doi:10.6041/j.issn.1000-1298.2026.04.027

基于多尺度注意力视觉 Mamba U－Net 的耕地遥感分割方法

侯新刚　王　勤　令伟锋
（西京学院会计学院，西安 710123）

摘要：耕地遥感影像的准确分割对产量预测、农业经营和国家粮食安全至关重要。由于遥感农田图像分辨率高、尺寸大、种类多、边界不规则、背景复杂等特点，以及遥感图像分割中广泛应用的卷积神经网络和 Transformer 存在难以提取远程依赖关系和计算复杂度高等局限性，使得农田遥感图像分割研究仍具有一定挑战性。针对当前耕地遥感分割任务中存在的边界模糊、地类混杂等问题，本文提出一种新型多尺度注意力视觉 Mamba U－Net（MSAVM－UNet）模型。该模型通过 3 个模块实现性能突破：首先，改进视觉状态空间模块采用双向选择性扫描机制，在保持线性计算复杂度的同时实现长程依赖建模；其次，通道感知注意力状态空间模块通过动态光谱-空间特征重标定，有效提升耕地与背景地物的区分度；最后，构建多尺度跨层级特征金字塔特征聚合模块，实现多粒度信息融合。在公开耕地数据集的试验表明，MSAVM－UNet 在分割精度和计算效率方面均显著优于现有方法，平均分割精度和相似系数分别达到 85.60% 和 84.46%。研究结果为智慧农业耕地精准监测提供了可靠技术支撑。

关键词：耕地遥感图像分割；通道感知注意力视觉状态空间；多尺度注意力聚合；MSAVM－UNet
中图分类号：TP391.41　　**文献标识码**：A　　**文章编号**：1000-1298(2026)04-0279-08　　**OSID**：

Remote Sensing Segmentation of Cultivated Land Based on Multi-scale Attention Vision Mamba U－Net

HOU Xin'gang　WANG Qin　LING Weifeng
(*School of Accounting*, *Xijing University*, *Xi'an* 710123, *China*)

Abstract: Accurate remote sensing image segmentation of cultivated land (CLRSIS) is crucial for yield prediction, agricultural management, and national food security. However, it remains challenging due to the high-resolution, large size and various remote sensing farmland images with irregularly boundaries and complex background. Convolutional neural networks(CNNs) and Transformers have been widely applied to RSI segmentation, but both of them have limited ability to handle long-range dependencies because of inherent locality or computational complexity. Aiming at the limitation of CNNs and Transformers, and the technical difficulties in CLRSIS, a multi-scale attention visual Mamba U－Net (MSAVM－UNet) model for CLRSIS was proposed. This model achieved performance breakthroughs through three innovative modules: firstly, modified visual state space module (MVSS) adopted a bidirectional selective scanning mechanism, enabling long-range dependency modeling while maintaining linear computational complexity. Secondly, channel-aware attention visual state-space (CAAVSS) effectively enhanced the discrimination between cultivated land and background features through dynamic spectral-spatial feature recalibration. Finally, multi-scale feature aggregation module (MSAA) built a cross-level feature pyramid to achieve multi-granularity information fusion. Experiments on public cultivated land datasets showed that this method was significantly superior to existing methods in terms of segmentation accuracy and computational efficiency, with the average segmentation precision accuracy and DSC achieving 85.60% and 84.46%, respectively. The research result can provide reliable technical support for the precise monitoring of cultivated land in smart agriculture.

Key words: remote sensing image segmentation of cultivated land; channel-aware attention VSS; multi-scale attention aggregation; multi-scale attention Mamba U－Net

收稿日期：2025－08－07　修回日期：2025－08－29
基金项目：国家自然科学基金项目(62172338)、陕西省科技厅重点研发计划项目(2025NC－YBXM－216)和陕西省教育厅重点研发项目(24JR158)
作者简介：侯新刚(1981—)，男，副教授，博士，主要从事农业物联网和农业数字化研究，E-mail：houxingang@xijing.edu.cn

0 引言

耕地作为农业生产的基础资源,耕地资源信息快速、精准提取是农业资源调查、产量估算与灾害监测的核心前提,其空间分布是粮食安全评价和土地资源管理等领域的主要依据[1-2]。随着城镇化进程加速、住区扩张、林地侵占和交通建设等导致耕地边界日益复杂化与破碎化。这一问题在西南山区尤为突出,狭窄条带状分布、零散化耕作模式及梯田式地形结构,使得传统面向对象分类的耕地信息提取方法难以准确刻画农田边界细节[3]。尽管遥感探测影像(RSI)分割技术已成为耕地信息提取的主流方法,但在实际应用中仍面临显著挑战[4-5]。高分辨率 RSI 数据量大、特征复杂,而且大尺度影像中的耕地具有复杂的相位和特征。近年来,RSI 与深度学习(Deep Learning, DL)相结合显著提升耕地遥感图像信息提取(Cultivated land remote sensing image information extraction, CLRSIS)精度[6-8]。张新长等[9]梳理和分析了传统耕地提取算法和基于 DL 的智能化耕地提取算法研究现状,阐述了 DL 支持下 CLRSIS 的基本原理、方法以及实验流程,并归纳了主要的智能耕地提取算法。魏青迪等[10]根据华北地区耕地具体特点,提出了一种耕地图像语义分割模型,形成了一种提取耕地信息的网络结构。李倩楠等[11]探讨金字塔场景解析模型(Pyramid scene parsing network, PSPNet)和 U-Net 在大规模耕地自动提取中的适用性,并与传统的面向对象方法进行了比较。结果表明,两模型在不同区域、不同时相影像中能得到较好的耕地提取效果,具有较强泛化能力和时空迁移能力。

以上 DL 模型能有效提取 RSI 局部特征,但难以提取 RSI 的长距离依赖关系和细化特征,不能准确表示农田地块。为解决传统 U-Net 在长程依赖建模上的局限性,CAO 等[12]将 Transformer 与 U-Net 相结合的混合架构(TransUNet),通过窗口移位注意力机制在图像分割任务中提升平均并交比达到 12.7%。为实现耕地 RSI 边缘特征和深度区域特征间的信息互补,董张玉等[13]构建了一种由上下文 Transformer、门控卷积、空间通道隔离与激励和注意机制组成的边缘分支子网络,提高了模型对边界信息处理的性能,该网络比 YOLO 和 U-Net 更有效地捕获全局依赖性。BO 等[14]根据农田尺度特征,提出了一种基于多频掩码 Transformer(Multi-frequency mask Transformer, MFMTrans)的农田提取方法,在科大讯飞挑战赛 2021 耕地提取大赛数据集的结果表明,该网络能有效缓解端到端实例分割网络在密集场景下造成的重叠问题。

尽管 Transformer 和视觉 Transformer(ViTs)能有效捕获全局依赖关系,在各种计算机视觉任务中表现出强大性能,但由于其自注意力机制的二次方计算复杂度,特别是在高分辨率和实时应用中,面临计算效率的挑战[15]。Mamba 是一种新兴的复杂图像与检测分割技术,能够捕获长距离依赖关系和提取全局上下文信息,且具有线性复杂度[16]。视觉 Mamba(Visual Mamba, VMamba)通过将状态空间模型(State space model, SSM)与注意机制相结合,在视频理解和遥感目标检测任务中取得了显著效果[17]。U-Net 和 Mamba 相结合模型 MambaUNet 结合 U-Net 和 Mamba 优点,采用 U-Net 的编码器-解码器体系结构,并利用跳转连接保留不同大小网络的空间信息,能捕获医学图像复杂细节和更广泛的语义上下文[18]。LIAO 等[19]提出了一种轻量化 Mamba UNet (LightM-UNet),取得了显著的图像分割效果。与 YOLO、U-Net、ViTs 和最新的混合 DL 模型相比,MambaUNet 和 VMamba 表现出很大应用潜力。

针对耕地 RSI 中存在光照变化、遮挡、边界模糊、小耕地块密集等使耕地信息难检测问题,在 MambaUNet 和 VMamba 启发下,本文提出一种基于多尺度注意力 VM-UNet 的耕地 RSI 语义分割方法(Multi-scale attention VM-UNet, MSAVM-UNet),并在公开耕地 RSI 数据集上验证其可行性。

1 多尺度注意力 VM-UNet

利用 VM-UNet 的线性复杂性和全局建模能力,构建多尺度注意力 VM-UNet 模型(MSAVM-UNet),其架构如图 1 所示,图中 H 为 RSI 高度,W 为 RSI 宽度。该模型由编码器、解码器、通道感知注意力 VSS(Channel-aware attention VSS, CAAVSS)模块和多尺度注意力聚合(MSAA)模块组成。

1.1 编码器和解码器

在 MSAVM-UNet 中,编码器由 1 个块划分与线性嵌入模块、4 个改进 VSS(MVSS)模块和 3 个块合并模块组成,解码器由 3 个 MVSS 模块和 4 个块扩展模块组成。设输入 2D 耕地 RSI 尺寸为 $H\times W\times 3$。与 ViT 和 VMamba 类似,首先将输入 RSI 分割为 16 个尺寸为 4×4 不重叠小块,并将其转换成尺寸为 $H/4\times W/4\times 16$ 的一维序列。线性嵌入层将特征维度调整到 C,C 为通道数。编码器各层输出分辨率分别为 $H/4\times W/4\times C$、$H/8\times W/8\times 2C$、$H/16\times W/16\times 4C$ 和 $H/32\times W/32\times 8C$。与编码器类似,解码器利用 2 个连续的 MVSS 模块进行特征重建,使

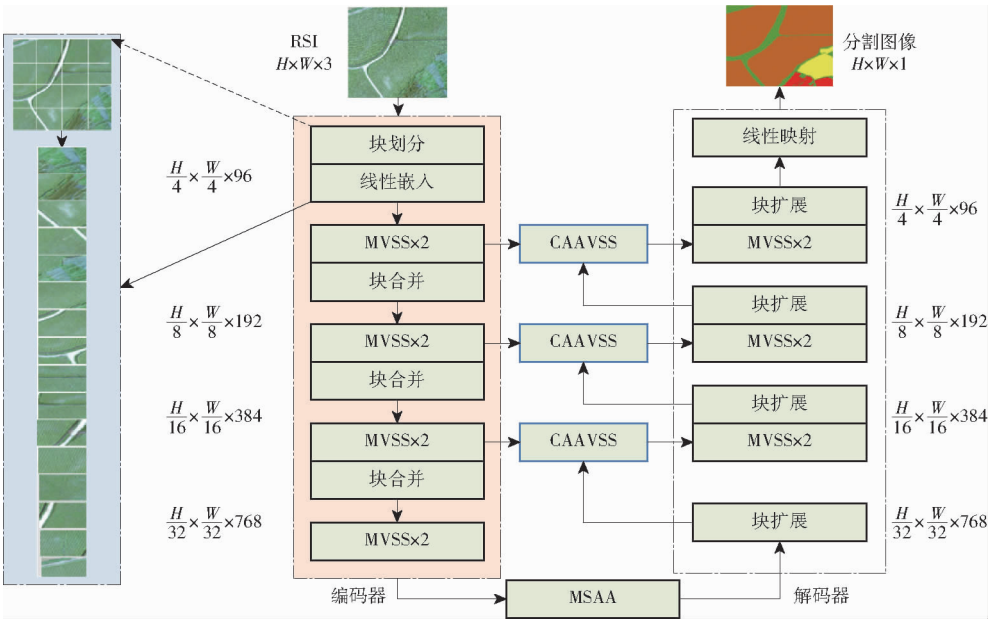

图 1　MSAVM－UNet 架构

Fig. 1　Architecture of MSAVM－UNet

用块扩展模块对深度特征进行上采样。这些模块提高了分辨率(2 倍升级),同时将特征尺寸减半 1/2 以提高分辨率,块合并层通过将相邻 2×2 图像块的特征合并成线性投影进行下采样,降低空间分辨率($H\times W\rightarrow H/2\times W/2$),同时增加通道维数($C\rightarrow 2C$),实现高效的多尺度特征提取。与传统的池化和上采样操作相比,块合并和块扩展操作可以提高 CLRSIS 检测精度,同时减少计算量。MSAVM－UNet 主要组成部分如图 2 所示,图中 C_1 为特征图通道数,C_2 为空洞多尺度卷积输入通道数。

(1)MVSS 结构如图 2a 所示。由层归一化、VSS 模块、映射层、残差连接、逐元素相加和缩放组成。其过程描述为

$$F_{\mathrm{MVSS}}=\mathrm{linear}(\mathrm{LN}(\alpha F_{\mathrm{in}}+F_{\mathrm{VSS}})) \tag{1}$$

其中

$$F_{\mathrm{VSS}}=\mathrm{VSS}(\mathrm{LN}(F_{\mathrm{in}})) \tag{2}$$

式中　F_{in}、F_{MVSS}——MVSS 输入和输出

α——残差连接的缩放因子,取 0.4

linear(·)、LN(·)和 VSS(·)为 MVSS 线性化、层归一化和 VSS 操作。

在 VSS 中,2D 选择性扫描模块(2D selective scanning, SS2D)由 3 个关键组件组成:扫描展开、S6 分块和扫描合并,其中扫描展开将输入图像沿上、下、左、右 4 个方向分解为独立序列,保证特征的空间覆盖并捕获多方向特征。其中 S6 分块在捕获序列的时间动态方面起着关键作用,其利用选择性机制准确识别和提取有用信息,同时过滤掉无关部分。

在 MVSS 中,块合并和块扩展模块将输入图像分割成互不重叠的块,通过线性投影压缩空间维度,并利用 SSM 保持全局依赖关系,避免了传统池化的信息丢失。块扩展作为一种转置卷积,在扩展过程中引入跳跃连接的多尺度 SSM 状态,以提高小块耕地分割精度。

(2)MSAA 结构如图 2b 所示,主要由 3 个模块组成:多尺度空间注意力、通道注意力和残差连接。其中,多尺度空间注意力的空洞多尺度卷积用于提取多尺度特征(图 2c),并拼接通道维度,形成更丰富的特征表示。对空洞多尺度卷积的特征图进行全局池化,再通过 7×7 卷积生成空间注意力特征图,然后逐元素与初始特征图相乘,增强显著空间区域特征;通道注意力对特征图进行全局池化,通过 2 次 1×1 卷积生成通道注意图,并与空间注意力输出逐元素相乘,增强重要通道特征;将其添加到输入特征映射中,在保留原始信息的同时增强特征表示。MSAA 输入为编码器特征的整合,表示为

$$F_{\mathrm{In}}=\mathrm{Concaten}(F_{r=1}(F''_{\mathrm{in}}),F_{r=2}(F''_{\mathrm{in}}),F_{r=3}(F''_{\mathrm{in}}),F_{r=5}(F''_{\mathrm{in}}))+F''_{\mathrm{in}} \tag{3}$$

其中

$$F''_{\mathrm{in}}=\mathrm{Conv}_{1\times 1}(F'_{\mathrm{in}}) \tag{4}$$

式中　F'_{in}——空洞多尺度卷积输入

$F_{r=i}(F''_{\mathrm{in}})$——空洞率为 $r=i$ 得到的卷积特征($i=1,2,3,5$)

MSAA 空间注意力表示为

$$F_{\mathrm{Spatial}}=\mathrm{Sigmoid}(\mathrm{Conv}_{7\times 7}(\mathrm{Maxpooling}(F_{\mathrm{In}}))) \tag{5}$$

式中　Sigmoid(·)——Sigmoid 激活

Maxpooling(·)——最大池化

MSAA 通道注意力表示为

$$F_{\mathrm{Channel}}=\mathrm{Conv}_{1\times 1}(\mathrm{ReLU}(\mathrm{Conv}_{1\times 1}(\mathrm{Avgpooling}(F_{\mathrm{In}})))) \tag{6}$$

式中　ReLU(·)——ReLU 激活

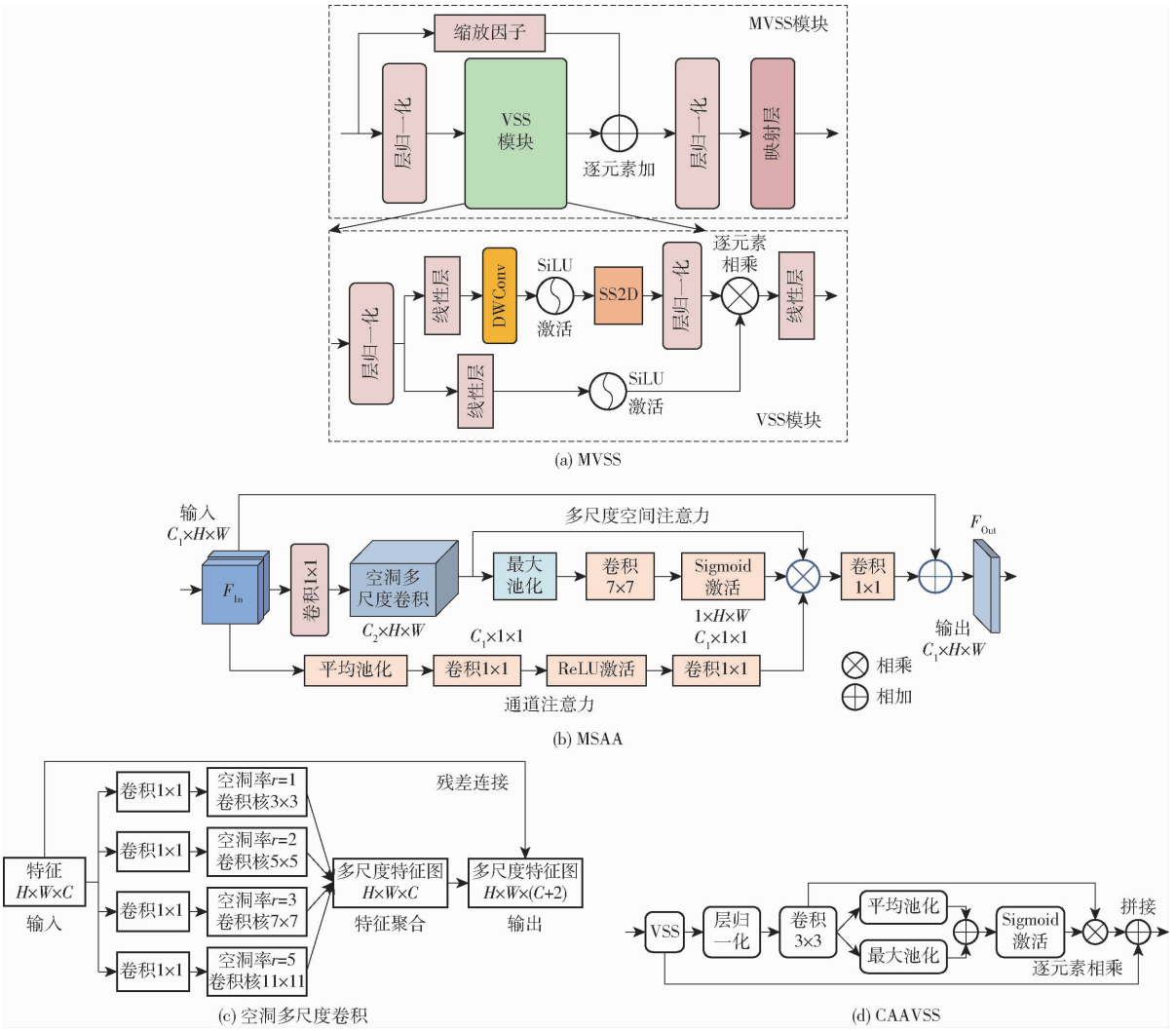

图 2 MVSS、MSAA、空洞多尺度卷积和 CAAVSS 模块

Fig. 2 MVSS, MSAA, void multi-scale convolution, and CAAVSS modules

MSAA 输出 F_{Out} 表示为

$$F_{Out}=\mathrm{Conv}_{1\times1}(F_{Spatial}\otimes F_{Channel}\otimes F_{In})\oplus F_{in} \quad (7)$$

MSAA 的 3 路径特征聚合操作能增强模型在空间域和通道域的表征能力,提高 CLRSIS 准确性。

(3)CAAVSS 结构如图 2d 所示,主要由 VSS、层归一化、卷积、平均池化与最大池化、Sigmoid 激活和 2 个残差连接组成。其过程描述为

$$F_{Spatial}=\mathrm{Sigmoid}(\mathrm{Avgpooling}(F_v)\oplus \mathrm{Maxpooling}(F_v))\otimes F_v\oplus \mathrm{VSS}(F_{in}^{CA}) \quad (8)$$

其中

$$F_v=\mathrm{Conv}_{3\times3}(\mathrm{LN}(\mathrm{VSS}(F_{in}^{CA}))) \quad (9)$$

式中 F_{in}^{CA}——CAAVSS 输入

1.2 损失函数

在训练阶段,采用二元两类交叉熵作为目标函数。假设 Pre_i 表示第 i 个模型预测图像,GT_i 表示第 i 个 RSI 的 Ground Truth,则二值交叉熵损失表示为

$$Loss=Loss(Pre_i,GT_i)=-\sum_{i=1}^{W}\sum_{j=1}^{H}(GT_{i,j}\ln Pre_{i,j})+(1-GT_{i,j})\ln(1-Pre_{i,j}) \quad (10)$$

式中 $GT_{i,j}$——Ground Truth 在第 i 个和第 j 个位置像素值

$Pre_{i,j}$——模型预测 RSI 的第 i 个和第 j 个位置像素值

2 试验结果与分析

2.1 数据来源与预处理

耕地 RSI 数据集(https://tianchi.aliyun.com/competition/entrance/231717/information)为基于 DL 的作物分类和产量预测方法提供了数据基础。该数据集有 4 096 幅农田地块 RSI,包含烟草、玉米、大麦和水稻 4 种农田地块。所有 RSI 均由无人机在贵州省兴仁市作物产区拍摄,该区域具有典型地理环境、气候条件和文化特征。每幅农田 RSI 为 RGBA 4 通道图像格式(Red、Green、Blue、Alpha, RGBA),约为 30 000 像素。图 3a 为大麦地块,图 3b 中黄色为大麦区域分割。从图 3a 可以看出 RSI 不规则、边界模糊、多尺度。由于无人机不可能固定高度和角度拍

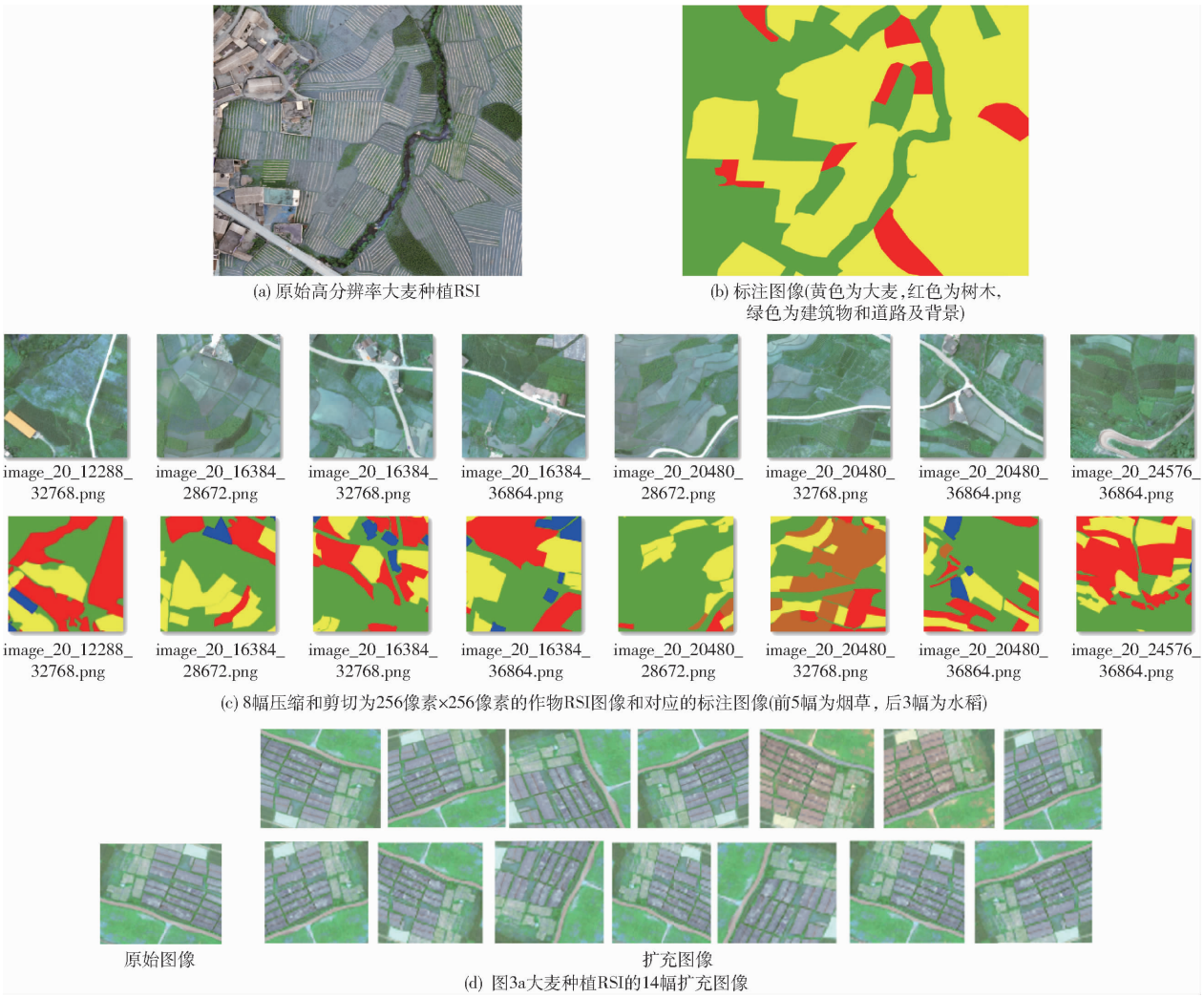

图 3　作物耕地 RSI 实例

Fig. 3　Examples of RSI for crop farmland

摄,所以数据标注不够精细,丢失了一定图像细节,不能在感受野中获得更多有用数据。为提高模型训练效果,将每副 RSI 压缩并裁剪为 256 像素 × 256 像素,如图 3c 所示。

为避免过度拟合和梯度消失问题,对原始 RSI 进行扩充,以生成更多 RSI。首先,每个原始 RSI 被随机裁剪为 256 像素 × 256 像素,然后通过左右翻转、上下翻转、0° ~ 30°随机旋转、随机移位 10 像素等 RSI 扩充策略,将每幅 RSI 扩充为 14 幅,如图 3d 所示,由原始 RSI 和扩充 RSI 构建数据集,该数据集包含 4 096 幅原始图像和 57 344 幅扩充图像。通过 5 折交叉验证(Five-fold cross-validation, FFCV)策略将数据集划分为训练集和测试集,分别用于模型训练和测试。在模型训练前对每副图像进行归一化,将像素值限制为[0,1],由此在保持分割准确率的前提下加速模型迭代。

2.2　试验环境和参数设置

在耕地 RSI 数据集上评估 MSAVM－UNet 性能,并与 4 种最新模型进行对比,包括边缘约束与学习模型(Edge constraint and learning, ECLDeeplab)[10]、MPSPNet 和 UNet 结合模型(MPSPNet－UNet)[11]、多频掩码 Transformer(Multi-Swin Mask Transformer, MSMTransformer)[14]、Mamba－UNet[18]。随机设置模型初始值。为确保公平合理的对比,所有试验均采用相同设置:Windows 10、Intel(R) Xeon(R)、CPU E5－2667v @ 3.20 GHz、NVidia Quadro M4000、Python 3.8.8、PyTorch 1.10 和 CUDA 11.3。对模型超参数使用相同设置:初始学习率设置为 0.001、批量大小设置为 15、动量设置为 0.5、权重衰减设置为 0.000 1、迭代次数设置为 3 000。在耕地 RSI 数据集上通过 5 折交叉验证对其超参数进行微调。

由于 MSAVM－UNet 很大程度上涉及训练样本复杂和类不平衡问题,采用精度(Precision)和相似系数(DSC)作为 MSAVM－UNet 评价指标。

2.3　检测结果可视化

比较 MSAVM－UNet 与 ECLDeeplab、MPSPNet－

UNet、MSMTransformer 和 Mamba - UNet 分割性能。第 3 000 次迭代后模型选择为训练好的模型,用于模型性能测试。图 4 为简单耕地 RSI 的 CLRSIS 结果。从图 4a 可看出,每幅 RSI 都很简单,耕地大而清晰。从图 4c ~ 4g 可发现,所有模型都可以有效地从每幅 RSI 中提取耕地,并且 MSAVM - UNet 提取的耕地信息比较完整,有清晰边缘。

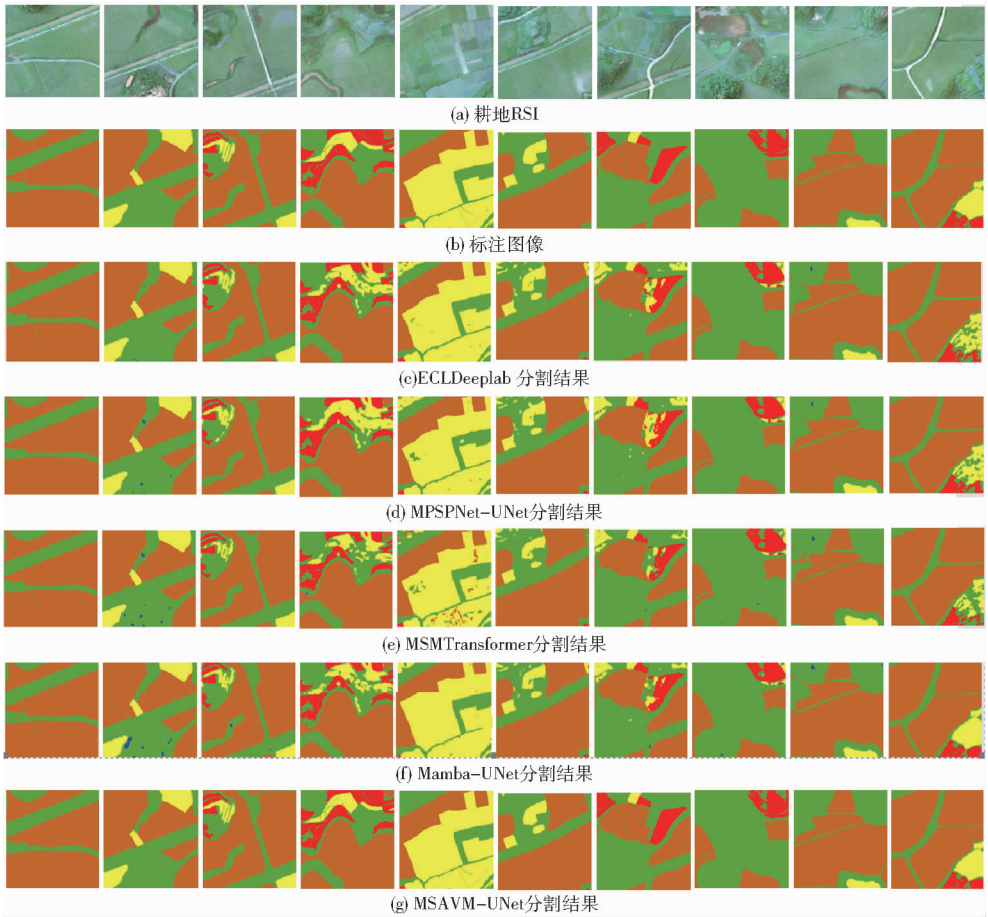

图 4 5 种模型 CLRSIS 结果

Fig. 4 CLRSIS results for five models

为进一步验证 MSAVM - UNet 泛化性,图 5 为 10 幅复杂耕地 RSI 以及 5 种模型检测到的耕地。从图 5a 可看出,这些 RSI 具有目标多尺度、边界模糊、背景复杂、耕地与背景对比度小,并且后 2 幅 RSI 中有许多微小耕地。所以传统的 RSI 分割和检测方法难以实现 CLRSIS。

从图 5c ~ 5g 可看出,MSAVM - UNet 的 CLRSIS 效果最好,可以检测到接近真实的小块耕地,而 ECLDeeplab 和 BECU - Net 的耕地信息不完整,存在重叠和假的耕地信息,并且有一些小块耕地缺失。还可以看到,Mamba - UNet 效果比 ECLDeeplab、MPSPNet - UNet 和 BECU - Net 好,ECLDeeplab 出现过分割问题,这可能是由于卷积运算的局部特性引起的。

2.4 定量检测结果

为验证 MSAVM - UNet 模型有效性,将其与 U - Net 和上述 4 种模型进行定量比较。表 1 为 6 种模型精度、DSC 和训练时间。

由表 1 可知,MSAVM - UNet 精度和 DSC 最高,分别为 85.60% 和 84.46%,但其训练时间稍长,原因为 MSAVM - UNet 的 MVSS、MSAA 和 CAAVSS 能提取耕地 RSI 的多尺度语义特征,且具有较低计算复杂度;Mamba - UNet 结果次之,其训练时间很短,因为它利用了 Mamba 和 UNet 的优势。U - Net 结果最差,主要原因是 U - Net 存在最大池化,导致该模型对耕地的检测和定位不准确。MSMTransformer 训练时间最长,原因为 Transformer 的计算复杂度为二次增长。

2.5 消融试验

为验证 MSAVM - UNet 算法鲁棒性,在相似条件下,对 MSAVM - UNet 进行消融试验。与 Mamba - UNet 不同,MSAVM - UNet 编码解码采用 MVSS 模块,并引入 MSAA 和 CAAVSS 模块,用于提取多尺度特征。为直观地展示 MVSS、MSAA 和 CAAVSS 对 CLRSIS 性能的影响,在上述相同试验条件下,采用 MSAVM - UNet 的变体模型进行试验,得到CLRSIS

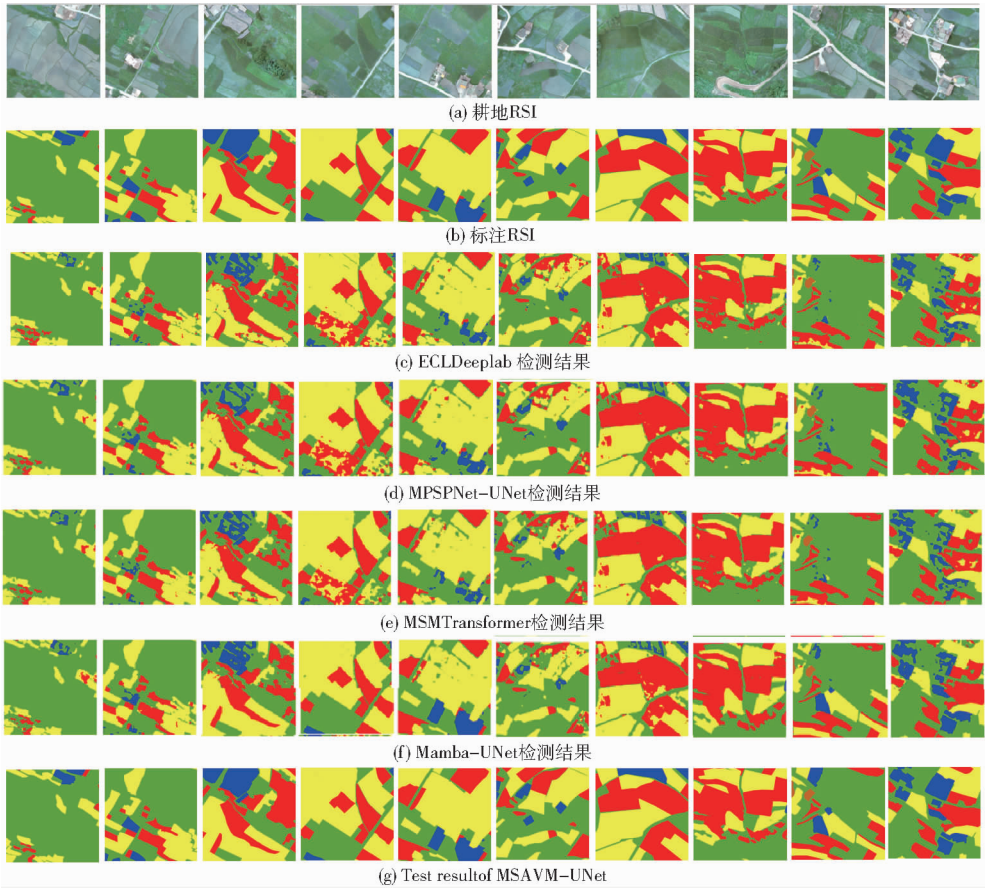

图 5　5 种模型检测到的耕地 RSI 样本视觉对比

Fig. 5　Visual comparison of cultivated land RSI samples detected by five models

表 1　6 种模型 CLRSIS 结果

Tab. 1　CLRSIS results for six models

模型	精度/%	DSC/%	训练时间/h
U－Net	66.74 ± 1.75	63.54 ± 1.74	12.23
ECLDeeplab	80.51 ± 1.68	75.56 ± 1.53	11.64
MPSPNet－UNet	83.54 ± 1.62	81.91 ± 1.41	12.18
MSMTransformer	84.20 ± 1.50	82.51 ± 1.36	14.76
Mamba－UNet	84.66 ± 1.43	82.09 ± 1.32	6.73
MSAVM－UNet	85.60 ± 1.11	84.46 ± 1.05	7.14

结果如表 2 所示。

表 2　消融试验结果

Tab. 2　Results of ablation test

MSAVM－UNet 变体	精度/%	DSC/%	训练时间/h
VSS 代替 MVSS	81.23 ± 1.24	81.07 ± 1.30	6.85
去掉 MSAA	80.62 ± 1.36	80.16 ± 1.31	6.12
跳跃连接代替 CAAVSS	81.41 ± 1.25	80.70 ± 1.22	6.73

从表 2 可看出，MVSS、MSAA 和 CAAVSS 对检测结果都有一定贡献，MSAA 贡献较大，可以减少训练时间，提高 CLRSIS 精度。

然而，该模型对复杂地形的边缘分割存在锯齿现象，需进一步优化多尺度特征融合策略；本文模型只是对耕地进行语义分割，未来工作将所提出模型应用于耕地面积和粮食产量估计，以提升模型在数字化农业中的应用价值。

3　结束语

由于田间耕地 RSI 和周围环境的复杂性，使得基于 RSI 的耕地信息提取研究仍具有挑战性。利用 Mamba 结构在长距离依赖关系建模和全局上下文特征捕获方面的能力，提出一种多尺度注意力视觉 Mamba U－Net（MSAVM－UNet）的耕地遥感分割方法。通过改进视觉状态空间模块（MVSS）、通道感知注意力 VSS 模块（CAAVSS）和多尺度特征聚合模块（MSAA）的协同作用，有效解决了耕地边界模糊、小田块漏分等地类混杂问题。在公开数据集分割精度和 DSC 分别达到 85.60% 和 84.46%，较传统 U－Net 方法分别提升 18.86、20.92 个百分点，训练时间减少 5.09 h。

参　考　文　献

[1]　TANG F, WANG L, FU M, et al. Quantitative characterization of China's farmland scale utilization level and driving factors: a 30-year comprehensive evaluation perspective[J]. Habitat International, 2025, 158: 103335.

[2] 王华,陈梦奇,蔡恩香,等.基于多目标粒子群优化算法的永久基本农田划定模型[J].农业机械学报,2021,52(8):138 - 146,171.
WANG Hua, CHEN Mengqi, CAI Enxiang, et al. Permanent basic farmland delineation model based on multi-objective particle swarm optimization algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(8): 138 - 146, 171. (in Chinese)

[3] 陈文广,廖宇波,孔祥斌,等.基于地块尺度的永久基本农田非粮化特征、驱动与管控[J].农业机械学报,2023,54(6):114 - 124.
CHEN Wenguang, LIAO Yubo, KONG Xiangbin, et al. Characteristics, driviers and control of non-grain production of permanent basic farmland based on plot scale[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 114 - 124. (in Chinese)

[4] 林建平,黄坤,邓爱珍,等.耦合耕地综合质量和空间连通性的基本农田划定研究[J].农业机械学报,2024,55(7):221 - 231,251.
LIN Jianping, HUANG Kun, DENG Aizhen, et al. Basic farmland demarcation coupling comprehensive quality and spatial connectivity of cultivated land[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(7): 221 - 231, 251. (in Chinese)

[5] LEI L, YANG Q, YANG L, et al. DL implementation of image segmentation in agricultural applications: a comprehensive review[J]. Artificial Intelligence Review, 2024, 57:149.

[6] TIWARI T, SARASWAT M. Analysis of UNet-based semantic segmentation models[J]. Lecture Notes in Networks and Systems, 2024,812: 421 - 431.

[7] 张超,高璐璐,郧文聚,等.遥感技术获取耕地质量评价指标的研究进展分析[J].农业机械学报, 2022,53(1):1 - 13.
ZHANG Chao, GAO Lulu, YUN Wenju, et al. Research progress on obtaining cultivated land quality evaluation indexes by remote sensing technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(1): 1 - 13. (in Chinese)

[8] 陈玲玲,施政,廖凯涛,等.基于卷积神经网络的高分遥感影像耕地提取研究[J].农业机械学报,2022, 53(9):168 - 177.
CHEN Lingling, SHI Zheng, LIAO Kaitao, et al. Cultivated land extraction from high-resolution remote sensing images based on convolutional neural networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 168 - 177. (in Chinese)

[9] 张新长,黄健锋,宁婷.高分辨率遥感图像耕地提取研究进展与展望[J]. 武汉大学学报(信息科学版), 2023, 48(10): 1582 - 1590.
ZHANG Xinchang, HUANG Jianfeng, NING Ting. Research progress and prospect of cultivated land extraction from high-resolution remote sensing images[J]. Geomatics and Information Science of Wuhan University, 2023, 48(10): 1582 - 1590. (in Chinese)

[10] 魏青迪,范昊,张承明.基于 ECLDeeplab 模型提取华北地区耕地的方法[J].江苏农业科学,2020,48(4):209 - 215.
WEI Qingdi, FAN Hao, ZHANG Chengming. Method for extracting cultivated land in North China based on ECLDeeplab model[J]. Jiangsu Agricultural Sciences, 2020, 48(4): 209 - 215. (in Chinese)

[11] 李倩楠,张杜娟,潘耀忠,等. MPSPNet 和 UNet 网络下山东省高分辨耕地遥感提取[J].遥感学报,2023, 27(2):471 - 491.
LI Qiannan, ZHANG Dujuan, PAN Yaozhong, et al. Remote sensing extraction of high-resolution cultivated land in Shandong Province using MPSPNet and Nnet networks[J]. Journal of Remote Sensing, 2023, 27(2): 471 - 491. (in Chinese)

[12] CAO H, WANG Y, CHEN J, et al. Swin - Unet: Unet-like pure transformer for medical image segmentation[J]. arXiv: 2105.05537, 2021.

[13] 董张玉,李金徽,张晋,等.边缘增强的 BECU - Net 模型高分辨率遥感图像耕地提取[J].遥感学报,2023,27(12):2847 - 2859.
DONG Zhangyu, LI Jinhui, ZHANG Jin, et al. Edge-enhanced BECU - Net model for cultivated land extraction from high-resolution remote sensing images[J]. Journal of Remote Sensing, 2023, 27(12): 2847 - 2859. (in Chinese)

[14] BO Z, WEI T, LUO X, et al. Multi-swin mask transformer for instance segmentation of agricultural field extraction[J]. Remote Sensing, 2023, 15(3): 549.

[15] LI X, DING H, YUAN H, et al. Transformer-based visual segmentation: a survey[J]. arXiv:2304.09854,2023.

[16] RAHMAN M, TUTUL A, NATH A, et al. Mamba in vision: a comprehensive survey of techniques and applications[J]. arXiv:2410.03105, 2024.

[17] LIU Y, TIAN Y, ZHAO Y, et al. VMamba: visual state space model[J]. arXiv:2401.10166,2024.

[18] WANG Z, ZHENG J, ZHANG Y, et al. Mamba - UNet: UNet-like pure visual Mamba for medical image segmentation[J]. arXiv:2402.05079, 2024.

[19] LIAO W, ZHU Y, WANG X, et al. LightM - UNet: Mamba assists in lightweight UNet for medical image segmentation[J]. arXiv:2403.05246, 2024.