

文章编号: 2096-1618(2021)04-0374-06

基于卷积神经网络的垃圾图像分类模型研究应用

唐康健, 文展, 李文藻

(成都信息工程大学通信工程学院, 四川 成都 610225)

摘要:传统的垃圾分类方法往往借助于传感器完成垃圾识别分类,存在分类的准确率不高、模型复杂、缺乏高效操作性等问题。为解决这一问题,提出结合卷积神经网络的垃圾分类方法。使用具有高效特征提取性能的 Inception-V3 和 ResNet50 两种卷积神经网络对华为公开垃圾数据集 Garbage Date 进行训练,建立垃圾分类模型。实验表明,在训练集上的 Inception-V3 和 ResNet50 训练的准确率分别为89.9%和95.1%,交叉熵损失函数分别为1.463和1.363。使用可视化界面验证测试集中随机6类单张图片,ResNet50的准确率均高于Inception-V3。但ResNet50却不及Inception-V3稳定,Inception-V3准确率曲线更平滑。Inception-V3的收敛速度也比ResNet50快,消耗资源更少。

关键词:垃圾分类;卷积神经网络;准确率;交叉熵;Inception-V3;ResNet50

中图分类号:TP301.6

文献标志码:A

doi:10.16836/j.cnki.jcuit.2021.04.004

0 引言

垃圾分类处理问题一直与人民生活息息相关,准确高效的垃圾分类处理方式能够促进社会可持续发展,对节约资源保护环境有深远的影响。由于垃圾分类意识的缺乏和垃圾分类设备有限等因素的影响,传统的垃圾分类方法具有分类效率低、成本高、人员需求量的缺点,很难做到准确高效的垃圾分类,从而导致垃圾分类效率和准确率都很低,环境问题和资源问题也日趋严重。科技快速进步和发展,人类已逐步地迈入了大数据、云计算、人工智能的时代。利用机器学习进行垃圾分类能够有效地弥补传统垃圾分类方式的不足,形成高效准确的垃圾分类模式。文献[1]将图像分类技术与物联网结合,提出了智能垃圾分类回收系统,文献[2]运用大数据、传感器和图像识别技术设计了智能分类垃圾箱,文献[3]提出了一种基于纹理特征融合的图像识别技术,对道路垃圾进行特征提取和分类。文献[4]利用迁移学习提出了一种基于卷积神经网络的垃圾图像分类方法。这些研究也为垃圾分类处理提供了新的研究方向。

深度学习作为最有效的机器学习技术之一,其中卷积神经网络(CNN)在图像分类、目标检测等方面取得了很大的成就。研究者们提出了很多性能优越的卷积神经网络模型,其中Inception系列网络的提出就是CNN分类器发展史上的一个里程碑。在Inception系列网络提出之前,通用的CNN都是将网络中的卷积层进行多

次堆叠来获得更好的分类效果^[5-6]。Krizhevsky^[7]提出的AlexNet网络就以这样的方法扩展了网络结构的深度。后来由Simonyan等提出的VGG-Net网络也同样有了更深层次的网络结构。但是随着网络结构的加深会导致网络参数量十分庞大,消耗更多的计算资源。随之提出了一种使用Inception模块的深层卷积神经网络,该网络结构主要特点提高了计算资源的利用率,在保持计算预算不变的情况下增加了网络的宽度和深度。Inception深度卷积架构首先由Szegedy等引入,被称为GoogleNet或Inception-V1^[9]。Google也凭借GoogleNet在2014年的ImageNet图像识别挑战赛(ILSVRC)中大放光彩。随着Inception体系结构地不断完善,Google接连提出了Inception-V1/V2/V3/V4网络^[10-13]。此外,由何凯明团队提出的深度残差网络(ResNet)在2015年的图像识别挑战赛中夺魁,为进一步增大CNN网络的深度,在训练中引入连跳结构来帮助梯度反向传播。王文成、蒋慧等2019年运用ResNet50网络模型框架,采用优化算法来分类识别10种鱼类图像,并且得到了较高的准确率,说明该网络模型能得到较好的图像分类识别效果。

文中主要研究了基于卷积神经网络实现垃圾分类的方法,利用Inception-V3和ResNet50建立了两种垃圾分类模型,并将模型用于华为的垃圾公开数据集Garbage Date进行训练,得到两种具有较高的准确率和较低的损失函数值的垃圾分类模型,并将两种训练好的分类模型应用在具体垃圾分类系统中。通过对垃圾进行拍照,搭建可视化平台观察图片识别的准确率的方式,对模型性能做进一步的检验。

1 卷积神经网络

1.1 卷积神经网络结构

卷积神经网络是从传统网络中获得的多层网络结构的人工神经网络,其特点:权值共享、局部连接、下采样,减少不重要的参数,自动提取图像特征,提高学习效果,得到较高的预测准确率。因此通常应用于图像识别系统。

卷积神经网络的基本架构有:输入层、卷积层、池化层、全连接层和输出层。

卷积神经网络的第一层是输入层。图像通过输入层直接作为输入训练网络,为了后续的卷积和池化运算,输入层图像的尺寸大小需要做预处理。

卷积神经网络有很多卷积层,每个卷积层又有很多卷积核。每个卷积核只能提取一个不同的输入图像特征,通过与输入执行卷积运算,反向传播优化卷积核来获得最优解,激活函数映射变换后得到新的由输入到输出的特征映射关系图像,因此卷积核是卷积层的核心部分。

池化层又叫下采样层,通常接在卷积层后。经过卷积运算后得到的特征图像的维度很大,会增加 CPU/GPU 运行的负担。但池化层会对上一层输出的特征图像进行特征聚合统计,取其最大值或平均值,对特征图像降维,在保留有效数据的基础上,达到减少计算量、加快网络训练速度的目的。

在经过一系列的卷积、池化运算后,网络会提取到输入图像最显著的特征作为输出。然后经过接在网络后面的全连接层,整合精炼出前面提取图像的特征值。最后将特征值用于图像分类计算中。

分类器常用于输出层,获得图像分类得到各类别的概率,概率总和为 1。

1.2 卷积神经网络网络模型

1.2.1 Inception-V3 网络模型

Inception-V3 是第三代 GoogleNet。Inception 系列

网络之所以拥有如此优良的性能,是因为网络中使用了不同类型的 Inception 模块。Inception 模块在同一层级上并行了多个尺寸的滤波器,能够很好地采集不同大小输入图像中的特征信息,能够很好地解决因为网络过深带来的容易过拟合和梯度更新传输到整个网络困难等问题,并且很大程度上降低计算时所消耗的计算资源。

Inception-V3 网络引入了 Factorization into small convolution 的思想,将一个较大的二维卷积拆成两个较小的非对称一维卷积,这样一方面节约了大量参数,加速运算并减轻了过拟合,同时还增加了一层非线性扩展模型表达能力。另一方面 Inception-V3 优化了 Inception 模块的结构,在模块的分支中使用分支。这样的结构不仅使 Inception-V3 网络的训练速度得到了提高,还使分类准确率得到了提升。

1.2.2 Resnet 网络模型

在 CNN 的深度结构中,达到一定的网络层深度时,一系列梯度弥耗和爆炸等梯度问题会造成网络越来越难以训练,梯度信息在传递过程中出现了接近于 0 或者梯度值特别大的现象,导致深层的网络训练准确率低于浅层网络的训练准确度。ResNet 的提出可以解决这类深度问题,在输入和输出间添加直接连接的 Skip Connection 使深层网络具有回退到浅层网络的能力,并且加快训练超深的神经网络的速度,在深度增加的同时网络的性能也得以极大的提升。

2015 年,何凯明团队提出了 18 层、34 层、50 层、101 层、152 层的模型,甚至达到 1202 层的极深层神经网络,这些成果至今仍有巨大的研究价值和影响力。文中主要介绍其中的一类 ResNet50 网络。该网络首先进行图像预处理,将不同尺寸的图像数据进行归一化处理,得到相同尺寸的图像。然后,4 个残差块进行连续卷积,其图像像素矩阵沿通道维度相加,深度随着加深。最后,图像像素矩阵输入全连接层,并通过 Softmax 分类器输出图像相应的分类概率。

1.2.3 两类模型对比

Inception-V3 和 ResNet50 的对比如表 1 所示。

表 1 Inception-V3 和 ResNet50 的对比

	Inception-V3	ResNet50
时间	2015-12	2015-12
提出目的	扩展网络宽度	扩展网络深度
核心思想	引入因子分解思想	重复叠加卷积块和识别块
工作思路	Inception 模块为多个卷积与池化层构成的网络结构。这种网络模型将较大的卷积核做了非对称的卷积结构拆分,通过把 $1 \times n$ 的卷积核卷积后接上一个 $n \times 1$ 的卷积核卷积来替代任意一个 $n \times n$ 的较大的卷积核。这种做法能够增加非线性的特性,减轻过拟合现象,可以处理更加复杂的空间特征。Inception-V3 模型整合前面历代模型的优点,进行优化,新增 8×8 、 17×17 、 35×35 这 3 种不同的卷积核尺寸,能有效地将梯度传播到底层,又能起到正则化的作用,使最终的识别效果更好。	
	首先,进行图像预处理,将不同尺寸的图像数据进行归一化处理,得到相同尺寸的图像。4 个残差块进行连续卷积后,其图像像素矩阵沿通道维度相加,深度随着加深。最后,图像像素矩阵输入全连接层,并通过 Softmax 分类器输出图像相应的分类概率。	

2 数据预处理和网络模型构建

2.1 数据预处理

首先运用数据归一化方法,将不同尺寸的图片都转化为相同尺寸的图片,方便后续的数据处理。然后采用随机增强变换来扩充数据,避免因为数据集较小的情况下产生过拟合等现象。

利用 keras 内置的图片生成器 Image Data Generator 来完成图像的增广。在训练中,设置要用到的随机变换,然后通过 flow_from_directory 生成经过数据归一化和增广处理的数据,一次进行增强的数量要与 Batch_size 一致,keras 方法中使用 fit_generator 函数。共使用了 7 种数据增强技术,如表 2 所示。

表 2 图片增强技术说明	
技术名称	具体作用
rotation_range	对输入图像随机指定旋转一定的角度范围,最大不超过 180°。
width_shift_range	对输入图像作水平偏移,改变图像的宽度。
shear_range	对输入图像做逆时针方向错切变换,发生平移变换。
zoom_range	对输入图像在长高两个方向随机缩放,参数大于 1 执行图片缩小操作;参数大于 0 而小于 1,执行图片放大操作。
channel_shift_range	输入图像颜色通道做偏移操作,更改图像的整体颜色。
horizontal_flip	在随机情况下,对输入图像做水平翻转。
fill_mode	当图片进行随机变换时出现一些缺失部分用该方式来填充补全。

通过以上这些变换可以增加图像的多样性,运用数据增强技术增加小数据集的图像数量,从而增强了网络的泛化能力,减轻网络的过拟合现象,使网络模型可以发挥更好的效果。

2.2 卷积神经网络模型的构建

2.2.1 网络模型的训练

卷积神经网络模型的训练过程如下:

步骤 1 把模型的最后一层全连接层删去,应用到新数据集中,新增加符合本课题的垃圾分类数据集类别数的全连接层,同时输出层要替换为 softmax 层,同时追加 Dropout 层,防止过拟合。

步骤 2 对垃圾数据集进行数据增强处理,通过随机旋转、水平翻转、垂直翻转、剪切变换、随机放缩等操作来扩充数据,将原来的图片扩充至多倍,促进模型

的推广。

步骤 3 设置合适的超参数值,根据实验环境配置和数据集大小,将 Batch_size 设为 64,epochs 设为 100,初始学习率设为 0.001。

步骤 4 分别采用 Inception-V3 和 ResNet50 作为特征提取器,将 Garbage Date 数据集输入修改好的网络模型中重新训练,得到新的权重。

步骤 5 保存训练好的模型权重。

2.2.2 网络模型的测试

卷积神经网络模型的测试过程如下:

步骤 1 使用训练好的网络模型权重。

步骤 2 程序开始运行后,会自动弹出“垃圾分类”可视化界面,打开测试图片数据集,从测试集中随机抽取一个测试样本,输入网络。

步骤 3 测试样本通过网络反向传播,网络逐层进行计算输出。

步骤 4 将网络输出与测试样本的标签进行比较,判断每张图片的输出类别是否正确,对比两类模型并对分类结果进行统计。

步骤 5 重复步骤 2~4,可以将测试集所有图片都测试和分类完毕。

3 实验结果

3.1 网络模型性能分析

实验选用的数据集是华为垃圾公开数据集。该数据包括 40 类垃圾,一共有 5087 张图片。分别使用了 Inception-V3 和 ResNet50 卷积神经网络对 Garbage Date 数据集进行训练,建立了垃圾分类模型。通过分析垃圾分类模型的准确率和交叉熵损失函数值来判定分类模型的性能。准确率越接近 1,说明模型的准确性越高。交叉熵值描述了两个概率分布之间的距离,当交叉熵值越小说明二者之间越接近,其值越接近于 0,说明分类结果越准确。

为了对比 Inception-V3 和 ResNet50 的性能,对这两种神经网络训练的分类模型的基本参数采用了相同的参数设置。选取 Garbage Date 数据集作为模型的输入,将训练集和测试集以 3:1 的比例划分,设置 epochs 为 20,学习率为 0.001,训练的 Batch_size 为 64。Dropout 的断开概率设置为 0.5。选取了 ReLu 函数为激活函数。

训练时神经网络先检索数据集里的图片的类别,形成 label 标签,作为分类的类别,此数据集中一共有 40 个类别,因此形成 40 个标签。然后依次检索每个

标签对应文件夹下的图片,并且生成对应于每张图片的 txt 文本文件。训练时,每次向神经网络中输入大小为 Batch_size 的图片,这样可以减少网络训练的时间。经过 20 次的迭代训练,最后得到模型的训练结果。图 1~3 分别是训练集准确率、测试集准确率和交叉熵函数值在两个不同模型中的对比结果。

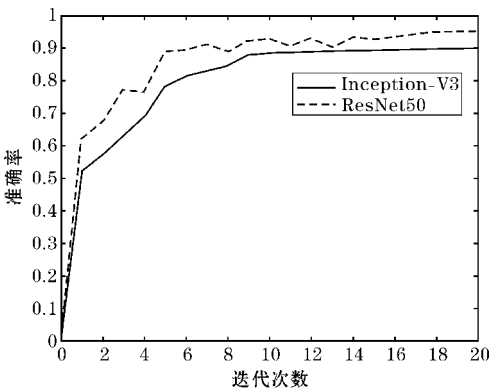


图 1 训练集准确率

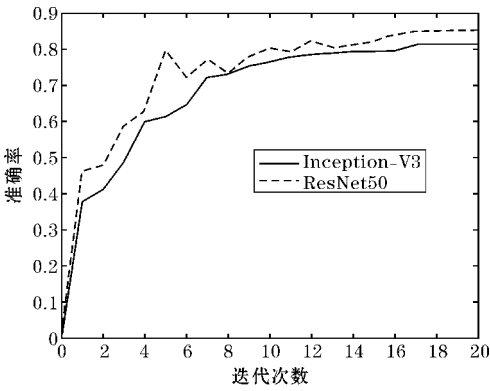


图 2 测试集准确率

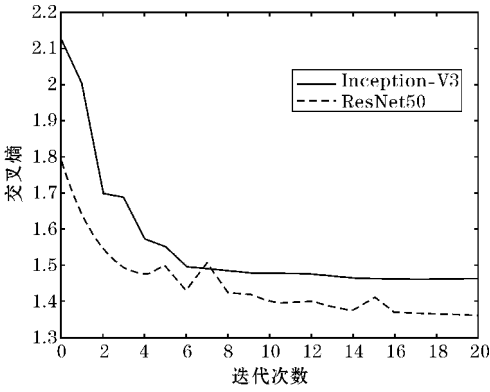


图 3 损失函数图像

从图 1~3 可以得知,随着迭代次数的不断增加,两个模型在训练集和测试集上的准确率均呈上升趋势。迭代次数达到 50 次左右时,在 Inception-V3 的训练集和测试集的准确率上升幅度趋于平稳,ResNet50 的训练集和测试集上的准确率还存在波动。在训练结束后,Inception-V3 在训练集上的准确率达到89.9%,在测试集上的准确率达到81.4%;ResNet50 在训练集上的准确

率达到了95.1%,在测试集上的准确率达到85.2%。另外,两个网络的交叉熵函数值随着迭代次数呈下降趋势,其中 Inception-V3 的交叉熵函数值为1.463;ResNet50 的交叉熵函数值为1.363。可见,ResNet50 的准确率高于 Inception-V3,但稳定性不足 Inception-V3。

为了对比两类模型消耗资源的情况,对两种模型的参数量进行了对比。Inception-V3、ResNet50 的参数量如表 3、4 所示。

表 3 Inception-V3 参数量

层	参数量
Conv0	864
Conv1	9216
Conv2	18432
Conv3	5120
Conv4	138240
Inception 模块 1	815616
Inception 模块 2	10672128
Inception 模块 3	12656640
fc	417792
总计	2.4×10^6

表 4 ResNet50 参数量

层	参数量
Conv1	$7 \times 7, 64$
	3×3
Conv2	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv5	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2028 \end{bmatrix} \times 3$
总计	25.5×10^6

可以看出,Inception-V3 的参数量是小于 ResNet50 的。另外,采用了对比模型消耗时间的方法,在同一硬件设备的条件下,两类模型具有相同的计算能力,因此消耗时间与消耗的计算量资源成正比。实验结果如表 5 所示。

表 5 消耗时间对比 单位:s

模型名称	迭代次数			
	5	10	15	20
Inception-V3	95	197	307	402
ResNet50	160	315	480	630

表5分别给出了两类模型随迭代次数增加消耗时间的变化情况。可以看出在迭代20次后 Inception-V3 消耗的时间为402 s,远小于 ResNet50 消耗的630 s,说明 Inception-V3 节约了计算资源,模型收敛更快。

3.2 模型应用研究分析

考虑将训练好的模型应用于智能垃圾桶中。智能垃圾桶是通过在垃圾桶上搭载摄像头等外设设备,对投入其中的垃圾进行拍照,将获取的图片作为垃圾分类模型的输入,利用训练好的分类模型对其进行分类,并将分类结果显示在外设屏幕上,以此提醒使用者对垃圾进行分类投放。实验依据这样的过程,对实物垃圾进行拍照并搭建可视化界面模拟智能垃圾桶的工作流程。

安装 Qt 工具,利用 Qt Designer 制作 GUI 界面,搭建一个可视化界面测试各类图像分类识别的准确率,更直观展现网络模型的分类识别效果。测试时,自行选择切换加载训练好的 Inception-V3 或者 ResNet50 网络模型。通过可视化界面,任意选择测试数据集的图片输进网络中,即可得到图片所属的类别的概率及预测其类别,通过一系列预测评估网络模型。

随机选取测试集中的6类垃圾进行实验,从每个类别中任意选择一张图片,分别在训练好的 Inception-V3 和 ResNet50 中测试,获得该图像预测的类别及其在该类别的最大准确率,测试结果如表6所示。可以直观地比较出:对于同一张图片,ResNet50 中的测试准确率优于 Inception-V3。

表6 6类垃圾准确率测试结果 单位:%

类别	Inception-V3	ResNet50
cardboard	85.43	96.58
glass	87.59	97.82
metal	88.47	93.24
paper	87.15	96.33
plastic	86.93	95.42
trash	85.41	95.73

以 glass 类别为例,Inception-V3 和 ResNet50 的可视化界面如图4~5所示。



图4 Inception-V3 的 glass 类测试结果



图5 ResNet50 的 glass 类测试结果

4 结束语

将人工智能与垃圾分类问题相结合是可以实现的。与传统的垃圾分类方法相比,利用深度学习的卷积神经网络的垃圾分类方法具有较高的准确率,能够实现垃圾分类的目的,并且能够节约大量的时间和人工成本。文中探讨了两种卷积神经网络的垃圾分类模型,结果表明 Inception-V3 和 ResNet50 都能够实现垃圾分类的目的,两种网络模型均能用于垃圾分类领域更深层次的研究中。对两种网络的结构进行分析和实验对比,可以看出,使用 Inception-V3 和 ResNet50 网络的垃圾分类模型都能得到较高准确率和较低损失函数值。其中,ResNet50 的准确率高于 Inception-V3,但 ResNet50 网络的性能没有 Inception-V3 平稳,且 Inception-V3 的参数数量小于 ResNet50,模型收敛速度更快。

致谢:感谢电子科技大学网络与数据安全四川省重点实验室开放课题项目(NDS2021-7)对本文的资助

参考文献:

[1] 张怀予. 基于物联网和图像识别的垃圾分类回收系统[C]. 物联网与无线通信-2018 年全国物联网技术与应用大会,2018(11):111-113.

[2] 叶志祥. 智能分类垃圾桶设计研究[J]. 中国资源综合利用,2019,37(4):191-193.

[3] 黄兴华,叶军一,熊杰. 基于纹理特征融合的道路垃圾图像识别及提取[J]. 计算机工程与设计,2019,40(11):3212-3218.

[4] 秦斌斌,何级,基于卷积神经网络的垃圾分类研究[J]. 无线通信技术,2019(3):51-56.

[5] 张斌,王强. 一种改进型卷积神经网络的图像分类方法[J]. 成都信息工程大学学报,2019,34(1):39-43.

[6] 王文文,陶宏才. 基于优化 VGG19 卷积神经网络的异常检测模型研究[J]. 成都信息工程大学学报,2020,35(3):253-258.

- [7] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks[C]. Proceeding of the Advances in Neural Information Processing Systems. LakeTanhoe, USA, 2012:1097–1105.
- [8] K Simonyan, A Zisserman. Very deep convolutional networks for large-scale image recognition [C]. arXiv preprint arXiv, 2014, 1409–1556.
- [9] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Boston, Massachusetts, USA: IEEE, 2015.
- [10] S Ioffe, C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]. In Proceedings of The 32nd International Conference on Machine Learning, 2015:448–456.
- [11] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2818–2826.
- [12] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning [J]. arXiv:1602.07261, 2016.
- [13] K He, X Zhang, S et al. Deep residual learning for image recognition [J]. arXiv preprint arXiv: 1512.03385, 2015.
- [14] 金钊. 基于 TensorFlow 的不同深层卷积神经网络的对比与分析[J]. 电子世界, 2018(6): 25–26.
- [15] 王文成, 蒋慧, 乔倩, 等. 基于 ResNet50 网络的十种鱼类图像分类识别研究[J]. 农村经济与科技, 2019, 30(19): 60–62.

Research and Application of Garbage Image Classification Model based on Convolutional Neural Network

TANG Kangjian, WEN Zhan, LI Wenzao

(College of Communication Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Traditional garbage classification methods often rely on sensors to complete garbage identification and classification, but there are problems such as low classification accuracy, complex models, and lack of efficient operability. In order to solve this problem, a garbage classification method combined with Convolutional Neural Networks (CNN) is proposed. Two convolutional neural networks, Inception-V3 and ResNet50, with high-efficiency feature extraction performance, were used to train Huawei's public garbage data set Garbage Data, and a garbage classification model was established. Experiments show that the accuracy rates of Inception-V3 and ResNet50 training on the training set are 89.9% and 95.1%, respectively, and the cross entropy loss functions are 1.463 and 1.363, respectively. Using the visual interface to verify the six random single images in the test set, the accuracy of ResNet50 is higher than that of Inception-V3. But ResNet50 is not as stable as Inception-V3, and the accuracy curve of Inception-V3 is smoother. Inception-V3 converges faster than ResNet50 and consumes less resources.

Keywords: garbage classification; convolutional neural network; accuracy; the cross entropy; Inception-V3; ResNet50