

张剑飞, 倪俊文. 基于改进 PointRend 的遥感图像分割边界优化[J]. 智能计算机与应用, 2025, 15(12): 9-16. DOI: 10.20169/j. issn. 2095-2163. 24052401

## 基于改进 PointRend 的遥感图像分割边界优化

张剑飞, 倪俊文

(黑龙江科技大学 计算机与信息工程学院, 哈尔滨 150022)

**摘要:** 遥感可以在远距离、不与目标接触的情况下通过传感器对地面目标物体的电磁波信息进行采集处理并成像, 语义分割是遥感智能解译的一个重要内容, 在城市规划、灾害评估及农业生产等领域具有十分重要的作用。基于深度学习的方法具备良好的鲁棒性和泛化性, 已成为遥感图像语义分割领域的主流方法, 但常用的分割方法在分割边界处识别精度不高。针对这种问题, 本文提出了基于改进 PointRend(点渲染方法)的遥感图像边界优化方法。首先, 引入 PointRend 对不准确的图像分割边界进行优化。然后, 通过融合不同阶段的特征来增强 PointRend 中细粒度特征的语义信息和上下文信息。最后, 采用连续空间特征表示方法来增强 PointRend 中细粒度特征的空间位置信息。实验结果表明, 改进的方法能够修正和校准图像分割中的边界, 提高图像边界的分割精度。

**关键词:** 遥感图像; 语义分割; 边界优化

中图分类号: TP751

文献标志码: A

文章编号: 2095-2163(2025)12-0009-08

### Segmentation refinement of remote sensing image based on improved PointRend

ZHANG Jianfei, NI Junwen

(School of Computer and Information Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China)

**Abstract:** Remote sensing can collect and process electromagnetic wave information of ground target objects at a distance without contacting them through sensors and generate images. Semantic segmentation is an important content of remote sensing intelligent interpretation. It has very important roles in urban planning, disaster assessment, and agricultural production. The method based on deep learning has good robustness and generalization, and has become the mainstream method in the field of remote sensing image semantic segmentation. However, the commonly used segmentation methods have low recognition accuracy at the boundary of segmentation. To solve this problem, this paper proposes a remote sensing image segmentation refinement method based on the improved PointRend method. First, the inaccurate image segmentation boundary is optimized by introducing PointRend. Then, the fine-grained feature's semantic information and context information are enhanced by fusing features from different stages. Finally, the spatial position information of the fine-grained feature is enhanced by using the continuous spatial feature representation method. The experimental results show that the improved method can correct and calibrate the boundaries of image segmentation and improve the segmentation accuracy of image boundaries.

**Key words:** remote sensing image; semantic segmentation; segmentation refinement

## 0 引言

语义分割是智能遥感影像解译技术的一个重要内容, 其目的是对遥感图像进行全场景的像素级的语义解析, 在许多领域都有广泛的应用。例如, 在城市规划中, 语义分割可以对城市遥感图像进行处理,

得到城市区域的建筑密度、绿地分布、交通状况等信息; 在生态环境领域, 语义分割可以识别出污染源、生态环境变化等<sup>[1]</sup>。

深度学习方法可以提取到具有强力语义表征能力的深层特征, 相应的图像分割方法具备更好的鲁棒性与泛化性, 在遥感图像语义分割领域逐渐成为

**基金项目:** 国家自然科学基金(61803148); 黑龙江省高等学校基本科研业务经费项目(2024-KYYWF-1099)。

**作者简介:** 倪俊文(1995—), 男, 硕士研究生, 主要研究方向: 人工智能与模式识别。

**通信作者:** 张剑飞(1978—), 女, 博士, 教授, 主要研究方向: 人工智能与模式识别。Email: zjfnfu2008@163.com。

收稿日期: 2024-05-24

主流<sup>[2]</sup>。2015年,Long等<sup>[3]</sup>提出的全卷积神经网络(Fully Convolutional Network, FCN)实现了端到端、像素到像素的图像语义分割。后续的大量研究基于FCN的架构进行,如金字塔场景解析网络(Pyramid Scene Parsing Network, PSPNet)<sup>[4]</sup>、U形网络(U-Net)<sup>[5]</sup>、李梓瑜等<sup>[6]</sup>对遥感图像道路提取方法研究等。

基于FCN的图像分割方法主要利用图像的深层特征进行分割工作,深层特征的空间分辨率较低,缺少图像的细节信息,分割的边界存在不精确的问题。针对该问题,研究人员提出了一些图像分割边界优化方法。DeepLab<sup>[7]</sup>中通过全连接的条件随机场(Fully-Connected CRF)建立图像和网络输出之间的联系,有效缓解了边缘定位不准确的问题。点渲染方法(PointRend)<sup>[8]</sup>是一种逐点的优化分割边界的方法,其模拟计算机图形学中的细分渲染算法,会由粗到细地从预测图中找出重难点进行重新预测,可以得到更准确的分割边界。级联金字塔场景解析网络(CascadePSP)<sup>[9]</sup>提出了一种级联式的、与模型无关的图像分割边界优化方案。在该方案中有一个细化模块,细化模块会迭代地进行分割边界优化工作。Chuong Huynh等提出了一种渐进式的、级联的图像分割边界优化框架MagNet<sup>[10]</sup>。该框架会由粗糙到细节地对图像进行分割,并且可以调节分割速度和分割精度的比重。连续细化模型CRM(Continuous Refinement Model)<sup>[11]</sup>通过图像超分辨率领域的局部隐式图像函数LIIF(Local Implicit Image Function)<sup>[12]</sup>来对图像分割结果进行优化。Li等<sup>[13]</sup>将局部分割结果与其上下文语义相结合,减少了分割边界伪影。SegRefiner<sup>[14]</sup>提出了一种使用扩散模型优化图像结果的方法,通过去噪扩散步骤来进行分割边界的优化。

总的来说,目前遥感图像的语义分割存在以下问题:(1)由于计算量的限制,模型会将特征的分辨率降低,造成图像细节信息的丢失,导致图像分割边界不准确;(2)点渲染方法(PointRend)虽然可以找出图像分割边界区域的点进行优化,受限于其输入是感受野小、语义信息少的浅层特征,对分割边界的优化效果有限。

针对这些问题,本文提出一种基于改进PointRend(点渲染方法)的分割边界优化方法。创新点在于:(1)引入PointRend对PSPNet的分割预测中的边界区域进行识别和优化;(2)通过融合不同阶段特征强化PointRend中使用的特征的语义信

息和上下文信息,提升边界优化效果;(3)通过连续空间的特征表示强化空间位置信息,进一步提升PointRend对分割边界的优化效果。

## 1 相关理论基础

### 1.1 金字塔场景解析网络 PSPNet

金字塔场景解析网络PSPNet(Pyramid Scene Parsing Network)是一种端到端的图像语义分割网络,相比于全卷积神经网络FCN(Fully Convolutional Network),PSPNet在主干网络从图像中提取到有高级语义信息的特征后,使用金字塔池化模块PPM(Pyramid Pooling Module)进一步提取了含有多尺度上下文信息的特征,有效提升了语义分割效果。

PSPNet的结构如图1所示,其中的PPM中有4个分支,每个分支包含池化和点卷积(Pointwise Conv)<sup>[15]</sup>。各分支中的池化将特征的空间分辨率分别压缩到 $1\times 1$ 、 $2\times 2$ 、 $3\times 3$ 和 $6\times 6$ (分别对应于图中的POOL模块的S值),而点卷积则将特征在通道上降维到512。空间压缩后的特征最后被双线性插值恢复空间分辨率后在通道维度上连接,生成的特征具有更大的感受野,可以提升语义分割效果,并且因为只使用了无参数的池化和参数量较少的点卷积,所以性能上的开销并不大。

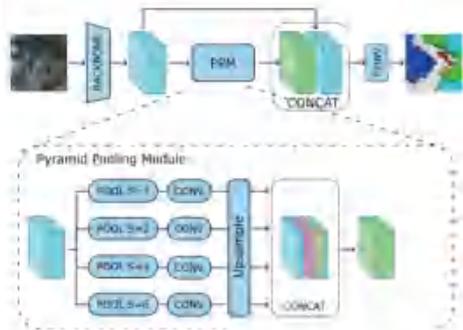


图1 PSPNet模型结构图

Fig. 1 Structure chart of PSPNet

### 1.2 点渲染 PointRend

PointRend(点渲染)从计算机图像学的渲染算法中受到启发,将图像分割视为渲染问题,构建了一种逐点优化的图像边界优化方法。PointRend的基本思想就是对图像分割中容易出错的点进行修正,包含如图2所示的关键流程:

(1)选点:从待优化的分割掩码图(称为粗糙预测图)中选出要优化的点。

(2)特征表示:基于选出的点,分别从粗糙预测图和细粒度特征图中取出特征,并进行组合,得到逐点的特征表示。具体来说就是用选出的点的位置,

对粗糙预测图和细粒度特征图分别进行双线性采样, 采样结果在通道上连接, 形成组合特征。

(3) 再预测: 用一个点头 (Point Head) 对组合特征进行再一次的语义分割预测, 得到修正后的预测结果。

(4) 替换: 用修正后的预测结果替换粗糙预测结果。

在训练时, 上述 PointRend 模块内部运行过程的第 4 步替换步骤是不必要进行的, 直接将第 3 步的精细预测结果和真实标签进行损失评估, 然后优化网络。而在推理时, 上述 PointRend 模块的运行过程是迭代进行的。在每一次迭代开始前, 粗糙预测图会经过上采样操作扩张到 2 倍大小再进行处理, 直到结果和原图大小一样。

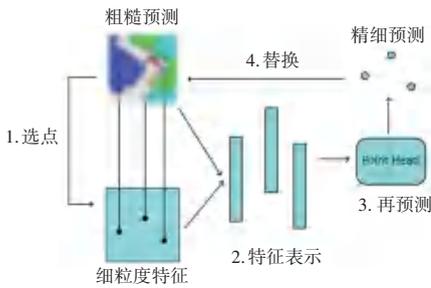


图 2 PointRend 的流程

Fig. 2 Process in PointRend

PointRend 中的点头 (Point Head) 是一个 3 层的、通道数为 256 的、不带归一化操作的、使用 ReLU 激活的多层感知机 MLP。点头的网络结构如图 3 所示。

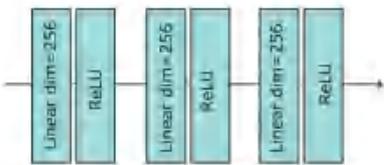


图 3 PointRend 的点头结构

Fig. 3 Point Head structure of PointRend

### 1.3 局部隐式图像函数

真实世界是连续的, 但是数码相机在捕获一个场景的图像时, 会进行离散化的采样以得到像素构成的图像, 也就是数字图像。将离散化前的连续图像称为“连续图像”, 用符号  $I$  表示。若能用一个隐编码表示一整幅图像, 就可以构造出一个隐式函数  $f$  来对这个隐编码进行解码, 得到图像在任意坐标  $x$  上的像素值。对应下面的公式:

$$s = f(z, x) \quad (1)$$

其中:  $z$  表示隐编码, 其可以表示一整幅图像的信息;  $x$  表示查询坐标;  $f$  表示隐式函数。该函数通

过隐编码和坐标, 可以查询到对应坐标处的图像像素值;  $s$  表示图像的像素值。对一般的 RGB 图像, 像素值就是 RGB 值。

但是用一个隐编码  $z$  来表示一整张图像非常困难, 图像的尺寸只要大一些, 隐编码就无法保存图像的细节信息。可以将这套方案局部化, 只对图像的局部区域套用, 那么隐编码表达图像细节信息的能力就会大幅提升。深度学习中可以使用卷积神经网络等编码器对图像  $I$  进行编码, 得到  $H \times W$  个  $D$  维的特征向量。这些特征向量均匀地分布在二维图像空间中, 构成了特征图  $M$ 。将特征图  $M$  中的特征向量当作图像局部区域的隐编码, 那么就可以利用其来查询局部区域图像的像素值。对应公式如下:

$$s = f_{\theta}(z^*, x_q - v^*) \quad (2)$$

其中:  $x_q$  表示查询坐标, 该坐标的范围是整幅图像的二维空间, 而且是连续的, 即可以位于图像的任意一点上, 而不仅限于离散图像的网格点上;  $z^*$  表示特征图  $M$  中, 距离  $x_q$  处最近的特征向量;  $v^*$  表示  $z^*$  对应的坐标。所以  $x_q - v^*$  其实就是查询坐标和特征向量的相对坐标;  $f_{\theta}$  表示局部隐式函数;  $s$  表示图像的像素值。

上述过程可以表述为: 对于一个查询坐标  $x_q$ , 可以通过距离其最近的特征向量, 以及向量之间的相对坐标, 使用隐式函数解出像素值。但是该过程会存在一个跳变问题: 当查询坐标  $x_q$  越过 2 个特征向量的中间位置时, 会使  $z^*$  和  $v^*$  瞬间改变, 即突然更换了特征向量进行求解。这会使得解出的像素值不连续。

可以采用加权求和的方法处理跳变问题: 上述像素查询过程对距离最近的 4 个特征向量同时进行, 最后将各自的结果加权求和。该方法如图 4 所示。图中  $z_{00}^*, z_{01}^*, z_{10}^*, z_{11}^*$  是距离  $x_q$  最近的 4 个特征向量。将其当作隐编码, 分别用隐式函数解出像素值, 然后加权求和。权值分别是  $S_{11}, S_{10}, S_{01}, S_{00}$ , 是公式 (2) 中向量  $x_q - v^*$  所对应的矩形面积。

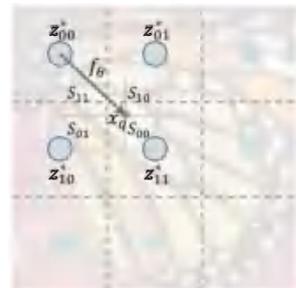


图 4 局部隐式图像函数中的像素值加权求和

Fig. 4 Weighted sum of pixel values in local implicit image functions

局部隐式图像函数的特点是查询坐标是在连续空间中的,而不仅仅只能位于网格点上。这对于超分辨率任务而言,可以在保留更多图像细节信息的情况下,对图像实现高质量的非整数倍、非固定倍数的超分辨率。

## 2 方法

本章节使用 PSPNet 作为基础图像分割模型,结合 PointRend 等改进方法,对图像分割边界进行优化。主要工作如下:(1)向 PSPNet 中引入 PointRend,完成对图像分割边界的优化;(2)通过融合不同阶段的特征,对 PointRend 所使用的细粒度特征进行语义信息和上下文信息的增强,提升分割边界优化效果;(3)使用连续空间特征表示方法增强细粒度特征的空间位置信息,进一步提高分割边界精度。

### 2.1 引入 PointRend

本节对 PSPNet 进行改进,称改进后的网络为 PSPNet-P。

图 5 是 PSPNet 的结构图,图中隐藏了解码部分细节,展示出主干网络细节结构。PSPNet 使用 ResNet-d8<sup>[16]</sup>类型的主干网络,其第 4、5 阶段特征的空间分辨率都是原图的 1/8。PSPNet 的解码器部分是在第 5 阶段(stage5)的特征上计算的,直接输出的分割掩码图的空间分辨率同样是原图的 1/8,使用了双线性插值上采样的方法来恢复空间分辨率。插值上采样操作不能凭空向图像中添加不存在的细节信息,因此得到的分割结果在边界处精度会下降,所以在图 5 中被绘制成模糊的图像。将 PSPNet 的分割结果称为“粗糙预测”。

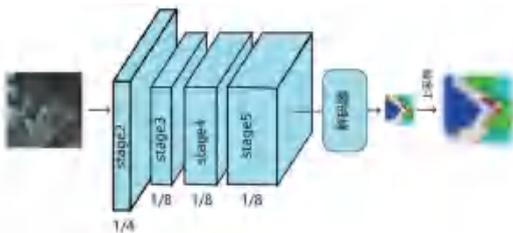


图 5 PSPNet 结构

Fig. 5 PSPNet structure

针对 PSPNet 的分割边界精度不足的问题,将 PointRend 引入到 PSPNet 中,得到改进网络模型 PSPNet-P,结构如图 6 所示。其中 PointRend 使用主干网络的第 2 阶段特征作为细粒度特征,从 PSPNet 的解码器生成的粗糙预测图中选点并进行优化。

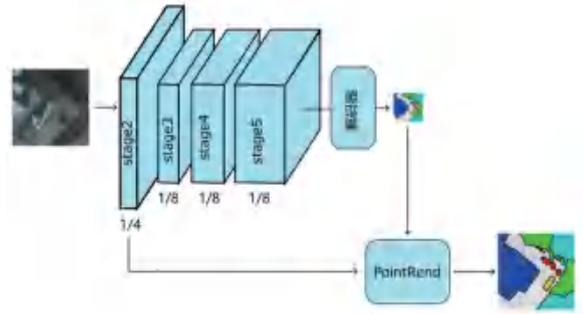


图 6 PSPNet-P 的模型结构

Fig. 6 PSPNet-P structure

### 2.2 融合不同阶段特征

本节通过融合不同阶段特征来提高 PSPNet-P 中 PointRend 的边界优化效果,称改进后的网络为 PSPNet-PM。

前一节中,PSPNet-P 中 PointRend 的细粒度特征来源于第 2 阶段的浅层特征,其有更高的空间分辨率,图像的细节信息更丰富,但是存在局部性严重(即感受野很小)、语义信息量少的问题,对于纹理和细节相似但类别不同的目标无法作出区分。因此本节受特征金字塔网络 FPN (Feature Pyramid Network)<sup>[17]</sup>的启发,通过特征融合来增强细粒度特征,在保证其空间分辨率不变的前提下,让其拥有更多的语义信息和上下文信息。改进后的网络模型 PSPNet-PM 的结构如图 7 所示,网络中添加了一个 MergeFPN(融合特征金字塔网络)模块,将第 2、3、4 阶段的特征融合后作为 PointRend 的细粒度特征。

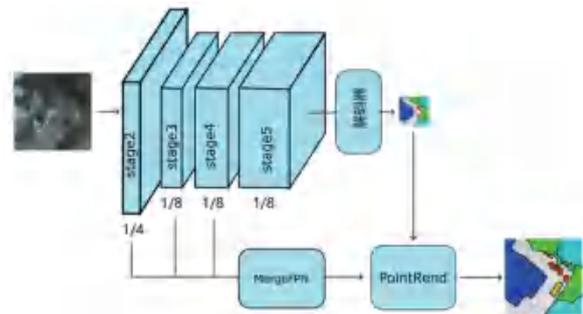


图 7 PSPNet-PM 的网络结构

Fig. 7 PSPNet-PM network structure

MergeFPN 的结构如图 8 所示,先进行和 FPN 类似的特征融合,从深层到浅层、自顶底向下层地将深层特征信息融合到浅层特征中。然后再将各层特征图在通道上连接在一起(对应图中的“C”),最后用一个卷积层再次进行特征融合。

### 2.3 基于连续空间的特征表示

本节对前一节中 PSPNet-PM 进行进一步改进,称改进后的模型为 PSPNet-PML。

前一节的 PSPNet-PM 中, PointRender 通过低分辨率的粗糙预测图生成高分辨率的精细预测, 这相当于是超分辨率的过程, 所以可以将原本用于超分辨率任务的局部隐式函数引入到 PointRender 中, 在连续空间中进行分割边界的优化, 避免离散空间中的网格不对齐的问题, 最终得到更精细的分割边界优化结果。PSPNet - PML 中使用改进后的 PointRender 模块, 其结构如图 9 所示, 原本的“特征表示”被改进为“连续空间特征表示”, 而点头被替换为 LIIF(局部隐式图像函数)模块。

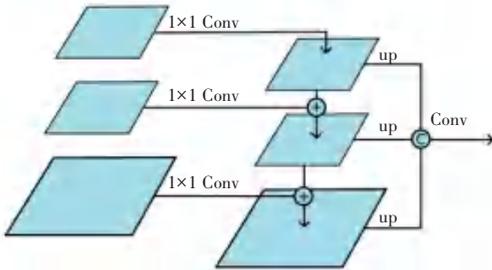


图 8 MergeFPN 结构

Fig. 8 MergeFPN structure

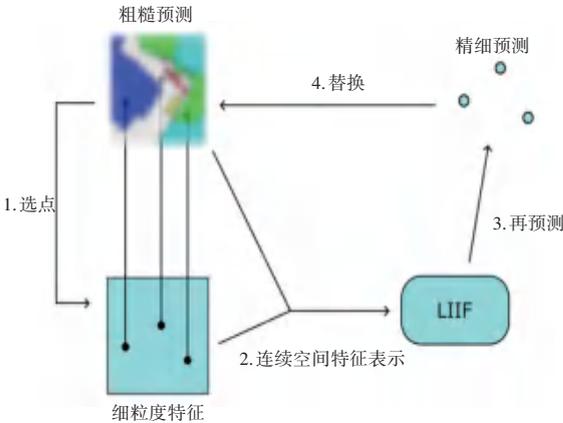


图 9 改进的 PointRender 模块

Fig. 9 Improved PointRender module

连续空间特征表示如图 10 所示。PointRender 模块中第一个“选点”步骤会产生一组待优化点的坐标位置, 对其中的任意一个坐标位置  $x$ , 都可以找出特征图上距离最近的 4 个特征向量  $F^0, F^1, F^2, F^3$  和其坐标  $v_0, v_1, v_2, v_3$ , 并计算出  $x$  与各特征向量的相对坐标  $r_0, r_1, r_2, r_3$ 。将各个特征向量、特征向量坐标和相对坐标分别连接在一起, 每个坐标位置  $x$  就得到了 4 个连续空间特征表示  $F_{cont}^0, F_{cont}^1, F_{cont}^2, F_{cont}^3$ 。

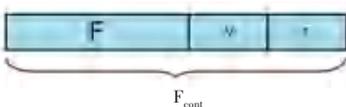


图 10 连续空间特征表示

Fig. 10 Continuous spatial feature representation

连续空间特征表示复合了特征向量本身和坐标信息, 可以使用 PointRender 中的点头充当局部的隐式函数进行解码, 并且解码完成后像图 4 中那样加权求和。将这些过程实现成 LIIF 模块, 如图 11 所示。

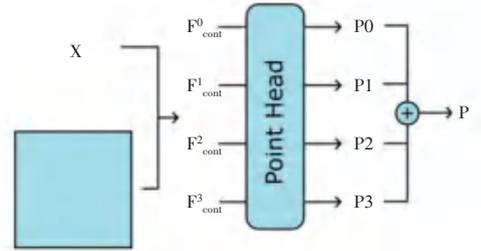


图 11 LIIF 模块

Fig. 11 LIIF module

### 3 实验与分析

#### 3.1 实验内容

为研究改进方法的图像分割边界优化效果, 实验内容安排如下:

(1) 软硬件环境。软件环境为 MMSegmentation1.1.1 + Pytorch1.11 + cuda11.3。硬件上使用 2 张 RTX 3080 显卡来训练和评估神经网络模型。

(2) 实验模型。以 PSPNet 作为基准模型, 逐步添加各项改动, 构建模型进行对比实验。各模型均采用 50 层的 ResNet-d8 作为主干网络。

(3) 数据集选择。首先在 ISPRS Potsdam 数据集<sup>[18]</sup>上进行消融实验, 以研究本章各个改进点的作用。然后在 ISPRS Potsdam 数据集和 LoveDA 数据集<sup>[19]</sup>上进行了图像分割的对比实验, 以验证改进的有效性和泛化性。

(4) 数据增强。在训练时使用随机等比缩放、裁剪、翻转、光度失真(随机改变图像的饱和度、亮度、色调)的数据增强方法。

(5) 训练策略。首先在数据集上预训练 PSPNet 作为基准图像分割模型, 然后将权重复制到添加了 PointRender 相关模块的网络模型之中并且冻结, 再进行训练。这种策略降低了训练难度、加快了训练速度, 并且通过和基准模型的分割结果对比, 能更精确、容易地观察到 PointRender 相关的网络结构所起的作用。

(6) 损失函数。使用多分类交叉熵函数作为损失函数。

(7) 评价指标。使用单位为百分比的交并比 (Intersection over Union, IoU) 与平均交并比 (mean

Intersection over Union, mIoU) 评估图像分割精度。使用边界精度 (boundary Accuracy, bAcc)<sup>[9]</sup> 来评估边界优化效果。bAcc 只在分割图像的边界区域进行平均准确率 (mean Accuracy, mAcc) 的评估, 可以更直接地反映出图像分割边界的优化情况。在实际实验中, 提取出分割标注图中直径为 3 个像素的边界作为边界区域来计算边界精度 bAcc。

### 3.2 ISPRS Potsdam 数据集实验结果与分析

ISPRS Potsdam 数据集是国际摄影测量及遥感探测学会 (ISPRS) 组织发布的德国波茨坦地区的遥感影像语义分割数据集。数据集有 6 个分类: 不透水表面 (Imp surf)、建筑 (Build)、低植被 (Low veg)、树木 (Tree)、汽车 (Car) 和杂物 (Clutter)。其中杂物类别的像素数量显著少于其余类别并且没有显著特征, 部分研究直接忽略该分类, 不统计其分割精度<sup>[20-21]</sup>。为了全面比较网络面对复杂情况下的能力, 6 个分类在实验中都会进行训练并统计分割精度。

ISPRS Potsdam 数据集就带不带边界 (Boundary) 提供两套标注信息。边界是指标注中的不同分类掩码的交界区域。带边界就是标注数据中存在交界, 而不带边界 (无边界) 就是用黑色遮盖了交界处, 以表示在训练时和评价时要忽略的区域。本实验选择带边界的标注进行训练并评价网络性能。

ISPRS Potsdam 数据集除提供红、绿、蓝和近红外共 4 个光谱波段的数据外, 还提供数字地面模型 (Digital Surface Model, DSM), 包含地表高度信息。实验中只使用数据集中的 RGB 波段图像, 并没有使用近红外和 DSM 等数据。

实验中使用滑动窗口裁剪的方法, 将数据集的影像和标注统一裁剪成 512×512 的大小进行实验,

表 2 ISPRS Potsdam 数据集实验结果

Table 2 Experiment result on ISPRS Potsdam

模型	IoU						mIoU	bAcc
	Imp surf	Build	Low veg	Tree	Car	Clutter		
PSPNet	83.18	91.26	72.71	75.00	81.91	39.62	73.95	49.43
PSPNet-PML	<b>83.13</b>	<b>91.37</b>	<b>72.18</b>	<b>75.56</b>	<b>82.90</b>	<b>41.35</b>	<b>74.41</b>	<b>50.25</b>

分割结果可视化对比如图 12 所示, RGB 表示待分割的影像, GT 表示标注。对其进行分析: (1) 从第 1 行与第 2 行可以看出改进后的网络模型 PSPNet-PML 对杂物有更好的分割效果, 其边界更加清晰和准确; (2) 从第 3 行可以看出 PSPNet-PML 对树木的分割更加准确, 覆盖率更高。证明改进后的方法能得到更准确、更精细的图像分割边界。

裁剪时不会超出图像边界部分。得到了共计 5 272 幅的图像, 将其中的 3 456 幅图像作为训练集, 剩余 2 016 幅图像作为验证和测试集。

各改进点的消融实验结果见表 1, 对其进行分析:

(1) PSPNet-P 和 PSPNet 进行对比, 边界精度指标 bAcc 增长了 0.47%, 这说明 PointRend 模块比较准确地找到了图像分割的边界区域进行优化。但是 mIoU 指标几乎没有增长, 这说明 PointRend 模块虽然能找到边界区域, 但是做出的预测相比于原始网络 PSPNet 并没有明显的精度提升。

表 1 Potsdam 数据集消融实验结果

Table 1 Ablation experiment results on ISPRS Potsdam

模型	引入 PointRend	融合不同阶段特征	基于连续空间间的特征表示	mIoU	bAcc
PSPNet	×	×	×	73.95	49.43
PSPNet-P	✓	×	×	74.06	49.90
PSPNet-PM	✓	✓	×	74.17	49.93
PSPNet-PML	✓	✓	✓	<b>74.41</b>	<b>50.25</b>

(2) PSPNet-PM 与 PSPNet-P 相比, mIoU 指标有所提升, 说明通过融合特征来增强细粒度特征的语义信息和上下文信息是有作用的, 可以让 PointRend 模块的优化预测结果更精确。

(3) PSPNet-PML 与 PSPNet-PM 相比, mIoU 指标和 bAcc 指标都提升了, 说明连续空间特征表示通过增强特征的空间位置信息, 可以得到更准确的边界优化结果, 有效提升了分割精度。

改进前后的详细实验结果见表 2。PSPNet-PML 相比于 PSPNet, 除低植被的精度稍有下降外, 树木、汽车和杂物类别的精度都有较明显提升, 最终提升了整体的分割精度指标 mIoU。

### 3.3 LoveDA 数据集实验结果与分析

LoveDA (Land - cOVE dataset for Domain Adaptation) 数据集是一个城市-农村的域自适应地表覆盖数据集, 来自于武汉大学测绘遥感信息工程国家重点实验室 RSIDEA 团队, 该数据集可用于图像语义分割任务或者迁移学习任务。LoveDA 数据集包含了来自南京、常州和武汉地区的遥感影像数

据,包含 7 个土地覆盖类别:背景 (Background)、建筑 (Building)、道路 (Road)、水体 (Water)、荒地 (Barren)、森林 (Forest) 和农业 (Agricultural)。LoveDA 数据集有以下挑战:(1) 包含多尺度对象,要求模型有更好的多尺度特征提取能力;(2) 有复杂的背景类别,数据集的背景类别包含非常多种类的土地覆盖类型,比如草地、无覆盖的裸露土地、空地等,类内方差较大;(3) 类别分布不一致,影像数据中的城市地区的地貌与农村区域有较大差异,要求模型有良好的泛化性,能够适应不同地区的类别分布。本实验将数据集中的 2 522 幅图像用于训练,1 669 幅图像用于验证和评估。

实验结果见表 3。除森林和农业外,各类别的 IoU 指标都有提升,其中最为困难的荒地类别提升明显, IoU 指标提升达到 2.91%。总体的 mIoU 指

标提升了 0.83%,说明本文改进在 LoveDA 数据集上也能对分割的困难点、错误点进行修正,以完成优化分割边界优化。

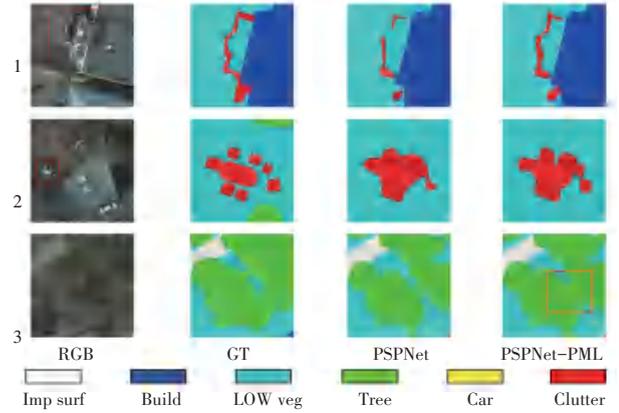


图 12 可视化对比,基于 ISPRS Potsdam 数据集

Fig. 12 Visual comparison on ISPRS Potsdam

表 3 实验结果,基于 LoveDA 数据集

Table 3 Experiment results on LoveDA

模型	IoU							mIoU	bAcc
	背景	建筑	道路	水体	荒地	森林	农业		
PSPNet	50.72	62.98	52.90	66.30	25.26	39.76	58.17	50.87	43.98
PSPNet-PML	<b>52.19</b>	<b>63.95</b>	<b>54.36</b>	<b>67.38</b>	<b>28.17</b>	<b>38.98</b>	<b>56.90</b>	<b>51.70</b>	<b>44.11</b>

可视化对比实验结果如图 13 所示。对其进行分析:(1)从第 1、2、3 行的对比图中可以看出,本文改进模型 PSPNet-PML 对水体都有更精确的分割效果;(2)第 4 行中,PSPNet-PML 对建筑的分割更加精细,建筑的边界部分的贴合度更高。综合来看,本文改进方法能在 LoveDA 数据集上达到优化分割边界的效果,说明改进方法具有良好的泛化性。

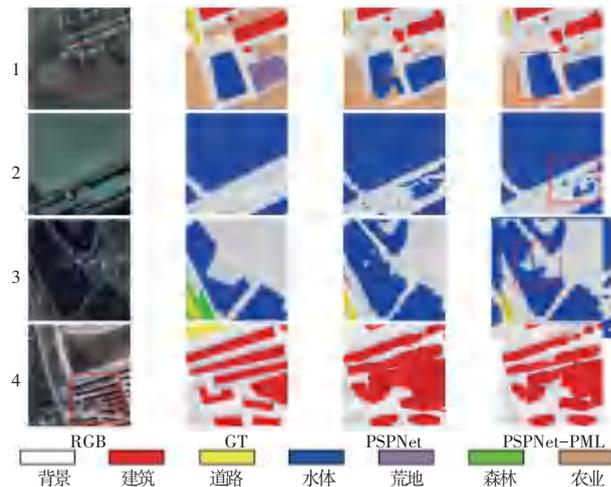


图 13 可视化对比,基于 LoveDA 数据集

Fig. 13 Visual comparison on LoveDA

## 4 结束语

针对遥感图像的分割边界精度低的问题,提出了基于改进 PointRend 的分割边界优化方法。首先将 PointRend(点渲染方法)引入到 PSPNet 中以识别和优化分割边界,然后通过融合不同阶段特征来强化细粒度特征的语义信息和上下文信息、通过连续空间的特征表示方法强化细粒度特征的空间位置信息,以提升 PointRend 的分割边界优化效果。与改进前的网络模型 PSPNet 相比,在 ISPRS Potsdam 数据集上,mIoU 提升了 0.46%,边界精度 bAcc 提升了 0.82%;在 LoveDA 数据集上,mIoU 提升了 0.83%,边界精度 bAcc 提升了 0.13%。实验结果表明,改进的 PointRend 能够修正和校准图像分割中的边界,得到更加精细的分割结果,并且具有良好的泛化性。

## 参考文献

- [1] 张兵. 遥感大数据时代与智能信息提取[J]. 武汉大学学报(信息科学版),2018, 43(12): 1861-1871.
- [2] 刘伟权,王程,臧斌,等. 基于遥感大数据的信息提取技术综述[J]. 大数据,2022, 8(2): 28-57.
- [3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE

- Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015; 3431–3440.
- [4] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017; 2881–2890.
- [5] RONNEBERGER O, FISCHER P, BROX T. U-Net; Convolutional networks for biomedical image segmentation [C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015; 234–241.
- [6] 李梓瑜, 王大东, 于晓鹏. 基于U型残差网络的遥感图像道路提取方法研究[J]. 智能计算机与应用, 2023, 13(9): 51–58.
- [7] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834–848.
- [8] KIRILLOV A, WU Y, HE K, et al. Pointrend: Image segmentation as rendering [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020; 9799–9808.
- [9] CHENG H K, CHUNG J, TAI Y W, et al. Cascadepsp: Toward class-agnostic and very high-resolution segmentation via global and local refinement [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020; 8890–8899.
- [10] HUYNH C, TRAN A T, LUU K, et al. Progressive semantic segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021; 16755–16764.
- [11] SHEN T, ZHANG Y, QI L, et al. High quality segmentation for ultra high-resolution images [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022; 1310–1319.
- [12] CHEN Y, LIU S, WANG X. Learning continuous image representation with local implicit image function [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021; 8628–8638.
- [13] LI Q, YANG W, LIU W, et al. From contexts to locality: Ultra-high resolution image segmentation via locality-aware contextual correlation [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021; 7252–7261.
- [14] WANG M, DING H, LIEW J H, et al. SegRefiner: Towards model-agnostic segmentation refinement with discrete diffusion process [J]. arXiv preprint arXiv, 2312.12425, 2023.
- [15] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint arXiv, 1704.04861, 2017.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016; 770–778.
- [17] KIM S W, KOOK H K, SUN J Y, et al. Parallel feature pyramid network for object detection [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018; 234–250.
- [18] ISPRS. 2D Semantic labeling contest – potsdam [EB/OL]. (2017-09-01) [2024-12-15]. <https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-potsdam.aspx>
- [19] MA A, ZHENG C, WANG J, et al. Domain adaptive land-cover classification via local consistency and global diversity [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1–12. DOI: 10.1109/TGRS.2023.3289347
- [20] LI H, QIU K, CHEN L, et al. SCAAttNet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(5): 905–909.
- [21] LIU Y, MINH NGUYEN D, DELIGIANNIS N, et al. Hourglass-shapenetwork based semantic segmentation for high resolution aerial imagery [J]. Remote Sensing, 2017, 9(6): 522.