引文格式:隋百凯,曹云刚,程海波,等.基于几何先验约束的高点多视角损毁建筑物检测方法[J].武汉大学学报(信息科学版),2025,50(6):1088-1099.DOI:10.13203/j.whugis20240297

Citation:SUI Baikai,CAO Yungang,CHENG Haibo,et al.Damaged Building Detection Method in Multi-view Scenes Based on Geometric Prior Constraints[J].Geomatics and Information Science of Wuhan University,2025,50(6):1088-1099.DOI:10.13203/j.whugis20240297

基于几何先验约束的高点多视角损毁建筑物检测方法

隋百凯[1] 曹云刚[1] 程海波[1] 杨 鹏[1] 曾雅慧[1] 谢亚坤[1] 朱 军[1]

1 西南交通大学地球科学与工程学院,四川 成都,611756

摘 要:建筑物是重要的受灾体之一,损毁建筑物检测与空间映射是应急救援的关键。高点多视角遥感装备为损毁建筑物信息提取提供了一种有效手段,但面临着多视角特征表达不一、建筑信息不全等问题,提出了一种几何先验约束的多视角场景下的损毁建筑物检测方法。首先,基于长距离高点监测摄像机、无人机以及网络搜集等方式构建多视角场景损毁建筑物实例分割数据集;其次,设计了一种结合阶段感知特征增强和细粒度特征对齐的跨视角域建筑特征同步模型,对不同视角下的建筑物进行微观特征同步;然后,提出了一种基于Canny边缘检测和熵混乱程度的实例分割模型,对损毁建筑物进行精准检测;最后,设计了一种基于垂直/水平视场切割的几何约束的空间映射方法,将目标检测信息映射到地理空间场景中。实验结果表明,与现有方法相比,所提方法对于不同视角下的建筑及损毁建筑检测效果更好,边界框平均精度均值和像素平均精度均值分别达到50.33%和46.69%;在预测边界框与真实边界框的交并比阈值为50%时,所提方法在目标检测和实例分割任务中的平均精度均值分别达到83.10%和81.91%。在空间映射方面,所提方法能够较为精准地将图像检测的损毁建筑映射到真实地理位置,为应急救援实时指挥提供技术支持。

关键词:高点监测;损毁建筑物;多视角场景;特征对齐;空间映射

中图分类号:P237 **文献标识码**:A **收稿日期**:2024-12-04

DOI:10.13203/j.whugis20240297 **文章编号**:1671-8860(2025)06-1088-12

Damaged Building Detection Method in Multi-view Scenes Based on Geometric Prior Constraints

SUI Baikai[1] *CAO Yungang*[1] *CHENG Haibo*[1] *YANG Peng*[1] *ZENG Yahui*[1]
XIE Yakun[1] *ZHU Jun*[1]

1 Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China

Abstract: **Objectives**: Buildings constitute a critical component that is adversely impacted by disasters. The intelligent detection of damaged building information, coupled with the alignment of 2D to 3D scenes, is pivotal in emergency rescue efforts. **Methods**: This paper proposes an innovative method for detecting damaged buildings from multiple perspectives, leveraging geometric prior constraints across various fields of view. First, a large amount of data collected from long-distance high-point monitoring cameras, drones and online gathering are generated by the simulated image generation to construct a multi-view scene dataset for damaged building instance segmentation. This dataset includes data from various perspectives, such as ground-level view, high-point ground view, and low-altitude high-point view. Then, a cross-view domain building feature synchronization model that combines phase-aware feature enhancement and fine-grained feature alignment is designed. This model performs micro-feature synchronization on buildings under different perspectives, and it consists of a multi-branch phase-aware feature enhancement module and a

基金项目:国家重点研发计划(2022YFC3005703)。

第一作者:隋百凯,博士,研究方向为遥感图像智能处理与分析。13012482890@163.com

通信作者:曹云刚,博士,教授。yungang@swjtu.edu.cn

deep-shallow feature synchronization module that aggregates perception and spatial-channel attention. The aim is to enhance deep features rich in semantic information such as building shape and category, as well as deeper features, and to focus on shallow features rich in spatial information such as building edges, textures, and lines. It addresses the spatial and semantic differences brought by features from different perspectives. Subsequently, based on the transformer attention network, an instance segmentation model that takes into account Canny edge detection and entropy disorder is proposed, further enhancing the feature expression of damaged buildings and achieving precise detection of damaged structures. The network consists of three modules: a pixel-level feature extraction module, a transformer attention extraction module, and a detection module. The detection results include both the pixel-level segmentation categories of buildings and the bounding box target detection information for each building. It provides rich data support for subsequent 2D to 3D geographic scene matching and mapping. Finally, a geometric constraint-based position matching method using vertical and horizontal field of view segmentation is designed. This method projects target detection information onto a 3D geographic information scene. It is tailored for cameras rich in angular information, including ground cameras, high-point gimbal cameras, and drone cameras. By leveraging the actual geographic location of cameras and the horizontal and tilt angles of the observed targets, it infers and calculates the real geographic locations of detection targets, and then performs scene matching and mapping based on a 3D model using ray cutting techniques. **Results:** The empirical findings reveal that the proposed method outperforms current approaches in detecting buildings and damaged structures from multiple perspectives. The mean average precision of bounding boxes (bbox_mAP) and mean average precision of segmentation (seg_mAP) achieve 50.33% and 46.69%, respectively. At intersection over union (IoU) threshold of 50%, bbox_mAP50 and seg_mAP50 are 83.10% and 81.91%, respectively, showing a significant enhancement in detection accuracy. In the case studies, high-point long-distance monitoring cameras are used to capture image data, and instance segmentation of damaged and normal buildings is performed to obtain detection results. The detected buildings are then matched to a 3D geographic scene using the proposed geographic matching method, which utilizes vertical and horizontal field of view segmentation and enables the precise matching and mapping of 2D detection results to 3D geographic scene. **Conclusions:** The proposed method can not only effectively improve the detection effect of damaged buildings in the high-point multi-view images, but also accurately match with the 3D geographic scene to provide technical support for the emergency rescue command at the disaster site.

Key words: high-point monitoring; damaged buildings; multi-view scene; feature alignment; spatial mapping

地震、洪水等自然灾害常对城镇造成毁灭性破坏,其中建筑结构往往是受灾最严重的部分。受损建筑的检测与评估过程复杂且耗时,即使针对单体建筑亦不例外。然而,灾害的实际影响范围通常远超单个建筑或局部区域,亟需高效的大范围评估手段。因此,从高点多视角遥感监测数据中自动发现受损建筑物并将其映射到地理场景能够为应急救援决策、规划提供足够的数据信息。高点监测包括地面高点和低空无人机高点,与卫星遥感影像不同,其成像角度、方式不一,监测影像呈现多视角,且不同视角建筑的空间分布特征不同,为高点多视角下影像损毁建筑物的检测与识别带来极大挑战。

损毁检测已有较多研究,学者们从不同角度对其进行了深入探讨[1-3]。现有损毁检测方法主要可分为监督学习与无监督学习两类。随着技术发展,计算机视觉与机器学习方法在结构损伤检测领域得到广泛应用。学者们尝试采用神经网络技术进行建筑物损伤识别,如文献[4]通过神经网络确定未损伤和受损结构,文献[5]采用基于神经网络的方法,通过整合多源监测数据实现建筑物健康状态评估。此外,较多研究聚焦于自然灾害(如地震)造成的建筑损伤检测。地震灾害破坏的建筑外观主要根据4个参数发生变化,即平台和传感器的类型、捕获的图像质量、图像的角度、建筑物损坏的类型[6],因此,基于神经网络所创建模型的性能评价也主要基于这4个参数。

随着科技和人工智能的发展,卷积神经网络(convolutional neural network,CNN)等一些深度学习方法可以自动从受损建筑图像中提取信息特征,被广泛应用于现代彩色航空图像中受损建筑物的自动检测。文献[7]提出了一种基于CNN的海啸灾后区域检测方法,利用海啸前后的航拍图像进行对比分析,在测试集上实现了94%~96%的检测准确率,但研究主要针对完全损毁建筑物的识别,而非部分受损建筑结构的检测;文献[8]提出了一个基于深度卷积神经网络的损毁房屋检测的完备框架,可以使用条件随机场进行改进,该框架在检测单个房屋方面表现良好,但由于卷积操作的局部性,在识别大型建筑物方面表现不佳;文献[9]利用多分辨率CNN在地震等自然灾害后对受损建筑物进行图像分类;文献[10]首先利用地理信息系统提取灾后建筑损伤信息,然后利用改进的CNN对建筑群进行损伤程度分类;文献[11]基于支持向量机算法获取了积石山地震同震滑坡易发性空间分布,同时通过震前震后的高分辨率光学卫星影像,对同震滑坡进行了应急识别,并探讨了地震、地形地貌和人类活动等因素对同震滑坡的影响。综上所述,现有灾后损害检测研究主要基于卫星遥感影像数据,通常仅提供垂直视角的观测信息,这种单一的成像几何特性在一定程度上限制了建筑物损伤检测的精度和可靠性[12-13],导致侧重于建筑顶面信息进行损毁评判,考虑信息不全面,此外,云层干扰的固有缺陷会显著影响灾后建筑物实时损伤监测系统的可靠性。相比之下,利用高点多视角监测手段能够监测到建筑墙面信息,从而实现对墙体裂缝、结构倾斜等损伤特征的精确识别。这种多维度的监测方式不仅扩展了可检测的损伤类型范围,还显著提升了建筑损伤评估的系统性能。在对于建筑墙面信息的识别和提取研究中,文献[14]使用已有的卷积神经网络模型VGG(visual geometry group)进行墙面特征提取,并结合其他方法如受限玻尔兹曼机获取墙面上下文约束,以实现更优的、包含全局特征的墙面像素级分类。随着深度学习的快速发展,建筑墙面检测精度进一步提升,文献[15]结合阶段感知模块和辅助训练策略提出了一种阶段感知特征对齐网络,该网络基于编码器-解码器结构,用于街景的实时语义分割。文献[12]针对街景图像语义分割任务,首先将空洞卷积与引入残差学习单元的深度可分离卷积结合,优化了编码器结构,降低了计算复杂度,同时缓解了梯度消失的问题;然后利用层传递的迭代空洞空间金字塔将自顶向下的特征信息依次融合,提高了上下文信息的有效交互能力。目前研究主要基于单一视角影像,而不同视角的建筑信息特征差别较大,特征对齐较为困难,使得建筑物信息提取困难。对于损毁检测,将建筑墙面破损作为损坏检测也被广泛探讨[16]。文献[17]开发了一种基于立体视觉的裂缝宽度检测方法,利用两台相机恢复裂缝边缘坐标;文献[18]建议使用无人驾驶飞行器自主检测建筑病变,并结合各种图像处理算法进行裂缝检测;文献[19]采用各种深度学习网络进行砖砌体的自动裂缝检测,优于传统的图像处理功能,取得较好的提取效果。

目前,以无人机倾斜摄影视角下的高点视角损毁建筑检测研究较多[20-21],但是地面高点视角下损毁建筑提取研究较少,损毁种类复杂多样,数据集严重缺乏,现有方法无法满足该条件下的损毁建筑检测。针对以上问题,为了满足灾害现场应急救援的需求,本文提出一种基于几何先验约束的多视角损毁建筑物检测方法和技术流程,有效提取灾害现场高点多视角下的建筑及损毁建筑信息。

1 几何先验约束的多视角损毁建筑物检测与映射方法

对图像上损毁建筑的定义与判定如图1所示。本文将其主要归类于3种类型:(1)倒塌建筑:在图像上无法保持建筑形态,多呈现废墟形态,其边界轮廓极其不规则,且在部分图像上倒塌部分伴随烟雾产生;(2)倾斜建筑:在高点监测图像上(含有墙面信息),与正常建筑形成对比,其立面4个边界轮廓均呈现非垂直状态,且顶面与地面非平行状态;(3)破损建筑:主要包括建筑轮廓断裂、表面不同程度的破损等,这些信息通过浅层到深层语义特征表达更为明显。

本文方法的总体技术流程如图2所示。首先,基于高点监测技术获取大量灾害现场损毁建筑视频和影像,并结合网络搜集的影像制作多视角下的损毁建筑实例分割数据集。其次,针对多视角识别难的问题,提出了一种结合阶段感知特征增强和细粒度特征对齐的跨视角域建筑特征同步模型。该模型首先通过阶段感知输入对编码网络生成的不同等级特征进行特征增强,旨在提升模型深层和浅层特征的语义表

达能力;通过聚合感知和空间-通道注意力的深层-浅层特征同步方法,利用细粒度的特征对齐弥合不同建筑物数据域之间的差异,旨在对齐多层特征提升模型在视角差异环境下的建筑物信息提取能力。然后,在损毁建筑检测部分,将Canny边缘检测算法和熵混乱表征模块融入Transformer注意力实例分割模型进行引导和约束,对高点监测下的损毁建筑物进行精准检测。最后,设计一种基于垂直/水平视场切割的几何约束的空间映射方法,将检测的建筑信息映射到三维地理空间场景中,满足应急救援的需求。



(a) 倒塌建筑(废墟)　　(b) 倾斜建筑　　(c) 断裂或表面破损建筑

图1　损毁建筑判定

Fig. 1　Determination of Damaged Buildings

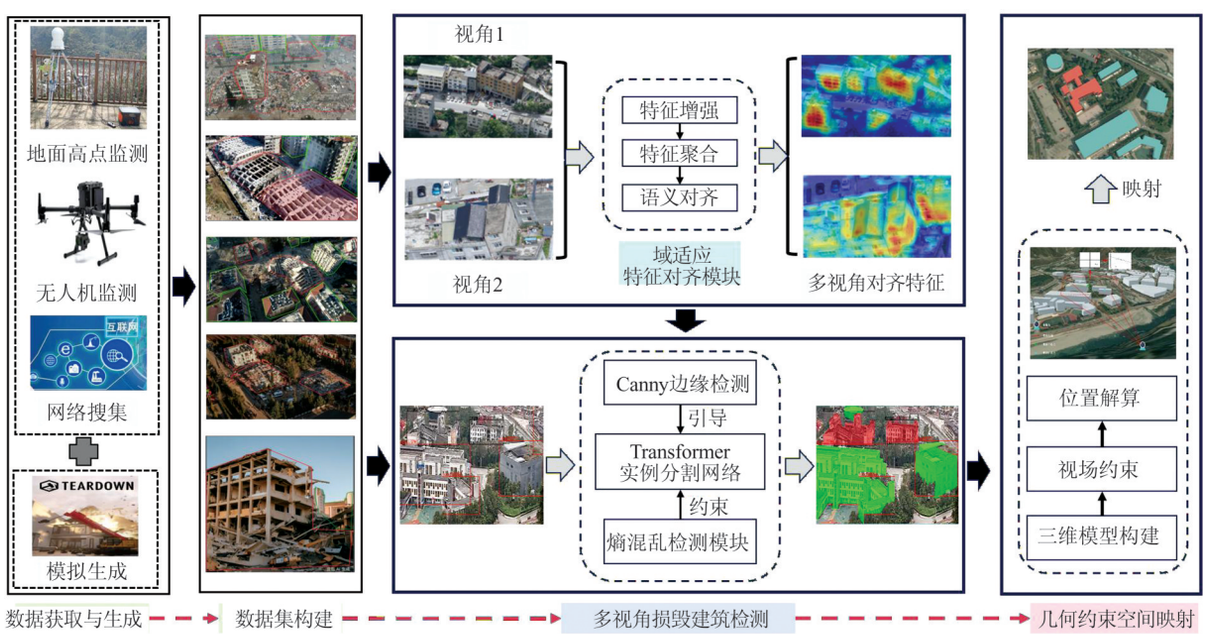


图2　本文方法的总体技术流程

Fig. 2　Overall Technical Process of the Proposed Method

1.1　基于特征增强和细粒度特征对齐的多视角域特征同步模型

针对高点建筑物信息提取中背景复杂和视角差异造成的网络特征表达能力有限的问题,构建基于阶段感知特征增强和细粒度特征对齐的跨视角域建筑特征同步模型,其核心主要包括多分支阶段感知特征增强模块、联合聚合感知和空间-通道注意力的深层-浅层特征同步模块。

多分支阶段感知特征增强模块如图3所示。对于富含建筑物边缘、纹理、线条等空间信息的浅层特征,使用分支1以保留更多的空间信息;对于富含语义信息如建筑物形状和类别的深层特征和较深层特征,使用分支2和分支3进行语义特征的映射和增强表达。浅层特征增强分支由卷积层、批量归一化(batch normalization,BN)层和ReLU激活层组成,深层特征增强分支由深度可分离卷积(depthwise separable convolution,DWConv)层、BN层和ReLU激活层组成。此外,模块中还添加了一个跳跃分支,保持原有的输入特征,避免有用信息丢失。最终,将原始特征、浅层增强特征以及深层增强特征进行特征融合得到的增强后的特征,可大幅提高网络的特征表达能力。

联合聚合感知和空间-通道注意力的深层-浅层特征同步模块如图4所示。首先,将增强后的深层(形状)和浅层(边缘、角点)特征进行聚合感知处理,使得特征的多尺度信息能够充分流动,对聚合感知后的特征进行卷积操作来预

测每个位置的偏移量,使用特征同步函数S对各个位置上的偏移语义完成对齐,实现粗略语义特征同步。然后,粗略语义同步特征与浅层增强特征经过Warp操作进行数据变换,提升特征对齐表达,并与深层特征相加得到中间对齐特征。最后,将得到的特征相加,通过空间-通道注意力模块进一步提升两个维度的特征映射表达能力,实现特征同步。

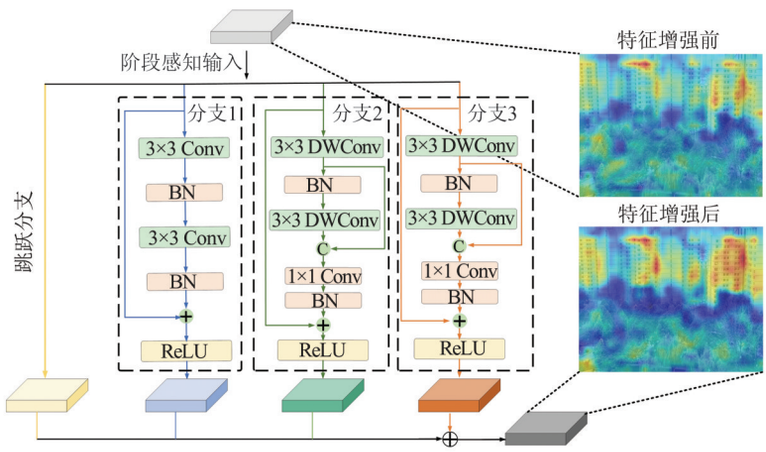


图3 阶段感知特征增强模块

Fig. 3 Stage-Aware Feature Enhancement Module

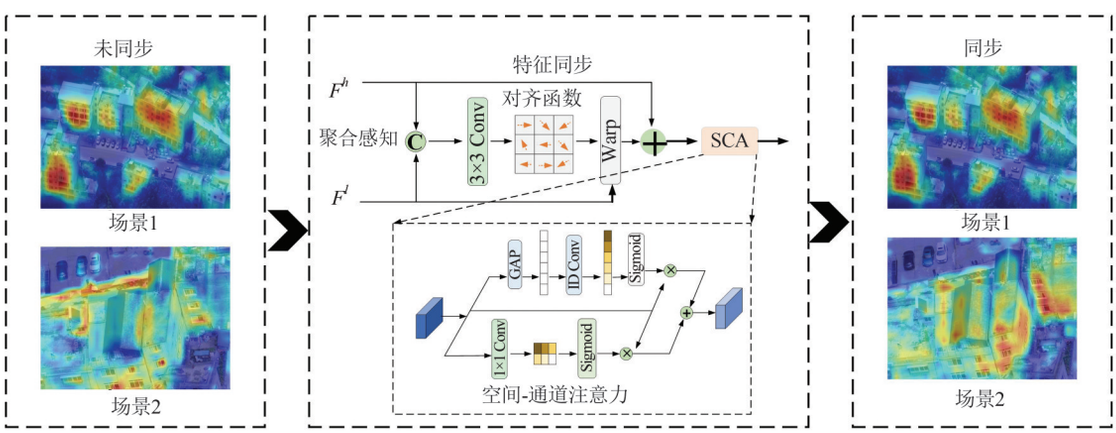


图4 跨视角域特征同步模块

Fig. 4 Cross-View Domain Feature Synchronization Module

该过程的数学计算式为:

$$F_{\text{out}}=S(\text{Conv}_{3\times3}(\text{cat}(\tilde{F}^h,\tilde{F}^l))) \tag{1}$$

$$S=\sum_{h'}^{H}\sum_{w'}^{W}F_{h'}F_{w'}\cdot\max(0,1-|h+\Delta_{1hw}-h'|)\cdot\max(0,1-|w+\Delta_{2hw}-w'|) \tag{2}$$

式中,F_{out}表示对齐后的特征;$\tilde{F}^h$、$\tilde{F}^l$分别为深层特征和浅层特征;$\text{Conv}_{3\times3}$表示3×3的卷积层;cat表示将多个张量沿特定的维度合并;S表示特征同步函数;H、W分别为特征图的长度和宽度;(h,w)、(h',w')分别表示该目标深层、浅层特征的具体位置;Δ_{1hw}、Δ_{2hw}分别表示该位置的二维变换偏移量。

1.2 基于边缘特征和熵混乱程度的损毁建筑检测网络

建筑正常与损毁状态下表达特征会有不同,但是差异并不明显。正常建筑与损毁建筑的边缘特征和熵混乱程度表达如图5所示。由图5可以看出,损坏建筑表面的边缘不规则,其熵特征极其混乱,由此认为建筑损毁部分的边缘和熵混乱程度与正常建筑表面的差异可能会帮助提升检测效果。因此,本文提出了结合建筑边缘细节特征和熵混乱程度的Mask-Transformer实例分割网络。如图6所示,该网络由3个模块组成,分别是像素级特征提取模块、Transformer注意力提取模块以及检测模块。

首先,对齐后的特征经过像素级提取模块,该模块由U-Net编码解码结构、Canny边缘检测分支和熵检测分支组成。对齐后的特征与两个分支得到的边缘特征和熵混乱特征融合,经过编

码网络提取深层特征,再经过解码网络得到逐像素嵌入特征。Canny边缘检测的关键步骤包括噪声降低、梯度计算、非极大值抑制、滞后阈值处理以及边缘跟踪。熵检测分支主要通过图像熵混乱检测函数实现,对输入图像的对齐特征进行计算,得到离散数据分布中事件出现的概率的混乱程度,熵值越高,数据越无序,损毁概率越高。熵混乱检测是一种基于信息熵理论的数据异常检测方法,通过量化数据的混乱程度(不确定性)来识别异常模式。

然后,深层特征还要输入到Transformer注意力模块中的解码网络提取序列特征。使用标准的Mask2Former解码网络[22]从图像特征和N个可学习的位置嵌入(即查询)中计算其输出,即深层影像特征与序列位置编码分别经过掩膜注意力层、自注意力层和前馈网络提取序列掩膜特征,每层后面均需要添加一个相加和归一化操作。

最后,在检测模块中,应用线性分类器在每段嵌入Q上进行Softmax激活,以产生每个段的类别概率预测。对于掩膜预测,由Transformer注意力解码网络输出的序列特征经过多层感知器(multilayer perceptron,MLP),将每段嵌入Q转换为$C\times N$的二维掩膜嵌入特征,与逐像素嵌入进行点积,再通过Sigmoid激活函数得到$N\times H\times W$的预测掩膜,之后与二维预测类别经过点积操作得到最终的实例分割结果,并通过类别损失和二值掩膜损失分别约束类别预测和掩膜预测的参数优化。



图5　建筑与损毁建筑边缘特征和熵混乱程度表达

Fig. 5　Edge Features and Entropy Chaos Expression of Buildings and Damaged Buildings

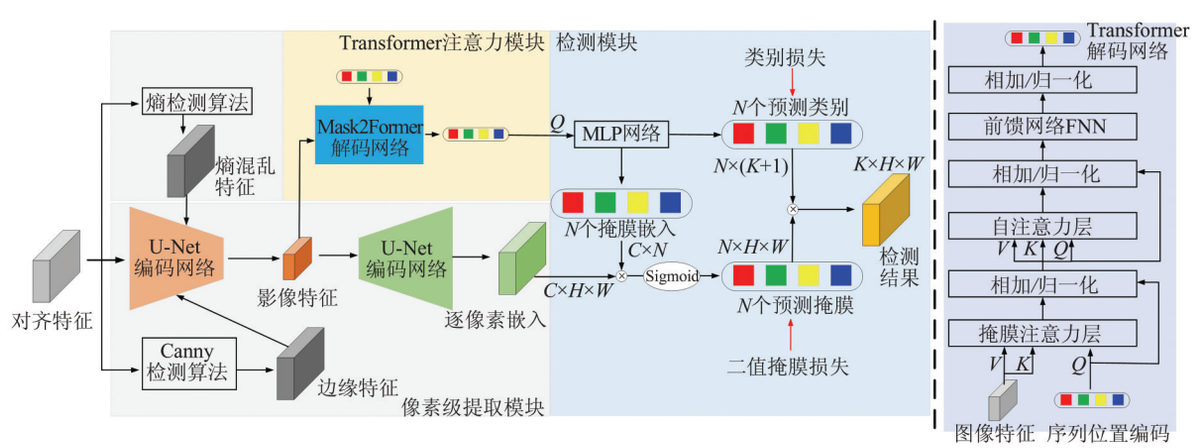


图6　损毁建筑实例分割网络

Fig. 6　Damaged Building Instance Segmentation Network

1.3　基于垂直/水平视场切割的几何约束的空间映射方法

一般来说,摄像机云台传回的水平旋转角α和俯仰角β均为图像中心点所对应的角度,当建筑物检测框的中心位于图像中心时,直接采用云台传回的角度,以摄像机位置为原点建立一条射线,这条射线与建立的建筑物图层的第一个相交点即为此刻检测到的建筑物。当建筑物检测框的中心不在图像中心时,需要重新计算当这一建筑物位于图像中心时所对应的水平旋转角和俯仰角。如图7所示,坐标系以图像左上角为原点,图像宽方向为x轴,图像高方向为y轴,$C(x_1,y_1)$是建筑物检测框中心在坐标系的位置。

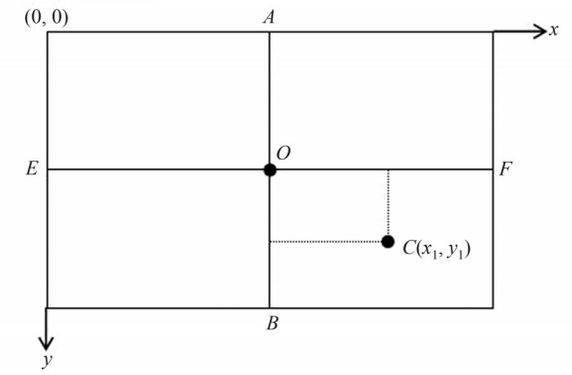
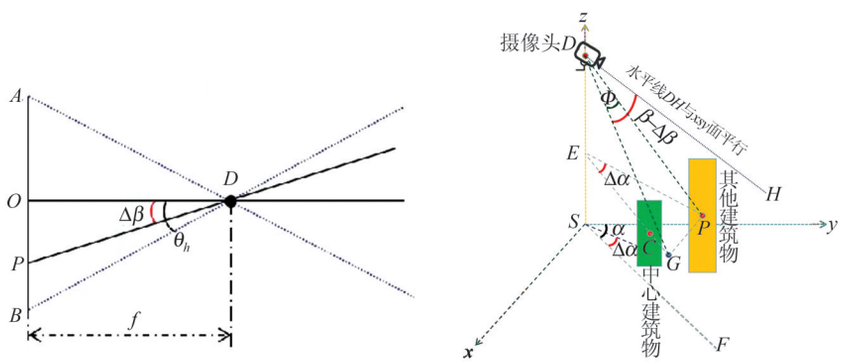


图7 任意像素点在图像坐标系的位置

Fig. 7 Position of Any Pixel in Image Coordinate System

图8为建筑物定位原理图。在图8(a)中,θ_h表示摄像头的垂直视场角的1/2,$\Delta\beta$表示其他建筑物与中心建筑物的俯仰角偏差,A、B分别为图像坐标系中沿y轴方向的最高点、最低点,O为AB线段的中点,f为摄像头的焦距。在图8(b)中,D为摄像头位置;C、P分别表示中心建筑物和邻近其他建筑物的中心,CE平行于SF,与z轴交于E点;β表示云台传回的摄像机的俯仰角;α表示云台传回的摄像机的水平旋转角,$\Delta\alpha$表示其他建筑物与中心建筑物的水平旋转角偏差;Φ表示摄像机光轴沿着水平方向旋转到P点的角度。



(a) 摄像机成像原理图(垂直方向) (b) 建筑物定位原理示意图

图8 建筑物定位原理图

Fig. 8 Schematic Diagram of Building Positioning

根据摄像成像原理以及三角形AOD,可以计算OA的实际高度;利用三角形POD,可以计算OP的实际高度,计算式分别为:

$$OA = f \times \tan\theta_h \tag{3}$$

$$OP = f \times \tan\Delta\beta \tag{4}$$

令h_y表示图像的像素高度,OP与OA的实际高度比值等于像素高度比值,计算式为:

$$\frac{OP}{OA} = \frac{f \times \tan\Delta\beta}{f \times \tan\theta_h} = \frac{\frac{h_y}{2} - y_1}{\frac{h_y}{2}} \tag{5}$$

由式(5)和反三角函数可以求得到俯仰角的偏移角$\Delta\beta$,计算式为:

$$\Delta\beta = \arctan\left(\frac{\frac{h_y}{2} - y_1}{\frac{h_y}{2}} \times \tan\theta_h\right) \tag{6}$$

当$\Delta\beta>0$时,表示其他建筑物在图像中位于中心建筑物的上方,因此需要在中心建筑物俯仰角的基础上减去$\Delta\beta$的绝对值;当$\Delta\beta<0$时,表示其他建筑物在图像中位于中心建筑物的下方,因此需要在中心建筑物俯仰角的基础上加上$\Delta\beta$的绝对值。由此可以求得当其他建筑物位于图像中心时摄像头的俯仰角,同理亦可求得当其他建筑物位于图像中心时摄像头的水平旋转角。

令h_x表示图像的像素宽度,θ_P表示摄像机水平视场角的1/2,摄像机光轴沿着水平方向旋转到P点的角度为Φ,计算式为:

$$\Phi = \arctan\left(\frac{\frac{h_x}{2} - y_1}{\frac{h_x}{2}} \times \tan\theta_P\right) \tag{7}$$

然而,此时求得的Φ并不是实际上的水平旋转角偏差。G为CE延长线上的一点,$PG\perp GE$,因此可以将$\Delta\alpha$表示出来,由于$PG\perp GE$且与z轴垂直,因此$PG\perp GD$,从而可以计算出Φ,同时计算三角形EGD中的GD,联立表达后求出实际水平旋转角的偏移角$\Delta\alpha$,计算式为:

$$\begin{cases} \tan\Delta\alpha = \dfrac{PG}{GE} \\ \tan\Phi = \dfrac{PG}{GD} \\ GD = GE \times \cos\left(\beta - \Delta\beta\right) \end{cases} \tag{8}$$

$$\Delta\alpha=\arctan\left(\frac{\tan\Phi}{\cos(\beta-\Delta\beta)}\right) \tag{9}$$

当$\Delta\alpha>0$时,表示其他建筑物在图像中位于中心建筑物的左边,因此需要在中心建筑物水平旋转角的基础上减去$\Delta\alpha$的绝对值;当$\Delta\alpha<0$时,表示其他建筑物在图像中位于中心建筑物的右边,因此需要在中心建筑物水平旋转角的基础上加上$\Delta\alpha$的绝对值。由此可以求得当其他建筑物位于图像中心时摄像头的水平旋转角。

最后,结合研究区域建筑物高度数据与对应建筑物矢量构建单体建筑物三维模型,通过摄像机所在点位地理坐标以及计算后的俯仰角偏差,得到一条射线与三维模型进行相交判断,判断为真时,则返回相交建筑的三维坐标,该坐标位置即为所计算检测框中心损毁建筑物的真实地理坐标。

2　实验与结果分析

2.1　高点损毁建筑数据集构建

本文利用地面高点监测设备长距离摄像机以及大疆M300无人机倾斜摄影测量采集四川省雅安、汶川、泸定以及消防实验基地等多个地区的建筑和损毁建筑数据,并通过网络搜集多个灾害地区的损毁建筑数据及模拟生成等方式来构建损毁建筑实例分割数据集,如图9所示。本文收集的高点数据集包含损毁建筑图像1 500张,正常建筑图像1 500张。值得注意的是,数据集中每个图像尺寸大小不一,在512×512到3 000×2 000之间。



图9　高点损毁建筑数据集

Fig. 9　High-Point Damaged Buildings Dataset

2.2　评价指标

本文使用像元级评价指标,主要以像元为最小单位,计算正确分类与错误分类的像元个数。使用交并比(intersection over union,IoU)和平均精度均值(mean average precision,mAP)来评价各方法对损毁建筑的提取效果。评价指标的计算式为:

$$\begin{cases} P=\dfrac{N_{TP}}{N_{TP}+N_{FP}}\times 100\% \\ R=\dfrac{N_{TP}}{N_{TP}+N_{FN}}\times 100\% \\ I_{IoU}=\dfrac{P\times R}{P+R-P\times R}\times 100\% \\ P_{mAP}=\dfrac{1}{N}\cdot\sum\limits_{i=1}^{N}\dfrac{1}{R_{max}-R_{min}}\int_{R_{min}}^{R_{max}}P(R)\mathrm{d}R \end{cases} \tag{10}$$

式中,P为精确率;R为召回率;TP(true positive)为真正例,N_{TP}为正样本预测为真的像素个数;FP(false positive)为假正例,N_{FP}为正样本预测为假的像素个数;FN(false negative)为假负例,N_{FN}为负样本预测为假的像素个数;$P(R)$代表在召回率为R时的精度,R_{min}和R_{max}分别为召回率的最小值和最大值;N代表测试样本类别数,i代表第i个类别。

2.3　实验结果与评价

2.3.1　损毁建筑实例分割

为了验证本文所提损毁建筑检测方法的有效性,采用Mask R-CNN[23]、Swin-Mask R-CNN[24]、SCNet[25]和Mask2Former[22]等多种先进的实例分割方法作为对比方法,在多个视角下的高点图像上进行了损毁建筑检测的对比实验,结果如图10所示,其中红色代表损毁建筑,绿色代表正常建筑。

由图10可知,其他对比方法虽然大体上能够提取损毁建筑与正常建筑,但是误检、漏检现象较为严重,均出现较多的混淆现象,即同一个建筑出现多个实例,且像素级识别导致边缘轮廓

粗糙。这些方法在无人机空对地观测视角下对损毁建筑的检测效果显著,但在地面及地面高点视角下的检测性能则相对有限。Mask2Former检测效果相对较好,而本文方法在3种视角下的提取效果均与地面真值相似度最高,误检、漏检减少且实例内部完整性最高。此外,本文还在地面高点视角和无人机高点视角下分别选取3张影像,统计了不同视角下的实例提取个数,结果见表1。由表1可以看出,本文方法提取出的损毁建筑和正常建筑数量与真实数量最为接近,证明了其在高点多视角下提取损毁建筑的有效性。

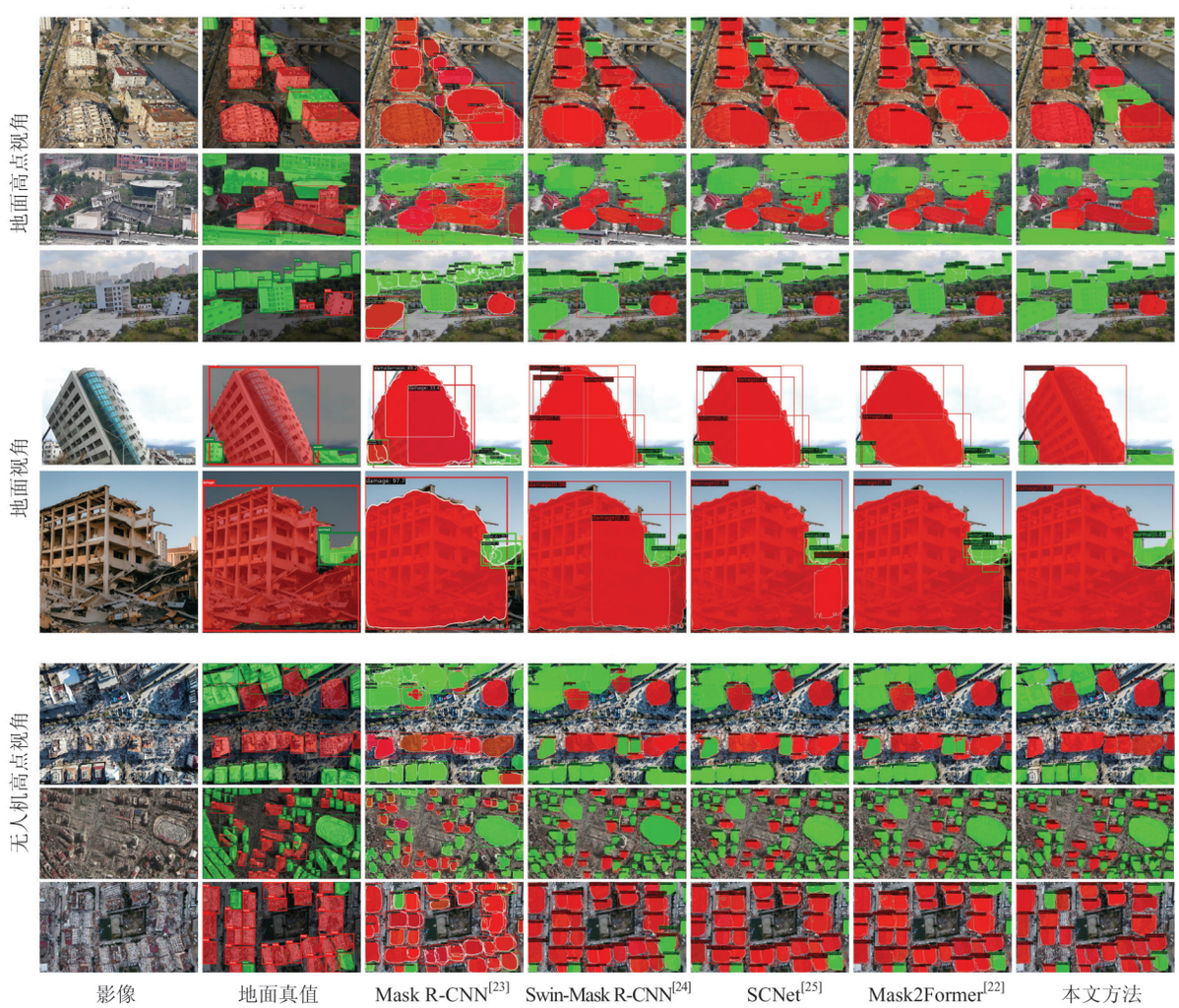


图10 不同方法的损毁与正常建筑物实例分割结果

Fig. 10 Instance Segmentation Results of Damaged and Normal Buildings Using Different Methods

表1 不同方法的损毁与正常建筑提取数量对比

Table 1 Comparison of Number of Damaged and Normal Buildings Extracted Using Different Methods

方法	地面高点视角				无人机高点视角			
	损毁建筑		正常建筑		损毁建筑		正常建筑	
	真实数量	提取数量	真实数量	提取数量	真实数量	提取数量	真实数量	提取数量
Mask R-CNN[23]	15	25	23	43	59	80	75	107
Swin-Mask R-CNN[24]		27		36		73		96
SCNet[25]		23		39		78		101
Mask2Former[22]		22		33		69		92
本文方法		17		24		62		84

为了定量比较本文方法与其他方法之间的差异,在所有视角下统计了边界框平均精度均值bbox_mAP、bbox_mAP50以及像素平均精度seg_mAP、seg_mAP50,其中mAP50指的是在预

测边界框与真实边界框的交并比阈值为50%时的平均精度,它可以提供一个平衡的视角来看待模型的检测能力。不同方法的损毁与正常建筑提取精度结果见表2。由表2可以看出,本文方法在4个评价指标上均得到了最高的得分,相比于Mask R-CNN、Swin-Mask R-CNN、SCNet和Mask2Former,本文方法在4种评价指标上均有明显提升,说明所提特征对齐模块可以较好地适应不同视角下的建筑提取,而且提出的检测网络更针对损毁建筑提取。除此之外,本文方法检测结果中对实例进行了优化,减少了大量重叠冗余,这也使得检测效果得到了较大的提升。

表2　不同方法的损毁与正常建筑提取精度评价/%

Table 2　Evaluation of Extraction Accuracy for Damaged and Normal Buildings Using Different Methods/%

方法	bbox_mAP	bbox_mAP50	seg_mAP	seg_mAP50
Mask R-CNN[23]	40.51	72.66	35.82	68.56
Swin-Mask R-CNN[24]	43.87	76.92	39.23	74.01
SCNet[25]	43.02	75.88	38.27	72.05
Mask2Former[22]	47.52	80.81	43.86	78.33
本文方法	50.33	83.10	46.69	81.91

消融组件包括多视角域泛化特征对齐模块、边缘检测模块和熵混乱检测模块,为了验证损毁检测网络内部各模块的有效性,本文开展了消融实验,结果见表3。由表3可以看出,3个模块的加入均能有效提升高点视角下损毁建筑检测的精度,证明所提方法在损毁建筑检测方面的有效性。

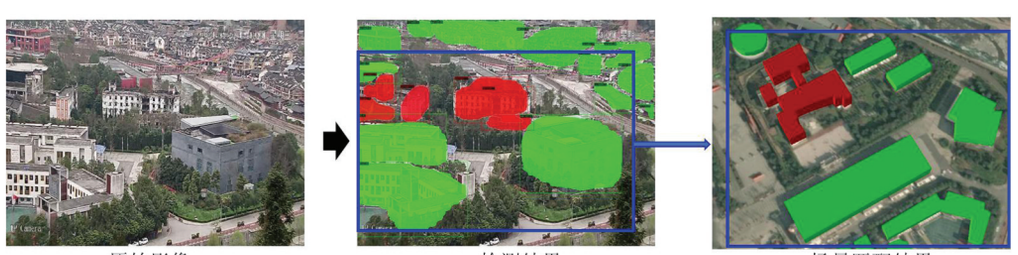
表3　不同模块消融实验精度/%

Table 3　Accuracy of Ablation Experiments with Different Modules/%

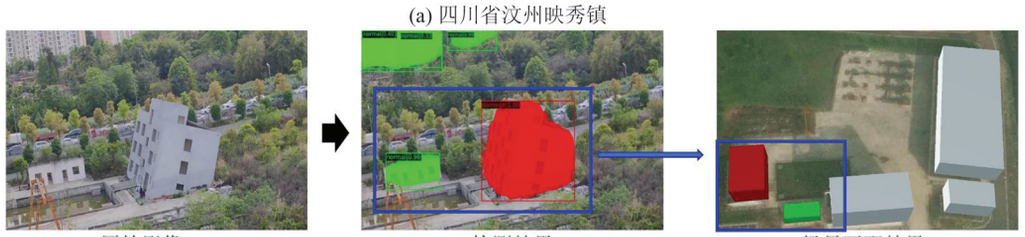
特征对齐模块	边缘检测模块	熵混乱检测模块	bbox_mAP50	seg_mAP50
			80.27	78.15
√			81.59	80.06
√	√		82.25	81.66
√	√	√	83.10	81.91

2.3.2　地面高点视角建筑空间几何映射

本文通过野外调查收集数据,并选择四川省汶川县映秀镇和四川省防灾减灾技术实验服务基地两处典型区域进行实验与验证,通过现场高点长距离监测摄像机获取的影像数据进行损毁与正常建筑实例分割,得到检测结果,该检测结果无空间坐标。因此,通过本文所提基于垂直和水平视场切割的几何约束的空间映射方法将检测到的建筑映射到三维空间地理场景,如图11所示,其中绿色表示正常建筑,红色表示损毁建筑。由图11可以看出,本文方法可以有效地将检测到的所有建筑精准映射到地理空间场景中,适用于带有水平角和俯仰角信息参数的地面高点和无人机高点监测摄像机获取的现场数据。



(a) 四川省汶州映秀镇



(b) 四川省防灾减灾技术服务基地

图11　四川省典型区域的建筑匹配结果

Fig. 11　Building Matching Results at Typical Areas in Sichuan Province

3 结 语

本文研究了地面/低空高点多视角影像的损毁建筑信息提取,并提出了一种多视场几何先验约束的多视角场景下的损毁建筑物检测方法。利用长距离高点监测摄像机、无人机、网络搜集以及图像生成等方式构建多视角场景(损毁)建筑物实例分割数据集;设计了一种结合阶段感知特征增强和细粒度特征对齐的跨视角域建筑特征同步模型,解决多视角场景下建筑物特征差异的问题;设计了一种基于Canny边缘检测和熵混乱程度的实例分割模型,提取高点影像下损毁/正常建筑物目标框坐标和像素信息;设计了一种基于垂直/水平视场切割的几何约束的空间映射方法,将建筑检测信息映射到地理空间场景中。与Mask R-CNN、Swin-Mask R-CNN、SCNet和Mask2Former方法相比,本文方法对于不同视角下的建筑及损毁建筑检测效果更好,边界框平均精度均值bbox_mAP和像素平均精度均值seg_mAP分别达到了50.33%和46.69%,在预测边界框与真实边界框的交并比阈值为50%时,bbox_mAP50和seg_mAP50分别达到83.10%和81.91%。本文方法不仅能有效提升高点多视角影像下的损毁建筑检测效果,而且能够精准与三维地理场景进行几何映射,获取检测损毁建筑的真实地理位置,为灾害现场应急救援指挥提供技术支持。

参 考 文 献

[1] RUPNIK E, NEX F, TOSCHI I, et al. Contextual Classification Using Photometry and Elevation Data for Damage Detection After an Earthquake Event [J]. *European Journal of Remote Sensing*, 2018, 51(1): 543-557.

[2] DUBOIS D, LEPAGE R. Fast and Efficient Evaluation of Building Damage from very High Resolution Optical Satellite Images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(10): 4167-4176.

[3] LU C H, NI C F, CHANG C P, et al. Coherence Difference Analysis of Sentinel-1 SAR Interferogram to Identify Earthquake-Induced Disasters in Urban Areas [J]. *Remote Sensing*, 2018, 10(8): 1318.

[4] WU X, GHABOUSSI J, GARRETT J H. Use of Neural Networks in Detection of Structural Damage [J]. *Computers & Structures*, 1992, 42(4): 649-659.

[5] FANG X, LUO H, TANG J. Structural Damage Detection Using Neural Network with Learning Rate Improvement[J]. *Computers & Structures*, 2005, 83(25/26): 2150-2161.

[6] KERLE N, NEX F, DUARTE D, et al. UAV-Based Structural Damage Mapping - Results from 6 Years of Research in Two European Projects[J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, XLII-3/W8: 187-194.

[7] FUJITA A, SAKURADA K, IMAIZUMI T, et al. Damage Detection from Aerial Images via Convolutional Neural Networks[C]//The 15th IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 2017.

[8] VOLPI M, TUIA D. Dense Semantic Labeling of Subdecimeter Resolution Images with Convolutional Neural Networks [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(2): 881-893.

[9] DUARTE D, NEX F, KERLE N, et al. Multi-resolution Feature Fusion for Image Classification of Building Damages with Convolutional Neural Networks[J]. *Remote Sensing*, 2018, 10: 1636.

[10] MA H J, LIU Y L, REN Y H, et al. Improved CNN Classification Method for Groups of Buildings Damaged by Earthquake, Based on High Resolution Remote Sensing Images [J]. *Remote Sensing*, 2020, 12(2): 260.

[11] 陈博, 宋闯, 陈毅, 等. 2023年甘肃积石山Ms 6.2地震同震滑坡和建筑物损毁情况应急识别与影响因素研究[J]. 武汉大学学报(信息科学版), 2025, 50(2): 322-332.
CHEN Bo, SONG Chuang, CHEN Yi, et al. Emergency Identification and Influencing Factor Analysis of Coseismic Landslides and Building Damages Induced by the 2023 Ms 6.2 Jishishan (Gansu, China) Earthquake[J]. *Geomatics and Information Science of Wuhan University*, 2025, 50(2): 322-332.

[12] KERLE N, HOFFMAN R R. Collaborative Damage Mapping for Emergency Response: The Role of Cognitive Systems Engineering[J]. *Natural Hazards and Earth System Sciences*, 2013, 13(1): 97-113.

[13] KERLE N. Satellite-Based Damage Mapping Following the 2006 Indonesia Earthquake—How Accurate Was It?[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2010, 12(6): 466-476.

[14] FATHALLA R, VOGIATZIS G. A Deep Learning Pipeline for Semantic Facade Segmentation [C]//British Machine Vision Conference, London, UK, 2017.

[15] WENG X, YAN Y, CHEN S, et al. Stage-Aware Feature Alignment Network for Real-Time Semantic Segmentation of Street Scenes[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4444-4459.

[16] MOHAN A, POOBAL S. Crack Detection Using Image Processing: A Critical Review and Analysis [J]. *Alexandria Engineering Journal*, 2018, 57(2): 787-798.

[17] SHAN B H, ZHENG S J, OU J P. A Stereovision-Based Crack Width Detection Approach for Concrete Surface Assessment[J]. *KSCE Journal of Civil Engineering*, 2016, 20(2): 803-812.

[18] PEREIRA F C, PEREIRA C E. Embedded Image Processing Systems for Automatic Recognition of Cracks Using UAVs [J]. *IFAC—PapersOnLine*, 2015, 48(10): 16-21.

[19] LOVERDOS D, SARHOSIS V. Automatic Image-Based Brick Segmentation and Crack Detection of Masonry Walls Using Machine Learning [J]. *Automation in Construction*, 2022, 140: 104389.

[20] 涂继辉, 眭海刚, 吕枘蓬, 等. 基于基尼系数的倾斜航空影像中建筑物立面损毁检测[J]. 武汉大学学报(信息科学版), 2017, 42(12): 1744-1748.
TU Jihui, SUI Haigang, LÜ Ruipeng, et al. Building Facade Damage Detection Based on the Gini Index from Oblique Aerial Images[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(12): 1744-1748.

[21] 眭海刚, 黄立洪, 刘超贤. 利用具有注意力的Mask R-CNN检测震害建筑物立面损毁[J]. 武汉大学学报(信息科学版), 2020, 45(11): 1660-1668.
SUI Haigang, HUANG Lihong, LIU Chaoxian. Detecting Building Facade Damage Caused by Earthquake Using CBAM-Improved Mask R-CNN[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(11): 1660-1668.

[22] CHENG B W, MISRA I, SCHWING A G, et al. Masked-Attention Mask Transformer for Universal Image Segmentation [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022.

[23] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.

[24] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows[C]//IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021.

[25] LIU J J, HOU Q B, CHENG M M, et al. Improving Convolutional Networks with Self-Calibrated Convolutions [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020.