

DOI:10.19651/j.cnki.emt.2416684

基于注意力改进 RTformer 的滑坡遥感图像语义分割^{*}

唐海林¹ 张俊¹ 李屹旭² 李升海¹

(1. 贵州大学矿业学院 贵阳 550025; 2. 贵州大学农学院 贵阳 550025)

摘要: 针对现有的遥感影像滑坡语义分割网络存在模型参数量大、训练速度较慢, 滑坡边界区域识别模糊、遥感影像多尺度语义信息分类差异化等问题, 本文提出一种改进的 RTformer 轻量级语义分割模型, 在模型不同层级模块间嵌入空洞卷积注意力 ASPP 模块和通道注意力 SE 模块, 以捕捉不同尺度的语义信息和通过计算通道关系从而增强特征表示能力, 提高模型特征提取能力, 使其更加适用于滑坡遥感影像识别任务。利用 Cityscapes 数据集针对模型中空洞卷积的膨胀率设置和不同批量大小进行对比试验以得到最优解, 以毕节滑坡灾害数据集做为预训练数据集设计一个自监督训练任务, 并使用其进行模型微调并检验模型针对滑坡灾害遥感影像的分割性能。最终得到的模型在 Cityscapes 数据集和毕节市滑坡灾害数据集上均获得了最优表现, 相比原始 RTformer 模型, 两个数据集的平均交并比(mIoU)分别提升了 2.26% 和 4.34%。并且与 FCN、U-Net、DeeplabV3、SegFormer 等经典语义分割模型相比, 改进模型以最少的参数和最快的推理速度实现了识别任务, 并达到了最优分割效果。

关键词: 图像处理; 滑坡检测; 遥感图像; 语义分割; 注意力机制

中图分类号: TP751; TN20 文献标识码: A 国家标准学科分类代码: 420.20

Semantic segmentation of landslide remote sensing image based on improved attention RTformer

Tang Hailin¹ Zhang Jun¹ Li Yixu² Li Shenghai¹

(1. School of Mining, Guizhou University, Guiyang 550025, China;

2. School of Agriculture, Guizhou University, Guiyang 550025, China)

Abstract: Aiming at the existing problems of landslide semantic segmentation network of remote sensing image, such as large number of model parameters, slow training speed, fuzzy recognition of landslide boundary region, and differentiation of multi-scale semantic information classification of remote sensing image, this paper proposes an improved lightweight semantic segmentation model of RTformer. The cavity convolution attention ASPP module and channel attention SE module were embedded among the modules at different levels of the model to capture semantic information at different scales and to enhance the feature representation ability and improve the feature extraction ability of the model, making it more suitable for landslide remote sensing image recognition. Cityscapes data set was used to conduct comparative experiments on the expansion rate setting of the cavity convolution in the model and different batch sizes to obtain the optimal solution. A self-supervised training task was designed using the Bijie landslide disaster data set as the pre-training data set, and the model was fine-tuned and the segmentation performance of the model against the landslide disaster remote sensing images was tested. The resulting model achieved the best performance on both Cityscapes dataset and Bijie landslide disaster dataset. Compared with the original RTformer model, the mean crossover ratio (mIoU) of the two datasets increased by 2.26% and 4.34%, respectively. Compared with the classical semantic segmentation models such as FCN, U-Net, DeeplabV3 and SegFormer, the improved model realizes the recognition task with the fewest parameters and the fastest reasoning speed, and achieves the optimal segmentation effect.

Keywords: image processing; landslide detection; remote sensing image; semantic segmentation; attention mechanism

0 引言

滑坡地质灾害因为其巨大的破坏性和难以预料的突发

性得到世界各国相关研究人员的广泛关注^[1]。滑坡测绘是调查滑坡灾害规模、分布规律及其危害性评价等任务的重要依据^[2]。遥感图像基于遥感技术时效性好、覆盖范围广

收稿日期: 2024-08-17

* 基金项目: 贵州省省级科技计划项目(黔科合支撑[2022]一般 204, 黔科合基础-ZK[2024]一般 093)资助

以及信息丰富等特点,已经发展成为安全、准确、高效获取丰富地物信息的重要途径,为滑坡测绘提供了大量的图像数据,是滑坡检测识别的基础^[3-5]。滑坡识别早期的主流方法包括人工视觉解译法和变化检测的图像分析法等,但是存在时间成本高、主观依赖性强、效率低以及定量描述不准确等问题^[6-7]。近年来,深度学习模型的快速发展,其表现出显著的分层特征表示、自动化和智能能力,语义分割作为深度学习的一项重要任务,因其端到端的像素级分割能力,在滑坡遥感影像的识别任务中具有巨大潜力,许多学者在此领域做了深入的研究。其中,Lei 等^[8]提出基于金字塔全卷积网络方法(FCN-PP),平衡了更广泛的感受野和上下文利用,实现了滑坡的单绘制任务;Bragagnolo 等^[9]利用 U-Net 的编码解码能力,有效地还原边界信息,获得更加清晰的滑坡边界;Lu 等^[10]提出了一种基于光探测和测距数据的轻量级注意力 U-Net,用于历史滑坡的自动探测;在 Yi 等^[11]提出的模型中,LandsNet 语义分割网络得到了增强,融合注意力模块、残差块和多尺度融合操作,可以从单相 RapidEye rsi 中智能地检测和描绘地震诱发的滑坡,对滑坡图像识别取得较高的准确率;陆续有学者分别提出了特征融合模块来实现不同特征的融合,以提高语义分割模型的准确性^[12-13];结合卷积神经网络(convolutional neural network,CNN)和 Transformer 的优点,Wu 等^[14]提出了一种 SCDUNet++ 的语义分割模型,以增强对滑坡特征的判别和提取;Chen 等^[15]提出轻量级模型 BisDeNet,使用更优的上下文路径,加快了滑坡检测速度。

上述大部分网络都是基于卷积神经网络模型,考虑到这些网络结构的模型参数量大、训练速度较慢,本文针对滑坡边界区域识别模糊、遥感影像多尺度语义信息分类差异化等问题,基于 Transformer 提出的轻量级实时语义分割模型(efficient design for real-time semantic segmentation with transformer, RTformer)^[16]进行改进,在模型框架的第 3 阶段加入空洞空间卷积池化金字塔(atrous spatial pyramid pooling, ASPP)模块^[17],以更好地捕捉不同尺度的语义信息,在保证不降低分辨率的情况下扩大感受野,提升模型特征提取能力,并在 RTformer 模块高低分辨率分支均加入通道注意力机制(squeeze and excitation, SE)模块^[18],通过计算通道关系从而增强特征表示能力,使其更加适用于滑坡遥感影像识别任务。对改进模型进行了训练和测试,实验结果表明,改进的 RTformer 算法,相较于其他语义分割模型对滑坡图像识别,滑坡特征识别能力更好,效率更高,并且具有更高的滑坡检测精度。

1 研究方法

本文利用 RTformer 网络作为基本框架,加入空洞卷积注意力 ASPP 模块获取不同尺度遥感影像特征,加入通道注意力 SE 模块,提高识别精度,提出一种适用于滑坡图像识别的模型。

1.1 RTformer 网络结构和原理

RTformer 模型是采用卷积块和 RTformer 块结合构建的骨干网络,如图 1 所示。前两个阶段使用卷积块提取局部信息,stem 阶段使用两个 3×3 conv 和对输入图像做特征提取,stage1 和 stage2 阶段都是基于 basic conv block 操作,包括 3×3 conv、BN、ReLU、残差结构等对图像进一步做局部特征提取。

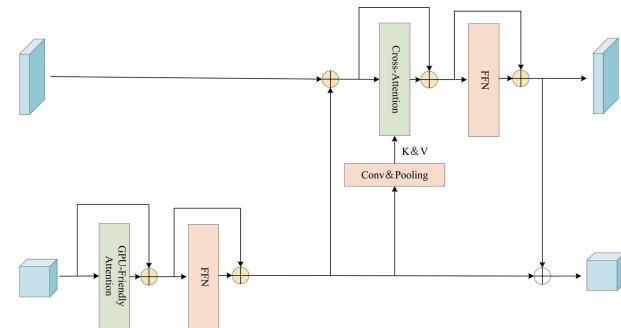


图 1 RTformer 网络结构

Fig. 1 RTformer network structure

第 3 阶段开始通过卷积采用双分辨率模块进行特征交换,具体操作是让通道增加到 2 倍, h 和 w 各减少到原来的 $1/2$,得到低分辨率特征图,具有更广阔上下文信息,将输入 x_2 与经过 conv, pooling 将 C, H, W 调整为与 x_2 一致的 x_3 相加,得到高分辨率特征图。第四和第五阶段使用模型的核心部分 RTformer 模块进行特征聚合,不同分支的输出结果将进入两个连续的 RTformer Block 进行处理,在低分辨率分支中,使用 GPU 友好型注意机制(GPU-friendly attention, GFA)来学习全局上下文,而在高分辨率分支中,使用跨分辨率注意力机制,它使用低分辨率分支提取的键值对,帮助高分辨率特征获取全局语义信息,两个分辨率分支采用分步排列,低分辨率特征先计算,然后助推高分辨率特征学习。最后利用 DAPPM 模块将多尺度特征图进行融合得到分割结果。

1.2 通道注意力 SE 模块

SE 模块通过显示地建模出卷积特征通道之间的相互依赖性,使网络能够自适应地分配不同通道的权重,从而关注重要特征。SE 模块分为 Squeeze、Excitation 以及 Scale 3 种操作,如图 2 所示。其中 Ftr 是一个标准的卷积算子,公式定义如下:

$$U_c = V_c * X = \sum_{s=1}^{c'} V_c^s * X^s \quad (1)$$

式中: $X \in R^{H' \times W' \times C'}$ 表示输入特征图, $U \in R^{H \times W \times C}$ 为输出特征图, V 代表卷积核, X 的每一层经过 2D 空间核的卷积得到 C 个特征图,组成 U 。

Squeeze 操作部分,使用全局平均池化,将包含全局信息的 $W \times H \times C$ 的特征图压缩成一个 $1 \times 1 \times C$ 的特征向量 Z ,该特征图具有全局感受野,其过程如下式所示:

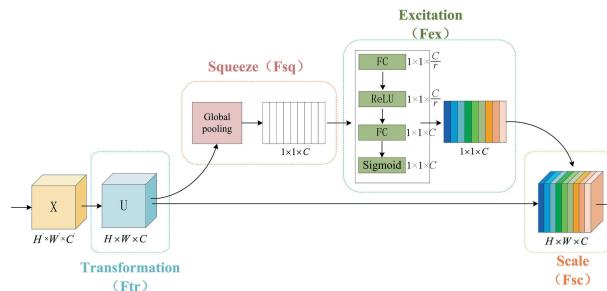


图 2 SE 模块结构

Fig. 2 SE module structure diagram

$$\mathbf{Z}_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2)$$

式中： F_{sq} 表示 Squeeze 操作， H 和 W 表示 U 的空间维度， u_c 表示 U 的第 c 个通道的输出。

在 Excitation 操作中，使用上一步得到的压缩信息，采用两层全连接层和 ReLU 函数执行非线性变换，全面捕获通道依耐性，再通过 Sigmoid 函数激活生成每个通道的特征权重 s ，公式如下：

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

式中： F_{ex} 表示 Excitation 操作， δ 代表 ReLU 函数， σ 表示 Sigmoid 函数。最后是 Scale 操作，将前面得到的注意力权重加权到每个通道的特征上，即 $\tilde{X}_c = F_{scale}(U_c, S_c) = S_c U_c$ ，最终得到 SE 模块的输出 \tilde{X} 。

1.3 空洞卷积注意力 ASPP 模块

本文借鉴 DeeplabV2 系列中使用空洞卷积池化模块 ASPP 的思路^[17]，在 RTformer 的卷积特征提取阶段加入 ASPP 模块，为适应模型特征图的大小，仅使用两个 3×3 卷积，其结构如图 3 所示。ASPP 总体思路是对于给定的输入以不同的采样率的空洞卷积进行并行采样，将得到的结果 concat 到一起，以扩大通道数，然后再通过 1×1 Conv 将通道数降低到预期的数值，这样就相当于以多个比例捕捉到图像的上下文。

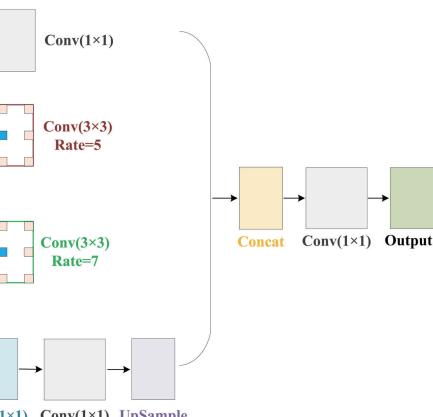
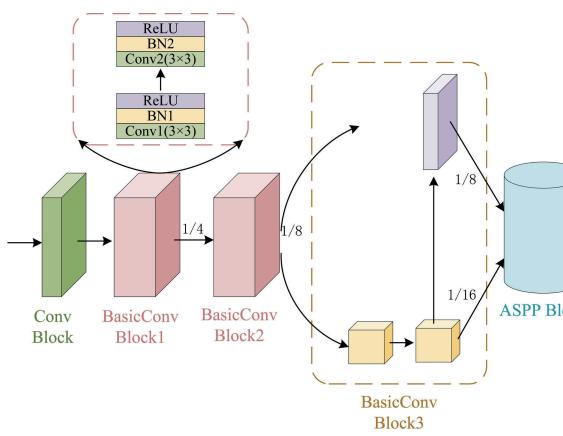


图 3 ASPP 模块结构

Fig. 3 ASPP model structure

ASPP 是基于空洞卷积和空间金字塔池化两个方法结合提出的，空洞卷积可以扩大感受野，捕捉不同尺度的上下文信息，空间金字塔池化则可以在不同分辨率下利用多个不同尺度的池化层进行特征提取，弥补了空洞卷积只能在一个尺度下工作的问题。

具体来说，输入的特征会经过 1×1 卷积层得到输出，同时经过两个不同空洞率的 3×3 卷积层得到输出，一般而言，空洞率的不同直接影响着特征提取的精度，此外，还会经过一个池化卷积层，并进行上采样到同一大小，得到输出，并将这 4 个输出进行拼接，经过 1×1 卷积后输出，还原为输入特征大小。

1.4 改进的 RTformer 滑坡分割模型

为顾及遥感影像多尺度语义信息，在 stage3 阶段后引入 ASPP 模块，强化模型对不同尺度影像特征的识别能力，在 RTformer 板块高低分辨率分支注意力机制前引入通道注意力机制 SE 模块，一方面缓解因 ASPP 结构造成的特点通道数问题，另一方面可以自适应地调整通道权重，强化模型特征学习能力，并得到含多尺度语义信息的语义信息，改进后的网络结构如图 4 所示。

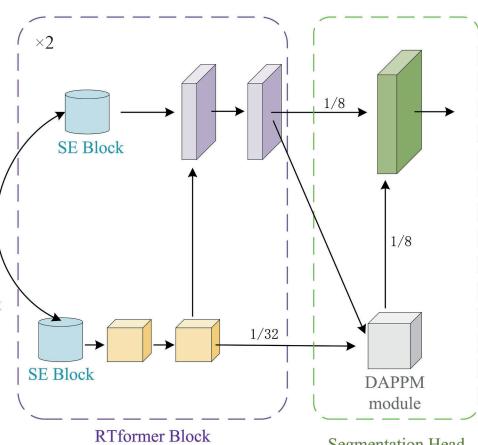


图 4 改进 RTformer 网络结构

Fig. 4 Improved RTformer network structure

网络的工作流程如下:图像输入网络之后,首先会经过一个卷积块 ConvBlock,包含两个步长为 2 的 3×3 卷积($k, s, p=3, 2, 1$),进行初步的特征提取之后进入两个类似的基础卷积模块 Basic Conv Block(3×3 卷积中 $k, s, p=3, 1, 1$);之后相较于原模型增加一个额外的学习层空洞卷积池化模块 ASPP 模块,它可以让模型学习到多个比例的图像上下文信息。

接下来,网络开始使用双分辨率模块,同时提取高分辨率分支和低分辨率分支的特征。不同分支的输出结果将进入两个连续的 RTformer Block 进行处理,在高低分支的注意力机制前分别加入 SE 模块,使得模型可以捕获更广阔的上下文信息,在低分辨率分支中,使用 GPU 友好注意力机制来学习全局上下文,而在高分辨率分支中,使用跨分辨率注意力机制,它使用低分辨率分支提取的键值对,帮助高分辨率特征获取全局语义信息,两个分辨率分支采用分步排列,低分辨率特征先计算,然后助推高分辨率特征学习。最后利用 DAPPN 模块将多尺度特征图进行融合,进行语义预测,得到分割结果。

2 实验过程与结果分析

基于改进后的模型,分别在两个数据集上进行了实验,并对实验结果进行分析和评估,最终实现了滑坡图像的语义分割任务。

2.1 数据集与实验环境

1) 数据集介绍

为了验证提出的改进模型的有效性和泛化性,在城市场景数据集 Cityscapes 和贵州省毕节市的滑坡遥感数据集上进行了测试,由于滑坡数据集的影像数量较少,为了验证模型的鲁棒性,先将城市场景数据集 Cityscapes 作为改进模型的测试数据集,检验模型性能,并以滑坡数据集做为预训练数据集,设计一个自监督训练任务,提高模型对滑坡数据的特征提取能力,然后使用毕节市滑坡灾害数据集进行模型微调并检验模型针对滑坡灾害遥感影像的分割性能。

Cityscapes 是语义分割领域常用的数据集,包含超过 50 个城市的 5 000 个高分辨率图像,这些图像拍摄于不同时间、不同天气情况下的城市街景,包括昼夜变化、阴天和晴天等,图像中包含多种交通工具、行人、建筑物以及路标等元素,使得数据集具有很高的视觉复杂性。贵州省毕节市的滑坡遥感数据集包括从 TripleSat 卫星 2018 年 5~8 月的图像中裁剪的 770 个滑坡样本和标签,包含岩石崩落、岩滑和少量岩屑滑动,每个样本图像截取滑坡的边界框加上 40 m 的边界部分,以及 2 003 个覆盖各种背景负样本对。边界部分的背景和负样本图像可以使网络模型更好地学习滑坡本身的特点。数据集中卫星 RGB 图像的地表分辨率为 0.8 m,每个滑坡的形状矢量由研究人员使用 ArcGIS 结合官方资料手动划定。

2) 实验环境

本文实验在 Windows11 操作系统下进行,采用 PaddleSeg 深度学习框架,所使用计算机配置 CPU 为 i7-13650HX, GPU 为 RTX4070, 使用 CUDA11.2 作为并行计算结构,将其与 GPU 相结合进行训练加速,在 Paddlepaddle2.6 环境下进行模型的训练。训练过程使用交叉熵损失函数,并使用 AdamW 优化器进行网络梯度更新,训练计数采用迭代次数,初始学习率设置为 0.000 06,一阶和二阶矩估计指数衰减率设为 0.9 和 0.999,权重衰减系数设置为 0.01,衰减策略使用 Polynomial Decay,数据集的 Batchsize 设置为 8,对于两个数据集,在训练时采用随机裁剪方式进行数据增强,均按照 7:2:1 的比例划分训练集、验证集和测试集。

数据集具体配置如表 1 所示。

表 1 数据集具体配置

Table 1 Configure data sets

项	Cityscapes 数据集	毕节滑坡数据集
train_numbers	3 500	6 020
valid_numbers	1 000	1 720
test_numbers	500	860
image_size	512×512	512×512
iterations	120 000	120 000
optimizer	AdamW	AdamW
strategy	Polynomial Decay	Polynomial Decay
Batch size	8	8

2.2 评价指标

为了客观地对比提出模型在改进前后的优劣,并验证其在滑坡遥感影像识别中的性能,评价指标采用图像语义分割领域常用的指标:平均交并比(mIOU)、像素精度(PA)、Dice 相似系数和 Kappa 系数,在二分类数据集毕节滑坡影像数据集中,还添加了 F1 分数指标来评价模型性能,它们的计算公式如下:

$$IOU_i = \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (4)$$

$$mIOU = \frac{1}{k+1} \sum_{i=0}^k IOU_i \quad (5)$$

$$PA = \frac{\sum_{i=0}^k P_{ii}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}} \quad (6)$$

式中: k 表示分类的类别数量,表示将第 i 类分为第 i 类(即正确分类)的像素数量,表示将第 i 类分为第 j 类(即所有像素数量)的像素数量。

$$Dice = \frac{2 \times T_p}{2 \times T_p + F_p + F_N} \quad (7)$$

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \quad (8)$$

$$F1 = \frac{2 \times (precision \times recall)}{precision + recall} \quad (9)$$

式中: T_p 表示真正例, F_p 表示假正例, F_N 表示假负例。 P_o 代表观测到的一致性比例, P_r 代表预测的随机一致性比例, precision 代表精确率, recall 代表召回率。

2.3 实验结果与分析

1) 膨胀率设置对比

考虑到贵州省毕节市滑坡数据集的遥感图像数量较少,且 cityscapes 数据集也属于卫星遥感影像分类任务数据集,具有一定相似性,故用 cityscapes 分类数据集验证改进模型性能。

遥感影像语义分割最优的空洞空间卷积池化金字塔膨胀率组合是影响模型性能的一项重要因素,合理选择膨胀率能够平衡特征图的分辨率和感受野大小,实现模型最佳特征提取能力,因此在 cityscapes 数据集上进行试验,为了适配模型参数,在 ASPP 中只加入两个空洞卷积块,将 ASPP 模块中的空洞卷积的膨胀率组合设置为 $n=1, 3, 5, 7, 9, 11$ 分别两两组合进行对比实验,实验结果如表 2 所示。由表 2 可以看出,在使用 5、7 空洞卷积组合时模型具有更好的性能,故后文实验均使用此膨胀率组合。

表 2 不同膨胀率下改进模型精度

Table 2 Improved model accuracy under different expansion rates

膨胀率	Kappa	Dice	PA	mIOU	%
1,3	94.43	85.00	95.72	75.06	
3,5	94.42	84.98	95.70	75.12	
5,7	94.51	85.15	95.77	75.37	
5,9	94.22	84.42	95.55	74.40	
5,11	94.42	85.10	95.70	75.23	
7,11	94.48	84.94	95.75	75.05	
9,11	94.46	85.08	95.74	75.30	

2) 超参数设置对比

本文采用批量随机梯度下降算法进行模型参数优化,模型的权重更新取决于批量大小(Batch Size)参数的影响,批量大小是指每次参数更新时所使用的样本数量,它直接影响模型的泛化能力。

在一定范围内增大批量大小可以减少训练时间,提高模型稳定性的同时带来一定的性能提升,但若超出范围则不仅会导致模型性能下降,还可能因显存超限导致模型崩溃。因此需要结合硬件条件设置合适的批量大小。基于上述原因,本文设置不同的批量大小进行模型的性能对比实验,分别设置成 2、4、6、8 的批量大小,超过 8 模型就因为显存超限而崩溃,不同 Batch Size 的模型性能如表 3 所示,在批量大小设置为 8 的时候各个指标都达到最优,因此后面的实验批量大小参数均设置为 8。

表 3 不同 Batch Size 下模型性能对比

Table 3 Comparison of model performance under different Batch Sizes

Batch Size	Kappa	Dice	PA	mIOU	%
2	93.07	80.06	94.67	68.47	
4	94.51	85.15	95.77	75.37	
6	94.81	85.94	96.00	76.43	
8	95.00	86.92	96.15	77.87	

3) 消融实验及各模型实验结果

为了验证提出模型的性能,在相同实验条件下将本文方法分别与 FCN、Unet、Deeplabv3、Segformer、RTformer 等当下主流网络进行对比实验及改进模型的消融实验,在 Cityscapes 数据集和毕节滑坡数据集上的实验结果如下。

由表 4 可以看到,本文提出的改进模型在 Cityscapes 数据集上的分割效果达到了最优,mIOU 达到了 77.87%,相较于传统经典模型 FCN、Deeplabv3、Segformer,分别提升了 10.16%、9.31% 和 6.80%,在基础模型 RTformer 上提升了 2.26%,表示模型能够准确地预测图像中的目标物体,并与真实标签之间具有较高的重叠度。

表 4 Cityscapes 数据集下各模型精度

Table 4 Precision of each model in Cityscapes dataset

模型	Kappa	Dice	PA	mIOU	%
FCN	93.53	76.51	95.01	67.71	
Unet	88.59	59.57	91.28	51.23	
Deeplabv3	92.88	77.98	94.51	68.56	
SegFormer	92.99	80.15	94.61	71.07	
RTformer	93.68	83.60	95.14	75.61	
RT+ASPP	94.44	84.84	95.72	77.40	
RT+SE	94.39	85.96	96.68	77.55	
Tish Study	95.00	86.92	96.15	77.87(2.26↑)	

其他指标 Kappa 系数、Dice 相似系数和像素精度(PA)分别为 95.00%、86.92%、96.15%,说明本文的模型具有较高的一致性和准确性、保持目标物体的形状和位置,并且能够准确地标记像素。消融实验结果表明,ASPP 模块对模型 mIOU 提升了 1.79%,说明其多尺度特征提取能力能够帮助模型更好地完成识别任务。加入的 SE 模块通过计算通道之间的关系给模型 mIOU 带来 1.94% 的提升。

对于毕节滑坡语义分割任务,在 Cityscapes 数据集上对模型进行了初步参数调优,考虑到 Cityscapes 数据集中主要是城市街景元素,而滑坡数据主要是滑坡区域和滑坡背景两个元素,存在较大差异,且滑坡图像数据较少,为提高模型的性能和泛化能力,因此,以滑坡数据集作为预训

练数据集,并通过几何变换进行数据增强,得到大量预训练数据,基于改进的 RTformer 模型,设计了一个自监督任务,训练得到最优权重参数,再利用标注数据进行监督微调,最后与未经预训练的模型进行预测图像对比结果如表 5 所示。

表 5 有无预训练数据改进模型各项指标对比

Table 5 Comparison of indicators of the improved model with or without pre-training data

指标	无预训练模型	预训练模型	相对提升
mIOU/%	98.03	99.52	+1.49
PA/%	99.57	99.89	+0.32
FLOPs/G	16.91	16.91	—
参数量/M	16.87	16.87	—
收敛轮数/次	8 500	8 300	+200

从表中可以看出,经过预训练处理模型在平均交并比(mIOU)和像素精确率(PA)都有所上升,说明模型在预训练过程中学习到的特征具有更好的泛化能力,有利于模型在不同条件下的滑坡图像分割,收敛轮数有所减少,说明模型的学习效率也随之提升,缩短了训练时间。

由于毕节滑坡数据是二分类任务,因此各项指标有更好地表现,从表 6 可以看出,在滑坡数据集上本文提出的方法性能最优,改进模型 mIOU 达到 99.52%,均超过其他主流网络,对比于 FCN、Deeplabv3、Segformer 网络,分别提升了 10.55%、7.69% 和 4.38%,并在基础模型 RTformer 上提升了 4.34%,说明改进模型对滑坡细节有更好的感知能力。其中 F1 分数达到了 99.20% 的较高水准,说明模型在保持较高精确度的同时,能够具备较高的召回率,在正

确分类的同时可以尽可能地找到更多的正例。从消融实验结果来看,在滑坡数据集中,加入 ASPP 模块使得模型的 mIOU 提升了 1.87%,可以验证其给模型性能的提升起到了积极作用;高低分辨率分支加入的 SE 模块给模型带来 1.88% 的提升,可以验证关注模型通道之间关系可以对滑坡分割任务的有效性;最终 ASPP 模块与 SE 模块共同为模型提供了 4.34% 的性能提升,可以证明每个模块的有效性。

表 6 毕节滑坡数据集下各模型精度

Table 6 Accuracy of each model in Bijie landslide data set

模型	Kappa	Dice	PA	mIOU	F1	%
FCN	87.93	93.97	97.45	88.97	93.98	
Unet	76.05	88.03	94.81	79.78	88.03	
Deeplabv3	94.27	95.64	98.15	91.83	95.64	
SegFormer	94.95	97.47	98.91	95.14	97.47	
RTformer	95.00	97.50	98.92	95.18	97.50	
RT+ASPP	96.92	98.46	99.34	96.99	98.46	
RT+SE	96.93	98.46	99.34	97.00	98.39	
This study	98.56	99.25	99.89	99.52(4.34↑)	99.20	

另外,由分割结果可以看出 FCN、Deeplabv3 + 和 Segformer 网络在 Cityscapes 数据集上均未能全部正确识别圈选区域内的目标,而本文方法对这些目标进行了完整细致的分割,展现出更强的特征提取能力。如图 5 所示,可以明显看出本文提出的方法在与其他方法的比较中表现出明显的优势。例如在图 5 第 2、3 行区域,FCN、Deeplabv3 和 Segformer 未能正确识别圈选区域内的道路,

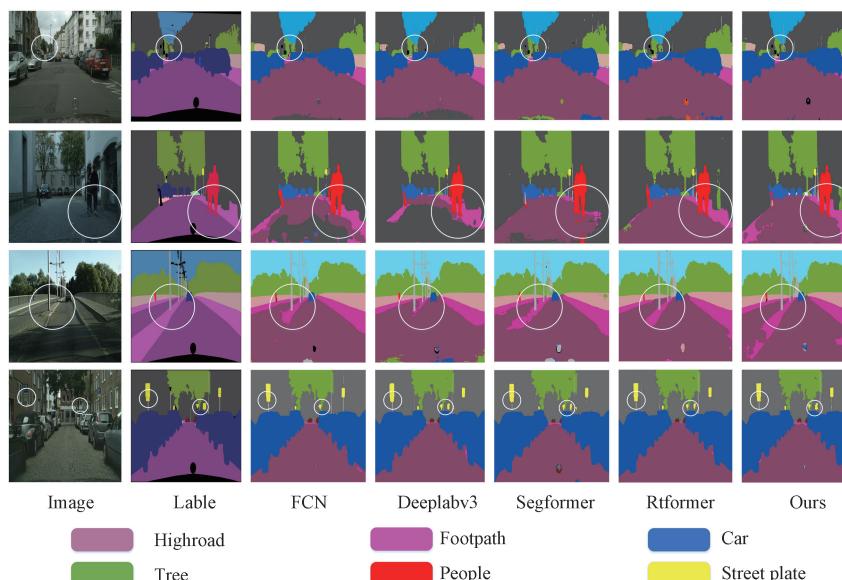


图 5 Cityscapes 数据集实验结果

Fig. 5 Experimental results of Cityscapes dataset

而本文方法对该道路进行了较为完整细致的分割,展现出更强的特征提取能力。在图 5 第 4 行区域,圈选范围对路边的路牌的边缘部分 FCN 不能完整的提取,提出地的改进 RTformer 模型与原始标签达到了最高的相似性。在滑坡数据集上,本文方法相比于其他模型对滑坡轮廓的特征提取更加完对滑坡边缘呈锯齿状,Segformer 模型对滑坡区域细节分割不完整,而改进能够对滑坡的边缘部分进行

更精细地区整,对滑坡边缘区域识别更加准确,再次验证了改进模型可以有效地完成基于遥感影像的滑坡语义分割任务,并且有着较高的准确率。如图 6 所示,在图 6 第 1、2 行区域,FCN 和 Deeplabv3 对滑坡的分割都出现间隙孔洞、边缘轮廓曲折等问题,而本文方法在圈选范围内与原始标签保持了高度一致。此外,进一步观察图 6 第 3 行区域,Deeplabv3 分,从而展现出了更高的分割精度。

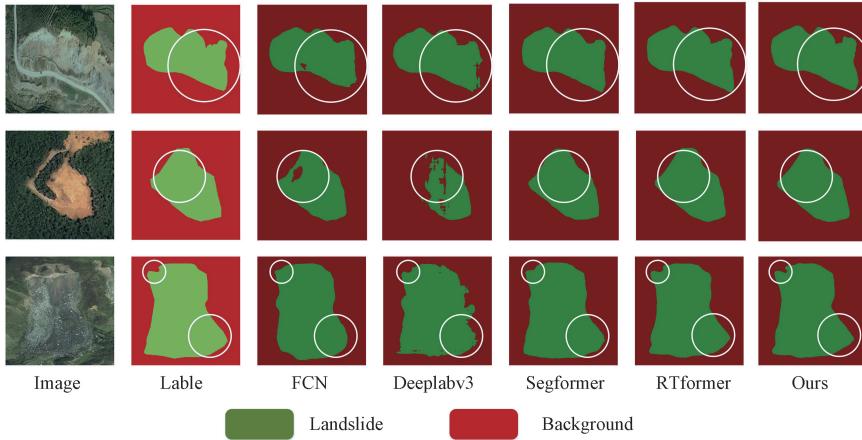
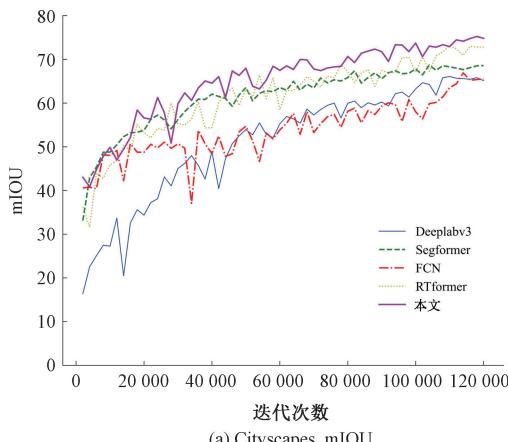


图 6 毕节滑坡数据集实验结果

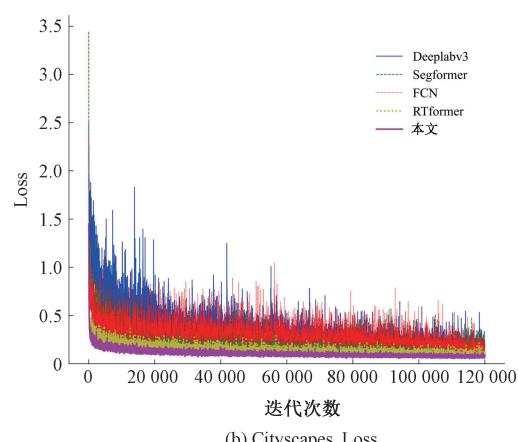
Fig. 6 Experimental results of Bijie landslide dataset

此外,本文对各模型的训练过程进行了可视化,展示了平均交并比(mIOU)和损失值(Loss)变化情况,如图 7 和 8 所示,Loss 值是用来评估模型预测结果与实际结果之

间差异的数量,一般而言,较低的 Loss 值意味着模型的预测结果与实际结果更接近。



(a) Cityscapes_mIOU



(b) Cityscapes_Loss

图 7 Cityscapes 数据集实验过程可视化

Fig. 7 Visualization of experimental process of Cityscapes dataset

总体来看,本文所提出的方法在两个数据集上均表现出最小的 Loss 值变化趋势,且并逐渐趋于稳定,避免了过拟合的风险,说明改进模型具有更高的分割精度。值得注意的是,在训练过程中,虽然其他模型在不同数据集上的 mIoU 表现各不相同,但总体上本文所提出的方法的 mIoU 值最高,尤其在毕节滑坡数据集上表现出明显优势,这表明本文提出的方法具有更优的分割精度和稳定性。

4) 模型复杂度对比分析

如表 7 所示,对传统的几种热门模型和本文提出的方法的模型参数和运行效率进行了对比。其中 UNet、Deeplabv3、Segformer 网络的参数数量较多,每秒千兆次浮点运算(GFLOPS)分别为 128.88、162.71、85.53 G,而 RTformer 为 16.89 G,改进模型为 16.91 G;模型内存消耗情况 Segformer 模型最大,为 64.01 M,Deeplabv3 模型次之,为 39.13 M,RTformer 模型和改进模型均为 16.87 M。

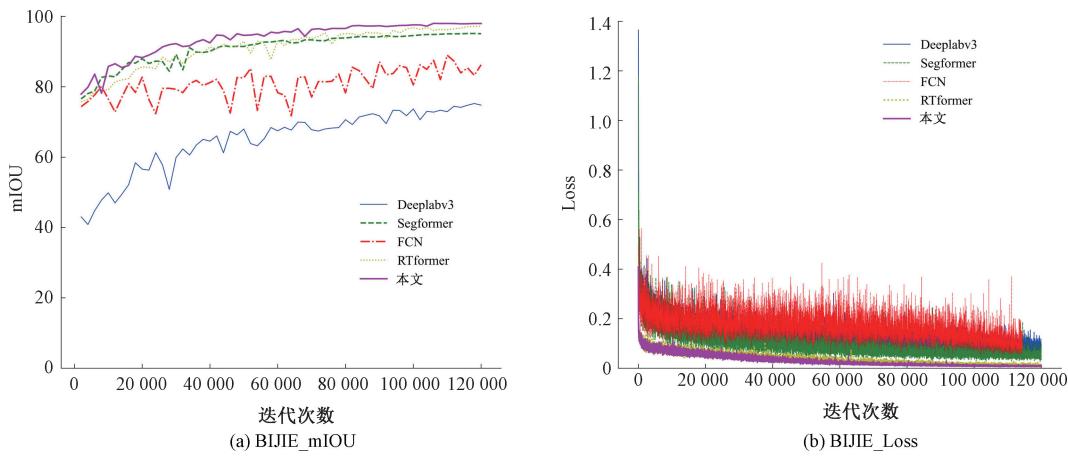


图 8 毕节滑坡数据集实验过程可视化

Fig. 8 Visualization of experimental process of Bijie landslide dataset

表 7 模型复杂度对比

Table 7 Comparison of model complexity

模型	GFLOPS/G	Parameters/M	Computation time(sec/10iters)	mIOU/%
FCN	18.59	9.67	3	88.97
UNet	128.88	13.41	5	79.78
Deeplabv3	162.71	39.13	3	91.83
Segformer	85.53	64.01	5	95.14
RTformer	16.89	16.87	2	95.18
本文	16.91	16.87	2	99.52

因此,本文的改进 RTformer 模型在保持轻量化的同时,成功实现了性能的全面提升。首先,相较几个经典语义分割网络,在参数数量和模型内存消耗方面,改进模型实现了大幅度降低,更低的硬件要求为在资源受限的边缘设备上部署提供了可能。其次,模型的推理速度得到了一定提升,能更快实现实时滑坡区域识别,为及时灾害预警和响应提供了有力支持。

模型在增加了极少量的 GFLOPS 参数,训练时间却几乎保持不变的情况下,模型的平均交并比(mIOU)从原来的 95.18% 提升至 99.52%,这一提升充分展示了改进模型在滑坡区域精确分类方面的良好能力。模型的识别准确率明显提升,这充分证明了本文优化策略的有效性。更少的计算资源和更快的模型迭代和优化周期,体现出改进模型良好的可扩展性,它能够更好地适应高分辨率遥感图像,同时有能力处理更大规模的遥感图像数据集,这对于提高滑坡识别的精度和覆盖范围具有重要意义。

3 结 论

本文在 RTformer 网络基础上,提出了一种适用于滑坡遥感图像改进语义分割模型,在 RTformer 模型中嵌入空洞卷积注意力 ASPP 模块和通道注意力 SE 模块,旨在提升网络对空间信息和通道信息的特征捕捉能力。提出的模型在毕节滑坡数据集上获得了最优表现,像素精确度

为 99.89%,平均交并比为 99.52%,F1 为 99.20%。相比经典语义分割网络 FCN、Unet、Deeplabv3、SegFormer 等,改进网络在滑坡特征提取能力以及效率明显更优,对滑坡数据集的边缘区域分割精度更高,边缘信息处理更平滑,能更有效地对滑坡影像实现精确分割。此外,消融实验的结果也验证了所提出的各个模块的有效性。改进模型不仅在计算效率和资源利用上取得了提升,在识别准确度上也得到了提高。这些特性使得该模型在实际滑坡监测应用中具有广阔的应用前景,为提高灾害预防和应对能力提供了有力工具。这一研究进一步证实了注意力机制在提高实时语义分割任务性能中的巨大潜力,并为轻量级网络在滑坡图像语义分割提供了新的思路。

参考文献

- [1] 许嘉慧,孙德亮,张虹,等.多尺度滑坡灾害生态风险评价与风险管理研究——以三峡库区万州-巫山段为例[J].生态学报,2023,43(11): 4594-4603.
XU J H, SUN D L, ZHANG H, et al. Ecological risk assessment and risk management of multi-scale landslide disaster: A case study of Wanzhou-Wushan section in the Three Gorges Reservoir Area[J]. Acta Ecologica Sinica, 2023, 43(11): 4594-4603.
- [2] 梅伟,顾世祥,刘鑫,等.基于滑坡大变形过程的滑坡定

- 量风险评估方法[J]. 武汉大学学报(工学版), 2022, 55(5):443-453.
- MEI W, GU SH X, LIU X, et al. Quantitative risk assessment of landslide based on large deformation process[J]. Journal of Wuhan University(Engineering Edition), 2022,55(5):443-453.
- [3] 许强,董秀军,李为乐. 基于天-空-地一体化的重大地质灾害隐患早期识别与监测预警[J]. 武汉大学学报(信息科学版), 2019,44(7):957-966.
- XU Q, DONG X J, LI W L. Early identification, monitoring and early warning of major geological hazards based on the integration of space-air-ground[J]. Journal of Wuhan University (Information Science Edition), 2019,44(7):957-966.
- [4] GUZZETTI F, MONDINI A C, CARDINALI M, et al. Landslide inventory maps: New tools for an old problem[J]. Earth-Science Reviews, 2012,112(1-2): 42-66.
- CHEN H X, ZHANG S, PENG M, et al. A physically-based multi-hazard risk assessment platform for regional rainfall-induced slope failures and debris flows[J]. Engineering Geology, 2016,203:15-29.
- [6] 胡芳驰,樊雅婧. 遥感影像滑坡灾害识别应用研究[J]. 农业灾害研究, 2021,11(4):98-99.
- HU F CH, FAN Y J. Research on the application of remote sensing image to landslide hazard recognition[J]. Agricultural Disaster Research, 2021,11(4):98-99.
- [7] XU CH. Preparation of earthquake-triggered landslide inventory maps using remote sensing and GIS technologies: Principles and case studies [J]. Geoscience Frontiers, 2015,6(6):825-836.
- LEI T, ZHANG Y X, LYU ZH Y, et al. Landslide inventory mapping from bitemporal images using deep convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2019,26(6): 892-896.
- [9] BRAGAGNOLO L, REZENDE L R, SILVA R V, et al. Convolutional neural networks applied to semantic segmentation of landslide scars [J]. CATENA, 2021,201:105189.
- [10] LU Z L, PENG Y X, LI W, et al. An iterative classification and semantic segmentation network for old landslide detection using high-resolution remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-13.
- [11] YI Y N, ZHANG W CH. A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal RapidEye satellite imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020,13:6166-6176.
- [12] LIU X R, PENG Y X, LU Z L, et al. Feature-fusion segmentation network for landslide detection using high-resolution remote sensing images and digital elevation model data [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.
- [13] NIU CH Y, GAO OU Y, LU W J, et al. Reg-SA-UNet++: A lightweight landslide detection network based on single-temporal images captured postlandslide[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 9746-9759.
- [14] WU L, LIU R, JU N P, et al. Landslide mapping based on a hybrid CNN-transformer network and deep transfer learning using remote sensing images with topographic and spectral features [J]. International Journal of Applied Earth Observation and Geoinformation, 2024,126:103612.
- [15] CHEN T, GAO X, LIU G, et al. BisDeNet: A new lightweight deep learning-based framework for efficient landslide detection [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024,17:3648-3663.
- [16] WANG J, GOU CH H, WU Q M, et al. RTFormer: Efficient design for real-time semantic segmentation with transformer[J]. Advances in Neural Information Processing Systems, 2022, 35: 7423-7436.
- [17] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. 2018,40(4):834-848.
- [18] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

作者简介

唐海林,硕士研究生,主要研究方向为计算机视觉与遥感图像处理与识别。

E-mail:1941249759@qq.com

张俊(通信作者),副教授,硕士生导师,主要研究方向为现代测绘数据处理教学、研究。

E-mail:jzhang13@gzu.edu.cn

李屹旭,讲师,主要研究方向为遥感科学与技术、农业信息技术教学。

E-mail:lyxgis@163.com

李升海,硕士研究生,主要研究方向为遥感图像信息处理。

E-mail:3277685445@qq.com