齐鲁工业大学学报 第40卷 第4期 2026年8月出版
JOURNAL OF QILU UNIVERSITY OF TECHNOLOGY Vol.40 No.4 Aug.2026
DOI:10.16442/j.cnki.qlgydxxb.2026.04.006 【机电与信息工程】

一种基于 VMamba 的小麦生长期检测模型研究

朱先远[1,2],李秀茹[1]
1.安徽应用技术职业大学 信息与人工智能学院,安徽 芜湖 241000;
2.芜湖市物联网智慧交通工程技术研究中心,安徽 芜湖 241000

摘要:小麦是一种重要的粮食作物,对其生长期等生长信息准确检测是保障品质和产量的重要基础。近年来,计算机视觉等人工智能技术在农业领域得到广泛应用,但传统卷积神经网络(CNN)在复杂光照、背景变化及长时序特征提取方面存在局限。本文以小麦的“发芽”、“分蘖”、“抽穗”和“成熟”4个生长期为研究对象,将新型状态空间建模网络 VMamba 引入小麦生长期检测任务中,构建了一种基于 VMamba 的小麦生长期检测模型,并与 ResNet50、MobileNetV3 和 ViT 等主流模型进行对比实验。实验结果表明,VMamba 模型在准确率、F1 值指标上均优于对比模型,且推理速度、参数量方面也具有优势,整体表现出更优的分类检测性能。此外,通过在 Nvidia Jetson Orin 平台实际部署验证了 VMamba 模型在农业视觉检测的可行性与有效性,为智能化作物监测提供了一种实现思路。

关键词:卷积神经网络;VMamba 模型;生长期检测;计算机视觉;ResNet50
中图分类号:TP391.4 **文献标志码**:A **文章编号**:2097-2792(2026)04-52-10

Research on wheat growth stage detection model based on VMamba

ZHU Xianyuan[1,2], LI Xiuru[1]
1.Anhui University of Applied Sciences, College of Information and Artificial Intelligence, Wuhu 241000, China;
2.Wuhu IoT Intelligent Transportation Engineering Technology Research Center, Wuhu 241000, China

Abstract: Wheat is an important cereal crop, and accurate detection of its growth period and other growth information is an important foundation for ensuring quality and yield. In recent years, artificial intelligence technologies such as computer vision have been widely applied in the field of agriculture, but traditional convolutional neural networks (CNNs) have limitations in extracting global features under complex lighting and background variations. This article takes the four growth stages of wheat, namely “germination,” “tillering,” “heading,” and “maturity,” as the research objects. A new state space modeling network VMamba is introduced into the wheat growth stage detection task, and a wheat growth stage detection model based on VMamba is constructed. Comparative experiments are conducted with mainstream models such as ResNet50, MobileNetV3, and ViT. The experimental results show that the VMamba model outperforms the comparison model in terms of accuracy and F1 score indicators, and also has advantages in inference speed and parameter quantity. Overall, it exhibits better classification and detection performance. In addition, the feasibility and effectiveness of the VMamba model in agricultural visual inspection were verified through actual deployment on the Nvidia Jetson Orin platform, providing

收稿日期:2025-10-31 **网络出版时间**:2026-08-31
基金项目:安徽省教育厅自然科学研究重点项目(2023AH052295);安徽省教育厅物联网应用技术教学创新团队项目(2024CXTD209)
作者简介:朱先远,博士、副教授,研究方向:人工智能、物联网;xyzhu@abc.edu.cn

an implementation approach for intelligent crop monitoring.

Key words: convolutional neural network; VMamba model; growth stage detection; computer vision; ResNet50

小麦生长期检测旨在识别小麦当前所处的生长阶段。小麦处在不同的生长期,需要不同的水分补给、养分补给和病虫害防治。准确检测小麦所处的生长期,并进行合理科学的水分、养分补给和病虫害防治,不仅有助于提高小麦的产量,还能保障籽粒质量。对小麦生长期进行准确的检测研究,对国家的粮食安全及农业经济发展具有重要意义。

现阶段小麦生长期检测主要依赖人工检查,劳动密集且效率低,随着现代化农业人均种植规模的不断扩大,仅依靠人工观测已难以满足现代农业对高精度和高效率的需求[1]。近年来,随着农业劳动力短缺现象加剧,已有越来越多的研究者将人工智能、无人机等技术应用到农作物种植管理等方面。在农作物生长期监测方面,徐建鹏等[2]提出一种基于 Rectified Adam(RAdam)优化器的 ResNet50 卷积神经网络图像识别方法,开展水稻关键生育期的自动识别,正确率达到 97.33%;刘平等[3]提出一种基于综合颜色特征和超像素分割算法的小麦开花期判定方法,小花平均识别精度为 91%,小穗平均识别精度为 90.9%,预测开花率与实际开花率的平均差值仅为 1.16%;赵春江等[4]采用改进的基于特征金字塔花束提取神经网络(Flower Extraction Feature Pyramid Networks,FE-FPN)实现番茄花束的局部区域提取,并采用 Prim 最小生成树对提取的花束区域图像进行识别优先级排序,然后按序将其输入到改进的 Yolov3 网络,实现番茄花朵不同花期的精准辨识检测,平均检测精度达到了 82.79%,平均单张检测时间为 12.54 ms;在玉米生长期识别和品种分类方面,Yu 等[5]使用数码相机对玉米进行监测,提出 AP-HI 算法判断苗期和三叶期,最高精度可以达到 96.68%;韩悦婷等[6]提出了基于轮廓和骨架提取的玉米出苗期和三叶期的自动识别算法,取得较好的识别结果;张芸德等[7]基于深度卷积神经网络特征提取的玉米生长期识别方法,利用深度学习和 SVM 分类建模的方法进行玉米生长期的识别,分类准确率为 94.81%;Zheng 等[8]使用阈值分割提取图片特征,建立 SVDD(Support Vector Data Description)算法实现玉米与杂草的分类识别,最高精度可达 93.87%。针对玉米产后品质检测需求,赵仲文等[9]提出基于改进 SS-YOLOv8 的轻量化鲜食玉米果穗优劣检测模型,在压缩模型体积的同时保持了优异的检测性能,为边缘端玉米品质快速检测提供了可行方案。温长吉等[10]提出一种大规模多类别精细病虫害识别网络模型 PD-Net,通过引入卷积块注意力模型和跨层非局部模块,增强网络对差异化特征的提取和表示能力和在多个特征提取层之间对于多尺度特征的融合,Top1 识别准确率在病害和虫害集上分别达到 88.62%和 74.67%。刘梦姝等[11]给出一种基于 YOLO v8n 改进的小麦病害检测系统,该系统使用 FasterNet 中的 FasterNet Block 替换 C2f 中 Bottleneck 模块,降低模型内存占用量的同时,提高模型特征融合能力和检测精度。张旺等[12]围绕小麦赤霉病检测任务构建了轻量化卷积神经网络改进模型,可实现小麦赤霉病的精准识别与病害程度评估,进一步拓展了轻量化深度模型在小麦病害防控领域的应用场景。针对无人机大田巡检的应用场景,毛锐等[13]提出双流空频协同的 DSFNet 模型,实现了小麦黄矮病的高效检测,为空基平台下小麦病害的大范围快速筛查提供了新的技术路径。

卷积神经网络(Convolutional Neural Network,CNN)能够自动提取图像特征,实现对作物病害、果实成熟度及生长期的分类检测。然而,传统 CNN 由于感受野有限,难以捕获图像的全局特征;同时在复杂田间场景下,光照变化、背景杂乱等因素会影响模型的鲁棒性。为增强模型的全局特征建模能力,研究者提出了 Transformer 结构[14]。其中,Vision Transformer(ViT)[15]通过将输入图像划分为若干固定大小的图像块(Patch),并将这些图像块序列化后输入多头自注意力(Multi-Head Self-Attention)模块,实现了视觉特征的全局建模能力。ViT 模型在大规模数据训练下能显著提升图像分类与检测任务的准确率,成为视觉任务的重要基础结构。然而,ViT 在计算复杂度和显存开销方面仍较高,难以直接应用于边缘设备上的实时检测任务。

VMamba 模型[16]基于二维选择性扫描结构(Selective Scan 2D,SS2D),能够高效捕获图片中长程依赖特征,并在计算复杂度上优于传统 Transformer。该模型在序列建模任务中表现出较强的泛化能力和高效性。本文将 VMamba 网络应用于小麦生长期图像识别任务中,构建了基于 VMamba 的小麦生长期检测模型,并通过对比实验验证其性能优势。

1 相关工作

Transformer 模型,它在 2017 年由 Vaswani 等[14]提出。当时主要用于解决序列建模任务的长距离依赖问题,在处理长序列数据更加高效、准确,在自然语言的机器翻译任务中取得优异效果。其也是 OpenAI 公司的 GPT 等大模型的前身。Transformer 模型是由多组编码器 Encoder、解码器 Decoder 组成。对应到自然语言翻译任务时,Encoder 将输入数据(一个句子)转化为一组特征表示,Decoder 则将这组特征表示转化为输出序列,即转化为另一种语言所表示的句子。Encoder 由多个编码层堆叠而成,每一层都有相似的结构,主要包含输入嵌入层、位置编码、多头自注意力机制、前馈神经网络和归一化层。其中多头自注意力机制是 Encoder 核心。它允许模型同时关注输入序列中不同位置的词。通过这种机制,模型可以捕捉到词与词之间的关系,无论它们之间的距离多远。$\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 矩阵是在这个模块中通过输入生成的,并用于计算词之间的相关性。每个词的自注意力权重基于其与其他词的相似度来分配。自注意力机制的计算如式 (1)所示。

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}, \tag{1}$$

其中,d_k 表示查询矩阵和键矩阵中查询向量和键向量的特征维数。$\sqrt{d_k}$ 为缩放因子,用于避免点积结果随特征维数增大而幅值过大,从而提高模型训练的稳定性。

Decoder 也由多个解码层堆叠而成,每一层的结构和 Encoder 有相似之处。但也有一些区别:其输入嵌入层为目标序列(如翻译任务的目标语言)。自注意力机制采用的为掩码操作,避免看到翻译的尚未被翻译的词。存在一个跨注意力机制 Encoder-Decoder,即通过跨注意力将 Encoder 的输出特征和 Decoder 进行交互,确保 Decoder 可以对整个输入序列进行关注,从输入序列中提取与当前位置相关的上下文信息,帮助完成基于整个输入序列生成当前的输出词。

视觉 Transformer(Vision Transformer,ViT)以图像块(Patch)为基本输入单元,将图像序列化后输入多头自注意力(Multi-head Self-Attention)模块,实现了全局特征建模,能在大规模数据训练下显著提升分类与检测精度,为作物生长监测等任务提供了新的解决思路。然而,ViT 结构计算量大,对显存与算力依赖较强,在农业边缘设备上部署仍面临实时性与资源受限的问题。

为兼顾全局依赖建模与计算效率,VMamba 将 Mamba 的选择性状态空间模型引入视觉任务,并设计二维选择性扫描(2D selective scan,SS2D)机制。SS2D 首先通过交叉扫描将二维特征图沿 4 条互补路径展开为一维序列,各序列分别经过选择性扫描处理,随后通过交叉合并恢复并融合为二维特征图。该机制使每个空间位置能够从不同扫描方向聚合上下文信息,从而建立全局感受野。对于单通道状态空间模型,其离散状态更新过程可表示为(2)所示。

$$\begin{cases}\boldsymbol{h}_p=\bar{\boldsymbol{A}}h_{p-1}+\bar{\boldsymbol{B}}x_p\\ y_p=\boldsymbol{C}_p\boldsymbol{h}_p+Dx_p\end{cases}, \tag{2}$$

其中,p 表示序列位置,x_p、y_p分别表示第 p 个位置的输入和输出,$\boldsymbol{h}_p\in\boldsymbol{R}^N$表示对应的隐藏状态,$N$ 为状态空间维数;$\bar{\boldsymbol{A}}\in\boldsymbol{R}^{N\times N}$和$\bar{\boldsymbol{B}}\in\boldsymbol{R}^{N\times 1}$分别为离散化后的状态转移矩阵和输入矩阵,$\boldsymbol{C}_p\in\boldsymbol{R}^{1\times N}$为输出映射矩阵,$D\in\boldsymbol{R}$ 为输入到输出的直接传递参数,用于保留当前输入信息,可视为状态空间模型中的跳跃连接项。在演化过程中,$\bar{\boldsymbol{A}}$、$\bar{\boldsymbol{B}}$、$\boldsymbol{C}_p$由输入动态生成,使模型能够根据当前特征选择性地保留或遗忘历史信息。由于 SS2D 包含 4 条扫描路径,其主要选择性扫描计算量可近似表示为式(3)所示。

$$\Omega(\mathrm{SS2D})\approx 4HWNC_f, \tag{3}$$

其中,H 和 W 分别为特征图的高和宽,C_f为特征通道数,N 为状态空间维数;当状态维数 N 固定时,SS2D 的计算复杂度随像素数量 $L=HW$ 线性增长。因此,在序列长度较大时,SS2D 的计算复杂度要显著低于自注意

力机制二次复杂度,使得 VMamba 能够在高分辨率输入下高效推理。该机制使 VMamba 在图像分类、目标检测和语义分割等视觉任务中表现出较好的建模能力,并在高分辨率输入条件下具有较好的计算扩展性。

2　基于 VMamba 的小麦生长期检测系统实现

2.1　系统架构设计

本文设计并实现了一套基于 VMamba 网络的小麦生长期检测系统,其总体框架如图 1 所示。系统包括四个主要模块:图像采集与预处理模块、特征提取与推断模块(基于 VMamba 分类器)、可视化与存储和系统运行参数监测模块。

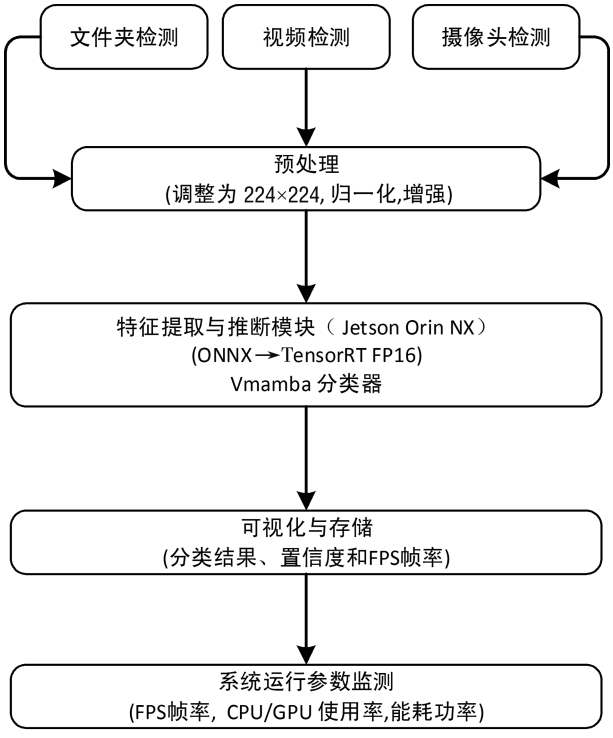


图 1　系统整体框架图

在系统运行过程中,首先通过无人机或田间摄像设备获取小麦不同生长期的图像或视频帧。图像经尺寸统一、归一化及增强处理后,被输入至 VMamba 检测网络中进行深度特征提取与分类推理。网络输出的预测结果包括小麦生长期类别及其置信度。检测结果随后通过 PyQt5 可视化界面进行显示与标注,实现田间环境下的实时检测与监测。

为满足农业场景的实时性与部署需求,系统在模型训练完成后部署至 Jetson Orin NX 等嵌入式设备。通过 ONNX 与 TensorRT 优化,模型推理速度得到显著提升,能够在低功耗条件下实现流畅检测。整体系统结构兼顾了算法精度、工程实用性与边缘计算可行性。

2.2　基于 VMamba 的检测模型

由于 Transformer 自注意力机制在大空间分辨率任务中具有二次复杂度,计算开销随输入尺度急剧增长,从而在高分辨率视觉建模中面临效率瓶颈。针对这一问题,VMamba 模型在视觉领域中引入了二维选择性扫描机制 SS2D,为具备线性复杂度的长序列建模提供了高效方案。本文将 VMamba 模型应用于小麦生长期检测任务,通过 SS2D 机制增强模型的空间结构建模能力,从而实现对小麦生长期图像的高效特征提取与精确阶段识别。

2.2.1　模型总体结构

VMamba 模型的总体结构如图 2 所示,由多个分辨率阶段组成,每个阶段包含一个下采样层(除第一阶

段外)及若干堆叠的视觉状态空间(Visual State Space,VSS)块。输入图像 $I \in R^{H\times W\times 3}$ 被划分为若干 patch 后,经多层变换生成不同分辨率的特征表示 $\frac{H}{8}\times\frac{W}{8}$、$\frac{H}{16}\times\frac{W}{16}$和$\frac{H}{32}\times\frac{W}{32}$。在每个阶段中 VSS 块在 Mamba 模型的 S6 模块基础上引入了二维选择性扫描机制,用于实现动态权重选择与全局感受野建模。同时,VMamba 通过引入二维选择性扫描机制(SS2D),避免了传统 Transformer 自注意力机制的二次计算复杂度,使模型能够在保持长距离依赖建模能力的同时降低计算开销,从而适用于高分辨率视觉任务和边缘设备部署。

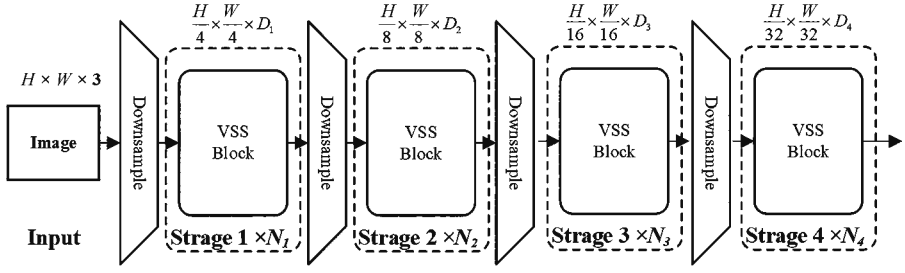


图 2 VMamba 网络结构示意图

2.2.2 二维选择性扫描(SS2D)机制

Mamba 的核心算法——选择性扫描(Selective Scan,S6)最初用于一维序列数据建模,其扫描顺序与自然语言的时间结构高度契合。然而,一个二维图片的特征数据不同于具有前后因果性质的自然语言序列,如果直接将其展平为一维序列利用 SSM 将丢失图片中上下左右间的依赖关系,削弱了模型解释空间关系的能力。VMamba 在通过图片构建序列时,采用 2D 选择扫描模块 SS2D 沿 4 个不同方向进行扫描:从左上到右下,从右下到左上,从右上到左下,从左下到右上,从不同方向扫描收集图片信息建立 4 个序列,最后再将 4 个序列融合为一个序列。SS2D 和 SSM 模块融合后构成 VSS Block 结构。VSS Block 结构如图 3 所示,经过两个线性投影层,输入被分为两个信息流,其中一支信息流在经过深度可分离卷积 DWConv、一个激活函数和一个 SS2D 等模块后,加到另一支信息流上,最后完成 VSS Block 的输出。

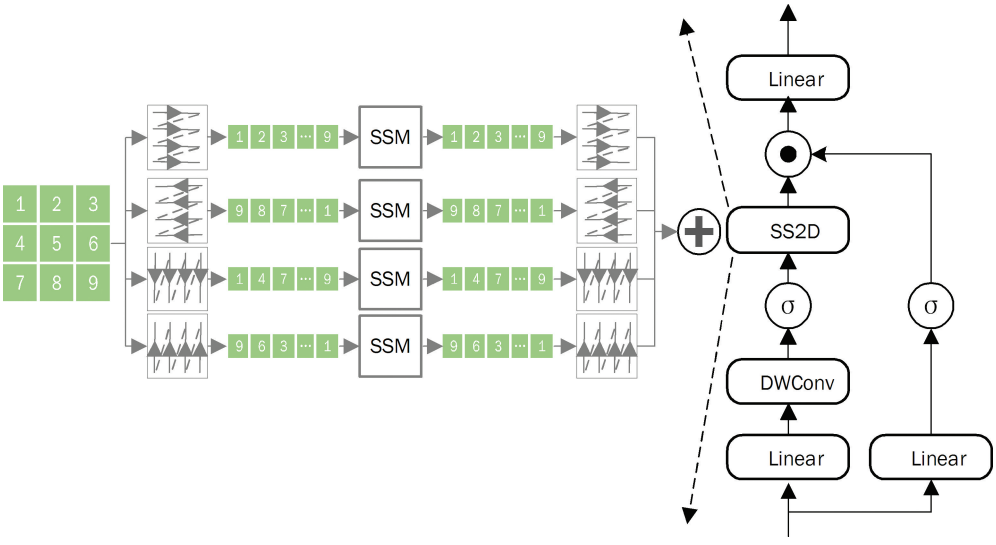


图 3 VSS Block 结构示意图

该机制确保每个像素点能够从多个方向感知全局上下文信息,从而显著提升了模型在复杂背景下对小麦叶片、穗部和株型结构的判别能力。这种结构既能够利用卷积操作提取局部空间特征,又能够通过状态空间模型实现长距离依赖关系建模,相比传统 Transformer 自注意力机制具有更低的计算复杂度,更适合复杂农业场景下的视觉识别任务。

2.2.3 网络优化与参数配置

为提高推理速度与部署效率,本文选用 VMamba-Tiny 结构作为小麦生长期检测的主干网络。模型训练

在 Ubuntu 22.04 系统下使用 PyTorch 框架完成,GPU 为 NVIDIA RTX 4090。训练过程中采用 AdamW 优化器,初始学习率 1×10^{-4},batch size 为 16,权重衰减系数 0.05,学习率采用余弦退火策略,训练轮数 100 个 epoch。损失函数为交叉熵损失(Cross-Entropy Loss)。

为提升训练效率与稳定性,训练过程中启用了 PyTorch 的自动混合精度(AMP)机制,实现显存节省与加速收敛。每个 epoch 结束后计算训练与验证集的 Top-1 精度与宏平均 F1 值,保存验证精度最高的模型作为最终权重。若连续 10 个 epoch 未出现性能提升,则提前终止训练以避免过拟合。训练完成的最终 Mamba 模型被导出为 ONNX 格式,并利用 NVIDIA TensorRT 进行推理优化。采用 FP16 精度模式以降低延迟与内存占用,实现了高吞吐、低延时的推理效果。

3　实验与分析

3.1　数据集构建与预处理

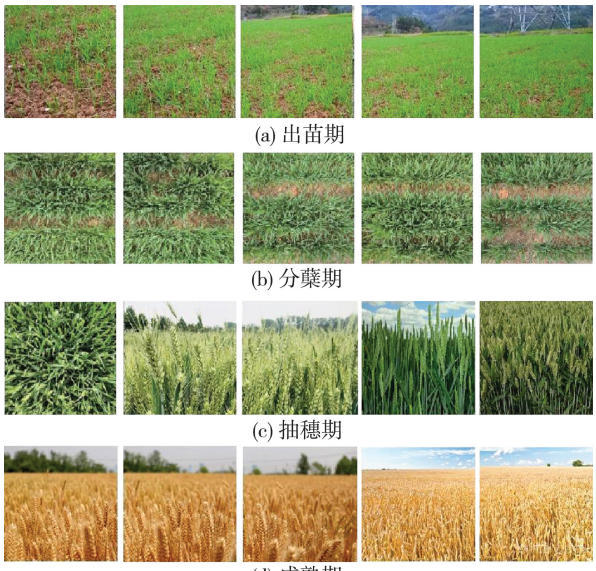
小麦的生长期划分根据 BBCH(Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie)标准,可以被划分为出苗、叶片发育、分蘖、茎秆伸长、拔节、抽穗、开花、果实发育和成熟阶段。本研究选择出苗期、分蘖期、抽穗期和成熟期作为检测阶段。出苗期特征为植株稀疏、幼苗叶片微弱。分蘖期特征为植株高度较低,但密度增大,叶片较密集。抽穗期特征为穗部明显可见,且穗部长度和密度成为主要判别点。成熟期特征表现为穗部干枯发黄,整体颜色偏棕黄。这些阶段小麦在种植管理方式方法差异比较大,是影响水肥管理的关键时期,具备分类识别价值。

现有公共小麦数据集,如全球小麦头检测(GWHD)数据集,包括 4 700 张高分辨率 RGB 图像和约 190 000个标记的小麦头,主要从抽穗到成熟,不包括出苗和分蘖等早期阶段。2021 年全球小麦头数据集同样只提供了小麦头(抽穗/成熟)的边界框,也不包含早期阶段小麦数据。现有的公共数据集无法满足小麦 4 个生长期检测任务的需要。本项目研究团队通过现场拍摄、在线下载和数据增强等方式建立了小麦 4 个生长期的数据集。数据增强包括随机旋转(±15°)、水平翻转(概率 0.5)、亮度调整(±20%)、对比度调整(±15%)及色调抖动(±10%)等操作,以提升模型对光照和视角变化的鲁棒性。数据集中的详细数据来源如表 1 所示。共收集了 9 871 张原始图片,其中现场拍摄田间样本 3 871 张(出苗期、分蘖期和抽穗期),在线下载分蘖期、抽穗期和成熟期照片 6 000 张。经数据增强处理后,每个生长期的图片扩充至 3 000 张。图 4 为数据集中不同生长期的样本。

表 1　小麦生长期数据集来源表

生长期	现场拍摄数量	在线下载数量	数据增强后数量
出苗期	3 000	0	3 000
分蘖期	395	2 000	3 000
抽穗期	476	2 000	3 000
成熟期	0	2 000	3 000
总计	3 871	6 000	12 000

数据集中田间图片的摄影地点位于安徽省芜湖市弋江区白马街道新大埠村(北纬 31.221°,东经 118.657°),其采集时间从 2024 年 11 月至 2025 年 4 月。拍摄环境考虑了一天中不同的时间段和不同的光照条件,以及晴天、阴天和雨天等情况,确保模型训练的泛化能力。数据标注由两名有农学背景的标注员独立完成,对两者存在分歧的样本,由第三位专家复核并结合 BBCH 标准最终确定分类,以确保对样本图片分类的准确性。数据集按照 8∶1∶1 被随机分为训练集、验证集和测试集。



(a) 出苗期

(b) 分蘖期

(c) 抽穗期

(d) 成熟期

图 4　数据集中 4 个生长期的小麦样本图片

3.2　小麦生长期数据集上对比试验

为了验证 VMamba 网络结构对小麦生长期的检测效果,本文将在自建小麦数据集上进行对比实验。实验在 PyTorch 框架下进行,硬件环境为 NVIDIA RTX 4090 GPU,输入图片尺寸统一为 224×224 像素。所有模型均采用相同的训练策略(训练 100 个 epoch,初始学习率 1×10^{-4},AdamW 优化器,采用 Cosine Annealing 调度策略),表 2 显示了所有实验中一致使用的核心超参数。模型性能通过分类准确率 Top-1 Acc、F1-score 和参数量等指标进行评估。比较模型包括:ResNet-50、MobileNetV3 和 Vision Transformer(ViT)。实验时采用 VMamba-tiny 作为主干网络,模型包含 4 个 stage,Patch 大小为 4×4,所有模型均采用 ImageNet 预训练权重,并在小麦数据集上进行统一微调。不同网络模型在小麦生长期分类检测实验结果如表 3 所示。

表 2　超参数表

超参数	参数值
Image Size	224×224
Batch Size	16
Optimizer	AdamW
Learning Rate	1×10^{-4}
Weight Decay	0.05
Learning Rate Scheduler	Cosine Annealing
Number of Epochs	100
Loss Function	Cross Entropy Loss

表 3　不同网络模型在小麦生长期分类任务上的性能对比

模型	Top-1 精度/%	F1 值	参数量
VMamba-tiny	98.4	0.984	22.9M
ResNet-50	94.6	0.945	25.6M
MobileNetV3	88.7	0.887	5.5M
ViT	97.5	0.971	86M

从表 3 可以看出,VMamba-tiny 模型在准确率(Top-1 精度 98.4%)方面取得最佳表现。在 F1 值 VMamba-tiny 也优于其他 3 种对比模型。在参数量方面,VMamba-tiny 的参数量为 22.9M,与 ResNet-50 相接近;MobileNetV3 虽然轻量级,但可能由于其简单的架构导致其准确性较低;ViT 的准确率与 VMamba-tiny 接

近,但参数数量太大。总体而言,VMamba-tiny 在对小麦生长期检测上的优势明显。

为进一步评估模型在小麦生长期上的分类性能,本文根据各个模型在小麦各个生长期检测的预测值与真实值的实验结果构建了混淆矩阵,图 5~8 分别为 ResNet-50、MobileNetV3、ViT 和 Vmamba-tiny 在小麦生长期分类任务中的混淆矩阵。混淆矩阵的横轴为真实类别,纵轴为预测类别。

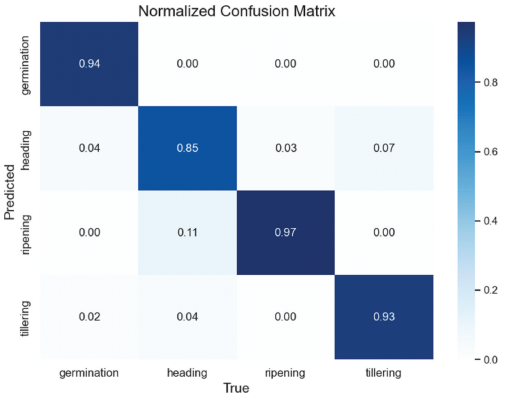


图 5　ResNet50 在小麦生长期分类任务中的混淆矩阵

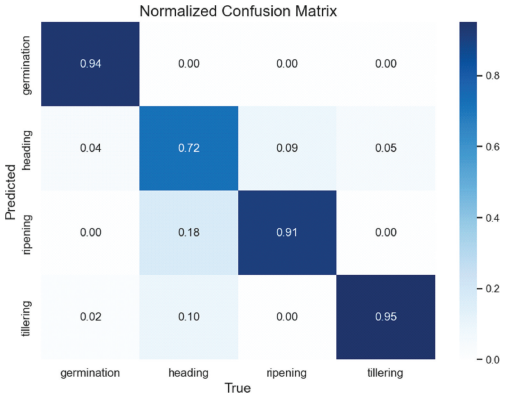


图 6　MobileNetV3 在小麦生长期分类任务中的混淆矩阵

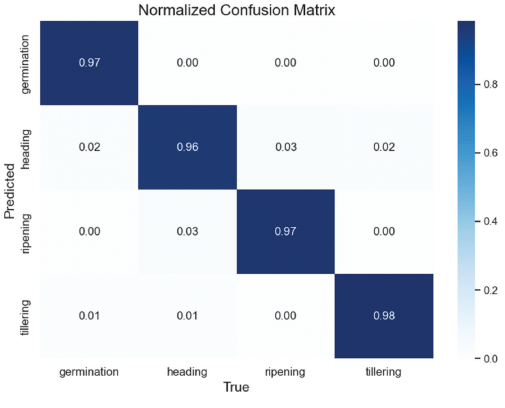


图 7　ViT 在小麦生长期分类任务中的混淆矩阵

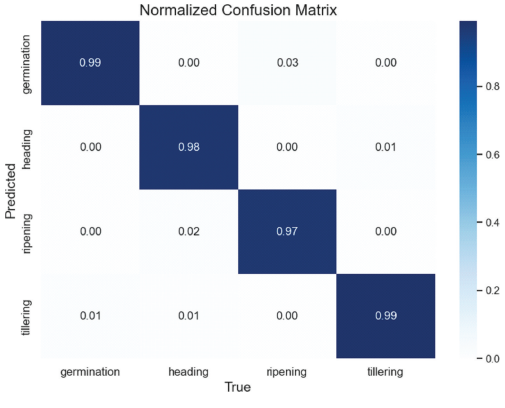


图 8　VMamba-tiny 在小麦生长期分类任务中的混淆矩阵

从图中可以看出,ResNet50 对标记为 heading 抽穗期的小麦错误预测为成熟期 ripening 比例为 11%,对标记为 tillerting 分蘖期的小麦错误预测为抽穗期 heading 比例为 7%;MobileNetV3 对标记为 heading 抽穗期的小麦错误预测为成熟期 ripening 比例为 18%,对标记为 tillerting 分蘖期的小麦错误预测为抽穗期 heading 比例为 5%;ViT 对标记为 heading 抽穗期的小麦错误预测为成熟期 ripening 比例为 3%,对分蘖期、出苗期的小麦错误预测为抽穗期比例为 2%;对于分蘖期和抽穗初期的小麦,由于两个阶段小麦的生长特征相似度较高,存在一定程度的混淆,导致如 MobileNetV3 和 ResNet50 等卷积神经网络检测失效。相比于 ResNet50 和 MobileNetV3,Vmamba-tiny 和 ViT 在分类预测上表现出了较低的误判率,特别是 Vmamba-tiny 在对小麦 4 个生长期误判率极低。究其原因,VMamba-tiny 能够在不同尺度下自适应捕捉图像中的局部与全局特征,有效弥补了传统 CNN 在长距离依赖建模上的局限,使模型在处理具有细粒度差异的图像(如分蘖期与抽穗期小麦)时具备更强的判别能力。因此,VMamba-tiny 在小麦图像特征中可以更好地区分边缘形态、穗部轮廓、叶片纹理等微小差异,从而显著降低了阶段间的误判率。

3.3　边缘设备部署测试

为验证本文给出的小麦生长期检测系统在实际应用场景中的可用性,本文将系统部署在 Jetson Orin NX 16 GB 边缘计算平台,并从图像、视频和摄像头 3 种输入模式出发,围绕系统运行过程中的吞吐量、检测精度、资源占用率及功耗等指标,开展定量分析实验。

为在 Jetson 平台上实现高效部署推理,实验中首先将训练得到的权重文件导出为 ONNX 格式模型,并进一步通过 TensorRT 优化生成 FP16 精度的.engine 文件,从而在边缘设备上实现高性能低延迟的推理运行。

系统部署于 NVIDIA Jetson Orin NX16G 开发板上,该平台集成了基于 Ampere 架构的 GPU、DLA 加速单元和多核 ARM CPU,可为边缘智能场景提供强大算力支持。系统运行于 Ubuntu 20.04 操作系统,基于 NVIDIA JetPack SDK 环境进行开发与部署。系统推理框架使用 PyTorch 结合 TensorRT 运行,以充分发挥硬件推理加速优势。图 9~12 分别为系统主界面、图片检测、视频检测和实时摄像头检测的系统截图。



图 9 系统主界面



图 10 图片检测



图 11 视频检测



图 12 实时摄像头检测

在实际部署监控中,由于实时采集中的噪声、视频压缩伪影、硬件解码延迟和 DVFS 动态频率调整等各种实际因素,与离线测试集的 98.4%相比,3 个输入源的端到端检测精度都会发生偏差。从图 10~12 可以发现,图片检测模式置信度为 97.0%,视频检测模式降至 88.0%,实时摄像头检测模式降至 95.0%。采用视频检测和实时摄像头检测由于 I/O 传输开销等因素 FPS 值基本可以保持 12 以上,FPS 略低于图片检测模式,但仍能保持稳定的实时性能。总体来说,在 TensorRT 的 FP16 加速下,VMamba-tiny 在 Jetson Orin NX 上推理帧率和检测精度满足实时农业检测需求,具备部署在农业现场的实用性。

4 结 论

本文给出了一种基于 VMamba 小麦生长期检测方法。通过在小麦生长期数据集上的对比实验,可以发现 VMamba 模型相比于常见的 ResNet、MobileNetV3 和 ViT 等网络模型,其在小麦生长期分类检测任务中取得了最高的准确率,对于小麦生长特征较相似的分蘖期、抽穗期也能保持较高的分类准确率和极低的错判率。在边缘部署方面,本文将 VMamba 模型转换为 TensorRT 引擎并在 Jetson Orin NX 上运行,系统测试结果表明该模型具有卓越的实时时性和低资源占用,能够在现场为农田监测提供即时、准确的生长期判断。本研究表明,基于 VMamba 的小麦生长期检测方法为智慧农业场景下的作物生长检测提供了一种高效可行的技术方案。

本研究未来将扩大数据集规模,增加更多区域、更多品种的农作物生长期样本,以增强模型的鲁棒性和泛化能力,以推广到其他作物(如水稻、玉米等)的生长期检测任务中,验证模型的跨作物通用性和应用前

景,为精准农业生产提供了重要技术支持。

参考文献:

[1] 扶兰兰,黄昊,王恒,等.基于 Swin Transformer 模型的玉米生长期分类[J].农业工程学报,2022,38(14):191-200.

[2] 徐建鹏,王杰,徐祥,等.基于 RAdam 卷积神经网络的水稻生育期图像识别[J].农业工程学报,2021,37(8):143-150.

[3] 刘平,刘立鹏,王春颖,等.基于机器视觉的田间小麦开花期判定方法[J].农业机械学报,2022,53(3):251-258.

[4] 赵春江,文朝武,林森,等.基于级联卷积神经网络的番茄花期识别检测方法[J].农业工程学报,2020,36(24):143-152.

[5] Yu Zhenghong, Cao Zhiguo, Wu Xi, et al. Automatic image-based detection technology for two critical growth stages of maize: Emergence and three-leaf stage[J]. Agricultural and Forest Meteorology, 2013, 174/175: 65-84.

[6] 韩悦婷,行鸿彦,金红伟.基于 openCV 的玉米出苗期和三叶期自动检测系统的设计[J].电子测量与仪器学报,2017,31(10):1574-1581.

[7] 张芸德,刘蓉,刘明,等.基于深度卷积特征的玉米生长期识别[J].电子测量技术,2018,41(16):79-84.

[8] Zheng Yang, Zhu Qibing, Huang Min, et al. Maize and weed classification using color indices with support vector data description in outdoor fields[J]. Computers and Electronics in Agriculture, 2017, 141: 215-222.

[9] 赵仲文,张永立,韩镇宇,等.基于改进的 SS-YOLOv8 轻量化鲜食玉米果穗优劣检测模型 [J].农业工程学报,2025,41(11):183-192.

[10] 温长吉,王启锐,陈洪锐,等.面向大规模多类别的病虫害识别模型[J].农业工程学报,2022,38(8):169-177.

[11] 刘梦姝,张春琪,晁金阳,等.基于 YOLO v8n 改进的小麦病害检测系统[J].农业机械学报,2024,55(S1):280-287.

[12] Zhang Wang, Ren Ri, Guo Zidi, et al. Improved lightweight convolutional neural network models for the detection and evaluation of Fusarium head blight in wheat[J]. Journal of Zhejiang University-SCIENCE B, 2026, 27(5): 450-465.

[13] 毛锐,郝思嘉,李书琴,等.DSFNet:双流空频协同的无人机小麦黄矮病高效检测模型 [J].农业工程学报,2026,42 (6):197-205.

[14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017, 30: 1-11.

[15] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[PP/OL]. V2. arXiv (2021-06-03)[2025-10-15]. https://iclr.cc/virtual/2021/poster/3013.

[16] Liu Yue, Tian Yunjie, Zhao Yuzhong, et al. VMamba: Visual state space model[C]//Advances in Neural Information Processing Systems 37. Vancouver, BC, Canada. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024: 103031-103063.

(责任编辑:王亚宾;校　对:孙洪清)