

基于 ResNet18 的垃圾分类实现

杨 东 周 丽

四川大学锦城学院 计算机与软件学院 四川 成都 611731

【摘 要】上海市垃圾分类要求提出之后,随着时间的不断推移,越来越具体的垃圾分类标准逐渐走入大众的视野,进一步要求民众不断提高垃圾分类知识。垃圾分类普遍推行,与之不匹配的是广大人民群众不熟练的垃圾分类技能,因此,研究垃圾分类模型具有社会意义和广泛应用价值。为了帮助人们快速识别手中的垃圾,有效地提高资源的回收利用率,减轻垃圾污染所带来的危害。本文将研究基于 ResNet18 模型实现垃圾分类,通过识别生活中的垃圾图片,将图片中的垃圾进行分类。帮助人们在生活中进行分类垃圾,实现资源的回收与利用。

【关键词】深度学习; 网络结构; ResNet18; 垃圾分类

1 当代社会下垃圾分类研究背景

当代社会不断发展的外卖行业以及各大线下实体商铺给予了人们更加舒适的购物体验 and 饮食服务,随之而来的是不断产生的大量生活垃圾。垃圾分类是一项大众都可参与以达到资源回收的有效途径,此举措不仅符合我国可持续发展战略,还可以很好地避免“垃圾围城”。垃圾分类正深度走入各大城市,与之不匹配的是人们的分类能力的不足和分类知识的匮乏,故本篇文章研究基于 ResNet18 实现垃圾分类,帮助大家更快速、准确实现生活中垃圾的分类。在当今社会人们普遍不熟悉垃圾分类知识的情况下,垃圾分类模型的研究具有广泛的社会意义和使用价值。

深度神经网络 (deep neural network, DNN) 目前已被成功地应用于图像分类^[1]、语音识别^[2-3]、自然语言处理^[4-5]等领域。但是,随着卷积神经网络 CNN 的发展和越来越复杂的图片数据集,传统神经网络在层数增加时出现的误差逐渐趋于稳定甚至出现显著增长的网络退化问题逐渐被大众发现。为了有效提高模型的精度,传统的卷积神经网络可以简单地增加一些网络的层级以达到目的,但是,在网络的层级达到一定的数量之后,模型的精度开始降低。同时,卷积神经网络在反向传播的过程中需要不断地传播梯度来调整权重,但是随着网络层级的不断增加,梯度会逐渐衰减直至消失,导致参数矩阵无法进行有效的调整^[6]。ResNet 网络模型就是为了有效解决这种退化问题而诞生。

2 基于 ResNet18 的深度网络训练

神经网络的分类和识别效果与网络深度有关,通常通过增加网络深度来提高模型分类和识别效果,而当传统的卷积神经网络网络深度加深到一定程度时,分类和识别效果会变差。其主要原因是,在学习过程中,网络越深其权值梯度弥散越慢,造成网络的训练效果越差,因此需要解决因网络的深度增加而产生的梯度弥散较慢的矛盾^[7]。所以,在使用普通的卷积神经网络来进行垃圾分类识别准确度不高的情况下,我们采用深度残差网络来解决这个问题。

深度残差网络 (deep residual networks, ResNet) 是由 Kaiming He 提出的一种深度学习网络,于 2015 年获得 ImageNet 目标检测、目标定位以及 COCO 目标检测和 COCO 图像分割的第一名^[8]。把网络简单重复地堆叠会导致严重的梯度消失问题,这样不仅不利于模型的训练,反而导致我们训练时间过长。在更深层次的网络上,出现了过拟合一类的问题:精确度几近饱和,模型的训练效果反而变得更加糟糕。既然梯度消失使得网络难以训练,增加网络的深度无法达到有效训练的目的,导致不仅无法提高性能,反而降低性能。所以不如直接在简单得模型上进行训练。我们在进行图像处理的时候使用残差向量编码来达到目的:通过信息重组,将一个问题分解成为多个相互关联的残差问题,这样,不仅优化了训练,还便于计算机计算。残差结构如图 1 所示。

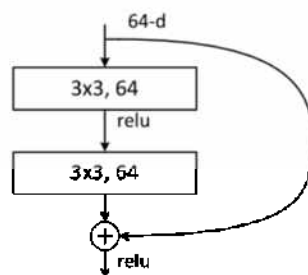


图 1: 残差结构图

目前 ResNet18 就是采用了跳跃连接让网络学习恒等映射,拥有带权重的 18 层网络,网络结构如图 2 所示,不同于其他的 ResNet 网络结构,ResNet18 中包含了 17 个卷积层和 1 个全连接层(其中不包括 BN 层和池化层)来进行学习。ResNet 受到 VGG 的启发,卷积层几乎都是 3*3 卷积。不难发现,所有的 ResNet 网络输入部分都是由一个大小为 7x7,步长为 2 的大卷积核,以及一个大小为 3x3,步长为 2 的最大池化组成。经过这样的预处理,大小为 224x224 的输入图像就变成了大小为 56x56 的特征图,这样的处理,可以减少对存储大小和计算的要求,方便了部分计算能力差的个人 PC 端

可以更好地使用这种网络结构。

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

图 2: ResNet 结构图

而相比于 VGG-16 网络,使用深度 ResNet 网络代替 VGG-16 网络,可以在一定程度上提取到更多的特征图,提高了这个检测模型的识别率准确度.该算法在 F 值和检测精度上都比其他广泛应用的算法有明显的优势^[9].提取更多特征图从而提高识别率和准确度是可行的手段,算法在可行性和准确度上也有相应优势.

3 实验结果及分析

3.1 训练和测试数据

本文使用的样本是来自 garbage_classify 数据包,

其中图片按照数字顺序命名,并采用一一对应的 txt 文件标记对应图片的分类.其中包括了 14802 个垃圾图像,其中共分为 40 类,最多的种类有 736 张,最少的种类也有 85 张.随机打乱之后按照 8: 2 的比列分成了训练集和测试集.训练样本包含 11824 张图片,测试样本中包含 2978 张图片.整个训练集和测试集具有样本数量多和种类复杂的特点.部分数据如图 3 所示,其中不少都是贴近生活中拍摄的图片,有利于训练结果更加贴近与实际使用时对照片中的垃圾分类的实际情况.

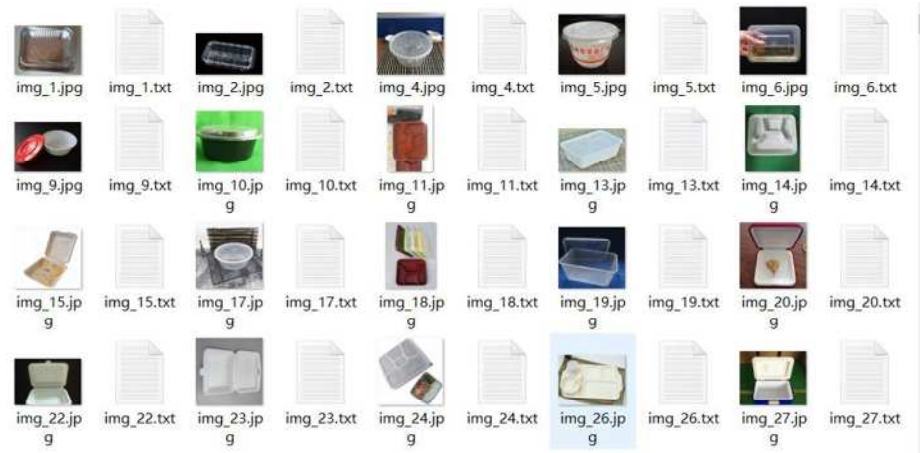


图 3: 部分数据截图

3.2 实验细节

遍历数据文件夹中所有标记的文件,取出其中对应的图片名和分类组成我们的数据集.数据集按照 8: 2 的比例分为训练集和测试集.

为了更好地检测模型的准确率,实验在模型训练前对训练数据集进行更多的预处理,以便于模型学习到更多的特征点,而在测试集上进行更少的预处理,以便于检测模型的泛化能力.故实验对于训练数据集和测试数据集进行了以下预处理:对所有的训练数据进行伸缩变化、中心裁剪、对比度调整、随机旋转、旋转、归一化.对测试集进行伸缩变换、中心裁剪和归一化.

用在 ImageNet 上预训练好的 ResNet18 进行迁移

学习,训练过程中,冻结了全连接层之前的所有卷积层,不参与训练和更新.模型之所以使用交叉熵损失函数来判断实际输出与期望输出的相近程度,是因为分类问题用交叉损失衡量预测值和真实值之间的距离更准确.采用随机梯度下降作为模型的优化器,并且为了防止无法收敛,学习率不宜过大,故将学习率设置为 0.03,动量因子设置为 0.9.将 epoch 设为 50, batch_size 设为 64.

3.4 测试结果及分析

在实验过程中,每轮 epoch 结束时会输出训练集上的平均损失函数值和准确度,还会在测试集上计算和打印平均损失函数值和准确度.为了便于观察到模型的适应程度,模型在训练集正确率和测试集正确率都趋于

稳定时停止训练。多次进行 50 轮的训练之后,我们发现每次结果都极为相似。其中一次模型在训练数据集(蓝色线)和测试数据集(黄色线)上的表现如图 4 所绘制的 loss 曲线所示:

```
plt.plot(np.arange(len(train_loss_arr)), train_loss_arr)
plt.plot(np.arange(len(test_loss_arr)), test_loss_arr)
```

[<matplotlib.lines.Line2D at 0x22eec962898>]

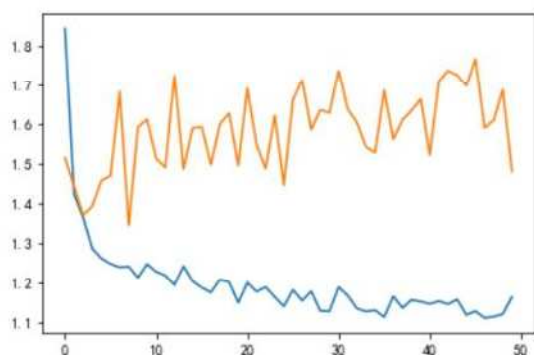


图 4: loss 曲线图



图 5: 检测结果

训练的结果不尽人意,最终训练数据集在第 30 轮左右时,准确率便稳定在 69% 左右,不再出现明显增长,而测试数据集的准确率在 20 轮左右时就稳定在 65% 左右。如图 5 所示,训练结束之后对随机拍摄的四张图片进行检测,其中两张被正确分类,两张被错误分类。通过实验得到,ResNet18 作为卷积神经网络的核心应用到垃圾分类问题上的效果还有待优化和改进。

结论

本文针对基于 ResNet18 进行垃圾分类实现进行了实验。在这种方法中,我们进行了数据预处理、冻结网络和随机梯度下降。但是实验结果表明 ResNet18 在垃圾分类的表现上有待改进。垃圾分类的模型实在复杂,即便数据集的垃圾在获取时尽量避免了噪音,但颜色相近的情况下难以分辨材质、容易混在一起,且垃圾照片中容易含有食物颜色特质,特征复杂,难以通过简单的网络结构分析出垃圾材质从而进行分类。ResNet18 虽然可以直接调用,也避免了梯度消失的问题。却也忽略了垃圾分类在事实层面难以通过简单的训练模型得到结果,仍然需要更加复杂的网络结构和训练模型才能

有效的对垃圾进行分类。实验结果表明本文所提出的在并不进行额外的网络结构优化的情况下,简单引用 ResNet18 难以实现垃圾分类。

【参考文献】

- [1]Krizhevsky A,Sutskever I,Hinton G E.Imagenet classification with deep convolutional neural networks[J].Communications of the ACM,2017,60(6):84- 90
- [2]Zhang Ying,Pezeshki M,Brakel P,et al.Towards end-to-end speech recognition with deep convolutional neural networks[J].arXiv preprint,arXiv:1701.02720,2017
- Graves A,Mohamed A,Hinton G.Speech recognition with deep recurrent neural networks[C] //Pro of the 2013 IEEE Int Conf on Acoustics,Speech and Signal Processing. Piscataway,NJ:IEEE,2013:6645- 6649
- [3]Kim Y.Convolutional neural networks for sentence classification[J].arXiv preprint,arXiv:1408.5882,2014
- Chen Danqi,Manning C D.A fast and accurate dependency parser using neural networks[C] //Pro of the 2014 Conf on Empirical Methods in Natural Language Processing. Stroudsburg,PA:ACL,2014:740- 750
- [4]梁栋,何佳,石陆魁,王松,刘佳.结合运动目标检测和 resnet 的车速车型智能识别 [J].北京交通大学学报,2019,43(05):10-19.
- [5]姜声.基于 resnet 的眼底图像分类及描述 [D].山东工商学院,2019:
- [6]朱超平,杨艺.基于 yolo2 和 resnet 算法的监控视频中的人脸检测与识别 [J].重庆理工大学学报 (自然科学),2018,32(08):170-175.
- [7]qubo xie,ke zhou,xiaohu fan.a scene text detection algorithm based on resnet and faster r-cnn. [C].2019.840-843. DOI:10.26914/c.cnkihy.2019.048646.