doi:10.6041/j.issn.1000-1298.2026.18.015

基于 SCBAM - BiGRU - ResNet50 的水稻缺素识别方法

林开颜[1,2]　张镇涛[1,2]　潘婧轩[1,2]　杨学军[1,2]　祝华军[1,2]

(1. 同济大学现代农业科学与工程研究院，上海 201804；2. 同济大学电子与信息工程学院，上海 201804)

摘要：在水稻生长过程中，氮、磷、钾等营养元素缺失会导致叶片变黄、发育迟缓等症状，严重影响水稻生长质量和最终产量。针对传统卷积神经网络在缺素类别识别上泛化能力不足，而新型架构(如 EfficientNetV2 等)因偏重全局建模而在局部病变区域的特征提取能力不足等问题，本文提出了基于卷积神经网络、循环神经网络及注意力机制相结合的识别模型 SCBAM - BiGRU - ResNet50。在 ResNet50 骨干网络中引入风格化卷积注意力机制模块(Style-based convolutional block attention module，SCBAM)，通过增强局部特征表达能力，使模型更聚焦于缺素状态下典型特征。同时结合双向门控循环单元(Bidirectional gated recurrent unit，BiGRU)，利用双向时序学习特性提升特征提取能力。试验结果表明，在水稻缺素特征识别任务中 SCBAM - BiGRU - ResNet50 准确率为 96.97%，表现最佳、显著优于对照组($p < 0.000\,1$)，同时保持了较小的波动范围，验证了本文提出的改进结构在特征表达与序列建模方面的有效性和稳定性。该模型通过融合注意力机制和双向时序建模，有效提升了局部特征响应和上下文理解能力，为当前水稻缺素诊断提供新的解决方案。

关键词：水稻缺素检测；卷积神经网络；风格化卷积注意力机制；双向门控循环单元

中图分类号：S126　　**文献标识码**：A　　**文章编号**：1000-1298(2026)18-0164-10　　**OSID**：

Rice Nutrient Deficiency Identification Method Based on SCBAM - BiGRU - ResNet50

Lin Kaiyan[1,2]　Zhang Zhentao[1,2]　Pan Jingxuan[1,2]　Yang Xuejun[1,2]　Zhu Huajun[1,2]

(1. *Modern Agricultural Science and Engineering Institute*, *Tongji University*, *Shanghai* 201804, *China*
2. *College of Electronic and Information Engineering*, *Tongji University*, *Shanghai* 201804, *China*)

Abstract: During the whole growth cycle of rice, insufficient supply of core essential elements, including nitrogen, phosphorus and potassium easily triggers typical abnormal physiological phenomena, such as leaf chlorosis, slow tillering and overall plant stunting, which not only deteriorate rice appearance quality and nutritional quality, but also cause substantial reduction in actual grain yield and restrict stable agricultural production. Aiming at the problems that traditional convolutional neural networks lacked sufficient generalization ability in identifying nutrient deficiency categories, and novel architectures such as EfficientNetV2 were biased toward global modeling and thus had insufficient feature extraction capability for local lesion areas, an intelligent identification model named SCBAM - BiGRU - ResNet50 was constructed by integrating convolutional structure, recurrent neural network and optimized attention mechanism. The proposed method embeded improved style-based convolutional block attention module into ResNet50 backbone network to strengthen local feature mining and guided the model to concentrate on distinguishable deficiency-related leaf traits. Simultaneously, the integration of bidirectional gated recurrent units (BiGRU) leveraged bidirectional temporal learning to further improve feature extraction. Experimental results demonstrated that SCBAM - BiGRU - ResNet50 achieved superior performance with the accuracy index of 96.97% in rice nutrient deficiency classification tasks, significantly outperforming benchmark models in accuracy ($p < 0.000\,1$) while maintaining low variability. This hybrid network structure effectively boosted local feature sensitivity and deepened contextual feature correlation learning. The research result can provide an efficient, accurate and stable technical reference and innovative intelligent diagnosis scheme for rapid and precise identification of rice nutrient deficiency in modern

收稿日期：2026 - 05 - 07　修回日期：2026 - 05 - 25

基金项目：上海市科委科技创新行动计划课题(23N21900400)

作者简介：林开颜(1975—)，男，副教授，主要从事农业计算机视觉研究，E-mail：linkaiyan@ tongji.edu.cn

precision agriculture.

Key words: rice nutrient deficiency detection; convolutional neural network; style-based convolutional block attention module; bidirectional gated recurrent unit

0　引言

作物病害与缺素症诊断已成为精准农业发展的关键问题，其精准识别对保障粮食安全具有重要意义。作为全球主要粮食作物，水稻因特殊的生理特性易受生物胁迫（病害）与非生物胁迫（缺素症）双重影响[1-2]。若未及时防控，病害可导致产量显著下降，严重时损失超 50%。而且，病害与缺素症叠加效应将导致水稻产量断崖式下降，凸显了构建早期水稻缺素诊断体系、实施精准农艺干预的紧迫性。

深度学习（Deep learning，DL）作为机器学习的重要分支，能够从大量数据中提取复杂特征，是农业图像分析中最有价值的工具[3]。近年来，基于卷积神经网络（Convolutional neural network，CNN）的植物缺素识别有了较多的研究成果。CNN 通过局部感知、权值共享和池化操作，能够高效解析叶片图像纹理、颜色及形态特征，从而实现对氮、磷、钾等营养元素缺乏症状的识别。

采用单一模型对农作物缺素进行识别是常见的研究方法。Dey 等[4]对比发现，在预测缺磷叶片方面相对简单的结构化卷积神经网络性能更好，而 VGG16 在叶枯病和氮缺乏症识别方面精度更高。在静态图像分析中，Islam 等[5]使用基于 ResNet101 的 Mask R－CNN（Mask region-based convolutional neural network）模型，通过迁移学习在小数据集上训练网络，缩短了处理时间。Muthusamy 等[6]基于集成 CNN 的图像分类器，对香蕉叶片微量营养素缺乏症进行诊断，实现了对香蕉缺素的高准确率自动诊断。Bevers 等[7]基于 DenseNet201 架构开发了 CNN 大豆病害分类器，测试准确率达 96.8%。熊俊涛等[8]利用 Mask R－CNN 模型对固定摄像头采集的叶片图像进行分割，去除背景特征，并利用 VGG16 模型进行缺素分类，发现该方法对复杂背景下大豆叶片缺素分类有较好的检测效果。朱红春等[9]探索了基于 Faster R－CNN 目标检测算法在复杂背景下茶叶嫩芽检测方面的应用，发现深度学习目标检测算法在检测精度和速度上明显优于传统目标检测算法，可以较好地识别复杂背景下的茶叶嫩芽。邹强等[10]构建轻量化深度学习模型，引入深度可分离卷积，并通过增加卷积运算步长实现下采样；同时引入 SE（Squeeze-and-excitation）注意力模块，提升病害识别能力。

通过整合多尺度信息、多模型结构以及增强关键特征区分度，可以提升模型识别准确率。谷伟等[11]以 Swin Transformer 作为基础模型，在核心模块中引入 DenseNet 思想，增强特征传递和梯度流动；使用 Outlook Attention 捕捉图像中细节特征，引入深度可分离卷积和膨胀卷积，实现较小参数量前提下捕捉不同尺度特征。Song 等[12]提出融合深度与浅层特征的 FA－Net 网络，通过多头注意力机制融合 ConvNeXt－Base 模型的深度特征与浅层特征，将梨树缺素分类平均准确率提升 15.34 个百分点。Xiao 等[13]引入纹理重建损失并结合 CycleGAN（TRL－GAN）生成柑橘病叶数据，提出 RCGTRL－GAN 算法提升黄龙病识别精度。Chen 等[14]在 ResNet101、MobileNetV2 和 DenseNet121 中集成卷积块注意力模块（Convolutional block attention module，CBAM）和 Focal 损失函数，增强棉花氮元素诊断能力。Zeng 等[15]提出了一种自注意力卷积神经网络（Self-attention convolutional neural network，SACNN）用于提取农作物病斑有效特征识别农作物病害，在 AES－CD9214 和 MK－D2 数据集识别准确率分别为 95.33% 和 98.00%。Li 等[16]使用 DenseNet121 作为主要提取网络，结合多重扩展模块和 CBAM 注意力机制构建了多重扩展 CBAM－DenseNet（MDCDenseNet）模型，在复杂背景下准确率可达 98.84%。张林鎧等[17]针对葡萄病害相似表型症状识别率低及小病斑检测困难的问题，提出了一种基于自适应鉴别器增强的样式生成对抗网络与改进的 YOLO v7 相结合的葡萄病斑检测方法。

近年来，研究者尝试将 CNN 与双向门控循环单元（BiGRU）结合，以增强模型对全局上下文信息的建模能力。Xiao 等[18]提出 ResNet50 与 BiGRU 的混合架构能够模拟人类视觉系统认知逻辑，利用 BiGRU 序列处理特性可丰富由卷积层提取的原始特征表达。Wan 等[19]在处理异常检测任务时，进一步验证了将二维图像特征图展平为一维序列并输入 BiGRU 的有效性，指出这种“空间数据序列化”通过递归机制有效弥补了卷积算子在全局视野上的不足。Zang 等[20]强调 BiGRU 的核心优势在于其“双向性”，这种结构并非为了追溯时间，而是为了在遍历图像特征空间时，通过前向和后向的相互补偿，消除由人工光栅扫描（Raster scan）带来的方向性偏置（Directional bias），从而确保每个空间像素点都能融

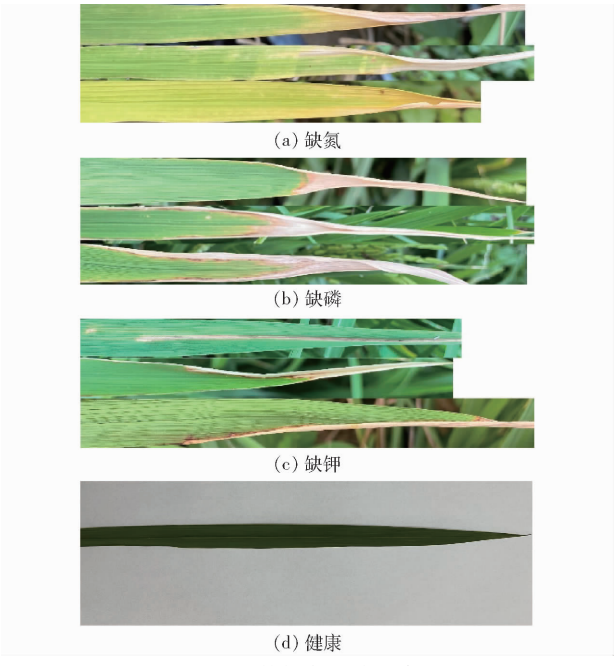
合来自整个图像平面的全局双向依赖信息。

综上,在作物缺素识别任务中,深度学习方法的优势在于能够自动提取高维特征且高效处理复杂非线性数据,无需人工干预,适用于大规模数据集;但部分方法过度依赖于在大规模图像上预训练的深度模型,导致在实际农业图像上泛化能力受限。针对不同缺素症状相似导致模型难以区分,以及因光照变化等多种环境因素干扰而降低模型泛化能力的问题,本文提出融合 ResNet50 和风格化卷积注意力机制模块 SCBAM 进行局部特征分析,使模型更加聚焦于缺素状态下典型缺素特征,以有效提升 ResNet50 模型识别能力;通过引入双向门控循环单元 BiGRU 从前、后 2 个方向学习全局特征信息,以提高模型识别精度。

1 材料和方法

1.1 数据集构建

本文数据集来自 2 个方面,一是 Kaggle 平台上 Nutrient Deficiency Symptoms in Rice[21] 中植物缺素叶片图像数据集,共计 1 156 幅(1.05 GB),其中缺氮叶片 440 幅、缺磷叶片 333 幅、缺钾叶片 383 幅;二是通过人工采摘拍摄的方式获取健康水稻叶片图像 172 幅,采摘地点位于江苏省镇江市上党镇义村附近。数据集如图 1 所示。



(a) 缺氮 (b) 缺磷 (c) 缺钾 (d) 健康

图 1 数据集内容示例

Fig. 1 Example content of dataset

由于采集设备、光照条件及背景环境存在差异,可能引入域偏移(Domain shift)问题。为减小客观差异带来的影响,在实验前对数据进行背景裁剪、统一分辨率以及随机亮度和对比度抖动等处理。

识别叶片可视为对图像进行显著性目标检测,即在复杂背景中找出图像主体并将背景分离。U^2-Net[22] 用于数据集去除背景,是一个 2 级嵌套 U 结构,专为显著目标检测(Salient object detection, SOD)设计,不使用任何来自图像分类的预训练主干。U^2-Net 通过嵌套 U 结构有效融合了不同尺度的特征,有助于模型同时捕捉全局和局部显著性信息。对原始数据集进行显著性目标提取后的效果如图 2 所示。为适应模型训练并缓解不同来源数据集域偏移问题,统一转换为 224 像素 ×224 像素并添加黑色背景。

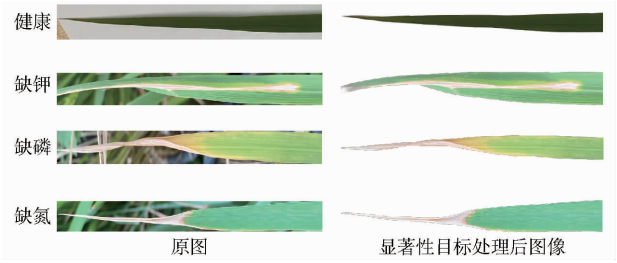


图 2 水稻叶片显著性目标提取效果

Fig. 2 Effectiveness of significant visual feature extraction from rice leaves

在真实环境中,拍摄时受天气、光照条件影响,同一株水稻叶片可能会有不同效果的图像,主要体现在对比度与亮度上。为模拟真实环境中光照变化对图像的影响,采用基于高斯随机扰动的数据增强方法。根据生成的正态分布随机变量

$$Z \sim N(0,1) \tag{1}$$

式中 Z——正态分布随机变量

使用 Sigmoid 函数进行映射

$$\delta = \frac{1}{1+e^{-Z}} \tag{2}$$

式中 δ——最终生成的亮度或对比度抖动系数,取 0 ~ 1

亮度和对比度调整计算式为

$$I_b(x,y) = k_1 I(x,y) \tag{3}$$

$$I_c(x,y) = k_2(I(x,y) - \bar{I}) + \bar{I} \tag{4}$$

式中 $I(x,y)$——原始像素

k_n——根据抖动系数 δ 生成的抖动范围,取 $1-\delta \sim 1+\delta$

$\bar{I}$——图像平均灰度

$I_b(x,y)$——亮度调整后像素

$I_c(x,y)$——对比度调整后像素

图 3 以健康水稻数据集为例,展示了图像经过亮度(式 3)与对比度抖动(式 4)后效果。

为定量评估数据预处理策略对域偏移的缓解效果,基于 ResNet 提取的深度特征,采用归一化均值距离(Normalized mean distance, NMD)对特征分布

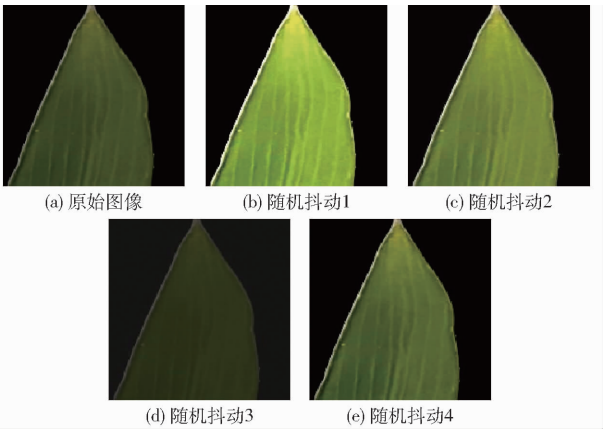


图 3　健康水稻图像亮度与对比度抖动增强示例

Fig. 3　Examples of brightness and contrast jitter augmentation for healthy rice image

差异进行度量。实验结果表明，预处理前 NMD 为 1.179 2，经过背景裁剪、分辨率统一及亮度抖动等处理后，NMD 降至 1.110 9，下降约 5.8%。结果表明，预处理操作在一定程度上减小了不同数据分布之间的差异，使特征空间中的分布更加紧凑，有助于提升模型训练稳定性与泛化能力。

由于健康水稻叶片图像与缺素水稻叶片图像来源不同，二者在整体图像风格上存在一定差异。为缓解分布不一致问题，并有效扩充缺素类别样本数量，引入基于图像风格迁移的 ADC－CycleGAN 模型[23]，以健康水稻叶片为输入，生成具有与目标缺素水稻图像风格一致的合成图像。图 4 为利用 ADC－CycleGAN 从健康水稻图像生成风格匹配的缺素图像效果。

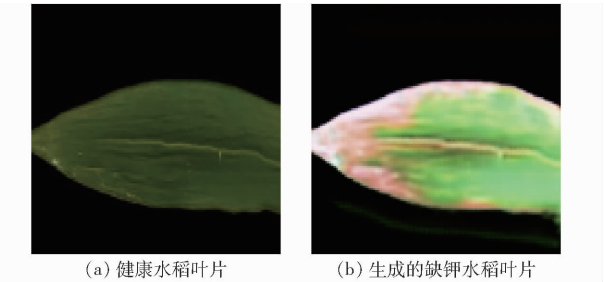


图 4　基于 ADC－CycleGAN 的缺素水稻叶片图像生成

Fig. 4　Generating nutrient-deficient rice leaf images with ADC－CycleGAN

为验证 ADC－CycleGAN 生成图像质量，采用弗雷歇起始距离（Frechet inception distance，FID）对生成图像与真实缺素图像进行分布一致性评估，并邀请农业领域专家进行人工判读。结果表明生成图像在视觉特征与症状表现上与真实图像的一致性可接受。表 1 为利用 ADC－CycleGAN 从健康叶片生成缺氮、缺磷、缺钾图像以及从缺素图像生成健康水稻图像的情况。表中“健康→缺氮”表示从健康叶片生成缺氮叶片图像，其他同理。

表 1　ADC－CycleGAN 从健康叶片生成缺氮、磷、钾图像情况

Tab. 1　ADC－CycleGAN for generating N, P and K deficiency images

图像生成方式	生成数量/幅	FID
健康→缺氮	600	37.13
健康→缺磷	600	30.11
健康→缺钾	600	39.72
缺氮→健康	200	19.75
缺磷→健康	200	20.65
缺钾→健康	200	13.32

在经过数据集构建、显著性目标提取、传统方法数据集扩充以及基于 GAN 方法对水稻数据集扩充后，水稻缺素叶片图像扩充情况如表 2 所示。样本数据按照训练集、验证集和测试集以比例 6∶2∶2 进行划分，其中 ADC－CycleGAN 生成模型所生成的图像仅用于扩充训练集，以提高模型鲁棒性和泛化能力。

表 2　图像扩充情况

Tab. 2　Image augmentation summary　　幅

图像内容	原始数量	扩充后数量	训练集	验证集	测试集
健康	172	2 600	1 560	520	520
缺氮	440	2 600	1 560	520	520
缺磷	333	2 600	1 560	520	520
缺钾	383	2 600	1 560	520	520

1.2　基于 SCBAM－BiGRU－ResNet50 模型的水稻缺素识别方法

提出了一种将卷积神经网络、循环神经网络（Recurrent neural network，RNN）以及注意力机制（Attention mechanism）相结合的模型 SCBAM－BiGRU－ResNet50，用于水稻缺素特征识别与诊断。如图 5 所示，为了高效提取特征并减少展平图像计算成本，采用 ResNet50 作为图像特征编码器，并引入 SCBAM 增强局部特征表达能力；同时，将 ResetNet50 中 Conv4 输出经 SCBAM 加权后，其特征从矩阵转换为序列数据后，输入到双向门控循环神经网络以从前、后 2 个方向学习叶片缺素的全局空间依赖关系，然后将输出数据（1 024 维）传回并与卷积神经网络输出（2 048 维）拼接以形成最终的决策数据（3 072 维），最终将其传递到分类层进行分类。

SCBAM－BiGRU－ResNet50 模型包含的主要模块：

（1）输入层：对训练集中图像进行预处理，处理后图像分辨率为 $\mathbb{R}^{224\times224\times3}$。

（2）CNN 层：以 ResNet50 作为主干网络，包括卷积层、池化层、残差块等。其中，Conv$N$_k 中的 N

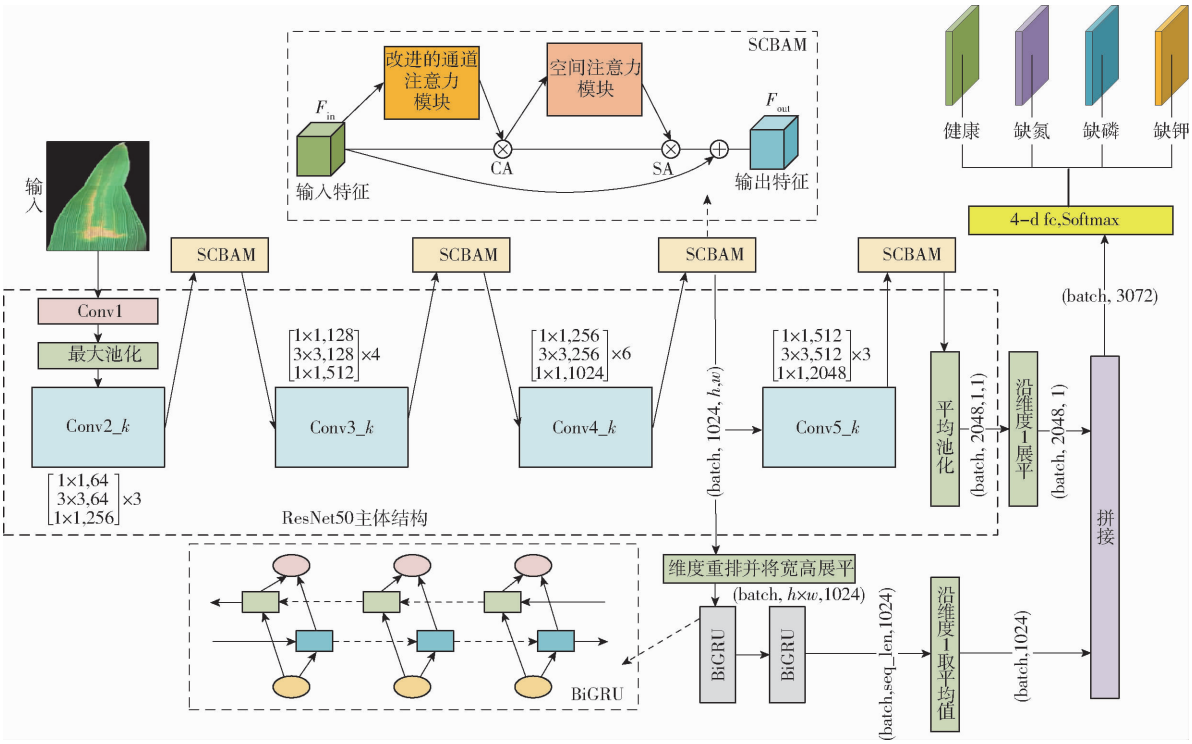


图 5 SCBAM－BiGRU－ResNet50 模型架构

Fig. 5 Architecture of SCBAM－BiGRU－ResNet50 model

表示模型中第几个卷积层，k 表示不同数量卷积层组。

(3) SCBAM：在 ResNet50 的每个残差块最后(除了 Conv1_k)，插入风格化的卷积注意力机制模块 SCBAM，通过 SCBAM 可得到不同输入特征权重。

(4) BiGRU：ResNet50 的 Conv4 输出经过 SCBAM 加权后，其特征输入至 2 个连续的 BiGRU 模块中。BiGRU 由 2 个 GRU 层(前向和反向)构成，支持信息双向流动，输出由前向和后向 GRU 模块共同作用产生。

(5)平均池化层：ResNet50 原模型平均池化。

(6)输出层：将平均池化后 2 048 维向量和 BiGRU 输出 1 024 维向量展平、拼接成 3 072 维向量后，输入到 fc 全连接层并通过 Softmax 函数进行分类，输出类别概率。

1.2.1 主干网络 ResNet

残差网络(Residual neural network，ResNet)是由 He 等[24]提出的用于图像识别的深度学习模型，旨在解决深层神经网络训练时梯度消失和梯度爆炸问题。ResNet 核心思想是通过残差块(Residual block)引入跳跃连接(Shortcut connection)，使网络学习残差映射，从而简化优化过程并提高训练稳定性。

在传统深度神经网络中，假设网络期望映射为 $H(x)$，则 ResNet 通过构造残差块，使网络学习的是 $H(x)-x$ 而非直接学习 $H(x)$。计算式为

$$H(x)=F(x)+x \tag{5}$$

式中 x——输入数据

$F(x)$——网络层学习的残差映射

跳跃连接使得输入 x 直接加到输出上，形成恒等映射(Identity mapping)。其优势在于当网络层数较深时，梯度更容易地反向传播，从而缓解梯度消失问题。此外，即使某些层的学习效果不佳，跳跃连接仍可保证信息流稳定传递，使深层网络不会出现退化问题。

ResNet 主要由基本残差块(Basic residual block)和瓶颈残差块(Bottleneck block)组成。基本残差块主要用于 ResNet18 和 ResNet34，由 2 个 3×3 卷积层组成，采用恒等映射跳跃连接，直接将输入加到输出上。瓶颈残差块用于 ResNet50 及以上，将特征图利用 1×1 卷积进行维度压缩后，进行 3×3 卷积，再使用 1×1 卷积恢复特征图维度。ResNet 包括多种版本，其中 ResNet50 采用瓶颈残差块，适用于更复杂的计算机视觉任务。ResNet50 网络结构如表 3 所示。

1.2.2 注意力机制 SCBAM

在水稻缺素识别任务中，叶子通常会反映缺素、病虫害、冻害等多种状态，识别模型应具有在复杂特征中过滤无关信息并提取主要特征的能力。卷积注意力机制模块 CBAM 结合了通道注意力和空间注意力，在通道和空间 2 个维度上依次生成注意力特征图，并对原输入特征图进行加权实现自适应特征

表 3　ResNet50 网络结构

Tab. 3　ResNet50 network architecture

层名称	输出大小	操作
Conv1	112×112	7×7 卷积，步长 2
最大池化层	56×56	3×3 最大池化，步长 2
Conv2_k	56×56	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$
Conv3_k	28×28	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$
Conv4_k	14×14	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1\,024\end{bmatrix}\times6$
Conv5_k	7×7	$\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2\,048\end{bmatrix}\times3$
平均池化层	1×1	全局平均池化
classifier		4－d fc，Softmax

修正，最终生成精炼特征图[25]，从而有效提升模型对重要特征的关注度和表示能力。

整合 CBAM 注意力机制模块与基于样式重校准的注意力机制模块（Style-based recalibration module，SRM）[26]的优点，本文提出一种风格化的卷积注意力机制模块（Style-based convolutional block attention module，SCBAM），如图 6 所示。

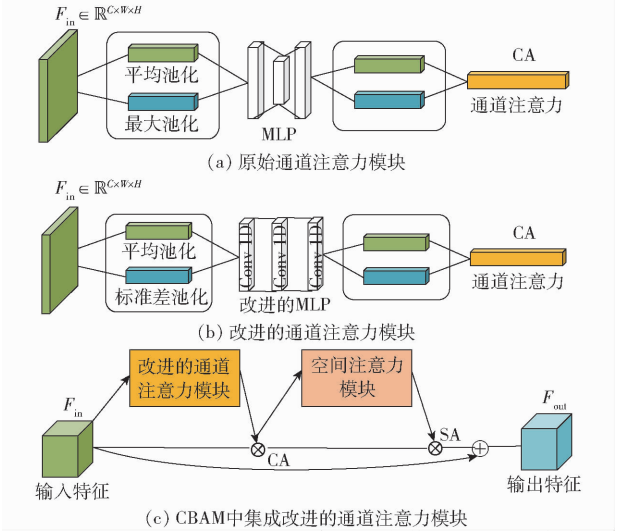


图 6　基于风格化的卷积注意力机制模块 SCBAM 架构

Fig. 6　Architectures of stylized convolutional attention module（SCBAM）

SCBAM 的改进体现在对 CBAM 中原通道注意力模块的改进，如图 6b 所示。由图 6b 可得

（1）将最大池化替换为标准差池化。标准差池化捕捉了局部区域内特征的波动和变化程度，可在特征中保留更多的信息。通道注意力模块将待识别特征F_{in}通过标准差池化（Standard pooling，StdPool）与平均池化（Average pooling，AvgPool）生成对应特征F_{std}^{c}及F_{avg}^{c}，计算式为

$$F_{std}^{c}=\sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(x_{nchw}-\mu_{nc}\right)^{2}} \tag{6}$$

$$F_{avg}^{c}=\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}x_{nchw} \tag{7}$$

式中　x_{nchw}——输入特征图 $\boldsymbol{X}$ 中位于批次 n、通道 c、高度 h、宽度 w 处的元素

μ_{nc}——第 n 个样本、第 c 个通道平均值

（2）将多层感知器替换为 3 个 1 维卷积。参数量小、权值共享，可高效地提取跨通道的依赖信息，避免多层感知器（Multilayer perceptron，MLP）层会丢失部分信息的问题。将 2 种注意力特征逐元素求和，生成通道注意力特征向量（Channel attention，CA），其形状为

$$CA\in\mathbb{R}^{C\times1\times1} \tag{8}$$

CA 与输入特征F_{in}相乘得到通道精炼特征（Channel-refined feature，F_{cr}）。对空间注意力机制向量进行推导：在通道维度进行平均池化和最大池化，将产生的特征图进行拼接，并使用卷积操作生成最终的空间注意力特征矩阵（Spatial attention，SA），最后将 SA 与F_{cr}相乘得到最终输出特征F_{out}。

1.2.3　双向门控循环单元 BiGRU

门控循环单元（Gated recurrent unit，GRU）[27]是一种循环神经网络（Recurrent neural network，RNN）的改进版本。GRU 解决了标准 RNN 的梯度消失问题，相对于 LSTM 结构更简单，计算效率更高，适用于较长序列的数据建模。

BiGRU 是一种双向循环神经网络（Bidirectional recurrent neural network，BiRNN）的变体，基于 GRU 架构。如图 7 所示，BiGRU 由 2 个 GRU 单元组成，一个前向传播（从左到右），另一个反向传播（从右到左）。2 个 GRU 输出通常被拼接或相加，以形成最终的隐层状态。

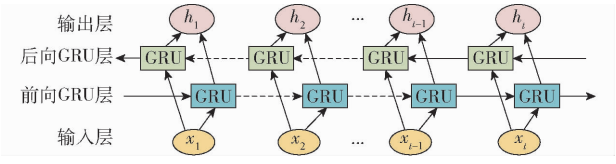


图 7　BiGRU 模块架构图

Fig. 7　Architecture diagram of BiGRU module

相比单向 GRU，BiGRU 能够更充分地利用序列的上下文信息，从而提高模型表达能力。ResNet50 的卷积操作受限于局部感受野，难以直接捕获距离较远的空间特征。将特征图展平并输入序列模型，是为了建立像素/特征块之间的长距离空间依赖关系。虽然图像是二维，但通过特定的空间遍历策略

(此处为逐行光栅扫描),可将二维空间信息映射为一维序列。这种做法将特征图视为一组特征词元(Feature tokens)的集合,使得网络能够以线性计算复杂度处理全局信息。因为人工强加的光栅扫描顺序(从左上到右下)不对称,会引入一种人为的"方向偏置",导致序列末端特征能看到前面的信息,而前端特征缺乏全局视野。BiGRU 反向传播(从右下到左上)并非为了寻找因果关系,而是为了对冲这种方向偏置,确保一维序列中每个特征点(Token)在更新状态时,能融合来自整个二维图像平面的全局上下文信息。在 SCBAM - BiGRU - ResNet50(图 5)中,BiGRU 模块参数设置如表 4 所示。

表 4 BiGRU 模块参数设置

Tab. 4 BiGRU module parameter settings

模块	输入序列维度	序列长度	隐层状态维度	层数	输出序列维度
BiGRU - 1	1 024	196	256	1	512
BiGRU - 2	512	196	256	1	512
BiGRU 输出拼接	BiGRU - 1 + BiGRU - 2	196			1 024

由于卷积网络处理的是二维特征图,而 BiGRU 双向循环网络处理的是序列数据,需将二维特征图进行"展平",变成序列数据,以利于 BiGRU 处理。具体地,将 ResNet50 的 Conv4_k 残差块最后一个输出首先调整维度为(batch, seq_len, 1 024),其中 batch 为批量大小;seq_len 为序列长度,在本模型中为 $H \times W$,即特征图空间维度展开的长度。在 ResNet50 中,由于 Conv4_k 输出为 14×14,故序列长度 seq_len 为 196,即对应图 7 中 x_t 序列长度。选择 Conv4_k 的理由是高层特征图输出尺寸较小,有利于减少双向循环网络输入序列长度,减少模型参数量。"1 024"表示特征通道维度,对应 Conv4_k 残差块输出的特征维度,展开后输入 2 个连续的 BiGRU 模块。同时 Conv4_k 最后一个输出特征经过 Conv5_k 以及 SCBAM 注意力机制模块后,进行全局池化得到一个 2 048 维的特征。最后对 BiGRU 输出的 1 024 维特征以及 RestNet50 - SCBAM 输出的 2 048 维特征进行拼接,得到 3 072 维的最终特征序列用于分类。

提出的算法在 ResNet50 中融合 BiGRU 双向门控循环单元的目的是利用其强大的序列建模能力,捕捉高层特征图感受野之间的空间特征关系,并结合卷积神经网络强大的局部特征提取能力,从局部和全局 2 个视角进行建模,提高模型对缺素特征的识别精度和抗干扰能力。

2 试验结果与分析

2.1 试验平台与超参数设置

操作系统为 Windows 10 专业版 22H2,配置如下:内存 32 GB,显卡为 NVIDIA GeForce RTX 4080,显存 16 GB,CPU 为 Intel(R) Core(TM) i7 - 12700K。所用的开发工具和库包括 CUDA 12.1.98、Python 3.9.13 和 PyTorch 1.12.1。

试验过程使用交叉熵损失函数(Cross-Entropy Loss)作为分类任务的优化目标,用于衡量预测结果与真实标签之间的差异。优化器选用 Adam,并设置权重衰减(Weight decay)以抑制模型过拟合,提高模型泛化能力。模型超参数设置为:初始学习率为 1×10^{-3},训练次数为 30 轮(epoch),每批次样本数(Batch size)为 32,输入图像尺寸为 224 像素 × 224 像素。在训练过程中,结合学习率衰减策略对学习率进行动态调整,以提高模型后期收敛的稳定性。为缓解过拟合问题,训练过程中引入权重衰减正则化项,并结合数据增强策略(包括随机裁剪、亮度扰动及水平翻转等),以提升模型泛化能力。所有实验参数配置相同,以保证不同模型之间对比的公平性与可重复性。

ResNet50 模块未采用 ImageNet 预训练权重,而是基于随机初始化参数进行端到端训练。考虑到水稻叶片图像与 ImageNet 自然图像在数据分布及纹理特征方面存在显著差异,直接迁移预训练权重可能影响模型对细粒度缺素特征的学习效果。因此,选择从零开始训练模型,以充分学习数据集自身的特征表达。

2.2 模型性能试验

为了验证本文提出的水稻缺素特征识别模型 SCBAM - BiGRU - ResNet50 的性能,将 SCBAM - BiGRU - ResNet50 模型与 3 个对照组模型进行对比。①经典模型对照组:对象为 VGG16[28]。②基于最新 CNN 模型对照组:基于 CNN 的图像分类模型不断演化,涌现出众多新型架构,对象为 FasterNet[29] 以及 EfficientNet - V2 - m[30]。③基于最新 CNN + Transformer 的混合结构(Hybrid)模型对照组。近年来,融合 CNN 与 Transformer 优势的 Hybrid 混合结构模型逐渐兴起,成为一种具有广泛研究潜力的新型架构范式,对象为 MaxViT[31]、MobileViT - V2[32] 以及 VanillaNet6[33]。为验证模型性能稳定性与统计显著性,对各模型进行 5 次独立训练,并进一步采用独立样本 t 检验对本文方法与次优模型进行显著性检验。

各模型准确率(Accuracy)、宏平均 F1 分数

(Macro-F1)与参数量试验结果如表 5 所示。可以看出,提出的 SCBAM－BiGRU－ResNet50 模型准确率(Accuracy)达 (96.97±0.21)% ,明显优于其他对比模型。其中,性能较优的 EfficientNet－V2－m 达到(94.31±0.28)%;在 Hybrid 模型中,MobileViT－V2 达到(93.18±0.44)%。与上述模型相比,本文模型分别提升 2.66、3.79 个百分点。相较于传统模型 VGG16 与 FasterNet,性能提升更为明显,所提出结构在特征表达能力与序列建模能力方面具有明显优势。此外,从标准差可以看出,SCBAM－BiGRU－ResNet50 的波动范围较小,说明模型在多次独立实验中具有良好的稳定性。

为验证模型性能差异统计显著性,采用等方差独立样本 t 检验($n_1=n_2=5$, $df=8$)对本文模型与各对比模型准确率进行比较,结果如表 6 所示。

表 5　各模型性能试验结果

Tab. 5　Performance experimental results of models

模型	准确率/%	宏平均 F1 分数/%	参数量
VGG16	87.45±0.42	88.69±0.6	1.34×10^8
FasterNet	87.39±0.40	87.32±0.8	3.90×10^6
EfficientNet－V2－m	94.31±0.28	93.07±0.5	5.41×10^7
MaxViT	90.12±0.30	91.49±0.6	9.76×10^6
MobileViT－V2	93.18±0.44	94.66±0.6	4.38×10^6
VanillaNet6	91.01±0.33	92.08±0.8	3.02×10^8
SCBAM－BiGRU－ResNet50	96.97±0.21	97.02±0.3	6.12×10^7

表 6　SCBAM－BiGRU－ResNet50 与各对比模型独立样本 t 检验结果

Tab. 6　Independent sample t-test results between SCBAM－BiGRU－ResNet50 and comparison models

模型	准确率/%		均值差/%	df	t	p	显著性
	对照模型	实验组					
SCBAM－BiGRU－ResNet50	—	96.97±0.21	—	—	—	—	—
VGG16	87.45±0.42	96.97±0.21	9.52	8	45.41	<0.000 1	极显著
FasterNet	87.39±0.40	96.97±0.21	9.58	8	47.11	<0.000 1	极显著
EfficientNet－V2－m	94.31±0.28	96.97±0.21	2.66	8	16.92	<0.000 1	极显著
MaxViT	90.12±0.30	96.97±0.21	6.85	8	41.87	<0.000 1	极显著
MobileViT－V2	93.18±0.44	96.97±0.21	3.79	8	17.51	<0.000 1	极显著
VanillaNet6	91.01±0.33	96.97±0.21	5.96	8	34.05	<0.000 1	极显著

由表 6 可知,SCBAM－BiGRU－ResNet50 在准确率上显著优于所有对比模型($p<0.000\ 1$)。具体而言,与 EfficientNet－V2－m 相比,$t=16.92$($p<0.000\ 1$);与 MobileViT－V2 相比,$t=17.51$($p<0.000\ 1$);与 MaxViT 相比,$t=41.87$($p<0.000\ 1$);与 VGG16 和 FasterNet 相比,差异更加显著($t>45$, $p<0.000\ 1$)。上述结果表明,本文模型在统计意义上具有显著且稳定的性能优势。综上所述,SCBAM－BiGRU－ResNet50 在保持较小波动范围的同时,实现了准确率显著提升,验证了所提出改进结构在特征提取与序列建模方面的有效性。

为进一步分析模型在不同类别之间的判别能力,引入混淆矩阵对分类结果进行可视化分析。如图 8 所示,模型在“健康”类别上的识别效果最佳,正确分类样本数达到 518,占该类别总样本的绝大多数,几乎不存在误分类现象。相比之下,“缺钾”“缺氮”和“缺磷”之间存在一定程度的混淆。其中,“缺钾”与“缺氮”之间存在双向误分类(10 次与 7 次),而“缺磷”与“缺氮”的混淆更为明显(21 次),说明这 2 类在特征空间中分布较为接近。造成上述现象的主要原因在于不同缺素症状在视觉表现上的相似性。“缺氮”和“缺钾”均表现为叶片黄化,颜色

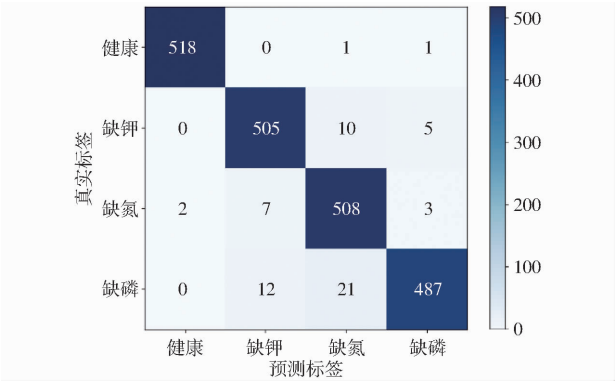


图 8　SCBAM－BiGRU－ResNet50 水稻缺素分类混淆矩阵

Fig. 8　Confusion matrix for rice nutrient deficiency classification using SCBAM－BiGRU－ResNet50

与纹理特征具有较高相似度;而“缺磷”在早期阶段特征不显著,容易被模型误判为其他缺素类型。总体而言,模型在各类别上识别准确率较高,混淆主要集中在少数视觉特征相似的类别之间,表明模型已具备较强的特征提取与判别能力。

2.3　消融试验

为了更全面地验证 SCBAM－BiGRU－ResNet50 模型在水稻缺素识别任务上的优越性,设计了消融试验。为验证模型性能稳定性与统计显著性,对各模型进行 5 次独立训练,采用独立样本 t 检验对本

文方法与次优模型进行显著性检验。各模型准确率和宏平均 F1 分数如表 7 所示。

各模型均进行 5 次独立重复试验($n_1 = n_2 = 5$),自由度 $df = 8$。显著性水平设定为 0.05,结果如表 8 所示。试验结果表明,SCBAM - BiGRU - ResNet50 与其他对比模型之间的性能差异均具有统计显著性,验证了所提出模型在水稻缺素识别任务中有效性与稳定性。

表 7 消融试验模型性能

Tab. 7 Model performance in ablation studies %

模型	准确率	宏平均 F1 分数
SCBAM - BiGRU - ResNet50	97.09 ± 0.45	97.03 ± 1.2
SCBAM - ResNet50	95.38 ± 0.39	96.57 ± 0.7
BiGRU - ResNet50	93.55 ± 0.36	93.45 ± 0.5
ResNet50	92.67 ± 0.44	92.58 ± 0.3

表 8 消融试验显著性分析

Tab. 8 Statistical significance of ablation results

模型	准确率/%		均值差/%	*df*	*t*	*p*	显著性
	对照模型	实验组					
SCBAM - BiGRU - ResNet50	—	97.09 ± 0.45	—	—	—	—	—
SCBAM - ResNet50	95.38 ± 0.39	97.09 ± 0.45	+1.71	8	6.12	< 0.01	显著
BiGRU - ResNet50	93.55 ± 0.36	97.09 ± 0.45	+3.54	8	10.87	< 0.01	显著
ResNet50	92.67 ± 0.44	97.09 ± 0.45	+4.42	8	13.25	< 0.01	显著

为进一步验证 SCBAM 注意力机制在特征提取过程的作用,对模型的注意力响应进行可视化分析。如图 9 所示,模型在不同类别图像中均能够自适应地关注叶片的关键区域。原始 ResNet50 模型提取的水稻缺素特征图大多特征不明显(绿色低热力区居多),且多为叶片本身纹路走向,无法准确定位缺素部分的特征。添加 SCBAM 注意力机制后,对缺素叶片特征的注意力有一定加强(红色高热力区增多)。

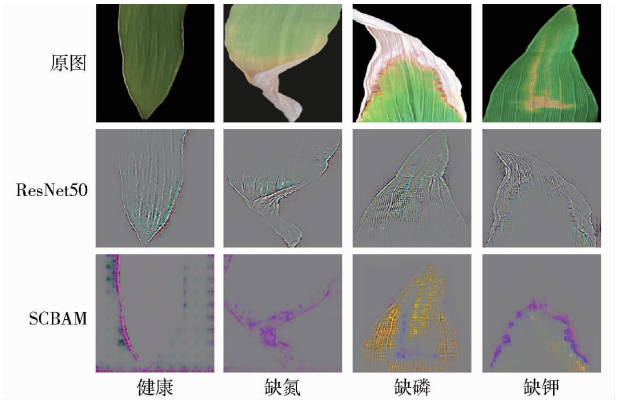


图 9 SCBAM 注意力机制可视化

Fig. 9 Visualization of SCBAM attention maps

引入 BiGRU 会因参数量增加导致模型推理时间延迟,为评估模型在实际应用中部署可行性,对模型推理效率进行了实验。在相同硬件环境下,统计模型对单幅 224 像素 ×224 像素输入图像平均推理时间(Latency)及帧率(Frames per second,FPS)。在推理时间测量过程中,首先进行若干次预热(warm-up)以消除初始运行开销;在 GPU 环境下采用同步机制(torch. cuda. synchronize( ))保证计时准确性;最后通过多次重复试验取平均值作为最终推理时间,以提高结果稳定性与可靠性。

试验结果如表 9 所示。由表 9 可知,未引入 BiGRU 模块时,模型单幅推理时间为 2.56 ms,对应帧率(取整)为 390 f/s;引入 BiGRU 模块后,单幅推理时间增至 9.77 ms,对应帧率为 102 f/s。尽管 BiGRU 模块引入了一定的计算开销,但模型仍能够保持超过 100 f/s 的处理速度,满足实时性要求。结合前文试验结果可知,引入 BiGRU 模块在带来有限推理延迟增加的同时,显著提升了模型识别精度,说明所提出方法在精度与效率之间取得了良好的平衡。

表 9 模型推理效率

Tab. 9 Comparison of model accuracy and inference efficiency

模型	平均推理时间/ms	帧率/(f·s^{-1})
BiGRU - ResNet50	9.77	102
ResNet50	2.56	390

3 结论

(1)提出了基于卷积神经网络、循环神经网络及注意力机制相结合的识别模型 SCBAM - BiGRU - ResNet50。

(2)通过添加风格化卷积注意力机制模块 SCBAM 增强局部特征表达,使模型更加聚焦于缺素状态下典型缺素特征,有效提升了 ResNet50 模型识别能力。通过引入双向门控循环单元 BiGRU 从前、后 2 个方向学习空间上下文信息,利用强大的序列建模能力,捕捉高层特征图感受野之间的空间特征关系;并结合卷积神经网络强大的局部特征提取能力,从局部和全局 2 个视野进行建模,提高模型对缺素特征的识别精度和抗干扰能力。

(3)性能试验结果表明,SCBAM - BiGRU -

ResNet50 在水稻缺素特征识别任务中表现最佳，准确率均显著优于对照组（$p < 0.000\ 1$），同时保持了较小的波动范围，验证了改进结构在特征表达与序列建模方面的有效性和稳定性。

（4）消融试验结果表明，每次删减 SCBAM－BiGRU－ResNet50 模型中某个模块后，识别准确率及其他指标均出现显著下降。当移除所有改进模块，使模型退回到原 ResNet50 结构时，识别准确率下降 4.42 个百分点。所提模型凭借融合注意力机制、双向时序建模和深度残差学习的混合架构，能够更精准地识别水稻缺素症状，为当前水稻缺素诊断提供新的技术方案。

参　考　文　献

[1] Anami B S, Malvade N N, Palaiah S. Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images[J]. Artificial Intelligence in Agriculture, 2020,4:12－20.

[2] Van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: a systematic literature review[J]. Computers and Electronics in Agriculture, 2020,177:105709.

[3] 郭祥云，台海江. 深度学习在大田种植中的应用及展望[J]. 中国农业大学学报，2019,24(1):119－129.
Guo Xiangyun, Tai Haijiang. Current situation and prospect of deep learning application in field planting[J]. Journal of China Agricultural University, 2019,24(1):119－129. (in Chinese)

[4] Dey B, Haque M M U, Khatun R, et al. Comparative performance of four CNN-based deep learning variants in detecting Hispa pest, two fungal diseases, and NPK deficiency symptoms of rice (*Oryza sativa*)[J]. Computers and Electronics in Agriculture, 2022,202:107340.

[5] Islam S, Reza M N, Ahmed S, et al. Nutrient stress symptom detection in cucumber seedlings using segmented regression and a mask region-based convolutional neural network model[J]. Agriculture, 2024,14(8):1390.

[6] Muthusamy S, Ramu S P. IncepV3Dense: deep ensemble based average learning strategy for identification of micro-nutrient deficiency in banana crop[J]. IEEE Access, 2024,12:73779－73792.

[7] Bevers N, Sikora E J, Hardy N B. Soybean disease identification using original field images and transfer learning with convolutional neural networks[J]. Computers and Electronics in Agriculture, 2022,203:107449.

[8] 熊俊涛，戴森鑫，区炯洪，等. 基于深度学习的大豆生长期叶片缺素症状检测方法[J]. 农业机械学报，2020,51(1):195－202.
Xiong Juntao, Dai Senxin, Ou Jionghong, et al. Leaf deficiency symptoms detection method of soybean based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(1):195－202. (in Chinese)

[9] 朱红春，李旭，孟炀，等. 基于 Faster R－CNN 网络的茶叶嫩芽检测[J]. 农业机械学报，2022,53(5):217－224.
Zhu Hongchun, Li Xu, Meng Yang, et al. Tea bud detection based on Faster R－CNN network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022,53(5):217－224. (in Chinese)

[10] 邹强，宋欣，宋中越，等. 基于残差连接的轻量化番茄作物病害检测模型[J]. 农业机械学报，2026,57(1):140－148.
Zou Qiang, Song Xin, Song Zhongyue, et al. Lightweight tomato crop disease detection model based on residual connections[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026,57(1):140－148. (in Chinese)

[11] 谷伟，叶峥，矫桂娥. 基于 DenseNet 优化 Swin Transformer 模型的苹果叶部病害多尺度特征分类[J]. 农业机械学报，2026,57(2):181－192.
Gu Wei, Ye Zheng, Jiao Guie. Swin Transformer model optimized with DenseNet for multi-scale feature classification of apple leaf diseases[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026,57(2):181－192. (in Chinese)

[12] Song Y, Liu L, Rao Y, et al. FA－Net: a fused feature for Multi-head Attention recoding network for pear leaf nutritional deficiency diagnosis with visual RGB-image depth and shallow features[J]. Sensors, 2023,23(9):4507.

[13] Xiao D, Zeng R, Liu Y, et al. Citrus greening disease recognition algorithm based on classification network using TRL－GAN[J]. Computers and Electronics in Agriculture, 2022,200:107206.

[14] Chen P P, Dai J G, Zhang G S, et al. Diagnosis of cotton nitrogen nutrient levels using ensemble MobileNetV2FC, ResNet101FC, and DenseNet121FC[J]. Agriculture-Basel, 2024,14(4):525.

[15] Zeng W, Li M. Crop leaf disease recognition based on Self-Attention convolutional neural network[J]. Computers and Electronics in Agriculture, 2020,172:105341.

[16] Li E, Wang L, Xie Q, et al. A novel deep learning method for maize disease identification based on small sample-size and complex background datasets[J]. Ecological Informatics, 2023,75:102011.

[17] 张林鎧，巴音塔娜，曾庆松. 基于 StyleGAN2－ADA 和改进 YOLO v7 的葡萄叶片早期病害检测方法[J]. 农业机械学报，2024,55(1):241－252.
Zhang Linxuan, Ba Yintana, Zeng Qingsong. Grape disease detection method based on StyleGAN2－ADA and improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(1):241－252. (in Chinese)

[18] Xiao D, Zhu F, Jiang J, et al. Leveraging natural cognitive systems in conjunction with ResNet50－BiGRU model and attention mechanism for enhanced medical image analysis and sports injury prediction[J]. Frontiers in Neuroscience, 2023,17:1273931.

（下转第 257 页）