

# 基于改进 YOLOv5s 的可回收垃圾检测算法

罗安能<sup>1</sup>, 万海斌<sup>1,2\*</sup>, 司志巍<sup>1</sup>, 覃团发<sup>1,2</sup>

<sup>1</sup>广西大学计算机与电子信息学院, 广西 南宁 530004;

<sup>2</sup>广西多媒体通信与网络技术重点实验室, 广西 南宁 530004

**摘要** 垃圾回收的好处有很多,有助于保护水土资源,提高居民的生活环境质量,加快绿色循环经济发展,然而传统的垃圾回收需要大量人力和物力。结合 ShuffleNet v2 与深度可分离卷积,提出一个更轻量化的 YOLOv5s 改进模型,将其用于对可回收垃圾的分类和定位。实验结果表明:改进模型的参数量仅为原始模型参数量的 38.98%;在输入分辨率为  $640 \times 640$  时,改进模型的平均精度均值(mAP)为 94.01%,比原始 YOLOv5s 高出 1.91 个百分点;在速度上,通过在 Jetson Nano 硬件上进行部署,改进模型的前传耗时比原始 YOLOv5s 少了 11.5%。另外,与目前常见的主流的目标检测模型对比,所提改进模型也具有很好的表达可回收垃圾特征的能力。

**关键词** 垃圾回收; YOLOv5s; ShuffleNet v2; 深度可分离卷积

中图分类号 TP391.4; X705

文献标志码 A

DOI: 10.3788/LOP220603

## Detection Algorithm of Recyclable Garbage Based on Improved YOLOv5s

Luo Anneng<sup>1</sup>, Wan Haibin<sup>1,2\*</sup>, Si Zhiwei<sup>1</sup>, Qin Tuanfa<sup>1,2</sup>

<sup>1</sup>School of Computer, Electronics and Information, Guangxi University, Nanning 530004, Guangxi, China;

<sup>2</sup>Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, Guangxi, China

**Abstract** Garbage recycling offers many benefits, e. g. , protection of water and soil resources, quality improvement of the living environment of residents, and accelerated development of green circular economy. However, traditional garbage recycling methods incur excessive labor and resource costs. In this work, we propose a lighter YOLOv5s improved model in which ShuffleNet v2 and deep separable convolution methods are combined to better solve the problems in garbage recycling by classifying and locating recyclable garbage more efficiently. Experimental results show that number of parameters of the improved model is only 38.98% of that of the original model. When the input resolution is  $640 \times 640$ , the mean average precision (mAP) of the improved model is 94.01%, which is 1.91 percentage points higher than the original YOLOv5s. With regard to the computing speed, the forward propagation time of the improved model is 11.5% greater than that of the original YOLOv5s by deploying on hardware of Jetson Nano. Moreover, compared with the current mainstream target detection models, the improved model has a good ability to express the characteristics of recyclable garbage.

**Key words** garbage collection; YOLOv5s; ShuffleNet v2; depthwise separable convolution

## 1 引言

目标检测,简而言之就是计算出图片或者视频中包含目标的类别和目标的定位,是各种智能机器人与外界感知的视觉基础。将目标检测技术应用于垃圾回收,可以大大提高垃圾回收效率。目标检测可分为传统的目标检测和深度学习的目标检测。近年来随着深

度学习的快速发展,很多研究人员提出了各种超越传统目标检测的方法。深度学习的主流目标检测算法按照完成的步骤可以分为两类:two-stage 目标检测算法(RCNN 系列)和 one-stage 目标检测算法(YOLO 系列)。Two-stage 检测算法包括候选区域生成和细致检测两个过程;one-stage 检测算法直接在特征解码的同时输出目标类别和坐标,在减小时间成本和提高计

收稿日期: 2022-01-21; 修回日期: 2022-02-17; 录用日期: 2022-03-01; 网络首发日期: 2022-03-15

基金项目: 国家自然科学基金(61961004)

通信作者: hbwan@gxu.edu.cn

算效率方面表现更好。RCNN 系列算法主要包括 Faster-RCNN<sup>[1]</sup>、Mask-RCNN<sup>[2]</sup>等。YOLO 系列算法主要包括 YOLOv3<sup>[3]</sup>、YOLOv4<sup>[4]</sup>、YOLOv5、YOLOX<sup>[5]</sup>等。

然而采用目标检测进行垃圾分类与定位还存在着许多困难。1)缺少垃圾检测的数据集且数据集有一定的局限性。例如数据集的规模较小,如 TrashNet 数据集仅有 2527 张图像<sup>[6]</sup>,TACO 数据集仅包含 1500 张图像<sup>[7]</sup>;同时不同数据集之间也容易存在类别定义的主观差异,因此简单融合多个数据集也有一定的难度;还有一些数据集只包含垃圾的分类信息,例如 2019 年华为公司垃圾分类大赛中提供了一个数据集,该数据集有 44 个类,总计 14964 张图像,数据集中只包含了分类的信息,并没有给目标的坐标进行标注<sup>[8]</sup>。2)现有大多数的目标检测模型参数量庞大,导致模型内存占用大、训练时间太长,不适合在廉价硬件设备上开发。3)垃圾检测难度大,在日常生活中的垃圾形态是多种多样的,例如垃圾的变形、污垢、老化、破损等,这些不规则和多变的形态都会对结果造成很大的影响。

针对这些问题,本文的主要工作和贡献如下:由于没有相关可回收垃圾数据集,创建常见的 10 类可回收

分类垃圾数据集;YOLOv5s 是 YOLOv5 四个版本网络中深度最浅和宽度最小的模型,通过实验对比,在常见的目标检测系列网络中也具有出色的速度和精度,因此选择 YOLOv5s 作为基准模型,并对网络进行改进和消融实验验证;Jetson Nano 是英伟达公司推出的低功耗、微型的嵌入式开发套件,开发用途非常广泛,因此将模型实际部署到嵌入式 Jetson Nano 上,进行实际推理速度的对比。

## 2 YOLOv5s 原理

### 2.1 YOLOv5s 结构

YOLOv5 架构包含 4 种不同的模型,这些模型的区别仅在于控制网络的深度和网络的宽度的不同。以 YOLOv5s 为例,如图 1 所示,网络结构主要包括 4 个部分:输入端(Input)、主干网络(Backbone)、颈部(Neck)、预测头(Head)。输入端主要的工作是进行图像数据的预处理工作。主干网络的主要功能是提取图像中一系列不同尺度的特征图。预测头用于检测的收尾阶段,通过将锚框机制与特征图数据关联,生成具有目标类别、目标类别概率和目标预测框的输出矩阵。

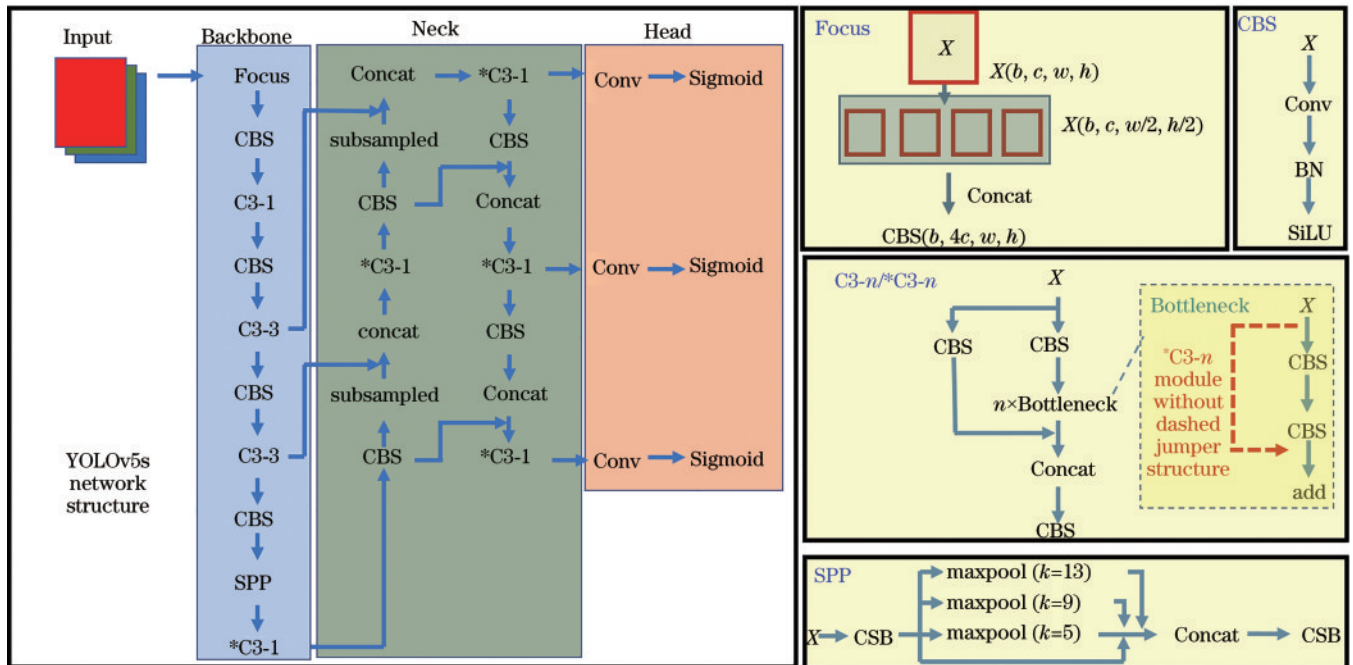


图 1 YOLOv5s 网络结构

Fig. 1 YOLOv5s network structure

### 2.2 YOLOv5s 损失函数

损失函数帮助调整到最小化网络误差。YOLOv5s 总的损失函数由 3 个部分组成,包含位置坐标预测损失、置信度预测损失和类别预测损失<sup>[9]</sup>。位置坐标预测损失采用完全交并比(CIoU)<sup>[10]</sup>度量,置信度预测损失和类别预测损失采用交叉熵损失度量。

## 3 YOLOv5s 网络结构改进

改进网络主要借鉴了 ShuffleNet v2 的基本单元与深度可分离卷积。

### 3.1 ShuffleNet v2 基本单元

2018 年 Zhang 等<sup>[11]</sup>首次提出 ShuffleNet v1,主要

是针对计算能力非常有限的设备的,与其他卷积神经网络(CNN)相比,该架构在保持准确性的同时大大降低计算成本。2018年他们依据设计高效网络结构的四个原则,进一步优化了 ShuffleNet v1 的基本单元结构,提出了 ShuffleNet v2<sup>[12]</sup>。

图 2 展示了 ShuffleNet v2,包含两个基本单元,其

中图 2(a)为提取特征的同时保持特征图大小不变的模块,图 2(b)为一个采用步长为 2 的单元进行 2 倍下采样的模块。从图 2 可以看出 ShuffleNet v2 的主要结构包括通道分割(channel split)、逐通道卷积(DWConv)和通道混洗(channel shuffle)。另外图 2(c)为通道混洗示意图,其还包含了分组卷积(GConv)。

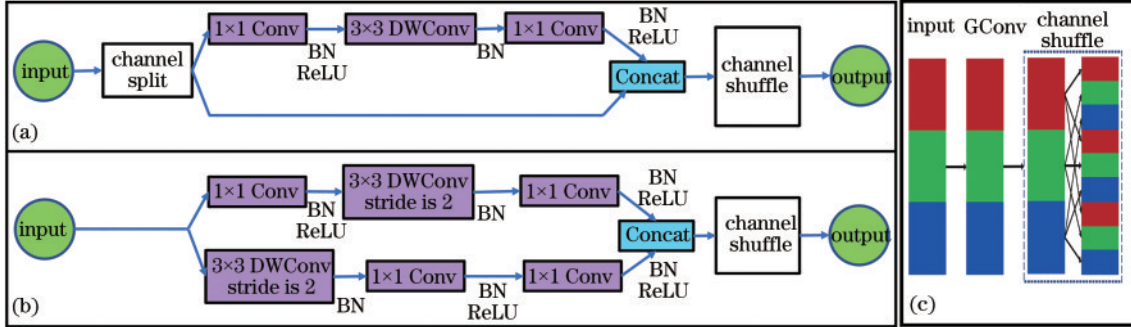


图 2 ShuffleNet v2 基本单元  
Fig. 2 Basic unit of ShuffleNet v2

通道分割把输入的通道分成两半。对于传统卷积运算,假设输入的特征图空间大小为  $D_w \times D_h \times C_{in}$ ,卷积核大小为  $D_K \times D_K$ ,卷积核个数为  $N$ 。传统卷积层的参数量  $P_C$  满足

$$P_C = D_K \times D_K \times C_{in} \times N. \quad (1)$$

分组卷积就是对输入特征图分小组后按各小组单独进行卷积运算的过程。当输入的特征图空间大小与式(1)传统卷积参数的输入的假设一样时,假设分组数为  $g$ ,则每小组的特征图空间大小为  $D_w \times D_h \times C_{in}/g$ ,该小组对应需要的卷积核的空间大小为  $D_K \times D_K \times N/g$ ,最后对各个小组的结果进行拼接。分组卷积的卷积层参数  $P_G$  满足

$$P_G = D_K \times D_K \times \frac{C_{in}}{g} \times \frac{N}{g} \times g = D_K \times D_K \times C_{in} \times N \times \frac{1}{g}. \quad (2)$$

对比式(1)与式(2),可以看出分组卷积可以将传统卷积的参数量压缩  $g$ 。逐通道卷积就是一个卷积核负责一个通道,是分组卷积中分组数等于输入通道数的一个特例。由于分组卷积各小组彼此之间没有信息的交流,图 2(c)中通道混洗方式能够实现跨小组的信息融合,增加信息的表达能力。

### 3.2 DW 模块单元

DW 模块借鉴 Xception<sup>[13]</sup> 与 MobileNet 系列网络<sup>[14-16]</sup> 采用深度可分离卷积(depthwise separable convolution)的成功经验,是由逐通道卷积和  $1 \times 1$  传统卷积组成的,如图 3 所示。逐通道卷积的每一个通道由一一对应的一个卷积核计算,所以逐通道卷积后输出的特征图在空间上显得很单薄。DW 模块进行逐通道卷积后第二部分采用传统卷积,对第一部分各自独立的通道进行组合,有利于通道的信息聚合,使用的卷积核大小为  $1 \times 1$ ,参数量也较小。进行逐通道卷积和



图 3 DW 模块单元  
Fig. 3 DW module unit

传统卷积后,都接着进行批归一化,防止过拟合,加快网络收敛速度,最后利用 ReLU 激活函数赋予网络非线性的表达能力。

当输入的特征图空间大小与式(1)传统卷积参数的输入的假设一样时,使用 DW 模块的卷积层参数量  $P_{DW}$  满足

$$P_{DW} = D_K \times D_K \times C_{in} + C_{in} \times N. \quad (3)$$

对比式(1)与式(3),CNN 中卷积核数量  $N$  通常较大,所提改进网络使用的逐通道卷积的卷积核大小为  $5 \times 5$ ,相比常用的大小为  $3 \times 3$  的卷积核,能够获得更大感受野,理论上大约能压缩  $1/25$  的参数量。

### 3.3 改进网络的实现

#### 3.3.1 主干网络的改进

ShuffleNet v2 论文提出的第 3 条准则要求应该尽量减少网络分支。类似于磁盘中的碎片,过多的分支网络导致网络碎片化,降低网络运行速度。原网络中的 C3 模块、\*C3 模块是由 Bottleneck 等较复杂部件组成的分支结构,考虑对其进行轻量化改进,避免这些结构导致冗余的计算。ShuffleNet v2 基本单元中也保持了类似于 C3 模块的两个分支的结构,分支中的设计相对简单,但也包含卷积、激活函数、批归一化等最关键运算,其中基本单元图 2(a)更是对一个分支直接进行拼接的映射,极大减少了计算量。因此本研究在主干网络中去除了 C3 模块和 \*C3 模块,使用 S-a 模块即 ShuffleNet v2 基本单元图 2(a)和 S-b 模块即

ShuffleNet v2 基本单元图 2(b) 重新构建主干网络。S-a 模块用于加强特征提取, S-b 模块在模型中作为向下采样的角色, 以压缩特征图长宽大小, 抽取出深层次的特征。

改进主干网络结构按表 1 的序号从小到大依次堆叠起来, 其中  $n$  表示模块重复的次数。中层次特征兼容性更好, 为加强对中层次特征的提取且序号为 4 和 6 的模块需要在颈部网络进行特征融合, 因此序号为 4 和 6 的模块重复次数更多, 设为 3; Params 表示在序号位置上模型的参数量; Module 为模块名称; 参数配置(Configuration) 中第一个参数表示输入的通道数, 第二个参数表示输出特征图的通道数, 第三个参数表示卷积核步长; SPP 参数配置稍有不同, 第一、第二参数的含义同其他模块, [5, 9, 13] 表示使用的池化核大

表 1 主干网络结构与参数的配置

Table 1 Configuration of backbone network structure and parameters

Serial number	$n$	Params	Module	Configuration	Output size
0	1	3520	Focus	[3, 32, 3]	$32 \times 320 \times 320$
1	1	3968	S-b	[32, 64, 2]	$64 \times 160 \times 160$
2	1	2528	S-a	[64, 64, 1]	$64 \times 160 \times 160$
3	1	14080	S-b	[64, 128, 2]	$128 \times 80 \times 80$
4	3	27456	S-a	[128, 128, 1]	$128 \times 80 \times 80$
5	1	52736	S-b	[128, 256, 2]	$256 \times 40 \times 40$
6	3	104064	S-a	[256, 256, 1]	$256 \times 40 \times 40$
7	1	203776	S-b	[256, 512, 2]	$512 \times 20 \times 20$
8	1	656896	SPP	[512, 512, 5, 9, 13]	$512 \times 20 \times 20$
9	1	134912	S-a	[512, 512, 1]	$512 \times 20 \times 20$

小为  $5 \times 5, 9 \times 9, 13 \times 13$ , 这三个不同池化核的最大池化操作能在特征图中只取出最主要的信息, 舍弃掉其他次要的信息, 从而降低参数量, 最后对这三个不同池化的输出与直接跳连没有进行池化的数据进行拼接, 能够加强对不同尺度特征的融合。假设输入尺寸为  $3 \times 640 \times 640$ , 从表中的输出特征图中可以看出, 经过这一系列步骤, 特征图的长宽尺度逐渐缩小, 通道数量逐渐扩张, 特征将由浅入深地一步一步得到表达。

3.3.2 颈部网络的改进

借鉴 MobileNet v2 的倒残差结构 (inverted residual block)<sup>[15]</sup>, 将深度可分离卷积应用于深层的特征上, 能够在减少参数量的同时对特征提取的影响较小。YOLOv5s 颈部特征融合网络包含很多深层次的特征, 其中 \*C3 模块包含复杂的分支结构, 不符合 ShuffleNet v2 提出的第 3 条准则, 因而去除颈部的 \*C3 模块, 用轻量化的 DW 模块进行重新构建。

颈部结构结合特征金字塔网络 (FPN)<sup>[17]</sup> 和金字塔注意力网络 (PAN)<sup>[18]</sup> 进行多尺度融合, 三个尺度特征融合分别来自主干网络中的序号 4、序号 6 和序号 9 的模块的输出特征图, 如图 4(a) 所示, 这些特征图的空间长宽尺寸分别为  $80 \times 80, 40 \times 40, 20 \times 20$ 。长宽为  $80 \times 80$  的特征图为较浅层特征, 通常包含丰富的位置信息, 利于检测小目标; 长宽为  $40 \times 40$  的特征图介于两者之间, 用于检测中等大小的目标; 长宽为  $20 \times 20$  的特征图为较为深层的特征, 通常包含更多的大目标特征信息, 用于检测大目标。从融合的整体结构上看, 特征图有从上向下的下采样, 又有从下向上的上采样, 又有相同空间长宽尺寸特征图的拼接融合, 能够进一步提升对不同大小目标的检测能力。

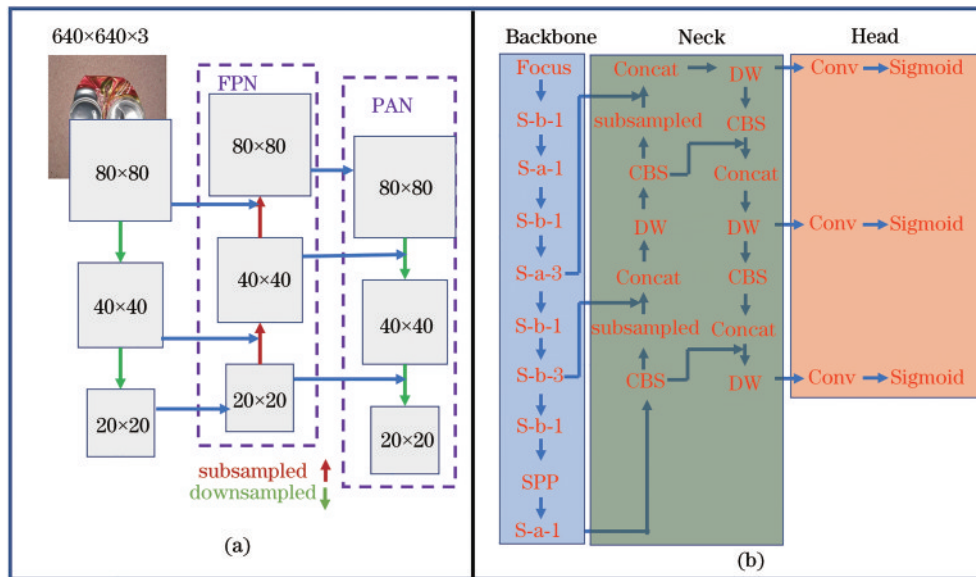


图 4 结构示意图。(a) 颈部融合示意图; (b) 改进 YOLOv5s 整体网络示意图

Fig. 4 Structure diagram. (a) Schematic of Neck fusion; (b) schematic of improved YOLOv5s overall network

综合上述改进,改进网络示意图如图 4(b)所示。特别地, S-a-n 和 S-b-n 的命名含义为: S 代表 ShuffleNet v2, a 代表基本单元图 2(a)、b 代表基本单元图 2(b)、参数 n 表示模块重复的次数,如 S-a-1,表示的结构为 ShuffleNet v2 基本单元图 2(a),并且重复 1 次,具体结构参照图 2。DW 即为图 3 DW 模块的结构; Focus 模块、SPP 模块、CBS 模块参照图 1; Focus 模块主要进行切片和拼接操作,输入的特征图先经过切片操作,长宽切为原来的一半,后对其进行拼接以保持原有数据的信息不变; CBS 模块为传统卷积、批归一化、SiLU 激活函数的组合操作; Concat 为拼接操作, Conv 为传统卷积, Sigmoid 为 Sigmoid 激活函数。

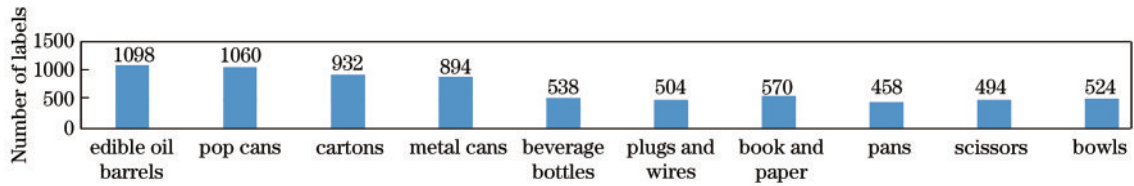


图 5 数据集中各类别的标签数量统计

Fig. 5 Statistic on number of labels in each class on the dataset

## 4.2 评价指标

为了评估模型的可回收垃圾检测性能,使用一些对比指标:精确率 (precision, 简称为 P)、召回率 (recall, 简称为 R)、平均精度 (AP)、平均精度均值 (mAP)、参数量 (Parameter)、所需内存 (Memory)、前传耗时 (FP time)。所需内存指模型权重文件占用的大小。前传耗时指从输入一张图像到输出最终结果所消耗的总时间,包括前预处理耗时 (preprocessing time)、网络前向传播耗时 (inference time)、非极大值抑制的后处理耗时 (NMS time)。

## 5 实验结果与分析

### 5.1 相关的消融实验

进行目标检测的消融实验,是为了研究特定的改进方法对结果的影响,如表 2 所示,根据是否对主

## 4 实验数据与评价指标

### 4.1 实验数据集

在已有数据集的基础上,还采用爬虫网站、手机等途径获取更多图片,对数据进行增强处理,通过随机旋转、随机亮度变化等方式扩充数据,数据变换后更加符合目标不同的形态,获取共 6632 张图片。数据包含 10 个类别:食用油桶 (edible oil barrels)、易拉罐 (pop cans)、纸盒纸箱 (cartons)、金属食品罐 (metal cans)、饮料瓶 (beverage bottles)、插头电线 (plugs and wires)、书籍纸张 (book and paper)、锅 (pans)、剪刀 (scissors)、碗 (bowls)。使用的标注软件为 LabelImg。数据集中各类的标签数量统计如图 5 所示。

干网络进行改进,是否对颈部特征融合网络进行改进,划分 4 个实验。第 1 个为未作改动的原始 YOLOv5s 模型,第 2 个为更改主干网络的改进方式,第 3 个为更改颈部网络的改进方式,第 4 个为改进主干网络的同时又改进颈部网络,即最终的所提改进模型。

所采用的实验训练环境是由本地主机 SSH 远程连接服务器搭建的。服务器配置:CPU 处理器为 Intel Xeon Gold 6230;显卡为 NVIDIA Tesla T4;显卡内存为 16G;操作系统为 CentOS;CUDA 版本为 10.1, PyTorch 版本为 1.7.1, Python 版本为 3.8.10。训练设置:batch\_size 为 16;输入图像分辨率为 640×640;epochs 为 500。在研究中,准备好的图片集应该分为训练集、验证集和测试集,具体为训练图片 4052 张,验证集图片 1013 张,测试图片 1267 张。

表 2 总类别的消融实验数据

Table 2 Ablation experiment data of general category

Backbone	Neck	P / %	R / %	mAP <sub>0.5</sub> / %	mAP <sub>0.5:0.95</sub> / %	Parameters / 10 <sup>6</sup>	Memory / MB
		89.37	85.18	92.09	68.88	7.09	13.7
✓		88.85	89.43	93.36	68.10	4.08	8.03
	✓	88.41	89.15	93.03	71.40	5.69	11.05
✓	✓	90.12	89.95	94.01	71.30	2.68	5.34

消融实验的测试数据如表 2 和表 3 所示, P、R、AP、mAP<sub>0.5</sub> 都是在 IoU 阈值为 0.5 时取得的。mAP<sub>0.5:0.95</sub> 表示 IoU 从 0.5 到 0.95, 间隔为 0.05 时的 mAP。从表 2 可知:第二、第三个实验的 P、R、mAP<sub>0.5</sub>、mAP<sub>0.5:0.95</sub> 个别指标值低于原始模型,

但在参数量和所需内存上,都比原始模型小很多;最终改进模型的 P、R、mAP<sub>0.5</sub>、mAP<sub>0.5:0.95</sub> 指标均优于原始模型,分别增加了 0.75 个百分点、4.77 个百分点、1.92 个百分点、2.42 个百分点,且参数量仅为原 YOLOv5s 参数量的 37.8%, 所需内存仅为原

表 3 单个类别的消融实验数据

Table 3 Experimental data for ablation of individual class

Class	P / %		R / %		AP / %	
	YOLOv5s	Proposed model	YOLOv5s	Proposed model	YOLOv5s	Proposed model
Edible oil barrels	93	94.1	91.5	97.9	97.2	97.4
Pop cans	84.4	90.1	76.3	85.9	89	91.8
Cartons	90.8	91.4	91.7	93.5	96.3	97.3
Metal cans	81.8	80.9	78.9	85.3	84	88.7
Beverage bottles	77.3	81.7	74.4	76.2	82.9	86.1
Plugs and wires	82.6	86.5	67.6	78.2	80.6	88.2
Book and paper	94.7	95	94.6	97.5	98.1	98.2
Pans	95.9	93.3	87.2	91.7	96.4	97.7
Scissors	97.2	90.4	89.6	93.3	97.1	95.2
Bowls	96	97.8	100	100	99.3	99.5

YOLOv5s 的 39.0%。在单个类别消融实验对比中, 最终改进模型对 metal cans、pans、scissors 三类的 P 指标低于原始模型, R 指标没有任何一类低于原始模型, 仅 scissors 类中的 AP 指标低于原始模型。

如图 6 所示, 在 S-b 模块和 CBS 模块获得的特征

图中, S-b 模块获得的特征图轮廓比 CBS 模块获得的特征图轮廓更为清晰, 同时能够看到更多的细节信息。

图 7 为最终改进模型与 YOLOv5s 模型的训练过程的指标对比, train/CIoU\_loss、train/conf\_loss、train/class\_loss 分别表示训练集的位置坐标预测损失、置信

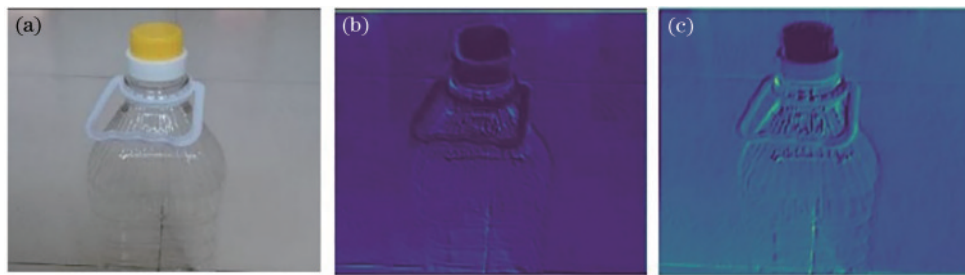


图 6 原图与模块获得的图像。(a) 原始图像; (b) CBS 模块获得的图像; (c) S-b 模块获得的图像

Fig. 6 Original image and image obtained by modules. (a) Original image; (b) image obtained by CBS module; (c) image obtained by S-b module

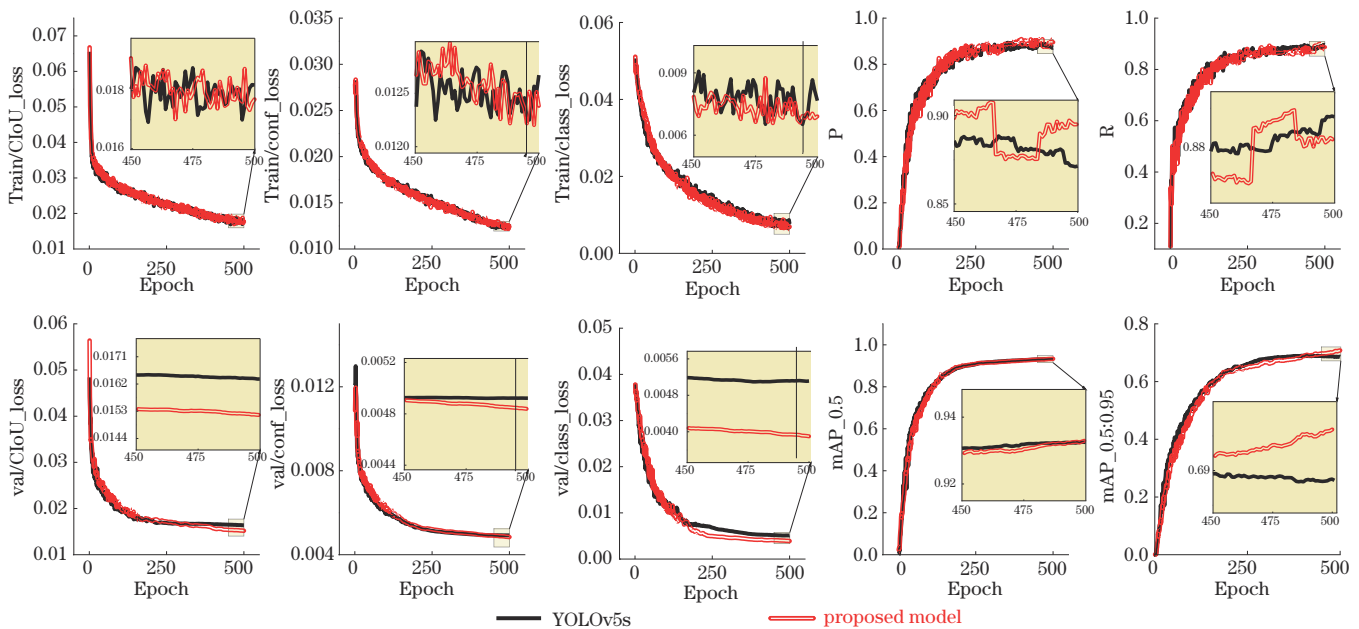


图 7 改进模型与 YOLOv5s 训练过程的指标对比

Fig. 7 Comparison of indicators between the improved model and YOLOv5s in training process

度预测损失和类别预测损失, val/CIoU\_loss、val/conf\_loss、val/class\_loss 分别表示验证集的位置坐标预测损失、置信度预测损失和类别预测损失。可以看出改进模型在收敛附近时, 训练集的三个损失曲线在波动中更低, 验证集的三个损失曲线更平缓, P 指标曲线在波动中更高, R 指标曲线在波动中稍低, mAP\_0.5 指标相近, mAP\_0.5:0.9 指标更高且还稍有上升趋势, 此时 YOLOv5s 模型已趋于平衡, 综合这些指标可以看出改进模型在众多指标性能上更具优势。

### 5.2 与其他常见模型对比实验

为了选择基准网络以及进行更为全面的对比, 选择一些常见其他架构的检测模型进行实验比较,

结果如表 4 所示。综合这些模型的规模与精度来看: 相比 Faster-RCNN (ResNet50)、YOLOv3、YOLOv4 模型, YOLOv5s 模型规模更小且精度更高; YOLOx-s 比 YOLOv5s 精度更高, 但是参数量更多和所需内存更大, 主要是因为 YOLOx-s 所需内存记录数据方式的差异, YOLOv5s 的所需内存为 YOLOx-s 的 40.00%; 改进的 YOLOv5s 模型的 P、R、mAP 稍差于 YOLOx-s 模型, 但在参数量和所需内存上小很多, 参数量仅为 YOLOx-s 的 30.00%, 更适用于在硬件配置差、体积小、功耗低的边缘终端中使用。总的来看, 所提改进模型考虑了精度与参数量的权衡, 具有优势。

表 4 与常见模型的对比实验

Table 4 Comparison experiments with common models

Model	P / %	R / %	mAP_0.5 / %	Parameter / 10 <sup>6</sup>	Memory / MB
Faster-RCNN(ResNet50)	68.50	94.15	92.32	28.3	108.48
YOLOv3	88.47	74.25	85.83	61.9	235.71
YOLOv4	87.83	85.60	88.50	63.9	244.48
YOLOv5s	89.40	85.20	92.10	7.09	13.70
YOLOx-s	92.13	92.53	94.40	8.94	34.29
Proposed model	90.12	89.95	94.01	2.68	5.34

### 5.3 部署在 Jetson Nano 上的速度对比实验

Jetson Nano 硬件配置: Jetson Nano 的整体架构是基于 ARM 架构芯片的, 结构紧凑、功耗低、成本低, 配备四核 ARM A57 @ 1.43 GHz CPU、128 核 Maxwell GPU 和 4 GB LPDDR4 RAM。

Jetson Nano 环境搭建: YOLOv5 是基于 PyTorch 框架开发的, 需要 PyTorch 1.6 以上版本, 为了满足需求, 根据 Nvidia 官方文档, 安装 JetPack SDK 4.4, 其提供了 Ubuntu 18.04 系统、CUDA 10.2、CUDNN 8.0、OpenCV 4.1.1; 安装 PyTorch 1.8.0 和 Torchvision 0.9.0, 其来源于 NVIDIA 官网提供的方式; 使用轻便的虚拟环境工具 Virtualenv 创建 YOLOv5s 的独立运

行环境, 其他的依赖包可采用普通方式安装, 官方系统默认交换分区 Swap 内存大小为 2G, 在安装一些软件或者运行较大规模的运算时, 常会有弹窗提醒 Swap 空间不足, 为了避免这种这个问题, 使用命令来增加 6G Swap 内存, 并开启最大功率模式, 此时功率为 10 W。

表 5 为不同模型在 Jetson Nano 上对不同图像的处理时间的对比。在图片分辨率为 640×640 时, 所提改进模型的前传输耗时比原始模型少 11.5%, 随着输入分辨率的减小, 从 640×640、512×512、416×416 到 320×320, 模型前传耗时逐渐接近, 但所提改进模型仍是最快的。

表 5 在 Jetson Nano 上的处理时间对比

Table 5 Comparison of processing time at Jetson Nano

Image size	Model	Preprocessing time / ms	Inference time / ms	NMS time / ms	FP time / ms
640×640	Yolov5s	1.6	182.3	10.1	194
	Proposed model	1.5	160.2	9.9	171.6
512×512	Yolov5s	1.1	123.3	9.2	133.6
	Proposed model	1.1	108.2	8.8	118.1
416×416	Yolov5s	0.9	82.9	8.0	91.8
	Proposed model	0.9	73.2	7.4	81.5
320×320	Yolov5s	0.8	56.0	6.5	63.3
	Proposed model	0.7	48.5	5.6	54.8

### 5.4 实验结果分析

综合结果分析: 主干的轻量化改进对主干特征的提取造成了一些不利的影响, 主要原因在于参数量下

降, 对一些特征不能给出更准确的抽象表达, 因此 mAP\_0.5:0.95 指标下降; 颈部网络的轻量化改进对结果的影响较小, 主要原因是将深度可分离卷积应用

于深层次特征上;从总体的改进上看,取得更优的效果,原因在于本轻量化的部分主要作用于 YOLOv5s 网络中一些复杂分支结构,大体上保持了原 YOLOv5s 架构的灵活性,使用的 ShuffleNet v2 基本单元与深度可分离卷积都为紧凑型的结构,使得总体的网络更加紧凑和高效,减少了过度提取重复特征的过程,降低网络的碎片化,有助于网络学习可回收垃圾特征和能更稳健地达到最优解。

## 6 结 论

针对可回收垃圾检测的精度低和模型参数量庞大的问题,首先解决没有数据集的问题,其次在原有常见的目标检测框架的基础上,通过对比实验,选择 YOLOv5s 为基础来进行进一步改进。通过在合适的位置上引入轻量化的高效单元模块,所提改进模型能够压缩参数量,也能保证有效的特征提取,比原始模型的精度更高,P、R、AP、mAP<sub>0.5</sub>、mAP<sub>0.5:0.95</sub> 的指标分别比原始模型高 0.75 个百分点、4.77 个百分点、1.92 个百分点、2.42 个百分点,最终模型大小仅为 5.34 MB,在 Jetson Nano 上具有比原始模型更快的运行速度,为后续的可回收垃圾系统的开发提供参考。由于垃圾的堆积,垃圾形态的破损、污垢、老化等问题,垃圾的分类与定位常常面临背景复杂和垃圾互相遮蔽的问题,接下来可以结合传送带分离技术,先将垃圾分散开,使用区分度高的背景环境,再结合机械部件构建系统,实现对垃圾更便捷的自动化回收处理。

### 参 考 文 献

- [1] 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.
- [2] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386-397.
- [3] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2022-02-04]. <https://arxiv.org/abs/1804.02767>.
- [4] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2022-01-02]. <https://arxiv.org/abs/2004.10934>.
- [5] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-07-18)[2022-02-04]. <https://arxiv.org/abs/2107.08430>.
- [6] 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.  
Li J Y, Chen X L, Zhang A H, et al. Survey of garbage classification methods based on deep learning[J]. *Computer Engineering*, 2022, 48(2): 1-9.
- [7] Prouença P F, Simões P. TACO: trash annotations in context for litter detection[EB/OL]. (2020-03-16)[2022-02-05]. <https://arxiv.org/abs/2003.06975>.
- [8] 刘齐锋. 基于光电图像识别的智能垃圾分类技术的研究[D]. 成都: 电子科技大学, 2021.  
Liu Q F. Research on intelligent garbage classification technology based on photoelectric image recognition[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [9] 李志军, 杨圣慧, 史德帅, 等. 基于轻量化改进 YOLOv5 的苹果树产量测定方法[J]. *智慧农业(中英文)*, 2021, 3(2): 100-114.  
Li Z J, Yang S H, Shi D S, et al. Yield estimation method of apple tree based on improved lightweight YOLOv5[J]. *Smart Agriculture*, 2021, 3(2): 100-114.
- [10] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993-13000.
- [11] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.
- [12] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet v2: practical guidelines for efficient CNN architecture design [M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11218: 122-138.
- [13] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [14] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2022-02-03]. <https://arxiv.org/abs/1704.04861>.
- [15] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [16] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1314-1324.
- [17] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [18] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.