doi:10. 3969/j.issn. 1672-4623.2023.02.006

多尺度特征融合神经网络的地表覆盖提取

罗　想[1]，张　荞[2]，黄　昶[1]
（1. 自然资源部第三航测遥感院，四川 成都 610100；2. 西南石油大学地球科学与技术学院，四川 成都 610500）

摘　要：深度学习的快速发展，为高分辨率卫星遥感影像解译提供了更好的技术手段和应用前景。围绕高分卫星遥感影像地表覆盖信息提取，利用高分辨率卫星遥感影像制作5种常见的地表覆盖类型的像素级样本数据集，并提出一种基于注意力增强与多尺度特征融合的语义分割方法，实现地表覆盖自动提取。通过影像波段选择、预训练模型迁移学习、损失函数改进等方法，提升语义分割模型识别精度，最优的提取结果中，5种地表覆盖类型的 F_1 均值、IoU均值和总体精度分别达到了78.6%、66.8%、85.0%，除道路之外，耕草、林地、建筑、水体的 F_1 均超过80%，且分类图斑边界能够与影像中的地物边界很好套合。实验表明建立的卫星影像地表覆盖分类样本数据集和分类方法，能够应用于高分辨率卫星影像地表覆盖信息提取。

关键词：深度学习；注意力机制；多尺度特征融合；卫星影像；地表覆盖
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2023）02-0029-06

Land Coverage Extraction Based on Multi-scale Features Fusion Neural Network

LUO Xiang[1], ZHANG Qiao[2], HUANG Chang[1]
(1. The Third Institute of Photogrammetry and Remote Sensing, Ministry of Natural Resources, Chengdu 610100, China;
2. School of Geoscience and Technology, Southwest Petroleum University, Chengdu 610500, China)

Abstract: The rapid development of deep learning provides a more effective approach and brighter prospects for interpretation of high-resolution satellite remote sensing imageries. In this paper, we extracted land coverage information from high-resolution satellite remote sensing imageries by producing a pixel wise dataset containing five common land coverage classes and using a semantic segmentation model based on the attention enhanced and multi-scale features fusion neural network. The accuracy of the semantic segmentation model could be promoted by several approaches such as choosing special spectral bands, transfer learning with pre-trained models and modifying loss function. For the optimal extraction result, the mean F1, mean IoU and overall accuracy of the five land coverage classes are 78.6%, 66.8% and 85.0%, respectively. Except for road, the F1 of crop and grass land, forest, building and water are all surpass 80%, and the boundaries of terrain objects in the classification chart can be well matched with those in the images, which shows that the pixel wise dataset and semantic segmentation method established in this paper can be used for the land coverage extraction from high-resolution satellite imageries.

Key words: deep learning, attention mechanism, multi-scale features fusion, satellite imagery, land coverage

随着深度学习在机器视觉领域取得突破，深度学习理论和方法开始被引入遥感影像的处理与分析中[1-5]。目前，基于FCNs的众多网络模型很快被引入到遥感影像分类中[6-15]，但是要实现像素级别分类，需要构建逐像素类别标记的样本数据集，这也阻碍了FCNs在遥感影像分类中的实际应用。

为了提高图像语义分割精度，众多学者在深度学习网络中引入注意力机制，注意机制能够自动发现重要信息、增强语义，起到提高图像语义分割精度的效果。Jie[16]提出了基于通道注意力的网络SENet，通过自动学习各个特征通道的重要程度，注意力机制也已经被学者用于遥感影像的分类中，确定了较好的分类效果[17]。

基于上述分析，本文围绕卫星遥感影像地表覆盖信息提取，使用高分辨率卫星影像建立一套地表覆盖分类的像素级样本数据集，并设计一种基于全卷积神经网络的语义分割模型，实现地表覆盖信息高精度、自动化提取。

1　本文方法

1.1　网络框架

本文采用一种基于PSPNet[18]改进的语义分割网络PSP-SENet，该网络在特征图提取部分，采用SE模块[20]，关注特征信息量较大的通道，而在解码部分利

收稿日期：2021-08-18。
项目来源：自然资源部第三航测遥感院2020年科技项目（Y2020ZL07）；四川省自然资源科研项目（KJ-2020-4）。

用金字塔池化模块，融合多个尺度的特征图，主体网络框架如图1所示。

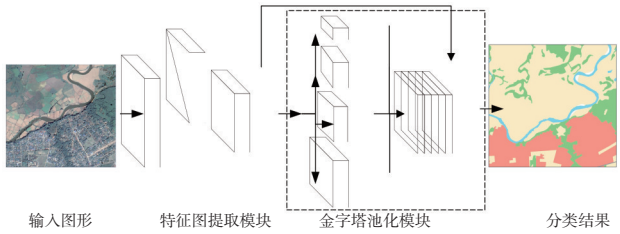


图1 PSP-SENet网络结构图

输入图像经过特征图提取模块提取到原始特征图，金字塔池化模块对原始特征图进行池化，金字塔池化模块是一个并行四层结构，4个池化分支的输出尺寸大小分别为1×1、2×2、3×3和6×6，再通过上采样，对4个池化分支结果进行融合，得到金字塔池化特征图，然后与原始特征图映射合并起来，再进行卷积，生成最终预测图（图2）。

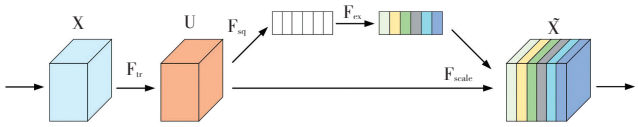


图2 SE模块单元

一个SE模块单元中，F_{tr}为传统的卷积结构，X和U分别为输入（$H'\times W'\times C'$）和输出（$H\times W\times C$），U=[u_1, u_2, ⋯, u_C]，即

$$u_c = v_c \times V = \sum_{s=1}^{C'} v_C^s \times x^s \tag{1}$$

式中，v_c为第c个卷积核的参数，$v_c=\left[v_c^1, v_c^2, \ldots, v_c^{C'}\right]$；$x^s$为第$s$个输入通道的特征图；$u_c$为第$c$个输出通道的特征图；$C'$为输入通道数量；$C$为输出通道数量。

增加SE模块后，首先对卷积层输出的特征图U，采用GPA进行压缩操作F_{sq}，即

$$z_c = F_{sq}(u_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{2}$$

然后对压缩后的Z，使用2个全连接层进行激励操作F_{ex}，第一个全连接把C个通道压缩成了C/r个通道来降低计算量，其中r是压缩的比例，后面跟了RELU函数，第二个全连接再恢复回C个通道，后面跟接Sigmoid函数，即

$$s = F_{ex}(z, \boldsymbol{W}) = \sigma(g(z, \boldsymbol{W}) = \sigma(\boldsymbol{W}_2\delta(\boldsymbol{W}_1 z)) \tag{3}$$

式中，δ为RELU函数；σ为Sigmoid函数；$\boldsymbol{W}_1$为$C/r\times C$的降维矩阵；$\boldsymbol{W}_2$为$C\times C/r$的升维矩阵。

通过压缩和激励操作，学习到各个通道间的关系，得到不同通道的权重，然后将SE模块计算出的各通道权重值分别与原特征图对应通道的二维矩阵相乘，得出的结果输出

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{4}$$

SE模块为网络模型提供了一种通道域注意力机制，让模型更加关注信息量较大的特征图，而抑制那些不重要特征图。

PSP-SENet网络结构如表1所示，表中卷积核大小用$n\times n$表示，卷积核大小之后的数字为卷积核个数，其他网络层后的数字代表输出通道数，全连接层后面中括号“[]”的数字是SE模块中2个全连接层的输出尺寸。

如表1所示，PSP-SENet首先对尺寸为480×480的输入影像进行特征图提取，特征图提取部分的结构取自文献[18]中的SE-ResNet-50，得到60×60特征图之后，使用平均池化，将特征图分别池化成1×1、2×2、3×3和6×6大小，然后再分别使用双线性插值上采样方法，上采样到60×60，并与之前的60×60特征图进行拼接，拼接之后再上采样到480×480的尺寸，最后连接到softmax激活层。

表1 PSP-SENet的网络结构和参数

输出尺寸	网络层类型
240×240	卷积层, 7×7, 64, 步长为2
120×120	最大池化层, 3×3, 64 步长为2
60×60	[卷积层, 1×1, 64 卷积层, 3×3, 64 卷积层, 1×1, 256 全连接层,[16, 256]] ×3
	[卷积层, 1×1, 128 卷积层, 3×3, 128 卷积层, 1×1, 512 全连接层,[32, 512]] ×4
1×1,2×2,3×3,6×6	平均池化层, 512
60×60	上采样, 512 拼接, 2 560 卷积层, 1×1, 512 卷积层, 3×3, 6 上采样, 6
480×480	softmax激活层

1.2 损失函数和精度评价

交叉熵损失函数是遥感影像分类任务中最常用的一种损失函数（L_cce），可从预测值（pr）与真值（gt）之间的交叉熵计算得到。

$$L_cce = -gt\cdot\log(pr) \tag{5}$$

本文引入dice系数构建损失函数，对PSP-SENet网络模型进行训练，dice损失函数（L_dice）定义为：

$$L_dice = 1 - F \tag{6}$$

其中，

$$F=(1+\beta^2)\frac{\text{precision}\cdot\text{recall}}{\beta^2\cdot\text{precision}+\text{recall}} \tag{7}$$

式中，β 为准确率和召回率之间的平衡系数，本文 β 取值为1。

同时，提供一种交叉熵损失与dice损失之和的损失函数（L_cce_dice）

$$L_cce_dice=L_cce+L_dice \tag{8}$$

精度评价中，F_1 和IoU（交并比）是最常用的语义分割评价指标[18]。

$$F_1=\frac{2\cdot\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}} \tag{9}$$

式中，precision和recall分别为准确率和召回率。

$$\text{precision}=\frac{TP}{TP+FP}\ ,\ \ \text{recall}=\frac{TP}{TP+FN} \tag{10}$$

$$\text{IoU}=\frac{TP}{TP+FP+FN} \tag{11}$$

式中，F_1 和IoU取值为0到1之间，值越大，分类精度越好。本文在多分类任务中，采用 F_1 均值和IoU均值评价分类模型的整体性能。

$$\text{M}F_1=\frac{1}{n}\sum_{i-1}^{n}F_1(i) \tag{12}$$

$$\text{MIoU}=\frac{1}{n}\sum_{i-1}^{n}\text{IoU}\,(i) \tag{13}$$

2　数据与实验

2.1　数　据

本文采用资源三号卫星影像，选取3个区域，制作了5种常见地表覆盖类型的样本数据，类别包括耕草、林地、建筑、道路、水体，影像覆盖区域位于东南亚，样本数据如图3所示。

训练样本数据采用资源三号全色与多光谱融合影像，地面分辨率为2.1 m，4个波段包含可见光（R、G、B）和近红外（IR）波段，样本影像覆盖的地表面积约3 400 km²，样本像素量达到千万级，未对样本进行增强处理前，样本数量约1 300个图像块（480×480），样本数量与国际上优秀的开源数据集差不多。样本数据中每个类别的像素数量如图4所示，5种地物覆盖类型的像素数量差异明显，耕草和林地的数量远远大于其他三类，道路的数量最少，训练样本数据存在数据不均衡问题。

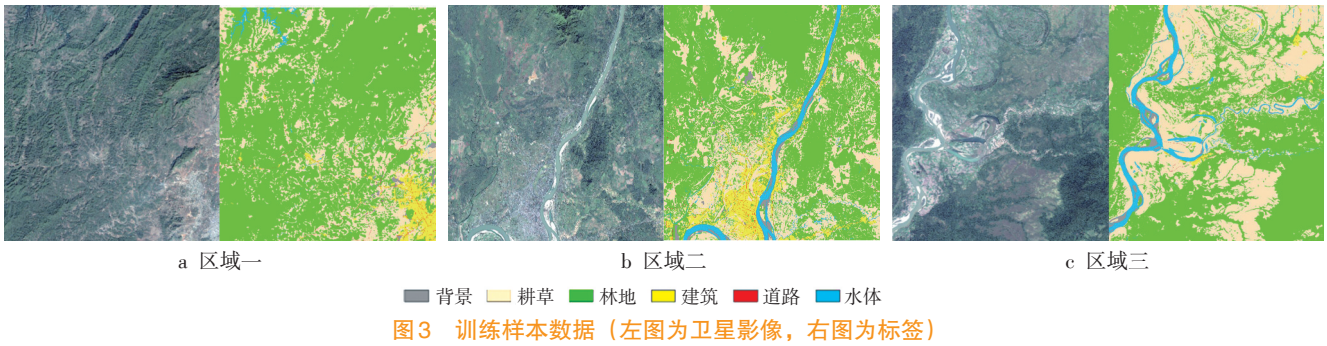


图3　训练样本数据（左图为卫星影像，右图为标签）

为了测试分类模型的性能，本文从东南亚地区的另外一景资源三号卫星影像中裁切出一块地物丰富的区域，测试影像覆盖面积约13 km²，对模型的分类结果进行精度测试，影像和参考标签如图5所示。

2.2　实验方案

本文方法基于TensorFlow和Keras深度学习开源框架，采用python语言编程实现。实验环境：操作系统为Ubuntu 16.04，GPU显卡为NVIDIA GTX1080TI。

为了测试PSP-SENet提取地表覆盖信息的性能，分别做如下实验与分析：①对比PSP-SENet和其他语义分割网络；②测试影像波段对PSP-SENet模型的影响；③测试ImageNet预训练模型对PSP-SENet的性能

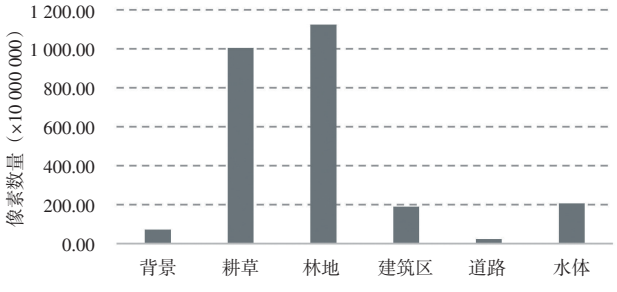


图4　训练样本中不同类别的像素数量

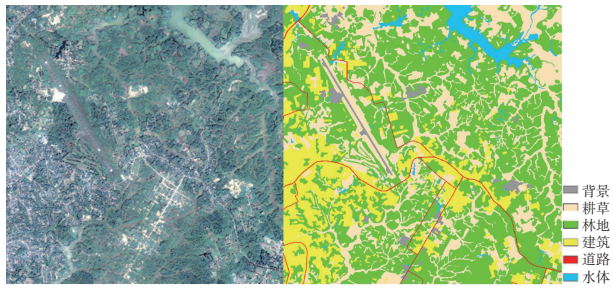


图5　测试数据(左图为卫星影像,右图为标签)

提升；④测试不同损失函数训练时，对PSP-SENet性能的影响。

2.3 实验结果

2.3.1 对比常用语义分割网络

Segnet和U-Net是经典的语义分割网络，近年来常被用于遥感影像的信息提取[20]，本文采用相同的训练和测试数据，分别对Segnet、U-Net和PSP-SENet进行训练和测试。图6为3种网络模型对测试影像的分类结果，表2给出了定量评价精度，耕草、林地、建筑、道路、水体单个类别的精度指标采用F_1值（本文其他表格中单个类别的精度指标均为F_1值），训练和测试影像的输入波段均为可见光波段（RGB）。

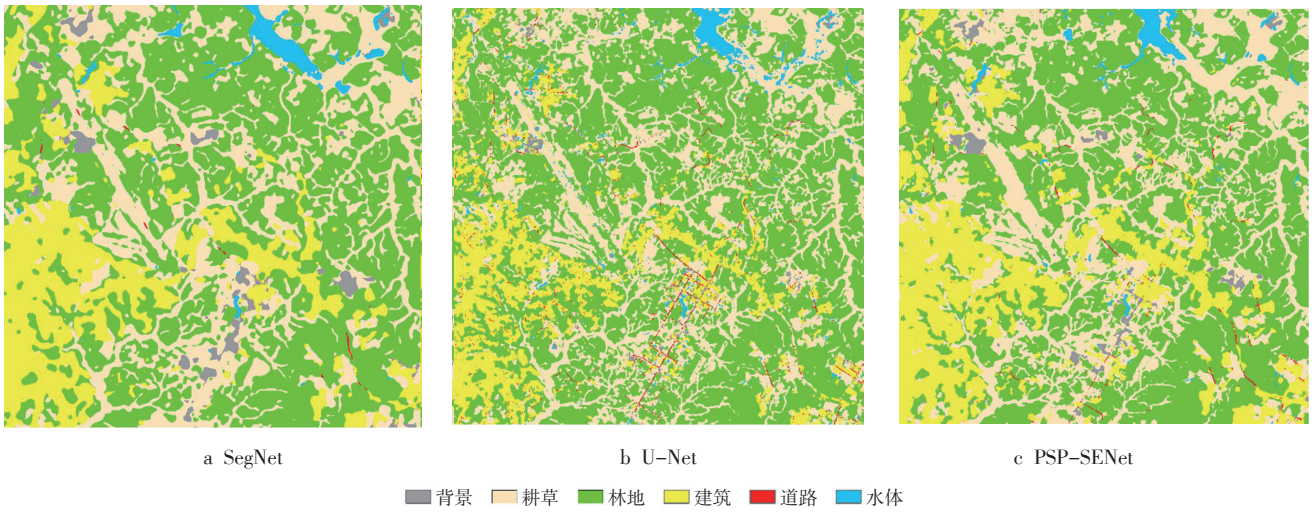


图6 不同网络模型的测试结果图

表2 不同网络模型的测试精度/%

网络	耕草	林地	建筑	道路	水体	MF_1	MIoU	OA
Segnet	73.2	84.7	79.4	5.7	61.5	60.9	48.9	78.8
U-Net	73.9	86.3	77.1	21.0	55.6	62.8	49.5	79.2
PSP-SENet	76.5	87.5	82.8	13.3	57.6	63.5	51.6	81.8

从图6可以看出，Segnet、U-Net和PSP-SENet均能够有效识别出耕草、林地、建筑，但是水体的提取精度只有60%左右，而且道路难以被识别出来。表2显示，3种网络模型对林地的识别精度最高，道路最低，PSP-SENet的OA、MF_1和MIoU均高于Segnet、U-Net，除了道路，PSP-SENet模型对单个地物类型的识别精度均高于其他2种网络模型。

2.3.2 比较影像的不同波段组合

卫星遥感影像的光谱范围通常包含可见光和近红外，为了充分挖掘卫星遥感影像波段丰富的优势，本文采用不同波段组合的卫星影像进行训练和测试，分析不同输入波段对地物的识别能力。

表3给出了3种波段组合的模型测试精度，IRRG（近红外、红色、绿色）波段组合和4波段（IRRGB）的MF_1相等、MIoU和OA相差仅为0.1%，RGB波段组合的精度指标远低于其他2种波段组合，MF_1比其他2种波段组合的低了8.5%。从不同地表类型的精度来看，IRRG和IRRGB2种波段组合对道路、水体的提取精度有很大提升，说明近红外波段在道路和水体识别中，有一定优势。

表3 不同输入波段的模型精度/%

波段组合	耕草	林地	建筑	道路	水体	MF_1	MIoU	OA
RGB	76.5	87.5	82.8	13.3	57.6	63.5	51.6	81.8
IRRG	78.2	87.9	83.3	21.7	88.8	72.0	61.2	83.2
IRRGB	78.3	88.2	83.3	22.3	87.7	72.0	61.1	83.1

2.3.3 采用ImageNet预训练模型初始化

本文采用在ImageNet数据集上训练后的模型，对PSP-SENet特征图提取部分进行初始化，实现网络模型参数迁移学习，即将从2012 ILSVRC ImageNet数据集上训练得到的SE-ResNet-50模型，对PSP-SENet特征提取部分对应的网络层参数进行初始化，然后在遥感数据集上进行训练，得到最终的分类模型。由于ImageNet数据集只有RGB波段，本文采用RGB波段的预训练模型分别对RGB和IRRG 2种影像波段组合的模型进行初始化，表4给出了采用预训练模型初始化后模型的精度表3对比，可以看出，采用ImageNet预训练模型初始化后的模型，分类精度有了较大提升，RGB波段模型的MF_1、MIoU分别提高了9.1%、8.8%，IRRG波段模型的MF_1、MIoU分别提高了3.7%、3.3%，由此可见，对于遥感影像小样本，采用ImageNet预训练模型对语义分割模型进行参数初始

化，能够有效提升模型分类精度。

表4　ImageNet预训练模型初始化的模型精度/%

波段组合	耕草	林地	建筑	道路	水体	MF_1	MIoU	OA
RGB	79.0	88.3	83.9	31.8	80.0	72.6	60.4	83.5
IRRG	80.3	88.8	84.5	36.3	88.6	75.7	64.4	84.8

2.3.4　比较不同损失函数

由于训练样本中不同类别之间的数据量不均衡，模型难以有效识别数据量的类别，这是遥感影像模型训练中常见的难题。本文制作的训练样本数据中，道路样本数量远远小于其他类别的样本数量，为了提高道路的识别能力，本文在1.3.3节的基础上，将dice系数引入到损失函数中，采用2种不同dice损失函数对模型进行训练。表5给出了不同损失函数模型的测试精度，其中，为了获取更好的分类效果，影像输入波段为IRRG，初始化采用了ImageNet预训练模型。

对比表5与表4，采用交叉熵和*L*_dice损失函数训练的模型，整体精度相差不大；从单个类别的F_1值来看，采用*L*_dice训练的模型，道路的提取精度提升了14.1%，但是其他类别的F_1值均有所下降。

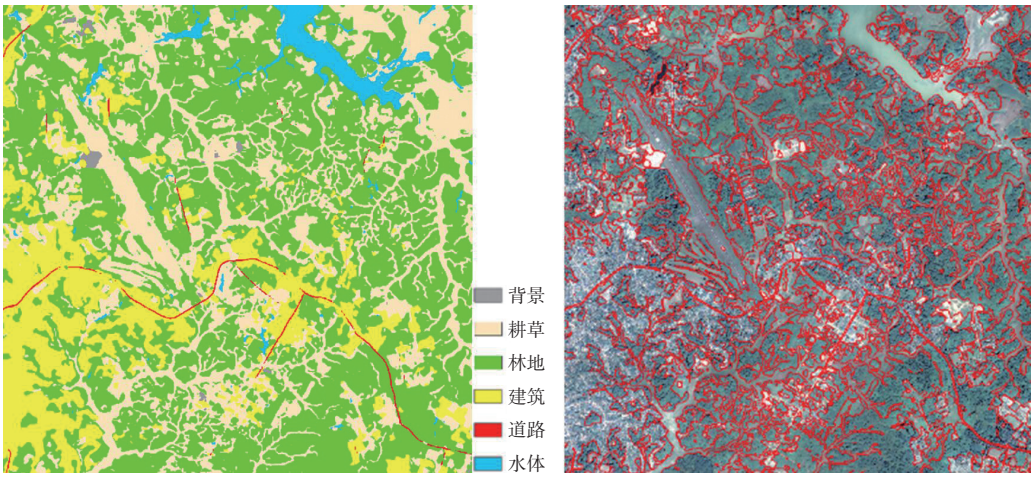
从表5可以看出，采用交叉熵与dice之和作为损失函数（*L*_cce_dice），模型在不降低其他类别识别精度的情况下，提升了对道路的识别精度，整体精度（MF_1、MIoU和OA）也有提升，说明对于类别不均衡的训练样本数据，可以采用交叉熵与dice之和作为损失函数，提高模型的分类精度。

表5　不同损失函数模型的测试精度/%

损失函数	耕草	林地	建筑	道路	水体	MF_1	MIoU	OA
*L*_dice	77.5	85.0	82.1	50.4	87.0	76.4	63.5	81.9
*L*_cce_dice	80.6	89.1	84.7	50.5	88.2	78.6	66.8	85.0

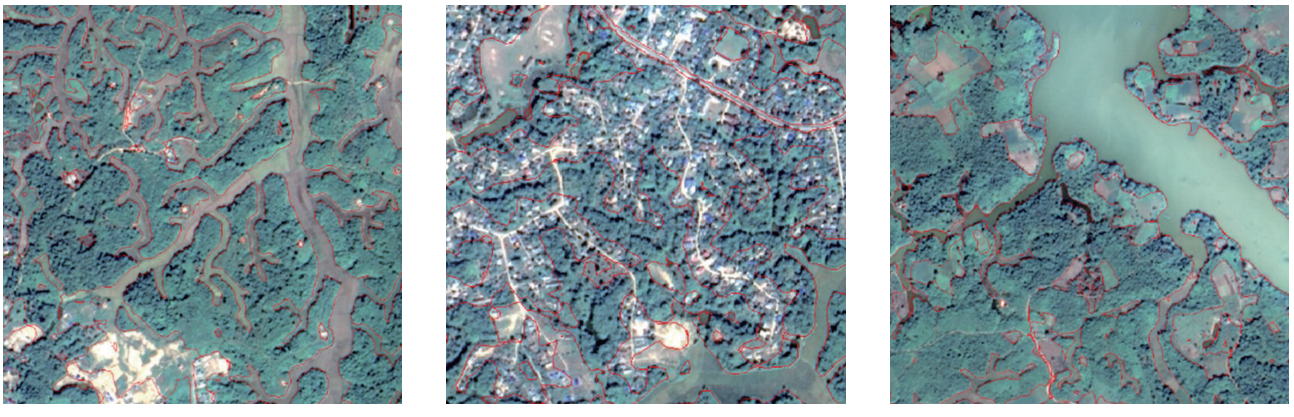
2.3.5　最优分类结果

PSP-SENet最优测试模型的分类图斑和图斑边界如图7所示，该结果采用的模型为表5中采用*L*_cce_dice损失函数训练的模型，从图中看出，PSP-SENet能够实现高分辨率卫星遥感影像的地表覆盖信息提取，图斑边界的细节如图8所示，耕草、林地、水体的图斑边界与影像中地物边界套合较好，房



a 分类图斑　　b 图斑边界

图7　PSP-SENet最优测试模型的分类图（红色线条为分类图斑的边界）



a 区域I　　b 区域II　　c 区域III

图8　PSP-SENet模型提取的图斑细节（红色线条为分类图斑的边界）

屋建筑和道路受到其他地类的遮挡影响，图斑边界与影像中地物边界套合存在一定误差，但是能够提取出建筑房屋区域和道路的整体轮廓。

3 结 语

本文围绕高分辨率卫星遥感影像地表覆盖信息提取，建立了一套涵盖耕草、林地、建筑、道路、水体的像素级分类数据集，并采用注意力增强与多尺度特征融合神经网络实现卫星影像语义分割，通过对资源三号卫星影像的测试，结果表明：①使用近红外、红色、绿色3个波段组合的影像，对地表覆盖的识别能力高于红色、绿色、蓝色3个波段组合的影像；②使用在大规模数据集（ImageNet）上预训练的模型对语义分割网络特征图提取部分进行初始化，能够提升网络在遥感影像小规模训练样本数据集上的分类精度；③采用交叉熵与dice系数之和作为损失函数，可以解决在类别不均衡样本数据集上训练模型的难题；④本文建立的遥感影像分类数据集和分类方法能够有效地从高分辨率卫星影像上提取地表覆盖信息，可为全球地表覆盖制图、生态环境保护等工作提供技术支持。

参考文献

[1] Blaschke T. Object Based Image Analysis for Remote Sensing[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2010,65(1): 2-16

[2] Yang Y,Newsam S. Geographic Image Retrieval Using Local Invariant Features[J]. IEEE Transactions on Geoscience and Remote Sensing,2013,51(2): 818-832

[3] LI E, Femiani J, XU S, et al. Robust Rooftop Extraction from Visible Band Images Using Higher Order CRF[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015,53(8): 4 483-4 495

[4] Melgani F,Bruzzone L. Classification of Hyperspectral Remote Sensing Images with Support Vector Machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004,42(8): 1 778-1 790

[5] Zhu X X, Tuia D, Mou L, et al. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources[J]. IEEE Geoscience and Remote Sensing Magazine, 2017,5(4): 8-36

[6] Nogueira K,Penatti O A B,Santos J A d. Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classication[J]. Pattern Recognition,2016,61(7): 539-556

[7] Tong X Y, Lu Q K, Xia G S, et al. Large-scale Land Cover Classification in GaoFen-2 Satellite Imagery[C]// IEEE International Geoscience and Remote Sensing Symposium Igarss,2018

[8] Längkvist M, Kiselev A, Alirezaie M, et al. Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks[J]. Remote Sensing, 2016, 8(4): 329-349

[9] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015,521(3): 436-444

[10] Guo Z, Shao X, Xu Y, et al. Identification of Village Building Via Google Earth Images and Aupervised Machine Learning Methods[J]. Remote Sensing,2016,8(4): 271-285

[11] Guo Z, Chen Q, Wu G, et al. Village Building Identification Based on Ensemble Convolutional Neural Networks[J]. Sensors,2017,17(11): 2 487-2 508

[12] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition,2015

[13] Wu G, Shao X, Guo Z, et al. Automatic Building Segmentation of Aerial Imagery Using Multi-constraint Fully Convolutional Networks[J]. Remote Sensing,2018,10(3): 407-424

[14] Zhang W, Huang H, Schmitz M, et al. Effective Fusion of Multi-modal Remote Sensing Data in a Fully Convolutional Network for Semantic Labeling[J]. Remote Sensing, 2018,10(52): 1-14

[15] 伍广明,陈奇,Shibasaki R,等. 基于U型卷积神经网络的航空影像建筑物检测[J]. 测绘学报,2018,47(6): 864-872

[16] Hu J, Shen L, Sun G. Squeeze and Excitation Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition,2018

[17] 李道纪,郭海涛,卢俊,等. 遥感影像地物分类多注意力融和U型网络法[J]. 测绘学报,2020,49(8): 1 051-1 064

[18] Maggiori E, Tarabalka Y, Charpiat G, et al. Convolutional Neural Networks for Large-scale Remote-sensing Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing,2017,55(10): 645-657

[19] 季顺平,魏世清. 遥感影像建筑物提取的卷积神经元网络与开源数据集方法[J]. 测绘学报,2019,48(4): 448-459

[20] 许慧敏,齐华,南轲,等. 结合nDSM的高分辨率遥感影像深度学习分类方法[J]. 测绘通报,2019(8): 63-67

第一作者简介：罗想（1988—），工程师，主要从事遥感信息提取与地理信息技术应用。