



## 基于多尺度特征融合井下猴车载人状态的智能识别算法与应用

解北京 李恒 董航 栾铮 张奔 李晓旭

### 引用本文:

解北京, 李恒, 董航, 等. 基于多尺度特征融合井下猴车载人状态的智能识别算法与应用[J]. 煤炭科学技术, 2024, 52(12): 272–286.

XIE Beijing, LI Heng, DONG Hang. Intelligent recognition algorithm and application of coal mine overhead passenger device based on multiscale feature fusion[J]. Coal Science and Technology, 2024, 52(12): 272–286.

在线阅读 View online: <https://doi.org/10.12438/cst.2024-0081>

### 您可能感兴趣的其他文章

#### Articles you may be interested in

#### 融合金字塔结构与注意力机制的煤矿井下巡检机器人PT目标检测算法

Research on mine underground inspection robot target detection algorithm based on pyramid structure and attention mechanism coupling

煤炭科学技术. 2024, 52(6): 206–215 <https://doi.org/10.12438/cst.2023-1071>

#### 改进YOLOv5s的采煤机滚筒与支架护帮板干涉状态智能识别

Interference state intelligent recognition method for shearer drum and hydraulic support guard plate of improved YOLOv5s algorithm

煤炭科学技术. 2024, 52(2): 253–263 <https://doi.org/10.12438/cst.2023-0757>

#### 基于CSPNet-YOLOv7目标检测算法的煤矸图像识别模型

Coal gangue image recognition model based on CSPNet-YOLOv7 target detection algorithm

煤炭科学技术. 2024, 52(S1): 238–248 <https://doi.org/10.12438/cst.2023-0546>

#### 基于CFS-YOLO算法的复杂工况环境下煤矸图像识别方法

Image recognition method of coal gangue in complex working conditions based on CES-YOLO algorithm

煤炭科学技术. 2024, 52(6): 226–237 <https://doi.org/10.12438/cst.2023-1967>

#### 基于图像与点云融合的巷道锚护孔位识别定位方法

Roadway anchor hole recognition and positioning method based on image and point cloud fusion

煤炭科学技术. 2024, 52(5): 249–261 <https://doi.org/10.12438/cst.2023-1050>

#### 机器视觉灰度化金字塔卷积模型的煤流异物识别

Recognition of unwanted objects in coal flow based on gray pyramid convolution model of machine vision

煤炭科学技术. 2022, 50(11): 194–201 <http://www.mtkxjs.com.cn/article/id/fb838c5c-a1b8-4ba5-98a1-343a9352252a>



关注微信公众号, 获得更多资讯信息



移动扫码阅读

解北京,李恒,董航,等.基于多尺度特征融合井下猴车载人状态的智能识别算法与应用[J].煤炭科学技术,2024,52(12):272-286.

XIE Beijing, LI Heng, DONG Hang, *et al.* Intelligent recognition algorithm and application of coal mine overhead passenger device based on multiscale feature fusion[J]. Coal Science and Technology, 2024, 52(12): 272-286.

## 基于多尺度特征融合井下猴车载人状态的智能识别算法与应用

解北京<sup>1</sup>,李恒<sup>1</sup>,董航<sup>2</sup>,栾铮<sup>1</sup>,张奔<sup>1,3</sup>,李晓旭<sup>1</sup>

(1. 中国矿业大学(北京) 应急管理与安全工程学院,北京 100083; 2. 贵州盘南煤炭开发有限责任公司,贵州 盘州 553000;

3. 国家矿山安全监察局贵州局,贵州 贵阳 550001)

**摘要:**井下猴车智能识别技术是实现猴车自动化巡检、实时监测与预警任务的基础,是推动煤矿智能化发展的重要支撑。针对猴车数据集样本缺乏、井下图像光照环境较差、运行猴车之间存在重叠遮挡、矿工坐姿多变、小目标猴车难以检测、模型部署困难、不同载人状态井下猴车传统识别方法难度大效率低等问题,通过在贵州多个煤矿不同机段自建的猴车数据集,将猴车载人状态划分为载人猴车(HC\_miner)和无载人猴车(HC\_nominer)2种,提出了一种基于多尺度特征融合的井下猴车载人状态智能识别算法。图像预处理阶段,采用自适应直方图均衡以增强图像质量,通过随机矩形遮挡以模拟运行猴车被井下物体遮挡的真实场景,解决了猴车图像数据集体量匮乏的同时降低了井下负环境的干扰;特征提取阶段,将主干网络C2f模块部分卷积替换为可变形卷积(DCN),设计了一种C2f\_DCN模块,增加不同载人状态猴车目标感受野的动态调整能力以获取复杂多变的尺度信息,使模型更好地学习到猴车矿工的耦合特征及适应矿工各类坐姿细节,提升模型对不同载人状态猴车目标的辨识能力;特征融合阶段,提出了一种基于坐标注意力机制跨层级连接的路径聚合网络—CLC-PAN-CA模块,实现了深层网络与浅层网络特征间多尺度信息的复用,可自适应捕捉全局关键信息,建立网络之间的多尺度依赖,提升模型对小目标猴车重要特征的提取,减少背景噪声干扰,降低猴车目标漏检误检率。试验结果表明:提出模型的精确率为95.8%,对比基线模型提高了7.4%,召回率为93.3%,提高了9.8%,平均精度均值为95.6%,提升了7.7%,参数量和模型大小分别仅为 $3.1 \times 10^6$ 和6.1 MB,识别速率为71帧。对比多种主流单阶段两阶段检测模型,提出模型可有效辨识有无载人猴车目标、显著提升井下猴车目标识别精度、减少漏检错检现象、具有较快的识别速度、更好的热度信息提取能力,可满足实际场景巡检需求,为不同载人状态的井下猴车精准识别提供了可行的方法。最后,将提出的猴车智能识别算法和井下监控视频流嵌入到设计的猴车智能识别系统中,构思了井上调度和井下监控“端到端”一体化的猴车智能识别系统,增加了面向煤矿智能化巡检应用的期望值,可为井下猴车载人运输安全提供实时预警。

**关键词:**井下猴车识别;YOLOv8;可变形卷积;跨层级连接;坐标注意力机制

**中图分类号:**TD76 **文献标志码:**A **文章编号:**0253-2336(2024)12-0272-15

### Intelligent recognition algorithm and application of coal mine overhead passenger device based on multiscale feature fusion

XIE Beijing<sup>1</sup>, LI Heng<sup>1</sup>, DONG Hang<sup>2</sup>, LUAN Zheng<sup>1</sup>, ZHANG Ben<sup>1,3</sup>, LI Xiaoxu<sup>1</sup>

收稿日期:2024-01-16

策划编辑:常琛

责任编辑:刘雅清

DOI: 10.12438/cst.2024-0081

基金项目:国家重点研发计划资助项目(2022YFC2904100);中央高校基本科研业务费专项资金资助项目(2023ZKPYAQ04);安徽省爆破器材与技术工程实验室资助项目(AHBP2022A-03)

作者简介:解北京(1984—),男,安徽滁州人,副教授,博士。E-mail:bjxie1984@163.com

通讯作者:李恒(1997—),男,贵州盘州人,博士研究生。E-mail:18084282133@163.com

(1. School of Emergency Management and Safety Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China;

2. Guizhou Pannan Coal Development Co., Ltd., Panzhou 553000, China; 3. Guizhou Bureau of State Mine

Safety Supervision Bureau, Guiyang 550001, China)

**Abstract:** The intelligent recognition technology for Coal mine overhead passenger devices(Cmopd) plays a crucial role in achieving automated inspection, real-time monitoring, and warning tasks for cmopd, thereby promoting the intelligent development of coal mines. However, there are several challenges that need to be addressed, such as the limited number of samples in the cmopd dataset, poor lighting conditions in underground images, overlapping and occlusion between operating cmopd, varying sitting postures of miners, difficulty in detecting small cmopd targets, complex model deployment, and low efficiency of traditional recognition methods for cmopd with different passenger-carrying statuses. To overcome these challenges, a cmopd dataset was created from various coal mines in Guizhou province. The passenger-carrying status of cmopd was classified into two categories: cmopd with passengers (HC\_miner) and cmopd without passengers (HC\_nominer). The YOLOv8n single-stage object detection algorithm was used as the baseline model, and a coal mine cmopd intelligent recognition algorithm based on multi-scale feature fusion was proposed. In the image preprocessing stage, adaptive histogram equalization was employed to enhance image quality, and random rectangle masking was applied to simulate real scenarios where cmopd is occluded by underground objects during operation. This approach addressed the scarcity of cmopd image datasets and reduced the interference from negative underground environments. In the feature extraction stage, the partial convolution of the backbone network C2f module is replaced by deformable convolution, and a novel C2f\_DCN module is designed. This enhancement increased the dynamic adjustment capability of the target receptive field for cmopd with different passenger-carrying statuses, allowing the model to capture different scale information and better learn the coupled features of cmopd and miners. As a result, the model became more adaptable to various sitting postures of miners and improved its ability to identify cmopd targets with different passenger-carrying statuses. In the feature fusion stage, a path aggregation network with a coordinate attention mechanism (CLC-PAN-CA) was proposed to achieve cross-level concat of features and adaptively capture the contextual information of cmopd. The CLC-PAN-CA module effectively integrated multi-scale features and improved the accuracy of cmopd recognition. The experimental results show that the proposed model achieves a precision of 95.8%, which is 7.4% higher than the baseline model. The recall is 93.3%, representing an improvement of 9.8%, and the mean average precision is 95.6%, indicating a 7.7% increase. Furthermore, the model parameters and size are only  $3.1 \times 10^6$  and 6.1 MB, respectively. The recognition speed is 71 frames per second. Compare to a variety of mainstream single-stage two-stage detection models, the proposed model demonstrated effective identification of cmopd targets with and without passengers, significantly improved the accuracy of cmopd recognition, reduced false positives and false negatives, and exhibited faster recognition speed and better extraction of contextual information. The proposed algorithm can meet the requirements of practical inspection scenarios and provide a feasible method for accurate recognition of cmopd with different passenger-carrying statuses. Finally, the proposed cmopd intelligent recognition algorithm and the underground monitoring video stream were embedded into the designed cmopd intelligent recognition system. Partial implementation approaches for deploying the video media stream into the cmopd intelligent recognition system were provided. The concept of an end-to-end integrated cmopd intelligent recognition system, which integrates the dispatching system on the ground and the monitoring system underground, was proposed. This increases the expectations for intelligent inspection applications in coal mines and provides real-time warnings for the safe transportation of cmopd with passengers.

**Key words:** coal mine overhead passenger devices detection; YOLOv8; deformable convolution; cross-level concatenate; coordinate attention mechanism

## 0 引 言

煤矿智能化对提升国家能源安全、推动煤炭行业转型升级、提高科技创新能力具有重要的战略意义,也是我国煤炭工业高质量发展的必由之路<sup>[1-2]</sup>。煤矿架空乘人装置<sup>[3]</sup>(俗称猴车)是井下载人运输的重要装备之一,猴车智能巡检技术对提升煤矿安全性、保证载人运输效率、减人提效具有重要意义,是煤矿安全高效运输的重要保障<sup>[4]</sup>。受井下复杂环境和预警技术水平等限制,猴车事故时常频发,如2023年6月山西某煤矿猴车乘人运输事故造成3死、

1重伤、15轻伤,严重危害了矿工生命安全及煤矿安全高效生产,开展猴车智能识别技术的研究势在必行。

随着煤矿智能化建设不断加速推进<sup>[1,2,5]</sup>,国内外学者借助计算机视觉技术(Computer Vision)对井下图像数据集开展多场景应用的智能识别研究<sup>[6]</sup>。如YANG等<sup>[7]</sup>为填补井下数据集空白,基于大量现场监控视频,构建了地下综采工作面数据集,包括井下行人、安全帽、锚杆、液压支架等,并以YOLO等算法验证了构建数据集的有效性。WANG等<sup>[8]</sup>、YAN等<sup>[9]</sup>、刘普壮<sup>[10]</sup>、张磊等<sup>[11]</sup>、徐志强等<sup>[12]</sup>基于



井下图像数据集和搭建室内仿工业的检测平台构建了各类煤矸石粗粒度和细粒度的图像数据集,采用各种不同深度学习模型对井下煤矸石进行了识别,通过对比试验分析验证了提出算法的精确性和实时性。程德强等<sup>[13-15]</sup>针对井下图像模糊和边缘特征较差问题,采用众多图像超分辨率重构算法对构建的数据集进行了图像精准还原及细节重构,弱化了井下图像振铃伪影等劣化现象。郝帅等<sup>[16]</sup>基于YOLO算法对输送带异物检测模型进行了改进,模型对大块煤、矸石、锚杆等异物的检测精度可达94.7%,检测速率达到30 FPS,实现了输送带异物高精度实时检测。闫志蕊<sup>[17]</sup>基于DeeplabV3和迁移学习,采用更加轻量化的深度可分离卷积和注意力机制对模型进行了改进,可较好地适应实际煤岩图像分类任务的复杂性和实时性。然而,目前对井下猴车载人状态进行智能识别相关研究鲜有报道,传统猴车巡检是调度人员基于监控画面实时观测猴车运行状态,不仅增加了人工巡检成本,而且人为辨识难度极大,费时费力,亟需开展对不同载人状态的井下猴车目标进行精准快速智能识别研究。

对于猴车检测任务,需要考虑视频监控中目标存在的复杂多变的尺度信息和背景干扰等挑战,这与井下行人检测任务相似。李伟山等<sup>[18]</sup>对Faster-RCNN网络的候选区域结构进行改进,提出了一种高精度井下行人检测算法,增强了井下模糊、遮挡和小目标挑战下的检测精度。张明臻<sup>[19]</sup>设计了一种Dense-YOLOv3网络的井下行人检测网络,通过引入残差块解决了网络提出消失等问题,降低了模型漏检率。邵小强等<sup>[20]</sup>以YOLOv5为基线模型,在主干网络引入ShuffleNetV2和Transformer自注意力模块,设计了一种低开销且轻量级的行人检测算法。

区别于行人检测,猴车载人状态的识别是针对猴车与矿工耦合目标的检测任务。然而,上述研究对于检测任务仍存在如下不足:Faster-RCNN两阶段检测网络检测性能较好,但模型体积过大,检测速度过慢,对于受限的煤矿边缘设备难以实现应用部署;YOLO系列模型中传统卷积中固定网格无法准确捕获矿工及载人猴车复杂的空间变化特征;PAN特征融合网络中,减少了高层特征信息损失,但不同层级间的猴车及矿工多尺度特征融合能力仍然欠缺。

基于上述问题,通过收集贵州多个煤矿井下猴车不同机段的监控视频,构建了井下猴车目标检测图像数据集,通过引入数据增强、设计的C2f\_DCN模块、跨层级连接的(Cross-level concat)PAN-CA模块对数据集及YOLOv8模型进行优化,提出了一种基于多尺度特征融合的井下猴车载人状态视觉识别算法,解决了井下猴车数据集样本量较贫、图像质量差、运行猴车之间存在重叠遮挡、矿工坐姿多变、数据集中小目标猴车容易漏检等问题。依托Pyside6库、提出的改进模型和对监控视频流解码,设计出了适合井下猴车智能识别检测的可视化界面,为煤矿智能巡检任务提供有效的决策支持,优化煤矿管理策略。提出的模型和部署平台可有效辨识不同载人状态的猴车目标,有着更好的识别精度,较快的识别速率,较好的部署效果,可满足井下巡检要求,促进了煤矿视觉巡检任务的研究和应用,加速算法的发展和部署落地,为煤矿智能化建设提供底层技术支持。

## 1 多尺度特征融合的井下猴车智能识别算法构建

多尺度特征融合的井下猴车载人状态智能识别算法构建具体思路及网络结构如图1所示。主要包

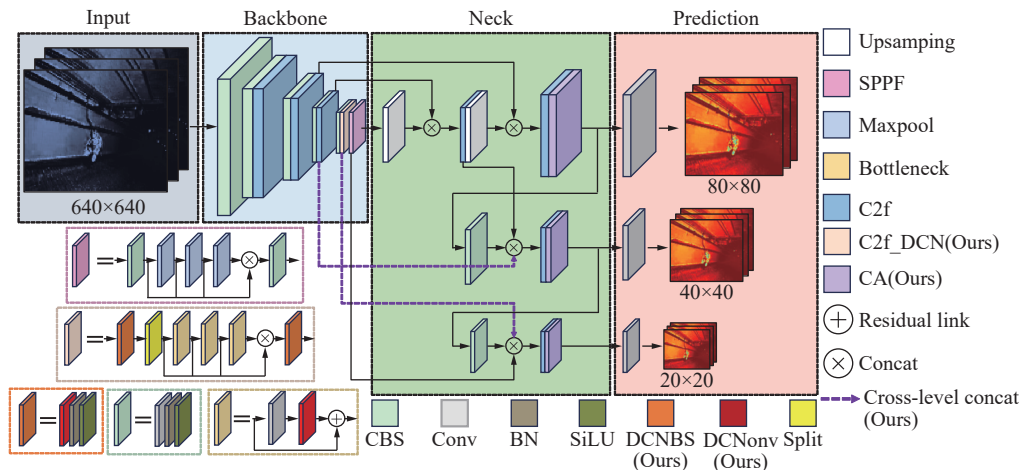


图1 井下猴车智能识别算法网络结构

Fig.1 Coal mine overhead passenger device intelligent recognition algorithm network structure

括猴车图像数据集采集及前处理过程、智能识别算法优化及构建。

**猴车图像数据集采集及前处理：**在贵州多个煤矿不同机段收集猴车图像、通过爬虫技术扩充数据集作为补充，实现多场景及普适性高的井下猴车图像数据集的构建；采用 Labelimg 标注软件对自建数据集进行不同载人状态猴车的标注；采用不同增强方式以增强图像对比度、模拟被遮挡真实场景、扩充图像数据集，为算法识别提供有效、足够的数据支撑。

**猴车智能识别算法优化及构建：**提出的算法是以单阶段目标检测算法 YOLOv8<sup>[21]</sup> 为基线构建，主要由主干网络、颈部层和检测头等关键模块构成（图 1），其中主干网络主要用于猴车图像的特征提取映射，颈部层主要作用是通过 PAN 特征金字塔操作来强化特征，增强模型多尺度检测能力，检测头主要用于预测特征映射中是否有载人猴车的边界框及标签等。检测头模块相比 YOLO 先前系列改动较大，由原先的耦合头变为目前主流的解耦头，将分类和检测头分离，同时 Anchor-Based 换成了 Anchor-Free 无锚点检测，无需对锚框进行预设处理，集成度更高，更适合实际部署。但井下环境复杂，实际巡检任务中，需要考虑多尺度因素开展不同载人状态的猴车目标识别研究，故需对基线模型进行一些重设计以提升模型在猴车识别任务中的辨识能力、检测精度及速度。

在特征提取阶段，设计了一种 C2f\_DCN 模块，增加不同运行状态猴车目标感受野的动态调整能力以获取不同尺度信息，解决猴车识别任务中存在的重叠遮挡、坐姿多变等空间尺度问题；特征融合阶段，在不同尺度特征图的检测头前引入坐标注意力机制以自适应捕捉全局关键信息，提出了跨层级连接的

PAN-CA 路径聚合方式，实现了网络之间的多尺度依赖，使得深层网络位置信息与浅层网络细节特征得到多尺度复用，增加模型对小目标猴车特征的提取，降低猴车目标漏检误检率，过滤背景干扰。各个模块设计细节及优化策略见下节。

### 1.1 C2f\_DCN 模块

主干网络的 C2f 模块主要负责将高层语义特征与低层细节特征进行融合，主要用于目标特征的提取。在实际猴车检测任务中，从空间尺度上看，存在如下问题：① 部分运行猴车之间存在遮挡，猴车的特征可能无法完全被观察到，模型难以准确区分遮挡车辆和其他物体，导致误检漏检；② 猴车在运行过程中会出现摆动，这会导致猴车在不同帧之间的位置和形状发生变化，使得模型难以捕捉到猴车的准确形状和位置信息，从而影响检测精度；③ 矿工的坐姿变化较大，使得模型在特征提取过程中难以准确地捕捉到猴车和矿工的耦合特征，辨识有无载人状态的猴车存在一定挑战。

针对这些问题，基于可变形卷积（Deformable convolution, DCN），笔者提出了一种融合形变尺度机制的新模块 C2f\_DCN。C2f\_DCN 网络结构如图 2 所示，由 DCNBS、Split、Bottleneck 等模块构成，有着更多的残差连接和分离操作，更丰富的梯度流。其中 DCNBS 模块由 DConv、Batchnorm 层和 SiLU 层构成。对于 Bottleneck 模块的设计，考虑到模型计算量问题，首先采用了 1 个  $1 \times 1$  的卷积将特征图通道数减半，然后再通过 1 个  $3 \times 3$  的可变形卷积以辅助模型增强猴车特征提取，并将通道数加倍，以对齐输入输出通道数，最后将输入输出特征进行残差连接，从而实现具有形变尺度机制的猴车特征融合。

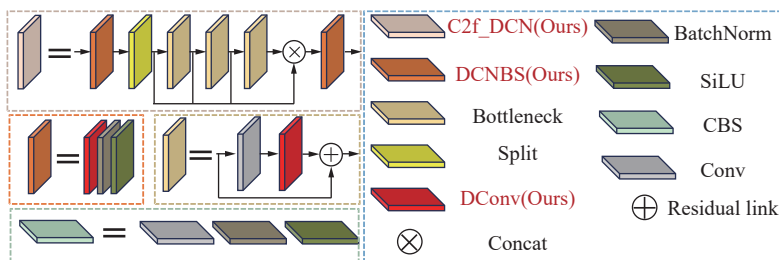


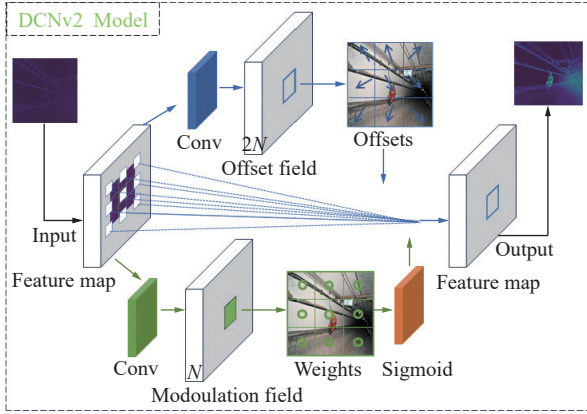
图 2 C2f\_DCN 网络结构

Fig.2 C2f\_DCN network structure

可变形卷积（Deformable Convolution, DCN）<sup>[22]</sup> 在传统卷积基础上引入了可学习偏移量，DCNv2 在学习偏移量的同时对学习到的特征幅度进行调制（调制标量），可有效解决传统卷积中固定网格无法

捕获检测对象复杂空间变化的问题。

DCNv2 模块（图 3）主要由 2 个独立的卷积块构成，总输出通道数为  $3N$ ，其中偏移层的卷积通过输入特征生成  $2N$  个通道数，对应  $N$  个二维偏移量

图3 可变形卷积网络结构<sup>[22]</sup>Fig.3 Deformable convolutional network structure<sup>[22]</sup>

( $x$ 方向和 $y$ 方向各1个),得到可学习偏移量;剩余的 $N$ 个通道送至sigmoid层得到调制标量,通道数 $N$ 对应卷积核感受野的权重。由此,在特征提取阶段将普通卷积替换为可变形卷积后,使模型自适应地调整卷积核尺寸,可有效增加感受野范围,提升模型对猴车的辨识能力及识别精度。可变形卷积计算式(1)如下:

$$y(p) = \sum_{k=1}^N w_k x(p + p_k + \Delta p_k) \Delta m_k \quad (1)$$

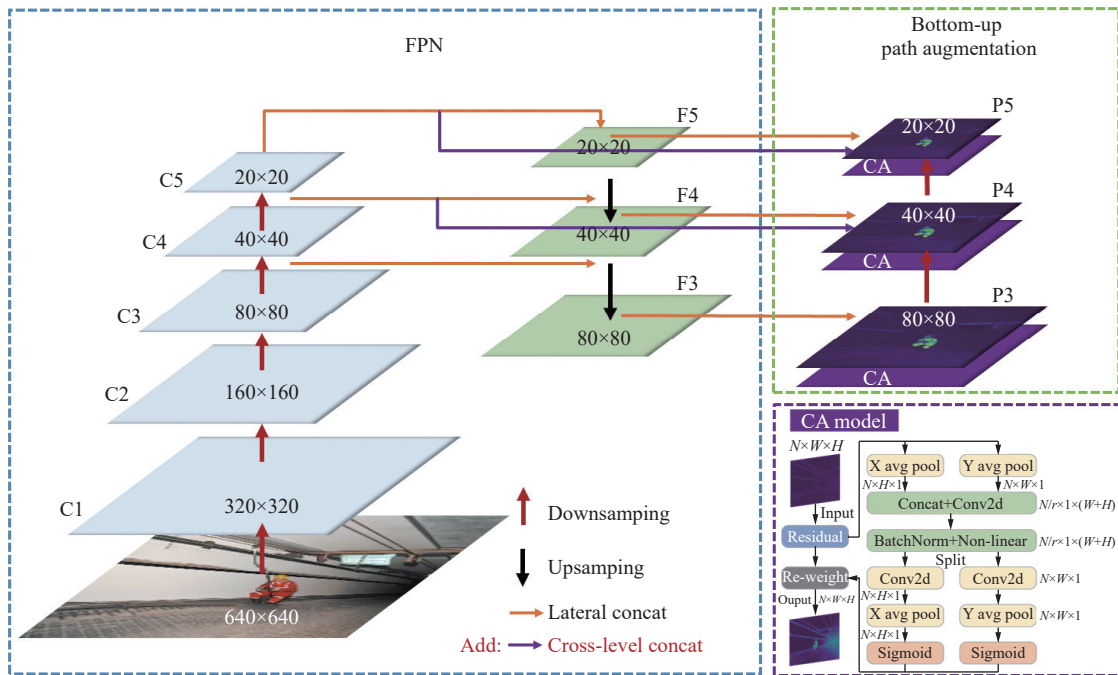
式中: $y(p)$ 为输出; $w_k$ 为卷积核第 $k$ 个位置的权重; $k$ 为卷积网格中表示位置的数字; $N$ 为特征图通道数; $x$ 为输入特征; $p$ 为输出特征图中的每个位置; $p_k$ 为

第 $k$ 个位置的元素; $\Delta p_k$ 为第 $k$ 个感受野位置的偏移量; $\Delta m_k$ 为第 $k$ 个位置的调制标量。

## 1.2 Cross-level concat PAN-CA 模块

YOLOv8中PAN<sup>[21]</sup>模块特征金字塔的构建是在FPN自上而下的上采样操作基础上增加了1条自底向上降采样操作的路径,可有效从低层特征图中获取目标细节信息。特征融合通过横向连接将不同层级的特征图进行融合,弥补了FPN在多尺度特征融合时部分信息丢失问题。但在实际猴车检测任务中,从时间尺度上看,存在如下问题:①监控视频不同帧间不同的猴车目标大小都不一致,Mosaic数据增强操作相当于变相扩充了数据集中小目标的样本数量;②机头机尾局部存在较大的背景干扰,PAN网络对关键位置信息的解码尚不够。在目标检测网络中浅层网络通常有较好的细节(颜色、纹理、边缘等抽象特征)表征能力,但语义抽象信息(如猴车类别、形状、位置信息等)表征较差,相反的深层网络有着较好的语义表征能力,而细节特征表征较弱,为进一步提升模型深层浅层多尺度信息复用能力,提出跨层级连接(Cross-level concat)和坐标注意力机制<sup>[23-24]</sup>(Coordinate Attention, CA)对PAN模块路径聚合网络进行优化。

构建的跨层级连接PAN-CA网络结构如图4所示,大小为 $640 \times 640$ 猴车输入图像经多个卷积下采样操作,生成5个不同尺度的特征图C1—C5,F3、



注: $r$ 为缩减参数,用于控制通道数的压缩比例。

图4 跨层级连接PAN-CA网络结构

Fig.4 Cross-Level Concat PAN-CA Network structure



F4 是通过 F4、F5 上采样得到的,  $40 \times 40$ 、 $20 \times 20$  特征图分别与具有同等分辨率的 C3、C4 拼接而成的新的特征图, 通过自上而下的路径及横向连接, 实现了深层语义特征的利用。P3、P4 和 P5 是通过轻量级 CA 注意力机制和卷积降采样操作得到的特征图与具有同等分辨率的 F3、F4、F5 拼接而成新的特征图, 在提升不同尺度的位置信息利用率的同时克服了 PAN 网络资源消耗较大等问题。在此基础上, 笔者提出了一种跨层级连接的方式, 通过复用深层浅层特征, 将深层网络语义信息丰富的 C5 和浅层网络细节特征丰富的 C4 分别与 P5、P4 在通道上进行张量拼接, 深层网络有着较好细节特征信息的同时, 浅层网络也能够有较好的语义信息。由此, 构建的跨层级连接 PAN-CA 网络中, 通过引入 CA 坐标注意力机制, 在模型的通道处嵌入坐标信息, 可自适应捕捉全局关键信息, 通过跨层级连接, 实现了多尺度特征信息的复用, 建立网络之间的多尺度依赖, 增强不同载人状态猴车的辨识能力, 提升小目标猴车检测能力, 降低噪声背景干扰。

CA<sup>[23-24]</sup> 是一种轻量级且高效的注意力机制, 主要通过嵌入位置信息到通道注意力, 从而使网络获取关键区域的信息且避免引入大的开销, 对于嵌入式设备几乎不产生额外的储存开销, 更适合模型部署。图 4 可见 CA 网络结构<sup>[23-24]</sup>, 首先对输入图像进行高度和宽度上 2 个方向的全局平均池化:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} |x_c(h, i)| \quad (2)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} |x_c(j, w)| \quad (3)$$

式中:  $z_c^h$  为高度为  $h$  的第  $c$  个通道输出;  $z_c^w$  为宽度为  $w$  的第  $c$  个通道输出;  $W$ 、 $H$  分别为特征图宽度和高

度;  $x_c$  为第  $c$  个通道的输入值;  $i$  为高度  $h$  时的宽度坐标值;  $j$  为宽度  $w$  时的高度坐标值。

然后将池化后的高度和宽度方向上的特征图进行拼接、归一化、卷积和非线性激活等操作, 随之通过分离操作将其分为宽度方向和高度方向的特征图, 每个图在对应空间方向上捕捉输入特征图的长程依赖关系, 从而保存位置信息。再对分离后的特征图分别进行卷积、平均池化和 sigmoid 激活函数等操作后, 得到宽度方向和高度方向的注意力权重  $g^w$  和  $g^h$ , 最后将这 2 个方向的注意力通过权重相乘应用于输入特征图, 在保留位置信息的同时, 可有效加强关键区域的表征, 降低背景噪声干扰。CA 计算式如下所示:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (4)$$

式中:  $y_c$  为 CA 模块的输出特征图;  $g_c^h$  为高度为  $h$  时第  $c$  个通道的注意力权重;  $g_c^w$  为宽度为  $w$  时第  $c$  个通道的注意力权重。

## 2 井下猴车数据集构建

自建井下猴车图像数据集来源于贵州多个煤矿猴车机头机中段机尾处, 采用 KBA165 矿用本安型防爆摄像头拍摄获取, 图像采集分辨率为  $1920 \times 1080$ , 摄像头电源供电方式为井下信号照明综保开关取 127 V 电源, 视频数据传输方式为光纤传输。图像采集地点包括 2 处猴车机头、2 处运输上山下山运行中部及 2 处猴车机尾处构建了具有多挑战任务的数据集(不同光照、遮挡、小目标), 机尾处巷道坡度为  $22^\circ \sim 29^\circ$ , 猴车运行速度为  $1.07 \sim 1.21$  m/s。根据采集到的猴车图像, 在不同运行机段选取猴车识别任务中存在的典型问题进行展示(测试样本中选取), 具体如图 5 所示。

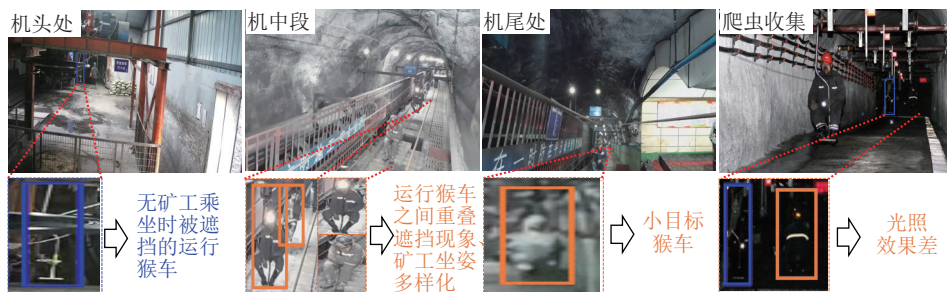


图 5 不同机段猴车图像构建示例及典型问题

Fig.5 Example of image construction of overhead passenger devices in coal mines in different segments

数据集的标注采用 Labelimg 标注软件, 将猴车图像标注为无人乘坐猴车状态(HC\_nominer)和有人乘坐猴车状态(HC\_miner)共计 2 类标签, 为保证数

据集构建的可靠及准确, 笔者和矿山猴车管理人员对构建的图像数据集和标注框信息进行了多次核查。此外, 由于领域内猴车数据集目前鲜见, 数据集虽收

集了多个矿不同运行地段的图像,但数据集体量较小会使得网络泛化能力略欠缺,为使得网络能够泛化到各类矿井猴车数据集中,采用爬虫技术,对“井下猴车”关键词进行检索,爬取了142张有效图像作为补充。最终得到748张猴车图像,其中包括606张自建数据集,142张补充数据集,按照7:2:1的比例划分为训练集、验证集和测试集。

由于井下复杂环境及工作面限制,大多学者<sup>[14-16]</sup>都以井下防爆摄像头作为获取图像数据集的主要途径,采集的图像数据大多存在低光照,模糊、遮挡严重等缺点,而数据集的质量决定了模型训练效果的上限,故通过图像增强(Data Augment)操作对构建的猴车数据集进行扩充,可得到多样、丰富的训练样本,增强了图像的随机性和多样性以减少模型过拟合风险。

考虑到井下环境煤尘低光照等影响,选用自适应直方图均衡<sup>[25]</sup>(Clahe)对井下图像对比度增强,弱化井下负环境影响;构建数据集中存在一定量的猴车被遮挡、重叠等现象,选用随机矩形框遮挡<sup>[26]</sup>(Cutout)来模拟识别目标被其他物体遮挡或部分遮挡的情况以增强对井下猴车的识别能力,使模型能够更好地适应各种复杂场景和处理遮挡情况;此外,还选用了常规增强垂直翻转(V-Flip)、水平翻转(H-Flip)对数据进行扩充。为确保数据集纯净度,只对训练样本进行增强,得到增强后训练集图像共计2 615张,随机选取中部1张图片为例,其增强效果如图6所示,本文的数据集分布如图7所示。

### 3 试验结果

试验在Win10 64位操作系统上进行,GPU为NVIDIA 4 070 ti,显存12 G,试验所用深度学习模型基于Python 3.9,Pytorch 2.0,Cuda 11.8框架搭建。模型输入尺寸为640×640,训练批处理大小设为16,核心数为8,最大训练轮次为300,优化器选用含动量随机梯度下降结合WarmUp(Adamw),学习率为0.001,动量设为0.937,权值衰减系数设置为0.000 5以控制正则化强度,在最后10个epoch关闭Mosaic操作以增强模型收敛稳定性。

#### 3.1 评价指标及损失函数

选用精确率(Precision,  $P$ )、召回率(Recall,  $R$ )、平均精度(Average precision,  $A^P$ )、平均精度均值( $m^{AP}$ )、每秒检测帧率(FPS)作为主要评判标准。其中 $P$ 为单个类别识别精确率,其值越大越好; $R$ 为单个类别识别查全率,其值越大,模型漏检率越低; $A^P$

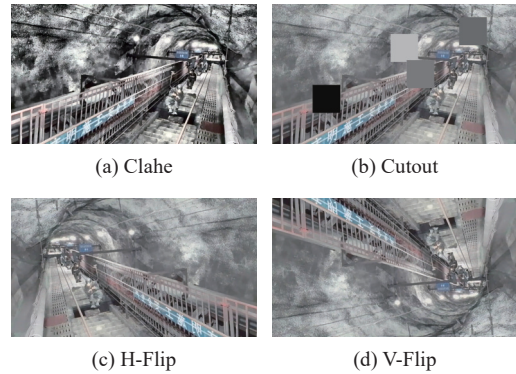
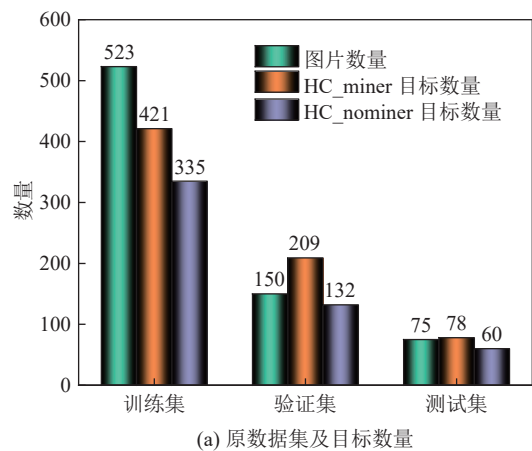
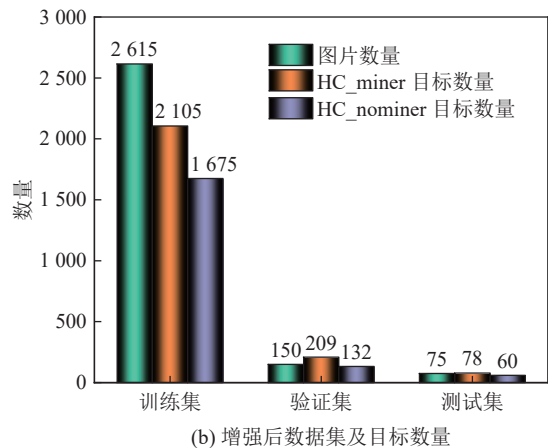


图6 猴车图像增强示例

Fig.6 Example of coal mine overhead passenger device image enhancement



(a) 原数据集及目标数量



(b) 增强后数据集及目标数量

图7 扩充前后的数据集及目标数量对比

Fig.7 Comparison of datasets and number of targets before and after expansion

为综合 $P$ 、 $R$ 的指标,用于衡量模型在不同类别的平均精度,其值越大越好; $m^{AP}$ 用于衡量模型在多个类别上的平均精度,综合考虑了所有类别的平均精度,其值越大,模型在识别任务中的精度就越高,选用检测框与真实标注框交并比阈值为0.5时<sup>[28]</sup>,计算 $m^{AP}$ ;FPS表示模型1 s内处理的图像帧数,其值越高,模型的识别速度就越快;



$P, R, A^P, m^{AP}$  式如下：

$$P = T^P / T^P + F^P \quad (5)$$

$$R = T^P / T^P + F^N \quad (6)$$

$$A^P = \int_0^1 P(R) d(R) \quad (7)$$

$$m^{AP} = \sum_{\lambda} A^{P_{\lambda}} / Z \quad (8)$$

式中： $P$ 、 $R$  分别为精准率和召回率，%； $T^P$ 、 $F^N$  和  $F^P$  分别为正检框、漏检框和误检框数目； $Z$  为检测类别是总数； $A^{P_{\lambda}}$  为第  $\lambda$  个类别的平均精度。

损失函数模块中，主要采用分类式和边界框回归解耦式组合损失，可有效提升模型权重利用率，使得标签能够更好指导模型对不同类别区分学习的能力。其中，分类损失采用二元交叉熵，每个类别判断为“是否为该类”，并输出置信度：

$$L_{BCE} = \frac{1}{S} \sum_i -[y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] \quad (9)$$

式中： $S$  为总样本数量； $y_i$  为第  $t$  个像素位置上模型预测值； $p_i$  为第  $t$  个像素位置上标注的实际值； $L_{BCE}$  为类别损失。

边界框回归损失采用  $C^{IOU} + D^{FL[27]}$  耦合损失， $C^{IOU}$  相比  $I^{OU}$  考虑了重叠面积、中心点距离、纵横比，解决了边界框不相交时 loss 等于 0、收敛慢、精度低等问题。该损失函数表示为

$$R_{Clou} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (10)$$

$$\alpha = \frac{v}{(1 - I^{OU}) + v} \quad (11)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (12)$$

$$L_{Clou} = 1 - I^{OU} + R_{Clou} \quad (13)$$

式中： $R_{Clou}$  为惩罚项； $I^{OU}$  为交并比； $L_{Clou}$  为总损失； $\rho^2(b)$  为预测框中心点欧氏距离； $\rho^2(b^{gt})$  为标注框中心点欧氏距离； $c$  为预测框和标注框外接矩形最小对角距离； $v$  为边界框长宽比； $\alpha$  为调节系数随  $v$  变化而变化； $w^{gt}$  和  $h^{gt}$  分别为标注框宽和高； $w$  和  $h$  分别为预测框宽和高。

$D^{FL}$  以交叉熵形式优化标签最接近的一左一右 2 个位置的概率，使网络更快的聚焦到目标位置及邻近区域的分布，能够增强模型在复杂情况下，如井下猴车被遮挡、猴车移动时的泛化性，式如下：

$$D^{FL}(S_u, S_{u+1}) = -[(y_{u+1} - y) \ln(S_u) + (y - y_u) \ln(S_{u+1})] \quad (14)$$

式中： $D^{FL}$  为分布焦点损失，关注的是边界框回归的分布差异； $y$  为实际值； $y_u$  和  $y_{u+1}$  为标签最接近的一左一右 2 个位置的预测值； $S_u$  和  $S_{u+1}$  分别为  $y_u$  和  $y_{u+1}$  对应的预测置信度。

### 3.2 基线模型试验

笔者侧重于在资源受限的部署场景下进行猴车识别，过大的模型在部署时对设备硬件要求较高、对网络带宽也较高，而在井下环境负效应影响下，难以实现较大模型和较大带宽网络部署。故网络结构在保持较高准确性的同时，需较低的计算和存储开销，在性能和计算效率之间能取得良好的平衡。对此，笔者对提出的改进模型在不同基线上 (n/s/m/l/x) 进行了试验，结果见表 1。

表 1 基准模型对比试验分析

Table 1 Baseline model comparison experiment analysis

基线	深度系数	宽度系数	参数量/ $10^6$	平均精度均值/%	检测帧数
/n	0.33	0.25	3.02	95.63	71
/s	0.33	0.50	11.02	95.95	59
/m	0.67	0.75	24.23	96.27	48
/l	1.00	1.00	41.25	96.21	35
/x	1.00	1.25	65.59	96.93	23

根据表 1 结果可知，随着网络宽度和深度的增加，YOLOv8-s/m/l/x 等模型的网络参数量也逐渐增加，对比 YOLOv8-n 模型，YOLOv8-s、YOLOv8-m、YOLOv8-l、YOLOv8-x 网络参数量增加了 3.57、7.86、13.39、21.29 倍，且推理速度有不同程度的下降，但精度最高为 96.93% 仅增加了约 1%。因此综合考虑到模型复杂度、推理速度、检测精度和实际应用部署，对于笔者自建的猴车数据集，选取 YOLOv8-n 网络结构作为基准模型进行改进较为合适。

### 3.3 消融试验

为验证提出及设计模块的有效性，笔者对数据增强 (DA)、特征提取优化 (C2f\_DCN)、特征融合路径优化 (Cross-level Concat, CA) 等模块逐一进行分析。模型 1 为直接采用 YOLOv8-n 预训练模型进行训练的结果，模型 2—模型 7 为单一消融或多元消融得到的结果，模型 8 为笔者提出模型的试验结果。总体改进模块的消融试验结果见表 2。

根据表 2 结果，相比基线模型，在采用单一模块对模型进行消融试验时，综合衡量模型检测精度的平均精度均值都得到了不同程度的涨点效果。采用了笔者的数据增强后，平均精度均值提升了 1.07%，验证了笔者设计的数据增强操作的有效性；添加坐标注意力模块后，模型浮点运算量和模型大小等衡

表 2 总体改进模块消融试验  
Table 2 Overall improvement module ablation experiment

模型	改进模块				平均精度均值/%	浮点运算量/ $10^9$	参数量/ $10^6$	模型大小/MB	检测帧数
	DA	C2f_DCN	Cross-level Concat	CA					
1					87.99	8.22	3.02	5.94	89
2	√				89.06	8.22	3.02	5.96	85
3		√			90.75	8.13	3.04	5.97	69
4			√		88.87	8.21	3.06	6.07	87
5				√	89.37	7.98	3.02	5.87	78
6			√	√	92.38	8.17	3.07	5.98	79
7		√	√	√	93.74	8.21	3.09	6.09	72
8(Ours)	√	√	√	√	95.63	8.23	3.12	6.12	71

量复杂度指标降低的同时,平均精度均值提升了 1.38%;平均精度均值提升效果最优的为设计的 C2f\_DCN 模块,在添加了可变形卷积核后,模型能够动态调整卷积核尺寸,适应不同尺寸大小及形变尺度的猴车目标,精度提升了 2.76%。

对于复合模块的改进,单一采用跨层级连接的方式对精度提升效果最低(模型 4),但在融合了注意力模块后(模型 6),不同尺度特征图各方向的权重得到了聚合增强,增加了模型位置信息定位能力,使得模型深层浅层猴车关键信息得到进一步复用,模型精度提升了 4.39%。在采用合适的数据增强、使用设计的 C2f\_DCN 模块、优化特征融合路径、融合坐标注意力模块后,提出的模型 8 的平均精度均值为最高(95.63%),提升了 7.64%,模型复杂度仅有略微增幅,每 s 可检测 71 张猴车图像,可满足对于实际巡检任务需求,进一步说明了提出模型在猴车识别任务中的有效性。表 2 中结果进一步表明提出及设计的模块对模型均有提升效果,且复合模块改进效果优于单个模块改进效果。

4 讨 论

4.1 主流模型对比试验

为进一步验证提出的猴车检测算法的有效性,笔者在 mmdetection 平台和 ultralytics 库进行了多种主流 SOTA 检测算法的对比试验。其中,两阶段的目标检测算法有 Cascade-RCNN<sup>[28]</sup> 和 Faster-RCNN<sup>[29]</sup> 模型,输入尺寸和主干网络-特征融合都分别为 1 333×800, ResNet50-FPN。单阶段的检测算法有 SSD<sup>[30]</sup>、Retina-net<sup>[31]</sup>、YOLOv5-n<sup>[9]</sup>、YOLOv6-s<sup>[32]</sup>、YOLOX-tiny<sup>[33]</sup>、YOLOv8-n<sup>[21]</sup>,对应的输入尺寸和主干网络-特征融合方式见表 3,对比试验的训练均在同等数据集上进行,轮次均为 300,指标通过最优权重得到。笔者提出模型的主干网络是基于 CSPDarknet53 和设计的 C2f\_DCN 构建,特征融合方式基于 PAN 提出了融合坐标注意力机制的跨层级连接路径(Cross-level concat-PAN-CA),在表 3 中将提出的主干网络-特征融合记为 CSP-CLC-PAN-CA。

通过表 3 对比试验结果可知,两阶段的 Cascade-RCNN 和 Faster-RCNN 模型检测精度都较高,但浮

表 3 主流模型的对比试验  
Table 3 Comparative experiments of mainstream models

类型	模型	输入尺寸	平均精度均值/%	浮点运算量/ $10^9$	参数量/ $10^6$	模型大小/MB	检测帧数
两阶段	Cascade-RCNN	1 333×800	94.2	236.2	69.2	542.3	18.8
	Faster-RCNN	1 333×800	93.7	204.8	41.3	351.8	25.0
单阶段	SSD	300×300	79.6	30.5	23.8	182.2	52.3
	Retina-net	1 333×800	83.8	152.7	19.7	150.3	41.8
	YOLOv5-n	640×640	85.5	7.2	2.5	5.2	79.0
	YOLOv6-s	640×640	90.3	16.3	44.2	32.8	51.0
	YOLOX-tiny	416×416	81.8	3.2	5.1	59.7	107.0
	YOLOv8-n	640×640	89.1	8.3	3.0	6.0	85.0
	Ours	640×640	95.6	8.2	3.1	6.1	71.0

点运算量、网络参数及模型大小较大,远大于提出模型,并且检测速度都较慢,不到提出模型的 40%,难以实现模型的边缘部署。单阶段 SSD 和 Retina-net 模型的检测速度对比两阶段检测模型有较大提升,浮点运算量、网络参数及模型大小有所下降,但检测精度较差,远低于提出的模型。YOLO 系列的单阶段检测模型,有着更低的浮点运算量、网络参数及模型大小,检测速度最高可达 107,但检测精度值远不如两阶段的检测模型和提出的检测模型。笔者提出模型在自建的数据集上精准、高效地完成了猴车检测识别任务。对比多种两阶段检测模型,提出模型

的网络结构更为轻量化,检测速度更快。对比多种单阶段检测模型,提出模型的参数量和模型大小仅有略微增加的情况下,有着更高的检测精度。提出模型的检测精度为最优(95.6%),浮点运算量、网络参数及模型大小分别为 8.2 G、3.1 M、6.1 MB,每 s 可达 71 张井下猴车图像识别速度,可满足实际巡检需求,进一步说明了提出模型在猴车识别任务中的精确性和实时性。

图 8 给出了部分主流模型在测试集上的推理预测可视化效果。图 9 给出了提出模型在自建数据集的各类场景的猴车目标推理可视化结果。根据

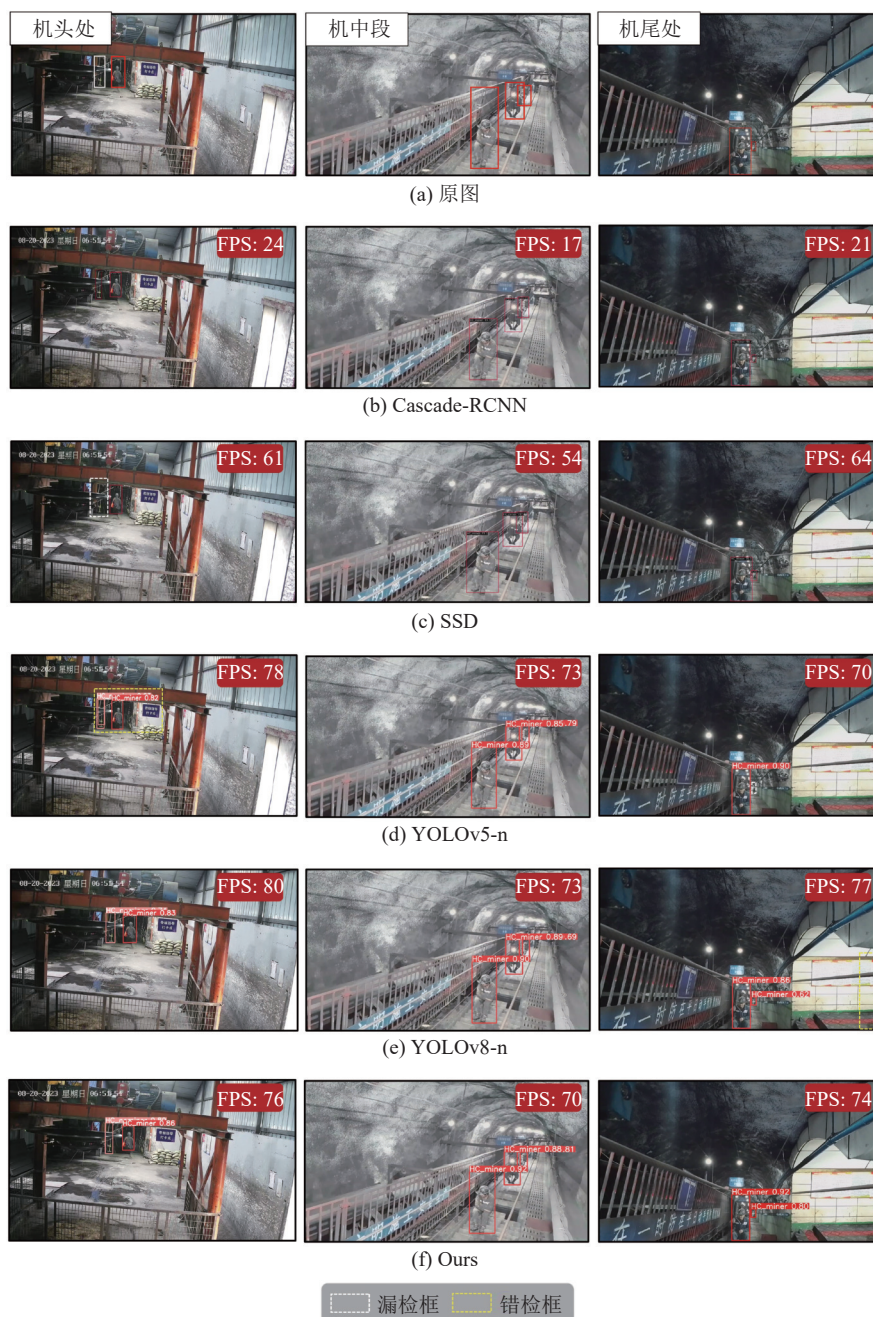


图 8 主流模型推理可视化

Fig.8 Mainstream model inference visualization



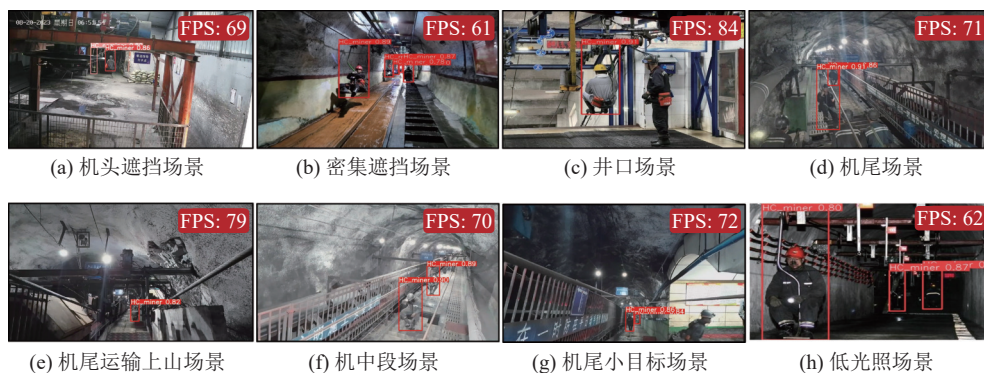


图9 提出模型在各类场景的推理可视化展示

Fig.9 Visualization of model's reasoning in various scenarios

Cascade-RCNN 推理结果(图 8b), 两阶段检测模型在无噪声情况下的检测置信度较高, 在有遮挡情况和小目标检测时, 检测效果出现了略微下降趋势; 根据 SSD、YOLOv5-n 和 YOLOv8-n 推理结果(图 8c—图 8e), 单阶段检测模型在无噪声情况下的检测效果较好, 在有遮挡情况和小目标检测时, 检测效果较差甚至出现了漏检重检的现象, 如 SSD 模型在机头处对无载人的猴车目标出现了漏检情况, 对机尾处载人小目标猴车的预测框存在定位不准确。YOLOv5-n 在机头处对无载人的猴车目标出现了重检现象, 对机尾处载人小目标猴车出现了漏检情况。YOLOv8-n 在机尾处将候车区的矿工识别为载人状态的猴车, 出现了错检现象。

综合图 8 和图 9 检测可视化结果, 对于各类场景下不同机段位置的猴车检测, 笔者提出模型没有出现漏检重检错检的情况, 对遮挡、低光照和小目标检测时, 模型也有着较好的识别效果。对预测框的定位更为精准, 检测置信度也更高, 模型的鲁棒性更好, 更适合实际巡检任务中复杂多变的煤矿场景。

## 4.2 模型改进前后对比试验

### 4.2.1 改进前后损失值分析

图 10 为基于 YOLOv8-n 模型改进前后的训练  $C^{IOU}$  Loss、 $B^{CE}$  Loss 和  $D^{FL}$  Loss 损失曲线对比, 改进的模型  $C^{IOU}$  Loss 在 100 轮左右时趋于收敛、最终约收敛于 0.41;  $B^{CE}$  Loss 在 70 轮时趋于收敛且最终收敛于 0.95;  $D^{FL}$  Loss 在 80 轮左右时趋于收敛且最终收敛于 0.91, 根据图 10, 对比基线模型, 提出模型的损失曲线有更快的收敛速度、更低的损失值。

### 4.2.2 改进前后识别精度分析

图 11a—图 11c 分别为改进前后精确率( $P$ ), 召回率( $R$ ), 平均精度均值对比曲线。根据图 10 结果, 采用设计的模块并融合注意力机制的跨层级连接路径优化方式后, 优化后模型的曲线收敛得更快, 基本

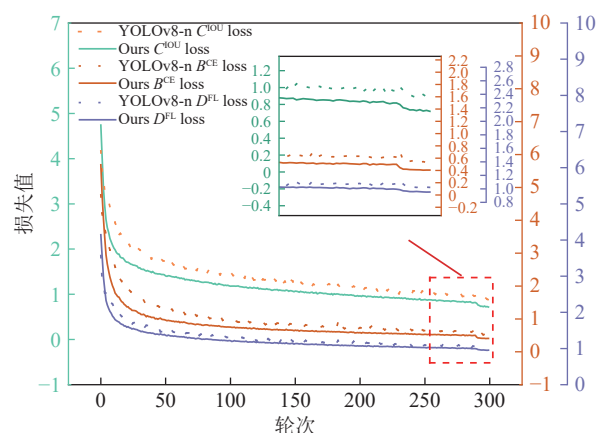


图10 改进前后损失曲线分析

Fig.10 Comparison of loss curves improvement

没有出现震荡现象。相比基线模型, 提出的优化模型精确率提升了 7.44%, 召回率提升了 9.82%, 平均精度均值提升了 7.72%, 进一步验证了设计模块对基线模型改进的合理性。

### 4.2.3 改进前后混淆矩阵分析

为进一步剖析改进前后模型对有无载人猴车目标的具体辨识能力, 图 12 给出了改进前后的混淆矩阵对比, 在图 7 的验证集样本上进行, 包括 209 个载人猴车目标, 132 个无载人猴车目标。对于 YOLOv8-n 模型中 209 个 HC\_miner 目标, 189 个 HC\_miner 目标正确识别, 3 个 HC\_miner 被识别成 HC\_nominer, 17 个 HC\_miner 被识别成背景(漏检); 对于 132 个 HC\_nominer 目标, 107 个 HC\_nominer 目标被正确识别到, 1 个 HC\_nominer 被识别成 HC\_miner, 漏检个数为 24 个。

提出的多尺度特征融合模型中, 205 个 HC\_miner 目标正确识别, 只有 4 个 HC\_miner 被识别成背景, 没有 HC\_miner 目标被识别成 HC\_nominer, 说明改进模型可有效区分有无载人状态的猴车特征。对于 HC\_nominer 目标, 改进模型仅有 2 个

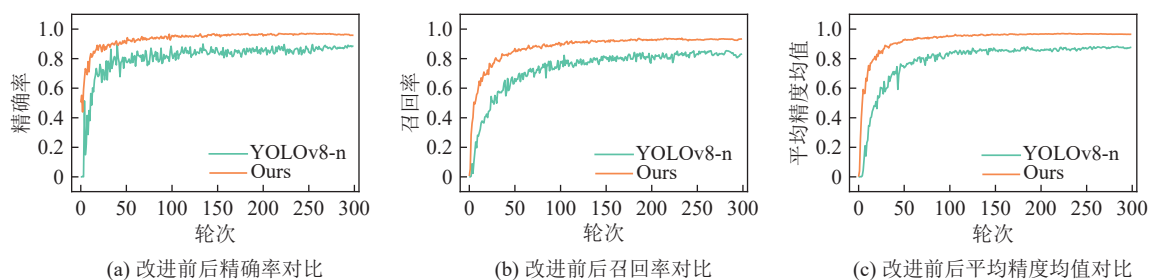


图 11 改进前后模型的精确率、召回率、平均精度均值对比曲线

Fig.11 Comparison curves of precision, recall and average precision of model before and after improvement

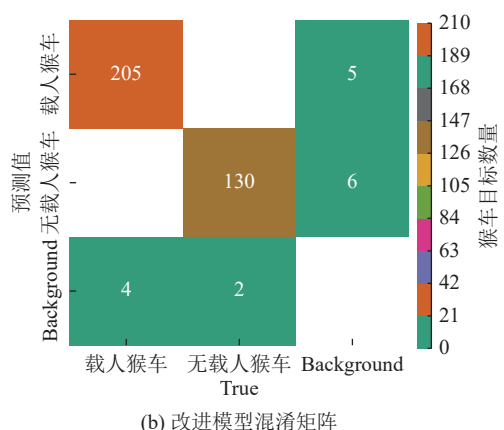
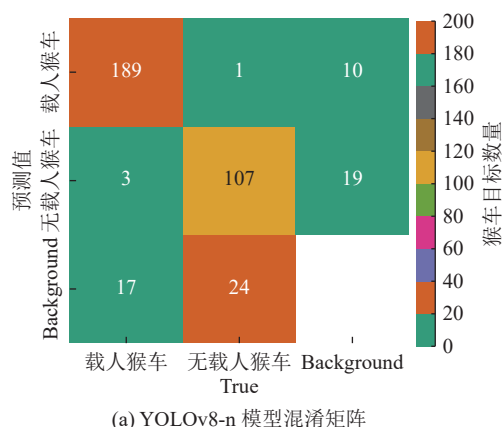


图 12 改进前后混淆矩阵对比

Fig.12 Comparison of confusion matrix before and after improvement

HC\_nominer 发生漏检。由此见得改进的多尺度井下猴车识别模型对比原模型精确度更高、漏检率更低,可有效辨识有无载人的井下猴车目标。

#### 4.2.4 改进前后热力图分析

为更直观了解模型对于猴车图像中重要区域的关注度情况,验证模型是否正确关注到了猴车目标位置,采用 Grad-CAM<sup>[34]</sup> 技术分析改进前后模型的热力图情况,结果如图 13 所示。YOLOv8-n 模型的热力图结果表明(图 13a),在矿井入口处虽然关注到了猴车的温度信息,但关注的有效区域范围较少,并且关注到了较多的背景干扰物;在猴车运行中部,只

提取到较少范围的猴车温度信息,多背景的过滤不足;在尾部,没有关注到小目标猴车特征,对背景的温度提取也较多。

笔者提出模型的热力图结果表明(图 13b),在矿井入口遮挡情况下提取到了较多猴车的温度信息,基本没有关注背景干扰物;在猴车运行中部,对于多个目标的猴车,模型对猴车关键区域关注得更多,识别到的背景温度信息更低;在尾部,模型有效关注到了不同猴车目标的重点区域,对于小目标猴车的特征温度更为明显,对背景的过滤效果更好。

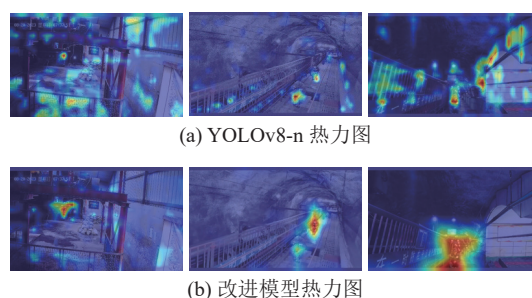


图 13 改进前后热力图对比

Fig.13 Comparison of heatmaps before and after improvement

#### 4.3 应用部署

采用 Pyside6 库设计一款适合井下猴车智能识别的可视化界面,将提出模型训练后的最优权重部署到该矿东一采区运输上山猴车运行中段处的视频流监控中,进行网络模型的应用测试,视频媒体流部署到井下猴车智能识别系统技术路线如图 14 所示。首先根据井下防爆摄像头设备 API 和网络协议获取猴车监控视频流,随之利用 OpenCV 库等对猴车监控视频流进行编码操作,再对编码后的猴车监控视频流创建 HTTP 或 RSTP 协议服务器,最后将创建 HTTP 或 RSTP 协议服务器作为处理和接收平台,以便将视频流实时传输给客户端软件(井下猴车智能识别系统)。该系统具体功能如图 14 所示。

设计的井下猴车智能识别系统支持本地文件



图14 井下猴车智能识别系统视频流部署

Fig.14 Video stream deployment of intelligent recognition system for overhead passenger devices in coal mines

(jpg、png、mp4 等格式)、摄像头、远程视频媒体流等输入待推理文件格式,在实时推理视频流的同时,还可实现类别数目计数、目标数目计数、帧率实时检测,工程人员还可根据实际需求调整检测置信度阈值、交并比阈值、抽帧间隔、保存推理文件和标签。井下猴车智能识别算法部署到监控视频流中可有效实现远程监控视频实时调度,有着提高监控准确性、提升巡检工作效率,减少人力巡检成本等多方面的实际工程应用价值。

#### 4.4 未来工作

井下猴车智能识别检测涉及多学科交叉,对数据集的构建及算法的设计需要考虑众多因素,笔者基于自建的井下猴车数据集,采用改进的 YOLOv8-n 模型对井下猴车进行了智能识别,初次对提出算法和远程监控视频流进行解码进行应用部署。未来对井下猴车智能识别研究将做出如下改进:①持续收集各类煤矿不同场景、不同运行条件的井下猴车图像以扩充笔者构建数据集,使算法能够更好地泛化于各类场景的井下猴车识别任务中;②采用更轻量更精准的改进策略,提升模型多尺度特征提取能力,实现算法识别精度与速度上的并进;③井下环境恶劣,存在粉尘烟雾、网络带宽有限、摄像头视野范围窄等问题,需考虑更鲁棒的模型、更好的网络带宽设备。

## 5 结 论

1)在自建数据集上设计了合适的数据增强策略,实现多机段及普适性高的井下猴车图像数据集的构建,为煤矿视觉巡检任务提供了基础支撑。

2)在主干网络引入可变形卷积,设计了一种新颖的特征提取模块,解决猴车识别任务中存在的重

叠遮挡和坐姿多变等空间尺度问题。在不同尺度特征图的检测头前引入坐标注意力机制以自适应捕捉全局关键信息,进一步提出了跨层级连接的路径聚合方式,实现了网络之间的多尺度依赖,使得深层网络位置信息与浅层网络细节特征得到多尺度复用,增加模型对小目标猴车特征的提取,降低猴车目标漏检误检率,过滤背景干扰。

3)在自建数据集上的试验结果表明,笔者提出模型的平均精度均值为 95.6%,相比基线模型提升了 7.7%,参数量和模型大小分别仅为  $3.1 \times 10^6$  和 6.1 MB,识别速率为 71 帧,满足边缘设备的部署要求。最后,基于 Pyside6 库,将提出的井下猴车智能识别算法和井下监控视频流部署到设计的猴车智能识别系统中,有效实现了远程监控视频实时调度,减少人力巡检成本,加速煤矿视觉巡检任务的发展与应用落地。

#### 参考文献(References):

- [1] 王国法. 煤矿智能化最新技术进展与问题探讨[J]. 煤炭科学技术, 2022, 50(1): 1-27.  
WANG Guofa. New technological progress of coal mine intelligence and its problems[J]. Coal Science and Technology, 2022, 50(1): 1-27.
- [2] 王国法, 赵国瑞, 任怀伟. 智慧煤矿与智能化开采关键核心技术分析[J]. 煤炭学报, 2019, 44(1): 34-41.  
WANG Guofa, ZHAO Guorui, REN Huaiwei. Analysis on key technologies of intelligent coal mine and intelligent mining[J]. Journal of China Coal Society, 2019, 44(1): 34-41.
- [3] 尹茂振. 煤矿架空乘人装置集中控制系统设计[D]. 徐州: 中国矿业大学, 2021.  
YIN Maozhen. Design of centralized control system for coal mine overhead passenger device[D]. Xuzhou: China University of Mining and Technology, 2021.
- [4] 王国法. 煤矿智能化十大“痛点”解析及对策[J]. 智能矿山,



- 2021, 2(3): 1-4.
- [5] 王国法, 杜毅博, 任怀伟, 等. 智能化煤矿顶层设计研究与实践[J]. 煤炭学报, 2020, 45(6): 1909-1924.  
WANG Guofa, DU Yibo, REN Huaiwei, et al. Top level design and practice of smart coal mines[J]. Journal of China Coal Society, 2020, 45(6): 1909-1924.
- [6] AL-KARKHI N K, ABBOOD W T, KHALID E A, et al. Intelligent robotic welding based on a computer vision technology approach[J]. *Computers*, 2022, 11(11): 155.
- [7] YANG W J, ZHANG X H, MA B, et al. An open dataset for intelligent recognition and classification of abnormal condition in long-wall mining[J]. *Scientific Data*, 2023, 10(1): 416.
- [8] WANG L Y, WANG X W, LI B. Data-driven model SSD-BSP for multi-target coal-gangue detection[J]. *Measurement*, 2023, 219: 113244.
- [9] YAN P C, SUN Q S, YIN N N, et al. Detection of coal and gangue based on improved YOLOv5.1 which embedded scSE module[J]. *Measurement*, 2022, 188: 110530.
- [10] 刘普壮. 基于改进 YOLO 算法的煤矸识别方法与实验研究[D]. 淮南: 安徽理工大学, 2022.  
LIU Puzhuang. Experimental research on coal gangue recognition method based on improved YOLO algorithm[D]. Huainan: Anhui University of Science & Technology, 2022.
- [11] 张磊, 王浩盛, 雷伟强, 等. 基于 YOLOv5s-SDE 的带式输送机煤矸目标检测[J]. 工矿自动化, 2023, 49(4): 106-112.  
ZHANG Lei, WANG Haosheng, LEI Weiqiang, et al. Coal gangue target detection of belt conveyor based on YOLOv5s-SDE[J]. Journal of Mine Automation, 2023, 49(4): 106-112.
- [12] 徐志强, 吕子奇, 王卫东, 等. 煤矸智能分选的机器视觉识别方法与优化[J]. 煤炭学报, 2020, 45(6): 2207-2216.  
XU Zhiqiang, LYU Ziqi, WANG Weidong, et al. Machine vision recognition method and optimization for intelligent separation of coal and gangue[J]. Journal of China Coal Society, 2020, 45(6): 2207-2216.
- [13] 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 Deqiang, QIAN Jiansheng, GUO Xingge, et al. Review on key technologies of AI recognition for videos in coal mine[J]. Coal Science and Technology, 2023, 51(2): 349-365.
- [14] 程德强, 徐进洋, 寇旗旗, 等. 融合残差信息轻量级网络的运煤皮带异物分类[J]. 煤炭学报, 2022, 47(3): 1361-1369.  
CHENG Deqiang, XU Jinyang, KOU Qiqi, et al. Lightweight network based on residual information for foreign body classification on coal conveyor belt[J]. Journal of China Coal Society, 2022, 47(3): 1361-1369.
- [15] 程德强, 陈杰, 寇旗旗, 等. 融合层次特征和注意力机制的轻量化矿井图像超分辨率重建方法[J]. 仪器仪表学报, 2022, 43(8): 73-84.  
CHENG Deqiang, CHEN Jie, KOU Qiqi, et al. Lightweight super-resolution reconstruction method based on hierarchical features fusion and attention mechanism for mine image[J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 73-84.
- [16] 郝帅, 张旭, 马旭, 等. 基于 CBAM-YOLOv5 的煤矿输送带异物检测[J]. 煤炭学报, 2022, 47(11): 4147-4156.  
HAO Shuai, ZHANG Xu, MA Xu, et al. Foreign object detection in coal mine conveyor belt based on CBAM-YOLOv5[J]. Journal of China Coal Society, 2022, 47(11): 4147-4156.
- [17] 闫志蕊, 王宏伟, 耿毅德. 基于改进 DeeplabV<sup>3+</sup>和迁移学习的煤岩界面图像识别方法[J]. 煤炭科学技术, 2023, 51(S1): 429-439.  
YAN Zhirui, WANG Hongwei, GENG Yide. Coal-rock interface image recognition method based on improved DeeplabV<sup>3+</sup> and transfer learning[J]. Coal Science and Technology, 2023, 51(S1): 429-439.
- [18] 李伟山, 卫晨, 王琳. 改进的 Faster RCNN 煤矿井下行人检测算法[J]. 计算机工程与应用, 2019, 55(4): 200-207.  
LI Weishan, WEI Chen, WANG Lin. Improved faster RCNN approach for pedestrian detection in underground coal mine[J]. *Computer Engineering and Applications*, 2019, 55(4): 200-207.
- [19] 张明臻. 基于 Dense-YOLO 网络的井下行人检测模型[J]. 工矿自动化, 2022, 48(3): 86-90.  
ZHANG Mingzhen. Underground pedestrian detection model based on Dense-YOLO network[J]. Industry and Mine Automation, 2022, 48(3): 86-90.
- [20] 邵小强, 李鑫, 杨涛, 等. 改进 YOLOv5s 和 DeepSORT 的井下人员检测及跟踪算法[J]. 煤炭科学技术, 2023, 51(10): 291-301.  
SHAO Xiaoqiang, LI Xin, YANG Tao, et al. Underground personnel detection and tracking based on improved YOLOv5s and DeepSORT[J]. Coal Science and Technology, 2023, 51(10): 291-301.
- [21] YANG W J, WU J C, ZHANG J L, et al. Deformable convolution and coordinate attention for fast cattle detection[J]. *Computers and Electronics in Agriculture*, 2023, 211: 108006.
- [22] ZHU X Z, HU H, LIN S, et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019.
- [23] 范瑶瑶, 王兴芬, 刘亚辉. 改进 DeepLabv3+网络的钢板表面缺陷检测研究[J]. 计算机工程与应用, 2023, 59(16): 150-158.  
FAN Yaoyao, WANG Xingfen, LIU Yahui. Improved Deeplabv3+Model for surface defect detection on steel plates[J]. *Computer Engineering and Applications*, 2023, 59(16): 150-158.
- [24] 颜玉松, 尹芳洁, 王彩玲. 融合 Xception 特征提取和坐标注意力机制的血细胞分割[J]. 计算机系统应用, 2023, 32(1): 275-280.  
YAN Yusong, YIN Fangjie, WANG Cailing. Blood cell segmentation fusing xception feature extraction and coordinate attention mechanism[J]. Computer Systems & Applications, 2023, 32(1): 275-280.
- [25] OMAR N, SENGUR A, AL-ALI S G S. Cascaded deep learning-based efficient approach for license plate detection and recognition[J]. *Expert Systems with Applications*, 2020, 149: 113280.
- [26] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with dropout[EB/OL]. 2017. [2024-01-06]. <https://arxiv.org/abs/2107.08430>.

- [ 27 ] DING C, CHEN Y F, LI R Z, et al. Integrating hybrid pyramid feature fusion and coordinate attention for effective small sample hyperspectral image classification[J]. *Remote Sensing*, 2022, 14(10): 2355.
- [ 28 ] ZHANG J M, XIE Z P, SUN J, et al. A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection[J]. *IEEE Access*, 2020, 8: 29742–29754.
- [ 29 ] REN S Q, HE K M, 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*, 2017, 39(6): 1137–1149.
- [ 30 ] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[M]//Computer vision—ECCV 2016. Cham: Springer International Publishing, 2016: 21–37.
- [ 31 ] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318–327.
- [ 32 ] DANG F Y, CHEN D, LU Y Z, et al. YOLOWeeds: a novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems[J]. *Computers and Electronics in Agriculture*, 2023, 205: 107655.
- [ 33 ] GE Z , LIU S , WANG F ,et al. YOLOX: Exceeding YOLO Series in 2021[EB/OL]. [2024–01–06]. <https://arxiv.org/abs/2107.08430>.
- [ 34 ] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision*, 2020, 128(2): 336–359.