

李亚,陈晓东,王海瑞,等.基于优化MobileViT模型的轻量化田间杂草识别[J].华中农业大学学报,2025,44(4):192-203.
DOI:10.13300/j.cnki.hnlkxb.2025.04.019

基于优化 MobileViT 模型的轻量化田间杂草识别

李亚¹,陈晓东¹,王海瑞¹,朱贵富^{2,3}

1.昆明理工大学信息工程与自动化学院,昆明 650504; 2.昆明理工大学信息化建设管理中心,昆明 650504;
3.昆明理工大学一曙光信息产业股份有限公司 AI 联合研究中心,昆明 650504

摘要 针对农业环境中杂草与作物幼苗的识别挑战,提出一种基于优化 MobileViT 模型的轻量化识别方法。首先引入 SimAM 注意力机制,增强模型对特征的注意力能力,使用 SCConv 卷积模块减少卷积神经网络中特征的空间和通道冗余来降低计算成本和模型存储,同时提高卷积模块性能;提出联合使用 Label Smoothing Loss 和 Cross Entropy Loss 的损失函数策略,提升模型的泛化性能,降低过拟合风险,并加速模型的收敛过程。选取 12 种田间常见作物幼苗与杂草图像作为训练数据集评估改进模型 MobileViT-SS 的性能,结果显示,改进模型的平均识别准确率、精确度、召回率、 F_1 分数分别达到 95.91%、95.97%、95.46%、95.69%,均优于当前广泛使用的深度神经网络模型 VGG-16、ResNet-18 和 MobileNetv3。结果表明,改进模型 MobileViT-SS 能够精准、快速区分多种形态相似的杂草与作物幼苗。

关键词 作物幼苗;杂草识别;智能农业;MobileViT;轻量化;SimAM 注意力

中图分类号 S451; TP391.41 **文献标识码** A **文章编号** 1000-2421(2025)04-0192-12

杂草的生长具有繁殖迅速和周期短的特点,它们会在早期与作物争夺关键的生长要素,如光照、水分和养分^[1]。若未能及时进行有效管理,可能会对农作物的产量和品质造成严重影响^[2],进而对农业生产构成重大威胁,特别是对于玉米和小麦等主要粮食作物^[3]。杂草的存在不仅与作物幼苗竞争必需的养分、光照和水分,还可能阻碍作物的正常生长^[4-5]。此外,杂草还为病虫害提供了滋生和扩散的环境,进一步危害作物的健康和生长^[6]。目前除草的方法主要包括人工除草和化学除草^[7]。人工除草需要耗费大量人力,效率低下;化学除草则是通过使用各种除草剂来清除杂草,虽然效果明显,但会导致化学药物残留,进而影响农产品的质量和土壤环境^[8]。

随着人工智能科技的快速进步,深度学习模型展现了其重要的作用。深度学习的方法能够有效区分农作物和杂草,帮助减轻因杂草引起的损失,同时减少农药使用的负面影响,其卓越的准确性和出色的泛化性能,成为当前备受关注的农田作物检测技术^[9-11]。国内外的研究者正积极探索深度学习技术

在杂草识别领域的应用^[12-14]。Olsen 等^[15]通过采用深度学习技术,对卷积神经网络(convolutional neural networks, CNN)在杂草检测和识别任务中的应用进行了深入研究。蔡竹轩等^[16]评估了 YOLOv、SSD 及 Faster R-CNN 等多种网络架构在田间水稻稻穗精准检测中的性能及应用效果。张乐等^[17]提出了一种利用 VGG-16 深度学习架构的油菜识别技术,利用该方法对油菜及杂草图像进行目标检测的准确率可达 83.9%。孙俊等^[18]基于改进的 MobileNet 模型识别复杂背景下的农作物叶部病害,引入了轻量型的坐标注意力机制,构建通道注意力机制与空间位置信息之间的关联,改进模型的识别准确率在背景复杂的农作物叶片病害数据集中达到了 92.20%。郭小燕等^[19]在 ResNet 模型中引入了双层注意力机制和通道特征提取机制,设计基于全局特征提取的深度神经网络,改善网络的全局特征提取能力,在农作物病害识别中平均准确率可达 88.98%。刘志等^[20]提出一种以 Cascade R-CNN 为基础框架的田间害虫检测模型,融合了跨通道注意力机制,并使用统一目标检测头,有效提升了模型对害虫特征的提取能力。

收稿日期:2024-12-24

基金项目:国家自然科学基金项目(61863016)

李亚, E-mail: 59515091@qq.com

通信作者: 朱贵富, E-mail: zhuguifu@kust.edu.cn

改进后的模型在田间害虫检测中的 F_1 分数可以达到86.2%。尽管卷积神经网络(CNN)功能强大,但在复杂的杂草识别场景中,CNN难以建模和提取全局级别特征。

而与CNN相比,Vision Transform 作为一种新兴的深度学习架构,通过使用自注意力机制,在图像块序列中高效地识别和提取重要的特征,能够捕捉到更全面的语义信息^[21]。因其在图像处理和分析方面的出色表现,正在农业领域获得越来越多的关注和应用^[22-23]。

当前的模型普遍具有大量的参数和复杂的网络结构,这意味着它们在训练和推理时需要处理的信息量较大,在移动端和嵌入式设备上的应用受到限制^[24-25]。MobileNet 系列网络在图像分类任务中表现优秀,其设计初衷是为了在资源受限的环境中(如移动设备)有效运作。这使得它既能保持较高的识别精度,又能有效利用较少的计算资源,成为移动设备上图像处理的理想选择^[26]。亢洁等^[27]基于轻量级神经网络 MobileNet,结合通道注意力机制和多尺

度特征融合,提出了一种精准且高效的杂草检测方法,平均精确率可达88.84%。

然而,现有的轻量化模型在卷积操作中主要侧重于局部特征的提取,难以同时捕捉到整体形态和细节的全局语义特征^[28]。为了克服这一局限,本研究提出1种基于优化 MobileViT 的田间杂草识别方法,结合轻量化设计和全局特征提取能力,用于数据规模较小且计算资源受限的实际应用场景,旨在获得高准确率与实时性的识别性能,为农业自动化领域提供一种有效的技术解决方案。

1 材料与方法

1.1 数据来源

本研究选取的数据来源于Kaggle平台的植物幼苗图像数据库(Plant Seedlings Dataset),该数据集包含12种植物类别(图1),共5 646幅图像(小麦258幅、玉米260幅、甜菜464幅、无香菊608幅、荠菜276幅、天竺葵580幅、野芥子454幅、猪殃殃348幅、繁缕716幅、黑草334幅、藜543幅、雀麦805幅)。

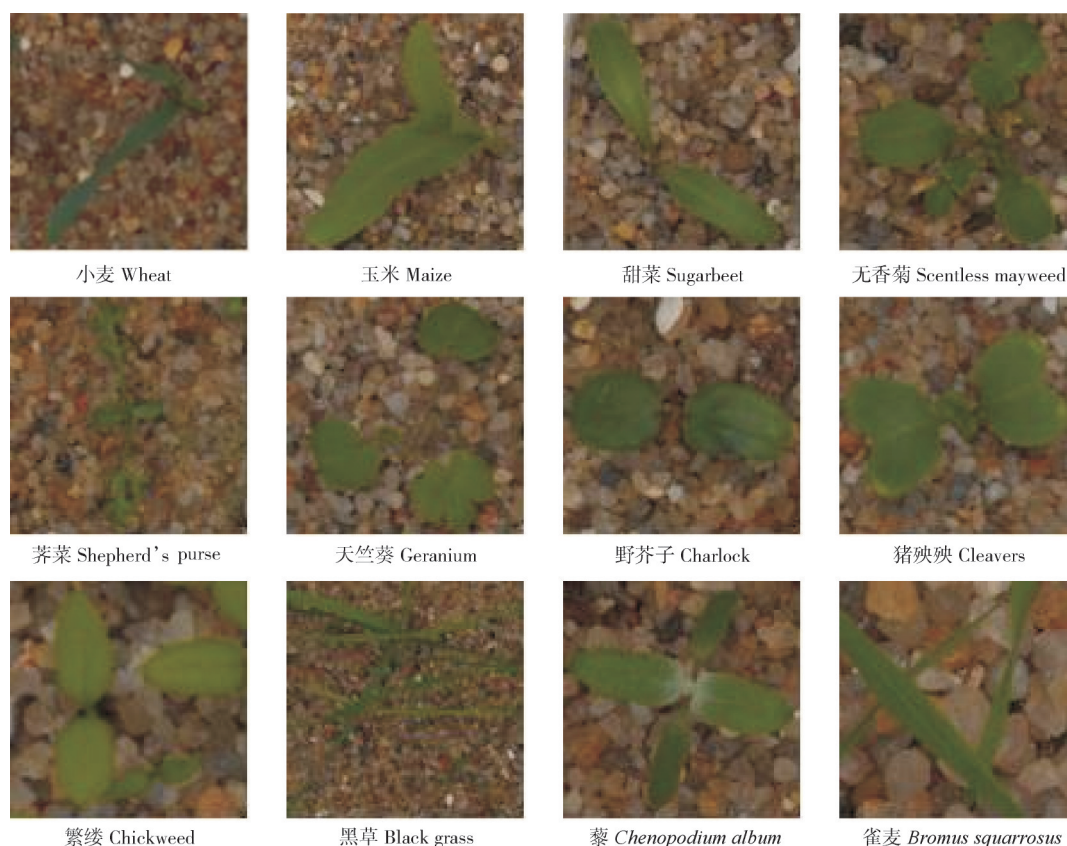


图1 数据集中部分作物幼苗与杂草示例

Fig. 1 Sample of crop seedlings and weeds in a dataset

1.2 数据处理

为了符合模型的输入要求,数据集中各种大小

的图像分辨率通过直接缩放调整为 $256 \times 256 \times 3$ 。从样本集合中随机挑选70%的样本用作训练数据,

用于模型参数的学习和超参数的调整;其余 30% 的样本用于模型性能的评估。

本研究的样本数据涵盖了作物幼苗与杂草,这些样本采集自不同的时间段、光照条件以及土壤环境,充分体现了图像的复杂性和多样性。数据集具有充足的数据量和高质量的图像,能够有效地支持学习过程,因此未采取数据增强技术以扩充数据集。

1.3 MobileViT 模块

MobileViT 是一种创新的轻量级神经网络架构,它融合了卷积神经网络(CNN)和视觉变换器(vision transformer, ViT)的优点,提高了图像特征的提取和理解能力。该混合模型尤其注重空间信息的归纳,以及对特征图信息的自适应加权,使其在保持轻量化的同时能够有效捕捉丰富的局部和全局特征。MobileViT 模型的主要结构由三部分组成:卷积层、MobileNetV2 倒残差模块以及核心的 MobileViT block。

在局部特征提取阶段,模型使用 3×3 的卷积核来捕捉局部特征,并通过 1×1 卷积核调整通道数,有助于提取局部细节并整合全局特征。连接初始层的是 MV2 层(图 2)。MV2 层作为 MobileNetV2 模型中的倒残差网络结构,通过先扩展后缩减特征图维度减少信息损失^[29]。捷径连接在输入输出维度相同时直接将输入添加到残差块输出上,不同步幅(stride, S)条件下采取不同策略。步幅为 1 时直接添加,步幅为 2 时通过下采样操作并增设卷积层匹配维度,确保捷径连接与主路径特征图尺寸匹配。

MobileViT block 是模型的核心(图 3)。它通过结合 CNN 的局部特征和 ViT 的全局特征建模,实现

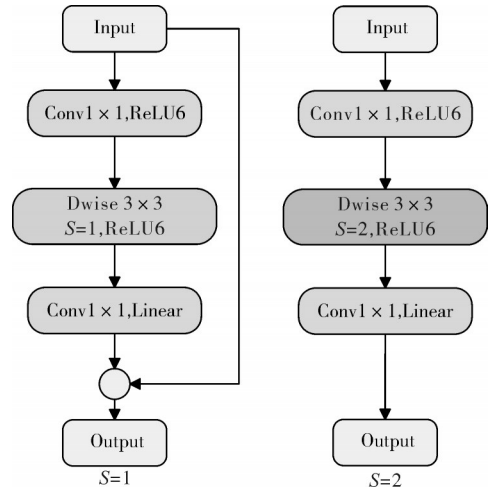


图 2 MV2 模块

Fig. 2 MV2 module

了对视觉信息的深入理解。该过程包括 Unfold、Transformer 和 Fold 操作,使模型能够对局部特征进行精细建模,并通过 Transformer 架构捕捉全局上下文信息。在特征融合阶段 MobileViT block 使用 1×1 卷积核恢复通道数,并利用跳跃连接将全局特征与局部特征结合,最终通过 3×3 卷积核完成特征的融合,生成最终的特征表示。这一融合过程不仅增强了模型对复杂视觉场景的理解能力,还提高了其在各种视觉识别任务中的泛化性能。

MobileViT 模型通过其创新的结构设计,有效地结合了 CNN 的空间归纳能力与 ViT 的全局感受能力,同时保持了模型的轻量化^[30]。这种设计不仅提升了模型对视觉信息的捕捉和理解能力,也使其能够在资源受限的设备上高效运行,具有广泛的应用前景。

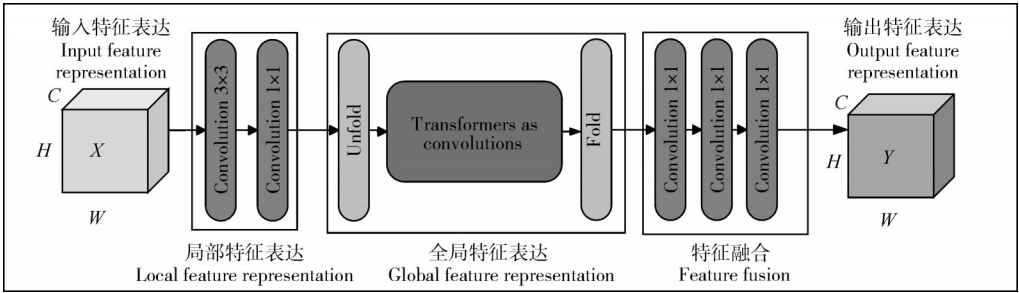


图 3 MobileViT 模块的结构

Fig. 3 Structure of the MobileViT module

1.4 优化的 MobileViT 模块

本研究对 MobileViT 模型结构进行了创新性改进,以进一步提升模型在图像分类任务中的表现。改进模型 MobileViT-SS 融合了 SimAM 注意力机制(simple attention module)和空间与通道重构卷积

(spatial and channel reconstruction convolution, SC-Conv),实现了更高效的特征提取和更优化的计算效率,同时引入了联合 Label Smoothing Cross-Entropy 损失函数,这种损失函数结合了标签平滑技术,对传统交叉熵损失进行了改进。首先将 SimAM 注意力

机制集成到 MobileViT 网络中,形成 MobileViT-SimAM 网络。SimAM 模块通过优化能量函数来推断每个神经元的重要性,无需额外参数即可实现对特征图的自适应加权。这种机制在网络的初期阶段增强了对关键物体信息的局部注意力,有效利用图像浅层语义中的关键信息,弥补了在初期阶段对局部特征关注的不足。

在 MobileViT-SimAM 的基础上引入了 SCConv 卷积模块,形成 MobileViT-SS 网络模块。SCConv 模块由空间重构单元 (spatial reconstruction unit, SRU) 和通道重构单元 (channel reconstruction unit, CRU) 组成,SRU 通过独立重组和优化空间信息减少重复信息,而 CRU 则通过将通道分割为更小部分并对其进行转换和融合,以减少冗余信息。这种设计不仅减少了计算量,还促进了特征的学习。如图4所示,MobileViT-SS 网络首先通过一系列普通卷积层对输入图像进行下采样。然后,特征图被输入到

Layer 1 后的 SimAM 注意力模块,以增强局部特征。接着,特征图依次通过 Layer 2 后的 SCConv 卷积块,使得模型在特征通道和空间维度上加强网络对细节特征的提取能力,从而提高分类准确率。在网络的最后阶段,通过 1×1 的普通卷积层和全局池化层,将提取的特征映射到全连接层,最终实现对目标类别的分类。

MobileViT-SS 网络在保持轻量化的同时,通过 SimAM 和 SCConv 模块的协同作用,有效提升模型的分类准确率。这种改进不仅减少了冗余计算,还增强了模型对关键特征的捕捉能力,使其在各种视觉任务中表现出色。此外,联合 Label Smoothing Cross-Entropy 损失函数的引入进一步提高了模型的泛化能力,减少了对训练数据的过拟合。得益于这种创新性的结构设计,MobileViT-SS 网络模型在图像分类任务中实现了性能和效率的双重优化。

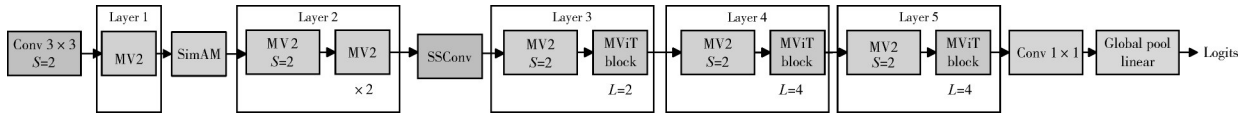


图4 基于改进的 MobileViT 模型网络结构

Fig. 4 Network structure diagram based on improved MobileViT model

1) SimAM 模块。SimAM 通过全局平均池化来提取每个通道的特征,并通过一维卷积实现通道之间的信息传递。在没有增加模型参数的情况下,通过更有效地整合通道和空间维度的信息来增强特征学习的能力,这种方法使得模型能够推断卷积层中特征映射的三维注意力权重,从而学习到更复杂的深层特征。

通过设定神经元的最小能量函数,SimAM 能够评估单个神经元的相对重要性。最小能量函数定义为:

$$e_i^* = -\frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (1)$$

式(1)中, $\hat{\mu}$ 是通道内所有神经元的均值, $\hat{\sigma}^2$ 是通道内神经元的方差, λ 是正则化系数避免分母出现零值, t 是目标神经元。如果 e_i^* 较小,则表示神经元 k 与

其他神经元的差异显著,这使得该神经元在视觉处理中的作用更为重要。

根据计算的能量值,SimAM 通过 Sigmoid 函数归一化将能量值 e_i^* 转换为权重 w_i :

$$w_i = \text{Sigmoid}(e_i^*) \quad (2)$$

最终,输出特征图 Y 由输入特征图 X 与计算得出的权重 w_i 相乘得到:

$$Y = w_i \cdot X \quad (3)$$

SimAM 模块的整体架构如图5所示。该模块包含 $H \times W$ 个独立的注意力子模块,每个子模块对其对应的神经元进行重要性评估和加权处理。通过 Sigmoid 函数对权重进行归一化,确保权重值在 $0 \sim 1$ 。最终,计算得到的权重与原始特征图逐元素相乘生成输出特征图,从而增强模型对重要特征的关注,提高特征表达的有效性。

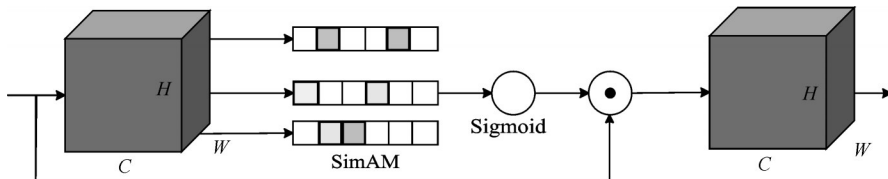


图5 SimAM 模块

Fig. 5 SimAM module

2)SCConv 卷积模块。卷积神经网络的研究与应用中,尽管取得了卓越的性能,但计算资源的巨大需求往往限制了其在资源受限环境中的部署。为了解决这一问题,本研究利用特征之间的空间和通道冗余来进行CNN压缩,采用了一种高效SCConv卷积模块,以减少冗余计算并促进代表性特征的学习。

如图6所示,SCConv模块由2个子单元组成,分

别为空间重构单元(SRU)和通道重构单元(CRU)。具体而言,对于瓶颈残差块中的中间输入特征 X ,首先通过SRU得到空间细化特征 X^w ,然后利用CRU处理获得通道细化特征 Y ,在SCConv模块中利用特征之间的空间冗余和通道冗余,可以无缝集成到任何CNN架构中,以减少中间特征映射之间的冗余并增强CNN的特征表示。

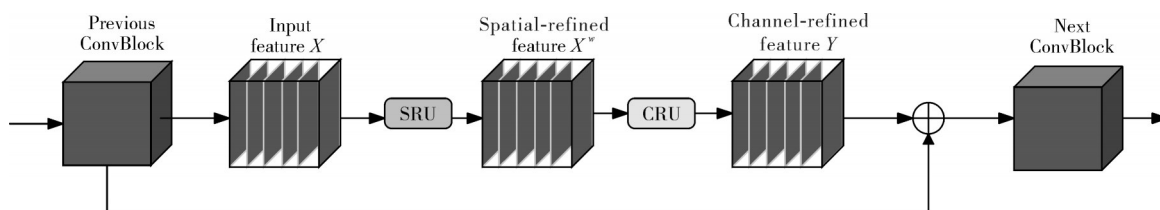


图6 SCConv模块

Fig. 6 SCConv module

如图7所示,SRU通过实施特征空间维度上的分离操作,有效减少输入特征图中的冗余信息,将信息丰富的特征图与信息较少的特征图分离。对输入特征图 X 利用组归一化(group normalization, GN)层中的缩放因子来评估不同特征图的重要性。计算公式如下:

$$X_{out} = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (4)$$

式(4)中, ϵ 是一个为了保证除法稳定性而添加的小正常数; μ 和 σ 是 X 的平均值和标准差; γ 和 β 是可训练的线性变换。

归一化的相关权重 W_γ 表示不同特征图的重要性。计算公式如下:

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^C \gamma_j}, i, j = 1, 2, \dots, C \quad (5)$$

通过 W_γ 重新加权的特征图权重值经过Sigmoid函数映射至(0,1)的范围内,并通过设定的阈值进行门控处理。将高于阈值的权重设置为1,以获得有信息的权重 W_1 ;而对于低于阈值的权重,则设置为0,形成无信息的权重 W_2 (在试验中,阈值设定为0.5)。W的计算公式如下:

$$W = \text{Gate}(\text{Sigmoid}(W_\gamma(\text{GN}(X)))) \quad (6)$$

重构操作将输入特征图 X_1^w 和 X_2^w 分别分离为信息丰富的特征图 X_{11}^w 和 X_{21}^w ,以及信息较少的特征图 X_{12}^w 和 X_{22}^w 。接着,利用组归一化(GN)层中的缩放因子对分离后的特征图进行归一化处理,得到归一化后的特征图 X^{w1} 和 X^{w2} ,通过特征图融合生成更具表达力的特征,从而有效减少了特征空间中的冗余信息。整个过程计算公式如下:

$$\begin{aligned} X_1^w &= W_1 \otimes X \\ X_2^w &= W_2 \otimes X \\ X_{11}^w \oplus X_{22}^w &= X^{w1} \\ X_{21}^w \oplus X_{12}^w &= X^{w2} \\ X^{w1} \cup X^{w2} &= X^w \end{aligned} \quad (7)$$

如图8所示,CRU通过将输入特征图 X^w 分割为信息丰富的通道 X_{up} 和信息较少的通道 X_{low} ,然后分别对 X_{up} 应用压缩卷积和分组卷积得到 Y_1 ,对 X_{low} 应用逐点卷积并与自身拼接得到 Y_2 。在融合阶段,CRU利用全局平均池化收集 Y_1 和 Y_2 的全局空间信息,并通过通道级软注意力机制计算特征重要性向量 β_1 和 β_2 ,以指导 Y_1 和 Y_2 在通道维度上的融合,最终得到通道细化特征 Y 。实现了特征图在通道维度上的优化,显著提升了特征的表达能力,并增强了模型的整体性能和计算效率。

SCConv的设计允许它作为一个插件式卷积模块无缝集成到现有的CNN架构中。通过替代标准卷积操作,SCConv能够在保持性能的同时降低计算复杂性,从而减少视觉任务中的计算成本。这一模块的提出,为轻量级模型的开发和资源受限环境下的深度学习应用提供了新的解决方案。

3)损失函数。在杂草识别的图像分类任务中,MobileViT模型中的标准交叉熵损失函数可能因对标签噪声过于敏感、导致模型过度自信以及泛化能力有限而表现不足。为了克服这些问题,采用了Label Smoothing Cross-Entropy联合损失函数通过引入标签平滑技术,减少了对硬标签的依赖,提高了模型的泛化能力和鲁棒性,同时通过正则化减少了过拟合,改善了模型的校准。

Label Smoothing Loss通过将目标标签从硬标签

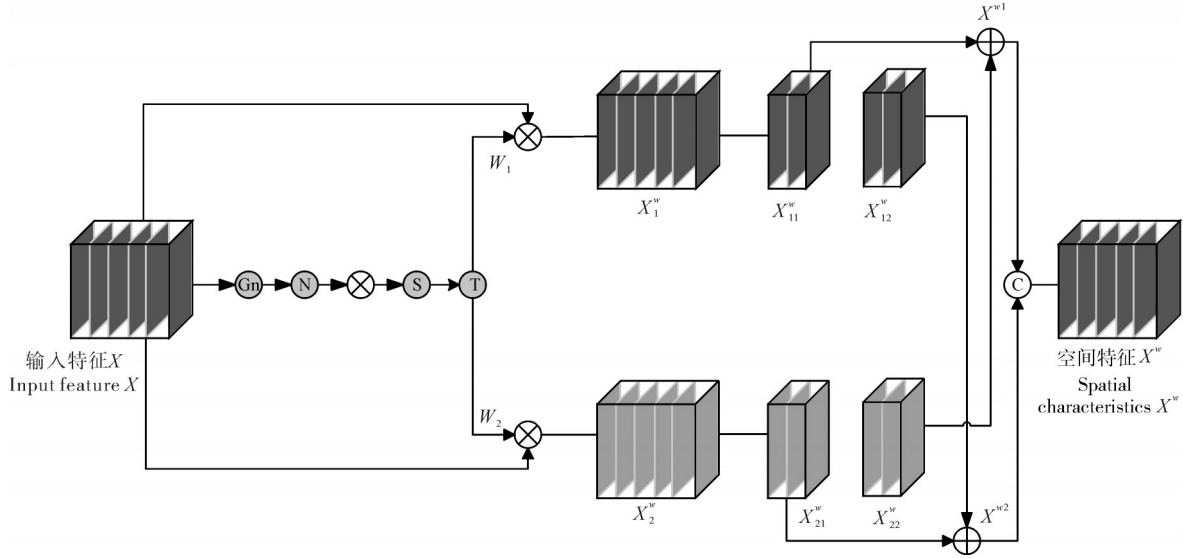


图7 空间重建单元

Fig. 7 Spatial reconstruction unit

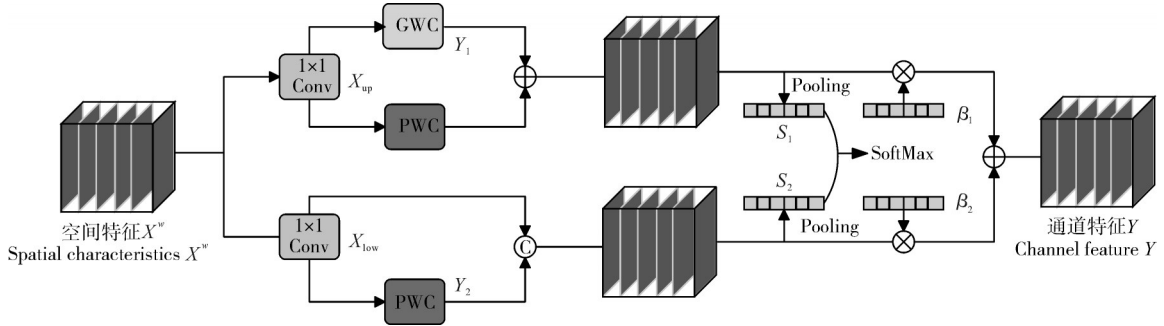


图8 通道重建单元

Fig. 8 Channel reconstruction unit

转变为一个平滑的分布来实现。公式如下:

$$L_{LS} = -\sum_{i=1}^K \left(\frac{\epsilon}{K-1} \log(p_i) \right) + (1-\epsilon) \log(p_y) \quad (8)$$

式(8)中, p_i 是模型预测样本属于类别 i 的概率, p_y 是模型预测样本属于真实类别的概率, ϵ 是平滑参数。

将 Label Smoothing Loss 与 Cross Entropy Loss 结合起来, 联合损失函数可以表示为:

$$L = (1-\lambda)L_{CE} + \lambda L_{LS} \quad (9)$$

其(9)中, 代表 L_{CE} 交叉熵损失, L_{LS} 代表标签平滑损失, 而 λ 是一个用于平衡两个损失项的权重系数。这种方法不仅提高了模型对标签噪声的鲁棒性, 还改善了模型的校准, 使得模型在预测时更加谨慎。

1.5 实验环境与参数设置

为验证模型性能, 作物与杂草识别模型的训练和测试均基于深度学习框架 PyTorch 进行, 系统硬件设施为 GPU: RTX 4090 (24 GB), CPU: 16 vCPU

Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10 GHz 显存。操作系统为 Windows 10, 程序使用 Python 3.8.3 编写, 并在 Anaconda 集成开发环境中进行开发。所用深度学习框架为 PyTorch 1.10.0, 支持 Python 3.8 和 Cuda 11.3。根据实验的训练效果, 将模型的超参数设置如下: 批量大小 (batch size) 设定为 32, 学习率 (learning rate) 为 0.01, 采用参数优化器 AdaW, 网络模型的训练迭代次数 (epoch) 设置为 100 次。

1.6 评价指标

本研究利用识别的准确率 (accuracy, A)、召回率 (recall, R)、精确率 (precision, P)、 F_1 分数以及混淆矩阵, 系统性地评估模型的有效性, 具体计算方式如下:

$$A = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (10)$$

$$R = \frac{T_p}{T_p + F_n} \quad (11)$$

$$P=\frac{T_p}{T_p+F_p}$$
(12)

$$F_1=2\cdot\frac{P\cdot R}{P+R}$$
(13)

式(10)~(13)中, T_p 表示真正例, T_n 表示真负例, F_p 表示假正例, F_n 表示假负例。

2 结果与分析

2.1 不同杂草识别模型对比试验

为验证所提出轻量化杂草识别算法的有效性, 本研究选取多种深度卷积神经网络进行试验分析,

包括 VGG-16^[31]、ResNet-18^[32]模型, 它们在杂草识别任务中已取得较好的效果。此外, 还引入了一些性能优异的轻量化图像识别模型进行对比试验, 如 ShuffleNet^[33]、EfficientNet^[34]、MobileNetv3^[35]模型。由表1可知, 本研究使用的 Mobilevit-SS 模型在多个性能指标上均呈现显著优势, 识别准确率达到 95.91%, 相较于 VGG-16 模型, 识别准确率提高了 3.90 百分点, 同时较轻量化卷积网络 MobileNetv3 也提升了 3.46 百分点。单张图像推理时间仅为 86 ms, 其推理速度高于通用卷积神经网络, 显示出模型在田间杂草识别上的优势。

表1 多种卷积神经网络在杂草识别中的性能比较

Table 1 Performance comparison of convolutional neural networks in weed recognition

模型 Model	模型大小/MB Model size	准确率/% Accuracy	精确率/% Precision	召回率/% Recall	F ₁ 分数/% F ₁ score	推理时间/ms Inference time
VGG-16	621	92.01	91.91	91.56	91.69	640
ResNet-18	84	88.99	89.15	87.11	87.90	174
ShuffleNet	8	89.96	90.45	88.44	89.24	76
EfficientNet	9	91.03	91.18	90.36	90.63	78
MobileNetv3	10	92.45	92.85	91.21	91.89	80
MobileViT-SS	12	95.91	95.97	95.46	95.69	86

由图9可知, 在杂草识别任务中, MobileViT-SS 模型在分类准确率这一性能指标上普遍优于 MobileViT, 展现出更高的性能。MobileViT 模型对雀麦和黑草 2 种杂草的分类错误率较高。这是因为这 2 类杂草在形态特征上非常相似, 增加了模型区分的难

度。得益于无参注意力机制和空间、通道重构单元的加入, MobileViT-SS 模型对雀麦和黑草的识别准确率显著提高, 在区分高度相似的作物与杂草时, 模型展现出了更优的识别精度。

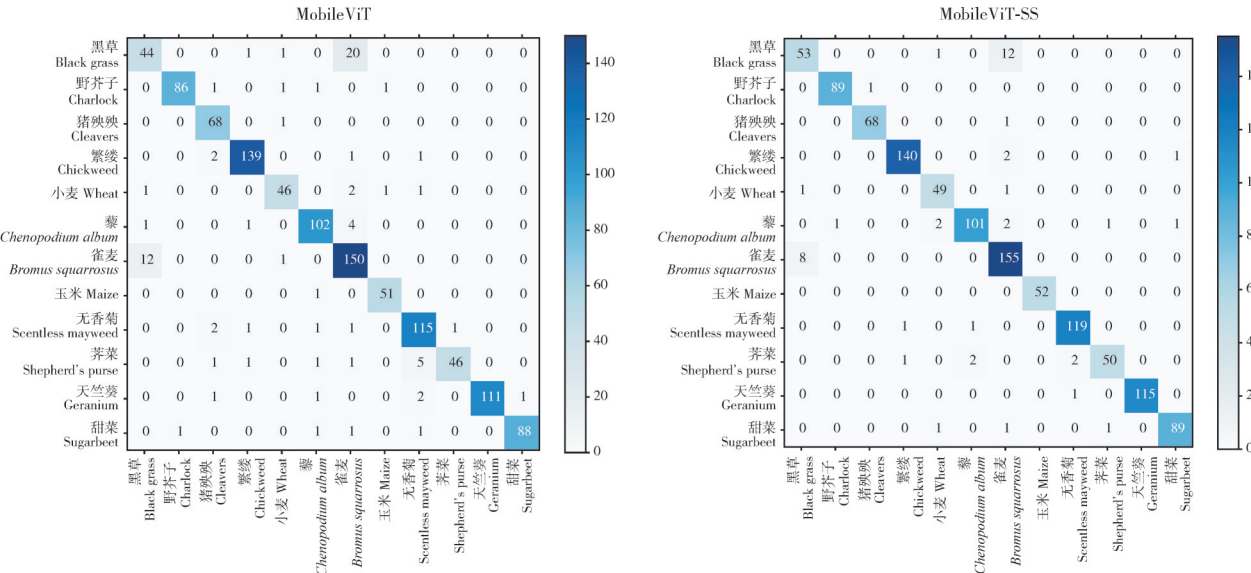


图9 不同模型混淆矩阵

Fig. 9 Different model confusion matrix

2.2 消融实验

为展示 SimAM 在不同注意力机制中的相对优

势, 本研究在 MobileViT 模型的相同网络层中分别引入了 EmA 注意力模块和 SeNet 注意力模块。由表 2

可知,4 种模型在测试集上的识别准确率分别为 93.96%、94.40%、94.58% 和 94.76%。总体来看,加入 SimAM 注意力模块后,模型的识别性能得到显著提升。这一结果表明,与其他注意力机制相比,SimAM 注意力模块在提升模型整体性能方面具有更强的有效性。

表 2 不同注意力机制结果对比

Table 2 Comparison of results of different attention mechanisms

模型 Model	准确率 Accuracy	精确率 Precision	召回率 Recall	F ₁ 分数 F ₁ score
MobileViT	93.96	94.21	92.95	93.52
MobileViT+EmA	94.40	95.06	93.36	94.09
MobileViT+SeNet	94.58	94.55	94.03	94.26
MobileViT+SimAM	94.76	95.65	94.05	94.72

为进一步改善模型精度,本研究引入深度可分离卷积(depthwise separable convolution, DSC)和 SCConv 卷积神经网络结构。由表 3 可知,添加 2 种不同的卷积神经网络后,模型的准确率均得到了提升。其中,加入 DSC 模型的准确率提高了 0.27 百分

点,加入 SCConv 模型的准确率提高了 0.71 百分点,表明 SCConv 卷积网络在提升模型整体性能方面与 DSC 卷积神经网络相比更为有效。

表 3 不同卷积网络结果对比

Table 3 Comparison of results of different convolutional networks

模型 Model	准确率 Accuracy	精确率 Precision	召回率 Recall	F ₁ 分数 F ₁ score
MobileViT	93.96	94.21	92.95	93.52
MobileViT+DSC	94.23	94.96	93.21	93.99
MobileViT+SSConv	94.67	95.33	93.70	94.73

为了探究 SimAM、SCConv 卷积和损失函数 LSCE 对所提出模型性能提升的影响,本研究针对不同变体模型进行消融实验,并进行对比分析。由表 4 可知,改进模型 MobileViT-SS 准确率最高为 95.91%,较模型 MobileViT+SimAM、MobileViT+SCConv、MobileViT+SimAM+SCConv 分别提升 1.15、1.24、0.88 百分点,表明改进方法对模型的性能均有实质提升,可以更有效地解决作物幼苗与杂草的识别问题。

表 4 不同模型的识别性能比较

Table 4 Comparison of recognition performance of various models

模型 Model	SimAM	SCConv	LSCE	准确率 Accuracy	精确率 Precision	召回率 Recall
MobileViT+SimAM	✓			94.76	95.65	94.05
MobileViT+SSConv		✓		94.67	95.33	93.70
MobileViT+SimAM+SSConv	✓	✓		95.03	95.67	94.61
MobileViT-SS	✓	✓	✓	95.91	95.97	95.46

2.3 MobileViT 泛化实验

为了进一步验证 MobileViT 模型的泛化能力,本研究使用来自 Roboflow 上的公开数据集 Weed Species Dataset 进行试验。该数据集包含 2 208 张图片,涵盖 12 种杂草,这些图片均是在各类杂草混杂生长环境下,具有复杂背景的田间杂草图像。在与前述试验相同环境的前提下,将数据集按照 7:3 的比例划分为训练集和验证集,分别使用 MobileNetv3、MobileViT、MobileViT-SS 模型进行试验。由表 5 可知,MobileViT-SS 模型准确率为 93.44%,相较于 MobileNetv3、MobileViT 模型分别提高了 5.10、1.90 百分点,表明改进后的 MobileViT-SS 模型在田间杂草混杂生长的复杂环境下,依然能够展现出良好的识别性能,具备实用价值,可用于识别真实田间环境中的作物幼苗与杂草。

表 5 在 Weed Species Dataset 上的对比试验

Table 5 Comparative experiment on Weed Species Dataset

模型 Model	准确率 Accuracy	精确率 Precision	召回率 Recall	F ₁ 分数 F ₁ score
MobileNetv3	88.34	88.69	87.55	88.25
MobileViT	91.54	91.58	91.55	91.54
MobileViT-SS	93.44	93.20	93.46	93.22

2.4 可视化分析

本研究利用 Grad-CAM (gradient-weighted class activation mapping) 生成模型的可视化热力图,以进一步验证模型对目标形态特征的精确关注程度。由图 10 可知,对于某些高度相似的杂草(如雀麦和黑草),热力图高亮了其重要特征,使模型能够更好地聚焦于目标区域,显著提升模型对形态相似杂草

和作物的区分能力。这种特征聚焦能力亦表明,模型具备有效提取类别相关信息、忽略无关信息的能力。

这一特性进一步增强了分类的准确性和鲁棒性,尤其是在区分形态相似的杂草与作物方面表现突出。

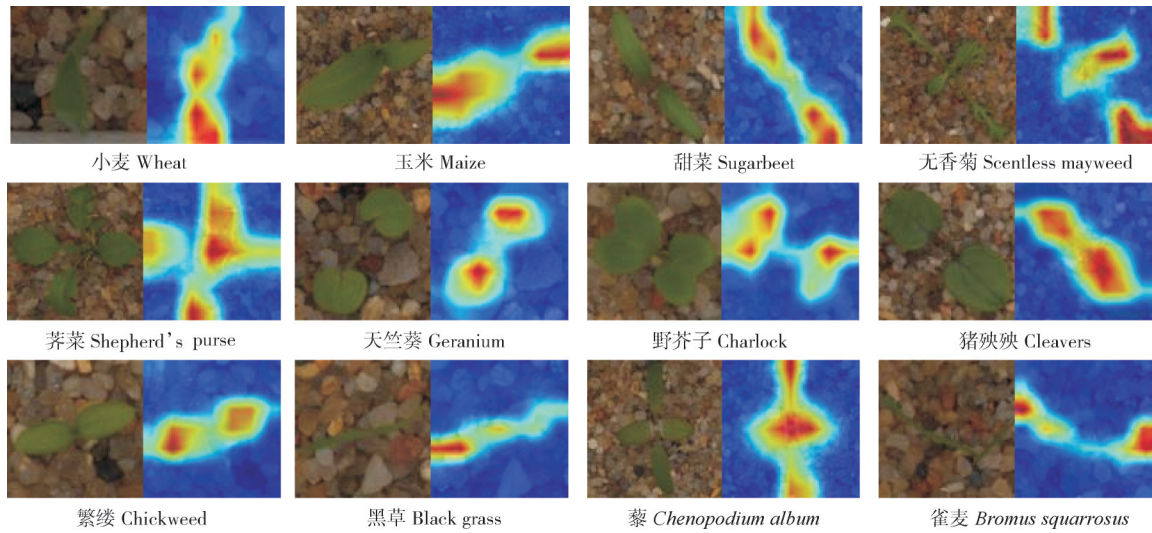


图 10 可视化热力图结果

Fig. 10 Visualized heat map results

2.5 损失分析

损失值是衡量模型预测性能的重要指标,反映了模型预测结果与真实标签之间的差距。在模型训练过程中,损失值通常随着优化的进行而逐渐下降。由图 11 可知,在前 10 次训练周期中,MobileViT-SS 模型在训练集和验证集上的损失值显著下降,说明随着训练的进行,模型在这两者上的预测误差不断减小;随后的迭代过程中,损失值逐渐趋于稳定;经过 100 次迭代后,训练集的损失值稳定在约 0.15,而验证集的损失值则稳定在 0.18 左右,表明改进后 MobileViT-SS 模型训练效果较为理想。

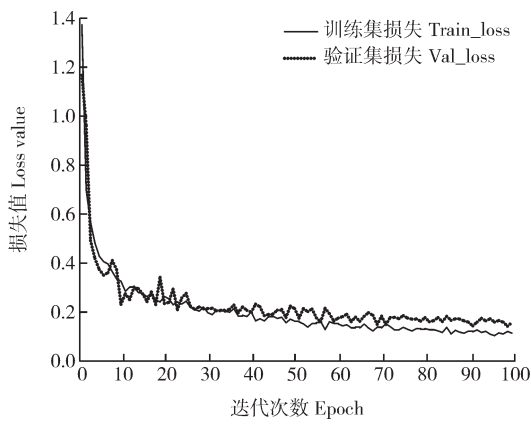


图 11 MobileViT-SS 模型损失函数曲线

Fig. 11 MobileViT-SS model loss function curve

3 讨 论

本研究针对传统农田杂草识别模型在复杂性与

对杂草幼苗识别准确率方面的不足,提出了一种基于优化 MobileViT 架构的轻量级卷积神经网络方法。以提升实际田间环境中杂草幼苗识别的准确性和实时性。通过嵌入轻量级的无参数 SimAM 注意力机制和 SCConv 卷积模块,模型在特征提取和计算复杂度方面得到了显著优化。同时,联合损失函数的引入进一步提升了模型的识别性能。在与原始 MobileViT 模型进行比较时,改进后的模型识别准确率提升 1.51 百分点。相较于传统卷积神经网络架构,改进后的模型展现出更优异的识别性能、更加稳定的收敛过程及更低的参数内存需求。

本研究提出的改进模型在复杂田间环境中有效提高了杂草识别的准确性,为深度学习技术在真实农业场景中的应用提供了切实可行的解决方案。然而仍存在一些局限性,训练数据主要来自特定田间环境,覆盖性不足;极端光照或天气条件下的鲁棒性有待验证。未来研究将扩展数据集多样性,探索多模态数据融合,以增强模型的泛用性和适应性,为农业智能化发展提供更全面的解决方案。

参考文献 References

- [1] MACHLEB J, PETEINATOS G G, KOLLEND A B L, et al. Sensor-based mechanical weed control: present state and prospects[J/OL]. Computers and electronics in agriculture, 2020,

- 176; 105638 [2024-12-24]. <https://doi.org/10.1016/j.compag.2020.105638>.
- [2] 胡炼, 刘海龙, 何杰, 等. 智能除草机器人研究现状与展望[J]. 华南农业大学学报, 2023, 44(1): 34-42. HU L, LIU H L, HE J, et al. Research progress and prospect of intelligent weeding robot [J]. Journal of South China Agricultural University, 2023, 44(1): 34-42 (in Chinese with English abstract).
- [3] CHEN D, LU Y Z, LI Z J, et al. Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems[J/OL]. Computers and electronics in agriculture, 2022, 198: 107091 [2024-12-24]. <https://doi.org/10.1016/j.compag.2022.107091>.
- [4] HASAN A S M M, SOHEL F, DIEPEVEEN D, et al. A survey of deep learning techniques for weed detection from images [J/OL]. Computers and electronics in agriculture, 2021, 184: 106067 [2024-12-24]. <https://doi.org/10.1016/j.compag.2021.106067>.
- [5] GAO J F, FRENCH A P, POUND M P, et al. Deep convolutional neural networks for image-based *Convolvulus sepium* detection in sugar beet fields[J/OL]. Plant methods, 2020, 16: 29 [2024-12-24]. <https://doi.org/10.1186/s13007-020-00570-z>.
- [6] 吴兰兰, 徐恺, 熊利荣. 基于视觉注意模型的苗期油菜田间杂草检测[J]. 华中农业大学学报, 2018, 37(2): 96-102. WU L L, XU K, XIONG L R. Detecting weed in seedling rapeseed oil field based on visual-attention model [J]. Journal of Huazhong Agricultural University, 2018, 37(2): 96-102 (in Chinese with English abstract).
- [7] 疏雅丽, 张国伟, 王博, 等. 基于深层连接注意力机制的田间杂草识别方法[J]. 计算机工程与应用, 2022, 58(6): 271-277. SHU Y L, ZHANG G W, WANG B, et al. Field weed identification method based on deep connection attention mechanism [J]. Computer engineering and applications, 2022, 58(6): 271-277 (in Chinese with English abstract).
- [8] ZHUANG J Y, LI X H, BAGAVATHIANNAN M, et al. Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat [J]. Pest management science, 2022, 78(2): 521-529.
- [9] LUO T Y, ZHAO J Y, GU Y J, et al. Classification of weed seeds based on visual images and deep learning[J]. Information processing in agriculture, 2023, 10(1): 40-51.
- [10] 满超, 饶元, 张敬尧, 等. 基于改进 Inception 网络的复杂环境下小样本黄瓜叶片病害识别[J]. 华中农业大学学报, 2023, 42(3): 152-160. MAN C, RAO Y, ZHANG J Y, et al. Recognition of cucumber leaf disease with small samples in complex environment based on improved Inception network[J]. Journal of Huazhong Agricultural University, 2023, 42(3): 152-160 (in Chinese with English abstract).
- [11] ABADÉ A, FERREIRA P A, DE BARROS VIDAL F. Plant diseases recognition on images using convolutional neural networks: a systematic review[J/OL]. Computers and electronics in agriculture, 2021, 185: 106125 [2024-12-24]. <https://doi.org/10.1016/j.compag.2021.106125>.
- [12] 金小俊, 孙艳霞, 于佳琳, 等. 基于深度学习与图像处理的蔬菜苗期杂草识别方法[J]. 吉林大学学报(工学版), 2023, 53(8): 2421-2429. JIN X J, SUN Y X, YU J L, et al. Weed recognition in vegetable at seedling stage based on deep learning and image processing[J]. Journal of Jilin University (engineering and technology edition), 2023, 53(8): 2421-2429 (in Chinese with English abstract).
- [13] YU J L, SHARPE S M, SCHUMANN A W, et al. Deep learning for image-based weed detection in turfgrass[J]. European journal of agronomy, 2019, 104: 78-84.
- [14] GARCÍA-NAVARRETE O L, CORREA-GUIMARAES A, NAVAS-GRACIA L M. Application of convolutional neural networks in weed detection and identification: a systematic review[J/OL]. Agriculture, 2024, 14(4): 568 [2024-12-24]. <https://doi.org/10.3390/agriculture14040568>.
- [15] OLSEN A, KONOVALOV D A, PHILIPPA B, et al. DeepWeeds: a multiclass weed species image dataset for deep learning[J/OL]. Scientific reports, 2019, 9(1): 2058 [2024-12-24]. <https://doi.org/10.1038/s41598-018-38343-3>.
- [16] 蔡竹轩, 蔡雨霖, 曾凡国, 等. 基于改进 YOLOv5l 的田间水稻稻穗识别[J]. 华南农业大学学报, 2024, 45(1): 108-115. CAI Z X, CAI Y L, ZENG F G, et al. Rice panicle recognition in field based on improved YOLOv5l model[J]. Journal of South China Agricultural University, 2024, 45(1): 108-115 (in Chinese with English abstract).
- [17] 张乐, 金秀, 傅雷扬, 等. 基于 Faster R-CNN 深度网络的油菜田间杂草识别方法[J]. 激光与光电子学进展, 2020, 57(2): 21508. ZHANG L, JIN X, FU L Y, et al. Recognition method for weeds in rapeseed field based on faster R-CNN deep network [J]. Laser & optoelectronics progress, 2020, 57(2): 21508 (in Chinese with English abstract).
- [18] 孙俊, 朱伟栋, 罗元秋, 等. 基于改进 MobileNet-V2 的田间农作物叶片病害识别[J]. 农业工程学报, 2021, 37(22): 161-169. SUN J, ZHU W D, LUO Y Q, et al. Recognizing the diseases of crop leaves in fields using improved MobileNet-V2 [J]. Transactions of the CSAE, 2021, 37(22): 161-169 (in Chinese with English abstract).
- [19] 郭小燕, 于帅卿, 沈航驰, 等. 基于全局特征提取的农作物病害识别模型[J]. 农业机械学报, 2022, 53(12): 301-307. GUO X Y, YU S Q, SHEN H C, et al. Deep learning network for crop disease recognition with global feature extraction [J]. Transactions of the CSAM, 2022, 53(12): 301-307 (in Chinese with English abstract).
- [20] 刘志, 翟瑞芳, 彭万伟, 等. 融合注意力机制的 Cascade R-CNN 田间害虫检测方法[J]. 华中农业大学学报, 2023, 42(3): 133-142. LIU Z, ZHAI R F, PENG W W, et al. Field pest detection method based on improved Cascade R-CNN by incorporating attention mechanism [J]. Journal of Huazhong Ag-

- ricultural University, 2023, 42(3): 133-142 (in Chinese with English abstract).
- [21] CHEN Y J, WU Z N, ZHAO B, et al. Weed and corn seedling detection in field based on multi feature fusion and support vector machine[J/OL]. *Sensors*, 2020, 21(1): 212[2024-12-24]. <https://doi.org/10.3390/s21010212>.
- [22] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: a survey [J]. *Computational visual media*, 2022, 8(3): 331-368.
- [23] 刘文婷, 卢新明. 基于计算机视觉的Transformer研究进展[J]. *计算机工程与应用*, 2022, 58(6): 1-16. LIU W T, LU X M. Research progress of transformer based on computer vision [J]. *Computer engineering and applications*, 2022, 58(6): 1-16 (in Chinese with English abstract).
- [24] 田永林, 王雨桐, 王建功, 等. 视觉Transformer研究的关键问题: 现状及展望[J]. *自动化学报*, 2022, 48(4): 957-979. TIAN Y L, WANG Y T, WANG J G, et al. Key problems and progress of vision transformers: the state of the art and prospects [J]. *Acta automatica sinica*, 2022, 48(4): 957-979 (in Chinese with English abstract).
- [25] LIU Y, ZHANG Y, WANG Y X, et al. A survey of visual transformers [J]. *IEEE transactions on neural networks and learning systems*, 2024, 35(6): 7478-7498.
- [26] 李健, 王晨, 马振宇, 等. MobileNet-CAL: 基于迁移学习和注意力机制的番茄病虫害分类方法[J/OL]. *吉林大学学报(工学版)*, 2023: 1-9[2024-12-24]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20230828>. LI J, WANG C, MA Z Y, et al. MobileNet-CAL: a classification method of tomato diseases and pests based on transfer learning and attention mechanism [J/OL]. *Journal of Jilin University (engineering and technology edition)*, 2023: 1-9[2024-12-24]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20230828> (in Chinese with English abstract).
- [27] 亢洁, 刘港, 郭国法. 基于多尺度融合模块和特征增强的杂草检测方法[J]. *农业机械学报*, 2022, 53(4): 254-260. KANG J, LIU G, GUO G F. Weed detection based on multi-scale fusion module and feature enhancement [J]. *Transactions of the CSAM*, 2022, 53(4): 254-260 (in Chinese with English abstract).
- [28] 徐艳蕾, 何润, 翟钰婷, 等. 基于轻量卷积网络的田间自然环境杂草识别方法[J]. *吉林大学学报(工学版)*, 2021, 51(6): 2304-2312. XU Y L, HE R, ZHAI Y T, et al. Weed identification method based on deep transfer learning in field natural environment [J]. *Journal of Jilin University (engineering and technology edition)*, 2021, 51(6): 2304-2312 (in Chinese with English abstract).
- [29] 杨志凯, 扶兰兰, 唐灿, 等. 基于MobileViT模型的小麦收获机喂入密度分类方法[J]. *农业机械学报*, 2023, 54(S1): 172-180. YANG Z K, FU L L, TANG C, et al. Classification method of feeding density of wheat harvester based on MobileViT model [J]. *Transactions of the CSAM*, 2023, 54(S1): 172-180 (in Chinese with English abstract).
- [30] JIANG H H, ZHANG C Y, QIAO Y L, et al. CNN feature based graph convolutional network for weed and crop recognition in smart farming [J/OL]. *Computers and electronics in agriculture*, 2020, 174: 105450 [2024-12-24]. <https://doi.org/10.1016/j.compag.2020.105450>.
- [31] SENGUPTA A, YE Y T, WANG R, et al. Going deeper in spiking neural networks: VGG and residual architectures [J/OL]. *Frontiers in neuroscience*, 2019, 13: 95 [2024-12-24]. <https://doi.org/10.3389/fnins.2019.00095>.
- [32] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. Las Vegas: IEEE, 2016: 770-778.
- [33] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. Salt Lake City: IEEE, 2018: 6848-6856.
- [34] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]//International conference on machine learning. June 9-15, 2019, Long Beach, California, USA. New York: PMLR, 2019: 6105-6114.
- [35] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27-November 2, 2019. Seoul, Korea. Seoul: IEEE, 2019: 1314-1324.

Lightweighted identification of weed in field based on optimized MobileViT model

LI Ya¹, CHEN Xiaodong¹, WANG Hairui¹, ZHU Guifu^{2,3}

1.School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China;

2.Information Construction Management Center, Kunming University of Science and Technology, Kunming 650504, China;

3.AI Joint Research Center, Kunming University of Science and Technology - Shuguang Information Industry Co., Ltd., Kunming 650504, China

Abstract A lightweighted method based on the optimized MobileViT model was proposed to solve the challenges in identifying weeds from crop seedlings in agricultural environments. SimAM attention mechanism was introduced to enhance the model's ability to pay attention to features. SCConv convolution module was used to reduce the spatial and channel redundancy of features in convolutional neural networks to lower computational costs and model storage, while improving the performance of the convolution module. A loss function strategy combining Label Smoothing Loss and Cross Entropy Loss was proposed to improve the generalization performance of the model, reduce the risk of overfitting, and accelerate the convergence process of the model. Images of 12 common crop seedlings and weeds in the field were used as the training dataset to evaluate the performance of the improved model MobileViT-SS. The results showed that the average recognition accuracy, precision, recall rate, and the F_1 score of the improved model reached 95.91%, 95.97%, 95.46%, and 95.69%, respectively, all of which were superior to that in the widely used deep neural network models including VGG-16, ResNet-18, and MobileNetv3. It is indicated that the improved model MobileViT-SS can accurately and quickly distinguish various weeds from crop seedlings with similar morphology. It will provide technical reference for the identification of weeds from crop seedlings with similar morphology.

Keywords crop seedlings; identification of weed; intelligent agriculture; MobileViT; lightweight; SimAM attention

(责任编辑:葛晓霞)