

文章编号:1671-6833(2025)04-0040-07

基于改进 YOLOv8 的遗留物品检测算法

张震¹, 葛帅兵², 陈可鑫³, 李友好⁴, 黄伟涛⁴

(1. 郑州大学 电气与信息工程学院, 河南 郑州 450001; 2. 郑州大学 河南先进技术研究院, 河南 郑州 450003; 3. 珠海优特电力科技股份有限公司, 广东 珠海 519000; 4. 河南汇融油气技术有限公司, 河南 郑州 450001)

摘要: 针对传统基于背景减法的遗留物品检测算法难以应对人流拥挤、小目标、物品遮挡和光线变化等环境, 以及基于深度学习方法中的模型准确率低等问题, 提出了一种基于改进 YOLOv8 的遗留物品检测算法。首先, 使用动态上采样 DySample 替换最近邻上采样, 优化上采样过程, 增强模型的泛化能力。其次, 将高效轻量的 ADown 下采样模块替代普通的下采样卷积, 在降低整个模型参数数量的同时, 提升算法的检测精度。最后, 引入 EMA 注意力机制, 优化特征提取过程, 增强特征提取能力, 提升对小目标检测的效果。实验结果表明: 改进后的模型 YOLO-DAE 在自建数据集上取得的准确率 P 、召回率 R 、 $mAP@50$ 和 $mAP@50:95$ 分别为 93.4%, 87.7%, 91.7% 和 80.2%, 相比于改进前的 YOLOv8s 模型在模型参数数量和计算量减少的同时, 分别提高了 1.8 个百分点、1.6 个百分点、1.2 个百分点和 2.1 个百分点, 并且 $mAP@50$ 和 $mAP@50:95$ 均高于 YOLOv5s r6.0、YOLOv6s v3.0、YOLOv7s AF 和 YOLOv9s, 有效提升了遗留物品检测能力。

关键词: 遗留物品检测; YOLOv8 算法; EMA 注意力机制; DySample 模块; ADown 模块

中图分类号: TP391 **文献标志码:** A **doi:** 10.13705/j.issn.1671-6833.2025.04.010

遗留物品检测作为智能视频监控系统的核心功能模块, 在公共安全、生产安全和安防等领域中扮演着重要角色。燃气生产行业(如天然气分输场站)对工人的作业行为和现场物品管理有着严格要求, 工人遗留的工具(如防爆扳手、螺丝刀等)可能被其他人员误触或人员被绊倒, 导致人员受伤或设备损坏, 而且疏散通道遗留的小型运输车辆在发生事故时严重影响人员撤离。因此, 针对遗留物品检测的研究具有重大意义。

目标检测是计算机视觉的基本任务, 当前主流的目标检测算法是基于深度学习的目标检测算法, 分为单阶段目标检测算法和双阶段目标检测算法。双阶段目标检测算法首先是 Girshick 等^[1]提出的 R-CNN, 该算法先根据图像提取候选框, 然后基于候选区域做二次修正得到检测结果, 检测精度较高, 但检测速度较慢, 难以达到实时处理效果。单阶段目标检测算法直接生成检测结果, 虽然检测速度快, 但检测精度低, 这类算法有 YOLO^[2]、SSD^[3]、FCOS^[4]等。

国内外学者对遗留物品检测技术探索出了不同

的研究方案。传统遗留物品检测算法多使用背景减法, Lin 等^[5]结合短期和长期背景模型来提取前景对象, 将输入图像中的每个像素分为不同的状态, 随后引入一个基于状态转换的静态前景区域识别框架, 并通过分析遗留物品拥有者的回溯轨迹来确定候选区域是否包含遗留物品。Su 等^[6]利用时域中值滤波和时域最小值滤波分别提取前景目标和静态目标, 同时提出了一种基于梯度的图像处理算法来消除夜间车灯的干扰, 该方法可以在短时间内检测出遗留物品。Russel 等^[7]使用基于深度特征构建的先验模型检索前景目标, 同时利用核典型相关分析和余弦相似度指数的深度特征来跟踪行人, 为被跟踪者生成步态能量图以判断其所携带行李的状态。传统遗留物品检测算法难以应对人流拥挤、小目标、物品遮挡和光线变化等环境。近年来也有不少学者使用深度学习的方法。Park 等^[8]结合双背景模型和 mask R-CNN, 考虑当前和过去的掩码信息, 检查当前视频帧或前一背景帧中是否存在相应的分割对象, 但这样无法达到对视频帧的实时处理。Liu

收稿日期: 2024-09-20; 修订日期: 2024-11-30

基金项目: 河南省重点研发专项(231111211600); 河南省重大公益专项(201300311200)

作者简介: 张震(1966—), 男, 河南郑州人, 郑州大学教授, 博士, 博士生导师, 主要从事计算机视觉研究, E-mail: zhang-zhen66@126.com。

引用本文: 张震, 葛帅兵, 陈可鑫, 等. 基于改进 YOLOv8 的遗留物品检测算法[J]. 郑州大学学报(工学版), 2025, 46(4): 40-46. (ZHANG Z, GE S B, CHEN K X, et al. Abandoned object detection algorithm based on improved YOLOv8 [J]. Journal of Zhengzhou University (Engineering Science), 2025, 46(4): 40-46.)

等^[9]使用 YOLO 检测器来检测监控视频中行人通过场景前后的关键帧,并对关键帧进行了比较和分析,以获得遗留物品的位置。蒋晓可^[10]使用 YOLOv5 和双背景模型检测前景目标,然后使用 DeepSort 算法跟踪前景目标,根据目标运动特点将物品与所属者进行配对,物品失配时将被识别为遗留物品。以上文献使用的 YOLO 系列算法的精度是有提升空间的。林德钰等^[11]首先利用高斯混合模型背景建模进行前景检测,根据移动区域与静止区域的分离距离与时间得到可疑静止区域,然后将判定为分离时刻的帧图像传入轻量化的 YOLO-G 模型进行检测与识别,但该模型的参数量较高。

综上所述,为了实现更高的检测精度和相对较低的模型参数量,本文提出了一种改进的 YOLOv8 算法用于实时检测监控视频中的遗留物品。首先,使用动态上采样 DySample 替换原来的最近邻上采样,优化了上采样过程,增加了模型的泛化能力。其次,引入高效轻量的 ADown 模块缓解模型过拟合,不仅降低了模型参数量和计算量,还提升了算法的准确率。此外,引入一种高效的多尺度注意力机制 EMA,对目标检测和特征提取过程进行优化,从而更好地提取行为特征,提升对小目标检测的效果。

1 YOLOv8 模型

YOLOv8^[12]是在 YOLOv5^[13]的基础上改进而来的,二者都是 Ultralytics 开源的实时目标检测算法。图 1 展示了 YOLOv8 模型结构及其模块,该算法数据预处理采用了马赛克增强、混合增强和空间扰动等。与 YOLOv5 不同的是,YOLOv8 在整个网络中使用梯度流信息更丰富的 C2f 模块替代 C3,Head 部分替换成了解耦头结构,将分类和检测头分离,同时也从 anchor-based 换成了 anchor-free。损失计算方面采用了 task aligned assigner^[14]正样本分配策略,根据分类与回归加权的分数选择正样本,其中分类损失使用了二值交叉熵(binary cross-entropy, BCE)损失,回归损失中引入了 distribution focal loss^[15],让网络快速聚焦到标签附近的数值,使标签处的概率密度尽量大。训练阶段的数据增强部分在最后 10 个 epoch 关闭马赛克数据增强的操作,可以有效地提升精度。Neck 部分依然是采用 PAN+FPN 结构,充分将深层的语义特征和浅层特征融合,增强多个尺度上的语义信息。Backbone 部分依旧使用了快速空间金字塔池化(spatial pyramid pooling fast, SPPF)模块, SPPF 使用 3 个 5×5 的最大池化代替 SPP 的 5×5, 9×9, 13×13 最大池化操作,使用较小的

卷积核能够提升计算速度。YOLOv8 算法适合在工业场景中应用,容易部署,在保持较小参数量的同时能达到较高的精度。YOLOv8 官方提供了 n、s、m、l、x 尺度的不同大小模型,本文采用 YOLOv8s 作为基础模型。

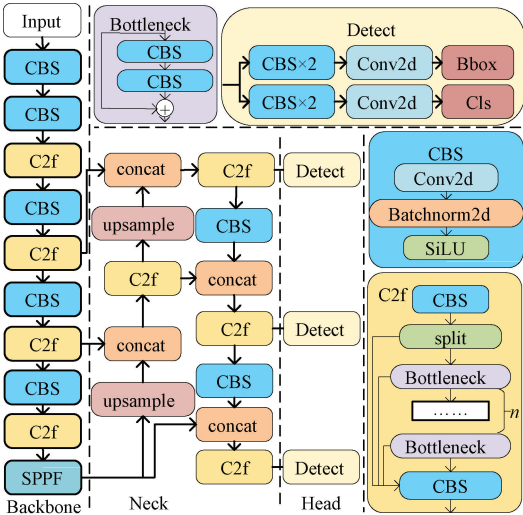


图 1 YOLOv8 模型结构

Figure 1 Structure of YOLOv8 model

2 模型改进

2.1 DySample 模块

DySample 是 Liu 等^[16]提出的一种高效、轻量的动态上采样算子。其过程如图 2 所示,首先从输入的特征图 X 中学习得到上采样点集,然后利用该点集对原特征图进行网格采样得到输出特征图 X' 。

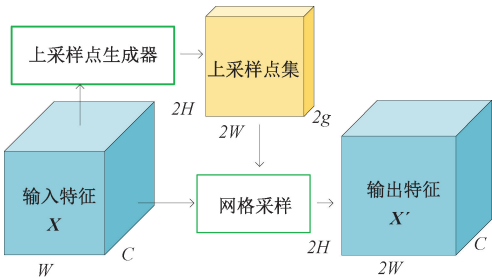


图 2 动态上采样流程图

Figure 2 Dynamic upsampling flowchart

上采样点就是上采样后的特征图中每个点在原特征图中的采样坐标。图 2 中 g 表示原始特征图的组数,同组特征共享相同的采样点集,图 3 展示了该坐标的生成过程。采样点集由原始网格和偏移量相加而成,偏移量有两种生成方式:静态调节方式首先从原始特征图中学习得到调节前偏移量,然后乘以静态调节因子 0.25 调整其大小,通过 Pixel Shuffle 操作上升到原始网格的维度;动态调节方式的差别在于调节因子不同,其调节因子是从原始特征图中学习,并经过 Sigmoid 函数然后乘以 0.5 得到的。

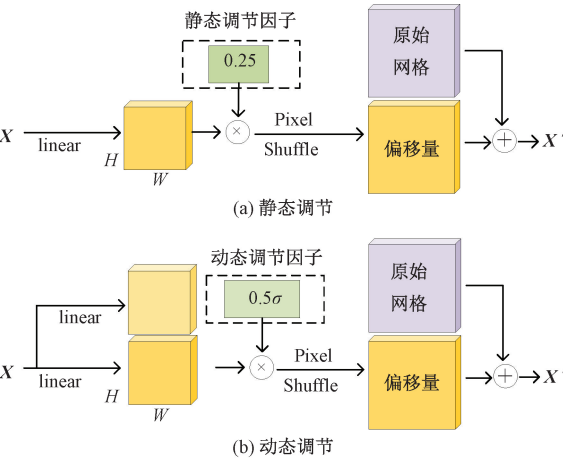


图 3 采样点生成器

Figure 3 Sampling point generator

YOLOv8 中使用最近邻上采样,该方法虽然简单快速,但只是将原始像素复制,没有考虑周围图像的信息,可能破坏图像中的线性关系。本文将具有动态调节因子的 DySample 模块替换原始网络中的最近邻上采样,设置 $g=4$ 。

2.2 ADown 模块

ADown 是 YOLOv9^[17] 算法中使用的一种取代普通卷积的轻量化、低参数的下采样模块,该模块优化了下采样过程,防止模型过拟合,如图 4 所示。首先将输入的特征图进行平均池化操作;然后拆分成两个部分,第一部分采用步长为 2 的普通卷积,第二部分先使用步长为 2 的最大池化,然后使用一个普通卷积;最后将得到的结果拼接到一起。

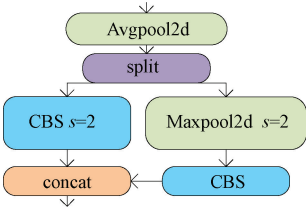


图 4 ADown 模块结构

Figure 4 Structure of ADown module

YOLOv8 中使用步长为 2 的普通卷积进行下采样,该操作计算代价高且参数多。本文将 ADown 模块替换原始模型中 Backbone 和 Neck 的部分步长为 2 的 Conv 模块。

2.3 EMA 注意力机制

EMA 是 Ouyang 等^[18]提出的一种高效多尺度注意力机制,能优化特征提取过程,增强模型性能,同时在保留每个通道信息的同时也减少了计算开销。图 5 所示为 EMA 注意力机制的结构。

(1)特征分组。将维度为 $C \times H \times W$ 的特征图 X 分成 G 组,拆分后子特征图维度为 $(C/G) \times H \times W$ 。

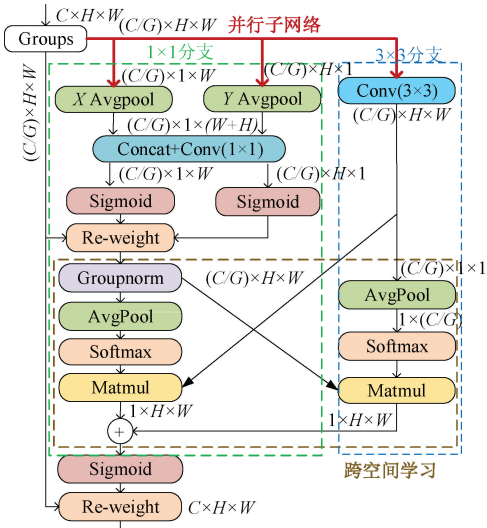


图 5 EMA 注意力机制

Figure 5 EMA attention mechanisms

(2)并行子网络。EMA 利用 3 个子网络并行提取注意力权重,其中 1×1 的分支分别沿着两个空间方向进行全局平均池化,结果拼接后通过跨通道信息交互进行建模,随后将其输出拆分为两个向量,使用 Sigmoid 函数将其提取为注意力权重并作用于原始特征图上。 3×3 分支中的卷积捕捉局部跨通道信息,扩大特征空间。并行子网络捕捉不同尺度的特征信息。

(3)跨空间学习。在 1×1 分支中利用全局平均池化和 Softmax 函数提取出注意力权重后与 3×3 分支中的特征图融合, 3×3 分支使用相同操作提取出注意力权重并和 1×1 分支中经过组归一化的特征图结合,将融合后的特征相加后就得到了不同空间的多尺度信息。

3 遗留物品检测

改进后的 YOLOv8s 模型 YOLO-DAE 如图 6 所示。

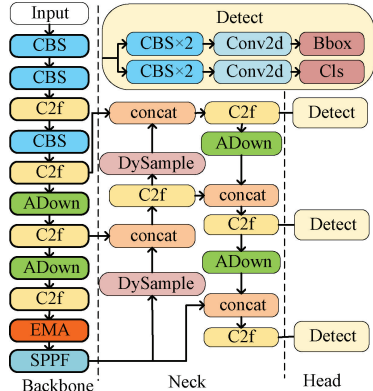


图 6 改进后的 YOLOv8 模型结构

Figure 6 Structure of improved YOLOv8 model

使用该模型检测视频流中的前景目标,利用多目标跟踪算法 ByteTrack^[19],为检测的每个目标分

配 ID,同时将每个跟踪到的目标的信息记录到跟踪队列中,根据分配的 ID 信息,判断当前帧某一目标与前一帧该目标的检测框的交并比 (IoU) 是否大于阈值 S ,其中 $S \in [0,1]$,若大于 S ,则说明该目标静止一帧,连续静止 N 帧的目标即可判定为遗留物品。本文的重点是提升遗留物品的检测能力,故不详细阐述遗留物品的判定规则。

4 实验结果与分析

4.1 数据集采集及预处理

目前开源的遗留物品检测数据集多为商场、车

站站台和机场,并且遗留物品为背包、行李箱等,缺乏工业场景的数据集。因此,根据实际的业务需求,在天然气分输场站现场取景,同时也从互联网视频中收集了一些数据,构建的数据集共有 4 828 张图片,如图 7 所示,包含工人、防爆工具箱、几种防爆工具、小型运输叉车和丢弃的垃圾包裹等 10 个类别,数据中包含室内、室外、不同光照的背景,同时含有一定数量的人流拥挤、小尺度目标的图片,以数据驱动的方式增强模型对复杂背景、光照变化、密集目标等情况的干扰,使用 Labellmg 对数据打标签,将数据集划分为 8:2 的训练集与验证集。



图 7 10 个遗留物品类别
Figure 7 10 categories of abandoned object

4.2 评价指标

模型性能评价指标采用准确率 P 、召回率 R 、平均准确率 mAP 。准确率 P 为预测正确的前景目标占有预测为前景目标数量的百分比,反映是否能准确检测出前景目标;召回率 R 表示被正确检测出的前景目标占有所有前景目标的百分比,反映前景目标的漏检程度; mAP 表示所有类别的 P - R 曲线下面积 AP 的平均值。

4.3 实验平台

本次实验的设备使用 Windows11 操作系统,其中 GPU 为 GRID T4-8Q,显存 8 GB,CPU 型号为 QE-MU Virtual CPU,内存 32 GB。采用 Pytorch 深度学习框架,其版本为 1.10.1,Python 版本为 3.9,将 $batch-size$ 设置为 8,训练 150 个 $epoch$,初始学习率设置为 0.01,使用 AdamW 优化器,初始动量为 0.9,权重衰减系数设置为 0.000 5。

4.4 DySample 有效性分析

为探究 g 取何值时 mAP 为最优以及比较动态、静态调节因子的优劣,本文在不同的组数和不同的调节因子下进行实验,实验结果如图 8 所示。可以看出,当 $g=4$,使用动态因子时模型的准确度增益更高。

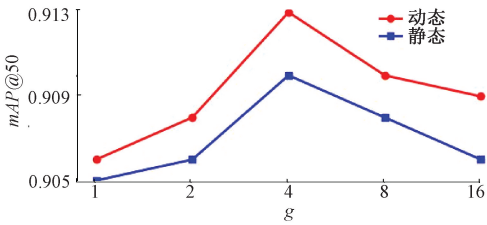


图 8 不同组数与调节因子下的 $mAP@50$
Figure 8 $mAP@50$ with different groups and regulatory factors

为进一步验证 DySample 上采样的优越性,本文采取用不同的上采样策略,包括双线性插值、Pixel Shuffle^[20]、CARAFE^[21]、FADE^[22]和 SAPA^[23],实验结果如表 1 所示。可以看出,基础 YOLOv8 模型使用最近邻插值算法,在替换成 DySample 之后达到了最高的准确率。

4.5 不同注意力机制对比

为探究 EMA 注意力中组数 G 对准确率的影响,本文做了不同组数下的对比实验。当 G 为 1,2,4,8,16 和 32 时, $mAP@50:95$ 分别为 0.904,0.910,0.913,0.914,0.915 和 0.913,可以得出, $G=16$ 时模型达到最高的准确率。

为进一步验证 EMA 注意力机制的优势,本文使

用不同注意力进行对比实验,包括通道注意力机制 SE^[24] 和 ECA^[25];空间和通道结合的注意力机制 CBAM^[26]、CPCA^[27]。实验结果如表 2 所示。可以看出,在使用 EAM 之后达到了最高的准确率。

表 1 不同上采样策略对比

Table 1 Comparison of different upsampling strategies		
策略	<i>mAP@ 50</i>	<i>mAP@ 50:95</i>
最近邻插值	0.905	0.781
双线性插值	0.906	0.784
Pixel Shuffle	0.906	0.787
CARAFE	0.909	0.789
FADE	0.910	0.789
SAPA	0.912	0.792
DySample	0.913	0.800

表 2 不同注意力机制对比

Table 2 Comparison of different attention mechanisms			
模型	<i>mAP@ 50</i>	参数量/ 10^6	计算量/GFLOPs
YOLOv8s	0.905	11.1	28.7
YOLOv8s+SE	0.908	11.2	28.8
YOLOv8s+ECA	0.909	11.2	28.8
YOLOv8s+CBAM	0.911	11.5	29.0
YOLOv8s+CPCA	0.913	11.6	29.7
YOLOv8s +EMA	0.915	11.1	28.8

表 3 消融实验结果分析

Table 3 Analysis of ablation experiment results								
DySample	ADown	EMA	<i>P</i>	<i>R</i>	<i>mAP@ 50</i>	<i>mAP@ 50:95</i>	参数量/ 10^6	计算量/GFLOPs
✓	✓	✓	0.916	0.861	0.905	0.781	11.1	28.7
			0.931	0.868	0.913	0.800	11.1	28.7
			0.930	0.874	0.916	0.796	9.5	26.6
✓	✓	✓	0.913	0.881	0.915	0.794	11.1	28.8
			0.930	0.875	0.914	0.798	9.6	26.7
			0.925	0.876	0.915	0.798	11.3	28.9
✓	✓	✓	0.919	0.878	0.915	0.800	9.6	26.8
			0.934	0.877	0.917	0.802	9.6	26.8

表 4 不同算法性能对比

Table 4 Comparison of the performance of different algorithms			
模型	<i>mAP@ 50</i>	<i>mAP@ 50:95</i>	参数量/ 10^6
YOLO-G	0.919	0.807	52.5
YOLOv5s	0.880	0.742	7.5
YOLOv5s r6.0	0.893	0.767	9.1
YOLOv6s v3.0	0.905	0.782	16.2
YOLOv7s AF	0.904	0.780	11.0
YOLOv9s	0.906	0.783	7.1
本文模型	0.917	0.802	9.6

由表 4 可知,用于遗留物品检测的 YOLOv5s 虽

4.6 实验结果及分析

本文通过消融实验研究每个模块的加入对模型性能的影响,具体实验结果数据如表 3 所示,原始模型取得的 *P*、*R*、*mAP@ 50* 和 *mAP@ 50:95* 分别为 91.6%、86.1%、90.5%和 78.1%。使用动态上采样 DySample 能够优化上采样过程、增强模型性能,评价指标分别提升了 1.5 百分点、0.7 百分点、0.8 百分点和 1.9 百分点。ADown 模块取代普通步长为 2 的卷积,防止模型过拟合,评价指标分别提升了 1.4 百分点、1.3 百分点、1.1 百分点和 1.5 百分点,同时参数量与计算量均有减少。EMA 注意力能优化特征提取过程,增强特征提取能力,召回率 *R* 提升了 2.0 百分点。改进后的模型 YOLO-DAE 在自建数据集上取得的评价指标分别为 93.4%、87.7%、91.7%和 80.2%,相比于改进前的 YOLOv8s 模型分别提高了 1.8 百分点、1.6 百分点、1.2 百分点和 2.1 百分点。

将用于遗留物品检测的 YOLO-G^[11]、YOLOv5s^[10]算法与近几年提出的新的一阶段算法 YOLOv5s r6.0^[13]、YOLOv6s v3.0^[28]、YOLOv7s AF^[17]、YOLOv9s^[17]和本文提出的算法进行对比,实验结果如表 4 所示。

然参数更少,但其准确率较低;YOLO-G 取得了最高的准确率,但其参数量约为本文模型的 5.5 倍;YOLOv5s r6.0 与 YOLOv7s 的准确率较低;YOLOv6s v3.0 的参数量较高;最新提出的 YOLOv9s 的综合性能较好;而本文模型在保持较低的模型参数量的同时达到了较高的检测准确率。

模型改进前后检测效果如图 9 所示。使用 LayerCAM^[29]生成模型可视化类激活热力图,可以看出,原模型更关注深红色区域,经过改进后置信度从 76%提升到 80%;同时从热力图中可以看出,改进后的模型提取了更多目标区域的特征,不过度依赖无关区域特征,有效提升了遗留物品检测能力。

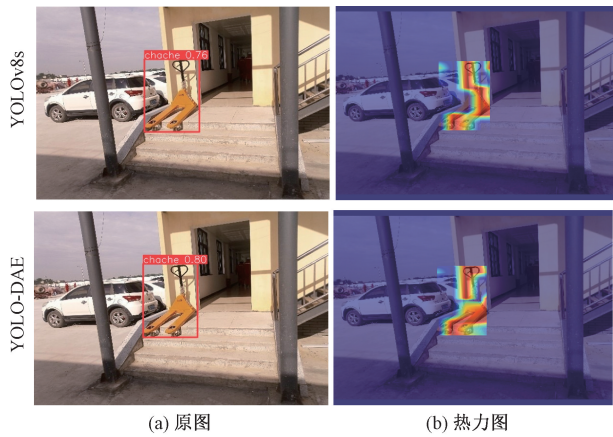


图 9 改进前后检测结果与热力图对比

Figure 9 Results and heat map before and after improvement

5 结论

针对传统遗留物品检测算法难以应对人流拥挤、小目标、物品遮挡和光线变化等环境,以及基于深度学习方法中模型准确率低等问题,提出了一种基于改进 YOLOv8 的遗留物品检测算法。使用具有动态调节因子的 DySample 优化了上采样,增加了模型的泛化能力。利用轻量高效的 ADown 下采样模块缓解网络过拟合,不仅能提升模型准确率,还降低了模型的参数量和计算量。引入 EMA 注意力机制,优化特征提取过程,有效地提升模型性能。改进后模型的准确率 P 、召回率 R 、 $mAP@50$ 和 $mAP@50$:95 分别为 93.4%,87.7%,91.7%和 80.2%,相比于改进前的 YOLOv8s 模型分别提高了 1.8 百分点、1.6 百分点、1.2 百分点和 2.1 百分点,有效提升了遗留物品检测能力。

参考文献:

[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 580-587.

[2] 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: IEEE, 2016: 779-788.

[3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//14th European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[4] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection [C] //2019 IEEE/ CVF International Conference on Computer Vision (IC-

CV). Piscataway: IEEE, 2019: 9626-9635.

[5] LIN K, CHEN S C, CHEN C S, et al. Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance [J]. IEEE Transactions on Information Forensics and Security, 2015, 10(7): 1359-1370.

[6] SU H, WANG W, WANG S W. A robust all-weather abandoned objects detection algorithm based on dual background and gradient operator [J]. Multimedia Tools and Applications, 2023, 82(19): 29477-29499.

[7] RUSSEL N S, SELVARAJ A. Ownership of abandoned object detection by integrating carried object recognition and context sensing[J]. The Visual Computer, 2024, 40(6): 4401-4426.

[8] PARK H, PARK S, JOO Y. Detection of abandoned and stolen objects based on dual background model and mask R-CNN[J]. IEEE Access, 2020, 8: 80010-80019.

[9] LIU W P, LIU P, XIAO C X, et al. General-purpose abandoned object detection method without background modeling[C]//2021 IEEE International Conference on Imaging Systems and Techniques (IST). Piscataway: IEEE, 2021: 1-5.

[10] 蒋晓可. 工业监控视频中的遗留物品检测与识别方法研究[D]. 武汉: 华中科技大学, 2021.

JIANG X K. Research on method of abandoned object detection and recognition in industrial surveillance videos [D]. Wuhan: Huazhong University of Science and Technology, 2021.

[11] 林德钰, 周卓彤, 过斌, 等. 高斯混合模型与 GhostNet 结合的 YOLO-G 遗留物检测方法 [J]. 计算机辅助设计与图形学学报, 2023, 35(1): 99-107.

LIN D Y, ZHOU Z T, GUO B, et al. YOLO-G abandoned object detection method combined with Gaussian mixture model and GhostNet [J]. Journal of Computer-Aided Design & Computer Graphics, 2023, 35(1): 99-107.

[12] Ultralytics. YOLOv8 [EB/OL]. (2023-01-10) [2024-09-10]. <https://github.com/ultralytics/ultralytics>.

[13] Ultralytics. YOLOv5 [EB/OL]. (2020-05-18) [2024-09-10]. <https://github.com/ultralytics/yolov5>.

[14] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: task-aligned one-stage object detection [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 3490-3499.

[15] LI X, WANG W H, WU L J, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [EB/OL]. (2020-06-08) [2024-09-10]. <https://arxiv.org/abs/2006.04388>.

[16] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample [C]//2023 IEEE/CVF International

Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023; 6004–6014.

[17] WANG C Y, YE H I, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024–02–29) [2024–09–10]. <https://arxiv.org/abs/2402.13616>.

[18] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning [C] // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2023; 1–5.

[19] ZHANG Y F, SUN P Z, JIANG Y, et al. ByteTrack: multi-object tracking by associating every detection box [EB/OL]. (2022–04–07) [2024–09–10]. <https://arxiv.org/abs/2110.06864>.

[20] SHI W Z, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016; 1874–1883.

[21] WANG J Q, CHEN K, XU R, et al. CARAFE: content-aware reassembly of features [C] //2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019; 3007–3016.

[22] LU H, LIU W Z, FU H T, et al. FADE: fusing the assets of decoder and encoder for task-agnostic upsampling [EB/OL]. (2022–12–27) [2024–09–10]. <https://arxiv.org/abs/2207.10392>.

[23] LU HAO, LIU W Z, YE Z X, et al. SAPA: similarity-aware point affiliation for feature upsampling [EB/OL]. (2022–12–27) [2024–09–10]. <https://arxiv.org/abs/2209.12866>.

[24] HU J, SHEN L, ALBANIE S. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42(8):2011–2023.

[25] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020; 11531–11539.

[26] WOO S Y, PARK J C, LEE J Y, et al. CBAM: convolutional block attention module [J]. Lecture Notes in Computer Science, 2018, 11211; 3–19.

[27] HUANG H J, CHEN Z G, ZOU Y, et al. Channel prior convolutional attention for medical image segmentation [EB/OL]. (2023–06–08) [2024–09–10]. <https://arxiv.org/abs/2306.05196v1>.

[28] LI C Y, LI L L, GENG Y F, et al. YOLOv6 v3.0: a full-scale reloading [EB/OL]. (2023–01–13) [2024–09–10]. <https://arxiv.org/abs/2301.05586v1>.

[29] JIANG P T, ZHANG C B, HOU Q B, et al. LayerCAM: exploring hierarchical class activation maps for localization [J]. IEEE Transactions on Image Processing, 2021, 30; 5875–5888.

Abandoned Object Detection Algorithm Based on Improved YOLOv8

ZHANG Zhen¹, GE Shuaibing², CHEN Kexin³, LI Youhao⁴, HUANG Weitao⁴

(1. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. Henan Institute of Advanced Technology, Zhengzhou University, Zhengzhou 450003, China; 3. Zhuhai Unitech Power Technology Co., Ltd., Zhuhai 519000, China; 4. Henan Huirong Oil and Gas Technology Co., Ltd., Zhengzhou 450001, China)

Abstract: An abandoned object detection algorithm based on improved YOLOv8 was proposed to address the difficulties of traditional background subtraction based abandoned object detection algorithms in dealing with crowded environments, small targets, occlusion, and light changes, as well as the low accuracy of models based on deep learning methods. Firstly, dynamic upsampling DySample was used to replace the nearest neighbor upsampling, optimizing the upsampling process, and increasing the model’s generalization ability. Secondly, the downsampling convolution was replaced with the efficient lightweight ADown module which reduced the overall model parameters while improving the detection accuracy of the algorithm. In addition, the introduction of EMA attention mechanism optimized the feature extraction process, enhanced feature extraction capabilities, and improved the effectiveness of small object detection. The experimental results showed that the improved model YOLO-DAE achieved *P*, *R*, and *mAP@50* and *mAP@50:95* was 93.4%, 87.7%, 91.7%, and 80.2%, respectively, which was 1.8, 1.6, 1.2, and 2.1 percentage points higher than the original YOLOv8s. And the average accuracy *mAP@50* and *mAP@50:95* was higher than YOLOv5s r6.0, YOLOv6s v3.0, YOLOv7s AF, and YOLOv9s, effectively improving the ability to detect abandoned object.

Keywords: abandoned object detection; YOLOv8 algorithm; EMA attention mechanisms; DySample module; A-Down module