多注意力融合的环高原湖泊遥感影像分割

何自芬[1],史本杰[1],张印辉[1],李素敏[2]
(1. 昆明理工大学机电工程学院,云南昆明 650500;2. 昆明理工大学国土资源工程学院,云南昆明 650500)

摘 要: 环高原湖泊区域土地类别监测为湖泊生态保护和土地资源规划提供了决策依据. 针对此区域遥感影像中河流、建筑物及植被目标分布零散、尺度不均导致分割精度较低的问题,设计了融合类别与多尺度注意力的遥感语义分割网络. 该网络采用编码-解码的端到端结构并以深度残差神经网络为基础构建类别与多尺度注意力模块. 类别注意力对网络特征层进行初步分类与空间信息过滤,有利于网络关注类别信息以降低像素分类误差;多尺度注意力将混合域注意力和多尺度特征进行融合,为不同尺度特征建立上下文联系,改善分布零散小尺度目标固有的分割消弥问题. 实验结果表明,在建立的环滇池区域遥感影像语义分割数据集上,本文设计的注意力融合语义分割网络测试精度在平均交并比和平均像素精度指标下分别达到77.4%和86.3%. 从整体分割效果来看,融合类别与多尺度注意力分割网络在一定程度上解决了分布零散小尺度目标区域的分割消弥问题,对环高原湖泊区域精准监测和科学规划提供了有效依据.

关键词: 语义分割;深度学习;高原湖泊;注意力机制;多尺度;遥感影像
基金项目: 国家自然科学基金(No.62171206,No.62061022,No.61761024)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2023)04-0885-11
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220085

Remote Sensing Image Segmentation of Around Plateau Lakes Based on Multi-Attention Fusion

HE Zi-fen[1], SHI Ben-jie[1], ZHANG Yin-hui[1], LI Su-min[2]
(1. *College of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, Yunnan* 650500, *China*;
2. *School of Land and Resources Engineering, Kunming University of Science and Technology, Kunming, Yunnan* 650500, *China*)

Abstract: Land category monitoring in lake region around plateau provides decision-making basis for lake ecological protection and land resource planning. Aiming at the problem of low segmentation accuracy caused by scattered distribution and uneven scale of rivers, buildings and vegetation in remote sensing images of this region, a remote sensing semantic segmentation network integrating category and multi-scale attention is designed. The network adopts encoding-decoding end-to-end structure and constructs class and multi-scale attention modules based on depth residual neural network. Category attention makes a preliminary classification and spatial information filtering for the network feature layer, which is beneficial for the network to pay attention to category information and reduce pixel classification error; multi-scale attention combines mixed domain attention with multi-scale features, establishes context connection for different scale features, and improves the inherent segmentation and elimination problem of scattered small-scale targets. Experiments are performed on the semantic segmentation data set of remote sensing images around Dianchi Lake, and the test accuracy of the attention fusion semantic segmentation network designed in this paper reaches 77.4% and 86.3% under the average intersection ratio and average pixel accuracy index, respectively. From the overall segmentation effect, the fusion category and multi-scale attention segmentation network solve the segmentation and elimination problem of scattered small-scale target areas to a certain extent, and provide an effective basis for accurate monitoring and scientific planning of lakes around plateau.

Key words: semantic segmentation; deep learning; plateau lake; attention mechanism; multi-scale; remote sensing image
Foundation Item(s): National Natural Science Foundation of China (No.62171206, No.62061022, No.61761024)

收稿日期:2022-01-17;修回日期:2022-04-10;责任编辑:朱梅玉

1 引言

高原湖泊作为地球生态系统重要组成部分,具备独特地理特性的同时也蕴藏稀缺的自然资源,为研究地球自然环境变化与生态系统科学提供了宝贵资料,而对高原湖泊进行有效保护是研究高原湖泊生态系统的重要一环[1,2]. 目前,高原湖泊监测除实地调研外还依靠大量遥感影像进行数据分析. 相对于实地调研,通过遥感影像数据分析能有效掌握高原湖泊在时空上的全局演化信息[3],对环湖泊区域土地进行类别划分,可以为环湖泊区域土地资源优化提供重要参考依据.

作为提取遥感影像信息的重要手段,遥感影像分割能获取目标的轮廓和类别信息,其分割方法主要分为传统分割算法和基于卷积神经网络的深度学习算法. 传统遥感影像分割主要采用图像阈值分割[4]、像素聚类和多光谱影像融合[5]等算法,传统算法分割过程需要人工进行流程设计和参数调整,且单一分割方法在复杂遥感场景下无法实现对密集语义目标区域的轮廓和类别信息进行有效提取.

基于深度学习的语义分割算法通过卷积神经网络自适应提取图像特征和语义信息进行像素分类[6],减少人为流程设计和先验参数设置,极大提高了分割效率以及网络模型对不同场景的泛化能力. 在语义分割网络模型设计方面,U-Net[7]采用与FCN[8]相同的编码-解码网络结构,并将浅层特征叠加到较深特征层中实现远距离特征信息融合,以缓解网络深层特征细节信息丢失问题. PSPNet[9]采用金字塔池化模块获取多尺度特征信息,提高网络模型对尺度不均目标的分割性能. DeepLabV3+[10]中使用空洞空间卷积池化金字塔增强网络获取全局信息的能力,通过空洞卷积扩大感受野实现单个像素关联更大范围的上下文信息. 在遥感影像语义分割领域中,文献[11]搭建了基于Segnet[12]改进的语义分割网络,通过卷积与池化索引来融合网络浅层和深层特征. 文献[13]中通过图卷积神经网络和独立循环神经网络挖掘分割目标区域上下文信息. 文献[14]通过反卷积融合结构获取多尺度信息,采用全连接条件随机场引入空间信息上下文,提高网络模型的处理精度.

上述基于扩大上下文信息关联的语义分割算法在一定程度上能提高网络对全局信息的获取,但缺乏对目标区域的集中关注导致信息冗余从而影响分割精度. 针对上述问题,注意力机制能自适应对上下文信息建立有效联系并减少信息冗余,通过自主分配注意力权重来关联网络前后特征层的空间与通道信息. 在注意力机制构建方面,空间与通道压缩激励注意力机制(spatial and channel Sequeeze & Excitation,scSE)[15]在压缩激励注意力(Squeeze and Excitation,SE)[16]基础上增加空间注意力,同时在特征空间与通道上提取注意力权重来获得更精细的上下文信息. 卷积注意力模块(Convolutional Block Attention Module,CBAM)[17]将通道和空间注意力进行串联,并使用全局平均池化和全局最大池化对并行通道聚合信息进行特征映射,增强了注意力模块的表征能力. 有效通道注意力(Efficient Channel Attention,ECA)[18]提出一种通道不降维的信息交互策略,使用一维卷积分配通道注意力权重以减小通道降维导致的信息丢失. 此外,非局部自注意力[19]也被广泛引入到语义分割任务当中,其中,CCNet[20]对非局部自注意力模块进行注意力结构优化,通过复用交叉自注意力模块替代原始非局部自注意力实现降低计算复杂度. 极化自注意力[21](Polarized Self-Attention,PSA)将特征层的空间和通道分别进行自注意力关注,同时生成自注意力索引信息获得密集上下文联系. 综上所述,注意力机制通过自适应提取注意力权重来联系网络上下文信息,实现特征提取过程对重要目标信息的关注. 但单一尺度注意力关联策略严重制约遥感影像中分布零散、尺度不均区域的语义分割性能.

本文建立环滇池区域遥感影像语义分割数据集(Semantic segmentation data set of remote sensing around Dianchi Lake,SDL),人工标注建筑用地、植被、河流湖泊、农业用地及其他用地语义区域. 针对SDL遥感影像中水系、建筑物及植被等目标尺度不均、特征相似及分布零散问题,提出融合类别与多尺度注意力网络(Fusion Category and Multi-scale Attention Network,FCMA-Net). 类别注意模块通过自适应分配类别注意力权重建立空间上下文类别信息联系,实现类别关注与空间信息过滤以降低目标特征相似导致的像素分类误差. 多尺度注意力将混合域注意力融合到多尺度特征信息中,获取不同尺度区域上下文信息来加强网络对全局信息的获取能力,从而改善分布尺度不均、分布零散导致的分割消弥问题.

2 FCMANet模型架构

遥感影像语义分割作为复杂分割场景之一,环高原湖泊较其他地形的地物分割都存在分割目标分布零散、尺度不均等分割难题,此外环高原湖泊中湖泊地块与非湖泊区域河流地块存在特征高度相似导致的难分割问题,为有效提高遥感影像分割效果,本文构建了FCMANet模型. 首先,层次较深且复杂度高的网络模型具备强表征能力,从而能拟合更复杂的非线性问题. FCMANet网络模型选择深度残差神经网络[22]作为主干网络提取图像深层语义特征. 其次,为解决环高原湖泊遥感影像中存在地块特征高度相似导致的分割难题,

在主干网络中穿插类别注意模块,实现提取类别注意力权重引导网络像素分类与空间信息过滤,通过为特征分类提供指导进而减少像素分类错误,以提高河流与湖泊地块的分割精度. 最后,为解决目标分割区域分布零散、尺度不均导致的分割消弥问题,网络融合多尺度特征信息和混合注意力以增强网络对目标尺度变化的鲁棒性,改善网络分割性能. FCMANet 网络整体架构如图 1 所示.

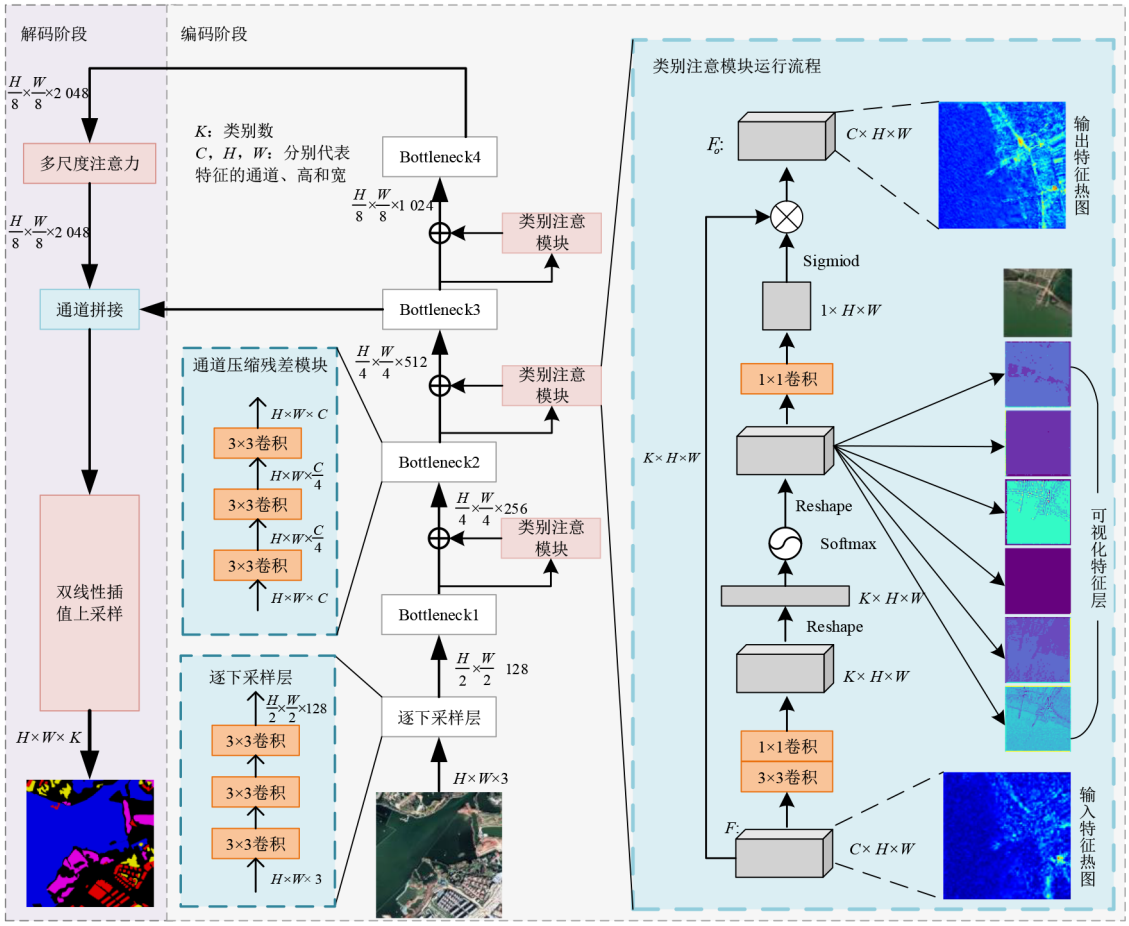


图 1　FCMANet 网络架构

2.1　网络主干设计

FCMANet 采用编码解码语义分割结构. 首先在编码部分应用含瓶颈结构的通道压缩残差模块,如图 1 中通道压缩残差模块所示,将特征通道数量收缩至原来的四分之一以生成密集特征信息;再使用 3×3 卷积和 1×1 卷积对密集特征采样与空间信息映射,达到减少参数量并提升信息密度的目的. 此外将 ResNet 中 7×7 卷积的单下采样层替换为 3 个 3×3 卷积形成逐下采样层,使得同一特征的单个像素感受野扩大,以此获取网络浅层中丰富的轮廓、位置等细节信息,该结构如图 1 中逐下采样层所示. 其次,在主干网络 Bottleneck 结构间穿插类别注意力模块,在网络特提取特征过程中实现类别引导. 在网络解码部分,在 Bottleneck4 后添加多尺度注意力模块,将 Bottleneck4 输出为 2 048 个通道的特征层拆分为四个 512 通道特征层. 对每个特征层分别使用不同大小的卷积核进行采样以获取多尺度特征区域,并对多尺度特征添加双通道串联注意力(Two-Channel Tandem Attention,TCTA),将 TCTA 所有输出特征层在通道上进行叠加,进而建立多尺度特征区域上下文信息联系. 然后将多尺度注意力输出特征与 Bottleneck3 输出特征进行通道拼接来融合深层与浅层特征信息,利用浅层特征来丰富网络最终输出层的细节信息,进一步改善特征信息丢失导致的分割消弥问题,再利用双线性插值上采样解码,最终形成编码解码结构语义分割网络.

2.2　类别注意模块

采用深度学习方法对图像进行语义分割时,特征层缺少空间或通道的上下文信息联系会限制网络分割性能. 注意力机制通过自适应提取特征层间的注意力权重来关联网络上下文,但大多注意力在提取注意力权重时并未对关键信息过滤,造成信息冗余. 本文主要通过在网络主干中穿插类别注意模块来过滤网络中的

冗余信息,该注意力首先将输入的特征信息进行通道压缩生成密集空间信息,对密集的空间信息按照类别数目使用Softmax进行信息过滤并映射到原输入特征中,此过程会从密集空间信息筛选出有关类别的特征信息从而抑制其他冗余信息的输出,从分类的角度对网络中冗余特征信息进行过滤以提高最终的分类精度. 如图1中类别注意模块所示,首先抽取Bottleneck输出特征$F(C\times H\times W)$,使用3×3卷积和1×1卷积将通道从C压缩到K(K表示语义分割的类别数)得到特征$F(K\times H\times W)$,其中每个通道包含的密集空间信息都代表一个类别粗略待分类特征. 然后将F变形到$F(K\times HW)$使空间特征信息被映射到一维向量上,通过Softmax函数对HW维度进行分类信息提取和过滤,该空间信息过滤方法如式(1)所示:

$$F_o=\mathrm{softmax}(F)=\sum_{i=1}^{k}\frac{\exp(x_{ij})}{\sum_{j=1}^{nm}\exp(x_{ij})} \tag{1}$$

其中,k为预设通道数;x_{ij}表示特征$F(K\times HW)$的单个像数信息;n、m分别是H和W的维度;$F_o(K\times HW)$是经过Softmax函数后的输出特征.

为提取类别注意力权重,特征层$F_o(K\times HW)$被还原到$F_o(K\times H\times W)$,使用1×1卷积将K个子空间类别特征信息聚合到一起,对聚合空间特征使用Sigmoid函数激活后得到类别注意力权重,并与输入特征相乘实现按照类别过滤信息的空间注意,从而增强网络提取特征过程中对类别信息的关注.

2.3 多尺度注意力

在语义分割网络中,常通过融合多尺度特征信息与扩大感受野来增强像素上下文信息,但未对网络中不同尺度区域使用注意力自适应关联上下文. 本文提出的多尺度注意力有效衔接了多尺度特征提取和注意力机制. 如图2多尺度注意力所示,对Bottleneck4输出特征$T(C\times\frac{H}{8}\times\frac{W}{8})$进行多尺度信息提取,即特征$T$按照不同卷积核大小进行卷积并将特征通道压缩到原来四分之一. 经过该方法得到四个不同尺度特征,对所有尺度特征添加TCTA提取不同尺度注意力权重,实现多个尺度区域特征在通道和空间上联系上下文信息,最后将TCTA所有输出特征按通道堆叠. 多尺度注意力实现如式(2)和式(3)所示:

$$T_i=\mathrm{TCTA}(\mathrm{Conv}^{i}_{k\times k}(T)) \tag{2}$$

$$T_{\mathrm{out}}=\mathrm{Cat}(T_1,T_2,T_3,T_4) \tag{3}$$

其中,k同时表示卷积核大小,$k\in\{3,5,7,9\}$;i表示不同尺度特征层代号,$i\in\{1,2,3,4\}$;$T_i(\frac{C}{4}\times\frac{H}{8}\times\frac{W}{8})$表示单个尺度TCTA输出特征;$T(C\times\frac{H}{8}\times\frac{W}{8})$代表输入特征;$T_{\mathrm{out}}(C\times\frac{H}{8}\times\frac{W}{8})$表示所有尺度特征$T_i$通道拼接后的输出特征.

如图2中TCTA结构图所示,TCTA采用全局最大池化和全局平均池化分别提取特征通道信息,输出特征为$G_m(\frac{C}{4}\times1\times1)$和$G_a(\frac{C}{4}\times1\times1)$. 通过ECA中提出的通道信息交互策略使用一维卷积为特征G_m和G_a建立通道信息联系,在不降低特征通道维度情况下建立权重映射关系,从而避免特征信息损耗. 然后将两通道输出特征进行特征融合,使全局最大池化和全局平均池化输出特征的通道信息互补,对融合后的特征使用Sigmoid函数提取通道注意力权重,再将权重映射到原特征f上形成并行双通道注意力. 并行双通道注意力实现如式(4)所示:

$$T_c=\mathrm{Sigmoid}\left[\mathrm{Avgpool}(\partial(f))+\mathrm{Maxpool}(\partial(f))\right]\times f \tag{4}$$

其中,$f(\frac{C}{4}\times\frac{H}{8}\times\frac{W}{8})$上为输入特征;$\partial$表示一维卷积通道交互策略;AvgPool和MaxPool代表全局平均池化和全局最大池化操作;$T_c\left(\frac{C}{4}\times\frac{H}{8}\times\frac{W}{8}\right)$表示双通道注意力输出特征. 此外,TCTA采用与CBAM中串联的方式形成混合域注意力,对上一级输出特征T_c使用1×1卷积聚合所有通道的空间信息,通过Sigmiod函数激活空间注意力权重信息并映射到特征T_c上,最终双通道注意力和空间域注意力串联形成TCTA模块. 图2中展示的TCTA输入和输出特征热力映射图表明TCTA会加强对目标重要信息区域的关注,通过与多尺度分支的有效结合来挖掘不同尺度间的上下文联系,缓解目标尺度变化大带来的分割消弥问题,有利于网络在复杂场景下进行分割.

3 实验结果及分析

3.1 实验准备

3.1.1 数据集建立

近年来云贵高原各大湖泊部分水体面积呈现逐年下降趋势,而湖泊周围建筑面积逐年扩张[23]. 因此,云贵高原湖泊在湖泊变化上具有显著研究价值. 滇池周围区域土地覆盖类别较全,数据信息较丰富,同时各类别土地相互穿插,在针对高原湖泊语义分割上具有一定的代表性和复杂度.

综上所述,本文选定环滇池区域作为研究对象. 采集环滇池区域光学遥感影像,数据主要来源于World-View-2卫星,对采集后的遥感数据划分为400张1 000×1 000像素值图像,使用Labelme进行标注,设置类别包括建筑用地、河流、湖泊、农业用地、植被以及其他用地(背景). 对标注完成的数据进行切分,得到像素值为

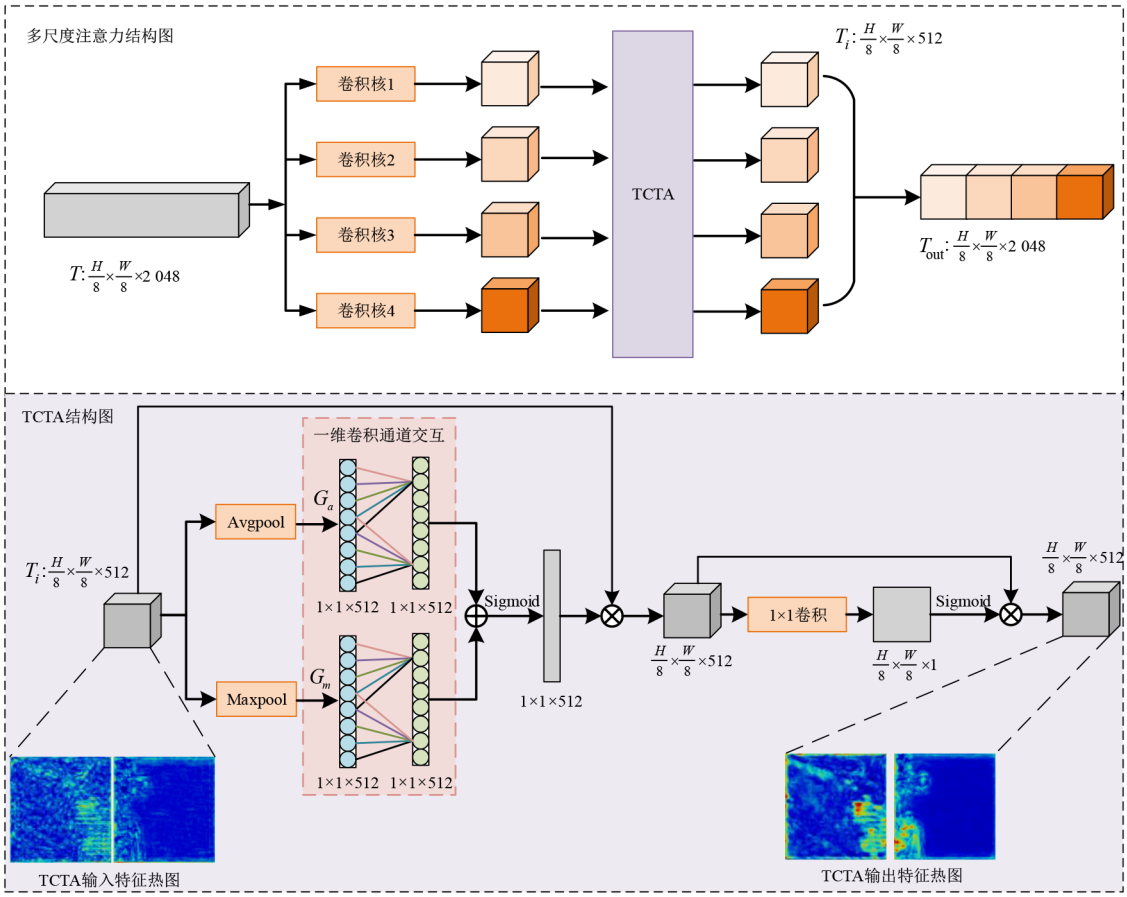


图2　多尺度注意力

500×500的训练图像和标签图像各1 600张,采用4∶1的比例设置训练集和测试集,训练过程中测试集不参与训练. 图3为本文SDL数据集中的原图像和标签图像,图4为各类别像素占比情况,其中河流、建筑物和植被存在小目标过多、分布零散的情况.

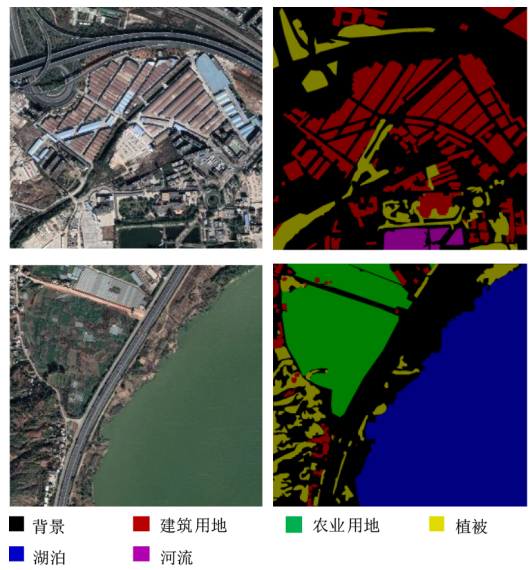


图3　SDL数据集示例

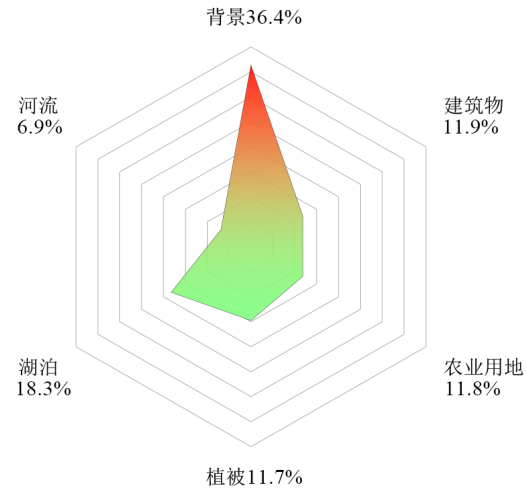


图4　类别像素占比统计

3.1.2　实验环境

本文实验环境基于Ubuntu 18.04和python 3.6,使用CPU为AMDR5-3600,内存16 GB,GPU为NVIDIA GeForce RTX2080ti的硬件平台,深度学习框架为PyTorch 1.20,使用CUDA 10.0和cudnn 7.6.5加速模型训练和测试.

为确定网络最佳学习率，采用不同学习率对模型进行分析测试. 如图5所示，对比4组不同学习率训练得到loss下降曲线，其中学习率(lr)与学习衰减策略的衰减值相差20倍，综合比对loss曲线图后确定学习率为0.005. 在模型训练过程中，预设的各种超参数以及loss函数与优化器的选择如表1所示. 为通过实验验证FCMANet各个模块有效性和消除超参数的影响，在实验分析中，模型训练的超参数设置和学习策略不变.

表1 训练网络的各项超参数及策略

超参数及学习策略	参数值
学习率	0.005
图像输入尺寸	500×500
Batch Size	4
动量	0.9
迭代步数	50 000
数据载入线程数	4
优化器	SGD
损失函数	交叉熵

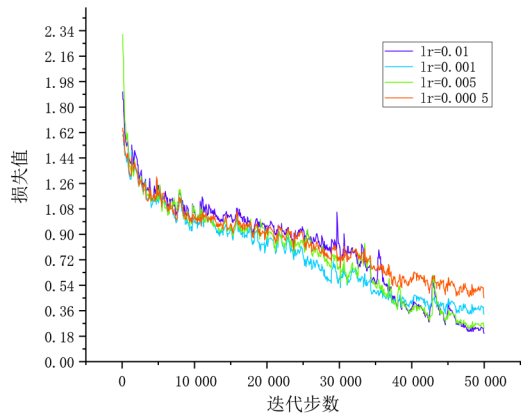


图5 不同学习率下的loss曲线

3.1.3 评价指标

本文目的在于准确并快速分割出遥感影像目标，因此采用平均交并比(mean Inetersection over Union，mIoU)和平均像素准确率(mean Pixel Accuracy，mPA)作为网络模型分割精度评价指标，模型测试时间作为模型速度评价指标，以此来综合评价模型性能. 平均交并比代表每个类别预测值与该类别真实值交集与并集的比值，采用混淆矩阵实现，mIoU计算如式(5)所示：

$$\mathrm{mIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{5}$$

平均像素准确率的计算要先计算每一类预测值当中正确分类的像素数量与所有像素数量的比值，再将每个类的PA求平均，计算如式(6)所示：

$$\mathrm{mPA}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}} \tag{6}$$

在式(5)和式(6)中，$k+1$代表分割类别总数(包含背景)；p_{ij}表示真实值是第i类，但预测值是第j类的像素数量，称为真负例；p_{ii}表示真实值是第i类，预测值也是第i类的像素数量，称为真正例；同理，p_{ji}表示假正例.

3.2 实验

3.2.1 模块消融实验及分析

为探究本文各模块在网络模型中对分割精度的影响，在网络为编码-解码结构框架和骨干网络为ResNet50前提下，分别添加类别注意模块和多尺度注意力(Multi-Scale Attention，MSA)进行消融实验.

实验定量结果如表2所示，在基础网络架构上分别添加类别注意模块和多尺度注意力，测试结果中mIoU分别提升了4.3%和4.1%，mPA分别增加3%和3.1%. 从定量结果来看，类别注意模块和MSA结构都能对分割效果起到积极作用，在二者叠加后的综合结果相比于原始网络模型mIoU和mPA分别提高了5.7%和4.3%.

表2 不同模块消融实验的结果

骨干网络	类别注意模块	MSA	mIoU/%	mPA/%
ResNet50	—	—	71.7	82.0
ResNet50	√	—	76.0	85.0
ResNet50	—	√	75.8	85.1
ResNet50	√	√	77.4	86.3

除定量分析外，还利用热力图对特征进行可视化，热力图能反映网络模型的特征权重占比，相比普通特征可视化结果更能凸显神经网络在学习过程中重点关注的特征信息区域. 如图6中展示的热力映射结果，其中图6(a)~(c)分别是网络模型输入图像与经过类别注意模块前后特征的热力映射图，对比图6(b)(c)的差异可以看出，类别注意模块会增强图像部分特征信息，引导网络在特征提取过程中重点关注类别信息. 图6中(d)~(f)分别代表了输入原图像、TCTA输入和输出特征的热力映射图，对比图6(e)(f)中的热点区域，在建筑物区域的网络权重占比更高，反映了经过TCTA模块后特征权重更集中在主要目标信息上，弱化了局部信息对高级语义信息的影响，从而增强了网络在复杂场景下自适应信息提取能力.

3.2.2 类别注意模块实验

注意力模块在网络中不同位置会导致精度差异，为验证类别注意模块不同位置和数量对网络分割精度的影响，本文在主干网络Bottleneck结构中穿插添加不同数量和位置的类别注意模块. 实验结果如表3所示.

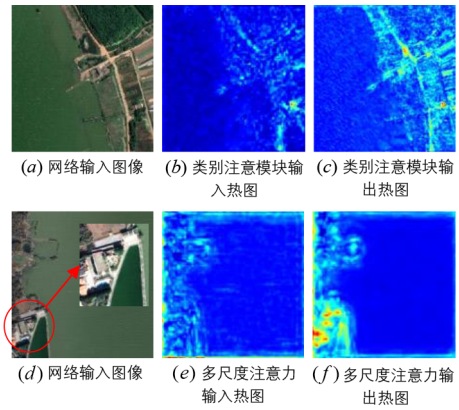


图6　热力图定性对比

表3中类别注意力位置使用0和1分别表示网络中四个位置是否添加了类别注意模块,0表示未添加,1表示添加.

表3　不同位置注意力实验结果

类别注意力位置	mIoU/%	mPA/%
(1, 1, 1, 1)	76.6	85.8
(0, 1, 1, 1)	77.4	86.3
(0, 0, 1, 1)	76.9	85.9
(0, 0, 0, 1)	77.3	86.3
(0, 0, 1, 0)	77.2	85.5
(0, 1, 0, 0)	76.8	85.9

从表3中结果可以得出,对类别注意模块的简单叠加并不能起到积极作用,相反注意力叠加过多会导致网络过于依赖注意力而损失网络主干信息,从而会造成分割精度下降. 另外,从网络不同位置添加类别注意模块的结果来看,由于浅层特征包含较多细节信息而缺少高级语义信息,类别注意模块提取的低级语义注意力权重不能充分对像素分类进行指导. 因此在高层语义特征中嵌入类别注意模块能实现最优化分类信息引导.

3.2.3　骨干网络消融实验

在网络结构设计上,为确定最优骨干网络以及不同骨干网络对分割精度的影响,在编码解码结构和注意力模块嵌入下仅对骨干网络进行替换,实验结果如表4所示.

表4　不同骨干网络实验结果

骨干网络	mIoU/%	mPA/%	FLOPs/G
ResNet-50	77.4	86.3	1 179
ResNet-101	75.0	84.7	1 480
ResNext-50	76.2	85.5	1 146
MobileNetV3	67.0	79.9	541

根据实验结果可知,ResNet101比ResNet50网络层次更深,深层次网络结构会提取到更高维度的语义特征,但浅层网络输出的特征信息较少从而导致分割精度上比ResNet50骨干更低. ResNext50[24]在ResNet50基础上对每个残差块使用分组卷积来拓展网络宽度,通过扩大特征信息通路来增加模型表征能力,但也包含更多冗余信息,导致类别注意模块提取较多错误类别信息,抑制了网络分割性能,造成分割精度下降. MobileNetV3[25]作为轻量化网络模型,如表4中浮点数(FLOPs)结果所示,网络参数量是其他网络FLOPs的一半,因此推理速度更快,但由于网络深度较低、信息通道较少并不能完全拟合数据,因此分割精度不高.

3.2.4　不同网络对比实验

为验证不同网络在SDL数据集上的分割效果,本文对比7种语义分割网络的综合分割能力,实验结果如表5所示. 分别用7种网络在相同实验环境下进行训练和测试,表5对比了各网络模型的mIoU和mPA精度、FLOPs、测试时间,以及各类IoU结果,加粗数据表示最优结果.

表5　与其他模型的对比实验结果

网络名称	mIoU/%	mPA/%	背景IoU/%	建筑用地IoU/%	农业用地IoU/%	植被IoU/%	湖泊IoU/%	河流IoU/%	FLOPs/G	测试时间/ms
PSPNet	73.2	79.5	70.4	54.4	79.5	58.9	84.0	56.2	687	187.5
SegNet	70.5	80.4	72.9	56.3	81.4	65.5	85.1	50.6	510	196.8
U-Net	73.5	85.0	74.4	70.5	83.7	66.1	85.7	60.7	957	203.1
DeeplabV3+	73.8	84.0	76.3	70.6	81.1	68.8	86.9	59.3	658	206.3
CCNet	75.2	84.8	76.9	72.6	84.1	71.2	86.2	60.5	874	200.0
ResNet50+PSA	75.6	86.0	76.5	72.0	84.1	70.3	87.5	63.6	466	225.0
FCMANet	**77.4**	**86.3**	77.7	73.9	85.9	71.9	88.8	65.3	1 179	196.8

从7种网络的分割精度结果对比可知,U-Net、DeeplabV3+,PSPNet和SegNet都属于编码解码结构网络,且都未嵌入注意力机制. 其中DeeplabV3+和PSPNet分别通过金字塔池化和空洞空间卷积池化金字塔增加多尺度特征信息,U-Net则通过融合高低层特征建立上下文联系,这些非注意力联系上下文信息方法的mIoU都在73%左右. 而Resnet50+PSA和CCNet都通过引入非局部自注意力获取网络全局语义信息,mIoU则达到

75%左右.FCMANet网络融合混合域注意力和多尺度信息构建上下文信息关联,并通过添加类别注意模块引导像素分类,最终mIoU达到77.4%.

从各类分割精度可知,湖泊IoU在80%以上,分割精度最优.而河流由于目标分布零散且在特征上与湖泊地块高度相似,导致分割精度在所有类别中最低,其中FCMANet通过引入类别注意力模块来减少分类误差,从而使得河流地块的分割精度较其他模型大幅上升.植被及建筑用地由于包含较多零散小目标且目标尺度差异大使得分割难度增加,因此分割精度有所下降.

在模型参数量上,通过计算所有网络在相同输入数据量下的FLOPs来衡量网络计算复杂度,FCMANet由于融合多尺度和注意力,在网络参数量上最高,训练时间和占用计算资源最多.在推理时间上,引入自注意力的Resnet50+PSA和CCNet在推理时间上比FCMANet采用局部混合注意力更久,FCMANet相比于PSPNet添加了注意力和多尺度信息,导致网络推理时间变长.

图7展示了7种网络模型在测试集上的分割结果,挑选6种不同场景下高原湖泊的土地分割进行分析,其中从左到右分别是场景1到6.对比总体分割效果,PSPNet,U-Net和Segnet出现了不同情况的错误分类情况,同时分割目标的边缘轮廓也比较模糊.Deeplabv3+采用空洞卷积提高感受野,拓展了网络获取全局信息的能力,降低了像素分类误差.Resnet50+PSA和CCNet通过自注意力捕获全局信息,但忽略了目标尺度不均问题,导致未分割出小目标区域.FCMANet采用类别注意模块引导网络像素分类,减少分类误差,且多尺度注意力提升了模型对多尺度目标的分割能力.在场景1中,植被类别的小目标过多,目标分布零散,FCMANet相比其他网络,有效保留了零散和小尺度目标的细节信息.在场景3和4中,FCMANet对目标轮廓边界轮廓处理更清晰.场景6中,建筑物部分目标小且密集,导致各目标边界轮廓较难被分割,而FCMANet分割效果表现良好,具备处理复杂分割场景的能力.

3.2.5 ISPRS Vaihingen 数据集实验

为验证FCMANet在其他地域遥感影像上分的割性能和模型泛化能力,选用ISPRS Vaihingen数据集进行模型性能验证.Vaihingen数据集一共包括16张有标注图像,包含不透水面、建筑物、树木、低矮植被、汽车、背景等类,同时该数据集存在类别不均衡问题.本次实验将数据集原图像和标注图像切分成250×250进行训练和测试,训练集与验证集比例为4∶1,在四种网络上分别进行和测试,训练时不进行数据扩充.四种网络在

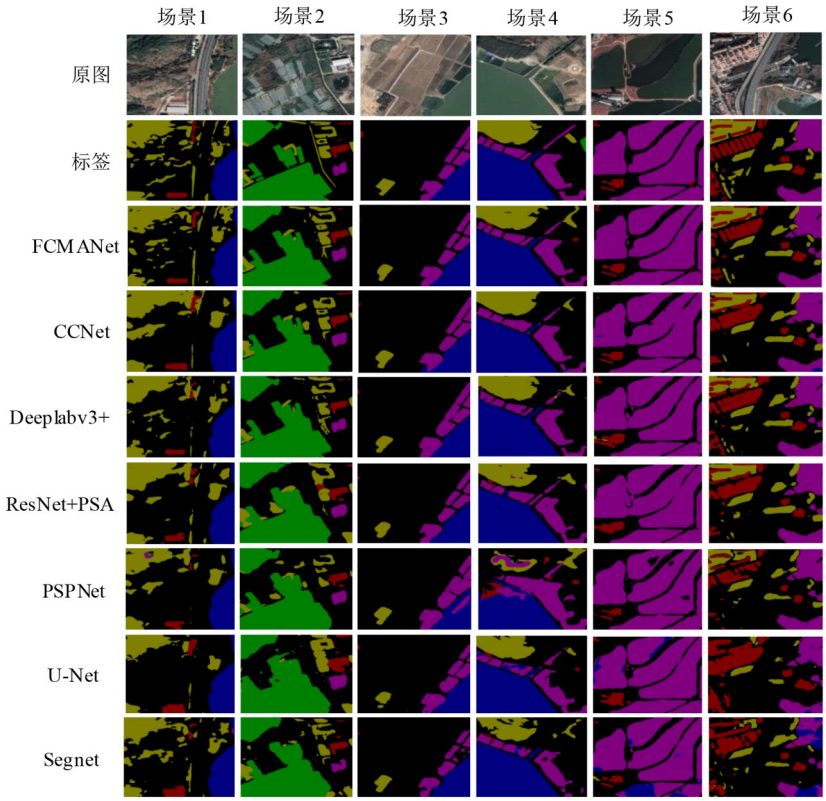
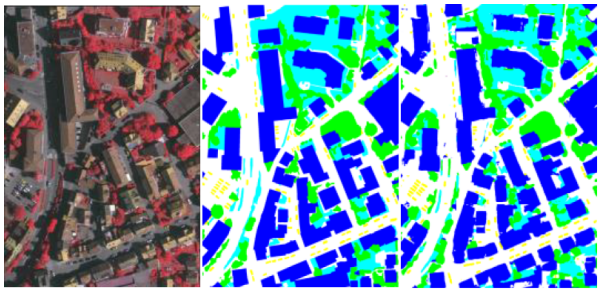


图7 不同网络SDL数据集上的分割效果

ISPRS Vaihingen数据集上的测试结果如表6所示,实际分割效果如图8所示. 从实验结果可以得出,FCMANet相较其他三个模型在数据类别不均衡且无数据扩充的情况下能进行准确的分割,验证了FCMANet模型对其他地域遥感影像也具有良好的泛化性.

表6　与其他模型的对比实验结果

网络模型	mIoU/%	mPA/%	测试时间/ms
SegNet	61.28	73.10	42.25
CCNet	68.56	81.50	78.57
ResNet+PSA	68.04	79.90	57.14
FCMANet	70.76	82.08	71.43



原图　　标签　　FCMANet

图8　Vaihingen数据集上的分割结果

3.2.6　WHDLD数据集实验

WHDLD[26]是一个多分类遥感语义分割数据集,该数据集总共包含4 950张256×256有标注图像,包括街道、裸露土地、建筑、公路、水体和植被六大分割类别. 本文随机选取20%图像作为测试集,即训练集与测试集按照4∶1设置进行训练,本次实验对比了四种模型的分割结果,表7为四种模型在WHDLD数据集上的测试精度,FCMANet的分割效果如图9所示. 由于WHDLD数据集图像分辨率较低导致感受野获取信息太密集造成细小目标分割困难,因此从整体分割效果上来看大区域地块分割效果较好,细小区域地块轮廓丢失较严重. 实验数据表明,FCMANet的在测试集上的mIoU和mPA分别为62.4%和74.38%,与其他三个模型相比,FCMA使用了多分支的多尺度注意力,能在一定程度上缓解低分辨率密集遥感影像中目标尺度不均问题,因此该网络模型具有更好的分割效果.

表7　WHDLD数据集实验结果

网络模型	mIoU/%	mPA/%	测试时间/ms
SegNet	58.98	71.27	41.50
CCNet	61.65	73.30	66.80
ResNet+PSA	60.05	71.75	61.74
FCMANet	62.40	74.38	73.87

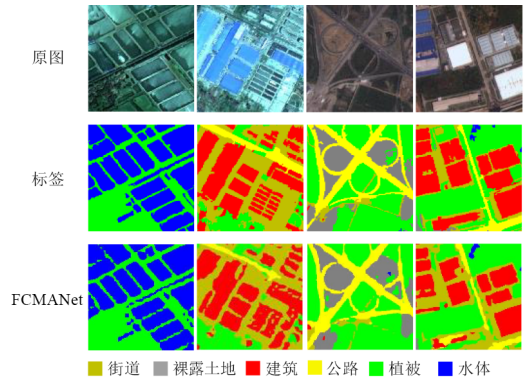


图9　WHDLD数据集上的分割结果

4　结论

本文建立了SDL数据集,并设计了融合类别与多尺度注意力的语义分割网络. 其中类别注意模块通过空间注意力和类别信息过滤引导网络关注分类信息,缓解了异类相似特征类别信息模糊问题,从而降低了像素分割误差. 通过实验发现,类别注意模块简单叠加会累积错误分类信息,引起分割精度下降. 另外,网络浅层特征包含丰富位置、轮廓等细节信息,而深层特征包含更高级的语义信息,因此将类别注意模块嵌入到深层特征以提取更精确有效的类别注意力权重. 多尺度注意力通过生成四个不同尺度特征层,提升网络对不同尺度目标的分割能力,其中双通道串联混合注意力为每个特征尺度的空间和通道信息建立上下文联系,通过可视化多尺度注意力前后特征热力映射图和模块消融实验对比,该注意力能在复杂遥感场景下有效提取不同尺度目标的重要特征信息.

从分割结果来看,多尺度注意力针对不同尺度特征分别建立上下文信息联系,在复杂分割场景下能准确分割出尺度不均、分布零散的目标区域,同时类别注意模块引导网络正确分类,提升网络分割精度. 但多尺度特征提取与注意力结合导致网络参数量较大,模型训练会占用较多计算资源. 因此,未来改进方案会考虑应用更简单有效的注意力模块,以及探索多尺度特征提取与注意力机制更简洁有效的结合方法,在保证精度的基础上进一步减少模型参数量和提高分割速度.

参考文献

[1] 闫立娟, 郑绵平, 魏乐军. 近40年来青藏高原湖泊变迁及其对气候变化的响应[J]. 地学前缘, 2016, 23(4): 310-323. YAN L J, ZHENG M P, WEI L J. Change of the lakes in Tibetan Plateau and its response to climate in the past forty years[J]. Earth Science Frontiers, 2016, 23(4): 310-323. (in Chinese)

[2] 朱立平, 鞠建廷, 乔宝晋, 等. “亚洲水塔”的近期湖泊变

化及气候响应: 进展、问题与展望[J]. 科学通报, 2019, 64(27): 2796-2806.
ZHU L P, JU J T, QIAO B J, et al. Recent Lake changes of the Asia Water Tower and their climate response: Progress, problems and prospects[J]. Chinese Science Bulletin, 2019, 64(27): 2796-2806. (in Chinese)

[3] 董一凡, 郑文秀, 张晨雪, 等. 中国湖泊生态系统突变时空差异[J]. 湖泊科学, 2021, 33(4): 992-1005.
DONG Y F, ZHENG W X, ZHANG C X, et al. Temporal and spatial differences of lake ecosystem regime shift in China[J]. Journal of Lake Sciences, 2021, 33(4): 992-1005. (in Chinese)

[4] 杨蕴, 李玉, 赵泉华. 高分辨率全色遥感图像多级阈值分割[J]. 光学精密工程, 2020, 28(10): 2370-2383.
YANG Y, LI Y, ZHAO Q H. Multi-level threshold segmentation of high-resolution panchromatic remote sensing imagery[J]. Optics and Precision Engineering, 2020, 28(10): 2370-2383. (in Chinese)

[5] 杨泽楠, 牛海鹏, 黄亮, 等. 基于MSR-cut的高空间分辨率遥感影像边缘检测分割[J]. 农业机械学报, 2021, 52(8): 154-162.
YANG Z N, NIU H P, HUANG L, et al. Edge detection segmentation method for high spatial resolution remote sensing image based on MSR-cut[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(8): 154-162. (in Chinese)

[6] 罗会兰, 张云. 基于深度网络的图像语义分割综述[J]. 电子学报, 2019, 47(10): 2211-2220.
LUO H L, ZHANG Y. A survey of image semantic segmentation based on deep network[J]. Acta Electronica Sinica, 2019, 47(10): 2211-2220. (in Chinese)

[7] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Munich: Springer, 2015: 234-241.

[8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 3431-3440.

[9] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6230-6239.

[10] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//European Conference on Computer Vision. Munich: Springer, 2018: 833-851.

[11] 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): 46-55.
ZHANG Z H, FANG W, DU L L, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network[J]. Acta Optica Sinica, 2020, 40(3): 46-55. (in Chinese)

[12] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[13] 尤洪峰, 田生伟, 禹龙, 等. 基于Word Embedding的遥感影像检测分割[J]. 电子学报, 2020, 48(1): 75-83.
YOU H F, TIAN S W, YU L, et al. Remote sensing image detection and segmentation based on word embedding [J]. Acta Electronica Sinica, 2020, 48(1): 75-83. (in Chinese)

[14] 肖春姣, 李宇, 张洪群, 等. 深度融合网结合条件随机场的遥感图像语义分割[J]. 遥感学报, 2020, 24(3): 254-264.
XIAO C J, LI Y, ZHANG H Q, et al. Semantic segmentation of remote sensing image based on deep fusion networks and conditional random field[J]. Journal of Remote Sensing, 2020, 24(3): 254-264. (in Chinese)

[15] ROY A G, NAVAB N, WACHINGER C. Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks[C]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2018. Granada: Springer, 2018: 421-429.

[16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132-7141.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]///European Conference on Computer Vision - ECCV 2018. Munich: Springer, 2018: 3-19.

[18] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Glasgow: IEEE, 2020: 11531-11539.

[19] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on

Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7794-7803.

[20] HUANG Z L, WANG X G, WEI Y C, et al. CCNet: Criss-cross attention for semantic segmentation[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. DOI: 10.1109/TPAMI.2020.3007032.

[21] LIU H J, LIU F Q, FAN X Y, et al. Polarized self-attention: Towards high-quality pixel-wise regression[EB/OL]. (2021-07-02)[2022-01]. https://arxiv.org/abs/2107.00782.

[22] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2016: 770-778.

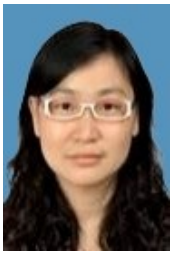
[23] 肖茜, 杨昆, 洪亮. 近30 a云贵高原湖泊表面水体面积变化遥感监测与时空分析[J]. 湖泊科学, 2018, 30(4): 1083-1096.
XIAO Q, YANG K, HONG L. Remote sensing monitoring and temporal-spatial analysis of surface water body area changes of lakes on the Yunnan-Guizhou Plateau over the past 30 years[J]. Journal of Lake Sciences, 2018, 30(4): 1083-1096. (in Chinese)

[24] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 5987-5995.

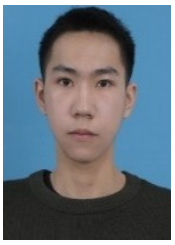
[25] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2020: 1314-1324.

[26] SHAO Z F, ZHOU W X, DENG X Q, et al. Multilabel remote sensing image retrieval based on fully convolutional network[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 318-328.

作者简介



何自芬　女,1976年10月出生于河北省南宫市. 现为昆明理工大学机电工程学院副教授、硕士生导师. 主要研究方向为计算机视觉、图像处理.
E-mail: zyhhzf1998@163.com



史本杰　男,1997年10月出生于四川省成都市. 现在为昆明理工大学硕士研究生. 主要研究方向为遥感图像处理.
E-mail: 1217432073@qq.com

张印辉(通讯作者)　男,1977年9月出生于河北省衡水市. 现为昆明理工大学机电工程学院教授、博士生导师. 主要研究方向为图像处理、机器视觉和机器智能.
E-mail: yinhui_z@163.com