融入注意力机制的轻量化可回收垃圾检测方法

郭洲^{1,2}, 黄诗浩^{1,2}, 谢文明², 吕晖^{1,2}, 张旋旋^{1,2}, 陈哲^{1,2}

(1.福建省汽车电子与电驱动技术重点实验室, 福州 350118;

2.福建工程学院 电子与电气物理学院, 福州 350118)

摘要:目的 针对目前智能垃圾分类设备使用的垃圾检测方法存在检测速度慢且模型权重文件较大等问题,提出一种基于 YOLOv4 的轻量化方法,以实现可回收垃圾的检测。方法 采用 MobileNetV2 轻量级 网络为 YOLOv4 的主干网络,用深度可分离卷积来优化颈部和头部网络,以减少参数量和计算量,提高检测速度;在颈部网络中融入 CBAM 注意力模块,提高模型对目标特征信息的敏感度;使用 K-means 算法重新聚类,得到适合自建可回收数据集中检测目标的先验框。结果 实验结果表明,改进后模型的参数 量减少为原始 YOLOv4 模型的 17.0%,检测的平均精度达到 96.78%,模型权重文件的大小为 46.6 MB,约为 YOLOv4 模型的 19.1%,检测速度为 20.46 帧/s,提高了约 25.4%,检测精度和检测速度均 满足实时检测要求。结论 改进的 YOLOv4 模型能够在检测可回收垃圾时保证较高的检测精度,同时具有较好的实时性。

关键词:可回收垃圾检测; MobileNetV2; YOLOv4; 注意力机制; 深度学习 中图分类号: TB487; TP391 文献标识码: A 文章编号: 1001-3563(2023)09-0243-11 DOI: 10.19554/j.cnki.1001-3563.2023.09.030

Lightweight Recyclable Garbage Detection Method Incorporating Attention Mechanism

GUO Zhou^{1,2}, HUANG Shi-hao^{1,2}, XIE Wen-ming², LYU Hui^{1,2}, ZHANG Xuan-xuan^{1,2}, CHEN Zhe^{1,2}

(1. Fujian Key Laboratory of Automotive Electronics and Electric Drive, Fuzhou 350118, China;
 2. School of Electronic, Electrical Engineering and Physics, Fujian University of Technology, Fuzhou 350118, China)

ABSTRACT: The work aims to propose a lightweight method based on YOLOv4 to detect recyclable garbage, so as to address the problems of slow detection speed and large model weight files in the current garbage detection methods used by smart garbage sorting devices. The MobileNetV2 lightweight network was used as the backbone network of YOLOv4 and the depth-separable convolution was used to optimize the neck and head networks to reduce the parameters and computation to accelerate detection. The CBAM attention module was incorporated into the neck network to improve the sensitivity of the model to the target feature information. The K-means algorithm was used to re-cluster to get suitable self-built recyclable data with a priori frame for focused detection of targets. The experimental results showed that: the parameters were reduced to 17.0% of the original YOLOv4 model. The detected mAP reached 96.78%. The model weight file size was 46.6 MB, which was about 19.1% of the YOLOv4 model weight file. The detection speed was 20.46 frames/s, which was improved by 25.4%. Both the detection accuracy and the detection speed met the real-time detection requirements. The improved YOLOv4 model can guar-

收稿日期: 2022-05-13

基金项目:国家自然科学基金(61604041);教育部产学研协同育人项目(201901021014);福建省教育厅基金项目 (JT180352)

作者简介:郭洲(1996—),男,硕士生,主攻机器视觉、图像处理。

通信作者:黄诗浩(1985-),男,博士,副教授,主要研究方向为光电信息材料与器件、机器视觉等。

antee high detection accuracy and good real-time performance in detection of recyclable garbage. **KEY WORDS:** recyclable garbage detection; MobileNetV2; YOLOv4; attention mechanism; deep learning

随着经济的快速发展,人们的生活水平进一步提 高,伴随而来的垃圾产出数量也在逐年增加,垃圾对 环境、人们的健康造成的影响也日益凸显,垃圾处理 面临着巨大的挑战。垃圾分类是减少垃圾处理量的一 种有效方式。由于目前垃圾的种类繁多,很难实现准 确分类。随着人工智能的快速发展及计算机算力的不 断提升,垃圾分类研究已经受到国内外学者的广泛关 注,尤其是对能高效分拣垃圾的智能垃圾分拣设备的 研究。比如,美国光学分类设备生产公司设计的 Max-AI 智能分类机器人,通过扫描物体形状来实现 分类,其准确率较高;日本的 FANUC 分拣机器人, 利用视觉分析系统对物品进行跟踪和分类;中国的 Picking AI 垃圾分类机器人,利用人工智能算法、机 器人控制等技术实现垃圾分拣。上述分拣机器人的分 拣功能较单一、占用面积较大、价格昂贵,因此无法 大规模应用。随着深度学习技术的快速发展,基于深 度学习模型解决的问题也越来越多。由此,可以将深 度学习技术应用于智能垃圾分拣设备,以解决当前垃 圾分类困难和人工分拣效率低等问题。

近年来,基于深度学习技术的目标检测算法发展 迅速[1]。目标检测算法主要分为两阶段算法和一阶段 算法。两阶段算法首先采用传统的选择性搜索 (Selective Search)及后来更新的区域生成网络 (Region Proposal Network, RPN) 生成候选区域, 然后对候选区域特征进行提取和分类,得到最终的检 测结果,其特点是精度较高,缺点是检测速度较慢, 以 Fast R-CNN (Fast Region-Convolutional Neural Network)^[2]、Faster R-CNN^[3]等算法为代表。一阶段 算法直接在整张图片中生成若干候选框,即可同时得 到目标的位置和类别信息。虽然一阶段算法的准确率 相对于两阶段算法较低,但其检测速度更快。目前, 通常将一阶段算法用于实时检测的场景,以 SSD (Single Shot Multibox Detector)^[4], YOLO(You Only Look Once)^[5-8]系列, 以及 RetinaNet^[9]等算法为代表。 其中, RetinaNet 提出的 Focal Loss 解决了一阶段算 法正负样本不均衡的问题,提高了一阶段算法的检测 精度。当前,深度学习目标检测方法在垃圾目标检测 方面开展了一系列研究。赵珊等^[10]提出了一种基于 IFPN+MobilenetV2-SSD 模型的垃圾实时分类检测方 法,使用 MobileNetV2 作为 SSD 的主干网络,加入 带有空洞卷积的空间金字塔池化模块,提高了模型的 检测精度和检测速度。马雯等^[11]提出了改进的 Faster R-CNN 垃圾目标检测模型,实验结果表明,与传统 Faster R-CNN 算法相比, 其平均精确度提高了 8.26%,综合识别率达到 81.77%。许伟等^[12]提出了一 种基于 YOLOv3 算法的轻量级垃圾目标检测算法,

能有效地对垃圾目标进行检测。李庆等^[13]基于 YOLOv4 提出了嵌入注意力机制的目标检测算法 Attn-YOLOv4,经实验验证,比原始 YOLOv4 算法 的平均精度(Mean Average Precision, mAP)提高了 0.16%,实现了对运动垃圾的快速稳定跟踪,在 20 mm 误差范围内达到 0.945 的精确度。Kumar 等^[14]建立了 一个垃圾数据集,共包含数量为 6317 张的垃圾图像, 在 YoLov4 模型中训练,平均精度达到 94.99%。

上述方法虽然在一定程度上有效地提高了垃圾 分类检测的精度和速度,但很多检测方法的参数过 多,导致内存占用较大,其检测精度和速度存在较大 的改进空间;存在不注意特定场景和标准数据集场景 之间差异的问题,不能达到与通用数据集相同的结 果;存在因垃圾目标样本中的小目标样本导致的正、 负样本不平衡,从而出现垃圾分类检测精度低的问 题。针对以上问题,文中提出了一种基于 YOLOv4 改进的轻量化可回收垃圾检测方法,通过优化网络模 型结构和减少模型参数来提升检测精度和检测速度, 使网络模型在满足检测精度的同时保证了检测速度。

1 相关网络模型及原理介绍

1.1 YOLOv4 网络模型

YOLOv4 是一种端到端的目标检测模型,在 YOLOv3 的基础上进行改进,并经过不断的模型优 化,模型的检测精度和速度达到了不错的水平。 YOLOv4 整体结构大致分为3个部分。

1)主干特征提取网络。YOLOv4 采用 CSPDarknet53 作为主干(Backbone),包含由 29 个卷积层堆叠而成的 5 组(Cross Stage Partial Network, CSPNet)模块^[15]。 CSPNet 结构可以增强卷积网络的学习能力,减少模型的 计算量。通过对主干网络进行特征提取,得到了 3 个有 效特征图,尺度分别为 13×13、26×26、52×52。不同尺 度的特征图包含不同维度的目标语义信息。

 2)颈部特征融合网络。包含空间金字塔池化 (Space Pyramid Pool, SPP)模块^[16]和路径聚合网络 (Path Aggregation Network, PANet)^[17]2个部分, SPP 网络对 Backbone 输出的 13×13 特征图进行了 1×1、5×5、9×9、13×13 等 4 种尺度的最大池化 (Maxpooling)操作,有效提高了网络的感受野。 PANet 相较于特征金字塔网络(Feature Pyramid Network, FPN),增加了一条自下而上的增强路径, 加强了对浅层信息的提取,提高了模型的检测精度。

3)头部预测网络。对特征融合网络输出的 3 个 不同大小特征图的信息进行解码,分别检测小、中、 大3个目标,在原图上输出检测目标的位置和类别。

1.2 MobileNetV2 网络结构

MobileNetV1 是 Google 团队于 2017 年提出的一 种轻量化的卷积神经网络模型, 轻量化的卷积网络与 标准卷积的区别在于使用了深度可分离卷积, 具备能 够在保持精度的同时减少网络参数和计算量的优点, 并且还引入了宽度 α 和分辨率 ρ 2 种缩放因子, 通过 调节 2 种缩放因子可以进一步缩小模型。在相同条件 下进行测试时, 虽然 MobileNetV1 与 VGG16 相比, 其准确率降低了 0.9%, 但参数量和计算量分别只有

VGG16的 $\frac{1}{32}$ 和 $\frac{1}{27}$ 。

深度可分离卷积将卷积分为 2 个过程: 深度卷 积,采用 3×3 的卷积核进行 DWConv 操作; 逐点卷 积,采用 1×1 的卷积核进行普通卷积操作。深度可分 离卷积操作在参数量和计算量上比标准卷积操作更 少,标准卷积和深度可分离卷积两者的卷积过程如图 1—2 所示。



图 2 深度可分离卷积 Fig.2 Deep separable convolution

对于输入为 $D_{\rm F} \times D_{\rm F} \times M$ 的特征图, $D_{\rm F}$ 为输入特征 图的高度或宽度, M 为通道数。对特征图进行卷积操 作,卷积核大小为 $D_{\rm K} \times D_{\rm K}$,卷积核的数量为 N,通道 数为 1,输出特征图的大小为 $D_{\rm G} \times D_{\rm G} \times N$, $D_{\rm G}$ 为输出 特征图的高度或宽度, N 为通道数。

对特征图进行标准卷积的过程中的计算量 C_1 和 参数量 P_1 如式 (1) - (2) 所示。

$$C_{1} = D_{F} \times D_{F} \times D_{K} \times D_{K} \times M \times N$$
(1)
$$P_{1} = D_{K} \times D_{K} \times M \times N$$
(2)

在对特征图进行深度可分离卷积过程中的计算 量 C_2 和参数量 P_2 如式(3)—(4) 所示。

$$C_2 = D_{\rm F} \times D_{\rm F} \times D_{\rm K} \times D_{\rm K} \times M + D_{\rm K} \times D_{\rm K} \times M \times N \quad (3)$$

$$P_2 = D_{\rm K} \times D_{\rm K} \times M \times N + D_{\rm K} \times D_{\rm K} \times N \tag{4}$$

深度可分离卷积与标准卷积计算量之比的计算 如式(5)所示。

$$\frac{C_1}{C_2} = \frac{D_F \times D_F \times D_K \times D_K \times M + D_K \times D_K \times M \times N}{D_F \times D_F \times D_K \times D_K \times M \times N} =$$

$$\frac{1}{N} + \frac{1}{D_F^2}$$
(5)

将卷积核大小设为 3×3、步长设为 1 时,相较于 普通卷积,其计算量下降为原来的 ¹/₉~¹/₈,使得模型 的检测速度得到进一步提高。

MobileNetV2 是在 MobileNetV1 的基础上引入具 有线性瓶颈的倒残差结构^[18]。残差模块先对特征图进 行压缩,然后再扩张,而倒残差结构与之相反,先对 特征图进行扩张,然后再压缩。MobileNetV2 有 2 种 倒残差模块,如图 3 所示。当 stride=1 时,特征图首 先通过一个 1×1 的卷积来提升通道维度,且激活函数 使用 ReLU6,然后通过一个 3×3 的深度卷积进行特 征提取,且激活函数使用 ReLU6,最后通过一个 1×1 的卷积来降低通道维度,并采用线性激活函数。为了 避免特征信息的损失,采用 Linear 线性激活函数,接 着将输出结果与输入进行 shortcut 拼接。当 stride=2 时,与 stride=1 时的原理差不多,唯一不同的是无 shortcut 拼接。因为输出和输入的特征图尺度不一样, 所以无法进行 shortcut 拼接。





CSPDarknet 53 与 MobileNetV2 各层计算量和参数量的对比如表 1 所示。2 种网络结构均使用了 2 种 大小为 3×3 和 1×1 的卷积核。在不考虑激活函数层、 BN 层和全连接层的影响下,设置输入图片的尺寸为 416×416×3,各层的计算量和参数量如表 1—2 所示。

| Tab.1 Computational and parametric quantities for each layer of CSPDarknet 53 | | | | | | | | |
|---|---------|-------------------|---------------------|---------------------|---------------------|--|--|--|
| Туре | Filters | Size | Output | 计算量/106 | 参数量/10 ³ | | | |
| Conv2d | 32 | 3×3 | 416×416×32 | 416×416×32 149.52 | | | | |
| Conv2d | 64 | 3×3/2 | 208×208×64 | 1 594.88 | 18.43 | | | |
| Conv2d | 32 | 1×1 | 208×208×32 | 2 | | | | |
| Conv2d | 64 | 3×3 | 208×208×64 1 417.67 | | 32.77 | | | |
| Resblock_body1 \times 1 | | | 208×208×64 | | | | | |
| Conv2d | 128 | 3×3/2 | 104×104×128 | 1 594.88 | 73.73 | | | |
| Conv2d | 64 | 1×1 | 104×104×64 | | | | | |
| Conv2d | 64 | 3×3 | 104×104×64 | 104×104×64 1 107.56 | | | | |
| Resblock_body2 \times 2 | | | 104×104×64 | | | | | |
| Conv2d | 256 | 3×3/2 | 52×52×256 | 1 594.88 | 294.91 | | | |
| Conv2d | 128 | 1×1 | 52×52×128 | | | | | |
| Conv2d | 128 | 3×3 | 52×52×128 | 3 765.70 | 1 392.64 | | | |
| Resblock_body3 \times 8 | | | 52×52×128 | | | | | |
| Conv2d | 512 | 3×3/2 | 26×26×512 | 1 594.88 | 1 179.65 | | | |
| Conv2d | 256 | 1×1 | 26×26×256 | | | | | |
| Conv2d | 256 | 3×3 | 26×26×256 | 3 765.70 | 5 570.56 | | | |
| Resblock_body4 \times 1 | | | 26×26×256 | | | | | |
| Conv2d | 1 024 | 3×3/2 | 13×13×1 024 | 1 594.88 | 4 718.59 | | | |
| Conv2d | 512 | 1×1 | 13×13×512 | | | | | |
| Conv2d | 512 | 3×3 | 13×13×512 | 1 993.61 | 11 796.48 | | | |
| Resblock_body5 \times 4 | | | 13×13×512 | | | | | |
| 合计 | | CSPDarknet53 网络结构 | | 20 174.16 | 25 181.02 | | | |

表 1 CSPDarknet 53 各层的计算量与参数量 Tab 1 Computational and parametric quantities for each layer of CSPDarkn

注: 3×3/2 表示步长为2的卷积,其余步长均为1。

| 表 2 | MobileNetV2 各层的计算量与参数量 |
|-----------------------|--|
| Tab.2 Computational a | nd parametric quantities for each layer of MobileNetV2 |

| Input | Operator | 升维倍数 t | 输出通道数 c | 重复次数 n | 步长 c | 计算量/10 ⁶ | 参数量/10 ³ |
|---------|---------------|--------|--------------|--------|------|---------------------|---------------------|
| 416×416 | Conv2d 3×3 | | 32 | 1 | 2 | 74.76 | 0.86 |
| 208×208 | bottleneck | 1 | 16 | 1 | 1 | 78.91 | 1.82 |
| 208×208 | bottleneck | 6 | 24 | 2 | 2 | 220.13 | 9.41 |
| 104×104 | bottleneck | 6 | 32 | 3 | 2 | 170.55 | 29.81 |
| 52×52 | bottleneck | 6 | 64 | 4 | 2 | 109.03 | 80.64 |
| 26×26 | bottleneck | 6 | 96 | 3 | 1 | 131.61 | 194.69 |
| 26×26 | bottleneck | 6 | 160 | 3 | 2 | 164.12 | 457.92 |
| 13×13 | bottleneck | 6 | 320 | 1 | 1 | 79.34 | 469.44 |
| 13×13 | Conv2d 1×1 | | 1 280 | 1 | 1 | 69.22 | 409.60 |
| 合计 | | Mobi | leNetV2 网络结构 | | | 1 097.67 | 1 654.19 |

2 YOLOv4 模型轻量化

2.1 主干网络优化

原始的 YOLOv4 模型存在参数过多、计算量大等缺点,因此为了减少参数量、提高检测速度,使减少参数量与提高检测速度这两者之间达到最优平衡, 文中采用参数量少的轻量级网络 MobileNetV2 来作 为模型 Backbone。由 MobileNetV2 替换 CSPDarknet53 重新构建的主干网络结构如图 4 所示,经过主干网络 特征提取后,获得了 3 个有效特征层,分别为 52×52×32、26×26×96、13×13×320,这 3 个有效特征 层将作为颈部网络的输入。





2.2 颈部网络和头部网络优化

为了进一步减少模型的参数,对颈部和头部网络 进行了优化。因为在颈部和头部网络中存在许多三次 和五次卷积块,并且在三次和五次卷积块中存在大量 步长为1的3×3卷积,同时在颈部网络中的二倍下采 样(DownSampling)模块中也存在步长为2的3×3 卷积,这都将造成大量的卷积运算,影响模型的推理 速度,所以将三次和五次卷积块及二倍下采样模块中 的 3×3 卷积修改为 3×3 的深度可分离卷积。同时,为 了使网络模型更加关注待检测目标的高层语义特征, 文中受到 CBAM 的启发,采用通道注意力模块来增 大感兴趣区域特征通道的权重,并且通过空间注意力 模块来关注感兴趣区域的空间位置,增大有意义特征 区域的权重,减少无效区域的权重^[19]。CBAM 模型 如图 5 所示。通道注意力模块的主要实现过程:首先 对输入的特征图 F 进行全局空间最大池化和平均池 化,得到2个维度为1×1×C的特征图,然后将它们 输入1个2层的共享神经网络(Multilayer Perceptron, MLP),得到2个特征向量,再将它们求和,通过激 活函数得到通道注意力权重参数 W_c(F),最后将权重 系数 $W_{c}(F)$ 与原特征图 F 相乘,得到新的特征图 F_{1} , 见式(6)—(7)。

 $W_{c}(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) (6)$ $F_{1} = W_{c}(F) \times F$ (7)

式中: $\sigma(\cdot)$ 为 sigmoid()函数; MLP(\cdot)为多层感知 机网络模型函数; AvgPool 为平均池化; MaxPool 为 最大池化。

空间注意力模块的主要实现过程:将特征图 F_1 作为本模块的输入特征图,首先进行全局空间最大池 化和平均池化,得到 2 个 $H \times W \times 1$ 的特征图,然后将 这 2 个特征图进行堆叠拼接,之后经过卷积核为 7×7 的卷积操作,再经过激活函数后得到空间注意力权重 参数 $W_s(F_1)$,最后将权重参数 $W_s(F_1)$ 与特征图 F_1 相 乘,得到最终的特征图 F_2 ,见式(8)—(9)。融入 具有双重注意力机制的颈部网络,如图 5 所示。

$$W_{s}(F_{1}) = \sigma(f^{3 \times 7}(\operatorname{AvgPool}(F_{1}); \operatorname{MaxPool}(F_{1})))$$
(8)
$$F_{2} = W_{s}(F_{1}) \times F$$
(9)



图 5 CBAM 结构 Fig.5 CBAM structure

模型的特征图可视化结果如图 6 所示,可知融入 CBAM 注意力机制后可以更好地覆盖目标区域,增加 目标区域的显著度,因此模型能够更好地学习目标区 域的特征。为了保证模型的检测精度和速度,最终在 颈部网络融入了3个CBAM模块。







a 原图 Original image

b 未融入CBAM c 融入CBAM Not integrated into CBAM

Integrate into CBAM

图 6 特征图可视化结果 Fig.6 Visualization results of feature map

经修改后模型的参数量得到大幅度减少,与原始

YOLOv4 模型参数量的对比见表 3。

表 3 模型参数的对比 **Tab.4** Comparison of model parameters

| 模型 | 参数量 |
|-----------------|------------|
| 原始 YOLOv4 模型 | 64 363 104 |
| 文中改进的 YOLOv4 模型 | 10 973 415 |

由表 3 可知,改进后 YOLOv4 模型的参数量为 10 973 415, 参数量仅为原始 YOLOv4 模型的约 17.0%,参数量的减少使得模型更加轻量化,从而加 快了模型的推理速度。改进后的模型整体结构如图 7 所示。

2.3 先验框重新聚类

原始 YOLOv4 模型中的先验框尺寸采用 K-means 算法^[20]在 PASCAL VOC 数据集聚类时得 到, PASCAL VOC 数据集中包含 20 类目标, 而自建 的可回收垃圾数据集只有5类目标,与PASCAL VOC 数据集中所包含的目标种类和数量都存在较大差异, 锚框尺寸不一定适合可回收垃圾的检测。为了得到更 加匹配的先验框^[21],采用 K-means 算法对自建的可 回收垃圾中 5 种类型标注框的宽高维度进行重新聚 类。K-means 算法的步骤: 首先随机选取 K 个初始



图 7 改进后的模型整体结构 Fig.7 Overall structure of the improved model

的聚类中心;其次,计算其他目标与聚类中心的距离, 根据距离度量形成新的 *K* 个簇,并重新调整聚类中 心;最后,通过循环迭代调整,使群中各个目标向各 聚类中心聚集,使群之间的距离变大。K-means 算法 通常以欧氏距离为距离度量,文中将用标注框和聚类 中心框的面积重叠度作为距离度量,距离度量计算见 式(10)—(11)。

$$d(b_i, b_c) = 1 - I_{\text{IOU}}(b_i, b_c)$$
(10)

$$I_{\rm IOU}(b_i, b_c) = \frac{(b_i) \cap (b_c)}{(b_i) \cup (b_c)}$$
(11)

式中: $d(b_i, b_c)$ 为第 i 个标注框到聚类中心 c 的距 离; $I_{IOU}(b_i, b_c)$ 为第 i 个标注框与聚类中心框的交并 比(Intersection over Union, IOU)^[22], 2 个框的重 合度越高, 交并比值越大。

检验 K-means 聚类生成的先验框的准确性常常 使用平均交并比(AvgIOU)进行评估,一般来说 AvgIOU 值越大,说明聚类算法生成的先验框越准确。 对已经标注好的可回收数据集进行聚类分析,得到 *K* 与 AvgIOU 之间的关系,如图 8 所示。



由图 8 可知, K=9 为 AvgIOU 曲线上的一个拐点; 在 K>9 时曲线变化的幅度非常小。由此,在考虑计 算速度和检测精度的情况下选取先验框的数量为 9, 经聚类选择的 9 个先验框分别为(146,132)、(177, 378)、(182,233)、(259,297)、(305,387)、(368, 227), (383, 299), (394, 359), $(400, 401)_{\circ}$

3 实验结果与分析

3.1 实验数据集及评价指标

文中实验使用的数据集来自自建的可回收垃圾 数据集,该数据集共有 5 类标签,分别为废纸类 (cardboard)、玻璃类(glass)、塑料类(plastic)、金 属类(metal)、纺织类(textile)。这里分别使用水平 翻转、亮度调整、添加噪声、随机裁剪和随机旋转等 5 种数据扩充方法对可回收垃圾数据集进行样本扩 充,扩充后的图片总数量为 5 048 张。采用标注工具 LabelImg 软件对可回收垃圾目标进行标注,各类可回 收垃圾图片样本数量如表 4 所示,数据集样例见图 9。 训练集、测试集和验证集的划分比例为 8:1:1。

在对实验结果进行分析时,主要利用模型的精确 率(Precision, P)、召回率(Recall, R)、平均准确 率均值(P_{mAP})和每秒传输帧数(Frames Per Second, FPS)这4个参数对网络模型的检测效果进行评价。 各种指标的计算方法如式(12)—(14)所示。

$$P = \frac{T_{\rm p}}{T_{\rm p} + F_{\rm p}} \tag{12}$$

$$R = \frac{T_{\rm p}}{F_{\rm p} + T_{\rm N}} \tag{13}$$

$$P_{\text{mAP}} = \frac{1}{N} \sum_{i=1}^{N} [P(R)d(R)]_i$$
(14)

式中: *T*_P 为检测结果与实际目标结果一致的样本 总数目; *F*_N 为存在实际目标但未被检测出的实际目 标结果,即漏检样本的总数目; *F*_P 为检测结果与实 际目标结果不一致的样本总数目; *T*_N 为不存在实际 目标且也未被检测出实际目标结果的样本总数目。

平均准确率均值为多个类别的平均准确率 (Average Precision, AP)求和后再取平均值。AP 表示以召回率为横轴,以精确率为纵轴,所绘制的 P-R曲线的面积。FPS表示在模型检测速度时每秒钟 能够处理的图片数量。

3.2 模型训练

1

在训练过程中涉及的具体软硬件环境:操作系统为 Windows 10 中文版, CPU 型号为 Intel Core i5-10200H

| Tab.4 Number of each sample category | | | | | | | |
|--------------------------------------|--------------|---------|-----------------|------------------|-----------------|--|--|
| 类别 | 废纸类 | 玻璃类 | 塑料类 | 金属类 | 纺织类 | | |
| 数量 | 976 | 957 | 1 042 | 1 128 | 945 | | |
| 包括 | 纸箱、信封、 纸盒 | 玻璃瓶、玻璃杯 | 塑料瓶、塑料盆、 塑料桶 | 易拉罐、铁锅、金属电 饭煲 | 衣服、背包、床单、 被单 | | |
| 标注类别 | cardboard | glass | plastic | metal | textile | | |

表 4 各类别样本数量 ab.4 Number of each sample category



图 9 收集的可回收垃圾数据集样例 Fig.9 Sample of collected recyclable garbage dataset

2.40 GHz, GPU 型号为 Nvidia GeForce GTX1650Ti, 内存为 16 GB, 显存为 4 GB, 深度学习框架选用 Pytorch1.7, 加速库为 Cuda11.2、Cudnn11.2。

在训练过程中,将输入图片尺寸设置为 416×416。在模型训练时,将momentum(动量系数) 设为0.9,初始learning_rate(学习率)设为0.001, 总训练代数为100个epoch。训练时分为2个阶段, 先进行60个epoch冻结训练,再进行40个epoch解 冻训练。冻结阶段具体为冻结主干网络,因为该阶段 网络占用的显存较少,此时设置batch_size为8。解 冻阶段具体为不冻结主干网络,采用Adam优化学习 率,将学习率降至0.0001。该阶段网络的参数变多, 占用的内存较大,此时设置 batch_size 为 4,置信度 为 0.5。

3.3 实验结果分析

为了验证 MobileNetV2、K-means 算法和 CBAM 模块对模型检测性能的影响,使用相同的训练数据集 和测试数据集,相同的迭代次数和学习速率,对优化 的模块进行消融实验,实验结果见表 5—6。

由表 5 可知,实验 1 是将 YOLOv4 的主干网络 替换成 MobileNetV2, 虽然 mAP 值只有 85.56%, 但 其检测速度最快,达到了 22.23 帧/s; 实验 2 和实验 3 是在原始 YOLOv4 的基础上分别使用 K-means 算法 优化先验框和在颈部网络融入3个 CBAM 模块,可 知这 2 种方法都有助于提高模型的 mAP 值, 但会使 模型的检测速度降低,尤其是加入 CBAM 模块后其 检测速度下降得最为明显,原因是融入 CBAM 模块 后,模型的计算量相对增加;实验4结合了3种改进 方法, 该模型的 mAP 值达到了 96.78%, 相较于实验 1 提高了约 13.1%, 检测速度为 20.46 帧/s, 相较于实 验2、3提高幅度较大。由表6可知,结合了3种改 进方法的模型的 AP 值都达到 95%以上,说明 K-means 聚类算法重新聚类先验框,提高了模型的检 测精度, CBAM 的注意力机制有效地抑制了干扰信 息,提高了模型对目标特征信息的敏感度。

为了验证文中改进模型在可回收垃圾检测中的 性能,选取了目前一些主流目标检测模型 (Faster-RCNN、YOLOv4、SSD、YOLOv5s)与文 中的改进YOLOv4模型在相同的数据集上进行训练, 性能的对比结果见表 7。

| 实验序号 | MobileNetV2 | K-means 算法 | CBAM | mPrecision/% | mRecall/% | mAP/% | FPS/ (帧·s ⁻¹) |
|------|--------------|---------------|--------------|--------------|-----------|-------|------------------------------|
| 1 | \checkmark | | | 89.08 | 82.94 | 85.56 | 22.23 |
| 2 | | \checkmark | | 94.98 | 88.55 | 92.65 | 16.84 |
| 3 | | | \checkmark | 96.79 | 90.26 | 94.29 | 15.87 |
| 4 | \checkmark | \checkmark | \checkmark | 97.95 | 92.34 | 96.78 | 20.46 |

表 5 消融实验检测效果 Tab.5 Detection results of ablation experiment

注: 打钩表示采用了该方法。

表 6 各模块对模型的 AP 值对比 Tab.6 Comparison of AP values of each module to model

| 实验序号 | AP 值/% | | | | | | | |
|------|--------|---------|-----------|-------|---------|--|--|--|
| | glass | textile | cardboard | metal | plastic | | | |
| 1 | 88 | 84 | 91 | 81 | 84 | | | |
| 2 | 92 | 87 | 95 | 93 | 96 | | | |
| 3 | 92 | 92 | 96 | 95 | 97 | | | |
| 4 | 96 | 95 | 99 | 96 | 98 | | | |

由表 7 可知,两阶段检测网络 Faster-RCNN 的 检测精度相对较高,但其模型体积也相对较大,检测 速度最低仅为 5.54 帧/s,难以满足可回收垃圾检测实 时性的要求。虽然 SSD 模型的检测精度最低,但其 模型文件大小和检测速度都优于 Faster-RCNN 模型。 原始 YOLOv4 模型的 mAP 值为 91.65%,权重文件的 大小为 244.48 MB,比其他模型大。虽然 YOLOv5s 在当前模型中的检测速度最高且模型文件最小,但是 其检测精度却不高,比原始 YOLOv4 模型低 1.6%。 文中的改进模型与原始 YOLOv4 模型相比,其 mAP 值提高了 5.6%,模型权重文件大小为 46.6 MB,相较 于 YOLOv4 大幅减少,仅为 YOLOv4 的 19.1%,检 测速度为 20.46 帧/s。相较于 YOLOv4,提高了约 25.4%,检测速度和精度均满足实时性需求。

文中的改进模型与其他检测模型之间的检测效 果对比如图 10 所示。由对比检测结果可知,虽然各 模型均能检测出可回收垃圾,但是在检测结果的置信 度值和拟合度上,文中的改进模型优于其他模型。

| 齿刑 | | AP 值/% | | | | | 权重文件 | FPS/ |
|-------------|-------|---------|-----------|-------|---------|------------|--------|----------------------|
| 快空 | glass | textile | cardboard | metal | plastic | - IIIAF/70 | 大小/MB | (帧·s ⁻¹) |
| Faster RCNN | 88 | 92 | 98 | 95 | 96 | 93.64 | 108.33 | 5.54 |
| YOLOv4 | 88 | 88 | 93 | 96 | 93 | 91.65 | 244.48 | 16.32 |
| SSD | 95 | 92 | 91 | 84 | 84 | 89.19 | 92.65 | 16.04 |
| YOLOv5s | 89 | 86 | 92 | 90 | 93 | 90.17 | 27.1 | 27.87 |
| Ours | 96 | 95 | 99 | 96 | 98 | 96.78 | 46.6 | 20.46 |

表 7 各模型的对比结果 Tab.7 Comparison results of models



图 10 文中的改进模型与其他模型检测效果的对比

Fig.10 Comparison of detection result between the improved model and other models

4 结语

针对目前垃圾检测方法存在的检测速度慢且权 重文件较大等问题,提出了一种改进的 YOLOv4 检 测方法。为了减少模型的计算量和参数量,将 YOLOv4 的主干网络替换为 MobileNetV2,并使用深 度可分离卷积对网络进行优化;融入 CBAM 注意力 机制,提高了模型对目标特征信息的敏感度,抑制了 干扰信息,从而提升了模型的检测精度。为了得到适 合数据集的先验框,采用 K-means 算法对自建可回 收垃圾数据集进行重新聚类。实验结果表明,参数量 和模型权重文件大小分别减小为原始 YOLOv4 模型 的17.0%和19.1%,检测精度为96.78%,提高了 5.6%, 检测速度为 20.46 帧/s,提高了 25.4%。未来应进一 步对模型进行优化,在保证检测速度的同时提高模型 的检测精度。

参考文献:

 [1] 鞠默然,罗海波,王仲博,等.改进的 YOLOV3 算法 及其在小目标检测中的应用[J].光学学报,2019, 39(7):253-260.

JU Mo-ran, LUO Hai-bo, WANG Zhong-bo, et al. Improved YOLOV3 Algorithm and Its Application in Small Target Detection[J]. Acta Optica Sinica, 2019, 39(7): 253-260.

- [2] GIRSHICK R. Fast R-CNN[C]// IEEE International Conference on Computer Vision (ICCV), 2016: 1440-1448.
- [3] REN Shao-qing, HE Kai-ming, 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.
- [4] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector[C]// European Conference on Computer Vision, 2016: 21-37.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[EB/OL]. (2015-06-08)[2022-03-15]. https://arxiv.org/abs/1506.02640
- [6] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.
- [7] REDMON J, FARHADI A. Yolov3: An Incremental Improvement[EB/OL]. (2018-04-08)[2021-01-15]. https:// arxiv.org/1804.02767.
- [8] BOCHKOVSKIY A, WANG Chien-yao, LIAO H Y M.

YOLOv4: Optimal Speed and Accuracy of Object Detection[EB/OL]. (2020-04-23)[20220-03-15]. https://arxiv.org/ abs/2004.10934

- [9] LIN Tsung-yi, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]// IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017: 2999-3007.
- [10] 赵珊,刘子路,郑爱玲,等. 基于 MobileNetV2 和 IFPN 改进的 SSD 垃圾实时分类检测方法[J]. 计算机 应用, 2022, 42(S1): 106-111.
 ZHAO Shan, LIU Zi-lu, ZHENG Ai-ling, et al. Real-time Classification and Detection Method of Garbage Based on SSD Improved with MobileNetV2 and IFPN[J]. Journal of Computer Applications, 2022, 42(S1): 106-111.
 [11] 马雯, 于炯, 王潇,等. 基于改进 Faster R-CNN 的垃
- 城, 王城, 守, 坐了仗近Taster R Child Hage 坂检测与分类方法[J]. 计算机工程, 2021, 47(8): 294-300.
 MA Wen, YU Jiong, WANG Xiao, et al. Garbage Detection and Classification Method Based on Improved Faster R-CNN[J]. Computer Engineering, 2021, 47(8):
- 294-300.
 [12] 许伟,熊卫华,姚杰,等. 基于改进 YOLOv3 算法在 垃圾检测上的应用[J]. 光电子·激光, 2020, 31(9): 928-938.
 XU Wei, XIONG Wei-hua, YAO Jie, et al. Application of Garbage Detection Based on Improved YOLOv3 Algorithm[J]. Journal of Optoelectronics·Laser, 2020, 31(9): 928-938.
- [13] 李庆, 龚远强, 张玮, 等. 用于智能垃圾分拣的注意 力 YOLOv4 算法[J]. 计算机工程与应用, 2022, 58(11): 260-268.
 LI Qing, GONG Yuan-qiang, ZHANG Wei, et al. Attention YOLOv4 Algorithm for Intelligent Waste Sorting[J]. Computer Engineering and Applications, 2022, 58(11): 260-268.
- [14] KUMAR S, YADAV D, GUPTA H, et al. A Novel YO-LOv3 Algorithm-based Deep Learning Approach for Waste Segregation: Towards Smart Waste Management[J]. Electronics, 2020, 10(1): 14.
- [15] WANG Chien-yao, LIAO H Y M, WU Yueh-hua, et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 1571-1580.
- [16] HE Kai-ming, ZHANG Xiang-yu, REN Shao-qing, et al.

Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

- [17] LIU Shu, QI Lu, QIN Hai-fang, et al. Path Aggregation Network for Instance Segmentation[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [18] SANDLER M, HOWARD A, ZHU Meng-Long, et al. Mobilenetv2: Inverted Residuals and Linear Bottlenecks[EB/OL]. (2018-12-16)[2022-03-15]. https://arxiv.org/abs/1801.04381.
- [19] 张宸嘉,朱磊,俞璐. 卷积神经网络中的注意力机制 综述[J]. 计算机工程与应用, 2021, 57(20): 64-72.
 ZHANG Chen-jia ZHU Lei, YU Lu. Review of Attention Mechanism in Convolutional Neural Networks[J].
 Computer Engineering and Applications, 2021, 57(20): 64-72.

- [20] SINAGA K P, YANG Min-shen. Unsupervised K-means Clustering Algorithm[J]. IEEE Access, 2020(8): 80716-80727.
- [21] 王子鹏,张荣芬,刘宇红,等.面向边缘计算设备的 改进型 YOLOv3 垃圾分类检测模型[J].激光与光电子 学进展,2022,59(4):291-300.
 WANG Zi-peng, ZHANG Rong-fen, LIU Yu-hong, et al. Improved YOLOv3 Garbage Classification and Detection Model for Edge Computing Devices[J]. Laser & Optoelectronics Progress, 2022, 59(4): 291-300.
- [22] 董豪,李少波,杨静,等. 基于 YOLOv4 算法的药用 空心胶囊表面缺陷检测方法[J]. 包装工程, 2022, 43(7): 254-261.

DONG Hao, LI Shao-bo, YANG Jing, et al. Surface Defect Detection Method for Pharmaceutical Hollow Capsules Based on YOLOv4 Algorithm[J]. Packaging Engineering, 2022, 43(7): 254-261.

责任编辑: 彭颋