

基于 MobileViT - PC - ASPP 和迁移学习的 果树害虫识别方法

张欢^{1,2} 周毅³ 王克俭^{1,2} 王超^{1,2} 李会平²

(1. 河北农业大学信息科学与技术学院, 保定 071001; 2. 河北省城市森林健康技术创新中心, 保定 071001;
3. 河北金融学院金融科技学院, 保定 071000)

摘要: 为提高果树害虫识别效果,及时做好防治措施,本研究以6种对果树危害程度较大的害虫为研究对象,针对自然环境下果树害虫识别背景复杂、害虫目标小检测难度大、与不同类别间特征相似度高问题,提出一种改进的轻量化 MobileViT - PC - ASPP 识别模型。该模型用 PConv (Partial convolution) 模块代替原模型 MobileViT 模块中部分标准卷积模块,其次修改 MobileViT 模块的特征融合策略,将输入特征、局部表达特征、全局表达特征进行拼接融合;删除网络第10层 MV2 模块和第11层 MobileViT 模块,使用改进空洞空间池化金字塔 (Atrous spatial pyramid pooling, ASPP) 模块进行代替,形成多尺度融合特征;此外,模型用 SiLU 激活函数代替 ReLU6 激活函数进行计算,最后基于 ImageNet 数据集进行迁移学习。实验结果表明,6类果树害虫图像识别准确率达 93.77%,参数量为 8.40×10^5 ,与改进前相比,识别准确率提高 7.5 个百分点,参数量降低 33.86%;与常用害虫 CNN 识别模型 AlexNet、ResNet50、MobileNetV2、ShuffleNetV2 相比识别准确率分别提高 8.25、4.78、7.27、7.41 个百分点,参数量分别减少 6.03×10^7 、 2.48×10^7 、 2.66×10^6 、 5.30×10^5 ;与 Transformer 识别模型 ViT、Swin Transformer 相比识别准确率分别提高 19.03、9.8 个百分点,参数量分别减少 8.56×10^7 、 2.75×10^7 。本研究适合部署在移动终端等有限资源环境,并且有助于实现对复杂背景下小目标果树害虫进行识别检测。

关键词: 果树害虫; 识别模型; PConv 模块; 融合策略; SiLU 激活函数; 空洞空间池化金字塔

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2024)11-0057-11

OSID:



Fruit Tree Pest Identification Method Based on MobileViT - PC - ASPP and Transfer Learning

ZHANG Huan^{1,2} ZHOU Yi³ WANG Kejian^{1,2} WANG Chao^{1,2} LI Huiping²

(1. College of Information Science and Technology, Hebei Agricultural University, Baoding 071001, China

2. Hebei Urban Forest Health Technology Innovation Center, Baoding 071001, China

3. School of Financial Technology, Hebei Finance University, Baoding 071000, China)

Abstract: In order to enhance the effectiveness of identifying pests in fruit trees and promptly implement preventive measures, focusing on six major pests that pose a significant threat to fruit trees, an improved lightweight MobileViT recognition model was proposed for the problems of complex background of fruit tree pest recognition in the natural environment, high difficulty of detecting the small target of the pests, and high feature similarity with the features between different categories. In enhancing the model, the partial convolution (PConv) module was employed to replace certain standard convolution modules in the original MobileViT module. Additionally, modifications were made to the feature fusion strategy within the MobileViT module, involving the concatenation fusion of input features, local expressive features, and global expressive features. The tenth layer MV2 module and the eleventh layer MobileViT module were removed, introducing an improved atrous spatial pyramid pooling (ASPP) module as a replacement, aiming to create multi-scale fusion features. Furthermore, the model adopted the SiLU activation function in lieu of the ReLU6 activation function for computations. Finally, the model underwent transfer learning

收稿日期: 2023-12-09 修回日期: 2024-01-11

基金项目: 国家自然科学基金项目(32171799)和河北省重点研发计划项目(22327404D)

作者简介: 张欢(1999—),男,硕士生,主要从事图像处理研究,E-mail: 2457581882@qq.com

通信作者: 王克俭(1971—),女,教授,主要从事图像处理研究,E-mail: wkj71@163.com

based on the ImageNet dataset. The experimental results indicated that the recognition accuracy of six categories of fruit tree pest images reached 93.77%, with a parameter count of 8.40×10^5 . In comparison with the previous version, the recognition accuracy was improved by 7.5 percentage points, while the parameter count was decreased by 33.86%. When compared with commonly used pest CNN recognition models, namely AlexNet, ResNet 50, MobileNetV2, and ShuffleNetV2, the proposed model achieved higher recognition accuracy by 8.25, 4.78, 7.27 and 7.41 percentage points, respectively, with parameter counts lowered by 6.03×10^7 , 2.48×10^7 , 2.66×10^6 and 5.30×10^5 , respectively. Compared with Transformer recognition models such as ViT and Swin Transformer, the accuracy was higher by 19.03 and 9.8 percentage points, respectively, with parameter counts lowered by 8.56×10^7 and 2.75×10^7 . The research was suitable for deployment in environments with limited resources, such as mobile terminals, which can contribute to the effective identification and detection of small target pests in fruit trees amidst complex backgrounds.

Key words: fruit tree pests; recognition model; PConv module; convergence strategy; SiLU activation function; atrous spatial pyramid pooling

0 引言

加强果树全程病虫害管理是提升果品的有效措施,同时还能提高果树树势,延缓树体衰老^[1]。近年来,北方虫灾频繁出现,不仅为果树种植带来巨大经济损失,对于果树生存的生态环境也是一种很大的阻碍^[2],依靠人工观察并鉴定识别虫害耗费大量精力和时间,且效率低下^[3-5],随着计算机技术的发展,利用计算机图像处理技术获取植物病虫害信息可以大大提高植物病虫害的识别效率^[6]。

基于机器学习的识别研究受到广泛关注,其有效解决了人工识别的难题^[7-10]。传统的基于机器学习的方法需要手工设计特征,其过程繁琐费时费力,特征提取困难^[11],且算法具有专用性,泛化能力较差^[12]。近年来,深度学习技术越来越受到青睐,卷积神经网络(Convolutional neural network, CNNs)作为深度学习的代表算法之一^[13-14],具有良好的特征提取能力和泛化能力。FUENTES等^[15]提出一种基于深度学习的番茄植株虫害检测方法;陈洋^[16]通过比较 VGG-16 和多功能卷积神经网络,实现了 21 种虫害疾病检测;余小东等^[17]提出一种基于 ResNet50 模型的 CDCNNV2 算法,实现了对病虫害严重程度的分类。

上述基于深度学习技术的识别任务尽管具有较好的识别效果,但其计算成本较高,需要在保证识别效果的同时使模型轻量化,文献[18-22]使用轻量化模型有效节约了计算成本和模型内存占用量,对病虫害的识别效果较好。然而,在自然环境下,害虫生存在果树枝干、树叶、果实等不同部位中,且受光照、温度和遮挡等因素干扰,使害虫目标较小且难以发现。因此,需要进一步研究复杂背景下害虫识别,并在保持模型轻量化的同时提高识别准确率。

本文以自然环境下果树常见的 6 种害虫为研究对象,提出一种基于 MobileViT-PC-ASPP 的果树害虫识别模型。对 MobileViT 模块的轻量化和特征提取方面进行改进,引入改进的空洞空间池化金字塔(Atrous spatial pyramid pooling, ASPP)^[23]模块进行多尺度融合特征提取,有效避免害虫细节特征丢失,使用 SiLU 激活函数^[24]解决训练过程中神经元坏死问题,使改进后模型保持轻量化特点且提高对复杂背景下害虫识别效果,以实现在移动终端上对复杂背景下小目标果树害虫识别检测。

1 数据采集与处理

1.1 害虫数据集构建

由于不同成长阶段的害虫对果树危害程度不同,采集危害程度较大的阶段害虫进行研究,选用幼虫阶段的春尺蠖(*Apocheima cinerarius*)、黄刺蛾(*Cindocampa flavescens*)、褐边绿刺蛾(*Latoia consocia*)、舞毒蛾(*Lymantria dispar*)、黄褐天幕毛虫(*Malacosoma neustria testacea*)、咖啡木蠹蛾(*Neuzera coffeae* Niether) 6 种害虫作为研究对象,图像采集地点为雄安新区“千年秀林”(39.00515°N, 116.047255°E),针对其果树区域进行害虫样本采集,使用华为 Mate30、iPhone 11、小米 12 型移动设备进行拍摄。为使害虫样本更丰富、更接近自然环境,采集受到不同光照、温度等自然环境影响的害虫,采集时间分为 08:00—11:00 和 17:00—18:00,采集人员使用自主开发的 APP“雄安新区千年秀林生物调查系统”采集图像并自动上传至服务器。另外,为丰富实验样本,充分利用 Baidu、Google 等搜索引擎搜集数据,对样本进行补充。

共采集 6 类害虫图像 2 010 幅,分别为春尺蠖幼虫 350 幅、黄刺蛾幼虫 334 幅、褐边绿刺蛾幼虫

330 幅、舞毒蛾幼虫 332 幅、黄褐天幕毛虫幼虫 332 幅、咖啡木蠹蛾幼虫 332 幅。

害虫图像样例如图 1 所示,可以看出害虫处在复杂背景下且害虫目标较小,黄刺蛾和褐边绿刺蛾、舞毒蛾和黄褐天幕毛虫本特征相似度高,而在复杂背景干扰下更加难以识别。



图 1 害虫样例

Fig. 1 Pest examples

1.2 数据集扩充

将原始数据按照 8:2 划分为训练集和验证集,保证训练集和验证集的完全独立且没有重复。由于果树害虫的发生是动态的,因此对于大样本数据的采集具有挑战性,为解决模型在小样本数据集训练时易发生的过拟合问题,本研究采用离线数据增强方式将划分后的训练集和验证集采用水平翻转、垂直翻转、旋转、平移、模糊、调整亮度、椒盐噪声等方法扩充。为进一步提高模型泛化能力和鲁棒性,训

练前随机裁剪训练集图像并调整为 256 像素 × 256 像素,此外,为更好评估训练好的模型在现实场景中的表现,对验证集图像使用中心裁剪并调整为 256 像素 × 256 像素。扩充后数据集共 11 059 幅,其中训练集为 8 925 幅,验证集为 2 134 幅,扩充结果如表 1 所示。

表 1 果树害虫数据

Tab. 1 Data on fruit tree pests

害虫类型	原数据		扩充后数据	
	训练集 图像数	验证集 图像数	训练集 图像数	验证集 图像数
春尺蠖	282	68	1 555	374
黄刺蛾	269	65	1 485	352
褐边绿刺蛾	266	64	1 463	352
舞毒蛾	266	66	1 474	352
黄褐天幕毛虫	266	66	1 474	352
咖啡木蠹蛾	266	66	1 474	352

2 识别方法

2.1 基于 MobileViT-PC-ASPP 的果树害虫识别方法

MobileViT^[25] 是 CNN 与 Transformer 的混合架构模型,在 ImageNet 数据集有着良好的分类效果,然而,在害虫识别时,由于害虫个体目标小,且普遍与周围复杂背景有着类似的保护色,导致害虫检测难度大,特征提取难度高,因此,本文改进 MobileViT 网络模型,提出 MobileViT-PC-ASPP 网络模型,结构如图 2 所示。

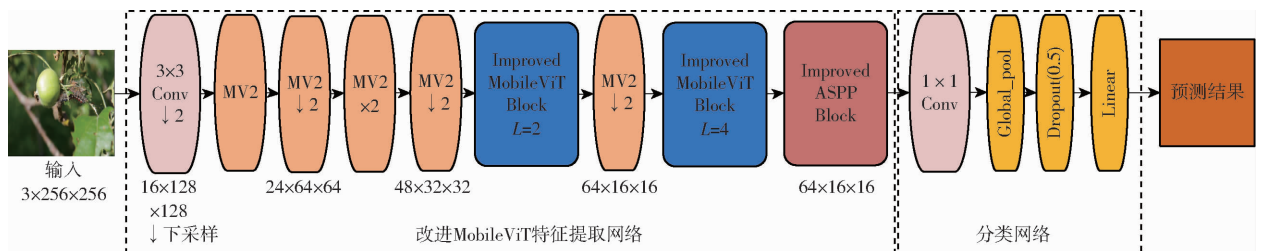


图 2 MobileViT-PC-ASPP 网络模型结构图

Fig. 2 MobileViT-PC-ASPP network model structure

(1) 将原模型中 MobileViT 模块局部特征提取中 3×3 标准卷积替换为 PConv 模块^[26]以降低模型计算量。此外,改进 MobileViT 模块中的特征融合策略,将输入特征、局部表达特征、全局表达特征进行拼接融合,强化模型对局部特征提取能力。

(2) 删除原网络第 10 层 MV2 模块和第 11 层 MobileViT 模块,引入改进的空洞空间卷积池化金字塔 ASPP 模块代替其进行特征提取,使模型在不降低图像分辨率的情况下增大感受野,从而更好地学

习图像上下文信息和全局特征。同时,充分利用并融合多个尺度提取特征,提高对小目标害虫细节特征提取。

(3) 使用 SiLU 激活函数代替 ReLU6 激活函数进行计算,解决训练过程中神经元坏死问题。

(4) 采用迁移学习训练方法,强化网络对小样本数据集训练效果,提升模型泛化能力。

2.2 MobileViT 模块改进

为减少网络计算量、提高模型识别速度、强化

模型特征提取能力,本文对 MobileViT 模块进行改进,改进后模块如图 3 所示,其中 L 表示

Transformer 数量,Unfold 为展开操作,Fold 为折叠操作。

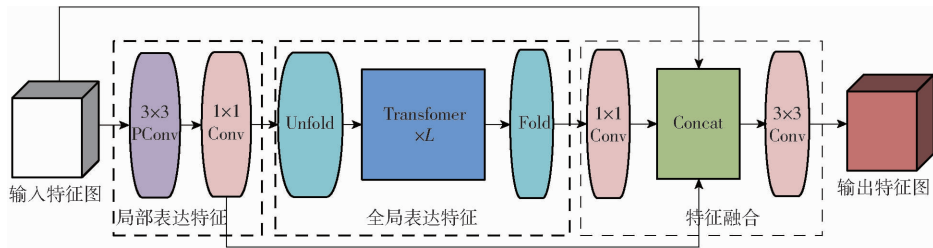


图 3 改进后 MobileViT 模块

Fig. 3 Improved MobileViT block

首先,将原模块局部特征表达中 3×3 标准卷积替换为 PConv 模块,减少网络计算量,其次为进一步强化网络特征提取能力,在特征融合中改变融合策略,将输入特征、局部表达特征、全局表达特征三者进行 Concat 融合强化 MobileViT 模块总体特征表达。

2.2.1 PConv 模块

PConv 模块如图 4 所示,与标准卷积不同,PConv 对其中部分通道进行卷积操作,此操作避免了不同通道产生的高度冗余。

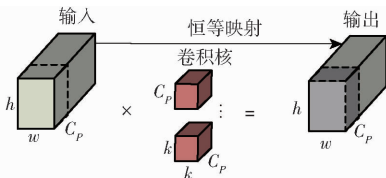


图 4 PConv 模块结构图

Fig. 4 Structure diagram of PConv module

本研究中选取输入特征图前段连续 $1/4$ 通道作为部分通道,通道数为 C_p ,为降低模型在害虫识别任务中计算复杂程度,保证任务过程中内存访问的连续性和规则性,将此部分通道作为特定的通道代表整个特征图进行卷积操作。PConv 的浮点运算量为 $h \times w \times k^2 \times C_p^2$,为标准卷积的 $1/16$;内存访问量为 $h \times w \times 12C_p$,为标准卷积的 $1/4$ 。因此将原模块局部特征表达中 3×3 标准卷积替换为 PConv 模块的方法能够有效降低网络的计算复杂度,使网络更加轻量化,提升模型对害虫识别的性能。

2.2.2 融合策略改进

原模型融合策略将输入特征和全局表达特征后的特征进行 Concat 融合,此融合策略得到的特征图缺少对图像局部特征的表达能力,因此改变融合策略,通过将输入特征、局部表达特征、全局表达特征三者进行 Concat 融合的方式,增强模型的特征提取能力。

假设 MobileViT 模块输入特征图 X 尺寸为 $H \times W \times C$,其中 H 、 W 、 C 分别为输入特征图的高度、宽

度、通道数,经过局部特征表达的 3×3 PConv 和 1×1 标准卷积提取局部特征,得到特征图尺寸为 $H \times W \times d$ 的 X_p , d 为更高维度的特征空间,然后 X_p 经过全局特征表达 Transformer(个数为 L) 得到特征图尺寸为 $H \times W \times d$ 的 X_T ,之后将 X_T 经过 1×1 标准卷积映射到与输入特征图相同维度 C 的特征图 X_F ,其尺寸为 $H \times W \times C$,将 X 、 X_p 、 X_F 进行 Concat 得到特征图 X_L ,即

$$X_L = \text{Concat}(X, X_p, X_F) \quad (1)$$

最后将特征图 X_L 经过一个 3×3 标准卷积进行特征融合得到最终特征图尺寸为 $H \times W \times C$ 的 Y 。

2.3 改进 ASPP 模块引入

由于在实际果树害虫识别任务中,害虫处在背景复杂的环境下且害虫个体目标较小,且在模型训练中分辨率的降低导致害虫细节特征难以提取,为此本文提出删除原网络第 10 层 MV2 模块和第 11 层 MobileViT 模块,使用 ASPP 模块进行特征提取,为使 ASPP 模块更加高效,能够更好地融合各尺度特征信息形成多尺度特征,提出一种改进的 ASPP 模块,结构如图 5 所示。

改进后 ASPP 模块由 3 个空洞率分别为 6、12、18 的 3×3 部分空洞卷积、 1×1 的卷积和 ASPP Pooling 3 部分组成,卷积步长大小均为 1,使用 1×1 卷积和不同空洞率(rate1、rate2 和 rate3)的 3×3 部分空洞卷积以捕捉不同尺度的语义信息,在不降低分辨率的同时增大感受野,最后将各部分特征图使用改进的融合方法进行融合,充分利用各尺度特征信息,实现害虫细节特征的提取。感受野尺寸为

$$s = (d - 1)(k - 1) \quad (2)$$

式中 s ——感受野尺寸

d ——空洞率

k ——卷积核大小

2.3.1 部分空洞卷积

将 ASPP 模块中空洞卷积中的 3×3 标准卷积替换为 3×3 部分卷积,即部分空洞卷积,如图 6 所

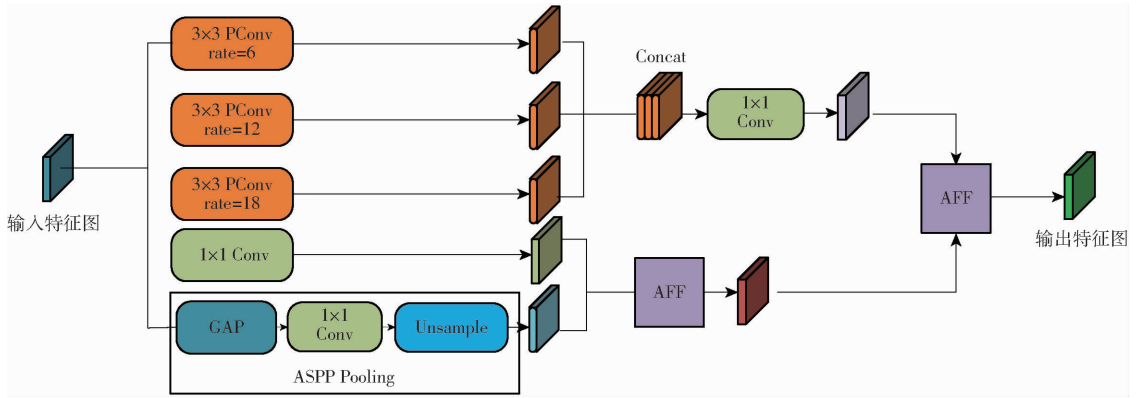


图 5 改进 ASPP 模块结构图

Fig. 5 Structure diagram of improved ASPP module

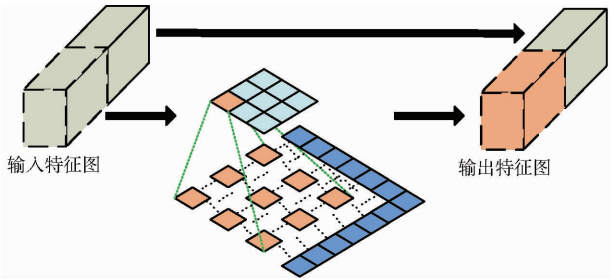


图 6 部分空洞卷积

Fig. 6 Partially atrous convolution

示,选取输入特征图前段连续 1/2 通道作为特定通道进行空洞卷积,将此特定通道代表整个特征图进行卷积操作,剩余 1/2 通道不进行卷积。此方法避免了通道间高度冗余,降低了 ASPP 计算复杂度,使 ASPP 更加高效。

2.3.2 ASPP Pooling

ASPP Pooling 部分由全局平均池化、1 × 1 的卷积、Unsample 上采样组成,对输入特征图进行全局平均池化操作,获得整个特征图的全局统计信息,通过 1 × 1 的卷积调整特征图维度,Unsample 将低分辨率的特征图恢复到原始输入图像的分辨率以便与其他分辨率特征图进行融合,本文利用双线性插值方法进行上采样,公式为

$$f(i+u, j+v) = (1-u)(1-v)f(i, j) + (1-u)vf(i, j+1) + u(1-v)f(i+1, j) + uvf(i+1, j+1) \quad (3)$$

式中 $f(i+u, j+v)$ ——目标位置 $(i+u, j+v)$ 处通过插值计算得到的像素值

$f(i, j)$ ——原始特征图上位置 (i, j) 处像素值

(i, j) ——原始位置的整数坐标

u, v ——相对偏移量,表示目标位置与原始位置 (i, j) 之间的相对偏移量

通过将每个像素值与对应的权重相乘,然后在公式中进行加权累加,得到目标位置 $(i+u, j+v)$ 处的像素值,实现特征图的上采样操作。

2.3.3 融合方法改进

原 ASPP 模块将各部分特征图进行 Concat 融合得到融合特征,这对各尺度特征图信息的利用具有一定的限制,为充分利用各尺度特征信息,本研究对 ASPP 模块的多尺度融合方法进行改进。

将 3 个空洞率分别为 6、12、18 的 3 × 3 部分空洞卷积进行 Concat 融合,捕捉不同尺度的语义信息得到特征 X_c ,将 1 × 1 卷积和 ASPP Pooling 部分得到的特征图进行注意力特征融合 (Attentional feature fusion, AFF)^[27],提高网络对原始输入特征和全局特征的利用能力,得到更加全面和丰富的特征 X_A ,将 X_c 和 X_A 再进行 AFF 融合得到最终的 ASPP 表示。

注意力特征融合如图 7 所示,首先,对输入特征 X 和特征 Y 进行逐元素求和和初始融合,接下来,通过两个分支提取注意力权重,其中一个分支使用全局平均池化和逐点卷积提取全局注意力,另一个分支直接使用逐点卷积提取局部注意力。全局特征的通道注意力计算式为

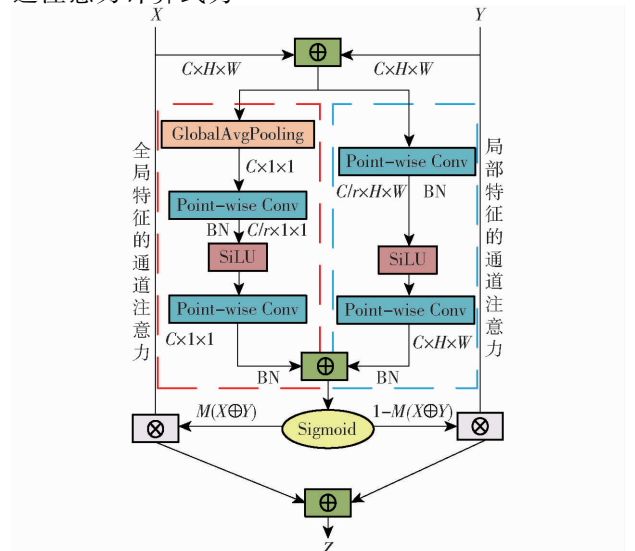


图 7 注意力特征融合

Fig. 7 Attentional feature fusion

$$G(X) = B(\text{PWconv2}(\delta(B(\text{PWconv1}(\text{GAP}(X)))))) \quad (4)$$

式中 GAP——全局平均池化操作

PWconv—— 1×1 逐点卷积

B ——BatchNorm 层

δ ——SiLU 激活函数

PWconv1 将输入特征通道数减少为原来的 $1/r$,

PWconv2 将通道数恢复为原输入通道数。

局部特征通道注意力计算式为

$$L(X) = B(\text{PWconv2}(\delta(B(\text{PWconv1}(X)))))) \quad (5)$$

与全局特征的通道注意力 $G(X)$ 不同的是,不对输入特征 X 进行全局平均池化操作。

将这两个分支得到的注意力权重进行逐元素求和,并经过 Sigmoid 运算,得到融合权重。最后,利用融合权重对输入特征进行加权平均,得到最终的融合特征 Z ,表达式为

$$Z = M(X \oplus Y) \otimes X + (1 - M(X \oplus Y)) \otimes Y \quad (6)$$

式中 \oplus ——特征图逐元素求和

\otimes ——特征图对应元素相乘

$M(X \oplus Y)$ ——融合权重,其值为 $0 \sim 1$ 之间的实数

2.4 SiLU 激活函数

ReLU6 计算公式为

$$\text{ReLU6}(x) = \min(6, \max(0, x)) \quad (7)$$

ReLU6 激活函数对于输入小于 0 或大于 6 的情况,输出为 0;对于输入大于 0 且小于 6 情况,输出为 x 本身,其具有更强的稀疏性和鲁棒性,但会导致一些梯度问题,例如在接近截止值 6 时,梯度为 0,这会影响模型的收敛性,且仍然存在某些神经元在整个训练过程中一直无法被激活的问题。

为解决训练过程中神经元坏死问题,使用 SiLU 激活函数进行计算。SiLU (Sigmoid - Weighted Linear Unit) 计算公式为

$$\text{SiLU}(x) = x \text{sigmoid}(x) \quad (8)$$

SiLU 与 ReLU6 激活函数对比如图 8 所示。其在整个实数域都可导和平滑,在许多情况下达到自归一化的效果,即使在深度网络中使用也可以保持较稳定的输出,有利于确保在不同层之间保持激活的平衡,保留 ReLU6 激活函数优点的同时能够避免某些神经元在整个训练过程中一直无法被激活,即神经元坏死的问题。

2.5 迁移学习训练方法

为强化网络对小样本数据训练效果,提升模型泛化能力,使本文模型更加适应在不同背景下的害虫识别,使用迁移学习训练方法。

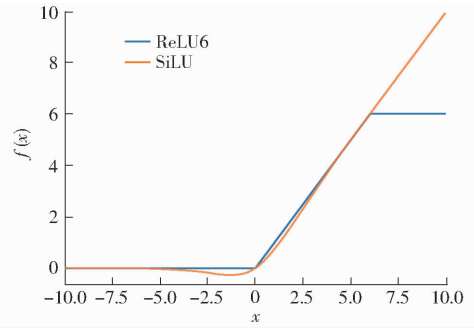


图 8 SiLU 与 ReLU6 激活函数比较

Fig. 8 Comparison of activation function between SiLU and ReLU6

采用全部权重迁移的迁移方法,即在 ImageNet 数据集预训练的模型基础上,不冻结权重,继续训练和优化模型,以适应果树害虫的识别。这种方法利用了 ImageNet 数据集上学到的通用特征表示,这些特征对于各种图像识别任务都具有较强的表达能力。通过在害虫识别任务上进行进一步训练,模型通过调整权重来适应特定任务的要求。

3 结果分析

3.1 实验环境及参数设置

在 Ubuntu 18.04、64 位操作系统上进行实验,采用 Pytorch 1.7.1 框架,使用的 GPU 显卡为 NVIDIA Quadro RTX 5000,运行内存为 16 GB,中央处理器为 Intel (R) Xeon (R) Gold 5220,主频为 2.20 GHz,编程语言为 Python 3.8, CUDA 版本为 10.0。为最大化发挥硬件性能,使模型训练处于最优环境,经多次反复实验,设置初始学习率 (Learning rate) 为 0.000 2,采用 AdamW 优化器实现对模型参数的优化和正则化,其权重衰减设置为 0.01。批处理大小 (Batch size) 设置为 32,训练周期 (epoch) 设置为 100,为缓解过拟合的发生,在全连接层网络随机失活 (Dropout) 参数设置为 0.5。

3.2 评价指标

使用参数量衡量模型大小,使用验证集准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、准确率与召回率加权调和平均值 (F1 值) 来评价模型识别能力。

3.3 改进前后 ASPP 模块性能对比实验

为验证改进后 ASPP 模块性能,在原模型基础上仅使用原 ASPP 模块和改进后 ASPP 模块对原模型进行改进,并使用 SiLU 激活函数进行计算,改进前后模型在验证集上的实验结果如表 2 所示。

由表 2 可知,由于只对特征图 $1/2$ 的通道进行空洞卷积,模型参数量减少 9.0×10^4 ,对 ASPP 模块

表 2 改进前后 ASPP 模块对比结果

Tab. 2 Comparison results of ASPP modules before and after improvement

模块	准确率/ %	精确率/ %	召回率/ %	F1 值/ %	参数量
ASPP	88.33	88.52	88.38	88.45	9.1×10^5
改进 ASPP	89.13	89.72	89.20	89.45	8.2×10^5

融合方法进行改进后,有效地整合了不同尺度的特征信息,准确率、精确率、召回率、F1 值分别提升 0.80、1.20、0.82、1.00 个百分点,由此验证本文提出的改进 ASPP 模块的有效性。

3.4 SiLU 激活函数对模型影响

为验证 SiLU 激活函数对 MobileViT 模型影响,分别使用 SiLU 激活函数和 ReLU6 激活函数在自建数据集上进行果树害虫识别,将两组实验结果进行对比,模型准确率如图 9 所示,使用 SiLU 激活函数的模型训练效果较好,准确率较高、收敛较快。在验证集准确率、精确率、召回率、F1 值结果如表 3 所示。

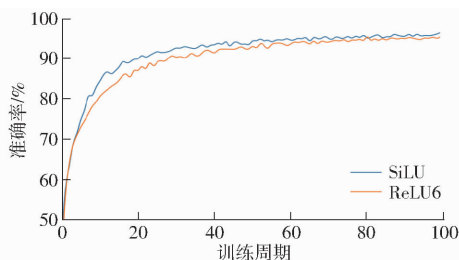


图 9 SiLU 和 ReLU6 激活函数训练准确率曲线

Fig. 9 Training accuracy curve of SiLU and ReLU6 activation functions

表 3 不同激活函数结果对比

Tab. 3 Comparison of results of different activation functions

激活函数	准确率	精确率	召回率	F1 值
ReLU6	86.27	86.37	86.32	86.34
SiLU	87.49	88.10	87.45	87.77

由表 3 可知,使用 SiLU 激活函数模型准确率、精确率、召回率、F1 值分别为 87.49%、88.10%、87.45%、87.77%,与 ReLU6 相比,分别提升 1.22、1.73、1.13、1.43 个百分点。结果表明 SiLU 激活函数能有效解决 ReLU6 激活函数产生神经元死亡和梯度消失的问题,使模型拥有更强大的表达能力和非线性映射能力。

3.5 消融实验

为探究 PConv 模块、改进融合策略、改进 ASPP 模块对模型 MobileViT 在识别果树害虫上的性能提升,在均使用 SiLU 激活函数基础上进行消融实验,

结果如表 4 所示。

表 4 消融实验结果

Tab. 4 Results of ablation experiment

PConv 模块	改进融 合策略	改进 ASPP 模块	准确 率/%	F1 值/ %	参数量
×	×	×	87.49	87.77	1.27×10^6
√	×	×	87.44	87.69	1.16×10^6
×	√	×	88.43	88.59	1.41×10^6
×	×	√	89.13	89.45	8.20×10^5
√	√	×	88.85	88.89	1.31×10^6
√	×	√	88.38	88.39	7.60×10^5
×	√	√	89.18	89.30	8.90×10^5
√	√	√	89.50	89.52	8.40×10^5

注:√表示使用此方法,×表示未使用此方法。

由表 4 可知,MobileViT 模块中使用 PConv 模块代替局部特征表达中的 3×3 标准卷积仅使准确率损失 0.05 个百分点的情况下让模型更加轻量化,参数量减少 1.10×10^5 ,验证了 PConv 模块对部分通道进行卷积能够减小模型内存占用量的能力。在原网络中只改变 MobileViT 模块的融合策略后,提升了模型对局部特征提取的能力,使得模型准确率提升 0.94 个百分点,F1 值提升 0.82 个百分点,但是由于增加了对局部特征的融合,影响了模型内存占用量,使参数量提升 1.40×10^5 。在删除原网络第 10 层 MV2 模块和第 11 层 MobileViT 模块,引入改进的 ASPP 模块之后,由于在不降低图像分辨率的情况下增大感受野,更好地学习了图像的上下文信息和全局特征。同时,模块更好地利用并融合了多尺度特征,增强了对害虫细节特征的提取,使得模型准确率提升 1.64 个百分点,F1 值提升 1.68 个百分点,此外,在删除操作后引入轻量化改进后的 ASPP 模块,使模型参数量减少 4.50×10^5 。

通过结合不同方法对模型进行改进后,改进模型的性能得到更均衡的提升。PConv 模块、删除网络部分层以及使用轻量化改进的 ASPP 模块成功减小了模型内存占用量,改进的融合策略和引入改进融合方法的 ASPP 模块增强了模型特征提取能力,与原模型相比,准确率提升 2.01 个百分点,F1 值提升 1.75 个百分点,参数量减少 4.30×10^5 。

3.6 不同模型对比实验

为进一步探究本文模型的有效性,选择 AlexNet^[28]、ResNet50^[29]、MobileNetV2^[30]、ShuffleNetV2_x0.5^[31] 4 种 CNN 模型和 ViT^[32]、Swin Transformer^[33] 2 种 Transformer 模型对本文害虫数据集进行识别,与本研究模型进行对比实验,结果如表 5 所示。

表5 不同模型结果对比

Tab.5 Comparison of results from different models

模型	准确 率/%	精确 率/%	召回 率/%	F1 值/ %	参数量
AlexNet	85.52	85.83	85.63	85.73	6.11×10^7
ResNet50	88.99	89.40	89.02	89.21	2.56×10^7
MobileNetV2	86.50	87.30	86.62	86.96	3.50×10^6
ShuffleNetV2_x0.5	86.36	87.07	86.41	86.74	1.37×10^6
ViT	74.74	75.48	74.83	75.16	8.64×10^7
Swin Transformer	83.97	84.72	84.00	84.36	2.83×10^7
MobileViT	86.27	86.37	86.32	86.34	1.27×10^6
MobileViT-PC-ASPP (未使用迁移学习)	89.50	89.50	89.55	89.52	8.40×10^5
MobileViT-PC-ASPP	93.77	94.05	93.75	93.90	8.40×10^5

由表5可知,本研究提出模型 MobileViT-PC-ASPP 使用迁移学习训练方法前后都具有最佳的性能,使用迁移学习训练方法后模型准确率达 93.77%,精确率达 94.05%,召回率达 93.75%,F1 值达 93.90%,参数量为 8.40×10^5 ,与 MobileViT 即改进前模型相比,识别准确率提高 7.5 个百分点,参数量降低 33.86%,与常用害虫 CNN 识别模型 AlexNet、ResNet50、MobileNetV2、ShuffleNetV2 相比识别准确率分别提高 8.25、4.78、7.27、7.41 个百分点,参数量分别减少 6.03×10^7 、 2.48×10^7 、 2.66×10^6 、 5.30×10^5 ,与 Transformer 识别模型 ViT、Swin Transformer 相比识别准确率分别提高 19.03、9.8 个百分点,参数量分别减少 8.56×10^7 、 2.75×10^7 。这是由于与 CNN 网络 AlexNet、ResNet50、MobileNetV2、ShuffleNetV2_x0.5 和 Transformer 网络

ViT、Swin Transformer 相比,展现了 MobileViT 网络结合 Transformer 和 CNN 混合架构的优势,CNN 的轻量化和 Transformer 的自注意力机制与全局感受野,使得模型更加轻量化并且兼顾准确率,在此基础上本文提出的改进网络中 MobileViT 模块的融合策略增强了学习局部特征的能力,引入改进 ASPP 模块增强了模型对害虫细节特征的提取,替换激活函数解决原激活函数存在的训练过程中神经元坏死问题,使用迁移学习训练方法使模型更加适应在不同背景下的害虫识别,以上方法有效提升了模型识别准确率。使用 PConv 模块、删除网络部分层、对 ASPP 模块提出使用部分空洞卷积,有效减少了模型参数量。综上所述,本文研究方法性能更优,实现了对害虫更高效准确地识别。

3.7 可视化分析

为进一步验证本文模型的有效性,采用 Grad-CAM 技术对害虫识别进行可视化分析,结果如图 10 所示,对原模型、本文模型中 ASPP 模块改进前和改进后 3 种激活特征图进行对比分析。

由可视化结果可知,原模型在识别害虫时容易受到复杂背景的干扰,聚焦部位包含过多背景信息,甚至由于害虫目标小且与周围环境具有相似的保护色,导致模型难以准确聚焦到害虫区域。以图中春尺蠖为例,由于其个体小且与周围环境颜色相似,模型在识别中未能准确聚焦到害虫区域,严重影响识别效果。在本文模型中引入的 ASPP 模块未改进时,模型改善了分辨率降低导致

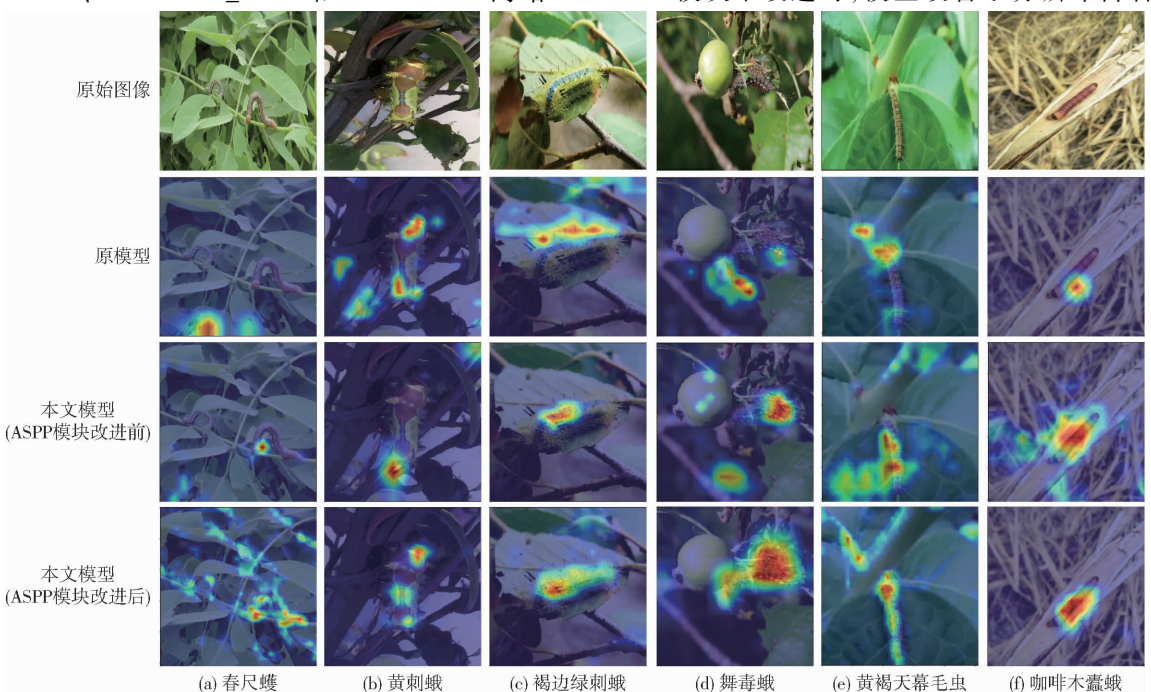


图10 可视化结果

Fig.10 Visualized results

的提取害虫细节特征不足的问题,降低了复杂背景干扰的影响,聚焦区域逐渐集中到害虫区域,但仍存在聚焦区域较小的问题。本文模型引入的 ASPP 模块改进后,模型保留了原 ASPP 模块的优点,同时更加充分地利用多个尺度的特征信息,这使得聚焦区域能够更集中地定位在害虫所在位置,并更关注害虫细节特征。

可视化结果表明,本文模型具备抗复杂背景干扰能力,对小目标害虫有着更加准确高效的特征提取能力,实现了对复杂背景下小目标害虫的识别。

3.8 迁移学习训练方法性能验证

为验证本文研究模型采用迁移学习训练方法后的性能,对本文模型在使用与不使用迁移学习训练方法下的结果进行对比,在验证集识别结果如图 11 所示,采用迁移学习训练方法的改进模型准确率增长更快,短时间内获得的准确率更高,最终达到的准确率更高。因此,改进模型采用迁移训练学习可以提高模型训练速率,提高模型准确率。

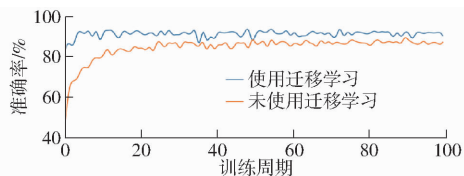


图 11 改进模型采用迁移学习训练方法前后结果对比

Fig. 11 Comparison of results before and after using transfer learning training method to improve model

为进一步验证迁移学习后本文模型的性能,与原模型在验证集上的混淆矩阵如图 12 所示,其中 0 表示春尺蠖,1 表示黄刺蛾,2 表示褐边绿刺蛾,3 表示舞毒蛾,4 表示黄褐天幕毛虫,5 表示咖啡木蠹蛾。

通过比较混淆矩阵对角线预测正确样本数量可知,改进后模型提升了对害虫的识别能力,对春尺蠖、褐边绿刺蛾、舞毒蛾、黄褐天幕毛虫的识别准确率分别提升 13.10、9.94、11.08、12.50 个百分点,但对黄刺蛾和咖啡木蠹蛾的识别准确率下降 0.28、1.70 个百分点,原因为模型提高了对害虫细节特征提取,使得改进后模型将处在复杂环境下受到光照不均和不同角度的害虫错误识别为其他害虫。

4 结论

(1)通过引入 PConv 模块、改进融合策略、引入

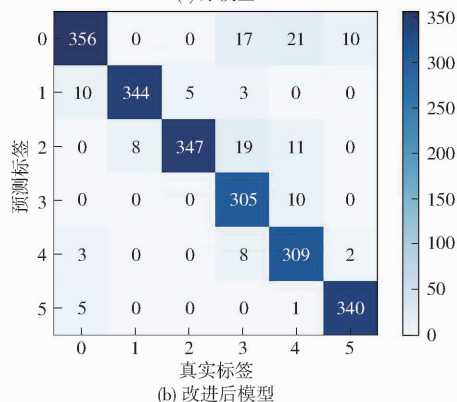
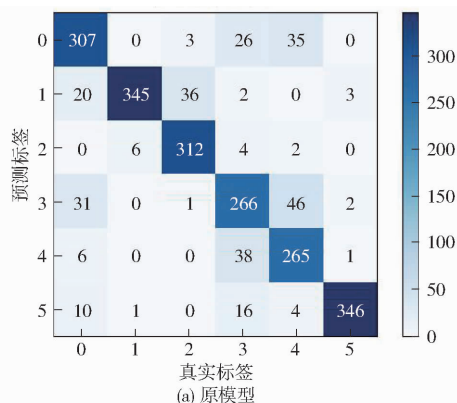


图 12 混淆矩阵

Fig. 12 Confusion matrices

改进空洞空间池化金字塔 ASPP 模块对 MobileViT 进行改进,并且网络采用 SiLU 激活函数进行计算,提出一种基于 MobileViT-PC-ASPP 的果树害虫识别模型,并使用迁移学习方法训练该模型。

(2)实验结果表明,本文模型在保证模型轻量化的同时能够提高对 6 种果树主要害虫的识别准确率,该模型参数量为 8.40×10^5 ,识别准确率可达 93.77%,与改进前相比,识别准确率提高 7.5 个百分点,参数量降低 33.86%;与常用害虫 CNN 识别模型 AlexNet、ResNet50、MobileNetV2、ShuffleNetV2 相比识别准确率分别提高 8.25、4.78、7.27、7.41 个百分点,参数量分别减少 6.03×10^7 、 2.48×10^7 、 2.66×10^6 、 5.30×10^5 ;与 Transformer 识别模型 ViT、Swin Transformer 相比,识别准确率分别提高 19.03、9.8 个百分点,参数量分别减少 8.56×10^7 、 2.75×10^7 。

(3)本文模型在实现轻量化的前提下,能有效识别果树害虫,且具有良好识别效果,为果树害虫防治提供了新思路。

参考文献

- [1] 金阔布. 北方果树病虫害防治的措施研究[J]. 河北农业, 2022(12): 83-84.
- [2] 郑海祥. 北方果树害虫主要天敌种类及其保护利用[J]. 新农业, 2021(19): 32.
- [3] 刘玉耀, 彭琼尹. 基于卷积神经网络和迁移学习的瓯柑病虫害识别研究[J]. 热带农业科学, 2022, 42(9): 64-70. LIU Yuyao, PENG Qiongyin. Research on diseases and insect pests identification of ougan based on convolution neural network and transfer learning[J]. Chinese Journal of Tropical Agriculture, 2022, 42(9): 64-70. (in Chinese)

- [4] 黄林生,罗耀武,杨小冬,等. 基于注意力机制和多尺度残差网络的农作物病害识别[J]. 农业机械学报,2021,52(10):264-271.
HUANG Linsheng, LUO Yaowu, YANG Xiaodong, et al. Crop disease recognition based on attention mechanism and multi-scale residual network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(10):264-271. (in Chinese)
- [5] 尚皓玺,郭小燕,朱恒宇. 基于迁移学习与 GhostNet 模型的农业害虫图像识别研究[J]. 软件导刊,2022,21(11):137-143.
SHANG Haoxi, GUO Xiaoyan, ZHU Hengyu. Research on image recognition of agricultural pests based on transfer learning and GhostNet model[J]. Software Guide, 2022, 21(11):137-143. (in Chinese)
- [6] 安强强,张峰,李赵兴,等. 基于深度学习的植物病虫害图像识别[J]. 农业工程,2018,8(7):38-40.
AN Qiangqiang, ZHANG Feng, LI Zhaoxing, et al. Plant diseases and insect pests images identification based on deep learning[J]. Agricultural Engineering, 2018, 8(7):38-40. (in Chinese)
- [7] 张一倩. 基于计算机视觉的草莓叶片病虫害识别研究[D]. 晋中:山西农业大学,2020.
ZHANG Yiqian. Identification of strawberry leaf diseases and insect pests based on computer vision[D]. Jinzhong: Shanxi Agricultural University, 2020. (in Chinese)
- [8] 马佳佳,陈友鹏,王克强,等. 基于优化 SVM 的虫害图像识别研究[J]. 中国粮油学报,2022,37(5):10-15.
MA Jiajia, CHEN Youpeng, WANG Keqiang, et al. Research on insect pest image recognition based on optimized SVM[J]. Journal of the Chinese Cereals and Oils Association, 2022, 37(5):10-15. (in Chinese)
- [9] 朱莉,罗靖,徐胜勇,等. 基于颜色特征的油菜害虫机器视觉诊断研究[J]. 农机化研究,2016,38(6):55-58,121.
ZHU Li, LUO Jing, XU Shengyong, et al. Machine vision recognition of rapeseed pests based on color feature[J]. Journal of Agricultural Mechanization Research, 2016, 38(6):55-58, 121. (in Chinese)
- [10] 陈林琳,姜大庆,蔡银杰,等. 基于计算机视觉的芋头病害监测与识别[J]. 农机化研究,2020,42(6):224-229.
CHEN Linlin, JIANG Daqing, CAI Yinjie, et al. Monitoring and identification of the taro disease based on the computer vision[J]. Journal of Agricultural Mechanization Research, 2020, 42(6):224-229. (in Chinese)
- [11] 翟肇裕,曹益飞,徐焕良,等. 农作物病虫害识别关键技术研究综述[J]. 农业机械学报,2021,52(7):1-18.
ZHAI Zhaoyu, CAO Yifei, XU Huanliang, et al. Review of key techniques for crop disease and pest detection[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(7):1-18. (in Chinese)
- [12] 孙亮,柯宇航,刘辉,等. 计算机视觉技术在植物病害识别上的研究进展[J]. 热带生物学报,2022,13(6):651-658.
SUN Liang, KE Yuhang, LIU Hui, et al. Advances in recognition of plant diseases based on computer vision[J]. Journal of Tropical Biology, 2022, 13(6):651-658. (in Chinese)
- [13] HUANG G, LIU Z, VANDERMAATEN L, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:4700-4708.
- [14] GU J, WANG Z, KUEN J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77:354-377.
- [15] FUENTES A, YOON S, KIM S C, et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition[J]. Sensors, 2017, 17(9):2022.
- [16] 陈洋. 基于卷积神经网络的农作物病虫害图像分类研究[D]. 南昌:江西农业大学,2019.
CHEN Yang. Image classification of crop diseases and pests based on convolutional neural network[D]. Nanchang: Jiangxi Agricultural University, 2019. (in Chinese)
- [17] 余小东,杨孟辑,张海清,等. 基于迁移学习的农作物病虫害检测方法研究与应用[J]. 农业机械学报,2020,51(10):252-258.
YU Xiaodong, YANG Mengji, ZHANG Haiqing, et al. Research and application of crop diseases detection method based on transfer learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10):252-258. (in Chinese)
- [18] 张鹏程,余勇华,陈传武,等. 基于改进 MobileNetV2 的柑橘害虫分类识别方法[J]. 华中农业大学学报,2023,42(3):161-168.
ZHANG Pengcheng, YU Yonghua, CHEN Chuanwu, et al. A classification and recognition method for citrus insect pests based on improved MobileNetV2[J]. Journal of Huazhong Agricultural University, 2023, 42(3):161-168. (in Chinese)
- [19] 彭红星,何慧君,高宗梅,等. 基于改进 ShuffleNetV2 模型的荔枝病虫害识别方法[J]. 农业机械学报,2022,53(12):290-300.
PENG Hongxing, HE Huijun, GAO Zongmei, et al. Litchi diseases and insect pests identification method based on improved ShuffleNetV2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12):290-300. (in Chinese)
- [20] CHEN J, CHEN W, ZEB A, et al. Crop pest recognition using attention-embedded lightweight network under field conditions[J]. Applied Entomology and Zoology, 2021, 56(4):427-442.
- [21] 曹欢,方睿. 基于深度学习的芒果病虫害分类识别[J]. 计算机技术与发展,2023,33(10):115-119.
CAO Huan, FANG Rui. Classification and identification of mango diseases and pests based on deep learning[J]. Computer Technology and Development, 2023, 33(10):115-119. (in Chinese)
- [22] 马维娣,吴钦木. 基于改进轻量化网络 MobileViT 的苹果树叶病害识别[J]. 江苏农业科学,2024,52(3):229-236.
- [23] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018:801-818.
- [24] 齐向明,董旭. 改进 Yolov7-tiny 的钢材表面缺陷检测算法[J]. 计算机工程与应用,2023,59(12):176-183.
QI Xiangming, DONG Xu. Improved Yolov7-tiny algorithm for steel surface defect detection[J]. Computer Engineering and Applications, 2023, 59(12):176-183. (in Chinese)

- [25] MEHTA S, RASTEGARI M. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer [J]. arXiv Preprint, arXiv:2110.02178, 2021.
- [26] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021 – 12031.
- [27] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3560 – 3569.
- [28] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84 – 90.
- [29] 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.
- [30] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510 – 4520.
- [31] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 116 – 131.
- [32] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words; transformers for image recognition at scale [J]. arXiv Preprint, arXiv:2010.11929, 2020.
- [33] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012 – 10022.

(上接第 56 页)

- [20] SONG Y, ZHANG H, LI J, et al. High-accuracy maize disease detection based on attention generative adversarial network and few-shot learning [J]. Plants, 2023, 12(17): 3105.
- [21] 翟雪东, 韩文霆, 马伟童, 等. 基于改进 AdvSemiSeg 的半监督遥感影像作物制图方法 [J]. 农业机械学报, 2024, 55(8): 196 – 204.
ZHAI Xuedong, HAN Wenting, MA Weitong, et al. Semi-supervised network for remote sensing crop mapping based on improved AdvSemiSeg [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(8): 196 – 204. (in Chinese)
- [22] ZHOU H, FANG Z, WANG Y, et al. Image generation of tomato leaf disease identification based on small-ACGAN [J]. Computers, Materials & Continua, 2023, 76(1): 175 – 194.
- [23] CHEN J, HU H, YANG J. Plant leaf disease recognition based on improved SinGAN and improved ResNet34 [J]. Frontiers in Artificial Intelligence, 2024, 7: 1414274.
- [24] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 286 – 301.
- [25] LIU Y, WANG Y, LI N, et al. An attention-based approach for single image super resolution [C] // 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018: 2777 – 2784.
- [26] 周齐, 杨晓君, 林浩申, 等. 基于 WGAN-GP 的搜索式路径规划算法 [J]. 计算机应用研究, 2022, 39(12): 3626 – 3630, 3638.
ZHOU Qi, YANG Xiaojun, LIN Haoshen, et al. Search-based path planning algorithm based on WGAN-GP [J]. Application Research of Computers, 2022, 39(12): 3626 – 3630, 3638. (in Chinese)
- [27] WOO S H, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.
- [28] CHEN Z, WANG J, JIA C, et al. Pathological image super-resolution using mix-attention generative adversarial network [J]. International Journal of Machine Learning and Cybernetics, 2024, 15(1): 149 – 159.
- [29] LV M, ZHOU G, HE M, et al. Maize leaf disease identification based on feature enhancement and DMS-robust AlexNet [J]. IEEE Access, 2020, 8: 57952 – 57966.
- [30] SYARIEF M, SETIAWAN W. Convolutional neural network for maize leaf disease image classification [J]. Telkomnika (Telecommunication Computing Electronics and Control), 2020, 18(3): 1376 – 1381.
- [31] KRISHNAMOORTHY M, SANKAVI R, AISHWARYA V, et al. Maize leaf diseases identification using data augmentation and convolutional neural network [C] // 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2021: 1672 – 1677.
- [32] STEPHEN A, PUNITHA A, CHANDRASEKAR A. Designing self attention-based ResNet architecture for rice leaf disease classification [J]. Neural Computing and Applications, 2023, 35(9): 6737 – 6751.