

基于多尺度特征融合的遥感图像水体分割

李益[†]

(江苏科技大学, 江苏 镇江 212000)

摘要: 水体语义分割在遥感和地球观测领域发挥着至关重要的作用, 现有方法难以在复杂情景下有效分割多尺度水体的细小边界, 并且面临标注数据成本高的问题。为此, 提出了一种基于多尺度特征融合的水体遥感图像语义分割算法。基于双注意力模块和转换注意力模块设计了水体分割网络主体架构, 并设计一种多尺度特征融合模块减少编码器和解码器中不同尺度特征之间的语义差距。在模型训练之前, 首先通过基于掩码-重建的自监督学习对编码器进行预训练, 以获得具有丰富特征表示的初始化权重。在FloodNet数据集上所提出的方法达到了最高的89.77% mIoU, 相比于目前最先进的语义分割算法DAE-Former实现了1.31%的提升。实验结果表明, 所提出方法具有较高的准确性。

关键词: 多尺度特征融合; 注意力; 语义分割; 遥感图像

中图分类号: TP183

文献标识码: A

Remote Sensing Image Water Body Segmentation Based on Multi-scale Feature Fusion

LI Yi[†]

(Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212000, China)

Abstract: Water area semantic segmentation plays a crucial role in remote sensing and earth observation. Existing methods can hardly effectively segment small boundaries of multi-scale water areas in complex scenarios, and face the challenge of high annotation data costs. Consequently, a multi-scale fusion based water area segmentation method for remote sensing images is proposed. The proposed water segmentation network is designed based on dual attention modules and transformed attention modules. A multi-scale feature fusion module is introduced to reduce the semantic gap between different scale features in the encoder and decoder. Prior to model training, the encoder is pre-trained with mask-reconstruction based self-supervised learning to obtain initial weights with rich feature representations. Experimental evaluation was conducted on the FloodNet dataset. The proposed method achieved the highest mIoU of 89.77%, which is a 1.31% improvement compared to the state-of-the-art semantic segmentation algorithm DAE-Former. The experimental results demonstrate the accuracy of the proposed method.

Key words: multi-scale feature fusion; attention; semantic segmentation; remote sensing images

遥感图像水体分割对湿地生态系统保护与恢复、河流域监管与污染治理、自然灾害评估等领域研究具有重要意义^[1], 其主要任务在于将各类

型遥感图像中的水体和陆地区分出来, 具体包含河湖分割^[2]、海岸线提取^[3]、河流提取^[4]等。

目前, 水体遥感图像语义分割方法主要分为机

器学习和深度学习两类。基于机器学习的方法包括边缘检测、活动轮廓线模型和聚类等。然而,机器学习方法试图制定某种映射关系来确定像素的类别,这导致其缺乏灵活性和智能性。近年来,以 UNet^[5]、DeepLabv3+^[6] 等算法为代表的卷积神经网络(CNN)^[7] 由于其平移不变性和权值共享的特点,在遥感图像分割领域取得了显著的效果。基于 CNN 的方法提取全局信息的能力较差,而 ViT^[8]、MissFormer^[9] 等基于 Transformer^[10] 的方法在提取长程全局信息上表现优异,但在高分辨率图像的细特征提取上能力受限。此外,陆地水体通常是不规则的连通区域,其表面粗糙度相对较低,缺乏显著的纹理,并且水体边界模糊且不规则,高准确率的水体分割具有较大难度。

因此,提出了一种基于多尺度特征融合的水体遥感图像语义分割方法。首先在分割网络中加入双注意力模块和转换注意力模块以提取水域细小边界的特征,并提出一种多尺度特征融合模块减少不同尺度特征之间的语义差距。为使模型在训练时具有更优的初始化权重以降低对标主数据的需求,通过掩码-重建的自监督学习方式对模型编码器的权重进行预训练。在 FloodNet 数据集上的实验结果表明,与其他现有方法相比,所提出方法达到了最高的水体分割准确率,验证了其在遥感图像水体分割任务中的有效性。

1 相关工作

1.1 机器学习分割方法

基于边缘检测的机器学习算法利用水体独特的边缘信息实现水体与陆地边缘的区分。文献[11]基于变分演算和区域水平集获得纹理特征,随后使用支持向量机(Support Vector Machine, SVM)对 COSMO-SkyMed X 波段图像中的水体进行分割。基于活动轮廓线模型的遥感图像分割算法将图像分割问题转化为寻找能量通用函数的最小值。文献[12]在连接边缘段的主动轮廓之前,通过局部边缘选择和沿小波尺度传播来细化水体边缘。文献[13]以丹江口水库为研究实例,首先通过粗分割方法来提取初始水线。其次,给出了局部区域划分的策略,以加快后续过程。然后将散斑滤波器和改进的几何活动轮廓模型相结合用于精细分割。最后,利用变化检测方法对湖泊的变化进行了研究。基于聚类的分割算法根据特定的标准(比如距离规则),将图像中的像素分成不同的类,同类

像素尽量聚在一起,不同像素尽量分开。文献[14]结合水体指数 NDWI 和改进的 FCM 聚类算法实现在城市地表水、河流的自动提取算法,不仅较好地保持了水体区域完整性,对复杂背景也具有更好的抑制作用。文献[15]采用面向对象方法对 Freeman-Durden 极化分解后的分量图像进行分割,然后使用随机森林(RF)分类方法对图像进行分类,提取开放水域。文献[16]采用 Pansharpening 融合法获得光谱扭曲小的高分辨率卫星融合影像,并通过标记分水岭算法对高分辨率卫星融合影像进行水体分割。

1.2 深度学习分割方法

卷积神经网络可以自动提取遥感图像中的多层次特征,其中浅层网络可以学习到更直观的图像特征,而深层网络具有更强的非线性表达能力。文献[17]利用原始的 CNN 进行河道区域提取,该方法在高分一号 2 m 图像中准确地提取了永定河的河道范围。文献[18]融合了光谱和空间信息,通过 CNN 进行特征提取后利用 SVM 对图像像素进行分类。文献[19]利用 UNet 融合 Sentinel-1 和 Sentinel-2 图像,探索多源遥感图像用于水体分割的可能性。

注意力机制可以在有限的计算资源下解决语义分割任务中的信息过载问题。文献[20]首先引入了通道注意力的概念,通道注意力由压缩和激励模块组成,用于收集全局信息,捕捉通道之间的关系,提高特征的表示能力。文献[21]在常用特征提取骨干网络的基础上引入了通道注意力和空间注意力混合模块,对长江三角洲部分地区的高分六号卫星多光谱影像提取水体范围。

2 网络结构

2.1 整体架构

所提出水体分割网络的整体结构如图 1 所示,在网络的编码器部分,首先利用可变形卷积(Deformable Convolution, DConv)将输入图像的通道扩展至 64,以自适应调整卷积采样点,更好地提取不规则的水体边缘特征。随后,特征进入可变形残差网络(Deformable ResNet, Dres)进行多尺度特征提取,其在传统 ResNet 结构的基础上将普通卷积替换为可变形卷积。共计包含四个可变形残差模块,每个模块中对输入特征的分辨率减半,通道数加倍。

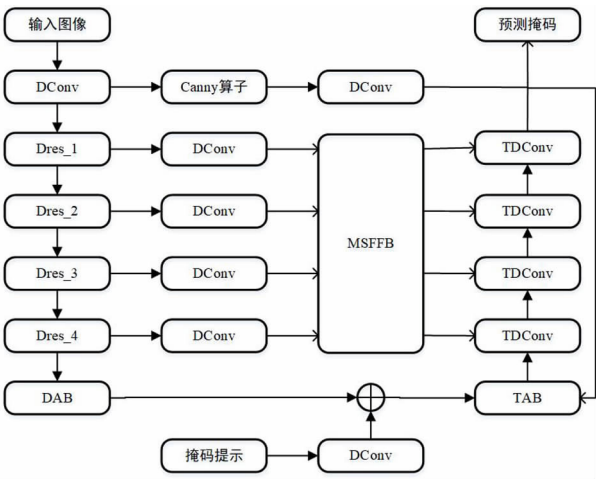


图 1 水体分割网络整体结构图

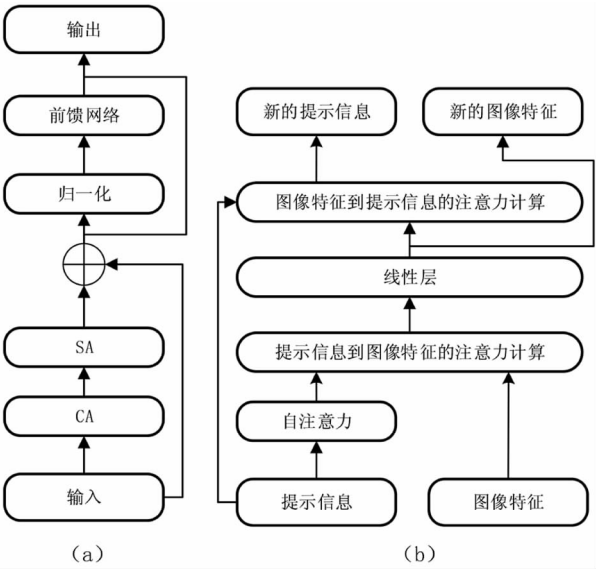


图 2 双注意力模块和转换注意力模块

特征经过双注意力模块 (Dual Attention Block, DAB)[图 2(a)]在通道维度和空间维度上分别进行注意力计算,更好地提取高级语义信息。在双注意力模块中,输入特征首先在通道维度进行通道注意力(Channel Attention,CA)计算,然后进行空间注意力(Spatial Attention,SA)的计算。通道注意力和通道注意力的计算如公式(1)和公式(2)所示:

$$C(Q,K,V)VK^TQ$$

(1)

$$S(Q,K,V)=QK^TV$$

(2)

其中,Q,K,V 分别代表查询、键和值,由同一个输入复制三份得到。随后特征进行归一化并经过前馈网络进行线性映射。前馈网络中进行的计

算过程如公式(3)所示。

$$F(X)=FC(GELU(DWConv(FC(X))))$$

(3)

其中,FC 代表全连接层,GELU 表示 GELU 激活函数,DWConv 代表深度可分离卷积。

在所提取的特征输入解码器之前,利用 Canny 算子进行边缘提取,结合可变形卷积调整特征维度得到稀疏提示,并利用可变形卷积直接对掩模操作得到稠密提示。实现了利用图像边缘信息和掩模信息自动生成稀疏提示和稠密提示,进而调整网络的关注目标。

在解码器部分,首先利用设计的转换注意力模块(Transform Attention Block,TAB)和多个转置可变形卷积(Transpose Deformable Convolution,TDCnv)作为解码器,在保留双向注意力计算的同时,实现了多尺度的特征还原并最终输出预测掩模。转换注意力模块的结构如图 2(b)所示,其输入包括提示信息和图像特征两部分。首先对提示信息进行自注意力的计算,然后将提示信息作为查询,图像特征复制两份分别作为键和值,做提示信息到图像特征的注意力计算。随后经过线性层后得到新的图像特征,并将新的图像特征作为查询,原始的提示信息复制两份作为键和值,做图像特征到提示信息的注意力计算,得到新的提示信息。最终将新的图像特征和新的提示信息共同输出。

同时,为了缓解多尺度特征之间的语义差异、编码器和解码器之间的语义差异,构建了多尺度特征融合模块(Multi-scale Feature Fusion Block,MSFFB),以替代传统的跳跃连接进行多尺度特征融合,避免非必要跳跃连接引起的参数过大问题并弥补其语义差距。

2.2 多尺度特征融合模块

编码器多尺度的特征统一至相同尺寸后,输入图 3 所示的多尺度特征融合模块。不同尺度编码器特征按照通道维度拼接,进行通道注意力计算后重新拆分。随后将不同尺度编特征按照长和宽所在维度拼接,计算空间交叉注意力(Spatial Cross Attention,SCA),依次将各尺度特征作为查询,将拼接后的特征复制两份作为键和值以进行注意力计算。然后各尺度特征分别进行层归一化(Layer Normalization, LN)和可变形卷积后,恢复各尺度原有的特征维度,并与解码器中对应维度的特征进行矩阵相加,最终实现多尺度特征之间、编码器和解码器之间的信息融合。

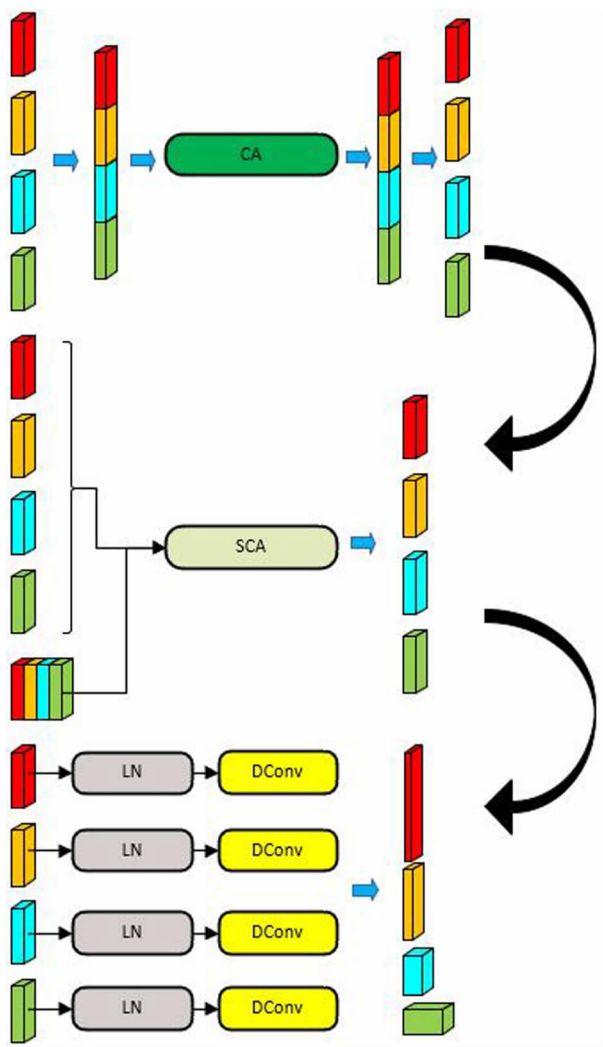


图 3 多尺度特征融合模块结构图

3 基于掩码-重建自监督的预训练策略

为促进分割网络训练收敛,提升数据利用效率和提升模型泛化能力,在利用标注数据进行有监督训练之前首先通过掩码-重建的自监督学习方法训练模型初始权重。与基于 Vision Transformer 的掩码重建自监督训练方式^[22]不同,在 CNN 中将输入图像切片后直接进行掩码会造成像素分布偏移问题,并且原始的掩码-重建方式无法在多个尺度上进行有效训练,因此提出了图 4 所示的自监督预训练策略。

由于对图像切片后掩码会导致像素分布偏移,因此本文通过子流形稀疏卷积跳过稀疏特征中的掩码位置并计算空白位置的掩码嵌入得到密集特征。在掩码-重建自监督预训练过程中,先将输入

图像进行随机掩码和子流形稀疏卷积操作,再通过可变形卷积操作得到特征图,随后特征图经过阶段 1—4。阶段 1 的操作为先对特征图切片后进行随机掩码(将数值置为 0)和子流形稀疏卷积操作,后经过可变形残差网络,阶段 2—4 亦是如此。相较于原始的掩码-重建策略,本文所采用的掩码策略不仅在原始图像的像素级别上进行,还在更高的抽象特征级别进行。

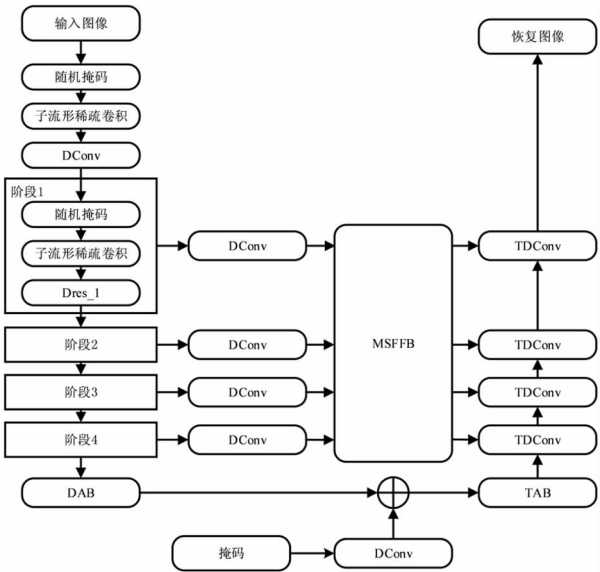


图 4 掩码-重建自监督预训练流程图

解码器和多尺度特征融合模块结构保持不变,但预测输出为三通道图像,通过预测图像与输入图像的 L2 损失更新网络参数。在利用标注样本进行有监督训练时,利用自检监督训练的可变形卷积和四个可变形残差网络模块参数初始化编码器权重,网络其他部分参数则通过高斯分布进行取值。

4 实验结果与分析

4.1 实验数据

实验所用数据集是公开数据集 FloodNet。该数据集是使用小型无人机平台 DJI Mavic Pro 四轴飞行器收集的。数据集包含 2343 张图像。其中,训练集占比 60%,验证集占比 20%,测试集占比 20%。语义分割标签包括被淹建筑物、未淹建筑物、被淹道路、未淹道路、水、树、车辆、水池和草共计 9 种类别,数据集的部分样本如图 5 所示。

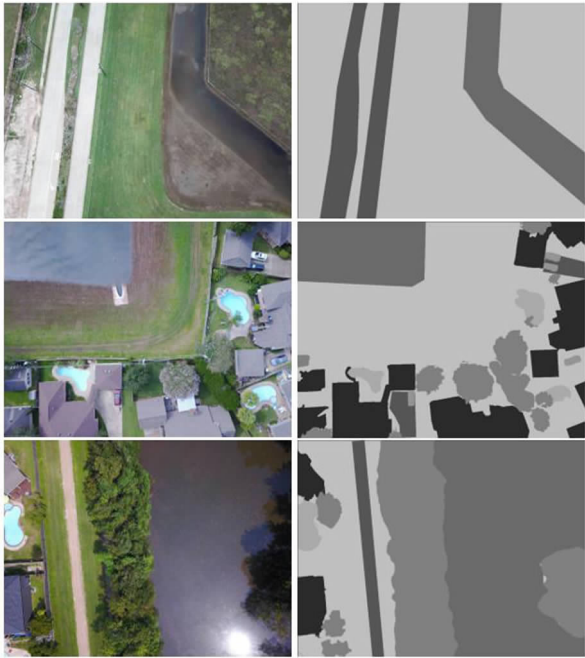


图 5 FloodNet 数据集样例

4.2 实验设置

实验在 Ubuntu20.04 LTS 系统上进行,CPU

为 Intel Core i5-13600KF,显卡为 NVIDIA RTX 3090,显存为 24 GB。实验软件为 PyCharm,语言为 Python 3.8。Batchsize 设置为 12,损失函数为 Dice,初始学习率为 0.001,最大轮次设置为 150,评估指标为平均交并比 (mean Intersection over Union, mIoU)。

4.3 对比实验

参与对比实验的算法共有七种,对比实验结果如表 1 所示。实验结果表明,所提出方法的平均交并比显著优于其余基线方法。与传统的图像分割算法(E-Net、U-Net 和 PSP-Net)相比,所提出方法实现了 21.43%~49.93%的平均交并比(mIoU)提升。与最新的先进方法(MissFormer、HiFormer 和 DAE-Former)相比,所提出方法实现了 1.31%~2.89%的平均交并比(mIoU)提升。

被淹建筑物、被淹道路、水三种类别常常具有不规则的边界,导致分割难度较大。所提出方法在上述三种类别上同样实现了较高的交并比。综上所述,所提出方法可以在多类别、边界不规则等复杂情况下实现水体遥感图像的精准分割。

表 1 对比实验结果

方法	被淹 建筑物	未淹 建筑物	被淹 道路	未淹 道路	水	树	车辆	水池	草	mIoU
ENet ^[23]	21.82	41.41	14.76	52.53	47.14	62.56	26.21	16.57	75.57	39.84
DeepLabv3+ ^[6]	28.10	78.10	32.00	81.10	73.00	74.50	33.60	40.00	87.10	58.61
UNet ^[5]	30.11	81.94	52.18	88.69	81.74	82.82	50.14	56.23	91.22	68.34
PSPNet ^[24]	65.61	90.92	78.69	90.90	91.25	89.17	54.83	66.37	95.45	80.35
MissFormer ^[9]	70.34	93.21	79.03	86.11	92.34	91.53	68.17	71.34	90.53	85.88
HiFormer ^[25]	78.52	94.01	80.21	84.67	92.14	92.68	80.32	73.11	92.54	87.80
DAE-Former ^[26]	81.34	94.62	81.88	86.57	94.82	94.01	81.02	85.68	93.22	88.46
所提出方法	82.74	95.36	82.34	85.88	95.81	93.34	82.28	91.67	96.59	89.77

4.4 消融实验

针对网络各模块有效性的消融实验结果如表 2 所示。由表 2 可知,对分割性能影响最大的是 MSFFB 模块,其次是自动生成稀疏和稠密提示。利用 Canny 算子获取边缘特征作为稀疏提示,并从掩码中使用可变形卷积获取稠密提示,进而实现自动的提示生成,这可以极大地提高模型的分割性能。而通过多种注意力计算,可以解决多尺度特征之间以及编码器和解码器之间的语义差距,进一步提升模型效果。本文还对下采样次数进行了验证,证明四次下采样的模型性能最佳。

表 2 网络各模块消融实验结果

操作	mIoU
去除多尺度模块	87.78
去除 DAB	88.12
去除 TAB	88.45
去除 MSFFB	86.98
去除掩码提示	87.43
下采样次数为 2	88.98
下采样次数为 3	89.03
下采样次数为 4	89.77
下采样次数为 5	89.65

自监督预训练初始化消融实验结果如表 3 所示,实验结果表明,不同初始化策略对网络性能有

显著影响。随机初始化和采用高斯分布取值的策略提供了相似的基线性能。采用 MoCo^[27] 和 SimCLR^[28] 这两种对比式自监督训练方法预训练的编码器分别能够将 mIoU 提高到 86.03% 和 87.48%，所提出的预训练方法显著提高了模型性能，达到了最高的 89.77% 的 mIoU，证明了所提出预训练方法在提高网络性能方面的显著效果和潜力。

表 3 自监督预训练初始化消融实验结果

初始化策略	mIoU
随机初始化	85.15
高斯分布取值	85.22
MoCo 预训练编码器	86.03
SimCLR 预训练编码器	87.48
所提出预训练方法	89.77

5 结 论

针对遥感图像水体分割任务中多尺度水体细小边界的分割难题，提出了一种基于多尺度特征融合的水体遥感图像语义分割方法。首先利用多种注意力机制和多尺度特征融合模块构建了水体分割网络，并采用掩码-重建自监督预训练机制对网络编码器权重进行初始化。全面的对比实验和消融实验验证了所提出方法的有效性。

遥感图像通常覆盖广阔的地理区域，处理这些大尺度数据需要高效的算法和计算资源，尤其是对于实时或近实时的水域监测应用，因此未来的研究将会聚焦于高分辨率遥感图像的性能优化。

参考文献

[1] 温泉,李璐,熊立,等. 基于深度学习的遥感图像水体提取综述[J/OL]. 自然资源遥感,1-16[2024-02-16].

[2] ARISTIZABAL F, JUDGE J. Mapping fluvial inundation extents with graph signal filtering of river depths determined from unsupervised clustering of synthetic aperture radar imagery[C]//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. 2021; 6124-6127.

[3] DING X, NUNZIATA F, LI X, et al. Performance analysis and validation of waterline extraction approaches using single-and dual-polarimetric SAR data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(3): 1019-1027.

[4] GASNIER N, DENIS L, FJØRTOFT R, et al. Narrow river extraction from SAR images using exogenous information[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5720-5734.

[5] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015; 18th International Conference. 2015; 234-241.

[6] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018; 801-818.

[7] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[8] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[9] HUANG X, DENG Z, LI D, et al. Missformer: An effective medical image segmentation transformer[J]. arXiv preprint arXiv:2109.07162, 2021.

[10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30.

[11] CAFARO B, CANALE S, PIRRI F. X-SAR SpotLigh images feature selection and water segmentation[C]//2012 IEEE International Conference on Imaging Systems and Techniques Proceedings. 2012; 217-222.

[12] NIEDERMEIER A, ROMANEESSEN E, LEHNER S. Detection of coastlines in SAR images using wavelet methods[J]. IEEE Transactions on Geoscience and Remote Sensing, 2000, 38(5): 2270-2281.

[13] LI N, NIU S, GUO Z, et al. Dynamic waterline mapping of inland great lakes using time-series SAR data from GF-3 and S-1A satellites: a case study of DJK reservoir, China[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(11): 4297-4314.

[14] 洪亮,黄雅君,杨昆,等. 复杂环境下高分二号遥感影像的城市地表水体提取[J]. 遥感学报,2019,23(5):871-882.

[15] SHEN G, FU W. Water body extraction using GF-3 polar data: a case study in Poyang lake[C]//IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2020; 4762-4765.

[16] 贾永红,冯在梅,沈庭乐. 基于标记分水岭分割的高分辨率影像水体信息提取[J]. 测绘通报,2014(9):31-42.

[17] 王惠英,孙中平,孙志伟,等. 基于深度学习的河道提取与变化监测应用——以永定河为例[J]. 北京测绘,2019,33(2):173-178.

[18] 吕亚龙,田生伟,禹龙,等. 空谱联合特征的 CNN_SVM 水体识别[J]. 计算机工程与设计,2019,40(5):1435-1439.

[19] KONAPALA G, KUMAR S V, AHMAD S K. Exploring sentinel-1 and sentinel-2 diversity for flood inundation mapping using deep learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 180: 163-173.

[20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018; 7132-7141.

[21] 沈骏翔,马梦婷,宋致远,等. 基于深度学习语义分割模型的高分辨率遥感图像水体提取[J]. 自然资源遥感,2022,34(4): 129-135.

[22] HE K, CHEN X, XIE S, et al. Masked auto-encoders are scalable vision learners[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022; 16000-16009.

[23] PASZKE A, CHAURASIA A, KIM S, et al. Enet: a deep neural network architecture for real-time semantic segmentation[J]. arXiv preprint arXiv:1606.02147, 2016..

[24] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017; 2881-2890.

[25] HEIDARI M, KAZEROONI A, SOLTANY M, et al. Hiformer: hierarchical multi-scale representations using transformers for medical image segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023; 6202-6212.

[26] AZAD R, ARIMOND R, AGHDAM E K, et al. Dae-former: dual attention-guided efficient transformer for medical image segmentation[C]//International Workshop on Predictive Intelligence In Medicine. Cham; Springer Nature Switzerland, 2023; 83-95.

[27] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020; 9729-9738.

[28] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]// International Conference on Machine Learning. PMLR, 2020; 1597-1607.