

杜俊. 基于 YOLOv8 和迁移学习的垃圾分类方法[J]. 智能计算机与应用, 2024, 14(9): 63-69. DOI: 10. 20169/j. issn. 2095-2163. 240909

基于 YOLOv8 和迁移学习的垃圾分类方法

杜俊

(陇东学院 数学与信息工程学院, 甘肃 庆阳 745000)

摘要: 垃圾分类对于有效的废物管理、促进回收利用和减少环境污染发挥着至关重要的作用。本文提出了一种利用 YOLOv8 图像分类模型进行垃圾分类的方法, 结合迁移学习和数据增强方法, 在华为垃圾分类数据集上训练 YOLOv8 模型。实验结果表明, 采用基于 ImageNet 上的 YOLOv8 图像分类模型在华为垃圾数据集上进行迁移学习, 收敛速度快, 分类准确率达到 89.79%, 在对数据集进行数据增强后, 采用迁移学习方法训练, 分类准确率进一步提高到 91.02%。该模型集成到垃圾分类回收设备中, 可提高垃圾分类回收的效率, 减少回收成本。

关键词: 垃圾分类; YOLOv8; 迁移学习; 数据增强

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2024)09-0063-07

Garbage classification method based on YOLOv8 and transfer learning

DU Jun

(School of Mathematics and Information Engineering, Longdong University, Qingyang 745000, Gansu, China)

Abstract: Garbage classification plays a vital role in effective waste management, promoting recycling, and reducing environmental pollution. This paper proposes a method of using the YOLOv8 image classification model for garbage classification. It combines transfer learning and data enhancement methods to train the YOLOv8 model on the Huawei garbage classification data set. The experimental results show that the YOLOv8 image classification model based on ImageNet is used for transfer learning on the Huawei garbage dataset, the convergence speed is fast, and the classification accuracy reaches 89.79%. After data enhancement on the Huawei garbage dataset, the classification accuracy is further improved to 91.02% using transfer learning method training. The model integrated into garbage sorting and recycling equipment can improve the efficiency of garbage sorting and recycling and reduce the cost of recycling.

Key words: garbage classification; YOLOv8; transfer learning; data augmentation

0 引言

随着城市化、收入和消费的增加, 垃圾的产生量增加, 预计到 2025 年, 全球垃圾量预计将增加到 22 亿吨, 造成持续环境污染加剧、土地占用增多, 因此有效、合理收集和处置废物是市政管理的一项重要任务^[1-3]。废物分类回收是合理处置废物的关键环节, 然而当前废物回收主要依赖人工分拣回收, 不仅工作量大、效率低, 还需要大量人力资源, 且回收成本高。利用现代信息和通信技术实现垃圾分类回收全过程的自动化和无人化, 尽可能地减少人工干预, 提高分类回收的效率和可靠性, 降低分类回收成本, 一直是废物处置领域追求的目标, 应用在多场景下、

满足不同需求、服务不同对象的垃圾回收和分类的自动化设备, 是环保工程工业界的重要任务^[4]。如: 室内、楼宇、道路、水面及恶劣环境中的垃圾清扫机器人, 公共场所或家庭的智能垃圾桶, 以及垃圾自动分拣机器人均具有广泛的应用前景。

尽管这些设备的结构、技术、工作机制不尽相同, 但是其能够工作的基础都是以垃圾图像特征提取, 作为实现垃圾自动分类回收任务的关键环节, 发挥着越来越重要的作用, 具有非常重要的研究意义^[5-6]。卷积神经网络在图像特征提取方面展现出了独特的优势, 各种架构的卷积网络在不同情境下的分类任务上取得了突出成绩, 因其具有自动提取特征能力、更高的准确率以及更好的泛化能力, 已逐

基金项目: 甘肃省教育科技创新项目(2022A-132); 庆阳市科技计划项目(QY2021A-S074); 陇东学院青年科技创新项目(XYZK2107)。

作者简介: 杜俊(1983-), 男, 硕士, 讲师, 主要研究方向: 嵌入式系统, 深度学习。Email: djzqb@foxmail.com

收稿日期: 2024-07-08

渐取代了传统的特征提取方法。然而在垃圾分类识别中,即使同类型垃圾,垃圾之间的特征差异也非常大,为准确检测垃圾类型带来了困难,捕获图像中目标的形态、尺寸、对比度、光照度、背景复杂等问题,增加了物体检测任务的难度。因此,将上述主流检测算法直接用于实际场景的垃圾分类任务难以符合预期。因此,本文在华为垃圾分类挑战赛数据集的基础上,通过数据增强方法扩展训练数据集,改变训练样本数量不足和样本不平衡的问题,利用 YOLOv8 架构^[7]结合迁移学习,训练垃圾分类模型。

1 相关工作

1.1 传统图像特征提取方法

传统的图像特征提取方法主要有 LBP^[8]、Haar^[9]、HOG^[10]、SURF^[11]、PCA^[12]、SIFT^[13]等,这些方法在描述图像的纹理、特征点、边缘等局部特征描述,以及仿射变换的不变性、抗噪等方面具有各自的优势,且计算效率较高。尽管这些传统方法给特征提取带来了显著的效果,但仍有一些需要改进问题。一方面,此类传统算法在复杂场景下效果不佳;另一方面,其只提取图像的低级语义特征,而忽略了高级语义特征,这导致一些特定问题特征的丢失。因此,使用传统的特征提取器很难从对象层提取高级语义特征,从而导致识别性能的瓶颈。

1.2 卷积神经网络

卷积神经网络(Convolutional Neural Network, CNN)通过模拟生物视觉系统,具有强大的特征提取和表示能力,能够自动学习图像中的复杂模式,在图像分类领域取得了显著进展,极大地推动了计算机视觉的发展。自20世纪90年代起,CNN从LeNet-5、AlexNet、GoogleNet、VGG到ResNet逐渐演变和成熟,成为图像识别、目标检测和图像分割等任务的核心技术。

CNN主要由卷积层、激活函数、池化层、全连接层和输出层构成。其中,卷积层是CNN的核心,通过卷积操作提取输入数据的局部特征;卷积核(或滤波器)在输入数据上滑动,计算内积生成特征图。激活函数使网络能够学习复杂的非线性关系,提升梯度传递效率和训练速度,提升网络的表达能力和性能;常用的激活函数包括Sigmoid、ReLU和Tanh。池化层用于下采样特征图,减少数据维度和计算量;常见的池化操作有最大池化和平均池化。全连接层类似于传统神经网络中的隐藏层,将提取的特征映射到输出类别。输出层通常采用Softmax函数,用

于多分类任务的概率输出。

LeNet-5由LECUN Y^[14]等在1998年提出,是最早成功应用于手写数字识别的卷积神经网络之一。LeNet-5由两层卷积层和两层子采样(池化)层交替组成,随后是全连接层和输出层。卷积层通过局部连接和权值共享机制有效提取图像的局部特征,池化层通过下采样减少特征图尺寸和计算复杂度;采用Sigmoid激活函数,引入非线性,提高模型的表达能力。LeNet-5的局部连接与权值共享,减少参数数量,提高计算效率,采用Sigmoid激活函数,引入非线性,使得神经网络能够学习和表示复杂的非线性关系,但其网络深度较浅,特征提取能力有限,Sigmoid函数在深层网络中容易导致梯度消失问题。2012年,AlexNet^[15]在ImageNet大规模视觉识别挑战赛(ILSVRC)中赢得了冠军,其由8层网络组成,通过增加网络深度增强特征提取能力,采用ReLU激活函数解决了梯度消失问题,加快训练速度;使用Dropout技术防止过拟合,提高模型的泛化能力,利用数据增强扩展训练集,提升模型鲁棒性,但较多的模型参数提高了计算和存储需求,对硬件设备要求高。GoogleNet^[16]在ILSVRC 2014中获得冠军,提出的Inception模块具备多尺度特征提取能力,同时通过 1×1 卷积减少计算量;使用全局平均池化替代全连接层,减少参数数量,降低过拟合风险,但Inception模块结构复杂,调参难度大。2014年,牛津大学的Visual Geometry Group提出VGG网络^[17],采用统一的 3×3 卷积核,通过堆叠卷积层增加网络深度。统一卷积核简化网络设计,提高可扩展性,增加网络深度显著提升特征提取能力,提高图像分类准确性,但网络参数量大,导致计算和存储需求高,训练时间长。ResNet由HE K等^[18]在2015年提出,其通过引入残差连接,解决了深层网络中的梯度消失问题,使得训练更深层次的网络成为可能。ResNet的残差连接通过跳跃连接学习残差,缓解深层网络的梯度消失问题,超过100层的超深网络显著提升图像分类准确性,残差块的模块化设计提高了网络结构扩展性。ResNet虽然解决了深度网络的训练问题,但仍需大量计算资源,结构复杂,调参难度大。

2 模型设计

2.1 YOLOv8 分类任务模型

深度学习模型由Backbone、Neck和Head3部分组成,YOLOv8基本遵循这种结构,其模型结构如

图 1 所示。其中, Backbone 网络主要负责从输入图像中提取形状、颜色、位置等语义信息,并将其聚合为高维特征映射; Neck 对骨干网的特征进行融合,优化提取的高维特征映射,有助于 Head 解码更高

质量的特征; Head 是模型的最后一层,其结构会根据不同的任务而有所不同,在目标检测任务中使用边界框回归器和分类器,而在图像分类任务中使用 Softmax 分类器。

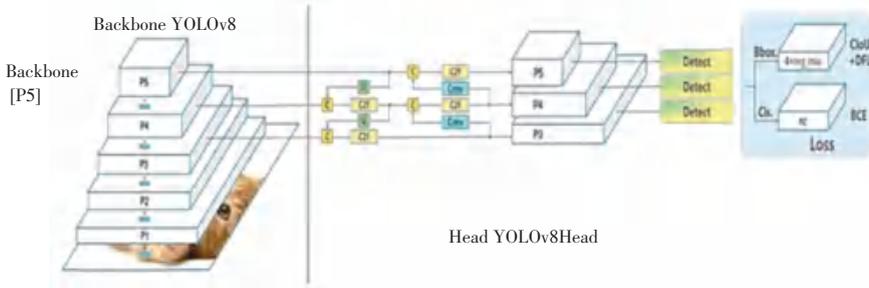


图 1 YOLOv8 检测模型结构

Fig. 1 YOLOv8 detection model structure

YOLOv8 的图像分类模型如图 2 所示,其网络由 Input、Backbone、Softmax 和 Output 构成。Input 负责标准化处理输入图像,将所有输入图像统一调整为固定尺寸。Backbone 由 5 个卷积层(P1—P5)和 4 个 C2F 层构成,卷积层用于提取图像的各种特征,从低级的边缘特征到高级的语义特征,通过池化操作(如最大池化或平均池化)降低特征图的空间尺

寸,从而减少计算量,同时保留重要特征; C2F 内部是残差块(Residual Blocks),引入跳跃连接,缓解深层网络中的梯度消失问题,增强特征的传递和复用。Softmax 层将卷积层和池化层提取到的特征进行整合和映射,最终映射到分类任务的标签空间,输出图像的分类结果。

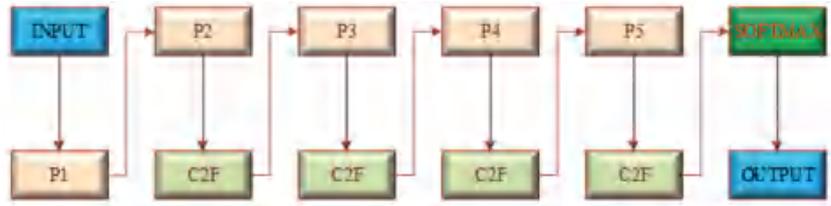


图 2 YOLOv8 分类模型

Fig. 2 YOLOv8 classification model

2.2 数据增强

本文采用华为垃圾分类挑战赛数据集,数据集共分为 40 种垃圾类别,同类型样本数量最多有 441 张图像,最少 51 张图像。在深度学习中,需要更大的数据集才能训练出更准确的模型,训练集中样本数量较少且分布不均衡,可能产生过拟合、预测的偏向、精度下降等问题。数据增强能够对原始数据进行变换或扩充,生成新的训练样本来增加数据集规模和多样性。本文对华为垃圾分类挑战赛数据集进行数据增强,如图 3 所示,首先对原始图像(编号 1)分别进行添加高斯噪声、椒盐噪声、变亮、变暗的变换,输出结果为第 1 列,将原始图像增强 5 倍,然后对这 5 张图像分别进行水平翻转、垂直翻转、水平垂直翻转、转置、转置后的水平翻转、转置后的垂直翻转、转置后的水平垂直翻转,输出结果依次为第 2 列至第 8 列,将原始图像增强 40 倍。由于数据集过

大,对增强后的每类图像随机抽样 2 000 张,建立新的训练数据集。

2.3 迁移学习

随着深度学习在计算机视觉中广泛应用,大数据集上建立的预训练模型,在相似的其它子任务建模时,通过载入预训练模型进行参数训练,可以节省改进模型的训练时间,提高改进模型的性能。迁移学习方法是在 ImageNet 等大数据集上训练网络,然后使用这些权重作为新分类任务中的初始权重。通常,仅复制卷积层中的权重,而不复制包括全连接层在内的整个网络。因为许多图像数据集共享低级空间特征,这些特征可以通过大数据更好地学习。在 CNN 模型训练中,当训练效果不理想,训练样本不够丰富,重新调整参数建立 CNN 模型很麻烦时,可以考虑使用迁移学习方法。在图像分类问题中,迁移学习不仅可以学习边缘、颜色等底层特征,还可以

学习数据的深层特征,可以有效提高网络的分类效果。与传统的 CNN 模型相比,使用转移学习网络拟

合数据的效果更好,网络的识别精度和收敛速度也会提高。

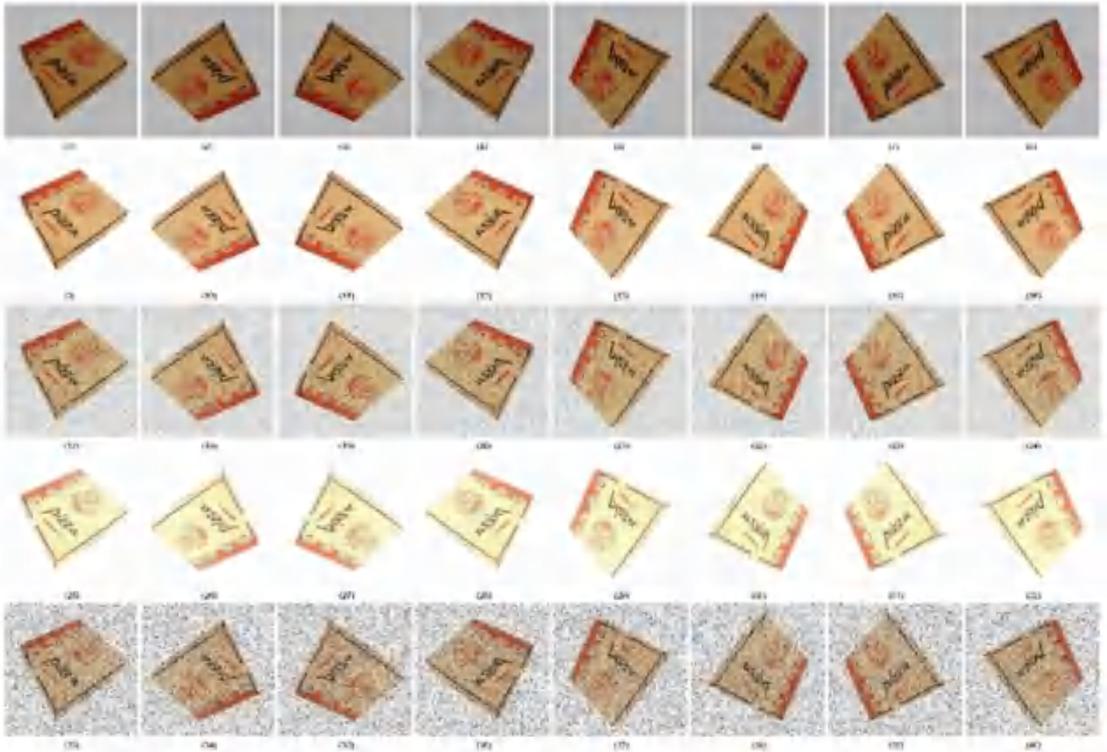


图3 对原始图像的数据增强

Fig. 3 Data enhancement of the original image

3 实验

3.1 数据集

本实验采用“华为云人工智能大赛·垃圾分类挑战赛”的数据集,该数据集共有 14 802 张垃圾图像,将数据集中包含的 40 种垃圾划分为可回收垃圾、厨余垃圾、有害垃圾和其它垃圾等 4 类(见表

1)。其中,可回收物品 23 种,厨余垃圾 8 种,有害垃圾 3 种,其它垃圾 6 种。数据集为每张图像标注了垃圾种类和对应的分类类型。

将整个数据集划分为训练集、验证集和测试集。其中训练集包括 8 869 张图像,验证集有 2 978 张,测试集有 2 956 张。

表1 华为垃圾分类挑战赛数据集垃圾类别图像数量分布统计

Table 1 Number distribution statistics of category images in Huawei Garbage dataset

序号	种类	数量	序号	种类	数量	序号	种类	数量	序号	种类	数量
0	快餐盒	145	10	茶叶渣	232	20	快递纸袋	135	30	纸板箱	192
1	污损塑料	222	11	菜叶菜根	441	21	插头电线	394	31	调料瓶	267
2	烟蒂	167	12	蛋壳	198	22	旧衣服	225	32	酒瓶	168
3	牙签	51	13	鱼骨	245	23	易拉罐	185	33	金属食品罐	193
4	碎花盆及碟碗	232	14	充电宝	214	24	枕头	190	34	锅	237
5	竹筷	173	15	包	251	25	毛绒玩具	330	35	食用油桶	210
6	剩饭剩菜	237	16	化妆品瓶	211	26	洗发水瓶	210	36	饮料瓶	159
7	大骨头	217	17	塑料玩具	185	27	玻璃杯	321	37	干电池	193
8	水果果皮	228	18	塑料碗盆	217	28	皮鞋	229	38	软膏	234
9	水果果肉	233	19	塑料衣架	187	29	砧板	249	39	过期药物	262

3.2 实验环境设置

本文所有实验都是在算力云服务器(7核, Xeon (R) E5-2680 v4, RTX 3080x2 / 20 GB, 60 GB RAM, Ubuntu-64 bit)上进行。深度学习框架为 Torch2.0, CUDA11.8 版本和 CUDN7.6.4 神经网络加速库。训练过程使用学习率预热、多尺度和迭代学习策略。其中,所有实验都设置优化算法为 SGD, 初始学习率设置为 0.01, 学习率衰减系数设置为 0.1, 权重衰减系数为 0.0005, 动量设置为 0.9。

3.3 实验评价指标

模型训练完成后,选择正确的指标来评估模型,可更好评估模型的性能。在图像分类领域,常见的模型评估标准包括精确率(P)、召回率(R)、评价准确率($Accuracy$),计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

其中, TP 是预测正样本的正确数量; FP 是被误认为阳性样本的阴性样本数; TN 是预测的负样本正确数量; FN 是被误认为阴性样本的阳性样本数。

3.4 实验结果分析

YOLOv8 有 5 个尺度分类模型,分别为 YOLOv8n-cls、YOLOv8s-cls、YOLOv8m-cls、YOLOv8l-cls 和 YOLOv8x-cls,从 n、s、m、l 到 x 的尺度分类,模型参数越来越多。使用 YOLOv8 在 ImageNet 上的预训练权重,在华为垃圾数据集上对 YOLOv8 的 5 个尺度模型进行微调,迭代 100 个 epoch,使用 SGD 优化算法,初始学习率为 0.01,采用余弦曲线策略。

实验结果见表 2,其中,YOLOv8 * -cls 是 ImageNet 数据集训练的模型,YOLOv8 * -cls+TF 是 YOLO 各尺度模型加载 ImageNet 预训练权重在华为垃圾数据集上的训练模型,YOLOv8 * -cls+TF+DA 是 YOLOv8 各尺度模型加载 ImageNet 预训练权重在对华为垃圾数据集增强后的训练模型。

表 2 YOLOV8 分类模型准确率

Table 2 Accuracy of YOLOv8 classification model

模型	TOP1/%	TOP5/%	速度/(m·s ⁻¹)	参数/M	FLOPs/B
YOLOv8n-cls	69.00	88.30	0.31	2.7	4.3
YOLOv8n-cls +TF	85.86	97.38	0.31	2.7	4.3
YOLOv8n-cls +TF+DA	86.02	97.50	0.31	2.7	4.3
YOLOv8s-cls	73.80	91.70	0.35	6.4	13.5
YOLOv8s-cls +TF	87.71	97.78	0.35	6.4	13.5
YOLOv8s-cls +TF+DA	88.63	98.06	0.35	6.4	13.5
YOLOv8m-cls	76.80	93.50	0.62	17.0	42.7
YOLOv8m-cls+TF	89.36	98.25	0.62	17.0	42.7
YOLOv8m-cls+TF+DA	89.87	98.85	0.62	17.0	42.7
YOLOv8l-cls	76.80	93.50	0.87	37.5	99.7
YOLOv8l-cls+TF	89.79	98.59	0.87	37.5	99.7
YOLOv8l-cls +TF+DA	91.02	98.99	0.87	37.5	99.7
YOLOv8x-cls	79.00	94.60	1.01	57.4	154.8
YOLOv8x-cls+TF	88.99	98.42	1.01	57.4	154.8
YOLOv8x-cls +TF+DA	89.08	98.51	1.01	57.4	154.8

在 YOLOv8 的 5 个尺度模型中,ImageNet 数据集上,样本充足且丰富,模型越复杂,参数越多,准确率越高。采用 ImageNet 的预训练权重在华为垃圾数据集上微调,模型准确率相比原模型显著提高,n 模型提高了 16.86%,s 模型提高了 13.91%,m 模型提高了 12.56%,l 模型提高了 12.99%,x 模型提高

了 9.99%。实验结果表明,迁移学习加速模型收敛,且显著提高模型准确率,小模型 n 准确率提升比例最高,而较大模型 x 准确率提升比例最低。由此表明,模型越大微调难度越大。采用加载 ImageNet 预训练权重,在对华为垃圾数据集增强后训练的模型,n 模型提高了 0.16%,m 模型提高了 0.92%,s 模

型提高了 0.51%, l 模型提高了 1.23%, x 模型提高了 0.09%。在所有的实验策略中, YOLOv8l 在数据增强后进行迁移学习训练模型 YOLOv8l-cls+TF+DA 的准确率最高, 达到了 91.02%。

YOLOv8l-cls +TF+DA 分类模型最好和最坏的 5 个类别的 *Precision* 和 *Recall* 见表 3。其中, 污损塑料(1)、茶叶渣(10)、鱼骨(13)、菜叶菜根(11)、插头电线(21), 这些类别内目标差异较大, 因而预测

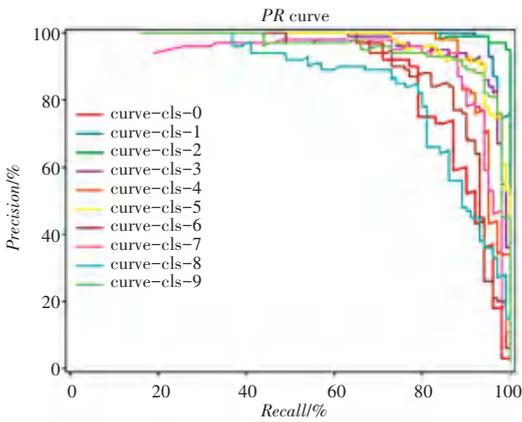
的精准率和召回率较高; 而洗发水瓶(25)上的卡通目标与毛绒玩具(26)较为相似, 软膏(38)与过期药物(39)两个类的目标图像相似性很大, 导致一些相似图像预测的精准率和召回率很低; 纸板箱(30)往往存在于其他类别图像的背景中, 将其他类别图像预测为纸板箱, 导致纸板箱的召回率非常低。其余类别的 *PR* 曲线如图 4 所示。

表 3 YOLOv8l-cls +TF+DA 分类模型的 *Precision* 和 *Recall*

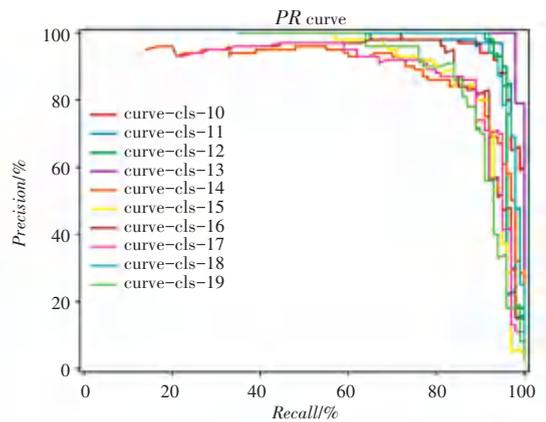
Table 3 *Precision* and *recall* of YOLOv8l-cls +TF+DA

%

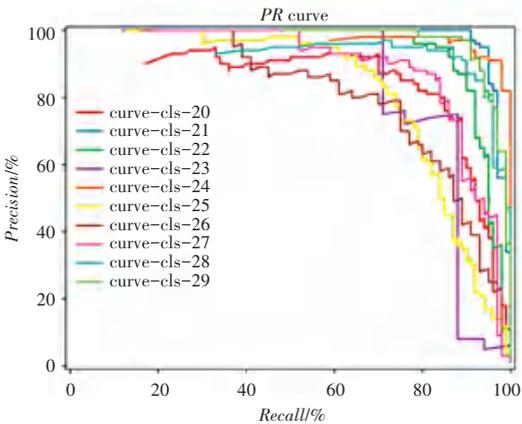
类别标签	PR 和 RE 值最大的 5 个类别					PR 和 RE 值最小的 5 个类别				
	1	10	13	11	21	26	25	38	39	30
<i>Recall</i>	94.60	94.44	97.78	95.16	94.74	73.21	71.91	75.00	78.21	77.36
<i>Precision</i>	94.60	94.44	100.00	92.19	91.14	78.85	83.12	79.17	80.26	82.00



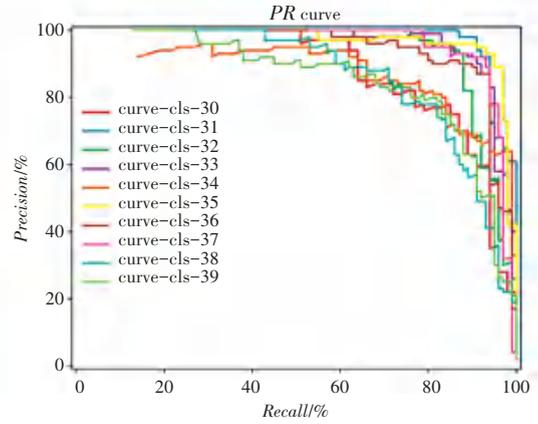
(a) 类别 0~9 的 PR 曲线



(b) 类别 10~19 的 PR 曲线



(c) 类别 20~29 的 PR 曲线



(d) 类别 30~39 的 PR 曲线

图 4 YOLOv8-cls+TF+DA 分类模型的 PR 曲线

Fig. 4 PR curve of YOLOv8-cls+TF+DA

4 结束语

本文对 YOLOv8 分类模型从迁移学习和数据增强两个方面进行了研究, 在华为垃圾分类数据集上进行了评估。研究发现, 基于 ImageNet 的预训练的 YOLOv8 分类模型通过迁移学习方法在华为垃圾数

据集上收敛速度快, 准确率高。在此基础上, 采用数据增强方法后, 进一步提高了分类准确率, 但准确率增幅不太显著, 这是因为华为垃圾数据集一些垃圾类别相似度极高, YOLOv8 分类模型难以提取这些图像的类间细节显著性特征, 未来可融合图像细分方法, 进一步提升垃圾图像分类准确率。

参考文献

- [1] SOSUNOVA I, PORRAS J. IoT-enabled smart waste management systems for smart cities: A systematic review[J]. IEEE Access, 2022, 10: 73326-73363.
- [2] ELIA V, GNONI M G, TORNESE F. Improving logistic efficiency of WEEE collection through dynamic scheduling using simulation modeling[J]. Waste Management, 2018, 72: 78-86.
- [3] AL DUHAYYIM M, MOHAMED H G, ALJEBREEN M, et al. Artificial ecosystem-based optimization with an improved deep learning model for IoT-assisted sustainable waste management[J]. Sustainability, 2022, 14(18): 11704.
- [4] 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[J]. 智能计算机与应用, 2021, 11(10): 5-11, 19.
- [5] ADEDEJI O, WANG Z. Intelligent waste classification system using deep learning convolutional neural network[J]. Procedia Manufacturing, 2019, 35: 607-612.
- [6] FU B, LI S, WEI J, et al. A novel intelligent garbage classification system based on deep learning and an embedded linux system[J]. IEEE Access, 2021, 9: 131134-131146.
- [7] TERVEN J, CORDOVA - ESPARZA D - M, ROMERO - GONZÁLEZ J - A. A comprehensive review of YOLO architectures in computer vision; From YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
- [8] OJALA T, PIETIKAINEN M, HARWOOD D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions[C]//Proceedings of 12th International Conference on Pattern Recognition. IEEE, 1994: 582-585.
- [9] PAPAGEORGIU C P, OREN M, POGGIO T. A general framework for object detection[C]// Proceedings of 6th International Conference on Computer Vision. IEEE, 1998: 555-562.
- [10] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2005: 886-893.
- [11] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [12] KAK A C, MARTÍNEZ A M. PCA versus LDA [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(3-4): 228-233.
- [13] LOWE D. Distinctive image features from scale-invariant key points[J]. International Journal of Computer Vision, 2003, 20: 91-110.
- [14] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [16] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 1-9.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.