# 基于编解码网络 UNet3十的遥感影像建筑变化检测

梁燕<sup>1),2)</sup> 易春霞<sup>1),2)</sup> 王光宇<sup>1),3)</sup>

<sup>1)</sup>(重庆邮电大学通信与信息工程学院 重庆 400065)
 <sup>2)</sup>(信号与信息处理重庆市重点实验室 重庆 400065)
 <sup>3)</sup>(移动通信教育部工程研究中心 重庆 400065)

遥感影像建筑变化检测需解决两个重要问题:一是双时相影像本身存在的时间依赖性问题;其二是由 摘 粟 于建筑物密集分布,阴影效应及各对象之间的相似性导致的特征鉴别问题,该文分析现有处理方案,提出了基于 UNet3+网络的边缘引导变换检测网络(Edge-Guided Change Detection Base on UNet3+, EGCD-UNet3+). UNet3+利用全尺度的跳跃连接把来自不同尺度特征图中的深层语义与浅层语义直接结合,从多尺度聚合的特征 图中学习层次表示,但是在特征提取时忽略了对象尺度规模,导致感受野与尺度不匹配.因此 EGCD-UNet3+首先 设计了一种具有自适应感受些的选择性核 Block(Selective Kernel Block, SKB)代替 UNet3+原始的 Block, 使影像 对在提取深、浅层特征时具有自适应感受野属性. EGCD-UNet3+由编码与解码两部分构成,在编码端利用长短期 记忆网络(Long Short-Term Memox, LSTM)捕捉长期依赖关系,建模像素之间的关系上下文,设计差分增强模 块(Difference Enhance Module, DEM),分析影像对之间的时间相关性,解决双时相本身存在的时间依赖性问题.在 解码端,EGCD-UNet3+提出边缘引导上下文模块(Edge-Guided Context Module,EGCM)进一步改善建筑检测 边界的性能,在更细粒度水平上有效提取多长安空间边缘信息.最后,EGCD-UNet3+利用同时具备像素分割误差 和边缘分割误差的复合损失函数,使网络能够充分学习有效的特征进行准确的标签预测.所提模型在 LEVIR-CD 与 WHU-CD 数据集上验证,精准率(P)分别达到 90.73%、91.75%,召回率(R)可分别增长到 96.68%、92.42%, F1-score(F1)分别增加到 93.15%、92.08%,总体分割宿确度(OA)分别达到 99.12%、98.96%,且交并比(IoU)分 别增加到 83.96%、74.91%.

关键词 变化检测;差分增强;长短期记忆;选择性核;边缘引导, )
 中图法分类号 TP391 DOI 号 10.11897/SP. J. 1016.2023.01720

## Detection of Building Change in Remote Sensing Image Based on Encoder-Decoder Network UNet3+

LIANG Yan<sup>1),2)</sup> YI Chun-Xia<sup>1),2)</sup> WANG Guang-Yu<sup>1),3)</sup>

<sup>1)</sup> (School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065)
 <sup>2)</sup> (Chongqing Key Laboratory of Signal and Information Processing, Chongqing 400065)

<sup>3)</sup> (Engineering Research Center of Mobile Communications of the Ministry of Education, Chongqing 400065)

**Abstract** There are two important problems in building change detection of remote sensing image: one is the time dependence problem of the bi-temporal image itself, the other is the object identification problem due to the dense distribution of buildings, shadow effect and the similarity between the features. This paper proposes an Edge-Guided Change Detection Base on UNet3+ (EGCD-UNet3+) network after analyzing the existing schemes. The UNet3+ uses the full-scale skip connections to directly combine the deep and shallow semantics from different scale

收稿日期:2022-08-08;在线发布日期:2023-04-07.本课题得到国家自然科学基金(61702066)、重庆市教委科学技术重点研究项目基金(KJZD-M201900601)资助.梁 燕(通信作者),硕士,高级工程师,主要研究方向为计算机视觉、深度学习、自然语言处理.E-mail: liangyan @cqupt.edu.cn.易春霞,硕士研究生,主要研究方向为计算机视觉、深度学习.**王光**字,博士,教授,主要研究领域为新型多址接入技术、多载波调制技术.

feature maps, and learns the hierarchical representation from the multi-scale aggregated feature maps. However, it ignores the object scale in feature extraction, leading to the mismatch between the receptive field and the scale. Therefore, the EGCD-UNet3+ first designs a Selective Kernel Block (SKB) with adaptive receptive field to replace the original UNet3+ block, so that the image of deep and shallow features is extracted with adaptive receptive field properties. The EGCD-UNet3+ consists of the encoding part and the decoding part. In the encoding part, the Long Short Term Memory (LSTM) is used to design a Difference Enhancement Module (DEM) for analyzing the temporal correlation between image pairs, and solving the time dependency problem of the bi-temporal itself, because the LSTM's abilities of capturing long-term dependency, modeling the context relationship. In the decoding part, the EGCD-UNet3+ proposes an Edge Guided Context Module (EGCM) to further improve the performance of building detection boundary, and effectively extract multi-scale spatial edge information at a finer-grained level. Finally, the EGCD-UNet3+ defines the composite loss function with pixel segmentation error and edge segmentation error, so that the network can fully learn effective features for accurate label prediction. The proposed model has been experimentally verified on the LEVIR-CD and WHU-CD dataset. The results show that the precision (P) reached 90.75% and 91.75% respectively, the recall (R) increased to 96.68% and 92.42% respectively, the F1-score (F1) increased to 93.15% and 92.08% respectively, the overall segmentation accuracy (OA) reached 99.12% and 98. 96% respectively, and the intersection over union (IoU) increased to 83. 96% and 74.91% respectively.

Keywords change detection; difference enhance, long short-term memory; selective kernel; edge-guided context

## 1 引 言

近年来,随着大量搭载高分辨率影像获取设备 的卫星广泛投入军事和民用领域,针对海量的高分 辨率或极高分辨率的卫星遥感图像分析成为研究热 点.在军事领域,遥感图像主要用于敌方目标定位和 识别,以提升军事打击精准性.在民用领域,遥感图 像主要在土地资源规划、城市化监测、建设用地监管 和自然灾害监测与评估<sup>[1]</sup>等方面发挥作用.遥感图 像表现出高空间分辨率、高光谱分辨率、高时间分辨 率的特点,但地球系统具有开放性、复杂性和不确定 性的特点,通过对不同时相采集的遥感图像的变化 检测(Change Detection, CD),例如建筑物变化检 测,侧重于从遥感图像中识别已更改的建筑,是该领 域重要的研究内容之一.

变化检测(CD)旨在从同一地理区域不同时间 采集的多时相遥感图像中识别和定位感兴趣对象 的变化,可以观察和分析地理空间对象的详细结构 和变换信息.传统的变化检测方法可分为三类:基于 图像算法的方法、基于图像变换的方法和基于后分 类的方法.图像算术方法包括图像差分法[2]、图像比 率法<sup>[3]</sup>、变化向量分析法(Change Vector Analysis, CVA)<sup>[4]</sup>等.这些方法通过减法或除法获得特征图, 然后确定分割阈值来生成变化图,但是这类方法忽 略了上下文信息,容易造成大量噪声干扰.基于图像 变换的方法包括主成分分析(Principal Component Analysis, PCA)<sup>[5]</sup>、多元变化检测(Multivariate Alteration Detection, MAD)<sup>[6]</sup>、迭代加权多元变化检测 (Iterative Reweighted Multivariate Alteration Detection, IR-MAD)<sup>[7]</sup>、核慢特征分析(Kernel Slow Feature Analysis,KSFA)<sup>[8]</sup>等.它们将图像转换为特定的特 征空间,突出变化区域,抑制不变区域.然而,要针对 不同地区选择最合适的方法是一个困难的过程.基 于分类后变化检测方法首先对图像中的地物进行分 类,然后对这些图像进行对比分析,生成变化图.变 化检测精度取决于分类精度,缺点是分类过程需要 手动标记样本.由于遥感图像包含丰富光谱信息及 纹理信息且地物结构多样性,存在多层次、多尺度、 多类别特征提取问题,传统变化方法不能取得很好

的检测效果.

随着深度学习理论和技术的蓬勃发展,深度卷 积神经网络因其在特征提取方面的独特优势,在 变化检测领域也得到了广泛应用.现有研究主要 从是否使用孪生网络架构作为编码器而分为单流 方法和双流方法.图1(a)所示为单流方法<sup>[9]</sup>:单流 方法将双时态图像波段叠加(Layer stacking)为通 道维度上的一个输入,再送入差分提取网络以生成变 化图.例如,在文献[10-11]中,分别提出 CD-Net 与 Unet++\_MSOF 网络用于遥感图像变化检测.基 于压缩块和扩展块的叠加思想形成单流网络.文献 [12]提出了将 Unet++<sup>[13]</sup>和微分金字塔相结合的 DifUnet++,它通过数据依赖上采样方法,而不是 传统的双线性插值.然而,单流方法依然存在共性问 题,特别是融合的双时态图像对被输入到 CD 网络, 使得单个原始图像的特征不可用,难以从差异特征 中提取单个图像信息.

双流方法<sup>[9]</sup>如图 1(b)所示:双流方法即孪生网 络体系结构.将双时态图像分别送到两个具有相同 权重的流编码器架构中.解码时,差异提取网络用于 捕获差异特征以生成变化图<sup>[14-17]</sup>.例如在文献[14] 首次提出两种具有跳跃连接的全卷积孪生网络体 系结构 FC Siam conc 和 FC Siam diff.文献[16]提 出完全卷积孪生度量网络,通过定制隐式度量来检 测变化.文献[17]提出的深度监督图像融合网络 (Image Fusion Network, IFN),利用各种注意机制 来融合异质特征.文献[15]在文献[18]基础上将孪 生网络与 Unet++相结合提出了紧密连接孪生网 络 SNUNet.双流方法可分别保留原始双时态图像 中具有位置信息的特征,然而与单流方法相比,双流 方法通常具有更多网络层,这将在一定程度上带来 梯度消失的可能性,因此网络的训练更加困难.



变化检测中的单流和双流方法均需完成池化操 作,必不可少的多速率采样会对变化图有直接影响. 例如,在每次下采样过程后,部分详细位置信息将丢 失;在每次上采样过程中,难以生成分辨率更高的特 征图.从视觉角度来看,这会导致预测变化图的性能 变差,尤其是在小区域变化和对象边缘变化时.综上 所述,遥感变化检测中主要存在多层次、多尺度特征 提取问题,单一策略效果不足,需要综合运用多种自 适应处理方法才可能有效解决.本文从提取各个层 次精细特征、丰富多尺度上下文及边界信息改善等 关键点出发,提出了一种综合运用多项技术的基于 UNet3+<sup>[19]</sup>的边缘引导变换检测网络(Edge-Guided Change Detection Base on UNet3+, EGCD-UNet3+). UNet3+由 UNet<sup>[20]</sup>和 UNet++ 设计而得的编码 器-解码器体系结构,该结构作为语义特征和差分特 征提取网络,利用了全尺度跳跃连接方式,较好的结 合了不同尺度特征映射的低层细节和高层语义. UNet3+同时也具有轻量化结构,可以利用较少的 网络参数,生成具有更精细位置信息的特征,这是恢 复变化映射和实现更高性能的关键.本文所提 EGCD-UN03个网络主要分为编码与解码两个阶段,其具有以下特征:

(1)编码阶段.针双时相遥感影像对之间的 时间相关性,在差分信息提取时设计了差分增强模 块(Difference Enhance Module,DEM),其引入了长 短时记忆(Long Short-Term memory,LSTM)网 络<sup>[21]</sup>技术;针对遥感影像的多层次、多尺度及类别 对象的自适应上下文信息,在对影像进行深、浅层特 征提取时,设计了一种选择性核 Block(Selective Kernel Block,SKB),使特征提取时具有自适应感受 野属性,更好的进行上下文建模.

(2)解码阶段.针对特征图空间位置、边缘部分 细节信息丢失的问题,提出边缘引导上下文模块 (Edge-Guided Context Module,EGCM)进一步改 善建筑检测边界的性能,提升分割精度及分割边界.

(3)利用同时具备像素分割误差和边缘分割误差的复合损失函数,使网络能够充分学习有效的特征进行准确的标签预测.

## 2 相关工作

变化检测是遥感领域的一项基本任务,检测方 法通常包括特征提取和变化区域识别.前者的目标 是提取有意义的特征,如颜色分布、纹理特征和上下 文信息.后者的目标是使用技术算法分析先前提取 的特征,以识别多时相遥感图像中的变化区域.深度 神经网络在特征学习方面极具优势,在遥感变化检 测领域已有广泛应用.变换检测存在的本质问题有 两类,一类是双时相遥感图受光照、季节变化等自然 因素的影响,多时相图像往往表现出较大的空间-光 谱差异.因此,线性差分分析方法很难实现准确的变 化检测结果.另一类是由于遥感地物场景复杂,在提 取特征时深层特征抽象程度高,浅层空间信息保持 完整,由于建筑物密集分布,阴影效应及各对象之间 的相似性导致在鉴别时受到影响.

如表1所示,为解决变换检测存在的本质性问题,基于 UNet 系列的网络架构大量的被用于遥感 变化检测方向.文献[22]通过结合空洞卷积扩展感 受域,提出了基于对象的检测网络来对土地覆盖 进行分类.文献[23]提出时间建模的深度框架,该框 架将完全卷积网络和 LSTM 相结合.为了融合不同 层次的变化图,文献[11]提出具有基于嵌套 U 网络 的多侧面输出融合策略,文献[24]将金字塔池引入 到卷积网络中,以克服复杂特征和空间不确定性.然 而,UNet 采用普通跳跃连接,UNet++采用嵌套和 稠密跳跃连接.它们并没有直接从多尺度信息中 提取足够多的信息,从本质上看基本都是短连接, 对解码特征进行了再次处理.由于各个连接的融 合,多尺度信息的原始特征几乎没有得到特别好 的利用.

表 1 变化检测方案分析

变化检测		已存在解决	e方案优势及问题	本文解决方案				
本质问题	方案	优势和问题	已不承决方案优势和存在问题	方案	解决方法			
(1) 次 影取之性 影取之性 的需要像 之 的 的 的 等 分 析 影 时 分 后 影 密 府 が 时 的 宗 分 析 影 间 的 需 图 像 炎 分 析 影 时 的 的 寄 分 析 影 时 的 前 宗 一 約 時 前 約 新 約 時 前 的 前 宗 一 約 時 前 的 前 宗 一 第 四 素 の 新 約 時 前 的 前 寄 一 第 四 書 (2) 章 分 析 影 句 所 前 問 寄 四 新 一 約 時 同 的 問 宗 四 結 四 第 一 的 時 同 的 寄 四 第 四 第 四 第 四 第 四 第 四 第 四 第 四 第 四 第 四	UNet 系列	优势:U-Net 系列独 特的跳跃连接和轻 量级网络架构在分割 任务中具有优异的性 能 <sup>[11,14,22:24]</sup> . 问题:UNet 采用普 通跳过连、UNet ++ 采用了嵌套和稠密跳 过连接结构,但并没 有直接从多尺度信息 中提取足够多的信 息,多尺度信息的原 始特征几乎没有得到 特别好的利用.	方案①, 计意力机制 优势:自动加权特征图以增强变化 特征 <sup>[25-26]</sup> ,增强由在于提取. 问题:注意力机制强过捕捉全局或 局部信息增强上下文信息,导致了 很大的计算复杂度和显存容量 方案②:孪生网络 优势:可分别保留原始双时态图像 中具有精细位置信息的特征, 在 孪生网络中引人LSTM或RNN,以 探索时空关系 <sup>[27-28]</sup> . 问题:孪生网络通常具有更多的网 络层,将在一定程度上带来梯度消 失的可能性,网络的训练困难.	方案:UNet3+ 优势:UNet3+利用全 尺度的跳跃连接把来 自不同尺度特征图中的 深层语义与浅层语义 直接结合,从多尺度聚 合的特征图中学习层次 表示. 问题:UNet3+可从全 尺度捕获细粒度细节和 粗松的语义.但是在特 他说取时忽略了对象尺 以其做,导致感受野与 尺度不匹配.	<ol> <li>①问题:UNet3+感受野与尺度不匹配问题. 方法:基于选择性核Block的UNet3+网络.</li> <li>②问题:双时相遥感图像本身时间依赖性问题(本质问题中的(1)). 方法:基于LSTM的差分增强模块.</li> <li>③问题:变化检测方法依赖于无效的鉴别特征问题(本质问题中的(2)). 方法:边缘引导上下文模块.</li> </ol>			

因此众多研究者主要从两个方向来解决上述存 在的问题.一是注意力机制,其对于区分物体和背景 的重要特征非常有用.通过注意机制,网络可以自动 加权特征图以增强变化特征,这大大提高了网络的 效率和准确性.目前,一些学者已将注意力机制应用 于变化检测<sup>[25-26]</sup>.文献[25]通过将空间和通道注意 力引入网络,提高伪变化信息的鲁棒性.文献[26]在 信息传输模块之后引入了注意力机制,以增强变化信 息.然而,注意力机制通过捕捉全局或局部信息增强 上下文信息,会导致很大的计算复杂度和显存容量, 故本文在设计 EGCD-UNet3+时舍弃了此方案.

其二是利用孪生网络.变化检测旨在区分多时 相图像中的变化像素和不变像素.从这一方面来看, 孪生网络是变化检测任务的良好解决方案.文献[14] 将图 1 中(a)所示的早期融合方法与孪生网络进行 了比较.比较结果表明了孪生网络可分别保留原始 双时态图像中具有位置信息的特征,以此证明了孪 生网络的有效性.为了进一步提高检测精度,一些方 法在此基础上引入 LSTM 网络或递归神经网络 (Recurrent Neural Networks, RNN),以探索时空 关系<sup>[27-28]</sup>.但是,孪生网络在一定程度上带来梯度消 失的可能性.故在设计 EGCD-UNet3+网络时需关 注梯度问题.

本文所设计 EGCD-UNet3+以 UNet3+为基础 框架.UNet3+利用了全尺度的跳跃连接(Full-scale skip connection).全尺度的跳跃连接把来自不同尺 度特征图中的深层语义与浅层语义直接结合,从多 尺度聚合的特征图中学习层次表示.由于 UNet3+ 特征图中的深层语义与浅层语义直接结合,从多尺 度聚合的特征图中学习层次表示.且在 UNet3+上 引入孪生网络来实现这一特征提取任务,孪生网络 将有效地提取深层抽象特征,这可以很好地描述每 个多时相图像,且UNet3+可以从全尺度捕获细粒 度的细节和粗粒度的语义.但是在特征提取时忽略 了对象尺度规模,且在进行下采样时会丢失高分辨 率位置信息,难以将微小变化建筑检测出来,导致感 受野与尺度不匹配,因此,提出了一种选择性核 Block(Selective Kernel Block,SKB)使影像对在提 取深、浅层特征时具有自适应感受野属性.

然而,UNet3+在应用于变化检测问题时,还面 临许多挑战:例如高分辨率遥感图像中的比例变化、 类别不平衡和伪变化等;不同地区或国家的建筑 规模和外观不同导致的多尺度问题.因此,在方案设 计时,我们先利用长短期记忆网络LSTM建模像素 之间的关系上下文,设计差分增强模块(Difference Enhance Module,DEM),分析影像对之间的时间相 关性,解决双时相本身存在的时间依赖关系问题.其 次,网络深层高级特征抽象程度高,网络浅层空间信 息保持完整,在变化检测任务中两者同等重要,由于 建筑物密集分布,阴影效应及各对象之间的执识性 导致在鉴别特征时受到影响,于是,提出边缘引导上 下文模块(Edge-Guided Context Module, EGCM) 进一步改善建筑检测边界性能,在更细粒度水平上 有效提取多尺度空间边缘信息.

## 3 基于 UNet3+的边缘引导变化 检测网络 EGCD-UNet3+

图 2 所示为本文所提出的基于 UNet3+的边 缘引导变化检测网络 EGCD-UNet3+架构.该结构 以 UNet3+网络为主干,主要包含 2 个部分:实线 框为编码阶段,虚线框为解码阶段.编码阶段,白色 粗箭头表示进入差分增强模块 DEM 中,该模块利 用 LSTM 在双时相遥感影像差分信息提取时分析 影像对之间的时间相关性,解决双时相本身存在的 时间依赖关系问题,从而建模像素之间的关系上下 文.黑色粗箭头为选择性核 SKB 模块,其模块能增 强多尺度建筑结构特征表征能力,每个对象与其他 对象具有不同的关系,可以利用这些关系进行更好 的上下文建模,平衡多尺度感受野.解码阶段灰色框 为边缘引导上下文模块 EGCM.在变化检测网络中 空间细节信息包括多尺度(大、小和一般尺度的建



图 2 基于 UNet3+的边缘引导变化检测网络 EGCD-UNet3+结构

筑)、多形状(包方形、矩形、圆形和其他不规则形状)、边缘细节信息的建筑,因此需要 EGCM 进一步改善建筑检测边界的性能,在更细粒度水平上有效提取多尺度空间边缘信息,建立多尺度长期通道及空间之间的依赖性.图2中双向箭头代表双分支结构具有完全相同的网络结构和共享权重(Weight Sharing),虚线箭头代表全尺度输入或输出跳跃连接(Full-scale inter /intra skip connection),图中数字表示在进行特征提取时的图像维度.最后,为了使边缘引导上下文模块的提出更具有实际意义,利用了同时具备像素分割误差和边缘分割误差的复合损失函数,使网络能够充分学习有效的特征进行准确的标签预测.

## 4 编码阶段

## 4.1 基于 LSTM 的差分增强模块

主干网提取深层特征和浅层特征. 深层特征表示语义信息,有助于定位变化区域. 浅层特征点含了详细的信息,可以更好地重建变化后建筑的、间结构. 为了进一步从多层次特征中挖掘更精确的差分信息,所设计的差分增强模块(DEM)如图 3 所示其模块对应着图 2 中的白色箭头. DEM 实质上是一种自适应加权机制的差分信息增强模块,根据差分信息自适应调整权重. 为了从多级特征中挖掘出更准确的差异信息,通过从"粗"到"精"的方式分析特征对之间的相关性. 在"粗"阶段,LSTM 用于初步测量特征 *F*<sub>11</sub>和*F*<sub>2i</sub>之间的相似性,从而产生初步的判别特征 *L*<sub>i</sub>. 在"精"阶段,设计自适应加权机制以增强差异信息,其中根据 *L*<sub>i</sub>自适应地调整权重. 如果初始特征对之间的距离较大,意味着样本对属于

变化像素的概率越高则会分配更高的权重,反之亦 然.这样,DEM可以有效地增强差分信息,自适应调 整权重能缓解梯度问题.具体地,多级判别特征 D<sub>i</sub> 可以表述如下:D<sub>i</sub>=|F<sub>1i</sub>-F<sub>2i</sub>|⊗L<sub>i</sub>.在差分增强模 块(DEM)中设计 LSTM 主要基于两点原因:(1)多 时相图像属于一种序列数据,LSTM 擅长建模序列 数据之间的时间相关性;(2)双时相遥感图受光照条 件、季节变化等自然因素的影响,往往表现出较大的 空间-光谱差异.在这种情况下,线性差分分析方法很 难实现准确的变化检测结果.例如,不同季节植被颜 色的伪变化,容易导致检测结果不准确.LSTM 是一 种非线性差分分析方法,有助于增强网络的鲁棒性.



图 3 差分增强模块 DEM 结构图

#### 4.2 基于选择性核 Block 的 UNet3+网络

UNet3+(图 4(c))的设计参考了 UNet(图 4 (a))和 UNet++(图 4(b))两个网络结构,主要区 别在于连接方式. UNet 采用普通跳过连接(Plain skip connections),UNet++采用了嵌套和稠密跳 过连接(Nested and dense skip connections).从本 质上看基本都是短连接,多尺度信息的原始特征几 乎没有得到较好的利用,均没有直接从多尺度信息 中提取足够多的有用信息.UNet3+利用了全尺度 的跳跃连接.全尺度的跳跃连接把来自不同尺度特 征图中的深层语义与浅层语义直接结合,从多尺度 聚合的特征图中学习层次表示.高分辨率遥感图像





中的建筑物往往具有不同的大小和形状,各层次的特征不能很好地反映建筑物的变化特征.基于这一考虑,需要提取包含深层语义特征和浅层外观特征的多级特征.由于 UNet3+特定的网络结构,引入孪生网络来实现这一特征提取任务.孪生网络将有效地提取深层抽象特征,可以很好地描述每个多时相图像.

遥感影像特征提取,当单一感受野比较大时有

足够的上下文信息,但对象特征会被不相关对象覆盖.当特征映射有较小感受野时,又缺乏上下文信息.如果使用固定扩张速率或跨距,忽略了对象尺度规模,导致感受野与尺度不匹配.为了进一步从全尺寸的聚合特征图中学习层次表示,本文设计了一种具有自适应感受野的选择性核<sup>[29]</sup>Block(SKB)代替UNet3+原始的 Block(图 5).SKB 结构如图 6 所示,其与图 2 中的黑色粗箭头所对应.



SKB 结构通过拆分(Split)、融合(Fuse)、选择(Select)三个操作符来实现.图中卷积核大小分别为 3×3 和 5×5.

拆分(Split):对于任何给定的特征映射  $X \in \mathbb{R}^{H' \times W' \times C'}$ ,默认情况下分别执行内核大小为 3×3、 5×5 的两个转换  $\tilde{F}: X \rightarrow \tilde{U} \in \mathbb{R}^{H \times W \times C}$ 和  $\hat{F}: X \rightarrow \hat{U} \in \mathbb{R}^{H \times W \times C}$ ,注意  $\tilde{F}$  和  $\hat{F}$  都是有效的分组卷积、BN 归 一化、Relu 激活函数组成的.

融合(Fuse):通过元素融合两分支的结果,整合 两分支的结构信息.然后通过使用全局平局池化(图 中 gp)嵌入全局信息,以生成通道统计信息  $s \in \mathbb{R}^{c}$ ,s 的第 c 个元素是通过空间尺寸  $H \times W$  压缩 U 计算. 最后,通过一个完全连接层(fc)创建了完整的特征  $z \in \mathbb{R}^{d \times 1}$ 以实现精确的自适应选择引导.全连接层把 以前的局部特征重新通过权值矩阵组装成完整的特 征图,通过降低维数来提高效率:

$$\boldsymbol{U} = \boldsymbol{\tilde{U}} + \boldsymbol{\hat{U}} \tag{1}$$

$$\mathbf{s}_{c} = \mathbf{F}_{gp}(\mathbf{U}_{c}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{U}_{c}(i,j) \qquad (2)$$

$$\boldsymbol{z} = \boldsymbol{F}_{fc}(s) = \delta(\boldsymbol{\beta}(\boldsymbol{W}_s)) \tag{3}$$

其中  $\delta \in \text{ReLU}$  函数, $\beta \in \text{BN}$  归一化, $W \in \mathbb{R}^{d \times C}$ ,r 用于控制 d 值,

$$d = \max(C/r, L) \tag{4}$$

其中 L 表示 d 的最小值(实验中设置 L=32).

选择(Select): 在特征 z 的指导下, 自适应选择 不同的信息空间尺度, 进行 Softmax 运算:

$$\boldsymbol{a}_{c} = \frac{e^{A_{c}z}}{e^{A_{c}z} + e^{B_{c}z}}, \ \boldsymbol{b}_{c} = \frac{e^{B_{c}z}}{e^{A_{c}z} + e^{B_{c}z}}$$
(5)

其中A  $B \in \mathbb{R}^{C \times d}$ , a, b 表示  $\tilde{U}$  和 $\hat{U}$  的向量,注意 $A_c \in \mathbb{R}^{1 \times d}$  是A 的第c 个行,  $a_c \in A$  的第c 元素,  $B_c$ 和 $b_c$ 同理.  $a_c + b_c = I$ , 最终的特征映射 V 是通过各种核上的注意权重获得的,其中  $V = [V_1, V_2, \dots, V_C], V_C \in \mathbb{R}^{H \times W}$ .

$$\boldsymbol{V}_{c} = \boldsymbol{a}_{c} \cdot \boldsymbol{\widetilde{U}}_{c} + \boldsymbol{b}_{c} \cdot \boldsymbol{\widehat{U}}_{c}$$
(6)

SKB 自适应加权多尺度特征图,根据各个不同 对象的尺度,与感受野匹配的将被增强,而其他的将 被抑制,确保多尺度语境信息有效捕获.

### 5 解码阶段

#### 5.1 边缘引导上下文模块

由于遥感场景的复杂性,对建筑物的边缘像素 进行准确分割至关重要.一方面,边缘信息不仅可以 反映建筑物的几何轮廓形状,还可以反映建筑物的 地理位置信息.有效地提取边缘信息将有利于建筑 物变化检测,尤其是在边界区域周围.另一方面,更 好的变化检测结果将有助于改进边缘预测结果,因 为这里的边缘信息是从变化估计特征图中提取的, 即变化检测和边缘预测这两项任务密切相关,可以 在检测性能方面相互促进.目前,精确的边缘检测仍 然是一个挑战性问题,受建筑物密集分布、阴影效应 及建筑物与道路之间的相似性等因素影响,常常使 变化检测方法依赖于无效的鉴别特征.

在编码阶段,已设计了多级特征提取和差分信息提取,以更好地提取变化区域和不变区域之间的区别特征.为了进一步提高已变化建筑检测边界的性能,在解码阶段设计了具有上下文感知水平<sup>[30]</sup>的边缘引导上下文模块(EGCM).在解码过程中,因不同阶段特征图维度都是 80,但特征图的大小(宽、高)不同,因此根据是否需要进行 interpolate 上采样将边缘引导上下文模块分为EGCM1 与 EGCM2.

图 7 所示为 EGCM1 结构图,输入特征图为 80

通道,宽与长为128. EGCM1 由两个分支组成:第一 分支通过1×1卷积块、interpolate上采样操作从输 入特征图中估计边缘信息,在分支1中引入了边缘 检测 Canny 算法提取特征图边缘信息;第二分支将 第一分支1×1卷积块的边缘信息与输入特征图相 结合,然后通过1×1卷积块、interpolate估计变化 信息.最后,将边缘信息和变化信息通过Softmax 层 得到变化概率图和边缘概率图.EGCM 不仅可以预 测变化的建筑物及其边缘,还可将边缘特征直接集 成到判别特征中,利用此模块提取到的上下文信息 进行长距离建模,进一步提高预测结果的质量.

EGCM2 结构如图 8 所示,它与 EGCM1 的区 别在于输入特征图的长宽均为 256,因此在进行一 系列操作时不需要进行 interpolate 上采样.





在 EGCD-UNet3+框架中,EGCM 预测解码 阶段前两层的输出,以获得变化概率图和边缘概率 图.由于其他三层的输出具有较低的空间分辨率和 缺乏结构信息,因此很难预测建筑物的边缘.这些层 的输出仅使用 1×1 卷积块来预测变化概率图.因 此,输出是五个变化概率图和两个边缘概率图.

#### 5.2 复合损失函数

为使边缘引导上下文模块更有意义,设计了同 时具备像素分割误差和边缘分割误差的复合损失 函数.像素分割误差主要针对遥感图像建筑变化 区域与不变化区域的变化检测,保证网络能够充 分学习有效的特征进行准确的标签预测.遥感影像 中不变区域比变化区域更大,因而采用 Focal loss<sup>[31]</sup> 函数解决变化检测中正负训练样本不平衡问题. Focal loss 函数可表示为

$$L_{1} = -\frac{1}{N} \sum_{n=1}^{N} \left[ \alpha (1 - \hat{y}_{n})^{\gamma} y_{n} \log \hat{y}_{n} + (1 - \alpha) \hat{y}_{n}^{\gamma} (1 - y_{n}) \log (1 - \hat{y}_{n}) \right]$$
(7)

其中,N是样本数,ŷ<sub>n</sub>是预测概率值,y<sub>n</sub>是真实值.通 过调整权重因子 α 的值和关注参数 γ 的值来减少负 样本的权重,保证网络在训练过程中更加注重正面 样本.为了定义边缘检测损失,使用均方误差(Mean Squared Error,MSE)损失函数,可以表达为

$$L_2 = \frac{1}{N} \sum_{n=1}^{N} (z_n - \hat{z}_n)^2$$
(8)

其中 z<sub>n</sub>是边缘检测中的真实值, ź<sub>n</sub>是边缘检测中的 概率值.最终复合函数的损失:

$$L = \lambda_1 L_1 + \lambda_2 L_2 \tag{9}$$

其中 $λ_1$ 和 $λ_2$ 是预定义的权重值.

## 6 实验与分析

### 6.1 实验环境与对比模型

本文基于 Pytorch 搭建框架,所有实验都基于 NVIDIA-SMI 460.67 GPU(11 G2和 Python3.7 实现.基本学习率初始值设为 0.001、当验证集的 F1 连续 15 个 epoch 没有增加,学习率下降到关新学习 率的 0.1 倍,整个训练过程包括 100 个 epoch 复合 函数的参数  $\alpha$ 和  $\gamma$  分别设置为 0.25 和 2,将复合函 数的权重比设置为  $\lambda_1:\lambda_2=1:10$ ,batch size 设置为 12,采用优化器为 Adam 和学习速率衰减策略来训 练网络,输入图像具有数据增强功能,包括随机灰度 化和高度和对比度的随机调整.为了便于比较,所有 实验都是在相同训练和测试设置下进行的.为证明 本文网络有效性,与以下网络模型进行比较和分析.

(1) FC-EF<sup>[14]</sup>.图像级融合方法,将双时态图像 串联为 FCN<sup>[32]</sup>的单个输入.

(2) FC Siam Di<sup>[14]</sup>. 特征级融合方法,用孪生 FCN 来提取多级特征,并用特征差分融合双时态 信息.

(3) FC Siam Conc<sup>[14]</sup>. 特征级融合方法,用孪生 FCN 提取多级特征,并用特征拼接融合双时态 信息.

(4) STANet<sup>[33]</sup>. 基于度量的孪生 FCN 方法,该 方法集成了时空注意机制,以获得更多的鉴别特征.

(5) DTCDSCN<sup>[34]</sup>. 多尺度特征拼接方法,该方 法将通道、空间注意添加到深孪生 FCN 中,从而获 得更具辨别力的特征.注意,他们还在每个时态的标 签映射的监督下训练了另外两个语义分段解码器.

(6) IFNet<sup>[17]</sup>. 多尺度特征拼接方法,该方法将 信道注意和空间注意应用于解码器各级拼接的双时 态特征.深度监督(即,计算解码器每一级的监督损 失)用于更好地训练中间层.

(7) SNUNet<sup>[15]</sup>. 多尺度特征拼接方法,该方法 将孪生网络和 NestedUNet<sup>[13]</sup>相结合,以提取图像 高级特征.信道注意应用于解码器每一级的特征.深 度监控还可用于增强中间特征的识别能力.

(8) BIT<sup>[35]</sup>. 基于 transformer 变换检测方法. 将双时态图像表示为几个标记,并使用 transformer 编码器在基于标记的时空中建模上下文.再将学习 的上下文丰富标记反馈到像素空间,以便通过解码 器细化原始特征.

(9) EGRCNN<sup>[30]</sup>. 是边缘引导递归卷积神经网络,思想是将鉴别信息和边缘结构先验信息结合到一个框架中,以改进变化检测结果,尤其是生成更精确的建筑边界.

#### 6.2 数据集

为了证实 EGCD-UNet3+的有效性,我们选择 文献[14-15,17,30,33-35]中所用的公共数据集 LEVIR-CD 和 WHU-CD 进行验证比较.

LEVIR-CD 是一个公共的大型建筑变化检测数据集. 它包含 637 对大小为 1024×1024 的空间分辨率为 5 m 的高分辨率遥感图像. 我们遵循其默认的数据集分割:70%的样本用于训练,10%用于验证,20%用于测试. 由于 GPU 的内存限制,将每个样本载势为 16 个 256×256 大小的小块,没有重叠. 因此获得了 7120、1024、2048 的 256×256 的图像, 分别用于训练,验证、测试.

WHUCO是一个公共建筑变化检测数据集. 它包含一对大小为 32507×15354 的空间分辨率为 0.075m 的航拍图像 按照默认数据集分割将图像 剪裁成 256×256 大小的小块,没有重叠,并将其随 机分为训练集 5205、验证集 742 和测试集 1486,相 应的比例为 7:1:2.

#### 6.3 评价指标

为了评价 EGCD-UNet3 + 在 LEVIR-CD 和 WHU-CD 数据集的精细分割变化性能,在实验中 使用了 5 个常用评估指标: precision(P)、recall(R)、 F1 score(F1), overall accuracy(OA)和 intersection over union(IoU),这些评估指标为通用指标,表述 如下:

$$P = \frac{TP}{TP + FP}, \ R = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = \frac{2PR}{P+R} \tag{11}$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$IoU = \frac{TP}{TP + FN + FP} \tag{13}$$

其中,TP 是真阳性数、FP 是假阳性数、TN 是真阴性数、FN 是假阴性数.

#### 6.4 实验结果与分析

表 2 和表 3 分别展示了在 LEVIR-CD 和 WHU-CD数据集上的对比实验结果.表中灰色标注表示在 所有对比网络模型中各项指标最好的数据, ↑ 与 ↓ 分别表示所提出网络的验证数据其相比是增长还是 下降.

如表 2 所示, EGCD-UNet3+在 LEVIR-CD 数 据集上 P 和 R 分别可以达到 90.75%、96.68%, F1 为 93.15%, OA 可以获得 99.12%的改善, IoU 为 83.96%.如表 3 所示, EGCD-UNet3+在 WHU-CD 数据集中, P 和 R 分别增长到 91.75%、92.42%, 且 F1 达到 92.08%, OA 和 IoU 分别改善到 98.96% 和 74.91%.综合分析表 2 与表 3 的结果,本文提出 的运用多项技术的 EGCD-UNet3+模型可从不同 角度自适应挖掘遥感特征图上下文信息、能充分有 效的进行准确的标签预测, 对多尺度的对象都有较 好检测效果.

表 2 在 LEVIR-CD 数据集上的变化检测结果比较

方法	P/%	R/%	F1/%	OA/%	IoU/%
FC-EF <sup>[14]</sup>	86.91	80.17	83.40	98.39	71.53
FC-Siam-Di <sup>[14]</sup>	89.53	83.31	86.31	98.67	75.92
FC-Siam-Conc <sup>[14]</sup>	91.99	76.77	83.69	98.49	71.96
STANet <sup>[33]</sup>	83.81	91.00	87.26	98.66	77.40
DTCDSCN <sup>[34]</sup>	88.53	86.83	87.67	98.77	78.05
$IFNet^{[17]}$	94.02	82.93	88.13	98.87	78.77
SNUNet <sup>[15]</sup>	89.18	87.17	88.16	98.82	78.83
BIT [35]	89.24	89.37	89.31	98.92	80.68
EGRCNN <sup>[30]</sup>	88.32	91.97	90.11	98.94	80.32
EGCD-UNet3+	<b>90.75</b> ↓	96.68 1	93.15	99.12	83.96 🛉

表 3 在 WHU-CD 数据集上的变化检测结果比较

方法	P/%	$R/rac{0}{10}$	F1/%	OA/%	IoU/%
FC-EF <sup>[14]</sup>	71.63	67.25	69.37	97.61	53.11
FC-Siam-Di <sup>[14]</sup>	47.33	77.66	58.81	95.63	41.66
FC-Siam-Conc <sup>[14]</sup>	60.88	73.58	66.63	97.04	49.95
STANet <sup>[33]</sup>	79.37	85.50	82.32	98.52	69.95
DTCDSCN <sup>[34]</sup>	63.92	82.30	71.95	97.42	56.19
IFNet <sup>[17]</sup>	96.91	73.19	83.40	98.83	71.52
SNUNet <sup>[15]</sup>	85.60	81.49	83.50	98.71	71.67
BIT [35]	86.64	81.48	83.98	98.75	72.39
EGRCNN <sup>[30]</sup>	90.88	89.19	89.33	99.07	73.58
EGCD-UNet3+	91.75↓	92. 42 †	92.08 1	98.96↓	74. 91 ↑

#### 6.5 可视化分析

为了更好地描述 6.4 节的结果,利用表 2 和表 3 实验数据完成可视化分析(图 9).图 9 中有明显改

Image1Image2Ground TruthIFNetSNUNCtBITEGRCNNEGCD-Unet3Image2Image

善的区域用虚线框突出显示.如图 9 所示. IFNet、 SNUNet、BIT 和 ECRCNN 会产生部分建筑物漏检 测,如第 2、3、6 行虚线框区域所示.可以看出,所提 EGCD-UNet3+网络可捕获多级自适应比例特征 图,实现较好建筑物检测图.这可归功 SKB 模块为对 象提取适当上下文信息,自适应增强特征表征能力, 不仅减少了不完整和不规则语义对象,而且更好保留 了几何细节和复杂轮廓.同时,由于引入了 DEM、 EGCM 模块,利用特征间时间相关性关系及语义位 置边界信息获取更准确特征,使建筑物变化检测更 加准确,如图 9 第 1、4、7 行,空间位置信息得以保留. 这表明,该方法对遥感图像变化检测有了显著改善.

#### 6.6 消融实验

在提出的 EGCD-UNet3+网络中,各个模块都 针对性的被用来改善目前变 E检测所存在的问题, 其目的都是为了增强高分辨率遥感图像特征提取和 表征能力.为了评估每个模块的性能及更直观地体 现每个单独模块的贡献,分别按表/和表 5 设置在 数据集 LEVIR-CD 和 WHU-CD 上完成消读实验. 其中×表示未加入对应模块, v表示加入对应模块.

表4结果显示 DEM、EGCM 模块在 EGCD-UNet3+中的作用,可以利用特征间时间相关性关 系及语义位置边界信息获取更准确特征.相比于 基线,在LEVIR-CD数据集上P、R、F1、IoU各指标 分别增长 2%以上, 而 SKB 模块可自适应增强特征 表征能力,也证实了感受野平衡对特征图多尺度空 间细节信息的重要性.在LEVIR-CD上,相比于引 入 DEM 和 EGCM 而言, P、R、F1、OA、IoU 各指标 分别增长到 90.75%、96.68%、93.15%、99.12%、 83.96%.在WHU-CD数据集中,各项评价指标都 得到提升, 而 P、R、F1、OA、IoU 各指标都得到提 升,而P、R、F1、OA、IoU各指标分别达到 91.75%、 92.42%、92.08%、98.96%、74.91%.因此,实验数 据表明,所提出的 EGCD-UNet3+对高分辨率遥感 图像变化检测有一定的成效.表5更直观地比较 DEM、EGCM 和 SKB 每个模块单独的贡献. 通过 表5结果可以明显看出各模块可以进一步改善遥感 图像建筑变化检测的性能,提升分割精度.

表 4 基于编解码网络 UNet3+的边缘引导变化检测网络 EGCD-UNet3+各模块消融实验对比

	模块			数据集									
方法				LEVIR-CD					WHU-CD				
	DEM	EGCM	SKB	P/%	$R/\frac{V_0}{0}$	F1/%	OA/%	IoU/%	P/%	R/%	F1/%	OA/%	IoU/%
EGCD-UNet3+(Ours)	$\times$	$\times$	×	86.77	93.34	89.94	98.28	77.45	86.49	87.31	86.90	97.42	70.55
EGCD-UNet3+(Ours)	$\checkmark$	$\times$	$\times$	88.09	94.45	91.16	98.63	79.62	89.65	89.92	89.81	97.83	71.96
EGCD-UNet3+(Ours)	$\checkmark$	$\checkmark$	$\times$	88.86	95.80	92.20	98, 90	80.32	90.38	91.09	90.73	98.54	72.85
EGCD-UNet3+(Ours)	$\checkmark$	$\checkmark$	$\checkmark$	90.75	96.68	93.15	99.12	83.96	91.75	92.42	92.08	98.96	74.91

表 5 基于编解码网络 UNet3十的边缘引导变化检测网络 EGCD-UNet3十单模块消融实验对比

		齿地		数据集									
方法	快状			LEVIR-CD						WHU-CD			
	DEM	EGCM	SKB	P/%	R/%	F1/%	OA/%	IoU/%	P/%	R/%	F1/%	OA/%	IoU/%
EGCD-UNet3+(Ours)	$\times$	$\times$	$\times$	86.77	93.34	89.94	98.28	77.45	86.49	87.31	86.90	97.42	70.55
EGCD-UNet3+(Ours)	$\checkmark$	$\times$	$\times$	88.09	94.45	91.16	98.63	79.62	89.65	89.92	89.81	97.83	71.96
EGCD-UNet3+(Ours)	$\times$	$\checkmark$	$\times$	88.32	93.94	91.04	98.71	7855	88.93	88.75	88.84	98.01	72.14
EGCD-UNet3+(Ours)	$\times$	×	$\checkmark$	88.96	95.03	92.39	98.58	80.21	90.32	90.58	90.45	98.28	72.69

## 7 结束语

为了提高遥感图像变化检测精度,本文提出了 基于编解码网络 UNet3+的边缘引导变化检测网 络 EGCD-UNet3+.该网络在编、解码阶段,采用了 DEM、SKB、EGCM 等针对性的设计,加强不同特征 图中多层次的深层、浅层语义边界信息的提取和自 适应特征图多尺度上下文感知水平.提出的网络模 型在 LEVIR-CD 和 WHU-CD 数据集上取得了较 好的效果.在 LEVIR-CD 数据集上精准率 P 可达到 90.75%,召回率 R 可增长到 96.68%,F1-score 增加到 93.15%,总体分割精确度 OA 达到 99.12%, 且交并比 IoU 显著增加到 83.96%.在 WHU-CD 数据集上精准率 P 可达到 91.75%,召回率 R 可增 长到 92.42%,F1-score 增加到 92.08%,总体分割 精确度 OA 达到 98.96%,且交并比 IoU 显著增加 到 74.91%.通过可视化分析,所提方案实现了更精 确语义边界改善以及多类别各尺度对象的变换检 测.在未来的研究中,我们将把 EGCD-UNet3+中 的各个模块推广到更多的数据集,并通过调整或提 出新模块来适应多类提取、道路检测和土地覆盖分

#### 类等任务,提高遥感图像检测性能.

#### 参考文献

- [1] Song C, Huang B, Ke L H, et al. Remote sensing of alpine lake water environment changes on the Tibetan Plateau and surroundings: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 92: 26-37
- [2] Singh B. Change detection in the tropical forest environment of northeastern India using Landsat. Remote Sensing and Tropical, 1986, 44: 254-273
- [3] Howarth P J, Wickware G M. Procedures for change detection using Landsat digital data. International Journal of Remote Sensing, 1981, 2(3): 277-291
- [4] Chen J, Chen X H, Cui X H, et a. Change vector analysis in posterior probability space: A new method for land cover change detection. IEEE Generation and Remote Sensing Letters, 2011, 8(2): 317-311
- [5] Huang W, Huang J, Wang L, et al. Remote sensing image change detection based on change sector analysis of PCA component. Remote Sensing for Land & Resources, 2016, 28(1): 22-27
- [6] Allan A, Nielsen, Conardsen K, et al. Multivated treation detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. Remote Sensing of Environment: An Interdisciplinary Journal, 1998, 64(1): 1-19
- [7] Xu G, Li H, Zang Y, et al. Change detection based on IR-MAD model for GF-5 remote sensing imagery//Proceedings of the IOP Conference Series Materials Science and Engineering. Shanghai, China, 2019: 072073-072081
- [8] Chen W, Zhang L P, Du B. Kernel slow feature analysis for scene change detection. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(4): 2367-2384
- [9] Pan F, Wu Z B, Liu Q, et al. DCFF-Net: A densely connected feature fusion network for change detection in high-resolution remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 11974-11985
- [10] Alcantarilla P F, Stent S, Ros G, et al. Street-view change detection with deconvolutional networks. Autonomous Robots, 2018, 42(7): 1301-1322
- Peng D F, Zhang Y J, Guan H Y. End-to-end change detection for high resolution satellite images using improved UNet++. Remote Sensing, 2019, 11(11): 1382-1398
- [12] Zhang X W, Yue Y Z, Gao W X, et al. DifUnet++: A satellite images change detection network based on UNet++ and differential pyramid. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5
- [13] Zhou Z W, Siddiquee M R, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation// Proceedings of the 4th Deep Learning in Medical Image Analysis. Springer, Cham, 2018; 775-789

- [14] Daudt R C, Saux B L, Boulch A. Fully convolutional Siamese networks for change detection//Proceedings of the 2018 25th IEEE International Conference on Image Processing. Athens, Ereece, 2018: 4063-4067
- [15] Fang S, Li K Y, Shao J Y, et al. SNUNet-CD: A densely connected Siamese network for change detection of VHR images. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5
- [16] Guo E, Fu X, Zhu J, et al. Learning to measure change: Fully convolutional Siamese metric networks for scene change detection. IEEE Transactions on Multimedia, 2018, 18: 1-10
- [17] Zhang C X, Yue P, Tapete D, et al. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 166: 183-200
- [18] Huang G, Liu Z, Weinberger K Q, et al. Densely connected convolutional networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017; 2261-2269
- [19] Huang H M, Lin L F, Tong R F, et al. UNet3+: A full-scale connected UNet for medical image segmentation//Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain, 2020: 1055-1059
- [20] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, Cham, 2015: 234-241
- [21] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735-1780
- [22] Chen HAR Wu, Du B, et al. Change detection in multisource VHR images via deep Siamese convolutional multiplelayers recurrent neural network. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(4): 2848-2864
- [23] Zhang C, Wei S, Ji S, et al. Detecting large-scale urban land cover changes from very high resolution remote sensing images using CNN-based classification. International Journal of Geo-Information, 2019, 8(4): 189-205
- [24] Papadomanolaki M, Verma S, Vakalopooulou M, et al. Detecting urban changes with recurrent neural networks from multitemporal sentinel-2 data//Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Yokohama, Japan, 2019: 214-217
- [25] Lei T, Zhang Y X, Lv Z Y, et al. Landslide inventory mapping from bitemporal images using deep convolutional neural networks. IEEE Geoscience and Remote Sensing Letters, 2019, 16(6): 982-986
- [26] Chen J, Yuan Z, Peng J, et al. DASNet: Dual attentive fully convolutional Siamese networks for change detection of high resolution satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 1194-1206

- [27] Liu R C, Cheng Z H, Zhang L L. et al. Remote sensing image change detection based on information transmission and attention mechanism. IEEE Access, 2019, 7: 156349-156359
- [28] Mou L C, Bruzzone L, Zhu X X. Learning spectral-spatialtemporal features via a recurrent convolutional neural network for change detection in multispectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2): 924-935
- [29] Li X, Wang W H, Hu X L, et al. Selective kernel networks //Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, USA, 2020: 510-519
- [30] Bai B F, Fu W, Lu T, et a. Edge-Guided recurrent convolutional neural network for multitemporal remote sensing image building change detection. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60, 3
- [31] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense



**LIANG Yan**, M. S., senior engineer. Her research interests include computer vision, deep learning and natural language processing. object detection//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Italy, 2017: 2999-3007

- [32] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA, 2015; 640-651
- [33] Chen H, Shi Z W. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sensing, 2020, 12(10): 1662-1685
- [34] 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. IEEE Geoscience and Remote Sensing Letters, 2021, 18(5); 811-815
- [35] Chen H, Qi Z P, Shi Z W. Remote sensing image hange detection with transformers. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14

YI Chun-Xia, M. S. candidate. Her research interests include computer vision and deep learning.

WANG Guang-Yu, Ph. D., professor. His research interests include new multiple access technology and multi-carrier nodulation technology.

#### Background

In the field of remote sensing, the change detection is a basic task to identify and locate objects of interest from multi-temporal images collected in the same geographic area at different times. There are two types of important problems to be solved, One is that the bi-temporal remote sensing images are affected by natural factors such as illumination conditions, seasonal changes, thus showing large spatial spectral differences. In this case, the linear difference method is difficult to achieve accurate results. The other is the feature identification problem caused by the dense distribution of buildings, the shadow effect, the similarity of objects and etc. Due to the mismatch between the high level of abstraction of deep features and the comprehensive shallow spatial information.

Currently the UNet-based networks such as UNet, UNet++ is used for change detection in remote sensing because of its excellent performance. However, the UNet uses ordinary skip connections, and UNet++ uses a network structure of nested and dense skip connection. Those connec-

tions are in essence short connection, which does not directly extract enough information from multi-scale information. Furthermore, it is difficult to make good use of the original features of multi-scale information due to the problem of information fusion from individual connections. Therefore, some researchers introduce the attention mechanism to change detection, which is helpful to enhance changing features by automatically weighted to feature maps. But this results in large computational complexity and memory capacity. The Siamese network is another choice. It is proved that the Siamese network is a good solution for the change detection task for distinguishing changing pixels from invariant pixels in multi-temporal images detection. But the Siamese network usually has more network layers, which will bring the possibility of gradient disappearance, so the training of the network is more difficult.

This paper proposes an Edge-Guided Change Detection Base on UNet3+(EGCD-UNet3+) after analyzing the existing processing schemes. The UNet3+ uses the full-scale skip connections to directly combine both the deep and shallow semantics from different scale feature maps, and learns the hierarchical representation from the multi-scale aggregated feature maps. However it ignores the object scale during feature extraction, resulting in the mismatch between the receptive field and the scale. Therefore, the EGCD-UNet3+ first designed a Selective Kernel Block (SKB) with adaptive receptive field to replace the original block of the UNet3+, so that the image pairs have properties of adaptive receptive field, when extracting deep and shallow features. The EGCD-UNet3+ is composed of two parts: the encoding and the decoding. On the encoding side, due to the ability of capturing the long-term dependencies and modeling the context relationship of pixels, the Long Short Term Memory (LSTM) is used to design a Difference Enhancement Module (DEM) for analyzing the temporal correlation between image pairs, and solving the long-term dependency problem of the

bi-temporal itself. On the decoding side, the EGCD-UNet3+ proposes an Edge Guided Context Module (EGCM) to further improve the performance of building detection boundaries and effectively extract multi-scale spatial edge information at a finer-grained level. Finally, the EGCD-UNet3+ uses the composite loss function with both pixel segmentation error and edge segmentation error, so that the network can fully learn effective features for accurate label prediction.

The proposed model is validated on the LEVIR-CD dataset, the precision (P), the recall (R), the F1-score (F1), the overall segmentation accuracy (OA) can be increased to 90.75%, 96.68%, 93.15%, 99.12% respectively and the mean intersection over union (IoU) is significantly increased to 83.96%. By visual analysing, the proposed scheme achieves more accurate semantic boundary improvement and transformation detection of objects of multiple categories and scale.