

多分支分组卷积的特征级联农作物病害识别^{*}

吕佳^{1,2}, 彭港建^{1,2}, 巫若愚^{1,2}

(1. 重庆师范大学 计算机与信息科学学院; 2. 重庆市数字农业服务工程技术研究中心, 重庆 401331)

摘要:针对现有小样本学习中农作物病害识别参数量高、网络易出现梯度消失等问题,提出一种多分支分组卷积的特征级联农作物病害识别模型。首先,该模型将分组卷积应用于残差块,利用残差块缓解网络的梯度消失问题。其次,将网络的输入拆分成同构多分支的嵌入,通过不同分支间的特征补充,获取更加丰富的特征并降低网络参数量。最后,设计了一种多层特征级联的骨干网络,用于挖掘图像中不同层次的信息,将浅层特征中更多位置、细节信息和深层特征中更强的语义信息级联输出,提升网络的泛化能力。在 PlantVillage 数据集上,该模型在 5-way 1-shot 与 5-way 5-shot 任务上的识别准确率分别达到了 $(79.68 \pm 0.35)\%$ 和 $(93.25 \pm 0.21)\%$,比原型网络分别提升了 3.74% 和 4.32%。实验结果表明,本文模型能较好地识别农作物病害,并且在加深网络深度的同时降低了网络参数量、缓解了梯度消失以及挖掘出了不同层次的语义信息,有效解决了农作物病害小样本学习问题。

关键词:农作物病害;图像识别;小样本学习;分组卷积;特征级联

中图分类号:TP391

文献标志码:A

文章编号:1672-6693(2024)05-0115-14

中国是农业大国,在推进农业现代化的进程中,农业生产方式正在从传统人工模式逐步向智能化方向转变,人工智能技术在农业领域得到了广泛应用^[1]。在农作物种植中,病害是影响产量和质量的关键因素之一^[2]。当前农作物病害识别大多仍需依靠具有丰富知识经验的专家来进行,因而难以满足大面积且快速准确诊断农作物病害的实际生产需求。近些年来,深度学习在计算机视觉领域取得了巨大成功,而利用深度学习识别农作物病害的智慧农业也展现出广阔的应用前景^[3-5]。然而,在实际农业生产中,由于患病农作物的比例较低,采集到的病害图像样本数量也有限,加之病害种类多样且部分病害发生概率较低,因此在有限的病害数据集上利用深度学习模型进行训练时容易出现模型过拟合的问题^[6]。

为了解决深度神经网络数据依赖的问题,受人类学习方式的启发,小样本学习(few-shot learning, FSL)^[7-8]应运而生。小样本学习旨在通过少量有标记样本的训练,使模型对从未见过的样本具备快速理解和泛化的能力。其中,元学习是现有小样本学习相关研究中讨论得最多的训练策略,并且在农作物分类和识别中得到了有效应用^[9-11]。任胜男等人^[12]针对 PlantVillage 数据集中样本量较少的 8 类病害图像,提出了一种基于关系网络^[13]的农作物病害模型,采用焦点损失函数来增强模型对困难样本的关注度,即通过调整损失函数的超参数,使分类器在训练过程中更聚焦于识别难以分类的样本。最终相较于原始关系网络模型和匹配网络^[14],该方法的识别准确率分别提高了 4.69% 和 25.02%。Argüeso 等人^[15]提出了一种具有 Triplet 损失的孪生神经网络,首先在 Inception V3 网络上进行微调,用以提取树叶图像特征,然后训练 1 个多类 SVM 分类器来学习不同农作物叶片的映射,在 PlantVillage 数据集上对 6 种病害进行实验,结果表明该网络的识别准确率均超过 90%。但该方法需要调整不同类别间的距离,使用 Triplet 损失也会增加网络参数量,且 Triplet 损失在计算速度上比交叉熵损失慢,导致网络识别病害的效率较低。Chen 等人^[10]提出了一种基于元学习的局部特征匹配条件神经自适应模型,能够自适应地调整 2 个类别之间的差异,在苹果叶片病害数据集和 PlantVillage 数据集上进行了实验,识别准确率达到 93.9%,然而该方法在处理规模较大的数据集时存在识别效率低的问题。

^{*} 收稿日期:2023-02-24 修回日期:2024-09-04 网络出版时间:2024-11-15T10:43

资助项目:重庆市高校创新研究群体项目(No. CXQT20015)

第一作者简介:吕佳,女,教授,博士,研究方向为机器学习、深度学习、数字农业,E-mail:lvjia@cqnu.edu.cn;通信作者:彭港建,男,E-mail:penggj98@163.com

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20241115.0841.002

与元学习不同,文献[16-18]从数据增强角度提出了扩充样本的策略,以应对样本量不足的问题。吴昊昱^[16]以 3 种茶叶病害为研究对象,首先对病害进行分割,并对分割后的病害数据集进行扩充,然后采用深度迁移和 Cayley-Klein 度量相结合的方法对茶叶病害进行识别,识别准确率接近 100%。Dhivyaa 等人^[17]使用图像颜色直方图变换技术生成合成图像,用改进的 MobileNetV2 识别 5 类木薯病害,实验结果表明该模型的识别准确率为 99.7%,比使用无数据增强的数据集训练的模型识别准确率提高了 2%。除此之外,还有通过剪切方式增加图片数量的策略,Pan 等人^[18]使用 Faster R-CNN 和小样本学习自动估计草莓叶焦的严重程度,提出一种两阶段模型,第 1 阶段使用 Faster R-CNN 标记草莓叶片位置,并从原始草莓叶片图像中剪切每个单片草莓叶,组成新的草莓叶片数据集;第 2 阶段使用原型网络(prototypical network,PN)^[19]进行分类,最终在包含了 60 张草莓叶片图像的新数据集上达到了 88.83% 的识别准确率。上述方法虽然增加了样本数量,对病害识别的准确率也较高,但在特征提取阶段却没有充分利用到样本不同层次的特征信息,且检测速度较慢,对设备要求也较高。

为了解决这些问题,本文提出一种多分支分组卷积的特征级联农作物病害识别模型,该模型以原型网络^[19]作为小样本学习框架,主要包含以下创新点:

1) 针对已有相关方法参数量多、训练开销大的问题,在 ResNeXt 基础上,对单一残差块进行了扩展,形成多分支结构,并采用分组卷积的策略,使模型以并行的方式在多个图形处理器上训练,从而让网络在每个步骤上均能处理更多的图像,以减少网络参数量,提高模型效率。

2) 针对样本特征信息未被充分利用的问题,在骨干网络的基础上,提出了一种多层特征归一化级联的方法,以便更充分地利用样本特征信息。该方法基于特征金字塔网络,为应对特征提取过程中物体的多尺度变化问题,有选择地提取每层特征图中能够提升特征提取效果的部分来进行横向融合。这样在计算量增加较少的情况下,整合了低分辨率、语义信息较强的特征图与高分辨率、空间信息丰富的特征图,实现了更高效的多尺度特征表达,提高了模型的特征提取精度。

1 本文模型

1.1 元学习

在基于元学习的小样本学习训练过程中,网络分为元训练、元验证和元测试这 3 个阶段,每个阶段使用不同类别的样本数据来提高模型的泛化能力,对应的数据集分别记作 D_{train} 、 D_{val} 、 D_{test} 且 $D_{\text{train}} \cap D_{\text{val}} \cap D_{\text{test}} = \emptyset$ 。元学习以元任务为基本单元,每个元任务的数据集由支持集(support set)和查询集(query set)组成,以便网络通过少量样本学习类间特征。元训练阶段,通过多个元任务进行训练,使网络能够反复在小样本条件下提升特征识别和泛化能力。元验证在每轮元训练后进行,用于判断是否保存本次训练学到的参数。元验证与元测试阶段使用元训练阶段所学习到的参数,并且所构建的元任务也相同。经过元训练与元验证后,元测试采用元训练阶段训练出的网络来进行目标类样本的识别。通过上述方式的学习,使得网络能够获得在面对未知的新类样本时也能够表现出良好的学习能力,而不仅仅是判别特定类别。

在小样本识别任务中,训练集包含了大量带标签样本,称为基类(base class)样本,用于模型训练;验证集与测试集中为现实中较难获得的无标签样本,称为新类(novel class)样本,用于预测标签。在训练阶段,每轮从训练集中随机选择 N 个类别,每类选择 K 个样本,选择的 $N \times K$ 个样本为支持集,记为 $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_{N \times K}, y_{N \times K})\}$,其中 $\mathbf{x}$ 是 D 维原始数据的向量化表示, y 为对应的类别,这样构成了 N -way K -shot 任务,再从 N 个类别中随机选择 M 个样本组成 $N \times M$ 个样本的查询集。

1.2 原型网络

Snell 等人^[19]提出了原型网络。在原型网络中,每个类别都存在 1 个原型,各类的原型点 C_k 是该类在嵌入空间中的均值,数据集中距离该原型点越近的样本,标签与该原型点对应的标签相同的概率也就越大。因此,对样本的识别问题就变成了目标样本在嵌入空间中的最近邻问题。原型网络基本概念如图 1 所示。

图 1 中 C_1, C_2, C_3 分别是 3 个类别的原型点,目标样本通过 1 个嵌入函数 $f_\varphi: \mathbf{R}^D \rightarrow \mathbf{R}^M$ 将 D 维的样本映射到 M 维的空间上,随后计算与这 3 个中心点的距离,从而获得 $\mathbf{x}$ 的类别。以元训练的 1 个任务 D_{train} 的支持集 $D_{\text{train_support}}$ 为例,原型点 C_k 的计算公式为:

$$C_k = \frac{1}{|D_k^{\text{train_support}}|} \sum_{x_i \in D_k^{\text{train_support}}} f_{\varphi}(x_i), \quad (1)$$

式中: $D_k^{\text{train_support}}$ 为 $D^{\text{train_support}}$ 中待标记类别序号为 k 的样本集合, x_i 表示支持集 $D^{\text{train_support}}$ 中的每个样本。

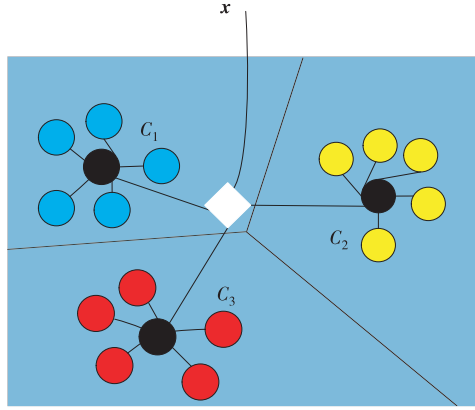


图 1 原型网络

Fig. 1 Prototype network

以相同映射作用于查询集 $D^{\text{train_query}}$ 中的数据, 给定距离函数 d , 计算支持集与查询集数据的相似度, 对待检测样本 x 利用 softmax 输出类别分布, 计算公式为:

$$p(y = k | x) = \frac{\exp(-d(f(x), C_k))}{\sum_{k'} \exp(-d(f(x), C_{k'}))}. \quad (2)$$

其中: k' 指样本序号。

最后采用交叉熵损失计算预测值与真实值之间的差异, 然后反向传播更新模型参数, 计算公式为:

$$L = -\frac{1}{i} \sum_i \sum_{k=1}^N y_{ik} \log p_{\varphi}(y_i = k | x_i), \quad (3)$$

式中: i 为查询集中样本个数; N 为查询集中类别个数; y_{ik} 表示第 i 个样本的标签是否为 k , 若为真, 则取值为 0, 否则为 1; $p_{\varphi}(y_i = k | x_i)$ 表示第 i 个样本标签为 k 的概率值。

1.3 多分支分组卷积残差块

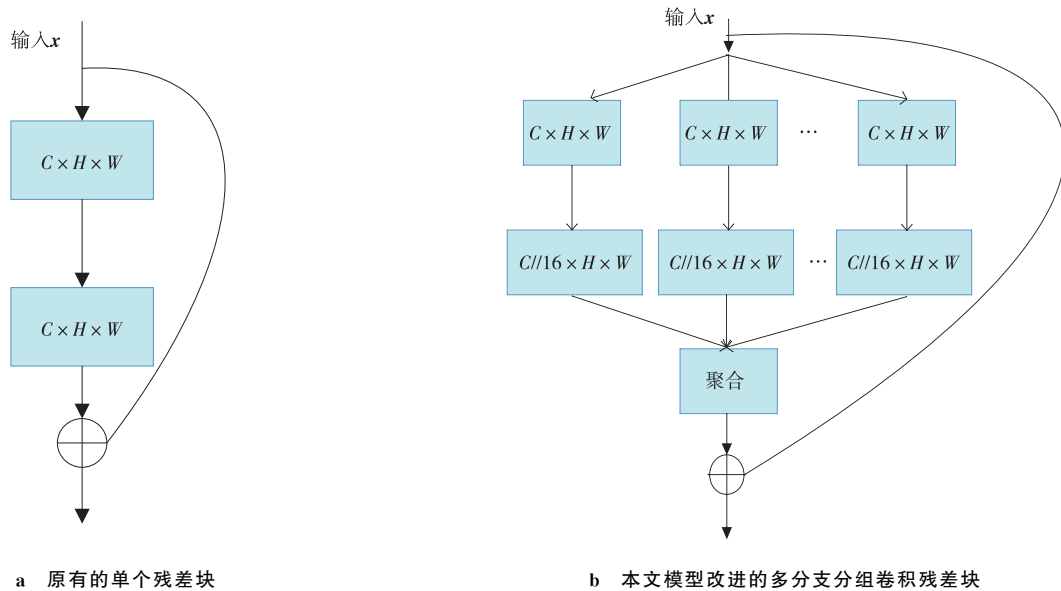
Li 等人^[20]的研究发现, 随着网络深度的增加, 测试误差和训练误差逐渐增大, 并且更深的网络会带来梯度消失问题, 从而阻碍了网络的收敛。为解决该问题, He 等人^[21]提出了 ResNet 网络, 通过给非线性的卷积层增加直连边的方式, 即直接使用步长为 2 的卷积做下采样, 用全局平均池化替换全连接层, 添加残差模块来提高信息的传播效率, 尽可能在不加深网络的情况下, 提升模型的识别准确率。

Chen 等人^[22]在 ResNet 中添加了 Inception 分支网络, 这种方法将网络的输入拆分成同构多分支嵌入, 每个分支结构相同, 通过不同分支间的相互补充, 再级联合并以取得更加丰富的特征, 获得优于 ResNet 模型的效果, 同时还减少了参数量。Xie 等人^[23]提出了 ResNet 模型的改进模型 ResNeXt; 与 Chen 等人^[22]在 ResNet 中添加 Inception 分支网络的方式类似, ResNeXt 通过不同分支之间的特征互补实现了更丰富的特征表达, 并在提升模型性能的同时还有效减少了参数量。这种设计在保持高效特征学习的基础上进一步缓解了网络深度增加带来的参数冗余问题。受上述文献启发, 本文在 ResNeXt 基础上, 扩展原有的单个残差块, 将多个残差块加以聚合, 再将网络每一层的残差块调整为适合于浅层网络的分组卷积残差块。原有的单个残差块与本文模型改进残差块的对比见图 2, 并且本文将特征图分为 16 组 3×3 卷积拼接特征图。

1.4 多层特征归一化级联的骨干网络

识别不同尺度的物体一直是计算机视觉领域的一个挑战, 面对多尺度目标检测问题, 学者们经常使用特征金字塔网络来解决, 但由于特征金字塔计算量大, 一定程度上降低了整体检测速度。为了提高检测速度, 如果只使用高层的特征来进行预测, 虽然丰富了语义信息, 但会导致分辨率降低, 很难准确地保存物体的位置信息。因

此,为平衡检测速度与物体位置信息,Zuo 等人^[24]在高层网络得到了语义化程度较高的特征后,通过多次上采样逐步恢复空间分辨率,并将每次上采样得到的特征与网络中的浅层特征进行融合,再与网络中的浅层特征融合。融合后的特征不仅包含了高级语义信息,网络还能保持较高的空间分辨率和物体位置信息,从而提高了检测精度和效率。



注: C 表示提取到的特征图通道个数; H, W 分别表示特征图的高、宽,均以像素值计。

图 2 残差块对比

Fig. 2 Residual block comparison diagram

受此启发,本文在特征金字塔网络的基础上,设计了 1 个包含多层特征归一化级联的骨干网络。首先从网络结构中提取多层特征图,然后将它输入 1 个可更新参数的线性层,该线性层可对每层中能够提升特征提取效果的部分进行横向提取。接着,将每层有效特征通过 1×1 卷积和最大池化进行下采样,统一为相同大小和通道数的结构,并通过调整各层特征的权重来优化融合效果,最终将所有层的特征级联输出。特征金字塔网络和本文设计的多层特征归一化级联骨干网络在特征提取方面的性能对比如图 3 所示。

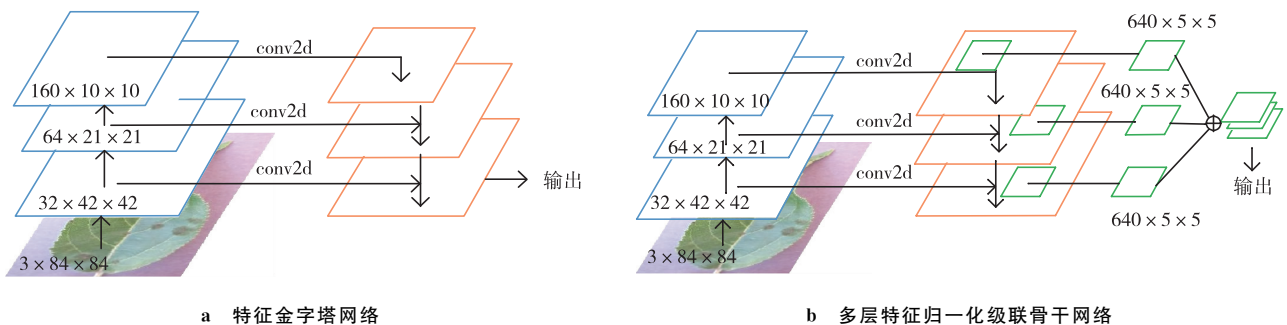


图 3 骨干网络对比

Fig. 3 Backbone network comparison

1.5 多分支分组卷积的特征级联农作物病害识别模型

综上所述,得到本文模型框架如图 4 所示。本文模型的特征提取阶段使用多分支分组卷积和多层特征归一化级联的骨干网络,也即 f_{φ} 部分,提升模型的鲁棒性,然后采用原型网络对样本进行识别,使用交叉熵损失计算模型损失,再反向传播给编码器部分。训练结束后,对模型进行知识迁移,使用测试集进行测试,最终得到识别结果。

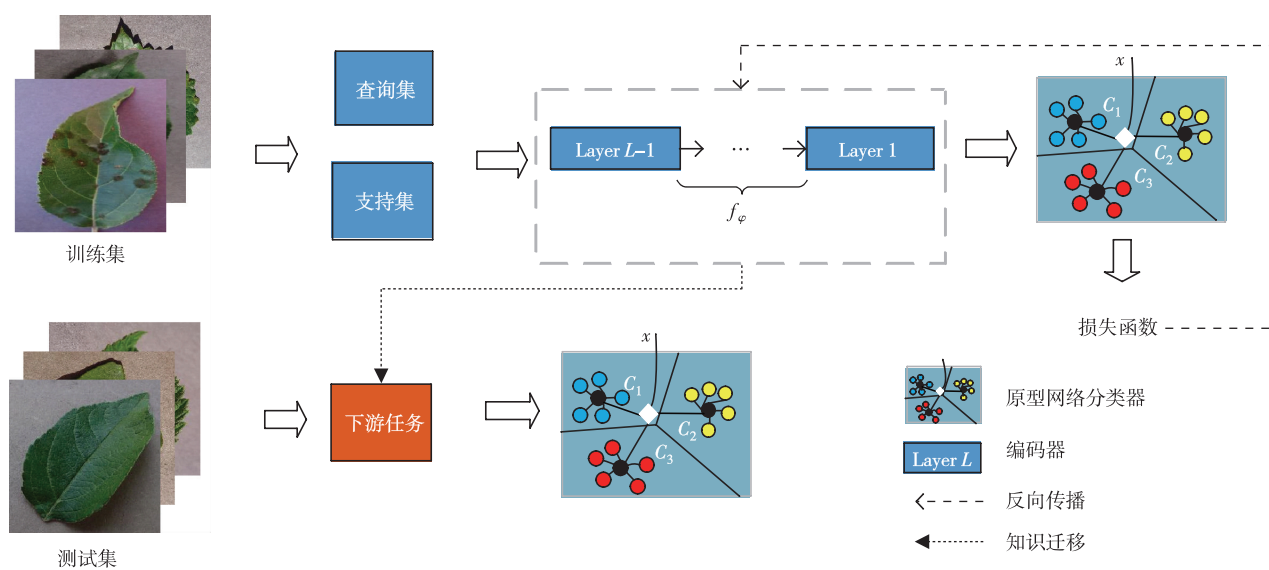


图 4 本文模型框架

Fig. 4 The framework of the proposed model

2 实验结果与分析

2.1 实验数据集

本文使用了 PlantVillage^[24]、miniImageNet^[13]、tieredImageNet^[25]与 CUB^[26]共 4 个数据集进行实验。其中,PlantVillage 数据集来源于 AI Challenge 2018 比赛,用于农作物病害识别任务。该数据集根据农作物不同的病害程度来划分种类,其中包含 14 种农作物(苹果、蓝莓、樱桃、玉米、葡萄、橘子、桃子、辣椒、土豆、树莓、黄豆、南瓜、草莓和番茄)的 38 种病害,共 54 306 张样本图像。

在深度学习中,样本数量越多,训练出来的模型效果越好,模型的泛化性越强。为了提升样本的多样性与质量,从而提高模型的泛化能力,本文对 PlantVillage 数据集进行了数据增强。图 5 显示了在训练数据中样本图像增强后的效果,其中图 5a 为原图,图 5b、图 5c、图 5d、图 5e、图 5f 和图 5g 分别对应将原数据集进行图像上下翻转、旋转、放大、PCA 颜色增强、伽马校正和噪声注入 6 种操作后得到的图像。

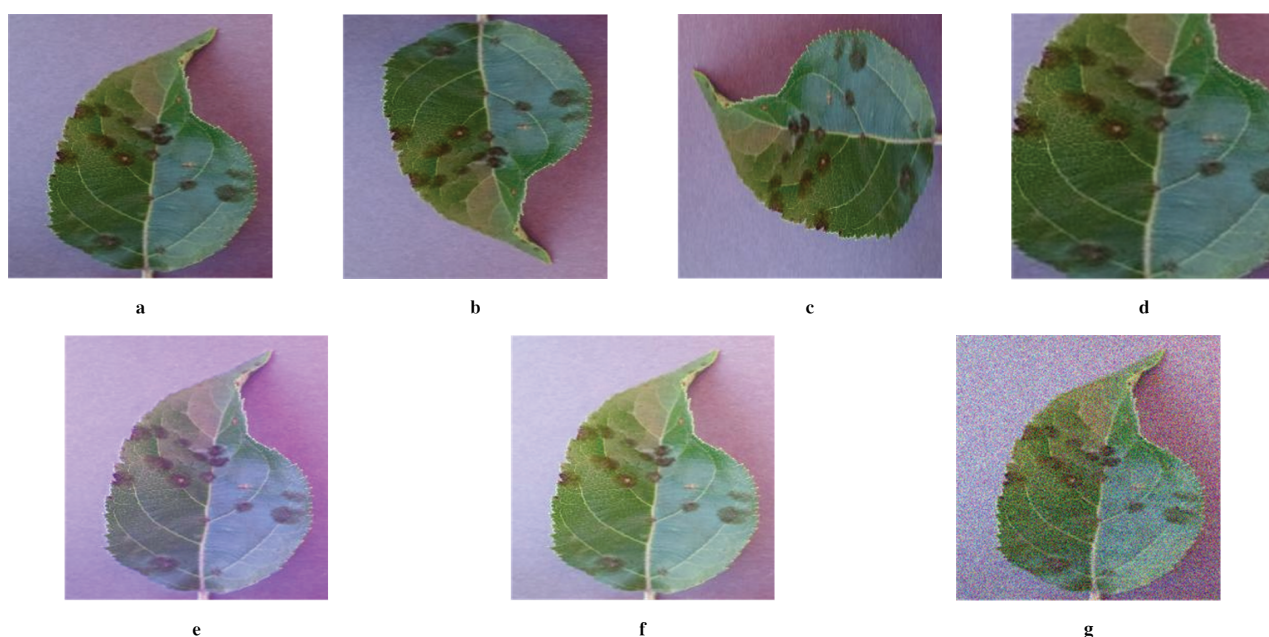


图 5 增强后的 PlantVillage 数据集

Fig. 5 The enhanced PlantVillage dataset

增强后的 PlantVillage 数据集包含 38 个不同类别的农作物叶片和 1 类无叶片的背景图,共计 61 486 张图片,详细信息见表 1。

表 1 PlantVillage 数据集信息
Tab. 1 Image information of PlantVillage dataset

农作物类别	类名称	图像数量	农作物类别	类名称	图像数量
苹果	疮痂病	1 000	土豆	早疫病	1 000
	黑腐病	1 000		晚疫病	1 000
	锈病	1 000		健康叶	1 000
蓝莓	健康叶	1 645	树莓	健康叶	1 000
	健康叶	1 502	黄豆	健康叶	5 090
樱桃	白粉病	1 052	南瓜	白粉病	1 835
	健康叶	1 000	草莓	叶烧病	1 181
玉米	灰斑病	1 000		健康叶	1 028
	锈病	1 192	细菌性斑点病	2 127	
	叶枯病	1 000	早疫病	1 000	
	健康叶	1 162	晚疫病	1 681	
葡萄	黑腐病	1 180	番茄	叶霉病	1 194
	黑痘病	1 383		斑枯病	1 747
	叶枯病	1 076		二斑叶螨病	1 608
	健康叶	1 000		轮斑病	2 392
橘子	黄龙病	5 507		花叶病	1 000
桃子	细菌性斑点病	2 297		黄曲叶病	4 416
	健康叶	1 000	健康叶	1 670	
辣椒	细菌性斑点病	1 000	无叶片的背景图	1 143	
	健康叶	1 478	总计	61 486	

本文采用小样本学习方法,模拟了元学习训练策略,即从 39 个类别的农作物叶片图像中每类随机抽取 N 张样本,组成 1 个包含 $39 \times N$ 张样本的模拟数据集。把该数据集划分为训练集、验证集和测试集。将数据集的病害样本随机取 19 类作为训练集,其中包含苹果疮痂病、苹果锈病、蓝莓健康叶、樱桃白粉病、玉米锈病、玉米叶枯病、葡萄黑痘病、葡萄叶枯病、桃子细菌性斑点、辣椒细菌性斑点、土豆早疫病、土豆晚疫病、黄豆健康叶、草莓健康叶、番茄细菌性斑点、番茄健康叶、番茄叶霉病、番茄二斑叶螨病和番茄花叶病。在余下的病害样本中再随机取 10 类组成验证集,其中包含苹果黑腐病、樱桃健康叶、玉米健康叶、葡萄健康叶、桃子健康叶、土豆健康叶、南瓜白粉病、番茄早疫病、番茄斑枯病和番茄黄曲叶病。测试集由剩余的 9 类病害和无叶片的背景图样本所组成,其中包含苹果健康叶、玉米灰斑病、葡萄黑腐病、橘子黄龙病、辣椒健康叶、树莓健康叶、草莓叶烧病、番茄晚疫病、番茄轮斑病和无叶片背景图。

为检测本文模型在不同粒度数据集上的识别能力,还选取了小样本学习中常见的 miniImageNet、tieredImageNet 和 CUB 这 3 个数据集,分别在 2 种训练策略下验证本文模型在小样本场景下的识别效果。其中,miniImageNet 数据集是 Google DeepMind 团队 Vinyals 等人^[14]从 ImageNet 数据集中提取出来的,包含 100 类共 60 000 张图片,其中每类有 600 个样本,每张图片的长、宽均为 84 像素。tieredImageNet 数据集与 miniImageNet 数据集类似,是 Ren 等人^[26]从 ImageNet 数据集中提取出来,不同的是,tieredImageNet 数据集是更大的子集,包含了 608 个类。CUB 数据集是加州理工学院于 2010 年提出鸟类数据集^[27],共有 11 788 张鸟类图像,包含 200 类鸟类子类。具体的划分方式见表 2。

表 2 不同数据集的划分方式

Tab. 2 The partition method on different datasets

数据集	变量	训练集	验证集	测试集
miniImageNet	类别数/个	64	16	20
	裁剪尺寸	84×84	84×84	84×84
	模型输入	(3,84,84)	(3,84,84)	(3,84,84)
tieredImageNet	类别数/个	351	97	160
	裁剪尺寸	84×84	84×84	84×84
	模型输入	(3,84,84)	(3,84,84)	(3,84,84)
CUB	类别数/个	100	50	50
	裁剪尺寸	84×84	84×84	84×84
	模型输入	(3,84,84)	(3,84,84)	(3,84,84)

注:裁剪尺寸以图像长、宽的像素值计。

2.2 实验结果与分析

2.2.1 实验环境与参数设置

本文实验所用计算机的操作系统为 Windows 10 20H2 专业版 64 位,搭载了 AMD Ryzen 5 3600xCPU, 32 GB 内存,显卡为 Nvidia GeForce RTX 1050ti,显存为 4 GB;采用的模型基于 Pytorch(version 1.8.0),并配置了 CUDA(version 10)加速。

采用本文提出的多分支分组卷积的特征级联农作物病害识别模型来高效提取病害特征。模型输入图片的尺寸为 3×84×84,输出特征图的尺寸为 640×50×50。在训练时采用 5-way 1-shot 和 5-way 5-shot 这 2 种方式,共实验 6 000 轮,在元训练阶段,每个轮次包含 200 轮元训练任务和 240 轮元验证任务。每个任务包含 $N \times K$ 个支持集样本与 $N \times 15$ 个查询集样本。使用随机梯度下降优化器对网络进行参数调整,初始学习率为 0.025,模型的权重衰减系数为 0.000 5。训练过程中,保存元验证任务中识别准确率最高的一轮模型参数,直至训练结束。在元测试阶段,从测试集样本中随机采样组成 1 000 个元任务,用于测试模型的识别准确率。

2.2.2 骨干网络参数分析

为验证本文框架在模型参数量、检测速度等多个参数上的优势,在不同的骨干网络上统一使用原型网络在 PlantVillage 数据集上进行 5-way 5-shot 实验,选取每秒浮点运算次数、参数量、单张图片处理时间和识别准确率作为指标进行分析,对比结果如表 3 所示。从 Conv4 到 ResNet12,随着网络层数的加深,骨干网络的浮点运算次数与参数量都在增加,但是准确率提升却不大,图片的处理时间也在增加。这是由于简单加深网络层数虽然能提高模型性能,但易导致过拟合现象。本文在 ResNet12 的基础上对特征图进行归一化级联操作,与 ResNet10 相比,参数量和图片处理时间分别降低了 2.19 M 和 0.001 s,但网络的每秒浮点运算次数仅增加了 0.464 G,准确率也超过其他模型,达到了 93.25%。说明本文模型通过减少网络参数量,提升了检测速度,同时在保持高效性的基础上,提高了病害的识别准确率。

表 3 骨干网络参数对比

Tab. 3 Backbone comparison of parameters

骨干网络类型	每秒浮点运算次数	参数量	单张图片处理时间/s	识别准确率
Conv4	0.865 G	2.836 M	0.011	(78.20±0.66)%
ResNet10	1.166 G	7.879 M	0.055	(87.58±0.78)%
ResNet12	3.518 G	12.424 M	0.089	(88.93±0.70)%
本文骨干网络	1.630 G	5.689 M	0.054	(93.25±0.21)%

2.2.3 多层特征级联层数分析

在小样本学习中,模型的性能易受骨干网络提取到的浅层与深层语义信息的影响。Yang 等人^[28]发现:浅层

网络的感受野较小,感受野的重叠区域也较小;浅层网络捕获的细节更多,分辨率更高,但由于它的卷积次数较少,导致浅层网络语义性偏低,且容易产生比深层网络更多的噪声。相比之下,深层网络提取的特征靠近输出层,主要包含图像整体性和粗粒度的语义信息。这是由于感受野的扩大以及感受野之间重叠区域的增加导致了深层网络的特征分辨率较低、对细节的感知能力较弱。为探究骨干网络层数对网络识别准确率的影响,通过级联不同深度的特征以充分利用各层特征,在 PlantVillage 数据集上使用原型网络并采用 5-way 5-shot 方式进行实验,对比不同层数的卷积层效果。实验结果见表 4。

表 4 特征级联层数对实验性能的影响

Tab. 4 Influence of the number of characteristic cascade layers on experimental performance

特征级联层数	5-way 5-shot	特征级联层数	5-way 5-shot
1	(88.79±0.23)%	3	(93.25±0.21)%
2	(90.42±0.45)%	4	(90.69±0.39)%

由表 4 可知,使用单层卷积进行特征提取的模型性能明显低于多层卷积层级联的模型。然而,随着级联层数的增加,模型性能并未得到持续提升。实验结果表明,采用 3 层卷积层级联能够获得最佳效果,此时识别准确率达到最高值。因此,本文选择使用 3 层卷积层级联的方式来输出特征。

2.2.4 收敛曲线分析

为验证本文模型的收敛性,以 5-way 5-shot 在 PlantVillage 数据集上元训练与元验证的实验结果作为对比,以 PN 作为基准模型,分别对本文模型与 PN 绘制收敛曲线,实验结果见图 6 和图 7。

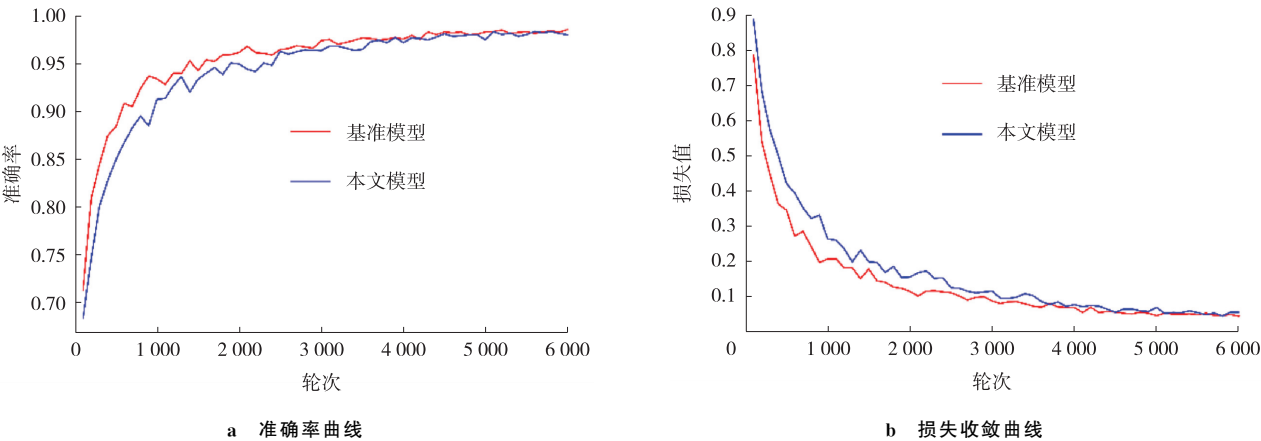


图 6 元训练收敛曲线对比
Fig. 6 Comparison of convergence curves for meta-training

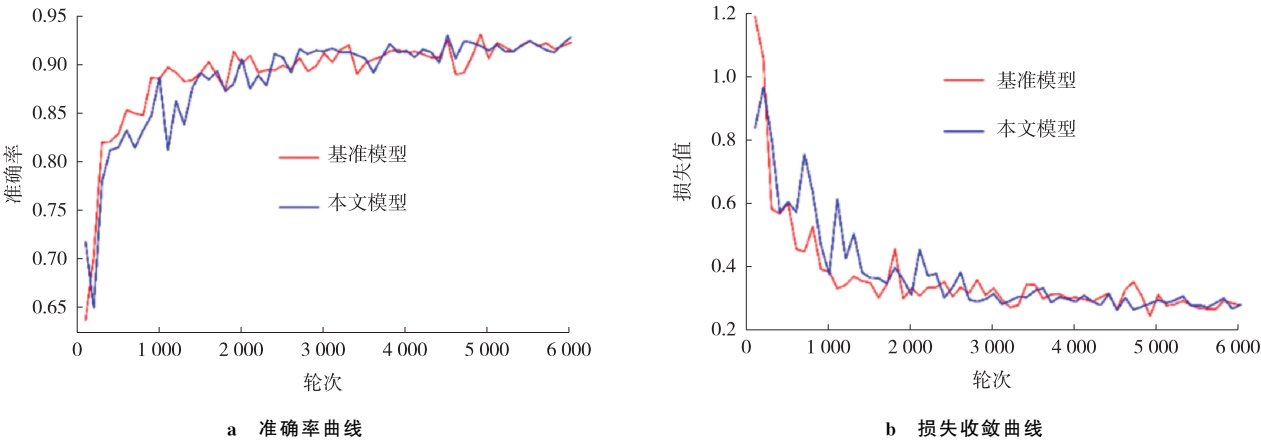


图 7 元验证收敛曲线对比
Fig. 7 Comparison of convergence curves for meta-validation

由图 6a 可知,尽管 PN 在元训练阶段时收敛速度比较快,且在大部分训练过程中识别准确率都大于本文模型,但随着训练轮次的增加,本文模型通过多层特征级联的方式获取了更多图像细节与语义信息,识别准确率逐渐逼近直至超过 PN。最终,本文模型的准确率达到 90%左右,而 PN 的准确率仅有 87%左右。此外,在小样本学习的元验证阶段,由于选择的样本未出现在元训练中,过高的识别准确率可能会降低模型的迁移能力,从而影响元验证的识别效果。图 6b 显示,本文模型损失最终下降至 0.06。

由于本文使用的 PlantVillage 数据集是一个细粒度的数据集,样本间的差异较小,若在元验证过程中单纯提高模型对样本的识别能力或简单增加网络深度,容易导致过拟合及梯度消失等问题。本文模型采用了多分支分组卷积残差块,在前向传播中通过引入恒等映射,有效避免了网络退化;在反向传播中,使得误差信息能够不经中间权重矩阵变换直接传递至低层,从而解决了梯度消失问题。因此,尽管本文模型的参数量和网络深度有所增加,但并未产生过拟合和梯度消失现象。最终,在收敛时,本文模型的元验证识别准确率和损失值分别为 92.71%和 0.273 8,与 PN 的这两个参数值(分别为 92.15%和 0.273 9)相差较小(图 7)。

2.2.5 消融实验分析

为验证本文所采用网络在提升模型识别能力方面的有效性,本节设计了 1 组消融实验,以 PN 作为基准模型进行对比。实验结果见表 5。

表 5 消融实验结果
Tab. 5 Result of ablation experiment

模型	5-way 1-shot	5-way 5-shot
PN	(74.15±0.98)%	(87.58±0.78)%
PN+多分支分组卷积残差块	(75.77±0.33)%	(88.79±0.23)%
PN+多层特征归一化级联	(76.76±0.52)%	(88.65±0.43)%
本文模型	(79.68±0.35)%	(93.25±0.21)%

由表 5 可知,单独为 PN 加入多分支分组卷积残差结构或多层特征归一化级联骨干网络均能提高模型性能。单独加入多分支分组卷积残差块时,模型在 5-way 1-shot 和 5-way 5-shot 任务中的识别准确率相较 PN 分别提升了 1.62%和 1.21%。提升幅度较小主要因为该模块的作用集中于降低模型参数量、缓解网络退化和梯度消失问题;单独加入多层特征归一化级联骨干网络时,模型在 5-way 1-shot 和 5-way 5-shot 任务中的识别准确率分别提升了 2.61%和 1.07%,表明多层特征归一化级联骨干网络明显增强了模型对不同大小和尺度信息的提取能力。对比发现,在 5-way 1-shot 场景下,多层特征归一化级联骨干网络的提升幅度优于多分支分组卷积残差块,说明该网络在小样本学习中对模型性能提升更为明显;而在 5-way 5-shot 场景下,多分支分组卷积残差块的提升幅度略高于多层特征归一化级联骨干网络,但总体提升程度接近。同时加入 2 种方法时,模型在 5-way 1-shot 和 5-way 5-shot 任务中的识别准确率分别提高了 5.53%和 5.67%,较单独使用任一方法的提升效果更加明显,表明 2 种方法的结合在提升模型性能方面具有协同增效作用。

2.2.6 识别实验结果与分析

为验证本文模型在不同粒度数据集上的识别效果,进行了如下实验。首先在 PlantVillage 细粒度数据集上进行了实验,将本文模型与 3 种小样本学习模型在不同的骨干网络上的表现进行了对比,模型性能的评价指标采用识别准确率。实验结果如表 6 所示。

由表 6 可知,本文模型与 3 种小样本学习模型 MN、RN 与 PN 在 5-way 1-shot 和 5-way 5-shot 的任务上相比,本文模型均取得了最优值,相较于 PN 中的最优值,本文模型在这 2 项任务上分别提升了 3.74%和 4.32%。通过分析对比模型的实验结果可以发现,简单地加深网络的深度并不总能带来更好的效果,反而可能降低模型的识别能力。这表明,本文模型在 PlantVillage 数据集上能够实现更优的识别效果。

表 6 PlantVillage 数据集上各模型性能对比结果

Tab. 6 Performance comparisons of the models on plantvillage dataset

模型	骨干网络类型	5-way 1-shot	5-way 5-shot
MN ^[14]	Conv4	(73.13±0.73)%	(88.91±0.43)%
	ResNet10	(72.81±0.77)%	(87.44±0.45)%
	ResNet18	(72.11±0.81)%	(88.09±0.44)%
	ResNet34	(72.15±0.81)%	(86.55±0.47)%
RN ^[13]	Conv4	(73.01±0.89)%	(85.82±0.47)%
	ResNet10	(65.32±0.86)%	(83.15±0.54)%
	ResNet18	(68.86±0.87)%	(82.12±0.53)%
	ResNet34	(68.70±0.90)%	(80.77±0.60)%
PN ^[19]	Conv4	(72.25±0.76)%	(88.30±0.49)%
	ResNet10	(74.15±0.98)%	(87.58±0.78)%
	ResNet12	(75.94±0.87)%	(88.93±0.70)%
	ResNet18	(72.11±0.81)%	(88.09±0.44)%
本文模型	ResNeXt12	(79.68±0.35)%	(93.25±0.21)%

为进一步探讨本文模型在粗粒度数据集上的表现,选择在 minImageNet 数据集和 tieredImageNet 数据集上分别进行实验。这 2 个数据集分别具有较大的类间差异和更丰富的类样本分布,是常用于评估小样本学习模型在更复杂、更加多样化的环境下的性能。本文模型与其他 7 种小样本学习模型的对比实验结果见表 7 和表 8。

分析表 7 和表 8 可知,本文模型在处理 miniImageNet 和 tieredImageNet 这类粗粒度数据集时,相较于其他主流模型,识别准确率虽然有所提升,但均未取得最优值。这是由于在训练过程中,经常会遇到源域与目标域的数据分布不同,存在明显的差异,而当输入数据的域发生了变化,目标域的标签保持不变时,就会造成本文模型较难适应目标域中的新任务。

为验证本文模型在其他细粒度数据集上的效果,在 CUB 数据集上进行了实验,实验结果见表 9。由表 9 可知,在 CUB 数据集上,本文模型取得了比其他模型更优的性能。

表 7 miniImageNet 数据集上各模型性能对比结果

Tab. 7 Performance comparisons of the models on miniimagenet dataset

模型	骨干网络类型	5-way 1-shot	5-way 5-shot
SNAIL ^[29]	ResNet12	(55.71±0.99)%	(68.88±0.92)%
AdaResNet ^[30]	ResNet12	(56.88±0.62)%	(71.94±0.57)%
RN	ResNet12	(57.02±0.92)%	(71.07±0.69)%
K-tuplet Nets ^[31]	ResNet12	(58.30±0.84)%	(72.37±0.63)%
INNA ^[32]	ResNet10	(54.95±0.30)%	(63.04±0.30)%
PNCL ^[29]	WRN-28-10	(59.54±0.47)%	(74.46±0.52)%
PN	ResNet10	(54.83±0.52)%	(70.35±0.30)%
本文模型	ResNeXt12	(56.97±0.18)%	(72.59±0.33)%

表 8 tieredImageNet 数据集上各模型性能对比结果

Tab. 8 Performance comparisons of the models on tieredimagenet dataset

模型	骨干网络类型	5-way 1-shot	5-way 5-shot
MAML ^[33]	Conv4	(51.67±1.81)%	(70.30±1.75)%
RN	Conv4	(54.48±0.93)%	(71.32±0.78)%
Cosine classifier ^[34]	ResNet12	(55.43±0.81)%	(77.18±0.61)%
E ³ BM ^[35]	ResNet12	(63.80±0.40)%	(80.10±0.30)%
CTM ^[36]	ResNet18	(64.78±0.11)%	(81.05±0.52)%
MetaOptNet ^[37]	ResNet12	(65.99±0.72)%	(81.56±0.53)%
PN	ResNet10	(56.39±1.00)%	(70.14±0.34)%
本文模型	ResNeXt12	(64.77±0.42)%	(81.40±0.41)%

表 9 CUB 数据集上各模型性能对比结果

Tab. 9 Performance comparisons of the models on CUB dataset

模型	骨干网络类型	5-way 1-shot	5-way 5-shot
MN	ResNet12	(71.87±0.85)%	85.08±0.57)%
Cosine classifier	ResNet12	(67.30±0.86)%	84.75±0.60)%
TOAN ^[38]	ResNet12	(67.17±0.81)%	(82.09±0.56)%
FEAT ^[39]	ResNet12	(68.87±0.22)%	(82.90±0.15)%
RN	ResNet34	(66.20±0.99)%	(82.30±0.58)%
UDS ^[40]	ResNet34	(72.23±0.88)%	(86.57±0.50)%
PN	ResNet10	(71.94±0.45)%	(85.03±0.51)%
本文模型	ResNeXt12	(74.03±0.16)%	(87.37±0.16)%

综上所述,本文模型在处理 PlantVillage 与 CUB 这类细粒度数据集时,表现出较强的识别能力。在 PlantVillage 数据集与 CUB 数据集上,模型能够有效地捕捉到不同农作物病害与鸟类物种的细微差异,模型性能明显优于其他小样本学习模型。相较之下,本文模型在面对 minImageNet 数据集和 tieredImageNet 数据集这类粗粒度数据集时,模型的表现则较为一般。这 2 个数据集具有较大的类间差异和更丰富的类别样本分布,虽然本文模型在处理此类数据集时仍保持了一定的准确性,但性能未能达到在细粒度数据集上的水平。这一差异表明,尽管模型在处理细粒度数据集时具有较强的表现力,但在类间差异较大的粗粒度数据集上,仍有进一步提升的空间。

3 结论

农作物种植过程中,由于部分病害的发生率较低,现有图像样本数量有限,不适合直接使用深度网络模型进行病害识别,而使用小样本学习方法进行训练,又容易出现过拟合、梯度消失以及过度依赖骨干网络提取信息能力等问题。为解决这些问题,本文提出了一种多分支分组卷积的特征级联农作物病害识别模型。其中多分支分组卷积残差块在网络深度加深时,有效降低了网络参数量、缓解梯度消失现象;多层特征归一化级联的骨干网络则有助于挖掘网络不同层次的语义信息,保留高分辨率的语义信息以及更多位置和细节的信息。本文在 PlantVillage 数据集上进行了骨干网络参数分析、多层特征级联层数分析、收敛曲线分析以及消融实验分析,结果表明本文模型在农作物病害图像样本量较少的情况下具有较高的识别有效性。

本研究仍存在需进一步探讨和完善之处。首先,PlantVillage 数据集是一个细粒度的数据集,主要包含各类

农作物的叶片图像,域间隙较小。然而,本文所提出的模型在 miniImageNet、tieredImageNet 等具有较大域间隙的数据集上并未获得理想的效果。其次,在不同任务中学习迁移知识的过程中,梯度下降速度较慢,导致模型在迁移至从未接触的新任务时难以快速收敛。未来的研究将继续深入探索不同类别之间的域间隙问题,设计能够快速调整并适应不同任务的基础模型,以实现跨域任务的有效迁移。

参考文献:

- [1] 邢献芳,王玉彩. 人工智能背景下的智慧生态农业发展模式:评《智能农业—智能时代的农业生产方式变革》[J]. 热带作物学报,2021,42(1):329.
XING X F, WANG Y C. The development model of intelligent ecological agriculture under the background of artificial intelligence:Comments on intelligent agriculture;the transformation of agricultural production mode in the intelligent age[J]. Chinese Journal of Tropical Crops,2021,42(1):329.
- [2] 王东方,汪军. 基于迁移学习和残差网络的农作物病害分类[J]. 农业工程学报,2021,37(4):199-207.
WANG D F, WANG J. Crop disease classification with transfer learning and residual networks[J]. Transactions of the Chinese Society of Agricultural Engineering,2021,37(4):199-207.
- [3] RAJ A P S S, VAJRARELU S K. DDLA: dual deep learning architecture for classification of plant species[J]. IET Image Processing,2019,13(12):2176-2182.
- [4] RUSSEL N S, SELVARAJ A. Leaf species and disease classification using multiscale parallel deep CNN architecture[J]. Neural Computing and Applications,2022,34(21):19217-19237.
- [5] LIU J, WANG X W. Plant diseases and pests detection based on deep learning;a review[J]. Plant Methods,2021,17:22.
- [6] LI Y, NIE J, CHAO X W, et al. Do we really need deep CNN for plant diseases identification?[J]. Computers and Electronics in Agriculture,2020,178:105803.
- [7] JANKOWSKI N, DUCH W, GRABCEWSKI K. Meta-learning in computational intelligence[M]. Berlin: Springer, 2011: 97-115.
- [8] LAKE B M, SALAKHUTDINOV R, TENENBAUM J B. One-shot learning by inverting a compositional causal process[C]//BURGES C J C, BOTTOU L, WELLING M, et al. NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2013: 2526-2534.
- [9] ZHONG F M, CHEN Z K, ZHANG Y C, et al. Zero- and few-shot learning for diseases recognition of Citrus aurantium L. using conditional adversarial autoencoders[J]. Computers and Electronics in Agriculture,2020,179:105828.
- [10] CHEN L Z, CUI X H, LI W. Meta-learning for few-shot plant disease detection[J]. Foods,2021,10(10):2441.
- [11] LI Y, YANG J C. Meta-learning baselines and database for few-shot classification in agriculture[J]. Computers and Electronics in Agriculture,2021,182:106055.
- [12] 任胜男,孙钰,张海燕,等. 基于 one-shot 学习的小样本植物病害识别[J]. 江苏农业学报,2019,35(5):1061-1067.
REN S N, SUN Y, ZHANG H Y, et al. Plant disease identification for small sample based on one-shot learning[J]. Jiangsu Journal of Agricultural Sciences,2019,35(5):1061-1067.
- [13] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: relation network for few-shot learning[C]//O'CONNER L. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1199-1208.
- [14] VINVALS O, BLUNDELL C, LILLCRAP T, et al. Matching networks for one shot learning[J]. Advances in Neural Information Processing Systems,2016,29:3630-3638.
- [15] ARGÜESO D, PICON A, IRUSTA U, et al. Few-shot learning approach for plant disease classification using images taken in the field[J]. Computers and Electronics in Agriculture, 2020,175:105542.
- [16] 吴昊昱. 基于小样本学习的自然场景图像中茶叶病害识别[D]. 合肥:安徽大学,2020.
WU H Y. Identification of tea leaf's diseases in natural scene images based on low shot learning[D]. Hefei: Anhui University, 2020.
- [17] DHIVYAA C R, KANDASAMY N, RAIENDRAN S. Integration of dilated convolution with residual dense block network and multilevel feature detection network for cassava plant leaf disease identification[J]. Concurrency and Computation: Practice and Experience,2022,34(11):e6879.

- [18] PAN J, XIA L, WU Q, et al. Automatic strawberry leaf scorch severity estimation via faster R-CNN and few-shot learning[J]. *Ecological Informatics*, 2022, 70: 101706.
- [19] SNELL J, SWERSKY K, ZEMEL R. Prototypical network for few-shot learning[J]. *Advances in neural information processing systems*, 2017, 30: 4077-4087.
- [20] LI Z, OUYANG F, ZHOU C, et al. Few-Shot relation classification research based on prototypical network and causal intervention[J]. *IEEE Access*, 2022, 10: 36995-37002.
- [21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//O'CONNER L. *Proceedings of 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2016: 770-778.
- [22] CHEN F, WEI J S, XUE B, et al. Feature fusion and kernel selective in Inception-v4 network[J]. *Applied Soft Computing*, 2022, 119: 108582.
- [23] XIE S N, GIRSHICK R, DOLLÁR P, et al. A ggregated residual transformations for deep neural networks[C]//O'CONNER L. *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2017: 1492-1500.
- [24] ZUO Z, TONG X Z, WEI J Y, et al. AFFPN: attention fusion feature pyramid network for small infrared target detection[J]. *Remote Sensing*, 2022, 14(14): 3412.
- [25] LI Z M, OUYANG F F, ZHOU C L, et al. Few-shot relation classification research based on prototypical network and causal intervention[J]. *IEEE Access*, 2022, 10: 36995-37002.
- [26] REN M Y, TRIANTAFILLOU E, RAVI S, et al. Meta-learning for semi-supervised few-shot classification[EB/OL]. (2018-03-02)[2023-02-24]. <https://arxiv.org/abs/1803.00676>.
- [27] WAH C, BRANSON S, WELINDER P, et al. The caltech-UCSD birds-200-2011 dataset[R]. California: California Institute of Technology, 2011.
- [28] YANG D W, DU Y, YAO H L, et al. Image semantic segmentation with hierarchical feature fusion based on deep neural network[J]. *Connection Science*, 2022, 34(1): 1772-1784.
- [29] LUO X, CHEN Y X, WEN L J, et al. Boosting few-shot classification with view-learnable contrastive learning[C]//O'CONNER L. *Proceedings of 2021 IEEE Conference on Multimedia and Expo*. Piscataway: IEEE, 2021: 1-6.
- [30] MUNKHDALAI T, YUAN X D, MEHRI S, et al. Rapid adaptation with conditionally shifted neurons[C]//DY J G, KRAUSE A. *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018: 3661-3670.
- [31] LI X M, YU L Q, FU C-W, et al. Revisiting metric learning for few-shot image classification[J]. *Neurocomputing*, 2020, 406: 49-58.
- [32] REN M Y, LIAO R J, FETAYA E, et al. Incremental few-shot learning with attention attractor networks[M]//KAYEJO S, MOHAMED S, AGARWAL A, et al. *Advances in Neural Information Processing Systems 35*. San Mateo: Morgan Kaufmann Publishers, 2019: 5276-5286.
- [33] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//PRECUP D, TEH Y W. *Proceedings of the 34th International Conference on Machine Learning*. Cambridge: MIT Press, 2017: 1126-1135.
- [34] CHEN W Y, LIU Y C, KIRA Z, et al. A closer look at few-shot classification[EB/OL]. (2020-01-12)[2023-02-24]. <https://arxiv.org/abs/1904.04232>.
- [35] LIU Y, SCHIELE B, SUN Q. An ensemble of epoch-wise empirical bayes for few-shot learning[C]//VEDALDI A, BISCHOF H, BROX T, et al. *Proceedings of the European Conference on Computer Vision*. Berlin: Springer, 2020: 404-421.
- [36] LI H Y, EIGEN D, DODGE S, et al. Finding task-relevant features for few-shot learning by category traversal[C]//O'CONNER L. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2019: 1-10.
- [37] LEE K, MAJI S, RAVICHANDRAN A, et al. Meta-learning with differentiable convex optimization[C]//O'CONNER L. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2019: 10649-10657.
- [38] HUANG H X, ZHANG J J, YU L T, et al. TOAN: target-oriented alignment network for fine-grained image categorization with few labeled samples[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(2): 853-866.
- [39] YE H J, HU H X, ZHAN D C, et al. Few-shot learning via embedding adaptation with set-to-set functions[C]//O'CONNER

- L. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE, 2020: 8805-8814.
- [40] HU Z P, LI Z J, WANG X Y, et al. Unsupervised descriptor selection based meta-learning networks for few-shot classification [J]. Pattern Recognition, 2022, 122: 108304.

Multi-Branch Grouping Convolution Feature Cascaded Crop Disease Recognition

LÜ Jia^{1,2}, PENG Gangjian^{1,2}, WU Ruoyu^{1,2}

(1. School of Computer and Information Science, Chongqing Normal University;

2. Chongqing Digital Agriculture Service Engineering Technology Research Center, Chongqing 401331, China)

Abstract: Aiming at the problems of the high number of crop disease recognition parameters and gradient disappearance of the network in the existing few-shot learning, a multi-branch grouping convolution feature cascade crop disease recognition model is proposed. Firstly, the model applies grouped convolution to the residual block, which is used to alleviate the gradient disappearance problem of the network. Secondly, the input of the network is divided into isomorphic multi-branch embeddings, and more abundant features are obtained and the number of network parameters is reduced through the feature supplementation between different branches. Finally, a multi-layer feature cascade backbone network is designed to mine different levels of information in the image, and more location and detail information in the shallow features and stronger semantic information in the deep features were cascaded out to improve the generalization ability of the network. On the PlantVillage dataset, the accuracy of the model on the 5-way 1-shot and 5-way 5-shot tasks reached $(79.68 \pm 0.35)\%$ and $(93.25 \pm 0.21)\%$, respectively, which were 3.74% and 4.32% higher than the prototype network. The experimental results show that the proposed model can better identify crop diseases, and reduce the number of network parameters while deepening the depth of the network, alleviating the gradient disappearance and mining different levels of semantic information, which effectively solves the problem of few-shot learning of crop diseases.

Keywords: crop disease; image recognition; few-shot learning; grouped convolution; feature cascade

(责任编辑 黄 颖)