

关注语义一致性的遥感影像变化检测

吴虎承¹, 王仁芳^{1,2}, 邱虹², 王峰², 高广^{2,3}, 吴敦^{2,3}

1. 中国海洋大学 计算机科学与技术学院, 青岛 266100;

2. 浙江万里学院 大数据与软件工程学院, 宁波 315100;

3. 宝略科技(浙江)有限公司, 宁波 315042

摘要: 遥感影像语义变化检测在生态环境、土地利用、土地覆盖监测等方面发挥着重要作用。近年来基于深度学习的变化检测方法是遥感智能解译关注的热点, 然而现有的三分支语义变化检测方法缺少对变化分支和语义分支一致性建模, 导致双时相语义变化检测的自相矛盾。针对该问题, 本文提出一种基于孪生 CNN 与 Transformer 的遥感影像语义变化检测算法。在编码阶段, 首先设计孪生 ResNet34 网络提取影像的多尺度特征, 并嵌入差异增强模块来提高变化信息的关注度; 然后利用语义标记器将特征图映射为紧凑的语义 Token, 并通过 Transformer 编码器联合双时相语义和变化信息来建模“语义—变化”一致性。在解码阶段, 通过 Transformer 解码器利用跳跃连接, 将不同细粒度的语义信息融合, 生成细化的语义特征图; 经过上采样恢复、掩码相乘得到双时相语义变化结果。在遥感语义变化检测公开数据集 SECOND 和 Landsat-SCD 上的实验结果表明, 本文提出的算法能够有效地关注变化区域, 保持变化结果和语义结果的一致性, 且达到了优异的评价指标和视觉效果。

关键词: 遥感影像, 变化检测, 语义一致性, 差异增强, 多尺度特征, 孪生网络, ResNet34, Transformer

中图分类号: P2

引用格式: 吴虎承, 王仁芳, 邱虹, 王峰, 高广, 吴敦. 2025. 关注语义一致性的遥感影像变化检测. 遥感学报, 29(4): 1031–1045

Wu H C, Wang R F, Qiu H, Wang F, Gao G and Wu D. 2025. Semantic consistency assisted change detection in remote sensing images. National Remote Sensing Bulletin, 29(4): 1031–1045 [DOI:10.11834/jrs.20243419]

1 引言

遥感影像变化检测 (Shi 等, 2020; Wang 等, 2023; 杨彬 等, 2023) 是通过对不同时间在同一区域获取的一对 (或多幅) 遥感影像进行联合分析来获取土地利用/土地覆盖 LCLU (Land Cover and Land Use) 类型变化信息的技术, 已被广泛应用于测绘制图 (周承乐 等, 2024; 柳思聪 等, 2023)、城市规划 (Luo 等, 2018)、国土空间规划 (Lv 等, 2022)、资源环境、各类动态监测 (Li 等, 2021) 等领域。变化检测通常分为二值变化检测 BCD (Binary Change Detection) (刘英 等, 2024) 和语义变化检测 SCD (Semantic Change Detection) 任务 (Zhu 等, 2022; Yuan 等, 2022)。BCD 旨在区分变化区域和未变化区域的像素, 主要关注变

化目标的定位与几何形状的刻画; 而 SCD 不仅需要区分变化区域, 还需要进一步识别变化区域中的不同类别, 同时完成定位和分类任务。

传统的语义变化检测方法主要是分类后的变化检测 (Bruzzone 和 Serpico, 1997; El-Hattab, 2016; 朱传海 等, 2023), 首先对地物类别进行分类, 生成 LCLU 语义分类图, 然后计算每个类别之间的相互转化统计信息, 其主要局限是双时相 LCLU 分类图独立产生且忽略了双时相遥感图像间的时间相关性。为了提高语义变化的分类准确性, Bruzzone 等 (1999) 采用贝叶斯先验概率估计来衡量语义类别的转移概率, Wu 等 (2017) 利用慢特征分析来计算变化概率。Singh 等 (1989) 提出了多数据直接分类的 SCD 方法, 对双时相遥感影像联合训练, 解决忽略了时间相关性的问题, 但需

收稿日期: 2023-10-16; 预印本: 2024-04-15

基金项目: 国家自然科学基金 (编号: 61906170); 浙江省科技计划项目 (编号: ZCLY24F0301); 宁波市科技计划项目 (编号: 2022Z233, 2021Z050, 2022S002, 2023J403)

第一作者简介: 吴虎承, 研究方向为遥感智能解译。E-mail: wuhucheng1@163.com

通信作者简介: 王仁芳, 研究方向为数字图像处理、机器学习、遥感智能解译。E-mail: renfang_wangac@126.com

要大量的训练数据来模拟所有可能的类型转换。Demir等(2012)引入复合分类技术,计算并最大化类型转换的后验联合概率,通过对转换概率的迭代半监督估计来建模两个遥感图像之间的时间依赖性,避免了对每个可能的转换进行训练数据。上述分类后变化检测方法的准确性取决于语义信息提取和变化识别算法,受制于阈值选择和误差积累问题,检测效果并不理想。

近年来,深度学习因其强大的特征学习能力,在遥感影像SCD中得到了广泛关注。Daudt等(2018)率先引入一种三分支结构的SCD框架,两个语义分割分支将双时相图像分割成LCLU映射,变化分支负责检测变化信息。Ding等(2022)综述分析了适用于SCD的4种架构,通过大量实验证明了将语义分割任务和变化检测任务进行后期融合的方法(即SSCD-I三分支架构)是最适合SCD的网络结构。基于三分支架构,Yang等(2022)提出了一个不对称孪生网络ASNet,通过在解码器中引入门控和加权设计来改进特征表示,从而提高模型定位和识别语义变化的能力。Xiang等(2021)提出了一种端到端的双任务SCD网络,利用生成式变化域模块引导双时语义分支进行定位和分割。Zhao等(2022)提出了一种端到端的空间和语义增强的孪生网络SSESN,通过空间和语义特征聚合模块增强VHR图像的空间信息和语义信息,并采用变化感知模块解耦聚合特征,将二值变化分支中的特征融合到语义分支提高检测的准确率。ChangeMask(Zheng等,2022)将SCD解耦为时序语义分割和BCD,然后将这两个任务整合为一个通用的编码器—转换器—解码器框架。在编码器部分,设计了一个语义感知编码器来建模语义变化的因果关系,然后通过后期的Transformer模块从语义表示中学习变化表示,同时引入正则化来降低过拟合的风险。MCDNet(Deng等,2022)将特征引导策略引入多任务变化检测网络中,实现任务间的相互引导同时进行变化信息的整合和增强。上述的三分支结构在一定程度上解决了定位和分割问题,但普遍存在以下问题:对差异变化信息的关注不够,难以从复杂背景中忽略无关伪变化;双时相语义特征被单独嵌入,未考虑双时相图像语义的时间相关性;三分支网络中语义分支和变化分支之间的联系较弱,导致双时相预测结果可能存在矛盾。

针对上述问题,受语义分割中基于CNN和Transformer的网络模型性能表现常常超过单纯的CNN模型和Transformer模型的启发,本文从增强差异信息以及时空依赖性建模两个方面入手,提出一种孪生CNN与Transformer相结合的遥感影像SCD网络模型(简称ReTNet)。针对变化信息,构造差异增强模块来增强多尺度的双时特征图,从而有效地增强双时图像的差异特征;针对双时相图像之间的时间相关性和“语义—变化”一致性,引入Transformer将差异信息和双时语义信息映射成语义Token并融合,在融合Token空间中对整个时空域的“语义—变化”信息进行联合建模,从而有效地建模图像中程长程依赖及双时相上下文相关性。

2 研究方法

在本节中,介绍本文提出的基于孪生CNN与Transformer的遥感影像语义变化检测方法的整体网络模型结构、设计的孪生ResNet34模块、差异增强模块DE(Difference Enhancement)、Transformer编码器、语义标记器、Transformer解码器和构造的联合损失函数。

2.1 网络框架

基于孪生CNN与Transformer的遥感影像语义变化检测模型主要包括3部分:用于多尺度特征提取的孪生ResNet34主干模块、用于“语义—变化”一致性建模的双时Transformer模块以及轻量级的CNN预测头。

孪生ResNet34结构如图1中CNN backbone部分所示,为了增强网络对变化特征的关注,孪生ResNet34特征提取模块相邻层之间嵌入设计DE其结构如图2所示。双时Transformer模块的结构如图1中Bitemporal Transformer所示,主要由语义标记器(Semantic Tokenizer)、Transformer编码器和Transformer解码器3部分构成,语义标记器将特征提取模块输出的双时语义特征和变化特征进行紧凑映射,转化为语义Token,Transformer编码器将双时语义特征和变化特征级联后在整个语义空间下联合建模,构建双时语义关联和“语义—变化”一致性。经过使用跳跃连接融合策略的Transformer解码器对三分支的特征进行语义增强,并将语义Token映射像素空间获得3个分支的细化特征。最

后经过轻量级的预测头进行上采样恢复和分类, 得到双时语义分割图和二值变化检测图, 并掩码

相乘获得最终的双时相语义变化图, 其结构如图1中 Predict Head 所示。

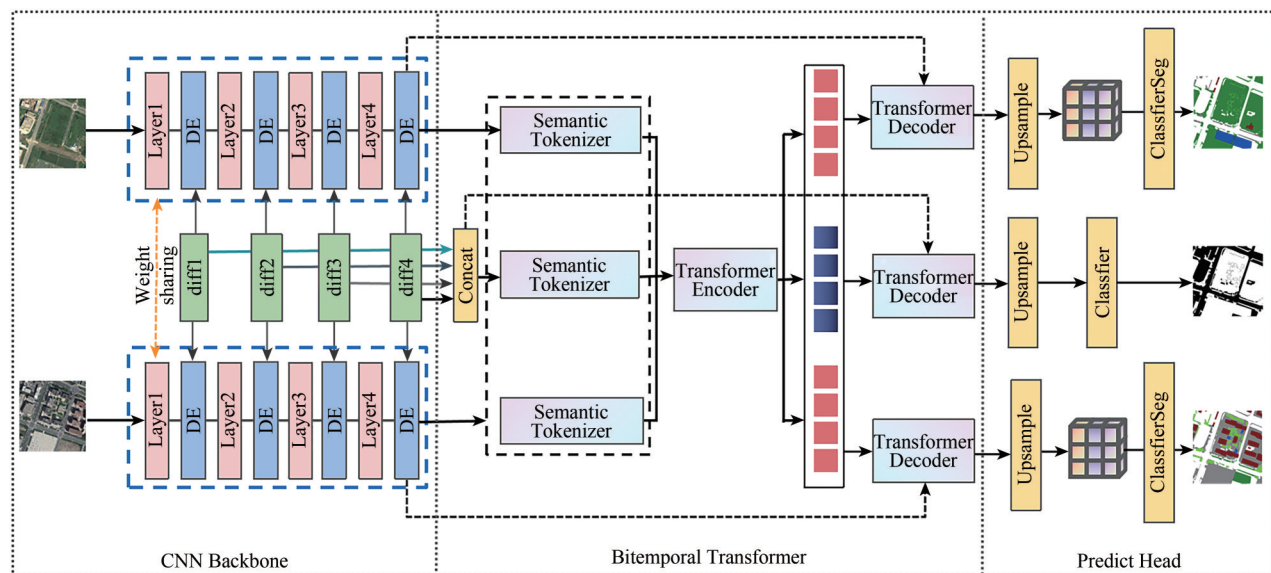


图1 ReTNet网络结构

Fig. 1 ReTNet network architecture

2.2 差异增强模块

在SCD中, 因背景占大部分使得网络更多地关注了无变化区域, 鉴于此本文设计差异特征增强模块, 使网络更多关注变化区域, 如图2所示, 由平均池化和最大池化以及多层感知器组成。首

先双时相特征图相减得到差异特征图, 公式为

$$D_i = |F_{i1} - F_{i2}| \quad (1)$$

式中, F_{i1} 是第一时相第 i 层的特征图, F_{i2} 是第二时相第 i 层的特征图。

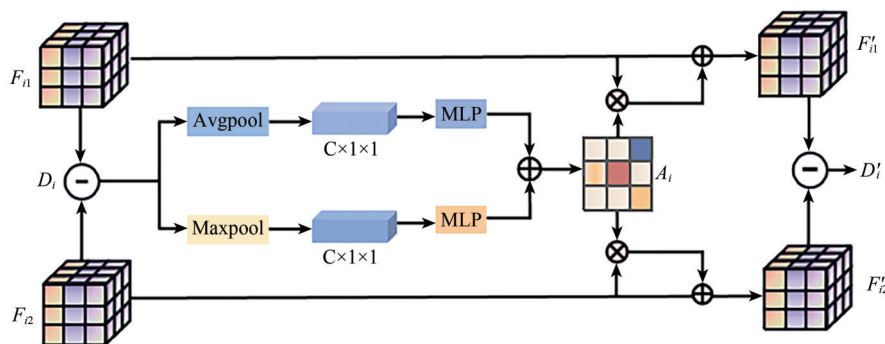


图2 差异增强模块

Fig. 2 Difference enhancement module

直接获得的差异特征图通常存在严重的噪声, 不利于网络检测变化, 为此本文充分地利用差异特征图的信道特征, 对其进行通道注意力操作:

$$A_i = \sigma(\text{MLP}(\text{Avgpool}(D_i)) + \text{MLP}(\text{Maxpool}(D_i))) \quad (2)$$

式中, MLP为多层感知器, Avgpool和Maxpool分别为平均池化和最大池化, σ 为Sigmoid函数。通过平均池化和最大池化操对差异特征进行压缩操作Avgpool有助于保留更多的全局信息, 而Maxpool

有助于捕捉局部细节, 结合二者的信息可以提高对图像变化和干扰具有不同的敏感性, 使其对于各种变化和季节光照等干扰更具鲁棒性。然后经过共享权重的多层感知器网络并进行元素求和操作得到通道注意力图。将双时相的语义特征分别与通道注意力图相乘后相加, 得到增强后的语义特征和差异特征。

2.3 语义标记器

语义标记器将像素分组为高级语义，为两个不同时间输入特征生成一组紧凑的语义标记，将得到的语义特征图分为若干个子图，通过空间注意力提取空间信息，将特征子图映射成Token，从而转化为具有高维语义信息的Token。具体过程是将特征图送入一个 1×1 的卷积核得到通道数为 L 的Attention maps，然后通过Softmax操作得到 L 个通道数为 C 的语义Tokens，用公式表示为

$$\mathbf{AM} = \text{Softmax}(\text{reshape}(\text{Conv}(\mathbf{F}))) \quad (3)$$

$$\text{SemanticToken} = \text{reshape}(\mathbf{F}) \times \mathbf{AM} \quad (4)$$

式中， $\mathbf{AM}$ 表示注意力图，Conv表示 1×1 卷积， $\mathbf{F}$ 表示孪生ResNet34特征提取模块输出的语义特征图（图3）。

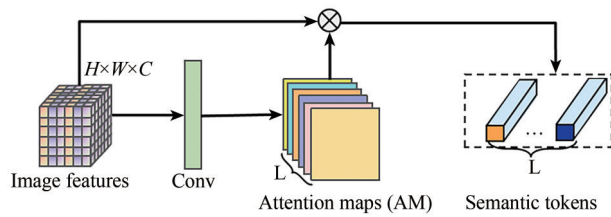


图3 语义标记器

Fig. 3 Semantic tokenizer module

2.4 Transformer 编码器

本文设计的Transformer编码器主要有位置编码模块（Positional Encoding）、多头自注意力MSA（Multi-head Self-Attention）和多层感知器模块构成，如图4所示。

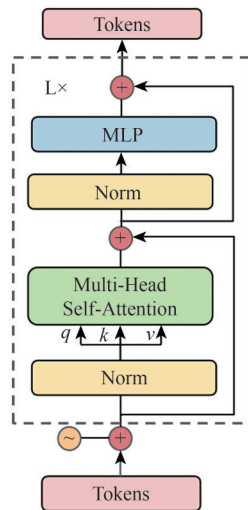


图4 Transformer 编码器

Fig. 4 Transformer encoder

融合语义标记器输出的两个语义Token和差异Token，传入Transformer编码器，首先添加位置编码信息：

$$\text{Tokens} = \text{Cat}(\text{token}_1, \text{token}_2, \text{token}_{\text{diff}}) + E_{\text{pos}} \quad (5)$$

式中， token_1 , token_2 , $\text{token}_{\text{diff}}$ 分别表示双时相语义Token和差异Token， E_{pos} 表示嵌入的位置编码。

经过多层多头注意力和多层感知器，计算双时相语义和差异特征之间的相关性，获得包含“语义—变化”一致性的全局上下文语义信息。第 l 层的输入为

$$\mathbf{Z}' = \text{MSA}(\text{Norm}(\mathbf{Z}_{i-1})) + \mathbf{Z}_{i-1} \quad (6)$$

$$\mathbf{Z}_i = \text{MLP}(\text{Norm}(\mathbf{Z}'_i)) + \mathbf{Z}'_i \quad (7)$$

式中，Norm为层归一化算子， $\mathbf{Z}_i$ 为Transformer编码器第 i 层编码特征， $i \in \{1, 2, \dots, L\}$ ，MSA是多头自注意力算子

$$\text{MSA} = \text{MSA}(T^{(i-1)}) = \quad (8)$$

$$\text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n) \mathbf{W}_0$$

$$\text{head}_j = \text{Att}(\mathbf{Q}_j, \mathbf{K}_j, \mathbf{V}_j) \quad (9)$$

式中， head_j 表示第 j 个自注意力头的输出， $\mathbf{Q}_j$, $\mathbf{K}_j$, $\mathbf{V}_j$ 是第 j 个自注意力头的权重矩阵， $\mathbf{W}_0$ 表示可训练的权重矩阵， n 表示自注意力头的总数。Att是自注意力算子

$$\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \sigma\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right) \times \mathbf{V} \quad (10)$$

式中， σ 表示softmax激活层， d 表示 $\mathbf{Q}$ 和 $\mathbf{K}$ 权重矩阵的维度。在Transformer结构中，每个多头自注意力模块由多个自注意力层构成，多头自注意力层的输入为 $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ 矩阵，可通过权重矩阵 $\mathbf{w}^q$ 、 $\mathbf{w}^k$ 、 $\mathbf{w}^v$ 分别与上一层Transformer块输出相乘获得

$$\begin{aligned} \mathbf{Q} &= T^{(i-1)} \mathbf{w}^q \\ \mathbf{K} &= T^{(i-1)} \mathbf{w}^k \\ \mathbf{V} &= T^{(i-1)} \mathbf{w}^v \end{aligned} \quad (11)$$

式中， $T^{(i-1)}$ 表示 $i-1$ 层Transformer块的输出。

2.5 Transformer 解码器

在Transformer Decoder中，本文将ViT（Dosovitskiy等，2021）中的多头自注意力替换为多头交叉注意力，构造出由M层的多头交叉注意力MA（Multi-Head Cross-Attention）和多层感知器MLP（Multi-Layer Perceptron）模块组成的Transformer Decode，移除了MSA块避免像素之间密集关系的大量计算，交叉注意力允许模型在不同的表示空间之间建立关联，能够提高模型对复

杂关系的建模能力。将含有丰富信息的双时语义 Token 和孪生 ResNet34 骨干提取的深层图像特征输入 Transformer 解码器, 得到细化后的双时语义特征和差异特征。经过 Transformer Encoder 后, 双时相图像获得了全局语义信息丰富的标签 Token, 这些富含全局信息的标记包含紧凑的高级语义信息, 很好地揭示了感兴趣的变化, 并将抽象的语义表示投影回像素空间获得像素级的特征。使用改进的 Transformer 解码器来细化双时相的图像特征。

如图 5 所示, 将图像特征作为查询 (Q), 将增强后的语义 Token 作为键和值 (K/V), 在每层 l 上的 MA 的计算过程为

$$\mathbf{MA}_l = \text{MA}(X^{i,(l-1)}, T_{\text{new}}^i) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) \mathbf{w}^o \quad (12)$$

$$\text{head}_j = \text{Att}(X^{i,(l-1)} \mathbf{w}_j^q, T_{\text{new}}^i \mathbf{w}_j^k, T_{\text{new}}^i \mathbf{w}_j^v) \quad (13)$$

式中, $\mathbf{MA}_l$ 表示第 l 上 MA 的输出, $\mathbf{w}_j^q$ 、 $\mathbf{w}_j^k$ 、 $\mathbf{w}_j^v$ 、 $\mathbf{w}^o$ 为线性投影矩阵, h 为注意力头的数量, $i = \{1, 2\}$ 表示时相一和二。然后经过层归一化和多层感知器

$$X^{i,l} = \text{MLP}(\text{Norm}(\mathbf{MA}_l)) + \mathbf{MA}_l \quad (14)$$

式中, $l \in \{1, 2, \dots, M\}$ 表示注意力块的层数, Norm 表示层归一化算子。

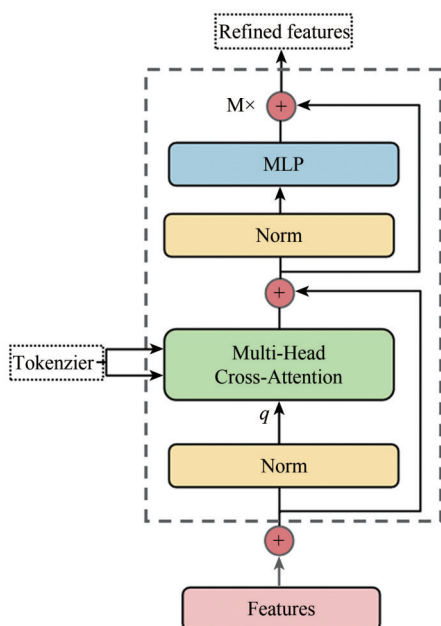


图 5 Transformer 解码器

Fig. 5 Transformer decoder

2.6 损失函数

本文提出的 ReTNet 网络的联合损失函数包括语义损失、二值变化损失和语义一致性损失, 两

个语义分割分支语义损失使用交叉熵损失

$$L_{\text{seg}} = -\frac{1}{n} \sum_{i=1}^N y_i \log(p_i) \quad (15)$$

式中, N 为语义类别数目, y_i 和 p_i 分别表示第 i 个类别的 Ground Truth 和预测概率。二值变化损失使用二值交叉熵损失, 计算公式为

$$L_{\text{change}} = -y_c \log(p_c) - (1 - y_c) \log(1 - p_c) \quad (16)$$

式中, y_c 和 p_c 分别表示第 i 个类别的 Ground Truth 和预测概率。此外引入一个语义一致性损失, 将双时相语义和变化信息进行了对齐, 采用余弦相似度函数, 计算为

$$L_{\text{sc}} = \begin{cases} 1 - \cos(x_1, x_2), & y_c = 1 \\ \cos(x_1, x_2), & y_c = 0 \end{cases} \quad (17)$$

式中, x_1 和 x_2 分别为两个语义图相同位置点的特征向量, y_c 是像素在二值变化图上相同位置的值。网络的总体损失函数为

$$L_{\text{scd}} = L_{\text{seg}} + L_{\text{change}} + L_{\text{sc}} \quad (18)$$

3 数据结果处理与分析

为了验证提出方法的有效性, 本文与主流的语义变化检测方法 HRSCD-str3、HRSCD-str4、Siam-UNet、Siam-Deeplabv3+、SSCDI、Bi-SRNet 在 SECOND 数据集和 Landsat-SCD 数据集上进行了实验比较。

3.1 数据集

本文实验数据利用用于高分辨率遥感影像中 SCD 大型的、有良好注释的基准数据集 SECOND (Yang 等, 2022), 该数据集从多个平台和传感器上收集了 4662 对分布在杭州、成都和上海等城市的航空图像, 每幅图像的大小为 512×512 像素, 并在像素级别上进行标注。在标注的标签中, 有 1 个无变化类和非植被地表、树木、低植被、水、建筑物、操场等 5 个 LC 类。本文采取翻转、旋转等操作扩增样本, 实验中训练集和测试集样本比设为 4:1。

另一个实验数据为 Landsat-SCD 数据集 (Yuan 等, 2022), Landsat-SCD 中的 RSIs 已被注释为无变化类和 4 个 LC 188 类, 包括农田、沙漠、建筑和水体 (被注释的变化区域仅有 189 个)。Landsat-SCD 数据有 1908468 个 RSI, 空间分辨率为 416×416。按照 8:1:1 的数值比例将 RSI 拆分为训练集、验证集和测试集, 分别有 7468 个、560 个和 560 个 RSI。

3.2 训练环境与参数配置

本文使用 NVIDIA GeForce RTX 3060 12 GB GPU 在工作站上进行网络训练和测试, 所有对比实验测试设置相同的超参数 (表1), 批处理大小设为2, 训练迭代次数设置为100, 初始学习率设置为0.01, 采用随机梯度下降 (SGD) 方法对权值进行了优化。

表1 实验环境及超参数

Table 1 Experimental environment and hyperparameters

环境	型号	超参数	超参值
CPU	Intel Core i7-12700	学习率	0.00015
GPU	NVIDIA GeForce RTX 3060	Epoch	100
内存	16 G	参数优化	SGD
操作系统	Windows 11	动量	0.9
深度学习框架	Pytorch	批大小	2
CUDA	11.6		
语言	Python 3.7		

3.3 评价指标

本文采用总体精度 OA (Overall Accuracy)、平均交并比 mIoU (mean Intersection over Union)、Sek 系数 Sek (Separated Kappa Coefficient) 和用于 SCD 任务的 F1 分数 (F_{scd}) 来对网络的预测结果进行评价。

假设 $Q = \{q_{ij}\}$ 是混淆矩阵, 其中 q_{ij} 表示真实类别为 j 类, 预测为第 i 类的像素数, 其中 $i, j \in \{0, 1, 2, \dots, N\}$, 0 表示无变化, 1— N 分别为不同的 LC 类。

OA 表示正确分类像素与总的图像像素的数值比, 可定义为

$$OA = \frac{\sum_{i=0}^N q_{ii}}{\sum_{i=0}^N \sum_{j=0}^N q_{ij}} \quad (19)$$

mIoU 通过求未改变类和改变类的 IoU 的平均值来确定为

$$mIoU = \frac{IoU_{nc} + IoU_c}{2} \quad (20)$$

式中, IoU_{nc} 、 IoU_c 分别表示未改变类、改变类的交并比, 通过混淆矩阵计算获得

$$IoU_{nc} = \frac{q_{00}}{\sum_{i=0}^N q_{i0} + \sum_{i=0}^N q_{0i} - q_{00}} \quad (21)$$

$$IoU_c = \frac{\sum_{i=1}^N \sum_{j=1}^N q_{ij}}{\sum_{i=0}^N \sum_{j=0}^N q_{ij} - q_{00}} \quad (22)$$

F_{scd} 用于评估变化区域的分割精度, 其计算公式为

$$F_{scd} = \frac{2 \times P_{scd} \times R_{scd}}{P_{scd} + R_{scd}} \quad (23)$$

其中

$$P_{scd} = \frac{\sum_{i=1}^N q_{ij}}{\sum_{i=1}^N \sum_{j=0}^N q_{ij}} \quad (24)$$

$$R_{scd} = \frac{\sum_{i=1}^N q_{ij}}{\sum_{i=0}^N \sum_{j=1}^N q_{ij}}$$

Sek 评估语义类的分类情况, 特别是在变化区域上的语义类, 根据变式后的混淆矩阵 $Q' = \{q'_{ij}\}$ 来计算, 其中 $q'_{i,j} = q_{ij}$, $q'_{00} = 0$, 其计算为

$$Sek = Kappa \times e^{IoU_c - 1} \quad (25)$$

式中

$$Kappa = \frac{P_o - P_e}{1 - P_e},$$

$$P_o = \frac{\sum_{i=0}^N q'_{ii}}{\sum_{i=0}^N \sum_{j=0}^N q'_{ij}}, \quad (26)$$

$$P_e = \frac{\sum_{i=0}^N (\sum_{j=0}^N q'_{ij} \times \sum_{j=0}^N q'_{ji})}{(\sum_{i=0}^N \sum_{j=0}^N q'_{ij})^2}$$

采用准确率 P (Precision)、召回率 R (Recall)、F1 值、交并比 IoU (Intersection over Union)、总体精度 OA (Overall Accuracy)、Kappa 系数 (Kappa Coefficient) 作为衡量二值变化检测结果质量的指标, 对各个网络产生的二值变化图进行定量评价, 来评价网络的二值变化检测能力是否是制约它整个语义变化检测的因素。

3.4 对比试验

首先对 6 种对比模型方法的特点做简要的说明, 如何对比分析。(1) HRSCD-str3。双分支的解耦变化检测和语义分割孪生网络。为了避免网络的输入进行早期融合, 造成的信息冗余, 采用了后融合策略, 让网络提取更适配的变化特征, 提高变化检测的性能。(2) HRSCD-str4。双分支的耦合变化检测和语义分割孪生网络。其主要改进在于融合语义分割网络编码器的双时特征图和 BCD 网络编码器的特征图。融合后的特征图被送入 BCD 网络的解

码器, 以获得更好的二进制变化图, 从而提高 SCD 的性能。(3) Siam-UNet。采用孪生结构的 UNet++-CD 网络模型。通过多尺度特征融合, 提高网络对多尺度变化的敏感度。(4) Siam-Deeplabv3。采用孪生 DeeplabV3+ 的变化检测模型。通过添加 ASPP 模块, 增大网络感受野, 提升对多尺度特征的融合能力, 无关伪变化的影响。(5) SSCDI。三支的解耦后融合策略的 CD 模型, 将提取后的双时语义特征进行融合, 送入单独的二值变化检测单元进行二值变化检测。二值变化检测的结果与双时语义分割结果进行相乘, 得到最终语义变化检测结果。这种结构进一步提升了变化检测的性能。(6) Bi-SRNet。基于三支解耦变化检测和语义分割的方法, 采用交叉 Non-local 机制增加双时语义分支和变化分支之间的通信, 建立双时语义关联, 从而提高对变化类型的识别能力。上述算法虽然取得了不错的性能指标, 但是均没有考虑双时语义关联和分割任务与二值分类任务的语义一致性建模, 导致变

化检测结果与双时语义分割结果的自相矛盾现象。

7 种算法实验在 SECOND 数据集和 Landsat-SCD 数据集上的可视化结果如图 6 和图 7 所示。从左到右, 每一列分别表示真实标签、HRSCD-str3 检测结果、HRSCD-str4 检测结果、Siam-UNet 检测结果、Siam-Deeplabv3+ 检测结果、SSCDI 检测结果、Bi-SRNet 检测结果。HRSCD-str3 和 HRSCD-str4 方法 (Daudt 等, 2018) 的预测结果出现大面积错检和误检, 这两种方法对复杂环境下的伪变化的判别缺乏鲁棒; Siam-UNet (Moustafa 等, 2021)、Siam-Deeplabv3、SSCDI (Ding 等, 2022)、Bi-SRNet (Ding 等, 2022) 这 4 种方法虽然对无关变化的鲁棒性提高, 但是仍然会存在如图 6 和图 7 矩形框中显示的双时相预测结果的自相矛盾问题 (即预测为变化区域, 但是双时相结果中该区域变化前和变化后的地物类型却相同)。而本文提出的算法, 不仅减少了双时相语义变化预测的自相矛盾现象, 并且在变化边缘的预测上更为精细。

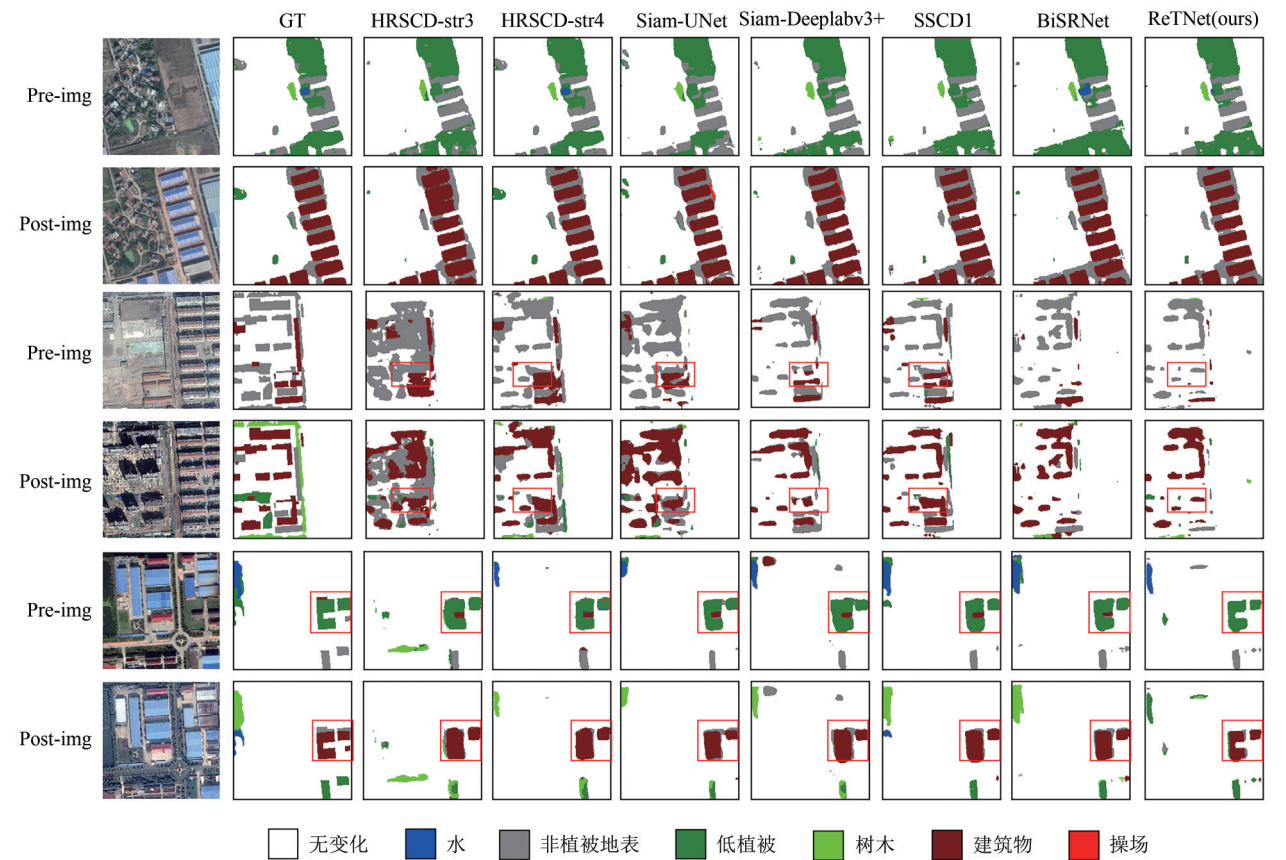


图 6 SECOND 上对比实验结果图

Fig. 6 Comparison of experimental results on the SECOND dataset

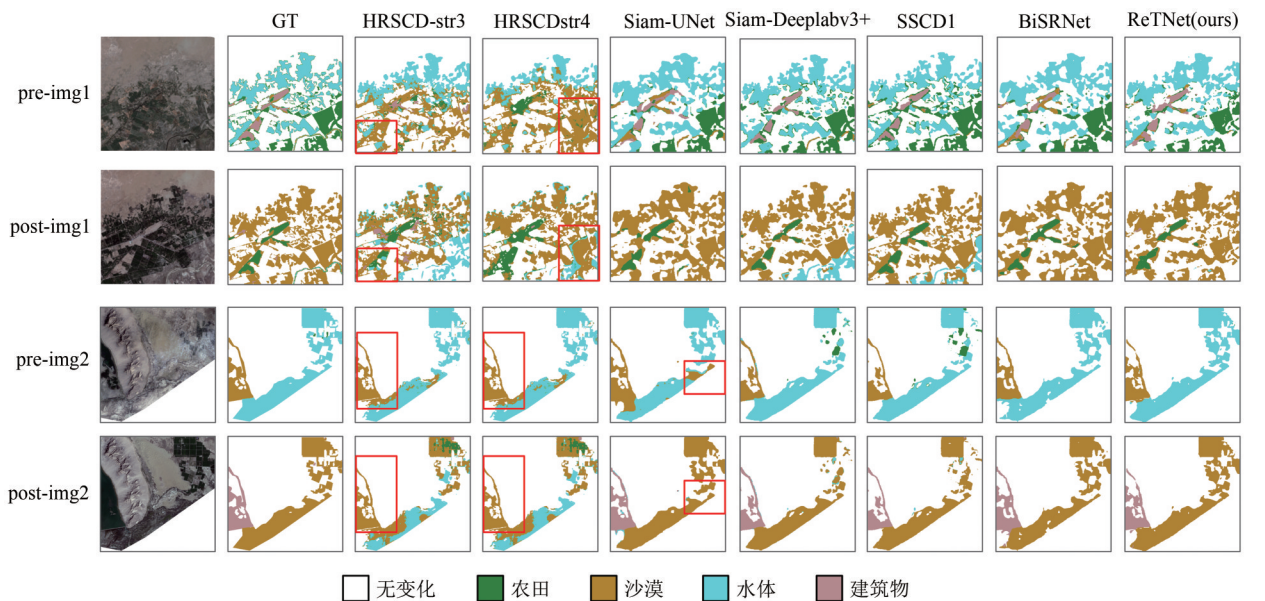


图7 Landsat-SCD数据集上对比实验结果图

Fig. 7 Comparison of experimental results on the Landsat-SCD dataset

7种算法的在SECOND数据集和Landsat-SCD数据集上语义变化检测4个评价指标如表2和表3所示。从表2中可以看出本文提出的ReTNet方法在SECOND数据集上的评价指标OA、mIoU、Sek、 F_{scd} 分别为86.79%、70.24%、17.84%、58.00%，在对比的几个方法在均达到最佳效果。如表3所示，在Landsat-SCD数据集上的评价指标OA、mIoU、Sek、 F_{scd} 分别为91.40%、83.52%、44.44%、78.81%在对比的几个方法均达到最佳效果。同时在两个数据集上的预测结果的“语义—变化”非一致率均有所降低。这主要得益于差异增强模块的嵌入使得网络更加敏锐地关注于变化信息；在CNN-Transformer混合架构中，局部和全局信息得以有机地交织，为网络提供了联合建模的能力；Transformer编码器对语义标记和差异特征标记的联合特征编码实现了整个时空域“语义—变化”一致性的建模，进一步增强了语义变化检测的性能。

语义变化检测任务被解耦为语义分割和二值变化检测任务，因此模型的性能与这两个分支任务的性能紧密相关，下面对这两个子任务的表现进行评估。网络对变化细节的刻画能力通过二值变化图来展现，可视化结果如图8和图9所示，对SECOND数据集和Landsat-SCD数据集的二值变化检测可视化结果进行了着色渲染，白色代表真阳性、红色为假阳性、黑色为真阴性、蓝色为假阴性。从图8红色矩形和图9可以看到，本文的方法能更好地辨别伪变化，对变化区域地形状刻画更

为精细，对变化区域的检测最接近实际情况。二值变化的检测性能定量评价如表4所示可以看出在SECOND数据集上ReTNet网络在F1、IoU、Kappa指标上达到最优，分别提高1.8%、7.1%、1.6%，如表5所示可以看出在Landsat-SCD数据集上ReTNet网络在F1、IoU、Kappa指标上达到最优，分别提高0.6%、0.9%、0.2%，归因于差异增强模块的嵌入使得网络更加敏锐地关注于变化信息，极大地降低了无关变化的影响。

表2 SECOND数据集语义变化检测评价指标

Table 2 Evaluation metrics for semantic change detection on SECOND dataset

	/%				
	OA	mIoU	Sek	F_{scd}	非一致率
HRSCD-str3	82.97	63.91	8.52	49.11	4.81
HRSCD-str4	85.10	67.69	13.09	54.12	3.30
Siam-UNet	85.63	68.54	14.40	55.42	2.80
Siam-Deeplabv3+	85.22	68.50	14.43	54.83	3.00
SSCD1	86.46	69.57	16.40	57.46	1.61
Bi-SRNet	<u>86.56</u>	<u>69.87</u>	<u>16.62</u>	<u>57.53</u>	<u>1.25</u>
ReTNet(Ours)	86.79	70.24	17.84	58.00	1.00

注：加粗数据为最优值，加下划线数据为次优值。

同时通过可视化语义分割分支的双时相语义分割图直观的评价网络的地物分类水平，可视化结果如图10和图11所示。从图10、图11中可以看出本文设计的模型对地面物体的分割边界较为清晰，完整度连续性保持较好。

表3 Landsat-SCD数据集语义变化检测评价指标
Table 3 Evaluation metrics for semantic change detection on Landsat-SCD dataset

	OA	mIoU	Sek	F_{scd}	非一致率
HRSCD-str3	82.64	76.47	11.96	50.36	10.4
HRSCD-str4	83.35	<u>78.28</u>	12.55	50.11	9.9
Siam-UNet	80.52	72.07	4.98	46.17	17.1
Siam-Deeplabv3+	84.80	76.54	17.23	55.69	6.46
SSCD1	87.09	75.93	28.19	68.66	<u>0.901</u>
Bi-SRNet	<u>87.74</u>	76.32	<u>30.01</u>	<u>70.48</u>	0.924
ReTNet(Ours)	91.40	83.52	44.44	78.81	0.809

注：加粗数据为最优值，加下划线数据为次优值。

3.5 消融实验

为了评估差异增强模块和Bitemporal Transformer对语义变化检测性能的影响，本文按照网络结构和模块创新的思路在SECOND和Landsat-SCD两个数据集上设计了消融实验，每个数据集上分别包含4个实验：（1）Base表示使用仅使用孪生ResNet34构造的语义变化检测网络，未加入差异增强模块和Bitemporal Transformer语义一致性机制。（2）DEBase

表示在Base网络基础上，将差异增强模块（DE）嵌入到ResNet34相邻层之间构造出的网络，不包含Bitemporal Transformer语义一致性机制。（3）BTBase表示在Base网络基础上，加入Bitemporal Transformer语义一致性机制，不包含Bitemporal Transformer语义一致性机制。（4）ReTNet表示在Base网络基础上，加入差异增强模块和Bitemporal Transformer语义一致性机制。在SECOND和Landsat-SCD的消融实验结果分别如表6和表7所示。从表6和表7中可以看到DEBase的各项指标较Base均有所提升，验证了差异增强模块确实能够帮助网络更好的关注变化信息，减少复杂背景下伪变化的影响。BTBase模型各项指标也都比Base模型的对应指标要高，表明基于双时Transformer的“语义—变化”一致性机制能够提高网络对双时相图像时间相关性和语义一致性的建模能力，从而提高模型的检测精度。从ReTNet模型的结果可以看出差异增强模块和Bitemporal Transformer的使用相对于Base有了明显的提升，说明了差异增强模块和基于Bitemporal Transformer的“语义—变化”一致性机制在遥感影像语义变化检测上的有效性。

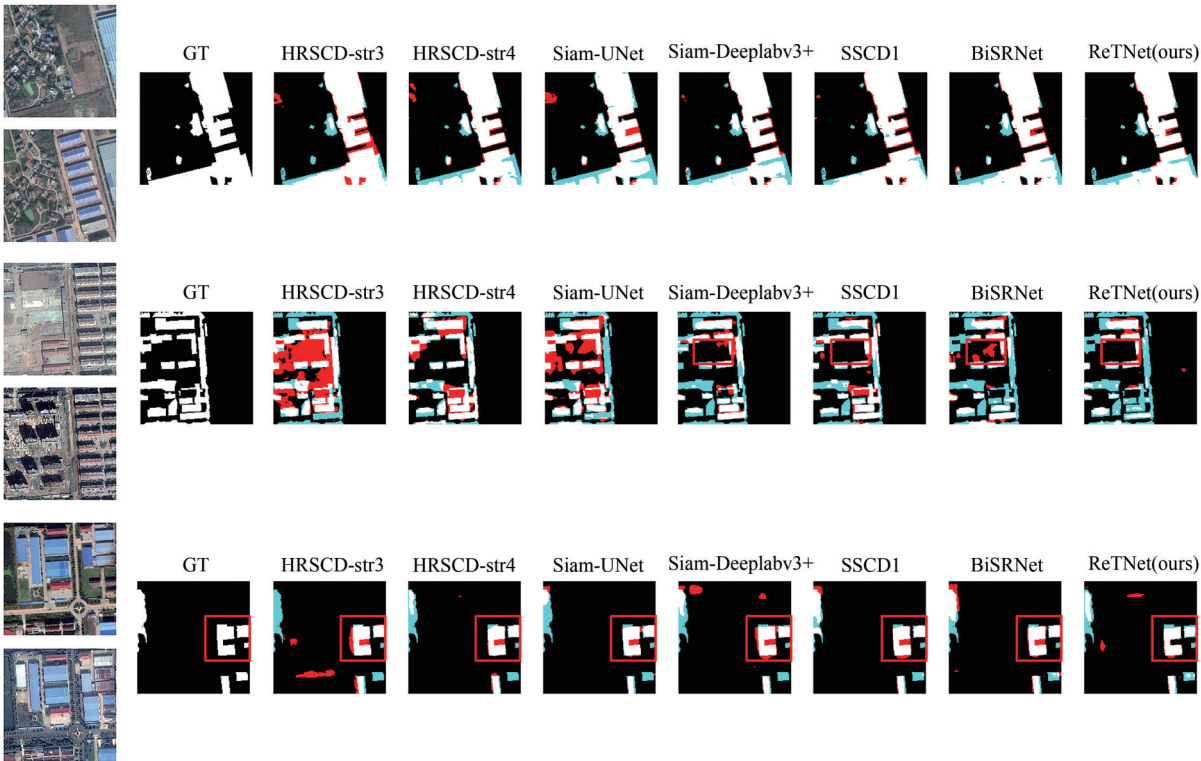


图8 SECOND数据集二值变化检测结果
Fig. 8 Binary change detection results on the SECOND dataset

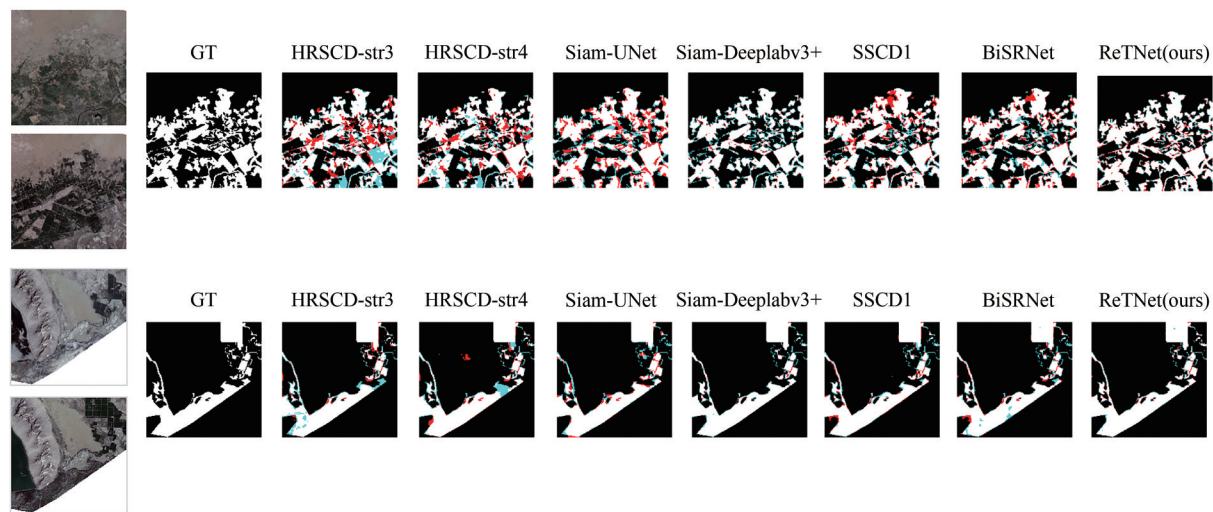


图9 Landsat-SCD数据集二值变化检测结果

Fig. 9 Binary change detection results on the Landsat-SCD dataset

表4 SECOND数据集二值变化检测性能评价指标
Table 4 Evaluation index of binary change detection performance on the SECOND dataset

	OA	F1	P	R	IoU	Kappa
HRSCD-str3	0.853	0.615	0.617	0.613	0.444	0.525
HRSCD-str4	0.874	0.665	0.657	0.673	0.498	0.587
Siam-UNet	0.879	0.675	0.662	0.688	0.509	0.601
Siam-Deeplabv3+	0.877	0.677	0.679	0.674	0.512	0.601
SSCD1	0.887	<u>0.684</u>	0.640	0.731	<u>0.520</u>	0.616
Bi-SRNet	<u>0.887</u>	0.695	<u>0.676</u>	0.715	0.532	<u>0.626</u>
ReTNet(Ours)	0.888	0.695	0.669	<u>0.723</u>	0.532	0.627

注：加粗数据为最优值，加下划线数据为次优值。

表5 Landsat-SCD数据集二值变化检测性能评价指标
Table 5 Evaluation index of binary change detection performance on the Landsat-SCD dataset

	OA	F1	P	R	IoU	Kappa
HRSCD-str3	0.898	0.793	0.796	0.789	0.657	0.725
HRSCD-str4	0.908	0.810	<u>0.796</u>	0.824	0.680	0.749
Siam-UNet	0.875	0.746	0.746	0.745	0.594	0.663
Siam-Deeplabv3+	0.935	<u>0.856</u>	0.784	0.942	<u>0.748</u>	<u>0.814</u>
SSCD1	0.895	0.787	0.785	0.788	0.648	0.717
Bi-SRNet	0.898	0.790	0.780	0.800	0.653	0.723
ReTNet(Ours)	<u>0.914</u>	0.862	0.861	<u>0.862</u>	0.757	0.816

注：加粗数据为最优值，加下划线数据为次优值。

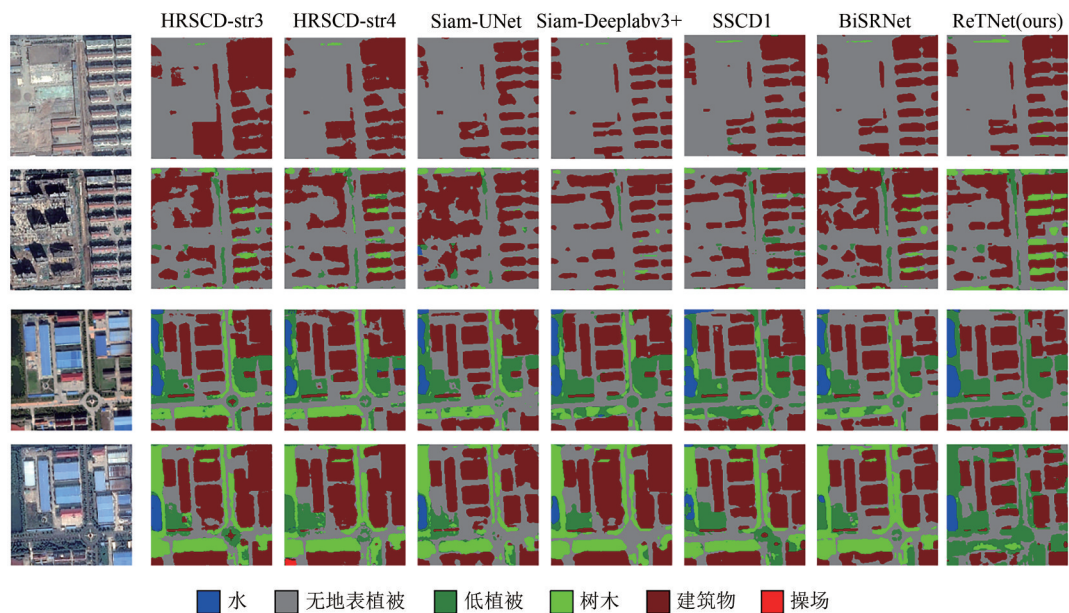


图10 SECOND数据集样例双时相语义分割结果

Fig. 10 Bi-Temporal semantic segmentation results of SECOND dataset

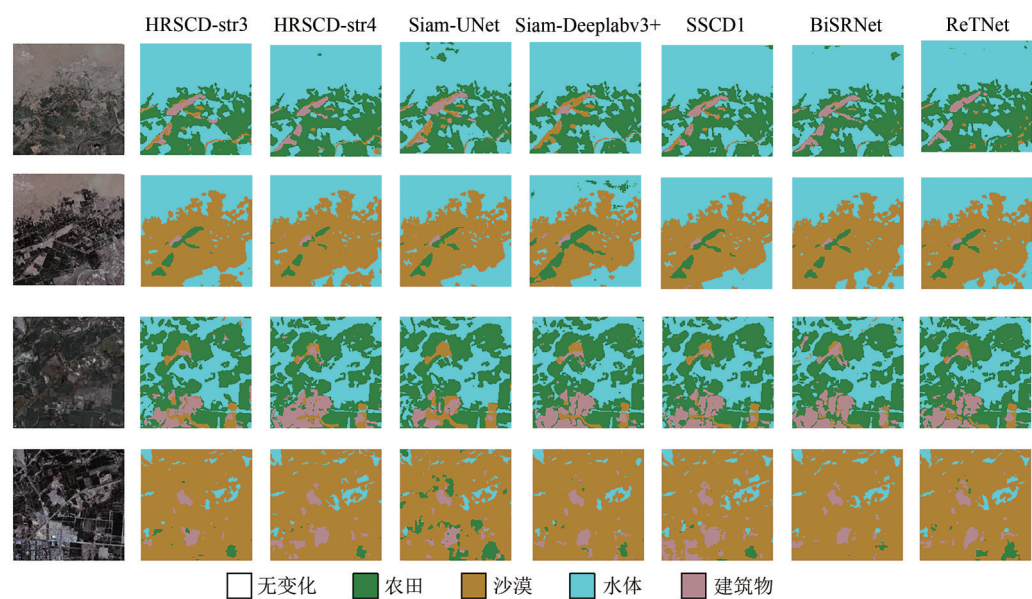


图 11 Landsat-SCD 数据集样例双时相语义分割结果

Fig. 11 Bi-Temporal semantic segmentation results of Landsat-SCD dataset

表 6 SECOND 数据集上的消融实验

Table 6 Ablation experiments on the SECOND dataset

	/%			
	OA	mIoU	Sek	F_{scd}
Baseline	85.50	68.56	15.28	54.98
Baseline+DE	85.97	69.21	16.02	56.09
ReTNet(withoutDE)	86.34	69.76	17.56	58.04
ReTNet	86.79	70.24	17.84	58.00

表 7 Landsat-SCD 数据集上的消融实验

Table 7 Ablation experiments on the Landsat-SCD dataset

	/%			
	OA	mIoU	Sek	F_{scd}
Baseline	90.23	82.61	40.77	75.76
Baseline+DE	90.69	83.65	42.38	76.28
ReTNet(withoutDE)	91.07	82.45	41.73	77.31
ReTNet	91.40	83.52	44.44	78.81

为了探索合适的 CNN Backbone 和 Transformer Block 层的最佳的超参数, 本文进一步尝试了 CNN 和 Transformer 的不同组合方式, 如表 8 所示, 分别孪生 ResNet18 特征提取模块、3 层 MSA 堆叠的 Transformer 编码器和单层 MA 堆叠的 Transformer 解码器构成的网络, 孪生 ResNet18 特征提取模块、3 层 MSA 堆叠的 Transformer 编码器和 3 层 MA 堆叠的 Transformer 解码器构成的网络、孪生 ResNet34 特征提取模块 3 层 MSA 单层 MA 构成的网络、孪生 ResNet34 特征提取模块、3 层 MSA 堆叠的 Transformer 编码器和 3 层 MA 堆叠的 Transformer 解码器构成的

网络和孪生 ResNet34 特征提取模块、6 层 MSA 堆叠的 Transformer 编码器和 6 层 MA 堆叠的 Transformer 解码器构成的网络。经实验验证: SECOND 数据集上使用孪生 ResNet34 设计的特征提取模块和由 6 层多头自注意力 (MSA)、多层感知器 (MLP) 堆叠的 Transformer 编码器, 6 层 (多头交叉注意力 (MA) 与 MLP 块堆叠构成 Transformer 解码器构成的网络模型性能达到最优。在 Landsat-SCD 数据集上使用孪生 ResNet34 设计的特征提取模块和由 6 层多头自注意力 (MSA)、多层感知器 (MLP) 堆叠的 Transformer 编码器, 6 层 (多头交叉注意力 (MA) 与 MLP 块堆叠构成 Transformer 解码器构成的网络模型性能达到最优。SECOND 数据集的可视化结果、评价指标分别如图 12、表 8 所示, Landsat-SCD 数据集上的可视化结果、评价指标分别如图 13、表 9 所示。

表 8 SECOND 数据集上 5 种孪生 CNN 与 Transformer SCD 性能评价指标

Table 8 Performance evaluation metrics of five siamese CNN and Transformer combination approaches on the SECOND dataset

	/%			
	OA	mIoU	Sek	F_{scd}
Res18_Trans31	86.19	68.42	15.11	56.51
Res18_Trans33	86.18	67.76	14.10	55.72
Res34_Trans33	86.79	70.24	17.84	58.00
Res34_Trans31	86.43	69.31	16.39	57.68
Res34_Trans66	86.60	70.01	16.91	57.62

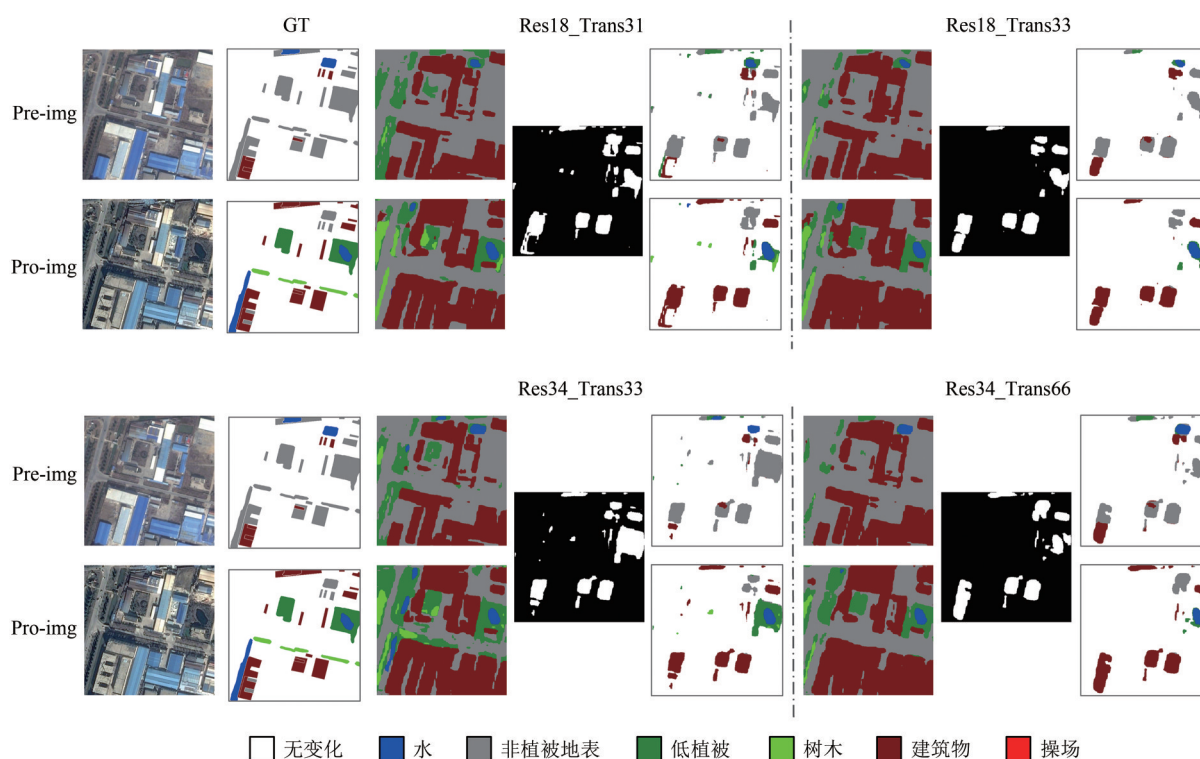


图 12 SECOND 数据集上不同孪生 CNN 和 Transformer 组合方式的探索

Fig.12 Exploration of different siamese CNN and Transformer combination approaches on SECOND dataset

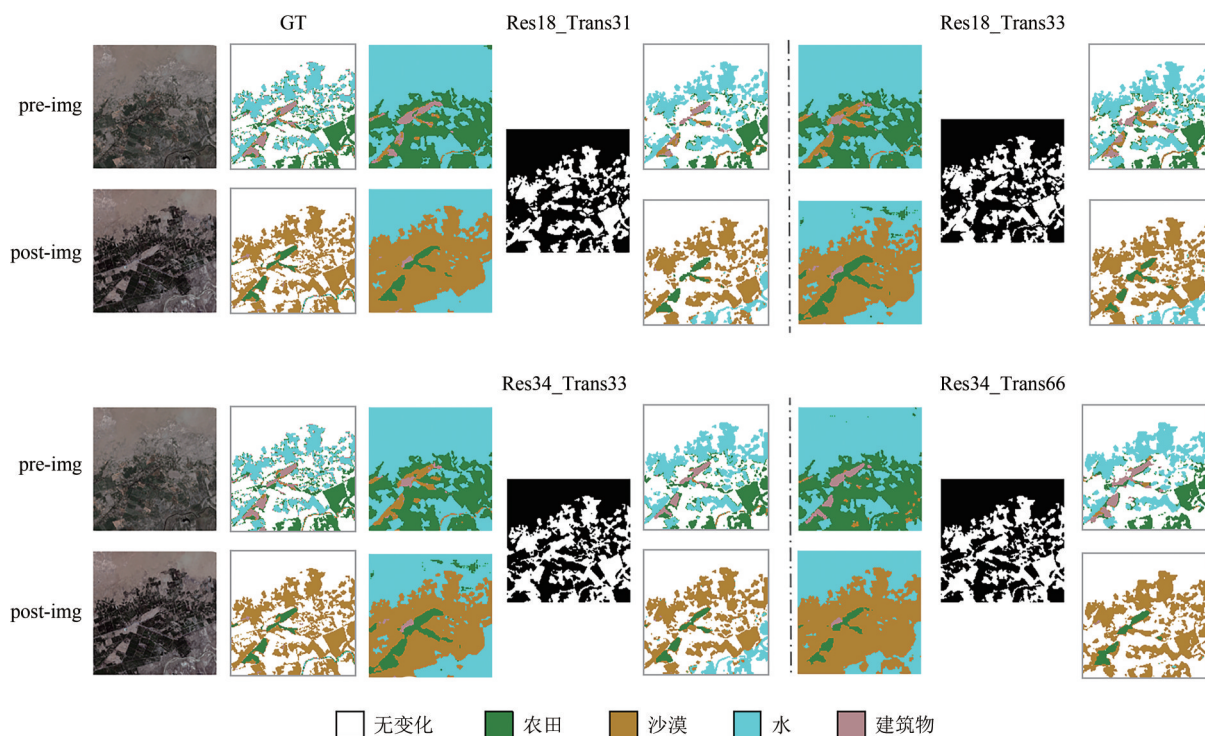


图 13 Landsat-SCD 数据集上不同孪生 CNN 和 Transformer 组合方式的探索

Fig. 13 Exploration of different siamese CNN and Transformer combination approaches on Landsat-SCD dataset

表9 Landsat-SCD数据集上5种孪生CNN与Transformer SCD性能评价指标

Table 9 Performance evaluation metrics of four siamese CNN and Transformer combination approaches on the Landsat-SCD dataset

	OA	mIoU	Sek	F_{scd} /%
Res18_Trans31	90.22	80.90	38.65	75.93
Res18_Trans33	92.12	81.55	39.87	76.07
Res34_Trans33	90.25	82.04	40.51	76.00
Res34_Trans31	92.74	82.75	43.11	78.19
Res34_Trans66	91.40	83.52	44.44	78.81

4 结 论

本文设计了一种CNN-Transformer联合的神经网络模型,该模型通过差异增强模块,增强双时图像的差异特征,提高网络对变化信息的关注度;设计双时Transformer模块将差异信息和双时语义信息映射成语义Token并融合,在Token空间上对整个时空域的“语义—变化”信息进行联合建模,有效地建模了图像之中的长程依赖,建模了双时间上下文相关性。据此设计的ReTNet网络更加关注变化区域,可准确地检测双时相遥感影像的变化位置及识别变化的要素类型。

参考文献(References)

- Bruzzone L, Prieto D F and Serpico S B. 1999. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3): 1350-1359 [DOI: 10.1109/36.763299]
- Bruzzone L and Serpico S B. 1997. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4): 858-867 [DOI: 10.1109/36.602528]
- Daudt R C, Le Saux B and Boulch A. 2018. Fully convolutional siamese networks for change detection//2018 25th IEEE International Conference on Image Processing (ICIP). Athens: IEEE: 4063-4067 [DOI: 10.1109/ICIP.2018.8451652]
- Demir B, Bovolo F and Bruzzone L. 2012. Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5): 1930-1941 [DOI: 10.1109/TGRS.2011.2168534]
- Deng Y P, Chen J S, Yi S M, Yue A Z, Meng Y, Chen J B and Zhang Y. 2022. Feature-guided multitask change detection network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 9667-9679 [DOI: 10.1109/JSTARS.2022.3215773]
- Ding L, Guo H T, Liu S C, Mou L C, Zhang J and Bruzzone L. 2022. Bi-temporal semantic reasoning for the semantic change detection in HR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5620014 [DOI: 10.1109/TGRS.2022.3154390]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houshy N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[A]. arXiv.
- El-Hattab M M. 2016. Applying post classification change detection technique to monitor an Egyptian coastal zone (Abu Qir Bay). *The Egyptian Journal of Remote Sensing and Space Science*, 19(1): 23-36 [DOI: 10.1016/j.ejrs.2016.02.002]
- Li T, Cui L Z, Xu Z H, Hu R H, Joshi P K, Song X F, Tang L, Xia A Q, Wang Y F, Guo D, Zhu J P, Hao Y B, Song L and Cui X Y. 2021. Quantitative analysis of the research trends and areas in grassland remote sensing: a scientometrics analysis of web of science from 1980 to 2020. *Remote Sensing*, 13(7): 1279 [DOI: 10.3390/rs13071279]
- Liu S C, Du K C, Zheng Y J, Chen J, Du P J and Tong X H. 2023. Remote sensing change detection technology in the Era of artificial intelligence: inheritance, development and challenges. *National Remote Sensing Bulletin*, 27(9): 1975-1987 (柳思聪, 都科丞, 郑永杰, 陈晋, 杜培军, 童小华. 2023. 人工智能时代的遥感变化检测技术: 继承、发展与挑战. *遥感学报*, 27(9): 1975-1987) [DOI: 10.11834/jrs.20222199]
- Liu Y, He X, Li D Y, Yue H and Wei J L. 2024. CBAM UNet++: attention mechanism to guide change detection studies of full-scale connected networks. *National Remote Sensing Bulletin*, 28(4): 1052-1065 (刘英, 何雪, 李单阳, 岳辉, 魏嘉莉. 2024. 注意力引导全尺度连接网络的高分辨率影像变化检测. *遥感学报*, 28(4): 1052-1065) [DOI: 10.11834/jrs.20221548]
- Luo H, Liu C, Wu C and Guo X. 2018. Urban change detection based on dempster-Shafer theory for multitemporal very high-resolution imagery. *Remote Sensing*, 10(7): 980 [DOI: 10.3390/rs10070980]
- Lv Z Y, Wang F J, Cui G Q, Benediktsson J A, Lei T and Sun W W. 2022. Spatial-spectral attention network guided with change magnitude image for land cover change detection using remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 4412712 [DOI: 10.1109/TGRS.2022.3197901]
- Moustafa M S, Mohamed S A, Ahmed S and Nasr A H. 2021. Hyperspectral change detection based on modification of UNet neural networks. *Journal of Applied Remote Sensing*, 15(2): 028505 [DOI: 10.1117/1.JRS.15.028505]

- Shi W Z, Zhang M, Zhang R, Chen S X and Zhan Z. 2020. Change detection based on artificial intelligence: state-of-the-art and challenges. *Remote Sensing*, 12(10): 1688 [DOI: 10.3390/rs12101688]
- Singh A. 1989. Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10(6): 989-1003 [DOI: 10.1080/01431168908903939]
- Wang Y, Shao Z F, Lu T, Liu L F, Huang X, Wang J M, Jiang K and Zeng K L. 2023. A lightweight distillation CNN-transformer architecture for remote sensing image super-resolution. *International Journal of Digital Earth*, 16(1): 3560-3579 [DOI: 10.1080/17538947.2023.2252393]
- Wu C, Du B, Cui X H and Zhang L P. 2017. A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion. *Remote Sensing of Environment*, 199: 241-255 [DOI: 10.1016/j.rse.2017.07.009]
- Xiang S, Wang M, Jiang X F, Xie G Q, Zhang Z Q and Tang P. 2021. Dual-task semantic change detection for remote sensing images using the generative change field module. *Remote Sensing*, 13(16): 3336 [DOI: 10.3390/rs13163336]
- Yang B, Mao Y, Chen J, Liu J Q, Chen J and Yan K. 2023. Review of remote sensing change detection in deep learning: bibliometric and analysis. *National Remote Sensing Bulletin*, 27(9): 1988-2005 (杨彬, 毛银, 陈晋, 刘建强, 陈杰, 闫凯. 2023. 深度学习的遥感变化检测综述: 文献计量与分析. *遥感学报*, 27(9): 1988-2005) [DOI: 10.11834/jrs.20222156]
- Yang K P, Xia G S, Liu Z C, Du B, Yang W, Pelillo M, Zhang L P. 2022. Asymmetric Siamese networks for semantic change detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5609818 [DOI: 10.1109/TGRS.2021.3113912]
- Yuan P L, Zhao Q Z, Zhao X B, Wang X W, Long X F and Zheng Y C. 2022. A transformer-based Siamese network and an open optical dataset for semantic change detection of remote sensing images. *International Journal of Digital Earth*, 15(1): 1506-1525 [DOI: 10.1080/17538947.2022.2111470]
- Zhao M Q, Zhao Z F, Gong S, Liu Y F, Yang J, Xiong X and Li S Y. 2022. Spatially and semantically enhanced siamese network for semantic change detection in high-resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 2563-2573 [DOI: 10.1109/JSTARS.2022.3159528]
- Zheng Z, Zhong Y F, Tian S Q, Ma A L and Zhang L P. 2022. ChangeMask: Deep multi-task encoder-transformer-decoder architecture for semantic change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183: 228-239 [DOI: 10.1016/j.isprsjprs.2021.10.015]
- Zhou C L, Shi Q, Li J and Zhang X C. 2024. Spectral-frequency domain attribute pattern fusion for hyperspectral image change detection. *National Remote Sensing Bulletin*, 28(1): 105-120 (周承乐, 石茜, 李军, 张新长. 2024. 光谱-频域属性模式融合的高光谱遥感图像变化检测. *遥感学报*, 28(1): 105-120) [DOI: 10.11834/jrs.20232600]
- Zhu C H, Chen X H, Chen J, Yuan Y H and Tang K. 2023. A siamese Nested-UNet for change detection in posterior probability space (SNU-PS). *National Remote Sensing Bulletin*, 27(9): 2006-2023 (朱传海, 陈学泓, 陈晋, 袁宇恒, 唐凯. 2023. 基于分类后验概率空间的孪生Nested-UNet(SNU-PS)变化检测网络. *遥感学报*, 27(9): 2006-2023) [DOI: 10.11834/jrs.20233070]
- Zhu Q Q, Guo X, Deng W H, Shi S N, Guan Q F, Zhong Y F, Zhang L P and Li D R. 2022. Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184: 63-78 [DOI: 10.1016/j.isprsjprs.2021.12.005]

Semantic consistency assisted change detection in remote sensing images

WU Hucheng¹, WANG Renfang^{1,2}, QIU Hong², WANG Feng², GAO Guang^{2,3}, WU Dun^{2,3}

1. College of Computer Science and Technology, Ocean University of China, Qingdao 266100, China;

2. College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315100, China;

3. Popsmart Technology(Zhejiang) Co., Ltd., Ningbo 315042, China

Abstract: Semantic change detection of remote sensing images plays an important role in ecological environment, land use, and land cover monitoring. In recent years, deep learning-based change detection methods are the hotspot of remote sensing intelligent interpretation concern; however, the existing three-branch semantic change detection methods lack consistency of change branch and semantic branches, which leads to the self-contradiction of the bitemporal semantic change detection.

To address this problem, this study proposes a remote sensing image semantic change detection algorithm based on Siamese CNN and Transformer. In the encoding stage, the Siamese ResNet34 network is first designed to extract the multiscale features of the image, and the difference enhancement module is embedded to increase the attention of the change information; then, the semantic tokenizer is used to map the feature map into compact semantic tokens, and the Transformer encoder is used to combine the bitemporal semantics and the change information to model the "semantic change" consistency. In the decoding stage, different fine-grained semantic information is fused by the

Transformer decoder using a hopping connection to generate a refined semantic feature map. Finally, the result of bitemporal semantic change is obtained after upsampling recovery and mask multiplication.

Experimental results on the remote sensing semantic change detection public dataset SECOND and Landsat-SCD show that the proposed algorithm can effectively focus on the change region, maintain the consistency between the change results and the semantic results, and achieve excellent evaluation indexes and visual effects.

We can draw the following conclusions: (1) The proposed difference enhancement module can enhance the difference characteristics of bitemporal remote sensing images and improve the network's focus on change information. (2) The proposed bitemporal Transformer module maps the difference information and bitemporal semantic information into semantic tokens and fuses them to model jointly the "semantic change" information of the whole spatiotemporal domain in the token space, effectively modeling the long-range dependencies in the images and the bitemporal contextual correlations. The long-range dependency in the image is effectively modeled, and the bitemporal contextual relevance is modeled. The ReTNet network designed accordingly pays more attention to the change area and can accurately detect the change location and recognize the change element type of the bitemporal remote sensing image.

Key words: remote sensing image, change detection, semantic consistency, difference enhancement, multi-scale features, siamese network, ResNet34, Transformer

Supported by National Natural Science Foundation of China (No.61906170); Science and Technology Plan for Zhejiang Province (No. ZCLY24F0301); Plan Project of Ningbo Municipal Science and Technology (No. 2022Z233, 2021Z050, 2022S002, 2023J403)