



武汉大学学报(信息科学版)

*Geomatics and Information Science of Wuhan University*

ISSN 1671-8860, CN 42-1676/TN

## 《武汉大学学报(信息科学版)》网络首发论文

题目: 联合全局-局部特征和动态错误监督的遥感影像建筑物损伤评估方法  
作者: 刘梓航, 刘欣怡, 张永军  
DOI: 10.13203/j.whugis20240346  
收稿日期: 2024-11-18  
网络首发日期: 2024-11-29  
引用格式: 刘梓航, 刘欣怡, 张永军. 联合全局-局部特征和动态错误监督的遥感影像建筑物损伤评估方法[J/OL]. 武汉大学学报(信息科学版).  
<https://doi.org/10.13203/j.whugis20240346>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20240346

### 引用格式：

刘梓航, 刘欣怡, 张永军. 联合全局-局部特征和动态错误监督的遥感影像建筑物损伤评估方法[J]. 武汉大学学报(信息科学版), 2024, DOI:10.13203/J.whugis20240346 (LIU Zihang, LIU Xinyi, ZHANG Yongjun. Building Damage Assessment from Satellite Images Combining Global-Local Features and Dynamic Error Supervision[J]. Geomatics and Information Science of Wuhan University, 2024, DOI:10.13203/J.whugis20240346)

## 联合全局-局部特征和动态错误监督的遥感影像建筑物损伤评估方法

刘梓航<sup>1</sup> 刘欣怡<sup>1</sup> 张永军<sup>1</sup>

<sup>1</sup> 武汉大学遥感信息工程学院, 湖北 武汉, 430079

**摘要：**在灾害发生后,快速准确地评估灾害区域的范围和严重程度对于后续的救援和重建至关重要。目前针对遥感影像的深度学习建筑物损伤评估方法面临特征差异建模不足、全局-局部特征利用不充分和困难样本感知能力缺乏等问题。为此,提出一种基于全局-局部特征融合和动态错误监督网络(global-local feature fusion and dynamic error supervision network, GLENet)的双时相遥感影像建筑物损伤评估方法。在编码阶段,采用共享权重的编码器提取双时相影像特征,将双时相影像特征送入差异增强融合模块(difference enhancement fusion module, DEFM)增强特征间差异并获取融合特征;在解码阶段,融合特征先后经过全局-局部特征融合模块(global-local feature fusion module, GLFFM)和动态错误感知解码器(dynamic error aware decoder, DEAD)并输出评估结果,以实现兼顾全局-局部特征的融合解码和困难样本感知的学习。在目前最大的全球级别建筑物损伤评估高分辨率遥感影像数据集xBD上进行实验,GLENet取得了建筑提取F1分数86.03%,损伤分类F1分数75.20%,综合评价F1分数78.45%的结果,总体指标优于多个对比方法。在Ida-BD和LEVIR-CD数据集上进行了迁移实验和变化检测实验,验证了GLENet的泛化性和不同任务适用性。

**关键词：**建筑物损伤评估; 变化检测; 孪生网络; 特征融合; 遥感影像

## Building Damage Assessment from Satellite Images Combining Global-Local Features and Dynamic Error Supervision

LIU Zihang<sup>1</sup> LIU Xinyi<sup>1</sup> ZHANG Yongjun<sup>1</sup>

<sup>1</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

**Abstract: Objectives:** After a disaster, it is essential to quickly and accurately assess the extent and severity of the disaster area for subsequent humanitarian relief and reconstruction. Traditional damage assessment methods are constrained by

收稿日期: 2024-11-18

项目资助: 国家自然科学基金(42192581, 42201474)。

第一作者: 刘梓航, 硕士生, 研究方向为地表异常遥感检测与评估。zhliu2022@whu.edu.cn。

通讯作者: 刘欣怡, 博士, 副研究员。liuxy0319@whu.edu.cn。

time efficiency, labor cost, and accessibility. In contrast, satellite images can quickly obtain the real situation of a wide range of disaster areas, and gradually become an important data source for building damage assessment. Automated building damage assessment from satellite images relies on deep learning methods, but current deep learning building damage assessment methods for satellite images face challenges such as insufficient modeling of feature differences, inadequate utilization of global-local features, and lack of difficult sample perception ability. **Methods:** To address these problems, a building damage assessment method based on global-local feature fusion and dynamic error supervision network (GLESNet) is proposed. At the encoding stage, the dual-temporal image features were extracted by a shared weight backbone, and the features were sent to the difference enhancement fusion module (DEFM) to enhance the difference between the features, filter out spurious changes, and obtain the fusion features. At the decoding stage, the fusion features are passed by the vertical and horizontal global-local feature fusion modules (GLFFM) and the dynamic error aware decoder (DEAD), to fuse the global and local features and percept the difficult samples. **Results:** The proposed GLESNet achieves 86.03% F1-score of building extraction, 75.20% F1-score of damage classification, and 78.45% overall F1-score on xBD, the largest global level high-resolution satellite image dataset for building damage assessment. **Conclusions:** The quantitative evaluation and visualization results are better than other advanced comparison methods. Ablation study verifies the effectiveness of each module. Transfer experiments and change detection experiments carried out on the IdABD and LEVIR-CD datasets verify the generalization of the proposed GLESNet to different data and tasks.

**Key words:** building damage assessment, change detection, siamese network, feature fusion, satellite imagery

地震、台风、洪水、火灾等自然和人为灾害严重威胁人们的生命财产安全<sup>[1]</sup>。建筑物作为一种重要承灾体，其倒塌是灾害中人员伤亡的重要原因之一，评估灾后建筑物的损伤程度对受灾区域确定、灾害应急响应、灾后重建与灾情评估具有重要意义<sup>[2]</sup>。传统的评估过程主要基于地面调查进行，可获得较为详细的损伤信息，但时间周期长、人力物力成本高<sup>[3]</sup>，具有一定的风险，并且在交通不便的情况下数据获取将会受限<sup>[4]</sup>，无法满足灾害应急等任务的需求。随着遥感技术的发展，高分辨率遥感卫星可在短时间内获得灾区大范围的影像数据，反映真实的地表信息<sup>[5]</sup>，已逐步应用于受灾区域确定和灾害应急响应<sup>[6]</sup>，为高效的大范围建筑物损伤评估提供了新的手段。因此，基于高分辨率卫星遥感影像的建筑物损伤评估方法具有广阔的应用前景和重要的研究价值。

传统的遥感影像建筑物损伤评估方法主要基于颜色纹理等特征和机器学习技术进行<sup>[7-9]</sup>。这些方法主要考虑建筑物本身的变化，集中于倒塌和未倒塌两种类别，产生的评估结果较为粗糙，受损建筑物中受到周围环境影响的部分难以准确提取，如被洪水部分淹没的建筑物。此外，这些方法往往只适用于单一的灾害类型，限制于特定的地理区域，难以迁移应用<sup>[10]</sup>。近年来，深度学习的蓬勃发展大大推进了建筑物损伤评估方法的研究。基于深度学习技术，已经探索了多种类别的建筑物损伤评估方法，按照所使用的数据，可分为仅使用灾后影像的方法和同时使用灾前灾后双时相影像的方法，按照所使用的评估技术，可分为基于对象级图像分类的方法和基于像素级变化检测的方法。其中，仅使用灾后影像的方法<sup>[11]</sup>难以获取建筑物的精确轮廓及变化特征，基于图像分类的方法<sup>[12]</sup>丢失了大量场景信息，且容易受到双时相影像偏移的影响，因此效果不佳。目前大部分方法将损伤评估视为一对多变化检测任务<sup>[13]</sup>，使用双时相影像进行像素级评估。将损伤评估分解为建筑提取和损伤分类两个子任务，根据两个子任务的耦合程度，又可将这些方法分为三类：（1）构建两个独立的模型，分两阶段进行建筑提取和损伤分类<sup>[14]</sup>；（2）将两个子任务合并到同一模型中，通过共享编码器和独立解码器分别进行建筑

提取和损伤分类<sup>[13]</sup>；(3) 不对两个子任务进行区分，采用孪生网络架构同时进行建筑提取和损伤分类<sup>[15]</sup>。这三类方法性能相近，但基于孪生网络架构的方法可实现端到端的输出，更加灵活高效。

为了进一步提升损伤评估方法的性能，近期提出的方法从双时相特征关系建模、建筑物位置先验引导、类别不平衡和语义模糊性缓解等角度进行了改进。文献[16]提出了 BDANet，针对两阶段方法特征交互不够充分的问题，开发了交叉关注模块，实现双时相特征间的通信。文献[17]在第二阶段引入分组洗牌注意力对双时相特征进行交换。文献[13]提出了 ChangeOS，通过灾前影像获取深度对象特征指导损伤分类，并进行对象化后处理。文献[18]将建筑物定位信息引入损伤分类阶段，通过建筑物位置先验约束上下文信息，实现可解释的损伤分类增强。文献[19]提出了 Dual-HRNet，并通过改进损失函数缓解类别不平衡问题。文献[20]通过对比学习缓解相邻损伤类别的语义模糊性，提高损伤评估性能。

损伤评估需要综合建筑物本身的外观和建筑物周围的环境变化，因此部分方法结合从局部到全局的多尺度特征以提升性能。文献[21]基于卷积网络，采用带膨胀卷积的 ResNet<sup>[22]</sup>骨干和空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 模块<sup>[23]</sup>获得多尺度特征。然而卷积的感受野有限<sup>[24]</sup>，难以提取全局特征，因此后续研究引入了自注意力机制。文献[25]提出了 DAHiTra，基于双时相差异特征，利用自注意力模块，捕获受损建筑物间的远程信息和不同局部区域间的多层次变化关系。文献[26]提出了 DamFormer，采用 SegFormer 骨干提取特征以获取全局上下文，显著提升了损伤分类效果。

现有的损伤评估方法主要通过特征交互以优化双时相特征，但交互过程中主要考虑特征之间的互补关系，对特征差异的建模不足。此外，不同建筑物样本的难易程度有较大差异，现有方法欠缺对困难样本的感知能力，难以进一步提高模型效果。而在全局和局部特征结合方面，现有方法仅使用卷积或自注意力中的一种，或仅在网络的某一特定位置使用自注意力，未能充分结合两种机制的优势。造成该问题的重要原因之一是，原始的自注意力机制具有二次复杂度，内存和计算开销高昂，一些改进的自注意力模型如 Swin Transformer<sup>[27]</sup>将内存和计算开销减小到可接受的范围，但也削弱了全局特征提取能力。相比之下，状态空间模型 (state space models, SSM) 具有线性复杂度下的全局特征提取能力，近年来受到研究者的高度关注。文献[28]改进了原始的 SSM，提出了选择性扫描 (selective scan, SS)，并基于此构建了 Mamba 架构，该架构在多项自然语言处理任务中表现出优于经典自注意力模型的效果。文献[29]将选择性扫描拓展到图像，提出了二维选择性扫描 (2D selective scanning, SS2D)，其衍生的模型在自然图像和遥感影像处理任务<sup>[30-31]</sup>中取得了优异表现，显示出在遥感领域的巨大应用潜力。因此，有必要考虑使用 SSM 替代自注意力机制，以实现更加充分的全局特征提取。

综合以上分析，针对以往建筑物损伤评估模型特征差异建模不足、全局-局部特征利用不充分以及困难样本感知能力缺乏等问题，提出了一种基于全局-局部特征融合和动态错误监督网络 (global-local feature fusion and dynamic error supervision network, GLENet) 的遥感影像建筑物损伤评估方法。设计差异增强融合模块 (difference enhancement fusion module, DEFM) 增强双时相特征的差异；引入二维选择性扫描，将其以串联和并联两种方式与卷积相结合，构建全局-局部特征融合模块 (global-local feature fusion module, GLFFM) 以增强全局-局部特征的利用；动态错误感知解码器 (dynamic error aware decoder, DEAD) 通过动态监督感知错分区域，激发网络对困难样本的关注。各模块之间紧密配合，最终获得良好的损伤评估结果。



# 1 基本原理

## 1.1 模型整体框架

本文提出的全局-局部特征融合和动态错误监督网络采用孪生网络架构，输入配对的双时相影像，最终输出损伤评估结果，其整体结构如图 1 所示。网络可分为四个单元，编码阶段包含权重共享编码器和特征融合单元，解码阶段包含全局-局部融合解码器和动态错误感知解码器。权重共享编码器采用 ResNet-34 骨干，具有 5 个编码阶段，双时相影像分别经过编码器获得多层次特征，权重共享能将双时相影像映射到同一特征空间内，同时减少网络的计算量和参数量。特征融合单元由差异增强融合模块组成，编码器产生的双时相特征被送入特征融合单元内，增强双时相特征的差异并融合。全局-局部融合解码器位于解码阶段的深层，由全局-局部特征融合模块组成，特征融合单元产生的差异增强融合特征首先被送入全局-局部融合解码器中，充分利用全局和局部特征提取高级语义信息。动态错误感知解码器位于解码阶段的中层和浅层，由跨层融合解码模块 (cross-layer fusion and decoding module, CFDM) 组成，全局-局部融合解码器输出的特征被送入动态错误感知解码器中，以困难样本感知的方式学习，通过两个子分支分别输出辅助的损伤评估和错误预测结果，将两个子分支末端跨层融合解码模块输出的特征进行通道叠加，并通过一个额外的跨层融合解码模块融合，输出最终的损伤评估结果。

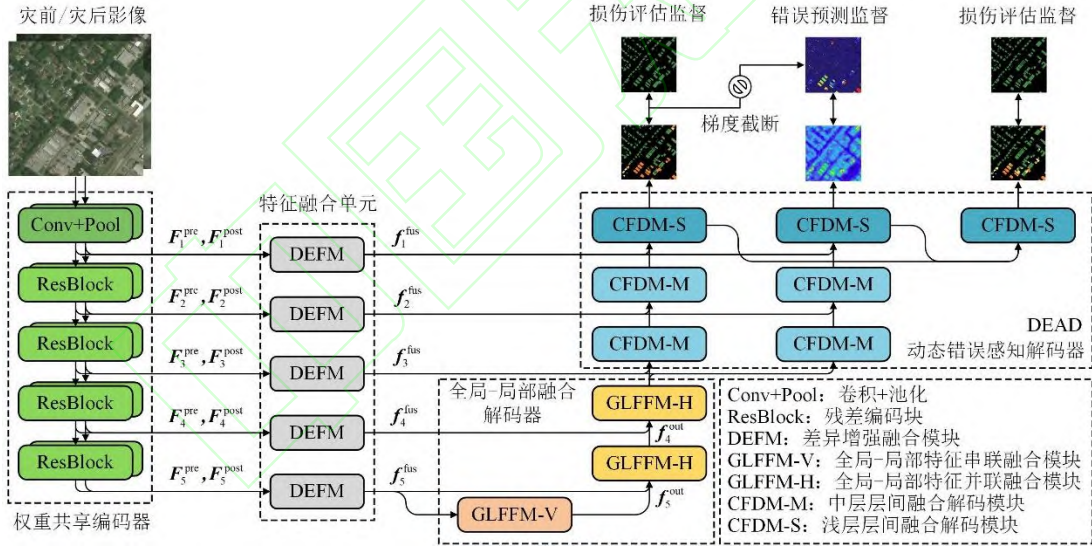


图 1 全局-局部特征融合和动态错误监督网络

Fig. 1 Global-Local Feature Fusion and Dynamic Error Supervision Network

## 1.2 差异增强融合模块

由于损伤评估任务需要深入挖掘影像的变化特征，因此差异增强融合模块通过注意力的方式对双时相特征进行增强，放大有效差异，剔除伪变化造成的差异，使得网络关注关键变化区域，过滤背景噪声引起的伪变化。将编码器获得的双时相特征分别记为  $F_i^{pre}$ ,  $F_i^{post} \in \mathbf{R}^{H_i \times W_i \times C_i}$ ,  $\{i=1,2,3,4,5\}$ , 其中  $H_i = H / 2^i$ ,  $W_i = W / 2^i$ ,  $H$  和  $W$  分别为输入影像的长和宽,  $C_i$  为通道数,  $C_1 = 64$ ,  $C_i = C_1 \cdot 2^{i-2}$ ,  $\{i=2,3,4,5\}$ 。二者相减并取绝对值获得差分特征  $F_i^{sub}$ , 通过通道注意力获得其通道注意力特征  $Ac_i^{sub} \in \mathbf{R}^{1 \times 1 \times C_i}$ , 对  $F_i^{pre}$  和  $F_i^{post}$  也按相同的过程获取通道注意力特征  $Ac_i^{pre}$ ,  $Ac_i^{post} \in \mathbf{R}^{1 \times 1 \times C_i}$ 。这里的通道注意力不包含 Sigmoid 函数:

$$\mathbf{Ac} = \text{MLP}(\text{GAP}(\mathbf{F})) + \text{MLP}(\text{GMP}(\mathbf{F})) \quad (1)$$

上式中  $\mathbf{F}$  代表输入特征， $\text{MLP}(\cdot)$  代表共享参数的多层感知机， $\text{GAP}(\cdot)$  代表全局平均池化， $\text{GMP}(\cdot)$  代表全局最大池化， $\mathbf{Ac}$  为获得的通道注意力特征。将  $\mathbf{Ac}_i^{\text{pre}}$ ， $\mathbf{Ac}_i^{\text{post}}$  和  $\mathbf{Ac}_i^{\text{sub}}$  由  $1 \times 1 \times C_i$  转置为  $1 \times C_i \times 1$ ，延通道维度叠加，并通过  $1 \times 7$  卷积自适应融合，获得差分增强通道注意力特征  $\mathbf{Ac}_i^{\text{pre}'}$  和  $\mathbf{Ac}_i^{\text{post}'}$ ：

$$\mathbf{Ac}_i^{\text{pre}'} = \sigma(\text{Conv}_{1 \times 7}([\mathbf{Ac}_i^{\text{pre}}]^T, [\mathbf{Ac}_i^{\text{sub}}]^T)) \quad (2)$$

$$\mathbf{Ac}_i^{\text{post}'} = \sigma(\text{Conv}_{1 \times 7}([\mathbf{Ac}_i^{\text{post}}]^T, [\mathbf{Ac}_i^{\text{sub}}]^T)) \quad (3)$$

式中  $\text{Conv}_{1 \times 7}(\cdot)$  代表  $1 \times 7$  卷积， $[\cdot, \cdot]$  代表延通道维度叠加， $\sigma(\cdot)$  代表 Sigmoid 函数。将获得的差分增强通道注意力特征  $\mathbf{Ac}_i^{\text{pre}'}$  和  $\mathbf{Ac}_i^{\text{post}'}$  转置回  $1 \times 1 \times C_i$ ，并与原始特征相乘，为了保持原始特征信息，利用残差连接的方式生成通道差分增强的特征  $\mathbf{Fc}_i^{\text{pre}}$  和  $\mathbf{Fc}_i^{\text{post}}$ ， $\otimes$  为逐元素相乘：

$$\mathbf{Fc}_i^{\text{pre}} = \mathbf{F}_i^{\text{pre}} \otimes (\mathbf{Ac}_i^{\text{pre}'})^T + \mathbf{F}_i^{\text{pre}} \quad (4)$$

$$\mathbf{Fc}_i^{\text{post}} = \mathbf{F}_i^{\text{post}} \otimes (\mathbf{Ac}_i^{\text{post}'})^T + \mathbf{F}_i^{\text{post}} \quad (5)$$

空间差异增强与通道差异增强类似，首先将通道差分增强的双时相特征  $\mathbf{Fc}_i^{\text{pre}}$  和  $\mathbf{Fc}_i^{\text{post}}$  相减并取绝对值，获得差分特征  $\mathbf{Fc}_i^{\text{sub}}$ ，通过空间注意力获得  $\mathbf{Fc}_i^{\text{pre}}$ ， $\mathbf{Fc}_i^{\text{post}}$  和  $\mathbf{Fc}_i^{\text{sub}}$  的空间注意力特征  $\mathbf{As}_i^{\text{pre}}$ ， $\mathbf{As}_i^{\text{post}}$  和  $\mathbf{As}_i^{\text{sub}} \in \mathbf{R}^{H_i \times W_i \times 1}$ ，这里的空间注意力过程如下所示， $\mathbf{As}$  为获得的空间注意力特征：

$$\mathbf{As} = \text{Conv}_{7 \times 7}([\text{GAP}(\mathbf{F}), \text{GMP}(\mathbf{F})]) \quad (6)$$

将  $\mathbf{As}_i^{\text{sub}}$  与  $\mathbf{As}_i^{\text{pre}}$  和  $\mathbf{As}_i^{\text{post}}$  延通道维度叠加，通过  $3 \times 3$  卷积和 Sigmoid 函数生成差分增强空间注意力特征  $\mathbf{As}_i^{\text{pre}'}$  和  $\mathbf{As}_i^{\text{post}'}$ ，并通过残差连接方式生成空间差分增强的特征  $\mathbf{Fs}_i^{\text{pre}}$  和  $\mathbf{Fs}_i^{\text{post}}$ ，整个过程如下：

$$\mathbf{As}_i^{\text{pre}'} = \sigma(\text{Conv}_{3 \times 3}([\mathbf{As}_i^{\text{pre}}, \mathbf{As}_i^{\text{sub}}])) \quad (7)$$

$$\mathbf{As}_i^{\text{post}'} = \sigma(\text{Conv}_{3 \times 3}([\mathbf{As}_i^{\text{post}}, \mathbf{As}_i^{\text{sub}}])) \quad (8)$$

$$\mathbf{Fs}_i^{\text{pre}} = \mathbf{Fc}_i^{\text{pre}} \otimes \mathbf{As}_i^{\text{pre}'} + \mathbf{Fc}_i^{\text{pre}} \quad (9)$$

$$\mathbf{Fs}_i^{\text{post}} = \mathbf{Fc}_i^{\text{post}} \otimes \mathbf{As}_i^{\text{post}'} + \mathbf{Fc}_i^{\text{post}} \quad (10)$$

最终，为了充分保持双时相特征中的原始信息，将  $\mathbf{Fs}_i^{\text{pre}}$  和  $\mathbf{Fs}_i^{\text{post}}$  进行通道叠加，通过两个连续的卷积层获得最终的融合特征：

$$\mathbf{f}_i^{\text{inter}} = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}([\mathbf{Fs}_i^{\text{pre}}, \mathbf{Fs}_i^{\text{post}}]))) \quad (11)$$

$$\mathbf{f}_i^{\text{fus}} = \text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(\mathbf{f}_i^{\text{inter}}))) \quad (12)$$

式中  $\text{BN}(\cdot)$  代表批归一化， $\text{ReLU}(\cdot)$  代表 ReLU 激活，通过  $1 \times 1$  卷积充分交互双时相特征，获得中间特征  $\mathbf{f}_i^{\text{inter}} \in \mathbf{R}^{H_i \times W_i \times 2C_i}$ ，并通过  $3 \times 3$  卷积进行维度压缩，最终输出融合特征  $\mathbf{f}_i^{\text{fus}} \in \mathbf{R}^{H_i \times W_i \times C_i}$ 。

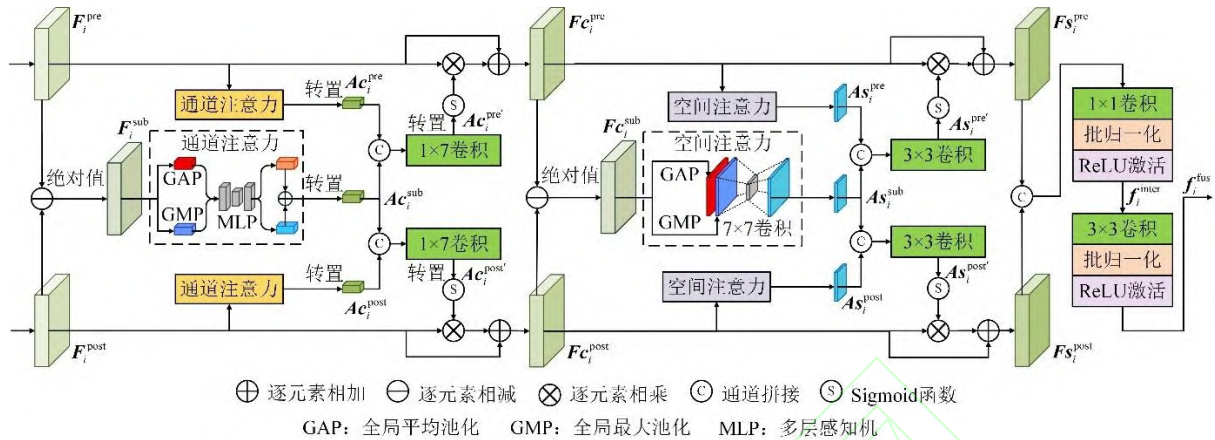


图2 差异增强融合模块

Fig. 2 G Difference Enhancement Fusion Module

### 1.3 全局-局部特征融合模块

建筑物损伤程度的评估同时依赖全局和局部线索，全局线索主要指建筑物周围的环境变化，包括建筑物周围大范围区域的地物类别和状态变化，局部线索主要指建筑物本身的外观变化，如建筑物的结构破坏、顶面材料剥落等。此外，建筑物损伤程度的评估还需要尽量充分地综合各项线索，因此还依赖综合性更强的高级语义特征，以应对评估任务的复杂性。基于以上考虑，使用全局-局部特征融合模块分别提取全局特征和局部特征并融合，以挖掘并协同全局和局部线索，同时将模块放置于解码器的深层，增强网络的高级语义提取能力，以提升网络对更复杂、模糊的伪变化情形的适应能力。在全局-局部特征融合模块中，全局特征和局部特征分别通过选择性扫描块（selection scan block, SSB）和分层多尺度卷积块（hierarchical multi-scale convolutional block, HMCB）两个核心模块提取。

#### 1.3.1 模块总体结构

根据选择扫描块和分层多尺度卷积块的结合方式，全局-局部特征融合模块可分为全局-局部特征并联融合模块（GLFFM-H）和全局-局部特征串联融合模块（GLFFM-V）。对于最深层特征  $f_5^{fus}$ ，首先通过 GLFFM-V 模块，生成语义增强的最深层特征  $f_5^{out}$ ，然后 GLFFM-H 模块利用  $f_5^{out}$  与  $f_5^{fus}$ ，以全局和局部协同的方式提取语义信息。两个模块的结构图如图 3 所示。

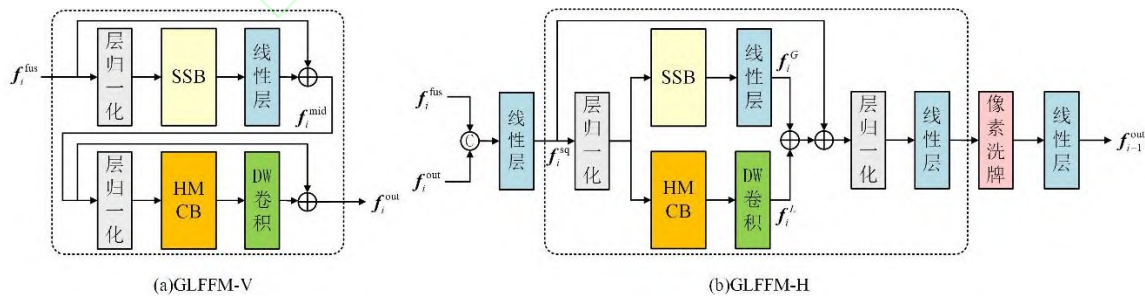


图3 全局-局部特征融合模块

Fig. 3 Global-Local Feature Fusion Module

GLFFM-V 设置在网络最深层，输入为  $f_i^{fus}$ ，输出为  $f_i^{out} \in \mathbf{R}^{H_i \times W_i \times C_i}$ ， $i=5$ 。SSB 通过中间特征  $f_i^{mid} \in \mathbf{R}^{H_i \times W_i \times C_i}$  与 HMCB 串联，引入残差结构改善梯度传播，并通过线性层和 DW 卷积增强模块衔接。

GLFFM-V 的计算过程如下，其中  $\text{Lin}(\cdot)$  为线性层， $\text{DWConv}(\cdot)$  为 DW 卷积， $\text{LN}(\cdot)$  为层归一化：

$$f_i^{\text{mid}} = \text{Lin}(\text{SSB}(\text{LN}(f_i^{\text{fus}}))) + f_i^{\text{fus}} \quad (13)$$

$$f_i^{\text{out}} = \text{DWConv}(\text{HMCB}(\text{LN}(f_i^{\text{mid}}))) + f_i^{\text{mid}} \quad (14)$$

GLFFM-H 的输入为  $f_i^{\text{fus}}$  和  $f_i^{\text{out}}$ ，输出为  $f_{i-1}^{\text{out}}$ ， $i \in \{4, 5\}$ 。首先通过线性层进行特征融合并压缩，获得压缩特征  $f_i^{\text{sq}} \in \mathbf{R}^{H_i \times W_i \times C_i}$ ，将  $f_i^{\text{sq}}$  归一化并分别送入 SSB 和 HCMB 分支，获得全局特征  $f_i^G$  和局部特征  $f_i^L$ ，两类特征分别通过线性层和 DW 卷积层进行调整并相加，并引入残差连接结构。相加后的特征通过归一化层和线性层进行深度融合，通过像素洗牌和线性层进行上采样和维度变换，减少上采样过程中语义信息的损失并获得输出特征  $f_{i-1}^{\text{out}}$ 。GLFFM-H 的计算过程如下， $\text{PS}(\cdot)$  为像素洗牌操作：

$$f_i^{\text{sq}} = \text{Lin}([f_i^{\text{fus}}, f_i^{\text{out}}]) \quad (15)$$

$$f_i^G = \text{Lin}(\text{SSB}(\text{LN}(f_i^{\text{sq}}))) \quad (16)$$

$$f_i^L = \text{DWConv}(\text{HMCB}(\text{LN}(f_i^{\text{sq}}))) \quad (17)$$

$$f_{i-1}^{\text{out}} = \text{Lin}(\text{PS}(\text{Lin}(\text{LN}(f_i^G + f_i^L))) + f_i^{\text{sq}}) \quad (18)$$

### 1.3.2 选择性扫描块

选择性扫描块通过选择性扫描获取全局信息，选择性扫描块的结构如图 4 上部所示。选择性扫描块的核心是 SS2D，遵循文献[29]中的做法，将输入特征  $X_i^{\text{in}} \in \{\text{LN}(f_i^{\text{mid}}), \text{LN}(f_i^{\text{sq}})\}$ ， $i \in \{4, 5\}$  输入两个分支，一个分支通过线性层获得门控特征  $Z_i$ ，另一个分支通过 SS2D 获得全局特征  $Y_i$ ，二者通过逐元素相乘获得输出  $X_i^{\text{out}} \in \mathbf{R}^{H_i \times W_i \times C_i}$ ，由于线性层欠缺空间信息的提取能力，而 SS2D 对通道信息的提取能力相对较弱<sup>[30]</sup>，因此加入了空间和通道注意力进行补充，主要过程如下：

$$Z_i = \text{SA}(\text{SiLU}(\text{Lin}(X_i^{\text{in}}))) \quad (19)$$

$$Y_i = \text{LN}(\text{SS2D}(\text{SiLU}(\text{DWConv}(\text{Lin}(X_i^{\text{in}})))))) \quad (20)$$

$$X_i^{\text{out}} = \text{CA}(Z_i \otimes Y_i) \quad (21)$$

式中  $\text{SiLU}(\cdot)$  代表 SiLU 激活， $\text{CA}(\cdot)$  和  $\text{SA}(\cdot)$  分别代表完整的通道和空间注意力。SS2D 的主要流程如图 4 下部所示，首先将二维特征沿纵横两个方向正反展开为一维序列，并在四个一维序列上分别应用一维选择性扫描，收集不同方向上的全局上下文。最后，将扫描后的一维序列折叠回二维并相加，以获得全局性特征。

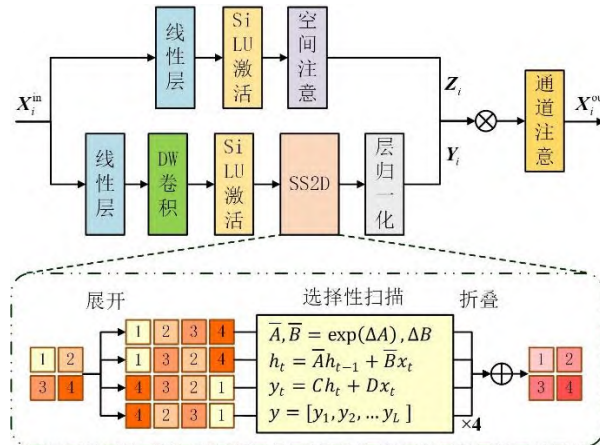


图 4 选择性扫描块



Fig. 4 Selection Scan Block

### 1.3.3 分层多尺度卷积块

分层多尺度卷积块采用分层多尺度卷积代替原始的  $3 \times 3$  卷积，通过分层卷积进一步增强卷积的多尺度性能，通过引入多尺度特征减小局部特征与全局特征间的差异，缓解结合时由于两种特征差异过大所带来的混乱，分层多尺度卷积块的结构如图 5 所示。

分层多尺度卷积块包含两个支路，多尺度卷积支路和  $1 \times 1$  卷积支路。多尺度卷积支路负责收集多尺度特征，侧重空间维度。将输入特征  $\mathbf{x}_i^{\text{in}} \in \{\text{LN}(\mathbf{f}_i^{\text{mid}}), \text{LN}(\mathbf{f}_i^{\text{sq}})\}$ ,  $i \in \{4, 5\}$  延通道维度均分为四份  $\{\mathbf{y}_i^j\}$ ,  $j \in \{1, 2, 3, 4\}$ ，对每份特征以不同膨胀率进行卷积，并加入跨通道连接，增强模块的跨通道和尺度交互能力，分层卷积得到的特征  $\{z_i^j\}$ ,  $j \in \{1, 2, 3, 4\}$  沿通道维度叠加，然后通过  $1 \times 1$  卷积进行通道整合，获得输出特征  $\mathbf{x}_i^{3 \times 3}$ 。公式如下：

$$\mathbf{z}_i^j = \text{Conv}_{3 \times 3}^{d=j}(\mathbf{z}_i^{j-1} + \mathbf{y}_i^j + \mathbf{y}_i^{j+1}) \quad (22)$$

$$\mathbf{x}_i^{3 \times 3} = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}([\mathbf{z}_i^1, \mathbf{z}_i^2, \mathbf{z}_i^3, \mathbf{z}_i^4]))) \quad (23)$$

式中， $\text{Conv}_{3 \times 3}^{d=j}(\cdot)$  代表膨胀率为  $j$  的  $3 \times 3$  卷积。 $1 \times 1$  卷积支路负责收集局部特征，侧重通道维度。 $1 \times 1$  卷积支路的特征  $\mathbf{x}_i^{1 \times 1}$  与多尺度卷积支路的输出相加，得到最终输出特征  $\mathbf{x}_i^{\text{out}} \in \mathbf{R}^{H_i \times W_i \times C_i}$ ：

$$\mathbf{x}_i^{1 \times 1} = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(\mathbf{x}_i^{\text{in}}))) \quad (24)$$

$$\mathbf{x}_i^{\text{out}} = \mathbf{x}_i^{3 \times 3} + \mathbf{x}_i^{1 \times 1} \quad (25)$$

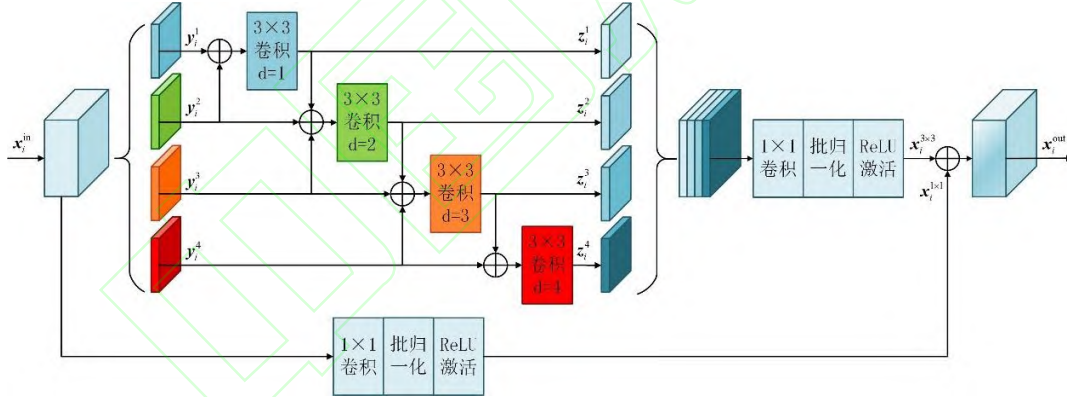


图 5 分层多尺度卷积块

Fig. 5 Hierarchical Multi-scale Convolutional Block

### 1.4 动态错误感知解码器

如图 6 所示，动态错误感知解码器具有主分支和副分支两个子分支，将 GLFFM-H 的输出  $\mathbf{f}_3^{\text{out}}$  分别输入两个子分支中，通过跳跃连接结合编码特征  $\mathbf{f}_1^{\text{fus}}$ ,  $\mathbf{f}_2^{\text{fus}}$ ,  $\mathbf{f}_3^{\text{fus}}$ ，分别完成损伤评估任务和错误预测任务。错误预测标签根据主分支的损伤评估预测获得，并作为副分支的监督，使副分支学习主分支出现的错误预测，进行错误样本识别的任务，从而增强网络对于易错困难样本的感知能力，对于由伪变化影响导致的困难样本，网络在学习过程中会进行额外的关注，以减少伪变化造成的影响。副分支的监督信号由主分支预测动态生成，副分支在训练过程中始终对主分支的状态进行动态追踪。动态错误感知解码器的基本单元是 CFDM，根据模块的位置，CFDM 可分为中层层间融合解码模块 (CFDM-

M) 和浅层层间融合解码模块 (CFDM-S)。CFDM-M 包含两个连续的卷积块 (3×3 卷积+批归一化+ReLU 激活), 第一个块将通道叠加的层间特征进行融合, 第二个块则对融合后特征进行语义提取。CFDM-S 将第二个块改为 3×3 卷积以输出结果。CFDM-M 和 CFDM-S 的区别如图 6 右侧所示。

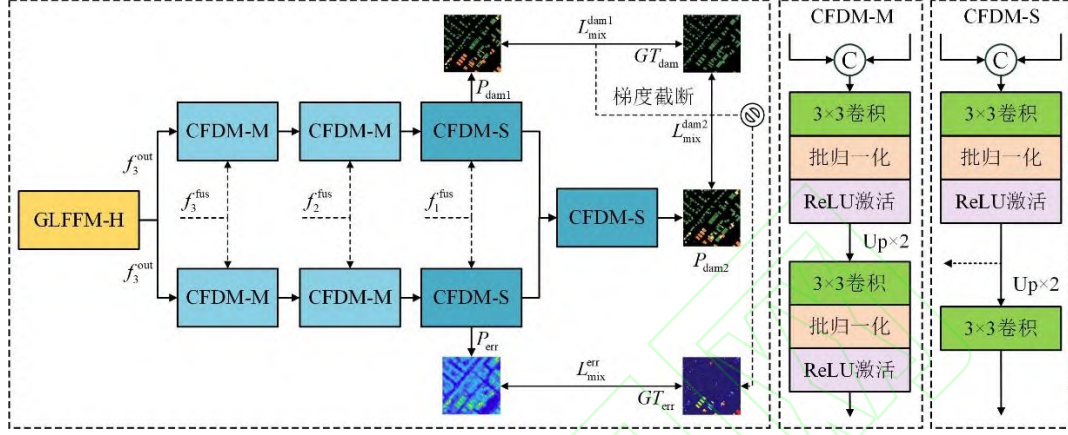


图 6 动态错误感知解码器

Fig. 6 Dynamic Error Aware Decoder

动态错误感知解码器在两个子分支末尾分别输出一个损伤评估结果  $P_{dam1}$  和一个错误预测结果  $P_{err}$ ，对损伤评估真值  $GT_{dam}$  和  $P_{dam1}$  作差并取绝对值，归一化后作为错误预测的真值  $GT_{err}$ ，计算过程如下式所示。作差过程进行梯度截断，防止副分支的梯度传播干扰主分支的训练。

$$GT_{err} = \lambda \cdot \text{SUM}(|GT_{dam1} - \delta(P_{dam1})|) \quad (26)$$

式中  $P_{dam1}$ ， $GT_{dam1} \in \mathbf{R}^{H \times W \times C_{cls}}$ ， $GT_{err} \in \mathbf{R}^{H \times W \times 1}$ ， $C_{cls}$  为总输出类别数。SUM( $\cdot$ ) 代表延通道维度相加， $\delta(\cdot)$  代表延通道维度的 Softmax 函数， $\lambda$  为归一化系数，这里设置为 0.5。两个子分支末尾 CFDM-S 模块输出的特征延通道叠加后送入一层额外的 CFDM-S 模块中，结合错误预测特征进一步细化，并输出最终的损伤评估结果  $P_{dam2}$ 。最终网络由两个损伤评估损失和一个错误预测损失联合优化：

$$L_{mix}^{all} = L_{mix}^{dam1} + L_{mix}^{dam2} + L_{mix}^{err} \quad (27)$$

式中  $L_{mix}^{dam1}$  和  $L_{mix}^{dam2}$  为两个损伤评估损失， $L_{mix}^{err}$  为错误评估损失。 $L_{mix}$  为类不平衡损失函数，该函数以交叉熵损失  $L_{ce}$  为主要损失，为了缓解类别不平衡问题，又加入了 Lovasz 损失<sup>[32]</sup>  $L_{Lovasz}$  作为修正：

$$L_{mix} = L_{ce} + L_{Lovasz} \quad (28)$$

## 2 实验结果与分析

### 2.1 实验数据

本文使用 xBD 数据集<sup>[12]</sup>开展对比和消融实验，使用 Ida-BD 数据集<sup>[33]</sup>开展迁移实验，并在 LEVIR-CD 数据集<sup>[34]</sup>上进行了变化检测实验，以充分验证所提 GLENet 的综合性能。

xBD 数据集是目前规模最大的建筑物损伤评估数据集，包含全球范围内 19 个灾害事件的 11034 对灾前和灾后影像，覆盖了 45361.79 km<sup>2</sup> 范围内的 850736 个建筑物实例，涵盖了地震、洪水、野火、海啸等常见的灾害类型。影像大小为 1024 × 1024 像素，空间分辨率为 0.8m。包含无损伤、轻微损伤、中度损伤和严重损伤四种建筑物标签。其训练/验证/测试集分别包含 9168/933/933 对影像。

Ida-BD 数据集提供了 2021 年 8 月飓风 Ida 期间路易斯安那州新奥尔良部分区域的 87 对灾前和灾后影像。影像大小为  $1024 \times 1024$  像素，空间分辨率为 0.5 m。建筑物标签按 xBD 数据集标准标注为无损伤、轻微损伤、中度损伤和严重损伤四种类别。Ida-BD 数据集专为测试损伤评估方法的域适应能力开发，如表 1 所示，其类分布与 xBD 数据集有较大差异。

LEVIR-CD 数据集包含 637 对影像，影像大小为  $1024 \times 1024$  像素，空间分辨率为 0.5m，影像对之间的时间跨度为 5 至 14 年不等。该数据集中有约 31000 个建筑物实例，包括仓库、车库和公寓大楼等各种建筑物类型。其训练/验证/测试集分别包含 445/64/128 对影像。

表 1 xBD 数据集与 Ida-BD 数据集中每类像素的数量分布/%  
Tab. 1 Class-wise Pixels Count Distribution in xBD Dataset and Ida-BD Dataset/%

| 数据集    | 背景   | 无损伤  | 轻微损伤 | 中度损伤 | 严重损伤 |
|--------|------|------|------|------|------|
| xBD    | 96.1 | 2.7  | 0.1  | 0.1  | 0.1  |
| Ida-BD | 81.7 | 11.9 | 4.6  | 1.6  | 0.05 |

## 2.2 实验设置

本文实验在 64 位 Ubuntu20.04 环境下进行，CPU 型号为 Intel(R) Xeon(R) Gold 6258R CPU @ 2.70GHz，内存 128GB，显卡为单个 GeForce GTX 3090，显存 24GB。网络采用 Pytorch 框架实现。对于 xBD 数据集，将影像裁剪为  $512 \times 512$  像素输入，批量大小为 8，并训练 50 轮，训练期间使用 Adam 优化器，初始学习率 0.0001，并采用多步学习率衰减策略，单步衰减率为 0.5。迁移实验时以同样配置在 Ida-BD 数据集上微调 25 轮。对于 LEVIR-CD 数据集，将影像裁剪为  $256 \times 256$  像素输入，批量大小为 8，并训练 200 轮，训练期间使用 Adam 优化器，初始学习率 0.01，采用线性学习率衰减策略。所有训练和测试均使用官方的数据集划分进行，并在训练过程中通过随机旋转、翻转、缩放、高斯模糊等方法进行数据增强。

## 2.3 评价指标

为了更好地与其他模型进行比较，本文采用 xBD 数据集官方评价指标。该指标采用建筑提取 F1 分数评价建筑提取结果，各损伤类别 F1 分数的调和均值评价损伤分类结果。总体 F1 分数通过建筑提取 F1 分数和损伤分类 F1 分数的加权和计算，由于更关注损伤分类结果，因此给予损伤分类 F1 分数更高的权重。各类 F1 分数定义如下：

$$F1_b = \frac{2TP}{2TP + FP + FN} \quad (29)$$

$$F1_d = \frac{m}{\sum_{n=1}^m 1/F1_{c_n}} \quad (30)$$

$$F1_s = 0.3 \times F1_b + 0.7 \times F1_d \quad (31)$$

式中  $F1_b$ 、 $F1_d$  和  $F1_s$  分别为建筑提取、损伤分类和总体 F1 分数。TP、FP、FN 分别为真阳性、假阳性、假阴性像素数， $F1_{c_n}$  为各损伤类别 F1 分数， $m$  为总类别数。F1 分数同时考虑了准确率和召回率，能够更好地评估类别不平衡时的建筑提取与损伤分类效果。

## 2.4 对比实验

为了验证 GLEsNet 在损伤评估任务上的优势,在 xBD 数据集上开展了对比实验,选择 Siam-UNet<sup>[14]</sup>、ChangeOS<sup>[13]</sup>、DamFormer<sup>[26]</sup>、Dual-HRNet<sup>[19]</sup>、DAHiTra<sup>[25]</sup> 和 BDANet<sup>[16]</sup>和六种先进的损伤评估方法作为对比。表 2 显示了在 xBD 数据集上 GLEsNet 与其他先进方法结果的定量对比,图 7 显示了与其他先进方法的可视化对比。由表 2 可以看到,GLEsNet 在总体 F1 和损伤分类 F1 上取得了最佳的结果,在总体结果上相对第二名提升 1.26%。而在建筑提取上,由于 Siam-UNet 和 BDANet 是二阶段方法,通过专门的建筑提取网络,使用灾前图像提取建筑物,因此 F1 比 GLEsNet 高出 0.06%和 0.26%。但在更重要的损伤分类结果上,GLEsNet 相对第二名提升了 1.91%。在各损伤级别上,GLEsNet 在轻微损伤和严重损伤类别上优势明显,领先其他方法 1.62%和 0.96%,在中度损伤类别上优于其他方法 0.22%,在无损伤类别上,也取得了仅次于 BDANet 的结果。

表 2 xBD 数据集对比实验定量评价结果/%  
Tab. 2 Quantitative Evaluation Results of Comparative Experiments on xBD dataset/%

| 网络模型       | 总体F1         | 建筑提取F1       | 损伤分类F1       | 各损伤级别F1      |              |              |              |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|            |              |              |              | 无损伤          | 轻微损伤         | 中度损伤         | 严重损伤         |
| Siam-UNet  | 75.13        | 86.09        | 70.44        | 93.71        | 51.23        | 70.95        | 79.99        |
| ChangeOS   | 75.50        | 85.69        | 71.14        | 89.11        | 53.11        | 72.45        | 80.79        |
| DamFormer  | 76.64        | 85.69        | 72.76        | 93.98        | 54.22        | 72.83        | 82.25        |
| Dual-HRNet | 76.76        | 85.85        | 72.69        | 94.42        | 54.81        | 73.66        | 80.10        |
| DAHiTra    | 77.10        | 85.91        | 73.31        | 92.90        | 56.28        | 72.54        | 81.64        |
| BDANet     | 77.19        | <b>86.29</b> | 73.29        | <b>94.96</b> | 54.50        | 72.67        | 83.75        |
| GLEsNet    | <b>78.45</b> | 86.03        | <b>75.20</b> | 94.49        | <b>57.90</b> | <b>73.88</b> | <b>84.71</b> |

图 7 展示了对比实验的部分定性可视化结果。在第一行中,大部分对比方法由于提取全局信息和高级语义的能力不足,无法在建筑物内部取得一致性的结果,将部分建筑物由中度损伤误判为轻微损伤。此外,除 GLEsNet 外,所有对比方法都在建筑提取上产生了一定误判,将附近的停车场判断为建筑物,也体现了 GLEsNet 在提取高级语义特征方面的优越性。对于第二行左下角位于森林区域的建筑物,建筑物周围的水体被遮挡,但由于 GLEsNet 能够更加充分地充分利用长距离上下文,因此能够正确区分遮挡建筑物的损伤类别。对于第三行左下部的小目标建筑物,由于 GLEsNet 能够同时获取局部特征,因此能够区分出小目标建筑物的类别,并且由于更加关注困难样本,对右上角其他方法均无法正确区分的建筑物,也能得到正确结果。对于第四行的大范围粘连建筑物,GLEsNet 整合了全局到局部的各个尺度的特征,并采用差异增强模块关注变化区域,因此能够正确判断出各个损伤类别的大致范围,在此基础上准确提取出各个类别间的边界。对于第五行中多个类别混合的情况,GLEsNet 也可以较为准确地提取出其边界,并保证各类别间的平滑过渡,在复杂场景下仍然具有较好的判别能力。



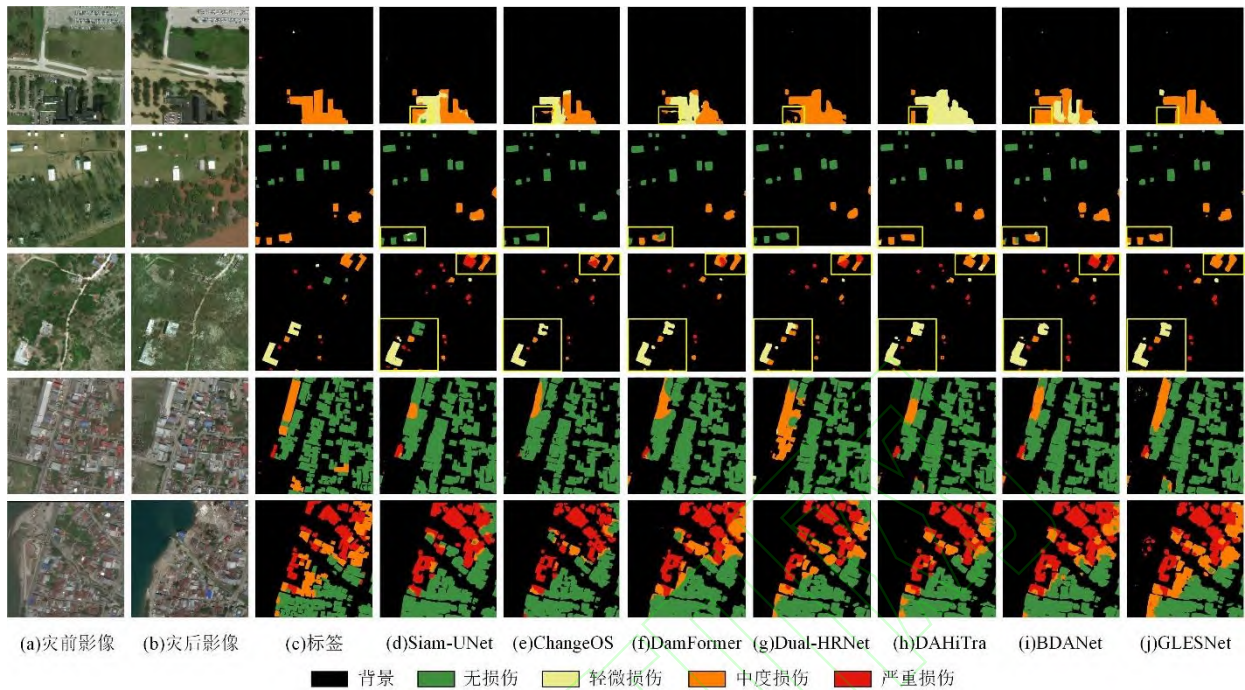


图 7 xBD 数据集对比实验可视化结果  
 Fig.7 Visualization Results of Comparative Experiments on the xBD Dataset

## 2.5 消融实验

为了进一步验证所提模块在 GLEsNet 中的有效性，对网络中的关键模块进行消融实验和分析，消融实验的设置如表 3 所示。模型 1 为基础网络，使用常规卷积代替 DFEM 模块和 GLFFM-H 模块，并去除了额外的错误预测分支和 GLFFM-V 模块，其他结构参数与 GLEsNet 一致。模型 2 采用 DFEM 进行特征融合，使得总体 F1 增加 0.69%，损伤分类 F1 增加 1.01%，尤其是轻微损伤和严重损伤类别有了较大的提升，分别达 2.26%和 1.02%。模型 3 和模型 4 逐步加入了 GLFFM-H 和 GLFFM-V 模块，分别较模型 2 在总体 F1 上提升 0.8%和 1.38%，在损伤分类上提升 1.09%和 1.93%。并在无损伤和轻微损伤方面提升显著，在无损伤方面提升 0.95%和 0.85%，在轻微损伤方面提升 2.31%和 3.89%。完整的 GLEsNet 进一步加入了 DEAD，较模型 4 总体 F1 提升 0.89%，损伤分类 F1 提升 1.14%，且在建筑提取 F1 上提升 0.29%。完整的 GLEsNet 一定程度上克服了语义模糊问题，增强了中度损伤和严重损伤类别的 F1，提升幅度达 1.95%和 1.61%。总的来说，各个模块的加入使得网络获得了稳定的性能提升，完整的 GLEsNet 补充了所有消融模型的不足，取得了最为均衡的结果。

表 3 消融实验设置  
 Tab. 3 Ablation Experimental Setup

| 网络模型      | Base | DFEM | GLFFM-H | GLFFM-V | DEAD |
|-----------|------|------|---------|---------|------|
| GLEsNet-1 | √    |      |         |         |      |
| GLEsNet-2 | √    | √    |         |         |      |
| GLEsNet-3 | √    | √    | √       |         |      |
| GLEsNet-4 | √    | √    | √       | √       |      |

GLESNet    √    √    √    √    √

表 4 xBD 数据集消融实验定量评价结果/%  
Tab. 4 Quantitative Evaluation Results of Ablation Experiments on xBD dataset/%

| 网络模型      | 总体F1  | 建筑提取F1 | 损伤分类F1 | 各损伤级别F1 |       |       |       |
|-----------|-------|--------|--------|---------|-------|-------|-------|
|           |       |        |        | 无损伤     | 轻微损伤  | 中度损伤  | 严重损伤  |
| GLESNet-1 | 75.49 | 85.67  | 71.12  | 93.70   | 51.05 | 72.58 | 81.95 |
| GLESNet-2 | 76.18 | 85.61  | 72.13  | 93.57   | 53.31 | 71.68 | 82.97 |
| GLESNet-3 | 76.98 | 85.75  | 73.22  | 94.52   | 55.62 | 71.69 | 82.49 |
| GLESNet-4 | 77.56 | 85.74  | 74.06  | 94.42   | 57.20 | 71.93 | 83.10 |
| GLESNet   | 78.45 | 86.03  | 75.20  | 94.49   | 57.90 | 73.88 | 84.71 |

图 8 展示了消融实验的部分可视化结果。从第一、二行可以看到，DFEM 的加入改善了轻微损伤和无损伤类别的误分，然而由于中度损伤具有更强的语义模糊性，简单的差异增强无法改善中度损伤类别的分类效果，GLFFM 能够获取多尺度上下文中的丰富语义信息，缓解了对中等损伤类别的误分，并保证了结果的平滑，DEAD 进一步对语义信息进行挖掘，最终正确分类了全部建筑物区域。在第三行中，对于左下角的中度损伤建筑物，所有消融模型都无法对损伤类别进行正确分类，而加入 DEAD 的模型增强了训练过程中对此类困难样本的学习，因此能够得到正确的分类结果。第四行中 DFEM 的加入剔除了部分伪变化，改善了原始模型对于损伤程度的过高估计，GLFFM 增强了特征提取能力，将部分建筑物正确判断为无损伤类别，在此基础上，DEAD 聚焦类别模糊的建筑物，进一步补充了无损伤类别的区域，同时也提升了建筑提取的效果。

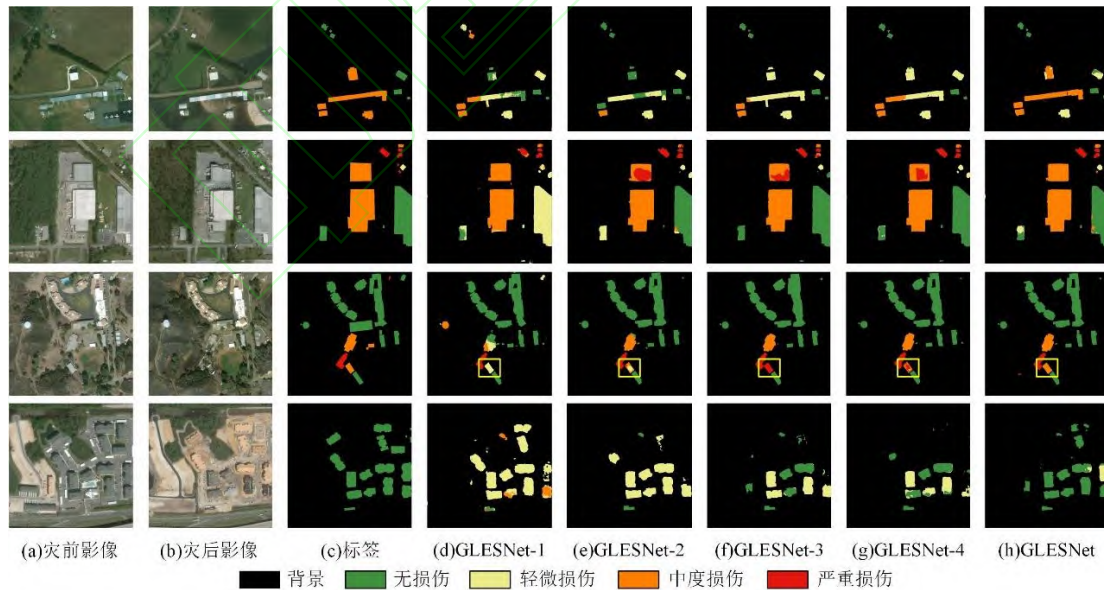


图 8 xBD 数据集消融实验可视化结果

Fig.8 Visualization Results of Ablation Experiments on xBD Dataset

为了进一步证明 DEAD 的有效性，对加入 DEAD 前网络难以处理的一些典型错误分类结果，以及加入 DEAD 之后的变化进行了展示，结果如图 9 所示。第一行灾害前后影像的光照变化较大，导致网

网络将部分正常建筑物识别为损伤建筑物。第二行灾害前后影像的拍摄倾角不同，导致双时相影像中建筑物的位置无法对齐，干扰了网络的预测。在第三行中，由于树木在灾前影像中对建筑物的遮挡，以及树木在灾害前后影像中的变化，从而使网络将部分建筑物误判为中度损伤。在第四行中，灾后影像较为模糊，网络将图像的模糊混淆为建筑物本身的变化，从而过高估计建筑物的损伤程度。加入 DEAD 的网络由于在训练过程中对易错的困难样本施加了更多关注，强化了对这些样本的判别能力，从而提升了损伤分类的效果。

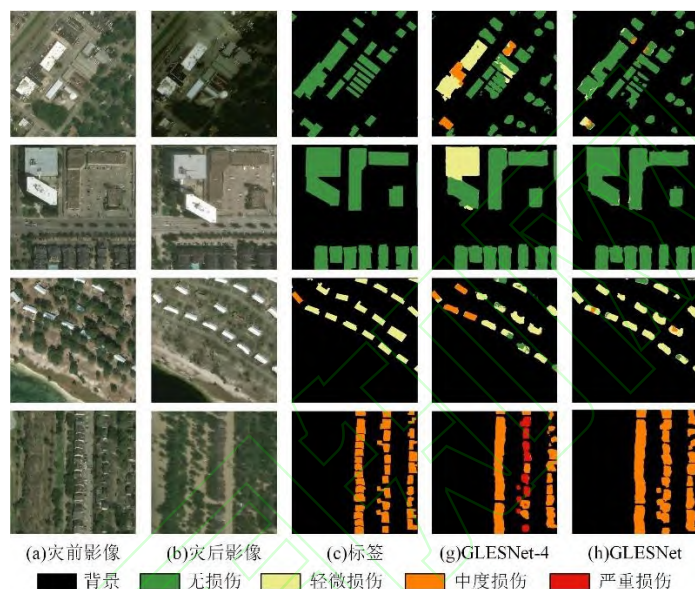


图 9 部分典型错误分类结果可视化  
Fig.9 Visualization of Some Typical Misclassification Results

## 2.6 迁移实验

为了探究在大范围数据集上训练的模型在真实场景下的迁移性能，以及后续微调对于迁移性能的影响，在 Ida-BD 数据开展了迁移实验。随机抽取 50%的 Ida-BD 数据固定为测试集，从剩下的数据中分别获取总数据量的 10%，20%，30%，40%和 50%进行训练或微调，结果如图 10 所示。图 10(a)是总体 F1、建筑提取 F1 和损伤分类 F1 的变化曲线，图 10(b)则是每个损伤类别 F1 各自的变化曲线。由于 Ida-BD 数据集中严重损伤类别样本量过少，因此只计算其他三个类别的 F1。如图 10 所示，预训练权重初始化的模型具有明显更高的初始 F1。对于建筑提取 F1，预训练权重初始化模型在 30%数据微调时基本达到了 xBD 数据集上的水平，并在后续保持稳定，无权重初始化模型在 50%数据微调时基本稳定，且与预训练权重初始化模型差异较小。对于损伤分类 F1，预训练权重初始化模型一直保持稳步增长，无权重初始化模型在前 30%数据量的微调中增长迅速，但最终仍与预训练权重初始化模型保持较大差距。各损伤类别 F1 的变化趋势与总体损伤分类 F1 的变化趋势基本一致。无损伤类别的 F1 明显高于轻微和中度损伤类别的 F1。相对无权重初始化模型，预训练权重初始化模型在轻微损伤和中度损伤类别的 F1 优势明显，而在无损伤类别上差异较小。此外，预训练权重初始化模型的训练过程也更加平稳。



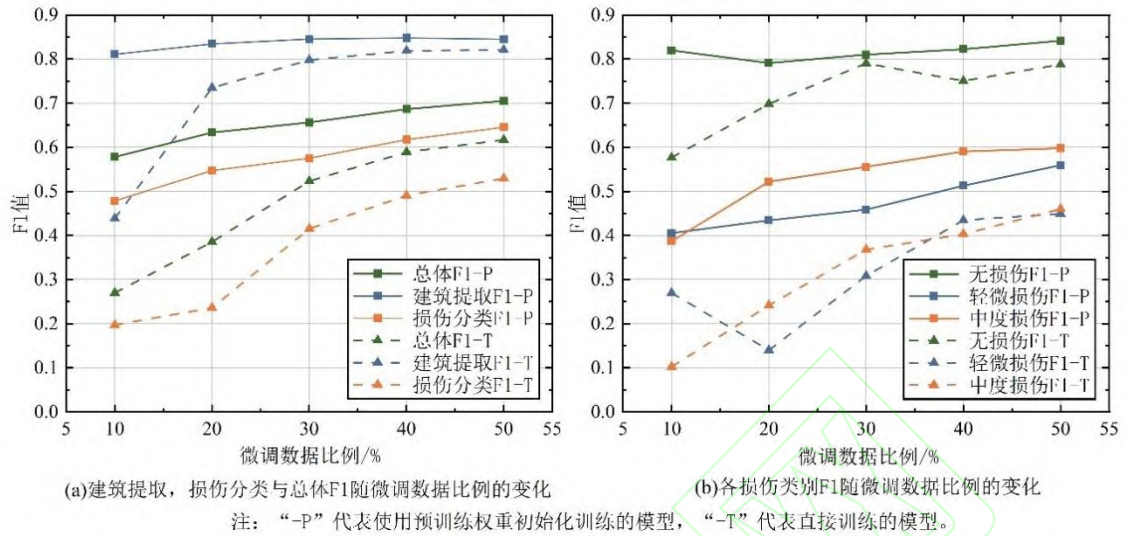


图 10 迁移实验中各评价指标随微调数据比例的变化  
Fig.10 The Change of Each Evaluation Index with the Proportion of Fine-tuned Data in the Transfer Experiment

为了进一步评估 GLENet 的迁移性能, 分别在未进行微调和使用 30%数据进行微调的情况下, 与 Siam-UNet 的效果进行了对比。在未进行微调时, 二者的性能均有较为严重的退化。GLENet 的损伤分类 F1 和总体 F1 略有优势, 分别优于 Siam-UNet 1.68%和 3.15%。更大的差距源于建筑提取 F1, Siam-UNet 与 GLENet 的差距达 6.6%, 这是由于 Siam-UNet 仅使用灾前影像训练独立的模型进行建筑提取, 因此学习到的建筑物特征泛化性不足导致的。而 GLENet 同时使用双时相影像, 联合损伤评估任务训练, 因此学习到了较为鲁棒的特征。在经过 30%数据微调后, 两个方法的效果均有了较大提升, GLENet 的总体 F1 分数达到了 65.62%, Siam-UNet 模型的总体 F1 也达到了 60%以上。

表 5 Ida-BD 数据集迁移实验对比结果定量评价/%  
Tab. 5 Quantitative Evaluation of the Transfer Experiment Results of the Ida-BD Dataset/%

| 微调数据比例 | 网络模型      | 总体F1         | 建筑提取F1       | 损伤分类F1       | 各损伤级别F1      |              |              |
|--------|-----------|--------------|--------------|--------------|--------------|--------------|--------------|
|        |           |              |              |              | 无损伤          | 轻微损伤         | 中度损伤         |
| 0%     | Siam-UNet | 41.56        | 71.48        | 28.74        | 82.14        | 23.35        | 20.25        |
|        | GLENet    | <b>44.71</b> | <b>78.08</b> | <b>30.42</b> | <b>82.72</b> | <b>31.53</b> | 18.24        |
| 30%    | Siam-UNet | 61.05        | 76.27        | 54.53        | <b>81.36</b> | 44.54        | 49.34        |
|        | GLENet    | <b>65.62</b> | <b>84.54</b> | <b>57.52</b> | 81.01        | <b>45.85</b> | <b>55.54</b> |

图 11 展示了经过 30%数据微调后 GLENet 和 Siam-UNet 的部分可视化结果。第一行显示了阴影条件下的结果。由于 Siam-UNet 仅采用前时相影像获取建筑物位置, 因此在右侧阴影区域的建筑提取上出现了大片的漏检, 而 GLENet 由于同时使用双时相影像, 不仅完整获取了建筑物区域, 在损伤分类上也具有一定的准确度。在第二行存在遮挡的区域, GLENet 也能够提取出 Siam-UNet 的漏检区域, 并给出正确的分类结果。对于第三行中面积较大的建筑物, GLENet 能保证所提建筑物轮廓的完整性, 以及内部损伤分类结果的一致性, 而 Siam-UNet 提取结果的轮廓较为模糊, 同时分类结果中椒盐噪声严重。对于第四行中的密集微小建筑物, GLENet 能够完整提取出建筑物的精细轮廓, 每个建筑物的



类别也能保持独立，而 Siam-UNet 的损伤分类结果存在粘连，难以区分每个建筑物的具体类别。

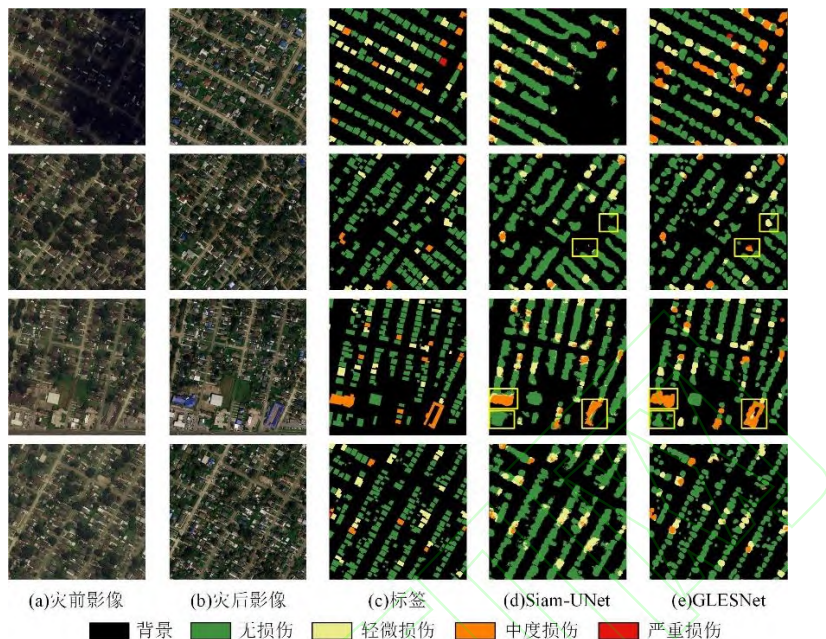


图 11 Ida-BD 数据集迁移实验对比结果可视化  
Fig.11 Visualization of the Transfer Experiment Results of the Ida-BD Dataset

## 2.7 变化检测实验

为了进一步说明 GLENet 对多种任务的适用性，在 LEVIR-CD 数据集上开展了二值变化检测实验。并与六种变化检测领域的先进方法对比，其中 DMINet<sup>[35]</sup>、ChangeFormer<sup>[36]</sup>和 BIT<sup>[37]</sup>是基于自注意力的方法，IFNet<sup>[38]</sup>、SNUNet<sup>[39]</sup>、TFI-GR<sup>[40]</sup>是基于卷积的方法。表 6 展示了各方法在 LEVIR-CD 数据集上的定量结果。尽管在召回率上略低于第二名 IFNet，但 GLENet 在其余的所有指标上均取得了最佳结果，尤其是在更综合的 F1 分数上高出其他方法 0.96%，在更严格的交并比上高出其他方法 1.41%，证明 GLENet 在变化检测任务上也表现出具有竞争力的性能。

表 6 LEVIR-CD 数据集变化检测实验结果定量评价/%  
Tab. 6 Quantitative Evaluation of the Experimental Results of Change Detection on the LEVIR-CD Dataset/%

| 网络模型          | OA           | P            | R            | F1           | IoU          |
|---------------|--------------|--------------|--------------|--------------|--------------|
| SNUNet        | 98.32        | 90.99        | 85.05        | 87.77        | 80.05        |
| ChangeFormer  | 98.37        | 89.70        | 87.72        | 88.68        | 81.27        |
| DMINet        | 98.43        | 92.40        | 85.03        | 88.36        | 80.82        |
| BIT           | 98.51        | 92.13        | 86.64        | 89.18        | 81.96        |
| TFI-GR        | 98.56        | 93.00        | 86.51        | 89.47        | 82.37        |
| IFNet         | 98.62        | 90.62        | <b>90.71</b> | 90.67        | 84.07        |
| <b>GLENet</b> | <b>98.68</b> | <b>94.48</b> | 89.16        | <b>91.63</b> | <b>85.48</b> |

图 12 显示了变化检测实验的部分定性可视化结果。第一、二行显示了背景较为简单，建筑物排列较为整齐时的结果，几种对比方法均能较为准确地提取出变化区域，而 GLENet 对于轮廓的提取更为

准确。第三、四行显示了背景噪声较多，建筑物排列较为散乱时的结果，基于卷积和自注意力的方法由于缺少有效上下文的指导，容易受到背景噪声的影响，造成了部分漏检和误检。而 GLEsNet 能够充分整合上下文信息，实现变化区域的准确提取。第五行展示了对大面积建筑物的提取结果，基于卷积的方法因为缺少长距离上下文关系，在建筑物内部出现了大面积的漏检，基于自注意力的方法则因为缺少细节信息，对于边缘的提取不够准确，且存在一定的噪声。而 GLEsNet 同时结合了全局信息和细节信息，可以完整提取出主要的变化区域，并具有较为精细的轮廓。

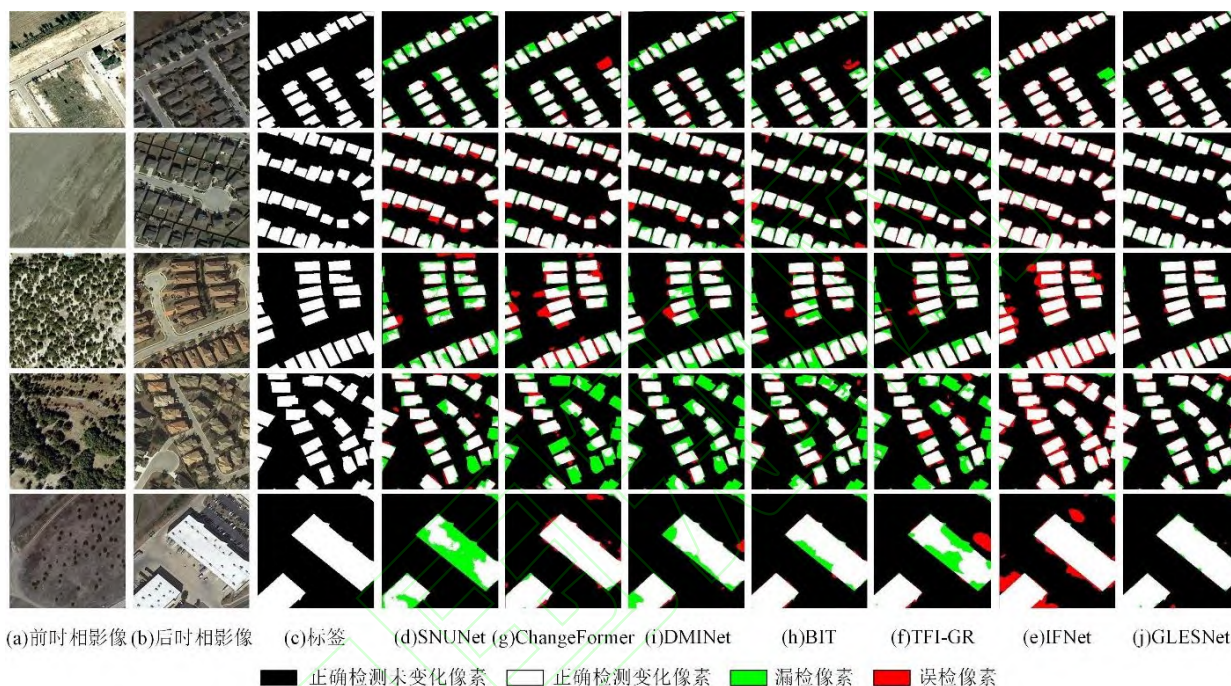


图 12 LEVIR-CD 数据集变化检测实验结果可视化  
 Fig.12 Visualization of the Experimental Results of Change Detection on the LEVIR-CD Dataset

## 2.8 模型复杂度与效率分析

本文进一步测试了不同方法的模型复杂度和运行效率。测试中，以  $512 \times 512$  像素作为数据的输入大小，以 ResNet-34 作为 GLEsNet 的骨干网络。本文通过参数量 (Params) 和浮点运算数 (FLOPs)，从参数和计算量两个维度衡量模型复杂度，并使用吞吐量 (Throughput) 衡量模型的实际运行效率。其中，吞吐量定义为模型在单位时间内 (例如，1s) 可处理的最大图像样本输入数量。不同方法的模型复杂度和运行效率测试结果如表 7 所示。测试结果表明，由于 GLEsNet 采用轻量化的注意力进行特征融合，并采用分组卷积替换原始卷积，因此在参数量和计算量上并未有过多增加，均处于可接受的范围内。由于选择性扫描具有在 CUDA 层面的专门优化，因此 GLEsNet 的具有较高的吞吐量，仅次于 Siam-UNet。GLEsNet 在模型复杂度、运行效率和精度上取得了较好的平衡。

表 7 不同方法的模型复杂度和运行效率比较  
 Tab. 7 Comparison of Model Complexity and Efficiency of Different Methods

| 网络模型      | 参数量/M | FLOPs/G | 吞吐量/ (img/s) |
|-----------|-------|---------|--------------|
| Siam-UNet | 26.44 | 66.56   | 137.50       |

|            |       |        |       |
|------------|-------|--------|-------|
| ChangeOS   | 54.75 | 218.28 | 45.42 |
| DamFormer  | 43.23 | 74.29  | 58.02 |
| Dual-HRNet | 59.54 | 91.27  | 42.21 |
| DAHiTra    | 14.18 | 47.60  | 27.53 |
| BDANet     | 44.33 | 158.35 | 63.85 |
| GLSENet    | 40.49 | 81.74  | 82.00 |

### 3 结语

本文针对目前高分辨率遥感影像建筑物损伤评估方法中存在的差异特征建模不足、全局-局部特征利用不充分和困难样本感知能力欠缺等问题,提出了一种基于全局-局部特征融合和动态错误监督网络 GLSENet 的遥感影像建筑物损伤评估方法。通过在特征融合阶段在通道和空间维度增强双时相特征的差异性,在解码阶段的深层同时利用全局-局部特征增强高级语义提取能力,中层和浅层使用多任务和动态监督的方式实现困难样本感知的学习,以提升方法的综合性能。相比其他先进方法, GLSENet 在总体 F1 上提升 1.26%,在损伤分类 F1 上提升 1.91%。对比原始模型,各消融模型的性能稳步提升,完整模型在总体 F1 上提升 2.96%,在损伤分类 F1 上提升 4.08%。迁移实验和变化检测实验证明了 GLSENet 在不同数据上的泛化性和在其他任务上的适用性。然而,目前的网络训练需要大量数据支撑,且在真实案例数据上的迁移性能有待进一步提升。因此,设计针对建筑物损伤评估任务特点的半监督和弱监督方法以减轻数据依赖、设计跨域方法以提升迁移性能是未来研究的重点。

### 参考文献

- [1] Gao Zhi, Hu Aohan, Chen Boan, et al. A Hierarchical Geometry-to-semantic Fusion GNN Framework for Earth Surface Anomalies Detection[J]. National Remote Sensing Bulletin, 2024, 28(7): 1760-1770. (高智,胡傲涵,陈泊安,等.多层次几何-语义融合的图神经网络地表异常检测框架[J].遥感学报,2024,28(07):1760-1770.)
- [2] Ge Xiaosan, Chen Xi, Zhao Wenzhi, et al. Detection of Damaged Buildings Based on Generative Adversarial Networks[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(2): 238-247. (葛小三,陈曦,赵文智,等.基于生成对抗网络的建筑物损毁检测[J].测绘学报,2022,51(02):238-247.)
- [3] Sui Haigang, Liu Chaoxian, Huang Lihong, et al. Application of Remote Sensing Technology in Earthquake-Induced Building Damage Detection[J]. Geomatics and Information Science of Wuhan University, 2019, 44(7): 1008-1019. (眭海刚,刘超贤,黄立洪,等.遥感技术在震后建筑物损毁检测中的应用[J].武汉大学学报(信息科学版),2019,44(07):1008-1019.)
- [4] Rehor M, Bähr H P, Tarsha - Kurdi F, et al. Contribution of Two Plane Detection Algorithms to Recognition of Intact and Damaged Buildings in Lidar Data[J]. The Photogrammetric Record, 2008, 23(124): 441-456.
- [5] Wang Qiao. Research Framework of Remote Sensing Monitoring and Real-Time Diagnosis of Earth Surface Anomalies[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(7): 1141-1152. (王桥.地表异常遥感探测与即时诊断方法研究框架[J].测绘学报,2022,51(07):1141-1152.)
- [6] Li Jiayi, Huang Xin, Hu Yuping, et al. Fusion of Optical Daily and Night-time Light Remote Sensing Images for Collapsed Building Detection: A Case in Turkey Mw 7.8 Earthquake[J]. Geomatics and Information Science of Wuhan University, 2023, 48(10): 1706-1714. (李家艺,黄昕,胡宇平,等.夜光影像和高分辨率影像耦合的土耳其 Mw 7.8 地震建筑倒塌智能解译[J].武汉大学学报(信息科学版),2023,48(10):1706-1714.)
- [7] Liu Ying, Tao Chao, Yan Pei, et al. Graph Cut Energy Driven Earthquake-damaged Building Detection from High-resolution Remote Sensing Images[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(7): 910-917. (刘莹,陶超,闫培,等.图割能量驱动的高分辨率遥感影像震害损毁建筑物检测[J].测绘学报,2017,46(07):910-917.)
- [8] Xie Jiali, Li Yongshu, Li Hechao, et al. Recognition of Damage Buildings in Hollow Village Based on Texture Feature of Gray Level Co-occurrence Matrix[J]. Bulletin of Surveying and Mapping, 2017, (12): 90-93,102. (谢嘉丽,李永树,李何超,等.利用灰度共生矩阵纹理特征识别空心村损毁建筑物的方法[J].测绘通报,2017,(12):90-93+102.)
- [9] Ye Xin, Wang Jun, Qin Qiming. Damaged Building Detection Based on GF-1 Satellite Remote Sensing Image: A Case Study For Nepal MS.1 Earthquake [J]. Acta Seismologica Sinica, 2016, 38(3): 477-485, 509. (叶昕,王俊,秦其明.基于高分一号卫星遥感图像的建筑物震害损毁检测研



究——以 2015 年尼泊尔 M<sub>s</sub>8.1 地震为例[J].地震学报,2016,38(03):477-485+509.)

- [10] Vetrivel A, Gerke M, Kerle N, et al. Disaster Damage Detection Through Synergistic Use of Deep Learning and 3D Point Cloud Features Derived from Very High Resolution Oblique Aerial Images, and Multiple-Kernel-Learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 140: 45-59.
- [11] Zhou Yang, Zhang Yunsheng, Chen Siyang, et al. Disaster Damage Detection in Building Areas Based on DCNN Features. Remote Sensing for Land & Resources[J], 2019, 31(02): 44-50. (周阳,张云生,陈斯飏,等.基于 DCNN 特征的建筑物震害损毁区域检测[J].国土资源遥感,2019,31(02):44-50.)
- [12] Gupta R, Goodman B, Patel N, et al. Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, USA, 2019.
- [13] Zheng Z, Zhong Y, Wang J, et al. Building Damage Assessment for Rapid Disaster Response with A Deep Object-based Semantic Change Detection Framework: From Natural Disasters to Man-made Disasters[J]. Remote Sensing of Environment, 2021, 265: 112636.
- [14] V. Durnov. Xview2 First Place Solution. [EB/OL]. [2024-10-16] [https://github.com/DIUX-xView/xView2\\_first\\_place](https://github.com/DIUX-xView/xView2_first_place).
- [15] Weber E ,Kané, Hassan. Building Disaster Damage Assessment in Satellite Imagery with Multi-temporal Fusion. [EB/OL]. [2024-10-16] <https://arxiv.org/abs/2004.05525>.
- [16] Shen Y, Zhu S, Yang T, et al. BDANet: Multiscale Convolutional Neural Network with Cross-directional Attention for Building Damage Assessment from Satellite Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.
- [17] Deng L, Wang Y. Post-disaster Building Damage Assessment Based on Improved U-Net[J]. Scientific Reports, 2022, 12(1): 15862.
- [18] Wang J, Guo H, Su X, et al. PCDasNet: Position-Constrained Differential Attention Siamese Network for Building Damage Assessment[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024. 62:1-18.
- [19] Ku, J., Seo, J., & Jeon, T. Dual-HRNet for Building Localization and Damage Classification. [EB/OL]. [2024-10-16] [https://github.com/SIAnalytics/dual-hrnet/blob/master/figures/xView2\\_White\\_Paper\\_SI\\_Analytics.pdf](https://github.com/SIAnalytics/dual-hrnet/blob/master/figures/xView2_White_Paper_SI_Analytics.pdf).
- [20] Xie H, Hu X, Jiang H, et al. Bssnet: Building Subclass Segmentation from Satellite Images Using Boundary Guidance and Contrastive Learning[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 7700-7711.
- [21] Gupta R, Shah M. RescNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery[C]//2020 25th International Conference on Pattern Recognition, Taichung, Taiwan, 2021.
- [22] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016.
- [23] 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.
- [24] Luo W, Li Y, Urtasun R, et al. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2016, 29.
- [25] Kaur N, Lee C C, Mostafavi A, et al. Large-scale Building Damage Assessment Using a Novel Hierarchical Transformer Architecture on Satellite Images[J]. Computer-Aided Civil and Infrastructure Engineering, 2023, 38(15): 2072-2091.
- [26] Chen H, Nemni E, Vallecorsa S, et al. Dual-tasks Siamese Transformer Framework for Building Damage Assessment[C]//IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022.
- [27] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021.
- [28] Gu A, Dao T. Mamba: Linear-time Sequence Modeling with Selective State Spaces. [EB/OL]. [2024-10-16] <https://arxiv.org/abs/2312.00752>.
- [29] Liu Y, Tian Y J, Zhao Y Z, et al. Vmamba: Visual State Space Model. [EB/OL]. [2024-10-16] <https://arxiv.org/abs/2401.10166>.
- [30] Wan Z, Wang Y, Yong S, et al. Sigma: Siamese Mamba Network for Multi-modal Semantic Segmentation. [EB/OL]. [2024-10-16] <https://arxiv.org/abs/2404.04256>.
- [31] Ma X, Zhang X, Pun M O. RS 3 Mamba: Visual State Space Model for Remote Sensing Image Semantic Segmentation[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 1-5.
- [32] Berman M, Triki A R, Blaschko M B. The Lovász-softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-over-union Measure in Neural Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.
- [33] Lee C C, Kaur N, Mahdavi-Amiri A, et al. Ida-BD: Pre- and Post-Disaster High-Resolution Satellite Imagery for Building Damage Assessment from Hurricane Ida. Designsafe-CI. [EB/OL]. [2024-10-16] <https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-3563>.
- [34] Chen H, Shi Z. A Spatial-temporal Attention-based Method and A New Dataset for Remote Sensing Image Change Detection[J]. Remote Sensing, 2020, 12(10): 1662.
- [35] Feng Y, Jiang J, Xu H, et al. Change Detection on Remote Sensing Images Using Dual-branch Multilevel Intertemporal Network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.
- [36] Bandara W G C, Patel V M. A Transformer-based Siamese Network for Change Detection[C]// IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022.
- [37] Chen H, Qi Z, Shi Z. Remote Sensing Image Change Detection with Transformers[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021,



60: 1-14.

- [38] Zhang C, Yue P, Tapete D, et al. A Deeply Supervised Image Fusion Network for Change Detection in High Resolution Bi-temporal Remote Sensing Images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 166: 183-200.
- [39] Fang S, Li K, Shao J, et al. Snunet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.
- [40] Li Z, Tang C, Wang L, et al. Remote Sensing Change Detection Via Temporal Feature Interaction and Guided Refinement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-11.

### 网络首发:

标题: 联合全局-局部特征和动态错误监督的遥感影像建筑物损伤评估方法

作者: 刘梓航, 刘欣怡, 张永军

收稿日期: 2024-11-18

DOI:10.13203/j.whugis20240346

### 引用格式:

刘梓航, 刘欣怡, 张永军. 联合全局-局部特征和动态错误监督的遥感影像建筑物损伤评估方法[J]. 武汉大学学报(信息科学版), 2024, DOI:10.13203/J.whugis20240346 (LIU Zihang, LIU Xinyi, ZHANG Yongjun. Building Damage Assessment from Satellite Images Combining Global-Local Features and Dynamic Error Supervision[J]. Geomatics and Information Science of Wuhan University, 2024, DOI:10.13203/J.whugis20240346)

网络首发文章内容和格式与正式出版会有细微差别, 请以正式出版文件为准!

---

### 您感兴趣的其他相关论文:

#### InSAR 数据约束下的 2023 年赫拉特地震序列发震断层探讨及其建筑物损毁评估

刘振江, 韩炳权, 刘海辉, 李振洪, 能懿菡, 陈博, 彭建兵

武汉大学学报(信息科学版), 2024, 49(5): 722-733.

<http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20230382>

#### 基于双流并行全向扫描 Mamba 的遥感影像建筑物变化检测

厉芳婷, 张过, 周方栋, 王上, 梁思, 李志炜

武汉大学学报(信息科学版), 2024, 49(9): 1574-1585.

<http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20240270>

#### 基于双路细节关注网络的遥感影像建筑物提取

张卓尔, 潘俊, 舒奇迪

武汉大学学报(信息科学版), 2024, 49(3): 376-388.

<http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20220613>