

基于卷积神经网络的植物图像分类

黄泳深 刘苏卉*

江汉大学人工智能学院数学与大数据系

DOI:10.12238/acair.v3i3.15562

[摘要] 图像分类的效果受到很多因素的影响,如图像的复杂性、噪声干扰等都影响着图像的分类效果。使用传统的图像分类方法往往难以处理这些复杂情况,使得图像分类的效果不理想,因此需要借助机器学习算法来提高图像分类的准确性。本论文以卷积神经网络模型为基础,研究算法在植物图像上的分类效果,并通过图像分类实验来实现。这包括数据集的准备、算法的选择以及参数的调整等方面。通过详细的实验设计,实现了卷积神经网络在植物图像分类任务中的分类,并使用准确率、精确率等技术指标对模型的效果进行评估和比较。获得的实验结果显示,CNN模型Resnet-18达到了98%,说明效果良好。

[关键词] 植物图像分类; 卷积神经网络; 基础研究算法

中图分类号: TP183 **文献标识码:** A

Plant image classification based on convolutional neural networks

Yongshen Huang Suhui Liu*

Department of Mathematics and Big Data, School of Artificial Intelligence, Jiangnan University

[Abstract] The effectiveness of image classification is influenced by many factors, such as the complexity of the image and noise interference, which both affect the classification results. Traditional image classification methods often struggle to handle these complex situations, resulting in less-than-ideal classification outcomes. Therefore, it is necessary to employ machine learning algorithms to enhance the accuracy of image classification. This paper investigates the classification performance of convolutional neural network (CNN) models on plant images, implemented through image classification experiments. This includes aspects such as dataset preparation, algorithm selection, and parameter tuning. Through detailed experimental design, the classification of plant images using CNN was achieved, and the classification performance of the model was evaluated and compared using technical metrics such as accuracy and precision. The experimental results obtained show that the CNN model Resnet-18 achieved an accuracy of 98%, indicating good performance.

[Key words] Plant image classification Convolutional Neural networks: Fundamental Research algorithms

1 绪论

图像分类效果受图像复杂性和噪声干扰等因素显著影响。传统方法依赖手工特征提取和简单分类器,难以应对复杂场景,导致分类效果不佳。为克服这些挑战,机器学习算法逐渐引入图像分类领域。支持向量机(SVM)和决策树等算法虽有应用,但在大规模数据集和复杂图像数据处理上存在局限性^{[1][2][3]}。

近年来,深度学习算法,尤其是卷积神经网络(CNN),在图像分类领域取得突破性进展。CNN通过模拟人脑神经元网络结构,自动学习图像特征,避免手工设计特征的局限性。其核心的卷积层和池化层结构,能够有效处理图像复杂性和噪声干扰。例如,在植物图像分类任务中,CNN可自动学习叶片形状、纹理和颜色等特征,即使背景复杂或存在噪声,也能准确分类^{[4][5]}。

实验表明,CNN在图像分类任务中表现优于传统机器学习算

法,分类准确率可达90%以上。随着计算能力提升和深度学习框架完善,CNN训练和优化更高效。综上,图像分类受多种因素影响,传统方法有局限性,而CNN为提高分类准确性提供有效解决方案,应用前景广阔。

2 构架技术

2.1 卷积神经网络

卷积神经网络是一种广泛应用于图像分类深度学习模型。CNN包含卷积层、池化层以及全连接层等层级结构。CNN模型以其优秀的特征提取能力和泛化能力,在图像识别领域取得了巨大成功。其层级结构和数学原理使其能够有效地学习和表示复杂的图像数据。

2.1.1 卷积层(Convolutional Layer)

卷积层通过滤波器在输入图像上滑动进行乘法和求和操作,

提取重要特征并转化为高级特征表示, 多层堆叠可学习更复杂特征, 实现图像理解和分类, 精准捕捉与表达图像特征。

对于输入图像 X 和卷积核 W , 卷积操作的数学表达式为:

$$(X * W)(i, j) = \sum_m \sum_n X(m, n) W(i-m, j-n) \quad (2-1)$$

其中, $X(m, n)$ 是输入图像中位置 (m, n) 处的像素值, $W(i-m, j-n)$ 是卷积核中的权重值。

通过在输入图像上滑动卷积核, 可以得到输出特征图。卷积层通常会使用多个不同的卷积核, 以提取多个特征。

2.1.2 池化层 (Pooling Layer)

池化层通过降低特征图空间维度减少参数, 降低过拟合风险, 增强空间不变性。最大池化取窗口最大值, 平均池化取窗口平均值, 分别保留显著特征和整体特征。数学表达式为:

$$\text{MaxPooling}(X)(i, j) = \max_{m,n} X(i \times s + m, i \times s + n) \quad (2-2)$$

其中, s 是池化核的大小, 通常为2或3。

2.1.3 激活函数 (Activation Function)

在卷积神经网络中, 激活函数能够将输入信号进行非线性映射, 使得网络能够学习到更加复杂的特征和模式, 从而提高了网络的拟合能力和表达能力。常用的激活函数包括ReLU、Sigmoid和Tanh等。ReLU(Rectified Linear Unit)激活函数是最常用的激活函数之一, 数学表达式为:

$$\text{ReLU}(x) = \max(0, x) \quad (2-3)$$

2.1.4 全连接层 (Fully Connected Layer)

全连接层中的神经元与前一层的所有神经元相连接。这种连接方式使得全连接层能够接收并加权求和前一层的输出。通过激活函数进行非线性转换后, 全连接层将生成输出。全连接层一般位于卷积神经网络的末端, 负责对卷积层提取的特征进行分类或回归。

对于具有 n 个神经元的全连接层和输入特征向量 X , 输出结果 Y 的计算公式为:

$$Y = \text{ReLU}(XW + b) \quad (2-4)$$

其中, W 是全连接层的权重矩阵, b 是偏置向量。

2.2 ResNet

传统的深层网络在进行反向传播时, 梯度会逐层地缩小或放大, 导致深层网络的训练变得困难。而ResNet通过在网络中引入跨层的残差连接, 将输入信号直接传递到输出, 从而使得梯度能够更好地传播。这种残差连接不仅能够解决梯度消失和爆炸问题, 还能够提高网络的收敛速度和训练效果。从而提高网络的训练效果和性能。

2.2.1 ResNet-18残差块的结构

残差块是ResNet的核心组件, 它允许网络直接学习残差函

数。典型的残差块结构包括两个主要分支: 一个是恒等映射(identity mapping), 另一个是残差映射(residual mapping)。假设输入特征为 x , 残差块的输出可以表示为:

$$\text{Output} = F(x) + x \quad (2-5)$$

其中, $F(x)$ 表示残差映射, x 表示输入特征。此设计优化了网络学习恒等映射的能力, 进而有效降低了训练过程中的优化难度。

2.2.2 ResNet-18的架构

ResNet-18由多个残差块组成, 这些残差块构成了网络的主干结构。具体来说, ResNet18的网络结构包含以下部分: 一个初始的卷积层, 用以提取输入数据的初步特征; 四个残差块构成的网络主体, 负责进行深层次的特征学习和提取; 通过全局平均池化层对特征进行汇总, 并通过全连接层实现分类任务。在每个残差块中, 通常包含两到三个卷积层和一个跳跃连接。

2.2.3 跳跃连接的作用

跳跃连接允许梯度直接通过整个网络进行反向传播, 从而解决了深层网络训练中的梯度消失和梯度爆炸问题。这种连接方式使得网络能够更容易地优化深层模型。

2.2.4 损失函数和优化算法

在训练ResNet-18模型时, 常使用交叉熵损失函数来计算预测结果与真实标签之间的差异。交叉熵损失函数是多分类任务中常用的损失函数, 它能够有效地衡量模型的性能。在算法优化的过程中, 一般会选择随机梯度下降(SGD)或其衍生算法, 如Adam, 来实现损失函数的最小化。

3 实施方法

3.1 收集数据

用于分类的图像数据集来自kaggle网站下载, 研究所使用的数据集包含了30个独特植物类别的图像样本, 每个类别均包含1000张图像, 是一个总计30000张图像的数据集。数据集部分图像如图1所示, 由图可以看出图像较为清晰, 没有明显的噪声, 其整个数据集图像也是如此, 各个类别的图像都包含了该类别的各种不同的形状和不同部分, 适用于进行图像分类任务。



图1 数据集部分图像

3.2 实验过程说明 (变量和参数设置)

本次实验采用深度学习框架pytorch, 使用CNN模型中的ResNet-18预训练模型进行植物图像分类实验。

ResNet-18模型的预训练权重: 我们选择使用在ImageNet数据集上预训练的ResNet-18模型, 并加载其预训练权重作为初始模型参数。

损失函数: 交叉熵损失函数是模型优化的重要工具。其显著优势之一在于, 对于梯度消失问题, 它能够进行妥善的处理。具体而言, 当预测值与真实值之间存在较大差距时, 交叉熵损失函数的梯度将变得相对较大。这一特性使得在反向传播过程中, 梯度能够更为有效地进行传递, 从而优化模型的性能。

此外, 交叉熵损失函数在分类问题中通常表现出色, 并且与softmax激活函数结合使用时, 可以直接衡量模型输出的概率分布与真实标签的差异。

优化器: 我们采用Adam优化器, 这是一种基于梯度的优化算法, 通过计算每个参数的梯度, 并根据这些梯度来更新参数。Adam优化器的核心在于它结合了动量优化和自适应学习率方法, 能够自适应地调整学习率, 并且通常表现出良好的性能。

学习率设定为0.001。

3.3数据预处理

本实验使用30类植物图像数据集, 每类1000张, 其中800张用于训练, 200张用于测试。数据集的图片大小分布如图2所示, 可以看出大部分的图片大小都在小于500×500像素大小的范围, 也有相当一部分的图片在这个范围之外, 图像大小分布不是很均匀, 后续图像预处理需对图像大小进行统一, 便于图像数据的训练。由于数据集未带标签, 所以需要手动获取标签, 为方便获取标签, 需对数据集图像进行重命名操作, 数据集图像重命名后, 获取各个类别图像的标签, 并保存为CSV文件, 方便后续进行图像分类。

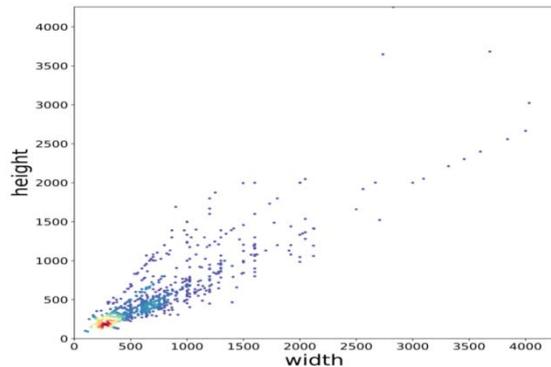


图2 数据集图像大小分布

图像预处理包括调整大小、颜色空间调整和边缘增强。特征提取涉及颜色直方图、LBP纹理和Canny边缘检测。颜色直方图描述颜色分布, LBP通过比较像素及其邻域的灰度值来获取纹理特征, 而Canny算法通过高斯滤波、梯度计算等步骤来识别图像边缘。使用transforms.Resize()方法将图像大小进行统一, 大小调整为224×224像素, 使用RGB颜色三通道; 使用transforms.ToTensor()方法将图像数据转换为pytorch张量(Tensor), 将图像的像素值从[0, 255]范围(在uin类型中)调整到[0.0, 1.0]范围(在f1oa2或f1oa4类型中)。这种归一化处理可以帮助模型

更好地学习数据的特征, 同时有助于避免梯度爆炸或消失的问题; 使用transforms.Normalize()方法对输入的图像数据进行均值和标准差的标准化处理, mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225], 以便适应ResNet-18模型的输入要求。

3.4模型训练

将数据集划分为三个子集: 训练集、验证集和测试集。具体比例为: 训练集占75%, 验证集和测试集各占15%。这样的划分方式有助于我们充分利用数据资源。

数据集划分情况如图3所示:

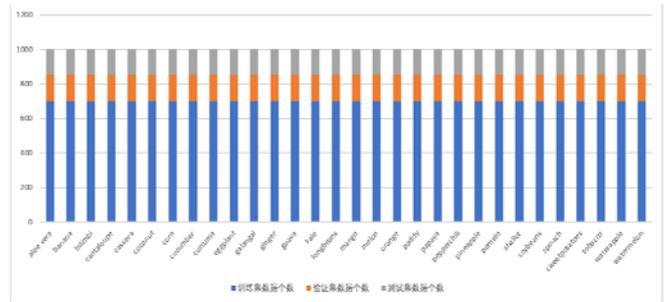


图3 各类别图像数量

使用训练集对ResNet-18模型进行训练, 采用交叉熵损失函数进行优化, 训练轮次10轮, 在训练中将数据传输到GPU上, 使用GPU加速训练过程, 在训练时使用scheduler和adam优化器一起动态调整学习率, 步长step_size=5, gamma=0.1, 即在每经过5个epoch(或者步数)之后, 优化器的学习率将被乘以0.1, 即降低为原来的0.1倍, 使用学习率调度器, 可以更好地控制模型的训练过程, 有助于模型更快地收敛或更好地适应训练数据。在训练时记录模型在训练集上的准确率和损失值。

为确保模型性能的优化, 我们将依据验证集上的准确率和损失值对模型进行性能评估。在此基础上对模型参数进行细致的调整。其训练及验证过程如表1所示:

表1 训练及验证过程

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1	0.8371	78.61%	0.8655	79.60%
2	0.276	90.89%	0.9122	81.78%
3	0.4009	87.68%	0.6691	85.29%
4	0.2585	91.79%	0.6161	84.27%
5	0.1726	94.30%	0.3794	95.40%
6	0.0387	98.77%	0.0831	97.29%
7	0.0139	99.59%	0.0934	97.82%
8	0.0127	99.77%	0.1015	97.71%
9	0.0035	99.91%	0.1087	97.67%
10	0.0018	99.95%	0.1219	97.87%

training time: 2096.580204515457s.

最终, 在测试集上评估经过训练的模型的性能。

3.5 实验结果评价标准

本文使用准确率、精确率等技术指标对不同模型的效果进行评估和比较。准确率(Accuracy): 准确率是指分类器正确预测的样本数与总样本数之比。其计算公式为:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3-1)$$

其中, TP, 即真正例, TN, 即真负例, FP, 即假正例。而FN, 即假负例。

精确率(Precision): 精确率是指分类器在预测为正类别的样本中, 真正例的比例。其计算公式为:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3-2)$$

召回率(Recall): 召回率, 是指在实际为正类别的样本中, 被分类器正确预测为正类别的样本所占的比例。其计算公式为:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3-3)$$

F1分数: F1分数是精确率和召回率的调和平均值, 其计算公式为:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3-4)$$

4 结语

从表2可以看出resnet18模型在本次植物图像分类实验中准确率达到了98.22%, 精确率为98.229%, 召回率为98.222%, 分数为98.212%, 总花费时间为2081.6057s。这些指标表明了模型在植物图像分类任务中表现出很高的性能, 模型对于分类任务的执行非常准确和可靠。

卷积神经网络算法在图像特征提取和分类方面具有优势。这是因为卷积神经网络能够自动学习图像中的特征, 并通过多层

网络进行分类。然而, 训练和调优卷积神经网络需要更多的时间和计算资源。为了提高训练效率, 可以采用并行计算、分布式训练等技术。此外, 还可以利用预训练的模型来加速网络的收敛过程。

表2 resnet18模型图像分类结果

模型	Accuracy	precision	recall	F1	Total_time(s)
Resnet18	0.9822	0.98229	0.98222	0.98212	2081.6057

[参考文献]

- [1]李蕾.一种改进的SVM遥感图像分类方法[J].北京测绘,2023,37(06):903-907.
- [2]张伟.基于改进果蝇算法的支持向量机参数优化研究[D].安徽理工大学,2022.75.
- [3]张粮,王璇之.基于决策树算法的水果分类识别[J].科技创新与应用,2024,14(06):18-23.
- [4]陈雪鑫.基于多特征融合的双路深度信念神经网络水果分类技术研究[D].青岛大学,2020.75.
- [5]王爱丽,丁姗姗,刘和.空谱域适应与XGBoost结合的跨场景高光谱图像分类[J].光学精密工程,2023,31(13):1950-1961.

作者简介:

黄泳深(2001--),男,汉族,广西北流市人,本科,研究方向:数学与大数据。

*通讯作者:

刘苏卉(1988--),女,汉族,江苏靖江市人,博士,职称:讲师,研究方向:数值线性代数,机器学习。