

李西兴,陈佳豪,吴锐,等.基于改进MaxViT的辣椒病害识别分类方法[J].华中农业大学学报,2024,43(2):123-133.
DOI:10.13300/j.cnki.hnlkxb.2024.02.015

基于改进MaxViT的辣椒病害识别分类方法

李西兴,陈佳豪,吴锐,杨睿

湖北工业大学机械工程学院/湖北工业大学现代制造质量工程湖北省重点实验室,武汉 430068

摘要 为实现复杂环境下辣椒病害的精准识别和分类,设计了一种适用于辣椒病害识别分类的方法。以辣椒在生长过程中常见的6种病害为分类研究的对象,使用数据增强的方法扩充数据集,提出一种基于MaxViT改进的MaxViT-DF模型,将MaxViT模型中的普通卷积替换为可变形卷积,使模型在提取特征时能更贴近复杂环境下的识别目标;同时在MaxViT模型施加注意力时引入特征融合模块,提高模型的全局感知能力。结果显示,改进的MaxViT-DF模型识别分类准确率达到98.10%,对6种辣椒病害的分类精度均高于95%。与ResNet-34、EfficientNetv2和VGG-16等模型相比,改进模型在收敛速度和分类精度上具有明显优势。以上结果表明,MaxViT-DF模型能够对不同种类的辣椒常见病害进行有效的分类识别。

关键词 MaxViT-DF; 辣椒病害分类; 可变形卷积; 特征融合; 深度学习

中图分类号 TP391.41; S436.418.1 **文献标识码** A **文章编号** 1000-2421(2024)02-0123-11

21世纪以来,全球辣椒种植面积与产量稳步增长,中国辣椒产业也进入了迅速发展阶段。据联合国粮食及农业组织的统计数据显示,2020年中国辣椒产量约占全球产量的1/2,是全球最大的辣椒主产国^[1]。在辣椒的实际种植生产过程中,各种病害阻碍了辣椒的正常生长,从而造成辣椒减产。因此,对辣椒常见病害的精准识别分类可以起到及时发现病害、指导农药使用和减少经济损失等作用。传统的辣椒病害识别主要是基于高光谱图像和基于表面纹理特征的方法,上述方法依赖人工并且不利于后续农业智能化的发展,因此亟需一种能精准识别辣椒病害的方法。

近些年来深度学习得到了快速发展,深度学习中的卷积神经网络^[2](convolutional neural network, CNN)能充分利用计算机的计算能力,高效地提取图像特征,在植物病害识别分类方面取得了良好的效果^[3]。Sladojevic等^[4]首次利用CNN将植物叶片与周围环境区分开来,并且在CaffeNet模型上微调改进,改进模型的平均识别精度达到了96.3%。张帅等^[5]针对人工分析提取特征时造成的不同植物种属差异性问题的,利用一个8层卷积神经网络深度学习系

统分别对简单背景和复杂背景叶片图像进行训练和识别,结果表明,使用CNN+SVM和CNN+Softmax分类器识别方法对单一背景叶片图像识别率分别高达91.11%和90.90%,复杂背景叶片图像的识别率也能达到34.38%。张善文等^[6]针对在复杂的病害叶片图像中很难选择出对病害类型识别贡献较大的有用特征这一问题,构建了一种深度CNN模型,直接从归一化后的彩色病害叶片图像中提取高层次的抽象特征,并在输出层进行病害识别,结果表明,与基于特征提取的传统病害识别方法相比,该方法的识别性能较高,识别分类准确率达到90.32%。虽然基于CNN识别植物病害具有识别准确率高的优势,但是受制于网络架构因素、模型参数量大、训练时间长等特点,模型不易使用。

针对上述CNN识别植物病害时模型训练收敛时间长和参数庞大的问题。Vishnoi等^[7]开发了一种CNN,该网络由较少数量的层组成以降低计算负担,所提出的模型仅需要较少的存储和计算资源便可以达到98%的识别分类准确率。Liu等^[8]使用MobileNetV2模型作为骨干网络实现了对柑橘病害的识别,该方法在保持良好分类精度的同时减少了模型

收稿日期:2023-06-05

基金项目:国家自然科学基金项目(51805152);湖北工业大学绿色工业引领计划项目(XJ2021005001);湖北省自然科学基金项目(2022CFB445);湖北省重点研发计划项目(2021BAA203)

李西兴, E-mail: li_xi_xing@126.com

通信作者:吴锐, E-mail: wurui@hbut.edu.cn

的预测时间和模型大小。为了使分类模型更容易部署到手机端,张鹏程等^[9]对 MobileNetV2 模型进行了进一步的改进,将 ECA 注意力机制嵌入 MobileNetV2 网络的反残差结构尾部,以增强原网络的跨通道信息交互能力,提升原网络的特征提取能力,并基于该改进模型开发了一款边缘计算 APP,改进模型对柑橘虫害的分类准确率达到 93.63%。为了提高检测速度的同时提高识别分类准确率,苏俊楷等^[10]在 YOLOv5 模型的主干网络中添加 CA 注意力机制,改善目标漏检问题;在颈部使用 BiFPN 替代原有的 PANet;并且引入 Focal-EIOU Loss 损失函数,在保持模型较低计算量的情况下,提升了模型的检测速度和算法性能,改进后的 YOLOv5 模型和传统的 YOLOv5s 相比,平均精度上升了 4.5%。为了提高模型识别病害的泛化性和精准度,孙道宗等^[11]在 EfficientNetv2 网络模型引入了迁移学习机制,并且选取 MultiMarginLoss 作为模型的损失函数,结果显示,改进模型在参数量和模型大小均下降的情况下,在存储空间大小、准确率和训练时间 3 个指标上具有明显优势。

值得注意的是,以上的植物病害识别结果都是建立在具有良好的环境和数据集的背景下得出的,其数据集背景简单并且检测目标较易被识别,而在日常农业生产中很难有如此理想的环境。因此,为了应对背景杂乱或有干扰物等的复杂环境下对辣椒病害进行精准识别分类的问题,本研究选用 MaxViT^[12]作为骨干网络,建立了 MaxViT-DF 模型,旨在保证病害特征提取能力的同时使网络兼顾注意力

机制以提高分类准确率。

1 材料与方法

1.1 数据集

本研究以炭疽病、病毒病等对辣椒生长危害最严重的 6 种病害作为研究对象,收集图像组成基础数据集,图像背景为辣椒自然生长环境,图片背景复杂并且有干扰物,如图 1 所示。

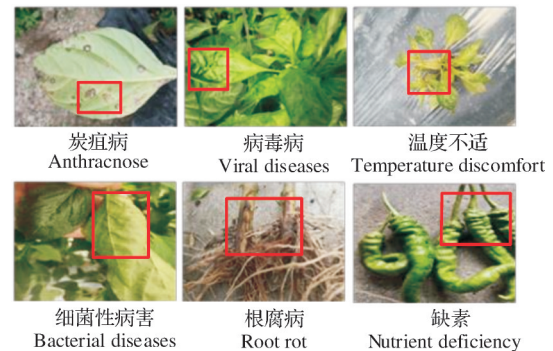


图 1 辣椒常见 6 种病害

Fig. 1 Common diseases of pepper

基础数据集共收集病害 2 107 张图片(随机划分测试集 421 张,训练集 1 686 张)。为了避免模型没有足够样例区分特征,导致模型在训练过程中出现过拟合现象,通过图像翻转、调整对比度、调整饱和度和调整锐度 4 类数据增强方法对训练集进行处理(图 2),获得本识别分类方法的最终训练集(表 1)。为了保证试验数据的均衡性,将 6 种病害的训练集数量扩大至同等规模,每种病害数据集包括 560 张图片,共 3 360 张图片。

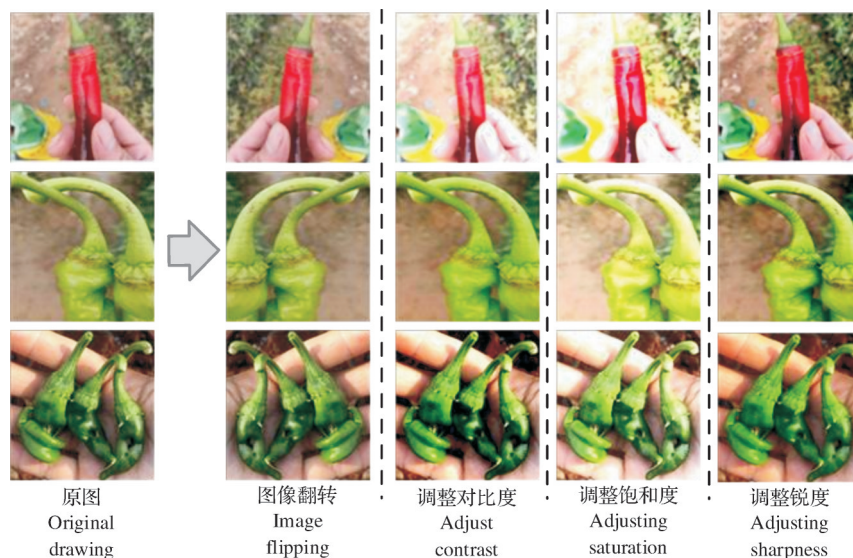


图 2 数据增强结果

Fig. 2 Data enhancement results

表 1 各病害训练集来源

Table 1 Source of each disease training set

种类 Type	训练集 Training set	图像翻转 Image flipping	调整对比度 Adjust contrast	调整锐度 Adjusting sharpness	调整饱和度 Adjusting saturation	合计 Total
d1	271	271	18	0	0	560
d2	560	0	0	0	0	560
d3	276	276	8	0	0	560
d4	187	187	186	0	0	560
d5	280	280	0	0	0	560
d6	112	112	112	112	112	560

注:d1~d6 分别表示炭疽病、病毒病、温度不适、细菌性病害、根腐病和缺素。Note:d1-d6 represent anthracnose, viral diseases, temperature discomfort, bacterial diseases, root rot, and nutrient deficiency respectively.

1.2 MaxViT 模型

MaxViT 是 2022 年谷歌提出的分层 Transformer 模型^[13],融合了注意力机制和卷积,并且提出了 1 种新颖的独立多轴注意力模块,由阻塞的局部注意力和扩张的全局注意力组成,具有线性复杂性的感知能力,MaxViT 在图像分类和目标检测方面表现良好。

MaxViT 模型如图 3 所示,输入网络的是 1 张 224 像素 × 224 像素的三通道 RGB 图片,输出为 6,

对应辣椒的 6 种病害。输入的图片首先通过 2 个卷积层进行特征提取,然后经过 4 个 MaxViT block 添加注意力,在 MaxViT block 中将 MBCov 与注意力一起使用进一步提高了网络的泛化性和可训练性,再经过池化层和全连接层输出结果,最后将网络处理后的结果输入 Softmax 分类器计算病害类别预测概率,将其中概率最大的病害作为网络识别这张图片的预测种类。

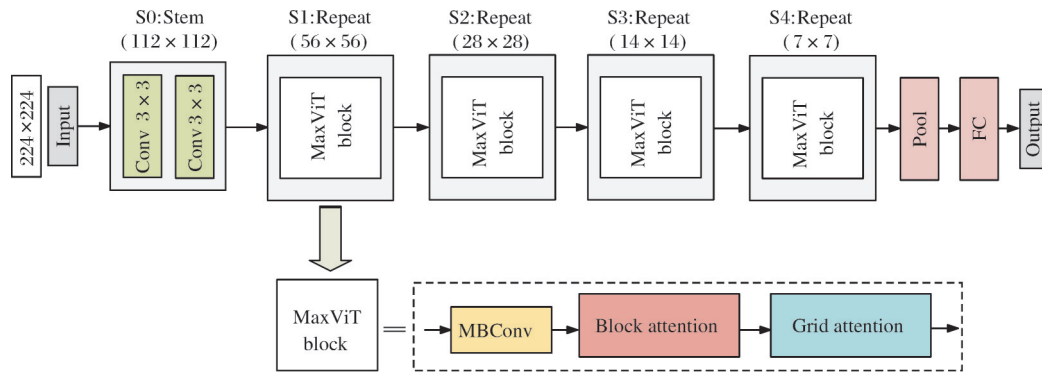


图 3 MaxViT 模型

Fig. 3 MaxViT model

1.3 可变形卷积模块

针对辣椒表面形状复杂和病害缺陷范围不规则的问题,在 MaxViT 模型的基础上将提取特征的普通卷积层替换为可以偏移位置的可变形卷积^[14]。可变形卷积在采样位置增加了 1 个偏移量,相比于普通卷积,可变形卷积在采样时可以更贴近物体的形状和尺寸,具有更强的鲁棒性(图 4)。

可变形卷积结构如图 5 所示,由普通卷积和偏移模块组成。偏移模块中的偏差通过 1 个卷积层获得,输入特征图,输出偏差,生成通道维度是“2N”,其中“2”分别对应平面上 x 值方向和 y 值方向这 2 个 2D 偏移,“N”是通道数,一共有 2 种卷积核,包括普通卷积中的卷积核以及卷积核学习 offset 对应的卷积层内

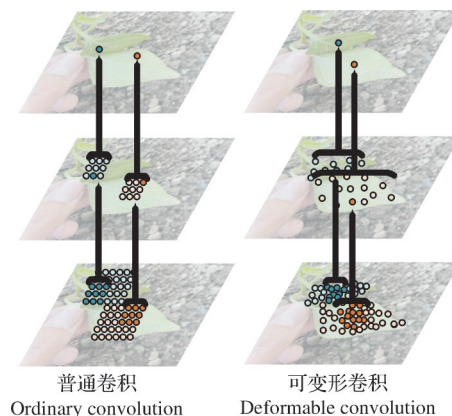


图 4 普通卷积和可变形卷积对比

Fig. 4 Comparison between ordinary convolutions and deformable convolutions

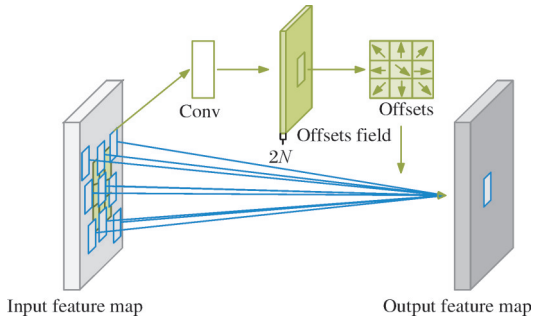


图5 可变形卷积结构

Fig. 5 Structure of deformable convolution

的卷积核,二者可通过双线性插值反向传播同时进行参数更新。

1.4 特征融合模块

MaxViT 模块在添加注意力的同时也对特征图进行尺寸的缩减,为了缓解这种因为尺寸缩减而带来的特征损失问题,本研究在MaxViT模型的框架中引入了特征融合模块。特征融合模块结构如图6所示,首先将输入的特征图分为2部分,分别进入MaxViT模块施加注意力和CBR模块提取特征,CBR模块由卷积、批量归一化和ReLU激活函数组成;其次将MaxViT模块的输出作为下一个MaxViT模块的输入继续添加注意力,并且把MaxViT模块的输出与

CBR模块的输出相加,将相加得到的结果作为下一个CBR模块的输入继续进行下采样过程;最后将MaxViT模块和CBR模块的最终输出结果相加,作为整个特征融合模块的输出结果。

1.5 MaxViT-DF模型

为了提高辣椒病害的识别准确率,将可变形卷积和特征融合模块融合到原始MaxViT模型中。MaxViT-DF模型结构如图7所示,用可变形卷积替代原始MaxViT中的2个卷积层进行特征提取,和原始MaxViT中的2个卷积层一样,可变形卷积将输入大小为224像素×224像素的图片提取为112像素×112像素的特征图,并且在提取特征的过程中更加贴近识别目标的位置与形状,有利于模型在后续对提取到的特征进行准确有效的学习;将经过特征提取后的特征图放入MaxViT block中施加注意力,同时也把特征图放入特征融合模块中融合全局信息,在每个特征融合的过程中,利用卷积层对数据进行下采样,以保证采样后的数据大小与MaxViT block处理后的数据大小相匹配;将4个MaxViT block处理后的结果和特征融合模块处理结果相加,经过池化层和全连接层处理输出分类结果。

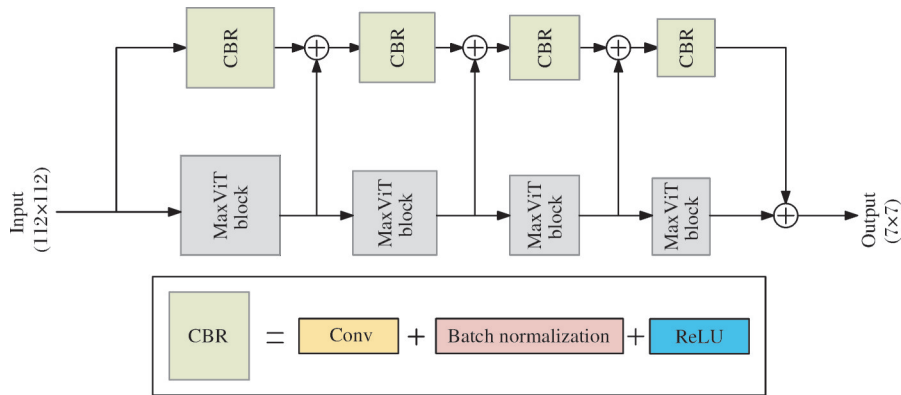


图6 特征融合模块结构

Fig. 6 Structure of feature fusion module

MaxViT block由MBConv和多轴注意力模块组成。在MBConv中输入特征在经过卷积升维后进行深度可分离卷积^[15],SE模块^[16]用于增强模型对重要通道的关注能力,之后再次使用逐点卷积恢复维度;由于直接沿整个空间施加注意力在计算上是不可行的,因此使用多轴注意力模块来施加注意力,多轴注意力模块包括窗格注意力和网格注意力,从而形成了局部和稀疏2种注意力形式,结构如图8所示。

窗口注意力对于输入特征图 $X \in R^{H \times W \times C}$,转化为形状张量 $\left(\frac{H}{P} \times P, \frac{W}{P} \times P, C\right)$,以表示划分为不

重叠的窗口,其中每个窗口的大小为 $P \times P$,最后在每一个窗口中执行自注意力计算,如式(1)所示:

$$(H, W, C) \rightarrow \left(\frac{H}{P} \times P, \frac{W}{P} \times P, C\right) \rightarrow \left(\frac{HW}{P^2}, P^2, C\right) \quad (1)$$

不同于传统使用固定窗口大小来划分特征图的操作,网格注意力使用固定的 $G \times G$,均匀网格将输入张量网格化为 $\left(G \times G, \frac{H}{G} \times \frac{H}{G}, C\right)$,得到自适应

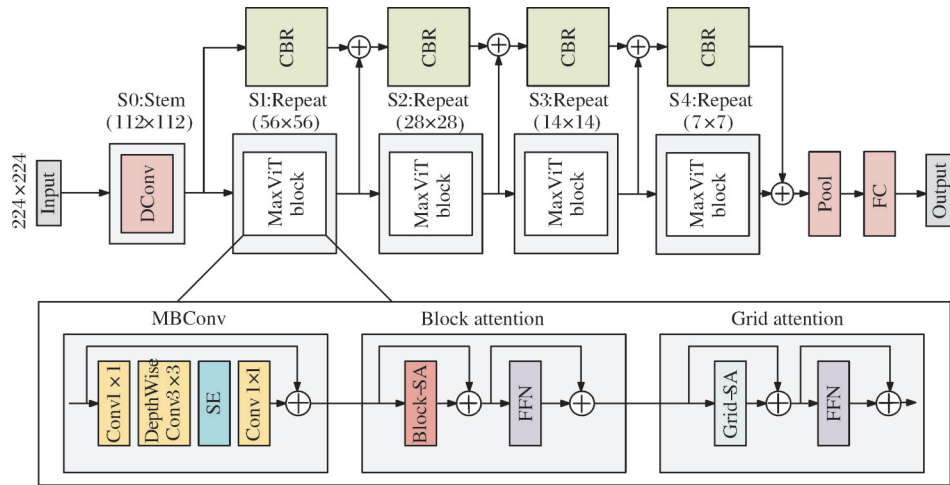


图 7 MaxViT-DF 模型结构

Fig. 7 MaxViT-DF model structure

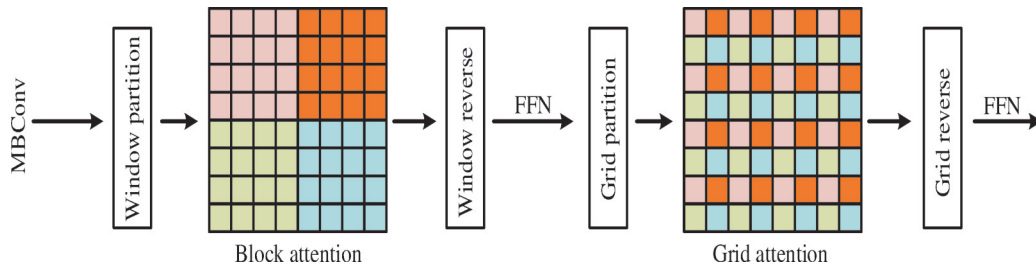


图 8 多轴注意力模块结构

Fig. 8 Structure diagram of multi-axis attention module

窗口 $\frac{H}{G} \times \frac{H}{G}$, 最后在 $G \times G$ 上使用自注意力计算, 通过使用相同的窗口 $P \times P$ 和网格 $G \times G$, 可以有效平衡局部和全局之间的计算, 且仅具有线性复杂度, 如式(2)所示:

$$\begin{aligned} (H, W, C) &\rightarrow \left(G \times \frac{H}{G}, G \times \frac{W}{G}, C\right) \\ &\rightarrow \left(G^2, \frac{HW}{G^2}, C\right) \rightarrow \left(\frac{HW}{G^2}, G^2, C\right) \end{aligned} \quad (2)$$

在窗格注意力和网格注意力之间通过前馈网络^[17-18]进行连接。

1.6 试验配置及训练参数

试验所用中央处理器型号为 AMD EPYC 7551P 256G 64 核、内存大小 64 GB, 图形处理器 GPU 为 NVIDIA GeForce RTX 3090。操作系统为 ubuntu20.04, 深度学习框架使用 Pytorch1.8.1、Python3.7、CUDA 11.3。

为了验证改进方法的有效性, 将 MaxViT-DF 模型与其他 5 个主流分类网络模型进行对比测试, 包括 ResNet-34^[19]、EfficientNetv2^[20]、VGG-16^[21]、MaxViT 和 MaxViT (DConv)。各分类模型在训练时除了网络结构不同外, 训练数据集、预处理等均保持一

致。所有分类模型均采用自适应梯度下降法 AdamW 优化器进行优化。网络训练初始学习率为 0.000 1, 设定最大训练步数为 200。

辣椒病害识别分类任务使用 BCEWithLogitsLoss 损失函数。先将预测种类经过 Sigmoid 函数归一化, 如式(3), 然后用真实标签和预测种类的概率计算损失, BCEWithLogitsLoss 损失函数计算公式如式(4)。

$$S(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

$$L_c = -\frac{1}{n} \sum_i (t[i] \times \log(S(o[i])) + (1 - t[i]) \times \log(1 - S(o[i]))) \quad (4)$$

式(3)~(4)中 $S(i)$ 为 Sigmoid 函数; L_c 为 BCEWithLogitsLoss 损失函数; $t[i]$ 为第 i 类的真实标签, $o[i]$ 为第 i 类预测的概率。

1.7 评价指标

试验在分类测试图像上分别测试模型在各类别上的准确率 (accuracy, A)、召回率 (recall, R)、精度 (precision, P) 和 F_1 值, 各评价指标的计算公式分别为:

$$A = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \quad (5)$$

$$P = \frac{T_p}{T_p + F_p} \quad (6)$$

$$R = \frac{T_p}{T_p + F_n} \quad (7)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (8)$$

其中, T_p 表示预测为正例实际上也是正例; F_p 表示预测为正例实际上却是负例; F_n 表示预测为负例实际上却是正例; T_n 表示预测为负例实际上也是负例。

2 结果与分析

2.1 损失函数分析

在训练过程中各模型的损失曲线、准确率曲线

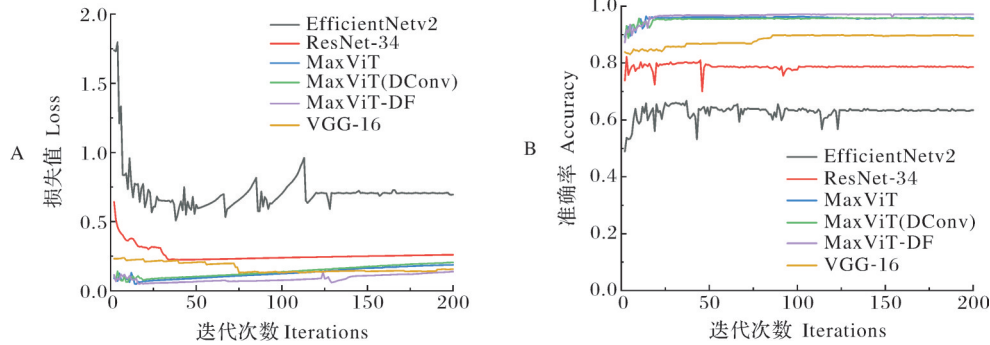


图9 各分类模型损失曲线(A)和准确率曲线(B)

Fig. 9 Loss curves (A) and accuracy curve (B) in each classification model

2.2 模型分类性能对比

不同辣椒病害分类模型的结果、平均单幅图像的推理时间和模型大小如表2所示。由表2可以看出,将普通卷积替换为可变形卷积后,模型对辣椒病害的分类准确率为95.00%,在此基础上再加入特征

融合模块,模型对辣椒病害的分类准确率为98.10%,优于原始的MaxViT模型(94.50%)。结果表明,可变形卷积可以更加精准地定位到辣椒病害的缺陷部位,并且特征融合模块可以提取输入特征的全局信息,从而提高模型的分类准确率。

表2 不同模型评估标准对比

Table 2 Comparison of evaluation standards for different models

模型 Model	准确率/% Accuracy	精度/% Precision	召回率/% Recall	F_1 值/% F_1 value	时间/ms Time	模型大小/MB Model size
ResNet-34	82.19	82.91	80.7	81.79	78.38	81.34
EfficientNetv2	66.75	62.67	64.72	63.68	79.35	77.86
VGG-16	90.97	91.46	92.35	91.90	118.76	512.27
MaxViT	94.50	94.57	94.82	94.69	102.14	109.32
MaxViT(DConv)	95.00	95.02	94.35	94.68	102.03	108.68
MaxViT-DF	98.10	98.48	98.06	98.27	106.89	109.66

为了进一步分析各分类模型对辣椒各病害的分类差异,列出各分类模型在辣椒各病害的分类召回率、精度和 F_1 值(表3)。由表3可知,相比原始MaxViT模型,MaxViT-DF模型在病毒病、温度不适、细

菌性病害和缺素上的分类精度分别提高了4.96、5.83、2.04、6.90百分点。从辣椒病害识别类别来看,MaxViT-DF模型对炭疽病、病毒病、温度不适、细菌性病害、根腐病和缺素这6种辣椒病害的分类精度均

表 3 不同模型在各病害中的评估标准对比

Table 3 Comparison of evaluation criteria for different models in different diseases

%

病害 Disease	模型 Model	召回率 Recall	精度 Precision	F_1 值 F_1 value
炭疽病 Anthracnose	ResNet-34	82.09	84.62	83.34
	EfficientNetv2	50.75	56.67	53.55
	VGG-16	97.01	94.20	95.58
	MaxViT	91.04	100.00	95.31
	MaxViT(DConv)	92.54	96.88	94.66
	MaxViT-DF	95.92	100.00	97.71
病毒病 Viral diseases	ResNet-34	87.86	76.88	82.00
	EfficientNetv2	70.00	73.68	71.79
	VGG-16	82.86	94.31	88.22
	MaxViT	92.86	92.20	92.53
	MaxViT(DConv)	93.57	92.25	92.91
	MaxViT-DF	97.86	97.16	97.51
温度不适 Temperature discomfort	ResNet-34	51.43	81.82	63.16
	EfficientNetv2	50.00	56.45	53.03
	VGG-16	91.43	76.19	83.12
	MaxViT	86.30	90.00	88.11
	MaxViT(DConv)	87.14	87.14	87.14
	MaxViT-DF	98.57	95.83	97.18
细菌性病害 Bacterial diseases	ResNet-34	97.83	81.82	89.11
	EfficientNetv2	78.26	63.16	69.90
	VGG-16	93.48	95.56	94.51
	MaxViT	100.00	95.83	97.87
	MaxViT(DConv)	100.00	93.88	96.84
	MaxViT-DF	100.00	97.87	98.92
根腐病 Root rot	ResNet-34	97.14	93.15	95.10
	EfficientNetv2	92.86	81.25	86.67
	VGG-16	100.00	95.89	97.90
	MaxViT	98.57	100.00	99.28
	MaxViT(DConv)	100.00	100.00	100.00
	MaxViT-DF	100.00	100.00	100.00
缺素 Nutrient deficiency	ResNet-34	67.86	79.17	73.08
	EfficientNetv2	46.43	44.83	45.62
	VGG-16	89.29	92.59	90.91
	MaxViT	96.43	93.10	94.74
	MaxViT(DConv)	100.00	92.86	96.30
	MaxViT-DF	96.43	100.00	98.18

高于 95%, 特别是对部分辣椒病害如炭疽病、根腐病和缺素的分类精度达到了 100%。

2.3 分类结果展示

由图 10 可知, 模型在预测缺陷集中在叶片上或者根部时准确率较高, 比如炭疽病和根腐病, 而在预测病毒病和温度不适这类缺陷分布在整株和大量叶片上的病害时错分情况较多。这说明模型在处理缺陷目标明显的图片时可以准确的定位到缺陷部

位并提取到相应的特征信息, 而在处理小目标或者分散目标的图片则会出现错分的情况。

为进一步了解各辣椒病害的识别情况, 根据各分类模型各病害分类准确率作混淆矩阵(其中 d1~d6 分别表示炭疽病、病毒病、温度不适、细菌性病害、根腐病和缺素)。由图 11 可知, 各分类模型普遍对病毒病和温度不适这 2 种病害的分类准确率较低, EfficientNetv2 模型中病毒病和温度不适 2 种病害

的分类准确率均低于75%，原始MaxViT模型中这2种病害分类准确率分别为92%和86%，而本研究提出的MaxViT-DF模型对上述2种病害的错分情况

有一定的缓解，分类准确率分别达到了97%和96%。表明本研究提出的方法提高了模型在处理病毒病和温度不适这类背景复杂病害时的预测能力。

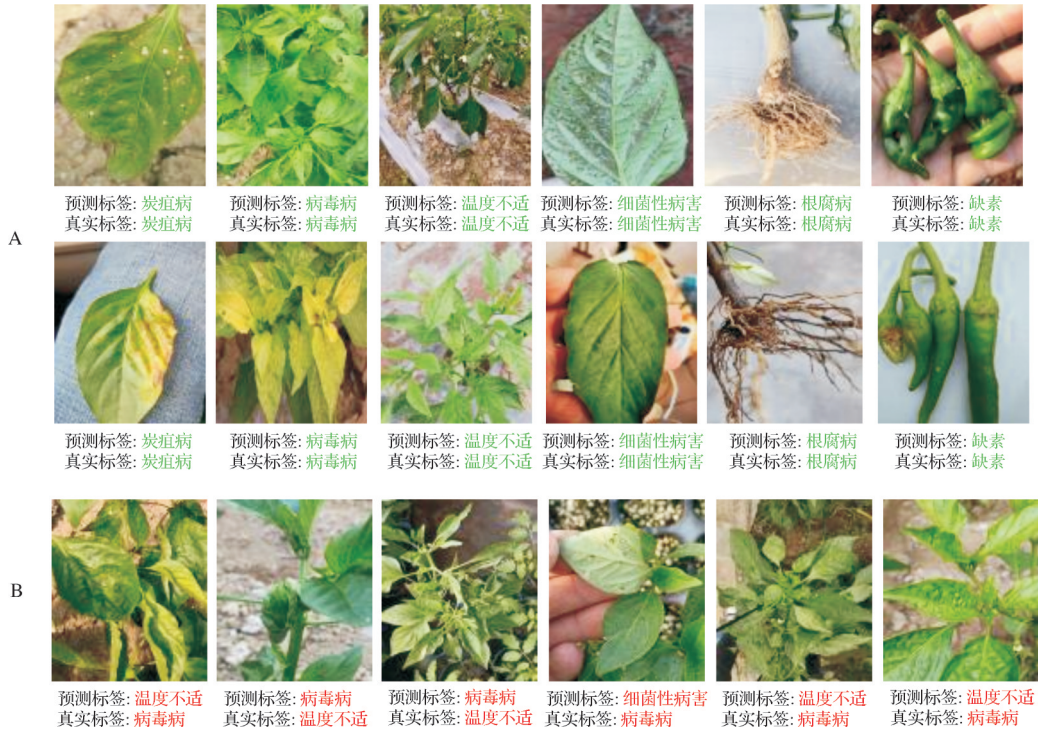


图 10 基于MaxViT-DF模型的辣椒病害分类结果

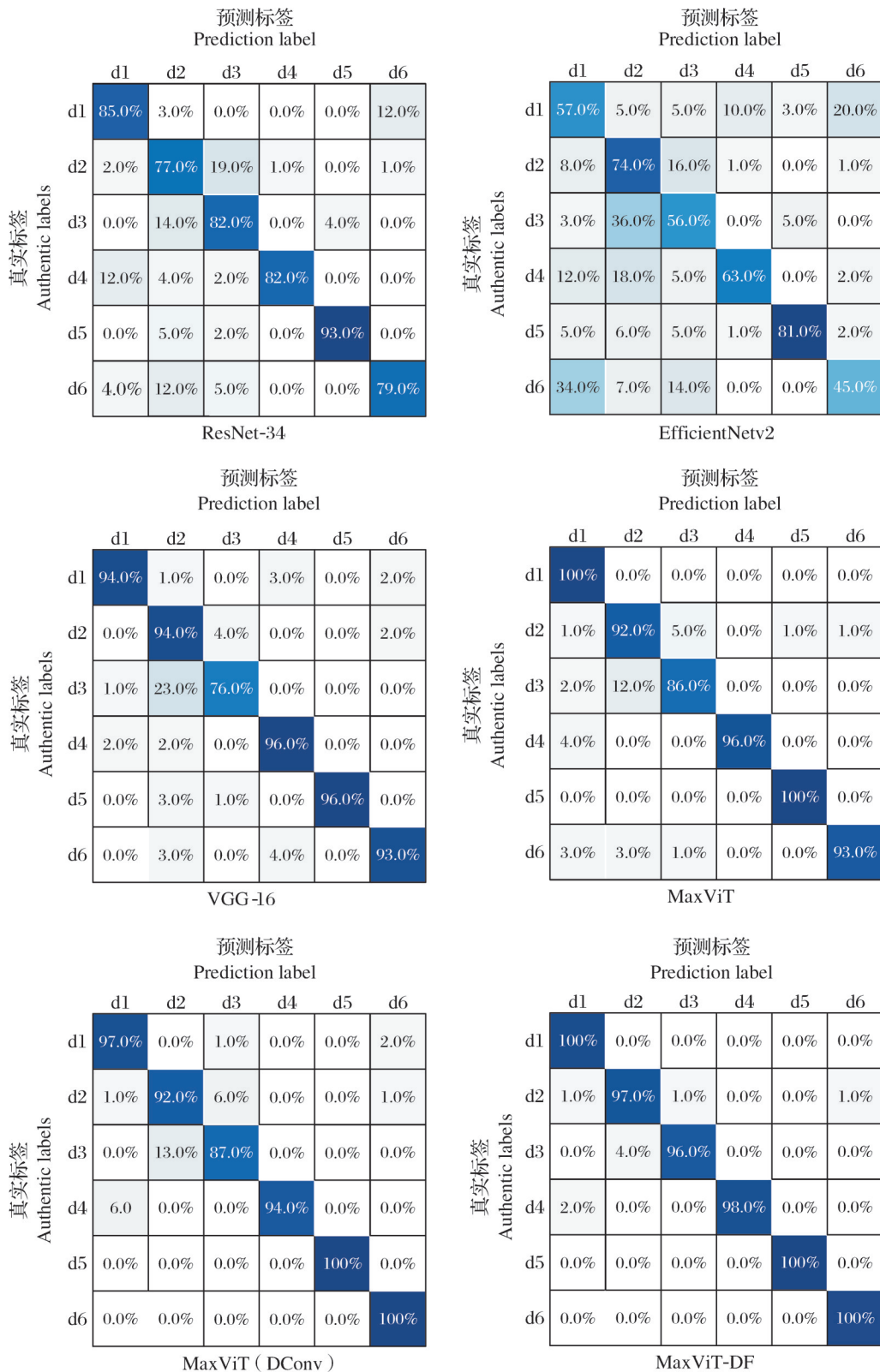
Fig. 10 Classification results of pepper diseases based on MaxViT-DF model

3 讨论

本研究提出了一种在复杂环境下(如背景杂乱或有干扰物等)对辣椒病害进行识别分类的方法,提出的MaxViT-DF模型相比于原始的MaxViT模型改进点有2个:(1)改进后的MaxViT-DF模型相比原始MaxViT模型,将用于提取特征的普通卷积替换为可变形卷积,使模型在提取特征时可以更贴近物体的形状和尺寸,提高了模型的分类准确率;(2)改进后的MaxViT-DF模型在原始MaxViT模型施加注意力机制时引入了特征融合模块,增加了模型在处理输入特征时的全局信息感知能力,使模型在应对小目标缺陷时保持较高的分类性能。通过与其他分类模型的对比,本研究所提出的辣椒病害识别分类方法的识别准确率达到98.10%,高于MaxViT(94.50%)、ResNet-34(82.19%)、EfficientNetv2(66.75%)和VGG-16(90.97%),分别提高3.6、15.91、31.35和7.13个百分点,表明改进MaxViT-DF

模型具有高准确率和较高的泛化能力。

同时,改进MaxViT-DF模型对各病害的识别结果分析发现,改进模型对炭疽病、细菌性病害、根腐病和缺素4种病害进行目标识别时,能够减少复杂环境对其识别性能的干扰,容易提取到有效特征,但改进模型对辣椒病毒病和温度不适病害的识别分类准确率相对较低,一方面是因为病毒病和温度不适病害多发生在辣椒叶片上,特征不如其他病害明显,而模型在应对病毒病和温度不适病害时容易被环境干扰从而学习到无效特征;另一方面病毒病和温度不适病害图片中可识别的部位过多,导致模型不能识别到真正有病害缺陷的叶片。针对上述错分问题,在今后的研究中应该对图片的复杂背景进行相应的前处理,将背景与病害部位分开提取,并且进一步加深模型的多尺度特征融合能力,增强模型在小目标和相似目标上的识别能力,从而提高模型在识别多种病害时的整体准确率。



d1~d6 分别表示炭疽病、病毒病、温度不适、细菌性病害、根腐病和缺素。d1-d6 represent anthracnose, viral diseases, temperature discomfort, bacterial diseases, root rot, and nutrient deficiency respectively.

图 11 各模型分类结果的混淆矩阵

Fig. 11 Confusion matrix of classification results of each model

参考文献References

- [1] Food and Agriculture Organization of the United Nations. FAOSTAT [DB/OL]. [2023-06-05]. <https://www.fao.org/faostat/zh/#data/QCL>.
- [2] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]//FLEET D, PAJDLA T, SCHIELE B, et al. Computer Vision - ECCV: lecture notes in computer science. Cham: Springer, 2014: 818-833.
- [3] 王燕, 吕艳萍. 混合深度CNN联合注意力的高光谱图像分类[J]. 计算机科学与探索, 2023, 17(2): 385-395. WANG Y, LÜ Y P. Hybrid deep CNN-attention for hyperspectral image classification [J]. Journal of frontiers of computer science and technology, 2023, 17(2): 385-395 (in Chinese with English abstract).
- [4] SLADOJEVIC S, ARSENOVIC M, ANDERLA A, et al. Deep neural networks based recognition of plant diseases by leaf image classification [J/OL]. Computational intelligence and neuroscience, 2016, 2016: 3289801 [2023-06-05]. <https://doi.org/10.1155/2016/3289801>.
- [5] 张帅, 淮永建. 基于分层卷积深度学习系统的植物叶片识别研究[J]. 北京林业大学学报, 2016, 38(9): 108-115. ZHANG S, HUAI Y J. Leaf image recognition based on layered convolutions neural network deep learning [J]. Journal of Beijing Forestry University, 2016, 38(9): 108-115 (in Chinese with English abstract).
- [6] 张善文, 谢泽奇, 张晴晴. 卷积神经网络在黄瓜叶部病害识别中的应用[J]. 江苏农业学报, 2018, 34(1): 56-61. ZHANG S W, XIE Z Q, ZHANG Q Q. Application research on convolutional neural network for cucumber leaf disease recognition [J]. Jiangsu journal of agricultural sciences, 2018, 34(1): 56-61 (in Chinese with English abstract).
- [7] VISHNOI V K, KUMAR K, KUMAR B, et al. Detection of apple plant diseases using leaf images through convolutional neural network [J]. IEEE access, 2022, 11: 6594-6609.
- [8] LIU Z, XIANG X, QIN J, et al. Image recognition of citrus diseases based on deep learning [J]. Computers, materials and continua, 2021, 66(1): 457-466.
- [9] 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 P C, YU Y H, CHEN C W, 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 with English abstract).
- [10] 苏俊楷, 段先华, 叶赵兵. 改进YOLOv5算法的玉米病害检测研究[J]. 计算机科学与探索, 2023, 17(4): 933-941. SU J K, DUAN X H, YE Z B. Research on corn disease detection based on improved YOLOv5 algorithm [J]. Journal of frontiers of computer science and technology, 2023, 17(4): 933-941 (in Chinese with English abstract).
- [11] 孙道宗, 刘锦源, 丁郑, 等. 基于改进 EfficientNetv2 模型的多品种南药叶片分类方法[J]. 华中农业大学学报, 2023, 42(1): 258-267. SUN D Z, LIU J Y, DING Z, et al. Classification of leaves of multi-variety southern traditional Chinese medicine based on improved EfficientNetv2 model [J]. Journal of Huazhong Agricultural University, 2023, 42(1): 258-267 (in Chinese with English abstract).
- [12] TU Z Z, TALEBI H, ZHANG H, et al. MaxViT: multi-axis vision transformer [DB/OL]. arXiv, 2022: 2204.01697 [2023-06-05]. <https://arxiv.org/abs/2204.01697>.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st international conference on neural information processing systems (NIPS 2017), December 4-9, 2017, Long Beach, USA. New York: Curran Associates Inc., 2017: 6000-6010.
- [14] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C]//Proceedings of the IEEE international conference on computer vision (ICCV), October 22-29, 2017, Venice, Italy. Venice: IEEE, 2017: 764-773.
- [15] JANG J G, QUAN C, LEE H D, et al. Falcon: lightweight and accurate convolution based on depthwise separable convolution [J]. Knowledge and information systems, 2023, 65(5): 2225-2249.
- [16] ZHANG J W, MA P L, JIANG T, et al. SEM-RCNN: a squeeze-and-excitation-based mask region convolutional neural network for multi-class environmental microorganism detection [J/OL]. Applied sciences, 2022, 12(19): 9902 [2023-06-05]. <https://doi.org/10.3390/app12199902>.
- [17] AGUIAR M A D, DIAS A, FIELD M. Feedforward networks: adaptation, feedback, and synchrony [J]. Journal of nonlinear science, 2019, 29(3): 1129-1164.
- [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale [DB/OL]. arXiv, 2020: 2010.11929 [2023-06-05]. <https://doi.org/10.48550/arXiv.2010.11929>.
- [19] GAO M Y, QI D W, MU H B, et al. A transfer residual neural network based on ResNet-34 for detection of wood knot defects [J/OL]. Forests, 2021, 12(2): 212 [2023-06-05]. <https://doi.org/10.3390/f12020212>.
- [20] DENG L W, SUO H F, LI D J. Deepfake video detection based on EfficientNet-V2 network [J/OL]. Computational intelligence and neuroscience, 2022, 2022: 3441549 [2023-06-05]. <https://doi.org/10.1155/2022/3441549>.
- [21] CAMPOS-LEAL J A, YEE-RENDÓN A, VEGA-LÓPEZ I F. Simplifying VGG-16 for plant species identification [J]. IEEE Latin America transactions, 2022, 20(11): 2330-2338.

A method for identifying and classifying pepper diseases based on improved MaxViT

LI Xixing, CHEN Jiahao, WU Rui, YANG Rui

College of Mechanical Engineering/Key Laboratory of Modern Manufacturing Quality Engineering, Hubei University of Technology, Wuhan 430068, China

Abstract A method suitable for identifying and classifying pepper diseases was designed to achieve precise recognition and classification of pepper diseases in complex environments including background clutter or interference. Six common diseases in the growth process of peppers were used to expand the dataset with data augmentation methods. An improved MaxViT-DF model was proposed based on MaxViT. The ordinary convolution in the MaxViT model was replaced with deformable convolution to enable the model to extract features closer to the recognition target in complex environments. A feature fusion module was introduced when applying attention to the MaxViT model to improve the model's global perception ability. The results showed that the improved MaxViT-DF model had an identification and classification accuracy of 98.10%, and the classification accuracy for six common pepper diseases was higher than 95%. The improved model had significant advantages in convergence speed and classification accuracy compared with models including ResNet-34, EfficientNetv2, and VGG-16. It is indicated that the MaxViT-DF model can effectively identify and classify common diseases in different types of peppers.

Keywords MaxViT-DF; classification of pepper disease; deformable convolution; feature fusion; deep learning

(责任编辑:葛晓霞)