

DOI:10.19651/j.cnki.emt.2519560

改进 RT-DETR 矿工不安全行为检测方法^{*}张红^{1,2} 陈晓彤¹ 许永炎¹ 高奎程¹ 王媛彬³(1. 西安科技大学通信与信息工程学院 西安 710054; 2. 西安科技大学西安市网络融合通信重点实验室 西安 710054;
3. 西安科技大学电气与控制工程学院 西安 710054)

摘要: 针对煤矿井下背景复杂、矿工行为尺度差异大和频繁遮挡等因素导致的现有行为检测模型精度低、鲁棒性不足的问题,提出一种改进 RT-DETR 矿工不安全行为检测方法。该方法构建了具备多路径特征提取与双分支下采样结构的主干网络 CANet,通过融合深浅层级特征并保留边缘细节,提升了模型在复杂背景下对行为细节的感知能力。同时,设计扩散感知特征金字塔网络 DAFPN,通过结合维度感知选择性集成模块与跨层扩散策略,构建两阶段融合—扩散机制,以增强多尺度行为特征间的语义交互,显著提升了模型对姿态多变和大尺度差异场景的适应能力。此外,引入可变核卷积模块 AKConv,通过动态调整采样位置,使网络在存在遮挡时仍能自适应聚焦于行为关键区域,增强矿工行为检测的鲁棒性。实验结果表明,改进后的 RT-DETR 模型在 mAP@0.5 和 mAP@0.5:0.95 上分别达到 92.9% 和 66.1%,较原模型分别提升 2.9% 和 1.9%,参数量减少 18%,计算量降低 13%。与 Faster R-CNN、SSD、YOLOv5m、YOLOv8m、YOLOv10m 等主流检测算法相比,整体性能更优,充分证明了该模型在复杂煤矿场景下不安全行为检测的有效性与工程应用价值。

关键词: 行为检测;RT-DETR;重参数化;维度感知模块;可变核卷积

中图分类号: TP391.41; TN957.52 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Improved RT-DETR method for detecting unsafe behaviors of miners

Zhang Hong^{1,2} Chen Xiaotong¹ Xu Yongyan¹ Gao Xicheng¹ Wang Yuanbin³(1. School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710054, China;
2. Xi'an Laboratory of Network Convergence Communication, Xi'an University of Science and Technology, Xi'an 710054, China;
3. School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: To address the issues of low accuracy and poor robustness in existing behavior detection models caused by complex underground backgrounds, large variations in miner behavior scales, and frequent occlusions, an improved RT-DETR-based unsafe behavior detection method for miners is proposed. The proposed method constructs a backbone network, CANet, featuring multi-path feature extraction and a dual-branch downsampling structure. By effectively fusing deep and shallow features while preserving edge details, CANet enhances the model's ability to perceive fine-grained behavior details in complex backgrounds. Meanwhile, a Diffusion-Aware Feature Pyramid Network (DAFPN) is designed by integrating a dimension-aware selective integration module with a cross-layer diffusion strategy, forming a two-stage fusion-diffusion mechanism to strengthen semantic interactions among multi-scale behavior features. This design significantly improves the model's adaptability to diverse postures and large-scale variations. In addition, a variable kernel convolution module (AKConv) is introduced, which dynamically adjusts sampling positions to enable the network to focus adaptively on key behavior regions under occlusion, thereby enhancing the robustness of miner behavior detection. Experimental results show that the improved RT-DETR model achieves 92.9% mAP@0.5 and 66.1% mAP@0.5:0.95, improving by 2.9% and 1.9% over the original model, while reducing parameters by 18% and computational cost by 13%. Compared with mainstream detection algorithms such as Faster R-CNN, SSD, YOLOv5m, YOLOv8m and YOLOv10m, the proposed model demonstrates superior overall performance, validating its effectiveness and engineering applicability for unsafe behavior detection in complex coal mine environments.

Keywords: behavior detection; real-time detection transformer; re-parameterization; dimension-aware module; alterable kernel convolution

0 引言

煤炭作为我国主要能源,在保障国家能源安全和推动

经济发展中具有不可替代的作用^[1]。然而,矿井下作业环境复杂,常伴随低照度、粉尘干扰与设备遮挡,极易诱发安全事故。据统计,约 74% 的煤矿事故都源于矿工的不安全

收稿日期:2025-08-07

* 基金项目:陕西省重点研发计划项目(2023YBSF-133)、西安市科技计划项目(24GXFW0049)资助

行为。因此,面向矿工行为的智能检测与实时监控是提升煤矿安全生产水平的关键手段^[2-3]。

近年来,深度学习技术在矿工行为检测领域取得显著进展,主流方法可以分为 3 类:基于骨骼关键点^[4-7]、基于时序建模^[8-10]以及基于单帧图像的方法^[11-13]。骨骼关键点方法通过姿态估计提取人体关键特征,具有一定的抗光照和背景干扰能力,但在井下易受遮挡影响,导致检测精度下降,且两阶段流程复杂,难以兼顾实时性。时序建模方法借助帧间动态信息进行连续行为检测,但对算力和帧率要求大,不利于在井下设备中部署。

相比之下,单帧图像的方法因其结构简洁、推理速度更快更适合井下算力有限且环境复杂的场景。如 Wang 等^[14]针对矿工行为检测中精度与计算效率难以兼顾的问题,提出在 YOLOv8 模型中引入可变形卷积与无参注意力机制,并优化损失函数设计,从而在保证检测精度的同时提升计算效率。陈伟等^[15]提出 MAC-YOLO 模型,在 YOLOv8l 基础上引入感受野注意力卷积,增强关键信息感知与抗干扰能力,优化回归损失提升定位精度,有效缓解井下异常行为检测中的特征提取与定位困难。骆津津等^[16]针对现有行为检测模型精度低、计算量大的问题,提出改进型 YOLOv8-ECW 算法,实现高效的行为检测。

尽管上述基于卷积神经网络(convolutional neural networks,CNN)的方法取得了一定的进展,但其固有的局部感受野限制了模型对井下复杂空间的理解。近年来,Transformer 凭借其强大的全局建模能力为该领域提供了新思路。RT-DETR(real-time detection transformer)^[17]作为代表性的实时检测模型,通过其独特的架构设计,在通用检测任务中展现出卓越性能^[18-20]。然而,将该模型直接迁移至矿井场景时,会暴露其固有不足:其主干网络在复杂背景下的特征提取能力有限,导致整体表征能力下降;特征融合结构对行为幅度剧变的适应性有限,影响多尺度感知;标准卷积算子对不规则遮挡的稳定性欠缺,造成漏检。这些挑战共同制约了模型在井下的实际应用。值得注意的是,目前尚无研究将 RT-DETR 框架专门用于矿工行为检测。现有改进多集中于通用场景优化。例如,陈星星等^[21]通过构建新型主干网络 CGSNet 以增强 RT-DETR 主干特征提取能力,虽利用多重门控提升了特征表达,但过多的注意力层叠导致模型参数量增加。徐仟祥等^[22]通过改进 BIFPN 结构以优化 RT-DETR 的多尺度特征融合能力,但固定的加权融合策略难以动态适应大尺度变化。李青云等^[23]引入 Haar 小波下采样以提升 RT-DETR 对遮挡目标的检测精度,但其固定的感受野无法应对遮挡物的多样性与不规则性。由此可见,当前不仅缺乏针对矿工行为检测的 RT-DETR 应用探索,而且现有的通用改进策略也因其固有的设计局限,难以直接迁移并有效解决矿井下的特有难题。

为此,本文采用问题驱动的设计思路,紧密围绕前述挑战,提出改进型检测模型——IRT-DETR(improved RT-

DETR),具体而言:1)设计深浅多路径特征提取与双分支下采样 ADown^[24]结构的主干网络 CANet,以增强在复杂环境下的行为细节捕捉能力。2)提出扩散感知特征金字塔网络(diffusion-aware feature pyramid network, DAFPN),结合维度感知选择性融合模块(dimension aware selective integration,DASI)^[25],实现多尺度行为特征的自适应动态融合,跨层扩散策略强化不同语义层级间的信息交互,以应对矿工行为尺度差异与姿态复杂性。3)引入可变核卷积模块(alterable kernel convolution, AKConv)^[26],利用动态偏移应对遮挡与形变,保留关键特征。在总体思路,本文与已有改进工作一脉相承,均从主干增强、特征融合与算子改进 3 方面着手。但根本区别在于,本文并非对通用模型的简单调优,而是面向矿井特殊性的深度定制,3 大模块协同工作,使 IRT-DETR 在矿工不安全行为检测任务中具备更高的精度与鲁棒性。

1 矿工不安全行为检测模型

本文算法以 RT-DETR-R18 为基准进行改进。该模型继承了 DETR 系列的端到端设计理念,摒弃了锚框与非极大值抑制等人工先验设计,通过 Transformer 编码器-解码器架构实现全局关系建模。其核心创新在于引入混合编码器,将主干网络提取的多尺度特征转换为序列表示,并采用 IoU 感知的查询选择机制,显著提升了训练收敛速度与检测精度。

针对煤矿井下背景复杂、矿工行为尺度差异大和频繁遮挡等挑战,本文在 RT-DETR 框架基础上构建了矿工不安全行为检测模型 IRT-DETR 如图 1 所示,首先,将原 ResNet-18 主干替换为自主设计的 CANet,通过多路径提取深浅层特征,双分支下采样机制增强边缘表达能力,实现对矿工肢体关键部位的细粒度建模,为区分不同行为提供基础特征;随后,设计扩散感知特征金字塔网络(DAFPN),通过维度感知选择性融合模块(DASI)计算自适应权重,并利用扩散机制在各层级间传递上下文信息,增强模型对多尺度目标和复杂姿态的适应能力;在此基础上,引入可变核卷积(AKConv)模块执行动态下采样与关键区域感知,通过可偏移卷积核聚焦肢体变化和遮挡区域;最终,优化后的特征通过 IoU 感知查询选择机制生成候选目标,由 Transformer 解码器完成高精度的定位与分类,实现在单帧图像中精准检测不安全行为。

1.1 CANet 主干网络

为应对矿工行为检测任务中设备资源受限、井下背景复杂等挑战,主干网络需在保证高效特征提取能力的同时具备较低的计算量。然而,RT-DETR 原采用的 ResNet-18 主干网络结构冗余度高,且在特征提取过程中缺乏对井下复杂背景下的适应能力。为此,本文提出 CANet 主干网络,其核心由深浅多路径特征提取与双分支下采样模块构成。

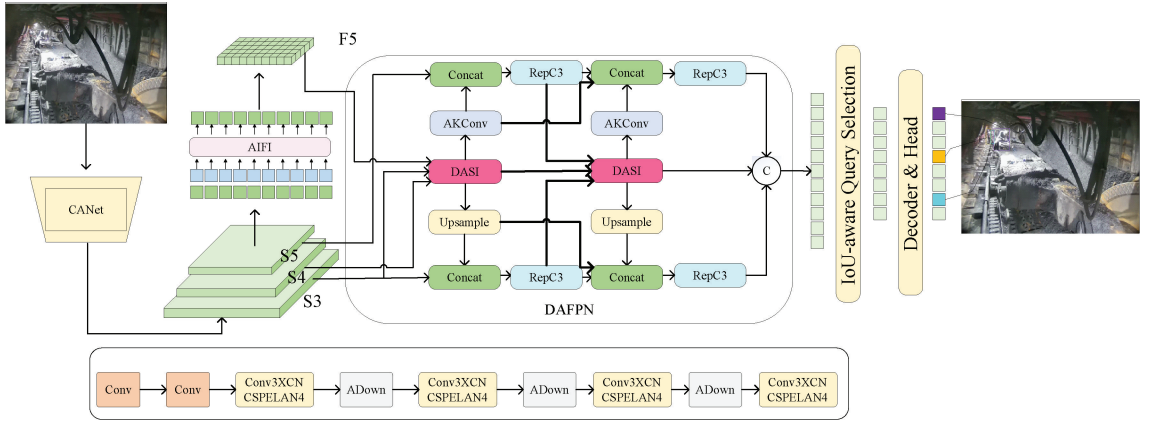


图 1 矿工不安全行为检测模型 IRT-DETR

Fig.1 Miner unsafe behavior detection model IRT-DETR

首先,在特征提取方面,本文设计了 Conv3XCNCSPELAN4 模块。该模块是在 GELAN^[24] 结构上的改进设计,在其特征提取单元引入重参数化的 Conv3XC^[27] 模块,结构如图 2 所示,旨在通过多路径特征捕捉矿工行为的全局语义与局部细节信息。具体而言,深层路径依次经多级 Conv3XC 重参数化卷积提取全局行为语义特征,用于提取“翻越”与“弯腰”等姿态幅度差异明显的行为特征;浅层路径则直接保留输入特征,重点捕捉局部边缘与细粒度信息,以辨析“摔倒”与“坐下”等行为间的微妙差异的特征。多路径的输出在通道维度进行融合,从而实现了更为细致与全面的行为特征表达。

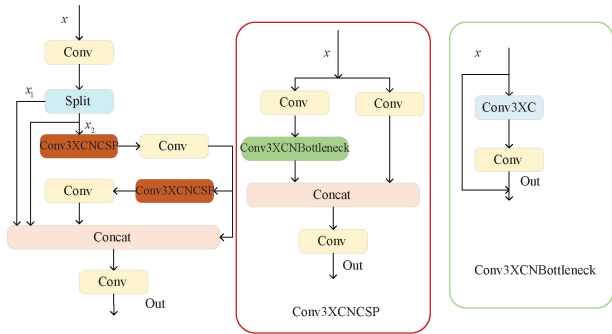


图 2 Conv3XCNCSPELAN4 模型结构

Fig.2 Conv3XCNCSPELAN4 model structure

Conv3XC 模块如图 3 所示,其通过重参数化结构降低推理计算量来适应资源受限的矿井部署环境,其在训练阶段,主路径由多个卷积层组成,推理阶段通过合并卷积核和偏置实现结构简化。假设主路径的三层卷积核分别为 W_1, W_2, W_3 , 3 个偏置项为 B_1, B_2, B_3 , 其合并过程如式(1)、(2)所示。

$$W' = \text{Trans}(\text{Trans}(W_1) \otimes W_2) \otimes W_3 \quad (1)$$

$$B' = W_3 \cdot (W_2 \cdot B_1 + B_2) + B_3 \quad (2)$$

式中: W', B' 分别表示主路径卷积核和权重的合并结果, Trans 表示对卷积核进行翻转和通道置换, \otimes 表示通过矩

阵乘法合并操作。

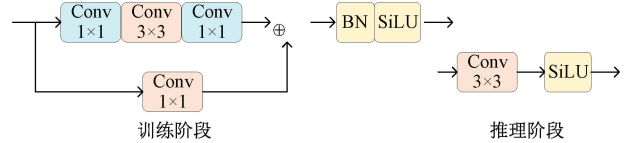


图 3 Conv3XC 模型结构

Fig.3 Conv3XC model structure

之后再跳跃连接的卷积核 W_{sk} 通过零填充为 3×3 卷积,合并后卷积核如式(3)所示。同时将主路径合并后的偏置 B' 和跳跃连接的偏置 B_{sk} 进行合并,合并后偏置如式(4)所示。

$$W'' = W' + \text{pad}(W_{sk}) \quad (3)$$

$$B'' = B' + B_{sk} \quad (4)$$

最后消除推理时的 BN 层计算,进一步减少计算量,最终融合结果如式(5)所示,其中 $W_{\text{fused}}, B_{\text{fused}}$ 是融合 BN 层后的结果。

$$Y = W_{\text{fused}} * X + B_{\text{fused}} \quad (5)$$

ADown 模块如图 4 所示,在降低分辨率的同时最大限度保留对矿工行为判别重要的细节与语义信息。输入特征首先经过平均池化抑制井下背景噪声。随后,特征沿通道维度分为两条互补路径:最大池化分支侧重提取显著边缘特征,有利于捕捉摔倒或翻越等含明显地面或物体接触的行为; 3×3 卷积分支强化局部姿态建模,能够表征肢体角度及细微行为差异。由于两分支分别处理部分通道特征,卷积操作的通道数显著减少,从而降低了计算量。最终,两分支特征融合后,整合了空间边缘与局部姿态等关键信息,在压缩特征尺寸与计算开销的同时,有效保留了对行为检测至关重要的细节与语义特征。

综上所述,Conv3XCNCSPELAN4 与 ADown 模块协同构成的 CANet 主干网络,能够在复杂井下环境和算力受限条件下,为区分不同行为提供基础特征。

1.2 扩散感知特征金字塔网络

煤矿井下空间狭窄、拍摄视角多变,矿工行为幅度大

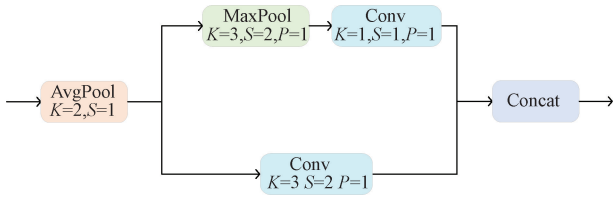


图 4 ADown 模型结构
Fig. 4 ADown model structure

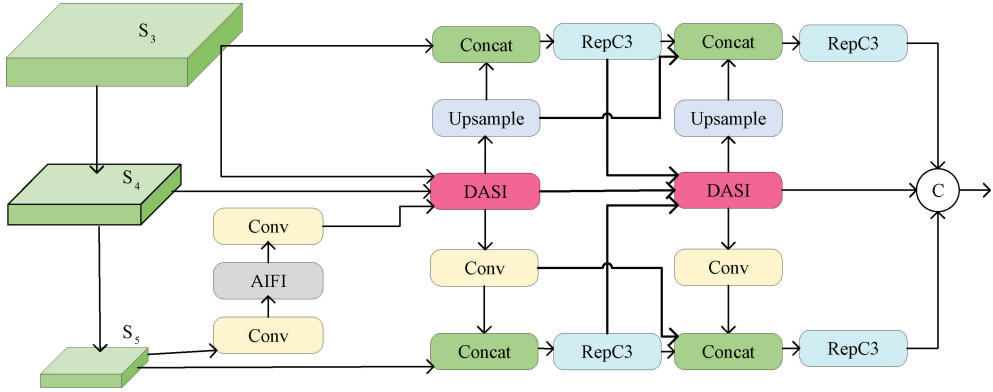


图 5 DAFPN 模型结构
Fig. 5 DAFPN model structure

经 AIFI 模块增强的深层特征,采用两阶段聚合和扩散机制强化跨尺度特征的信息交互。第 1 阶段通过 DASI 模块对不同层级特征进行融合,自适应调整贡献比例。随后,结合上下采样扩散并与原始浅层和深层进行特征拼接,并通过 RepC3 模块进一步提取语义信息,聚合后的特征再次输入 DASI 模块进行第 2 次选择性融合,之后结合上下采样将特征扩散至各个层,实现跨尺度的特征信息融合。这样的设计使浅层特征在保留手脚、头部等局部行为细节的同时,通过与深层语义特征融合能够感知整体姿态变化;深层特征在拥有全局语义的基础上,也可以通过细粒度信息补充增强对局部行为的敏感性。

DASI 如图 6 所示。其通过感知矿工行为的尺度与形

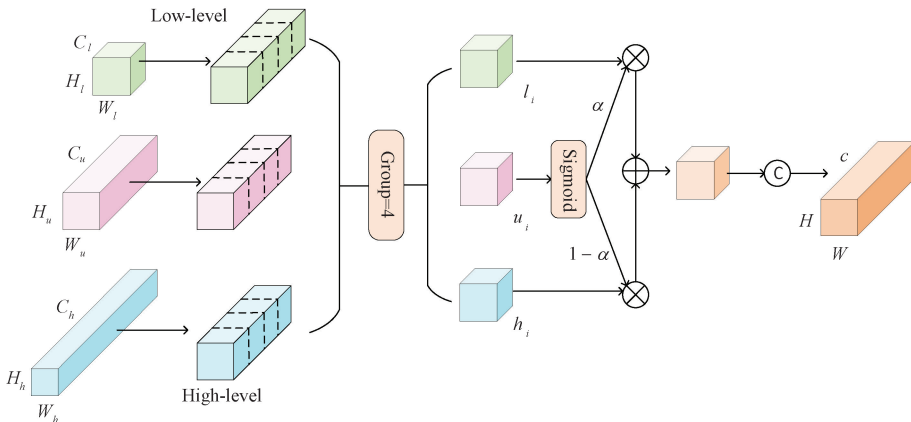


图 6 DASI 模型结构
Fig. 6 DASI model structure

且姿态多样,使图像中目标呈现明显的尺度和形态差异,增加了行为检测难度。针对该场景下现有特征金字塔网络存在的交互深度不足问题,本文提出扩散感知特征金字塔网络 DAFPN,如图 5 所示,其核心创新在于深度融合维度感知模块 DASI 与跨层扩散机制,形成两阶段融合—扩散框架,从而显著提升模型对多尺度矿工行为的特征建模与表征能力。

具体而言,DAFPN 融合主干网络提取的浅层、中层及

态变化调整权重,具体而言,DASI 模块将输入的高维特征 F_h 、低维特征 F_l 及当前特征 F_u 进行空间对齐,并在通道维度划分为 4 组,并利用当前特征生成一组融合权重,如式(6)所示。

$$\alpha = \text{sigmoid}(u_i) \tag{6}$$

式中: α 是当前特征层 u_i 的激活响应, $i = 0, 1, 2, 3$, 此机制对于幅度大的行为,更倾向融合强化深层语义,而对于局部特征明显的行为,则更多依赖浅层细节信息如式(7)所示。

$$u'_i = \alpha l_i + (1 - \alpha)h_i \tag{7}$$

最终融合特征 u'_i 在通道维度拼接后送入后续卷积层,形成对矿工行为敏感的多尺度特征表征,为分类任务提供坚实基础。

1.3 AKConv 下采样模块

井下作业环境中遮挡与形变频繁,矿工在作业过程中关键动作区域常因遮挡或姿态变化而出现部分缺失或显著变形。RT-DETR 原下采样模块,其固定的采样模式难以适应此类不规则形变,导致关键特征被丢失。为此,本文引入可变核卷积模块 AKConv 替代原下采样结构,通过其动态自适应的采样点分布,有效聚焦于未遮挡的关键部位,显著提升了网络在复杂矿井环境中的遮挡鲁棒性。

具体而言,AKConv 在标准卷积规则采样的基础上,引

入空间偏移机制,使网络能够动态调整采样位置,自适应聚焦于行为关键区域,而非依赖固定结构的感受野。如图 7 所示,AKConv 以输入特征图(C, H, W)为基础,先通过卷积操作学习到每个采样点的的偏移量($B, 2N, H, W$),结合初始坐标形成新的采样位置。偏移后的采样点能够绕过被遮挡区域,自适应聚焦在仍然可见的如手部头部等关键部位。由于偏移后的坐标可能为非整数,AKConv 采用双线性插值在原始特征图上重采样获取特征值,并经过重塑和卷积核归一化后生成输出特征图。

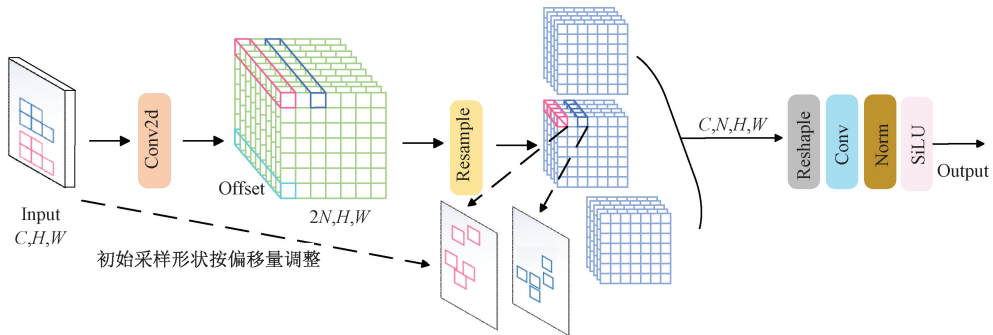


图 7 AKConv 模型结构

Fig. 7 AKConv model structure

相较于传统的卷积或下采样结构,AKConv 通过单一卷积核结合偏移预测实现可变感受野建模,聚焦关键信息,避免冗余计算。在保持特征表达力的同时降低计算量与参数开销,并在遮挡场景下增强了自适应性与关键行为的建模精度,为行为判别提供了更可靠的特征支撑。

2 实验结果与分析

2.1 数据集介绍

本实验所使用的矿工行为数据集由 3 部分组成:第 1 部分为开源的井下人员行为数据集 DsLMF+^[28];第 2 部分为 CUMT-HelmeT 数据集^[29];第 3 部分通过网络视频提取帧并筛选获得,如图 8 所示。构建的数据集共包含 4 251 张图像,按 8:1:1 的比例划分为训练集、验证集和测试集。为了提升模型的泛化能力,训练集中引入了多种数据增强手段,包括缩放、翻转、颜色抖动和噪声添加等操作,增强后共 10 269 张。数据集共包含 7 类矿工行为,各个类别的样本数量如图 9 所示。依据行为的安全风险特征,将违规乘坐(illegal ride)、未佩戴安全帽(unhelmet)、翻越(climb over)、摔倒(tumble)等行为归为不安全行为,将弯腰(stoop)、坐着(sitting)、站立(standing)等常规行为归为安全行为。上述分类结合井下作业的典型行为模式与安全管理经验设定,有助于提升模型在多类别场景下的判别能力,为矿工行为预警与智能安全监管提供有效的数据支持。

2.2 实验环境

本实验在 Ubuntu20.04 操作系统上进行,编程环境为



图 8 部分数据集展示

Fig. 8 Partial dataset display

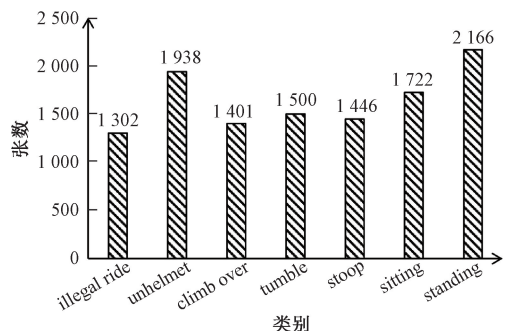


图 9 类别样本数量统计图

Fig. 9 Number of samples per category

Python3.9,基于 PyTorch 2.0.1 深度学习框架,CUDA 版本为 11.7,显卡为 RTX 3090 GPU,显存为 24 GB,模型训练过程中的参数如表 1 所示。

表 1 实验参数设置

Table 1 Experimental parameter settings

参数名称	数值
图片大小	640×640
批次大小	8
优化器	AdamW
学习率	0.001
训练轮数	200
动量系数	0.9

2.3 评价指标

本文选取平均精度均值(mean average precision, mAP)、模型参数量(parameters,Params)以及浮点运算次数(GFLOPs)作为模型性能评估指标,该 3 项指标分别从检测性能、模型规模和计算量 3 个维度对模型进行全面评估。

2.4 行为检测实验

为验证所提方法在矿工行为数据集上的有效性,本文采用 RT-DETR 与改进模型 IRT-DETR 进行了对比实验,

各类别检测精度如图 10 所示。由图 10 可知,IRT-DETR 整体精度较 RT-DETR 提升了 2.9%,且各行为类别均表现出不同程度的提升,其中,对于如“违规乘坐”、“翻越”等关键不安全行为类别,检测精度分别提升了 5%与 5.9%,显示出其在不安全行为检测中的显著优势,进一步验证了改进设计的有效性。

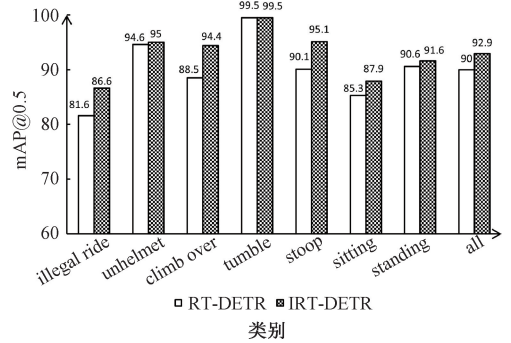


图 10 不同模型各类别精度对比

Fig.10 Per-class accuracy comparison of different models

2.5 消融实验

为验证本文问题驱动设计思路,以 RT-DETR-R18(主干网络为 ResNet-18)为基线模型,在矿工行为数据集上开展消融实验,系统性评估 3 个核心模块的针对性改进效果,结果如表 2 所示。

表 2 消融实验结果

Table 2 Results of ablation experiment

序号	RT-DETR	CANet	DAFPN	AKConv	mAP@0.5	mAP@0.5;0.95	GFLOPs	Params/M
1	✓				90.0	64.2	57.0	20.0
2	✓	✓			90.3	64.2	42.8	15.5
3	✓	✓	✓		91.9	65.4	50.2	17.2
4	✓	✓	✓	✓	92.9	66.1	49.6	16.4

实验表明,各模块均有效解决了其对应的核心难题:首先,采用基于重参数化与多路径融合的 CANet 替换原主干,通过将训练阶段的多分支卷积在推理时融合为单卷积,大幅减少计算量,在 mAP@0.5 提升的同时实现运算量降低 25%、参数量减少 22.5%。体现了该结构在复杂场景下兼顾特征提取效率与模型轻量化的优势。其次,为进一步应对行为尺度多样性,引入具备两阶段自适应融合与特征扩散机制的 DAFPn,使 mAP@0.5 进一步提升 1.6%,验证了其在多尺度特征融合方面的增强能力。最后,为缓解遮挡导致的关键特征丢失,将下采样模块替换为 AKConv,让卷积核能自适应地聚焦于关键特征区域,自动忽略不重要的背景。在 mAP@0.5 再提升 1%的同时进一步降低计算开销,凸显了其动态感知机制在复杂环境下的有效性。

综上,通过 3 个模块的协同优化,IRT-DETR 在精度

与效率方面均取得显著提升:mAP@0.5 提升 2.9%,mAP@0.5;0.95 提升 1.9%,参数量减少 18%,计算量降低 13%,实现了面向矿井场景的高性能检测。

2.6 对比实验

1) 不同特征提取模块对比实验

为解决主干网络在矿井场景下面临的特征提取能力与计算效率难以兼顾问题,本文对重参数化技术进行了验证。选取 OREPA (online re-parameterization)、DBB (diverse branch block)与 DRB(dilated reparam block)典型模块进行验证。如表 3 所示,OREPA 与 DRB 虽有效降低了模型复杂度,但在复杂背景下的特征提取能力不足,导致精度下降;而 Conv3XC 通过独特训练期多分支、推理期单分支的重参数化策略,实现了性能与效率的最佳平衡,在 mAP@0.5 上表现最优,验证了多分支特征提取作为主干网络组成的优势。

表 3 不同特征提取模块对比实验结果

Table 3 Comparative experimental results of different feature extraction modules

模型	mAP@0.5	mAP@0.5;0.95	Params/M	GFLOPs
ResNet18(baseline)	90.0	64.2	57.0	20.0
+OREPA	89.8	63.4	42.8	15.5
+DBB	89.6	62.0	42.8	15.5
+DRB	89.1	63.1	42.3	15.1
+Conv3XC(本文)	90.3	64.2	42.8	15.5

2) 不同特征融合模块对比实验

为验证 DAFPN 在应对矿工行为多尺度特性方面的有效性,本文将其与主流特征融合网络进行对比分析。如表 4 所示,相较于 CCFF、CSFCN 等依赖固定加权策略且缺乏扩散机制的融合方法,DAFPN 通过设计的两

阶段融合—扩散机制,实现了多尺度特征的自适应感知与逐层扩散融合。实验结果表明,DAFPN 在仅略微增加参数数量的前提下,使 mAP@0.5 显著提升 1.5%,充分验证了该机制在井下复杂尺度变化场景中的特征建模优势。

表 4 不同特征融合模块对比实验结果

Table 4 Comparative experimental results of different feature fusion modules

模型	mAP@0.5	mAP@0.5;0.95	Params/M	GFLOPs
CCFF(baseline)	90.0	64.2	57.0	20.0
+CSFCN	90.7	63.2	65.5	21.1
+GoldYOLO	89.7	64.8	60.0	22.3
+MAFPN	90.4	64.9	56.3	22.9
+DAFPN(本文)	91.5	65.5	63.4	21.7

3) 不同下采样模块对比实验

为克服井下设备遮挡与行为形变导致特征不完整的难题,本文引入具有动态感受野的 AKConv 模块。区别于 Conv、HWD 等依赖固定结构的卷积方式,AKConv 通过自

适应的采样点偏移机制,能够主动规避遮挡区域并聚焦有效特征区域。由表 5 可知,在相近计算开销下,仅 AKConv 实现 mAP@0.5 的稳定提升,验证了其可变核机制在应对井下不规则遮挡与形变问题中的独特优势。

表 5 不同下采样模块网络对比实验结果

Table 5 Comparative experimental results of networks with different downsampling modules

模型	mAP@0.5	mAP@0.5;0.95	Params/M	GFLOPs
Conv(baseline)	90.0	64.2	57.0	20.0
+HWD	89.1	64.6	55.7	19.2
+ContextGuidedDown	89.7	64.8	60.0	22.3
+SRFD	89.6	64.8	55.0	19.6
+AKConv(本文)	90.8	64.8	55.5	19.1

4) 不同模型对比实验

为了验证本文算法改进的有效性,将改进后的模型与其他主流检测模型进行对比,对比结果如表 6 所示。结果表明,IRT-DETR 在精度、参数量、计算开销等多个关键指标上均表现优越。相较于 Faster R-CNN 和 SSD,IRT-DETR 在各项指标上全面领先,与 YOLO 系列相比,其 mAP@0.5 分别较 YOLOv5m、YOLOv8m、YOLOv10m 和 YOLOv11m 提高了 8.5%、5.3%、6.4% 和 3.8%。尽管与 RT-DETR-R34 的检测精度接近,但 IRT-DETR 在参

数量和计算开销方面更具优势。此外,与该领域的最新研究 YOLO-EWS 和 YOLO-DCF 相比,本文算法在 mAP@0.5 与效率上均表现更优;相较于 YOLO-EWS,本文模型在 mAP@0.5 最高的同时计算量更小;相较于 YOLO-DCF,在计算量和参数量相近的条件下,本文模型依然取得了最高 mAP@0.5。进一步地,与采用相似方法的 RT-DETR-MLD 相比,虽然其通过轻量化降低了计算开销,但精度明显下降。综上所述,IRT-DETR 在保持较低资源消耗的同时,实现了更强的检测性能,展现出突出的综合优势。

表 6 不同模型对比实验结果

Table 6 Comparative experimental results of different models

模型	mAP@0.5	mAP@0.5;0.95	Params/M	GFLOPs
Faster-RCNN	79.3	42.8	136.8	369.9
SSD	86.4	52.9	24.4	61.4
YOLOv5m	84.4	55.4	21.0	50.3
YOLOv8m	87.6	58.4	25.8	75.7
YOLOv10m	86.5	59.5	16.5	63.5
YOLO11m	89.1	60.4	20.0	67.7
YOLOv8m-EWS ^[30]	91.2	60.7	23.3	72.3
YOLOv8s-DCF ^[12]	91.8	62.9	17.1	48.4
RTDETR-MLD ^[31]	87.6	61.5	9.7	26.3
RT-DETR-R18	90.0	64.2	20.0	57.0
RT-DETR-R34	90.6	65.2	31.0	88.8
IRT-DETR	92.9	66.1	16.4	49.6

2.7 可视化分析

图 11~14 展示了不同算法在矿工行为检测任务中的可视化对比结果。图 11 中,在人员尺度变化显著的场景下,IRT-DETR 下依然保持较高精度,而 YOLOv8m 出现漏检,稳定性较差。图 12 中在远距离与复杂背景场景中,YOLOv8m 将“坐着”误识别为“站着”,YOLOv11m 则存

在漏检,而 IRT-DETR 在此类场景下表现更为稳健。图 13 中粉尘干扰环境下 IRT-DETR 的置信度最高,表现最好。图 14 中,面对遮挡情况,IRT-DETR 仍保持较高的置信度与准确性。综上,IRT-DETR 在复杂背景、人员尺度变化及遮挡等多种挑战下均展现出更强的检测性能与鲁棒性。

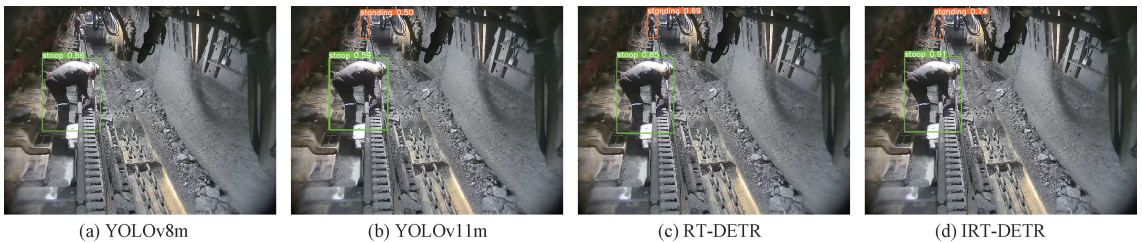


图 11 不同尺度下的行为检测结果对比

Fig. 11 Comparison of target behavior detection results at different scales

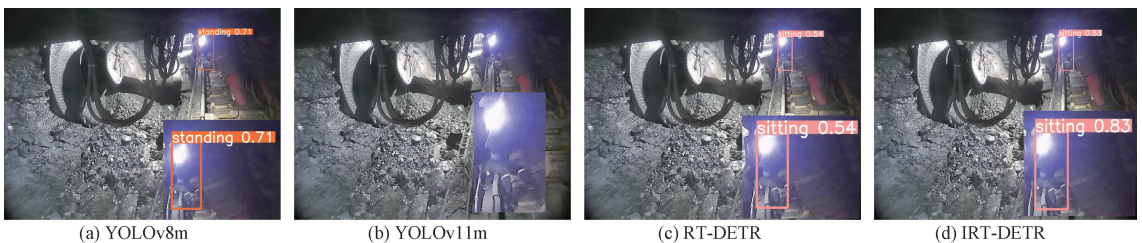


图 12 复杂背景下的行为检测结果对比

Fig. 12 Comparison of behavior detection results in complex backgrounds

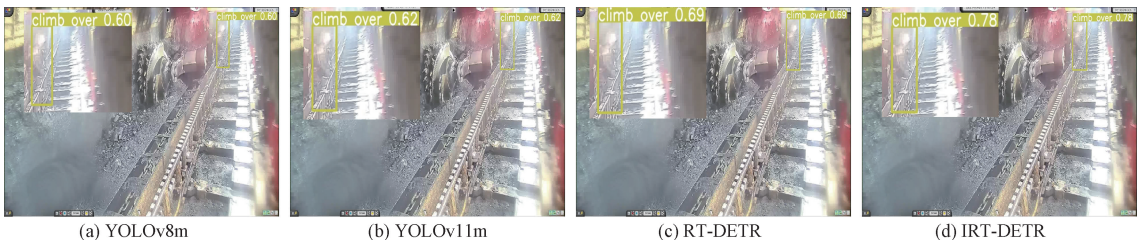


图 13 粉尘干扰环境下的行为检测结果对比

Fig. 13 Comparison of behavior detection results in a dust-disturbed environment

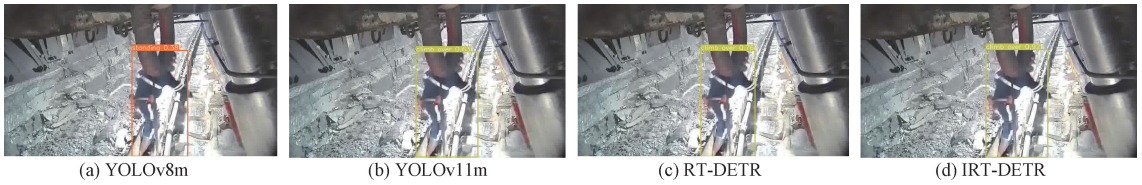


图 14 部分遮挡条件下的行为检测结果对比

Fig. 14 Comparison of behavior detection results under partial occlusion conditions

3 结 论

针对煤矿井下环境中存在的背景复杂、人员尺度变化大及频繁遮挡等问题,本文在 RT-DETR 模型基础上进行了系统性结构优化。通过设计主干网络 CANet,增强模型在复杂场景下对矿工关键行为特征的提取能力;构建结合维度感知与扩散机制的 DAFPN 特征融合网络,提高在多尺度与姿态变化下的特征表达与信息融合效果;并引入 AKConv 下采样模块,使模型能够动态感知行为形态与位置偏移,从而强化对遮挡区域中关键行为的捕捉能力。改进后的模型在矿工行为数据集上取得显著性能提升, $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别较基准模型提高 2.9% 和 1.9%,同时参数量减少 18%,计算量降低 13%,实现了检测精度与效率的双重优化。在复杂井下场景中,IRT-DETR 在不同尺度、粉尘干扰及局部遮挡等条件下均表现出良好的检测性能。未来工作将进一步探索模型的轻量化与场景自适应优化,以推动该方法在煤矿井下安全监测系统中的工程化应用。

参考文献

- [1] 王国法,任世华,庞义辉,等.煤炭工业“十三五”发展成效与“双碳”目标实施路径[J].煤炭科学技术,2021,49(9):1-8.
WANG G F, REN SH H, PANG Y H, et al. Development achievements of China's coal industry during the 13th Five-Year Plan period and implementation path of the “dual carbon” target[J]. Coal Science and Technology, 2021, 49(9):1-8.
- [2] 王海军,齐庆杰,梁运涛,等.我国煤矿重特大事故统计分析及对策建议[J].中国安全科学学报,2024,34(9):9-18.
WANG H J, QI Q J, LIANG Y T, et al. Statistical analysis and counter measure of major accidents in coal mines in China [J]. China Safety Science Journal, 2024, 34(9): 9-18.
- [3] 苏晨阳,武文红,牛恒茂,等.深度学习的工人多种不安全行为检测方法综述[J].计算机工程与应用,2024,60(5):30-46.
SU CH Y, WU W H, NIU H M, et al. Review of deep learning approaches for recognizing multiple unsafe behaviors of workers [J]. Computer

- Engineering and Applications, 2024, 60(5): 30-46.
- [4] CAO X G, ZHANG CH Y, WANG P, et al. Unsafe mining behavior identification method based on an improved ST-GCN[J]. Sustainability, 2023, 15(2): 1041.
- [5] 韩康,李敬兆,陶荣颖.基于改进 YOLOv7 和 ByteTrack 的煤矿关键岗位人员不安全行为识别[J].工矿自动化,2024,50(3):82-91.
HAN K, LI J ZH, TAO R Y. Recognition of unsafe behaviors of key position personnel in coal mines based on improved YOLOv7 and ByteTrack [J]. Journal of Mine Automation, 2024, 50(3): 82-91.
- [6] 王建芳,段思源,潘红光,等.轻量化姿态估计时空增强图卷积模型下的矿工行为识别[J].工矿自动化,2024,50(11):34-42.
WANG J F, DUAN S Y, PAN H G, et al. Lightweight pose estimation spatial-temporal enhanced graph convolutional model for miner behavior recognition [J]. Journal of Mine Automation, 2024, 50(11): 34-42.
- [7] 李雯静,刘鑫.基于深度学习的井下人员不安全行为识别与预警系统研究[J].金属矿山,2023(3):177-184.
LI W J, LIU X. Research on underground personnel unsafe behavior identification and early warning system based on deep learning [J]. Metal Mine, 2023 (3): 177-184.
- [8] 李占利,权锦成,靳红梅.基于 3D-Attention 与多尺度的矿井人员行为检测算法[J].国外电子测量技术,2023,42(7):95-104.
LI ZH L, QUAN J CH, JIN H M. Mine personnel behavior recognition algorithm based on 3D-Attention and multi-scale [J]. Foreign Electronic Measurement Technology, 2023, 42(7): 95-104.
- [9] WANG ZH, LIU Y, YANG Y, et al. Dual-branch deep learning architecture enabling miner behavior recognition [J]. Multimedia Tools and Applications, 2024, 83(37): 84523-84538.
- [10] WANG Y, CHEN X Q, LI J Q, et al. Convolutional block attention module multimodal feature-fusion action recognition: Enabling miner unsafe action recognition [J]. Sensors, 2024, 24(14): 4557.
- [11] XIN F F, HE X Y, YAO CH X, et al. A real-time

- detection for miner behavior via DYS-YOLOv8n model[J]. Journal of Real-Time Image Processing, 2024, 21(3): 92.
- [12] 董芳凯,赵美卿,黄伟龙.煤矿井下暗光环境人员行为检测研究[J].工矿自动化,2025,51(1):21-30,144.
DONG F K, ZHAO M Q, HUANG W L. Research on mine worker behavior detection in low-light underground coalmine environments [J]. Journal of Mine Automation, 2025, 51(1): 21-30, 144.
- [13] 温廷新,王贵通,孔祥博,等.基于迁移学习与残差网络的矿工不安全行为检测[J].中国安全科学学报,2020, 30(3):41-46.
WEN T X, WANG G T, KONG X B, et al. Identification of miners' unsafe behaviors based on transfer learning and residual network [J]. China Safety Science Journal, 2020, 30(3): 41-46.
- [14] WANG ZH, LIU Y, DUAN S Y, et al. An efficient detection of non-standard miner behavior using improved YOLOv8 [J]. Computers and Electrical Engineering, 2023, 112: 109021.
- [15] 陈伟,江志成,田子建,等.基于 YOLOv8 的煤矿井下人员不安全动作检测算法[J].煤炭科学技术,2024, 52(S2):267-283.
CHEN W, JIANG ZH CH, TIAN Z J, et al. Unsafe action detection algorithm of underground personnel in coal mine based on YOLOv8 [J]. Coal Science and Technology, 2024, 52(S2): 267-283.
- [16] 骆津津,陈伟,田子建,等.基于 YOLOv8-ECW 的井下人员行为实时检测算法[J].矿业科学学报,2025, 10(2):316-327.
LUO J J, CHEN W, TIAN Z J, et al. Real-time detection algorithm for underground personnel behavior based on YOLOv8-ECW [J]. Journal of Mining Science and Technology, 2025, 10 (2): 316-327.
- [17] ZHAO Y AN, LYU W Y, XU SH L, et al. DETRs beat yolos on real-time object detection [C]. IEEE/ CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [18] 孙光灵,王薪博,李艳秋.改进 RT-DETR 的双轮车头盔检测算法[J].电子测量与仪器学报,2025,39(4): 62-73.
SUN G L, WANG X B, LI Y Q. Improved helmet detection algorithm for two-wheeled vehicles of RT-DETR [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(4): 62-73.
- [19] 张永宏,许鑫豪,尹贺峰,等. MFSF-DETR: 一种基于多尺度特征移位融合的 PCB 缺陷与元件检测算法[J]. 仪器仪表学报, 2025, 46(8): 266-285.
ZHANG Y H, XU X H, YIN H F, et al. MFSF-DETR: A PCB defect and component detection algorithm based on multi-scale feature shift fusion [J]. Chinese Journal of Scientific Instrument, 2025, 46(8): 266-285.
- [20] 秦伦明,张云起,崔昊杨,等.基于改进 RT-DETR 的极端天气下交通标志检测方法[J].电子测量技术,2025, 48(9):56-64.
QIN L M, ZHANG Y Q, CUI H Y, et al. Improved RT-DETR based method for traffic sign recognition in extreme weather [J]. Electronic Measurement Technology, 2025, 48(9): 56-64.
- [21] 陈星星,李生林,周香伶.改进 RT-DETR 的水下目标检测算法[J].激光与光电子学进展,2025,62(24): 352-361.
CHEN X X, LI SH L, ZHOU X L, Improved underwater object detection algorithm based on RT-DETR [J]. Laser & Optoelectronics Progress, 2025, 62(24): 352-361.
- [22] 徐仁祥,曾勇,卢倩,等.基于改进 RT-DETR 的车门内拉手表面缺陷检测方法[J].电子测量技术,2024, 47(18):172-181.
XU Q X, ZENG Y, LU Q, et al. Surface defect detection method for inner handle of car door based on improved RT-DETR [J]. Electronic Measurement Technology, 2024, 47(18): 172-181.
- [23] 李青云,魏佳.改进 RT-DETR 的密集行人检测算法[J].电子测量技术,2025,48(21):148-156.
LI Q Y, WEI J. Improvement of the dense pedestrian detection algorithm of RT-DETR [J]. Electronic Measurement Technology, 2025, 48(21): 148-156.
- [24] WANG C Y, YE H I H, MARK L H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 1-21.
- [25] XU SH B, ZHENG SH CH, XU W H, et al. Hcf-net: Hierarchical context fusion network for infrared small object detection [C]. 2024 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2024: 1-6.
- [26] ZHANG X, SONG Y Z, SONG T T, et al. AKConv: Convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters [J]. ArXiv preprint arXiv:2311.11587, 2023.
- [27] WAN CH, YU H Y, LI ZH Q, et al. Swift parameter-free attention network for efficient super-

- resolution[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 6246-6256.
- [28] YANG W J, ZHANG X H, MA B, et al. An open dataset for intelligent recognition and classification of abnormal condition in longwall mining[J]. Scientific Data, 2023, 10(1): 416.
- [29] 程德强, 钱建生, 郭星歌, 等. 煤矿安全生产视频 AI 识别关键技术研究综述[J]. 煤炭科学技术, 2023, 51(2): 349-365.
- CHENG D Q, QIAN J SH, GUO X G, et al. Review on key technologies of AI recognition for videos in coal mine[J]. Coal Science and Technology, 2023, 51(2): 349-365.
- [30] 陈伟, 穆华星, 管彦允, 等. 改进 YOLOv8s 的煤矿井下矿工行为检测方法[J]. 辽宁工程技术大学学报(自然科学版), 2025, 44(3): 257-264.
- CHEN W, MU H X, GUAN Y Y, et al. Improving YOLOv8s for behavior detection of underground miners in coal mine[J]. Journal of Liaoning Technical University(Natural Science), 2025, 44(3): 257-264.
- [31] LIN L H, YANG H D, XU Q CH, et al. Research on student classroom behavior detection based on the real-time detection transformer algorithm [J]. Applied Sciences, 2024, 14(14): 6153.

作者简介

张红, 博士, 副教授, 研究生导师, 主要研究方向为计算机视觉。

E-mail: 757485672@qq.com

陈晓彤(通信作者), 硕士研究生, 主要研究方向为矿工行为检测。

E-mail: 1722739389@qq.com

许永炎, 硕士研究生, 主要研究方向为目标检测、缺陷检测。

E-mail: 1296743227@qq.com

高玺程, 硕士研究生, 主要研究方向为目标检测、图像处理。

E-mail: 2274396124@qq.com

王媛彬, 博士, 副教授, 研究生导师, 主要研究方向为矿井下视频监控与装备监测。

E-mail: wangyb998@163.com