

# 面向边缘计算设备的改进型 YOLOv3 垃圾分类检测模型

王子鹏, 张荣芬\*, 刘宇红, 黄继辉, 陈至栩

贵州大学大数据与信息工程学院, 贵州 贵阳 550025

**摘要** 为提高垃圾分类的自主化和智能化程度, 垃圾桶需要配备视觉传感器和搭载有效的垃圾检测与分类算法的智能硬件。针对该需求, 提出了一种基于改进型 YOLOv3 的智能化垃圾识别分类算法。首先, 引入 MobileNetv3 网络代替 YOLOv3 的主干网络 Darknet53, 并加入空间金字塔池化结构, 在减少网络模型计算复杂度的同时保证模型准确率; 其次, 采用 4 个不同的尺度检测加强模型对小目标的检测能力; 然后, 采用 complete intersection over union (CIoU) 损失函数替换原有 YOLOv3 模型的损失函数, 进一步提升模型的精确度; 最后, 搭建家用垃圾桶测试平台, 并将所提算法移植到边缘计算模块 NVIDIA Xavier NX 上。实验结果表明, 所提轻量化优化算法在服务器和 NVIDIA Xavier NX 平台上对于自主建立的垃圾数据集平均精度一致, 达到了 72.1%, 比 YOLOv3 提高了 4.9 个百分点, 比 YOLOv4 略低 1.6 个百分点; 检测速度分别达到了 74, 19 frame/s, 远高于 YOLOv3 算法的 43, 8 frame/s 及 YOLOv4 算法的 50, 11 frame/s, 表明所提算法满足边缘计算设备的要求, 具备潜在的应用价值。

**关键词** 机器视觉; 垃圾分类; YOLOv3; 深度可分离卷积; 空间金字塔池化; 边缘计算

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202259.0415002

## Improved YOLOv3 Garbage Classification and Detection Model for Edge Computing Devices

Wang Zipeng, Zhang Rongfen\*, Liu Yuhong, Huang Jihui, Chen Zhixu

College of Big Data and Information Engineering, Guizhou University, Guiyang, Guizhou 550025, China

**Abstract** In order to promote the degree of autonomy and intelligence of garbage classification, the garbage bin need be equipped with visual sensor and intelligent hardware carrying effective garbage detection and classification algorithm. To meet this demand, an improved garbage identification and classification algorithm based on YOLOv3 is proposed. First, MobileNetv3 network is introduced to replace Darknet53, the backbone network of YOLOv3, and spatial pyramid pooling structure is added to reduce the computational complexity of the network model and ensure the accuracy of the model. Second, four different scales are used to enhance the detection ability of the model to small targets. Then, the loss function of the original YOLOv3 model is replaced by the complete intersection over union (CIoU) loss function to improve the accuracy of the network model. Finally, a household trash can test platform is built, and the proposed algorithm is transplanted to the edge computing module NVIDIA Xavier NX. The experimental results show that the average accuracy of the proposed optimization algorithm is consistent on the server and NVIDIA Xavier NX platform in the self-made garbage dataset, reaches 72.1%, which is 4.9 percentage point higher than that of YOLOv3 and 1.6 percentage point lower than that of YOLOv4; detection speed is

收稿日期: 2021-02-05; 修回日期: 2021-02-25; 录用日期: 2021-03-25

基金项目: 贵州省科学技术基金资助项目(黔科合基础[2019]1099号)

通信作者: \*rfzhang@gzu.edu.cn

74, 19 frame/s, which is much higher than 43, 8 frame/s of YOLOv3 algorithm and 50, 11 frame/s of YOLOv4 algorithm, indicating that proposed algorithm meets the requirements of edge computing equipment and has potential application value.

**Key words** machine vision; garbage classification; YOLOv3; depthwise separable convolution; spatial pyramid pooling; edge computing

## 1 引言

近年来,我国城市化进程不断推进,传统的对生活垃圾进行填埋和焚烧的方式对环境造成了巨大影响,并且导致了可回收资源的浪费,如何有效处理生活垃圾成为了一个不容忽视的问题。随着人工智能的发展,国内外对于垃圾检测与分类的关注越来越多,例如,谷歌公司 Google X 项目研发的“日常机器人”可以把垃圾分类、把分错类的垃圾放到正确的地方。美国 National Recycling Technologies 公司也推了一款名为 MAX-AI 的垃圾分拣系统,该系统基于深度学习技术,拥有较高的准确率。但是这些垃圾分拣设备占用面积巨大且价格昂贵,无法在国内进行大面积推广。而国内现有的自动分类垃圾桶的分类方式有两种。1) 传统分类方式,该方式依赖垃圾重量进行分类<sup>[1]</sup>,或将垃圾分为金属垃圾和塑料垃圾<sup>[2]</sup>;2) 采用现代技术的分类方式,例如采用摄像机+大数据平台实现自动分类的系统<sup>[3]</sup>,但该系统并未按照国家规定的 4 种分类进行处理,且感知型摄像机价格昂贵;采用分类神经网络初步分类的系统<sup>[4]</sup>,但该系统置信度较低时,需要使用光敏传感器进一步分类,分类过程较为繁琐且不确定性较高,无法保证垃圾分类的实时性及准确性。深度学习的发展给基于计算机视觉的垃圾分类方法带来了新的发展方向。

基于计算机视觉的垃圾检测与分类属于目标检测与分类问题。在目标检测领域,传统的目标检测方法依赖手工设计特征,鲁棒性较差,难以处理复杂背景下的检测问题<sup>[5]</sup>。Krizhevsky 等<sup>[6]</sup>提出了 AlexNet 模型,并在 ImageNet<sup>[7]</sup>比赛中取得了冠军。卷积神经网络(CNN)开始迅猛发展,并在计算机视觉领域广泛应用<sup>[8]</sup>。2014 年, Girshick 等<sup>[9]</sup>提出了 R-CNN, R-CNN 使用基于候选的方法在准确率上取得了巨大提升。随后 Fast R-CNN<sup>[10]</sup>、Faster R-CNN<sup>[11]</sup>、Mask R-CNN<sup>[12]</sup>等相继被提出。但这些网络需要在检测前生成候选区域,计算量较大,无法满足实时检测的要求。为了解决这一问题,Redmon 等<sup>[13-15]</sup>提出了基于回归的 you only look once

(YOLO)系列检测模型,Liu 等<sup>[16]</sup>提出了 SSD 系列模型。然而这些检测模型在提升检测精度的同时对硬件的处理能力要求也逐步提升,从而导致很多模型无法在边缘计算设备上很好地运行。因此关于网络模型的轻量化研究逐渐成为国内外研究者们关注的热点问题,例如对 YOLO 网络进行轻量化的改进<sup>[17-18]</sup>及更换 SSD 的基础网络使其可以搭载到边缘计算平台上<sup>[19]</sup>。此外,各种轻量化网络模型也相继被提出,如 MobileNet<sup>[20]</sup>、SqueezeNet<sup>[21]</sup>等,为实现实时智能化垃圾分类奠定了技术基础。

日常生活中垃圾堆叠互相遮挡,图片较为复杂,且一些垃圾体积较小。现阶段的目标检测与分类网络无法对垃圾进行准确的识别及分类,且模型复杂度较高,在移植到边缘计算设备上时无法满足检测的实时性。因此,为实现对生活垃圾的快速、准确检测与分类并解决模型搭载于嵌入式边缘计算设备的问题,本文针对家庭生活场景,提出了一种改进型 YOLOv3 网络模型,在保证垃圾分类检测精度的情况下,降低了模型的参数量和计算复杂度,以满足搭载到边缘计算设备上的需求。首先,采用 MobileNetv3 网络模型代替 Darknet53 作为 YOLOv3 的主干网络,并引入空间金字塔池化结构(SPP)<sup>[22]</sup>在保证模型检测速度的前提下提升模型的检测精度;其次,在模型中增加了检测尺度,并使用 K-Means 聚类算法对数据集进行聚类分析,加强模型对小目标的检测能力;然后,为进一步提升模型的精确度,在原 YOLOv3 损失函数的基础上引入了 complete intersection over union (CIoU)<sup>[23]</sup>,使得目标检测框的回归更加稳定,也更符合目标检测框回归的机制;最后,针对日常家用场景搭建了基于 NVIDIA Xavier NX 平台的智能分类垃圾桶,并在平台上进行算法移植与测试验证。实验结果表明,所提模型在提高检测精确度的情况下,减少了计算复杂度与参数量,能够搭载到边缘计算设备上有效运行。

## 2 YOLOv3 网络模型

YOLO 系列网络模型借鉴了 GoogLeNet 结构实现了端到端(end-to-end)的学习过程。YOLOv3

采用 Darknet53 作为主干网络,相较于 YOLO 及 YOLOv2,网络的准确率及检测速度均有很大提升。YOLOv4 整体检测效果比 YOLOv3 有所提升,但检测层部分较为复杂,更加难以满足在边缘计算设备上的实时性要求。因此所提垃圾分类优化模型使用 YOLOv3。

表 1 为 YOLOv3 Darknet53 网络框架,不同于以往的卷积神经网络通过控制卷积核步长的方式来实现张量尺寸的变换,网络结构主要由卷积层、批标准化层及 shortcut 层组成,采用 Leaky ReLU 激活函数。相对于 ResNet-152 和 ResNet-101,Darknet53 在保证分类精度的同时大大降低了网络层数和计算复杂度。

表 1 Darknet53 网络结构图

Table 1 Network structure of Darknet53

	Type	Number of filters	Size	Output
	Convolutional	32	3×3	416×416
	Convolutional	64	3×3 / 2	208×208
1×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			208×208
	Convolutional	128	3×3/2	104×104
2×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			104×104
	Convolutional	256	3×3/2	52×52
4×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			52×52
	Convolutional	512	3×3/2	26×26
8×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			26×26
	Convolutional	1024	3×3/2	13×13
16×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			13×13
	Avgpool		Global	
	Connected		1000	
	Softmax			

YOLOv2 采用的 Softmax 分类方法仅能输出一个目标类别,YOLOv3 采用多个 logistic 回归来计算 objectness score 中最高 anchor priors 并输出,从而实现了多标签分类。此外,YOLOv3 采用了类似 feature pyramid networks (FPN) 的 upsample 及特征

融合的方法,在多个尺度的特征图上进行检测,从而极大地加强了算法对于小目标检测的精确度。

### 3 改进型 YOLOv3 网络模型

针对现有 YOLOv3 目标检测模型算法参数量较大,计算较为复杂,无法搭载到边缘计算设备上,且对小目标的检测识别率较低等问题,对 YOLOv3 算法在网络结构、检测尺度及损失函数等方面进行改进,加强了模型对小目标物体的检测能力,在降低模型计算复杂度的同时提升了模型的准确率;同时借鉴 YOLOv4 结构,在主干网络后引入空间金字塔池化(SPP),在计算复杂度略微增长的情况下进一步加强对每张图片的多尺度融合,提高了模型的检测精度,使其能够搭载于嵌入式边缘计算设备,用于家用垃圾桶智能硬件实现实时垃圾分类检测。

#### 3.1 改进网络结构

Darknet53 采用的残差结构使得网络模型获得很好的精度及检测速度,并且通过步长为 2 的卷积核代替池化层降低网络参数的数量,但是由于网络结构过于庞大,参数数量较多,运算较为复杂,该网络不适于搭载到智能硬件用于垃圾图像的检测,且难以满足智能硬件设备对垃圾分类的实时性要求。

为了提高智能硬件设备的检测速度,减小运算复杂度,并保持模型的准确度,所提算法采用 MobileNetv3 模型代替 Darknet53 作为 YOLOv3 的主干网络,其网络结构<sup>[22]</sup>如表 2 所示。

在 MobileNetv3 网络中引入深度可分离卷积来替代传统卷积层,能够极大降低模型参数量及计算量,从而在检测精度没有明显变化的情况下,提高检测速度。设输入数据为  $D_F \times D_F \times M$ ,卷积核为  $D_K \times D_K \times N$ ,并且步长为 1 时,输出特征图尺寸为  $D_F \times D_F \times N$ ,标准卷积参数量为

$$S = D_K \times D_K \times M \times N. \quad (1)$$

对应的计算量为

$$C = D_F \times D_F \times D_K \times D_K \times M \times N. \quad (2)$$

而深度可分离卷积的参数量为

$$S' = D_K \times D_K \times M + M \times N. \quad (3)$$

对应的计算量为

$$C' = D_F \times D_F \times D_K \times D_K \times M + D_F \times D_F \times M \times N. \quad (4)$$

因此,两种结构对应参数量和计算量的比值分别为

$$\frac{S'}{S} = \frac{1}{N} + \frac{1}{D_K^2}, \quad (5)$$

表 2 MobileNetv3-Large 网络结构图<sup>[22]</sup>Table 2 Network structure of MobileNetv3-Large<sup>[22]</sup>

Input	Operator	Exp size	#Out	SE	NL	Step
416×416×3	Conv2d		16		HS	2
208×208×16	bneck, 3×3	16	16		RE	1
208×208×16	bneck, 3×3	64	24		RE	2
104×104×24	bneck, 3×3	72	24		RE	1
104×104×24	bneck, 5×5	72	40	0	RE	2
52×52×40	bneck, 5×5	120	40		RE	1
52×52×40	bneck, 5×5	120	40		RE	1
52×52×40	bneck, 3×3	240	80		HS	2
26×26×80	bneck, 3×3	200	80	0	HS	1
26×26×80	bneck, 3×3	184	80		HS	1
26×26×80	bneck, 3×3	184	80		HS	1
26×26×80	bneck, 3×3	480	112	✓	HS	1
26×26×112	bneck, 3×3	672	112	✓	HS	1
26×26×112	bneck, 5×5	672	160	✓	HS	2
13×13×160	bneck, 5×5	960	160	✓	HS	1
13×13×160	bneck, 5×5	960	160	✓	HS	1
13×13×160	Conv2d, 1×1		960		HS	1
13×13×960	pool, 7×7					1
1×1×960	Conv2d, 1×1 NBN		1280		HS	1
1×1×1280	Conv2d, 1×1 NBN		k			1

$$\frac{C'}{C} = \frac{1}{N} + \frac{1}{D_K^2} \quad (6)$$

较高分辨率的垃圾图像在预处理及进行多尺度预测时可能会造成图像关键信息的缺失,从而影响检测效果。而空间金字塔池化可以对一张图片进行不同的分块池化,在每个池化块中提取出图像的特征作为一个维度,确保最后得到特征的维度一致,从而解决了关键信息丢失及尺度不统一的问题。因此,为了进一步提高模型的精确度,在检测层前引入不同大小的池化核来加强图像局部及全局特征的信息融合。空间金字塔池化结构的最大池化核与需要池化的特征图大小尽可能相同,因此将引入的最大池化核大小设置为 13×13,其他池化核分别设置为 9×9, 5×5, 1×1。空间金字塔池化的引入进一步加强每张图片的多尺度融合,提高模型的检测精度,如图 2 所示。

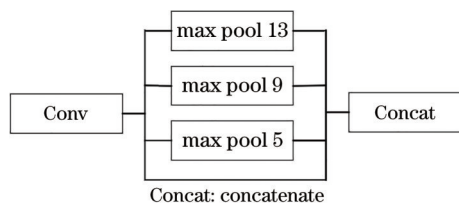


图 2 空间金字塔池化结构

Fig. 2 Structure of spatial pyramid pooling

虽然加入空间金字塔池化在一定程度上增加了模型的计算复杂度,降低了模型的检测速度,但实验结果表明,在加入空间金字塔池化后,模型的整体检测精度进一步提高,且检测速度仅略微下降。

### 3.2 增加检测尺度

YOLOv3 借鉴了 FPN 网络的思路,采用多尺度的检测模块来对不同尺寸大小的物体进行检测,越为精细的网络单元格(grid cell)可以检测出尺寸越小的目标。YOLOv3 采用上采样的方法提取深层特征,并对不同尺度的特征图进行融合,最终输出 3 个不同的尺度检测。由于在对垃圾图像进行检测和识别的过程中,摄像头离垃圾的远近及垃圾本身尺寸的大小不同,因此目标物体在图像当中的占比较小的情况时常会出现。为了加强模型对小目标物体的检测能力,将原模型中的 3 个尺度检测增加为 4 个,增加的尺度为 104×104。对 13×13 的特征图进行下采样获取浅层的图像信息,对 26×26, 52×52, 104×104 的特征图以 2 倍大小执行上采样以获取深层较为复杂的图像信息。为了增强特征金字塔的表征能力,加强图像中的信息融合,将主干网络中相对应尺寸的特征图和尺度预测中的特征图进行级联,并通过 K-means 聚类获得更加准确的锚框,从而加强检测精度。改进后的网络模型如图 3 所示。

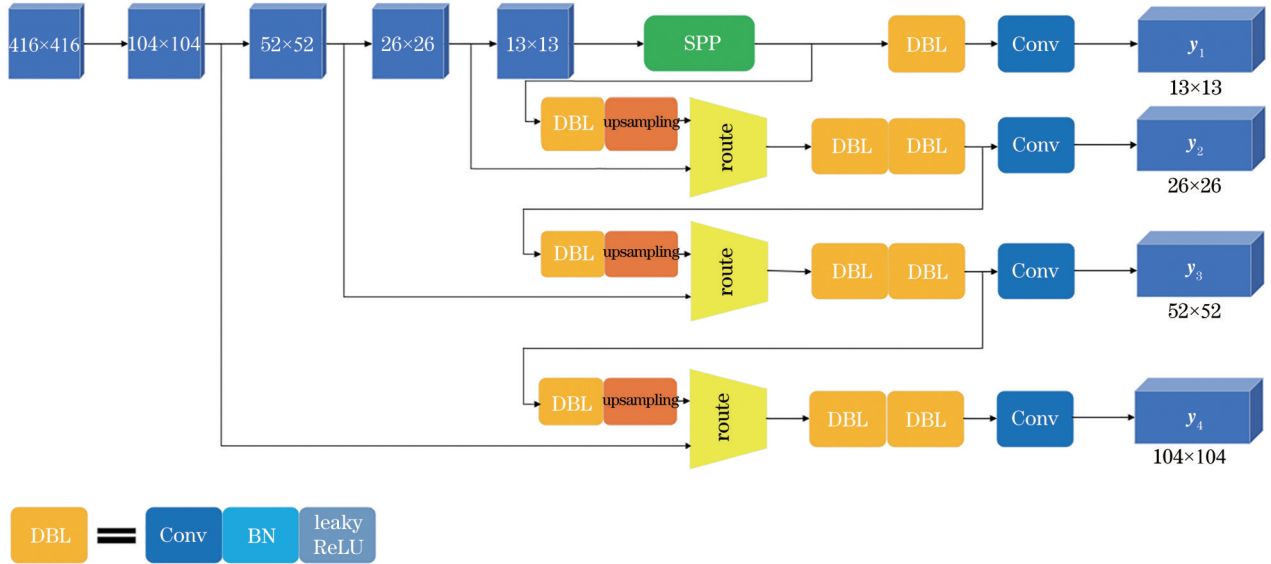


图 3 改进的 YOLOv3 模型  
Fig. 3 Improved YOLOv3 Model

### 3.3 引入新的损失函数

在 YOLOv3 中,采用交并比 (IOU) 来衡量预测检测框的准确度, IOU 为预测检测框和真实检测框的重叠面积与总面积的比值,表达式为

$$R_{IOU} = \frac{|A \cap B|}{|A \cup B|}, \quad (7)$$

式中:  $A$  和  $B$  分别为候选框的面积和真实框的面积。

YOLOv3 的损失函数关于回归预测的部分是采用多个均方差 (MSE) 损失函数相加来进行的,关于分类部分和置信度是采用交叉熵作为损失函数来进行的,总的损失函数表达式为

$$L_{tot} = L_{xy} + L_{wh} + L_{conf} + L_{class}, \quad (8)$$

式中:  $L_{xy}$  表示预测检测框和真实检测框中心点的距离损失;  $L_{wh}$  表示预测检测框和真实检测框宽和高的误差损失;  $L_{conf}$  表示目标的置信度损失;  $L_{class}$  则表示预测的类别损失。但是以 MSE 损失作为回归预测的损失有两个缺点: 1) 如果预测检测框和真实检测框的重叠面积为 0, 根据 (7) 式 IOU 也为 0, 因此不能反映两者的重合度; 2) 同样由于损失为 0, 没有梯度回传, 学习训练也无法进行。

IOU 无法精确地反映预测检测框和真实检测框的重合度。图 4 中, 3 种情况下的重合度不同但 IOU 相同, 左边图回归的效果最好, 右边的最差。

针对 IOU 存在的问题, 兼顾预测检测框和真实检测框的重叠率、尺度问题及对于目标检测框长宽比一致性的计算, 所提算法引入了 CIOU。

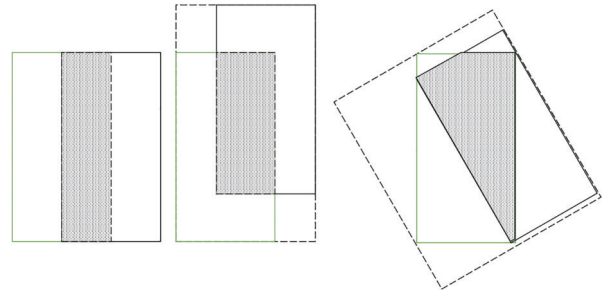


图 4 三种情况对比图  
Fig. 4 Comparison of three situations

$$R_{CIOU} = 1 - R_{IOU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu, \quad (9)$$

$$\nu = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, \quad (10)$$

式中:  $\alpha$  为权重函数;  $\nu$  代表检测框长宽比的一致性;  $w, h$  和  $w^{gt}, h^{gt}$  分别代表预测检测框和真实检测框的宽和高;  $\rho$  为预测检测框和真实检测框中心点坐标的欧氏距离;  $b, b^{gt}$  分别代表预测检测框和真实检测框;  $c$  为包含预测检测框和真实检测框的最小方框的对角线距离。

所提算法使用 CIOU, 与原来的 IOU 相比, 在数据集上的平均精度均值 (mAP) 上的提升较为明显。

## 4 实验结果与分析

### 4.1 实验数据集

由于目前并没有公开的垃圾图像数据集, 所用实验数据均为在家庭日常生活中通过摄像机所采

集的垃圾置于地面上的视频及通过手机拍摄的垃圾置于各种环境下的图片。图像的采集过程充分考虑了视角高低、角度旋转、拍摄距离和光照变化等因素。视频中包含多类垃圾物体,从中共提取关键图片 3500 张;手机拍摄的图片包含 1 类或 1 类以上的垃圾物体,从中共挑选关键图片 5100 张,总计 8600 张图片。各类垃圾具体数量如表 3 所示。

表 3 各类垃圾数量  
Table 3 Quantity of various garbage

Garbage category	Number
Cigarette case	460
Glass bottles	580
Milk box	720
Plastic bottle	810
Screw	370
Batteries	330
Cigarette end	420
Disposable paper cup	650
Packing bag	460
Pen	280
Tissue paper ball	310
Toothpick	650
Walnut shell	390
Banana peel	410
Vegetable leaf	360
Egg shell	390
Orange peel	420
Shells of Sunflower seed	410
Tea leaves	420
Cotton swab	370
Drugs	360
Lipstick	530
Mask	850

为增加数据集的样本规模,使用图像数据增强工具 ImageDataGenerator 对采集的图片样本进行了扩充,扩充方式包括图片放缩、随机旋转、裁剪及镜像旋转,最终得到 21000 张图片,其中 85% 用于模型训练,15% 用于模型测试,并用 LabelImg 标注工具对目标进行标注。对实验采集的垃圾类别依照国家颁布的垃圾分类标准共分为四大类 23 小类。

#### 4.2 实验环境及参数设置

训练均在服务器 Ubuntu 16.04 操作系统下运行,CPU 为 Intel(R) Core(TM) i7-9700K@GHz, GPU 为两块 NVIDIA GTX 1080Ti 显卡,内存为 64 GB。对比实验的硬件配置环境相同。为了加快

网络模型在训练中的收敛速度,提高模型的泛化能力,对模型部分卷积层的起始参数进行了迁移学习。首先使用 MobileNetv3+SPP 作为特征提取模型在 COCO 数据集上以 80 种不同种类物体的预训练参数作为模型部分卷积层的共享参数,并在训练中加以微调。

实验的训练批次为 64,动量为 0.9,衰减为 0.0005,最大迭代次数为 40000。开始训练时学习率为  $1 \times 10^{-3}$ ,迭代 20000 次时学习率调整为  $1 \times 10^{-4}$ ,迭代 30000 次时学习率调整为  $1 \times 10^{-5}$ 。网络训练过程中,损失变化曲线如图 5 所示。

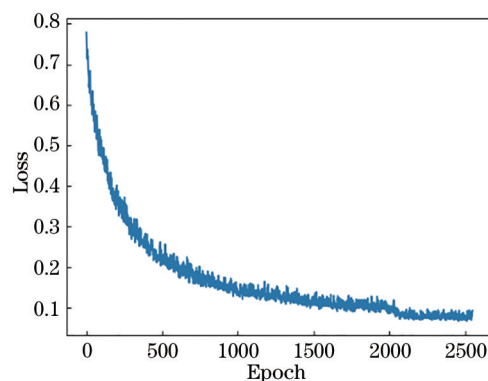


图 5 训练损失曲线

Fig. 5 Training loss curve

#### 4.3 实验结果与分析

使用 mAP、检测速度和模型大小等作为模型的检测指标。mAP 是通过计算每个类别的平均精度 (AP),再对所有类别的平均精度求取均值得到的,表达式为

$$P_{\text{mAP}} = \frac{\sum P_{\text{classAve}}}{N_{\text{classtot}}}, \quad (11)$$

$$P_{\text{classAve}} = \frac{\sum P_{\text{class}}}{N_{\text{class}}}, \quad (12)$$

式中:  $P_{\text{classAve}}$  为不同类别的平均精度;  $P_{\text{class}}$  为某一类别每幅图像的平均检测精度;  $N_{\text{classtot}}$  为总类别个数;  $N_{\text{class}}$  为某一类别物体的测试数目。

##### 4.3.1 模型优化对比

表 4 为仅引入 MobileNetv3 网络及 SSP 结构与原 YOLOv3 模型的 mAP 和检测速度对比结果。从表中可以明显看出:相对于原 YOLOv3 模型,引入 MobileNetv3 网络后的模型,由于采用了深度可分离卷积来替代传统的标准卷积层,参数量和模型的计算量大幅降低,模型的大小有较大的提升,仅为原 YOLOv3 模型的 1/9;从检测速度上看,引入

表 4 不同网络的性能对比

Table 4 Performance comparison of different networks

Network	mAP / %	Model size / MB	Detection speed / (frame·s <sup>-1</sup> )
MobileNetv3	68.8	26.1	76
MobileNetv3+SPP	70.9	27.6	74
YOLOv3	67.2	246.8	43

MobileNetv3 的网络后的模型相对于原 YOLOv3 模型提升了 76.7%；在 mAP 指标上，在采用 MobileNetv3 替代 Darknet53 作为主干网络后，引入 SPP 结构能够进一步加强每张图片的多尺度融合，提高模型的检测精度，使得模型相对于原 YOLOv3 模型提升了 5.5%。实验结果表明，引入 MobileNetv3 网络并加入 SPP 层的方法是可行的。

表 5 为引入 MobileNetv3 的网络模型与原 YOLOv3 模型在引入 CIUO 前后的对比结果。从表中可以看出，在引入 CIUO 后，引入 MobileNetv3 的网络模型与原 YOLOv3 模型相比，mAP 在一定程度

表 5 引入 CIUO 函数前后的对比

Table 5 Comparison before and after the introduction of CIUO function

Network	mAP / %
MobileNetv3+SPP+CIUO	72.1
YOLOv3+CIUO	69.3
MobileNetv3+SPP	70.9
YOLOv3	67.2

上有所提升。对比实验结果表明，CIUO 可以从根本上解决 IOU 中存在的明显缺陷，并同时兼顾预测检测框和真实检测框的重叠率及尺度问题，因此在模型中引入 CIUO 可以在一定程度上提升模型的准确率。

#### 4.3.2 与现有算法对比

将所提改进型 YOLOv3 算法与 YOLOv3、YOLOv3-Tiny 及 YOLOv4 算法对每个垃圾类别的 AP 值进行对比，对比结果如表 6 所示。进一步针对 YOLOv4、YOLOv3-Tiny 及 SSD 算法的 mAP、模型大小及检测速度 3 项指标进行对比，对比结果如表 7 所示。

表 6 四种算法的 AP 值对比

Table 6 Comparison of AP values of four algorithms

unit: %

Garbage category	Improved YOLOv3	YOLOv3	YOLOv3-Tiny	YOLOv4
Cigarette case	59.58	53.85	42.01	60.9
Glass bottles	69.85	61.84	54.03	73.94
Milk box	70.06	59.84	50.18	72.22
Plastic bottle	79.98	79.91	66.86	79.97
Screw	63.95	46.27	37.42	60.19
Batteries	76.21	71.21	52.57	78.7
Cigarette end	74.63	72.44	66.32	77.55
Disposable paper cup	75.52	72.64	67.56	80.45
Packing bag	77.95	70.95	62.12	78.35
Pen	74.49	73.28	60.58	76.5
Tissue paper ball	73.2	70.1	55.01	74.24
Toothpick	76.98	74.61	64.4	77.11
Walnut shell	57.6	51.57	30.53	60.09
Banana peel	78.85	75.75	66.88	79.85
Vegetable leaf	66.14	62.68	47.49	71.8
Egg shell	71.95	71.52	56.49	72.8
Orange peel	75.35	65.25	63.28	75.87
Shells of Sunflower seed	75.51	73.88	64.5	76.61
Tea leaves	72.66	70.43	57.57	73.22
Cotton swab	65.97	64.47	63.39	68.64
Drugs	75.27	71.25	67.53	77.95
Lipstick	77.93	74.62	61.33	78.37
Mask	68.8	68.37	53.53	69.94

表 7 七种算法整体检测性能对比

Table 7 Comparison of overall detection performance of seven algorithms

Network	mAP / %	Model size / MB	Detection speed / (frame·s <sup>-1</sup> )
Improved YOLOv3	72.1	27.6	74
YOLOv3	67.2	246.8	43
YOLOv3-Tiny	56.6	34.9	77
YOLOv4	73.7	256.6	50
SSD	65.4	97.3	48
MobileNet+YOLOv3	66.8	25.3	75
MobileNetv3+YOLOv3-Tiny	55.8	11.7	82

从表 6 中可以看出,在 23 个垃圾类别上,所提改进型 YOLOv3 的检测精度均高于 YOLOv3 和 YOLOv3-Tiny,略低于 YOLOv4。且在小目标的检测当中,如 Screw(螺丝钉),所提改进型 YOLOv3 的检测精度相较于 YOLOv4、YOLOv3 和 YOLOv3-Tiny 分别提升 6.25%、38.21%、70.9%,表明在改进型 YOLOv3 的检测模型中增加小尺度检测的方法可以增强对小目标的检测能力。对比表 6 和表 7 可以看出,在检测精度上,由于采用了 SPP 结构并引入 CIUO,所提改进型 YOLOv3 的检测效果要明显优于 YOLOv3、YOLOv3-Tiny、MobileNet+YOLOv3 及 SSD 算法,但与 YOLOv4 还有一定差距;在检测速度方面,由于采用了 MobileNetv3 网络替代 Darknet53 作为主干网络的策略,模型在模型参数及计算量上面大幅减少,同时保持了模型原有的精度,所提改进型 YOLOv3 的检测速度相对于原 YOLOv3 提升 72.1%,相对于 YOLOv4 提升 48%,略低于 YOLOv3-Tiny。同样,在模型大小上,所提改进型 YOLOv3 仅为原 YOLOv3 模型大小的 1/9,大于 YOLOv3-Tiny。

MobileNet+YOLOv3 和 MobileNetv3+YOLOv3-Tiny 为常见的基于 YOLO 系列的轻量化方式,从表 7 可以看出,仅引入 MobileNet 网络虽可以大大减少模型大小,提高检测速度,但模型的 mAP 略有下降,并不能够实现提升模型的 mAP 的同时加强对小目标物体的检测。

图 6 为所提改进型 YOLOv3 和 YOLOv3、YOLOv3-Tiny、YOLOv4 对几类垃圾的检测效果图。从图中可以明显看出,所提改进型 YOLOv3 的预测检测框与真实检测框的重合度更高、对于目标

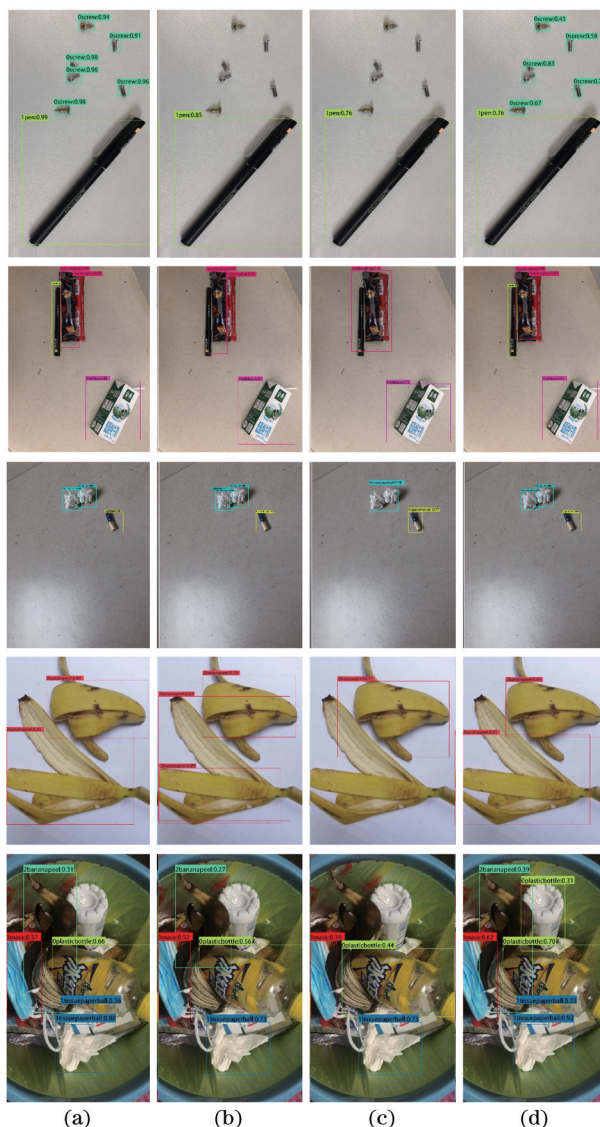


图 6 四种算法检测效果对比。(a)改进型 YOLOv3; (b) YOLOv3; (c) YOLOv3-Tiny; (d) YOLOv4  
Fig. 6 Comparison of detection effects of four algorithms. (a) Improved YOLOv3; (b) YOLOv3; (c) YOLOv3-Tiny; (d) YOLOv4

的定位更加准确。在针对小目标垃圾物体的检测中, YOLOv3、YOLOv3-Tiny、YOLOv4 均存在漏检、误检及重复检测的情况,由于模型中加强了对小目标物体的检测,所提改进型 YOLOv3 具有更好的检测效果。

### 5 平台移植与验证

为进一步在边缘计算设备上开展对比验证实验,将服务器上训练完成的算法移植到边缘计算平台上进行实验。实验搭建的家用移动自主分类垃圾桶包括高性能 NVIDIA Xavier NX 边缘计算板(垃圾分类模块)、可以全方向运动的四轮驱动机器人底盘、基于



STM32 的驱动控制板、车载摄像头及安装支架、蓝牙通信模块与红外采集模块、语音交互装置、Ubuntu 18.04 操作系统、robot operating system (ROS) 系统、

OpenCV3.4、TensorFlow1.15、caffe、darknet 等工具与框架。图 7 为系统整体工作流程示意图。

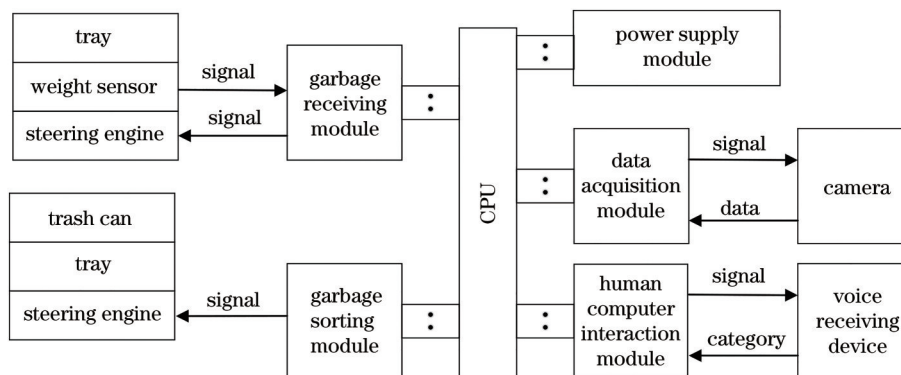


图 7 系统整体工作流程示意图

Fig. 7 Overall work flow diagram of system

将算法移植到实物平台上,各算法识别准确度与表 7 相同,检测速度对比效果如表 8 所示。从表 8 中可以看出,所提改进型 YOLOv3 算法搭载到 NVIDIA Xavier NX 后检测速度明显高于 YOLOv4 及 YOLOv3,与 YOLOv3-Tiny 检测速度相当,但检测准确度远远高于 YOLOv3-Tiny,同时满足了垃圾分类实时性和准确度的要求。总的来说,所提改进型 YOLOv3 算法在垃圾数据集上可以在一定程度上提高垃圾分类准确率,降低算法的复杂程度,从而提高检测的整体效果,并且满足边缘计算设备对模型的各项要求。

表 8 四种算法在 NVIDIA Xavier NX 上的检测速度对比  
Table 8 Comparison of detection speed of four algorithms on NVIDIA Xavier NX

Network	Detection speed / (frame·s <sup>-1</sup> )	mAP / %
Improved YOLOv3	19	72.1
YOLOv3	8	67.2
YOLOv3-Tiny	20	56.6
YOLOv4	11	73.7

## 6 结 论

提出了一种改进型 YOLOv3 垃圾分类算法,该算法可以搭载于边缘计算设备上对垃圾进行检测,满足了边缘设备对于常用家庭垃圾检测与分类的准确性和实时性要求。所提算法引入 MobileNetv3 网络代替 YOLOv3 的主干网络 Darknet53 并加入空间金字塔池化结构,实现了多层信息的融合与特征

重用,从而提高了模型的准确率;在损失函数方面,采用了 CIUO 损失函数,使得对于目标的定位更加精准,检测的精度也有了进一步的提升;此外,模型增加了检测尺度,提升了对于小目标物体的检测精度。实验结果表明,所提改进型 YOLOv3 算法在保证检测准确率的同时,在检测速度方面明显优于现有的 YOLOv3、YOLOv4 算法,满足边缘计算设备应用需求,具有实际应用潜能。后续工作可以进一步优化算法并增加可识别垃圾的种类,搭载到服务器云端,实现家用、办公场所及商场等各种环境下的垃圾自动分类。

## 参 考 文 献

- [1] Wang S C. An intelligent garbage sorting box: CN105905477A[P]. 2016-08-31.  
王世晨. 一种智能垃圾分类箱: CN105905477A[P]. 2016-08-31.
- [2] Miao J H. A garbage sorting device: CN108313562A [P] 2018-07-24.  
缪佳禾. 一种垃圾分类装置: CN108313562A[P]. 2018-07-24.
- [3] Ge J Z, Li J X, Wu K W, et al. Interactive garbage automatic sorting and recycling bin: CN107600791A [P]. 2018-01-19.  
葛建中, 李婧璇, 吴克伟, 等. 交互式垃圾自动分类回收箱: CN107600791A[P]. 2018-01-19.
- [4] Liu B, Gan J H, Wen B, Wang J, et al. An automatic sorting and recycling method for campus garbage based on deep learning was proposed: CN109308479A[P]. 2019-02-05.

- 刘波, 甘健侯, 文斌, 等. 一种基于深度学习的校园垃圾自动分类回收方法: CN109308479A[P]. 2019-02-05.
- [5] Ning K, Zhang D B, Yin F, et al. Garbage detection and classification of intelligent sweeping robot based on visual perception[J]. *Journal of Image and Graphics*, 2019, 24(8): 1358-1368.  
宁凯, 张东波, 印峰, 等. 基于视觉感知的智能扫地机器人的垃圾检测与分类[J]. *中国图象图形学报*, 2019, 24(8): 1358-1368.
- [6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [7] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [8] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [9] 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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [10] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [11] 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.
- [12] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [13] 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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [14] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [15] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-02-01]. <https://arxiv.org/abs/1804.02767>.
- [16] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [17] Li C Y, Yao J M, Lin Z X, et al. Object detection method based on improved YOLO lightweight network [J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141003.  
李成跃, 姚剑敏, 林志贤, 等. 基于改进 YOLO 轻量化网络的目标检测方法[J]. *激光与光电子学进展*, 2020, 57(14): 141003.
- [18] Cui J H, Zhang Y Z, Wang Z, et al. Light-weight object detection networks for embedded platform[J]. *Acta Optica Sinica*, 2019, 39(4): 0415006.  
崔家华, 张云洲, 王争, 等. 面向嵌入式平台的轻量级目标检测网络 [J]. *光学学报*, 2019, 39(4): 0415006.
- [19] Chen L L, Zhang Z D, Peng L. Real-time detection based on improved single shot MultiBox detector[J]. *Laser & Optoelectronics Progress*, 2019, 56(1): 011002.  
陈立里, 张正道, 彭力. 基于改进 SSD 的实时检测方法[J]. *激光与光电子学进展*, 2019, 56(1): 011002.
- [20] 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.
- [21] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [22] He K M, Zhang X Y, Ren S Q, 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.
- [23] 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.