

深度学习作物分类模型空间泛化能力分析

盖爽^{1,2,3,4}, 张锦水^{1,3,4}, 朱爽⁵

1. 北京师范大学 地理科学学部 遥感科学国家重点实验室, 北京 100875;

2. 南京大学 江苏省地理信息技术重点实验室/国际地球系统科学研究所, 南京 210023;

3. 北京师范大学 地理科学学部 北京市陆表遥感数据产品工程技术研究中心, 北京 100875;

4. 北京师范大学 地理科学学部遥感科学与工程研究院, 北京 100875;

5. 北京工业职业技术学院, 北京 100042

摘要: 大数据驱动训练的深度学习模型是当今农作物分类的最新方法。当前研究仍然主要关注该模型方法的创新性, 其在特定时间、特定地区的作物分类模型的泛化能力分析经常被忽略。因此, 提高遥感分类模型在大尺度空间范围的有效迁移能力是遥感技术支撑地球系统科学研究和社会应用的关键。本研究通过设计实验分析了模型架构、作物物候特征、农区地块尺度、数据类型等因素对作物分类模型泛化能力的影响。结果表明: 一方面当训练区和测试区地块大小发生明显变化时, MultiResUNet 相对于 SegNet, DeepLab V3+ 和 U-Net 具有更好的泛化性能。然而, 单纯依靠 MultiResUNet 的泛化能力依然无法完全克服地块空间形态的变化对模型迁移的不利影响, 为获得更高精度的华北玉米分布信息, 需要优先使用与华北地区农业景观更相似的东北作物分布数据产品进行深度学习模型训练; 另一方面, 相对于 TOA (Top of Atmosphere) 数据, 采用 SR (Surface Reflectance) 数据更有利于模型在跨洲际尺度进行空间迁移, 因此, 在大尺度作物制图研究中, 应优先考虑使用 SR 数据。综上, 本研究从一定程度上验证了影响农作物分类模型迁移性能的内在因素, 可为大尺度作物制图提供科学依据。

关键词: 模型泛化, 深度学习, SegNet, DeepLab V3+, U-Net, MultiResUNet, 作物制图

中图分类号: TP701/P2

引用格式: 盖爽, 张锦水, 朱爽. 2023. 深度学习作物分类模型空间泛化能力分析. 遥感学报, 27(12): 2796–2814

Ge S, Zhang J S and Zhu S. 2023. Spatial generalization ability analysis of deep learning crop classification models. National Remote Sensing Bulletin, 27(12): 2796–2814 [DOI: 10.11834/jrs.20221408]

1 引言

精准、及时的全球作物制图对于粮食安全评估具有重要意义 Wang 等 (2019)。随着高分辨遥感卫星观测能力不断增强, 利用卫星遥感数据并结合机器学习算法成为主流的作物制图范式 (Azzari 和 Lobell, 2017), 即“特征—样本—分类器”模式。但是这些分类模型往往只针对一个特定地区或者特定时期 (Zhong 等, 2014; 姜伊兰等, 2021), 在大范围空间尺度的泛化性能缺少全面的评估 (Zhang 等, 2019)。此外, 由于训练样本的获取和标注费时费力, 甚至在有些地区难以获得真值样本 (Song 等, 2017), 这是作物自动化

分类面临的最大挑战。因此, 开发具有较强时空泛化性能的模型十分必要, 即在样本充足可靠的地区训练模型, 然后将其应用到样本缺乏或者获取困难的地区进行作物制图, 以辅助当地粮食生产决策 (Hao 等, 2020)。

传统分类方法依靠人工选取特征, 这些特征并不一定是有利于区分地物类别的最优特征, 其泛化性能能力弱 (Zhong 等, 2019; 杨蜀秦等, 2021)。近 10 年来, 深度学习在海量标记样本 (大数据) 和并行计算的支持下, 迅速发展起来, 能够从大数据中自动学习特征 (LeCun 等, 2015), 不再需要人为干预。卷积神经网络是深度学习领域应用最广泛的网络架构之一, 由于其可以从训

收稿日期: 2021-06-21; 预印本: 2022-07-11

基金项目: “高分辨率对地观测系统重大专项支持项目”民用部分 (编号: 20-Y30F10-9001-20/22)

第一作者简介: 盖爽, 研究方向为农业遥感、作物分类模型泛化。E-mail: 201921051175@mail.bnu.edu.cn

通信作者简介: 张锦水, 研究方向为农业遥感。E-mail: zhangjs@bnu.edu.cn

练集中自动学习多层次代表性的深度特征,因此具有强泛化性能,已广泛应用于高分辨率遥感影像目标检测和语义分割任务当中(Li等, 2019; Zhong等, 2020; 赵红伟等, 2020; 李前景等, 2021)。

在遥感影像分类中,模型泛化能力一般指模型在训练区之外的其他区域或其他时期的测试数据上的分类性能。当前进行作物分类模型泛化能力的研究虽然取得了一定进展,但是依然存在局限性。该类研究通常采用固定长度的时间序列数据,使用传统机器学习或深度学习方法,后者的泛化性能往往更优(Zhong等, 2019; Xu等, 2020)。由于不同地区作物生长周期的长短不一致,因此这些研究的分类模型的泛化能力有限。已有研究通过增添不同年份(Konduri等, 2020)或者不同区域(Hamrouni等, 2021)的训练样本来提升模型在测试区的性能,但是在跨时间或空间的样本中目标作物与非目标作物的混淆的可能会增大,并且这些样本通常难以获取。此外,已有研究通过修改分类模型来提升模型的泛化能力(Xu等, 2020; Zhang等, 2020),但是到目前为止并没有公认的泛化性能最强的作物分类模型。由于在大尺度区域,作物的物候、种植结构和地块尺度等特征会发生显著变化,这使得同种作物在遥感影像上呈现的特征出现差异,限制了模型的泛化性能(Kluger等, 2021)。

作物分类模型泛化研究往往从“样本”、“特征”或“分类器”3个方面入手(图1(a))。通过增加样本数量或提高样本质量可以提升样本的代表性,从而提升模型在测试区的分类性能(Hamrouni等, 2021)。在样本质量方面,CNN可以在一定程度上削弱具有噪声样本的不利影响(Hamdi等, 2019; Kattenborn等, 2020)。在样本数量方面,如果目标类别在遥感数据中易于区分,则可能需要较少的样本;为区分那些仅在细微特征上存在差异的类别,可用样本的数量可能是模型训练收敛成功的关键因素(Kattenborn等, 2021)。Weinstein等(2020)在利用深度学习方法进行树冠识别的研究中得出分类准确性首先随着样本量的增加而迅速增加,随后不再增加。Liu等(2018)发现样本数量对不同架构模型的影响不同。因此,训练样本量对模型泛化能力的影响不能一概而论。

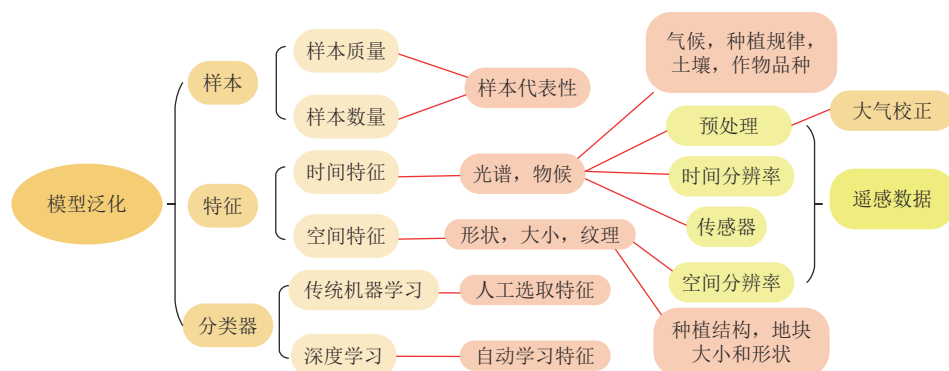
在特征方面,作物相对于其他的土地类别呈

现特有的特征,例如在时间特征方面,作物生长有着明显的光谱变化或物候规律,具体表现为作物的植被指数曲线在生长周期内随着时间的推移呈现规律的变化。由图1(b)可见作物物候在大尺度地区存在空间异质性,这是由于不同地区的气候条件、土壤状况、种植规律以及作物品种等因素的差异,很大程度上影响了作物分类模型的空间泛化能力(Xu等, 2020)。Ge等(2021)分析了不同地区作物物候变化对模型泛化能力的影响,发现物候变化使得同一作物在不同地区同一日期的遥感影像上呈现的光谱特征具有异质性,严重影响着模型的泛化性能,因此提出将各地区数据合成时间窗口进行平移的方法,以实现模型在不同地区的有效迁移。在空间特征方面,作物地块内部光谱特征通常较均质,因此呈现出一定的形状、大小和纹理,但是这些空间特征也具有空间异质性,例如不同地区的地块形状和大小不一致,详见图1(c)。一部分地区种植单一作物,而另一部分地区同时种植多种作物,因此,这些地区作物的纹理和上下文信息存在差异,这也会影响模型的泛化性能(王协等, 2020)。此外,不同的遥感数据捕捉到的作物时空特征不同。高时间分辨率的遥感数据能够捕捉作物完整的生长过程,并且随着时间序列数据的增加,模型的性能也会提升(Zhang等, 2018; Xu等, 2020)。高空间分辨率的遥感数据能够捕捉更精细的空间信息,如作物地块边界和邻域信息。Neupane等(2019)发现当数据空间分辨率从40 cm下降到60 cm时,香蕉棕榈树的识别精度下降了20%。Schiefer等(2020)在树种识别过程中,发现当数据空间分辨率从2 cm下降到8 cm时,识别结果的F1值下降了25%。不同传感器数据以及数据经过大气校正后,同种地物的反射率也会存在差异。在作物制图方面,多光谱数据往往优于多时相合成孔径雷达(SAR)数据,而将这两者结合可以提升模型性能(Nezami等, 2020)。在“分类器”方面,从依靠人工建立特征的传统机器学习算法到由端到端自动学习深度特征的深度学习模型转变,使得分类模型的泛化性能得到了显著提升(Zhong等, 2019)。

综上,虽然现有的针对影响模型泛化能力因素的研究虽然取得了一定的进展,但是在如何解决训练样本缺失问题、不同深度学习模型的作物分类性能、数据类型变化和地块大小对模型泛化

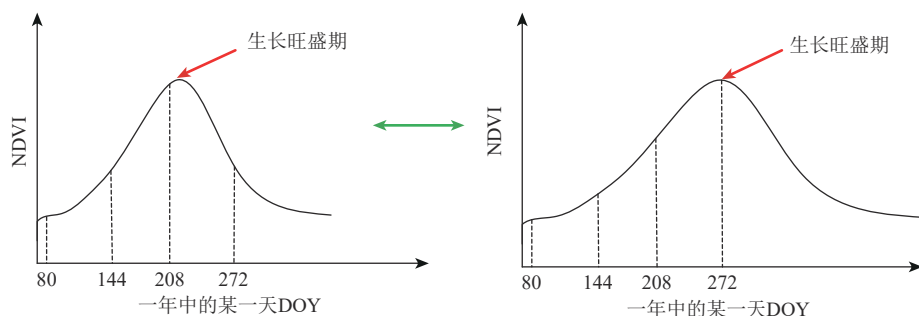
能力的影响等方面缺少充分的讨论。为进一步总结影响作物分类模型性能的因素,从而为大尺度作物制图研究提供参考,本文重点关注的问题包括:(1)利用现有的作物分布数据产品作为深度学习模型的真值样本来解决深度学习模型训练样本缺乏的问题,分析利用美国 Cropland Data Layer

(CDL)以及 You 等(2021)发布的东北作物分布数据产品所训练的模型在中国华北的适用性、不同网络架构的深度学习模型的泛化性能以及地块大小和数据类型变化对模型泛化能力的影响,以期为大尺度作物制图提供基础。



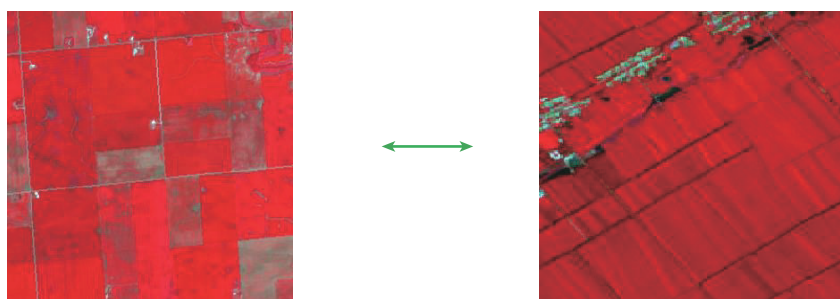
(a) 作物模型泛化能力影响因素

(a) Factors affecting the generalization capability of crop models



(b) 时间特征差异

(b) Differences in temporal characteristics



(c) 空间特征差异

(c) Differences in spatial characteristics

图1 作物分类模型泛化能力影响因素及时间和空间特征差异

Fig. 1 Factors affecting the generalization ability of crop classification models and differences in temporal and spatial characteristics

2 数据与方法

2.1 研究区

在中国吉林省和美国伊利诺伊州各选择一个玉米种植区域作为训练区,在中国河北省及江苏

省各选择一个玉米种植区域作为测试区,用于提取测试区的夏播玉米。每个训练区和测试区范围均为 50 km×50 km (图2)。

训练区 T1 处于中国东北地区的吉林省,年均降水量为 480—600 mm,夏季高温多雨。该省土壤

肥沃，有利于玉米种植，玉米产量占全国总产量的15%左右（齐晓宁 等，2002）。该省玉米为春播

玉米，播种日期为4月下旬—5月上旬，8月下旬开始收获。

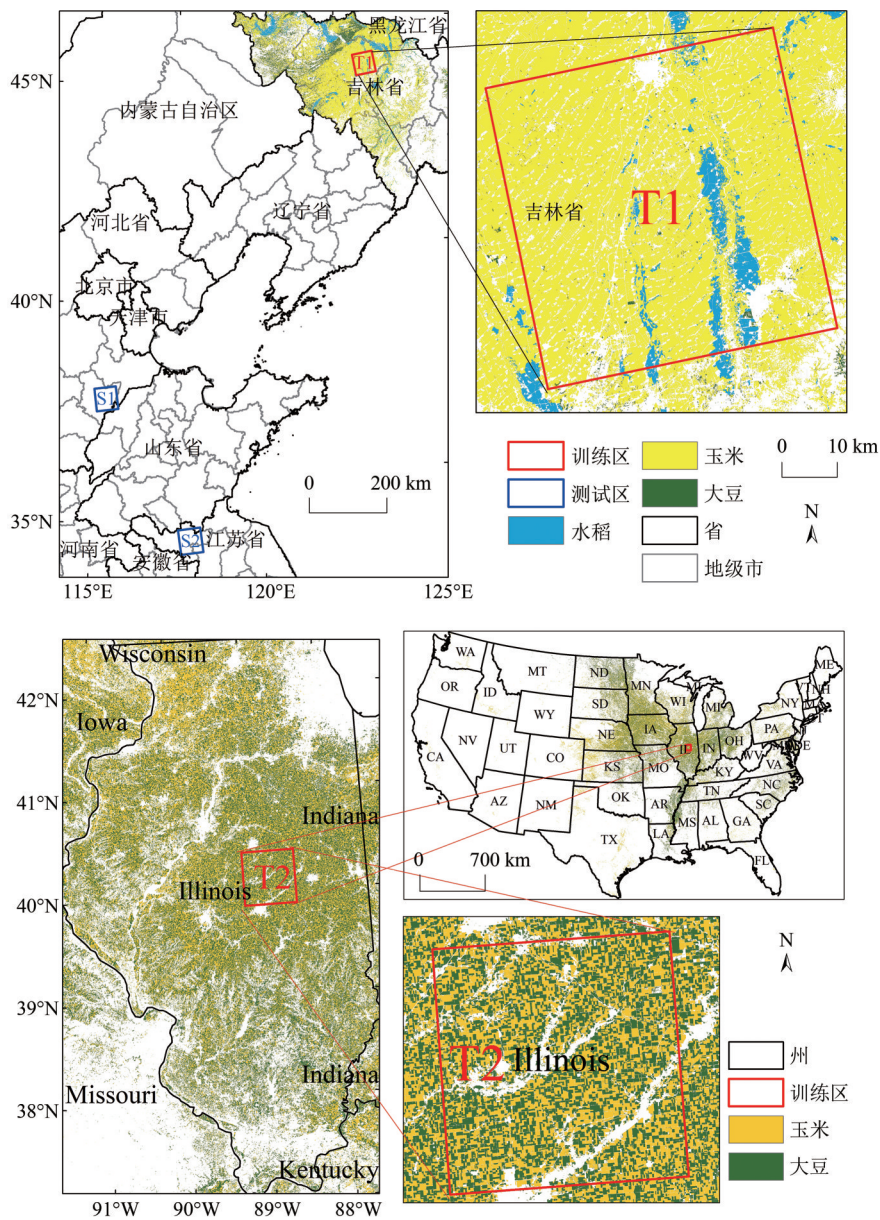


图2 研究区分布概况图

Fig. 2 Location overview map of the study areas

训练区T2位于美国中西部地区的伊利诺伊州，是美国农产品出口的主要来源地。该州南部地区年平均降水为1200 mm左右，北部为890 mm左右。夏季平均气温23℃左右，冬季平均气温0℃左右。该州种植的两大农作物是玉米和大豆，通常连年轮作，在雨养条件下种植（Diao, 2020）。该地区玉米的生长周期为4月下旬到9月中旬（Xu 等，2020）。

测试区S1位于河北省，玉米是该省主要的粮

食作物。该省属暖温带半干旱半湿润大陆性季风气候，年均气温约为10℃，年降水量为341—750 mm（李元华和车少静，2005）。河北省北部种植春播玉米，南部种植夏播玉米。春玉米在4月下旬—5月上旬播种，8月下旬开始收获。夏玉米为6月播种，10月中下旬收获。

测试区S2位于江苏省，玉米是该省第3大粮食作物。该省属亚热带季风气候，年均降水量为约1022 mm，年均气温约为15.1℃（傅帅 等，

2017)。该省种植春玉米和夏玉米，春玉米在4月播种，在7月底到8月中旬收获。夏玉米在6月份播种，在9月底到10月初收获。

2.2 数据

现有的进行作物分类模型泛化性研究的方法多采用固定长度时间序列数据，而欠考虑作物生长周期在不同地区的长短变化以及时间上的偏移，因此这些模型在大尺度地区进行作物制图存在局限性(Xu等, 2020)。除此之外，在大尺度地区进行作物制图研究还面临一系列难题，例如不同地区同一作物具有不同物候，而不同作物却具有相似的物候，这导致了在不同地区、同一时间获取的遥感影像上，同一作物会呈现不同的特征，而不同作物会呈现相似的特征。因此，为实现模型在不同地区有效地识别同一作物，需要在不同地区灵活采用不同时期获取的影像，以保证同一作物在不同地区所筛选的影像上呈现相似的特征，从而帮助分类模型能够准确地识别作物。

运用从谷歌地球引擎GEE (Google Earth Engine) 下载的2019年MODIS (Moderate Resolution Imaging Spectroradiometer) MCD43A4数据产品用于确定玉米物候，以及Sentinel-2卫星影像用于分类。其中MCD43A4为经过二向反射分布函数调整的地表反射率数据，空间分辨率为500 m，时间分辨率为1 d。该数据具有时空一致性，对大气噪声和传感器噪声具有鲁棒性(Diao, 2020)。本研究在每个研究区玉米种植广泛且连续的地区采集100个玉米样本点，并通过MCD43A4计算归一化差值植被指数NDVI (Normalized Difference Vegetation Index)，结合SG滤波得到时间序列曲线来确定玉米物候信息(图3)。如图3所示，不同地区玉米的NDVI曲线的形态存在着明显差异，并且在时间上存在偏移。本研究在不同地区作物生长周期内的NDVI曲线的最低值对应的日期范围内筛选无云Sentinel-2影像作为播种期影像，NDVI曲线的最高值对应的日期范围内筛选无云Sentinel-2影像作为生长旺盛期影像。将这2期影像作为训练和测试数据，以保证训练数据与测试数据尽可能相似，从而削弱因不同地区物候差异导致的模型泛化受限。其中训练区T1的2期影像的筛选日期分别为2019年5月29日和7月28日，训练区T2分别为2019年4月26日和7月25日，测试区S1分别为2019年6月14日和

8月21日，测试区S2分别为2019年6月11日和8月20日。由此可知T1、T2的影像日期相对于S1和S2均有所提前，表明训练区与测试区之间存在显著的物候差异。

Sentinel-2数据大气表观TOA (Top of Atmosphere) 反射率和地表反射率SR (Surface Reflectance) 2种数据分别进行模型训练和测试，来对比分析数据类型变化对模型泛化能力的影响。每景影像选取6个波段(蓝、绿、红、近红、短波红外1和短波红外2)，并将其空间分辨率重采样为10 m，作为模型输入。其中生长旺盛期影像的短波红外波段有利于区分玉米和大豆(Cai等, 2018)。

深度学习模型通常需要大量训练样本来驱动(Rußwurm和Körner, 2017)。为解决深度学习模型训练所需地面真值样本不足的问题，已有研究利用作物类别数据产品作为真值标签进行深度学习模型训练，并证明该类模型具有一定的泛化能力(Xu等, 2020)。其中，主要使用作物类别产品为CDL (Cropland Data Layer)。CDL是由美国农业部USDA (United States Department of Agriculture) 发布的土地覆盖图，其提供了美国本土非作物类型和100多种作物类型的高精度空间分布信息，包括玉米、大豆和水稻。从2008年开始提供覆盖全国30 m空间分辨率制图产品(Boryan等, 2011)。CDL是目前允许任意访问、免费共享的精度最高的作物类型分类产品(Cai等, 2018)。并且，已有相关研究表明利用CDL作为地面真实数据用于作物制图是可靠的(Xu等, 2020)。在中国，You等(2021)制作并发布了2017年—2019年中国东北地区的水稻，玉米和大豆的空间分布数据产品。该产品是中国首套10 m空间分辨率的区域年度作物制图，其总体精度在0.81—0.86。然而，基于该数据进行深度学习模型训练并尝试将其应用在我国其他地区的可行性需要进一步研究。因此，本研究使用2019年美国CDL数据(重采样为10 m)和2019年中国东北作物分布数据作为训练数据的标签，对比其所训练模型的性能。本研究类别体系为“玉米”、“其他类型”和“无数据”。

本研究通过在每个测试区进行实地调查获取测试样本点，用于精度验证。其中测试区S1的“玉米”和“其他类型”的测试样本点数分别为230和246，测试区S2的2种测试样本点数分别为215和219。将T1和T2内的训练数据裁剪成256×

256个像元大小并互相有25%重叠的tiles,以降低在训练过程中对内存的需求(Wieland等, 2019)。最后,将训练区的数据划分成训练(80%)和验证数据(20%),样本数分别为2864和713。在模型

训练过程中,使用Adam(Kingma和Ba, 2014)来优化更新权重,设置初始学习率为0.00001, batch size为10,模型训练30个epoch。

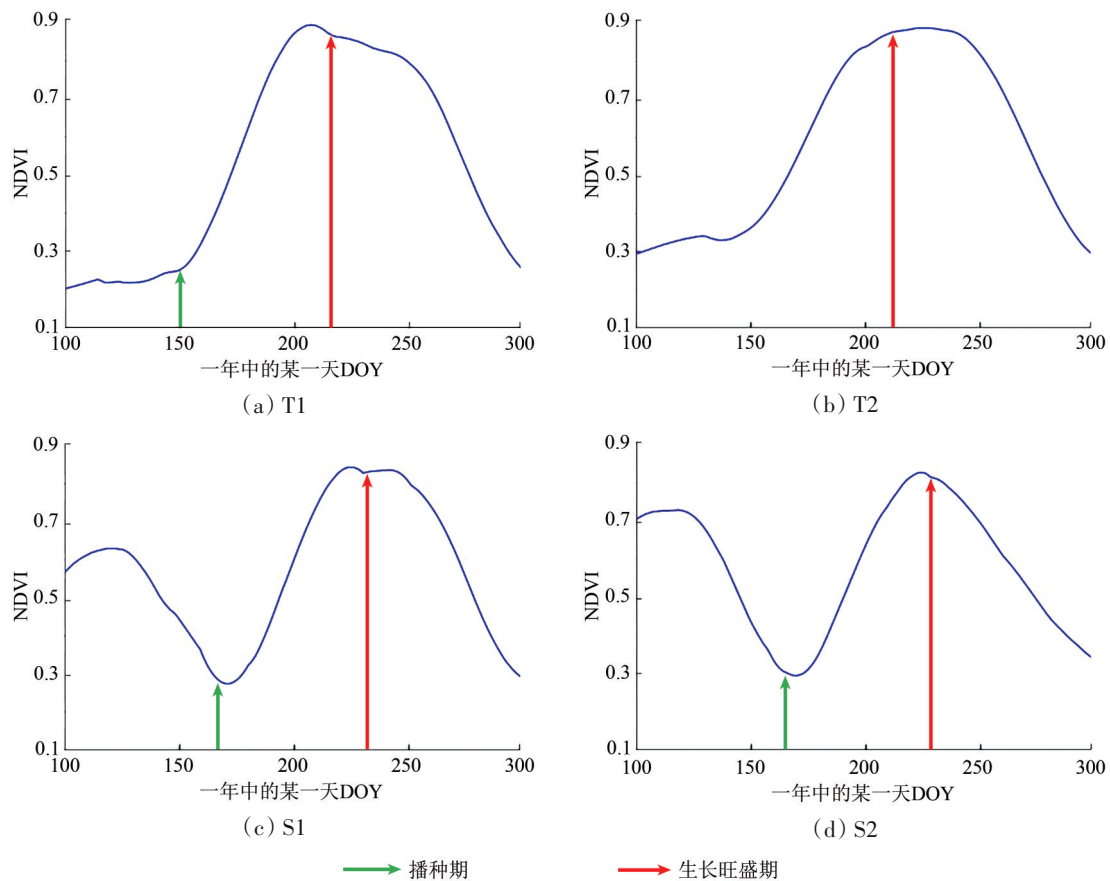


图3 各地区玉米时间序列NDVI曲线(箭头指示关键期影像日期)

Fig. 3 NDVI curve of corn time series in various regions (Arrows indicate the date of the image during critical period)

2.3 方法

本研究技术流程图详见图4(a)。具体流程如下:首先,利用MCD43A4数据产品计算各个地区玉米的NDVI时间序列曲线,以确定播种期和生长旺盛期;然后在这2个关键期内筛选无云Sentinel-2影像,结合CDL数据或You等(2021)发布的作物产品作为真值标签进行模型训练。将训练完成的模型在测试区S1和S2进行夏玉米制图,并结合野外采集的样本点进行精度验证。

2.3.1 分类模型

卷积神经网络CNNs(Convolutional Neural Networks)结合卷积层和池化层,直接从原始图像中提取特征,能够在多个上下文级别处理输入数据,与传统机器学习算法相比具有较强的泛化能

力(Zhu等, 2017; Zhang等, 2020)。本研究采用广泛应用于遥感影像分类的CNN架构SegNet、DeepLabv3+、U-Net,以及U-Net的衍生网络“多分辨率U-Net”(Multi-resolution U-Net, MultiResUNet),测试这4个网络在不同地区的泛化能力。

U-Net架构最初用于生物医学图像分割,选择U-Net因为它能够从非常少的数据中学习,并且在语义分割基准数据集上实现最高水准的结果(Ronneberger等, 2015),有利于解决大尺度作物制图所面临的因样本缺乏而导致分类精度较低的问题。U-Net架构(图5)由编码器和解码器两部分组成。编码器有3个子模块,每个子模块包含2个卷积层,每个子模块之后有一个通过最大池化实现的下采样层。解码器包含3个子模块,分辨率通

过上采样操作依次上升,直到与输入图像的分辨率一致。该网络还使用了跳跃连接,将上采样结

果与编码器中具有相同分辨率的子模块的输出进行连接,作为解码器中下一个子模块的输入。

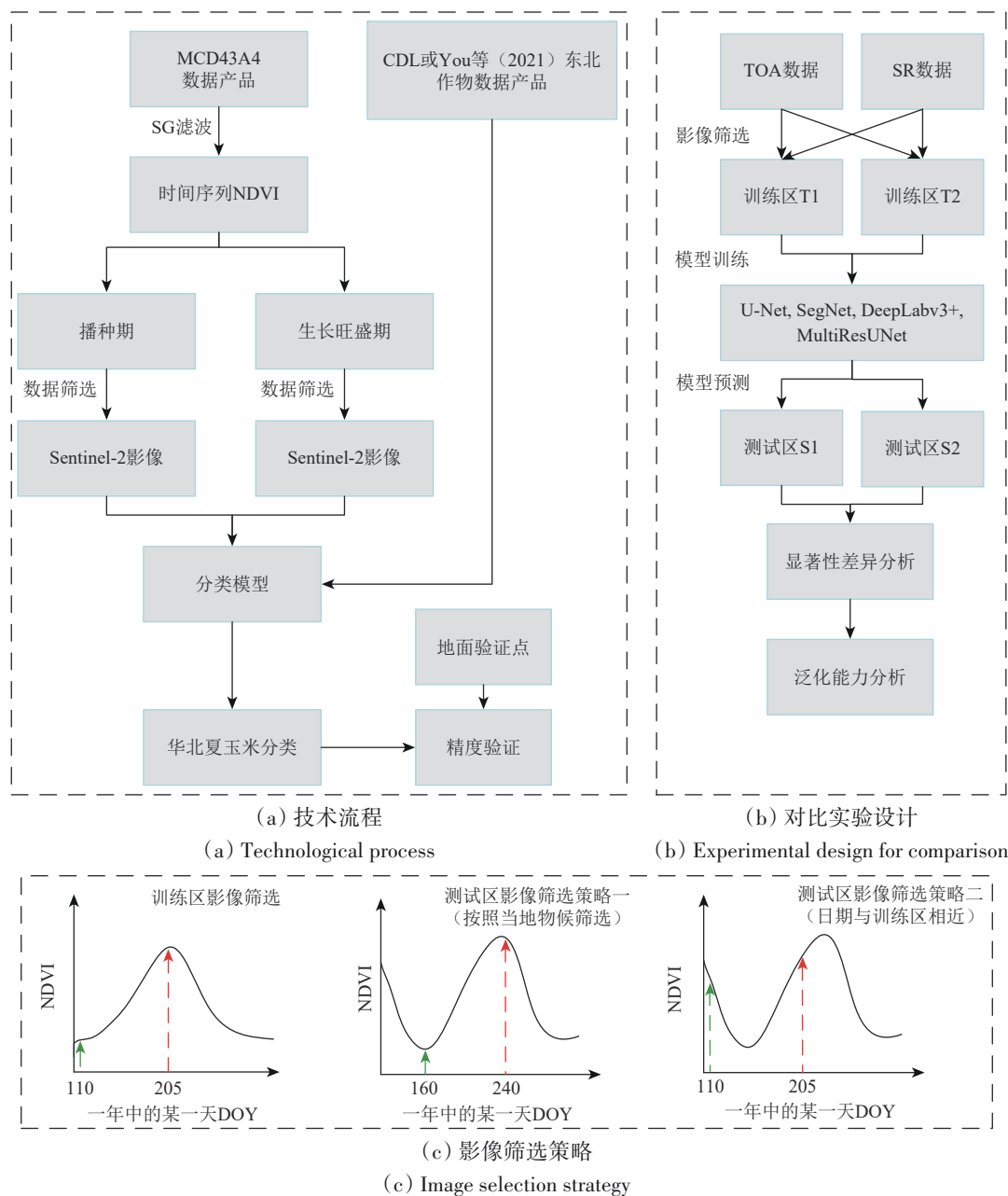


图4 技术路线图

Fig. 4 Flowchart of this research

SegNet结构与U-Net相似(图5),两者区别在于解码器部分,使用相应的编码器特征图中记忆的最大池化索引对其输入特征图进行上采样(Yang等,2020)。因此,SegNet减少了训练过程中参数的数量,在卫星图像的边界划分和目标检测方面具有良好的性能(Panboonyuen等,2017)。

DeepLab V3+(图6)是DeepLab家族中的一种网络。DeepLab扩展了全卷积神经网络(FCN),

并使用空洞卷积扩大了卷积核的范围,将图像上下文纳入更大的感受野,可以明确控制特征响应的分辨率(Chen等,2018a)。空洞卷积的一个优点是能够使用更大的卷积核的同时保持网络参数个数不变。DeepLab通过空洞空间金字塔池化ASPP(Atrous Spatial Pyramid Pooling)结构应用多个空洞卷积,并使用相同的卷积核,但是有不同的采样率。所有卷积的输出都进行相加操作。此

外, 由于使用下采样和最大池化操作使得分割结果容易丢失细节。为了提高分割的空间定位, 特别是边界细节, DeepLab为双线性插值后的网络输出采用了全连接的条件随机场CRF (Conditional Random Field)。CRF设法通过二元势函数寻找像素之间的联系, 并使得相似的像素属于同一类别的概率较大。除DeepLab外, DeepLab V3 (Chen等, 2017) 和DeepLab V3+ (Chen等, 2018b) 模型被开发。与DeepLab不同, DeepLab V3使用不同

采样率的多个连续的空洞卷积, 其通过ASPP对模型的最后一个特征图应用全局平均池化, 用256个卷积核对特征进行 1×1 卷积, 并进行双线性上采样以恢复所需的空间大小。此外, DeepLab V3放弃了CRF, 在计算预测分数之前, 特征图是通过 1×1 卷积连接和相加的。DeepLab V3+在DeepLab V3网络中增加了一个解码器模块, 以优化边界细节。该模型在解码器模块和ASPP中应用了可分离的深度卷积, 对于每个输入通道, 执行一个深度的空间卷积。

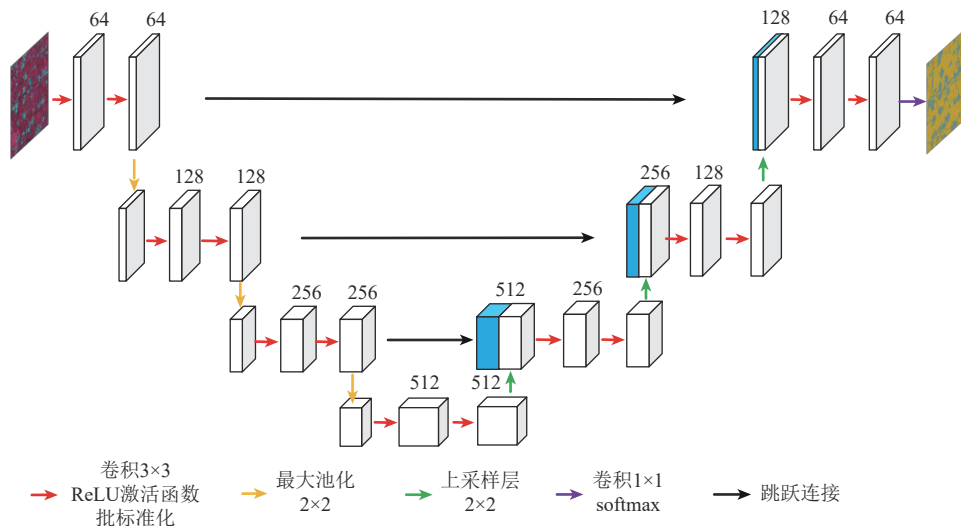


图5 U-Net架构

Fig. 5 Architecture of U-Net

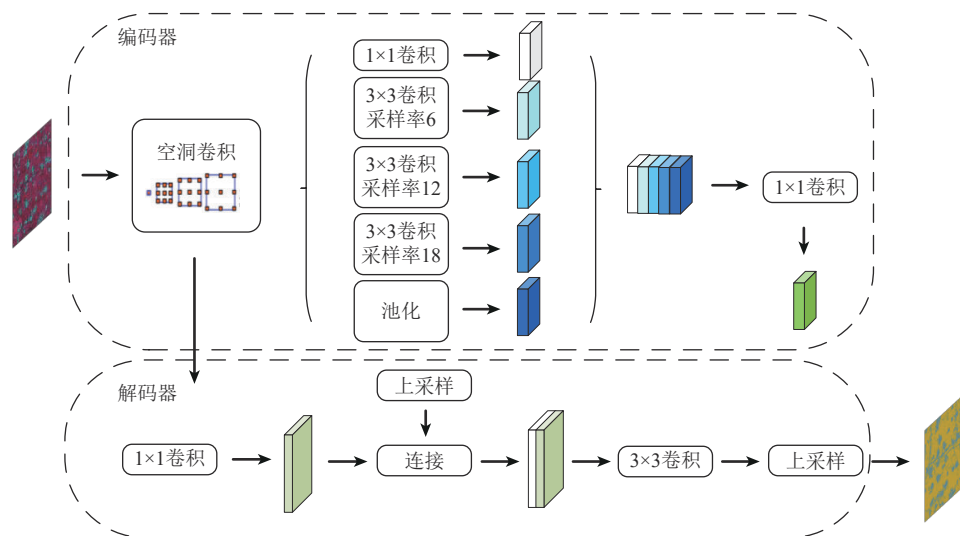


图6 DeepLab V3+架构

Fig. 6 The architecture of DeepLab V3+

MultiResUNet是源于医学影像分割任务, 由于医学影像中细胞和器官具有不同的形状和大小, 该网络旨在提升U-Net在不同尺度上识别目标的能力 (Ibtehaz和Rahman, 2020)。其相对于U-Net的

改进主要在2个方面 (图7): 1. 修改编码器和解码器的子模块为“多分辨率模块” (图7 (b))。该模块包含3个卷积核个数逐渐递增的 3×3 卷积层和一个 1×1 卷积层, 并将卷积层的输出特征进行拼接

和叠加操作。该模块的改进思路源于 GoogleNet 的 Inception 结构 (Szegedy 等, 2015), 使模型从不同尺度上检测目标; 2. 修改“跳跃连接”为“残差路径” (图 7 (c))。U-Net 结构中的“跳跃连接”用于融合编码器所提取的低层特征和解码器所提取的高层特征, 但是低层和高层特征存在语义差异, 直接将这二者融合不利于模型的学习。因此, “残差路径”通过添加一系列卷积层来增加非线性变换, 使得编码器特征和解码器特征之间的语义差异得以减弱。在另一方面, “残差路径”引入“残差连接”以加速模型的学习过程 (He 等, 2016)。在作物制图过程中, 由于不同地区作物地块大小和形状会发生显著变化, 即同样存在尺度问题, 因此针对医学影像多尺度问题而设计且应用效果显著的 MultiResUNet 在作物制图方面的适用性需要深入地分析。

2.3.2 实验场景设计

为分析各因素对模型泛化能力的影响, 本研究设置对比实验流程如图 4 (b) 所示, 具体内容如表 1 所示。由于训练样本量对模型泛化能力的影响不能一概而论, 需要依据具体问题 (目标作物的类型、数量和可分性, 模型架构等) 进行具体分析, 即在本研究场景中所总结得到的样本数量对模型泛化的影响可能并不适用于其他研究场景。因此, 本研究未进行样本数量对模型泛化能力影响的探究。实验所用计算机系统为 Linux Ubuntu 16.04 版本, 内存大小为 64 G, 使用基于 Python 的 Tensorflow 进行模型训练和测试。

2.3.3 精度评价方法

本研究利用在河北和江苏的 2 个测试区 S1 和 S2 采集的地面验证点进行精度评价。为整体分析夏玉米识别的生产者精度 PA (Producer's Accuracy) 和用户精度 UA (User's Accuracy), 采用 F1-score 指标来进行精度评价。由于 Kappa 系数在评价作物制图精度方面存在局限性, 因此, 本研究未采用 Kappa 系数 (Foody, 2020)。测试区验证样本指标 (F1) 计算公式如下:

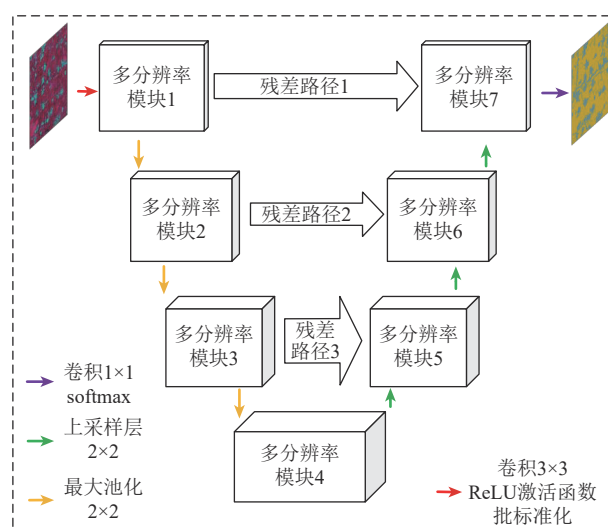
$$F1 = 2 \times \frac{PA \times UA}{PA + UA} \quad (1)$$

本研究采用配对卡方检验 (McNemar's test) (De Wit 和 Clevers, 2004) 来评估 U-Net 和 MultiResUNet 的性能之间是否存在显著性差异。其计算公式

如下:

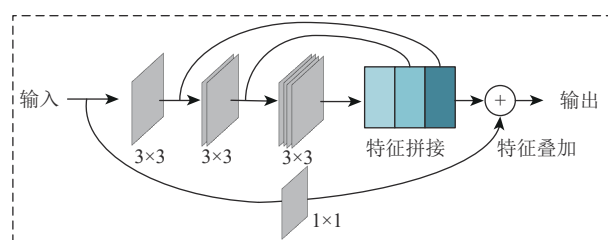
$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (2)$$

式中, f_{12} 表示分类器 1 (U-Net) 正确识别而分类器 2 (MultiResUNet) 未正确识别的样本数, f_{21} 表示分类器 2 正确识别而分类器 1 未正确识别的样本数。若 $-1.96 < Z < 1.96$, 则证明 2 个分类器之间不存在显著性差异。若 $Z > 1.96$, 则证明 2 个分类器之间存在正向的显著性差异, 即分类器 1 比分类器 2 有更好的性能。若 $Z < -1.96$, 则证明 2 个分类器之间存在负向的显著性差异, 即分类器 2 比分类器 1 有更好的性能。



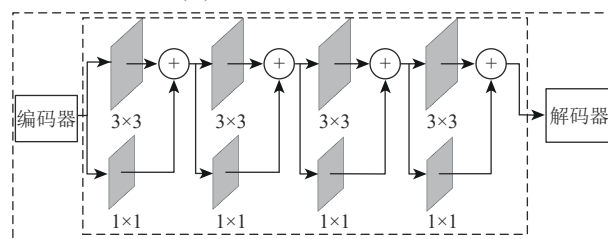
(a) MultiResUNet

(a) MultiResUNet



(b) 多分辨率模块

(b) Multi-resolution module



(c) 残差路径

(c) Residual pathway

图 7 MultiResUNet 架构

Fig. 7 Architecture of MultiResUNet

表1 探究各因素对模型泛化能力影响的实验设计内容

Table 1 The content of experimental design that explores the influence of various factors on the generalization ability of the model

因素	实验设计
数据类型变化对模型泛化能力的影响	分别使用TOA 和SR 数据来训练和测试模型
地块尺度变化对模型泛化能力的影响	分别在位于中国吉林的T1 和美国伊利诺伊州的T2 训练模型,对比模型在测试区的性能
对比不同架构模型的泛化能力	各个训练区同时训练UNet、SegNet、DeepLab V3+和U-Net,对比这四种模型的性能
作物物候变化对模型泛化的影响	在测试区根据不同的影像筛选策略来筛选测试影像(图4(c)),策略1 为按照测试区的物候筛选影像,策略2 为采用与训练影像相近日期的影像(在相同日期,可能无影像覆盖测试区),对比模型在不同测试影像上的性能

本研究通过平均地块面积 S 来反映某一区域内的地块平均大小,以此来分析地块大小与模型泛化性能的关系。其计算公式如下:

$$S = \frac{\sum_{i=1}^n a_i}{n}$$

(3)

式中, n 表示某一区域内的地块个数, a_i 表示某一地块的面积。

3 结果与分析

3.1 模型收敛速度及显著性差异分析

不同模型在不同训练数据下的训练过程的损

失曲线如图8所示。可见:在不同的训练区使用不同类型的训练数据(TOA 或SR), MultiResUNet、SegNet 和 DeepLab V3+的 Loss 下降速度均快于U-Net;在同样训练30个轮数之后,U-Net的 Loss 最大,而其他3种网络的 Loss 相差不大。不同模型的性能显著性差异分析如图9所示,可知:在T1 训练的4种网络性能均无显著性差异,而在T2 训练的MultiResUNet 同其他网络则出现了显著性差异,且性能强于其他网络。该结果表明 MultiResUNet 所进行的结构改动对其性能有一定的提升作用;并且T1 和T2 区域的训练数据存在差异。该差异对模型泛化能力的影响需要进一步分析。

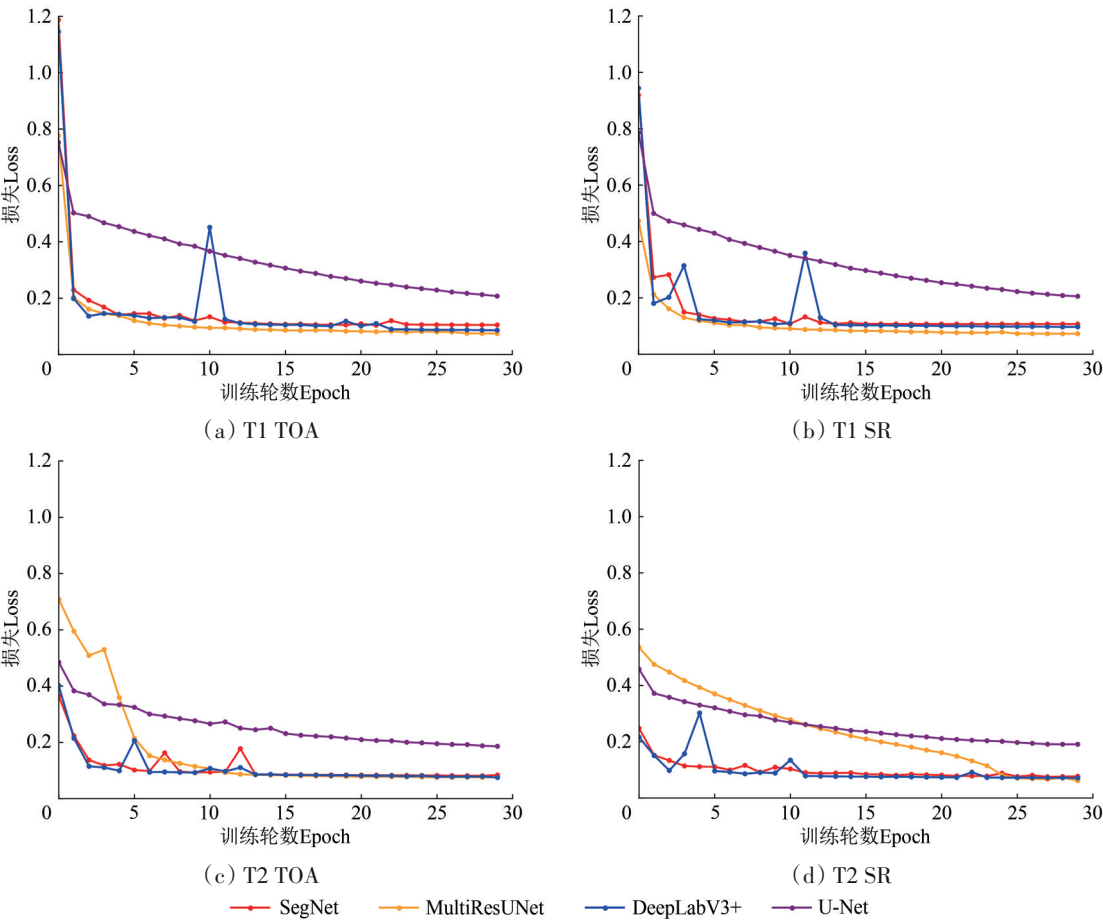


图8 模型在训练过程中的损失值曲线

Fig. 8 Loss curves of the different CNN models during the training process

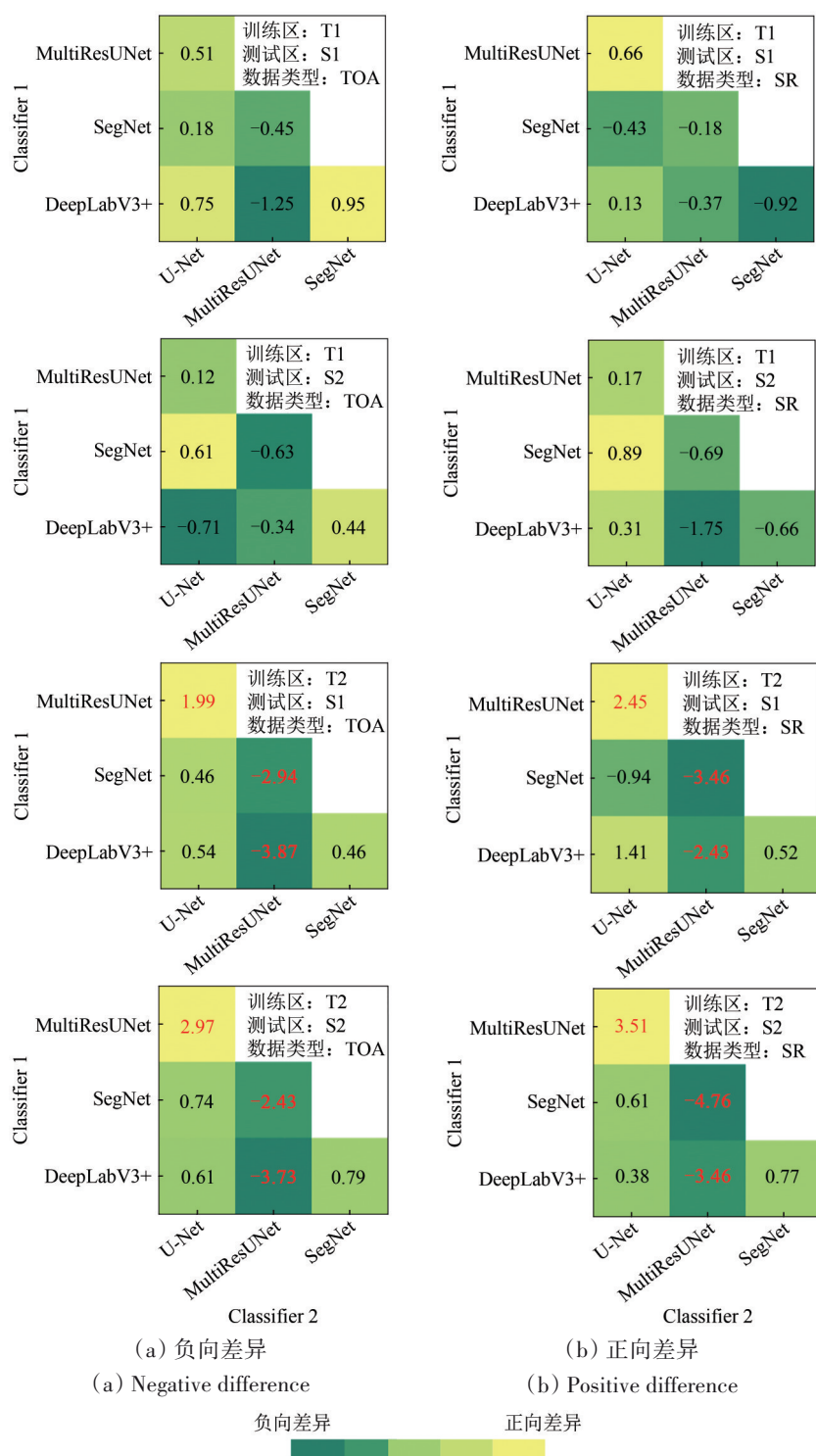


图9 模型性能显著性差异分析结果

Fig. 9 Significant difference analysis results between different models

3.2 数据类型变化对模型泛化能力的影响

数据类型变化对模型泛化能力的影响如表2所示。可知：采用SR数据进行模型训练和测试的效果均优于TOA数据。值得注意的是，当训练区为伊利诺伊州T2，测试区为位于江苏的S2时，采用SR数据对于U-Net和SegNet迁移效果的提升显著，

其F1值比采用TOA数据分别提升了0.39 (0.31—0.70) 和0.23 (0.62—0.85)。当训练区为T2，测试区为S1，SR数据对于U-Net和SegNet的F1值的提升也达到了0.14 (0.76—0.90) 和0.11 (0.77—0.88)。DeepLab V3+结果的F1值随着数据类型的变化在0.06以内，相对于上述两种网络较小。而MultiResUNet

对于数据类型 的变化并不敏感，采用SR数据对于模型迁移性能的提升较小（F1 值增加0.01 左右）。这是由于 MultiResUNet 通过在残差路径添加的非线性操作使得其对数据扰动具有鲁棒性（Qi 等，2020）。

3.3 地块尺度对模型泛化能力的影响

根据表 2 可知，不论采用 TOA 数据还是 SR 数据，在 T1 训练的模型在 S1 和 S2 的迁移效果均优于在 T2 训练的模型。值得注意的是，在测试区 S2，这一差异更为显著（表 2）。例如，当采用 SR 数据时，虽然 T1 和 T2 的训练数据分布较为相似（图 10（a）），但是在 T1 训练的 U-Net 的 F1 值为 0.92，比在 T2 训练的 U-Net 的 F1 值高出 0.22。对于 MultiResUNet，

F1 值随着训练区域的变化浮动相对于 U-Net 更小，因此具有更强的泛化能力。

表 2 不同训练区和测试区组合下的精度评价结果
Table 2 Accuracy evaluation results under different combinations of training areas and test areas

测试区	数据 类型	训练区	F1 值			
			U-Net	MultiResUNet	SegNet	DeepLab V3+
S1	TOA	T1	0.96	0.97	0.93	0.89
S1	TOA	T2	0.76	0.95	0.77	0.80
S1	SR	T1	0.98	0.98	0.95	0.91
S1	SR	T2	0.90	0.96	0.88	0.86
S2	TOA	T1	0.87	0.95	0.91	0.84
S2	TOA	T2	0.31	0.84	0.62	0.75
S2	SR	T1	0.92	0.97	0.93	0.86
S2	SR	T2	0.7	0.85	0.85	0.81

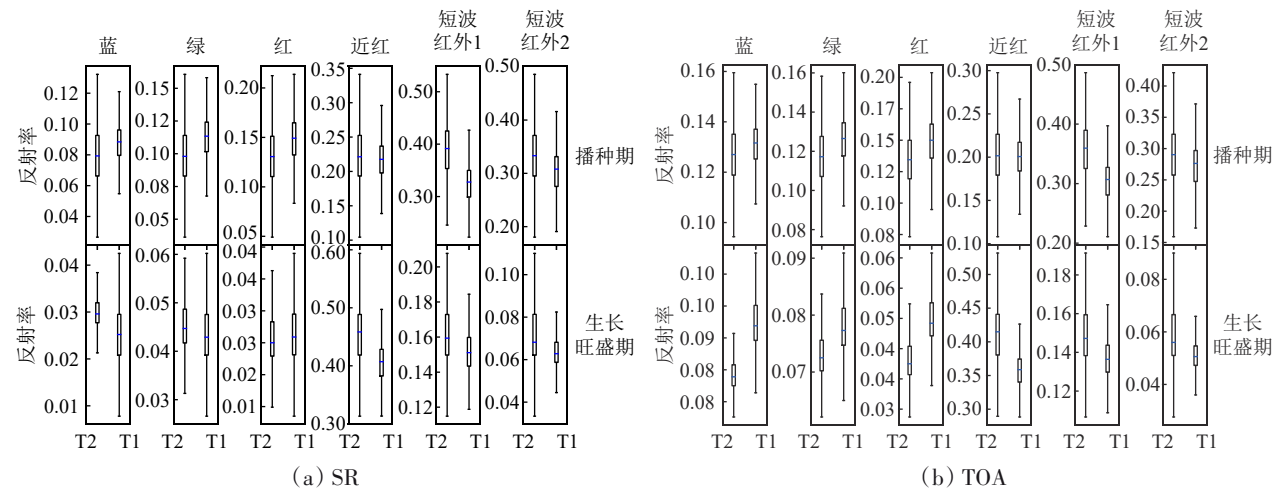


图 10 作物播种期和生长旺盛期训练数据各光谱波段反射率分布
Fig. 10 Reflectance distribution of each spectral band of training data during crop planting and peak growth periods

根据图 11 可知，不同地区的地块尺度大小存在明显差异，位于美国的 T2 的地块更大更规整。而位于中国的 T1、S1 和 S2 的地块更小且形状不规则。其中位于江苏省的 S2 的地块最小。根据图 12 和图 13 可知，在 T1 训练的模型在各个测试区相对

于 T2 有更好的目视效果，田间道路可以更好地得以区分。在地块大小更小的 S2，在 T2 训练的 U-Net 几乎无法准确确定地块边界，目视效果相对于其他 3 种网络较差。





图 11 不同地区地块示意图

Fig. 11 Schematic diagram of plots in different regions

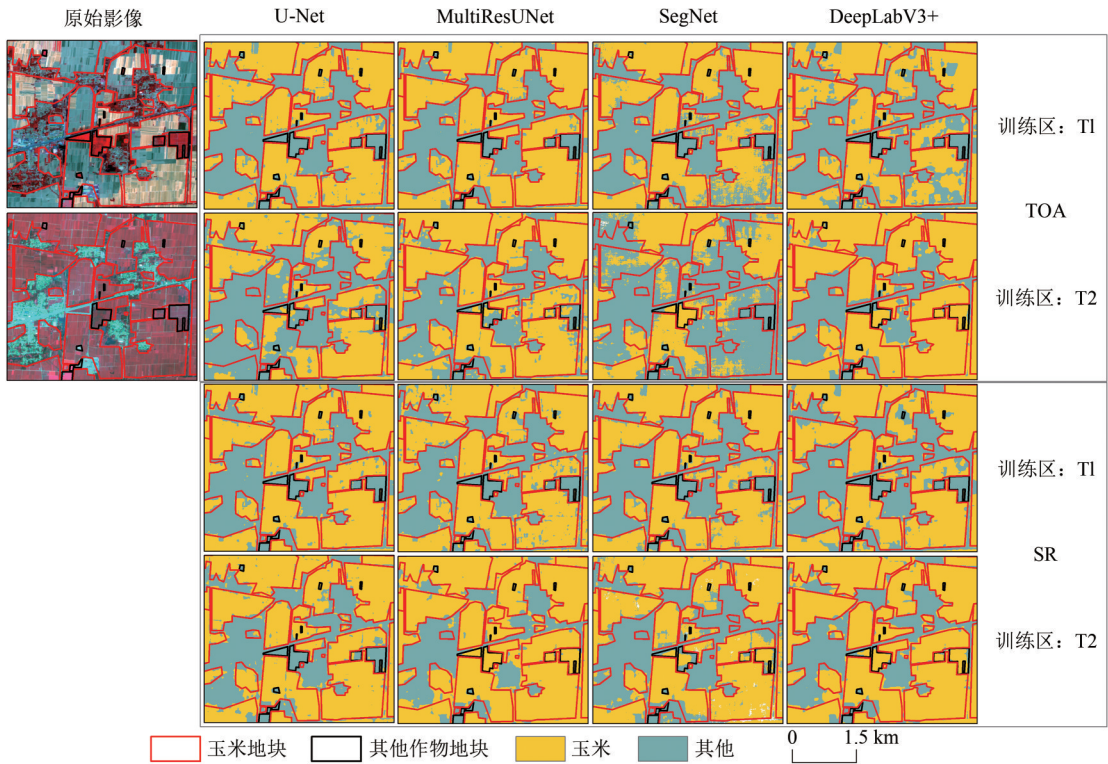


图 12 基于不同的训练区和不同数据类型数据训练的模型在测试区 S1 的分类结果(原始影像从上到下分别为播种期和生长旺盛期影像)

Fig. 12 The classification results for test area S1, using models trained on various training regions and different data types (Original images for the planting and peak growth periods are displayed from top to bottom)

本研究进一步定量分析了地块大小与模型泛化能力的关系 (图 14)。在 S1 和 S2 各筛选 3 个共 6 个大小为 1×1 km 的子区, 并结合野外调查点人工勾绘出子区的地块范围和夏玉米种植区, 分析模型在这些子区的测试精度与地块平均面积的关系。以训练数据为 SR 数据为例, 利用 T1 和 T2 训练的 MultiResUNet、SegNet 和 DeepLab V3+ 和 U-Net

在这 6 个子区进行测试, 测试结果与地块大小的相关性分析如图 14 所示。结果表明随着平均地块面积的增大, 各模型的泛化性能均得到明显提升。在训练区保持一致时, MultiResUNet 在各子区的表现均优于 U-Net, 大部分情况下优于 SegNet 和 DeepLab V3+, 尤其是在地块相对较小时。

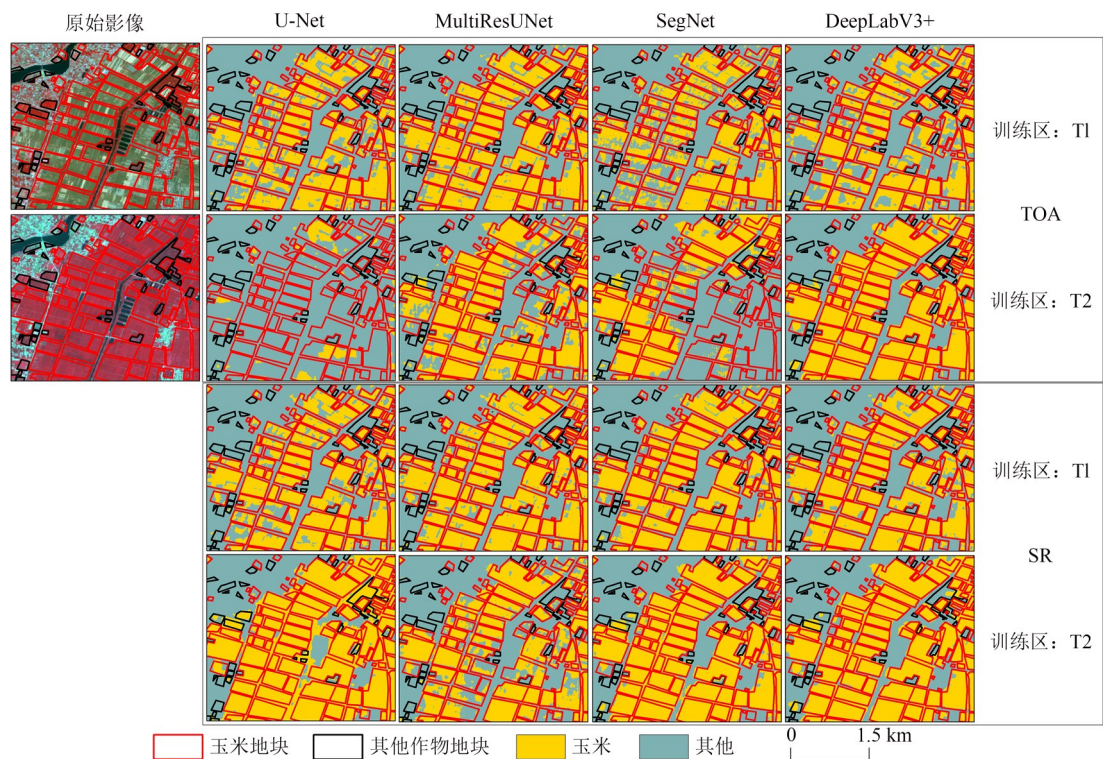


图 13 基于不同的训练区和不同数据类型数据训练的模型在测试区 S2 的分类结果(原始影像从上到下分别为播种期和生长旺盛期影像)

Fig. 13 The classification results for test area S2, using models trained on various training regions and different data types (Original images for the planting and peak growth periods are displayed from top to bottom)

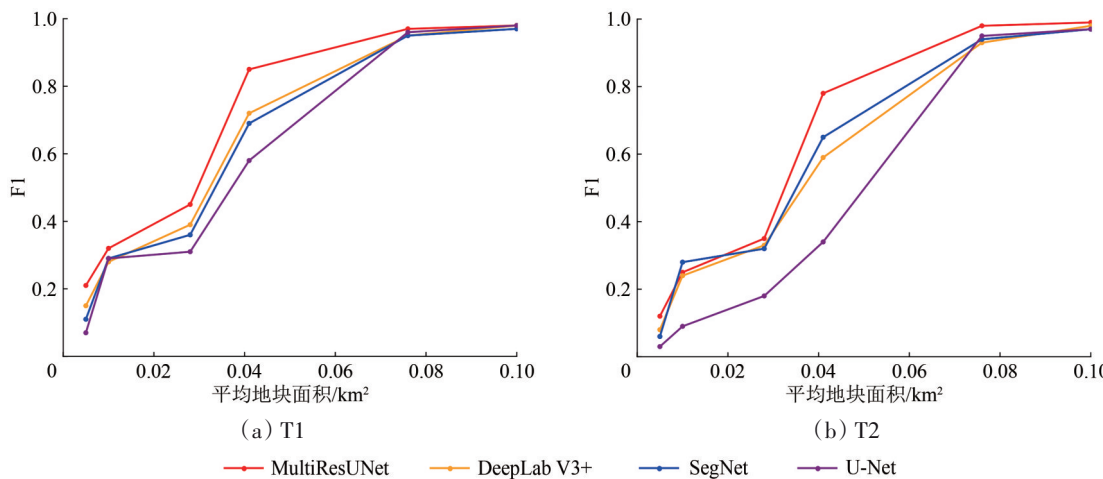


图 14 模型预测精度与平均地块面积的关系

Fig. 14 The relationship between model prediction accuracy and average plot area

3.4 作物物候变化对模型泛化能力的影响

根据表 3 可知当测试区采用与训练区相近日期的影像时（策略二），4 种网络测试结果的 F1 值均小于 0.3，无法准确识别夏玉米。而在测试区通过当地物候来筛选影像时（策略一），模型性能得到较大提升，F1 值均高于 0.86。本研究进一步分析了不同筛选策略所得到的测试影像的玉米类别的

数据分布与训练数据的差异（图 15），可知根据策略二得到的测试数据的分布范围相对于训练数据存在较大差异，该测试数据的分布范围远大于训练数据，因此训练的模型无法准确识别该测试数据中的夏玉米。而当采用策略一时，该差异较小。这表明，在不同区域作物物候变化导致的测试数据与训练数据分布不一致使得模型泛化能力受限，

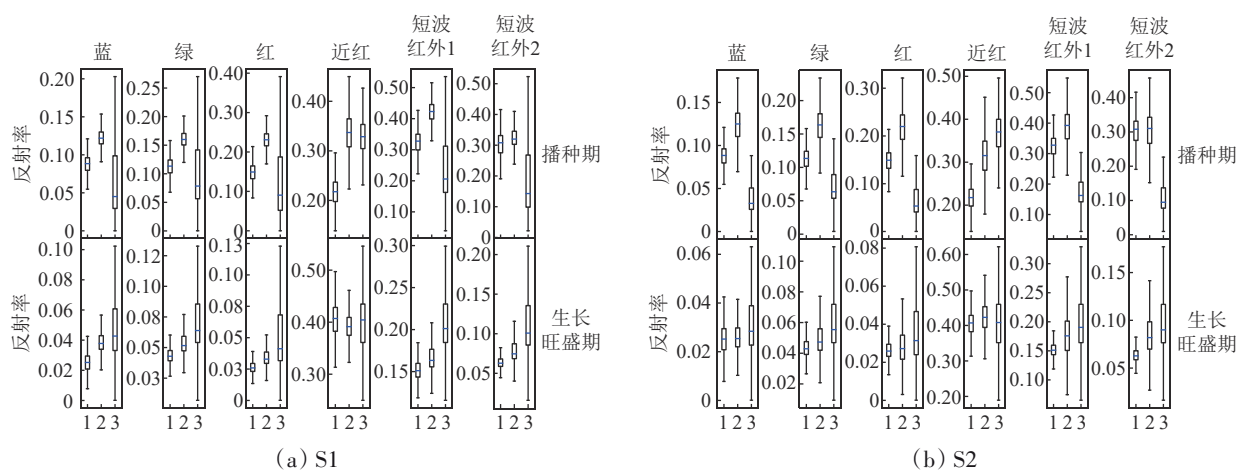
在该情况下模型架构的改进对于精度的提升并不明显。在测试区结合目标作物的物候来筛选关键期内的影像作为测试数据,可以使得测试数据与

相同关键期内的训练数据保持相似,这使得模型性能得到显著提升,从而削弱作物物候变化对模型泛化的不利影响。

表3 不同测试影像筛选策略对模型泛化能力的影响

Table 3 The influence of different test image selection strategies on the model's generalization ability

影像筛选策略	训练区	F1 值			
		U-Net	MultiResUNet	SegNet	DeepLab V3+
测试影像按照当地物候筛选	T1	0.98	0.98	0.95	0.91
	T2	0.92	0.97	0.93	0.86
测试影像日期与训练区相近	T1	0.26	0.27	0.24	0.26
	T2	0.20	0.23	0.22	0.21



1: 训练区T1 SR数据分布 2: 测试区策略一所得数据的分布 3: 测试区策略二所得数据的分布

图15 不同测试影像筛选策略所得测试数据的分布

Fig. 15 Distribution of test data obtained from different test image selection strategies

4 讨论

由于大气校正并非遥感影像分类的先决条件,因此TOA数据经常用于遥感影像分类任务中(Song等,2001)。但是在跨洲际尺度的不同地区,其大气状况可能会存在较大的差异(You和Dong,2020)。在本研究中T1、S1和S2位于中国东部,T2位于美国中部。结果表明当采用在T1训练的模型向S1和S2迁移时,F1值受数据类型的影响较小,变化在0.05以内(表2)。当采用T2训练的模型向S1和S2跨洲际尺度迁移时,采用TOA数据使得U-Net结果的F1值较低(0.76和0.31)。

本研究进一步分析了采用不同的数据类型时,代表玉米类别的训练数据各光谱波段的反射率分布(图10)。可知当采用SR数据时,T1和T2的训练数据分布高度相似。这是由于本研究通过分析

不同地区玉米的NDVI时间序列曲线,确定播种期和生长旺盛期,然后据此在不同地区筛选影像,可以使得不同地区的玉米在这些影像上表现出更相似的遥感特征。然而,当采用TOA数据时,T1和T2的训练数据在生长旺盛期影像的可见光波段(蓝、绿和红)出现了明显差异,这是由于可见光波段波长较短,受大气影响较大(徐春燕和冯学智,2007)。因此在不同地区同种地物的TOA反射率并不一致,这会潜在地影响模型泛化性能。因此,在大尺度作物制图的研究实践中,需要优先使用SR数据,以避免大气状况变化对模型迁移的不利影响。

在地块尺度对不同模型的影响方面,结果表明MultiResUNet相对于其他3种网络可以识别更小的地物,这归因于MultiResUNet针对多尺度问题所设计的多分辨率模块和残差路径(Ibtehaz和

Rahman, 2020)。然而，在地块面积小于0.03 km²时，MultiResUNet的F1值小于0.5（表3），即单纯依靠MultiResUNet的泛化能力依然无法完全克服地块空间尺度的变化对模型迁移的不利影响。值得注意的是，在T2训练的模型的性能依然要劣于T1训练的模型，这表明对于中国华北玉米识别，应优先考虑与之农业景观更相似的You等（2021）提供的东北作物分布数据作为训练数据。

由于作物生长旺盛期通常为夏季，该时期遥感影像通常伴随着频繁且大面积的云覆盖，因此本研究尚未进行整个华北地区的夏玉米识别。在未来的研究需要考虑使用不受云影响的合成孔径雷达SAR（Synthetic Aperture Radar）数据，以进一步扩大测试范围。本研究所筛选的测试区为夏玉米集中种植的地区，暂未考虑与夏玉米物候相似的其他小规模作物对模型泛化能力的影响，在未来的研究中可尝试增加这些作物的训练样本，以获得效果更好的作物分类模型。

5 结 论

本研究通过设置对比实验分析了遥感影像数据类型、地块尺度变化、作物物候变化以及不同模型架构对作物分类模型泛化能力的影响，主要结果如下：（1）当训练区和测试区地块大小发生明显变化时，MultiResUNet相对于其他网络具有更好的泛化性能。然而，单纯依靠MultiResUNet的泛化能力依然无法完全克服地块空间形态的变化对模型迁移的不利影响，为获得更高精度的华北玉米分布信息，需要优先使用与华北地区农业景观更相似的东北作物分布数据产品进行深度学习模型训练；（2）对于TOA数据，采用SR数据更有利于模型在跨洲际尺度进行空间迁移。因此，在大尺度作物制图研究中，应优先考虑使用SR数据。

参考文献(References)

Azzari G, Lobell D B. 2017. Landsat-based classification in the cloud: an opportunity for a paradigm shift in land cover monitoring. *Remote Sensing of Environment*, 202: 64-74 [DOI: 10.1016/j.rse.2017.05.025]

Boryan C, Yang Z W, Mueller R and Craig M. 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, 26(5): 341-358 [DOI: 10.1080/10106049.2011.562309]

Cai Y P, Guan K Y, Peng J, Wang S W, Seifert C, Wardlow B and Li Z. 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, 210: 35-47 [DOI: 10.1016/j.rse.2018.02.045]

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018a. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]

Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* [DOI: 10.48550/arXiv.1706.05587]

Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation//*Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]

De Wit A J W and Clevers J G P W. 2004. Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing*, 25(20): 4091-4112 [DOI: 10.1080/01431160310001619580]

Diao C Y. 2020. Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. *Remote Sensing of Environment*, 248: 111960 [DOI: 10.1016/j.rse.2020.111960]

Footy G M. 2020. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sensing of Environment*, 239: 111630 [DOI: 10.1016/j.rse.2019.111630]

Fu S, Chen S H, Zhang X Q and Xu S Q. 2017. Variations of precipitation and air temperature during 1960-2014 in Jiangsu province. *Environmental Protection and Technology*, 23(2): 26-33 (傅帅, 陈舒慧, 张小泉, 徐士琦. 2017. 江苏省1960-2014年降水及气温变化特征. *环保科技*, 23(2): 26-33) [DOI: 10.3969/j.issn.1674-0254.2017.02.006]

Ge S, Zhang J S, Pan Y Z, Yang Z and Zhu S. 2021. Transferable deep learning model based on the phenological matching principle for mapping crop extent. *International Journal of Applied Earth Observation and Geoinformation*, 102: 102451 [DOI: 10.1016/j.jag.2021.102451]

Hamdi Z M, Brandmeier M and Straub C. 2019. Forest damage assessment using deep learning on high resolution remote sensing data. *Remote Sensing*, 11(17): 1976 [DOI: 10.3390/rs11171976]

Hamrouni Y, Paillassa E, Chéret V, Monteil C and Sheeren D. 2021. From local to global: a transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171: 76-100 [DOI: 10.1016/j.isprsjprs.2020.10.018]

Hao P Y, Di L P, Zhang C and Guo L Y. 2020. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. *Science of the Total Environment*, 733: 138869 [DOI: 10.1016/j.scitotenv.2020.138869]

- 10.1016/j.scitotenv.2020.138869]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Ibtehaz N and Rahman M S. 2020. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121: 74-87 [DOI: 10.1016/j.neunet.2019.08.025]
- Jiang Y L, Chen B W, Huang Y F, Cui J Q and Guo Y L. 2021. Crop planting area extraction based on google earth engine and NDVI time series difference index. *Journal of Geo-Information Science*, 23(5): 938-947 (姜伊兰, 陈保旺, 黄玉芳, 崔佳琪, 郭宇龙. 2021. 基于 Google Earth Engine 和 NDVI 时序差异指数的作物种植区提取. *地球信息科学学报*, 23(5): 938-947) [DOI: 10.12082/dqxxkx.2021.200291]
- Kattenborn T, Eichel J, Wiser S, Burrows L, Fassnacht F E and Schmidtlein S. 2020. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sensing in Ecology and Conservation*, 6(4): 472-486 [DOI: 10.1002/rse2.146]
- Kattenborn T, Leitloff J, Schiefer F and Hinz S. 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173: 24-49 [DOI: 10.1016/j.isprsjprs.2020.12.010]
- Kingma D P and Ba J. 2014. Adam: a method for stochastic optimization//3rd International Conference on Learning Representations. San Diego: ICLR
- Kluger D M, Wang S and Lobell D B. 2021. Two shifts for crop mapping: leveraging aggregate crop statistics to improve satellite-based maps in new regions. *Remote Sensing of Environment*, 262: 112488 [DOI: 10.1016/j.rse.2021.112488]
- Konduri V S, Kumar J, Hargrove W W, Hoffman F M and Ganguly A R. 2020. Mapping crops within the growing season across the United States. *Remote Sensing of Environment*, 251: 112048 [DOI: 10.1016/j.rse.2020.112048]
- LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436-444 [DOI: 10.1038/nature14539]
- Li Q J, Liu J, Mi X F, Yang J and Yu T. 2021. Object-oriented crop classification for GF-6 WFV remote sensing images based on Convolutional Neural Network. *Journal of Remote Sensing*, 25(2): 549-558 (李前景, 刘珺, 米晓飞, 杨健, 余涛. 2021. 面向对象与卷积神经网络模型的 GF-6 WFV 影像作物分类. *遥感学报*, 25(2): 549-558) [DOI: 10.11834/jrs.20219347]
- Li Y H and Che S J. 2005. Impacts of changes of temperature and precipitation on agriculture in Hebei regions. *Chinese Journal of Agrometeorology*, 26(4): 224-228 (李元华, 车少静. 2005. 河北省温度和降水变化对农业的影响. *中国农业气象*, 26(4): 224-228) [DOI: 10.3969/j.issn.1000-6362.2005.04.005]
- Li Z W, Shen H F, Cheng Q, Liu Y H, You S C and He Z Y. 2019. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150: 197-212 [DOI: 10.1016/j.isprsjprs.2019.02.017]
- Liu T, Abd-Elrahman A, Morton J and Wilhelm V L. 2018. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience and Remote Sensing*, 55(2): 243-264 [DOI: 10.1080/15481603.2018.1426091]
- Neupane B, Horanont T and Hung N D. 2019. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PLoS One*, 14(10): e0223906 [DOI: 10.1371/journal.pone.0223906]
- Nezami S, Khoramshahi E, Nevalainen O, Pölonen I and Honkavaara E. 2020. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. *Remote Sensing*, 12(7): 1070 [DOI: 10.3390/rs12071070]
- Panboonyuen T, Jitkajornwanich K, Lawawirojwong S, Srestasathien P and Vateekul P. 2017. Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields. *Remote Sensing*, 9(7): 680 [DOI: 10.3390/rs9070680]
- Qi X M, Zhu P P, Wang Y B, Zhang L Q, Peng J H, Wu M F, Chen J L, Zhao X D, Zang N and Mathiopoulos P T. 2020. MLRSNet: a multi-label high spatial resolution remote sensing dataset for semantic scene understanding. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169: 337-350 [DOI: 10.1016/j.isprsjprs.2020.09.020]
- Qi X N, Wang Y, Wang Q C, Liu Z Y and Bao Q. 2002. The situation and developing prospect of corn zone in Jinlin province. *Scientia Geographica Sinica*, 22(3): 379-384 (齐晓宁, 王洋, 王其存, 刘兆永, 鲍强. 2002. 吉林玉米带的地位与发展前景. *地理科学*, 22(3): 379-384) [DOI: 10.3969/j.issn.1000-0690.2002.03.022]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Rußwurm M and Körner M. 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu: IEEE: 1496-1504 [DOI: 10.1109/CVPRW.2017.193]
- Schiefer F, Kattenborn T, Frick A, Frey J, Schall P, Koch B and Schmidtlein S. 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 170: 205-215 [DOI: 10.1016/j.isprsjprs.2020.10.015]
- Song C H, Woodcock C E, Seto K C, Lenney M P and Macomber S A. 2001. Classification and change detection using landsat TM data:

- when and how to correct atmospheric effects?. *Remote Sensing of Environment*, 75(2): 230-244 [DOI: 10.1016/S0034-4257(00)00169-3]
- Song X P, Potapov P V, Krylov A, King L A, Di Bella C M, Hudson A, Khan A, Adusei B, Stehman S V and Hansen M C. 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sensing of Environment*, 190: 383-395 [DOI: 10.1016/j.rse.2017.01.008]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]
- Wang S, Azzari G and Lobell D B. 2019. Crop type mapping without field-level labels: random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, 222: 303-317 [DOI: 10.1016/j.rse.2018.12.026]
- Wang X, Zhang X C and Su C. 2020. Land use classification of remote sensing images based on multi-scale learning and deep convolution neural. *Journal of Zhejiang University (Science Edition)*, 47(6): 715-723 (王协, 章孝灿, 苏程. 2020. 基于多尺度学习与深度卷积神经网络的遥感图像土地利用分类. *浙江大学学报(理学版)*, 47(6): 715-723) [DOI: 10.3785/j.issn.1008-9497.2020.06.009]
- Weinstein B G, Marconi S, Bohlman S A, Zare A and White E P. 2020. Cross-site learning in deep learning RGB tree crown detection. *Ecological Informatics*, 56: 101061 [DOI: 10.1016/j.ecoinf.2020.101061]
- Wieland M, Li Y and Martinis S. 2019. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sensing of Environment*, 230: 111203 [DOI: 10.1016/j.rse.2019.05.022]
- Xu C Y and Feng X Z. 2007. Atmospheric correction on TM image and its influence analysis on spectral response characteristics. *Journal of Nanjing University (Natural Sciences)*, 43(3): 309-317 (徐春燕, 冯学智. 2007. TM图像大气校正及其对地物光谱响应特征的影响分析. *南京大学学报(自然科学)*, 43(3): 309-317)
- Xu J F, Zhu Y, Zhong R H, Lin Z X, Xu J L, Jiang H, Huang J F, Li H F and Lin T. 2020. DeepCropMapping: a multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sensing of Environment*, 247: 111946 [DOI: 10.1016/j.rse.2020.111946]
- Yang R Y, Ahmed Z U, Schulthess U C, Kamal M and Rai R. 2020. Detecting functional field units from satellite images in smallholder farming systems using a deep learning based computer vision approach: a case study from Bangladesh. *Remote Sensing Applications: Society and Environment*, 20: 100413 [DOI: 10.1016/j.rsase.2020.100413]
- Yang S Q, Song Z S, Yin H P, Zhang Z T and Ning J F. 2021. Crop classification method of UVA multispectral remote sensing based on deep semantic segmentation. *Transactions of the Chinese Society for Agricultural Machinery*, 52(3): 185-192 (杨蜀秦, 宋志双, 尹瀚平, 张智韬, 宁纪锋. 2021. 基于深度语义分割的无人机多光谱遥感作物分类方法. *农业机械学报*, 52(3): 185-192) [DOI: 10.6041/j.issn.1000-1298.2021.03.020]
- You N S and Dong J W. 2020. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161: 109-123 [DOI: 10.1016/j.isprsjprs.2020.01.001]
- You N S, Dong J W, Huang J X, Du G M, Zhang G L, He Y L, Yang T, Di Y Y and Xiao X M. 2021. The 10-m crop type maps in Northeast China during 2017—2019. *Scientific Data*, 8(1): 41 [DOI: 10.1038/s41597-021-00827-9]
- Zhang D J, Pan Y Z, Zhang J S, Hu T G, Zhao J H, Li N and Chen Q. 2020. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sensing of Environment*, 247: 111912 [DOI: 10.1016/j.rse.2020.111912]
- Zhang L, Liu Z, Liu D Y, Xiong Q, Yang N, Ren T W, Zhang C, Zhang X D and Li S M. 2019. Crop mapping based on historical samples and new training samples generation in Heilongjiang province, China. *Sustainability*, 11(18): 5052 [DOI: 10.3390/su11185052]
- Zhang M, Lin H, Wang G X, Sun H and Fu J. 2018. Mapping paddy rice using a convolutional neural network (CNN) with landsat 8 datasets in the Dongting lake area, China. *Remote Sensing*, 10(11): 1840 [DOI: 10.3390/rs10111840]
- Zhao H W, Chen Z X, Jiang H and Liu J. 2020. Early growing stage crop species identification in southern China based on sentinel-1A time series imagery and one-dimensional CNN. *Transactions of the Chinese Society of Agricultural Engineering*, 36(3): 169-177 (赵红伟, 陈仲新, 姜浩, 刘佳. 2020. 基于Sentinel-1A影像和一维CNN的中国南方生长季早期作物种类识别. *农业工程学报*, 36(3): 169-177) [DOI: 10.11975/j.issn.1002-6819.2020.03.021]
- Zhong L H, Gong P and Biging G S. 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, 140: 1-13 [DOI: 10.1016/j.rse.2013.08.023]
- Zhong L H, Hu L N and Zhou H. 2019. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221: 430-443 [DOI: 10.1016/j.rse.2018.11.032]
- Zhong Y F, Hu X, Luo C, Wang X Y, Zhao J and Zhang L P. 2020. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H^2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sensing of Environment*, 250: 112012 [DOI: 10.1016/j.rse.2020.112012]
- Zhu X X, Tuia D, Mou L C, Xia G S, Zhang L P, Xu F and Fraundorfer F. 2017. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4): 8-36 [DOI: 10.1109/MGRS.2017.2762307]

Spatial generalization ability analysis of deep learning crop classification models

GE Shuang^{1,2,3,4}, ZHANG JinShui^{1,3,4}, ZHU Shuang⁵

1. State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China;
2. Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, International Institute for Earth System Science, Nanjing University, Nanjing 210023, China;
3. Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China;
4. Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China;
5. Beijing Polytechnic College, Beijing 100042, China

Abstract: Timely and accurate global crop mapping is important for global food security assessment. However, existing crop classification models are often targeted at specific regions, and their performance in other regions has not been fully evaluated. This study determined the critical period of crop growth in different regions to realize the effective transfer of the model in large-scale regions, and the remote sensing data during these critical periods were filtered such that the same crop in different regions showed similar characteristics on these remote sensing images. This way helps the model achieve a better transfer effect. In this study, the MultiResUNet, SegNet, DeepLab V3+, and U-Net models were trained using data from Northeast China, and the optimal F1 value for summer corn recognition in the study areas in North China can reach more than 0.97. This research also analyzed the factors that affected the generalization ability of the model. The issues addressed in this article include (1) using existing crop distribution data products as the ground truth samples for model training to solve the problem of lack of training samples for the deep learning model. We compared and analyzed the applicability of the models trained using the US Cropland Data Layer and Northeast crop distribution data products in North China. (2) We compared the generalization performance of depth models with different architectures. (3) We compared and analyzed the influence of different data types on the generalization ability of the model. (4) We comparatively analyzed the impact of crop phenology changes on the generalization ability of the model. Results show that MultiResUNet has better generalization performance than other networks when the plot size in the training and test areas varies significantly. However, the generalization ability of MultiResUNet alone still cannot completely overcome the adverse effect of the change in plot spatial morphology on model migration. The crop distribution data products in Northeast China, which are more similar to the agricultural landscape in North China, need to be used for deep learning model training to obtain more accurate information of maize distribution in North China. Compared with TOA data, we found that SR data are more conducive to the spatial migration of the model at the transcontinental scale. Therefore, SR data should be given priority in large-scale crop mapping. This research provides a useful discussion for large-scale crop mapping research using only local samples.

Key words: model generalization, deep learning, SegNet, DeepLab V3+, U-Net, MultiResUNet, crop mapping

Supported by Civil Part of the High-Resolution Earth Observation System Major Special Project (No. 20-Y30F10-9001-20/22)