doi:10.6041/j.issn.1000-1298.2026.23.028

深度卷积与状态空间协同优化的双分支田间杂草分类方法研究

邬开俊[1]　豆佳蓉[1]　白晨帅[1]　徐小容[1]　吴晓强[2]　张红娜[2]

(1.兰州交通大学电子与信息工程学院,兰州 730070;2.内蒙古民族大学物理与电子信息学院,通辽 028000)

摘要:针对同种类杂草间存在显著的类内差异性和类间相似性,以及复杂田间环境干扰导致杂草分类精度受限的问题,本文提出一种基于 Mamba 的多尺度高效杂草分类网络。为提升网络特征提取能力,设计双分支提取模块,整合卷积操作的局部特征捕捉优势与状态空间模型的全局依赖建模能力。为增强模型对多尺度特征的表达能力,构建多尺度过滤融合金字塔网络,通过通道注意力机制实现高层语义特征对低层空间特征的动态加权融合。最后,引入自适应多核分类器,通过自适应权重调整机制有效抑制背景干扰。在 CottonWeedID15 和 DeepWeeds 数据集上,该方法分别取得 96.4% 和 89.9% 的 Top-1 分类准确率。实验结果表明,本文方法在分类性能和模型复杂度之间取得很好的平衡。

关键词:田间杂草分类;深度学习;Mamba;双分支特征提取;多尺度特征融合

中图分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1000-1298(2026)23-0296-11　　**OSID**:

Double-branch Field Weed Classification Method via Collaborative Optimization of Deep Convolution and State Space

Wu Kaijun[1]　Dou Jiarong[1]　Bai Chenshuai[1]　Xu Xiaorong[1]　Wu Xiaoqiang[2]　Zhang Hongna[2]

(1. *School of Electronic and Information Engineering*, *Lanzhou Jiaotong University*, *Lanzhou* 730070, *China*
2. *School of Physics and Electronic Information*, *Inner Mongolia University for Nationalities*, *Tongliao* 028000, *China*)

Abstract: Based on the challenges of significant intra-class variability and inter-class similarity among weed species, as well as the interference from complex field environments that limited weed classification accuracy, a multi-scale efficient weed classification network named WeedMamba was proposed, which was based on the Mamba architecture. To enhance the feature extraction capability, a dual-branch extraction module was designed, integrating the local feature extraction advantages of convolutional operations with the global dependency modeling ability of state space models. To improve the multi-scale feature representation, a multi-scale filtering and fusion pyramid network was constructed, which employed a channel attention mechanism to achieve dynamic weighted fusion of high-level semantic features and low-level spatial features. Finally, an adaptive multi-kernel classifier was introduced to effectively suppress background interference through an adaptive weight adjustment mechanism. The proposed method was evaluated on two public datasets, CottonWeedID15 (15 classes) and DeepWeeds (9 classes), with images captured under diverse lighting, growth stages, and occlusion conditions. Training was conducted on an NVIDIA RTX 4060Ti GPU with a batch size of 64 and an initial learning rate of 0.000 3, using a cosine decay schedule and early stopping. On these datasets, the method achieved Top-1 accuracies of 96.4% and 89.9%, respectively, along with F1 scores of 96.3% and 89.2%. Ablation studies further verified the individual contributions of each core component. Experimental results demonstrated that the proposed method achieved an excellent balance between classification performance and model complexity.

Key words: field weed classification; deep learning; Mamba; dual-branch feature extraction; multi-scale feature fusion

收稿日期:2025-08-27　修回日期:2025-10-28

基金项目:甘肃省自然科学基金项目(23JRRA913)、甘肃省重点研发项目(25YFGA047)、内蒙古自治区重点研发与成果转化计划项目(2023YFDZ0043、2023YFDZ0054、2023YFSH0043)和兰州交通大学重点研发项目(ZDYF2304)

作者简介:邬开俊(1978—),男,教授,博士,主要从事深度学习和计算机视觉研究,E-mail:wkj@mail.lzjtu.cn

0　引言

田间杂草图像分类的目的是通过快速、准确识别图像中占主导地位的杂草物种，为田块尺度的杂草群落分析、大面积除草剂精准选型及生态化植被管理提供决策支撑[1]。早期卷积神经网络(Convolutional neural network，CNN)从时延神经网络[2]、LeNet－5[3]奠定基础，经 AlexNet[4]验证深层网络有效性，再到 VGG[5]深化网络层次、ResNet[6]解决梯度消失问题，逐步形成层次化特征提取架构，可逐层捕捉杂草叶片纹理、边缘等低级特征与物种形态等高级语义特征[7]；MobileNet[8]、ShuffleNet[9]通过深度可分离卷积[10]和通道重排[11]等技术，显著提升模型的计算效率[12]。但局部感受野特性难以建模杂草群落的长程空间依赖，对形态相似物种的分类精度受限。

2020 年 Vision Transformer(ViT)[13]突破 CNN 局部限制，以图像分块与位置编码实现全局依赖建模，其注意力机制可捕捉杂草整体分布模式，但计算复杂度随分块数量平方增长，高分辨率下推理速度骤降，且对局部精细特征的建模效率低[14－16]。近年结构化状态空间模型(Structured state space model，SSM)[17]成为新方向：经典 SSM[18]因计算复杂长期受限，经 HiPPO 初始化[19]、S4 模型[20]优化后逐步实用化，Mamba[21]、VMamba[22]进一步提升效率，但现有工作如 Vim[23]仅将 SSM 作为补充，未充分协同 CNN 与 SSM 优势，且模型吞吐量低，难以平衡精度、复杂度和速度，制约田间边缘设备部署。针对复杂田间场景下杂草分类面临的局部特征捕捉不细、全局依赖建模不足、多尺度适配性差、类别不平衡、鲁棒性低的问题，本文设计兼顾分类精度、轻量化与实时性的 WeedMamba 模型。

1　WeedMamba 模型

WeedMamba 将 SSM 架构引入农业视觉领域，通过搭建 CNNSSM 混合编码器实现局部细节与全局上下文的高效协同。该结构采用通道分割机制进行并行计算：卷积分支集成 ECA 注意力模块，以自适应感受野强化局部纹理特征；SSM 分支则基于离散化状态方程建模长程依赖关系。为增强多尺度表达能力，设计了层级化多尺度特征融合金字塔模块，利用深度可分离卷积与 ECA 注意力协同实现动态权重调整，通过一维卷积建模通道关系，避免降维损失，提升对不同生长阶段杂草形态的适应能力。针对类别不平衡问题，引入基于动态路由机制的自适应多核分类器，包含 K 个并行分类核，经两层 MLP 生成路由分数，并采用熵正则化约束的 Softmax 注意力实现核间动态权重分配，从而有效增强模型对少数类样本的识别能力与整体分类鲁棒性。

1.1　WeedMamba 模型构建

本文提出的 WeedMamba 模型构建一种层次化特征表示架构，通过多阶段特征提取与融合策略，有效应对 DeepWeeds 数据集中杂草形态多样性和尺度变化带来的分类挑战。模型采用 4×4 非重叠图像块嵌入作为初始特征表示，该设计将输入图像划分为 16 像素×16 像素区域(每个 4 像素×4 像素块对应一个特征向量)，在保留原始空间结构的同时，实现 16 倍下采样率的空间维度压缩。通过分层堆叠 CSM(Conv－SSM Block)模块，模型逐步实现特征的深层提取与精炼。每个 CSM 模块采用双分支架构：CNN 分支利用深度可分离卷积并结合高效通道注意力(ECA)机制[24]，专注于捕捉叶片纹理、边缘细节等局部特征；SSM 分支则基于离散化状态空间方程，借助递归计算建模植株间的长程依赖关系，从而有效捕获整体形态特征。两个分支的输出经过通道拼接与混洗操作，在保持二维空间结构完整性的同时，通过跨尺度信息的通道融合实现了多层级特征的充分整合。

在多尺度特征构建方面，模型通过块合并层逐步下采样形成 4 个分辨率递减的特征阶段。MSFPN 模块采用层级化融合策略，通过转置卷积上采样恢复高层特征的空间分辨率，同时利用通道注意力机制过滤低层特征中缺乏判别性的细节，最终构建出兼具语义丰富性和空间精确性的特征金字塔。该过程采用双线性插值法实现亚像素级特征图对齐，并通过深度可分离卷积显著提升计算效率。

在分类决策阶段，模型设计自适应多核动态聚合机制。全局平均池化后的特征向量经由路由网络动态分配各分类器核的权重分布，通过熵正则化约束确保所有分类器核都参与预测过程。这种创新的分类策略使模型能够根据输入特征依据路由分数动态加权融合多分类器输出，提升模型对形态相似杂草物种的鉴别能力。整个网络架构凭借局部与全局特征协同提取的创新交互模式、多尺度特征动态融合的精准适配机制以及分类决策自适应优化的智能聚合策略，最终实现对复杂杂草图像的分类。

1.2　CSM 模块

在特征提取任务中，需要兼顾局部细节与全局依赖关系的捕捉。传统单一模块往往难以在这两方面都做到出色。为解决这一问题，本文提出 CSM(Conv－SSM Block)模块，将 CNN 的局部特征提取

优势与 SSM 的长序列建模能力相结合。该模块采用分层维度策略，通过最小状态维度阈值机制动态调整特征表达能力：在浅层网络保留 32 维特征通道以捕捉叶片纹理等细粒度细节，在深层网络扩展至 64 维以增强高级语义特征的表示能力，这种自适应的状态投影器通过 1 ×1 卷积和批归一化实现无信息损失的维度变换，当输入维度满足要求时跳过调整层。

如图 1 所示，CSM 模块首先通过通道分割（Split，SP）将输入特征沿通道维度划分为两组，分别由 CNN 分支和 SSM 分支构建独立且互补的特征处理路径：前者采用深度可分离卷积架构捕获局部细节，并引入 ECA 注意力机制强化关键通道特征。后者通过 SS2D 模块建模长程依赖，其状态空间维度通过 d_state 调控（本文设置为 16，与 Mamba 及 VMamba 等主流工作保持一致，以在建模能力和计算负担间取得平衡），使用可配置的时间步秩优化状态转移矩阵参数效率。两个分支的输出通过通道连接恢复原始维度，采用通道洗牌（Shuffle，SF）操作可打乱通道顺序，促进跨分支信息交互，有效避免分组卷积导致的通道间信息闭塞问题。本文遵循经典 CNN 和 ViT 的设置，将 CNN - Branch 和 SSM - Branch 中的激活函数分别设置为 GELU[25] 和 SiLU[26]，其中 SiLU 函数的平滑特性更适合状态空间模型的连续信号建模，CNN 分支选用 GELU 激活函数。通过这种设计，模型在浅层优先保留局部细节，在深层侧重全局关系建模，最终形成层次化的特征学习机制。

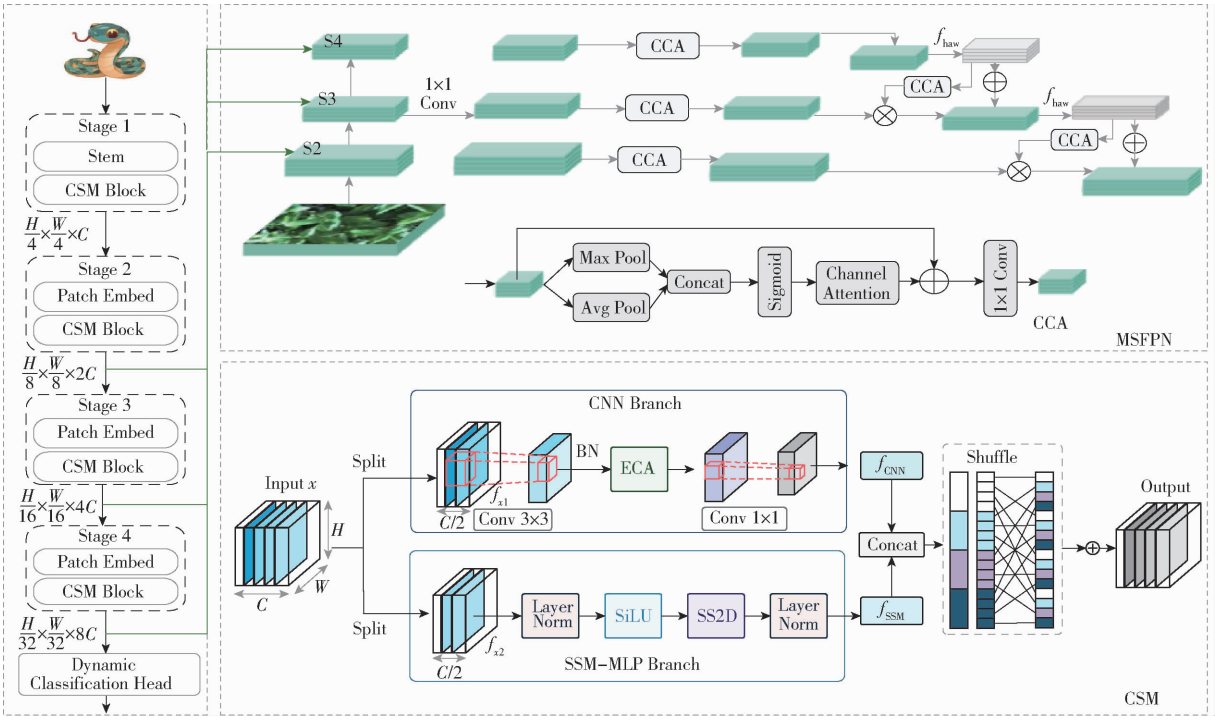

图 1 WeedMamba 网络框架图

Fig. 1 WeedMamba network framework diagram

1.2.1 局部特征提取

在 DeepWeeds 杂草数据集的图像分类任务中，由于数据集中杂草种类多样，图像特征复杂，若采用未经优化的模型，可能会出现训练困难和特征提取不准确的情况。借鉴在图像领域的相关经验，CNN 分支采用深度可分离卷积来提取局部特征。深度可分离卷积把标准卷积拆分为深度卷积和逐点卷积，能在显著降低计算量的同时，有效捕捉杂草图像中的局部细微特征，例如不同杂草叶片的独特纹理、锯齿状边缘等细节。

CSM 模块针对上述问题进行针对性设计。该模块首先通过通道分割将输入的杂草图像特征划分为两个子特征 x_1、$x_2 \in \mathbf{R}^{H\times W\times (C/2)}$。在 CNN 分支中，采用深度可分离卷积进行局部特征提取，利用空间可分离卷积核实现高效计算，并对输出特征进行批归一化处理以稳定训练过程并加速收敛。在此基础上，引入高效通道注意力（ECA）机制，通过一维卷积（卷积核大小 k 根据通道数 C 自适应确定）直接建模通道间依赖关系，显著增强特征的通道感知能力。该处理流程在保证计算效率的同时，有效提升局部特征的判别性表征能力。ECA 注意力机制作为一种轻量级的通道注意力模块，能够自适应地调整各通道的权重，使模型更加聚焦于对杂草分类关键的特征通道。最后，通过卷积层调整通道数量以实现特征的融合与压缩，并采用 GELU 激活函数进一步增强非线性表达能力，使所提取的局部特征更

具区分性，从而为后续杂草种类的准确分类提供有效支持。其过程为

$$x_1' = \mathrm{DWConv}_{3\times3}(x_1) \tag{1}$$

$$x_1'' = \mathrm{BN}(x_1') \tag{2}$$

$$x_1''' = \mathrm{ECA}(x_1'') \tag{3}$$

$$x_{c1} = \mathrm{GELU}(\mathrm{Conv}_{1\times1}(x_1''')) \tag{4}$$

式中　$\mathrm{DWConv}_{3\times3}$——核大小为 3×3 的深度可分离卷积

BN——批归一化操作

ECA——高效通道注意力模块

$\mathrm{Conv}_{1\times1}$——1×1 卷积，用于通道数调整

GELU——高斯误差线性单元激活函数

x_1'——经深度可分离卷积（卷积核大小 3×3）提取的局部空间特征

x_1''——批归一化后的特征

x_1'''——经高效通道注意力（ECA）加权增强的特征

x_{c1}——CNN 分支最终输出的局部增强特征

1.2.2　全局特征提取

在处理大规模、复杂的图像时，CNN 难以有效捕捉全局结构信息或长程依赖关系。而 SSM 分支采用的状态空间模型，尤其是其中的 SS2D，与 CNN 分支形成特征提取能力的互补。近期基于 SSM 的模型，如 Structured state space sequence models（S4）和 Mamba，均依赖于一个经典的连续系统，该系统通过中间隐式状态 $h(t)\in\mathbf{R}^N$ 将一维输入函数或序列 $x(t)\in\mathbf{R}$ 映射到输出 $y(t)\in\mathbf{R}$。这一过程通常可以表示为一个线性常微分方程

$$\begin{cases} h'(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t) \\ y(t) = \boldsymbol{O}h(t) \end{cases} \tag{5}$$

其中，$\boldsymbol{A}\in\mathbf{R}^{N\times N}$表示状态矩阵，$\boldsymbol{B}\in\mathbf{R}^{N\times1}$、$\boldsymbol{O}\in\mathbf{R}^{1\times N}$分别为输入到状态和状态到输出的投影参数。$h(t)$为系统在时刻 t 的记忆或内部表征，$h'(t)$为系统更新记忆的速度和方向。由于传统 SSM 仅适用于连续数据，而深度学习任务常同时包含连续与离散输入，因此需将连续 SSM 转换为离散形式以适应深度学习场景。具体而言，引入时间尺度参数 Δ，通过固定的离散化规则，如零阶保持技术将矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 转换为离散参数 $\overline{\boldsymbol{A}}$ 和 $\overline{\boldsymbol{B}}$。其离散化规则定义为

$$\begin{cases} \overline{\boldsymbol{A}} = e^{\Delta\boldsymbol{A}} \\ \overline{\boldsymbol{B}} = (\Delta\boldsymbol{A})^{-1}(e^{\Delta\boldsymbol{A}} - I)\Delta\boldsymbol{B} \end{cases} \tag{6}$$

式中：$e^{\Delta\boldsymbol{A}}$表示矩阵指数运算，用于计算状态转移的指数衰减/增长；$\boldsymbol{I}$ 表示单位矩阵；$\Delta\boldsymbol{A}^{-1}$表示矩阵 $\Delta\boldsymbol{A}$ 的逆矩阵，该逆矩阵存在的条件是 $\Delta\boldsymbol{A}$ 为非奇异矩阵（即行列式不为零）。离散化后，该系统在步长为 Δ 时可递归计算为

$$\begin{cases} \boldsymbol{h}_t = \overline{\boldsymbol{A}}\boldsymbol{h}_{t-1} + \overline{\boldsymbol{B}}x_t \\ y_t = \boldsymbol{O}\boldsymbol{h}_t \end{cases} \tag{7}$$

式中　t——离散时间步索引（对应于连续域中的采样时刻）

$\boldsymbol{h}_t$——第 t 步的隐状态向量

$\boldsymbol{h}_{t-1}$——上一步的隐状态

x_t——第 t 步的输入标量

y_t——第 t 步的输出标量

$\boldsymbol{O}$——输出投影矩阵，负责将隐状态映射为输出

该递推式消除了连续域中的导数运算，以代数迭代方式逐步更新状态，使 SSM 可直接适用于离散序列数据。最终，基于 SSM 的模型可通过全局卷积计算输出为

$$y = x\overline{K} \quad (\overline{K} = (\boldsymbol{O}\boldsymbol{B}, \boldsymbol{O}\,\overline{\boldsymbol{A}\boldsymbol{B}}, \cdots, \boldsymbol{O}\,\overline{\boldsymbol{A}}^{L-1}\boldsymbol{B})) \tag{8}$$

式中　y——模型输出的全局卷积结果

$\overline{K}$——由状态空间模型导出的结构化卷积核，$\overline{K}\in\mathbf{R}^L$

L——输入序列 x 的长度

SSM 分支采用状态空间模型 SS2D 建模长程依赖关系，并辅以批归一化（BN）以稳定训练过程。SS2D 通过离散状态空间方程递归计算全局特征，能够有效捕捉图像中元素之间的长距离关联。该分支特别擅长提取全局上下文信息，从而弥补 CNN 分支在处理长程依赖方面的不足。在特征融合阶段，SSM 分支与 CNN 分支的输出经通道拼接恢复原始通道维度，并通过通道混洗操作促进跨组信息流动，以避免分组卷积可能引发的通道间信息丢失，其融合公式为

$$y = \mathrm{Channel-Shuffle}(\mathrm{Concat}(x_{c1}, x_2'')) \tag{9}$$

式中，x_2''表示 SSM 状态空间分支经 SS2D 模块建模长程依赖后输出的全局上下文特征（即 SSM 分支的最终输出）；若输入与输出的空间尺寸不一致，采用双线性插值对齐空间分辨率，以确保后续处理的准确性与有效性。

1.3　多尺度过滤融合金字塔

针对 DeepWeeds 数据集中杂草分类任务面临的形态多样性和尺度变化挑战，本文提出一种多尺度过滤融合金字塔结构（Multi-scale filtering pyramid network，MSFPN）模块。该模块通过特征过滤和层级融合的双阶段处理机制，实现对不同尺度特征的自适应优化。

在特征过滤阶段，MSFPN 采用 CCA 模块进行通道特征选择。对于输入特征图 $f_{\mathrm{in}}\in\mathbf{R}^{C\times H\times W}$，其通

道注意力权重计算式为

$$f_{\mathrm{ECA}}(c)=\sigma\left(\sum_{i=-\left\lfloor\frac{k-1}{2}\right\rfloor}^{\left\lfloor\frac{k-1}{2}\right\rfloor}w_i\ f_{\mathrm{in}}(c+i)\right)\tag{10}$$

其中，σ 表示 Sigmoid 激活函数，w_i 表示一维卷积中第 i 个位置的卷积核权重。$f_{\mathrm{in}}(c+i)$ 表示输入通道描述向量中第 $(c+i)$ 个通道对应的响应值。本模块舍弃传统通道注意力机制中的降维操作，在保持较高计算效率的同时有效避免信息损失。过滤后的特征进一步通过深度可分离卷积进行尺度适配，显著降低计算复杂度。

在特征融合阶段，MSFPN 采用自顶向下的层级融合策略。网络第 2、3、4 阶段生成的多尺度特征图具有不同特性：高层特征语义信息丰富但空间细节粗糙，低层特征则保留精细的空间信息却缺乏语义表达能力。本网络 4 个特征阶段（Stage 1 ~ 4）的输出通道数分别设置为：$C_1=64$，$C_2=128$，$C_3=256$，$C_4=512$。该设计遵循常见的特征金字塔缩放原则，即随着空间分辨率的降低，通道数逐层翻倍，以容纳更高级的语义信息。MSFPN 的中间特征层通道数与此对齐，以确保融合时维度匹配。传统融合方法（如直接像素相加）难以充分挖掘多尺度特征间的互补性。而 MSFPN 设计一种层级化融合机制，以高层语义特征为引导，对低层特征进行过滤与增强。以第 3 阶段特征 $f_{\mathrm{high}}\in\mathbf{R}^{C\times H\times W}$ 和第 2 阶段特征 $f_{\mathrm{low}}\in\mathbf{R}^{C\times H_1\times W_1}$ 融合为例，首先使用步长为 2、卷积核为 3×3 的转置卷积对高层特征进行上采样，使其空间尺寸接近低层特征；再通过双线性插值进一步统一维度，得到 $f_{\mathrm{haw}}\in\mathbf{R}^{C\times H_1\times W_1}$；随后利用 ECA 模块计算通道权重，对低层特征进行过滤，突出与高层语义相关的细节信息；最后将过滤后的低层特征与上采样后的高层特征相加，得到融合后的特征，其融合过程为

$$f_{\mathrm{haw}}=\mathrm{BL}(T-\mathrm{Conv}(f_{\mathrm{high}}))\tag{11}$$

式中，BL 表示双线性插值上采样操作，用于将特征图的空间分辨率调整至与低层特征图相匹配；T 表示目标空间尺寸参数。类似的，第 4 阶段特征也按照此机制与第 3 阶段融合结果进行交互，最终生成包含融合后的特征图，同时保持高层语义信息和低层空间细节。

1.4 自适应多核分类器

针对传统图像分类方法中静态分类头结构难以适应输入特征动态变化、导致模型表达能力受限的问题，本文提出一种自适应多核分类器（Adaptive multi-kernel classifier），通过多核动态聚合机制和熵正则化约束实现分类决策边界的自适应调整。在多核动态聚合机制方面，定义 K 个并行的分类器核 $\{W_k,b_k\}_{k=1}^{K}$，每个分类器核独立计算输入特征 x 的分类结果（本文根据实验验证，设定 $K=4$。过多的核会导致参数量激增而收益递减，过少的核则无法体现动态聚合的优势）。通过动态路由网络生成注意力权重 $\pi_k(x)$，最终的输出是各分类器核的加权组合，即

$$y=\sum_{k=1}^{K}\pi_k(x)(\widetilde{\boldsymbol{W}}_k^{\mathrm{T}}x+\widetilde{\boldsymbol{b}}_k)\tag{12}$$

式中，$\widetilde{\boldsymbol{W}}_k$、$\widetilde{\boldsymbol{b}}_k$ 分别表示第 k 个分类器核的权重矩阵和偏置项；$\widetilde{\boldsymbol{W}}_k^{\mathrm{T}}x+\widetilde{\boldsymbol{b}}_k$ 表示第 k 个线性分类器核的原始决策分数；$\sum_{k=1}^{K}$ 表示对 k 个分类器核的输出进行加权求和，该机制使得模型能够学习多个不同的决策边界，并借助动态权重 $\pi_k(x)$ 自适应地选择最合适的分类策略，使模型参数量增加 K 倍（K 为分类器核数量）。而熵正则化约束则是为优化动态路由权重 $\pi_k(x)$。动态路由权重 $\pi_k(x)$ 由标准 Softmax 计算得出，即

$$\pi_k(x)=\frac{\exp(z_k)}{\sum_{j=1}^{K}\exp(z_j)}\tag{13}$$

式中，z_k 为第 k 个分量，$\sum_{j=1}^{K}\exp(z_j)$ 表示对所有 k 个路由分数取指数后的累加和，作为归一化分母。为防止在训练过程中某些分类器核被完全抑制，引入熵正则化损失以促进权重分布的合理性与多样性，其表达式为

$$L_{\mathrm{entropy}}=-\lambda\sum_{k=1}^{K}\pi_k(x)\ln\pi_k(x)\tag{14}$$

式中，L_{entropy} 表示熵正则化损失项，用于约束动态路由权重的分布特性；λ 表示熵正则化系数，用于调节该损失项在总损失函数中的权重占比。在特征处理流程中，输入特征图 x 经全局平均池化处理，被压缩为向量 $\boldsymbol{z}$，从而保留通道级信息。在之后的动态路由环节，压缩后的特征 $\boldsymbol{z}$ 会通过两层 MLP 计算路由分数 $a\in\mathbf{R}^K$，即 $a=\mathrm{MLP}(\boldsymbol{z})$。在注意力计算步骤，对路由分数 a 应用 Softmax 函数，得到各分类器核的权重 π，即 $\pi=\mathrm{Softmax}(a)$。最后在多核聚合输出阶段，将各分类器核的输出进行加权求和，得到最终的分类结果为

$$y=\sum_{k=1}^{K}\pi_k(\widetilde{\boldsymbol{W}}_k^{\mathrm{T}}\boldsymbol{z}+\widetilde{\boldsymbol{b}}_k)\tag{15}$$

2 实验结果与分析

为验证所提出方法的有效性，本研究设计对比

实验、消融实验及可视化分析。对比实验将 WeedMamba 与当前主流的图像分类方法进行比较，以验证其性能优势；消融实验通过逐步引入关键组件，定量分析各组件对最终性能的贡献。可视化分析则直观展示模型对背景干扰的抑制效果及对关键特征的定位能力，从而全面验证算法的有效性。

2.1　实验设置

本文所有模型的训练和测试过程在 Microsoft Windows 11 操作系统下进行，CPU 型号为 13th Gen Intel（R）Core（TM）i5 - 13400F，GPU 型号为 NVIDIA GeForce RTX 4060Ti，GPU 可用显存为 16 GB。学习率按照“余弦衰减”策略调整，初始学习率设为 0.000 3，Batch Size 设为 64，采用 AutoAugment 进行数据增强，优化器为 AdamW，训练轮次为 300，训练过程中采用早停策略以防止过拟合。

2.2　数据集

DeepWeeds 是一个公开的杂草识别与分类图像数据集，由澳大利亚昆士兰大学研究团队构建。该数据集共包含 17 509 幅高分辨率图像，涵盖 8 类澳大利亚常见杂草及 1 类背景植被。图像采集覆盖不同物候期、光照条件和拍摄视角，具有较高的场景多样性。由于杂草与作物间存在较高的视觉相似性，且杂草形态随生长阶段变化显著，该数据集的分类任务具有较强的挑战性，数据及图像如图 2 所示。



(a) DeepWeeds数据集图像　　(b) CottonWeedID15数据集图像

图 2　DeepWeeds 和 CottonWeedID15 数据集图像

Fig. 2　DeepWeeds dataset and CottonWeedID15 dataset images

CottonWeedID15 是由美国奥本大学构建的棉花田间杂草识别数据集，包括 5 187 幅高质量图像，涵盖 15 种常见棉花杂草类别，每幅图像均经过精确标注。图像采集于不同光照条件、生长阶段及遮挡环境下，进一步增加了识别任务的复杂性。为满足算法训练需求并保证模型泛化能力，以上两个数据集均按比例 8∶2随机划分为训练集和测试集。

本研究所使用的两个数据集，其标注方式为每幅图像对应一个主要的杂草类别标签。因此，任务定义为基于主导物种的图像分类。

2.3　评价指标

在评估杂草数据集图像分类模型时，采用参数量（Params）、浮点运算次数、吞吐量（Throughput）、Top - 1 准确率和 F1 分数指标进行多维评价，分别从模型复杂度、计算开销、推理速度和分类性能等不同角度综合评价模型的表现。

其中，Params 作为模型中可训练参数的总量，直接反映模型的容量和结构复杂度；浮点运算次数通过量化前向推理过程中所需的浮点运算次数，客观衡量计算资源的消耗程度；Throughput 表示单位时间内模型能够处理的图像数量，是评估实际部署效率的关键指标；而 Top - 1 准确率作为分类任务的核心评价指标，其高低直接决定了模型在杂草类别判别上的准确性。F1 分数综合精确率和召回率，能在类别分布不均衡的杂草数据集中更全面地衡量模型对少数类杂草的识别能力。

2.4　对比实验

在 DeepWeeds 数据集上的实验结果表明，WeedMamba 模型仅使用 1.93×10^7 个参数和 2.5×10^9 的浮点运算次数，实现 89.9% 的 Top - 1 分类准确率（表 1）。MobileOne_S4[27] 采用深度可分离卷积优化计算效率，其参数量与本文模型相近，但准确率低 0.3 个百分点；FastViT - SA24[28] 通过空间注意力机制增强特征表达能力，准确率差距进一步扩大至 0.7 个百分点。参数量更大的 LocalVmamba - T 和 InceptionNext - T[29] 分别达到 2.60×10^7 和 2.80×10^7，其准确率仍低于本文方法。在推理效率方面，WeedMamba 每秒可处理 503.5 幅图像，其实时处理能力显著优于 FastViT - SA24 和 Caformer - s18[30] 等对比模型。

表 1 在 DeepWeeds 和 CottonWeedID15 数据集上与先进方法的对比结果

Tab. 1 Comparison with advanced methods on DeepWeeds and CottonWeedID15 datasets

数据集	模型	年份	参数量	浮点运算次数	吞吐量/(幅·s^{-1})	准确率/%	F1 分数/%
DeepWeeds	FastViT - SA24	2023	2.06×10^7	3.8×10^9	453.1	89.2	88.5
	MobileOne_S4	2023	2.03×10^7	3.3×10^9	829.0	89.6	89.0
	Caformer - s18	2023	2.45×10^7	4.1×10^9	512.6	87.2	86.5
	Agent - PVT - S	2024	2.06×10^7	4.0×10^9	681.2	87.6	86.9
	ViM - S	2024	2.60×10^7	5.1×10^9		88.4	87.7
	InceptionNext - T	2024	2.80×10^7	4.2×10^9	265.2	88.9	88.2
	LocalVmamba - T	2024	2.60×10^7	5.7×10^9		89.1	88.4
	VMamba - T	2025	2.24×10^7	3.8×10^9	420.0	89.0	88.3
	MambaVision - S	2025	2.51×10^7	4.2×10^9	380.0	88.7	87.9
	WeedMamba		1.93×10^7	2.5×10^9	503.5	89.9	89.2
CottonWeedID15	FastViT - SA24	2023	2.06×10^7	3.8×10^9	521.3	95.1	94.9
	MobileOne_S4	2023	2.03×10^7	3.3×10^9	678.6	96.2	96.0
	Caformer - s18	2023	2.45×10^7	4.1×10^9	580.4	95.1	94.8
	Agent - PVT - S	2024	2.06×10^7	4.0×10^9	588.3	94.9	94.7
	ViM - S	2024	2.60×10^7	5.1×10^9		95.6	95.4
	InceptionNext - T	2024	2.80×10^7	4.2×10^9	234.1	96.0	95.8
	LocalVmamba - T	2024	2.60×10^7	5.7×10^9		96.1	95.9
	VMamba - T	2025	2.24×10^7	3.8×10^9	516.0	94.2	93.9
	MambaVision - S	2025	2.51×10^7	4.2×10^9	497.0	95.8	95.5
	WeedMamba		1.93×10^7	2.5×10^9	576.2	96.4	96.3

在 CottonWeedID15 数据集上,该模型的处理速度达到 576.2 幅/s,远高于传统 CNN 架构的 InceptionNext - T,也明显优于轻量级 Transformer 模型 Agent - PVT - S[31]。在仅使用 VIM - S[32] 四分之三参数的情况下,WeedMamba 实现更高的分类准确率,超出 0.8 个百分点。与采用复杂特征金字塔结构的 Caformer - s18 相比,WeedMamba 在参数量减少 21.2% 的同时,准确率提升 1.3 个百分点。与 VMamba - T[33] 相比,在参数量减少 13.8% 的同时,准确率提升 2.2 个百分点;与 MambaVision - S[34] 相比,参数量减少 23.1%,准确率提升 0.6 个百分点。

综合上述对比可看出,WeedMamba 在保持轻量化的同时,其分类准确率优于许多参数量更大的模型,而推理效率又显著高于同类轻量模型。该优势源于其独特的架构设计:一方面借助 Mamba 模块实现有效的全局特征建模,另一方面通过精简结构减少冗余计算,从而在模型复杂度、推理速度与分类精度三者间实现了更优的平衡。

2.5 消融实验

在 DeepWeeds 数据集上,SSM 模块作为基础架构单独使用时即展现出较强的特征提取能力,获得 88.1% 的 Top - 1 准确率。引入 Conv 模块后,模型性能出现显著提升,准确率提高 0.7 个百分点至 88.8%(表 2),验证卷积操作在局部特征提取方面的有效性。然而,SP 模块的加入使模型参数量和计算量急剧减少,而准确率仅出现轻微波动。该模块的主要贡献在于大幅降低了模型的复杂度和计算开销。但会引起一定的性能波动,其根本原因在于通道分割削弱了跨通道的全局协同作用,在分支处理阶段,两组特征被通道分割操作隔离,缺乏跨组信息交互,直至通道拼接和混洗后才得以融合。这种阶段性隔离削弱了跨通道的全局协同作用,导致模型在提取复杂语义或细粒度判别特征时受限,最终导致整体表现下降。MSFPN 模块的整合使模型性能得到明显改善,准确率回升至 89.2%,较基础模型提升 1.1 个百分点,充分验证多尺度特征融合的有效性。最终加入 DCH 模块后,模型在仅增加 1.1×10^6 参数量的情况下,准确率进一步提升 0.7 个百分点至 89.9%,体现出自适应特征聚合机制的优越性。在 CottonWeedID15 数据集上,各组件表现出相似的影响趋势,但整体准确率更高,这与该数据集相对简单的分类任务特性相符。值得注意的是,DCH 模块在该数据集上的性能提升更为显著,准确率增幅达到 0.5 个百分点。

从模型复杂度角度分析,基础 SSM 架构在 DeepWeeds 数据集上仅需参数 1.32×10^7 和 1.8×10^9 浮点运算次数。随着各模块的逐步加入,模型复杂度呈现阶梯式增长:Conv 模块使参数量增加 7.6×10^6,计算量增加 8×10^8 浮点运算次数;

表 2　在 DeepWeeds 和 CottonWeedID15 数据集上的消融实验结果

Tab. 2　Ablation results on DeepWeeds and CottonWeedID15 datasets

数据集	模块						评价指标				
	SSM	Conv	SP	SF	MSFPN	DCH	参数量	浮点运算次数	吞吐量/(幅·s^{-1})	准确率/%	F1 分数/%
DeepWeeds	√						1.32×10^{7}	1.8×10^{9}	476.9	88.1	87.4
	√	√					2.08×10^{7}	2.6×10^{9}	364.7	88.8	88.1
	√	√	√				1.79×10^{7}	2.2×10^{9}	526.7	87.3	86.5
	√	√	√	√			1.79×10^{7}	2.2×10^{9}	499.9	88.9	88.2
	√	√	√	√	√		1.82×10^{7}	2.6×10^{9}	498.2	89.2	88.5
	√	√	√	√	√	√	1.93×10^{7}	2.5×10^{9}	503.5	89.9	89.2
CottonWeedID15	√						1.32×10^{7}	1.8×10^{9}	399.3	95.7	96.4
	√	√					2.08×10^{7}	2.6×10^{9}	348.7	96.2	94.7
	√	√	√				1.79×10^{7}	2.2×10^{9}	462.8	94.9	95.2
	√	√	√	√			1.79×10^{7}	2.2×10^{9}	459.5	95.4	95.7
	√	√	√	√	√		1.82×10^{7}	2.6×10^{9}	460.7	95.9	96.3
	√	√	√	√	√	√	1.93×10^{7}	2.5×10^{9}	476.2	96.4	96.3

注：“√”表示包含对应结构或模块。

MSFPN 模块带来 3×10^{5} 参数量和 4×10^{8} 浮点运算次数的增长；DCH 模块最终增加参数 1.1×10^{6} 和减少 1×10^{8} 浮点运算次数，这种模块化的复杂度增长方式为模型优化提供灵活的调整空间。实验结果表明，WeedMamba 通过精心设计的模块，在模型性能和计算效率之间能实现良好的平衡。特别是 MSFPN 和 DCH 模块的协同作用，在有限增加模型复杂度的前提下，显著提升分类准确率。

2.6　超参数搜索与分析

为探究超参数对模型性能的影响，本文采用网格搜索，以 Top－1 准确率和 F1 分数为核心指标，在 DeepWeeds 和 CottonWeedID15 数据集开展超参数搜索与敏感性分析（见表 3）。网络结构上，状态空间维度取 16，既满足 SSM 分支对杂草群落长程空间关联的建模，又避免参数冗余过拟合，与 Mamba 等主流模型思路一致；CSM 模块通道分割比例 1∶1，让 CNN 局部纹理提取与 SSM 全局依赖建模的特征量均衡互补；MSFPN 模块 3×3 卷积因感受野适中，能精准捕获杂草细粒度特征并减少背景噪声，对相似类别区分更准。训练超参数中，初始学习率 0.000 3

表 3　在 DeepWeeds 和 CottonWeedID15 数据集上的超参数分析

Tab. 3　Hyperparameter analysis on DeepWeeds and CottonWeedID15 datasets

超参数类别	超参数名称	搜索范围	最优取值	DeepWeeds		CottonWeedID15	
				准确率/%	F1 分数/%	准确率/%	F1 分数/%
网络结构超参数	状态空间维度	{8,16,32,64}	16	89.2	88.5	95.9	96.3
	CSM 通道分割	{1∶1, 1∶2, 2∶1}	1∶1	89.9	89.2	96.4	96.3
	MSFPN 卷积核	{3×3, 5×5, 7×7}	3×3	89.2	88.5	95.9	96.3
训练超参数	初始学习率	{0.000 1,0.000 3,0.000 5}	0.000 3	89.9	89.2	96.4	96.3
	批大小	{32,64,128}	64	89.9	89.2	96.4	96.3
分类器超参数	分类器核数量 K	{2,4,6,8}	4	89.9	89.2	96.4	96.3
	熵正则化系数	{0.01,0.05,0.1}	0.05	89.9	89.2	96.4	96.3

平衡了收敛速度与稳定性，批大小 64 则兼顾权重更新稳定性与显存占用。自适应多核分类器核数 4 可提供合适动态决策边界，避免核数过多引发的参数冗余与泛化下降；熵正则化系数 0.05 能调节核权重分布，聚焦关键特征，兼顾多类性能。

2.7　可视化分析

为直观验证本文算法在杂草图像分类任务中的有效性，本研究基于 Grad－CAM 技术开展可视化分析。实验选取 DeepWeeds（图 3 第 1 行）与 CottonWeedID15（图 3 第 2 行）数据集的验证集样本，通过逐步引入各组件，采用类激活映射方法对模型最后一层卷积层的特征响应进行可视化，以热力图形式揭示模型在分类判别过程中的注意力分布，为理解其决策机制提供直观依据。

图 3 所示为 DeepWeeds 数据集中杂草图像的注意力热力图结果。热力图中，红色与黄色区域表示模型对该区域高度关注，这些区域对最终分类决策贡献显著；蓝色与绿色区域则表示关注度较低，贡

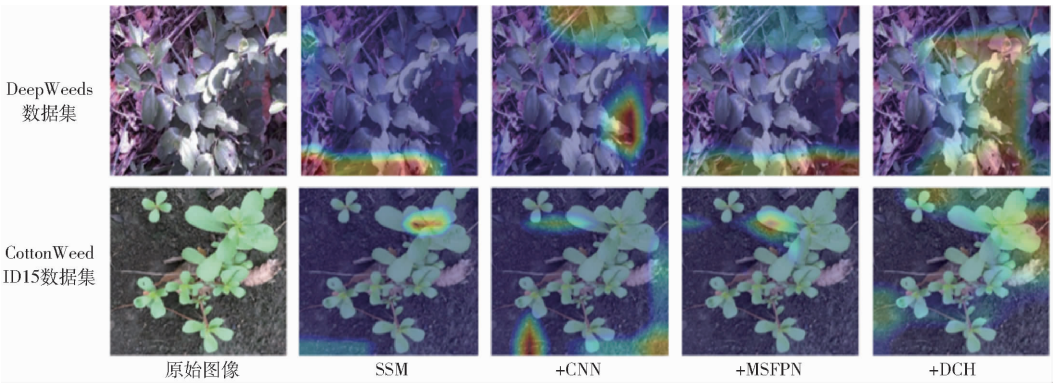

图 3 各部分组件在两个数据集上的可视化结果

Fig. 3 Visualization results of each component on two datasets

献相对有限。可视化结果表明,对于被标注为特定主导类别的图像,WeedMamba 能够有效抑制背景干扰,并将注意力集中在对判断该主导类别最具鉴别性的特征区域(如特定杂草的叶片形态),这印证了模型在主导物种分类任务上的决策合理性。相比之下,传统基线模型的注意力分布往往呈全局性,倾向于覆盖杂草的整体轮廓;而 WeedMamba 更符合人类视觉认知习惯,能够精准聚焦于叶片脉络、茎部表皮等鉴别性细节特征,从而提取更丰富的视觉线索,减轻相似类别间的混淆。

为了进一步验证 WeedMamba 模型在杂草识别任务中的分类性能,在 DeepWeeds 和 CottonWeedID15 两个标准数据集上进行了混淆矩阵分析。在 DeepWeeds 数据集中,横纵坐标代表 9 类澳大利亚常见杂草。从混淆矩阵可以看出(图 4),本文方法在识别 Lantana、Parthenium、Rubber Vine 和 Siam Weed 等类别上均能达到 97% 以上的分类准确率,其中 Siam Weed 的识别准确率最高,达到 99.0%。然而,在识别 Chinee apple 和 Parkinsonia 时出现相对较多的分类错误,准确率分别为 93.8% 和 95.1%,这可能是这些杂草在不同生长阶段形态变化较大,或与背景植被存在较高的视觉相似性所致。

在 CottonWeedID15 数据集中,横纵坐标分别代表 15 种棉花田间常见杂草类别(图 5)。分析结果表明,本文算法对 Sicklepod 类别实现了 100% 的完美识别,显示出模型对具有显著形态特征杂草的强大辨别能力。特别是在 Purslane、Sicklepod 等叶形独特的杂草识别上表现优异。相比之下,模型在识别 Carpetweed、TallMorningglory 等类别时出现轻微误差,这可能是田间复杂背景下叶片遮挡或光照变化导致的。总体而言,模型通过有效学习不同杂草的鉴别性特征,在多个类别上均取得了优异的分类性能。

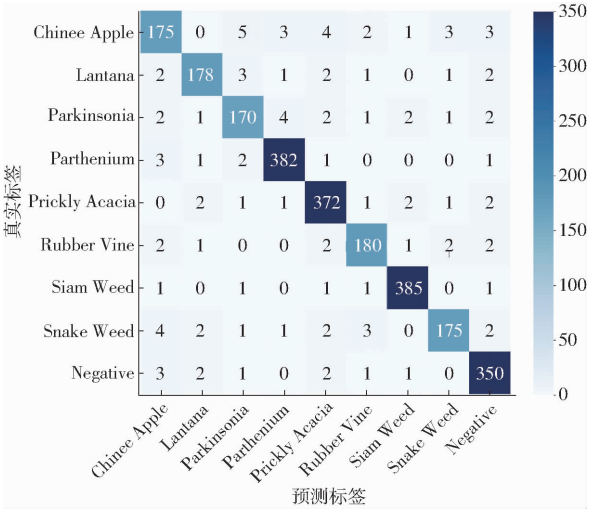

图 4 DeepWeeds 数据集的混淆矩阵

Fig. 4 Confusion matrix of DeepWeeds dataset

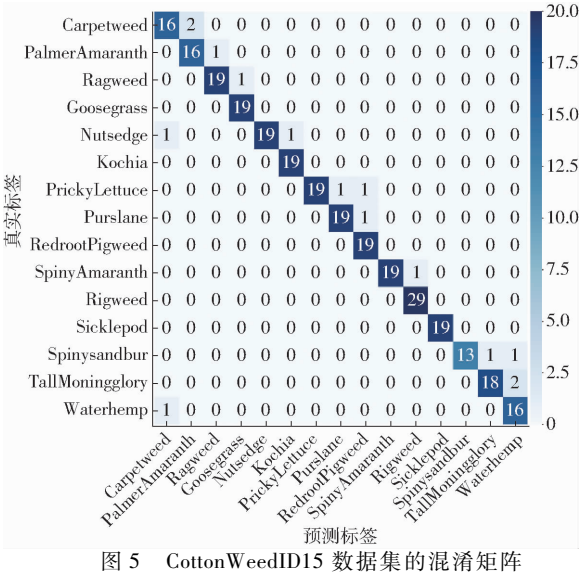

图 5 CottonWeedID15 数据集的混淆矩阵

Fig. 5 Confusion matrix of CottonWeedID15 dataset

2.8 可行性分析

WeedMamba 模型的轻量化设计为其实际田间部署奠定了坚实基础。评估模型部署可行性的核心在于其计算效率与资源消耗。WeedMamba 仅需 2.5×10^{9} 浮点运算次数的计算复杂度和 1.93 ×

10^7 的参数量，这一优势在对比模型中尤为突出。在 NVIDIA GeForce RTX 4060Ti GPU 上的测试结果表明，其推理吞吐量可达 503.5 幅/s。在实际田间部署中，模型需要在具备至少 RTX 3060 等算力相当的 GPU 平台上运行，以支持高分辨率输入和多分支网络结构。同时，为保证动态场景下的目标捕获，系统可实现 30 f/s 的推理帧率。在嵌入式或移动端平台，可通过降低输入分辨率至 512 像素 ×512 像素、通道裁剪、轻量化卷积以及模型量化等优化手段，进一步控制显存占用和计算延迟，从而实现稳定的田间部署。综上所述，本研究在模型效率与分类性能之间取得优异平衡，证实了在轻量化架构下兼顾高精度与高效率的可行性，为其实际应用提供了有力支撑。

3　结束语

针对杂草类内差异性显著、类间相似及田间环境干扰导致分类精度受限的问题，本文提出基于 Mamba 的多尺度高效杂草分类网络 WeedMamba，设计包含 CNN－SSM 双分支提取模块、多尺度过滤融合金字塔、自适应多核分类器；通过消融实验量化了各核心组件对分类性能的贡献，为杂草分类模型设计提供理论支撑；在 CottonWeedID15 和 DeepWeeds 数据集上，该模型的分类准确率分别达 96.4% 和 89.9%，实现了分类性能与模型复杂度的优异平衡；综上，WeedMamba 在单类主导杂草分类任务中表现出强环境适应性与鲁棒性，可支撑精准农业杂草识别与管理需求。

参　考　文　献

[1] 孙杰，杨静，丁书杰，等. 基于多注意力机制与编译图神经网络的高光谱图像分类[J]. 农业机械学报，2024，55(3)：183－192，212.
Sun Jie, Yang Jing, Ding Shujie, et al. Hyperspectral image classification based on multi-attention mechanism and compiled graph neural networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 183－192, 212. (in Chinese)

[2] Han J Y, Zhou T H, Huang D, et al. TD－Net: a time-dilation structured CNN for phoneme recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 2402－2416.

[3] Liu Y C, Sun Y N, Xue B, et al. A survey and performance evaluation of deep learning methods for small sample image classification[J]. Artificial Intelligence Review, 2023, 56(12): 14631－14686.

[4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84－90.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]// Proceedings of the 3rd International Conference on Learning Representations(ICLR), 2015: 1－14.

[6] He K, Zhang X, Ren S, et al. Deep residual learning forimage recognition[PP/OL]. V1. arXiv (2015－10－10)[2025－08－26]. https://arxiv.org/abs/1512.03385.

[7] Gu J, Wang Z, Kuen J, et al. Recent advances inconvolutional neural networks[J]. Pattern Recognition, 2018, 77: 354－377.

[8] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[PP/OL]. V2. arXiv(2017－04－11)[2025－08－26]. https://arxiv.org/abs/1611.05431.

[9] Liu Z, Mao H, Wu C Y, et al. A ConvNet for the 2020s[PP/OL]. V2. arXiv(2022－03－02)[2025－08－26]. https://arxiv.org/abs/2201.03545.

[10] Chollet F. Xception: deep learning with depthwiseseparable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, 2017: 1251－1258.

[11] Zhang X, Zhou X, Lin M, et al. ShuffleNet: anextremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition. Piscataway, 2018: 6848－6856.

[12] Howard A G, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[PP/OL]. V1. arXiv(2017－04－17)[2025－08－26]. https://arxiv.org/abs/1704.04861.

[13] 田甜，程志友，鞠薇，等. 基于 SimAM－ConvNeXt－FL 的茶叶病害小样本分类方法研究[J]. 农业机械学报，2024，55(3)：275－281.
Tian Tian, Cheng Zhiyou, Ju Wei, et al. Few-shot classification of tea diseases based on SimAM－ConvNeXt－FL[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 275－281. (in Chinese)

[14] Akkaya I B, Kathiresan S S, Arani E, et al. Enhancing performance of vision transformers on small datasets through local inductive bias incorporation[J]. Pattern Recognition, 2024, 153: 110510.

[15] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchicalvision transformer using shifted windows[PP/OL]. V2. arXiv (2021－03－23)[2025－08－26]. https://arxiv.org/abs/2011.05203.

[16] Dong X, Bao J, Chen D, et al. CSWin Transformer: a general vision transformer backbone with cross-shaped windows[PP/

OL]. V2. arXiv(2021 - 07 - 02)[2025 - 08 - 26]. https://arxiv.org/abs/2107.00032.
[17] Yu W, Luo M, Zhou P, et al. Metaformer is actually what you need for vision[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, 2022: 10819 - 10829.
[18] Han Kai, Wang Yunhe, Chen Hanting, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 87 - 112.
[19] Park C W, Seo S W, Kang N, et al. Artificial intelligence in health care: current applications and issues[J]. Journal of Korean Medical Science, 2020, 35(42):e379.
[20] Gu A, Goel K, Re C. Efficiently modeling long sequences with structured state spaces[J]. International Conference on Learning Representations (ICLR), 2022: 1124 - 1135.
[21] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces[J]. Advances in Neural Information Processing Systems 36 (NeurIPS 2023), 2023: 2534 - 2546.
[22] Liu Yue, Tian Yunjie, Jiao Jianbin, et al. VMamba: visual state space model[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(3): 1567 - 1582.
[23] Zhu Lianghui, Liao Bencheng, Zhang Qian, et al. Vision Mamba: efficient visual representation learning with bidirectional state space model[J]. International Journal of Computer Vision, 2024, 132(5): 2101 - 2118.
[24] Wang Q, Wu B, Zhu P, et al. ECA - Net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020: 11531 - 11539.
[25] 陈超, 齐峰. 卷积神经网络的发展及其在计算机视觉领域中的应用综述[J]. 计算机科学, 2019, 46(3): 63 - 73.
Chen Chao, Qi Feng. Development of convolutional neural networks and their applications in computer vision: a survey[J]. Computer Science, 2019, 46(3): 63 - 73. (in Chinese)
[26] Elfwing S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural Networks, 2018, 107: 3 - 11.
[27] Vasu P K A, Gabriel J, Zhu J, et al. MobileOne: an improved one millisecond mobile backbone[C]// Proceedings of the IEEE/CVF Conference on Computer Visionand Pattern Recognition (CVPR), 2023: 7907 - 7917.
[28] Vasu P K A, Gabriel J, Zhu J, et al. FastViT: a fast hybrid vision transformer using structural reparameterization[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 5785 - 5795.
[29] Yu W H, Zhou P, Yan S C, et al. InceptionNeXt: When inception meets ConvNeXt[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 5672 - 5683.
[30] Yu W H, Si C Y, Zhou P, et al. MetaFormer baselines for vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(2): 896 - 912.
[31] Han D C, Ye T Z, Han Y Z, et al. Agent attention: on the integration of softmax and linear attention[C]// Proceedings of the 17th European Conference on Computer Vision (ECCV), 2024: 124 - 140.
[32] Zhu L H, Liao B C, Zhang Q, et al. Vision Mamba: efficient visual representation learning with bidirectional state space model[C]// Proceedings of the 41st International Conference on Machine Learning (ICML), 2024.
[33] Liu Y, Tian Y, Zhao Y, et al. VMamba: visual state space model[PP/OL]. V1. arXiv(2024 - 01 - 18)[2025 - 08 - 26]. https://arxiv.org/abs/2401.10166.
[34] Chen X, Wu J, Lin T, et al. MambaVision: hierarchical state space models for multi-scale image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2025:12345 - 12355.