

DOI:10.19651/j.cnki.emt.2210123

基于 Yolo_ES 的垃圾分类目标检测模型^{*}

范金豪^{1,2} 崔立志^{1,2}

(1.河南理工大学电气工程与自动化学院 焦作 454000; 2.河南理工大学河南省煤矿装备智能检测与控制重点实验室 焦作 454000)

摘要: 垃圾分类问题目前已被政府和社会广泛关注,面对分拣过程中对垃圾类别实时性和准确性判断的需求,提出了一种 Yolo_ES 目标检测算法。该算法以 Yolov4 为基础网络,首先使用 EfficientNet 作为主干特征提取网络,实现算法的轻量化;其次通过注意力机制 ECA 对 MBConv 模块进行重构,筛选出高质量的信息,增强模型的特征提取能力并降低了参数量;同时针对最大池化易丢失细节信息的问题,使用 SoftPool 对 SPP 模块中的 MaxPool 层进行替换,保留更多细粒度特征信息。在自制的 HPU_WASTE 垃圾分类数据集上进行实验,结果表明,Yolo_ES 模型相比于 Yolov4 基础网络 mAP 从 91.81% 提升到了 96.06%,模型大小压缩了 75.45% 同时每张图片处理时间为 58ms;与其他目标检测网络相比,该模型具有更好的鲁棒性和更佳检测性能。

关键词: 目标检测;Yolov4;轻量化;注意力机制;SoftPool

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6

Garbage classification target detection model based on Yolo_ES

Fan Jinhao^{1,2} Cui Lizhi^{1,2}(1. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China;
2. Henan Key Laboratory of Intelligent Detection and Control of Coal Mine Equipment, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: At present, garbage classification has been widely concerned by the government and society. Facing the demand for real-time and accurate judgment of waste classification in the sorting process, a Yolo_ES target detection algorithm was proposed. The algorithm takes Yolov4 as the basic network. Firstly, EfficientNet is used as the backbone feature extraction network to realize the lightweight of the algorithm; secondly, the MBConv module is reconstructed by the attention mechanism ECA to filter out the high-quality information, enhance the feature extraction ability of the model and reduce the number of parameters; at the same time, aiming at the problem that it is easy to lose detailed information in the max-pooling, the SoftPool is used to replace the MaxPool layer in the SPP module to retain more fine-grained feature information. The experiments are conducted on the self-made HPU_WASTE garbage classification dataset, and the results show that compared with the Yolov4 basic network, Yolo_ES model increases the map from 91.81% to 96.06%, and the model size is compressed by 75.45%. Meanwhile, the processing time of each image is 58ms; Compared with other target detection networks, this model has better robustness and better detection performance.

Keywords: object detection; Yolov4; lightweight; attention mechanism; SoftPool

0 引言

随着我国经济快速发展和人民生活水平的不断提高,城市生活垃圾增长迅速。2019年上海“最严”垃圾分类条例^[1]发布,使得垃圾分类概念在国内迅速普及并被越来越多人熟

知。对垃圾进行有效且正确的划分是垃圾回收分类的前提,但是目前人工分拣垃圾在一致性、稳定性和卫生条件方面难以达到令人满意的效果。因此,将采集到的垃圾图像结合现有技术,采用智能化、自动化的方式进行处理,不仅可以有效降低人工成本,同时还可以提高资源的利用率。

收稿日期:2022-05-25

* 基金项目:国家自然科学基金(U1804417)项目资助

近些年来,随之深度学习迅猛发展,国内外的学者在垃圾检测领域也开展了广泛研究。Zeng 等^[2]针对高光谱图像垃圾检测问题,提出了一种多尺度卷积神经网络(MSCNN)对 HIS(hue, intensity, saturatuin)数据像素进行分类并生成二值化垃圾分割图,在大面积垃圾检测中具有良好性能。Zhang 等^[3]采集低像素街道垃圾图像作为数据集,使用改进后的 Faste R_CNN 进行训练,模型精度达到了 82%。Liang 等^[4]提出了一种基于卷积神经网络的多任务学习架构(MTLA),此架构对图像中垃圾目标进行识别和定位,在实验中准确度达到了 81.50%。Ma 等^[5]通过对 SSD 特征融合模块进行改进、使用 Focal Loss 解决正负样本失衡问题、采用主干网路更换等措施,改进后的模型在精度和速度方面都有较大提升。陈智超等^[6]使用改进的 MobileNet_v2 用于垃圾图像分类,通过对 MobileNet_v2 基础网络调节宽度因子、嵌入注意力模块和设计多尺度特征融合,在自制数据集上进行测试,网路推理时间降低的同

时准确度为 94.6%。Mao 等^[7]制作了一个基于台湾地区的生活垃圾数据集,使用 YOLOv3 网络进行训练,准确率达到 92.12%。Zhang 等^[8]提出一种 YOLO_WASTE 垃圾分类模型,以 YOLOv4 网络为基础利用迁移学习进行训练,在自建数据集上取得良好效果。

上述方法在一定程度上有效解决了垃圾分类问题,但是仍然存在着垃圾图像单一、数据集种类少以及网络参数过多、模型冗余等缺点。针对这些问题,本文构建了一个多目标生活垃圾数据集,提出了一种基于 YOLO_ES 的垃圾分类目标检测模型。

1 YOLOv4 算法

YOLOv4 目标检测模型主要是由特征提取网络、外加模块和输出三部分组成,结构如图 1 所示。其原理是将固定大小的图像输入网络,基于回归获得边界框位置和其所属类别,从而实现目标的检测。

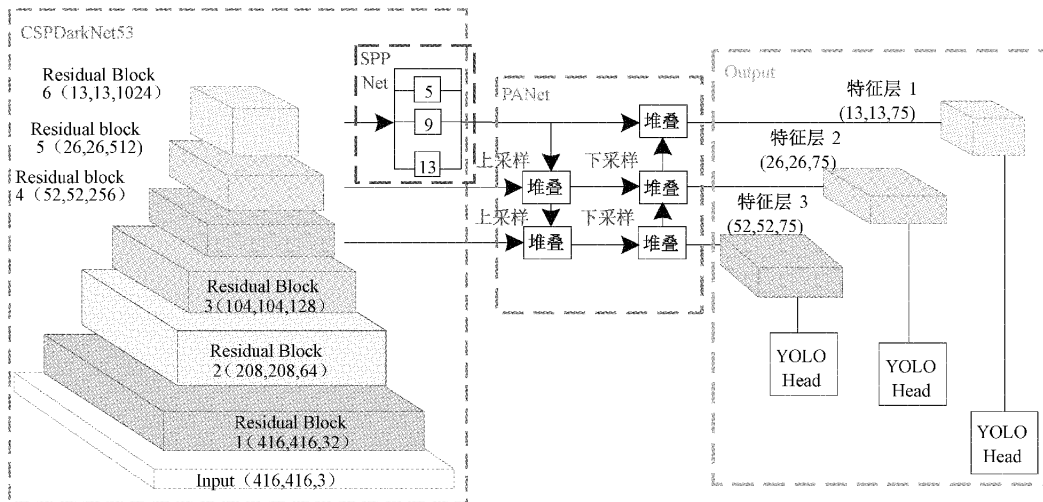


图 1 YOLOv4 网络结构

YOLOv4 主干特征提取网络采用 CSPDarknet53, 经过 8 倍、16 倍和 32 倍下采样输出维度为 $52 \times 52 \times 256$ 、 $26 \times 26 \times 512$ 和 $13 \times 13 \times 1024$ 的特征图。输出特征图经过外加模块 SPP (spatial pyramid pooling) 和 PANet (pyramid aggregation network) 处理。SPP 模块只对残差块 6 (Residual block 6) 输出特征图进行操作, 该特征经过卷积核大小为 1×1 、 5×5 、 9×9 和 13×13 的最大池化层, 将池化后的结果堆叠后再输出。PAN 模块对高层信息进行上采样, 使得特征图维度翻倍后与上层同维度特征进行堆叠; 对底层信息进行下采样, 使得特征图维度减半后与下层同维度特征进行堆叠, 进行充分的特征融合。最后将融合后的特征图输入到 YOLOhead 进行解码预测。

2 本文方法

本文根据 YOLOv4 目标检测模型进行改进, 主要工作有以下几个方面:

1) 数据准备工作。制作 HPU-WASTE 数据集, 涵盖了日常生活中常见的 4 类垃圾, 每张图像包含多个目标且目标种类不一。

2) 采用轻量级网络 EfficientNet-b2 作为 YOLOv4 主干特征提取网络, 降低模型参数、实现算法轻量化。

3) 使用 ECA 模块对 MBConv 进行重构, 提出 R_MBConv 结构。

4) 引入 SoftPool 模块对 MaxPool 进行替换, 避免了池化过程中重要信息丢失的问题。

2.1 数据集

在深度学习任务当中, 数据集的优劣对模型效果起着至关重要的作用^[9]。目前, 现有的垃圾分类研究开源的数据集较少且多数图片只含有单个目标^[10]。本次实验构建一个名为 HPU_WASTE 多目标生活垃圾数据集, 数据集照片由海康威视 DS-2CD3T45D 摄像机手工拍摄, 包括可回收垃圾(标签 0)、其他垃圾(标签 1)、有害垃圾(标签 2)

和厨余垃圾(标签 3)四类,涵盖了烟蒂、电池(小目标样本)及塑料瓶、易拉罐、香蕉皮(中目标样本)和纸箱、一次性手套(大目标样本)等生活垃圾。共 1 694 张 RGB 彩色图像,如表 1 所示。

表 1 HPU_WASTE 垃圾数据集

标签	类别	种类
0	可回收垃圾	塑料瓶,易拉罐,纸箱
1	其他垃圾	烟蒂,一次性手套,废弃口罩
2	有害垃圾	电池,灯泡,废弃药品
3	厨余垃圾	香蕉皮,树叶,橘子皮

HPU_WASTE 数据集示例如图 2 所示。采集好的照片使用图像标注工具 LabelImg 进行打标,并仿照 VOC2007 数据集格式进行存储。

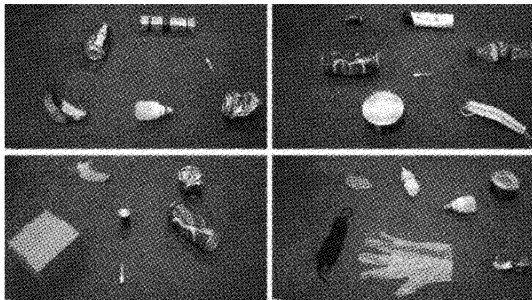


图 2 HPU_WASTE 数据集示例

2.2 EfficientNet 特征提取网络

YOLO 目标检测算法是直接通过回归进行预测,因此选择合适的特征提取网络对检十分关键。EfficientNet^[11]是 Google 在 2019 年提出的轻量级网络,为了使得垃圾分类目标检测更加实时、高效,本文选用 EfficientNet-b2 作为 YOLOv4 的主干特征提取网络,其网络结构如表 2 所示。

表 2 EfficientNet-b2 网络结构

阶段 (stage)	算子 (Operator)	$k \times k$	步距	输出	层数
Stem	Conv	3×3	2	$208 \times 208 \times 32$	1
Black1	MBCConv1	3×3	1	$104 \times 104 \times 16$	2
Black2	MBCConv6	3×3	2	$104 \times 104 \times 24$	3
Black3	MBCConv6	5×5	2	$52 \times 52 \times 48$	3
Black4	MBCConv6	3×3	2	$26 \times 26 \times 88$	3
Black5	MBCConv6	5×5	1	$26 \times 26 \times 120$	4
Black6	MBCConv6	5×5	2	$13 \times 13 \times 208$	5
Black7	MBCConv6	3×3	1	$13 \times 13 \times 352$	2

通过表 2 可知,其有 EfficientNet-B2 由 1 个 Stem 层和 7 个 Block 层组成。Stem 是 1 个卷积核大小为 3×3 卷积层,Block 层由移动翻转瓶颈卷积 (mobile inverted bottleneck convolution, MBCConv) 堆叠而成, MBCConv 网络

结果如图 3 所示。

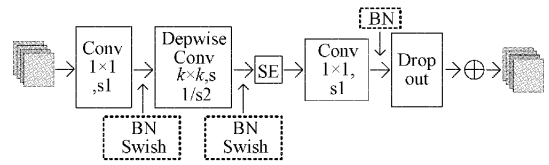
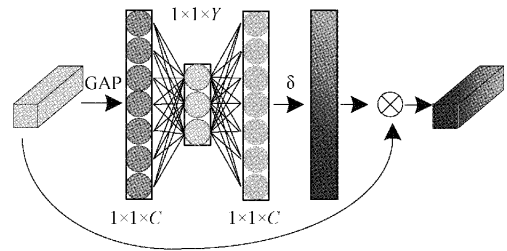


图 3 MBCConv 结构图

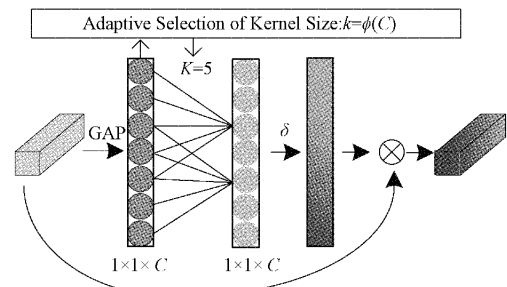
MBCConv 首先使用 1×1 大小的卷积核改变输出通道数,然后通过加入 SE(squeeze-and-excitation)注意力模块的深度可分离卷积进行特征提取,最后通过 Dropout 层并使用残差连接,得到最终的特征图输出。深度可分离卷积减少了模型参数量;通道注意力 SE^[12]模块有利于提高对显著特征特征的表达能;残差连接可以有效解决网络过深引起的梯度消失问题。

2.3 重构 MBCConv 模块

在 MBCConv 的基础上为了进一步探索效率更高、参数更少的网络,本文结合 EAC^[13] 模块对 SE 模块进行替换,提出 R_MBCConv 结构,用于 EfficientNet-b2 特征提取网络。SE 模块和 ECA 模块分别如图 4(a)和(b)所示,其中 SE 模块先使用全局平均池化对输入进行压缩,然后通过两个全连接层进行激励,最后使用 Sigmoid 函数生成通道权重与原输入对应的通道相乘,得到特征图的输出。而 ECA 模块先对每个通道单独全局平均池化,然后通过大小为 K 的快速 1 维卷积实现本地跨通道交互,其中 K 也表示局部跨通道交互的范围,最后使用 Sigmoid 函数生成通道权重。为了避免通过交叉验证进行手动调节 k 造成的计算资源浪费,ECA 采用一种非线性自适应的方法确定 K 的大小,其中一维卷积大小 K (跨通道交互范围)与通道维度成比例。



(a) SE 模块结构



(b) ECA 模块结构

图 4 注意力机制模块结构

ECA 模块的引入使用更少的参数,相较于 SE 模块, ECA 模块避免了降维,并能够有效的捕捉跨通道交互,同时 S 个一维卷积结构复杂度更低,推理速度更快。

2.4 SoftPool 池化

池化操作的主要作用是对特征进行下采样,减少后续层中的参数数量,缓解网络过拟合的风险^[14]。目前常用的池化层主要包换两种类型:最大池化层和平均池化层,前者取特点区域中的最大值进行输出,后者取特定点中的平均值进行输出。在垃圾分类目标检测中,由于部分垃圾特征信息相似度较高,这两种池化类型都容易在池化过程损失重要的信息造成误检。SoftPool^[15]可以有效解决这一问题,其以自然指数 e 为基础,基于 SoftMax 加权方法使得池化内核中的重要信息得到保留。在池化过程中,SoftPool 为内核邻域 R 内的每一个像素 a_i 分配一个权重 w_i , 权重 w_i 为像素 a_i 的自然指数除以邻域 R 中所有像素的自然指数之和,如式(1)所示。

$$w_i = \frac{e^{a_i}}{\sum_{j \in R} e^{a_j}} \quad (1)$$

SoftPool 输出值 \tilde{a} 为内核邻域 R 内所有像素的加权求和所得,由像素值 a_i 和其所对的权重 w_i 共同实现,如式(2)所示。

$$\tilde{a} = \sum_{i \in R} w_i a_i \quad (2)$$

SoftPool 与其他池化相比,使用 SoftMax 具有可微性,因此在每一次反向传播时可以更新梯度值。将 YOLOv4 中 SPP 模块中 MaxPool 层替换为 SoftPool,如图 5 所示,图 6 表示 3 种池化对比图。

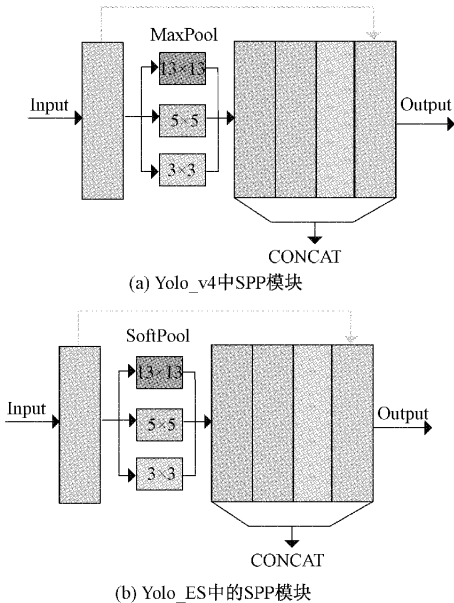


图 5 SPP 模块对比图

2.5 Yolo_ES 网络模型

针对垃圾分类目标检测,在 YOLOv4 的基础上,结合



图 6 池化效果对比图

3.2、3.3 和 3.4 节改进,得到改进后的 Yolo_ES 模型如图 7 所示。

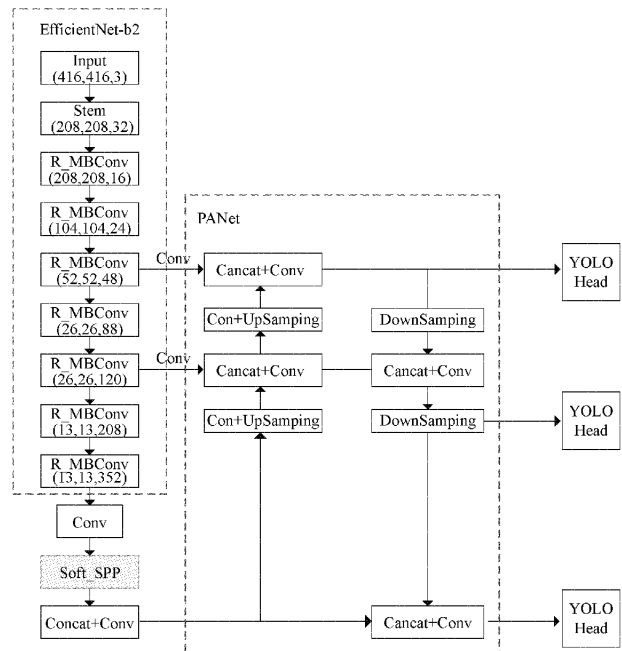


图 7 Yolo_ES 网络结构图

R_MBCConv 表示重构后的 MBCConv 模块,Soft_SPP 表示使用 SoftPool 池化的 SPP 模块。新的 Yolo_ES 网络架构模型可以快速的实现垃圾目标的检测和分类,在减少模型复杂度的同时对精度进行提升。

3 实验与结果分析

本文使用的服务器操作系统为 64 位 Ubuntu18.04,处理器为 Inter Xeon CPU E5-2696v4,内存为 12 G,搭配 480 G 固态硬盘,GPU 为 NVIDIA GeForce 2080ti,显存为

11 G,使用 Pytorch 深度学习框架,编程语言为 Python。

为了增加数据样本的多样性,本文采用数据增强 imgaug augmenters 工具对数据集进行扩充。通过对图片使用反转、平移、缩放以及高斯模糊、亮度变换等策略^[16],每张图片扩充为原来的 3 倍,获得共计 5 082 张新样本数据集,按照 9 : 1 的比例分为训练集和测试集,应用于所有实验。

本文模型训练迭代次数共计 150 个 Epoch,使用 Adam 优化器,初始学习率为 0.001,动量大小为 0.875,权重衰减为 0.000 5,batchsizes 设置为 16。对比优化前后的模型训练过程如图 8 所示,可以看出改进后的 Yolo_ES 在收敛速度和定位精度方面都优于原始 YOLOv4 模型。

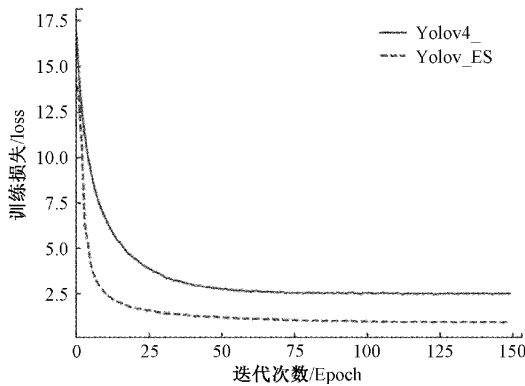


图 8 模型训练损失函数变化曲线

3.1 评价指标

本次研究选取平均精度 (average precision, AP)、平均精度均值 (mean average precision, mAP)、参数量 (params) 和单张图片的检测时间作为模型评价指标。其中 AP 和 mAP 需要对精确度 (precision, P) 和召回率 (recall, R) 计算所得,公式如下:

$$P = TP / (TP + FP) \tag{3}$$

$$R = TP / (FP + FN) \tag{4}$$

$$AP = \int_0^1 P dR \tag{5}$$

$$mAP = \frac{1}{n} \sum_{i=0}^n AP_i \tag{6}$$

式中: TP 表示被正确判定的正样本数, FP 表示被误检样本数, FN 表示漏检样本数; AP 为精确度 P 和召回率 R 所组成的曲线与坐标轴围成的面积; mAP 是所有 AP 的平均值, n 为训练样本中总类别数, i 为当前类别,当 mAP 越高时,表示模型预测性能越好。

3.2 消融实验

为了验证不同改进措施对算法的影响,本文设置了一组消融实验。实验包括:1)原始 YOLOv4 模型;2)EfficientNet 作为主干特征提取网络的 YOLOv4 模型;3)EfficientNet + ECA 作为主干特征提取网络的 YOLOv4 模型;4)EfficientNet 作为主干特征提取网络同时加入 SoftPool 的 YOLOv4 模型;5)Yolo_ES (EfficientNet + ECA + SoftPool) 模型。实验结果如表 3 所示。

表 3 消融实验

模型	AP/%				mAP%	参数量 (Params)/M	检测时间/ms
	可回收垃圾 (0)	其他垃圾 (1)	有害垃圾 (2)	厨余垃圾 (3)			
Yolov4	95	89	87	96	91.81	256.3	68
Yolo+EfficientNet	95	88	87	95	91.08	64	63
Yolo+EfficientNet+ECA	97	94	91	97	94.58	62.9	55
Yolo+EfficientNet+SoftPool	97	94	93	96	94.69	64	60
Yolo_ES (EfficientNet+ECA+SoftPool)	97	96	94	97	96.06	62.9	58

从表 3 实验结果可以看出,EfficientNet 中深度可分离卷积的作用使得模型的参数量降低为原来的 1/4;ECA 模块中一维卷积操作代替全连接和跨通道交互作用,使得模型大小进一步减少 1.1 M 的同时 mAP 提升了 2.77%;使用 SoftPool 取代 SPP 模块中的 MaxPool 池化,保留更多细腻度信息,对包含小目标物体的其他垃圾和有害垃圾检测效果提升明显,平均精度均值 mAP 为 94.69%。

最终,在原始 YOLOv4 模型基础上,通过引入 EfficientNet 网络、ECA 注意力机制和 SoftPool 池化模块,本研究所提出算法在 HPU_WASTE 生活垃圾数据集中相较于基础 YOLOv4 模型大小降低了 75%,检测时长为 58 ms,平均精度均值 mAP 提高了 4.25%,达到了

96.06%,实现轻量化的同时提高了模型的准确率。

图 9 所示为部分检测样本在模型改进前后效果对比图。从左至右分别为标注图、YOLOv4 和 Yolo_ES 模型检测效果图。

通过图 9 可知,原始的 YOLOv4 模型在第二组实验中出现了漏检的烟蒂的情况,在第三组实验中,将有害垃圾电池误检为其他垃圾;改进后的 Yolo_ES 模型避免了小目标物体漏检和误检的问题,同时对中大型目标在置信度和目标框定位方面也有稳定的提高。

3.3 不同模型对比分析

为了进一步验证本文所提出算法的先进性,将 Yolo_ES 与其他网络模型进行对比,并选取平均精度均值、参数

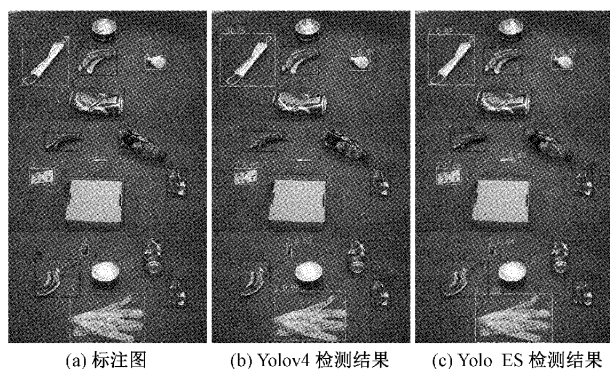


图 9 模型改进前后结果对比图

量和单张图片检测时间作为评价指标,对比实验结果如表 4 所示。

表 4 不同算法实验对比

模型	mAP/ %	参数量 (Params)/M	检测时间/ ms
Faste R_CNN	94.07	547	172
Yolov3	90.03	246.9	74
SSD	87.26	105.2	51
Yolov3(EfficientNet-b3)	91.67	81.1	68
FSSD	90.93	130	59
Yolov4(DenseNet-121)	88.81	64.8	65
Yolo_ES(ours)	96.06	62.9	58

由表 4 可得, Yolo_ES 与传统的目标检测算法 Faste R_CNN^[17]、Yolov3^[18] 和 SSD^[19] 相比平均精度均值 mAP 和参数量方面都有明显优势,虽然检测时长略高于 SSD,但依然满足实时检测的需求。对比于使用 EfficientNet-b3 作为主干网络的 Yolov3 模型、改进 SSD 特征融合模块的 FSSD 网络和使用 DenseNet-121 作为主干网络的 Yolov4 模型,本文所提出的算法综合表现更佳,具有更强的检测性能。

4 结 论

针对目前垃圾分类数据集样本种类少、目标单一以及垃圾分类网络模型参数冗余、精度不高等问题,本文构建了一个 HPU_WASTE 多目标垃圾数据集,并以 Yolov4 网络为基础进行改进,提出 Yolo_ES 模型用于垃圾分类目标检测任务当中。该模型主要通过引入 EfficientNet-b2 作为主干特征提取网络,实现算法的轻量化;使用 ECA 注意力重构 MBConv 模块,增强特征提取;利用 SoftPool 替代最大池化,保留更多细粒度信息,提高对小目标检测能力。实验结果表明,本文提出的 Yolo_ES 模型能够实时、准确的对垃圾目标进行定位与识别,具有实际应用的潜能。

下一步工作将使用 Delta 机器人与 Yolo_ES 模型进行

手眼结合,推动行业迈向智能化方向发展。

参考文献

- [1] 上海市生活垃圾管理条例[N]. 解放日报,2019-02-20(008).
- [2] ZENG D, ZHANG S, CHEN F, et al. Multi-scale CNN based garbage detection of airborne hyperspectral data[J]. IEEE Access,2019, 7: 104514-104527.
- [3] ZHANG R, YIN D, DING J, et al. A detection method for low-pixel ratio object [J]. Multimedia Tools and Applications, 2019, 78(9): 11655-11674.
- [4] LIANG S, GU Y. A deep convolutional neural network to simultaneously localize and recognize waste types in images[J]. Waste Management, 2021, 126: 247-257.
- [5] MA W, WANG X, YU J. A lightweight feature fusion single shot multibox detector for garbage detection[J]. IEEE Access, 2020, 8: 188577-188586.
- [6] 陈智超,焦海宁,杨杰,等. 基于改进 MobileNet v2 的垃圾图像分类算法[J]. 浙江大学学报(工学版),2021, 55(8): 1490-1499.
- [7] MAO W L, CHEN W C, FATHURRAHMAN H I K, et al. Deep learning networks for real-time regional domestic waste detection [J]. Journal of Cleaner Production, 2022, 344: 131096.
- [8] ZHANG Q, YANG Q, ZHANG X, et al. A multi-label waste detection model based on transfer learning [J]. Resources, Conservation and Recycling, 2022, 181: 106235.
- [9] 赵勇,雷欢,马敬奇,等. 迁移模式下卷积神经网络在虹膜识别中的应用[J]. 电子测量技术, 2020, 43(9): 114-120.
- [10] 李金玉,陈晓雷,张爱华,等. 基于深度学习的垃圾分类方法综述[J]. 计算机工程, 2022, 48(2): 1-9.
- [11] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International conference on machine learning, 2019: 6105-6114.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 7132-7141.
- [13] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [14] 雷江波,王泽民,李静. 基于 Faster R-CNN 的碎片群图像目标检测研究[J]. 国外电子测量技术, 2021,

- 40(1): 70-74.
- [15] STERGIU A, POPPE R, KALLIATAKIS G. Refining activation downsampling with SoftPool[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10357-10366.
- [16] 史朋飞, 韩松, 倪建军, 等. 结合数据增强和改进 YOLOv4 的水下目标检测算法[J]. 电子测量与仪器学报, 2022, 36(3): 113-121.
- [17] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [18] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.
- [19] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European conference on computer vision, 2016: 21-37.

作者简介

范金豪, 硕士, 主要研究方向为计算机视觉、目标检测。

E-mail: 1002198611@qq.com

崔立志(通信作者), 副教授, 博士, 主要研究方向为人工智能信号处理、控制工程。

E-mail: clzh0308@126.com