

卢晓蒙, 张明. 基于深度残差网络和注意力机制的植物识别方法[J]. 智能计算机与应用, 2024, 14(12): 151-156. DOI: 10.20169/j. issn. 2095-2163. 241221

基于深度残差网络和注意力机制的植物识别方法

卢晓蒙, 张明

(江苏科技大学 计算机学院, 江苏 镇江 212003)

摘要: 本文提出了一种基于注意力机制和 BN-Activation-Conv 结构的植物识别模型, 旨在解决传统植物识别过程中提取特征步骤复杂、耗时长、精度低等问题。本文模型在残差网络中引入 SE-Net 模块, 提高模型识别准确率, 并将传统卷积神经网络常用的 Conv-BN-Activation 结构替换为 BN-Activation-Conv 结构, 加速了模型收敛。此外, 本文还使用迁移学习的训练方法加速训练过程。实验表明, 该模型在识别精度和速度方面表现均较好。

关键词: 植物图像识别; 残差网络; 特征提取; 注意力机制; 迁移学习

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2024)12-0151-06

Plant recognition method based on deep residual network and attention mechanism

LU Xiaomeng, ZHANG Ming

(School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212003, Jiangsu, Chian)

Abstract: This article presents a plant recognition model, utilizing attention mechanism and the BN-Activation-Conv structure, to address the difficulties of intricate feature extraction steps, prolonged time consumption, and low precision in traditional plant recognition procedures. This paper introduces the SE-Net module to the residual network, thereby enhancing the precision of model recognition. The Conv-BN-Activation structure, which is typically employed in the classic Convolutional neural network, is supplanted by the BN-Activation-Conv structure. Consequently, model convergence is accelerated. The transfer learning training method is employed in this paper to expedite the training process. The experiment reveals that the model is well in terms of recognition accuracy and speed.

Key words: plant image recognition; residual network; feature extraction; attention mechanism; transfer learning

0 引言

中国植物种类资源丰富^[1]。除了常见的农作物与观赏性绿植外, 野生植物也是大自然生态链的重要组成部分, 对于保护自然环境和推动经济持续发展都起着至关重要的作用^[2]。近年来, 由于国内野生植物品种的多样化和公众对其了解的不足, 许多植物已然面临濒危的境地, 因而亟需采取有效的保护措施。为了提高植物保护的效果, 就需要解决植物物种识别工作主要依赖于人工识别、效率低和精度差等问题。

深度学习技术逐渐被应用到各个领域^[3-4]。1993年, Guyer 等学者^[5]提取了 17 种叶片形状特征, 并对 40 类植物进行分组, 取得了良好效果。孟

亮等学者^[6]、曹香滢等学者^[7]、于慧伶等学者^[8]通过卷积神经网络对多种植物图像进行识别, 取得了可观的研究成果。目前, 基于卷积神经网络的植物分类方法在识别效率和精度方面仍有待提升。这一状况可能是由于植物的形态多样性和图像复杂性等因素所导致。为了提高野生植物分类的准确性和效率, 研究人员正在尝试结合多种先进的图像处理技术和机器学习算法, 以进一步提升植物图像分类的性能。这些措施将会为野生植物保护提供更好的支持和保障。

基于此, 本文提出一种将 SE-Net 模块和 BN-Activation-Conv 结构相融合的植物识别模型。首先, 在残差网络中引入 SE-Net 模块来加强有效特征的权重, 降低输入数据中的噪声。接下来, 将传统

作者简介: 卢晓蒙(1986—), 男, 硕士研究生, 主要研究方向: 计算机视觉, 模式识别。Email: 674633393@qq.com; 张明(1978—), 男, 教授, 主要研究方向: 人工智能与模式识别, 机器学习与深度学习, 嵌入式智能终端硬件/软件开发, 人工智能及其应用的软硬件集成开发等。

收稿日期: 2023-09-05

哈尔滨工业大学主办 ◆ 专题设计与应用

卷积神经网络中的 Conv-BN-Activation 结构替换为 BN-Activation-Conv 结构,加快模型收敛速度。

1 基础网络模型

传统卷积神经网络 (Convolutional Neural Networks, CNN) 在面临网络模型层数不断增加的情况时,网络的性能常常出现不增加反降的情况、即网络退化,残差网络^[9-10]可以很好地解决这一问题。

残差网络和传统卷积网络相比结构中添加跨层恒等路径,使得残差映射趋近于 0,保留了恒等映射。通过这种方式使模型不会随着网络加深而在性能表现上变差。残差网络往往包括多个残差模块,在残差模块内,恒等映射和残差映射被用于构建网络结构,其中残差映射表示对输入进行一定处理的映射,而恒等映射则表示不对输入进行任何变换的映射。当残差映射接近 0 时,网络性能最佳,接下来如果继续增加网络层数,网络性能也将保持在最佳水平。图 1 展示了残差块的结构。不妨设最优解为 $H(x) = x$, 残差映射是指映射 $H(x)$ 和 x 的差值,用 $F(x)$ 表示,即:

$$F(x) = H(x) - x \quad (1)$$

当 $F(x) = H(x) - x$ 无限接近于 0 时,网络达到最佳状态,继续加深网络深度,网络也将一直处于最佳性能。当残差块的输入为 x_n 时,可得计算后的输出为:

$$x_{n+1} = f(x_n + F(x_n, W_n)) \quad (2)$$

其中, $F(\cdot)$ 表示残差映射; W_n 表示其对应的权值; $f(\cdot)$ 表示激活函数。由图 1 可见,不同残差块之间可能存在维度不匹配的情况,此时只需对恒等映射 x_n 进行线性变换 W_s 即可得到:

$$X_{n+1} = f(W_s x_n + F(x_n, W_n)) \quad (3)$$

其中, W_s 表示权重参数。

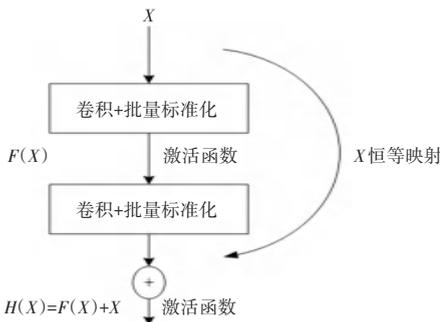


图 1 残差块结构示意图

Fig. 1 Schematic diagram of residual block structure

接下来,本文选取 ResNet-18 作为基础网络进

行改进。ResNet-18 模型如图 2 所示。

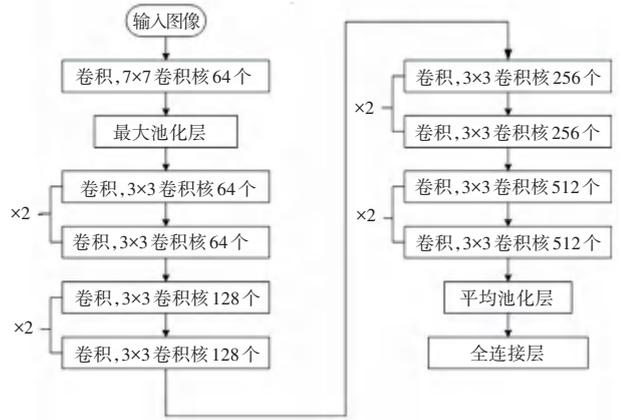


图 2 ResNet-18 模型示意图

Fig. 2 Schematic diagram of ResNet-18 model

2 网络模型构建与优化

2.1 注意力机制

SE-Net 模块作为注意力机制的常见算法之一,在各种深度学习任务中已被广泛采用^[11]。SE-Net 的核心功能是通过挤压 (Squeeze) 和激励 (Excitation) 两个方法来处理数据。挤压方法是通过输入特征图进行全局平均池化,从而获得每个通道的空间信息。而激励方法则是通过使用 2 个完整的连接层来学习通道之间的相互依存关系。这两个操作共同帮助 SENet 自适应地调整通道间的关系,进而提高卷积神经网络的性能^[12]。

SE-Net 的关键贡献在于可以自动评估并赋予不同信道不同的权重,这有助于模型捕获更多对识别任务有益的信息,抑制噪声的影响^[13],有效地提高了模型的识别性能。结合 SE-Net 模块的残差模块示意如图 3 所示。

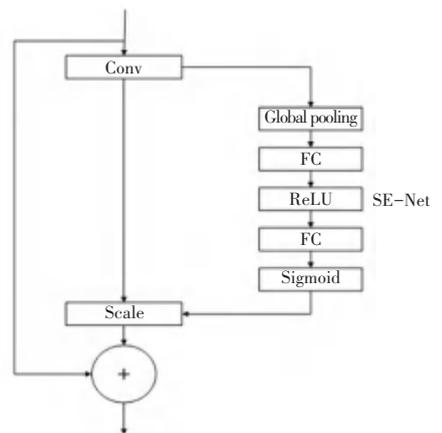


图 3 加入注意力机制的残差块结构

Fig. 3 Residual block diagram with added attention mechanism

2.2 迁移学习

迁移学习是一种在卷积神经网络中常用的技术,这是通过利用在原任务上学到的知识(特征表示和参数)来加速新任务学习过程实现的,即针对新任务的特点对先期已经在大型数据集上进行预训练操作的网络模型进行调整。这往往包括冻结模型的底层,因为底层学到的特征较为通用,替换顶层,并在新任务的数据集上训练顶层。这样可以在较短时间内获得针对新任务的高性能模型。

本文采用改进后的 ResNet18 为基础构架,将在 ImageNet 数据集上预训练得到的参数应用到一个新的网络中,以便在较小数据集上进行训练。这一策略的优势在于利用了在大规模网络上学到的特征参数,同时规避了直接在小数据集上进行模型训练所可能遇到的挑战,从而有助于提高模型在分类任务上的准确性。

2.3 网络结构的改进

2.3.1 模型调整

传统残差网络 ResNet18 的流程顺序先后为卷积层、批量标准化层、激活函数层。本文对网络结构进行调整,使输入数据先经过批量标准化层,再进入激活函数层,最后进入卷积层。这种调整通常称为 BN-Activation-Conv 结构,相较于传统的卷积神经网络结构 Conv-BN-Activation,由此带来的改进主要有以下 3 点:

(1) 加快模型收敛速度。将批量标准化放在卷积层前,可以使输入在进入激活函数之前就完成规范化,避免了激活函数的非线性特性对输入的扰动,从而有利于模型快速收敛。

(2) 减少模型训练时间和参数量。在传统的结构中,批量标准化层需要额外的参数和计算,而且在每个批次中都要计算均值和方差,这会增加训练时间和模型的参数量。而在 BN-Activation-Conv 结构中,由于批量标准化层已经对输入进行了规范化,因此可以省略掉传统结构中的批量标准化层,从而减少了模型的参数量和计算量。

(3) 提高模型泛化性能。批量标准化层的作用是对特征进行规范化处理,使得输入分布更加稳定,同时将批量标准化层放在卷积层前,还可以更好地控制输入数据的分布,从而有助于提高模型的泛化性能。此外,实验在保证对网络识别准确率无明显影响的前提下减少了卷积核的数量,使得网络轻量化,从而大幅提升了网络训练速度。传统残差网络和改进后的残差网络结构示意图如图 4 所示。

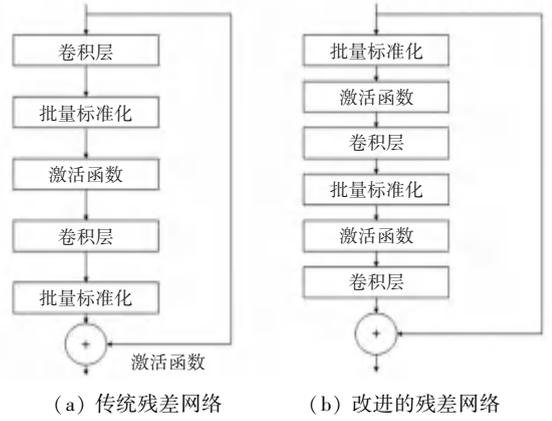


图 4 残差网络结构示意图

Fig. 4 Schematic diagram of residual network structure

调整后的网络各层参数如下:

- (1) 输入层。图像尺寸为 224×224 。
- (2) 卷积层。64 个 7×7 卷积核,卷积步长为 2, BN-Activation-Conv 结构。
- (3) 最大池化层。池化核尺寸为 3×3 ,步长为 2。
- (4) 残差块 1。包含 2 个卷积层和 1 个跳跃连接,64 个 3×3 卷积核, BN-Activation-Conv 结构。
- (5) 残差块 2。包含 2 个卷积层和 1 个跳跃连接,64 个 3×3 卷积核, BN-Activation-Conv 结构。
- (6) 残差块 3。包含 2 个卷积层和 1 个跳跃连接,128 个 3×3 卷积核, BN-Activation-Conv 结构。
- (7) 残差块 4。包含 2 个卷积层和 1 个跳跃连接,256 个 3×3 卷积核, BN-Activation-Conv 结构。
- (8) 平均池化层。池化核尺寸为 7×7 。
- (9) 全连接层。输出为 62 个类别。
- (10) Softmax 层。

2.3.2 模型优化

(1) 优化器对网络模型表现起着重要作用。研究可知,带动量的随机梯度下降(SGDM)算法是众多流行的自适应学习优化器中的一种,在卷积神经网络训练中具有出色的表现。

与传统的随机梯度下降(SGD)相比,SGDM 利用一阶动量 Momentum,将参数 β 调整至 0.9,而梯度变化曾经用到的公式为:

$$m_{t_0} = g_{t_0} \quad (4)$$

本次研究中用到的梯度变化公式为:

$$m_t = \beta m_{t-1} + (1 - \beta) g_t \quad (5)$$

其中, m_t, m_{t-1} 分别表示当前时刻与前一时刻的梯度更新; g_t 表示目标函数的梯度,其计算公式如下:

$$g_t = \frac{\partial(\text{LOSS})}{\partial \theta_t} \quad (6)$$

这样一来,梯度的变化不仅受当前时刻的影响,也受到历史梯度累积的作用,大大提高了模型的收敛速度,同时也使得基本梯度下降法可能出现的局部鞍点和噪声梯度的问题得以减轻。

(2)在深度学习中,损失函数对于优化模型性能和降低预测误差起着关键性作用。在神经网络多分类任务中,交叉熵损失函数(Cross Entropy Loss)是最常使用的损失函数之一。为了计算交叉熵损失,首先需要将神经网络的原始输出通过激活函数(如Softmax)映射到0~1范围内,得到概率分布。然后,根据预测概率分布和真实概率分布,计算两者之间的交叉熵。

通过最小化交叉熵损失,模型能够在训练过程中更快地收敛,预测误差也有所降低。而这将会有效提高神经网络在分类任务中的准确性和泛化能力。

2.4 图像分类流程

本文网络模型在处理植物图像的分类流程如图5所示。

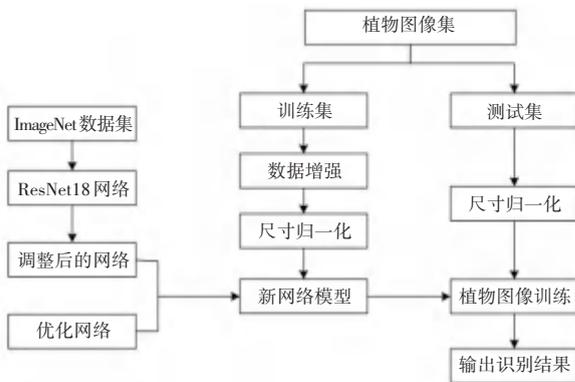


图5 植物分类流程图

Fig. 5 Flowchart of plant classification

3 实验结果与分析

3.1 数据集及其预处理

本文使用的植物图片数据源自于kaggle的野生植物图像数据集,包含62种植物类别,共计6558张图片,如银杉、蓝马鞭草和车前草等。

由于数据集相对较小,可能会导致模型过拟合,从而影响模型的识别准确率。因此,研究使用了平移、错切变换、翻转、亮度调整和剪切变换等多种数据增强方法来扩充数据^[14-17],将原始数据集扩充为原来的5倍,总共32790张植物图片。为了进行实验,将这些图片按照4:1的比例分成训练集和测试集,其中训练集包含26232张图片,测试集包含658张图片。最后,研究则将处理后的数据规范化为大

小为224×224的图片。这一数据增强处理方法为本文后续实验的有效性提供了重要基础。

3.2 实验环境及参数设置

实验在Windows10系统环境下完成,CPU处理器选择Intel(R)Core(TM)i5-10210U,GPU为NVIDIA RTX2080Ti,深度学习框架为PyTorch1.6.0。

综合比较评估热门的神经网络,选择综合评分较高的ResNet18为基础网络模型进行改进和训练。

3.3 实验结果及分析

3.3.1 数据扩充对实验的影响

为考察数据增强对实验的影响,文中进行了对比实验。以改进后的ResNet18为模型网络,在保证其余等条件相同的情况下,分别在原始数据集和增强后的数据集上进行了训练,数据增强操作包括平移、翻转、剪切等,测试精度对比如图6所示。

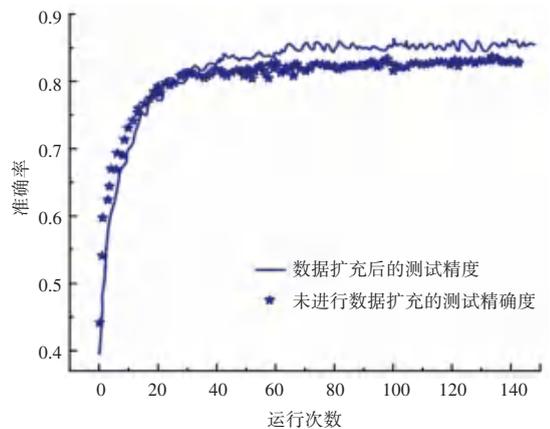


图6 数据增强前后准确率对比

Fig. 6 Comparison of accuracy before and after data enhancement

分析图6可得,采用数据增强方法,使模型的测试精度提高约3个百分点,说明了数据增强对于小样本数据的有效性。

3.3.2 改进策略对实验结果的影响

(1)ResNet18网络自对比实验。为了探究不同的调整策略对模型表现的影响,在迁移学习的基础上对原始ResNet18模型、BN-Activation-Conv ResNet18、添加SENet模块的ResNet18模型以及添加SENet模块的BN-Activation-Conv ResNet18进行了测试实验。损失函数选用交叉熵损失函数,经过150次迭代后,根据测试准确率最大值进行模型性能评估。表1展示了不同模型识别准确率与损失的对比情况,模型测试精度和损失随迭代次数的变化情况如图7、图8所示。

表 1 改进模型训练结果对比

Table 1 Comparison of improved model training results

| 模型 | 测试精度/% | 测试损失 |
|------------------------|--------|------|
| ResNet18 | 78.6 | 1.4 |
| ResNet18+注意力机制 | 84.2 | 1.2 |
| 改进网络的 ResNet18 | 79.4 | 0.5 |
| 改进网络的 ResNet18 + 注意力机制 | 86.6 | 0.4 |

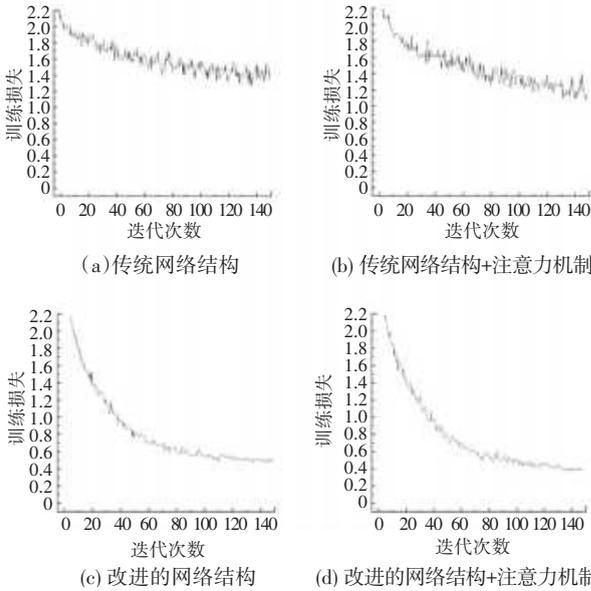


图 7 4 种不同改进算法的训练损失

Fig. 7 Training losses of four different improved algorithms

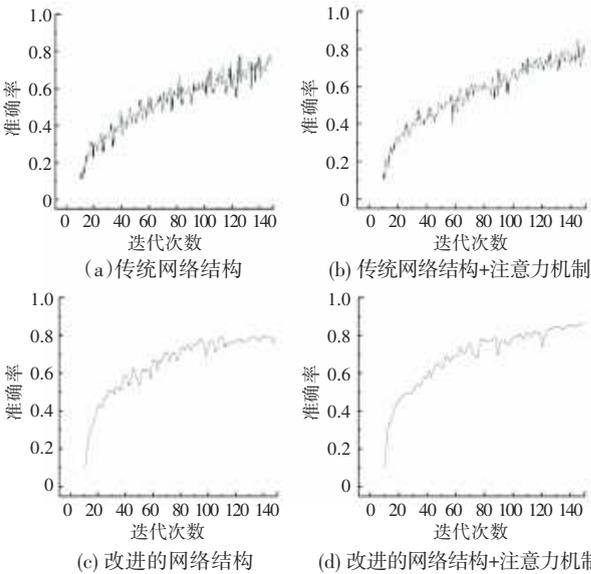


图 8 4 种不同改进算法的识别准确率

Fig. 8 Recognition accuracy of four different improved algorithms

讨论表 1 实验结果可知,与原始 ResNet18 模型相比,加入 SENet 后模型准确率提高 5.6%,模型损失降低 0.2;调整网络后准确率提高 0.8%,模型损失降低 0.9,最终改进模型准确率达 86.6%,较原始模型 ResNet18 准确率提升了 8.0%,损失降低了 1.0。

分析原因可得:

① SENet 模块增强了和识别任务相关的特征信息的权重,提升了网络模型的抗干扰能力。

② BN-Activation-Conv 结构中,梯度能够更自然地传递,并加强模型的正则化,降低模型过拟合风险。观察图 7,相同迭代次数下,引入 BN-Activation-Conv 结构的模型收敛速度更快,能够更快地达到较好的训练效果。

(2)与其他网络模型的对比试验。为验证改进后的网络结构的性能,选择热门网络 VGGNet16 和 VGGNet19 以及未改进的 ResNet18 作为对照组。经过 150 次迭代,训练结果见表 2。同时研究得到的不同网络准确率对比结果如图 9 所示。与其他模型比较可知,本文模型的识别准确率更高,且模型损失更低,验证了本文提出模型的有效性。

表 2 不同模型训练结果对比

Table 2 Comparison of training results of different models

| 模型 | 测试精度/% | 测试损失 |
|------------------------|--------|------|
| VGGNet16 | 70.0 | 1.5 |
| VGGNet19 | 67.7 | 1.5 |
| ResNet18 | 78.6 | 1.4 |
| 改进网络的 ResNet18 + 注意力机制 | 86.6 | 0.8 |

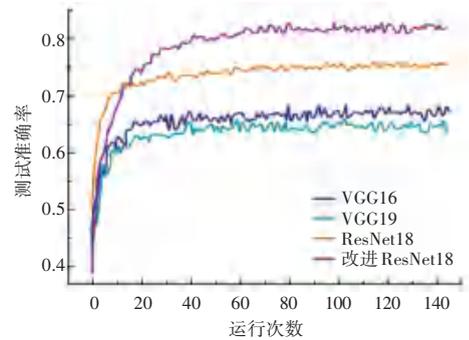


图 9 不同网络准确率对比

Fig. 9 Comparison chart of accuracy of different networks

3.3.3 优化器和学习率对模型性能的影响

优化器的选择对模型表现有着显著影响。为了找到能让模型达到最佳状态的优化器和学习率,本文在改进后的 ResNet18 的基础上,将 SGDM、Adam 和 RMSprop 三种优化器进行对比,初始学习率分别设置为 0.01、0.001 和 0.000 1。实验结果见表 3。

表 3 不同优化器的识别准确率

Table 3 Recognition accuracy of different optimizers

| 学习率 | 优化器 | | |
|---------|------|------|---------|
| | SGDM | Adam | RMSprop |
| 0.01 | 85.2 | 84.6 | 85.2 |
| 0.001 | 86.6 | 84.9 | 84.8 |
| 0.000 1 | 76.6 | 85.2 | 85.3 |

结果显示,优化器选择 SGDM,学习率为 0.001,使得模型能表现最佳,测试准确率最高达到 86.6%。

4 结束语

综上,为了解决传统识别方法识别精度低、速度慢的问题,本文以 ResNet18 为基础网络模型,引入注意力模块,降低了噪声的干扰。接着将传统 Conv-BN-Activation 结构改为 BN-Activation-Conv 结构,加强模型的正则化,减少过拟合的影响。网络还采用了迁移学习的训练方法,降低网络表现对数据的依赖,提高模型整体性能。该模型在植物种类识别任务中表现优异,准确率较高,可以作为一个实用的工具用来解决实际问题。

参考文献

- [1] 张殷波,杜昊东,金效华,等. 中国野生兰科植物物种多样性与地理分布[J]. 科学通报,2015,60(2):179-188.
- [2] 马艳,李韵. 虚拟价值理论及现代性分析[J]. 复旦学报(社会科学版),2012(1):103-110.
- [3] 张卡,宿东,王蓬勃,等. 深度学习技术在影像密集匹配方面的进展与应用[J]. 科学技术与工程,2020,20(30):12268-12275.
- [4] 赫磊,邵展鹏,张剑华,等. 基于深度学习的行为识别算法综述[J]. 计算机科学,2020,47(S1):139-147.
- [5] GUYER D E, MILES G E, GAULTNEY L D, et al. Application of machine to shape analysis in leaf and plant identification [J]. Transactions of the Asae, 1993, 36(1): 163-171.
- [6] 孟亮,郭小燕,杜佳举,等. 一种轻量级 CNN 农作物病害图像识别模型 [J]. 江苏农业学报, 2021, 37(5): 1143-1150.
- [7] 曹香滢,孙卫民,朱悠翔,等. 基于科优先策略的植物图像识别 [J]. 计算机应用, 2018, 38(11): 3241-3245.
- [8] 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.
- [9] PAN S J, YANG Qiang. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [10] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016: 770-778.
- [11] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [12] 刘学平,李琦乾,刘励,等. 嵌入 SENet 结构的改进 YOLOV3 目标识别算法 [J]. 计算机工程, 2019, 45(11): 243-248.
- [13] ZHU Hongyu, XIE Chao, FEI Yeqi, et al. Attention mechanisms in CNN-based single image super-resolution: A brief review and a new perspective [J]. Electronics, 2021, 10(10): 1187.
- [14] LIN M, CHEN Q, YAN S. Network in network [J]. arXiv preprint arXiv, 1312.4400, 2014.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- [16] JIA Deng, WEI Dong, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248-255.
- [17] 蒋芸,张海,陈莉,等. 基于卷积神经网络的图像数据增强算法 [J]. 计算机工程与科学, 2019, 41(11): 2007-2016.