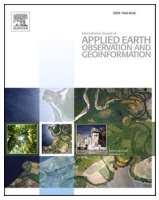
International Journal of Applied Earth Observation and Geoinformation 141 (2025) 104629

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Post-disaster building damage assessment based on gated adaptive multi-scale spatial-frequency fusion network

Bo Yu [a,b,c], Yao Sun [a,d], Jiansong Hu [c,e], Fang Chen [a,b,c,*], Lei Wang [a,b,c,*]

[a] International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China
[b] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[c] Chenjiang Laboratory, Chenzhou 423000, China
[d] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
[e] Longhudong Reservoir Management Office, Chenzhou 423000, China

ARTICLE INFO

Keywords:
Building damage assessment
Adaptive attention
Convolutional neural network
Satellite imagery
Multi-scale feature extraction

ABSTRACT

Accurate building damage assessment is crucial for post-disaster response, yet existing methods struggle to capture complex spatial relationships and contextual features needed for distinguishing damage levels. To address this, we propose the Gated Adaptive Multi-scale Spatial-frequency Fusion Network (GAMSF), a two-phase framework for building localization and damage classification. GAMSF integrates three key innovations: (1) Adaptive Attention (AA) to dynamically prioritize critical regions, (2) Gated Multi-scale Feed-Forward Network (GMFFN) to enhance robustness by emphasizing prominent damage features, and (3) Multi-Scale Wavelet Fusion (MWF) to extract fine-grained structural details using wavelet transforms. Rigorous evaluations on the datasets, including xBD and xFBD, demonstrates that GAMSF achieves the state-of-the-art performance, with a 1.7% improvement in F1-score, a 2.1% gain in Kappa, and a 3.7% increase in minor damage identification accuracy compared to existing approaches. Furthermore, transferability experiments on the high-resolution Ida-BD dataset validate GAMSF's superior generalization capabilities, outperforming four advanced models. These results highlight the practical value of GAMSF in enhancing disaster management, emergency response, and resource allocation strategies.

1. Introduction

Timely and accurate assessment of building damage is crucial for effective post-disaster response. Traditional ground-based inspections, although reliable, are constrained by limited scalability, time inefficiency, and inaccessibility in disaster-stricken regions. In contrast, remote sensing imagery, particularly high-spatial-resolution (HSR) satellite, aerial, and drone data, has become a powerful tool for facilitating rapid post-earthquake response and rescue operations, offering efficient and timely damage assessment capabilities (Khankeshizadeh et al., 2024; Xing et al., 2024). As the demand for rapid and large-scale damage assessments grows, the development of automated methods leveraging HSR imagery becomes increasingly vital (Entezami et al., 2024; Sarmadi et al., 2022).

Currently, building damage assessment methods can be categorized into three categories: (1) manual inspection, (2) machine learning-based methods, and (3) deep learning-based methods. Manual inspection is highly accurate but time-consuming and dependent on expert evaluators (Chen et al., 2024a; Spencer et al., 2019). Machine learning-based algorithms, such as Support Vector Machines (SVM) and Random Forests (RF) (Breiman, 2001), automate assessments but rely on handcrafted features, reducing adaptability. Anniballe et al. (2018) used an SVM classifier with color, texture, and statistical features to distinguish pre- and post-disaster building damage. Natarajan et al. (2023) applied Elastic Net Regression to integrate seismic and structural data for rapid earthquake damage assessment. Other works, such as Wang et al. (2022), have focused on combining SAR and optical data to improve damage assessment accuracy, and Ghimire et al. (2022) achieved 68 % accuracy using RF Regression Model on the 2015 Gorkha earthquake dataset. However, the reliance on feature engineering makes traditional machine learning costly and lack the flexibility for heterogeneous disaster scenarios.

* Corresponding authors at: International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China.
E-mail addresses: chenfang@radi.ac.cn (F. Chen), wanglei@radi.ac.cn (L. Wang).

https://doi.org/10.1016/j.jag.2025.104629
Received 21 March 2025; Received in revised form 23 May 2025; Accepted 24 May 2025
Available online 28 May 2025

Deep learning significantly advances damage assessment by improving feature extraction and classification (Xu et al., 2023; Yue et al., 2021). CNN-based methods have proven effective in capturing local structural features (e.g., collapses), but often struggle with fine-grained damage (e.g., minor cracks) due to their limited receptive fields. Sidharta et al. (2022) analyzed deep learning trends in building damage assessment, while Zhang et al. (2023) introduced U-BDD++ to incorporate vision-language models for more nuanced understanding. Zaryabi et al. (2022) developed MSBDA-Net, a multi-scale Siamese network that enhances correlation between pre- and post-disaster images via cross-attention fusion. Ci et al. (2019) combined CNNs with ordinal regression to refine damage feature extraction, and Jamshidi et al. (2024) proposed the Strong Baseline Method, simplifying the xView2-winning model while maintaining strong performance. Despite their promise, CNNs often fail to generalize across diverse scenarios.

To address this, attention-based deep learning models have been introduced to improve long-range dependency modeling and adaptive feature weighting. Hu et al. (2019) introduced SENet, which recalibrates channels to highlight key features, while Woo et al. (2018) proposed CBAM, integrating spatial and channel attention for refined feature selection. Shen et al. (2022) developed BDANet, incorporating Cross-Directional Attention Blocks to strengthen correlations between pre- and post-disaster images, improving robustness. Hou et al. (Hou et al., 2024), developed Multi-Scale Attention Fusion Module, enhancing key feature perception while suppressing irrelevant information.

With the rise of Transformers, self-attention mechanisms have transformed building damage assessment by enabling global context modeling. Da et al. (2022) introduced SDAFormer, a Siamese hierarchical Transformer that achieved state-of-the-art results on xBD using high-resolution satellite imagery. Kaur et al. (Kaur et al., 2023) proposed DAHiTrA to capture temporal and spatial variations for better classification. Gomroki et al. (2025) developed WETUM for generating Building Damage Maps using drone imagery, achieving a 78.26 % Damage Detection Rate (DDR) when integrating spectral and geometric features. However, Transformer-based methods still face challenges such as inadequate modeling of spatial relationships and limited sensitivity to subtle damage features (Yu et al., 2023).

Wavelet transforms have proven effective in analyzing signals at multiple scales, capturing fine-grained textural and structural details often overlooked by conventional deep learning models. Their integration with deep learning enhances robustness and generalization, particularly for complex datasets. Chen et al. (2021) combined wavelet transforms with CNNs for aerial damage assessment, while Liu et al. (2019) introduced a wavelet scattering network, significantly improving remote sensing classification accuracy. Jamshidi and El-Badry (2023) demonstrated that wavelet-based models enhance generalization, leading to more precise post-disaster damage assessments. Barbosh and Sadhu (2025) further advanced this approach with a damage localization framework, integrating acoustic signals with wavelet transforms and deep learning for high-precision structural damage extraction. Although these studies confirm the promise of wavelet integration, most models often struggle to differentiate damage levels across building types and disaster scenarios, particularly in multi-scale contexts.

Despite considerable progress, several key challenges remain unaddressed in the building damage assessment domain. First, current methods often lack robust integration of multi-scale features, limiting their ability to extract both large structural failures and subtle damage. Second, many models fail to effectively incorporate temporal and spatial contextual information, which is essential for accurate post-disaster analysis. Third, existing approaches frequently struggle with generalization across diverse disaster types and environments, reducing their practical utility. Furthermore, while attention mechanisms and wavelet transforms offer promise, their combined potential has not been fully explored in an end-to-end damage assessment framework. These gaps highlight the need for novel architectures that are both accurate and adaptable across a wide range of disaster conditions.

To address these limitations, we propose GAMSF, a novel Transformer-based network framework tailored for post-disaster building damage assessment. GAMSF introduces three key innovations: (1) Multi-Scale Wavelet Fusion (MWF) module for fine-grained feature extraction through frequency decomposition; (2) Adaptive Attention (AA) module to selectively prioritize critical damage-relevant regions; and (3) Gated Multi-Scale Feed-Forward Network (GMFFN) to refine spatial and channel-level feature representation. The framework employs a two-phase approach, ensuring both precise localization of buildings and accurate classification of damage severity. In Phase 1: Building Localization, a Transformer-based segmentation network is employed to extract detailed building structures from pre-disaster images. This step establishes a reliable spatial foundation, enhancing the accuracy of damage assessment by providing well-defined building boundaries. In Phase 2: Damage Classification, a Siamese fusion network processes both pre- and post-disaster images, integrating the localized information from Phase 1. This fusion enables a more precise comparison of structural changes, facilitating improved damage classification. Additionally, leveraging the pre-trained weights from Phase 1 enhances training efficiency and improves model generalization, ensuring robustness across different disaster scenarios.

2. Method

2.1. Overall architecture

The GAMSF framework introduces a two-phase approach for post-disaster building damage assessment, differing from widely used change detection architectures (Guo et al., 2018), as illustrated in Fig. 1. In Phase 1, the framework focuses on building localization. It processes pre-disaster images through a multi-stage sequence (Stages 1 to 4), applying refined feature extraction and enhancement to accurately generate a binary building mask. This mask serves as the spatial foundation for the next phase. In Phase 2, both pre- and post-disaster images are processed through parallel multi-stage networks to extract features. These features are fused using concatenation and 1×1 convolution, allowing effective comparison across time. The building mask from Phase 1 is integrated into this process to guide damage classification. By combining structural information and temporal comparison, GAMSF classifies buildings based on damage severity. This two-phase workflow ensures accurate, robust, and scalable damage assessment.

2.2. Phase 1: Building localization

As illustrated in Fig. 2, Phase 1 extracts building boundaries from pre-disaster images, establishing the spatial reference required for accurate damage classification. The network architecture integrates several key components that optimize feature extraction and localization accuracy. The Multi-Scale Wavelet Fusion (MWF) module, introduced in Stage 1, decomposes the input image into multiple frequency components using wavelet transforms. This enhances the model's ability to capture fine structural details, textures, and edges at various scales, improving building boundary delineation. In Stages 2 to 4, the Adaptive Attention (AA) module dynamically refines focus on critical regions by adjusting attention weights based on the spatial importance of different areas. This selective focus is particularly beneficial in complex urban settings, where building occlusions and structural variations can hinder localization. In addition, multi-resolution feature fusion ensures consistent spatial representation across scales. Together, these techniques yield a precise segmentation map, forming a strong foundation for damage classification.

2.3. Phase 2: Damage classification

As shown in Fig. 3, Phase 2 uses a dual-branch network to process both pre-disaster and post-disaster images for damage classification. The

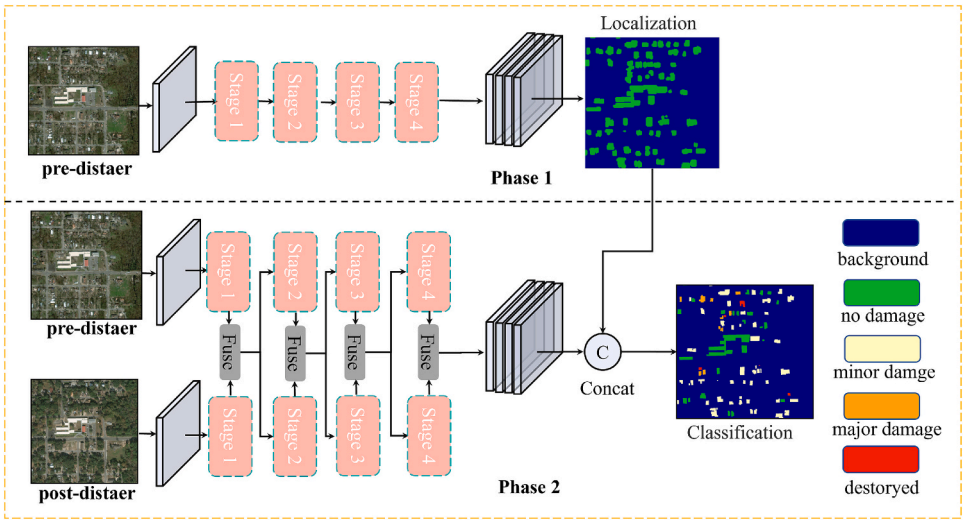


Fig. 1. Overall structure of GAMSF, composed of two phases for building localization and damage classification respectively.

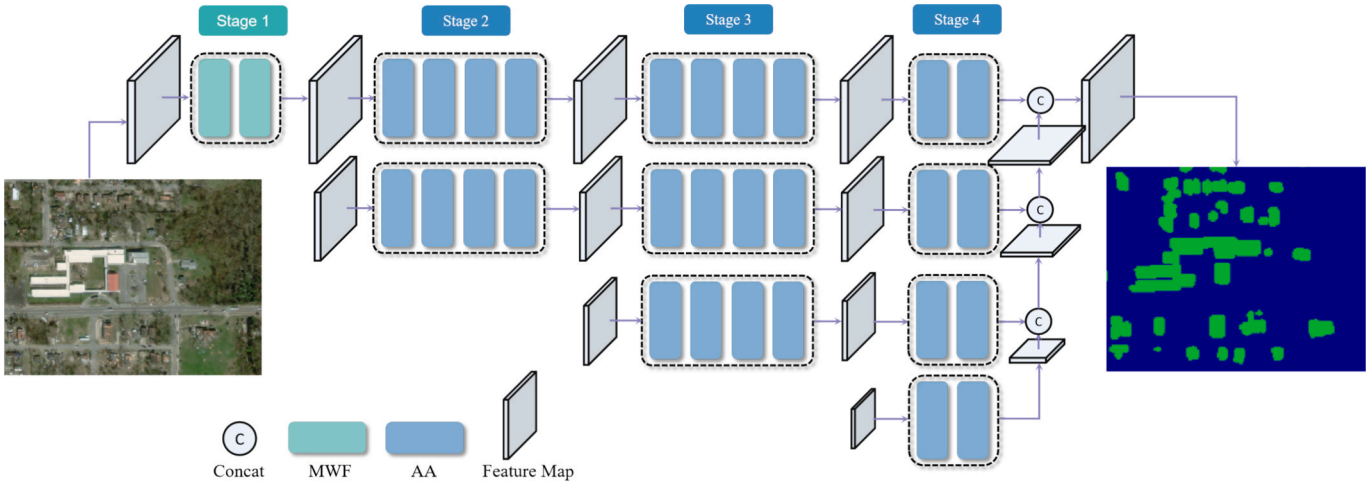


Fig. 2. The Phase 1 of the GAMSF for Building Localization, composed of Multi-Scale Wavelet Fusion (MWF) module and Adaptive Attention (AA) module.

network mirrors Phase 1′s structure to maintain consistency in feature extraction while allowing separate treatment of the two image types. The MWF module in Stage 1 captures multi-scale structural and damage-specific patterns by decomposing the images into frequency components. Features from each branch are fused via concatenation and 1×1 convolution to incorporate contextual cues from both time points. The AA module, used in later stages, emphasizes regions with significant damage and filters out irrelevant background information. Finally, the fused and refined features are used to classify buildings by damage severity, leveraging both temporal context and spatial focus. This approach enhances classification accuracy and robustness across diverse disaster scenarios.

2.4. Multi-Scale wavelet fusion (MWF) module

Building damage occurs at multiple spatial scales, necessitating features that capture both local and global patterns. The MWF module, shown in Fig. 4, addresses this by embedding multi-resolution processing into the encoder. It performs wavelet transforms to separate input images into low-frequency components (representing structural integrity) and high-frequency components (capturing fine damage details). The sparse representation of wavelet outputs reduces noise and sharpens the focus on relevant damage areas, making the model more effective across varying disaster conditions. Specifically, a single-level Haar wavelet decomposition is applied. This choice balances detail retention with computational efficiency and minimizes the propagation of noise from high-frequency components.

The MWF module incorporates two parallel processing streams to extract features from images at different scales following Equations (1) to (4). Let I represent the original image. The first stream operates on the image at its original resolution, utilizing convolutional layers and batch normalization to generate feature maps. The other stream performs wavelet decomposition into four sub-bands: LL (low-low), LH (low–high), HL (high-low), and HH (high-high). These components capture structural and textural characteristics at different frequencies. The low-frequency component retains the overall structure and primary features of the image, while the high-frequency component captures finer details and boundary features. However, it is acknowledged that the HH sub-band, which contains the highest-frequency details, may introduce noise into the damage evaluation process due to the sensitivity to image artifacts and environmental noise. To mitigate this, the network architecture is designed to learn and selectively emphasize informative frequency components during training, reducing the potential negative impact of HH-induced noise on classification accuracy.

The feature maps from both streams are concatenated along the channel dimension to generate I_h. In the second stream, I is first

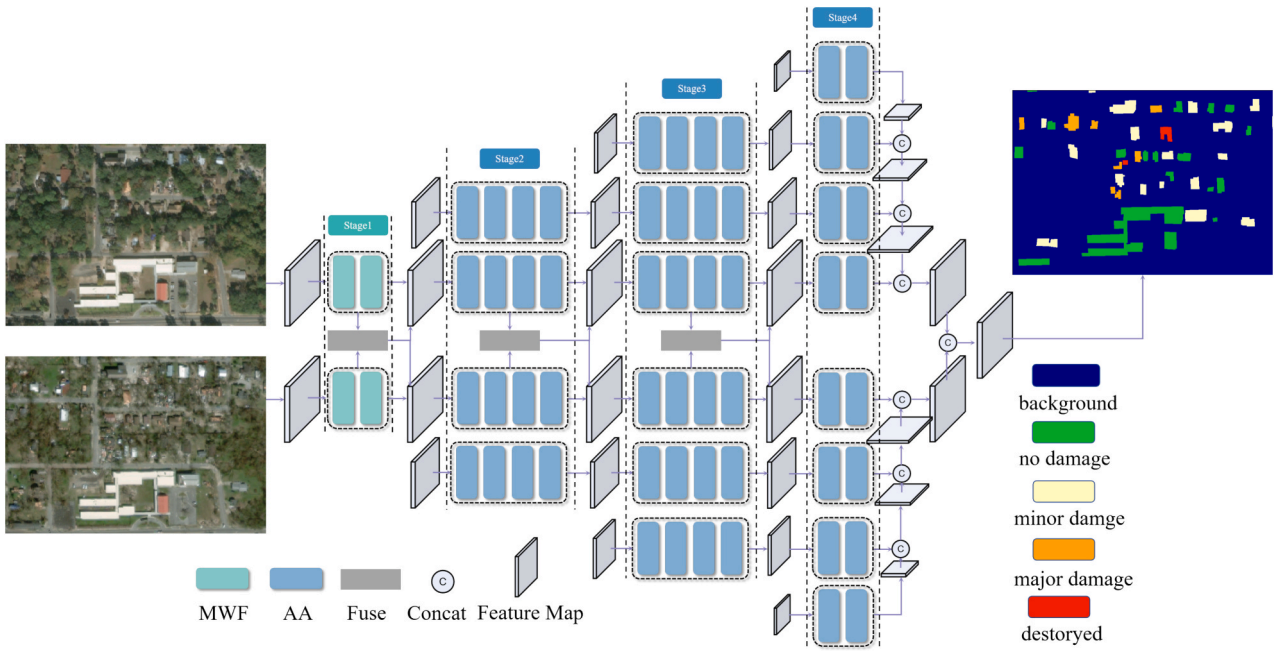


Fig. 3. Diagram of Network Structure in Phase 2, composed of Multi-Scale Wavelet Fusion (MWF) module and Adaptive Attention (AA) module.

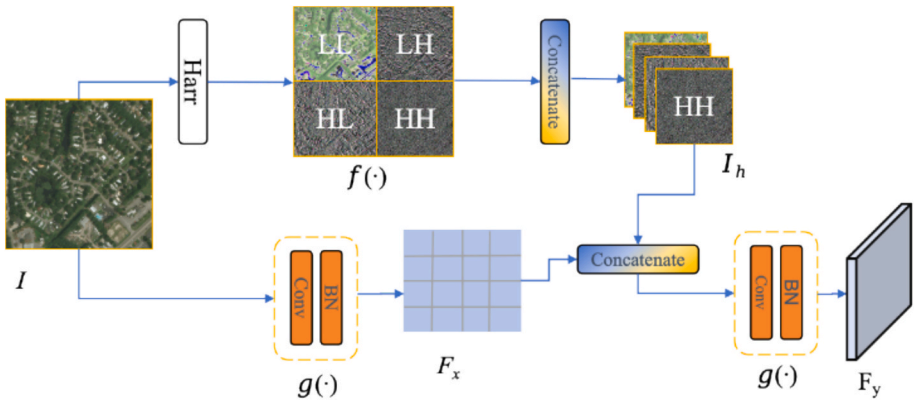


Fig. 4. Diagram of MWF module.

processed through $g(\cdot)$ to obtain F_x, which reduces the size of the feature map and ensures consistency with the dimensions of I_h. Finally, I_h and F_x are concatenated along the channel dimension and processed through $g(\cdot)$ to generate the final output F_y. This effectively integrates image features at different scales, allowing the model to capture both subtle damage features and broader contextual information.

$$LL, LH, HL, HH = f(I) \tag{1}$$

$$I_h = concat(LL, LH, HL, HH) \tag{2}$$

$$F_x = g(I) \tag{3}$$

$$F_y = g(concat(I_h, F_x)) \tag{4}$$

2.5. Adaptive attention (AA) module

To enhance the robustness of building damage feature learning, we introduce an Adaptive Attention (AA) module, combining CNN and Transformer strengths through adaptive cross-attention (Fig. 5). This module dynamically balances computational efficiency with global feature extraction by employing crisscross window-based self-attention.

Given an input feature map $F_x \in R^{2C\times H1\times W1}$, it is divided into horizontal and vertical stripe regions of dimensions H and W. Each stripe region is further segmented into multiple $K \times K$ windows, where $K \leq \min(H, W)$, allowing localized feature extraction at different scales. A neighborhood attention mechanism is then applied to each window to capture fine-grained structural details. Specifically, for a pixel centered at point (x, y) within a $K \times K$ window, its local features are extracted to form the matrix $F_{pi} \in R^{C\times H1\times W1}$, where $i \in \{1, 2\}$.

To effectively capture both local structural deformations and global contextual dependencies, we calculate attention scores that dynamically adjust the model's focus on critical regions, ensuring that damage-related features are prioritized while suppressing irrelevant background information. The attention scores are computed as follows. First, learnable transformations are applied to the local feature matrices to generate the Query, Key, and Value matrices, denoted as $Q_{pi(x,y)}$, $K_{pi(x,y)}$, and $V_{pi(x,y)}$. Within each $K \times K$ window centered at $p_i(x, y)$, we compute the attention relevance in three steps:

1. Compute attention weights: Perform matrix multiplication between the Query matrix $Q_{pi}(x,y)$ with the transpose of the Key matrix $K_{pi}(x, y)$;

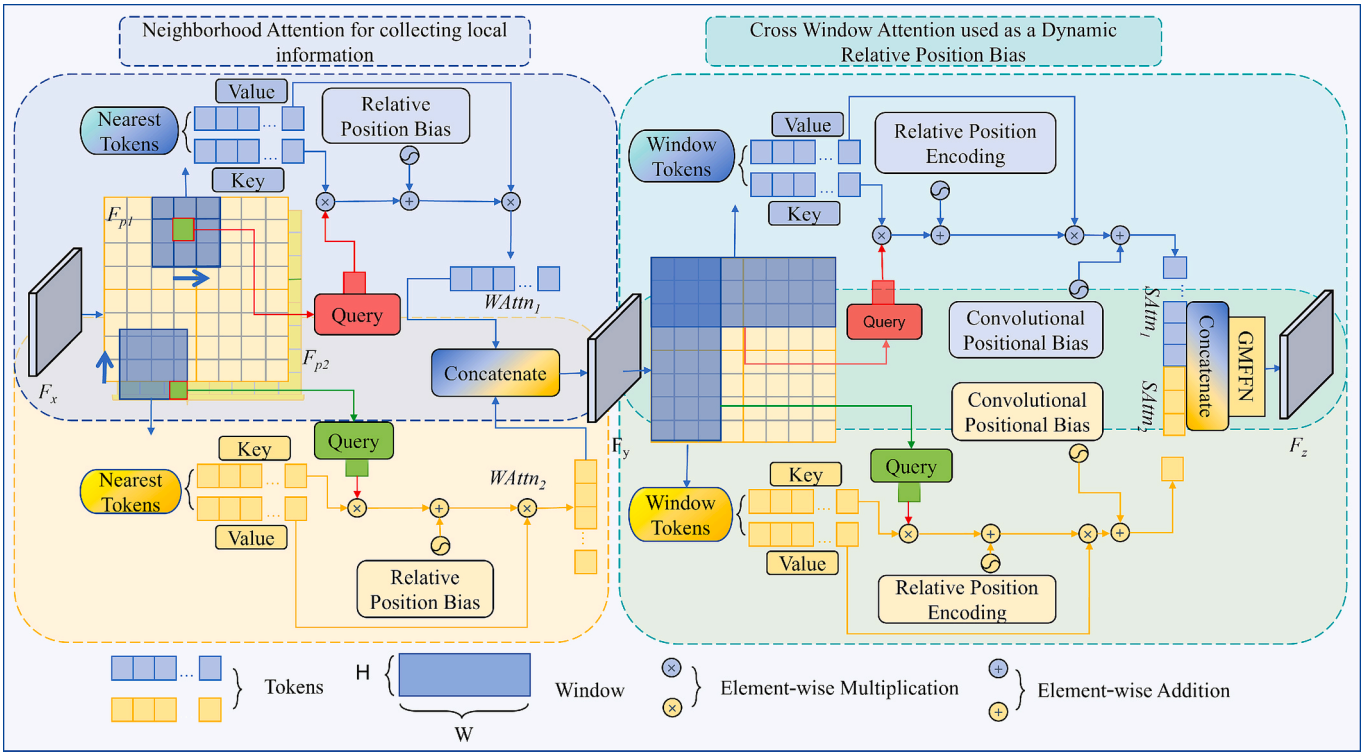


Fig. 5. Diagram of network structure adaptive attention.

2. Incorporate positional information: Add the position bias $B_i(x,y)$ to the result to signify the attention relevance of point $pi(x, y)$;
3. Normalized and scale: To prevent gradient vanishing in the softmax function, we normalize values and scale the attention intensity using $dk_{pi}(x, y)$, where $dk_{pi}(x, y)$ represents the dimensionality of the Key matrix. This procedure generates the attention scores $WAttn_i$ using Eq. (5). The final window attention feature map F_y is then obtained by concatenating attention scores along the channel dimension following Eq. (6).

$$WAttn_i = Softmax\left(\frac{Q_{p_i}(x_i,y_i)K_{p_i}(x_i,y_i)^T + B_i(x_i,y_i)}{d_{kp_i}(x_i,y_i)}\right)V_{p_i}(x_i,y_i), i \in \{1,2\} \tag{5}$$

$$F_y = concat(WAttn_1, WAttn_{2)}) \tag{6}$$

Once the window-level attention scores are computed, the feature map $F_y \in R^{2C\times H1\times W1}$ is further partitioned into two sub-feature maps, each containing C channels. One sub-feature map is divided into stripe regions of $C \times H_1 \times W$, where $W \le W_1$, while the second is divided into $C \times H \times W_1$, where $H \le H_1$. This partitioning mechanism dynamically adjusts stripe sizes based on the input feature map dimensions, allowing the model to capture multi-scale damage patterns.

Next, we calculate the stripe-level attention scores $SAttn_i$, similar to the window-level attention but at a broader spatial scale. As shown in Eqs. (7) and (8), instead of computing attention for individual pixel neighborhoods, this step aggregates information across entire stripe regions. This is achieved by introducing relative position encoding for better modeling of long-range dependencies. Finally, to integrate local (window-level) and global (stripe-level) representations, the attention scores from both stripe regions are concatenated along the channel dimension, producing the final feature map F_z. This integration enables the model to combine fine-grained damage cues using window attention with broader contextual cues from stripe attention, resulting in a comprehensive and precise feature representation for post-disaster building damage classification.

$$SAttn_i = \text{Softmax}\left(\frac{Q_iK_i^T + B_i}{\sqrt{d_{ki}}}\right)V_i, i \in (1,2) \tag{7}$$

$$F_z = GMFFN(concat(SAttn_1, SAttn_2)) \tag{8}$$

2.6. Gated multi-scale Feed-Forward network (GMFFN)

To improve the Transformer's capability in capturing both fine-grained local details and global context for building damage assessment, we propose the Gated Multi-scale Feed-Forward Network (GMFFN), shown in Fig. 6. GMFFN selectively refines feature representations using spatial reduction attention and multi-scale feature fusion.

Given an input tensor $F_y \in R^{2C\times H\times W}$, the GMFFN first splits the feature map along the channel dimension into two components: F_{p1} and F_{p2}. The first component, F_{p1}, undergoes spatial reduction attention $f(\cdot)$ to emphasize critical local features and suppress redundancy following Equation (9). The second component, F_{p2}, applies a linear transformation $g(\cdot)$ to preserve global structural relationships referring to Equation (10), ensuring that broader structural relationships are maintained.

$$T_{p3} = f(F_{p1}) \tag{9}$$

$$T_{p4} = g(reshape(F_{p2})) \tag{10}$$

To effectively merge local and global information, the two transformed feature sets are combined through element-wise multiplication followed by a linear projection, creating feature map F_z. This fusion mechanism strengthens the model's ability to differentiate between varying damage levels by emphasizing relevant features while filtering out noise. By effectively combining fine-grained structural details and broader context, GMFFN enhances the model's robustness and accuracy in classifying varying degrees of building damage.

$$F_z = reshape(concat(T_{p3}, Projection(T_{p4} \times T_{p3}))) \tag{11}$$

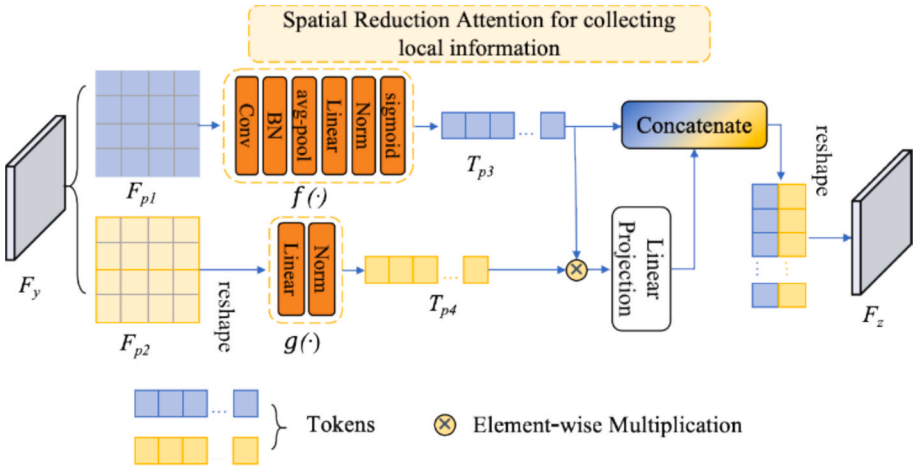


Fig. 6. GMFFN network structure diagram.

3. Datasets

To evaluate the proposed method, we utilize three publicly available datasets: xBD (Gupta et al., 2019), xFBD (Melamed et al., 2023), and Ida-BD (Kaur et al., 2023). The xBD and xFBD datasets are combined for training and evaluation, while the Ida-BD dataset serves as an independent benchmark for evaluating the model's generalization capability. A visual comparison, including pre-disaster and post-disaster imagery along with damage annotations, is provided in Fig. 7 to illustrate differences across these datasets.

3.1. xBD dataset

The xBD dataset (Gupta et al., 2019) is one of the largest benchmarks for building damage assessment, containing 22,068 pre- and post-

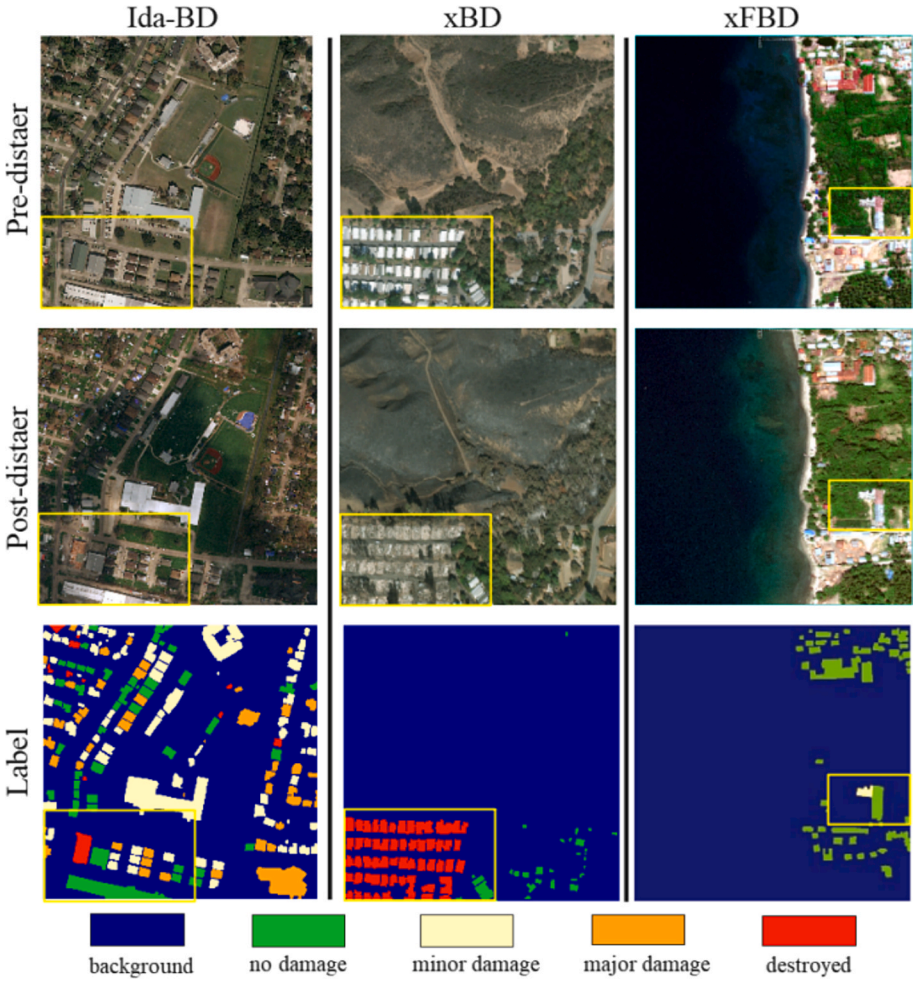


Fig. 7. Comparison of different datasets with images before and after the disaster, and the annotations of actual building damage.

disaster image pairs from WorldView and GeoEye satellites. The dataset covers multiple disasters, including earthquakes, hurricanes, floods, and volcanic eruptions, providing images with a spatial resolution of approximately 1.2 m. Each image is 1024 × 1024 pixels in size, with a ground sampling distance (GSD) of < 0.8 m. Building damage in xBD is categorized into four levels: no damage, minor damage, major damage, and destroyed. The dataset poses significant challenges due to its imbalanced class distribution, where certain damage categories, particularly minor and major damage, are underrepresented. Additionally, geographical and environmental variations across different regions introduce further complexity, requiring models to demonstrate high adaptability and context-awareness to achieve robust performance.

3.2. xFBD dataset

The xFBD dataset (Melamed et al., 2023) extends the xView2 challenge dataset with 16,234 image pairs, each with a size of 1024 × 1024 pixels and a GSD of < 0.8 m, captured from WorldView and GeoEye satellites, similar with xBD dataset. Unlike xBD, xFBD incorporates additional environmental context, including roads, vegetation, and water. It emphasizes subtle structural changes, increasing task difficulty. This dataset also suffers from class imbalance, demanding robust learning strategies. Its increased complexity and fine-grained spatial detail make xFBD a more challenging and demanding benchmark for evaluating model robustness in damage classification.

3.3. Ida-BD dataset

The Ida-BD dataset (Kaur et al., 2023) contains 87 high-resolution pre- and post-disaster image pairs from the WorldView-2 satellite. It covers areas most affected by Hurricane Ida near New Orleans, Louisiana, in August 2021. The dataset includes panchromatic imagery with a spatial resolution of 46 cm, which was orthorectified to 0.5 m/pixel for precise geolocation alignment. Unlike xBD and xFBD, Ida-BD provides a significantly higher spatial resolution, enabling the capture of fine-grained damage patterns such as partial collapses, roof deformations, and facade damages. This dataset is used solely for testing to evaluate cross-domain generalization.

3.4. Loss function

To ensure effective model training for both building localization (Phase 1) and damage classification (Phase 2), we employ a composite loss function combining focal loss (Lin et al., 2018) and dice loss (Milletari et al., 2016). Focal Loss addresses class imbalance by emphasizing hard-to-classify samples. It is calculated using Equation (12), where p_t represents the predicted probability of the correct class, while α and γ are modulation parameters controlling the contribution of hard samples. Dice Loss, based on overlap between prediction and ground truth masks, enhances boundary accuracy. It is calculated according to Equation (13), where P and G represent the predicted segmentation mask and ground truth mask, respectively. The final loss function is formulated as Equation (14).

$$Focal\,Loss(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{12}$$

$$Dice\,Loss = 1 - \frac{2|P \cap G|}{|P| + |G|} \tag{13}$$

$$Loss = Focal\,Loss + Dice\,Loss \tag{14}$$

3.5. Performance evaluation metrics

To evaluate the model's performance in both segmentation and classification, we adopt standard metrics, including Precision, Recall, and F1-score (Weber and Kané, 2020), as well as the XView2 Challenge evaluation framework (Zheng et al., 2021). For building segmentation, we use F1-score for localization ($F1_{loc}$), computed as Equation (15). For damage classification, we use the harmonic mean F1-score ($F1_{cls}$), aggregating per-class scores following Equation (16). In this context, True Positives (TP) represent the number of correctly classified pixels, while False Positives (FP) are incorrect predictions. Finally, a comprehensive evaluation metric is introduced to reflect overall performance by integrating both segmentation and classification performance using Equation (17). Additionally, Kappa statistic is used to evaluate overall classification performance while accounting for chance agreement, which is crucial in multi-class settings with imbalanced classes. It is computed using Equation (18), where P_o represents the observed agreement, calculated as the ratio of correctly classified samples, as shown in Equation (19). The notation P_e is the expected agreement by chance, calculated referring to Eq. (20). The inclusion of the Kappa provides a robust measure of the model's classification capability.

$$F1_{loc} = \frac{2TP}{2TP + FP + FN} \tag{15}$$

$$F1_{cls} = \frac{n}{\sum_{i=1}^{n}\frac{1}{F1_{C_i}}} \tag{16}$$

$$F1_{score} = 0.3 \times F1_{loc} + 0.7 \times F1_{cls} \tag{17}$$

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \tag{18}$$

$$P_o = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$P_e = \frac{(TP + FN) \times (TP + FN) + (FP + TN) \times (FN + TN)}{N^2} \tag{20}$$

3.6. Implementation details

All the experiments are conducted using PyTorch 2.2.0 on an NVIDIA RTX-3090 GPU. The combined dataset of xBD and xFBD is split into training (14,545 image pairs), validation (2,207 pairs), and testing (2,399 pairs). All the image pairs are cropped to 512 × 512-pixel patches. To enhance generalization, data augmentation techniques, including rotation, cropping, Gaussian noise injection, and flipping, are applied.

The AdamW optimizer (Nex et al., 2019) is used for model training with an initial learning rate of $1e^{-4}$, and a batch size of 7. The building localization model (Phase 1) is trained for 150 epochs, while the damage classification model (Phase 2) is fine-tuned for 300 epochs, initialized with Phase 1 weights. To mitigate overfitting, we employ early stopping, data augmentation, and L2 regularization. The Ida-BD dataset is used exclusively for testing to evaluate generalization. Sigmoid activation is used for binary localization in Phase 1, and softmax for multi-class classification in Phase 2.

4. Experimental results

4.1. Comparison with other methods

To evaluate the effectiveness of the proposed GAMSF framework, we conducted comparing experiments against four state-of-the-art methods on the combined dataset of xBD and xFBD. The methods include BiT (Chen et al., 2022), ChangeFormer (Bandara and Patel, 2022), DAHiTra (Kaur et al., 2023) and MambaBDA (H. Chen et al., 2024b). The comparative results summarized in Table 1 demonstrate GAMSF's superior performance across the evaluation metrics, notably in building localization and damage classification. GAMSF achieves the highest overall F1-score (80.4 %) and Kappa score (77.9 %), outperforming DAHiTra by 1.7 % and 2.1 %, respectively. Its superior performance

Table 1
Statistical comparison of our proposed GAMSF with other methods on the combined dataset, ND is short of No damage, MiD and MaD represent Minor damage, Major damage, D is for Destroyed.

Method	Kappa	F1-score	$F1_{loc}$	$F1_{cls}$	Class F1-scores			
					ND	MiD	MaD	D
BiT	55.7	61.6	83.9	52.1	97.4	34.2	44.9	74.7
MambaBDA	48.6	54.4	83.2	40.9	97.3	25.9	46.9	62.3
ChangeFormer	68.5	72.8	83.5	68.2	98.6	49.2	65.9	77.0
DAHiTra	75.8	78.7	83.9	76.5	99.1	59.3	74.8	83.5
Our	**77.9**	**80.4**	**84.5**	**78.7**	**99.2**	**63.0**	**76.4**	**84.7**

results from effective integration of multi-scale feature extraction with Transformer-based temporal modeling.

For building localization, GAMSF attains an F1-score of 84.5 %, surpassing other methods due to precise multi-resolution feature fusion. In comparison, DAHiTra and MambaBDA demonstrate lower boundary accuracy due to less refined spatial modeling.

In damage classification, GAMSF achieves an F1-score of 78.7 %, significantly exceeding existing methods by at least 10.5 %. Its multi-scale architecture effectively differentiates between minor, major, and destroyed damage. Specifically, GAMSF improves minor damage classification by 3.7 % and major damage by 1.6 % compared to DAHiTra. For destroyed structures, GAMSF attains the highest F1-score (84.7 %), clearly differentiating complete destruction from severe damage. In contrast, ChangeFormer, which relies heavily on a change-detection-driven approach, struggles to differentiate between complete destruction and severe but non-total damage, leading to increased misclassification errors. MambaBDA, leveraging a Selective State Space Model (SSM) with self-attention, captures temporal dependencies but lacks the multi-scale feature fusion required to distinguish fine-grained damage variations. BiT, employing a single Transformer encoder, struggles with spatial coherence, leading to higher misclassification rates for minor and major damage categories.

Alongside quantitative evaluation, we conducted a qualitative analysis using four representative examples from the combined dataset (Fig. 8), comparing GAMSF with other methods. BiT (Fig. 8(e*)) delineates building boundaries clearly but struggles with minor and major damage classification due to its single-scale Transformer encoder and spatial inaccuracies from upsampling. ChangeFormer (Fig. 8(f*)) frequently mislabels minor damage as either no damage or major damage, as its change-detection approach lacks specific severity discrimination. DAHiTra (Fig. 8(h*)) also exhibits classification inconsistencies for minor damage, due to spectral discrepancies between pre- and post-disaster imagery. MambaBDA (Fig. 8(g*)) produces blurry boundaries and imprecise damage classifications, limited by insufficient multi-scale feature extraction.

In contrast, GAMSF (Fig. 8(d*)) demonstrates clear segmentation boundaries and accurate damage classification, effectively capturing fine-grained structural variations through multi-scale feature extraction. These results illustrate GAMSF's consistent superiority in accuracy and robustness across diverse disaster scenarios.

4.2. Transferability evaluation on the Ida-BD dataset

To assess the generalization capability of the proposed GAMSF framework, we conducted an evaluation on the Ida-BD dataset without any fine-tuning, using models trained solely on xBD and xFBD datasets.

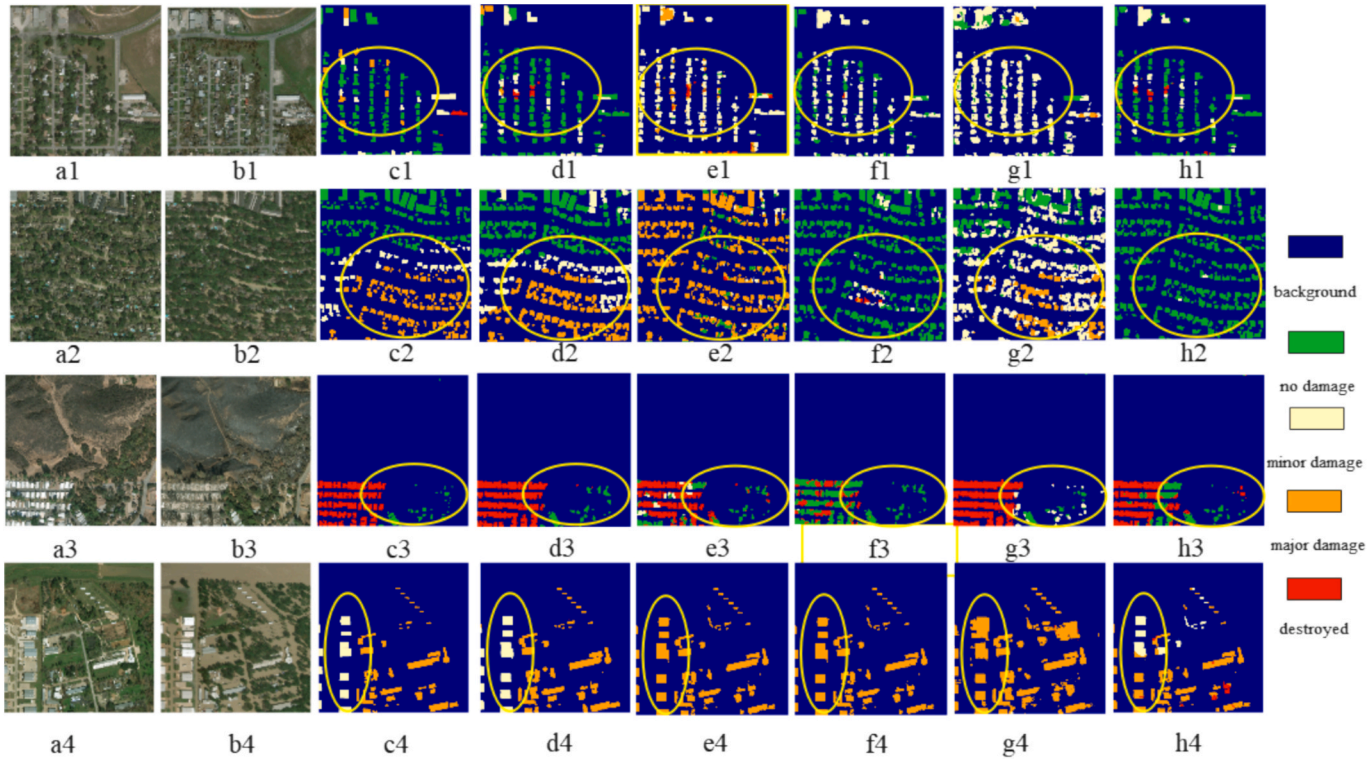


Fig. 8. Visual comparison of prediction results on the combined dataset: (a*) is the pre-disaster image; (b*) is the post-disaster image; (c*) is the ground truth; (d*) is the result of our method; (e*), (f*) (g*), and (h*) are the results of Bit, ChangeFormer and MambaBDA, and DAHiTra, respectively.

As shown in Table 2, GAMSF achieves the highest overall F1-score of 28.4 %, outperforming MambaBDA (25.9 %), BiT (23.0 %), ChangeFormer (23.3 %), and DAHiTra (22.8 %). Notably, GAMSF also achieves the highest Kappa score of 18.2 %, exceeding BiT's 16.6 % by 1.6 %, DAHiTra's 14.9 % by 3.3 %, and MambaBDA's 8.9 % by a substantial 9.3 %. This demonstrates GAMSF's superior generalization and consistency under varying disaster scenarios.

In terms of damage classification, GAMSF excels across all the severity levels. For minor damage, GAMSF achieves an F1-score of 37.7 %, significantly surpassing DAHiTra (24.0 %), MambaBDA (3.7 %) and ChangeFormer (1.9 %). This improvement demonstrates GAMSF's superior capability in distinguishing fine-grained structural damage features, which is critical for accurate post-disaster assessment. For major damage, GAMSF achieves 9.0 %, outperforming MambaBDA (4.5 %), DAHiTra (0.5 %), and ChangeFormer (1.1 %). The multi-scale feature fusion and Transformer-based temporal modeling within GAMSF allow for a more effective differentiation between partial and severe structural damage, leading to improved classification accuracy. However, ChangeFormer and DAHiTra exhibit spectral inconsistencies, resulting in frequent misclassifications between minor and major damage levels. Despite GAMSF's strong performance in damage classification, its building localization F1-score is slightly lower than BiT and DAHiTra, due to urban complexities, shadows, and geometric distortions in oblique views.

To further investigate GAMSF's effectiveness, we conducted a visual comparison on the representative samples from Ida-BD dataset in Fig. 9. It highlights GAMSF's more accurate segmentation maps, especially in distinguishing minor from major damage. The comparing methods often misclassify damage severity levels due to insufficient fine-grained feature extraction, spectral inconsistencies, and structural-change biases.

Specifically, BiT often incorrectly identifies damaged areas as undamaged due to limited fine-grained feature extraction, as shown in Fig. 9(e4). Similarly, ChangeFormer (Fig. 9(f*)) and MambaBDA (Fig. 9 (g*)) struggle with minor damage classification, tending toward over-classification by emphasizing structural changes without distinguishing damage severities. DAHiTra (Fig. 9(h*)) also demonstrates inconsistencies, particularly with minor damage identification, due to heavy reliance on subtle spectral variations.

Conversely, GAMSF demonstrates enhanced accuracy in damage classification and localization, effectively capturing fine-grained structural variations through multi-scale feature extraction. GAMSF shows clearer boundary delineation and better differentiation among minor, major, and destroyed buildings, as evident in Fig. 9(d*), highlighting its reliability in realistic disaster scenarios.

Nonetheless, GAMSF still encounters localization challenges in dense urban environments with occlusions from high-rise structures. Shadows and oblique viewing angles introduce geometric distortions and obscure critical structural details, complicating precise localization. Future research will thus focus on improving spatial alignment and incorporating adaptive feature fusion strategies to enhance localization accuracy in complex urban settings.

4.3. Ablation study

4.3.1. Module contribution

We conducted a comprehensive ablation study focused on the three core components: the Adaptive Attention (AA) module, the Gated Multi-Scale Feed-Forward Network (GMFFN), and the Multi-Scale Wavelet Fusion (MWF) module. Each component was systematically removed and replaced with a simpler alternative to assess its specific role in enhancing model performance. The fully integrated model served as the performance baseline for comparison. All the evaluation statistics are summarized in Table 3.

The fully integrated GAMSF model achieves an overall F1-score of 80.4 % and a Kappa score of 77.9 %, confirming its effectiveness. Omitting the AA module reduces the F1-score to 79.3 % and Kappa to 76.8 %. The F1-score for destroyed buildings drops from 84.7 % to 82.8 %, and for minor damage from 63.0 % to 62.2 %. These results highlight the importance of AA in selective feature weighting. Replacing the MWF module with a standard 3×3 convolution lowers the F1-score to 79.9 % and Kappa to 77.3 %, underscoring the importance of wavelet-based multiscale decomposition for capturing fine structural details. Substituting GMFFN with a conventional feed-forward network yields an F1-score of 80.0 % and Kappa of 77.6 %. Minor and major damage F1-scores also decline to 62.6 % and 76.0 % respectively, demonstrating GMFFN's contribution to contextual propagation. Removing any two modules further degrades performance. Without AA and GMFFN, the F1-score drops to 78.4 %. Eliminating all the three modules yields the lowest performance, with an F1-score of 77.1 % and Kappa of 75.3 %, reducing the model to a basic CNN. These results confirm that AA, MWF and GMFFN each contribute uniquely and synergistically to GAMSF's accuracy and robust generalization in complex post-disaster scenarios.

4.3.2. Module Layer Optimization across Stages

As shown in Fig. 2 and 3, Stages 2, 3, and 4 in both Phase 1 and 2 adopt a "four-four-two" arrangement of AA modules. To explore more cost-effective alternatives, we evaluated four different configurations: two-two-two, two-four-two, four-two-two, and four-four-four, with the results summarized in Table 4.

The baseline "four-four-two" setup delivers the best performance, with a Kappa score of 77.9 % and F1-score of 80.4 %, balancing localization and classification. It also achieves top scores for minor (63.0 %) and major damage (76.4 %), which are harder to classify. Among alternatives, two-four-two and four-four-four perform competitively. Four-four-four gives the highest destroyed-class F1-score (86.7 %) and best localization accuracy (84.6 %), but at higher computational cost. Two-four-two offers a good compromise (Kappa 77.6 %, F1-score 79.3 %) for resource-limited scenarios. In contrast, two-two-two and four-two-two show lower scores, especially for minor and major damage, underscoring the need for sufficient attention depth. These results confirm that "four-four-two" is the most effective and balanced configuration for both accuracy and efficiency.

Table 2
Statistical comparison of our proposed GAMSF with other methods on the Ida-BD dataset, ND is short of No damage, MiD and MaD represent Minor damage, Major damage, D is for Destroyed.

					Class F1-scores			
Method	Kappa	F1-score	$F1_{loc}$	$F1_{cls}$	ND	MD	MD	D
BiT	16.6	23.0	**75.0**	0.71	77.5	13.9	0.20	2.50
MambaBDA	8.9	25.9	70.5	6.78	**79.5**	3.70	4.50	**11.9**
ChangeFormer	13.4	23.3	72.3	2.30	78.7	1.90	1.10	2.90
DAHiTra	14.9	22.8	72.0	1.80	80.2	24.0	0.50	9.00
Our	**18.2**	**28.4**	73.9	**9.00**	75.2	**37.7**	**9.00**	4.00

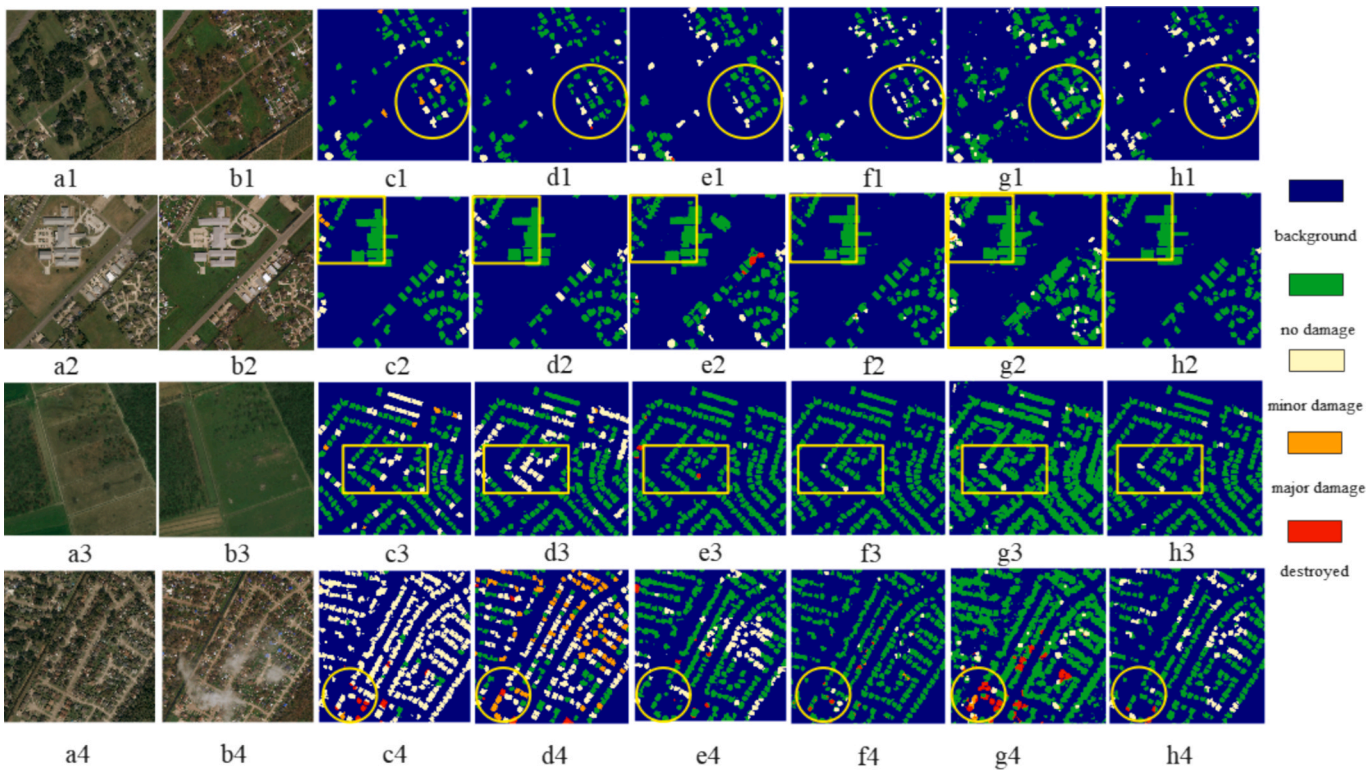


Fig. 9. Visual comparison of prediction results on the Ida-BD dataset: (a*) is a pre-disaster image; (b*) is a post-disaster image; (c*) is the ground truth; (d*) is the result of our method; (e*), (f*) and (g*) are the results of Bit, ChangeFormer, MambaBDA and DAHiTra, respectively.

Table 3
Performance comparison of GAMSF with different module removals, ND is short of No damage, MiD and MaD represent Minor damage, Major damage, D is for Destroyed cases.

Method	Kappa	F1-score	$F1_{loc}$	$F1_{cls}$	Class F1-scores			
					ND	MiD	MaD	D
Baseline	**77.9**	**80.4**	**84.5**	**78.7**	**99.2**	**63.0**	**76.4**	**84.7**
Baseline-AA	76.8	79.3	84.1	77.3	99.1	62.2	74.1	82.8
Baseline-MWF	77.3	79.9	84.3	78.0	99.1	62.8	75.0	84.0
Baseline-GMFFN	77.6	80.0	83.9	78.3	99.1	62.6	76.0	84.5
Baseline-AA-MWF	76.6	79.1	82.8	77.6	98.8	62.3	74.7	82.7
Baseline-AA-GMFFN	76.3	78.4	82.7	76.6	98.2	61.7	73.6	81.9
Baseline-AA-GMFFN-MWF	75.3	77.1	81.8	75.1	98.7	58.3	72.8	82.1

5. Discussion

5.1. Training dynamics and model convergence

Our proposed GAMSF was trained in two phases with distinct objectives. In Phase 1 (segmentation only), we used a combined focal-dice loss over 150 epochs. Training loss fell from 0.28 to 0.16 and validation loss stabilized at 0.14–0.15 (Fig. 10a). The higher training loss reflects strong regularization and broader scenario diversity. Phase 2 added damage classification, extending to 300 epochs. Training loss declined from 0.90 to 0.45, while validation loss, after early fluctuations, flattened (Fig. 10b). Early stopping was employed to prevent overfitting once validation ceased improving. These results demonstrate that GAMSF reliably converges under both single-task and multi-task regimes, underscoring the value of adaptive loss scheduling and convergence monitoring.

5.2. Computational efficiency and architectural trade-offs

The computational efficiency of our proposed GAMSF model were systematically evaluated against the comparing methods using four key indicators, number of parameters, inference speed (images per second), floating-point operations (FLOPs), and peak GPU memory consumption (in megabytes). As summarized in Table 5, our model achieves a

Table 4
Performance comparison for different AA module arrangements in Stages 2–4 across Phase 1–2.

Method	Kappa	F1-score	$F1_{loc}$	$F1_{cls}$	Class F1-scores			
					ND	MiD	MaD	D
Baseline	77.9	**80.4**	84.5	**78.7**	99.2	**63.0**	76.4	84.7
Two-two-two	76.7	78.7	82.5	77.2	99.0	60.8	74.6	84.3
Two-four-two	77.6	79.3	83.4	77.6	99.4	60.3	76.2	84.9
Four-two-two	77.1	78.9	82.8	77.3	99.4	60.5	75.6	83.9
Four-four-four	**78.0**	79.7	**84.6**	77.7	**99.4**	59.5	**76.5**	**86.7**

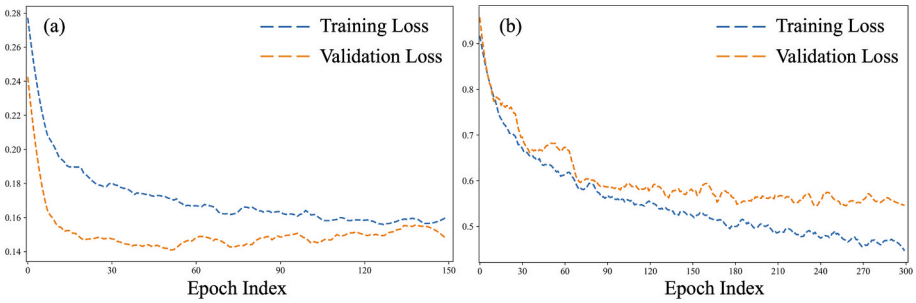


Fig. 10. Training loss curves across different phases. (a) Loss curve during Phase 1 training. (b) Loss curve during Phase 2 training.

favorable trade-off between performance and computational cost, making it suitable for real-world deployment scenarios that require both accuracy and efficiency.

Specifically, GAMSF has 17.7 million parameters, requires 121.68 GFLOPs, and runs at 8.71 images per second. Its memory usage is 2381 MB. In comparison, lightweight CNN models like BiT and DAHiTra run faster at 16.22 and 13.65 images per second, with 14.5 and 13.3 million parameters. However, they lack strong temporal and contextual modeling. On the other end, the Transformer-heavy MambaBDA uses 523.2 GFLOPs but runs only at 3.97 images per second. This makes it unsuitable for limited-resource environments. ChangeFormer runs the fastest at 17.87 images per second but uses 38.2 million parameters and 186.6 GFLOPs, which can be too demanding for real-time or embedded systems.

GAMSF reduces GFLOPs by nearly 35 % compared with MambaBDA and maintains half the parameter count of ChangeFormer. Despite being more efficient, GAMSF delivers rich features through gated multiscale fusion and dual temporal attention. This balance makes GAMSF both practical for deployment and reliable for large-scale post-disaster analysis.

5.3. Robustness and generalization across subsets

To assess stability across diverse conditions, we divided the combined test dataset into ten non-overlapping subsets and the Ida-BD dataset into five. For each subset we computed Kappa score, overall F1-score, and F1-score for each damage class (No Damage, Minor Damage, Major Damage, Destroyed). Fig. 11 presents boxplots of these metrics for BiT, ChangeFormer, MambaBDA, DAHiTra and our proposed GAMSF, with detailed statistics listed in Supplementary Tables 1–10 (see Supplementary File).

On the combined test subsets, GAMSF achieves the highest median Kappa score and overall F1-score and exhibits the tightest interquartile ranges, with paired t-tests confirming its superiority. Class-wise, GAMSF consistently leads in distinguishing Minor Damage and Major Damage categories where other methods show wider variability. On the Ida-BD subsets, although formal significance testing is not conducted due to the small sample count, GAMSF maintains top-ranked medians and low variance across all metrics, demonstrating its strong cross-domain generalization.

Table 5
Computational efficiency comparison of different models.

Method	Params(M)	Inference(im/s)	FLOPs(G)	Memory(M)
BiT	14.5	16.22	170.69	1017.27
MambaBDA	51.1	3.97	523.21	1176.38
ChangeFormer	38.2	17.87	186.61	1868.82
DAHiTra	13.3	13.65	159.41	2622.74
Our	17.7	8.71	121.68	2381.32

5.4. Limitations and Outlook

We illustrate our model's performance across diverse disaster scenarios using three close-up examples (Fig. 12). In moderately damaged hurricane areas (Fig. 12(d1)), the model identifies structures accurately but occasionally mislabels swimming pools as roofs. In sparse regions with vegetation interference (Fig. 12(d2)), natural elements sometimes cause missed detections. In dense, earthquake-affected urban scenarios (Fig. 12(d3)), GAMSF successfully distinguishes severely damaged structures amid complex debris, highlighting its robustness.

Despite these strengths, our model faces several limitations:

- The dual-temporal Transformer architecture of GAMSF demands substantial computational resources, restricting rapid deployment. Future work will explore model simplification and compression to improve efficiency without accuracy loss.
- Accurately differentiating minor from major damage remains challenging due to subtle visual distinctions and limited image resolution. Enhancing resolution, refining attention mechanisms, and incorporating damage-specific modules will help address these intermediate classifications.
- Cross-domain generalization also requires improvement. Performance declined on Ida-BD due to variations in spatial resolution, sensor types, and context. Diversifying training data, advanced augmentation, and domain adaptation strategies will enhance model robustness. Automating hyperparameter tuning could further improve reproducibility.
- Integrating complementary data sources (e.g., synthetic aperture radar, LiDAR, ground-level imagery) can mitigate occlusion, weather, and viewing-angle issues. Addressing ethical considerations, such as fair and transparent deployment, will guide GAMSF toward becoming a more reliable, equitable tool for disaster response and urban resilience.

6. Conclusion

This study introduced GAMSF, a novel framework for building damage assessment using high-resolution remote sensing imagery. By integrating multi-scale feature extraction, Transformer-based temporal learning, and adaptive attention mechanisms, GAMSF overcomes CNNs' limitations in global context modeling and Transformers' challenges in fine-grained feature extraction.

Experimental results show that GAMSF outperforms state-of-the-art methods, achieving a 2 % increase in F1-score and Kappa score, with 3 % and 1.5 % improvements in minor and major damage classification, respectively. It also enhances localization precision, reducing boundary detection errors and improving segmentation clarity. Evaluation on the Ida-BD dataset further confirms its superior transferability, with a 3 % F1-score improvement, demonstrating its robustness for cross-domain disaster assessment.

Despite its strengths, GAMSF faces challenges in urban environments, particularly with shadow occlusions and oblique views, which

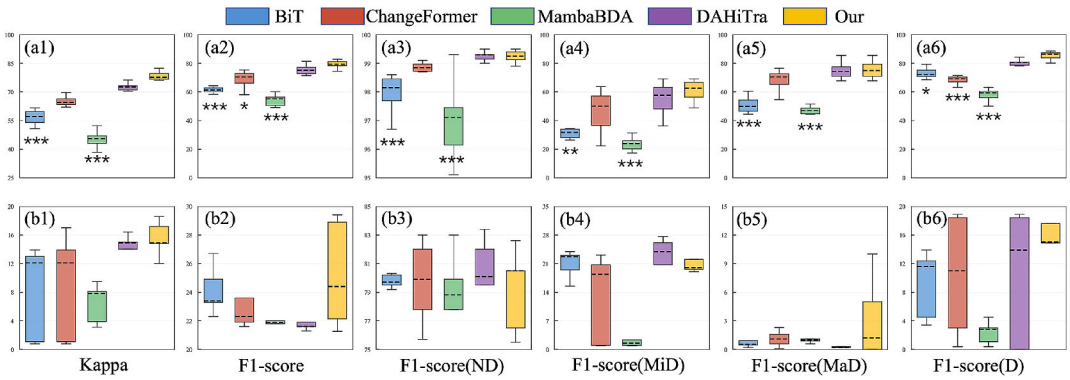


Fig. 11. Boxplots of evaluation metrics for BiT, ChangeFormer, MambaBDA, DAHiTra and GAMSF across test subsets: (a1–a6) the combined test subsets (n = 10) and (b1–b6) the Ida-BD subsets (n = 5). Metrics shown are Kappa score, overall F1-score, and class-specific F1-score for No Damage, Minor Damage, Major Damage and Destroyed. Asterisks indicate paired *t*-test significance between GAMSF and other methods on the combined test sets (*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$). Detailed statistics in Supplementary Tables 1–10.

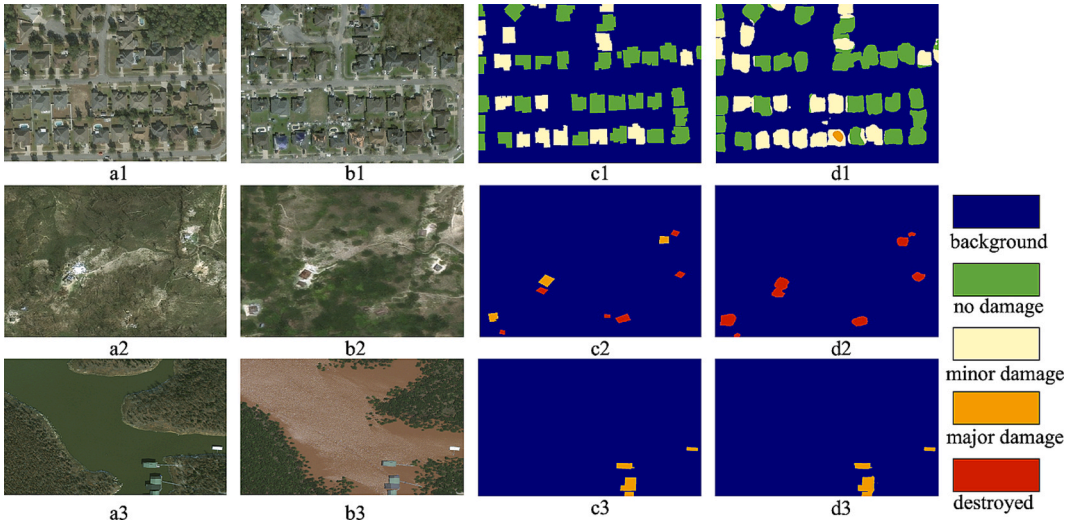


Fig. 12. Building damage assessment close-up demonstration: (a*) is a pre-disaster image; (b*) is a post-disaster image; (c*) is the ground truth; (d*) is the result of our method.

introduce minor localization errors. Future work will focus on spatial alignment improvements and multi-modal data integration (SAR, LiDAR) to further enhance model robustness and accuracy. Addressing these challenges will solidify GAMSF as a highly reliable framework for high-precision building damage assessment.

7. Code availability

The implementation code will be released upon the acceptance of the manuscript.

CRediT authorship contribution statement

Bo Yu: Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yao Sun:** Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Jiansong Hu:** Visualization, Validation, Resources. **Fang Chen:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Funding acquisition. **Lei Wang:** Writing – review & editing, Supervision, Resources, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research is financially supported by the National Natural Science Foundation of China (No. 42425103), the Provincial Special Funding for the Construction of Chenzhou National Sustainable Development Agenda Innovation Demonstration Zone (No. 2023sfq69), the Joint HKU-CAS Laboratory for iEarth (No. 313GJHZ2022074MI, E4F3050300), and the Youth Innovation Promotion Association, CAS (2022122). The work is also supported by CAS-TWAS Centre of Excellence on Space Technology for Disaster Mitigation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2025.104629.

Data availability

Data will be made available on request.

References

Anniballe, R., Noto, F., Scalia, T., Bignami, C., Stramondo, S., Chini, M., Pierdicca, N., 2018. Earthquake damage mapping: An overall assessment of ground surveys and VHR image change detection after L'Aquila 2009 earthquake. Remote Sens. Environ. 210, 166–178. https://doi.org/10.1016/j.rse.2018.03.004.

Bandara, W.G.C., Patel, V.M., 2022. A Transformer-Based Siamese Network for Change Detection. In: IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 207–210.

Barbosh, M., Sadhu, A., 2025. Wavelet packet transformation-based improved acoustic emission method for structural damage identification. Smart Mater. Struct. 34, 015036. https://doi.org/10.1088/1361-665X/ad9dc8.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, F., Wang, L., Wang, N., Guo, H., Chen, C., Ye, C., Dong, Y., Liu, T., Yu, B., 2024a. Evaluation of road network power conservation based on SDGSAT-1 glimmer imagery. Remote Sens. Environ. 311, 114273. https://doi.org/10.1016/j.rse.2024.114273.

Chen, H., Qi, Z., Shi, Z., 2022. Remote Sensing Image Change Detection with Transformers. IEEE Trans. Geosci. Remote Sensing 60, 5607514. https://doi.org/10.1109/TGRS.2021.3095166.

Chen, H., Song, J., Han, C., Xia, J., Yokoya, N., 2024b. ChangeMamba: Remote Sensing Change Detection With Spatiotemporal State Space Model. IEEE Trans. Geosci. Remote Sensing 62, 4409720. https://doi.org/10.1109/TGRS.2024.3417253.

Chen, Z., Wang, Y., Wu, J., Deng, C., Hu, K., 2021. Sensor data-driven structural damage detection based on deep convolutional neural networks and continuous wavelet transform. Appl. Intell. 51, 5598–5609. https://doi.org/10.1007/s10489-020-02092-6.

Ci, T., Liu, Z., Wang, Y., 2019. Assessment of the Degree of Building Damage Caused by Disaster Using Convolutional Neural Networks in Combination with Ordinal Regression. Remote Sens. (Basel) 11, 2858. https://doi.org/10.3390/rs11232858.

Da, Y., Ji, Z., Zhou, Y., 2022. Building Damage Assessment Based on Siamese Hierarchical Transformer Framework. Mathematics 10, 1898. https://doi.org/10.3390/math10111898.

Khankeshizadeh, E., Mohammadzadeh, A., Arefi, H., Mohsenifar, A., Pirasteh, S., Fan, E., Li, H., Li, J., 2024. A Novel Weighted Ensemble Transferred U-Net Based Model (WETUM) for Postearthquake Building Damage Assessment From UAV Data: A Comparison of Deep Learning- and Machine Learning-Based Approaches. IEEE Trans. Geosci. Remote Sens. 62, 1–17. https://doi.org/10.1109/TGRS.2024.3354737.

Entezami, A., Sarmadi, H., Behkamal, B., 2024. Short-term damage alarming with limited vibration data in bridge structures: A fully non-parametric machine learning technique. Measurement 235, 114935. https://doi.org/10.1016/j.measurement.2024.114935.

Ghimire, S., Guéguen, P., Giffard-Roisin, S., Schorlemmer, D., 2022. Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake. Earthq. Spectra 38, 2970–2993. https://doi.org/10.1177/87552930221106495.

Gomroki, M., Hasanlou, M., Chanussot, J., Hong, D., 2025. UNet-GCViT: a UNet-based framework with global context vision transformer blocks for building damage detection. Int. J. Remote Sens. 46, 2587–2610. https://doi.org/10.1080/01431161.2025.2454531.

Guo, E., Fu, X., Zhu, J., Deng, M., Liu, Y., Zhu, Q., Li, H., 2018. Learning to Measure Change: Fully Convolutional Siamese Metric Networks for Scene Change Detection. Remote Sens. (Basel) 10 (11), 1827. https://doi.org/10.3390/rs10111827.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M., 2019. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 10–17. https://doi.org/10.1109/CVPRW.2019.00012.

Hou, Y., Liu, K., Zhai, X., Chen, Z., 2024. Mbda-net: a building damage assessment model based on a multi-scale fusion network. SIViP 18, 9363–9374. https://doi.org/10.1007/s11760-024-03551-0.

Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2019. Squeeze-and-Excitation Networks. IEEE Trans. Pattern Anal. Mach. Intell. 42 (8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372.

Jamshidi, M., El-Badry, M., 2023. Structural Damage Identification from Acceleration Wavelet Data Using Convolutional Neural Networks. In: Walbridge, S., Nik-Bakht, M., Ng, K.T.W., Shome, M., Alam, M.S., El Damatty, A., Lovegrove, G. (Eds.), Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021. Springer Nature Singapore, pp. 457–469.

Kaur, N., Lee, C., Mostafavi, A., Mahdavi-Amiri, A., 2023. Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. Computer Aided Civil Eng 38, 2072–2091. https://doi.org/10.1111/mice.12981.

Liu, L., Wu, J., Li, D., Senhadji, L., Shu, H., 2019. Fractional Wavelet Scattering Network and Applications. IEEE Trans. Biomed. Eng. 66, 553–563. https://doi.org/10.1109/TBME.2018.2850356.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2018. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42 (2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826M.

Melamed, D., Johnson, C., Zhao, C., Blue, R., Morrone, P., Hoogs, A., Clipp, B., 2023. xFBD: Focused Building Damage Dataset and Analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 1238–1247. https://doi.org/10.1109/CVPRW59228.2023.00128.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: Proceedings of the 4th International Conference on 3D Vision (3DV), pp. 565–571. https://doi.org/10.1109/3DV.2016.79.

Natarajan, Y., Wadhwa, G., Ranganathan, P.A., Natarajan, K., 2023. In: Earthquake Damage Prediction and Rapid Assessment of Building Damage Using Machine Learning. IEEE, Goa, India, pp. 1–5. https://doi.org/10.1109/ICONAT57137.2023.10080586.

Nex, F., Duarte, D., Tonolo, F.G., Kerle, N., 2019. Structural Building Damage Detection with Deep Learning: Assessment of a State-of-the-Art CNN in Operational Conditions. Remote Sens. (Basel) 11, 2765. https://doi.org/10.3390/rs11232765.

Sarmadi, H., Entezami, A., Behkamal, B., De Michele, C., 2022. Partially online damage detection using long-term modal data under severe environmental effects by unsupervised feature selection and local metric learning. J. Civ. Struct. Heal. Monit. 12, 1043–1066. https://doi.org/10.1007/s13349-022-00596-y.

Shen, Y., Zhu, S., Yang, T., Chen, C., Pan, D., Chen, J., Xiao, L., Du, Q., 2022. BDANet: Multiscale Convolutional Neural Network with Cross-directional Attention for Building Damage Assessment from Satellite Images. IEEE Trans. Geosci. Remote Sensing 60, 5402114. https://doi.org/10.1109/TGRS.2021.3080580.

Sidharta, S., Warnars, H.L.H.S., Gaol, F.L., Soewito, B., 2022. In: Building Damage Assessment Using Deep Learning: Bibliometric Analysis. IEEE, Yogyakarta, Indonesia, pp. 1–6. https://doi.org/10.1109/ICITDA55840.2022.9971269.

Spencer, B.F., Hoskere, V., Narazaki, Y., 2019. Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. Engineering 5, 199–222. https://doi.org/10.1016/j.eng.2018.11.030.

Wang, C., Zhang, Y., Xie, T., Guo, L., Chen, S., Li, J., Shi, F., 2022. A Detection Method for Collapsed Buildings Combining Post-Earthquake High-Resolution Optical and Synthetic Aperture Radar Images. Remote Sens. (Basel) 14 (5), 1100. https://doi.org/10.3390/rs14051100.

Weber, E., Kané, H., 2020. Building Disaster Damage Assessment in Satellite Imagery with Multi-Temporal Fusion. IEEE Transactions on Geoscience and Remote Sensing 58 (8), 5949–5962. https://doi.org/10.1109/TGRS.2020.2976981.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: Convolutional Block Attention Module. arXiv preprint arXiv:1807.06521. Doi: 10.48550/arXiv.1807.06521.

Xing, Q., Wu, C., Chen, F., Liu, J., Pradhan, P., Bryan, B.A., Schaubroeck, T., Carrasco, L. R., Gonsamo, A., Li, Y., Chen, X., Deng, X., Albanese, A., Li, Y., Xu, Z., 2024. Intranational synergies and trade-offs reveal common and differentiated priorities of sustainable development goals in China. Nat Commun 15, 2251. https://doi.org/10.1038/s41467-024-46491-6.

Xu, M., Yoon, S., Fuentes, A., Park, D.S., 2023. A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. Pattern Recogn. 137, 109347. https://doi.org/10.1016/j.patcog.2023.109347.

Yu, B., Chen, F., Ye, C., Li, Z., Dong, Y., Wang, N., Wang, L., 2023. Temporal expansion of the nighttime light images of SDGSAT-1 satellite in illuminating ground object extraction by joint observation of NPP-VIIRS and sentinel-2A images. Remote Sensing of Environment 295, 113691. https://doi.org/10.1016/j.rse.2023.113691.

Yue, Z., Gao, F., Xiong, Q., Wang, J., Huang, T., Yang, E., Zhou, H., 2021. A Novel Semi-Supervised Convolutional Neural Network Method for Synthetic Aperture Radar Image Recognition. Cogn. Comput. 13, 795–806. https://doi.org/10.1007/s12559-019-09639-x.

Zaryabi, E.H., Kalantar, B., Moradi, L., Halin, A.A., Ueda, N., 2022. In: MSBDA-Net: Multi-Scale Siamese Building Damage Assessment Network. IEEE, Gold Coast, Australia, pp. 1–6. https://doi.org/10.1109/CSDE56538.2022.10089353.

Zhang, Y., Wang, Z., Luo, Y., Yu, X., Huang, Z., 2023. Learning Efficient Unsupervised Satellite Image-based Building Damage Detection. arXiv preprint arXiv:2312.01576. https://arxiv.org/abs/2312.01576.

Zheng, Z., Zhong, Y., Wang, J., Ma, A., Zhang, L., 2021. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. Remote Sens. Environ. 265, 112636. https://doi.org/10.1016/j.rse.2021.112636.