remote sensing

MDPI

Article

Pyramid Pooling Module-Based Semi-Siamese Network: A Benchmark Model for Assessing Building Damage from xBD Satellite Imagery Datasets

Yanbing Bai [1,*], Junjie Hu [2], Jinhua Su [1], Xing Liu [3], Haoyu Liu [1], Xianwen He [1], Shengwang Meng [1], Erick Mas [4] and Shunichi Koshimura [4]

[1] Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China; chasesu@ruc.edu.cn (J.S.); 2017201665@ruc.edu.cn (H.L.); 2017201647@ruc.edu.cn (X.H.); mengshw@ruc.edu.cn (S.M.)

[2] Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518172, China; hujunjie@cuhk.edu.cn

[3] Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan; ryu@vision.is.tohoku.ac.jp

[4] International Research Institute of Disaster Science, Tohoku University, Sendai 980-8572, Japan; mas@irides.tohoku.ac.jp (E.M.); koshimura@irides.tohoku.ac.jp (S.K.)

* Correspondence: ybbai@ruc.edu.cn; Tel.: +86-10-6251-1318

Received: 17 October 2020; Accepted: 7 December 2020; Published: 11 December 2020

Abstract: Most mainstream research on assessing building damage using satellite imagery is based on scattered datasets and lacks unified standards and methods to quantify and compare the performance of different models. To mitigate these problems, the present study develops a novel end-to-end benchmark model, termed the pyramid pooling module semi-Siamese network (PPM-SSNet), based on a large-scale xBD satellite imagery dataset. The high precision of the proposed model is achieved by adding residual blocks with dilated convolution and squeeze-and-excitation blocks into the network. Simultaneously, the highly automated process of satellite imagery input and damage classification result output is reached by employing concurrent learned attention mechanisms through a semi-Siamese network for end-to-end input and output purposes. Our proposed method achieves F1 scores of 0.90, 0.41, 0.65, and 0.70 for the undamaged, minor-damaged, major-damaged, and destroyed building classes, respectively. From the perspective of end-to-end methods, the ablation experiments and comparative analysis confirm the effectiveness and originality of the proposed method. Finally, the consistent prediction results of our model for data from the 2011 Great East Japan Earthquake verify the high performance of our model in terms of the domain shift problem, which implies that it is effective for evaluating future disasters.

Keywords: pyramid pooling module; semi-Siamese; benchmark model; damage assessment; end-to-end; xBD dataset

1. Introduction

Natural disasters, which have been occurring frequently in recent years [1], pose a huge threat to the safety of residential buildings as well as life and property. Therefore, it is of great significance to obtain accurate information on damaged buildings to carry out interventions after natural disasters [2,3]. Satellite remote sensing technology is used to obtain disaster information because it can acquire rapid and large-scale surface information [4–9]. In particular, the recent development of deep convolutional neural network algorithms has improved disaster assessment accuracy based on satellite imagery [7,10–14]. Nevertheless, the practicability of disaster assessment methods also needs to be considered and the development of a high-precision and practical disaster assessment method is of great significance for emergency rescues during disasters.

The key factor to obtaining disaster information is assessing building damage. Mainstream building damage assessment methods include two main steps: building localization and damage classification. First, building localization is unnecessary if building footprint information is provided; however, this information is rarely available in disaster events, especially those in underdeveloped areas. Second, damage classification relies heavily on building footprint information; therefore, the accuracy of building localization information directly affects this classification. Xu et al. [12] proposed a two-stage architecture for assessing damage. A faster region-based convolutional neural network (R-CNN) model was used to localize building information, followed by a change detection network to identify building damage from both pre- and post-satellite imagery. Gupta et al. developed a two-stage baseline model [15] based on the xBD dataset [16] in which a U-Net model was implemented to first detect building areas and then classify damage based on detecting change. However, the separation of building localization and damage classification necessitates the greedy stage-wise training,which finds the local optimization for each stage using parameters with the temporal results of previous stages fixed. Consequently, such two-stage methods often suffer from low operability during actual disaster responses. The main reason behind this low operability is that although each stage needs an input and a ground truth label, there is no corresponding ground truth for the building localization results. The alternative is to use the building localization ground truth as the damage classification input, which lowers performance when predicting the damage level using a bad building localization result. The shortcomings of the two-stage method have been discussed and the end-to-end method is more popular than other methods because of its better performance and convenience in one-step training.

To solve the above challenges, some researchers apply five-class semantic segmentation, which simply regards “no building” as a damage class [17,18]. This approach solves the problem that the classification of damage level depends highly on the precision of building localization under the two-stage architecture. Adopting the end-to-end strategy usually improves the classification of damage level greatly; however, building localization performance may worsen slightly. Weber et al. [19] used the Mask R-CNN with the FPN architecture and the same model architecture for both building localization and per-pixel damage classification. Further, instead of working with full images, they trained the architecture on both the pre- and the post-image quadrants and fused the final segmentation layer to draw building boundaries more accurately. Hao [20] designed a Siam-U-Net-Attn model end-to-end for both damage classification and building segmentation, which indicated that embedding building segmentation helped classify damage. In detail, the U-Net model was used for both the pre-disaster and the post-disaster images to produce binary masks. The two features produced by the U-Net encoder were merged using different fusion methods in the Siamese network to compare the features of the two input frames to detect building damage. Meanwhile, the features extracted from the encoder regions also assisted in damage classification.The baseline achieved an appreciable intersection over union (IoU) score for localization and performed well when classifying buildings into not damaged and destroyed. Hence, end-to-end methods need to balance building detection with damage classification. Gupta et al. [17] proposed a novel loss function that consisted of a binary cross-entropy loss for building segmentation and a foreground only selective categorical cross-entropy loss for damage classification.

However, these baseline models cannot accurately distinguish between minor- and major-damaged buildings. Indeed, five-class semantic segmentation is a harder task than building localization. Using the transfer learning technique, the performance of some end-to-end models can be improved by initializing the final model with pre-trained building localization weights. Nia and Mori [21] proposed a novel damage assessment deep model for buildings using only post-disaster images. The model transferred three neural networks: DilatedNet, LeNet, and VGG. VGG and LeNet extracted deep features from the input source, while DilatedNet preprocessed the input data. Combinations of these networks were then distributed among three separate feature streams. The extracted features were summarized into a continuous value denoting the damage level. The transfer learning mechanism can thus benefit all end-to-end models; however, we do not apply the mechanism to compare the performance of model structures in this study.

In addition to those works discussed above, Valentijn et al. [22] addressed the problem of automated building damage assessment based on the xBD dataset. The authors proposed a CNN consisting of two inception-v3 blocks for extracting features from pre-/post-disaster images and a stack of fully connected layers for the classifier. To overcome the overfitting problem, they employed a batch normalization layer and a dropout layer for each fully connected layer and analyzed the generalizability and transferability of the CNN. Harirchian et al. [23] addressed the problem of risk assessment using SVM and data on the Düzce Earthquake in Turkey. They employed 22 building features such as system type, year of construction, and ground floor area as inputs to the SVM for the estimation. Compared with CNN-based methods, this method is a "white box". However, it relies more on carefully chosen parameter(s) for the SVM and may perform worse than CNN-based methods. Zhuo et al. [24] focused on evaluating the risk of the subsidence of reclaimed land at the Xiamen Xi'an New Airport in China. They showed that SAR data are a powerful information source for analyzing reclaimed land subsidence as well as estimating the risk of future subsidence, which is valuable for land use planning. Morfidis et al. [25] used an artificial neural network (ANN) to estimate seismic damage to structures. This study provided a good explanation for civil engineers unfamiliar with ANNs. Harirchian et al. [26] addressed the problem of predicting damage to reinforced concrete buildings when an earthquake occurs. They employed six human-defined features to represent a building. A shallow neural network was then used as the estimator, which was trained and tested based on the representation vectors consisting of the six features for each sample. The dataset employed for this work was obtained from the Düzce Earthquake in Turkey. Morfidis et al. [27] addressed the problem of estimating damage to reinforced concrete buildings using ANNs. The authors employed human-defined features (i.e., seismic and structural parameters) to train a shallow neural network consisting of linear production layers and activation layers and then analyzed the network's hyper-parameters and human-defined features, providing a good guide for applying ANNs experimentally.

In this study, we design a concurrent learned attention network, which is an end-to-end trainable, unified model, to localize buildings and classify damage jointly. This network is built on a semi-Siamese strategy that can learn collectively. We use a pixel-level segmentation-based approach as well as residual blocks (RBs) with dilated convolution and squeeze-and-excitation (SE) blocks to detect damage to the segmented buildings. To model the global contextual prior, we also introduce the pyramid-pooling module (PPM) that enhances the scale invariance of images, while lowering over-fitting risk.

To benchmark our method, we develop our model based on the large-scale xBD dataset, which contains satellite images from multiple disaster types worldwide such as earthquakes, hurricanes, floods, and wildfires. To verify our method's effectiveness and practicality, we compare its performance with that of the published baseline model based on the xBD dataset. To demonstrate its usefulness, we use data from the 2011 Great East Japan Earthquake.

We contribute to the body of knowledge in four main ways. First, redwe propose a benchmark model for assessing building damage based on a large-scale xBD satellite imagery dataset. Second, we put forward an end-to-end model for assessing building damage, termed PPM-SSNet, which adopts the semi-Siamese technique, the PPM, and an attention mechanism. To overcome the difficulty of multi-target learning, we use the weighted combined losses of dice, focal, and cross-entropy. Third, we use efficient five data augmentation methods and four class balance strategies designed for these tasks to improve the task performance of all the mainstream models. Finally, we use different disaster images, including severely damaged images and rare disaster images, to test our model's robustness by comparing it with two strong baseline models.

2. Data

The xBD dataset [16] used in this study comes from xView 2 challenge (https://xview2.org/dataset). It contains over 850,000 building polygons from six types of disasters (earthquake, tsunami, flood, volcanic eruption, wildfire, and wind) worldwide, covering 45,000 km^2. The building polygons and damage scales are included. Following the joint damage scale (JDS) based on EMS-98, the building damage scales are visually interpreted from satellite imagery and categorized into undamaged, minor-damaged, major-damaged, and destroyed buildings. The training dataset contains 9168 pairs of pre-event/post-event three-band images with a spatial resolution of 1024 × 1024 pixels. Moreover, segmented ground truth masks with building polygons and building damage class labels are provided in the JSON file format. Figure 1 shows the details of the xBD dataset. Approximately 96.7% of the pixels are in the non-building area, as shown in Table 1, which indicates the sample imbalance among our original data.

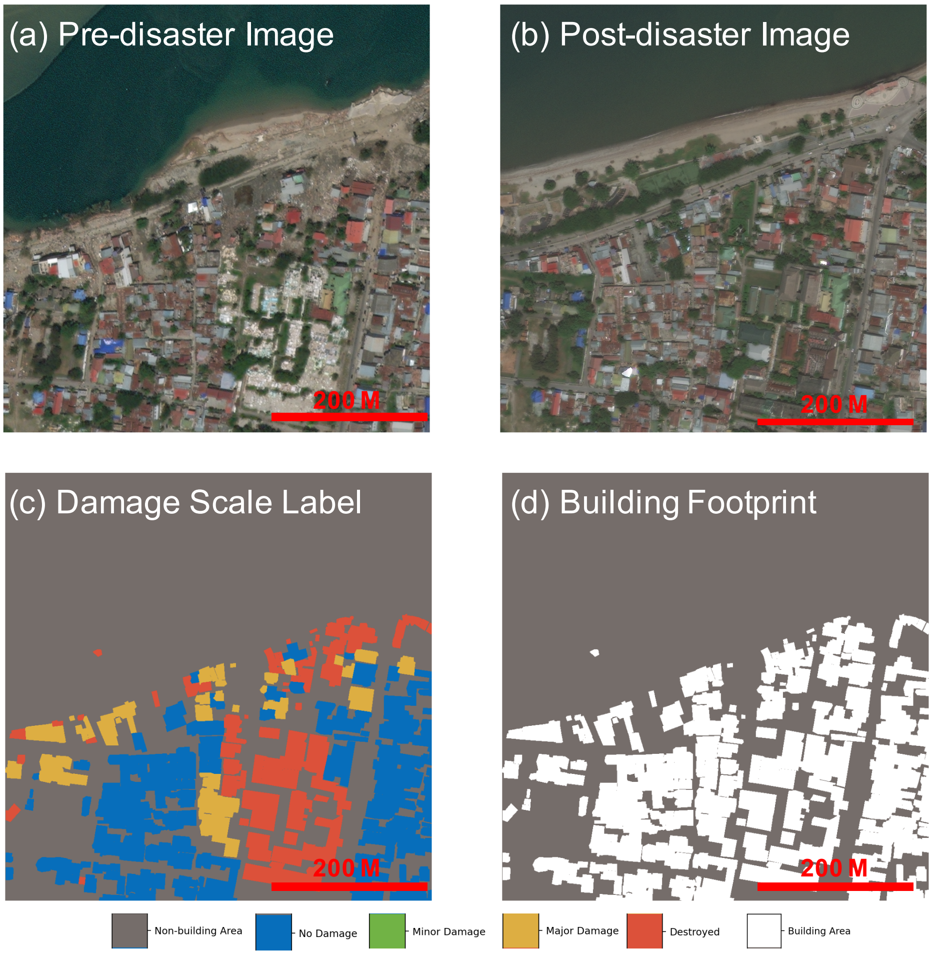


Figure 1. Example of the xBD dataset: Tsunami in Palu, Indonesia. From left to right: (**a**) Pre-disaster image, (**b**) Post-disaster image, (**c**) Damage scale, and (**d**) Building footprint.

Table 1. Non-building area to building area ratio at the pixel level.

Non-Building Area	Building Area
96.97%	3.03%

Consistent with real-world disaster case scenarios, the xBD dataset presents severe class imbalance. In terms of the building area/non-building area ratio at the pixel level, the non-building pixel occupies 97% of the image pixels, as shown in Table 1. Regarding the proportional distribution of the damage class at the pixel level, the number of undamaged building pixels far exceeds that of the other three classes, with a ratio of up to 76%. Only 6% of pixels belong to the class of destroyed. The minor-damaged and major-damaged categories account for almost the same proportion. Figure 2 compares the class balance.

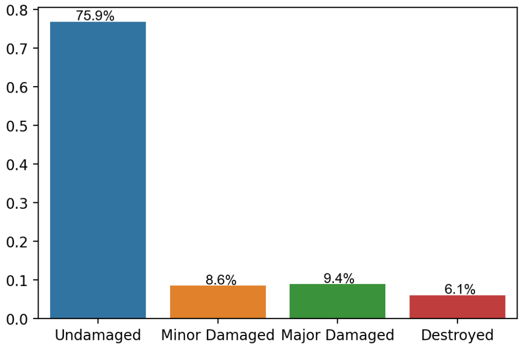


Figure 2. Ratio of damage class at the pixel level.

To verify our method's transferability, we test other satellite imagery with the developed model based on the xBD dataset. Two areas in Higashi Matsushima severely affected by the 2011 Great East Japan Earthquake are used for testing, as shown in Figure 3a–c. These two areas are selected because the xBD dataset does not contain any disaster data from Japan and data on the tsunami in the xBD dataset are scarce. This design can test the ability of our model for to evaluate and predict unknown disasters.

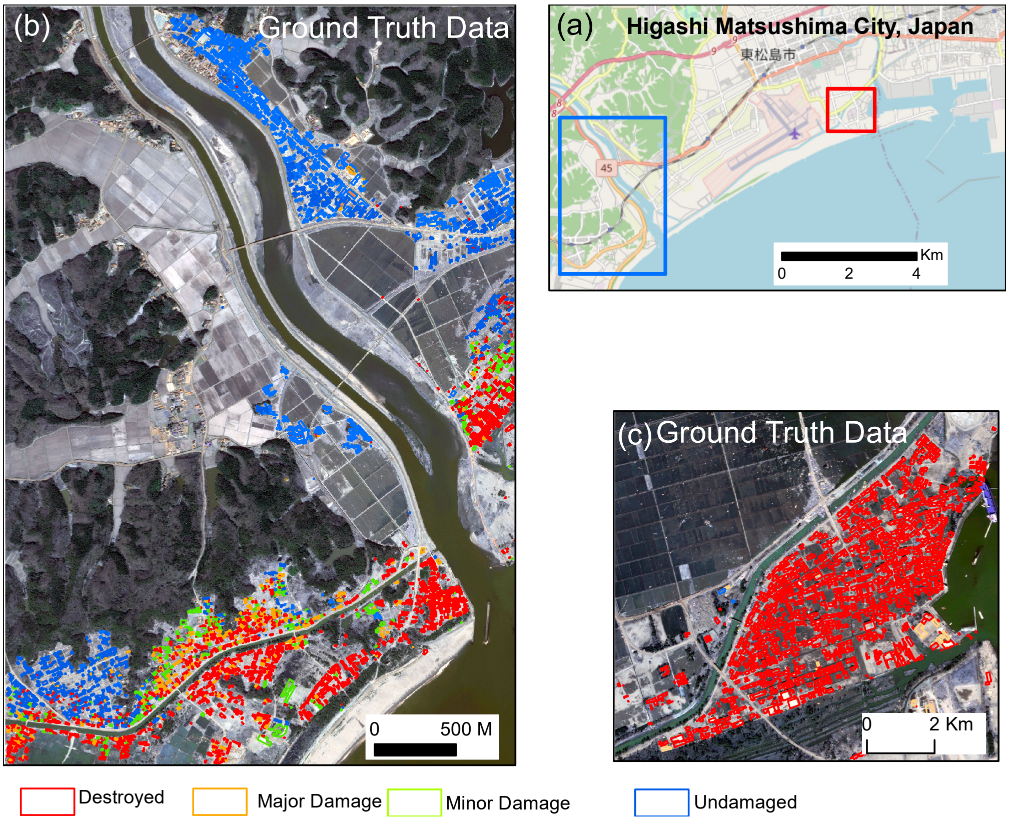


Figure 3. Validation area. (**a**) Higashi Matsushima in the Tohoku region of Japan; the rectangular areas marked in blue and red are the selected validation areas; (**b**) The close-up of the blue area as shown in Figure 10a with the ground truth data of building damage; and (**c**) The close-up of the red area as shown in Figure 10a with the ground truth data of building damage.

The building damage ground truth data for the testing area are based on the field investigation conducted by TTJS [28]. To retain consistency with the xBD data label as much as possible to facilitate the comparative analysis, we recategorize the TTJS building damage data into four classes: "undamaged", "minor damage" (including "moderate damage and" "minor damage in the" TTJS standard), "major damage," and "destroyed" (including "washed away," "collapsed," and "completely damaged" in the TTJS standard) as shown in Figure 3b,c. We implement this classification standard because standards based on field surveys are much stricter than the visual interpretation based on satellite images.

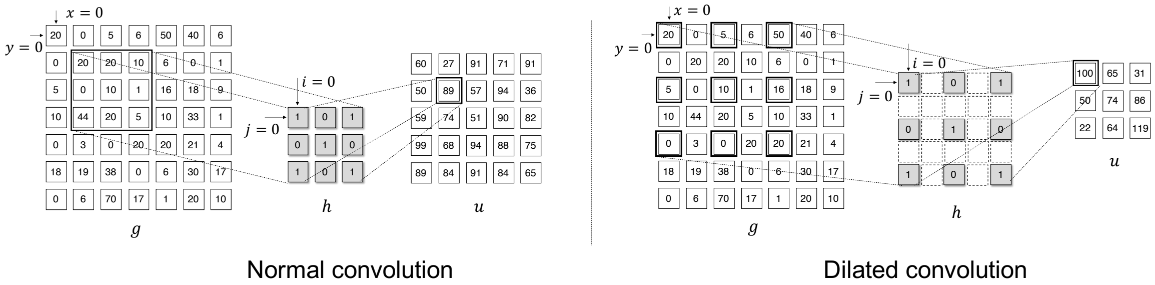
The four-band multispectral high-resolution Worldview-2 images with a spatial resolution of 0.6 m, collected before and after the 2011 Great East Japan Earthquake, were used for validation as shown in background of Figure 3b,c.

3. Methodology

The PPM-SSNet model developed in this research employed dilated convolution, the SE mechanism for attention, and the PPM, as detailed below.

3.1. Dilated Convolution for Large Receptive Fields

Collectively leveraging the global and local features of an input image is effective at solving computer vision problems [29–32]. Because of the nature of images, the different characters of an image are represented on different scales. A large field in an image includes global appearances such as objects' contours, whereas a small field includes local appearances such as local textures. This also applies to building localization and damage assessment. One way to realize this idea is with image down-sampling, which reduces the size of an image. This is equivalent to enlarging the receptive field of a convolutional unit in a specific location of an image. Although down-sampling an image leads to less information compared with a reduction in the resolution, it is still used when computing resources (e.g., GPU memory) are limited. Another way to enlarge the convolutional receptive field is by employing dilated convolution [29]. A dilated convolutional unit performs in the same way as normal convolution on an image. The difference is that it has dilated convolutional kernels. A high-dilated rate enables us to have a large convolutional receptive field for the unit. Further, no information is lost with an increasing receptive field under dilated convolution. Figure 4 shows an example of dilated convolution with a dilated rate of 2.



Normal convolution Dilated convolution

Figure 4. Dilated convolution with dilated rates of 1 (i.e., normal convolution; left side of the figure) and 2 (right side of the figure). g, h, and u mean the input image (or activation map), convolutional kernel, and output. An output u is calculated by summing the multiplications of each value (i, j) at the kernel h and its corresponding value (x, y) at g.

3.2. SE Mechanism for Attention

The SE mechanism was originally developed to improve the performance of image classification on ImageNet [33]. It is a weighting system that produces and applies channel-wise weights on a feature map (i.e., the output from an intermediate layer in a CNN). To determine the weight on each channel, it computes the average activation values of the channels; then, these are converted by two linear production layers with ReLU and Sigmoid activation functions to generate the channel-wise weights. The aggregation of the activation values is equivalent to global average pooling, as shown in Figure 5. A CNN, which is equipped with a number of attentional mechanisms, can perform feature recalibration; it learns to selectively emphasize informative features and suppress less useful features, which helps reduce ambiguity when estimating the correct damage level and thus improves the accuracy of building assessment.

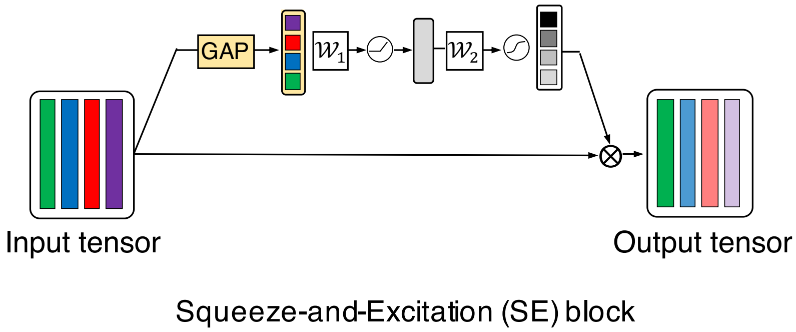


Figure 5. Squeeze-and-excitation (SE) blocks produce and apply channel-wise attention on the activation maps. GAP means global average pooling. w_i denotes the ith linear production layer. ReLU and Sigmoid are employed following w_1 and w_2 for the activation functions. The columns depicted in different colors represent the activation map of each channel of the input/output tensor.

3.3. PPM

The PPM pools the activation map of each channel in a pyramidal fashion [34]. It makes N × N (N = 1, 2, 4, ...) grids on the activation map of each channel. Each cell of a grid overlaps with a square region of the activation map. Each grid for the channel perfectly covers the whole activation map. On the region covered by each cell of a grid, a user-defined pooling process such as global max pooling or global average pooling is employed to pool the region into a single value. This process quantifies each activation map into a vector with a length equal to N × N. The vectors produced with different N (e.g., 1, 2, and 4) are then concatenated into a representation vector for the channel. The above process is applied to all the channels to produce their representation vectors. The final output of this module is generated by concatenating these representation vectors, as shown in Figure 6. The PPM is a simple yet effective feature aggregation mechanism. It aggregates features from multiple scales. Global features such as the shapes of buildings are covered with a small N (e.g., N = 2), whereas local features such as the details of damaged buildings are covered with a large N (e.g., N = 4). Then, the final output of this mechanism becomes a representative vector of the input sample, which improves the accuracy of building localization and damage assessment.

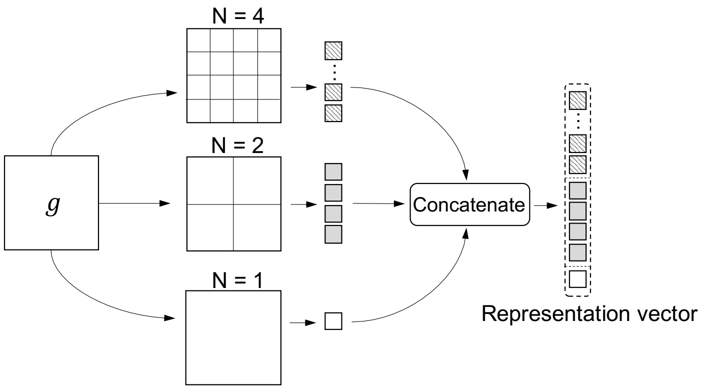


Figure 6. The pyramid pooling module (PPM) g represents an activation map of a single channel. N is the number of cells in a row/column of a pooling grid.

3.4. Pyramid Pooling Module-Based Semi-Siamese Network b(PPM-SSNet)

The task of estimating the damage assessment of buildings is divided into two stages. The first stage identifies the buildings on an image. This can be treated as a localization problem in which a system such as a CNN is employed to estimate the binary localization map for an input image. A location with 1 or 0 on the map indicates whether it is a building or not. The localization map is then employed as a prior for the second stage to estimate the damage assessment of a location with a value equal to 1. Based on this idea, we design a network to jointly estimate buildings' locations and assess their damage. We use the pre-image alone to estimate the location map and then use both the pre- and the post-images to estimate the assessment result. To leverage the localization map to produce an accurate assessment result, we directly multiply it by the output of the assessment estimator. This process corrects the assessment result, improving its quality from a coarse to a fine level (see Figure 7).

Figure 7 shows the architecture. The network is built on a semi-Siamese strategy. We let the weights at the shallow layers of the network share the two input images (i.e., pre-/post-images) to enable it to produce a good "filters' bank" by collectively learning the low-level features from both. As the layers go deeper, we stop sharing weights and use independent branches for the two inputs instead. The two branches are merged by subtracting one from the other along their channels, which encourages the network to learn the differences between the pre- and post-images. For the tail of the network, we use a single branch of the layers to produce the final estimation result. In the network, we employ RBs with dilated convolution and SE blocks. Our motivation for using RBs is that the network can extract features from large and small receptive fields by employing the large and small dilated rates used in RBs, which may improve its representation ability for the estimations. In addition, SE blocks are employed to encourage the network to focus on the important features, while suppressing the less useful ones. We employ a PPM at the end of the network, immediately before an SE block, and a convolutional layer to aggregate the features.

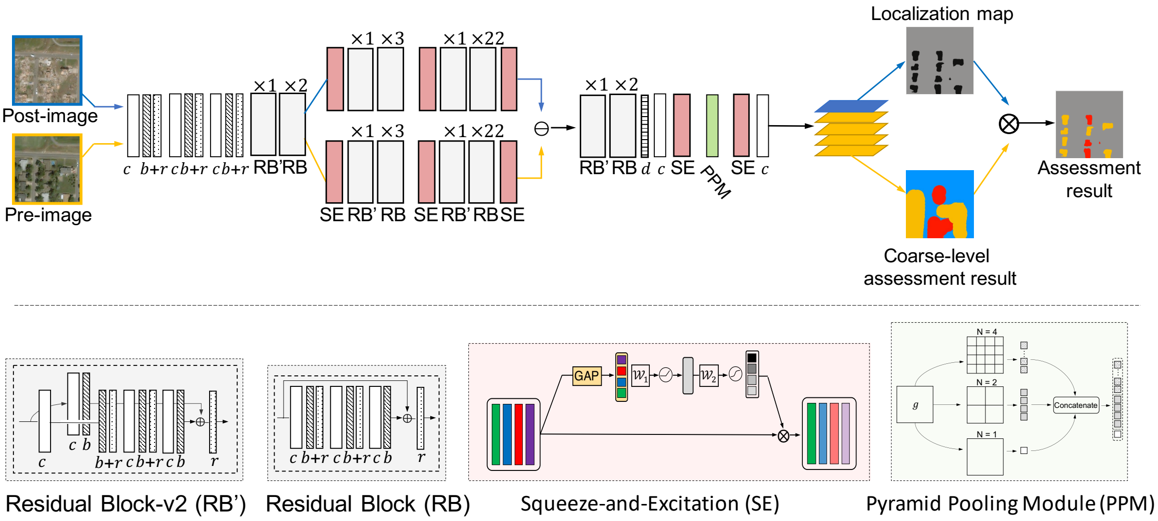


Figure 7. The architecture of the proposed network. *c*, *b*, *d*, and *r* represent the convolutional layer, batch normalization layer, dropout layer, and ReLU layer. SE, RB', RB, and PPM represent the modules illustrated at the bottom of this figure. The difference between RB' and RB is that RB' has an additional convolutional layer + batch normalization layer, which is designed to change the number of channels or size of the input tensor if needed. See Table 2 for more details.

Table 2. Details of the proposed network. Conv., RB′, RB, SE, Drop, and PPM mean the convolutional layer (c), RB-v2 (RB′), RB, SE, dropout layer (d), and PPM (see Figure 7). For the convolutional layer (Conv./conv.), in, out, $stride$, and $dila$ mean the input's dimension, output's dimension, stride, and dilation rate for the layer. $k \times k$ means the size of the convolutional kernel. For an SE module, in, mid, and out mean the input's dimension, dimension of the output of the middle layer, and output's dimension. For the PPM, out means the output's dimension.

	Layer	Parameters	Number
Share	Conv.	$[\ 7 \times 7,\ in = 3,\ out = 16,\ stride = 1,\ dila = 1\ ]$	×1
	Conv.	$[\ 3 \times 3,\ in = 16,\ out = 16,\ stride = 1,\ dila = 1\ ]$	×1
	Conv.	$[\ 3 \times 3,\ in = 16,\ out = 32,\ stride = 2,\ dila = 1\ ]$	×1
Share	RB′	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 32,\ out = 64,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 64,\ out = 64,\ stride = 2,\ dila = 1 \\ \text{conv.},\ 1 \times 1,\ in = 64,\ out = 256,\ stride = 1,\ dila = 1 \\ \text{down.},\ 1 \times 1,\ in = 32,\ out = 256,\ stride = 2,\ dila = 1 \end{bmatrix}$	×1
	RB	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 256,\ out = 64,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 64,\ out = 64,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 1 \times 1,\ in = 64,\ out = 256,\ stride = 1,\ dila = 1 \end{bmatrix}$	×2
Independent	SE	$[\ in = 256,\ mid = 16,\ out = 256\ ]$	×1
	RB′	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 256,\ out = 128,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 128,\ out = 128,\ stride = 2,\ dila = 1 \\ \text{conv.},\ 1 \times 1,\ in = 128,\ out = 512,\ stride = 1,\ dila = 1 \\ \text{down.},\ 1 \times 1,\ in = 256,\ out = 512,\ stride = 2,\ dila = 1 \end{bmatrix}$	×1
	RB	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 512,\ out = 128,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 128,\ out = 128,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 1 \times 1,\ in = 128,\ out = 512,\ stride = 1,\ dila = 1 \end{bmatrix}$	×3
	SE	$[\ in = 512,\ mid = 32,\ out = 512\ ]$	×1
	RB′	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 512,\ out = 256,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 256,\ out = 256,\ stride = 1,\ dila = 2 \\ \text{conv.},\ 1 \times 1,\ in = 256,\ out = 1024,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 1 \times 1,\ in = 512,\ out = 1024,\ stride = 1,\ dila = 1 \end{bmatrix}$	×1
	RB	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 1024,\ out = 256,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 256,\ out = 256,\ stride = 1,\ dila = 2 \\ \text{conv.},\ 1 \times 1,\ in = 256,\ out = 1024,\ stride = 1,\ dila = 1 \end{bmatrix}$	×22
	SE	$[\ in = 1024,\ mid = 64,\ out = 1024\ ]$	×1
Single	RB′	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 1024,\ out = 512,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 512,\ out = 512,\ stride = 1,\ dila = 4 \\ \text{conv.},\ 1 \times 1,\ in = 512,\ out = 2048,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 1 \times 1,\ in = 1024,\ out = 2048,\ stride = 1,\ dila = 1 \end{bmatrix}$	×1
	RB	$\begin{bmatrix} \text{conv.},\ 1 \times 1,\ in = 2048,\ out = 512,\ stride = 1,\ dila = 1 \\ \text{conv.},\ 3 \times 3,\ in = 512,\ out = 512,\ stride = 1,\ dila = 4 \\ \text{conv.},\ 1 \times 1,\ in = 512,\ out = 2048,\ stride = 1,\ dila = 1 \end{bmatrix}$	×2
	Drop	–	×1
Single	Conv.	$3 \times 3,\ in = 2048,\ out = 512,\ stride = 1,\ dila = 2$	×1
	SE	$[\ in = 512,\ mid = 16,\ out = 512\ ]$	×1
	PPM	$[\ out = 512]$	×1
	SE	$[\ in = 1024,\ mid = 64,\ out = 1024\ ]$	×1
	Conv.	$1 \times 1,\ in = 1024,\ out = 5,\ stride = 1,\ dila = 1$	×1

4. Experimental Analysis

Over-sampling and data augmentation are adopted in this study. The assessment metrics as well as loss and mask dilation parameter settings are detailed below.

4.1. Resampling

Building damage detection networks based on xBD generally perform badly when detecting minor and major damage, resulting in comparatively low recalls and F1 scores for these two categories because of imbalanced training data. To overcome this problem, we devise several methods to increase the number of minor damage and major damage instances, one of which is over-sampling the training dataset. Since our model is designed to generate pixel-level classification results, we suggest using a *main label* to decide how many times a picture containing multi-label pixels should be repeated in the training dataset. A weight vector $w = (w_0, w_1, w_2, w_3)^T$ is given based on experience, each element of which represents the relative importance of the corresponding category. For picture i, n_i is the vector recording the number of pixels of each category and its *main label* is defined as

$$Main\ Label_i = \underset{j \in \{0,1,2,3\}}{\arg\max}\ w_j n_{ij} \tag{1}$$

where category 0 denotes no damage, category 1 denotes minor damage, and so on. Table 3 shows the *main label* categories and corresponding repeated times.

Table 3. Main labels and corresponding repeated times.

Main Label	No Damage	Minor Damage	Major Damage	Destroyed
Repeated Times	0	3	2	1

Since the images are cropped and randomly augmented later, there is no concern that the repeated pictures are identical to the original ones.

After over-sampling, we perform a cropping-and-selecting process with discrimination. Similar to above, we reweight each pixel as inversely proportional to the frequency of its corresponding damage level. The original image size is 1024 × 1024. We uniformly sample several 512 × 512 crops from each image and choose the one with the largest sum of pixel weights. Without increasing the volume of the training data, such a process further alleviates the data imbalance of the xBD dataset.

4.2. Data Augmentation

To enhance the generalizability of our model, we apply the following data augmentation methods sequentially to each image. As shown in Table 4, every method is assigned a value, indicating the probability of occurrence. In other words, the sequence of augmentation methods applied to an image is determined randomly and the higher the order, the earlier is the execution.

Table 4. Data augmentation methods and probabilities.

Method	**Pre to Post**	**Flip**	**Rotate by 90 Degree**	**Shift Pnt**
Probability	0.015	0.5	0.95	0.1
Method	Rotation	Scale	Color shifts	Change hsv
Probability	0.1	0.7	0.01	0.01
Method	CLAHE	Blur	Noise	Saturation
Probability	0.0001	0.0001	0.0001	0.0001
Method	Brightness	Contrast		
Probability	0.0001	0.0001		

4.3. Assessment Metrics

End-to-end building damage assessment includes two progressive tasks: building localization and damage classification. The former can be regarded as a binary segmentation, while the latter is a multi-classification task. This study adopts F1 scores, precision, recall, and IoU to evaluate our network's performance. For the localization task, the F1 score is used:

$$loc\ F_1 = \frac{2TP_{loc}}{2TP_{loc} + FN_{loc} + FP_{loc}} \tag{2}$$

where TP_{loc} denotes the number of pixels correctly classified as buildings, FP_{loc} indicates the number of pixels misclassified as buildings, and FN_{loc} means the number of pixels misclassified as background. For the classification task, the F1 scores, precision, and recall for each damage category are calculated. A macro-IoU is also implemented to quantify accuracy when data are imbalanced:

$$precision_j = \frac{TP_j}{TP_j + FP_j} \tag{3}$$

$$recall_j = \frac{TP_j}{TP_j + FN_j} \tag{4}$$

$$cls\ F_{1j} = \frac{2 \times precision_j \times recall_j}{precision_j + recall_j} \tag{5}$$

$$IoU_j = \frac{TP_j}{TP_j + FP_j + FN_j} \tag{6}$$

$$IoU = \frac{1}{4}\sum_{j=1}^{4} IoU_j \tag{7}$$

where $j \in \{0, 1, 2, 3\}$, TP_j denotes the number of pixels (or instances) correctly classified as category j, FP_j indicates the number misclassified as category j, and FN_j means the number misclassified as other categories.

4.4. Loss and Mask Dilation

The output damage scale classification mask has five channels: the four damage levels and no-building label. We adopt a weighted mixed loss that consists of dice loss and focal loss for the damage scale classification loss L_d and weighted binary cross-entropy loss for the building segmentation loss L_d, which are defined as

$$L_s = -[w_{s,1} \times y_s \log P_s + w_{s,0} \times (1 - y_s) \log(1 - p_s)] \tag{8}$$

$$Seg_c = w_1 \times Dice_c(m_p, m_t) + w_2 \times Focal_c(m_p, m_t) \tag{9}$$

$$L_d = \sum_{c=1}^{5} w_c \times Seg_c \tag{10}$$

where y_p and y_s are the ground truth label and detected building segmentation probability, respectively, while m_p and m_t are the true mask and predicted mask for damage scale c, respectively. As most samples contain no buildings, we use a larger weight for the building class, as indicated by $w_{s,1}$ in segmentation loss L_s. Additionally, minor-damaged and-major damaged buildings are uncommon in our samples. Therefore, we select larger weights for them (c = 2, 3) in damage scale classification loss. We also use weighted mixed loss in which focal loss accounts for a larger proportion to improve category imbalance.

To achieve better classification at the boundary, we expand the building damage scale labels. Given the overlap in pixel' labels, we prioritize minor damaged and major damaged buildings (c = 2, 3), which are relatively vulnerable in the classification.

5. Results and Discussion

5.1. Experimental Setting

In this work, we use PyTorch, a Python package that provides Tensor computation with strong GPU acceleration and deep neural networks built on a tape-based autograd system, as the deep learning framework. PyTorch is designed to be intuitive, linear in thought, and easy to use. Equipped with acceleration libraries such as Intel MKL and NVIDIA and custom memory allocators for GPU, PyTorch enables users to train larger deep learning models than with other Python packages.

All the experimentation and modeling tasks are implemented in the public cluster in the x64 Linux environment with the public computing cloud at the Renmin University of China. This computing cloud is equipped with the Simple Linux Utility for Resource Management (Slurm) scheduling system. Computations are performed on the node titan, which is configured with 128 GB of RAM, two Intel Gold 5218 CPUs, and two NVIDIA Titan RTX GPUs.

5.2. Ablation Study

In this study, we use an ablation experiment to demonstrate the effectiveness of our proposed method. An ablation study typically refers to subtracting a "feature" of the model or algorithm and verifying how this affects performance. Instead of subtracting, however, we gradually add modules such as Siamese, attention, and pyramid pooling into our proposed baseline network to verify its performance. Nevertheless, the improvement of the model performance is incompatible with different sectional tasks. Conducting experiments over several rounds guarantees that the modules of interest boost model performance.

Tables 5 and 6 show the results of the ablation experiment. The shaded row in the tables represents the performance of our proposed baseline model. The second row in Table 5 indicates that deploying the Siamese network module to the baseline model leads to a significant increment in all the metrics. Adding the attention module into the model results in a slight decline in all the metrics except the recall rate. The increase in the recall rate might be a consequence of the scale-aware semantic image segmentation that arises with an attention mechanism. We then introduce the PPM, which raises all the metrics except the recall rate slightly. This variation can be attributed to pyramid pooling, which enhances the scale invariance of images, while lowering the risk of over-fitting.

Table 5. Ablation experiments of the location methods with different modules (the shaded row represents the results of the ablated model).

	$\mathbf{IOU_{Non-building}}$(%)	$\mathbf{IOU_{Building}}$(%)	**Mean IoU**(%)	$\mathbf{Precision_{loc}}$(%)	$\mathbf{Recall_{loc}}$(%)	$\mathbf{F1_{loc}}$(%)	$\mathbf{Dice_{loc}}$(%)	$\mathbf{Score_{loc}}$(%)
Baseline model	94.91	52.57	73.74	54.70	75.27	63.36	95.14	56.07
+Siamese	96.98	66.07	81.53	73.93	82.42	77.95	95.98	61.97
+Siamese + Attention	96.60	65.45	81.03	64.98	87.26	74.49	96.15	60.90
+Siamese + PPM + Attention	97.00	67.33	82.17	71.15	85.58	77.70	95.95	66.40

Table 6. Ablation experiments of the multi-classification methods with different modules (the shaded row represents the results of the ablated model).

	P_{clf_0}(%)	R_{clf_0}(%)	$F1_{clf_0}$(%)	P_{clf_1}(%)	R_{clf_1}(%)	$F1_{clf_1}$(%)	P_{clf_2}(%)	R_{clf_2}(%)	$F1_{clf_2}$(%)	P_{clf_3}(%)	R_{clf_3}(%)	$F1_{clf_3}$(%)	$F1_{clf}$(%)
Baseline Model	87.22	93.04	90.04	54.64	26.20	35.43	48.14	56.41	51.95	85.41	45.02	58.96	52.95
+Siamese	90.19	79.10	84.28	22.59	55.14	32.05	67.24	65.25	66.23	92.07	55.73	69.44	55.12
+Siamese + Attention	91.35	77.26	83.72	22.52	56.60	32.22	61.73	66.64	64.10	83.07	62.31	71.21	55.08
+Siamese + PPM + Attention	90.64	89.07	89.85	35.51	49.50	41.36	65.80	64.93	65.36	87.08	57.89	69.55	61.55

Table 6 shows that sequentially applied modules improve overall performance since total F1, the harmonic mean of the F1 of each category, increases gradually with a mere recession. As for the irregular increment in the metrics, a trade-off between the precision rate and recall rate and the respective F1s of the different classes often results. For instance, deploying the Siamese network raises $F1_{clf_2}$ and $F1_{clf_3}$ and lowers $F1_{clf_0}$ and $F1_{clf_1}$. This is based on the decision boundaries, mutually exclusive in hyperspace, and generated by the recently attached module that changes when another module is consequently applied, leading to fluctuations in the metrics. Finally, the introduction of pyramid pooling, which is noteworthy for its scale-adaptive feature extracting ability, enables the model to yield rather satisfactory metrics for all the categories.

Table 7 also shows the confusion matrix of our final PPM-SSNet. Our model performs well overall and the non-building category holds the highest accuracy of 96.52%; whereas accuracy for the minor damage pixel is only 30.29%.

Table 8 compares the results between the post-and-pre strategy (both pre-disaster and post-disaster images are available) and the post-only strategy (only the post-disaster images are applied). According to the results, The performance of using only post-disaster image is lower than using both pre and post disaster images to locate buildings and assess damage levels,this demonstrates the important role of pre-disaster image in improving building localization and damage classification.

Table 7. Confusion matrix.

		Ground Truth				
		Non-Building	**No-Damage**	**Minor Damage**	**Major Damage**	**Destoryed**
Prediction	Non-building	2.88×10^8	2.16×10^8	2.60×10^8	2.84×10^8	2.05×10^6
	No-damage	2.22×10^7	3.67×10^7	8.31×10^5	3.76×10^5	7.43×10^4
	Minor damage	4.26×10^6	2.53×10^6	2.06×10^6	3.81×10^5	1.50×10^4
	Major damage	4.93×10^6	1.60×10^6	1.21×10^6	4.15×10^6	2.06×10^5
	Destoryed	1.39×10^6	4.09×10^5	1.12×10^5	1.28×10^5	1.95×10^6
Total		9.20×10^8	6.29×10^8	6.80×10^6	7.91×10^6	4.30×10^6
Accuracy(%)		96.52	58.35	30.29	52.47	45.35

Table 8. Comparison between the pre-and-post strategy and the post-only strategy.

Strategy	$\boldsymbol{Mean}\mathbf{IoU}_{\mathbf{Non-building}}(\%)$	$\boldsymbol{Mean}\mathbf{IoU}_{\mathbf{Building}}(\%)$	$\boldsymbol{Mean}\mathbf{IoU}_{\mathbf{loc}}(\%)$	$\mathbf{F1}_{\mathbf{loc}}(\%)$	$\mathbf{Score}_{\mathbf{loc}}(\%)$	$\mathbf{F1}_{\mathbf{clf_0}}(\%)$	$\mathbf{F1}_{\mathbf{clf_1}}(\%)$	$\mathbf{F1}_{\mathbf{clf_2}}(\%)$	$\mathbf{F1}_{\mathbf{clf_3}}(\%)$	$\mathbf{F1}_{\mathbf{clf}}(\%)$
post-only	91.88	47.32	69.60	56.94	58.16	82.84	38.16	63.23	71.10	58.69
pre-and-post	97.00	67.33	82.17	77.70	66.40	89.85	41.36	65.36	69.55	61.55

5.3. Comparisons with Other Methods

Since the release of the xBD dataset, some studies have divided a share of its data for training and achieved good results, whereas others use different evaluation metrics to assess accuracy. Moreover, some work is not strictly an end-to-end study, preventing us from being able to compare these published results with ours. To solve this problem, we reproduce previous research results and carry out comparative experiments under uniform experimental conditions. A Mask R-CNN network [19] and Siam-U-Net-Attention network [20] are compared.

Weber et al. [19] used the Mask R-CNN with the FPN architecture as well as the same model architecture for both building localization and damage classification. However, instead of working with full images, they trained the architecture on both the pre- and the post-image quadrants and fused the final segmentation layer to draw building boundaries more accurately. For the class imbalance problem, they engineered their loss function to weight errors on classes inversely proportional to their occurrence on the dataset. However, this is insufficient to address the problem. In practice, to solve class imbalance, we usually combine multiple approaches such as over-sampling and reweight operation with the weighted loss functions used in our experiment. Figure 8 shows the details of the Mask R-CNN network.

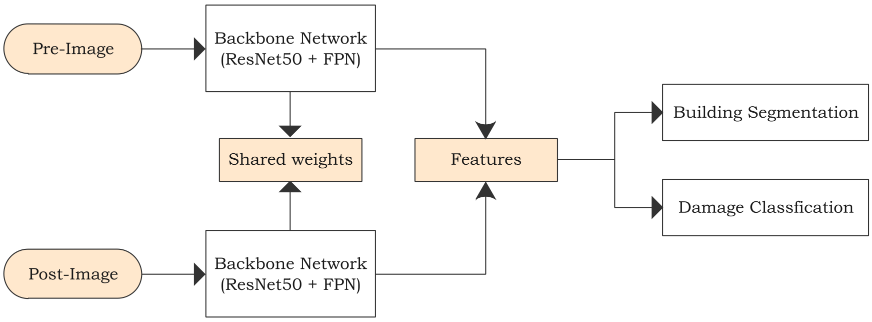


Figure 8. FPN R-CNN network.

Hao [20] designed a Siam-U-Net-Attn model end-to-end for both damage classification and building segmentation. One element of this architecture was a U-Net model that analyzed a single input image and produced a segmentation mask showing building locations. The same U-Net model was used for both the pre-disaster and the post-disaster images to produce binary masks. The features extracted from the encoder regions of the U-Net model also helped classify the damage scale. The two features produced by both the pre-image and the post-image U-Net encoder would be used by the middle part: a new separate decoder constitute in the Siamese network that compared the features from the two input frames to detect damage to buildings. The network achieved an appreciable IoU score on localization and performed well when classifying undamaged and destroyed buildings. However, the model could not identify minor-damaged and major-damaged buildings accurately. Figure 9 shows the Siam-U-Net-Attention network.

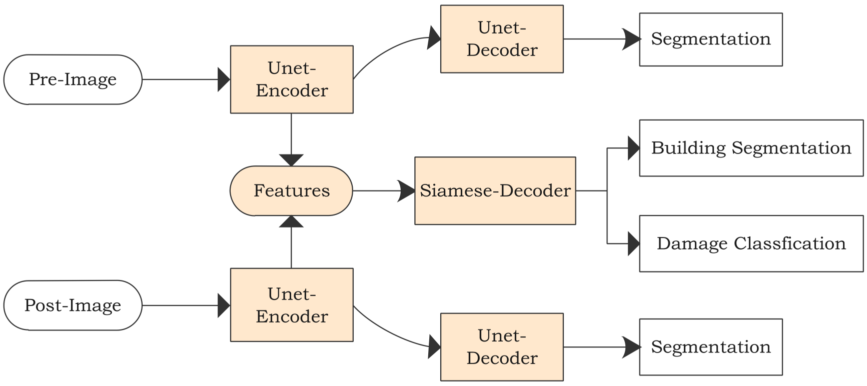


Figure 9. Siam-U-Net-Attention network model.

We train and test our network and other methods using the same datasets described above and same parameter settings. The results show that our proposed network easily outperforms the other approaches, as shown in Tables 9 and 10. We also compare the classification results of earthquakes, tsunamis, floods, typhoons, and volcanic eruptions, as shown in Figure 10. The results again verify the superiority of our method over previous approaches.

Table 9. Comparison with other methods on the location task.

Networks	Mean IOU$_{Non-building}$(%)	Mean IOU$_{Building}$(%)	Mean IoU(%)	Precision$_{loc}$(%)	Recall$_{loc}$(%)	F1$_{loc}$(%)
Siam-U-Net-Diff	96.50	44.57	70.54	52.75	90.75	66.72
Weber et al.	95.63	48.62	72.13	85.30	82.90	84.10
PPM-SSNet	97.00	67.33	82.17	71.15	85.58	77.70

Table 10. Comparison with other methods on the classification task.

Networks	P$_{clf_1}$(%)	R$_{clf_1}$(%)	F1$_{clf_1}$(%)	P$_{clf_2}$(%)	R$_{clf_2}$(%)	F1$_{clf_2}$(%)	P$_{clf_3}$(%)	R$_{clf_3}$(%)	F1$_{clf_3}$(%)	P$_{clf_4}$(%)	R$_{clf_4}$(%)	F1$_{clf_4}$(%)	F1$_{clf}$(%)
Siam-U-Net-Diff	80.58	49.64	60.51	28.69	26.32	27.45	51.31	27.60	35.89	75.00	33.03	45.86	39.01
Weber et al.	94.80	56.90	71.10	58.90	22.00	32.00	70.10	38.00	49.30	89.50	40.03	60.71	48.73
PPM-SSNet	90.64	89.07	89.85	35.51	49.50	41.36	65.80	64.93	65.36	87.08	57.89	69.55	61.55

Further, our model outperforms baseline models when predicting building localization and damage classification. Post-disaster images with destroyed buildings make a noise to building localization since the edges of destroyed buildings may be vague. FPN-R-CNN classified the majority of destroyed buildings into the no building category, while the U-Net-Siam-Attn's prediction of destroyed buildings is not robust. In these cases, our model can easily distinguish undamaged and destroyed buildings, but it is hard to distinguish minor from major damage.

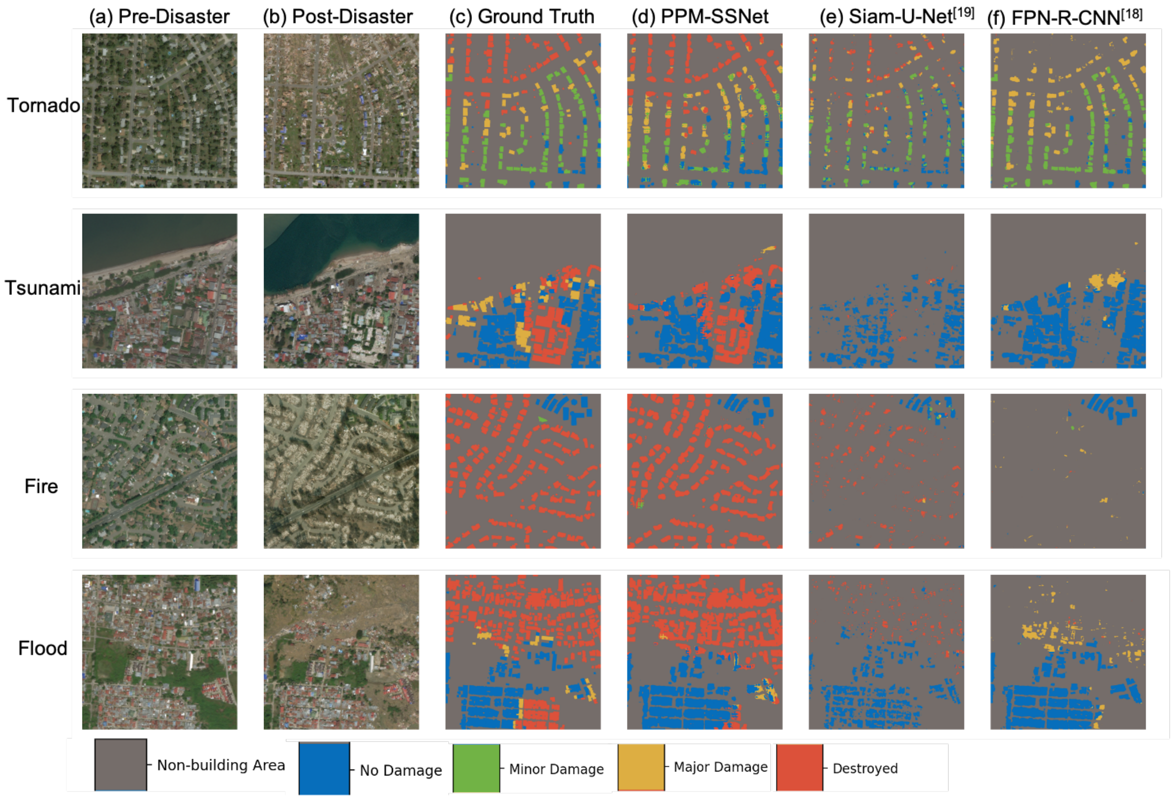


Figure 10. The results from our proposed method and comparisons with others. (**a**) Pre-disaster image; (**b**) Post-disaster image; (**c**) Ground truth; (**d**) Proposed PPM-SSNet model; (**e**) Siam-U-Net model; and (**f**) FPN-R-CNN model.

5.4. Robustness of the Method

The validation areas are characterized by a great diversity of environmental settings, building structures and spatial distributions, tsunami processes, and image acquisition conditions, as shown in Figure 11(a1,b1,a2,b2), respectively.

The predicted results show that the proposed model detects destroyed and undamaged buildings, but separating minor damage from major damage is still challenging, as shown in Figure 11(c1,d1,c2,d2) and Table 11. Partly because the Tohoku tsunami's annotation standard and that of the xBD dataset are not uniform. The Tohoku tsunami's building label is from a field survey, while the label of the xBD dataset comes from a visual interpretation, leading to an error in the like-for-like comparison. As the small validation area as shown in Figure 3c contains almost destroyed buildings, therefore we only did quantitative confusion matrix (Table 11) analysis for the larger validation area with variety of damage types as shown in Figure 3b. Still, We can visually interpret that the prediction results of the small validation area as shown in Figure 11(c1) are quite consistent with the ground truth data as shown in Figure 3c. Further, satellite remote sensing is limited when detecting fine-scale building damage because of its lower spatial resolution. Therefore, the method's inability to distinguish major and minor damage is logical. One way to solve this challenge would be to use high-resolution drone images. In general, our prediction results are consistent with the field observation data.

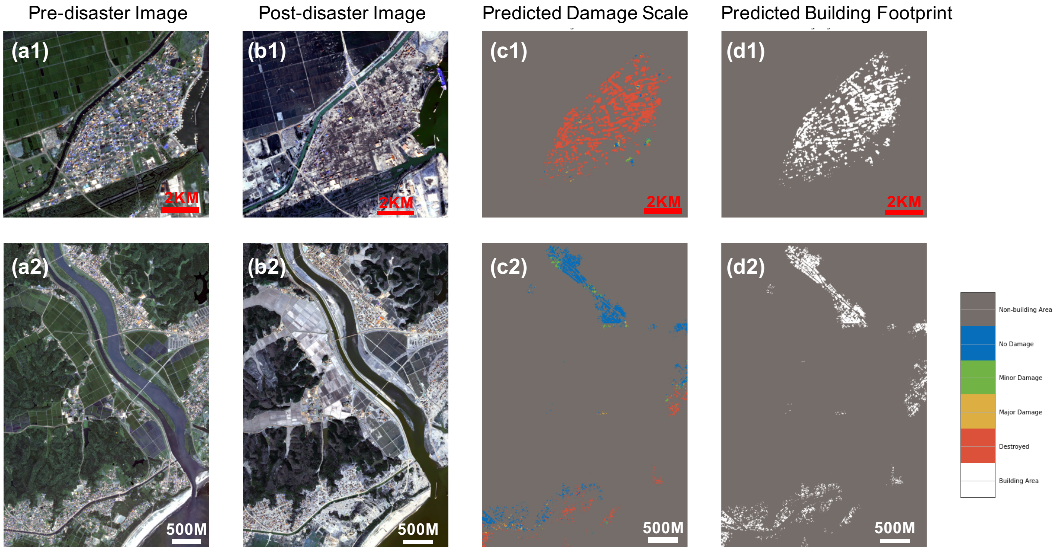


Figure 11. Prediction results from our proposed method in the validation areas. (**a1**,**a2**) Pre-disaster image; (**b1**,**b2**) Post-disaster image; (**c1**,**c2**) Predicted damage scale by the PPM-SSNet model; and (**d1**,**d2**) Prediction building footprint by the PPM-SSNet model.

Table 11. Confusion Matrix of Tohoku Tsunami Building Damage Prediction Experiment.

		Prediction				
		Non-Building	**No-Damage**	**Minor Damage**	**Major Damage**	**Destoryed**
Ground Truth	Non-building	38,960,379	66,366	50,870	19,195	34,488
	No-damage	215,480	368,283	862	1962	39,889
	Minor damage	58,680	2841	34,629	1736	8293
	Major damage	86,002	8	4331	43,611	3272
	Destoryed	196,579	80,942	12,550	6839	314,583
Total		39,517,120	518,080	103,242	73,343	400,525
Accuracy(%)		98.59	71.04	33.54	59.46	78.54

6. Conclusions

In this study, we developed an end-to-end attention-guided semi-Siamese network with a pyramid-pooling module. Our proposed model yielded satisfactory results when focusing on building localization and damage classification compared with other methods. Employing dilated convolution, the method leveraged the global and local features of an input image. To improve damage classification performance, we adopted a squeeze-and-excitation mechanism, a weighting system that produces and applies channel-wise weights on a feature map. Our ablation experiments on the xBD dataset demonstrated that the proposed semi-Siamese network, dilated convolution, and squeeze-and-excitation mechanism were both necessary and effective. Meanwhile, the demonstration with 2011 Great East Japan Earthquake data revealed consistent results with the ground truth data, confirming the effectiveness of evaluating future disasters using our proposed method. Further, it achieved true end-to-end input and output. Thanks to the open source of the large-scale high-precision xBD dataset, which used to be the main challenge of training deep learning models for building damage assessment from satellite imagery, it has become unnecessary to xxxx. Nevertheless, the contribution of this research is developing a damage detection algorithm based on large-scale benchmark data from multiple types of disasters. Therefore, we do not provide targeted solutions for a specific type of disaster.

Our research has some limitations. It is based on the visual information of optical images, meaning that it may be unable to measure extensive flood damage under an intact roof. To address this, researchers could consider using synthetic aperture radar images to detect bottom or sidewall damage [35]. In addition, wall ruptures caused by earthquakes may not be effectively measured, which could be overcome using higher resolution drone images to detect this type of damage [36]. These limitations suggest that despite the contributions of the proposed approach, a highly robust and transplant deep learning model for assessing building damage with high precision is still urgently needed. Since domain shift is still an important challenge in deep learning, satellite imagery is particularly problematic in this field, and this will be the direction of our future efforts.

Author Contributions: Conceptualization, Y.B.; methodology, Y.B., J.H., J.S., X.L.; software, X.H., H.L.; validation, X.H., H.L.; formal analysis, Y.B., J.H., J.S., X.H., H.L., X.L.; investigation, Y.B., E.M., S.K.; resources, Y.B., S.M., E.M., S.K.; data curation, Y.B., E.M., and S.K.; writing—original draft preparation, Y.B., J.H., J.S., X.H., H.L., X.L.; visualization, X.H., H.L.; supervision, Y.B., S.M., X.L., E.M., S.K.; project administration, Y.B., S.M., E.M., S.K.; funding acquisition, Y.B., S.M., E.M., S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the Fundamental Research Funds for the Central Universities, Research Funds of Renmin University of China (20XNF022), fund for building world-class universities (disciplines) of Renmin University of China, Major projects of the National Social Science Fund (16ZDA052), Japan Society for the Promotion of Science Kakenhi Program (17H06108), and Core Research Cluster of Disaster Science and Tough Cyberphysical AI Research Center at Tohoku University.

Acknowledgments: This work was supported by the Public Computing Cloud, Renmin University of China. We also thank the SmartData Club, an Entrepreneurship Incubation Team lead by Jinhua Su of Renmin University of China; Wenqi Wu, students from Renmin University of China; and the Core Research Cluster of Disaster Science at Tohoku University (a Designated National University) for their support. We thank the two reviewers for their helpful and constructive comments on our work. The author gratefully acknowledges the support of K.C. Wong Education Foundation, Hong Kong.

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

Abbreviations

The following abbreviations are used in this manuscript:

PPM-SSNet	Pyramid Pooling Module-based Semi-Siamese Network
PPM	Pyramid Pooling Module
CNN	Convolutional Neural Network
IoU	Intersection over Union
SE	Squeeze-and-Excitation
RBs	Residual Blocks

References

1. Hillier, J.K.; Matthews, T.; Wilby, R.L.; Murphy, C. Multi-hazard dependencies can increase or decrease risk. *Nat. Clim. Chang.* **2020**, *10*, 595–598. [CrossRef]
2. Koshimura, S.; Shuto, N. Response to the 2011 great East Japan earthquake and tsunami disaster. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *373*, 20140373. [CrossRef] [PubMed]

3. Mascort-Albea, E.J.; Canivell, J.; Jaramillo-Morilla, A.; Romero-Hernández, R.; Ruiz-Jaramillo, J.; Soriano-Cuesta, C. Action protocols for seismic evaluation of structures and damage restoration of residential buildings in Andalusia (Spain): "IT-Sismo" APP. *Buildings* **2019**, *9*, 104. [CrossRef]
4. 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]
5. Suppasri, A.; Koshimura, S.; Matsuoka, M.; Gokon, H.; Kamthonkiat, D. Application of remote sensing for tsunami disaster. *Remote Sens. Planet Earth* **2012**,143–168. [CrossRef]
6. Bai, Y.; Adriano, B.; Mas, E.; Koshimura, S. Building damage assessment in the 2015 Gorkha, Nepal, earthquake using only post-event dual polarization synthetic aperture radar imagery. *Earthq. Spectra* **2017**, *33*, 185–195. [CrossRef]
7. Bai, Y.; Gao, C.; Singh, S.; Koch, M.; Adriano, B.; Mas, E.; Koshimura, S. A framework of rapid regional tsunami damage recognition from post-event TerraSAR-X imagery using deep neural networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 43–47. [CrossRef]
8. Moya, L.; Mas, E.; Koshimura, S. Learning from the 2018 Western Japan Heavy Rains to Detect Floods during the 2019 Hagibis Typhoon. *Remote Sens.* **2020**, *12*, 2244. [CrossRef]
9. Koshimura, S.; Moya, L.; Mas, E.; Bai, Y. Tsunami Damage Detection with Remote Sensing: A Review. *Geosciences* **2020**, *10*, 177. [CrossRef]
10. 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]
11. Nex, F.; Duarte, D.; Tonolo, F.G.; Kerle, N. Structural building damage detection with deep learning: Assessment of a state-of-the-art cnn in operational conditions. *Remote Sens.* **2019**, *11*, 2765. [CrossRef]
12. Xu, J.Z.; Lu, W.; Li, Z.; Khaitan, P.; Zaytseva, V. Building damage detection in satellite imagery using convolutional neural networks. *arXiv* **2019**, arXiv:1910.06444.
13. Rudner, T.G.; Rußwurm, M.; Fil, J.; Pelich, R.; Bischke, B.; Kopačková, V.; Biliński, P. Multi3Net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 702–709.
14. Doshi, J.; Basu, S.; Pang, G. From satellite imagery to disaster insights. *arXiv* **2018**, arXiv:1812.07033.
15. Gupta, R.; Hosfelt, R.; Sajeev, S.; Patel, N.; Goodman, B.; Doshi, J.; Heim, E.; Choset, H.; Gaston, M. xbd: A dataset for assessing building damage from satellite imagery. *arXiv* **2019**, arXiv:1911.09296.
16. Gupta, R.; Goodman, B.; Patel, N.; Hosfelt, R.; Sajeev, S.; Heim, E.; Doshi, J.; Lucas, K.; Choset, H.; Gaston, M. Creating 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, Long Beach, CA, USA, 16–20 June 2019.
17. Gupta, R.; Shah, M. RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery. *arXiv* **2020**, arXiv:2004.07312.
18. Cooner, A.J.; Shao, Y.; Campbell, J.B. Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake. *Remote Sens.* **2016**, *8*, 868. [CrossRef]
19. Weber, E.; Kané, H. Building Disaster Damage Assessment in Satellite Imagery with Multi-Temporal Fusion. *arXiv* **2020**, arXiv:2004.05525.
20. Hao, H.; Baireddy, S.; Bartusiak, E.R.; Konz, L.; LaTourette, K.; Gribbons, M.; Chan, M.; Comer, M.L.; Delp, E.J. An Attention-Based System for Damage Assessment Using Satellite Imagery. *arXiv* **2020**, arXiv:2004.05525.
21. Nia, K.R.; Mori, G. Building damage assessment using deep learning and ground-level image data. In Proceedings of the 2017 14th Conference on Computer and Robot Vision (CRV), Edmonton, AB, Canada, 16–19 May 2017; pp. 95–102.
22. Valentijn, T.; Margutti, J.; van den Homberg, M.; Laaksonen, J. Multi-Hazard and Spatial Transferability of a CNN for Automated Building Damage Assessment. *Remote Sens.* **2020**, *12*, 2839. [CrossRef]
23. Harirchian, E.; Lahmer, T.; Kumari, V.; Jadhav, K. Application of Support Vector Machine Modeling for the Rapid Seismic Hazard Safety Evaluation of Existing Buildings. *Energies* **2020**, *13*, 3340. [CrossRef]

24. Zhuo, G.; Dai, K.; Huang, H.; Li, S.; Shi, X.; Feng, Y.; Li, T.; Dong, X.; Deng, J. Evaluating potential ground subsidence geo-hazard of Xiamen Xiang'an new airport on reclaimed land by SAR interferometry. *Sustainability* **2020**, *12*, 6991. [CrossRef]
25. Morfidis, K.E.; Kostinakis, K.G. Use of Artificial Neural Networks in the R/C Buildings'seismic Vulnerabilty Assessment: The Practical Point of View. In Proceedings of the 7th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Crete, Greece, 24–26 June 2019.
26. Harirchian, E.; Lahmer, T.; Rasulzade, S. Earthquake Hazard Safety Assessment of Existing Buildings Using Optimized Multi-Layer Perceptron Neural Network. *Energies* **2020**, *13*, 2060. [CrossRef]
27. Morfidis, K.; Kostinakis, K. Seismic parameters' combinations for the optimum prediction of the damage state of R/C buildings using neural networks. *Adv. Eng. Softw.* **2017**, *106*, 1–16. [CrossRef]
28. Takahashi, T.; Mori, N.; Yasuda, M.; Suzuki, S.; Azuma, K. The 2011 Tohoku Earthquake Tsunami Joint Survey (TTJS) Group. Available online: http://www.coastal.jp/tsunami2011 (accessed on 30 November 2020).
29. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
30. Zhang, H.; Patel, V.M. Densely connected pyramid dehazing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3194–3203.
31. Liu, X.; Suganuma, M.; Sun, Z.; Okatani, T. Dual residual networks leveraging the potential of paired operations for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 7007–7016.
32. Li, Y.; Song, L.; Chen, Y.; Li, Z.; Zhang, X.; Wang, X.; Sun, J. Learning Dynamic Routing for Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–28 June 2020.; pp. 8553–8562.
33. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
34. 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]
35. Yamazaki, F.; Iwasaki, Y.; Liu, W.; Nonaka, T.; Sasagawa, T. Detection of damage to building side-walls in the 2011 Tohoku, Japan earthquake using high-resolution TerraSAR-X images. In Proceedings of the Image and Signal Processing for Remote Sensing XIX, Dresden, Germany, 23–26 September 2013; Volume 8892, p. 889212.
36. Duarte, D.; Nex, F.; Kerle, N.; Vosselman, G. Towards a more efficient detection of earthquake induced facade damages using oblique UAV imagery. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 93. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.