

基于BIT网络的遥感图像变化检测算法

彭海峰, 傅迎华

(上海理工大学光电信息与计算机工程学院, 上海 200093)

摘要: 变化检测是通过分析同一地点两个不同阶段的遥感图像来获取地表变化的过程。卷积神经网络难以解决复杂场景下的变化检测任务。BIT网络在卷积后串联Transformer, 通过捕获全局信息改善检测效果。然而, 直接串联Transformer会削弱局部信息特征表示能力, 从而导致漏检现象。为此, 在BIT网络的基础上使用残差网络将Transformer前后对应的局部与全局信息融合起来, 以弥补直接串联Transformer对局部信息造成的损失, 达到提升精度的目的。同时针对遥感图像中变化与不变化类不平衡问题, 加入新的损失函数Dice来优化训练。实验结果表明, 在公开数据集LEVIR-CD与BCDD上, 改进后网络在参数量增加0.02 M的情况下, F1指标分别提升了0.51%与0.69%; 召回率分别提升了0.45%与0.50%。

关键词: 卷积神经网络; 变化检测; Transformer; 残差网络; 遥感图像

DOI: 10.11907/rjdk.241001

开放科学(资源服务)标识码(OSID):



中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)001-0150-08

Remote Sensing Image Change Detection Based on BIT Network

PENG Haifeng, FU Yinghua

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Change detection is the process of obtaining surface changes by analyzing remote sensing images from two different stages at the same location. Convolutional neural networks are difficult to solve change detection tasks in complex scenes. The BIT network concatenates the Transformer after convolution to improve detection performance by capturing global information. However, directly concatenating Transformers will weaken the ability to represent local information features, leading to missed detections. Therefore, based on the BIT network, a residual network is used to fuse the local and global information before and after the Transformer, in order to compensate for the loss of local information caused by directly concatenating Transformers and achieve the goal of improving accuracy. At the same time, a new loss function Dice is added to optimize training for the imbalance between changing and unchanged classes in remote sensing images. The experimental results showed that on the publicly available datasets LEVIR-CD and BCDD, the improved network improved the F1 index by 0.51% and 0.69%, respectively, with a parameter increase of 0.02M; The recall rates have increased by 0.45% and 0.50% respectively.

Key Words: convolutional neural network; change detection; Transformer; residual network; remote sensing images

0 引言

变化检测是根据对同一物体或现象在不同时间的观测结果来确定其不同之处的处理过程。具体来说是利用多时相获取的覆盖同一地表区域的遥感影像及其他辅助数据, 通过计算机图像处理系统识别、分析不同时段目标

或现象状态的变化情况, 提供地物的空间分布及其变化的定性与定量信息, 目的是发现感兴趣的变化信息, 并剔除作为干扰因素的不相干变化信息^[1]。我国已将地理国情普查监测作为一项重要战略需求, 目标是逐步建立与完善国家级地理国情动态监测信息系统, 为各行各业提供科学可靠的空间信息服务^[2]。目前, 自动与半自动化的遥感影像变化检测技术已经广泛应用于土地调查、城市研究、生

收稿日期: 2023-12-29

扫描二维码阅读全文:



作者简介: 彭海峰(1996-), 男, 上海理工大学光电信息与计算机工程学院硕士研究生, 研究方向为图像处理与计算机视觉; 傅迎华(1976-), 女, 博士, 上海理工大学光电信息与计算机工程学院副教授, 研究方向为数据挖掘与计算机视觉。本文通讯作者: 傅迎华。

态系统监测、灾害监测等领域,对社会的可持续发展起到重要作用^[3]。

1 相关工作

传统的遥感图像变化检测方法大致可分为图像代数、图像变换与分类后比较 3 类。这些方法在小规模数据集上可以实现较高的识别精度,但是需依赖手工设计的低级特征,并且无法减少同物异谱与异物同谱的影响,难以实现真正的智能化与自动化。

近年来,深度学习技术由于具备强大的特征表示能力,在计算机视觉的许多领域取得了比传统方法更好的效果,如语义分割、目标检测、图像分类等^[4-6]。根据网络架构的不同,深度学习方法主要分为循环神经网络(Recurrent Neural Network, RNN)、生成对抗网络(Generative Adversarial Networks, GAN)与卷积神经网络(Generative Adversarial Networks, GAN)几类。RNN 是处理变化检测任务中多个时间序列数据之间时间依赖性的有效手段。例如, Lyu 等^[7]采用长短期记忆网络(Long Short-Term Memory, LSTM)作为基础模型,将前后时期的影像光谱值分别作为时间序列的特征,提取光谱值在时序中的特征,通过全连接层输出检测结果。相较于常规神经网络,该方法通过对图像的时间序列信息进行建模获得了更为出色的结果,然而其并未充分考虑图像在空间中的特征信息。GAN 是一种生成模型,由生成器与判别器两个部分组成。例如, Gong 等^[8]提出一种基于 GAN 的无监督变化检测模型,采用传统变化检测方法生成初始差异图像,将其与两个不同时期图像的联合分布一起输入到 GAN 中。通过持续对抗学习,GAN 的生成器建立了训练数据与其相应图像块之间的共享映射函数,从而改进 DI(Difference Image)。应用无监督聚类算法对改进 DI 进行分析,最终得到二值变化图。

基于 CNN 的遥感图像变化检测方法可以分为前期融合结构与后期融合结构两类。前期融合结构是将双时相原始图像沿通道拼接起来,以获得单个图像作为网络的输入;后期融合结构采用并行的双流网络从双时相图像中提取特征,再融合各级特征的差异信息,若双流网络的结构与权重相同,则称该网络为孪生网络。例如, Daudt 等^[9]提出 3 种不同的端到端变化检测架构,即 FC-early fusion、FC-Siam-conc 与 FC-Siam-diff,其区别在于双时相图像融合策略与跳跃连接方式有所不同。受 UNet++^[10]的启发, Fang 等^[11]设计了一个密集连接的孪生网络 SNUNet-CD,通过聚合不同层次语义特征来减轻深层特征定位信息的丢失; Zhang 等^[12]提出 IFNet 网络,通过融合不同层次的差异信息来提升变化检测性能。然而,上述几种网络均通过堆积更多卷积层或使用更大卷积核的方式获得对象特征的全局信息,参数量巨大,对硬件性能有较高要求。

传统 CNN 使用固定大小的滑动窗口处理图像,未充分

考虑大范围的全局信息。最近,Transformer 模型在图像处理领域得到了广泛应用^[13-16]。Transformer 中用于提取全局信息的核心部分为自注意力机制,其允许模型在处理输入序列时动态地分配不同位置的注意力权重,能够捕捉输入序列的全局相关性。受光照角度与季节变化的影响,同类型的物体在遥感图像不同空间位置往往表现出不同的亮度。为此, Chen 等^[17]提出 BIT(Bitemporal Image Transformer)网络解决这一类问题,通过引入 Transformer 进行全局信息建模,将高级特征映射为语义标记,送入 Transformer 中来捕获远距离时空中的相关信息,排除不相关的伪变化。尽管 BIT 网络通过引入 Transformer 模块弥补了 CNN 在全局建模上的不足,但是直接串联 Transformer 模块可能会导致 BIT 在信息流动过程中丢失一些卷积提取的重要局部信息。为解决这一问题,本文提出名为 Res-CDNet 的网络架构,以 BIT 网络为基础,在卷积输出与 Transformer 输出之间引入 ResNet 残差块,代替原有的 Transformer 串联方式,旨在保留卷积网络提取的局部信息,同时允许 Transformer 更好地利用全局信息。这种方式平衡了全局与局部信息的提取,提高了网络整体性能。此外,由于变化区域与未变区域比例相差较大,存在类不平衡的问题,本文在二元交叉熵(Binary Cross-Entropy, BCE)损失的基础上加入 Dice 损失函数来优化训练^[18]。在两个公开数据集上对所提网络进行全面验证,并与一些主流方法进行了比较,证实了所提网络的有效性。

2 Res-CDNet 网络

Res-CDNet 是一种端到端的孪生网络,结构如图 1(彩图扫 OSID 码可见,下同)所示。其主要由 5 个模块组成,分别为卷积特征提取模块、语义标记模块、Transformer 模块、残差模块与预测模块。首先,通过卷积操作获取原始双时相图像的特征;其次,通过 Transformer 模块增强特征的全局信息;再次,通过残差模块将卷积特征提取模块的差异信息与 Transformer 提取的差异信息连接起来;最后,通过预测模块输出预测图。

2.1 卷积特征提取模块

与大多数卷积网络相比, ResNet-18 可以在较少的参数下达到很好的性能,适合硬件资源受限的环境。因此,本文将 ResNet-18 作为网络特征提取模块,并去除了池化模块与全连接层,仅采用 ResNet 的前 5 个模块,其中第 4、5 个模块的卷积步长设置为 1。5 个模块的输出大小分别为 $F_1 \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times c_1}$ 、 $F_2 \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times c_2}$ 、 $F_3 \in R^{\frac{H_0}{8} \times \frac{W_0}{8} \times c_3}$ 、 $F_4 \in R^{\frac{H_0}{8} \times \frac{W_0}{8} \times c_4}$ 与 $F_5 \in R^{\frac{H_0}{8} \times \frac{W_0}{8} \times c_5}$,其中 H_0 与 W_0 分别为输入双时相图像的高度与宽度; c_1, \dots, c_5 为通道数,分别为 64、64、128、256 与 256。

经过 ResNet 模块后,采用双线性插值法 *Bilinear* 进行 2 倍的上采样,并利用 3×3 卷积将通道数降为 32。具体公式

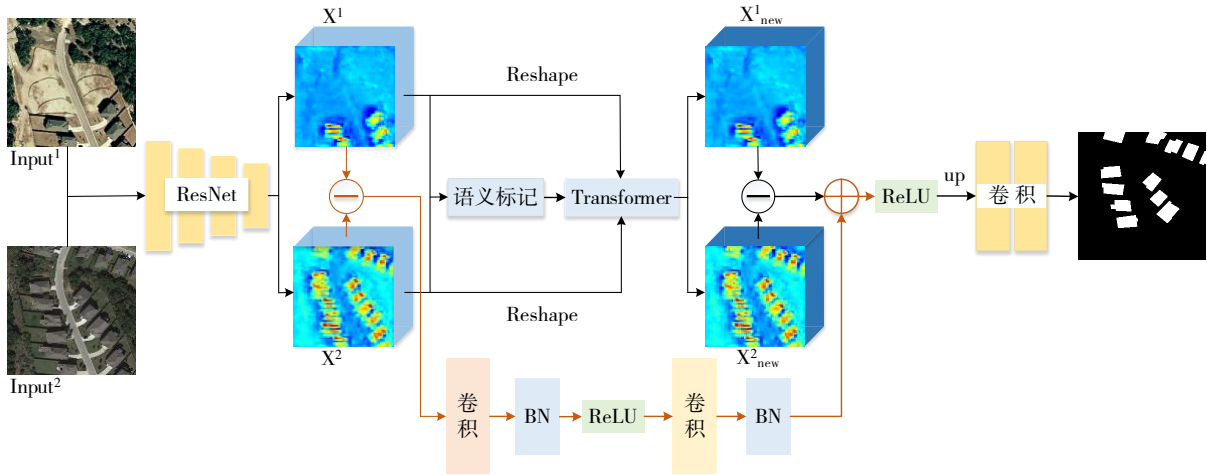


Fig. 1 Res-CDNet overall structure

图1 Res-CDNet 总体结构

如下：

$$X = Conv(Bilinear(F_s)) \quad (1)$$

因此，卷积特征提取模块的输出大小为 $X \in R^{64 \times 64 \times 32}$ 。

2.2 语义标记模块

在图像领域，Transformer的常用输入处理方式为将输入图像不重叠切分成固定大小的块(patch)，然后将这些块的像素点灰度值依次展开作为矩阵行向量(或列向量)，那么行数(或列数)即为patch的数量，构成了一个二维矩阵作为Transformer编码器的输入。这种方式的每一个patch信息代表着一个Token，整个二维矩阵代表Token集合。BIT网络没有采用直接将图像切分成patch再展开的方式作为Transformer输入，而是利用卷积构建一个二维矩阵作为输入。语义标记模块结构如图2所示。

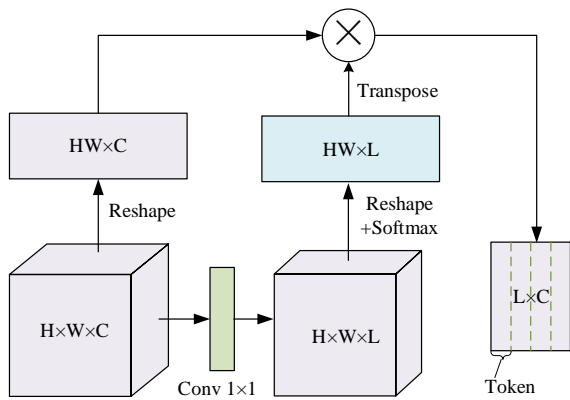


Fig. 2 Semantic markup module structure

图2 语义标记模块结构

假设上个卷积特征提取模块的输出为 $X \in R^{H \times W \times C}$ ，首先，通过逐点卷积将输入 f 的通道数变为 L ，得到 $X' \in R^{H \times W \times L}$ ；其次，将 f_1 按通道维度进行展开，得到 $X'' \in R^{HW \times L}$ ；再次，对 f_2 在 HW 维度上进行 Softmax 操作，得到注意力图 $p \in R^{HW \times L}$ ；接着，按通道对输入 f 进行直接展开，得到特征 $R \in R^{HW \times C}$ ；最后，将 p 转置后与 f' 进行矩阵相

乘，得到 $T \in R^{L \times C}$ ，作为Transformer的输入。具体公式为：

$$T = (p)^T R = (\sigma((\phi(X; w))^{reshape}))^T X^{reshape} \quad (2)$$

式中： ϕ 为采用卷积核为 $w \in R^{L \times C}$ 的逐点卷积操作； σ 为 Softmax 函数。

可以看出，网络中的Token数量 L 由逐点卷积中卷积核的数量决定，可以自由设定；每个Token长度为输入 X 的通道数 C 。

2.3 Transformer 模块

Transformer 模块结构如图3所示，由位置编码、编码器与解码器组成。其中， ϕ 表示采用卷积核为 $w \in R^{L \times C}$ 的逐点卷积操作； σ 为 Softmax 函数。

2.3.1 位置编码

位置信息有助于上下文建模，可以指导Transformer利用上下文信息。双时相图像经过语义标记模块后得到 $T_1 \in R^{L \times C}$ 与 $T_2 \in R^{L \times C}$ 两个输出，对两者进行拼接得到 $T' \in R^{2L \times C}$ ，通过加入位置信息得到新的Token集合 $T'' \in R^{2L \times C}$ 。

2.3.2 编码器

编码器由归一化(Normalization, Norm)、多头自注意力(Multi-Head Self Attention, MSA)与多层感知器(Multilayer Perceptron, MLP)层组成，结构如图3(a)所示。为提高MSA与MLP训练的稳定性，在两者前面分别加入Norm来帮助稳定梯度流，缓解内部协变量偏移情况。MSA通过引入多个注意力头使模型能够在不同的表示子空间中学习各自不同的关系，从而提升模型的代表能力。自注意力模块为根据输入 $T'' \in R^{2L \times C}$ 计算得出的3个矩阵：查询 Q 、键 K 与值 V 。具体表示为：

$$Q = T'' W^q \quad (3)$$

$$K = T'' W^k \quad (4)$$

$$V = T'' W^v \quad (5)$$

式中： $W^q, W^k, W^v \in R^{C \times d}$ 为3个可学习的线性投影层。参照BIT网络，维度 d 设置为8。

一个注意力头可表示为：

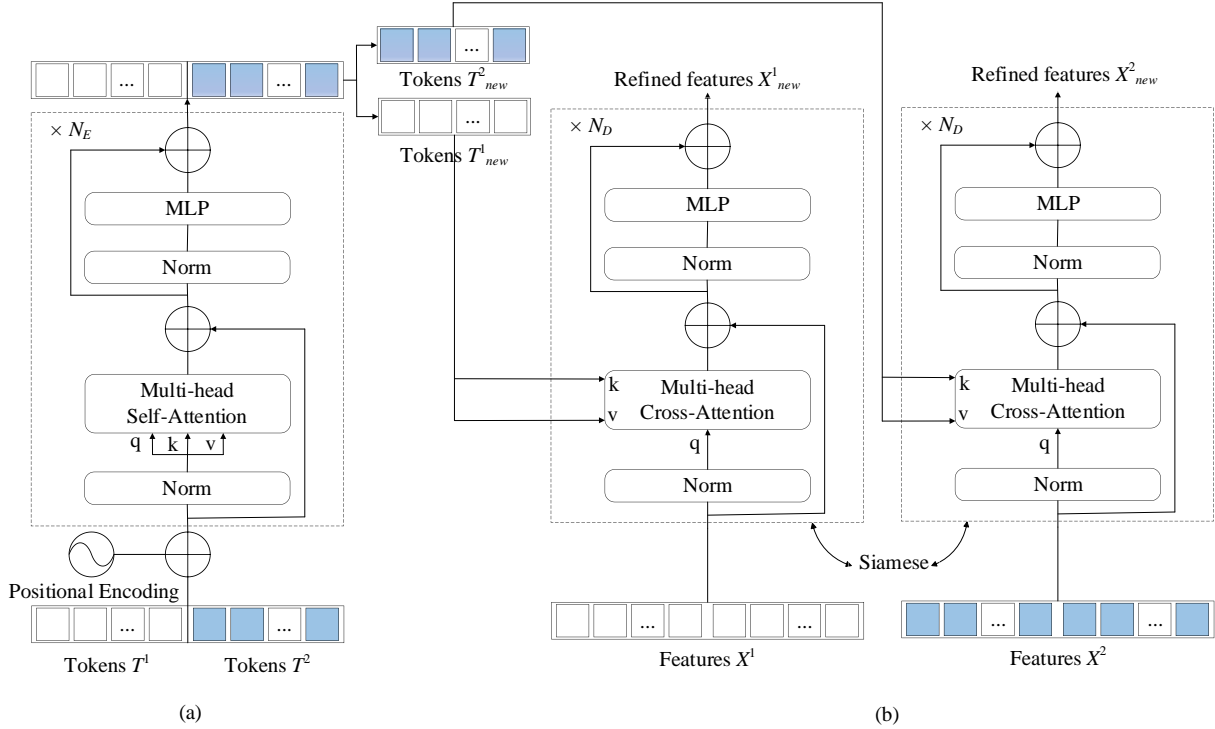


Fig. 3 Transformer structure

图 3 Transformer 结构

$$Att(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (6)$$

式中: $\sigma(\cdot)$ 为在维度 d 上的 Softmax 函数。

MSA 通过并行多个自注意力头, 将不同注意力头的输出融合起来投影成一个新的输出。具体公式为:

$$head_j = Att(T^q W_j^q, T^k W_j^k, T^v W_j^v) \quad (7)$$

$$MSA(T^m) = Contact(head_1, \dots, head_h)W^O \quad (8)$$

式中: h 为自注意力头的数量, 设定为 8; $Contact$ 为维度 d 方向上进行的拼接操作; $W_j^q, W_j^k, W_j^v \in R^{C \times d}$ 为第 j 个头的 3 个线性投影层; $W^O \in R^{hd \times C}$ 为一个线性投影层。

MLP 由两个线性变换层组成, 其中第一层加入 GELU [53] 激活函数, 以提升非线性建模能力。公式为:

$$MLP(T^m) = GELU(T^m W_1) W_2 \quad (9)$$

式中: $W_1 \in R^{C \times 2C}, W_2 \in R^{2C \times C}$ 为线性投影层。

2.3.3 解码器

解码器由 Norm、多头交叉注意力 (Multi-Head Cross Attention, MCA) 与 MLP 层组成, 结构如图 3(b) 所示。编码器的输出为 $T_{new} \in R^{2L \times C}$, 对其在第一维度方向上进行分离, 得到 $T_{new}^1 \in R^{L \times C}, T_{new}^2 \in R^{L \times C}$, T_{new}^1 与 T_{new}^2 中包含的高级语义信息揭示了兴趣的变化。将这些高级信息表示投影回像素空间以获取像素级特征。与编码器的 MSA 不同, MCA 中的查询 Q 与键 K 来自于 Transformer 编码器的输出; 值 V 来自于上阶段卷积模块 Resnet 的输出, 为 $X \in R^{H \times W \times C}$, 按通道展平, 转化为 $X \in R^{HW \times C}$ 作为值 V 的输入。表示为:

$$head_j = Att(X^i W_j^q, T_{new}^i W_j^k, T_{new}^i W_j^v) \quad (10)$$

$$MSA(X^i, T_{new}^i) = Contact(head_1, \dots, head_h)W^O \quad (11)$$

式中: $W_j^q, W_j^k, W_j^v \in R^{C \times d}, W^O \in R^{hd \times C}$ 为线性投影层; h 为注意力头的个数; 维度 d 设置为 8。

2.4 残差模块

ResNet 中残差块的作用是通过引入跳跃连接与残差学习的机制解决深度神经网络中的梯度问题, 促进网络更好地学习与优化。这种模块化学习机制使网络可以更灵活地保留与传递卷积以及 Transformer 提取的信息, 从而平衡对局部与全局信息的利用。

设双时相图像经过第一阶段卷积模块后的输出为 $X_1 \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times C}$ 与 $X_2 \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times C}$, 经过 Transformer 解码器的输出为 $F_1' \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times C}$ 与 $F_2' \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times C}$ (H_0 与 W_0 分别为原始图像的高度和宽度)。分别对 X_1 与 X_2, F_1' 与 F_2' 进行差分, 得到局部与全局信息两个阶段的差异信息。利用 ResNet 基本残差块将两阶段差异信息融合起来, 得到输出 $K \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times C}$ 。具体公式为:

$$\bar{X} = |X_1 - X_2| \quad (12)$$

$$\bar{F} = |F_1' - F_2'| \quad (13)$$

$$K = ReLU(BN(Conv(Conv_{3 \times 3}, \bar{X})) + \bar{F}) \quad (14)$$

式中: $|\cdot|$ 为绝对值; $Conv$ 为普通 3×3 卷积, 卷积核个数为 C ; BN 为批标准化; $ReLU$ 为激活函数; $Conv_{3 \times 3}$ 为基本的卷积块, 由 $Conv, BN$ 与 $ReLU$ 组成。

2.5 预测模块

预测模块的输入为 $K \in R^{\frac{H_0}{4} \times \frac{W_0}{4} \times C}$, 即来自残差模块的输出。首先, 采用双线性差值法 *Bilinear* 对 K 进行上采样

恢复到原始图像大小;其次,经过卷积进行深度调整,使其通道数降为1;最后,采用 *sigmoid* 函数对其进行逐像素的预测生成 $P \in R^{h_0 \times w_0}$ 。具体公式为:

$$P = \text{sigmoid}(\text{Conv}(\text{Bilinear}(K))) \quad (15)$$

2.6 损失函数

在变化检测任务中,处理变化区域与未变化区域之间的类不平衡是一个常见挑战。为此,本文提出一种混合损失函数,将BCE损失与Dice损失结合起来。

BCE损失可以表示为:

$$L_{bce}(p, g) = - \sum_{n=1}^N [g_n \cdot \log p_n + (1 - g_n) \cdot \log(1 - p_n)] \quad (16)$$

Dice损失可以表示为:

$$L_{dice}(p, g) = 1 - \frac{2 \sum_{n=1}^N (p_n \times g_n)}{\sum_{n=1}^N p_n + \sum_{n=1}^N g_n} \quad (17)$$

式中: N 为一个样本中总的像素数; n 为样本中的一个像素数; p_n 为像素点 n 发生变化的预测概率值,来自预测头 *sigmoid* 的输出, $p_n \in [0, 1]$; g_n 为实际标签的值,对应0或1。

3 实验方法与结果分析

3.1 实验设置

实验运行环境基于PyTorch框架实现,采用RTX 2080Ti显卡进行训练,显存为11 GB。训练选用Adam优化器,训练轮次为140,批次大小设置为32,初始学习率设置为0.000 5,90个训练轮次后,学习率调整为原来的0.1倍。

3.2 数据集

LEVIR-CD是一个建筑变化检测数据集,由637对大小为1 024×1 024的高分辨率(0.5 m/pixel)遥感图像组成,涵盖了各种类型的建筑,如别墅住宅、高层公寓、小车库与大型仓库^[19]。参照BIT,对数据集进行不重叠裁剪,大小设置为256×256,并随机划分为训练集、验证集与测试集,分别有7 120、1 024和2 048对图像。

BCDD是一个建筑变化检测数据集,原始图像为一对32 507×15 354的高分辨率遥感图像,前期图像中包含的建筑数量有12 796个,后期图像中包含的建筑数量有16 077个^[20]。参照BIT,对数据集进行不重叠裁剪,大小设置为256×256,并随机划分为训练集、验证集与测试集,分别有6 095、762和763对图像。

3.3 评价指标

采用变化检测网络广泛使用的评价指标,包括Kappa系数(κ)、交并比(IoU)、F1分数(F1)、召回率(Rec)与精度(Pre),对所提网络性能进行全面评价^[21]。各评价指标的计算公式为:

$$Pre = \frac{TP}{TP + FP} \quad (18)$$

$$Rec = \frac{TP}{TP + FN} \quad (19)$$

$$F1 = \frac{2Pre \cdot Rec}{Pre + Rec} \quad (20)$$

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (21)$$

$$P_e = \frac{(TN + FN) \cdot (TN + FP) + (FP + TP) \cdot (FN + TP)}{(TP + FP + TN + FN)^2} \quad (22)$$

$$\kappa = \frac{OA - P_e}{1 - P_e} \quad (23)$$

$$IoU = \frac{TP}{TP + FP} \quad (24)$$

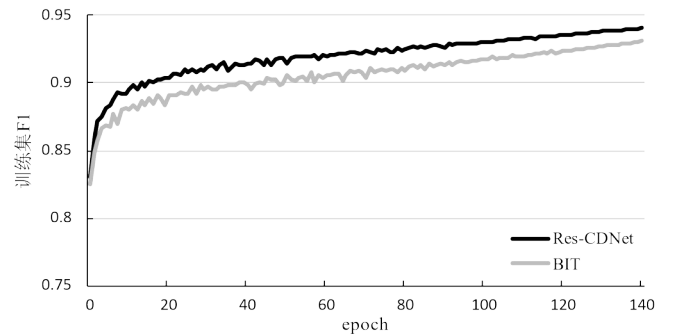
式中:TP、FP、TN与FN分别为真阳性、假阳性、真阴性与假阴性的数量。此外,本文还提供了模型参数与计算成本的详细信息以供参考。

3.4 实验结果与分析

为验证Res-CDNet网络的有效性,将其与FC-Siam-conc^[9]、FC-Siam-diff^[9]、SNUNet^[11]、IFNet^[12]、L-UNet^[22]以及BIT 6种主流变化检测方法进行比较。

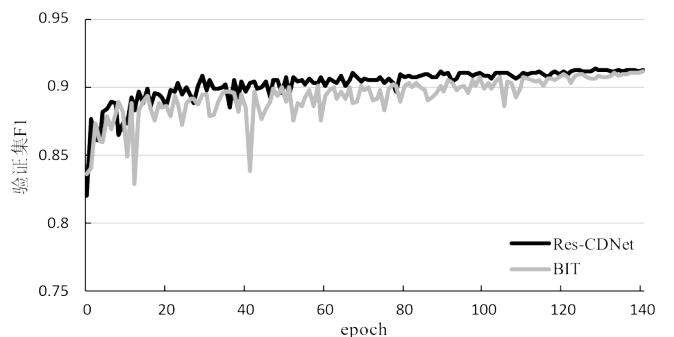
3.4.1 训练过程

在LEVIR-CD数据集上进行训练,图4为每个训练epoch训练/验证集的平均F1分数。可以看出,与BIT网络相比,Res-CDNet网络的训练精度和验证精度更高,表明Res-CDNet的训练更加稳定高效,泛化能力更强。



(a) F1 of training set

(a) 训练集F1



(b) F1 of verification set

(b) 验证集F1

Fig. 4 F1 scores of the training set and validation set in each training epoch

图4 每个训练epoch中训练集和验证集的F1分数

3.4.2 LEVIR-CD 数据集实验结果

表 1 为不同网络在 LEVIR-CD 数据集上的检测结果。可以看出,在处理 LEVIR-CD 这样的小目标建筑时,Res-CDNet 网络的召回率优于其他模型,表明其变化漏检的情况较少。与 BIT 网络相比,本文网络在参数量增加不到 0.2M 的情况下,Rec、K、IoU 与 F1 4 个指标均更优,表明本文改进方法有效。然而,本文方法与 IFNet 网络有一定差距,尤其是准确率 Pre 低了 1.71%,说明其在伪变化检测方面有很大不足。然而,IFNet 网络的优势也伴随着较大参数量与计算量的劣势,且对硬件有着较高要求。FC-Siam-diff 与 FC-Siam-conc 虽然参数较少,但过于简单的网络结构影响了其检测性能。

图 5 为不同网络在 LEVIR-CD 数据集上的检测结果示

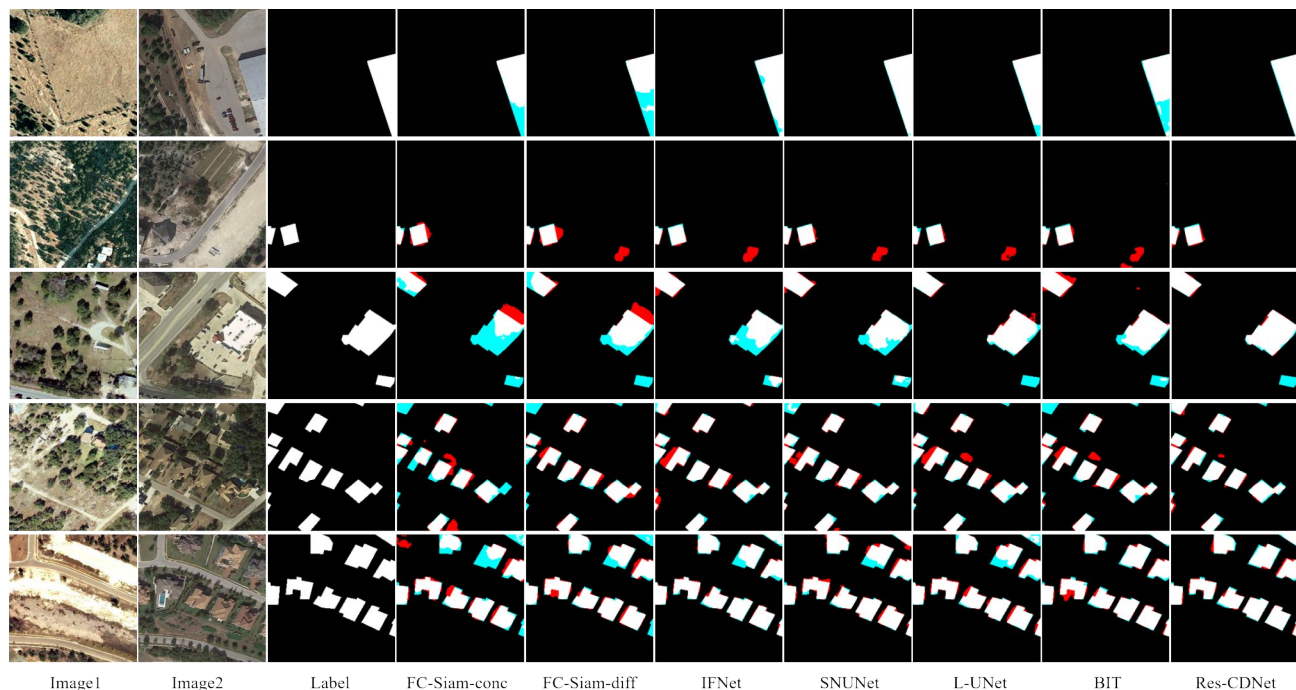


Fig. 5 Examples of detection results of different networks on the LEVIR-CD dataset

图 5 不同网络在 LEVIR-CD 数据集上的检测结果示例

3.4.3 BCDD 数据集实验结果

表 2 为不同网络在 BCDD 数据集上的检测结果。可以看出,在处理 BCDD 这样的大目标建筑数据集时,Res-CDNet 网络在 K、IoU、F1、Rec 与 Pre 几项指标上均优于其他模型,分别比第二名高出 0.71%、1.23%、0.69%、0.50% 与 0.58%,其中 IoU 指标提升显著,表明所提网络可以更准确地捕捉到目标区域,预测结果与实际目标更为一致。

图 6 为不同网络在 BCDD 数据集上的检测结果示例。可以看出,与其他几个网络相比,本文网络假阳性与假阴性的比例均较低,尤其是对于第四行这种亮度不均的建筑具有更少的漏检区域与错误区域。

3.4.4 消融实验

为验证本文网络与损失函数的有效性,在 LEVIR-CD 数据集上进行消融实验。结果见表 3。可以看出,分别加

Table 1 Detection results of different networks on the LEVIR-CD dataset

网络	K	IoU	F1	Rec	Pre	Params/M	FLOPs/G
FC-Siam-conc	85.64	75.99	86.36	85.11	87.65	1.545	10.643
FC-Siam-diff	87.85	79.31	88.46	87.12	89.84	1.350	9.435
IFNet	90.82	83.95	91.28	89.20	93.45	50.442	164.513
SNUNet	90.02	82.68	90.52	89.42	91.65	10.198	88.740
L-UNet	89.63	82.07	90.15	88.68	91.67	8.448	34.662
BIT	90.01	82.67	90.51	88.86	92.23	3.006	16.968
Res-CDNet	90.54	83.52	91.02	90.31	91.74	3.025	17.121

例,其中白色代表真阳性,红色代表假阳性,黑色代表真阴性,蓝色代表假阴性。虽然 Res-CDNet 与 IFNet 在定量指标上存在差距,但是其对小尺寸建筑目标的检测结果更加整齐平滑。

入残块模块和 Dice 损失函数均能提高 K、IoU、F1 和召回率,同时加入两者可进一步提升大部分指标,但对准确率有一定削弱。

4 结语

本文提出一种基于 BIT 的遥感图像变化检测网络,通过加入残差网络将卷积部分的差异信息与 Transformer 后的差异信息融合起来,平衡了全局与局部信息的提取。与 BIT 网络相比,所提网络在 LEVIR-CD 与 BCDD 数据集上的 F1 分数均有所提升,而参数量仅增加 0.02 M。在 BCDD 数据集上的实验结果表明,所提网络对较大建筑对象的变化检测效果较其他主流网络有一定优势。受光照以及数据质量的影响,建筑边缘变化检测往往存在困难。后续拟

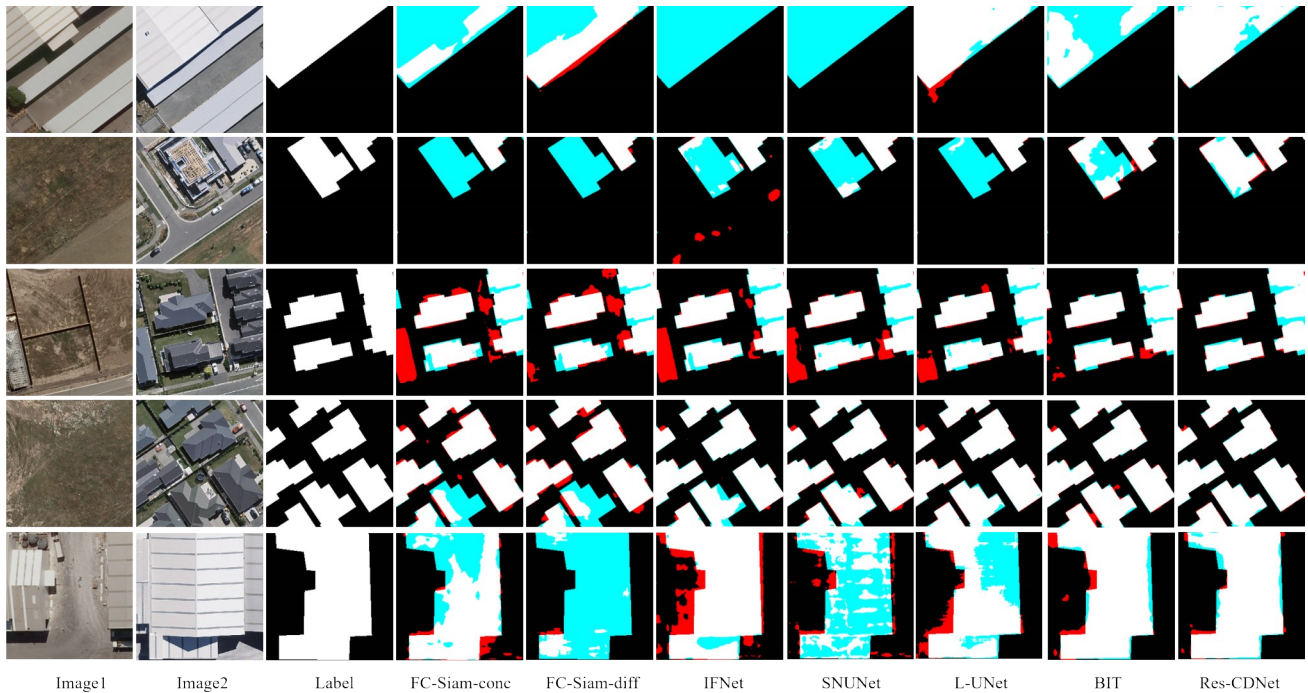


图 6 不同网络在 BCDD 数据集上的检测结果示例

图 6 不同网络在 BCDD 数据集上的检测结果示例

Table 2 Detection results of different networks on the BCDD dataset

表 2 不同网络在 BCDD 数据集上的检测结果

网络	K	IoU	F1	Rec	Pre	Params/M	FLOPs/G
FC-Siam-conc	80.81	68.92	81.60	79.18	84.18	1.545	10.643
FC-Siam-diff	82.17	70.79	82.89	78.80	87.43	1.350	9.435
IFNet	88.21	79.66	88.68	82.80	95.45	50.442	164.513
SNUNet	90.61	83.49	91.00	88.07	94.13	10.198	88.740
L-UNet	91.56	85.02	91.90	88.83	95.20	8.448	34.662
BIT	93.64	88.51	93.90	92.69	95.15	3.006	16.968
Res-CDNet	94.35	89.74	94.59	93.19	96.03	3.025	17.121

Table 3 Ablation test results

表 3 消融实验结果

网络	损失函数	K	IoU	F1	Rec	Pre
BIT	BCE	90.01	82.67	90.51	88.86	92.23
BIT	BCE+Dice	90.15	82.88	90.64	89.16	92.17
Res-CDNet	BCE	90.44	83.35	90.92	89.70	92.17
Res-CDNet	BCE+Dice	90.54	83.52	91.02	90.31	91.74

通过以下方式改善:一是通过多源数据融合(如结合高分辨率影像和激光雷达数据)来丰富边缘信息^[23];二是引入深度学习和自适应滤波等技术,提升在复杂环境下的边缘提取精度^[24-25];三是优化预处理算法以减少光照变化带来的误差^[26],从而提高检测效果。

参考文献:

[1] ZHANG L P, WU C. Current status and prospects of change detection in multi-temporal remote sensing images[J]. Journal of Surveying and Mapping, 2017, 46(10): 1447-1459.
张良培, 武辰. 多时相遥感影像变化检测的现状与展望[J]. 测绘学报, 2017, 46(10): 1447-1459.

[2] SUI H G, FENG W Q, LI W Z, et al. A review of change detection methods

in multi-temporal remote sensing images[J]. Journal of Wuhan University (Information Science Edition), 2018, 43(12): 1885-1898.
眭海刚, 冯文卿, 李文卓, 等. 多时相遥感影像变化检测方法综述[J]. 武汉大学学报(信息科学版), 2018, 43(12): 1885-1898.

[3] SINGH A. Review article digital change detection techniques using remotely-sensed data[J]. International Journal of Remote Sensing, 1989, 10(6): 989-1003.

[4] CHEN J, LU J, ZHU X, et al. Generative semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7111-7120.

[5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of the Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[6] LIU H, ZHANG C, DENG Y, et al. TransIFC: invariant cues-aware feature concentration learning for efficient fine-grained bird image classification[J]. Transactions on Multimedia, 2023, 1: 1-14.

[7] LYU H, LU H, MOU L. Learning a transferable change rule from a recurrent neural network for land cover change detection[J]. Remote Sensing, 2016, 8(6): 506.

[8] GONG M, NIU X, ZHANG P, et al. Generative adversarial networks for change detection in multispectral imagery [J]. Geoscience and Remote Sensing Letters, 2017, 14(12): 2310-2314.

[9] DAUDT R C, LE SAUX B, BOULCH A. Fully convolutional siamese networks for change detection[C]//2018 25th International Conference on Image Processing, 2018: 4063-4067.

[10] ZHOU Z, RAHMAN S M M, TAJBAKHSH N, et al. Unet++: a nested u-net architecture for medical image segmentation [C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, 2018: 3-11.

[11] FANG S, LI K, SHAO J, et al. SNUNet-CD: a densely connected siamese

- network for change detection of VHR images[J]. *Geoscience and Remote Sensing Letters*, 2021, 19: 1-5.
- [12] ZHANG C, YUE P, TAPETE D, et al. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images[J]. *Journal of Photogrammetry and Remote Sensing*, 2020, 166: 183-200.
- [13] CHEN C F R, FAN Q, PANDA R. Crossvit: cross-attention multi-scale vision transformer for image classification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 357-366.
- [14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[DB/OL]. <http://arxiv.org/pdf/2010.11929>.
- [15] LIANG J, CAO J, SUN G, et al. Swinir: image restoration using swin transformer[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 1833-1844.
- [16] HAN K, XIAO A, WU E, et al. Transformer in transformer[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 15908-15919.
- [17] CHEN H, QI Z, SHI Z. Remote sensing image change detection with transformers[J]. *Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-14.
- [18] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//*2016 Fourth International Conference on 3D Vision (3DV)*, 2016: 565-571.
- [19] CHEN H, SHI Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection[J]. *Remote Sensing*, 2020, 12(10): 1662.
- [20] JI S, WEI S, LU M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set[J]. *Transactions on Geoscience and Remote Sensing*, 2018, 57(1): 574-586.
- [21] JIANG M, ZHANG X C, SUN Y, et al. High resolution remote sensing image change detection network based on full-scale feature aggregation[J]. *Journal of Surveying and Mapping*, 2023, 52(10): 1738-1748.
- 姜明,张新长,孙颖,等. 全尺度特征聚合的高分辨率遥感影像变化检测网络[J]. *测绘学报*, 2023, 52(10): 1738-1748.
- [22] PAPADOMANOLAKI M, VAKALOPOULOU M, KARANTZALOS K. A deep multitask learning framework coupling semantic segmentation and fully convolutional LSTM networks for urban change detection[J]. *Transactions on Geoscience and Remote Sensing*, 2021, 59(9): 7651-7668.
- [23] LIU X L, WANG Z H, XING J H, et al. Identification and monitoring of tailings ponds by integrating multi-source geographic data and high-resolution remote sensing images: a case study of Gejiu City, Yunnan Province[J]. *Natural Resources Remote Sensing*, 2024, 36(1): 103-109.
- 刘晓亮,王志华,邢江河,等. 融合多源地理数据与高分辨率遥感影像的尾矿库识别与监测——以云南省个旧市为例[J]. *自然资源遥感*, 2024, 36(1): 103-109.
- [24] TANG P, LIU X, JIN X, et al. Neural network generation of adaptive filters and new applications in remote sensing image processing[J]. *Journal of Remote Sensing*, 2023, 27(7): 1523-1533.
- 唐婷,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[J]. *遥感学报*, 2023, 27(7): 1523-1533.
- [25] WU C Z, DONG H, FANG J. Adaptive filtering remote sensing image segmentation network based on attention mechanism[J]. *Computer Engineering and Science*, 2022, 44(11): 2010-2018.
- 吴从中,董浩,方静. 基于注意力机制的自适应滤波遥感图像分割网络[J]. *计算机工程与科学*, 2022, 44(11): 2010-2018.
- [26] HE J X. Research on preprocessing algorithm for infrared remote sensing images and its FPGA implementation[D]. Wuhan: Huazhong University of Science and Technology, 2022.
- 何佳欣. 红外遥感图像预处理算法研究及其FPGA实现[D]. 武汉: 华中科技大学, 2022.

(责任编辑:尹晨茹,毛宛婷)