

汤文亮, 黄梓锋. 基于知识蒸馏的轻量级番茄叶部病害识别模型[J]. 江苏农业学报, 2021, 37(3): 570-578.

doi: 10.3969/j.issn.1000-4440.2021.03.004

基于知识蒸馏的轻量级番茄叶部病害识别模型

汤文亮, 黄梓锋

(华东交通大学信息工程学院, 江西 南昌 330013)

摘要: 目前, 基于迁移学习诊断农作物病害已经成为一种趋势, 然而大多数研究使用的模型参数众多, 占用了大量设备空间并且推理演算耗时较长, 导致对存储和计算资源有严格限制的设备无法利用深度神经网络的优势。为此, 本研究以 PlantVillage 数据集中的番茄病害样本为研究对象, 基于条件卷积及通道注意力机制, 提出 1 种新颖的轻量级模型, 同时使用知识蒸馏法指导模型训练, 在保证模型性能的前提下压缩模型大小。将 AlexNet、VGG16、GoogLeNet、ResNet50 及 DenseNet121 进行对比, 并利用类激活图(CAM)可视化模型分类决策的图像区域。结果表明, 经过蒸馏的自定义模型可以精准定位番茄病叶的发病区域, 在测试集中的平均识别准确率达 97.6%, 不仅优于其他模型, 而且模型大小仅为 4.4 M。

关键词: 番茄病害; 识别模型; 条件卷积; 注意力机制; 知识蒸馏

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1000-4440(2021)03-0570-09

Lightweight model of tomato leaf diseases identification based on knowledge distillation

TANG Wen-liang, HUANG Zi-feng

(School of Information Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: At present, crop disease diagnosis based on transfer learning has become a trend. However, models used in most studies had a large number of parameters, which occupied a lot of equipment space and took much time for inference. The above conditions make devices with strict restrictions on storage and computing resources cannot take advantage of deep neural networks. Thus, tomato disease samples from the PlantVillage dataset were used as the research object in this study, a novel lightweight model based on conditional convolution and channel attention mechanism was proposed. At the same time, knowledge distillation method was used to train the custom model, which could greatly reduce the model volume while ensuring the performance of the model. The AlexNet, VGG16, GoogLeNet, ResNet50 and DenseNet121 were compared, and the class activation map (CAM) was used to visualize the image area in the model classification decision. The results showed that the distilled user-defined model could accurately locate the diseased tomato leaves. The average recognition accuracy in the test set was 97.6%, which was higher than other models and the model size was only 4.4 M.

Key words: tomato disease; identification model; conditional convolution; attention mechanism; knowledge distillation

收稿日期: 2020-09-10

基金项目: 江西省重点研发计划重点项目(20192ACB50027); 江西省教育厅科技重点项目(GJJ190296); 江西省交通厅项目(2019X0016)

作者简介: 汤文亮(1969-), 男, 江西南昌人, 硕士, 教授, 硕士生导师, 主要从事大数据分析 with 深度学习研究。(E-mail) 1402855900@qq.com

通讯作者: 黄梓锋, (E-mail) fshzfxqeng@163.com

中国番茄的常年种植面积达 6×10^4 hm^2 , 对经济发展和提高农户收入具有重要意义^[1]。然而, 由于番茄易受多种病虫害威胁, 导致其产量降低歉收^[2]。传统的人工鉴定农作物病害的方式依赖菜农或指导专

家的经验,容易出现错诊、漏诊,导致农民错过防控病害的最佳时间^[3]。目前,人们更偏向于选择高效精准的图像识别技术来诊断农作物病害。早期智能化诊断农作物病害的方式需要人工针对特定病害提取病斑纹理、形状及颜色等建立特征向量,再利用支持向量机法^[4]、随机森林法^[5]、K-均值法^[6]等机器学习算法进行分类。然而,这类方法过度依赖农业疾病领域的专家知识,每种病害识别都需要进行仔细而复杂的特征工程,前期工作繁重且模型泛化能力差。而在过去几年,深度学习(Deep learning, DL)的快速发展^[7]引起了大量研究人员的关注。DL通过分层体系结构从原始数据中自动学习有助于解决问题的表示形式^[8],其中卷积神经网络(Convolutional neural network, CNN)是一种学习效率很高的深度算法^[9],被广泛应用于农作物病害诊断中。

赵立新等^[10]先利用 PlantVillage 数据集训练改进的 AlexNet 模型,再使用棉花病虫害图像微调模型参数,其平均准确率达到 97.16%。黄双萍等^[11]以田间采集的 1 467 株水稻稻穗为研究对象拍摄高光光谱图像,使用 GoogLeNet 得出在验证集上的最高预测准确率为 92%。许景辉等^[12]采集 600 幅玉米大斑病、锈病和健康叶片图像,通过微调 VGG16 的全连接层后,对 3 种复杂田间背景下的玉米叶片平均识别准确率为 95.33%。龙满生等^[13]基于迁移学习,利用 AlexNet 对 4 种油茶病叶及健康叶片进行识别,平均识别准确率为 96.53%。Prabhakar 等^[14]将 PlantVillage 数据集中的番茄早疫病图像按照感染程度划分为早、中、晚 3 个阶段,再使用 ResNet101 进行病害识别,结果显示,该模型在测试集上的平均准确率达 94.60%。Mohanty 等^[15]使用 AlexNet 和 GoogLeNet 识别 PlantVillage 数据集中的 26 种农作物疾病,其最高准确率达 99.35%。郭小清等^[16]构建了基于 AlexNet 的多感受野识别模型,对番茄叶部病害的平均识别准确率为 92.70%。

虽然目前利用 CNN 识别农作物病害的研究已取得较好成果,然而这些研究大多直接采用迁移学习方式,虽然模型性能优异但参数众多,如 AlexNet 的参数数量有 6.0×10^7 个,VGG16 的参数数量更是多达 1.38×10^8 个,需要消耗大量设备资源,不能满足实际应用中的部署需求。此外,深度学习过程并不透明,即使模型在测试集上取得较高准确率,也不代表模型学习到具有区分度的特征。针对上述问题,本研究结合条件参

数化卷积^[17]及注意力^[18]机制,提出 1 种番茄叶部病害识别模型,同时利用知识蒸馏^[19]训练模型,使用梯度加权类激活^[20]映射的方式生成热力图以及可视化模型分类决策的图像区域。该轻量化模型可以在移动设备上部署,菜农根据设备中模型的识别结果精准施用农药,有效防止病害,从而降低经济损失。

1 材料与方法

1.1 数据筛选及预处理

番茄病害样本均来自 PlantVillage 数据集^[21],包含细菌性斑点病、早疫病、晚疫病、叶霉病、斑枯病、二斑叶螨病、轮斑病、花叶病、黄曲叶病 9 种番茄病叶及 1 种健康番茄叶片,详见图 1。原始数据集中各类样本数量分布不均,如黄曲叶病有 5 357 张图像样本,而叶霉病、花叶病图像样本数量分别仅为 952 张、373 张,因此模型学习的权重将倾向于黄曲叶病。为了避免该问题,使用水平翻转、垂直翻转及亮度增强 3 种数据增强方法将叶霉病、花叶病样本数量各扩充至 1 000 张;其余 8 种叶样本经过人工筛选后各种类型保留 1 000 张,总计 10 000 张图像。试验随机划分 6 400 张图像作为训练集拟合模型、1 600 张图像作为验证集评估模型的性能,并观察模型指标的变化趋势以监测模型是否过拟合,将剩余 2 000 张图像作为测试集,以评估模型的泛化能力。试验时将分辨率为 256×256 的图像裁剪成 224×224 ,并作归一化及标准化处理。

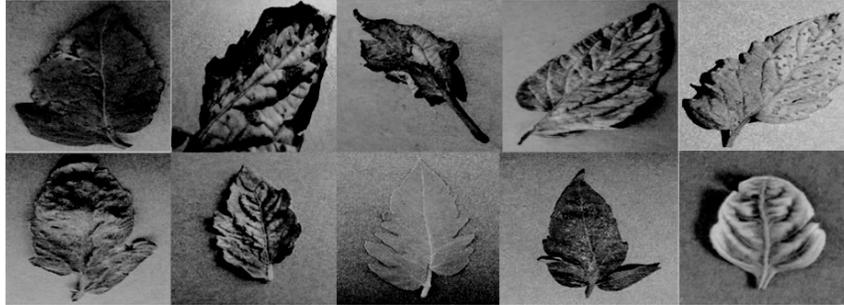
结合图 1 可以得出,本研究主要有 3 个难点:(1)番茄病叶的病斑表征相似,不同类别间仅有细微差异,而不同病叶样本的姿态及拍摄角度不同,在类别内存在较大差异,属于细粒度图像分类问题,分类难度大;(2)部分番茄病叶病斑的颜色与阴影相近,容易干扰模型的判别决策;(3)同种病害的图像背景相似,容易误导模型将背景作为判别依据。

1.2 病叶识别模型

本研究基于条件卷积(Conditionally parameterized convolutions, Cond Conv)和通道注意力模块(Squeeze-and-excitation block, SE block)构建 1 种轻量级模型 CondConvSENet,该模型包含 1 个输入层、5 个卷积结构、2 个全连接层及 1 个输出层。首先,输入层接收分辨率为 224×224 的番茄病叶彩色图像作为输入。然后,卷积结构逐层抽取输入图像的高级语义特征,该结构包含 Cond Conv、批量正则化、ReLU 激

活函数、SE block 及最大池化。最后,将卷积的输出张量转化为 1 维向量,输入全连接层中融合学习到的深度特征,并在输出层得到 10 元素向量,各个元素表

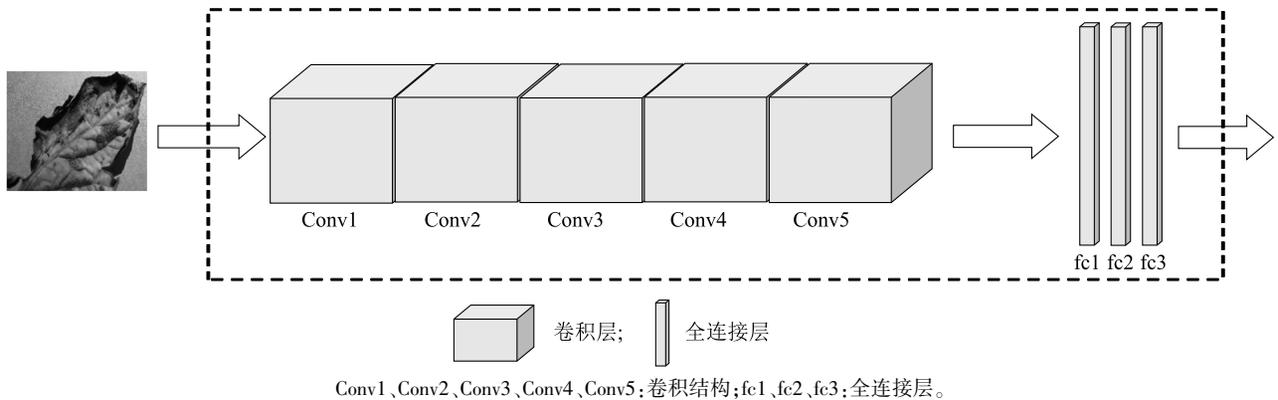
示模型对输入进行预测后分配给各个类别的概率。试验前先构建基础模型再进行改进,基础模型结构如图 2 所示,各个卷积结构的详细信息见表 1。



第 1 行图片从左到右分别为细菌性斑点病叶片、早疫病叶片、晚疫病叶片、叶霉病叶片、斑枯病叶片,第 2 行图片从左到右分别为轮斑病叶片、花叶病叶片、健康叶片、二斑叶螨病叶片、黄曲叶病叶片。

图 1 番茄叶片示例

Fig.1 Samples of tomato leaves



Conv1、Conv2、Conv3、Conv4、Conv5: 卷积结构; fc1、fc2、fc3: 全连接层。

图 2 基础模型结构

Fig.2 Basic model architecture

表 1 卷积结构的详细信息

Table 1 Details of convolution structures

名称	卷积核尺寸	步长	填充	输出张量
Input				3×224×224
Conv1	3×3	1	1	16×112×112
Conv2	3×3	1	1	24×56×56
Conv3	3×3	1	1	32×28×28
Conv4	3×3	1	1	48×14×14
Conv5	3×3	1	1	64×7×7

Input: 输入; Conv1、Conv2、Conv3、Conv4、Conv5: 卷积结构。

假设现有 1 组卷积核 $K = [k_1, k_2, \dots, k_c]$, 常规的卷积计算通过卷积核 K 将输入 $X = [x_1, x_2, \dots, x_c] \in R^{C \times H \times W}$ 映射为输出 $Y = [y_1, y_2, \dots, y_c] \in R^{\hat{C} \times \hat{H} \times \hat{W}}$, 其中第 i 通道输出的具体计算公式如下:

$$y_i = k_i * X = \sum_{j=1}^c k_i^j * x_j \quad (1)$$

式中, * 表示卷积计算; k_i 表示第 i 卷积核, 并且 $k_i = [k_i^1, k_i^2, \dots, k_i^c]$ 。各个通道的卷积输出按公式(1)进行计算, 这意味着通过扩展卷积核尺寸或增加其通道数量以提升模型性能的方式带来的额外计算量均将成正比地增加, 不切合实际需求。不同于常规的卷积计算, Cond Conv 先将卷积核参数化为若干个专家的线性组合, 如公式(2)所示。此时通过增加专家数量, 可以提升模型容量, 而这仅带来 1 个额外乘法累加的计算量, 相比于改变卷积核的方式, Cond Conv 显得更加高效。

$$Output(x) = (\alpha_1 W_1 + \alpha_2 W_2 + \dots + \alpha_n W_n) \times x \quad (2)$$

式中, $\alpha_i = \gamma_i(x)$, 是依赖于样本并可以通过梯度下降更新的加权参数。路由函数的结构与 SE block 类似, 但全连接层只有 1 层。先使用全局平均池化压缩输入, 并将结果输入全连接层, 可以得到路

由权重(α_i),将其与初始化的权重向量(W_i)进行乘积累加,获得卷积参数,最后与压缩的输入相乘,通过 Sigmoid 函数激活得到输出:

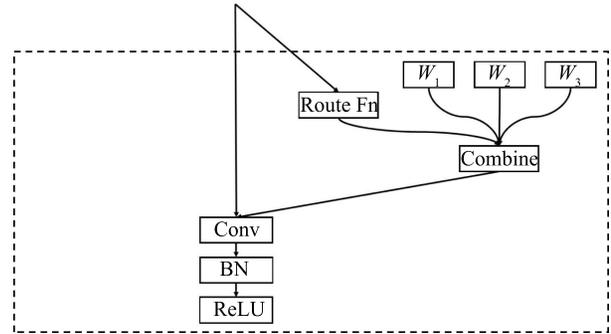
$$\gamma(x) = \text{Sigmoid}[\text{GAP}(x) \times R] \quad (3)$$

式中, GAP 表示全局平均池化, R 为参数矩阵。从数学角度看, Cond Conv 的计算等价于公式(4), 与 Inception^[22]、ResNext^[23] 等多分支卷积结构等效, 其中各个分支是单个卷积, 但不同的是 Cond Conv 仅需要计算 1 个卷积, 因此效率更高。Cond Conv 的结构如图 3 所示。

$$\text{Output}(x) = [\alpha_1 (W_1 \times x) + \alpha_2 (W_2 \times x) + \dots + \alpha_n (W_n \times x)] \quad (4)$$

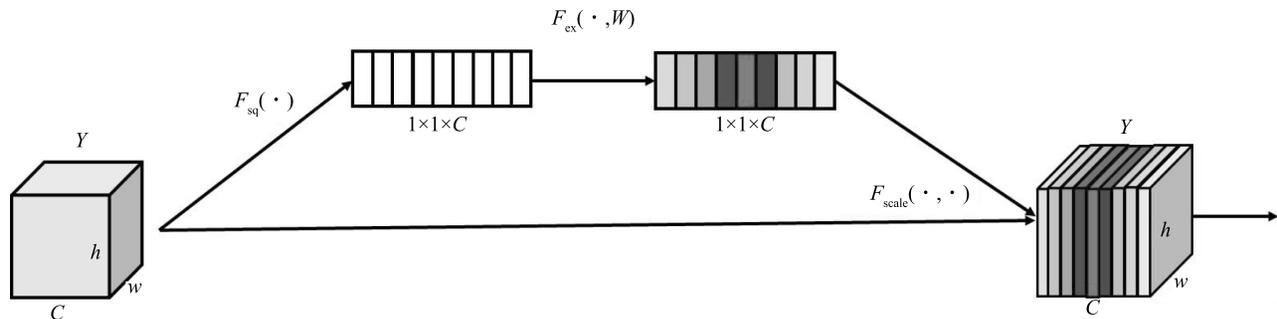
试验模型用 Cond Conv 代替图 2 卷积结构中的常规卷积, 以降低增加卷积核数量时所带来的额外计算量。随后模型嵌入了注意力模块 SE block, 通

过特征重标定操作强化重要特征, 抑制无关信息以引导模型学习得到更好的结果。SE block 的结构如图 4 所示。



Route Fn:路由函数;Conv:卷积;BN:批量正则化;ReLU:ReLU 激活函数; $W_1 \sim W_3$: 3 个专家;Combine:将专家进行线性组合。

图 3 Cond Conv 的结构
Fig.3 Structure of Cond Conv



$F_{sq}(\cdot)$:对特征图进行 Squeeze 操作; $F_{ex}(\cdot, W)$:对特征图进行 Excitation 操作; $F_{scale}(\cdot, \cdot)$:特征重标定操作。Y:图像统称; C:通道数; h, w :特征图尺寸。

图 4 SE block 的结构
Fig.4 Structure of SE block

图 4 中的 Y 可以看作经过卷积计算后由 C 张分辨率为 $h \times w$ 的特征图 Y_i 组成。由于 SE block 主要对特征通道间的相互依赖关系进行建模, 需要尽可能排除来自空间维度的信息干扰。 F_{sq} 使用全局平均池化方法将所有 Y_i 沿着空间维度进行压缩, 最终得到 $1 \times 1 \times C$ 的特征向量, 表征在特征通道上响应的全局分布, 其计算过程相当于将 1 个二维矩阵中所有数值求和后再求平均:

$$F_{sq}(Y_i) = \frac{1}{h \times w} \sum_{k=1}^h \sum_{j=1}^w Y_i(k, j) \quad (5)$$

F_{ex} 通过可学习参数 (W) 为每个特征通道生成权重, 在代码实现上使用 2 个全连接层代替 1 个全连接层。首先, 第 1 个全连接层通过放缩参数 (r) 将输入的特征维度 (C) 压缩成 C/r 以实现降维。然后, 使用 ReLU 激活后与第 2 个全连接

层连接以还原特征维度。这样模型可以学习更多非线性关系, 更好地拟合通道间复杂的相关性, 并减少参数量及计算量。最后, 利用 Sigmoid 函数获取 $[0, 1]$ 范围的归一化权重, 通过 Fscale 操作将其加权到各个通道的特征上, 完成特征重标定操作。

本研究使用 Cond Conv、SE block 改进图 2 的基础卷积结构, 但并不改变各个卷积层的输出, 因此其输出的张量仍与表 1 保持一致。

1.3 模型训练方法

模型训练的的目的是得到 1 组合适的参数, 这些参数蕴含模型从数据集中学习的知识, 然而在保证不损失知识的前提下, 压缩模型参数是十分困难的。而知识蒸馏法将模型学习的知识看作由输入向量到输出向量的映射, 其输出体现了模型所学的知识, 利

用复杂模型的输出指导小模型训练是一种更加可行的方案。运用知识蒸馏需要2个模型:教师模型和学生模型,该训练过程的目标函数由2个交叉熵函数组成,最终的损失值为两者的加权和,该训练流程如图5所示。若将教师模型的输出记为 f_q ,学生模型的输出记为 f_p ,真实标签记为 y ,交叉熵函数记为 CE ,调节损失函数比重的超参数记为 λ, μ ,那么得到损失函数如下:

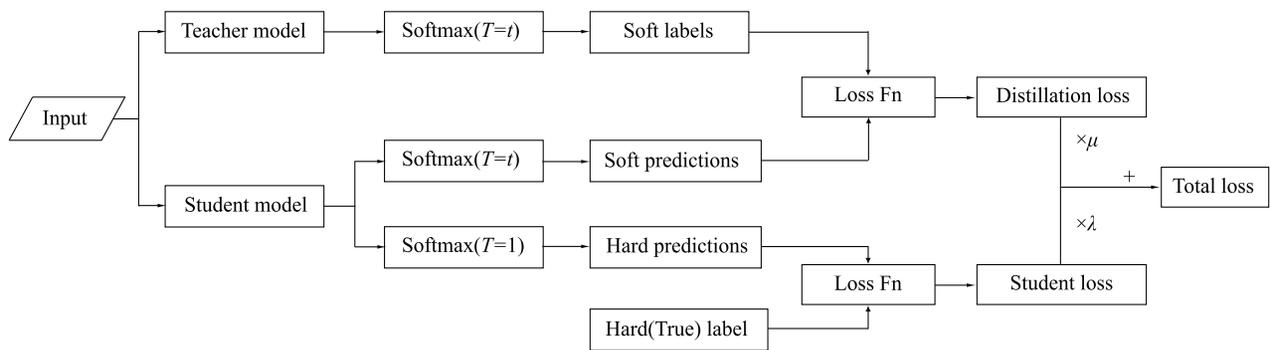
$$Loss = \lambda CE(y, f_p) + \mu CE(f_q, f_p) \quad (6)$$

通常模型使用 Softmax 输出层输出类别概率,这与经过 one-hot 编码的向量不同,其输出为各个类别分配了概率,蕴含很多有价值的信息,比如现假设某一试验得到的向量包含 A、B 及 C 3 个类别概率,其中 A 的概率是 0.7, B 的概率是 0.2, C 的概率是

0.1, 这表明相比于类别 C, 类别 B 与类别 A 更相似。然而模型对于预测为真值的类别标签有很高的置信度,如果 f_q 直接使用 Softmax 的输出结果,则无法将教师模型输出概率中丰富的知识迁移到学生模型中,因此文献[19]引入了温度系数(T)平滑标签,此时 f_q 的计算公式如下:

$$f_q = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (7)$$

式中,温度系数(T)越高则输出结果分布越平缓,当 $T=1$ 时, f_q 为 Softmax 函数。本试验使用的教师模型是在 ImageNet^[24]完成训练的 SEResNet101。由于 ImageNet 与番茄病害数据集的相似度较低,需要先使用番茄病害图像微调 SEResNet101 参数;学生模型为自定义的 CondConvSENet。



Input: 输入; Teacher model/Student model: 教师模型/学生模型; Softmax($T=t$ 或 $T=1$): 将温度系数设置为 t 或 1 以调整 Softmax 的输出分布; Soft labels/Hard(True) label: 经过调整的标签值/实际的标签值; Soft predictions/Hard predictions: 经过调整的 Softmax 输出向量/原始的 Softmax 输出向量; Loss Fn: 损失函数; Distillation loss/ Student loss: 蒸馏的损失/未经蒸馏的损失; Total loss: 对图中上、下 2 个损失值进行加权求和(μ 和 λ 均为超参数)计算得到的总体损失。

图5 利用知识蒸馏训练的模型

Fig.5 Model trained by knowledge distillation

1.4 损失函数与优化方法

本研究为单标签多分类问题,使用 pyTorch 的交叉熵函数 CrossEntropyLoss() 作为损失函数。假定未经过 Softmax 的输出 $z = [z_0, z_1, \dots, z_{class-1}]$, 样本标签记为 $class$, 此时学生模型与标签的损失计算过程见公式(8), 该公式(以 e 为对数的底数)结合 Softmax 和负对数损失, 故模型在输出层输出预测值后不再使用 Softmax 函数。

$$L(z, class) = -\ln \frac{\exp(z[class])}{\sum_{j=0}^{C-1} \exp(z[j])} = -z[class] + \ln \sum_{j=0}^{C-1} \exp(z[j]) \quad (8)$$

优化器采用基于动量因子的随机梯度下降(Stochastic gradient descent, SGD)算法进行梯度更新。公式(9)为 t 时刻动量的计算方法:

$$m_t = \gamma m_{t-1} + \omega \frac{\partial Loss}{\partial \theta_t} \quad (9)$$

式中, γ 表示超参数, t 表示时刻, m_{t-1} 表示 $t-1$ 时刻的动量, ω 表示学习率, θ 表示权重, $Loss$ 表示损失函数值。

利用公式(10)更新权重:

$$\theta_{t+1} = \theta_t - m_t \quad (10)$$

相比于 SGD 算法,基于动量因子的 SGD 算法包含过去的历史信息,因而计算梯度更接近真实梯

度,提高了算法的稳定性。

2 结果与分析

2.1 试验设定

在配有 1 块 NVIDIA GeForce RTX 2080Ti GPU 的计算机上进行本试验的运行测试,运行环境为 64 位 Windows 10 操作系统,实现算法的编程语言版本为 Python 3.7.6,深度学习框架版本为 pyTorch 1.3.1,并以准确率作为参考指标。试验时,所有模型的训练轮次及数据批次均为 100 个 epoches,将 Cond Conv 的专家数量设置为 3 个,将 SE block 的放缩参数(r)设置为 16,将 SGD 算法的动量因子(γ)设为 0.9,将学习率(ω)初始化为 0.1,并采用固定步长调整学习率的策略,每间隔 30 个 epoches 将 ω 缩小为原来的 10%。为了抑制模型过拟合,采用权重衰减和 Dropout 技术,并将权重衰减系数设置为 0.000 1,Dropout 随机舍弃神经元的概率为 0.3。在模型训练时将知识蒸馏的温度系数(T)设置为 5,超参数(λ)设为 1, μ 设为 0.8。

2.2 模型训练与测试分析

首先使用知识蒸馏训练 CondConvSENet,如图 6 所示,经过蒸馏的模型曲线在未经蒸馏的模型曲线上方收敛,表明使用知识蒸馏可以有效提升模型的性能。随后,微调 5 种经典模型 AlexNet^[25]、VGG16^[26]、GoogLeNet^[22]、ResNet50^[27] 及 DenseNet121^[28] 的全连接层,将其替换为 1 个全连接层,输出 10 个类别概率,形成试验对比。由于 CondConvSENet 使用知识蒸馏从头开始训练,耗时相对较长,模型完成 100 个 epoches 的训练时间大约为 200 min,其余模型大约花费 2 h。所有模型在验证集上识别准确率的变化趋势如图 7 所示。

由图 7 可以看出,5 种经典模型拥有很高的初期识别准确率,这是由于 CNN 浅层提取的是输入图像的边缘、纹理等更具有泛化性的特征,同时这些模型均在百万级图像数据库 ImageNet 中得到充分训练与评估,具有很强的泛化能力。DenseNet121、ResNet50 及 GoogLeNet 模型的初期识别准确率超过 70.00%,高于 AlexNet 和 VGG16 模型。其中可能的原因是后 2 个模型的网络结构主要基于模型深度的角度,通过简单堆叠卷积层和全连接层构建而成,而 GoogLeNet 通过 Inception 模块扩展模型的深度和宽度以融合更多特征,因此其性能相对更好,但是这 3

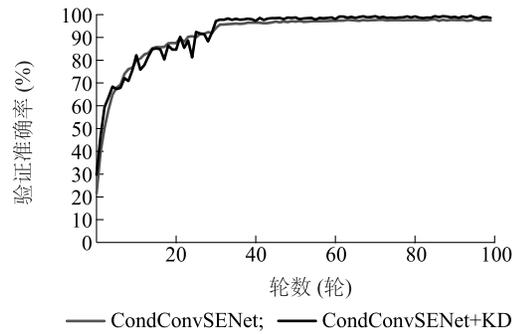


图 6 知识蒸馏前后的对比

Fig.6 Comparison of results before and after knowledge distillation

种模型并未解决梯度消失问题,导致模型的深度受到限制。ResNet50 及 DenseNet121 模型通过短连接,有效避免了模型训练因梯度消失而导致的性能下降,并保证了第 L 层卷积的输出不差于第 $L-1$ 层,因此可见,模型构建的结构层次越深,就越能学习到更多的非线性表示,使得初期的识别准确率高过前 3 种模型。由于 CondConvSENet 是从头开始学习拟合数据集的特征参数,因此初期识别准确率略低于 30.00%,如果没有使用知识蒸馏指导训练,该指标会更低。与此同时,CondConvSENet 模型受益于 SE block,其准确率在随后的训练中快速稳定地上升,并在第 29 个 epoch 时超过 5 种经典模型,其最佳验证准确率达 99.44%。

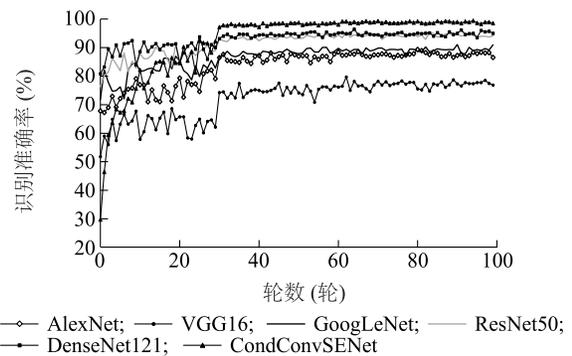


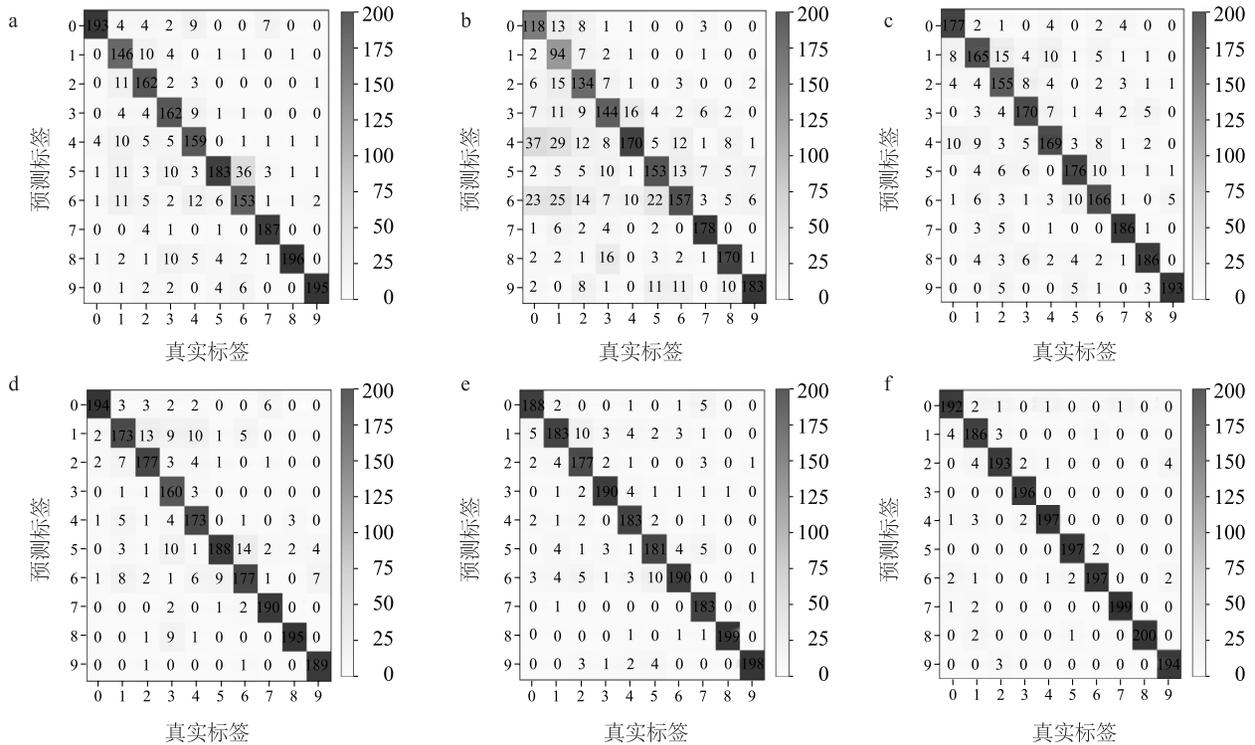
图 7 模型在验证集上识别准确率的变化曲线

Fig.7 Changing curves of recognition accuracy of models in validation set

模型训练结束后需要在测试集上评估其泛化能力。本试验使用如图 8 所示的混淆矩阵表示模型对验证集中图像的识别结果。通过混淆矩阵计算 6 种模型对各种病害的识别准确率。由表 2 可以看出,

这 6 种模型的平均识别准确率与验证集上的表现相近,具有良好的泛化能力。其中,VGG16 模型的平均识别准确率仅有 75.1%,与使用 VGG16 模型解决类似问题的文献[16]中的结果大相径庭,这可能是由于试验时只将模型全连接层替换为 1 层输出 10 个类别,如要得到更好的结果,需要进一步调整优化。ResNet50、DenseNet121 模型的效果较优,平均

识别准确率均超过 90.0%,而 CondConvSENet 模型除了细菌性斑点病的识别准确率略低于 ResNet50 模型的 97.0%以及对黄曲叶病的识别准确率略低于 DenseNet121 模型的 99.0%外,对其余 8 种叶片的识别准确率均达到最高值,其中对花叶病的识别准确率更是高达 100.0%,即 200 个测试样本全部被正确识别。



a: AlexNet; b: VGG16; c: GoogLeNet; d: ResNet50; e: DenseNet121; f: CondConvSENet。横纵坐标的数字 0~9 表示番茄病害数据集的 10 个类别。

图 8 本试验模型的混淆矩阵

Fig.8 Confusion matrix of the experimental model

上述试验结果表明,本研究中使用的 6 种模型具有较好的分类准确率,但是端到端的学习方式并不透明,在试验时将数据输入模型得到输出结果,却无法知晓其决策分类过程,不免让人质疑。然而对于 CNN 而言,该过程是可视化的,试验使用类激活图(Class activation map, CAM)可视化 CondConvSENet 的结果,以便观察模型识别番茄病害的特征是否具有区分度。

从表 2 可以看出,VGG16 对花叶病的识别准确率较高,达 85.0%,然而试验随机抽取 2 张图片进行预测并将结果可视化,发现 VGG16 将图像背景作为判别依据。使用 CondConvSENet 对同样的

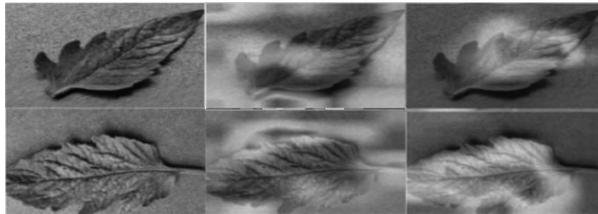
样本进行预测,发现模型对番茄发病部位的激活度很高,表明模型准确定位到了病害区域,模型学习得到了真正有价值的特征,该对比结果如图 9 所示。同样,本试验随机抽取另外 8 种病叶图像各 1 张,用 CondConvSENet 进行预测并可视化结果。如图 10 所示,CondConvSENet 忽视了图像背景,只关注番茄病叶的发病部位,模型预测结果的可信度高。本试验采用的 6 种模型的最终大小见表 3。

由此可见,CondConvSENet 不仅保证了优秀的识别准确率,还极大程度地减少了模型的参数量,缩小了模型体积。

表 2 CondConvSENet 模型在测试集上的识别准确率

Table 2 Recognition accuracy of CondConvSENet model in the test set

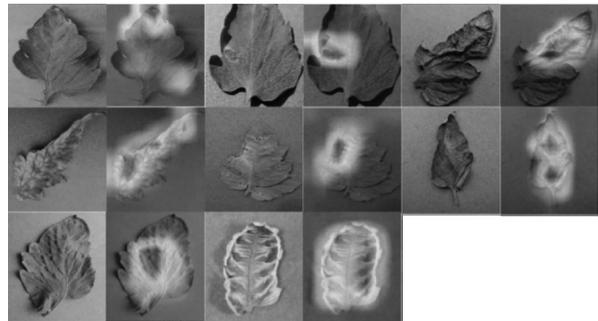
模型类别	识别准确率 (%)										
	细菌性斑点病	早疫病	健康叶片	晚疫病	叶霉病	斑枯病	二斑叶螨病	轮斑病	花叶病	黄曲叶病	平均
AlexNet	96.5	73.0	81.0	81.0	79.5	91.5	76.5	93.5	98.0	97.5	86.8
VGG16	59.0	47.0	67.0	72.0	85.0	76.5	78.5	89.0	85.0	91.5	75.1
GoogLeNet	88.5	82.5	77.5	85.0	84.5	88.0	83.0	93.0	93.0	96.5	87.2
ResNet50	97.0	86.5	88.5	80.0	86.5	94.0	88.5	95.0	97.5	94.5	90.8
DenseNet121	94.0	91.5	88.5	95.0	91.5	90.5	95.0	91.5	99.5	99.0	93.6
CondConvSENet	96.0	93.0	96.5	98.0	98.5	98.5	98.5	99.5	100.0	97.0	97.6



同一行 3 张图片为 1 组,从左到右分别为原始图像、VGG16 及 CondConvSENet 的可视化结果。

图 9 VGG16 及 CondConvSENet 可视化结果对比

Fig.9 Comparison of visualization results between VGG16 and CondConvSENet



同一行 2 张图片为 1 组,1 组中左边为原始图像,右边为可视化结果,从左到右分别为细菌性斑点病、早疫病、晚疫病、叶霉病、斑枯病、二斑叶螨病、轮斑病、黄曲叶病。

图 10 可视化 CondConvSENet 识别结果

Fig.10 Recognition result of visualized CondConvSENet

3 结论

使用深度学习算法^[29]进行番茄病害识别,不需要事先针对特定病害类别进行复杂的特征工程,模型可以自动从原始数据中学习具有区分度的特征,

省

表 3 本研究中所有模型的大小

Table 3 Sizes of all models in the study

模型类别	模型大小(M)
AlexNet	217.8
VGG16	512.5
GoogLeNet	21.6
ResNet50	90.1
DenseNet121	27.1
CondConvSENet	4.4

去了大量前期工作。相比于重新训练模型,迁移学习可以大幅度缩短模型的训练时间,同时预训练模型拥有极强的泛化能力,即便只微调模型的全连接层,模型也能在番茄病叶分类问题中取得较好的识别效果,但是模型的最终大小和参数量庞大,而利用知识蒸馏可以在减少模型参数的同时提升模型的性能。

番茄病叶分类属于细粒度图像分类任务,同时受到背景及光照阴影等干扰因素影响,模型难以准确捕捉具有区分度的特征,而利用注意力机制可以改善模型的性能,引导模型学习重要的分类特征。模型在数据集中取得满意的指标后,需要进一步使用 Grad-CAM 等技术进行可视化验证,以确保模型利用有价值的特征作为决策依据。

参考文献:

[1] 吴琼,梁巧兰,张娜. 番茄主要病害病原菌培养条件及室内药剂筛选[J]. 甘肃农业大学学报, 2018, 53(5): 79-86.
 [2] 谭海文,吴永琼,秦莉,等. 我国番茄侵染性病害种类变迁及

- 其发生概况[J]. 中国蔬菜, 2019, 359(1):86-90.
- [3] 贾少鹏,高红菊,杭 潇. 基于深度学习的农作物病虫害图像识别技术研究进展[J]. 农业机械学报, 2019, 50(增刊1):313-317.
- [4] 李 颀,王 康,强 华,等. 基于颜色和纹理特征的异常玉米种穗分类识别方法[J]. 江苏农业学报, 2020, 36(1):24-31.
- [5] 张经纬,贡 亮,黄亦翔,等. 基于随机森林算法的黄瓜种子腔图像分割方法[J]. 农机化研究, 2017, 39(10): 163-168.
- [6] GINNE M, PUNITHA S C. Tomato disease segmentation using K-means clustering[J]. International Journal of Computer Applications, 2016, 144(5):25-29.
- [7] GUO Y, LIU Y, OERLEMANS A, et al. Deep learning for visual understanding: a review[J]. Neurocomputing, 2016, 187(26): 27-48.
- [8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436.
- [9] 周飞燕,金林鹏,董 军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6):1229-1251.
- [10] 赵立新,侯发东,吕正超,等. 基于迁移学习的棉花叶部病虫害图像识别[J]. 农业工程学报, 2020, 36(7):184-191.
- [11] 黄双萍,孙 超,齐 龙,等. 基于深度卷积神经网络的水稻穗瘟病检测方法[J]. 农业工程学报, 2017, 33(20):169-176.
- [12] 许景辉,邵明焯,王一琛,等. 基于迁移学习的卷积神经网络玉米病害图像识别[J]. 农业机械学报, 2020, 51(2): 25.
- [13] 龙满生,欧阳春娟,刘 欢,等. 基于卷积神经网络与迁移学习的油茶病害图像识别[J]. 农业工程学报, 2018, 34(18):194-201.
- [14] PRABHAKAR M, PURUSHOTHAMAN R, AWASTHI D P. Deep learning based assessment of disease severity for early blight in tomato crop[J]. Multimedia Tools and Applications, 2020, 79: 1-12.
- [15] MOHANTY S P, HUGHES D P, SALATHÉ M. Using deep learning for image-based plant disease detection[J]. Frontiers in Plant Science, 2016, 7:1419.
- [16] 郭小清,范涛杰,舒 欣. 基于改进 Multi-Scale AlexNet的番茄叶部病害图像识别[J]. 农业工程学报, 2019(13):162-169.
- [17] YANG B, BENDER G, LE Q V, et al. Condconv: conditionally parameterized convolutions for efficient inference [C]//SCHÖLKOPF B, PLATT J, HOFMANN T. Advances in Neural Information Processing Systems. Vancouver, Canada; 2019:1307-1318.
- [18] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 42: 2011-202399.
- [19] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7):38-39.
- [20] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2019, 128(8): 336-359.
- [21] HUGHES D, SALATHÉ M. An open access repository of images on plant health to enable the development of mobile disease diagnostics[EB/OL]. (2015-11-25) [2020-09-01]. <https://arxiv.org/abs/1511.08060>.
- [22] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//IEEE. Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, MA:IEEE, 2015: 1-9.
- [23] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//IEEE. Proceedings of the IEEE conference on computer vision and pattern recognition. New York:IEEE Press, 2017: 1492-1500.
- [24] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [25] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [26] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2020-09-01]. <https://arxiv.org/abs/1712.031497>.
- [27] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE conference on computer vision and pattern recognition. New York: IEEE Press, 2016: 770-778.
- [28] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]//IEEE. Proceedings of the IEEE conference on computer vision and pattern recognition. New York:IEEE Press, 2017: 4700-4708.
- [29] 李懿超,沈润平,黄安奇. 基于深度学习的湘赣鄂地区植被变化及其影响因子关系模型[J]. 江苏农业科学, 2019, 47(3): 213-218.

(责任编辑:徐 艳)