

# 基于深度学习和模型压缩技术的轻量级 煤矿人车检测模型 ——以贵州地区煤矿为例

解北京<sup>1</sup>, 李恒<sup>1</sup>, 栾铮<sup>1</sup>, 雷振<sup>2</sup>, 李晓旭<sup>1</sup>, 李卓<sup>3</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. 贵州理工学院矿业工程学院, 贵州贵阳 550003;  
3. 北京工业大学城市建设学部, 北京 100124)

**摘要:** 煤矿工人和载人车辆(煤矿人车)的智能识别是视频监控系统的重要组成部分, 也是煤矿智能化发展的关键任务。然而, 煤矿人车检测场景较为复杂, 大型人车检测模型部署在有限的计算设备上难以实现, 如何在模型检测性能和检测效率之间取得平衡存在诸多挑战。以贵州地区煤矿视频监控数据集为例, 提出了一种基于深度学习和模型压缩技术的轻量级煤矿人车检测模型, 该模型精准实时的完成了煤矿人车检测任务, 对网络进行瘦身的同时几乎没有损失检测性能。具体来说, 在网络模型设计阶段, 以 YOLOv8s 为基线提出了一种名为 FCW-YOLO 的煤矿人车轻量级检测模型, 首先将 Faster-Block 和坐标注意力和开发到网络的特征提取模块中, 设计了一种新颖的 C2f-Faster-CA 轻量级架构, 旨在减少网络的冗余通道同时自适应捕捉全局关键信息; 其次, 采用了 WIOU 边界回归损失函数以增加模型对普通质量样本的关注, 降低了训练样本不平衡带来的回归误差等问题。在模型压缩阶段, 联动剪枝算法对提出的 FCW-YOLO 模型进行通道级别的稀疏, 模型可自动识别不重要的通道并对其进行删减, 实现了煤矿人车检测模型二次轻量化设计 FCWP-YOLO。在自建的煤矿人车检测数据集上的结果表明, 提出的模型参数量, 计算量和模型大小分别为 2.3 M, 4.0 GFLOPs, 6.0 MB, 对比基线模型分别实现了 4.9、4.7、4.4 倍的压缩效果, 平均检测精度为 88.7%, 提高了 1.1%, 每张图像的处理速度仅为 5.6 ms。对比多种轻量级架构和先进的检测模型, 该方法精度表现优异, 计算成本更低, 实时性能更好, 为资源受限的煤矿场景提供了一种可行的煤矿人车检测方法, 满足煤矿视频监控部署要求, 可为煤矿人车智能巡检任务提供实时预警。

**关键词:** 煤矿工人检测; 煤矿载人车辆检测; 深度学习; 模型压缩; 轻量级架构

中图分类号: TD82; TP18; TP391 文献标志码: A 文章编号: 0253-9993(2025)02-1383-16

## Lightweight coal miners and manned vehicles detection model based on deep learning and model compression techniques: A case study of coal mines in Guizhou region

XIE Beijing<sup>1</sup>, LI Heng<sup>1</sup>, LUAN Zheng<sup>1</sup>, LEI Zhen<sup>2</sup>, LI Xiaoxu<sup>1</sup>, LI Zhuo<sup>3</sup>

收稿日期: 2024-07-21 策划编辑: 郭晓炜 责任编辑: 宫在芹 DOI: 10.13225/j.cnki.jccs.2024.0459

基金项目: 国家重点研发计划资助项目(2022YFC2904100, 2023YFC3011300); 中央高校基本科研业务费专项资金资助项目(2023ZKPYAQ04)

作者简介: 解北京(1984—), 男, 安徽滁州人, 副教授, 博士, Email: bpxie1984@163.com

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

引用格式: 解北京, 李恒, 栾铮, 等. 基于深度学习和模型压缩技术的轻量级煤矿人车检测模型——以贵州地区煤矿为例[J]. 煤炭学报, 2025, 50(2): 1383-1398.

XIE Beijing, LI Heng, LUAN Zheng, et al. Lightweight coal miners and manned vehicles detection model based on deep learning and model compression techniques: A case study of coal mines in Guizhou region[J]. Journal of China Coal Society, 2025, 50(2): 1383-1398.



移动阅读

(1.School of Emergency Management and Safety Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China; 2.Guizhou Institute of Technology, College of Mining Engineering, Guiyang 550003, China; 3.Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100124, China)

**Abstract:** Intelligent recognition of coal mine workers and manned vehicles (coal mine pedestrian-vehicles) is an important component of video surveillance systems and a key task in the development of coal mine intelligence. However, the detection scene of coal mine pedestrian-vehicles is complex, and deploying large pedestrian-vehicle detection models on limited computing devices is challenging. Balancing between model detection performance and efficiency poses many challenges. This paper proposes a lightweight coal mine pedestrian detection model based on deep learning and model compression techniques. Taking the coal mine video surveillance dataset in Guizhou region as an example. The model accurately and in real-time completes the task of detecting coal mine pedestrian-vehicles, achieving a balance between model detection performance and efficiency. Specifically, in the network model design phase, a lightweight detection model named FCW-YOLO is proposed based on YOLOv8s as the baseline. Faster-Block and coordinate attention are integrated into the feature extraction module of the network, designing a novel C2f-Faster-CA lightweight architecture to reduce redundant channels of the network while adaptively capturing global key information. Furthermore, the WIOU boundary regression loss function is employed to increase the model's focus on common quality samples, addressing issues such as regression errors caused by imbalanced training samples. In the model compression phase, the proposed FCW-YOLO model undergoes channel-level sparsity through a collaborative pruning algorithm, automatically identifying unimportant channels and reducing them, resulting in the FCWP-YOLO model, achieving secondary lightweight design of the coal mine pedestrian-vehicle detection model. Results on a self-built coal mine pedestrian-vehicle detection dataset show that the proposed model has parameters, computational load, and model size of 2.3 M, 4.0 GFLOPs, and 6.0 MB, respectively, achieving compression ratios of 4.9 times, 4.7 times, and 4.4 times compared to the baseline model. The average detection accuracy is 88.7%, an improvement of 1.1%, with a processing speed of only 5.6ms per image. Compared to various lightweight architectures and advanced detection models, this method demonstrates excellent accuracy, lower computational costs, and better real-time performance, providing a feasible coal mine pedestrian-vehicle detection method for resource-constrained coal mine scenarios, meeting the deployment requirements of coal mine video surveillance and enabling real-time alerts for intelligent inspection of coal mine pedestrian-vehicles.

**Key words:** coal miners detection; coal manned vehicles detection; deep learning; model compression; lightweight architecture

## 0 引 言

煤矿作为我国主体能源产业,目前正处于机械化过渡到智能化的关键时期<sup>[1-2]</sup>。《煤矿智能化建设指南》<sup>[3-6]</sup>指出,要以人为本,坚持把减人、增安和提效作为智能化煤矿建设的根本目标,重点突破智能主辅运输和智能安全监控系列的关键技术和装备。煤矿工人和载人运输车辆(煤矿人车)是煤炭开采的前提与基础,在推动经济发展的同时也面临着诸多安全隐患,其智能化管理程度决定了煤矿安全高效生产效率。煤矿人车检测技术作为视频监控系统的重要组成部分,可为煤矿工人事故的发生和载人运输车辆跑偏提供实时预警,对于提高煤矿安全管理水平具有重要意义。然而,传统的基于人工的和机器学习算法在煤矿行人检测中存在一系列问题。人工检测存在主观性强、工作效率低下等缺点;传统机器学习算法在复杂

环境下的准确率和鲁棒性有限。这些挑战制约了煤矿监控预警水平的提升,也增加了煤矿事故发生的风险,亟需开展新一代信息技术的煤矿人车智能检测研究。

目前基于深度学习技术主流的物体检测算法主要有3个类别:第一是以Faster-RCNN<sup>[7]</sup>为代表的两阶段检测算法。第二是以DETR<sup>[8]</sup>为代表的基于Transformer架构的序列建模检测算法。第三是以SSD<sup>[9]</sup>,YOLO系列<sup>[10-15]</sup>为代表的一阶段检测算法。然而,两阶段检测算法和基于Transformer架构的检测算法检测精度高,仍然无法避免计算成本大速度慢等缺点,难以实际应用于煤矿场景。相反,一阶段的YOLO系列检测算法因其实现了检测速度和检测精度的平衡,在煤矿行人和运输车辆检测模型开发过程中受到了更多研究人员的关注。张明臻等<sup>[16]</sup>将含有残差块的Dense开发到YOLOv3中,避免了训练过程

中梯度消失的影响,有效降低了行人目标的漏检率,提升了煤矿行人检测性能。邵小强等<sup>[17]</sup>采用了MobileOne 和 RepVGG 结合的轻量级骨干架构,并在YOLOv7 网络的末端引入具有通道注意力的交叉垂直注意力模块以增强检测性能,获得了 83.1% 的平均精度和 12.6 ms 图像处理速度。徐志等<sup>[18]</sup>提出了一种由深度可分离卷积和下采样逆残差块组成的骨干,并开发到 YOLOv3 中,实现了 0.767 的检测精度和 4.6 ms 的处理速度,模型大小为 16.1 MB。江帆等<sup>[19]</sup>针对井工煤矿辅助运输车辆的通行性评估问题,采用传统机器学习支持向量机分类模型结合 HOG 和 LBP 特征构建了煤矿行人检测算法,实现了 82.1% 的检测精度。

以上研究人员通过一系列的改进策略提高了煤矿行人检测精度,降低了模型的计算成本。但对于轻量级的煤矿人车检测任务仍存在以下几点不足:①煤矿人车的智能检测能有效提升巡检效率和实现减人增安。然而,目前的研究仅局限于煤矿行人单类别检测任务,开展煤矿人车多类别智能检测技术的相关研究鲜见,并且针对轻量级的煤矿检测任务研究也较少;②一些研究人员都以轻量级网络替换为思路以开发计算成本低的 YOLO 系列检测模型,虽然处理速度有所降低,模型体积仍然较大,其轻量级表现性能仍然有限。并且一些研究人员是以牺牲过多精度为代价实现轻量级的煤矿行人检测,例如徐志等的研究中模型的图像处理速度非常快为 4.6 ms,但检测精度较低仅为 76.7%,同时牺牲了 5.7% 的精度;③最重要的局限是,大多物体检测网络中,都存在过参数化的不利影响,即网络的计算成本远超于待解决问题所需的最小计算成本,以致生成冗余权重及网络层,这是替换网络结构无法解决的问题。在计算资源受限的煤矿场景,这给模型部署在煤矿视频监控系统里带来诸多不利影响。近年来,研究人员<sup>[20-21]</sup>提出了模型修剪技术以解决过参数化问题,旨在几乎不以损失精度为代价最大程度的压缩模型,但是目前还没有相关的研究将模型剪枝技术开发到煤矿人车检测模型中。因此为应对煤矿实际条件的复杂性和有限的计算设备,亟需开发一种高精度,轻量,实时的煤矿人车检测方法。

鉴于此,笔者以煤矿工人,平巷人车,煤矿架空乘人装置为代表构建了 4 个类别的煤矿人车检测数据集,以实时性,模型体积,集成度等方面表现出色的YOLOv8s<sup>[22]</sup>为基线,设计了一种轻量级架构融合坐标注意力的特征提取模块,采用边界回归误差更低的损失函数以增加模型检测性能,提出了一种轻量级的煤矿人车检测模型 FCW-YOLO。进一步结合通道修剪

方法对 FCW-YOLO 模型进行压缩,减少模型的冗余通道,得到几乎不损精度的低成本模型,命名为 FCWP-YOLO。本文工作是首次尝试将其应用于煤矿人车检测领域,该方法在轻量级,实时性方面表现优异,检测性能达到了先进的水平,可满足煤矿视频监控及井下机器人实际部署的需求,为复杂条件和计算资源受限的煤矿场景提供了高效实时的人车检测方法。

## 1 轻量级煤矿人车检测模型构建

在煤矿人车检测任务中,实际输入为视频流,通过将视频切分为一帧一帧的图像以实现等价训练的目的。然而对于视频流输入而言,所需底层算法的要求较高,不仅需要考虑检测精度,还需较低延迟和较小体积的模型才能保证视频流的煤矿人车实时检测,同时实际应用中还需考虑系统集成和部署兼容问题。据此,笔者选取了合理的网络模型、设计了高效的轻量化架构、选择了兼容性较好的模型压缩技术,该方法充分考虑实际应用场景的要求,提高了工程应用的可靠性。以上具体措施将在下节展开详细说明。同时,图 1 给出了设计的轻量级人车检测模型实现思路,主要包括煤矿人车图像采集及前处理、轻量级煤矿人车检测模型设计(FCW-YOLO)、采用通道剪枝对 FCW-YOLO 进行二次轻量化设计。

**煤矿人车图像采集及前处理:**采用井下防爆摄像头获取视频信息,利用调度室的视频数据将其切分为图片,对采集到的图片进行 YOLO 格式标注,划分数据集,为检测任务提供夯实的数据支持。

**基线选取:**本文的工作是以 YOLOv8 检测器为基础开展, YOLOv8 是目前 YOLO 系列最新的先进检测器,相较以往系列,主要引进了 C2f 模块,解耦头,无锚点检测等。C2f 模块的分离操作使得特征信息的交换更为丰富,解耦头将目标位置和类别信息分离使得目标定位更加准确,且无锚框的设计不需要采用聚类等额外操作对锚框进行聚类,省去了中间操作,模型集成度更好,更适合边缘设备的部署。YOLOv8s 是 YOLOv8 中计算开销较小的模型,模型复杂度较低,推理速度较快,且检测精度也有着较好表现,本文旨在设计极具轻量化、用于未来能够实际部署于煤矿场景的人车检测器,因此我们选择了 YOLOv8s 作为本文工作的基线模型。

**轻量级煤矿人车检测模型设计:**在煤矿人车特征提取过程中,不同通道的特征具有较高相似性以致特征的重复提取,增加了网络计算成本,据此笔者将轻量级架构(Faster-Block)开发到基线模型中以降低网

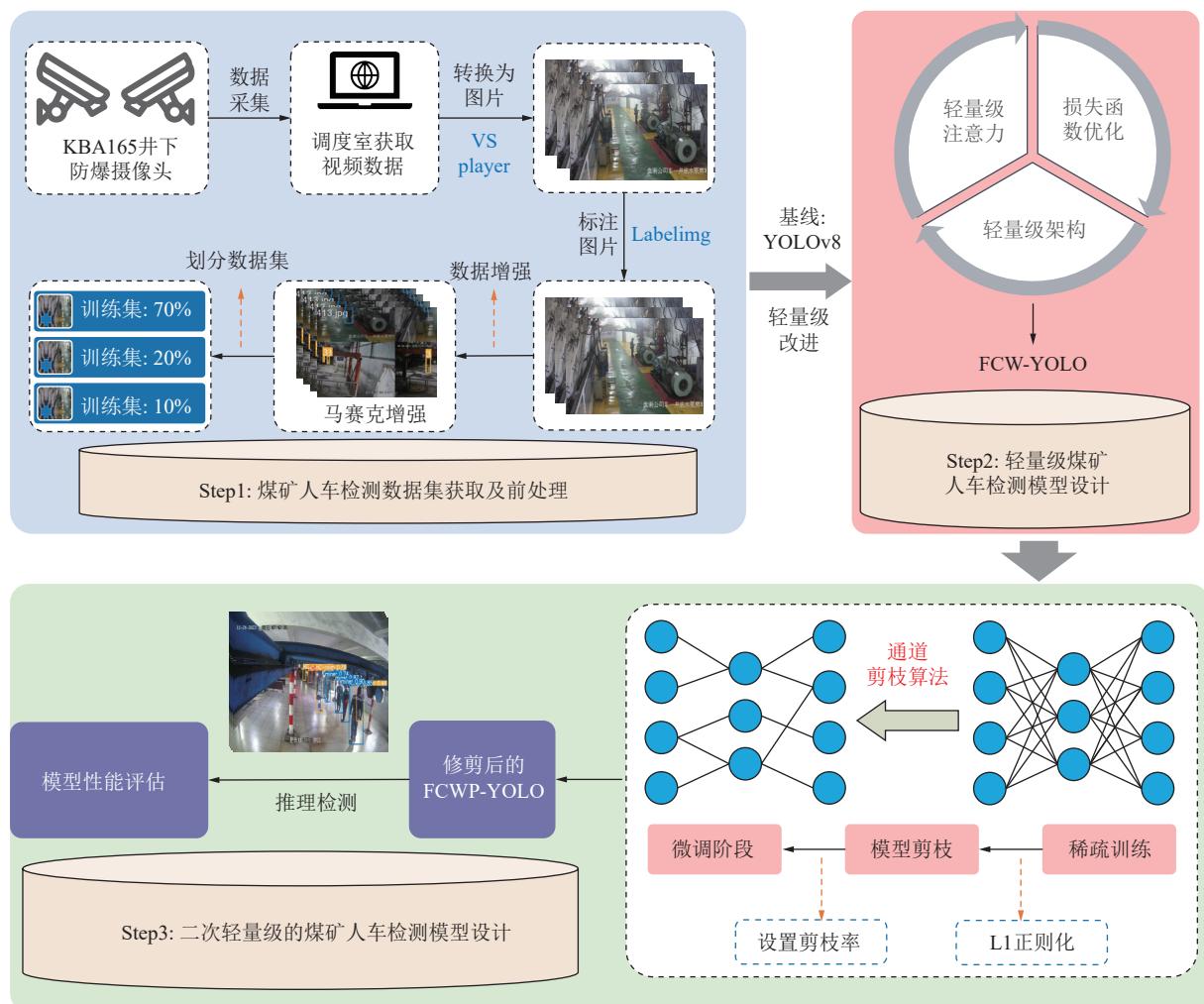


图1 轻量级煤矿人车检测模型实现的技术路线

Fig.1 Technical roadmap of lightweight coal mine pedestrian - vehicle detection model implementation

络开销；同时，煤矿开采环境存在大量的背景噪声，煤矿人车关键特征的提取存在挑战，故引入坐标注意力模块以降低噪声干扰同时增强煤矿人车关键特征的提取；另一方面煤矿人车检测任务中，光照不均匀、遮挡等挑战产生了较多低质量样本，数据集中的高质量样本容易主导训练过程的梯度，导致了低质量样本的拟合有限，从而使训练产生较大的回归误差，故引入 WIOU 损失函数以进一步增强模型检测性能。总体而言，笔者综合了轻量级架构 (Faster-Block) 和轻量级注意力 (CA) 2 个模块的各自优势，并将其引入 C2f 模块中，设计了一种新型的特征提取模块 C2f-Faster-CA，该网络架构即轻量又提点。进一步将边界回归误差更低的损失函数 (WIOU) 开发到基线模型中，得到轻量高效的煤矿人车检测模型 FCW-YOLO。该网络结构如图 2 所示。

采用通道剪枝算法对设计的 FCW-YOLO 煤矿人车检测模型进行二次轻量化设计：通过上述训练得到最优 FCW-YOLO 模型权重文件，然后采用通道剪枝

对 FCW-YOLO 检测模型权重文件的 BN 层进行稀疏训练，得到通道级别的稀疏模型，其次根据实际裁剪比例要求，人工设置一个全局剪枝阈值，模型可自动识别不重要的通道并对其进行删除，最后对剪枝后的模型进行微调重新训练，以恢复检测精度，这样即可得到不损失精度的高压缩模型 FCWP-YOLO。我们的方法可为煤矿智能化硬件装备的研制大大减少模型的开销成本，实现高精度、实时的煤矿人车检测任务。采用的通道剪枝算法将在下节详细展开。

### 1.1 C2f-Faster-CA 模块设计

C2f 作为 YOLOv8 网络的核心模块，是模型特征提取和特征融合阶段的重要部件，由于其结构上堆叠了大量的卷积操作，造成了网络的特征冗余较大。图 3 给出了本文检测数据集中 P2 层 C2f 模块的特征输出可视化示例，在煤矿人车检测任务中，这些特征在不同通道中具有较高的相似性，造成了特征的重复提取，增加了不必要的计算成本。为解决该问题，笔者受到 CHEN 等<sup>[23]</sup>工作的启发，引入了一种 Faster-

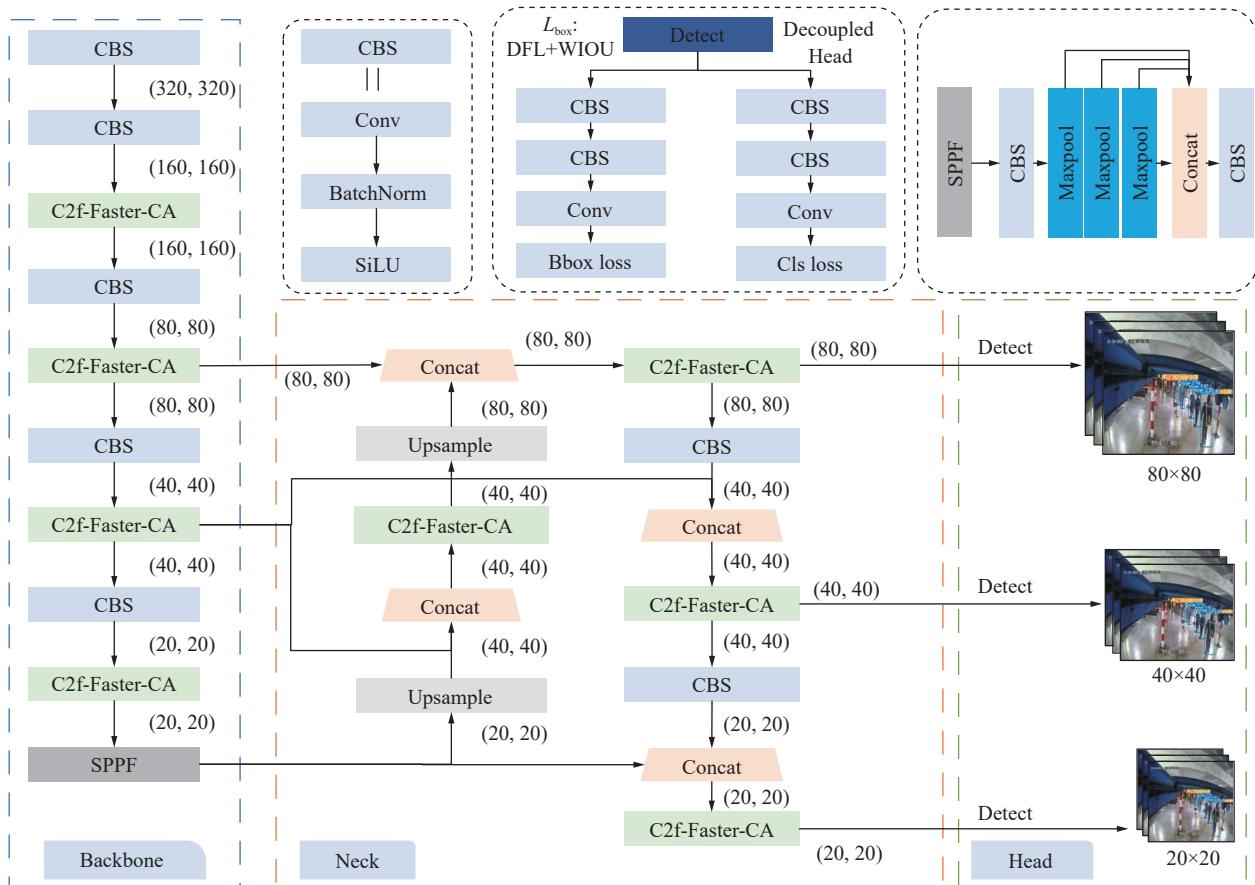


图2 FCW-YOLO 煤矿人车检测模型网络架构

Fig.2 Network architecture of FCW-YOLO coal mine pedestrian - vehicle detection model

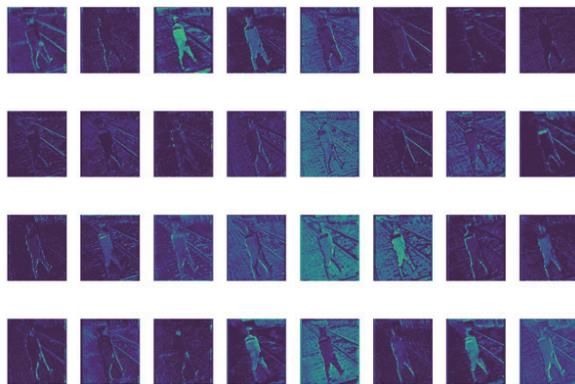


图3 第1个C2f模块中不同通道特征可视化

Fig.3 Visualization of different channel features in the first C2f module

Block 以降低网络的计算成本。进一步将坐标注意力 (CA) 开发到 Faster-Block 中 (Faster-Block-CA), 以增加网络的多尺度依赖, 提升检测性能。最后, 将 C2f 原有的 Bottleneck 全部替换为 Faster-Block-CA, 得到设计的 C2f-Faster-CA 轻量级架构。该轻量级架构的具体细节如下展开。

### 1.1.1 Faster-Block 模块

Faster-Block 由 1 个  $3 \times 3$  的部分卷积 (Partial Con-

volution, PConv) 和 2 个  $1 \times 1$  的 Conv 层构成。在 Bottleneck 的堆叠网络中, 每一个 Conv 层后都有一个批量归一化层和激活层, 网络中过度使用这些层可能会限制特征多样性, 并且造成冗余现象。因此, Faster-Block 中仅在中间的 Conv 层后使用批量归一化和激活操作, 以确保特征多样性和更低的延迟。

Faster-Block 的关键组件是 PConv, 原理可见图 4a。PConv 核心是对第 1 个或最后一个连续的通道作为整个特征图的代表进行计算, 相比传统卷积 (Conv), 即仅需要对一部分输入通道数量 ( $C_p$ ) 的卷积进行空间上的特征提取, 并保留其余通道不变, 通过恒等映射使得该层的特征信息能够流过所有通道, 在有效提取特征的同时减少了计算冗余和内存访问。式(1)和式(2)分别为 Conv 和 PConv 的计算量对比, 以典型的缩放因子  $r = C_p / C$  为例, 当  $r$  为  $1/4$  时, PConv 的计算量仅为 Conv 的  $1/16$ 。

$$F_{\text{Conv}} = HWk^2C^2 \quad (1)$$

$$F_{\text{PConv}} = HWk^2C_p^2 \quad (2)$$

式中:  $H, W$  分别为特征图高和宽;  $k$  为卷积核尺寸;  $C$  为普通卷积的通道数;  $C_p$  为连续的通道数。

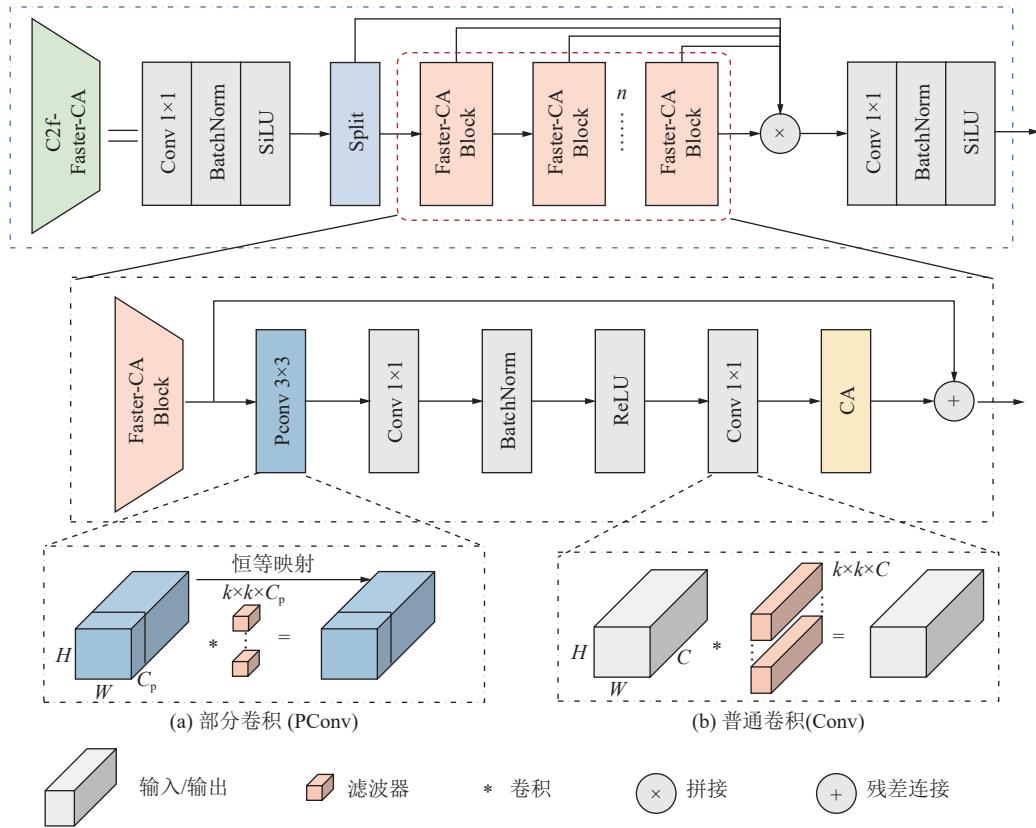


图 4 C2f-Faster-CA 网络架构  
Fig.4 C2f-Faster-CA network architecture

### 1.1.2 坐标注意力模块

为降低 Faster-Block 造成的一部分特征损失影响, 引入了一种轻量化且易嵌入的坐标注意力<sup>[24]</sup>(CA)机制对其进行改进, 旨在加强对重要特征信息提取的同时保持模型轻量高效。CA 网络架构见图 5。

首先将输入特征图分为宽度( $X$ 方向)和高度( $Y$ 方向)2个方向的同时进行全局平均池化, 如下式表示:

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

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (4)$$

式中:  $z_c^h(h)$  为高度为  $h$  的第  $c$  个通道输出;  $z_c^w(w)$  为宽度为  $w$  的第  $c$  个通道输出;  $W, H$  分别为特征图宽和高;  $x_c$  为第  $c$  个通道的输入;  $i$  为高度  $h$  时的宽度坐标值;  $j$  为宽度  $w$  时的高度坐标值。

其次, 将获得全局感受野范围  $X$  方向和  $Y$  方向的特征图进行拼接以实现信息的交换, 进一步经过批量归一化处理后的特征图送入 Sigmoid 激活函数得到特征图  $f$ :

$$f = \delta(F_1([z^h, z^w])) \quad (5)$$

式中:  $F_1$  为拼接操作;  $\delta$  为非线性激活操作。

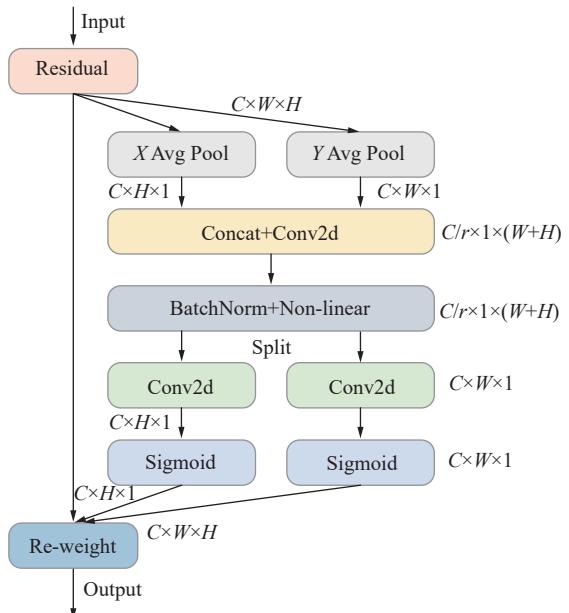


图 5 坐标注意力网络架构  
Fig.5 Coordinate Attention Network Architecture

得到的各方向的特征图能够自适应捕捉输入图像的长程依赖关系, 保存各自的位置信息, 同时采用卷积采用对齐通道数, 得到  $X$  方向注意力权重  $g^w$  和  $Y$  方向的注意力权重  $g^h$ 。最后将各方向注意力权重与输入特征权重进行乘法加权计算, 得到一组新的保留

位置信息的通道注意力特征图  $F'$ , CA 主要的计算公式如下:

$$F' = x_c g_c^h g_c^w \quad (6)$$

式中:  $F'$  为 CA 模块的输出特征图;  $x_c$  为输入特征图权重;  $g_c^w$  为宽度为  $w$  时的注意力权重;  $g_c^h$  为高度为  $h$  时的注意力权重。

## 1.2 损失函数优化

YOLOv8 网络中, 损失函数由分类损失 (BCE) 和边界回归解耦的综合损失 (CIOU+DFL) 组成。CIOU 回归损失考虑了边界框重叠面积和纵横比, 解决了边界框不相交情况下损失值不收敛的现象, 但在煤矿人车检测任务中, 煤矿的地质条件差异较大、环境复杂、遮挡等挑战较为严重, 导致训练样本不平衡, 从而使训练产生较大的回归误差。为应对该挑战, 我们采用 WIOU 损失函数<sup>[25]</sup>作为回归损失替换 CIOU 损失函数。图 6 为 WIOU 的预测框与真实框。

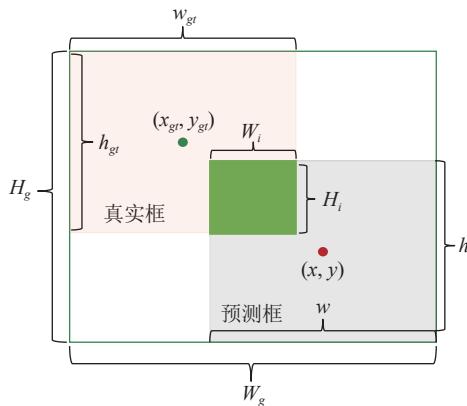


图 6 WIOU 损失函数真实框与预测框

Fig.6 Visualization of true and predicted boxes in wiou loss function

WIOU 首先根据距离度量, 通过消除纵横比因子减少了锚框与目标框重合良好时几何因素的惩罚, 同时对  $\mathcal{L}_{\text{IOU}}$  损失函数指数化得到  $\mathcal{R}_{\text{WIOU}} \in [1, e]$ , 公式如下:

$$\mathcal{R}_{\text{WIOU}} = \exp \left[ \frac{(x - x_{\text{gt}})^2 + (y - y_{\text{gt}})^2}{W_g^2 + H_g^2} \right] \quad (7)$$

$$\mathcal{L}_{\text{IOU}} = 1 - \frac{W_i H_i}{w h + w_{\text{gt}} h_{\text{gt}} - W_i H_i} \quad (8)$$

式中:  $x$  为预测框中心的横坐标,  $y$  为纵坐标,  $w$  为宽,  $y$  为高;  $x_{\text{gt}}$ 、 $y_{\text{gt}}$ 、 $w_{\text{gt}}$ 、 $h_{\text{gt}}$  分别为实际框中心横坐标、纵坐标、宽和高;  $W_i$  和  $H_i$  为预测框和实际框交集的宽和高;  $\mathcal{R}_{\text{WIOU}} \in [1, e]$  显著放大了普通质量锚框损失, 可有效降低样本质量不平衡所导致的误差。由此得到

2 层注意力机制的 WIOU v1:

$$\mathcal{L}_{\text{WIOUv1}} = \mathcal{R}_{\text{WIOU}} \mathcal{L}_{\text{IOU}} \quad (9)$$

## 1.3 模型剪枝

模型修剪主要分为非结构化修剪(权重修剪)和结构化修剪(通道、卷积核、网络层等修剪), 而非结构化修剪主要针对模型权重进行修剪, 需要特定硬件才能实现压缩和加速效果, 难以在煤矿场景的固定设备下实现工业应用, 不作为本文研究重点。

在开发的 FCW-YOLO 检测器中, 每个卷积后大多都要经过一次批量归一化操作, 大量的批量归一化操作以保证数据稳定性, 但同时造成了网络的冗余。然而, 当某个卷积通道的输出经过批归一化层时, 根据式 (10), 通道输出的大小取决于  $\gamma$  的取值, 当  $\gamma$  的值较小时, 意味着该通道的输出对于整个网络的贡献较小, 因此我们可以考虑将该通道进行修剪。

$$\begin{cases} \mu = \frac{1}{m} \sum_i^m x_i \\ \sigma^2 = \frac{1}{m} \sum_i^m (x_i - \mu)^2 \\ \hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \\ y_i = \gamma \hat{x}_i + \beta \end{cases} \quad (10)$$

式中:  $m$  为批次大小。 $\mu$  为输入数据均值,  $\sigma$  为输入数据的标准差,  $x_i$  为输入数据,  $\hat{x}_i$  为经过批量归一化后的输入,  $y_i$  为通道输出。 $\gamma$  和  $\beta$  是训练过程中被不断更新的可学习参数, 分别是尺度因子和偏移量。

根据上述描述, 本文在设计的 FCW-YOLO 煤矿人车轻量级检测器基础上, 进一步采用通道修剪技术对网络进行压缩, 实现更轻量化的检测模型设计。图 7 中红色通道表示贡献较高的保留通道, 灰色通道表示贡献较低的删除通道, 具体步骤主要分为 3 步:

1) 稀疏训练, 在网络训练过程中对批量归一化层(BN)的  $\gamma$  因子施加 L1 正则化, 图 7c 为 L1 正则化稀疏图解,  $\sum_{\gamma \in \Gamma} g(\gamma) = \sum_{\gamma \in \Gamma} |\gamma|$  即所有  $\gamma$  元素绝对值的和, 其函数图像为正方形, 而原函数的损失函数为一个椭圆形状, 通过施加一个惩罚项(稀疏项)  $\lambda \sum_{\gamma \in \Gamma} g(\gamma)$ , 与原来的损失函数产生竞争作用, 两者相交时使得整体损失最小, 即可将较多  $\gamma$  逼近 0。最后将其乘以对应层的输出通道, 以达到稀疏化的作用, 稀疏后网络(图 7a)的冗余通道因子大多趋向于 0, 稀疏公式如下:

$$L = \sum_{(x,y)} l(f(x,W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (11)$$

式中:  $\sum_{(x,y)} l(f(x,W), y)$  为基线模型的损失函数;  $x, y$  分别

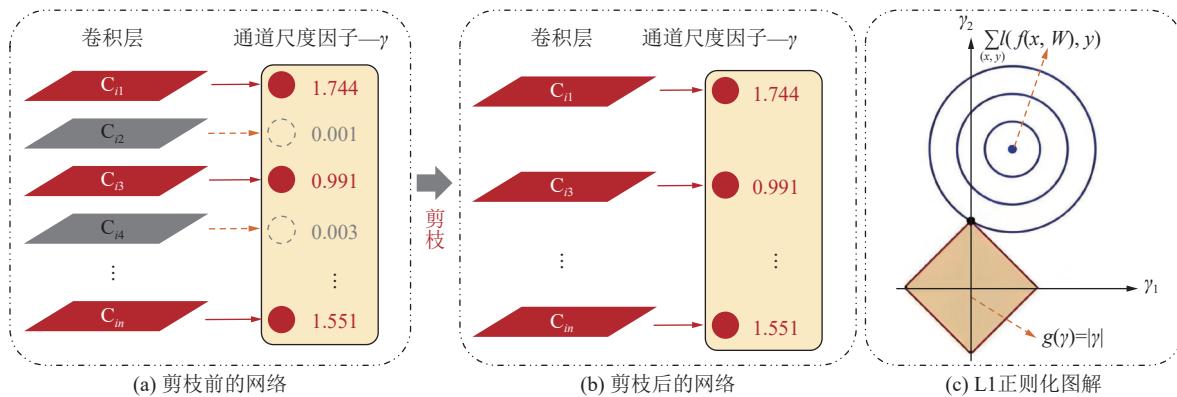


图 7 通道剪枝示意

Fig.7 Channel pruning illustration

为目标输入输出;  $W$  为训练权重;  $\lambda$  为稀疏因子;  $g(\gamma)$  为稀疏化的稀疏函数, 其中  $g(\gamma) = |\gamma|$ ;

2) 修剪, 经过稀疏化训练后的网络, 在每个卷积层中的通道中都有着对应的缩放因子  $\gamma$ , 对  $\gamma$  因子按从小到大排序, 进一步设置一个全局剪枝阈值(剪枝率), 小于该阈值的通道将被裁剪, 用以控制需要裁剪通道的比例。图 7b 展示了裁剪后的网络模型,  $\gamma$  因子较小的通道被裁剪后, 得到一个稀疏紧凑的网络。

3) 微调, 经过修剪后的网络由于卷积层通道结构的改变, 网络变得紧凑, 造成了部分精度损失, 因为需要对剪枝后网络重新训练微调, 即将剪枝后的权重文件作为微调阶段的预训练权重, 具体原理为: 首先生成一个与参数矩阵尺寸一致的 mask 矩阵, 根据上述设置的全局修剪阈值将删减的通道中的权值设为 0, 其余为 1, 在计算前将该 mask 与参数矩阵相乘, mask 为 1 的值继续训练并通过反向传播调整梯度, mask 为 0 的部分输出始终为 0, 因此不会对模型的其余计算产生影响, 由此即可得到几乎不损失精度的高压缩煤矿人车检测模型。

## 2 煤矿人车数据集构建

自建的煤矿人车(煤矿工人和载人车辆)(Coal mine personnel and manned vehicle detection, Cpmv)检测数据集来源于贵州省多个煤矿, 因此将其命名为Cpmv, 选用煤矿本质安全型号为KBA165的防爆监控摄像头获取, 采集分辨率为 $1920\times1080$ , 采集的各个摄像头高度为 $2.5\sim3$  m, 摄像头俯角在 $40^\circ\sim70^\circ$ , 以确保最大化监控范围, 避免拍摄到过多噪声影响图像质量。为实现具有众多挑战的煤矿场景数据集构建, 在主平硐入口(白天场景和夜晚场景)、水泵房、皮带机口、架空乘人机头、机尾、中段等多个摄像头处, 通过VS player软件将采集到的监控视频切分为图片。同时我们尽量考虑符合实际情况的密集, 遮挡场景,

在煤矿工程师的指导下从中筛选特征信息相似度较低的 2 272 张图片。自建的 Cpmv 数据集中包含了煤矿工人、平巷人车和煤矿架空乘人装置(俗称猴车)等具代表性的煤矿人车样本。共计 4 个类别, 分别为煤矿工人, 平巷人车, 载人状态的猴车, 无载人状态的猴车, 采用 labelimg 标注软件分别将其标注为 miner、PHRC、HC-m、HC-nom, 按照 7 : 2 : 1 的比例随机划分为训练集(1 590 张)、验证集(454 张)和测试集(228 张), 其中验证集用于模型的评估, 测试集用于模型的推理可视化展示。Cpmv 数据集分布情况见表 1。图 8a 给出了自建数据集标注示例, 图 8b 给出了标注数据集相对宽高比例的统计分布。

表 1 煤矿人车数据集划分及分布  
Table 1 Division and distribution of coal mine pedestrian-vehicle dataset

类别	标注	训练集	验证集	测试集	合计
煤矿工人	miner	2 700	812	372	3 884
平巷人车	PHRC	344	76	46	466
载人的猴车	HC-m	816	228	136	1 180
无载人的猴车	HC-nom	724	232	112	1 068
	合计	4 584	1 348	666	6 598

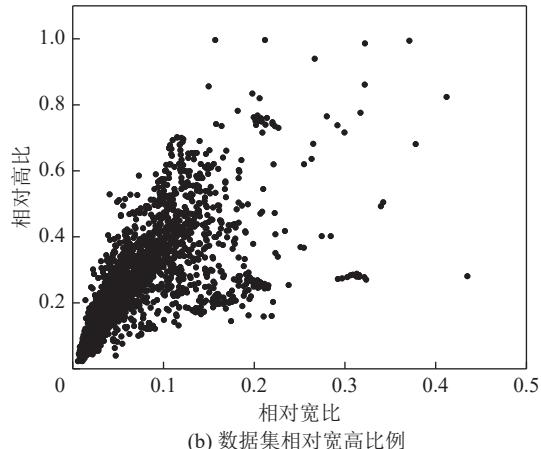
## 3 实验与结果

### 3.1 实验环境及评估指标

本文工作开展的主要硬件配置和软件配置如表 2 所述。图像输入尺寸为 $640\times640$ , 训练轮次为 200, 批次为 32, 学习率为 0.001, 优化器为 Adamw, 在最后 20 个轮次关闭马赛克增强以保证训练的稳定。同时采用了式(12)—式(15)的指标以评估模型的检测性能。另外, 一些通用指标如网络参数量, 计算量和模型大小被用以衡量网络复杂度。推理时间被用于衡量模型的检测速度, 即模型处理每张图像的用时。



(a) 数据集标注及各场景示例



(b) 数据集相对高宽比例

图8 自建的数据集标注及统计分布

Fig.8 Self-constructed dataset labeling and statistical distributions

表2 实验环境配置

Table 2 Experimental Environment Configuration

配置	名称	型号
硬件	CPU×2	Intel Xeon 8275CL
	GPU	NVIDIA RTX A6000(48G)
	RAM×8	32G 3 200 DDR4 RECC(256G)
软件	Python	3.9
	Pytorch	2.0
	Cuda	11.8

$$P = \text{TP}/(\text{TP} + \text{FP}) \quad (12)$$

$$R = \text{TP}/(\text{TP} + \text{FN}) \quad (13)$$

$$\text{AP} = \int_0^1 P(R)d(R) \quad (14)$$

$$\text{mAP} = \sum_i^N \text{AP}_i / N \quad (15)$$

式中:  $P$  为精确率, 用于评判检测为正样本的概率;  $R$  为召回率, 用于评判检测框漏检情况, 其值越大模型漏检率越低;  $\text{AP}$  为平均精度, 是 P-R 曲线下的面积, 表示模型在不同召回率下的平均精确率;  $\text{mAP}$  为平均精度均值, 衡量了模型在多个检测类别上的平均精度, 是综合评价模型检测性能的重要指标, 本文选取了 IOU 阈值为 0.5 时计算 AP 及 mAP;  $\text{TP}$ 、 $\text{FN}$  和  $\text{FP}$  分别为正检框、漏检框和误检框数目;  $N$  为检测类别数。

### 3.2 FCW-YOLO 实验分析

为验证设计及改进模块的有效性, 对基线模型开展了表3所示的消融实验, 其中模型1为基线模型(YOLOv8s), 模型2—模型8为改进模块不同的排列组合所得到。单独采用 Faster-Block 对基线改进后(模型1), 模型复杂度指标大幅降低, 同时检测精度并没有明显降低。在单独采用注意力模块对基线改进后(模型2), 模型复杂度指标仅有少许增幅, 但检测精度有明显上涨。进一步将 CA 融合到 Faster-Block 中对基线改进后得到模型5, 此时模型复杂度降低的同时, 检测性能也得到了有效提升。根据模型4结果, 采用 WIOU 损失函数改进后, 模型总体检测精度也有较大改善, 采用 Faster-Block 联动 WIOU 后(模型5), 模型检测性能进一步提升, 并且模型更加轻量化。通过模型7和模型8的对比发现, 模型7对煤矿工人类别的检测精度为消融实验中最高, 但平均精度均值却不如模型8, 总体而言在单独采用 Faster-Block 时精度虽有略微丢失现象, 但在结合了 CA 和 WIOU 模块的优势后, CA 一定程度弥补了 Faster-Block 中部分卷积丢弃通道信息损失的影响, 使得提出的模型8即轻量又高效准确, 证明了 Faster-Block 在耦合使用时不仅能增加检测性能还能有效降低网络开销。此外, 根据模型4, 6, 7, 8结果对比可知, WIOU 的引入对模型复杂度和推理时间没有影响, 同时与各模块组合优化时, 检测性能都得到了提升。图9为基线模型和改进模型的边界回归损失对比, 采用 WIOU 改进后, 模型的损失值收敛得更快, 最终的边界回归误差也更低。

本文提出的模型参数量为 8.3 M, 计算量为 22.0 GFLOPs, 模型大小为 16.1 MB, 对比基线模型, 分别降低了 25.9%, 22.5%, 24.8%。根据表3消融实验结果和图10改进前后 P-R 曲线对比, 改进后 P-R 曲线表现更好, 煤矿工人, 平巷人车, 载人猴车和无载人猴车的精度分别提升了 3.0%, 3.4%, 2.5%, 1.7%, 总体精度提升了 2.6%, 这说明了笔者提出方法是合理的, 在煤矿人车检测任务中能较好的识别和定位目标。同时

表 3 消融实验

Table 3 Ablation Experiments

模型	改进模块			模型复杂度评估指标			模型精度评估指标				模型速度评估指标	
	Faster-Block	CA	WIOU	参数量/M	计算量/GFLOPs	模型大小/MB	AP <sub>miner</sub> /%	AP <sub>PHRC</sub> /%	AP <sub>HC-m</sub> /%	AP <sub>HC-nom</sub> /%	mAP/%	推理时间/ms
1(基线)				11.2	28.4	21.4	91.5	95.1	83.2	80.6	87.6	6.4
2	√			8.2	21.4	16.0	92.7	92.3	84.9	79.2	87.3	6.9
3		√		11.3	28.6	21.6	92.9	96.9	81.3	82.1	88.3	8.3
4			√	11.2	28.4	21.4	93.1	94.4	86.1	81.1	88.7	6.4
5	√	√		8.3	22.0	16.1	93.5	95.7	82.8	83.7	88.9	9.8
6	√		√	8.2	21.4	16.0	93.7	95.6	84.9	83.1	88.8	6.9
7		√	√	11.3	28.6	21.6	96.1	95.2	84.9	81.9	89.6	8.3
8	√	√	√	8.3	22.0	16.1	94.5	98.5	85.7	82.3	90.2	
本文				(-25.9%)	(-22.5%)	(-24.8%)	(+3.0%)	(+3.4%)	(+2.5%)	(+1.7%)	(+2.6%)	9.8

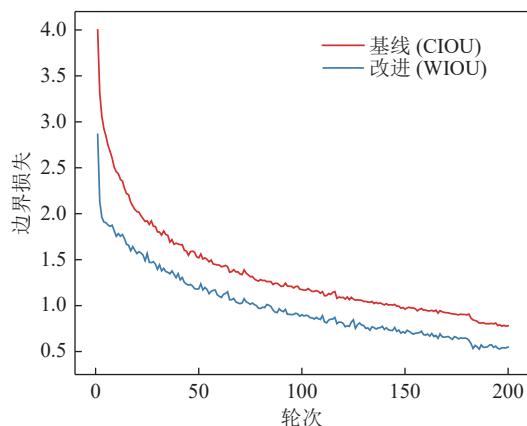


图 9 边界回归损失函数曲线对比

Fig.9 Comparison of boundary regression loss function curves

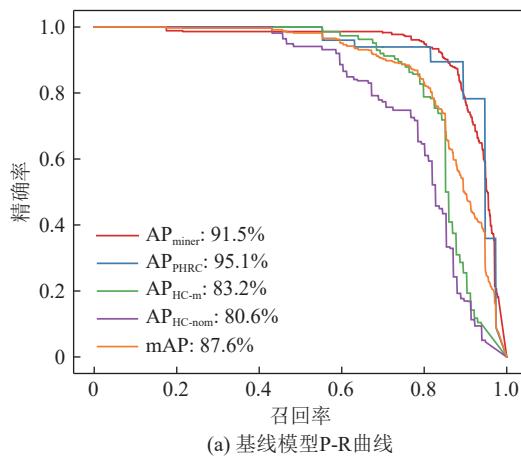
由于改进模块无法避免网络堆叠次数的增加,模型内存访问量会因此增加,推理时间随之变慢,笔者的方法推理时间为9.8 ms,虽然比基线高了3.2 ms,但仍然是低于30 ms的实时检测要求。综上,消融实验结果表明,单个模块改进对模型的复杂度和检测精度都有改善,并且复合模块的改进效果更好,优于单一模块改进,我们提出的模型从模型轻量化,检测性能,速度多方面考虑更加优越。

### 3.3 通道剪枝实验分析

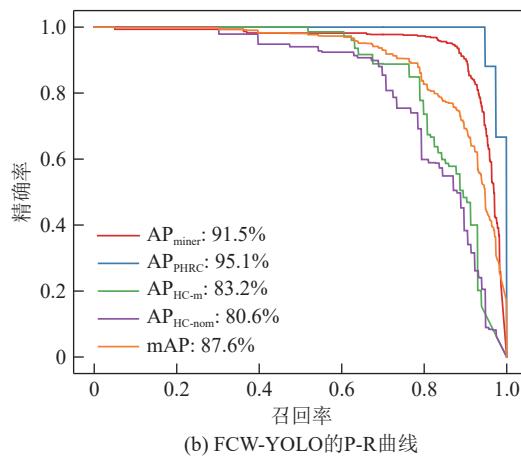
表4给出了对提出的FCW-YOLO检测模型在通道修剪阶段的实验设置,将剪枝后的模型记为FCWP-YOLO。除剪枝率外,其余超参数根据刘等<sup>[22]</sup>的工作确认。通过在0.1~0.9范围内迭代以测试模型的综合性能来确定最优剪枝率,其选择原则为几乎不损失精度的情况下,最大化的压缩模型。

#### 3.3.1 稀疏训练阶段

图11为Tensor-board可视化软件导出的FCW-YOLO稀疏训练前后的BN层通道尺度因子 $\gamma$ 分布对



(a) 基线模型P-R曲线



(b) FCW-YOLO的P-R曲线

图 10 改进前后的P-R曲线

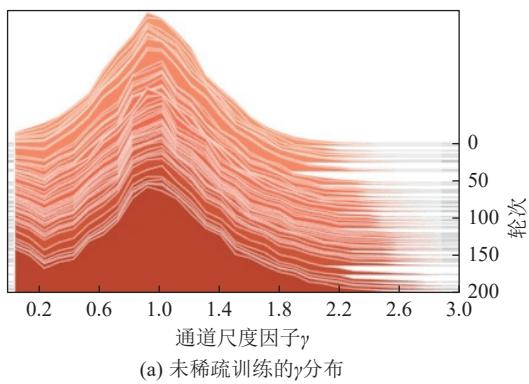
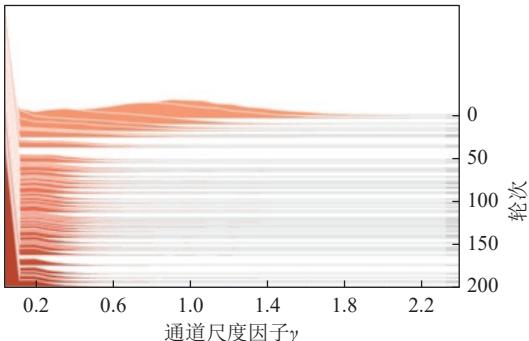
Fig.10 P-R Curve Comparison Before and After Improvement

比。FCW-YOLO没有稀疏训练时(图11a),此时 $\gamma$ 呈正态分布,仅有少部分 $\gamma$ 值趋于0,模型无法根据 $\gamma$ 大小筛选是否为重要通道,无法进行下一步剪枝操作。FCW-YOLO稀疏训练后(图11b), $\gamma$ 从正态分布逐渐过渡到基本趋近于0,模型变得稀疏,并且到后期 $\gamma$ 分布变化不大,说明稀疏训练达到稳定状态,此时即可

表4 FCW-YOLO 在通道剪枝阶段的实验超参数

Table 4 Experimental hyperparameters of FCW-YOLO in channel pruning stage

实验阶段	超参数	数值
稀疏训练	学习率	0.001
	批次大小	32
	稀疏因子( $\lambda$ )	$10^{-4}$
	轮次	200
剪枝	剪枝率*	[0.1, 0.9]
微调	与训练阶段一致	

(a) 未稀疏训练的 $\gamma$ 分布(b) 稀疏训练后的 $\gamma$ 分布图11 FCW-YOLO 模型稀疏训练前后的 $\gamma$ 分布Fig.11 Distribution of  $\gamma$  before and after sparse training of FCW-YOLO model

根据剪枝率修剪不重要的通道。

### 3.3.2 剪枝和微调

在剪枝和微调阶段, 对不同剪枝率([0.1-0.9])的FCW-YOLO模型进行了实验, 评估结果见表5, 表5中mAP为经过微调后模型的检测精度。随着剪枝率的增加, 模型的参数量、计算量、模型体积和推理时间不断下降。然而, 模型的mAP随着剪枝率的增加不断波动, 在修剪率为0.4时mAP达到最高。修剪率为0.7时, 对比没有修剪的FCW-YOLO模型, mAP为88.7%, 在仅损失1.5%精度的情况下模型的开销大幅降低, 推理时间仅为5.6 ms。但在修剪率达到0.8以后, 模型的检测性能急剧下降。剪枝率为0.9时, 模

型压缩效果最好, 但损失过多检测性能, 精度仅为32.6%。因此, FCW-YOLO模型的剪枝率设置为0.7最为合适。

表5 不同剪枝率下FCW-YOLO模型的性能指标

Table 5 Performance metrics of FCW-YOLO model at different pruning rates

剪枝率	参数量/ M	计算量/ GFLOPs	模型大小/ MB	mAP/ %	推理时间/ ms
0	8.3	22.0	16.1	90.2	9.8
0.1	7.3	18.9	14.3	89.6	8.6
0.2	6.2	16.3	12.2	88.5	8.1
0.3	5.5	14.1	10.9	89.7	7.9
0.4	4.5	11.7	8.7	91.4	7.2
0.5	3.7	9.6	7.4	90.1	6.8
0.6	3.2	8.2	6.4	89.6	6.1
<b>0.7</b>	<b>2.3</b>	<b>6.0</b>	<b>4.9</b>	<b>88.7</b>	<b>5.6</b>
0.8	1.6	4.2	3.6	76.7	5.4
0.9	0.8	2.1	2.3	32.6	4.6

注: 剪枝率为0时表示没有进行剪枝的FCW-YOLO模型。

图12为剪枝率0.7时FCW-YOLO模型剪枝前后各网络层通道数对比。剪枝后, 各层通道数量大幅减少, 尤其是18~62层的网络通道得到大幅修剪, 同时可以看出网络层数越深的网络, 修剪比例就越大, 说明了网络层数较深的层存在了大量的冗余通道。修剪后, 通道总数从11 781减少到了3 443, 通道删减比例为70.1%, 这与设置的剪枝比例基本一致。这表明了通道剪枝方法使我们设计的模型更加紧凑, 并且这几乎不损失模型检测性能, 该方法能够有效简化FCW-YOLO网络结构, 实现模型压缩, 验证了我们提出方法的有效性。

表6给出了基线与改进阶段和模型压缩阶段性能对比的总结。对比基线, 笔者设计的FCW-YOLO展现了出色的检测性能, 并减少了计算成本, 但推理速度有所增加。在通道剪枝阶段, 模型计算成本大幅降低, 由于模型通道结构的改变, 同时检测性能也一定下降趋势。但通过微调以后, 模型的检测性能几乎恢复到了没有修剪之前, 虽然这是牺牲了1.5%的检测精度实现的, 但其检测性能仍然是优于基线, 并且推理速度也比基线要快。此时得到了高度压缩且检测性能优异的煤矿人车检测器, 即FCWP-YOLO。总的来说, 对比基线, 笔者提出方法的参数量、计算量和模型大小的压缩比分别达到了4.9、4.7、4.4倍, 压缩效果远优于设计的轻量级FCW-YOLO, 检测精度为88.7%, 增加了1.1%, 推理速度仅为5.6 ms。这是目前

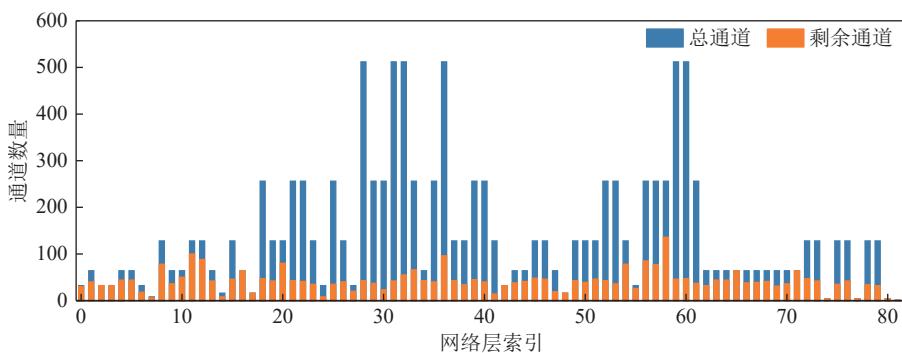


图 12 剪枝前后 FCW-YOLO 的通道数量对比

Fig.12 Comparison of channel numbers before and after pruning FCW-YOLO

表 6 基线与改进阶段和模型压缩阶段的性能对比。

Table 6 Performance Comparison between Baseline, Improvement, and Model Compression Stages

阶段	模型	参数量/M	压缩比	计算量/GFLOPs	压缩比	模型大小/MB	压缩比	mAP/%	推理时间/ms
基线	YOLOv8s	11.2	/	28.4	/	21.4	/	87.6	6.4
轻量级改进	FCW-YOLO*	8.3	1.4×	22.0	1.3×	16.1	1.3×	90.2(+2.6%)	9.8
剪枝	FCWP-YOLO*	2.3	4.9×	6.0	4.7×	4.9	4.4×	80.5(-7.1%)	5.6
微调	FCWP-YOLO*	2.3	4.9×	6.0	4.7×	4.9	4.4×	88.7(+1.1%)	5.6

注：带\*的模型均为本文提出的方法，压缩比计算为基线网络开销与改进后的网络开销比值。

替换网络无法达到的效果，验证了笔者对基线二次轻量化设计的有效性。

## 4 讨 论

### 4.1 轻量级网络对比

轻量级网络在能有效减少模型参数量和计算复杂度，这包括一些轻量级主干和轻量级颈部的开发。对此在表 7 中，采用了多个先进的轻量级骨干网络和颈部对 YOLOv8s 网络进行了替换，这包括 ShuffleNetV2<sup>[26]</sup>，MobileNetV3-s<sup>[27]</sup>，EfficientNet-lite0<sup>[28]</sup>，Slim-neck<sup>[29]</sup>，这些轻量级网络用于对比笔者提出的 FCWP-YOLO 检测模型性能。其中，参数量、计算量、模型大小、推理时间越小越好；mAP 越大越好。对比基线，轻量级网络的替换使得网络开销都得到了不同程度

的优化，但检测速度变化却不一样。ShuffleNetV2 和 EfficientNet-lite0 的推理时间减少了，MobileNetV3-s 和 Slim-neck 推理时间却增加了，这表明了轻量级网络的规模对推理时间是有影响的，但并不是成正比关系。另一方面，大多轻量级网络有效减少了计算成本，但 ShuffleNetV2 和 MobileNetV3-s 检测性能下降较多，这是以牺牲了较多精度为代价实现的，并且它们的轻量级程度也远不如我们提出的模型。EfficientNet-lite0 检测精度是高于基线的，但是计算成本远高于我们提出的方法。这表明了轻量级网络的替换对于 YOLOv8s 的改进作用是有限的，网络开销的降低明显不如通道剪枝方法带来的收益大。然而，我们的方法 FCWP-YOLO 考虑了轻量级架构，注意力模块，损失函数，通道修剪等多维度因素，各项指标都得到了正

表 7 轻量级网络对比实验

Table 7 Comparative experiment of lightweight networks

轻量级网络	参数量/M	计算量/GFLOPs	模型大小/MB	mAP/%	推理时间/ms
YOLOv8-s(基线)	11.2	28.4	21.4	87.6	6.4
ShuffleNetV2	<b>6.1(-45.5%)</b>	<b>15.9(-44.0%)</b>	<b>11.9(-44.4%)</b>	82.3(-5.3%)	<b>5.1(-20.3%)</b>
MobileNetV3-s	<b>6.5(-42.0%)</b>	<b>16.3(-42.6%)</b>	<b>12.8(-40.2%)</b>	84.3(-3.3%)	6.7(+4.7%)
EfficientNet-L0	<b>9.3(-17.0%)</b>	<b>22.0(-22.5%)</b>	<b>18.1(-15.5%)</b>	<b>87.9(+0.3%)</b>	<b>5.7(-10.9%)</b>
Slim-neck	<b>10.5(-6.2%)</b>	<b>25.6(-9.9%)</b>	<b>20.4(-4.7%)</b>	87.1(-0.5%)	6.9(+7.8%)
FCWP-YOLO(Ours)	<b>2.3(-79.5%)</b>	<b>6.0(-78.9%)</b>	<b>4.9(-77.1%)</b>	<b>88.7(+1.1%)</b>	<b>5.6(-12.5%)</b>

注：表中加粗表示模型指标相比基线得到正优化。

优化,实现了检测精度和处理效率之间的平衡,进一步表明了通道剪枝方法比轻量级网络的替换更为有效,仅用较少成本的网络开销即可实现先进的检测性能。

#### 4.2 主流检测模型对比

为验证提出方法的优越,将提出的FCWP-YOLO模型与多个主流的先进检测器进行了对比实验,包括

两阶段的Faster-RCNN<sup>[7]</sup>、一阶段的SSD<sup>[9]</sup>、YOLOv3-tiny<sup>[30]</sup>、YOLOv5<sup>[13]</sup>、YOLO6<sup>[14]</sup>、YOLOv8<sup>[22]</sup>、基于Transformer架构的DETR<sup>[8]</sup>、Deformable-DETR<sup>[31]</sup>等。为保证实验的一致性,采用相同的实验环境,数据集和超参数设置。其中选用的Faster-RCNN, DETR和Deformable-DETR主干网络均为Resnet50, SSD主干网络为VGG16。结果如表8所示。

表8 主流检测网络对比实验

Table 8 Comparative Experiment of Mainstream Detection Networks

模型	参数量/M	计算量/GFLOPs	模型大小/MB	mAP/%	推理时间/ms
YOLOv8-s(基线)	11.2	28.4	21.4	87.6	6.4
Faster-RCNN	40.6(+262.5%)	208.4(+633.8%)	320.4(+1 397.2%)	74.5(-13.1%)	52.3(+717.2%)
SSD	23.8(+112.5%)	30.3(+6.7%)	183.8(+758.9%)	79.6(-8.0%)	18.9(+195.3%)
YOLOv3-tiny	103.6(+825.0%)	282.2(+893.7%)	207.8(+871.0%)	85.6(-2.0%)	7.9(+23.4%)
YOLOv5-s	<b>9.1(-18.8%)</b>	<b>23.8(-16.2%)</b>	<b>17.4(-18.7%)</b>	<b>87.7(+0.1%)</b>	<b>6.0(-6.3%)</b>
YOLOv6-n	<b>4.2(-62.5%)</b>	<b>11.8(-58.5%)</b>	<b>8.3(-61.2%)</b>	84.2(-3.4%)	6.5(+1.6%)
YOLOv8-x	68.1(+508.0%)	258.1(+808.8%)	274.9(+1 184.6%)	<b>89.8(+2.2%)</b>	12.3(+92.2%)
DETR	41.5(+270.5%)	94.9(+234.2%)	477.2(+2 129.9%)	83.6(-4.0%)	78.1(+1 120.3%)
Deformable-DETR	40.1(+258.0%)	190.0(+569.0%)	400.7(+1 772.4%)	84.9(-2.7%)	57.1(+792.2%)
FCWP-YOLO(Ours)	<b>2.3(-79.5%)</b>	<b>6.0(-78.9%)</b>	<b>4.9(-77.1%)</b>	<b>88.7(+1.1%)</b>	<b>5.6(-12.5%)</b>

注: 表中加粗表示模型指标相比基线得到了正优化。各指标的增减百分比计算为“(对比模型指标-基线指标)/基线指标”。

根据上表实验结果可知,两阶段的算法Faster-RCNN不仅检测性能较差,而且网络的开销也较大,推理速度较慢,说明了早期的两阶段算法难以适应煤矿复杂条件的人车检测任务。基于Transformer架构端到端的DETR检测器不仅检测精度较低,模型体积也较大,推理速度为所有对比模型中最慢,尽管Deformable-DETR通过可变形卷积改善了DETR的检测性能,但模型体积和推理速度仍然较慢,显然不满足本文实时检测要求(30 ms),以Transformer为架构的检测算法对于大型数据集全局信息处理更有优势,但并不适用于本文中小型数据集,此外自注意力机制使得模型的复杂度呈指数增加,推理延迟过长,不满足视频流信息的实时推理。在实验的单阶段检测算法中,SSD的检测精度是最低的,仅为79.6%,相比其他单阶段等比规模的网络,SSD的模型也更为复杂,检测速度虽达到了实时检测要求,但仍不如YOLO系列。大多YOLO系列的先进检测算法的精度基本都在85%以上,并且检测速度都较快,但难以实现检测性能和检测效率的平衡。如YOLOv6n的网络较为轻量,但检测性能是不如较大规模的x系列的,YOLOv8x的检测精度为对比实验中最高的,但网络计算成本仍然较高,对于资源受限的煤矿场景而言,部署这类较

大模型仍然存在挑战。然而,我们的方法网络计算成本均为最低,同时检测性能也达到了头部水平,仅低于YOLOv8x,但提出的模型参数量,计算量和模型体积远远小于YOLOv8x,推理速度为最快,证明了我们提出的模型在人车检测任务中轻量化、实时性和准确性方面的优越。

图13展示了部分主流模型和提出模型在自建的煤矿人车检测数据集中的推理可视化,其中图13a代表密集场景的主井闸机入口场景,仅有煤矿工人类别;图13b和图13c分别代表主平硐白天和夜晚场景,包含煤矿工人和平巷人车类别;图13d代表水泵房场景,仅有煤矿工人类别;图13e代表皮带机头场景,仅有煤矿工人类别;图13f和图13g代表猴车机头场景,包含煤矿工人、载人猴车和无载人猴车类别;图13h代表机头中段场景,仅有载人猴车类别。根据图13检测结果可知,Faster-RCNN在图13a中密集场景出现冗余框现象,在图13b中遮挡场景出现定位不准现象;YOLOv3-tiny在图13a和图13b中不仅出现定位不准,还出现了漏检的现象,在图13c中的夜晚场景也出现了冗余框现象;SSD的检测表现较差,在图13f、图13g和图13h等密集遮挡场景都出现了漏检现象和部分定位框不准现象,同时在图13(f)中还将载人猴车

识别成了无载人猴车;本文的基线模型 YOLOv8s 仅在图 13f 中出现了小目标载人猴车的漏检;在图 13d、图 13e 等几乎无遮挡,密集的简单场景,所有模型都正确检测到目标;笔者提出的模型 FCWP-YOLO 即使在大多背景复杂的情况下也可准确的预测和定位各类场景下的煤矿人车目标,仅在图 13a 密集遮挡场景

发生了一个目标漏检的情况,对比其他模型的检测可视化仍然是较为精准的。综合上述大量的对比实验和可视化对比,提出的方法在煤矿各类场景出色的完成了人车检测任务,在轻量化和实时性方面表现出了巨大的优势,同时检测性能达到了先进的水平,可有效提升煤矿智能巡检工作的水平和效率。

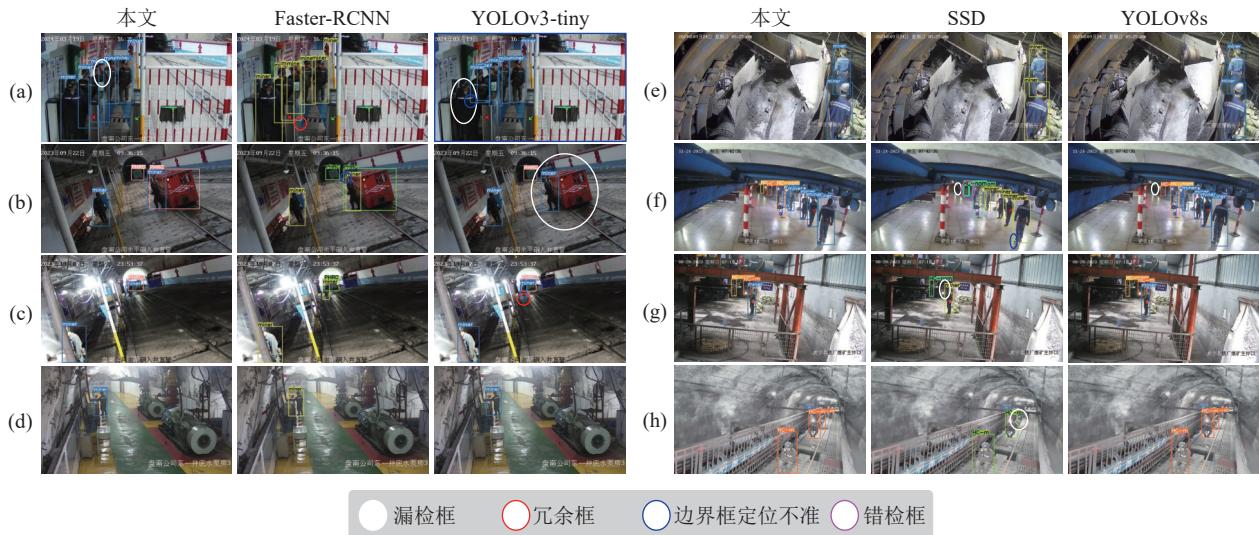


图 13 主流模型与提出模型的检测可视化

Fig.13 Detection visualization of mainstream models and proposed models

#### 4.3 视频流推理验证

为验证提出模型在实际应用中的效果,我们在测试视频中的多个典型场景进行了视频流的推理测试实验,并在每个场景中随机选取了 3 个片段的推理画面进行展示,同时给出了视频流推理的平均速度,结果见图 14。图 14a 为主井入口场景,共计 2 450 帧视频流画面,包括煤矿工人类别和平巷人车类别,平均每帧图像推理速度为 6.0 ms;图 14b—图 14c 为猴车机头场景,共计 1 703 帧视频流画面,包括煤矿工人类别和载人猴车和无载人猴车类别,平均每帧图像推理速度为 6.2 ms;图 14d 为主井闸机场景,共计 451 帧视频流画面,仅有煤矿工人类别,平均每帧图像推理速度为 6.3 ms。在多个场景的视频流推理可视化展示表明,提出的 FCWP-YOLO 模型可适应大多场景的煤矿人车检测任务,模型能够有效辨识检测任务中的多目标物体,即煤矿工人和载人车辆的差异,几乎没有出现漏检错检的现象,模型的鲁棒性能较为稳定,检测性能也较好。同时,提出模型在多个场景下的平均检测推理速度为 6.17 ms,保证了实际视频流检测任务中的实时性,此外通过与上节得到的在 0.7 剪枝率条件下的 FCWP-YOLO 推理速度相比 (5.6 ms),视频流检测增加了一定的推理延迟,这是由于视频流推理时需

要一定的解码时间,但模型在各场景下的推理时间波动不大,基本都稳定在 6.2 ms 左右,推理时间仍然是低于 30 ms,远满足实时检测要求,进一步说明了提出模型在实际应用中准确性和实时性方面的强大表现力。

#### 4.4 局限与未来工作

提出的方法是初次应用于煤矿人车检测任务,该方法不以牺牲检测精度为代价实现了网络模型的有效压缩,但还存在一些不足,在未来的工作的将做出如下改进:①在密集场景及遮挡场景检测任务中出现了少许的漏检情况,因此在检测网络设计阶段,可引入采样范围较大的卷积如空洞卷积,可变形卷积以应对复杂多变的空间特征;②本文仅对煤矿工人,平巷人车,载人和无载人猴车类别开展了轻量级的智能检测研究,下一步将细分煤矿人车状态类别,如煤矿工人的行走状态,工作状态,危险行为等,旨在进一步提升煤矿安全管理水;③该工作仅调研了贵州地区煤矿视频监控数据集,并开发了一种新型的轻量级人车检测模型。然而,南方煤矿和北方煤矿地质和生产条件存在一定的差异性,故下一步将对北方煤矿的视频监控进行调研,采集更多数据集并集成到我们设计的模型中,使得我们的方法能够泛化到各类煤矿场景。



图 14 FCWP-YOLO 在不同场景的视频流推理展示

Fig.14 Demonstration of FCWP-YOLO's video stream inference in different scenarios

## 5 结 论

1) Faster-Block 中的部分卷积可减少重复特征的提取,降低计算成本,坐标注意力的引入可抑制 Faster-Block 仅计算部分通道造成的特征损失影响,加强关键信息的提取,降低煤矿复杂背景干扰。WIOU 边界损失函数增加了对训练过程中普通质量样本的关注,降低了由高质量样本主导梯度带来的回归误差。

2) 采用通道剪枝技术对设计的 FCW-YOLO 网络中的批量归一化层的尺度因子施加正则化稀疏训练后,可筛选趋于零的冗余通道,模型通过剪枝和微调后,相比基线,模型的计算成本得到了有效降低,同时检测精度提升了 1.1%,这是替换轻量级架构无法达到的效果。

3) 自建的数据集实验结果表明,对比多种轻量级网络和主流先进检测网络,提出的 FCWP-YOLO 轻量级人车检测模型在煤矿各类场景下出色完成了检测任务,模型在轻量级,实时性方面的表现均为最优,同

时检测性能也达到了先进水平。我们的方法可有效降低煤矿智能化装备的研发成本,在未来我们将使用提出的方法探索更多煤矿实际应用场景下的有效性和可行性,推动煤矿智能化发展进程。

4) 通道剪枝算法可有效压缩网络模型,但在稀疏训练阶段需耗费较多的时间和显存,未来将进一步结合知识蒸馏技术,应用通道剪枝方法对学生模型进行修剪。在修剪过程中,可以利用教师模型的知识作为剪枝的依据,保留对教师模型预测影响较大的通道,减少通道数量的同时提高网络的训练效率。

## 参考文献(References):

- [1] 王国法,庞义辉,任怀伟,等.智慧矿山系统工程及关键技术研究与实践[J].煤炭学报,2024,49(1): 181–202.  
WANG Guofa, PANG Yihui, REN Huaiwei, et al. System engineering and key technologies research and practice of smart mine[J]. Journal of China Coal Society, 2024, 49(1): 181–202.
- [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 Huawei. Analysis on key technologies of intelligent coal mine and intelligent mining[J]. Journal of China Coal Society, 2019, 44(1): 34–41.
- [3] 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.
- [4] 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.
- [5] 鲍久圣, 张可琨, 王茂森, 等. 矿山数字孪生 MiDT: 模型架构、关键技术及研究展望[J]. 绿色矿山, 2023, 1(1): 166–177.
- BAO Jiusheng, ZHANG Kekun, WANG Maosen, et al. Mine Digital Twin: Model architecture, key technologies and research prospects[J]. Journal of Green Mine, 2023, 1(1): 166–177.
- [6] 王学文, 王孝亭, 谢嘉成, 等. 综采工作面 XR 技术发展综述: 从虚拟 3D 可视化到数字孪生的演化[J]. 绿色矿山, 2024, 2(1): 75–84.
- WANG Xuewen, WANG Xiaotong, XIE Jiacheng, et al. Review of XR technology development in fully mechanized mining faces: From 3D visualization to digital twin[J]. Journal of Green Mine, 2024, 2(1): 75–84.
- [7] 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.
- [8] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 213–229.
- [9] BERG A C, FU C Y, SZEGEDY C, et al. SSD: Single Shot MultiBox Detector[J]. 2015. DOI: [10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [10] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 6517–6525.
- [11] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. arXiv e-prints, 2018. DOI: [10.48550/arXiv.1804.02767](https://doi.org/10.48550/arXiv.1804.02767).
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. 2020. DOI: [10.48550/arXiv.2004.10934](https://doi.org/10.48550/arXiv.2004.10934).
- [13] Glenn J. YOLOv5 by Ultralytics[EB/OL]. <https://github.com/ultralytics/yolov5>.
- [14] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv: 2209.02976, 2022.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016: 779–788.
- [16] 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.
- [17] SHAO X Q, LIU S B, LI X, et al. Rep-YOLO: An efficient detection method for mine personnel[J]. Journal of Real-Time Image Processing, 2024, 21(2): 28.
- [18] XU Z, LI J Z, ZHANG M. A surveillance video real-time analysis system based on edge-cloud and FL-YOLO cooperation in coal mine[J]. IEEE Access, 2021, 9: 68482–68497.
- [19] 江帆, 刘一凡, 朱真才, 等. 基于多特征融合的煤矿井下辅助运输车辆可行性评估方法[J]. 煤炭学报, 2024, 49(12): 4986–5001.
- JIANG Fan, LIU Yifan, ZHU Zhencai, et al. Accessibility evaluation method of auxiliary transport vehicles in coal mine based on multi-feature fusion[J]. China Industrial Economics, 2024, 49(12): 4986–5001.
- [20] LIU Z, LI J G, SHEN Z Q, et al. Learning efficient convolutional networks through network slimming[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017: 2755–2763.
- [21] LI H , KADAV A , DURDANOVIC I , et al. Pruning Filters for Efficient ConvNets[J]. 2016. DOI: [10.48550/arXiv.1608.08710](https://doi.org/10.48550/arXiv.1608.08710).
- [22] GLENN J. YOLOv8 by Ultralytics[EB/OL]. <https://github.com/ultralytics/ultralytics>.
- [23] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2023: 12021–12031.
- [24] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 13708–13717.
- [25] Tong Z, Chen Y, Xu Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[J]. arXiv preprint arXiv: 2301.10051, 2023.
- [26] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design[C]// Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 122–138.
- [27] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 1314–1324.
- [28] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[J]. ArXiv e-Prints, 2019: arXiv: 1905.11946.
- [29] LI H, LI J, WEI H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv: 2206.02424, 2022.
- [30] Adarsh P , Rathi P , Kumar M . YOLO v3-Tiny: Object Detection and Recognition using one stage improved model[C]//2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2020. DOI: [10.1109/ICACCS48705.2020.9074315](https://doi.org/10.1109/ICACCS48705.2020.9074315).
- [31] Zhu X , Su W , Lu L , et al. Deformable DETR: Deformable Transformers for End-to-End Object Detection[C]//International Conference on Learning Representations. 2021. DOI: [10.48550/arXiv.2010.04159](https://doi.org/10.48550/arXiv.2010.04159).