Article

Discrimination of Earthquake-Induced Building Destruction from Space Using a Pretrained CNN Model

Min Ji [1,2], **Lanfa Liu** [1,2,*], **Rongchun Zhang** [3] **and Manfred F. Buchroithner** [2]

1   School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China; min.ji@tu-dresden.de
2   Institute of Cartography, Dresden University of Technology, 01069 Dresden, Germany;
    Manfred.Buchroithner@tu-dresden.de
3   School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications,
    Nanjing 210023, China; rongchunzhang@njupt.edu.cn
*   Correspondence: lanfa.liu@outlook.com

check for updates

Abstract: The building is an indispensable part of human life which provides a place for people to live, study, work, and engage in various cultural and social activities. People are exposed to earthquakes, and damaged buildings caused by earthquakes are one of the main threats. It is essential to retrieve the detailed information of affected buildings after earthquakes. Very high-resolution satellite imagery plays a key role in retrieving building damage information since it captures imagery quickly and effectively after the disaster. In this paper, the pretrained Visual Geometry Group (VGG)Net model was applied for identifying collapsed buildings induced by the 2010 Haiti earthquake using pre- and post-event remotely sensed space imagery, and the fine-tuned pretrained VGGNet model was compared with the VGGNet model trained from scratch. The effects of dataset augmentation and freezing different intermediate layers were also explored. The experimental results demonstrated that the fine-tuned VGGNet model outperformed the VGGNet model trained from scratch with increasing overall accuracy (OA) from 83.38% to 85.19% and Kappa from 60.69% to 67.14%. By taking advantage of dataset augmentation, OA and Kappa went up to 88.83% and 75.33% respectively, and the collapsed buildings were better recognized with a larger producer accuracy of 86.31%. The present study showed the potential of using the pretrained Convolutional Neural Network (CNN) model to identify collapsed buildings caused by earthquakes using very high-resolution satellite imagery.

Keywords: VGGNet; buildings; earthquake; dataset augmentation; pretrained CNNs

1. Introduction

Earthquakes have a consistently high frequency of occurrence and usually lead to secondary damage such as landslide and tsunami. The building is an essential part of modern human lives, which is vulnerable to earthquakes. Damaged buildings commonly lead to the main fatalities. Therefore, it is important to monitor structural health and rapidly assess building damage after earthquakes. Remote sensing can quickly and accurately capture the surface change of the earth, and the application of remote sensing for assessing damaged buildings has shown its values for post-event emergency response and reconstruction. For instance, near real-time satellite mapping played a vital role to support the government in emergency response of post-disaster relief after the 2015 Gorkha earthquake in Nepal [1].

With the development of remote sensing technologies, the amount of data obtained after an earthquake has been gradually increasing. Feature extraction is the key to make use of remotely sensed data to automatically identify damaged buildings after earthquakes. The common methods for

image features can be roughly divided into 2 categories (the hand-crafted and learned features) when dealing with images. Hand-crafted features are typically derived using statistical functions based on expert knowledge. The commonly applied hand-crafted features are local binary pattern (LBP) [2], scale-invariant feature transform (SIFT) [3], and histogram of oriented gradients (HOG) [4] features. Hand-crafted features have achieved remarkable performance on various tasks. For deep learning, features can be automatically learned by training a deep neural network model. Learnt features are more effective than hand-crafted features in image classification and object representation [5,6]. The learnt features from the Convolutional Neural Network (CNN) model showed better performance than six typical texture features calculated from pre- and post-event satellite data obtained before and after the 2010 Haiti earthquake [7].

In recent studies, CNNs have been applied in the field of identifying damaged buildings caused by the earthquake using remote sensing imagery and shown its potential in automatically discriminating damaged buildings [8]. CNNs were applied for collapse classification and spalling detection in concrete structures caused by earthquakes [9]. The results demonstrated that the proposed method achieved accurate and rapid identification of visual contents in a large volume of real-world imagery. A deep CNN was proposed to estimate the pre-event digital height model (DHM) from the single satellite image, and the DHM was trained by the post-event satellite imagery and light detection and ranging (LiDAR) data [10]. The collapsed buildings were successfully identified by analyzing the difference between pre- and post-event DHMs. The single-short multibit detector (SSD) [11] based on CNN pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) classification-localization (CLS-LOC) dataset was applied to detect building damage using extremely few training samples [12]. The experiment proved that the pre-training method can effectively increase various indicators of the model. Deep transfer learning (TL) based on Visual Geometry Group (VGG)Net was applied for image-based structure damage recognition, and the results revealed the potential to use deep TL in image-based structural damage recognition [13]. While CNN features are more effective compared to conventional hand-crafted features, it is also possible to further improve the performance in damage detection by combing 3-dimensional (3D) point cloud features. The integration of CNN and 3D point cloud features derived from very high-resolution (VHR) oblique aerial images significantly improved the model transferability accuracy compared to the accuracy achieved by CNN features alone [14].

While CNNs have great potential for the task of identifying the damaged buildings, there are some problems in training a deep CNN model from scratch. First, it requires a large amount of labeled data, while the number of training data is usually not enough to train the CNN model in the real world. Second, the training process is often time consuming if there are not extensive computational and memory resources. Pretrained CNNs have been proposed to function as generic feature extractors. In the training phase, there is a process, named as fine-tuning to take a network model that has already been trained for a given task, and make it perform on a new task. Fine-tuning a pretrained CNN is a promising alternative to training a CNN from scratch, especially when there is a limited dataset, and fine-tuned pretrained CNNs has an advantage of the speed of convergence. Pretrained CNNs have been successfully applied in imagery classification as a feature extractor or as a baseline for transfer learning. The pretrained CaffeNet [15] was fine-tuned to classify the remote sensing scenes solving the overfitting problem due to the lack of labeled images [16]. A pretrained model was applied to classify SAR targets with limited labeled data and achieved a superior performance than the methods based on CNNs [17]. Moreover, pretrained CNNs have been applied in disaster identification and achieved satisfactory performance. In order to achieve a good recognition performance toward small dataset tasks, instead of training deep CNN from scratch, deep transfer learning (TL) with a Visual Geometry Group (VGG) [18] pretrained model is implemented, and feature extractor and fine-tuning as two TL strategies. The results showed the potential use of deep TL in image-based structural damage recognition [13].

CNNs have been widely adopted in image classification, while there are still limited studies related to the detection of damaged buildings after an earthquake. VGGNet is a popular CNN structure. It achieved excellent performance on ImageNet Large Scale Visual Recognition Challenge (ILSVRC) classification and localization tasks, and it can generalize well to a wide range of tasks and datasets [18]. In remote sensing, it is difficult to obtain a large number of labeled samples. VGGNet could be adopted as the base model which was pretrained using a large amount of labeled data, and then a small amount of target data was applied to fine-tune the pretrained network for tasks including remotely sensed imagery classification and target detection [19,20]. In this study, a pretrained VGGNet model was proposed to recognize collapsed buildings caused by the 2010 Haiti earthquake using pre- and post-event remotely sensed overhead imagery, and the performance of fine-tuned pretrained VGGNet model and VGGNet model trained from scratch were compared. Dataset augmentation was also considered to enlarge the training dataset to improve identification accuracy. The study area and data sources were described in Section 2. The detailed methodology is illustrated in Section 3. The results achieved by the applied methods and discussion were shown in Section 4. Section 5 demonstrated the conclusion.

2. Study Area and Data Sources

The present study focused on the capital Port-au-Prince area, Haiti, which was severely damaged by the 2010 Haiti earthquake. More than 100,000 houses were heavily destroyed. One of the main reasons that led to more than 100,000 buildings being heavily destroyed is that these buildings had no or little consideration for seismic design. The post- and pre-earthquake satellite images were acquired after the 2010 Haiti earthquake through WorldView-2 on 9 January 2010 and QuickBird on 15 January 2010, respectively. The damaged level of buildings can be divided into five grades based on the European Macroseismic Scale 1988 (EMS-98) [21]. The building damage inventory was provided by UNITAR/UNOSAT [22] in which only four building damage levels existed (G 1, 3, 4, and 5). Figure 1 shows the comparison of pre- and post-event remote sensing imagery for each damage level. There is no difference of G1 buildings between pre- and post-event imagery. Even though G3 buildings have significant change on the roofs and its surroundings between pre- and post-event, the change is still difficult to be identified by remote sensing imagery. The debris around G4 can be identified, however, it is difficult to identify the difference of the G4 roofs. On the contrary, the roofs of G5 buildings are completely destroyed which indicated that it is feasible to identify the collapsed buildings from the images. Thus, it is difficult to find any edges and house boundaries using remote sensing imagery even with VHR 0.5 m. As mentioned above, heavy damage grades such as collapsed buildings are generally detected, and it remains challenges to identify low damage grades using the overhead imagery [23]. Thus, several low damage grades are usually aggregated as one grade. In this study, buildings with the damage grade from G1 to G4 were grouped as the non-collapsed category, and buildings with the damage grade of G5 were labeled as collapsed ones. There are 1789 buildings in the study area, and 610 buildings were labeled as collapsed ones. Figure 2 displays the distribution of collapsed and non-collapsed buildings in the study area.
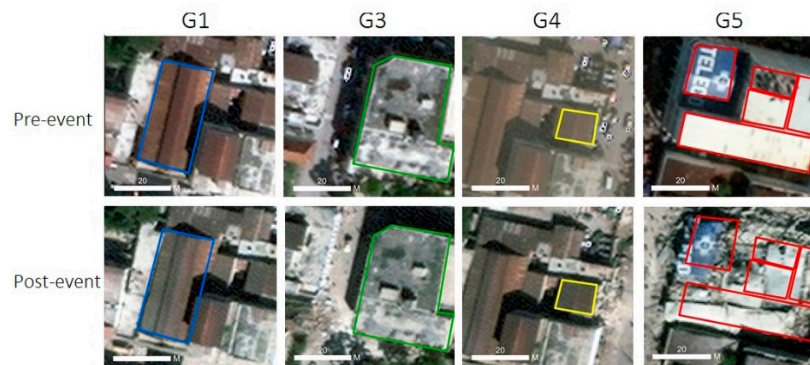
Figure 1. Comparison of pre- and post-event remote sensing imagery for each damage level. (G1: negligible to slight damage; G3: substantial to heavy damage; G4: very heavy damage; G5: destruction.) [24].

Figure 2. Distribution of collapsed and non-collapsed buildings in the study area. Red indicates collapsed buildings; green indicates non-collapsed buildings.

3. Methodology

In this section, the methods for detecting collapsed buildings are described in detail. Firstly, building patches were extracted from pre- and post-event satellite images with manually prepared building footprints. Then, the building patches with 96 × 96 pixels are functioned as input data to be applied to explore the effect of dataset augmentation and the number of fine-tuning layers for the performance of applied pretrained VGGNet. Section 3.1 describes the basic concept of CNNs and the adopted VGGNet structure. The concept of fine-tuning method and the detailed fine-tuning process is demonstrated in Section 3.2. Six incremental fine-tuning methods were applied to study the effect

of the number of fine-tuning layers for the performance of CNNs. Section 3.3 shows what dataset augmentation is and its characters. To assess the performance of applied methods, the evaluation metrics are described in Section 3.4.

3.1. CNNs

CNNs usually consists of several convolutional, pooling, and fully-connected layers, and it has shown its effective performance in analyzing images due to their remarkable performance on benchmark datasets such as ImageNet. CNNs have the ability to extract features and classify classes with a large number of parameters learned from the training dataset. VGGNet [18] can achieve good classification accuracy and it is relatively simpler compared to ResNet [25] and Inception [26]. Pretrained VGGNet model is easy to be fine-tuned for different classification tasks, and it can be used as the base model instead of training a model from scratch. The process of fine-tuning a pretrained CNN can be shown in Figure 3. In this study, the VGGNet-16 model was chosen as the basic architecture for detecting collapsed buildings using pre- and post-event VHR remote sensing imagery. The VGGNet-16 model contains 16 weight layers including 13 convolutional layers and 3 fully-connected layers, and it utilizes small-size convolutional filters (3×3) with a fixed stride of 1 for convolutional layers and 2×2 pixels with a stride of 2 for max-pooling from the beginning to the end. The dropout probability is set at 0.5 to reduce overfitting. Each convolutional block contains two or three convolutional layers and a pooling layer. There are five convolutional blocks, in which the first and second convolutional blocks have two convolutional layers, and the last three blocks have three convolutional layers.
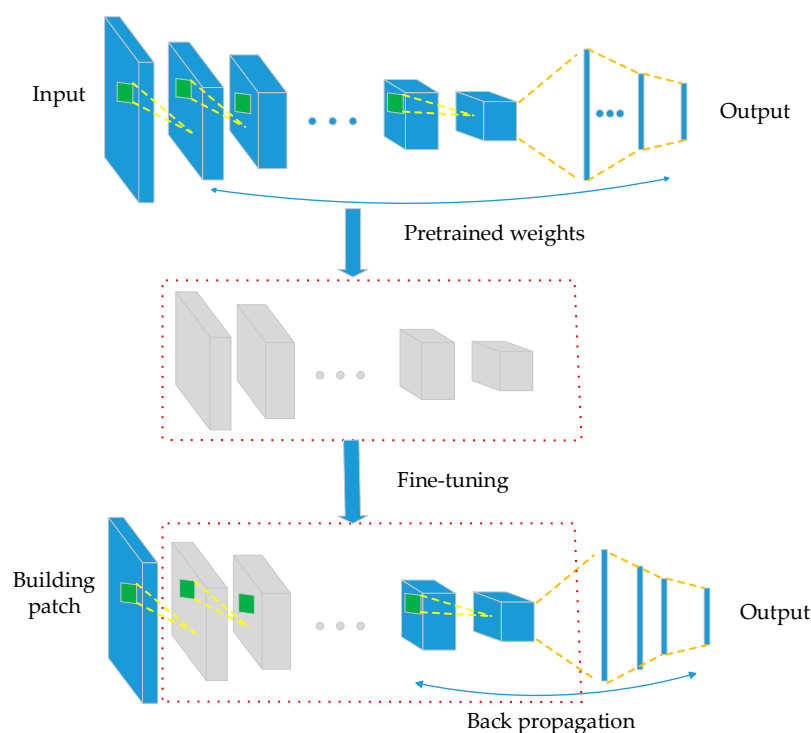

Figure 3. Illustration of fine-tuning pretrained Visual Geometry Group (VGG)Net for discriminating collapsed buildings after an earthquake.

3.2. Fine-Tuning with the Pretrained CNN Model

Figure 3 illustrates the process of fine-tuning a pretrained CNN. The VGGNet model was pretrained with the ImageNet dataset to learn features which were transferrable to further training on the target dataset. ImageNet dataset consists of 1.2 million images with over 1000 categories which could be divided into two big groups (animals and objects) and has been applied to several popular architectures, including VGGNet, ResNet, and Inception. The weights and biases were updated by

retrained different convolutional blocks using the target dataset, and the output layer was replaced with two neurons which respond to the binary classification of buildings. The comparison between incremental fine-tuning was applied to determine the number of fine-tuning layers with optimal performance. When fine-tuning part of layers, the rest of weights are frozen during training in this network, meaning that they only participate in the forward propagation of the network and will not be updated via backpropagation. The learning rate determines how much weights change at each iteration, and a higher learning rate comes with faster-changing weights. The network was trained with a small learning-rate, since our dataset is small and very different from the ImageNet. The network has been trained via 200 epochs with the learning-rate equal to 0.001, momentum 0.9.

3.3. Dataset Augmentation

CNN models are prone to overfitting when the training dataset is limited. To alleviate the problem, the easiest and most common method is to artificially enlarge the dataset. Dataset augmentation is an effective way to reduce the effect of overfitting for CNN, which generates more training data by making minor alterations (flip, translation, and rotation) to an existing dataset. Moreover, the architecture of CNNs enables extraction of scale, translation, and rotation tolerant features for classifying images or object categories [9]. The key concept of dataset augmentation is that the method applied to the labeled data does not change the semantic meaning of the labels while producing additional training data. When an additional dataset was trained in the network, the network can become invariant to those deformations and generalizes better for unseen data [27].

3.4. Evaluation Metrics

In this study, overall accuracy (OA), Kappa, producer accuracy (PA), and user accuracy (UA) were chosen as evaluation metrics. OA can be used as a standard that the proportion of non-collapsed and collapsed buildings were correctly identified. However, there is no thresholds of OA for image classification. In addition, it cannot tell the exact identified proportion of non-collapsed and collapsed buildings, respectively. Kappa has threshold provided by Landis and Koch [28], which values between 0.01 and 0.20 are slight, between 0.21 and 0.40 are fair, between 0.41 and 0.60 are moderate, between 0.61 and 0.80 are substantial, and between 0.81 and 1.00 are almost perfect. PA is the probability that collapsed or non-collapsed buildings are classified correctly. UA is the probability of predicting that it is indeed of this type in collapsed or non-collapsed buildings.

4. Experimental Results and Discussion

4.1. Pretrained VGGNet Model for Collapsed Building Detection

For the VGGNet structure, we slightly modified the layers after convolutional layers. The flatten layer was replaced by a global average pooling layer to reduce the total number of parameters in the model. Correspondingly, the number of neurons in the dense layer was reduced to 64. As the target is to classify collapsed and non-collapsed buildings, the softmax layer was replaced by a new layer with two neurons. The total number of network weights is 1.47 million. More details about VGGNet can be found in [18]. In this study, pretrained CNN model was fine-tuned using the target dataset and the model trained from scratch was also obtained using the same dataset. There are 1789 buildings utilized in the present study, including 716 and 1073 for testing and training, respectively. The results were shown in Table 1. It is obvious that the pretrained VGGNet model outperformed the VGGNet model with improvements of overall accuracy (OA) values from 83.38% to 85.19% and Kappa values from 60.69% to 67.14%. Even though VGGNet model achieved a higher value of OA, nearly half-collapsed buildings were misclassified as non-collapsed ones, which could be partly caused by that the target training dataset was too small for training VGGNet model since deep CNN requires a large amount of training dataset to obtain a satisfied classifier. On the contrary, the collapsed and non-collapsed buildings were relatively well identified by the pretrained VGGNet model, which indicated that

pretrained CNNs can perform satisfactorily under a limited dataset. The pretrained weights have been proved to outperform randomly initialized weights [17,29,30]. Neural network models trained from scratch were compared to pretrained models fine-tuned by target dataset in [31], and the results indicated that fine-tuned pretrained models achieved better performance, which corresponds to the results obtained in the present study.

Table 1. Performance of identification of collapsed buildings using Visual Geometry Group (VGG)Net model and fine-tuned VGGNet model.

Method	OA [1] (%)	Kappa (%)	Collapsed		Non-Collapsed	
			PA [2] (%)	UA [3] (%)	PA (%)	UA (%)
Pretrained VGGNet	85.19	67.14	81.70	75.29	86.90	90.67
VGGNet	83.38	60.69	59.21	90.96	96.75	81.09

[1] Overall accuracy; [2] Producer accuracy; [3] User accuracy.

4.2. Impact of the Dataset Augmentation

In general, the model that fine-tunes the entire network (such as updating all the weights) is prone to suffer the overfitting problem, especially among the first few layers when the new dataset is not large enough [32]. Dataset augmentation is a simple but effective technique to reduce the effect of overfitting and to improve performance. While pretrained CNNs can be applied to reduce the effect of the problem of a limited dataset, dataset augmentation still has the potential to improve classification performance, which was verified by Hu et al., 2015 [31]. It is common that the more training dataset CNNs have, the better results could be obtained. Agrawal et al., 2014 [33] pointed out that fine-tuning a network pretrained with the ImageNet dataset has a positive effect on a target task, such as image classification and object detection, and this effect increases when more data is used for fine-tuning. Thus, there is still potential to improve the classification accuracy by increasing the number of training dataset, and dataset augmentation can be applied to reduce the effect of overfitting problem in the fine-tuning procedure.

Figure 4 shows the loss curves and consumed time when dataset augmentation was considered. The VGGNet model was using pre-trained weights rather than random weights. The loss curve can be used for supervised learning procedure, if the loss is decreasing, it indicates that the network is learning effective features. Fine-tuned CNNs has an advantage of the speed of convergence. The loss decreased sharply along with the number of increasing epochs at the beginning for all dataset augmentation size. For the 1× and 2× dataset augmentation (× represents the multiple of the data increased by dataset augmentation), the loss values significantly increased after 60 epochs which implies that overfitting significantly affects the performance of the fine-tuned pretrained VGGNet model, and slight overfitting took place for 4× and 8× dataset augmentation. After 160 epochs, both loss curves for 16× and 32× became flat and vibrates in a small range, which indicated that it is hard to obtain more improvement with further training based on the current parameter complexity. Figure 4B shows that the training time increased with the increase of training data. It was noteworthy that the model with 32× obtained the lowest loss value, however, it needs about twice as much time to perform the model compared to the model with 16×. Time is precious for planning a rescue after an earthquake. The experiment demonstrated the 16× method achieved relatively fast training and high accuracy. Therefore, 16× was chosen as the dataset augmentation coefficient.
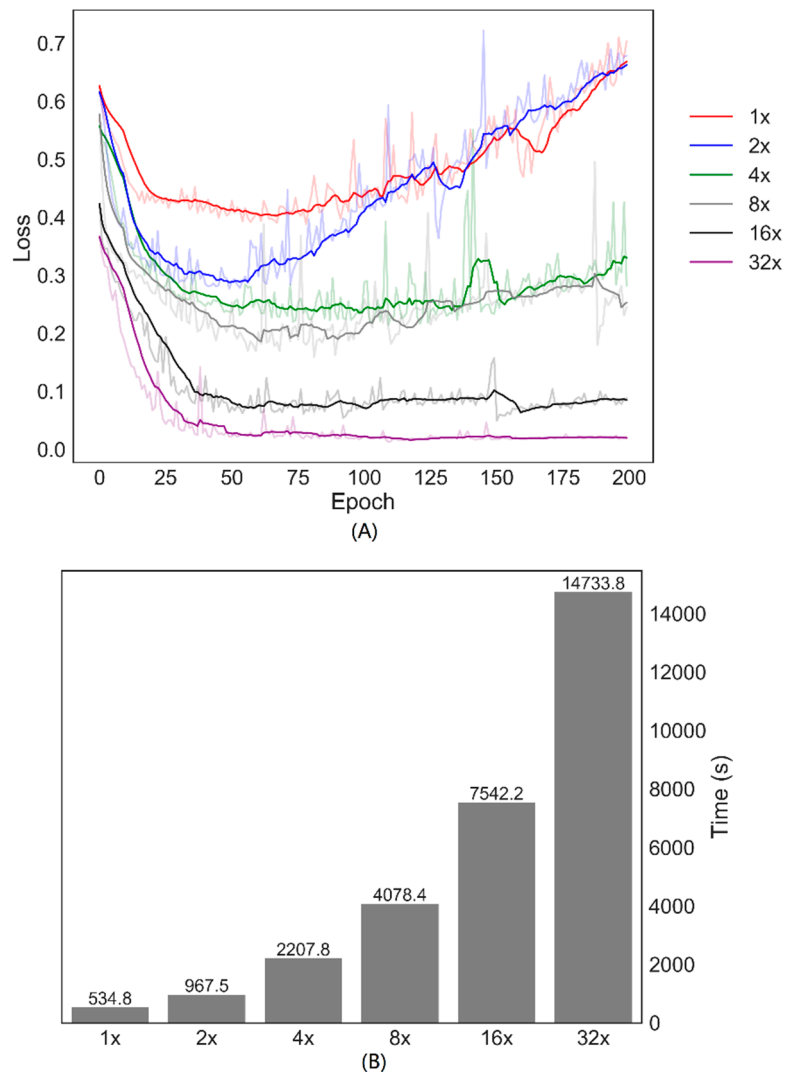
Figure 4. Performance of dataset augmentation. (**A**) Loss curves. (**B**) Time consumption for training the model.

4.3. Effect of Fine-Tuning Different Layers for the Detecting Collapsed Buildings

The comparison between incremental fine-tuning was applied to investigate the impact of the depth of fine-tuning on the performance of pretrained VGGNet model with dataset enlarged by 16 times (16×). The results were shown in Figure 5. It is obvious that the performance of pretrained VGGNet model increases with the increased of fine-tuned convolutional blocks. The worst result was obtained by only fine-tuning the fully-connected layer. One of the reasons is that the dataset for pre-training and the dataset for fine-tuning are significantly distant. The pretrained VGGNet model with fine-tuning each layer performed best with 88.83% of OA, 75.33% of Kappa, and 11.18% of total disagreement values. The results also indicate that, the more layers that are used for backpropagation, the better the learning process could be obtained.
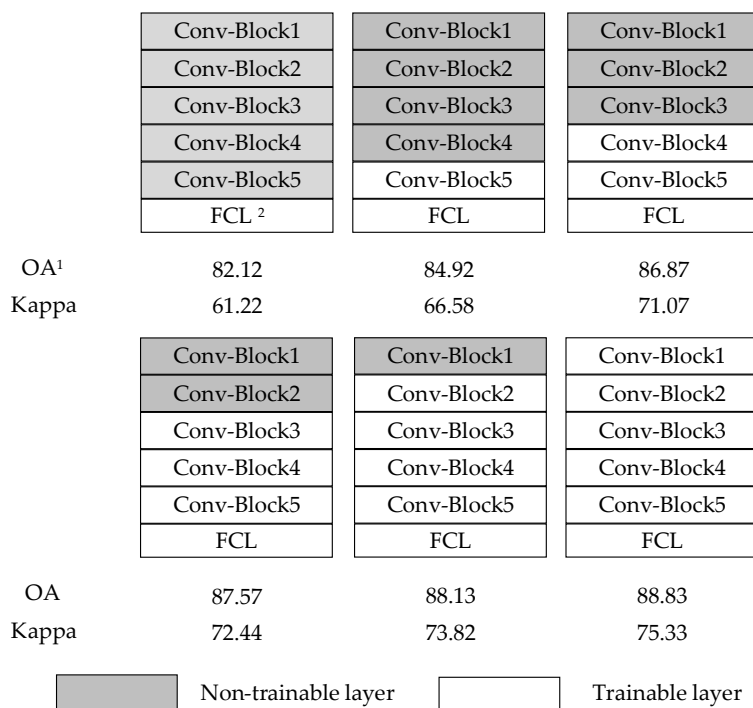
Conv-Block1	Conv-Block1	Conv-Block1
Conv-Block2	Conv-Block2	Conv-Block2
Conv-Block3	Conv-Block3	Conv-Block3
Conv-Block4	Conv-Block4	Conv-Block4
Conv-Block5	Conv-Block5	Conv-Block5
FCL [2]	FCL	FCL

OA[1]	82.12	84.92	86.87
Kappa	61.22	66.58	71.07

Conv-Block1	Conv-Block1	Conv-Block1
Conv-Block2	Conv-Block2	Conv-Block2
Conv-Block3	Conv-Block3	Conv-Block3
Conv-Block4	Conv-Block4	Conv-Block4
Conv-Block5	Conv-Block5	Conv-Block5
FCL	FCL	FCL

OA	87.57	88.13	88.83
Kappa	72.44	73.82	75.33

Non-trainable layer   Trainable layer

Figure 5. Fine-tuning of different layers. [1] Overall accuracy; [2] Fully-connected layer.

4.4. Performance of Pretrained VGGNet for the Detection of Collapsed Buildings

The classification results of collapsed buildings by means of the pretrained VGGNet model with fine-tuning all layers were shown in Table 2 and Figure 6. The overall accuracy is 88.83%, which is better than the results achieved in [34]. It can be seen that the applied model performed satisfactorily for identifying collapsed and non-collapsed buildings with relatively high producer accuracy (PA) and user accuracy (UA) values, which could be utilized as a guidance for the monitoring of disaster conditions and the responding emergency rescue. The non-collapsed buildings were better classified than collapsed buildings, which is partly caused by the effect of imbalanced data in which there were more non-collapsed buildings than collapsed ones. In addition, there were many steel or wooden frame buildings with metal sheet roofs, which had no visible deformation or textural change of the metal sheet roofs on the overhead satellite imagery when the buildings had been collapsed.

Table 2. Results of fine-tuned pretrained VGGNet model with dataset augmentation.

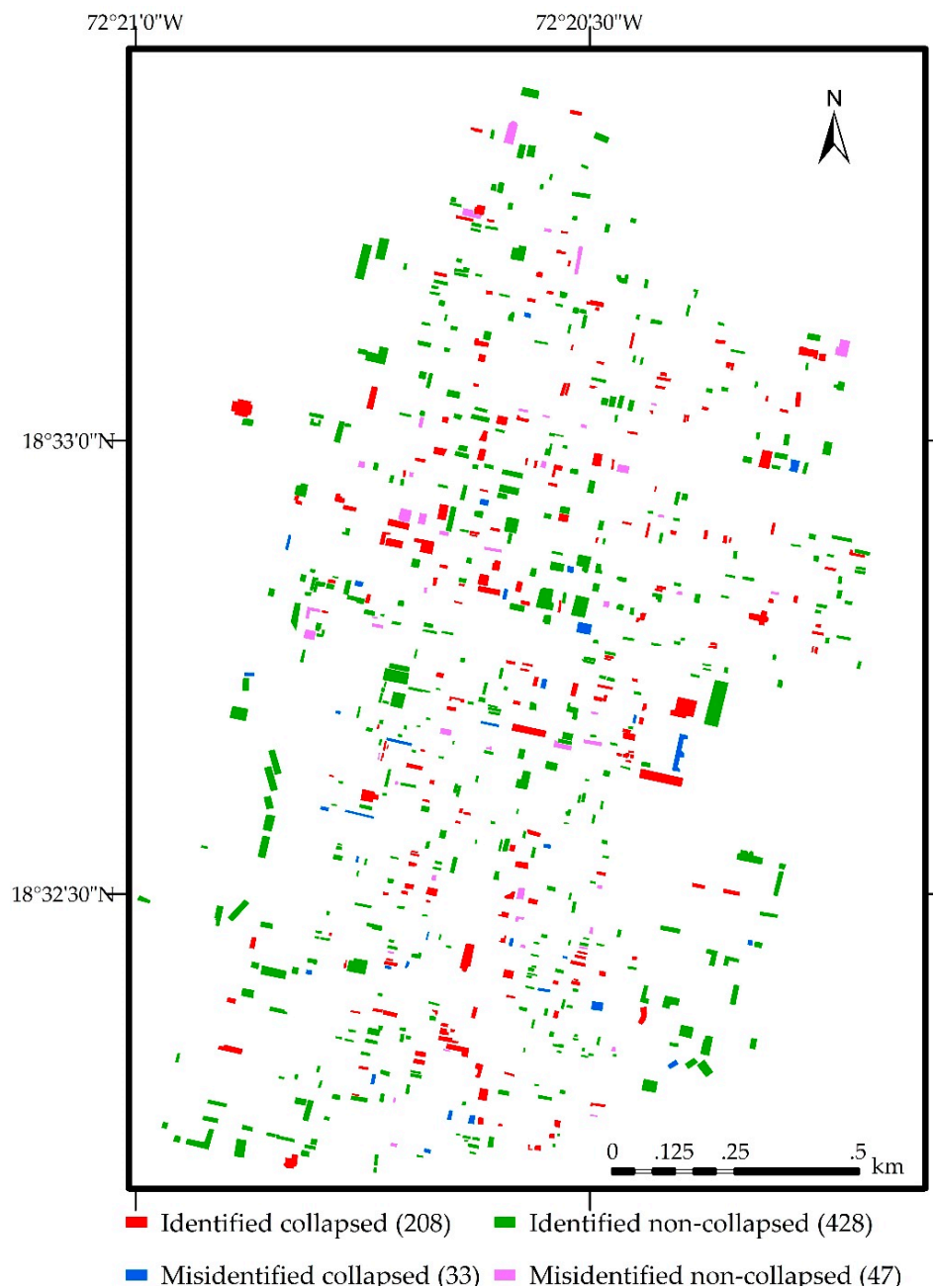| Confusion Matrix	Ground Truth			UA (%)
	Building Damage Grade	Collapsed	Non-Collapsed	
Predicted	Collapsed	208	47	81.57
	Non-collapsed	33	428	92.84
	PA (%)	86.31	90.11	

Figure 6.   Performance of pretrained VGGNet model for discriminating collapsed and non-collapsed buildings.

5. Conclusions

Deeply pretrained CNNs have shown their potential in many fields. In this study, to explore the capability of pretrained CNN models for classifying collapse and non-collapse buildings induced by the 2010 Haiti earthquake, a pretrained VGGNet model was applied and a CNN model trained from scratch was also obtained for comparison. The results demonstrated that the pretrained VGGNet model performed better than the one learned from scratch to classify collapsed and non-collapsed buildings. Dataset augmentation was used to reduce the effect of the overfitting problem induced by the limited dataset. The appropriate dataset augmentation multiple was ensured by comparing the effect of dataset augmentation for validation loss curves and consumed time. As the available dataset was far more different from ImageNet, the pretrained VGGNet model fully fine-tuned layers achieved

the best results with an OA of 88.83% and a Kappa of 75.33%. These findings not only can promote the use of CNNs that have been fine-tuned but also emphasize that large training sets are important to effective training and fine-tuning of CNNs. However, it should be pointed out that the building footprints were manually prepared in the present study, which is a limitation for rapid building damage mapping. An automatic approach should be considered to extract building footprints from high-resolution imagery. It is also possible to directly use an existing building database if available.

The advantages of using CNNs to detect damaged buildings include feature learning without feature engineering, the ability to extract invariant features, often with high accuracy. CNN technology based on remote sensing images is still in the early stages of detecting damaged buildings since the image features are complex and limited training dataset after earthquakes. There are some other factors that affect the accuracy of classification. For instance, the buildings on remote sensing imagery might be obscured by other objects such as trees, and there might be no visible distortion of collapsed buildings with metal roofs. Therefore, some additional datasets could be applied to improve the classification accuracy if it is available, such as airborne oblique dataset which can reveal cracks of a building facade and LiDAR data which can acquire failure geometrics of earthquake-affected buildings. In addition, it is still interesting to explore how to combine CNNs with conventional methods to take full advantage of existing techniques in order to improve the quality of derived building-damage information from remotely sensed imagery.

Author Contributions: M.J. conceived, designed, and performed the research and wrote the manuscript. L.L. made contributions to the analysis of the data. M.J., L.L., R.Z., and M.F.B. discussed the basic structure of the manuscript. M.F.B. reviewed the manuscript and supervised the study at all stages. M.J., L.L., R.Z., and M.F.B. read and approved the submitted manuscript. All authors have read and agreed to the published version of the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

CNNs	Convolutional Neural Networks
DHM	Digital Height Model
EMS	European Macroseismic Scale
FCL	Fully-Connected Layer
OA	Overall Accuracy
PA	Producer Accuracy
UA	User Accuracy

References

1. Ge, L.; Ng, A.H.M.; Li, X.; Liu, Y.; Du, Z.; Liu, Q. Near real-time satellite mapping of the 2015 Gorkha earthquake, Nepal. *Ann. GIS* **2015**, *21*, 175–190. [CrossRef]
2. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [CrossRef]
3. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Kerkyra, Greece, 20–27 September 1999; Volume 99, pp. 1150–1157.
4. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005.

5.   Zhou, W.; Newsam, S.; Li, C.; Shao, Z. Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval. *Remote Sens.* **2017**, *9*, 489. [CrossRef]

6.   Antipov, G.; Berrani, S.A.; Ruchaud, N.; Dugelay, J.L. Learned vs. hand-crafted features for pedestrian gender recognition. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; ACM: New York, NY, USA, 2015; pp. 1263–1266.

7.   Ji, M.; Liu, L.; Du, R.; Buchroithner, M.F. A comparative study of texture and convolutional neural network features for detecting collapsed buildings after earthquakes using pre- and post-event satellite imagery. *Remote Sens.* **2019**, *11*, 1202. [CrossRef]

8.   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]

9.   Yeum, C.M.; Dyke, S.J.; Ramirez, J. Visual data classification in post-event building reconnaissance. *Eng. Struct.* **2018**, *155*, 16–24. [CrossRef]

10.   Amirkolaee, H.A.; Arefi, H. CNN-based estimation of pre- and post-earthquake height models from single optical images for identification of collapsed buildings. *Remote Sens. Lett.* **2019**, *10*, 679–688. [CrossRef]

11.   Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Germany, 2016; pp. 21–37.

12.   Li, Y.; Hu, W.; Dong, H.; Zhang, X. Building damage detection from post-event aerial imagery using single shot multibox detector. *Appl. Sci.* **2019**, *9*, 1128. [CrossRef]

13.   Gao, Y.; Mosalam, K.M. Deep transfer learning for image-based structural damage recognition. *Comput. Aided Civ. Infrastruct. Eng.* **2018**, *33*, 748–768. [CrossRef]

14.   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]

15.   Qin, Z. How convolutional neural network see the world-A survey of convolutional neural network visualization methods. *arXiv* **2018**, arXiv:1804.11191. [CrossRef]

16.   Fang, Z.; Li, W.; Zou, J.; 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; pp. 2610–2613.

17.   Huang, Z.; Pan, Z.; Lei, B. Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. *Remote Sens.* **2017**, *9*, 907. [CrossRef]

18.   Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

19.   Chen, Z.; Zhang, T.; Ouyang, C. End-to-end airplane detection using transfer learning in remote sensing images. *Remote Sens.* **2018**, *10*, 139. [CrossRef]

20.   Liu, X.; Chi, M.; Zhang, Y.; Qin, Y. Classifying high resolution remote sensing images by fine-tuned VGG deep networks. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 7141–7144.

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

22.   United Nations Institute for Training and Research (UNITAR) Operational Satellite Applications Programme (UNOSAT). Available online: http://www.unitar.org/unosat/ (accessed on 10 May 2017).

23.   Ehrlich, D.; Guo, H.D.; Molch, K.; Ma, J.W.; Pesaresi, M. Identifying damage caused by the 2008 Wenchuan earthquake from VHR remote sensing data. *Int. J. Digit. Earth* **2009**, *2*, 309–326. [CrossRef]

24.   Miura, H.; Midorikawa, S.; Soh, H. Building damage detection of the 2010 Haiti earthquake based on texture analysis of high-resolution satellite images. In Proceedings of the 15th World Conference on Earthquake Engineering (15WCEE), Lisbon, Portugal, 24–28 September 2012; Volume 14, pp. 10703–10711.

25.   He, K.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

26.   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.

27. Salamon, J.; Bello, J.P. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [CrossRef]

28. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [CrossRef]

29. Erhan, D.; Bengio, Y.; Courville, A.; Manzagol, P.A.; Vincent, P.; Bengio, S. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* **2010**, *11*, 625–660.

30. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2014; pp. 3320–3328.

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

32. 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]

33. Agrawal, P.; Girshick, R.; Malik, J. Analyzing the performance of multilayer neural networks for object recognition. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Germany, 2014; pp. 329–344.

34. Cooner, A.; Shao, Y.; Campbell, J. Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake. *Remote Sens.* **2016**, *8*, 868. [CrossRef]