doi:10. 3969/j.issn. 1672-4623.2023.01.011

一种新的基于深度学习的遥感影像变化检测算法
——H-BIT方法的提出与应用

傅绘锦[1]
（1. 武汉大学 遥感信息工程学院，湖北 武汉 430079）

摘　要：对遥感影像变化检测进行了研究，提出了一种融合HRNet与BIT的H-BIT方法，能兼顾遥感影像高分辨率与高语义特征，在场景复杂、目标尺度跨度大时表现优异。该方法经HRNet网络、词元分析器处理后，引入Transformer，通过注意力机制，能从更大的感受野解译变化结果。在LEVIR-CD数据集上进行了实验，结果表明H-BIT方法能完整提取目标，对目标边缘的处理更平滑，变化检测的精确率、召回度、F1得分和总体精度分别为86.95%、90.24%、87.93%和96.62%，均高于原始BIT算法的表现，说明该方法能适应多尺度目标与复杂场景，变化检测精度高、计算速度快、鲁棒性高、泛化性强。

关键词：变化检测；深度学习；HRNet网络；并行结构；注意力机制
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2023）01-0053-05

A New Change Detection Algorithm for Remote Sensing Images Based on Deep Learning: The Proposal and Application of H-BIT Method

FU Huijin[1]
(1. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China)

Abstract: In this paper, we studied the change detection of remote sensing images, and proposed an H-BIT method combining HRNet with BIT, which could take into account the high resolution and high semantic features of remote sensing images, and perform well when the scene was complex and the target scale span was large. Firstly, H-BIT method extracted image features by HRNet. Then, the tokenizer transformed features into tokens. With an attention mechanism model, change detection results were calculated from a larger receptive field. In our experiments, we qualitatively and quantitatively demonstrated that H-BIT method outperformed the classical BIT on LEVIR-CD dataset with higher accuracy (86.95% accuracy, 90.24% recall, 87.93% F1 score, 96.62% overall accuracy). To be specific, with less error detection, H-BIT method adapt to multi-resolution targets and complex scenarios with high level of robustness and generalization.

Key words: change detection, deep learning, HRNet network, parallel structure, attention mechanism

利用遥感影像动态掌握国土资源变化，可为国家提供地理国情信息决策支撑。传统人工变化检测对高分辨率卫星遥感图像的特征刻画能力较差，且工作量巨大，而基于深度学习的解译方法可自动分析变化信息，但如何提升检测效率和准确性仍是业界重要讨论的主题。变化检测主要分为两步，即先进行图像内的目标提取，再计算图像间的信息差，因此高效正确的目标提取对后续信息差计算影响很大。

在深度学习普及前，传统方法根据像素阈值、颜色等信息进行目标提取，如袁敏[1]等利用最大流、最小割函数进行图像分割，但无法消除遥感影像中“同物异谱、同谱异物”的影响；肖明虹[2]等提出的超像素协同分割变化检测方法有效克服了椒盐噪声，但忽略了像素块之间的关系，检测效果不理想。随着深度学习的快速发展，变化检测方法得到了相关算法的支持与改进，如LYU H[3]等利用卷积神经网络提取遥感影像抽象特征，但只重点关注了高维语义信息，重要边界细节易丢失；GONG M[4]等利用分类图获取训练样本，通过深度学习训练得到变化检测图，但特征提取较少，预分类结果缺少空间结构信息。业界提出利用语义分割提取目标的方法，语义分割网络形成两大流派：①以SVM、UNet为代表的下采样后复原方法，如孙红岩[5]提出的SVM与多特征融合的方法，该方法受噪声影响较大，在特征图分辨率缩小的过程中仍有空间信息流失；②以DeepLab为代表的保持特征高分辨率的方法，如赵祥[6]等提出的改进的DeepLab3+孪生网络，计算繁杂、效率不高。相较于传统目视方法，深度学习降低了检测成本，但目前只有少数研究关注到保持高空间语义信息对模型性能的影响，为了做出改进，SUN K[7]等在研究人类姿势时提出了一种并行结构

收稿日期：2022-08-09。

网络HRNet，不同于传统语义分割网络，其通常基于从高到低的编码器提取特征图，使特征提取同时拥有高分辨率和高语义，能保留重要的形状和边界细节。受上述研究启发，本文以CHEN H[8]等提出的BIT-CD框架为基础，设计了一种基于HRNet方法的H-BIT遥感影像变化检测方法，通过HRNet融合高分辨率和高语义信息，Transformer整合特征向量与语义特征输出强化特征，最后生成二值变化结果；并在LEVIR-CD数据集上进行了对比实验，验证了本文方法的有效性和鲁棒性，还通过消融实验获取了最佳模型。

1 H-BIT遥感影像变化检测方法

1.1 网络结构

H-BIT变化检测网络结构如图1所示，由HRNet-V2特征提取主网络、双时态图像自注意力变换网络Transformer和差异计算预测Predction Head三个模块组成。

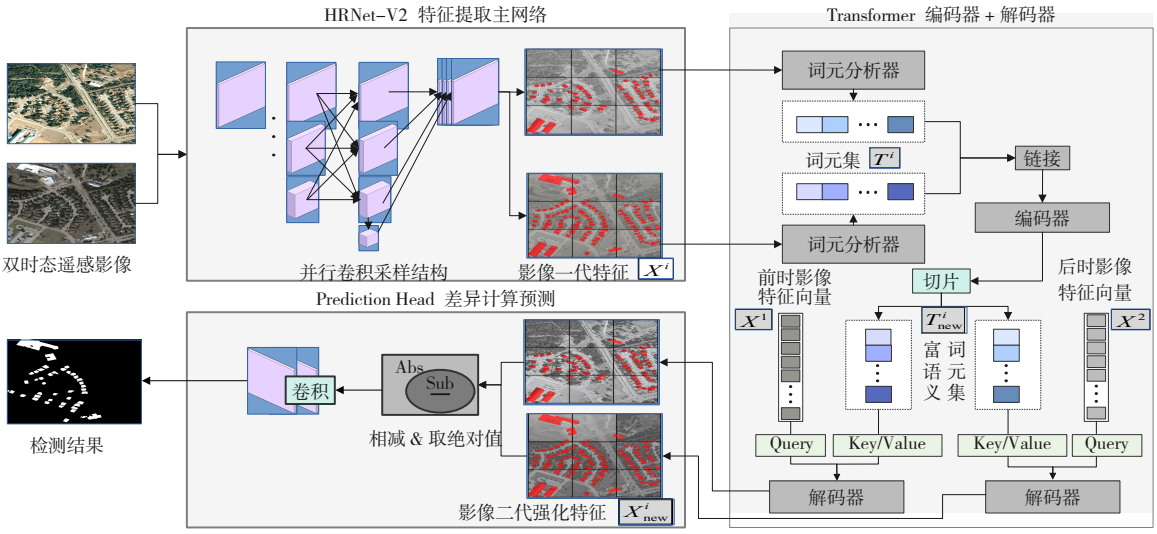


图1 H-BIT变化检测网络结构

1.1.1 HRNet-V2特征提取主网络

受HRNet[7]启发，本文采用多分辨率并行的HRNet-V2网络，对不同分辨率的特征图进行提取后再跨分辨率融合，从而解决传统串联网络提取语义高维信息时高分辨率空间信息被压缩、对小尺度目标特征提取能力有限的问题，保证在高分辨率特征图下较精确完整的空间位置信息，对位置敏感的语义分割任务友好。

如图2所示，HRNet-V2网络对图像特征进行并行提取，特征单元块通过下采样使语义聚合、上采样恢复高分辨率，再通过层间融合得到一代特征图 X^i，即可在提取高维语义信息的同时，保留高分辨率特征。

1.1.2 双时态图像自注意力变换网络Transformer

Transformer主要包括词元分析器、编码器和解码器，用于处理一代特征图。如图3所示，词元分析器可将HRNet-V2输出的一代特征图转换为词元，类似自然语言处理，将语句切分后用词元去表达，CHEN H[8]等设计了孪生形的词元集，将 X^1、X^2 两幅特征图送入同一个词元分析器网络分别进行卷积操作，将提取的特征图也分为 A^1、A^2 两套子图，并分别映射为 T^1、T^2 两套词元集。为了汇聚空间层面上的信

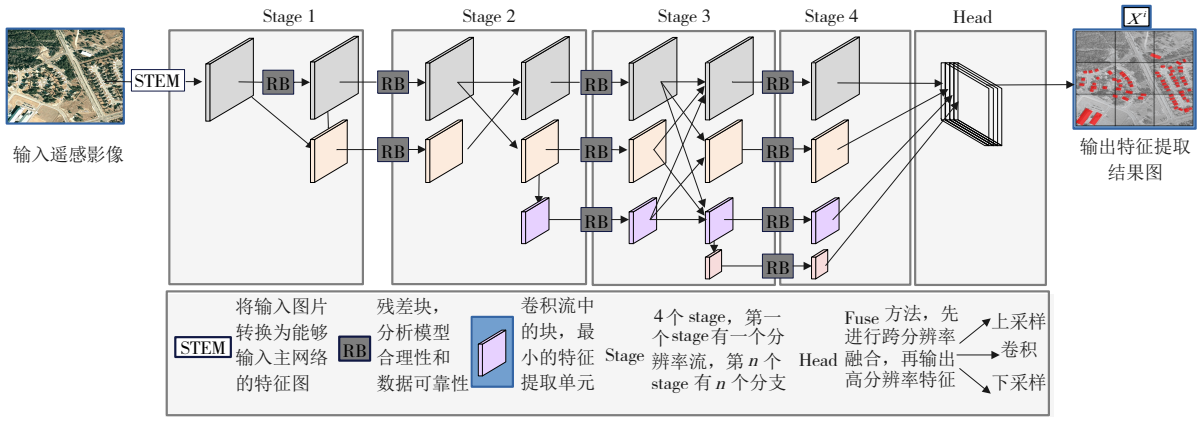


图2 HRNet-V2特征提取主网络结构

息，本文利用Woo S[9]等提出的Convolutional Block Attention Module中的空间注意力模块来压缩通道，得到语义丰富的紧凑词元。

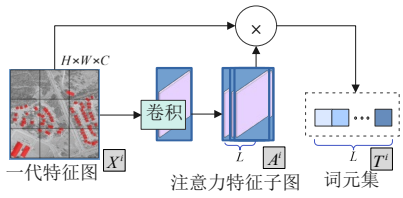


图3　词元分析器结构

编码器与Dosovitskiy A[10]等提出的ViT相似，但本文模块更小，注意力模块只堆叠了4次。如图4所示，将Transformer输出的两套词元集 T^1 、T^2 串联成为一个词元集后进行编码、范数计算，得到Query、Key和Value。多头自我注意力机制模块并行处理，串联后计算范数，送入多层感知机模块。多头自我注意力机制利用不同位置的多种子图信息表达，拥有异源信息共同作用的优势。多层感知机模块则通过GELU激活函数[11]连接两个线性转换层构成，最终将输出的词元集切片还原为富含高级语义信息的词元集 T^1_{new} 、T^2_{new} 。

解码器采用孪生网络形式，结合词元信息对原始特征进行改进。为了将编码器得到的富语义词元集重投影到像素空间，解码器利用原始影像特征向量 X^1 、X^2 每个像素与富语义词元集 T^1_{new} 、T^2_{new} 之间的关系，得到优化后的特征向量 X^1_{new} 、X^2_{new} 。如图4所示，解码器由多头交叉注意力模块和多层感知机模块组成，不同于Siamese解码器[12]，利用多头交叉注意力模块代替原始多头自我注意力模块，可有效避免像素间富连接造成的不良影响，此时Query、Key和Value分别来自不同的输入序列，即Query来自一代特征图，Key和Value来自富语义词元集。

1.1.3　差异计算预测 Predction Head

在最后的预测模块，Prediction Head对改进后得到的两幅特征图作差再取绝对值得到差异特征张量，利用FCN的思想，在分类器中通过两个卷积核大小为3×3的卷积层，得到最终预测输出的二值变化图。

1.2　损失函数

在训练阶段，为优化网络参数，选择最大程度降低交叉熵损失 L ，计算公式为：

$$L=\frac{1}{H_0\times W_0}\sum_{h=1,w=1}^{H,W} l\left(P_{hw},Y_{hw}\right) \tag{1}$$

式中，$l\left(P_{hw},y\right)=-\log\left(P_{hwy}\right)$ 为交叉熵损失函数；Y_{hw} 为位于 (h,w) 上像素的标签。

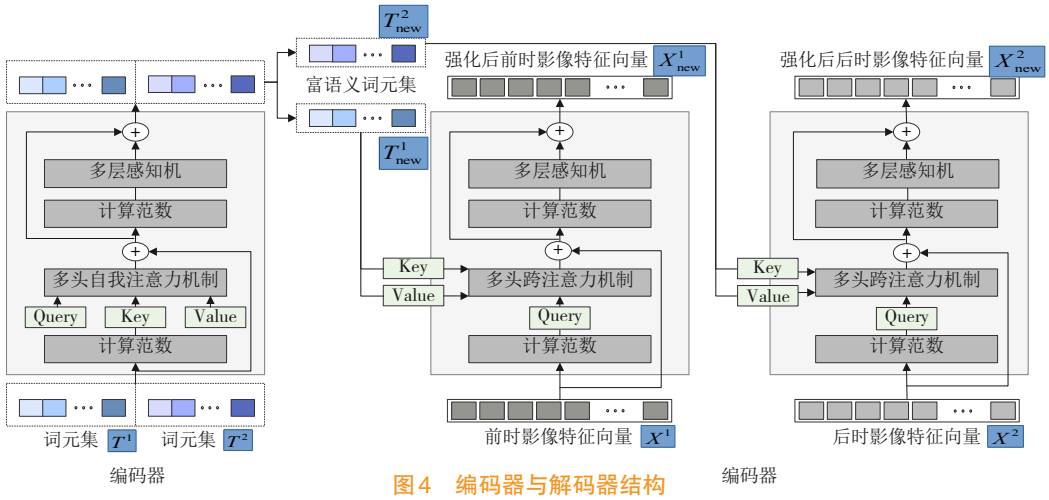


图4　编码器与解码器结构

1.3　网络训练

本文在PyTorch上实现了H-BIT方法。训练时采用动量梯度随机下降（SGD）优化器，动量设置为0.99，权重衰减设置为 5×10^{-4} ，学习率最大值max_lr=0.01，并在前30%的迭代轮次中进行学习率预热，从0缓慢上升到max_lr，再随迭代轮次逐步衰减到max_lr/75。每轮训练后对模型进行验证，验证集中的最佳模型用于测试集预测。实验在武汉大学超级计算中心的GPU服务器上完成模型训练，具体硬件配置为两块Intel(R) Xeon(R) E5-2640 v4 x86_64、24 GHz、20核心，4块Nvidia Tesla V100 16GB和128GB DDR4 2 400 MHz ECC，共迭代100轮，训练635幅遥感影像数据。模型每迭代一轮大约需要花费5 min。

2　实验结果与分析

2.1　数据集

LEVIR-CD数据集源于北航LEVIR团队的公开论文[8]，包含637对由航拍器采集的遥感影像对，时间跨度为5~14 a，每幅图像由近红外、红色和绿色3个波段组成，大小均为1 024像素×1 024像素，影像分辨率为

0.5 m，共有31 333个变化建筑实例作为数据集的变化信息，平均变化大小为987像素。由于GPU内存限制，将图像切成256×256的无重叠小尺寸图像块，按照训练集、验证集、测试集的划分，分别得到7 120对训练图像块、1 024对验证图像块和2 048对测试图像块。

2.2 评价指标

鉴于遥感影像变化检测可看作像素点的二分类问题，本文的评价指标选取机器学习统计学中用以衡量二分类模型精确度的F1得分（可兼顾分类模型的精确率和召回率）、精确率（precision）、召回度（recall）、交并比（IoU）、总体精度（*OA*）。

$$\text{precision} = TP/(TP+FP) \tag{2}$$

$$\text{recall} = TP/(TP+FN) \tag{3}$$

$$F1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \tag{4}$$

$$\text{IoU} = TP/(TP+FN+FP) \tag{5}$$

$$OA = (TP+TN)/(TP+TN+FN+FP) \tag{6}$$

TP 表示将正类预测为正类，*FP* 表示将负类预测为正类，*TN* 表示将负类预测为负类，*FN* 表示将正类预测为负类，构成变化检测混淆矩阵如表1所示。

表1 变化检测混淆矩阵

	预测为变化类	预测为未变化类
实际为变化类	*TP*	*FT*
实际为未变化类	*FP*	*TN*

2.3 实验结果

利用训练好的模型对测试集进行端对端的输出预测；在LEVIR-CD数据集上，将H-BIT方法与原始BIT方法进行比较实验，数据集和实验环境均相同。定量和定性结果如表2和图5所示，可以看出，H-BIT方法各项指标均优于原始BIT方法，说明H-BIT方法能有效提升模型变化检测精度；两种方法均能识别检测目标的变化，但H-BIT方法效果更好。图5中白色代表建筑变化，黑色代表未变化，黄色标记圈出了H-BIT方法的优点，与标准变化标签图相比，原始BIT方法存在错检、漏检现象，大型变化建筑检测结果结构丢失、部分内部有较大空洞，小型变化建筑检测结果边界模糊、形态缺陷；而H-BIT方法在复杂背景下也表现优越，建筑边界更平滑，去除了噪点影响，能完整提取目标、有效区分变化像素与伪变化像素，极少出现错检、漏检现象，对于不同尺度目标的提取均能发挥较好的效果，因此在定性结果上，H-BIT方法也优于原始BIT方法。

2.4 消融实验

为了获得性能最优的H-BIT模型，本文控制一些参数的设置，仍采用F1得分、precision、recall、IoU和*OA*五个评价指标，通过调整学习率预热所占总体

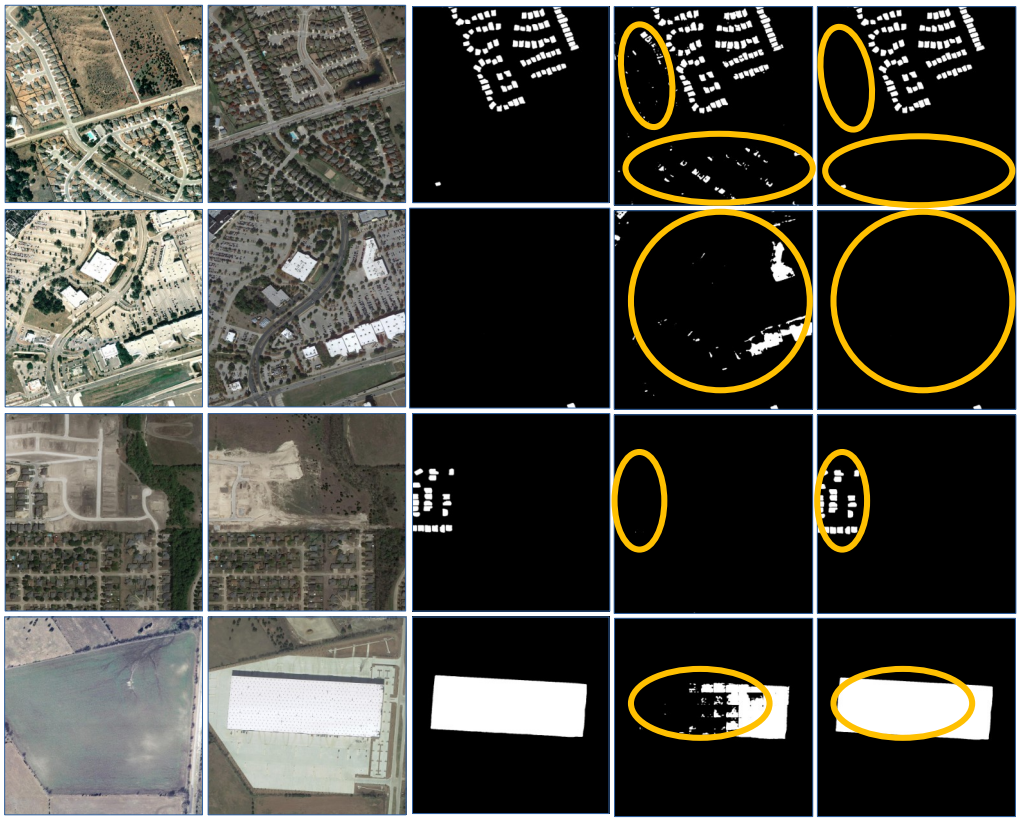


图5 H-BIT方法与原始BIT方法检测结果对比

表2　H-BIT与BIT方法的精度对比/%

方法	*F*1	precision	recall	IoU	*OA*
BIT	77.07	75.93	80.02	71.37	92.67
H-BIT	87.93	86.95	90.24	82.36	96.62

训练轮次的比例和Token_length，获得一个较稳定的模型。本文分别在预热比例为0.3、0.5、0.7和Token_length为4、6的情况下训练模型，具体结果如表3所示，可以看出，模型在预热比例为0.3和Token_length为6时的表现较好；预热比例为0.5和0.7时，虽然模型训练经过预热已经稳定，但剩下较少的训练轮次，模型无法取得足够的训练条件；模型在Token_length为6时，更加符合数据集语义信息，因此学习率预热比例为0.3，Token_length为6的H-BIT模型在LEVIR-CD上表现最优。

表3　消融实验结果统计

预热比例	Token_length	*F*1/%	precision/%	recall/%	IoU/%	*OA*/%
0.3	6	87.93	86.95	90.24	82.36	96.62
0.5	6	87.03	85.84	88.54	79.02	94.43
0.7	6	85.04	83.64	86.28	78.63	92.09
0.3	4	86.94	85.84	86.75	80.46	95.46
0.5	4	84.93	84.48	83.02	77.59	92.04
0.7	4	80.68	81.26	78.92	74.38	88.63

3　结　语

本文将遥感影像变化检测视为目标提取与差异计算两大块，基于BIT变化检测框架提出了一种改进后以HRNet为目标提取主干网络的H-BIT模型；在公开的遥感影像变化检测数据集LEVIR-CD上进行训练与测试；并与原始BIT模型进行定性和定量的测试对比。同时，为了获得性能最优的模型，本文对H-BIT模型进行了参数消融实验，结果表明：

1）以HRNet为目标提取的主干网络，在很大程度上降低了错检漏、检概率，提高了变化检测精度。H-BIT方法的precision、recall、F1得分和*OA*分别达到了86.95%、90.24%、87.93%和96.62%，较原始BIT方法分别提高了11.02%、10.22%、10.86%和3.95%。HRNet有利于目标边界的平滑处理，对于背景与目标模糊的影像也能完整提取出目标。

2）不同于原始卷积神经网络降低分辨率的做法，经HRNet网络提取的特征图像融合了高分辨率和高语义信息两大优势，高分辨率有利于后续差异计算，对小目标的变化也很敏感，即使影像复杂，存在多种尺度目标，H-BIT方法也能表现出较高的变化检测性能。

3）相较于原始方法，H-BIT方法的参数数量和计算量更少，训练速度更快，拥有时间优势。

本文提出的基于HRNet的H-BIT模型仍存在一些不足，如训练样本变化检测目标种类单一等，下一步将获取拥有不同种类目标的数据进行检测，以进一步提高该方法的目标可拓展性。

参考文献

[1] 袁敏,肖鹏峰,冯学智,等. 基于协同分割的高分辨率遥感图像变化检测[J]. 南京大学学报(自然科学),2015,51(5):1 039-1 048

[2] 肖明虹,冯文卿,眭海刚. 超像素分割和多方法融合的遥感影像变化检测方法[J]. 测绘通报,2018(10):93-97

[3] LYU H, LU H, MOU L. Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection[J]. Remote Sensing,2016,8(6):506

[4] GONG M, ZHAN T, ZHANG P, et al. Superpixel-based Difference Representation Learning for Change Detection in Multispectral Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5):2 658-2 673

[5] 孙红岩. 基于多特征融合的遥感图像特征提取方法[J]. 计算机仿真,2016,33(10):334-337

[6] 赵祥,王涛,张艳,等. 基于改进DeepLabv3+孪生网络的遥感影像变化检测方法[J]. 地球信息科学学报,2022,24(8):1 604-1 616

[7] SUN K, XIAO B, LIU D, et al. Deep High-resolution Representation Learning for Human Pose Estimation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:5 693-5 703

[8] CHEN H, QI Z, SHI Z. Remote Sensing Image Change Detection with Transformers[J]. IEEE Transactions on Geoscience and Remote Sensing,2022,60:1-14

[9] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[C]//Proceedings of the European Conference on Computer Vision (ECCV),2018:3-19

[10] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale[C]. International Conference on Learning Representations,2021

[11] Hendrycks D, Gimpel K. Gaussian Error Linear Units(GELUs)[M]. arXiv,2020

[12] Vaswani A, Shazeer N, Parmar N, et al. Attention is All You Need[C]. Advances in Neural Information Processing Systems: 30. Curran Associates, Inc. ,2017

作者简介：傅绘锦（2001—），主要研究方向为地理计算与服务。