

用于智能垃圾分选的轻量级检测算法^①



王 林, 刘靖贇

(西安理工大学 自动化与信息工程学院, 西安 710048)

通信作者: 刘靖贇, E-mail: 1505537593@qq.com

摘 要: 为实现垃圾分选自动化, 确保垃圾正确分类, 提出了一种基于 YOLOv4 的轻量级垃圾检测算法. 算法对 YOLOv4 中的主干网络 CSPDarknet53, 使用层级调整后的 MobileNetV3 网络进行替换, 使得网络架构更适用于 YOLOv4 网络, 并提升网络的检测速度; 同时结合 Ghost 模块和 MobileNeXt 网络结构思想, 设计了一种全新的 bottleneck, 用以替换主干网络中的 bottleneck, 以提升模型的检测精度; 接着在主干网络中添加大残差边结构, 以提升网络的检测精度; 然后在颈部网络之前添加 CA (coordinate attention) 注意力机制, 进一步提升网络的检测精度; 最后为避免 K-means 算法在聚类过程中陷入局部极值, 使用二分 K-means 算法对垃圾检测数据集进行 anchor box 的重新聚类. 实验结果表明, 重新设计的网络与 YOLOv4 网络的 mAP 值相近, 但参数量减少了 89%, 检测速度提升了 51%, FPS 值达到了 67.5 (on NVIDIA GeForce RTX 3060), 可实现部署到算力和内存较低的嵌入式设备中.

关键词: 垃圾检测; YOLOv4; MobileNetV3; 大残差边; 注意力机制; 嵌入式设备

引用格式: 王林, 刘靖贇. 用于智能垃圾分选的轻量级检测算法. 计算机系统应用, 2023, 32(4): 231-240. <http://www.c-s-a.org.cn/1003-3254/9060.html>

Lightweight Detection Algorithm for Intelligent Garbage Sorting

WANG Lin, LIU Jing-Yun

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: To achieve automatic garbage sorting and ensure correct garbage classification, this study proposes a lightweight garbage detection algorithm based on you only look once version 4 (YOLOv4). The algorithm replaces CSPDarknet53, the backbone network in YOLOv4, with a MobileNetV3 network after level adjustment, thereby making the network architecture more suitable for the YOLOv4 network and improving the detection speed of the network. Subsequently, a new bottleneck is designed by drawing on the idea of the Ghost module and the MobileNeXt network structure to replace the bottleneck in the backbone network and boost the detection accuracy of the model. Then, a large residual edge structure is added to the backbone network to raise the detection accuracy of the network. Furthermore, a coordinate attention (CA) mechanism is added before the neck network to further enhance the detection accuracy of the network. Finally, to prevent the K-means algorithm from falling into a local extremum in the clustering process, this study employs the bisecting K-means algorithm to re-cluster anchor boxes for the garbage detection dataset. Experimental results show that the redesigned network achieves a mean average precision (mAP) close to that of the YOLOv4 network but reduces the number of parameters by 89%. Moreover, it improves detection speed by 51% to 67.5 FPS (on NVIDIA GeForce RTX 3060). It can thus be deployed to embedded devices with low computing power and memory.

Key words: garbage detection; YOLOv4; MobileNetV3; large residual edge; attention mechanism; embedded device

① 基金项目: 陕西省科技计划重点项目 (2017ZDCXL-GY-05-03)

收稿时间: 2022-09-19; 修改时间: 2022-10-27; 采用时间: 2022-11-16; csa 在线出版时间: 2023-02-10

CNKI 网络首发时间: 2023-02-13

垃圾是放错地方的资源, 未经分类的垃圾直接掩埋或焚烧, 造成的危害是难以想象的. 我国早在 2017 年就提出了《生活垃圾分类制度实施方案》^[1], 将垃圾分为了 4 大类进行分类处理, 以减轻环境污染. 在垃圾分类过程中, 我们可以在垃圾产生源头和在垃圾中端处理厂对垃圾进行分类, 但由于现阶段居民垃圾分类意识不强, 源头垃圾不能得到有效分类, 故需要在垃圾分选站对错分的垃圾进行二次分类. 目前中端垃圾分选站多依赖于人工分选, 在分拣精度和速度上都不够高, 同时恶劣的环境对工作者的身心健康影响巨大, 故分选自动化是垃圾分选站未来的发展趋势^[2].

针对中端垃圾分选的实际场景, 本文设计了一套适用于垃圾分选站的自动化分选系统. 首先通过工业摄像机拍摄传送带上的垃圾, 将视频传送到检测设备上; 然后通过 OpenCV 自带的接口读取视频中的每一帧图片, 将图片传递给经 TensorFlow 推理框架部署过的检测模型, 模型进行推理得到目标检测结果; 最后检测设备将检测结果中的物品类别和位置信息下发至机械臂, 机械臂进行垃圾夹取操作, 具体操作流程如图 1 所示. 现有的图像传输和机械臂控制技术足以满足上述要求, 故在自动化分选中最需要解决的是垃圾种类和位置的检测.

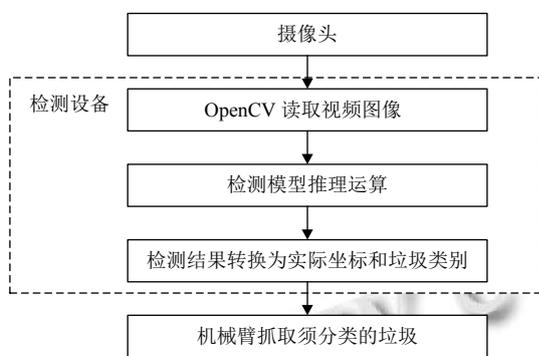


图 1 中端垃圾分选系统工作流程图

随着深度学习的高速发展, 各类目标检测算法被应用在垃圾检测中. 董霄霄等人^[3]基于 YOLOv3 提出了一种 RP-YOLO 的垃圾检测算法, 该算法使用 RepVGG 网络去替换 YOLOv3 的骨干网络, 并加入空洞卷积, 实现了复杂场景下的垃圾检测. 耿丽婷等人^[4]以 SSD 网络为基础, 更换了轻量级的主干网络, 并引入 SK 模块, 使网络在保证检测精度的前提下具有较快的检测速度. 魏铨磊等人^[5]在 YOLOv3 中引入多尺度感受野模块, 提高了网络的检测精度, 并更改了主干网络, 使得最终

的网络能够对密集垃圾有较好的检测效果.

本文使用基于深度学习的目标检测算法实现垃圾的精准检测. 将目标检测算法运用在实际场景中, 最重要的就是要保证算法的检测精度和速度与实际场景相匹配. 由于垃圾分选站恶劣的工作环境和设备连续工作的机制, 传统的 PC 电脑无法适应在这种环境中, 故算法需要移植到嵌入式设备中. 但由于嵌入式设备算力较为低下, 故检测算法需在保证检测精度的前提下具备较快的推理速度. 本文以 YOLOv4 目标检测网络为基础, 搭建了一个新的检测网络, 该网络可以更好地移植到算力和内存较低的嵌入式设备上.

1 背景知识

1.1 YOLOv4 算法简介

YOLOv4 网络在 YOLOv3^[6] 的基础上对其主干网络、颈部网络和损失函数等多方面进行改进, 在保证检测速度的前提下, 极大地提升了检测精度, 同时对小目标问题有较好的检测能力. YOLOv4 网络主要由 3 大部分组成, 分别是进行特征提取的 CSPDarknet53 主干网络, 进行特征融合的颈部网络 SPP 网络和 PAN 网络, 以及用于目标检测的检测头网络, 其具体的网络结构如图 2 所示. 但在垃圾检测中, YOLOv4^[7] 因其网络参数量较大、检测速度较慢, 不适合移植在嵌入式设备上, 本文预对 YOLOv4 网络进行改进, 在保证网络精度的前提下, 尽可能压缩网络参数量、提升检测速度.

1.2 MobileNetV3 简介

MobileNetV3^[8] 是一个轻量级网络, 适合部署在嵌入式设备和算力有限的移动设备上. MobileNetV3 由 1×1 普通卷积和深度可分离卷积堆叠形成的倒残差结构搭建而成, 其部分倒残差结构中添加 SE 注意力机制以提升网络的特征提取能力, 具体的网络结构如图 3 所示. 本文将以 MobileNetV3 为基础, 重新设计 YOLOv4 中的主干网络, 使得检测网络更为轻量化.

1.3 Ghost 模块简介

Ghost 模块^[9] 由华为在 2020 年提出, 其本质是在特征提取过程中, 首先使用 1×1 卷积对输入特征通道数进行压缩, 然后利用逐层卷积对其进行特征提取, 最后将前两步得到的特征层进行堆叠, 完成特征提取, 其特征提取过程如图 4 和图 5 所示. 实验发现这种特征提取方式在避免提取重复特征的同时, 大大降低了计算量, 加快了模型的推理速度.

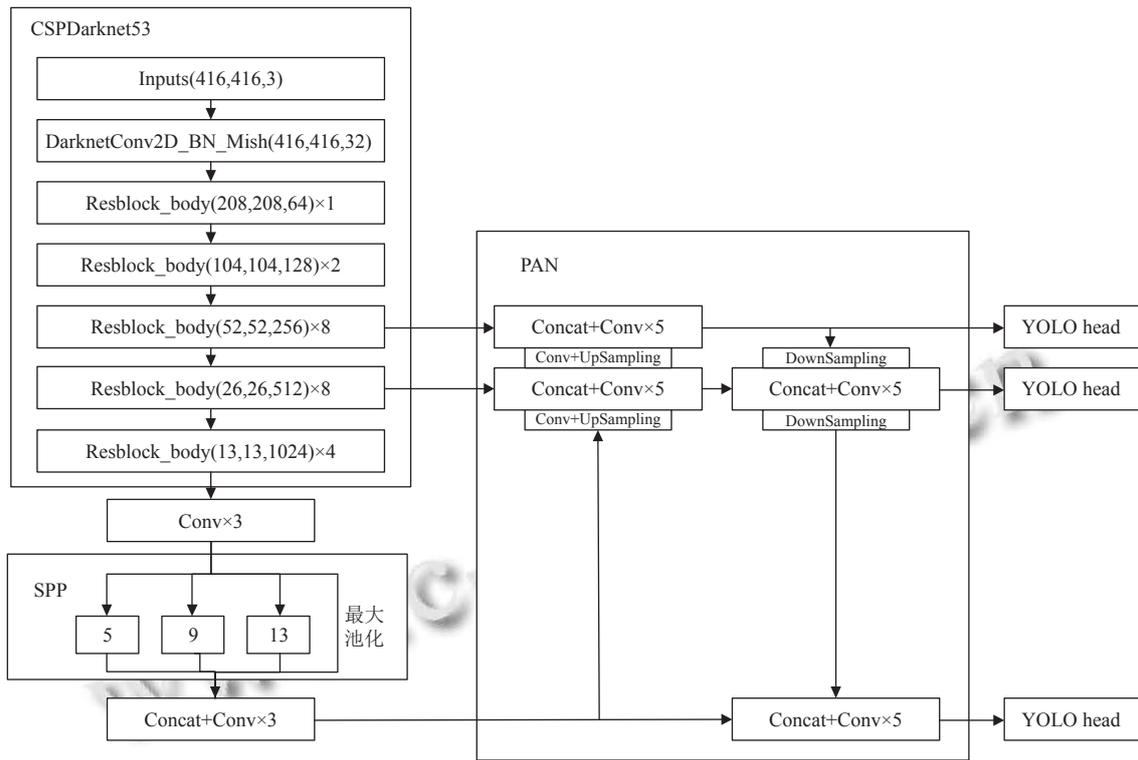


图2 YOLOv4网络结构图

Input	Operator	Exp size	Out	s
416 ² ×3	Conv2d	—	16	2
208 ² ×16	Bneck, 3×3	16	16	1
208 ² ×16	Bneck, 3×3	64	24	2
104 ² ×24	Bneck, 3×3	72	24	1
104 ² ×24	Bneck, 5×5	72	40	2
52 ² ×40	Bneck, 5×5	120	40	1
52 ² ×40	Bneck, 5×5	120	40	1
52 ² ×40	Bneck, 3×3	240	80	2
26 ² ×80	Bneck, 3×3	200	80	1
26 ² ×80	Bneck, 3×3	184	80	1
26 ² ×80	Bneck, 3×3	184	80	1
26 ² ×80	Bneck, 3×3	480	112	1
26 ² ×112	Bneck, 3×3	672	112	1
26 ² ×112	Bneck, 5×5	672	160	2
13 ² ×160	Bneck, 5×5	960	160	1
13 ² ×160	Bneck, 5×5	960	160	1

图3 MobileNetV3网络结构

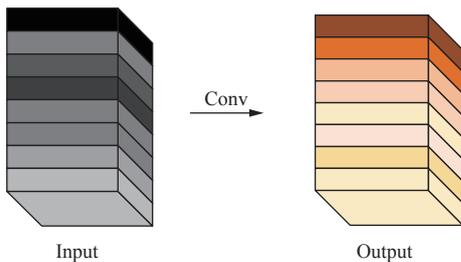


图4 传统卷积模块示意图

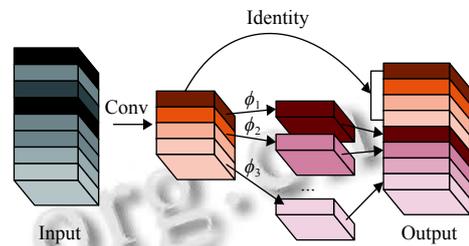


图5 Ghost模块示意图

2 本文工作

在 YOLOv4 中, 原始的骨干网络 CSPDarknet53 为一个中量级的网络, 其大量使用 3×3 普通卷积, 导致参数数量和网络架构较为庞大, 并不适合部署在资源有限的嵌入式设备上. 而 MobileNetV3^[10] 中使用深度可分离卷积代替传统的卷积模块, 在保证特征提取精度的前提下, 减少了模型的参数数量, 提升了模型的推理速度. 基于以上, 本文首先将 YOLOv4 中的 CSPDarknet53 网络替换为 MobileNetV3 网络; 同时对 MobileNetV3 在网络层级上进行调整, 使得主干网络能更好地匹配 YOLOv4 检测网络^[11]; 接着设计了一种全新的 bottleneck 结构替换掉层级调整后的 MobileNetV3 中的

bottleneck 结构; 然后在主干网络的第 3、4、5 层级上添加大残差边结构; 最后在颈部网络之前添加 CA 注

意力机制模块, 构建更适合于嵌入式设备的检测网络, 网络总体框图如图 6 所示。

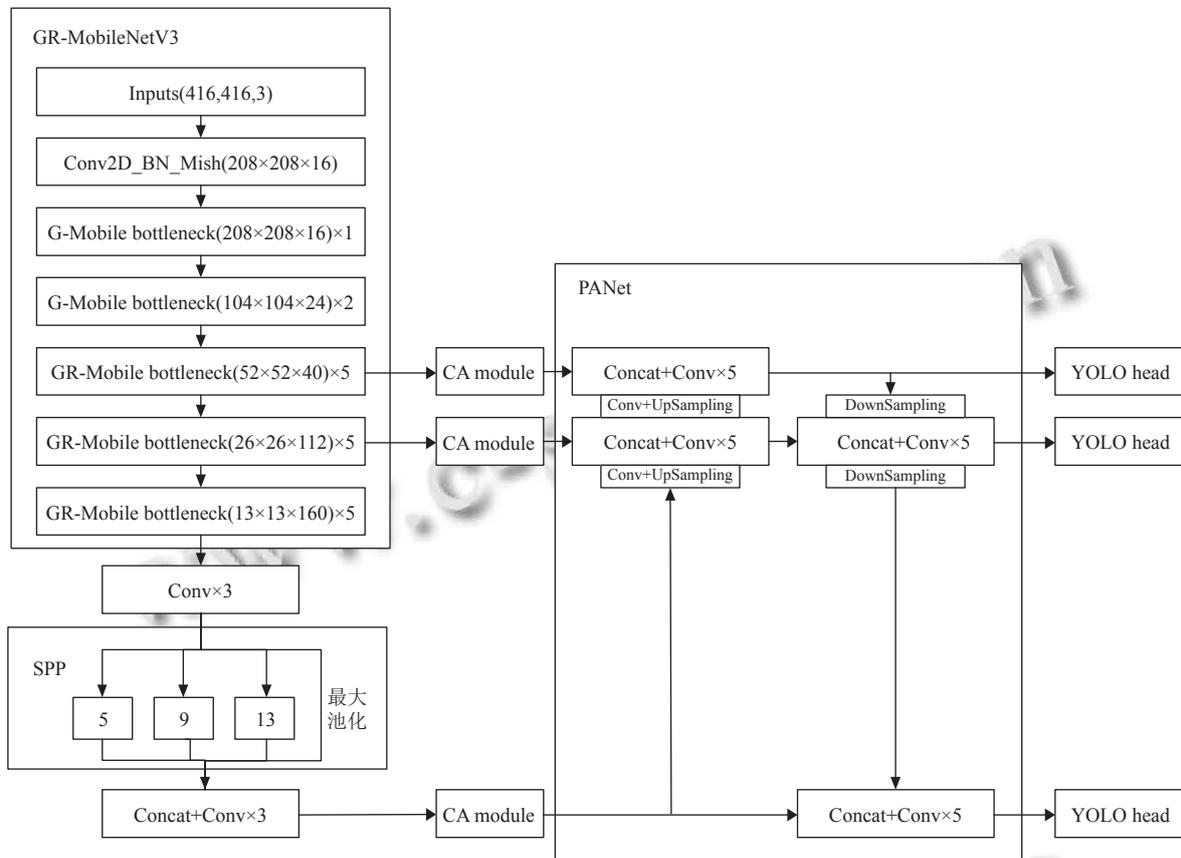


图 6 本文检测网络总体框图

2.1 改进的主干网络

2.1.1 层级均衡的 MobileNetV3 网络

MobileNetV3 网络主要由 5 大层组成, 分别对应提取到的 5 种不同的 feature map. 对 416×416 的输入图片, YOLOv4 使用 3 个检测头对其进行检测, 故本文选取 MobileNetV3 的后 3 层作为特征提取的输出. 输出 feature map 为 52×52 和 13×13 的第 3 层和第 5 层仅有 2 个 bottleneck 进行特征提取, 会造成图片特征提取不均衡, 影响检测精度, 故本文为这两层各增加 3 个 5×5 的 bottleneck, 达到图片特征提取均衡的效果, 具体网络结构如图 7 所示。

2.1.2 基于 Ghost 模块的 G-Mobile bottleneck

MobileNetV3 中的深度可分离卷积相较于传统卷积在模型参数数量和推理速度上有了很大的提升^[12], 但研究发现深度可分离卷积在特征提取过程中会出现特征重复, 重复的特征可以由其余特征经过线性变换得

到. 针对上述研究, 发现通过少量的卷积和堆叠的方式可得到相较于深度可分离卷积参数量更小, 速度更快的 Ghost 模块^[13], 其计算量仅为深度可分离卷积模块的 1/s, 具体论证结果如式 (1) 所示:

$$r_s = \frac{n \times h' \times w' \times c \times k \times k}{n/s \times h' \times w' \times c \times k \times k + (s-1) \times n/s \times h' \times w' \times d \times d} = \frac{c \times k \times k}{1/s \times c \times k \times k + (s-1) \times (s-1)/s \times d \times d} \approx \frac{c \times s}{s + c - 1} \approx s \quad (1)$$

其中, 分子为普通卷积的复杂度, 分母为 Ghost 模块的复杂度, $n \times h' \times w'$ 为输出的特征尺寸, s 代表每个通道产生的总映射, c 为输入特征图且数值一般较大, n/s 指经过普通卷积输出的特征图数, $k \times k$ 为卷积核的大小, $d \times d$ 为线性操作核的大小, 与 $k \times k$ 大小相似。

研究表明^[14], MobileNetV3 中纺锤形结构并不有利于梯度回传, 会使得网络提早出现过拟合的现象. 针

对上述结论, MobileNeXt 网络构建了一种两头大中间小的沙漏型结构来代替 MobileNet 系列网络中使用的两头小中间大的纺锤形结构, 并将深度可分离卷积层放置在 bottleneck 两头, 使其有更多的信息可以从底层传递到顶层, 有利于梯度回传, 同时可以提取到多种空间信息, 具体结构如图 8 所示。

Input	Operator	Exp size	Out	s
416 ² ×3	Conv2d	-	16	2
208 ² ×16	Bneck,3×3	16	16	1
208 ² ×16	Bneck,3×3	64	24	2
104 ² ×24	Bneck,3×3	72	24	1
104 ² ×24	Bneck,5×5	72	40	2
52 ² ×40	Bneck,5×5	120	40	1
52 ² ×40	Bneck,5×5	120	40	1
52 ² ×40	Bneck,5×5	160	40	1
52 ² ×40	Bneck,5×5	160	40	1
52 ² ×40	Bneck,5×5	200	40	1
52 ² ×40	Bneck,3×3	240	80	2
26 ² ×80	Bneck,3×3	200	80	1
26 ² ×80	Bneck,3×3	184	80	1
26 ² ×80	Bneck,3×3	184	80	1
26 ² ×80	Bneck,3×3	480	112	1
26 ² ×112	Bneck,3×3	672	112	1
26 ² ×112	Bneck,5×5	672	160	2
13 ² ×112	Bneck,5×5	672	160	1
13 ² ×112	Bneck,5×5	672	160	1
13 ² ×112	Bneck,5×5	672	160	1
13 ² ×160	Bneck,5×5	960	160	1
13 ² ×160	Bneck,5×5	960	160	1

图 7 层级均衡的 MobileNetV3 网络结构

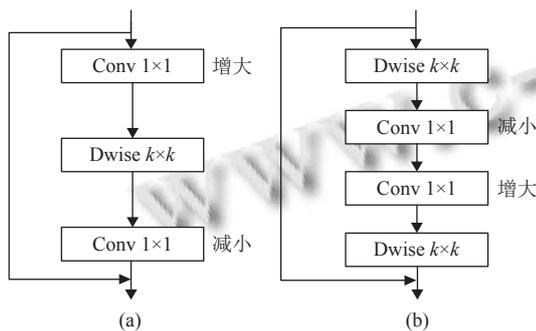


图 8 MobileNeXt 中改进的 bottleneck 结构

本文综合 Ghost 模块中的特征提取思想和 MobileNeXt 网络中的 bottleneck 结构设计思想, 搭建了一种全新的 bottleneck 结构, 用以替换层级调整后的 MobileNetV3 中的 bottleneck 结构, 具体模型结构如图 9 所示, 按照特征尺寸是否需要压缩可分为步长为 1 和 2 的

两种 bottleneck 结构. 本文设计的 bottleneck 融合了 Ghost 模块的思想性, 采用少量逐层卷积和堆叠的方式作为特征提取的模块, 同时将 MobileNetV3 中先大后小的纺锤形 bottleneck, 设计成先小后大的沙漏型结构, 在一个 bottleneck 中采用两个特征提取模块, 以提取更多的空间特征信息. 采用本文设计的 bottleneck 可以有效缓解 MobileNetV3 中纺锤形网络结构带来的信息损失和梯度混淆, 同时可以进一步压缩模型计算量, 使得模型精度和速度得到进一步的提升.

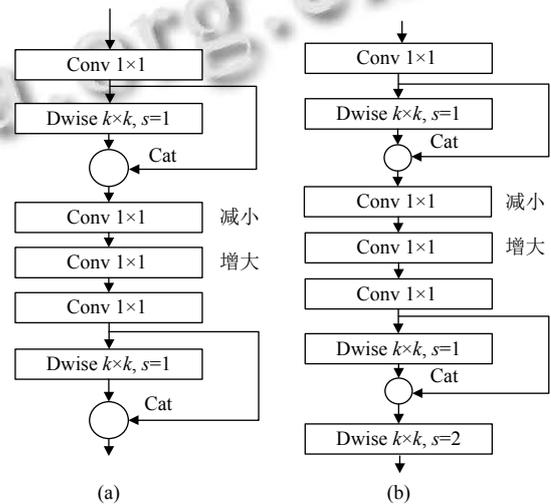


图 9 G-Mobile bottleneck 结构

2.1.3 添加大残差边的主干网络

CSP 结构^[15]的本质是一个大的残差网络, 使用在主干网络的每一大层上. 在主干网络上增加 CSP 结构, 可以在防止模型过拟合的基础上, 有效地保留前一层提取到的图像特征, 从而进一步提升网络的检测精度. 本文借鉴了 CSPNet 网络的思想, 在主干网络的后 3 层添加了大残差边结构, 为使得残差边输出 feature map 数和该层的 feature map 数相同, 该结构在跳跃分支上添加了一个 1×1 卷积, 以保证网络拼接的平衡性. 添加大残差边的目的是在尽量不加深轻量级 G-MobileNet 网络深度的前提下, 获取更多的样本特征, 提高主干网络的特征提取能力, 进而提升检测网络的检测精度. 未在网络前两层添加大残差边结构的原因是前两层主干网络中 bottleneck 较少, 且每一个 bottleneck 都自带一个残差网络, 再添加大残差边结构会出现冗余的现象, 且会导致网络参数量增加. 以 G-MobileNetV3 网络第 3 层为例, 改进后的 GR-MobileNetV3 网络如图 10 所示.

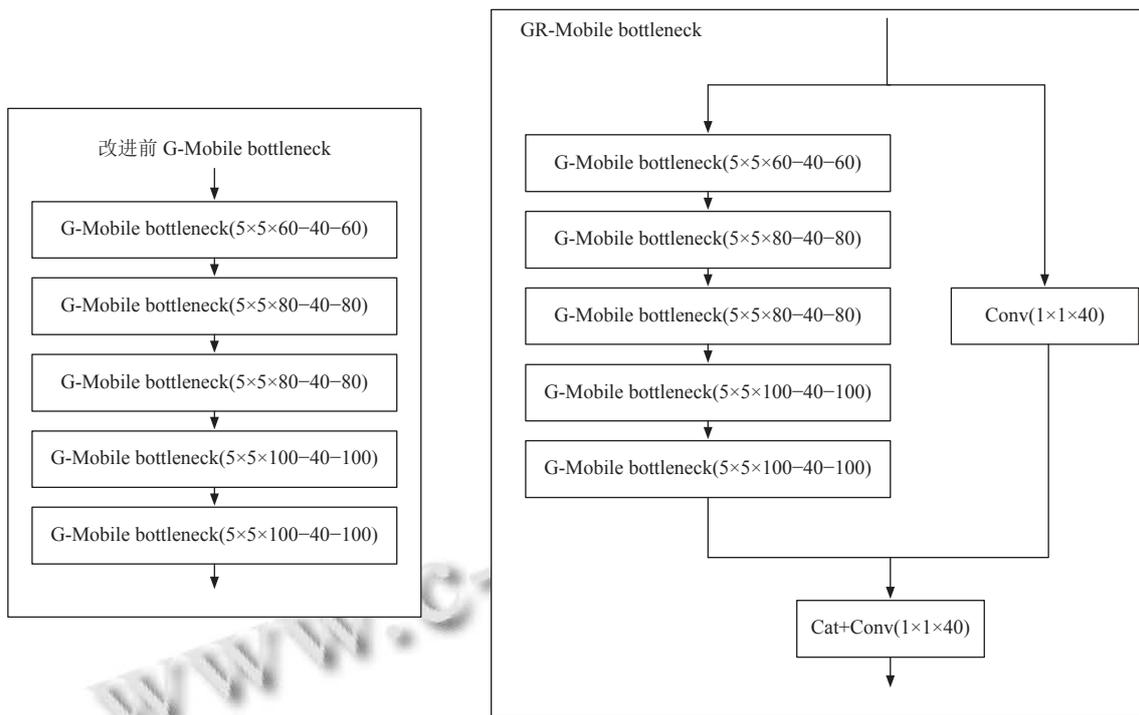


图 10 GR-MobileNetV3 bottleneck 网络

2.2 改进的颈部网络

在 YOLOv4 中, 颈部网络的主要作用是通过反复上采样和下采样对提取到的特征信息进行融合, 使得检测网络的每个检测头都可以得到不同层级的图像信息, 因此在特征融合之前获得更为准确的特征信息尤为重要, 注意力机制作为一种即插即用的网络模块, 其主要作用是使得网络在训练过程中更能注意到需要的特征信息。

CA 注意力机制^[16]是一种通道注意力机制, CA 注意力机制将通道注意力沿 W 和 H 分解为两个 1 维特征编码, 分别沿 2 个空间方向捕获长度依赖关系和位置信息. 两种信息互补的应用于特征图像, 可以增强网络对需要特征的学习^[17]. 同时 CA 注意力机制相较于其余的通道注意力机制, 几乎没有额外的运算成本的开销, 很适合运用在轻量级网络上. 其具体模型图如图 11 所示.

本文将 CA 注意力机制添加在主干网络输出的特征提取层之后, 颈部网络之前, 得到 GR-MobileNetV3-CAYOLOv4 检测网络, 使得网络忽略训练图片中的背景干扰信息, 更多的注意到图片中需要被注意到的信息.

2.3 二分 K-means 算法优化 anchor box

在 YOLOv4 算法中, 对于每个检测头需要 3 个初

始的 anchor box, 故需要在数据集中选出 9 个不同的 anchor box. 对于不同的数据集来说, anchor box 也不尽相同. 在 YOLOv4 中, 我们经常使用 K-means 聚类算法, 通过输入数据集的标定位置, 聚类生成 9 个不同的 anchor box. K-means 算法的本质是随机生成 9 个不同的 anchor box, 通过数据集中各个标记框距 anchor box 中心的距离将其余标记框划分为 9 簇; 然后再找出每簇标记框最中心的 anchor box, 作为新的 9 个 anchor box; 接着重新计算中心距离, 划分为新的 9 簇; 如此循环往复, 直到生成的 9 个不同的 anchor box 位置不在变换为止, 即可认为聚类完成.

K-means 算法过分依赖于初始聚类中心的选择, 如果初始聚类中心选择不好, 可能会导致聚类生成的 anchor box 不是最优解而是局部最优解, 针对 K-means 算法这一缺陷, 本文使用二分 K-means 算法^[18]对 anchor box 进行求解.

在聚类中, 我们认为一个簇中的各个元素靠得越近, 则该簇的聚类越为准确, 二分 K-means 算法则采用了这种思想. 二分 K-means 算法首先将整个数据集作为一个簇, 将该簇进行 K 为 2 的 K-means 聚类, 即将数据集分为两个簇; 接着计算 2 个新簇中各个样本到中心 anchor box 距离的算术平方和, 选择算术平方和较

大的那个簇类使用 K 为 2 的 K-means 聚类算法再次划分; 然后计算现有的 3 个新簇中各个样本到中心 anchor box 距离的算术平方和, 接着进行二分 K-means 操作; 最后将整个数据集划分为 9 个簇, 选择 9 个簇中的中心 anchor box 作为目标检测算法的最终 anchor box. 二分 K-means 算法解决了 K-means 算法过分依赖初值的缺陷, 不断地使用 K 为 2 的 K-means 操作, 使得聚类划分更为科学, 算法不会陷入局部极值, 从而提升检测算法的检测精度.

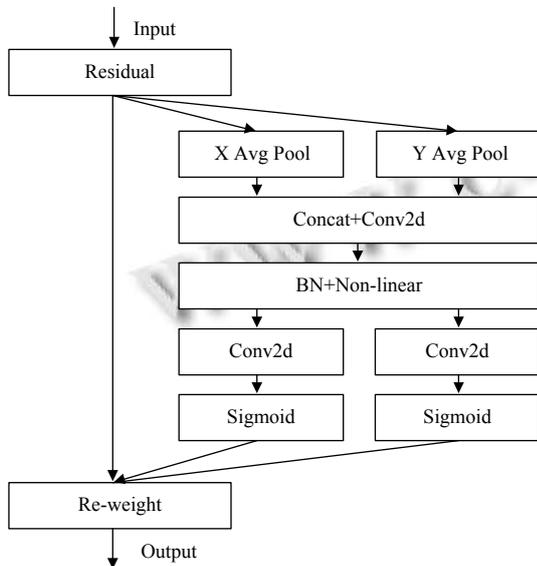


图 11 CA 注意力机制网络结构

3 实验分析

本文实验使用 Windows 10 操作系统的主机; 内存为 16 GB; GPU 为 NVIDIA GeForce RTX 3060; 使用 Python 3.8; 学习框架为 PyTorch 1.71+cuda 11.0. 实验中通过对比主干网络的分类能力验证本文主干网络的性能, 通过将本文设计的检测网络同 YOLOv4 和 YOLOv4-tiny 等算法进行对比, 验证本文网络的检测能力.

3.1 数据集建立

本文对检测网络的总体结构进行了重新设计, 所以 YOLOv4 中提供的预训练权重不能直接进行使用, 故本文须通过分类训练获得主干网络的预训练权重. 本文使用华为提供的垃圾分类数据集对主干网络构成的分类网络进行训练, 该数据集中包含 40 种垃圾图片, 图片总数为 17 000 余张. 通过分类网络训练后, 可以得

到分类网络的垃圾分类识别率和相应主干网络的预训练权重.

在垃圾检测网络的训练中, 本文将标注好的华为垃圾分类数据集以及自行扩充的垃圾检测数据集进行结合, 得到了本文使用的垃圾检测数据集, 该数据集中包含 43 种垃圾图片, 图片总数为 23 000 余张.

3.2 数据集划分

根据本文数据集样本数量和检测难度, 可将数据集分为简单样本、中等样本和困难样本 3 种类别. 其中简单样本指样本数量较多, 类内差异较小、类间差异较大的样本; 困难样本指样本数量较少, 类内差异较大、类间差异较小的样本; 其余为中等样本. 图 12 所示的茶叶渣和食用油桶, 其形状颜色固定, 较容易检测, 故被定义为简单样本; 而图 13 的饮料瓶, 因其颜色种类较多, 且掺杂有形变, 较难进行检测, 故被定义为困难样本. 按照上述定义可将本文数据集进行统一划分, 得到图 14 所示的 3 类样本占比图, 由图可知本文使用的数据集困难样本占比较大.



图 12 简单样本示例



图 13 困难样本示例

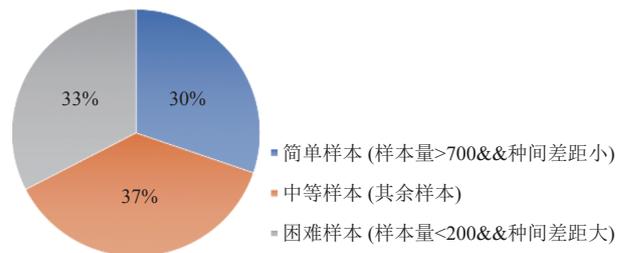


图 14 数据集样本种类占比图

3.3 模型训练

本文实验首先使用垃圾分类数据集训练分类网络,然后剔除分类网络中的全连接层和分类器权重,得到主干网络的预训练权重,将预训练权重添加到检测网络中,使用垃圾检测数据集进行检测网络的训练。

分类网络训练中,采用统一的训练学习率 0.000 1,模型输入分辨率统一转换成 (416, 416, 3),统一训练 400 代。检测网络训练中,采用自适应学习率,模型输入分辨率统一转换为 (416, 416, 3),冻结和不冻结主干网络各训练 100 代。

3.4 评价指标

针对主干网络的分类能力,本文选取测试集上分类效果最好的 5 代权重的识别率,求取平均值作为分类网络的平均识别率 (mean recognition, MR),具体如式 (2) 所示,以衡量主干网络的分类能力,使用模型参数量 Params 来衡量模型的大小。

$$MR = 0.2 \times \sum_{i=1}^5 \frac{CS(i)}{TS(i)} \quad (2)$$

其中, $CS(i)$ 和 $TS(i)$ 分别为测试集中第 i 代分类正确的样本数和总样本数。

本文以 (mean average precision, mAP) 作为模型检测精度的评价指标,以每秒钟图片的传输帧数 (frames per second, FPS) 和参数量 Params 作为检测速度的评价指标^[19]。

3.5 垃圾分类网络实验

本文使用垃圾分类数据集对各主干网络构造的分类网络进行训练,具体结果如表 1 所示。

表 1 分类网络相关实验结果

分类网络名称	MR	Params (MB)
CSPDarknet53 (YOLOv4 的主干网络)	0.775	101.54
层级均衡的 MobileNetV3	0.734	20.92
G-MobileNetV3	0.753	14.64
GR-MobileNetV3 (本文使用的主干网络)	0.796	21.29

从表 1 可以发现, CSPDarknet53 网络参数量大,具有较好的分类能力,但推理速度较慢;层级均衡的 MobileNetV3 参数量降低了 80%,但其检测精度同样下降明显;G-MobileNetV3 网络的模型参数量进一步降低,同时检测精度有所提升,但精度与 CSPDarknet53 仍有一定差距;本文使用的 GR-MobileNetV3 网络,模型参数量虽然略有增加,但精度却得到了很大的提升,

相较于 CSPDarknet53 网络精度提升了 2.1%,故本文将 YOLOv4 中的主干网络替换为 GR-MobileNetV3 网络。

3.6 垃圾检测网络实验

3.6.1 检测网络消融实验

对检测网络进行消融实验,第 1 组为 MobileNetV3-YOLOv4 网络,其余组别依次添加√标注的网络结构,可以得到表 2 所示的实验结果。

组别 2 层级调整后的 MobileNetV3-YOLOv4 网络的推理速度相较于 MobileNetV3-YOLOv4 略微下降,但 mAP 值增加 2.7%;将 bottleneck 更换为本文设计的 G-Mobile bottleneck,组别 3 的 FPS 和 mAP 均得到了提升;接着在主干网络的第 3、4、5 层添加大残差边结构,组别 4 网络检测精度提升 3.7%,但推理速度略有下降;组别 5 是添加 CA 注意力机制的本文网络,其网络运行速度上几乎没有下降,但 mAP 提升 0.7%;组别 6 是在本文设计的检测网络基础上,使用二分 K-means 聚类算法进行聚类操作,相比于使用 K-means 算法在检测精度上提升了 0.3%。综合分析表 2 中数据可以发现,本文采用的网络以较小的检测速度,换取了 9.1% 的检测精度提升。

表 2 垃圾检测消融实验结果

组别	网络设置				mAP@0.5 (%)	FPS
	网络层级调整	更换 bottleneck	大残差边结构	二分 K-means		
1					34.7	75.7
2	√				37.4	72.5
3	√	√			39.1	73.6
4	√	√	√		42.8	69.3
5	√	√	√	√	43.5	67.5
6	√	√	√	√	43.8	67.5

3.6.2 检测网络对比实验

对比不同检测网络在本文数据集上的训练效果,可得到如表 3 所示实验结果。

表 3 相关网络对比结果

检测网络名称	mAP@0.5 (%)	FPS	Params (MB)
YOLOv3	33.1	51.2	235.5
YOLOv4	43.9	44.5	243.3
YOLOv4-tiny	25.0	139.4	23.8
GR-MobileNetV3-CAYOLOv4 (本文构建的网络)	43.8	67.5	26.3

分析表 3,由 Params 指标可以看出本文构建的网络与 YOLOv4-tiny 网络处于同一量级,但 mAP 值明显

优于 YOLOv4-tiny 网络. 本文网络与 YOLOv4 和 YOLOv3 相比, 参数量约为它们的 1/10, 且 FPS 提升了 51% 和 31%, mAP 值和 YOLOv4 网络相当, 比 YOLOv3 网络提升了 10.7%. 综合实验结果可以看出, 本文构建的网络在保证检测精度的前提下, 能够极大的提高检测速度, 可以部署在算力和内存较低的垃圾抓取嵌入式设备上.

3.6.3 不同类别样本检测能力对比实验

根据样本数量和检测难度, 数据集被分为简单样本、中等样本和困难样本 3 种类别, 将这 3 种类别的测试样本图片输入到各检测网络中, 可得到不同网络下各类样本的 mAP 值, 此处提取典型的简单样本、中等样本和困难样本的 mAP 值, 如表 4 所示. 其中茶叶渣、食用油桶、锅是典型的简单样本, 毛绒玩具、包、剩饭剩菜是典型的中等样本, 插头电线、饮料瓶、金属食品罐头是典型的困难样本.

分析表 4 发现, 本文网络对于茶叶渣、食用油桶和锅等简单样本的检测能力较高, mAP 值可以达到 65% 以上, 对于插头电线、金属食品罐等困难样本的

检测能力较弱, mAP 值仅为 10% 左右. 综合数据集样本数据发现, 检测效果不好的多为样本数据量不足, 或类内差异大的困难样本, 如本文数据集中的插头电线、金属食品罐, 它们的训练图片数量仅为 150 余张, 饮料瓶因其类内差异过大, 且无大量数据集样本支撑, 因此检测效果均不理想.

表 4 不同网络下各类样本的检测精度 mAP@0.5 (%)

样本种类	YOLOv3	YOLOv4	本文网络
茶叶渣	73.2	79.6	81.2
食用油桶	71.6	79.3	79.7
锅	69.4	75.8	76.3
毛绒玩具	39.3	47.6	46.8
包	37.1	45.7	46.1
剩饭剩菜	36.3	44.9	45.1
饮料瓶	13.6	18.7	17.6
插头电线	9.2	11.5	11.3
金属食品罐	6.3	9.1	8.7

3.6.4 检测网络实例对比

我们使用 5 种不同种类的垃圾图片输入到 YOLOv4、YOLOv4-tiny 和本文构建的检测网络中, 可以得到如图 15 的检测结果.



图 15 不同算法垃圾检测能力对比

对比图 15 的检测结果, 可以发现 YOLOv4-tiny 在第 3、4 两张图片上未检测到物体, 第 2 张图片将茶叶错误识别为菜叶, 第 1、5 张图片检测框定位不准确, 虽然其 FPS 很高, 但不适用于垃圾检测. 对比 YOLOv4 和本文检测网络可以发现, 本文检测网络在第 3 张图

的检测效果低于 YOLOv4 网络, 其余图片中检测能力均高于 YOLOv4 网络, 尤其是在第 2 张这种有遮挡的图片中, 本文网络能同时检测出茶叶和陶瓷器皿两种物体, 并且定位准确, 说明本文的检测网络有较强的抗遮挡能力, 同时具备更快的推理速度, 更适用于算力较

差的嵌入式设备上,满足现阶段垃圾检测的要求。

4 结束语

基于垃圾检测现状,本文对YOLOv4进行改进,设计了一种适用于垃圾检测的网络。相较于YOLOv4网络,本文网络的参数量减少了89%,FPS提升了51%,同时具有和YOLOv4相似的mAP值。通过对大量图片进行实验,发现本文检测网络在垃圾相互遮挡方面具有较好的检测效果,适合部署在算力和内存较低的嵌入式设备上。针对数据集,后续研究中可通过增加训练样本的数量来提高网络的检测精度;同时针对检测网络的颈部网络进行进一步的改进,以提升网络的检测精度。

参考文献

- 1 张树静,郑毅.城市生活垃圾综合处理技术的探讨——方案的设计与说明.内蒙古师范大学学报(自然科学汉文版),2002,31(4):404-409,413.
- 2 高根树,宋开宇,翁沁文.城市生活垃圾分选技术新进展与瓶颈.环境卫生工程,2016,24(6):11-13,17. [doi: 10.3969/j.issn.1005-8206.2016.06.004]
- 3 董霄霄,刘振国,宋滕滕,等.基于RB-YOLO的生活垃圾实时检测算法.智能计算机与应用,2021,11(10):5-11,19. [doi: 10.3969/j.issn.2095-2163.2021.10.003]
- 4 耿丽婷,阿里甫·库尔班,米娜瓦尔·阿不拉,等.改进SSD的可回收垃圾检测方法.计算机工程与应用,2022,58(23):293-299. [doi: 10.3778/j.issn.1002-8331.2105-0443]
- 5 魏铨磊,南新元,李成荣,等.一种具有多尺度感受视野注意力机制的生活垃圾单阶段目标检测方法.环境工程,2022,40(1):175-183. [doi: 10.13205/j.hjgc.202201026]
- 6 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- 7 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.
- 8 张忠民,张高成.基于YOLOv4的轻量化优化算法.应用科技,2022,49(5):74-80.
- 9 Han K, Wang YH, Tian Q, *et al.* GhostNet: More features from cheap operations. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1577-1586.
- 10 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- 11 项新建,宋晓敏,郑永平,等.基于MobileNet-YOLO的嵌入式人脸检测研究.中国农机化学报,2022,43(4):124-130.
- 12 肖振久,杨玥莹,孔祥旭.基于改进YOLOv4的遥感图像目标检测方法.激光与光电子学进展,2022:1-14. <http://kns.cnki.net/kcms/detail/31.1690.TN.20220214.0853.002.html> [2022-11-17].
- 13 王晨,齐华,史建利.基于YOLOv4的安全帽佩戴检测及工种身份识别.计算机系统应用,2022,31(7):272-277. [doi: 10.15888/j.cnki.csa.008561]
- 14 Zhou DQ, Hou QB, Chen YP, *et al.* Rethinking bottleneck structure for efficient mobile network design. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 680-697.
- 15 Wang CY, Liao HYM, Wu YH, *et al.* CSPNet: A new backbone that can enhance learning capability of CNN. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020. 1571-1580.
- 16 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13713-13722.
- 17 齐永康.基于MobileNet与YOLOv3的路面障碍检测轻量化算法.计算机系统应用,2022,31(2):176-184. [doi: 10.15888/j.cnki.csa.008331]
- 18 张军伟,王念滨,黄少滨,等.二分K均值聚类算法优化及并行化研究.计算机工程,2011,37(17):23-25. [doi: 10.3969/j.issn.1000-3428.2011.17.006]
- 19 曹秋阳,邵叶秦,尹和.基于注意力YOLOv5模型的自动水果识别.计算机系统应用,2022,31(7):333-340. [doi: 10.15888/j.cnki.csa.008576]

(校对责编:孙君艳)