YOLOv10为基线模型开展消融实验,通过逐步集成WaveletPool、LSKA_SPPF及SEAMHead模块,定量分析各组件对检测性能与模型复杂度的影响,实验结果如表2所示。实验表明,单独引入WaveletPool与LSKA_SPPF模块时,模型在mAP@0.5指标上由基准的0.829分别提升至0.839与0.835,验证了多尺度特征提取与高效空间注意力机制对检测精度的正向作用。然而,当二者并行集成时,精度仅提升至0.831,未呈现预期的叠加增益。分析认为,这一现象主要归因于模块间的特征竞争与梯度干扰:WaveletPool通过小波变换保持高频细节,而LSKA_SPPF侧重于大核空间感受野下的全局上下文建模,二者在特征表达上存在响应冲突,导致联合优化时难以形成协同互补。此外,

引入 SEAMHead 后,三者组合模型精度达到 0.840,说明检测头模块能够有效缓解前述干扰,实现特征补偿与语义对齐,最终发挥出各模块的联合增益。整体而言,消融结果不仅验证了各模块的有效性,也揭示了在轻量化结构下多模块协同优化所面临的非线性耦合现象,为复杂工业场景下的模型设计提供了有益参考。

Table 2. Ablation experiment
表 2. 消融实验

模块			Params/10 ⁶	GFLOPs	mAP@0.5	mAP@0.5-0.95
WaveletPool	LSKA_SPPF	SEAMHead				
			2.2M	6.5	0.829	0.583
√			2.3M	6.1	0.839	0.589
	√		2.5M	6.7	0.835	0.590
		√	2.1M	6.0	0.837	0.588
√	√		2.5M	6.3	0.831	0.584
√		√	2.2M	5.5	0.832	0.584
√	√	√	2.5M	5.8	0.840	0.589

5. 结论

本文针对车间生产场景下工人不安全行为检测任务中,因目标尺度多变、遮挡频繁及背景干扰严重所导致的细粒度特征提取不足、多尺度感知能力受限及遮挡区域特征表达困难等复合型难题,提出一种基于改进 YOLOv10 的检测模型。具体而言,引入小波池化模块重构上下采样结构,旨在解决传统下采样操作易造成高频细节丢失的问题,增强模型对小尺度目标与边缘特征的感知能力;提出 LSKA-SPPF 模块,在不显著增加计算负担的前提下扩大有效感受野,缓解单一尺度特征对尺度剧烈变化目标表征能力不足的问题;设计遮挡感知检测头 SEAMHead,通过显式建模通道依赖与空间上下文信息,弥补遮挡区域特征抑制所导致的漏检与定位偏差。三者分别从特征保真度、多尺度上下文建模及遮挡特征补偿三个维度切入,有效应对了工业场景下多种干扰因素耦合的复杂检测需求。同时,针对车间场景构建了专用数据集 WorkerBehavior,涵盖安全帽、吸烟、使用手机、穿着工服等多类不安全行为样本。实验结果表明,改进后的模型取得了优于基线模型的性能,且未显著增加部署负担。本研究为复杂工业环境下的人员不安全行为实时监测提供了一种有效的技术方案,后续工作将围绕模型轻量化与多任务行为识别展开深入探索。

参考文献

- [1] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 24-27 June 2014, 580-587. <https://doi.org/10.1109/cvpr.2014.81>
- [2] 易泓. 基于改进 Faster RCNN 的行人检测算法研究与应用[D]: [硕士学位论文]. 武汉: 湖北大学, 2024.
- [3] 邵延华, 张铎, 楚红雨, 等. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报, 2022, 44(10): 3697-3708.
- [4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., et al., Eds., *Computer Vision—ECCV 2016*, Springer International Publishing, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [5] Chen, H., Chen, K., Ding, G.G., et al. (2024) YOLOv10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems 37*, Vancouver, 10-15 December 2024, 107984-108011. <https://doi.org/10.52202/079017-3429>

-
- [6] Gao, L. (2023) Algorithm of Detection and Tracking for Laboratory Unsafe Behavior Based on Computer Vision. 2023 *International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, Bengaluru, 20-21 October 2023, 1-6. <https://doi.org/10.1109/easct59475.2023.10393584>
- [7] Wu, J., Li, G., Zhang, J., Wang, F. and Huang, Q. (2023) Recognition of Unsafe Behavior of Electrical Workers Based on Spatial-Temporal Visual Attention. 2023 *Panda Forum on Power and Energy (PandaFPE)*, Chengdu, 27-30 April 2023, 1839-1843. <https://doi.org/10.1109/pandafpe57779.2023.10141251>
- [8] Tian, F., Song, C., and Liu, X. (2025) Research on Miner's Unsafe Behaviour Detection Algorithm Based on Improved AlphaPose. 2025 *5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)*, Xi'an, 28-30 March 2025, 589-593.
- [9] Williams, T. and Li, R. (2018) Wavelet Pooling for Convolutional Neural Networks. *6th International Conference on Learning Representations*, Vancouver, 30 April-3 May 2018.
- [10] Lau, K.W., Po, L.M. and Rehman, Y.A.U. (2024) Large Separable Kernel Attention: Rethinking the Large Kernel Attention Design in CNN. *Expert Systems with Applications*, **236**, Article ID: 121352. <https://doi.org/10.1016/j.eswa.2023.121352>