

中图法分类号: P2 文献标识码: A 文章编号: 1006-8961(2025)03-0870-13

论文引用格式: Wang R F, Yang Z J, Qiu H, Wang F, Gao G and Wu D. 2025. Lightweight change detection network integrating feature interaction and fusion. Journal of Image and Graphics, 30(3):0870-0882(王仁芳, 杨梓健, 邱虹, 王峰, 高广, 吴敦. 2025. 结合特征交互和融合的轻量级变化检测网络. 中国图象图形学报, 30(3):0870-0882)[DOI:10.11834/jig.240280]

结合特征交互和融合的轻量级变化检测网络

王仁芳^{1,2}, 杨梓健², 邱虹^{1*}, 王峰¹, 高广^{1,3}, 吴敦^{1,3}

1. 浙江万里学院大数据与软件工程学院, 宁波 315100; 2. 上海海洋大学信息学院, 上海 201306;

3. 宝略科技(浙江)有限公司, 宁波 315042

摘要: 目的 利用深度学习开展变化检测是遥感智能解译热点研究方向之一。针对基于Transformer变化检测模型结构复杂、参数过多、训练耗时的问题,设计了一种融合特征交互和融合的轻量级变化检测网络(feature interaction and fusion lightweight network, FIFLNet)。方法 解码器中采用EfficientNet作为特征提取网络,其能利用模型的放缩(model scaling)能力来扩大模型的感受野。然后通过设计通道、像素交互模块(spatial and channel interact block)和浅层跳跃连接(low-level skip-connection)来实现浅层双时相的细节特征交互和上采样阶段的传递,以此增加模型对局部特征的判别精度。此外,利用特征融合分组卷积模块(feature fusion and groups convolution block, FFGCB)对双时相数据进行降维融合,降低了模型计算量。最后,设计了融合上采样模块(fusion upsampling block, FUB)对局部特征与全局特征进行融合还原,同时利用局部特征的细节、纹理来补偿全局特征细节的缺失。**结果** 本文方法在两个遥感影像数据集LEVIR-CD(landearth view image retrieval building change detection dataset)和SYSU-CD(Sun Yat-sen University change detection dataset)上与13种SOTA(state-of-the-art)方法进行比较。实验结果表明,本文方法在变化检测任务中展现出显著优势。在定量评估方面,本文方法在LEVIR-CD和SYSU-CD数据集上的F1值分别达到91.51%和82.19%,较现有最优方法分别提升0.43%和1.58%。同时,模型计算效率优异,仅需1.66 GFLOPs的运算量和0.56 M的参数规模,显著低于所有对比方法。在定性分析方面,本文方法展现出更精确的检测区域定位、更低的漏检率,并能够保留更丰富的细节特征。**结论** 本文提出的轻量级变化检测网络FIFLNet以较少的参数量和每秒浮点运算量获得了优越的性能,改善了小目标漏检、边界误检的情况,能够获得高质量的变化检测结果。

关键词: 遥感影像;变化检测;局部特征;特征交互;轻量级网络

Lightweight change detection network integrating feature interaction and fusion

Wang Renfang^{1,2}, Yang Zijian², Qiu Hong^{1*}, Wang Feng¹, Gao Guang^{1,3}, Wu Dun^{1,3}

1. College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315100, China; 2. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; 3. Popsmart Technology (Zhejiang) Co., Ltd., Ningbo 315042, China

Abstract: Objective Change detection in remote sensing imagery is a process that leverages remote sensing technology to

收稿日期:2024-05-28;修回日期:2024-08-13;预印本日期:2024-08-20

*通信作者:邱虹 qiuHong@zhu.edu.cn

基金项目:国家自然科学基金项目(61906170);浙江省科技计划项目(LGF21F020023,ZCLY24F0301);宁波市科技计划项目(2022Z233, 2021Z050, 2022S002, 2023J403)

Supported by: National Natural Science Foundation of China(61906170); Science and Technology Plan for Zhejiang Province (LGF21F020023, ZCLY24F0301); Plan Project of Ningbo Municipal Science and Technology(2022Z233, 2021Z050, 2022S002, 2023J403)

compare and analyze images from the same geographical area but captured at different time intervals. This process mainly aims to identify changes on the Earth's surface. The main challenge in this process lies in the extraction of effective change features from a large volume of image data and subsequently mapping them onto pixel-level change labels for high-precision detection. Detection methods for changes in remote sensing imagery can be broadly divided into traditional and deep learning-based methods. Traditional methods primarily rely on image processing and pattern recognition techniques. However, these methods often require manual selection of suitable features and thresholds, which can introduce a degree of subjectivity and limitations. By contrast, deep learning methods can automatically learn abstract and high-level change features from remote sensing images, thereby enabling end-to-end change detection. This approach notably enhances the accuracy and efficiency of change detection. Among them, change detection models based on convolutional neural networks (CNNs) and Transformer architectures have shown remarkable performance. Models that utilize these mechanisms have demonstrated notable advancements in recent years due to the extensive research conducted by scholars worldwide. However, for the currently effective models based on the Transformer architecture, the complexity of the model also increases as the detection accuracy of the model improves. Therefore, designing a change detection method with a small parameter size, low computational cost, and high detection accuracy is a pressing issue that must be urgently addressed in this field.

Method This paper proposes a lightweight change detection method for remote sensing images based on feature interaction and fusion. The main idea of this method is to use EfficientNet B7 as a lightweight backbone network to extract deep- and low-level features from bi-temporal remote sensing images. Channel and pixel swap modules are introduced to enable the interaction and combination of bi-temporal features, enhancing the spatiotemporal feature fusion. Low-level skip-connections are employed to transfer the original image information to the upsampling phase, aiming to preserve additional edge and texture details and reduce artifact generation. A feature fusion group convolution module that reduces the computational overhead and the number of parameters is designed to effectively fuse the deep- and low-level features obtained in the downsampling stage. Finally, the feature fusion group convolution and upsampling modules are used to fuse and recover the deep- and low-level features, and the pixel-level change detection map is generated. **Result** In this paper, experiments on two datasets are conducted for remote sensing image change detection landearth view image retrieval building change detection dataset (LEVIR-CD) and Sun Yat-sen University change detection dataset (SYSU-CD). Each dataset is split into 7:1:2 for training, validation, and testing, respectively, and each image is segmented into 256×256 pixels. This approach facilitated the processing and increased the generalization capability of the model. The binary cross-entropy (BCELoss) is used as the loss function, and the performance of the proposed method is evaluated using three metrics: F1 score (F1), intersection over union (IoU), and overall accuracy (OA). The proposed method achieved F1 scores of 91.51% and 82.19%, IoU of 84.35% and 69.76%, and OA of 99.14% and 91.99% on the LEVIR-CD and SYSU-CD datasets, respectively. Compared with previous classical methods, the model obtained the best change detection results, especially preserving additional details on the change boundary. Ablation experiments are performed on the same dataset to illustrate the effect of low-level skip-connections and channel and spatial exchange modules. The results showed that the channel and spatial exchange module substantially optimized the utilization and representation of spatiotemporal features in the network, while the low-level skip-connection compensated for the loss of detailed features in the downsampling process and further enhanced the feature learning capability of the network. **Conclusion** The network used channel and spatial exchange modules to increase the utilization and understanding of spatiotemporal features and low-level skip-connections to focus the model on local detailed features. Finally, a binary cross-entropy loss function at the output layer is utilized to achieve optimal change detection performance. Experiments show that the method proposed in this paper can improve the capability of recognizing changing regions while ensuring a light network, enhancing the detection performance of change detection in various environments and terrains.

Key words: remote sensing image; change detection; local feature; feature interact; light network

0 引言

遥感影像变化检测是利用数理模型及遥感图像处理理论,结合地物特性和遥感成像原理,通过分析同一地理区域的多时期遥感影像和相关地理空间数据,来识别区域地物变化(位置、范围、性质和状态的变化)的技术(眭海刚等,2018)。该技术在环境监测、城市规划、灾害评估、资源管理和覆盖变化分析等领域具有重要的应用价值(Zhang等,2021; Silva等,2020;王蓉芳等,2023)。

遥感影像变化检测方法按照技术手段不同可以分为传统变化检测和基于深度学习的变化检测。前者主要包括可视化分析方法、基于代数的方法和基于变换的方法等(杨胜等,2009;杜培军和柳思聪,2012)。这些方法应用范围相对有限,且易受大气条件、季节变化及成像时间等影响。目前,得益于深度学习(deep learning)强大的学习和表征能力,基于深度学习的变化检测方法已逐渐成为遥感领域的研究热点。自Lecun等人(1998)提出卷积神经网络(convolutional neural network, CNN),并成功应用于变化检测后,学者们提出一系列基于CNN的变化检测方法,并且取得了显著的成就和进展。2018年,Daudt等人(2018)基于全卷积网络(fully convolutional network, FCN)提出FC-Sima-conc(fully convolutional siamese-concatenation)网络,该网络首次将孪生(siamese)网络引入变化检测任务中,实现双输入特征提取,并通过特征拼接来获取变化信息。基于FCN模型,Daudt等人还提出了FC-Sima-diff(fully convolutional siamese difference)网络,其与FC-Sima-conc的主要区别在于,特征还原阶段采用差值连接。两种方法中引入的孪生网络变化检测模型能够有效地保留影像的高维特征,使得方法在变化检测中能取得不错的效果。但是,由于CNN的感受野往往受限于卷积核的尺度,使得基于CNN的变化检测方法无法有效捕获全局上下文信息。

为克服这一挑战,学者们将注意力机制(attention mechanism)引入到变化检测任务中。2020年,Chen和Shi(2020)提出基于孪生结构的时空注意力网络(spatial-temporal attention network, STANet),该网络引入自注意力机制,以建模像素间的时空关系,从而获取更具判别性的特征。Zhang等人(2020)提

出深度监督影像融合网络(deeply supervised image fusion network, DSIFN),该网络在解码器中采用通道和空间注意力机制,以细化双时相影像特征,并通过深度监督的方式进行网络训练。Liu等人(2021)提出一个双任务约束的孪生卷积神经网络(dual task constrained deep siamese convolutional network, DTCDSN),其由3个子网络构成,包括一个变化检测网络和两个语义分割网络。此外,该模型引入了双注意力模块,并改进多标签分类损失函数以解决样本不均衡的问题。Fang等人(2022)基于CNN和U-Net++(Zhou等,2018)结构提出稠密连接孪生网络(siamese nested U-Net network, SNUNet),该网络利用U-Net++网络的密集连接结构来缓解神经网络全局特征定位信息的丢失问题,同时通过集成通道注意力模块来增强不同语义层次特征的识别能力。Lei等人(2023)提出超轻量级光谱特征协作网络(ultralight-weight spatial-spectral feature cooperation network, USSFC-Net),设计多尺度膨胀卷积,减少参数量和计算冗余,并引入空间光谱特征协作机制以获取更丰富的特征。Noman等人(2024)提出高效局部全局上下文聚合网络(efficient local-global context aggregation network, ELGC-Net),该模型在编码器中引入高效局部全局上下文聚合模块,分别通过全新的池化转置和深度卷积捕获和增强全局上下文和局部空间信息。Zhong等人(2024)提出使用定位并细化特征策略的网络(localization-then-refinement network, LRNet),其使用三支编码器 and 可学习池化来减少特征丢失的现象,并提出3种变化区域对齐模块,细化对变化区域的判别。尽管这些方法在一定程度上取得了优异的效果,但感受野受限的问题仍然存在,这意味着模型无法充分捕获全局特征。

为解决上述问题,谷歌旗下人工智能团队提出Transformer(Vaswani等,2017)模型,其感受野可以覆盖整幅影像,能够对全局上下文信息进行有效建模。Li等人(2024)提出由一系列级联注意模块组成的级联Transformer架构(a cascaded Transformer architecture, CasFormer),利用局部光谱和空间相似性以及全局相互关联实现高空间分辨率RGB图像的融合,更紧凑、准确地恢复光谱序列信息。这些方法都为注意力机制在变化检测中的应用提供新的解决方案和启发。为此,学者们纷纷开始探索将

Transformer 应用于变化检测任务。Chen 等人 (2022) 提出的 BiT (Bi-temporal Transformer) 网络是最早将 Transformer 架构用于变化检测的工作之一, BiT 利用孪生 Transformer 来改进 CNN 提取的特征, 实现 CNN 和 Transformer 的协同训练, 从而在保持轻量模型结构的同时得到准确的变化检测结果。Bandara 等人 (2022b) 提出基于 Transformer 的孪生变化检测网络 (Transformer-based siamese network for change detection, Changeformer), 该网络将 Transformer 编码器与孪生网络中的多层感知机相结合, 通过有效渲染多尺度长程细节以获得更优精度。Bandara 等人 (2022a) 利用生成模型 (generative model) 的概率推断和处理噪声数据的能力, 提出基于去噪扩散概率的变化检测模型 (denoising diffusion probabilistic models change detection, DDPM-CD), 该模型使用去噪扩散概率进行特征重生成, 极大地减缓了由传感器和气候等因素造成的遥感变化检测中的噪声干扰问题。但基于 Transformer 方法的多层感知机 (multilayer perceptron, MLP) 和多头注意力机制会导致计算量、参数量激增, 而 Diffusion 方法通常结构过于复杂, 且参数过多, 这使得该方法的训练过程更加复杂和耗时。综上所述, 如何设计一种参数量小、计算量低且检测精度高的变化检测方法是该领域亟需解决的问题。

针对上述问题, 本文的主要工作和贡献如下: 1) 本文提出一种融合特征交互和融合的轻量级变化检测方法, 该网络参数量和浮点运算量小, 并实现了高精度的变化检测。在两个数据集上的实验表明, 该方法以较少的参数和浮点计算实现了 SOTA (state-of-the-art) 的变化检测结果。2) 设计了通道交互模块 (channel interact block, CIB) 和像素交互模块 (spatial interact block, SIB) 来实现双时相特征的互动和融合, 增强对时空信息的利用和表达, 以降低模型误检率。同时加入浅层跳跃连接 (low-level skip-connection), 将局部细节特征传递到上采样阶段, 增强模型对局部细节特征的判别能力。此外, 通过设计特征融合分组卷积模块 (feature fusion group convolution block, FFGCB) 有效地融合下采样阶段得到的双时相特征图, 并通过分组卷积降低数据维度, 减少计算负担。3) 在解码器部分, 设计了融合上采样模块 (fusion upsampling block, FUB), 该模块将下采

样阶段得到的全局特征和局部特征进行融合和还原, 并利用局部特征的细节、纹理来弥补全局特征细节的缺失。

1 研究方法

本文所提出的方法的核心设计是融合特征交互和特征融合的孪生 EfficientNet 遥感影像变化检测方法。方法的总体框架如图 1 所示。

1.1 编码器部分

本模型采用在 Imagenet-1k (Deng 等, 2009) 中已预训练的 EfficientNet B7 (Tan 和 Le, 2019) 网络作为主干网络, 该网络通过模型放缩改善传统 CNN 感受野受限问题, 同时使用 MBConv (Tan 和 Le, 2021) 优化模型的参数量和每秒浮点计算次数。实验中, 两层主干网络下采样会使模型的特征表达能力受限, 无法提取高级语义特征, 导致分类精度较低; 四层主干网络下采样会导致模型过拟合, 参数量冗余等问题; 经实验表明, 三层主干网络组成的编码器特征表达能力较强, 分类精度较高。

设 $f(x)$ 为第 k 层 ($k \in [1, 3]$) 主干网络。将第 k 层主干网络输出的双时相特征图表示为 X_k 和 Y_k 。其中, X_k 和 Y_k 为两幅具有相同尺寸 (包括通道数、宽度和高度) 的图像, 具体为

$$\begin{aligned} X_k &= f_k(X_{k-1}) \\ Y_k &= f_k(Y_{k-1}) \end{aligned} \quad (1)$$

1.1.1 通道、像素交互模块

因为变化检测的核心是对象的变化, 而非对象的出现或消失 (Fang 等, 2023), 所以双时相特征交互可作为提升数据时空特征利用率的手段。为使模型充分利用双时相图像中的时空信息, 本方法在编码器部分分别设计了通道交互模块和空间交互模块。如图 2 所示, 通道交互模块从通道信息交互的角度进行特征交换, 通过交换除去双时相数据首尾通道外的中间通道来提升数据通道级时空信息的利用率。而像素交互模块从像素出发, 通过交换双时相数据偶数行和列的像素来优化数据像素级时空信息的利用率。

$$X'_c = \begin{cases} X_c & c = 0 \text{ 或 } -1 \\ Y_c & \text{其他} \end{cases} \quad (2)$$

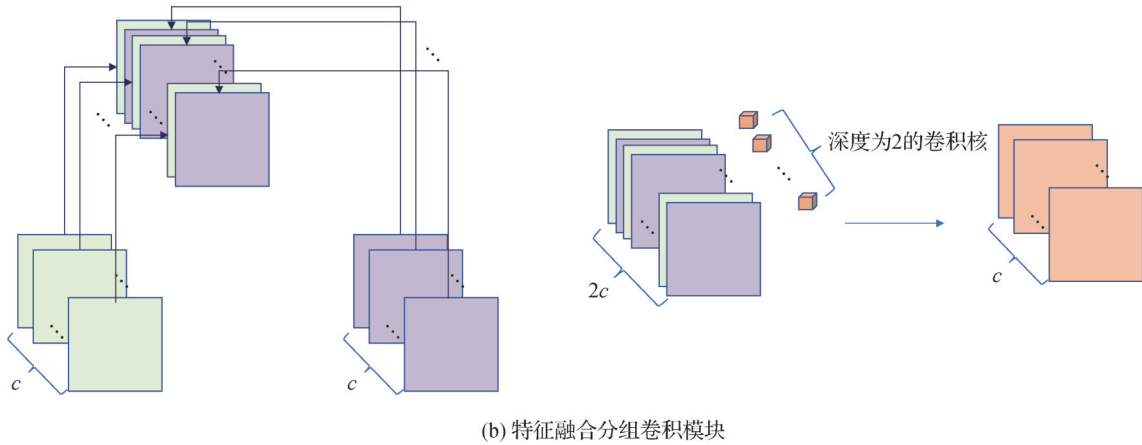
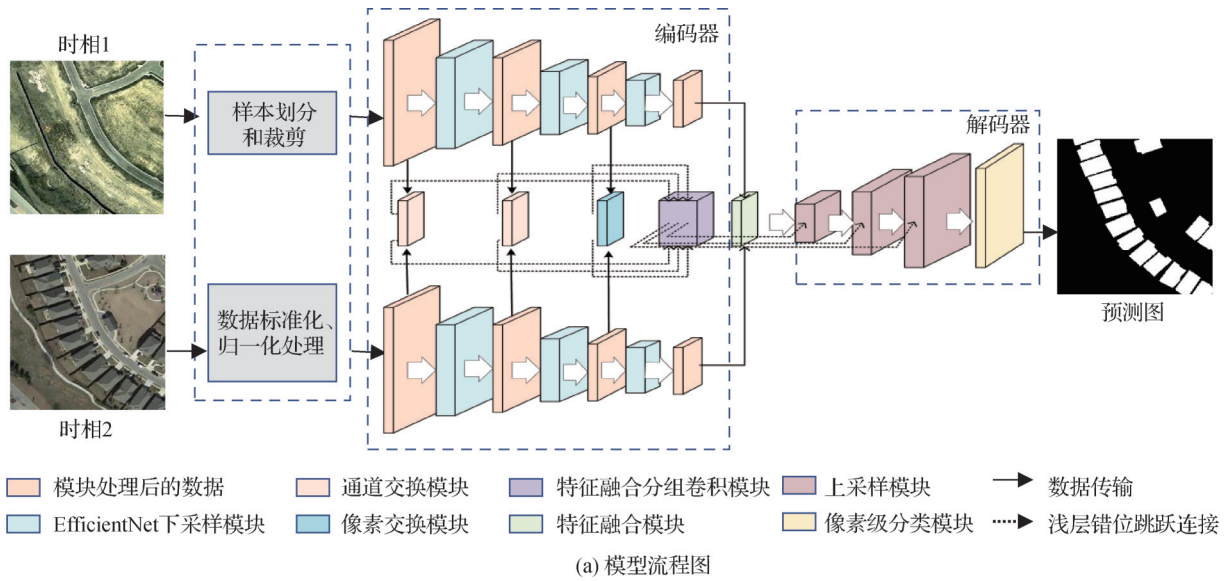


图1 本文方法的总体流程

Fig. 1 The overall flowchart of the proposed method ((a) flowchart of our model; (b) feature fusion and group convolution block)

$$X'_c = \begin{cases} X_c & c \text{ 为奇数} \\ Y_c & c \text{ 为偶数} \end{cases} \quad (3)$$

式(2)为通道交互,式(3)为像素交互; X 为T1数据, Y 为T2数据, c 为当前通道。

1.1.2 浅层跳跃连接

为更有效增强模型对局部特征的识别能力,本文在网络中加入浅层跳跃连接。如图3所示,相较于普通跳跃连接,浅层跳跃连接强调数据的局部特征。其能有效利用局部特征中的细节、纹理和棱角信息,并将这些特征传入融合上采样模块内,与全局特征进行融合上采样。

1.2 解码器部分

传统上采样模块采用对称跳跃连接数据以及全局特征数据作为输入,缺乏对局部特征的利用,使网

络丢失大量细节特征,容易出现边界干扰等问题。本方法中的解码器由3个融合上采样模块组成,融合上采样模块的输入包括两部分:全局特征数据和浅层跳跃连接数据。融合上采样模块使用网络处理层数较少的浅层数据,能有效抑制边界干扰。在上采样过程中,需对全局特征数据进行一次反卷积操作来解决浅层跳跃连接数据与全局特征数据尺寸不匹配的问题。具体为

$$U_k = U_p(U_{k-1}, M_k) = \text{PReLU}[\text{IN2d}[U_p(U_{k-1}) \odot M_k]] \quad (4)$$

式中, U_p 为反卷积操作, U_k 为第 k 层上采样模块的输出数据, M_k 为浅层跳跃连接数据。 PReLU 为线性参数校正激活函数(parametric rectified linear unit)(He等,2015), IN2d 为实例归一化操作(Ulyanov等,2017), $\odot$ 为哈达玛积。

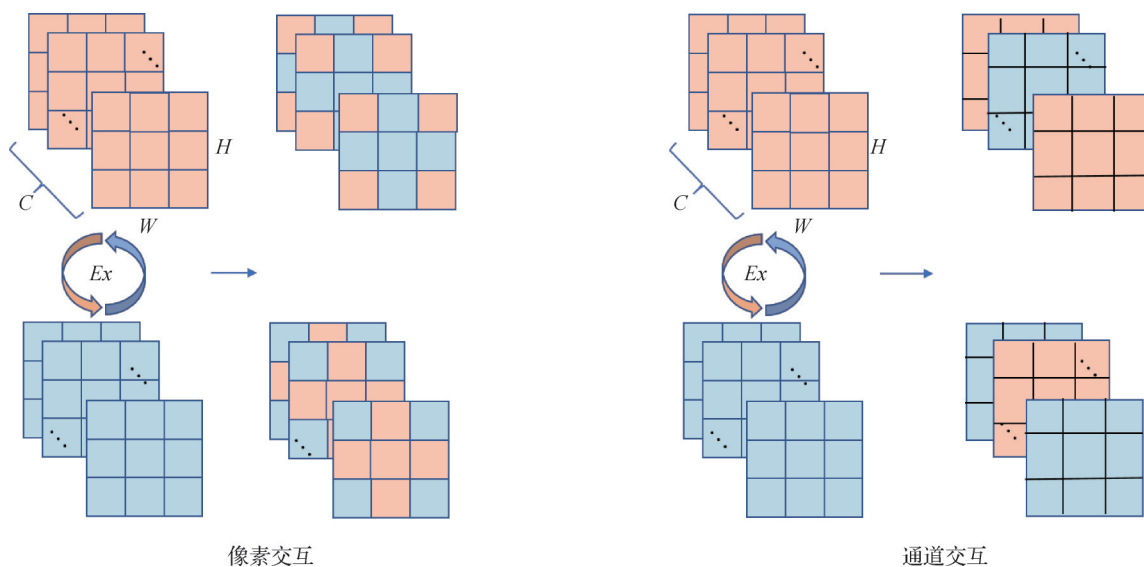


图2 通道、空间交互模块

Fig. 2 Channel exchange and spatial exchange modules

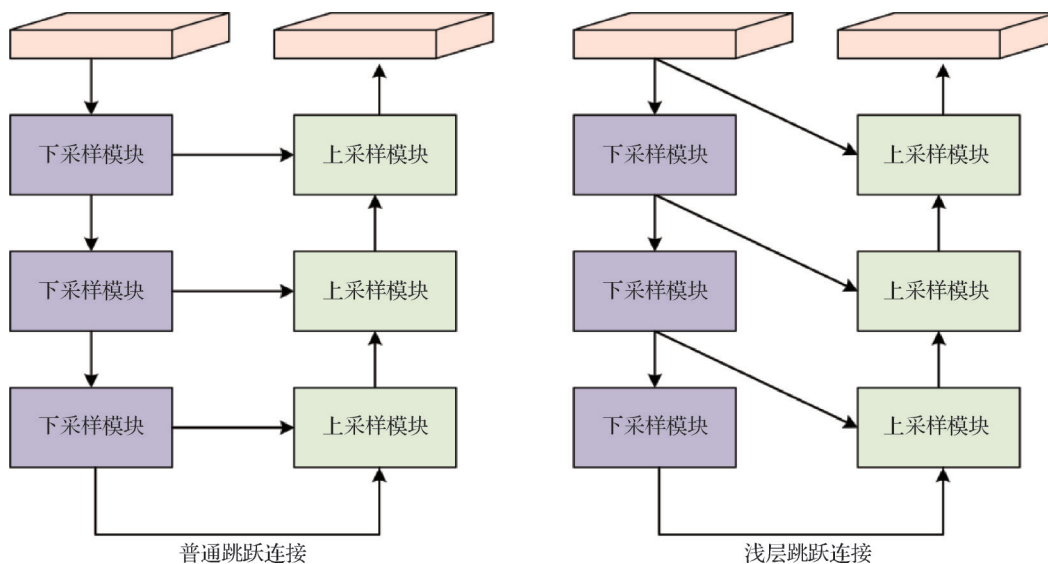


图3 普通跳跃连接和浅层跳跃连接

Fig. 3 Normal skip-connection and low-level skip-connection

1.3 特征融合分组卷积模块

为在上采样之前将双时相特征融合为单个张量,设计特征融合模块将双时相特征进行通道级融合,具体为

$$\mathbf{Z}_k^c = \begin{cases} \mathbf{X}_k^{c/2} & c \text{ 为偶数} \\ \mathbf{Y}_k^{(c-1)/2} & c \text{ 为奇数} \end{cases} \quad (5)$$

式中, $\mathbf{Z}_k^c$ 代表混合后的数据, c 为数据的通道数。

在特征融合模块的基础上增加分组卷积,以进一步降低模型的计算量并对数据进行降维。则第 k 层网络的输出 $\mathbf{Z}_{k_{out}}$ 可表示为

$$\mathbf{Z}_{k_{out}} = \text{Mix}(\mathbf{X}_k, \mathbf{Y}_k) = \text{PReLU}[\text{IN2d}[\text{Conv}(\mathbf{Z}_k, ch_{in} = 2C_k, ch_{out} = C_k, groups = C_k)]] \quad (6)$$

式中, Mix 为数据融合操作; ch_{in} 为输入通道数, ch_{out} 为输出通道数; $groups$ 为分组卷积的通道数。

1.4 像素级分类

为提升分类的准确率,本文采用像素级分类。采用 1×1 卷积核的卷积调整图像通道数和特征图的深度,同时进行特征的线性组合和非线性变换。然后通过 sigmoid 函数,将预测值映射到 $[0, 1]$ 的范围内。最后,将映射后的值使用阈值 0.5 进行二分类,生成最终分类结果。

1.5 损失函数

本文采用在二分类任务中表现优秀的二元交叉熵损失函数,该函数对概率分布的差异性具有高度

敏感性。通过将模型的分类结果与真实值进行比较来更新神经元的权重,从而优化模型的性能,具体为

$$L(\mathbf{M}, \mathbf{P}) = -\frac{1}{|H| \times |W|} = \sum_{h \in H, w \in W} M_{h,w} \times \log(P_{h \times w}) + (1 - M_{h \times w}) \log(1 - P_{h \times w}) \quad (7)$$

式中, $\mathbf{M}$ 为标签值, $\mathbf{P}$ 为模型的预测输出, H 和 W 代表图像的高度和宽度。

2 实验设置

本文选取 LEVIR-CD (landearth view image retrieval building change detection dataset) (Chen 和 Shi, 2020) 和 SYSU-CD (Sun Yat-sen University change detection dataset) (Shi 等, 2022) 两个遥感影像数据集进行实验验证。样本划分和裁剪方法遵循 2.1 节的方法。实验采用 PyTorch (Paszke 等, 2019) 框架, 并选用 1.5 节中的二值交叉熵函数作为损失函数。模型的优化器经过对比选用适合本文任务的 Adamw (Loshchilov 和 Hutter, 2019), 且采用智能神经网络 (neural network intelligence, NNI) (Yan 等, 2022) 作为超参数调整工具。实验共进行 120 个 epoch 训练, 从中选取出最佳模型。选取 13 种现有先进模型作为对比方法, 包括: FC-EF (fully convolutional early fusion) (Daudt 等, 2018)、FC-Siam-Diff (Daudt 等, 2018)、FC-Siam-Conc (Daudt 等, 2018)、DSIFN (deeply supervised image fusion network) (Zhang 等, 2020)、DTCDCN (dual-task constrained deep siamese convolutional network) (Liu 等, 2021)、STANet (spatio-temporal attention network) (Chen 和 Shi, 2020)、SNUNet (siamese nested U-Net network) (Fang 等, 2022)、ELGC-Net (efficient local-global context aggregation network) (Noman 等, 2024)、USSFC-Net (ultralightweight spatial-spectral feature cooperation network) (Lei 等, 2023)、LRNet (localization-then-refinement network) (Zhong 等, 2024)、BiT (Bi-temporal Transformer) (Chen 等, 2022)、Changeformer (Bandara 等, 2022b) 和 DDPM-CD (denoising diffusion probabilistic models change detection) (Bandara 等, 2022a)。此外, 本文还在所提模型的基础上, 采用 6 种不同模块组合进行实验, 以寻求模型的最优配置方案。

2.1 数据集

LEVIR-CD 数据集由 637 个超高分辨率 (very high resolution, VHR, 0.5 m/像素) 的谷歌地球 (Google Earth, GE) 图像补丁对组成, 尺寸为 $1\,024 \times 1\,024$ 像素。SYSU-CD 数据集由 20 000 个超高分辨率 (VHR, 0.5 m/像素) 的图像补丁对组成, 尺寸为 256×256 像素。数据集中均包含陆地、建筑物和道路等元素, 受到季节变化、光照强度和天气条件等多种因素的影响。LEVIR-CD 主要关注建筑物的变化检测, 而 SYSU-CD 除了关注建筑物外还关注河流、水面和森林的变化内容。

训练样本的数量对模型的训练和评估结果有重要影响。如果训练样本过多 (超过 70%), 可能会导致模型过拟合, 从而影响评估结果的可靠性; 如果训练样本过少, 可能会导致模型训练不足。同时, 验证样本的数量也应适中, 以避免对模型性能的过度优化。测试样本占 20%, 可以提供一个全面且可靠的最终性能评估。因此, 本文将两个遥感数据集按照 7:1:2 的比例划分为训练样本、验证样本和测试样本, 旨在确保充足的训练、适当的验证以及可靠的性能评估。实验过程中, 为优化训练过程并提高模型对数据的学习能力, 将每幅 LEVIR-CD 中 $1\,024 \times 1\,024$ 像素的数据裁剪成 16 幅 256×256 像素的子图, 用于训练、验证和测试。

2.2 定量评价

本文采用准确率 F1 score、全局精度 (overall accuracy, OA) 和交并比 (intersection over union, IoU) 3 个评价指标验证所提模型的有效性。其中, F1 score 反映模型的综合性能; OA 和 IoU 分别用来衡量模型的整体精度以及预测结果与真实结果的重叠程度。

3 实验分析

实验对本文所提方法和选取的对比方法的变化检测结果进行定量评价。使用首次提出孪生网络的 FC-EF、FC-Siam-Diff 和 FC-Siam-Conc 这 3 个模型作为所有对比实验的基准模型。为验证本模型与其他采用空间和通道注意力机制模型的优势, 选取 DTCDCN、STANet、DSIFN、SNUNet、USSFC-Net、ELGC-Net 和 LRNet 这 7 个模型进行对比实验。鉴于 Transformer 在计算机视觉领域的成功应用, 将 BiT

和Changeformer与本模型进行实验对比。考虑到生成模型在特征生成方面的潜力,将DDPM-CD也纳入对比实验中,以全面对比各模型在该任务上的性能表现。为直观展示实验结果,以下将具有代表性的对比实验与所提模型的部分数据进行可视化。图中的黑色部分代表未变化区域,白色部分代表变化区域,虚线框则代表重点关注区域。

3.1 对比实验

如图 4 和图 5 所示,尽管各种方法均能有效检测变化区域,但 DSIFN、STANet 和 SNUNet 所采用的主干网络泛化能力较差,提取出的目标特征丢失较为严重,特别是在检测小目标的变化时,容易出现漏

检现象。而本文方法通过浅层跳跃连接,在保证普通目标检测率的同时,提升了小目标变化检测的准确率。DTCDCN、BiT 和 DDPM-CD 方法在小目标的变化检测中表现良好,但在地物变化较为复杂的情况下,仍然存在误检率较高、检测精度较低的问题。本网络通过构建通道、像素交互模块和特征融合分组卷积模块,使网络提升了对该情况下的特征利用率,降低了模型的误检率,提高了变化检测精度。现有方法较为依赖全局特征导致相邻的变化区域互相干扰严重,出现难以区分独立变化区域的问题。本文通过构建融合上采样模块,将局部特征与全局特征进行融合上采样,有效抑制了边界干扰的问题。

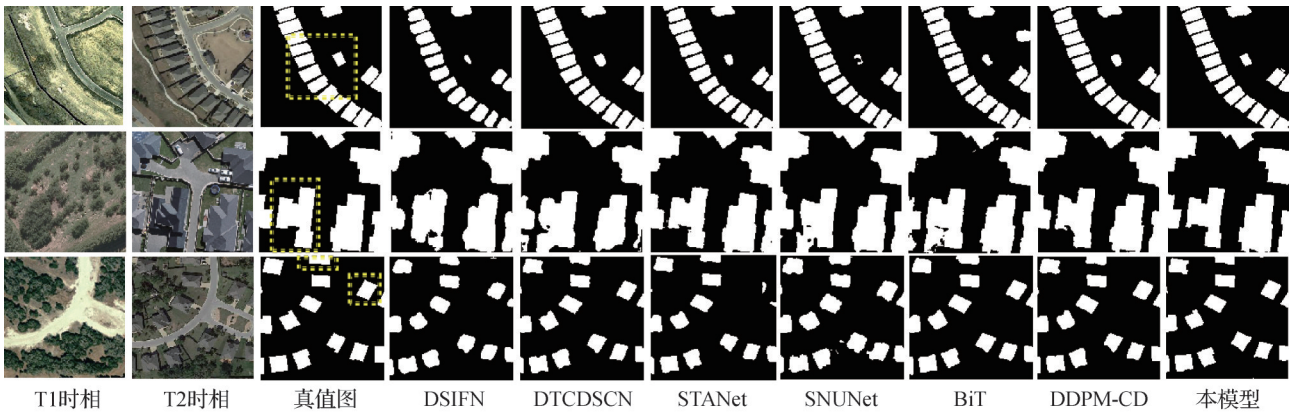


图 4 LEVIR-CD 数据集上部分对比实验变化检测结果可视化

Fig. 4 Visual results of the comparison experiments between our method and the SOTA models on the LEVIR-CD dataset

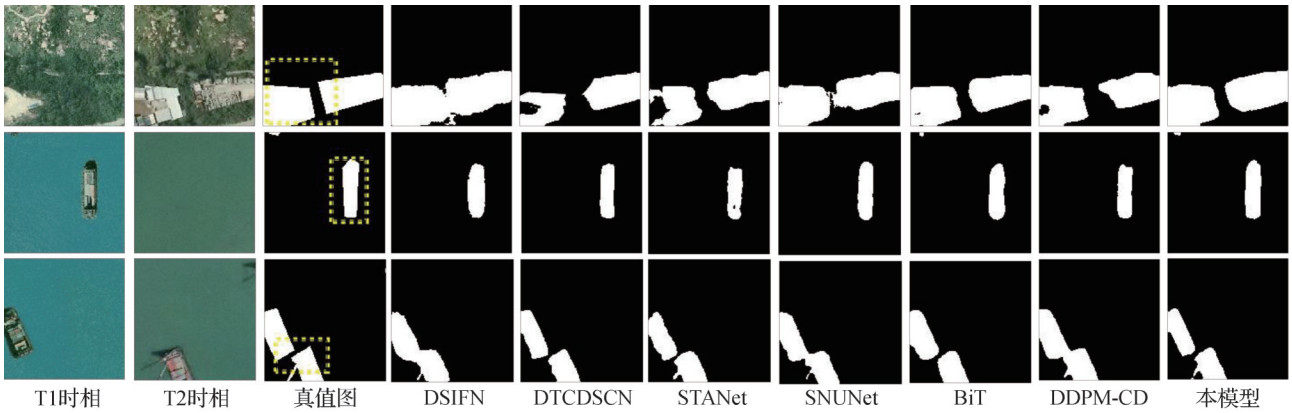


图 5 SYSU-CD 数据集上部分对比实验变化检测结果可视化

Fig. 5 Visual results of the comparison experiments between our method and the SOTA models on the SYSU-CD dataset

14 种方法在 LEVIR-CD 和 SYSU-CD 数据集上的定量评价指标如表 1 所示。从表 1 可以看出,本文所提 FIFLNet 方法在两个数据集上的 F1、IoU 和 OA 分别为 91.51%、84.35%、99.14% 和 82.19%、69.76%、91.99%,在所有对比方法中取得最优。这主要得益

于通道、像素交互模块的嵌入,使得网络更敏锐地关注于变化信息,以及浅层跳跃连接提升了网络捕捉局部细节特征的能力,融合上采样模块有效地结合局部特征和全局特征,抑制边界干扰现象。如表 2 所示,本文从参数量(parameters)、每秒浮点运算次

表 1 各模型在 LEVIR-CD 和 SYSU-CD 数据集上训练结果的评价对比

Table 1 Quantitative evaluation of the training results of various model on the LEVIR-CD and SYSU-CD datasets

		/%					
模型	方法	LEVIR-CD			SYSU-CD		
		F1	IoU	OA	F1	IoU	OA
CNN	FC-EF(Daudt 等, 2018)	83.40	71.53	98.39	75.07	60.09	88.46
	FC-Siam-Diff(Daudt 等, 2018)	86.31	75.92	98.67	72.58	56.96	88.13
	FC-Siam-Conc(Daudt 等, 2018)	83.69	71.96	98.49	76.35	61.75	88.52
	DSIFN(Zhang 等, 2020)	84.45	73.12	98.03	77.26	62.95	90.51
	DTCDCSCN(Liu 等, 2021)	87.67	77.40	98.66	76.01	61.55	89.69
	STANet(Chen 和 Shi, 2020)	87.54	77.84	98.83	67.93	51.43	86.72
	SNUNet(Fang 等, 2022)	87.78	77.98	98.62	75.57	61.79	89.01
	ELGC-Net(Noman 等, 2024)	91.02	83.22	99.03	78.52	64.32	89.26
	USSFC-Net(Lei 等, 2023)	91.04	83.24	99.06	79.66	65.34	89.62
	LRNet(Zhong 等, 2024)	91.08	83.22	99.10	80.21	67.41	90.43
Transformer	FIFLNet(本文)	91.51	84.35	99.14	82.19	69.76	91.99
	BiT(Chen 等, 2022)	89.33	80.72	98.94	77.59	63.39	89.27
	Changeformer(Bandara 等, 2022b)	90.40	82.48	99.04	78.79	64.78	89.92
生成模型	DDPM-CD(Bandara 等, 2022a)	90.90	83.31	99.09	80.61	67.52	90.82

注:加粗字体表示各列最优结果。

表 2 不同变化检测方法的参数量、每秒浮点运算次数、
单个 epoch 训练时间和 F1 score 的对比
Table 2 Comparison of number of parameters, FLOPs,
single epoch training time and F1 score of different
change detection methods

方法	参数 量/M	FLOPs /G	时间 /min	F1/%	
				LEVIR- CD	SYSU- CD
DTCDCSCN	41.07	7.21	16.45	87.67	76.01
DSIFN	35.99	5.64	12.38	84.45	77.26
STANet	16.93	18.07	6.91	87.54	67.93
SNUNet	12.03	54.83	7.36	87.78	75.57
BiT	3.55	8.75	4.71	89.33	77.59
Changeformer	41.03	13.88	22.13	90.4	78.79
DDPM-CD	120.33	71.36	1 671.23	90.9	80.61
ELGC-Net	10.57	12.36	9.19	91.02	78.52
USSFC-Net	1.52	4.86	3.01	91.04	79.66
LRNet	48.71	92.23	46.51	91.08	80.21
FIFLNet(本文)	0.56	1.66	1.73	91.51	82.19

注:加粗字体表示各列最优结果。

数(floating point operations per second, FLOPs)和单个 epoch 训练时间 3 个维度对现有模型进行了对比,以展示所提模型在资源消耗方面的优势。从表 2 中的定量评价指标可以看出,现有基于 CNN 和注意力机制的模型存在参数量大、每秒浮点运算次数高和单个 epoch 训练时间长的问题。基于 Transformer 的模型和生成类模型也存在类似问题。而轻量级模型如 BiT、USSFC-Net 等,虽然在资源消耗上表现良好,但在变化区域检测的精确度上却不尽如人意。本文所提方法在保持更少的参数、更低的每秒浮点计算次数和更短的单个 epoch 训练时间的同时,实现了 SOTA 的变化检测结果。综上所述,本文方法的性能优于对比网络。

3.2 消融实验

3.2.1 浅层跳跃连接消融实验

为验证浅层跳跃连接的有效性,进行消融实验,包括无跳跃连接、使用普通跳跃连接及使用浅层跳跃连接 3 种情况的模型。如表 3 所示,使用局部特征的模型在变化检测任务中的评价指标上表现出更优的结果。其中, No skip 代表未使用跳跃连接, Skip

代表普通跳跃连接,SD-Skip代表浅层跳跃连接。从表3得出,在不使用跳跃连接时,本模型在LEVIR-CD和SYSU-CD数据集上的F1分别为90.91%和80.84%。然而,在上采样过程中缺少局部特征,容易导致模型整体出现过拟合,进而降低在测试数据中的精度。为缓解过拟合问题,模型加入了跳跃连接,以改善模型梯度的反向传播,减少信息丢失,使得深层网络的训练更加稳定。在两个数据集上的F1分别为91.23%和81.91%。虽然模型性能得到了一定提升,但仍缺少对数据局部特征的关注;为使模型更关注局部特征,模型加入了浅层跳跃连接,以达到在跳跃连接的基础上,增加模型对局部细节特征的关注。在两个数据集上的F1则分别提升至91.51%和82.19%,过拟合问题也得到了进一步的改善,各项评价指标均有提高。这些结果证实了浅层跳跃连接的有效性。

表 3 跳跃连接消融实验

Table 3 Ablation experiment of skip-connection

/%

模型	LEVIR-CD			SYSU-CD		
	F1	IoU	OA	F1	IoU	OA
No skip	90.91	83.11	98.76	80.84	66.78	90.11
Skip	91.23	83.79	99.01	81.91	68.45	91.73
SD-skip	91.51	84.35	99.14	82.19	69.76	91.99

注:加粗字体表示各列最优结果。

3.2.2 通道、像素交互模块消融实验

对通道、像素交互模块进行消融实验,结果如表4所示。表4中,“No exchange”表示模型未使用通道和像素交互模块,“Exchange”表示模型使用通道和像素交互模块。特征交互模块能够捕获不同特征之间的相互作用和依赖关系,这对于准确区分地物变化目标的边界非常关键。与未使用该模块的模型相比,使用该模块的模型在两个数据集上的F1分别高0.21%和0.18%。从实验数据中得出,通过在模型中引入此模块,可有效提升模型对双时相数据中关联信息的学习能力,使模型能够有效地利用时空信息。实验数据表明,相比于未引入通道和像素交互模块的模型,引入该模块的模型在学习变化区域特征方面具有更强的学习能力,能够准确区分地物变化目标边界,降低误检率。

表 4 特征交互模块消融实验

Table 4 Ablation experiment of feature exchange

/%

模型	LEVIR-CD			SYSU-CD		
	F1	IoU	OA	F1	IoU	OA
No exchange	91.30	84.16	99.01	82.01	69.52	91.65
Exchange	91.51	84.35	99.14	82.19	69.76	91.99

注:加粗字体表示各列最优结果。

3.3 模块组合实验分析

为寻找通道和像素交互模块的最优嵌入组合,在模型其他条件不变的前提下,进行以下6种模型组合实验,如表5所示。6种组合分别是:1)在数据输入层插入像素交互模块,第1层和第2层下采样操作后插入通道交互模块(spatial-channel-channel, SPCHCH);2)在数据输入层和第1层下采样后插入像素交互模块,第2层下采样操作后插入通道交互模块(spatial-spatia-channel, SPSPCH);3)在数据输入层和第2层下采样后插入通道交互模块,第1层下采样后插入像素交互模块(channel-spatial-channel, CHSPCH);4)在数据输入层后插入通道交互,第1层和第2层下采样后插入像素交互模块(channel-spatial-spatial, CHSPSP);5)在数据输入层和第2层下采样后插入像素交互模块,第1层下采样后插入通道交互模块(spatial-channel-spatial, SPCHSP);6)在数据输入层和第1层下采样后插入通道交互模块,第2层下采样后插入像素交互模块(channel-channel-spatial, CHCHSP)。

表 5 各种模型组合在LEVIR-CD数据集上的
定量评价结果

Table 5 Quantitative evaluation of the results of various
model combinations on the LEVIR-CD dataset

/%

模型	F1	IoU	OA
SPCHCH	91.49	84.31	99.14
SPSPCH	91.38	84.12	99.13
CHSPCH	91.41	84.18	99.14
CHSPSP	91.35	84.08	99.14
SPCHSP	91.17	83.77	99.12
CHCHSP	91.51	84.35	99.14

注:加粗字体表示各列最优结果。

由于 LEVIR-CD 数据集具有分辨率高、数据质量高等特点,本小节选取 LEVIR-CD 数据集进行实验,定量评价指标结果如表 5 所示,部分可视化结果如图 6 所示。从表 5 中得出,组合 CHCHSP 的模型效果最好,在 LEVIR-CD 数据集上 F1、IoU 和 OA

最高。鉴于在 LEVIR-CD 数据集上的定量评价指标结果,本文采用组合 CHCHSP 作为本文优选模型。从图 6 可以直观得出组合 CHCHSP 具有更强大的时空信息利用能力和对变化区域的识别能力。

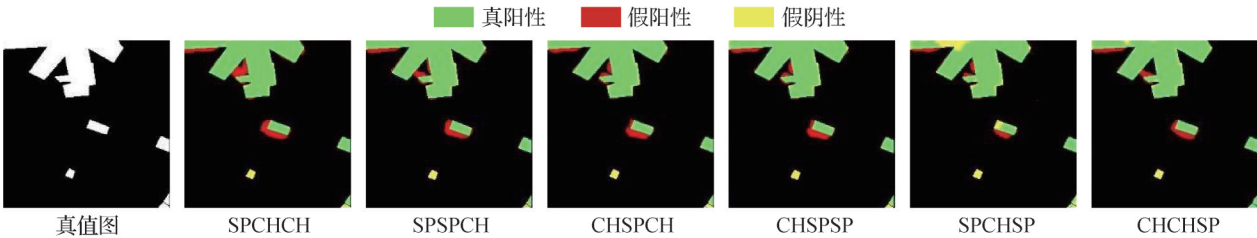


图6 不同模块组合在 LEVIR-CD 数据集的可视化结果

Fig. 6 Visualization of the results of the combination of different modules on the LEVIR-CD dataset

4 结 论

针对深度学习遥感影像变化检测任务中普遍存在的模型复杂度高、参数过多、精度欠佳等问题,本文提出了一种融合特征交互和特征融合的轻量级变化检测网络遥感影像变化检测方法(FIFLNet)。该方法的核心思想是通过优化网络结构和算法设计,提高模型的性能和效率。首先,FIFLNet 采用 EfficientNet 作为特征提取网络,以降低训练成本和参数数量。该网络的强大模型放缩能力能有效扩大模型的感受野。然后,通过设计浅层跳跃连接使模型聚焦于数据的局部特征,改善小目标漏检问题。在跳跃连接过程中还引入了通道和像素交互模块,能实现双时相特征的交互和融合,使得模型充分学习数据的时空特征,从而降低网络的误检率。接着,本文还设计了特征融合分组卷积模块。该模块能够有效融合下采样阶段的各级特征,并通过分组卷积降低计算开销,进一步细化特征以提高数据的利用率。此外,本文还通过设计融合上采样模块将网络中的浅层特征和深层特征进行融合上采样,以浅层特征的细节和纹理信息来改善预测数据边界的干扰。最后,利用像素级分类器生成最终的变化图。在两个遥感影像数据集上与多种 SOTA 方法进行实验对比可知,FIFLNet 能够显著提高对变化区域的识别能力,在各种环境和地形的变化检测中均获得较高的检测精度。

尽管 FIFLNet 在当前的实验中表现出色,但模

型训练仍然依赖于大量的标注数据,这对于高分辨率和动态变化的遥感影像检测可能存在挑战。为此,未来将探索更多无监督或半监督的变化检测方法,以减少方法对标注数据的依赖,提升模型的可扩展性和适应性。另一方面,计划借鉴相关的遥感特征提取和融合方法,通过引入级联注意模块等来更有效地捕捉和利用高光谱图像的频谱和空间信息,从而提高变化检测的准确性和效率。此外,还将探索如何优化模型结构和参数,以进一步降低计算复杂性,使模型更适合于实际应用。

参考文献(References)

Bandara W G C, Nair N G and Patel V M. 2022a. DDPM-CD: denoising diffusion probabilistic models as feature extractors for change detection [EB/OL]. [2024-05-28] <https://arxiv.org/pdf/2206.11892.pdf>

Bandara W G C and Patel V M. 2022b. A Transformer-based Siamese network for change detection//Proceedings of 2022 IEEE International Geoscience and Remote Sensing Symposium. Kuala Lumpur, Malaysia: IEEE: 207-210 [DOI: 10.1109/IGARSS46834.2022.9883686]

Chen H and Shi Z W. 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sensing, 12(10): #1662 [DOI: 10.3390/rs12101662]

Chen H, Qi Z P and Shi Z W. 2022. Remote sensing image change detection with Transformers. IEEE Transactions on Geoscience and Remote Sensing, 60: #5607514 [DOI: 10.1109/TGRS. 2021. 3095166]

Daudt R C, Le Saux L and Boulch A. 2018. Fully convolutional Siamese networks for change detection//Proceedings of the 25th IEEE International Conference on Image Processing. Athens, Greece: IEEE;

- 4063-4067 [DOI: 10.1109/ICIP.2018.8451652]
- Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 248-255 [DOI: 10.1109/CVPR.2009.5206848]
- Du P J and Liu S C. 2012. Change detection from multi-temporal remote sensing images by integrating multiple features. *Journal of Remote Sensing*, 16(4): 663-677 (杜培军, 柳思聪. 2012. 融合多特征的遥感影像变化检测. *遥感学报*, 16(4): 663-677) [DOI: 10.11834/jrs.20121168]
- Fang S, Li K Y and Li Z. 2023. Changer: feature interaction is what you need for change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5610111 [DOI: 10.1109/TGRS.2023.3277496]
- Fang S, Li K Y, Shao J Y and Li Z. 2022. SNUNet-CD: a densely connected Siamese network for change detection of VHR images. *IEEE Geoscience and Remote Sensing Letters*, 19: #8007805 [DOI: 10.1109/LGRS.2021.3056416]
- He K M, Zhang X Y, Ren S Q and Sun J. 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1026-1034 [DOI: 10.1109/ICCV.2015.123]
- Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324 [DOI: 10.1109/5.726791]
- Lei T, Geng X Z, Ning H L, Lv Z Y, Gong M G, Jin Y C and Nandi A K. 2023. Ultralightweight spatial-spectral feature cooperation network for change detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #4402114 [DOI: 10.1109/TGRS.2023.3261273]
- Li C Y, Zhang B, Hong D F, Zhou J, Vivone G, Li S T and Chanussot J. 2024. CasFormer: cascaded Transformers for fusion-aware computational hyperspectral imaging. *Information Fusion*, 108: #102408 [DOI: 10.1016/j.inffus.2024.102408]
- Liu Y, Pang C, Zhan Z Q, Zhang X M and Yang X. 2021. Building change detection for remote sensing images using a dual-task constrained deep Siamese convolutional network model. *IEEE Geoscience and Remote Sensing Letters*, 18(5): 811-815 [DOI: 10.1109/LGRS.2020.2988032]
- Loshchilov I and Hutter F. 2019. Decoupled weight decay regularization [EB/OL]. [2024-05-28]. <https://arxiv.org/pdf/1711.05101.pdf>
- Noman M, Fiaz M, Cholakkal H, Khan S and Khan F S. 2024. ELGC-Net: efficient local-global context aggregation for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #4701611 [DOI: 10.1109/TGRS.2024.3362914]
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z M, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J J and Chintala S. 2019. PyTorch: an imperative style, high-performance deep learning library//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 8026-8037
- Shi Q, Liu M X, Li S C, Liu X P, Wang F and Zhang L P. 2022. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5604816 [DOI: 10.1109/TGRS.2021.3085870]
- Silva C, Marino A, Lopez-Sanchez J M and Cameron I. 2020. Agricultural fields monitoring with multi-temporal polarimetric SAR (MT-POLSAR) change detection//Proceedings of 2020 IEEE International Geoscience and Remote Sensing Symposium. Waikoloa, USA: IEEE: 4554-4557 [DOI: 10.1109/IGARSS39084.2020.9324626]
- Sui H G, Feng W Q, Li W Z, Sun K M and Xu C. 2018. Review of change detection methods for multi-temporal remote sensing imagery. *Geomatics and Information Science of Wuhan University*, 43(12): 1885-1898 (眭海刚, 冯文卿, 李文卓, 孙开敏, 徐川. 2018. 多时相遥感影像变化检测方法综述. *武汉大学学报(信息科学版)*, 43(12): 1885-1898) [DOI: 10.13203/j.whugis.20180251]
- Tan M X and Le Q V. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//Proceeding of the 36th International Conference on Machine Learning. Long Beach, USA: PMLR: 6105-6114
- Tan M X and Le Q V. 2021. EfficientNetV2: smaller models and faster training [EB/OL]. [2024-05-28]. <https://arxiv.org/pdf/2104.00298.pdf>
- Ulyanov D, Vedaldi A and Lempitsky V. 2017. Instance normalization: the missing ingredient for fast stylization [EB/OL]. [2024-05-28]. <https://arxiv.org/pdf/1607.08022.pdf>
- Vaswani S, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Wang R F, Wang L, Li C, Huo C L and Chen J W. 2023. IIQ-CNN-based cross-domain change detection of SAR images. *Journal of Image and Graphics*, 28(7): 2208-2220 (王蓉芳, 王良, 李畅, 霍春雷, 陈佳伟. 2023. 整型推理量化CNN的SAR图像跨域变化检测. *中国图象图形学报*, 28(7): 2208-2220) [DOI: 10.11834/jig.211159]
- Yan C Q, Zhang Y G, Zhang Q L, Yang Y M, Jiang X Y, Yang Y Q and Wang B Y. 2022. Privacy-preserving online AutoML for Domain-Specific face detection//Proceeding of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 4124-4134 [DOI: 10.1109/CVPR52688.2022.00410]

- Yang S, Li M, Peng Z G and Feng C. 2009. A novel multi-band remote sensing image change detection algorithm. *Journal of Image and Graphics*, 14(4): 572-578 (杨胜, 李敏, 彭振国, 冯春. 2009. 一种新的多波段遥感影像变化检测方法. *中国图象图形学报*, 14(4): 572-578) [DOI: 10.11834/jig.20090402]
- Zhang C X, Yue P, Tapete D, Jiang L C, Shangguan B Y, Huang L and Liu G C. 2020. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166: 183-200 [DOI: 10.1016/j.isprsjprs.2020.06.003]
- Zhang K Y, Fu X K, Lv X L and Yuan J L. 2021. Unsupervised multi-temporal building change detection framework based on cosegmentation using time-series SAR. *Remote Sensing*, 13 (3) : #471 [DOI: 10.3390/rs13030471]
- Zhong H, Wu C and Xiao Z Q. 2024. LRNet: change detection of high-resolution remote sensing imagery via strategy of localization-then-refinement [EB/OL]. [2024-05-28].
<https://arxiv.org/pdf/2404.04884.pdf>
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. U-Net++: a nested U-Net architecture for medical image segmenta-

tion//*Proceedings of 2018 Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介

王仁芳,男,教授,主要研究方向为数字图像处理、机器学习和遥感智能解译。E-mail:renfang_wangac@126.com

邱虹,通信作者,女,副教授,主要研究方向为图像处理和模式识别。E-mail:qiu hong@z wu.edu.cn

杨梓健,男,硕士研究生,主要研究方向为遥感智能解译。E-mail:a1057697923@gmail.com

王峰,男,讲师,主要研究方向为机器学习、地球物理反演和遥感图像处理。E-mail:wangf_721@zju.edu.cn

高广,男,高级工程师,主要研究方向为多视几何与三维重建、遥感解译和数字孪生。E-mail:gaoguang@popsmart.com

吴敦,男,高级工程师,主要研究方向为测绘与地理信息和数字孪生。E-mail:wudun@popsmart.com