

# 基于边缘引导和动态可变形Transformer的遥感图像变化检测

雷涛<sup>1,2</sup>, 翟钰杰<sup>1,2</sup>, 许叶彤<sup>1,2</sup>, 王营博<sup>1,2</sup>, 公茂果<sup>3</sup>

(1. 陕西省人工智能联合实验室(陕西科技大学), 陕西西安 710021; 2. 陕西科技大学电子信息与人工智能学院, 陕西西安 710021; 3. 西安电子科技大学协同智能系统教育部重点实验室, 陕西西安 710071)

**摘要:** 卷积神经网络(Convolutional Neural Network, CNN)和Transformer的混合架构能够有效建模图像的局部与全局特征, 已成为遥感图像变化检测任务的主流网络. 然而这类网络仍面临着一些挑战. CNN分支中的卷积和池化运算通常会抑制遥感图像中的高频信息, 降低目标边界的精度; 此外, Transformer分支对图像像素进行等长程依赖关系建模, 忽略了变化目标的形状及语义关联信息, 导致网络对变化目标特征的表达不足. 为解决上述问题, 提出了基于边缘引导和动态可变形Transformer的遥感图像变化检测网络. 在CNN分支中设计了边缘信息引导模块, 利用高频信息增强目标区域的边缘信息, 从而改善变化目标的轮廓精度. 同时设计了一种新颖的动态可变形Transformer, 能够自适应地匹配形状不同的变化目标, 选择与变化相关的特征建模长程依赖关系, 以提高网络的特征表达能力. 实验结果表明, 提出的方法在三个公开数据集LEVIR-CD、CDD和DSIFN-CD上显著提高了检测精度, 在变化目标的边界精度和内部完整性方面都明显优于当前的主流网络.

**关键词:** 遥感图像; 变化检测; 高频信息; 边缘信息; 动态可变形Transformer

**基金项目:** 国家自然科学基金(No.62271296, No.62201334); 陕西省杰出青年科学基金(No.2021JC-47); 陕西省重点研发计划(No.2022GY-436, No.2021ZDLGY08-07); 陕西省自然科学基金基础研究计划(No.2022JQ-634, No.2022JQ-018); 陕西省创新能力支撑计划(No.2020SS-03)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2024)01-0107-11

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230583

## Edge Guided and Dynamically Deformable Transformer Network for Remote Sensing Images Change Detection

LEI Tao<sup>1,2</sup>, ZHAI Yu-jie<sup>1,2</sup>, XU Ye-tong<sup>1,2</sup>, WANG Ying-bo<sup>1,2</sup>, GONG Mao-guo<sup>3</sup>

(1. Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China;

2. School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China;

3. Key Laboratory of Collaborative intelligent Systems, Ministry of Education, Xidian University, Xi'an, Shaanxi 710071, China)

**Abstract:** The hybrid architecture of convolutional neural network (CNN) and Transformer can effectively model local and global features of images, and has emerged as the predominant choice for remote sensing images change detection tasks. Nevertheless, these networks still confront challenges. The convolution and pooling operations employed by the CNN branch typically suppress the high-frequency information of remote sensing images, resulting in decreased precision of object boundaries in change detection results. Additionally, the Transformer branch equivalently models long-range dependencies for all pixels in remote sensing images, thereby disregarding shape information and semantic associations of objects, which limits the network's feature representation ability on changed objects in remote sensing images. To address these challenges, a remote sensing images change detection network is proposed based on edge guidance and dynamic deformable Transformer. In the CNN branch, an edge information guidance module (EIG) is designed to enhance the edge in-

formation of changed objects by leveraging the high-frequency details of images. This enhancement improves the edge accuracy of the changed objects. Simultaneously, an innovative dynamically deformable Transformer (DDaT) is designed to adaptively match changed objects with different shapes, selecting features relevant to changes to model long-range dependency relationships and enhance the network's feature expression capability. Experimental results show that the proposed method significantly improves the detection accuracy on three public datasets: LEVIR-CD, CDD and DSIFN-CD, and is significantly better than the current mainstream networks in terms of edge accuracy and internal integrity of changed objects.

**Key words:** remote sensing images; change detection; high-frequency information; edge information; dynamically deformable Transformer

**Foundation Item(s):** National Natural Science Foundation of China (No.62271296, No.62201334); Shaanxi Province Outstanding Youth Science Fund (No.2021JC-47); Shaanxi Province Key Research and Development Program (No.2022GY-436, No.2021ZDLGY08-07); Shaanxi Province Natural Science Basic Research Program (No.2022JQ-634, No.2022JQ-018); Shaanxi Province Innovation Capability Support Program (No.2020SS-03)

## 1 引言

遥感图像变化检测旨在对同一地区在不同时期拍摄的遥感图像进行比较,以识别和分析图像之间的变化信息<sup>[1]</sup>。这一技术已被广泛应用于土地利用、城市规划<sup>[2]</sup>、灾害评估<sup>[3]</sup>等领域。

在早期遥感图像变化检测任务中,通常采用基于图像代数运算、特征提取与降维、后分类等方法,例如差值法<sup>[4]</sup>、比值法<sup>[5]</sup>、主成分分析<sup>[6]</sup>、独立成分分析<sup>[7]</sup>、多元变化检测<sup>[8]</sup>、迭代加权多元变化检测<sup>[9]</sup>等。这些传统方法在低分辨率图像和前景背景信息对比明显的图像上表现良好。但随着成像技术的迅速发展,各领域对变化检测任务提出了更高的实时性和准确性要求,而传统方法提取的变化目标特征表达能力有限,且对噪声敏感,在识别精度和鲁棒性方面表现不佳。

近年来,随着深度学习技术的快速发展,CNN(Convolutional Neural Network)能够自动学习复杂地面目标的抽象和多级特征,在变化检测任务中展现出卓越性能。U-net<sup>[10]</sup>以其卓越的分割能力成为基准模型,随后,Daudt<sup>[11]</sup>等人将孪生网络引入高分辨率遥感图像变化检测任务,充分利用双时相图像更丰富的特征,显著提高变化检测精度,使孪生网络成为该任务的标杆网络。该工作激发了一系列基于孪生网络的遥感图像变化检测方法<sup>[12-18]</sup>。但由于CNN局部特征建模的固有局限性,这类网络无法实现图像的长程依赖建模,在一定程度上限制了其性能。

为克服CNN模型的固有缺陷,学者们将Transformer<sup>[19]</sup>模型引入到遥感变化检测任务中,取得了显著成效。Transformer采用全局自注意力机制,有效地解决了CNN仅能获取图像局部特征的缺陷,提高了遥感图像变化检测精度。基于此涌现出一系列变化检测方法<sup>[20-25]</sup>,分为纯Transformer架构和CNN与Transformer的混合架构。纯Transformer架构通过自注意力实现了

对遥感图像的长程依赖建模,但由于缺少CNN的归纳偏置,对于小目标变化的检测效果有限。此外,Transformer在建模时容易忽略局部细节特征,降低了微小目标和尺度变化大的目标的检测精度。而CNN与Transformer的混合架构充分利用了两者优势,实现了对遥感图像的局部与全局信息建模,获得了更佳的变化检测效果。然而,这些网络仍然存在以下两个问题:首先,它们在对遥感图像的局部信息建模时通常忽略了图像高频信息,难以有效捕捉变化区域的边缘信息,从而导致边缘像素的不确定性。其次,上述网络在对遥感图像进行全局特征建模时采用形状和大小固定的图像块输入,忽略了不同形状大小的变化区域对模型精度的影响,难以避免不相关区域对结果的负面影响,并且忽略了空间与通道之间的相关性,导致自注意力表达不充分。为解决上述问题,本文提出一种新型的变化检测模型,提高检测性能。

## 2 相关工作

### 2.1 传统变化检测方法

在早期阶段中,主要采用无监督数学方法完成变化检测,如差值法<sup>[4]</sup>、比值法<sup>[5]</sup>和变化向量分析法<sup>[26]</sup>等。这些方法通过减除运算生成特征图,但忽略了不同成像条件对变化检测精度的影响。基于特征提取与降维的方法,如主成分分析、独立成分分析、多元变化检测和迭代加权多元变化检测,将图像转换到特定的特征空间,突出变化区域并抑制背景信息,但需要根据数据类型的不同选择合适的变化检测方法。此外,基于后分类的变化检测方法首先对双时相的遥感图像进行独立分类,然后比较分类结果,但获得高精度分类结果通常较为困难,导致变化检测结果的不确定性。

## 2.2 基于CNN的变化检测方法

随着深度学习的不断发展,基于CNN的方法也被应用于变化检测任务.这些方法通过多层网络结构学习图像的高维特征信息,减少噪声对图像的干扰,从而提高检测精度.Daudt<sup>[11]</sup>等人将孪生网络引入变化检测任务中,以准确检测变化区域.Chen<sup>[27]</sup>等人通过并行使用卷积和池化提取多尺度空间特征,有效避免卷积运算的平滑效应.Lei<sup>[28]</sup>等人采用金字塔结构聚合全局上下文关系,提高变化检测精度.Peng<sup>[29]</sup>等人采用端到端的网络架构,将残差连接引入Unet++模型中,提高了模型对高层次特征的捕捉能力.Chen<sup>[30]</sup>等人提出时空注意力机制,利用时空依赖关系生成更优的特征表示,减少因配准误差而造成的错误检测.但由于CNN中卷积运算的固有局限性以及其对高频信息的削减效应,最终导致变化目标边缘模糊,限制了主流变化检测网络的能力.

## 2.3 基于Transformer的变化检测方法

近年来,Transformer逐渐被应用于遥感图像变化检测任务中,并取得了显著成就.Transformer采用全局自注意力机制,在处理遥感变化检测任务时能够充分利用遥感图像的全局特征,克服了CNN仅能获取图像局部特征的缺陷,提高变化检测精度.BIT<sup>[20]</sup>首次将Transformer引入到变化检测任务中,通过在紧凑的语义标记上进行时空上下文建模,有效处理同一物体的不同语义概念,完成复杂场景下的变化检测任务.Swin-SUNet<sup>[23]</sup>采用基于Swin-Transformer<sup>[31]</sup>的孪生网络架

构,有效提取图像全局信息.ChangeMask<sup>[32]</sup>提出语义感知解码器用于学习语义表示,从而获得更优的遥感图像变化检测结果.ICIF<sup>[33]</sup>通过有效交互局部与全局信息,以获得精准的变化检测结果.上述方法虽能对遥感图像的全局空间信息进行建模,但在建模过程中存在着无关像素对模型造成的负面影响.此外,由于Transformer接受固定大小和位置的图像块作为输入,导致完整目标被分割成不同位置输入,破坏了图像本身的空间语义信息,一定程度上限制了遥感图像的变化检测精度.

## 3 本文方法

### 3.1 网络整体架构

局部和全局特征的有效融合有助于提升遥感图像变化检测精度,CNN通过卷积运算和分层特征表示来捕获遥感图像中的局部特征,而Transformer网络通过级联自注意力机制实现图像中全局特征的提取.为了充分利用遥感图像中的局部细节和全局语义特征,本文提出了基于边缘信息引导模块(Edge Information Guidance module, EIG)和动态可变形Transformer(Dynamically Deformable Transformer, DDaT)的遥感图像变化检测网络(Edge Guided and Dynamically Deformable Transformer, EGDDT).网络的整体架构如图1所示.简单而言,EGDDT将双时相图像输入到权重共享的编码器后得到各阶段的差值图像,并将其馈送至解码器,最终得到变化检测结果.

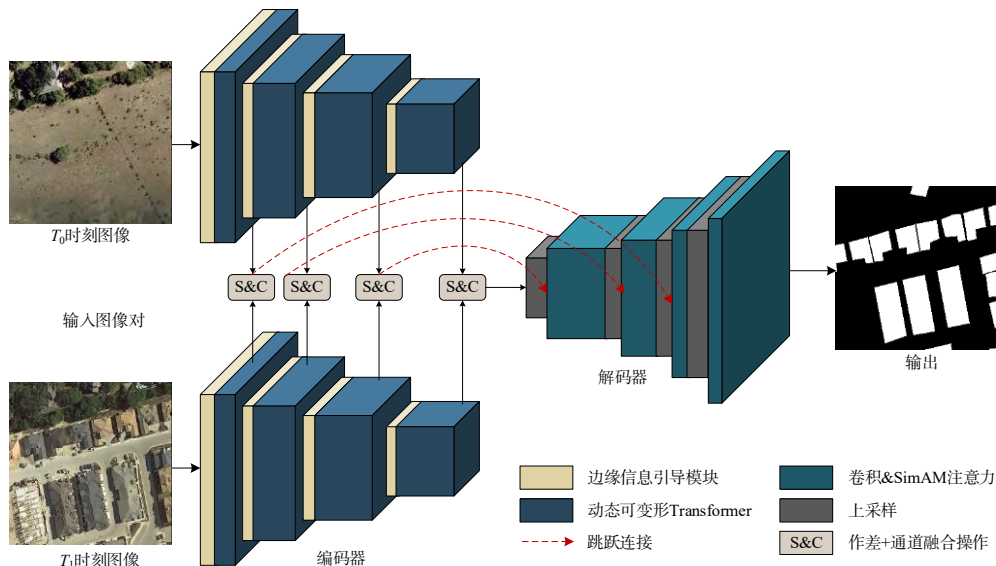


图1 EGDDT的整体架构

EGDDT将CNN提取到的局部细节信息馈送到Transformer中以增强全局信息特征,从而得到更为优

秀的特征表示. EGDDT采用标准的编解码结构,其编码器权重共享,由EIG和DDaT两个模块级联组成.在

编码阶段,首先通过EIG增强图像变化区域的边缘完整性,然后通过DDaT更好地匹配变化目标区域,降低背景像素对网络精度的负面影响.在跳跃连接的作差融合模块中,首先将每个编码层获取的双时相特征图进行通道融合和降维,随后将其与通过作差操作得到的差分图相加.此模块可以提供更加丰富的双时相信息,解决了传统作差操作导致的信息丢失问题.在解码阶段,通过双线性插值对具有丰富语义信息的差分图进行上采样解码.然后将其输入到卷积解码器中,该解码器由卷积层和SimAM<sup>[34]</sup>注意力组成.与传统的注意力机制不同,SimAM注意力同时关注目标区域的空间和通道维度,通过优化能量函数来发掘每个神经元的重要性,这是一种简单而高效的注意力机制,无需引入额外的网络参数,进一步增强语义信息.最后,将编码层获取的浅层特征信息融合到每个输出层,以丰富细节信息,最终形成编解码结构变化检测网络.

### 3.2 边缘信息引导模块

传统CNN网络在获取图像特征时通常会导致高频信息的削减<sup>[35]</sup>.这主要是由于卷积和池化操作中平滑滤波器在局部区域内平均或加权平均输入像素的值,从而丢失了图像中的一部分高频细节信息,如纹理、边缘和细节区域.这些高频信息中包括图像的边缘信息,而图像的边缘信息对于变化检测任务非常重要.因此,传统CNN网络可能导致最终结果的变化区域边缘模糊,影响变化检测的结果.本文在不同类型的图像上对卷积和池化的高频削减效应进行了验证,结果如图2所示,可明显观察到由于卷积和池化操作导致的图像中边缘及细节信息的丢失.

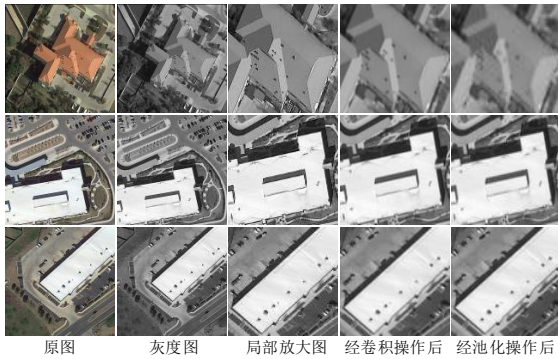


图2 卷积和池化操作对高频信息的衰减结果

为了解决这一问题,本文设计了边缘信息引导模块,利用EIG补充被削减的高频信息,更有效地捕获变化检测任务中变化区域的边缘信息.由于EIG的输入输出维度相同,因此它可以灵活高效地嵌入到其他网络中.EIG由多尺度空间注意力模块和边缘检测模块组成,其结构如图3所示.

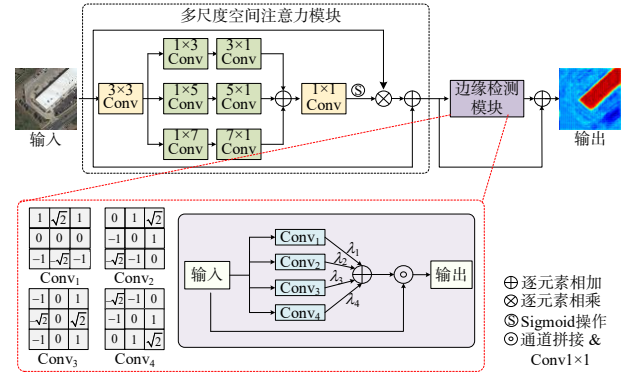


图3 EIG的整体结构

首先,多尺度空间注意力模块用于引导网络更加关注遥感图像中的变化区域,并通过多分支的非对称卷积来处理输入特征图 $X_{in}$ ,以捕获遥感图像不同尺度的上下文信息.这种方法不仅能够保证网络复杂度较低,还有效地融合不同尺度的特征图信息,提高网络的特征表达能力.通过上述的卷积操作在空间维度上重新校准特征图,以确保网络更好地关注真实的变化区域,减少未变化区域对网络建模的影响.可以表示为

$$X_{att} = \delta \left( \text{Conv}_{1 \times 1} \left( \sum_i \text{Scale}_i \left( \text{As-Conv} \left( \text{Conv}_{3 \times 3} \left( X_{in} \right) \right) \right) \right) \right) \quad (1)$$

$$X_{out} = (X_{att} \otimes X_{in}) + X_{in} \quad (2)$$

其中, $X_{in}$ , $X_{att}$ 和 $X_{out}$ 分别代表输入特征图,注意力权重和输出结果.As-Conv代表非对称卷积, $\sum_i \text{Scale}_i()$ 代表不同尺度卷积的相加结果, $i \in \{3, 5, 7\}$ 代表卷积核的大小, $\delta(x)$ 代表Sigmoid操作.

其次,在得到注意力输出结果图后,通过边缘检测模块进一步提取图像变化区域的高频边缘信息,其结构如图3红框中所示.传统的图像处理方法可分为空间域方法和频域方法,频域方法可以从不同的角度分析图像并提取不同频域的特征.在进行空间域注意力操作后,引入边缘检测算子作为频域方法的一部分.鉴于Sobel算子能够以较低的计算成本提取特征图的高频特征分量<sup>[36]</sup>,因此我们选择Sobel算子用于边缘检测模块.将多尺度空间注意力输出的结果图送入卷积核尺寸为 $3 \times 3$ 的各向同性Sobel算子中,利用算子更准确的加权系数,从四个不同的方向提取特征图的高频信息,以确保变化区域的边缘信息完整性.在得到特征图的高频信息后,采用通道融合的方式将这些信息补充到原特征图中,具体操作可表示为

$$X_{out} = \text{Conv}_{1 \times 1} \left( \text{Concat} \left( \sum_{i=1}^4 \lambda_i \text{Conv}_i \left( X_{in}, X_{in} \right) \right) \right) \quad (3)$$

$\lambda_i$ 是可学习的参数,其中, $i \in \{1, 2, 3, 4\}$ ,在端到端

的训练过程中自适应地控制每个方向上的边缘信息重要程度.

此外,为了细化特征图并减少梯度消失对训练的影响,每个阶段的输出结果中引入了残差连接,以提高网络在反向传播过程中训练的稳定性.

### 3.3 动态可变形Transformer

现有基于Transformer的变化检测方法通常将图像划分为多个图像块后对其进行自注意力计算.这种方法使用固定大小和位置的图像块作为输入,破坏了图像中目标区域的空间语义信息完整性,难以处理形状

多变的变化目标区域.此外,在全局等同依赖建模的过程中,Transformer难以避免无关像素对结果的影响.鉴于此,受可变形卷积<sup>[37,38]</sup>和动态卷积<sup>[39]</sup>的启发,本文设计了DDaT,以可变形的学习方式灵活地对变化区域进行建模,以忽略背景区域的干扰.DDaT通过学习图像内容的动态偏移,将感受野变成不规则形状,从而能够更好地匹配遥感图像中形状不同的目标区域,并充分捕捉通道与空间之间的相关性,提高现有模型的特征表达能力.DDaT的整体结构如图4所示,其中 $W_{sq}$ ,  $W_{sk}$ ,  $W_{sv}$ 和 $W_{cq}$ ,  $W_{ck}$ ,  $W_{cv}$ 分别为空间和通道层面上的线性投影矩阵.

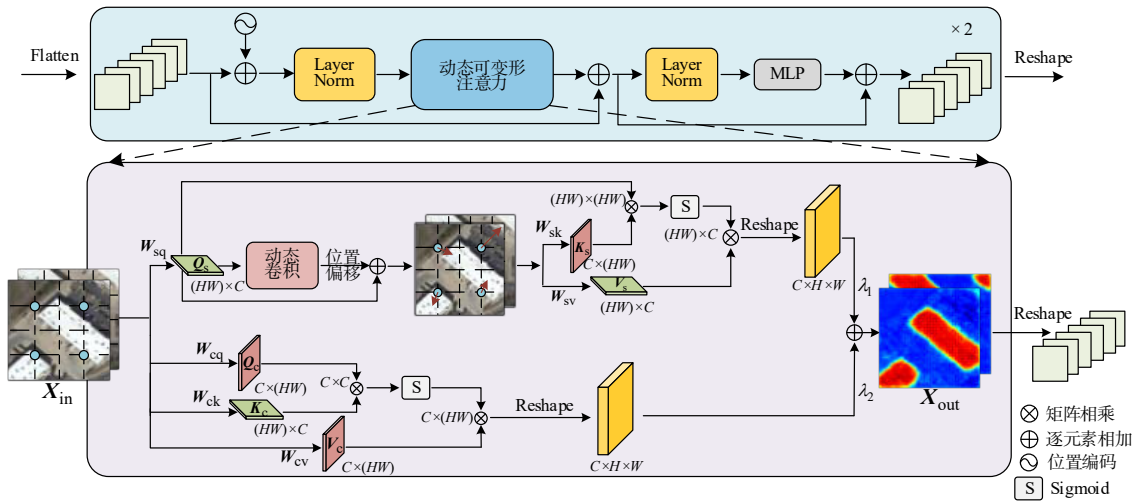


图4 DDaT的整体结构

首先,将输入特征经过线性映射后得到查询值 $Q_s$ ,然后将其输入动态卷积中,根据每个输入图像的不同,生成其特定的偏移量 $\Delta p$ .如下所示:

$$\Delta p = \text{DConv}(Q_s) \quad (4)$$

DConv()代表动态卷积.具体来说,根据图像本身的不同特征,使用不同的权重系数来对卷积核进行调整,以捕获与内容相关的语义特征,学习可信的偏移量 $\Delta p$ .动态卷积中的权重变化是通过引入权重系数来决定的,定义如下:

$$y = \sigma((\alpha_1 W_1 + \dots + \alpha_n W_n)x) \quad (5)$$

其中,输入特征为 $x$ ,输出特征为 $y$ , $n$ 为权重系数的个数, $\alpha_n$ 是具有可学习参数的权重系数, $W_n$ 为卷积核权重 $\sigma$ 为激活函数.在进入卷积核之前,不同的权重矩阵按照对应的系数线性组合,实现卷积核的动态调整.

接下来,根据输入特征将偏移量嵌入原查询值 $Q_s$ 后生成动态偏移的特征图.但学习到的偏移量可能是小数,并且在位置上难以对应原输入特征上的实际位置的像素,因此需要使用双线性插值来避免最终生成

小数,并使其可微分,操作可表示为

$$\phi(\mathbf{G}; (p_x, p_y)) = \sum_{h_x, h_y} f(p_x, h_x) f(p_y, h_y) \mathbf{G}[h_x, h_y, :] \quad (6)$$

其中, $f(a, b) = \max(0, 1 - |a - b|)$ ,  $(h_x, h_y)$ 代表特征图 $\mathbf{G}$ 上的所有位置.在偏移位置 $p + \Delta p$ 使用采样函数 $\phi$ 对特征进行采样,得到偏移后的特征图为

$$\hat{x} = \phi(x; p + \Delta p) \quad (7)$$

之后,通过对此特征图进行线性投影操作生成Key值和Value值,通过此,我们将生成的Key值以及Value值灵活地应用于不同位置和形变的目标区域,从而增强原有的特征表达.通过可变形的Key值 $K_s$ 和Value值 $V_s$ ,实现对不同尺度以及目标位置信息的有效关注,从而得到最终空间自注意力结果表示 $A_1$ .除此之外,还引入通道自注意力 $A_2$ 对传统Transformer进行补充,提高网络模型对遥感图像的光谱特征表达能力,最终特征融合输出的结果为

$$A_1(Q_s, K_s, V_s) = \text{Softmax}\left(\frac{Q_s K_s^T}{\sqrt{d}}\right) V_s \quad (8)$$

$$A_2(Q_c, K_c, V_c) = \text{Softmax}\left(\frac{Q_c K_c^T}{\sqrt{d}}\right) V_c \quad (9)$$

$$X_{\text{out}} = \lambda_1 A_1 + \lambda_2 A_2 \quad (10)$$

其中,  $\lambda_1$  和  $\lambda_2$  被设计为可学习的参数, 用来控制每个注意力分支对当前任务中的空间和通道信息的重要程度.

## 4 实验结果与分析

### 4.1 实验数据集与评估指标

#### 4.1.1 数据集

为了验证本文提出方法的有效性, 本文在三个大型公开遥感变化检测数据集 LEVIR-CD<sup>[29]</sup>、CDD<sup>[40]</sup> 和 DSIFN-CD<sup>[12]</sup> 上进行了实验. 各数据集介绍如下.

LEVIR-CD 数据集中包含 637 对超高分辨率 Google Earth 图像, 分辨率为 0.5 m, 每个图像大小为 1 024×1 024. 为了防止过拟合, 本文进行了随机旋转和随机裁剪等数据增强操作, 并将图像随机裁剪为 256×256 的大小. 在最终的实验中, 选取了 10 000 对作为训练集, 1 024 对作为验证集, 2 048 对作为测试集.

CDD 数据集是由 Google Earth 获取的同一地区具有季节性变化的遥感图像, 分辨率从 0.03 m 至 1 m, 通过随机裁剪等数据增强操作, 共得到 16 000 对大小为 256×256 的图像对, 其中 10 000 对用于训练集, 3 000 对用于验证集, 剩余的 3 000 对作为测试集.

DSIFN-CD 数据集由 Google Earth 手动收集, 包含来自中国 6 个不同城市的高分辨率图像. 经过简单的数据增强操作后, 最终得到 3 600 对训练集, 340 对验证集以及 48 对测试集.

#### 4.1.2 评估标准

本文采用精度 (Precision, Pre)、召回率 (Recall, Rec)、 $F_1$  分数 ( $F$ -score,  $F_1$ ) 和距离理想位置的距离 (Distance from the Ideal Position, DIP)<sup>[41]</sup> 这四个主要评价指标, 综合评估该网络的性能.

Pre 反映的是预测结果与目标结果的接近程度, 表示在预测结果为正的数据中预测结果正确的样本所占的比例. 计算公式如下:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

Rec 指的是在总的标签为正的样本中被预测正确所占的比例, 衡量的是预测结果的查全率. 计算方式为

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

在理想情况下, Pre 和 Rec 都尽可能高, 但在某些情况下, 两者相互制约. 为了提高 Pre, 模型会过于保守而漏掉很多正样本, 导致 Rec 值降低. 单凭这两个指标来评定网络的性能并不充分. 因此引入  $F_1$  分数和 DIP 对

网络的性能进行综合性的评价. 其中  $F_1$  分数的计算公式如下:

$$F_1 = 2 \times \frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \quad (13)$$

DIP 旨在衡量预测结果本身作为最终目标的理想位置的接近程度, 计算方式为

$$\text{DIP} = 1 - \sqrt{\frac{(1 - \text{Pre})^2 + (1 - \text{Rec})^2}{2}} \quad (14)$$

在上述公式中, TP 和 TN 分别表示对正类标签样本和负类标签样本的正确预测数量; FP 和 FN 分别表示对标签图中的正类和负类样本的错误预测个数.

### 4.2 实验细节

本研究方法基于 Pytorch 框架实现, 在 NVIDIA GeForce RTX 3090 显卡上进行了 200 个 epoch 的训练. 采用 AdamW 优化器, 将初始学习率设置为 0.000 1, 权重衰减设置为 0.000 5, 动量设置为 0.99. 为解决前后背景不平衡的问题, 本文采用二元交叉熵损失<sup>[42]</sup> (BCE Loss) 和 Dice Loss<sup>[43]</sup> 来优化网络权重. 最终损失函数可表示为

$$L = L_{\text{BCE}} + L_{\text{Dice}} \quad (15)$$

### 4.3 实验结果分析

#### 4.3.1 与其他算法的对比结果

为了验证本文所提方法的有效性, 选取了 8 种当前主流的变化检测方法来进行比较, 并对比分析了对比方法和所提出算法在三个数据集上的变化检测结果.

##### (1) LEVIR-CD

表 1 展示了本文提出的 EGDDT 模型与目前主流的 CNN 和 Transformer 网络在 LEVIR-CD 数据集上的定量分析结果, 每个度量的最佳结果已加粗显示. 实验结果表明, 本文所提方法几乎达到了最佳性能水平, 并取得了最佳的 Rec、 $F_1$  和 DIP 指标. 相对于主流方法 SNUNet、BIT 和 Hybrid-TransCD, EGDDT 在 Rec 上分别提高了 4.23%、2.09% 和 2.84%; 在  $F_1$  上分别高出 3.01%、1.92% 和 1.08%.  $F_1$  和 DIP 指标的结果表明, EGDDT 能更有效地完成遥感图像变化检测任务.

图 5 的可视化结果进一步展示了所提出方法的卓越性能, 其中红框突出重点细节区域. 从图中可以清晰地看出, 在边缘信息引导模块的协助下, EGDDT 成功地避免了其他方法, 如 SNUNet、ChangeFormer 和 Hybrid-TransCD 等存在的边缘模糊问题, 保证了边缘信息的完整性. 此外, 在动态可变形 Transformer 的协同作用下, EGDDT 通过高度关注变化目标区域, 确保了空间语义信息的完整性. 相较于其他方法, EGDDT 能够有效实现对变化目标内部区域的准确检测, 呈现出更为精确的检测结果. 特别地, ChangeFormer 模型的 Pre 极高但 Rec 极低, 从第三行细小区域的变化结果可明显看出其

存在严重的误检问题. 而EGDDT通过边缘信息引导模块和动态可变形Transformer的串联,能够更准确地检测变化检测任务中的小目标和变化区域的边缘,确保目标区域的完整性和准确性,从而避免了漏检和误检情况的发生.

表1 不同算法在LEVIR-CD数据集上的定量比较结果 单位:%

网络类型	方法	Pre	Rec	$F_1$	DIP
CNN	FCN-PP <sup>[28]</sup>	82.09	84.48	83.27	83.24
	STANet <sup>[30]</sup>	86.14	89.39	87.73	87.65
	IFN <sup>[12]</sup>	86.00	89.73	87.03	87.72
	SNUNet <sup>[16]</sup>	89.06	87.53	88.29	88.27
Transformer	BIT <sup>[20]</sup>	89.24	89.37	89.31	89.30
	ChangeFormer <sup>[21]</sup>	<b>92.05</b>	88.80	90.40	90.28
	SwinSUNet <sup>[23]</sup>	90.51	89.72	90.11	90.10
	Hybrid-TransCD <sup>[24]</sup>	91.45	88.72	90.06	89.99
上二者混合	EGDDT(ours)	90.83	<b>91.24</b>	<b>91.03</b>	<b>91.03</b>

(2) CDD

表2详细比较了本文方法与目前主流方法在CDD数据集上的表现,最佳的结果已以粗体突出显示. 在CDD数据集上,本文方法表现卓越,相比于主流方法如SNUNet、BIT和Hybrid-TransCD,本文方法在 $F_1$ 分数上分别高出1.86%、1.18%和1.19%,在DIP上分别高出2.10%、1.17%和1.18%.  $F_1$ 和DIP的结果充分证明了本

文所提方法的卓越性.

CDD数据集上的实验结果可视化分析如图6所示,其中红框突出重点细节区域. 在复杂的变化区域,如第一行所示,其他方法的结果边缘不够光滑,甚至存在着错误的边缘检测结果. EGDDT则通过EIG有效关注和补充边缘信息,实现了对边缘信息的准确定位和检测. 在第二行小目标区域的变化检测中,其他方法均未能准确定位目标,存在着不同程度的误检和漏检. EGDDT通过DDaT实现了对变化目标区域的有效关注,能够准确关注和检测图像中的变化区域,提高其空间完整性. 即使是在复杂的变化区域中,其性能依然卓越,保证了目标区域的边缘完整性和整体准确性,进一步证实了EGDDT架构的有效性.

表2 不同算法在CDD数据集上的定量比较结果 单位:%

网络类型	方法	Pre	Rec	$F_1$	DIP
CNN	FCN-PP <sup>[28]</sup>	81.69	90.31	85.78	85.35
	STANet <sup>[30]</sup>	88.98	93.11	91.00	90.81
	IFN <sup>[12]</sup>	85.33	91.76	88.43	88.10
	SNUNet <sup>[16]</sup>	90.92	94.75	92.79	92.58
Transformer	BIT <sup>[20]</sup>	92.89	94.02	93.45	93.43
	ChangeFormer <sup>[21]</sup>	94.26	93.46	93.84	93.84
	SwinSUNet <sup>[23]</sup>	<b>95.70</b>	92.30	94.00	93.76
	Hybrid-TransCD <sup>[24]</sup>	93.86	93.02	93.43	93.42
上二者混合	EGDDT(ours)	93.94	<b>95.17</b>	<b>94.55</b>	<b>94.52</b>

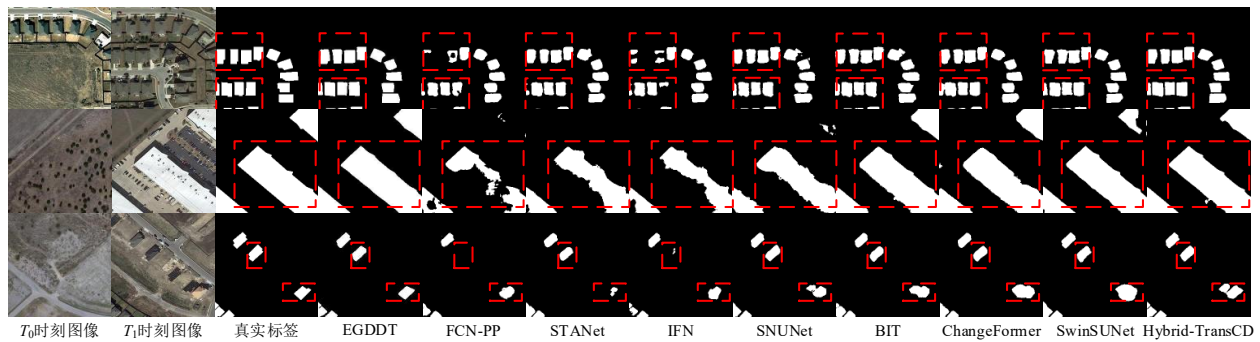


图5 LEVIR-CD数据集下的可视化结果

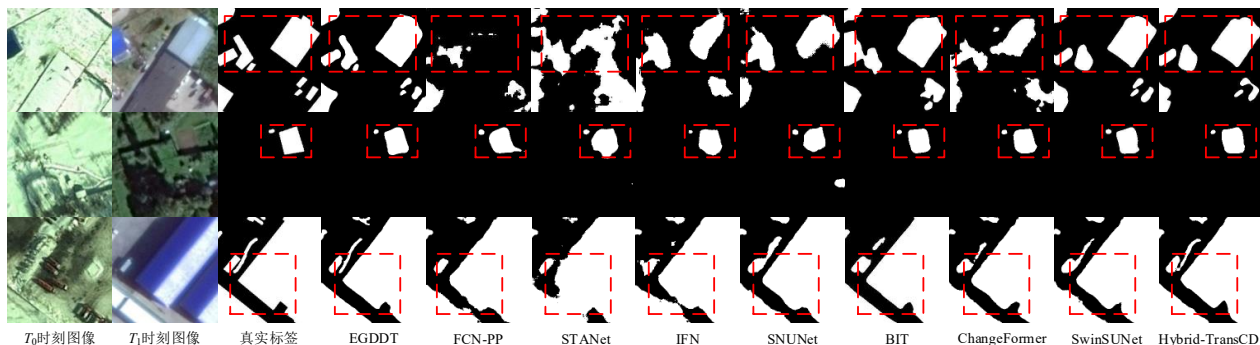


图6 CDD数据集下的可视化结果

### (3) DSIFN-CD

在 DSIFN-CD 数据集上对本文方法和其他方法进行了定量比较,结果如表 3 所示. 其中最佳的结果以粗体显示. 同样,本文方法在 DSIFN-CD 数据集上也取得了最优结果. 相比于目前主流的三种架构 SNUNet、BIT 和 Hybrid-TransCD,本文方法在  $F_1$  分数上分别高出 7.90%、2.67% 和 2.03%,在 DIP 上分别高出 7.87%、2.67% 和 2.17%,从数据上进一步证实 EGDDT 的有效性.

图 7 展示了在 DSIFN-CD 数据集上的实验可视化结果,红框突出重点细节区域,可以明显观察到其他方法对图像中变化区域的错误判断. 如第一行和第三行所示,对于细小变化区域和其内部完整性而言,其他方法存在着严重的漏检和误检问题,并且边缘检测也不够准确,存在着错误的边缘定位. 相反,EGDDT 通过 EIG

和 DDaT 的串联使用,能够有效地保证图像中变化区域的边缘准确性和空间完整性,在复杂的变化检测任务中仍表现出良好的性能,对于变化区域的边缘和整体准确性有着较好的检测性能.

表 3 不同算法在 DSIFN-CD 数据集上的定量比较结果 单位:%

网络类型	方法	Pre	Rec	$F_1$	DIP
CNN	FCN-PP <sup>[28]</sup>	56.42	59.25	57.80	57.81
	STANet <sup>[30]</sup>	66.22	67.16	66.69	66.68
	IFN <sup>[12]</sup>	<b>72.36</b>	63.86	67.85	67.82
	SNUNet <sup>[16]</sup>	62.47	69.74	65.90	65.91
Transformer	BIT <sup>[20]</sup>	68.36	70.18	69.26	69.25
	ChangeFormer <sup>[21]</sup>	69.38	70.51	69.94	69.93
	SwinSUNet <sup>[23]</sup>	68.72	71.68	70.17	70.16
	Hybrid-TransCD <sup>[24]</sup>	68.79	70.42	69.69	69.59
上二者混合	EGDDT(ours)	70.28	<b>71.96</b>	<b>71.11</b>	<b>71.10</b>

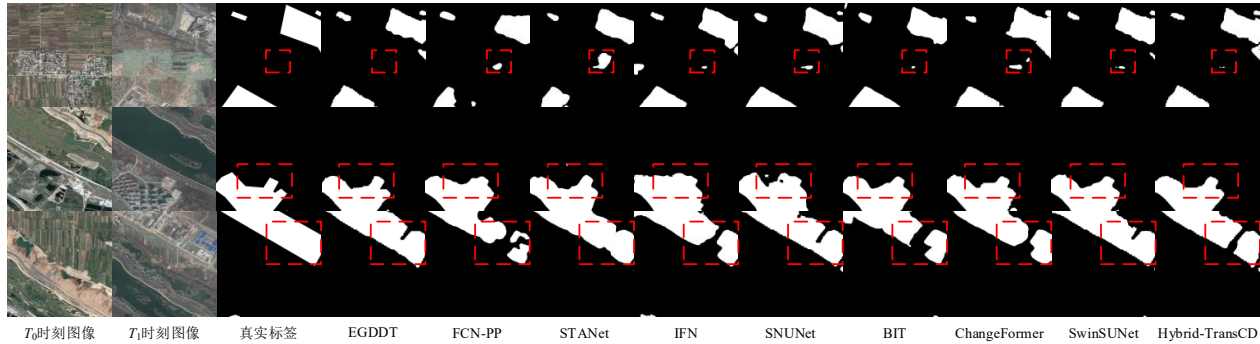


图 7 DSIFN-CD 数据集下的可视化结果

### (4) 计算效率分析

我们从多个角度测试和分析提出的 EGDDT 模型与其他对比方法在遥感图像变化检测中的计算效率,使用了四个评估指标: $F_1$ 、DIP、模型参数量(Params)和 FLOPs,具体的结果如表 4 所示. 实验结果表明,相较于主流的 Transformer 网络,EGDDT 几乎保持了最低的

参数量和 FLOPs,仅高于 BIT 模型. 这是因为 BIT 模型把图像转化为简单的语义信息后进行 Transformer 建模,导致检测精度低,因此与其进行复杂度比较不具备现实意义. 显然,我们的方法在  $F_1$  和 DIP 方面表现最优,实现了结果精确度与计算复杂度之间的动态平衡.

表 4 不同算法的计算效率比较结果

网络类型	方法	$F_1$ /%			DIP/%			FLOPs/G	Params/M
		LEVIR-CD	CDD	DSIFN-CD	LEVIR-CD	CDD	DSIFN-CD		
CNN	FCN-PP <sup>[28]</sup>	83.27	85.78	57.80	83.24	85.35	57.81	34.65	28.13
	STANet <sup>[30]</sup>	87.73	91.00	66.69	87.65	90.81	66.68	6.58	16.93
	IFN <sup>[12]</sup>	87.03	88.43	67.85	87.72	88.10	67.82	41.18	50.71
	SNUNet <sup>[16]</sup>	88.29	92.79	65.90	88.27	92.58	65.91	33.04	12.03
Transformer	BIT <sup>[20]</sup>	89.31	93.45	69.26	89.30	93.43	69.25	8.44	6.93
	ChangeFormer <sup>[21]</sup>	90.40	93.84	69.94	90.28	93.84	69.93	202.83	41.01
	SwinSUNet <sup>[23]</sup>	90.11	94.00	70.17	90.10	93.76	70.16	21.19	50.95
	Hybrid-TransCD <sup>[24]</sup>	90.06	93.43	69.69	89.99	93.42	69.59	51.38	166.57
上二者混合	EGDDT(ours)	<b>91.03</b>	<b>94.55</b>	<b>71.11</b>	<b>91.03</b>	<b>94.52</b>	<b>71.10</b>	11.68	21.46

(5) 热力图可视化结果分析

在 LEVIR-CD 数据集上对变化区域结果的热力图可视化分析如图 8 所示. 从图中明显可见,EGDDT 能有效聚焦于遥感图像中的变化区域的边缘信息,并实现其精准定位与恢复,确保了边缘信息的准确性. 此外,EGDDT 还更加准确地关注了不规则的目标变化区域,保证了其内部的完整性. 显著提高了遥感图像变化检测任务的性能,验证了本文提出的结合边缘信息引导模块和动态可变形 Transformer 的方法能够产生更准确的变化检测结果.

4.3.2 消融实验

(1) 定量分析比较结果

为了验证网络中不同模块的有效性,本小节在 LEVIR-CD 数据集上进行了一系列的消融实验分析,结果如表 5 所示,最佳结果以粗体显示. 本文以纯 Transformer 块搭建的网络作为基线网络. 实验结果表明,

表 5 不同模块在 LEVIR-CD 数据集上的实验结果

方法	Pre/%	Rec/%	$F_1$ /%	DIP/%	FLOPs/G	Params/M
Base	86.00	88.09	87.03	87.03	9.98	17.76
Base+EIG	88.91	89.31	89.11	89.11	10.39	18.29
Base+DDaT	89.58	90.87	90.22	90.20	11.27	20.94
Base+EIG+DDaT	<b>90.83</b>	<b>91.24</b>	<b>91.03</b>	<b>91.03</b>	11.68	21.46

(2) 定性分析比较结果

为了进一步验证本文提出的各个模块的有效性,我们进行了可视化分析,如图 9 所示. 从结果可以清晰地看出,在增加了 EIG 模块后,网络能够更准确地识别变化区域的边缘信息,从而提高边缘完整性. 此外,DDaT 的引入也提高了网络对变化区域内部完整性的检测. 最终,通过两个模块的协同作用,网络实现了出色的性能表现.

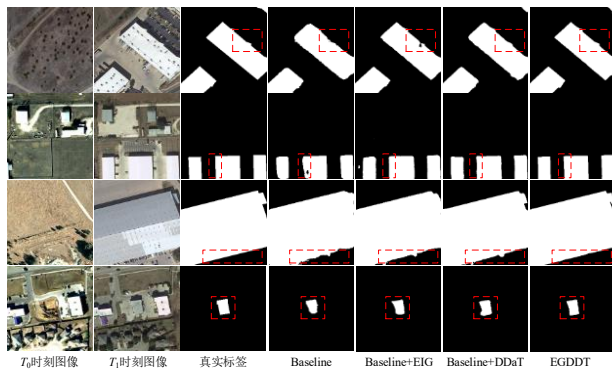


图 9 不同模块在 LEVIR-CD 数据集下的可视化结果

5 结论

本文提出一种用于遥感图像变化检测的新架构 EGDDT 网络. EGDDT 通过边缘信息引导模块和动态可

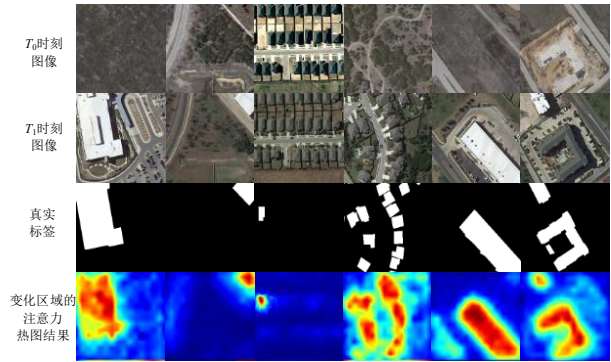


图 8 在 LEVIR-CD 数据集上的热力图结果

EIG 和 DDaT 均可提高变化检测精度,而两种方法的组合可以进一步提升检测性能. 此外,从网络参数量可以看出,两种方法的加入并未给网络带来太大负担. 消融实验充分证明了本文所提出的 EIG,DDaT 以及整体网络在遥感图像变化检测任务中的有效性.

变形 Transformer 的串联使用,有效地保留了遥感图像中的局部细节特征和全局语义特征,解决了当前遥感图像变化检测任务的主要问题. 在三个公开的变化检测数据集上的实验结果表明,EGDDT 能够在增加少量网络参数的情况下便取得了最先进的检测结果,显著提高了变化检测的性能. 在后续研究中,我们计划将 EGDDT 中的各个模块推广应用到更广泛的数据集上,并通过调整或引入新模块来解决更高分辨率图像、更为复杂的任务,以及更高精度需求的挑战,提高遥感图像变化检测的性能.

参考文献

[1] SEBASTIAN A, TUMA T, PAPANDREOU N, et al. Temporal correlation detection using computational phase-change memory[J]. Nature Communications, 2017, 8(1): 1115.

[2] DEMIR B, BOVOLO F, BRUZZONE L. Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 51(1): 300-312.

[3] FAN Y D, WEN Q, WANG W, et al. Quantifying disaster physical damage using remote sensing data—A technical work flow and case study of the 2014 Ludian earthquake in

- China[J]. *International Journal of Disaster Risk Science*, 2017, 8(4): 471-488.
- [4] SINGH A. Change detection in the tropical forest environment of northeastern India using Landsat[J]. *Remote Sensing and Tropical Land Management*, 1986, 44: 273-254.
- [5] HOWARTH P J, WICKWARE G M. Procedures for change detection using Landsat digital data[J]. *International Journal of Remote Sensing*, 1981, 2(3): 277-291.
- [6] DENG J S, WANG K, DENG Y H, et al. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data[J]. *International Journal of Remote Sensing*, 2008, 29(16): 4823-4838.
- [7] MARCHESI S, BRUZZONE L. ICA and kernel ICA for change detection in multispectral remote sensing images [C]//2009 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE, 2009: II-980-II-983.
- [8] NIELSEN A A, CONRADSEN K, SIMPSON J J. Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies[J]. *Remote Sensing of Environment*, 1998, 64(1): 1-19.
- [9] XU G B, LI H F, ZANG Y W, et al. Change detection based on IR-MAD model for GF-5 remote sensing imagery [J]. *IOP Conference Series: Materials Science and Engineering*, 2020, 768(7): 072073.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference. Cham: Springer International Publishing, 2015: 234-241.
- [11] CAYE DAUDT R, LE SAUX B, BOULCH A. Fully convolutional Siamese networks for change detection[C]//2018 25th IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2018: 4063-4067.
- [12] ZHANG C X, YUE P, TAPETE D, et al. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 166: 183-200.
- [13] HOU B, LIU Q J, WANG H, et al. From W-Net to CDGAN: Bitemporal change detection via deep learning techniques[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 58(3): 1790-1802.
- [14] ZHANG M, SHI W Z. A feature difference convolutional neural network-based change detection method[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(10): 7232-7246.
- [15] CHENG H Q, WU H Y, ZHENG J, et al. A hierarchical self-attention augmented Laplacian pyramid expanding network for change detection in high-resolution remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 182: 52-66.
- [16] FANG S, LI K Y, SHAO J Y, et al. SNUNet-CD: A densely connected Siamese network for change detection of VHR images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 1-5.
- [17] LEI T, WANG J, NING H L, et al. Difference enhancement and spatial—Spectral nonlocal network for change detection in VHR remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-13.
- [18] LEI T, GENG X Z, NING H L, et al. Ultralightweight spatial—Spectral feature cooperation network for change detection in remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-14.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.
- [20] CHEN H, QI Z P, SHI Z W. Remote sensing image change detection with Transformers[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-14.
- [21] BANDARA W G C, PATEL V M. A Transformer-based Siamese network for change detection[C]//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE, 2022: 207-210.
- [22] LIU M X, CHAI Z Q, DENG H J, et al. A CNN-Transformer network with multiscale context aggregation for fine-grained cropland change detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 4297-4306.
- [23] ZHANG C, WANG L J, CHENG S L, et al. SwinSUNet: Pure Transformer network for remote sensing image change detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-13.
- [24] KE Q T, ZHANG P. Hybrid-TransCD: A hybrid Transformer remote sensing image change detection network via token aggregation[J]. *ISPRS International Journal of Geo-Information*, 2022, 11(4): 263.
- [25] LI Q Y, ZHONG R F, DU X, et al. TransUNetCD: A hy-

- brid Transformer network for change detection in optical remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-19.
- [26] LU D, MAUSEL P, BRONDÍZIO E, et al. Change detection techniques[J]. International Journal of Remote Sensing, 2004, 25(12): 2365-2401.
- [27] CHEN H, WU C, DU B, et al. Change detection in multi-temporal VHR images based on deep siamese multi-scale convolutional networks[EB/OL]. (2020-07-10) [2023-06-26]. <https://arxiv.org/abs/1906.11479>.
- [28] LEI T, ZHANG Y X, LV Z Y, et al. Landslide inventory mapping from bitemporal images using deep convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(6): 982-986.
- [29] PENG D F, ZHANG Y J, GUAN H Y. End-to-end change detection for high resolution satellite images using improved UNet++[J]. Remote Sensing, 2019, 11(11): 1382.
- [30] CHEN H, SHI Z W. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection[J]. Remote Sensing, 2020, 12(10): 1662.
- [31] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 10012-10022.
- [32] ZHENG Z, ZHONG Y F, TIAN S, et al. ChangeMask: Deep multi-task encoder-Transformer-decoder architecture for semantic change detection[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 183: 228-239.
- [33] FENG Y C, XU H H, JIANG J W, et al. ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.
- [34] YANG L, ZHANG R Y, LI L D, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning. New York: PMLR, 2021: 11863-11874.
- [35] LUO T, MA Z, XU Z Q J, et al. Theory of the frequency principle for general deep neural networks[EB/OL]. (2019-07-02)[2023-06-26]. <https://arxiv.org/abs/1906.09235>.
- [36] AL-SUMAIDAE S, ABDULLAH M, AL-NIMA R, et al. Multi-gradient features and elongated quinary pattern encoding for image-based facial expression recognition [J]. Pattern Recognition, 2017, 71: 249-263.
- [37] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 764-773.
- [38] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable Transformers for end-to-end object detection[EB/OL]. (2021-03-18)[2023-06-26]. <https://arxiv.org/abs/2010.04159>.
- [39] CHEN Y P, DAI X Y, LIU M C, et al. Dynamic convolution: Attention over convolution kernels[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11030-11039.
- [40] LEBEDEV M A, VIZILTER Y V, VYGOLOV O V, et al. Change detection in remote sensing images using conditional adversarial networks[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2018, 42: 565-571.
- [41] NANDI A K. From multiple independent metrics to single performance measure based on objective function[J]. IEEE Access, 2023, 11: 3899-3913.
- [42] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. (2016-05-19)[2023-06-26]. <https://arxiv.org/abs/1409.0473>.
- [43] LI X, SUN X, MENG Y, et al. Dice loss for data-imbalanced NLP tasks[EB/OL]. (2019-11-07) [2023-06-26]. <https://arxiv.org/abs/1911.02855>.

#### 作者简介



雷涛 男, 1981年11月出生, 陕西大荔人。2011年在西北工业大学获得博士学位, 现为陕西科技大学教授, 博士生导师。主要从事图像处理、模式识别和计算机视觉等方面的研究工作。  
E-mail: leitao@sust.edu.cn



翟钰杰 男, 1999年11月出生, 山西运城人。现为陕西科技大学电子信息与人工智能学院硕士研究生。主要研究方向为计算机视觉。  
E-mail: 211612114@sust.edu.cn