

基于高阶残差和注意力机制的轻量型作物害虫识别^①



阮炬全, 刘 朔

(武汉轻工大学 数学与计算机学院, 武汉 430048)
通信作者: 刘 朔, E-mail: 874477154@qq.com

摘 要: 精准识别作物害虫对作物进行适时地防护和治理具有重要意义. 在面向自然环境时, 由于作物害虫体积小、与环境颜色的差异性不大, 同时又面临着各种噪声和复杂背景等因素的影响, 目前与深度学习相关的作物害虫识别模型存在难以兼顾识别准确率和鲁棒性的要求, 难以部署在计算资源有限和低性能的移动端等缺陷. 因此选取 ShuffleNetV2 网络结构中模型参数最少的 ShuffleNetV2 0.5× 为基准网络, 设计了一个基于高阶残差和注意力机制的轻量型作物害虫识别模型 (HOR-Shuffle-CANet). 其中, 前期的高阶残差可以为后面的网络层提供丰富的害虫特征, 有效提高模型的识别准确率; 坐标注意力 (coordinate attention, CA) 机制能够进一步抑制背景噪声和对作物害虫重点信息的关注, 有效增强模型的鲁棒性; 带标签平滑正则化 (label smoothing regularization, LSR) 的双稳态逻辑损失函数可以解决训练含噪数据集时逻辑损失函数的两个缺点, 使得模型对噪声的适应能力更强. 试验结果表明, HOR-Shuffle-CANet 模型在自然场景中 10 类常见作物害虫图像的测试数据集上达到了 91.22% 的识别准确率, 较基准网络提升了 3.54 个百分点. 在保持轻量化计算的基础上, 其识别准确率也高于现有的经典卷积神经网络 AlexNet、VGG-16、GoogLeNet、Xception、ResNet-34 和轻量级网络模型 MobileNetV3-Small、EfficientNet-B0 等. 该模型具有识别准确率高、鲁棒性强和抗干扰性能好等特点, 能够很好地适应作物害虫识别的实际应用需求.

关键词: 作物; 高阶残差; 坐标注意力; 轻量化; 鲁棒性; 害虫识别

引用格式: 阮炬全, 刘朔. 基于高阶残差和注意力机制的轻量型作物害虫识别. 计算机系统应用, 2023, 32(3): 104-115. <http://www.c-s-a.org.cn/1003-3254/8987.html>

Lightweight Recognition of Crop Pests Based on High-order Residual and Attention Mechanism

RUAN Ju-Quan, LIU Shuo

(School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan 430048, China)

Abstract: Accurate recognition of crop pests is essential for timely crop protection and treatment. However, crop pests in natural environment are small in size and have almost the same color as the environment. Moreover, crop pest images are affected by various factors such as noise and complex background. Therefore, it is difficult for existing crop pest recognition models related to deep learning to balance the requirements of recognition accuracy and robustness and be deployed on mobile devices with limited computational resources and low performance. In this study, ShuffleNetV2 0.5×, which has the fewest model parameters in the ShuffleNetV2 network structure, is selected as the benchmark network, and a lightweight crop pest recognition model based on high-order residual and attention mechanism (HOR-Shuffle-CANet) is designed. Specifically, the high-order residual in the early stage can provide rich pest features for the subsequent network layer, which significantly improves the recognition accuracy of the model. The coordinate attention (CA) mechanism can further suppress the background noise and focus on the key information about crop pests, which effectively enhances the

^① 收稿时间: 2022-08-15; 修改时间: 2022-09-15; 采用时间: 2022-09-27; csa 在线出版时间: 2022-11-16
CNKI 网络首发时间: 2022-11-18

robustness of the model. The bi-tempered logistic loss function with label smoothing regularization (LSR) can solve two shortcomings of logistic loss functions in training noisy data sets and make the model more adaptable to noise. The experimental results show that the HOR-Shuffle-CANet model achieves a recognition accuracy of 91.22% on the test dataset of ten types of common crop pest images in natural scenarios, which is 3.54 percentage points higher than the benchmark network. On the basis of maintaining lightweight computing, its recognition accuracy is also higher than that of the existing classical convolutional neural networks such as AlexNet, VGG-16, GoogLeNet, Xception, and ResNet-34, as well as lightweight network models such as MobileNetV3-Small, EfficientNet-B0, etc. Due to its high recognition accuracy, strong robustness, and excellent anti-interference performance, the proposed model can meet the practical application requirements of crop pest recognition.

Key words: crop; high-order residual (HOR); coordinate attention (CA); lightweight; robustness; pest recognition

我国是一个农业大国, 农业是国家的根本, 人民的粮食是第一位的. 粮食安全问题一直是全球重点关注话题, 我国“十四五”规划纲要明确提出“实施粮食安全战略”, 重点是确保国家粮食安全. 近年来, 由于气候变化、耕作方式的改变和作物复种指数的增加, 导致作物病虫害时常暴发成灾. 通过对害虫源基数、作物布局、种植规律和气象条件等方面的综合分析, 预测2022年我国主要粮食作物病虫害呈重发性, 重大检疫病虫发生程度比2021年和2016–2020年均值分别增加11.9%和8.7%^[1]. 众所周知, 农业害虫是影响作物稳定和增加产量的重要因素, 作物害虫识别是农业领域的主要挑战之一^[2]. 作物一旦受到害虫感染, 会对产量造成巨大损失, 必须及时采取有效的防治措施以避免害虫的扩散. 因此, 如何准确、鲁棒、抗干扰地对自然场景中作物害虫进行有效的辨识, 是保证粮食安全的先决条件, 也是目前农业生产中亟待解决的关键问题, 值得我们共同关注和重视.

随着互联网技术迅速发展, 许多国内外研究者使用传统图像处理 and 深度学习等技术来对作物害虫识别进行研究. 例如, 卢柳江等^[3]利用 Haar-like 特征描述子提取的害虫特征和训练级联 AdaBoost 分类器, 实现了大螟的智能识别. Umar 等^[4]通过提取害虫的颜色直方图和形状特征, 并将其输入到贝叶斯网络进行训练, 实现了在简单背景下对4类小样本害虫的自动识别, 平均精度可达到80%. Liu 等^[5]通过提取小麦蚜虫的梯度方向直方图特征和训练支持向量机, 取得了较好的识别效果. 李静等^[6]改进了基于 GoogLeNet^[7]的卷积神经网络 (convolutional neural network, CNN) 模型来识别玉米螟虫害, 并引入批标准化 (batch normalization,

BN)^[8]来加速优化网络模型训练, 使其具有很强的适用性. 孙鹏等^[9]在普通 CNN 的基础上, 引入了注意力机制, 研究出了一种基于注意力机制与 CNN 的大豆害虫识别方法, 识别准确率相比于传统网络提升了6.53%. Cheng 等^[10]通过借鉴 ResNet^[11]中的残差模块, 基于 AlexNet^[12]来加深网络设计了一个新型的害虫识别卷积神经网络模型, 实现对自然环境中10类害虫的自动识别. 陈峰等^[13]结合机器视觉与 CNN 的方法来智能识别了东北寒地中5种玉米害虫. Pattnaik 等^[14]提出了一种基于预训练深度 CNN 框架的迁移学习, 用于对10种小样本的番茄害虫进行分类, 并使用 DenseNet^[15]模型获得了88.83%的最高识别准确率. 然而, 这些研究存在以下局限性: 基于传统图像处理的识别方法特别依赖于人类专家的经验 and 肉眼观测能力, 其速度慢、设计特征时主观性强, 而且误判率高; 基于深度学习的识别方法可以自动提取特征, 具有更强的学习能力, 以端对端方式识别作物害虫, 存在大型网络和轻量级网络两种趋势. 由于大型网络的庞大参数数量和计算量, 通常只能在具有较高计算能力的 GPU 上运行, 难以在实际生产中部署和应用, 促使深度学习分支逐渐转向轻量级研究. 这便于边缘设备直接运行深度学习模型, 但同时由于自然环境中作物害虫体型较小、与环境颜色的差异性不大, 而且图像本身包含各种噪声和复杂背景等因素干扰, 造成网络模型的识别准确率和鲁棒性都不高.

针对这些问题, 本文选取自然场景中10种常见作物害虫作为研究对象, 以轻量化网络 ShuffleNetV2^[16]为基础, 结合作物害虫图像的特点, 设计了一个基于高阶残差和注意力机制的轻量级作物害虫识别模型

(HOR-Shuffle-CANet), 并使用带 LSR^[17] 的双稳态逻辑损失 (bi-tempered logistic loss)^[18] 函数进行模型优化, 旨在构建一个简单准确、鲁棒性强、抗干扰性能好的轻量级 CNN 模型来实现对自然场景中常见作物害虫图像的高精度识别, 以满足现代农业生产中实际需求. 本文的主要贡献如下.

(1) 设计了一个基于高阶残差和注意力机制的轻量级 CNN 模型 (HOR-Shuffle-CANet) 用于自然场景中常见作物害虫识别. 其中, 高阶残差和坐标注意力机制分别用于提高网络模型的识别准确率和鲁棒性.

(2) 为了使得 HOR-Shuffle-CANet 模型对噪声的适应能力更强, 本文采用带 LSR 的双稳态逻辑损失函数进行模型优化.

(3) 大量试验结果表明, HOR-Shuffle-CANet 模型较现有的经典卷积神经网络和轻量级网络模型在参数

量、计算量和准确率之间达到了良好的平衡, 并使得模型的性能保持在较高水平, 具有识别准确率高、鲁棒性强和抗干扰性能好等特点.

1 数据来源与处理

1.1 数据集

从自然环境中获取的图像总是具有复杂的背景干扰, 这与固定背景颜色标本的图像不同, 在复杂背景下具有更高的难度和实用意义. 作为试验数据的自然场景中常见作物害虫图像数据集来源于文献 [19], 并将其 10 种常见作物害虫作为本文的研究对象, 如图 1 所示, 该数据集总共有 5629 幅 RGB 图像. 根据样本总数和类别样本的分布情况, 随机抽取数据集的 70% 作为训练集、20% 作为验证集、10% 作为测试集, 用于后续试验.

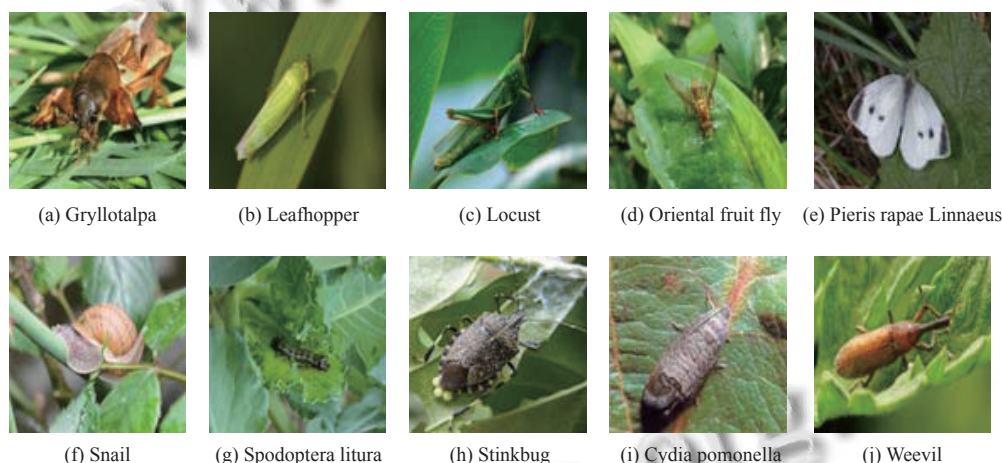


图 1 10 类常见作物害虫样本

1.2 数据增强

作物害虫数据的多样性、旋转不变性和对称性, 使得可以对作物害虫数据集进行数据增强. 数据增强是深度学习中扩充更多数据的常用技术^[20]. 数据增强包括两种方式, 即离线增强和在线增强. 对于样本量相对较小的数据集, 使用离线数据增强技术在有限的数据上进行多种函数映射, 将函数映射后获得的图像保存至本地数据集中, 实现数据扩充. 对于大型数据集, 在线数据增强是一种更合适的方法. 数据增强有两个主要优点: (1) 减少 CNN 模型训练中的过拟合风险, 增强其泛化能力; (2) 加入噪声数据, 增强模型的鲁棒性^[19], 由于本文研究的数据集相对较小, 所以采用离线数据增强的训练策略, 主要分为几何变换和颜色变换两大

类. 其中, 几何变换并不改变图像自身的内容, 而是增加了害虫样本的随机多样性, 包括翻转、旋转、随机裁剪、仿射变换、错切变换. 颜色变换主要有亮度变换、对比度变换、HSV 增强、高斯模糊、添加噪声. 这些数据增强方法相对通用且计算成本较低, 采用 Python 图形库和开源计算机视觉库中的相关函数即可实现, 可以有效地训练深度学习模型. 经过上述数据增强操作后, 如图 2 所示, 训练集数据扩增为原来的 12 倍, 增加至 47220 幅图像. 为了满足后续网络模型的输入要求, 将作物害虫图像统一调整为 224 像素×224 像素, 图像格式均为 png 格式. 表 1 列出了用于试验的作物害虫数据集图像的标签和种类名称以及数量等详细信息.



图2 数据增强示例

表1 作物害虫图像数据集信息

标签	种类名称	训练集		验证集	测试集
		增强前	增强后		
0	Grylotalpa (蝼蛄)	353	4236	101	51
1	Leafhopper (叶蝉)	300	3600	86	43
2	Locust (蝗虫)	434	5208	124	63
3	Oriental fruit fly (东方果蝇)	322	3864	92	47
4	Pieris rapae linnaeus (菜粉蝶)	378	4536	108	55
5	Snail (蜗牛)	751	9012	215	108
6	Spodoptera litura (斜纹夜蛾)	281	3372	80	41
7	Stinkbug (臭蝽)	475	5700	136	68
8	Cydia pomonella (苹果蠹蛾)	290	3480	83	42
9	Weevil (象鼻虫)	351	4212	100	51
图像总数		3935	47220	1125	569

2 轻量型作物害虫识别模型

本节先介绍了基于高阶残差和注意力机制的轻量型作物害虫识别模型 (HOR-Shuffle-CANet) 的网络架构和参数详情, 然后从注意力机制、高阶残差模块与双稳态逻辑损失函数 3 个方面来详细分述模型设计和优化.

2.1 网络架构

由于自然环境中作物害虫体积小、颜色与背景相近, 而且图像本身具有各种噪声和复杂背景等因素, 不利于网络模型对害虫类别的辨识. 面向作物害虫识别的实际应用需求, 本文选取轻量化网络 ShuffleNetV2 0.5 \times 为基准, 兼顾识别准确率和抗干扰能力, 设计了 HOR-Shuffle-CANet 网络模型, 该网络的整体架构如图 3 所示. 图 3 中, 在 HOR-Shuffle-CANet 中输入待识别的作物害虫图像后, 先通过第 1 层卷积层提取低级特征; 然后通过 3 个嵌入 CA 的残差模块 (CA-Res) 级联, 进一步提取更为丰富的特征; 接着使用最大池化层进行下采样, 在保留主要特征的前提下, 降低参数量和计算量; 再使用 3 个嵌入 CA 的 Stage 模块层 (CA-Stage) 来提取图像的高级特征, 随后一个卷积层用来混合特征; 最后经过一个全局平均池化层 (GAP) 和一个全连接层 (FC) 来获取作物害虫图像的识别结果. 其中, 蓝色模块代表 CA-Res, 绿色模块代表 CA-Stage, 3 个 CA-Stage 是由 Shuffle-CA Unit2 和 Shuffle-CA Unit1 组成的模块层, 模块层中单元 2 和单元 1 的数量分别为 1:3、1:7 和 1:3.

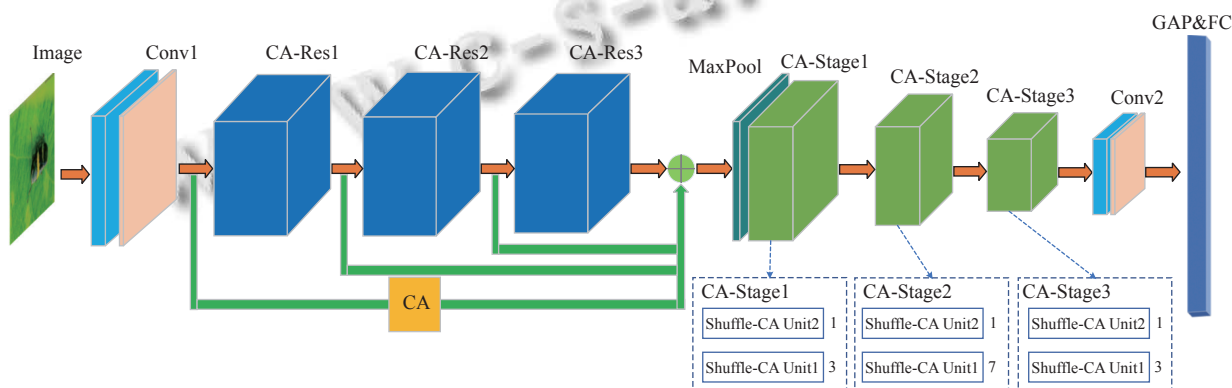


图3 HOR-Shuffle-CANet 网络架构图

2.2 参数详情

本文所设计的 HOR-Shuffle-CANet 模型是一个简单的轻量型卷积神经网络, 表 2 中列出了该模型的主

要参数. 由表 2 可见, 在设计 HOR-Shuffle-CANet 时, 仅采用了 1 \times 1 和 3 \times 3 两种类型的卷积核, 以降低网络参数量. 此外, 3 个 CA-Res 的输入和输出特征图的

通道数保持一致,此时内存访问成本 (memory access cost, MAC) 最低、运行速度最快,满足轻量化模型的设计原则^[16]. 最后,在全连接层中还使用了随机失活 (dropout)^[12],它可以有效减少隐含层节点间的相互作用,有助于避免过拟合,在一定程度上达到正则化的效果,使得模型的泛化性能更强.

表 2 HOR-Shuffle-CANet 网络参数详情

名称	滤波器/步长	输出尺寸
Image	—	224×224×3
Conv1	3×3/2	112×112×24
CA-Res1	—	112×112×24
CA-Res2	—	112×112×24
CA-Res3	—	112×112×24
MaxPool	3×3/2	56×56×24
CA-Stage1	—	28×28×48
CA-Stage2	—	14×14×96
CA-Stage3	—	7×7×192
Conv2	1×1/1	7×7×1024
GAP	—	1×1024
FC	—	1×10

2.3 注意力机制

运用在深度学习中的注意力机制与人类的注意力机制类似,皆是从众多信息中选择出重点关注的信息,并投入更多的注意力资源^[21]. 注意力机制能够灵活地捕获全局信息和局部信息之间的联系,它的目的就是突出显著有用特征,抑制和忽略无关特征.

本研究使用坐标注意力机制,CA^[22]模块是一种针对轻量级网络设计的新型混合域注意力模块,将空间注意力^[23]和通道注意力^[24]组合起来,其结构如图4所示. CA模块采用了有效的方法来获取通道关系和位置信息,实现了进一步抑制背景噪声和对作物害虫重点信息的关注,并输出更具害虫本质特性的精细特征. 它不仅捕捉了网络通道间的长期依赖关系,而且保留了精确位置信息的特性,使得网络能够在不失去精确位置信息的情况下,同时关注与害虫有关的区域.

具体而言,其过程分为以下3步.

Step 1. 坐标信息嵌入: 给定一个输入特征图,将全局平均池化为两步操作(一对1D池化),即用两个池化核 $H \times 1$ 和 $1 \times W$ 沿着特征图的两个不同方向(水平和垂直)进行池化,得到两个嵌入后的信息特征图.

Step 2. 坐标注意力生成: 将得到的两个嵌入特征图沿着空间维度进行拼接 (concat), 经过共享 1×1 卷积降维变换和 H-Swish 函数^[25]激活. 随后沿着空间维度

进行切分 (split) 操作得到两个分离的特征图,再对它们进行 1×1 卷积变换和 Sigmoid 函数激活,最后分别得到水平方向和垂直方向的注意力向量.

Step 3. 特征图重校准: 将输入的特征图与具有位置信息的两个方向上的特征图,通过按照元素相乘的方式来获取 CA 模块的增强特征表达,从而完成对特征图的重校准.

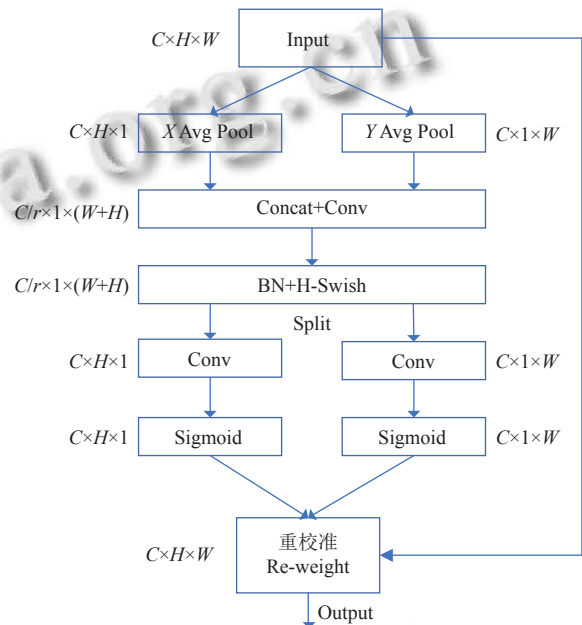


图 4 坐标注意力机制 (CA 模块)

本文所设计的 HOR-Shuffle-CANet 模型中嵌入 CA 模块的 3 处地方.

1) 在网络的第 1 层卷积层的捷径分支处连接 CA 模块,如图 3 所示.

2) 在 ShuffleNetV2 Unit1 的深度卷积 (depthwise convolution, DW 卷积) 输出的后面连接 CA 模块,即 Shuffle-CA Unit1 (基本单元), ShuffleNetV2 Unit2 作为 Shuffle-CA Unit2 (下采样单元),如图 5 所示.

3) 在高阶残差模块的子模块内嵌 CA 模块,即 CA-Res 模块,如图 6 所示.

这是因为 CA 模块结合了卷积层提取的特征来改变权重参数,通过赋予更大的权重来强化重要通道的特征,通过赋予较小的权重来削弱非重要通道的特征,从而使网络不会因为卷积操作而丢失更多关键的害虫信息,从而增强了模型识别不同类别作物害虫的能力.

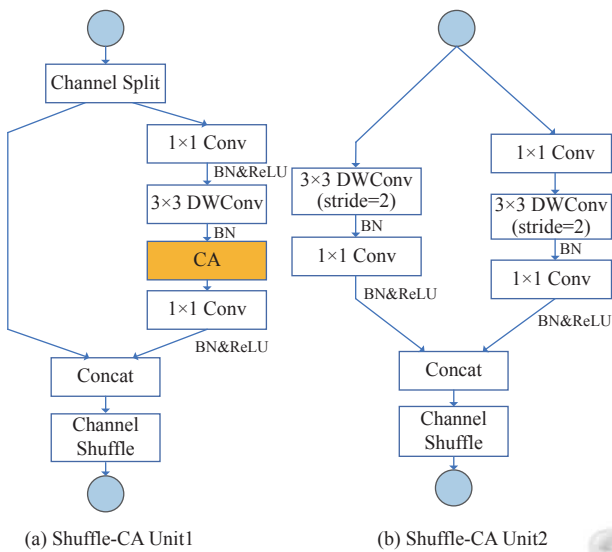


图5 Shuffle-CA Unit

2.4 高阶残差模块

2015年, He等^[11]提出了一种残差块结构, 它可以通过捷径连接的方式来获得网络中某一层的激活, 然后迅速反馈到另外一层, 甚至是神经网络的更深层. 这有效地解决了深层网络训练时的退化问题, 利用残差块设计的 ResNet 模型在 ILSVRC 上取得了很好的成绩. ResNet 可以加速网络拟合, 同时能够达到更高的识别准确率.

基于已有工作和后续的消融试验分析表明, 使用 3 个残差模块的级联操作^[26,27], 即高阶残差模块 (high-order residual, HOR) 可以有效地提高 ShuffleNetV2 网络的识别准确率. 在图 3 中, 为了有效地利用不同层的特征, 将第 1 层卷积操作后的特征映射和经过捷径分支处 CA 模块后所得到的增强特征表达进行特征图的重校准, 新的输出特征映射和前 3 个 CA-Res 模块的输出一起连接, 实现特征复用. 这样不仅能为后面的网络层提供更为丰富的特征, 使得模型的识别准确率更高, 还能间接提高模型的鲁棒性, 增强其抗干扰性能. 因此, 本研究考虑将 HOR 和 CA 模块融合, 其结构如图 7 所示. 此外, 并将嵌入 CA 的 HOR 引入 ShuffleNetV2 网络中, 其在 HOR-Shuffle-CANet 网络中的结构, 如图 3 所示.

由图 6 可得, CA-Res 模块是由 Conv1、Conv2、Conv3 和一个内嵌 CA 模块这 4 部分构成. 输入经过这 3 个卷积层后的输出分别为 X_1 、 X_2 、 X_3 , 其参数设置如图 6 所示. 其中, X_1 和 X_3 的通道数相同, 可以直接

级联起来, 将 X_1 经过捷径分支处 CA 模块增强后所得到的具有注意力信息的特征映射和 X_3 之和作为整个 CA-Res 模块的总输出.

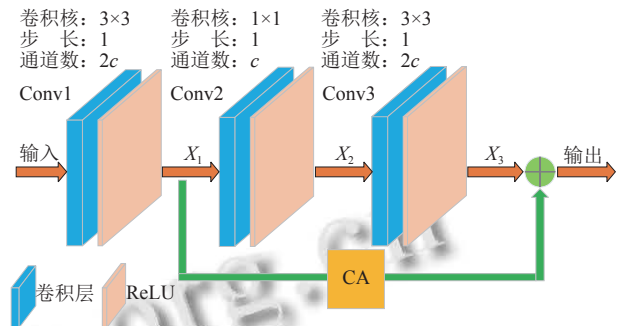


图6 CA-Res 模块结构图

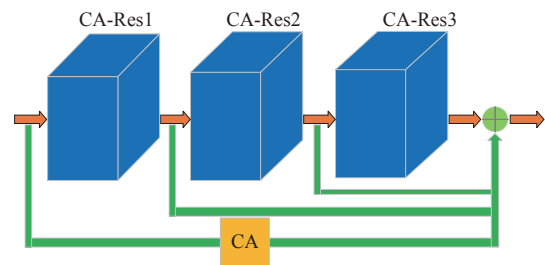


图7 嵌入 CA 的高阶残差模块结构图

2.5 双稳态逻辑损失函数

对于含有噪声的训练集, 在训练过程中使用的损失函数对模型的效果具有很大影响. 而逻辑损失函数, 一直是一个标准的选择, 用来训练深度神经网络进行分类. 这个损失涉及在最后一层的激活层上使用 Softmax 函数来形成类别概率, 然后获得真实标签和预测标签之间的相对熵.

逻辑损失是最后一层的激活凸函数, 但是在使用这个函数训练含噪数据集时存在以下两个问题: ① 错误标记的数据所产生的影响会扩展到决策边界, 即在分类边界附近存在错误标记的训练样本的情况下, 由于 Softmax 函数对这些类别进行概率分配的尾部呈指数衰减, Softmax 函数的短尾会强制分类器将决策边界扩展到嘈杂的训练示例; ② 离群点会对总损失产生很大的影响, 即从优化的角度来看虽然是可取的, 但凸损失已经被证明很容易出现异常值^[28], 这是因为每个样本在激活函数作用下的损失会无限增大. 这些异常值可能对应于造成大梯度的极端示例, 或远离决策边界的错误分类训练示例^[18]. 因此, 在输出层采用凸损失函

数,对含噪数据集来说显得并不可取。

标签平滑是一种用于分类问题的正则化技术,通过修正损失函数,可使模型在训练时不会对其所预测的训练数据标签过度自信,从而确保模型的泛化性。

因此,本研究使用带 LSR 的双稳态逻辑损失函数来替换神经网络中的 Softmax 输出层,调整温度 t_1 和尾部重量 t_2 两个参数^[18],解决训练含噪数据集时逻辑损失函数的两个缺陷,使得模型对噪声的适应能力更强。一般的分类问题采用的是 Softmax 函数,计算公式如式(1):

$$\hat{y}_i = \frac{\exp(\hat{a}_i)}{\sum_{j=1}^k \exp(\hat{a}_j)} = \exp\left(\hat{a}_i - \log \sum_{j=1}^k \exp(\hat{a}_j)\right) \quad (1)$$

其中, \hat{a}_i 为类别 i 的线性激活。

首先,分别定义带 t_1 参数的对数函数 $\log_{t_1}(x)$ 和带 t_2 参数的指数函数 $\exp_{t_2}(x)$,即:

$$\log_{t_1}(x) := \frac{1}{1-t_1} (x^{1-t_1} - 1) \quad (2)$$

$$\exp_{t_2}(x) := [1 + (1-t_2)x]_+^{1/(1-t_2)} \quad (3)$$

然后,将式(1)替换为使用式(3)的稳态版本,即:

$$\begin{cases} \hat{y}_i = \exp_{t_2}(\hat{a}_i - \lambda_{t_2}(\hat{\mathbf{a}})), \text{ where } \lambda_{t_2}(\hat{\mathbf{a}}) \in R \\ \text{s.t. } \sum_{j=1}^k \exp_{t_2}(\hat{a}_j - \lambda_{t_2}(\hat{\mathbf{a}})) = 1 \end{cases}$$

其中, $\lambda_{t_2}(\hat{\mathbf{a}})$ 代表每个样本的归一化值。

接着,使用标签平滑编码来生成新的真实标签 y_i ,编码方式如下:

$$y_i = (1 - \varepsilon \times k / (k - 1)) \times y_i + \varepsilon / (k - 1)$$

其中, ε 是权重因子,范围为 $0 \leq \varepsilon \leq 1$, k 是类别数。

最后,使用式(2)的稳态版本替换新的真实标签 y 和预测标签 \hat{y} 之间的相对熵,得到带 LSR 的双稳态逻辑损失函数,计算公式如式(4):

$$\sum_{i=1}^k \left(y_i (\log_{t_1} y_i - \log_{t_1} \hat{y}_i) - \frac{1}{2-t_1} (y_i^{2-t_1} - \hat{y}_i^{2-t_1}) \right) \quad (4)$$

其中,当 $t_1 = t_2 = 1$ 时,双稳态逻辑损失减少到普通的相对熵损失;当 $0 \leq t_1 \leq 1$ 时,双稳态逻辑损失存在有界性,可以防止那些大边距噪声数据将决策边界拉开;当 $t_2 > 1$ 时,双稳态逻辑损失具有重尾性,可以保持决策边界远离小边距噪声样本。

3 模型试验与评价标准

3.1 试验设置

所有试验均是在 Windows 11 64 位操作系统下实现的。硬件环境:计算机内存为 32 GB,搭载 12th Gen Intel(R) Core(TM) i7-12700 @ 2.10 GHz 处理器;GPU 采用 NVIDIA GeForce RTX 3080,显存容量为 12 GB。软件环境:Python 3.6.13;开源的深度学习计算框架 PyTorch 1.7.0。

综合考虑试验设备的性能和预试验的训练效果,每批次样本数(batch size)设定为 64。初始学习率设置为 0.01,学习率更新采用 Warmup^[11]和余弦退火衰减策略^[29];后续模型中损失函数未标注说明则默认使用带 LSR 的交叉熵损失函数,已标注说明则使用带 LSR 的双稳态逻辑损失函数,参数取值为 $t_1 = 0.6$ 和 $t_2 = 1.8$;使用 AdamW^[30]最优化算法来优化损失函数;迭代次数(epoch)设置为 50 次。在测试阶段,为了证明所设计的作物害虫识别模型 HOR-Shuffle-CANet 的抗干扰能力,将不同程度的高斯噪声和椒盐噪声添加至测试集图像中,以测试模型的识别准确率和评估其鲁棒性。

3.2 评价指标

本研究选用准确率(Accuracy)、精确度(Precision)、召回率(Recall)和 F1 值^[31]这 4 个主要指标来衡量 HOR-Shuffle-CANet 模型的识别性能。

对于图像分类模型,准确率是最为重要的评价指标。准确率反映的是被模型正确识别的样本数占总样本数的概率值;精确度反映的是模型预测为正样本的数量中实际上被正确识别的概率值;召回率反映的是原始样本中的正样本被正确识别的概率值;F1 值为精确度和召回率的调和平均数。以上各指标的对应计算公式分别为式(5)–式(8):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

其中,TP(true positive)、FP(false positive)、TN(true negative)和 FN(false negative)分别表示真阳性、假阳性、真阴性和假阴性的样本数量。

为了更好地评价 HOR-Shuffle-CANet 模型的开销,

通过参数量和浮点运算量 (FLOPs)^[31] 指标与其他优秀模型进行对比, 参数量和浮点运算量常用来衡量模型的复杂度^[32].

混淆矩阵 (confusion matrix)^[33] 是一个评估模型预测结果的度量标准, 常用来评价分类模型的优劣. 本研究将分类模型对每个类别的预测结果是否正确以矩阵的形式进行可视化, 其中对角线上的元素表示各类别作物害虫被正确分类的数量, 其他元素则表示被错误分类的数量.

4 结果与分析

4.1 消融试验

为了实现模型的轻量化, 本文选取了 ShuffleNetV2 网络结构中模型参数最少的 ShuffleNetV2 0.5×作为基准网络. 如表 3 所示, 为了验证本文所设计的 HOR-Shuffle-CANet 模型的可行性和有效性, 在相同试验条件下, 采用以下 5 种消融试验方案: 1) 仅采用 ShuffleNetV2 0.5×基准网络; 2) 在方案 1 的基础上嵌入 CA 模块; 3) 在方案 1 的基础上加入 HOR; 4) 在方案 2 和方案 3 的基础上进行融合形成本文所设计的轻量级网络模型 HOR-Shuffle-CANet; 5) 在方案 4 的基础上采用带 LSR 的双稳态逻辑损失函数对 HOR-Shuffle-CANet 模型进行优化训练.

从表 3 的 5 种消融试验方案的试验结果中, 比较方案 1 和 2 可知, 嵌入 CA 模块使得模型的准确率提高了 0.72 个百分点. 从方案 1 和 3 的试验结果可知, 当加入了 HOR 时, 模型的识别准确率获得了 1.82% 的提升, 这验证了加入 HOR 可以有效地提高识别准确率. 对比方案 1 和 4 的结果表明, 当同时使用 HOR 和 CA

模块时, 在仅增加部分参数量和计算量的情况下, 模型的准确率提升了 1.43%. 此外, 比较方案 4 和 5 可知, 采用带 LSR 的双稳态逻辑损失进行优化模型训练, 在不增加模型参数量的前提下, 模型的准确率整体提高了 2.11 个百分点. 最后, 从方案 1 和 5 的试验结果可知, 本文 HOR-Shuffle-CANet 模型的识别效果良好, F1 值可达 90.57%, 准确率提升到 91.22%. 与基准网络相比, HOR-Shuffle-CANet 模型在自然场景下常见作物害虫的识别准确率提升了 3.54 个百分点. 因此, 本研究的模型改进方案和网络训练方案是有效和可行的.

表 3 HOR-Shuffle-CANet 模型消融试验对比

编号	改进策略	参数量 (M)	F1 值 (%)	准确率 (%)
1	ShuffleNetV2 0.5×	0.34	87.21	87.68
2	ShuffleNetV2+CA	0.36	87.74	88.40
3	ShuffleNetV2+HOR	0.37	88.80	89.50
4	HOR-Shuffle-CANet	0.39	88.65	89.11
5	HOR-Shuffle-CANet (带LSR的双稳态逻辑损失)	0.39	90.57	91.22

4.2 与优秀模型对比

为了进一步检验 HOR-Shuffle-CANet 模型的识别效果, 在试验条件保持一致的情况下, 利用自然场景中 10 类常见作物害虫图像数据集, 将本文所设计的 HOR-Shuffle-CANet 模型与 AlexNet、VGG-16^[34]、GoogLeNet、Xception^[35]、ResNet-34 这 5 种经典卷积神经网络以及 SqueezeNet^[36]、MobileNetV3-Small^[25]、EfficientNet-B0^[37] 这 3 种轻量级网络进行对比试验. 表 4 分别显示了这些模型的复杂度和测试集识别准确率, 并与同类优秀轻量级网络的验证集准确率曲线进行了可视化对比, 如图 8 所示.

表 4 不同 CNN 模型识别准确率对比

模型	参数量 (M)	FLOPs (G)	精确度 (%)	召回率 (%)	F1 值 (%)	准确率 (%)
VGG-16	134.30	15.47	69.48	69.97	69.47	70.11
AlexNet	57.05	0.69	79.52	79.42	79.27	80.52
Xception	21.35	4.66	88.06	88.81	88.36	88.83
ResNet-34	21.29	3.68	87.66	87.78	87.65	88.35
GoogLeNet	5.61	1.51	88.46	88.74	88.53	88.81
EfficientNet-B0	4.02	0.40	86.58	86.63	86.50	87.22
MobileNetV3-Small	1.53	0.06	85.15	84.86	84.89	85.77
SqueezeNet	0.72	0.72	83.96	83.44	83.44	83.80
HOR-Shuffle-CANet (带LSR的双稳态逻辑损失)	0.39	0.35	90.52	90.95	90.57	91.22

由表 4 和图 8 可见, 针对在自然环境下常见作物害虫识别问题时, 不同层数和结构的 CNN 模型具有

不一样的识别效果. 本文设计的 HOR-Shuffle-CANet 模型识别准确率要高于其他模型. 对比各模型的参数

量和浮点运算量可知, HOR-Shuffle-CANet 模型的参数量仅为 0.39 M, 浮点运算量也仅高于 MobileNetV3-Small 轻量级网络. 除了 SqueezeNet、MobileNetV3-Small 和 EfficientNet-B0 轻量级网络以外, HOR-Shuffle-CANet 模型参数量和浮点运算量都远少于其他经典卷积神经网络模型. 综上所述, HOR-Shuffle-CANet 模型在保证模型性能的同时, 又保证了模型的参数量和运算量, 因此该模型可实现移动端的作物害虫识别.

4.3 抗干扰性能对比

为了评估 HOR-Shuffle-CANet 模型的鲁棒性, 在 10 类常见作物害虫图像数据的测试集中加入了不同等

级的多种噪声, 并列出了各模型的识别准确率结果, 如表 5 所示.

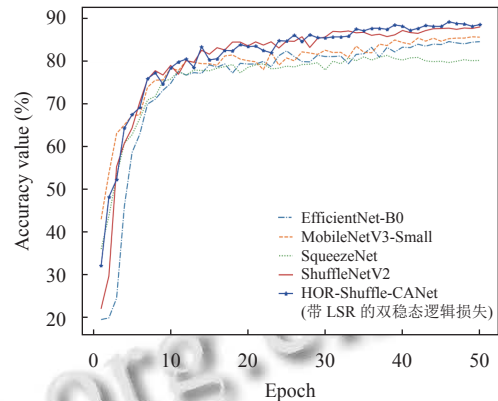


图 8 准确率曲线对比图

表 5 添加不同噪声时的识别准确率对比 (%)

噪声	模型	噪声等级					
		0.005	0.01	0.015	0.02	0.025	0.03
高斯噪声	ShuffleNetV2 0.5×	85.94	85.55	83.80	82.36	81.15	80.46
	ShuffleNetV2+HOR	86.51	88.01	85.38	82.39	81.21	79.50
	ShuffleNetV2+CA	88.59	86.44	84.21	83.25	82.71	79.96
	HOR-Shuffle-CANet (带LSR的双稳态逻辑损失)	90.13	88.55	87.77	87.12	84.75	83.17
椒盐噪声	ShuffleNetV2 0.5×	87.70	85.49	83.45	83.30	81.19	79.74
	ShuffleNetV2+HOR	87.53	86.99	85.77	85.25	83.88	81.90
	ShuffleNetV2+CA	85.92	85.53	84.12	81.87	81.82	81.19
	HOR-Shuffle-CANet (带LSR的双稳态逻辑损失)	89.76	88.37	87.68	85.60	83.97	81.91
高斯+椒盐噪声	ShuffleNetV2 0.5×	77.66	75.55	76.03	75.01	74.12	73.51
	ShuffleNetV2+HOR	78.13	76.37	74.88	75.27	74.10	72.34
	ShuffleNetV2+CA	77.89	75.86	74.97	73.25	72.90	72.38
	HOR-Shuffle-CANet (带LSR的双稳态逻辑损失)	80.24	80.98	77.63	76.92	75.64	75.34

表 5 中添加不同噪声时最高识别准确率是以加粗字体来表示的. 显而易见, 当添加了不同等级的高斯、椒盐、高斯+椒盐噪声后, HOR-Shuffle-CANet 模型的识别准确率均高于其他对比模型. 在加入不同等级的噪声干扰后, HOR-Shuffle-CANet 模型的识别准确率衰减较慢, 同时保证准确率比其他 3 种模型更高, 对噪声的抗干扰性能最强. 这是因为前期的高阶残差可以提供更为丰富的害虫特征给后面的网络层, 有效提高模型的识别准确率; CA 模块能够进一步抑制背景噪声和对作物害虫重点信息的关注, 有效增强模型的鲁棒性; 带 LSR 的双稳态逻辑损失函数能够解决训练含噪数据集时逻辑损失函数的两个缺点, 使得模型对噪声的适应能力更强. 上述试验结果表明, 本文设计的 HOR-Shuffle-CANet 模型在面向自然环境下作物害虫图像识别问题时, 具有高识别准确率、

强鲁棒性和较好的抗干扰性能等特点, 可以满足实际应用需求.

4.4 类激活图可视化分析

为了直观地展示 HOR-Shuffle-CANet 模型的有效性, 使用 Grad-CAM 技术^[38], 选取了部分作物害虫的测试集图像对 ShuffleNetV2 0.5×基准网络和 HOR-Shuffle-CANet 模型分别进行了类激活图可视化对比分析. 如图 9 所示, 第 1 行作物害虫是 locust (蝗虫), 第 2 行是 gryllotalpa (蝼蛄), 第 3 行是 spodoptera litura (斜纹夜蛾). 其中, 激活图色条权重越大越有利于害虫的正确识别.

从图 9 可以看到, 使用同一张原始作物害虫图像时, 本文选取的 ShuffleNetV2 0.5×基准网络虽然也能注意到害虫相关区域, 但定位不够准确, 仅集中在局部区域. 而 HOR-Shuffle-CANet 模型由于受背景区域

的关注较少,可以精确定位到目标害虫在图像中所处的主要位置,使得复杂背景的干扰较小.这说明了CA模块能够对特征的关键区域进行有效提取,对背景区域的噪声干扰起到抑制作用.因此,该模型具有较强的注意力学习能力,可以有效提高作物害虫的识别准确率.

4.5 不同类别作物害虫识别结果

为了更清晰地展现HOR-Shuffle-CANet模型识别不同类别作物害虫的效果,绘制了该模型在测试集上识别结果的混淆矩阵,如图10所示.其中,混淆矩阵对角线表示正确识别各类别作物害虫的数量.

最后,还详细列出了该模型在测试集上各类别作物害虫的识别精确度、召回率和F1值,如表6所示.

试验结果表明,HOR-Shuffle-CANet模型在测试集的大部分作物害虫上都获得了比较好的识别结果,这说明该模型的可靠性较高.

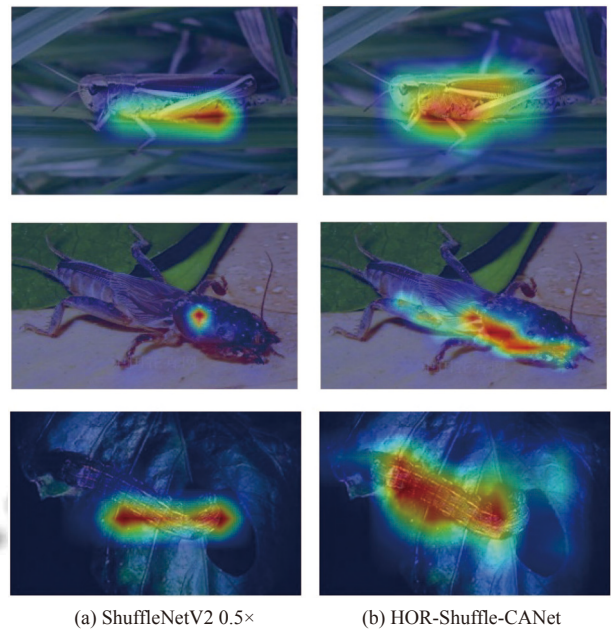


图9 不同模型的害虫图像类激活图

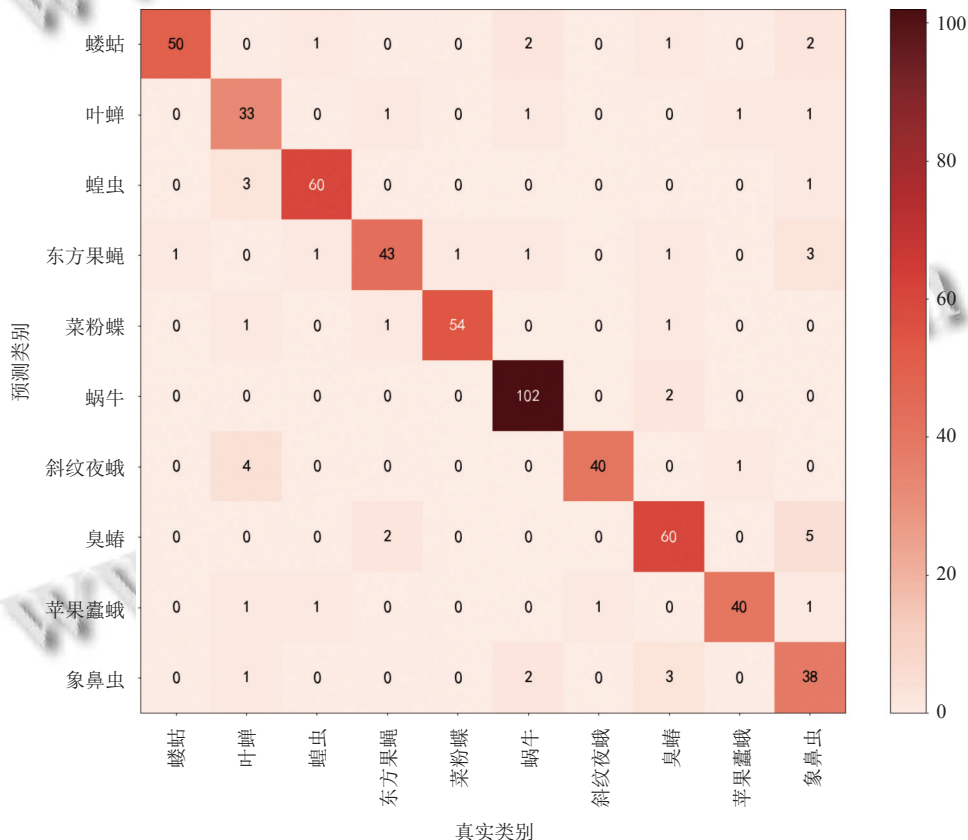


图10 作物害虫混淆矩阵

5 总结与展望

面向作物害虫识别的实际应用需求,本文设计了一个基于高阶残差和注意力机制的轻量级CNN模型

(HOR-Shuffle-CANet),用于解决农业生产实践中作物害虫识别模型难以部署在计算资源有限和低性能的移动端等问题.为了验证此模型,本文采用带LSR的双

稳态逻辑损失函数进行模型优化,在自然场景中10类常见作物害虫图像数据集上进行了大量的相关试验。试验结果表明,HOR-Shuffle-CANet模型在参数量、计算量和准确率之间达到了良好的平衡并使得模型性能保持在较高水平,在自然环境下10类常见作物害虫的识别准确率可达91.22%,具有高识别准确率、强鲁棒性和较好的抗干扰能力等特点。

表6 测试集上的识别精确度、召回率和F1值(%)

种类名称	Precision	Recall	F1
蝼蛄	89.3	98.0	93.4
叶蝉	89.2	76.7	82.5
蝗虫	93.8	95.2	94.5
东方果蝇	84.3	91.5	87.8
菜粉蝶	94.7	98.2	96.4
蜗牛	98.1	94.4	96.2
斜纹夜蛾	88.9	97.6	93.0
臭蝽	89.6	88.2	88.9
苹果蠹蛾	90.9	95.2	93.0
象鼻虫	86.4	74.5	80.0

在今后的研究中,将进一步扩充和完善自然场景中作物害虫数据集的种类,考虑将高阶残差模块和其他新兴技术更好地融合,继续完善模型结构来提升网络的识别准确率和鲁棒性,促进智慧农业的发展。

参考文献

- 2022年全国农作物重大病虫害发生趋势预报. 中国植保导刊, 2022, 42(4): 107-108.
- Thenmozhi K, Reddy US. Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, 2019, 164: 104906. [doi: 10.1016/j.compag.2019.104906]
- 卢柳江, 匡迎春, 陈兰鑫, 等. 基于级联AdaBoost分类器的农作物害虫图像识别研究. *中国农机化学报*, 2019, 40(8): 127-131.
- Umar HGA, Abbas Q, Gulzar F. Insect classification using image processing and Bayesian network. *Journal of Entomology and Zoology Studies*, 2017, 5(6): 1079-1082.
- Liu T, Chen W, Wu W, et al. Detection of aphids in wheat fields using a computer vision technique. *Biosystems Engineering*, 2016, 141: 82-93. [doi: 10.1016/j.biosystemseng.2015.11.005]
- 李静, 陈桂芬, 安宇. 基于优化卷积神经网络的玉米螟虫害图像识别. *华南农业大学学报*, 2020, 41(3): 110-116. [doi: 10.7671/j.issn.1001-411X.201907017]
- Szegedy C, Liu W, Jia YQ, et al. Going deeper with convolutions. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 1-9.
- Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*. Lille: ICML, 2015. 448-456.
- 孙鹏, 陈桂芬, 曹丽英. 基于注意力卷积神经网络的大豆害虫图像识别. *中国农机化学报*, 2020, 41(2): 171-176. [doi: 10.13733/j.jcam.issn.2095-5553.2020.02.26]
- Cheng X, Zhang YH, Chen YQ, et al. Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, 2017, 141: 351-356. [doi: 10.1016/j.compag.2017.08.005]
- He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770-778.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe. 2012. 1097-1105.
- 陈峰, 谷俊涛, 李玉磊, 等. 基于机器视觉和卷积神经网络的东北寒地玉米害虫识别方法. *江苏农业科学*, 2020, 48(18): 237-244.
- Pattnaik G, Shrivastava VK, Parvathi K. Transfer learning-based framework for classification of pest in tomato plants. *Applied Artificial Intelligence*, 2020, 34(13): 981-993. [doi: 10.1080/08839514.2020.1792034]
- Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2261-2269.
- Ma NN, Zhang XY, Zheng HT, et al. ShuffleNetV2: Practical guidelines for efficient CNN architecture design. *Proceedings of the 15th European Conference on Computer Vision*. Munich: IEEE, 2018. 122-138.
- Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 2818-2826.
- Amid E, Warmuth MK, Anil R, et al. Robust bi-tempered logistic loss based on Bregman divergences. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver: NeurIPS, 2019. 1344.
- Li YF, Wang HX, Dang LM, et al. Crop pest recognition in natural scenes using convolutional neural networks.

- Computers and Electronics in Agriculture, 2020, 169: 105174. [doi: [10.1016/j.compag.2019.105174](https://doi.org/10.1016/j.compag.2019.105174)]
- 20 Montserrat DM, Lin Q, Allebach J, *et al.* Training object detection and recognition CNN models using data augmentation. *Electronic Imaging*, 2017, 29(10): 27–36. [doi: [10.2352/ISSN.2470-1173.2017.10.IMAWM-163](https://doi.org/10.2352/ISSN.2470-1173.2017.10.IMAWM-163)]
- 21 Xu K, Ba J, Kiros R, *et al.* Show, attend and tell: Neural image caption generation with visual attention. *Proceedings of the 32nd International Conference on Machine Learning*. Lille: ICML, 2015. 2048–2057.
- 22 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. *Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 13708–13717.
- 23 Fu JL, Zheng HL, Mei T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 4476–4484.
- 24 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 25 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 1314–1324.
- 26 曾伟辉. 面向农作物叶片病害鲁棒性识别的深度卷积神经网络研究 [博士学位论文]. 合肥: 中国科学技术大学, 2018.
- 27 岳有军, 李雪松, 赵辉, 等. 基于改进 VGG 网络的农作物病害图像识别. *农机化研究*, 2022, 44(6): 18–24. [doi: [10.3969/j.issn.1003-188X.2022.06.003](https://doi.org/10.3969/j.issn.1003-188X.2022.06.003)]
- 28 Long PM, Servedio RA. Random classification noise defeats all convex potential boosters. *Machine Learning*, 2010, 78(3): 287–304. [doi: [10.1007/s10994-009-5165-z](https://doi.org/10.1007/s10994-009-5165-z)]
- 29 Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts. arXiv:1608.03983, 2016.
- 30 Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv:1711.05101, 2017.
- 31 周志华. 机器学习. 北京: 清华大学出版社, 2016. 230–245.
- 32 Saini R, Jha NK, Das B, *et al.* ULSAM: Ultra-lightweight subspace attention module for compact convolutional neural networks. *Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision*. Snowmass: IEEE, 2020. 1616–1625.
- 33 Zomahoun DE. A semantic collaborative clustering approach based on confusion matrix. *Proceedings of the 2019 15th International Conference on Signal-Image Technology & Internet-based Systems (SITIS)*. Sorrento: IEEE, 2019. 688–692.
- 34 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
- 35 Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 1800–1807.
- 36 Iandola FN, Han S, Moskewicz MW, *et al.* SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size. arXiv:1602.07360, 2016.
- 37 Tan MX, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*. Long Beach: ICML, 2019. 6105–6114.
- 38 Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 618–626.

(校对责编: 牛欣悦)