

改进 YOLOv11 的复杂煤矿井环境目标检测模型

张建辉^{1,2}, 蔡小航¹, 王瑞民³, 曾俊杰¹, 罗旭东¹

(1. 郑州大学 网络空间安全学院, 河南 郑州 450002; 2. 嵩山实验室, 河南 郑州 450002; 3. 郑州大学 计算机与人工智能学院, 河南 郑州 450001)

摘要: 针对煤矿施工环境中光照分布不均、目标严重遮挡、粉尘干扰等工况导致的目标检测精度不足的问题, 提出一种基于 DIM 和 YOLOv11 的复杂煤矿井环境目标检测模型 DME-YOLO。在主干网络设计动态感知混合卷积模块(DIM), 通过动态权重机制实现多尺度特征的自适应融合, 提升复杂背景下的特征表征能力; 在检测头部分引入动态多注意力机制检测头(DMA-Head), 利用多尺度注意力模块增强对小目标和弱纹理目标的感知; 在颈部网络中嵌入高效上采样卷积模块(EUCB), 通过双线性插值与深度可分离卷积结合优化上采样路径。实验结果表明: DME-YOLO 在自建矿井数据集上的 $mAP@50$ 达到 93.7%, 较原始 YOLOv11 提升 3.0 个百分点; $mAP@50-95$ 达到 66.8%, 较原始 YOLOv11 提升 5.2 个百分点。与 YOLOv9s、YOLOv12 等模型对比, 所提模型收敛速度更快且检测精度更优, 适合煤矿井施工现场监测。

关键词: YOLOv11; 目标检测; 动态感知混合卷积; 多注意力机制; 高效上采样

中图分类号: TP391; TD76 文献标志码: A doi: 10.13705/j.issn.1671-6833.2026.04.009

随着“智能矿山”战略的深入推进, 煤矿井下作业正朝着自动化与智能化方向加速转型, 基于视觉感知的安全监测系统已成为智慧矿井建设的核心环节。作为计算机视觉的基础任务, 目标检测算法性能直接影响智慧矿井系统对井下目标识别、设备状态监控及安全隐患预警的可靠性^[1-2]。

深度学习目标检测算法分为两阶段(如 Fast R-CNN^[3]、Mask R-CNN^[4])与单阶段(如 YOLO 系列^[5-9]、SSD^[10]), 在通用场景下取得显著应用效果。

近年来, 研究者们针对目标检测算法在复杂施工场景中的应用开展了一系列改进工作。Zhang 等^[11]通过大分离卷积优化多尺度特征提取, 引入双步策略提升特征融合能力; Ramyadevi 等^[12]在 YOLOv8 算法模型中融入长短期记忆网络, 构建实时检测与行为分析系统; Zhang 等^[13]采用分组混洗卷积替代标准卷积提升检测精度与速度; Shao 等^[14]通过捕捉长距离依赖与强化通道注意力提升特征表达; Fu 等^[15]通过注意力尺度序列融合与新增检测层提升小尺度目标的检测能力。这些方法能够在通用场景下取得显著效果, 但在矿井某些特有的复杂

环境的下仍存在泛化能力不足或检测精度较低的问题, 原因在于现有方法的骨干网络多采用固定卷积核或简单分组卷积, 无法根据矿井环境的动态干扰调整特征提取策略, 且现有方法的颈部网络部分在特征上采样过程中易丢失小目标细节。这些问题使得模型在实际应用中面临检测精度不足、鲁棒性差的瓶颈。

针对现有技术的局限性, 本文对 YOLOv11 进行改进, 提出复杂煤矿井环境目标检测模型 DME-YOLO, 主要贡献包括: 构建包含 4 023 张井下作业图像的矿井数据集, 覆盖光照不均、强遮挡等多元场景; 设计动态感知混合卷积模块(dynamic inception mixer, DIM)嵌入骨干网络, 通过动态权重机制优化多尺度特征自适应融合, 提升复杂背景下的特征表征能力; 设计动态多注意力机制检测头(dynamic multi-attention detection head, DMA-Head), 借助多尺度注意力模块强化对小目标与弱纹理目标的感知能力; 在颈部网络中引入高效上采样卷积模块(efficient upsampling convolutional block, EUCB), 通过双线性插值与深度可分离卷积的结合优化上采样路

收稿日期: 2026-01-12; 修订日期: 2026-02-13

基金项目: 国家重点研发计划(2023YFB2906400); 河南省重大科技专项(221100210900)

作者简介: 张建辉(1977—), 男, 河南平顶山人, 郑州大学副研究员, 博士, 主要从事网络信息安全与人工智能技术研究,

E-mail: ndsczjh@163.com。

会与输入特征图进行残差加和,保证了网络中低层和高层特征的有效传递。这不仅使得模型能够充分利用多层次的信息,还能减少由于网络层次过深导致的梯度消失或特征丢失问题。

动态特征混合(dynamic feature mixer,DFM)模块通过组合多个具有不同核尺寸的动态深度卷积(dynamic depth-wise convolution,DDC)模块,实现多尺度特征提取。输入通道被拆分为多个组,每组由1个DDC模块处理,处理后的特征通过 1×1 卷积融合。这种设计能捕获不同尺度和方向的特征,增强模型对目标多样性的适应能力。

DIM的自适应性来源于其中的DDC机制。该机制能够为每个卷积核分配一个自适应的权重,使得网络能够根据输入特征图的不同需求调整卷积核的使用。这种机制打破了传统固定卷积核的限制,使得卷积操作能够在每次前向传播时根据输入的不同特征进行自适应调整,进而有效捕捉多尺度和多样化的特征信息^[16]。

DDC机制将输入特征图的空间维度压缩至 1×1 ,提取全局上下文信息,再生成3条卷积路径的权重,输出通道数为3,最后将权重归一化为概率分布,确保各路径的权重和为1。在前向传播中,输入特征图 \mathbf{x} 分别通过3条卷积路径处理,各路径输出由对应的权重加权后求和,生成聚合特征图。随后,该特征图经过批归一化BN和CGLU激活函数处理并输出最终结果:

$$\mathbf{y} = \text{CGLU}\left(\text{BN}\left(\sum_{i=0}^2 w_i \cdot \text{Conv}_i(\mathbf{x})\right)\right)。(1)$$

式中: w_i 为动态权重; $\text{Conv}_i(\mathbf{x})$ 为第*i*条卷积路径的输出;CGLU激活函数用于对输入特征进行拆分、门控筛选后输出,表达式为

$$\text{CGLU}(\mathbf{j}) = (\mathbf{j}_{:, \dots, d/2}) \times \sigma(\mathbf{j}_{:, \dots, d/2})。(2)$$

式中: $\mathbf{j}_{:, \dots, d/2}$ 表示取输入最后一维的前半部分; $\mathbf{j}_{:, \dots, d/2}$ 表示取最后一维的后半部分; σ 为Sigmoid函数。

1.4 DMA-Head 检测头

为提升煤矿井作业场景中目标检测模型对小目标、弱纹理目标以及复杂光照环境的鲁棒性,本文在YOLOv11检测框架的基础上,引入了多尺度注意力模块SEAM^[17],其结构如图3所示。利用该模块对检测头结构进行了模块化重构,分别在回归分支与分类分支中嵌入SEAM,构建出一种动态多注意力机制检测头DMA-Head,其结构见图1。

SEAM原用于解决人脸检测中因遮挡导致的部分数据丢失,模型难以提取足够特征进行精确检测

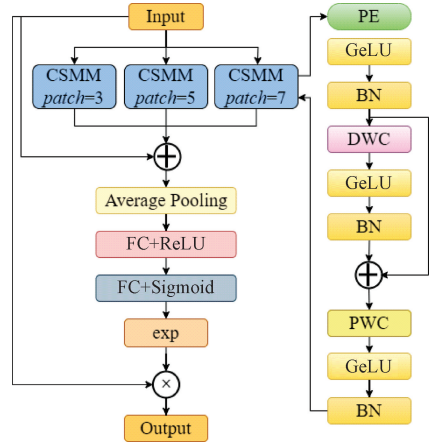


图3 SEAM模块

Figure 3 Module of SEAM

的问题。其核心在于聚焦被遮挡区域,增强被遮挡区域的特征表示。而在煤矿井场景中,存在设备对人员的遮挡、人员之间的密集遮挡等情况,这与人脸检测面临的遮挡问题类似,因而该模块在矿井复杂环境中具有良好的适配性。

SEAM通过3组不同patch大小的CSMM(cross-scale multi-module)分支并行建模多尺度上下文。CSMM以深度可分离卷积DWC+GeLU激活+批归一化BN为核心,借助残差结构提升特征通道表达稳定性。随后,3组并行输出利用全局平均池化将特征图压缩为通道维度的统计量,利用全连接网络学习通道间的依赖关系,生成通道注意力权重。随后通过指数激活函数扩大通道间的响应差异,强化关键通道的影响,提升对显著目标区域的聚焦能力。与传统的卷积结构相比,SEAM能够在保持计算效率的同时,增强模型对目标局部细节的感知能力,提升复杂场景下的检测精度。

1.5 EUCB 模块

为提升模型在复杂矿井环境下对特征的上采样效率与跨通道交互能力,本文引入高效上采样卷积模块EUCB^[18]。该模块针对矿井图像中多尺度目标特征融合需求进行改进,核心思想是以高效的方式完成特征图的精确上采样和跨通道融合。

与传统的上采样加卷积模块相比,EUCB构建了一套高效且有序的信息处理链路。首先对输入进行上采样,针对矿井环境下目标尺度变化多样、特征易受干扰的特点,采用了自适应权重的双线性插值上采样方法。其次是 3×3 深度可分离卷积,更有利于捕捉目标关键边缘和纹理的模式,同时在训练过程中采用了基于目标特征显著性的参数更新机制,使得卷积核能更高效地提取到目标的独特特征。最后特征图会依次进入批归一化、ReLU激活和 1×1

卷积,这一环节将根据矿井环境下特征的分布范围,动态调整批归一化的参数,再由逐点卷积实现不同通道信息的有效融合,并映射到所需的输出通道数,形成了一种高效的特征融合与优化策略。

算法 1 DME-YOLO 算法。

输入:自建矿井数据集、迭代次数、批次大小、初始学习率、预训练权重等;

输出:目标边界框、分类置信度、目标类别。

- ① 初始化 DME-YOLO 模型,加载预训练权重;
- ② 对输入数据集进行预处理,图像尺寸调整为 640×640 ,归一化为 $[0, 1]$,数据增强包括马赛克增强、水平翻转增强;
- ③ 读取预处理后的批量图像及对应标签;
- ④ 执行模型前向传播,动态混合卷积模块利用动态权重机制实现多尺度特征自适应融合,高效上采样卷积模块,保留多尺度特征完整性,动态多注意力机制检测头增强小目标、弱纹理目标感知;
- ⑤ 生成初步检测特征图与预测结果;
- ⑥ 计算多任务损失函数,优化模型预测精度,其中边界框回归损失 L_{reg} 提升目标边界框定位精度,减少位置偏差,分类损失 L_{cls} 优化目标类别的预测准确性,置信度损失 L_{conf} 提高检测结果的分类置信度,降低误检率;
- ⑦ 使用 SGD 优化器更新模型参数,采用阶梯式学习率衰减策略;
- ⑧ 重复步骤 2~7,直至达到迭代次数;
- ⑨ 在测试数据集上进行推理,采用非最大抑制算法,对存在冗余的检测框加以筛选,去除重复的预测输出,设定置信度阈值,仅保留超出该阈值的有效检测信息;
- ⑩ 输出最终检测结果。

2 实验结果与分析

2.1 数据集介绍

本实验数据集为自建数据集,通过网络爬虫技术收集煤矿井下作业照片并上传至数据标注平台,经由清洗、标注、划分而成,数据集示例如图 4 所示。考虑到井下作业特有的复杂工况,本文自建的数据集在场景覆盖维度涵盖了矿井的核心作业区域,如掘进工作面、运输巷道以及设备检修区域;在目标类别方面,包含了矿井作业的关键目标及其在不同角度、不同光照反射情况,不同动作姿态(弯腰、站立、行走等)的图像数据;对于干扰因素,在光照条件下,有矿灯直射形成的强光区域、巷道角落的低照度区域,在遮挡情况方面,包含设备对人员的遮挡、人

员之间的密集遮挡以及不同粉尘浓度下的图像。数据集包含 4 023 张图像,标注了 4 个类别:矿工(miner)、头盔(helmet)、牵引链(towline)、支撑板(support plate)。按 8:1:1 的比例随机划分为训练集、测试集和验证集,用于评估所提算法的性能表现。本文所有实验均基于该自制数据集进行,确保评估结果的统一性和可靠性。



(a) 不同光照环境 (b) 不同遮挡(粉尘) (c) 不同动作姿态

图 4 煤矿井下作业现场数据集示例

Figure 4 Example of dataset for coal mine underground operation site

2.2 实验平台

本文实验使用 Windows11 操作系统,其中 GPU 为 NVIDIA GeForce RTX 2060,显存 6 GB,采用 CUDA12.9 进行加速,CPU 型号为 AMD Ryzen7 4800,内存 16 GB。采用 PyTorch 深度学习框架,版本为 2.2.1,Python 版本为 3.10。实验设置 *batchsize* 为 4,训练 300 个 *epoch*。采用预训练模型 *yolo11n.pt*,马赛克增强概率为 1.0,水平翻转概率为 0.5,初始学习率设置为 0.01,最终学习率系数 0.01,使用 SGD 优化器,初始动量为 0.9,采用阶梯式学习率衰减,权重衰减系数设置为 0.0005。

2.3 评价指标

模型采用准确率 P 、召回率 R 、平均准确率 AP 、平均准确率均值 mAP 、帧率、计算量、参数量作为评价指标。准确率 P 是预测正确的前景目标占有所有预测为前景目标数量的百分比;召回率 R 是被正确检测出的前景目标占有所有前景目标的百分比; AP 是每个类别对应的 $P-R$ 曲线的面积, mAP 是所有类别的 $P-R$ 曲线下面积的平均值;帧率是模型每秒钟处理图片的数量;计算量、参数量分别用于评估模型计算复杂度、轻量化水平。

2.4 模型检测效果对比

分别使用 YOLOv11(base)和 DME-YOLO 模型对自建的煤矿井下作业现场数据集进行实验,实验结果如表 1 所示。可知 DME-YOLO 整体 $mAP@50$ 提升了 3.0 百分点,其中 *miner*、*helmet*、*towline*、*support plate* 的 AP 值分别提高了 2.7、6.7、2.4、0.1 百

分点,说明模型对复杂矿井环境下多类别目标的检测准确率提升明显。尤其在 miner 类和 helmet 类的检测中效果较为显著,这两类目标较小,容易受到光线和遮挡的影响,验证了本文模型对弱纹理小目标的检测能力强,同时对 topline 类和 support plate 类的检测率亦有不同幅度的提升,说明本文模型在提升小目标检测准确率的同时保持了对大目标的高检测准确率。

表 1 模型检测效果对比

Table 1 Comparison of model detection performance

| 模型 | AP | | | | mAP |
|----------|-------|--------|---------|---------------|-------|
| | miner | helmet | topline | support plate | @ 50 |
| base | 0.908 | 0.820 | 0.925 | 0.976 | 0.907 |
| DME-YOLO | 0.935 | 0.887 | 0.949 | 0.977 | 0.937 |

2.5 各模块有效性分析

2.5.1 DIM 有效性分析

为验证本文提出的动态感知混合卷积模块 DIM 在目标检测任务中的实际效果,本文设计了多组对比实验,分别将 DIM 模块嵌入 YOLOv11 的不同网络位置(Backbone、Neck 以及两者结合),以评估其在不同结构层级上的贡献。实验结果如表 2 所示。

表 2 DIM 有效性实验

Table 2 DIM validity experiment

| 模型 | P | R | mAP | mAP@ |
|---------------------|-------------|-------------|-------------|-------------|
| | | | @ 50 | 50-95 |
| base | 91.8 | 85.4 | 90.7 | 61.6 |
| +DIM(Backbone) | 92.0 | 87.5 | 92.1 | 63.5 |
| +DIM(Neck) | 91.6 | 88.1 | 90.5 | 62.2 |
| +DIM(Neck、Backbone) | 91.9 | 86.9 | 91.4 | 63.0 |

由表 2 可知,未引入 DIM 模块的 YOLO11 (base)的 $mAP@50$ 和 $mAP@50-95$ 分别为 90.7%和 61.6%。当 DIM 模块集嵌入 Backbone 时, $mAP@50$ 提升至 92.1%, $mAP@50-95$ 提升至 63.5%,准确率与召回率也均有所上升,表明 DIM 能有效提升模型的整体检测性能,尤其在特征提取阶段具备明显优势。将 DIM 模块应用于 Neck 时, $mAP@50-95$ 虽有提升至 62.2%,但 $mAP@50$ 下降至 90.5%,表明在该结构层级中 DIM 对部分尺度特征的表达仍存在一定的局限性。进一步将 DIM 同时应用于 Backbone 和 Neck 层, $mAP@50$ 和 $mAP@50-95$ 分别为 91.4%和 63.0%,性能优于 base 模型但略逊于仅在 Backbone 引入 DIM 的情况。

综合分析可知,DIM 模块在增强多尺度特征表示能力、提升目标检测准确率方面效果显著,尤其在

主干网络中表现最佳。

2.5.2 EUCB 有效性分析

为验证所提出的 EUCB 在目标检测性能上的提升作用,本文设计了不同上采样策略的对比实验,评估各策略对检测准确率的影响。对比对象包括常见的上采样方法,例如 Upsample、CARAFE^[19]、Dysample^[20]、SAPA^[21]、FADE^[22],实验结果如表 3 所示。

表 3 不同上采样策略对比

Table 3 Comparison of different upsampling strategies

| 策略 | mAP@ 50 | mAP@ 50-95 |
|--------------------------|-------------|-------------|
| Upsample | 90.7 | 61.6 |
| CARAFE ^[19] | 90.8 | 61.7 |
| Dysample ^[20] | 91.0 | 62.4 |
| SAPA ^[21] | 91.1 | 63.0 |
| FADE ^[22] | 91.2 | 63.1 |
| EUCB | 91.3 | 63.6 |

由表 3 可知,Upsample 的 $mAP@50$ 为 90.7%、 $mAP@50-95$ 为 61.6%,难以满足复杂场景中多尺度目标的精细还原需求。CARAFE 和 Dysample 引入卷积融合操作后,分别获得略微提升, $mAP@50$ 分别提高至 90.8%和 91.0%。SAPA 和 FADE 在空间特征建模方面更进一步,检测准确率分别提升至 91.1 和 91.2%。相比之下,EUCB 上采样卷积模块结合双线性插值、深度可分离卷积、通道洗牌与逐点卷积等机制, $mAP@50$ 为 91.3%、 $mAP@50-95$ 为 63.6%,优于其他主流上采样方法。

2.5.3 DMA-Head 有效性分析

本文设计了多组对比实验,涵盖当前主流检测头改进策略,包括 EfficientHead^[23]、DyHead^[24]、LADH^[25]、LSCD-Head^[26]等,并将结果与 DMA-Head 进行系统比较,实验结果如表 4 所示。

表 4 DMA-Head 有效性实验

Table 4 DMA-Head validity experiment

| 策略 | P | R | mAP@ 50 |
|--------------------------------|-------------|-------------|-------------|
| base | 91.8 | 85.4 | 90.7 |
| +EfficientHead ^[23] | 91.7 | 86.0 | 90.9 |
| +DyHead ^[24] | 91.9 | 88.4 | 90.9 |
| +LADH ^[25] | 91.8 | 86.7 | 91.2 |
| +LSCD-Head ^[26] | 91.8 | 87.3 | 91.4 |
| +DMA-Head | 92.0 | 88.1 | 91.8 |

由表 4 可知,YOLO11(base)的准确率、召回率和 $mAP@50$ 分别为 91.8%、85.4%和 90.7%。引入 EfficientHead、DyHead、LADH 和 LSCD-Head 后, $mAP@50$ 分别提升至 90.9%、90.9%、91.2%和 91.4%,表现出一定的性能增益。相比之下,DMA-Head 在准确率

(92.0%)、召回率(88.1%)以及 $mAP@50$ (91.8%) 这3项指标上实现综合最优性能。

2.6 消融实验

为验证提出的改进部分(DIM、DMA-Head和EUCB)对YOLOv11模型性能的提升作用,进行了消融实验,实验结果如表5所示,可知逐步去除各模块会降低模型的性能。引入DIM模块后,模型的 $mAP@50$ 从90.7%提升至92.1%,表明DIM在特征提取和多尺度特征融合方面具有显著效果。DMA-Head模块的引入使得 $mAP@50$ 提升至91.8%,并在小目标和弱纹理目标的检测能力上表现突出,尤

其在复杂光照和遮挡环境下尤为明显。EUCB模块提高了模型的上采样效率,将原模型的 $mAP@50$ 提升至91.3%,有效增强了多尺度目标的检测能力。综合来看,结合DIM、DMA-Head和EUCB三者后,模型的准确率、召回率和 $mAP@50$ 等指标均显著提升。此外,当引入3个模块后,计算量提升至6.3GFLOPs,较于原始模型增幅8.6%,且模型检测帧率降低至119.1帧·s⁻¹,这表明3个模块的引入虽带来一定计算量增加,但整体增幅可控,模型推理速度仅受到小幅负面影响。实验验证了各模块在复杂矿井环境下的有效性与互补性。

表5 消融实验

Table 5 Ablation experiments

| DIM | DMA-Head | EUCB | P/% | R/% | $mAP@50$ /% | $mAP@50-95$ /% | 参数量/10 ⁶ | 计算量/GFLOPs | 帧率/(帧·s ⁻¹) |
|-----|----------|------|-------------|-------------|-------------|----------------|---------------------|------------|-------------------------|
| | | | 91.8 | 85.4 | 90.7 | 61.6 | 2.5 | 5.8 | 126.2 |
| √ | | | 92.0 | 87.5 | 92.1 | 63.5 | 2.6 | 6.0 | 122.3 |
| | √ | | 92.0 | 88.1 | 91.8 | 63.7 | 2.5 | 5.9 | 124.7 |
| | | √ | 91.8 | 86.6 | 91.3 | 63.6 | 2.5 | 5.9 | 125.0 |
| √ | √ | | 92.1 | 87.7 | 92.4 | 66.1 | 2.6 | 6.2 | 120.4 |
| √ | | √ | 93.0 | 89.2 | 93.1 | 66.5 | 2.6 | 6.2 | 119.3 |
| | √ | √ | 91.9 | 88.4 | 93.3 | 65.9 | 2.7 | 6.1 | 121.5 |
| √ | √ | √ | 92.9 | 90.0 | 93.7 | 66.8 | 2.6 | 6.3 | 119.1 |

2.7 DME-YOLO 与其他模型对比

将DME-YOLO与主流目标检测模型针对 $mAP@50$ 开展对比实验,结果如图5所示。训练至300轮时,YOLOv3、YOLOv5s、YOLOv9s、YOLOv12、YOLOv7、YOLOv8s的 $mAP@50$ 分别为89.9%、91.3%、91.7%、91.9%、92.1%、92.1%,而DME-YOLO收敛速度更快,100轮训练后 $mAP@50$ 稳定在90%以上,最终达93.7%,且训练后期呈稳定上升趋势。

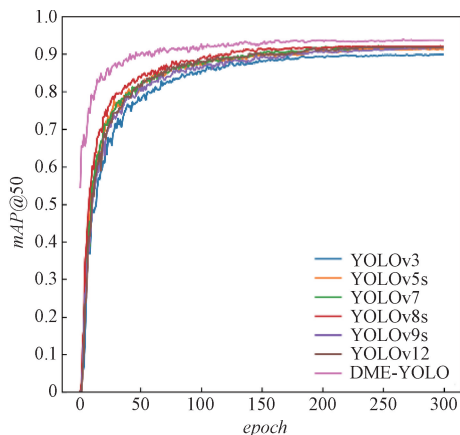
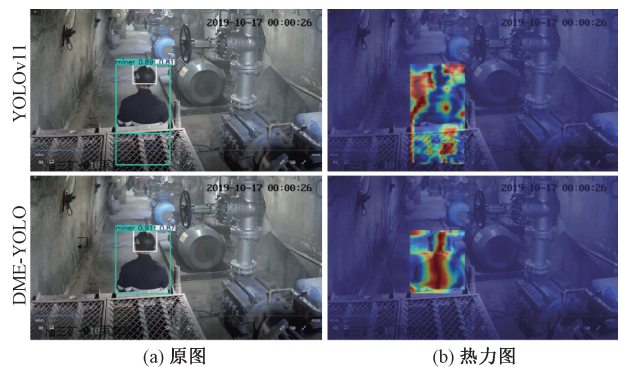


图5 不同网络模型 $mAP@50$ 对比图

Figure 5 Comparison chart of $mAP@50$ of different network models

2.8 检测效果分析

改进前后的检测结果与热力图对比如图6所示。图6(a)清晰地展现了模型性能的提升,原始YOLOv11对目标矿工(miner)以及头盔(helmet)的检测边界略显松散,置信度分为0.89、0.81,而DME-YOLO定位更精准,置信度分别提升至0.91、0.87,表明改进模型在特征提取与分类判断上更为稳定和准确。图6(b)热力图中,改进前的模型在对矿工区域进行检测时,热力图显示其对目标区域的关注不够集中,背景干扰区域也存在一定程度的响



(a) 原图

(b) 热力图

图6 改进前后检测结果与热力图对比

Figure 6 Comparison of detection results and heatmap before and after improvement

应。相比之下, DME-YOLO 对矿工区域的关注明显更为集中, 热力图中响应区域显著增强, 表明模型能够更敏锐地捕捉到目标所在位置及特征。

3 结论

本文提出一种基于 DIM 和 YOLOv11 的复杂煤矿井环境目标检测模型 DME-YOLO, 针对矿井复杂环境下目标检测问题进行了优化, 构建了包含矿工、头盔、牵引链、支撑板 4 类目标的自建数据集。引入动态感知混合卷积模块、动态多注意力机制检测头, 高效上采样卷积模块以提升模型在光照不均、强遮挡等极端场景下的检测性能。实验结果表明, 改进算法在自建数据集上的 $mAP@50$ 达到 93.7%, $mAP@50-95$ 达到 66.8%, 实验验证了模型在有效提升大目标检测准确率的同时对小目标和弱纹理目标的检测能力显著增强。对比实验显示, DME-YOLO 检测模型收敛速度更快, 且最终检测准确率更高, 且对矿井环境中的设备遮挡、光照不均等问题表现出更强的适应性。本文模型为煤矿井内复杂环境下目标检测提供了可靠的依据。

参考文献:

- [1] Yuan Zhi, Jiang Qingyou, Pang Zhenzhong. Application status and development thinking of intelligent mining technology and equipment in coal mines in China[J]. Coal Science and Technology, 2024, 52(9): 189-198. [袁智, 蒋庆友, 庞振忠. 我国煤矿智能化综采开采技术装备应用现状与发展思考[J]. 煤炭科学技术, 2024, 52(9): 189-198.]
- [2] Jung D, Choi Y. Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation[J]. Minerals, 2021, 11(2): 148.
- [3] Girshick R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: 1440-1448.
- [4] He Kaiming, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.
- [5] Redmon J, Farhadi A. YOLOv3: an incremental improvement[PP/OL]. V1. arXiv(2018-04-08)[2025-09-01]. <https://arxiv.org/abs/1804.02767>.
- [6] Ultralytics. YOLOv5 [EB/OL]. (2020-05-18)[2025-09-01]. <https://github.com/ultralytics/yolov5>.
- [7] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 7464-7475.
- [8] Ultralytics. Ultralytics [EB/OL]. (2023-01-10)[2025-09-01]. <https://github.com/ultralytics/ultralytics>.
- [9] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[PP/OL]. V2. arXiv(2024-02-29)[2025-09-01]. <https://arxiv.org/abs/2402.13616>.
- [10] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//Computer Vision - ECCV 2016. Cham: Springer, 2016: 21-37.
- [11] Zhang Zhihao, Tao Lei, Yao Linhu, et al. LDSI-YOLOv8: Real-time detection method for multiple targets in coal mine excavation scenes[J]. IEEE Access, 2024, 12: 132592-132604.
- [12] Ramyadevi R, Keerthy S, Catherina J S J, et al. Helmet and equipment detection with worker's mobility tracker in mining sector using YOLOv8 & LSTM[C]//Proceedings of the 2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Piscataway: IEEE, 2025: 1-6.
- [13] Zhang Lei, Sun Zhipeng, Tao Hongjing, et al. Research on mine-personnel helmet detection based on multi-strategy-improved YOLOv11[J]. Sensors, 2025, 25(1): 170.
- [14] Shao Xiaoqiang, Liu Shibo, Li Xin, et al. Rep-YOLO: an efficient detection method for mine personnel[J]. Journal of Real-Time Image Processing, 2024, 21(2): 28.
- [15] Fu Zhibo, Ling Jierui, Yuan Xinpeng, et al. Yolov8n-FADS: a study for enhancing miners' helmet detection accuracy in complex underground environments[J]. Sensors, 2024, 24(12): 3767.
- [16] Yu Weihao, Si Chenyang, Zhou Pan, et al. MetaFormer baselines for vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(2): 896-912.
- [17] Yu Ziping, Huang Hongbo, Chen Weijun, et al. YOLO-FaceV2: a scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.
- [18] Rahman M M, Munir M, Marculescu R. EMCAD: efficient multi-scale convolutional attention decoding for medical image segmentation[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 11769-11779.
- [19] Wang Jiaqi, Chen Kai, Xu Rui, et al. CARAFE: content-aware reassembly of FEatures[C]//Proceedings of the 2019 IEEE/CVF International Conference on Comput-

- er Vision (ICCV). Piscataway: IEEE, 2019: 3007–3016.
- [20] Liu Wenze, Lu Hao, Fu Hongtao, et al. Learning to up-sample by learning to sample [C] // Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2024: 6004–6014.
- [21] Lu Hao, Liu Wenze, Ye Zixuan, et al. SAPA: similarity-aware point affiliation for feature upsampling [PP/OL]. V2. arXiv(2022-12-27) [2025-10-20]. <https://arxiv.org/abs/2209.12866>.
- [22] Lu Hao, Liu Wenze, Fu Hongtao, et al. FADE: fusing the assets of decoder and encoder for task-agnostic upsampling [PP/OL]. V2. arXiv(2022-12-27) [2025-09-01]. <https://arxiv.org/abs/2207.10392>.
- [23] Xie Weining, Ma Weifeng, Sun Xiaoyong. An efficient re-parameterization feature pyramid network on YOLOv8 to the detection of steel surface defect [J]. Neurocomputing, 2025, 614: 128775.
- [24] Dai Xiyang, Chen Yinpeng, Xiao Bin, et al. Dynamic head: unifying object detection heads with attentions [C] // Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 7369–7378.
- [25] Zhang Jiarui, Chen Zhifeng, Yan Guoxu, et al. Faster and lightweight: an improved YOLOv5 object detector for remote sensing images [J]. Remote Sensing, 2023, 15 (20): 4974.
- [26] Gao Lin, Yu Pengwei, Dong Hongjuan, et al. Multi-scale fusion lightweight target detection method for coal and gangue based on EMBS-YOLOv8s [J]. Sensors, 2025, 25(6): 1734.

Improved YOLOv11 Target Detection Model for Complex Coal Mine Environments

ZHANG Jianhui^{1,2}, CAI Xiaohang¹, WANG Ruimin³, ZENG Junjie¹, LUO Xudong¹

(1. School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China; 2. Songshan Laboratory, Zhengzhou 450002, China; 3. School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China)

Abstract: To address the insufficient target detection accuracy caused by harsh working conditions in coal mine construction environments, such as uneven illumination distribution, severe target occlusion, and dust interference, a target detection model named DME-YOLO was proposed for coal mine in complex environments based on DIM and YOLOv11. In the backbone network of DME-YOLO, a dynamic inception mixer convolution module (DIM) was designed. This module achieved adaptive fusion of multi-scale features through a dynamic weight mechanism, thereby enhancing the model's capability of feature representation in complex backgrounds. For the detection head, a dynamic multi-attention detection head (DMA-Head) was introduced, which leveraged a multi-scale attention module to strengthen the perception of small targets and targets with weak textures. Additionally, an efficient up-sampling convolutional block (EUCB) was embedded into the neck network optimizing the upsampling path by combining bilinear interpolation with depthwise separable convolution. Experimental results demonstrated that DME-YOLO achieved a $mAP@50$ of 93.7% on the self-constructed mine dataset, representing 3.0 percentage points improvement compared to the original YOLOv11. Its $mAP@50-95$ reached 66.8%, which was 5.2 percentage points increase relative to the original YOLOv11. When compared with models such as YOLOv9s and YOLOv12, DME-YOLO exhibited faster convergence speed and superior detection accuracy, making it well-suited for safety monitoring in coal mine construction sites.

Keywords: YOLOv11; target detection; dynamic inception mixer convolution; multi-attention mechanism; efficient upsampling