

基于ResNet网络的垃圾分类模型研究

王世屹¹, 许静怡²

¹中央民族大学信息工程学院, 北京

²北京林业大学信息学院, 北京

收稿日期: 2024年3月26日; 录用日期: 2024年4月30日; 发布日期: 2024年4月30日

摘要

垃圾分类问题对于环境保护和可持续发展至关重要。传统的垃圾分类模型在处理复杂场景时存在性能瓶颈, 而ResNet网络以其深度残差学习结构在图像识别任务中取得了显著的成功。为进一步提升垃圾分类模型的性能, 本文在ResNet网络基础上进行了改进, 引入了注意力机制和增强的特征提取策略。通过对垃圾分类数据集的训练, 验证了改进ResNet网络在垃圾分类准确性和泛化性能上的优势, 所提出的模型在不同垃圾类别上均取得了较高的分类准确度, 实验结果表明相较于ResNet50, 本文提出的垃圾分类模型提高了1.9%的准确率。本研究为垃圾分类问题的解决提供了一种更有效的解决方案, 也为其他复杂图像分类问题提供了有益的借鉴。

关键词

垃圾分类, 图像识别, 深度学习, ResNet网络

Research on Garbage Classification Model Based on ResNet Network

Shiyi Wang¹, Jingyi Xu²

¹School of Information Engineering, Minzu University of China, Beijing

²School of Information Science and Technology, Beijing Forestry University, Beijing

Received: Mar. 26th, 2024; accepted: Apr. 30th, 2024; published: Apr. 30th, 2024

Abstract

The problem of garbage classification is crucial for environmental protection and sustainable development. Traditional garbage classification models face performance bottlenecks when dealing

文章引用: 王世屹, 许静怡. 基于 ResNet 网络的垃圾分类模型研究[J]. 计算机科学与应用, 2024, 14(4): 368-382.

DOI: 10.12677/csa.2024.144106

with complex scenes, while the ResNet network has achieved significant success in image recognition tasks due to its deep residual learning structure. In order to further enhance the performance of garbage classification models, this paper improved upon the ResNet network by introducing attention mechanisms and enhanced feature extraction strategies. Through training on a garbage classification dataset, the improved ResNet network's advantages in accuracy and generalization performance for garbage classification were verified. The proposed model achieved higher classification accuracy for different garbage categories, demonstrating a 1.9% improvement in accuracy compared to ResNet50. This research provides a more effective solution for the problem of garbage classification and offers beneficial insights for other complex image classification problems.

Keywords

Garbage Classification, Image Recognition, Deep Learning, ResNet Network

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人们对环境保护意识的逐渐增强, 垃圾分类[1]成为了一个备受关注的问题。垃圾分类不仅可以有效减少垃圾的数量, 还可以提高资源的利用率, 降低对环境的影响。近年来, 深度学习技术在图像识别领域取得了显著进展, 被广泛应用于垃圾分类任务中[2] [3] [4]。

ResNet 是一种深度残差网络, 因其在 ImageNet 比赛上的优异表现而备受瞩目。ResNet 通过引入跨层连接和残差块来解决深度神经网络中梯度消失和过拟合等问题。然而, 在实际应用中, 传统的 ResNet 模型仍然存在一些局限性, 如特征表达能力有限、训练效率低下等。因此, 需要采用一些改进措施来提高模型的性能。

本文在 ResNet 的基础上, 针对垃圾识别中背景颜色易混淆的问题, 提出在主网络的 BottleNeck 层加入 CBAM 注意力机制的解决方案; 为克服部分样本单一化挑战, 采用多种数据增强策略; 为改善样本不平衡而导致的模型过拟合, 提出使用标签平滑和 Focal Loss 结合的损失函数; 从而构建一种基于改进 ResNet 网络的垃圾分类模型。最后在一个公开的垃圾分类数据集上进行实验, 比较本文构建的模型和传统的 ResNet 模型在分类准确率方面的表现。

2. 数据集

2.1. 基础数据集

本文采用图片质量较高、图片数据量较大的华为垃圾分类数据集作为基础数据集。

华为垃圾分类数据集共包含 19,735 张图片, 图片数据共分为 4 个大类和 40 个小类。垃圾分类数据小类类别及数量直方图如图 1 所示。垃圾分类数据大类类别及数量表如表 1 所示。

Table 1. Garbage classification data categories and quantity table

表 1. 垃圾分类数据大类类别及数量表

| 类别 | 可回收垃圾 | 其它垃圾 | 厨余垃圾 | 有害垃圾 |
|----|-------|------|------|------|
| 数量 | 8494 | 1684 | 3360 | 1145 |

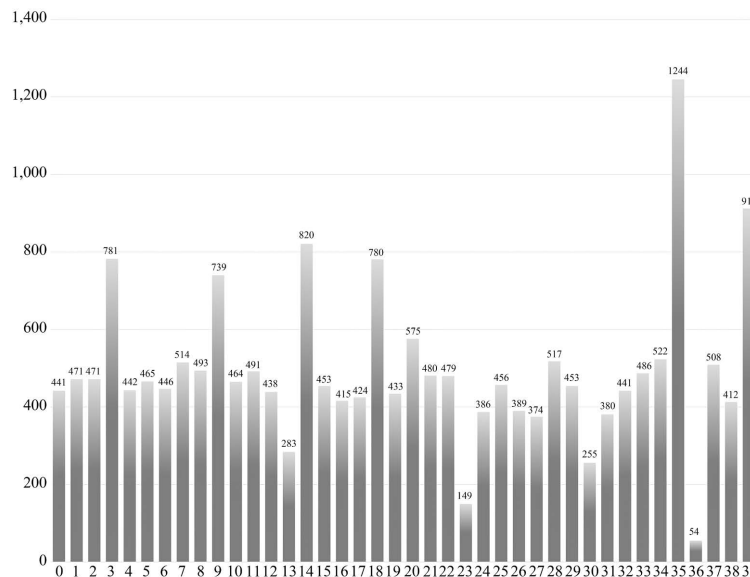


Figure 1. Histogram of subcategories and quantities of garbage classification data
图 1. 垃圾分类数据小类类别及数量直方图

2.2. 修改数据集

华为垃圾数据集如图 2 所示。原始数据集中，所有的图片均放在一个文件下并以 img_ 图片序号命名。每张图片都对应着一个记录了该图片所属的垃圾类别序号的 txt 文件。华为垃圾数据集将 40 个垃圾类别与其序号的 K-V 对应关系存放于 garbage_classify_rule.json 文件中，其中 key 值为垃圾类别标号，value 值的格式为“大类类别/小类类别”。

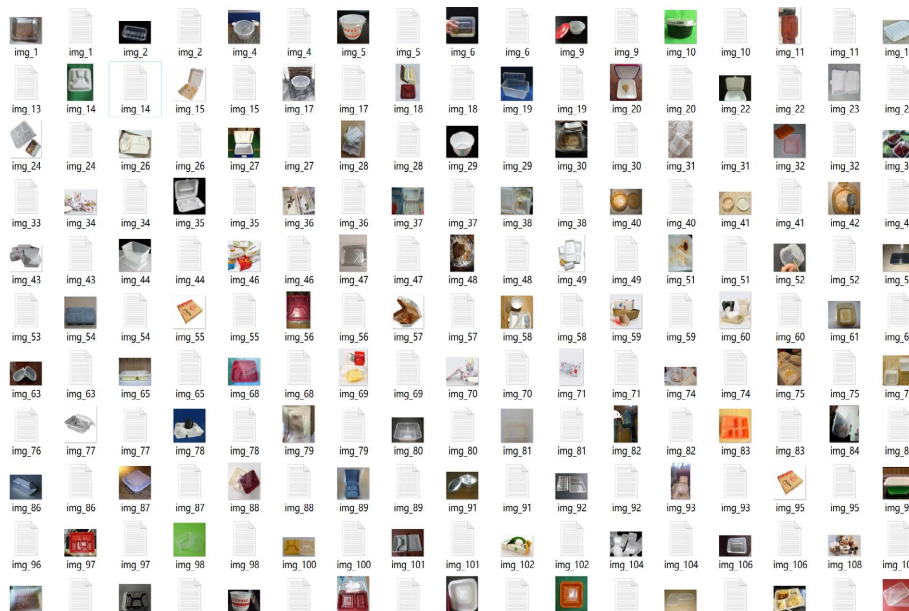


Figure 2. Huawei garbage classification dataset
图 2. 华为垃圾分类数据集

为了优化华为垃圾数据集增删图片时的操作流程，本文对该数据集的存储结构进行了改进。首先为

每一个小类创建一个文件夹并命名为该小类对应的序号, 然后将该类别下的所有图片转移到文件夹中。修改后的华为垃圾数据集如图 3 所示。

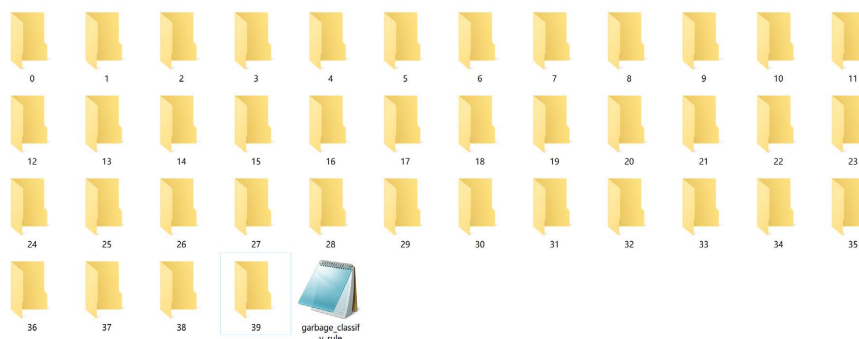


Figure 3. Modified Huawei garbage classification dataset

图 3. 修改后的华为垃圾分类数据集

由于华为垃圾分类数据集的种类和数量均有不足。因此, 本文收集标注了 4265 张百度图片并将其加入到华为垃圾分类数据集中, 最终得到的垃圾分类数据集新增了一次性棉签、纸巾、毛巾、笔、护肤品、肉类、袋子、坚果壳、粉笔和书本 10 类垃圾, 共包含 24,000 张图片。修改后的垃圾分类数据小类类别及数量直方图如图 4 所示。

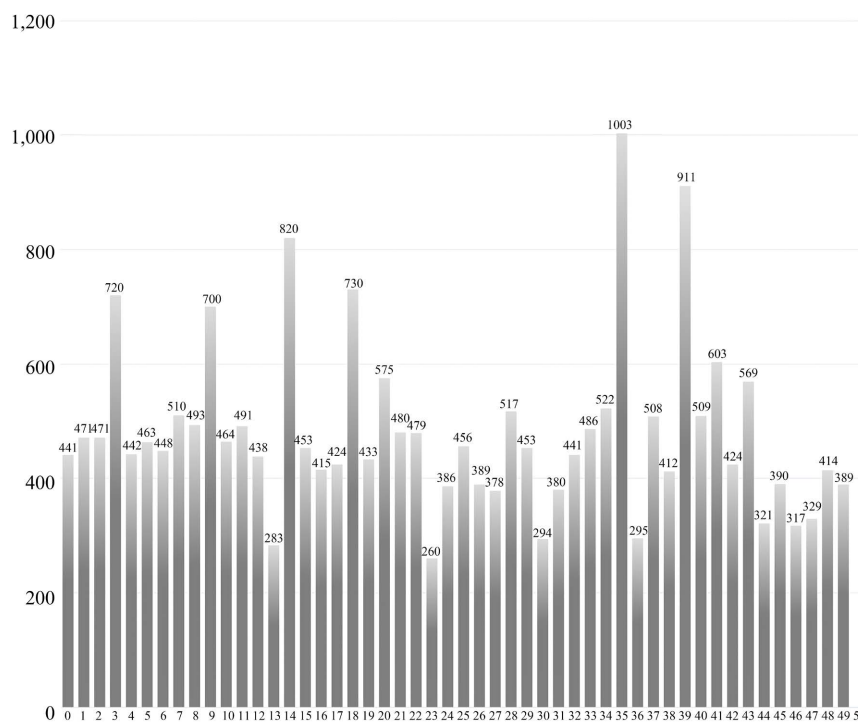


Figure 4. Histogram of subcategories and quantities of modified garbage classification data

图 4. 修改后的垃圾分类数据小类类别及数量直方图

优化后的垃圾分类数据集有以下特点:

- 1) 受拍照设备的影响, 同类图片的明暗度、对比度和颜色饱和度皆存在差异。
- 2) 部分图片存在多种垃圾互相遮挡、重叠的情况。

- 3) 部分图片的背景颜色和垃圾本体互相混淆。
- 4) 不同类别的垃圾的部分样本图像在形状、大小和颜色上较为相似。
- 5) 图片类别分布不均匀。

2.3. 数据增强

如前文所述, 本文采用的垃圾分类数据集中存在部分同类别下的图片具有明暗度、对比度和颜色饱和度上的差异。为了降低该差异对模型准确率的影响, 训练的数据必须多元化。增加数据多元化的常见解决方案是数据增强。数据增强一方面可以增加图片的数量, 另一方面可以增加图片的质量。

数据增强可以根据增强后的图片是否存储在本地分为在线增强和离线增强。前者在训练模型时随机选取图片进行增强, 无需存放在本地。在线增强可以增加数据的多元化, 但无法单独对某一类别的图片进行扩增。后者需要将增强后的图片存放在本地, 若单独增加数量较少的类别下的图片, 则可以降低模型的过拟合。本文采用标签平滑正则化和 Focal Loss 的方式降低模型的过拟合, 因此最终选择在线增强。

1) 常规数据增强

常规的数据增强方法有: 翻转、旋转、平移和调整明暗度等。常规数据增强效果示意图如图 5 所示。第一层从左至右分别为经过水平翻转、垂直翻转和同时经过水平、垂直翻转后的图片。第二层从左至右分别为经过水平平移、垂直平移和同时经过水平、垂直平移后的图片。第三层从左至右分别为亮度增加 1.5 倍、颜色增加 1.5 倍和对比度增加 1.5 倍后的图片。



Figure 5. Schematic of general data enhancement effects

图 5. 常规数据增强效果示意图

2) 图像插值方法

除了常规的数据增强方法外, 图像插值方法同样被广泛用于深度学习数据集的数据增强中。垃圾识别模型的训练图像由规范的华为垃圾分类数据集和从百度上收集到的图片共同组成, 后者收集到的图片尺寸不唯一, 而垃圾识别模型要求所有的图片尺寸均为 $380 \times 380 \times 3$, 因此需要对部分图片进行缩放。由于垃圾分类图片并不是矢量图, 因此在图片缩放的过程中会伴随一定的失真。常用的图像插值方法均

可以减少信息的丢失。本文对比了经过最近临边插值和双立方插值方法的图像的清晰度之后, 最终选定了双立方插值方法。

双立方插值涉及到中心点及其领域, 该插值方法将中心点附近的 15 个点作为领域。双立方插值涉及的像素点及邻域如图 6 所示, 其具体步骤如下。

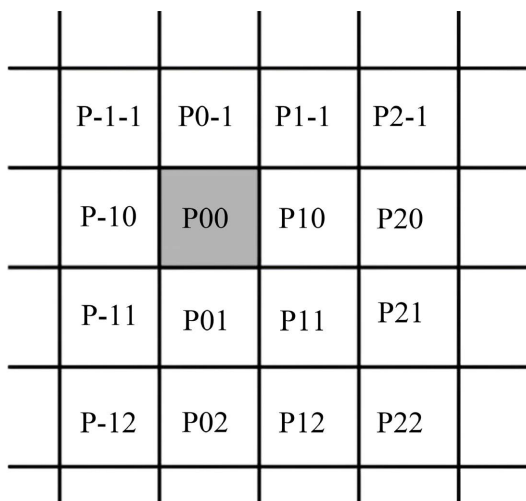


Figure 6. Pixels and neighborhoods involved in bicubic interpolation
图 6. 双立方插值涉及到的像素点及邻域

- ① 找到离目标差值点映射在原图的坐标点的最近点作为原点, 然后依次找到其周围的领域并记录。
- ② 将上述像素点带入到公式 1 中, 转换从原图到目标图的像素点。

$$F(i+v, j+u) = \sum_{row=-1}^2 \sum_{col=-1}^2 f(i+row, j+col) S(row-v) S(col-u) \quad (1)$$

其中, $(i+v, j+u)$ 为转换后的像素点, $f(x)$ 为中心点和领域的值, $S(x)$ 为卷积公式, 其具体公式如公式 2 所示:

$$S(x) = \begin{cases} (a+2)|x|^3 - (a+3)|x|^2 + 1|x| \leq 1, \\ (a)|x|^3 - 5(a)|x|^2 + 8a|x| - 4a1 < |x| < 2, \\ \text{Ootherwise} \end{cases} \quad (2)$$

其中, a 可以取任意值, 常用值为 -0.5 和 -0.75 , 本文取 $a = -0.5$ 进行插值, 则上述公式转换为:

$$S(x) = \begin{cases} 1 - 2.5|x|^2 + 1.5|x|^3 |x| \leq 1, \\ 2 - 4|x| + 2.5|x|^2 - 0.5|x|^3 1 < |x| < 2 \end{cases} \quad (3)$$

3. 模型构建

3.1. 主干网络选取

综合考虑分类模型准确率和推理时间, 本文选取了 EfficientNet-B6 [5], Inception-V3 [6] 和 ResNet50 三种主流深度学习网络进行对比, 并分别在垃圾数据集上进行了训练, 不同模型的准确率曲线图如图 7 所示。对比三个网络在数据集上的各方面指标可知, Inception-V3 网络收敛最快, 但收敛后的模型准确率最低, 仅为 90% 左右; EfficientNet-B6 收敛速度最慢, 收敛后的模型准确率为 92% 左右; 而 ResNet50 网

络收敛速度稍慢, 但模型准确率达到到了 93%, 排名第一。综合考虑网络的收敛速度和准确率, 最终选择 ResNet50 作为主干网络。

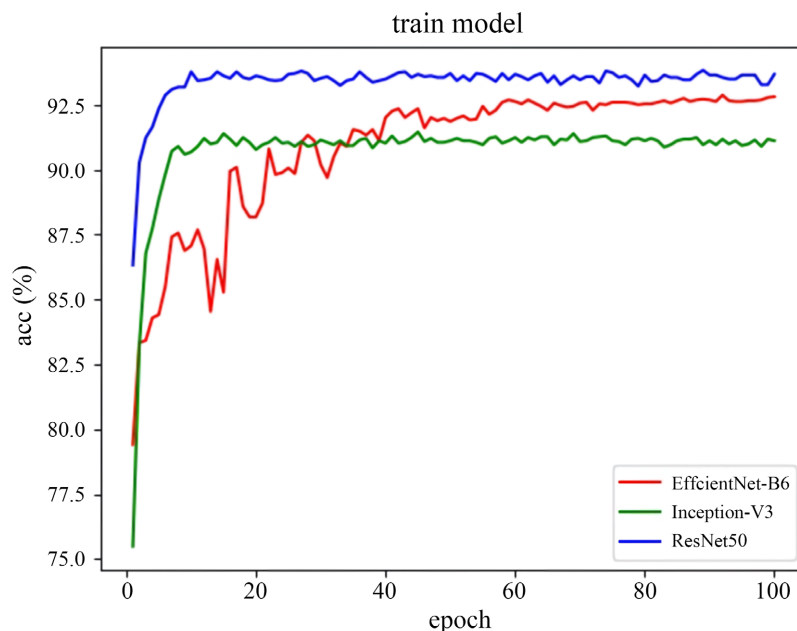


Figure 7. Accuracy curve graph of different model
图 7. 不同模型的准确率曲线图

3.2. 模型设计

本文以 ResNet50 作为主干网络并对该网络进行了改进: 首先将 CBAM 模块插入到原网络的 BottleNeck 层中, 结合上下文对特征进行聚焦并抑制背景干扰对垃圾识别模型准确率的影响; 然后使用标签平滑正则化[6]和 Focal Loss [7]相结合而产生的新损失函数替换交叉熵损失函数, 有效得缓解了垃圾分类数据集图像不均衡所造成的模型过拟合; 最后使用 softmax 作为分类器, 其输出的结果是维数为 50 的一维向量, 向量的索引对应着 0~49 类垃圾, 向量索引对应的值为该样本在该类别下的分类分数。垃圾识别模型架构如图 8 所示。

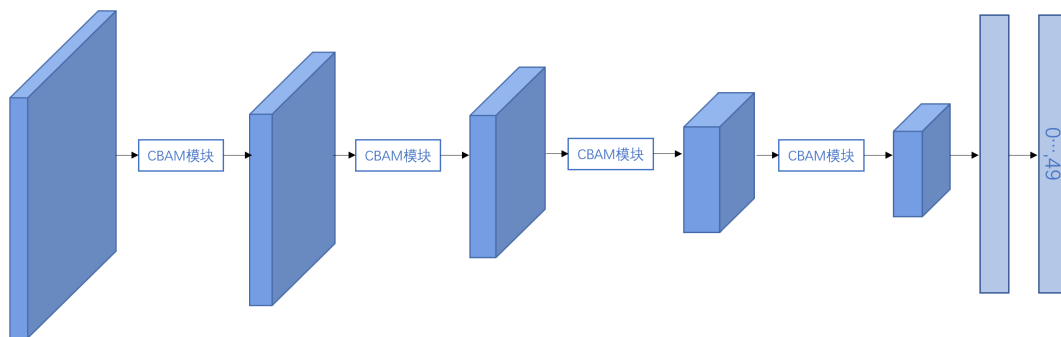


Figure 8. Garbage recognition model architecture diagram
图 8. 垃圾识别模型架构图

3.3. CBAM 注意力机制

本文采用的垃圾分类数据集中的部分图片存在背景干扰, 为了使垃圾识别模型能够结合上下文将注意力集中在垃圾主体上, 从而关注更为重要的特征并抑制与背景颜色相关的特征, 本文采用了 Woo S 提出的 CBAM 融合注意力机制模块[8]辅助特征提取, 该模块同时具有空间域和通道域, 其中前者可以消除图片的旋转, 噪声等干扰因素带来的误差, 后者可以帮助模型提升对当前任务有用的特征图的通道, 并抑制对当前任务用处不大的特征通道。当一张特征图输入之后, 首先经过通道注意力机制模块, 该模块提取出一组权重并将其与输入特征图相乘, 此时得到的特征图更为关注重要的通道; 接着将改变后的特征图输入到空间注意力模块, 同样得到一组权重并与输入特征图相乘, 经过两个模块提取后的特征称为 Refined 特征。

1) 通道注意力模块(Channel Attention Module)

通道注意力模块架构图如图 9 所示, 其提取权重的步骤如下:

① 假设输入的特征图 F 的维度为 $H \times W \times C$, 为了聚合同一通道的特征信息, 将这组特征分别进行平均池化和最大值池化。最终得到两个维度为 $1 \times 1 \times C$ 的池化特征并分别命名为 F_{avg} 和 F_{max} 。

② 将 F_{avg} 和 F_{max} 输入到包含一层隐层的多层感知器 MLP 里, 得到两个 $1 \times 1 \times C$ 的通道注意力图。为了减少参数量, 隐层神经元的个数为 C/r , 其中 r 也被称作压缩比。

③ 将经过 MLP 得到的两个通道注意力图相加, 并输入到 sigmoid 激活方法进行激活从而得到最终的权重。

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^c)) + \sigma(W_1(W_0(F_{max}^c)))) \end{aligned} \quad (4)$$

其中, F_{avg}^c 指的是特征图进行平均池化的值, F_{max}^c 指的是特征图进行最大池化的值, W_0 指的是多层感知机第一层的权值矩阵, W_1 指的是多层感知机第二层的权值矩阵, $\sigma(x)$ 为激活方法, 通常选用 sigmoid 方法。

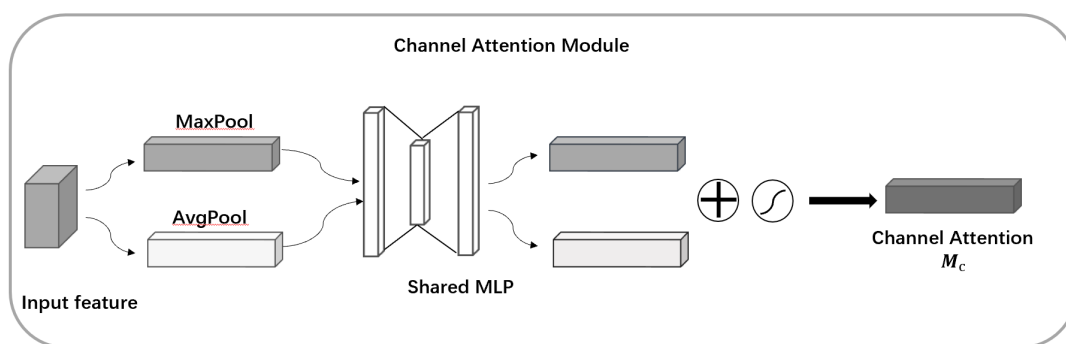


Figure 9. Channel attention module architecture diagram

图 9. 通道注意力模块架构图

2) 空间注意力模块(Channel Attention Module)

空间注意力模块架构图如图 10 所示, 其提取权重的步骤如下:

假设输入的经过通道注意力机制提取后的特征图 F' 的维度为 $H \times W \times C$, 将其沿着通道方向进行最大池化和平均池化。最终得到两个维度为 $H \times W \times 1$ 的池化特征, 分别命名为 F'_{avg} 和 F'_{max} 。

将 F'_{avg} 和 F'_{max} 在通道维度上进行级联, 然后对其求卷积, 并将求卷积后的结果输入到 sigmoid

激活方法进行激活从而得到最终的权重。

具体的计算公式如下：

$$M_s(F) = \sigma\left(f^{7 \times 7}\left(\left[\text{AvgPool}(F'); \text{MaxPool}(F')\right]\right)\right) = \sigma\left(f^{7 \times 7}\left(\left[F_{avg}^{rs}; F_{max}^{rs}\right]\right)\right) \quad (5)$$

其中, F_{avg}^{rs} 指的是特征图进行平均池化的值, F_{max}^{rs} 指的是特征图进行最大池化的值, $f^{7 \times 7}(x)$ 指的是卷积, 卷积的维度与级联后的池化层的值相同, $\sigma(x)$ 为激活方法, 通常选用 sigmoid 方法。

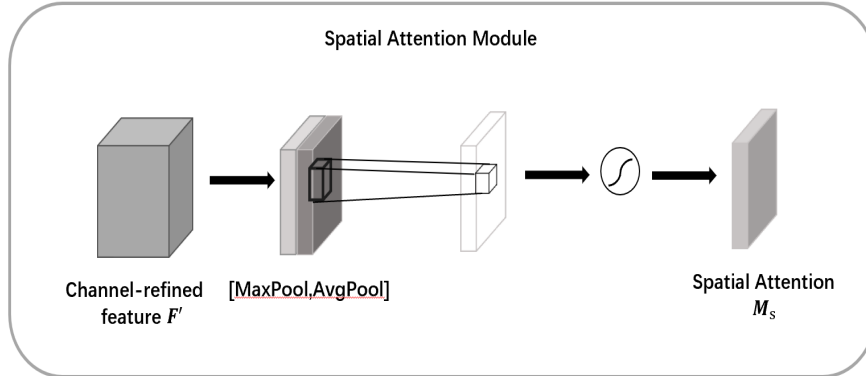


Figure 10. Spatial attention module architecture diagram
图 10. 空间注意力模块架构图

3.4. 损失方法

本文采用的垃圾分类数据集中的样本数量不均衡, 如果采用普通的交叉熵损失方法则容易出现过拟合的情况。常用的抑制过拟合方法有 DropOut 方法、标签平滑正则化等。由于 ResNet50 网络使用了大量的残差层, 全连接层较少, 导致 DropOut 方法效果较差。因此本文选用标签平滑正则化方法结合 Focal Loss 方法对损失函数进行改进。

标签平滑的核心思想是为每一个标签分配一个小概率, 从而抑制部分类别标签对模型产生过大影响。由于原始类别数组为一个 one-hot 向量, 即除了正确标签的位置的值为 1 外, 其余标签的值均为 0。从而导致原始的交叉熵损失函数只计算正确标签位置的损失, 而不计算其他标签位置的损失, 这就会使得模型朝着增大正确位置的预测结果的方向训练, 最后导致过拟合。平滑过后的类别数组的各个标签均为小于 1 的小数, 则损失函数的结果与所有标签均相关, 迫使模型往增大正确分类概率并且同时减小错误分类概率的方向前进。

Focal Loss 损失函数的思想是将样本划分为简单样本和困难样本并提升困难样本的权值, 从而缓解简单样本多、困难样本少造成的过拟合问题。最终得到的损失函数如下：

$$\begin{cases} y_i = (1-e)y_i^0 + e \times \frac{ones_{y_i^0}}{num_{classes}} \\ loss = -y_i \times (1-y_p)^\gamma \times \log(y_p) \end{cases} \quad (6)$$

其中, y_i^0 表示图片的原始类别数组, y_i 表示经过标签平滑后的类别数组, e 表示平滑因子, $ones_{y_i^0}$ 表示一个全 1 的类别数组, $num_{classes}$ 表示数据集内小类的个数, y_p 为所有类别对应的经过激活函数后输出的概率数组, γ 为调整简单样本和困难样本的系数。本文经实验最终选定 $e = 0.1$, $\gamma = 2$ 。

4. 模型训练与评估

4.1. 模型训练

本文的实验训练环境配置如表 2 所示。

Table 2. Experimental training environment configuration

表 2. 实验训练环境配置

| 硬件环境 | | 软件环境 | |
|------|-----------------|------------|-------------|
| CPU | Intel I7-8700 | 操作系统 | Ubuntu16.04 |
| 内存 | 16 G | Python 版本 | 3.8 |
| GPU | NVIDIA RTX 2080 | Pytorch 版本 | 1.1 |

本文将数据集按照 7:2:1 的比例划分为训练集、验证集和测试集。在模型训练刚开始时, 由于模型所学习到的垃圾分类数据集的知识较少, 可能会造成准确率曲线的振荡。为使模型训练更为稳定, 本文先设置学习率为 0.001, 目标准确率为 80% 进行 warm up。预训练结束后, 将最终输出的权重输入到主干网络中作为初始权重。

本文在训练时, 采用 ADAM 优化算法。本文共设置 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 0.9, 二阶矩估计的指数衰减率为 0.999。

4.2. 模型评估指标

模型评估的常见指标为混淆矩阵(Confusion Matrix)、准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数(F1-Score)等。本文采用混淆矩阵、准确率和精确率作为评估指标。

1) 混淆矩阵

混淆矩阵将实验结果进行汇总, 矩阵的行表示真实值, 矩阵的列表示预测值。以三分类为例, 混淆矩阵示例如表 3 所示。

Table 3. Example of a confusion matrix

表 3. 混淆矩阵示例

| 真实值 \ 预测值 | 卫生纸 | 塑料袋 | 湿纸巾 |
|-----------|-----|-----|-----|
| | 卫生纸 | 90 | 10 |
| 塑料袋 | 2 | 98 | 0 |
| 湿纸巾 | 0 | 0 | 100 |

2) 准确率

准确率和精确率的计算过程涉及到 TP、TN、FP、FN 四个概念, 在进行模型评估的时候, 如果选定一类垃圾为正样本, 则其余垃圾为负样本。TP、TN、FP、FN 的具体解释如下:

TP: 被判定为正样本, 事实上也是正样本。

TN: 被判定为负样本, 事实上也是负样本。

FP: 被判定为正样本, 但事实上是负样本。

FN: 被判定为负样本, 但事实上是正样本。

准确率用来展示在目前的所有分类中, 模型预测对了的个数在总体预测个数中的占比。其公式如公式 7 所示。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

对于上述的情况:

$$Accuracy = \frac{90 + 98 + 100}{90 + 10 + 2 + 98 + 100} = 0.96$$

在实际使用模型时, 往往会将排名前几的预测结果全部返回给用户, 只要其中包含正确答案, 则判断此次预测为正确。因此, 衍生出了两个常用的准确率指标 TOP1 和 TOP5, TOP1 指的是只返回预测结果中最大的值给用户, 如果这个值不是正确分类, 则预测错误。而 TOP5 指的是将分类结果中的排名前五的结果返回给用户, 一旦中间包含正确分类, 则预测正确。

3) 精确率

精确率用来展示正确结果为某中类别的样本中, 预测结果也为该类别的样本的数量, 反映出模型对某个具体类别的精确程度。精确率的公式如公式 8 所示。

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

对于上述的情况, 分类别计算精确率:

$$Precision_{\text{卫生纸}} = \frac{90}{10 + 90} = 0.9$$

$$Precision_{\text{塑料袋}} = \frac{98}{2 + 98} = 0.98$$

$$Precision_{\text{湿纸巾}} = \frac{100}{100} = 1$$

根据上述计算可以看出, 模型对湿纸巾的精确率最高, 即模型对湿纸巾的识别效果最好

4.3. 模型评估

1) CBAM 模块插入位置的选取

本文依次在 4 个 BottleNeck 层中间插入 CBAM 模块进行测试。CBAM 消融实验结果对照表如表 4 所示。由表中数据可以看出, 当分别在 BottleNeck 的 1、2、3、4 的位置插入空间注意力模块的时候, 分类准确率相差并不大, 最高为 93.42%, 对模型的提升较小。而当在各个位置同时插入 CBAM 模块的时候, 随着插入个数的增多, 模型的最终准确率出现了明显变化, 当插入的 CBAM 模块的数量为 4 时, 其准确率相较原始网络的 93% 提高了 1.19%。因此, 最终选择插入四层 CBAM 模块。

Table 4. Comparison table of CBAM ablation experiment results

表 4. CBAM 消融实验结果对照表

| 实验代号 | 模块插入位置 | 准确率 |
|------|---------|--------|
| 1 | 1 | 93.12% |
| 2 | 2 | 93.17% |
| 3 | 3 | 93.42% |
| 4 | 4 | 93.37% |
| 5 | 1、2 | 93.63% |
| 6 | 1、2、3 | 93.91% |
| 7 | 1、2、3、4 | 94.19% |

2) 损失方法的选取

本文选取了交叉熵损失方法, 标签平滑正则化方法和 Focal Loss 方法进行消融实验, 损失方法消融实验结果对照表如表 5 所示。当选取标签平滑损失方法之后, 相较于交叉熵损失方法, 模型在测试集上的 TOP1 准确率提升了 0.2%, TOP5 准确率提升了 0.71%。当选取 Focal Loss 损失方法之后, 相较于交叉熵损失方法, 模型在测试集上的 TOP1 准确率提升了 0.57%, TOP5 准确率提升了 0.89%。而将两者结合之后, 相较于交叉熵损失方法, 模型在测试集上的 TOP1 准确率提升了 0.91%, TOP5 准确率提升了 1.87%。皆有明显提升。

Table 5. Comparison table of loss method ablation experiment results

表 5. 损失方法消融实验结果对照表

| 实验标号 | 损失方法 | TOP1 准确率 | TOP5 准确率 | 耗时 |
|------|------------------------------|----------|----------|--------|
| 1 | Cross Entropy Loss | 94.19% | 97.23% | 112 ms |
| 2 | Label Smoothing | 94.39% | 97.94% | 106 ms |
| 3 | Focal Loss | 94.76% | 98.12% | 109 ms |
| 4 | Focal Loss + Label Smoothing | 94.9% | 99.1% | 110 ms |

3) 模型对比

本文在准确率、占用内存和模型预测时间的维度上分别对 ResNet50 和垃圾分类模型进行了测试, 模型对比结果如表 6 所示。从表中数据可以看出, ResNet50 的内存比垃圾分类模型低 1.3 M, 且平均耗时比垃圾分类模型少 13 ms。在这两个维度上二者的差异较小。但垃圾分类模型的准确率相较于 ResNet50 提高了 1.9%, 提升较大。

Table 6. Model comparison results

表 6. 模型对比结果

| 模型 | 模型内存 | 准确率 | 耗时 |
|----------|---------|-------|--------|
| ResNet50 | 107.6 M | 93% | 97 ms |
| 垃圾分类模型 | 108.9 M | 94.9% | 110 ms |

4) 在测试集上的结果

使用最终训练得到的垃圾识别模型在 2400 张测试集进行测验, 最终得到的结果如下:

① 混淆矩阵

根据最终的实验结果, 绘制出的混淆矩阵如图 11 所示。图中的数字代表对应行列代表的结果类型的数量, 数字的颜色越深则代表数字越大。图中的混淆矩阵的对角线上的颜色均较深, 证明了模型在大部分测试样本下的实验结果为正确。

② 准确率

垃圾分类模型准确率如表 7 所示, 在 2400 张测试集上, TOP1 的准确率为 93%。TOP5 的准确率为 99.1%。由于 TOP5 准确率为垃圾分类系统的图像识别准确率, 且该值较高, 因此本节设计的垃圾识别模型能够应用于实践中。

③ 精确率

垃圾分类模型精确率如表 8 所示, 在测试的 50 类垃圾中, 有 19 个类别的精确率为 100%。有 31 类垃圾的精确率未达到 100%, 其中大多数类别的精确率都大于等于 90%, 仅有 6 类样本的精确率在 90%

以下, 分别为水果果皮、肉类、旧衣服、调料瓶、酒瓶和金属食品罐。其中有 3 张真实值为肉类的图片被识别为剩饭剩菜, 同时有 2 张分类结果为水果果皮的图片被识别为水果果肉, 这可能是因为这两组垃圾图片互相遮盖, 产生了混淆。有 2 张真实值为旧衣服的图片被识别为包, 这可能是因为折叠后的旧衣服形态、材质和颜色均与包较为相似。有 3 张金属食品罐真实值被识别为易拉罐, 这可能是因为两者的颜色、形态、大小、材质均较为相似。总体来看, 垃圾识别模型对大多数类别的识别效果较好。

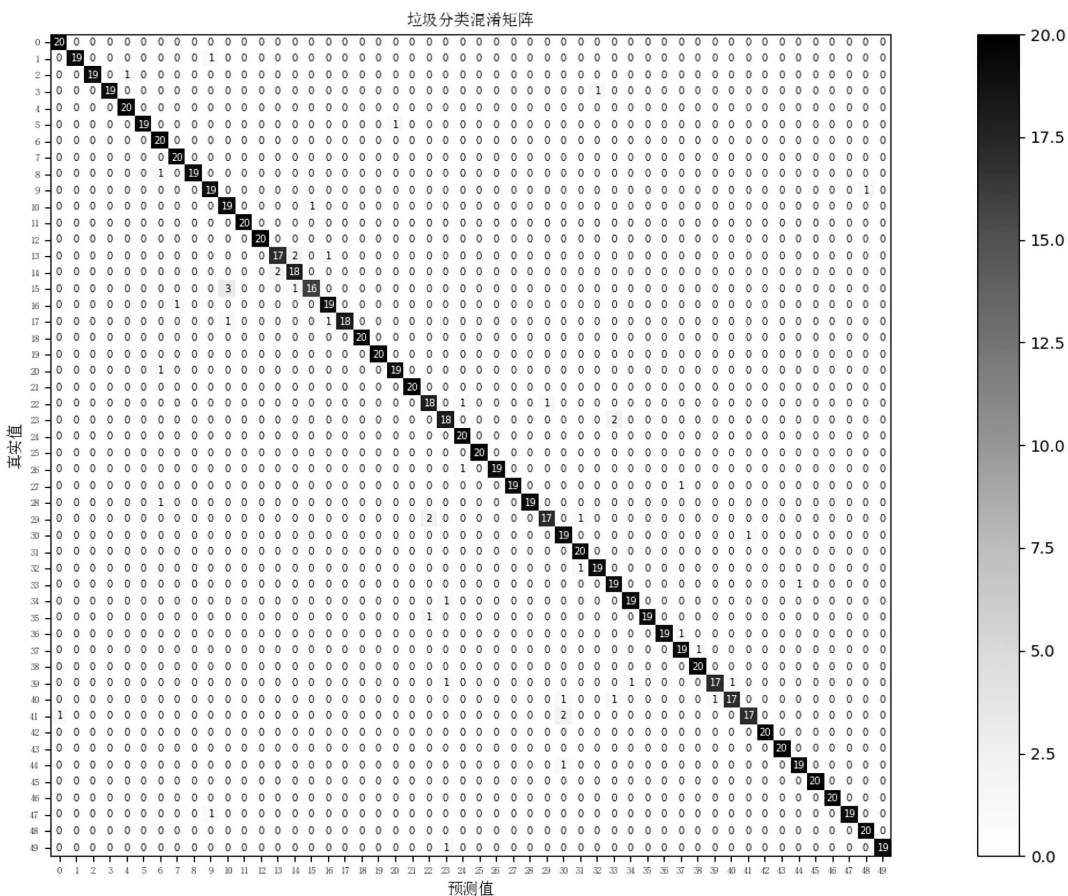


Figure 11. Confusion matrix
图 11. 混淆矩阵

Table 7. Classification model accuracy
表 7. 垃圾分类模型准确率

| 数据集 | TOP1 准确率 | TOP5 准确率 |
|-------------|----------|----------|
| 2400 张测试数据集 | 93% | 99.1% |

Table 8. Classification model precision
表 8. 垃圾分类模型精确率

| 类别 | 精确率 |
|--------|------|
| 一次性快餐盒 | 100% |
| 一次性棉签 | 95% |
| 卫生纸 | 95% |

续表

| | |
|---------|------|
| 毛巾 | 95% |
| 污损塑料 | 100% |
| 烟蒂 | 95% |
| 牙签 | 100% |
| 破碎花盆及碟碗 | 100% |
| 竹筷 | 95% |
| 笔 | 95% |
| 剩饭剩菜 | 95% |
| 坚果壳 | 100% |
| 大骨头 | 100% |
| 水果果皮 | 85% |
| 水果果肉 | 90% |
| 肉类 | 80% |
| 茶叶渣 | 95% |
| 菜叶菜根 | 90% |
| 蛋壳 | 100% |
| 面包 | 100% |
| 鱼骨 | 95% |
| 充电宝 | 100% |
| 包 | 90% |
| 化妆品瓶 | 90% |
| 塑料玩具 | 100% |
| 塑料碗盆 | 100% |
| 塑料衣架 | 95% |
| 快递纸袋 | 95% |
| 插头电线 | 95% |
| 旧衣服 | 85% |
| 易拉罐 | 95% |
| 枕头 | 100% |
| 毛绒玩具 | 95% |
| 洗发水瓶 | 95% |
| 玻璃杯 | 95% |
| 皮鞋 | 95% |
| 砧板 | 95% |
| 纸制品 | 95% |
| 纸板箱 | 100% |

续表

| | |
|-------|------|
| 调料瓶 | 85% |
| 酒瓶 | 85% |
| 金属食品罐 | 85% |
| 锅 | 100% |
| 食用油桶 | 100% |
| 饮料瓶 | 95% |
| 干电池 | 100% |
| 杀虫剂 | 100% |
| 胶水 | 95% |
| 软膏 | 100% |
| 过期药物 | 95% |

5. 结论

本文首先针对华为垃圾数据集的各个维度展开了介绍, 并详细阐述了对应的修改流程。其次在比较了三种网络模型在垃圾分类数据集上的训练速度和准确度的基础上, 最终选定了 ResNet50 作为垃圾识别模型的主网络。针对垃圾分类数据集中同一类图片的明暗度、对比度和颜色饱和度差异较大的问题, 本文采取了多种在线增强方法; 针对垃圾分类数据集中部分图片的垃圾主体和背景颜色较为相似的问题, 本文采用了 CBAM 注意力机制模块, 该模块能够结合上下文将注意力聚焦在目标物体上; 针对垃圾分类数据集图像分布不均匀的问题, 本文将标签平滑正则化方法与 Focal Loss 方法相结合, 有效得降低了模型的过拟合。最后本文针对垃圾识别模型进行了多个消融实验, 并在测试集上对模型进行了测试, 实验结果证明模型兼具高准确率、低耗时和高精确率, 能够满足垃圾分类的实用需求。

参考文献

- [1] 刁艳杰. 什么是垃圾分类? [J]. 走向世界, 2019(32): 41.
- [2] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, Lake Tahoe, 1106-1114.
- [3] He, K.M., Zhang, X.Y., Ren, S.Q., et al. (2016) Deep Residual Learning for Image Recognition. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [4] Aral, R.A., Keskin, S.R., Kaya, M., et al. (2018) Classification of TrashNet Dataset Based on Deep Learning Models. *IEEE International Conference on Big Data*, Seattle, 10-13 December 2018, 2058-2062. <https://doi.org/10.1109/BigData.2018.8622212>
- [5] Tan, M. and Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*, Long Beach, 9-15 June 2019, 6105-6114.
- [6] Szegedy, C., Liu, W., Jia, Y.Q., et al. (2015) Going deeper with Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [7] Lin, T.Y., Goyal, P., Girshick, R., et al. (2017) Focal Loss for Dense Object Detection. *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2999-3007. <https://doi.org/10.1109/ICCV.2017.324>
- [8] Woo, S., Park, J., Lee, J.Y., et al. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1