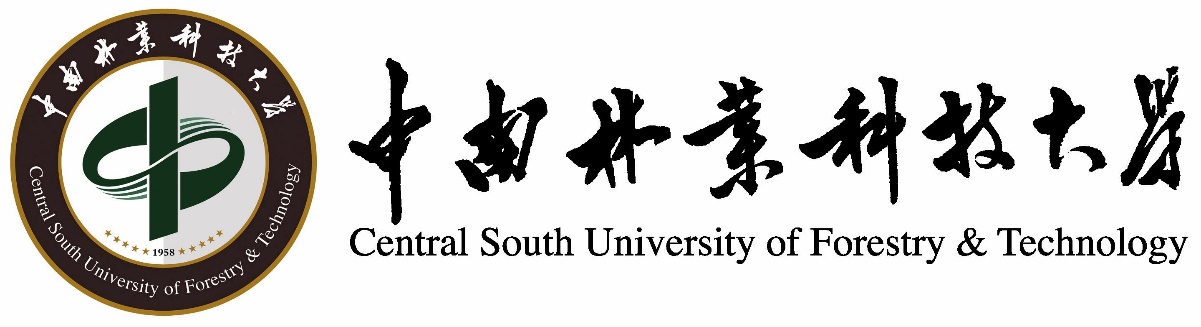
分类号： 密 级：

学校代号： 10538 学 号： 20202524





本科毕业（学士学位）设计

基于多特征融合和深度学习的植物叶片识别

学生姓名：  曾祥文

指导教师：邝祝芳教授

培养学院：计算机与信息工程学院

年级专业：2020级计算机科学与技术

提交日期：2024年6月1日

中南林业科技大学

毕业设计原创性声明

本人郑重声明：所呈交的设计是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本设计不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

学生签名： 日期：2024年 月 日

毕业设计版权使用授权书

本毕业设计作者完全了解学校有关保留、使用设计的规定，同意学校保留并向国家有关部门或机构送交设计的复印件和电子版，允许设计被查阅和借阅。本人授权中南林业科技大学可以将本设计的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本设计。

本设计属于

1、保 密 ，在 年解密后适用本授权书。

2、不保密 。

（请在以上相应方框内打“√”）

学生签名： 日期：20 年 月 日

导师签名： 日期：20 年 月 日

基于多特征融合和深度学习的植物叶片识别

摘 要

植物在医学、农业、环境保护以及生态保护中起着极其重要的作用。近年来，随着环境污染加剧，每年都有大量植物物种不断死亡甚至灭绝，保护植物多样性变得越来越重要，而保护植物的前提就是识别植物。因此，伴随着计算机视觉相关技术的发展，基于叶片图像的植物识别方法已经成为研究的热点。

本文针对 Flavia植物叶片数据集，使用GoogLeNet模型和ResNet模型来对植物叶片进行分类识别。

Plant Leaf lmage Recognition Based on Multi-feature Integration and Deep Learning

**Abstract**

目 录

[目 录 IV](#_Toc167283258)

[**1 引言** 1](#_Toc167283259)

[**1.1 研究背景和意义** 1](#_Toc167283260)

[**1.2 国内外研究现状** 2](#_Toc167283261)

[**1.3 论文结构** 7](#_Toc167283262)

[**2 相关基础知识** 7](#_Toc167283263)

[**2.1 深度学习算法研究** 7](#_Toc167283264)

[2.2 植物叶片图像特征处理与计算 14](#_Toc167283265)

[2.3 植物叶片识别常用网络模型介绍 17](#_Toc167283266)

[2.4 总结 24](#_Toc167283267)

[3 基于GoogLeNet网络模型的植物叶片识别 25](#_Toc167283268)

[**3.1 数据集及预处理** 25](#_Toc167283269)

[3.2模型设计 26](#_Toc167283270)

[3.3 实验过程 27](#_Toc167283271)

[3.4 实验结果与分析 28](#_Toc167283272)

[**3.5 总结** 31](#_Toc167283273)

[4 基于ResNet网络模型的植物叶片识别 31](#_Toc167283274)

[**4.1 数据集及预处理** 31](#_Toc167283275)

[4.2模型设计 33](#_Toc167283276)

[4.3 实验过程 35](#_Toc167283277)

[4.4 实验结果与分析 36](#_Toc167283278)

[参 考 文 献 40](#_Toc167283279)

[致 谢 43](#_Toc167283280)

**1 引言**

* 1. **研究背景和意义**

1.1.1 研究背景

植物在食品、医学、工业和日常生活中扮演着至关重要的角色，例如供应食物和维持大气中的氧气平衡。然而，由于人类的生产活动不断增加，城市化进程加速，过度开发、过度砍伐、全球气候变暖、环境破坏和污染等因素，导致植物的生态环境受到严重破坏，每年都有大量植物物种灭绝。如果不采取措施，这将对人类社会造成严重后果。全球生物多样性正在迅速减少[25]。根据Knapp等人[26]的调查，已知的植物物种约有25万种，其中近13%濒临灭绝。植物物种大规模消失将给人类社会和生态系统带来负面后果，比如导致土地荒漠化、洪水频发以及极端气候事件增多。同时，不同物种之间相互依赖，一种植物的消失可能导致其他物种的灭绝。越来越多的证据表明，生物多样性的减少会对生态系统带来不利影响。因此，保护植物物种显得尤为迫切和重要[27]。

对于未来的植物多样性保护，准确掌握植物的特性和地理分布信息至关重要[28]。为了保护植物物种，首要任务是认识它们，了解它们的特征和分布情况。无论是农业信息化建设还是生态保护，都需要对植物进行分类研究。然而，鉴别植物物种是一项艰巨的任务，因为地球上已知的植物超过27万种，其中许多尚未得到充分研究。人类通过传统的人工方法识别植物物种是复杂且耗时的，而且容易出错。此外，由于需要使用特定的植物学术语，使得非专业人士难以理解，因此人工识别大量现存植物物种几乎是不可能的。

植物的自动识别系统对植物分类学的研究者以及广大群众来说都具有重要意义[29]。随着计算机技术的进步，包括软硬件的提升、便携设备的普及、图像处理和机器学习技术的发展，植物物种的自动识别已经变得可行。近年来，越来越多的个人和团队投入到植物物种自动识别方法的研究中，主要参与者包括计算机视觉等相关领域的专家学者。植物物种的分类和识别可以通过植物的不同器官特征来实现，例如花、茎、果实、叶、根和种子。植物叶片的形状和结构相对稳定，通常呈扁平状，接近平面。叶片具有生命周期长、易于获取和特征明显等优点，因此专家学者们往往选择植物叶片图像进行分类研究。



图1-1 植物的六大器官

植物物种的识别可以通过叶片的形状、颜色和纹理等特征来实现。尽管叶片通常呈绿色，但由于自然环境的影响，颜色特征并不常用于识别。不同种类的植物叶片在形状上往往存在显著差异，因此形状特征成为识别植物物种的重要依据。叶片的纹理特征主要涉及叶脉及其内部结构信息。同一种植物的叶脉结构通常相似，而不同种植物的叶脉结构差异较大，这使得纹理特征也成为植物识别的重要依据。

传统的机器学习方法通常利用形状和纹理等特征来设计植物识别算法，但这些方法通常需要大量的预处理，且设计的算法通常只能在特定或某类数据集上表现良好。在拥有足够多的叶片图像样本时，深度学习能够自动学习到高辨识度的叶片特征，从而实现对植物叶片图像的有效识别。因此，基于卷积神经网络（Convolutional Neural Networks, CNN）[30]和叶片图像的植物物种识别算法正在逐渐兴起。通过卷积神经网络，系统能够自动提取并学习叶片的复杂特征，这显著提高了识别的准确性和效率。与传统方法相比，基于CNN的深度学习方法减少了对人工特征设计的依赖，能够在更广泛的数据集上表现出色。随着大量叶片图像数据的积累和深度学习技术的不断发展，基于CNN的植物识别方法在研究和应用中正变得越来越重要。

**1.2 国内外研究现状**

科技的不断进步促使不同学科之间建立了密切联系，形成了一种融合的趋势。计算机图像识别技术在农业领域得到了广泛应用，并取得了长足的发展。对植物叶片进行分类识别主要利用图像识别技术，通过对叶片图像的基本特征进行建模和种类判别，以解决人眼识别效率低下的问题。近年来，植物叶片识别研究已成为国内外科研专家、学者以及相关领域的研究热点。基于深度学习的植物叶片识别研究也不断涌现出新的思路和方法。

1.2.1 植物叶片识别领域的研究现状

图像识别技术正日益成熟，并在不断扩大应用范围。早在1986年，国外就已经开始对植物叶片识别进行了一系列研究。Imgrouile采用多特征融合的方法提取橡树叶片特征，利用降维算法进行橡树分类[13]。Osika及其团队将叶片形态特征和面积参数作为识别标志，运用BP前馈神经网络算法进行分类[32]。Rashad等人从植物图像数据库中选择10种不同类型的叶片，提取纹理特征并进行训练，取得了良好的识别效果[33]。Lee提出了一种新型的植物叶片分类方法，利用傅里叶变换将频域信号作为特征依据，提取叶脉特征，从而提高了识别率[34]。Mehdipour Ghazi及其团队采用PCANet算法对叶片进行分类，其单一背景叶片的识别率远高于复杂背景叶片，达到了90%的识别率[35]。DivyaTomar等人利用基于双梯度最小二乘支持向量机(DAG-MLSTSVM)的分类器构建了植物种类叶片识别系统，其识别准确率为84.70%[36]。Krizhevsky等人构建了AlexNet卷积神经网络，并以ImageNet数据集为训练对象，测试失误率为15.4%[37]。SoonJyeKho等人通过提取叶片形态特征，结合图像处理和模式识别技术，利用人工神经网络和SVM训练分类模型，开发了三种榕属植物的自动识别系统，其识别准确率达到83.3%[39]。EMImah提出了一种基于竞争学习算法GRLVO，利用全特征属性对15种植物进行自动分类，其精确度可达92.98%[39]。

在国内，植物叶片识别技术起步较晚。在2001年，傅弘等人为提高叶片识别准确率，利用叶片脉络纹理作为特征，运用边缘梯度等统计特征方法进行分类识别。经过训练，准确度有了显著提高，为植物叶片识别奠定了基础[1]。杜吉祥针对植物叶片识别研究，提出了一种移动中值重心超球分类器(MMC)的方法，对选取的15种叶片特征进行提取和分类[2]。王晓峰等人基于传统识别方法，根据叶片图像特征及形状参数进行分类识别，取得了良好的识别率[3]。朱宁利用LBP方法提取植物叶片图像的纹理特征，结合模式识别技术和数字图像处理技术，设计实现了树叶识别系统[4]。贺鹏首次将概率神经网络(PNN)及分类器应用于阔叶树识别研究，准确率达到98.3%[5]。阚江明等人提出了多种图像处理方法，对叶片图像进行预处理后识别，平均正确率为70.83%[6]。张宁结合图像分析技术，深入研究植物叶片识别相关理论方法，介绍了植物识别技术的现状和发展趋势[7]。王丽君利用支持向量机(SVM)分类器提取了植物的26个特征，将观叶叶片识别准确率提高至91.41%[8]。杨天天等人对几种柳属(Salix)植物的单一背景叶片进行分析，融合多种叶片特征信息，通过分类提取后识别效果显著提升[9]。于慧伶等人利用双路卷积神经网络结构对Flavia公共叶片数据集进行分类，提高了植物叶片的识别准确率[10]。朱良宽等人利用CVL研究领域中的瑞典植物叶片集，提出一种融合DCGAN与迁移学习的识别方法[11]。李龙龙等人利用LBP算法度量和提取图像局部的纹理信息，提高了训练速度[12]。

传统的图像识别技术是当前对植物叶片进行识别研究的主流方法。这些方法通常是将提取到的植物叶片特征输入到机器学习模型中进行训练，取得了令人满意的效果。然而，在图像识别过程中，常常使用人工设定的特征进行提取，并且所使用的数据集也是特定的。如果在后续实验中更换数据集，结果可能会有较大差异。相比之下，深度学习通过独特的深度网络结构能够自动学习数据特征信息，从而加快了识别速度，并且更好地适用于各个领域。因此，随着机器学习领域的发展，专家学者们越来越广泛地将深度学习应用于图像识别研究中。

1.2.2 深度学习在图像识别领域的研究现状

人工智能的崛起是信息时代的亮点之一，而机器学习是人工智能的重要组成部分，其中深度学习算法尤为突出。在传统机器学习方法中，植物叶片识别方式相对固定，通常通过提取叶片的形状、纹理和脉络等特征，然后将这些特征输入分类器进行识别。然而，传统方法对图像数据集的要求较高，特征提取方式的差异可能会影响识别效果，导致模型泛化能力差，准确度低。随着时间的推移和硬件技术的进步，深度学习得到了广泛发展，并在图像识别领域得到了广泛应用。其独特的架构模式和高效的学习能力引起了国内外的广泛关注，受到了世界各地相关领域研究专家的重视。

在国外，深度学习在2006年迎来了热潮。Hinton等人构建了深度信念网络(DBN)，采用无监督训练方法，并结合反向传播算法来优化模型参数[40]。随后，2012年，杰弗里等人提出了具有多个隐藏层的深度学习模型AlexNet, 哥伦比亚大学、马里兰大学和史密森尼学会联合发起了Leafsnap项目，开发了一个植物叶片识别的APP，采用了美国东北部的185种叶片。APP使用的截图如图1- 所示。Jeon等人对Flavia数据库中的植物叶片进行分类训练，构建了GoogLeNet模型来提取特征信息，并通过调整参数来提高识别率[41]。Ghazi等人在PlantCLEF2016比赛中发现，使用改进后的AlexNet、VggNet和GoogLeNet模型识别植物种类的方法优于其他方法[42]。

在国内，许多研究者将深度学习算法应用于植物叶片识别领域。尽管国内在深度学习方面的起步较晚，但其发展迅速，在农业领域取得了显著成就，例如病虫害识别、农作物自动化采摘和植物叶片识别等方面。朱静等人通过对图像进行处理并提取叶片特征信息，将14种植物的识别准确率提高到了92%[14]。龚丁禧等人基于深度学习理论，构建了卷积神经网络模型，对Swedish叶片数据集中的15种叶片进行训练，并不断调整参数，结果优于传统的叶片识别算法[15]。丰晓霞等人引入了深度学习中的常用模型分类方法，并将softmax分类器替换为SVM对提取到的特征进行分类[16]。张帅以PlantNet叶片库和自扩展的叶片作为训练样本集，采用分割算法对复杂和单一背景的叶片进行处理，通过CNN与分类器相结合的方式对叶片特征进行训练，其中SVM分类器的识别率高于softmax分类器，单一背景取得了较好的分类效果，但在复杂背景下识别率较低[17]。薄琪苇利用深度学习算法构建了植物叶片识别系统并应用于Android端，不仅对15种叶片图像取得了不错的识别率，还创新了识别模式[18]。程曦等人为了提高储粮害虫图像的识别率，构建了AlexNet与GoogLeNet卷积神经网络模型进行训练，取得了较好的分类效果[19]。2019年，黄志国将中草药叶片作为识别对象，对LeNet-5模型进行了改进，通过对比分析，改进后的模型不仅降低了训练时间，还提高了识别正确率[20]。王艳等人利用深度学习相关算法对中草药图像进行了分类识别，取得了良好的效果[22]。孙颖异等人针对叶片识别提出了一种新的方法，即在AlexNet的基础上进行改进，得到基于残差连接的CNN模型，并对PlantVillage提供的12种植物叶片数据集进行分类和识别，证明了带有残差连接的AlexNet卷积神经网络能够进行高效识别[23]。樊湘鹏等人利用复杂背景的玉米病害图像，进行了图像预处理操作，设计了一个具有5层卷积、4层池化和2个全连接层的卷积神经网络结构，开发了基于移动端的玉米田间病害识别系统，实验平均识别准确率达到了83.33%[24]。

综上所述，本研究计划采用多特征融合和深度学习的方法，旨在解决传统方法在植物叶片识别中的局限性。通过综合利用形状、纹理和颜色等多个特征，并通过深度学习进行训练和分类，本研究旨在提高植物叶片识别的准确性和鲁棒性，从而推动植物叶片识别技术的发展，提高植物资源的利用效率，保护生态环境。

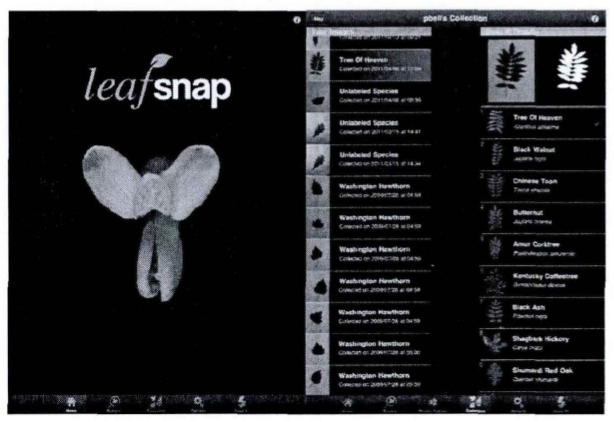


图1-2 ipad版本Leafsnap的截屏

* 1. 论文结构

本论文是基于多特征融合和深度学习的植物叶片识别，文章的组织结构如下：

第一章，引言部分。本章首先介绍的是植物叶片识别的研究背景和意义；然后，对国内外在该领域的研究历程进行了梳理，罗列了一些具有代表性的成果。

第二章，植物叶片识别和深度学习的相关基础知识部分，本章节首先介绍的是深度学习的基础知识，详细阐述了CNN 的各个组成部分。然后介绍了深度学习的两种框架Pytorch和TensorFlow；最后，介绍了常用的植物叶片识别的网络模型，包括AlexNet，VGG，GoogLeNet和ResNet模型。

第三章，详细介绍了GoogLeNet的网络模型结构并且使用GoogLeNet模型对植物叶片进行识别实验。

第四章，首先使用ResNet18模型对植物叶片进行识别，为了提升植物叶片识别的准确率对ResNet18模型进行改进；同时也采用了ResNet34网络模型和ResNet18模型做对比试验；最后通过改变ResNet34模型的实验参数Batch\_Size来获得识别率最高的Batch\_Size。

**2 相关基础知识**

**2.1 深度学习算法研究**

2.1.1 深度学习相关概述

随着人工智能的迅速发展，机器学习成为了实现人工智能的重要方法之一。机器学习旨在从数据中获取预测函数，通过对目标数据的分析来学习数据之间的规律，从而实现对真实事件的识别和决策。在图像识别领域，机器学习应用广泛，其主要思想是通过获取图像数据并提取特征，再利用分类器学习历史数据中的规律，以实现对真实场景的识别。

然而，传统的机器学习方法通常需要人工设计特征，如尺度不变特征（SIFT）、DOG金字塔（Difference of Gaussian）和全局特征（HOG）等，这在特征选择和分析方面存在一定的挑战。为了解决这一问题，近年来出现了一种新的研究方向——深度学习。深度学习利用多层神经网络来自动提取特征，从而减少了特征提取的时间和工程投入。

深度学习的发展源远流长。自2006年以来，深度学习在机器学习领域取得了巨大的进展，特别是在图像识别领域。2012年，Hinton教授和他的学生提出了AlexNet网络模型，并在ImageNet大规模视觉识别竞赛中夺得了冠军，标志着深度学习在图像识别领域的成功应用。之后，VGG模型和ResNet模型相继出现，进一步推动了深度学习在图像识别领域的发展。

相较于传统的神经网络，深度学习算法通过增加网络层数来自主学习主要特征，减少了特征提取的工程复杂性。深度学习网络结构包含多个隐藏层，其中每一层都以相邻节点相连，并且具有不同的激活函数，从而实现了更好的泛化能力。深度学习算法的设计理念借鉴了人脑的分层结构，使得其结构更加接近人脑神经系统的运作方式。

深度学习的发展将进一步推动人工智能技术的进步，为解决各种现实世界的问题提供了新的思路和方法。

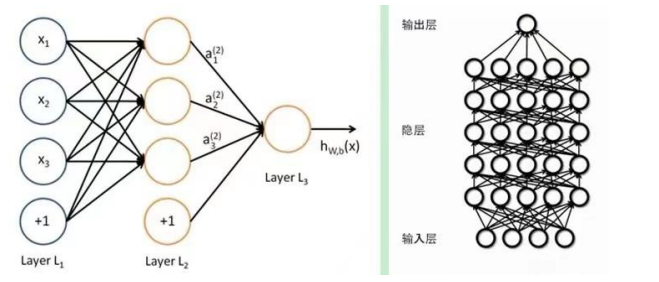


图2-1 单隐层和多隐层深度学习模型对比图

深度学习的核心思想在于利用深度网络模型作为主要手段，通过构建特征来自主训练海量数据，从而不断提高模型在识别和决策方面的准确率。深度学习的优势集中表现在以下两个方面：其一是网络架构的“深度化”，即模型含有多个隐层和节点，能够逐层地提取数据信息，用较少的参数来表示复杂的函数，从而使得分类结果更加准确；其二是特征提取的“关键性”，即模型能够自主学习特征规律，丰富数据的内在信息，从而提高了模型的识别能力。

深度学习在人工智能领域已经被证明是最先进的技术之一，其发展势头强劲。从模式识别技术到分类任务，从个性化推荐系统到购物助手，从医疗预防诊断到疾病预测等等，深度学习都在不同领域展现出了强大的应用潜力，为我们的生活带来了意想不到的惊喜和便利。

2.1.2 深度学习训练过程

在传统的神经网络训练中，一般采用梯度下降法。这种方法通过设定初始参数，计算网络输出与标签之间的差值，并不断调整参数直至达到收敛。然而，在训练过程中存在一些不足之处。例如，BP神经网络在训练过程中自上而下的顺序可能导致梯度下降变得稀疏，从而造成较大的误差，并且对于没有标签的数据样本存在一定的局限性。而深度学习网络能够在训练过程中修复这些不足。

神经网络同时对每一层进行训练可能会导致训练时间过长，难度增加。若每次只训练一层，很容易产生偏差，导致过拟合现象。为了解决这一问题，2006年，Hinton率先采用了无监督学习的方法构建多层神经网络。该方法主要分为两步：首先，设置层与层之间的单元节点；其次，引入wake-sleep算法调整模型参数，其中wake阶段用于参数调整记录，sleep阶段用于反馈结果输出。最终，通过调优算法构建深度学习网络模型。

在图像识别领域，深度学习网络的训练过程通常遵循以下五个步骤：

1. 数据预处理：对训练数据进行标记和归一化处理，以提高模型的识别性能。
2. 正向传播法：采用有监督学习方式，将标定好的数据输入神经网络中逐层训练，调整每个神经元的权重，并引入激活函数计算中间变量。
3. 计算误差函数：降低损失函数的值，防止模型过拟合，通过准确值与实际识别率之间的误差来判断识别效果。
4. 反向传播法：采用无监督学习方式，对有标签或无标签的数据进行训练，通过反向传播算法得到误差函数的最小值，确定梯度向量，并整合参数值。
5. SGD算法：通过梯度向量调整每个权值，使得误差趋于最小值，对于非凸损失函数，调节收敛趋势以达到更好的训练效果。

2.1.3 卷积神经网络

卷积神经网络(CNN)是一种前馈神经网络、强大的神经网络，在深度学习领域中具有广泛应用且备受瞩目。CNN 的主要优势是在于其能够自动学习相关特征但不需要进行人工监督。目前，CNN 在多个领域中得到了广泛应用，尤其是在计算机视觉等领域。

卷积神经网络（CNN）是一种层层相连的多层网络结构，其中每一层都包含多个特征图和独立的神经元。一般来说，CNN由三部分组成：输入层用于接收原始数据信息，中间包含多个卷积层和池化层，主要用于对输入图像进行卷积操作以提取有用信息和学习特征，最后是全连接层，将提取的特征输入到分类器中进行分类，然后由输出层展示结果。在卷积层中，滤波器用于提取输入图像的特征值，而池化层则对输入图像进行降维操作，以避免模型过拟合。如图2-所示，是卷积神经网络结构示意图

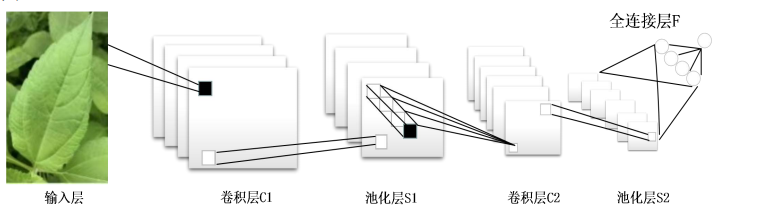


图2-2 卷积神经网络结构示意图

(1)卷积层

卷积层(convolutional layer)是CNN中的最重要的组成部分，可以用于提取输入图像中的特征。卷积层主要是通过滑动一个固定大小的滤波器（filter）或卷积核（kernel）在输入图像上进行卷积操作，生成一个新的特征图（feature map）。在这个过程中，滤波器中的权重与输入图像的一部分进行乘积运算，并将结果相加，得到特征图中对应像素的值。图2-10展示输入大小为5\*5，卷积核大小为3\*3的卷积运算。

卷积层通常包含一个或多个卷积核，每个卷积核可以捕获输入图像中的不同特征。例如，一些卷积核可能会捕获边缘、线条或纹理等低级特征，而另一些卷积核则可能会捕获更高层次的语义特征，如形状、物体部分或场景。

卷积层包括一些重要的超参数，如卷积核的大小、步长（stride）和填充（padding）等。卷积核的大小决定了每个卷积核能够捕获的特征的空间范围。步长决定了每个滤波器在输入图像上滑动的距离，影响输出特征图的空间分辨率。填充是在输入图像的边缘周围添加一些像素，以控制输出特征图的大小和形状。

卷积层的输出可以通过应用非线性激活函数，以增加模型的非线性能力。其中常用的激活函数之一是ReLU函数。在训练期间，通过反向传播算法可以更新卷积层中的权重，以最小化损失函数并提高模型的准确性。

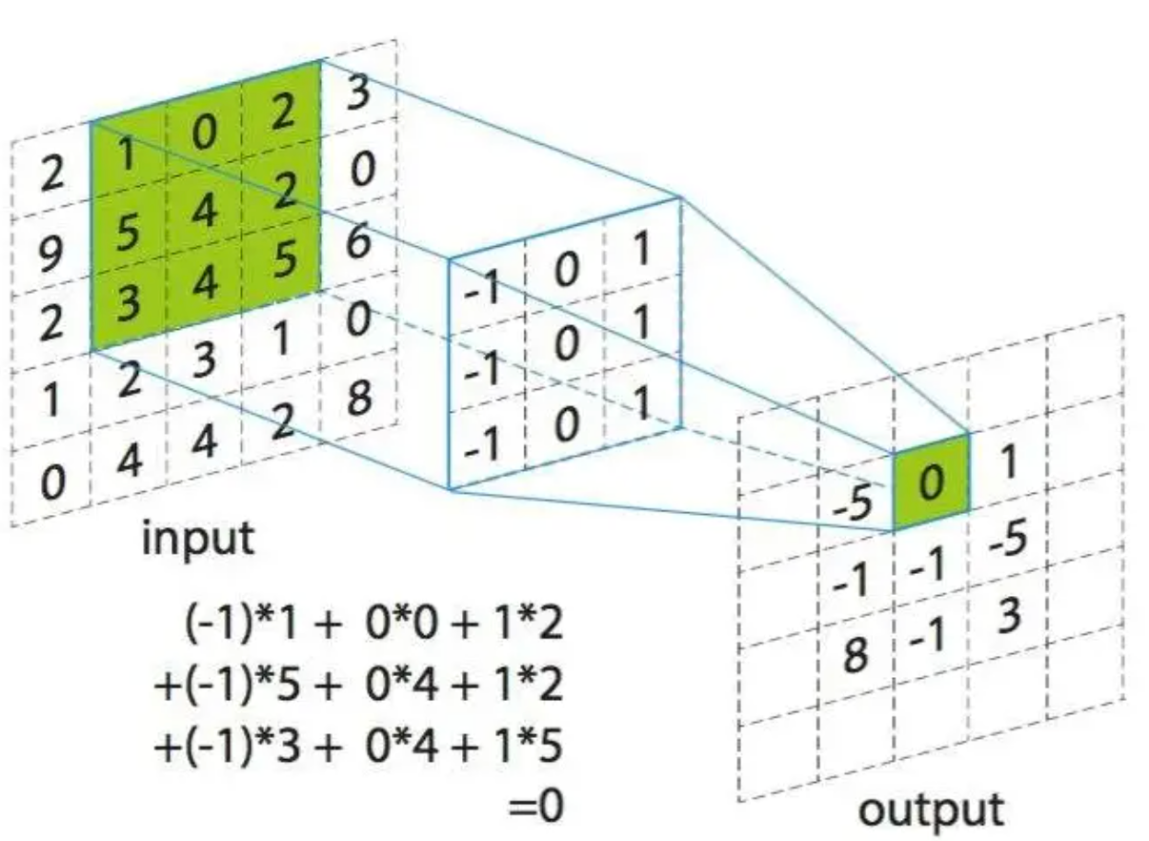


图2-3 卷积运算示意图

卷积层在CNN中具有权重共享的优势，这意味着相邻层的神经元之间没有独立的权重分配。相反，整个卷积层使用一组权重与输入矩阵的所有像素一起进行运算。这种共享权重的机制大大减少了训练过程中所需的参数数量，降低了模型的复杂度，并显著减少了训练时间和计算成本。

(2)池化层

池化层（Pooling layer）是卷积神经网络（CNN）中的一种常见层次。它通常紧接着卷积层使用，用于缩小特征图来尺寸，减少参数数量同时降低过拟合的风险。在池化过程中，主要信息或特征被保留在每个池化步骤中。

常见的池化操作主要包括最大池化（Max Pooling）、平均池化（Average Pooling）和全局平均池化（Global Average Pooling, GAP）。最大池化是在一个池化窗口内选取窗口中的最大值作为输出，这种方法能够突出显著特征，有助于保留图像中的重要信息。平均池化是指在一个池化窗口内，将窗口中的像素值的平均值，并将其作为输出值，这有助于平滑图像，并减少对细节的过度关注，从而提高模型的鲁棒性。全局平均池化（Global Average Pooling，GAP）是一种池化层的形式，其将输入的每个特征图的每个通道的元素平均化，得到一个标量输出。利用图2-11来说明这三种池化操作。

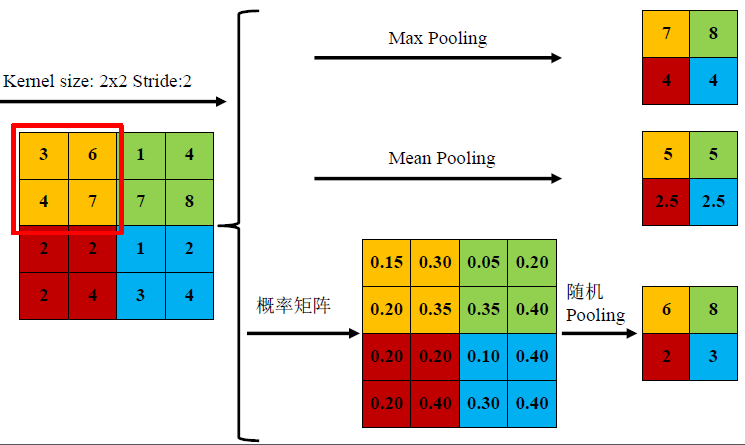


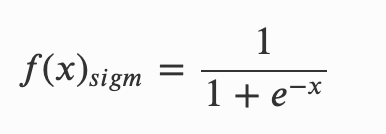
图2-4 三种池化操作

在上图2-4中，输入特征图为4\*4的矩阵，步长为 2，池化窗口大小为 2\*2，所以最大池化和平均池化后输出的特征图为 2\*2 的矩阵，全局平均池化的输出为一个值。

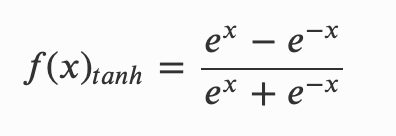
(3)激活函数

激活函数在神经网络中起到将输入映射到输出的非线性转换的关键作用。在卷积神经网络（CNN）架构中，非线性激活函数通常用于可学习层（例如卷积层）之后。这些激活函数的非线性性质使得输入与输出之间的映射成为非线性关系，使得CNN能够学习到更复杂的特征。激活函数还需要具备可区分性的特点，这对于使用误差反向传播来训练网络是非常重要的。以下是在CNN最常用的激活函数：

Sigmoid：如图2-5，输入是实数，而输出范围在0到1之间。sigmoid 函数曲线是 S 形的，表达式如下：

 (2-4)

Tanh：如图2-6，它类似于 sigmoid 函数，输入是实数，但输出范围在-1和1之间，表达式如下：

 (2-5)

ReLU：如图2-7，是一种简单且使用频率较高的激活函数。当输入值大于零时直接返回输入值，而在输入值小于等于零时返回零。表达式如下：

 (2-6)

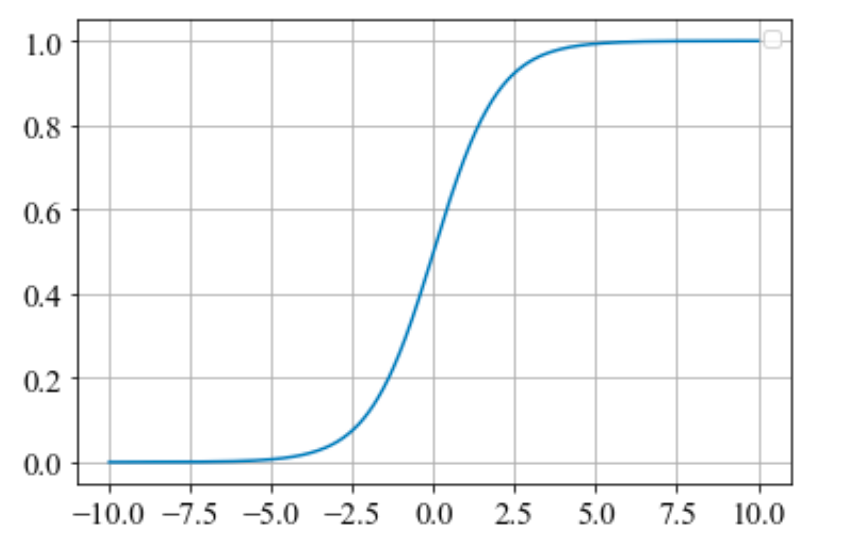


图2-5 Sigmoid 函数图

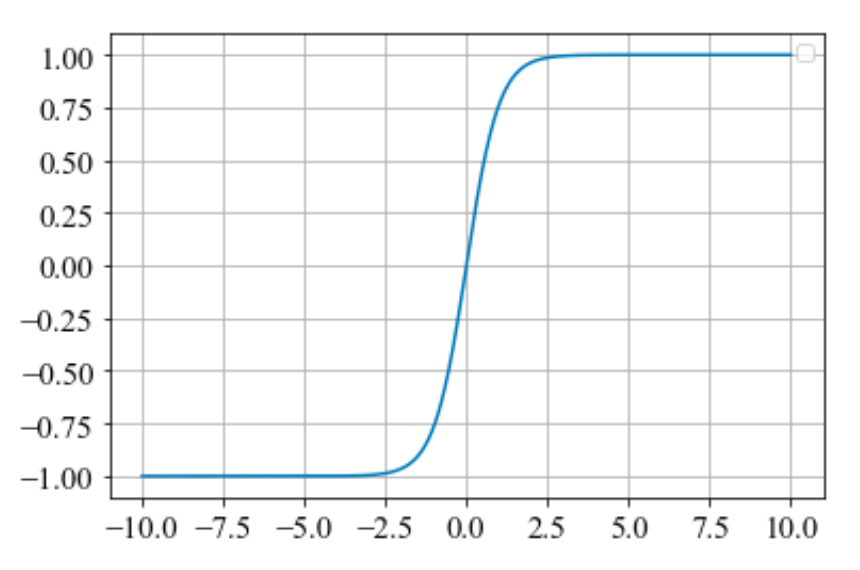


图2-6 Tanh函数图

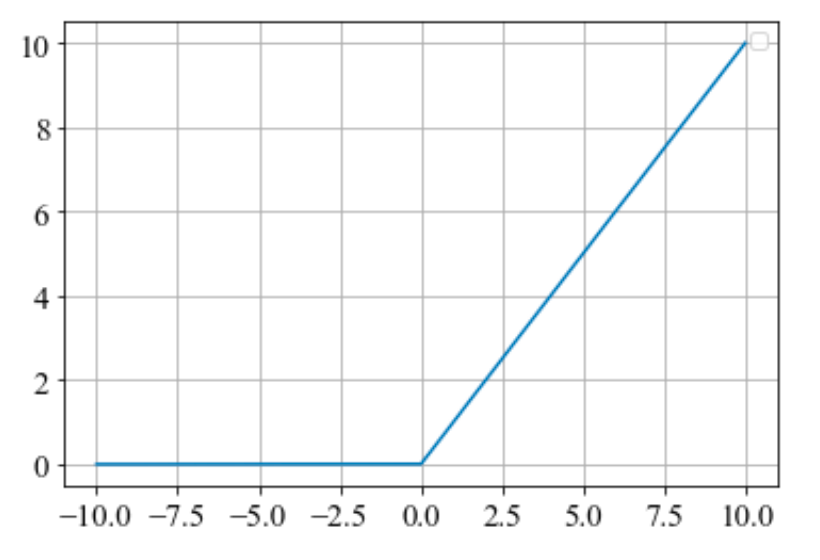


图2-7 ReLU函数图

(4)全连接层

全连接层是卷积神经网络（CNN）中常用的一种神经网络层，通常，它会被设计在CNN网络的末端，如图2-8所示。在全连接层中，每个神经元都与前一层的所有神经元相连接，这被称为的全连接 (FC) 方法。在全连接层中的每个神经元都连接到前一层的所有神经元，这意味着它可以考虑前一层中所有特征的组合，使得全连接层能够对提取到的特征进行更高级别的抽象，从而更好地区分不同的类别。但正因此导致全连接层的参数数量非常大，增加模型的计算和存储开销。

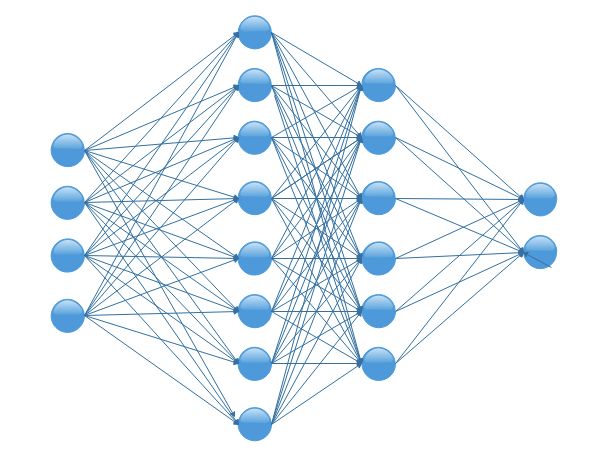


图2-8 全连接层示意图

(5)损失函数

在深度学习中，损失函数(Loss function)是用于衡量模型预测与真是值之间差异的函数。CNN 中的损失函数通常被用来指导模型训练，并且可以在训练期间被优化。目前被广泛使用的交叉熵损失函数，可表示为：

其中，N表示样本数，k表示总类别数，表示类别为m对应的真实标签，表示模型对类别m的预测概率。损失函数的损失越小，模型越好

2.2 深度学习框架

在深度学习领域，选择合适的框架至关重要。框架的选择不仅影响着使用者的体验，也直接影响着开发者的效率。因此，一个优秀的框架不仅对使用者友好，同时也对开发者十分友好。为了实现研究目标，研究人员们开发了各种各样的框架来满足不同的需求。目前，深度学习领域有许多框架可供选择，其中包括PyTorch、TensorFlow、Theano等。在本节中，我们将重点介绍两个重要的深度学习框架：PyTorch和TensorFlow。

2.2.1 Pytorch

PyTorch和TensorFlow、Caffe、CNTK等框架一样，用于构建深度神经网络，如多层感知机、卷积神经网络等。这些框架就像建造房子时的地基一样重要，如果地基不牢固，整个房子都会不稳定。因此，在深度学习领域，选择一个优秀的框架至关重要。在特定任务中选择适合的框架可以促进任务的顺利完成。PyTorch不仅支持CPU计算，还支持GPU计算，利用GPU可以实现高速运算。

PyTorch有许多优点，其中之一就是其设计追求最少的封装，这体现在其简洁性上。这种设计让使用者能够更轻松地阅读和理解代码。换句话说，相比于TensorFlow，PyTorch在实现相同功能时，代码量只有十分之一左右。PyTorch的动态计算图也非常简洁，当需要构建简单的深度神经网络时，只需几行代码就可以轻松实现复杂的神经网络结构。例如，要搭建一个多层感知机，只需调用torch.nn.Linear即可实现一层神经网络。如果需要构建图神经网络，可能需要使用卷积层和池化层，而相应的代码也非常简洁，可直接调用torch.nn.Conv2d和torch.nn.MaxPool。

PyTorch不像TensorFlow那样需要重新学习复杂的计算规则和语法结构，它更像是Python的扩展，因此人们只需要花费更少的时间就能很快适应。它具有简单的调试机制和一些通俗易懂的API，这对于初学者来说非常有帮助。此外，PyTorch还能对输入的数据进行并行化处理，这种并行化处理使得数据的执行效率比串行化更高。它允许将数据分成多个批次，然后将这些批次同时发送到多个图形处理单元进行并行运算。通过这样的操作，PyTorch逐渐将原本由CPU执行的大量任务转移到GPU，从而大大减轻了CPU的使用压力。

2.2.2 TensorFlow

在深度学习领域中，TensorFlow同样也是最常用的框架之一。它备受学生和研究者的欢迎并非没有原因，因为它是专门为方便人们使用而设计的框架。TensorFlow不仅是一种深度学习框架，还是一个端到端的学习平台。在这个平台上，人们可以创建用于机器学习或深度学习的库或软件。它也可以被理解为一个库，本质上是一个包含各种符号和定义的数学库。通过这个库中的各种符号和定义，人们可以构建和计算深度学习领域中各种复杂的任务。

TensorFlow是由谷歌的研究与开发人员构建的。在TensorFlow问世之前，深度学习凭借其出色的求解和计算性能在很长一段时间内遥遥领先于其他传统机器学习算法。当时，谷歌的研发人员也意识到了深度学习的优势，他们希望将这种优势应用于改进谷歌自身的软件功能，例如谷歌搜索引擎、谷歌地图和谷歌邮箱等服务。为此，他们开发了一种名为TensorFlow的框架。这个框架不仅适用于谷歌公司自身的服务和应用程序，也可以被其他开发人员使用。由于其开源性，TensorFlow得到了快速的发展，现在已经成为一个规模越来越庞大的项目。

TensorFlow的优势之一是其强大的社区功能。由于其早期出现以及代码的开源性，TensorFlow的社区不断壮大。其库中包含许多简单易用的接口供开发人员选择，如循环神经网络（RNN）、长短期记忆网络（LSTM）和残差网络等。然而，TensorFlow仍然存在一些缺点，主要体现在代码方面。由于其代码相对底层，用户需要编写大量代码，这可能导致代码的重复性，对于初学者来说并不友好。

2.3 植物叶片识别常用网络模型介绍

2.3.1 AlexNet模型

AlexNet[44]是CNN中的经典网络模型，它包含8个深层网络，其中包括5个卷积层和3个全连接层。在卷积层中，引入了ReLU（Rectified Linear Unit）激活函数，以加快模型的训练速度，并有效地解决了梯度消失的问题。此外，还利用局部响应归一化（LRN）技术，对卷积层输出进行归一化处理，以防止较小的神经元输出对模型的干扰，从而提高了模型的拟合效果。在全连接层之后，引入了Dropout技术，以避免网络过拟合的问题。总体而言，AlexNet在图像分类和识别领域具有显著的性能优势。AlexNet模型的网络示意图所示。表为 Alexnet 模型的网络参数

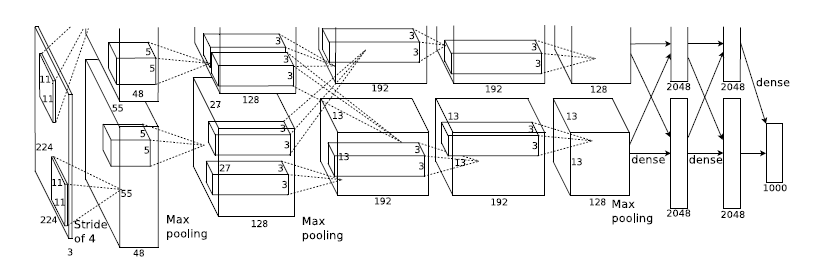


图2 AlexNet卷积神经网络架构示意图

表3 Alexnet 模型的网络参数



在AlexNet的第一个卷积层中，输入尺寸大小为224×224×3的RGB彩色图片，其中有3个通道。接着，使用96个尺寸为11×11×3的卷积核，设置步长为4，对输入的图像进行卷积操作。然后，将卷积操作的输出作为第二个卷积层的输入。在第二个卷积层中，采用256个尺寸为5×5×48的卷积核，步长设置为1，通过池化操作来降低参数量。接下来，第三个卷积层包含384个尺寸为3×3×256的卷积核，然后是384个尺寸为3×3×192的卷积核，第五层使用256个尺寸为3×3×192的卷积核，并进行池化降维操作。

在接下来的三个全连接层中，第六层、第七层和第八层分别包含4096个通道，将之前学习到的图像特征相融合。最后一层设置为1000输出，并通过softmax函数进行分类识别。这一结构能够有效地提取图像特征，并在最后的全连接层中实现分类。

2.3.2 VGG模型

VggNet[45]的提出起源于 2014年，这一网络模型主要通过加深网络结构以提升性能。VggNet的架构包括Vgg-11、Vgg-13、Vgg-16和Vgg-19，它们都以多个3x3的小型卷积核和2x2的max pooling为特征。其中，Vgg-16和Vgg-19因其简洁结构和出色分类性能，在深度学习领域广受欢迎。在训练过程中，VggNet要求输入图像的尺寸为224x224，并进行批量均值处理。它采用连续的3x3卷积核和13个卷积层来提取图像特征信息，最后利用3个全连接层和softmax分类器对图像进行分类，以最大程度地提取图像的细微特征。为保持图像分辨率，VggNet还使用1x1卷积滤波器对图像进行线性变换。

相较于AlexNet模型，Vgg-16在模型架构方面做出了重要改进，主要体现在卷积核的选取上。Vgg-16采用了多个3x3的轻量级卷积核，而不是像AlexNet那样使用较大的11x11、7x7、5x5的卷积核，这一改进有助于降低分类错误率。采用小卷积核的优势在于它能够增加网络的层次结构，使得网络可以更好地提取抽象的高阶特征。这样，代价函数的值就能够更快地收敛于最小值，有利于复杂网络学习模式，从而在一定程度上提升了模型性能和训练精度。举例来说，通过将三个步长为1的3x3小型卷积核叠加，可以得到与7x7卷积核相同的像素感受野大小。整个网络的参数量为3x(9xC2)，其中C为模型中输入层和输出层的通道数。

Vgg-16是一个深度网络结构，共包含16层，其中13层为卷积层，3层为全连接层。这一网络架构是在VggNet基础上延伸而来的。相对于Vgg-19，在分类任务中，Vgg-16具有更好的识别率。下图展示了Vgg-16的网络架构，而具体的模型网络参数可见于表4.2。

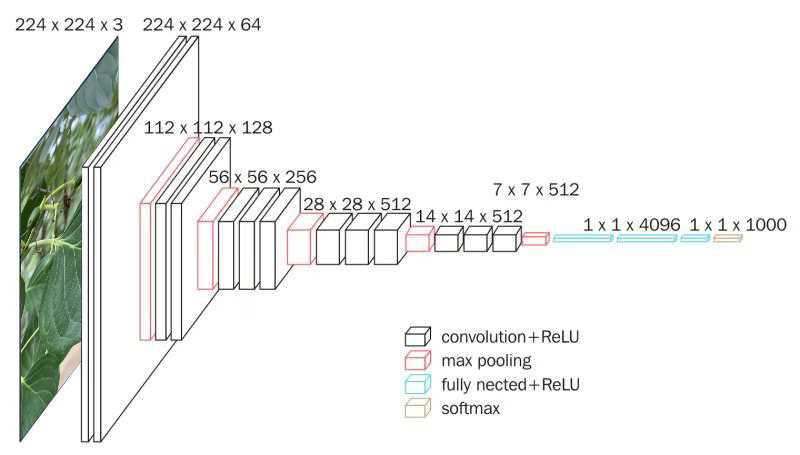
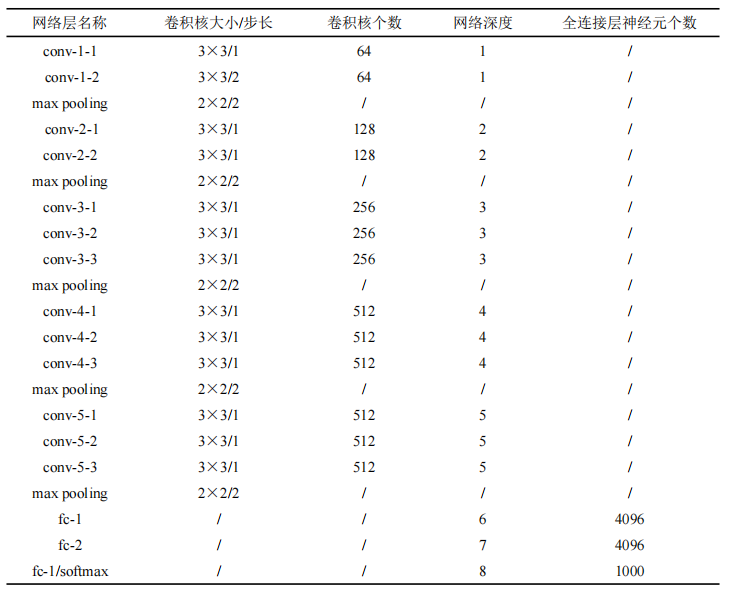


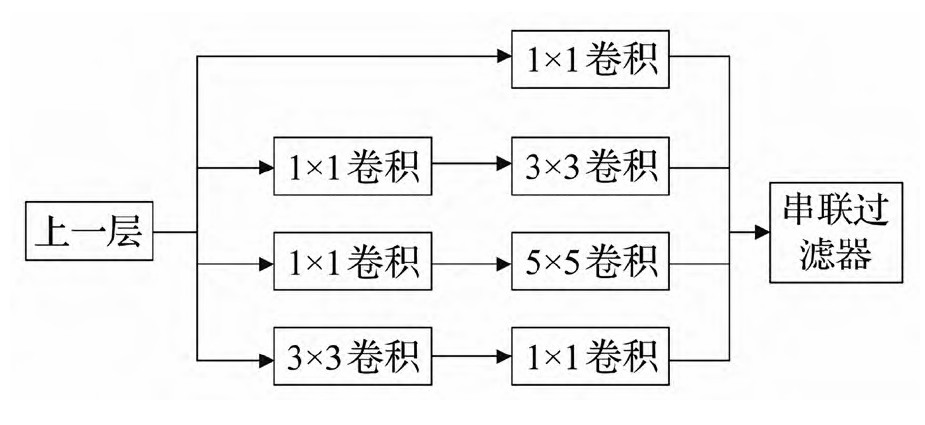
图2 VGG-16的网络架构图

表2 Vgg-16 模型网络参数



2.3.3 GoogLeNet模型

GoogLeNet[46]是一种新型的网络结构，于2014年提出，并在ImageNet比赛中夺得第一名。尽管GoogLeNet有22层的深度，但其参数量仅为6.8百万，远少于传统网络如AlexNet和VGG等。然而，从模型表现来看，GoogLeNet的性能却更胜一筹。这是因为GoogLeNet采用了一种全新的Inception结构，Inception模块结构采用了四个并行路径，其中前三个路径分别通过1x1、3x3和5x5的卷积核获取不同区域的感受野信息，而第四个路径则使用3x3的最大值池化核提取特征点数据信息。在卷积过程中，引入了1x1的卷积核，虽然增加了深度，但有助于防止过拟合。为了降低模型复杂性和数据维度，第二条和第三条路径中增加了1x1的卷积操作。尽管这加深了网络结构，但却能有效提高精确度。Inception模块的结构如图2所示。



因此，使用GoogLeNet模型提取植物叶片图像的特征有助于迅速识别植物叶片。在CNN结构中，全连接层在目标图像的分布式特征表示中发挥着关键作用，能够充分呈现图像的全局特征信息。下图是GoogleNet的网络架构示意图

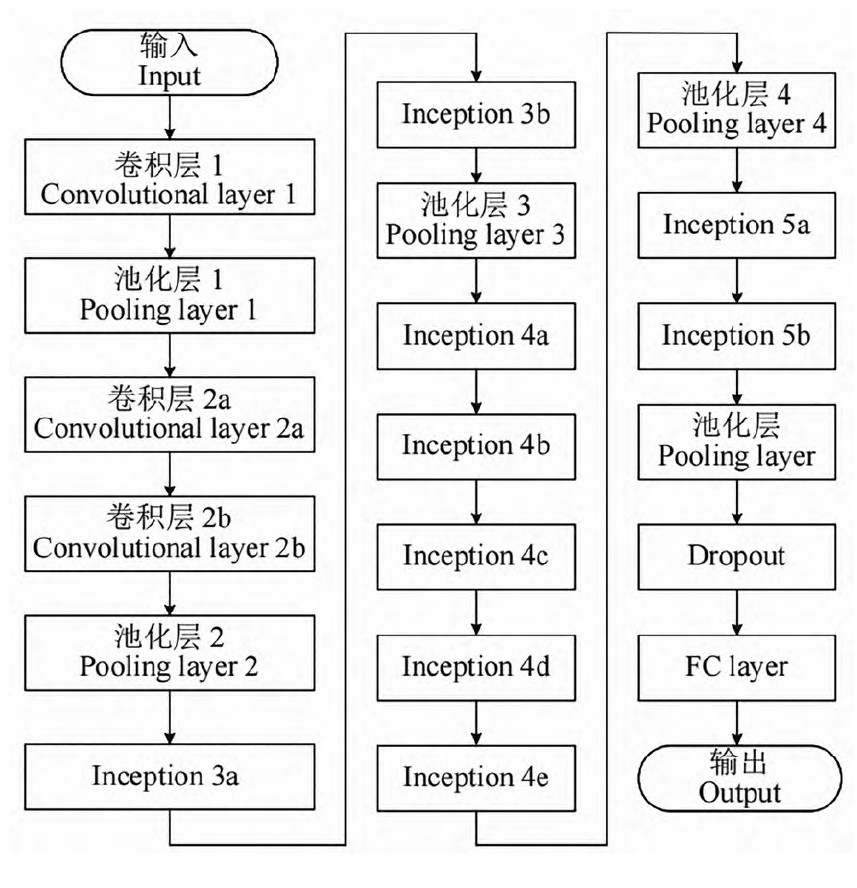


图2 经典GoogleNet网络结构

2.3.4 ResNet模型

2015年，何凯明等人提出了ResNet[47]网络，这一创新标志着卷积神经网络领域的又一次突破。该网络模型最多可达152层，具有相当深的层次结构，在当年在多个领域取得了显著的成绩，包括目标检测、语义分割和图像识别等方面，取得了具有历史意义的突破，其成果引人注目。与传统神经网络模型相比，ResNet不仅具有良好的学习效果，而且参数数量较低。ResNet的问世为研究者们提供了更深层次的训练方法。传统的深度学习方法通过反复堆叠卷积层来获取更好的特征，但随着层次深度的增加，模型的精度往往会出现先增加后下降的趋势。为了解决这一问题，ResNet团队引入了残差网络的概念。残差网络的核心思想是“恒等映射连接”，即每一层的输出保持不变，这样做的好处是减少了数据的冗余，更容易进行特征学习，并且残差映射比原始映射更容易优化。残差块的结构如图2-所示。

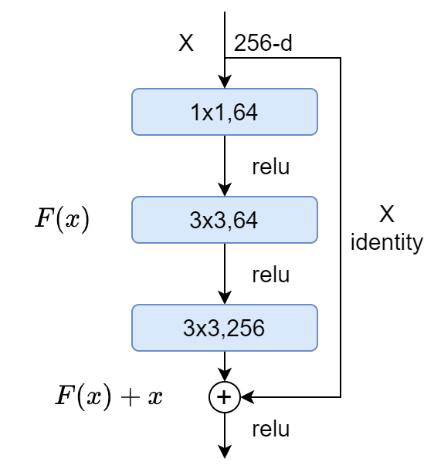


图2- 残差块结构

在残差块中，假设x为模型的输入。由于存在Identity Mapping，H(x)=F(x)+x，其中F(x)代表模型的真实学习部分，可表达为F(x)=H(x)-x。

残差块在某些情况下能够增强网络模型中的特征提取，而不会带来负面影响。因此，残差网络能够在减少模型参数的同时提升神经网络的性能，这正是近年来残差网络被广泛应用的关键原因之一。

ResNet可以根据其深度分为不同版本，包括ResNet18（18层）、ResNet34（34层）、ResNet50（50层）、ResNet101（101层）和ResNet152（152层）。每个版本的ResNet具有不同的结构参数，如图2-中所示。以ResNet18为例，其网络结构如图2-所示。。

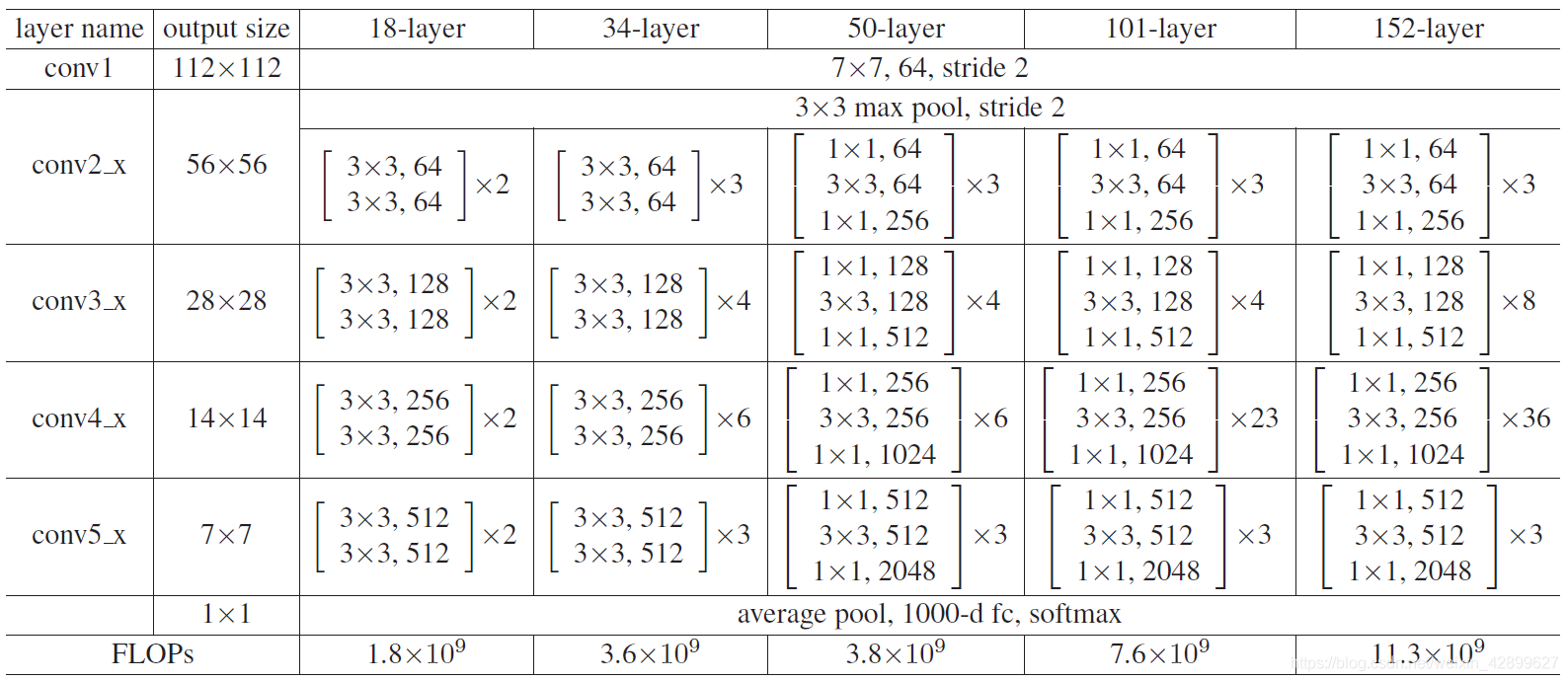


图2- 不同层数的ResNet网络参数

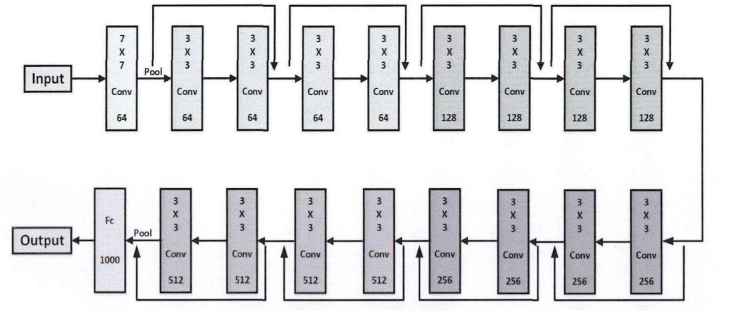


图2- ResNet-18结构图

2.4 总结

本章节首先介绍的是深度学习的基础知识，详细阐述了卷积神经网络的各个组成部分。然后介绍了深度学习的两种框架Pytorch和TensorFlow；最后，介绍了常用的植物叶片识别的网络模型，包括AlexNet，VGG，GoogLeNet和ResNet模型。

3 基于GoogLeNet网络模型的植物叶片识别

在深度学习领域，卷积神经网络（CNN）被广泛应用于植物叶片的识别任务。相比传统的基于特征识别方法，CNN模型能够自动学习和提取叶片的特征，从而实现更加精准的识别结果。为了训练CNN模型，需要大量的标注好的训练样本集作为网络的输入数据。因此，在进行训练之前，通常需要对植物叶片图像进行图像预处理，以减少噪声、增强特征等，然后将处理过的图像输入到相应的CNN模型中进行训练。在训练过程中，通过不断地调整模型的参数，可以逐步提高植物叶片的识别率[1]。

本章的主要研究目标是以GoogleNet网络模型为基础，将其应用到植物叶片识别的研究中，探索其在识别植物叶片方面的性能和适用性。

**3.1 数据集及预处理**

3.1.1 数据集描述

Flavia数据库[43]是一个被广泛应用于植物叶片分类和识别领域的公共数据库。该数据库包含了来自32个不同植物物种的1907个样本。每个样本都是彩色图像，其大小均为1600x1200像素，分辨率为300-DPI。在我们的研究中，为了保证样本的多样性和代表性，我们按照80%的比例将这些样本划分为训练集，20%划分为验证集，同时在对实验进行测试时选择20%作为测试集。这样的数据划分方式旨在充分利用数据库中的样本信息，同时保证了训练、验证和测试集之间的独立性和公平性。

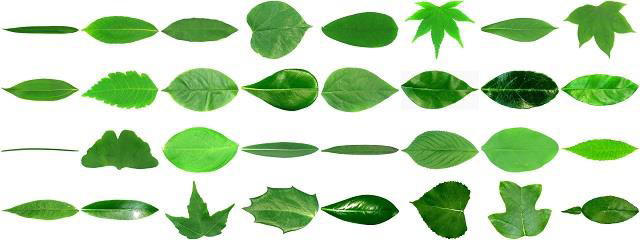


图3- Flavia数据库示例

3.1.2 数据预处理和增强

在计算机视觉领域，图像预处理是为了在将图像输入到CNN或其他深度网络模型中进行学习和识别之前对其进行处理。由于图像中的纹理特征受图像大小和方向的影响较大，因此在提取纹理特征之前，需要对图像进行适当的预处理。

在本次植物叶片分类识别实验中，数据预处理过程对于提升模型的训练效果至关重要。本文采用了一系列图像增强技术来丰富训练数据的多样性，并使用标准化方法来保证模型的输入一致性。首先，调整所有输入图像的大小至224x224像素，以适应模型的输入要求。统一的图像尺寸不仅减少了计算复杂度，还能确保模型对每个输入图像的处理一致性。接下来，为了增加训练数据的多样性，本文采用了随机垂直翻转和随机水平翻转两种图像增强技术。这种随机翻转操作可以有效防止模型对图像方向的过度依赖，从而提升模型的泛化能力。此外，对图像进行随机旋转，最大旋转角度为180度。这一处理进一步增加了数据集的多样性，使模型能够更好地适应不同角度下的叶片图像。最后，对植物叶片图像数据进行归一化处理。归一化操作使用预先计算的均值和标准差，使得每个通道的图像数据分布更加均匀。归一化不仅可以加速模型收敛，还能避免因输入数据差异过大导致的训练不稳定问题。

综上所述，这些预处理步骤通过多样化和标准化输入数据，提高了模型的训练效果和鲁棒性，为后续的植物叶片分类识别任务奠定了坚实的基础。

3.2模型设计

GoogLeNet模型在训练前，叶片图像输入大小设置为299x299像素。配合GoogLeNet的特征处理器，提取图像特征值，并结合softmax 分类器对提取到的特征信息进行分类。训练过程中采用正则化中的 dropout(随机失活)设置该层节点的保留概率来防止模型过拟合，最后由输入层将训练结果展示。不断调整模型参数，并对图像数据及进行整理和改进，通过测试对比，引入Checkpoint机制保存最优模型。

3.3 实验过程

3.4.1 实验环境

在本研究中，我们选择了Python作为主要开发语言，因其具备跨平台兼容性和丰富的第三方库支持。我们采用了Anaconda3作为编译器，并结合PyCharm作为主要开发工具，以便于项目环境配置和代码调试。为了进行图像预处理，我们引入了OpenCV计算机视觉库，用于实现图像的灰度化、降噪和分割等处理方法。在深度学习方面，我们选择了PyTorch作为主要框架，以便于构建深层神经网络模型。在实验结果的展示方面，我们利用可视化工具matplotlib，根据训练数据的权重文件，绘制了准确率和损失函数曲线，以直观地呈现模型训练的效果；同时为了更好地观察实验结果，将混淆矩阵可视化。这样的工具组合和方法选择，旨在提高开发效率和实验结果的可解释性，为研究成果的展示提供了良好的支持。

表3 实验环境

|  |  |
| --- | --- |
| 硬件环境 | 软件环境 |
| CPU：AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz  GPU：NVIDIA GeForce RTX 3050 Laptop GPu  内存：16G | 操作系统：Windows11 64位  开发工具：Pycharm 2023  编译器：Anaconda3, Python3.10 |

3.4.2 参数设置

在模型训练过程中，网络使用随机梯度下降（SGD）优化算法，并且使用L2范数作为正则化系数，L2范数的权重衰减系数初始值为0.01, 可以帮助减少模型过拟合。学习率的初始值为0.02，使用指数衰减调整学习率 ExponentialLR来衰减学习率。将学习率在每个周期内按照以下调整公式进行衰减。

上述表达式中e是epoch计数器，即当前训练所处的epoch，gamma参数来控制学习率的衰减速度，设定为0.95。批处理量Batch\_size设定为32，共训练100个epoch。训练一个epoch需要花费1～3min。

4.5.3 算法评价指标

在我们的研究中，我们使用了多个评价指标来全面评估我们提出的植物叶片识别算法的性能。在数据集中，如果各个类别的分布相对平衡，通常会选择准确率（Accuracy）作为评估模型性能的指标。在我们的实验中，准确率反映了我们提出的算法在识别植物叶片时的整体分类准确性。假设总样本数为S，而正确分类的样本数为m，则准确率可表示为：

当数据集中各类别样本数量不平衡时，即各类别样本数量存在显著差异，模型在训练过程中会倾向于学习样本数量较多的类别特征。这会导致模型在识别样本数量多的类别时表现较好，而在识别样本数量少的类别时表现较差。在这种情况下，准确率并不是评估模型分类性能的最佳指标。

因此，为了在样本分布不均衡的情况下更准确地评估模型的分类性能，通常采用精确率（Precision）、召回率（Recall）和F1值（F1 Score）等指标。这些评价指标通常通过混淆矩阵计算得到，假设一个二分类的混淆矩阵如表3-所示。

表3- 二分类混淆矩阵

|  |  |  |
| --- | --- | --- |
| 混淆矩阵 | 预测为正 | 预测为负 |
| 实际为正 | TP | FN |
| 实际为负 | FP | TN |

其中TP表示真正例，即实际为正样本且被模型正确预测为正样本的样本数量；FN表示假负例，即实际为正样本但被模型错误预测为负样本的样本数量；FP表示假正例，即实际为负样本但被模型错误预测为正样本的样本数量；TN表示真负例，即实际为负样本且被模型正确预测为负样本的样本数量。

根据混淆矩阵和准确率的定义，准确率还可以表示为下式（3-3），它表示预测正确的样本数量在总样本数量中所占的比例。

精确率（Precision）可用以下公式（3-4）表示，它指的是真实值为正且预测值也为正的样本数量在所有预测值为正的样本中的比例。在植物叶片识别的任务中，精确率衡量了模型在预测为某一类别时的准确性，高精确率意味着模型的预测结果中真正为正类的比例较高。

召回率（Recall）可由公式（3-4）表示，表示真实值为正且预测值也为正的样本数量与真实值为正的样本数量之比。在植物叶片识别的任务中，召回率衡量了模型对正样本的识别能力。高召回率表示模型能够较好地捕获正样本。

根据公式（3-4）和（3-5），可以得知，精确度和召回率这两个评估标准是相互对立的。当模型的精确度提高时，召回率就会降低；而当精确度下降时，召回率会增加。因此，需要一个合适的指标来综合精确度和召回率，这个评估指标就是F1值。

F1-score值可由方程式（3-6）表示，它代表着精确率和召回率的调和平均，综合考虑了这两个指标。F1值越高，F1-score的取值范围在0到1之间，越接近1表示模型的性能越好。

混淆矩阵也可叫做误差矩阵，以有n行n列的矩阵形式显示准确度的评价。这种精确度指标体现了不同类型的图像分类精度。在图像的准确性评价中，重点研究了分类结果与实际情况的对比，从而可以得出分类结果的准确性。混淆矩阵通过对图像中的每个特征的位置和顺序进行对比，对分类图像中的对应位置和顺序进行对比。

在评估我们的算法时，我们不仅关注了单一指标的表现，还综合考虑了多个指标，以确保对模型性能的全面评估。通过分析这些评价指标，我们能够更好地理解我们提出的算法在植物叶片识别任务中的优劣势，并且为进一步改进和优化算法提供了指导。

3.4 实验结果与分析

在上述实验过程中，将植物叶片数据集输入到GoogLeNet模型中训练，对模型参数进行设置和调整，经过训练后得到不同的识别率,并保存识别精度高的模型。利用Python语言引入了 OpenCV计算机视觉中的 matplotlib 库绘制模型的准确率及损失函数变化曲线图，同时为了更好地观测测试集识别的效果，将混淆矩阵进行可视化。

3.4.1 基于GoogLeNet网络模型的训练结果

根据模型最初设置的参数,得到训练结果。如图3-为GoogLeNet模型的训练过程，可以看出实验的测试准确率约为92.88%，精确率Precision约为92.92%，召回率Recall约为92.61%，F1-score约为92.63%；

图3-为GoogLeNet模型的准确率与 Loss曲线，从图中可以看出，GoogLeNet的准确率随着迭代次数的增加在第40轮有了明显提升且呈缓慢上升状态，损失值迭代到 40 轮，Loss 曲线有了较为明显的下降趋势。由于验证集图像样本较少，从图中能够看到验证集的准确率波动较大。

图3-为GoogLeNet模型是对于测试集32种植物叶片进行测试的混淆矩阵可视化，在混淆矩阵中，横轴上的标签标示样本的预测label，纵轴上的标签标示样本的实际label，对角线上的数字表示预测label和真是label一致的数量，也就是预测正确的数量。对于其他位置的数字就表示预测错误的；从图中可以看到植物叶片Beale’s barberry和Peach有两次被错误的预测为pubescent bamboo，Ford Woodlotus有两次错误地预测为Japanese Flowering Cherry, 整体上来看对于测试集的预测正确率较高

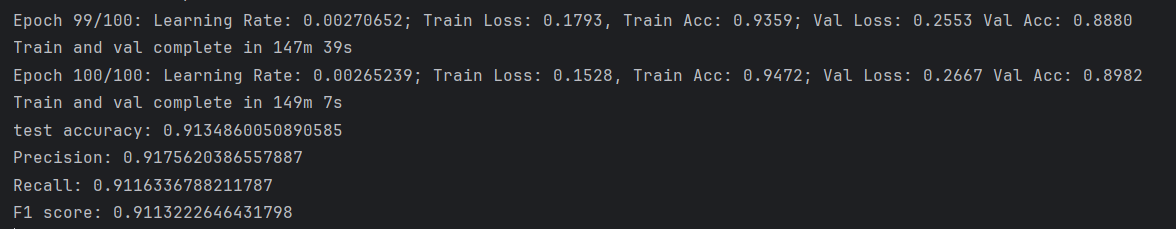




图3- GoogLeNet模型的实验结果

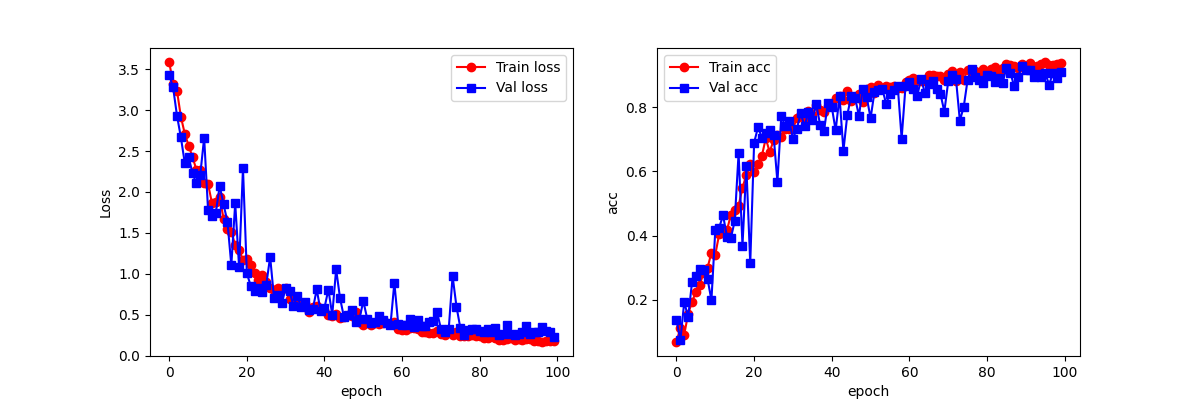


图3- GoogLeNet模型的准确率与Loss变化曲线

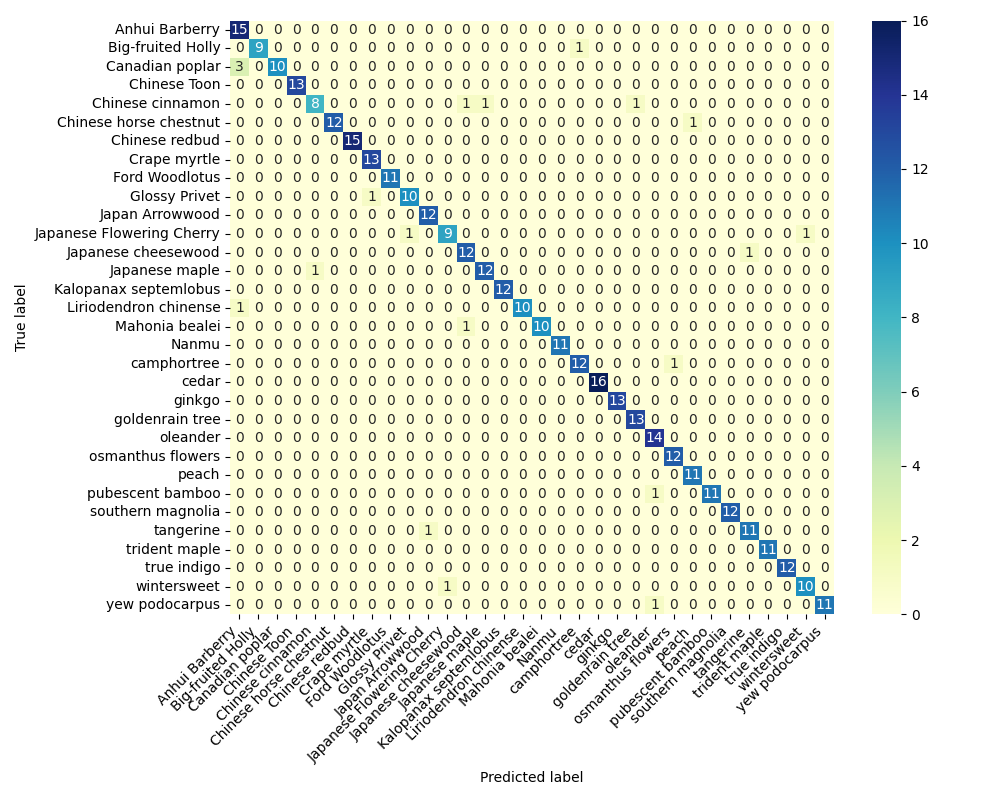


图3- GoogLeNet模型的混淆矩阵可视化

表3- 是GoogLeNet模型下的 Flavia 数据集中各类植物叶片识别准确率。

表3-1 GoogLeNet模型下各类植物叶片识别准确率

|  |  |  |  |
| --- | --- | --- | --- |
| 植物种类 | 识别率 | 植物种类 | 识别率 |
| 毛竹 | 91.67% | 雪松 | 100.00% |
| 中国七叶树 | 92.31% | 银杏 | 100.00% |
| 安徽小檗 | 100.00% | 紫薇 | 100.00% |
| 中国紫荆花 | 100.00% | 夹竹桃 | 100.00% |
| 蓝靛果 | 100.00% | 紫杉梅松 | 90.91% |
| 日本枫树 | 92.31% | 日本樱花 | 81.82% |
| 楠木 | 100.00% | 女贞 | 90.91% |
| 龙牙 | 100.00% | 中国香椿 | 100.00% |
| 中国肉桂 | 72.73% | 桃树 | 100.00% |
| 金雨树 | 100.00% | 木莲 | 100.00% |
| 大果冬青 | 80% | 三叉戟枫 | 100.00% |
| 日本干酪木 | 92.31% | 阔叶十大功劳 | 90.91% |
| 梅花 | 90.91% | 南方玉兰 | 100.00% |
| 樟树 | 92.31% | 加拿大杨树 | 76.92% |
| 日本箭树 | 100.00% | 鹅掌楸 | 90.91% |
| 桂花 | 100.00% | 柑橘 | 91.67% |

从表 3-1 中可看出，对于大部分物种GoogLeNet都可以得到很好的识别效果，但同样

也会出现误分类情况，比如日本樱花，加拿大杨树，中国肉桂，大果冬青

**3.5 总结**

4 基于ResNet网络模型的植物叶片识别

**4.1 数据集及预处理**

4.1.1 数据集描述

Flavia数据库是一个被广泛应用于植物叶片分类和识别领域的公共数据库。该数据库包含了来自32个不同植物物种的1907个样本。每个样本都是彩色图像，其大小均为1600x1200像素，分辨率为300-DPI。在我们的研究中，为了保证样本的多样性和代表性，我们按照80%的比例将这些样本划分为训练集，20%划分为验证集，同时在对实验进行验证时选择20%作为测试集。这样的数据划分方式旨在充分利用数据库中的样本信息，同时保证了训练、验证和测试集之间的独立性和公平性。

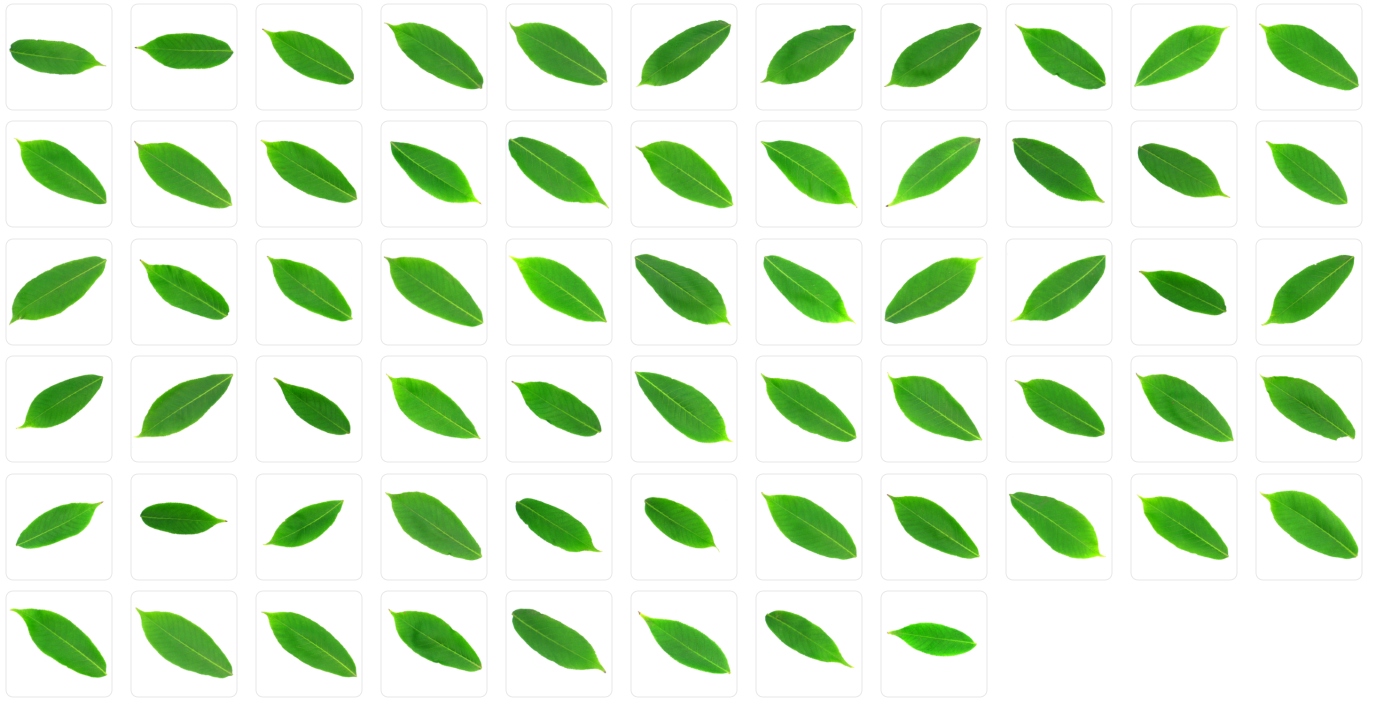


图3-1 flavia数据库叶片图像样例

4.1.2 数据预处理和增强

在计算机视觉领域，图像预处理是为了在将图像输入到CNN或其他深度网络模型中进行学习和识别之前对其进行处理。由于图像中的纹理特征受图像大小和方向的影响较大，因此在提取纹理特征之前，需要对图像进行适当的预处理。

在计算机视觉领域，图像预处理是为了在将图像输入到CNN或其他深度网络模型中进行学习和识别之前对其进行处理。由于图像中的纹理特征受图像大小和方向的影响较大，因此在提取纹理特征之前，需要对图像进行适当的预处理。

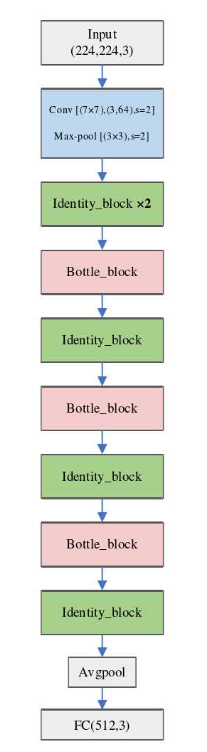
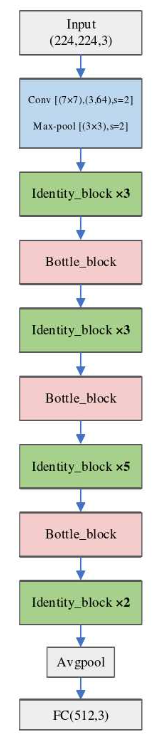
在本次植物叶片分类识别实验中，数据预处理过程对于提升模型的训练效果至关重要。本文采用了一系列图像增强技术来丰富训练数据的多样性，并使用标准化方法来保证模型的输入一致性。首先，调整所有输入图像的大小至224x224像素，以适应模型的输入要求。统一的图像尺寸不仅减少了计算复杂度，还能确保模型对每个输入图像的处理一致性。接下来，为了增加训练数据的多样性，本文采用了随机垂直翻转和随机水平翻转两种图像增强技术。这种随机翻转操作可以有效防止模型对图像方向的过度依赖，从而提升模型的泛化能力。此外，对图像进行随机旋转，最大旋转角度为180度。这一处理进一步增加了数据集的多样性，使模型能够更好地适应不同角度下的叶片图像。最后，对图像数据进行了归一化处理。归一化操作使用预先计算的均值和标准差，使得每个通道的图像数据分布更加均匀。归一化不仅可以加速模型收敛，还能避免因输入数据差异过大导致的训练不稳定问题。

综上所述，这些预处理步骤通过多样化和标准化输入数据，提高了模型的训练效果和鲁棒性，为后续的植物叶片分类识别任务奠定了坚实的基础。

4.2模型设计

4.2.1 ResNet18模型结构

Resnet18的网络结构如图4- a)所示，总共包含18个层（仅考虑卷积层和全连接层），其中包括17个卷积层和1个全连接层。输入数据首先通过一个卷积核尺寸为7x7、步长为2的卷积层，然后经过一个最大池化层，这导致特征图的尺寸缩小为原始的1/4，即从224变为56，并且特征图的通道数从3增加到64。接下来，特征图通过两个Identityblock残差块，通道数和特征图的尺寸保持不变。然后，特征图通过一个Bottleblock残差块，导致特征图的尺寸缩小为原始的1/2，通道数增加为原来的2倍，然后再经过一个Identityblock残差块，通道数和特征图的尺寸保持不变。交替经过三个Bottleblock和Identityblock后，特征图的尺寸缩小为原始的1/8，从56变为7，通道数增加为原来的8倍，从64变为512。最后，特征图通过一个全局平均池化层被展平成一维数据，然后输入到一个包含512个神经元的全连接层，在全连接层后面添加一个三分类的分类器来完成植物叶片分类任务。

a) ResNet18模型结构 b)ResNet34模型结构

图4- ResNet18和ResNet34模型结构图

4.2.2 对ResNet18模型进行改进

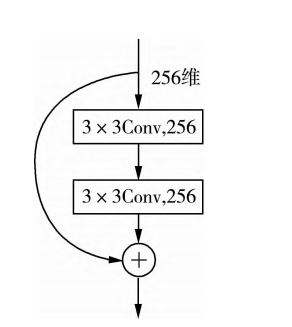
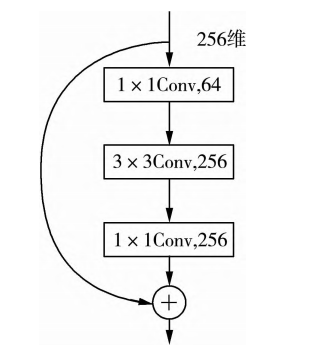
为了精确识别植物叶片图像类别，ResNet18模型表现出了明显的优势。然而，在实验中发现，ResNet18仍然存在一些无法避免的问题：(1)底层卷积层参数过多；(2)模型训练速度慢。为了解决这些问题，我们对ResNet18模型进行了调整，包括调整底层卷积尺寸和改进模型残差块结构。

1. 卷积核尺寸调整

在深度卷积神经网络中，虽然大尺寸卷积核能够捕获图像中更为复杂的特征，但其参数量较大，不适合用于植物叶片分类识别研究。相反，小尺寸卷积核参数量较小，但可能无法充分表达图像信息。为了在尽量减少模型精度损失的前提下降低参数量，我们选择将原有的7x7卷积替换为5x5卷积，以平衡模型参数和特征表达。

1. 残差块结构优化

为了减少ResNet18网络模型的参数量，我们对残差块的结构进行了优化。我们采用了1x1卷积、3x3卷积和1x1卷积串联的瓶颈结构，来替代原有的两个3x3卷积串联的结构。改进前和改进后的残差块结构如图4所示。

（a）改进前的结构 （b）改进后的结构

图4- 残差快结构

以输入256维数据为例，使用改进前的残差块结构进行特征提取需要1179648个参数；而使用改进后的残差结构进行特征提取仅需要69632个参数。通过数据对比可知，本文改进的残差块结构显著降低了参数量，降幅达到了94%。改进后的残差块结构在计算量上有明显减少，有助于提高模型训练速度、减小模型大小，并且提升了泛化性能。

经由上述卷积核尺寸调整、残差块优化后，所设计改进后的植物叶片识别模型如图4-所示。

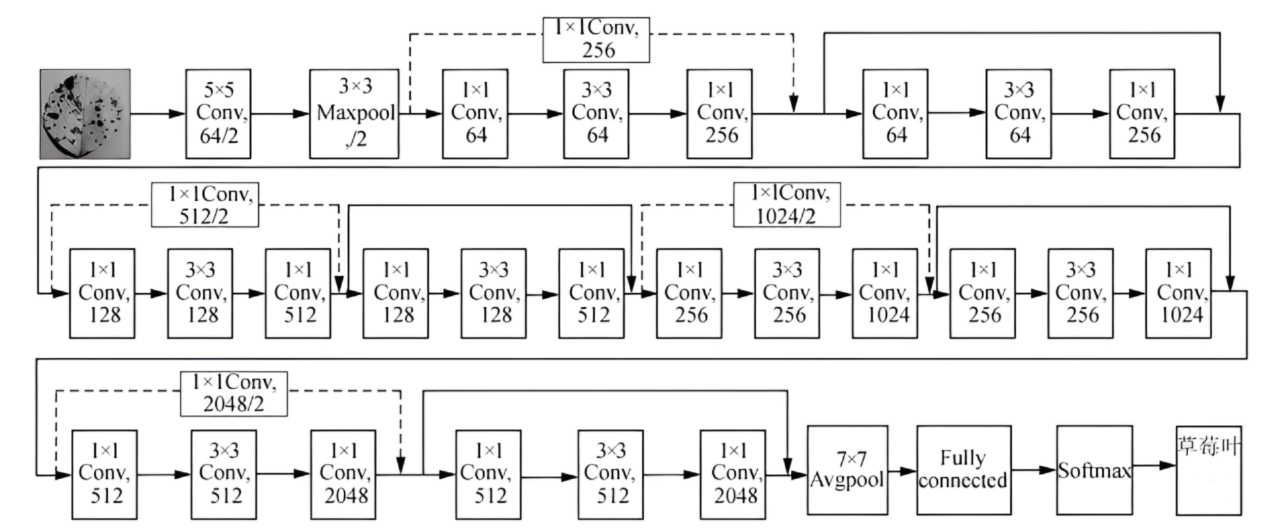


图4- 改进后的ResNet18模型结构示意图

4.2.3 ResNet34模型结构

Resnet34的网络结构如图4-2 b)所示，与ResNet18相似。输入数据首先通过一个7x7的卷积核，步长为2的卷积层，然后经过一个3x3大小的最大池化层。这使得特征图的尺寸从224缩小到56，通道数从3增加到64。不同于Resnet18，Resnet34在相应位置增加了Identityblock的数量，以增加网络深度，而具有下采样功能的Bottleblock数量保持不变。在经过一个Identity block残差块后，通道数和特征图的尺寸保持不变；在经过一个Bottleblock残差块后，特征图的尺寸缩小为原来的1/2，通道数增加为原来的2倍。因此，经过3个Identity block、1个Bottle block、3个Identity block、1个Bottle block、5个Bottle block、1个Bottle block、2个Identity block的顺序后，特征图的尺寸从56缩小到7，通道数增加为原来的8倍，从64增加到512。接着，特征图通过一个全局平均池化层被展平成一维数据，然后输入到一个包含512个神经元的全连接层，最后在全连接层后添加一个三分类的分类器，完成植物叶片识别分类任务。

4.3 实验过程

4.4.1 实验环境

在本研究中，我们选择了Python作为主要开发语言，并结合PyCharm作为主要开发工具，选择PyTorch作为主要框架。

表3 实验环境

|  |  |
| --- | --- |
| 硬件环境 | 软件环境 |
| CPU：AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz  GPU：NVIDIA GeForce RTX 3050 Laptop GPu  内存：16G | 操作系统：Windows11 64位  开发工具：Pycharm 2023  编译器：Anaconda3, Python3.10 |

4.4.2 参数设置

在模型训练过程中，网络使用随机梯度下降（SGD）优化算法，并且使用L2范数作为正则化系数，L2范数的权重衰减系数初始值为0.01, 可以帮助减少模型过拟合。学习率的初始值为0.02，使用指数衰减调整学习率 ExponentialLR来衰减学习率。将学习率在每个周期内按照以下调整公式进行衰减。

上述表达式中e是epoch计数器，即当前训练所处的epoch，gamma参数来控制学习率的衰减速度，设定为0.95。批处理量Batch\_size设定为32，共训练100个epoch。训练一个epoch需要花费1～3min。

4.4.3 算法评价指标

在我们的研究中，我们使用准确率（Accuracy）和混淆矩阵来评估我们提出的植物叶片识别算法的性能。下面将对每个指标进行详细说明：

（1）准确率（Accuracy）：准确率是评价分类模型整体性能的指标，它表示模型正确分类的样本数量与总样本数量之比。在我们的实验中，准确率反映了我们提出的算法在识别植物叶片时的整体分类准确性。公式表示为：

这里，TP表示的含义是真实值为positive且模型判断正确数量；TN表示的含义是真实值为negative 且模型判断正确数量；FP的含义是真实值为negative而模型判断错误的数量；FN的含义是真实值为positive而模型判断错误的数量。

（2）混淆矩阵:混淆矩阵也可叫做误差矩阵，以有n行n列的矩阵形式显示准确度的评价。这种精确度指标体现了不同类型的图像分类精度。在图像的准确性评价中，重点研究了分类结果与实际情况的对比，从而可以得出分类结果的准确性。混淆矩阵通过对图像中的每个特征的位置和顺序进行对比，对分类图像中的对应位置和顺序进行对比。

4.4 实验结果与分析

在上述实验过程中，将植物叶片数据集输入到ResNet18以及ResNet34模型中训练，对模型参数进行设置和调整，经过训练后得到不同的识别率,并保存识别精度高的模型。利用Python语言引入了 OpenCV计算机视觉中的 matplotlib 库绘制模型的准确率及损失函数变化曲线图，同时为了更好地观测测试集识别的效果，将混淆矩阵进行可视化。

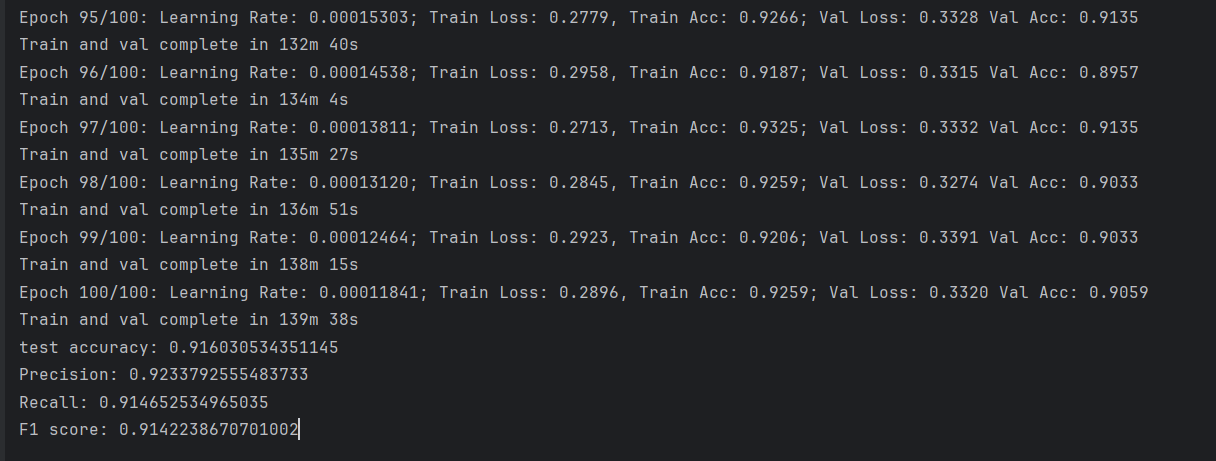


图4- ResNet18模型的实验结果

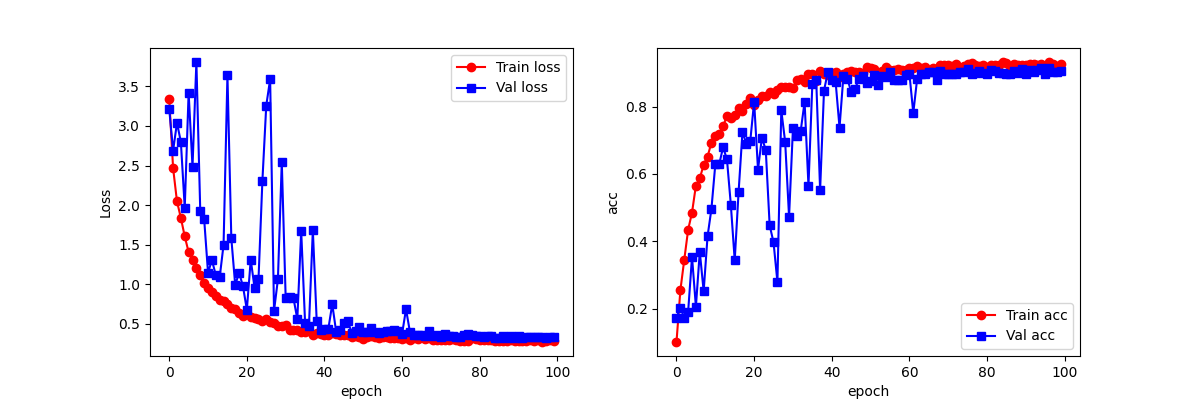


图4- ResNet18模型的准确率与Loss变化曲线

4.4.1 ResNet18改进前后的实验对比

为确定ResNet18模型改进中所采用的卷积核尺寸调整、残差块优化对ResNet18模型植物叶片识别能力的具体提升效果，设计测试试验如表1所示，其中方案5为本文 Simplify-ResNet 模型:

表4- 测试设计方案

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 底层卷积核 | | 残差快 | |
| 7x7卷积核 | 5x5卷积核 | 捷径结构 | 瓶颈结构 |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |

表4- 各方案实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 准确率 | 精确率 | 召回率 | F1分数 |
| 1  2 | 88.21%  90.19% | 88.34%  90.41% | 88.94%  89.62% | 88.83%  89.35% |
| 3  4 | 91.14%  91.60% | 91.22%  92.34% | 90.54%  91.47% | 90.36%  91.42% |

由根据表4- 的测试结果，方案4改进后的ResNet18模型在植物叶片识别方面表现最佳，准确率达到了91.60%，同时模型的内存占用仅为35.34 Mb，有效地实现了在保持模型准确率的同时进行轻量化设计的目标。

相比方案2，在采用瓶颈结构替代捷径结构后，模型的准确率提升了6.32%，但内存大小增加了5.2 Mb。而在方案3中，通过使用5x5卷积核替代原有的7x7卷积核，模型的准确率略微下降了1.63%，但内存大小减小了23 Mb。

综合而言，方案4采用了5x5卷积核替代原有的7x7卷积核，并采用了瓶颈结构替代捷径结构，使模型在准确率和内存大小之间达到了良好的平衡。

4.4.2基于ResNet网络模型的训练结果

在本实验中，还使用了ResNet18和ResNet34两个网络模型进行对比实验，实验中的实验环境、使用到的训练集、测试集、验证集和参数设置均保持相同，以此来确保实验结果的可靠性和可比性。

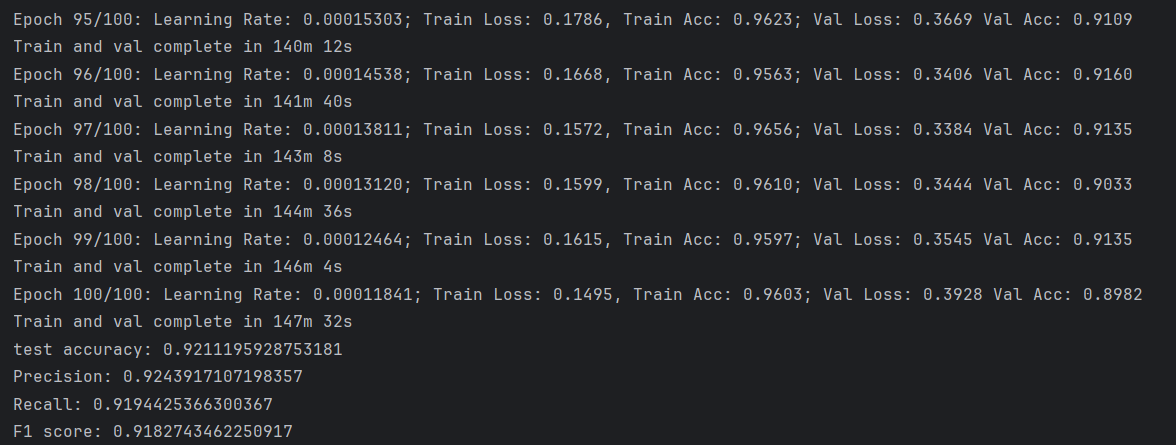


图4- ResNet34模型的实验结果

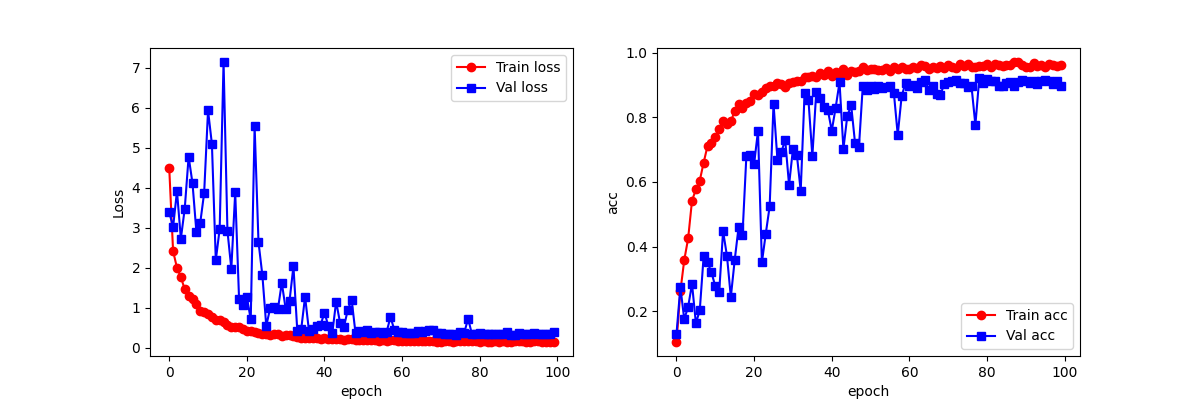


图4- ResNet34模型的准确率与Loss变化曲线

表4- ResNet18和ResNet34模型实验结果对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 准确率 | 精确率 | 召回率 | F1分数 |
| ResNet18 | 91.60% | 92.34% | 91.47% | 91.42% |
| ResNet34 | 92.11% | 92.44% | 91.94% | 91.83% |

通过实验可以看出，ResNet34模型相比于ResNet18模型对植物叶片识别准确率平均提升了约0.5个百分比

4.4.2 不同实验参数下的ResNet34模型的实验分析

实验中，采用批量训练的方法，合理的批次可以确保梯度向量上模型的准确性，保证以最快的速度下降到局部最小值。初始时每批次训练24张图片,根据训练过程中loss值的变化进行调整，调整每批次训练到32，48。设置初始学习率为 0.02，训练过程中做微调优化

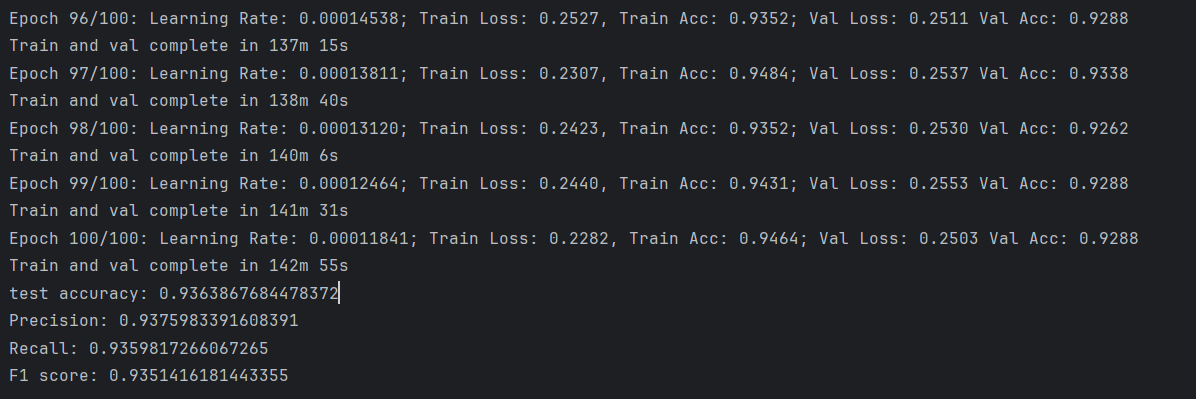


图4- ResNet34模型(batch\_size=24)的实验结果

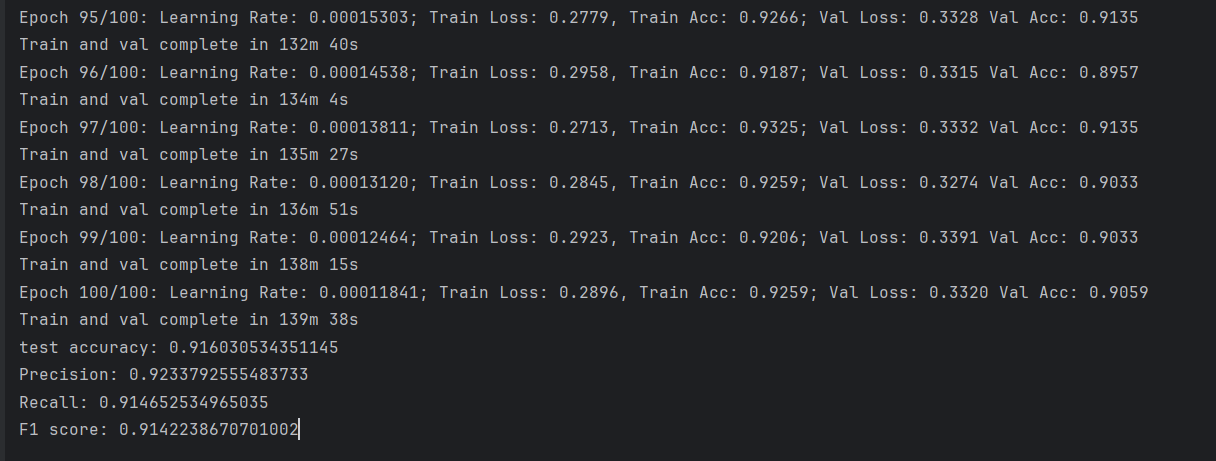


图4- ResNet34模型(batch\_size=32)的实验结果

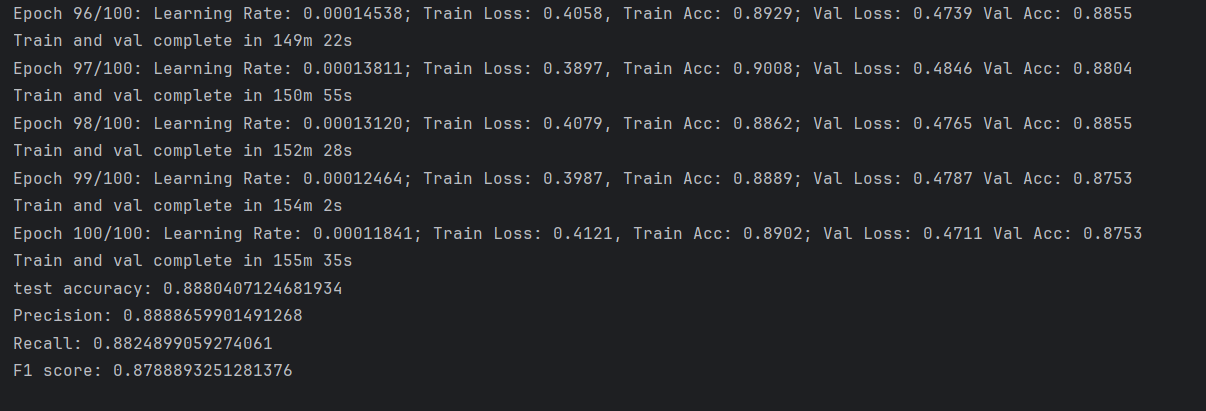


图4- ResNet34模型(batch\_size=32)的实验结果

表4- ResNet34模型不同Batch\_Size实验结果对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Batch\_Size | 准确率 | 精确率 | 召回率 | F1分数 |
| 24 | 93.64% | 93.76% | 93.60% | 93.51% |
| 32 | 91.60% | 92.34% | 91.47% | 91.42% |
| 48 | 88.80% | 88.89% | 88.25% | 87.89% |

每批次训练数量(batch\_size)对模型性能的影响。batch\_size 对模型训练结果有一定的影响，从实验结果可以看出，batch\_size调整的越大，准确率和精确率都会有一定程度的下降。合理的batch\_size能得到较高的准确率,通过实验可以看出训练模型在 batch\_size为 24 时训练准确率最高，为93.64%。

参 考 文 献

1. 傅弘, 池哲儒, 常杰. 基于人工神经网络的叶脉信息提取——植物活体机器识别研究Ⅰ[J]. 植物学通报, 2004, (04): 429-436.
2. 杜吉样, 植物物种机器识别技术的研究[D]. 安徽: 中国科学技术大学，2005.
3. 王晓峰, 黄德双, 杜吉祥, 张国军. 叶片图像特征提取与识别技术的研究[J].计算机工程与应用, 2006, (03): 190-193.
4. 朱宁. 基于 LBP 的树叶识别系统研究与实现[D]. 北京林业大学, 2008.
5. 贺鹏. 基于叶片综合特征的阔叶树机器识别研究[D]. 西北农林科技大学, 2008.
6. 阚江明, 王怡萱, 杨晓微, 冷萃. 基于叶片图像的植物识别方法[J]. 科技导报, 2010, 28(23):81-85.
7. 张宁. 基于图像分析的植物叶片识别算法研究[D]. 北京林业大学, 2013.
8. 王丽君. 基于叶片图像多特征提取的观叶植物种类识别[D]. 北京林业大学, 2014.
9. 杨天天, 潘晓星, 穆立蔷. 基于叶片图像特征数字化信息识别 7 种柳属植物[J]. 东北林业大学学报, 2014, 42(12): 75-79.
10. 于慧伶, 麻峻玮, 张怡卓.基于双路卷积神经网络的植物叶片识别模型[J]. 北京林业大学学报, 2018, 40(12): 132-137.
11. 朱良宽, 晏铭, 黄建平. 一种新型卷积神经网络植物叶片识别方法[J]. 东北林业大学学报, 2020, 48(04): 50-53.
12. 李龙龙, 何东健, 王美丽. 基于改进型 LBP 算法的植物叶片图像识别研究[J/OL]. 计算机工程与应用: 1-10[2021-01-24].
13. 张露. 基于深度学习的植物叶片图像识别方法研究[D]. 北京林业大学, 2019
14. 朱静, 田兴军, 陈彬, 吕劲紫. 植物叶形的计算机识别系统[J]. 植物学通报, 2005(05): 89-94.
15. 龚丁禧, 曹长荣. 基于卷积神经网络的植物叶片分类[J]. 计算机与现代化, 2014(04): 12-15.
16. 丰晓霞. 基于深度学习的图像识别算法研究[D]. 太原理工大学, 2015.
17. 张帅. 基于深度学习的植物叶片识别算法研究[D]. 北京林业大学, 2016.
18. 薄琪苇. 基于卷积神经网络的植物叶片识别研究[D]. 浙江农林大学, 2018.
19. 程曦, 吴云志, 张友华. 基于深度卷积神经网络的储粮害虫图像识别[J]. 中国农学通报, 2018, 34(01): 154-158.
20. 黄志国. 基于深度学习的中草药叶片识别算法的研究与应用[D]. 湖北工业大学, 2019.
21. 王艳, 孙薇, 周小平. 基于深度学习的中草药植物图像识别方法研究[J]. 医药信息, 2020, 37(06):21-25.
22. 孙颖异, 李健, 时天等. 基于改进的 AlexNet 卷积神经网络的植物叶片识别[J]. 种子, 2020, 39(02): 77-81.
23. 樊湘鹏, 周建平, 许燕, 彭炫. 基于改进卷积神经网络的复杂背景下玉米病害识别[J]. 农业机械学报:1-12[2021-01-11].
24. Pimm S. L., Jenkins C. N., Abell R., et al. The biodiversity of species and their rates of extinction, distribution, and protection[J]. Science, 2014, 344(6187): 1246752.
25. Knapp A. K., Fay P. A., Blair J. M. , et al. Rainfall Variability, Carbon Cycling, and Plant Species Diversity in a Mesic Grassland[J]. Science, 2002, 298(5601): 2202-2205.
26. Zhang S., Huang W., Huang Y., et al. Plant species recognition methods using leaf image:Overview[J]. Neurocomputing, 2020, 408: 246-272.
27. Wldchen J., Mder P.. Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review[J]. Archives of Computational Methods in Engineering, 2018, 25(2):507–543.
28. Thyagharajan K. K., Raji I. K.. A review of visual descriptors and classification techniques used in leaf species identification[J]. Archives of Computational Methods in Engineering, 2019, 26(4):933-960.
29. Tan C., Sun F., Kong T., et al. A survey on deep transfer learning[A]. International conference on artificial neural networks[C]. Cham: Springer, 2018: 270-279.
30. Ingrouillem J, Lairds M. A quantitative approach to oak variability in some north London woodlands[J]. The London Naturalist, 1986, 65: 35-46.
31. Osikar J O. Computer vision classification of leaves from Swedish trees[J]. Linkoping: Linkoping University, 2001, 181-186.
32. Rashad M Z, Eldesouky B S, Khawasik M S. Plants Images Classification Based on Textural Features using Combined Classifier[J]. International Journal of Computer Science & Information Technology, 2011, 3(4): 93-100.
33. Lee K B, Hong K S. An Implementation of Leaf Recognition System using Leaf Vein and Shape[J]. International Journal of Bio-Science & Bio-Technology, 2013, 214(2): 109-116.
34. Mehdipour Ghazi, Mostafa, Berrin, et al. Sabanci-Okan system in LifeCLEF 2015 plant identification competition[C]. CLEF 2015: 23-30.
35. Divya Tomar, Sonali Agarwal. Leaf Recognition for Plant Classification Using Direct Acyclic Graph Based Multi-Class Least Squares Twin Support Vector Machine[J]. International Journal of Image and Graphics, 2016, 16(3).
36. Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.
37. Kho S J , Manickam S , Malek S , et al. Automated plant identification using artificial neural network and support vector machine[J]. Frontiers in Life Science, 2017, 10(1):98-107.
38. Imah E M, Rahayu Y S, Wintarti A. Plant Leaf Recognition Using Competitive Based Learning Algorithm[J]. Iop Conference, 2018, 288(1).
39. Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neuralco mputation, 2006, 18(7): 1527-1554.
40. Jeon W, Rhee S. Plant Leaf Recognition Using a Convolution Neural Network[J]. International Journal of Fuzzy Logic and Intelligent Systems, 2017, 17(1): 26-34.
41. Ghazi, MM, Yanikoglu, et al. Plant identification using deep neural networks via optimization of transfer learning parameters[J]. Neurocomputing, 2017,685(2): 132-231.
42. Wu S G,Bao F S,Xu E Y,et al.A leaf recognition algorithm for plant classificationusing probabilistic neural network[C]//2007 IEEE international symposium on signalprocessing and information technology.IEEE,2007:11-16.
43. Boya Zhao, Mingjiang Wang,Ming Liu. An energy-efiicient coarse grained spatial architecture for convolutional neural networks AlexNet[J. IEICE Electronics Express,2017,14(15):595
44. Qu X, Lu H, Tang W, et al. A VGG attention vision transformer network for benign and malignant classification of breast ultrasound images[J]. Medical Physics, 2022, 49(09): 5787-5798.
45. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]Proceedings of the lEEE conference on computer vision and pattern recognition. 2015: 1-9.
46. 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. 2016: 770-778.

致 谢