

注意力引导全尺度连接网络的高分辨率影像变化检测

刘英¹, 何雪¹, 李单阳², 岳辉¹, 魏嘉莉¹

1. 西安科技大学 测绘科学与技术学院, 西安 710054;

2. 中国地质大学(武汉) 国家地理信息系统工程技术研究中心, 武汉 430000

摘要: 针对普通跳跃连接缺乏从全尺度角度获取变化信息及编码器特征提取能力不足的问题, 本文设计了耦合注意力机制 CBAM (Convolutional Block Attention Module) 的 UNet+++ 高分辨率遥感影像变化检测网络 CBAM UNet+++。CBAM UNet+++ 基于融合全尺度特征的语义分割结构 UNet+++，同时替换基本卷积单元为残差注意力模块 ResBlock_CBAM (Residual Block_CBAM) 抑制背景影响, 增强编码器对显著特征的学习能力, 并利用两种不同变化类型的高分辨率遥感影像变化检测数据集进行验证。结果表明: 该方法在 LEBEDEV 多地物变化数据集上取得最高精度, F1 (F1-Score) 和 OA (Overall Accuracy) 值分别为 88.9%、97.3%; 在 LEVIR-CD 建筑物变化数据集上取得次高精度, F1 和 OA 值分别为 86.7%、96.8%; 同时, 该法能有针对性的获取深层语义, 定性结果优于其他基准网络。

关键词: 遥感, 变化检测, UNet+++，注意力机制, 编码解码

中图分类号: TP701/P2

引用格式: 刘英, 何雪, 李单阳, 岳辉, 魏嘉莉. 2024. 注意力引导全尺度连接网络的高分辨率影像变化检测. 遥感学报, 28(4): 1052-1065

Liu Y, He X, Li D Y, Yue H and Wei J L. 2024. CBAM UNet+++ : Attention mechanism to guide change detection studies of full-scale connected networks. National Remote Sensing Bulletin, 28(4) : 1052-1065 [DOI: 10.11834/jrs.20221548]

1 引言

遥感影像变化检测 CD (Change Detection) 是对同一场景不同时间获取的两幅或多幅影像进行变化分析的技术, 被广泛应用于土地资源管理、农林监测、城市扩张及军事监管等领域 (刘统飞, 2020)。该技术易受传感器噪声、光线变化、季节变迁等因素影响 (Stramondo 等, 2006), 因此, 许多传统像素级方法如图像差分、主成分分析、变化矢量分析、迭代慢特征分析等可在简单场景中获得有效变化检测结果, 但在复杂场景容易造成结果细碎化, 不利于后续应用 (袁洲 等, 2021)。深度学习方法能从监督样本中自适应学习低层视觉特征到高层语义特征的层次化表达, 避免人为

设计特征对先验知识的依赖 (Tian 等, 2013), 得益于其天然优势, 近年来被广泛应用于变化检测。早期卷积神经网络如 AlexNet (Krizhevsky 等, 2012)、VGG Net (Simonyan 和 Zisserman, 2015) 包含全连接层, 存储开销大、计算效率低。Long 等 (2015) 提出全卷积神经网络 FCN (Fully Convolutional Networks), 用卷积替代全连接, 保留空间信息并提升计算效率, 接受任意大小的输入影像, 实现对图像像素端到端的分类, 但精细化程度不够; Ronneberger 等 (2015) 继 FCN 提出 UNet, UNet 使用编码解码结构及逐层跳连接, 提高对浅层位置特征利用率和像素定位精度, 但编码和解码器特征跨度过大, 容易生成边界粗糙及内部紧凑性差的结果图; 受 DenseNet 启发, Peng 等 (2019) 改

收稿日期: 2021-08-10; 预印本: 2022-04-07

基金项目: 陕西省自然科学基金基础研究计划 (编号: 2020JM-514); 国家自然科学基金 (编号: 41401496); 西安科技大学优秀青年科技基金 (编号: 2019YQ3-04)

第一作者简介: 刘英, 研究方向为环境遥感。E-mail: liuying712100@163.com

进具有密集连接的变化检测网络 UNet++, 改变卷积块为特征提取能力更强的残差块, 同时在输出端引入深度监督和多侧输出融合机制, 但该方法缺乏从全尺度角度探索足够信息。上述方法都被应用于 CD 任务并取得不错的效果, 但关键问题在于缺乏足够的空间分辨率信息及语义信息帮助变化图重建。Huang 等 (2020) 等设计全尺度连接网络 UNet+++ 并成功应用于医学影像分割任务, 全尺度跳连接能够充分学习低层位置信息和高层语义信息, 避免信息丢失。进一步研究中, 研究者发现与上下文相关的特征对 CD 任务至关重要, 在人类视觉中被称为注意力机制。注意力机制强调对变化类的学习, 能提升网络对变化类的敏感度, 还不会增加过多参数, 因此常被集成到神经网络中以优化网络性能。Zhang 等 (2020) 提出了深度监督融合网络, 通过孪生结构提取深度特征并输入到深度监督差分网络中, 交替使用空间和通道注意力模块以提高结果的边界完整性和内部紧凑性; 韩星等 (2022) 等在编码阶段利用 ResNet50 逐级提取图像特征, 应用空洞卷积及金字塔池化模块对编码器最深层特征进行多尺度特征提取, 在解码器阶段利用 CBAM 精细恢复图像分辨率特征。上述方法将注意力机制嵌入网络解码阶段, 能够有效筛选来自编码器的高层语义信息及低层空间信息, 提高检测精度, 但网络依赖于逐层卷积提取特征信息, 而缺少对显著特征的关注, 尤其对样本占比较少的类别易出现漏分、错分现象。

综合上述研究及分割任务和 CD 任务的相似性, 提出注意力引导的 UNet+++ 高分辨率遥感影像变化检测网络 CBAM UNet+++。首次将分割网络 UNet+++ 用于 CD 任务, 通过 ResBlock_CBAM 模块来自适应模拟空间上下文特征和通道语义特征, 避免巨大计算开销的同时聚焦对目标任务有利区域的学习。并在多地物和单一建筑物两种不同类型变化数据集上进行验证, 选择经典语义分割网络和 CD 网络进行对比分析, 验证设计有效性。

2 CBAM UNet+++ 网络结构设计

2.1 CBAM 模块

通道注意力模块 SE (Squeeze-and-Excitation-block) 只考虑通道信息, 而忽略位置信息重要性 (Wang 等, 2020)。CBAM (Woo 等, 2018) 尝试

在通道上进行全局池化来引入位置信息, 通道注意力模块可以学习到关键特征的具体内容, 空间注意力模块可以学习到关键特征的位置信息, 通过混合跨通道和空间信息对特征进行更好的自适应细化 (许文慧 等, 2021)。CBAM 整体结构如图 1 所示, 给定中间特征图作为输入, 首先利用特征间的通道关系生成通道注意力图 (式 (1)), 使用平均池化和最大池化操作沿着特征的宽度和高度方向聚合特征映射的空间信息, 生成特征转入到一个包含隐层的多层感知机 MLP (Multilayer Perceptron) 进行特征的降维升维操作, 使用求和和 ‘Sigmoid’ 操作来合并输出通道注意力权重系数, 将该权重系数和输入特征进行逐元素相乘, 生成空间注意模块需要的输入特征; 然后利用特征间的空间关系生成空间注意力图 (式 (2)), 在通道维度上进行平均池化和最大池化操作压缩得到一维通道特征图, 将结果进行连接并经过一个 7×7 卷积操作, 经过 ‘Sigmoid’ 激活生成空间注意力权重系数。将权重系数与该模块输入特征相乘, 得到经过空间注意力细化后的特征图。通过双注意力的使用, 从通道和空间两个维度上关注到重要特征, 过滤不重要特征 (向阳, 2020)。

$$F_c = F \times \left(\sigma \left(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)) \right) \right) = F \times \left(\sigma \left(\mathbf{W}_1 \left(\mathbf{W}_0(F_{\text{avg}}^c) \right) + \mathbf{W}_1 \left(\mathbf{W}_0(F_{\text{max}}^c) \right) \right) \right) \quad (1)$$

$$F_s = F_c \times \left(\sigma \left(f^{7 \times 7} \left([\text{AvgPool}(F); \text{MaxPool}(F)] \right) \right) \right) = F_c \times \left(\sigma \left(f^{7 \times 7} \left([F_{\text{avg}}^s; F_{\text{max}}^s] \right) \right) \right) \quad (2)$$

式中, F_c 和 F_s 分别是通道和空间注意力图; F 为输入特征; F_{avg}^c 和 F_{max}^c 分别表示平均池化和最大池化操作; σ 表示 Sigmoid 激活函数; $\mathbf{W}_0$ 为 $\frac{C}{r} \times C$ 的矩阵; $\mathbf{W}_1$ 为 $C \times \frac{C}{r}$ 的矩阵; $f^{7 \times 7}$ 表示卷积核大小为 7×7 的卷积操作。

2.2 CBAM UNet+++

普通跳连接或密集跳连接难以获取全尺度的语义信息, 且特征提取阶段往往仅采用卷积操作, 未能充分学习对目标任务有用特征。针对上述问题, 本文提出注意力引导的 UNet+++ 变化检测网络 CBAM UNet+++。

CBAM UNet+++ 结构如图 2 (a) 所示, 采用编解码结构, 在编码阶段利用 ResBlock_CBAM 进

行特征提取, 帮助网络对影像的空间场景和变化信息生成更高层次的语义理解。在解码阶段对来自编码器底端的稀疏特征图进行上采样操作, 同时连接来自于编码器的不同尺度特征信息, 帮助

网络获取来自全尺度特征的空间位置信息, 提高特征利用率及变化检测精度。网络各部分构造详细介绍如下。

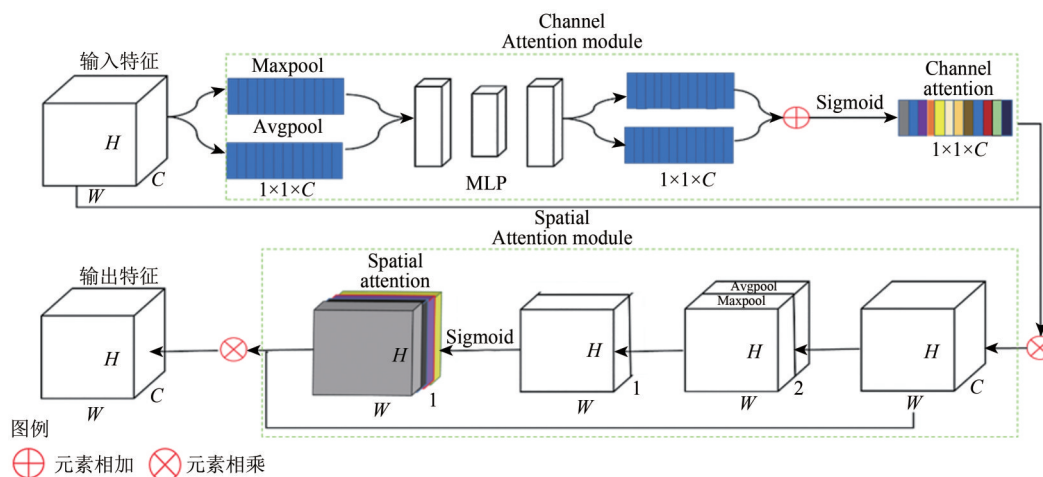


图1 CBAM模块

Fig. 1 CBAM module

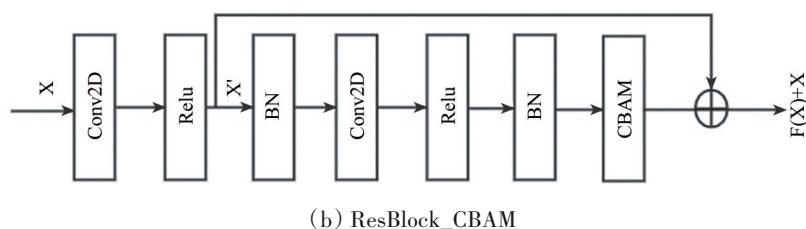
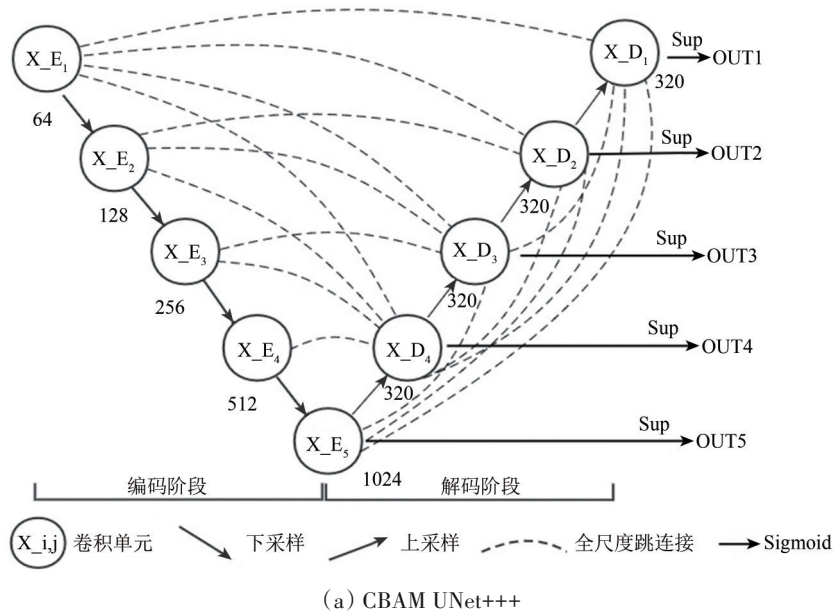


图2 CBAM UNet+++网络结构

Fig. 2 CBAM UNet+++ network structure

(1) 编码阶段利用5个 ResBlock_CBAM 生成输入图像对不同尺度大小的高级语义, ResBlock_

CBAM 的构建如图2 (b), 给定输入特征 X 要通过两层卷积单元和 CBAM 模块, 从高级特征图的通

道和空间聚合信息中突出特定尺度的显著特征,自适应的实现空间和通道的校准,生成注意力图 $F(x)$ (阮晨等, 2021),同时 X 在经过卷积和激活操作后(X')直接映射传输到后面层的输出结果中,最后通过相加操作得到输出结果。分别比较不同激活函数构建的ResBlock对于网络精度的影响(表1),最后选择Relu激活函数。

表1 不同残差块下UNet+++的精度

Table 1 Accuracy of UNet+++ under different residual blocks

类别	精度	召回率	总体精度	F1-Score
Res_module(Selu)	86.1	81.2	96.1	83.6
Res_module(Relu)	87.4	88.3	97.0	87.9

(2) 解码阶段需要构造全尺度特征图来融合编码器上不同尺度特征图的高级语义和空间位置信息。全尺度特征图的构造如图2(a)所示,以 X_{D_3} 为例,将较小尺度特征 X_{E_1} 和 X_{E_2} 通过不同步长(stride=4, stride=2)的最大池化操作进行下采样,以便传递底层位置信息,同时为了统一特征图的分辨率;将较大尺度特征 X_{D_4} 和 X_{E_5} 通过双线性插值法进行上采样从而放大特征图的分辨率;将上述统一大小后的4张特征图与同尺度特征图 X_{E_3} 进行通道数量的统一,减少冗余信息;然后通过通道拼接操作,产生320通道的全尺度特征图,再经过卷积、归一化、池化操作得到 X_{D_3} 。

(3) CBAM UNet+++进一步采用全尺度深度监督机制,通过对学习过程附加软约束,使网络能够增强浅层特征的表示和识别能力,加速收敛。CBAM UNet+++每个解码阶段都有一个侧输出,将其送入一个普通 3×3 卷积,然后是一个双线性上采样和Sigmoid函数来得到网络输出,通过损失函数计算损失值调整模型参数,损耗函数权值均设为1。此外,增加卷积滤波器数量,使得CBAM UNet+++能对有效特征进行充分学习,避免细粒度信息丢失。

3 实验设计

3.1 实验数据及预处理

3.1.1 LEBEDEV数据集

LEBEDEV(Lebedev等, 2018)数据集(图3(a))

是由谷歌Earth获取的具有季节变化的真实光学卫星图像。其包含7对没有附加物体的4725像素 $\times$ 2200像素的季节变化图像和4对有附加物体的1900像素 $\times$ 1000像素的季节变化图像。为便于训练,将原始图像剪切成约16000张大小为256 $\times$ 256的实时季节图像集,其中训练集10000张,验证集和测试集各3000张。

3.1.2 LEVIR-CD数据集

LEVIR-CD是一个大规模遥感建筑物变化检测数据集(图3(b)),包含637个超高分辨率(VHR, 0.5 m/像素)的Google Earth图像对,大小为1024 $\times$ 1024像素(Chen和Shi, 2020)。由于GPU显存限制,将LEVIR-CD数据集其分割为256 $\times$ 256像素的影像对,并剔除无变化像对。其中,训练集包括2917个影像对,验证集包括872个影像对,测试集包括430个影像对。文中实验均在训练集与测试集上执行,精度评价均在验证集上执行。

表2 数据集样本量

Table 2 Data set sample size

尺寸	训练/张	测试/张	验证/张	总计
256 $\times$ 256	10000	3000	3000	16000
256 $\times$ 256	2917	430	872	4219

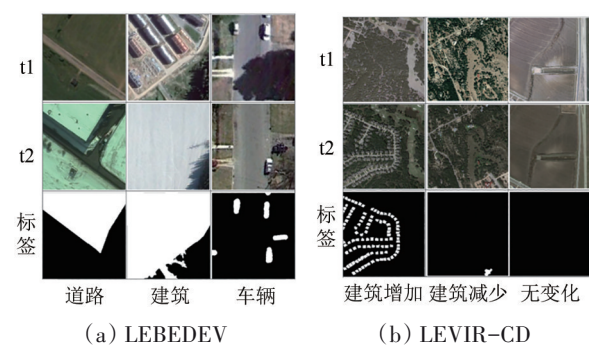


图3 数据集

Fig. 3 Data set

3.2 实验环境及参数设置

深度学习框架以Keras作前端,Tensorflow为后端,软件环境包括:操作系统为windows server2019, tensorflow版本为2.1, GPU加速库采用CUDA 10.1。此外,使用型号Tesla p100/16 g的GPU对训练过程进行加速,硬件环境包括:CPU为Intel Xeon Gold 6154, 64 G内存, 2 t固态硬盘。

使用Adam优化器,学习率设为 1×10^{-4} ,批处

理个数设为8, 共训练15个epoch。实验环境及超参设置见表3。

表3 实验环境及超参设置

Table 3 Experimental environment and hyperparameter settings

环境	型号	仿真参数	仿真值
CPU	Intel Xeon Gold 6154	学习率	1×10^{-4}
GPU	Tesla p100/16g	epoch	15
内存	64 G	动量	0.9
操作系统	windows server2019	参数优化	Adam
深度学习框架	Keras		
CUDA	10.1	Batch Size	8
语言	Python 3.7		

3.3 损失函数及精度评价

3.3.1 损失函数

(1) 平衡二元交叉熵 Bce (Balanced Binary Crossentropy)。为解决正负样本数量不平衡问题, Bce通过权重参数对正负样本进行加权处理。公式如下:

$$L_{bce} = \sum_{j \in \gamma_+} \log \Pr(y_j = 1) - (1 - \beta) \sum_{j \in \gamma_-} \log \Pr(y_j = 0) \quad (3)$$

式中, $\beta = |\gamma_-| / (|\gamma_+| + |\gamma_-|)$, $(1 - \beta) = |\gamma_+| / (|\gamma_+| + |\gamma_-|)$; $|\gamma_+|$ 和 $|\gamma_-|$ 分别代表标签值中变化和不变像素数量; $\Pr(y_j = 1)$ 和 $\Pr(y_j = 0)$ 分别代表像素 j 处预测结果是变化和不变像素的概率。

(2) 筛子损失函数 (Dice Loss)。Bce的权重需要人为敲定, 对正负样本的公正处理导致网络倾向把样本判为负例。为缓解数据类不平衡问题从而提高分类性能, 选择Dice Loss作为平衡二元交叉熵函数的补充, Dice loss倾向于对前景区域的挖掘, 从而解决前景占比较少的问题, 公式如下:

$$L_{dice} = 1 - \frac{2\gamma\hat{\gamma}}{\gamma + \hat{\gamma}} = 1 - L_{dice_coef} \quad (4)$$

式中, γ 和 $\hat{\gamma}$ 分别表示预测结果概率值和地面标签真值。

(3) 混合损失函数。结合Dice Loss和Bce设计混合损失函数, 公式如下:

$$L_{side} = L_{bce} + \omega L_{dice} \quad (5)$$

式中, L_{bce} 表示平衡二元交叉熵, L_{dice} 表示筛子损失, ω 起到平衡二者的作用。

3.3.2 精度评价指标

本文采用精度 P (Precision)、召回率 R

(Recall)、综合评价指标 F1 (F1-Score) 及总体精度 OA (Overall Accuracy) 4个定量评估指标。其中, 虚警率越低, P值越高, 漏检率越低, R值越高, F1、OA显示整体性能, 其值越大越好。

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (8)$$

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

式中, TP、TN和FP、FN分别表示变化检测结果中被正确划分为变化类和未变化类以及被错误划分为变化类和未变化类的像素个数。

3.3.3 损失函数的影响

ω 用于平衡Dice loss和Bce, 而损失函数对网络训练起着举足轻重的作用, 因此, 为验证 ω 的灵敏度, 将 ω 值从0变到1, 统计相应指标, 寻找最优值。如图4所示, 当 ω 设为0时, 只计算Bce, 其F1和OA值都维持在较低水平, 随着 ω 不断增加, F1和OA也不断增大, ω 在0.5时各指标达到最大值, 然后呈减小趋势。因此, 为提高CD性能, 本文设 ω 为0.5。

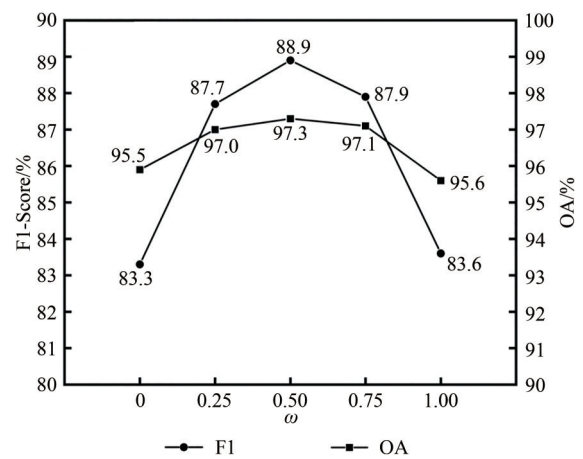


图4 ω 大小对于网络精度的影响

Fig. 4 The impact of ω on network accuracy

3.4 对比方法

为验证本文方法的优越性和有效性, 选择部分变化检测方法进行比较, 具体如下:

(1) UNet。UNet包含两条串联路径, 下采样路径用来提取图像特征, 捕捉上下文语义信息, 将图像压缩为特征图; 上采样路径用来精确定位,

将提取特征还原为原始图像尺寸。利用上采样和下采样过程中相同大小特征图进行跳连接,以整合更多图像信息进行分类。

(2) UNet MSOF。UNet++由不同深度U网组成,解码器引入密集跳连接,将相同大小特征图进行连接。考虑反向传播过程中中间上采样部分接收不到梯度信息,因此采用深度监督策略。利用多侧输出融合策略连接不同层的输出结果,产生最终精细化分类后的变化图。

(3) Attention UNet。Attention UNet (Oktay等, 2018)模型以UNet模型为基础,编码阶段和UNet完全相同,区别在于Attention UNet在解码时,添加了注意力门,生成对应层级权重图,将权重图和和编码器上结构化特征进行相乘,从而强调本层显著性区域特征。

(4) Attention UNet++。Attention UNet++ (Li等, 2020)使用UNet++作为基本网络框架。不同的是,在下采样和跳连接路径横向连接的拼接操作后嵌入注意力门,使用不同层次提取的特征在拓展路径中进行集中选择合并,减少冗余特征数量,提升UNet++准确性。

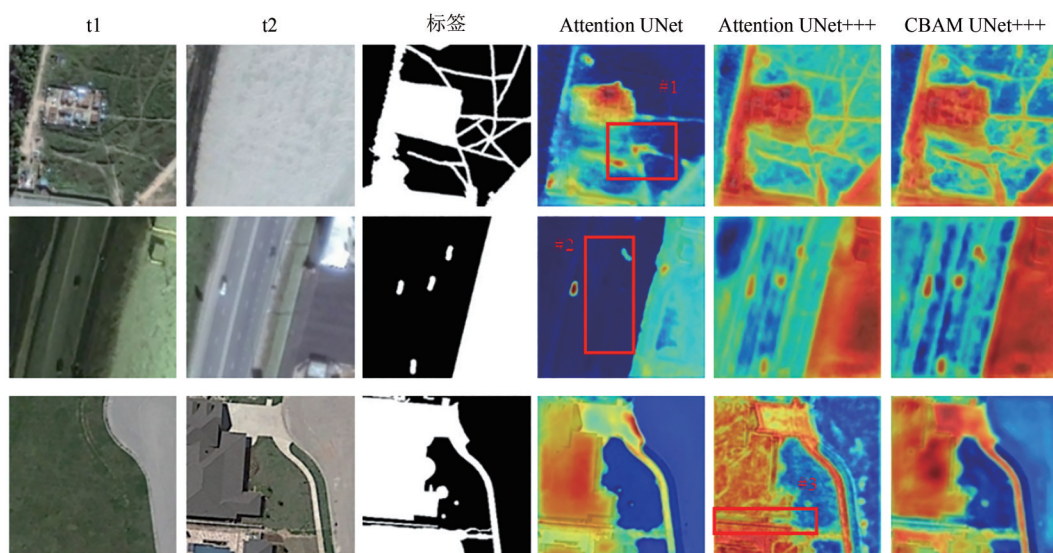
(5) FC Siam diff。FC Siam diff (Daudt等, 2018)基于FC EF架构,套用Siamese结构,编码器部分利用孪生结构提取不同时相特征及差异性信息,解码过程中对特征进行重构,利用转置卷积进行上采样。将差分图像和重构图像连接后通过反卷积恢复图像大小。

(6) ResNet50_ppm。ResNet50_ppm基于残差网络ResNet50提取图像特征,并为解码器最深层特征引入空洞卷积及金字塔池化模块来增大感受野提取多尺度特征,在解码阶段运用CBAM模块进行信息筛选,通过聚合不同尺度的特征实现信息重用。注意ResNet50_ppm为本文对其网络的称呼,原文中并未对所设计网络进行命名。

4 结果分析

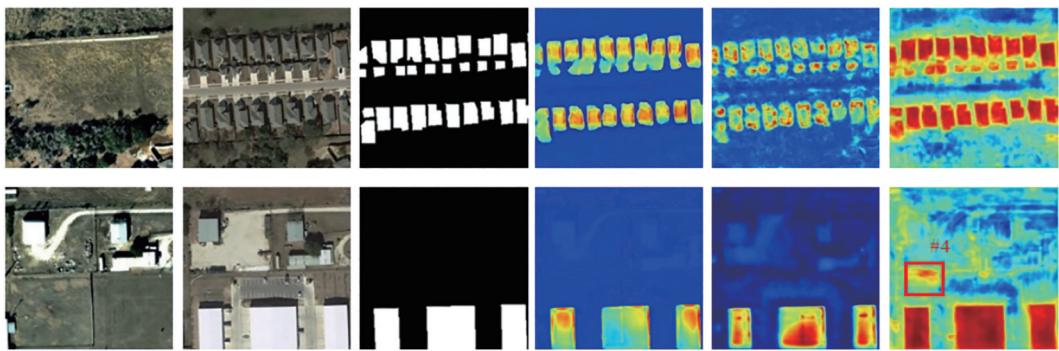
4.1 注意力机制响应结果分析

在进行变化检测结果分析前,文章比较注意力集成网络(Attention UNet、Attention UNet++、CBAM UNet++)可视化结果来说明CBAM模块有效性。运用梯度加权激活映射(Selvaraju等, 2017)方法分别在两个数据集上对变化像素进行定位,通过热图和特征图图叠加以达到可视化的目的。由图5可知,无论对多地物或单一建筑物的变化检测,不同网络在图像中寻找的是变化区域。针对LEBEDEV数据集,CBAM UNet+++凸显了道路、建筑所在大部分区域,Attention UNet未能关注到道路及部分车辆变化(#1、#2),Attention UNet++则错误的定位到未变化草地(#3);针对LEVIR-CD数据集,CBAM UNet+++对目标区域的关注程度高于其他方法,但未能察觉#4处伪变化,原因是该数据集多由小尺度密集建筑物群构成,一定程度上会增加网络在大尺度建筑物场景下的误判。



(a) LEBEDEV 数据集

(a) LEBEDEV datasets



(b) LEVIR-CD数据集
(b) LEVIR-CD datasets

图5 注意力机制响应结果图

Fig. 5 The response result of the attention mechanism

综上，集成CBAM优势的UNet+++方法能更好的利用目标区域的信息并从中聚合特征，受伪变化干扰小，对多地物变化场景更有效，在单一建筑物变化场景下，CBAM UNet+++错误定位到部分伪变化区域，但该方法在上述两个数据集上对变化区域的响应程度仍优于基准方法。

4.2 LEBEDEV 结果分析

由表4可知CBAM UNet+++方法的精度、召回率、F1、OA分别是88.9%、88.9%、88.9%、97.3%，在LEBEDEV数据集上展现最优性能。相较UNet、Attention UNet、FC Siam diff方法，其F1提高13.2%、8.1%、10.9%，OA提高3.4%、2%、2.6%。与性能优越的UNet++ MSOF、Attention UNet++和ResNet50_ppm相比，F1提高3.9%、1.4%、8.1%，OA提高1.1%、0.4%、1.8%。与使用普通卷积的UNet+++相比，其F1提高3.2%，OA提高0.8%。

表4 不同方法的变化检测精度评价

Table 4 Evaluation of change detection accuracy of different methods

/%					
类别	方法	精度	召回率	F1-Score	OA
不添加注意力机制	UNet	74.9	76.5	75.7	93.9
	FC Siam diff	80.5	75.6	78.0	94.7
	UNet++ MSOF	84.8	85.2	85.0	96.2
	UNet+++	86.5	85.0	85.7	96.5
添加注意力机制	Attention UNet	81.8	79.9	80.8	95.3
	Attention UNet++	86.8	88.3	87.5	96.9
	ResNet50_ppm	85.6	76.5	80.8	95.5
	CBAM UNet+++	88.9	88.9	88.9	97.3

注：粗体为精度最优。

为对各方法的优劣性做出进一步的评价，本文从图像级别上统计各范围内的图片张数，定位低精度结果图上的变化类型。将各评价指标按数值大小划分为5个等级，依次表示检测效果由差到好。由图6可知CBAM UNet+++方法的OA主要集中在0.95—1.0范围内，其余3项指标主要集中在0.8—1.0范围内；UNet方法在低精度区间内的图片张数较多，该模型的定量结果也是最差；Attention UNet++方法在0—0.2范围内的图片数量多于CBAM UNet+++，这也极大影响Attention UNet++的整体评价精度。查看验证集中处于低精度范围内图像变化类型，发现Attention UNet对道路的检测存在弱势；UNet和FC Siam diff在各类别的检测结果上都存在误差，CBAM UNet+++在大部分变化图像上检测精度都是最优的。

为直观展示检测结果，图7给出车辆、道路、建筑物变化的可视化结果和双时相影像及其对应标签。分析结果如下：

(1) 从整体上来看，无论是车辆、道路还是建筑物变化，CBAM UNet+++结果图与标签最相似，而其他基准方法的结果存在漏检、误检、边缘模糊等部分或全部问题。

(2) 对于道路变化（图7（a）），UNet和ResNet50_ppm的结果误差较大，在#1出现断连；UNet++MSOF、Attention UNet、Attention UNet++、FC Siam diff、UNet+++受到噪声干扰出现误检现象（#2、#3）；CBAM UNet+++不存在上述现象，且完美刻画道路形状，但不难发现，上述方法对道路锯齿化边界都存在着一定的过平滑现象。

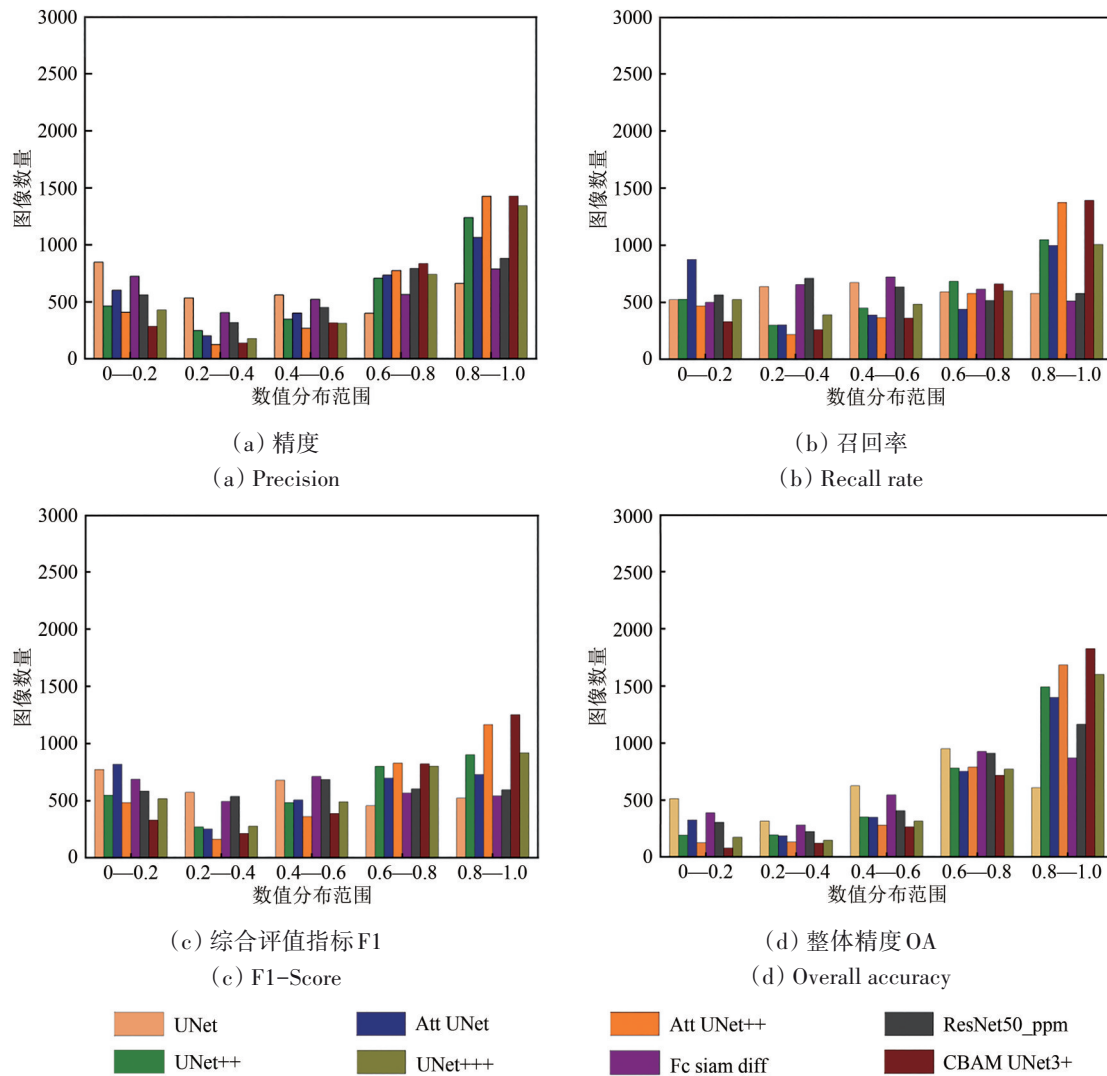
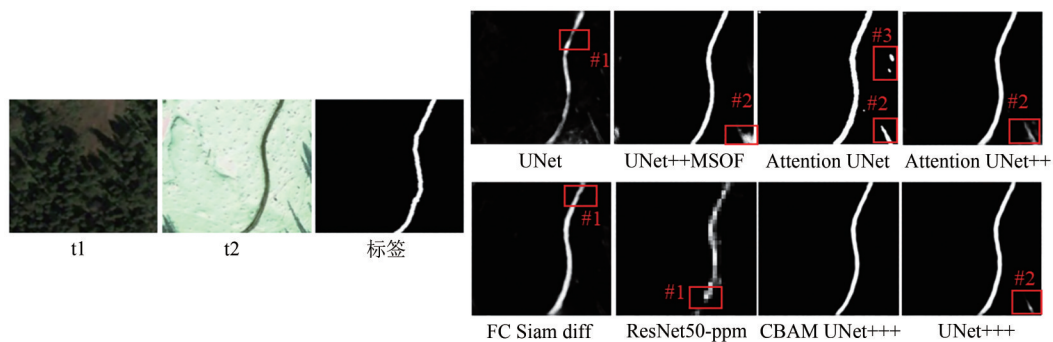


图6 从图像级别出发不同方法在验证集上的精度分布

Fig. 6 Accuracy distribution of different methods on the verification set from image level

(3) 对于不规则建筑物变化 (图 7 (b)), UNet 和 ResNet50_ppm 结果存在边界模糊现象, 配有空洞卷积和金字塔池化模块的 ResNet50_ppm 出现了边界丢失现象 (#1), 可能是由于过大的下采样率所致; UNet、UNet++ MSOF、FC Siam diff 和

UNet+++ 结果出现空洞 (#2), UNet++ MSOF、UNet+++ 方法在 #3 处存在断连; Attention UNet 在 #4 处存在大片误检且未能检测出 #5 处房屋的轮廓, 说明 Attention UNet 大尺度特征的局部变化难以模拟。CBAM UNet+++ 在复杂建筑物场景下更适用。



(a) 道路变化

(a) Road changes

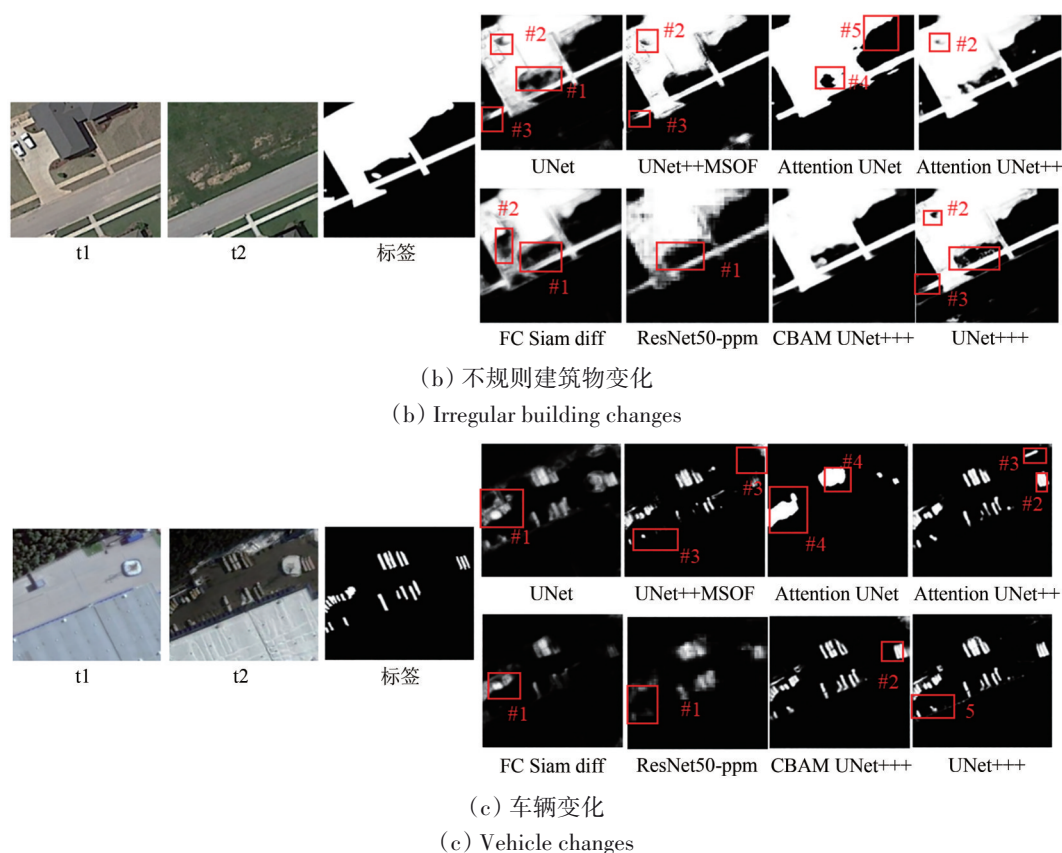


图7 变化检测结果可视化比较

Fig. 7 Visual comparison of change detection results

(4) 对于独立车辆变化 (图7 (c)), UNet、FC Siam diff 及 ResNet50_ppm 方法存在漏检、误检、粘连等部分及全部问题 (#1), 说明上述方法对小尺度特征学习能力不足; UNet++ MSOF、Attention UNet++存在粘连或误检现象 (#2、#3); Attention UNet 方法粘连最严重 (#4), 已不能辨认车辆轮廓; UNet+++在#5处存在漏检, CBAM UNet+++在#2处出现边界粘连。

通过对上述实验结果的定量及定性分析, 可以得出结论: 定量角度上, CBAM UNet+++方法相较基准方法达到最优精度。同时, 通过自对比方法 UNet+++, 说明在编码器中使用 ResBlock_CBAM 有助于提升网络精度; 定性角度上, CBAM UNet+++对细部变化的复刻效果优于基准方法, 说明主干网络 UNet+++能够从全尺度角度学习多层语义, 从而最大程度减少信息丢失。

4.3 消融实验分析

为进一步验证 CBAM UNet+++设计的合理性, 本文在网络不同位置集成 ResBlock_CBAM, 设计

系列消融实验。如图8所示, 图8 (a) XR1 中 ResBlock_CBAM 置于编码器特征池化操作后, 全尺度特征重构前; 图8 (b) XR2 中 ResBlock_CBAM 置于全尺度特征后, 最终变化结果前; 图8 (c) XR3 中 ResBlock_CBAM 置于解码器上采样后。各模型精度如表5所示, ResBlock_CBAM 在网络中不同位置会对模型精度产生一定影响, XR1 较本文 CBAM UNet+++方法 F1 下降 2.1%, XR3 方法 F1 下降 2.3%, XR2 方法 F1 最低, 下降 8.3%, 这也说明将 ResBlock_CBAM 用作编码器端特征提取能获取更密集有效的特征, 帮助实现 CD 任务。

4.4 LEVIR-CD 结果分析

本文还测试了各模型在建筑物变化数据集 LEVIR-CD 上的性能, 以证明本文方法的有效性和泛化性。不同于 LEBEDEV 数据集, LEVIR-CD 关注的是单一建筑类的生长和衰败。网络超参和实验环境同 4.1 节, 在此不再赘述。

由表6可知, CBAM UNet+++在该数据集上精度最高, 精度、召回率、F1、OA 分别是 86.3%、

87.1%、86.7%、96.8%，相较UNet、UNet++ MSOF、Attention UNet、Attention UNet++、FC Siam diff、ResNet50_ppm、UNet+++方法，F1分别提高9.9%、

2%、0.7%、1.6%、2%、7.8%、1.4%；OA分别提高2.7%、0.4%、0.2%、0.4%、0.4%、1.7%、0.3%。

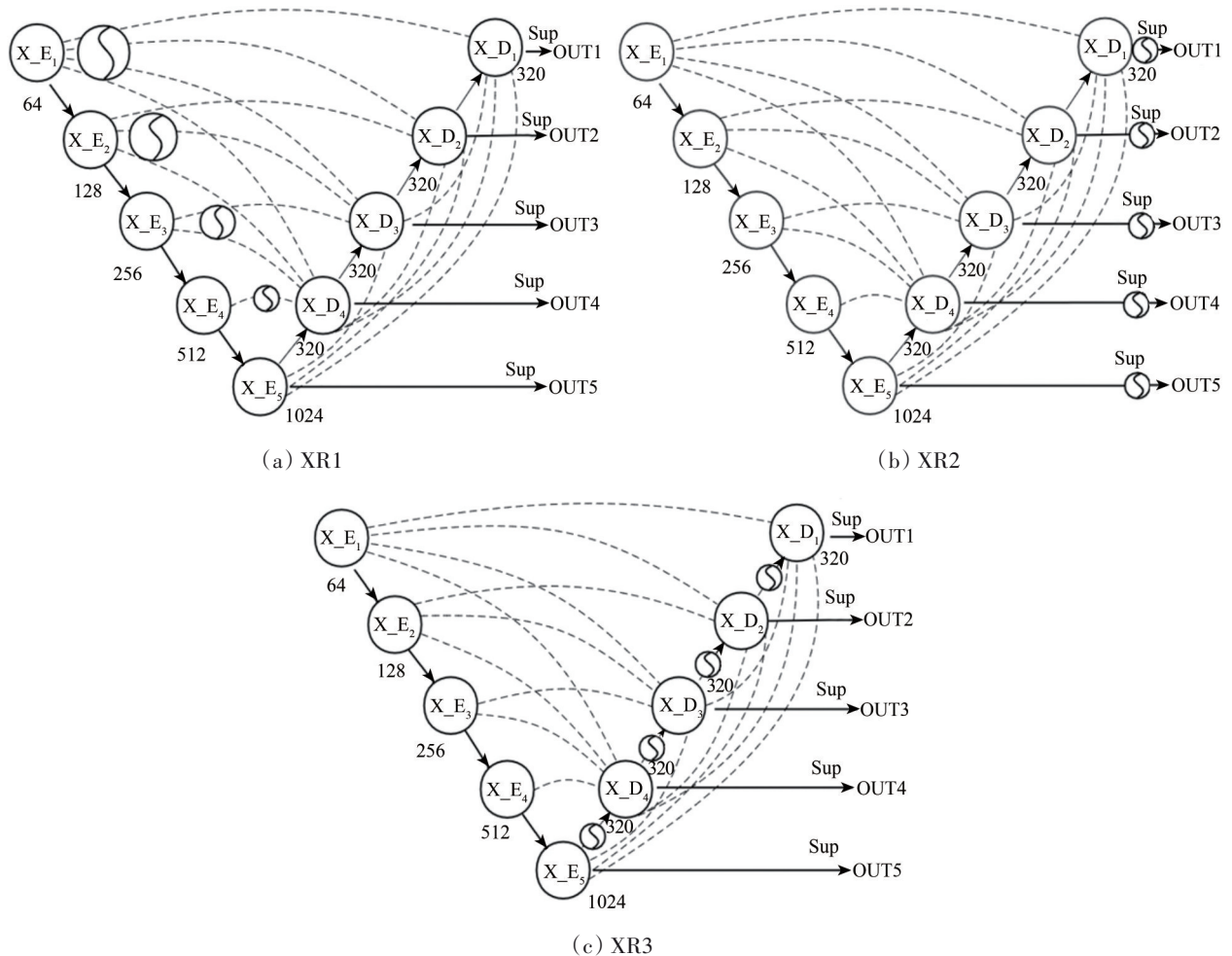


图8 不同CBAM UNet+++构造方式

Fig. 8 Different CBAM UNet+++ construction methods

表5 各模型检测精度评价

Table 5 Evaluation of accuracy of change detection by different methods

名称	精度	召回率	总体精度	F1-Score
XR1	86.3	87.3	96.7	86.8
XR2	75.8	86.1	94.9	80.6
XR3	87.3	85.9	96.7	86.6
CBAM UNet+++	88.9	88.9	97.3	88.9

LEVIR-CD数据集上的检测结果如图9所示，对于稀疏建筑物变化（图9（a）），所有变化建筑都被识别，但UNet++MSOF、Attention UNet、Attention UNet++、ResNet50_ppm方法在#1处存在

粘连，FC Siam diff方法出现误检（#2）；对于密集建筑物变化（图9（b）），UNet、UNet++ MSOF、ResNet50_ppm和UNet+++方法存在漏检（#1），UNet++ MSOF、Attention UNet++和FC Siam diff则出现“过检”（#2），上述方法对密集建筑物的敏感程度均有所下降，相比较下CBAM UNet+++对建筑物边界响应程度更高；当检测到大尺度建筑物时（图9（c）），CBAM UNet+++成功返回边界清晰、内部紧凑性高的变化区域，UNet检测边界模糊（#1），ResNet50_ppm则是出现边界丢失现象，UNet++MSOF、FC Siam diff检测结果出现空洞，Attention UNet、Attention UNet++、UNet+++则出现“过检”现象。

表6 不同方法变化检测精度评价

Table 6 Evaluation of accuracy of change detection by different methods

类别	方法	精度	召回率	F1-Score	OA
不添加 注意力机制	UNet	71.8	82.4	76.8	94.1
	FC Siam diff	85.1	84.4	84.7	96.4
	UNet++ MSOF	85.9	83.5	84.7	96.4
	UNet+++	84.6	86.0	85.3	96.5
添加 注意力机制	Attention UNet	85.4	86.5	86.0	96.6
	Attention UNet++	84.1	86.2	85.1	96.4
	ResNet50_ppm	81.3	76.6	78.9	95.1
	CBAM UNet+++	86.3	87.1	86.7	96.8

注：粗体为精度最优。

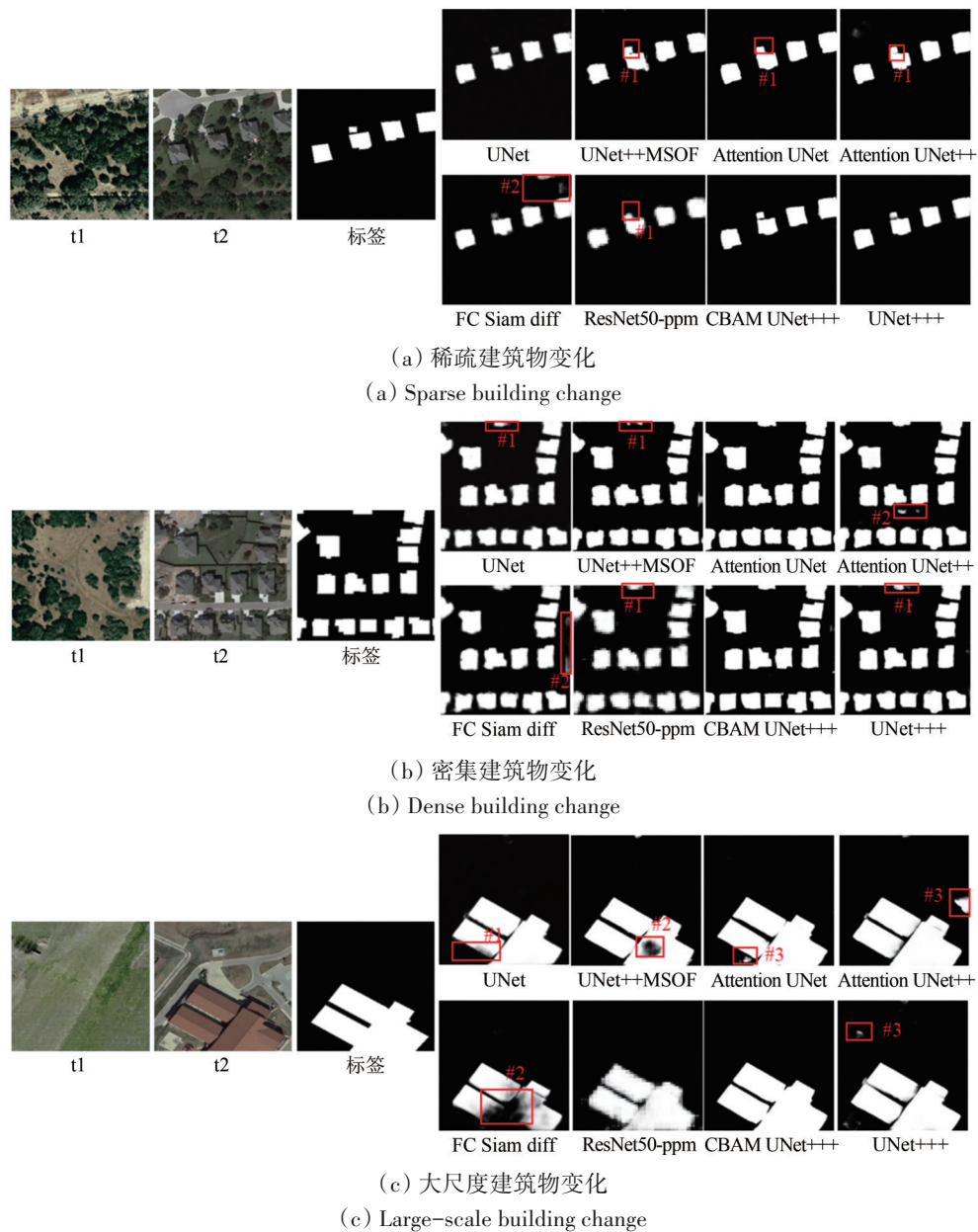


图9 变化检测结果可视化比较

Fig. 9 Visual comparison of change detection results

综上, 通过在单一建筑物数据集上的实验结果, 验证了CBAM UNet+++对CD任务的有效性和适用性, 定量角度上, CBAM UNet+++在LEVIR-CD数据集上的精度略低于LEBEDEV数据集; 定性角度上, CBAM UNet+++方法在两个数据集上均返回边界清晰、位置准确、吻合度高的变化图, 优于基准方法。

各网络参数量及其在LEVIR-CD验证集上的速率如表7所示, CBAM UNet+++模型参数低于UNet、Attention UNet、Attention UNet++, 同时, CBAM UNet+++在同类算法中取得了较快的运行效率 ($9.36 \text{ f}\cdot\text{s}^{-1}$), 较好的保证了算法的实时性。说明CBAM UNet+++方法在可应用性与实时性起到了更好的平衡作用, 相较于基准方法有更好的应用前景。

表7 不同方法参数量及运行时间
Table 7 Parameters and running time of different methods

方法	运行效率/($\text{f}\cdot\text{s}^{-1}$)	参数量/Mb
UNet	6.62	355
UNet++ MSOF	6.63	104
Attention UNet	5.49	365
Attention UNet++	4.16	427
ResNet50_PPM	12.33	140
FC Siam diff	8.78	62.3
CBAM UNet+++	9.36	313

5 结 论

针对现有CD网络存在的编码阶段特征提取能力不足、抗噪能力差以及解码阶段特征利用率低等问题, 提出由注意力机制引导的UNet+++高分辨遥感影像变化检测方法CBAM UNet+++ , 通过ResBlock_CBAM增强编码器对语义信息及位置信息的聚合能力, 结合分割网络UNet+++独有的全尺度信息重构优势, 成功应用于高分辨率遥感影像CD任务中; 利用Dice loss和平衡二元交叉熵组成混合损失函数对网络进行深度监督并在两个不同类型CD数据集上进行实验。结果表明, CBAM UNet+++方法能够准确定位并检测变化区域, 检测效果及精度优于基准方法, 在两个数据集中的精度表现存在轻微差异, 但体现在结果图上并无二致, CBAM UNet+++在建筑物数据集上受部分伪变

化信息干扰。未来工作将考虑该网络在异构双时相图像中变化检测的可用性, 并进一步解决早期融合给变化检测精度带来的影响。

参考文献(References)

- 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]
- Daudt R C, Saux B L and Boulch A. 2018. Fully convolutional siamese networks for change detection//2018 25th IEEE International Conference on Image Processing (ICIP). Athens: IEEE: 4063-4067 [DOI: 10.1109/ICIP.2018.8451652]
- Han X, Han L, Li L Z and Li H H. 2022. Building change detection in high-resolution remote-sensing images based on deep learning. *Laser and Optoelectronics Progress*, 59(10): 1001003 (韩星, 韩玲, 李良志, 李慧慧. 2022. 基于深度学习的高分辨率遥感图像建筑物变化检测. *激光与光电子学进展*, 59(10): 1001003) [DOI: 10.3788/LOP202259.1001003]
- Huang H M, Lin L F, Tong R F, Hu H J, Zhang Q W, Iwamoto Y, Han X H, Chen Y W and Wu J. 2020. Unet 3+: a full-scale connected unet for medical image segmentation//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE: 1055-1059 [DOI: 10.1109/ICASSP40776.2020.9053405]
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe: NIPS: 1106-1114
- Li C, Tan Y S, Chen W, Luo X, Gao Y M, Jia X G and Wang Z Y. 2020. Attention Unet++: a nested attention-aware U-Net for liver CT image segmentation//2020 IEEE International Conference on Image Processing (ICIP). Abu Dhabi: IEEE: 345-349 [DOI: 10.1109/ICIP40778.2020.9190761]
- Liu T F. 2020. Research on Land Cover Change Detection Method Based on Spatial-Spectral Feature Combination Using High Resolution Remote Sensing Image. Xi'an: Xi'an University of Technology (刘统飞. 2020. 基于高分辨率遥感影像空-谱特征联合的地表覆盖变化检测方法研究. 西安: 西安理工大学) [DOI: 10.27398/d.cnki.gxalu.2020.001280]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Oktay O, Schlemper J, Folgoc L L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N Y, Kainz B, Glocker B and Rueckert D. 2018. Attention U-Net: learning where to look for the pancreas. *arXiv: 1804.03999* [DOI: 10.48550/arXiv.1804.03999]
- Peng D F, Zhang Y J and Guan H Y. 2019. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sensing*, 11(11): 1382 [DOI: 10.3390/rs11111382]

- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Ruan C, Guo H and An J B. 2021. SAR inshore ship detection algorithm in complex background. *Journal of Image and Graphics*, 26(5): 1058-1066 (阮晨, 郭浩, 安居白. 2021. 复杂背景下 SAR 近岸舰船检测. *中国图象图形学报*, 26(5): 1058-1066) [DOI: 10.11834/jig.200266]
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition//3rd International Conference on Learning Representations. San Diego: ICLR [DOI: 10.48550/arXiv.1409.1556]
- Stramondo S, Bignami C, Chini M, Pierdicca N and Tertuliani A. 2006. Satellite radar and optical remote sensing for earthquake damage detection: results from different case studies. *International Journal of Remote Sensing*, 27(20): 4433-4447 [DOI: 10.1080/01431160600675895]
- Tian J, Reinartz P, d' Angelo P and Ehlers M. 2013. Region-based automatic building and forest change detection on Cartosat-1 stereo imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79: 226-239 [DOI: 10.1016/j.isprsjprs.2013.02.017]
- Wang H Y, Zhu Y K, Green B, Adam H, Yuille A and Chen L C. 2020. Axial-DeepLab: stand-alone axial-attention for panoptic segmentation//16th European Conference on Computer Vision. Glasgow: Springer: 108-126 [DOI: 10.1007/978-3-030-58548-8_7]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xiang Y. 2020. Change Detection Based on Encoder Decoder for High-Resolution Remote Sensing Image. Xuzhou: China University of Mining and Technology (向阳. 2020. 基于编码解码的高分辨率遥感图像变化检测. 徐州: 中国矿业大学) [DOI: 10.27623/d.cnki.gzkyu.2020.001720]
- Xu W H, Pei Y J, Gao D L, Zhu J D and Liu Y K. 2021. Mass classification of breast mammogram based on attention mechanism and transfer learning. *Laser and Optoelectronics Progress*, 58(4): 0410007 (许文慧, 裴以建, 邵冬林, 朱久德, 刘云凯. 2021. 基于注意力机制与迁移学习的乳腺钼靶肿块分类. *激光与光电子学进展*, 58(4): 0410007) [DOI: 10.3788/LOP202158.0410007]
- Yuan Z, Guo H T, Lu J, Lu W and Lin Y Z. 2021. High-resolution remote sensing image change detection technology based on UNet++ and attention mechanism. *Journal of Geomatics Science and Technology*, 38(2): 155-159 (袁洲, 郭海涛, 卢俊, 路威, 林雨淮. 2021. 融合 UNet++ 网络和注意力机制的高分辨率遥感影像变化检测算法. *测绘科学技术学报*, 38(2): 155-159) [DOI: 10.3969/j.issn.1673-6338.2021.02.008]
- Zhang C X, Yue P, Tapete D, Jiang L C, Shangguan B, 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]

CBAM UNet+++ : Attention mechanism to guide change detection studies of full-scale connected networks

LIU Ying¹, HE Xue¹, LI Danyang², YUE Hui¹, WEI Jiali¹

1. School of Surveying and Mapping Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China;
2. National Geographic Information System Engineering Technology Research Center, China University of Geosciences, Wuhan 430000, China

Abstract: Existing change detection networks rely heavily on layer-by-layer convolution for feature extraction. However, the use of this method leads to a loss of information, and it lacks the ability to mine important change features. Therefore, knowing how to effectively suppress the influence of the background and identifying ways to increase the ability of the network to learn salient features and generate recognizable feature information are highly important for change detection tasks. Traditional skip connections lack the ability to obtain change information from a full-scale perspective and perform encoder feature extraction. Thus, a UNet+++ high-resolution remote sensing image change detection network called CBAM UNet+++ combined with a coupled attention mechanism (i. e., a convolutional block attention module [CBAM]) was designed in this research.

CBAM UNet+++ is based on the semantic segmentation structure UNet++. The unique full-scale concatenation operation of UNet+++ effectively fuses the semantic and spatial information from the full-scale perspective to avoid information loss. The basic convolutional unit

can be replaced by a residual attention module (Residual Block_CBAM and ResBlock_CBAM) to suppress background effects and enhance the learning ability of the encoder to handle significant features. The residual attention module was validated on two remote sensing image change detection datasets—LEBEDV and LEVIR-CD—involving different high-resolution change regions.

The proposed method has the highest accuracy on the LEBEDEV multifeature change dataset, with F1 and OA values of 88.9% and 97.3%, respectively, and the second highest accuracy on the LEVIR-CD building change dataset, with F1 and OA values of 86.7% and 96.8%, respectively. The proposed method can obtain deep semantics in a targeted manner, and its qualitative results are better than those of other benchmark networks.

The CBAM UNet+++ method can accurately locate and detect change regions with better detection and accuracy than can the benchmark method. The accuracy results of the two selected datasets were slightly different, but they were not inconsistent. The accuracy of the CBAM UNet+++ model was disrupted by pseudochange information in the building dataset. Future work may focus on the usability of this network for change detection in heterogeneous dual-temporal images to further address the impact of early fusion on change detection accuracy.

Key words: remote sensing, change detection, UNet+++, attention mechanism, encoding and decoding

Supported by Natural Science Basic Research Program of Shaanxi (No. 2020JM-514); National Natural Science Foundation of China (No. 41401496); Xi'an University of Science and Technology (No. 2019YQ3-04)