融合高分影像与时序 NDVI 的农作物语义分割模型

赵 旭[1]，李 浩[1※]，朱益虎[2]，王胜利[2,3]，何燕兰[2]

（1. 河海大学地球科学与工程学院，南京 211100；2. 江苏省地质测绘大队，南京 211102；3. 中国矿业大学环境与测绘学院，徐州，221116）

摘 要：通过遥感技术准确及时地掌握农作物分类信息，对农业生产的管理、预估产量以及调整种植结构等方面至关重要。随着光学传感器性能的不断提升，遥感影像的分辨率也在持续提高，农业遥感正逐步进入高精度时代。然而，目前的高分辨率农作物语义分割模型在利用包含农作物物候信息的时序数据方面存在一定的困难，特别是在既有单季作物也有双季作物的复杂种植结构区域。针对此问题，该文提出了一种能够融合高分辨率遥感影像和中分辨率时序 NDVI 的语义分割模型 MCSNet（multi-source crops segmentation network），该模型采用双编码器结构，能够有针对性地同步挖掘高分辨率影像的空间细节与中分辨率时序影像的时空特征，并通过注意力机制引导的数据融合模块对时空信息进行充分融合，提高了农作物分类精度。试验表明，该模型加入了时序 NDVI 数据后分类精度大幅提高；在对比试验中，该模型分类结果的平均交并比和总体精度分别达到了最高的 77.75% 和 89.56%；在卷积长短期记忆单元和残差双注意力模块的联合作用下，该模型的分类结果在平均交并比和总体精度上分别提升 3.84、4.24 个百分点。将该模型应用到研究区盱眙县，得出了县域尺度的高分辨率农作物分类结果，制图效果优秀，且各项评价指标的精度均高于基于像素与面向对象的双向长短期记忆网络算法，为基于深度学习语义分割算法的大面积复杂种植结构区域农作物制图提供了可行的方案。

关键词：农作物；分类；高分辨率遥感；时序 NDVI；语义分割；深度学习

doi：10.11975/j.issn.1002-6819.202407098

中图分类号：TP79；S127 **文献标志码**：A **文章编号**：1002-6819(2025)-14-0216-12

赵旭，李浩，朱益虎，等. 融合高分影像与时序 NDVI 的农作物语义分割模型[J]. **农业工程学报**，2025，41(14)：216-227. doi: 10.11975/j.issn.1002-6819.202407098 http://www.tcsae.org

ZHAO Xu, LI Hao, ZHU Yihu, et al. A crop semantic segmentation model integrating high-resolution imagery and time-series NDVI[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(14): 216-227. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202407098 http://www.tcsae.org

0 引 言

及时精确地掌握农作物分类信息对精准农业各方面，如作物面积提取、生长状况监测和产量估计等至关重要，不仅能有效管理农业生产，还对保障粮食安全和促进社会经济持续发展具有战略意义[1-3]。传统农作物分类方法主要由人工实地调查完成，存在实时性差、效率低下等缺点。相比之下，遥感技术具有实时性强、范围大、成本低等优势，能够快速而准确地获取大范围区域内的农作物空间分布、长势以及产量等农业信息，为数字化农业应用提供基础数据[4-6]。

农作物的显著特征是物候特征，这些特征直接反映了农作物在其生长周期中的动态变化和对环境因素的响应。近年来，国内外学者对基于多时相遥感数据的农作物分类方法进行了大量研究，研究发现在构建长时序遥感数据时，使用全部的光谱信息会导致数据冗余，计算复杂，分类器对特征不敏感等问题[7]。针对这一问题，一些研究使用植被指数（vegetation indices，VI）作为特征，能够在一定程度上提高分类器对植被状态的敏感性，同时降低了数据量，提高了分类效率。其中，归一化植被指数（normalized difference vegetation index，NDVI）成为了农作物与植被分类最常用的指数[8-9]。李晓慧等[10]利用多时相 Landsat 8 OLI 影像构建 NDVI 时间序列，结合光谱角填图和决策树分类方法，完成了研究区内春玉米、谷物、大豆和马铃薯的分类，总体精度为 85.34%，Kappa 系数为 0.76。KHALIQ 等[11]利用多时相 Sentinel-2 构建的时序 NDVI，采用随机森林算法对意大利某区域农作物进行分类，总体精度达到 91.2%。时间序列数据在作物分类中的应用取得了重大进展，特别是在全年存在多季轮作和多作物种类的地区，长时序数据显得尤为重要。然而，这些研究大多基于重返周期较短的中低分辨率遥感卫星，而重返周期更长的高分辨率卫星则难以大面积获取多时相数据，导致这些研究成果在进行更高精度农作物分类时的适用性受到限制。

随着光学传感器性能不断改善，近年来，WorldView、国产高分系列卫星等多颗高空间分辨率卫星已先后投入商业运营，遥感影像进入米级、亚米级时代[12]，高分遥感影像也已被广泛应用于农作物生产管理、农作物面积提取等多个领域，使得农作物提取进入“精细化”时代，

收稿日期：2024-07-11 修订日期：2025-01-19
基金项目：江苏省地质局科研项目（2022KY15）
作者简介：赵旭，研究方向为深度学习遥感影像解译。
Email：zx0065@hhu.edu.cn
※通信作者：李浩，博士，教授，博士生导师，研究方向为摄影测量与遥感。Email：lihao@hhu.edu.cn

高分辨率遥感农作物分类方法也经历了长足的发展，包括分析单元从基于像素到面向对象的转变，从单一特征提取扩展到多源多尺度的特征提取，以及从简单分类器进化到复杂分类器。在各种应用场景中，这些方法能够实现较高的解译精度。徐新刚等[13]以 QuickBird 高分辨率影像为数据源，采用基于像元的最大似然法监督分类方法，对四川绵阳某区域进行农作物分类，分类准确率达到 95.3%。ESETLILI 等[14]利用 RapidEye 高分辨率数据和面向对象分类方法对土耳其某平原农作物分类，试验结果表明，面向对象方法分类精度 OA（overall accuracy）高达 94.12%，优于基于像素的 SVM 算法。这些基于传统的机器学习分类方法虽然对简单特征有较好的提取效果，但处理过程只经过较少层次的非线性变换组合，对影像中复杂特征的提取效果较差。相比之下，深度学习方法的发展为高精度的农作物分类提供了更强大的工具和技术支持，其主要优势在于其能够通过端到端的网络模型自动学习多层次特征。其中，卷积神经网络（convolutional neural networks，CNN）是一种专门用于处理图像数据的深度学习模型，能够很好地提取遥感影像中隐含的深层特征信息[15]；递归神经网络（recurrent neural network，RNN）是一种专门为处理多维时间序列的网络，可以准确捕捉时间序列中的时间相关性[16]；基于 RNN 框架扩展而成的长短期记忆网络[17]（long short term memory，LSTM）成为常用的分析时间序列的模型。这些神经网络模型强大的特征学习能力能够有效提升农作物遥感影像的解译精度。

随着深度学习技术在图像处理领域的迅速发展，语义分割逐渐成为图像像素级解译的主流方法，并被广泛应用于遥感影像分类任务中[18]，包括耕地提取、水体提取、建筑提取等地表单要素提取和地表全要素分类任务[19-22]。在高分辨率遥感影像作物分类任务，语义分割算法也得到了广泛的研究和应用。董秀春等[23]以 World-View-2 遥感影像为数据源，测试了 U-Net 和 DeepLabV3+ 语义分割模型在某研究区的小麦分类效果，测试结果表明，两种模型的小麦分类总精度和 Kappa 系数分别在 94% 和 0.89 以上，但试验仅针对单一作物类型。XIANG 等[24]使用无人机遥感影像对作物类型进行精确分割，提出了一种端到端的特征融合语义分割网络 CTFuseNet（cnn-transformer feature-fused network），在数据集上测试的平均交并比 85.33%，像素精度为 92.46%，但无人机影像研究区范围小，缺乏代表性。LU 等[25]以 GF-1 高分辨率遥感影像为数据源，利用 CSNet（crop segmentation network），对哈尔滨市某区域玉米、水稻、大豆进行语义分割，分类精度 OA 和 mIoU 分别达到了 90.6% 和 0.825，但模型仅用了单一数据源。XU 等[26]提出了多层金字塔作物语义分割网络（multi-layer pyramid crop classification network，MP-Net），基于 GF-6 和 Sentinel-2 数据，结合两种数据在分辨率和光谱数量上的优势完成作物分类，在两个研究区的分类总体精度 OA 分别达到了 94.17% 和 92.28%，模型虽融合了中分辨率卫星的光谱特征，但不包含时序信息特征。

综合上述内容，实现对农作物的精细分类高度依赖于高分辨率影像所提供的空间细节信息以及高时间分辨率所提供的时序信息。然而，现有的高分辨率语义分割模型在处理高分辨率影像时，往往难以兼顾长时序数据，因此通常只针对小面积、单一时间段的单季作物类型进行分类。针对复杂种植结构区域中同时存在单季、双季种植的情况，高分辨率语义分割网络在融合与挖掘时序信息方面仍有很大的发展潜力。

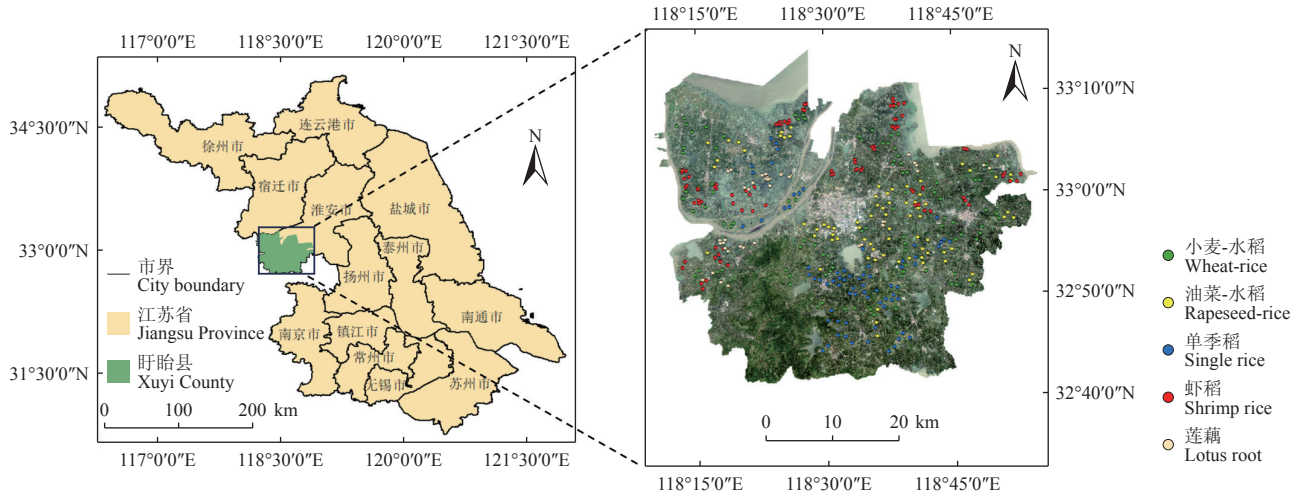
鉴于此，本文提出了一种能够融合高分辨率影像与时序 NDVI 的多源数据农作物语义分割网络（multi-source crops segmentation network，MCSNet）。该模型针对时序 NDVI 影像和高分辨率遥感影像的数据格式，分别构建了能够有效挖掘时序特征和高分辨率空间特征的编码器，并引入了注意力机制以充分融合时空特征，从而提高在复杂种植结构区域的高分辨率农作物分类精度。本文选择江苏省淮安市盱眙县作为研究区，验证了所提模型的可行性与有效性。同时，结合农作物分类结果分析了研究区内小麦-水稻、油菜-水稻、单季稻、虾稻等主要作物类型的空间特征，为作物种植结构的调整和优化提供了科学依据。

1　数据说明

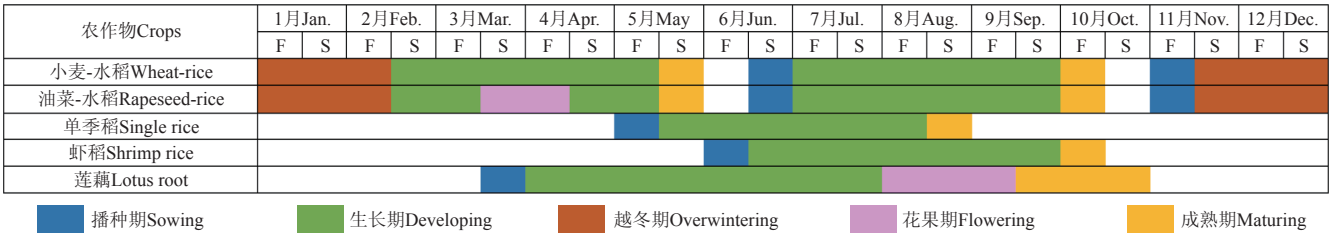
1.1　研究区概况

盱眙县位于江苏省北部，隶属淮安市，地处江淮平原，地势平坦，属温带季风气候，四季分明，雨量充沛，是中国重要的粮食生产基地，研究区位置见图 1a。通过对盱眙县的实地调研发现当地全年主要农作物有水稻、小麦、莲藕、油菜，花生，大豆等。其中，水稻、小麦、莲藕占主要耕地面积，且当地的种植结构类型丰富，存在如小麦-水稻、油菜-水稻的双季种植模式，及稻虾共作种植模式。将研究区以下几类主要农作物类型作为研究对象：小麦-水稻、油菜-水稻、单季稻、虾稻、莲藕。

图 1b 为研究区农作物物候历。小麦-水稻为双季种植模式，每年 11 月初播种小麦，越冬后，来年 2 月底返青，进入生长期，至 5 月下旬成熟并完成收割，6 月开始种植第二季作物水稻，成长期 3 个月左右，在 10 月上旬完成收割。双季种植模式还包含油菜-水稻，每年 11 月初完成对油菜的播种，经历越冬期后，于 2 月下旬开始进入生长期，并于 3 月下旬开花，花期通常结束于 4 月上旬，在 5 月底完成对油菜的收割后，进入第二季的水稻种植期，物候规律同小麦-水稻。单季种植模式主要包括：虾稻、早稻、莲藕。其中稻虾养殖模式主要为每年 3–4 月投放虾苗，5–6 月份第一季收虾，同时完成水稻的移栽，8–9 月第二季收虾，在 9 月底完成水稻收割；早稻种植模式，在 5 月上旬完成对水稻的移栽，经过 3 个月左右的成长期，在 8 月下旬成熟并收获；莲藕在每年 3 月份进行移栽，经过 3 个月左右的成长期，于 8 月至 9 月之间开花结藕，10 月完成收获。



a. 研究区位置以及农作物样本空间分布
a. Location of the research area and spatial distribution of crop samples



b. 研究区主要农作物物候历
b. Crop calendars for major crops in study area

注：F 和 S 分别表示每个月的上半月和下半月。
Note: F and S represent the first and second half of each month, respectively.

图 1 研究区概况
Fig.1 Overview of the research area

综上所述，选择盱眙县作为研究区是因为该区域种植模式多样，既有单季作物，也有双季作物，具有典型性和代表性。同时，盱眙县的种植面积较大，为进一步验证和探究模型的泛化性提供了良好的条件。

1.2 数据及预处理

本研究的高分辨率遥感数据采用国产高分 2 号（GF-2）卫星多光谱影像。GF-2 卫星搭载 2 个全色和多光谱传感器和 4 个宽视场传感器，其中多光谱传感器相机可以获取分辨率为 0.8 m 的全色波段和 3.2 m 的多光谱影像。由于 GF-2 重返周期长、单景影像覆盖面积有限、存在云雾遮挡等问题，需要获取不同时间的多景影像数据以实现对研究区的覆盖。本文共获取了 14 景有效影像实现对研究区的覆盖，影像包含 2021 年 4 月的 4 景、5 月的 7 景和 7 月的 3 景影像。对影像进行辐射定标、大气校正和正射矫正；使用最近邻扩散的全色锐化（nearest-neighbor diffusion-based pan sharpening）算法完成全色影像与多光谱影像的像素级图像融合；再经过镶嵌、裁剪操作，得到 1 幅空间分辨率为 0.8 m 覆盖盱眙全境的 4 波段（红、绿、蓝、近红外）遥感影像。

本研究的时序数据采用 Sentinel-2 多光谱影像构建的时序 NDVI 影像数据。Sentinel-2 是欧洲航天局研发的多光谱成像对地观测系统，该系统由 Sentinel-2A 和 Sentinel-2B 两颗卫星组成，二者结合可以实现 5 d 的时间分辨率。两颗卫星均搭载了多光谱成像仪，包括 13 个波段，涵盖可见光、红边、近红外和短波红外光谱，其中蓝、绿、红和近红外 4 个波段为 10 m 分辨率。本文选取了云量覆盖率低于 10% 的 12 幅影像，拍摄时间跨度为整个 2021 年，具体的拍摄时间及对应的云量覆盖率详见表 1。依次对 12 幅影像数据进行辐射定标和大气校正、裁剪操作；计算每一景影像计算归一化植被指数 NDVI，然后进行波段合成，合成一幅 12 波段的时序 NDVI 影像数据。

表 1 Sentinel-2 影像列表
Table 1 List of Sentinel-2 images

序号 No.	拍摄时间 Shooting time	云量覆盖 Cloud cover/%
1	2021-01-01	5.224
2	2021-02-20	0
3	2021-03-22	0
4	2021-04-06	1.538
5	2021-05-01	0.032
6	2021-06-05	0
7	2021-07-30	0
8	2021-08-19	3.765
9	2021-09-23	6.399
10	2021-10-03	3.421
11	2021-11-12	0
12	2021-12-17	0

预处理后的 GF-2 影像位于 CGCS2000 坐标系，而时序 NDVI 影像数据位于 WGS-84 坐标系，将 GF-2 重

投影至 WGS-84 坐标系，再使用地面控制点对两副影像配准，确保两幅影像几何位置上精确对应。

1.3　数据集制作

随机选择 3 个 5 km×5 km 农作物类别较为均衡的典型区域作为数据集标注位置，3 个区域在较大程度上包含了需要研究的 5 种农作物。借助实地调查信息及高清航空影像，绘制出专家解译标签，作为农作物真值数据。使用 ArcGIS 对选取出的区域进行标注，如图 2a 所示，图中 5 种色彩分别代表 5 种作物，黑色为背景。将 GF-2 影像及标签数据和时序 NDVI 影像进行地理配准，然后按照一定的重叠率对影像进行裁剪。将 GF-2 与标签数据裁剪为 512×512 像素大小，为了减少数据冗余及适配本文所提出网络的输入尺寸，将时序 NDVI 数据裁剪后，重采样至 128×128 像素大小，最终获得 GF-2、时序 NDVI、标签相互对应的共 2 832 对数据的数据集，部分数据集展示如图 2b 所示。

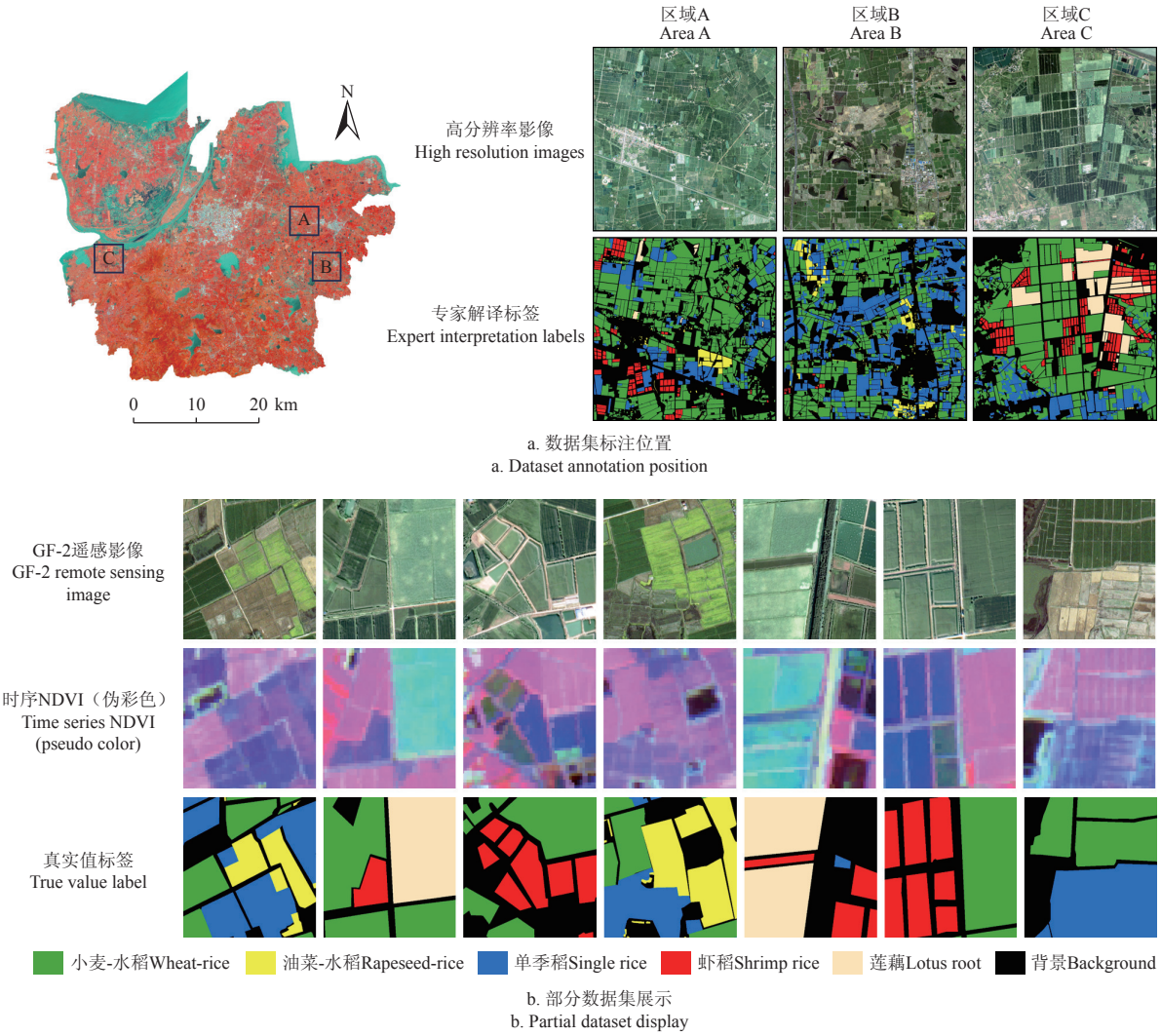


图 2　数据集制作

Fig.2　Dataset production

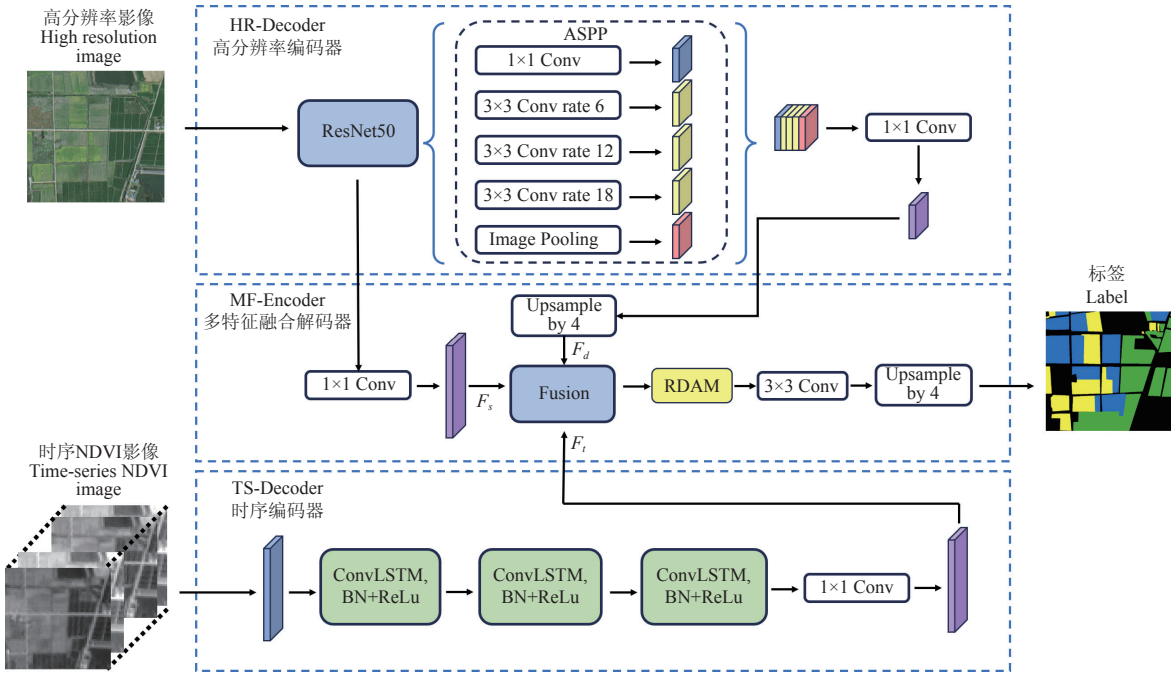
2　多源数据农作物语义分割网络——MCSNet

2.1　模型整体结构

本文提出的 MCSNet 模型整体结构如图 3 所示，模型主要由高分辨率编码器（high-resolution decoder，HR-Decoder）和时序编码器（temporal-sequence decoder，TS-Decoder）构成的双编码器结构，以及引入注意力机制的多特征融合解码器（multi-feature fusion encoder，MF-Encoder）结构组成。模型通过高分辨率编码器结构输入 512×512 像素大小的分辨率为 0.8 m 的高分辨率影像；时序编码器输入 128×128 像素大小，重采样至 3.2 m 分辨率的时序 NDVI 数据。之所以这样设计是出于以下两个原因的综合考量：1）重采样操作的只能改变数据的输入尺寸，并不改变数据的实际分辨率。如果将 10 m 分辨率的时序 NDVI 数据重采样至与高分辨率影像相同的 0.8 m 分辨率，将大大增加数据冗余与模型参数量，进而增加模型运算量，减缓模型推理速度。2）在不进行重采样的情况下，两种数据源虽然能以最小的数据量保留信息，但两种数据的分辨率不具备合适的倍数关系，会导致双编码器输入的数据中相同位置信息无法有效整合。

本文将两种数据源的输入分辨率设置为 1:4 的比例关系，不仅可以在一定程度上控制数据冗余，模型参数，

提升模型推理速度；同时可以利用卷积和池化结构在提取图像特征的同时降低了图像的分辨率，将高分辨率图像中的空间信息有效压缩，进而实现高分辨率的空间特征与中分辨率的时序特征在同一空间尺度上的融合。



注：ASPP 为空洞空间金字塔池化，Conv 为卷积，rate 为膨胀率，Image Pooling 为图像池化,ResNet50 表示以 ResNet50 作为骨干网络，Fusion 表示特征融合，RDAM 为卷积块注意力模块，ConvLSTM 为卷积长短期记忆单元，BN+ReLu 为 Batch Normalization 层和 ReLU 层，Upsample 为上采样，F_d、F_s 和 F_t 分别代表深层特征、浅层特征和时序特征。

Note: ASPP represents atrous spatial pyramid pooling, Conv represents convolution, rate represents the dilation rate, Image Pooling represents image pooling operation, ResNet50 represents the backbone network based on ResNet50, Fusion represents feature fusion, RDAM represents the Convolutional Block Attention Module, ConvLSTM represents convolutional long short-term memory unit, BN+ReLU represents batch normalization and ReLU layers, Upsample represents upsampling, F_d, F_s and F_t represent deep features, shallow features, and temporal features, respectively.

图 3 模型总体结构

Fig.3 Overall structure diagram of the model

2.2 高分辨率编码器

高分辨率编码器（HR-Decoder）承担着提取高分辨率遥感影像信息的作用，所输入的高分辨率遥感影像包含着农作物丰富的光谱、纹理以及边界信息。编码器的主干网络占据了模型大部分的参数量和计算量，它影响着整个模型的分割性能，因此选择合适的主干网络至关重要[27-28]。编码器采用了 ResNet50[29] 作为其主干网络，ResNet50 是一个深度残差网络，它通过引入残差连接来促进深层网络中信息的流动，有效地缓解了随网络加深而产生的梯度消失问题，极大地增强了模型在深层特征提取方面的能力，使得模型能够更加有效地从高分辨率遥感影像中提取出细致和复杂的农作物特征。

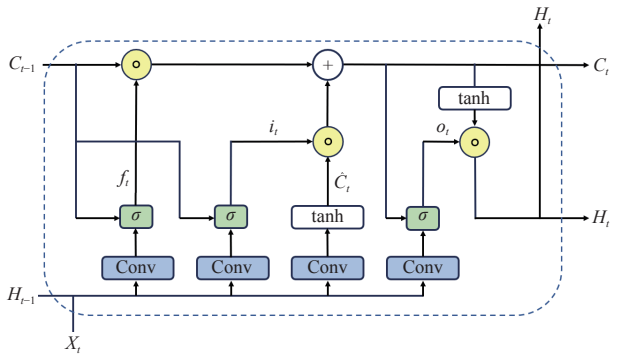
为了能在提取高分影像语义信息的同时保留其细节信息，在主干网络中加入分支，在完成 4 倍下采样后，保留一份高分辨率影像的细节特征。主干网络在完成 16 倍下采样后，为进一步深化语义特征，引入了空洞空间金字塔池化（atrous spatial pyramid pooling，ASPP）模块，ASPP 通过并行采用不同采样率的空洞卷积，能够捕捉不同尺度的图像信息，这有助于处理大小不一、破碎地块的农作物信息。

该编码器最终通过两个分支分别输出特征 F_s 和 F_d，F_s 为浅层特征图，浅层特征具有更高的分辨率和更丰富的细节信息，有助于改善物体边缘的预测精度。F_d 为深层语义特征图，经历了更深层的下采样以及空洞空间金字塔池化模块，拥有更加丰富的语义信息，更加有助于理解图像中的各种农作物和场景之间的关系，尤其是当这些农作物和场景在尺寸上有很大差异时。

2.3 时序编码器

时序编码器（TS-Decoder）旨在对 NDVI 时序数据进行高效的特征提取。NDVI 时序数据中包含着更加丰富的农作物物候信息，但普通卷积层在对时序 NDVI 数据的处理上，并不能兼顾通道间的时序关系。本文引入 ConvLSTM（convolutional long short term memory）[30] 单元，它是 LSTM 的变体，1 个 ConvLSTM 单元由 3 个门联合调制即输入门、遗忘门和输出门，由于融合了卷积运算，与传统 LSTM 相比，它能在处理时间序列的同时，保留空间信息，减少空间数据的冗余，ConvLSTM 单元示意图见图 4。

时序编码器（TS-Decoder）输入 128×128 尺寸、通道数为 12 的时序 NDVI 数据，首先通过 3×3 卷积层对空间特征进行初步提取，随后，经过 3 个串联的 ConvLSTM 单元，每个 ConvLSTM 单元后串联 Batch Normalization+ReLU 层用来稳定梯度分布并增强特征表达能力，充分提取数据的时间序列特征，随后将处理后的特征输入到普通卷积进行压缩，为后续的融合步骤做准备。



注：Conv 为卷积，X_t 为当前时间步输入，H_t 和 H_{t-1} 分别为当前和上一时间步隐藏状态，C_t 和 C_{t-1} 分别为当前和上一时间步记忆单元状态，f_t、i_t、o_t 分别为遗忘门、输入门和输出门，$\hat{C}_t$ 为候选记忆单元状态，σ 为 Sigmoid 激活函数，tanh 为双曲正切函数，∘为哈达玛积。
Note: Conv represents the convolution operation, X_t represents the input at the current time step, H_t and H_{t-1} represent the hidden states at the current and previous time steps, C_t and C_{t-1} represent the cell states at the current and previous time steps, f_t, i_t and o_t represent the forget gate, input gate, and output gate, $\hat{C}_t$ represents the candidate cell state, σ represents the Sigmoid activation function, tanh represents the hyperbolic tangent function, and ∘ represents the Hadamard product.

图 4　ConvLSTM 网络模型图
Fig.4　ConvLSTM network model diagram

2.4　多特征融合解码器

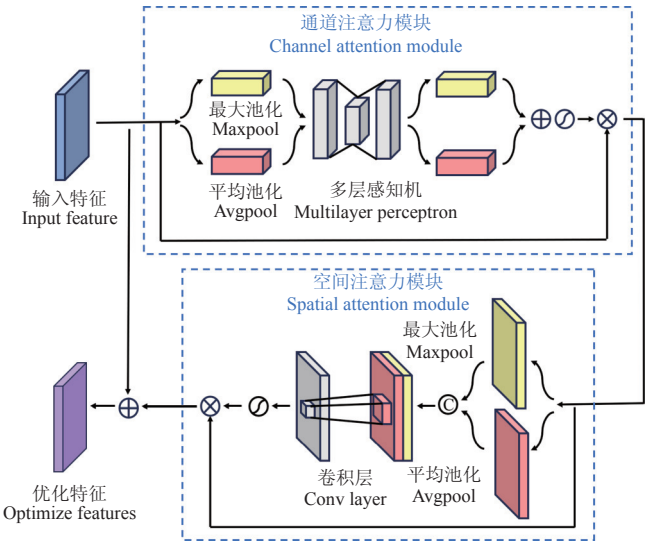
多特征融合解码器（MF-Encoder）的作用将高分辨率编码器的输出的深层特征图 F_d、浅层特征图 F_s 与时序编码器的输出时序特征图 F_t 进行融合，将多源遥感信息特征优势进行互补，再通过上采样输出结果。该模块中的融合原理为先通过乘法运算与拼接操作对特征进行初步融合，再输入所设计的残差双注意力机制模块，挖掘融合特征中的特征重要性。

具体流程为先将 F_d、F_s、F_t 与 $F_d{\cdot}F_s{\cdot}F_t$ 进行拼接操作，将这几个特征张量在同一维度上拼接起来，完成特征的初步融合。这种拼接方式保留了原始特征的独立信息，又通过逐点乘积提取了高阶交互特征，丰富了特征表达的多样性，拼接后的特征包含了更丰富的时空和空间信息，使得模型能够更好地进行特征学习和分类。

随后输入本文设计的残差双注意力模块（residual dual attention module, RDAM）对融合特征进行增强操作。模块基于 CBAM（convolutional block attention module）进行设计，并结合了残差连接。图 5 展示了 RDAM 的结构组成，通过串联通道空间注意模块（channel attention module, CAM）和空间注意模块（spatial attention module, SAM），使每个分支能够独立学习通道和空间轴上的关键特征，再引入残差连接，将原始特征与经过双注意力模块增强的特征相结合，不仅可以避免学习过程中出现梯度消失现象，同时也增强了模型的特征表达能力。

模块的上半部分是通道注意力机制，它首先对输入的特征在高度和宽度维度上进行全局最大池化和全局平均池化，生成与输入通道数相匹配的特征。随后分别被送入多层感知机（multilayer perceptron，MLP）中进行处理。多层感知机的结构为两层全连接层。然后使用 Sigmoid 函数对两个结果之和进行激活并与原始特征进行相乘，得到通道注意力特征。

模块的下半部分是空间注意力机制，该机制首先使用池化层对特征图进行最大值和平均值压缩以减少冗余信息。然后将这两个池化结果堆叠并通过 3×3 卷积层生成空间注意力图，以增强重要的空间特征。最后使用 sigmoid 函数进行激活并与输入特征相乘，得到空间注意力特征。



注：⊕代表加法运算，⊗表示乘法运算，⊘表示 sigmoid 函数，©表示 concat 操作。
Note: ⊕ represents addition operation, ⊗ represents multiplication operation, ⊘represents sigmoid function, © represents concat operation.

图 5　RDAM 模块结构
Fig.5　RDAM module structure

最后，经过残差链接，得到优化特征，将经过 RDAM 处理的优化特征 F 输入到 3×3 卷积，对数据进行降维，再通过反卷积操作完成 4 次上采样，恢复特征图的空间分辨率，最后一层是分类层，它通过 1×1 的卷积将特征图中的每个像素点映射到目标类别的概率上，输出最终的农作物分类结果。

3　试验结果与分析

3.1　试验设计

将本文的农作物类别数据集按照 2:1:1 的比例随机划分为训练集、验证集、测试集。训练集用于模型的训练，验证集用于更新优化模型的超参数，测试集用于对模型精度进行评价。本文所提模型基于 Pytorch 深度学习框架实现，训练配置如下：GPU 采用 RTX4070 ti（12G 显存），CPU 采用 i5-13400F，运行内存为 32G。采用 Adam 作为模型训练的优化器训练的初始学习率设置为 0.001，并采用指数衰减的学习率衰减策略。训练批次大小 Batch Size 设置为 4，迭代次数 epoch 设置为 200。

在本文使用的农作物样本中，小麦-水稻样本数量远高于其他几类，存在样本不平衡现象，且农作物样本形状、边界较为复杂，应当选择更加符合样本条件的训练损失函数。将 Focal Loss[31] 和 Dice Loss[32]。组成混合损失函数作为模型训练的损失函数。

Focal Loss 是一种在类别不平衡的数据集上训练深度学习模型时常用的损失函数，这个损失函数是对交叉熵

损失的改进，其目的是减少容易分类样本的相对重要性，从而让模型集中学习那些难以分类的样本，其算式如下：

$$L_f = -\sum_{C=1}^{C} \alpha_c (1-p_c)^{\gamma} \ln(p_c) \quad (1)$$

式中 L_f 为 Focal Loss 计算结果，C 是类别的总数。α_c 是平衡正负样本的权重系数，是一个针对每个类别的权重。p_c 是模型对于每个类别的预测概率。γ 是一个调节因子，用来减少易分类样本的权重并提升难分类样本的权重。

Dice Loss 损失函数是一种用于图像分割的损失函数，其式如下：

$$L_d = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (2)$$

式中 X 和 Y 分别表示预测值和真实标签，L_d 为 Dice Loss 损失函数计算结果。Dice Loss 的值越小，说明预测结果与真实标签的重叠度越高。

Focal Loss 和 Dice Loss 在损失函数的形式上具有互补性，前者专注于解决类别不平衡问题，后者关注于优化分割结果的空间相似性。两者结合作为损失函数可以更全面地指导模型训练，其式如下：

$$L = L_f + L_d \quad (3)$$

式中 L 为总损失，L_f 为 Focal Loss 计算结果，L_d 为 Dice Loss 损失函数计算结果。

3.2 对比试验

为定性定量评估本文所提 MCSNet 农作物分类任务上的分类效果，试验选用了经典语义分割模型 PSPNet、U-Net、U-Net++、DeepLabV3+，以及能够从不同尺度提取特征的 MA-Net[33]（multi-scale attention network）和结合 Transformer 架构全局特征捕捉能力与卷积网络局部特征提取优势的 SegFormer[34]，进行对比分析。同时，为了证明时序 NDVI 数据的引入对这些模型分类效果产生的影响，对这些模型分别进行两组试验，第一组仅使用高分辨率影像，第二组使用高分辨率影像并引入时序 NDVI。需要注意的是，为保证所对比模型能够同时输入高分辨率影像和时序 NDVI 数据，需对时序 NDVI 数据进行重采样，使其与高分辨率影像有相同的输入尺寸，将数据堆叠后输入模型。采用语义分割常用的评价指标交并比（intersection over union，IoU）、平均交并比（mean intersection over union，mIoU）以及总体精度（overall accuracy，OA）进行精度评价，试验结果如表 2 所示。根据表格可以看出，在融入时序 NDVI 之后，7 类模型的分类精度都显著提高，相较于只使用高分数据，PSPNet、U-Net、U-Net++、DeepLabv3+、MA-Net、SegFormer、MCSNet 的 mIoU 分别提高了 6.55、8.39、9.93、12.13、12.53、13.05、15.03 个百分点；OA 提高了 10.89、9.88、10.7、12.08、13.09、13.42、14.82 个百分点。本文模型在小麦-水稻，小麦-油菜，单季稻、虾稻、莲藕这几类具有明显物候特征的农作物分类上，精度提高了 17.58、19.49、16.03、13.57、9.95 个百分点。证明了时序 NDVI 的融入对农作物语义分割精度的提高具有作用。在融入时序 NDVI 之后，本文模型的 mIoU 和 OA 达到了 77.75% 和 89.56%，在对比试验中达到了最高的精度。

表 2 对比试验结果

Table 2 Comparative experimental results %

使用数据 Data use	方法 Method	IoU						mIoU	OA
		小麦-水稻 Wheat - rice	油菜-水稻 Rapeseed- rice	单季稻 Single rice	虾稻 Shrimp rice	莲藕 Lotus root	背景 Background		
高分影像 High-Resolution Imagery	PSPNet	59.98	48.32	59.57	63.19	65.06	60.23	59.39	68.82
	U-Net	60.54	49.65	60.84	64.74	66.60	60.59	60.49	71.21
	U-Net++	61.73	50.48	60.01	65.26	65.84	61.76	60.85	72.32
	DeepLabV3+	61.50	49.17	61.25	68.83	67.01	62.57	61.72	73.67
	MA-Net	60.91	51.45	61.33	67.77	67.79	62.68	61.99	73.34
	SegFormer	61.81	50.98	61.38	68.65	67.32	62.54	62.11	74.07
	MCSNet	62.18	52.83	61.62	68.08	68.48	63.11	62.72	74.56
高分影像+时序 NDVI High-Resolution Imagery + Temporal NDVI	PSPNet	70.32	56.23	67.53	68.66	67.49	65.41	65.94	76.82
	U-Net	72.12	58.87	71.58	69.96	71.32	69.45	68.88	80.21
	U-Net++	72.73	61.23	73.32	72.32	74.87	70.23	70.78	81.32
	DeepLabV3+	76.23	65.31	73.86	76.23	77.23	74.21	73.85	85.67
	MA-Net	75.87	68.43	75.23	76.12	74.65	76.84	74.52	86.34
	SegFormer	77.21	67.96	75.58	76.32	77.23	76.67	75.16	87.49
	MCSNet	79.76	72.32	77.65	81.65	78.43	76.69	77.75	89.56

注：IoU 为交并比，mIoU 为平均交并比，OA 为总体精度。

Note: IoU represents intersection over union, mIoU represents mean intersection over union, and OA represents overall accuracy.

图 6 展示了在同时使用高分数据和时序 NDVI 数据时，5 种模型在测试集上分割结果的部分示例。从左到右依次为 GF-2 原图、标签、PSPNet、U-Net、U-Net++、DeepLabV3+、MA-Net、SegFormer 及本文提出的 MCSNet，可以看出 PSPNet、U-Net、U-Net++存在较多破碎图斑，且农作物的边界分割效果不佳，DeepLabV3+、MA-Net、SegFormer 的整体分类效果较好，无明显破碎图斑现象，边缘分割效果有所提高，但由于数据融合方法过于简单，未能有效利用到时序 NDVI 数据，出现了较多误分现象，结果示例中展示了小麦-油菜与单季稻、虾稻与莲藕之间都有较多明显的误分情况。本文模型 MCSNet 不仅有最少的误分情况，其边缘分割效果与地块完整性也最为理想，说明模型结构在特征提取与特征融合上都具有一定优势。

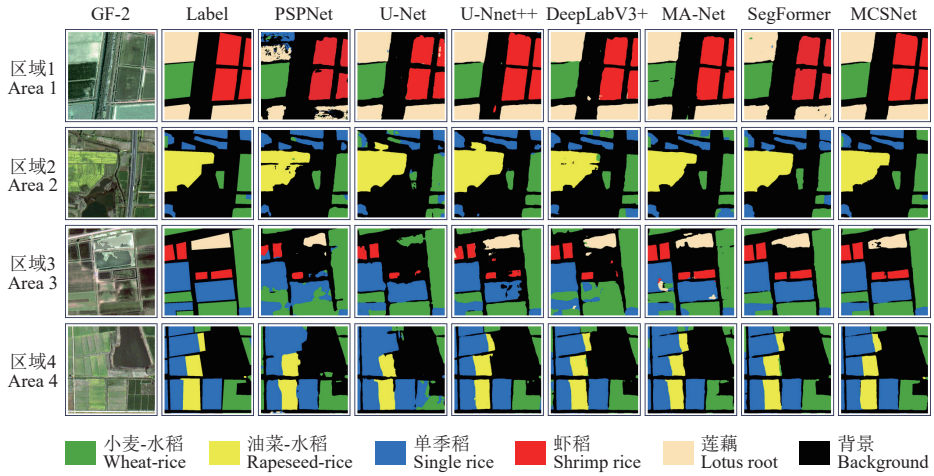


图 6　各模型结果对比图

Fig.6　Comparison of results from different model

3.3　消融试验

为了进一步验证本文所提 MCSNet 模型中引入的 ConvLSTM 单元及 RDAM 模块的有效性，本文设置了一组消融试验，对 4 个模型进行对比与评价：模型 1 删除了 RDAM 模块，且 ConvLSTM 单元用普通卷积层替代；模型 2 保留 RDAM 模块，ConvLSTM 单元用普通卷积层替代；模型 3 保留 ConvLSTM 单元，删除 RDAM 模块；模型 4 为本文所提出的 MCSNet。在相同条件下完成对 4 个模型的训练，图 7 为 4 个模型在测试集上输出特征的 t-SNE（t-distributed stochastic neighbor embedding）可视化结果，可以看出，在没有 RDAM 模块与 ConvLSTM 单元的情况下，数据点的分布比较散乱，且聚类中心距离较近，易产生误分现象。

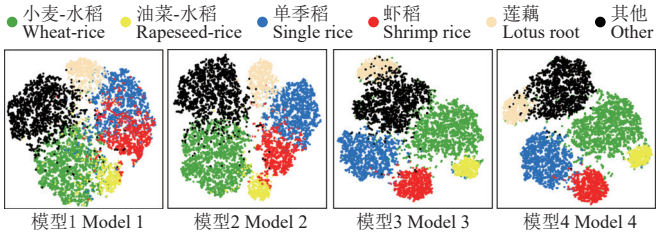


图 7　消融试验特征 t-SNE 可视化图

Fig.7　Visualization of t-SNE (t-distributed stochastic neighbor embedding) characteristics in ablation experiment

模型 2 在加入 RDAM 后，可以观察到各类别数据点聚类中心位置有一定的扩散效果，能为分类准确性带来正向影响。模型 3 在引入了 ConvLSTM 单元后，可以看出各类别数据点的聚集性有所提高，说明 ConvLSTM 单元提取到时空信息更为准确有效。本文提出的 MCSNet，综合了 ConvLSTM 结构及 RDAM 模块对特征提取带来的积极作用，特征点离散度较低，且聚类中心距离较远，使农作物种类能够得到更好的区分。表 3 为 4 种模型的定量评价结果，试验结果表明，RDAM 模块和 ConvLSTM 单元的引入对模型的分类效果有了较大的提高，其中 RDAM 主要作用是提高时空特征的融合效果，引入该模块后，模型 2 相对于模型 1 在 mIoU 和 OA 上分别提高了 0.28 和 0.85 个百分点，模型 4 相对于模型 3 在 mIoU 和 OA 上分别提高了 0.68 和 1.28 个百分点；ConvLSTM 结构的主要作用在于更加深入地挖掘时序 NDVI 数据中的时空信息，试验结果表明，引入 ConvLSTM 结构后，农作物分类精度有了很大提升，引入该模块后，模型 3 相对于模型 1 在 mIoU 和 OA 上分别提高了和 3.16 和 2.96 个百分点；模型 4 相对于模型 2 在 mIoU 和 OA 上分别提高了 3.56 和 3.39 个百分点。本文提出的 MCSNet，在 RDAM 模块和 ConvLSTM 单元的联合作用下，提高了模型在信息提取与信息融合的能力，对比无 RDAM 模块和 ConvLSTM 单元的模型 1，mIoU 和 OA 上分别提高了 3.84 和 4.24 个百分点。

表 3　消融试验定量评价结果

Table 3　Quantitative evaluation results of ablation experiments　　%

模型 Model	IoU						mIoU	OA
	小麦-水稻 Wheat - rice	油菜-水稻 Rapeseed- rice	单季稻 Single rice	虾稻 Shrimp rice	莲藕 Lotus root	背景 Background		
模型 1 Model 1	76.23	64.83	74.16	76.47	76.82	74.96	73.91	85.32
模型 2 Model 2	77.32	65.68	72.72	76.98	77.32	75.12	74.19	86.17
模型 3 Model 3	79.15	70.85	76.87	81.18	78.32	76.06	77.07	88.28
模型 4 Model 4	79.76	72.32	77.65	81.65	78.43	76.69	77.75	89.56

3.4　模型应用与分析

本文研究模型可以应用于研究区全域的农作物类别分类，获取研究区全域高分辨率农作物类别分布图。操作方法是将研究区高分影像与经过尺寸调整的时序 NDVI 影像在对应位置进行裁剪，将切片输入模型进行预测，最后将所有切片的预测结果拼接为研究区完整的农作物分类图。

为了验证本文方法在研究区农作物分类制图效果和精确度上的优越性，选择了近期结合深度学习的时空协同精细制图方法进行对比。黄翀等[35] 率先基于双向长短

期记忆网（bidirectional long short-term memory，Bi-LSTM）网络对时间序列遥感进行农作物分类，基于全年时序数据完成对黄河三角洲地区的农作物分类。张冬韵等[36]同样采用 Bi-LSTM 网络对时序遥感影像进行作物分类，同时引入高空间分辨率影像扩展出面向对象的 Bi-LSTM 分类方法，完成对宁夏引黄灌区进行种植结构提取。本文将加入 Bi-LSTM 的面向像素与面向对象方法进行对比，采用常应用于农作物制图精度评价的 OA、Kappa 系数、mF1(mean F1 score）和 mIoU 作为本次试验的评价指标。3 种方法在测试集上的精度评价结果见表 4，分类结果混淆矩阵见图 8。

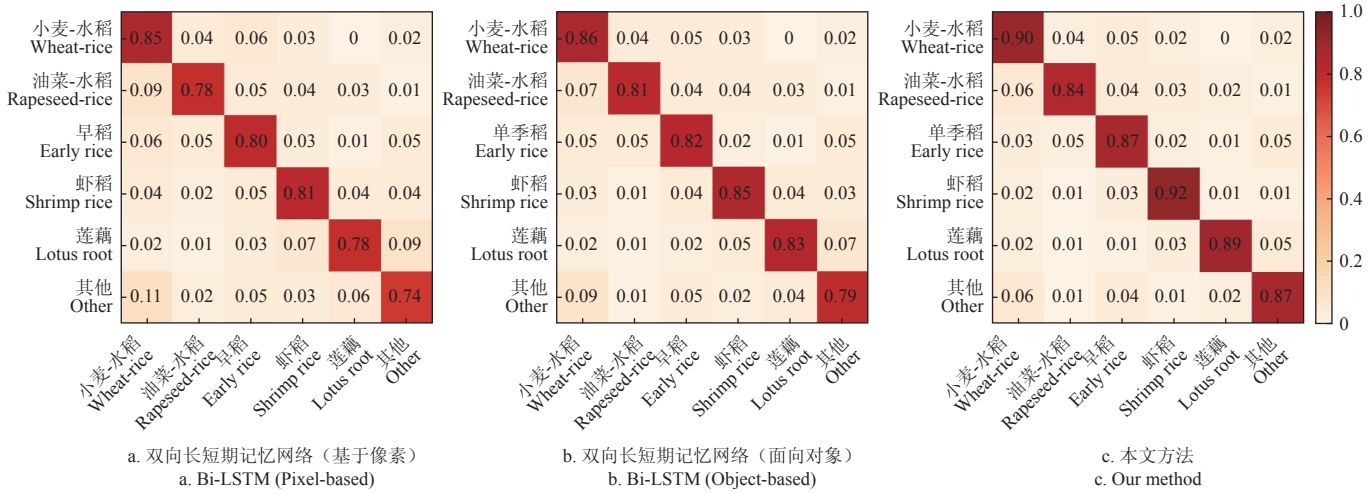
表 4 各方法的定量评价结果

Table 4 Quantitative evaluation results of each method

分类方法 Classification method	OA	Kappa	mF1	mIoU
双向长短期记忆网络（基于像素） Bi-LSTM (Pixel-Based)	0.79	0.75	0.79	0.66
双向长短期记忆网络（面向对象） Bi-LSTM (Object-Based)	0.83	0.80	0.83	0.71
本文方法 Our Method	0.89	0.85	0.89	0.78

注：OA 为总体精度，Kappa 为 Kappa 系数，mF1 为平均 F1 得分，mIoU 为平均交并比，

Note: OA represents overall accuracy, Kappa represents Kappa coefficient, mF1 represents mean F1 score, and mIoU represents mean intersection over union.



a. 双向长短期记忆网络（基于像素） a. Bi-LSTM (Pixel-based)

b. 双向长短期记忆网络（面向对象） b. Bi-LSTM (Object-based)

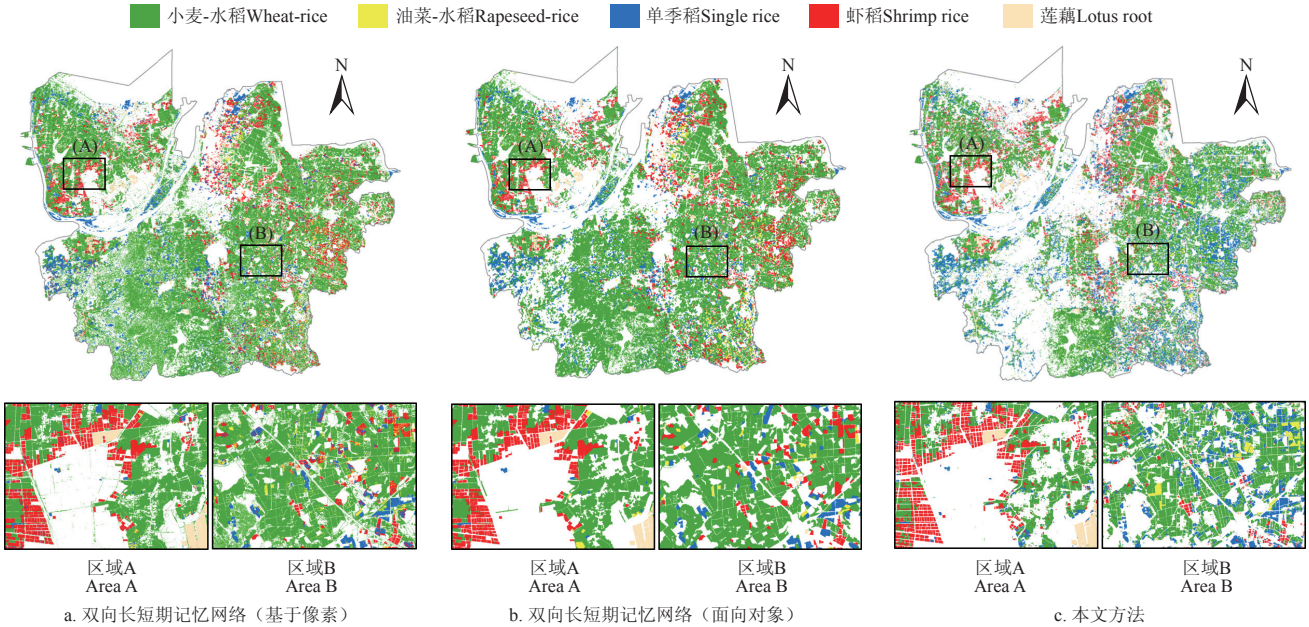
c. 本文方法 c. Our method

图 8 不同方法混淆矩阵

Fig.8 Confusing matrices with different methods

试验结果表明，本文分类方法的结果优于基于像素与面向对象的 Bi-LSTM 算法，测试结果 OA（0.89）、Kappa 系数（0.85）、mF1（0.89）和 mIoU（0.78）4 项指标均达到了最高。根据混淆矩阵可以直观看出预测准确率和错误分类情况，本文方法在小麦-水稻，小麦-油菜、单季稻、虾稻、莲藕这几类农作物的分类精度达到了 0.90、0.84、0.87、0. 92、0.89、0.87，效果优于所对比的两种网络。

使用 3 种方法得到了研究区高分辨率全域农作物分类图，随机选择了两块区域 A、B 进行细节展示（如图 9 所示）。



a. 双向长短期记忆网络（基于像素） a. Bi-LSTM (Pixel-based)

b. 双向长短期记忆网络（面向对象） b. Bi-LSTM (Object-based)

c. 本文方法 c. Our method

图 9 不同方法分类效果对比

Fig.9 Comparison of classification effects of different method

根据展示的结果可以看出，研究区最主要的种植模式为小麦-水稻，占据了较大的耕地面积，其次为单季稻与虾稻，两种农作物都具有较大的种植面积，莲藕和油菜-水稻种植类型所占面积则较少。通过 A 和 B 区域的细节展示可以观察到，基于像素的 Bi-LSTM 算法存在严重的“椒盐噪声”现象，这是由于基于像素的分类方法只考虑了像素本身的特征，没有利用到相邻像素在空间上的联系。面向对象的 Bi-LSTM 算法分类方法由于引入高分辨率影像分割的矢量边界，有效地解决了“椒盐噪声”问题，但仍存在较多错分现象，这是因为该方法的分类过程是两个独立的步骤，时序影像分类与高分辨率影像分割是两个独立的过程，不能同步挖掘多源数据所具有的时空信息。本文所提出的深度学习语义分割模型 MCSNet 的分类效果优于基于像素和面向对象的 Bi-LSTM 分类方法。MCSNet 不仅具有最高的分类精度，还通过所示区域的细节展示，可以观察到模型预测结果中农作物田块完整，田块之间清晰分隔，并且在不同区域表现出较强的泛化能力。本文方法展现了通过深度语义分割模型结合多源不同尺度的遥感时空数据进行农作物分类的在农业遥感监测中具有的潜力，为县域级别的农业生产管理、作物种植结构的优化调整以及农业的可持续发展，提供了理论基础和技术支持。

4　结　论

为了解决当前高分辨率农作物语义分割模型无法有效利用包含农作物物候信息的时序数据的问题，特别是在复杂种植结构区域（如存在单双季作物的情况）下进行农作物分类任务时所面临的困难与挑战，本文提出了一种能够融合高分辨率遥感影像和中分辨率时序 NDVI 的语义分割模型 MCSNet，模型的双编码器结构，能够做到对高分辨率影像空间细节与中分辨率时序影像时空特征的同步挖掘，并通过注意力机制引导的数据融合模块对时空信息进行融合，有效地融合了多源遥感时空数据，提高了农作物分类效果与精度。主要结论如下：

1）本文提出的 MCSNet 模型，分类效果和精度优于所对比的 PSPNet、U-Net、U-Net++、DeepLabV3+、MA-Net、SegFormer。在引入时序 NDVI 数据后，本文模型在小麦-水稻，小麦-油菜，单季稻、虾稻、莲藕这几类具有明显物候特征的作物类型的分类精度分别提高了 17.58、19.49、16.03、13.57、9.95 个百分点。

2）使用本文提出的 MCSNet 模型，在 ConvLSTM 结构单元和 RDAM 模块的联合作用下，精度得到了较大提高，消融试验结果表明在加入 ConvLSTM 结构单元和 RDAM 模块后，模型的分类结果 mIoU 和 OA 分别提高了 3.84 和 4.24 个百分点。

3）使用本文提出的 MCSNet 模型，完成了对盱眙县的农作物类别分类，并与基于像素和面向对象的 Bi-LSTM 分类方法进行了对比，无论是在制图效果还是分类精度方面，本文方法均展现出明显的优势，四项指标均达到了最高水平：总体精度为 0.89，Kappa 系数为 0.85，平均 F1 得分为 0.89，平均交并比为 0.78。

［参　考　文　献］

[1] LIU N, ZHAO Q, WILLIAMS R, et al. Enhanced crop classification through integrated optical and SAR data: A deep learning approach for multi-source image fusion[J]. International Journal of Remote Sensing, 2024, 45(19/20): 7605-7633.

[2] LONGCHAMPS L, PHILPOT W. Full-season crop phenology monitoring using two-dimensional normalized difference pairs[J]. Remote Sensing, 2023, 15(23): 5565.

[3] 陈媛媛，游炯，幸泽峰，等. 世界主要国家精准农业发展概况及对中国的发展建议[J]. 农业工程学报，2021，37(11)：315-324.
CHEN Yuanyuan, YOU Jiong, XING Zefeng, et al. Review of precision agriculture development situations in the main countries in the world and suggestions for China[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2021, 37(11): 315-324. (in Chinese with English abstract)

[4] TARIQ A, YAN J, GAGNON A S, et al. Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest[J]. Geo-Spatial Information Science, 2023, 26(3): 302-320.

[5] WANG H, YE Z, WANG Y, et al. Improving the crop classification performance by unlabeled remote sensing data[J]. Expert Systems with Applications, 2024, 236: 121283.

[6] 刘威，郑雪丽，马恒运. 农业数字化对粮食生产安全的影响机理与效应[J]. 中国农业大学学报，2024，29(7)：307-320.
LIU Wei, ZHENG Xueli, MA Hengyun. Influence mechanism and effect of agricultural digitization on food production security[J]. Journal of China Agricultural University, 2024, 29(7): 307-320. (in Chinese with English abstract)

[7] CHOUKRI M, LAAMRANI A, CHEHBOUNI A. Use of optical and radar imagery for crop type classification in Africa: A review[J]. Sensors, 2024, 24(11): 3618.

[8] SUN R, CHEN S, SU H, et al. The effect of NDVI time series density derived from spatiotemporal fusion of multisource remote sensing data on crop classification accuracy[J]. ISPRS International Journal of Geo-Information, 2019, 8(11): 502.

[9] ZHAO Z, ISLAM F, WASEEM L A, et al. Comparison of three machine learning algorithms using google earth engine for land use land cover classification[J]. Rangeland ecology & management, 2024, 92: 129-137.

[10] 李晓慧，王宏，李晓兵，等. 基于多时相 Landsat 8 OLI 影像的农作物遥感分类研究[J]. 遥感技术与应用，2019，34(2)：389-397.
LI Xiaohui, WANG Hong, LI Xiaobing, et al. Study on crops remote sensing classification based on multi-temporal Landsat 8 OLI images[J]. Remote Sensing Technology and Application, 2019, 34(2): 389-397. (in Chinese with English abstract)

[11] KHALIQ A, PERONI L, CHIABERGE M. Land cover and crop classification using multitemporal Sentinel-2 images based on crops phenological cycle[C]//2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems

(EESMS). Salerno, Italy:IEEE,2018: 1-5.

[12] 李德仁，王密. 高分辨率光学卫星测绘技术综述[J]. 航天返回与遥感，2020，41(2)：1-11.
LI Deren, WANG Mi. A review of high resolution optical satellite surveying and mapping technology[J]. Spacecraft Recovery & Remote Sensing, 2020, 41(2): 1-11. (in Chinese with English abstract)

[13] 徐新刚，李强子，周万村，等. 应用高分辨率遥感影像提取作物种植面积[J]. 遥感技术与应用，2008，1：17-23.
XU Xingang, LI Qiangzi, ZHOU Wancun, et al. Classification application of QuickBird imagery to obtain crop planting area[J]. Remote Sensing Technology and Application, 2008,1: 17-23. (in Chinese with English abstract)

[14] ESETLILI M T, BALCIK F B, SANLI F B, et al. Comparison of object and pixel-based classifications for mapping crops using Rapideye imagery: A case study of Menemen Plain, Turkey[J]. International Journal of Environment and Geoinformatics, 2018, 5(2): 231-243.

[15] BHOSLE K, MUSANDE V. Evaluation of deep learning CNN model for land use land cover classification and crop identification using hyperspectral remote sensing images[J]. Journal of the Indian Society of Remote Sensing, 2019, 47(11): 1949-1958.

[16] KHAKI S, WANG L, Archontoulis S V. A CNN-RNN framework for crop yield prediction[J]. Frontiers in Plant Science, 2020, 10: 201901750.

[17] GAFUROV A, MUKHARAMOVA S, SAVELIEV A, et al. Advancing agricultural crop recognition: The application of LSTM networks and spatial generalization in satellite data analysis[J]. Agriculture, 2023, 13(9): 1672.

[18] 马妍，古丽米拉·克孜尔别克. 图像语义分割方法在高分辨率遥感影像解译中的研究综述[J]. 计算机科学与探索，2023，17(7)：1526-1548.
MA Yan, GULIMILA Kezierbieke. Research review of image semantic segmentation method in high-resolution remote sensing image interpretation[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(7): 1526-1548. (in Chinese with English abstract)

[19] YAN G, JING H, LI H, et al. Enhancing building segmentation in remote sensing images: Advanced multi-scale boundary refinement with MBR-HRNet[J]. Remote Sensing, 2023, 15(15): 3766.

[20] WANG X, JING S, DAI H, et al. High-resolution remote sensing images semantic segmentation using improved UNet and SegNet[J]. Computers and Electrical Engineering, 2023, 108: 108734.

[21] ZHANG L, LU Y, SHI R, et al. Towards interpretability lightweight semantic segmentation model for waterbody extraction in large-scale high resolution remote sensing images[J]. International Journal of Remote Sensing, 2024, 45(8): 2721-2738.

[22] 张新长，黄健锋，宁婷. 高分辨率遥感影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. 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 with English abstract)

[23] 董秀春，刘忠友，蒋怡，等. 基于 WorldView-2 影像和语义分割模型的小麦分类提取[J]. 遥感技术与应用，2022，37(3)：564-570.
DONG Xiuchun, LIU Zhongyou, JIANG Yi, et al. Winter wheat extraction of WorldView-2 image based on semantic segmentation method[J]. Remote Sensing Technology and Application, 2022, 37(3): 564-570. (in Chinese with English abstract)

[24] XIANG J, LIU J, Du Chen, et al. CTFuseNet: A multi-scale CNN-Transformer feature fused network for crop type segmentation on UAV remote sensing imagery[J]. Remote Sensing, 2023, 15(4): 15041151.

[25] LU T, WAN L, WANG L. Fine crop classification in high resolution remote sensing based on deep learning[J]. Frontiers in Environmental Science, 2022, 10: 991173.

[26] XU C, GAO M, YAN J, et al. MP-Net: An efficient and precise multi-layer pyramid crop classification network for remote sensing images[J]. Computers and Electronics in Agriculture, 2023, 212: 108065.

[27] WANG S, ZHU Y, ZHENG N, et al. Change detection based on existing vector polygons and up-to-date images using an attention-based multi-scale ConvTransformer network[J]. Remote Sensing, 2024, 16(10): 16101736.

[28] ATIK S O, ATIK M E, IPBUKER C. Comparative research on different backbone architectures of DeepLabV3+ for building segmentation[J]. Journal of Applied Remote Sensing, 2022, 16(2): 024510.

[29] WANG M, ZHANG X, NIU X, et al. Scene classification of high-resolution remotely sensed image based on ResNet[J]. Journal of Geovisualization and Spatial Analysis, 2019, 3: 16-24.

[30] KARTAL S, IBAN M C, SEKERTEKIN A. Next-level vegetation health index forecasting: A convLSTM study using MODIS time series.[J]. Environmental Science and Pollution Research International, 2024, 31(12): 18932-18948.

[31] TSUNG-YI L, PRIYA G, ROSS G, et al. Focal loss for dense object detection.[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.

[32] ZHAO R, QIAN B, ZHANG X, et al. Rethinking dice loss for medical image segmentation[C]//2020 IEEE International Conference on Data Mining (ICDM). Sorrento, Italy:IEEE, 2020: 851-860.

[33] FAN T, WANG G, LI Y, et al. MA-Net: A multi-scale attention network for liver and tumor segmentation[J]. IEEE Access, 2020, 8: 179656-179665.

[34] XIE E, WANG W, YU Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.

[35] 黄翀，侯相君. 基于 Bi-LSTM 模型的时间序列遥感作物分类研究[J]. 中国农业科学，2022，55(21)：4144-4157.
HUANG Chong, HOU Xiangjun. Crop classification with time

series remote sensing based on Bi-LSTM model[J]. Scientia Agricultura Sinica, 2022, 55(21): 4144-4157. (in Chinese with English abstract)

[36] 张冬韵，吴田军，骆剑承，等. 时空协同的农业种植结构遥感精细制图[J]. 遥感学报，2024，28(8)：2014-2029. ZHANG Dongyun, WU Tianjun, LUO Jiancheng, et al. Precise crop planting structure mapping method based on spatial-temporal collaboration of remote sensing[J]. National Remote Sensing Bulletin, 2024, 28(8): 2014-2029. (in Chinese with English abstract)

A crop semantic segmentation model integrating high-resolution imagery and time-series NDVI

ZHAO Xu[1] , LI Hao[1※] , ZHU Yihu[2] , WANG Shengli[2,3] , HE Yanlan[2]

(1. *School of Earth Science and Engineering*, *Hohai University*, *Nanjing* 211100, *China*;　2. *Jiangsu Provincial Geological Surveying and Mapping Brigade*, *Nanjing* 211102, *China*;　3. *School of Environment and Spatial Informatics*, *China University of Mining and Technology*, *Xuzhou* 221116, *China*)

Abstract: High spatial resolution of remote sensing imagery has been increasing with the optical sensor performance. An accurate and rapid classification of the crops is often required for agricultural production, yield prediction, and structure adjustment. However, the traditional high-resolution imagery cannot fully meet the rich phenological information in the crop growth cycle, particularly in the complex planting structure regions with both single- and double-season crops. This limitation can significantly constrain the performance of the high-resolution crop semantic segmentation models. This study aims to propose a multi-source crop semantic segmentation model— MCSNet (multi-source crops segmentation network)— that integrates the high-resolution remote sensing imagery with the medium-resolution time-series normalized difference vegetation index (NDVI) data. A dual-encoder structure was composed of a high-resolution encoder (HR-Decoder) and a time-series encoder (TS-Decoder). The HR-Decoder was targeted at the high-resolution imagery in order to extract the spatial detail features, such as the crop plot boundaries and texture differences. Meanwhile, the TS-Decoder was focused on the medium-resolution time-series NDVI data. The vegetation indices were utilized to sensitively capture the spectral variations of crops throughout their growth cycles, thereby fully exploiting the phenological features to distinguish between single- and double-season crops. Furthermore, the network incorporated the convolutional long short-term memory (ConvLSTM) units within the TS-Decoder. The modeling capacity was enhanced for the complex temporal information. The local spatial features were extracted to effectively capture the dynamic changes along the temporal dimension. Subsequently, a multi-feature fusion encoder (MF-Encoder) was integrated to fuse the multi-source features from the HR-Decoder and TS-Decoder. The residual double attention was also utilized to emphasize the importance of the critical feature channels and spatial positions. Thereby, the temporal features and high-resolution spatial details were fused to ultimately strengthen the precision and robustness of the crop classification. The time-series NDVI data were also integrated with the MCSNet in the experimental phase. The accuracy of the crop classification was significantly improved, compared with the traditional only on the high-resolution imagery. The comparative experiments showed that the MCSNet shared the outstanding performance, with the mean intersection over union (mIoU) of 77.75% and an overall accuracy (OA) of 89.56%, indicating the highest levels. Furthermore, the ConvLSTM and the residual double attention in the network enhanced the modeling capability of the spatiotemporal features, thus increasing mIoU and OA by 3.84 percentage point and 4.24 percentage point, respectively. The MCSNet model was applied to the large-scale and complex study area of Xuyi County, Huaian City, Jiangsu Province, China. According to the pixel- and object-oriented classification, like Bi-LSTM (bidirectional long short-term memory), MCSNet exhibited significant advantages in both mapping and classification accuracy. Specifically, the MCSNet achieved an OA of 89%, a Kappa coefficient of 0.85, a mean weighted F1 score (mF1) of 0.89, and an mIoU of 0.78, thus outperforming the comparative data across all metrics. Therefore, there were the effectiveness and practicality of the MCSNet for the classification tasks in the large-scale and complex planting structure regions. In summary, the MCSNet can offer a viable technical pathway for multi-source data processing by integrating high-resolution imagery and time-series NDVI data. The dual-encoder structure (ConvLSTM) and residual double attention module (MCSNet) were introduced to effectively enhance the crop classification accuracy and stability in the complex planting structure regions. This finding can also provide a strong theoretical and technical solution to the multi-source remote sensing data fusion for crop production and structure optimization in sustainable agriculture.

Keywords: crop; classification; high-resolution remote sensing; time-series NDVI; semantic segmentation; deep learning