

多时态 BIT 遥感图像建筑物的变化检测^①



牟彦霖, 刘向阳

(河海大学 数学学院, 南京 210098)

通信作者: 刘向阳, E-mail: liuxy@hhu.edu.cn

摘要: 针对来自相同地理空间的不同时刻遥感图像之间的季节性和光度变化(色差)等因素所引起的干扰, 提出了多时态-BIT 遥感图像变化检测方法. 该方法引入了过去多个不同时刻的遥感图像, 融合当前遥感图像与过去时态遥感图像两两变化检测的结果, 该方法有助于排除季节性和光度变化引起的误报, 提高了变化检测的准确性; 并且利用过去多个不同时刻的遥感图像, 进一步消除非目标建筑变化的影响, 其变化点像素差值引入作为损失函数正则化项, 从而进一步提高变化检测的鲁棒性和可靠性. 本文以三时态(3个不同时刻的遥感图像)为例, 使用了遥感图像建筑物变化数据集进行了实验. 实验结果表明, 多时态-BIT 方法相对于仅考虑两个时态的变化检测方法, 在遥感图像建筑物变化检测任务中表现出更好的效果.

关键词: 深度学习; 变化检测; 注意力机制; 孪生网络; 并行结构; Transformer; 卷积神经网络; 正则化

引用格式: 牟彦霖, 刘向阳. 多时态 BIT 遥感图像建筑物的变化检测. 计算机系统应用, 2024, 33(6): 185-191. <http://www.c-s-a.org.cn/1003-3254/9546.html>

Change Detection for Buildings in Multitemporal BIT Remote Sensing Images

MU Yan-Lin, LIU Xiang-Yang

(College of Science, Hohai University, Nanjing 210098, China)

Abstract: A remote sensing image change detection method of multi-temporal binary change detection based on image transformation (BIT) is proposed to address issues related to seasonal and radiometric variations (color discrepancies) between remotely sensed images acquired at different times but from the same geographic area. This method incorporates remote sensing images from multiple past time points and combines the results of pairwise change detection between the current image and the past temporal images to obtain a stable change detection outcome. This method helps mitigate false alarms caused by seasonal and radiometric variations, thereby enhancing the accuracy of change detection. Multiple remote sensing images from different time points in the past are utilized to eliminate the influence of non-target building changes. The pixel difference value of change points is introduced as a regularization term in the loss function, further improving the robustness and reliability of change detection. In this study, a three-temporal (three images from different time points) example is provided, and experiments are conducted with a remote sensing image dataset of building changes. The experimental results demonstrate that the multi-temporal BIT method outperforms change detection methods that only consider two temporal images in the task of remote sensing image change detection.

Key words: deep learning; change detection; attention mechanism; Siamese network; parallel structure; Transformer; convolutional neural network (CNN); regularization

① 基金项目: 国家自然科学基金 (41830110)

收稿时间: 2023-12-18; 修改时间: 2024-01-23; 采用时间: 2024-02-23; csa 在线出版时间: 2024-04-30

CNKI 网络首发时间: 2024-05-07

近年来,随着大规模应用高分辨率遥感影像获取设备的卫星投入民用领域,研究关于海量高分辨率和极高分辨率卫星遥感图像的分析成为一个备受关注的领域.在民用领域,遥感图像主要在土地资源^[1]规划、城市扩展、土地利用以及森林管理等方面发挥着重要作用.遥感图像^[2]具有高空间分辨率和高时间分辨率的特点,但受地球系统的开放性、复杂性和不确定性的影响,因此,如何有效地识别和提取建筑物等目标成为变化检测的重要研究领域之一.

神经网络技术和组件在变化检测任务中的应用取得了显著进展.这些技术旨在提取更深层次的特征表示,从而改善变化检测性能.首先,U-Net^[3]被引入作为基准模型,通过图像分割,编码和解码全过程线性表达,缺点是全过程都是线性压缩,丢失不少特征.随后Siamese^[4-6]网络成为标准的变化检测方法,它采用CGB(context guide block)模块构建孪生的骨干网络,另外增加金字塔特征自底而上的融合路径,为任务提供了一种有效的方法.林娜等^[7]运用空洞残差卷积的思想提取浅层特征,将提取到的特征与深层特征进行融合,有效提高了遥感图像中飞机的检测精度.姚艳清等^[8]使用双尺度特征融合模块,缓解深层信息的丢失问题,有效提高了多尺度遥感目标的检测能力.为了进一步提高变化检测^[9-11]的性能,进行了大量工作,主要集中在深度特征提取和特征细化方面.其中一些方法包括使用金字塔模型来提取多尺度特征^[12]、采用深度监督来增强浅层特征的和识别能力^[13,14],以及使用注意力机制^[15]来细化特征以获取更好的表示,闵钰魁等^[15]构建高时空分辨率的归一化植被差异指数(NDVI)数据集,并利用密集时间序列遥感影像的互花米草清除进行动态监测;张久丹等^[16]以多时相影像为数据源,利用多尺度膨胀卷积模块改进D-UNet(deformable U-Net)网络,以提高模型对高分辨率遥感影像的多尺度学习能力.

上述方法基于双时态和多时态的遥感图像变化,如图1所示,T1和T2分别代表当前时态和过去时态的遥感图像.这些方法致力于在特征提取方面进行精炼,以更好地生成特征图来预测最终的变化检测结果.然而,它们未充分考虑遥感图像本身的特有因素,如色差和季节性变化等,可能会对结果产生一定的影响.因此,在双时态遥感变化检测BIT(binary-image Transformer)^[17]的基础上,我们提出了一种多时态-BIT遥感

变化检测模型,在T1和T2之间,引入了T3(过去其他时刻的图像)或过去多个不同时刻的图像,将它(或它们)与T1和T2两个时态图像进行组合,即最新时刻图像与过去时刻图像两两变化检测结果加权后生成更加稳定的特征图,过去时刻图像两两组合,其非目标建筑变化像素差作为损失函数的正则化项,以降低非目标建筑变化因素对模型参数的影响,从而降低总体损失,减小了外部因素的干扰影响,确保变化检测结果更加可靠.

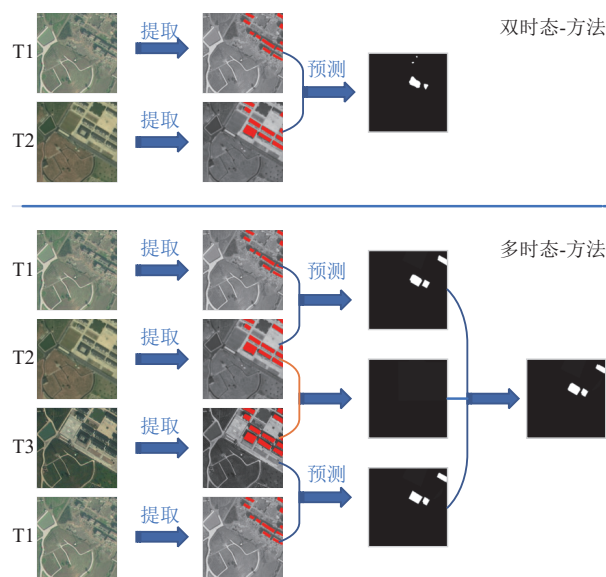


图1 多时态遥感变化检测示意图

1 多时态 BIT 方法原理

1.1 原始 BIT 相关原理.

在基于Transformer的遥感图像变化检测^[13]研究中,深入探讨了Transformer神经网络在遥感图像的二进制变化检测任务中所具有的潜在应用潜力,并提出了一种BIT的方法.而Transformer于2017年首次引入,已广泛应用于自然语言处理(natural language processing)领域,以解决序列到序列的任务,同时轻松处理远程依赖关系.在计算机视觉(computer vision)领域采用Transformer是最近的一个趋势.由于Transformer具有较强的表示能力,基于它的模型在图像分类、分割^[18]、目标检测^[19]、图像生成^[20]等各种视觉任务中表现出与卷积模型相当甚至更好的性能.

该方法能够高效地在时空中建模全局的语义关系,有助于更好地表示感兴趣区域的变化特征.将BIT整合到常规的CD(change detection)管道中,以充分利用

卷积神经网络和 Transformer 神经网络的优势. 该模型的工作流程首先使用卷积块来提取每个输入图像的特征映射, 然后将这些特征映射输入到 BIT 中, 以生成增强的双时间特征. 最后, 得到的特征映射被送入预测头部, 用于生成像素级的变化预测. 这项研究的关键观点在于 BIT 能够学习并关联高级语义概念, 具备全局上下文建模的能力, 并将这些信息反馈回原始的双时间

特征. 这意味着 BIT 能够理解图像中的更高级别的语义信息, 例如物体、结构和空间关系, 并将这些信息应用于变化检测任务, 以提高性能和准确性.

1.2 多时态 BIT 的网络结构

其网络结构如图 2 所示, 可划分为 3 个部分: CNN Backbone 提取主网络模块、三时态图像自注意力变化网络模块和差异计算预测模块

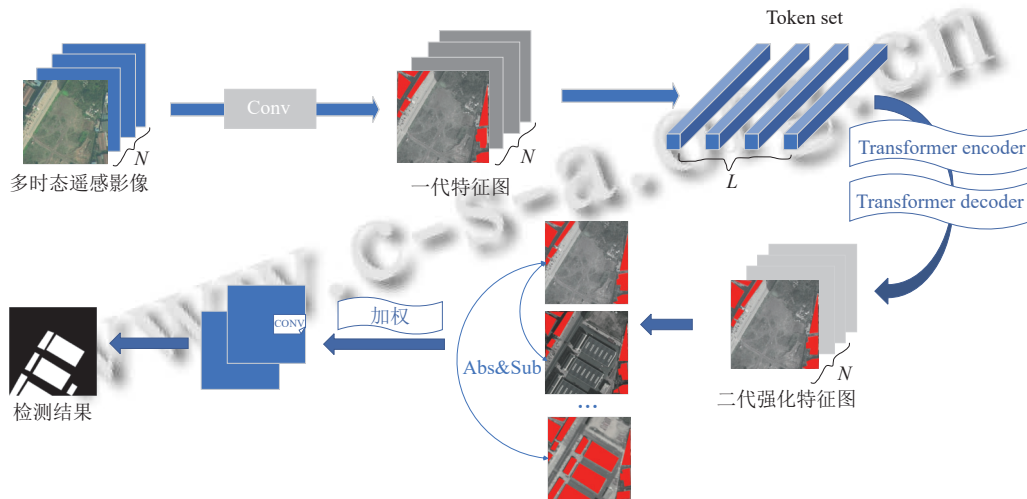


图 2 多时态 BIT 变化检测网络结构

1.3 T-BIT (三时态-BIT) 的网络结构为例

1.3.1 CNN Backbone 提取主网络

如图 3 所示, 输入图像首先通过 CNN 网络结构进行处理. 这个 CNN 阶段包含了 4 个卷积层, 每一层都负责从输入图像中提取不同级别的特征. 卷积操作通过滤波器检测图像中的各种特征, 从较低级别的边缘和纹理到更高级别的对象和结构. 经过 4 次卷积操作后, 得到了第 1 代特征图. 接下来, 第 1 代特征图通过 1 次额外的卷积操作进行编码. 编码可能会将特征图映射到更高维度的表示或者提取更抽象的特征. 然后, 对编码后的特征图进行分割操作, 以获得 token 组.

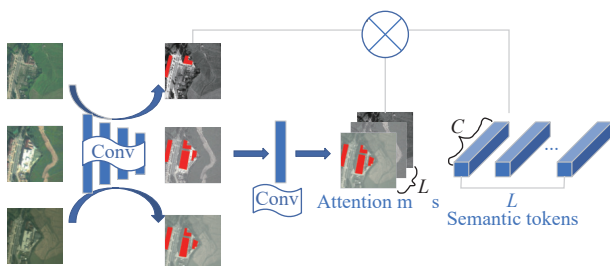


图 3 CNN Backbone 提取网络结构

但对于传统的 CNN Backbone 提取网络结构的改进在遥感图像变化检测任务中是非常重要的, 特别是为了提高性能和适应复杂场景. 本文对 ResNet18^[21]网络结构进行微小的修改, 特别是对最后一个 stages 的下采样改为 1. 这个改进保持了空间信息的完整性, 防止了特征图过小的问题, 同时也增强了感受野, 使模型更好地理解图像中的上下文信息. 此外, 在 ResNet18 后面接上双线性插值和逐点卷积, 增加了特征图的尺寸和改变了特征图的维度, 这有助于提高模型的分辨率和表达能力.

1.3.2 三时态图像自注意力变化网络

如图 4 所示, Transformer 主要包括 token 分析器、编码器和解码器, 用于处理一代特征图. Token 分析器将前面提取 A、B、C 特征图, 送入同一词元分析器网络分别进行卷积工作, 映射成词元集.

编码器工作原理与 ViT (vision Transformer)^[22]相似, 将上述输出出来的 3 个词元集串联成一个词元集进行编码、范数计算得到 Transformer 中的 Q (query)、K (key)、V (value) 矩阵. 此外, 词元集中的词元两两

组合生成新的 token, 而过去不同时刻的遥感图像之间的差异信息用于作为损失函数的正则化项, 以更新损失函数. 然后通过多层感知机模块和激活函数, 将输出的词元集切片还原成富含高级语义信息的词元集 T_1 、 T_2 .

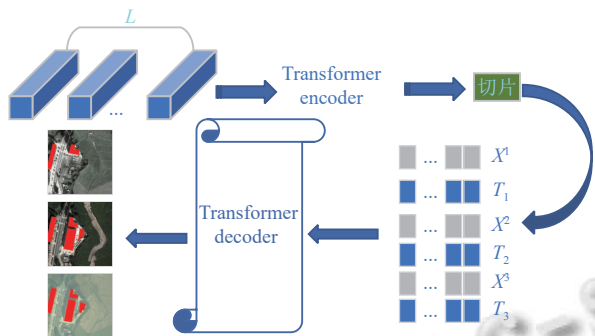


图4 三时态自注意力机制网络变化结构

解码器就是将上述词元集与原始影像特征向量 X^1 、 X^2 每个像素之间的关系结合起来, 得到优化后的特征向量 X_{new}^1 、 X_{new}^2 . 用于生成同等大小的新特征图.

1.3.3 差异计算预测

在预测模块如图5所示, 从模型中获得3幅特征图, 对这些特征图进行两两相减的操作, 再对其结果取绝对值, 以精确计算图像之间的差异. 接下来, 对计算得到的差异图像进行加权操作. 加权^[23]后的差异图像经过一系列卷积操作, 包括卷积层、激活函数等, 以深化特征的处理和提取. 最后, 通过一个分类器, 包括两个卷积核大小为 3×3 的卷积层, 以获得最终的预测结果, 即二值化的变化图.

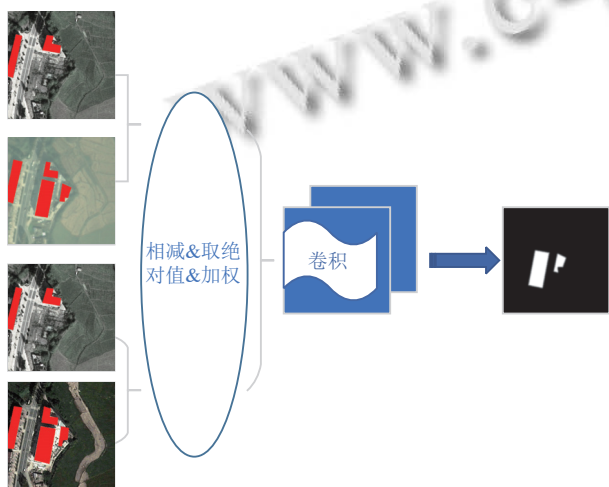


图5 差异计算预测结构

1.4 损失函数

1.4.1 BIT 的损失函数

损失函数是在训练深度学习模型时使用的一种度量, 用于衡量模型的预测与实际值之间的差异. 在遥感图像变化检测中, 损失函数的选择对模型的性能至关重要. BIT 文中采用最小化交叉熵损失函数来优化网络参数.

交叉熵损失函数 (cross-entropy loss function), 也称为 Softmax 损失函数, 是常用的损失函数, 可以用于多类别任务, 在两类别任务中其计算公式表示为:

$$L = \frac{1}{H_0 W_0} \sum_{h=1, w=1}^{H, W} l(P_{hw}, Y_{hw}) \quad (1)$$

其中, H_0 是图像的原始长度, W_0 是图像的原始宽度, P_{hw} 是图像位于 (h, w) 预测值, Y_{hw} 是图像位于 (h, w) 上像素的标签, $l(P_{hw}, Y_{hw}) = -\log(P_{hw})$ 为交叉熵损失.

1.4.2 T-BIT 的损失函数

为了提高模型的精度, 本文采用的损失计算方法旨在降低非目标建筑物变化对模型性能的负面影响, 并通过降低总体损失来改进模型性能. 在加权交叉熵损失函数的基础上增加了正则化函数损失, 其损失函数定义如下:

$$L = L_1 + L_{1-norm} \quad (2)$$

其中, L_1 代表加权交叉熵损失函数, L_{1-norm} 代表 L_1 正则化.

加权交叉熵损失函数, 是在交叉熵损失函数上进行一次加权操作, 使损失根据权重来关注区域, 其计算公式表示为:

$$L_1 = \frac{1}{2H_0 W_0} \sum_{h=1, W=1}^{H, W} (l(P_{hw}^{AB} + P_{hw}^{AC}, Y_{hw})) \quad (3)$$

其中, H_0 是图像的原始长度, W_0 是图像的原始宽度, P_{hw}^{AB} 是图像 A 和 B 位于 (h, w) 预测值, P_{hw}^{AC} 是图像 A 和 C 位于 (h, w) 预测值, Y_{hw} 是图像位于 (h, w) 上像素的标签.

L_{1-norm} 是正则化函数, 使 loss 下降速度变慢, 模型训练表现更平滑, 防止模型过拟合. 其计算公式表示为:

$$L_{1-norm} = \lambda |M_{BC}| \quad (4)$$

其中, λ 是正则化系数, M_{BC} 是过去时刻图像 B 和 C 经编码之后的非建筑物变化的像素差值.

2 实验结果及分析

2.1 构建数据集

为了使模型更贴近实际目标,本文所用数据集是来源于某一地区卫星拍摄的真实数据,并进行数据增强,从中挑选出720张具有干扰因素的图像,图像尺寸大小均为256×256像素,其中576张图像为训练集,72张图像为验证集,72张图像为测试集,具有RGB通道。

本实验是在PyTorch深度学习框架下进行,其中使用的PyTorch 1.10.0.代码使用语言为Python 3.7版本.实验过程中使用GPU加速,模型训练过程中,根据以往神经网络训练经验,采用参数学习率Learning_rate为0.1,epochs为100,batch-size为16进行训练,将T-BIT方法与原始的BIT方法比较实验,实验环境等配置均相同。

2.2 评价指标选取

变化检测实际上是一个像素级二分类任务,一般采用F1、Precision、Recall、IoU、OA作为评价变化检测效果的指标.其具体计算为:

$$Precision = TP / (TP + FP) \quad (5)$$

$$Recall = TP / (TP + FN) \quad (6)$$

$$IoU = TP / (TP + FN + FP) \quad (7)$$

$$OA = (TP + TN) / (TP + TN + FN + FP) \quad (8)$$

TP表示将正类预测为正类,FP表示将负类预测为正类,TN表示将负类预测为负类,FN表示将正类预测为负类,构成变化检测混淆矩阵。

2.3 正则化系数λ的选取

为了在训练过程中,更好地泛化模型的能力,根据前后项数量级,我们选取了λ的范围为0-1之间,经过测试,如图6显示在λ=0.1时,有明显的峰值,且模型的泛化能力达到最优。

2.4 对比实验

图7中,(a)是最新时态的图像,(b)是中间时态的图像,(c)是最初时态的图像,(d)是真实标签,(e)是原始BIT的检测结果,(f)是T-BIT的检测结果.图7和表1表明T-BIT方法相对于原始BIT方法在定性和定量方面都表现出更好的性能.在此数据集下,T-BIT方法在各项指标上均优于原始BIT方法.T-BIT方法在图像变化检测任务中具有以下优势:T-BIT方法能

够更准确地识别建筑物目标的变化.这意味着它在捕捉建筑物变化方面具有更高的精确度和找回率.相对于原始BIT方法,T-BIT方法减少了漏检的现象,即更全面地检测到了建筑物变化。

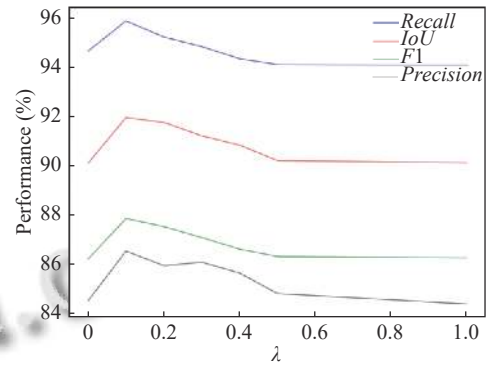


图6 正则化系数选取影响

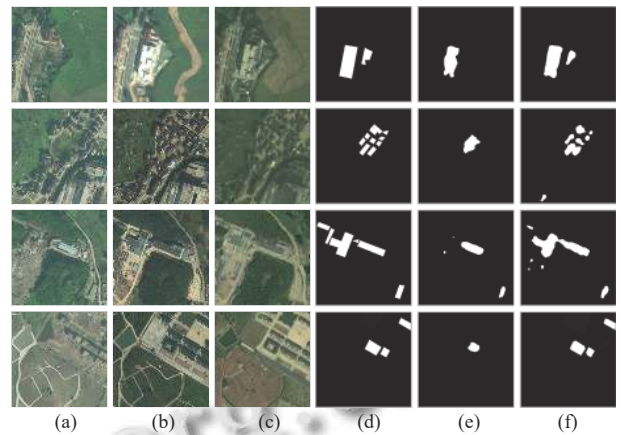


图7 T-BIT和原始BIT方法对比

表1 T-BIT与BIT精度对比(%)

方法	F1	Precision	Recall	OA	IoU
BIT	78.86	80.54	86.39	90.24	76.85
T-BIT	86.86	85.86	89.46	93.54	82.64

2.5 多时态建筑物提取指标对比

基于相同的数据集,以文中图7的数据作为提取影像,结果取平均值,评估各模型的提取结果.其结果表现见表2。

表2 多时态特征提取精度评估(%)

模型	OA	F1	IoU	Precision
T-BIT	95.46	91.38	87.25	92.14
D-ABNet	91.12	88.36	84.25	88.47
D-UNet	93.28	90.25	86.36	90.36

从指标来看,T-BIT在OA、F1、IoU、Precision表现最优,更能准确反映建筑物结构特征和分布.D-

ABNet 在模型边缘平滑区域处理不错,但在建筑物变化区域细节信息损失较大,导致漏检现象较多,而 D-UNet 对于建筑物提取边缘平滑性较好,由于数据存在光谱混淆和非目标建筑物变化干扰的问题,故在部分场景下存在误检现象较多。而 T-BIT 在准确性上有好的表现,显示出较强的鲁棒性和泛化能力,对于建筑物和非建筑物区分效果较好,是因为利用数据间融合特征和相互约束,进一步地处理了影像中非目标建筑物和光谱等干扰,从而达到效果最佳。

2.6 消融实验

为了获得性能最优的 T-BIT 模型,本文控制一些参数的设置,仍采用 $F1$ 得分、 $Precision$ 、 $Recall$ 、 IoU 和 OA 这 5 个评价指标,通过调整学习率与训练轮次的比例 M 和 $token_length$ 大小 N , 获得一个较稳定的模型。结果见表 3 和表 4。

表 3 学习率与训练轮次比例 M 消融实验 (%)

M	$F1$	$Precision$	$Recall$	IoU	OA
0.3	87.62	87.23	89.68	82.64	95.32
0.5	85.98	86.21	87.35	80.36	93.54
0.7	85.62	85.68	86.78	78.69	90.25

表 4 $token_length$ 大小 N 消融实验结果 (%)

N	$F1$	$Precision$	$Recall$	IoU	OA
2	88.47	86.78	90.34	80.67	95.68
4	88.96	87.82	92.75	82.63	93.54
8	88.26	87.53	88.36	76.67	94.36
16	87.89	86.89	84.32	74.25	90.25

2.7 Transformer 结构深度的选取

如表 5 所示,该结构可以通过单层编码器很好地学习,模型性能与解码器深度大致呈正相关。可能是因为通过考虑上下文丰富的标记,在解码器的每一层之后都对图像特征进行了改进。当解码器深度为 8 时,效果最好。虽然进一步增加解码器深度可能会提高性能,但为了在效率和精度之间进行权衡,我们将编码器深度设置为 1,将解码器深度设置为 8。

表 5 编码器和解码器深度测试 (%)

编码器深度	解码器深度	$F1$	IoU
1	1	87.63	81.05
2	1	88.13	81.39
4	1	87.87	81.13
8	1	87.93	81.12
1	2	87.96	81.01
1	4	88.16	81.69
1	8	88.34	81.78

3 结语

在传统的遥感图像变化检测任务中,通常使用的双时态遥感变化检测框架。本文在 BIT 的基础上,研究了多时态-BIT 框架在遥感图像建筑物变化检测中的效果。在提取特征阶段,采用多个图像进行特征提取以生成特征图,然后在编码解码过程,对其特征图进行二次增强,并引入过去不同时刻的图像的非目标检测变化改进损失函数,减弱干扰因素。在预测阶段,使最新时态与过去时态特征图两两检测变化的结果,融合得到二值化图像。通过在遥感建筑物变化的数据集上的训练和测试与原始 BIT 模型进行了定性和定量的对比研究和与多时态方法建筑物提取对比,研究结果表明:多时态-BIT 方法显著改善了非目标变化的影响,从而提高了建筑物变化检测的精度,即这一方法有效地减轻了图像受色差等因素干扰的问题。

由于时间问题,本文所使用的建筑物变化数据集较小,未来工作可以考虑在较大数据集上进行实验,让模型更好地应用到实际任务中。

参考文献

- Lu D, Mausel P, Brondizio E, et al. Change detection techniques. *International Journal of Remote Sensing*, 2004, 25(12): 2365–2401. [doi: 10.1080/0143116031000139863]
- Hussain M, Chen DM, Cheng A, et al. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013, 80: 91–106. [doi: 10.1016/j.isprsjprs.2013.03.006]
- Ailimujiang G, Jiaermuhamaiti Y, Jumahong H, et al. A Transformer-based network for change detection in remote sensing using multiscale difference-enhancement. *Computational Intelligence and Neuroscience*, 2022, 2022: 2189176.
- Zhang SL, Ang KK, Zheng DZ, et al. Learning EEG representations with weighted convolutional Siamese network: A large multi-session post-stroke rehabilitation study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, 30: 2824–2833. [doi: 10.1109/TNSRE.2022.3209155]
- Guo EQ, Fu XS, Zhu JW, et al. Learning to measure change: Fully convolutional Siamese metric networks for scene change detection. *arXiv:1810.09111*, 2018.
- 唐娉, 刘璇, 金兴, 等. 自适应滤波器的神经网络生成及遥

- 感图像处理新应用. 遥感学报, 2023, 27(7): 1523–1533. [doi: 10.11834/jrs.20232174]
- 7 林娜, 冯丽蓉, 张小青. 基于优化 Faster-RCNN 的遥感影像飞机检测. 遥感技术与应用, 2021, 36(2): 275–284.
- 8 姚艳清, 程臻, 谢星星, 等. 多分辨率特征融合的光学遥感图像目标检测. 遥感学报, 2021, 25(5): 1124–1137. [doi: 10.11834/jrs.20210505]
- 9 梁敏, 汪西莉. 结合超分辨率和域适应的遥感图像语义分割方法. 计算机学报, 2022, 45(12): 2619–2636. [doi: 10.11897/SP.J.1016.2022.02619]
- 10 Chen H, Qi ZP, Shi ZW. Remote sensing image change detection with Transformers. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5607514.
- 11 Pan JP, Cui W, An XY, *et al.* MapsNet: Multi-level feature constraint and fusion network for change detection. International Journal of Applied Earth Observation and Geoinformation, 2022, 108: 102676. [doi: 10.1016/j.jag.2022.102676]
- 12 Shen GJ, Yu KF, Zhang MY, *et al.* ST-AFN: A spatial-temporal attention based fusion network for lane-level traffic flow prediction. PeerJ Computer Science, 2021, 7: e470. [doi: 10.7717/peerj-cs.470]
- 13 Chen J, Yuan ZY, Peng J, *et al.* DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 1194–1206. [doi: 10.1109/JSTARS.2020.3037893]
- 14 Peng DF, Zhang YJ, Guan HY. End-to-end change detection for high resolution satellite images using improved UNet++. Remote Sensing, 2019, 11(11): 1382. [doi: 10.3390/rs11111382]
- 15 闵钰魁, 柯樱海, 韩月, 等. 融合 Sentinel-2 和 GF-1 时序影像的入侵植物互花米草清除动态监测. 遥感学报, 2023, 27(6): 1467–1479. [doi: 10.11834/jrs.20232279]
- 16 张久丹, 王浩宇, 李均力, 等. 基于改进 D-UNet 模型的典型洪泛湿地信息提取. 遥感学报, 2023. [doi: 10.11834/jrs.20233087]
- 17 陈浩, 齐子鹏, 史振伟. 基于 Transformer 的遥感图像变化检测. 遥感学报, 2022, 60(1): 1–14. [doi: 10.11834/j.issn.1007-4619.2022.1.ygxb202201002]
- 18 Li K, Guo D, Wang M. ViGT: Proposal-free video grounding with a learnable token in the Transformer. Science China Information Sciences, 2023, 66(10): 202102. [doi: 10.1007/s11432-022-3783-3]
- 19 Li SJ, Guo Q, Li A. Pan-sharpening based on CNN+ pyramid Transformer by using no-reference loss. Remote Sensing, 2022, 14(3): 624. [doi: 10.3390/rs14030624]
- 20 Zhang L, Ruan LD, Hu AW, *et al.* Multimodal pretraining from monolingual to multilingual. Machine Intelligence Research, 2023, 20(2): 220–232. [doi: 10.1007/s11633-022-1414-4]
- 21 Liu Y, Wang L, Li HF, *et al.* Multi-focus image fusion with deep residual learning and focus property detection. Information Fusion, 2022, 86–87: 1–16.
- 22 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16×16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.
- 23 Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10778–10787.

(校对责编: 张重毅)