

# 融合高分辨率网络与双分支结构的耕地范围建筑物遥感监测

刘圳<sup>1,2</sup>, 刘德儿<sup>1,2</sup>, 赵尘<sup>3</sup>

1. 江西理工大学 土木与测绘工程学院, 赣州 341000;
2. 河流源头水生态保护江西省重点实验室, 赣州 341000;
3. 福建省经纬数字科技有限公司, 福州 350001

**摘要:** 耕地是保证农业可持续发展的基础, 快速、精准监测耕地非农化行为对粮食生产和安全具有重要意义。为精准监测乱占耕地进行非农建设的行为, 本文提出一种准确的高分辨率遥感图像耕地非农化行为分割的新方法, 即耕地非农化行为监测深度学习模型 DHRformer (Dilation-enhanced High-Resolution former)。该模型由高分辨率网络和双分支解码结构组成, 通过多尺度融合和扩张策略增强非农建筑的特征信息, 从而获得更丰富的非农建筑细节信息。首先根据多时相遥感数据构建亚米级的潜在非农化区域的建筑物样本数据集; 然后, 使用本文设计的耕地非农化行为监测深度学习模型 (DHRformer) 完成对潜在非农化区域内建筑物的提取。最后, 以怀化市鹤城区的山区耕地为研究区域, 基于高分辨率遥感影像进行本文模型验证实验。结果表明: 本文改进算法的 MIoU、mAcc 及 F1 分数均有改善, 相比于 Segformer 网络, MIoU、mAcc 及 F1 分数分别提升了 2.53%、2.68%、3.05%。因此, 本文提出的方法可为耕地非农化行为监测提供技术参考, 为复杂的非农化行为提供理论支撑。

**关键词:** 耕地非农化, 建筑物遥感监测, 高分辨率网络, 双分支结构, 编码解码, 高分影像

**中图分类号:** TP751.1/P2

**引用格式:** 刘圳, 刘德儿, 赵尘. 2025. 融合高分辨率网络与双分支结构的耕地范围建筑物遥感监测. 遥感学报, 29(9): 2776–2787  
Liu Z, Liu D E and Zhao C. 2025. Remote sensing monitoring of cultivated land via integration of a high-resolution network and a dual-branch structure. National Remote Sensing Bulletin, 29(9): 2776–2787 [DOI: 10.11834/jrs.20244068]

## 1 引言

耕地是粮食生产的重要基础, 也是重要的生产资料。耕地“非农化”指的是耕地违规用于农业生产以外的生产经营活动, 其实质是耕地利用性质的转变, 耕地资源转变为工业、住宅或商业用地等用途的过程。耕地非农化是社会经济发展和快速城镇化背景下出现的现象, 体现出城镇化扩张与土地资源保护之间的矛盾。快速城镇化对耕地的影响主要体现在占用耕地从事非农建设 (Belgiu 和 Csillik, 2018)。根据《第三次全国国土调查主要数据公报》公布的数据显示, 截止2019年末, 全国耕地总面积为 1278666.67 km<sup>2</sup>, 较二调耕

地数据减少 79333.33 km<sup>2</sup>, 为确保农业生产的需要和国家粮食安全, 政府出台了一系列相应的耕地保护政策和措施 (Wu 等, 2017; Chen 等, 2021; 国务院办公厅, 2020), 2024 年中央一号文件明确指出要严格落实耕地保护制度, 举有力措施打击各种“耕地非农化行为”。然而中国地形多为山区, 人均耕地资源匮乏, 部分耕地资源匮乏的地区依然出现耕地非农化现象。因此, 遏制耕地非农化亟需高效的监测方法, 通过高效的监测方法能够提炼出管理耕地非农化的经验, 在耕地资源条件匮乏的山区, 有着重要的参考价值。这对于保障中国粮食安全和维持农业可持续性也具有重要意义 (Ji 等, 2019)。

收稿日期: 2024-03-13; 预印本: 2024-05-30

基金项目: 国家自然科学基金 (编号: 42271434)

第一作者简介: 刘圳, 研究方向为遥感图像目标检测、分割理论和应用。E-mail: 1041910447@qq.com

通信作者简介: 刘德儿, 研究方向为空间数据组织与管理、计算机视觉、深度学习和三维激光点云智能处理。E-mail: landserver@163.com

高分辨率遥感影像能够提供大范围的时间序列数据, 具有周期性短, 时效性强及高分辨率的特点, 已成为监测耕地非农化行为的一种有效手段。在早期研究中, 耕地非农化行为的监测主要依赖于传统方法, 如支持向量机 SVM (Support Vector Machine)、阈值分割、区域分割等方法 (Wu, 2012; Liasis 和 Stavrou, 2016)。O'Connell 等 (2015) 提出了一种与随机森林 RF (Random Forest) 相结合的创新的自适应图像分割技术, 在复杂的农业环境中能够准确提取耕地地块, 但是分割参数的选择容易影响结果。范胜龙等 (2018) 提出了一种结合 BP 神经网络和 SVM 的耕地分等方法, 但是耕地提取精度易受到先验知识和耕地特征选取的干扰。陈丹玲等 (2019) 针对大范围耕地利用效率的复杂性提出一种结合 RF 思想和随机抽样方法的耕地利用效率测度分类树模型, 但是该方法在构建大范围耕地分类树决策过程中会变得复杂。Useya 等 (2019) 为了确定耕地时空变化, 提出一种多分类系统和多指数阈值化的耕地检测方法以实现耕地自动分类, 然而, 该方法对精确校准数据和阈值选取有较高的依存性。这些方法侧重于监测目标的颜色、形状及纹理等低级语义特征的提取。虽然能够利用高分辨率遥感影像提供丰富的细节信息, 但容易受到天气、光照、传感器构造及建筑物结构等外部因素影响, 导致语义信息提取并不清晰。因此这些方法需要在特定数据下进行耕地非农化行为监测 (El-Naggar, 2018), 对于大范围的耕地非农化行为监测通用性不强。

随着深度学习技术不断发展, 语义分割作为计算机视觉领域一个经典任务, 它是对图像内目标逐像素分类的方法。通过堆叠多层卷积层下采样获得目标的高级语义信息, 从而取得令人满意的分割结果。基于深度学习的语义分割方法早已引入到遥感影像监测和分类领域中, 并取得了系列成果 (Liao 等, 2021; Peng 等, 2021; Hua 等, 2022)。Li 等 (2021) 提出一种注意力增强的 U-Net 模型, 基于 Google 和 WorldView2 高分辨遥感影像提取农田内的建筑物, 但在密集建筑区易出现粘连的现象。朱昱等 (2022) 结合深度卷积神经网络和分水岭方法, 弥补深度学习方法在提取耕地地块时丢失的位置信息, 实现对耕地边界更准

确的识别。针对耕地非农化类型的复杂性, 厉芳婷等 (2022) 基于地理国情监测成果构建出一种耕地非农化遥感解译体系, 并利用 EfficientNet 深度学习模型验证其分类效果, 然而构建该解译体系需要人工目视识别才能准确选择样本。谭敏等 (2023) 基于 Segformer 和注意力机制提出一种融入先验知识的潜在非农化图斑提取网络模型, 以耕地范围和类别属性作为深度学习模型监测耕地非农化行为的先验知识, 在大范围的耕地非农化行为监测中取得良好的效果。Wang 等 (2023) 提出一种自适应特征的 UNet 结构分割建筑物, 使用密集连接结构和空间注意力机制的方式连接编码—解码结构, 其模型参数量显著提升且大尺度建筑物分割结果依然存在空洞的情况。

综上, 虽然基于深度学习技术的卷积神经网络已大量运用于遥感图像的目标提取任务中, 但基于高分辨率网络和多尺度特征融合的分割模型去挖掘高分遥感影像在耕地非农化行为监测的研究还很少, 特别是针对耕地资源匮乏的地区, 耕地非农化行为的影响更加强烈。国务院办公厅在《关于坚决制止耕地“耕地非农化”行为的通知》中明确指出 6 种违法占用耕地行为, 包括严禁违规占用耕地绿化造林、违规占用耕地挖湖造景、超标准建设绿色通道、占用永久基本农田扩大自然保护地、违规占用耕地从事非农建设以及违法违规批地用地。而占用耕地从事非农建设的现象在上述行为中尤为常见, 特别是占用耕地新建、扩建非法建筑 (Shao 等, 2020) 和搭建棚房 (农发, 2018)。因此, 本文为进一步实现在有限的耕地资源下平衡农业生产和非农化需求, 即乱占耕地进行非农建设的准确提取, 尤其针对山区地形条件下非农化行为的判别难题。本文提出了一种面向遥感非农建筑物分割的 DHRFormer 深度网络模型, 该方法以 Segformer 为基础, 通过高分辨率网络重构编码器结构, 并在解码器中引入具有不同扩张率的空洞卷积, 构建双分支解码器结构, 通过多尺度融合和扩张的策略, 以达到提升非农建筑识别精度的目的, 并以湖南省怀化市鹤城区的耕地区域为例, 与 Segformer 等模型进行对比, 对本文提出的 DHRFormer 模型进行验证, 以期对现有的非农化行为提取方法中存在目标提取不完整问题提供理论支撑。

## 2 研究方法

本文提出DHRformer提取非农化区域内建筑物网络,如图1所示。在Segformer算法(Qiu C等, 2022)的基础上,选择HRNetV2(Sun等, 2019)作为特征映射的编码器,其目的是保持网络在整个处理过程中的高分辨表示,为生成精确的建筑

物边界提供支持,而“HR”则代表高分辨率网络(High-Resolution Network)。其次“D”指的是在解码器部分引入的扩张率(Dilation)技术,用于构建具有扩展感受野的双分支解码结构,利用该策略将主干映射网络提取到强语义特征进一步提升弱语义特征的输出,目的在于提高模型非农化区域内建筑物细粒度特征的表达能力。

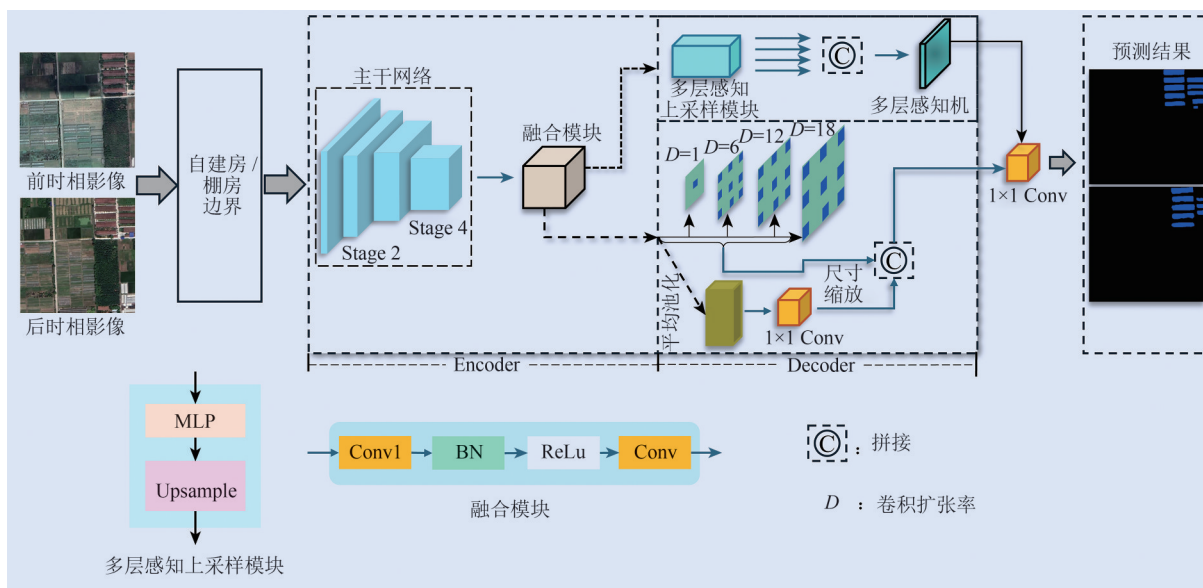


图1 DHRformer网络结构图

Fig. 1 DHRformer network structure diagram

### 2.1 Segformer算法

Segformer的整体结构为编码器—解码器结构,其编码器借助多个Transformer模块设计的分级结构,能有效提取不同层级的语义特征。与传统的Transformer模块不同,Segformer不需要位置编码,因此避免了训练与测试分辨率不同时,导致网络性能下降的问题。Segformer的解码器由多个多层感知机(MLP)构成,由主干输出特征图都通过一个MLP进行预处理,经过MLP的特征图上采样至相同尺寸大小后进行串联,最终通过一个卷积块融合,得到分割掩膜结果。但基于Transformer的编码器的计算复杂度远高于基

于CNN的编码器,不太适用于计算资源受限的环境。以Vision Transformer(ViT)作为编码器的主要结构,由于ViT是柱状结构,只能输出固定分辨率的特征图,如1/16,如此低的分辨率对提取非农化区域内建筑物边界不太友好。

### 2.2 高分辨率网络编码器

非农化区域内建筑物的边界提取精细程度取决于网络经过多次上下采样后,仍能高分辨率保持目标的强语义特征。但大多数的分割模型下采样聚合更广泛的信息得到强语义特征,然后通过上采样恢复原始图像相同的高分辨率,有助于生成更精细的边界分割结果。使用HRNet作为特征映射的主干提取网络的优势是能够在网络中从头到尾保持高分辨率的表示,其结构如图2所示。该网络由4个子网络(stage)组成,每个子网络负责提取不同层次的特征,多分辨率并行流的架构实现了高分辨率的保持,确保了空间位置信息的精确和完整,这样的结构能很好的满足遥感对象的分割。该结构通过将高分辨率和低分辨率并行连接,以实现同步推进,旨在有效整合不同分辨率的特征,从而提高网络对于空间位置信息的准确性和完整性的感知能力。

HRNet作为主干提取网络的第二个优势在于



多次重复利用多尺度融合策略, 每个子网络引入交换单元后重复接收其他子网络的语义信息。这种跨分辨率信息交互特征融合, 高分辨率和低分辨率之间不断交换信息, 实现语义融合, 在原空间信息基础上, 增加了丰富的语义信息。因此, 本文使用 $3\times 3$ 卷积进行下采样及 $1\times 1$ 卷积上采样实

现通道数的对齐。对于并行的4个子网络输出不同分辨率的特征图, 使用多尺度特征融合将3个低分辨率特征图进行上采样与高分辨特征进行尺度对齐, 使网络在提取建筑物边界过程中具有更强的感知能力。

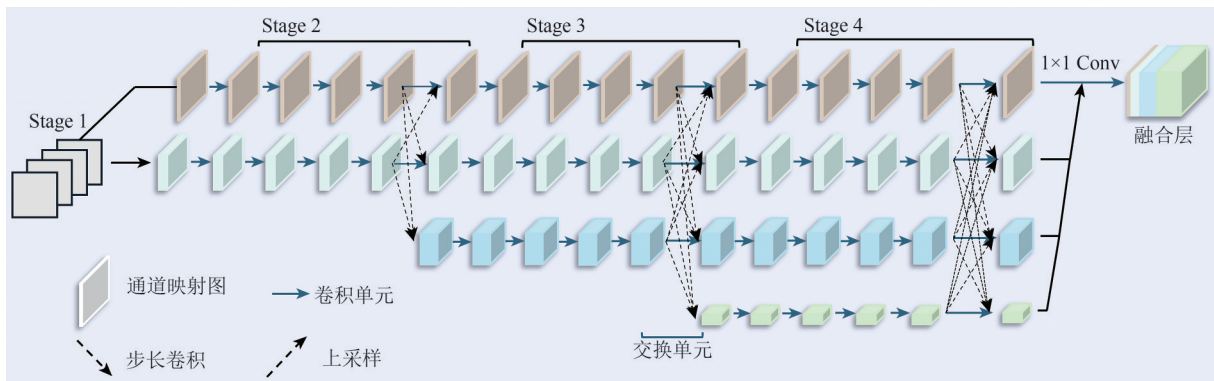


图2 HRNet 高分辨率网络结构

Fig. 2 HRNet high-resolution network architecture

### 2.3 双分支解码器

特征解码模块的功能主要是还原非农化区域内建筑物的细节语义特征, 生成与多时相遥感图像大小相同的建筑物分割掩膜图。在解码器中, 本文构建了一个双分支解码器结构, 主要包括MLP分支和带有不同扩张率的空洞卷积分支。MLP分支由多个MLP层组成, 以较少的参数量和达到较好的分割效果。带有不同扩张率的空洞卷积分支通过设置4种不同的扩张率, 以捕获非农化区域内多个尺度的建筑物, 获得更密集的特征响应后更好地还原建筑物分割细节特征。

本文设计的空洞卷积分支如图1 DHRformer网络结构图的Decoder所示, 空洞卷积的扩张率分别为1、6、12和18, 除了扩张率为1的空洞卷积的卷积核大小为 $1\times 1$ , 其余的空洞卷积的卷积核大小均为 $3\times 3$ , 因此, 将主干提取的信息输入到这4个不同扩张率的通道中, 生成不同尺度的信息流, 最后将它们连接, 得到多尺度语义特征信息。这样设置扩张率的目的在于通过不同的扩张率获得不同的感受野, 有利于不同通道适应于不同尺度的非农化区域内建筑物的分辨率, 从而在接收高分辨率编码器所输出的信息流中融合不同尺度的建筑物信息并提炼潜在的语义信息, 提升建筑物边界分割效果及减少空洞现象。

### 2.4 联合协作损失函数

为了更好地组合双分支解码器的输出, 本文将交叉熵损失函数 (CE Loss) 和 Soft Dice Coefficient 损失函数设计联合协作损失函数, 其联合协作损失函数 $L$ 的表达式为

$$L = \lambda_1 L_{CE} + \lambda_2 L_{Dice} \quad (1)$$

式中,  $L_{CE}$  为交叉熵损失函数,  $L_{Dice}$  为 Soft Dice Coefficient 损失函数,  $L$  为联合协作损失函数,  $\lambda$  为不同损失的权重因子。本文经过实验进行验证, 分别设置为3和1时效果最好。

在潜在耕地非农化区域内建筑物分割中, 建筑物与非建筑物的比例可能不平衡, CE Loss 对多类别进行分割时可能受到类别不平衡的影响。Dice Loss 在计算损失时对正类别 (建筑物) 的像素更为敏感, 通过联合使用有助于缓解类别不平衡问题。其次, Dice Loss 在计算损失时注重预测和真实标签的相似性, 对于目标边缘的匹配有一定的优势。在建筑物分割中, 边缘的准确性对于模型性能至关重要。通过联合使用 CE Loss 和 Dice Loss, 可以综合考虑类别内部的像素匹配和边缘准确性, 提高分割模型在边缘细节上的表现。 $L_{CE}$  和  $L_{Dice}$  计算公式如下:

$$L_{CE(p_i)} = - \sum_{i=0}^{N-1} p_i \log(\hat{p}_i) \quad (2)$$



$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N p_i \hat{p}_i + \varepsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N \hat{p}_i + \varepsilon} \quad (3)$$

式中,  $N$ 为类别数;  $p_i$ 和 $\hat{p}_i$ 分别为类别 $i$ 的真实值和预测值;  $\varepsilon$ 是一个很小的常数, 防止分母为零。

## 2.5 精度评价指标

为了定量评价潜在非农化区域内建筑物分割模型的效果, 本文采用F1-score、平均像素精度(mAcc)和平均交并比(MIoU)的3个度量指标。F1-score定义为精确率和召回率的调和平均值, 是衡量潜在非农化区域内建筑物分割模型的综合指标; mAcc是对于每个类别的正确分类像素占该类别的总像素比例, 然后将所有类别像素值准确度取平均; MIoU是计算每个类别真实值和预测值2个集合的交集和并集之比。

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (4)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (5)$$

$$\text{F1-score} = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \quad (6)$$

$$\text{mAcc} = \frac{1}{n} \frac{\sum_{i=0}^n q_{ii}}{\sum_{i=0}^n \sum_{j=0}^n q_{ij}} \quad (7)$$

$$\text{MIoU} = \frac{1}{n+1} \frac{\sum_{i=0}^n q_{ii}}{\sum_{i=0}^n \sum_{j=0}^n q_{ij} + \sum_{j=0}^n \sum_{i=0}^n q_{ji} - \sum_{i=0}^n q_{ii}} \quad (8)$$

式中,  $n$ 为类别数;  $i$ 和 $j$ 表示类别;  $q_{ii}$ 表示正确预测的像素数量,  $q_{ij}$ 表示原本属于 $i$ 类的像素预测为 $j$ 类, 属于错误预测的像素数量; TP为真阳性, 被预测为正样本, 实际为正样本, FP为假阳性, 被预测为负样本, 实际为正样本, FN为假阴性, 被预测为负样本, 实际为负样本。

## 2.5 研究区概况及数据源

以湖南省怀化市鹤城区为研究区, 研究区概况如图3所示。研究区境内地貌类型多样, 以山地为主, 约占全区总面积74%, 耕地资源却并不丰富。随着生产和建设活动不断增加, 耕地非农化现象逐渐凸显, 特别是乱占耕地建房及棚房搭建行为。

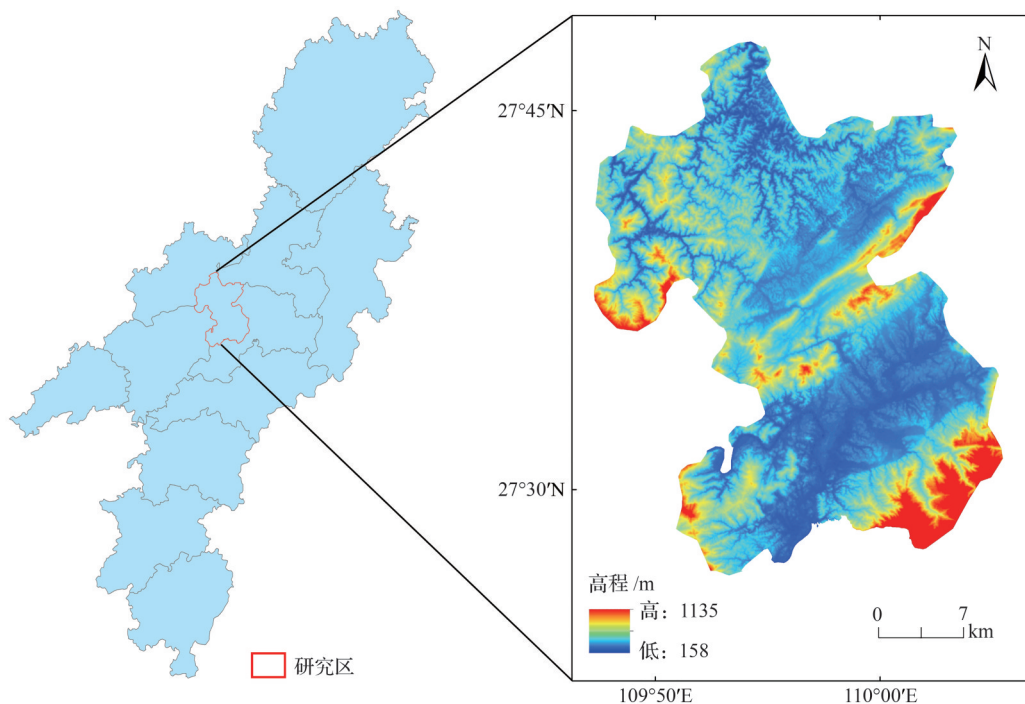


图3 研究区位置和高程示意图

Fig. 3 The location of study area and digital elevation mode

本文使用的耕地数据来自鹤城区“第三次全国国土调查”成果数据。该数据包含了耕地图斑矢量范围, 图4显示了研究区耕地与非耕地的空间分布情况。可见耕地资源分布存在着空间差异,

主要集中在中部和南部, 形成聚集带, 这些耕地区域分布与研究区地形平缓和水资源丰富有关, 适宜农业活动。在整体上, 研究区内的耕地分布呈现出较大的片段化, 片段化的耕地分布体现了

人类活动对耕地格局的显著作用。前时相和后时相数据分别采用2019和2022年Google Earth的遥感影像数据。经过裁剪等预处理方式，选取位于农村地区及含有耕地的图像作为基准数据，最后得到3000组512×512大小的影像数据。数据集是按照8：1：1的比例划分训练集、验证集和测试集。



图4 研究区耕地与非耕地的空间分布  
Fig. 4 Distribution of cultivated and non-cultivated land of the study are

3 结果与分析

3.1 实验参数设置

本文所有训练与测试实验均在Windows10系统上进行，该系统配置由AMD R3900x 12核处理器、48 GB内存和NVIDIA RTX 2080Ti 12 GB显卡组成。实验使用Python3.8版本内置的Pytorch框架进行训练，共迭代20000次，充足的迭代次数确保模型能够充分学习并收敛。在潜在耕地非农化区域建筑物提取模型训练过程中，模型训练初始学习率设置为0.01，更好地平衡收敛速度和模型性能的稳定性，避免过大的学习率在模型训练中出现振荡。对自适应矩估计（Adam）、随机梯度下降（SGD）及带权重衰减的自适应矩估计算法（AdamW）进

行性能对比实验，旨在筛选出对模型参数优化最为有效的优化器。MIoU计算了预测与真实结果之间重叠区域的平均比例，在进行图像分割任务时，MIoU能够提供更全面的性能评估，本文采用MIoU作为主要的评估指标，以确保选择优化器的客观性。表1展示了在不同迭代次数下，DHRformer模型采用上述3种优化器及各自权重衰减因子配置的MIoU结果。结果表明，当使用AdamW作为优化器，并将权重衰减因子WD（Weight Decay）设定为0.01时，DHRformer模型能够达到最优性能。

表1 不同衰减因子下使用3种优化器的DHRFormer的MIoU值  
Table 1 MIoU results for DHRFormer using three optimizers

| 迭代<br>次数 | Adam MIoU 值 |       | SGD MIoU 值 |       | AdamW MIoU 值 |       | /% |
|----------|-------------|-------|------------|-------|--------------|-------|----|
|          | WD=         | WD=   | WD=        | WD=   | WD=          | WD=   |    |
|          | 0.005       | 0.01  | 0.005      | 0.01  | 0.005        | 0.01  |    |
| 2000     | 63.21       | 55.43 | 52.74      | 47.1  | 63.83        | 67.83 |    |
| 8000     | 68.22       | 69.36 | 70.64      | 73.48 | 75.21        | 75.25 |    |
| 16000    | 75.16       | 74.06 | 73.68      | 74.85 | 77.95        | 79.13 |    |
| 20000    | 77.91       | 78.16 | 76.27      | 77.29 | 79.43        | 80.35 |    |

3.2 潜在非农化区域内建筑物提取实验结果及分析

为了验证模型对非农化区域内建筑物分割的有效性，与典型的语义分割模型和先进语义分割方法进行比较，如Deeplabv3+（Chen等，2018）、PSPNet（Zhao等，2017）、HRNet（Sun等，2019）和Segformer（Qiu等，2022），还包括2021年提出的ISA-Net（Huang等，2019）和K-Net（Zhang等，2021）分割模型。不同方法在非农化区域内建筑物分割结果如表2所示。可见：本文提出的DHRFormer方法相较于其他方法在F1-score、mAcc和MIoU取得了最高精度，分别为89.81%、89.37%及80.35%。其中：与Segformer对比分别提高了3.05%，2.68%及2.53%；比2021年提出的ISA-Net模型提高了3.46%、5%及3.11%；比K-Net模型分别提高了2.6%、4.8%及1.8%。此外，较之RF、SVM等传统算法F1-score分别提高了18.23%和20.51%，在MIoU方面提高25%以上。表明本文设计的优化方法可以有效提高对潜在非农化区域内建筑物的提取精度。

表2 多种分割方法在非农化区域建筑物分割精度结果评估

Table 2 Evaluation of the results of multiple segmentation methods for building segmentation accuracy in non-agriculturalised areas

| 方法         | 主干网络                 | F1-score | mAcc  | MIoU  |
|------------|----------------------|----------|-------|-------|
| DeepLabv3+ | Resnet101            | 86.61    | 84.98 | 77.60 |
| HRNet      | HRNetv2-w48          | 86.46    | 87.77 | 77.37 |
| ISA-Net    | Resnet50             | 86.35    | 84.37 | 77.24 |
| K-Net      | Swintransformer-tiny | 87.21    | 84.57 | 78.55 |
| PSPNet     | Resnet50             | 86.92    | 85.98 | 78.05 |
| Segformer  | VisionTransformer    | 86.76    | 86.69 | 77.82 |
| RF         |                      | 71.58    | 82.29 | 51.14 |
| SVM        |                      | 69.30    | 74.88 | 53.03 |
| DHRFormer  | HRNetv2-w48          | 89.81    | 89.37 | 80.35 |

3.3 与主流分割算法对比分割结果

图5直观显示了7个模型的可视化结果.从微观层面的视觉表现上可见总体上本文提出的方法更加准确地描述了潜在非农化区域内建筑物的空间信息,并且更接近真实标签。由图5第三行结

果可见：对于小型建筑物的提取，Deeplabv3+、ISANet、PSPNet、Segformer出现了明显的漏检，丢失了小尺寸建筑的分割结果。由图5第5行结果可见：HRNet和Segformer出现了明显的错误提取，其他模型也存在明显的漏检区域。由图5第3行和第5行对比结果可知，本文方法相较于其他方法能够准确识别小目标建筑，没有出现丢失其特征信息的现象，而且提取的建筑物斑块更加完整。这证明了以高分辨率主干网络和双分支解码高效融合多尺度信息，提升了网络对小目标建筑物提取能力。由图5第1行和第2行结果可见：Deeplabv3+、HRNet、K-Net及PSPNet模型出现了不同程度的粘连现象和空洞缺失情况。示例图像中建筑物与耕地相互交错，而本文方法却能够从交错环境中区分密集建筑物之间的边界，提取的建筑物并没有出现明显的小空洞和破碎现象。总体而言，本文方法比其他语义分割方法具有更好的分割效果，体现在小目标建筑物提取完整，建筑边界分割清晰，综合定性和定量结果证明了DHRformer模型在非农化区域内建筑物提取整体层面上的有效性。

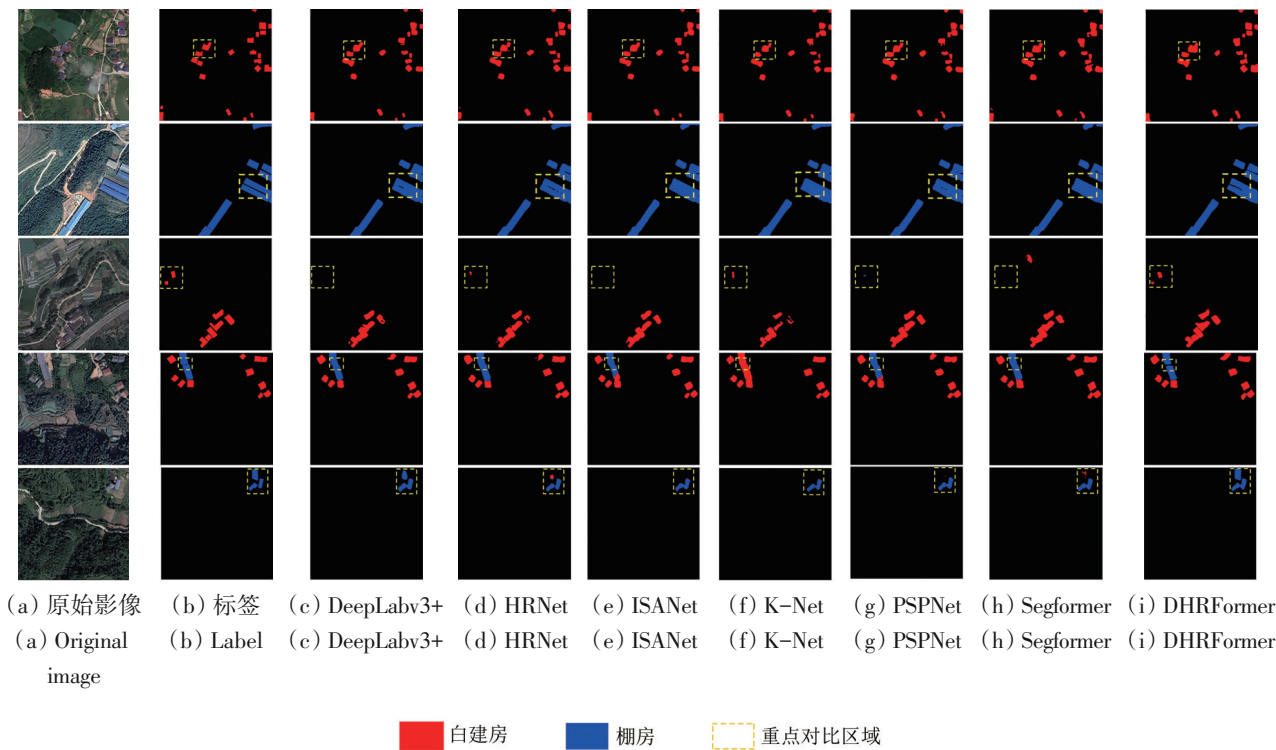


图5 不同网络的分割效果对比

Fig. 5 Comparison of segmentation results from different networks



3.4 与传统算法对比分析

图6显示了本文所提方法与RF、SVM等传统算法进行潜在非农化区域内建筑物提取结果示例。可见：RF算法提取的建筑物边界不够完整，且建筑物内部出现空洞情况，尤其在建筑物尺寸较小或与周围地物对比度较低时，存在明显的误检区域；SVM算法在潜在非农化区域建筑物分割任务中，表现得不理想，尤其是在建筑物与非建筑物

特征交叉显著的区域，产生了对建筑物的误判现象，出现了明显的错误提取；相比之下，使用本文提出的网络模型提取潜在非农化区域内建筑物较完整，最接近真实标签，并且保持较低的漏检和误检。因此，DHRformer模型与传统算法相比具有卓越的优势，体现在分割结果完整，很好地保留建筑物区域细节信息，能够避免漏检和错检的情况。

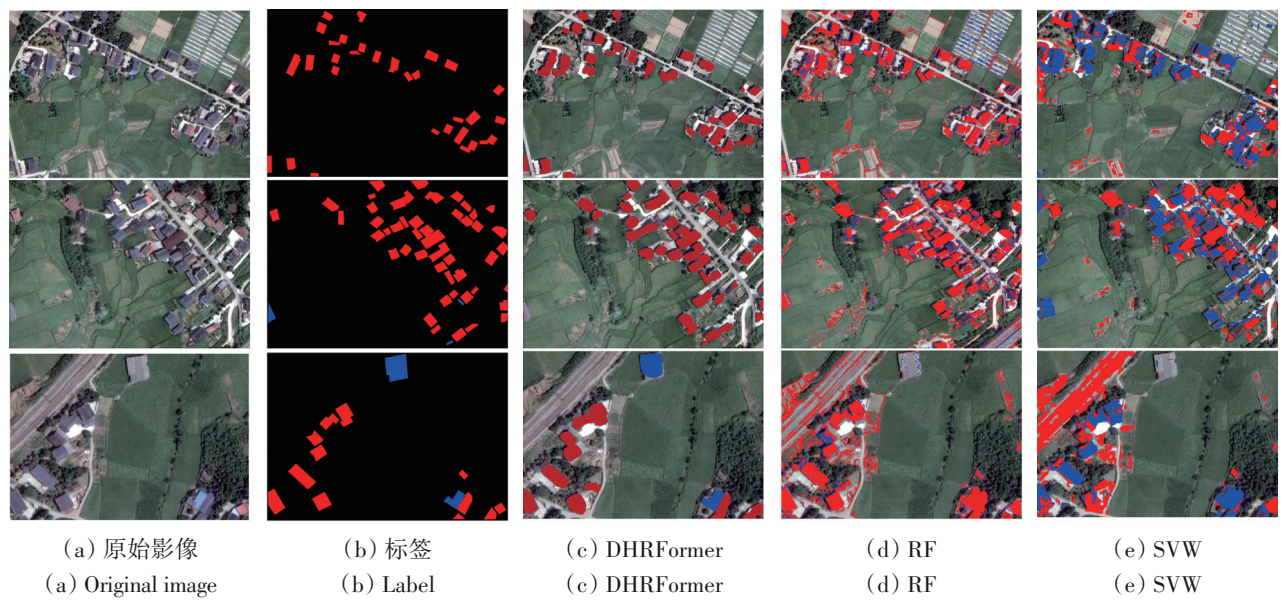


图6 与传统算法对比结果  
Fig. 6 Comparison results with traditional algorithm

3.5 消融实验

消融实验作为探索深度学习网络的研究手段之一，其目的是有计划的移除网络中特定模块或功能，以评估它们对网络性能的影响。为了了解DHRformer模型各个组件对于非农化区域内建筑物提取的具体效果，通过在网络编码器和解码器中分别移除高分辨率主干网络和空洞卷积分支，以联合语义损失函数作为网络的损失函数。消融实验结果精度结果见表3。可见：在编码器使用高分

辨率网络后，对潜在非农化区域内建筑物进行提取，精度有了一定提升，组合高分辨率网络和空洞卷积分支使模型的预测能力进一步得到提高。本文采用类激活图对消融实验结果进行可解释性分析，可视化结果见图7。可见DHRformer模型在类激活图上呈现更为明显的目标聚焦效果，与原模型相比，在非农化区域内提取建筑物的显著性更为突出，DHRformer模型提高了对潜在非农化区域内建筑物目标的关注度。

表3 消融实验结果精度评估  
Table 3 Evaluation of the precision of ablation test results

| 编号  |                   |           |         |       | /%    |       |
|-----|-------------------|-----------|---------|-------|-------|-------|
|     | VisionTransformer | HRNet-W48 | 双分支解码结构 | F1    | mAcc  | MIoU  |
| 实验1 | √                 | ×         | ×       | 86.76 | 86.69 | 77.82 |
| 实验2 | ×                 | √         | ×       | 86.85 | 88.24 | 79.97 |
| 实验3 | ×                 | √         | √       | 89.81 | 89.37 | 80.35 |

注：√代表DHRFormer网络使用该结构；×代表DHRFormer网络未使用该结构。

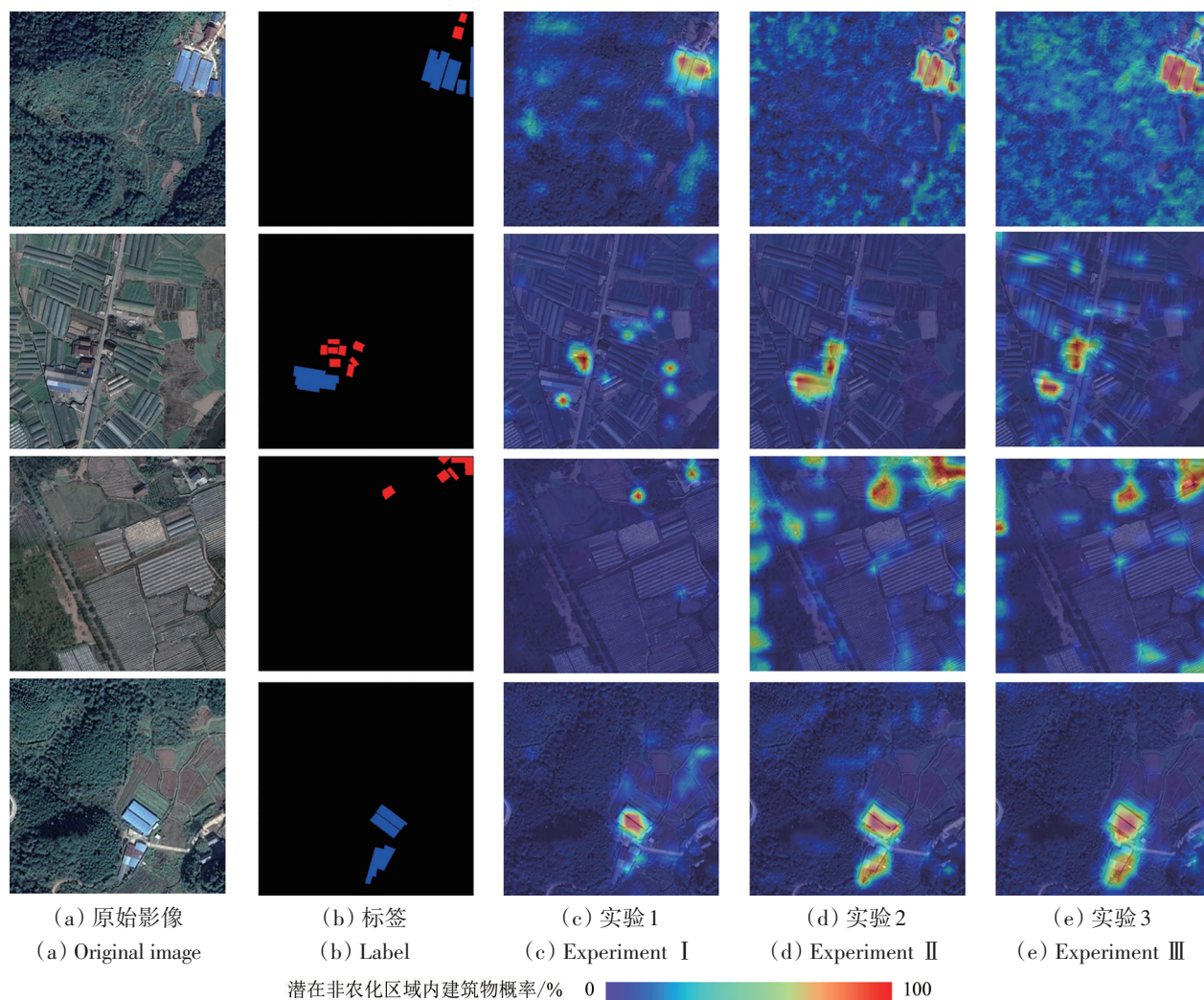


图7 消融实验结果可视化类激活图

Fig. 7 Visualization of class activation maps from ablation experiments

### 3.6 潜在非农化区域内建筑物监测应用

结合研究区后时相2022年Google Earth遥感影像数据和DHRformer模型进行大范围潜在非农化区域内建筑物快速监测,该影像数据分辨率为0.54 m,随机选取3个面积约为1.5 km<sup>2</sup>的区域。首先,将后时相遥感影像数据进行校正处理得到正射影像;其次,将耕地矢量范围与处理后的高分辨率影像叠加后使用DHRformer模型监测。监测结果如图8所示。可见本文提出的DHRformer模型对于大范围的潜在非农化区域内建筑物提取结果较为完整,且边界无明显破碎现象。因此,本文方法在大范围提取潜在非农化区域内建筑也能取得相对令人满意的性能。在耕地范围内搭建棚房的现象比自建房稍显突出,这需要结合当地实际生产状况裁定是否存在占用耕地进行生产建设活动。

## 4 结 论

本研究提出的DHRFormer模型在潜在耕地非农化区域建筑物提取任务中表现优于HRNet、Segformer、K-Net等当前主流深度语义分割方法和传统机器学习方法。与Segformer相比,DHRFormer在F1、mAcc和MIoU上分别提升了3.05%、2.68%和2.53%,说明其在非农建筑边界精细化提取方面具有更强的能力,验证了模型结构设计的合理性。与RF和SVM等传统方法相比,DHRFormer能够显著减少建筑物提取中的边界断裂、内部空洞以及相邻建筑物粘连等问题。这一优势主要源于深度特征表达能力对复杂场景的适应性,而传统依赖光谱和几何特征的方法在高密度、形态多样的建筑群提取中难以避免误判和漏判。



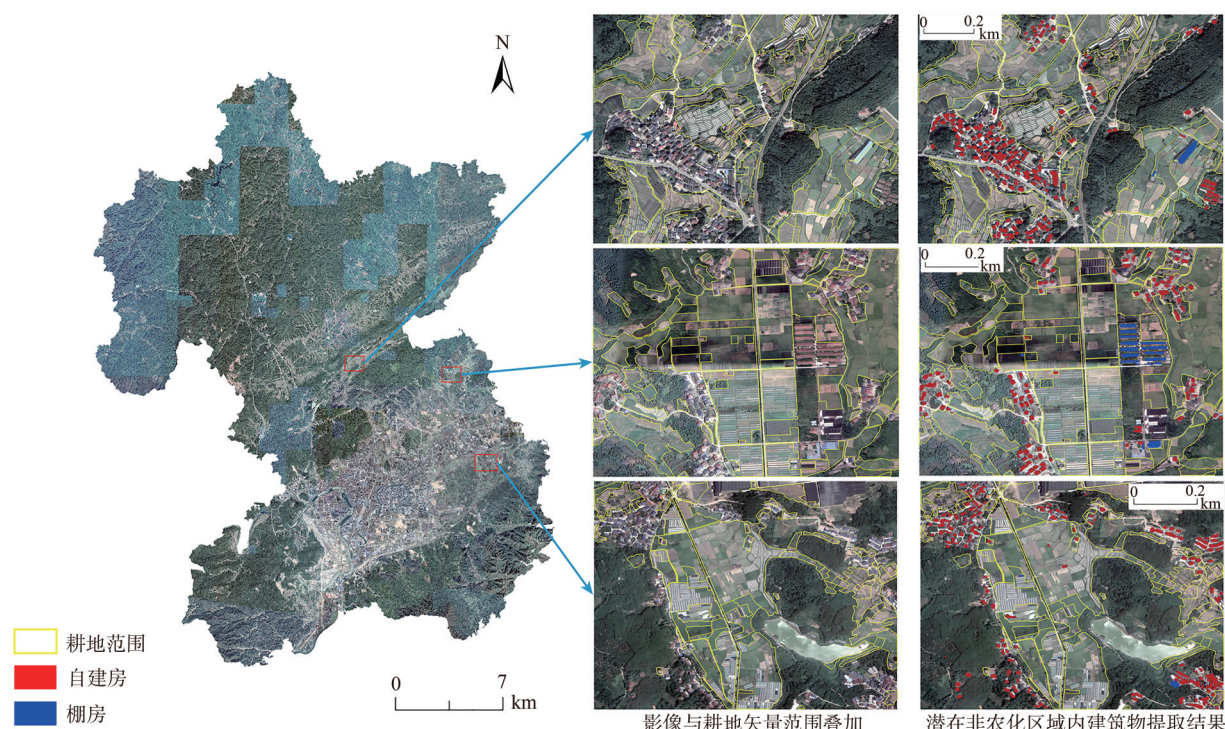


图8 研究区大范围耕地非农化区域内建筑物监测结果

Fig. 8 Monitoring results of buildings within large-scale cropland non-agriculturalized areas in the study region

不过，本文研究仍存在一定局限性：一是监测对象仅聚焦于潜在非农化区域内的非农建筑物，未涵盖绿化造林、挖湖造景等其他类型的非农化行为；二是数据源仅包含两期高分辨率遥感影像，时间序列有限，难以全面揭示动态变化过程。这些限制可能影响模型在更广泛非农化监测任务中的适用性。未来，可在保持现有框架优势的基础上，引入多时相、多源遥感数据及对象级时序分析方法，以提升监测的全面性和时效性。

为提升潜在耕地非农化行为中建筑物的提取精度，本文提出了基于高分辨率遥感影像的DHRFormer深度学习模型，并以湖南省怀化市鹤城区耕地区域为研究对象，验证了该方法的有效性。主要结论如下：

(1) 利用高分辨率网络重构编码器，以维持高分辨率遥感影像在模型运算全过程的高语义信息；

(2) 设计了不同扩张率的空洞卷积双分支解码器，实现多尺度特征逐步融合与细化，显著减少了相邻建筑物的粘连现象；

(3) 设计联合协作损失函数，优化了类别不平衡问题，以及非边缘像素主导损失函数的问题。

## 参考文献 (References)

- Belgiu M and Csillik O. 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204: 509-523 [DOI: 10.1016/j.rse.2017.10.005]
- Chen D L, Lu H B and Kuang B. 2019. Measurement of cultivated land utilization efficiency: construction and application of random forest. *Journal of Natural Resources*, 34(6): 1345-1356 (陈丹玲, 卢新海, 匡兵. 2019. 基于随机森林的耕地利用效率测度模型构建及其应用. *自然资源学报*, 34(6): 1345-1356) [DOI: 10.31497/zrzyxb.20190617]
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//15th European Conference on Computer Vision. Munich: Springer, 833-851 [DOI: 10.1007/978-3-030-01234-2\_49]
- Chen Y M, Yao M R, Zhao Q Q, Chen Z J, Jiang P H, Li M C and Chen D. 2021. Delineation of a basic farmland protection zone based on spatial connectivity and comprehensive quality evaluation: a case study of Changsha City, China. *Land Use Policy*, 101: 105145 [DOI: 10.1016/j.landusepol.2020.105145]
- El-Naggar A M. 2018. Determination of optimum segmentation parameter values for extracting building from remote sensing images. *Alexandria Engineering Journal*, 57(4): 3089-3097 [DOI: 10.1016/j.aej.2018.10.001]
- Fan S L, Qiu L J, Ru K L, Chen Q Y and Hu Y. 2018. Classification



- method of agricultural land quality based on back-propagation neural network and support vector machine. *Journal of China Agricultural University*, 23(12): 138-148 (范胜龙, 邱凌婧, 茹凯丽, 陈巧燕, 胡勇. 2018. 基于BP神经网络和支持向量机的农用地分等方法研究. *中国农业大学学报*, 23(12): 138-148) [DOI: 10.11841/j.issn.1007-4333.2018.12.17]
- General Office of the State Council of the People's Republic of China. 2020. Circular of the general office of the state council on resolutely stopping the "non-agriculturalization" of cultivated land. *Gazette of the State Council of the People's Republic of China*, (27): 36-37 (国务院办公厅. 2020. 国务院办公厅关于坚决制止耕地“非农化”行为的通知. *中华人民共和国国务院公报*, (27): 36-37)
- Hua Y S, Marcos D, Mou L C, Zhu X X and Tuia D. 2022. Semantic segmentation of remote sensing images with sparse annotations. *IEEE Geoscience and Remote Sensing Letters*, 19: 8006305 [DOI: 10.1109/LGRS.2021.3051053]
- Huang L, Yuan Y H, Guo J Y, Zhang C, Chen X L and Wang J D. 2019. Interlaced sparse self-attention for semantic segmentation [2024-02-22]. <https://arxiv.org/abs/1907.12273> [DOI: 10.48550/arXiv.1907.12273]
- Ji S P, Wei S Q and Lu M. 2019. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1): 574-586 [DOI: 10.1109/TGRS.2018.2858817]
- Li C N, Fu L, Zhu Q, Zhu J, Fang Z, Xie Y K, Guo Y K and Gong Y H. 2021. Attention enhanced U-net for building extraction from farmland based on Google and WorldView-2 remote sensing images. *Remote Sensing*, 13(21): 4411 [DOI: 10.3390/rs13214411]
- Li F T, Zhang G, Shi T T and Li L. 2022. Classification system and application of remote sensing interpretation samples of cultivated land non-agriculturalization. *Transactions of the Chinese Society of Agricultural Engineering*, 38(15): 297-304 (厉芳婷, 张过, 石婷婷, 李乐. 2022. 耕地“非农化”遥感解译样本分类体系及应用. *农业工程学报*, 38(15): 297-304) [DOI: 10.11975/j.issn.1002-6819.2022.15.032]
- Liao C, Hu H, Li H F, Ge X M, Chen M, Li C N and Zhu Q. 2021. Joint learning of contour and structure for boundary-preserved building extraction. *Remote Sensing*, 13(6): 1049 [DOI: 10.3390/rs13061049]
- Liasis G and Stavrou S. 2016. Building extraction in satellite images using active contours and colour features. *International Journal of Remote Sensing*, 37(5): 1127-1153 [DOI: 10.1080/01431161.2016.1148283]
- Notice on the plan to carry out the special clean-up and rectification action of "greenhouse" to resolutely curb the non-agriculturalization of agricultural land. *Gazette of the Ministry of Agriculture and Affairs of the People's Republic of China*, (10): 49-52 (2018. 农业农村部自然资源部印发《关于开展“大棚房”问题专项清理整治行动坚决遏制农地非农化的方案》的通知农农发[2018] 3号. *中华人民共和国农业农村部公报*, (10): 49-52)
- O'Connell J, Bradter U and Benton T G. 2015. Wide-area mapping of small-scale features in agricultural landscapes using airborne remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109: 165-177 [DOI: 10.1016/j.isprsjprs.2015.09.007]
- Peng C, Li Y Y, Jiao L C and Shang R H. 2021. Efficient convolutional neural architecture search for remote sensing image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7): 6092-6105 [DOI: 10.1109/TGRS.2020.3020424]
- Qiu C P, Li H, Guo W Y, Chen X, Yu A Z, Tong X C and Schmitt M. 2022. Transferring transformer-based models for cross-area building extraction from remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 4104-4116 [DOI: 10.1109/JSTARS.2022.3175200]
- Shao Z F, Li C M, Li D R, Altan O, Zhang L and Ding L. 2020. An accurate matching method for projecting vector data into surveillance video to monitor and protect cultivated land. *ISPRS International Journal of Geo-Information*, 9(7): 448 [DOI: 10.3390/ijgi9070448]
- Sun K, Zhao Y, Jiang B R, Cheng T H, Xiao B, Liu D, Mu Y D, Wang X G, Liu W Y and Wang J D. 2019. High-resolution representations for labeling pixels and regions[EB/OL]. *arXiv*. [2024-02-22]. <https://arxiv.org/abs/1904.04514> [DOI: 10.48550/arXiv.1904.04514]
- Tan M, Lin H J and Hao M. 2023. Prior knowledge guided deep learning for monitoring buildings and greenhouses within cultivated land. *Journal of Geo-information Science*, 25(11): 2293-2302 (谭敏, 林惠晶, 郝明. 2023. 先验知识引导深度学习的耕地范围建筑物和大棚房监测方法. *地球信息科学学报*, 25(11): 2293-2302) [DOI: 10.12082/dqxxkx.2023.230315]
- Useya J, Chen S B and Murefu M. 2019. Cropland mapping and change detection: toward Zimbabwean cropland inventory. *IEEE Access*, 7: 53603-53620 [DOI: 10.1109/ACCESS.2019.2912807]
- Wang X L, Hu Z R, Shi S H, Hou M, Xu L and Zhang X. 2023. A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved UNet. *Scientific Reports*, 13(1): 7600 [DOI: 10.1038/s41598-023-34379-2]
- Wu J X. 2012. Efficient HIK SVM learning for image classification. *IEEE Transactions on Image Processing*, 21(10): 4442-4453 [DOI: 10.1109/TIP.2012.2207392]
- Wu Y Z, Shan L P, Guo Z and Peng Y. 2017. Cultivated land protection policies in China facing 2030: dynamic balance system versus basic farmland zoning. *Habitat International*, 69: 126-138 [DOI: 10.1016/j.habitatint.2017.09.002]
- Zhang W W, Pang J M, Chen K and Loy C C. 2021. K-Net: towards unified image segmentation//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. Curran Associates Inc.: 790
- Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]
- Zhu Y, Pan Y Z and Zhang D J. 2022. An agriculture parcel identification method based on convolutional neural network and water-

shed segmentation. *Journal of Geo-information Science*, 24(12): 2389-2403 (朱昱, 潘耀忠, 张杜娟. 2022. 基于深度卷积网络和

分水岭分割的耕地地块识别方法. *地球信息科学*, 24(12): 2389-2403) [DOI: 10.12082/dqxxkx.2022.220447]

## Remote sensing monitoring of cultivated land via integration of a high-resolution network and a dual-branch structure

LIU Zhen<sup>1,2</sup>, LIU Deer<sup>1,2</sup>, ZHAO Chen<sup>3</sup>

1. School of Civil and Surveying Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China;

2. Jiangxi Provincial Key Laboratory of Water Ecological Conservation at Headwater Regions, Ganzhou 341000, China;

3. Fujian Jingwei Digital Technology Co., Ltd., Fuzhou 350001, China

**Abstract:** Arable land is the basis for ensuring sustainable agricultural development, and rapid and precise monitoring of the nonfarming behavior of arable land is of great significance to China's food production and security.

To accurately monitor illegal occupation of cultivated land for non-agricultural construction, this study proposes a novel high-resolution remote sensing image segmentation method for detecting non-agricultural use of cultivated land, namely the Dilation-enhanced High-Resolution Former (DHRformer). The model is composed of a high-resolution network and a dual-branch decoder structure, which enhances the feature representation of non-agricultural buildings through multi-scale fusion and dilation strategies, thereby capturing richer building detail information. First, a sub-meter-level building sample dataset of potential non-agricultural areas was constructed based on multi-temporal remote sensing data. Then, the proposed DHRformer model was used to extract buildings within these potential non-agricultural areas.

Finally, taking the mountainous cultivated land in Hecheng District, Huaihua City as the study area, validation experiments were conducted using high-resolution remote sensing imagery. The results show that the proposed algorithm achieved improvements in MIoU, mAcc, and F1-score, with increases of 2.53%, 2.68%, and 3.05%, respectively, compared with the Segformer network.

Therefore, the proposed method demonstrates high practical value in monitoring non-agricultural use of cultivated land, providing technical reference and theoretical support for the detection of complex non-agricultural activities.

**Key words:** non-agriculturalization of cultivated land, building segmentation, high resolution network, two-branch decoding structure, encoder-decoder, high resolution image

**Supported by** National Natural Science Foundation of China (No. 42271434)