

文章编号:1007-2780(2024)10-1361-19

基于 CNN-Transformer 结构的遥感影像变化检测

潘梦洋^{1,2}, 杨 航^{1*}, 范祥晖^{1,2}

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

摘要:现代高分辨率遥感图像变化检测借助卷积神经网络(Convolutional Neural Network, CNN)取得了显著成果。然而,卷积操作的感受野限制导致在学习全局上下文和远程空间关系方面存在不足。虽然视觉 Transformer 能有效捕获远程特征的依赖性,但其对影像变化细节的处理不足,导致空间定位能力有限且计算效率低下。为解决上述问题,本文提出了一种基于空间空洞金字塔池化的跨层级联线性融合端到端编解码混合 CNN-Transformer 的变化检测模型,兼具视觉 Transformer 和 CNN 的优势。首先,利用孪生 CNN 网络提取图像特征,并借助空洞金字塔池化模块对特征进行精细处理,从而更精准地捕获图像的细节特征信息。其次,将提取的特征转化为视觉单词,并通过 Transformer 编码器进行建模,以获取丰富的上下文信息。这些信息随后被反馈至视觉空间,通过 Transformer 解码器对原始特征进行强化,提升特征的表达效果。接着,采用跨层级联的方式将 CNN 提取的特征与 Transformer 编解码的特征进行融合,利用上采样技术联系不同分辨率的特征图,实现位置信息与语义信息的融合。最后,通过差异增强模块生成包含丰富变化信息的差异特征图。在 LEVIR、CDD、DSIFN 和 WHUCD 4 个公开遥感数据集上的广泛实验验证了本文方法的有效性。与其他先进方法相比,本文模型的性能更出色,有效改善了变化检测中的欠分割、过分割及边缘粗糙等问题。

关键词:遥感图像;变化检测;卷积神经网络;Transformer;空间空洞金字塔池化

中图分类号:TP391 文献标识码:A doi:10.37188/CJLCD.2024-0086

Remote sensing image change detection based on CNN-Transformer structure

PAN Mengyang^{1,2}, YANG Hang^{1*}, FAN Xianghui^{1,2}

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Modern high-resolution remote sensing images have achieved remarkable results in change detection with the aid of convolutional neural network (CNN). However, the limited receptive field of convolution operations leads to insufficient learning of global context and long-distance spatial relationships. While visual Transformers effectively capture dependencies in remote features, their handling of details in image changes is insufficient, resulting in limited spatial localization capabilities and low computational efficiency. To address these issues, this paper proposes a multi-level cross-layer linear fusion end-to-end

收稿日期:2024-03-15;修订日期:2024-05-01.

基金项目:中国科学院青年创新促进会(No.2020220)

Supported by Youth Innovation Promotion Association, Chinese Academy of Sciences (No.2020220)

*通信联系人, E-mail: yanghang@ciomp.ac.cn

encoding-decoding hybrid CNN-Transformer change detection model based on dilated spatial pyramid pooling, combining the advantages of visual Transformers and CNN. Firstly, image features are extracted using Siamese CNN, refined through dilated pyramid pooling to better capture detailed feature information. Secondly, the extracted attributes are converted into visual words, and a Transformer encoder models the compact visual words, feeding the learned context-rich labels back into visual space through a Transformer decoder to reinforce the original features. Thirdly, CNN features are fused with the features from Transformer encoding-decoding through skip connections, facilitating the fusion of position and semantic information by connecting features of different resolutions through upsampling. Finally, a difference enhancement module generates difference feature maps containing rich change information. Comprehensive experiments conducted on four publicly accessible remote sensing datasets, including LEVIR, CDD, DSIFN and WHUCD, confirm the efficacy of the proposed approach. Compared with other cutting-edge techniques for detecting changes, the model presented in this paper achieves superior classification performance, effectively addressing issues such as under-segmentation, over-segmentation and rough edge segmentation in change detection results.

Key words: remote sensing images; change detection; convolutional neural network; transformer; atrous spatial pyramid pooling

1 引 言

近年来,遥感技术取得了显著的进步,遥感卫星已具备同步观测地球大范围区域的能力,能快速捕捉地表状态信息,有效监测地表变化。这一技术的突破使得地球资源的利用更加高效合理^[1]。通过对比分析同一地区不同时间点的图像,我们能够精准识别场景变化^[2]。随着我国卫星发射次数的不断增加,大量遥感数据在太空中累积,这些宝贵的历史资料具有重要价值。变化检测作为对地观测领域的核心且极具挑战性的任务,在自然灾害评估、城市规划、资源管理、森林砍伐监测等领域得到了广泛应用,发挥着不可替代的作用^[3-8]。

随着卫星和机载传感器技术的持续升级,多时相遥感影像资料日益丰富。遥感图像中,空间纹理特征和场景语义信息更加精细,为地物数据的处理、解释和分析提供了强有力的技术支撑^[9]。如今,卫星遥感技术已实现了亚米级的高分辨率,使得对地物小目标的变化检测成为遥感影像领域的研究焦点^[10]。尽管高分辨率遥感影像提供了更为精细的信息,但同时也带来了海量的数据。这些影像虽几何结构鲜明、纹理复杂,但光谱信息相对匮乏,导致光谱波动、空间数据丢失以及双时态图像几何对准误差等问题频发,从而增加了遥感影像变化检测的难度^[11]。同时,

受到阳光照射、传感器性能及人为操作等因素的干扰,遥感影像数据质量参差不齐,变化检测的准确性面临严峻挑战,错误率居高不下^[12]。传统遥感影像变化检测方法的预处理过程繁琐,结果图易受到椒盐噪声的干扰,导致检测准确率低下且误差较大,实际应用效果受到较大限制^[13-14]。因此,提高遥感影像变化检测技术的精确度和稳健性至关重要。在大数据量和复杂图像特征背景下,我们亟需探索高效方法以应对光谱变化、空间信息丢失和配准误差等挑战,从而推动遥感影像变化检测技术的实际应用。

当前的变化检测研究主要聚焦于两个不同时期的影像中变化区域或者类型的识别。传统方法通常采用构建差异影像的方式以提取变化阈值,从而实现对变化区域的识别。常见方法包括变化向量分析法^[15](Change Vector Analysis, CVA)、主成分分析法^[16](Principal Component Analysis, PCA)、多元变化检测法^[17](Multivariate Alteration Detection, MAD)等。Malila 等人^[18]利用变化向量分析,通过计算像素强度和方向变化来区分变化与非变化区域。Mu 等人^[19]通过主成分分析融合差值图和比值图的特征,进而聚类得到差异检测图。Morton 等人^[20]利用 MAD 变化方法和正交线性回归的不变性,采用基于不变概率迭代重加权方法,对遥感图像变化进行检测。然而,这些传统的方法常依赖于手动选择阈值,

导致检测结果可能存在错检和漏检。此外,传统方法难以精准捕捉深层次的真实语义变化特征,因此在处理高精度变化检测任务时,其精确度和鲁棒性略显不足,面临严峻挑战。深度学习技术的蓬勃发展为变化检测算法领域带来了前所未有的创新。基于 CNN 的变化检测算法能够高效地将双时相遥感图像转化为特征图,并精准地提取出高度概括的变化区域语义特征,从而有效降低误差,避免预处理过程中可能产生的人为失误。基于变化检测独特的双时相特性,Zhan 等人^[21]创新性地引入了双分支结构的孪生卷积网络以更精准地进行变化检测。这种孪生结构由两个相同的 CNN 网络构成,能够高效地学习双时相图像之间的细微差异,从而精确提取出变化信息。因此,目前众多变化检测方法均将注意力放在孪生双分支网络架构上,致力于对该领域进行更为深入的探索与研究。Daudt 等人^[22]将孪生全卷积网络(Siamese Fully Convolutional Network, SFCN)成功应用于遥感影像变化检测的端到端任务中,同时创新性地提出了 3 种网络架构以进行高效的变化检测。其中,FC-EF(Fully Convolution-Early Fusion)架构通过早期拼接两张不同时刻的遥感图像进行输入,而 FC-Siam-conc 和 FC-Siam-diff 则通过提取特征后期融合的方式实现对遥感影像变化检测的高效处理。虽然上述变化检测方法均基于 CNN 网络,但卷积算子的局限性使其难以充分建模双时图像中的远程上下文信息,导致浅层难以学习有效特征。

为了应对上述问题并获取更为卓越的深度特征,设计更深、更复杂的 CNN 网络已成为主流趋势。然而,这同时也引发了参数爆炸的问题,给网络的训练和实际应用带来了挑战。为此,众多学者提出了各种改进模型的方法以获取更富判别性的特征表示。在对不同尺度的特征进行提取的深入探索中,DeepLab 系列网络^[23-24]设计了 ASPP(Atrous Spatial Pyramid Pooling)模块,巧妙运用了具有不同膨胀率的空洞卷积,有效拓宽模型的感受野,显著提升其对上下文信息的感知能力。PSPNet^[25]为了构建金字塔池化结构,采用了不同大小的池化模块,从而实现了多尺度特征的有效提取和处理以应对不同尺度地物的分割挑战。此外,基于双注意力机制的 DANet^[26]

和基于时空注意力神经网络 STANet^[27]通过提取输入影像的空间特征和通道权重,进一步强化了对图像细节特征的识别能力。Zhang 等人^[28]设计了一种名为 IFNet 的深度监督影像融合网络,它通过融合通道注意力和空间注意力机制,对双时相影像的特征进行了有效提炼。同时,该网络还采用了深度监督训练策略,显著提高了其性能和效果。Fang 等人^[29]提出了 SNUNet 网络,借助 UNet++ 的密集连接优势,有效减少了深层定位信息的丢失问题,通过整合通道注意力模块进一步提升了网络对多层次语义特征的识别能力。这些方法在提升变化检测精度方面取得了显著成果,然而,由于 CNN 的感受野受限于其卷积核的大小,其在时空维度上仍存在一定的局限性。尽管 CNN 注重提取深层变化的语义特征,但却常忽视了高分辨率浅层信息及其细颗粒特征的重要性,这导致了变化区域边缘像素的不确定性以及小物体的丢失问题。

Transformer 最初应用于自然语言处理(Natural Language Processing, NLP)^[30],其独特的自注意力机制显著扩展了模型的感知范围,进而强化了对变化检测特征的表征能力。由于其强大的表示能力,Transformer 目前在图像处理领域引起了越来越多的关注,例如图像分类^[31]、目标检测^[32]、语义分割^[33]等任务,Transformer 都展现出了卓越的性能,取得了显著的成果。相较于依赖局部算子提取信息的 CNN,Transformer 通过使用多头注意力机制来构建遥感图像全局关系模型,这使得它能够更加深入地洞察和理解数据可能存在的远程依赖性^[34]。由于其独特的多头自注意力机制,Transformer 巧妙地将二维图像任务变为一维序列任务,深入计算蕴含图像相关信息序列之间的相关性,从而有效地对图像的全局上下文进行建模。这种处理方式与 CNN 的局部处理策略存在显著区别,为图像处理带来了新的视角和可能性。ViT(Vision Transformer)是首个将 Transformer 架构引入视觉领域的模型,它将原始图像分割成一系列的图像块,并通过线性映射将这些图像块转换为一维向量,这些向量被输入到 Transformer 中,用于执行图像识别任务^[35]。Chen 等人^[36]提出的双时态图像变换器(Bitemporal Image Transformer, BIT)方法旨在建模时空域

内的上下文信息。然而,这种方法在处理过程中却未能充分重视高分辨率浅层信息的价值,且通过上采样操作获取更高分辨率变化图忽略了浅层细节,进而影响了变化检测的准确性。Feng 等人^[37]将自注意力(SelfAttention)和交叉注意力(CrossAttention)引导每个输入的全局特征分布,并通过串联和上采样操作对双时特征进行特征聚合。Lei 等人^[38]通过高斯核构建注意机制,逐步上采样差异特征,并使用跳跃连接完成特征聚合。需要注意的是,这些使用跳跃连接的方法缺乏多级特征交互的能力。Ying 等人^[39]提出一种新颖的差异引导多尺度聚合注意力网络 DG-MA²Net,该网络使用多尺度特征学习机制来融合双时态特征并生成差异特征。

为有效应对上述挑战,本文创新性地提出了一种基于 ASPP 多层级联线性融合的光学遥感图像变化检测方法。该方法在 Siam-Unet 网络架构的基础上进行了创新,通过引入 Transformer 对 CNN 特征图中的标记化图像块进行编码,从而实现对上下文信息的建模和全局上下文信息的提取。解码器对上采样后的特征进行处理,并通过跳跃连接与多尺度高分辨率特征融合,从而捕捉更丰富的局部与全局语义信息。最终,通过恢复特征图的全空间分辨率,实现了对变化区域的精确定位。本文的主要贡献如下:

(1)提出了一种端到端的网络架构模型,该模型通过结合 Transformer 结构和 Siamese CNN 网络,利用 Transformer 模块充分学习全局上下文并提取更具代表性的深层特征,减少伪变化的干扰。结合 Siamese CNN 能够充分学习浅层和深层特征,有效地处理多尺度图像特征,提高遥感图像变化识别的精确度。

(2)为了解决 CNN 感受野有限的问题,本文引入了 ASPP 空间空洞金字塔池化模块。该模块采用多分支结构,结合不同扩张率的膨胀卷积有效获取不同尺度下的特征信息,显著提升模型对复杂场景的感知与理解能力。

(3)针对输入噪声和干扰信息的问题,提出了一种差异增强模块(Difference Enhancement Module, DEM)。通过对变化图进行加权,获得更具辨别力的特征表示,生成包含丰富变化信息的差异变化图,提高模型的学习能力。

(4)针对级联或元素直接叠加的融合方式在特征表达能力上的局限性,本文提出了跨层级线性融合方法(Cross-Hierarchy Linear Fusion, CHLF)。该方法通过跳跃连接策略,巧妙地融合编码器生成的多尺度特征与上采样特征,既融入浅层细节丰富的特征,又结合深层语义丰富的特征,从而显著提升模型的特征表达能力,为后续的图像处理任务提供了更为精准和有效的特征支持。

实验结果表明,该模型能够更好地定位细节信息和纹理特征,获取到边界完整的变化区域,并实现更好的信息保留,在高分辨率和中分辨率遥感图像的变化检测上表现良好,尤其在建筑物、道路等地物变化检测方面具有优异的表现,能够很好地减弱季节变化的影响。在不同数据集集中进行实验验证,本文提出的模型在相关评价指标和检测精度优于对比模型。

2 本文提出的算法

2.1 孪生网络

孪生网络(图1)采用两个参数共享的主干网络作为特征提取器,旨在提取两张输入图片的特征图对。在训练过程中,通过将相似的图片 A 和 B 输入同一个网络,基于它们的差异(Label)建立损失,并进行反向传播训练。经过训练,孪生网络能够预测两张图片的输出特征图。这意味着越相似的输入图片,其特征图中相似处的向量矩阵越相似。在遥感图像处理中,可以充分利用这一特性进行变化检测等任务。特征提取器间的参数共享使模型能提取更一致的特征图对,同时大幅减少网络参数。值得一提的是,孪生网络作为一种灵活的架构设计方式,并不局限于特定的特征提取器。

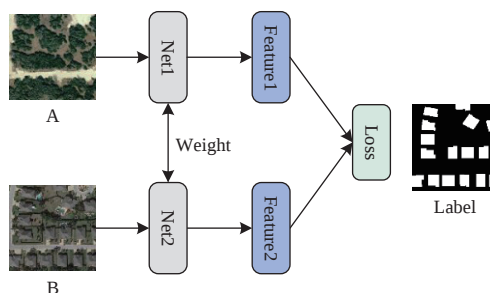


图1 孪生网络结构

Fig. 1 Siamese network structure

2.2 ASPP 模块的工作原理及改进

DeepLab 系列算法自 2014 年首次亮相以来便在语义分割领域一直受到广泛青睐。从 DeepLabV1 到后续的 V2、V3 以及 V3+, 研究者们不断优化与改进, 推动算法技术的革新, 并在开源数据集上取得了显著的性能提升^[23-24]。传统方法因固定尺寸的下采样导致感受野受限, 影响了分割精度。DeepLabV1 引入了空洞卷积, 通过调整空洞率扩大感受野, 从而增加了模型对图像局部和全局信息的理解能力。DeepLabV2 借鉴 SPPNet^[40] 思想提出了 ASPP 结构, 利用不同尺度的空洞卷积扩展多尺度感受野, 从而有效地融合了不同尺度下的语义信息。DeepLabV3 在 V2 的基础上进一步优化了空间金字塔的下采样结构, 引入 ASPP 模块在编码阶段的底层, 更好地捕获了图像的上下文和全局信息, 进一步提升了多尺度特征融合的效果。

本文没有采用常规的不同扩张率(Rate)的空洞卷积(Rate=6, 12, 18), 而是根据遥感图像的特性, 尝试了扩张率为 12, 24, 36 的空洞卷积, 在实验数据集上达到了较好的效果, 提高了模型对不同尺度变化区域的分割能力以及分割的准确性, 为 Transformer 模型的编码阶段提供了更丰富的视觉语义信息。这种丰富的语义信息对

于许多计算机视觉任务都具有重要意义, 如图像分类、目标检测等。这一系列改进使 DeepLab 系列网络(图 2)能够更好地应对不同尺度变化的图像区域, 为语义分割领域注入了新的活力, 推动了更多相关研究的涌现和进步。

2.3 本文网络结构

为了更好地适应各种遥感图像类型, 本文提出了一种基于 ASPP 多层次级联线性融合的光学遥感图像变化检测模型。该模型主要分为 4 个部分: CNN 主干编码网络、Transformer 模块、跨层级联线性融合模块和差异增强模块。在特征提取阶段, 本文利用高效的 ResNet18 网络来精准地抽取图像特征。同时, 借助 ASPP 模块有效地拓展了编码模块的感受野, 从而增强了模型对多尺度信息的感知能力。随后, 通过引入 Transformer 模块得到图像的全局上下文信息, 生成更为细致和丰富的特征表示。最后, 通过跨层级联线性融合与差异增强模块, 巧妙地将不同阶段的特征信息整合起来, 实现了对变化区域的精确分类与预测。模型结构如图 3 所示。

2.3.1 CNN 主干编码网络

孪生 ResNet18 网络凭借其独特的局部感知和参数共享机制展现出卓越的图像处理能力。它能够精准地提取不同时刻遥感图像的细节信息, 实现对小目标的精确检测, 并在像素级别上实现精细定位, 为遥感影像变化检测提供了强大的技术支持。如图 3 所示, 本文选取以 ResNet18 为主干网络的前 3 层, 分别采用下采样因子 2、4 和 8。通过这样的设置, 得到了 3 个不同大小的特征图, 分别为 $f_c^1 \in R^{\frac{H_0}{2} \times \frac{W_0}{2} \times C_1}$ 、 $f_c^2 \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times C_2}$ 、 $f_c^3 \in R^{\frac{H_0}{8} \times \frac{W_0}{8} \times C_3}$ 。其中 H_0 、 W_0 代表输入图像的高度和宽度; C 为通道数, C_1 、 C_2 、 C_3 分别为 64、128、256。随后, 在特征提取网络后加入 ASPP 模块, 并引入一个双线性插值来增加特征图的尺寸。这一设计的目的是防止特征图过小, 从而避免空间信息的丢失并增强感受野。接着, 通过 1×1 卷积进行降维, 最终得到大小为 $f_c \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times \frac{C_1}{2}}$ 的特征图。这些特征图随后被作为 Transformer 编码器的输入, 以进一步提取和利用其蕴含的高维信息。这一流程有助于提高对图像信息的全局把握能力。

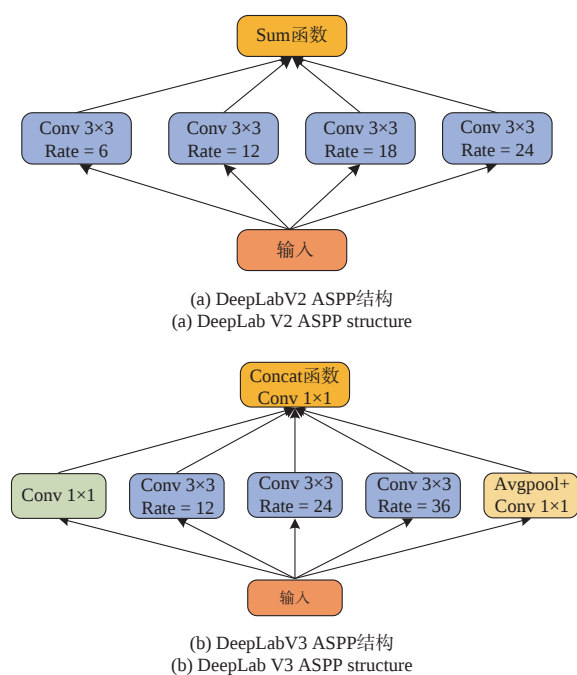


图 2 DeeLab 系列网络

Fig. 2 DeeLab series networks

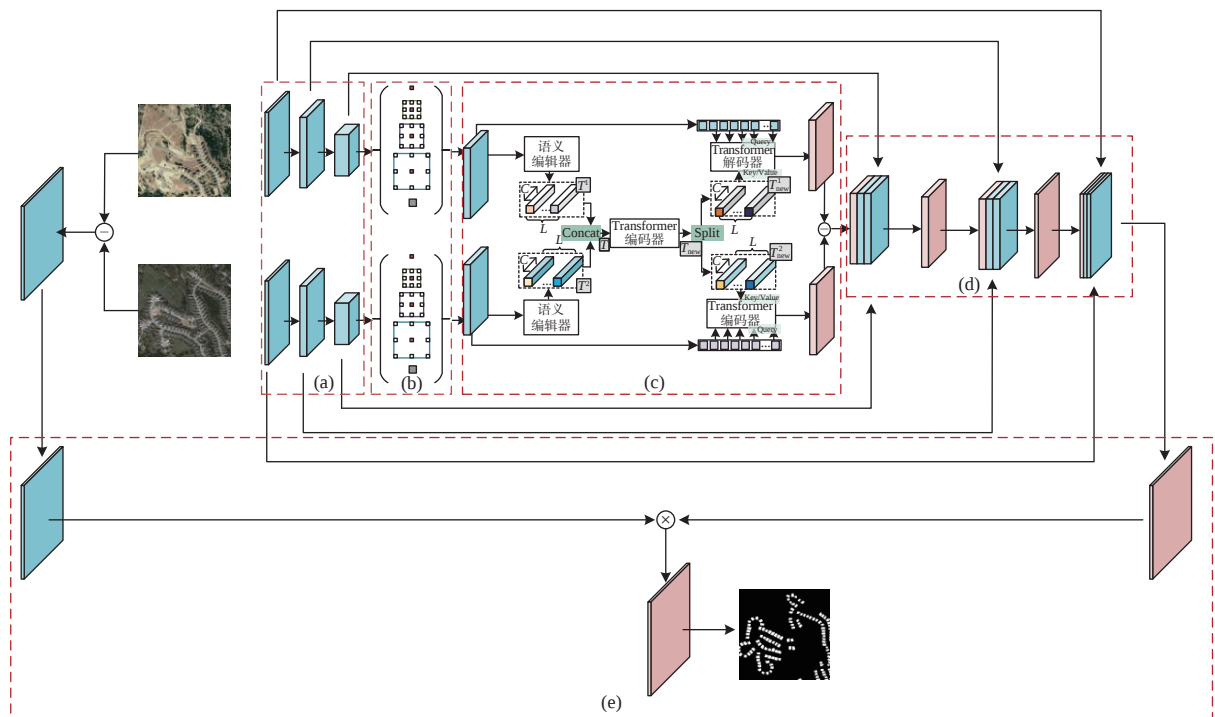


图3 基于CNN-Transformer结构的遥感图像变化检测模型。(a) Resnet18改进结构;(b) ASPP模块;(c) Transformer结构;(d) CHLF特征融合结构;(e) DEM差异增强模块。

Fig. 3 Remote sensing image change detection model based on CNN-Transformer structure. (a) Resnet18 improved structure; (b) ASPP module; (c) Transformer structure; (d) CHLF feature fusion structure; (e) DEM difference enhancement module.

2.3.2 Transformer编解码模块

如图4所示,整个Transformer编解码结构大体可以分为3个部分:语义编译器、编码器和解码器。

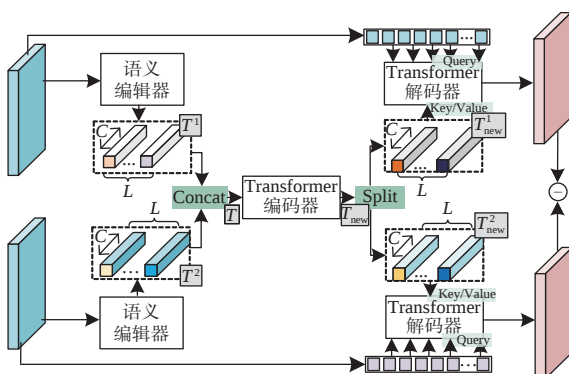


图4 Transformer结构

Fig. 4 Transformer structure

语义编译器将主干网络提取的图像特征图转化为富含语义信息的语义令牌(Token)表达形式,这一做法借鉴了NLP中将句子元素化并用

token表达的思路。在图像分类任务中,这些token能够有效地揭示图像间的类别差异及变化部分。通过将特征图分解为子图并映射为token,结合空间注意力机制,能够精准提取图像的关键空间信息,使token既紧凑又富含语义。这种方法显著提升了语义信息的准确性和维度,对图像处理具有显著价值。

Transformer的编码器和解码器结构在孪生型语义编译网络中发挥了关键作用。如图5所示,编码器接收拼接后的token进行特征提取,解码器则利用孪生网络思想处理。在解码过程中,token作为键(Key)-值(Value)输入,原始特征图作为查询(Query)输入,通过Transformer的注意力机制相互作用,实现特征图的翻译。这种结构生成的两张新特征图具有较强的区分度,凸显了图像间的异同。整合孪生网络和Transformer的模型结构提供了更灵活、强大的表达能力,能够深入捕捉图像语义信息,生成具有明显差异的特征图,有效区分图像的语义和结构差异。

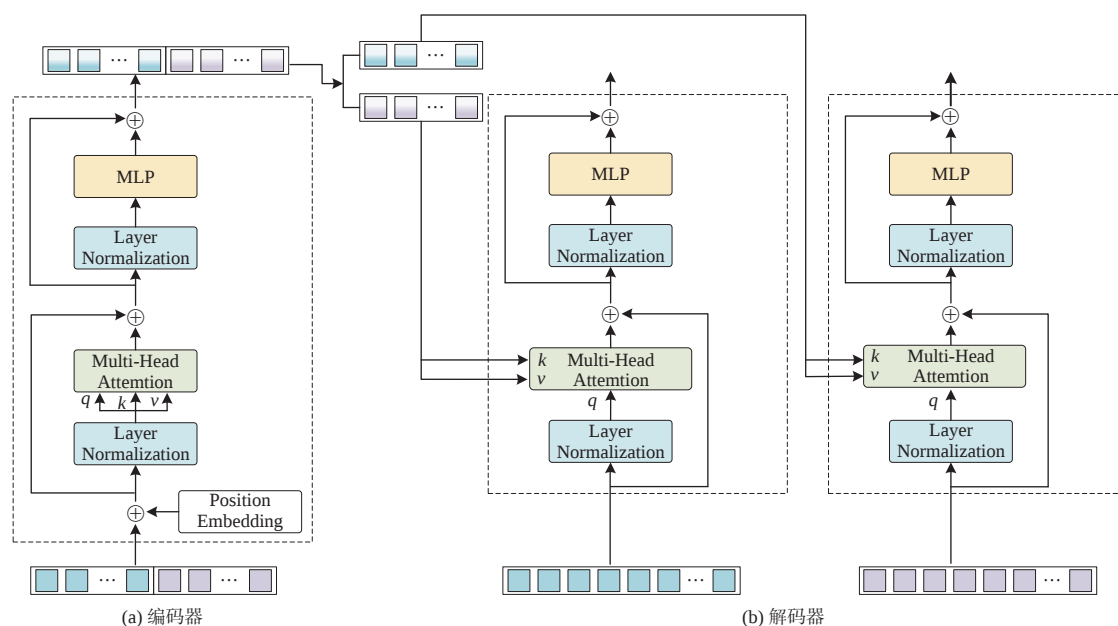


图 5 Transformer 的编码器 (a) 和解码器 (b) 结构

Fig. 5 Structure of Transformer encoder (a) and decoder (b)

2.3.3 跨层级联线性融合

在基于深度学习的算法模型中,浅层特征图通常捕获较少的抽象特征,更接近于原始图像的表达;而深层特征图则包含更多抽象特征,具有更高的语义信息表示能力。这意味着,浅层特征更适合预测小目标,而深层特征更适合预测大目标。然而,目前大多数变化检测算法模型的特征融合策略局限于简单的高级和低级特征拼接,或者是高级特征辅助低级特征的训练并选择性地融合低级特征。这些策略虽然考虑了不同层次的语义特征信息,但通常忽视了中间特征层,而这些层的信息在特征提取过程中经常与其他层信息相融合后重新传递给操作流,未能直接体现在最终的判别特征中,导致模型未能充分利用所有可用特征信息,从而限制了其性能和准确性。在遥感图像变化检测任务中,简单地将编码器得到的特征上采样到全分辨率并不能获得最佳效果,因为编码得到的特征图通常比原始图像分辨率要小得多,导致了浅层细节特征的丢失。为了解决这一问题,本文提出了跨层级联线性融合的方法(Cross Hierarchy Linear Fusion, CHLF),旨在精确定位变化目标并提高检测效果。

如图 6 所示,受 UNet 设计的启发,CHLF 采用跳跃连接的方式,将 CNN 网络生成的多尺度

特征与上采样特征融合,有效降低了由于下采样操作而导致的空间信息损失。CHLF 通过多个级联上采样块对编码特征进行解码。首先对编码器输出的特征图进行上采样操作,然后与 Transformer 解码器生成的特征图拼接,实现信息融合。最后采用级联上采样块恢复空间分辨率。每个上采样块包含一个 2 倍双线性插值上采样算子和一个卷积块,其中卷积块由两个 3×3 卷积层、一个 BN 层和一个 ReLU 激活层构成。通过这种 CHLF 的设计,变化检测转换器能够更有效地保留并恢复图像的空间信息,尤其在处理浅层细节时,取得了更为令人满意的检测结果。

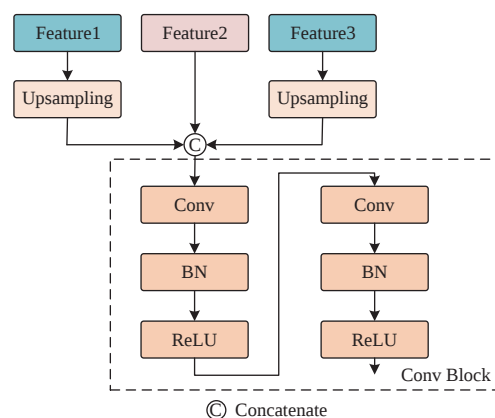


图 6 跨层级联线性融合

Fig. 6 Cross-hierarchy linear fusion

2.3.4 差异增强模块

尽管双时遥感图像的特征图能够清晰显示图像的整体变化,但由于其采用无差别像素计算,忽略了图像局部特征的差异,并且受到辐射和地理误差的影响,因此为了提高双时相遥感图像变化检测的准确性,本文设计了差异增强模块(Difference Enhancement Module, DEM)。该模块将处理后的图像特征图与卷积层相结合,将提取的差异特征图用于对原始双时相图像提取的特征进行加权相乘,从而得到最终的遥感变化检测结果。差异增强模块能够对不同时间差异图像进行建模,将差分图像映射到新的特征空间,充分挖掘变化信息并提取变化密度图并过滤噪声。通过引入该模块,网络能够额外关注双时图像的差异化特征,对特征图的每个像素进行加权,自动进行有选择的聚合特征,提高了网络的有效性和变化特征的提取能力。

如图7所示,差异增强模块由两个卷积单元和一个空间注意力(Spatial Attention, SA)模块组成。SA模块通过学习差异图像的特征图,有效地区分深层特征中变化的物体和背景。具体来说,SA模块能够在特征图上产生权重,以便网络能够更加关注重要的局部特征,同时抑制无关的背景信息。这种机制使得特征增强模块能够为深度学习网络提供更为准确和有针对性的权重,从而更好地捕捉局部特征的变化。通过引入DEM,增强了网络对于变化目标和背景的感知

能力,使网络能够更加准确地捕捉到局部特征的变化,针对性地学习和理解输入图像的变化信息,从而提高了网络整体的准确性和鲁棒性。

2.4 损失函数

遥感图像变化检测任务的特点是发生变化的对象是少数,变化的区域远小于未变化的区域,极大地影响了变化检测的精度。在训练阶段,本文使用二元交叉熵损失函数优化网络参数,用交叉熵评估当前训练得到的概率分布与正式分布的差异,如公式(1)所示:

$$L_{\text{bce}} = -\frac{1}{N} \sum_{n=1}^N (q_n \log(p_n) + (1 - q_n) \log(1 - p_n)), \quad (1)$$

其中: N 表示图像中的样本个数,即像素个数; q_n 表示像素 n 的真实值,如果像素属于变化类, q_n 为1,否则 q_n 为0; p_n 和 $1 - p_n$ 分别表示经模型预测后像素 n 属于变化类和不变化类的概率。

3 实验与分析

3.1 数据集

深度学习网络结构具备高度灵活性,能够自动提取图像中的特征,展现出强大的数据处理能力。相较于传统的变化检测方法,深度学习在性能和实际应用范围上具有显著的优势,它适用于高分辨率光学遥感图像变化检测,能精准识别和定位变化区域。但深度学习网络的检测效果受数据质量、数量和场景多样性的影响。深度学习模型的训练与优化离不开充足的数据支持,其中数据的质量和数量对于模型的鲁棒性和泛化能力具有至关重要的影响。因此,本文选择以下4个不同的变化检测数据集进行实验,以验证算法在不同分辨率、不同场景、季节变化、光谱变异和信息丢失下的鲁棒性。

(1) WHU建筑变化检测数据集(WHUCD)^[41]。WHUCD数据集包含新西兰克赖斯特彻奇地区的建筑数据,由一对高分辨率(0.2 m)航拍图像组成,分辨率为32 507像素×15 354像素。该数据集覆盖基督城,仅包含建筑物,不包括其他人造结构。实验将数据划分为256像素×256像素块,分别随机划分为5 947个训练集、743个验证集和744个测试集。

(2) CDD合成图像和真实季节变化检测数

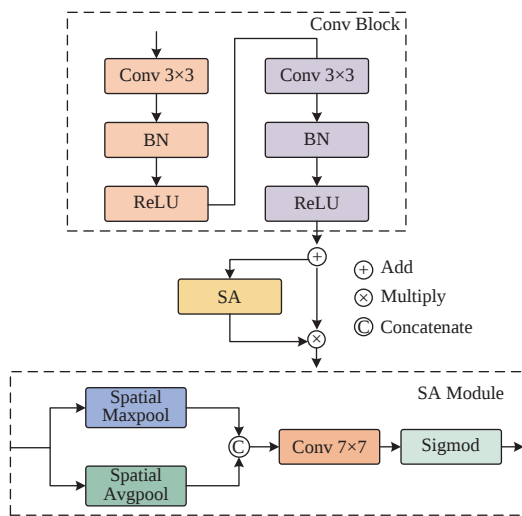


图7 差异增强模块

Fig. 7 Difference enhancement module

据集(CDD)^[42]。CDD数据集是 Google Earth 图像的季节变化数据集,包含 7 对 4 725 像素 $\times$ 2 700 像素和 4 对 1 900 像素 $\times$ 1 000 像素的季节变化图像。数据集的空间分辨率为 0.03~1 m/像素。该数据集考虑了不同对象的变化,同时忽略了季节差异和亮度带来的变化,对 CD 算法提出了挑战。图像的大小为 256 像素 $\times$ 256 像素,其中包括 10 000 个训练样本、3 000 个验证样本以及 3 000 个测试样本。

(3)LEVIR 建筑变化检测数据集(LEVIR)^[27]。LEVIR 数据集包含美国得克萨斯州不同地区的建筑数据,由 637 对高分辨率(0.5 m)图像构成,每幅图像均拥有 1 024 像素 $\times$ 1 024 像素的精细分辨率。数据集聚焦于土地利用变化,尤其关注建筑增长与衰落现象。其涵盖的建筑类型丰富多样,包括别墅住宅、高层公寓、小型车库及大型仓库等。数据集划分为 256 像素 $\times$ 256 像素块,分别获得 7 120 个训练集、1 024 个验证集和 2 048 个测试集。

(4)DSIFN 变化检测数据集(DSIFN)^[28]。DSIFN 数据集由中国 6 个主要城市的 6 个大型共配双时态图像对组成,空间分辨率为 2 m/像素。该数据集包括道路、建筑物、农田和水等变化,考虑了来自 6 个城市的不同亮度和对比度图像以及不同的陆地物体,使得变化检测变得更为复杂。数据集划分为 256 像素 $\times$ 256 像素块,分别获得 14 400 个训练集、1 360 个验证集和 192 个测试集。

如图 8 所示,通过使用这些不同的数据集,能

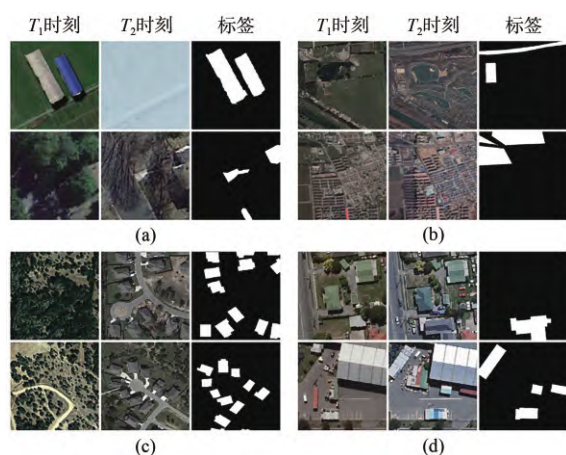


图 8 实验数据集。(a) CDD; (b) DSIFN; (c) LEVIR; (d) WHUCD。

Fig. 8 Experimental Datasets. (a) CDD; (b) DSIFN; (c) LEVIR; (d) WHUCD.

够全面验证算法在处理不同场景和变化类型时的鲁棒性和泛化能力,确保模型对于各种复杂情况都能够表现出色。

3.2 对比方法

为验证所提方法的优越性,实验对比了 8 种先进方法,分别是 FC-Siam-conc^[22]、FC-Siam-diff^[22]、STANet^[27]、SNUNet^[29]、BIT^[36]、DMINet^[37]、USFFC-Net^[38]、DGMA²-Net^[39]。

(1)FC-Siam-conc(2018)利用共享参数的分支结构处理图像,通过卷积层合并输出,从而充分利用两个输入图像的信息并保持高效计算。

(2)FC-Siam-diff(2018)通过计算特征图差异并结合跳跃连接,关注图像间的差异信息,提升目标跟踪等任务的准确性,并保留空间细节信息以增强性能。

(3)STANet(2022)通过引入时空注意力机制,增强了模型对关键信息和区域的关注能力,提高了模型在处理序列或动态数据时的性能。

(4)SNUNet(2022)为孪生密集连接网络,将密集连接结构的编码器设计为孪生网络结构,并引入集成通道注意力模块以聚合和优化解码器的 4 个特征输出。

(5)BIT(2022)为基于 Transformer 的网络,通过实施后期融合策略,打造出一种高效且精准的遥感影像变化检测方法。

(6)DMINet(2023)将自注意力机制和跨注意力机制统一在单个模块中,以引导全局特征分布,使用减法和串联以及多级差异聚合来检测变化。

(7)USFFC-Net(2023)是一个轻量级模型,实现了一种新的多尺度特征提取方法,并通过高斯核构建了注意力机制。

(8)DGMA²-Net(2024)是一种将双时态特征的语义信息融合到差异特征中的新方法。

3.3 评价标准

评价指标在遥感图像变化检测网络的性能衡量中发挥着至关重要的作用。通过对比检测结果与真实情况的数值,定量评估变化检测方法的优劣。由于变化检测结果通常为二值图的形式,因此本文选择采用混淆矩阵来评估遥感图像变化检测网络的性能。根据混淆矩阵中的数值指标,可以计算出各类评价指标的结果。在遥感

图像变化检测中,通常采用精确度(Precision, Pre)、召回率(Recall, Rec)、F1系数(F1-score)、总体准确率(Overall Accuracy, OA)和交并比(Intersection over Union, IoU)作为评价指标。为了更好地介绍各类评价指标的含义,本文首先对混淆矩阵进行介绍,具体见表1。

表1 混淆矩阵
Tab. 1 Confusion matrix

真实情况	预测情况	
	正例	负例
正例	TP(TruePositive)	FN(FalseNegative)
负例	FP(False Positive)	TN(True Negative)

(1)精确度(Pre):查准率高低体现了模型在识别正类样本时的准确性,数值越高,说明模型性能越优异。精确度可用式(2)计算:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (2)$$

(2)召回率(Rec):召回率是实际正样本中模型正确识别的比例,值越高说明模型预测正类样本越准确,效果越优。召回率可用式(3)计算:

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (3)$$

(3)F1系数(F1-score):精确度和召回率的综合评价指标,通过计算两者的调和平均数得出。其值介于0~1之间,其中1表示最佳性能,0表示最差性能。F1系数可用式(4)计算:

$$\text{F1} = \frac{2 \times \text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}}. \quad (4)$$

(4)总体准确率(OA):精确率作为衡量分类性能的核心指标,能够精准地展现模型的识别能力,准确反映模型识别能力。精度高则效果好,模型性能更卓越。总体准确率可用式(5)计算:

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (5)$$

(5)交并比(IoU):用于评估真实值与预测值之间的相关程度,随着相关度的提升,该标准值也会相应增高,表明预测结果更为接近真实情况。交并比可用式(6)计算:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}. \quad (6)$$

3.4 训练参数及环境配置

本文实验配置如表2所示。实验采用Pytorch

深度学习框架与Pycharm工具对模型进行详尽的训练与测试。在此过程中,本文针对256×256尺寸的图像块应用了常规的数据增强技术,包括翻转、缩放、裁剪以及高斯模糊等多种手段,以提升模型的泛化能力。本实验模型优化采用了带有动量的随机梯度下降(SGD)算法,其中动量设置为0.90,权重衰减参数为0.0005,以提高模型的收敛速度和稳定性。学习率最初设置为0.01并线性衰减到0,直到训练200个epoch。使用二元交叉熵损失函数进行训练,在每个训练周期后进行验证,并使用验证集上的最佳模型对测试集进行评估。

表2 实验配置
Tab. 2 Experimental configuration

Parameter	Configuration
OS	Windows 11
CPU	Inter i9-13900K
RAM	128G
GPU	NVIDIA GeForce RTX 4090
Language	Python 3.9.13
GPU Acceleration	Acceleration CUDA 11.8
CUDNN Version	V8.9.0

3.5 数据分析

3.5.1 对比实验

结果可视化比较使用本文变化检测模型检测WHUCD、CDD、LEVIR和DSIFN四个数据集的变化,并将结果与文献中选择的7种方法进行比较。4个数据集的可视化比较如图9~12所示。图中用不同颜色表示检测结果:TP(白色)、TN(黑色)、FP(红色)、FN(绿色)。

在季节变化数据集CDD中,本文提出的模型相较于其他模型取得了更为优越的结果,并且在适应季节变化所带来的影响方面表现更为出色。实验选择在该数据集上分析建筑物和道路变化。在图9的建筑物变化结果对比中,本文模型即使在极端天气下也能完整地提取变化的建筑物,并且变化的建筑物误检和漏检最少。此外,本文提出的模型面对对象大小的变化都能保持其内部结构的紧凑性和边界的完整性,展现出出色的鲁棒性。相较之下,FC-Nets(包含了FC-Siam-conc、

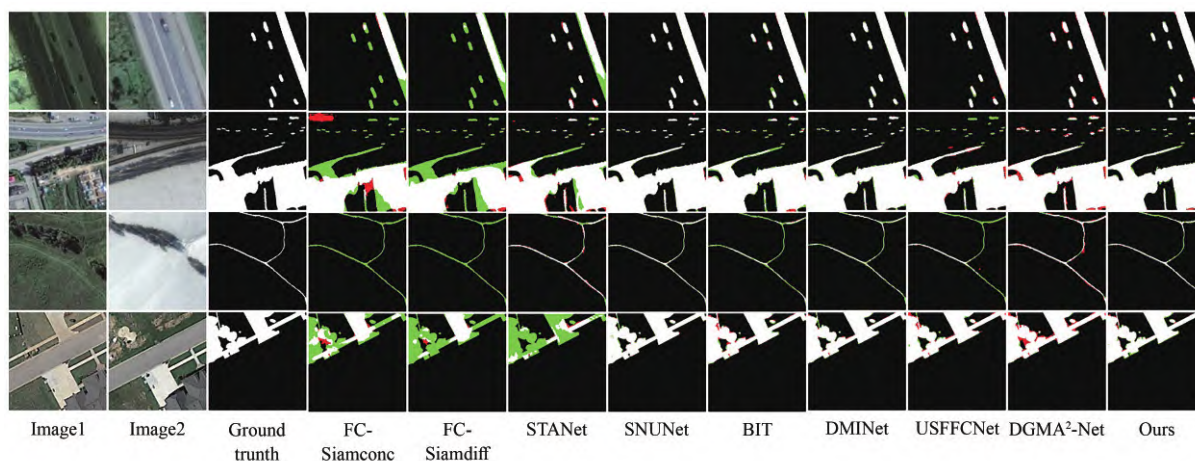


图9 不同方法在CDD数据集上的结果可视化:TP(白色)、TN(黑色)、FP(红色)、FN(绿色)。

Fig. 9 Visualization results of different methods for CDD datasets: TP(white), TN(black), FP(red), FN(green).

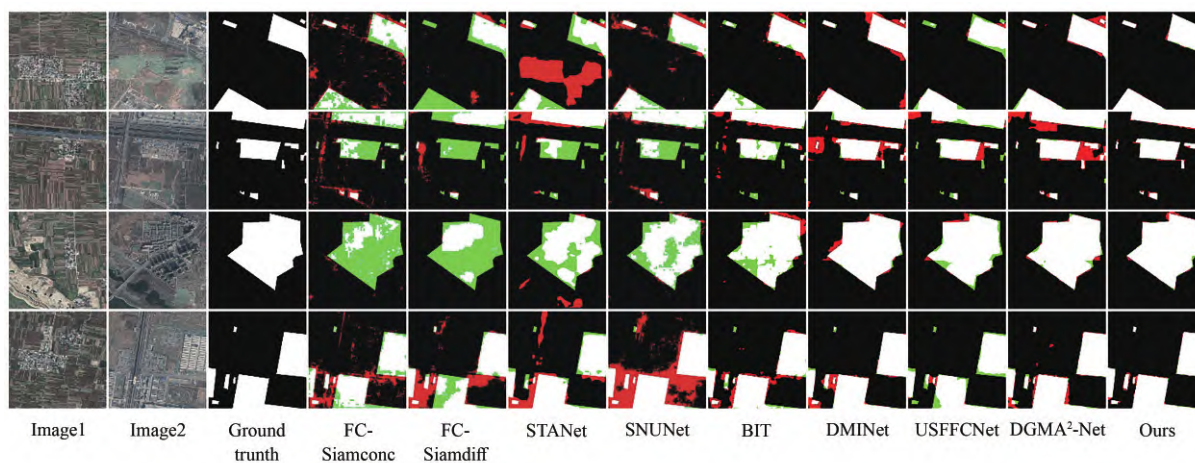


图10 不同方法在DSIFN数据集上的结果可视化:TP(白色)、TN(黑色)、FP(红色)、FN(绿色)。

Fig. 10 Visualization results of different methods for DSIFN datasets: TP(white), TN(black), FP(red), FN(green).

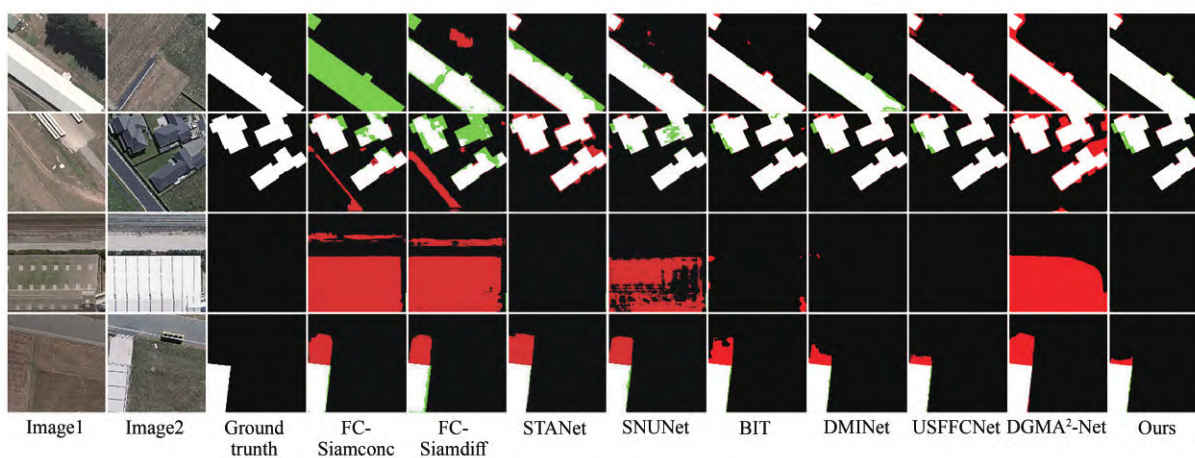


图11 不同方法在WHUCD数据集上的结果可视化:TP(白色)、TN(黑色)、FP(红色)、FN(绿色)。

Fig. 11 Visualization results of different methods for WHUCD datasets: TP(white), TN(black), FP(red), FN(green).

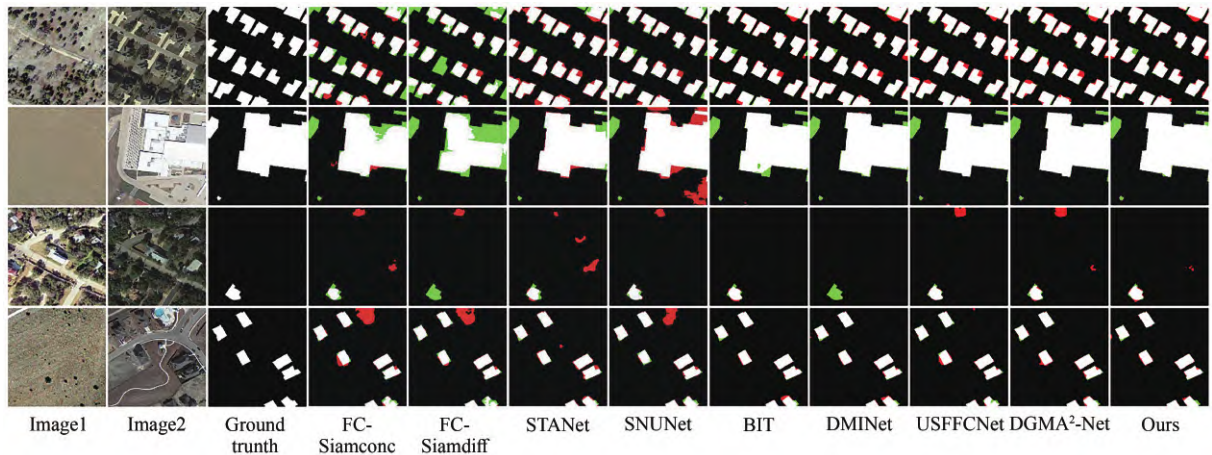


图 12 不同方法在 LEVIR 数据集上的结果可视化: TP(白色)、TN(黑色)、FP(红色)、FN(绿色)。

Fig. 12 Visualization results of different methods for LEVIR datasets: TP(white), TN(black), FP(red), FN(green).

FC-Siam-diff 模型)和 STANet 则破坏了大变化区域的内部结构,且其边界比真实标记的边界要大。这是因为双时态图像被串联起来作为网络的输入,并且图像差异辨别从网络的开头开始。尽管跳跃连接结合了中间特征图像,但很难提供原始图像的深层特征来帮助图像重建。DGMA²-Net 可以提取变化区域,但误检过多,复杂建筑物边界受损最为严重。SNUNet 模型在维持建筑物变化的边界和内部完整性方面表现优异,但直接将结果上采样至全分辨率,使变化区域的细节信息丢失,难以捕捉细微处变化。对于道路变化,仅有本文模型可以检测到狭窄道路的变化, BIT、DMINet 和 USFFC-Net 模型都对道路结构变化造成严重损失。本文模型获得的变化图最接近真实的道路变化结果,并且变化区域完整且准确,很好地抑制了伪变化。

在不同城市的变化数据集 DSIFN 中,除了本文提出的模型外,没有一个模型能够完全保持变化区域的边界,这主要是由于不同城市之间的图像差异较大。在图 10 中,与其他模型相比,本文模型在变化区域中的边缘和细节表现最佳。这是因为 Transformer 模块通过学习上下文信息,有效减少了伪变化现象,进而更精准地表达了真实的语义变化。然而,FC-Nets、STANet 和 SNUNet 等比较方法无法完全检测到许多变化,而且 BIT、DMINet 和 DMINet 通常对噪声敏感,导致检测结果出现较多伪变化。DGMA²-Net 虽然能检测到完整的变化区域,但是误检率较高。在对比实

验的可视化结果中,本文模型展现出对变化区域的完整检测能力,并能有效应对伪变化、错检和漏检等问题,充分验证了所提算法的优越性。

在建筑变化数据集 WHUCD 和 LEVIR 中,除了 FC-Nets 之外,大多数模型都可以检测到较大的建筑变化。其中,STANet 和 DMINet 产生缺失边缘,SNUNet、USFFC-Net 和 DGMA²-Net 遭受误报和复杂建筑物错检的困扰,而 BIT 和本文模型可以准确地检测变化并削弱噪声的影响。对于小的建筑变化,只有本文模型几乎可以区分所有小的变化对象。这是因为 FC-Nets、STANet 和 SNUNet 模型受到感受野的限制,而 BIT 模型则受到低层损失的限制,直接上采样中的细节导致伪变化和边界丢失。如图 12 所示,除了本文模型之外,其他模型都没有检测到 LEVIR 中白色小建筑物的变化。

表 3~表 6 比较了本文所提出模型和对比模型的定量验证结果。在 CDD、DSIFN、WHUCD 和 LEVIR 数据集中,所提模型在所有 4 个指标中都达到了最佳水平。在 CDD 数据集中准确度、F1 分数、IoU、精确度和召回率分别为 99.11%、92.75%、96.24%、96.43% 和 96.03%,优于同类型模型(BIT)0.56%、4.5%、3.83% 和 0.88%。在 DSIFN 数据集中,虽然所有方法的检测精度都不高,但本文模型可以检测更多变化的像素,这在精度和召回率之间做出了很好的权衡。在 LEVIR 和 WHUCD 建筑数据集中,本文模型比同类型模型(BIT)分别提高了 0.16%、2.61%、

1.57%、2.74%、0.43% 和 0.48%、9.36%、5.98%、11.28%、0.7%。在各种数据集的定量比较中,本文模型表现出了最好的性能,边界清晰,内部紧凑性高,漏检和误检较少。

表 3 CDD 数据集对比实验结果

Tab. 3 Comparison results of the methods on CDD dataset

%					
Methods	OA	IoU	F1	Pre	Rec
FC-Siam-conc	96.36	72.77	84.24	86.21	82.36
FC-Siam-diff	95.76	68.79	81.51	83.88	79.27
STANet	97.80	83.60	91.10	88.00	94.30
SNUNet	98.89	91.01	95.29	95.45	95.14
BIT	98.55	88.25	93.76	92.41	95.15
DMINet	98.78	72.36	83.97	90.18	78.55
USFFC-Net	98.66	89.24	94.31	94.37	94.26
DGMA ² -Net	98.39	87.79	93.50	89.36	98.04
Ours	99.11	92.75	96.24	96.43	96.03

表 4 DSIFN 数据集对比实验结果

Tab. 4 Comparison results of the methods on DSIFN dataset

%					
Methods	OA	IoU	F1	Pre	Rec
FC-Siam-conc	87.57	42.56	59.71	66.45	54.21
FC-Siam-diff	86.63	45.50	62.54	59.67	65.71
STANet	88.49	47.66	64.56	67.71	61.68
SNUNet	87.34	49.45	66.18	60.60	72.89
BIT	89.41	52.97	69.26	68.36	70.18
DMINet	94.45	73.46	84.70	79.62	90.47
USFFC-Net	93.23	64.49	78.41	85.67	72.29
DGMA ² -Net	93.33	69.66	82.12	75.42	90.12
Ours	96.71	82.65	90.50	88.79	92.27

表 5 LEVIR 数据集对比实验结果

Tab. 5 Comparison results of the methods on LEVIR dataset

%					
Methods	OA	IoU	F1	Pre	Rec
FC-Siam-conc	98.49	71.96	83.69	91.99	76.77
FC-Siam-diff	98.67	75.92	86.31	89.53	83.31
STANet	98.66	77.40	87.26	83.81	91.00
SNUNet	98.82	78.83	88.16	89.18	87.17
BIT	98.92	80.68	89.31	89.24	89.37
DMINet	98.97	82.11	90.18	87.74	92.75
USFFC-Net	99.04	82.54	90.44	89.37	91.53
DGMA ² -Net	99.03	82.29	90.29	91.82	88.81
Ours	99.08	83.29	90.88	91.98	89.80

表 6 WHUCD 数据集对比实验结果

Tab. 6 Comparison results of the methods on WHUCD dataset

%					
Methods	OA	IoU	F1	Pre	Rec
FC-Siam-conc	97.04	49.95	66.63	60.88	73.58
FC-Siam-diff	95.63	41.66	58.81	47.33	77.66
STANet	98.52	69.95	82.32	79.37	85.50
SNUNet	98.71	71.67	83.50	85.60	81.49
BIT	98.75	72.39	83.98	81.48	86.64
DMINet	98.97	79.68	88.69	93.84	86.25
USFFC-Net	98.93	80.49	89.19	88.67	85.23
DGMA ² -Net	97.93	67.20	80.38	68.04	80.17
Ours	99.23	81.75	89.96	92.76	87.34

3.5.2 消融实验

为验证模块的有效性,本文在多个数据集上进行了消融实验,深入探究了不同模块组合的应用效果。基础网络采用 Siam-Unet 编解码结构并省略了 ASPP(A)、DEM(D)和 Transformer(T)模块以简化网络。评估模型性能时,采用 F1 分数作为综合评价指标,其值越高表明模型性能越卓越,能更精准地反映整体表现。实验结果表明,所提模块对变化检测算法具有有效性和鲁棒性,为实际应用提供了有力支持。

如表 7 所示,在 DSIFN 数据集上的实验结果充分验证了所提出网络模块的有效性。通过引入 ASPP 模块,显著扩大了网络感受野,增强了捕获多尺度上下文信息的能力,使模型的 F1 分数相较于 Base 网络提升了 4.26%。Transformer 模块的加入,通过学习全局上下文并提取深层特征,进一步提升了模型性能,F1 分数相比 Base 网络提高了 2.23%。DEM 模块的引入则显著提高了模型精度并加速收敛过程,使得 F1 分数提升了 0.64%。此外,在 ASPP 基础上添加 DEM 或 Transformer 模块,F1 分数分别提升了 4.38% 和 3.9%。最终,3 个模块组合相较于 Base 网络,F1 分数高出 6.31%,充分证明了所提网络模块及其组合的优越性。这些改进不仅提升了模型性能,也为遥感影像变化检测提供了更可靠的技术支持。

在 LEVIR 数据集上的实验结果如表 8 所示。通过引入 ASPP 模块,成功地将 F1 分数相较于基础网络 Base 提升了 0.18%。DEM 模块的加入也展现出显著效果,使得 F1 分数提升了 0.19%。尽管单独使用 Transformer 模块时性能略有下降,但

表 7 DSIFN数据集消融实验结果

Tab. 7 Ablation experiment results of the methods on DSIFN dataset %

模型	ASPP	DEM	Transformer	OA	IoU	F1	Pre	Rec
Base	×	×	×	95.00	72.71	84.19	90.84	78.46
Base+A	√	×	×	96.12	79.29	88.45	89.43	87.49
Base+D	×	√	×	95.08	73.66	84.83	88.99	81.05
Base+T	×	×	√	95.55	76.09	86.42	89.71	83.36
Base+A+D	√	√	×	96.14	79.49	88.57	89.21	87.94
Base+A+T	√	×	√	95.84	78.71	88.09	85.80	90.50
Base+D+T	×	√	√	95.36	75.63	86.12	87.61	84.68
Ours	√	√	√	96.71	82.65	90.50	88.79	92.27

与其他模块结合使用时效果卓越。进一步的实验表明,ASPP与DEM或Transformer的组合分别使F1分数提升0.18%和0.11%。最终,3个模块共同使用时效果最为显著,F1分数相较于基础网络Base提高了3.28%。这一成果充分验证了我们所提网络模块的有效性及其组合使用的优越性能,为遥感影像变化检测技术的发展提供了有力支持。

在WHUCD数据集上的实验结果如表9所示。通过在基础网络上集成ASPP模块,成功地

将F1分数相较于基础网络Base提高了1.29%。此外,使用DEM模块也取得了显著成效,使得F1分数相较于基础网络Base提升了2.49%。在已经应用ASPP模块的基础上,我们进一步探索了与DEM和Transformer模块的组合效果。实验结果显示,这两种组合分别使F1分数提升了1.8%和0.89%。值得一提的是,在Transformer模块的基础上,在最后一层特征融合中增加了DEM模块,这一举措显著加强了变化对象的完

表 8 LEVIR数据集消融实验结果

Tab. 8 Ablation experiment results of the methods on LEVIR dataset %

模型	ASPP	DEM	Transformer	OA	IoU	F1	Pre	Rec
Base	×	×	×	99.05	82.66	90.51	92.45	88.64
Base+A	√	×	×	99.07	82.96	90.69	92.00	89.41
Base+D	×	√	×	99.07	82.98	90.70	92.19	89.25
Base+T	×	×	√	99.04	82.31	90.30	92.74	87.98
Base+A+D	√	√	×	99.07	82.96	90.69	92.25	89.18
Base+A+T	√	×	√	99.06	82.85	90.62	91.70	89.57
Base+D+T	×	√	√	99.05	82.66	90.51	92.19	88.89
Ours	√	√	√	99.08	83.29	90.88	91.98	89.80

表 9 WHUCD数据集消融实验结果

Tab. 9 Ablation experiment results of the methods on WHUCD dataset %

模型	ASPP	DEM	Transformer	OA	IoU	F1	Pre	Rec
Base	×	×	×	98.72	76.49	86.68	85.22	88.16
Base+A	√	×	×	99.04	78.52	87.97	87.72	88.22
Base+D	×	√	×	99.12	80.45	89.17	87.56	90.83
Base+T	×	×	√	98.79	74.33	85.28	82.17	88.62
Base+A+D	√	√	×	99.10	79.34	88.48	89.73	87.26
Base+A+T	√	×	√	99.01	77.89	87.57	86.87	88.29
Base+D+T	×	√	√	98.97	76.92	86.96	87.67	86.26
Ours	√	√	√	99.23	81.75	89.96	92.76	87.34

整性和内密性,进而使F1分数提高了1.37%。最终,当3个模块共同组合使用时,其效果最为显著,相较于基础网络Base,F1分数提高了3.28%。这一结果充分验证了所提网络各模块的有效性及其在组合使用时的优越性能。

3.5.3 学习曲线比较

图13对比了本文提出模型和BIT模型在训练集和验证集上的可视化结果,展示了经过200个

epoch的训练集和验证集的平均F1分数。从训练数据集的比较来看,本文模型展现出更快的收敛速度,能够以更高的训练精度更有效地达到最优的训练结果。在验证数据集方面,本文模型表现更为稳定,具有良好的泛化能力,而BIT模型随着epoch的增加振荡较大,导致鲁棒性较差。这也进一步证明了Transformer和UNet网络在遥感图像变化检测中具有强大的学习和泛化能力。

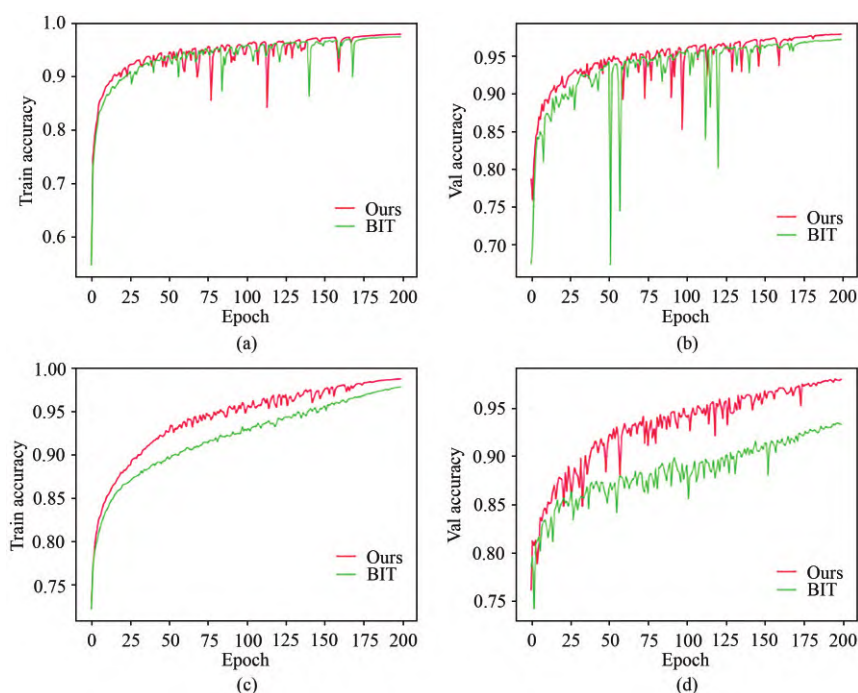


图13 学习曲线的可视化比较。(a) CDD训练学习曲线;(b) CDD验证学习曲线;(c) DSIFN训练学习曲线;(d) DSIFN验证学习曲线。

Fig. 13 Learning curve visualization comparison. (a) Learning curve from CDD training set; (b) Learning curve from CDD verification set; (c) Learning curve from WHUCD training set; (d) Learning curve from WHUCD verification set.

3.5.4 模型参数及效率分析

本文致力于在实现高精度检测的同时降低计算复杂度。为验证所提模型的有效性,我们在CDD、LEVIR和WHUCD数据集上进行了全面的比较和分析,涵盖了浮点运算量(FLOPs)、参数数量(Params)、F1分数(F1)以及交并比(IoU)等多个关键指标。如表10所示,本文提出的模型在FLOPs和Params方面优于STANet、SNUNet、DMINet和DGMA2Net,体现了本文模型在计算效率和模型复杂度上的优势。值得注意的是,本文模型在F1分数和交并比上也达到了最优,这进一步证明了本文模型在检测性能上

的卓越性。相比之下,FC-Siam-conc和FC-Siam-diff虽然通过减少特征提取层和最后一层通道数降低了模型大小,但在处理复杂检测任务时,它们的特征提取能力受限,导致准确率和鲁棒性不足。尽管BIT和USFFCNet在计算量和模型规模上表现出一定的优势,但它们的检测精度相对较低,难以满足高精度检测的需求。相比之下,本文所提模型不仅在参数数量和模型大小上有了显著的减少,更在检测性能上实现了最优,从而在确保高精度的同时降低了计算复杂度。这一成果充分验证了本文所提模型的有效性和实用性。

表 10 模型参数及效率分析
Tab. 10 Model parameters and efficiency analysis

Methods	Params/M	FLOPs/G	DSIFN	LEVIR	WHUCD
			IoU/F1/%	IoU/F1/%	IoU/F1/%
FC-Siam-conc	1.07	4.07	42.56/59.71	71.96/83.69	49.95/66.63
FC-Siam-diff	0.87	3.47	45.50/62.54	75.92/86.31	41.66/58.81
STANet	16.93	7.21	47.66/64.56	77.40/87.26	69.95/82.32
SNUNet	12.03	54.88	49.45/66.18	78.83/88.16	71.67/83.50
BIT	3.55	10.59	52.97/69.26	80.68/89.31	72.39/83.98
DMINet	6.24	14.55	73.46/84.70	82.11/90.18	79.68/88.69
USSFCNet	1.52	4.86	64.49/78.41	82.54/90.44	80.49/89.19
DGMA ² Net	37.10	18.10	69.66/82.12	82.29/90.29	67.20/80.38
Base+ASPP	4.85	14.58	79.29/88.45	82.96/90.69	78.52/87.97
Base+Transformer	4.18	13.35	76.09/86.42	82.31/90.30	80.45/85.28
Base+DEM	4.19	13.54	73.66/84.83	82.98/90.70	74.33/89.17
Ours	4.85	14.78	82.65/90.50	83.29/90.88	81.75/89.96

4 结 论

本文提出了一种用于遥感图像变化检测的混合 CNN-Transformer 架构网络。在网络结构中引入 Transformer 作为编码器-解码器的一部分,以更好地建模上下文、捕获全局上下文信息。通过与 UNet 解码器的融合,模型在增强浅层细节、实现精确定位和提高变化区域的边缘表达方面取得了显著效果。此外,引入 DEM 对变化图进行加权,实现了快速收敛,提高了网络的学习能力。

实验结果充分证明,本文模型在多个数据集上具有良好的鲁棒性和泛化性,明显优于所有基准比较方法,充分验证了 Transformer 在变化检测网络中的有效性。在季节变化数据集中,与其他方法相比,本文提出的模型更能准确识别伪季

节变化。在建筑数据集中,相比其他方法,本文模型更能保留建筑物的边缘和内部结构。在中等分辨率卫星图像数据集中,本文模型也保持了良好的检测精度,适用于不同的应用场景,如城市规划 and 土地利用/覆盖等。相较于没有 Transformer 的模型,本文模型通过消除冗余信息和完整建模特征,显著提高了精度。与带有 Transformer 的模型相比,本文模型能够更好地恢复变化区域的细节并保持变化区域的边缘。本文模型在训练集和验证集上的学习曲线证明了其泛化性和鲁棒性。

未来的工作将重点关注该模型在异构图像中的适应性,如光学卫星、合成孔径雷达(SAR)卫星和无人机图像。同时,仍需进一步研究建筑物的弱化、颜色近似的伪变化等问题。

参 考 文 献:

[1] 任秋如,杨文忠,汪传建,等. 遥感影像变化检测综述 [J]. 计算机应用, 2021, 41(8): 2294-2305.
REN Q R, YANG W Z, WANG C J, *et al.* Review of remote sensing image change detection [J]. *Journal of Computer Applications*, 2021, 41(8): 2294-2305. (in Chinese)

[2] PARELIUS E J. A review of deep-learning methods for change detection in multispectral remote sensing images [J]. *Remote Sensing*, 2023, 15(8): 2092.

[3] FANG B, CHEN G, PAN L, *et al.* GAN-based siamese framework for landslide inventory mapping using bi-temporal optical remote sensing images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(3): 391-395.

[4] SUBLIME J, KALINICHEVA E. Automatic post-disaster damage mapping using deep-learning techniques for change detection; case study of the Tohoku Tsunami [J]. *Remote Sensing*, 2019, 11(9): 1123.

- [5] SAHA S, BOVOLO F, BRUZZONE L. Building change detection in VHR SAR images via unsupervised deep transcoding [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(3): 1917-1929.
- [6] PAPADOMANOLAKI M, VAKALOPOULOU M, KARANTZALOS K. A deep multitask learning framework coupling semantic segmentation and fully convolutional LSTM networks for urban change detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(9): 7651-7668.
- [7] KHAN S H, HE X M, PORIKLI F, *et al.* Forest change detection in incomplete satellite images with deep neural networks [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(9): 5407-5423.
- [8] ISAIENKOV K, YUSHCHUK M, KHRAMTSOV V, *et al.* Deep learning for regular change detection in Ukrainian forest ecosystem with sentinel-2 [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 364-376.
- [9] 黄亮. 多时相遥感影像变化检测技术研究 [J]. 测绘学报, 2020, 49(6): 801.
HUANG L. Research on change detection technology in multi-temporal remote sensing images [J]. *Acta Geodaetica et Cartographica Sinica*, 2020, 49(6): 801. (in Chinese)
- [10] 刘红超, 张磊. 面向类型特征的自适应阈值遥感影像变化检测 [J]. 遥感学报, 2020, 24(6): 728-738.
LIU H C, ZHANG L. Adaptive threshold change detection based on type feature for remote sensing image [J]. *Journal of Remote Sensing*, 2020, 24(6): 728-738. (in Chinese)
- [11] 佃袁勇, 方圣辉, 姚崇怀. 一种面向地理对象的遥感影像变化检测方法 [J]. 武汉大学学报(信息科学版), 2014, 39(8): 906-912.
DIAN Y Y, FANG S H, YAO C H. The geographic object-based method for change detection with remote sensing imagery [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(8): 906-912. (in Chinese)
- [12] 黄亮, 左小青, 於雪琴. 遥感影像变化检测方法探讨 [J]. 测绘科学, 2013, 38(4): 203-206.
HUANG L, ZUO X Q, YU X Q. Review on change detection methods of remote sensing images [J]. *Science of Surveying and Mapping*, 2013, 38(4): 203-206. (in Chinese)
- [13] LV P Y, ZHONG Y F, ZHAO J, *et al.* Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(7): 4002-4015.
- [14] ZHUANG H F, DENG K Z, FAN H D, *et al.* Strategies combining spectral angle mapper and change vector analysis to unsupervised change detection in multispectral images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(5): 681-685.
- [15] HUSSAIN M, CHEN D M, CHENG A, *et al.* Change detection from remotely sensed images: From pixel-based to object-based approaches [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013, 80: 91-106.
- [16] 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.
- [17] 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.
- [18] MALILA W A. Change vector analysis: an approach for detecting forest changes with Landsat [C]//*Proceedings of LARS symposia*. West Lafayette: IEEE, 1980: 385.
- [19] 慕彩虹, 霍利利, 刘逸, 等. 基于小波融合和 PCA-核模糊聚类的遥感图像变化检测 [J]. 电子学报, 2015, 43(7): 1375-1381.
MU C H, HUO L L, LIU Y, *et al.* Change detection for remote sensing images based on wavelet fusion and PCA-kernel fuzzy clustering [J]. *Acta Electronica Sinica*, 2015, 43(7): 1375-1381. (in Chinese)
- [20] CANTY M J, NIELSEN A A. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation [J]. *Remote Sensing of Environment*, 2008, 112(3): 1025-1036.
- [21] ZHAN Y, FU K, YAN M L, *et al.* Change detection based on deep siamese convolutional network for optical aerial images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(10): 1845-1849.
- [22] DAUDT R C, LE SAUX B, BOULCH A. Fully convolutional siamese networks for change detection [C]//

- Proceedings of the 2018 25th IEEE International Conference on Image Processing*. Athens: IEEE, 2018: 4063-4067.
- [23] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [24] CHEN L C, ZHU Y K, PAPANDREOU G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018: 833-851.
- [25] ZHAO H S, SHI J P, QI X J, *et al.* Pyramid scene parsing network [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 6230-6239.
- [26] FU J, LIU J, TIAN H J, *et al.* Dual attention network for scene segmentation [C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 3141-3149.
- [27] 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.
- [28] 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.
- [29] 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*, 2022, 19: 8007805.
- [30] 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*. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [31] CHEN C F R, FAN Q F, PANDA R. CrossViT: cross-attention multi-scale vision transformer for image classification [C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 347-356.
- [32] HAN G X, MA J W, HUANG S Y, *et al.* Few-shot object detection with fully cross-transformer [C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE, 2022: 5311-5320.
- [33] WANG L B, LI R, ZHANG C, *et al.* UNetFormer: a UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 190: 196-214.
- [34] DING L, LIN D, LIN S F, *et al.* Looking outside the window: wide-context transformer for the semantic segmentation of high-resolution remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 4410313.
- [35] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16×16 words: transformers for image recognition at scale [C]//*Proceedings of the 9th International Conference on Learning Representations*. Online: OpenReview.net, 2021.
- [36] CHEN H, QI Z P, SHI Z W. Remote sensing image change detection with transformers [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5607514.
- [37] FENG Y C, JIANG J W, XU H H, *et al.* Change detection on remote sensing images using dual-branch multilevel intertemporal network [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 4401015.
- [38] 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: 4402114.
- [39] YING Z L, TAN Z J, ZHAI Y K, *et al.* DGMA²-Net: a difference-guided multiscale aggregation attention network for remote sensing change detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5619716.
- [40] HE K M, ZHANG X Y, REN S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [41] JI S P, WEI S Q, LU M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(1): 574-586.

- [42] 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, e XLII-2: 565-571.

作者简介:



潘梦洋,男,硕士研究生,2021年于河南大学获得学士学位,主要从事遥感图像变化检测算法方面的研究。
E-mail:panmy1479@163.com



杨航,男,博士,研究员,2012年于吉林大学获得博士学位,主要从事图像复原、图像增强、目标识别与跟踪方面的研究。E-mail:yanghang@ciomp.ac.cn