Article

A Comparative Study of Texture and Convolutional Neural Network Features for Detecting Collapsed Buildings After Earthquakes Using Pre- and Post-Event Satellite Imagery

Min Ji [1], Lanfa Liu [1],* [ID], Runlin Du [2] and Manfred F. Buchroithner [1] [ID]

[1]   Institute for Cartography, TU Dresden, 01062 Dresden, Germany; Min.Ji@tu-dresden.de (M.J.); Manfred.Buchroithner@tu-dresden.de (M.F.B.)
[2]   Qingdao Institute of Marine Geology, Qingdao 266071, China; durunlin123@163.com
*   Correspondence: Lanfa.liu@mailbox.tu-dresden.de

check for updates

Abstract: The accurate and quick derivation of the distribution of damaged building must be considered essential for the emergency response. With the success of deep learning, there is an increasing interest to apply it for earthquake-induced building damage mapping, and its performance has not been compared with conventional methods in detecting building damage after the earthquake. In the present study, the performance of grey-level co-occurrence matrix texture and convolutional neural network (CNN) features were comparatively evaluated with the random forest classifier. Pre- and post-event very high-resolution (VHR) remote sensing imagery were considered to identify collapsed buildings after the 2010 Haiti earthquake. Overall accuracy (OA), allocation disagreement (AD), quantity disagreement (QD), Kappa, user accuracy (UA), and producer accuracy (PA) were used as the evaluation metrics. The results showed that the CNN feature with random forest method had the best performance, achieving an OA of 87.6% and a total disagreement of 12.4%. CNNs have the potential to extract deep features for identifying collapsed buildings compared to the texture feature with random forest method by increasing Kappa from 61.7% to 69.5% and reducing the total disagreement from 16.6% to 14.1%. The accuracy for identifying buildings was improved by combining CNN features with random forest compared with the CNN approach. OA increased from 85.9% to 87.6%, and the total disagreement reduced from 14.1% to 12.4%. The results indicate that the learnt CNN features can outperform texture features for identifying collapsed buildings using VHR remotely sensed space imagery.

Keywords: earthquake; grey-level co-occurrence matrix texture; convolutional neural network; CNN; random forest

1. Introduction

Buildings are fundamental for human living. However, they are vulnerable to natural hazards. Buildings are usually seriously damaged or completely destroyed by earthquakes, such as the Sichuan (2008), Chile (2010), Haiti (2010), and Nepal (2015) earthquakes. Therefore, it is vital to monitor the status of buildings and provide high-precision building damage assessment after an earthquake at a detailed scale to support the emergency response and rescue activities. Considering the demand of the timely retrieval of disaster damage information, the implementation of satellite-based methods to assess damage in buildings has raised more and more attention after an earthquake, especially where the road connections are blocked or destroyed and access became thus difficult [1–3].

Remote sensing has been widely utilized for various disasters as it can capture affected areas from space [4,5]. Synthetic aperture radar (SAR), light detection and ranging (LiDAR), and optical techniques have been adopted to detect and assess damaged buildings after an earthquake and have achieved great success for comparatively low costs, minimal corresponding fieldwork, large coverage, digital processing, and quantitative results [6]. SAR is strongly sensitive to surface changes based on the backscatter coefficient and intensity correlation [7]. In the previous study, object-based image analysis (OBIA) was proposed in Reference [8] and evaluated on post-event ALOS-2/PALSAR-2 dual polarimetric SAR imagery after the 2015 Nepal earthquake [9]. LiDAR imagery can provide information of height change, while optical imagery has become an important way to identify damaged buildings with the improvement of spatial resolution and image quality. It is also possible to integrate different methods to detect damaged buildings and produce accurate and reliable results. A novel method was proposed to detect damaged buildings using high-resolution remote sensing images and three-dimensional GIS data by Tu et al. in Reference [10]. Remote sensing and GIS can be used to not only detect earthquake damage, but also to monitor the recovery after earthquakes, like how remote sensing and GIS were applied to monitor the recovery after the 2009 L'Aquila earthquake, Italy [11].

Compared with automatic methods, the visual method [12] is usually time-consuming, which is disadvantageous for planning rescue and for the generation of building damage maps using satellite or aerial imageries. OBIA has been applied to detect earthquake damage using remote sensing imagery since 1988 [13]. The OBIA approach is usually performed in two steps. First, the input image is segmented, and then each segmented object is assigned to a class by a classification algorithm [14]. OBIA provides an automated method for the analysis of high-resolution imagery by describing the object using spectral, textural, spatial, and topological properties. An adaptive-network-based fuzzy inference system (ANFIS) model was designed to attain the building damage degrees, and OBIA played a key role in detecting damaged buildings using high-resolution imagery after the earthquake in 2004 Bam, Iran [15]. OBIA was already incorporated with random forest classifier to identify damaged buildings [16]. In high-resolution imagery, the pixel size is significantly smaller than the average size of the object of interest. OBIA can group pixels into objects based on spectral similarity, and thus, showed better performance compared to pixel-based classification [17]. The commercial image segmentation and classification software eCognition has been widely used in detecting earthquake damage. For earthquake-collapsed building extraction from LiDAR and aerophotograph based on OBIA, eCognition software was used to segment imagery. Texture features (contrast, dissimilarity, and variance) were calculated based on the gray level co-occurrence matrix (GLCM), and support vector machine (SVM) was chosen as the classifier to identify the collapsed buildings [18,19]. eCognition software was also considered for image segmentation and classification to detect damaged buildings after the 2010 Haiti earthquake [20]. OBIA has shown the potential for earthquake damage detection, while the application of convolutional neural networks (CNNs) is still limited and worth to be explored to discover its advantages.

Deep learning is able to improve the efficiency of damage recognition because of its high ability of automatic feature learning and visual pattern recognition [21]. There are many kinds of CNN structures, including GoogleNet [22] with 22 layers for classification when counting only layers (or 27 layers when also counting pooling), VGG [23] with 11 to 19 weight layers for large image recognition, AlexNet [24] with five convolutional layers (some of them followed by max-pooling layers) and three full-connected layers, which were used to classify 1.2 million high-resolution images in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) into the 1000 different classes. Network in network (NiN) was proposed by replacing the liner filter and full connected layer with nonlinear multilinear perceptron and global average pooling, respectively [25]. CNNs have achieved superior results in various tasks including image classification, speech recognition, and object detection [24,26–29]. However, there are limited studies related to earthquake-induced building damage mapping using deep learning and its robustness and applicability of building damage mapping should be explored. A SqueezeNet-similar CNN structure was adopted to identify the collapsed buildings after the 2010 Haiti earthquake using

only post-event satellite imagery [30]. CNN and three-dimensional features, both independently and in combination, were considered to detect damaged buildings, and the results showed that the integration of CNN and 3D point cloud features significantly improved the model transferability and achieved accuracy was improved up to a maximum of 7% compared with the result achieved by CNN features alone [31].

The objective of the present study was to explore the performance of using texture and CNN features respectively integrated by random forest for building damage assessment using pre- and post-earthquake VHR satellite imagery. The dataset obtained after the 2010 Haiti earthquake were used in this study. A neural network containing three convolutional layers were implemented to automatically learn features. GLCM texture features (including contrast, dissimilarity, entropy, homogeneity, correlation, and angular second moment) were extracted from pre- and post-event data using eCognition. Subsequently, random forest was used as the classifier to evaluate the performance of these two different features. The rest of this paper is organized as follows. Section 2 describes the study area and dataset. Section 3 introduces the basic concepts of convolutional neural networks, textural feature extraction, random forest, and evaluation metrics. The results are then provided in Section 4 and discussed in Section 5. Conclusion is described in Section 6.

2. Study Area

After the earthquake in Haiti on January 12, 2010, there were about 105,000 houses completely destroyed, especially in its capital, Port-au-Prince, where some important buildings collapsed, such as the Port-au-Prince Cathedral and the Presidential Palace. Buildings in the study area were highly vulnerable to earthquakes since little or no seismic design was applied. Building structures in the capital mainly included concrete structures with flat roofs of varying heights and sizes, wooden or steel frame buildings with corrugated metal sheet roofs, and low metal sheet shelters (shanty housing) with very small dwellings [32]. Based on the European Macroseismic Scale 1988 (EMS-98) [33], the damage of buildings can be classified into five grades as shown in Figure 1.

Masonry buildings	Reinforced concrete buildings	Classification of damages
		Grade 1: Negligible to slight damage (no structural damage, slight non-structural damage)
		Grade 2: Moderate damage (slight structural damage, moderate non-structural damage)
		Grade 3: Substantial to heavy damage (moderate Structural damage, heavy non-structural damage)
		Grade 4: Very heavy damage (heavy structural damage, very heavy non-structural damage)
		Grade 5: Destruction (very heavy structural damage)

Figure 1. Classification of damage to masonry and reinforced buildings (taken from EMS-98).

The proposed method was applied to the area of Port-au-Prince to distinguish collapsed and noncollapsed buildings. Considering limited data availability, the pre-event WorldView-2 (WV-2) data acquired on January 9, 2010 (Figure 2A) and post-event QuickBird (QB) data acquired on January 15, 2010 (Figure 2B) were utilized in this study. The data were obtained via DigitalGlobe open data program with the identical spatial resampling resolution of 0.5 m. Furthermore, building damage information was also used to evaluate the performance of the model. The building damage information was acquired from UNITAR/UNOSAT and created by visual interpretation of high-resolution satellite images and aerial photos observed after the earthquake [34]. In the present study, the dataset is randomly divided into training and test dataset, and the number of training and testing buildings is 1074 and 716, respectively.

Figure 2. Very high-reslution (VHR) imagery in the study area before (**A**) and after (**B**) the 2010 Haiti earthquake and building footprints with damage grades (**C**). Grade 2 was not presented in the dataset.

3. Methodology

The purpose of this study is to explore the performance of texture and CNN features for identifying collapsed buildings after the earthquake. It should be mentioned that it is hard to identify small dwelling units with metal rooves, even when VHR imagery is used [15]. Therefore, the study mainly focuses on distinguishing between collapsed and noncollapsed buildings. The damage grades were classified into binary categories: Collapsed (G5) and noncollapsed (G1–G4) buildings as labels for the corresponding objects for further study. Building patches were extracted from pre- and post-event satellite images with manually prepared building footprints. In general, CNNs accept input image patches with the same dimension. Thus, building patches were scaled to have the same pixel size of 96 by 96 [30]. Ideally, the CNNs can learn useful task-oriented features from the training data. Thus, the pre- and post-event data were combined by simply concatenating them. The workflow was shown in Figure 3 and the main steps were further described in the following sub-sections. The CNN model was trained and then used as a feature extraction tool to derive CNN features. The basic concept of CNNs and the adopted CNN structure were presented in Section 3.1. For texture features as described in Section 3.2, a total of 12 features was calculated for each building object using pre- and post-event satellite images by means of eCognition software. Random forest was briefly explained in Section 3.3, which was chosen as the classifier to compare the performance of texture and CNN features for identifying collapsed buildings after the 2010 Haiti earthquake. Evaluation metrics were described in Section 3.4.
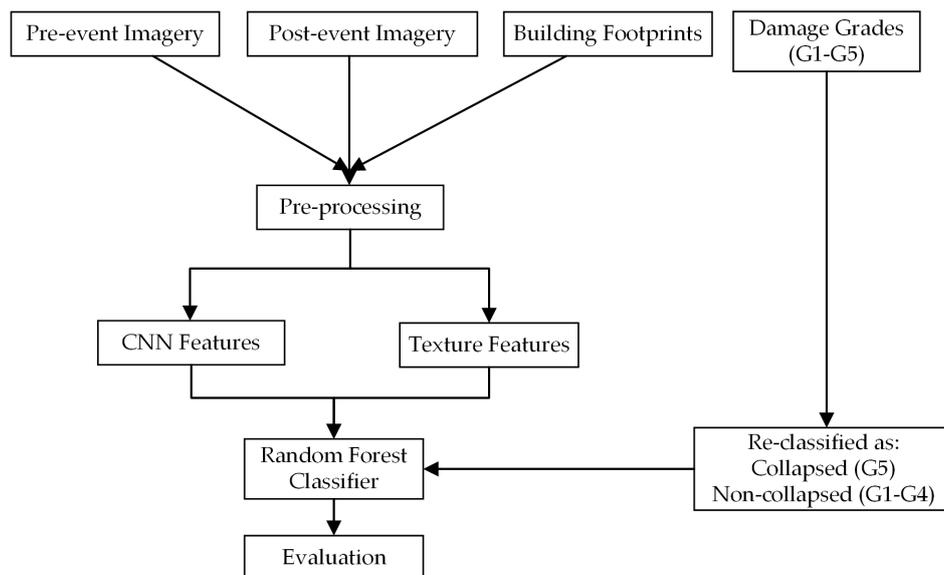

Figure 3. The flowchart of comparing convolutional neural network (CNN) and texture features with random forest for deriving collapsed building information after the earthquake.

3.1. CNNs

Compared to conventional classification methods, CNNs can provide better classification performance due to its capability of learning high-level features from a large number of training dataset [35]. To ensure the shift and distortion invariance, three architectural ideas were proposed in convolutional neural networks including local receptive fields, shared weights, and sometimes spatial or temporal subsampling [36]. The number of parameters could be reduced using the weight-sharing technique to solve the problems of sophisticated and complicated hyperparameters in deep neural networks. The standard CNN structure consists of convolutional, pooling, and fully connected layers. The convolution layers are used to generate feature maps by linear convolutional filters followed by nonlinear activation functions such as sigmoid, tanh, softmax, and Rectified Linear Unit (ReLU). The feature maps obtained after convolutional layers are often subsampled by pooling layers to reduce the

dimensionality. Max-pooling operation is commonly adopted in the pooling layer, which is a sort of downsampling where we can preserve the locations of the original image that showed the strongest correlation to the specific features. Fully connected layer is often used to combine the local features into the global features. Global average pooling is also proposed to minimize overfitting by reducing the total number of parameters in the model [37]. It averages all the features within the spatial region. Given a feature map of the size *height × width × depth*, global average pooling reduces that to a $1 \times 1 \times$ *depth* feature map. Compared to fully connected layers, it is more native to the convolution structure by enforcing correspondences between feature maps and target categories.

　　The adopted CNN structure in this study is shown in Table 1. It comprises three convolutional and activation layers, two max-pooling layers, a global average pooling layer, and a traditional fully connected layer. Max-pooling operation is utilized following activation layers. The input values of width, length, and band values are 96, 96, and 3, respectively. The stride is 1×1 in the first and second convolutional layers. The first convolutional layer has 32 filters with the window size of 3×3 followed by a pooling layer of 32 filters of window size 2×2, and the second convolutional layer has 64 filters with the window size of 3×3. The window size should be larger than the smallest size of building or the intermediate output. Thus, the output size after the first pooling operation was $47 \times 47 \times 32$. The activation feature maps acquired by the activation layer were pooled with a 2×2 max-pooling window, and would be used as the input data for the next convolutional layer. The drop probability was set to 0.5 in the dropout. ReLU was used as the activation function in the model. The CNN model was implemented using Keras 2.4 with tensorflow 1.7 as the backend. The model was trained using the prepared Haiti earthquake dataset and functioned as the feature extractor. The extracted CNN features were then used as input for the random forest classifier, which was term as CNN-RF for convenience, to distinguish collapsed and noncollapsed buildings.

Table 1. CNN structure in the present study.

Layer	Shape (N, Width, Length, Bands)	No. of Parameters
Conv2d_1	(N, 94, 94, 32)	896
Activation_1	(N, 94, 94, 32)	0
Max_pooling2d_1	(N, 47, 47, 32)	0
Conv2d_2	(N, 45, 45, 64)	18,496
Activation_2	(N, 45, 45, 64)	0
Max_pooling2d_2	(N, 22, 22, 64)	0
Conv2d_3	(N, 20, 20, 64)	36,928
Activation_3	(N, 20, 20, 64)	0
Global_average_pooling2d_1	(N, 64)	0
Dense_1	(N, 32)	2080
Dropout_1	(N, 32)	0
Softmax	(N, 2)	66
Total		58,466

3.2. GLCM Texture Features

　　GLCM has been proven to be a popular statistical method of extracting textural feature from images. A variety of GLCM derived features have been applied in previous studies related to damaged buildings induced by an earthquake [37–40]. According to the co-occurrence matrix, Haralick defined 14 texture features measured from the probability matrix to extract the characteristics of texture statistics of images [19]. In this paper, six texture features (contrast, dissimilarity, entropy, homogeneity, correlation, angular second moment) that have been correlated with earthquake-induced damaged buildings in previous studies [16,41–43] were selected and calculated with the eCognition software for pre- and post-event satellite imagery. A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels in the image. The eigenfunctions (Equations (1)–(6)) are as follows: Where i is the row number, j is the column number, $p(i, j)$ is the $(i, j)^{th}$ entry in a normalized GLCM. N

is the number of gray levels. σ_i and σ_j are the standard deviations for row i and column j. u_i and u_j are the means of row i and column j.

GLCM contrast measures local variation among neighbours in the image. If gray value difference among neighbors is high, the contrast values will be high. It can be calculated as follows:

$$\text{Contrast} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} p(i,j) * (i-j)^2 \tag{1}$$

GLCM dissimilarity is similar to the contrast and inversely related to homogeneity. Dissimilarity is also high when the contrast of area is high. The formula is as follows:

$$\text{Dissimilarity} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} p(i,j) * |i-j| \tag{2}$$

GLCM homogeneity reflects the texture homogeneity. A large value indicates strong texture homogeneity and elements concentrating on the main diagonal. The equation reads as follows:

$$\text{Homogeneity} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \frac{p(i,j)}{1 + (i-j)^2} \tag{3}$$

GLCM entropy measures the disorder in an image. High entropy indicates the heterogeneous texture of the image and low entropy indicates a homogeneous texture [40]. The value is large when the elements of GLCM are equal. The equation can be expressed as follows:

$$\text{Entropy} = -\sum_{i=1}^{N-1} \sum_{j=1}^{N-1} p(i,j) * \ln p(i,j) \tag{4}$$

GLCM correlation represents the linear dependence of gray levels on those of neighbors and is calculated as follows:

$$\text{Correlation} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \frac{(i-u_i)(i-u_j)p(i,j)}{\sigma_i \sigma_j} \tag{5}$$

GLCM angular second moment (ASM) is also called energy and measures textural uniformity. The value will be close to maximum when the image patch is homogeneous. It is expressed as follows:

$$\text{ASM} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} p(i,j)^2 \tag{6}$$

3.3. Random Forest

Random forest is an ensemble method consisting of classification and regression trees [44]. The final classification decision is taken by averaging the class assignment probabilities calculated by all produced trees [45]. There are two random procedures in random forest. First, the training set is created for each tree by sampling with replacement (bootstrapping) from the original training dataset. Second, random features are selected with nonreplacement from the total features when the nodes of the trees are split [46]. It has been adopted widely since it can increase the robustness and performance of classification based on bootstrap sampling and random feature combination strategy [9]. Furthermore, it has the capability of handing a number of variables, ranking those variables, evaluating the importance of variables based on the performance, and finding the computationally optimal number of trees through testing the algorithm [47,48]. Random forest can overcome the overfitting problem

of decision trees, and has a strong anti-interference ability for noise and outliers. It has already been demonstrated to be a powerful classifier to detect the damaged buildings caused by earthquakes using SAR imagery [49,50]. In this study, random forest was chosen as the classifier to identify collapsed and standing buildings, and implemented in a production-ready Python library scikit-learn.

3.4. Evaluation Metrics

In this study, in order to accurately assess the performance of building damage, allocation disagreement (AD) and quantity disagreement (QD), overall accuracy (OA), user accuracy (UA), producer accuracy (PA), and Kappa were adopted as evaluation metrics. OA is the proportion of buildings that are correctly identified. It is one of the most popular agreement measures. However, it is pointed out that it has different thresholds for image classification in different scenarios [51–53], so it is difficult to define acceptable threshold values of OA. PA is the probability that a value in a given class was classified correctly. UA is the probability that a value predicted to be in a certain class really is that class. Kappa indices are common evaluation parameters in the remote sensing literature, since they also compare two maps that show a set of categories. Recent studies, however, suggested that Kappa has some limitations, and Pontius Jr. and Millones [54] state that standard Kappa is frequently complicated to compute, difficult to understand and unhelpful to interpret, and recommend that Kappa should be replaced by QD and AD. The QD is defined as the amount of difference between the reference and the observed maps because of the imperfect match in the proportions of the damaged building classes, and the AD is the amount of difference between the reference and observed maps because of the imperfect match in the spatial allocation of damaged building classes [38]. AD and QD can be calculated by Equations (8) and (9), respectively. The proportion of agreement C is estimated by Equation (10). The total disagreement D is the sum of AD and QD. Where q_g and a_g are the quantity disagreement and the allocation disagreement of land use class g; J is the number of damaged building categories; p_{ij} is the number of sample classified as i and referenced as j; p_{ig} is the estimated proportion of study area classified as i and referenced as g; N_i is the number of buildings damage class i [55].

$$p_{ij} = \left(\frac{n_{ij}}{\sum_{j=1}^{J} n_{ij}} \right) \left(\frac{N_i}{\sum_{i=1}^{J} N_i} \right) \tag{7}$$

$$QD = \frac{1}{2} \sum_{g=1}^{J} q_g = \frac{1}{2} \sum_{g=1}^{J} \left| \left(\sum_{i=1}^{J} p_{ig} \right) - \left(\sum_{j=1}^{J} p_{gj} \right) \right| \tag{8}$$

$$AD = \frac{1}{2} \sum_{g=1}^{J} a_g = 2\min \left[\left(\sum_{i=1}^{J} p_{ig} - p_{gg} \right), \left(\sum_{j=1}^{J} p_{gj} \right) - p_{gg} \right] \tag{9}$$

$$C = \sum_{g=1}^{J} p_{jj} \tag{10}$$

$$D = 1 - C = QD + AD \tag{11}$$

4. Results

4.1. Performance of CNN-RF

The CNN model should be established before using it as a feature extractor. For the training parameters, the Adam optimizer [56] was used with a learning rate parameter of 0.001 and 500 epochs. The batch size was set to 64. Each building object was scale to 96 × 96 pixels. Validation/training losses and accuracies of the model are depicted in Figure 4. Both validation and training losses could be significantly reduced after 300 epochs, while the accuracies could increase correspondingly. Thus, an

early-stopping technique could be considered to reduce the amount of training time and also to avoid overfitting. The curves are fluctuating, as we used a mini-batch training method.
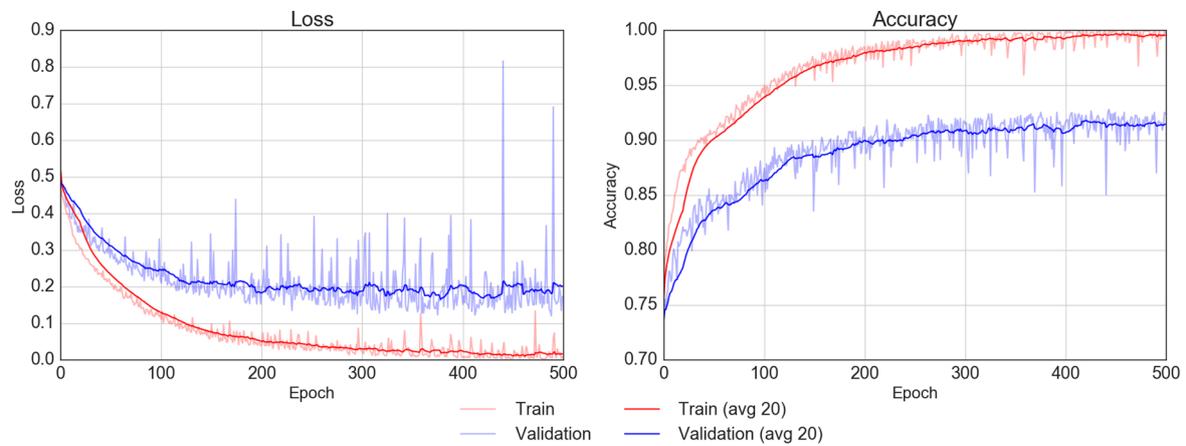

Figure 4. Loss and accuracy curves of the CNN model.

Convolutional neural networks can automatically learn robust and representative features layer by layer. The performance of output features from intermediate layers is shown in Figure 5. The accuracies with output features from the third layer, sixth layer, eighth layer, and ninth layer were obtained by the random forest classifier. It is clear that the values of OA and Kappa consistently increased along with the depth of layers, while D (the sum of AD and QD) values showed an opposite tendency. The QD and AD values gradually decreased except the AD in the eighth layer and QD in the sixth layer which increased compared to the previous layer. Especially, the performance improved significantly when intermediate layer outputs were combined with the aggressive global pooling operation (low resulting dimensionality). Finally, using features from the ninth layer (global average pooling layer) as input for the random forest classifier, the best result was achieved with 87.6%, 72.5%, and 12.4% for OA, Kappa, and D, respectively. Features of high-level layers are an abstraction of those of low-level layers and more discriminative for the classification task.
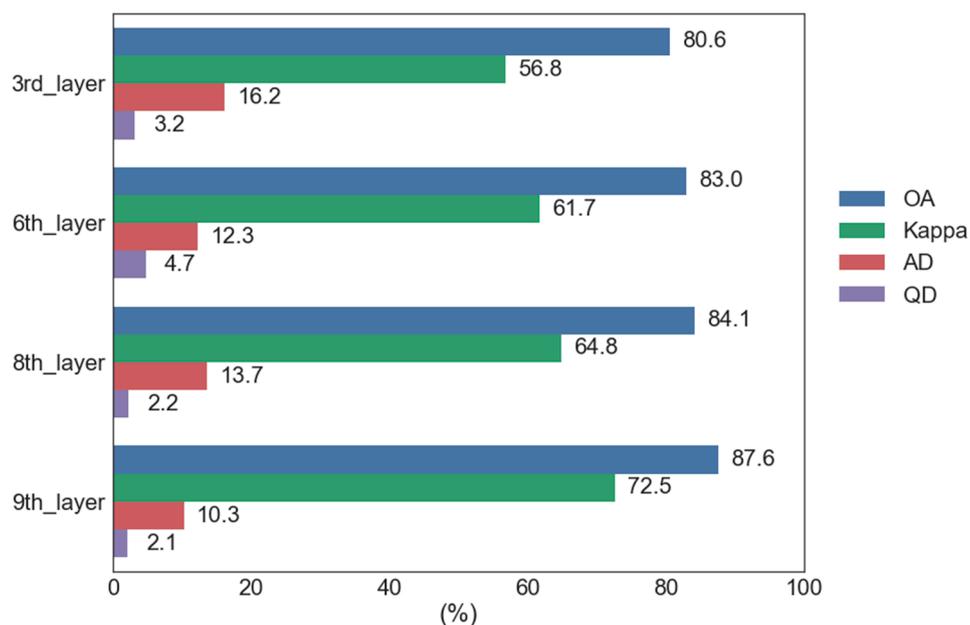

Figure 5. Classification performance of outputs from intermediate layers.

The output vector of the global average pooling layer was then used as the input for the random forest classifier and the results were shown in Table 2. The achieved accuracy of OA is 87.6% with a Kappa of 72.5%. There are 52 collapsed buildings misclassified as noncollapsed ones and 37 noncollapsed ones that failed to be identified. The UA and PA measures were obtained more than 70% for noncollapsed and collapsed building, which indicated the success of this classification technique applied in this study for damaged building detection. More collapsed buildings, which affected the corresponding PA percentage, were misclassified compared to noncollapsed ones. The greater values of PA (91.8%) and UA (89.1%) stemmed from noncollapsed buildings, which demonstrated that the model performed better on identifying noncollapsed buildings than collapsed ones. Regarding the building structures, some steel or wooden frame buildings collapsed during/after the earthquake with no visible deformation or textural changes visible on their roofs, in which cases the collapsed buildings were hard to discriminate from overhead imagery. Two components of disagreement (QD and AD) were calculated. The sum of them is the total disagreement, which equals to 1 minus OA. The lower QD and AD are, the better the model performs. The QD and AD were 2.1% and 10.3%, respectively. Figure 6 shows the classification map achieved by CNN-RF.

Table 2. Accuracy achieved by CNN-RF.

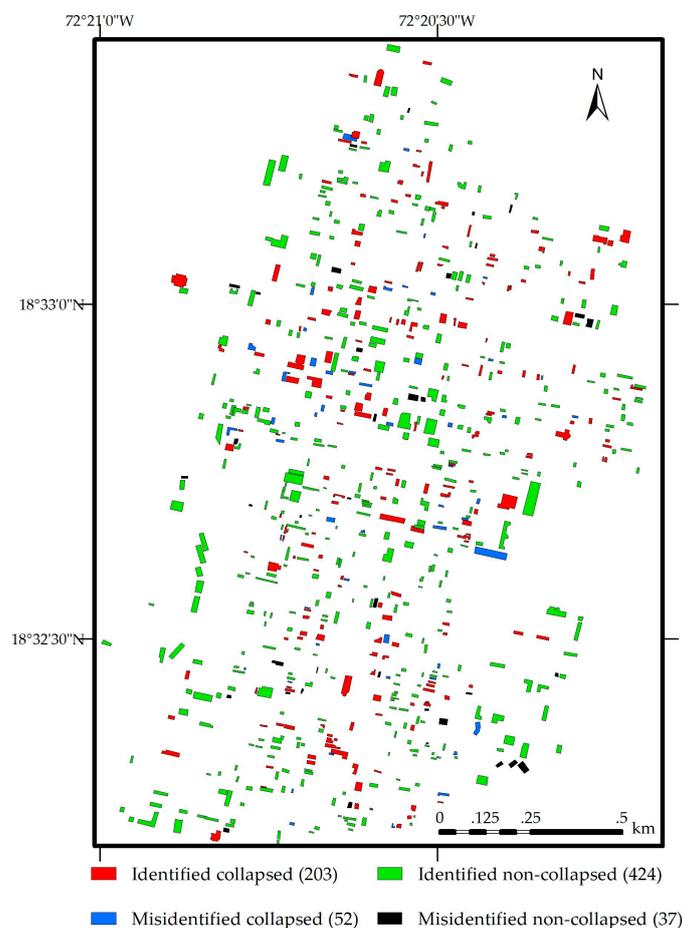|		Ground Truth						
		Collapsed	Noncollapsed	UA (%)	OA (%)	Kappa (%)	QD (%)	AD (%)
Predicted	**Collapsed**	203	37	84.6				
	Noncollapsed	52	424	89.1				
	PA (%)	79.6	91.8					
					87.6	72.5	2.1	10.3

Figure 6. Performance of CNN-RF on test dataset.

4.2. Performance of Texture-RF

Texture features (including contrast, dissimilarity, entropy, homogeneity, correlation, and ASM) of the pre- and post-event imagery based on GLCM were calculated. The violin plot synergistically combines the box plot and the density trace (or smoothed histogram) into a single display that reveals structure found within the data [57]. The distributions of derived pre- and post-event features based on the violin plot are shown in Figure 7. The violin plot shows quantiles for 0.25 (the first quantile) and 0.75 (the third quantile). It highlights data density and extends to the most extreme data points. The white dot indicates the median inside the plot. Edges of the vertical lines show the minimum and maximum values. The textural features (ASM and homogeneity) from pre-event imagery show greater distributions than those from post-event imagery. However, ASM and homogeneity from post-event imagery are more stable and concentrated than those from pre-event imagery.
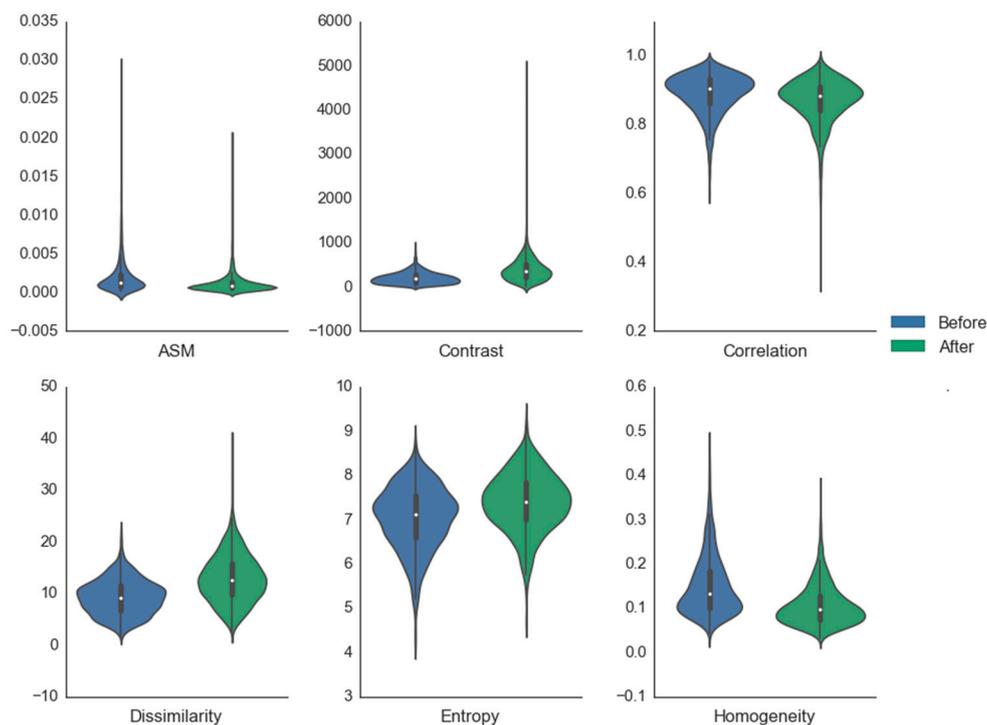

Figure 7. Violin plots showing the distribution of textural features obtained from pre- and post-event data.

Table 3 lists the error matrix and accuracy of collapsed and noncollapsed buildings using the Texture-RF method. The OA value of the result was 83.4% with a Kappa value of 61.7%. There were 29 for noncollapsed buildings and 90 for collapsed buildings were misclassified. It is possible to define noncollapsed buildings as the better-detected class by analyzing UA and PA for each class. The number of correctly identified collapsed buildings decreases due to the difficulty to identify changes of steel or wood frame buildings with steel roof. QD (8.5%) and AD (8.1%) are similarly important in the Texture-RF model. Figure 8 shows the classification map achieved by Texture-RF.

Table 3. Error matrix and accuracy achieved by Texture-RF.

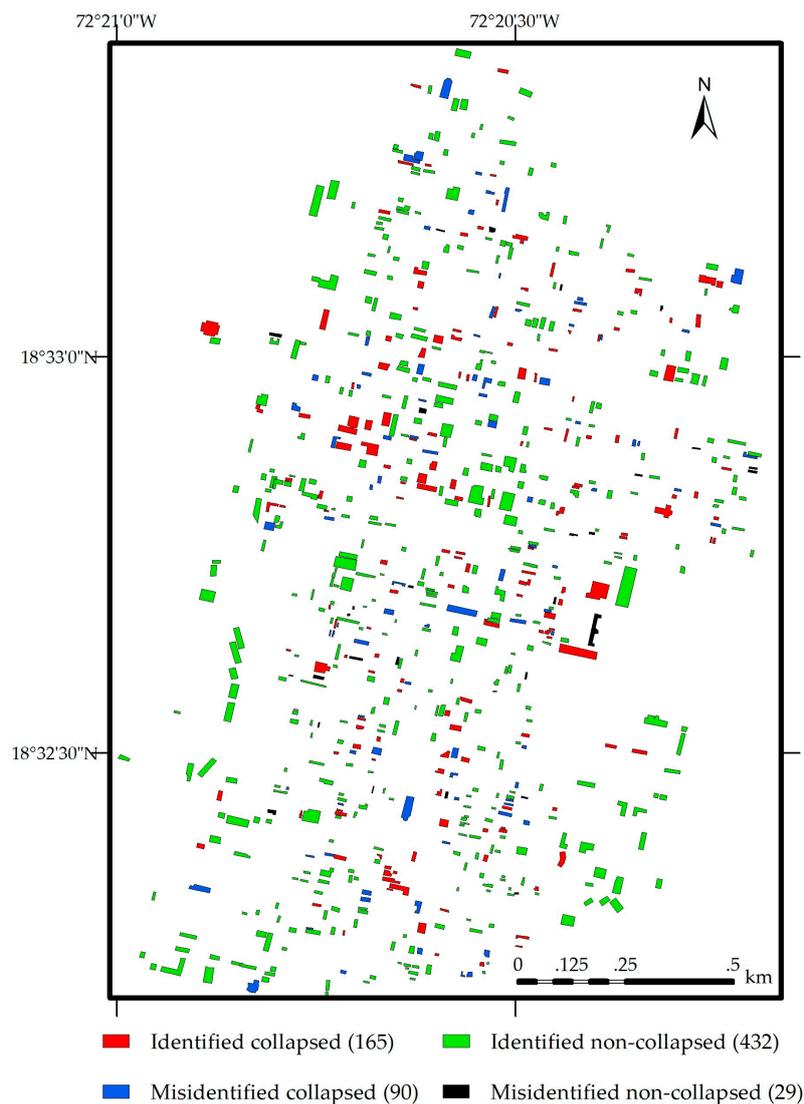| | | Ground Truth | | | | | | |
		Collapsed	Noncollapsed	UA (%)	OA (%)	Kappa (%)	QD (%)	AD (%)
Predicted	**Collapsed**	165	29	85.1				
	Noncollapsed	90	432	82.8				
	PA (%)	64.7	93.7					
					83.4	61.7	8.5	8.1

Figure 8. Performance of Texture-RF on test dataset.

5. Discussion

This study showed that random forest classifier with CNN features performed better than with GLCM texture features. To further explore the difference between these two kinds of features, Section 5.1 discussed the ability of separating the collapsed and noncollapsed buildings with different features via feature visualization. The intermediate CNN features for selected collapsed and noncollapsed buildings were also presented. Section 5.2 explored the relative important variables provided by the random forest classifier, and Section 5.3 further compared the performance of CNNs, CNN-RF, and Texture-RF for identifying collapsed buildings after the earthquake.

5.1. Feature Visualization

The t-distributed stochastic neighbour embedding (t-SNE) algorithm was proposed to visualize high-dimensional data by giving each data-point a location in a two- or three-dimensional map, which is similar to the principal component analysis by reducing data dimensionality [58]. To compare the difference between derived texture and CNN features, t-SNE was employed to visualize embedded texture and CNN features using the first three components (Figure 9). Raw data were also considered for the comparison. Red color represents collapsed buildings in the plots, and green color indicates noncollapsed ones.
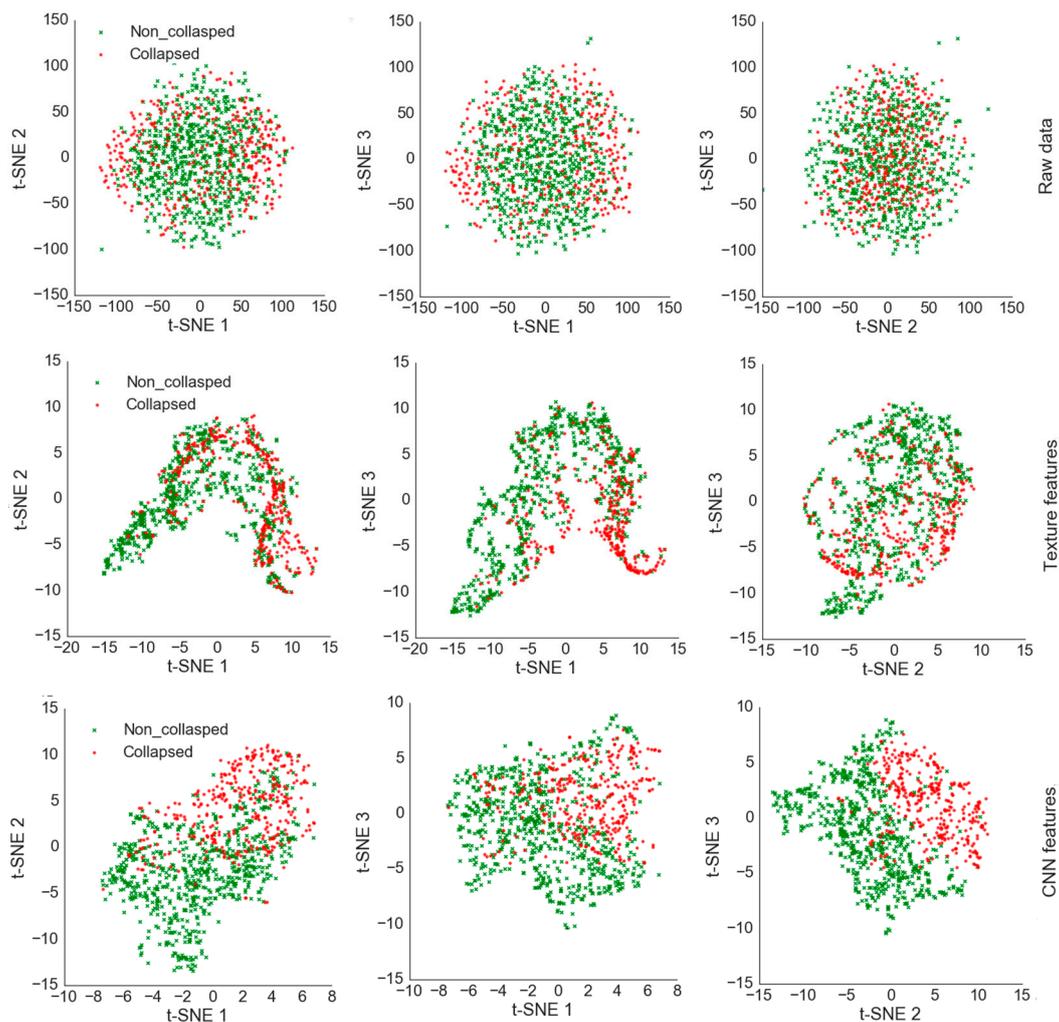
Figure 9. Visualization of the utilized dataset using t-distributed Stochastic Neighbor Embedding (t-SNE).

Basically, there are no visible clusters for t-SNE plots generated from raw data. All building objects are completely mixed up and it is hard to distinguish these two groups. Therefore, it is vital to extract informative features to identify collapsed buildings from satellite imagery. From the plots obtained by texture features, it can be seen that collapsed objects tend to cluster on the extreme left and right side of the first two plots. High levels regarding the features extracted from CNN contain more abstract information. T-SNE separates the different building objects and forms clustered groups of similar ones using CNN features. There is a slight improvement over texture features, especially for the plots using the second and third t-SNE components. Although texture and CNN features cn separate these two groups to some extent, there are still many buildings mixed up. It remains a challenge to distinguish collapsed and noncollapsed objects after earthquakes using satellite imagery.

Once the CNN model is established, it is also possible to visualize outputs of convolutional neural network layers, which provides a better understanding of what happened to the input building object after each operation. Figure 10 shows outputs of intermediate layers for selected collapsed and noncollapsed building objects. The third layer is the first pooling layer with 32 filters, which was used to reduce the dimensions of the image and keep the important features for further processing. There are 64 filters in the sixth layer (the second pooling layer) and eighth layer (the activation layer). The third layer acts as a collection of various change detectors by accepting the inputs of post-event data subtracted by pre-event data. It tends to retain the full shape of the building, although there are several filters that are not activated and left blank for the noncollapsed building objects. The obviously

changed pixels have higher values. As the layer goes deeper, the activations become increasingly abstract and less visually interpretable. The sparsity of the activations increases with the depth of the layer. In the third layer, most of the filters are activated by the input data. For the following layers, more and more filters are blank, which means the pattern encoded by the filter is not found in the input image. The features extracted by a layer become increasingly abstract with the depth of the layer. The activations of higher layers retain less information present in the specific input, and more information about the target.
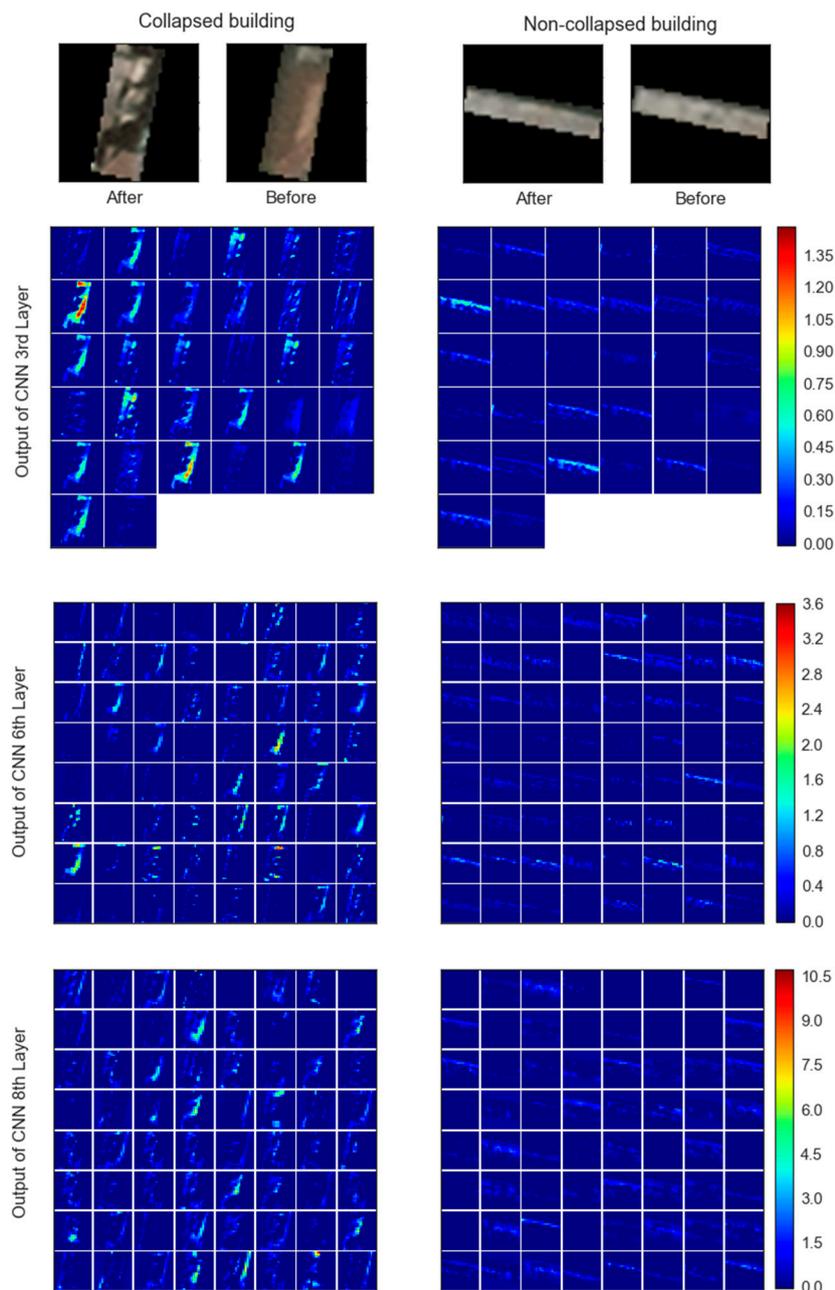

Figure 10. Visualization of outputs from intermediate layers.

5.2. Relative Importance Variables

A benefit of the random forest algorithm is that it can provide the estimation of variable importance from the trained model. For classification, the importance of input variable can be calculated based on mean decrease impurity, which is defined as the total decrease in node impurity (weighted by the

probability of reaching that node) averaged over all trees of the ensemble. For Texture-RF, Figure 11a demonstrates the relative importance of texture features generated from pre- and post-event imagery. It can be seen that the dissimilarity feature (except entropy) from the post-event data made the greatest contribution to the random forest model. Especially, four out of the top five variables are from post-event data. Features calculated from post-earthquake data are more important to distinguish collapsed and noncollapsed buildings. Looking at the variable, its importance can give you a sense of which variable has the highest effect on the model. It is possible to make use of such information to engineer new features or drop out features that look like noise. For CNN-RF, Figure 11b also displays the relative importance variables for neural nodes representing features extracted from the global average pooling layer. However, variables in the global average pooling layer are too abstract to be interpretable.
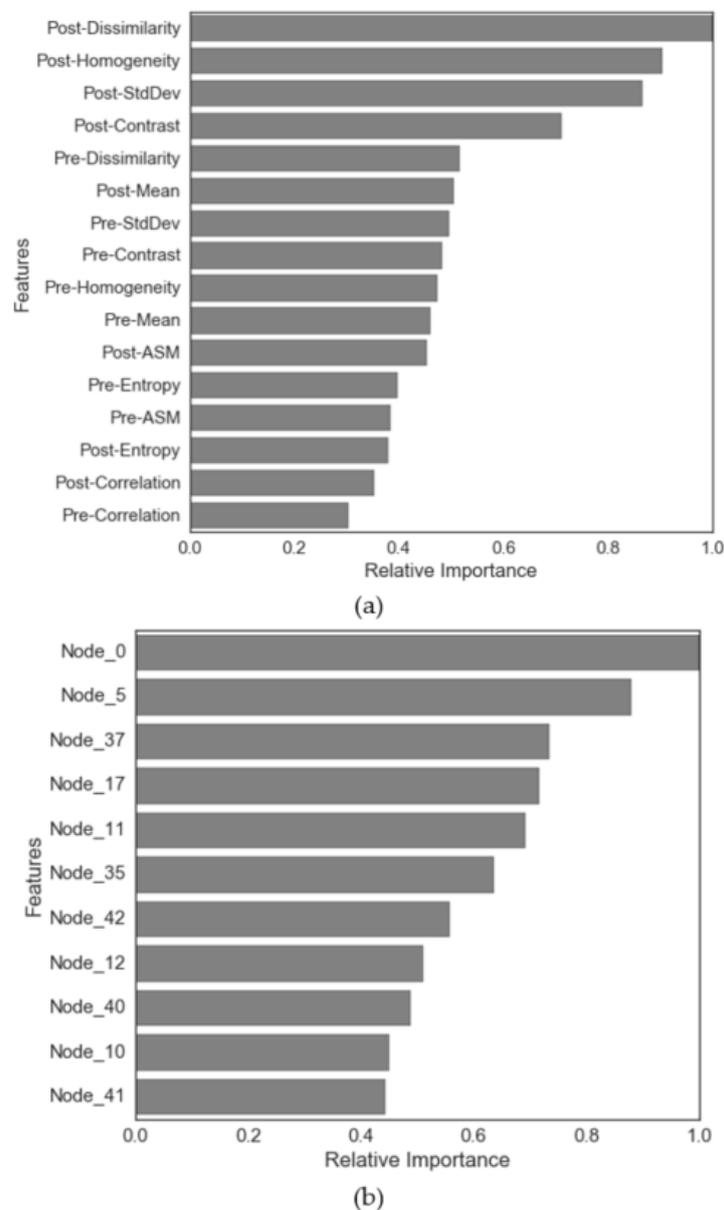

Figure 11. Relative importance variables derived from Texture-RF and CNN-RF. (**a**) shows the relative importance of texture features from pre- and post-event imagery. (**b**) shows the relative importance of features extracted by CNN.

5.3. Comparison between CNNs, CNN-RF, and Texture-RF

Considering the number of limited training data, we simply adopt three convolutional, activation, and pooling layers in the structure of CNN. According to the results, random forest classifier with features extracted from the sixth layer achieved similar performance compared to texture features with the sample Kappa value of 61.7% and OA values of 83.0% and 83.4%, respectively. The accuracy increases progressively along with the depth of the layers [26]. The number of convolutional filters for the first convolutional layer is 32, and the number for the second and third convolutional layers is 64. It is pointed out that larger number of filters can lead to an increase in performance. However, utilization of larger number of filters can increase training time and overfit the training data if the model is not regularized properly [29].

Figure 12 shows the comparison of the classification results achieved by CNNs, CNN-RF, and Texture-RF. Apart from Kappa and OA, in the present study, the metrics of allocation disagreement and quantity disagreement were also considered. The disagreement percentage for CNN and CNN-RF was mainly due to the component of AD (12.8% and 10.3% for CNN and CNN-RF, respectively) rather than QD (1.3% and 2.1% for CNN and CNN-RF, respectively). Although texture-RF achieved the lowest AD with 8.1%, its QD (8.5%) is the highest one which is also an important part to affect the classification results. A decrease of the D value that entails an increase of OA and Kappa. If the value of D is small, there is a high agreement between training data and test data. Although the lowest values of AD and QD were produced by Texture-RF and CNN respectively, the lowest D was obtained using CNN-RF. Thus, in the present study, when viewed from the perspective of evaluation metrics including D, OA and Kappa, CNN-RF outperformed CNNs and texture-RF.
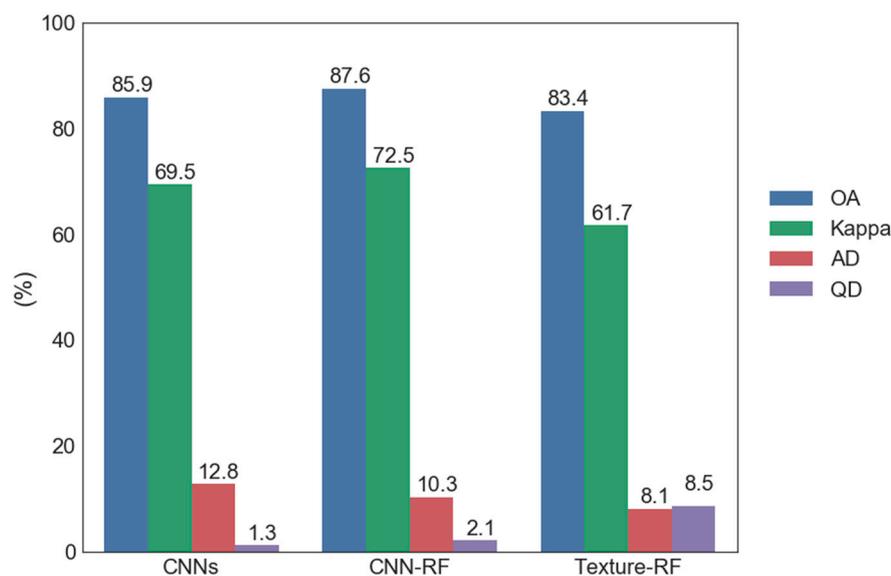

Figure 12. Performance comparison for CNNs, CNN-RF, and Texture-RF.

The result achieved by CNNs (OA = 85.9%, Kappa = 69.5%, and D = 14.1%) is slightly worse than CNN-RF. It demonstrated that CNN-RF is able to improve the accuracy by replacing the softmax function with random forest as the final classifier. Softmax is a generalization of the binary logistic function, and it is used as a cost function for probabilistic multi-class classification. As a classifier, the random forest classifier more complex than the softmax, and it is an excellent model for classification tasks. A simple decision tree is not very robust, so ensemble methods like random forest are proposed to run many decision trees and aggregate their outputs for prediction. This process controls overfitting and can often produce a very robust, high-performing model. Besides, the random forest classifier does not require much pre-processing, and can handle both categorical and numerical variables as the input.

Texture features play a key role in remote sensing image analysis. Texture-based descriptors characterize spectral variations that can provide supplementary information for high-resolution image analysis. In the present study, the result acquired by texture features with the random forest classifier was worse than CNNs and CNN-RF. Handcrafted features like texture features are designed based on expert knowledge about the problem, which only reflects limited aspects of the problem, while deep learning can automatically learn robust and representative features layer by layer. Deeply learnt features are generally more general and robust, and they were proven to be more effective for the identification of collapsed buildings after the 2010 Haiti earthquake.

6. Conclusions

Remote sensing imagery has been widely adopted to assess damaged buildings induced by an earthquake. In this study, we compared the performance of texture and CNN features with the random forest classifier to distinguish collapsed and noncollapsed buildings after the 2010 Haiti earthquake using pre- and post-event satellite imagery. The result directly obtained by a simple CNN model was also considered. Deep learning has proven its value for many problems, and is sometimes even able to surpass human ability to solve highly computational tasks, such as ImageNet Large Scale Visual Recognition Competition (ILSVRC) image classification and the highly mediatized Go match [59,60]. Motivated by these exciting advances, deep learning is becoming the model of choice in remote sensing, and has been successfully applied to land use and land cover classification, building detection, data fusion, and 3D reconstruction. Detailed summaries can be found in References [61,62]. It provides a promising approach for building damage assessment after the earthquake. The trained CNN model can be used as a feature descriptor and learnt features were combined with the random forest classifier. CNN-RF achieved the highest accuracy (OA = 87.6%, Kappa = 72.5%, and D = 12.4%). The learnt features from the CNN model showed better performance than texture features calculated from pre- and post-event satellite data. However, it still remains a challenge to identify collapsed buildings using remotely sensed data. Although buildings could be obscured by other features (e.g., trees, clouds) in the imagery, some collapsed buildings with metal sheet roofs showed basically no visible distortion in the overhead imagery. Thus, more data should be considered if available such as airborne oblique imagery revealing cracks on the building façade and LiDAR data providing failure geometrics of earthquake-affected buildings.

Author Contributions: All authors contributed in a substantial way to the manuscript. M.J. conceived, designed and performed the research and wrote the manuscript. L.L. and R.D. made contributions to the analysis of the data. All authors discussed the basic structure of the manuscript. M.F.B. reviewed the manuscript and supervised the study at all stages. All authors read and approved the submitted manuscript.

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

References

1.	Moya, L.; Yamazaki, F.; Liu, W.; Yamada, M. Detection of collapsed buildings from lidar data due to the 2016 Kumamoto earthquake in Japan. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 65. [CrossRef]

2.	Chen, S.W.; Wang, X.S.; Sato, M. Urban damage level mapping based on scattering mechanism investigation using fully polarimetric SAR data for the 3.11 east Japan earthquake. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6919–6929. [CrossRef]

3.	Zhai, W.; Shen, H.; Huang, C.; Pei, W. Building earthquake damage information extraction from a single post-earthquake PolSAR image. *Remote Sens.* **2016**, *8*, 171. [CrossRef]

4.	Shafique, M.; van der Meijde, M.; Khan, M.A. A review of the 2005 Kashmir earthquake-induced landslides; from a remote sensing prospective. *J. Asian Earth Sci.* **2016**, *118*, 68–80. [CrossRef]

5. Dong, Y.; Li, Q.; Dou, A.; Wang, X. Extracting damages caused by the 2008 Ms 8.0 Wenchuan earthquake from SAR remote sensing data. *J. Asian Earth Sci.* **2011**, *40*, 907–914. [CrossRef]

6. He, M.; Zhu, Q.; Du, Z.; Hu, H.; Ding, Y.; Chen, M. A 3D shape descriptor based on contour clusters for damaged roof detection using airborne LiDAR point clouds. *Remote Sens.* **2016**, *8*, 189. [CrossRef]

7. Chini, M.; Pierdicca, N.; Emery, W.J. Exploiting SAR and VHR optical images to quantify damage caused by the 2003 bam earthquake. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 145–152. [CrossRef]

8. Blaschke, T.; Strobl, J. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GeoBIT/GIS* **2001**, *6*, 12–17.

9. Bai, Y.; Adriano, B.; Mas, E.; Gokon, H.; Koshimura, S. Object-based building damage assessment methodology using only post event ALOS-2/PALSAR-2 dual polarimetric SAR intensity images. *J. Disaster Res.* **2017**, *12*, 259–271. [CrossRef]

10. Tu, J.; Sui, H.; Feng, W.; Song, Z. Automatic building damage detection method using high-resolution remote sensing images and 3D gis model. ISPRS Ann. Photogramm. *Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 43–50. [CrossRef]

11. Contreras, D.; Blaschke, T.; Tiede, D.; Jilge, M. Monitoring recovery after earthquakes through the integration of remote sensing, GIS, and ground observations: The case of L'Aquila (Italy). *Cartogr. Geogr. Inf. Sci.* **2016**, *43*, 115–133. [CrossRef]

12. Buchroithner, M.; Gevorkian, R.G.; Karachanian, A.S. Resultaty geologo-strukturnogo analisa aerokosmitscheskoj informatsii po sone spitakskogo semletrjasenija (Armenija)—Results of a structural-geological analysis of aero-cosmic data of the Spitak earthquake zone (Armenia). In Proceedings of the 10th Micro Symposium about Relativistic Planetology, Moscow, Russia, 7–11 August 1989; p. 2.

13. Gamba, P.; Casciati, F. GIS and image understanding for near-real-time earthquake damage assessment. Photogramm. *Eng. Remote Sens.* **1998**, *64*, 987–994.

14. Guirado, E.; Tabik, S.; Alcaraz-Segura, D.; Cabello, J.; Herrera, F. Deep-learning versus OBIA for scattered shrub detection with Google earth imagery: Ziziphus Lotus as case study. *Remote Sens.* **2017**, *9*, 1220. [CrossRef]

15. Janalipour, M.; Mohammadzadeh, A. Building damage detection using object-based image analysis and ANFIS from high-resolution image (case study: BAM earthquake, Iran). *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1937–1945. [CrossRef]

16. Gong, L.; Wang, C.; Wu, F.; Zhang, J.; Zhang, H.; Li, Q. Earthquake-induced building damage detection with post-event sub-meter VHR TerraSAR-X staring spotlight imagery. *Remote Sens.* **2016**, *8*, 887. [CrossRef]

17. Liu, T.; Abd-elrahman, A.; Morton, J.; Wilhelm, V.L.; Liu, T.; Abd-elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience Remote Sens.* **2018**, *55*, 243–264. [CrossRef]

18. Yu, H.; Cheng, G.; Ge, X. Earthquake-collapsed building extraction from LiDAR and aerophotograph based on OBIA. In Proceedings of the 2nd International Conference on Information Science and Engineering, Hangzhou, China, 4–6 December 2010; pp. 2034–2037.

19. Haralick, R.M. Statistical and structural approaches to texture. *Proc. IEEE* **1979**, *67*, 786–804. [CrossRef]

20. Pham, T.T.H.; Apparicio, P.; Gomez, C.; Weber, C.; Mathon, D. Towards a rapid automatic detection of building damage using remote sensing for disaster management: The 2010 Haiti earthquake. *Disaster Prev. Manag.* **2014**, *23*, 53–66. [CrossRef]

21. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

22. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

23. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 2015 International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.

24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, 1097–1105. [CrossRef]

25. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2014**, arXiv:1312.4400.

26. Hu, F.; Xia, G.; Hu, J.; Zhang, L.; Sensing, R. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]

27. Sainath, T.N.; Mohamed, A.R.; Kingsbury, B.; Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8614–8618.

28. Liu, L.; Ji, M.; Buchroithner, M. Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. *Sensors* **2018**, *18*, 3169. [CrossRef]

29. Sameen, M.I.; Pradhan, B.; Aziz, O.S. Classification of very high resolution aerial photos using spectral-spatial convolutional neural networks. *J. Sens.* **2018**, *2018*, 7195432. [CrossRef]

30. Ji, M.; Liu, L.; Buchroithner, M. Identifying collapsed buildings using post-earthquake satellite imagery and convolutional neural networks: A case study of the 2010 Haiti earthquake. *Remote Sens.* **2018**, *10*, 1689. [CrossRef]

31. Vetrivel, A.; Gerke, M.; Kerle, N.; Nex, F.; Vosselman, G. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 45–59. [CrossRef]

32. Ural, S.; Hussain, E.; Kim, K.; Fu, C.-S.; Shan, J. Building extraction and rubble mapping for city Port-au-Prince post-2010 earthquake with GeoEye-1 imagery and Lidar data. Photogramm. *Eng. Remote Sens.* **2011**, *77*, 1011–1023. [CrossRef]

33. Grünthal, G. *European Macroseismic Scale 1998*; Cahiers du Centre Europèen de Gèodynamique et de Seismologie, Conseil de l'Europe; Centre Europèen de Géodynamique et de Séismologie: Luxembourg, 1998.

34. UNITAR/UNOSAT; EC Joint Research Centre; The World Bank. Haiti Earthquake 2010: Remote Sensing Damage Assessment. 2010. Available online: http://www.unitar.org/unosat/haiti-earthquake-2010-remote-sensing-based-building-damage-assessment-data (accessed on 10 May 2017).

35. Fang, Z.; Li, W.; Du, Q. Using CNN-based high-level features for remote sensing scene classification. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016.

36. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time-series. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1998.

37. Bialas, J.; Oommen, T.; Rebbapragada, U.; Levin, E. Object-based classification of earthquake damage from high-resolution optical imagery using machine learning. *J. Appl. Remote Sens.* **2016**, *10*, 036025. [CrossRef]

38. Sun, W.; Shi, L.; Yang, J.; Li, P. Building collapse assessment in urban areas using texture information from postevent SAR data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3792–3808. [CrossRef]

39. Li, L.; Li, Z.; Zhang, R.; Ma, J.; Lei, L. Collapsed buildings extraction using morphological profiles and texture statistics—A case study in the 5.12 wenchuan earthquake. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010.

40. Rastiveis, H. Object-oriented analysis of satellite images using artificial neural networks for post-earthquake building change detection. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 139–144.

41. Geiss, C.; Taubenböck, H.; Tyagunov, S. Assessment of seismic building vulnerability from space. *Earthq. Spectra* **2014**, *30*, 1553–1583. [CrossRef]

42. Rastiveis, H.; Eslamizade, F.; Hosseini-Zirdoo, E. Building damage assessment after earthquake using post-event LiDAR data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 595–600. [CrossRef]

43. Park, J.; Harada, I.; Kwak, Y. A determination of the earthquake disaster area by object-based analysis using a single satellite image. *J. Remote Sens. Soc. Jpn.* **2018**, *38*, 14–29. [CrossRef]

44. Breiman, L.E.O. Randomf forest. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

45. Belgiu, M.; Dra, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

46. Ma, L.; Fan, S. CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests. *BMC Bioinform.* **2017**, *18*, 169. [CrossRef]

47. 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]

48. Mahmoud, A. *Plot-Based Land-Cover and Soil-Moisture Mapping Using X-/L-Band SAR Data*; Case Study Pirna-South: Saxony, Germany, 2011.

49. Geiß, C.; Aravena, P.; Marconcini, M.; Sengara, W.; Edwards, M. Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 175–188. [CrossRef]

50. Shi, L.; Sun, W.; Yang, J.; Li, P.; Lu, L. Building collapse assessment by the use of postearthquake Chinese VHR airborne SAR. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2021–2025. [CrossRef]

51. Anderson, J.R.; Hardy, E.E.; Roach, J.T.; Witmer, R.E. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*; US Government Printing Office: Washington, DC, USA, 1976.

52. Pringle, M.J.; Schmidt, M.; Muir, J.S. Geostatistical interpolation of SLC-off Landsat ETM+ images. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 654–664. [CrossRef]

53. Thomlinson, J.R.; Bolstad, P.V.; Cohen, W.B. Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products. *Remote Sens. Environ.* **1999**, *70*, 16–28. [CrossRef]

54. Pontius, R.G.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]

55. Ou, J.; Liu, X.; Li, X.; Chen, Y. Quantifying the relationship between urban forms and carbon emissions using panel data analysis. *Landsc. Ecol.* **2013**, *28*, 1889–1907. [CrossRef]

56. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. [CrossRef]

57. Nelson, R.D. Violin plots: A box plot-density trace synergism. *Am. Stat.* **2009**, *52*, 181–184. [CrossRef]

58. Van Der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

59. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. Mastering the game of go without human knowledge. *Nature* **2017**, *550*, 354–359. [CrossRef]

60. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 1026–1034.

61. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]

62. Ball, J.E.; Anderson, D.T.; Chan, C.S. A comprehensive survey of deep learning in remote sensing: Theories, tools and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]