

面向遥感语义变化检测的时空变化协同感知方法

荆纬^{1,2}, 燕庭轩¹, 靳翔沆¹, 宋彬彬¹, 王琦¹

1. 西北工业大学 光电与智能研究院, 西安 710072;

2. 西北工业大学 国家卓越工程师学院, 西安 710072

摘要: 语义变化检测的核心目标是精准识别地表覆盖的复杂变化过程, 并同时确定其对应的语义类别。现有研究表明, 采用分离的变化定位分支和语义识别分支来协同完成语义变化检测任务是解决该问题的可靠方案。然而, 已有框架往往未能充分整合和利用地物语义信息来指导变化识别过程, 这在一定程度上限制了模型对复杂场景变化的精准识别能力。针对该问题, 本文提出了语义引导的时空协同感知网络 SemSTNet (Semantic-guided Spatio-Temporal Collaborative Perception Network), 在考虑双时相特征时空交互的同时以语义指导变化区域的精确识别。具体来说, 构建了对称时空解调差分模块利用双向特征调制实现跨时相特征互校准, 通过三维差分卷积解耦时空维度变化响应; 其次, 设计了语义门控融合解码器引入语义相似性动态加权机制, 以余弦相似度图谱生成语义先验, 通过自适应门控融合增强真实变化区域的显著性表达; 最后, 设计深度语义对比损失构建双监督信号, 在特征空间约束未变化区域语义一致性, 扩大变化/未变化区域的类间距离。在2个语义变化检测数据集上的广泛实验表明, 提出的模型在定性和定量评估方面都显著优于其他最先进的语义变化检测方法。

关键词: 遥感图像, 地表覆盖变迁, 语义变化检测, 时空协同, 语义引导

中图分类号: TP751/P2

引用格式: 荆纬, 燕庭轩, 靳翔沆, 宋彬彬, 王琦. 2026. 面向遥感语义变化检测的时空变化协同感知方法. 遥感学报, 30(6): 1792-1806

Jing W, Yan T X, Jin X L, Song B B and Wang Q. 2026. Spatiotemporal change collaborative perception method for remote sensing semantic change detection. National Remote Sensing Bulletin, 30(6): 1792-1806 [DOI: 10.11834/jrs.20265265]

1 引言

近年来, 随着遥感大数据技术的迅猛发展, 遥感图像变化检测领域取得了显著进展 (Lei等, 2022a; Radke等, 2005; Ru等, 2021)。该技术在灾害评估、农业资源调查、城市规划等应用场景中发挥着日益重要的作用, 已成为地表环境监测和人类活动分析的关键技术手段。然而, 现有研究主要集中于发生变化的位置检测, 而在实际应用中, 决策者不仅需要识别变化发生的地理位置, 更需准确获取土地覆盖变化的类型信息。尤其是在地物类别分布不均衡、季节与光照条件变化显著且多尺度小目标变化并存的复杂场景中, 真实语义变化与由环境因素引起的伪变化高度交

织, 对变化检测与语义识别提出了更高要求。基于这一实际需求, 语义变化检测 SCD (Semantic Change Detection) 逐渐成为遥感领域的研究热点, 吸引了众多学者开展系统性研究 (Caye Daudt等, 2019; Ding等, 2022; Mou等, 2019; Yang等, 2022a; Zhao等, 2022)。

与传统的二元变化检测任务相比 (Du等, 2020; Wu等, 2014; Zhang等, 2020), 语义变化检测不仅需要检测变化发生的地理位置, 更重要的是要准确识别变化区域内的地物变迁的演化方向, 即识别变化前后地物的语义类别。目前, 语义变化检测任务的一个直观解决方案是将其视为基于二值变化检测模型的多类别分割任务, 如Caye等 (2019) 提出的 HRSCD.STR 1 模型。然而,

收稿日期: 2025-07-21; 预印本: 2026-03-30

基金项目: 国家自然科学基金(编号:62471394, 62301385, U21B2041); 西北工业大学博士论文创新基金项目(编号:CX2024108)

第一作者简介: 荆纬, 研究方向为遥感图像处理、变化检测。E-mail: wei_adam@mail.nwpu.edu.cn

通信作者简介: 王琦, 研究方向为遥感图像处理、计算机视觉和模式识别。E-mail: crabwq@gmail.com

这种方法存在明显的理论局限和实践挑战: 首先, 随着土地覆盖类型数量的增加, “从一到”映射的排列组合数量呈指数级增长, 显著增加了构建多功能多类别分割网络所需的训练数据规模, 给模型训练和优化带来了巨大挑战; 其次, 从模型优化机制来看, 在缺乏有效先验知识引导的情况下, 深度模型往往倾向于从训练数据中学习固定模式, 而非深入挖掘双时相图像间的深层语义关联。这种学习机制容易导致模型陷入局部最优解, 并引发严重的过拟合问题, 从而影响模型的泛化能力和实际应用效果。

鉴于上述原因, 研究者们将SCD任务视为多任务学习过程 (Ding 等, 2022; Zhao 等, 2022; Zheng 等, 2022)。该方案首先从双时相图像中采用深度特征提取来捕获差异信息并生成二值变化图; 随后, 通过掩膜操作, 迫使模型聚焦于变化区域的语义识别。在此过程中, 提取高判别力语义表征并进而从中建模差异特征是实现高精度语义变化检测的基础 (Lv 等, 2022; Yang 等, 2022b)。Yang 等 (2022b) 提出非对称孪生网络, 通过结构差异化模块提取特征实现语义变化的定位与识别。Zheng 等 (2022) 在设计中引入语义感知编码器建模语义变化的因果关系, 该编码器专门学习语义表征, 随后通过 Transformer 模块从语义特征中学习变化表征, 并辅以正则化防止过拟合。Ding 等 (2022) 将语义特征与时序特征整合到深度变化检测单元, 强化时相内与时相间的语义一致性, 同时设计新损失函数提升检测结果的语义连贯性。尽管研究人员从各种角度提取和增强差异特征, 然而, 许多现有变化检测方法仍主要依赖像素级或低层外观差异 (如光谱差异、纹理差异) 来判断变化, 常见的孪生网络结构仅在编码阶段共享参数, 在解码阶段主要对低层差分特征进行卷积处理。语义分支往往只用于输出语义分割结果, 而缺乏通过显式的语义相似性建模或门控机制, 将高层语义先验反馈到变化分支中, 导致在地物类别转换、伪变化抑制和小目标变化识别等场景下, 高层语义特征的指导作用没有被充分发挥。

针对上述问题, 本文提出了语义引导的时空协同感知网络。首先, 为了从双时相特征中提取有鉴别力的差异信息, 构建了基于双向动态调制与时空差分卷积的差异特征提取方法——对称时空解调差分模块, 通过双向特征调制实现跨时相

特征互校准, 然后利用 3D 差分卷积捕捉时空维度变化响应; 其次, 通过语义感知的动态加权机制, 将语义一致性先验与像素级差异特征深度融合, 有效解决复杂场景下的伪变化抑制难题; 最后, 设计了深层语义对比损失, 通过语义引导的对比学习策略, 增强特征空间的变化可辨识性, 与语义感知变化解码器形成端到端的协同优化, 显著提升复杂场景下的变化查全率与查准率, 以期为高分辨率遥感影像语义变化检测提供高效可靠的技术方案。

2 国内外研究现状

2.1 二值变化检测

早期的变化检测研究主要依赖于传统方法, 如阈值分割 (Zhu 等, 2022a)、形态学分析 (Javed 等, 2020) 和图像代数 (Yang 等, 2022a), 后分类变化检测等。这些方法通过手工设计的特征来识别时相遥感图像中的变化区域。然而, 这类方法在特征表达能力上存在明显局限, 难以有效应对遥感图像中的视角变化、光照变化等干扰因素。特别是在建筑物变化检测任务中, 传统方法在适应复杂场景和表达变化语义方面表现出明显不足。

崔斌等 (2020) 针对高分三号 SAR 影像提出的双阈值变化检测方法 D-GKIT, 有效解决了广义高斯分布下阈值重叠带来的分割误差问题, 并首次在 SAR 数据上实现了变化类型的区分。随着遥感大数据的积累与深度学习技术的发展, 基于深度神经网络的变化检测方法逐渐成为主流。这类方法通过端到端的特征学习, 能够同时提取低层次空间细节与高层次语义表示 (柳思聪 等, 2023; Chi 等, 2023; He 等, 2016; Jing 等, 2023; Lin 等, 2024; Liu 等, 2025b; Luo 等, 2023)。从网络架构角度来看, 主要可分为单分支和双分支两类。

传统的后分类变化检测方法通常先对不同时相遥感影像分别进行土地覆盖分类或语义分割, 再通过叠加分类结果构建“从一到”变化矩阵, 从而获得具有语义含义的变化信息。Serra 等 (2003) 利用不同传感器 (MSS 与 TM) 数据开展后分类变化检测, 并从分类精度、地物碎片化程度、平面定位精度和像元尺度等方面系统分析了后分类比较的误差来源, 提出了一套提高多源数据后分类变化检测可靠性的操作流程。Garcia-Aguirre 等

(2005) 针对墨西哥城西南部温带森林, 基于多时相分类结果进行后分类数字变化检测, 定量评估了1973年—1989年的16年间森林覆盖变化及年均砍伐率, 表明后分类方法在“从森林到灌丛或农地”等类型变化监测中的直观优势。在此基础上, Wu等(2017)提出将迭代慢特征分析与贝叶斯软融合相结合的后分类变化检测框架, 通过独立分类获得类别概率、利用ISFA提取连续变化概率图, 并在贝叶斯理论下融合类别概率与变化概率, 有效缓解了传统后分类方法中误分类误差累积和阈值分割导致变化信息损失的问题, 将后分类比较方法的精度和鲁棒性推向新水平。

单分支框架采用早期融合策略, 在图像输入网络之前, 通过图像差分或拼接方式对双时相图像进行预处理。例如, Sun等(2022)将扩展的ConvLSTM与U-Net结合, 并采用空洞卷积提取多尺度空间特征, 实现了端到端的变化检测; Lin等(2023)借鉴视频理解方法, 提出了P2V-CD模型, 将双时相图像的空间和时间维度解耦建模。

相比之下, 双分支结构(如典型的孪生网络)则分别从双时相图像中独立提取特征, 并在后续阶段进行特征融合(Liu等, 2025a; Zhang等, 2020)。Zhang等(2020)提出了一个深度监督的差异判别网络, 联合原始特征与差分特征, 有效提升了变化图的边界一致性。多数基于CNN的方法注重局部空间建模, 例如LGSAANet(Lei等, 2022b)引入了尺度自适应注意力模块, 建模变化信息与卷积核大小之间的关系; FCCDN(Chen等, 2022)采用特征金字塔结构提取多尺度特征。

而Transformer模型则通过自注意力机制, 突破了CNN感受野受限的局限。Bandara和Patel(2022)构建了分层Transformer架构, 以精细建模多尺度遥感特征。DiFormer(Lin等, 2024)则通过引入token差异评估模块, 从变化目标与背景之间生成差异表示。李健慷等(2025)提出的MFTSNet模型融合孪生结构和Transformer模块细化特征, 利用浅层CNN网络结构生成变化检测结果。

尽管当前方法已取得显著进展, 但仍面临以下若干关键技术瓶颈: 首先, 特征建模机制本身存在局限。CNN受限于固定卷积核大小, 难以捕捉全局上下文信息; 而Transformer虽擅长建模长距离依赖, 但易忽视token的静态空间感受范围, 导致变化边界表达能力不足(Lin等, 2024); 其

次, 现有特征融合策略尚缺乏统一的理论基础。例如, ACAHNet(Zhang等, 2023)采用中间融合策略, 引入语义图以保留关键特征; MDAFormer(Zhu等, 2023)则通过多层级Transformer模块对双时相特征进行融合。

2.2 语义变化检测

传统的二值变化检测方法仅能定位变化区域, 无法实现“从一到”语义映射, 无法满足语义层面的变化表达需求。为此, Caye Daudt等(2019)将变化检测中的二值分割任务重构为多类别分割问题。然而, 这一策略带来了2个主要挑战: 一是由于类别空间扩展, 模型参数量显著增加; 二是对标注数据的需求显著提升。

后续研究提出了将二值变化检测与语义分割相结合的集成框架, 进一步提升了检测性能。典型方法采用双编码器结构, 分别提取语义特征与变化特征, 配合全卷积网络FCN(Fully Convolutional Networks)与双解码器架构。其中, 语义分割图为变化检测提供先验指导, 而二值变化图用于定位变化区域, 最终通过后处理生成语义变化图(Caye Daudt等, 2019)。

在此基础上, Yang等(2022a)提出一种非对称孪生网络, 引入异构模块以提升语义变化定位能力。Zheng等(2022)构建了一个语义感知编码器, 用于建模语义变化的因果关系, 并引入Transformer模块提取变化表示, 同时采用正则化技术缓解过拟合风险。Ding等(2022)则通过融合语义与时序特征, 从而增强时间内的语义一致性, 并设计特定损失函数以优化语义协调性。

尽管取得了一定成效, 语义变化检测SCD(Semantic Change Detection)仍面临“类间相似性高、类内差异性大”的问题, 导致误检现象频发。当前, 多尺度特征融合已成为主流解决方案。HPCFNet基于孪生CNN网络, 采用分层多尺度融合策略, 实现了多类别变化检测(Lei等, 2021); KGLPD引入自适应梯度感知机制, 构建块分布先验与认知土地模式, 用于分析地物间的关系(Zhu等, 2022b)引; Y-Net利用双流网络提取双时相特征, 并通过全尺度跳跃连接重建多层次变化图(Wang等, 2022)。此外, 吴虎承等(2025)提出了一种“关注语义一致性”的遥感影像变化检测方法, 通过孪生CNN与Transformer的有机结合,

实现了变化结果和语义结果在空间与类别层面的高度对齐。汤玉奇等(2024)提出了一种无监督对称变化检测网络SCDN (Symmetric Change Detection Network), 将光学影像和SAR影像映射到近似同一特征空间以提取变化信息, 利用聚类分析提取变化区域并区分变化类别。Wang等(2024a)提出了SiamContrast, 该方法通过引入监督对比学习框架, 旨在学习具有时空不变性且判别力强的地物特征, 以解决高分辨率遥感影像中由于光谱差异导致的语义误判问题。Tang等(2024)提出了ClearSCD模型, 该模型引入了语义增强对比学习与双时相语义关联捕获机制, 实现了语义分割与变化检测任务的互惠增益, 显著提升了语义变化检测的识别精度。

尽管上述方法有效提升了语义检测精度, 在时序关系建模方面, 大部分方法仅采用特征拼接、逐像素差分或二维卷积来隐式表征时间差异, 并未显式刻画双时相特征在通道维度和空间维度上的演变轨迹; 少量引入自注意力或Transformer结构的工作, 也多侧重于单时相内部的空间依赖,

对跨时相语义演化过程的长距离时序依赖刻画不足。因此, 在季节变化、光照变化与真实地物变化交织的复杂场景下, 这类方法仍容易将背景光谱变化误判为语义变化, 难以全面捕捉双时相语义变化中的深层时序关联。此外, 尚缺乏一种协同优化机制, 能够将显著的二值变化信息与语义特征进行有效融合。因此, 构建全局时序依赖模型, 并实现二值变化与语义变化的联合优化, 是当前语义变化检测研究中的关键挑战, 也是本研究工作的核心目标。

3 研究方法

本文将系统阐述语义引导的时空协同感知网络的整体架构设计, 深入剖析其核心组件与工作机制。如图1所示, 该创新性架构主要由2个关键模块构成: 基于对称时空解调机制的差分特征提取模块用于提取双时相差分特征, 以及语义门控融合解码器用于从多维语义差异生成变化先验用于预测真实变化区域。

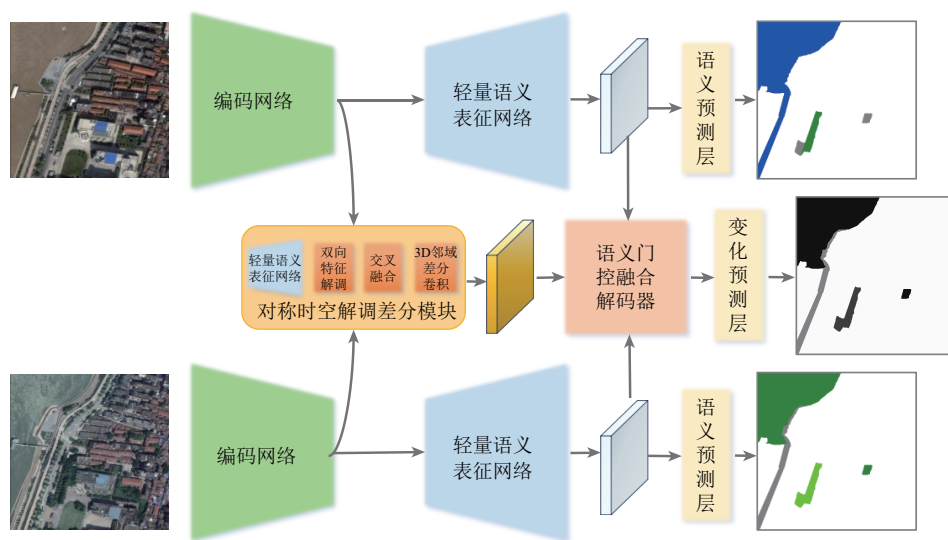


图1 SemSTNet的整体架构

Fig. 1 Overall architecture of SemSTNet

3.1 语义引导的时空协同感知网络架构

如图1所示, 本文提出的语义变化检测网络采用双时相遥感图像对作为输入, 生成像素级语义变化类型判别结果的同时输出精细化二值变化掩膜。具体来说, 采用孪生Transformer作为图像编码器, 该网络擅长从双时相图像中捕捉浅层细节和深层语义信息, 这些特征封装了图像的时空特

征。然后, 通过设计的对称时空解调差分模块从编码特征中显式提取高鉴别力差异信息。具体计算公式如下:

$$X_{\text{diff}} = X_1 \Theta X_2, \quad (1)$$

式中, X_{diff} 为双时相差分特征, X_1 和 X_2 分别为来自双时相图像的编码特征, Θ 表示对称时空解调差分模块。在提取差异特征的同时, 采用轻量语义

表征网络将来自编码网络的多尺度特征通过特征金字塔的结构融合为深层语义特征 X_{s1} 和 X_{s2} 。轻量语义表征网络的计算过程表示如下：

$$X = K_{1 \times 1}(\Pi_{i=1}^4 K_{1 \times 1}(U(X_i))) \quad (2)$$

式中, $K_{1 \times 1}$ 是线性卷积核, Π 表示拼接操作, U 表示上采样操作。

然后, 将语义特征馈送至由卷积层构成的语义预测层输出双时相图像的像素级语义类别。进一步, 为了通过语义表征增强真实变化区域的显著性表达, 将 X_{s1} , X_{s2} 输入至语义门控融合解码器中, 作为语义先验与差异特征 X_{diff} 进行深度融合:

$$X'_{diff} = G(X_{s1}, X_{diff}, X_{s2}) \quad (3)$$

式中, G 表示语义门控融合解码器。最后将增强后

的差异特征 X'_{diff} 通过变化预测层得到二值变化图, 变化预测层同样由卷积层构成。

3.2 对称时空解调差分模块

为了提升语义变化检测中差异特征的判别性与鲁棒性, 本文提出对称时空解调差分模块 SSDM (Symmetric spatiotemporal demodulation differential module)。如图 2 所示, 该模块由“时空解调域对齐单元”和“本质差分提取单元”两部分构成, 前者通过双向特征解调显式对齐双时相语义特征分布, 以消除域偏移问题; 后者在对齐后的特征上构建类时序特征流, 并利用中心差分卷积在时空维联合建模, 以捕获特征演变规律中的本质差异模式。二者协同作用, 在复杂场景下增强真实语义变化的响应强度并抑制背景伪变化。

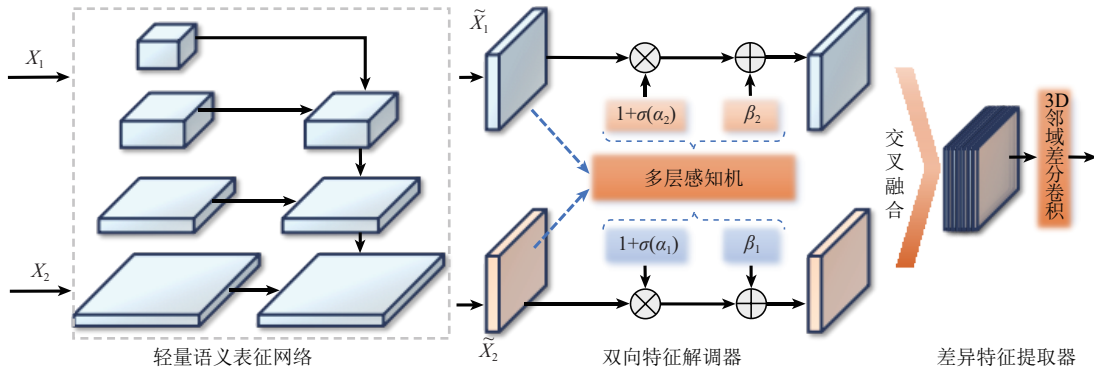


图2 对称时空解调差分模块

Fig. 2 Symmetric spatiotemporal demodulation differential module

首先, 来自编码网络的多尺度特征首先通过轻量语义表征网络进行多尺度融合获得双时相特征 X_1 和 X_2 。然后为了促进双时相特征的交互校准, 双向特征解调器基于参数动态生成的 Scale-Shift 机制对双时相特征进行对称互调制, 通过 Sigmoid 门控控制特征增强强度, 实现特征空间的对等语义校准。输入特征通过多层感知机 M 及切片操作 Split 生成解调参数:

$$\alpha, \beta = \text{Split}^*(M(X)) \quad (4)$$

式中, α 和 β 分别表示对特征图进行缩放和平移的参数。接下来, 利用特征图 X_2 的缩放和平移参数对特征图 X_1 进行调制:

$$X_1 = X_1 \cdot (1 + \sigma(\alpha_2)) + \beta_2 \quad (5)$$

为了保持对称性, 同样的调制过程也应用于特征图 X_2 , 从而在通道维度上实现双向、对称的特征互校准。上述过程等价于利用一个时相的全局语

义统计对另一时相特征进行条件重标定, 在语义特征空间内完成跨时相的条件标准化与互对齐, 有效削弱由成像条件差异引起的域偏移, 为后续差分建模提供语义上更一致的特征基础。

在完成语义域对齐后, SSDM 并不直接对 X_1 与 X_2 进行简单拼接或减法, 而是通过交叉缝合策略构造类时序特征 X_{cs} , 以显式刻画特征在“时序-通道”维度的演变轨迹。具体而言, 从两个时相特征中交替提取通道: X_1 提供奇数通道, X_2 提供偶数通道, 从而在通道维度形成 X_{cs} 的“交替时序”结构。然后, 将相邻 3 个通道的融合特征视为时间序列, 并利用中心差分卷积核来提取差异信息, 表示如下:

$$X_{diff}(p_0) = \sum_{p_n \in R} W(p_n) \cdot (X_{cs}(p_0 + p_n) - X_{cs}(p_0)) \quad (6)$$

式中, p_0 表示输入和输出特征图上的当前位置, p_n

枚举特征图中的所有位置。最后通过特征压缩与融合获得具有判别性的差异表征。该结构通过双向特征解调保持特征语义一致性,同时利用三维卷积捕获深层时空关联,有效抑制伪变化并增强真实语义变化的响应强度。

综上,SSDM通过“双向时空解调”显式缓解双时相特征间的域偏移,并通过“差异特征提取单元”在类时序特征流上利用中心差分卷积捕获深层时空关联与本质差异模式,从而有效抑制伪

变化、增强真实语义变化的响应强度,为变化解码提供更加可靠的差异先验。

3.3 语义门控融合解码器

语义门控融合解码器 SGFD (Semantic Gated Fusion Decoder) 旨在通过自适应融合差异特征与语义相似性特征,增强变化检测任务中的特征判别性。如图3所示,其以差异特征和双时相高层语义特征为输入,通过门控机制动态调节特征权重,最终输出变化概率图。

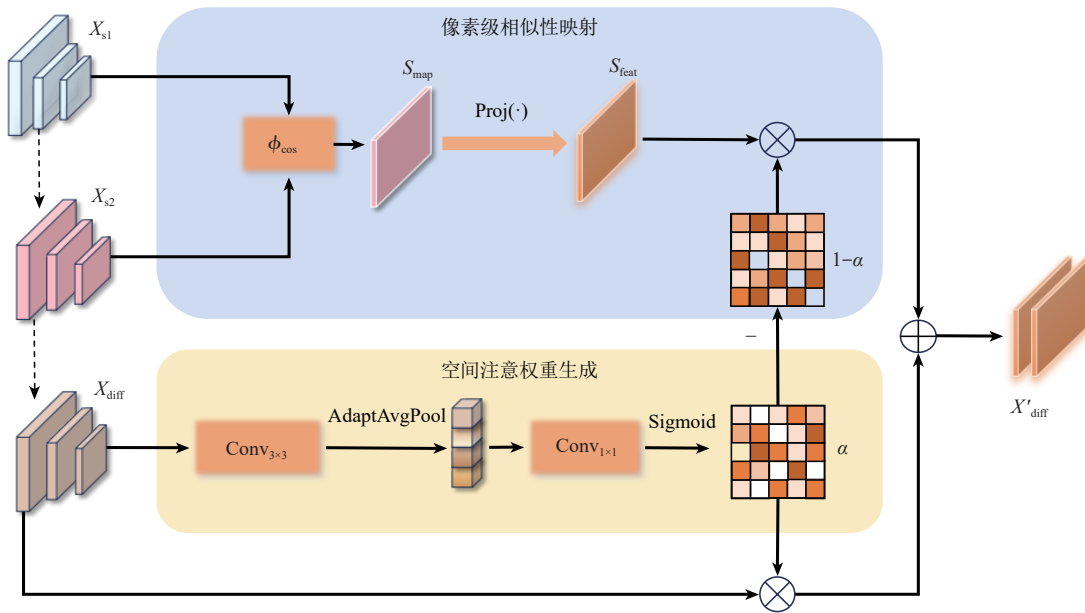


图3 语义门控融合解码器

Fig. 3 Semantic gated Fusion Decoder

首先,计算双时相特征的语义相似性矩阵以捕捉未变化区域的共性特征。基于余弦相似度准则,沿通道维度生成像素级相似性映射:

$$S_{\text{map}} = \phi_{\text{cos}}(X_{s1}, X_{s2}) \in \mathbb{R}^{1 \times H \times W} \quad (7)$$

$$\phi_{\text{cos}}(a, b) = \frac{a \cdot b}{\|a\| \|b\|} \quad (8)$$

式中, ϕ_{cos} 表示余弦相似度函数,通过投影层 $\text{Proj}(\cdot)$ 将 S_{map} 映射至与差异特征同维度的语义相似性特征:

$$S_{\text{feat}} = \text{Proj}(S_{\text{map}}) \in \mathbb{R}^{C \times H \times W} \quad (9)$$

然后,设计权重生成网络 $W(\cdot)$ 从差异特征中提取注意力权重 α , 其结构如下:

$$W(X_{\text{diff}}) = \sigma(K_{1 \times 1}(GAP(K_{3 \times 3}(X_{\text{diff}})))) \quad (10)$$

式中, GAP 为全局平均池化。通过空间注意力权重 α 根据局部上下文动态调整融合比例:

$$X'_{\text{diff}} = \alpha \odot X_{\text{diff}} + (1 - \alpha) \odot S_{\text{feat}} \quad (11)$$

式中, $\odot$ 表示通道广播的逐元素乘法。

3.4 损失函数

本文采用3个损失函数训练提出的模型,分别为语义分割损失 L_s 、变化损失 L_{BCE} 和提出的深度语义对比损失 L_{CL} 。

L_s 是双时相语义分割图和地面实况之间的交叉熵损失,表示如下:

$$L_s = - \sum_{i=1}^M y_i \log(p_i) \quad (12)$$

式中, M 表示类别的总数, i 用作每个类别的索引, y_i 表示对应于类别 i 的真实类别的概率,并且 p_i 表示预测的类别被分类为 i 的概率。

变化损失与语义分割损失类似,采用交叉熵损失函数约束预测的变化区域和真实变化区域间

的误差。在语义变化检测任务中, 现有方法往往仅关注变化区域的区分性, 而忽略了非变化区域语义特征的一致性约束。此外, 由于场景复杂性和样本不平衡问题, 模型容易对细微变化过拟合或对显著变化欠拟合。为解决上述问题, 本文提出一种基于对比学习的监督损失函数 CCL (Contrastive Change Loss), 通过联合优化双时相特征的语义差异与一致性: 对变化区域强制扩大双时相特征差异, 对非变化区域则约束其特征相似性。该损失结合基于二元交叉熵的语义相似度监督与对比学习的结构化约束, 增强模型对语义变化的敏感性和鲁棒性。具体来说, 给定双时相语义特征图 X_{s1} , X_{s2} 及二值变化标签 y , 首先对特征进行 L2 归一化, 计算像素级余弦相似度矩阵 S , 然后为了监督相似度与标签的对齐, 构建二元交叉熵损失强制非变化区域相似度趋近 1, 变化区域相似度趋近 0:

$$L_{BCE} = -\frac{1}{M} \sum_i (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (13)$$

式中, M 表示像素总数, p 表示预测的变化像素概率。此外, 在语义对比损失中引入间隔约束, 采样 K 个负样本 (变化区域), 要求非变化区域相似度至少高于变化区域相似度一个间隔 γ :

$$L_{CL} = \frac{1}{|P| \cdot K} \sum_{p \in P} \sum_{k=1}^K \max(0, \gamma - (S_p - S_{n_k})) \quad (14)$$

式中, P 为非变化区域样本集合, S_p 为非变化区域的样本相似度, S_{n_k} 为采样的负样本相似度, γ 为预设间隔。总损失为语义分割损失和变化损失的加权和:

$$L_{Total} = L_s + L_{BCE} + L_{CL} \quad (15)$$

该损失函数通过端到端训练, 使模型同时捕捉语义变化并保持非变化区域稳定性, 提升复杂场景下的检测精度。

4 实验结果与分析

4.1 数据描述

SECOND 数据集被广泛用于语义变化检测研究, 由 4662 对航空图像组成, 其中 2968 对用于训练, 1694 对用于测试。每张图像的分辨率为 512×512 像素, 由 RGB 3 个颜色通道构成, 空间分辨率在每像素 0.5—3.0 m。数据集中的标签分为 7 类: 裸土、树木、低植被、水体、建筑物、游乐场和无

变化区域。值得注意的是, 61.2% 的像素属于无变化类别, 29.9% 属于裸土, 其余类别的像素比例 < 5%。由于类别分布高度不平衡, 要求模型具有较强的泛化能力。

JL-1 数据集是农田变化检测的竞赛数据集, 图像来自吉林一号遥感卫星。它包括 6000 张图像, 其中 4000 张用于训练, 2000 张用于测试。每张图像的大小为 256×256 像素, 同时也包含 RGB 3 个通道, 空间分辨率为每像素 0.75 m。与 SECOND 数据集不同, JL1 数据集为双时相图像提供了 9 个“从一到”类型标签, 以表示两个时间点之间的变化。这些变化类型包括农田—道路、农田—森林/草原、农田—建筑、农田—其他、道路—农田、森林/草原—农田、建筑—农田、其他农田和无变化区域。在训练之前, 这些标签被转化为与 SECOND 数据集格式一致的标签。

4.2 评价指标

本文采用平均交并比 (mIoU)、Separated Kappa (SeK) 系数和 F1 分数 (F1) 3 种指标定量评估语义变化检测模型性能。

平均交并比 (mIoU) 是变化区域和未变化区域的交并比 (IoU) 的平均值。交并比 (IoU) 定义为预测结果与实际真实值 (ground truth) 之间的交集与并集的比值, 其计算公式如下:

$$mIoU = (IoU_0 + IoU_1) \times 0.5 \quad (16)$$

式中, IoU_0 和 IoU_1 则分别对应未变化区域和变化区域的交并比。

SeK 是能够通过自适应校正输出偏差来缓解标签不平衡影响的评估指标, 其计算公式如下:

$$SeK = e^{(IoU_1 - 1)} \cdot (\hat{\rho} - \hat{\eta}) / (1 - \hat{\eta}) \quad (17)$$

$$\hat{\rho} = \sum_{i=2}^C q_{ii} / \left(\sum_{i=1}^C \sum_{j=1}^C q_{ij} - q_{00} \right) \quad (18)$$

$$\hat{\eta} = \sum_{j=1}^C (\hat{q}_{j+} \cdot \hat{q}_{+j}) / \left(\sum_{i=1}^C \sum_{j=1}^C q_{ij} - q_{00} \right)^2 \quad (19)$$

式中, q_{ij} 表示被识别为第 i 种变化类型但实际属于第 j 种变化类型的像素数量, $\hat{q}_{j+}$ 和 $\hat{q}_{+j}$ 则分别表示混淆矩阵中 (不含 q_{00} 项) 的行和与列和。

4.3 本文算法与 7 种 SOTA 算法所得结果比较

为了评估 SemSTNet 的优越性, 本研究将其与 7 种 SOTA 算法进行了比较。7 种方法的简要介绍如下:

(1) HRSCD3是将语义变化检测拆分为“变化检测”和“语义预测”两个任务的早期代表方法,采用3个网络分别预测双时相影像的语义图与变化图,从而实现基于多网络的后分类变化检测范式(Caye Daudt等, 2019)。

(2) HRSCD4在HRSCD3的基础上进一步引入分支间的信息交互机制,通过跳跃连接将语义预测分支中的高层语义特征传递至变化检测分支,实现语义信息与变化特征的共享与融合(Caye Daudt等, 2019)。

(3) SSESNet是一种端到端的空间与语义增强型孪生网络,在孪生结构的基础上,设计了专门的空间特征增强模块和语义特征融合模块,能够充分聚合图像内部的细粒度空间纹理与高层语义信息(Zhao等, 2022)。

(4) Bi-SRNet面向语义变化检测任务,强调对语义时态特征的显式建模,通过两个语义推理模块分别推断单时相和跨时相的语义相关性,并在此基础上引入余弦相似度损失以约束变化检测结果与语义预测的一致性(Ding等, 2022)。

(5) HGINet是面向高分辨率遥感影像语义变化检测的多任务分层语义图交互网络,通过时间相关性建模模块刻画双时相之间的时序依赖,并结合图卷积交互模块实现语义图之间的多层级信息交互,从而联合利用语义特征与差异特征(Long等, 2024)。

(6) MambaSCD基于Mamba状态空间架构构建视觉变化检测框架,首次将状态空间模型系统性引入遥感变化检测领域,利用Visual Mamba编码器高效建模全局空间上下文,并设计时空关系解码机制在解码阶段充分交互多时相特征,以克服传统CNN局部感受野有限和Transformer计算开销较大的固有限制(Chen等, 2024)。

(7) CdSC是一种基于交叉差异语义一致性思想的语义变化检测框架,通过3D交叉差分模块在三维空间内挖掘双时相实例特征间的深层差异,并结合对比学习驱动的语义共对齐损失,显式增强双时相语义特征的一致性(Wang等, 2024b)。

图4为不同算法在SECOND数据集3组不同样本上的定性比较结果,表明语义引导的时空协同感知网络在语义变化检测方面的优越性能,图4中

红色虚线框表示易出现误分类问题的区域。可见:在第1对样本中,本文提出模型准确全面地识别了从裸地到住宅区的过渡,其他对比方法存在将低矮植被误识别为树木的微小错误。此外,尽管DAUDT3识别出双时相图像中未变化的裸土类别,但却并未将其归为未变化区域;第2对样本表现了从裸地到建筑群的转变,尽管所有方法都检测到这一点变化,本文提出模型在保持检测精度的同时,显著增强了变化区域边界轮廓的细节;第3对样本突出了密集小尺度地物变化带来的挑战,其中建筑物附近难以区分的低矮植被与树木导致对比算法中大量的错误检测。相比之下,提出的模型凭借其特征解调与语义引导机制,有效缓解了这种干扰,表现出更强的鲁棒性。表1为各种方法在SECOND数据集上的性能比较结果。值得注意的是,提出的方法在所有评估指标上都有显著改进。它在mIoU (73.62%)、SeK (24.19%)和F1 (64.27%)方面实现了最高性能,优于所有其他方法。特别是,本文方法显示出比次优方法CdSC的显著改进,mIoU增加了0.30%,SeK增加了0.67%,F1增加了0.57%。这凸显了SemSTNet在准确检测变化的同时,在所有关键指标上保持强劲的整体性能的卓越能力。

表1展示了各方法在SECOND数据集上的定量对比结果。整体来看,本文提出的SemSTNet在精度与效率之间取得了更优的平衡,并在主要检测指标上实现了领先表现。在精度方面,SemSTNet取得了最佳的mIoU为73.62%,SeK为24.19%,F1为64.27%,3项关键指标均优于所有对比方法。与排名第二的CdSC相比,mIoU提高0.30个百分点,SeK提高0.67个百分点,F1提高0.57个百分点。从模型复杂度与推理效率的角度看,SemSTNet的参数量为27.85 M,显存占用为424.17 MB,明显低于CdSC的38.6 M参数量与482.9 MB显存开销,同时推理时间也更短,为227.22 ms,而CdSC为258.74 ms。与此同时,部分方法虽然显存占用较低,但在mIoU、SeK与F1上存在明显差距,说明仅降低资源开销并不能保证可靠的变化检测效果。综上,SemSTNet不仅在SECOND数据集上实现了最优的检测性能,在参数规模、显存占用与推理速度方面展现出更强的综合竞争力,体现了其在实际部署场景中的应用价值。

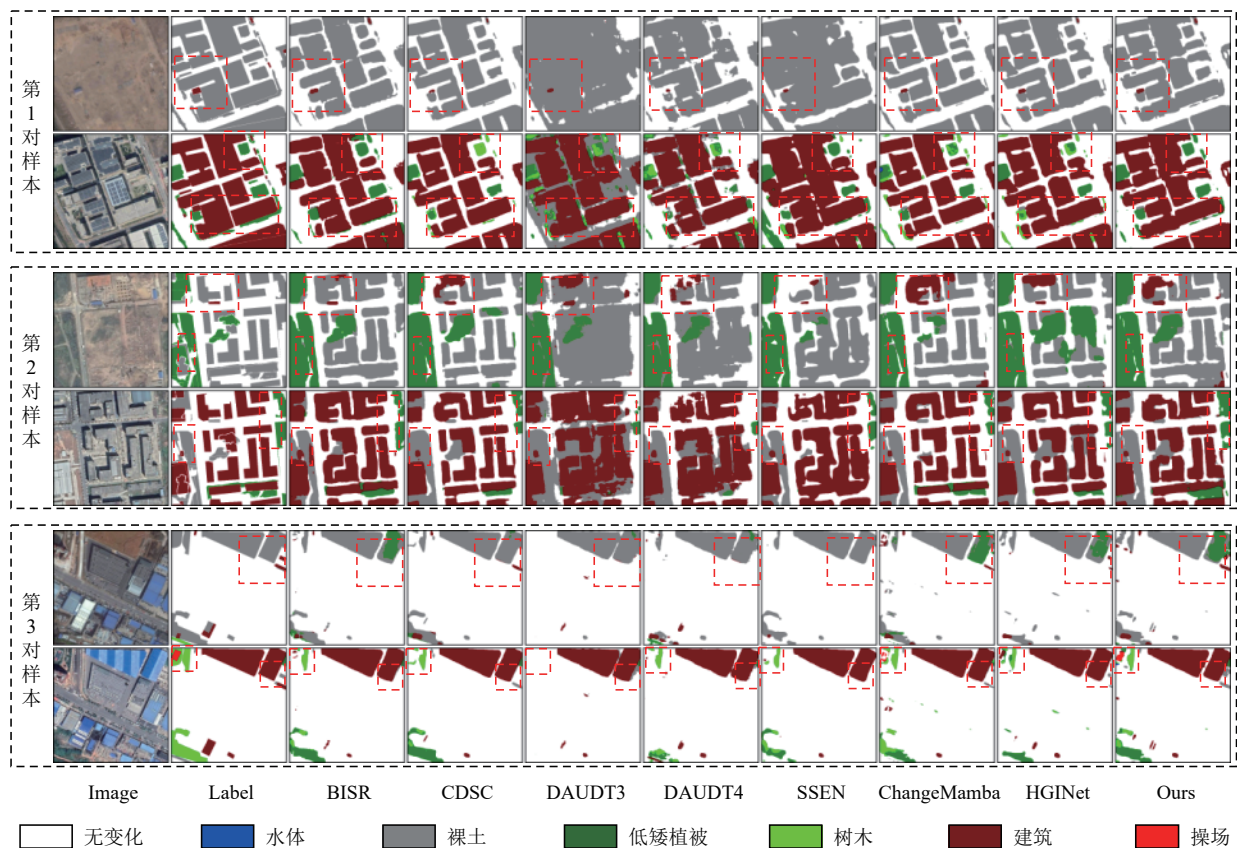


图4 不同方法在SECOND数据集3组不同样本上的语义变化检测结果

Fig. 4 Semantic change detection results of different methods on the SECOND dataset

表1 本文提出模型与其他7种模型在SECOND上的定量结果对比

Table 1 Quantitative results comparison between the proposed model and seven other models on SECOND							
模型	参数量/M	mIoU/%	SeK/%	F1/%	FLOPs/G	推理时间/ms	显存/MB
HRSCD.str3	12.77	63.34	8.27	48.54	42.94	42.95	180.83
HRSCD.str4	13.71	69.47	15.48	55.61	43.97	44.88	253.62
SSEN	51.31	71.26	20.45	61.58	280.14	36.91	605.72
HGINet	27.70	71.69	21.66	62.70	50.64	62.79	238.46
Bi-SRNet	23.38	72.65	21.91	62.01	190.30	21.38	377.80
MambaSCD	97.03	73.12	21.83	54.93	211.55	105.76	1104.46
CdSC	38.60	73.32	23.52	63.70	134.80	258.74	482.90
SemSTNet(Ours)	27.85	73.62	24.19	64.27	87.14	227.22	424.17

注：数值加粗代表最优结果。

图5给出了各种方法在农田变化数据集JL1上的可视化结果。第1个图像对展示了在农田中修建道路的变化过程，尽管大多数方法都成功检测到了这一变化，但提出的方法检测结果最为完整，边界保持更为精细。图5中的第2对图像表现了从农田到住宅区的土地使用变化。在该示例中，大部分对比方法出现了大面积类别分类错误，例如将非植被地面误识别为颜色相似的道路。同时，CdSC错误地在双时相语义映射中将变化区域分类到同一类别中，导致语义不一致。得益于提出的对比变化损失提供的有效监督，本文方法成功避

免了这一问题，确保了更精确的变化检测结果。第3对图像显示了农田向工业用地的转变。与其他竞争方法相比，提出的模型捕获了更多细节，并在检测建筑物和非植被地面变化方面保持了更高的准确性。表2显示了JL1数据集的结果。在这方面，SemSTNet再次表现出色，在所有指标上都取得了最佳表现：mIoU（86.51%）、SeK（59.02%）和F1（88.01%），证明了其在具有复杂季节性光谱变化的挑战性数据集上的出色稳健性。这种性能差异进一步强调了SemSTNet在处理不同数据和提供高度准确的变化检测方面的卓越能力。

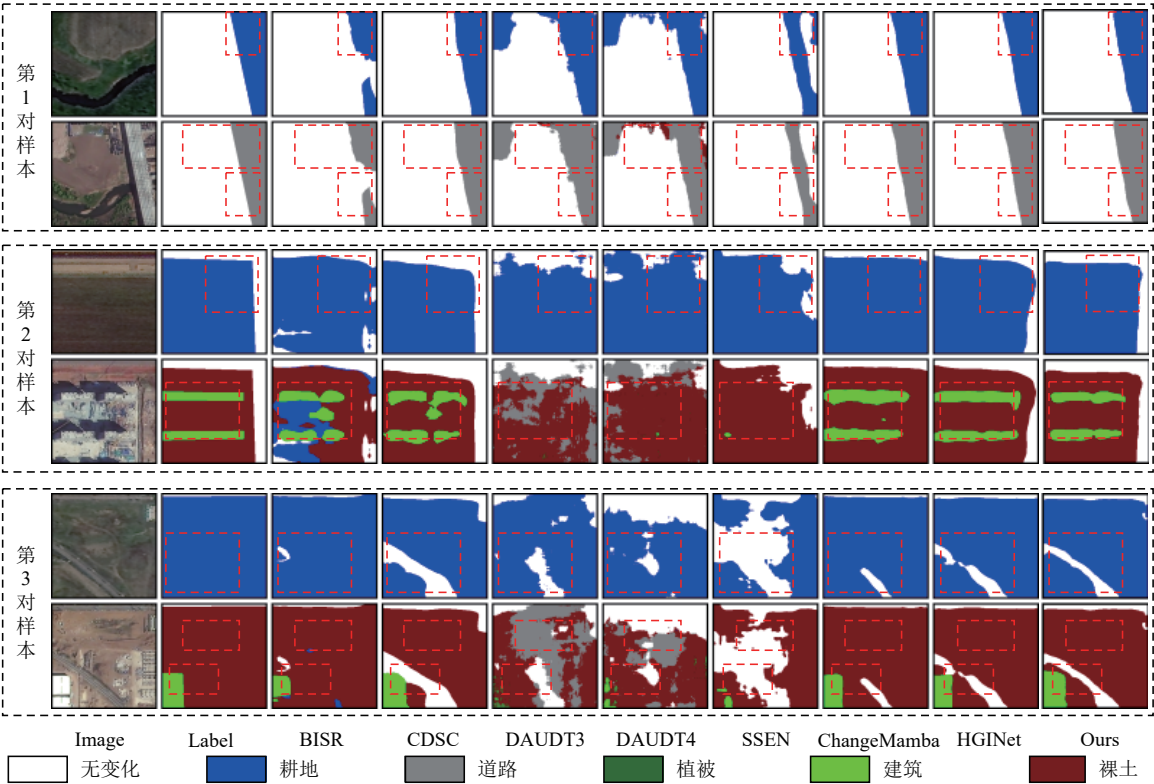


图5 不同方法在JL1数据集上的语义变化检测结果

Fig. 5 Semantic change detection results of different methods on the JL1 dataset

表2 本文提出模型与其他7种模型在JL1上的定量结果对比

Table 2 Quantitative results comparison between the proposed model and seven other models on JL1							
模型	参数量/M	mIoU/%	SeK/%	F1/%	FLOPs/G	推理时间/ms	显存/MB
HRSCD.str3	12.77	63.87	10.6	51.94	10.74	42.91	84.20
HRSCD.str4	13.71	64.86	10.54	51.74	10.99	45.66	142.06
SSEN	51.31	67.92	20.30	62.67	69.67	20.62	310.08
Bi-SRNet	23.38	81.03	45.39	81.39	47.54	15.68	118.04
HGINet	27.70	85.40	56.77	87.40	12.65	43.29	146.52
MambaSCD	97.03	86.44	58.85	86.62	52.89	73.07	561.77
CdSC	38.60	85.10	55.89	86.98	33.70	53.14	234.44
SemSTNet(Ours)	27.85	86.51	59.02	88.01	21.79	47.92	185.4

注：数值加粗代表最优结果。

为了评估提出的模块和损失函数对语义变化检测的影响，本研究使用SECOND数据集进行了一系列全面的消融研究。这些研究系统地评估了选择性去除特定成分的效果，以促进比较分析，结果如表3所示。可见随着每个附加组件的增量集成，模型性能不断提高。

(1) 基线：基线网络采用与SemSTNet相同的骨干架构。在解码器中，利用骨干特征来生成双时相语义图 and 变化图，而不通过SSDM模块合并差异特征。另外，HDC解码器被排除在外。使用交叉熵损失函数训练基线以确保公平比较，所有其他超参数在所有实验设置中保持恒定。在这些条

件下，基线模型在SECOND数据集上实现了70.60%的mIoU和19.09%的SeK。

表3 SECOND数据集各模块消融实验结果对比

Table 3 Ablation study results of different modules on the SECOND dataset					
SSDM	SGFD	CCL	mIoU/%	SeK/%	F1/%
			70.60	19.09	59.34
√			72.45	21.73	61.96
	√		72.03	21.66	61.74
		√	72.37	21.96	62.37
√	√		72.81	22.76	63.03
	√	√	72.84	22.59	63.10
√		√	72.81	22.58	63.07
√	√	√	73.62	24.19	64.27

注：“√”表示当前模型集成了相关模块。

(2) SSDM 模块的有效性: 为了增强模型对时空变化的敏感性引入 SSDM 模块。该模块通过利用双向特征调制和 3D 差分卷积技术, 提高了模型感知不同目标变化的能力。如表 3 所示, SSDM 模块显著增强了模型的整体性能, 特别是在 mIoU 和 SeK 指标方面实现了显著的改进。例如, 第 3 行中的结果表明, 在集成 SSDM 模块后, mIoU 从 70.60% 增加到 72.45%, SeK 从 19.09% 改善到 21.73%。为

了证明 SSDM 模块在提取鲁棒差分特征时的作用, 本节在有或没有 SSDM 模块的情况下可视化了梯度类激活图 (Grad-CAM), 其中不使用 SSDM 的模型使用普通卷积进行差分建模。如图 6 所示, 与传统卷积相比, SSDM 模块更全面地提取变化特征。此外, 它显著增强了对象变化区域, 突出了其在提取判别差异特征方面的优势。

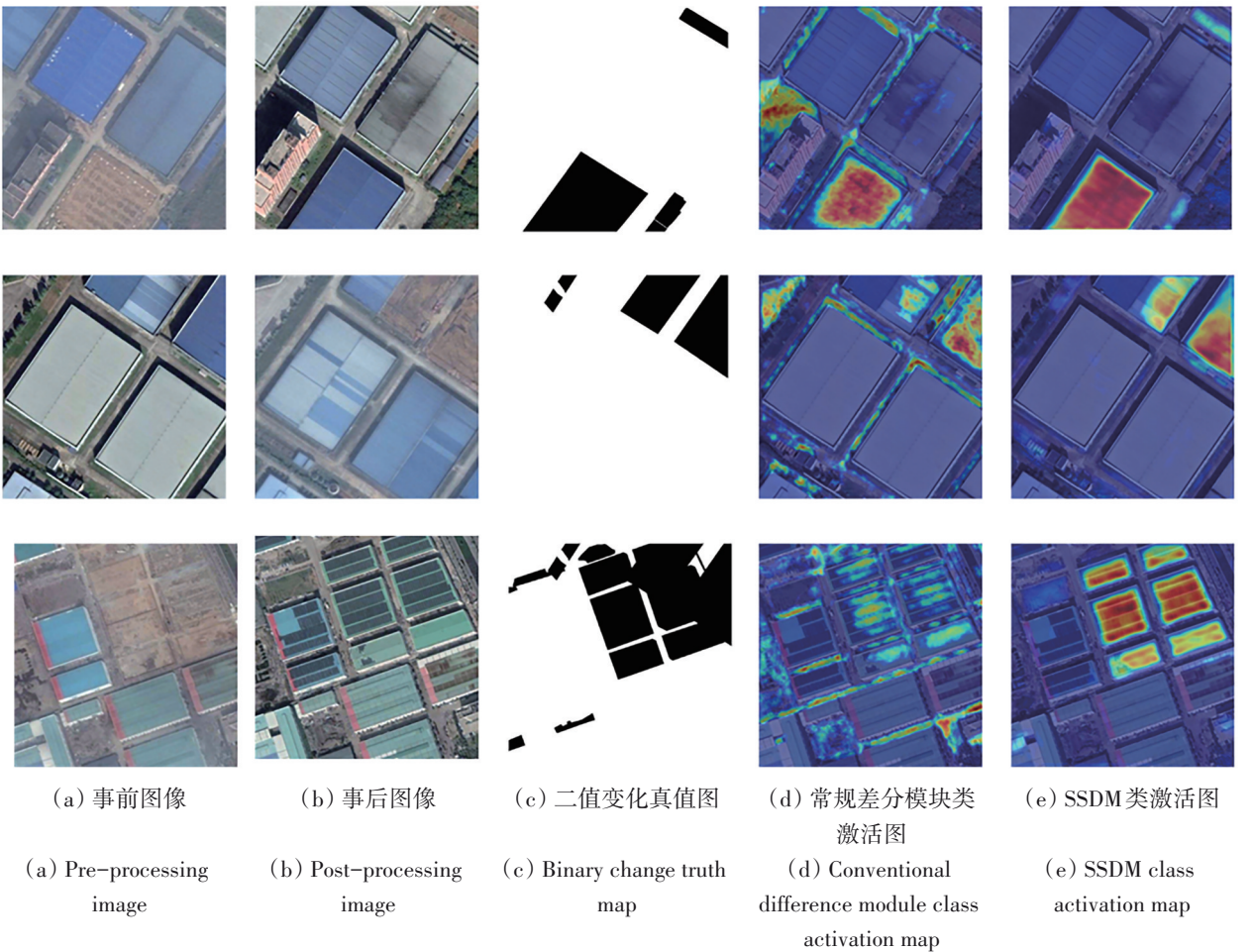


图 6 SSDM 模块的 Grad-CAM 可视化示例

Fig. 6 Grad-CAM visualisation examples of the SSDM module

(3) SGFD 变化解码器的有效性: 为了解决复杂场景下准确解码变化区域的挑战, 本文设计了 SGFD 变化解码器。该解码器采用动态加权机制来生成变化先验, 并通过自适应门控融合增强真实变化区域的表达。如表 3 所示, SGFD 解码器特别改进 SeK 度量。第 4 行的实验结果表明, 集成 SGFD 解码器后, SeK 度量从 19.09% 增加到 21.66%。此外, SGFD 解码器略微改进了模型的 mIoU 指标, 将其从 70.60% 提高到 72.03%。

(4) CCL 损失函数的有效性: 为了进一步减少语义冲突并优化模型的变化预测准确性, 引入了 CCL 损失函数。该损失函数通过加强正负样本之间的区分, 有效地提高了模型的变化检测性能。根据表 3 中的实验结果, CCL 显著增强了 mIoU 和 SeK 度量。可见在引入 CCL 损失后, mIoU 从 70.60% 增加到 72.37%, SeK 从 19.09% 改善到 21.96%。

(5) CCL 损失函数参数敏感性分析: 为进一步验证深度语义对比损失的鲁棒性并确定最优超

参数配置,本文在SECOND数据集上对间隔约束 γ 和负样本数量 K 进行了参数敏感性实验。如图7所示绿色框所示,当 $\gamma = 0.5$ 、 $K = 5$ 时,模型在两项指标上均达到最优(mIoU=73.62%, SeK=24.19%),最优点在热力图中呈现明显峰值。从热力图的分布特征可以观察到:对于间隔约束参数 γ ,当其过小(如0.1)时,变化区域与未变化区域在特征空间中的分离程度不足,导致对比约束效果较弱;当 γ 过大(如0.9)时,过强的间隔约束反而增加

了训练难度,性能出现下降。对于负样本数量 K ,当 K 过小(如1)时,对比样本不足,难以形成有效的特征约束;当 K 过大(如10)时,虽然计算开销增加,但性能收益递减,甚至可能因引入噪声样本而略有下降。值得注意的是,在参数中间区域($\gamma \in [0.3, 0.7]$, $K \in [3, 8]$)内,模型均能保持相对稳定的性能表现,表明所提出的深度语义对比损失对参数选择具有良好的鲁棒性。

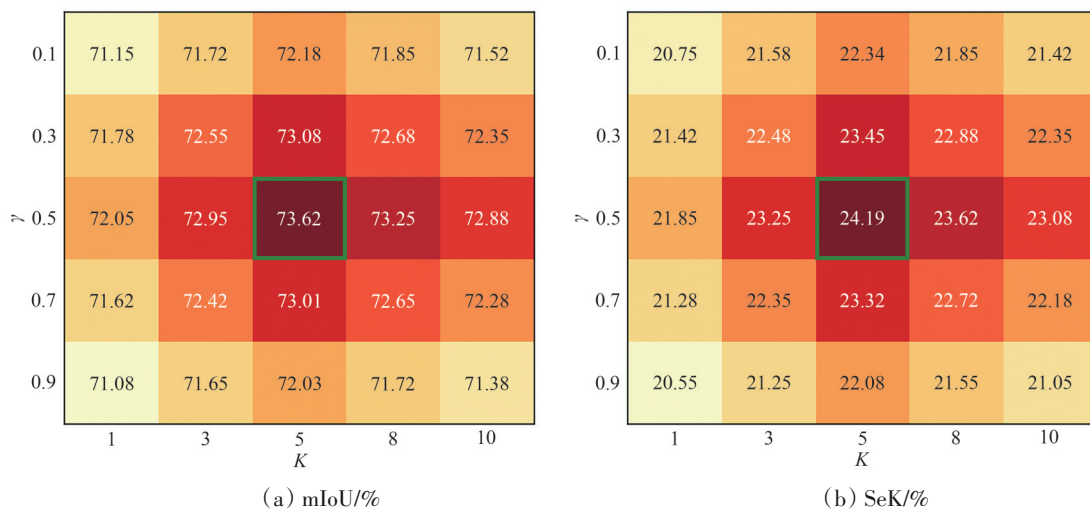


图7 CCL参数敏感性分析热力图

Fig. 7 Parameter sensitivity heatmap of CCL

5 结 论

为解决语义变化检测中对复杂地物变化识别能力不足的问题,本文提出了一种语义引导的时空协同感知网络。该方法通过引入对称时空解调差分模块,实现了跨时相语义特征的动态互校准,增强了模型对时空变化模式的刻画能力;语义门控融合解码器则进一步利用语义先验进行差异特征的动态加权,引导模型关注真实变化区域;配合深层语义对比损失函数在特征空间施加双重监督约束,从优化机制上提升了变化/未变化区域的判别能力。该协同设计有效打破了以往模型对变化与语义处理割裂的局限,显著提升了在多元复杂变化场景下的检测准确性。结果表明,本文方法在多个公开数据集上均取得了优于现有主流方法的性能,验证了所提出方法在变化区域定位精度和语义识别一致性方面的有效性。特别是在地

物类别转换频繁或变化尺度差异显著的场景中,提出的网络展现出了更强的鲁棒性和泛化能力。未来研究将从两个方面进一步推进:其一,探索轻量级协同结构设计,在保持检测精度的同时提升模型效率;其二,结合多模态遥感数据挖掘跨模态语义关联,引导模型更有效地区分伪变化与真实语义变更,从而进一步提升在真实场景中的应用能力。

参考文献(References)

- Bandara W G C and Patel V M. 2022. A transformer-based Siamese network for change detection//Proceedings of the 2022 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Kuala Lumpur: IEEE: 207-210 [DOI: 10.1109/IGARSS46834.2022.9883686]
- Caye Daudt R, Le Saux B, Boulch A and Gousseau Y. 2019. Multitask learning for large-scale semantic change detection. Computer Vision and Image Understanding, 187: 102783 [DOI: 10.1016/j.cviu.

- 2019.07.003]
- Chen H R X, Song J, Han C X, Xia J S and Yokoya N. 2024. Change-Mamba: remote sensing change detection with spatiotemporal state space model. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 4409720 [DOI: 10.1109/TGRS.2024.3417253]
- Chen P, Zhang B, Hong D F, Chen Z C, Yang X and Li B P. 2022. FCCDN: feature-constraint network for VHR image change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187: 101-119 [DOI: 10.1016/j.isprsjprs.2022.02.021]
- Chi K C, Yuan Y and Wang Q. 2023. Trinity-Net: gradient-guided Swin Transformer-based remote sensing image dehazing and beyond. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 4702914 [DOI: 10.1109/TGRS.2023.3285228]
- Cui B, Zhang Y H, Yan L and Wei J J. 2020. Dual-thresholds change detection in GF-3 SAR images. *Journal of Remote Sensing (in Chinese)*, 24(1): 1-10 (崔斌, 张永红, 闫利, 魏钜杰. 2020. 高分三号 SAR 影像双阈值变化检测. *遥感学报*, 24(1): 1-10) [DOI: 10.11834/jrs.20208179]
- 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]
- Du P J, Wang X, Chen D M, Liu S C, Lin C and Meng Y P. 2020. An improved change detection approach using tri-temporal logic-verified change vector analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161: 278-293 [DOI: 10.1016/j.isprsjprs.2020.01.026]
- Garcia-Aguirre M C, Alvarez R, Dirzo R and Bernal A. 2005. Post-classification digital change detection analysis of a temperate forest in the southwest basin of Mexico City, in a 16-year span//Proceedings of the 2005 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images. Biloxi: IEEE: 81-84 [DOI: 10.1109/AMTRSI.2005.1469845]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Javed A, Jung S, Lee W H and Han Y. 2020. Object-based building change detection by fusing pixel-level change detection results generated from morphological building index. *Remote Sensing*, 12(18): 2952 [DOI: 10.3390/rs12182952]
- Jing W, Yuan Y and Wang Q. 2023. Dual-field-of-view context aggregation and boundary perception for airport runway extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 4702412 [DOI: 10.1109/TGRS.2023.3271676]
- Lei T, Wang J, Ning H L, Wang X W, Xue D H, Wang Q and Nandi A K. 2022a. Difference enhancement and spatial-spectral nonlocal network for change detection in VHR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 4507013 [DOI: 10.1109/TGRS.2021.3134691]
- Lei T, Xue D H, Ning H L, Yang S M, Lv Z Y and Nandi A K. 2022b. Local and global feature learning with kernel scale-adaptive attention network for VHR remote sensing change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 7308-7322 [DOI: 10.1109/JSTARS.2022.3200997]
- Lei Y J, Peng D, Zhang P P, Ke Q H and Li H F. 2021. Hierarchical paired channel fusion network for street scene change detection. *IEEE Transactions on Image Processing*, 30: 55-67 [DOI: 10.1109/TIP.2020.3031173]
- Li J K, Zhang G X, Zhu S Y, Xu Y M and Li X Y. 2025. Change detection for high-resolution remote sensing images with multi-scale feature transformer. *National Remote Sensing Bulletin*, 29(1): 266-278 (李健慷, 张桂欣, 祝善友, 徐永明, 李湘雨. 2025. 融合多尺度特征 Transformer 的高分辨率遥感图像变化检测. *遥感学报*, 29(1): 266-278) [DOI: 10.11834/jrs.20243201]
- Lin H, Hang R L, Wang S M and Liu Q S. 2024. DiFormer: a difference Transformer network for remote sensing change detection. *IEEE Geoscience and Remote Sensing Letters*, 21: 6003905 [DOI: 10.1109/LGRS.2024.3359220]
- Lin M H, Yang G Y and Zhang H Y. 2023. Transition is a process: pair-to-video change detection networks for very high resolution remote sensing images. *IEEE Transactions on Image Processing*, 32: 57-71 [DOI: 10.1109/TIP.2022.3226418]
- 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 T F, Xu J J, Lei T, Wang Y B, Du X G, Zhang W C, Lv Z Y and Gong M G. 2025a. AEKAN: exploring Superpixel-based AutoEncoder Kolmogorov-Arnold network for unsupervised multimodal change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 5601114 [DOI: 10.1109/TGRS.2024.3515258]
- Liu T F, Zhang M Y, Gong M G, Zhang Q F, Jiang F L, Zheng H H and Lu D. 2025b. Commonality feature representation learning for unsupervised multimodal change detection. *IEEE Transactions on Image Processing*, 34: 1219-1233 [DOI: 10.1109/TIP.2025.3539461]
- Long J, Li M M, Wang X Q and Stein A. 2024. Semantic change detection using a hierarchical semantic graph interaction network from high-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211: 318-335 [DOI: 10.1016/j.isprsjprs.2024.04.012]
- Luo Z P, Gao L P, Xiang H D and Li J. 2023. Road object detection for HD map: full-element survey, analysis and perspectives. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197: 122-144 [DOI: 10.1016/j.isprsjprs.2023.01.009]
- 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]

- Mou L C, Bruzzone L and Zhu X X. 2019. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2): 924-935 [DOI: 10.1109/TGRS.2018.2863224]
- Radke R J, Andra S, Al-Kofahi O and Roysam B. 2005. Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing*, 14(3): 294-307 [DOI: 10.1109/TIP.2004.838698]
- Ru L X, Du B and Wu C. 2021. Multi-temporal scene classification and scene change detection with correlation based fusion. *IEEE Transactions on Image Processing*, 30: 1382-1394 [DOI: 10.1109/TIP.2020.3039328]
- Serra P, Pons X and Saurí D. 2003. Post-classification change detection with data from different sensors: some accuracy considerations. *International Journal of Remote Sensing*, 24(16): 3311-3340 [DOI: 10.1080/0143116021000021189]
- Sun S T, Mu L, Wang L Z and Liu P. 2022. L-UNet: an LSTM network for remote sensing image change detection. *IEEE Geoscience and Remote Sensing Letters*, 19: 8004505 [DOI: 10.1109/LGRS.2020.3041530]
- Tang K, Xu F, Chen X H, Dong Q, Yuan Y H and Chen J. 2024. The ClearSCD model: comprehensively leveraging semantics and change relationships for semantic change detection in high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211: 299-317 [DOI: 10.1016/j.isprsjprs.2024.04.013]
- Tang Y Q, Lin Z F, Han T, Yang X, Zou B and Feng H H. 2024. Optical and SAR image change detection based on a symmetric network. *National Remote Sensing Bulletin*, 28(6): 1560-1575 (汤玉奇, 林泽锋, 韩特, 杨欣, 邹滨, 冯徽徽. 2024. 基于对称网络的光学和 SAR 影像变化检测. *遥感学报*, 28(6): 1560-1575) [DOI: 10.11834/jrs.20232027]
- Wang D C, Zhao F, Wang C, Wang H Y, Zheng F J and Chen X N. 2022. Y-Net: a multiclass change detection network for Bi-temporal remote sensing images. *International Journal of Remote Sensing*, 43(2): 565-592 [DOI: 10.1080/01431161.2021.2022241]
- Wang J, Zhong Y F and Zhang L P. 2024a. Semantic change detection based on supervised contrastive learning for high-resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 5649720 [DOI: 10.1109/TGRS.2024.3484178]
- Wang Q, Jing W, Chi K C and Yuan Y. 2024b. Cross-difference semantic consistency network for semantic change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 4406312 [DOI: 10.1109/TGRS.2024.3386334]
- 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]
- Wu C, Du B and Zhang L P. 2014. Slow feature analysis for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5): 2858-2874 [DOI: 10.1109/TGRS.2013.2266673]
- 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 (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) [DOI: 10.11834/jrs.20243419]
- Yang K P, Xia G S, Liu Z C, Du B, Yang W, Pelillo M and Zhang L P. 2022a. 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]
- Yang Y F, Qu J H, Xiao S, Dong W Q, Li Y S and Du Q. 2022b. A deep multiscale pyramid network enhanced with spatial-spectral residual attention for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5525513 [DOI: 10.1109/TGRS.2022.3161386]
- Zhang C X, Yue P, Tapete D, Jiang L C, Shangguan B Y, Huang L and Liu G C. 2020. A deeply supervised image fusion network for change detection in high resolution Bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166: 183-200 [DOI: 10.1016/j.isprsjprs.2020.06.003]
- Zhang X F, Cheng S L, Wang L J and Li H J. 2023. Asymmetric cross-attention hierarchical network based on CNN and Transformer for bitemporal remote sensing images change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 2000415 [DOI: 10.1109/TGRS.2023.3245674]
- 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. Change-Mask: 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]
- Zhu P P, Xu H and Luo X B. 2023. MDAFormer: multi-level difference aggregation transformer for change detection of VHR optical imagery. *International Journal of Applied Earth Observation and Geoinformation*, 118: 103256 [DOI: 10.1016/j.jag.2023.103256]
- Zhu Q Q, Guo X, Deng W H, Shi S N, Guan Q F, Zhong Y F, Zhang L P and Li D R. 2022a. 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]
- Zhu Q Q, Lei Y, Sun X L, Guan Q F, Zhong Y F, Zhang L P and Li D R. 2022b. Knowledge-guided land pattern depiction for urban land use mapping: a case study of Chinese cities. *Remote Sensing of Environment*, 272: 112916 [DOI: 10.1016/j.rse.2022.112916]

Spatiotemporal change collaborative perception method for remote sensing semantic change detection

JING Wei^{1,2}, YAN Tingxuan¹, JIN Xianglong¹, SONG Binbin¹, WANG Qi¹

1. School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an 710072, China;

2. National Elite Institute of Engineering, Northwestern Polytechnical University, Xi'an 710072, China

Abstract: Semantic Change Detection (SCD) aims to accurately identify complex land cover changes and simultaneously determine their corresponding semantic categories from bitemporal remote sensing images. Existing methods typically employ separate change localization and semantic recognition branches to accomplish SCD tasks. However, these frameworks often fail to fully integrate and leverage semantic information to guide the change detection process, which reduces their robustness in complex scenarios, seasonal and illumination variations, and multiscale small-target changes. To address this limitation, this study proposes a Semantic-guided Spatio-Temporal Collaborative Perception Network (SemSTNet) that incorporates semantic guidance into the precise identification of change regions while considering spatiotemporal interactions of bitemporal features.

The proposed SemSTNet consists of three key components. First, a symmetric spatiotemporal demodulation differential module (SSDM) is constructed. It employs bidirectional feature modulation to achieve cross-temporal feature mutual calibration and explicitly model intrinsic spatiotemporal difference patterns using 3D central difference convolution, thereby extracting discriminative difference features from bitemporal images. Second, a Semantic Gated Fusion Decoder (SGFD), which introduces a semantic similarity dynamic weighting mechanism, is designed. It generates semantic priors through cosine similarity maps and enhances the saliency representation of genuine change regions via adaptive gated fusion. Third, a Contrastive Change Loss (CCL) is proposed to construct dual supervision signals. It constrains the semantic consistency of unchanged regions in the feature space and enlarges the interclass distance between changed and unchanged regions through margin-based contrastive learning.

Extensive experiments are conducted on two SCD datasets: the SECOND dataset and the JL1 dataset. On the SECOND dataset, SemSTNet achieves the best performance with an mIoU of 73.62%, a separated kappa (SeK) of 24.19%, and an F1 score of 64.27%, surpassing the second-best method CdSC by 0.30%, 0.67%, and 0.57%, respectively, while maintaining lower parameter count (27.85 M vs. 38.6 M) and reduced memory consumption (424.17 MB vs. 482.9 MB). On the JL1 dataset, SemSTNet again achieves the highest scores across all metrics, with an mIoU of 86.51%, a SeK of 59.02%, and an F1 of 88.01%, demonstrating strong robustness against complex seasonal spectral variations. Ablation studies on the SECOND dataset confirm the effectiveness of each proposed component, with the full model improving mIoU from 70.60% to 73.62% and SeK from 19.09% to 24.19% over the baseline.

The proposed SemSTNet effectively breaks the conventional separation between change detection and semantic recognition by establishing a collaborative optimization mechanism among SSDM, SGFD, and CCL. Experimental results on multiple public datasets validate the superiority of the proposed method in change localization accuracy and semantic recognition consistency, particularly in scenarios with frequent land cover type transitions and significant scale variations. Future work will focus on exploring lightweight collaborative architectures and incorporating multimodal remote sensing data to further enhance practical applicability.

Key words: remote sensing images, land cover change, semantic change detection, spatio-temporal collaboration, semantic guidance

Supported by National Natural Science Foundation of China (No. 62471394, 62301385 and U21B2041), the Innovation Foundation for Doctoral Dissertation of Northwestern Polytechnical University (No. CX2024108)