基于 GoogLeNet 的垃圾分类应用

赖 炫 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】近年来全国各地城市正在逐步实行垃圾分类，但是由于垃圾的种类过多，对于广大市民的科普程度很低，在对垃圾进行分类时不能够正确地判断垃圾类别，所以要把垃圾分类落实到位却是一个大的难题。针对这一难题，本文使用一种基于 GoogLeNet 的垃圾图像分类方法进行实验，经实验得到，该法对于测试数据准确率大约为 70%，在场景单一、物体单一的图片上识别准确率高达 99%，证明了该方法在未来城市垃圾分类中的可行性。

【关键词】垃圾分类；GoogLeNet；卷积神经网络；迁移学习

1 绪论

在城市的发展进程中，环保是不可忽略的一个部分，近几年来，各地的环保政策纷纷开始实施，国家在此情况下提出了实施垃圾分类的建议[1]，但是要合理地进行垃圾分类，必须对全体民众普及垃圾分类知识，显然，这不是一朝一夕就能做到的，因此，随着人工智能领域的日新月异，使用计算机来处理对垃圾地分类无疑是一个很好的解决办法，在进行垃圾分类处理时，可以使用摄像头捕捉垃圾图像并进行实时识别垃圾种类反馈给用户，让用户能够快速准确的得知垃圾地种类并进行分类，采用这种方法能对垃圾分类的普及速度有显著的提升，在对图像数据进行处理的时候我们通常会使用到卷积神经网络，其中经典卷积神经网络之一的 GoogLeNet 能够有效地分类垃圾并识别垃圾的种类，本文实验将使用 GoogLeNet 预训练模型在垃圾分类数据集上进行迁移学习并测试垃圾分类的准确率。

2 迁移学习与网络模型结构

2.1GoogLeNet 结构模型概述

本文实验采用 GoogLeNet 模型，GoogLeNet 的深度为二十二层[2]，在 2014 年 ImageNet 比赛中脱颖而出的是由知名的互联网巨头公司 Google 设计的划时代意义 GoogLeNet 模型，其拥有卓越的性能和丰富的结构参考设计意义[3]。GoogLeNet 基于卷积神经网络（Convolutional Neural Networks,CNN）[4]，在其基础上而设计和开发的，卷积神经网络是一种前馈神经网络[5]，与其他神经网络相比，图像和语音的分类识别是它擅长的领域，比如 VGGNet[6]，ResNet[7] 这两种经典网络都源于卷积神经网络，GoogLeNet 开创性地使用了不同于以往线性模型的框架，并行的分支结构标新立异，并解决传统网络特征提取不足的情况。简单地说，VGG 模型是增加网络的深度，但网络深度达到一定程度时，网络能力增长会到达瓶颈，GoogLeNet 从另一个维度来强化网络，它拓展了网络的宽度，每个单元有多个层并行运算。GoogLeNet 中包含有 Inception 模型，可以有效提升卷积核的感受野[8]。

2.2 GoogLeNet 结构模型原理

2.2.1Inception 模块示意图

如图 1，Inception 模型包含有四个互不相同的组合，四个分支里又含有 6 个不同的卷积层和 1 个池化层，然后将四个分支所提取到的特征合并到一起就构成了下一个模块的输入。如图 2，GoogLeNet 的整个层次比较深，其中就包含有多个像图 1 一样的 Inception 模块，多个 Inception 模块与网络开头和结尾的单个层连接在一起即为 GoogLeNet 模型。

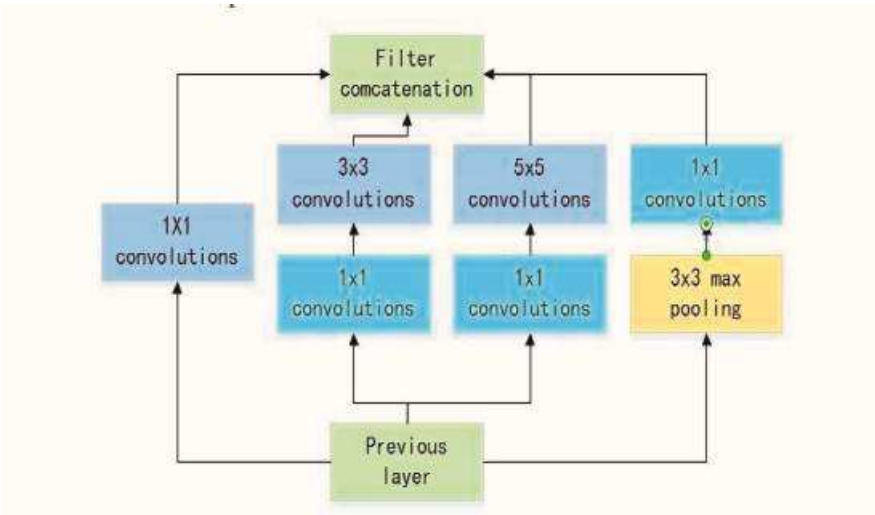

图 1 Inception 模型

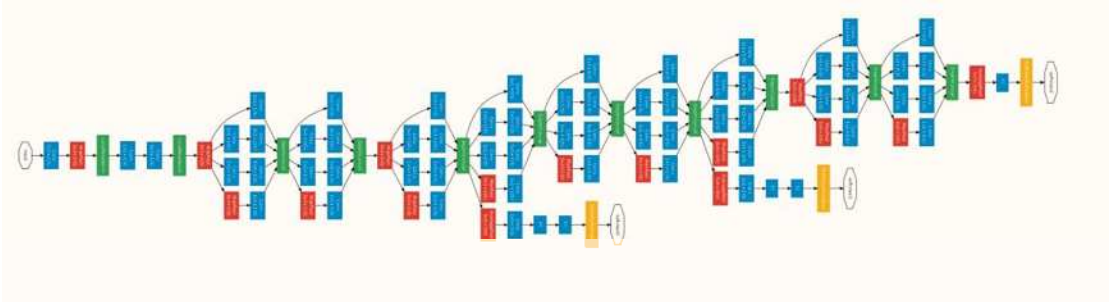

图 2GoogLeNet 模型结构

2.2.2 卷积层

显然，卷积层是整个卷积神经网络的核心，而卷积（Concolution）又是卷积层的核心。在卷积计算过程中，卷积核不断地在进行移动，而移动的步长可以自行设定，这个自行设定的步长称为步幅（strides），步幅在卷积计算中起到控制移动的作用，由一个元组构成，分别控制卷积核自左而右、自上而下移动，并且能够缩小卷积层的输出高、宽。卷积核在移动的时候如果输入的数据为 6×6，卷积核为 3×3，strides=2，第三次移动卷积核会超出输入数据矩阵的边界，针对此种情况，就会使用到填充矩阵边缘（padding）的方法，即扩展输入矩阵尺寸，扩展边界补零，卷积核移动时会刚好到达矩阵边界，便可将输入数据与卷积核进行完美的匹配运算。

2.2.3 池化层

图像在通过卷积层后如果直接进行特征训练分类，特征图像的大小会很大，导致计算的过程会耗费更多的时间和资源，并且如果对全部特征都进行训练，过拟合出现的概率也会较大，池化层能够较好地解决这一问题，在卷积层输出特征图之后，可以对它进行池化，即在卷积输出数据的基础上再次运算取值，常用的池化方式分为三类，最大池化、均值池化和全局均值池化，每一类对于数据的处理方式都有一定的不同，池化层可以减小由卷积层传入的特征大小，相应的可以提高整个运算过程的速度，并且降低过拟合发生的概率。因此，下采样可以总结为两个步骤：保留显著特征、降低特征维度。

2.3 迁移学习理论

迁移学习是深度学习中常用的方法，迁移学习的要点在于迁移，GooGleNet 是已经经过训练的模型，拥有良好的效果，使用它来对垃圾数据集进行新的训练就可以得到一个新的模型，简单来说就是将一个问题上训练好的模型通过简单的调整使其适用于一个新的问题[9]，将预训练模型作为我们新模型的起点，这些预训练模型不管是在网络层次设计还是模型通用性上都很好，如果我们实际需要解决的问题与预训练模型关联性较强，那么使用这些预训练模型能够极大地节约时间，并且训练出来的新模型的性能、泛化性和鲁棒性都能得到提高。

3 实验及结果分析

3.1 数据集

本文实验所采用数据集来自于华为垃圾分类挑战杯数据集[10]。如图 3，该数据集里每张图片都对应有一个标签，所以本次实验属于有监督学习，这些标签包含了四个大类，每个大类下面又细分为了总计 40 个小类，整个数据集一共包含 14802 张垃圾图像。实验第一阶段对数据集进行划分，将整个数据集的 80％图片作为实验训练集，20％图片作为实验测试集，此外，要增强测试的准确率，可以预先处理图像，例如进行对亮度、对比度进行调整，翻转，随机剪裁，长宽比重设，使模型的泛化性得到提升。

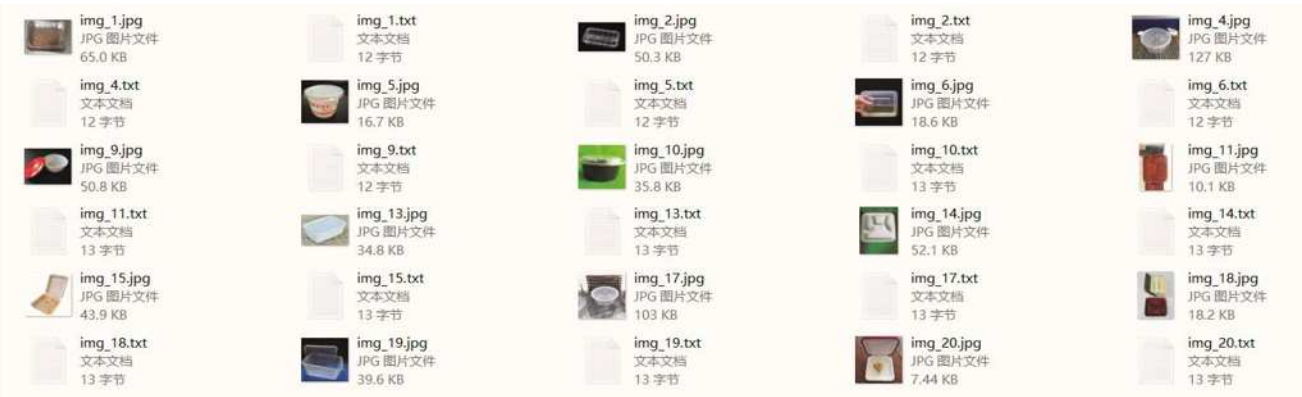

图 3 部分数据集展示

3.2 实验过程

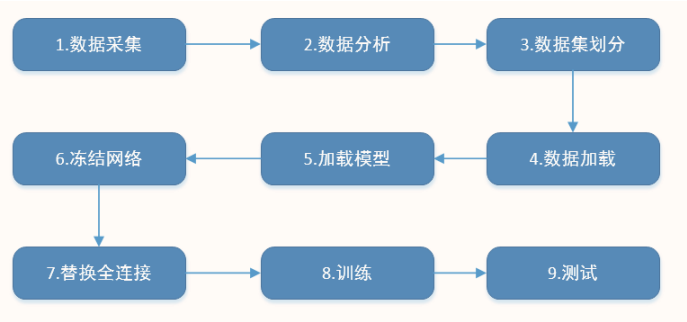

图 4 实验流程

实验流程说明：1. 首先采集我们所需要的数据集，或者使用在互联网上已采集好的数据 2. 对数据进行分析是否均匀。3. 将整个数据集的 80％图片作为实验训练集，20％图片作为实验测试集。4. 创建 dataset 并送入可迭代数据加载器中。5. 加载 GoogLeNet 来进行预训练。6. 保留末尾的一个全连接层，冻结其之上的所有层。7. 用一个权重是随机的新层来替换末尾全连接层。8. 进入训练模式，训练上一步的新层，通过前向传播和后向传播计算训练集上的损失函数值和准确率。9. 进入验证模式，计算测试集上的损失函数值和准确率。

3.4 结果分析

训练集的数据集比较庞大，所以实验的 epoch 仅设置为 20，前五轮的学习率置为 1e-3，第五轮过后缩小十倍，优化器选择 Adam，一阶矩衰减系数设为 0.9，二阶矩衰减系数设为 0.99，eps 设置为 10e-8。由图五可得，损失函数呈正常趋势下降，但在 1.25 时趋于平缓，因此可以设置更优的参数来使损失函数收敛。

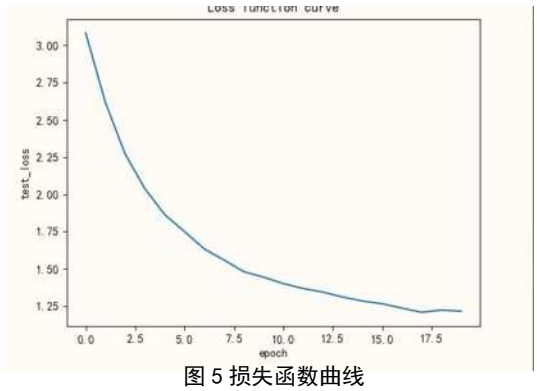

图 5 损失函数曲线

从表 1 可以看到，模型经过二十次迭代训练，测试集准确率基本在 0.7 左右浮动。

epoch	训练集准确率	测试集准确率
4	0.6782	0.6936
8	0.6898	0.6915
12	0.6879	0.7042
16	0.6944	0.7032
20	0.6898	0.7009

表 1 训练集和测试集准确率

在本实验中，使用 GoogLeNet 网络模型来对垃圾数据集进行迁移学习，从实验数据可以看出，由图 6 可得，除香蕉皮之外，其他三张图片准确率在 99％以上，因此可知，在垃圾图片清晰且场景单一时，模型得到的最高预测值都是准确的，由表一可以看到，epoch 在为 20 时测试集准确率在 70% 以上，并且所微调的参数比较单一，调优学习率，重设优化器的参数，对训练集图片做更多地调整，或者使用性能较好的设备进行成千上万次的迭代训练，可以使训练的新模型识别准确率进一步提高。

图 6 垃圾图片预测结果

论文总结

本文以研究基于GoogLeNet的垃圾分类方法为目的，进行了垃圾分类实验，使用GoogLeNet网络模型对垃圾数据集进行迁移学习得到了新的模型，并使用新模型进行了二十次的迭代训练,得出了测试集准确率之后，继续对随机的常见垃圾图片进行了测试，得到了比较高的准确率，效果良好。因此，在日常生活中，用户可以通过摄像头捕捉垃圾图像并送入模型预测，捕捉图像时尽量的使垃圾物体单一，场景选择空白单色场景来提升识别准确率，即可快速地得到垃圾的分类并准确地进行分类处理，该方法有较强的实用性。下一步的工作可以考虑使用性能更强的设备进行更高次数的迭代训练，或者加入动态学习率来使损失函数更好的收敛，以及添加使用更多的优化器来进一步提升识别准确率，提升该方法的准确率和可行性。

【参考文献】

[1] 王依欣 ， 胡冠九 ． 我国城市生活垃圾分类问题及对策建议 [J]. 产业与科技论坛,2020,19(9):201-202.

[2] 张振洲，熊凌，李克波，陈刚，但斌斌，吴怀宇．基于改进GoogLeNet的锌渣识别算法[J]. 武汉科技大学学报,2021,44(03):182-187.

[3] 李英，宋丽娟．基于GoogLeNet模型的遥感图像识别分类研究[J]. 电脑知识与技术,2021,17(12):4-6.

[4]CoudrayN,Ocampo P S,Sakellaropoulos T, et al.Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning[J].Nature Medicine,2018,24(10):1559-1567.

[5]Deng L, Abdel-Hamid O, Yu D. A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. May26-31,2013,Vancouver,BC,Canada. IEEE,2013:6669-6673.

[6] 程冉，史健芳．基于卷积神经网络的手势识别算法研究[J]. 电子设计工程,2021,29(02):179-184.

[7] 陈伟. 基于深度学习的垃圾分类算法研究[D]. 天津职业技术师范大学,2021.