

# 基于卷积神经网络的垃圾图像分类算法<sup>①</sup>



董子源<sup>1,2</sup>, 韩卫光<sup>2</sup>

<sup>1</sup>(中国科学院大学, 北京 100049)

<sup>2</sup>(中国科学院 沈阳计算技术研究所, 沈阳 110168)

通讯作者: 董子源, E-mail: ronaldo0819@126.com

**摘要:** 垃圾分类作为资源回收利用的重要环节之一, 可以有效地提高资源回收利用效率, 进一步减轻环境污染带来的危害. 随着现代工业逐步智能化, 传统的图像分类算法已经不能满足垃圾分拣设备的要求. 本文提出一种基于卷积神经网络的垃圾图像分类模型 (Garbage Classification Network, GCNet). 通过构建注意力机制, 模型完成局部和全局的特征提取, 能够获取到更加完善、有效的特征信息; 同时, 通过特征融合机制, 将不同层级、尺寸的特征进行融合, 更加有效地利用特征, 避免梯度消失现象. 实验结果证明, GCNet 在相关垃圾分类数据集上取得了优异的结果, 能够有效地提高垃圾识别精度.

**关键词:** 垃圾分类; 卷积神经网络; 图像分类; 注意力机制; 特征融合

引用格式: 董子源, 韩卫光. 基于卷积神经网络的垃圾图像分类算法. 计算机系统应用, 2020, 29(8): 199-204. <http://www.c-s-a.org.cn/1003-3254/7581.html>

## Classification Algorithm of Garbage Images Based on Convolutional Neural Network

DONG Zi-Yuan<sup>1,2</sup>, HAN Wei-Guang<sup>2</sup>

<sup>1</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

<sup>2</sup>(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

**Abstract:** Garbage classification, as one of the important links of resource recycling, can effectively improve the efficiency of resource recycling and further reduce the harm caused by environmental pollution. With the development of modern industry, traditional image classification algorithm cannot meet the requirements of garbage sorting equipment. This study proposes a garbage classification model based on convolutional neural networks (Garbage Classification Network, GCNet). By constructing the attention mechanism, the model completes extracting the local and global features and can obtain perfect and effective feature information. At the same time, the feature fusion mechanism is used to fuse features at different levels and sizes, which can effectively use features and prevent gradient from vanishing. The experimental results prove that GCNet has achieved excellent results on garbage classification datasets and can effectively improve the accuracy of garbage classification.

**Key words:** garbage classification; convolutional neural network; image classification; attention mechanism; feature fusion

### 1 引言

垃圾回收利用作为发展循环经济的必经之路, 是根治污染、提高环境治理效果的关键所在. 随着我国

生产力水平的发展, 生活垃圾、工业垃圾数量不断增加, 困扰着很多城市. 据统计, 仅 2018 年, 中国垃圾的清运量已经达到了 2.28 亿吨<sup>[1]</sup>. 在人们将垃圾投放进

① 收稿时间: 2020-02-08; 修改时间: 2020-03-03; 采用时间: 2020-03-20; csa 在线出版时间: 2020-07-29

垃圾箱之后,垃圾被运送到垃圾处理厂统一处理.当前国内的垃圾处理厂,更多依靠人工在流水线上作业去分拣垃圾,对作业者健康不利且分拣效率较低,已不能满足大量垃圾处理需求.此外,人工分拣的垃圾种类极为有限,大部分垃圾无法重新回收利用,造成很大浪费.随着深度学习技术的发展,卷积神经网络使图像分类算法在精度和速度上得到了巨大的提升,让我们看到了借助视觉技术自动分拣垃圾的可能性.通过摄像头拍摄垃圾图片,利用卷积神经网络检测出垃圾的类别,之后就可以借助机械手或推板自动完成分拣任务,可以降低人工成本,提高分拣效率.因此,开展垃圾图像分类算法的研究,具有重要的应用价值.

## 2 相关工作

早期,学者们只能借助经典的图像分类算法<sup>[2-5]</sup>完成垃圾图像分类任务,这要通过手动提取的图像特征并结合相应的分类器完成.吴健等<sup>[6]</sup>利用颜色和纹理特征,初步完成了废物垃圾识别.由于不同数据集的图像背景、尺寸、质量不尽相同,传统算法需要根据相应数据人工提取不同的特征,算法的鲁棒性较差,并且处理方式复杂,所需时间较长,无法达到实时的效果.随着卷积神经网络(Convolution Neural Network, CNN)的飞速发展,深度学习广泛应用于图像识别领域.作为数据驱动的算法,CNN具有强大的特征拟合能力,可以有效、自动地提取图像特征,并具有较快的运行速度.2012年,AlexNet<sup>[7]</sup>取得了ImageNet图像分类竞赛的冠军,标志着深度学习的崛起.随后几年,GoogleNet<sup>[8]</sup>、VGGNet<sup>[9]</sup>、ResNet<sup>[10]</sup>等算法提升了图像分类的精度,并成功应用于人脸识别、车辆检测等多个领域.垃圾图像分类,在深度学习算法的帮助下同样取得了较大的突破.斯坦福大学的Yang等建立了TrashNet Dataset公开数据集,包含6个类别,共计2527张图片.Ozkaya等<sup>[11]</sup>通过对比不同CNN网络的分类能力,搭建神经网络(本文称之为TrashNet)并进行参数微调,在数据集TrashNet Dataset上取得了97.86%的准确率,是目前这一数据集上最佳分类网络.在非公开数据集方面,Mittal等<sup>[12]</sup>自制了2561张的垃圾图片数据集GINI,使用GarbNet模型,得到了87.69%的准确率.国内方面,郑海龙等<sup>[13]</sup>用SVM方法进行了建筑垃圾分类方面的研究.向伟等<sup>[14]</sup>使用分类网络CaffeNet,调整卷积核尺寸和网络深度,使其适用于水面垃圾分

类,在其自制的1500张图片数据集上取得了95.75%的识别率.2019年,华为举办垃圾图像分类竞赛,构建了样本容量为一万余张的数据集,进一步促进了该领域的发展.

我国各地区生活垃圾分类标准有所不同,大致可分为可回收垃圾、有害垃圾、厨余垃圾和其他垃圾这4大类,且每个类别下又包含若干子类别,种类繁多且十分复杂.按照这样的分类标准做的垃圾图像识别研究,国内目前还处于起步阶段.现有的图形分类算法在垃圾处理领域的应用较少,且存在准确率不足、泛化性能差、处理效率低的缺点.针对现有方法的不足,本文提出一种基于卷积神经网络的垃圾图像分类算法(Garbage Classification Net, GCNet),在网络结构中融合了注意力机制模块与特征融合模块,提高了模型在垃圾分类任务上的准确性与鲁棒性.

## 3 算法设计

### 3.1 模型结构

本文构建的GCNet模型包括特征提取器、分类器两部分,整体结构如图1所示.图中特征提取器由Resnet101作为主干部分,共包括5个bottleneck,并在不同的bottleneck后加入注意力机制模块,同时对不同模块提取到的特征进行特征融合(如图1中虚线所示)以从输入 $x$ 中提取图像的特征信息 $F_1$ :

$$F_1 = M_e(x) \quad (1)$$

其中, $M_e$ 表示特征提取器.

分类器由两层全连接层和一个Softmax分类器组成,对提取到的特征信息 $F_1$ 进行分类,以得到图像在每个类别下的最终得分 $y_i$ :

$$y_i = M_c(F_1), i \in \{1, 2, \dots, n\} \quad (2)$$

其中, $M_c$ 表示分类器.

### 3.2 注意力机制

注意力机制源于对人类视觉的研究,人类会根据实际需要选择视网膜内特定区域进行集中关注,可以将有限的处理资源分配至重要的部分.由于相同类别垃圾的特征表征差异性可能较大,不利于图片的正确分类,这就要求准确地关注图像中的显著区域.受这一思想的启发,通过构建注意力机制模块,使网络模型重点关注有利于分类的特征区域,以实现更好的特征提取功能,其具体结构如图2所示.

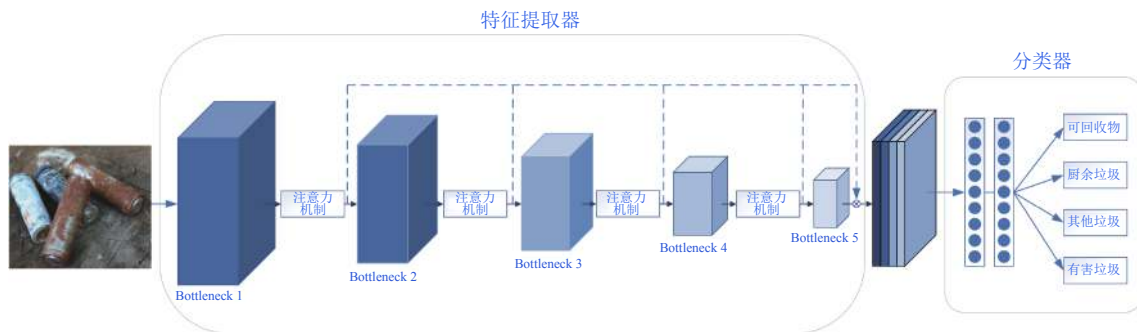


图1 GCNet网络结构图

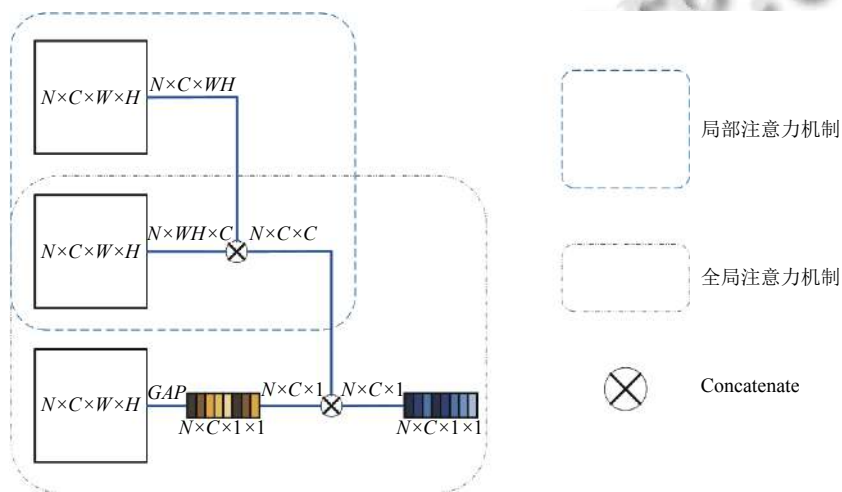


图2 注意力机制示意图

将每个 bottleneck 提取到的特征以  $F_i$  表示, 其大小为  $N \times C \times W \times H$ . 利用格拉姆矩阵<sup>[15]</sup> 构建局部注意力机制模块, 将  $F_i$  与其转置  $F_i^T$  相乘, 获得大小为  $N \times C \times C$  的局部特征  $F_{local}^i$ :

$$F_{local}^i = F_i^T \cdot F_i \quad (3)$$

该操作可以获得特征图中每个元素自身间的关联, 使原本数值较大的元素越大、数据较小的元素越小, 即凸显每个局部的重要程度, 使有助于判断类别的特征更加明显, 同时抑制影响判断的特征.

随后, 对  $F_i$  进行全局平均池化操作 (Global Average Pooling, GAP)<sup>[16]</sup>, 可以保留特征  $F_i$  提取到的空间信息和语义信息, 并获得大小为  $N \times C \times 1 \times 1$  的全局特征  $F_{global}^i$ :

$$F_{global}^i = GAP(F_i) \quad (4)$$

$F_{global}^i$  代表了特征  $F_i$  中最显著的特征, 这也符合所定义的全局注意力机制.

最后, 将局部特征  $F_{local}^i$  与全局特征  $F_{global}^i$  相乘, 从而获得包含局部和全局信息的总体特征  $F_{fusion}^i$ :

$$F_{fusion}^i = F_{local}^i \cdot F_{global}^i \quad (5)$$

### 3.3 特征融合机制

垃圾分类任务中, 从属于同一类别的垃圾往往差异性较大, 如可回收物中包括的纸板箱和玻璃杯, 二者存在较大的外观差异但却属于同一类别, 这会给分类器的判别增加一定的难度. 此外, 随着网络层数的加深, 单一的图像特征会丢失某些区域的信息, 从而导致模型的性能变差. 针对这一问题, 常用的方法是通过不同卷积核的池化操作构建多尺度的特征融合模块, 但这种操作会增大模型的运算量, 同时也会使模型训练更加困难.

由于构建的特征提取器中包含了不同的池化操作, 因此只需要提取不同 bottleneck 下的特征便能获得不同尺寸的特征. 在此基础上, 本文提出了改进的特征融合机制, 如图 1 中的虚线所示, 将每个 bottleneck 经过

注意力机制后所提取到的总体特征 $F_{fusion}^i$ 进行融合:

$$F_1 = w_c * Concat(F_{fusion}^1, \dots, F_{fusion}^5) + b_c \quad (6)$$

其中,  $w_c$ 为 $1 \times 1$ 的卷积核,  $b_c$ 为偏置,  $Concat$ 表示融合操作. 该操作旨在利用不同尺度的特征信息, 避免信息的丢失, 进一步提升模型的鲁棒性.

## 4 实验及结果分析

### 4.1 实验平台

本文实验在 Ubuntu16.04 系统下, 利用 Python 语言, Pytorch 深度学习框架完成. 硬件环境为 CPU Intel I7-9700K, 内存 32 GB, 显卡 Nvidia GeForce RTX 2080Ti.

### 4.2 实验数据

本文采用华为垃圾分类挑战杯数据集, 该数据集已全部标注类别, 包括厨余垃圾、可回收物、其他垃圾和有害垃圾 4 个大类, 每个大类下又包含若干子类别, 合计 40 个小类别, 共计 14 683 张图片.

在实验中, 采用 4:1 的比例划分数据集, 80% 作为训练数据, 20% 作为测试数据. 此外, 为了增强模型的泛化能力和鲁棒性, 还对训练样本进行了数据增强操作, 包括随机旋转、随机翻折、随机裁剪等.

### 4.3 结果分析

实验选用 ADAM<sup>[17]</sup> 优化算法训练模型, 动量系数为 0.9, 共设置 50 个迭代周期, 初始学习率设置为 0.01, 每隔 10 个迭代周期学习率衰减 0.1 倍, 一阶矩估计的指数衰减率为 0.99, 二阶矩估计的指数衰减率为 0.999. 此外, 使用交叉熵损失函数来训练优化模型.

在训练过程中, 对 GCNet 模型通过消融实验以分别验证注意力机制和特征融合机制的功能. 实验 1 (Experiment\_1) 为含有注意力机制和特征融合机制的模型, 实验 2 (Experiment\_2) 为只含有注意力机制的模型, 实验 3 (Experiment\_3) 为只含有特征融合机制的模型, 实验 4 (Experiment\_4) 为不含有注意力机制和特征融合机制的模型. 各个实验的训练过程迭代曲线如图 3 所示.

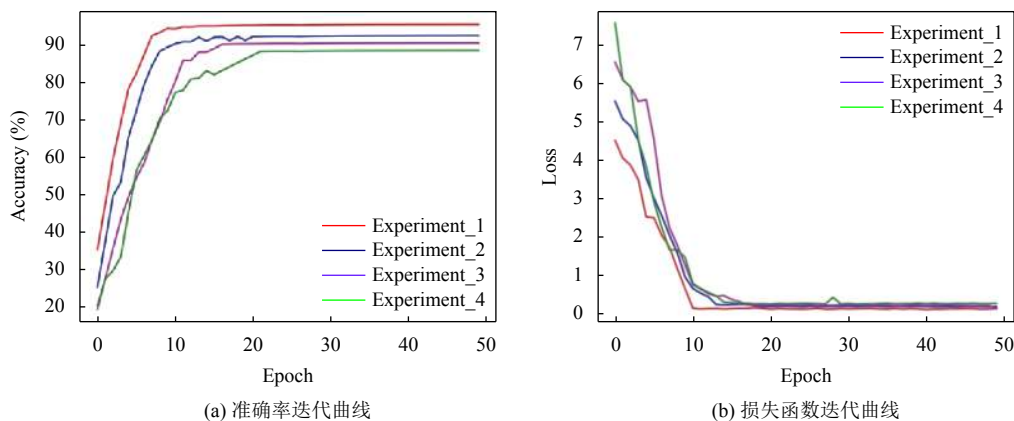


图3 GCNet 训练过程曲线

从图 3 中可以看到, 实验 1 的准确率提升与损失值下降都快于其他 3 个实验, 说明训练过程收敛更快. 实验 2 和实验 3 的速度几乎相近, 而实验 4 的速度最慢. 为了进一步验证上述训练的准确性, 在对应的测试集上进行测试, 实验结果如表 1 所示. 包含注意机制模块和特征融合模块的实验 1 取得了 96.73% 的最优准确率, 同时在各个子类别上均取得了最高的准确率, 说明本文构建的模型泛化能力较好. 在类内差异性较大的“其他垃圾”这一类别中, 注意力机制和特征融合机制均能够显著提高模型的准确率.

实验序号	可回收垃圾	有害垃圾	厨余垃圾	其他垃圾	平均准确率
1	95.21	96.35	97.82	97.54	96.73
2	94.53	90.35	93.12	90.52	92.13
3	93.63	90.65	92.36	88.48	91.28
4	90.15	90.34	88.12	86.06	88.67

本文模型与其他模型的准确率对比结果如表 2 所示. 可以看出, 本文构建的 GCNet 平均准确率最高, 在各个类别中也取得了最高的准确率, 说明本文构建的

注意力机制和特征融合机制充分提取了有利于图像分类的特征,使得分类结果更加准确. TrashNet 的平均准确率略差于 GCNet, 而 CaffeNet 结果最差.

表 2 不同模型实验准确率对比 (单位: %)

模型	可回收垃圾	有害垃圾	厨余垃圾	其他垃圾	平均准确率
GCNet	95.21	96.35	97.82	97.54	96.73
TrashNet	92.78	93.24	93.06	91.56	92.66
CaffeNet	90.53	89.19	93.81	87.43	90.24

图 4 为 3 种算法 GCNet, TrashNet 和 CaffeNet 对同样 4 张图片的测试结果, 显示了预测类别以及属于这一类别的概率. GCNet 在各个类别的样例中, 也取得了最好的结果. 特别是在第 4 幅 (其他垃圾) 的测试中, 由于该类别的图像类内差异过大, CaffeNet 将其识别为其他垃圾的概率仅为 62.75%, 将其错误地识别成有害垃圾的概率却高达 35.61%, 在某些干扰情况下, 这可能会导致判定错误, 对分类结果造成较大影响. 本文算法成功的改进了这一问题, 确保了在较难分辨的类别上的准确率.

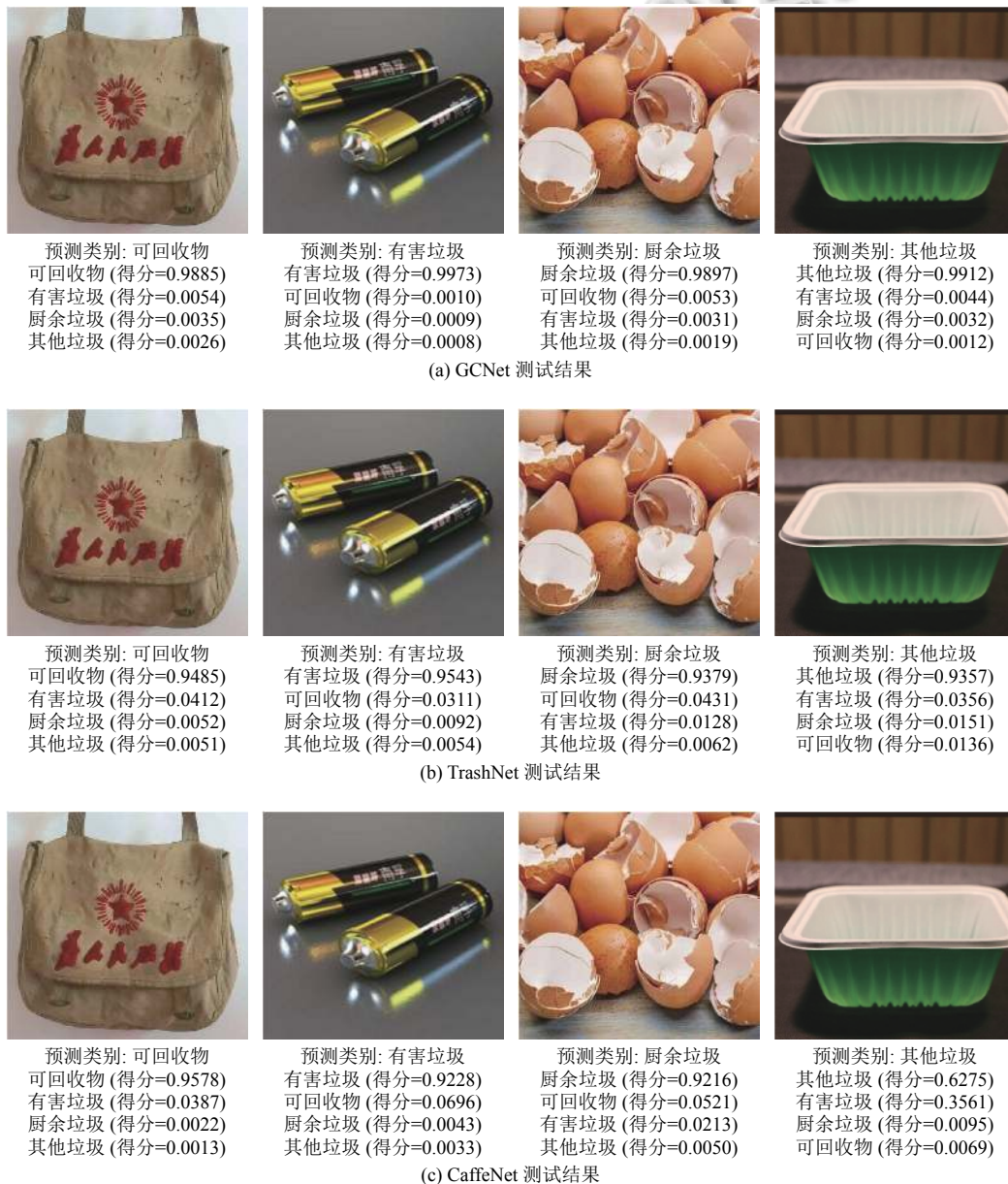


图 4 各模型测试结果对比图

## 5 结论

本文针对垃圾图像分类问题,构建了一种基于卷积神经网络的算法 GCNet,该网络通过构建注意力机制和特征融合机制,能够有效地提取图像特征、降低类别差异性带来的影响,并在相关数据集上取得了96.73%的平均准确率,相较于现有的分类算法提升了约4%的准确率,满足了实际的应用需求,具有良好的应用前景。

### 参考文献

- 1 吕思敏. 以史为鉴, 开启垃圾分类新时代. 城乡建设, 2020, (3): 30–32.
- 2 Abeywickrama T, Cheema MA, Taniar D. K-nearest neighbors on road networks: A journey in experimentation and in-memory implementation. Proceedings of the VLDB Endowment, 2016, 9(6): 492–503. [doi: [10.14778/2904121.2904125](https://doi.org/10.14778/2904121.2904125)]
- 3 Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91–110. [doi: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94)]
- 4 Harri C, Stephens M. A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference. Manchester, UK. 1988. 207–217.
- 5 Vapnik V. Statistical Learning Theory. New York: Wiley, 1998. 401–492.
- 6 吴健, 陈豪, 方武. 基于计算机视觉的废物垃圾分析与识别研究. 信息技术与信息化, 2016, (10): 81–83. [doi: [10.3969/j.issn.1672-9528.2016.10.020](https://doi.org/10.3969/j.issn.1672-9528.2016.10.020)]
- 7 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1106–1114.
- 8 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. arXiv: 1409.4842, 2014.
- 9 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2015.
- 10 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 11 Ozkaya U, Seyfi L. Fine-tuning models comparisons on garbage classification for recyclability. arXiv: 1908.04393, 2019.
- 12 Mittal G, Yagnik KB, Garg M, *et al.* SpotGarbage: Smartphone app to detect garbage using deep learning. Proceedings of 2016 ACM International Joint Conference. Heidelberg, Germany. 2016. 940–945.
- 13 郑龙海, 袁祖强, 殷晨波, 等. 基于机器视觉的建筑垃圾自动分类系统研究. 机械工程与自动化, 2019, (6): 16–18. [doi: [10.3969/j.issn.1672-6413.2019.06.006](https://doi.org/10.3969/j.issn.1672-6413.2019.06.006)]
- 14 向伟, 史晋芳, 刘桂华, 等. 改进 CaffeNet 模型在水面垃圾识别中的应用. 传感器与微系统, 2019, 38(8): 150–152, 156.
- 15 Zhang XK, Wang Y, Gou MR, *et al.* Efficient temporal sequence comparison and classification using gram matrix embeddings on a riemannian manifold. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 4498–4507.
- 16 M, Q, Yan SC. Network in network. arXiv: 1312.4400, 2014.
- 17 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980, 2017.