

结合空间注意力机制与多任务学习的耕地地块分割模型

田富有¹, 曹玉佩^{2,3}, 赵航^{1,4}, 吴炳方^{1,4}, 曾红伟^{1,4}, 刘亚洲³,
覃星力¹, 张淼¹, 朱亮¹, 朱伟伟¹

1. 中国科学院空天信息创新研究院 遥感科学国家重点实验室, 北京 100094;

2. 北京师范大学, 北京 100101;

3. 山东土地集团数字科技有限公司, 枣庄 277000;

4. 中国科学院大学, 北京 100049

摘要: 地块作为农业耕作的最小单元, 对其精准识别是国土资源监测、耕地利用监测的需要。现有的方法多使用手工勾绘的方式获取, 耗时费力, 成本高昂, 并且无法实现实时、近实时更新。本文设计了一种基于空间注意力机制与多任务学习的地块分割模型—Field-Net。模型基于UNet架构, 增加了空间注意力机制, 并采用多任务学习的策略, 在语义分割的基础上增加了边界、像素到地块边界的距离等任务。在山东省东营市利津县对模型的性能进行了测试, 结果发现耕地地块识别的交并比达到了87.05%, 总体精度为92.23%。Field-Net模型的性能优于几种高性能的深度学习框架, 交并比较Link-Net模型高出0.26%, 较DeepLab v3+高出7.59%。在空间泛化性能测试中, Field-Net模型的平均交并比Link-Net模型高出3.51%, 空间泛化性能明显提升。通过消融试验发现, 使用空间注意力机制的Field-Net较ResUNet模型F₁-Score提高了1.01%, 交并比提高了1.6%; 多任务学习策略使得Field-Net模型的F₁-Score提高了0.18%, 交并比提高了0.21%; 将模型权重特征进行可视化后发现空间注意力机制模块和多任务学习策略可以使模型学习到的特征更加聚集于地块边界和地块内部, 使学习到的特征更具代表性。总体而言, Field-Net模型可以支撑地块级别国土资源和耕地非农化、非粮化利用监测, 从而提高监测的效率和时效性。

关键词: 地块分割, Field-Net模型, 空间注意力机制, 多任务学习, 高分卫星数据

中图分类号: TP701/P2

引用格式: 田富有, 曹玉佩, 赵航, 吴炳方, 曾红伟, 刘亚洲, 覃星力, 张淼, 朱亮, 朱伟伟. 2024. 结合空间注意力机制与多任务学习的耕地地块分割模型. 遥感学报, 28(11): 2850-2864

Tian F Y, Cao Y P, Zhao H, Wu B F, Zeng H W, Liu Y Z, Qin X L, Zhang M, Zhu L and Zhu W W. 2024. Agricultural Field Segmentation using Spatial Attention Mechanism and Multi-task Learning Strategy. National Remote Sensing Bulletin, 28(11): 2850-2864 [DOI: 10.11834/jrs.20243191]

1 引言

地块是进行国土资源耕地利用监测的最小单元, 其准确提取对于耕地非农化、非粮化监测, 精准提供农业种植补贴, 保障国家粮食安全具有重要的意义 (杨颖频 等, 2021; 程锐 等, 2022; 朱昱 等, 2022)。但是现有的地块信息主要依赖人

工勾画, 比如历次国土资源调查, 耗费人力物力, 而且每10年更新一次, 更新速度和获取成本很难满足耕地精准监测需求 (韩衍欣和蒙继华, 2019; 李博, 2020)。

地块是农户进行播种、管理、收获等行为的最小单元 (宋茜 等, 2020), 理论上也应该成为耕地利用监测、作物面积、长势、产量估算的最小

收稿日期: 2023-06-02; 预印本: 2023-11-11

基金项目: 国家重点研发计划 (编号: 2019YFE0126900); 第三次新疆综合科学考察 (编号: 2022xjkk030106), 国家自然科学基金 (编号: 41861144019), 深化科研院所改革农业遥感创新团队项目 (编号: E33D0201-6)

第一作者简介: 田富有, 研究方向为遥感信息智能提取与基于地理大数据的水土资源监测。E-mail: tianfy@aircas.ac.cn

通信作者简介: 吴炳方, 研究方向为生态系统遥感, 农业遥感, 水资源遥感等。E-mail: wubf@aircas.ac.cn

单位(Wu等, 2023)。由于耕地的光谱高度异质性, 即使一个地块内, 光谱特征由于长势、土壤背景、物候等差异也不尽相同, 导致现有的耕地监测以面向像素、面向对象为主, 尚未能实现地块级别的精准识别(Persello等, 2019)。随着深度学习技术的迅速发展, “端到端”的语义分割模型可以通过学习地块的空间上下文信息实现地块分割(Tian等, 2023)。

利用深度学习方法进行地块分割成为耕地利用监测的热点之一(Waldner和Diakogiannis, 2020; 张乾坤等, 2022)。杨颖频等(2021)利用时空协同特征, 结合D-LinkNet模型实现广西扶绥县的耕地地块提取。张海天等利用改进Unet++模型, 实现丘陵地区耕地地块深度分割。周楠等(2021)将地块边缘检测作为地块提取的副线任务, 实现了湖南邵东县的地块提取。朱昱等(2022)利用D-LinkNet模型, 结合分水岭分割算法对地块边缘获取进行分水岭优化, 取得了较好的效果。但是现有模型仅仅考虑了地块边缘这一信息, 且未能在模型中有效利用地块形状边界特征, 同时存在严重依赖训练样本, 模型泛化能力较弱的问题, 严重制约耕地地块分割模型的广泛应用。已有的研究表明, 多任务学习的方法通过学习主线任务的相关特征, 提高模型的泛化性能。模型同时完成多个任务的学习, 可以从共享的特征中学习到相关信息, 从而引导模型学习到更多关联特征, 提高模型的通用性与可迁移性(Takikawa等, 2019; Waldner和Diakogiannis, 2020)。

随着深度学习技术的发展, 时空注意力机制模块成功被应用于遥感语义分割任务中(林缪鹏等, 2023)。U-net模型和类Unet模型在遥感影像分类任务中取得了巨大的成功。但是以Unet模型为代表的通用深度学习框架大多学习颜色、纹理等信息, 对地块分割任务中形状特征不敏感, 存在对地块边缘分割不准确, 边界特征未能有效利用的问题。受人类对图像理解过程的启发, 注意力机制被应用到语义分割模型中(Vaswani等, 2017)。人类不是一次性地处理整个图像, 而是在全局浏览之后关注其中的重要部分。最近, Woo等(2018)通过整合通道和空间注意力两种注意力机制, 提出了注意力模块(CBAM), 研究表明, 空间注意力机制可以引导模型专注于空间形状信息(Tian等, 2023)。

因此, 本文提出一种基于多任务学习与空间注意力机制的Field-Net模型, 旨在通过多任务学习的方法, 让模型学习到更多的地块边界相关特征, 提高地块边界的精度和模型的泛化能力。本研究在山东省东营市利津县展示试验, 分析Field-Net模型的适用性, 该区域位于华北平原这一地块破碎、“非粮化”监测的重点区域。

2 研究区介绍

山东省东营市利津县位于山东省东北部、渤海西南岸、黄河近口段左侧(图1)。地跨东经 $118^{\circ}07'E-118^{\circ}54'E$, $37^{\circ}22'N-38^{\circ}12'N$, 总面积 1665.6 km^2 。截至2021年, 区域内耕地面积为 55137.16 ha , 林地 538.42 ha , 常驻人口23.84万人(利津县政府, 2021)。利津县处于暖温带半湿润季风气候, 冬春干旱, 夏季多雨, 区域内主要种植小麦、玉米等粮食作物, 人均耕地面积3.7亩, 农田地块较为破碎, 农田生态系统种植结构复杂, 是华北平原典型的农田区, 适用于检验地块分割模型的性能。

3 数据

研究所用的数据以2021年9月份的高分系列和资源系列遥感影像资料为主, 其中高分系列主要选用了GF-1D获取的2021年9月21日的两景影像(空间分辨率为全色 2 m , 多光谱为 8 m), 同时选用了资源三号(空间分辨率为全色 2.1 m , 多光谱为 5.8 m)在2021年9月29日获取的一景影像。针对获取的影像, 我们利用ERDAS IMAGE 2010软件进行几何精纠正, 校正精度控制在0.5个像素之内, 同时对进行全色与多光谱数据的融合处理, 形成 2 m 分辨率的遥感影像。

所用的耕地地块边界是根据《第三次全国国土调查技术规程》在ArcGIS10.8中手工勾绘完成, 同时参照了盐窝镇的三调数据, 根据标准假彩色合成的遥感影像地块边界的可分性, 勾绘完成利津县盐窝镇 222.18 km^2 , 以及模型泛化性能测试区 71.75 km^2 的地块矢量边界, 经过2人按照10%的比例进行交叉检查, 一致性大于98%(图2)。

为了构造训练样本, 将绘制的矢量边界文件, 转换成与遥感影像同样分辨率的栅格文件, 然后与遥感影像一同进行切片处理。具体方法为: 将标签文件与遥感影像文件裁切成 256×256 像素的影

像块，同时横向和纵向分别设置 128 像素的重叠，增加样本的数量与代表性，最后制作成 3480 个 256×256 的影像块，其中 3000 个用于模型的训练，360 个用于验证，120 个用于模型的空间泛化性能测试（图 1）。

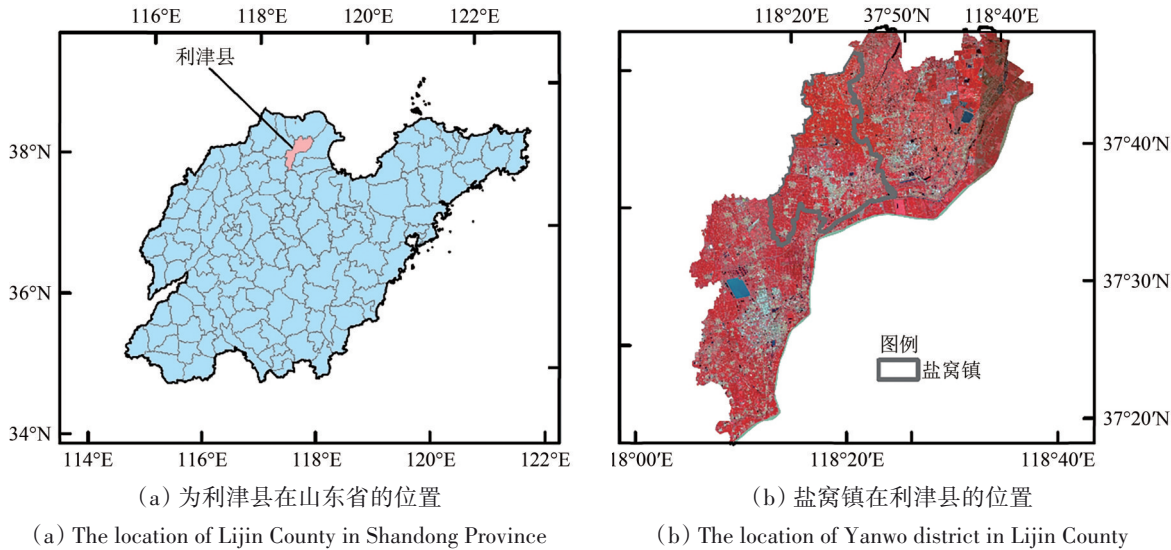


图1 研究区位置示意图

Fig. 1 The location of study area and satellite images

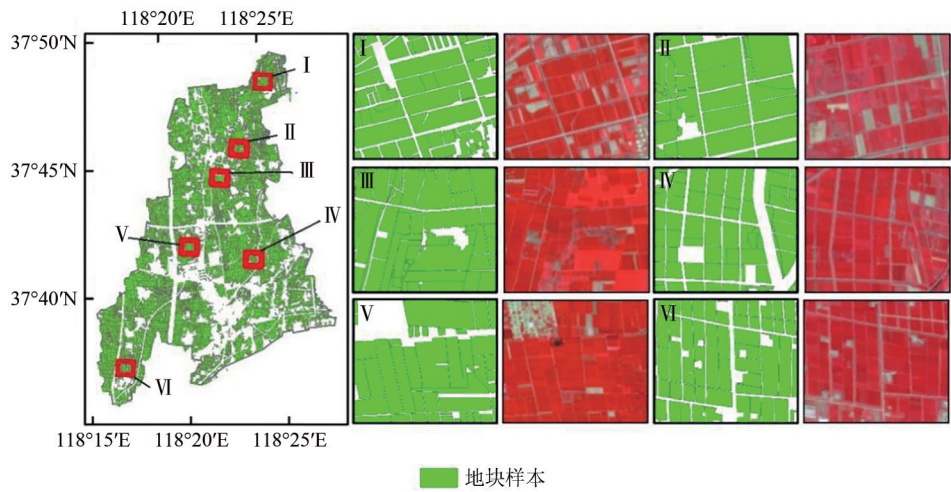


图2 地块样本示意图

Fig. 2 Example of agricultural filed samples

4 方法

4.1 Field-Net 网络结构

本文提出基于深度学习的遥感影像地块分割方法 Field-Net，该网络是在 U-Net 骨干网络上增加空间注意力机制和多任务学习策略构建而成的。U-Net 是经典的编码器—解码器语义分割网络 (Ronneberger 等，2015)，该网络被广泛应用于生物医学领域的分割任务，因其简洁的编解码结构及

优越的性能很快成为遥感影像语义分割任务的基础网络 (Falk 等，2019)。U-Net 和类 U-Net 的模型框架，例如 ResUNet (Zhang 等，2018；Waldner 和 Diakogiannis，2020)，在一些遥感图像分割比赛中显示出较高的精度 (Saraiva 等，2020)。

Field-Net 模型包含下采样和上采样两个阶段，分别对应编码器和解码器部分。编码器部分将数据降采样为高维图像，而解码器部分将恢复上采样压缩的信息成像素级的输出，同时保持其维度

与输入相同。因此, 它可以将高层次的空间信息和低层次特征进行空间串联 (Jeppesen 等, 2019)。模型总体框架图如图 3 所示, 模型中每个卷积层之后都有一个批处理归一化层。

对于一个输入影像 $X \in R_{w \times h \times c}$, 经过 Field-Net 计算输出之后, 得出一个掩膜图、边界图的概率结果, 其范围为 0—1, 以及像素到地块边界的距离。

输入影像为 $256 \times 256 \times 4$ (红、绿、蓝与近红外), 在模型的编码阶段, 经过了 5 个残差块 (Residual block) 和 4 个最大池化层 (Max Pooling), 经过编码器后, 模型输出为 $16 \times 16 \times 512$ 的特征图, 经过具有 4 个残差块和上采样的解码器后, 输出的图像被还原成 3 个 $256 \times 256 \times 1$ 的特征图, 分别为掩膜、边界以及到边界的距离图。

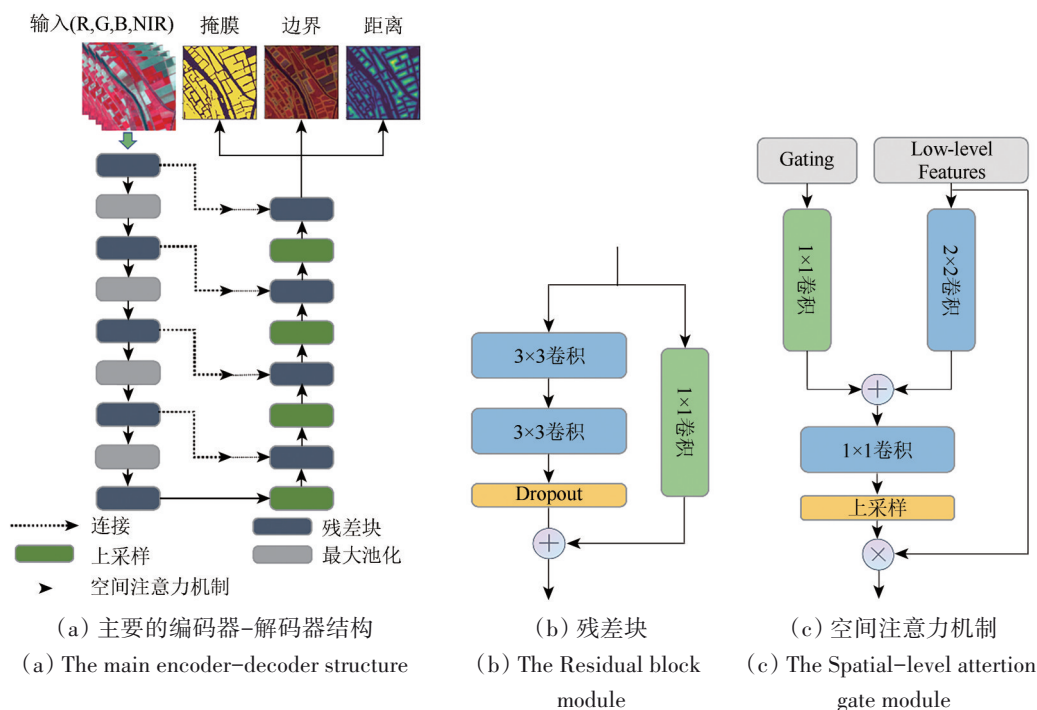


图 3 Field-Net 模型框架结构

Fig. 3 The Field-Net model architecture

4.2 残差块

受 Res-Net 的启发, 因为随着神经网络深度的增加, 会存在梯度爆炸和消失的问题 (Yi 等, 2019), 在 Field-Net 中用一个残差块取代了单一的卷积操作 (图 3(b))。残差块的目的并不是直接解析目标输出, 而是尝试解决输出和输入之间的差值, 这种方式称为“残差” (He 等, 2016)。如式 (1) 所示。

$$y_i = h(x_i) + F(x_i, \theta_i) \quad (1)$$

式中, x_i 表示残差块的输入, y_i 是输出目标, $F(x_i, \theta_i)$ 表示残差部分, 式中 $h(x_i) = \theta_i^1 x$, θ_i^1 表示 1×1 卷积核, 用来增加或减少特征层的通道数以匹配 $F(x_i, \theta_i)$ 中的特征图。 $F(x_i, \theta_i)$ 包含两个卷积层, 然后是两个批量归一化层和一个 dropout 层。每个卷积层的滤波器大小为 3×3 , 并具有相同

的填充。滤波器的数量等同于目标输出的维度。 dropout 层将随机地对每个神经元以一定的概率进行保留或者丢弃, 用以防止过拟合问题的发生 (Jeppesen 等, 2019), 这里将 dropout 设置为 0.25 (Tian 等, 2023)。

在 Field-Net 中, 有 9 个残差块。其中 5 个是在编码器部分, 4 个是在解码器部分。编码器部分残差的输入是原始图像或最大池化层的输出。最大池化层是为了得到 2×2 区域内的最大值, 因此宽度和高度将被减半, 但通道数将保持不变。残差块的输入尺寸将相应地改变。解码器中残差块的输入是前一个残差块上采样结果和相应的编码器的输出之间的连接。上采样是最大池化层的逆向操作, 将结果存储在 2×2 的区域内, 并将空值用 0, 因此宽度和高度将翻倍。

4.3 空间注意力机制

在人类脑的启发下,注意力机制在深度学习领域已被证明是一个有用的“核心设计”(Wang等, 2017; Woo等, 2018; Fu等, 2019)。人类在浏览一张图片之后,通常会关注更吸引人的地方。因此,学者试图将这一机制引入到卷积神经网络CNN架构中以提高语义分割任务的性能(Wang等, 2017)。受空间注意力机制的启发,我们在下采样和上采样之间的连接中添加了4个注意力模块(图3(c))。在空间注意力机制中,低维特征与高维特征一起更新(图3(a)左半部分如式(2)–(3)所示)。

$$y_l = x_l \otimes g_{x_l, x_h} \quad (2)$$

$$g_{x_l, x_h} = F_{\text{sigmoid}}\left(\left(\theta_l^1 x_h + F_{\text{relu}}(x_l, \theta_l^2)\right), \theta_l^1\right) \quad (3)$$

式中, $x_l \in \mathbb{R}^{w_l \times h_l \times c_l}$ 与 $x_h \in \mathbb{R}^{w_h \times h_h \times c_h}$ 是低维与高维输入特征, x_l 是编码器残差块的输出, 而 x_h 是解码器残差块的输出。 w_l, h_l 的通道数分别是 w_h, h_h 的两倍, $\theta_l^1 x_h$ 是 1×1 卷积层用来调整通道数。 $F_{\text{relu}}(x_l, \theta_l^2)$ 是一个卷积核为 2×2 的卷积层, 然后以 2×2 的步长进行填充, 作为激活函数的输入。得到的结果将加入到高维特征 $\theta_l^1 x_h$ 中。 $F_{\text{sigmoid}}(\cdot, \theta_l^1)$ 表示 1×1 的卷积核的 sigmoid 激活层。 g_{x_l, x_h} 经过先前的 1×1 的卷积核后变为 $w \times h \times 1$ 。 $\otimes$ 表示低维特征与学习到的特征之间的点积。最后, 原始的低维特征通过与学习到的空间特征 (g_{x_l, x_h}) 相乘将被更新。

4.4 多任务学习

多任务学习的定义为一个深度学习模型通过特征共享的方式完成多个任务, 已有研究表明, 多任务学习能够通过增加副线任务, 增加主线任务的精度和泛化性能。形状信息是识别耕地地块的另一个关键特征。Sun等(2020)提出了一个用于医学图像分割的形状注意力U-Net模型, 并在两个公共医学图像分割数据集上取得了较高精度的结果。Tian等(2023)提出了一个用于圆形喷灌地提取的Pivot-Net, 并完成美国圆形喷灌地的高精度提取。

在Field-Net中, 除了分割任务之外, 边界与像素到边界的距离作为另外的两项任务。通过增加边界预测任务, 可以增加模型对地块边界特征的提取能力。地块分割实际上是个例分割的任务,

现有的语义分割模型通过监测地块间的缝隙达到分割地块的任务, 从而减少计算量。因此增加像素到地块边界的距离这一项回归任务, 可保障相邻地块之间的可分性。

像素到地块边界的距离利用OpenCV库中的距离变换函数 `cv2.distanceTransform()` 进行计算, 该函数可以计算各个像素与值为0的像素点的归一化距离。然后再基于距离图, 生成一定宽度的地块边界图。通过反复试验, 地块边界宽度设置为4像素效果最佳(图4)。

4.5 损失函数

在Field-Net多任务学习中, 识别两种类别, 地块与背景。Dice系数(又称 F_1 -Score)是对两个集合之间的重叠和相似性的测量。相应的Dice损失函数(Dice loss), 可以使模型在精度和召回率方面有更好的表现。因此, Dice loss被用来预测耕地范围。Dice loss如式(4):

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (4)$$

式中, A 和 B 是可计算集合(countable sets)。那么带有平滑系数的Dice loss可以用式(5)表示:

$$L_{\text{Dice}} = 1 - \frac{2|A \cap B| + \text{smooth}}{|A| + |B| + \text{smooth}} \quad (5)$$

式中, smooth是一个平滑系数, 通常设置为1。如上所述, Field-Net是用来预测耕地范围和边界的双重任务。因此, 最终的损失函数被定义为分割和边界检测的加权损失之和。

像素到地块边界的距离这一回归任务的损失函数则直接使用平均绝对误差(MAE), 其定义如下:

$$L_{\text{MAE}} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (6)$$

式中, y_i 为像素 i 到地块边界的真实值, x_i 为像素 i 到地块边界的预测值

$$L_{\text{total}} = \lambda_1 L_{\text{Dice-seg}} + \lambda_2 L_{\text{Dice-boundary}} + \lambda_3 L_{\text{MAE}} \quad (7)$$

式中, L_{total} 表示总损失, λ_1, λ_2 与 λ_3 是分割任务、边界预测、像素到边界距离回归任务损失权重系数。分别针对3个参数进行梯度对比试验, 对掩膜损失权重 λ_1 设置为1, 2, 3, 4时, λ_2, λ_3 分别设置为1; 对边界预测损失权重 λ_2 设置为1, 2, 3, 4时, λ_1, λ_3 分别设置为1; 像素到边界距离损失权重 λ_3 同理。

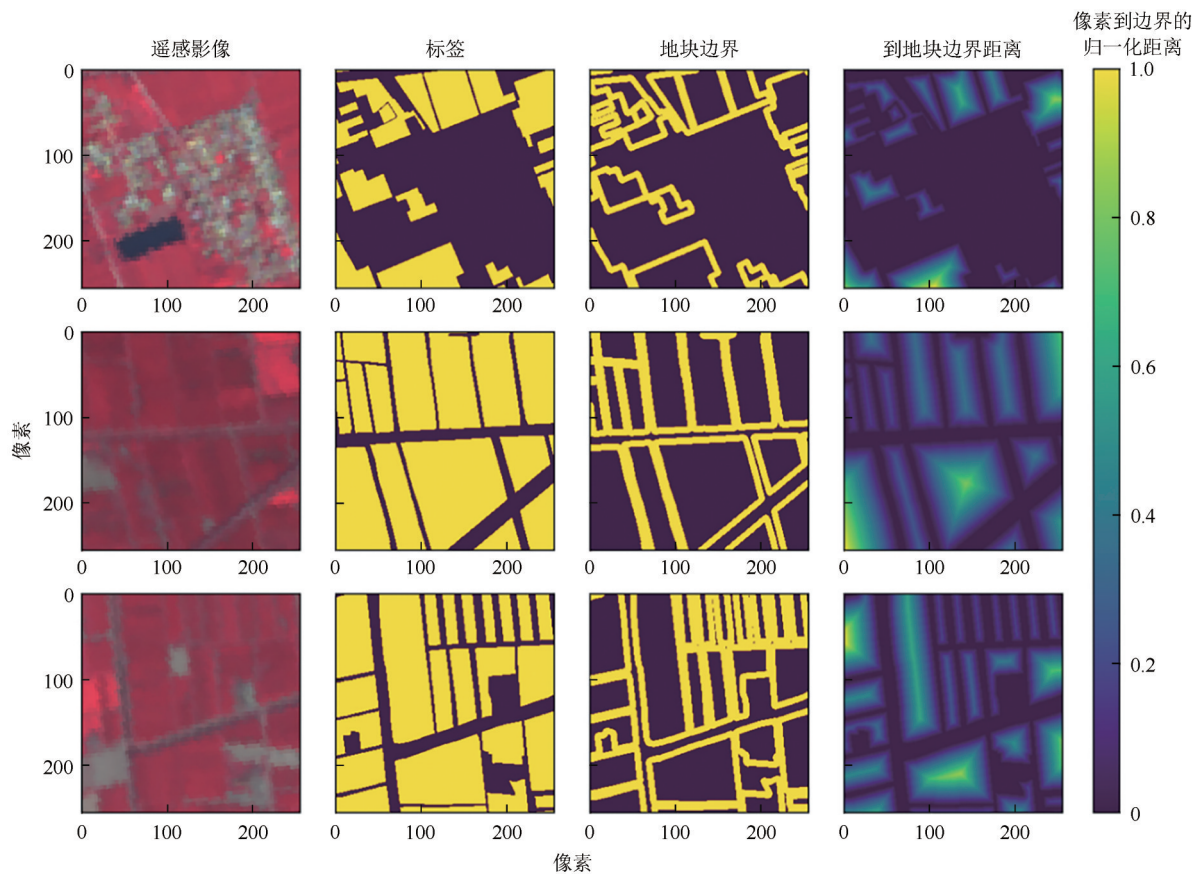


图4 输入样本多任务计算示例图

Fig. 4 Example of input samples with multi-task label

4.6 模型训练与精度评价

Field-Net 模型使用 Keras2.3 框架构建，其核心库为 Tensorflow 2.1。模型在配置有英伟达 Tesla V100 PCIe 显卡（显存 32 GB）的服务器上运行。在

模型的训练过程中运用了常见的放大、缩小、翻转、旋转、加高斯/随机噪声以及对比度变换等数据增广方法，每个影像块进行上述操作的概率为 25%。同时输入和测试数据进行了最大最小值归一化操作。

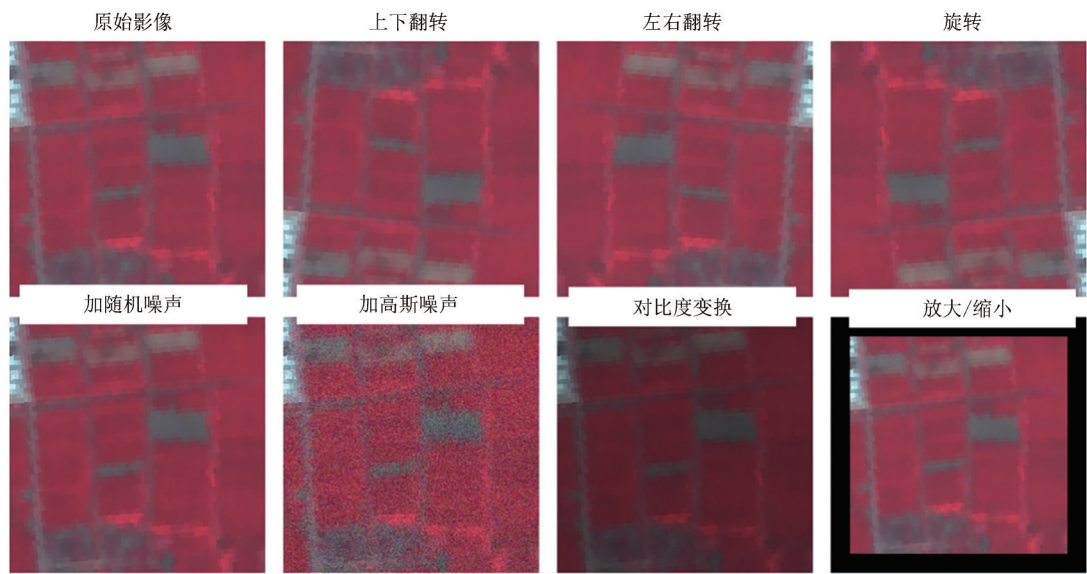


图5 数据增强示例

Fig. 5 Example of data augmentation images

在模型的训练过程中使用 Adam 优化算法进行训练, 该算法可以在每次迭代训练中动态调整学习率, 初始的学习率设置为 1×10^{-4} 。由于显存的限制, 每次训练输入 10 张影像块 (Batch size=10), 最大的迭代次数为 200 次。

为了评估地块分割模型的性能, 本文采取 4 个常用的指标进行精度评价: F_1 -Score, 交并比 (IOU) 和平均交并比 (Mean IOU) 以及总体精度 (OA) (式(8)–(11)), 同时参考常用的精度 (P) 以及召回率 (R)。

$$F_1\text{-Score} = \frac{2TP}{2TP + FP + FN} \quad (8)$$

$$IOU = \frac{TP}{TP + FP + FN} \quad (9)$$

$$\text{Mean IOU} = \frac{1}{n} \sum_{k=1}^n IOU_k \quad (10)$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

这里 TP, TN, FP, FN 分别代表正—正, 正—负, 负—正以及负—负样本的数量。

本文选择 5 个在语义分割任务中取得过最高精度的网络模型验证 Field-Net 模型的分类性能, 分别是: DeepLabv3+、HRNet、LinkNet、D-LinkNet 以及 Psi-Net。

DeepLabv3+ 是一个经典的分割架构, 它在编码器—解码器网络中引入了一个空洞空间卷积池化金字塔 (ASPP) 模块, 该模型可以在不损失图像分辨率的情况下获得更大的感受野。在 PASCAL VOC 挑战赛中取得了最高的精度, 并被用于建筑物识别 (Li 和 Dong, 2022) 和土地覆盖制图 (Liu 等, 2021) 等任务中。

HRNet 是一个并行的卷积结构, 在整个网络中保持高分辨率的特征 (Sun 等, 2020)。它在几个开放的数据集 (如 scapes 和 COCO) 的语义分割任务中取得了最高的精度 (Liu 等, 2021)。

LinkNet 是一种高效的分割神经网络, 它具有跳接、残差块和编码器—解码器体系结构的优势。可以在有限的参数下实现了高效、准确的语义分割 (Zhou 等, 2018)。D-LinkNet 采用带有预编码器的 Linknet 作为主干网络, 同样采用编码器—解码器结构, 其中心部分采用膨胀卷积并最终用于道路提取任务, 在 CVPR DeepGlobe 2018 道路提取挑战中取得了最先进的性能 (Zhou 等, 2018)。

选择 Psi-Net 模型作为多任务学习的模型对比。Psi-Net 一种新颖的由单个编码器和 3 个平行

解码器的卷积神经网络模型, 因模型形状类似 Ψ , 因此被命名为 Psi-Net。3 个平行解码器中, 一个解码器学习掩膜, 另外两个分别学习轮廓检测和距离图估计的辅助任务 (Murugesan 等, 2019)。

5 结 果

5.1 训练集精度测试

首先, 对多任务学习中的掩膜、边界、像素到边界的距离三项任务损失权重按照梯度试验的方法进行分析。在配备有英伟达 Tesla V100 PCIe 显卡的服务器上对模型进行训练, 最大迭代次数为 200, 每次迭代平均耗时 58 s, 每次训练平均耗时约 3.2 h, 结果如图 6 所示。在 3 个损失函数权重参数的梯度对比试验中, 当 3 个权重设置成相等, $\lambda_1=\lambda_2=\lambda_3=1$ 时, 模型取得了较好的效果, 总体的 F_1 -Score 达到了 93.1%; 对于边界预测任务权重从 1 增加到 4 时, 模型的精度呈现先下降后上升的趋势; 对像素到地块边界的任务损失权重从 1 增加到 4 时, 模型的精度呈现逐渐下降趋势; 对于地块掩膜分割任务的权重从 1 增加到 4 时, 模型的精度同样呈现先减少后增加的趋势, 当掩膜分割权重增加到 $\lambda_1=4$, $\lambda_2=\lambda_3=1$ 时, 耕地地块提取的 F_1 -Score 精度最高, 达到了 93.12%。

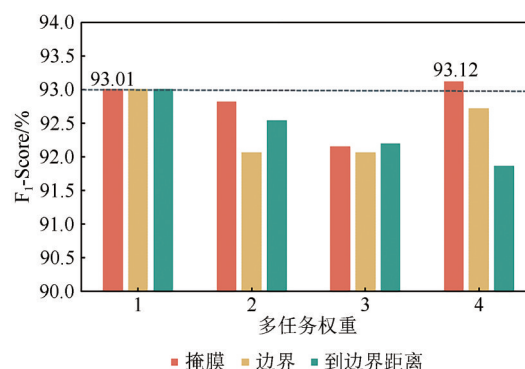


图 6 多任务权重与精度之间的关系

Fig. 6 The relationship between multitasking weights and F_1 -Score

模型训练次数与总体的损失函数之间的关系如图 7 所示。测试集的总体损失在 160 次迭代之后基本趋于稳定, 最小的损失出现在第 186 次迭代, 最小总体损失为 0.60, 地块掩膜任务损失为 0.08, 地块边界预测损失为 0.22, 像素到边界的距离损失为 0.06。在整个测试集上, Field-Net 的总体精度为 92.23%, 耕地地块的精度为 90.97%, 召回率为 95.37% (表 1)。

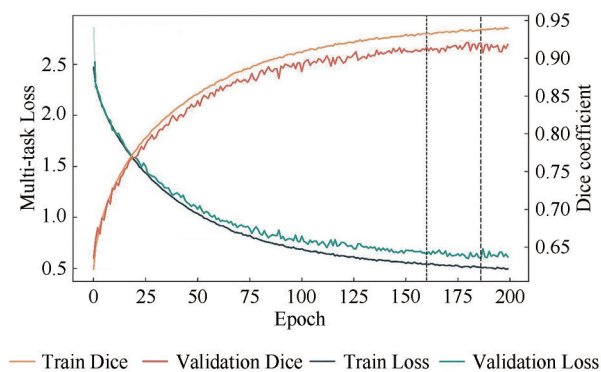


图7 模型训练迭代次数与总体损失之间的关系
Fig. 7 Relationship between the number of model training iterations and the overall loss

将预测结果进行了可视化分析，结果如图8所示，分别展示出原始遥感影像、标签、掩膜、边

界、距离等任务的预测结果，最后将原始影像与去掉边界的耕地地块进行叠加显示，有助于地块的分割处理，实现逐地块的分割，解决地块的“粘连”问题。结果的交并比处于78.03%—89.81%之间，说明Field-Net模型能够较好的实现耕地地块的分割。

表1 Field-Net在测试集的混淆矩阵

Table 1 Confusion matrix for Field-Net output in the test set

		预测结果		召回率/%
		其他	耕地地块	
标签	其他	6920800	909136	88.40
	耕地地块	444763	9157877	95.37
类别精度/%		93.96	90.97	—
总体精度/%		—	—	92.23

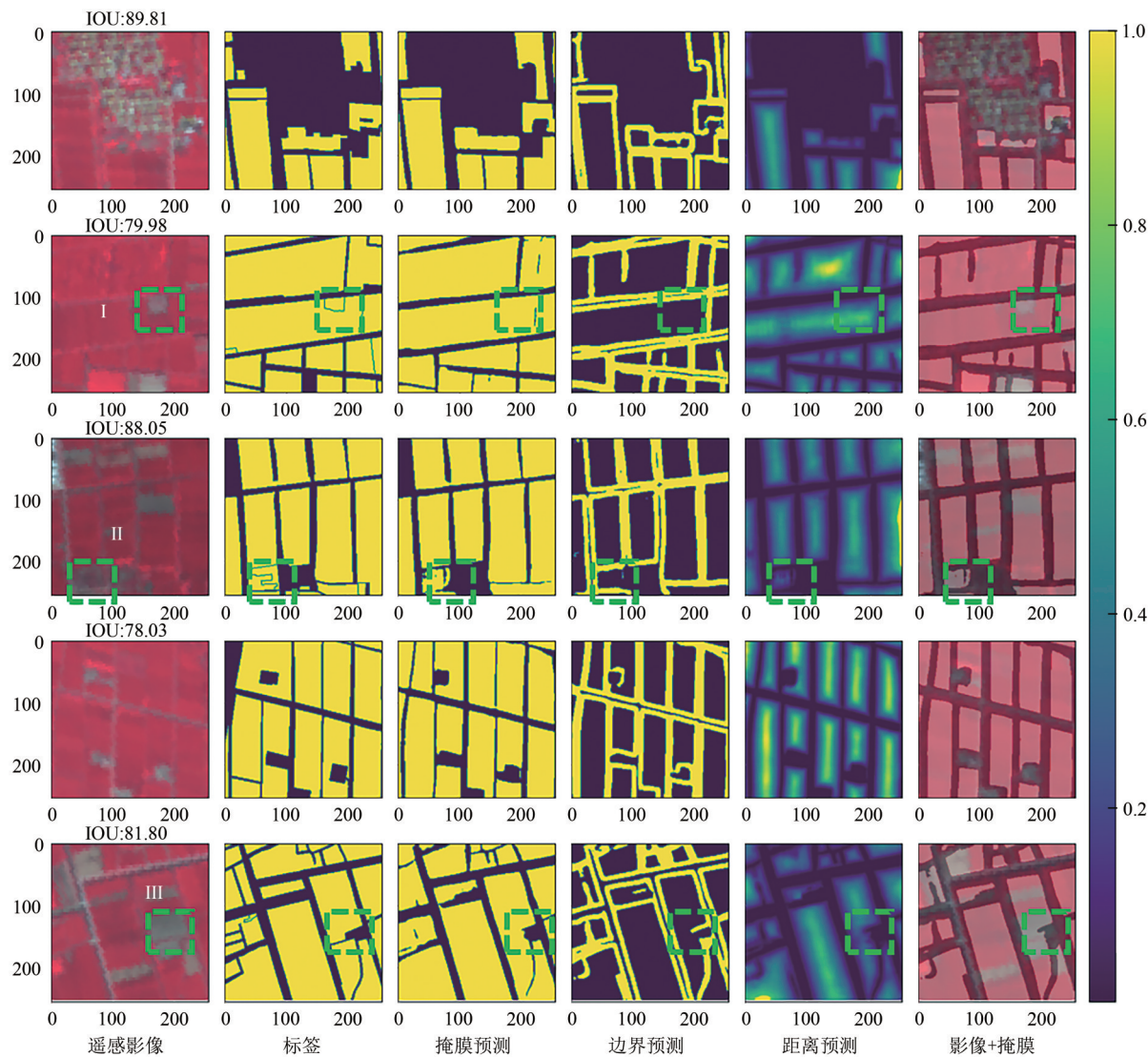


图8 Field-Net预测结果分析
Fig. 8 Display of Field-Net results map

5.2 模型精度对比

本文将 Field-Net 模型与 5 个高性能的深度学习模型进行了对比，其结果如表 2 所示。可以看到，增加了多任务学习与空间注意力机制的 Field-Net 模型较现有模型有不同程度的精度提高。在所选择的其他 4 个模型中 Link-Net 的精度最高，F₁-Score 达到了 92.93%，其次为 HR-Net 和

D-LinkNet。Deeplab v3+在耕地地块分割模型中的表现最差，较 Field-Net 模型低 4.51%。就耕地地块交并比而言，Field-Net 较 4 个模型中精度最高的 Link-Net 模型高出 0.26%，较 DeepLab v3+高出 7.59%。就参数量而言，Psi-Net 模型的参数量最小，其次为 HR-Net，Field-Net 模型的参数量也仅为 1105 万个参数左右。

表 2 Field-Net 与现有模型精度、参量、效率对比

Table 2 Accuracy, parameter, and efficiency comparison between Filed-Net existing models						
模型名称	F ₁ -Score/%	Mean IOU/%	IOU/%	OA/%	参数量/m	预测用时/min
Field-Net	93.07	85.40	87.05	92.23	11.5	20:23
HR-Net	91.91	83.19	85.03	90.95	9.5	23:04
D-LinkNet	91.29	81.95	83.98	90.22	32.8	16:28
LinkNet	92.93	85.26	86.79	92.14	28.3	29:58
DeepLab v3+	88.56	77.10	79.46	87.23	11.9	16:01
Psi-Net	87.71	75.60	78.11	86.27	7.85	6:01

注:加粗数字为最优结果;“—”为次优结果。

将各个模型的预测结果进行可视化，结果如图 9 所示。从图 9 中可以看出，DeepLab v3+未能区分不同的地块，表现出的性能不佳。其余的地块分割模型取得了较好的效果。HR-Net 未能较好的识别非耕地类型，从空间分布上看分割效果稍逊

于 D-LinkNet 模型与 LinkNet 模型。Link-Net 模型的效果与 Field-Net 效果相近，但是由于 Field-Net 增加了多任务学习的策略，Field-Net 获得的地块分割结果较 Link-Net 能更好区分相邻地块，从一定程度上解决了地块之间的“粘连”问题。

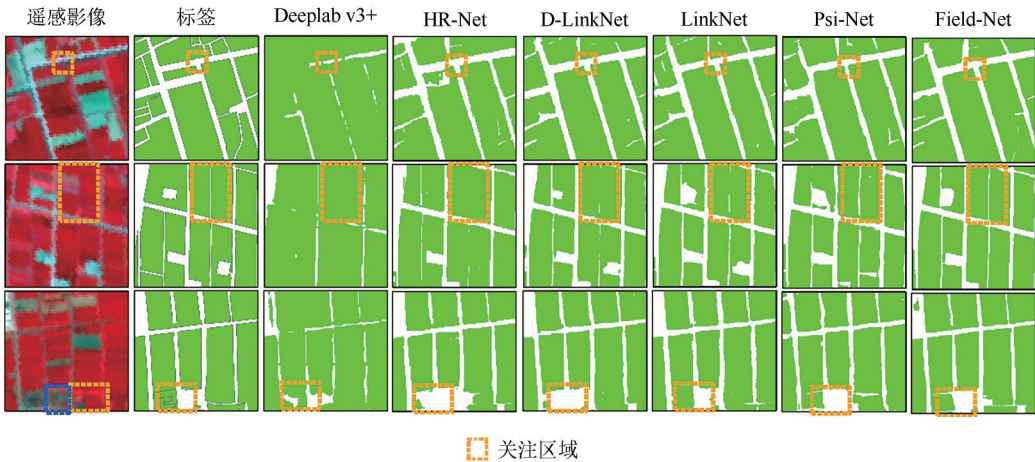


图 9 不同模型预测结果对比图

Fig. 9 Comparison between Filed-Net prediction results and other models

5.3 模型空间泛化性能分析

模型在训练和验证区域取得较高的精度，但是未必能在其他区域具有良好的泛化性能，这也是限制地块分割模型空间应用的重要难题。为了验证模型的空间可迁移性，将 Field-Net 模型与选择的 5 个高性能深度学习模型在模型训练之外

的区域，选择了 3 处空间泛化性能测试区，分析不同深度学习模型的空间可迁移性。

Field-Net 模型与其他的模型的精度对比如表 3 所示。在 3 个空间泛化性能测试区，Field-Net 模型的 F₁-Score 达到了 94.98%，平均交并比为 71.55%，较在测试集精度较高的 Link-Net 模型高出 3.51%；

地块的交并比为 80.45%，较 Link-Net 模型高出 2.72%。在泛化性能测试区，Psi-Net 的地块交并比在所比较的 5 个高性能模型中最高，仅比 Field-Net 模型低 1.36%。

表 3 各模型空间泛化测试精度对比
Table 3 Accuracy assessment for DL models on spatial generality

模型名称	F_1 -Score	Mean IOU	IOU	OA
Field-Net	94.98	71.55	90.45	91.36
HR-Net	93.61	67.70	87.99	89.17
D-LinkNet	92.39	65.98	85.85	87.38
LinkNet	93.46	<u>68.04</u>	87.73	88.99
DeepLab v3+	<u>93.83</u>	48.55	88.38	88.51
Psi-Net	94.23	60.26	<u>89.09</u>	<u>89.61</u>

注:加粗数字为最优结果;“—”为次优结果。

从各个模型地块分割的可视化结果图(图 10)中可以看出 Deeplab v3+模型未能区分不同地块的边界，

存在明显的地块粘连问题。其余模型存在一定的地块漏分现象，Field-Net 模型与 Psi-Net 模型的空间一致性较高，但是 Psi-Net 的地块漏分现象更为明显。

5.4 模型应用

基于本文构建的基于空间注意力机制与多任务学习的卷积神经网络，利用 GF-1D 与 ZY-3 卫星数据，通过处理分块 (256×256) 的方式完成 2021 年山东省利津县的耕地地块制图，为了准确获取影像块边界处的地块分布，在横向、纵向分别设置 128 像素的重叠区域，重叠区域的掩膜和边界概率以及到边界的距离取最大值处理，最终预测耗时 0.5 h，结果如图 11 所示。从耕地地块的空间分布可以看出，模型可以实现耕地地块的良好区分，能将建设用地、水体等类型与耕地区分开来，并去除耕地中的田间道路、防护林等基础设施，实现耕地地块的划分，模型共识别出 35641 个地块，平均地块大小为 13056 m²，约合 1.31 ha。

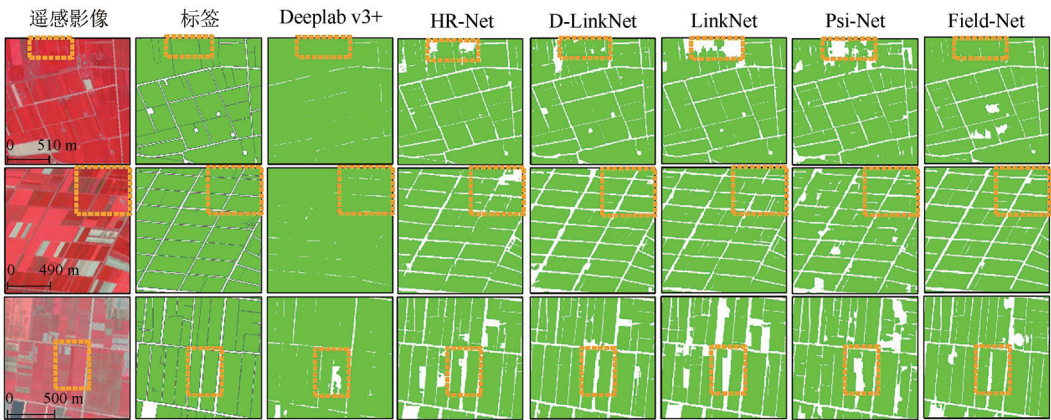


图 10 不同模型空间泛化性能结果对比图

Fig. 10 Comparison of spatial generalization performance for different models

6 讨论

本文设计了一种基于空间注意力机制与多任务学习的地块分割模型 Filed-Net，基于 UNet 架构，增加了空间注意力机制，并采用多任务学习的策略，在语义分割的基础上结合边界、像素到地块边界的距离任务。在山东省东营市利津县实现耕地地块的准确识别，交并比达到了 87.05%，总体精度为 92.23%，模型的性能优于 5 个高性能深度学习框架，有效提升了体块识别的性能。

6.1 空间注意力机制与多任务学习消融分析

Field-Net 模型的核心模块是空间注意力机制

和多任务学习策略。为了验证 Field-Net 模型中的空间注意力机制与多任务学习的效果，本文设置了 2 组对照试验。首先，将 Field-Net 与 ResUNet 模型进行对比分析，分析空间注意力机制模块的性能；其次分别在 Field-Net 模型与 ResUNet 模型中采取多任务学习和单任务学习的策略，分析多任务学习在地块分割任务中带来的精度提升。

在 Field-Net 模型的结构中，与 ResUNet 的主要区别就是增加了空间注意力机制。通过两个模型的对比可以发现，仅仅使用单一的耕地地块掩膜作为训练样本，使用了空间注意力机制的 Fied-Net 相比 ResUNet 模型 F_1 -Score 提高了 1.01%，交并

比提高了 1.6%；当同时对 ResUNet 模型和 Field-Net 模型使用多任务学习的策略时，使用空间注意力机制使 Field-Net 模型的 F_1 -Score 提高了 1.19%，交并比提高了 0.58%。

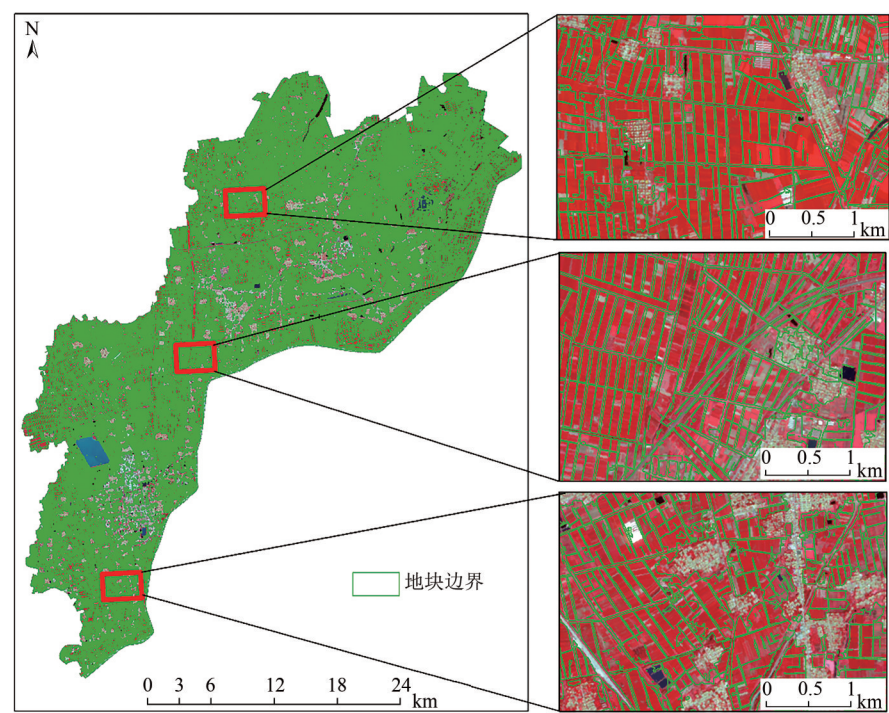


图 11 研究区耕地地块分割结果图
Fig. 11 The results of agricultural field in the study area

多任务学习可以通过学习相关任务的表征特征，从而起到提高原来任务的作用。简而言之，就是设置与主任务相关的其他任务。在本文设置了地块边界和到地块边界的距离这两个“副线”任务，使得模型聚焦任务中通用的特征，提高模型的泛化性能。通过将 Field-Net 与 ResUNet 模型分别进行单任务和多任务学习发现，多任务学习策略使得 Field-Net 模型的 F_1 -Score 提高了 0.18%，交并比提高了 0.21%；使得 ResUNet 模型 F_1 -Score 提高了 0.66%，交并比提高了 0.65%。进一步分析边界和像素到地块边界的距离两个任务对于交并比的提升，发现如果使用掩膜+边界任务进行训练，交并比提升了 0.61%，使用掩膜+到边界的距离任务，交并比仅仅提高了 0.1%。虽然像素到地块边界距离任务对于 IOU 的提升较为有限，但是通过像素到边界距离的回归预测，可以实现相邻地块的区分，将个例分割任务 (Instance Segmentation) 的任务，转化为语义分割 (Semantic Segmentation) 任务，降低逐个地块识别的计算量。

在多任务学习的过程中，不同任务的损失权重设置至关重要。通过本文的权重对比试验发现，对于多任务学习的权重设置应当以掩膜分割任务

为主，其他任务作为辅助任务，这样对与主要的分割任务精度最高。如果将其他的辅助任务设置的权重过高，则会导致主要任务的精度显著下降，如当把像素到边界的距离的任务损失权重设置为 4 时，在整个测试中出现了最低的精度。

表 4 空间注意力机制与多任务学习消融试验
Table 4 Spatial attention mechanisms and multi-task learning ablation experiments

模型	F_1 -Score	IOU	Mean IOU	/%
				OA
ResUNet	91.88	85.08	84.18	91.67
ResUNet+多任务学习	92.54	86.12	84.83	91.91
ResUNet+空间注意力机制	92.89	86.68	85.19	92.10
Field-Net(ResUNet+注意力机制+多任务学习)	93.07	87.05	85.40	92.23

6.2 特征可视化

在模型得精度对比过程中，Field-Net 取得了较高得精度。通过对 Field-Net 模型在输出 3 个任务之前的特征进行可视化，同时输入了 DeepLab v3+ 模型在空间升尺度之前 256 个特征中的 31 个，结果如图 12 所示。

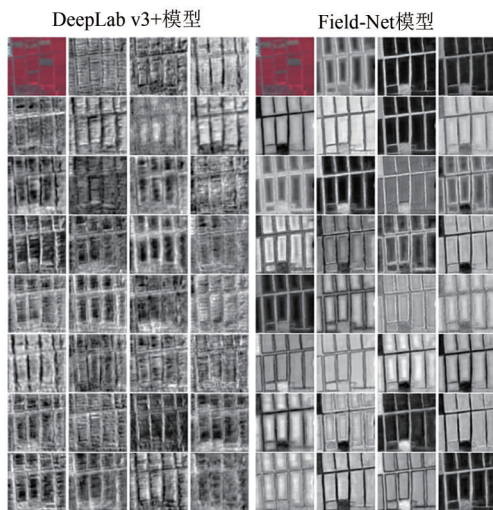


图 12 DeepLab v3+模型与Field-Net模型特征对比
Fig. 12 Comparison of DeepLab v3+ model and Field-Net model features

从两个模型的特征图中可以看出, 在增加空间注意力机制、采取多任务学习策略之后, Field-Net最后的特征部分聚焦于地块边界, 部分聚焦于地块内部 (图 12), 而 DeepLab v3+模型的特征则较为分散, 没有能够实现地块边界特征的准确提取。从侧面说明增加了边界和像素到边界的距离这两个任务之后, 模型能够学习到更加具有代表性的特征, 从而提高模型的精度, 更重要的是提高模型的泛化性能 (Xu 等, 2023)。

6.3 展望

地块是土地资源精准监测的基础, 也是进行高标准农田建设和粮食植补的对象, 因此对其准确识别对摸清土地资源和耕地利用现状具有重要意义。本文提出的 Field-Net 模型可以实现华北平原破碎农田生态系统的地块分割任务, 并且精度高于现有的模型。由于样本有限, 模型在其他农田生态系统景观中的适用性尚有待评价。随着深度学习“大模型”时代的到来, 针对地块分割任务, 同样需要从模型和数据集两个角度, 构建可以分割任何一个地块的分割模型 (Segment every field)。在中国的地块分布数据集的生成过程中, 复杂破碎的农田生态系统给这一任务带来了不小的挑战, 将来可以通过借鉴 Image-Net 的模式, 积累不同区域的地块分割数据集, 解决模型训练缺少样本的问题, 同时后续应该构造对通道、区域、传感器更加通用的模型。

在连片的地块分割识别方面, 由于地块间区别并不明显, 因此还存在一定的地块粘连问题, 如图 8 中标记的 I、II、III 等 3 处。导致这种差异的主要原因是本文主要针对的是耕地地块识别任务, 与农业种植地块存在一定的差异, 未来可以采用多时相、更高分辨率的遥感影像增强地块之间的特征, 实现种植地块的准确识别。

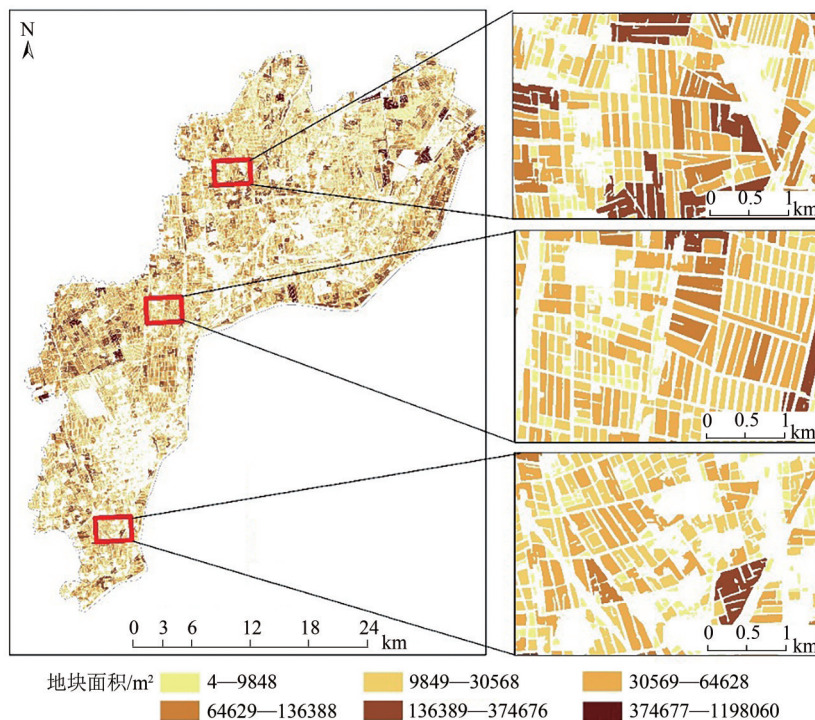


图 13 利津县耕地地块大小分布图
Fig. 13 Field size map of arable land in Lijin County

7 结 论

耕地地块的准确提取对于国土资源监测和粮食生产以及监测具有重要的意义。但是现有模型在地块破碎区域的适用性不明,存在模型的泛化性能不强的问题。本文提出了一种基于空间注意力机制与多任务学习的地块分割模型 Field-Net 模型,并在山东省东营市利津县进行应用。通过优选多任务学习的损失函数权重,模型在测试集的 F_1 -Socre 达到了 93.07%,耕地地块的交并比为 87.05%,精度较 Link-Net 模型高出 0.26%,DeepLab v3+ 高出 7.59%;在空间泛化性能测试中,Field-Net 模型的平均交并比为 71.55%,较 Link-Net 模型高出 3.51%,空间泛化性能明显提升。最后,本研究通过消融试验和特征可视化分析发现,空间注意力机制模块和多任务学习策略可以使模型学习的特征更加聚集于地块边界和地块内部具有代表性的特征,从而提升其地块识别性能。总体而言,Field-Net 模型可以支撑地块级别国土资源与耕地非农化、非粮化利用监测,提高监测的效率和时效性。

参考文献(References)

- Cheng R, Wei Y B, Lu M and Wu W B. 2022. Cropland field extraction based on ensemble deep learning model. *Chinese Journal of Agricultural Resources and Regional Planning*, 43(7): 273-281 (程锐, 魏妍冰, 陆苗, 吴文斌. 2022. 基于集成深度学习模型的耕地地块提取. *中国农业资源与区划*, 43(7): 273-281) [DOI: 10.7621/cjarrp.1005-9121.20220727]
- Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhme A, Böhme A, Deubner J, Jäckel Z, Seiwald K, Dovzhenko A, Tietz O, Dal Bosco C, Walsh S, Saltukoglu D, Tay T L, Prinz M, Palme K, Simons M, Diester I, Brox T and Ronneberger O. 2019. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1): 67-70 [DOI: 10.1038/s41592-018-0261-2]
- Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W and Lu H Q. 2019. Dual attention network for scene segmentation//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE: 3141-3149 [DOI: 10.1109/CVPR.2019.00326]
- Han Y X and Meng J H. 2019. A review of per-field crop classification using remote sensing. *Remote Sensing for Natural Resources*, 31(2): 1-9 (韩衍欣, 蒙继华. 2019. 面向地块的农作物遥感分类研究进展. *国土资源遥感*, 31(2): 1-9) [DOI: 10.6046/gtzyyg.2019.02.01]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Identity mappings in deep residual networks//*Proceedings of the 14th European conference on Computer Vision*. Amsterdam: Springer: 630-645 [DOI: 10.1007/978-3-319-46493-0_38]
- Jeppesen J H, Jacobsen R H, Inceoglu F and Toftegaard T S. 2019. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, 229: 247-259 [DOI: 10.1016/j.rse.2019.03.039]
- Li B. 2020. Study on the application of remote sensing technology in land resources survey--the third national land survey as an example. *Metallurgy and Materials*, 40(5): 51, 53 (李博. 2020. 遥感技术在国土资源调查中的应用研究——以第三次全国土地调查为例. *冶金与材料*, 40(5): 51, 53) [DOI: 10.3969/j.issn.1674-5183.2020.05.027]
- Li Z C and Dong J W. 2022. A framework integrating DeeplabV3+, transfer learning, active learning, and incremental learning for mapping building footprints. *Remote Sensing*, 14(19): 4738 [DOI: 10.3390/rs14194738]
- Lijin County Government. 2021. Natural Resources. [2023-05-13]. <http://www.lijin.gov.cn/col/col70883/index.html> (利津县政府. 2021. 自然资源. [2023-05-13]. <http://www.lijin.gov.cn/col/col70883/index.html>)
- Lin M P, Li J and Shen H F. 2023. PolSAR image deep learning super-resolution model based on multi-scale attention mechanism. *National Remote Sensing Bulletin*. [DOI: 10.11834/jrs.20233002/] (林缪鹏, 李杰, 沈焕锋. 2023. 基于多尺度注意力机制的 PolSAR 深度学习超分辨率模型. *遥感学报*. [DOI: 10.11834/jrs.20233002/])
- Liu M, Fu B L, Fan D L, Zuo P P, Xie S Y, He H C, Liu L L, Huang L K, Gao E T and Zhao M. 2021. Study on transfer learning ability for classifying marsh vegetation with multi-sensor images using DeepLabV3+ and HRNet deep learning algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 103: 102531 [DOI: 10.1016/j.jag.2021.102531]
- Murugesan B, Sarveswaran K, Shankaranarayana S M, Ram K, Joseph J and Sivaprakasam M. 2019. Psi-Net: shape and boundary aware joint multi-task deep network for medical image segmentation//*Proceedings of 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin: IEEE: 7223-7226 [DOI: 10.1109/EMBC.2019.8857339]
- Persello C, Tolpekin V A, Bergado J R and de By R A. 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, 231: 111253 [DOI: 10.1016/j.rse.2019.111253]
- Ronneberger O, Fischer P, Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//*Proceedings of 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Saraiva M, Protas É, Salgado M and Souza C. 2020. Automatic mapping of center pivot irrigation systems from satellite images using

- deep learning. *Remote Sensing*, 12(3): 558 [DOI: 10.3390/rs12030558]
- Song Q, Hu Q, Lu M, Yu Q Y, Yang P, Shi Y, Duan Y L and Wu W B. 2020. Prospect of crop mapping. *Chinese Journal of Agricultural Resources and Regional Planning*, 41(6): 57-65 (宋茜, 胡琼, 陆苗, 余强毅, 杨鹏, 史云, 段玉林, 吴文斌. 2020. 农作物空间分布遥感制图发展方向探讨. *中国农业资源与区划*, 41(6): 57-65) [DOI: 10.7621/cjarrp.1005-9121.20200607]
- Sun J, Darbehani F, Zaidi M and Wang B. 2020. SAUNet: shape attentive u-net for interpretable medical image segmentation//*Proceedings of 23rd International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2020*. Lima: Springer: 797-806 [DOI: 10.1007/978-3-030-59719-1_77]
- Takikawa T, Acuna D, Jampani V and Fidler S. 2019. Gated-SCNN: gated shape CNNs for semantic segmentation//*Proceedings of IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE: 5228-5237 [DOI: 10.1109/ICCV.2019.00533]
- Tian F Y, Wu B F, Zeng H W, Zhang M, Hu Y R, Xie Y, Wen C C, Wang Z D, Qin X L, Han W and Yang H H. 2023. A shape-attention Pivot-Net for identifying central pivot irrigation systems from satellite images using a cloud computing platform: an application in the contiguous US. *GIScience and Remote Sensing*, 60(1): 2165256 [DOI: 10.1080/15481603.2023.2165256]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc.
- Waldner F and Diakogiannis F I. 2020. Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network. *Remote Sensing of Environment*, 245: 111741 [DOI: 10.1016/j.rse.2020.111741]
- Wang F, M. Jiang M Q, Qian C, Yang S, Li C, Zhang H G, Wang X G and Tang X O. 2017. Residual attention network for image classification//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 6450-6458 [DOI: 10.1109/CVPR.2017.683]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Wu B F, Zhang M, Zeng H W, Tian F Y, Potgieter A B, Qin X L, Yan N N, Chang S, Zhao Y, Dong Q H, Boken V, Plotnikov D, Guo H D, Wu F M, Zhao H, Deronde B, Tits L and Loupian ES. 2023. Challenges and opportunities in remote sensing-based crop monitoring: a review. *National Science Review*, 10(4): nwac290 [DOI: 10.1093/nsr/nwac290]
- Xu L L, Yang P, Yu J J, Peng F, Xu J, Song S R and Wu Y X. 2023. Extraction of cropland field parcels with high resolution remote sensing using multi-task learning. *European Journal of Remote Sensing*, 56(1): 2181874 [DOI: 10.1080/22797254.2023.2181874]
- Yang Y P, Wu Z F, Luo J C, Huang Q T, Zhang D Y, Wu T J, Sun Y W, Cao Z, Dong W and Liu W. 2021. Parcel-based crop distribution extraction using the spatiotemporal collaboration of remote sensing data. *Transactions of the Chinese Society of Agricultural Engineering*, 37(7): 166-174 (杨颖频, 吴志峰, 骆剑承, 黄启厅, 张冬韵, 吴田军, 孙营伟, 曹峥, 董文, 刘巍. 2021. 时空协同的地块尺度作物分布遥感提取. *农业工程学报*, 37(7): 166-174) [DOI: 10.11975/j.issn.1002-6819.2021.07.020]
- Yi Y N, Zhang Z J, Zhang W C, Zhang C R, Li W D and Zhao T. 2019. Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network. *Remote Sensing*, 11(15): 1774 [DOI: 10.3390/rs11151774]
- Zhang H T, Gao H F and Ren C. 2022. Deep segmentation and extraction of cultivated land in hilly areas based on improved Unet++. *Spacecraft Recovery Remote Sensing*, 43(4): 36-45 (张海天, 高懋芳, 任超. 2022. 基于改进Unet++的丘陵地区耕地地块深度分割与提取. *航天返回与遥感*, 43(4): 36-45) [DOI: 10.3969/j.issn.1009-8518.2022.04.004]
- Zhang Q K, Meng J H and Ren C. 2022. Crop classification based on two-dimensional representation and CNN model from remote sensing. *National Remote Sensing Bulletin*, 26(7): 1437-1449 (张乾坤, 蒙继华, 任超. 2022. 构建地块二维表征及CNN模型的作物遥感分类. *遥感学报*, 26(7): 1437-1449) [DOI: 10.11834/jrs.20219432]
- Zhang Z X, Liu Q J and Wang Y H. 2018. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5): 749-753 [DOI: 10.1109/LGRS.2018.2802944]
- Zhou L C, Zhang C and Wu M. 2018. D-LinkNet: LinkNet with pre-trained encoder and dilated convolution for high resolution satellite imagery road extraction//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City: IEEE: 192-1924 [DOI: 10.1109/CVPRW.2018.00034]
- Zhou N, Yang P, Wei C S, Shen Z F, Yu J J, Ma X Y and Luo J C. 2021. Accurate extraction method for cropland in mountainous areas based on field parcel. *Transactions of the Chinese Society of Agricultural Engineering*, 37(19): 260-266 (周楠, 杨鹏, 魏春山, 沈占锋, 余娟娟, 马晓宇, 骆剑承. 2021. 地块尺度的山区耕地精准提取方法. *农业工程学报*, 37(19): 260-266) [DOI: 10.11975/j.issn.1002-6819.2021.19.030]
- Zhu Y, Pan Y Z and Zhang D J. 2022. An agriculture parcel identification method based on convolutional neural network and watershed 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. *地球信息科学学报*, 24(12): 2389-2403) [DOI: 10.12082/dqxxkx.2022.220447]

Agricultural field segmentation using spatial attention mechanism and multi-task learning strategy

TIAN Fuyou¹, CAO Yupei^{2,3}, ZHAO Hang^{1,4}, WU Bingfang^{1,4}, ZENG Hongwei^{1,4},
LIU Yazhou³, QIN Xingli¹, ZHANG Miao¹, ZHU Liang¹, ZHU Weiwei¹

1.State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute,
Chinese Academy of Science, Beijing 100094, China;

2.Beijing Normal University, Beijing 100101, China;

3.Shandong Land Group Digital Technology Co., Ltd., Zaozhuang 277000, China;

4.University of Chinese Academy of Science, Beijing 100049, China

Abstract: The accurate identification of agricultural fields, which are the smallest units of agricultural farming, is crucial for monitoring land resources and arable land utilization. Manual mapping methods are time-consuming and expensive, and they are incapable of real-time or near-real-time updates. To enhance agricultural field delineation efficiency, we introduce Field-Net, a field segmentation model that leverages a spatial attention mechanism and multitask learning in this research.

Field-Net is based on the UNet architecture. It combines a spatial attention mechanism and a multitask learning approach. In addition to the segmentation task, we incorporate boundary identification and distance to the field boundary as two additional tasks, enabling the model to learn representative features related to fields. The model's performance was evaluated using Gaofen-1 and Ziyuan-3 satellite images with a spatial resolution of 2 m in Lijin County, Dongying City, Shandong Province. We labeled 3,480 tiles that measured 256×256 pixels in YanWo District, with 3000 used for training, 360 for validation, and 120 for the spatial generalization performance test.

We initially analyzed loss weight for the three tasks, i.e., mask, boundary, and pixel-to-boundary distance, in multitask learning by using a gradient test. For multitask learning, loss weight should prioritize the mask segmentation task as the primary task, while other tasks should be considered secondary. Across the entire test set, Field-Net achieved an overall accuracy of 92.23% and an Intersection Over Union (IOU) of 87.05%. We compared Field-Net with four state-of-the-art architectures: DeepLab v3+, HRNet, LinkNet, and D-LinkNet. Field-Net outperformed all of them in semantic segmentation tasks, with an IOU that was 0.26% higher than LinkNet, the most accurate among the four compared models, and 7.59% higher than that of DeepLab v3+. In the spatial generalization performance test, the average IOU of the Field-Net model was 3.51% higher than that of the Link-Net model, and spatial generalization performance was significantly improved. Ablation tests demonstrated that the spatial attention mechanism and multitask learning strategy improved the F1 score by 1.01% and IOU by 1.6% compared with the ResUNet model. The multitask learning strategy led to an improvement of 0.18% in F1 score for Field-Net and an improvement of 0.21% in IOU.

Although challenges remain in identifying contiguous fields due to unclear boundaries, future enhancements can incorporate multitemporal and high-resolution remote sensing images to improve field feature discrimination. Feature visualization analysis revealed that the spatial attention mechanism and multitask learning strategy enabled the model to learn clustered features at field boundaries and within plots, enhancing feature representativeness. Overall, the Field-Net model supports the field-level monitoring of cropland use, including nonagricultural applications, such as grain production, enhancing the efficiency and timeliness of land resource monitoring. In generating the field dataset of China, complex and fragmented croplands pose considerable challenges to this task. In the future, the problem of lack of samples for model training can be solved by accumulating field segmentation datasets from different regions by borrowing the paradigm of Image-Net, while a general model for channels, regions, and sensors should be constructed subsequently. In the future, with the arrival of the “large model” era of deep learning, constructing a model for the task of parcel segmentation is also necessary to segment every field from the perspective of model and dataset.

Key words: field segmentation, Field-Net model, spatial attention mechanism, multi-task learning, GF satellite data

Supported by National Key Research and Development Program of China (No. 2019YFE0126900); The Third Xinjiang Scientific Expedition (No. 2022xjkk030106); National Natural Science Foundation of China (No. 41861144019); Agricultural Remote Sensing Innovation Team Project of AIRCAS (No. E33D0201-6)