

基于多时相 ChangeFormer 的遥感图像建筑物变化检测方法

姜有泽, 刘向阳

(河海大学数学学院, 江苏 南京 211106)

摘要: 针对相同地理空间、不同时相的高分辨率遥感图像之间受季节性变化、气候、光照等干扰因素影响的问题, 提出一种基于多时相 ChangeFormer 的遥感图像建筑物变化检测(CD)方法。该方法使用多个不同时相的遥感图像, 将最新时相遥感图像与变化前的多个遥感图像在特征差异提取上进行不同尺度下的融合, 分别关注图像的综合语义特征以及图像之间语义信息的细节。该方法有助于减少季节、光照等因素发生变化时引起的误检。同时, 考虑变化前多个不同时相的遥感图像, 将其特征差异进行融合并引入正则化项作为损失函数, 进一步消除非建筑物变化以及建筑物非变化区域光照阴影带来的干扰, 提高模型的泛化能力。构建从农业土地耕地到建筑用地变化的三时相遥感图像数据集, 实验结果表明, 相较于目前最优的 BIT 方法, 多时相 ChangeFormer 方法在 F1 值、交并比(IoU)、精确率和召回率指标上分别提升了 9.04%、9.87%、15.27% 和 3.4%, 显著提高了检测精度, 且在细节信息处理方面明显优于经典的 CD 方法。

关键词: 多时相; 干扰因素; 变化检测; 特征差异; 不同尺度融合; ChangeFormer; 正则化

中图分类号: TP751

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070443

Remote Sensing Image Building Change Detection Method Based on Multi-Temporal ChangeFormer

JIANG Youze, LIU Xiangyang

(School of Mathematics, Hohai University, Nanjing 211106, Jiangsu, China)

【Abstract】 To address the issue of interference factors such as seasonal changes, climate, and illumination that affect high-resolution remote sensing images of the same geographical space but different temporal phases, a remote sensing image building Change Detection (CD) method based on a multi-temporal ChangeFormer is proposed. This method uses multiple remote sensing images from different temporal phases and fuses the latest temporal remote sensing image with multiple prechange remote sensing images at different scales for feature difference extraction. Additionally, it focuses on both the comprehensive semantic features of the images and the details of the semantic information between the images. This approach helps reduce false detections caused by changes in factors such as season and illumination. Additionally, the method fuses the feature differences of multiple prechange remote sensing images from different temporal phases and introduces a regularization term as a loss function. This eliminates interference from nonbuilding changes and illumination shadows in the nonchanging areas of buildings, thereby enhancing the generalization ability of the model. A three-temporal remote sensing image dataset covering changes from agricultural land to construction land is constructed. The experimental results show that, compared to the current optimal BIT method, the multi-temporal ChangeFormer method improves the F1 value, Intersection over Union (IoU), precision, and recall by 9.04%, 9.87%, 15.27%, and 3.4%, respectively, thus significantly enhancing detection accuracy. Furthermore, it outperforms classical CD methods in terms of detailed information processing.

【Key words】 multi-temporal; interfering factors; Change Detection (CD); characteristic difference; different scale fusion; ChangeFormer; regularization

0 引言

近年来,随着自然环境的不断演变和人类生产活动的不断出现,城市建设建筑用地规模越来越大,农业土地耕地的滥用问题成为社会关注的一大热

点^[1-2]。然而,土地的滥用浪费治理主要依赖于人工监察执法,这在浪费大量人力物力资源的同时,导致土地资源利用效率低下,此时,遥感图像变化检测(CD)在当前土地资源利用和遥感应应用研究中变得尤为重要^[3-5]。遥感图像 CD 旨在利用高分辨率的

基金项目: 云南省重大科技专项(202002AE090010)。

作者简介: 姜有泽,男,硕士研究生,主研方向为变化检测、图像语义分割;刘向阳(通信作者),副教授、博士。

收稿日期: 2024-10-08

修回日期: 2024-12-06

E-mail: liuxy@hhu.edu.cn

遥感图像和覆盖不同时期相同地表的地理空间数据,确定和分析地物的变化,包括距离和状态的变化,将区域中的每个像素分类为“变化”和“无变化”^[6-8]两类。

本文针对土地资源出现的滥用问题,使用高分辨率的遥感图像执行 CD 任务,利用有监督学习,通过训练样本与标签的一一对应关系,使模型提取遥感图像 CD 的特征与细节语义信息,并将模型的性能表现对未见遥感数据进行泛化,达到 CD 的目的,从而提高土地资源监察执法的管理效率。

1 相关工作

文献[9]提出了 3 种全卷积神经网络架构 FC-EF、FC-Siam-Conc 和 FC-Siam-Di,它们使用关于当前问题的启发式指标,在 CD 数据集上取得了较好结果。对于具有辨别力的对象级特征,一种双任务约束的深度孪生卷积网络(DTCDCN)^[10]模型被提出,以获得完整的 CD 图。文献[11]提出了一个双时相图像 Transformer(BIT),以高效地对时空域内的上下文进行建模。文献[12]提出一种基于 Transformer 的孪生网络架构(ChangeFormer),该架构对深度卷积神经网络(ConvNets)的依赖性较低,并且具有轻量级解码器的分层编解码结构,可以很好地用于 CD 任务。

此外,建筑物 CD 采用了与像素相关的技术^[13],提出了一种将图像与特征融合的 CD 模型,并得出了良好的测量结论。针对大规模的遥感影像 CD 任务,一种基于全卷积系统所提供的高深度编解码的 CD 技术^[14]被提出,以有效地解决此类检测问题。上述经典的遥感图像 CD 方法基于双时相的遥感图像变化,旨在通过不同的深度学习方法,对遥感图像的特征进行提取和细化,从而更好地生成 CD 结果^[15-17]。然而,这些方法在双时相遥感图像中只关注目标地物的真实变化^[18],受到无关因素(天气、季节、光照等)的干扰影响,具体表现在:对于遥感图像 CD 数据集,变化前后的遥感图像在通过卫星拍摄时不能保证遥感区域的季节、气候、光照等完全一致。不同的季节、气候导致遥感图像非建筑区域出现不同的颜色变化,光照则导致建筑物非变化区域出现不同角度的阴影,同时也导致不同的遥感图像之间出现较大的亮暗变化。这些无关因素影响了模型的 CD 性能,在面对多场景遥感图像时难以很好地执行 CD 任务。

目前,高分辨率遥感图像的拍摄并不局限于双时相,随着时间的推移,形成了多时相遥感图像的时

间序列,逐渐出现了对多个不同季节、不同时间的遥感图像之间变化的处理与分析。文献[19]提出一种多时相遥感图像 CD 方法,通过引入多个不同时相的遥感图像,消除非建筑物变化对结果带来的影响,进一步通过两两加权提高 CD 的精确度。然而,这种方法依赖于 ConvNets 提取语义特征,并且两两加权得到的语义信息有限。针对季节性变化、光照等因素对 CD 结果产生的干扰影响以及对 ConvNets 的依赖问题,本文在双时相 ChangeFormer^[12]的基础上,提出一种多时相 ChangeFormer 遥感图像建筑物 CD 方法。这种方法在双时相遥感图像之间再次引入过去变化前的多个不同时相的遥感图像,在提取遥感图像特征后进行不同尺度的特征融合,结合整体特征差异与两两加权特征差异对 CD 结果的影响,提高 CD 结果的稳定性。对过去变化前的多个时相遥感图像进行特征提取,将其特征差异作为损失函数的正则化项,促使模型偏向检测建筑物的变化,降低非建筑物变化和建筑物非变化光照阴影带来的影响,减小外部因素的干扰,进一步提高模型的性能,确保 CD 结果更加可靠。

2 多时相 ChangeFormer 方法原理

2.1 双时相 ChangeFormer 基本原理

随着遥感图像 CD 方法的不断发展与改进,ConvNets 在关键特征提取方面表现出色,因此,目前最先进的 CD 方法主要基于 ConvNets。由于在空间和时间范围内捕获上下文信息以识别双时相图像中的相关变化至关重要,因此最新的 CD 研究主要集中在提高 CD 模型的接受度上^[12]。近年来,Transformer 在自然语言处理(NLP)中的应用取得了巨大的成功^[20-21],因此,人们在 NLP 之外的其他领域(如计算机视觉)中也逐渐加入 Transformer,包括图像分类和语义分割^[22-23],如 ViT(Vision Transformer)^[24]、SETR(SEgmentation Transformer)^[25]、SegFormer^[26]等。

在双时相 ChangeFormer 中,对 ConvNets 的依赖性是不必要的,并且带有轻量级 MLP 解码器的分层 Transformer 编码器可以很好地用于 CD 任务。如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同),双时相 ChangeFormer 模型在 Transformer 编码器中设计了 4 个阶段,每个阶段通过 Transformer 模块提取特征,同时进行下采样操作。

自注意力^[20]被定义为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_{\text{head}}}}\right)\mathbf{V} \quad (1)$$

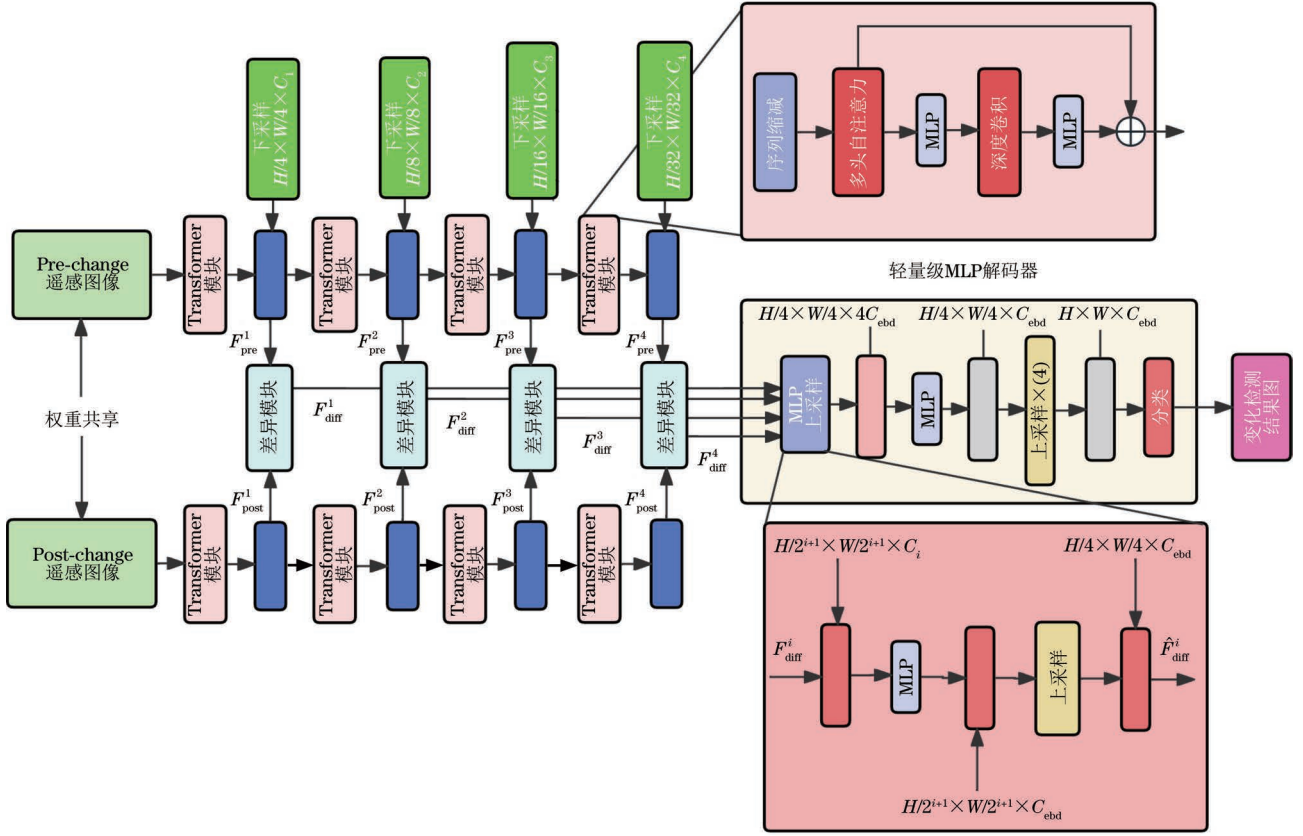


图 1 双时相 ChangeFormer 遥感图像 CD 网络结构

Fig. 1 CD network structure of the dual-temporal ChangeFormer remote sensing image

式中:Attention 表示自注意力; Q 、 K 、 V 分别表示 Query、Key、Value,并且具有相同的 $HW \times C$ 维度;Softmax 是激活函数; d_{head} 是键向量的维度。为了给 Transformer 提供位置信息,使用 MLP 层和 3×3 卷积,如下^[12]:

$$F_{out} = \text{MLP}(\text{GELU}(\text{Conv2D}_{3 \times 3}(\text{MLP}(F_{in})))) + F_{in} \quad (2)$$

式中: F_{out} 是输出位置信息; F_{in} 是自我注意力的特征;MLP 表示 MLP 层;GELU 表示高斯误差线性单位激活;Conv2D_{3×3} 表示卷积。对每一层的特征图进行长宽 4 倍的下采样,在维度层面翻倍。第一个特征层使用 7×7 、步幅为 4、Padding 为 3 的卷积层,其余使用 3×3 、步幅为 2、Padding 为 1 的卷积层。双时相 ChangeFormer 设计了计算每层编码器输出特征的特征差异模块,定义如下^[12]:

$$F_{diff}^i = \text{BN}(\text{ReLU}(\text{Conv2D}_{3 \times 3}(\text{Cat}(F_{pre}^i, F_{post}^i)))) \quad (3)$$

式中: F_{diff}^i 是第 i 阶段的特征差异; F_{pre}^i 、 F_{post}^i 分别表示变化前和变化后第 i 阶段的特征;BN 表示批归一化;ReLU 表示激活函数;Cat 表示张量连接。显而易见,双时相 ChangeFormer 把每层的 2 个特征图拼接

起来,经过一系列卷积进一步提取 2 个特征之间的差异,每一层特征差异传递到下一层继续计算。双时相 ChangeFormer 在 MLP 解码器中依据 FCN 模型^[27] 的思路,将解码器分为上采样和最后结果的输出,如图 1 所示。将从编码器中提取到的不同阶段的特征差异进行融合,将 4 个阶段的特征差异在维度上进行拼接,其经过一个 1×1 的卷积进行融合,然后利用 2D 转置卷积进行上采样操作,通过一系列残差块保持特征,最后经过一个 3×3 的分类卷积来生成最终的变化图,这个过程可以表示如下^[12]:

$$\hat{F}_{diff}^i = \text{Linear}(C_i, C_{ebd})(F_{diff}^i), \forall i \quad (4)$$

$$\hat{F}_{diff}^i = \text{UpSample}\left(\left(\frac{H}{4}, \frac{W}{4}\right), \text{bilinear}\right)(\hat{F}_{diff}^i) \quad (5)$$

$$F = \text{Linear}(4C_{ebd}, C_{ebd})(\text{Cat}(\hat{F}_{diff}^1, \hat{F}_{diff}^2, \hat{F}_{diff}^3, \hat{F}_{diff}^4)) \quad (6)$$

$$\hat{F} = \text{ConvTranspose2D}(S=4, K=3)(F) \quad (7)$$

$$C_{CM} = \text{Linear}(C_{ebd}, N_{cls})(\hat{F}) \quad (8)$$

式中: $\hat{F}_{diff}^i$ 表示连接后的第 i 阶段特征差异;Linear 表示线性层; C_{ebd} 表示嵌入维度;UpSample 表示上采样;bilinear 表示双线性插值;ConvTranspose2D 是 $S=4, K=3$ 的 2D 反置卷积层; C_{CM} 是预测掩码; N_{cls}

(值为 2)是类的数量,即“变化”和“不变化”两类。

2.2 多时相 ChangeFormer 基本框架

多时相 ChangeFormer 遥感图像建筑物 CD 模型在双时相 ChangeFormer 的基础上,引入多个过去变化前不同时期相的遥感图像,如图 2 所示。本文以三时相为例,将其记为 $T_ChangeFormer$ 。A 为

最新时相遥感图像,B、C 分别为过去变化前不同时期相的遥感图像。根据多时相遥感图像的时间序列特性,将不同尺度下的特征差异分为 2 个部分:一个是三时相遥感图像特征融合提取整体特征差异,记为 $T_1_ChangeFormer$;另一个是两两特征融合提取特征差异后加权,记为 $T_2_ChangeFormer$ 。

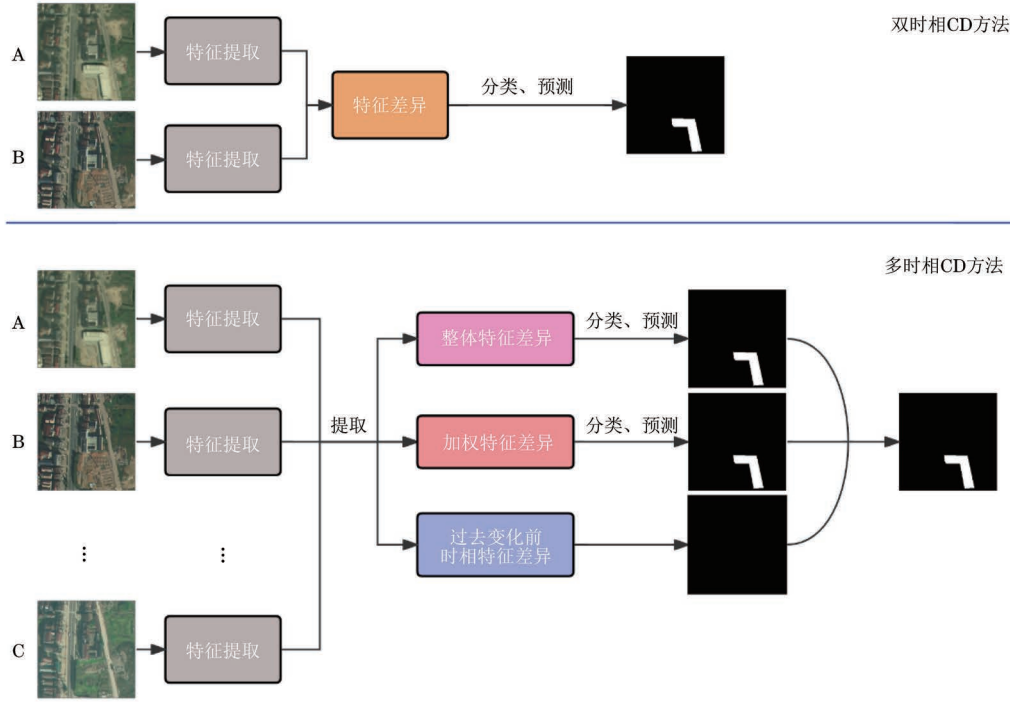


图 2 多时相遥感图像 CD 示意图

Fig.2 Schematic diagram of multi-temporal remote sensing image CD

2.2.1 整体差异提取($T_1_ChangeFormer$)

三时相遥感图像 A、B、C 根据双时相 ChangeFormer 的思想,不同时期的遥感图像利用式(1),通过 Transformer 模块进行特征提取和下采样操作,经过 4 次下采样,生成不同尺度的特征。对于 $T_1_ChangeFormer$,即将每个阶段 3 个不同时期的遥感图像生成的不同尺度特征沿着维度层面进行整体拼接融合,利用差异模块提取每阶段不同尺度的整体特征差异,计算如下:

$$F_{diff,T_1}^i = \text{BN}(\text{ReLU}(\text{Conv2D}_{3 \times 3}(\text{Cat}(F_A^i, F_B^i, F_C^i)))) \quad (9)$$

式中: F_A^i, F_B^i, F_C^i 分别表示第 i 层 A、B、C 图像所提取的特征; $F_{diff,T_1}^1, F_{diff,T_1}^2, F_{diff,T_1}^3, F_{diff,T_1}^4$ 是 $T_1_ChangeFormer$ 编码器的第一部分整体特征差异。

$T_1_ChangeFormer$ 关注 A、B、C 之间的整体特征差异,使模型学习到更加丰富的语义信息,在生成不同尺度的差异特征时,极大地降低了季节性变化、光照等带来的干扰,检测结果更加精确可靠。整体拼接

融合增加了通道数或其他维度,特征表达能力得到增强,有助于提升模型的表现,同时网络可以在后续层自动学习到如何使用这些组合的特征,使得模型具有更大的灵活性,能够自行决定特征之间的关系。

2.2.2 两两差异提取加权($T_2_ChangeFormer$)

$T_2_ChangeFormer$ 在 Transformer 编码器中,利用式(1),使用相同的 Transformer 模块进行特征提取和下采样操作,先对 A、B 生成的不同阶段、不同尺度的特征 F_A^i, F_B^i 进行维度层面的拼接融合,然后通过差异模块计算出 4 个不同阶段的特征差异 $F_{diff,AB}^1, F_{diff,AB}^2, F_{diff,AB}^3, F_{diff,AB}^4$,计算如下:

$$F_{diff,AB}^i = \text{BN}(\text{ReLU}(\text{Conv2D}_{3 \times 3}(\text{Cat}(F_A^i, F_B^i)))) \quad (10)$$

对 A、C 进行相同的操作,生成 4 个不同阶段、不同尺度的特征差异 $F_{diff,AC}^1, F_{diff,AC}^2, F_{diff,AC}^3, F_{diff,AC}^4$,计算如下:

$$F_{diff,AC}^i = \text{BN}(\text{ReLU}(\text{Conv2D}_{3 \times 3}(\text{Cat}(F_A^i, F_C^i)))) \quad (11)$$

每个阶段获得 2 个特征差异,遥感图像 A、B、C

特征权重共享,设置相同权重为 0.5,对 A、B 和 A、C 的特征差异进行平滑处理,生成 4 个加权后的特征差异,计算如下:

$$F_{\text{diff},T_2}^i = \frac{F_{\text{diff},AB}^i + F_{\text{diff},AC}^i}{2} \quad (12)$$

式中: F_{diff,T_2}^1 、 F_{diff,T_2}^2 、 F_{diff,T_2}^3 、 F_{diff,T_2}^4 是 T_ChangeFormer 编码器第二部分的整体特征差异。

T₂_ChangeFormer 在不同的特征差异组合之间进行平滑处理,能够更好地捕捉特征差异中的复杂关系,并且可以根据特征差异的相对关系进行优化,提高模型的性能。同时,两两加权组合可以利用相对信息,这有助于减小噪声数据对模型的影响,提高结果的鲁棒性。在特定情况下精细优化特征差异,可以增强模型对细微差异的区分能力,更好地提高模型的泛化能力。

2.2.3 三时相遥感图像建筑物变化检测模型(T_ChangeFormer)

T₁_ChangeFormer 和 T₂_ChangeFormer 在处

理特征差异关系时,前者保留和利用了大量的特征性质,后者在特定情况下对不同的特征进行精细优化。通过式(4)~式(7),利用 MLP 层分别将上采样的特征差异连接起来并融合,利用 2D 反置卷积层将融合特征上采样到 $H \times W$ 大小,分别为 $\hat{F}_{T_1}$ 、 $\hat{F}_{T_2}$ 。将 2 个上采样结果进行不同尺度下的融合,计算如下:

$$C_{\text{CM}} = \frac{\text{Linear}(C_{\text{ebd}}, N_{\text{cls}})(\hat{F}_{T_1}) + \text{Linear}(C_{\text{ebd}}, N_{\text{cls}})(\hat{F}_{T_2})}{2} \quad (13)$$

上述过程既保证了对 T₁_ChangeFormer 的优化,又将 T₂_ChangeFormer 的优化纳入考量,通过结合两部分特征差异,进一步提高了模型的性能和稳定性。由式(13)预测变化掩码生成最后的 CD 结果图,最终形成三时相遥感图像建筑物 CD 模型 T_ChangeFormer,如图 3 所示。根据 T_ChangeFormer 模型框架,本文多时相 ChangeFormer 遥感图像建

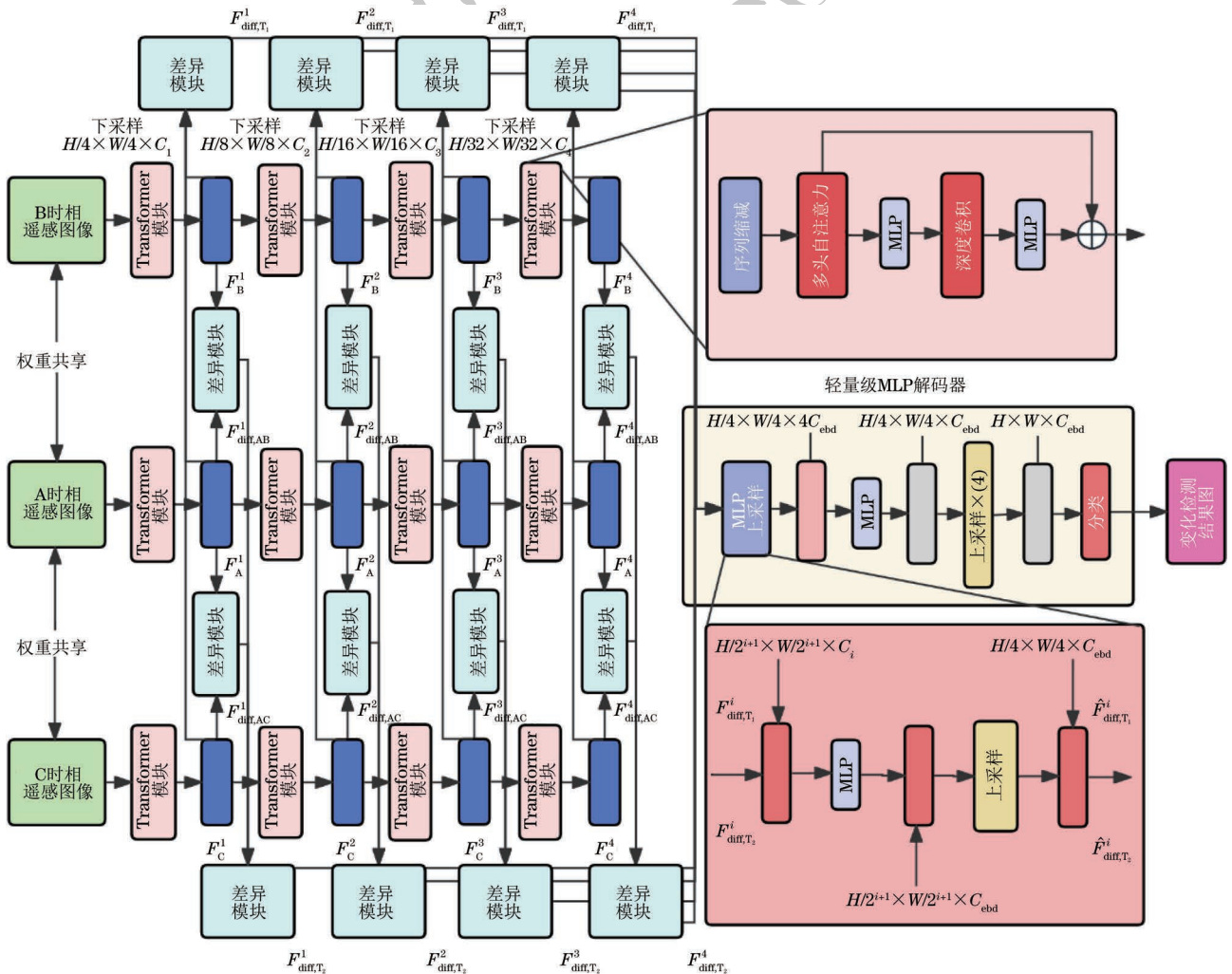


图 3 T_ChangeFormer 网络结构

Fig.3 Network structure of T_ChangeFormer

筑物 CD 模型利用多时期遥感图像提取更丰富的语义信息,同时获得更加稳定的 Transformer 编码器结构,充分考虑更多尺度下的特征差异以及非变化区域的有监督学习,进一步消除无关因素的影响,并通过轻量级 MLP 解码器,保证模型的稳定性与可靠性。

2.3 CD 的损失函数

2.3.1 双时相 ChangeFormer 损失函数

损失函数是机器学习和深度学习中的一个核心概念,用于量化模型的预测结果与真实结果之间的差异,它是模型训练过程中用于指导优化的依据。在双时相 ChangeFormer 中,采用交叉熵损失(CE Loss)来优化参数。CE Loss 是一种常用的损失函数,适用于分类任务,它衡量了模型输出的预测概率分布与真实类别之间的差异,其定义如下:

$$H(p, q) = - \sum_{i=1}^n p(x_i) \ln(q(x_i)) \quad (14)$$

式中: $p(x_i)$ 表示真实的概率分布,即真实的类别标签,如果是分类问题,正确类别的 $p(x_i)=1$,其余类别的 $p(x_i)=0$; $q(x_i)$ 表示模型预测的概率分布,它是模型输出的概率值(通过 Softmax 等激活函数获得); n 表示类别的总数; x_i 表示第 i 个类别。

2.3.2 三时相 T_ChangeFormer 损失函数

在许多检测任务中,正样本(目标类)和负样本(背景类)的数量常常严重不平衡。传统的 CE Loss 可能会受到大量负样本的影响,导致模型在训练时过于关注容易区分的负样本,而忽略了更难区分的正样本。如果继续使用 CE Loss,将不利于整个目标函数的收敛。在本文遥感图像 CD 任务中,根据实际遥感图像的质量、建筑物变化区域数量等因素,针对遥感数据构建时出现的类别不平衡问题,即将样本分为变化(change)和未变化(no-change)两类,其中未变化类占据大部分比例,变化类占据少数,在 T_ChangeFormer 模型中使用焦点损失(Focal Loss)^[28],其定义如下:

$$FL(p_t) = -\alpha_t (1-p_t)^\gamma \log_a(p_t) \quad (15)$$

式中: p_t 是模型对正确类别的预测概率,如果样本属于正类, $p_t=p$,否则 $p_t=1-p$; α_t 是平衡因子,用于平衡正负样本的重要性, α_t 通常用于在类别不平衡情况下调整正负样本的影响力; γ 是聚焦因子,控制损失函数对容易分类样本的惩罚程度。

为了进一步提高模型的性能和泛化能力,本文在焦点损失的基础上加入正则化损失。总损失定义如下:

$$L_{\text{总}} = L_T + L_{\text{BC}} \quad (16)$$

式中: L_T 代表 T_ChangeFormer 损失; L_{BC} 代表过去变化前遥感图像 B、C 之间特征差异的正则化损失。

针对实际土地资源利用中出现的滥用以及人工监察执法效率低下问题,本文重点关注农业土地耕地到建筑用地的 CD,并不考虑非建筑区域的变化情况。在理想状态下,过去变化前遥感图像 B 和 C 保持一致。根据实际遥感图像拍摄及构建情况,由于受到季节性变化、气候变化以及阳光照射阴影等干扰因素的影响,在实际 CD 时,在遥感图像 B、C 之间,非建筑物变化(如土地到土地的变化、河流之间的变化、地形之间的变化)以及光照对建筑物非变化区域产生的阴影,对模型的性能表现都会产生干扰,最终出现大量的误检现象。为了保证模型的稳定性与可靠性,本文在过去变化前遥感图像 B、C 之间提取不同尺度的特征差异,如下:

$$F_{\text{diff,BC}}^i = \text{BN}(\text{ReLU}(\text{Conv2D}_{3 \times 3}(\text{Cat}(F_B^i, F_C^i)))) \quad (17)$$

式中: F_B^i 、 F_C^i 分别为图像 B、C 第 i 阶段的特征。

经式(4)~式(7)的特征融合、MLP 解码器和 2D 反置卷积层,使融合的特征差异上采样到 $H \times W$ 大小,得到 $\hat{F}_{\text{BC}}$ 。根据有监督学习的特性,在 $\hat{F}_{\text{BC}}$ 与像素值为 0 的标签之间进行正则化损失参数优化,旨在向损失函数中添加一个惩罚项来约束模型,促使模型忽略非建筑物变化带来的影响,在一定程度上消除非建筑物变化带来的大量误检现象,其计算如下:

$$L_{\text{BC}} = \lambda |FL(l(\hat{F}_{\text{BC}}, Y_{\text{HW}}))| \quad (18)$$

式中: λ 是正则化强度的超参数,其控制正则化项的影响程度; Y_{HW} 是图像位于 (H, W) 上像素为 0 的标签; $\hat{F}_{\text{BC}}$ 是 B、C 之间在 (H, W) 上的特征差异; l 是 $\hat{F}_{\text{BC}}$ 与 Y_{HW} 之间的预测概率。

3 实验结果及分析

3.1 数据集构建

本文所构建的数据集来源于通过卫星拍摄的浙江省某一区域 2018 年—2022 年 5 期的高分辨率遥感图像,对于多时相数据的中间变化,如果出现了遥感数据的二次甚至更多的变化,将降低标签的鲁棒性,使模型学习更多错误的信息,从而出现更多的误检现象。为了解决上述问题,同时考虑此问题的实际意义,需要尽可能保证中间图像的未变化性,只保留最新时相发生变化的遥感数据集。对多时相遥感

数据集进行不重叠分割,使用 256×256 尺寸同时进行对齐。针对土地资源利用中出现的滥用导致人工监察执法效率低下的问题,进一步结合实际遥感图像建筑物之间的变化关系,挑选出从农业土地耕地到建筑用地变化的三时相遥感数据,并对其进行相应的数据增强,包括翻转、旋转、亮度、高斯模糊等,得到 1 095 张分辨率大小为 256×256 像素的 RGB 图像并进行对齐,其中以 8:1:1 的比例形成训练集、验证集、测试集。

3.2 实验环境及参数

本文实验在深度学习框架 PyTorch 1.9.0 版本下实现,使用 Python 版本为 3.8.19,并在 NVIDIA A10 GPU 下进行训练加速。根据深度学习的训练经验,本文使用 AdamW 优化器,权重衰减为 0.01,学习率初始值为 0.000 1,Epoch 为 100, Batch size 为 4。针对模型的特点,本文将模型层数设置为 [3, 3, 4, 3],注意力机制头数设为 [1, 2, 4, 8],将 MLP 中每一层的宽度与输入维度的比例设为 [4, 4, 4, 4],drop_path 正则化率设置为 0.1。其他实验环境均相同。

3.3 性能评价指标

为了评价各模型的性能,计算出模型在 CD 中的精确率(Precision, P)、召回率(Recall, R)、交并比(IoU, I_{IoU})、F1 值($F1$, F_1)等评价指标,并将关于变化类的 F1 和 IoU 作为主要的定量指标。各指标具体计算如下:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (19)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (20)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (21)$$

$$I_{IoU} = \frac{N_{TP}}{N_{TP} + N_{FN} + N_{FP}} \quad (22)$$

式中: N_{TP} 表示把正类预测成正类; N_{TN} 表示把负类预测成负类; N_{FP} 表示把负类预测为正类,简称误检; N_{FN} 表示把正类预测为负类,简称漏检。

3.4 正则化系数 λ 选取

正则化的目的是在模型中添加一个惩罚项来约束模型,使模型的泛化能力得到进一步提升。在引入多时相遥感图像后,非建筑物区域的变化对 CD 结果产生了干扰,此时考虑过去变化前遥感图像 B、C 之间的特征差异变得尤为重要,通过训练可以让模型趋向于非建筑物区域和建筑物非变化区域的无变化发展,本文将其作为正则化项来

优化参数,探究不同正则化强度下的模型性能表现。根据正则化项的数量级和实际调试过程,本文选取正则化系数 λ 在 0~1.5 范围,在相同的环境下进行实验,得到不同正则化系数下的实验结果,如图 4 所示。

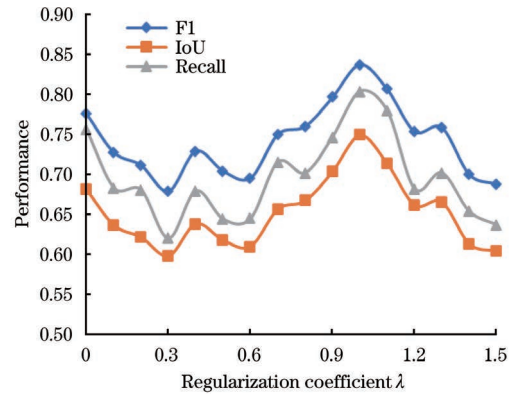


图 4 不同正则化系数下的实验结果

Fig.4 Experimental results under different regularization coefficients

由图 4 可知,当 $\lambda=0$ 时,F1 值为 0.776 28,IoU 为 0.681 37,模型未考虑正则化约束,B、C 之间非建筑物变化对结果带来一定的干扰,使模型在一定程度上关注了非建筑物之间的变化,对农业土地耕地到建筑物变化产生了较大干扰,降低了模型的总体性能。随着正则化影响程度的不断扩大,当 $\lambda=1.0$ 时,各指标达到峰值,F1 值为 0.837 92,IoU 为 0.749 70。继续增加正则化系数发现,模型的损失受到正则化项的主导影响,模型的性能呈下降趋势。通过正则化系数的调整,模型在 $\lambda=1.0$ 时的性能相较未使用正则化时的性能提升 6%,在一定程度上消除了非建筑物变化对模型的影响,提高了模型结果的可靠性。

3.5 对比实验

本文将 T_ChangeFormer 与现有部分经典遥感图像 CD 方法进行比较,结果如表 1 所示(最优结果加粗标注),可视化定性结果如图 5 所示。同时,将现有遥感图像 CD 方法改进为三时相输入 T_FC-EF、T_FC-Siam-Di、T_FC-Siam-Conc、T_DTCDSCN 和 T_BIT 中。

1) T_FC-EF。FC-EF^[9] 连接双时相遥感图像并通过 ConvNets 处理以检测变化,在本文改进为三时相输入,利用整体拼接融合三时相遥感图像来检测变化。

2) T_FC-Siam-Di。FC-Siam-Di^[9] 是一种特征差分方法,在本文改进为三时相输入,从 ConvNets 中提取三时相遥感图像的多级特征,并利用它们两

两差分加权来检测变化。

3) T_FC-Siam-Conc。FC-Siam-Conc^[9] 是一种特征连接方法,在本文改进为三时相输入,从 ConvNets 中提取三时相遥感图像的多级特征,并使用特征连接将特征两两加权以检测变化。

4) T_DTCDSN。DTCDSN^[10] 是一种基于注意力的方法,在本文改进为三时相输入,利用双注意力模块(DAM)和 ConvNets 特征的通道和空间位置之间的相互依赖关系,通过对三时相遥感图像特征两两加权来检测变化。

5) T_BIT。BIT^[11] 是一种基于 Transformer 的方法,在本文改进为三时相输入,使用 Transformer 编码器-解码器网络,通过语义标记增强 ConvNets 特征的上下文信息,然后对三时相遥感图像特征两两差分加权以获得变化图。

表 1 不同遥感图像 CD 方法的平均定量结果

Table 1 Average quantitative results of different remote sensing image CD methods

Method	F1	IoU	Precision	Recall
FC-EF ^[9]	0.709 67	0.611 17	0.666 42	0.802 03
FC-Siam-Di ^[9]	0.684 80	0.591 75	0.654 98	0.735 30
FC-Siam-Conc ^[9]	0.690 85	0.593 04	0.648 77	0.793 44
DTCDSN ^[10]	0.709 19	0.616 81	0.706 52	0.711 95
BIT ^[11]	0.706 24	0.613 50	0.695 20	0.718 91
ChangeFormer ^[12]	0.729 72	0.635 26	0.726 80	0.732 73
T_FC-EF	0.722 58	0.622 81	0.675 58	0.825 11
T_FC-Siam-Di	0.707 25	0.612 16	0.678 30	0.751 40
T_FC-Siam-Conc	0.705 69	0.608 45	0.666 17	0.783 18
T_DTCDSN	0.710 75	0.617 83	0.703 56	0.718 58
T_BIT	0.747 49	0.651 00	0.728 98	0.769 97
T_ChangeFormer	0.837 92	0.749 70	0.881 64	0.804 01

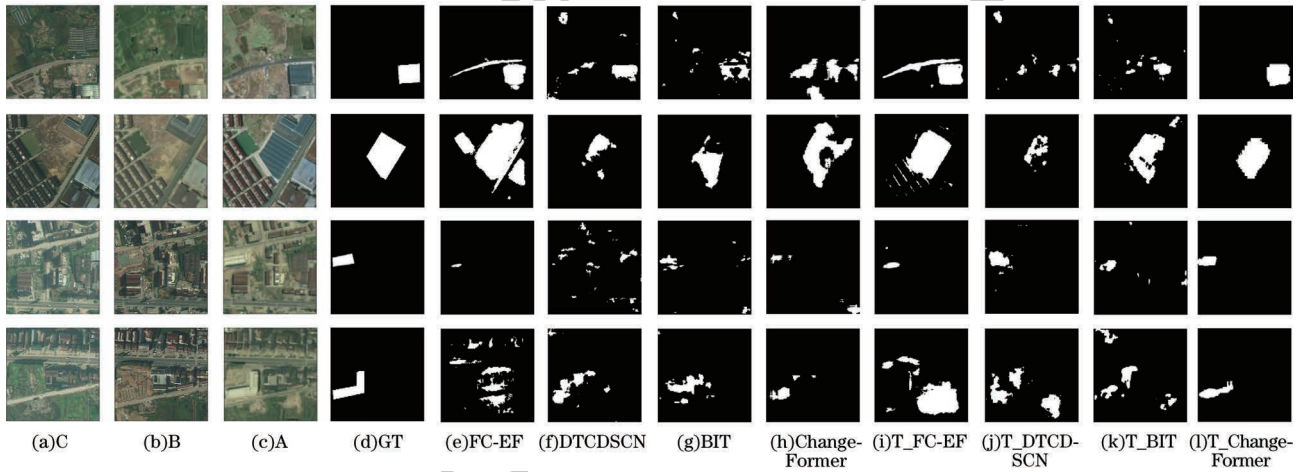


图 5 不同遥感图像 CD 方法的定性结果

Fig.5 Qualitative results of different remote sensing image CD methods

表 1 显示了不同遥感图像 CD 方法及三时相输入方法在测试集上的平均定量结果。从表中可以看出, T_ChangeFormer 模型在 F1 值、IoU、Precision 上实现了更好的 CD 性能,在 Recall 指标上仅次于 T_FC-EF。本文的 T_ChangeFormer 相比之前最优的 CD 方法,在 F1 值、IoU、Precision、Recall 指标上分别提升了 9.04%、9.87%、15.27%、3.4%。

图 5 展示了 T_ChangeFormer 方法与现有遥感图像 CD 方法(包括改进为三时相的方法)在具体数据测试集上的定性结果。从图 5 中可以看出,不同时期遥感图像之间整体亮度有明显变化,根据不同时期的光照影响,非建筑物区域出现了颜色、饱和度等变化,同时建筑物出现了不同角度的光照阴影,致使建筑物非变化区域出现了伪变化,即非变化的建筑物在不同角度的光照下出现了阴影变化,实际建筑物未变化。因此,对于现有的遥感图像 CD 方法,

其在建筑物变化区域细节信息丢失较多,出现了一些漏检,并且在非建筑物区域以及建筑物非变化区域受无关因素影响而出现大量误检现象。通过改进为三时相输入后,在一定程度上提升了建筑物变化区域的检测精度,但依旧受到无关因素的干扰,存在大量误检现象。本文 T_ChangeFormer 在捕获建筑物变化区域的细节上表现更加出色,利用特征差异在不同尺度下的融合,提供了大量的语义信息,极大地减少了漏检现象,同时结合正则化进一步降低了遥感图像中因季节性变化、光照以及非建筑物变化等干扰因素带来的误检现象。通过定性结果可以直观看出正则化的显著影响,这些定量和定性结果表明,本文 T_ChangeFormer 方法性能表现优于现有经典的 CD 方法。

3.6 消融实验

为了提高 T_ChangeFormer 在 CD 性能上的稳定

性,获得性能最优的 T_1 _ChangeFormer 模型,本文在 T_1 _ChangeFormer 中通过控制三时相遥感图像特征差

异之间的提取方式以及 B、C 正则化损失 L_{BC} ,进一步得到一个更加稳定的模型,具体结果如表 2 所示。

表 2 消融实验结果

Table 2 Ablation experiment results

T_1 _ChangeFormer	T_2 _ChangeFormer	L_{BC}	F1	IoU	Precision	Recall
✓	✓	✓	0.837 92	0.749 70	0.881 64	0.804 01
✓	✓	—	0.799 66	0.703 52	0.756 97	0.864 04
✓	—	✓	0.791 69	0.698 44	0.882 26	0.738 15
—	✓	✓	0.791 73	0.697 03	0.800 02	0.783 96
✓	—	—	0.772 07	0.675 13	0.747 18	0.803 79
—	✓	—	0.765 88	0.665 36	0.709 29	0.886 62

由表 2 可知,未考虑正则化时, T_1 _ChangeFormer 的 F1 值为 0.772 07, IoU 为 0.675 13, T_2 _ChangeFormer 的 F1 值为 0.765 88, IoU 为 0.665 36, 并且 Recall 为 0.886 62, 达到最优。将两者特征差异进行不同尺度的融合后, F1 值为 0.776 28, IoU 为 0.681 37, 提升 1%。随着正则化损失的加入, 在最优正则化系数 $\lambda=1.0$ 条件下, T_1 _ChangeFormer 的 F1 值和 IoU 分别为 0.791 69 和 0.698 44, T_2 _ChangeFormer 的 F1 值和 IoU 分别为 0.791 73 和 0.697 03, 相较于未使用正则化, 性能提升了 2%~2.5%。可以看出, 正则化对于不同特征差异提取的模型性能提升较大。三者结合后, T_1 _ChangeFormer 模型的 F1 值和 IoU 分别为 0.837 92 和 0.749 70, 性能提升 4.6%~7.2%。综合考虑不同尺度下的特征差异, 同时结合最优系数下的正则化损失, 可以使 T_1 _ChangeFormer 模型取得更高的稳定性和泛化性能。

4 结束语

本文针对实际应用中土地资源滥用以及监察执法效率低下的问题, 提出一种多时相 ChangeFormer 遥感图像建筑物 CD 模型。该模型对特征差异进行不同尺度下的融合, 并引入未变化遥感图像特征差异的正则化项, 在一定程度上消除了季节、光照等因素的干扰。本文在农业土地耕地到建筑用地的三时相变化数据集上进行实验, 结果表明, 相比基于 ConvNets、注意力、ConvNets+Transformer 的方法, 所提模型取得了更好的 F1 值、IoU 以及整体准确率, 获得了显著的精度改进。多时相 ChangeFormer 遥感图像建筑物 CD 方法显著降低了非建筑等无关因素对结果的影响, 进一步提升了建筑物 CD 的稳定性。在未来的工作中, 可将模型在其他区域的建筑物中进行进一步的推广与应用, 同时构建大规模的多时相建筑物 CD 数据集,

充分考虑不同时相遥感图像之间的细微关系, 从而提升模型的泛化能力。

参考文献

- [1] 李婷. 农村土地资源利用和保护存在问题及对策[J]. 河北农机, 2024(13): 67-69.
LI T. Problems and countermeasures in the utilization and protection of rural land resources [J]. Hebei Agricultural Machinery, 2024(13): 67-69. (in Chinese)
- [2] 俞科立. 乡村振兴背景下基层土地资源管理问题及优化策略——以浙江某村为例[J]. 农村科学实验, 2024(13): 12-14.
YU K L. Problems and optimization strategies of grass-roots land resources management under the background of rural revitalization—a case study of a village in Zhejiang Province [J]. Rural Scientific Experiment, 2024(13): 12-14. (in Chinese)
- [3] 黄莹莹, 郭传旭, 张婷, 等. 农村土地资源闲置问题研究[J]. 合作经济与科技, 2024(16): 4-7.
HUANG Y Y, GUO C X, ZHANG T, et al. Study on idle rural land resources [J]. Co-Operative Economy & Science, 2024(16): 4-7. (in Chinese)
- [4] 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)
- [5] SONG L, XIA M, XU Y, et al. Multi-granularity Siamese transformer-based change detection in remote sensing imagery [J]. Engineering Applications of Artificial Intelligence, 2024, 136: 108960.
- [6] 孙剑明, 赵梦鑫, 郝旭耀. 遥感图像变化检测方法研究综述[J]. 计算机工程与应用, 2024, 60(20): 30-48.
SUN J M, ZHAO M X, HAO X Y. Research review of remote sensing image change detection methods [J]. Computer Engineering and Applications, 2024, 60(20): 30-48. (in Chinese)
- [7] YU S B, TAO C, ZHANG G, et al. Remote sensing image change detection based on deep learning: multi-level feature cross-fusion with 3D-convolutional neural networks [J]. Applied Sciences, 2024, 14(14): 6269.
- [8] YANG B, QIN L, LIU J Q, et al. IRCNN: an irregular-time-distanced recurrent convolutional neural network for change detection in satellite time series [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 2503905.
- [9] CAYE D R, LE S B, BOULCH A. Fully convolutional Siamese networks for change detection [C]//Proceedings of the 25th IEEE International Conference on Image Processing

- (ICIP). Washington D. C., USA: IEEE Press, 2018: 4063-4067.
- [10] LIU Y, PANG C, ZHAN Z Q, et al. Building change detection for remote sensing images using a dual-task constrained deep Siamese convolutional network model[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(5): 811-815.
- [11] 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.
- [12] BANDARA W G C, PATEL V M. A Transformer-based Siamese network for change detection[C]//Proceedings of the 2022 IEEE International Geoscience and Remote Sensing Symposium. Washington D. C., USA: IEEE Press, 2022: 207-210.
- [13] 张永梅, 李立鹏, 姜明, 等. 综合像素级和特征级的建筑物变化检测方法[J]. 计算机科学, 2013, 40(1): 286-293.
ZHANG Y M, LI L P, JIANG M, et al. Change detection method for buildings based on pixel-level and feature-level[J]. Computer Science, 2013, 40(1): 286-293. (in Chinese)
- [14] 余江南. 基于深度编-解码结构的高分辨遥感影像变化检测研究[D]. 西安: 西安电子科技大学, 2021.
YU J N. A research on deep en-decoding structure based high-resolution remote sensing image change detection[D]. Xi'an: Xidian University, 2021. (in Chinese)
- [15] 冯星宇, 朱灵龙, 张永宏, 等. 基于多边特征引导聚合网络的变化检测算法[J]. 计算机工程与应用, 2025, 61(3): 264-274.
FENG X Y, ZHU L L, ZHANG Y H, et al. Change detection algorithm based on multilateral feature guided aggregation network [J]. Computer Engineering and Applications, 2025, 61(3): 264-274. (in Chinese)
- [16] 吴潮宇, 杨斌. 基于大核重参 U-Net 的遥感影像变化检测[J]. 计算机工程, 2025, 51(3): 261-273.
WU C Y, YANG B. Remote sensing image change detection based on large kernel re-parameter U-Net [J]. Computer Engineering, 2025, 51(3): 261-273. (in Chinese)
- [17] 徐雄, 杨欣宇, 朱学康, 等. 任意尺度超分辨率图像篡改掩膜生成方法[J]. 计算机工程, 2025, 51(12): 277-284.
XU X, YANG X Y, ZHU X K, et al. Manipulation mask manufacturer for arbitrary-scale super-resolution images[J]. Computer Engineering, 2025, 51(12): 277-284. (in Chinese)
- [18] 黄霞. 基于深度学习的高分辨率遥感图像变化检测方法[D]. 成都: 电子科技大学, 2022.
HUANG X. High-resolution remote sensing image change detection methods based on deep learning [D]. Chengdu: University of Electronic Science and Technology of China, 2022. (in Chinese)
- [19] 牟彦霖, 刘向阳. 多时态 BIT 遥感图像建筑物的变化检测[J]. 计算机系统应用, 2024, 33(6): 185-191.
MOU Y L, LIU X Y. Change detection for buildings in multitemporal BIT remote sensing images [J]. Computer Systems & Applications, 2024, 33 (6): 185-191. (in Chinese)
- [20] VASWANI A. Attention is all you need[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1706.03762>.
- [21] 鞠天杰, 刘功申, 张伟胜, 等. 自然语言处理中的探针可解释方法综述[J]. 计算机学报, 2024, 47(4): 733-758.
JU T J, LIU G S, ZHANG Z S, et al. A review of probe interpretable methods in natural language processing [J]. Chinese Journal of Computers, 2024, 47(4): 733-758. (in Chinese)
- [22] 智敏, 陆静芳. 面向图像分类的 Vision Transformer 研究综述[J]. 郑州大学学报(工学版), 2024, 45(4): 19-29.
ZHI M, LU J F. A review of Vision Transformer for image classification [J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(4): 19-29. (in Chinese)
- [23] 严毅, 邓超, 李琳, 等. 深度学习背景下的图像语义分割方法综述[J]. 中国图象图形学报, 2023, 28(11): 3342-3362.
YAN Y, DENG C, LI L, et al. Survey of image semantic segmentation methods in the deep learning era[J]. Journal of Image and Graphics, 2023, 28(11): 3342-3362. (in Chinese)
- [24] DOSOVITSKIY A. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2010.11929>.
- [25] ZHENG S X, LU J C, ZHAO H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 6877-6886.
- [26] XIE E Z, WANG W H, YU Z D, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2105.15203>.
- [27] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [C] // Proceedings of the IEEE Conference on Pattern Analysis and Machine Intelligence. Washington D. C., USA: IEEE Press, 2016: 640-651.
- [28] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for dense object detection [C] // Proceedings of the IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2980-2988.

文字编辑 吴云芳
栏目编辑 宋 圆