

光学信号 Token 引导的异源遥感变化检测网络

刘秦森¹, 孙帮勇^{1,2}

1. 西安理工大学 印刷包装与数字媒体学院, 西安 710054;

2. 中国科学院西安光学精密机械研究所 瞬态光学与光子技术国家重点实验室, 西安 710119

摘要: 基于遥感图像中的光学信号检测出一定时间内特定区域的变化状态的遥感图像变化检测方法, 在国防安全、环境监测、城市建设等领域具有重要应用价值。由于多时相异源图像在成像机理、光谱范围、空间分辨率等方面存在差异, 现阶段异源遥感图像变化检测仍存在精度不够高、漏检和误检等问题, 本文提出一种基于 Transformer 网络的异源变化检测网络框架, 该框架能够利用不同类别的异源遥感图像获得准确的变化检测结果。首先, 所提出检测网络为多时相遥感图像自适应生成对应的光学信号 Token (光信 Token); 然后, 以光信 Token 作为引导与对应图像块 Token 进行交互计算, 从而对双时相序列特征进行变化分析, 并且在交互学习过程中构建了差分放大模块以提高网络对特征间差分信息的提取精度; 最后, 利用多层感知机对输出的差分 Token 进行预测并分割出变化区域。采用 Sardinia、Shuguang 和 Bastrop 等 3 个不同类别的异源遥感图像数据集和 Farmland 同源高光谱图像数据集来验证本文提出的方法, 结果证明在选取有限训练样本数据情况下, 本文方法与现有主流变化检测方法相比, 在多个客观指标以及主观视觉上都表现出先进性。

关键词: 遥感, 异源图像, 变化检测, 多模态分析, 深度学习, Transformer

中图分类号: TP701/P2

引用格式: 刘秦森, 孙帮勇. 2024. 光学信号 Token 引导的异源遥感变化检测网络. 遥感学报, 28(1): 88-104

Liu Q S and Sun B Y. 2024. Optical-signal token guided change detection network for heterogeneous remote sensing image. National Remote Sensing Bulletin, 28(1): 88-104[DOI: 10.11834/jrs.20233067]

1 引言

遥感变化检测通过分析多时相遥感图像中的各像素数据差值, 能够快速、有效地检测地物目标的变化状态, 已被广泛用于建筑物、水体面积、农林覆盖等目标的监测中 (Adão 等, 2017; 柳思聪 等, 2022)。随着遥感探测技术的快速发展, 所获取的图像数量和质量都得到较大提升, 极大丰富了变化检测的数据源。根据输入的遥感图像数据类型, 变化检测可分为同源图像变化检测和异源图像变化检测。前者的数据源多由同一传感器在不同时间对特定区域采集获得, 而后者数据源一般由不同类型传感器在不同时间获取。由于异源图像存在成像类型和数据定义形式等显著差异, 图像中各通道的成像波段、分辨率以及物理意义都有所差

异。例如 RGB 图像和合成孔径雷达 (SAR) 图像组成的异源图像数据对中, RGB 图像包含红绿蓝 3 个通道的成像信息, 而 SAR 图像的多个通道则蕴含了地物目标的粗糙度、坡度和含水量等信息, 二者在每个像素位置上的定义显著不同。

一般认为, 基于遥感图像的变化检测方法可分为基于特定数学模型的传统方法和基于深度学习的方法两大类 (Yang 等, 2018; Sun 等, 2020)。传统变化检测方法主要包括基于代数、基于转换和基于分类等几类方法 (Hasanlou 和 Seydi, 2018)。基于代数的检测方法通过对图像进行差分计算图像比来提取发生变化的信息。其中, 变化向量分析 CVA (Change Vector Analysis) 是一种代表性的代数类变化检测方法。其通过计算两幅图像各像素之间的欧氏距离, 生成差值图像来表达变化信

收稿日期: 2023-04-13; 预印本: 2023-10-03

基金项目: 国家自然科学基金 (编号: 62076199); 陕西省重点研发计划 (编号: 2022ZDLGY01-03); 瞬态光学与光子技术国家重点实验室开放基金 (编号: SKLST202214); 陕西省教育厅重点科学研究计划 (编号: 23JY063)

第一作者简介: 刘秦森, 研究方向为人工智能与遥感图像变化检测方法。E-mail: 2371759407@qq.com

通信作者简介: 孙帮勇, 研究方向为光谱成像技术与光谱图像处理与复现。E-mail: sunbangyong@xaut.edu.cn

息, 已经广泛应用 (Bovolo 和 Bruzzone, 2007)。为了进一步提高这种方法的检测效果, 学者们提出了一系列的 CVA 改进方法。例如: Bovolo 等 (2012) 提出了压缩 CVA 方法 (C2VA), 该方法可以从处理序列信息的角度实现无监督变化检测; Pirrone 等 (2020) 提出基于 CVA 的多类别变化检测方法, 对双时相 SAR 图像实现无监督多类变化检测。然而, 以上方法存在多类别变化检测精度低、阈值选择难等问题。基于转换的变化检测方法则通过将双时相遥感图像输入到特定空间中来分析变化特征与未变化特征, 如主成分分析法 (PCA) (Ma 等, 2019)、慢特征分析法 (SFA) (Wu 等, 2014)、迭代重加权多元变化检测法 (IR-MAD) (Canty 和 Nielsen, 2008) 等。这类方法对高光谱图像降维或去噪较为有效, 但同时存在模型中阈值选择难等问题。基于分类的变化检测方法, 通过对分类后的图像进行分析以判断各位置是否为变化像素。支持向量机 (SVM) 与极限学习机 (ELM) 是较为典型的分类变化检测方法, 这类方法能够实现异源变化检测, 但检测精度很大程度依赖于分类器的性能。

目前, 基于深度学习的网络模型已在遥感数据中得到广泛应用 (Mandal 和 Vipparthi, 2022)。由于深度网络的特征学习能力较强, 能够有效提取图像中的光学特性和空间上下文信息, 目前已提出了多种基于特定骨干网络的变化检测方法 (杨彬 等, 2023)。例如: Lyu 等 (2016) 引入递归神经网络 (RNN) 实现双时相图像之间的变化分析。Zhan 等 (2017) 提出一种基于卷积神经网络 (CNN) 的变化检测模型。Qu 等 (2022) 引入图卷积网络 (GCN) 来实现特征提取与变化分析。不过, 以上网络模型在处理序列信息时仍存在一定问题, 如 CNN 与 GCN 较难精确提取图像中的序列信息, 特别是全局序列信息的获取 (Hong 等, 2022); 而 RNN 网络虽然能以一种有序方式学习各信号间序列特征, 但该方法受图像谱段数量影响较大, 存在梯度消失等问题 (Bengio 等, 1994), 这些问题很大程度上限制了深度学习变化检测模型的精度。

近年来, Dosovitskiy 等 (2020) 提出的 ViT (Vision Transformer) 在多种计算机视觉任务中已被证明具有较强的序列信息处理能力。并且, Bandara 和 Patel (2022) 提出基于 ViT 的变化检测网络—

Changeformer 在变化检测中表现十分优异。但是, 以上方法并未关注到异源遥感图像变化检测问题, 无法对双时相异源图像不同的信号进行自适应建模。

异源遥感图像从通道方向可看作由不同光学信号组成的序列信息, 例如: RGB 图像通常是指可见光波段传感器获取的影像数据。其由不同范围的光谱信号组成的序列信息。而获取 SAR 图像的传感器一般属于微波频段, 其可看作由微波信号组成的序列信息。因此, 考虑到 Transformer 采用多头注意力机制能够有效处理和分析序列信息, 本文针对异源遥感图像特征对现有的先进 Transformer 网络进行设计和改进, 提出一种光学信号 Token (光信 Token) 引导的异源图像变化检测网络。针对输入的异源遥感图像, 所提出的检测网络能够从通道方向自适应地对不同类型的图像进行建模, 生成对应的光信 Token, 结合图像丰富的空间上下文信息对双时相序列特征进行交互, 从而对双时相序列特征进行变化分析, 并利用所设计的差分放大模块 DAM (Differential Amplification Module) 增强双时相序列特征之间的差分信息, 准确计算各像素的变化信息。实验证明所提出的检测网络在多个异源遥感图像集上获得了较高的检测精度。

2 研究方法与实验原理

由于成像设备和拍摄条件的不同, 双时相异源遥感图像数据之间的数据特征和物理意义存在显著差异, 详见图 1。可知仅通过直接可视化对比双时相异源图像对应区域的差异很难实现准确的变化检测。因此, 在遥感图像数据多样化和复杂化的背景下, 本文以解析异源图像的深层信息并有效检测变化区域为出发点, 设计了一个新的基于 Transformer 网络的框架。该框架结构主要由光信 Token Transformer 即 OT-Former (Optical-signal Token transformer) 和跨时相 Transformer 即 CT-Former (Cross-temporal transformer) 两部分组成, 其流程如图 2 所示。具体流程如下: 首先, OT-Former 在通道方向对双时相异源图像光学信号进行自适应建模并生成对应的光信 Token; 然后, CT-Former 对双时相光信 Token 和图像块 Token 进行交互计算来对双时相序列特征进行变化分析, 并输出变化检测结果。本文所提出模型的典型特征是能够自适应地处理不同种类的遥感图像, 以下对检测流程和机理进行阐述。

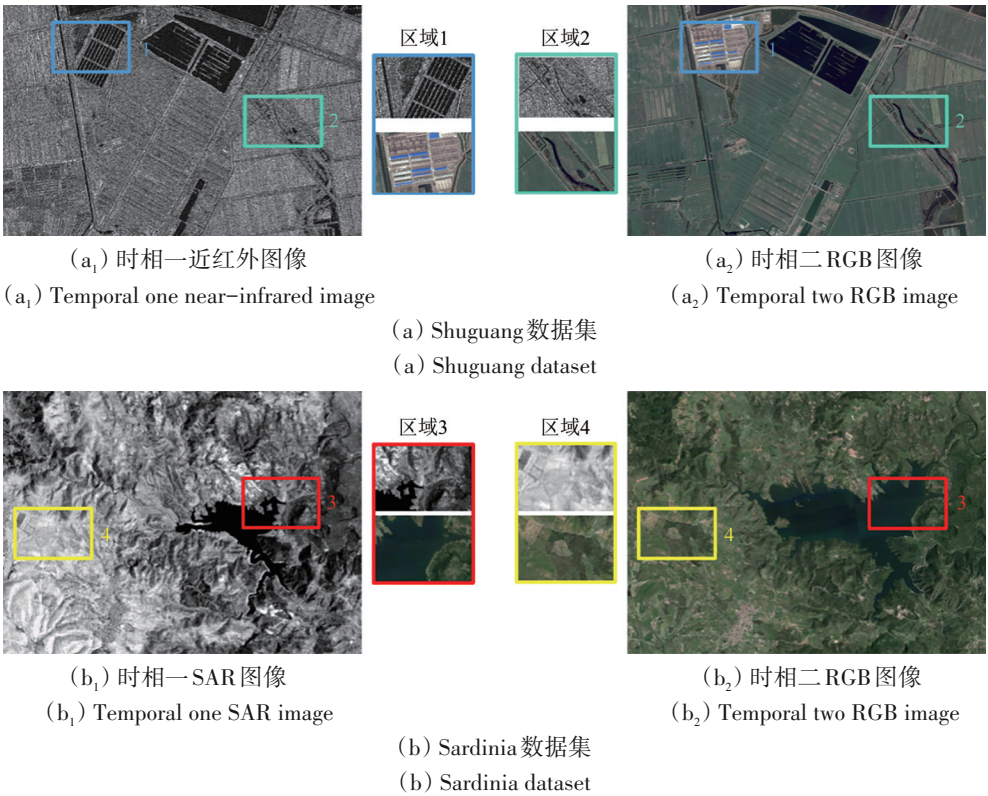


图1 异源图像光学特性差异示意图

Fig. 1 The difference of optical characteristics between heterogeneous images

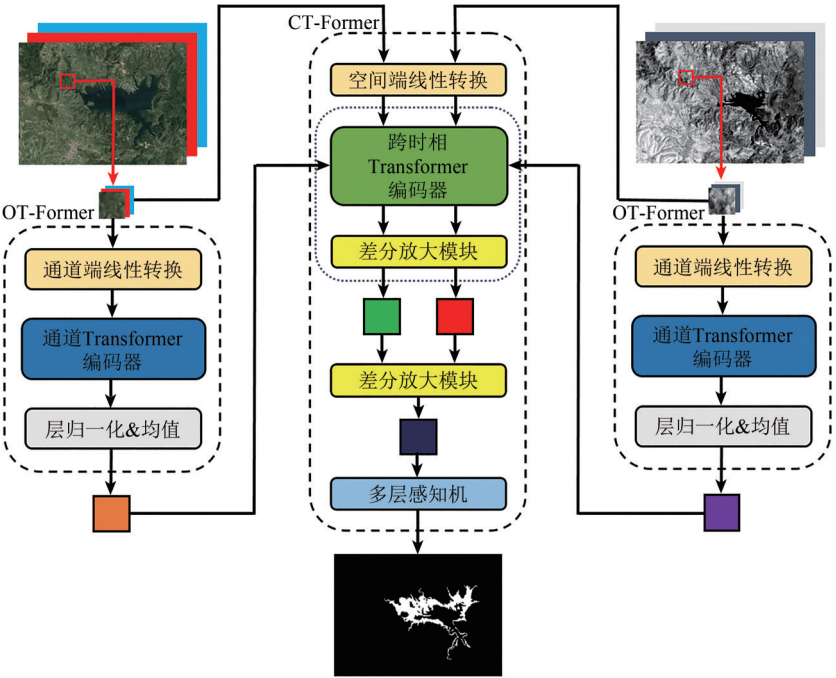


图2 本文方法的网络架构图

Fig. 2 Network architecture of our method.

2.1 OT-Former计算原理

双时相异源遥感数据对可看作为一对不同模态下的图像数据，如何对多模态数据进行自适应建模是变化检测需要解决的关键问题，而使用特

征Token表征不同模态数据已被证实是一种有效的特征建模方式。Wang等（2022b）使用ViT对多模态图像数据进行建模生成多模态Token并进行融合的方法已在多项图像任务中表现出了优异性能。

基于此，本文提出 OT-Former 在通道方向对多类异源图像数据对进行自适应建模生成对应的光信Token。

变化检测是一种像素级的检测任务，其信息输入形式主要为单像素输入和图像块输入。以像素形式输入时，检测网络通常只注重通道上信息的提取，易造成空间上下文信息的丢失；而以图像块输入则能够保持图像块空间信息与频谱信息的一致性，从而提取到更加丰富的光谱信息和空间信息（Hou 等，2021）。因此，本文将双时相图像划分为多个成对图像块进行输入。以单个时相图为例，图像块 $X \in \mathbb{R}^{c \times h \times w}$ 被转换为 $c \times h \times w$ 矩阵，并由线性转换层映射为 $X_0 \in \mathbb{R}^{c \times d}$ 的低维序列特征，其中 h 、 w 、 c 为输入图像块的高、宽以及通道数， d 为通道编码器特征维度。为了对于序列特征的提取进行有效建模，为其加入了位置信息矩阵 $PE \in \mathbb{R}^{c \times d}$ ， PE 为随机生成的可学习参数矩阵，可以以偏置的形式保留序列中向量间的顺序信息。

$$X_0 = \omega_1 X + PE \quad (1)$$

式中， ω_1 为线性映射层中可学习的网络参数。

在加入位置信息后，将序列特征 $X_0 \in \mathbb{R}^{c \times d}$ 输入至 OT-Former 中，并在通道方向上进行编码并生成对应的光信Token。

编码器的输出结果被进一步归一化和均值化，最终获得光信Token。如图3（a）所示，通道Transformer编码器主要包括多头注意力机制 MHSA（Multi-heads self-attention mechanism）与前馈网络 FF（Feed-Forward），OT-Former 通过 MHSA 来学习通道序列信息中的全局有效信息，其可以描述为

$$\begin{aligned} \text{MHSA}(Q, K, V) = & \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^0, \\ \text{head}_i = & \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d}}\right)V_i \end{aligned} \quad (2)$$

式中， h 为特征空间的数量， W^0 为可学习的参数矩阵， $Q_i = X_i W_i^q$ ， $K_i = X_i W_i^k$ ， $V_i = X_i W_i^v$ 。 W_i^q ， W_i^k 和 W_i^v 为3个可学习的参数矩阵，图3（b）为多头注意力机制计算的简要示意图。

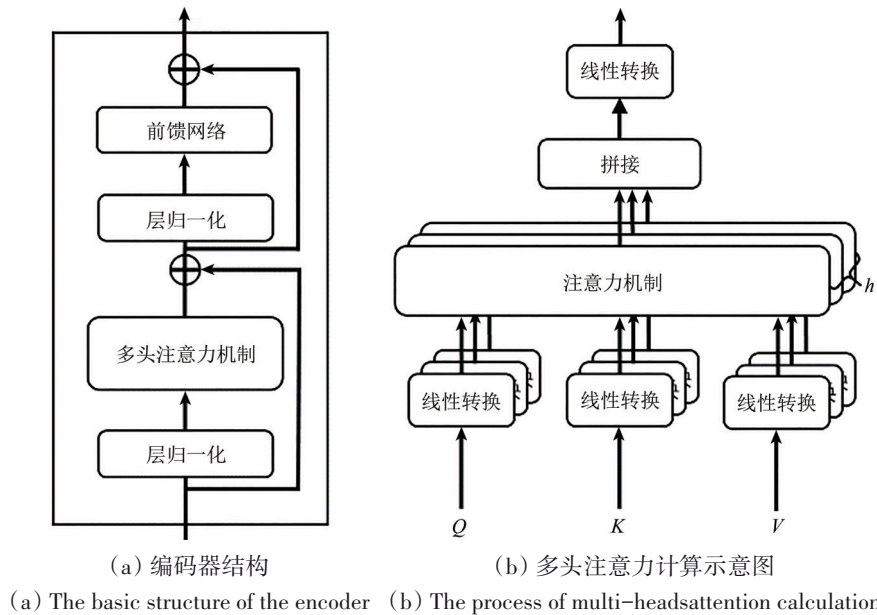


图3 Transformer编码器结构与多头注意力计算流程示意图

Fig. 3 The basic structure of the encoder and multi-heads attention calculation

在通用Transformer架构中，前馈网络FF（Feed-Forward）也是一项关键模块。FF主要由线性转换层与高斯误差线性函数GELU（Gaussian error linear unit function）组成，FF可以表示为

$$\text{FF}(X) = \text{GELU}(xW_1 + b_1)W_2 + b_2 \quad (3)$$

式中， W_1 和 W_2 为两项参数矩阵， b_1 和 b_2 为偏置。

通道Transformer编码器在通道端对双时相序列特征编码，从而有效地获取双时相图像各自独立的光信Token。结合MHSA与FF的计算过程，OT-Former的编码过程可表述为

$$X'_L = \text{MHSA}(\text{LN}(X_{L-1})), L \in \{1, 2, 3\}$$

$$X_L = \text{FF}(\text{LN}(X'_L)), L \in \{1, 2, 3\} \quad (4)$$

式中, X_L 为第 L 层编码器的序列特征, LN 为层归一化。

在经过 L 层的编码后, OT-Former 对 X_L 进行层归一化与均值处理并最终生成对应的光信 Token。

$$FT = \text{Mean}(\text{LN}(X_L)) \quad (5)$$

式中, $FT \in \mathbb{R}^{1 \times d}$ 为对应生成的光信 Token。

2.2 CT-Former 计算原理

多时相图像之间存在的变化是多样的, 如何从多样的变化中学习感兴趣目标的变化特征是本文的目的之一。大多数变化检测方法通过图像差分或拼接方式对多时相图像进行变化分析, 但不同数据集的变化类型、变化分布以及变化量都不尽相同。并且, 直接采用图像差分或拼接的方式不适用于计算序列特征间的变化信息 (Wang, 2022a)。基于此, 本文方法以光信 Token 为引导并与图像块 Token 进行交互学习实现异源图像间的变化分析, 并以监督学习的方式学习不同数据集的变化类型、变化分布以及变化量。

为了提取图像中丰富的空间语义信息, CT-Former 首先将图像块 Token 进行转置, 并在空间方

向进行线性映射以生成对应的空间序列特征 $F \in \mathbb{R}^{d \times d}$, 而对应的光信 Token 也经由线性映射层转化为 d_i 维度。然后, 转化后的光信 Token 与另一时相的空间序列特征 F 拼接, 最终生成 F_1^0, F_2^0 :

$$\begin{cases} F_1^0 = \text{Concat}(FT_1^0 \| F_1) \\ F_2^0 = \text{Concat}(FT_2^0 \| F_2) \end{cases} \quad (6)$$

式中, $FT_1^0 \in \mathbb{R}^{1 \times d_i}$ 与 $FT_2^0 \in \mathbb{R}^{1 \times d_i}$ 表示双时相图像对应的光信 Token, $F_1^0 \in \mathbb{R}^{(1+d) \times d_i}$ 和 $F_2^0 \in \mathbb{R}^{(1+d) \times d_i}$ 表示输入到 CT-Former 中对应的双分支空间序列特征。

跨时相 Transformer 编码器结构与前述编码器结构相同。本文将双时相序列特征 F_1^0 与 F_2^0 输入跨时相 Transformer 编码器, 通过计算跨时相注意力 CTA (Cross-Temporal Attention) 来统计变化信息并分析, 如图 4 (a) 所示。计算过程可用以下公式描述:

$$\begin{aligned} Q &= FT_1^0 W^q, K = F_2^0 W^k, V = F_2^0 W^v, \\ \text{Attn} &= \text{softmax}\left(\frac{QK^T}{\sqrt{d_i/r}}\right), \\ \text{CTA}(F_1^0) &= \text{Attn}V \end{aligned} \quad (7)$$

式中, W^q, W^k 与 W^v 为 3 个可训练的参数矩阵, r 为跨时相 Transformer 编码器的特征空间数量。

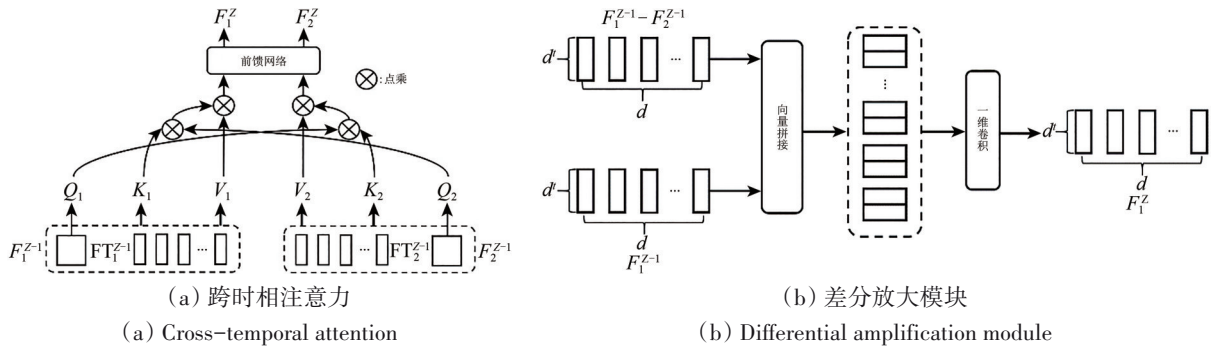


图 4 跨时相注意力与差分放大模块计算示意图

Fig. 4 The process of cross-temporal attention and difference amplification module

CTA 以光信 Token 作为引导, 将不同种类图像的光信 Token 与图像块 Token 交互计算以实现双时相序列特征之间的变化分析。前馈网络 (FF) 同样被应用于 CTA 计算后, 跨时相注意力计算过程如下:

$$\begin{cases} FT_1^z = FT_1^{z-1} + \text{CTA}(\text{LN}(F_2^{z-1})) \\ FT_2^z = FT_2^{z-1} + \text{CTA}(\text{LN}(F_1^{z-1})) \\ F_1^z = \text{FF}(\text{Concat}(FT_1^z \| F_2^0)) \\ F_2^z = \text{FF}(\text{Concat}(FT_2^z \| F_1^0)) \end{cases} \quad (8)$$

式中, z 为编码器层数, LN 为层归一化。

2.3 差分放大模块

为了进一步提高网络特征间差分信息的提取能力, 本文在每一层跨时相 Transformer 编码器后加入了差分放大模块 DAM (Differential Amplification Module) 来增强序列特征中的差分信息。DAM 可以实现低维序列特征与差分信息自适应融合, 在降低关键信息损失的同时提高变化检测性能。DAM 模块结构如图 4 (b) 所示, 其计算过程可表述为

$$\omega_2 \begin{bmatrix} (F_1^{z-1} - F_2^{z-1}) \\ F_1^{z-1} \end{bmatrix} \rightarrow F_1^z \quad (9)$$

式中， ω_2 为可学习的网络参数。

DAM 模块建立了大小为 1×2 的一维卷积核来实现差分信息与序列特征的自适应融合，其可以有效地保留并增强变化信息以增强网络性能。如图 2 所示，检测网络通过差分放大模块来获取最终的差分 Token，并且多层感知机对差分 Token 进行预测并由 Sigmoid 激活函数输出最终变化检测结果 $\mathbf{y} \in \mathbb{R}^{1 \times 2}$ ，以上过程可表述为

$$\text{DT} = \omega_3 \begin{bmatrix} \text{FT}_1^z \\ \text{FT}_2^z \end{bmatrix} \quad (10)$$

$$\mathbf{y} = \text{Sigmoid}(\omega_4 \text{LN}(\text{DT})) \quad (11)$$

式中， ω_3 ， ω_4 为对应的可学习网络参数，DT 为最终生成的差分 Token。 $\mathbf{y} \in \mathbb{R}^{1 \times 2}$ 是对应像素点变化或未变化的概率，若 $\mathbf{y}(0, 1) < \mathbf{y}(0, 0)$ ，则为未

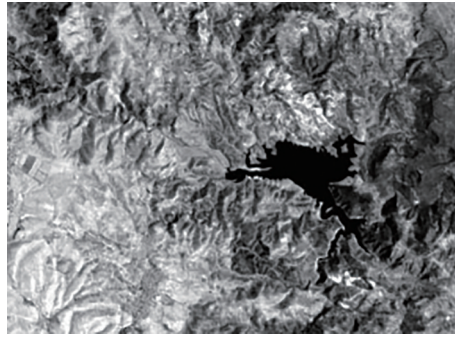
变化，否则为变化。

3 实验方法与结果分析

3.1 数据集

实验选用 3 组双时相异源遥感图像数据集与一组同源高光谱图像进行训练、验证以及测试。其中，训练样本用于训练网络模型，验证样本用于选择最优超参数，而测试样本用于检验网络有效性。

第 1 组数据集为“Sardinia”，其双时图像如图 5 (a) 和图 5 (b) 所示。时相 1 的近红外图像于 1995 年 9 月由 Landsat5 获取，时相 2 彩色图像为 1996 年 7 月的 Google earth 影像，所选区域均为意大利撒丁岛的某一区域，图像大小为 300×412 像素。双时图像空间分辨率一致为 30 m/pixel。



(a) 时相 1

(a) Temporal one



(b) 时相 2

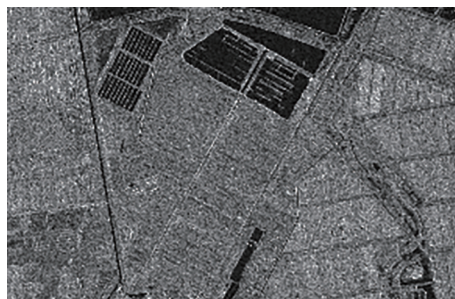
(b) Temporal two

图 5 Sardinia 数据集中时相 1 和时相 2 异源图像

Fig. 5 Temporal one and temporal two heterogeneous images of the Sardinia dataset

第 2 个数据集“Shuguang”，取自英格兰格洛斯特郡的某一农田区域。双时图像如图 6 (a) 和图 6 (b) 所示，时相一图像由 Radarsat-2 于 2008 年

6 月拍摄，时相 2 图像为 2012 年 9 月的 Google Earth 影像，图像大小为 593×921 像素。双时图像空间分辨率一致为 8 m/pixel。



(a) 时相 1

(a) Temporal one



(b) 时相 2

(b) Temporal two

图 6 Shuguang 时相 1 和时相 2 异源图像

Fig. 6 Temporal one and temporal two heterogeneous images of the Shuguang dataset

第3个数据集“Bastrop”取自美国加州巴斯特罗普的某一森林区域。多光谱双时伪彩色合成图如图7(a)和图7(b)所示。可知：时相1图像由Landsat5拍摄于2011年8月，含7个波段，光谱范围包涵0.45—2.35 μm 与10.40—12.50 μm 。时相2

图像由Earth Observing mission (EO-1 ALI) 于2011年9月获取，含10个波段，光谱范围为0.4—2.4 μm 。该数据集图像大小为1534×800像素。双时图像空间分辨率一致为30 m/pixel。

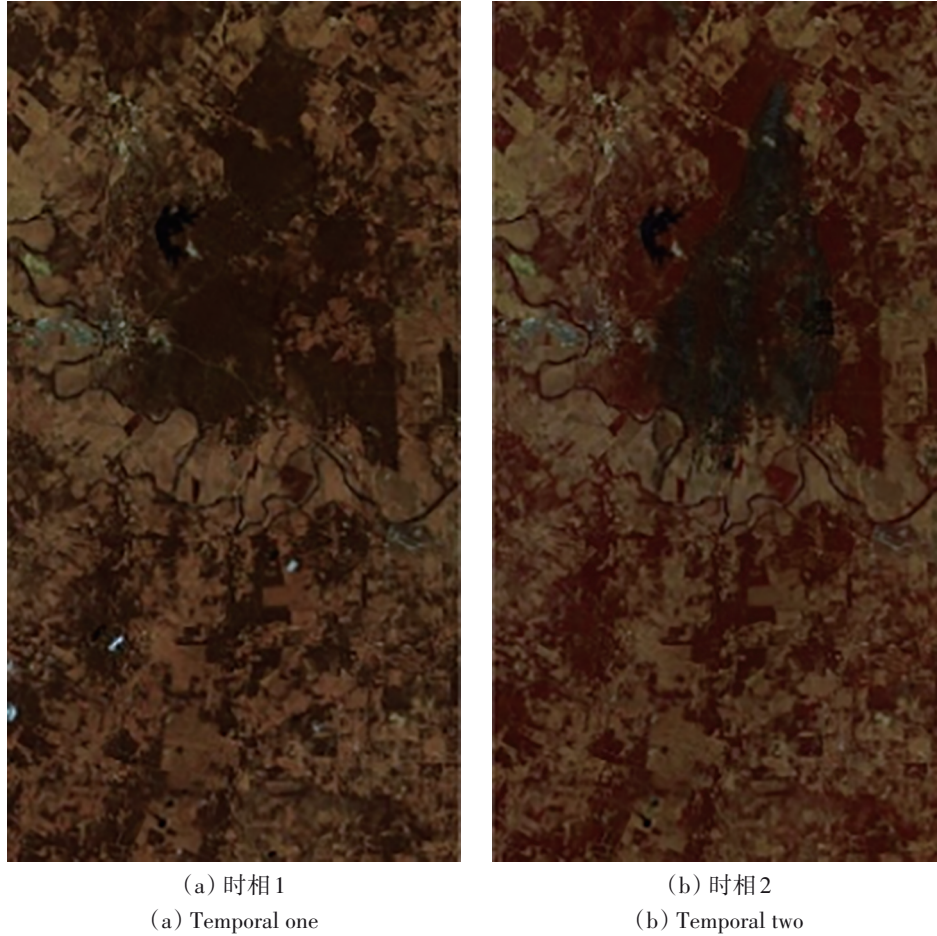


图7 Bastrop时相1和时相2异源图像
Fig. 7 Temporal one and temporal two heterogeneous images of the Bastrop dataset

第4个同源高光谱数据集“Farmland”取自中国江苏省盐城市附近的一片农田区域。高光谱双时伪彩色合成图如图8(a)和图8(b)所示，由EO-1高光谱传感器分别于拍摄于2006年5月与2007年4月，光谱范围为0.4—2.5 μm ，去除部分噪声后含155波段。该数据集图像大小为420×140像素。双时图像空间分辨率一致为30 m/pixel。

3.2 评价指标

实验选取的客观评价指标有：整体准确率(OA)、Kappa系数以及F1-Score(F1)，其计算公式为

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (12)$$

$$P_c = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + FP + TN + FN)^2} \quad (13)$$

$$Kappa = \frac{OA - P_c}{1 - P_c} \quad (14)$$

$$Re = \frac{TP}{TP + FN} \quad (15)$$

$$Pre = \frac{TP}{TP + FP} \quad (16)$$

$$F1 = 2 \times \frac{Pre \times Re}{Pre + Re} \quad (17)$$

式中，TP为正确检测到的变化像素个数，TN为正确检测到的未变化像素个数，FP为错误检测的变化像素个数，FN是错误检测的未变化像素个数。此外，Pre和Re是计算F1的相关参数。

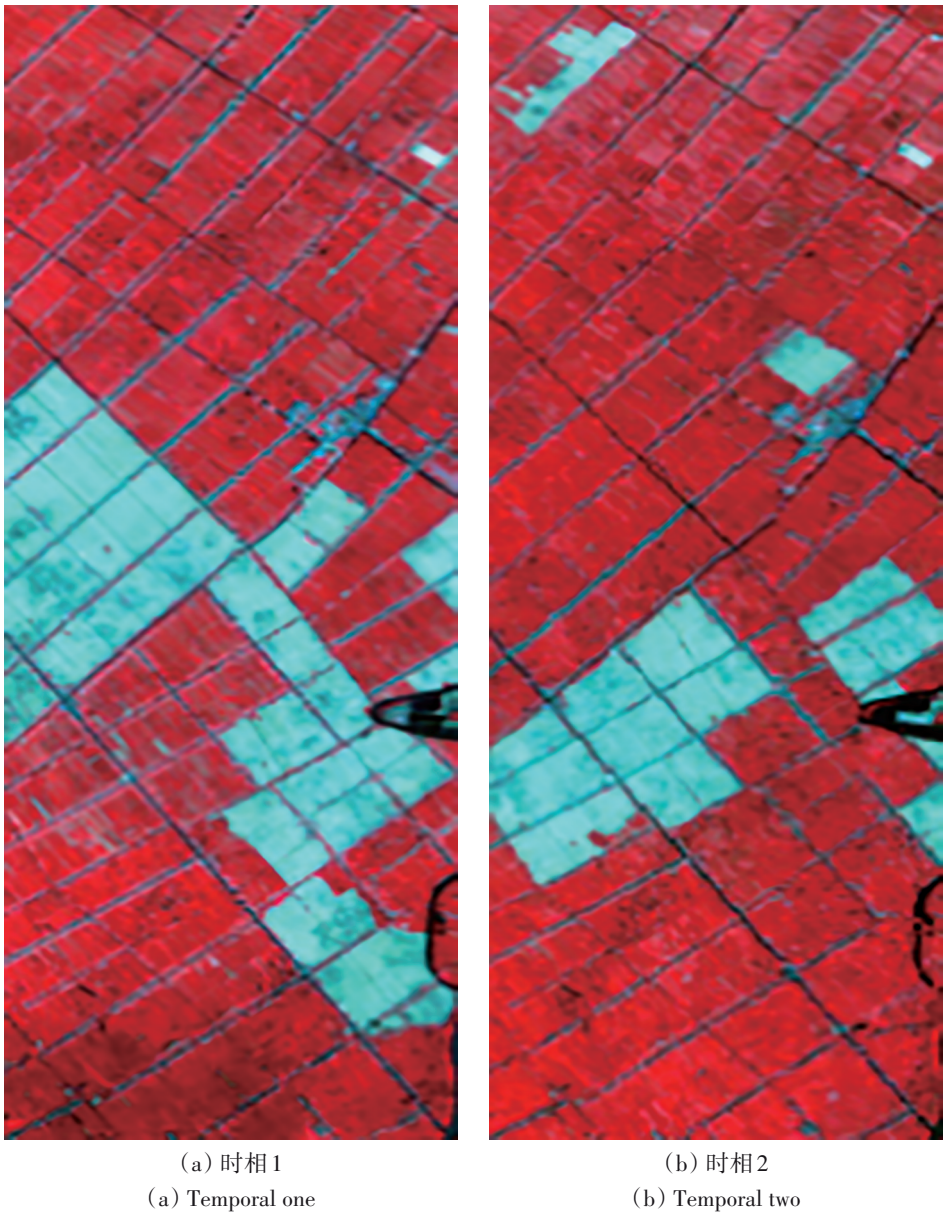


图 8 Farmland 时相 1 和时相 2 同源高光谱图像
Fig. 8 Temporal one and temporal two homogeneous hyperspectral images of the Farmland dataset

3.3 实验设置

为了验证本文提出网络的有效性，本研究将其与 6 种主流变化检测方法进行比较。其中包括基于有传统图形学的 GSP-VDF (Sun 等, 2022a) 与无监督学习的 SCASC (Sun 等, 2022b)，以及其他

4 种有监督学习的 SVM (linear)、ViT-Channel、ViT-Spatial 以及 SST-Former (Wang, 2022a)。为了保持公平性，实验各数据集都采用了相同数量与大小的图像块作为训练样本。具体如表 1 所示。

表 1 不同数据集选用的训练样本、验证样本以及测试样本的像素对数

Table 1 Numbers of pixel pairs in training samples and testing samples for the three datasets

数据集	训练样本			验证样本			测试样本		
	未变化样本	变化样本	总计样本	未变化样本	变化样本	总计样本	未变化样本	变化样本	总计样本
Sardinia	5798	381	6179	5798	381	6179	104378	6864	111242
Shuguang	26052	1254	27306	26052	1254	27306	468950	22591	491541
Bastrop	5538	659	6197	5538	659	6197	1096527	130551	1227078
Farmland	447	182	629	447	182	629	43829	17913	61742

由于训练样本数量与图像块输入大小对于网络模型精度有较大的影响,本文在Sardinia与Shuguang数据集中随机选用双时相图像中5%图像块作为训练样本,图像块输入大小为9。Bastrop数据集随机选用双时相图像中0.5%的图像块作为训练样本,图像块输入大小为5。在同源高光谱数据集Farmland

中选取了1%的图像块作为训练样本,图像块输入大小为7。对比实验设置如下所述。

GSP-VDF是一种基于图形信号处理的异源变化检测方法。该方法通过顶点域滤波和谱域分析来检测双重时间的异源图像变化。部分关键参数设置如表2所示,与GSP-VDF中描述相同。

表2 对比实验中各方法的详细参数设置

Table 2 The detailed parameter settings in comparison experiment

GSP-VDF	SCASC	SVM	ViT-Channel	ViT-Spatial	SST-Former
多项式阶数 $M=3$, 截止频率 $\gamma_{cr}=0.9$	选用k-adaptive策略	主成分数量选择为2, 选用线性SVM来实现 二分类	通道编码器数量为5, 特征维度为64	空间编码器数量为5, 特征维度为128	光谱编码器数量为2,特征 维度为25。空间编码器数量 为4,特征维度为256

SCASC是一种基于结构一致性的无监督异源变化检测方法。一方面,这种基于结构一致性的自适应图像回归模型可以减轻噪声的影响,另一方面此方法可以使用变化特征的先验知识来区分差分图像中的变化和不变。部分关键参数设置如表2所示,参考了SCASC中参数实验的最优参数。

在SVM的比较实验中,先使用主成分分析法提取差异图的主成分,然后用SVM进行分类,以检测变化与未变化的像素。表2中给出了本文采用的主成分数量以及二分类策略。

本文设计了一个双分支通道方向的ViT编码网络作为ViT-channel的架构,对ViT-channel的双时相编码结果进行拼接,以预测最终结果。类似于ViT-channel,ViT-spatial从空间方向对双时相输入进行编码,然后对编码结果进行拼接和预测,以实现变化检测。

在基于ViT的变化检测方法中,SST-Former可以提取图像中联合空间-光谱-时间序列信息,并且在同源高光谱变化检测中达到了优异的变化检测性能。因此,本文将其移用于异源变化检测任务中。表2给出了部分参数设置,参考了SST-Former参数实验中给出的最优参数。

本文方法在PyTorch1.10.1框架下实现,配备Intel Xeon Silver 4210R CPU和NVIDIA GeForce RTX 3090 24 G GPU。优化器选用Adam,学习率设定为0.0005。在每轮次后,当前轮次学习率将乘以0.997以更新下一轮次学习率。设定的总轮次为400轮,其他超参数与参数实验中讨论的相同。

表3对本文方法进行了从整体准确率(OA)、

Kappa系数等多个客观评价指标的定量分析。在3个异源变化检测数据集中,与传统的图形学检测方法、无监督学习方法以及其他几种深度学习主流检测网络相比,本文提出的网络模型在有限的训练样本条件下达到了最优效果。由于3个异源数据集中存在变化样本与未变化样本分布不均的问题,实验结果在OA指标上提升并不明显。但在Kappa指标上,本文相较于其他方法都有较大提升。在Sardinia数据集上,本文方法取得的Kappa相较于第二名高出了约1%。而在Shuguang与Bastrop数据集上,本文Kappa指标高出第二名2%到3%。同时,本文在F1 score上也表现出色,相比于其他几种主流变化检测方法,本文方法都达到了最优。在同源高光谱数据集中,SST-Former在各项指标上达到最优。本文方法相较于SST-Former在Kappa指标上低了0.5%,但仍然取得了较优的变化检测性能。这也从侧面表明本文方法可用于对多源遥感图像数据进行建模并达到较优的变化检测性能。除以上准确率相关指标外,不同方法的测试所用时间详见表3。

图9—图12展示了4个数据集上不同检测方法的比较实验可视化结果。图中白色部分表示变化区域,黑色部分表示未变化区域,红色部分表示误检区域,而绿色部分表示漏检区域。

在Sardinia数据集实验中,本文的方法表现出了比其他方法更加准确的检测结果,尤其是在容易漏检的微小变化区域方面。然而,在部分噪声区域仍然存在小部分误检,这也是大多数模型都存在的问题。

表3 客观评价指标结果											
Table 3 Quantitative evaluation results											
数据集	对比方法	OA/%	Kappa/%	F1/%	时间/s	数据集	对比方法	OA/%	Kappa/%	F1/%	时间/s
Sardinia	GSP-VDF	97.14	75.61	77.14	20.44	Shuguang	GSP-VDF	98.51	80.87	81.64	28.17
	SCASC	95.31	60.71	63.24	2.51		SCASC	97.34	64.59	65.92	2.86
	SVM	97.40	76.44	87.71	4.97		SVM	96.94	53.97	78.89	147.11
	ViT-channel	98.57	83.39	91.70	27.19		ViT-channel	98.57	83.39	91.70	279.13
	ViT-spatial	98.48	86.17	93.26	24.51		ViT-spatial	97.65	75.59	87.78	301.51
	SST-Former	98.49	86.41	93.20	43.46		SST-Former	98.72	85.51	92.86	315.14
	本研究提出方法	98.58	87.48	93.74	34.89		本研究提出方法	98.85	86.29	93.14	427.14
Bastrop	GSP-VDF	92.51	63.67	67.82	28.19	Farmland	GSP-VDF	90.62	78.01	84.81	53.01
	SCASC	91.81	45.88	49.73	3.85		SCASC	85.61	62.02	71.47	9.82
	SVM	98.51	91.82	95.84	296.06		SVM	91.79	80.97	88.43	11.42
	ViT-channel	98.68	93.17	97.12	283.03		ViT-channel	94.10	87.31	92.04	39.17
	ViT-spatial	98.54	92.81	96.73	261.12		ViT-spatial	92.61	83.51	90.87	31.50
	SST-Former	98.86	94.09	97.31	449.98		SST-Former	97.25	93.39	96.69	58.29
	本研究提出方法	99.25	95.99	98.47	327.82		本研究提出方法	97.03	92.86	96.07	40.52

注:表中粗体表示最优结果。

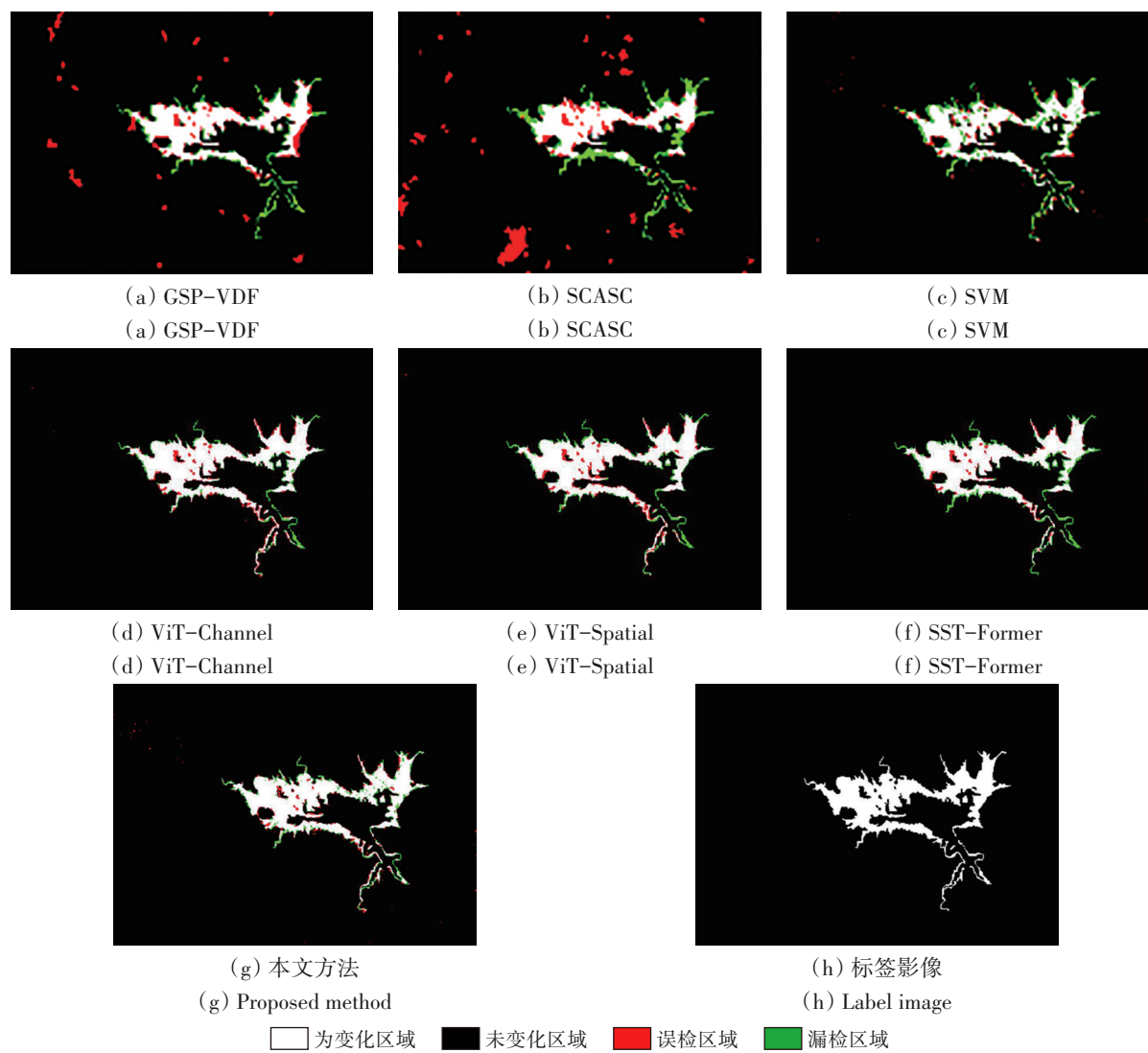


图9 Sardinia数据集中GSP-VDF、SCASC、SVM、ViT-Channel、ViT-Spatial、SST-Former、本文方法及标签影像的对比实验可视化结果

Fig. 9 Comparison experiment visual results of GSP-VDF, SCASC, SVM, ViT-Channel, ViT-Spatial, SST-Former, method in this study and label image for the Sardinia dataset

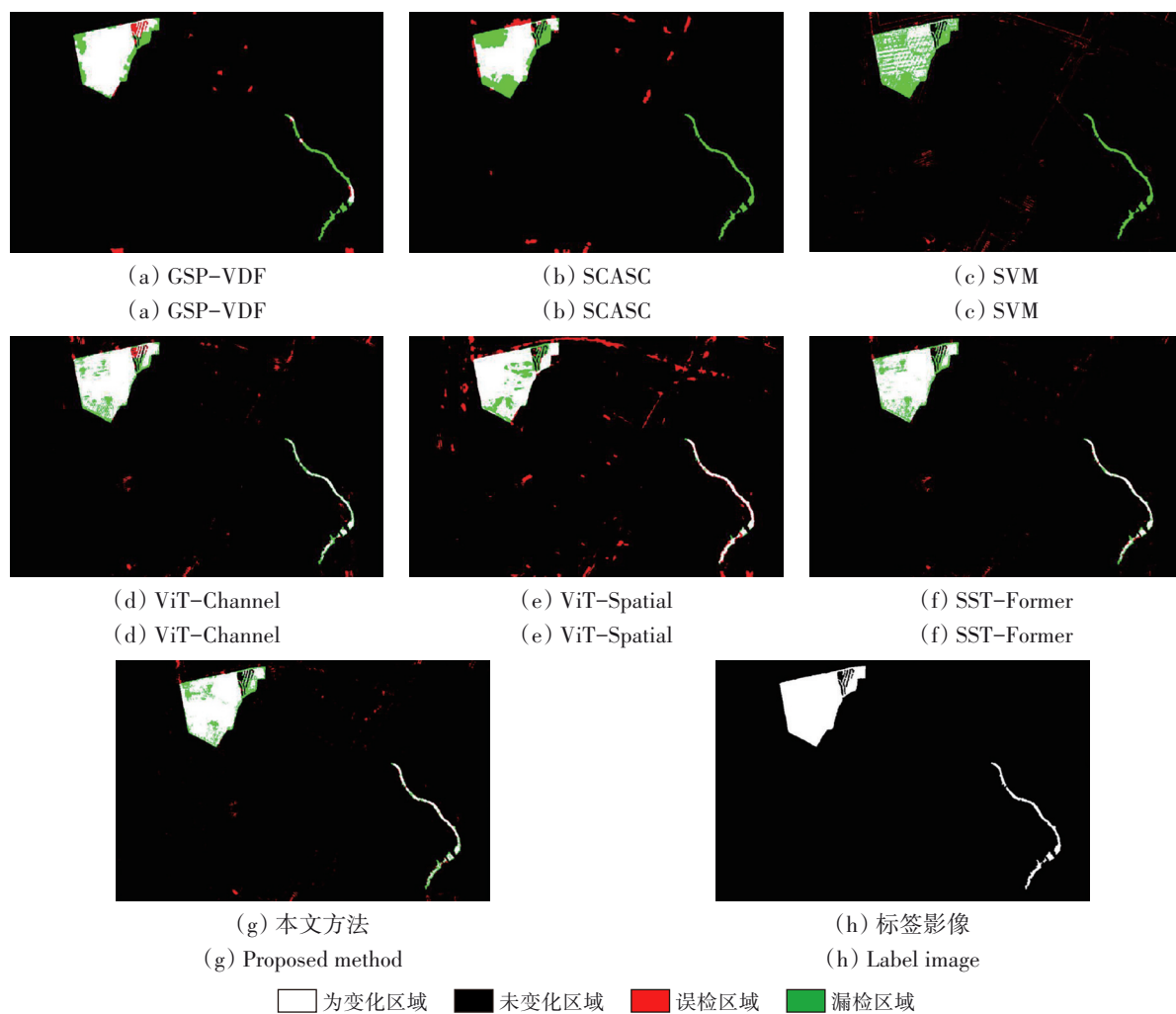
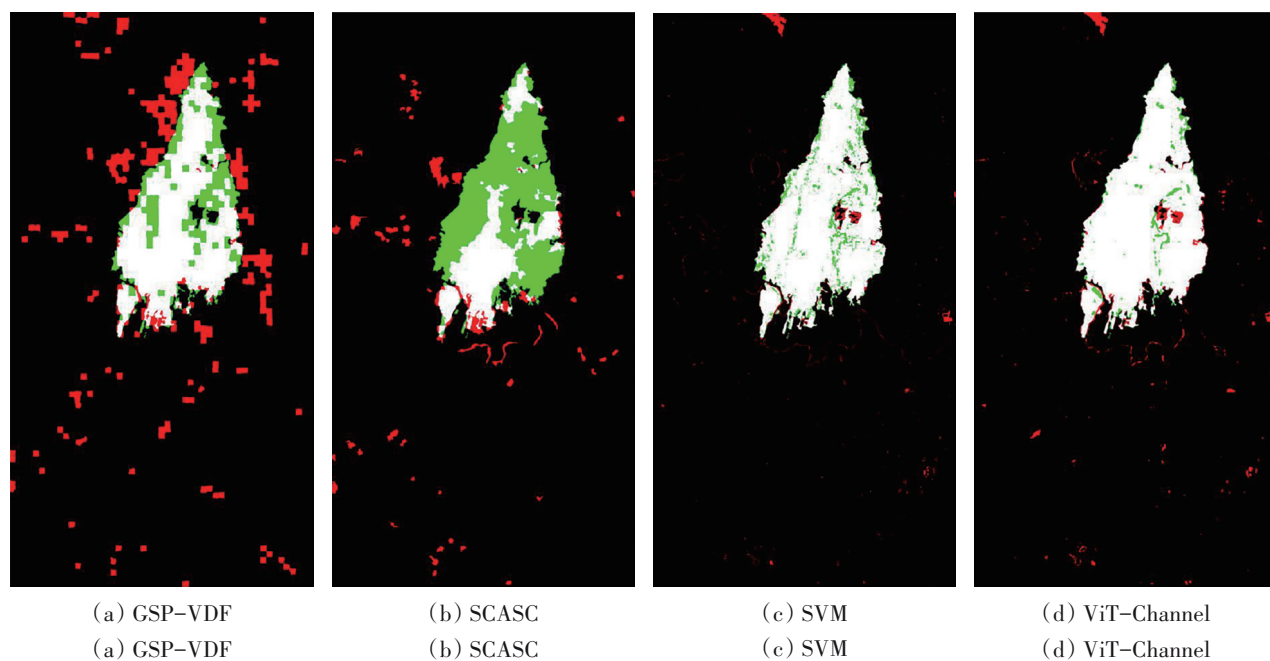


图 10 Shuguang 数据集中 GSP-VDF、SCASC、SVM、ViT-Channel、ViT-Spatial、SST-Former、本文方法和标签影像的对比实验可视化结果

Fig. 10 Comparison experiment visual results of GSP-VDF, SCASC, SVM, ViT-Channel, ViT-Spatial, SST-Former, method in this study and label image for the Shuguang dataset



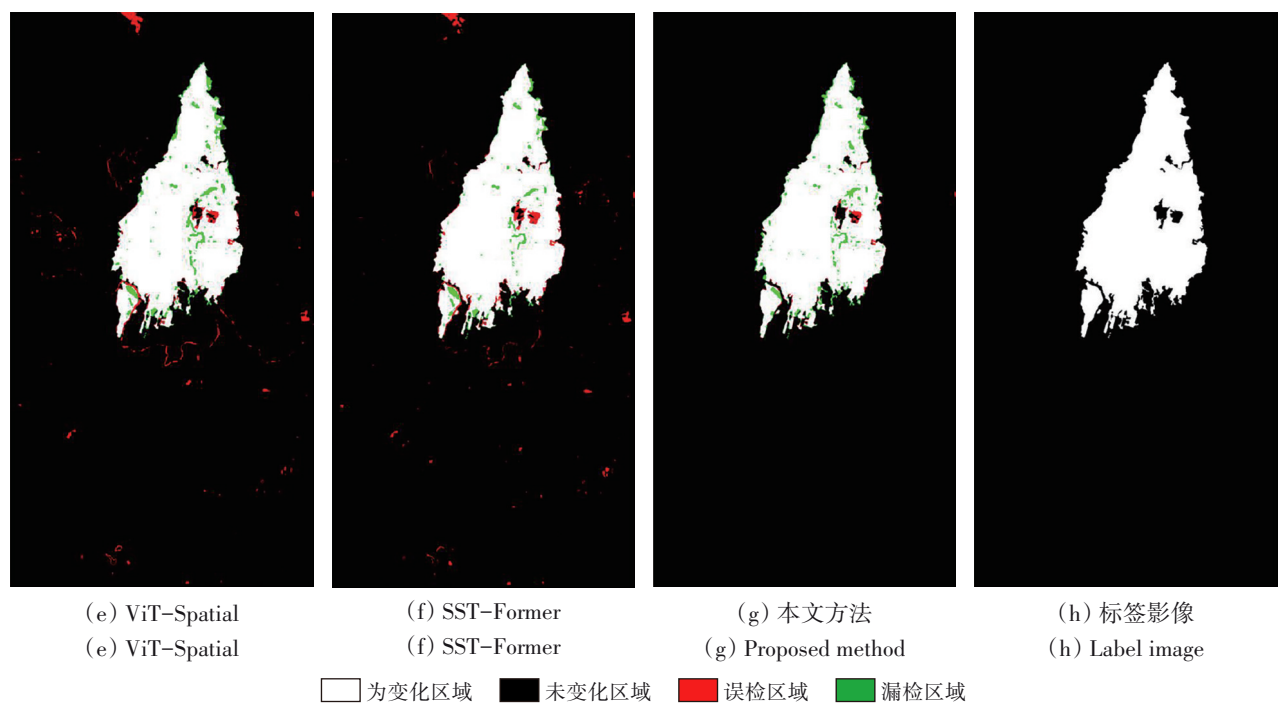
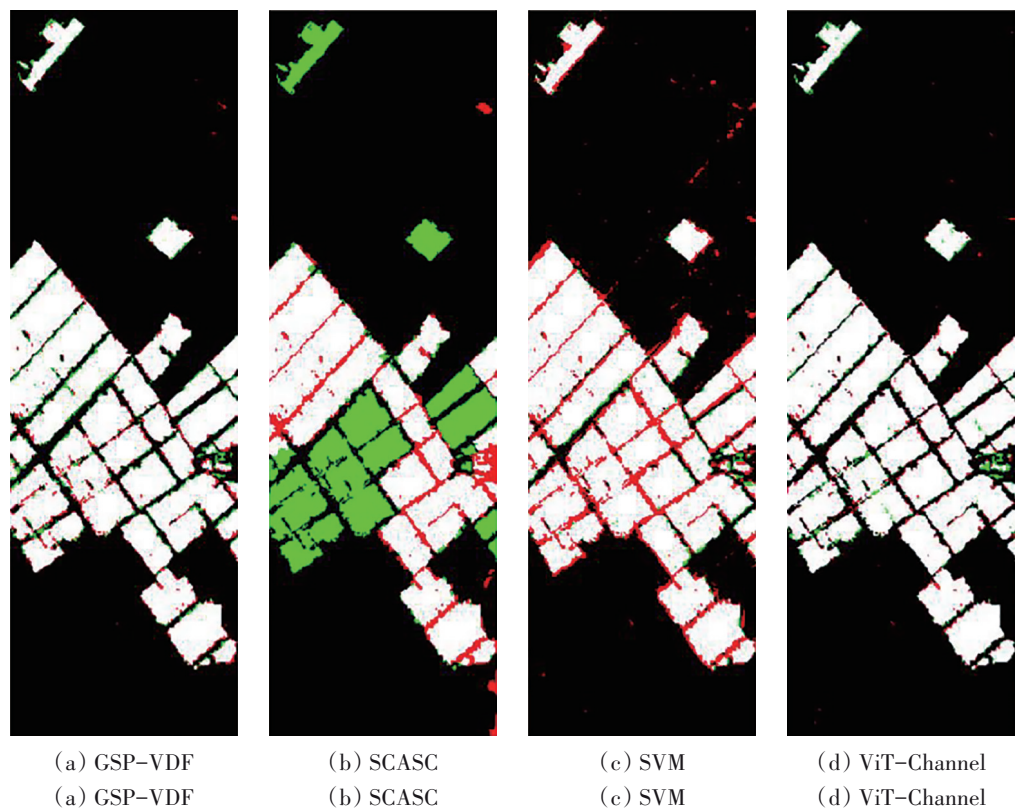


图 11 Bastrop 数据集中 GSP-VDF、SCASC、SVM、ViT-Channel、ViT-Spatial、SST-Former、本文方法、标签影像的对比实验可视化结果

Fig. 11 Comparison experiment visual results of GSP-VDF, SCASC, SVM, ViT-Channel, ViT-Spatial, SST-Former, method in this study and label image for the Bastrop dataset



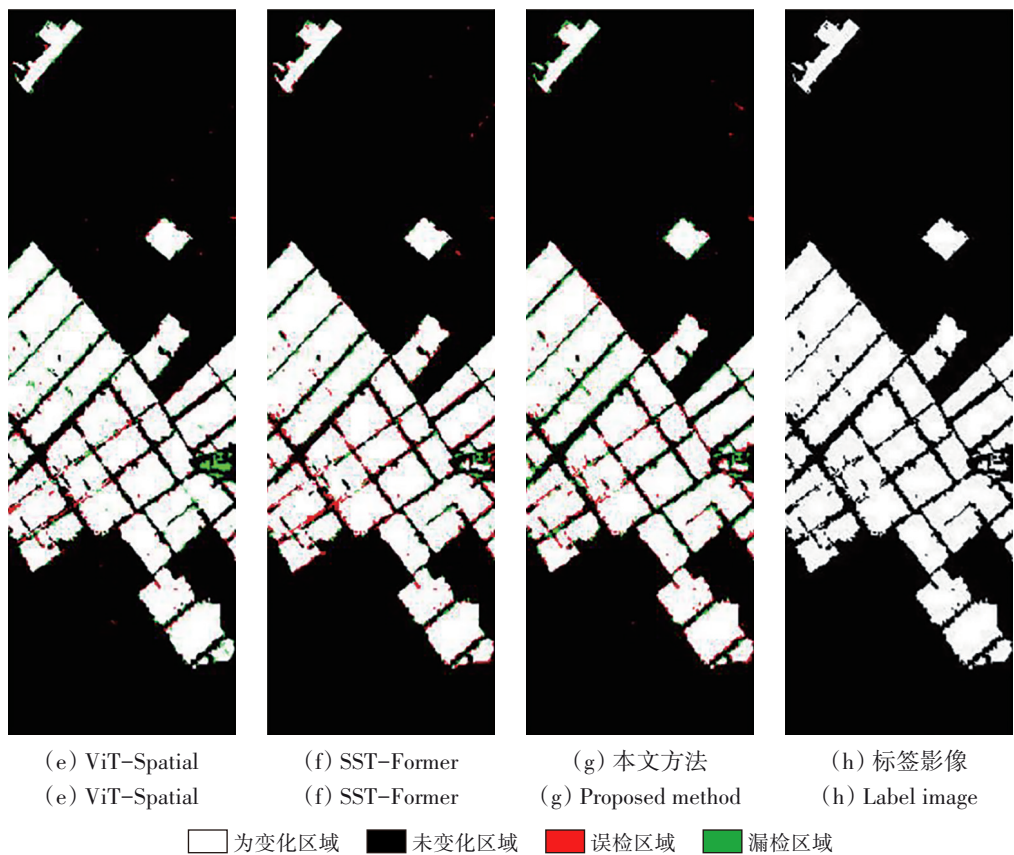


图 12 Farmland 数据集中 GSP-VDF、SCASC、SVM、ViT-Channel、ViT-Spatial、SST-Former、本文方法、标签影像的对比实验可视化结果

Fig. 12 Comparison experiment visual results of GSP-VDF, SCASC, SVM, ViT-Channel, ViT-Spatial, SST-Former, method in this study and label image for the Farmland dataset

在 Shuguang 主观实验中，本文方法表现与 SST-Former 相近，总体上优于其他方法，尤其是受到噪声区域影响相对较小。

在 Bastrop 数据集中，本文方法展现了优异的主观视觉效果，相比于其他方法，几乎没有出现杂乱的噪声区域，直观呈现了变化区域，与标签影像相近。在同源高光谱数据上，本文可视化效果与 SST-Former 相似，但在部分小区域出现漏检和误检。

3.4 消融实验

为了验证本文提出的各个模块在网络中的有效性，针对 OT-Former、CT-Former 以及 DAM 模块在 Bastrop 数据集上进行了消融实验，实验结果如表 4 所示。

消融实验结果如表 4 所示，在不加入 CT-Former 的情况下，OT-Former 自适应生成的光信 Token 也可以实现优异的变化检测性能。而使用可训练的 Class token 加入 CT-Former 并直接进行变化检测的

性能并不理想，在 OA 指标上相较于本文方法降低了接近 5%。在去除 DAM 的实验中，结果显示加入 DAM 可以在 Kappa 指标上提升近 0.5%，这也证实了 DAM 对变化检测性能确有提升。

表 4 Bastrop 数据集消融实验		
Table 4 Ablation experiment on Bastrop dataset		
		/%
方法	OA	Kappa
OT-Former	98.54	92.45
CT-Former	94.42	91.20
Removed DAM	99.15	95.49
本文所提方法	99.25	95.99

3.5 参数实验

为了验证不同大小的训练样本、编码器数量、特征维度、图像块输入大小以及批处理大小等参数对变化检测精度的影响，针对所提出网络结构进行了消融实验对比，在 4 个数据集上进行的多参数实验结果如图 13 所示。

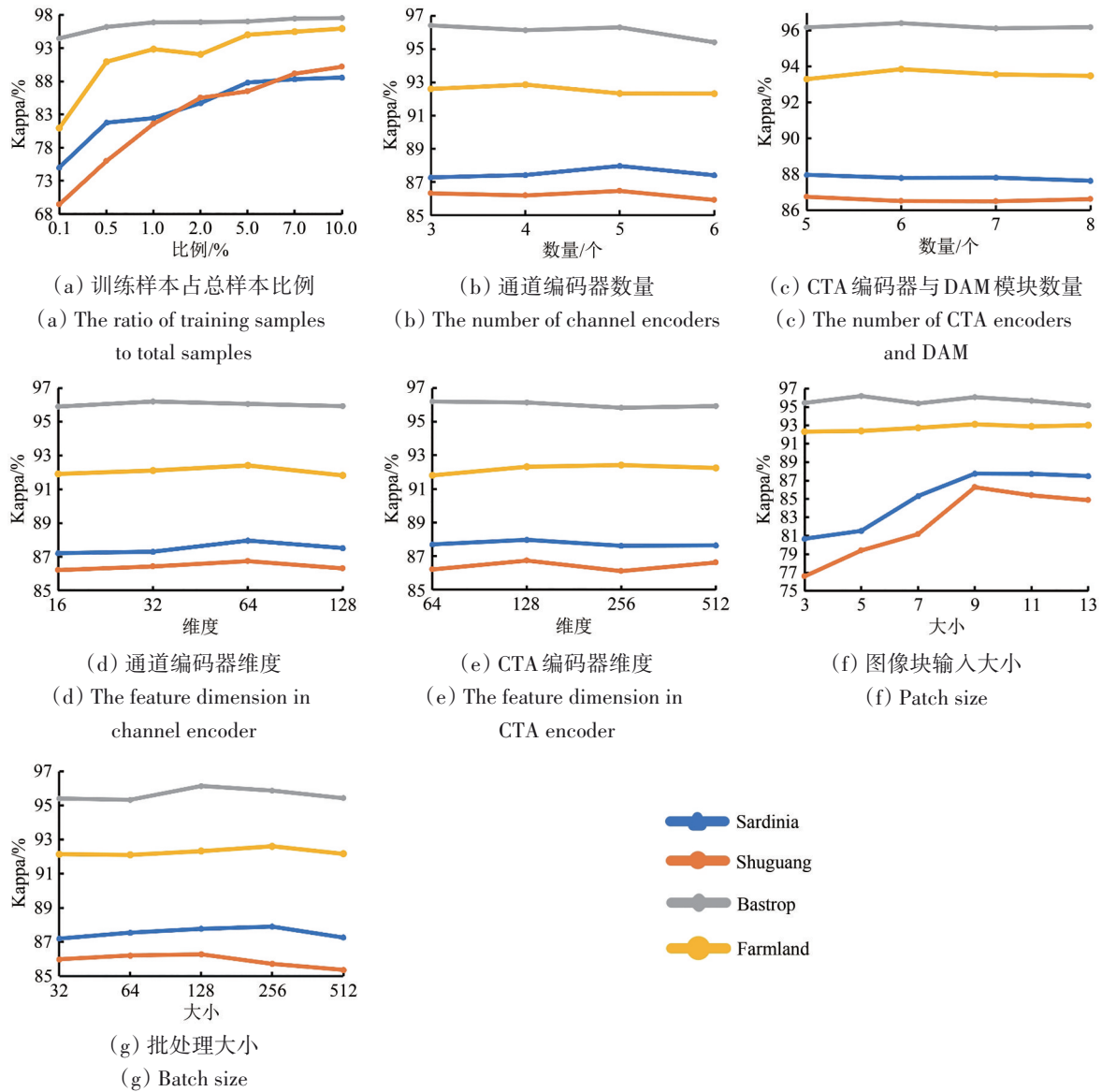


图 13 不同参数配置在各数据集上的 Kappa 系数检测结果

Fig. 13 Kappa coefficient across different parameter configurations on various datasets

参数实验结果显示, 不同批处理大小和特征维度对本文方法的检测精度影响较小。相反, 不同大小的训练样本、图像块输入大小和网络深度对检测精度影响较大。如图 13 (a) 所示, 输入训练样本比重越高, 本文方法达到的 Kappa 指标也越高。

图 13 (b) 和图 13 (c) 展示了在不同通道编码器、CTA 编码器和 DAM 模块数量下对检测精度的影响。其中, 编码器与各类模块的数量与网络深度直接相关, 不同深度的网络结构会对模型精度产生不同程度的影响。一般来说, 适当的网络结构更有利于检测。参数实验结果表明, 不同网络深度会导致本文方法的检测精度产生一定波动, 最高可达 1.2%。

此外, 图 13 (f) 还揭示了不同大小的图像块输入对模型精度的影响。由于遥感图像包含多种地物类型, 不同大小的图像块输入对网络精度产生不同程度的影响。总体而言, 图像块输入大小越大, 包含的地物信息也越多。但需要注意的是, 在不同类型的图像输入中, 适当的图像块输入大小更有利于检测, 而不一定是输入越大越好。

4 结 论

由于异源遥感图像本身成像机理的差异, 为对多时相异源图像进行多类型地物区域精准变化检测, 本文将异源图像从通道方向上序列化为信号序列的形式输入检测网络, 并基于 Transformer

模型提出了一种新的异源变化检测网络架构, 主要结论如下:

(1) 针对于异源多时相遥感数据在成像类型和数据表征形式上存在的显著差异, 本研究所提方法能自适应地对不同种类异源图像的光学信号特征在通道方向进行有效建模并表征;

(2) 为了实现异源多时相图像间高效的变化分析, 本研究提出了一种交互学习的方式, 旨在深入挖掘和学习序列特征中的潜在变化规律。此外, 在交互学习过程中, 本研究融入了差分放大模块来增强网络对于关键信息的捕获能力, 进一步优化变化检测的细节和准确性;

(3) 为验证所提方法有效性, 本研究在Sardinia、Shuguang以及Bastrop 3类异源变化检测数据集与同源高光谱变化检测数据集Farmland上进行了实验验证。实验结果表明, 与其他主流变化检测方法相比, 本文提出的方法获得了较高的变化检测性能。

此外, 本研究方法已在现有公开数据集上展现出了优异的变化检测性能, 但其仍存在一些未解决问题, 针对未解决问题的讨论以及未来研究的解决方案建议如下:

(1) 仍需要进行手动数据标注。对双时相异源图像进行变化区域标注的工作量十分巨大。虽然本文方法在多类遥感图像上表现良好, 但仍需要有经验的研究人员进行手动标注。为了解决这个问题, 无监督学习和自监督学习是当前的研究热点。在未来工作中, 应研究新的训练策略, 使用未标记或少量标记的数据集来训练网络;

(2) 提高模型可解释性。本文提出的方法在多个异源数据集上表现良好。然而, 对于不同信号的建模表征仍是一个复杂的数学过程。尽管深度学习方法在特征提取方面具有强大的能力, 但其可解释性却相对较弱。未来, 本研究将从各种异源遥感成像机理出发, 借鉴深度学习技术的优点, 研究更优且更具可解释性的检测方法;

(3) 对超参数较为敏感。为了达到较优的检测结果, 本研究在不同种类的异源数据集中设定了不同的超参数。然而, 本研究在参数实验过程中发现本文方法的一些超参数的选取会对检测结果产生较大的影响。例如, 通道编码器特征维度 d 的不同设置会使得检测精度Kappa系数波动约1%左右。因此, 今后研究将参考更多的先进深度学

习方法, 研究网络结构和参数设置, 以实现检测方法更优、更稳定。

参考文献(References)

- Adão T, Hruška J, Pádua L, Bessa J, Peres E, Morais R and Sousa J J. 2017. Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sensing*, 9(11): 1110 [DOI: 10.3390/rs9111110]
- Bandara W G C and Patel V M. 2022. A transformer-based Siamese network for change detection//IEEE International Geoscience and Remote Sensing Symposium. Kuala Lumpur: IEEE: 207-210 [DOI: 10.1109/IGARSS46834.2022.9883686]
- Bengio Y, Simard P and Frasconi P. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2): 157-166 [DOI: 10.1109/72.279181]
- Bovolo F and Bruzzone L. 2007. A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. *IEEE Transactions on Geoscience and Remote Sensing*, 45(1): 218-236 [DOI: 10.1109/TGRS.2006.885408]
- Bovolo F, Marchesi S and Bruzzone L. 2012. A framework for automatic and unsupervised detection of multiple changes in multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(6): 2196-2212 [DOI: 10.1109/TGRS.2011.2171493]
- Canty M J and Nielsen A A. 2008. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. *Remote Sensing of Environment*, 112(3): 1025-1036 [DOI: 10.1016/j.rse.2007.07.013]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Housley N. 2020. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint 2010.11929* [DOI: 10.48550/arXiv.2010.11929]
- Hasanlou M and Seydi S T. 2018. Hyperspectral change detection: an experimental comparative study. *International Journal of Remote Sensing*, 39(20): 7029-7083 [DOI: 10.1080/01431161.2018.1466079]
- Hong D F, Han Z, Yao J, Gao L R, Zhang B, Plaza A and Chanussot J. 2022. SpectralFormer: rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-15 [DOI: 10.1109/TGRS.2021.3130716]
- Hou Z F, Li W and Du Q. 2021. A patch tensor-based change detection method for hyperspectral images//2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Brussels: IEEE: 4328-4331 [DOI: 10.1109/IGARSS47720.2021.9554630]
- 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]

- Lyu H B, Lu H and Mou L C. 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. *Remote Sensing*, 8(6): 506 [DOI: 10.3390/rs8060506]
- Ma W P, Yang H, Wu Y, Xiong Y T, Hu T, Jiao L C and Hou B. 2019. Change detection based on multi-grained cascade forest and multi-scale fusion for SAR images. *Remote Sensing*, 11(2): 142 [DOI: 10.3390/rs11020142]
- Mandal M and Vipparthi S K. 2022. An empirical review of deep learning frameworks for change detection: model design, experimental frameworks, challenges and research needs. *IEEE Transactions on Intelligent Transportation Systems*, 23(7): 6101-6122 [DOI: 10.1109/TITS.2021.3077883]
- Pirrone D, Bovolo F and Bruzzone L. 2020. A novel framework based on polarimetric change vectors for unsupervised multiclass change detection in dual-pol intensity SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7): 4780-4795 [DOI: 10.1109/TGRS.2020.2966865]
- Qu J H, Xu Y S, Dong W Q, Li Y S and Du Q. 2022. Dual-branch difference amplification graph convolutional network for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-12 [DOI: 10.1109/TGRS.2021.3135567]
- Sun W W, Yang G, Peng J T and Du Q. 2020. Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(1): 107-111 [DOI: 10.1109/LGRS.2019.2915315]
- Sun Y L, Lei L, Guan D D, Kuang G Y and Liu L. 2022a. Graph signal processing for heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-23 [DOI: 10.1109/TGRS.2022.3221489]
- Sun Y L, Lei L, Guan D D, Li M and Kuang G Y. 2022b. Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-14 [DOI: 10.1109/TGRS.2021.3110998]
- Wang Y H, Hong D F, Sha J J, Gao L R, Liu L, Zhang Y G and Rong X H. 2022a. Spectral-spatial-temporal transformers for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-14 [DOI: 10.1109/TGRS.2022.3203075]
- Wang Y K, Chen X H, Gao L L, Huang W B, Sun F C and Wang Y H. 2022b. Multimodal token fusion for vision transformers//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12176-12185 [DOI: 10.1109/CVPR52688.2022.01187]
- 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]
- 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 X F, Ye Y M, Li X T, Lau R Y K, Zhang X F and Huang X H. 2018. Hyperspectral image classification with deep learning models. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9): 5408-5423 [DOI: 10.1109/TGRS.2018.2815613]
- Zhan Y, Fu K, Yan M L, Sun X, Wang H Q and Qiu X S. 2017. Change detection based on deep Siamese convolutional network for optical aerial images. *IEEE Geoscience and Remote Sensing Letters*, 14(10): 1845-1849 [DOI: 10.1109/LGRS.2017.2738149]

Optical-signal token guided change detection network for heterogeneous remote sensing image

LIU Qinsen¹, SUN Bangyong^{1,2}

1. Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710054, China;

2. State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China

Abstract: Change Detection (CD) is a vital technique for identifying and analyzing changes over time in a specific area using optical signals from remote sensing images. This technique has been extensively utilized in various fields, including national defense security, environmental monitoring, and urban construction. However, some challenges in achieving accurate and reliable CD are still encountered due to inherent disparities in imaging mechanisms, spectral ranges, and spatial resolutions among heterogeneous images. These challenges lead to issues such as inadequate accuracy, missed detections, and false detections. Heterogeneous remote sensing images can be regarded as sequences of different optical signals from the channel perspective. For example, RGB and infrared images can be regarded as sequences of spectral signals from different ranges. Transformers employ a multi-head attention mechanism that can effectively handle and analyze sequence information to achieve accurate heterogeneous CD. Thus, the paper proposes an optical signal token guided CD network for heterogeneous remote sensing images.

This paper presents a novel heterogeneous CD network, primarily comprising the optical-signal token transformer (OT-Former) and the

cross-temporal transformer (CT-Former). The proposed method demonstrates the capacity to effectively handle diverse remote sensing images of distinct categories and attain precise CD results. Specifically, OT-Former can encode diverse heterogeneous images in channel-wise for adaptively generating the optical-signal tokens. Meanwhile, CT-Former can use the optical-signal tokens as a guide to interact with the patch token for the learning of change rules. Moreover, a Difference Amplification Module (DAM) is embedded into the network to enhance the extraction of difference information. This module utilizes a 1×2 convolutional kernel to effectively fuse difference information. Finally, the differential token is predicted by multilayer perceptron to output the CD results.

Experiments were conducted on three heterogeneous datasets and one homogeneous dataset to evaluate the performance of the proposed method. Furthermore, the proposed method was compared with six typical CD methods and evaluated the performance using overall accuracy (OA), Kappa coefficient, and F1-score, among other evaluation metrics, to validate the effectiveness of the proposed network in this study. A limited number of samples were utilized for training during the experiment. Under identical experimental conditions, the proposed method demonstrated exceptional performance in homogeneous and heterogeneous CD. The results show that the proposed approach surpasses existing state-of-the-art methods in terms of qualitative and visual performance. Additionally, ablation experiments and parameter analyses were conducted to validate the effectiveness of the proposed methods, including the OT-Former, CT-Former, and DAM modules, and to assess the impact of various parameters within the network.

Overall, the current study presents a novel heterogeneous CD network based on the transformer framework. Within this network, OT-Former is proposed to achieve the adaptive generation of optical-signal tokens from diverse remote sensing images. Moreover, the CT-Former utilizes these optical-signal tokens as a guide to facilitate interaction with patch tokens for the learning of change rules. Additionally, DAM modules were embedded into the network to effectively extract the difference information. An extremely limited number of samples were utilized only for training in the experiments. Remarkably, the proposed method outperformed the existing state-of-the-art methods, achieving a significantly advanced performance in heterogeneous CD.

Key words: remote sensing, heterogeneous images, change detection, multimodal analysis, deep learning, transformer

Supported by National Natural Science Foundation of China (No. 62076199); The Key R&D project of Shaan'xi Province (No. 2022ZDLGY01-03); The Open Research Fund of State Key Laboratory of Transient Optics and Photonics (No. SKLST202214); Key scientific Research Program of Shan'xi Provincial Department of Education (No. 23JY063)