Rapid Building Damage Detection from Remote Sensing Images : a Novel Lightweight Network with Contrastive Learning

Wei Li, Guorui Ma, Lunjun Fan, Haiming Zhang

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - weili_rsai@whu.edu.cn; mgr@whu.edu.cn; 2023186190087@whu.edu.cn; zhanghaiming@whu.edu.cn

Keywords: Building Damage Detection, Rapid, Lightweight, Contrastive Learning, Remote Sensing Images.

Abstract

Accurate and timely building damage detection (BDD) is crucial for disaster emergency response. Although deep learning-based change detection methods have made significant progress in remote sensing, their practical application in disasters still faces two major challenges: (1) Existing high-accuracy models are typically computationally complex and difficult to deploy for real-time inference on edge devices.. (2) Model performance heavily relies on large amounts of annotated data, but disaster data are extremely scarce. To address these challenges, this paper proposes a novel lightweight Local-Global Interaction Network (LGINet) for efficient BDD. The core of LGINet is the proposed Local-Global Interaction Unit (LGIU), which achieves efficient fusion of detailed and contextual features through a dual-path architecture and channel-wise cross-attention mechanism. Furthermore, a Frequency Difference Enhancement Unit (FDEU) is proposed to generate more accurate damage features, and contrastive learning is employed to reduce the model’s sensitivity to weather conditions and its reliance on annotated data. Experimental results on the xBD and WBD datasets show that LGINet achieves F1-scores of 81.76% and 80.91%, respectively, with an inference speed of 47.83 FPS. It achieves the best balance between accuracy and efficiency, outperforming existing methods.

1. Introduction

Global climate change and geopolitical conflicts have increased the frequency of both natural and human-made disasters, posing severe challenges to human society (Li et al., 2026; Zhang et al., 2024). Following a disaster, accurate and timely acquisition of building damage information forms the critical decision-making foundation for effective life rescue and material distribution. In recent years, with the proliferation of high-resolution satellite and drone remote sensing, alongside breakthroughs in deep learning technologies, automated BDD based on remote sensing imagery has become a prominent research frontier. Mainstream approaches generally treat damage detection as a specialized change detection task, employing complex network architectures to improve the accuracy of identifying differences between pre- and post-event imagery (Olivar et al., 2024).

However, a significant application gap exists in current BDD research. Most studies focus primarily on improving model accuracy, failing to co-optimize computational efficiency and data utilization efficiency, which are critical for disaster emergency response. This research objective directly leads to two key bottlenecks. Computationally, complex models built for higher evaluation metrics typically have large parameter sizes and computational costs, making them difficult to deploy on edge devices with limited computing power, such as satellites, drones, or mobile terminals in disaster areas. This directly results in an inability to meet the timeliness requirements for rapid output of analytical results within hours after a disaster (Zheng et al., 2021). In terms of data, mainstream fully supervised learning paradigms heavily rely on large amounts of accurately annotated data. Such data are extremely difficult to acquire quickly for specific, sudden disaster events, and the labeling cost is high. This severely limits the practical utility of models in real disaster scenarios.

Furthermore, the two aforementioned critical bottlenecks are often addressed in isolation within existing research paradigms, lacking a co-designed approach. A large body of model lightweighting research primarily focuses on compressing network architectures and reducing parameters to improve inference speed. However, such efforts typically rely on the assumption of abundant annotated data. On the other hand, many studies aimed at reducing reliance on annotations (e.g., employing weakly supervised or self-supervised learning frameworks) may still be based on inherently complex models, without fully considering their deployability under extremely resource-constrained conditions. It is precisely this approach of treating computational efficiency and data utilization efficiency as two independent improvement paths that results in developed models struggling to simultaneously overcome the dual practical constraints of resources and data. Consequently, many algorithms that demonstrate excellent performance on standard benchmark datasets ultimately fail to bridge the "last mile" from the laboratory to the front lines of disaster relief.

To bridge this application gap, this paper proposes a lightweight network for rapid BDD. The network maintains accuracy while significantly improving inference speed and reducing reliance on annotated data. The main contributions of this paper are as follows: 1) A Local-Global Interaction Unit (LGIU) is proposed. Through an efficient dual-path architecture and a channel-wise cross-attention mechanism, it achieves efficient fusion of detailed and contextual features with minimal computational overhead. 2) A Frequency Difference Enhancement Unit (FDEU) is proposed. It performs adaptive filtering of change features in the frequency domain, effectively enhancing the frequency components associated with damage, thereby generating damage feature maps with stronger discriminative power. 3) A contrastive learning-based training strategy is incorporated. By constructing positive sample pairs containing weather perturbations, the model is guided to focus on the intrinsic damage features caused by the disaster. This not only suppresses the model's sensitivity to weather interferences such as illumination and clouds but also reduces its dependence on precise annotations. The proposed method is expected to provide reliable technical support for post-disaster rescue and reconstruction efforts.

2. Methodology

2.1 Network architecture

The overall design of LGINet is established to address the core trilemma in building damage detection, which involves balancing accuracy, inference speed, and data dependency under stringent computational constraints. The network adopts an encoder-decoder architecture as its foundational framework and integrates three key innovative modules to achieve this objective, as illustrated in Figure 1. First, the LGIU employs a parallel dual-path structure. One path utilizes lightweight operations to focus on capturing pixel-level details, while the other is dedicated to extracting regional contextual semantics. A novel channel-cross attention mechanism enables the adaptive and deep fusion of these two information streams, thereby constructing a powerful feature extractor with low computational overhead. Second, the FDEU acts as a feature purifier. It applies wavelet transform and learnable adaptive filtering to the initial feature difference, aiming to enhance the intrinsic patterns associated with structural damage while suppressing various types of irrelevant noise. Finally, a contrastive learning-based training strategy is introduced. By constructing positive sample pairs that incorporate synthetic weather perturbations, the strategy drives the network to learn feature representations that are robust to irrelevant imaging variations. This enhances the model's generalization capability and reduces its dependence on large volumes of precisely annotated data. The entire processing pipeline begins with a co-registered pair of pre- and post-disaster images. These images undergo multi-level feature extraction and fusion within the encoder. The resulting features are then subjected to difference computation and enhancement in the feature space. Finally, a lightweight decoder reconstructs a pixel-wise damage probability map. Consequently, this framework forms an efficient, practical, and end-to-end detection solution tailored for rapid emergency response.

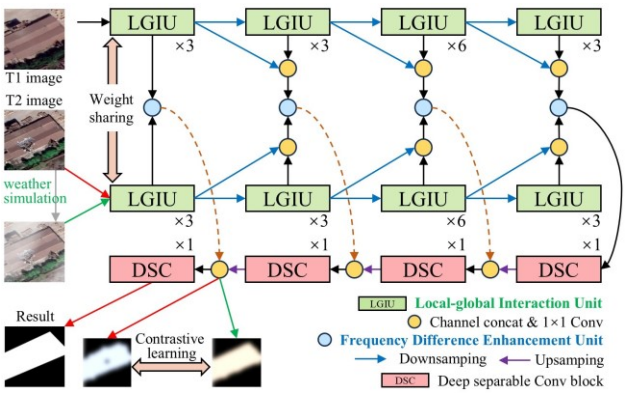

Figure 1. The architecture of LGINet.

2.2 Local-global interaction unit

The LGIU is the core component of the encoder, as illustrated in Figure 2. Its purpose is to accurately identify damaged buildings and comprehend their scene semantics with minimal computational cost. Specifically, the LGIU employs a parallel dual-path design to capture local details and global context, respectively, and achieves adaptive fusion of the two through a channel-wise cross-attention mechanism.

In the local detail branch, a simple 1×1 standard convolution is employed. This design aims to preserve the finest spatial details, such as building edges, textures, and roof structures, from high-resolution feature maps with minimal computational cost. These microfeatures are critical for accurately delineating damage boundaries. In the global context branch, a 7×7 depthwise separable convolution and an adaptive average pooling layer are processed in parallel, and their outputs are fused using a 1×1 convolution. The depthwise separable convolution substantially reduces the parameter burden introduced by large convolution kernels, while the large receptive field combined with pooling enables the network to capture broader scene semantics, such as the overall building outline and its relative positional relationship to surrounding roads or vegetation. This helps the model to assess building damage at a macro scale and mitigates false positives caused by local similarities.

Subsequently, to effectively fuse local details with global context, a channel-wise cross-attention mechanism is introduced. This mechanism first concatenates the features processed by the two branches along the channel dimension, and then employs a lightweight sub-network composed of two 1×1 convolutions to generate a set of channel attention weights. This set of weights is evenly split into detail attention weights and context attention weights. The detail attention weights are then applied to the global context features, recalibrating the global semantics from the perspective of local details, thereby emphasizing those contextual channels crucial for damage assessment. Conversely, the context attention weights are applied to the local detail features, using the semantic understanding of the global scene to filter and enhance local information, increasing the response of detail channels associated with damage. After this cross-modulation, the two sets of features are deeply fused, achieving profound interaction between local details and global context. Ultimately, this enables the network to adaptively allocate the contributions of local and global information at each level, forming a powerful feature extractor that not only focuses on microstructural changes but also comprehends macro-scene semantics.

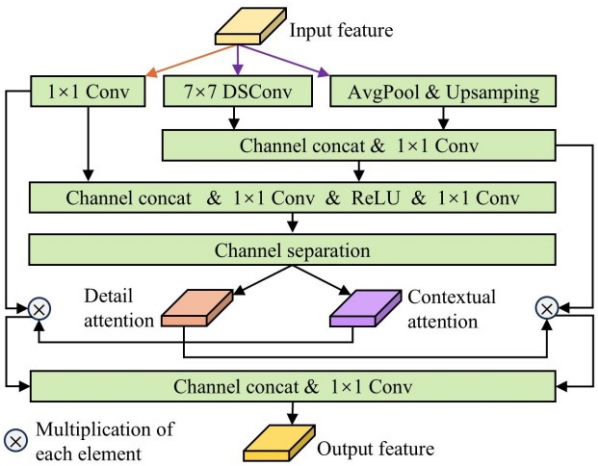

Figure 2. The structure of LGIU.

2.3 Frequency difference enhancement unit

After the encoder extracts features from the pre-disaster and post-disaster images respectively, a preliminary change feature map is first obtained through element-wise absolute difference operations. However, this feature map contains not only the essential damage information generated by structural collapse and fracturing of buildings but is also contaminated by substantial irrelevant change noise. This noise arises from factors such as weather variations (e.g., clouds, fog), illumination differences, seasonal vegetation changes, and

residual image registration errors. Spatially and spectrally overlapping with genuine damage features, this noise severely interferes with the model's judgment. To enhance the model's perception of intrinsic damage characteristics and effectively suppress interference, this paper proposes the FDEU, as illustrated in Figure 3. Departing from the traditional Fourier transform, this unit employs a two-dimensional discrete wavelet transform (2D DWT) for its superior time-frequency localization capability. This allows for the simultaneous analysis of features in both the frequency (scale) and spatial dimensions, enabling more precise localization and characterization of local high-frequency mutation patterns associated with building damage.

The unit first applies a two-dimensional discrete Wavelet transform (DWT) to the preliminary feature map, decomposing it into a low-frequency approximation subband (used to capture the global contours and smooth background of the image) and multiple high-frequency detail subbands (which respectively represent details in the horizontal, vertical, and diagonal directions). These high-frequency subbands primarily reflect abrupt changes in image details (such as edges and textures) and serve as the primary carriers of damage features. Subsequently, a lightweight learnable filtering network consisting of two consecutive 1×1 convolutional layers is specifically designed to process these high-frequency subbands. The first 1×1 convolutional layer is responsible for fusing information from subbands of different directions and extracting common feature patterns across channels. After introducing nonlinearity via the ReLU activation function, the second 1×1 convolutional layer outputs a smooth filtering weight within the range [0, 1] for each spatial location and feature channel, thereby forming a soft mask. Through end-to-end training, the network can automatically identify and amplify high-frequency components highly correlated with structural damage (such as linear cracks or block collapse), while suppressing random high-frequency responses caused by noise. Finally, the learned soft mask is used to adaptively modulate the initial variation features. The entire process is equivalent to a data-driven intelligent frequency-domain filter embedded within a neural network. It significantly cleans up the variation features, providing the subsequent decoder with clearer, more discriminative enhanced feature maps, thereby comprehensively improving the model’s detection accuracy and robustness.

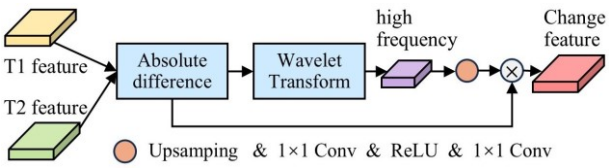

Figure 3. The structure of FDEU.

2.4 Lightweight decoder

The decoder is tasked with processing the highly abstracted and refined change feature maps generated by the encoder and the Feature Difference Enhancement Unit. Through a series of precise upsampling and feature fusion operations, it progressively restores the spatial resolution to match the original input image size, ultimately producing a pixel-wise damage probability map. In alignment with the network's overarching goal of lightweight design, the decoder forgoes complex structures in favor of an efficient, modular architecture. It consists of four cascaded decoding stages, each identical in structure. The core of each stage is a depthwise separable convolution block, a design chosen to maximize parameter and computational efficiency without compromising performance.

Each decoding stage executes a precisely defined sequence of operations to progressively enlarge the feature map and incorporate spatial detail. First, the stage applies 2× bilinear upsampling to its low-resolution input feature map. Bilinear interpolation provides a deterministic, parameter-free method for initial spatial expansion. This is immediately followed by a 3×3 depthwise separable convolution applied to the upsampled features. This operation serves a dual purpose: it refines local features that may have been blurred during upsampling, and it fuses them with multiscale features delivered via skip connections from the corresponding encoder level. These skip connections provide rich spatial details and intermediate semantic information captured during encoding, which are crucial for accurately reconstructing the boundaries and shapes of damaged structures. The depthwise separable convolution decomposes the standard convolution into a depthwise and a pointwise convolution, enabling efficient cross-channel information integration and local spatial context modeling with minimal computational overhead. This convolution is typically followed by batch normalization and a ReLU activation function to stabilize training and introduce non-linearity.

By repeating this "upsample-fuse-refine" process four times, the spatial scale of the feature map is gradually restored. At the network's terminus, the output of the final decoding stage is fed into a 1×1 standard convolution layer. This layer acts as the final projection head, mapping the multi-channel features to a single-channel output. A Sigmoid activation function is then applied to this single-channel map, compressing each pixel's value to the [0, 1] interval to generate the final pixel-wise building damage probability map. The value at each pixel intuitively represents the confidence of building damage at that location, completing the end-to-end mapping from abstract features to a concrete semantic segmentation result. The entire decoder design embodies the principles of lightness and efficiency. By leveraging skip connections for feature fusion, it ensures the reconstruction of clearly bounded and accurately localized damage detection results while maintaining strict control over model complexity.

2.5 Training strategy

To effectively train LGINet, a hybrid learning strategy combining supervised segmentation loss and a self-supervised contrastive auxiliary task is employed. The primary supervised objective directly addresses the pixel-wise building damage segmentation task. Given the severe foreground-background class imbalance inherent in change detection scenarios, where damaged pixels often constitute only a small fraction of the image, a composite loss function is utilized. This function is a linear combination of Cross-Entropy loss and Dice loss. The Cross-Entropy loss provides stable and precise gradient signals for each individual pixel, ensuring accurate per-pixel classification. Concurrently, the Dice loss, which computes the overlap between the prediction and the ground truth, directly optimizes for region-based similarity. It is particularly effective in mitigating the model's bias towards the dominant background class, which consists of undamaged areas, by focusing on the spatial coherence and volume of the predicted damaged regions. This dual-loss mechanism ensures the model achieves both accurate boundary delineation and robust region coverage for the damaged structures.

To further enhance the model's robustness against pervasive and confounding imaging variations, such as clouds, haze, and illumination shifts, a self-supervised contrastive learning auxiliary task is integrated into the training paradigm (Jiang et al., 2023). For each real pre-disaster and post-disaster image pair in a batch, multiple synthetic weather-perturbed versions of the post-disaster image are generated through data augmentation techniques. These augmented images serve as positive samples, representing the same underlying geographic scene under different environmental conditions. Both the original real image pair and its generated positive pairs are then processed through the shared-weight encoder of LGINet. The resulting high-level change feature representations are optimized using the InfoNCE (Noise-Contrastive Estimation) loss. This objective function operates on the feature representations, pulling the change features derived from the same scene (i.e., the real pair and its weather-augmented counterparts) closer together in the latent embedding space, while simultaneously pushing apart the representations from different, unrelated scene pairs within the batch. Consequently, this auxiliary task drives the encoder to distill weather-invariant, semantically meaningful feature representations. It teaches the model to ignore irrelevant atmospheric noise and to focus on the structural changes genuinely caused by damage. This approach not only improves the model's generalization capability to unseen weather conditions but also reduces its dependency on large volumes of perfectly curated, noise-free training data.

3. Dataset

To evaluate the BDD performance of the proposed method, experiments were conducted on two datasets, corresponding to natural disaster and human conflict scenarios, respectively. Both datasets are annotated with complete building damage outlines, without damage level grading. This annotation approach aligns precisely with the core objective of rapid BDD in this study, as the key to rapid response lies in first determining whether damage has occurred, rather than in detailed damage level assessment. Omitting the grading step simplifies task complexity, thereby providing timely damage assessment results for disaster emergency response. The details of the data are as follows:

xBD dataset (Gupta et al., 2019): This dataset contains 11,034 pairs of satellite remote sensing images (1024×1024 pixels, with a resolution better than 0.8 meters), annotated for building damage caused by 19 types of disasters worldwide, including earthquakes, floods, and wildfires. In the experiments, the images were cropped to 256×256 pixels, and were ultimately divided into training, validation, and test sets in a 7:1:2 ratio.

WBD dataset (Zhang et al., 2024): This dataset comprises 3,634 pairs of remote sensing images (256×256 pixels, with a resolution better than 0.6 meters), covering building damage caused by geopolitical conflicts in locations such as Mariupol and Deir ez-Zor. The data are similarly divided into training, validation, and test sets in a 7:1:2 ratio.

4. Results

4.1 Experimental setup

To evaluate the genuine performance of LGINet, we selected four advanced remote sensing image change detection models as benchmarks, including DMINet (Feng et al., 2023), GASNet(Zhang et al., 2023), BIT(Chen et al., 2022), and ChangeFormer(Bandara and Patel, 2022). In terms of training setup, the models were trained with the Adam optimizer. The initial learning rate was set to 0.0002, adjusted with a cosine annealing scheduler (Johnson et al., 2023). The number of epochs was set to 100, and the batch size was set to 8 based on GPU memory constraints. All experiments were conducted on a computing platform equipped with an NVIDIA RTX 8000 GPU.

4.2 Results of xBD and WBD

Figure 4 provides a visual comparison of the building damage detection results between LGINet and four state-of-the-art change detection models, namely DMINet, GASNet, BIT, and ChangeFormer, on both the xBD and WBD datasets. The overall visual comparison demonstrates that LGINet generates prediction results closest to the ground truth across various typical damage scenarios.

Specifically, in the natural disaster cases from the xBD dataset, Case A presents large-scale, contiguous building damage, while Case B involves localized, fine structural collapse and fragmentation. The comparative models tend to produce holes or discontinuous predictions within the damaged areas in Case A, and exhibit over-smoothing or blurred boundaries at the edges between damaged regions and the background. Under the cluttered background interference in Case B, the comparative models show noticeable missed detections for local, fine damage or confuse it with similarly textured backgrounds. In contrast, LGINet's prediction mask maintains better regional integrity and spatial coherence in Case A. In Case B, it more accurately captures local, subtle damage structures, delineates boundaries more sharply and clearly, and significantly reduces false alarms from background noise.

In the conflict damage scenarios represented by the WBD dataset, Case C corresponds to dense damage in high-density urban areas, and Case D involves complex, irregular building collapse patterns. In these high-density, structurally complex urban settings, the predictions of the comparative models are prone to two types of issues: first, misclassifying intact but texturally complex buildings as damaged; and second, causing fragmented predictions within actual damaged areas, breaking contiguous damage into multiple pieces. LGINet demonstrates stronger discriminative and generalization capabilities in such challenging scenarios. Its prediction results not only more completely cover various damage targets within high-density areas, effectively suppressing false positives for dense but intact buildings, but also better preserve the integrity and structural consistency of irregular damage patterns. Consequently, its outputs are visually more robust and accurate.

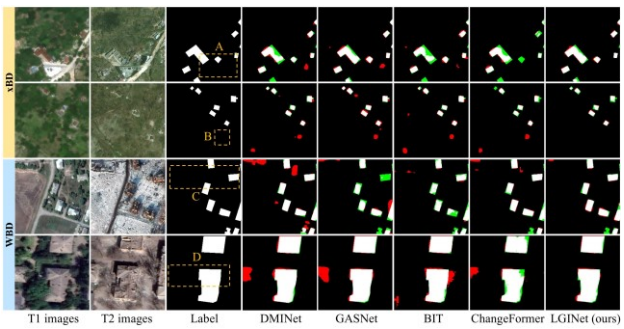

Figure 4. The building damage detection results of all models. Red indicates false positives, and green indicates false negatives.

Table 1 presents the accuracy, and table 2 presents the complexity, and efficiency of all models. It can be observed that LGINet achieves significant improvements in both accuracy and efficiency. In terms of detection accuracy, LGINet attains F1-scores of 81.76% and 80.91% on the xBD and WBD datasets, respectively. Compared to the second-best models (GASNet and DMINet), this represents improvements of 1.65% and 1.07%, respectively. Regarding complexity and efficiency, its parameter count (3.02M) and computational load (7.71G FLOPs) are considerably lower than those of all other models. Meanwhile, its inference speed (FPS) reaches the highest value of 47.83, reflecting an increase of at least 46.2%. Overall, LGINet achieves the best balance between accuracy and efficiency.

N/A	F1-score (%)	
	xBD	WBD
DMINet	79.58	79.84
GASNet	80.11	79.23
BIT	77.25	77.87
ChangeFormer	77.34	78.08
LGINet (Ours)	81.76	80.91

Table 1. The accuracy of all models.

N/A	Complexity and efficiency		
	Params (M)	FLOPs (G)	FPS
DMINet	6.24	14.55	32.71
GASNet	23.52	289.97	11.19
BIT	11.50	106.14	15.06
ChangeFormer	24.30	44.54	24.84
LGINet (Ours)	3.02	7.71	47.83

Table 2. The complexity, and efficiency of all models.

4.3 Ablation experiment

Table 3 presents the contributions of each core component to the performance of LGINet. First, when only the FDEU and contrastive learning (CL) are used while the LGIU is disabled, the model achieves F1-scores of 80.68% and 78.86% on the xBD and WBD datasets, respectively. When only the LGIU and CL are used, the F1-scores are 80.24% and 79.73%. Both sets of results are significantly superior to those obtained with CL disabled (F1-scores of 79.56% and 77.12%, respectively). This directly demonstrates the critical and independent role of the CL strategy in enhancing the model's discriminative capability and robustness. Finally, when the LGIU, FDEU, and CL are all enabled, the model achieves optimal performance on both datasets (F1-scores of 81.76% and 80.91%, respectively). This indicates that the LGIU, FDEU, and CL form an effective complementary relationship, working synergistically to achieve the best detection performance.

Each core component			F1-score (%)	
LGIU	FDEU	CL	xBD	WBD
×	√	√	80.68	78.86
√	×	√	80.24	79.73
√	√	×	79.56	77.12
√	√	√	81.76	80.91

Table 3. The ablation experiment results for LGINet.

4.4 Verification of real-world disaster cases

To evaluate the BDD capability of LGINet in real disaster scenarios, we used satellite imagery from the Maui wildfire in Hawaii, USA, on August 8, 2023, for assessment (Zhang et al., 2024). The images have a spatial resolution of 0.8 meters and dimensions of 5120×7168 pixels. The pre-disaster image is sourced from Google Earth dated March 24, 2023, and the post-disaster image was captured by the Jilin-1 satellite on August 13, 2023. In the experiment, the image pair was cropped into 256×256 pixels, and 10% of the data was randomly selected for training. Additionally, we compared the performance with GASNet, which achieved the best results on the xBD dataset. As shown in Figure 5, LGINet achieved an F1-score of 70.69% with an inference time of 11.78 seconds. In contrast, GASNet achieved an F1-score of 66.41% and required 19.21 seconds for processing. The results indicate that, owing to its lightweight design, LGINet demonstrates superior accuracy and efficiency in the real-disaster assessment. Thus, it is better suited to meet the urgent demands of emergency response for rapid and accurate assessment.

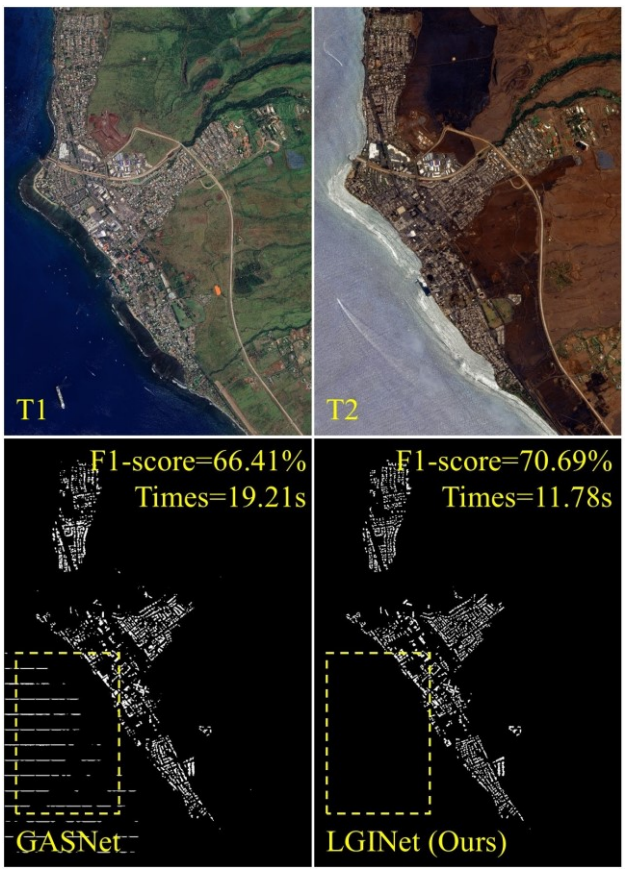

Figure 5. The BDD results of Hawaii wildfires.

5. Conclusion

This paper proposes LGINet, a lightweight network for rapid building damage detection in emergency response. The core innovations of the method lie in its Local-Global Integration Unit, which achieves efficient fusion of detailed and contextual semantic features with low computational cost. And Its Frequency Domain Enhancement Unit, which employs adaptive filtering in the wavelet domain to significantly improve the discriminative power of change features; and the introduction of a contrastive learning strategy to enhance model robustness

against irrelevant imaging variations. Comprehensive experiments show that LGINet outperforms existing state-of-the-art methods in both accuracy and inference speed while maintaining very low parameter counts and computational complexity. It demonstrates high detection accuracy and strong generalization capability on real-world imagery from disaster and conflict scenarios. These results validate the effectiveness of the proposed framework in balancing accuracy, efficiency, and practicality, offering a viable technical solution for real-time damage assessment in emergency response.

Despite the progress achieved with LGINet, this study has certain limitations that point to directions for future work. First, although the lightweight design improves efficiency, its deployment performance on extremely resource-constrained edge devices requires further validation in real-world settings. Second, while the training data covers multiple disaster types, the model’s ability to represent certain extreme damage patterns or rare geographical environments may be limited; incorporating more extensive and challenging global datasets is needed. Furthermore, the current method focuses on pixel-wise binary damage detection; extending it to fine-grained assessment of damage severity levels represents a valuable future direction. Future research will proceed along these three avenues.

Acknowledgements

This work was supported by the Guangxi Science and Technology Major Project (AA22068072).

References

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. Presented at the IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 207–210.

Chen, H., Qi, Z., Shi, Z., 2022. Remote Sensing Image Change Detection With Transformers. IEEE Transactions on Geoscience and Remote Sensing 60, 1–14.

Feng, Y., Jiang, J., Xu, H., Zheng, J., 2023. Change Detection on Remote Sensing Images Using Dual-Branch Multilevel Intertemporal Network. IEEE Transactions on Geoscience and Remote Sensing 61, 1–15.

Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeev, S., Heim, E., Doshi, J., Lucas, K., Choset, H., Gaston, M., 2019. Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 10–17.

Jiang, F., Gong, M., Zheng, H., Liu, T., Zhang, M., Liu, J., 2023. Self-Supervised Global–Local Contrastive Learning for Fine-Grained Change Detection in VHR Images. IEEE Transactions on Geoscience and Remote Sensing 61, 1–13.

Johnson, O.V., Xinying, C., Khaw, K.W., Lee, M.H., 2023. ps-CALR: Periodic-Shift Cosine Annealing Learning Rate for Deep Neural Networks. IEEE Access 11, 139171–139186.

Li, W., Ma, G., Zhang, H., Chen, P., Wang, D., Chen, R., 2026. Multi-scenario building change detection in remote sensing images using CNN-Mamba hybrid network and consistency enhancement learning. Expert Systems with Applications 298, 129843.

Olivar, F.R.M., Reyes, R.B., Nagai, M., 2024. Identifying Damaged Areas Caused by the Noto Peninsula 2024 Earthquake Using Change Detection from Optical Satellite Images. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences X-3–2024, 277–284.

Zhang, H., Ma, G., Fan, H., Gong, H., Wang, D., Zhang, Y., 2024. SDCINet: A novel cross-task integration network for segmentation and detection of damaged/changed building targets with optical remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing 218, 422–446.

Zhang, R., Zhang, H., Ning, X., Huang, X., Wang, J., Cui, W., 2023. Global-aware siamese network for change detection on remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing 199, 61–72.

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 Sensing of Environment 265, 112636.