引文格式：孙凌辉,赵丽科,李琛,等. 融入CBAM的Res-UNet高分辨率遥感影像语义分割模型[J]. 地理空间信息, 2024, 22(2): 68-70. doi:10.3969/j.issn. 1672-4623.2024.02.016

融入CBAM的Res-UNet高分辨率遥感影像语义分割模型

孙凌辉[1]，赵丽科[1]，李　琛[1]，成子怡[1]
（1. 河南工业大学 信息科学与工程学院，河南 郑州 450001）

摘　要：针对现有语义分割方法处理复杂遥感影像细节特征识别能力差、信息丢失等问题，提出一种融合注意力机制的遥感影像语义分割网络模型。模型主干网络采用编码器-解码器架构的U-Net模型，为了缓解梯度和网络退化问题，将残差结构嵌入到主干网络中；同时融入通道、空间注意力模块，兼顾影像的细节特征和模型鲁棒性。在ISPRS Potsdam数据集上进行分析验证，实验结果表明，在去除“噪声”、地物边缘“平滑”、细窄地物“连续”、细小目标分割等方面，融入CBAM模块的Res-UNet语义分割精度要优于传统网络模型。

关键词：Res-UNet；注意力机制；神经网络；语义分割
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）02-0068-03

Res-UNet High-resolution Remote Sensing Image Semantic Segmentation Model Integrated into CBAM

SUN Linghui[1], ZHAO Like[1], LI Chen[1], CHENG Ziyi[1]
(1. College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: Aiming at the problems of poor recognition ability and information loss of the existing semantic segmentation methods for complex remote sensing images, we proposed a semantic segmentation network model of remote sensing images based on attention mechanism. We adopted the U-NET model with encoder-decoder architecture in the model backbone network. In order to alleviate the gradient and network degradation, we embedded the residual structure into the backbone network, and integrated the channel and spatial attention modules to take into account the detail features of image and model robustness. We analyzed and verified this model on ISPRS Potsdam dataset. The experimental results show that the accuracy of Res-UNet semantic segmentation with CBAM module is better than that of traditional network model in terms of removing noise, smoothing edge of ground objects, continuing narrow ground objects and small object segmentation.

Key words: Res-UNet, attention mechanism, neural network, semantic segmentation

遥感图像具有的数据量大、空间相关性强、分辨率高的特点，适合采用卷积神经网络的方法开展语义分割的方法提取地物信息。本文根据其特点，采用融入残差模块的U-Net网络作为主干，避免模型发生退化，同时融入通道和空间注意力模块，增强地物的细节特征，改善高分辨率遥感影像语义分割的精度[1-3]。

1　融入CBAM的Res-UNet遥感影像语义分割模型

1.1　整体结构框架

鉴于采用编解码结构的UNet在语义分割中表现出来的优势，本文模型主干选用UNet网络结构，同时在编码器部分融入残差连接模块与通道空间注意力模块[4]。网络的主干由三部分组成，分别为获取地物特征的编码器、对影像进行分类的解码器、连接底层信息与高层信息的跳跃连接，在此基础上还加入了抑制网络退化的残差结构和提高网络聚焦特征的CBAM注意力模块，改善细节信息的特征抽取能力。

1.2　Res-UNet结构

编码器-解码器结构是研究语义分割方面普遍使用的一种结构，由编码器、过渡层和解码器共同组成。编码器负责通过下采样降低图像维度；过渡层存在于编码器和解码器之间，提取并传递高维特征信息；解码器负责通过上采样将特征图还原为原始维度。U-Net网络是基于编解码结构的经典模型，由编码结构、解码结构和跳跃连接构成，跳跃连接的本质就是将下采样过程中获得的特征图像融合到上采样中，所用的融合方式就是将特征图像通道叠加。

收稿日期：2022-09-29。
项目来源：国家自然科学基金资助项目（41901276）；河南工业大学自然科学创新基金资助项目（2021ZKCJ18）。
第一作者简介：孙凌辉（2002—），研究方向为深度学习、遥感图像处理，E-mail：kuaixi611011@foxmail.com。

由于在网络层数增加的过程中，预测效果越来越差，梯度消失和梯度爆炸产生了网络退化问题，“残差结构”跳过一些神经元节点，跨层相连，减弱各个网络层之间的强联系，提高网络训练效率。鉴于残差结构的优点，将残差结构融入到U-Net网络中，以达到提高精度的目的。

1.3　CBAM注意力模块

由于遥感影像的复杂性，语义分割经常会受到模糊或者是杂糅等区域的影响且影像本身的时相可能不同，以及数据来源、阴影等变化导致使用传统方法不能很好地进行各类地物的分割。因此在嵌入残差结构的U-Net模型中添加CBAM注意力模块以强化地物的细节特征，提高遥感影像语义分割的精度，如图1为CBAM模块的示意图[5]。

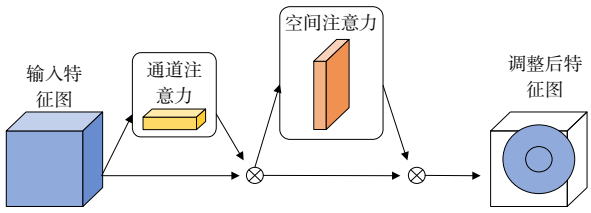


图1　CBAM示意图

CBAM凭借自身是一个轻量级通用注意力模块的特点，可以无缝集成到各种神经网络架构中。首先对于语义分割来说，有多个通道的特征输出，有些通道的特征对最终的分类结果影响很大，有必要使注意力聚集在这些通道中，具体做法是通过全局池化，在通道注意力模块中使用最大池化和均值池化，得到2个1×1×C的特征图，然后通过MLP进行非线性的特征变化，进行像素级相加后经激活函数激活得到对应的注意力权重，表达式为：

$$M_c(F)=\sigma(\mathrm{MLP}(\mathrm{AvgPool}(F))+\mathrm{MLP}(\mathrm{MaxPool}(F))) \quad (1)$$

对于空间注意力，它是使网络模型更集中于具有空间形状的特征上，它将通道注意力特征图作为输入，分别进行最大池化和均值池化，生成2个2D特征图，之后对其进行级联，经卷积层降维至单通道，最后由激活函数激活得到注意力权重，表达式为：

$$M_s(F)=\sigma(f^{7\times7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)])) \quad (2)$$

CBAM按照通道、空间域2个独立的维度对遥感影像进行注意力“注入”，然后将已获得“注意力”的图像乘以读取的特征图像进行图像特征细化。CBAM对特征图 $F\in R^{C\times H\times W}$ 分别进行通道、空间注意力计算，会分别得到一维的和二维的通道和注意力权重 $M_c\in R^{C\times1\times1}$ 和 $M_s\in R^{1\times H\times W}$，之后对特征图进行加权计算，表达式为（3）。

$$\left.\begin{aligned}F'&=M_c(F)\otimes F\\F'&=M_s(F')\otimes F'\end{aligned}\right\} \quad (3)$$

在公式（3）中，$\otimes$ 为矩阵与所对应的元素依次相乘，在这个过程中，按照对应的维度，注意力的值会进行复制，通道和空间注意力权重分别以通道、空间维度进行复制。最后的 F'' 为输出的特征图。CBAM注意力模块的嵌入，可以有效提升地物的细节特征，改善语义分割的精度。

2　实验结果与分析

为了验证本文方法的可靠性，选用ISPRS Potsdam[6]数据集开展实验分析，Potsdam遥感数据集一共包含38幅高分辨率遥感影像，影像的大小为6 000×6 000像素，其空间分辨率为5 cm，将每张影像划分为6类，分别是建筑物（蓝色）、非渗透表面（白色）、灌木植被（青色）、车辆（黄色）、树木（绿色）以及杂波层（红色）。本实验选取17幅图像用作训练、5幅图像用作验证、其余图像用作测试。

为了客观且全面地评估网络模型进行遥感影像语义分割的效果，采用了遥感影像语义分割比较常用的评价指标客观地评估语义分割效果[7]，使用的精度评价指标为精确度（Precision）、像素准确率（PA）、召回率（Recall）、特异度（Specificity）、均交并比（MIoU）频权交互比（FWIoU）以及F1分数（F1），各评价指标的数学表达为：

$$\mathrm{Precision}=\frac{TP}{TP+FP} \quad (4)$$

$$\mathrm{Pixel-accuracy}=\frac{(TP+TN)}{(TP+TN+FP+TN)} \quad (5)$$

$$\mathrm{Recall}=\frac{TP}{TP+FN} \quad (6)$$

$$\mathrm{Specificity}=\frac{TN}{TN+FP} \quad (7)$$

$$\mathrm{MIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{FN+FP+TP} \quad (8)$$

$$\mathrm{FWIoU}=\sum_{i=0}^{k}\left[\left(\frac{TP+FN}{TP+FP+TN+FN}\right)\times\left(\frac{TP}{TP+FP+FN}\right)\right] \quad (9)$$

$$\mathrm{F1}=\frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \quad (10)$$

式中，TP 为正样本预测正确；TN 为负样本预测正确；FP 为将负样本预测错误；FN 为将正样本预测错误。将2种模型的预测结果与原始的标签图像对比后分别计算。

Res-UNet网络模型与融入了CBAM注意力机制的

Res-UNet神经网络模型在精确度、像素准确率、F1、MIoU以及FWIoU 5个方面的整体平均值如表1所示。

表1 整体预测结果评估指标

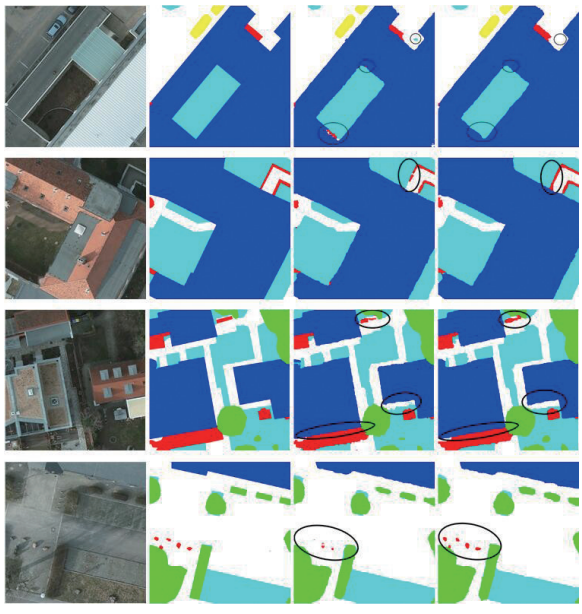
模型	Precision	PA	MIoU	FWIoU	F1
未融入CBAM	0.945 6	0.980 3	0.948 1	0.961 6	0.964 5
融入CBAM后	0.949 2	0.983 3	0.957 3	0.969 7	0.969 4

Res-UNet网络模型与融入了注意力机制的Res-UNet神经网络模型在精确度、MIoU、特异度以及召回率4个方面的整类别平均值如表2所示。

表2 类别预测结果评估指标

模型	精度评价	杂波层	车辆	灌木植被	树木	建筑物	非渗透表面
融入后	Precision	0.956 2	0.923 5	0.978 7	0.978 8	0.992 2	0.981 2
未融入		0.949 9	0.924 6	0.977 0	0.969 1	0.988 6	0.979 2
融入后	MIoU	0.945 2	0.945 2	0.971 0	0.974 5	0.990 2	0.974 3
未融入		0.932 8	0.943 5	0.964 9	0.969 6	0.987 5	0.970 5
融入后	Specificity	0.997 8	0.998 0	0.996 1	0.997 0	0.996 2	0.990 9
未融入		0.997 5	0.998 1	0.995 9	0.995 8	0.994 4	0.989 9
融入后	Recall	0.934 1	0.964 7	0.971 2	0.975 5	0.994 6	0.983 7
未融入		0.914 5	0.959 6	0.962 1	0.976 2	0.994 5	0.980 5
融入后	F1	0.945 0	0.943 6	0.974 9	0.977 2	0.993 4	0.982 5
未融入		0.931 9	0.941 8	0.969 5	0.972 6	0.991 5	0.979 8

从表1和表2中可以看出，通道与空间注意力的融入可以明显改善遥感影像语义分割的精度，Precision、PA、Precision、Recall、F1、MIoU及FWIoU各项指标均有所提升，通道与空间注意力模块能够强化遥感影像中的细节信息，提高语义分割的精度，表明本文方法的有效性。



a 原始图像 b 真值标签 c 未融入CBAM d 融入CBAM

图2 实验结果展示

为了定性观察融入CBAM模块后遥感影像语义分割的效果，如图2展示了典型的遥感图像分割结果，图中a为原始遥感影像，b为真值标签，c为未融入CBAM的分割结果，d融入CBAM的分割结果。比较第一行c与d的分割结果，影像中孤立的噪声被去除；第二行展示了遥感影像中细窄地物的分割结果，加入注意力模块后，对细节的识别能力有所提升，分割结果更为连续；第三行的图像则表明了未融入CBAM时，结果多呈“锯齿状”，融入注意力模块后会有明显改善；第四行展示了对面积较小地物的分割结果，融入注意力模块后，细节信息补全，提高影像语义分割的精度。

3 结 论

本文提出一种融入通道和空间注意力模块的Res-UNet模型，模型主干采用编解码结构的UNet模型，将残差结构嵌入到编码部分，有效避免模型退化问题，将CBAM模块连接到网络中，挖掘地物间不同特征信息，增强地物的细节，提升遥感影像语义分割精度。在ISPRS Potsdam数据集上进行分析验证，该模型能够强化地物的细节特征，有效去除噪声、改善地物边缘的分割结果，提升语义分割的精度，验证了本文提出方法的鲁棒性和科学性。但本文对地物边界提取仍有待进一步优化，后续考虑增加边缘感知模块，改善地物边缘特征。

参考文献

[1] 吴泽康,赵珊,李宏伟,等. 遥感图像语义分割空间全局上下文信息网络[J]. 浙江大学学报(工学版),2022,56(4):765-802

[2] 韩彬彬,张月婷,潘宗序,等. 残差密集空间金字塔网络的城市遥感图像分割[J]. 中国图象图形学报,2020,25(12):2 656-2 664

[3] 姚建华,吴加敏,杨勇,等. 全卷积神经网络下的多光谱遥感影像分割[J]. 中国图象图形学报,2020,25(1):180-192

[4] 黄旭,张世义,李军. 图像分割技术研究综述[J]. 装备机械,2021(2):6-9

[5] 张哲晗,方薇,杜丽丽,等. 基于编码-解码卷积神经网络的遥感图像语义分割[J]. 光学学报,2020,40(3):31-35

[6] 陆妍如,毛辉辉,贺琰,等. 基于Deeplab v3+的高分辨率遥感影像地物分类研究[J]. 地理空间信息,2022,20(6):1-6

[7] 苏耀,于濂,周伟. GCAT-U-Net嵌入全局坐标注意力机制的遥感地块分割网络[J]. 计算机测量与控制,2022,30(2):222-228