

基于迁移学习的多场景垃圾图像分类方法

彭 治¹, 刘 杨¹, 杜永萍¹, 常燕青², 韩红桂¹

(1. 北京工业大学 信息学部, 北京 100124;

2. 江苏省固体废物处理环保装备工程技术研究中心, 江苏 常州 213000)

摘 要:垃圾分类回收是资源循环再生链条中重要的一环,传统的人工分类方法效率低,而基于神经网络模型的图像识别技术是实现智能化分类回收的重要策略,具有准确度高、泛化性能好等优点。针对生活垃圾场景的多样性以及垃圾图像数据集缺乏的问题,该文在基于单一背景和复杂背景两种不同垃圾图像场景下,提出一种基于预训练 ResNeXt101 模型进行迁移学习的方法。利用大规模 ImageNet 图像数据集训练模型参数,并将参数迁移到垃圾图像数据分类模型中进行训练,解决缺乏大规模标注数据的垃圾图像数据集的分类问题,实现玻璃和纸类等 4 种可回收物的分类识别以及厨余垃圾中杂质的检测。实验结果表明,利用该模型在真实背景下的垃圾图像分类准确率可以达到 87.42%,单一背景下分类准确率达到 96.19%,实现了高精度的可回收物识别分类以及厨余垃圾中高效的杂质检测,同时有效提升了模型训练速度。

关键词:图像识别;迁移学习;残差网络;ResNeXt;垃圾分类

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)05-0106-06

doi:10.3969/j.issn.1673-629X.2022.05.018

Multi-scene Garbage Image Classification Method Based on Transfer Learning

PENG Zhi¹, LIU Yang¹, DU Yong-ping¹, CHANG Yan-qing², HAN Hong-gui¹

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

2. WELLE Environmental Group Co., Ltd, Changzhou 213000, China)

Abstract: Garbage classification and recycling is an important part of the resource recycling chain. Traditional classification depends on the manual intervention mostly, which is inefficient, while image recognition technology based on neural network model is an important strategy to realize intelligent classification and recycling with the advantages of high accuracy and good generalization performance. In response to the diversity of domestic waste scenes and the lack of waste image datasets, a transfer learning method based on a pre-trained ResNeXt101 model with ImageNet image dataset is proposed to solve the problem of classifying and recognizing four categories of recyclable household garbage, such as glass and paper, and detecting whether impurity is mixed in the kitchen garbage under two different image scenarios, including a single background and a complex background. The experimental results show that the garbage image classification based on this model can achieve 87.42% accuracy in the complex background and 96.19% accuracy in the flat background, achieving high accuracy in the identification and classification of recyclables, and efficient detection of impurities in kitchen garbage, while further the training speed of the model is improved.

Key words: image recognition; transfer learning; residual network; ResNeXt; garbage classification

0 引 言

随着日常生活中产生的垃圾不断增加,对垃圾进行分类管理是实现垃圾资源回收利用,同时达到垃圾减量的重要途径。传统的垃圾分类依靠人工手动进行,需要耗费大量的人力物力,分类效率低。图像识别是实现智能化垃圾分类的重要方法,基于神经网络模

型的图像识别与分类技术具有准确度高、泛化性能好等优点。

近年来在图像识别领域,深度学习网络应用较广,并在数据充足时有稳定的表现。目前,将图像识别技术应用于垃圾分类中也受到了学者广泛的关注与研究^[1]。对于规模较小的数据集,深度神经网络模型在

收稿日期:2021-05-27

修回日期:2021-09-27

基金项目:国家重点研发计划(2019YFC1906002,2018YFC1900804)

作者简介:彭 治(1997-),男,硕士研究生,通讯作者,研究方向为智能信息处理与信息检索;杜永萍,博士,教授,研究方向为信息检索、信息抽取与自然语言处理。

训练过程中容易产生过拟合的现象,而通过数据增强的方法可增加训练样本,有效缓解过拟合的问题^[2-3]。此外,提高模型的泛化能力是深度神经网络模型最困难的挑战之一^[3]。对于模型,不仅要求它对训练数据集有很好的拟合,同时也希望它可以在测试数据集上有很好的性能。数据增强可以实现数据更复杂的表征,让模型更好地学习迁移数据集上的数据分布^[4-5]。目前,在图像分类模型中都普遍应用了数据增强技术,Krizhevsky 等人提出的 AlexNet^[6]中,应用数据增强增加 ImageNet 数据集样本从而避免神经网络训练过程中产生过拟合。G. Huang 等人提出的 DenseNet^[7]中,对 CIFAR 数据集采用标准的数据增强方案(镜像/移位)提高模型的泛化能力。

随着传统神经网络模型的深度不断增加,训练过程中会出现梯度消失、梯度爆炸、模型过拟合等问题^[8]。通过正则化的方法对大量数据进行训练可以有效地解决过拟合问题,在一定程度上缓解浅层网络的梯度消失/梯度爆炸问题,更快地训练稳定的网络^[9-10]。如果网络层数继续简单增加,会出现网络退化,上百层的传统神经网络将难以训练。He 等人^[11]通过引入恒等映射,实现跳层连接,缓解了深度网络的退化问题,深度残差网络(deep residual networks, ResNet)逐渐成为了图像分类的核心技术。Inception^[12-13]使用3个不同大小的滤波器对输入执行卷积操作,所有子层的输出最后会被级联,并传送到下一个 Inception 模块,实现了特征的复用,提高了特征张量宽度以及对特征的囊括性^[13],也减缓了加深神经网络模型引起的梯度消失/梯度爆炸问题。

ResNeXt^[14]是 ResNet 和 Inception 的结合体,不同于 Inception v4^[13]的是,ResNeXt 不需要人工设计复杂的 Inception 结构细节,而是每个分支都采用相同的拓扑结构。ResNeXt 的本质是分组卷积^[15],通过变量基数来控制分组的数量。传统的模型通过加深或加宽网络来提高性能,但是随着超参数的增加,网络设计的难度和计算开销也会增加,而 ResNeXt 模型在不增加参数复杂度的前提下可以有效提高准确率,同时减少了超参数的数量。Lin 等人^[16]提出一种新的基于 ResNeXt 弱监督模型的染色体簇类型识别方法,识别准确率、灵敏度以及特异性都达到了较高的效果。

迁移学习是运用已有的知识对不同但相关领域问题进行求解的一种机器学习方法^[17]。深度学习模型能够从海量数据中学习高级特征,但数据依赖是一个严重问题。与传统的机器学习方法相比,深度学习模型需要大规模数据来学习数据间的潜在关联^[18]。在某些领域由于数据采集和标注费用高昂,构建大规模的高质量标注数据集非常困难。迁移学习放宽了训练

数据必须与测试数据独立同分布的假设,因此利用迁移学习可以有效缓解训练数据不足的问题。例如,胡等人^[19]提出了动态采样与迁移学习结合的疾病预测模型,在规模较大的多类别疾病数据集上训练预测模型,然后迁移到样本不足的疾病预测模型中,从而缓解数据不足的问题。迁移学习把训练好的模型(预训练模型)迁移到新的模型来进行训练^[20-21],将已经学到的模型参数通过某种方式分享给新模型从而加快并优化模型的学习效率,而不需要对目标域内的新模型从零开始训练,显著地降低了对目标域内训练数据需求^[22]。

该文将基于深度学习的图像识别技术应用于垃圾智能分类任务,采用 ResNeXt 模型提取有效特征,对于训练数据规模有限的问题,在利用数据增强技术扩大训练集规模的同时,进一步通过迁移学习的策略提升模型性能,实现垃圾图像的智能分类。

1 基于迁移学习的垃圾图像分类模型

该文提出了一种基于迁移学习技术的深度神经网络垃圾分类模型,ResNeXt 网络在图像分类中具有高效性,采用该网络进行可回收垃圾图像的识别分类,模型结构如图1所示。在大规模数据集 ImageNet 上进行模型训练,并将其迁移到垃圾图像数据集,实现可回收物的准确识别分类。

为了进一步提升模型的分类准确率,对采集到的垃圾图像采用如下的方式进行数据增强,包括比例缩放、水平/垂直翻转、图像随机裁剪等不同的数据增强技术,获得丰富的训练样本用于模型训练,提升模型的鲁棒性和泛化能力。

(1) 利用 transforms.Resize() 函数重新设置大小,统一输入图像大小;

(2) 利用 transforms.RandomCrop(size) 函数对输入的图像进行随机裁剪,设置裁剪的大小(size)为300,将输入的图随机裁剪为300*300的图像;

(3) 利用 transforms.RandomVerticalFlip() 和 transforms.RandomHorizontalFlip() 函数对随机裁剪后的图像分别依照默认的概率进行垂直翻转和水平翻转;

(4) 利用 transforms.RandomRotation() 函数对翻转后的图像进行随机旋转一定角度;

(5) 利用 transforms.Normalize() 和 transforms.ToTensor() 函数对旋转变换后的图像进行归一化,使模型训练时,梯度对每张图片的作用都是平均的。

经过数据预处理实现数据增强后,将图像输入模型进行训练,训练集更加丰富,使模型具有较好的泛化能力。

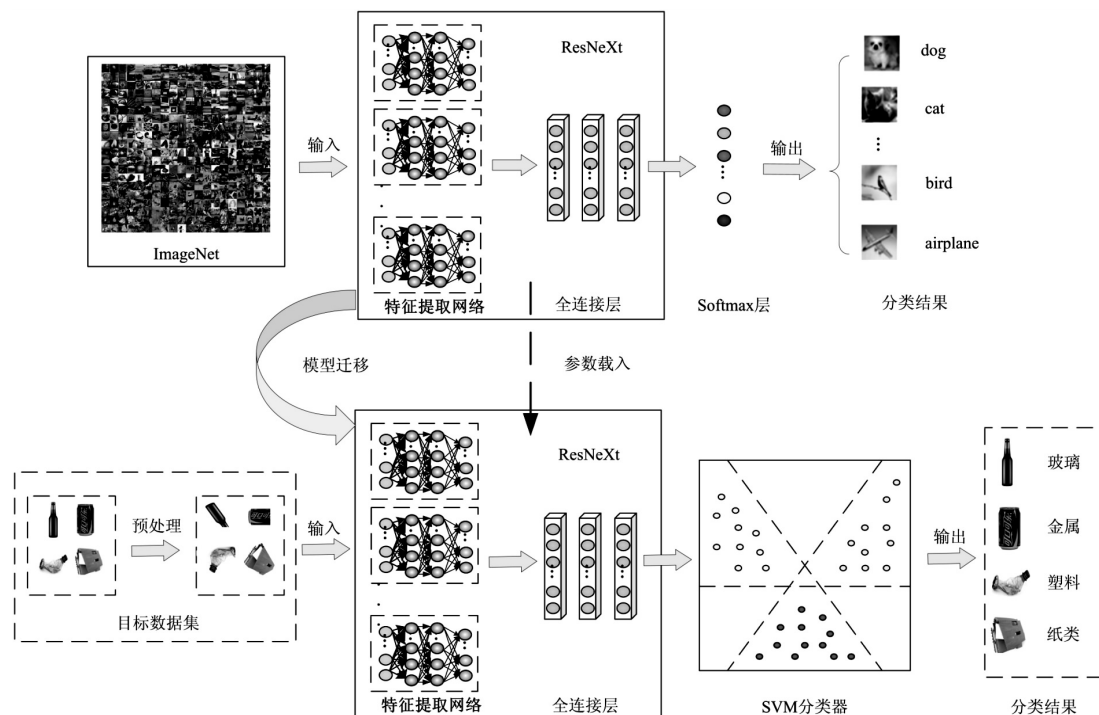


图1 基于迁移学习技术的垃圾图像分类模型

最终的形式,如公式(2):

$$y = x + \sum_{i=1}^c T_i(x) \quad (2)$$

其中, y 为输出。

ResNeXt 采用同构多分支的结构,使模型有更少的超参数;同时引入基数,基数增加可提高模型分类效果,比传统的加深或加宽网络模型有更好的性能。在模型迁移过程中调参过程更简单并具有更强的鲁棒性。

2.2 SVM 分类器实现图像分类

传统的卷积神经网络分类模型提取图像特征后,通常采用 softmax 分类层进行分类,该文使用 SVM 代替 softmax 分类层。

支持向量机属于线性分类器,它将一系列的输入数据映射到类别上,通过寻找超平面实现对样本的分类。支持向量机的预测输出是 $f(x) = Wx + b$ 。该文采用多类 SVM 损失函数作为目标函数:

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \nabla) \quad (3)$$

$$L = \frac{1}{N} \sum_{i=1}^N L_i \quad (4)$$

其中, y_i 表示真实的类别, s_{y_i} 表示在真实类别上的评分, s_j 表示预测错误的评分。

对于一个样本,模型预测的结果是一个评分,通过优化损失函数,分类器函数拟合这些样本,使大部分样本输出满足 $f(x) > 0$ 。若图像恰好落在分类超平面,其评分等于 0,离超平面越远,评分的绝对值也就越大,箭头方向指向评分的正增长方向,表示正确的分类

2 基于残差网络的特征提取与分类

2.1 特征提取网络

在分类模型中,采用残差网络 ResNet 和 Inception 的结合体,即 ResNeXt,实现特征提取。在大规模图像分类数据集 ImageNet 上进行预训练得到 ResNeXt 的网络结构和参数。

深度神经网络中最简单的神经元执行内积(加权求和)是全连接层和卷积层的基本变换。神经元的内积操作可以看作是一个分割-变换-聚合的过程:首先,向量 X 被分割成了 D 个低维嵌入 x_i ;其次,模型对低维表示进行变换,分别与相应的权重进行点积得到 $w_i x_i$;最后,把全部低维嵌入聚合得到 $\sum_{i=1}^D w_i x_i$ 。

ResNeXt 利用聚合变换的思想,采用一个更通用的函数代替初等变换 ($w_i x_i$),它本身也可以是一个网络,聚合转换可以表示为:

$$F(x) = \sum_{i=1}^c T_i(x) \quad (1)$$

其中, $T_i(x)$ 可以是任意函数,类似于一个简单的神经元, T_i 将 x 投影到一个(可选的低维)嵌入表示中,然后进行转换。

在公式(1)中, C 是要聚合的转换集的数量,称为基数。虽然宽度的维度与简单变换(内积)的数量有关,但基数的维数控制着更复杂的变换的数量。实验表明,基数是一个基本的维度,比宽度和深度的维度更有效。

最后,公式(1)中的聚合转换用残差函数,得到

结果。

3 实验结果与分析

针对垃圾图像数据中存在的单一背景与复杂背景两种不同场景,分别进行了实验。在单一背景中,每张垃圾图像只有一个目标物体;在复杂背景中,每张垃圾图像中有多个物体,包含复杂的背景干扰。该文同时针对厨余垃圾中是否混有杂质(如:可回收物)的问题,利用图像分类技术进行检测。

3.1 实验环境

模型训练与测试均是在 Pytorch 的深度学习框架下完成的。实验平台为 Ubuntu 系统、Nvidia GTX1080 Ti GPU,12 GB 显存和 CUDA 11.1。模型的训练与测试均通过 GPU 加速。

3.2 数据集

(1)单一背景数据。

在单一背景图像数据的实验中所用到的垃圾图像数据是在 kaggle 和华为云等平台进行采集,从中筛选出玻璃、金属、塑料、纸类 4 类可回收垃圾图像进行识别。

(2)复杂背景数据。

在复杂背景实验中所用到的垃圾图像数据是通过人工采集到的垃圾图像,包括纯厨余垃圾和非纯厨余垃圾 2 类图像数据。

采集到的 4 类单一背景的可回收垃圾图像样本共 9 620 张,2 类复杂背景图像样本共 4 681 张,图像数据分类统计如表 1 所示。

表 1 单一背景和复杂背景图像数据分类统计

类型	图像种类	图像数量
单一背景	玻璃	2 922
	金属	2 203
	塑料	1 973
	纸类	2 522
复杂背景	纯厨余	1 924
	非纯厨余	2 757

3.3 评价指标

采用准确率(Accuracy)来对测试结果进行评价。

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (5)$$

其中,TP(true positives)为正确地划分为正例的个数,即实际为正例且被分类器划分为正例的样本数;FP(false positives)为被错误地划分为正例的个数,即实际为负例但被分类器划分为正例的样本数;FN(false negatives)为被错误地划分为负例的个数,即实际为正例但被分类器划分为负例的样本数;TN(true negatives)为被正确地划分为负例的个数,即实际为负例且被分类器划分为负例的样本数。

3.4 实验结果与分析

将数据集按 7:3 分割为训练集和测试集,并将测试集随机均分为三组进行测试,同时采用 ResNet101 模型(基础的残差网络模型)进行对比实验。在训练过程中,设置超参 Epoch 为 100,训练数据集 100 次,在每次训练的图像输入阶段,将所选图像进行数据增强处理,提升训练数据量的同时增强模型鲁棒性。初始学习率为 0.001,每 10 次迭代进行一次验证评价。

单一背景下的图像数据在训练过程中损失和准确率变化如图 2 所示。损失值在前 700 次迭代下降较快。在 2 000 次迭代训练后,损失趋近于 0,训练准确率趋于稳定。

复杂背景下的图像数据在训练中的损失和准确率变化如图 3 所示,损失值在前 1 000 次迭代下降较快。在 2 500 次迭代训练后,损失趋近于 0,训练准确率趋于稳定。

单一背景图像数据测试集的实验结果如表 2 所示。

单一背景下,经过实验结果验证,基于 ResNeXt101 模型方法的准确率都稳定在 93% 以上,在纸类回收物模型平均最高准确率达到 98.15%,而 ResNet101 模型的准确率相对较低。这表明 ResNeXt 通过变量基数来控制分组的数量,在不增加参数复杂度的前提下可以有效提高准确率。

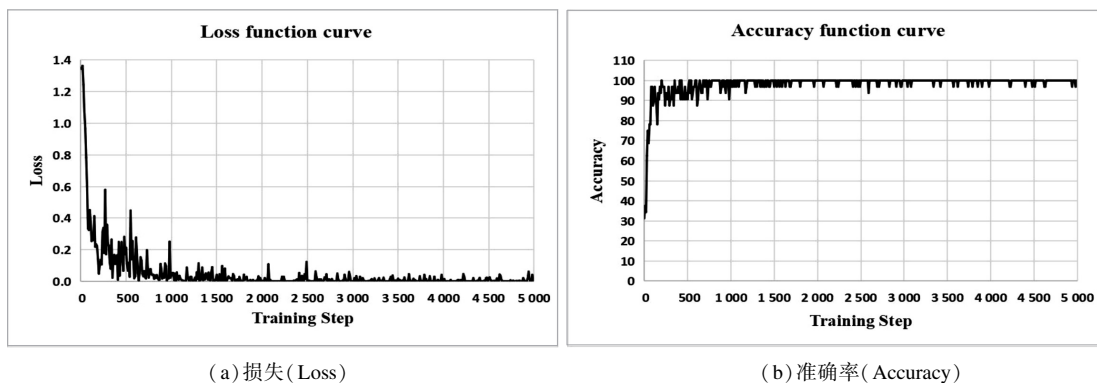
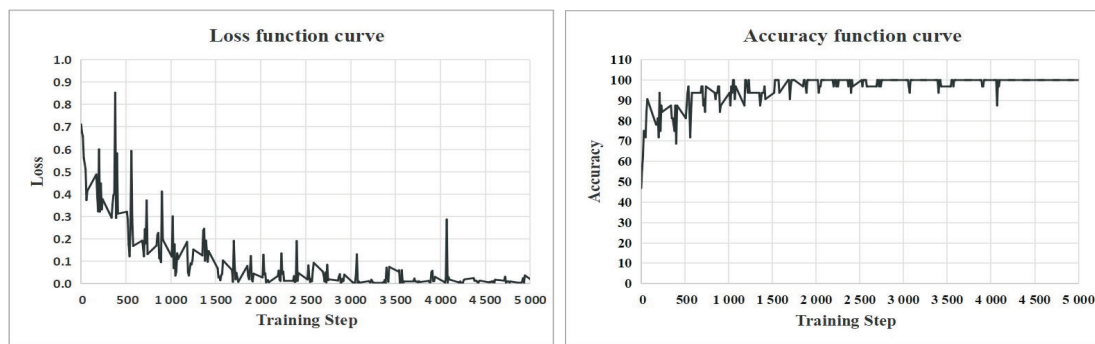


图 2 单一背景图像训练集损失值和准确率



(a) 损失 (Loss)

(b) 准确率 (Accuracy)

图 3 复杂背景图像训练集损失值和准确率

表 2 单一背景图像测试集准确率 (Accuracy)

模型	金属				玻璃			
	Group1	Group2	Group3	Avg	Group1	Group2	Group3	Avg
ResNet101	91.82	93.64	90.95	92.13	92.47	94.86	93.86	93.73
ResNeXt101	97.27	97.73	95.02	96.67	94.86	97.26	96.25	96.12

模型	纸类				塑料			
	Group1	Group2	Group3	Avg	Group1	Group2	Group3	Avg
ResNet101	97.62	97.22	98.02	97.62	90.86	83.25	89.90	88.01
ResNeXt101	98.41	98.02	98.02	98.15	94.42	90.86	94.44	93.24

复杂背景图像测试集测试的结果如表 3 所示。

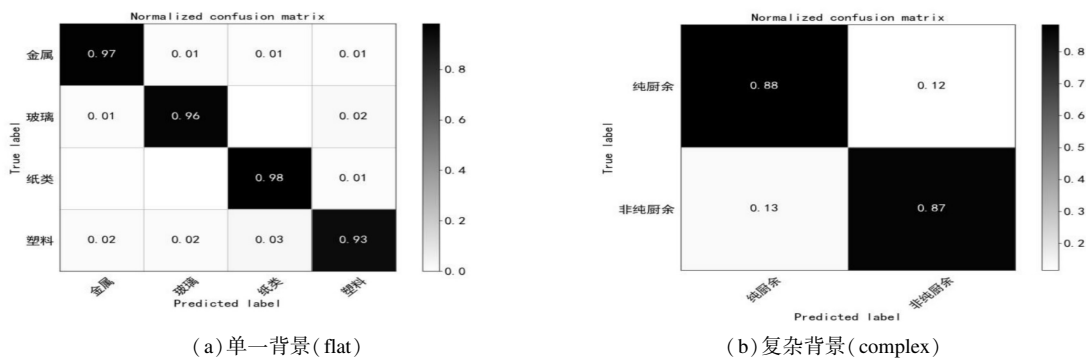
表 3 复杂背景图像测试集准确率 (Accuracy)

模型	纯厨余				非纯厨余			
	Group1	Group2	Group3	Avg	Group1	Group2	Group3	Avg
ResNet101	83.96	79.25	80.56	81.25	83.66	86.27	87.58	85.84
ResNeXt101	94.34	83.96	87.04	88.44	85.62	88.89	85.62	86.71

复杂背景下, 经过实验结果验证, 基于 ResNeXt101 模型方法的准确率稳定在 86% 以上, 总分分类准确率为 87.42%。模型能够在复杂背景下有效地检测厨余垃圾中是否包含杂质。

基于 ResNeXt101 模型方法得出的混淆矩阵如图

4(a)、(b)所示, 分别表示单一背景与复杂背景的不同场景分类结果, 纵坐标表示图像的真实标签, 横坐标表示模型的预测标签, 矩阵对角线上的值代表模型分类正确的结果, 其余数值分别代表每个类别的分类误差。



(a) 单一背景 (flat)

(b) 复杂背景 (complex)

图 4 模型混淆矩阵

图 4(a) 中可以发现纸类图像数据分类准确率最高, 达到了 98%, 而塑料类的图像数据分类误差相对较大, 这是由于塑料类的图像数据种类较多且复杂, 该类数据准确率相对较低, 这也将是下一阶段需要重点

解决的问题。图 4(b) 中模型对纯厨余垃圾分类准确率较高, 在非纯厨余垃圾分类错误样例中发现包含的杂质复杂多样, 且部分杂质体积较小, 颜色与厨余垃圾相似, 例如餐巾纸、塑料袋等, 这也是图像分类技术面

临的重点难点问题。

对比分类结果可以发现,在单一背景图像数据下,模型的分类准确率更高,同时损失值和准确率的变化能更快收敛并且能稳定地维持最优结果。在复杂背景下,图像分类模型对小目标物体的特征提取能力较低,存在干扰特征,影响模型的学习效果,具有更大的挑战性。

4 结束语

面向单一背景和复杂背景两种不同场景下的垃圾图像数据,利用迁移学习的策略进行模型训练与学习,实现了在缺乏大规模标注数据的可回收物图像准确分类和厨余垃圾杂质检测,在单一背景下模型分类准确率高达96.19%,验证了迁移模型的有效性。在复杂背景下,模型的分类准确率可以达到87.42%,这表明模型对于复杂背景中的杂质有较好的特征提取能力,以及对图像的背景与前景有较好的区分能力。其中,模型对体积较小存在干扰特征的杂质检测性能相对较低,对于图像分类模型具有挑战性,也将是下一阶段需要重点解决的问题。同时,该文通过数据增强的方法降低了图像成像要求,进一步提高了模型的鲁棒性和泛化能力。

参考文献:

- [1] 陈非予,杨婷婷,蒋铭阳. 基于深度学习技术的生活垃圾分类模型设计[J]. 电子元器件与信息技术,2020,4(7):94-96.
- [2] 高友文,周本君,胡晓飞. 基于数据增强的卷积神经网络图像识别研究[J]. 计算机技术与发展,2018,28(8):62-65.
- [3] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data,2019,6:60-108.
- [4] ZHANG C, BENGIO S, HARDT M, et al. Understanding deep learning requires rethinking generalization[C]//International conference on learning representations. Toulon, France:ICLR,2017.
- [5] ZHANG C, BENGIO S, HARDT M, et al. Understanding Deep Learning (Still) Requires Rethinking Generalization [C]//Communications of the ACM. New York, United States:COMMUN ACM,2021:107-115.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Neural information processing systems. Nevada, United States:NIPS,2012:1097-1105.
- [7] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//IEEE conference on computer vision and pattern recognition. Honolulu, United States:IEEE,2017:2261-2269.
- [8] BASODI S, JI C, ZHANG H, et al. Gradient amplification: an efficient way to train deep neural networks[J]. Big Data Mining and Analytics,2020,3(3):196-207.
- [9] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research,2014,15(1):1929-1958.
- [10] GARBIN C, ZHU X Q, MARQUES O. Dropout vs. batch normalization: an empirical study of their impact to deep learning[J]. Multimedia Tools and Applications,2020,79:12777-12815.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE conference on computer vision and pattern recognition. Las Vegas, United States:IEEE,2016:770-778.
- [12] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//IEEE conference on computer vision and pattern recognition. Boston, United States:IEEE,2015:1-9.
- [13] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, Inception-Resnet and the impact of residual connections on learning[C]//Association for the advance of artificial intelligence. San Francisco, United States:AAAI,2017:4-12.
- [14] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//IEEE conference on computer vision and pattern recognition. Honolulu, United States:IEEE,2017:5987-5995.
- [15] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861,2017.
- [16] LIN C C, ZHAO G S, YIN A H, et al. A novel chromosome cluster types identification method using ResNeXt WSL model[J]. Medical Image Analysis,2021,69:101943-101952.
- [17] 庄福振,罗平,何清,等. 迁移学习研究进展[J]. 软件学报,2015,26(1):26-39.
- [18] TAN C, SUN F, KONG T, et al. A survey on deep transfer learning[C]//International conference on artificial neural networks. Rhodes, Greece:Springer,2018:270-279.
- [19] 胡满满,陈旭,孙毓忠,等. 基于动态采样和迁移学习的疾病预测模型[J]. 计算机学报,2019,42(10):2339-2354.
- [20] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering,2010,22(10):1345-1359.
- [21] 张雪松,庄严,闫飞,等. 基于迁移学习的类别级物体识别与检测研究与进展[J]. 自动化学报,2019,45(7):1224-1243.
- [22] JASON Y, JEFF C, YOSHUA B, et al. How transferable are features in deep neural networks? [C]//Neural information processing systems. Montreal, Canada: NIPS, 2014: 3320-3328.