

Article

BDHE-Net: A Novel Building Damage Heterogeneity Enhancement Network for Accurate and Efficient Post-Earthquake Assessment Using Aerial and Remote Sensing Data

Jun Liu ^{1,2}, Yigang Luo ¹, Sha Chen ^{3,*}, Jidong Wu ⁴  and Ying Wang ²

¹ College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; liujun_eq@sina.com (J.L.); m210911523@st.shou.edu.cn (Y.L.)

² National Earthquake Response Support Service, Beijing 100049, China; wy_wangying2019@126.com

³ National Disaster Reduction Center of China, Ministry of Emergency Management of China, Beijing 100124, China

⁴ State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China; wujidong@bnu.edu.cn

* Correspondence: chensha@ndrcc.org.cn

Abstract: Accurate and efficient post-earthquake building damage assessment methods enable key building damage information to be obtained more quickly after an earthquake, providing strong support for rescue and reconstruction efforts. Although many methods have been proposed, most have limited effect on accurately extracting severely damaged and collapsed buildings, and they cannot meet the needs of emergency response and rescue operations. Therefore, in this paper, we develop a novel building damage heterogeneity enhancement network for pixel-level building damage classification of post-earthquake unmanned aerial vehicle (UAV) and remote sensing data. The proposed BDHE-Net includes the following three modules: a data augmentation module (DAM), a building damage attention module (BDAM), and a multilevel feature adaptive fusion module (MFAF), which are used to alleviate the weight deviation of intact and slightly damaged categories during model training, pay attention to the heterogeneous characteristics of damaged buildings, and enhance the extraction of house integrity contour information at different resolutions of the image. In addition, a combined loss function is used to focus more attention on the small number of severely damaged and collapsed classes. The proposed model was tested on remote sensing and UAV images acquired from the Afghanistan and Baoxing earthquakes, and the combined loss function and the role of the three modules were studied. The results show that compared with the state-of-the-art methods, the proposed BDHE-Net achieves the best results, with an F1 score improvement of 6.19–8.22%. By integrating the DBA, BDAM, and MFAF modules and combining the loss functions, the model's classification accuracy for severely damaged and collapsed categories can be improved.

Keywords: building damage assessment; deep learning; building damage heterogeneity; UAV and remote sensing data



Citation: Liu, J.; Luo, Y.; Chen, S.; Wu, J.; Wang, Y. BDHE-Net: A Novel Building Damage Heterogeneity Enhancement Network for Accurate and Efficient Post-Earthquake Assessment Using Aerial and Remote Sensing Data. *Appl. Sci.* **2024**, *14*, 3964. <https://doi.org/10.3390/app14103964>

Academic Editor: Nicholas Vassiliou Sarlis

Received: 31 March 2024

Revised: 3 May 2024

Accepted: 6 May 2024

Published: 7 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Earthquakes, as highly destructive natural disasters, can severely impair societal development and jeopardize the safety of human lives and property. Therefore, it is crucial to evaluate the extent of structural damage in buildings promptly and accurately following an earthquake. This assessment holds significant value in supporting government emergency response efforts and facilitating efficient rescue operations [1]. Remote sensing images are characterized by wide coverage, high spatial resolution, and rich spectral information, while UAV data have the advantages of high resolution, flexibility, and low

operating costs. Consequently, the utilization of remote sensing imagery and UAV imagery has become prevalent in studies pertaining to building damage assessment.

The conventional approaches for building damage assessment include visual interpretation and field investigation, which are highly accurate but time-consuming and labor-intensive, especially when the affected area is large [2]. Change detection, based on pre- and post-earthquake remote sensing images, is also an effective approach to assessing building damage. Gong [3] used an object-oriented classification method to extract building images before and after an earthquake and used a change detection method to analyze changes in the buildings. Although this method makes full use of multi-temporal features and obtains better accuracy, it requires pre-disaster images corresponding to the same location at the same time. In addition, this method can usually only extract images of collapsed buildings and cannot meet the need for a refined classification of building damage. Although this method fully utilizes multi-temporal features of the images, it cannot determine the damage level of buildings.

The progress in deep learning technologies has led to the development of various neural network models, including recurrent neural networks (RNNs) [4], convolutional neural networks (CNNs) [5], and graph neural networks (GNNs) [6]. In particular, CNNs have shown promising potential in image classification and semantic segmentation. Therefore, they have been widely used in building damage assessment research. For example, Duarte [7] used three different CNN-based feature fusion methods based on residual connections and dilated convolutions to assess building damage at different resolutions. The results showed that when multiple-resolution feature maps were fused, and the feature information from intermediate layers of each resolution-level network was considered, better accuracy and localization capability were demonstrated. Chowdhury [8] developed RescueNet, a high-resolution dataset tailored for natural disaster analysis. This dataset is meticulously annotated at the pixel level and categorizes features into 11 distinct types, such as debris, water, buildings, vehicles, roads, trees, ponds, and sand. Moreover, RescueNet includes four unique labels for segmenting buildings based on varying degrees of damage. Xie [9] proposed a network that considers heterogeneous features of damaged buildings. This network utilizes a local-global context attention module, which extracts features from multiple directions. The test results indicated that compared with excellent deep learning models, the proposed method achieved a joint intersection-over-union (IOU) growth of 0.03–7.39%. Gupta [10] developed an end-to-end model for building segmentation that integrates a unique location-aware loss function. This function combines binary cross-entropy loss with foreground-selective category cross-entropy loss to classify damage. This model outperforms those utilizing conventional cross-entropy loss in terms of building segmentation and damage classification. Additionally, it demonstrates enhanced generalization across diverse geographical regions and types of disasters. Shen [11] proposed a two-stage CNN designed for assessing building damage. Initially, a U-Net is employed to identify building locations. Subsequently, the second stage utilizes a dual-branch, multi-scale U-Net architecture as its core framework. Pre-disaster and post-disaster images were input into the network, and a cross-directional attention module was employed to explore correlations among these images. Zheng [12] developed the ChangeOS framework for semantic change detection, utilizing a deep object localization network to accurately identify building structures for damage assessment. Comparative studies demonstrated that ChangeOS outperformed existing methods in terms of speed and accuracy, also showing enhanced generalization capabilities for anthropogenic disasters. Shafique [13] proposed a new deep learning algorithm that replaces the upsampling layer in U-Net3+ with a sub-pixel count convolutional layer, thus improving the problem of poor segmentation including irrelevant change information and inconsistent boundaries present in building change detection. Bai [14] suggested employing a U-net convolutional network for the semantic segmentation of building damage in high-resolution remote sensing imagery. The effectiveness of the U-Net model was evaluated by comparing it with the deep residual U-Net model, with the 2011 Tohoku earthquake tsunami serving as a benchmark. Rudner [15] presented a new

method for fast and accurate disaster loss segmentation, which integrates multi-resolution, multi-sensor, and multi-temporal satellite imagery within a CNN framework. Hong [16] proposed a deep learning-based Multi-View Stereo (MVS) model for reconstructing 3D models of earthquake-damaged buildings, aimed at assisting in building damage assessment tasks. Hong [17] presented EBDC-Net (Earthquake Building Damage Classification Network). The network comprises a feature extraction module and a damage classification module and was designed to augment semantic information and differentiate various damage levels. Günen [18] introduced a new framework that accelerates building detection in ultra-high-resolution images. This approach employs the maximum correlation minimum redundancy method for feature selection, resulting in the generation of five distinct feature sets.

While previous studies on building damage classification have yielded valuable insights, numerous challenges still need resolution. Firstly, the issue of data sample category imbalance poses a significant problem. In most seismic hazards, significantly fewer buildings are damaged or collapsed by earthquakes than buildings that are undamaged or slightly damaged. This imbalance causes the model to be biased towards learning a larger number of categories during the training process, thus affecting the ability to recognize a smaller but important number of categories (e.g., severely damaged and collapsed houses). Secondly, current approaches to assessing building damage rely on semantic segmentation models that are standard in computer vision. However, these methods fail to consider the specific characteristics of building damage, resulting in suboptimal assessment outcomes. For instance, after a building collapses, debris, tiles, and other objects may be scattered around, and the collapse direction can be random. Directly applying existing semantic segmentation models may result in incomplete feature extraction and misclassification. Furthermore, there is a significant resolution difference between UAV images and satellite remote sensing images. Consequently, traditional convolutional neural networks may struggle to effectively capture the complete information of a house, leading to decreased accuracy in house classification [19]. Therefore, the objective of this study is to design an enhanced deep learning model that incorporates multi-directional convolution, data augmentation strategies, and deep and shallow feature fusion. The methodology presented in this paper is designed for post-earthquake building damage assessment utilizing UAV and satellite remote sensing imagery. The primary contributions of this study can be summarized as follows:

1. A novel data augmentation module (DAM) is presented, different from the commonly used data augmentation methods such as rotation and size scaling, by integrating oversampling techniques and label polygon dilation techniques, which can improve the situation where the model weights are biased towards a large number of categories.
2. A building damage attention module (BDAM) is proposed to enhance the accuracy of severely damaged and collapsed categories by considering the randomness of the collapse direction in collapsed buildings following earthquakes, as well as the heterogeneity in texture features in damaged houses and the ground.
3. A multilevel feature adaptive fusion module (MFAF) is introduced to search for optimal parameters on feature maps of different scales, focusing on extracting contour integrity information among houses of different sizes and enhancing the model's sensitivity to diverse house sizes.

2. Materials and Methods

2.1. Data Sources

This study utilizes datasets composed of post-earthquake images from two distinct locations. The first dataset includes UAV images captured after a magnitude 4.5 earthquake in Baoxing County, Ya'an City, Sichuan Province, on 1 June 2022. The second consists of post-earthquake remote sensing images acquired after a 6.2 magnitude earthquake on 22 June 2022 in Khost Province, Afghanistan [20]. The original remote sensing images of Afghanistan can be downloaded (license required) from the following URL: <https://>

[//resources.maxar.com/](https://resources.maxar.com/), accessed on 27 June 2022. Initially, a region of interest (ROI) was extracted from the original images. Subsequently, the region of interest (ROI) was divided into patches of various sizes based on the image resolution and the architectural characteristics of the buildings. Every patch was then evenly adjusted to dimensions of 512×512 pixels. Table 1 provides detailed information on the datasets.

Table 1. Image detail information on the datasets.

Dataset	Source	Resolution	Year	Number of Samples
Afghanistan Dataset	WorldView3	≤ 0.31 m	2022	5601
Baoxin Dataset	UAV	≤ 0.1 m	2022	4020

All buildings were categorized into four levels as follows: intact, slightly damaged, severely damaged, and collapsed. The criteria for judging the four categories were as follows: slightly damaged buildings are those where the roof exhibits uneven color tones due to partial tile loss, resulting in visible leaking areas. Severely damaged buildings have the characteristics of the overall outline of the building remaining intact, but one side of the house shows partial wall collapse, forming local ruins in the imagery. Fallen debris displays significant variations in brightness and color tones, indicating a collapse degree of 10–50%.

The outline of the collapsed buildings is incomplete, and the roof texture and color tones appear asymmetrical. There is a noticeable contrast between the collapsed corners of the walls and the roof texture in terms of brightness and color tones, indicating a collapse degree of over 50%. The proportion of intact, slightly damaged, severely damaged, and collapsed is 52%, 22%, 16%, and 10%, respectively. Table 2 shows an example of four building damage levels in the UAV and remote sensing images. Lastly, Table 3 illustrates the disparity in resolution between drone imagery and satellite remote sensing imagery. Notably, within the demarcated region highlighted by the red box, a significant discrepancy exists in the pixel area occupied by a single residential structure in the two image types [21].

Table 2. Examples of four building damage levels in UAV and remote sensing images.

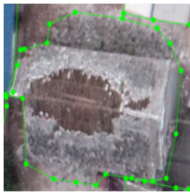
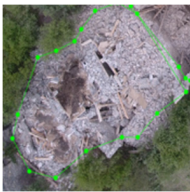
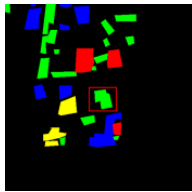
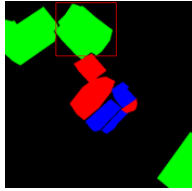
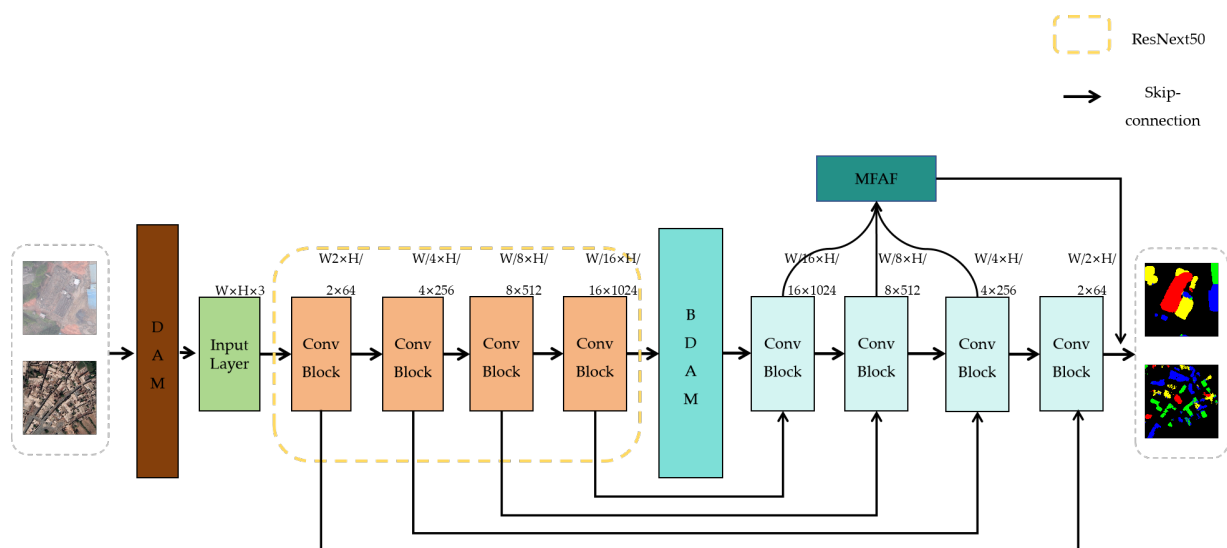
Image Category	Intact	Slightly Damaged	Severely Damaged	Collapsed
UAV				
Remote sensing				

Table 3. Examples of UAV and satellite remote sensing images of the same size.

Image Category	Image	GT
UAV		
Remote sensing		

2.2. Methods

Figure 1 depicts BDHE-Net, the proposed framework for classifying building damage. The framework incorporates a data augmentation module, a building damage attention module, and a multilevel feature adaptive fusion module, which is employed to suppress the model weights likely to favor the intact category, enhance the extraction of building damage features from the model, and strengthen the extraction of housing integrity information at different scales, respectively.

**Figure 1.** The structure of the proposed BDHE-Net framework.

2.2.1. Data Augmentation Module Based on Oversampling Techniques and Label Polygon Dilation Techniques

In earthquake disasters, it is common for the number of intact houses to be much higher than the number of damaged houses, with a serious imbalance in the sample size of categories. When traditional convolutional neural networks are used directly for building damage assessment, the model predictions may be biased towards a large number of categories, thus resulting in a decline in the classification accuracy of damaged houses. Therefore, we designed a data augmentation module to enhance its ability to perceive the damaged house category. The data augmentation module consists of the following two components: oversampling and label polygon dilation techniques.

Considering that the number of severely damaged and collapsed types is significantly less than that of the slightly damaged categories, we judge the input image and copy it once if there is a slightly damaged type, and twice if there is a severely damaged category or collapse category. Therefore, the model can learn multiple times from these images to better extract features of damaged houses.

Secondly, when a building suffers severe damage or collapse, there will be debris around the house, which is significantly different from intact or slightly damaged buildings. This characteristic is critical for building damage classification. Therefore, we used labeled polygon inflation techniques based on an inflation factor k . For the severely damaged category, k was set to 5–10% of the labeled area, while for the collapsed category, k was to 10–15% of the labeled area because of more debris and gravel around collapsed buildings than in the severely damaged category. In this way, the characteristics of the severely damaged and collapsed categories can be amplified. Figure 2 shows a comparison between before and after dilation.

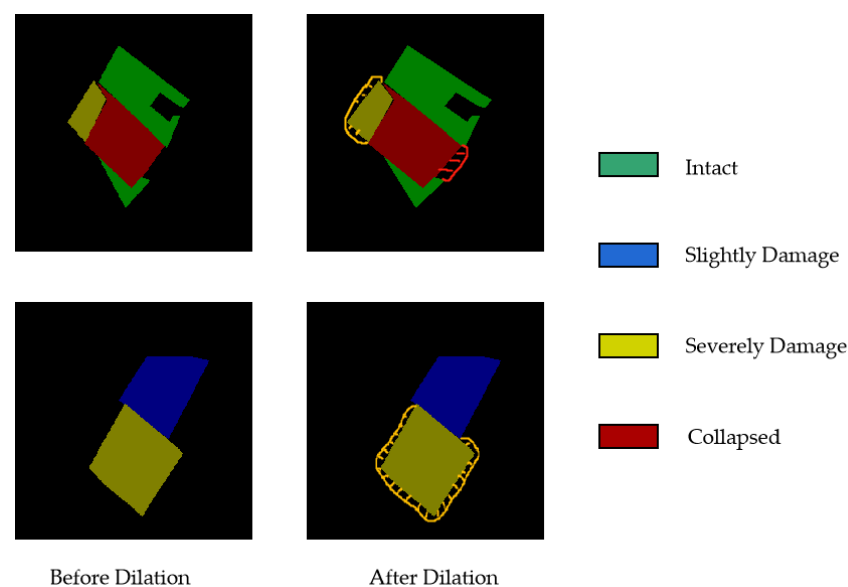


Figure 2. Label dilation technology.

2.2.2. Building Damage Attention Module Based on Dilated Convolution and Direction Convolution

The building damage attention module consists of an ASPP (Atrous Spatial Pyramid Pooling) module [22] and a DFE (Direction Feature Extraction) module [23]. Firstly, since our dataset includes both UAV and satellite imagery, the difference in the resolution scale between the two types of images can be very large. Therefore, an ASPP module was initially employed within the feature extraction module to enhance the network's capability to extract features at various scales. The design of the ASPP module is inspired by DeepLabv3 [24] and DeepLabv3+ [25]. Because of the different resolutions between UAV images and satellite remote sensing images, this module enhances the receptive field of the convolutional kernel by incorporating an adaptable dilation rate, enabling the capture of features at varying scales. The augmentation allows the network to carry out comprehensive mapping across different scales effectively. Furthermore, the entire feature map is transformed into a fixed-length feature vector that retains global information while minimizing spatial dimensions. This modification strengthens the network's capacity to derive both local and global information from images.

Secondly, in addition to the debris accumulation around damaged houses, there is evidence of random collapse directions and distinct textural characteristics compared with the surrounding ground. To take advantage of this distinctive feature, we incorporated a Direction Feature Extraction (DFE) module for extracting direction features. As shown in

Figure 3, the DFE module comprises two branches. In the first branch, a 1×1 convolution is initially applied to reduce dimensionality, followed by directional convolutions that employ four 1×1 convolutions to perform convolutions in the up, down, left, and right directions. This process generates four local spatial feature weight maps. Subsequently, the weight maps are concatenated to form a single feature aggregation weight map encompassing local spatial features from all four directions. The second branch incorporates global average pooling, 1×1 convolution, batch normalization (BN), and rectified linear unit (ReLU) activation, which collectively yields global spatial information weight maps. Finally, the local spatial information weight map and the global spatial information weight map are integrated using the element-wise multiplication (“mul” operation). F_i represents the input features processed by the convolution layer, and F_j represents the learned damaged buildings feature map. By incorporating building damage heterogeneity features such as random collapse directions, unique texture attributes compared to the surrounding ground, and debris accumulation around the affected structures, the network gains a deeper understanding and enhanced utilization of these features. Consequently, this leads to improved accuracy and robustness in the pixel-level classification of building damage.

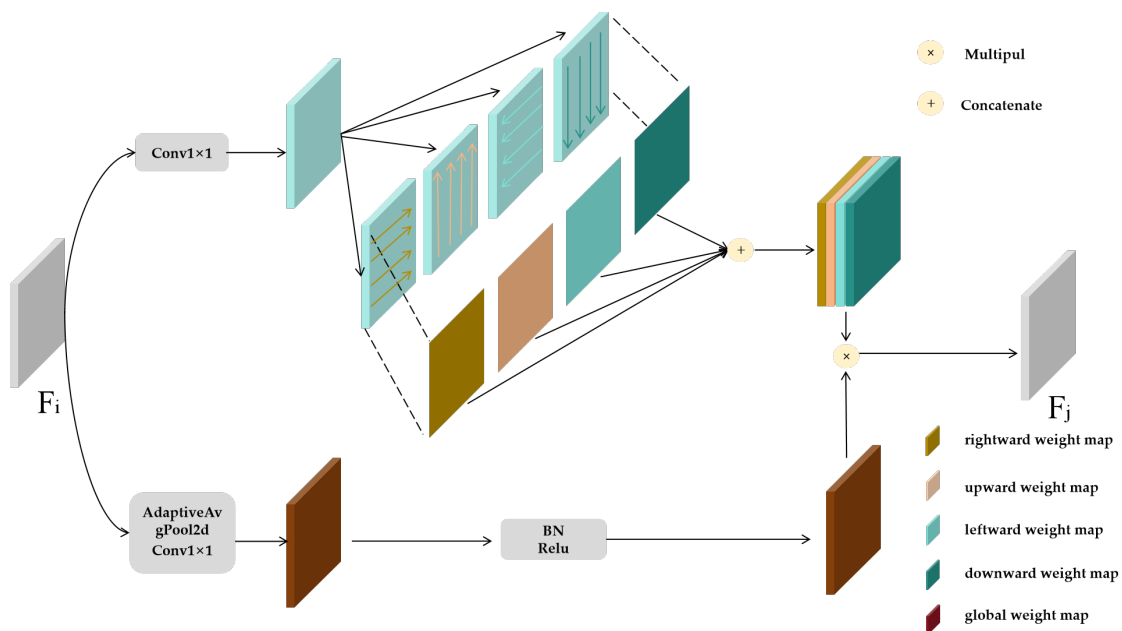


Figure 3. DFE module structure.

2.2.3. Multilevel Feature Adaptive Fusion Module Based on Multi-Scale Fusion

The design of the MFAF incorporates information about the integrity of a house at different scales [26]. Figure 4 displays the construction of the MFAF module. F_i represents the input features processed by the convolution layer and F_j represents the learned house integrity feature map. Global statistics of feature maps at different resolutions are captured through three different strategies. The first strategy involves the simultaneous modification of channel number and resolution by applying a 3×3 convolution layer with a stride of 2. This downsamples the input feature map resolution to $1/2$. The second strategy compresses the input feature map by employing a 1×1 convolution, preserving image information at the original scale. The third strategy increases the resolution to $2 \times$ by employing a 1×1 convolution and interpolation. This design captures global statistics of feature maps across various resolutions, aiding in the detection of buildings of different sizes in the input. Following the scaling operation, the output from each branch is processed through a fully connected network, resulting in a series of weighted feature vectors. The number of channels in the feature weight vector, denoted as l , represents the feature weight vector diagram for the three branches. Subsequently, the SoftMax operation is applied to fix

the values of the feature matrix between 0 and 1. These values are then used to multiply the original feature map, weighting the house features to amplify the impact of complete information. The features at the corresponding level l are fused as follows:

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l} \quad (1)$$

where $x_{ij}^{n \rightarrow l}$ denotes the feature vector at position i, j on the feature map, adjusted from level n to level l . The terms α_{ij}^l , β_{ij}^l , and γ_{ij}^l refer to the spatial importance weights from three different levels to level l , which the network adaptively learns. Additionally, y_{ij}^l indicates the i, j vector of the output feature map between channels.

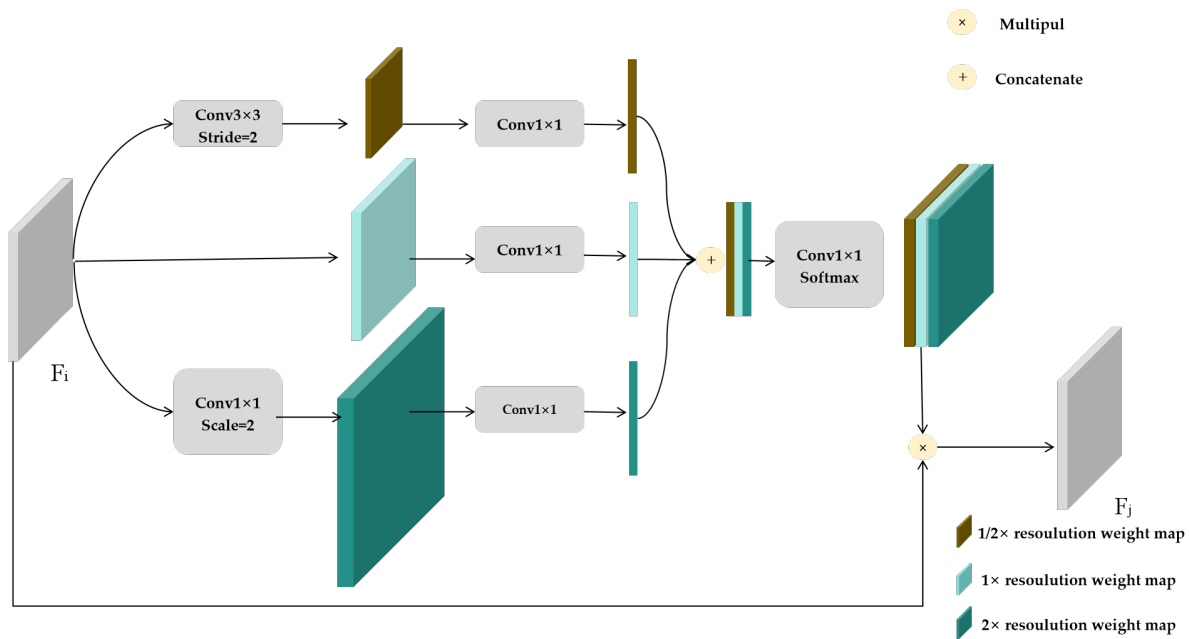


Figure 4. MFAF module structure.

2.2.4. Combination Loss Function Based on Focal and Dice Loss

Cross-entropy loss is a common loss function in deep learning. However, cross-entropy is a global loss function that only considers pixel-level loss, so when cross-entropy loss is used in our post-earthquake building damage dataset, it will lead to the model weights being biased in favor of categories with a high number of sample categories, whereas in the post-earthquake dataset, the number of samples in the generally largely intact category is much larger than in the other categories. As a result, it can easily lead to larger errors when extracting damaged houses in the models. Unlike cross-entropy loss, since Dice loss emphasizes the overlap between predicted results and true labels [27], focal loss focuses more on the strategy of hard-to-predict samples by adjusting the sample weights [28]. Therefore, both Dice loss and focal loss enable the model to concentrate more significantly on samples from a limited number of classes, thereby mitigating the issue of model weights disproportionately favoring intact classes to some extent. To address the imbalance where the number of damaged houses in the post-earthquake building damage dataset is significantly lower than that of undamaged houses, a combined loss function comprising Dice loss and focal loss, alongside a cross-entropy loss function, is proposed. This combined loss function is utilized to optimize the model during the training process.

The Dice coefficient is a metric utilized to quantify the degree of overlap between predicted and ground truth regions in semantic segmentation tasks. It serves as an evaluation measure, assessing model performance by comparing the similarity between predicted results and actual labels. The Dice coefficient ranges between 0 and 1, where a value of

1 indicates a perfect overlap, meaning the predicted region is identical to the real region, while a value of 0 implies no overlap, indicating that the prediction has no relevance. The Dice coefficient offers several advantages over other metrics. Firstly, it exhibits regional correlation, meaning the loss of a given pixel is not solely dependent on its predicted value but also on the values of neighboring pixels. This characteristic enables the Dice coefficient to provide a balanced assessment of the data, particularly in scenarios where there is class imbalance. Secondly, since larger Dice coefficients indicate better performance, this metric remains unaffected by data imbalance. In neural network training, the goal is often to minimize the loss function to optimize the model. However, as larger Dice coefficients are desirable, it is possible to utilize the complementary value of the Dice coefficient (1 minus the Dice coefficient) as the formulation for the Dice loss function, as depicted in Equation (2). This approach aims to prioritize an enhancement in the Dice coefficient during training, thereby improving the model's performance in semantic segmentation tasks.

Furthermore, Equation (3) presents the focal loss, which serves as an additional component to the balanced cross-entropy loss function to adjust the weights of samples that are easy or challenging to classify. The focal loss introduces an adjustable focusing parameter, γ , to decrease the weight of easily separable samples and emphasize the importance of difficult-to-categorize samples. By utilizing a γ value greater than 1, the focal loss reduces the weight assigned to conveniently separable samples, directing the model's attention towards more challenging instances. Conversely, an γ value less than 1 increases the weight of easily separable samples, ensuring a balanced consideration of all categories. This mechanism directs the model to focus more on samples that are challenging to classify, thereby enhancing the classification accuracy of minority categories.

$$dice\ loss = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (2)$$

$$focal\ loss = -\alpha(1 - p)^\gamma \log(p) \quad (3)$$

where X and Y represent the ground truth and predict_mask of segmentation, α represents the category weight, and γ represents the weight of difficult-to-distinguish samples.

3. Experiment and Analysis

3.1. Experimental Environment

An experiment was conducted using a single Nvidia GeForce RTX 3090 24G (GPU) and an Intel(R) Xeon(R) Gold 6248R CPU @ 3.00 GHz (CPU). The training and testing were all implemented on a Windows 10 system. The network model was constructed using the PyTorch1.9 deep learning framework [29], which is widely utilized in the field. PyTorch, an open-source machine learning framework that enjoys extensive adoption, provides a diverse range of pre-trained models and libraries, enabling time and computational resource savings.

3.2. Evaluation Metrics

To objectively assess the segmentation performance of the model for building damage assessment and enable effective comparisons with different approaches, four widely employed evaluation metrics for semantic segmentation, namely, P (Precision), R (Recall), the $F1$ score, and IOU (Intersection over Union), are adopted to evaluate the effectiveness of the introduced model. The evaluation is conducted using the following formulas:

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (6)$$

$$IOU = \frac{TP}{TP + FP + FN} \quad (7)$$

where TP , TN , FP , and FN represent the counts of true positive, true negative, false positive, and false negative samples for the respective classes, respectively.

3.3. Experimental Parameter Setting

The neural network's internal parameters are derived from iterative model training, while certain hyperparameters require manual configuration before training. In our experiments, the training settings included a batch size of 12, the Adam optimizer, an initial learning rate of 0.0001, and a weight decay of 0.00001. We use the parameters trained on ResNext-50 in ImageNet as the initial weights to improve the stability and generalization ability of the model [30]. The number of iterations was set to 30. As depicted in Figure 5, the model's loss value exhibited a decline from the initial 0.924 to 0.165 after 10 epochs of training. Subsequently, during the course of 30 training epochs, the model displayed a propensity to stabilize and converge.

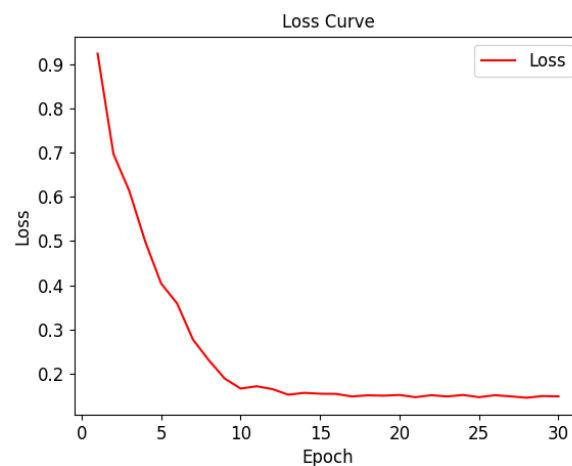


Figure 5. Training loss curves.

3.4. Comparative Analysis of Splitting Performance

In this paper, three classic semantic segmentation models including DeepLabv3+, ResNet-50 [31], and U-Net [32] are used for comparison with our proposed model. The comparison results for the satellite remote sensing images from the Afghanistan earthquake are presented in Figure 6, while those for the UAV images from the Baoxing earthquake are displayed in Figure 7. The figure reveals an incorrect categorization of damaged houses by the U-Net model. The ResNet-50 model fails to provide clear boundary information for different building categories. In addition, the DeepLabv3+ model does not predict the outline of houses completely. In contrast, the BDHE-Net model proposed in this paper not only comprehensively extracts the contours of buildings but also effectively differentiates between buildings with different damage classes.

To further assess the model's performance, a detailed analysis is presented in Figure 8 with satellite images and UAV images. The first row is the satellite remote sensing image, and the original image, zoomed in the red box, is shown in the second column. The main body of the house within the red box belongs to the slightly damaged category. However, the U-Net model misidentifies it as severely damaged. The ResNet-50 model struggles to differentiate between slight and severely damaged, categorizing it as a mix of both. The Deeplabv3+ model classifies it as severely damaged and fails to recognize the complete and regular shape of the main building. In contrast, our BDHE-Net model accurately identifies it as slight damage and recognizes the complete and regular shape of the main building.

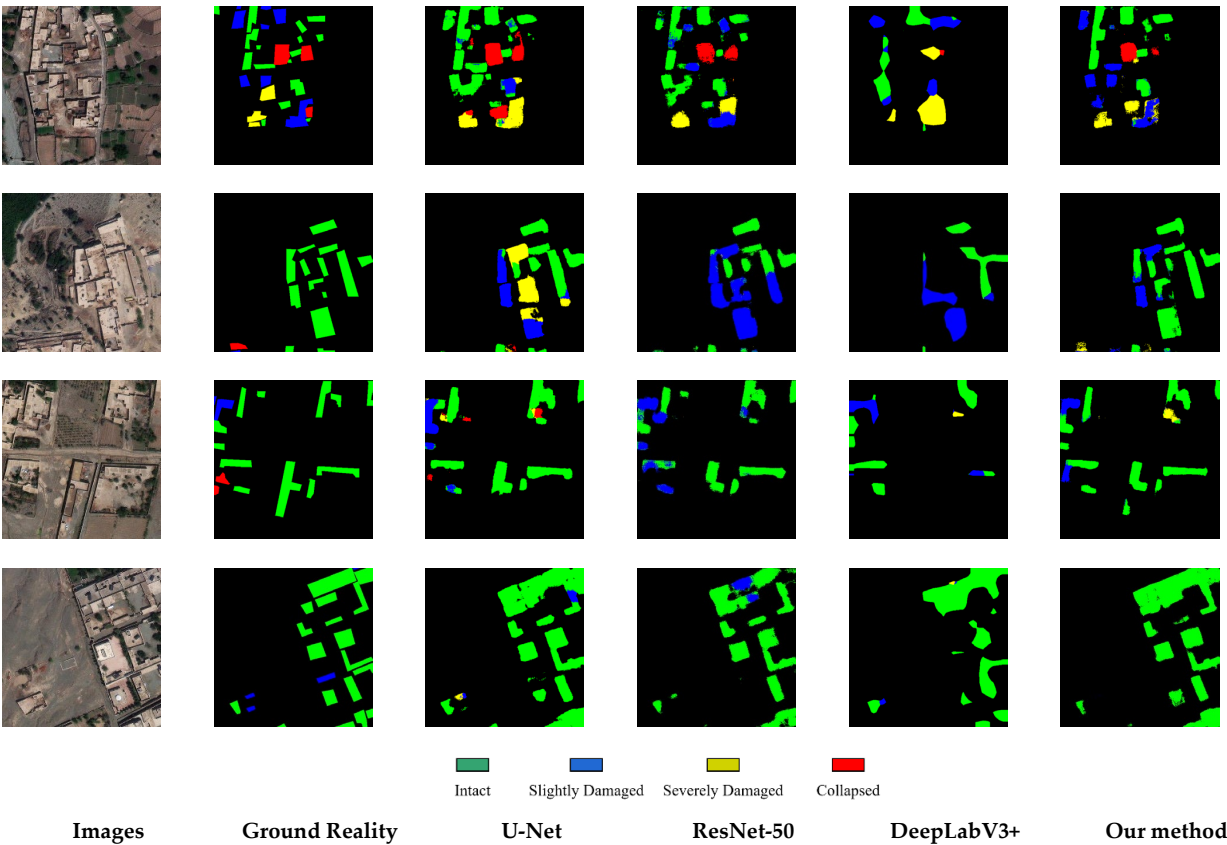


Figure 6. Comparison of different segmentation models for the Afghanistan earthquake.

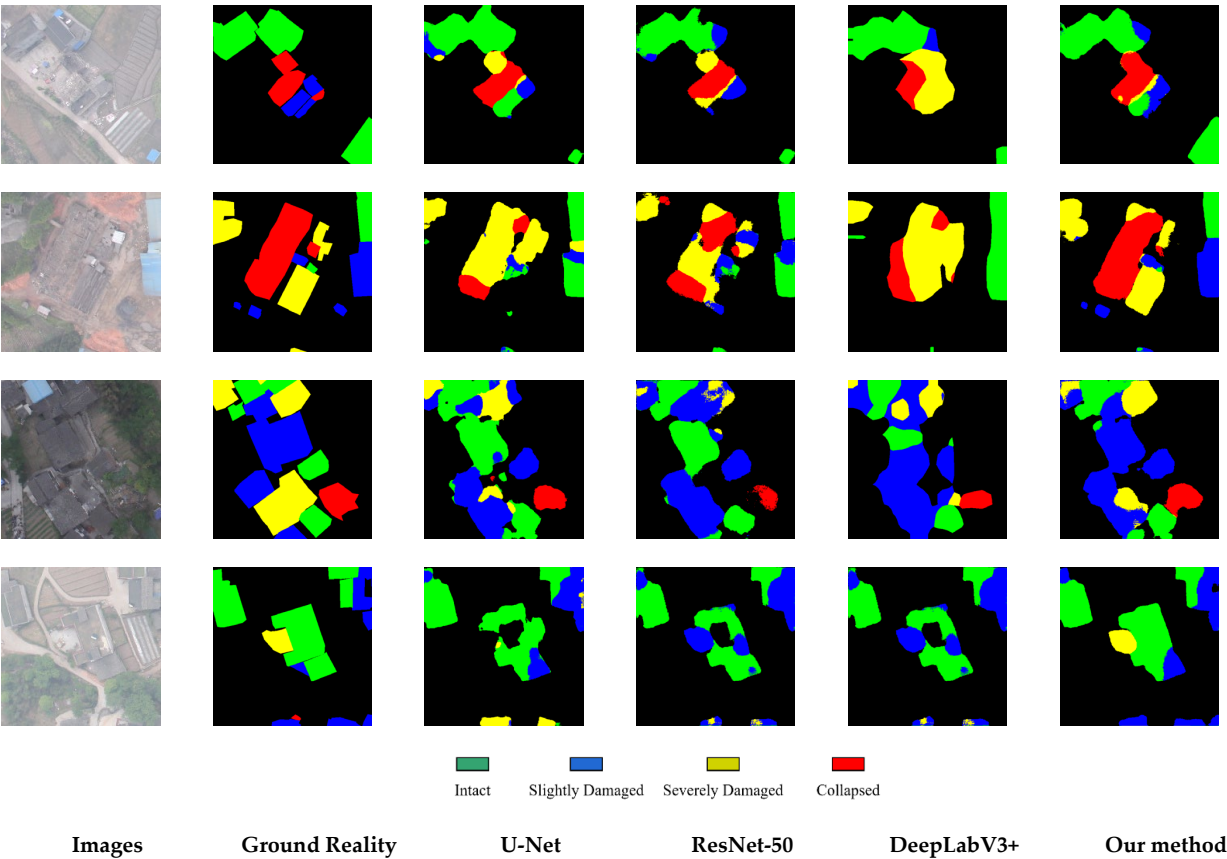


Figure 7. Comparison of different segmentation models for the Baoxing earthquake.

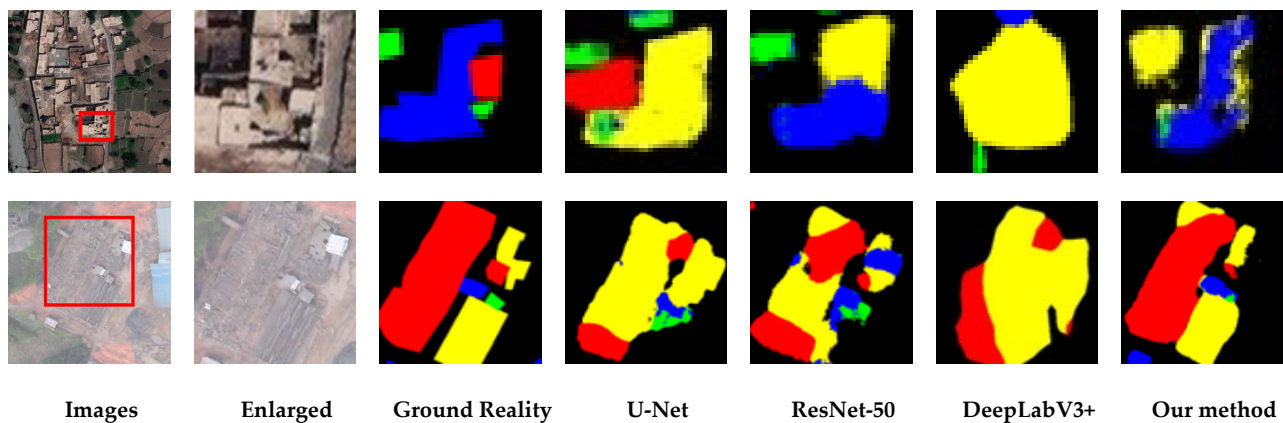


Figure 8. Detailed comparison of impact local amplification.

Similarly, the second row of Figure 8 contains a UAV image. The local image within the red box is enlarged and displayed in the second column. The left side of the building within the red box corresponds to the collapsed category, while the right side corresponds to the severely damaged category. It is evident that both the U-Net and ResNet-50 models predominantly classify the houses on the left and right sides as severely damaged, failing to differentiate between collapsed and severely damaged structures. The Deeplabv3+ model fails to recognize the complete and regular shape of the building. However, our BDHE-Net model accurately identifies the left side as collapsed and the right side as severely damaged while recognizing the complete and regular shape of the building. This demonstrates the effectiveness of our model in predicting building damage level classification.

Table 4 presents a quantitative comparison of building damage classification accuracy among various baseline models in our post-earthquake dataset. As indicated by the experimental results in Table 4, the proposed BDHE-Net surpasses the other models, achieving an average F1 score of 66.35% and an IOU of 47.15%. Compared with U-Net, ResNet-50, and DeepLabv3+, the average F1 score of BDHE-Net improved by 6.57%, 6.19%, and 8.22%, respectively, and the average IOU increased by 6.62%, 6.24%, and 8.09%, respectively. For the intact category, there was a slight difference in F1 between the four models, indicating that all models performed well in distinguishing intact buildings. However, for the severely damaged categories, BDHE-Net improved its F1 scores by 3.62% to 6.81% compared with the other models. This result proves the effectiveness of our proposed DAM module and combined loss function; by combining the method, BDHE-Net effectively improves the original model weights biased towards categories with the high number of categories, which results in good accuracy of the model in the lesser number of slightly damaged and severely damaged categories. In addition, for the collapsed category, our model improves the F1 scores by 13.29% to 21.98% compared with the other models, which proves that our proposed BDAM module can fully utilize building collapse characteristics, thus enabling the model to distinguish collapsed houses effectively. Finally, in terms of the average F1 scores of the four classes, our model improves the average F1 scores by 6.19% to 8.22% compared with the other models, which demonstrates that our proposed MFAF module plays a role in the dataset at different scales, which enables BDHE-Net to extract more complete information about houses of different sizes at different scales.

Table 4. Comparison of results among semantic segmentation networks.

Methods	Intact/F1 (%)	Slightly Damaged/F1 (%)	Severely Damaged/F1 (%)	Collapsed/F1 (%)	Mean/F1 (%)	Mean/P (%)	Mean/R (%)	Mean/IOU (%)
U-Net	81.93	50.05	51.51	56.48	57.78	56.82	58.67	40.53
ResNet-50	84.91	52.77	49.82	57.66	58.16	59.26	57.05	40.91
DeepLabv3+	81.13	51.02	53.01	48.97	56.13	55.27	57.17	39.06
Our method	83.62	53.31	56.63	70.95	64.35	64.36	63.73	47.15

3.5. Ablation Experiments

To validate the contribution of each module to the proposed method, detailed ablation experiments were conducted using the same data and experimental setup. Table 5 displays the results of the ablation experiments performed on the post-earthquake dataset, where B represents the BDAM module, M represents the MFAF module, D represents the DBA module, and C represents the combined loss function module. By adding the BDAM module, the MFAF module, the DBA, and combined loss function modules to the model, the present method achieves the highest overall accuracy relative to the baseline method. Specifically, when the MFAF module was added to the model, the average F1 score and IOU increased by 1.03% and 0.91%, respectively, compared with the baseline. In contrast, when the BDAM module was added to the model in this paper, the average F1 score and IOU increased by 1.71% and 1.80%, respectively. These results demonstrate that BDHE-Net offers tremendous benefits in the task of building damage assessment. The introduction of the MFAF module effectively integrates the integrity information of houses of different scales and enhances the ability to identify the integrity of houses of different scales. The BDAM module, on the other hand, fully considers different collapse directions and texture characteristics of damaged buildings, enhancing model performance for extracting damaged houses. In addition, when the DBA module and the combined loss function module were added to the model in this experiment, the average F1 and IOU scores increased by 0.67% and 0.51%, respectively. This suggests that the DBA module plays an active role in model training. To rectify the disparity in sample distribution among different classes of building damage in the dataset, the DBA module incorporates a strategy that involves augmenting the dataset with a smaller number of samples (e.g., in the severe damage and collapse categories) while simultaneously expanding the damaged house regions. By introducing more samples of fewer categories and expanding the damaged areas, the model is able to learn and classify the small number of samples better, which improves the refinement of building damage assessment. The combined loss function combines Dice and focal losses, which can balance the weights of different category samples during the training process. This approach prevents the model from excessively concentrating on samples with a large number of categories, such as the intact category, while neglecting those with fewer instances. By adjusting the weights of the samples, the combined loss function ensures the model allocates more attention to samples that are challenging to classify, thereby enhancing the performance of building damage assessment.

The results suggest that BDHE-Net shows significant advantages in the task of building damage assessment. This is because the BDAM module fully considers the characteristics of damaged buildings, e.g., different collapse directions and texture characteristics. Secondly, the DBA and combined loss function modules improve the model weight bias towards cases with a large number of categories. Finally, the MFAF module effectively fuses the house dimension information at different scales and enhances the integrity of houses at different scales. In summary, BDHE-Net demonstrates superior capabilities in conducting building damage assessments across multiple-resolution datasets.

Table 5. The impact of various modules within the BDHE-Net architecture on the accuracy of building damage classification.

Methods	Intact/F1 (%)	Slightly Damaged/F1 (%)	Severely Damaged/F1 (%)	Collapsed/F1 (%)	Mean/F1 (%)	Mean/P (%)	Mean/R (%)	Mean/IOU (%)
Baseline	82.18	52.04	51.43	67.41	60.94	60.68	61.46	43.93
Baseline + M	83.34	52.24	54.51	66.35	61.97	65.23	58.92	44.84
Baseline + M + B	83.83	56.51	53.45	69.63	63.68	65.73	61.42	46.64
Baseline + M + B + D + C	83.62	53.31	56.63	70.95	64.35	64.36	63.73	47.15

4. Conclusions

This study initially examined the heterogeneity in damaged buildings following an earthquake. Additionally, satellite remote sensing and UAV datasets from the Afghanistan and Baoxing earthquakes were compiled, classifying building damage into four categories as follows: intact, slightly damaged, severely damaged, and collapsed. To address the challenge of pixel-level assessment of building damage post-earthquake, a network named BDHE-Net was proposed to significantly enhance the model's accuracy in classifying severely damaged and collapsed buildings. The method was tested on our dataset and benchmarked against three state-of-the-art methods. Furthermore, the role of BDAM, MFAF, BDA, and combined loss function modules was explored. The experimental results show that the introduction of these four strategies improves the mean F1 and mean IOU value distribution by 3.41% and 3.22%, respectively, compared with the baseline model.

This paper presents the following key contributions:

1. A novel deep learning-based model is proposed to solve the pixel-level classification problem for post-earthquake building damage assessment, which is crucial for earthquake rescue and post-disaster damage assessment.
2. BDAM, MFAF, BDA, and combined loss function modules are incorporated into BDHE-Net, which enhance the model's capacity to discern varying levels of damage among buildings.

In the future, we will attempt to use multi-modal images, which refer to the fusion of data acquired by different sensors, such as optical imagery, radar data, hyperspectral data, and so on. By fusing data from different modalities, more comprehensive and multi-angle information can be obtained, thus enhancing the accuracy of building damage assessment. For example, optical images can provide information on the shape and texture of a building, while radar data can penetrate clouds and smoke to obtain structural information about a building, and hyperspectral data can provide rich spectral features for distinguishing buildings made of different materials. Pixel-level assessment of building damage under complex conditions can be further investigated to address the requirements of emergency rescue and post-disaster reconstruction efforts.

Author Contributions: Conceptualization, J.L. and Y.L.; methodology, Y.L.; software, Y.L.; validation, J.W., J.L., S.C., Y.W. and Y.L.; formal analysis, J.W., Y.W. and Y.L.; resources, J.W. and S.C.; data curation, J.L. and J.W.; writing—original draft preparation, J.L., Y.L. and S.C.; writing—review and editing, J.L., Y.L. and S.C.; supervision, J.W., J.L., Y.W. and S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key R&D Program of China, grant number 2022YFC3004405; the National Natural Science Foundation of China, grant number 42061073; and the Natural Science and Technology Foundation of Guizhou Province under Grant [2020]1Z056.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors are very grateful to Liu Jun and others at the National Earthquake Response Support Service for providing us with experimental data and data acquisition sites.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Saito, K.; Spence, R.J.; Going, C.; Markus, M. Using high-resolution satellite images for post-earthquake building damage assessment: A study following the 26 January 2001 Gujarat earthquake. *Earthq. Spectra* **2004**, *20*, 145–169. [\[CrossRef\]](#)
2. Mas, E.; Bricker, J.; Kure, S.; Adriano, B.; Yi, C.; Suppasri, A.; Koshimura, S. Field survey report and satellite image interpretation of the 2013 Super Typhoon Haiyan in the Philippines. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 805–816. [\[CrossRef\]](#)
3. Gong, L.; Li, Q.; Zhang, J. Earthquake building damage detection with object-oriented change detection. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Melbourne, Australia, 21–26 July 2013; pp. 3674–3677.
4. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv* **2015**, arXiv:1506.00019.
5. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
6. Sanchez-Lengeling, B.; Reif, E.; Pearce, A.; Wiltchko, A.B. A gentle introduction to graph neural networks. *Distill* **2021**, *6*, e33. [\[CrossRef\]](#)
7. Duarte, D.; Nex, F.; Kerle, N.; Vosselman, G. Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. *Remote Sens.* **2018**, *10*, 1636. [\[CrossRef\]](#)
8. Chowdhury, T.; Murphy, R.; Rahnemoonfar, M. RescueNet: A high resolution UAV semantic segmentation benchmark dataset for natural disaster damage assessment. *arXiv* **2022**, arXiv:2202.12361.
9. Xie, Y.; Feng, D.; Chen, H.; Liu, Z.; Mao, W.; Zhu, J.; Hu, Y.; Baik, S.W. Damaged building detection from post-earthquake remote sensing imagery considering heterogeneity characteristics. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4708417. [\[CrossRef\]](#)
10. Gupta, R.; Shah, M. RescueNet: Joint building segmentation and damage assessment from satellite imagery. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 4405–4411.
11. Shen, Y.; Zhu, S.; Yang, T.; Chen, C.; Pan, D.; Chen, J.; Xiao, L.; Du, Q. U-Net: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5402114.
12. Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A.; Zhang, L. 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.* **2021**, *265*, 112636. [\[CrossRef\]](#)
13. Shafique, A.; Seydi, S.T.; Cao, G. BCD-Net: Building change detection based on fully scale connected U-Net and subpixel convolution. *Int. J. Remote Sens.* **2023**, *44*, 7416–7438. [\[CrossRef\]](#)
14. Bai, Y.; Mas, E.; Koshimura, S. Towards operational satellite-based damage-mapping using u-net convolutional network: A case study of 2011 tohoku earthquake-tsunami. *Remote Sens.* **2018**, *10*, 1626. [\[CrossRef\]](#)
15. Rudner, T.G.; Rußwurm, M.; Fil, J.; Pelich, R.; Bischke, B.; Kopacková, V.; Bilinski, P. Rapid computer vision-aided disaster response via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In Proceedings of the First Workshop on AI for Social Good. Neural Information Processing Systems (NIPS-2018), Montreal, QC, Canada, 6 December 2018; pp. 3–8.
16. Hong, Z.; Yang, Y.; Liu, J.; Jiang, S.; Pan, H.; Zhou, R.; Zhang, Y.; Han, Y.; Wang, J.; Yang, S.; et al. Enhancing 3D reconstruction model by deep learning and its application in building damage assessment after earthquake. *Appl. Sci.* **2022**, *12*, 9790. [\[CrossRef\]](#)
17. Hong, Z.; Zhong, H.; Pan, H.; Liu, J.; Zhou, R.; Zhang, Y.; Han, Y.; Wang, J.; Yang, S.; Zhong, C. Classification of building damage using a novel convolutional neural network based on post-disaster aerial images. *Sensors* **2022**, *22*, 5920. [\[CrossRef\]](#)
18. Günen, M.A. Fast building detection using new feature sets derived from a very high-resolution image, digital elevation and surface model. *Int. J. Remote Sens.* **2024**, *45*, 1477–1497. [\[CrossRef\]](#)
19. Yu, Z.; Chen, Z.; Sun, Z.; Guo, H.; Leng, B.; He, Z.; Yang, J.; Xing, S. SegDetector: A Deep Learning Model for Detecting Small and Overlapping Damaged Buildings in Satellite Images. *Remote Sens.* **2022**, *14*, 6136. [\[CrossRef\]](#)
20. Chenna, R.; Patnala, N.; Vemuri, J.P.; Ramancharla, P.K. Insights on the June 21, 2022, Khost earthquake, Afghanistan. *Sadhana* **2023**, *48*, 144. [\[CrossRef\]](#)
21. He, D.; Shi, Q.; Liu, X.; Zhong, Y.; Zhang, X. Deep subpixel mapping based on semantic information modulated network for urban land use mapping. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10628–10646. [\[CrossRef\]](#)
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Poulencard, A.; Ovsjanikov, M. Multi-directional geodesic neural networks via equivariant convolution. *ACM Trans. Graph.* **2018**, *37*, 236. [\[CrossRef\]](#)
24. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.

25. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
26. Liu, S.; Huang, D.; Wang, Y. Learning spatial fusion for single-shot object detection. *arXiv* **2019**, arXiv:1911.09516.
27. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
28. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
29. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An imperative style, high-performance deep learning library. *arXiv* **2019**, arxiv:1912.01703.
30. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.