

轻量级双侧输入 D-WNet 航空影像建筑物变化检测

张枫幸¹, 黄健², 李浩^{1*}¹河海大学地球科学与工程学院, 江苏 南京 211000;²江苏省测绘工程院, 江苏 南京 211000

摘要 针对传统语义分割网络变化检测结果易受阴影及其他地物干扰、建筑物边界分割较为粗糙的问题,提出一种轻量级双侧输入的变化检测网络 D-WNet。新网络从 W-Net 出发,采用深度可分离卷积块和空洞空间金字塔池化模块代替原本繁琐的卷积和下采样过程,利用右侧线特征编码器加强高低维特征的融合,同时在解码器上采样部分引入通道和时空注意力机制获取网络在不同维度下的有效特征,得到的 D-WNet 在性能方面有明显提升。在公开的 WHU 和 LEVIR-CD 建筑物变化检测数据集上进行实验,并与 W-Net、U-Net、ResNet、SENet 和 DeepLabv3+ 语义分割网络进行对比。实验结果表明,D-WNet 在交并比、F1 值、召回率、准确率和运行时间等 5 项指标中综合表现优异,对阴影干扰及建筑物边缘区域具有更精确的变化检测结果。

关键词 航空遥感; 建筑物变化检测; D-WNet; 双侧输入; 轻量级网络

中图分类号 P237

文献标志码 A

DOI: 10.3788/LOP231478

Lightweight Bilateral Input D-WNet Aerial Image Building Change Detection

Zhang Fengxing¹, Huang Jian², Li Hao^{1*}¹College of Earth Science and Engineering, Hohai University, Nanjing 211000, Jiangsu, China;²Jiangsu Academy of Surveying and Mapping Engineering, Nanjing 211000, Jiangsu, China

Abstract A lightweight dual-input change detection network, D-WNet, is proposed to address the issues of traditional semantic segmentation networks being susceptible to interference from shadows and other ground objects, as well as the rough boundary segmentation of buildings. The new network starts with W-Net and uses deep separable convolutional blocks and hollow space pyramid pooling modules to replace the originally cumbersome convolutional and downsampling processes. It utilizes a right-line feature encoder to enhance the fusion of high-dimensional and high-dimensional features and introduces channels and spatiotemporal attention mechanisms in the sampling section of the decoder to obtain effective features of the network in different dimensions. The resulting D-WNet has significantly improved performance. Experiments were conducted on the publicly available WHU and LEVIR-CD building change detection datasets, and the results were compared with the W-Net, U-Net, ResNet, SENet, and DeepLabv3+ semantic segmentation networks. The experimental results show that D-WNet performs well in five indicators (intersection-to-intersection ratio, F1 value, recall rate, accuracy rate, and running time) and has more accurate change detection results for shadow interference and building edge areas.

Key words aerial remote sensing; building change detection; D-WNet; bilateral input; lightweight network

1 引言

在遥感领域,多期数据的变化信息提取是一个重要研究方向。尽管目前变化检测方法多样,但大多数学者认为其是复杂的综合处理过程,尚没有哪一种方

法能够较好适应于所有变化检测的需要^[1]。传统的变化检测主要有影像代数法、空间信息转化法、影像分类方法等,但分别存在变化阈值选取困难^[2]、难以寻求非线性的依赖关系^[3]、无法检测出细微变化^[4]等问题。这些方法大多以像元处理为基础,仅统计图像的光谱特

收稿日期: 2023-06-08; 修回日期: 2023-07-09; 录用日期: 2023-07-24; 网络首发日期: 2023-08-07

基金项目: 国家自然科学基金(41471276)

通信作者: *lihao@hhu.edu.cn

征值,没有利用地物的形状、结构、相互关系,缺乏语义信息的提取,从而导致精度受到制约^[5]。

近年来得益于新算法与数据集的引入,各种深度卷积神经网络(CNN),如VGG、ResNet、SENet、DenseNet、Inception等被广泛应用,通过对语义分割神经网络FCN和U-Net^[6]进行编码器优化、空洞卷积设计、特征融合和残差模块构建^[7],改进得到DeepLab系列和RefineNet等网络模型^[8],不断在变化检测领域取得突破性的进展。一些学者利用U-Net编码器与解码器融合浅深层特征的特性,改善了建筑物的变化检测结果^[9]。王明常等^[10]在此基础上引入ResNet18残差结构,弥补了语义分割网络对小目标建筑检测的欠缺。郝明等^[11]基于DeepLabv3+深度学习网络,以大疆精灵正射影像为数据源进行迁移学习,总精度达到97%。压缩-激励单元是一种简单有效的注意力单元,陈良轩等^[12]在解码阶段横向连接结构中引入注意力机制增强通道间的学习能力,其改进的SE-ResNet在建筑物变化数据集上的综合性能提升5.88%。臧晓敏等^[13]使用SuffleNetV2模块替换U-Net的特征提取编码块,引入通道时空注意力机制,性能上获得较大的提升。

上述网络模型改善了分割尺度单一、信息传播低效的问题,但缺少原始影像的衍生特征输入,通道间的特征依赖性尚待加强。范自柱等^[14]使用特征金字塔网络和全局上下文模块改进得到了W-Net,使用影像重叠策略在Massachusetts数据集上进行实验,建筑物分类总精度达到79%以上,但面对复杂场景中的阴影干扰问题仍需优化。Zhang等^[15]提出压缩-激励W-Net模型,通过卷积层的深入以及网络的双侧输入使得训练精度更高,并利用光谱、纹理、形状等特征作为输

入提高复杂建筑物的分割能力,但模型参数的增多导致网络训练效率低下。Wang等^[16]使用特征金字塔网络(FPN)改进W-Net,解决了遥感卫星图像分割任务中精度不足的问题,但如何将建筑物边缘分割得更加精细、提升检测效率有待进一步研究。詹光莉等^[17]提出一种改进注意力机制与焦点损失函数的IAW-Net,通过叠加注意力向量、调整网络权重项的分割比重,有效解决小目标物漏检、误检的问题。

针对以上W型网络在建筑物分割任务中边缘不够精细、运行效率低下的问题,本文从融入建筑物形状特征出发,设计双侧输入的轻量级D-WNet,通过影像边缘特征的限制提升语义分割的精度,并使用深度卷积块和空洞金字塔池化(ASPP)模块代替原本繁琐的卷积和下采样过程,提升W型网络的运行效率。另外,针对上述复杂场景下语义分割网络易受阴影及其他地物干扰的问题,在D-WNet的解码器上采样部分引入通道和时空的注意力机制模块(CBAM),保证在编码过程中不破坏特征提取部分权值的情况下,获取通道与空间中丰富的语义信息。

2 原理与方法

2.1 双侧输入的W-Net

双U结构是U-Net的一种变型,如图1所示,它对数据输入路径进行扩充、加深了卷积层,能满足多源数据和原始输入图像的重建与分割,故也被称为W-Net^[18]。对于变化检测而言,W-Net可以在两端分别输入前后时相的遥感影像,在解码器部分单独学习两个时期建筑物的变化特征,但在下采样过程中仅通过卷积提取单幅影像特征,两侧编码器缺少对变化信息的提取能力。其次,也可以将双时相影像叠加作为左侧

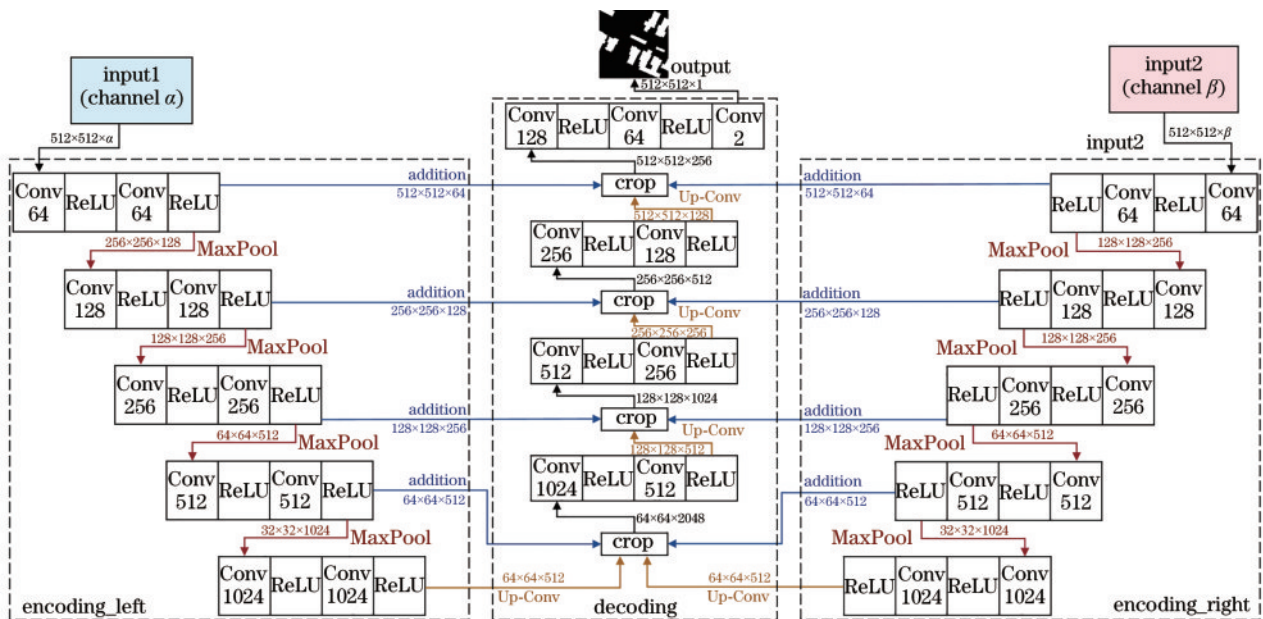


图1 双侧输入的W-Net

Fig. 1 Two-side input W-Net

的输入,右侧输入其多源特征信息,但是建筑物的衍生特征表现形式各异,需要针对性地选择合适的特征信息来提高建筑物的边缘分割精度。

网络两侧的编码器各由 5 个模块组成,每个模块进行两次 3×3 的卷积操作、非线性函数(ReLU)激活和标准化处理来获取建筑物变化的基本特征。模块尾部除了通过 2×2 的最大池化输出给下一层,还会在解码部分堆叠融合高低维特征,但加深网络提取到更多特征的同时也会导致堆叠的特征通道数成倍增加。单一的 U-Net 参数量仅为 94 MB,而 W-Net 两侧收缩路径的权值相互独立,虽达到了多源数据非线性模拟的目的,参数量却达到了 243 MB,当待寻优的参数过多时,不仅训练、收敛速度会变得缓慢,且容易造成参数的过拟合。

解码阶段的 4 次上采样结构与收缩路径类似,通过拼接相应层的输入来恢复可能因编码而丢失的空间信息,最后使用一个 1×1 的卷积和 Softmax 层将 64 维的特征向量映射回原始输入重构,输出每个像素建筑物变动的概率,得到变化检测的结果。图像缩小放大过程中采用的卷积和上、下采样虽然能去除部分干扰噪声,但难以充分学习阴影及其他干扰地物与建筑物在通道和空间轴上的不相关性,缺少在大量的深度学习样本中过滤无关信息、降低干扰权重的能力,会产生分割错误、遗漏等问题。

2.2 建筑物的形状特征

由遥感数据衍生出来的纹理、颜色、形状、高度等特征,往往表现出性质稳定、受辐射差异影响小、不易受时相变化干扰等特点。其中的形状特征更是识别建筑物的重要信息之一,它能够提供更清晰的边缘特征,保留图像中重要的结构属性,在检测小而复杂的建筑物

时,能够弥补光谱、纹理特征容易混淆、不易检测的不足。

Sobel 和 Canny 是两种经典的边缘检测算法,均通过计算梯度导数最大的方向寻找影像中的边界信息。Sobel 的抗噪能力较强,但是边缘定位具有多个像素,准确度较低;Canny 通过双阈值限制能够标识到真正的弱边缘,但是容易错误标记噪声。二者核心公式分别如下:

$$G = \sqrt{G_x^2 + G_y^2}, \quad (1)$$

$$g_{up}(i, j) = \left(1 - \frac{G_y}{G_x}\right) \cdot g_{xy}(i, j + 1) + \frac{G_y}{G_x} \cdot g_{xy}(i - 1, j + 1), \quad (2)$$

$$g_{down}(i, j) = \left(1 - \frac{G_y}{G_x}\right) \cdot g_{xy}(i, j - 1) + \frac{G_y}{G_x} \cdot g_{xy}(i + 1, j - 1), \quad (3)$$

式中: G 为 Sobel 所求的梯度图像; g_{up} 、 g_{down} 为 Canny 正负方向的两个梯度强度; G_x 、 G_y 分别为像素点 x 、 y 方向的梯度; g_{xy} 为中心像素点 (i, j) 的梯度强度。

将 Canny、Sobel 以及两种算法叠加放入网络右侧输入端的线特征提取模块进行消融实验,以此来探讨最适合 W 型网络的边缘检测算法,对特征进行加强的同时弥补 W-Net 建筑物边缘分割不够精细的缺点。

2.3 通道和时空注意力机制

注意力机制的核心就是自适应调节网络权重,集中到一部分输入特征上。CBAM^[19]将通道和空间注意模块串联,每个分支可以分别学习通道和空间轴上的重要特征,能有效削弱遮挡、阴影等复杂场景下的干扰。CBAM 对特征层的卷积运算通过混合跨通道和空间信息来提取,分别从通道和空间轴两个维度对网络重要特征进行学习,如图 2 所示。

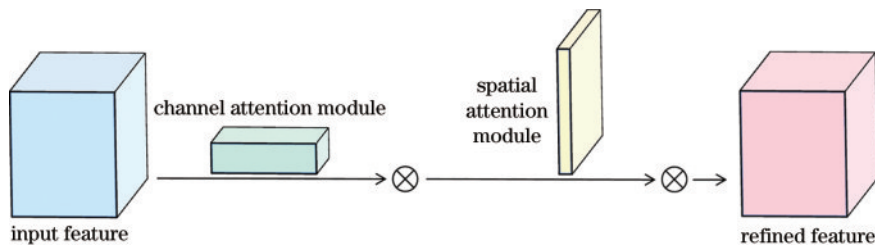


图 2 CBAM 模块的实现过程
Fig. 2 Implementation process of CBAM

模块前半部分的通道注意力机制(M_c)可以分为两个部分,首先对输入的特征图(F)在高和宽上分别进行全局最大池化和全局平均池化,得到与输入通道数一致的特征长条。再对两个池化结果利用共享的全连接层进行处理,第 1 层的神经元个数较少,第 2 层的神经元个数等于通道数。接着使用 Sigmoid(σ)对处理后的两个结果进行激活相加,便可获得输入特征图介于 0~1 之间每一个通道的权值。计算公式和示意图如式(4)和图 3 所示:

$$M_c(F) = \sigma \left\{ \text{MLP} \left[\text{AvgPool}(F) \right] + \text{MLP} \left[\text{MaxPool}(F) \right] \right\}. \quad (4)$$

与通道注意力所不同,模块后半部分的空间注意力机制(M_s)对输入进来的特征层每一个特征点的通道上取最大值和平均值,将池化堆叠后的结果利用 $k \times k$ 的卷积($f^{k \times k}$)调整通道数,再进行 Sigmoid(σ)激活。计算公式和示意图如式(5)和图 4 所示:

$$M_s(F) = \sigma \left\{ f^{k \times k} \left\{ \left[\text{AvgPool}(F); \text{MaxPool}(F) \right] \right\} \right\}. \quad (5)$$

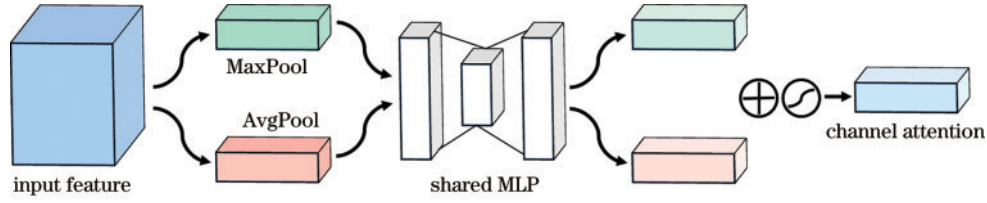


图 3 通道注意力模块

Fig. 3 Channel attention module

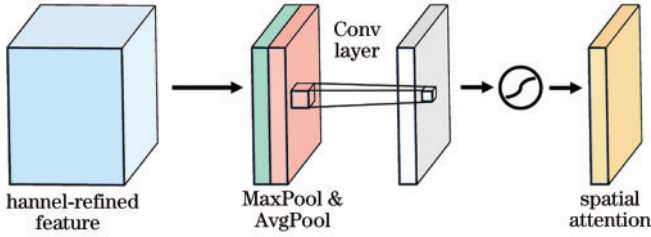


图 4 空间注意力模块

Fig. 4 Spatial attention module

2.4 改进的轻量级 D-WNet

随着特征层变多,权重参数会几何式倍增,所提 D-WNet 摒弃了 W 型网络双侧编码器中大量的卷积和下采样过程,使用深度可分离卷积块获取影像的低维特征,并在线性层后利用空洞空间金字塔池化模块获得图像更大的感受野,融合至解码器与高维特征共同学习 CBAM 的通道空间特征并参与上采样环节,有效提升网络分割性能的同时参数量由 243 MB 减少至 32 MB,网络结构如图 5 所示。

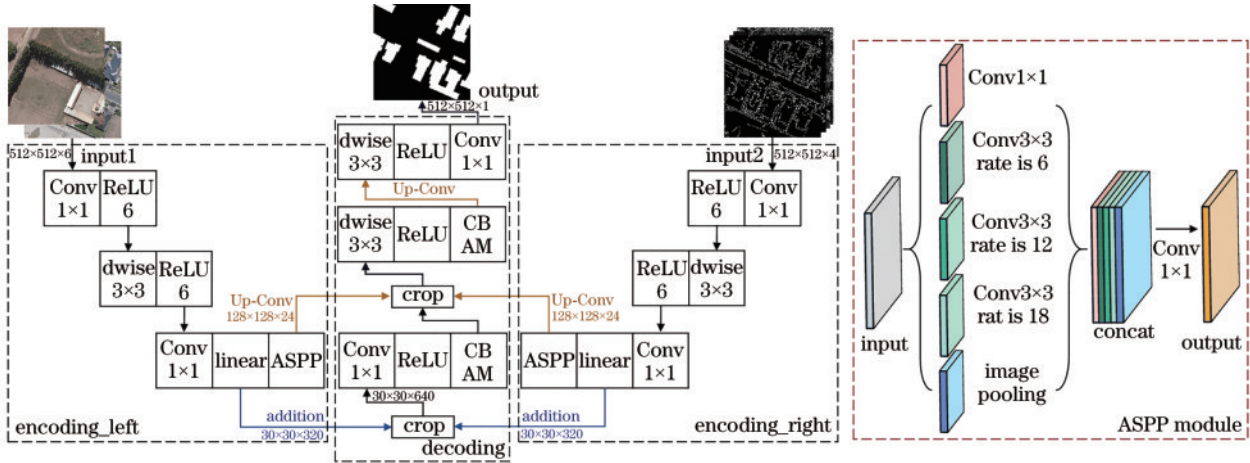


图 5 轻量级双侧输入 D-WNet

Fig. 5 Lightweight bilateral input D-WNet

为了弥补卷积层的减少导致的网络高级抽象特征的缺失,在线性层后使用 ASPP 模块^[20]对多尺度特征信息进行加强提取,采用 1 个 1×1 普通卷积层、3 个扩张率(rate)分别为 6、12、18 的 3×3 空洞卷积层和 1 个平均池化层对提取的特征进行联接。空洞卷积在参数量级不变的情况下比标准卷积拥有更大的感受野,并且可以通过设置 rate 的大小定义特征提取时卷积核的间距从而获得不同的感受野。因此 ASPP 对于变化检测下多尺度的目标分割信息,能有效提取出各尺度目标及边缘,输出更高级别的深层影像特征。

右侧特征编码器充分融入建筑物的形状特征,与左侧端口并行在倒残差部分使用 1×1 的卷积分别进行升维和降维,中间利用 3×3 的深度卷积提取特征,最后使用线性函数激活并接入 ASPP 模块进行加强特征提取。上采样的过程中会先融合两侧的浅层特征,通过 1×1 的卷积调整通道数与深特征层合并后,采用

两次深度可分离卷积块恢复原始图像,CBAM 则放置在浅特征层和第一个深度卷积块激活函数之后,充分学习深浅特征层的通道和空间关系来抵抗阴影等复杂环境的干扰。

3 结果与分析

3.1 数据集与图像预处理

本实验采用以下数据集:1) WHU Building^[21]下的建筑物变化检测数据集,20.5 km² 的区域中包含 2012 年 4 月航拍的 12796 座建筑物与 2016 年同一区域的 16077 座建筑物,分辨率为 0.075 m。通过步幅为 200 pixel 的滑动剪裁,得到 9360 张 512 pixel \times 512 pixel 的图像集,其中,6240 张用作训练,3120 张用作测试和验证。2) LEVIR-CD 数据集^[22],由 637 个 0.5 m 分辨率的图像补丁对组成,时间跨度为 5~14 年,具有显著的土地利用变化。剪裁得到 2548 张

512 pixel×512 pixel 的图像集,取 1698 张图像作为训练集,850 张图像作为测试集和验证集。同时在训练集的载入过程中,随机对 DataLoader 中加载的 Dataset

进行旋转、缩放、翻转、裁剪等数据增强操作,以此来提高模型的鲁棒性和泛化能力。图 6 是两个数据集的示例,其中,白色区域代表变化的建筑物。

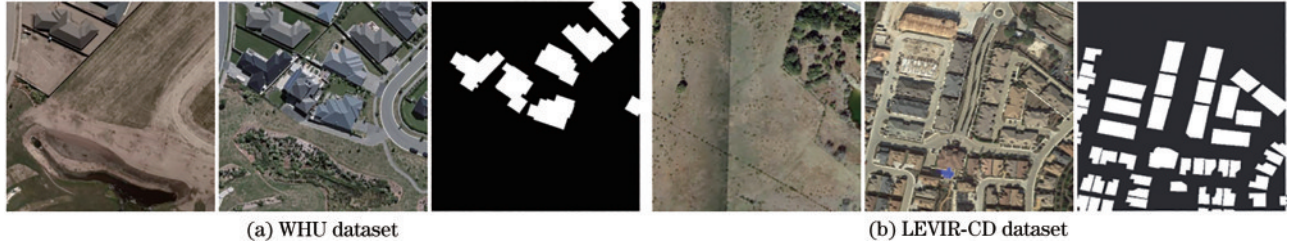


图 6 数据集中的双时期影像和相应标签

Fig. 6 Two-period images and corresponding labels in the datasets

图 7 为通过线特征提取模块对 W 型网络右侧分支进行预处理的结果,由于建筑物在影像上具有与其他地物区分明显的边缘特征,采用 7×7 高斯模糊后的

Canny 算子(双阈值为 160、65)和 Sobel 算子进行形状特征提取,在减少图像数据量的同时有效保留原始的结构属性。

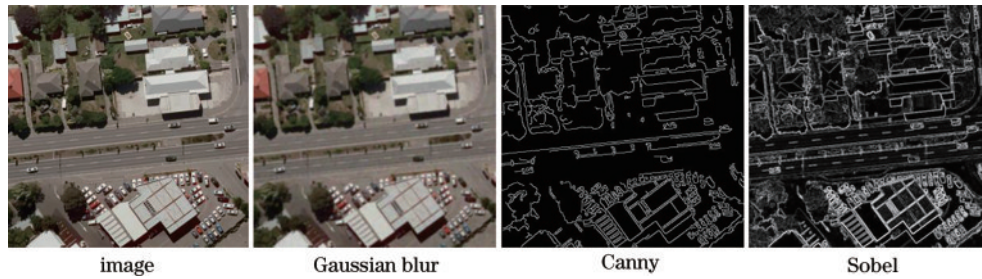


图 7 线特征提取模块图像处理结果

Fig. 7 Image processing results of line feature extraction module

3.2 网络精度评价指标

在语义分割网络中,交并比(R_{IoU})、召回率(R_{recall})、精准率($R_{precision}$)、F1 分数(s_{F1})足以从像素级别全面地量化该网络的分割结果。因此,采用上述 4 种精度指标来评价深度学习网络模型的性能。

$$R_{IoU} = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}}, \quad (6)$$

$$R_{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (7)$$

$$R_{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (8)$$

$$s_{F1} = 2 \times \frac{R_{precision} \times R_{recall}}{R_{precision} + R_{recall}}, \quad (9)$$

式中: N_{TP} 表示真阳性的数量; N_{FP} 表示假阳性的数量; N_{FN} 表示假阴性的数量; N_{TN} 表示真阴性的数量。

3.3 网络参数设置

基于 PyTorch 框架搭建深度学习网络模型,实验的运行硬件环境为 Intel(R) Core(TM) i7-12700KF CPU, NVIDIA GeForce RTX 3080 GPU(10 GB)。对网络参数做如下设置:各端输入训练样本大小为 512 pixel×512 pixel,学习率在 $10^{-4} \sim 10^{-6}$ 之间动态调整,冻结阶段和解冻阶段训练的数据量分别为 6 和 4,使用 Adam 优化器与交叉熵损失函数,激活函数采用

ReLU。为有效防止主干模型的权值被破坏,提升运行效率节约时间成本,每组实验均设置 100 epoch 的冻结训练和 200 epoch 的解冻训练。

3.4 形状特征选取消融实验

为了验证所提 D-WNet 融入建筑物形状特征的有效性,在保证网络结构和各模块均一致的前提下输入不同的数据进行对比。两侧端口组合的不同情况有:左右各输入前后时期影像,不融入形状特征,记为消融模型 A;左侧输入双时期叠加影像,右侧输入 Canny、Sobel 以及两种算子叠加的处理结果,分别记为消融模型 B、C 和 D-WNet。融入不同形状特征的建筑物评价指标对比如表 1 所示,变化分割结果对比如图 8 所示。

评价指标中的 R_{IoU} 值是对象类别分割问题的标准性能度量,能给出在该组图像中存在对象的预测区域和地面实况区域之间的相似性,因此以 R_{IoU} 作为评价最佳网络的性能指标。同时为了更加直观地比较各模块对结果的影响,使用不同的颜色来区分建筑物的变化检测结果。

图 8 表明,融入 Canny 和 Sobel 两种算子的 D-WNet 较模型 A 能分割出更多难以识别的小型建筑物与天台区域,比模型 B、C 的建筑物边缘分割效果要更好。由表 1 可见,D-WNet 在 WHU 和 LEVIR-CD 测试

表 1 形状特征选取实验评价指标对比

Table 1 Comparison of experimental evaluation indexes of shape feature selection

unit: %

Dataset	Network	Input_left	Input_right	R_{recall}	$R_{precision}$	s_{F1}	R_{IoU}
WHU	A	Former phase image	Post-temporal image	98.01	97.61	97.81	95.54
	B	Biphasic image	Canny	96.71	98.30	97.50	95.07
	C	Biphasic image	Sobel	97.42	98.17	97.80	95.56
	D-WNet	Biphasic image	Canny+Sobel	97.49	98.81	98.14	96.36
LEVIR-CD	A	Former phase image	Post-temporal image	90.82	91.40	91.11	83.36
	B	Biphasic image	Canny	89.76	91.84	90.79	82.94
	C	Biphasic image	Sobel	89.40	91.98	90.67	83.25
	D-WNet	Biphasic image	Canny+Sobel	90.23	92.25	91.23	83.49

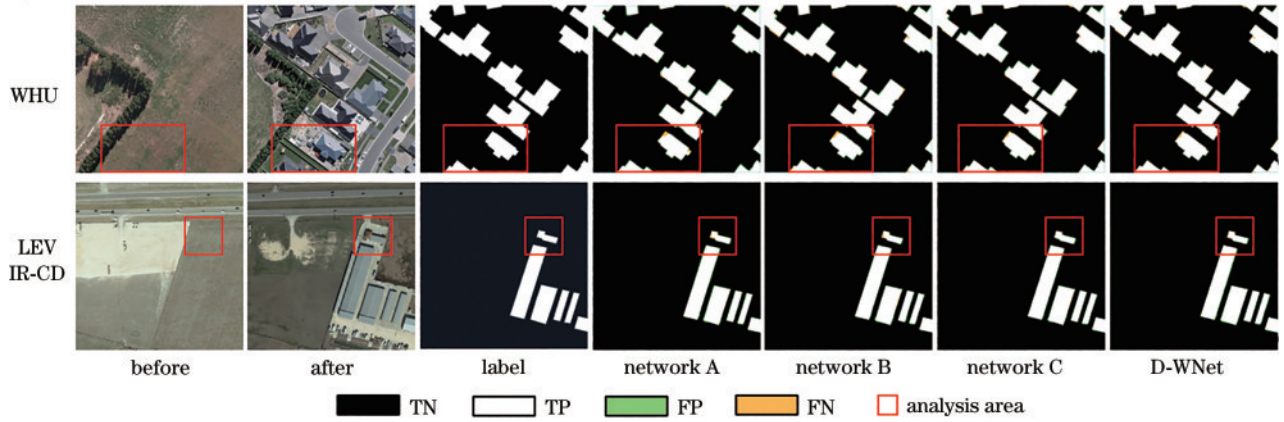


图 8 形状特征选取实验消融结果对比

Fig. 8 Comparison of ablation results of shape feature selection experiment

集上的 R_{IoU} 、 s_{F1} 和 $R_{precision}$ 均高于其他 3 种消融模型, 在 R_{recall} 上却比消融模型 A 分别低 0.52 个百分点和 0.59 百分点。模型 A 未融入形状特征, 但网络两侧对两幅影像独自进行卷积操作, 能获得更加精细的变化细节, 发现正样本的能力有所提升, 但 $R_{precision}$ 和 s_{F1} 值的降低表明缺少边缘限制的网络正样本相关性并不高。综合所有指标可知, 当同时叠加两种边缘提取算法时效果更佳。

3.5 网络结构改进消融实验

针对所提变化检测网络, 为了验证改进的轻量型 W 结构及 CBAM、ASPP 模块对网络整体性能的影响,

采用 3 种消融模型与 D-WNet 进行对比分析。在上采样部分放入 CBAM 模块的 W-Net 记为模型 D; 删去注意力 CBAM 模块, 保留改进的轻量型 W 结构的网络记为模型 E; 保留 CBAM 模块, 删去轻量级结构中 ASPP 模块的网络记为模型 F。表 2 和图 9 分别为网络结构改进消融模型在 WHU 和 LEVIR-CD 测试集上精度评价结果和建筑物变化检测结果。

从表 2 可以看出, 所提 D-WNet 使用的 CBAM、ASPP 模块和改进的轻量型 W 结构在 $R_{precision}$ 、 s_{F1} 值和 R_{IoU} 这 3 项精度指标中作用明显。缺少 CBAM 模块的网络 E 单轮训练时间最快, 但 R_{IoU} 值在 WHU 和

表 2 网络结构改进实验评价指标对比

Table 2 Comparison of experimental evaluation indexes of network structure improvement

Dataset	Network	CBAM	Lightweight structure	ASPP	$R_{recall} / \%$	$R_{precision} / \%$	$s_{F1} / \%$	$R_{IoU} / \%$	Epoch time / s
WHU	D	✓	×	×	97.96	97.90	98.03	96.02	1502
	E	×	✓	✓	96.60	98.55	97.56	95.25	438
	F	✓	✓	×	97.69	97.60	97.64	95.97	560
	D-WNet	✓	✓	✓	97.49	98.81	98.14	96.36	454
LEVIR-CD	D	✓	×	×	90.29	92.13	91.20	83.38	786
	E	×	✓	✓	88.21	92.20	90.16	82.67	235
	F	✓	✓	×	91.22	90.71	90.96	83.22	312
	D-WNet	✓	✓	✓	90.23	92.25	91.23	83.49	260

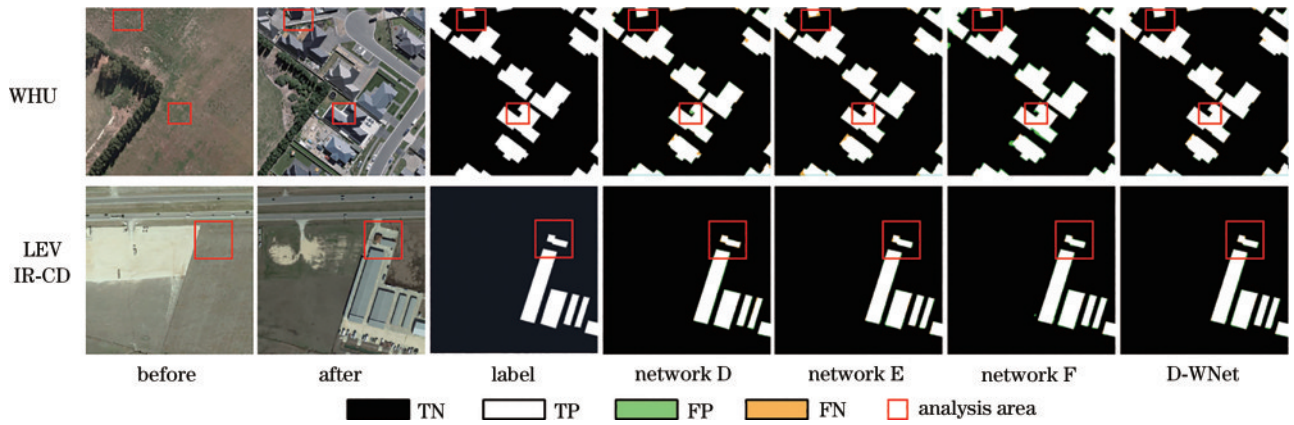


图 9 网络结构改进实验消融结果对比

Fig. 9 Comparison of ablation results of network structure improvement experiment

LEVIR-CD测试集上分别降低 1.11 个百分点和 1.18 百分点,影响显著;缺少改进轻量型结构的 W-Net 精度影响较小, R_{IoU} 值仅降低 0.34 个百分点和 0.11 百分点,但单次迭代运行时间弊端明显,增加约 2 倍;D-WNet 使用 ASPP 模块对多尺度特征信息进行加强,弥补了卷积层减少导致的网络高级抽象特征的缺失,以较小的 R_{recall} 损失为代价获得了更高的分割效率与 R_{IoU} 精度,网络 F 去除 ASPP 模块后,3 项精度指标出现了明显的下降,充分说明改进的轻量型结构中融入 ASPP 模块的必要性。消融实验结果表明, CBAM、ASPP 模块和改进的轻量型 W 结构在建筑物变化检测中的重要性,且 CBAM 模块的影响程度更大。

网络结构改进消融模型的变化检测结果如图 9 所示。从分析区域可以明显看到,去除 CBAM 模块之后的网络 D 对于小型建筑物的阴影部分检测能力明显减弱,无法完整地提取出包含阴影的建筑物变化范围。该现象出现的原因可能是由于失去了 CBAM 模块后网络难以在通道和空间多维度捕捉充足的信息,因此无法区分建筑物边界与其阴影。网络 F 的错分区域最多,表明缺少 ASPP 模块的轻量化网络在减少卷积缺失高级特征层后易受阴影等复杂场景干扰,无法获得

有效的变化检测结果。W-Net 较网络 E 与 F 检测效果有所提升,但与 D-WNet 相比,对于建筑物边界的检测能力尚有欠缺,导致此现象的原因可能是仅凭多次的卷积与下采样难以获得各尺度目标的多层影像特征,并且网络的超参数过多,运行效率的低下可能导致未达到有效的预训练程度。所提网络引入 CBAM、ASPP 模块和改进的轻量型 W 结构,通过与右侧线特征提取编码器的结合,有效解决了建筑物阴影干扰及其边缘识别问题,充分说明了上述两个模块结合的有效性。

图 10 为 4 种消融模型在测试集上的特征提取热图。可以看出:缺少 CBAM 模块的网络 E 特征区域关注较为杂乱,难以区分水泥路面与建筑物边界;W-Net、缺少 ASPP 模块的轻量化网络 F 由于无法从多尺度上获得建筑物的变化信息,而把部分关注点放在了道路上;D-WNet 获取的特征热图把大部分重心聚集到了产生变化的建筑物边缘区域,准确地凸显出各建筑物的主体,边界信息较为明显清晰,这表明网络中的 CBAM 模块能够对表示建筑物的特征进行加权强化,ASPP 模块也能从不同尺度获得建筑物的上下文信息,从而对产生变化的建筑物分割出指向性强的特征。

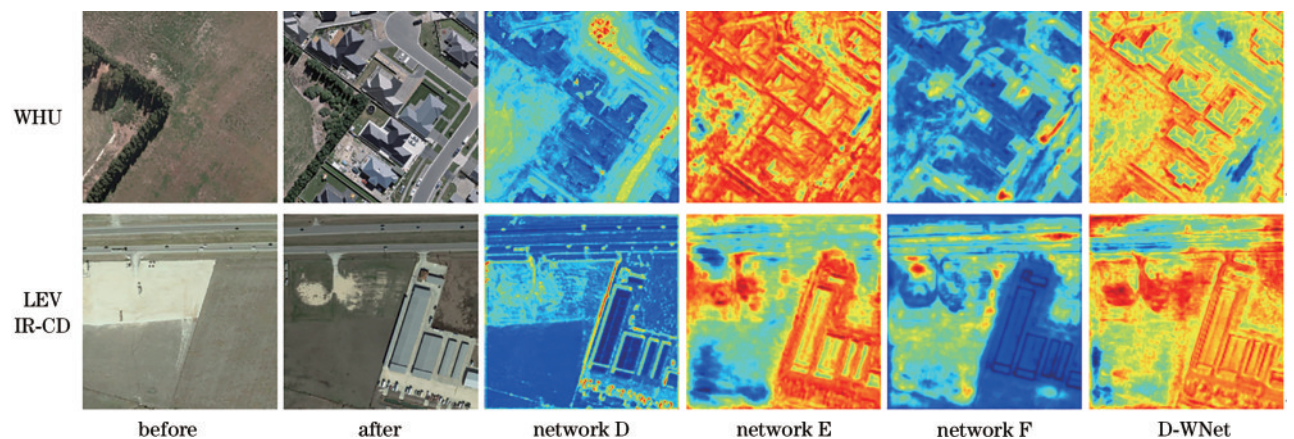


图 10 网络结构改进消融模型的特征提取热图

Fig. 10 Feature extraction heat map of the network structure improved ablation model

3.6 与其他方法比较

为了更好地评估 D-WNet 的性能,在保证超参数均相同的情况下,采用当前主流的语义分割网络:U-Net、SENet、ResNet 和 DeepLabv3+ 进行对比分析。表 3 列出了 5 种网络模型在 WHU 和 LEVIR-CD 测试集上的建筑物变化检测精度。可以看出,D-WNet 在 5 项精度指标中综合表现优异, $R_{precision}$ 、 S_{F1} 和 R_{IoU} 在 WHU 测

试集上提升 0.56 百分点~1.25 百分点、0.11 百分点~0.98 百分点和 0.05 百分点~1.87 百分点,在 LEVIR-CD 测试集上提升 0.04 百分点~0.52 百分点、0.63 百分点~1.20 百分点和 0.93 百分点~1.63 百分点,单次迭代运行时间较 SENet 最高减少 127 s 和 114 s,充分说明所提 D-WNet 性能得到有效提升。

表 3 不同网络在 WHU 和 LEVIR-CD 测试集上的建筑物变化检测精度

Table 3 Building change detection accuracy of different networks on WHU and LEVIR-CD test set

Dataset	Network	$R_{recall} / \%$	$R_{precision} / \%$	$S_{F1} / \%$	$R_{IoU} / \%$	Epoch time / s
WHU	U-Net	97.82	98.25	98.03	96.16	555
	SENet	97.95	98.13	98.04	96.31	581
	ResNet	97.75	98.04	97.89	95.88	455
	DeepLabv3+	96.77	97.56	97.16	94.49	224
	D-WNet	97.49	98.81	98.14	96.36	454
LEVIR-CD	U-Net	88.53	92.17	90.31	82.55	365
	SENet	89.04	92.21	90.60	82.56	374
	ResNet	88.91	91.73	90.30	82.14	262
	DeepLabv3+	88.06	92.08	90.03	81.86	160
	D-WNet	90.23	92.25	91.23	83.49	260

图 11 为 5 种网络在测试集上的部分变化检测结果。对于第 1 行影像前后时相变化较大的新旧建筑物,5 种网络都提取出了大致完整的变化区域,但 U-Net 和 DeepLabv3+ 错误地将变化建筑物之间的道路分割了出来,SENet 和 ResNet 受到了后时相新建建筑物覆盖的干扰,只有 D-WNet 在完整性方面做得最好。第 2 行影像的分析区域受到了明显的树木遮挡及阴影干扰,U-Net、SENet 和 ResNet 均错分了该建筑物的边

缘区域,DeepLabv3+ 甚至把部分阴影投影下的建筑物区域划分为负类出现了遗漏现象,只有 D-WNet 实现了更为准确的分割。第 3 行影像中有屋顶纹理较为复杂的建筑物,SENet 和 ResNet 难以识别,U-Net、DeepLabv3+ 和 D-WNet 均能分割出基本的正确形状,但 D-WNet 仍然具有更高的精确度。对于第 4 行影像,D-WNet 较其他网络分割出了更加精准的边缘区域,漏分、错分现象明显减少。

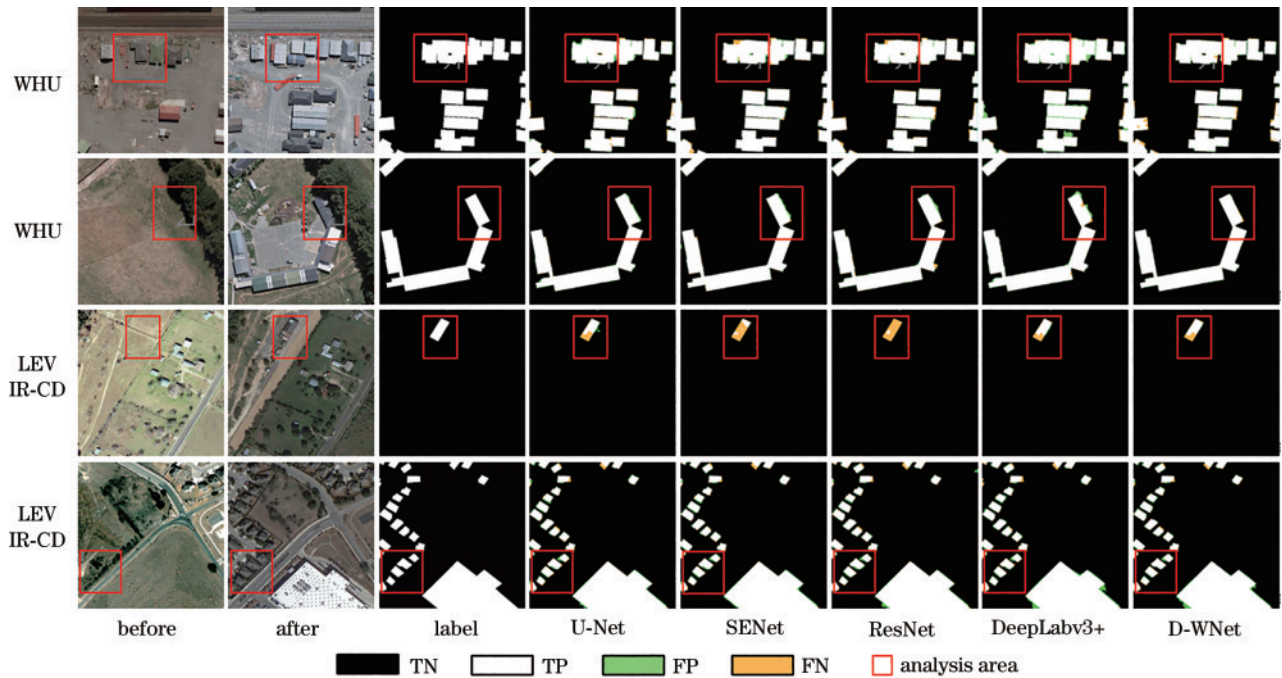


图 11 不同网络在 WHU 和 LEVIR-CD 测试集上的建筑物变化检测结果

Fig. 11 Building change detection results for different networks on WHU and LEVIR-CD test sets

从图 11 的结果来看,高分辨率航空影像中前后时期变化的道路与建筑物在外观上有许多的相似性,由于 U-Net 和 SENet 缺少在空间轴上学习重要特征的能力,仅从单一的通道维度对建筑物的变化进行分割,会受部分阴影、植被、道路等干扰出现错分的现象。ResNet 的残差块有效降低了网络参数的计算量,但对于变化建筑物的边界难以做出有效的限制。DeepLabv3+ 虽然采用了空洞卷积层扩大了不同尺度影像特征的感受野,但在实验中较其他网络错分、漏分现象最为严重,受干扰影响最大。与上述 4 个网络相比, D-WNet 在建筑物分割和抗干扰的性能上表现最优,变化检测结果与标签重合度最大。网络右侧解码器线特征提取模块的嵌入对建筑物的形状特征进行了约束加强,获得了更好的边界分割效果;双侧解码器中深度可分离卷积块和 ASPP 模块的替换缩短了网络单轮训练时间,提高了模型的整体效率;通过在上采样层添加 CBAM 模块,有效提升了网络学习通道与空间轴上重要特征的能力,获得了最为完整的双时相建筑物变化范围。实验结果充分表明,通过形状特征约束的 D-WNet 能从通道和时空维度更具针对性地获得建筑物边缘的变化特征信息,从而适应受阴影、植被、道路等干扰下的建筑物变化检测。

4 结 论

针对双时相高分辨率航空影像中建筑物边缘分割粗糙及阴影干扰的变化检测问题,融入建筑物的形状特征改进轻量级 W 网络结构并引入 CBAM 模块,提出一种轻量级双侧输入的 D-WNet。实验结果表明: 1) D-WNet 在 WHU 测试集上的 S_{F1} 与 R_{IoU} 达到 98.14% 和 96.36%, 在 LEVIR-CD 测试集上达到 91.23% 和 83.49%, 均超过对比网络; 2) 网络替换了原本繁琐的卷积和多次下采样过程, 在两个数据集上 D-WNet 单次迭代运行时间仅次于 DeepLabv3+, 为 454 s 和 267 s, 32 MB 的参数量较 W-Net 大大减少, 有效提升了 W 型网络的分割效率; 3) 从建筑物的变化检测结果来看, 各个网络均可以对干扰较少的明显变化建筑物实现精确分割, 但只有融入边缘特征的 D-WNet 在使用 CBAM、ASPP 模块后获取到通道与时空中丰富的多尺度信息, 在充分抑制树木、道路、阴影等环境干扰的同时更精确地分割出了建筑物的边缘区域。然而所提网络在公共数据集上存在过拟合的倾向, 后续的工作中将重点研究其在其他研究区域中的迁移性, 以提高双侧输入网络的泛化能力。

参 考 文 献

- [1] 徐亚瑾, 舒红. 基于 ISODATA 和变化矢量分析法的影像变化提取方法[J]. 地理空间信息, 2020, 18(1): 73-76, 9.
Xu Y J, Shu H. Image change extraction method based on ISODATA and change vector analysis[J]. Geospatial Information, 2020, 18(1): 73-76, 9.
- [2] 周婷, 邹俊, 崔玉环, 等. 基于主成分分析和支持向量机的遥感影像水体提取方法及应用[J/OL]. 水资源保护: 1-15[2022-07-04]. <http://kns.cnki.net/kcms/detail/32.1356.TV.20220407.1836.006.html>.
Zhou T, Zou J, Cui Y H, et al. Water extraction from Remote sensing image based on principal component analysis and support vector machine and its application[J/OL]. Water Resource Conservation: 1-15[2022-07-04]. <http://kns.cnki.net/kcms/detail/32.1356.TV.20220407.1836.006.html>.
- [3] Wang M C, Zhang H M, Sun W W, et al. A coarse-to-fine deep learning based land use change detection method for high-resolution remote sensing images[J]. Remote Sensing, 2020, 12(12): 1933.
- [4] 孙晓霞, 张继贤, 燕琴, 等. 遥感影像变化检测方法综述及展望[J]. 遥感信息, 2011, 26(1): 119-123.
Sun X X, Zhang J X, Yan Q, et al. A summary on current techniques and prospects of remote sensing change detection[J]. Remote Sensing Information, 2011, 26(1): 119-123.
- [5] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[EB/OL]. (2016-02-24)[2023-02-03]. <https://arxiv.org/abs/1602.07360>.
- [6] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [7] 罗松强, 李浩, 陈仁喜. 多尺度特征增强的 ResUNet+ 遥感影像建筑物提取[J]. 激光与光电子学进展, 2022, 59(8): 0828007.
Luo S Q, Li H, Chen R X. Building extraction of remote sensing images using ResUNet+ with enhanced multiscale features[J]. Laser & Optoelectronics Progress, 2022, 59(8): 0828007.
- [8] Rajaram R N, Ohn-Bar E, Trivedi M M. RefineNet: refining object detectors for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2016, 1(4): 358-368.
- [9] 汤富平, 吴臻涵, 蔡东燕, 等. 基于 UNet 的差值影像变化检测试验与分析[J]. 地理空间信息, 2022, 20(11): 25-28.
Tang F P, Wu Z H, Cai D Y, et al. Experiment and analysis of difference image change detection based on UNet[J]. Geospatial Information, 2022, 20(11): 25-28.
- [10] 王明常, 朱春宇, 陈学业, 等. 基于 FPN Res-Unet 的高分辨率遥感影像建筑物变化检测[J]. 吉林大学学报(地球科学版), 2021, 51(1): 296-306.
Wang M C, Zhu C Y, Chen X Y, et al. Building change detection in high resolution remote sensing images based on FPN Res-Unet[J]. Journal of Jilin University (Earth Science Edition), 2021, 51(1): 296-306.
- [11] 郝明, 田毅, 张华, 等. 基于 DeepLabV3+ 深度学习的无人机影像建筑物变化检测研究[J]. 现代测绘, 2021, 44

- (2): 1-4.
Hao M, Tian Y, Zhang H, et al. Research on building change detection of UAV images based on DeepLabV³⁺ deep learning[J]. *Modern Surveying and Mapping*, 2021, 44(2): 1-4.
- [12] 陈良轩, 于海洋, 李英成, 等. 一种融合注意力机制的建筑物变化检测模型[J]. *测绘科学*, 2022, 47(4): 153-159.
Chen L X, Yu H Y, Li Y C, et al. A model for detecting building changes incorporating attention mechanisms[J]. *Science of Surveying and Mapping*, 2022, 47(4): 153-159.
- [13] 臧晓敏, 李志鹏, 胡家昕. 基于孪生网络的遥感图像变化检测方法[J]. *信息与电脑(理论版)*, 2021, 33(18): 56-58.
Zang X M, Li Z P, Hu J X. A change detection method of remote sensing image based on Siamese network[J]. *China Computer & Communication*, 2021, 33(18): 56-58.
- [14] 范自柱, 王松, 张泓, 等. 基于 W-Net 的高分辨率遥感卫星图像分割[J]. *华南理工大学学报(自然科学版)*, 2020, 48(12): 114-124.
Fan Z Z, Wang S, Zhang H, et al. W-Net-based segmentation for remote sensing satellite image of high resolution[J]. *Journal of South China University of Technology (Natural Science Edition)*, 2020, 48(12): 114-124.
- [15] Zhang H M, Wang M C, Wang F Y, et al. A novel squeeze-and-excitation W-net for 2D and 3D building change detection with multi-source and multi-feature remote sensing data[J]. *Remote Sensing*, 2021, 13(3): 440.
- [16] Wang S, Fan Z Z, Li Z M, et al. An effective lunar crater recognition algorithm based on convolutional neural network[J]. *Remote Sensing*, 2020, 12(17): 2694.
- [17] 詹光莉, 刘辉, 杨路. 基于改进注意力 W-Net 的工业烟尘图像分割[J]. *计算机集成制造系统*, 2023, 29(2): 628-637.
Zhan G L, Liu H, Yang L. Industrial smoke image segmentation based on improved attention W-Net[J]. *Computer Integrated Manufacturing Systems*, 2023, 29(2): 628-637.
- [18] Xia X D, Kulis B. W-Net: a deep model for fully unsupervised image segmentation[EB/OL]. (2017-11-22) [2023-03-06]. <https://arxiv.org/abs/1711.08506>.
- [19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018*. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [20] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [21] Ji S P, Wei S Q, Lu M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(1): 574-586.
- [22] Chen H, Shi Z W. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection[J]. *Remote Sensing*, 2020, 12(10): 1662.