

杨淑婷, 李季, 刘正予, 等. 面向水稻杂草识别的高精度图像分类算法[J]. 江苏农业学报, 2025, 41(8): 1538-1552.

doi: 10.3969/j.issn.1000-4440.2025.08.010

面向水稻杂草识别的高精度图像分类算法

杨淑婷^{1,2}, 李季¹, 刘正予³, 马聪¹, 王蓉⁴

(1.宁夏农林科学院农业经济与信息技术研究所, 宁夏 银川 750002; 2.宁夏数智农业工程技术研究中心, 宁夏 银川 750002; 3.南京理工大学计算机科学与工程学院, 江苏 南京 210018; 4.农业农村部农产品质量安全监督检验测试中心<银川>, 宁夏 银川 750002)

摘要: 为实现稻田苗期杂草的精准识别与清除, 本研究开展了一项系统性的研究工作。通过对稻田实地环境的图像采集与数据整理, 构建了一个水稻杂草图像分类数据集。在此基础上, 提出了一种创新的高效水稻杂草图像分类算法 YOLOv8n-cls-Swift。在图像特征提取阶段, 采用 SwiftFormer 在复杂的田间场景下有效提取水稻植株与杂草的差异化特征。在分类预测阶段, 本研究设计了一种高效的加权分类层, 使模型能够更加精准地聚焦于区分性强的目标特征区域, 显著提升了模型对判别性目标特征的捕捉能力。结果表明, 本研究提出的模型能达到一个较高的识别准确率。本研究提出的一种水稻田杂草清除精准施药系统, 能够实现稻田杂草的全自动识别与清除, 有望在水稻种植的减药增效、环境保护等方面发挥重要作用。

关键词: 图像识别; 水稻; 深度学习; 卷积神经网络; 图像处理

中图分类号: S126 **文献标识码:** A **文章编号:** 1000-4440(2025)08-1538-15

High-precision image classification algorithm for recognition of rice weed

YANG Shuting^{1,2}, LI Ji¹, LIU Zhengyu³, MA Cong¹, WANG Rong⁴

(1. Institute of Agricultural Economics and Information Technology, Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan 750002, China; 2. Ningxia Digital and Intelligent Agricultural Engineering Technology Research Center, Yinchuan 750002, China; 3. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210018, China; 4. Supervision and Testing Center of Agricultural Products Quality and Safety (Yinchuan), Ministry of Agriculture and Rural Affairs, Yinchuan 750002, China)

Abstract: To achieve accurate identification and removal of weeds during the seedling stage in rice fields, this study carried out a systematic study. By collecting and organizing images from real rice field environments, a rice-weed image classification dataset was constructed. Based on this dataset, an innovative and efficient rice-weed image classification algorithm named YOLOv8n-cls-Swift was proposed. During the image feature extraction phase, SwiftFormer was employed to effectively extract discriminative features between rice plants and weeds under complex field conditions. In the classification prediction phase, an efficient weighted classification layer was designed to enable the model to focus more accurately on highly discrimi-

native target feature regions, and significantly enhanced its ability to capture distinguishing characteristics. The results demonstrated that, the proposed model achieved a high recognition accuracy. The precision herbicide application system for rice fields presented in this study can achieve fully automatic identification and removal of weeds, which is expected to play a significant role in reducing pesticide use, improving efficiency, and protecting the environment in rice

收稿日期: 2025-04-22

基金项目: 宁夏自然科学基金项目(2023AAC03411); 宁夏回族自治区重点研发项目(2023BCF01051, 2024BBF01013)

作者简介: 杨淑婷(1984-), 女, 宁夏盐池人, 硕士, 助理研究员, 研究方向为农业遥感、智能决策技术研究与应用。(E-mail: nxnkyyst@163.com)

通讯作者: 李季, (E-mail: lijinxny@163.com)

cultivation.

Key words: image recognition; rice; deep learning; convolutional neural network; image processing

水稻是全球重要的粮食作物之一,为全球超过50%的人口提供了主食^[1]。根据2022年联合国粮食及农业组织的数据,中国是世界上最大的水稻生产国。宁夏水稻种植历史悠久,引黄灌区的特殊地理条件和优越的自然环境使得该地区生产的稻米品质优良,享誉海内外。在稻田生态系统中,杂草的存在是一个普遍问题,它们不仅占据有限的种植面积,还与水稻竞争水分、养分和光照资源,严重影响水稻的生长和产量。传统的杂草处理方法往往依赖于人工除草或使用化学药剂,这些方法不仅成本高昂,还可能对环境造成负面影响。因此,精确确定杂草的具体位置,实现靶向管理和精准用药,已成为现代农业杂草管理的重要方向。近年来,深度学习技术为水稻杂草分类提供了新的解决方法^[2],通过无人机航拍技术获取稻田高清图像,图像经过预处理后被切割成多个小区域,再利用深度学习算法对每个小区域的水稻和杂草进行细粒度识别和分类,最终可实现稻田水稻杂草的细粒度识别分类,为除草剂的选择和靶向施药提供理论和技术依据。

图像分类任务是指输入一张图像并在输出图像中识别目标对象的类别^[3]。当前随着深度学习技术和计算机算力的快速发展,图像分类算法取得了突破性的进展,被广泛应用于自动驾驶、生产制造、智慧林业、农业检测等多个领域。当前基于深度学习的图像分类算法主要分为2类:基于卷积神经网络(CNN)的图像分类识别算法和基于Vision Transformer(ViT)的图像分类识别算法^[4]。卷积神经网络由卷积层、池化层和全连接层等组件构成。卷积层负责从输入数据中提取特征,其中每个神经元与上一层中位置相近的神经元形成局部连接,从而完整地表征输入数据的特征关联性。池化层用于降采样,以减少数据的维度和计算量。全连接层通常完成最后的图像分类。Krizhevsky等^[5]在2012年提出了AlexNet模型,并一举在ImageNet大尺度图像识别竞赛(ILSVRC)中夺魁,奠定了CNN在图像分类领域的主导地位。可变卷积、深度可分离卷积等技术的出现,使CNN的性能不断优化,ImageNet数据集上的图像识别准确率也逐步提高。基于Transformer模型^[6]在自然语言任务上的强势表现,

Vision Transformer^[7]被用于图像识别任务,并在大型图像识别基准数据集ImageNet榜单占据前列。ViT的核心结构包括块处理模块、位置编码、多头注意力机制和前馈神经网络。在块处理模块中,输入图像被处理成一块块不重叠的图像块。在位置编码过程中,ViT使用标准的可学习的一维位置嵌入向量,可由不同频率的正弦或余弦函数得到。在多头自注意力计算模块中,输入特征向量通过注意力模块后得到特征向量,计算每个块内部和块与块之间的特征相似性。前馈神经网络包含一层或多层的全连接层,主要用于变换维度与输出结果。后续方法以改进块处理模块、位置编码、多头注意力和前馈神经网络为主,得到各种变体模型,进一步优化了模型的特征提取能力,提高了模型的检测准确率^[8-10]。此外,还有一些方法结合卷积神经网络和ViT,降低ViT类模型的计算复杂度,同时结合2种方法的优势共同完成图像目标的分层特征提取,实现了图像识别速度和识别精度的高效统一。

近年来,国内外研究者利用深度学习技术针对水稻杂草图像分类提出多种方法。Deng等^[11]基于CNN实现水稻分蘖的自动检测与计数,并达到了较高的检测准确度。基于深度学习的图像分析技术在精准、高效地识别和计数水稻有效分蘖方面展现了巨大的应用潜力。孙俊等^[12]基于植物幼苗分类数据集提出一种融合空洞卷积与全局池化的多尺度特征融合的CNN模型,实现了对12种作物和杂草的高精度识别。然而,该数据集的图像样本是在实验室条件下拍摄采集得到的,背景较为单一,该模型在应对田间复杂背景下的稻田杂草识别问题时存在局限性。为解决自然光照和田间复杂背景下稻田苗期杂草的自动识别问题,邓向武等^[13]提出了一种结合CNN和迁移学习的方法,该方法将预训练的CNN模型迁移应用于稻田苗期杂草识别任务。该研究采集了鳢肠、丁香蓼、空心莲子草、千金子、野慈姑和稗草的共928张稻田苗期杂草图像,随机选取70%的样本作为训练集,其余30%作为测试集,并基于图像分类模型AlexNet、VGG16和GoogLeNet进行参数迁移学习,试验结果证明了方法的有效性。桂越等^[14]提出了一种基于CNN的农田杂草分类识别方法,并在农业基地构建

一个水稻和杂草的图像分类数据集,基于 CNN 训练并得到一个高效的水稻杂草分类模型,在测试集上有着不错的识别效果。为解决现有田间杂草识别方法在高准确率与实时性之间难以兼顾的问题,杨森森等^[15]提出一种基于改进 MobileViT 网络^[16]的轻量化杂草识别方法。这种方法通过 MobileViT 模块在特征提取网络中实现局部语义信息与全局语义信息的同步学习,仅需少量模块组合即可精准捕捉不同杂草和作物类别间的细微差异,最终模型的识别准确率优于大多数 CNN 方法。尽管上述方法在自建数据集上取得了较好的识别效果,但在实际应用场景中,存在真实世界的水稻和杂草在形态上相似度高,现有识别方法难以有效区分。

本研究拟以智能化分类识别水稻和杂草为目标,构建一个水稻杂草图像分类数据集,基于最新的 YOLOv8-cls 算法^[17]提出 YOLOv8-cls-Swift 算法,在图像特征提取阶段使用 SwiftFormer^[18]提取图像特征,其结合卷积和 Transformer 有效整合图像的整体特征与目标的局部特征信息,可实现复杂场景下水稻和杂草目标关键特征的高效区分性提取。在图像分类模块上,本研究设计一种高效的加权分类层,通过深度可分离卷积和动态权重分配机制,赋予水稻和杂草目标更高的权重,使得模型能够在细微差异中准确区分水稻与杂草目标。

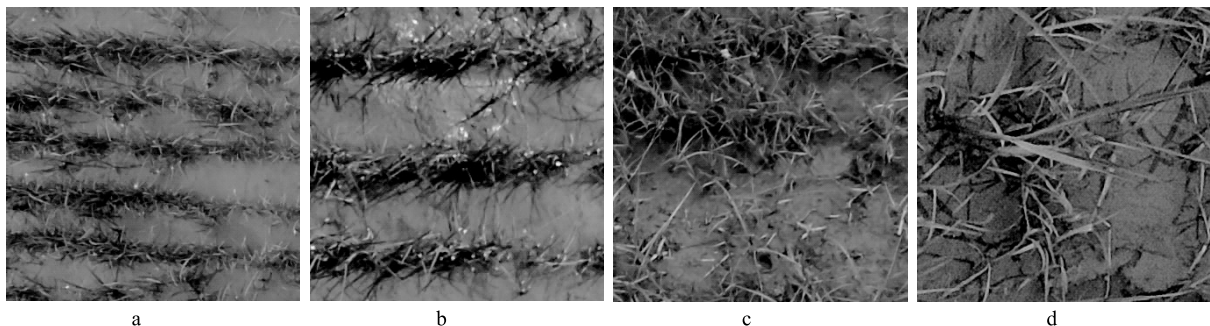
本研究拟通过计算机视觉技术实现田间杂草的精准识别与定位,进而实现水稻田杂草的精准清除。具体来说,不仅仅是为了找到杂草,更是为了通过人工智能技术实现对杂草的准确识别,并指导无人机

进行精准施药。在整个流程中,杂草的识别和位置确定是关键步骤,但最终目的还是要实现杂草的精准清除,以保护水稻的健康生长。基于此,本研究提出一种水稻田杂草精准施药系统,实现了从图像采集到杂草精准清除的全流程自动化作业,有望在水稻种植的减药增效、环境保护等方面发挥重要作用,实现水稻生产减施、节本、提质、增效,助力黄河流域生态保护和高质量发展先行区建设。

1 材料与方法

1.1 数据来源

本研究中水稻杂草图像分类数据集的图像采集地点为宁夏灵武市绿先锋水稻种植基地(37°934 114' N, 106°351 963' E),采集时间主要集中在 11:00-15:00,图像采集方式为无人机拍摄,采集设备为 DJI MAVIC 3 多光谱版,获得图像分辨率为 1 920×1 080。在对数据集的处理过程中,将原始大图以 336×336 的分辨率进行分割,得到 6 890 张图像,然后对得到的每张小图像单独做水稻和杂草的二分类标注,得到的图像大部分以单类别为主,仅有少量图像同时包含水稻和杂草目标,若杂草的比例小于水稻的比例则这部分图像被剔除,否则这些图像以杂草标签进行标注,同时若图像中大面积仅包含土地则这些图像同样被剔除,通过人工筛选最终得到 6 000 张有效图像。本研究以 8:1:1 的比例将自建的水稻杂草图像分类数据集划分为训练集、验证集和测试集,分别包含 4 800 张、600 张以及 600 张图像。图 1 是部分裁剪后的水稻和杂草图像分类数据集的图像示例。



a, b 为水稻图片;c, d 为杂草图片。

图 1 水稻和杂草示例

Fig.1 The examples of rice and weed

1.2 研究方法

本研究以智能化分类识别水稻和杂草为目标,

构建一个水稻杂草图像分类数据集,基于 YOLOv8-cls 算法提出 YOLOv8-cls-Swift,在图像特征提取阶

段使用 SwiftFormer 模型提取图像特征。在识别分类头上,本研究设计一种高效的加权分类层,使得模型能够在细微差异中准确区分水稻与杂草目标。

1.2.1 YOLOv8-cls YOLOv8 是 Ultralytics 公司提出的一款多任务模型框架,能够同时处理图像检测、目标分类、实例分割和旋转框目标检测等任务,也是 YOLO 系列新的算法,在准确率和速度上相较于前几代算法均有较大幅度的提升。YOLOv8-cls 是适用于目标分类任务的算法,与检测模型相似的是,其包含 YOLOv8n-cls、YOLOv8s-cls、YOLOv8m-cls、YOLOv8l-cls 以及 YOLOv8x-cls 5 个不同尺度的模型,其参数量依次增加,但所有模型的内部结构都是相同的,区别在于模型中模块的数量与通道数不同。图 2 是 YOLOv8s-cls 的整体模型结构。

YOLOv8s-cls 模型整体可分为图像特征提取和分类预测两大阶段。在图像特征提取阶段,输入图像经过堆叠的 CBS 和 C2f 模块来提取特征,其中 CBS 是

一个标准的卷积块,包含 1 个卷积层、1 个批量归一化层以及 1 个按比例指数线性单位激活函数 (SiLU)^[19]。C2f 模块首先用一个 CBS 模块进行特征提取,随后经过 split 操作将特征图分为 3 个子部分,即 CS_1 、 CS_2 和 CS_3 , CS_1 经过第 1 个 Bottleneck 模块得到 CB_1 ,其中 Bottleneck 包含 2 个 CBS 模块和 1 个残差连接层,接着 CS_2 和 CB_1 一同送入第 2 个 Bottleneck 层得到 CB_2 ,后续 CS_3 和 CB_2 一同送入第 3 个 Bottleneck 层得到 CB_3 ,最后将特征 CB_1 、 CB_2 和 CB_3 按通道维度拼接得到融合后的特征,将此特征送入一个 CBS 模块得到最后的特征向量。在分类预测阶段,图 2 中的 Classify 模块包含 1 个 CBS 模块、1 个自适应平均池化层、1 个线性层输出类别预测。在模型训练阶段,使用交叉熵损失计算目标类别预测值与真实标签值之间的偏差。在模型测试阶段,直接根据线性层的输出值返回预测结果。

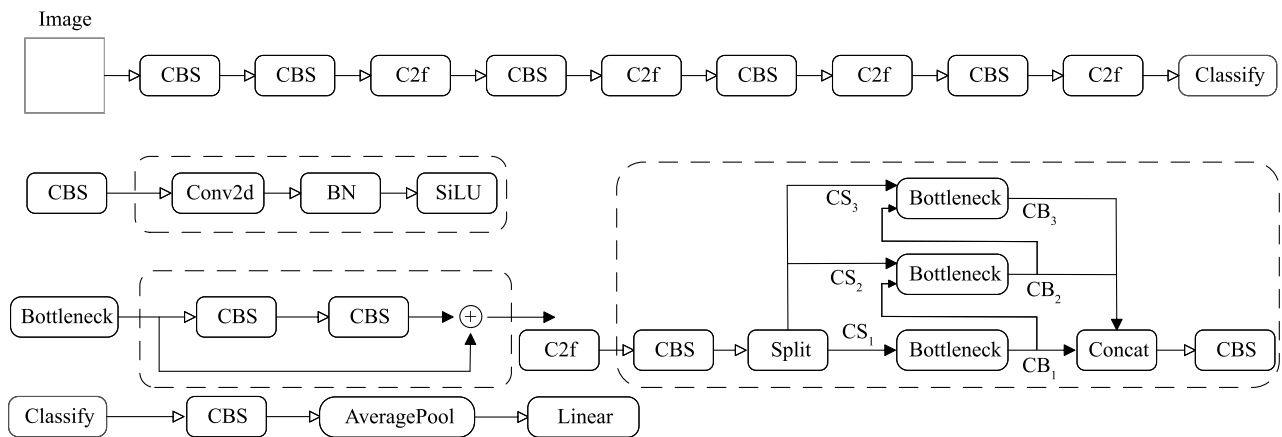


Image:输入图像;CBS:卷积模块;C2f:跨阶段特征融合模块;Classify:分类模块;Conv2d:卷积层;BN:批量归一化层;SiLU:按比例指数线性单位激活函数;Bottleneck:瓶颈块;Split:特征图分割模块;Concat:拼接模块;AveragePool:自适应平均池化层;Linear:线性层。

图 2 YOLOv8s-cls 的整体模型结构

Fig.2 The overall structure of YOLOv8s-cls model

1.2.2 YOLOv8-cls-Swift YOLOv8-cls 算法结构清晰,在公开数据集上表现出一定的竞争力,但其在复杂场景下(如细粒度目标分类任务中)目标识别性能较差。首先,YOLOv8-cls 算法对图像的特征提取不够充分,卷积块在复杂背景图像中无法充分捕捉多尺度目标特征,导致对目标区域的表征能力较弱。其次,YOLOv8-cls 算法的分类头设计简单,难以有效区分细粒度目标类别,在面对相似目标时容易出现误检的问题。在面对水稻杂草图像分类任务时,

YOLOv8-cls 算法存在难以从复杂的田间背景中准确提取水稻目标特征的问题,导致特征提取阶段激活目标区域过于分散。另外水稻与杂草相似度较高,简单的分类头设计难以高效区分水稻与杂草。因此尽管 YOLOv8-cls 在常规任务中表现良好,但在复杂场景和细粒度分类任务中,仍需进一步优化特征提取和分类头设计,以提高其整体性能。

针对以上问题,本研究基于 YOLOv8-cls 设计得到 YOLOv8-cls-Swift 算法,分别得到 YOLOv8n-cls-Swift、

YOLOv8s-cls-Swift、YOLOv8m-cls-Swift、YOLOv8l-cls-Swift 以及 YOLOv8x-cls-Swift 5 个不同尺度的模型,分别对应于 YOLOv8n-cls、YOLOv8s-cls、YOLOv8m-cls、YOLOv8l-cls 和 YOLOv8x-cls 5 个基准模型。首先为了提升特征提取模型对目标区域的表征能力,本研究使用 SwiftFormer 替换 YOLOv8-cls 中的卷积块,其高效地融合卷积层和 Transformer,实现图像上下文目标特征信息的关联,有效减少背景干扰,同时强化目标区域的

表征,显著提升了模型的特征提取效率和准确率。其次,为了提高模型分类头对不同场景图像分类目标分类的准确性,本研究提出一种高效的加权分类层(EWCA)。EWCA 结合深度可分离卷积和加权重参数设计,在不同感受野下提取多尺度目标特征,同时动态权重分配机制进一步优化了分类决策过程,使模型能够更加精准地聚焦于区分性强的特征区域。图 3 是本研究提出的 YOLOv8s-cls-Swift 的整体模型结构。

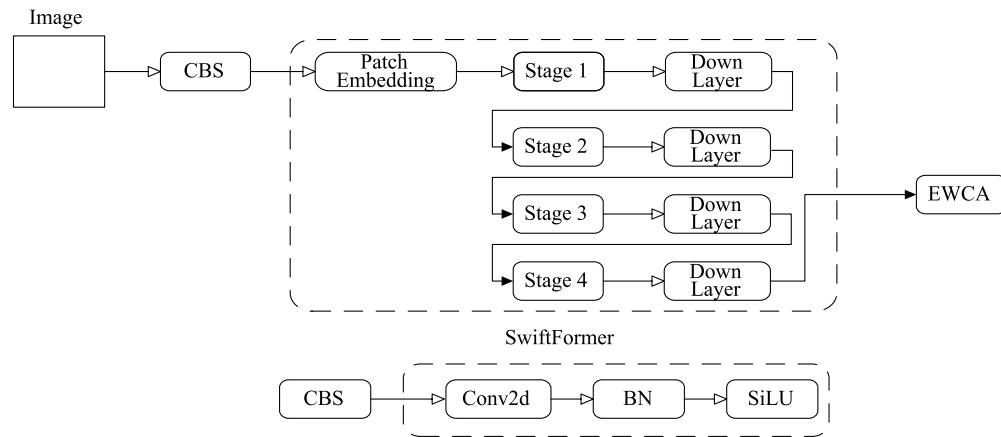


Image: 输入图像; CBS: 卷积模块; Conv2d: 卷积层; BN: 批量归一化层; SiLU: 按比例指数线性单位激活函数; Patch Embedding: 嵌入块向量; Stage 1: 第一阶段模块; Stage 2: 第二阶段模块; Stage 3: 第三阶段模块; Stage 4: 第四阶段模块; Down Layer: 下采样层; EWCA: 高效的加权分类层。

图 3 YOLOv8s-cls-Swift 整体模型结构

Fig.3 The overall structure of YOLOv8s-cls-Swift model

1.2.2.1 SwiftFormer 随着 Transformer 在自然语言任务中验证了其结构的高效性,ViT 被提出利用自注意力机制来提取图像特征。在主流的计算机视觉任务如图像分类、目标检测、实例分割上,ViT 及其变体^[20-22]在公开基准数据集如 ImageNet^[23]、MS COCO^[24]等上均表现出较好的检测性能。研究表明,ViT 中的自注意力机制已成为模型捕获图像上下文特征的一个较好的方法。然而自注意力机制对图像分辨率二次计算的复杂性限制了其在实时应用程序中的应用,尤其是在资源受限的设备上的部署^[25]。

为了减少模型在自注意力机制阶段的计算量,Shaker 等^[18]提出一种新的高效加性注意力机制模块,通过利用线性元素乘法取代原始注意力机制中的二次矩阵乘法计算,显著降低了模块的计算复杂性。一种新的高效加性注意力机制整体结构如图 4 所示。首先输入特征 X 分别经过 2 个线性层得到查询矩阵 Q 和键矩阵 K ,其中 $Q \in R^{n \times d}$, n 为输入向量的数量, d 为通道数。接着 Q 与可学习参数向量

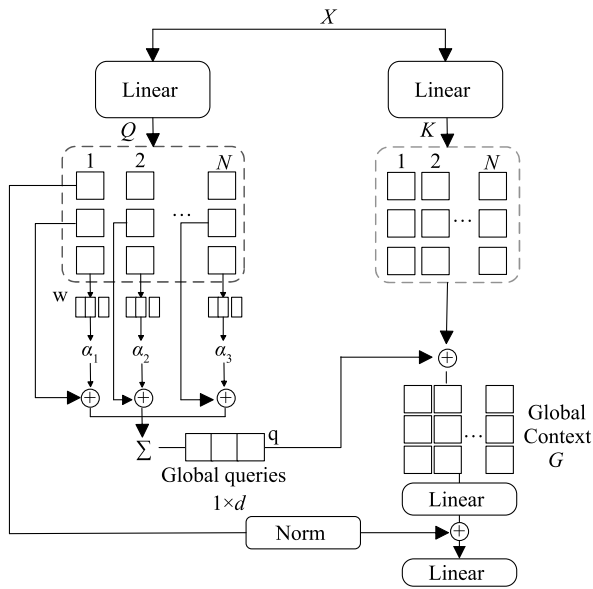
$W_\alpha \in R^d$ 相乘得到注意力权重 $\alpha \in R^d$,其计算方法如公式(1)所示。然后根据学习到的注意力权重以及 Q 进行池化,得到全局特征查询向量 $q \in R^d$,其计算方法如公式(2)所示。接下来,使用元素积方法对全局特征查询向量 q 和键矩阵 K 之间的交互进行编码,并最终得到全局上下文信息 $G \in R^{n \times d}$ 。最终利用得到的全局上下文信息 G 与查询矩阵 Q 作线性变换与归一化,得到最后的输出结果,其计算方法如公式(3)所示。

$$\alpha = Q \cdot W_\alpha / \sqrt{d} \quad (1)$$

$$q = \sum_{i=1}^n a_i \times Q_i \quad (2)$$

$$X_{out} = \text{Norm}(Q) + T(K \times q) \quad (3)$$

其中 Q 为查询矩阵, W_α 为可学习参数向量, d 为通道数, α 为注意力权重, q 为全局特征查询向量, Q_i 为第 i 层查询矩阵向量, α_i 为第 i 层的注意力权重, K 为键矩阵, Norm 为归一化操作, T 为线性变换操作。



Linear:线性层;Norm:归一化层;Global queries:全局序列;Global Context:全局上下文。 $\alpha_1, \alpha_2, \alpha_3$ 为加权系数。

图4 高效加性注意力模块

Fig.4 The module of Efficient Additive Attention

基于加性自注意力机制,Shaker 等^[18]提出了SwiftFormer,模型整体结构如图5所示。输入图像首先经过图像块嵌入层(Patch Embedding)降低输入特征分辨率,随后通过4个阶段分别提取不同层级的目标特征,其中每个阶段都包含1个卷积编码器(Conv Encoder)和1个SwiftFormer Encoder,然后送入不同的预测模块来适应不同的下游任务。在SwiftFormer中,输入图像 $X \in F^{H \times W \times 3}$,首先经过Patch Embedding层降低图像分辨率,得到 $X_1 \in F^{H/4 \times W/4 \times c_1}$,其中Patch Embedding层由2个步长为2的 3×3 卷积组成。然后,将特征图 X_1 送入模型的第一阶段,经过卷积编码器提取空间特征,通过SwiftFormer Encoder来学习图像全局-局部目标信息。图5分别画出了卷积编码器以及SwiftFormer Encoder的内部结构,其中卷积编码器由深度可分离卷积、归一化层、卷积层、激活函数GeLU、卷积层以及残差连接层组成。SwiftFormer Encoder包含1个深度可分离卷积和1个卷积层,可学习空间信息并对局部目标特征表征进行编码,然后将得到的编码特征送到高效的加性注意力中,该模块用于学习不同输入大小的每个尺度上的目标上下文信息,最后将输出的特征送入一个线性块中,其由2个卷积层、批处理归一化和激活函数GeLU组成。在每个阶段之间,SwiftFormer通过一个下采样层来降低特

征图的宽、高尺寸并提高特征图的通道数量。最终经过4个阶段的特征提取后,输入图像 X 经过SwiftFormer后得到4个不同尺寸的输出特征图 $X_1 \in F^{H/4 \times W/4 \times c_1}$ 、 $X_2 \in F^{H/8 \times W/8 \times c_2}$ 、 $X_3 \in F^{H/16 \times W/16 \times c_3}$ 和 $X_4 \in F^{H/32 \times W/32 \times c_4}$ (其中 H 和 W 分别是特征图的宽和高, c_1, c_2, c_3 和 c_4 分别是4个特征图的通道数)。SwiftFormer提出了一种新颖的加性注意力机制,通过线性乘法计算取代ViT中自注意力的二次矩阵计算,显著降低了模型的计算复杂度。此外,SwiftFormer结合卷积编码器提取的目标空间信息与Transformer提取的目标上下文信息,高效地表征了输入图像在不同分辨率下的特征信息,最终在复杂场景下实现目标特征的高效提取。

SwiftFormer作为一种基于Transformer的特征提取器,其核心优势是结合卷积和Transformer层有效整合图像的整体特征与目标的局部特征信息,最终在复杂场景下实现目标特征的高效提取。此外,SwiftFormer基于多尺度的图像特征提取有效弥补了之前方法在小目标和目标细节特征捕获上的不足,确保了模型对目标区域的全面感知能力。对于水稻杂草图像分类任务而言,这些特性尤为重要。水稻田间场景通常具有复杂的背景,如杂草、土壤和光照变化等,目标区域往往被遮挡或分布不规则。SwiftFormer通过强化对水稻和杂草区域的特征提取,不仅提升了分类精度,还增强了模型在多样化场景中的适应能力。

1.2.2.2 一种高效的加权分类层(EWCA) 现有大多数方法在目标分类层上的设计偏少,通常仅通过卷积层和线性分类层的简单叠加来实现。为了高效地区分图像背景与目标之间的特征差异,本研究提出一种高效的加权分类层(EWCA),整体结构如图6所示。输入特征图 F_{in} 首先分别经过 $3 \times 3, 5 \times 5$ 和 7×7 的深度可分离卷积^[26]来得到特征图 $F_{3 \times 3}, F_{5 \times 5}$ 以及 $F_{7 \times 7}$,具体操作如公式(4)、公式(5)和公式(6)所示,同时经过1个全局平均池化层和1个 1×1 的卷积层得到3层特征图的加权系数 α_1, α_2 和 α_3 ,具体操作如公式(7)所示。接着加权系数 α_1, α_2 和 α_3 分别与特征图 $F_{3 \times 3}, F_{5 \times 5}, F_{7 \times 7}$ 相乘得到加权后的特征图 F_{w1}, F_{w2}, F_{w3} 。最后将这3个特征图逐元素相加,得到最后的输出特征图 F_{out} ,具体操作如公式(8)所示。

$$F_{3 \times 3} = \text{DepthwiseConv}_{3 \times 3}(F_{in}) \quad (4)$$

$$F_{5 \times 5} = \text{DepthwiseConv}_{5 \times 5}(F_{in}) \quad (5)$$

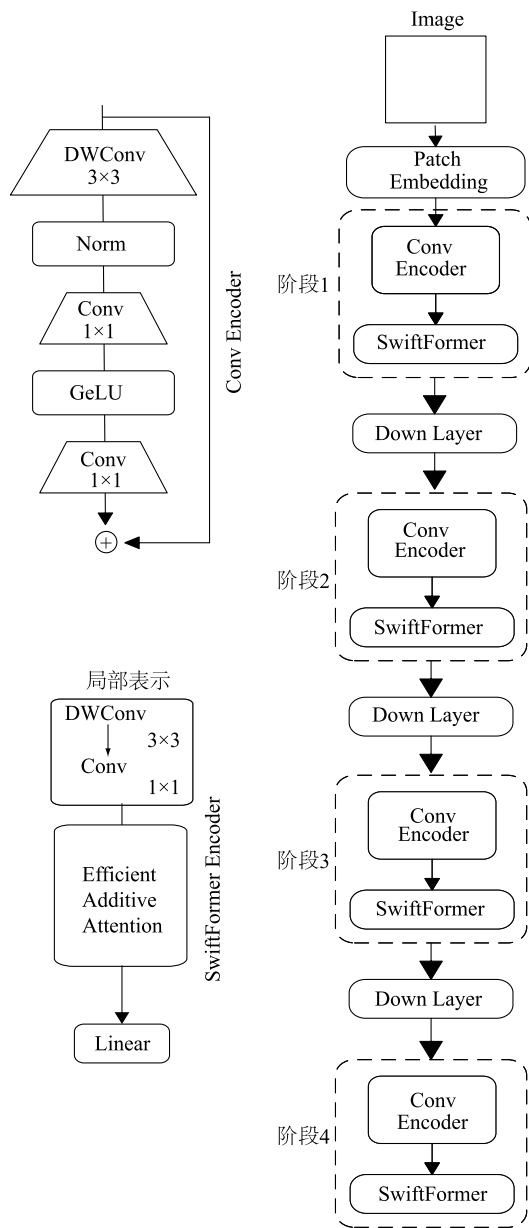


Image: 输入图像; Patch Embedding: 块嵌入层; Conv Encoder: 卷积编码器; Down Layer: 下采样层; DWConv: 深度可分离卷积层; Norm: 归一化层; Conv: 卷积层; GeLU: 非线性激活函数; Linear: 线性层; Efficient Additive Attention: 高效的加性注意力。

图 5 SwiftFormer 模型结构

Fig.5 The structure of SwiftFormer

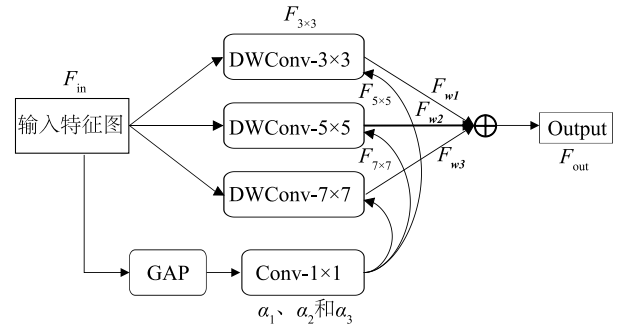
$$F_{7 \times 7} = \text{DepthwiseConv}_{7 \times 7}(F_{in}) \quad (6)$$

$$\alpha_1, \alpha_2, \alpha_3 = \text{Conv}_{1 \times 1}[\text{GlobalAvgPool}(F_{in})] \quad (7)$$

$$F_{out} = \alpha_1 \times F_{3 \times 3} + \alpha_2 \times F_{5 \times 5} + \alpha_3 \times F_{7 \times 7} \quad (8)$$

式中, F_{in} 是输入特征图, $\text{DepthwiseConv}_{i \times i}$ 是卷积核为 i 的深度可分离卷积, $\text{Conv}_{j \times j}$ 是卷积核为 j 的二维普通卷积, $F_{3 \times 3}$ 、 $F_{5 \times 5}$ 以及 $F_{7 \times 7}$ 是中间层输出特征

图, GlobalAvgPool 是全局平均池化操作, α_1 、 α_2 和 α_3 是 3 个加权系数, F_{out} 是最后的输出特征图。



DWConv: 深度可分离卷积; GAP: 全局平均池化层; Conv: 普通二维卷积层; α_1 、 α_2 和 α_3 为加权系数; F 表示特征图。

图 6 一种高效的加权分类层 EWCA 模块结构

Fig.6 An efficient weighted classification layer of EWCA module structure

EWCA 作为一种优化的分类头,其主要优点在于通过深度可分离卷积和动态权重分配机制,实现了不同感受野下特征的高效融合。传统分类头在处理目标与背景特征相似的场景时,常因模型提取到的目标特征不充分或冗余而导致分类性能下降。本研究提出的 EWCA 通过引入深度可分离卷积,在不同感受野下提取多尺度目标特征,同时动态权重分配机制进一步优化了分类决策过程,使模型能够更加精准地聚焦于区分性强的特征区域。这种设计不仅提升了模型分类效率,还增强了模型对细粒度类别的区分能力。在水稻杂草图像分类任务中,EWCA 的优势尤为突出。由于水稻与杂草目标常具有高度相似的纹理和形状,传统分类头容易在决策阶段混淆目标,导致误分类。而 EWCA 通过高效融合多层特征并赋予关键特征更高的权重,使得模型能够在细微差异中准确区分水稻与杂草目标。这种机制不仅提高了模型分类精度,还增强了其在复杂背景下的鲁棒性,从而在实际应用中展现出更高的实用价值。

1.3 试验环境和模型配置

1.3.1 试验环境 本研究模型训练时搭载的操作系统为 Ubuntu20.04,处理器为 Intel Xeon(R) Platinum 8352V CPU @ 2.1GHZ,显卡为 24 G 显存的 NVIDIA GeForce RTX 3090,搭配的统一计算设备架构(Compute unified device architecture, CUDA)和 CUDNN(CUDA 深度神经网络库)版本分别为 12.0

和 8.8.0。在训练环境上,本研究使用 Python3.9 作为基础,使用深度学习框架 PyTorch 来训练与测试模型,其版本为 2.0.1,其他一些重要的库如下:OpenCV 版本为 4.9.0.80, Pillow 版本为 10.4.0, PYYAML 版本为 6.0.1, Numpy 版本为 1.26.4, Matplotlib 版本为 3.9.1。

1.3.2 模型超参数 在训练阶段,本研究算法的总训练轮数为 100 轮,批次设置为 64,图像尺寸统一缩放到 224×224;使用随机梯度下降算法作为优化器,初始学习率、最终学习率、动量以及权重衰减分别设置为 0.010 0、0.000 1、0.937 0 以及 0.000 5。另外分别将图像色相、饱和度以及明度变换的概率设置为 0.015、0.700 以及 0.400。在几何数据增强策略上,本研究方法以 0.5 的概率对图像尺寸进行缩放,以 0.5 的概率对图像进行左右翻转。本研究选择 SwiftFormer_S 作为基准, YOLOv8n-cls-Swift 在 4 个阶段的模块数量分别为 1、1、1 和 1,4 个阶段特征图的输出通道数分别为 48、56、96 和 112。YOLOv8s-cls-Swift 在 4 个阶段的模块数量分别为 3、3、9 和 6,4 个阶段特征图的输出通道数分别为 48、64、168 和 224。YOLOv8m-cls-Swift 在 4 个阶段的模块数量分别为 3、3、12 和 6,4 个阶段特征图的输出通道数分别为 64、128、256 和 512。YOLOv8l-cls-Swift 在 4 个阶段的模块数量分别为 3、3、12 和 6,4 个阶段特征图的输出通道数分别为 96、192、386 和 768。YOLOv8x-cls-Swift 在 4 个阶段的模块数量分别为 3、3、12 和 3,4 个阶段特征图的输出通道数分别为 128、256、512 和 1 024。

1.4 试验评价指标

本研究选取准确率 (*Accuracy*)、Top1 准确率 (*Top1-Accuracy*)、精确率 (*Precision*)、召回率 (*Recall*) 以及精确率和召回率的调和平均 (*F1-score*) 作为试验指标^[27]。*Accuracy* 和 *Top1-Accuracy* 的计算分别如公式(9)和公式(10)所示。

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

$$Top1-Accuracy = \frac{\sum_{i=0}^{N-1} [f(x_i) = y_i]}{N} \quad (10)$$

其中, *TP* 是算法对目标类别预测为正类且实际目标的类别标签也为正类; *TN* 是算法对目标类别预测为正类但实际目标的类别标签为负类; *FP* 是算法

对目标类别预测为负类但实际目标的类别标签为正类; *FN* 是算法对目标类别预测为负类且实际目标的类别标签为负类。 *N* 为类别的个数(即目标类别为 0, 1, ..., *N*-1), 样本为 x_i , 其对应的类别标签为 y_i , 算法预测函数为 f 。

Precision、*Recall* 以及 *F1-score* 的计算方法如公式(11)、公式(12)和公式(13)所示。

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

2 结果与分析

2.1 水稻杂草分类数据集的试验结果

表 1 是 ViT-B、ViT-L、SwiftFormer-XS、SwiftFormer-S、SwiftFormer-L1、YOLOv8n-cls、YOLOv8s-cls、YOLOv8m-cls、YOLOv8l-cls、YOLOv8x-cls 和本研究提出的方法在水稻杂草分类数据集上的指标对比结果。ViT-B 是视觉变换器中参数较少的模型,其在水稻杂草分类数据验证集上的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 4 个指标分别能达到 88.9%、87.8%、90.4% 及 89.1%,对水稻和杂草具有一定的分辨能力。具有更多参数的 ViT-L 模型的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 4 个指标分别能达到 93.4%、91.0%、93.9% 及 92.4%,相较于 ViT-B 分别增加了 4.5 个百分点、3.2 个百分点、3.5 个百分点和 3.3 个百分点,这表明模型参数的增加和优化能提高图像识别准确率。SwiftFormer-XS 在 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 这 4 个指标上分别能达到 87.8%、87.1%、89.5% 和 88.3%。SwiftFormer-S 在这 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别较 SwiftFormer-XS 增加了 3.4 个百分点、2.9 个百分点、3.7 个百分点和 3.3 个百分点,SwiftFormer-L1 在这 4 个指标上较 SwiftFormer-XS 分别增加了 5.8 个百分点、4.3 个百分点、4.7 个百分点和 4.5 个百分点,基于多尺度的图像特征提取有效弥补了卷积层在小目标和目标细节特征捕获上的不足,增强了模型对目标区域的感知能力,提高了水稻杂草分类的性能。YOLOv8n-cls 是 YOLOv8-cls 算法中参数最少的模型,其在水稻杂草分类数据验证集上的

Top1-Accuracy、*Precision*、*Recall* 以及 *F1-score* 分别能达到 90.2%、88.4%、91.2% 以及 89.8%，对水稻和杂草有一定的分辨能力。本研究基于 YOLOv8n-cls 设计得到的 YOLOv8n-cls-Swift 模型的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别能达到 92.1%、89.8%、93.0% 以及 91.4%，相较于基准模型 YOLOv8n-cls 分别提高了 1.9 个百分点、1.4 个百分点、1.8 个百分点和 1.6 个百分点，这表明图像特征提取能力的增强能有效提高模型的检测准确率。SwiftFormer 基于多尺度的图像特征提取能有效弥补卷积层在小目标和目标细节特征捕获上的不足，确保了模型对目标区域的全面感知能力。YOLOv8s-cls 在水稻杂草分类数据验证集上的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别能达到 92.8%、90.4%、93.6% 以及 92.0%，而本研究提出的 YOLOv8s-cls-Swift 模型在这 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别能达到 93.3%、91.0%、93.9% 以及 92.4%，较 YOLOv8s-cls 分别提升了 0.5 个百分点、0.6 个百分点、0.3 个百分点以及 0.4 个百分点。值得注意的是，本研究提出的最轻量的 YOLOv8n-cls-Swift 仅比 YOLOv8s-cls 在 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别落后 0.7 个百分点、0.6 个百分点、0.6 个百分点和 0.6 个百分点。YOLOv8m-cls 在水稻杂草分类数据验证集上的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别能达到 93.6%、91.4%、94.4% 以及 92.9%，而本研究提出的 YOLOv8m-cls-Swift 在这 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别能达到 93.9%、91.7%、94.6% 以及 93.1%，较 YOLOv8m-cls 在 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别提升了 0.3 个百分点、0.3 个百分点、0.2 个百分点以及 0.2 个百分点。YOLOv8l-cls 在水稻杂草分类数据验证集上的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别能达到 94.4%、92.8%、95.0% 以及 93.9%，而本研究提出的 YOLOv8l-cls-Swift 在这 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别能达到 94.9%、92.9%、95.6% 以及 94.2%，较 YOLOv8l-cls 分别在 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上提升了 0.5 个百分点、0.1 个百分点、0.6 个百分点以及 0.3 个百分点。YOLOv8x-cls 在水稻杂草分类数据验证集上的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别能达

到 95.1%、93.3%、95.8% 以及 94.5%，而本研究提出的 YOLOv8x-cls-Swift 在这 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别能达到 95.3%、93.4%、95.9% 以及 94.6%，较 YOLOv8x-cls 在 4 个指标 (*Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score*) 上分别提升了 0.2 个百分点、0.1 个百分点、0.1 个百分点以及 0.1 个百分点。这些结果不仅证明了改进图像特征模型的有效性和必要性，同时改进的分类头进一步地提高了模型分类准确率。

表 1 ViT、SwiftFormer、YOLOv8-cls 与本研究提出的方法在自建水稻杂草分类数据验证集上的指标对比

Table 1 Comparison of metrics between ViT, SwiftFormer, YOLOv8-cls and the research method on a validation set of self-constructed dataset for rice weed classification

模型	前 1 准确率 (%)	精确率 (%)	召回率 (%)	精确率和召回率的调和平均 (%)
ViT-B	88.9	87.8	90.4	89.1
ViT-L	93.4	91.0	93.9	92.4
SwiftFormer-XS	87.8	87.1	89.5	88.3
SwiftFormer-S	91.2	90.0	93.2	91.6
SwiftFormer-L1	93.6	91.4	94.2	92.8
YOLOv8n-cls	90.2	88.4	91.2	89.8
YOLOv8n-cls-Swift	92.1	89.8	93.0	91.4
YOLOv8s-cls	92.8	90.4	93.6	92.0
YOLOv8s-cls-Swift	93.3	91.0	93.9	92.4
YOLOv8m-cls	93.6	91.4	94.4	92.9
YOLOv8m-cls-Swift	93.9	91.7	94.6	93.1
YOLOv8l-cls	94.4	92.8	95.0	93.9
YOLOv8l-cls-Swift	94.9	92.9	95.6	94.2
YOLOv8x-cls	95.1	93.3	95.8	94.5
YOLOv8x-cls-Swift	95.3	93.4	95.9	94.6

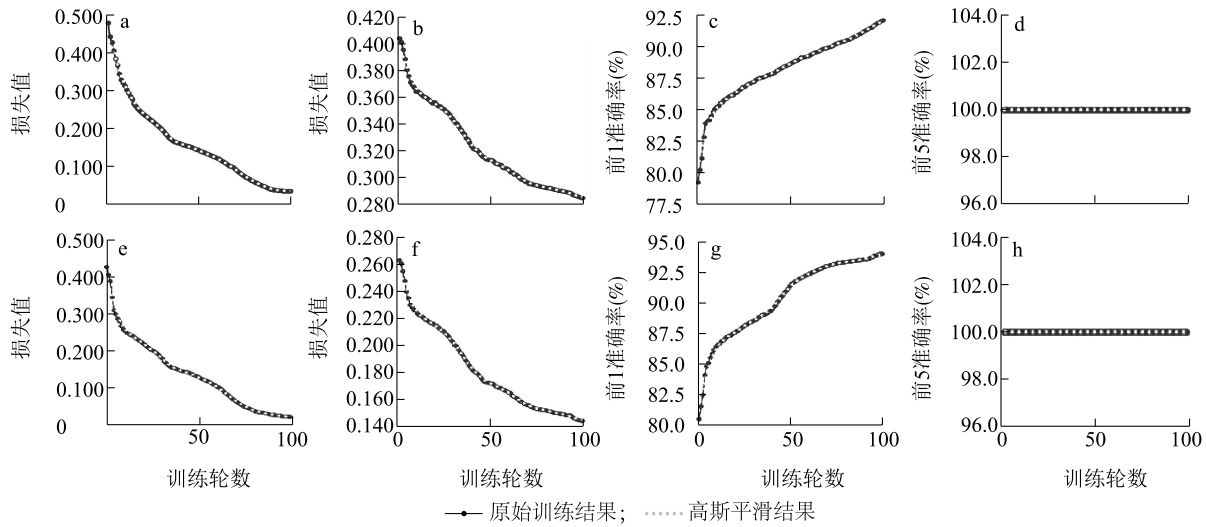
图 7 是 YOLOv8n-cls-Swift 和 YOLOv8m-cls-Swift 模型在本研究自建的水稻杂草图像分类数据验证集上的训练结果。从图 7 中可以看出，YOLOv8n-cls-Swift 模型的训练损失和验证损失在经过 100 轮模型训练后分别收敛到 0.033 6 和 0.284 4 (图 7a 和 7b)，而 YOLOv8m-cls-Swift 模型的训练损失和验证损失在经过 100 轮模型训练后分别收敛到 0.022 1 和 0.143 3 (图 7e 和 7f)。在 *Top1-Accuracy* 指标上，YOLOv8n-cls-Swift 和 YOLOv8m-cls-Swift 在 100 轮训练后分别能达到 92.1% 和 93.9% (图 7c 和 7g)，拥有更多参数的 YOLOv8m-cls-Swift 模型学习到

更具区分性的目标特征,最终能有一个较高的识别准确率。

2.2 不同改进模块的消融试验

表2是YOLOv8n-cls、YOLOv8s-cls模型与本研究提出的不同改进模块得到的变体模型在自建水稻杂草分类数据集的验证集上的试验结果指标对比。试

验基准模型YOLOv8n-cls和YOLOv8s-cls在自建水稻杂草分类数据集的验证集上的 $Top1-Accuracy$ 分别达到90.2%和92.8%,精确率分别为88.4%和90.4%,召回率分别为91.2%和93.6%, $F1-score$ 分别能达到89.8%和92.0%。虽然这2个模型整体表现都还不错,但仍存在进一步提升的空间。



a 为 YOLOv8n-cls-Swift 在训练集上的损失;b 为 YOLOv8n-cls-Swift 在验证集上的损失;c 为 YOLOv8n-cls-Swift 在验证集上前 1 准确率;d 为 YOLOv8n-cls-Swift 在验证集上前 5 准确率;e 为 YOLOv8m-cls-Swift 在训练集上的损失;f 为 YOLOv8m-cls-Swift 在验证集上的损失;g 为 YOLOv8m-cls-Swift 在验证集上前 1 准确率;h 为 YOLOv8m-cls-Swift 在验证集上前 5 准确率。

图 7 YOLOv8n-cls-Swift 和 YOLOv8m-cls-Swift 在水稻杂草分类数据集上的训练结果

Fig.7 Training results of YOLOv8n-cls-Swift and YOLOv8m-cls-Swift on rice weed classification validation dataset

表 2 不同改进模块在自建水稻杂草分类数据验证集上的指标对比

Table 2 Comparison of metrics of different improvement modules on self-constructed rice weed classification validation dataset

模型	SwiftFormer	EWCA	前 1 准确率 (%)	精确率 (%)	召回率 (%)	精确率和召回率的调和平均 (%)
YOLOv8n-cls	×	×	90.2	88.4	91.2	89.8
Model1	√	×	91.6	89.4	92.5	91.0
Model2	×	√	91.1	89.2	92.0	90.6
YOLOv8n-cls-Swift	√	√	92.1	89.8	93.0	91.4
YOLOv8s-cls	×	×	92.8	90.4	93.6	92.0
Model3	√	×	93.1	90.8	93.8	92.3
Model4	×	√	93.0	90.6	93.7	92.1
YOLOv8s-cls-Swift	√	√	93.3	91.0	93.9	92.4

EWCA:高效的加权分类层。

Model1 和 Model3 分别基于 YOLOv8n-cls 和 YOLOv8s-cls 模型,将原始卷积块替换为 SwiftFormer,在 $Top1-Accuracy$ 、 $Precision$ 、 $Recall$ 以及 $F1-score$ 方面均取得了提升。具体而言,Model1 的 $Top1-Accuracy$ 、 $Precision$ 、 $Recall$ 和 $F1-score$ 分别能达到 91.6%、89.4%、92.5%和 91.0%,较基准模型 YOLOv8n-cls 在

4 个指标上分别提高了 1.4 个百分点、1.0 个百分点、1.3 个百分点和 1.2 个百分点。Model3 的 $Top1-Accuracy$ 、 $Precision$ 、 $Recall$ 以及 $F1-score$ 分别能达到 93.1%、90.8%、93.8% 以及 92.3%,较基准模型 YOLOv8s-cls 在 4 个指标上分别提高了 0.3 个百分点、0.4 个百分点、0.2 个百分点和 0.3 个百分点。

SwiftFormer 通过其高效的图像特征提取能力提升了模型的判别力,在保持计算效率的同时有效增强了水稻图像的分类精度。Model2 和 Model4 在 YOLOv8n-cls 和 YOLOv8s-cls 模型基础上引入了本研究提出的 EWCA 分类头。与基准模型 YOLOv8n-cls 相比,Model2 的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别提高了 0.9 个百分点、0.8 个百分点、0.8 个百分点和 0.8 个百分点,分别达到了 91.1%、89.2%、92.0% 和 90.6%。与基准模型 YOLOv8s-cls 相比,Model4 的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别提高了 0.2 个百分点、0.2 个百分点、0.1 个百分点和 0.1 个百分点,分别达到了 93.0%、90.6%、93.7% 和 92.1%。EWCA 分类头通过结合深度可分离卷积和权重参数,增强了模型对背景和目标区域的区分能力,显著提高了模型在复杂背景下的分类表现。

YOLOv8n-cls-Swift 和 YOLOv8s-cls-Swift 模型将 SwiftFormer 与 EWCA 分类头联合应用,进一步提升了模型性能。相较于基准模型 YOLOv8n-cls, YOLOv8n-cls-Swift 的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别提高了 1.9 个百分点、1.4 个百分点、1.8 个百分点和 1.6 个百分点,在 4 个指标上分别能达到 92.1%、89.8%、93.0% 和 91.4%。相较于基准模型 YOLOv8s-cls, YOLOv8s-cls-Swift 的 *Top1-Accuracy*、*Precision*、*Recall* 以及 *F1-score* 分别提高了 0.5 个百分点、0.6 个百分点、0.3 个百分点和 0.4 个百分点,在 4 个指标上分别能达到 93.3%、91.0%、93.9% 和 92.4%,在各项指标中表现最佳。试验结果表明,SwiftFormer 高效的特征提取能力与 EWCA 分类头精确的目标分离能力在水稻图像分类任务中具有互补性,两者的结合显著提升了模型对目标对象的特征提取学习能力以及对目标和背景的判别能力,最终使得模型在分类性能上实现了最优表现。整体而言,SwiftFormer 与 EWCA 结构的引入显著改善了 YOLOv8-cls 系列模型在水稻图像分类任务中的表现,为提升复杂背景图像的目标分类精度提供了一种有效的模型优化方案。

2.3 图像热力图可视化

为了更直观地验证本研究提出的改进模型在特征提取与分类能力上的适用性,本研究使用 Grad-CAM^[28] 可视化 YOLOv8s-cls-Swift 在多个图像上不同层的热力图(图 8)。图 8 中每列的第一幅图像是原始图像,第二列、第三列和第四列为热力图,分别

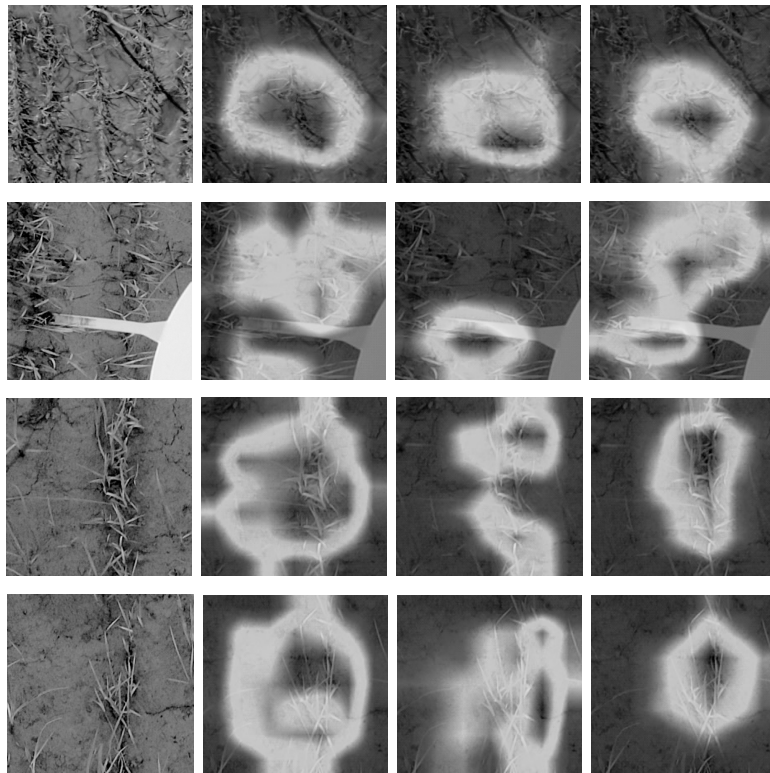
为基于 SwiftFormer 第二阶段的输出和第四阶段的输出以及最后分类头输出的热力图。从第二列图可以看出,输入图像经过 SwiftFormer 前 2 个阶段的特征提取之后激活的区域较为分散,热力分布不仅覆盖了水稻目标,还扩展到了大面积背景区域,表明其早期特征提取存在显著冗余。在第三列的热力图中,特征图的激活值范围开始逐渐缩小,但仍未能完全消除背景干扰,其特征聚焦依然存在模糊性。模型试图捕捉目标特征的同时,部分注意力仍分配到了非目标区域,表现出较强的局部依赖性。在第四列的热力图中,特征图的热力分布高度集中且强度均匀,清晰地覆盖了目标区域的关键特征,而对背景区域几乎没有激活。这一结果充分体现了 EWCA 分类头的贡献,通过深度可分离卷积减少了参数冗余,同时结合权重参数高效融合了目标特征,增强了分类器对不同目标类别的区分能力。

图 9 是 YOLOv8s-cls 与 YOLOv8s-scls-Swift 使用 Grad-CAM 工具可视化多个图像在特征提取网络后的热力图,其中第一列是原始输入图像,第二列和第三列分别是 YOLOv8s-cls 和 YOLOv8s-cls-Swift 在特征提取网络后的热力图。从图 9 中可以看出, YOLOv8s-cls 经过卷积特征提取后得到的输出特征图的激活区域较为分散,且关注的区域不止水稻目标,同时输出了部分图像背景信息。本研究提出的 YOLOv8s-cls-Swift 能较全面地捕捉到水稻目标,其多尺度特征捕捉和动态赋权机制有效剔除了背景冗余,使模型能够更准确地定位水稻目标。

图 10 是 YOLOv8s-cls 与 YOLOv8s-cls-Swift 使用 Grad-CAM 工具可视化多个图像在分类头后的热力图,其中第一列是原始输入图像,第二和第三列分别是 YOLOv8s-cls 和 YOLOv8s-scls-Swift 在分类头得到的热力图。从图 10 中可以看出,相较于基准模型 YOLOv8s-cls 的卷积检测头,本研究提出的 YOLOv8s-cls-Swift 方法中引入的高效加权分类头在目标区域特征融合方面展现出优越的性能,显著增强了模型对目标特征的判别能力。

2.4 模型参数和推理时间

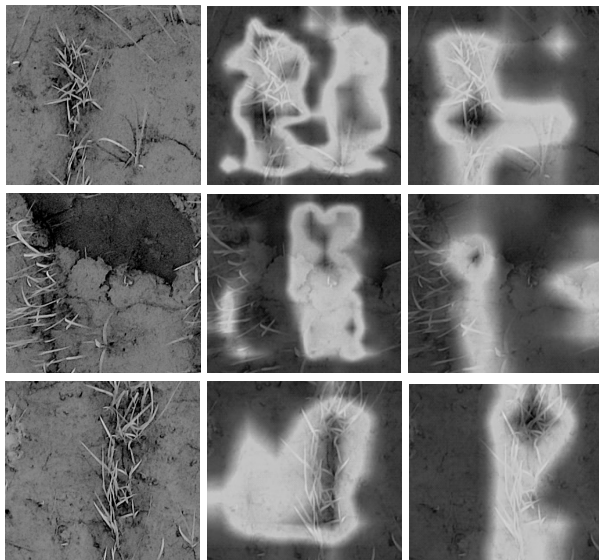
为了进一步评估 YOLOv8-cls-Swift 算法的实际应用价值,本研究对不同尺度模型在水稻杂草分类数据集上的计算复杂度与推理性能进行了试验对比。测试使用了 1 000 张测试图像,所有图像的尺寸统一缩放至 224×224,推理用时包括模型对输入图



每列的第一幅图像是原始图像,第二列、第三列和第四列均为热力图,分别为基于 SwiftFormer 第二阶段的输出和第四阶段的输出以及最后分类头输出的热力图。

图 8 YOLOv8s-clc-Swift 使用 Grad-CAM 可视化不同层得到的特征图

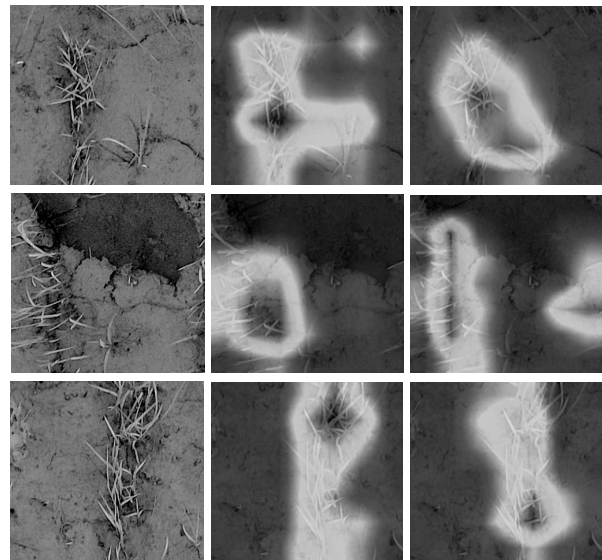
Fig.8 Feature maps obtained by YOLOv8s-clc-Swift through using Grad-CAM to visualize different layers



第一列是原始输入图像,第二列和第三列分别是 YOLOv8s-clc 和 YOLOv8s-sclc-Swift 在特征提取网络后的热力图。

图 9 YOLOv8s-clc 和 YOLOv8s-sclc-Swift 使用 Grad-CAM 可视化特征提取层得到的特征图

Fig.9 Feature maps obtained by YOLOv8s-clc and YOLOv8s-sclc-Swift through using Grad-CAM to visualize feature extraction layers



第一列是原始输入图像,第二列和第三列分别是 YOLOv8s-clc 和 YOLOv8s-sclc-Swift 在分类头得到的热力图。

图 10 YOLOv8s-clc 和 YOLOv8s-sclc-Swift 使用 Grad-CAM 可视化特征提取层引入分类头后得到的特征图

Fig.10 Feature maps obtained by YOLOv8s-clc and YOLOv8s-sclc-Swift through using Grad-CAM to visualize feature extraction layers and introducing classification head

像的前处理、模型推理及后处理的总时间,总耗时单位为帧率每秒(FPS)。如表 3 所示,YOLOv8n-cls 是参数量和计算量最少的模型,其计算量仅为 3.4 GFLOPs(1 s 1×10^9 次的浮点运算),在 NVIDIA GeForce RTX 3090 设备上模型推理仅耗时 5.3 ms。本研究提出的同等参数量级下的 YOLOv8n-cls-Swift 模型虽然参数量和计算量略有提高,但模型推理仅耗时 5.4 ms。YOLOv8s-cls 的参数量和计算量分别是 5 083 298 和 12.6 GFLOPs,在 NVIDIA GeForce RTX 3090 设备上模型推理仅耗时 5.6 ms,而本研究提出的同等参数量下的 YOLOv8s-cls-Swift 模型的参数量和计算量分别增加到 5 934 530 和 16.4 GFLOPs,在 NVIDIA GeForce RTX 3090 上模型推理仅耗时 5.8 ms。另外值得注意的是,最大尺寸的 YOLOv8x-cls 模型的参数量和计算量分别是 56 144 402 和 154.3 GFLOPs,在 NVIDIA GeForce RTX 3090 上模型推理仅耗时 8.3 ms。而本研究提出的同等参数量下的 YOLOv8x-cls-Swift 模型的参数量和计算量分别是 68 455 554 和 178.4 GFLOPs,在 NVIDIA GeForce RTX 3090 上模型推理仅耗时 8.6 ms。尽管本研究提出的 YOLOv8-cls-Swift 系列模型在参数量和计算量上有所增加,但其推理时间仅略有增加,在保持高效推理性能的同时,显著提升了模型的特征提取和分类能力。这表明 YOLOv8-cls-Swift 算法在实际应用中具有较高的实用价值。

2.5 水稻田杂草精准施药系统及其实际应用

针对水稻田杂草清除这一具体应用场景,本研究通过集成多项高新技术提出一种水稻田杂草清除精准施药系统,实现了从图像采集到杂草精准清除的全流程自动化作业。该系统由 5 个主要模块构成,即图像采集模块、图像处理模块、边缘计算模块、图像识别模块及水稻除草模块,整体结构如图 11 所示。

在图像采集模块中,利用无人机在水稻田上空进行飞行作业,负责大范围水稻田的图像数据采集。无人机携带高分辨率摄像头,能够覆盖广阔的农田区域,收集的图像资料为后续处理提供了基础信息。在图像处理模块中,对接收到的原始图像数据进行预处理,目的是将无人机采集的大图分割成便于处理的小图,为后续的图像识别工作提供清晰、准确的数据支持。在边缘计算模块中,集成了一台 NVIDIA Jetson AGX Orin 设备,该设备拥有强劲的计算性能和低延迟优势,能现场快速处理图像数据,减少对云端传输

的依赖,进一步提高系统的响应速度和处理效率。在图像识别模块中,完成分类模型的配置与加载初始化,采用深度学习技术对处理后的图像进行杂草和水稻的准确识别。识别出的杂草信息将以空间坐标的形式存储,并转换为地理坐标,为后续的除草操作提供精确的位置指导。基于图像识别模块提供的精确位置信息,无人机再次出动,执行除草作业。无人机系统会根据杂草的具体位置规划合理的航线,确保药物能够准确地覆盖到每块土地上,同时最大限度减少对水稻及其他作物的影响。需要注意的是,无人机获取图像后会裁剪成小图,之后将这些小图送入水稻杂草分类算法中预测得到标签,后续小图预测结果被拼接成大图,并得到标签为杂草的小图的地理位置坐标信息,无人机会根据这些地理位置坐标信息执行药物喷洒作业。该水稻田杂草精准施药系统通过各模块间的高效协作,实现了水稻田杂草的精准清除。通过融合无人机、边缘计算、图像处理与识别及精准施药等先进技术,该水稻田杂草精准施药系统不仅提高了农业生产的自动化水平,还有效保障了作物的健康成长,促进了农业的可持续发展。

表 3 YOLOv8-cls 与本研究方法在自建水稻杂草分类数据集上的参数对比

Table 3 Comparison of different parameters between YOLOv8-cls and proposed methods on the self-constructed dataset for rice weed classification

模型	参数量	计算量 (GFLOPs)	图像尺寸 (宽×高)	推理用时 (ms)
YOLOv8n-cls	1 440 850	3.4	224×224	5.3
YOLOv8n-cls-Swift	1 720 322	4.2	224×224	5.4
YOLOv8s-cls	5 083 298	12.6	224×224	5.6
YOLOv8s-cls-Swift	5 934 530	16.4	224×224	5.8
YOLOv8m-cls	15 774 898	41.9	224×224	6.0
YOLOv8m-cls-Swift	23 819 330	50.2	224×224	6.5
YOLOv8l-cls	36 202 306	99.1	224×224	7.3
YOLOv8l-cls-Swift	52 992 610	112.2	224×224	7.8
YOLOv8x-cls	56 144 402	154.3	224×224	8.3
YOLOv8x-cls-Swift	68 455 554	178.4	224×224	8.6

GFLOPs;每 1 s 1×10^9 次的浮点运算。

为了定量地评估减药增效指标,本研究引入一个关键指标即药剂使用量减少率,药剂使用量减少率的计算使用精准施药系统后药剂使用量相对于传统方法的减少比例,药剂使用量减少率等于传统方

法药剂使用量与精准施药系统药剂使用量的差值除以传统方法药剂使用量。本研究以 667 m² 水稻田为试验目标,传统方法药剂使用量大约为 150 g。本精准施药系统药剂使用量大约为 120 g,因此使用本

系统后药剂使用量减少率为 20%。这一结果表明,使用本系统不仅能够有效减少药剂使用量,还能够保持或提高除草效果,实现减药增效的目标。

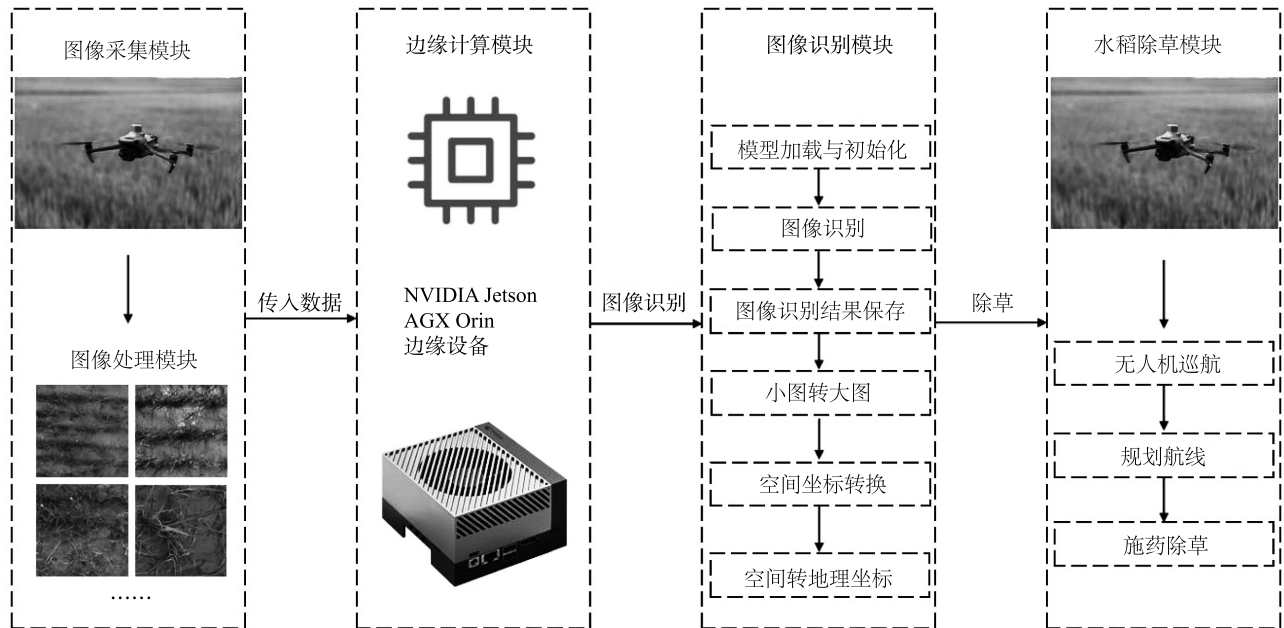


图 11 一种水稻田杂草清除精准施药系统

Fig.11 A precision herbicide application system for weed control in rice fields

3 结论

本研究通过采集稻田图像,构建了一个水稻杂草图像分类数据集,并设计了一种高效的分类算法 YOLOv8-cls-Swift。在特征提取阶段,本研究使用 SwiftFormer 提取图像特征,其结合卷积和自注意力机制全面完整地提取图像的整体信息与目标特征。在分类预测阶段,本研究提出一种高效的加权分类层,引入深度可分离卷积构建多尺度特征提取机制,并通过动态权重分配优化分类过程,显著提升模型对关键目标特征区域的聚焦能力。研究结果表明, YOLOv8-cls-Swift 在图像识别准确率、精确率、召回率等指标上大幅优于基准 YOLOv8-cls 算法。基于此,本研究提出一种清除水稻田杂草的精准施药系统,实现了从图像采集到杂草精准清除的全流程自动化作业,有助于实现水稻生产的减施增效,提升农业生态效益和经济效益,服务于黄河流域生态保护和高质量发展的战略需求。

农作物生长信息的及时获取是实际生产中提升种植管护水平的有效途径,其中,水稻生长状态的精

准监测与杂草的及时处理是生产中面临的迫切需要解决的问题。为实现水稻大面积种植中产量精准预测、生长状态评估及杂草及时处理的目标,建议采用低成本无人机遥感技术和人工智能驱动的水稻生长信息提取模型,对水稻的生长状态和田间杂草情况进行智能监测与分析。

本研究构建了一种全自动水稻田杂草清理系统。该系统以无人机遥感技术和人工智能算法为核心,结合多光谱影像数据和地面巡检信息,实现了水稻田间杂草的精准识别与清除。系统首先通过无人机搭载的多光谱相机获取稻田的高分辨率影像数据,利用 YOLOv8-cls-Swift 算法对影像中的水稻和杂草进行实时检测与分类。此外,该系统可集成地理信息系统技术,将识别结果与田间地理位置信息相结合,生成稻田杂草分布热图,为后续精准除草作业提供可视化指导。同时,系统支持与智能农机设备(如无人除草机)的联动,实现杂草识别与精准除草的全流程自动化。通过构建这一全自动水稻田杂草清理系统,不仅能够显著提高杂草识别的效率和

准确性,还能有效减少除草剂的使用量,降低农业生产成本,推动水稻种植向智能化、精准化和可持续发展方向,助力现代农业的高质量发展。

参考文献:

- [1] 宁运旺,张辉,张永春. 江苏省水稻种植行为的南北差异及其对化肥投入的影响:以苏州、盐城为例[J]. 江苏农业学报, 2018, 34(3): 533-539.
- [2] 王忠培,谢成军,董伟,等. 基于多维注意力机制的水稻病害识别模型[J]. 江苏农业学报, 2024, 40(4): 625-635.
- [3] ALOM M Z, TAHA T M, YAKOPCIC C, et al. The history began from alexnet: a comprehensive survey on deep learning approaches [J]. arXiv Preprint, 2018, DOI: 10.48550/arXiv.1803.01164.
- [4] 苏赋,吕沁,罗仁泽. 基于深度学习的图像分类研究综述[J]. 电信科学, 2019, 35(11): 58-74.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30(1): 5999-6009.
- [7] DOSOVITSKIY A. An image is worth 16x16 words: transformers for image recognition at scale [J]. arXiv Preprint, 2020, DOI: 10.48550/arXiv.2010.11929.
- [8] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision. October 11-17, 2021, Montreal, QC, Canada; IEEE, 2021.
- [9] TOUVRON H, CORD M, JEGOU H. DeiT III: revenge of the ViT [M]. Cham: Springer Nature Switzerland, 2022: 516-533.
- [10] WANG A, CHEN H, LIN Z J, et al. Rep ViT: revisiting mobile CNN from ViT perspective [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 16-22, 2024, Seattle, WA, USA; IEEE, 2024.
- [11] DENG R L, JIANG Y, TAO M, et al. Deep learning-based automatic detection of productive tillers in rice [J]. Computers and Electronics in Agriculture, 2020, 177: 105703.
- [12] 孙俊,谭文军,毛罕平,等. 基于改进卷积神经网络的多种植物叶片病害识别[J]. 农业工程学报, 2017, 33(19): 209-215.
- [13] 邓向武,马旭,齐龙,等. 基于卷积神经网络与迁移学习的水稻苗期杂草识别[J]. 农机化研究, 2021, 43(10): 167-171.
- [14] 桂越,范国华,唐贝贝,等. 基于卷积神经网络的农田杂草与水稻的分类识别[J]. 洛阳理工学院学报(自然科学版), 2019, 29(1): 78-82, 93.
- [15] 杨森森,张昊,兴陆,等. 改进 MobileViT 网络识别轻量化田间杂草[J]. 农业工程学报, 2023, 39(9): 152-160.
- [16] MEHTA S, RASTEGARI M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer [J]. arXiv Preprint, 2021. DOI: 10.48550/arXiv.2110.02178.
- [17] RAHIM A, YUAN F, BARABABY J. An ultralytics YOLOv8-based approach for road detection in snowy environments in the arctic region of norway [J]. Computers, Materials & Continua, 2025, 83(3): 1546-2218. DOI: 10.32604/cmc.2025.061575.
- [18] SHAKER A, MAAZ M, RASHEED H, et al. SwiftFormer: efficient additive attention for transformer-based real-time mobile vision applications [C]//2023 IEEE/CVF International Conference on Computer Vision. October 1-6, 2023, Paris, France; IEEE, 2023.
- [19] NISHIYAMA T, KUMAGAI A, KAMIYA K, et al. SILU: strategy involving large-scale unlabeled logs for improving malware detector [C]//2020 IEEE Symposium on Computers and Communications. July 7-10, 2020, Rennes, France; IEEE, 2020: 1-7.
- [20] FAN Q H, HUANG H B, CHEN M R, et al. RMT: retentive networks meet vision transformers [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 16-22, 2024, Seattle, WA, USA; IEEE, 2024: 5641-5651.
- [21] LI W, WANG X, XIA X, et al. Sepvit: separable vision transformer [J]. arXiv Preprint, 2022. DOI: 10.48550/arXiv.2203.15380.
- [22] GAO Q, HU X Y, HUANG Y H, et al. An improved EfficientFormerV2 network for bone marrow cell classification [C]//2024 5th International Conference on Computer Vision, Image and Deep Learning. April 19-21, 2024, Zhuhai, China; IEEE, 2024: 822-825.
- [23] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. June 20-25, 2009, Miami, FL, USA; IEEE, 2009: 248-255.
- [24] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context [C]//2014 13th European Conference on Computer Vision, September 6-12, 2014, Switzerland, Zurich; Springer, 2014: 740-755.
- [25] ISLAM K. Recent advances in vision transformer: a survey and outlook of recent work [J]. arXiv Preprint, 2022. DOI: 10.48550/arXiv.2203.01536.
- [26] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA; IEEE, 2018: 4510-4520.
- [27] LI Y H, EL HABIB DAHO M, CONZE P H, et al. A review of deep learning-based information fusion techniques for multimodal medical image classification [J]. Computers in Biology and Medicine, 2024, 177: 108635.
- [28] MARMOLEJO-SAUCEDO J A, KOSE U. Numerical grad-cam based explainable convolutional neural network for brain tumor diagnosis [J]. Mobile Networks and Applications, 2024, 29(1): 109-118.

(责任编辑:陈海霞)