

Article

On the Generalization Ability of a Global Model for Rapid Building Mapping from Heterogeneous Satellite Images of Multiple Natural Disaster Scenarios

Yijiang Hu  and Hong Tang * 

State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; huyijiang@mail.bnu.edu.cn

* Correspondence: hongtang@bnu.edu.cn

Abstract: Post-classification comparison using pre- and post-event remote-sensing images is a common way to quickly assess the impacts of a natural disaster on buildings. Both the effectiveness and efficiency of post-classification comparison heavily depend on the classifier's precision and generalization abilities. In practice, practitioners used to train a novel image classifier for an unexpected disaster from scratch in order to evaluate building damage. Recently, it has become feasible to train a deep learning model to recognize buildings from very high-resolution images from all over the world. In this paper, we first evaluate the generalization ability of a global model trained on aerial images using post-disaster satellite images. Then, we systemically analyse three kinds of method to promote its generalization ability for post-disaster satellite images, i.e., fine-tune the model using very few training samples randomly selected from each disaster, transfer the style of postdisaster satellite images using the CycleGAN, and perform feature transformation using domain adversarial training. The xBD satellite images used in our experiment consist of 14 different events from six kinds of frequently occurring disaster types around the world, i.e., hurricanes, tornadoes, earthquakes, tsunamis, floods and wildfires. The experimental results show that the three methods can significantly promote the accuracy of the global model in terms of building mapping, and it is promising to conduct post-classification comparison using an existing global model coupled with an advanced transfer-learning method to quickly extract the damage information of buildings.

Keywords: damage assessment; global building mapping; transfer learning



Citation: Hu, Y.; Tang, H. On the Generalization Ability of a Global Model for Rapid Building Mapping from Heterogeneous Satellite Images of Multiple Natural Disaster Scenarios. *Remote Sens.* **2021**, *13*, 984. <https://doi.org/10.3390/rs13050984>

Academic Editor: Bahareh Kalantar

Received: 9 January 2021

Accepted: 2 March 2021

Published: 5 March 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The frequent occurrence of various kinds of disaster around the world has caused unprecedented heavy losses to human life and property. As one of the main disaster-bearing bodies, buildings might collapse and/or be damaged due to earthquakes, typhoons and other major natural disasters. Most human casualties from disasters occur in collapsed buildings. As an important indicator of the severity of a disaster, damage assessment of buildings plays an important role in disaster relief and government decision making. With increasing numbers of remote-sensing images and higher spatial resolutions, the issue of how to use remote-sensing technology for damage assessment has become the focus of researchers. A common method for collapsed building detection after disasters is to extract building damage information from high-resolution remote-sensing images.

Many methods have been designed for building-damage detection, and they can roughly be divided into two groups: (1) methods that detect changes between pre- and post-event data and (2) methods that only interpret postevent data [1]. These two types of method mainly differ in their applicability and the accuracy of results. Generally, the method using pre- and post-event data can obtain more accurate results, since more information is available to utilize. However, it is still very challenging to quickly obtain uniform images with dual time phases in many areas. The spectral difference in the pre- and

post-event images also affects the accuracy of the results to a great extent. The method using only post-disaster data is more suitable for rapid assessment. However, the identification accuracy is not satisfactory due to the absence of pre-disaster information. Change detection using pre- and post-event optical images is a common method used to obtain building damage information. This type of method includes image enhancement, post-classification comparison and deep-learning models. The image enhancement method calculates the differences in image colour, spectrum, texture and morphological features in pre- and post-event data [2,3]. Post-classification comparison is performed by comparing independent classification results of pre- and post-event images for change detection [4]. Deep-learning methods input both pre- and post-event images into a deep-learning model. The model actively learns the difference between the two types of image to extract building damage information [5,6]. At present, detailed ground texture information and context information can be presented in very high-resolution remote-sensing images. It is possible to detect building damage by using only post-event data. Many studies make full use of the features in post-event images, such as the spectrum, edge, texture, shape, shadows and spatial relationships, to detect building damage [7]. Due to the rapid development of deep learning, more building damage identification models based on post-event images have also been proposed [8,9].

At present, an increasing number of deep-learning models have been applied in building damage identification. Many deep-learning models have achieved high accuracy using either pre- or post-event data or only post-disaster data [6,8,9]. However, in actual disaster emergency assessments, these methods can only train a model after obtaining the images and manual annotations of the disaster area. Acquiring a large amount of manually labelled data and training the model during an emergency rescue after a disaster will waste considerable time. Therefore, it is more reasonable and effective to make full use of an existing damage-identification model and existing disaster case data to extract the damage information after the disaster occurs. For example, Valentijn et al. and Nex et al. [10,11] consider that training using real-time data is difficult for disaster emergency rescue. These studies evaluated the geographic transferability of deep-learning models to images taken in different locations, so as to simulate the situation of using the prepared model to identify the images acquired after a disaster. The results show that the models can obtain reasonable accuracy in very limited cases. Therefore, it is very important to adopt the transfer-learning method to promote the generalization ability of the prepared model. Some other scholars have also performed research in this area. For example, Li et al. [12] promoted the generalization ability of deep-learning models to identify buildings in different disaster images by using the unsupervised domain adaptation method.

Furthermore, the present disaster-damage-extraction methods, based on deep learning, are all devoted to the extraction of the characteristics of damaged buildings. Few studies have applied deep-learning models to the post-classification comparison framework. Therefore, combined with the post-classification comparison method, this paper seeks to extract building damage by promoting the generalization ability of the building identification model on post-event images. When the model can accurately identify undamaged buildings, damaged buildings can be extracted by the post-classification comparison method.

The post-classification comparison of pre- and post-event optical images is a common method used to obtain building damage information. This method, which makes good use of the existing building identification model, can eliminate the impact of radiation difference between pre- and post-event images. However, both the effectiveness and efficiency of post-classification comparison heavily depend on the classifier's precision and generalization ability. Therefore, the generalization ability of the existing building identification model is particularly important when using this method in disaster-damage assessments. In this paper, an existing global building mapping model, i.e., U-NASNetMobile model, is used to identify undamaged buildings in the images of the xBD dataset [13]. The xBD dataset is the largest building damage assessment dataset to date. The U-NASNetMobile model is a building identification model trained on a global building dataset [14]. However,

as shown in Figure 1, the model fails to identify undamaged buildings in the post-disaster images of the xBD dataset, while the identification results for pre-disaster images are very good. The potential reason for this might be that the global model might be misled by some invisible factors caused by imaging conditions, environmental changes due to disaster impacts and other factors. Even if there is no damaged building in the post-disaster images, the change in imaging conditions caused by the disaster still makes the boundary of the building more blurred, while the boundary of the building is clearer and sharper in the pre-disaster images. The blurring of key information results in poor performance of the model. Therefore, in order to extract the damage information of buildings through this post-classification comparison change detection method, it is necessary to promote the generalization ability of the existing building-identification model on the post-disaster data of the xBD dataset.



Figure 1. xBD images and labels of undamaged buildings in which both the ground-truth and predicted labels are indicated by red and green colours, respectively. (a) Labels of buildings in pre-disaster image. (b) Identification results for pre-disaster images. (c) Labels of undamaged buildings in post-disaster image. (d) Identification results for post-disaster images.

In order to promote the generalization ability of the existing global building mapping model on xBD post-disaster images, damaged buildings are identified through post-classification comparison. In this paper, we first evaluate the generalization ability of a global model trained on aerial images using post-disaster satellite images. Then, we systematically analyse three types of methods, i.e., fine-tune the model using very few training samples randomly selected from each disaster, transfer the style of post-disaster satellite images using CycleGAN [15], and perform feature transformation using domain adversarial training [16], to promote the model's generalization ability on post-disaster satellite images.

The remainder of this paper is organized as follows. Section 2 describes the experimental data, the global building mapping model used in this paper, and the evaluation of the model generalization ability. Section 3 provides a description of the three transfer learning methods used in this paper and the results of promoting the generalization ability of the model. We discuss the results of the experiments in Section 4. Finally, we draw some conclusions in Section 5.

2. Evaluation

2.1. xBD Dataset

In this big data era, many disaster damage assessment datasets have been produced in the past few years. However, most satellite image datasets tend to be limited to a single type of disaster and only contain postdisaster images of disaster areas. The experimental data used in this paper are part of the xBD dataset [13]. The xBD dataset is the United States Department of Defense's open-source satellite image dataset for natural disasters, and it is the largest building damage assessment dataset to date.

2.1.1. Images in the xBD Dataset

All images in the dataset are from the Maxar/DigitGlobe Open Data Program for both building damage assessment and postdisaster reconstruction. The dataset contains 22,068 optical satellite images with RGB bands (size $1024 \times 1024 \times 3$). The spatial resolution is below 0.8 m. The xBD dataset includes pre- and post-event satellite images of 19 disasters, including a diverse set of disasters around the world—earthquakes, tsunamis, floods, volcanic eruptions, hurricanes, etc. In view of some quality problems in the dataset, we selected 14 out of 19 disasters in this paper. The distribution of the disaster locations is shown in Figure 2. Most of these disasters come from the United States, including wildfires and wind disasters. For disasters in the United States, some are named after the location, such as Woolsey wildfire and Joplin tornado, while some wind disasters are named after the wind, such as Harvey hurricanes and Michael hurricanes. The specific information of disasters, covering six major kinds of natural disaster, i.e., wildfires (WF), tornadoes (TN), hurricanes (HC), floods (FD), tsunamis (TM) and earthquakes (EQ), is shown in Table 1. These are also frequent natural disasters around the world in recent years, that have caused serious damage to people's lives and property. The pre- and post-event images of some disasters are shown in Figure 3.

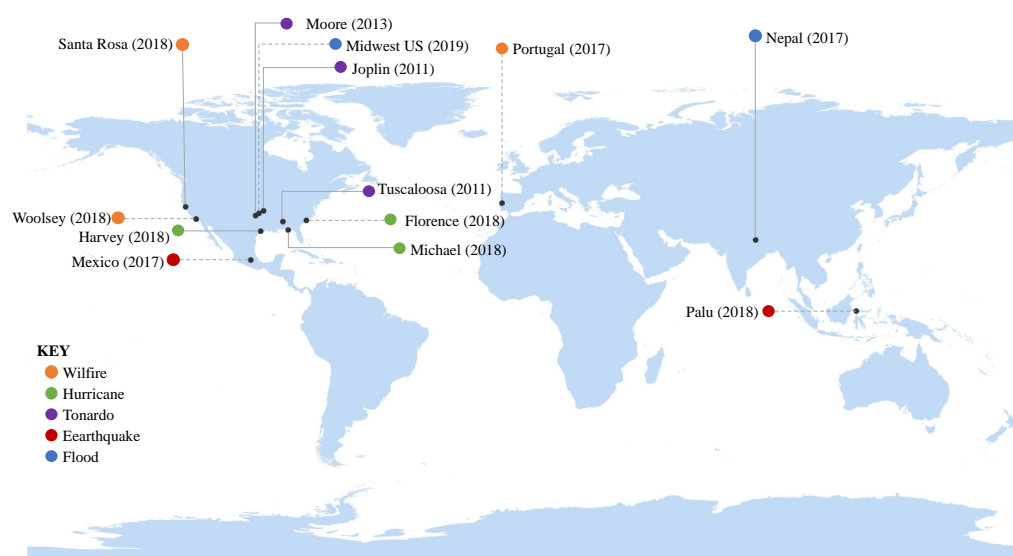
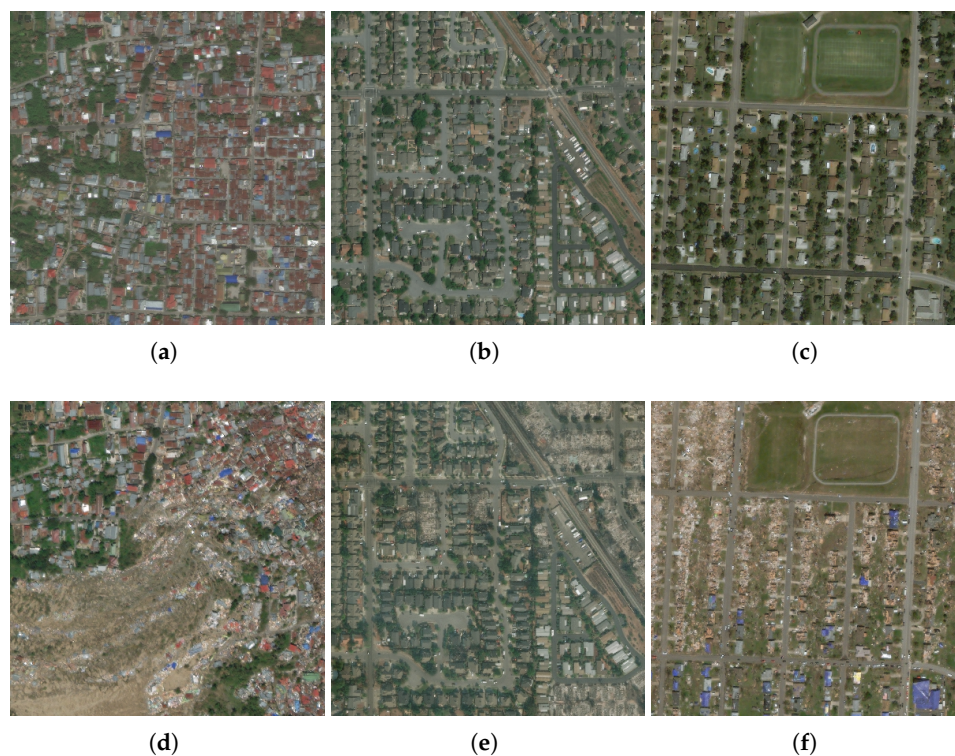


Figure 2. Distribution of the disaster locations.

Table 1. Details of the disasters included in this research.

Disaster Type	Disaster Location	Region	Event Dates
Tsunami(TM)	Palu	Asia	18 September 2018
Earthquake(EQ)	Mexico	America	19 September 2017
Flood (FD)	Nepal	Asia	July–September 2017
	Midwest of USA	America	3 January–31 May 2019
	Portugal	Europe	7–24 June 2017
Wildfire(WF)	Socal	America	23 July–30 August 2018
	Santarosa		8–31 October 2017
	Woolsey		9–28 November 2018
	Harvey		17 August–2 September 2017
Hurricane(HC)	Florence		10–19 September 2018
	Michael	America	7–16 October 2018
	Joplin		22 May 2011
Tornado(TD)	Tuscaloosa		27 April 2011
	Moore		20 May 2013

**Figure 3.** The images in the (a–c) subfigures represent the predisaster images from Palu-TM, Santarosa-WF, Joplin-TD, respectively. The images in the (d–f) subfigures represent the postdisaster images from Palu-TM, Santarosa-WF, Joplin-TD, respectively.

2.1.2. Damage Scales in xBD Dataset

In the process of annotating postdisaster images, Gupta et al. [13] proposed a joint damage scale, which was used to uniformly evaluate the degree of damage to buildings from the satellite images of different types of disaster. Table 2 shows the evaluation criteria for different levels of damage. Figure 4 shows the specific forms of buildings with different levels of damage.

Table 2. Joint damage scale descriptions on a four-level granularity scheme.

Disaster Level	Structure Description
0 (No Damage)	Undisturbed. No sign of water, structural or shingle damage, or burn marks.
1 (Minor Damage)	Building partially burnt, water surrounding structure, volcanic flow nearby, roof elements missing, or visible cracks.
2 (Major Damage)	Partial wall or roof collapse, encroaching volcanic flow, or surrounded by water/mud.
3 (Destroyed)	Scorched, completely collapsed, partially/completely covered with water/mud, or otherwise no longer present.

**Figure 4.** The different levels of damage. The images in the first, second, and third rows are from Moore-TD, Santarosa-WF and Harvey-HC, respectively.

There are 850,736 building annotations across 45,362 km² of imagery. In the predisaster images, two kinds of labels are provided, i.e., building or nonbuilding. The post-disaster images have labels with different degrees of damage, among which the polygon of the building was directly obtained from the predisaster image according to the projection of the geographical coordinates. There are four levels of building damage, i.e., no damage, minor damage, major damage and destroyed. The Figure 5 shows the post-disaster images and labels of some disasters.

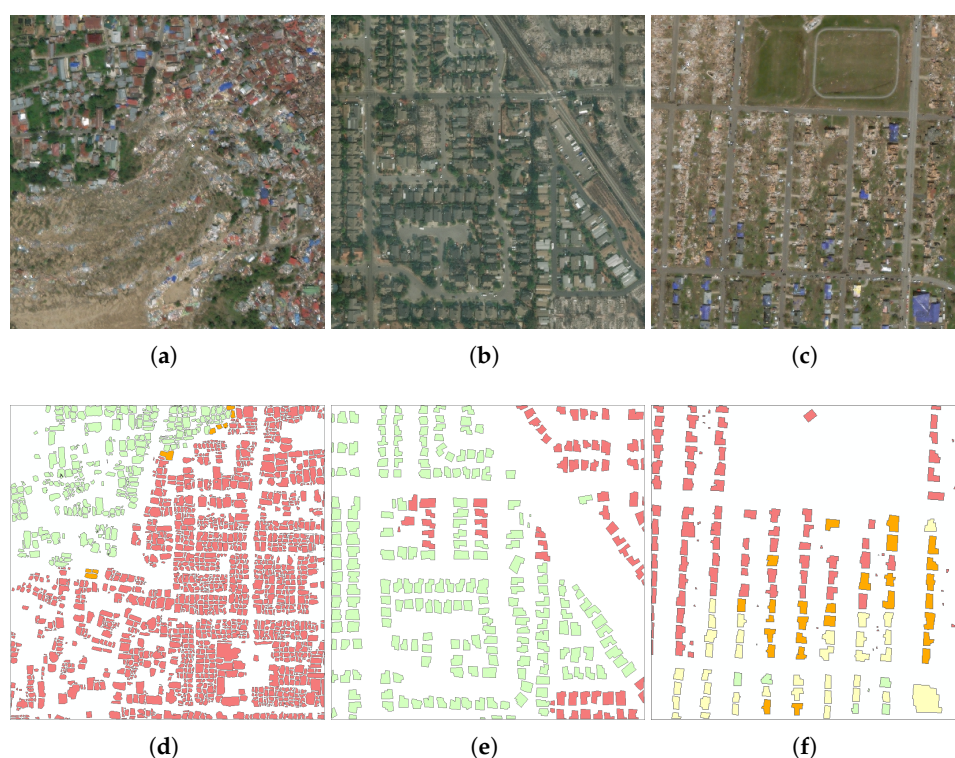


Figure 5. The images in the (a–c) subfigures represent the postdisaster images from Palu-TM, Santarosa-WF and Joplin-TD, respectively. The images in the (d–f) subfigures represent the labels of Palu-TM, Santarosa-WF and Joplin-TD, respectively. Green equals *no damage*, yellow *minor damage*, orange *major damage*, and red *destroyed*.

2.1.3. Quality of the xBD Dataset

Although the xBD dataset has made significant contributions to advance change detection and post-disaster damage assessment for humanitarian assistance and disaster recovery, some challenges remain. First, as shown in Figure 6, the occurrence of disasters is often accompanied by changes in the weather. Therefore, cloud cover cannot be ignored, especially in the images after a disaster. This is a major challenge to the existing building identification model. In this paper, the experiment manually screens out images with large-area cloud cover. Second, the post-disaster polygons are obtained directly from the pre-disaster images according to the projection of the geographical coordinates. Therefore, there are problems of dislocation between the polygons in the pre- and post-event images due to the difference in the times and angles of satellite imaging. Although we only evaluate and promote the generalization ability of U-NASNetMobile on post-disaster images, such annotation deviation still affects the accuracy assessment to a certain extent. Third, this paper aims to evaluate and promote the generalization ability of the building identification model on post-disaster images by using the deep-transfer-learning method. Therefore, it is necessary to consider the number of building samples of each disaster and the imbalance in the categories. In some images in the xBD dataset, there is a serious imbalance between building samples and background samples. For instance, more than 99% of the pixels in some images are labelled as background. Consequently, considering the challenges existing

in the xBD dataset, we selected partial data of 14 natural disasters from 19 natural disasters.

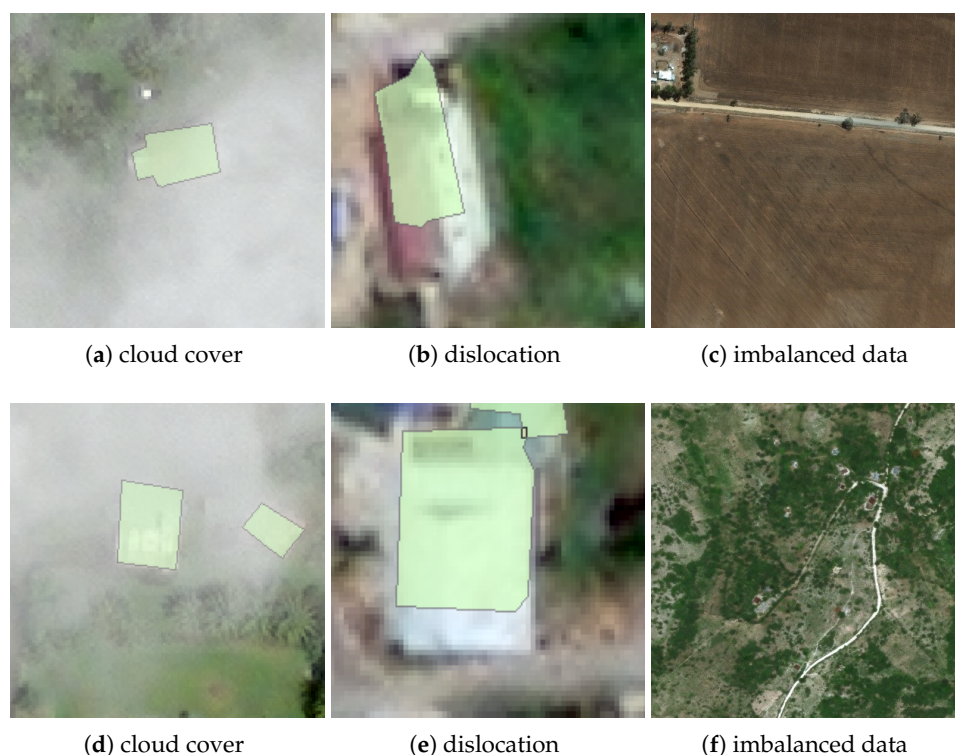


Figure 6. Examples of problems in the xBD dataset.

2.1.4. Differences between Disasters

When evaluating the generalization ability of a building identification model, the images of different disasters usually come from different regions, different times and different imaging conditions. There could be differences in building style or image quality. As shown in Figure 7, even though most of the buildings in the images are not damaged by disasters, there are still significant differences in the styles and sizes of buildings, and even the image quality, between images of different disasters. In general, the post-disaster images of Moore-TD and Mexico-EQ are relatively clear. The size of the buildings is relatively large, with obvious features. Especially in Moore-TD, most of the undamaged buildings after the disaster have a very regular form. The degree of separation between buildings is relatively high. There are few buildings that are closely connected to each other. In contrast, the buildings that remained intact after the Mexico EQ are more complex in terms of image features. There are some buildings closely which are connected to each other. The images of Nepal-FD are less clear, with smaller buildings, tighter interconnections and rougher roof textures. Such differences often affect the performance of existing building identification models in different disaster images to a large extent. Yang et al. [14] also confirmed that models trained on data from Europe have good results for data from Europe but do not perform as well using data from America. Therefore, this paper conducts experiments on 14 natural disasters to evaluate and promote the generalization ability of the U-NASNetMobile model on the images of each disaster.

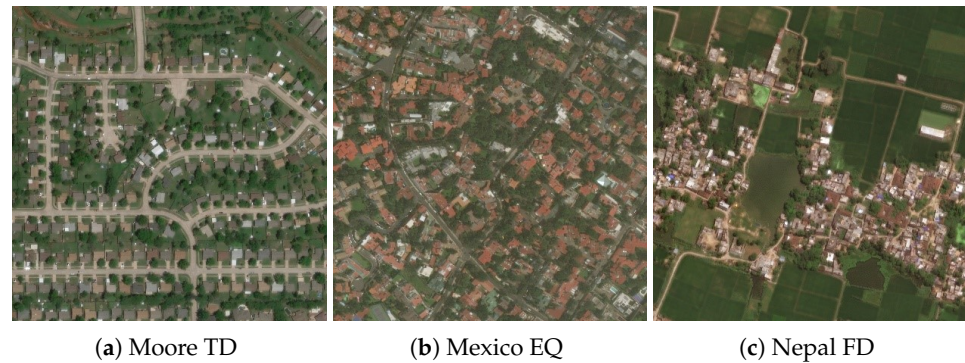


Figure 7. Post-disaster images of parts of the areas that were impacted by the Moore TD, the Mexico EQ and the Nepal FD.

2.2. The Global Model Trained on DREAM-B

In order to use very high-resolution (VHR) images for building mapping on a global scale, Yang et al. [14] constructed the so-called DREAM-B dataset, including the VHR image dataset of buildings from all over the world, and trained the U-NASNetMobile network using the DREAM-B dataset.

2.2.1. DREAM-B Dataset

Some commonly used datasets in image semantic segmentation are established on the basis of a few cities. This approach has difficulty meeting the needs of global building mapping. Yang et al. [14] used aerial images from more than 100 cities around the world to create the Disaster Mitigation and Emergency Management Construction Data Set (DREAM-B). DREAM-B contains 626, 4096×4096 image tiles, which are composed of three colour bands of red, green and blue, with a spatial resolution of 30 cm. The dataset contains two classes: the building and the non-building classes.

2.2.2. U-NASNetMobile

U-net [17] is a classical structure in the image segmentation field. It consists of a contraction path (encoder) and a symmetrical extension path (decoder) connected by a bottleneck. The encoder gradually reduces the spatial size of feature maps, which captures the context information and transmits it to the decoder. The decoder recovers the image details and spatial dimensions of the object through up-sampling and skip connections. Zoph et al. [18] proposed the NASNet-Mobile to learn the model architectures directly on the dataset of interest. In order to improve the computational efficiency, Yang et al. [14] combined the U-Net model with the NASNet-Mobile model, in which the neural cell obtained via neural architecture searching was used to replace the convolution modules in the U-Net. This model is called U-NASNetMobile in this paper. The architecture is shown in Figure 8.

Of the 626 image tiles in the DREAM-B dataset, 250 are used for training, 63 for validation and 313 for testing. The input size of U-NASNetMobile is 512×512 . The original image tiles of the DREAM-B dataset are divided into 51×12 pieces to match the model. Data enhancement, including random horizontal and vertical flips, random rotation, and random brightness jitter, is used during training to avoid overfitting. In the training process, the Adam optimizer [19] is used for optimization. In addition, the cosine attenuation learning rate [20] is used. The maximum and minimum learning rates are 3×10^{-4} and 1×10^{-6} , respectively. All the experiments are trained for 200 epochs with a minibatch size of 16. In addition, the Intersection over Union (IoU) was used to assess the accuracy of building areas [21]. The IoU is defined as

$$\text{IoU} = \frac{\text{Prediction} \cap \text{GroundTruth}}{\text{Prediction} \cup \text{GroundTruth}}. \quad (1)$$

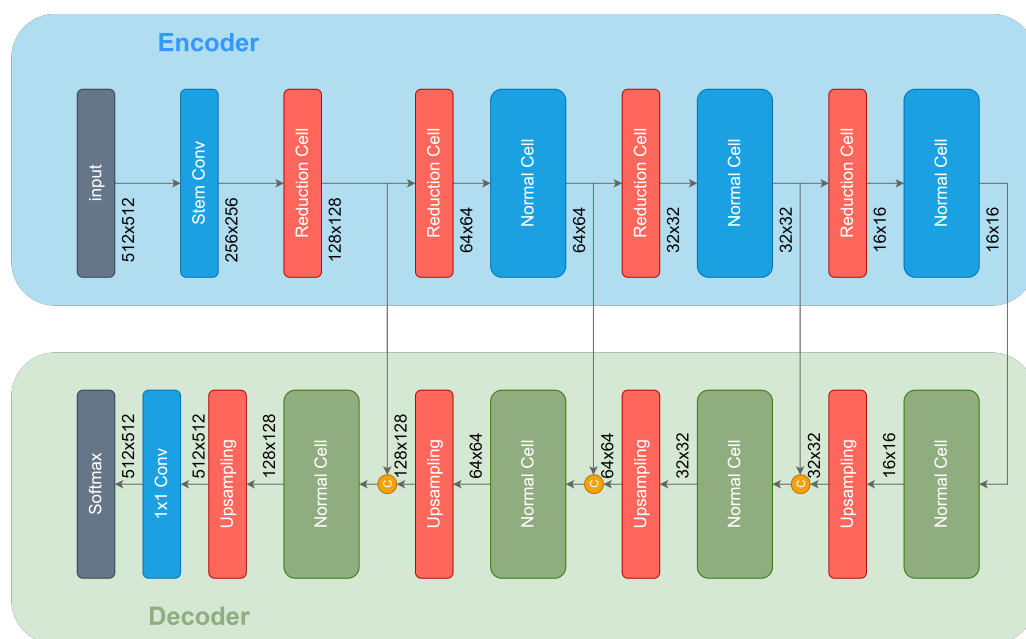


Figure 8. The architecture of U-NASNetMobile.

2.3. Evaluation

In this subsection, the trained model, i.e., U-NASNetMobile, is used to predict where buildings are in the post-disaster images and evaluate the generalization ability of this model. In view of the different damaged buildings in the post-disaster images, there are four kinds of modes used to evaluate the model performance. Mode 1 considers only the undamaged buildings in the post-disaster images to be recognizable buildings. Mode 2 considers only the undamaged and minor damage buildings in the post-disaster images to be recognizable buildings. Mode 3 considers only the undamaged, minor damage and major damage buildings in the post-disaster images to be recognizable buildings. Mode 4 considers buildings with any degree of damage to be recognizable buildings. We conducted experiments on different disasters with the four different evaluation modes. Taking the IoU of the buildings as an example, as shown in Figure 9, the identification accuracy of the model varies greatly from one disaster to another. Different evaluation modes in the same disaster have little influence on the evaluation results. Generally, when a building is marked as having major damage, its overall form will change significantly. If it is still considered to be a recognizable building, it will be quite different from the building form learned by the existing model, resulting in an unreliable accuracy evaluation. Therefore, this paper adopts Mode 2 to evaluate the identification results. It can also be seen from Figure 9 that Woolsey-WF, Tuscaloosa-TD, Moore-TD, Joplin-TD and Midwest-FD all showed the highest accuracy in the Mode 2 evaluation method. A disaster with relatively high accuracy indicates that the distribution of these images is closer to that of the training data. The image quality is better. Therefore, the degree of damage will have a more significant impact on its accuracy. Consequently, it is appropriate to consider the major damage and destroyed buildings as the background in the evaluation.

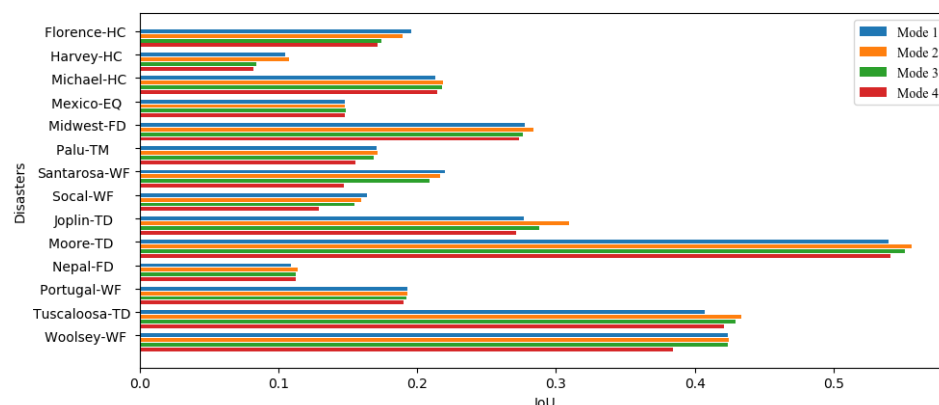


Figure 9. The IoU under different evaluation modes over 14 disaster events.

Table 3 shows the prediction results of all 14 disasters using the U-NASNetMobile model. In this paper, the recall, precision and IoU of the building class are used to evaluate the identification results. We also use the missed detection rate and false detection rate to evaluate the identification accuracy of a single building object. In the ground truth, we selected undamaged and minor damage buildings in the post-disaster images as the building class, while other types of damaged building are all included in the background class. Table 3 shows that the IoU is generally low. The recall is low, but the precision is relatively high. Regarding the detection accuracy of an individual building, the missed detection rate is low, but the false detection rate is relatively high. This result shows that the generalization ability of the existing U-NASNetMobile model on xBD postdisaster images is generally poor. This method has difficulty identifying the buildings from the images. However, most of the buildings identified by the model are correct. In addition, the generalization performance of the existing U-NASNetMobile model varies greatly among different disasters, with the highest IoU of 0.556 in Moore-TD and the lowest IoU of 0.107 in Harvey-HC.

Table 3. Accuracy evaluation of building identification results.

Disaster Name	Recall	Precision	IoU	Kappa	Missed Detection Rate	False Detection Rate
Florence-HC	0.210	0.689	0.189	0.283	70.54%	17.89%
Harvey-HC	0.121	0.587	0.107	0.149	80.49%	25.33%
Michael-HC	0.247	0.697	0.218	0.317	62.30%	14.81%
Mexico-EQ	0.160	0.729	0.148	0.176	66.52%	10.37%
Midwest-FD	0.307	0.726	0.284	0.393	60.87%	5.25%
Palu-TM	0.197	0.593	0.171	0.224	51.42%	11.23%
Santarosa-WF	0.259	0.522	0.216	0.286	37.04%	9.50%
Socal-WF	0.171	0.593	0.159	0.239	59.07%	6.97%
Joplin-TD	0.350	0.736	0.310	0.416	46.77%	10.97%
Moore-TD	0.625	0.852	0.556	0.666	28.15%	4.48%
Nepal-FD	0.125	0.646	0.114	0.179	78.27%	12.29%
Portugal-WF	0.197	0.904	0.193	0.292	72.46%	7.65%
Tuscaloosa-TD	0.499	0.779	0.434	0.568	39.10%	12.17%
Woolsey-WF	0.470	0.831	0.425	0.567	36.36%	9.10%

Figure 10 shows a typical phenomenon in the identification results of the U-NASNetMobile model. Some obvious undamaged buildings in the postdisaster images cannot be correctly identified by the model. In addition, Table 3 shows that buildings have low recall and high missed detection rates while achieving good precision and false detection rates. In the subsequent promotion, we can focus on the changes in the recall, missed detection rate

and IoU. In addition, there are certain differences between different kinds of disasters. The dataset used in this paper includes six different kinds of disasters, such as earthquakes, tsunamis, hurricanes, tornadoes, floods and wildfires. There were two or more cases for all disasters except earthquakes and tsunamis. These cases of disasters can be used to assess the impact of disaster types on the model's generalization ability. Figure 11 shows the average IoU of each disaster, indicating that the accuracies of tornadoes and wildfires are higher than those of both hurricanes and floods. The reason for this phenomenon may be that in the four disasters of Harvey-HC, Florence-HC, Nepal-FD and Midwest-FD, the damage to buildings is mainly caused by the inundation of water. Buildings with such characteristics have never been seen by models in the training process, which results in the low generalization ability of models in such disasters.

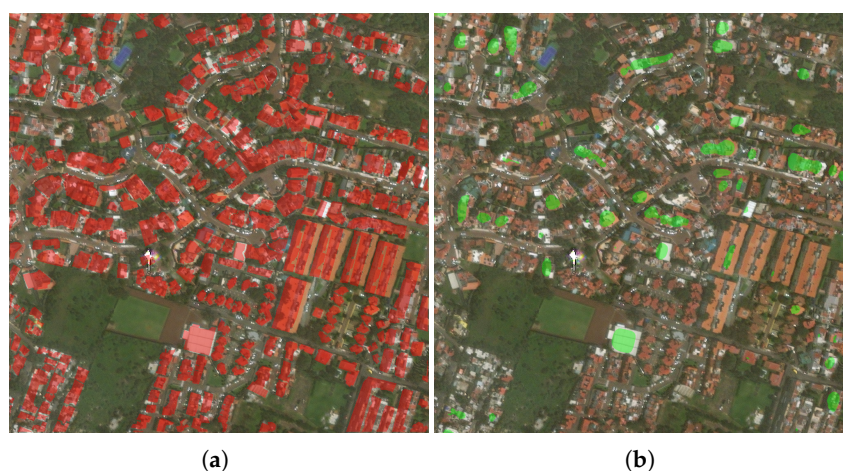


Figure 10. The images in the (a,b) subfigures represent the ground truth and the identification results, respectively.

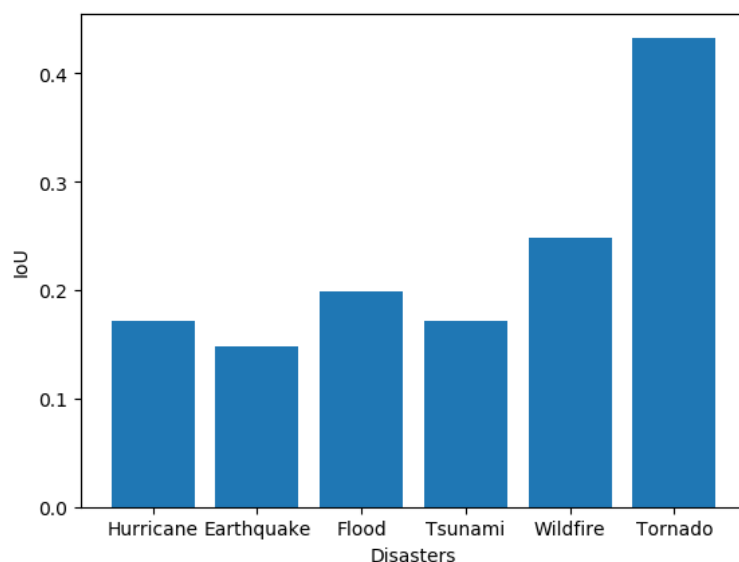


Figure 11. Average IoU for each disaster event.

3. Promotion

As shown in Figure 12, three transfer learning methods, i.e., the CycleGAN, fine-tuning and domain adversarial training, were used to promote the generalization ability of the U-NASNetMobile model in this paper. Transfer learning [22] aims to improve the performance of the task based on the target domain by discovering and transferring latent knowledge from the task based on the source domain. In this paper, the source domain

and target domain are the DREAM-B dataset and xBD dataset, respectively. The CycleGAN and domain adversarial training do not need the labels of xBD, while fine-tuning requires a small number of labels of xBD. The domain adversarial training and fine-tuning method require proper training of the original network model with xBD images, while the CycleGAN method does not need to adjust any parameters of the model.

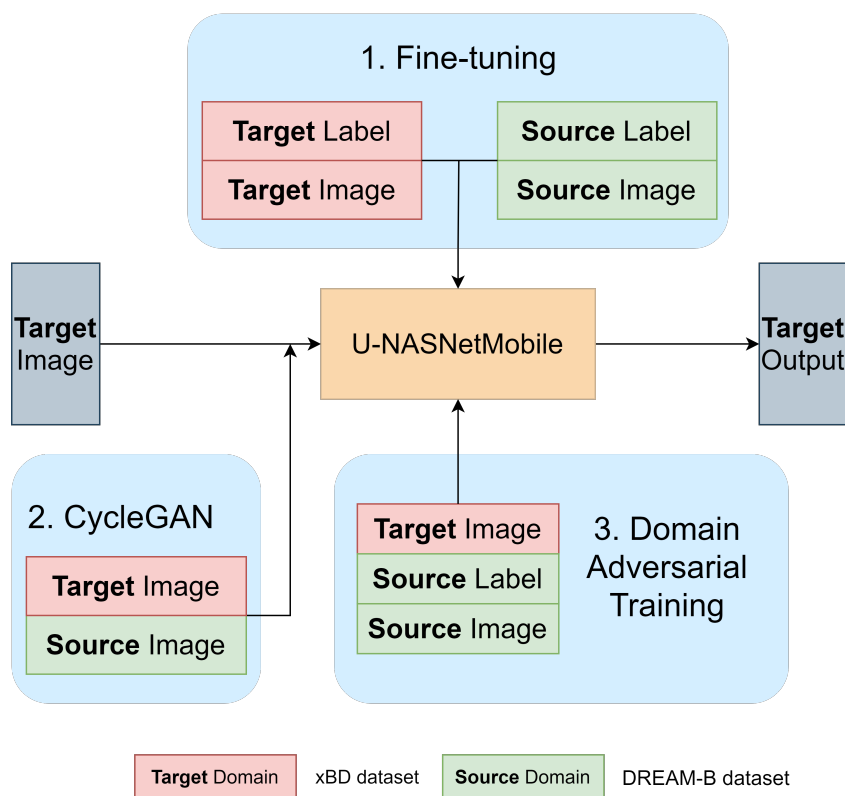


Figure 12. The three transfer learning methods used in this paper.

3.1. Fine-Tuning

3.1.1. Fine-Tuning Using Images from xBD

In this subsection, U-NASNetMobile is fine-tuned with a small set of training samples. This transfer-learning method is based on the pre-training network. When the size of the annotated target dataset is significantly smaller than that of the source dataset, transfer learning based on a pre-training network can be a powerful tool. The convolutional neural network has multiple layers to extract features. The lower layer is used to capture basic common features. The higher layer is applied to learn the advanced features corresponding to the input. Based on this characteristic, we can freeze the lower layer and retrain the specific layer parameters of the higher layer to meet the requirements of the new task. In the experiment, we freeze the encoder of U-NASNetMobile. We retrain the decoder of U-NASNetMobile using the post-disaster images in the xBD dataset. The models were executed using Python 3.7.0 on the platform of Intel(R) Xeon(R) Gold 5118 CPU @ 2.70 GHz system. All experiments were run on a single NVIDIA Tesla P40 GPU with 20 h.

In the actual damage assessment of buildings, the distribution of damaged buildings needs to be obtained as soon as possible due to the needs of post-disaster emergency rescue. However, it is difficult to obtain a large-scale and detailed degree of damage labels of buildings in a short time. However, in many cases, it is possible to obtain the labels of a small number of undamaged buildings. Based on this reality, this method uses a small number of post-disaster images and undamaged building labels to fine-tune U-NASNetMobile so as to promote the building identification accuracy of the network on the new post-disaster data. In order to achieve the goal of rapid damage assessment through a small portion of the labels acquired for post-disaster images, we use the fine-tuning method

to improve the identification accuracy of the U-NASNetMobile model on xBD postdisaster images. Since it is difficult to obtain the labels of post-disaster images on a large scale, we select only 10% of the data from each disaster for model training.

3.1.2. Quantitative Evaluation

Table 4 shows the identification results of the model on 14 disasters. The results show that the recall of all disasters increased after fine-tuning. In addition, the precision of Santarosa-WF, Harvey-HC, Joplin-TD and Moore-TD also increased, while the precision of other disasters decreased in exchange for the increase in the recall. Overall, the IoUs of all disasters increased to different degrees. In addition, the missed detection rates of building objects decreased obviously, while the false detection rates increased slightly. Figure 13 shows the increasing rate of the IoU for each disaster. The Palu-TM, Portugal-WF, Harvey-HC and Socal-WF increased by nearly 100%.

Table 4. Accuracy evaluation of building identification results before and after fine-tuning.

		Recall	Precision	IoU	Kappa	Missed Detection Rate	False Detection Rate
Florence-HC	before	0.211	0.712	0.189	0.284	70.22%	11.93%
	after	0.256	0.683	0.223	0.330	61.34%	16.17%
Harvey-HC	before	0.132	0.662	0.119	0.166	82.10%	15.60%
	after	0.270	0.671	0.229	0.312	48.29%	21.73%
Michael-HC	before	0.250	0.684	0.220	0.320	62.33%	14.91%
	after	0.391	0.648	0.321	0.446	41.18%	22.21%
Mexico-EQ	before	0.116	0.636	0.108	0.128	73.90%	10.20%
	after	0.225	0.552	0.183	0.198	52.84%	35.77%
Midwest-FD	before	0.317	0.722	0.292	0.405	61.28%	5.53%
	after	0.547	0.645	0.424	0.557	35.26%	14.91%
Palu-TM	before	0.176	0.575	0.154	0.203	51.72%	11.31%
	after	0.405	0.535	0.303	0.391	27.49%	18.64%
Santarosa-WF	before	0.219	0.520	0.190	0.255	47.10%	10.15%
	after	0.261	0.550	0.222	0.298	40.83%	12.26%
Socal-WF	before	0.158	0.580	0.148	0.225	67.07%	8.66%
	after	0.343	0.530	0.275	0.381	50.26%	15.76%
Joplin-TD	before	0.328	0.731	0.295	0.400	49.63%	11.58%
	after	0.345	0.745	0.321	0.435	36.35%	15.95%
Moore-TD	before	0.626	0.859	0.560	0.671	28.54%	4.23%
	after	0.686	0.868	0.618	0.724	23.52%	6.27%
Nepal-FD	before	0.126	0.635	0.114	0.178	77.14%	12.51%
	after	0.250	0.576	0.204	0.301	58.77%	17.27%
Portugal-WF	before	0.189	0.892	0.185	0.280	73.71%	7.99%
	after	0.383	0.805	0.354	0.487	48.53%	12.86%
Tuscaloosa-TD	before	0.502	0.780	0.438	0.572	38.61%	12.18%
	after	0.621	0.737	0.510	0.642	24.39%	17.94%
Woolsey-WF	before	0.487	0.823	0.437	0.579	34.51%	9.53%
	after	0.672	0.705	0.517	0.650	19.88%	24.68%

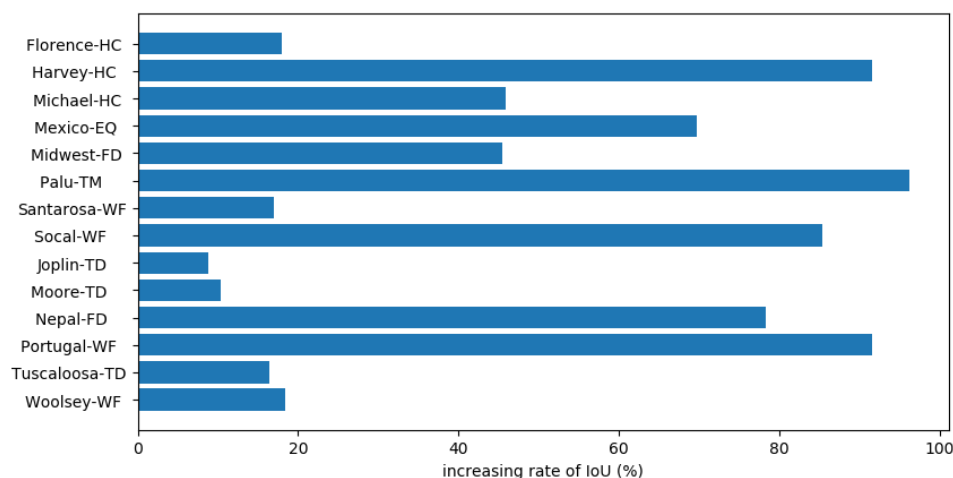


Figure 13. The increasing rate of the IoU for each disaster.

3.1.3. Qualitative Comparison

Figure 14 shows a post-disaster image from Harvey-HC with its ground truth and identification result before and after fine-tuning. From Figure 14a, highlighted in red, we can see that the buildings in this image are not affected by the hurricane at all. The U-NASNetMobile model missed most of the buildings. This suggests that, at least on this image, the generalization ability of U-NASNetMobile is very poor because most buildings are confused with the background. However, after fine-tuning, the building identification results were greatly improved. Only a small number of buildings were missed.

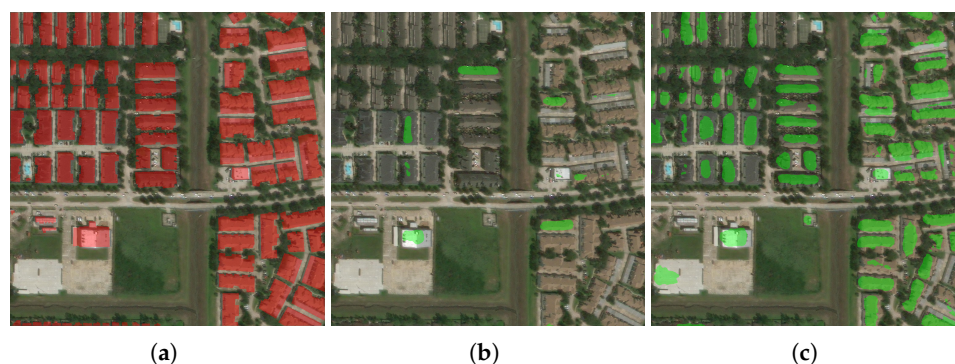


Figure 14. The images in the (a–c) subfigures represent the ground truth, the result before fine-tuning, and the result after fine-tuning, respectively.

3.2. CycleGAN

3.2.1. Image Translation from xBD to DREAM-B

The traditional image enhancement methods are generally divided into spatial domain enhancement and frequency domain enhancement. Spatial domain enhancement methods, such as histogram equalization algorithm, may cause image blurring and amplify image noise. Frequency domain enhancement methods, such as directional-filter methods, usually have problems of contrast and sharpness reduction and partial feature loss. The generation model based on deep learning can obtain features from the data at different levels by abstracting the original data layer by layer, which solves the limitations of traditional methods to a certain extent. The Generative Adversarial Network (GAN) [23] consists of a generator and a discriminator. The generator attempts to generate fake images. The discriminator seeks to distinguish the difference between fake images and real images. In the training process of a GAN, the performance of the generator and discriminator is

alternately improved. Finally, the generator can be used to generate images that approximate real samples. Based on pix2pix [24], Zhu et al. [15] proposed a cycle generative adversarial network (CycleGAN) method, which is a deformation of a conventional GAN. The CycleGAN can transform an image from source domain X to target domain Y without paired training data. Many studies have applied the CycleGAN to various fields [25–28].

After a disaster occurs, it is necessary to evaluate building damage. In many cases, there are many building identification models that have been trained on remote-sensing images without disasters. However, the training data and the actual postdisaster data are greatly different for different data sources or in different regions. Furthermore, because of the changes in the imaging conditions caused by disasters, the two kinds of data can be very different. Therefore, the generalization ability of an existing model on new post-disaster data is poor. In this part of this paper, image translation based on the CycleGAN is applied to make the xBD images have the style of the DREAM-B images. The CycleGAN can make the distribution of the two types of data as close as possible so as to promote the generalization ability of the models on new data without affecting the performance of the existing models.

As shown in Figure 15, the CycleGAN has two mirror GANs. The ring network is formed by two generators G and F and two discriminators D₁ and D₂. In this paper, generator G utilizes an image from DREAM-B to generate a fake xBD image, and vice versa for generator F. The function of discriminator D₂ is to distinguish real and fake xBD images, and vice versa for discriminator D₁. In the training process of the CycleGAN, generators G and F seek to generate real xBD and DREAM-B images, respectively. Discriminators D₁ and D₂ seek to distinguish real and fake images. Finally, we can use the generator F to transform the images from xBD to DREAM-B. The models were executed using Python 3.7.4 on the platform of Intel(R) Xeon(R) Gold 6226 CPU @ 2.70 GHz system. All experiments were run on a single NVIDIA Tesla v100 GPU with 30 h.

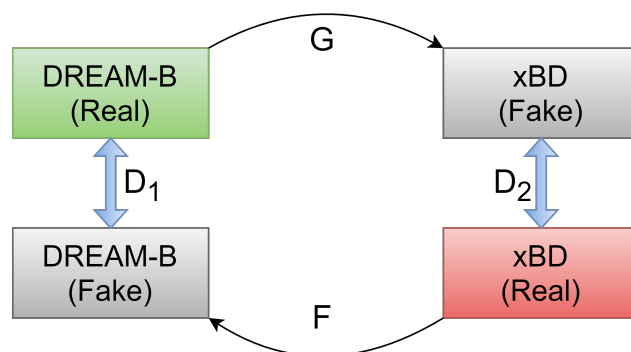


Figure 15. Structure of the CycleGAN. The two one-way GANs share two generators, and each has a discriminator.

3.2.2. Quantitative Evaluation

In order to maintain consistency with the resolution of the DREAM-B dataset, we converted the image size of the xBD dataset to 2048×2048 . After conducting training and prediction using the post-disaster images of 14 disasters, the quantitative evaluation is shown in Table 5. This evaluation shows that the missed classification and missed detection rates are lower than those before image translation. The CycleGAN results show that at the pixel level, the recall of most disasters increased. However, the precision decreased slightly after image translation based on the CycleGAN.

Table 5. Accuracy evaluation of the building identification results before and after CycleGAN translation.

		Recall	Precision	IoU	Kappa	Missed Detection Rate	False Detection Rate
Florence-HC	before	0.210	0.689	0.189	0.283	70.54%	17.89%
	after	0.303	0.629	0.257	0.369	60.10%	24.09%
Harvey-HC	before	0.121	0.587	0.107	0.149	80.49%	25.33%
	after	0.309	0.591	0.254	0.322	42.45%	31.87%
Michael-HC	before	0.247	0.697	0.218	0.317	62.30%	14.81%
	after	0.435	0.656	0.350	0.475	38.58%	19.43%
Mexico-EQ	before	0.160	0.729	0.148	0.176	66.52%	10.37%
	after	0.282	0.706	0.246	0.284	49.59%	11.62%
Midwest-FD	before	0.307	0.726	0.284	0.393	60.87%	5.25%
	after	0.401	0.680	0.351	0.474	46.86%	8.08%
Palu-TM	before	0.197	0.593	0.171	0.224	51.42%	11.23%
	after	0.218	0.551	0.185	0.241	46.89%	10.73%
Santarosa-WF	before	0.259	0.522	0.216	0.286	37.04%	9.50%
	after	0.421	0.550	0.329	0.428	24.96%	13.63%
Socal-WF	before	0.171	0.593	0.159	0.239	59.07%	6.97%
	after	0.172	0.557	0.161	0.234	58.67%	7.02%
Joplin-TD	before	0.350	0.736	0.310	0.416	46.77%	10.97%
	after	0.361	0.745	0.328	0.441	52.77%	8.73%
Moore-TD	before	0.625	0.852	0.556	0.666	28.15%	4.48%
	after	0.705	0.798	0.593	0.701	23.99%	5.87%
Nepal-FD	before	0.125	0.646	0.114	0.179	78.27%	12.29%
	after	0.059	0.559	0.054	0.085	91.00%	12.88%
Portugal-WF	before	0.197	0.904	0.193	0.292	72.46%	7.65%
	after	0.270	0.780	0.250	0.353	67.84%	8.06%
Tuscaloosa-TD	before	0.499	0.779	0.434	0.568	39.10%	12.17%
	after	0.550	0.762	0.467	0.597	40.11%	11.92%
Woolsey-WF	before	0.470	0.831	0.425	0.567	36.36%	9.10%
	after	0.585	0.761	0.494	0.635	29.21%	12.81%

In general, except for the Nepal-FD, other disasters also showed different degrees of increase in their IoU. The most significant improvement was in Harvey-HC, where the IoU increased by 0.147. The evaluation results at the building object level are consistent with those at the pixel level. The missed detection rates of most disasters decreased obviously, but the false detection rates increased slightly. Thus, in the absence of supervised information guidance, the CycleGAN image translation method can improve the building identification accuracy. In the training process, the CycleGAN network can learn the difference between two datasets, especially the specific information of the DREAM-B dataset learned by the U-NASNetMobile model. Figure 16 shows the increase in the IoU of each disaster. Except for Nepal-FD, all disasters have increased IoUs to different extents. Among the disasters, the increase in the IoU of Harvey-HC is over 100%.

3.2.3. Qualitative Comparison

Figure 17 shows a DREAM-B image and an xBD image before and after image translation. The DREAM-B image is from Shanghai, China. The xBD image is from Harvey-HC. It can be seen that the overall style of the image from xBD is green and grey with low contrast before image translation. Most of the images in the DREAM-B dataset are from China. The shooting time and imaging conditions are different from the Harvey-HC postevent images. As a result, the image tone and style of the two domains are different. The buildings in Figure 17e are closely connected while the distance between the buildings in Figure 17c is relatively larger. There are two main changes in the translated image. First, the overall tone changes from green and grey to blue. The buildings in the image become clearer, with sharper edges and a higher contrast to the background. It can also be seen from Figure 17d–f that all the buildings in this image were not affected by the hurricane disaster at all, while the U-NASNetMobile

model missed most of the buildings. However, the model missed only a small part of the buildings after image translation by the CycleGAN.

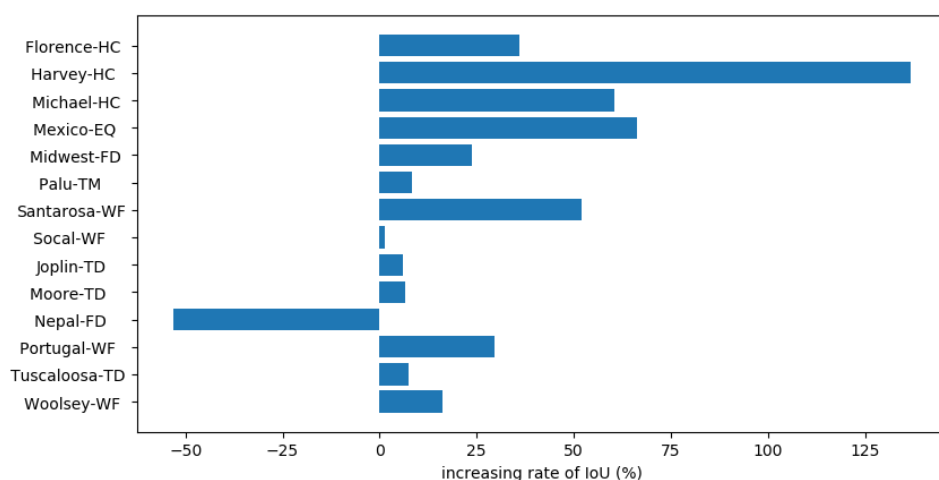


Figure 16. The increasing rate of the IoU for each disaster.

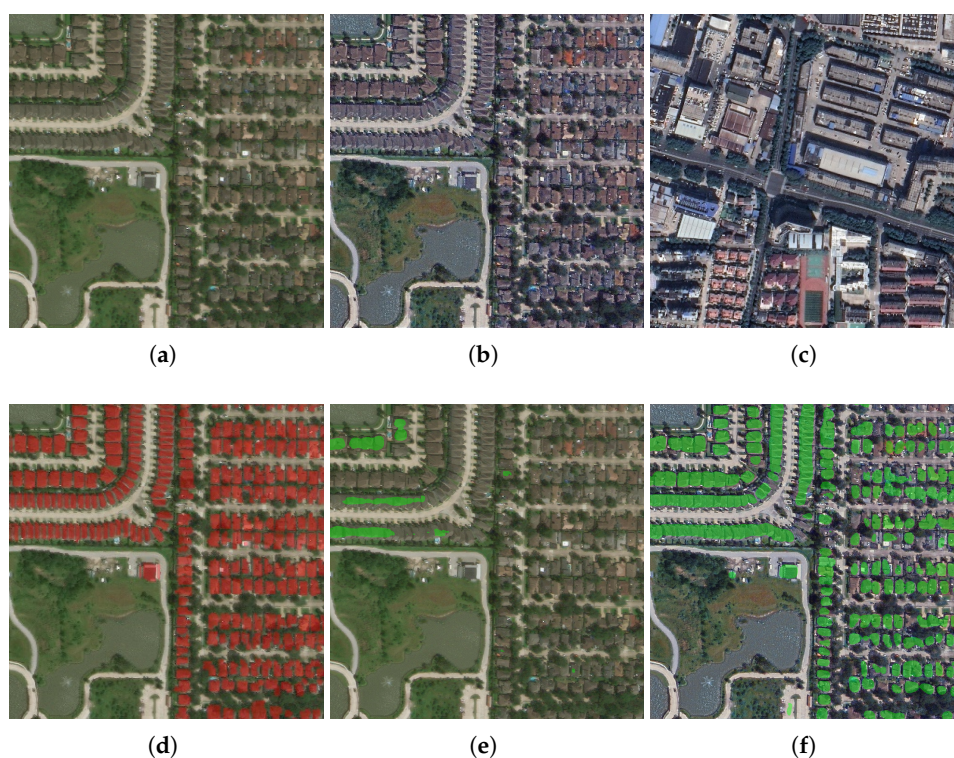


Figure 17. Harvey-HC images and identification results before and after image translation. (a) Image before translation, (b) Image after translation, (c) Image from DREAM-B, (d) Labels, (e) Identification results before translation, and (f) Identification results after translation.

3.3. Domain Adversarial Training

3.3.1. Domain Adversarial Training between xBD and DREAM-B

Domain adversarial training [16] is a classic method based on adversarial training for domain adaptive research. There is one more objective function on the basis of the existing model to encourage confusion between the two domains.

In this paper, the network structure of domain adversarial training is shown in Figure 18, and the structure is divided into three parts, i.e., a feature extractor, a label predictor and a domain classifier. We split U-NASNetMobile into a feature extractor and label predictor.

The domain classifier, which consists of a convolutional layer and a fully connected layer, is connected to the penultimate layer. The inputs of the network are two images from DREAM-B and xBD. The feature extractor should generate the same distributed features regardless of the input from DREAM-B or xBD. The label predictor is used to classify the buildings. The domain classifier should distinguish whether the data features come from DREAM-B or xBD.

The gradient reversal layer (GRL), which reverses the gradient in the backpropagation, is inserted between the feature extractor and domain classifier to achieve domain adversarial training. The domain classifier after the GRL minimizes the domain classification loss, while the feature extractor before the GRL maximizes the domain classification loss. Finally, the domain classifier is unable to distinguish the features from DREAM-B and xBD. The feature spaces of DREAM-B and xBD are completely mixed together. The gradient reversal layer will cause the model to maximize the error to some extent instead of only minimizing the objective function like the original model. This means that the model can learn the features that minimize the objective function without allowing the two domains to be distinguished.

Like the CycleGAN, the domain adversarial training transfer-learning method does not require the labelled information of xBD images. All the post-disaster images of different disasters are used for domain adversarial training. The same part of the domain adversarial network as U-NASNetMobile uses the weight pretrained on the DREAM-B dataset. The models were executed using Python 3.7.0 on the platform of Intel(R) Xeon(R) Gold 5118 CPU @ 2.70 GHz system. All experiments were run on a single NVIDIA Tesla P40 GPU with 30 h.

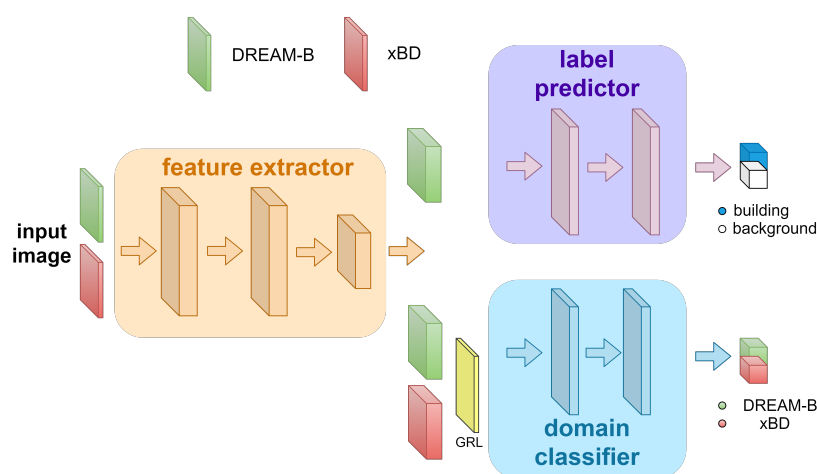


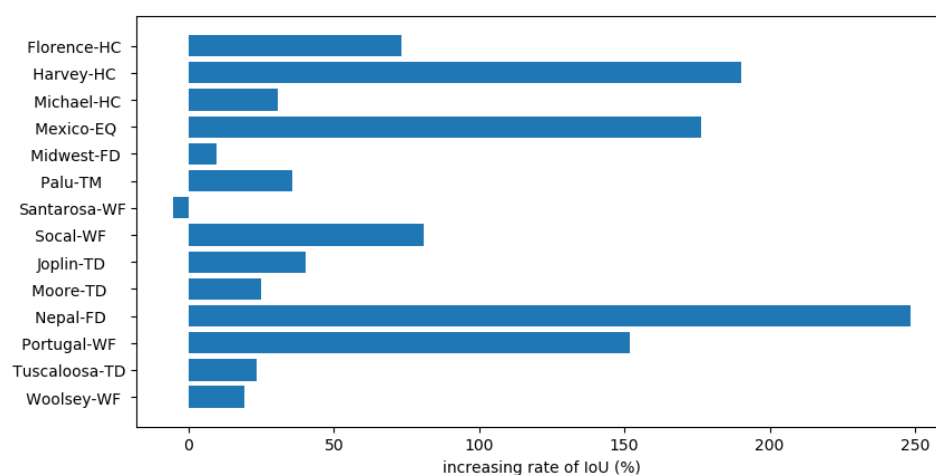
Figure 18. The structure of domain adversarial training.

3.3.2. Quantitative Evaluation

The accuracy evaluation after domain adversarial training and prediction is shown in Table 6. The increase in the IoU for each disaster is shown in Figure 19. Table 6 shows that the recall of all disasters has increased significantly. Among the disasters, the highest increase is in Nepal-FD, of 0.638. In addition, the precision of many disasters fell. The greatest reduction was in the Midwest-FD, with a reduction of 0.338. However, in general, as shown in Figure 19, except for Santarosa-WF, all the IoUs increased to varying degrees. The evaluation results at the building object level are consistent with those at the pixel level. The missed detection rates of most disasters decreased obviously. However, the false detection rates increased to different degrees.

Table 6. Accuracy evaluation of building identification results before and after domain adversarial training.

		Recall	Precision	IoU	Kappa	Missed Detection Rate	False Detection Rate
Florence-HC	before	0.217	0.721	0.195	0.292	70.06%	19.19%
	after	0.606	0.469	0.338	0.452	36.68%	71.89%
Harvey-HC	before	0.131	0.589	0.114	0.157	78.00%	25.16%
	after	0.452	0.546	0.331	0.413	32.66%	42.70%
Michael-HC	before	0.265	0.680	0.232	0.334	60.15%	15.76%
	after	0.853	0.324	0.304	0.400	7.12%	62.35%
Mexico-EQ	before	0.157	0.724	0.145	0.173	67.12%	11.57%
	after	0.494	0.690	0.401	0.450	32.84%	11.29%
Midwest-FD	before	0.313	0.729	0.289	0.401	59.89%	6.15%
	after	0.656	0.391	0.317	0.420	18.51%	55.96%
Palu-TM	before	0.214	0.613	0.186	0.246	49.66%	10.47%
	after	0.305	0.617	0.252	0.334	41.48%	13.18%
Santarosa-WF	before	0.261	0.518	0.218	0.287	36.61%	11.00%
	after	0.707	0.237	0.206	0.257	8.10%	70.59%
Socal-WF	before	0.172	0.577	0.159	0.237	58.38%	18.14%
	after	0.398	0.511	0.288	0.397	38.85%	48.64%
Joplin-TD	before	0.388	0.715	0.334	0.447	44.79%	12.86%
	after	0.828	0.509	0.469	0.567	13.14%	41.95%
Moore-TD	before	0.635	0.841	0.561	0.673	27.70%	5.10%
	after	0.863	0.789	0.702	0.793	16.97%	9.45%
Nepal-FD	before	0.115	0.623	0.106	0.167	79.36%	13.13%
	after	0.753	0.424	0.368	0.489	14.68%	40.98%
Portugal-WF	before	0.149	0.624	0.145	0.216	71.98%	40.97%
	after	0.436	0.554	0.366	0.460	37.72%	45.63%
Tuscaloosa-TD	before	0.511	0.770	0.443	0.577	37.04%	13.24%
	after	0.826	0.614	0.547	0.670	10.11%	37.72%
Woolsey-WF	before	0.496	0.830	0.446	0.591	34.43%	9.99%
	after	0.676	0.723	0.531	0.671	24.45%	20.97%

**Figure 19.** The increasing rate of the IoU for each disaster.

3.3.3. Qualitative Comparison

Figure 20 shows a post-disaster image from Mexico-EQ, with its label and the identification results before and after domain adversarial training. From Figure 20a highlighted in red, we can see that the buildings in this image are not affected by the earthquake at all. The global model only identified a few buildings and missed most of the buildings. However, after domain adversarial training, the result greatly improved

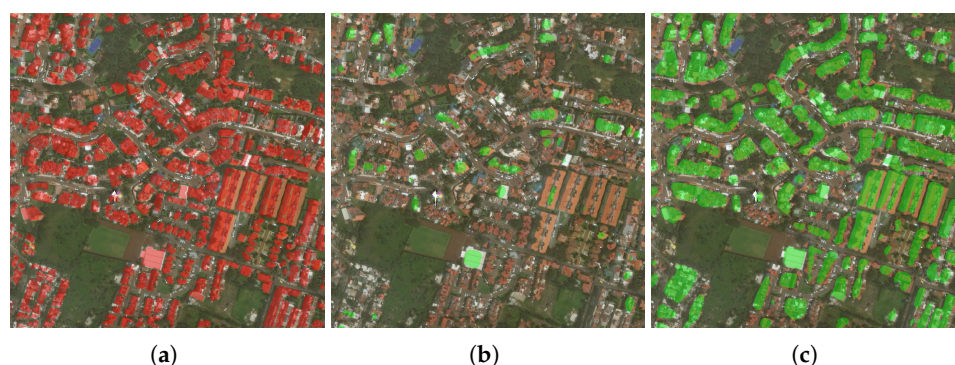


Figure 20. The images in the (a–c) subfigures represent the ground truth, the result before domain adversarial training, and the result after domain adversarial training, respectively.

4. Discussion

In order to improve the reliability and accuracy of building damage extraction via post-classification comparison, we aim to promote the generalization ability of a global model for building mapping using heterogeneous satellite images from multiple natural disaster scenarios. In this paper, we use the satellite images in the xBD dataset and its manual labels of the degree of building damage. Compared with Landsat and Sentinel data, the image resolution of the xBD dataset is much higher. In addition, this dataset has building labels with different degrees of damage, which can save the time and effort of manual annotating. The xBD dataset is the largest and best-quality building damage dataset currently available. Therefore, the results based on this dataset should also be reliable. The experimental results show that when the existing building identification model is directly applied to actual postdisaster images without any transfer learning method, the overall performance is poor. The difference between different disasters is particularly large. This may be caused by the type of disaster. It can be found in the experiment of this paper that the performance of the model is generally poor when there are a large number of water-covered damage buildings in an image after a disaster. In addition, in the accuracy evaluation of this paper, the major damage and destroyed buildings in the xBD dataset are all treated as the background, which may be a factor affecting the result. However, this is not the main factor affecting the low generalization ability of global building-mapping models in post-disaster images. Previous research suggested that image parameters, such as the off-nadir angle, can influence the performance [11,29]. We can see from the results of the qualitative analysis that the building damage and the changes in imaging conditions caused by disasters are the main factors that influence the performance of the model. After the disaster, some buildings were damaged and lost the building features that the global model learned. The changes in imaging conditions also blurred the features of the undamaged buildings in the post-disaster images, thus influencing the performance. Therefore, the global model performs poorly when applied to the post-disaster images of xBD dataset. In general, the U-NASNetMobile model trained on DREAM-B has a poor generalization ability in the post-disaster images of the xBD dataset. The post-classification comparison method is completely dependent on the performance of the existing building identification model. Consequently, it is not advisable to directly apply the existing model to actual post-disaster images.

In view of the low generalization ability of the existing global model, we use the transfer-learning method to seek to promote the generalization ability of the existing models, hoping to make the post-classification comparison method more feasible and reliable. We systemically analyse three kinds of methods, i.e., fine-tuning the model using very few training samples randomly selected from each disaster, transferring the style of post-disaster satellite images using the CycleGAN, and feature transformation using domain adversarial training, to promote the generalization ability using post-disaster satellite images. Image translation based on the CycleGAN does not need to use the

manual annotation information of the xBD dataset. It also does not make any changes to the existing model. By means of image translation between the two domains, the translated xBD image can acquire the image features that are conducive to classification. The training method based on domain adversarial training does not need to use the manual annotation information of the xBD dataset. However, it needs to conduct adversarial training on the basis of the existing model. The fine-tuning method not only needs to use the manual annotation information of the xBD dataset, but it also needs to retrain the existing model. The experimental results show that the three methods have obvious promotional effects on the performance of the model. There is a phenomenon where the recall tends to increase at the expense of the precision, which is not a bad thing for the post-classification comparison method. Since the generalization abilities of the three methods are different in principle, the performances of the three methods are also different.

First, the fine-tuning method uses the most information. This should be the best way to improve the generalization ability. However, only a limited amount of annotated data from the xBD dataset were used in the fine-tuning experiment. Therefore, the improvement in the generalization ability was not significantly higher than those of the other two methods. It is worth mentioning that this method of promotion is the most stable without a negative transfer phenomenon. The training process of the model is always conducted in the direction of improving the building identification accuracy of xBD data. All disasters showed varying degrees of increase in their IoU. There is no manual annotation information in the CycleGAN and domain adversarial training. Therefore, the impact on the model's generalization ability is not as stable with a certain degree of negative transfer phenomenon in a few disasters. Although they had positive impacts in most of the disasters, there were always 1 or 2 disasters whose building identification accuracy was reduced. Previous studies have found that the domain adversarial training, combined with cycle-consistency constraints, can improve the performance of semantic segmentation model [28]. However, the CycleGAN experiment in this paper does not change any parameters of the model, nor does it require any supervision information of the target domain. Therefore, it is a surprise that the image translation based on the CycleGAN could obtain such a good result. However, in the training process of the CycleGAN, a reduction in the loss may not mean an improvement in the performance of the existing building identification model. Given the potential for negative transfer and the unclear building features in the images of Nepal, as shown in Figure 7, there is a large decrease in IoU accuracy in the Nepal experiment. Except for Nepal-FD, the identification accuracy of all disasters is still generally improved. Overall, image translation based on the CycleGAN can improve the performance of the existing model on postdisaster images to some extent. Domain adversarial training causes the model to lose part of the unique characteristics on the DREAM-B dataset during training, which may reduce the performance on the DREAM-B dataset to a certain extent. However, after adversarial training, the recall of building identification on the postdisaster images from the xBD dataset is considerably increased. This increase is the largest increase among the three transfer-learning methods. Although the precision decreased considerably, in general, it improved the building identification performance on the xBD dataset.

The transfer-learning experiment in this paper conducts training on different disasters. Therefore, it is inevitable that the promotion of the generalization ability may be caused by the multiplicity of the model. When we transfer the model to 14 different disasters, there will be 14 different transfer directions to meet the characteristics of the current disaster. Therefore, it seems unfair to evaluate these 14 “models” together with the results of one previous model. However, due to the diversity of the disasters in the xBD dataset, such experiments are necessary. Moreover, in many cases, we only face one type of disaster at a time. It is more realistic to conduct experiments on different types of disasters.

5. Conclusions

In conclusion, in order to evaluate building damage via post-classification comparison, we first evaluate the generalization ability of a global model trained on aerial images

using post-disaster satellite images. Then, we systemically analyse three kinds of methods, i.e., fine-tune the model using very few training samples randomly selected from each disaster, transfer the style of post-disaster satellite images using the CycleGAN, and perform feature transformation using domain adversarial training, to promote the generalization ability for post-disaster satellite images.

The research results show that the performance of the existing global building mapping model is poor when directly applied to xBD post-disaster images. Even undamaged buildings are difficult for the model to recognize, i.e., the recall of the identification results is generally low. Furthermore, the model shows wide differences for various disasters. When there are a large number of water-covered damaged buildings in the post-disaster images, the performance of the model is generally poor. Therefore, when the generalization ability of the model is not guaranteed, it is not advisable to use the existing global building mapping model to assess damage via post-classification comparison.

The research in this paper mainly focuses on the generalization ability of the existing global building-mapping model on the post-disaster images of the xBD dataset. Overall, the promotion of the performance of the model is very obvious. From these results, we can find that even if the generalization ability of a building global-mapping model may not be satisfactory, some transfer-learning methods can also be used to promote the identification performance so as to provide strong support for the post-classification comparison method to assess damage. When the annotation information of post-disaster images is available, fine-tuning will be the most reliable transfer learning method to avoid unnecessary negative transfer. When the annotation information is not available, the image-translation method based on CycleGAN and the method of domain adversarial training will also be a good method to improve the generalization ability.

Author Contributions: Conceptualization, H.T.; Funding acquisition, H.T.; Investigation, H.T. and Y.H.; Methodology, Y.H.; Software, Y.H.; and Supervision, H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant No. 41971280 and in part by the National Key R&D Program of China under Grant No. 2017YFB0504104.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We would like to thank the high-performance computing support from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University (<https://gda.bnu.edu.cn/>), accessed on 10 September 2020).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Dong, L.; Shan, J. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS J. Photogramm. Remote Sens.* **2013**, *84*, 85–99. [[CrossRef](#)]
2. Tomowski, D.; Klonus, S.; Ehlers, M.; Michel, U.; Reinartz, P. Change visualization through a texture-based analysis approach for disaster applications. In Proceedings of the ISPRS Proceedings, Vienna, Austria, 5–7 July 2010; pp. 1–6.
3. Miura, H.; Modorikawa, S.; Chen, S.H. Texture characteristics of high-resolution satellite images in damaged areas of the 2010 Haiti earthquake. In Proceedings of the 9th International Workshop on Remote Sensing for Disaster Response, Stanford, CA, USA, 15–16 September 2011; pp. 15–16.
4. Chini, M.; Cinti, F.; Stramondo, S. Co-seismic surface effects from very high resolution panchromatic images: The case of the 2005 Kashmir (Pakistan) earthquake. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 931–943. [[CrossRef](#)]
5. Zhao, F.; Zhang, C. Building Damage Evaluation from Satellite Imagery using Deep Learning. In Proceedings of the 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), Las Vegas, NV, USA, 11–13 August 2020; pp. 82–89.

6. Kalantar, B.; Ueda, N.; Al-Najjar, H.A.; Halin, A.A. Assessment of Convolutional Neural Network Architectures for Earthquake-Induced Building Damage Detection based on Pre-and Post-Event Orthophoto Images. *Remote Sens.* **2020**, *12*, 3529. [\[CrossRef\]](#)
7. Ma, J.; Qin, S. Automatic depicting algorithm of earthquake collapsed buildings with airborne high resolution image. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 939–942.
8. Ci, T.; Liu, Z.; Wang, Y. Assessment of the Degree of Building Damage Caused by Disaster Using Convolutional Neural Networks in Combination with Ordinal Regression. *Remote Sens.* **2019**, *11*, 2858. [\[CrossRef\]](#)
9. Miura, H.; Aridome, T.; Matsuoka, M. Deep learning-based identification of collapsed, non-collapsed and blue tarp-covered buildings from post-disaster aerial images. *Remote Sens.* **2020**, *12*, 1924. [\[CrossRef\]](#)
10. Valentijn, T.; Margutti, J.; van den Homberg, M.; Laaksonen, J. Multi-hazard and spatial transferability of a cnn for automated building damage assessment. *Remote Sens.* **2020**, *12*, 2839. [\[CrossRef\]](#)
11. Nex, F.; Duarte, D.; Tonolo, F.G.; Kerle, N. Structural building damage detection with deep learning: Assessment of a state-of-the-art cnn in operational conditions. *Remote Sens.* **2019**, *11*, 2765. [\[CrossRef\]](#)
12. Li, Y.; Lin, C.; Li, H.; Hu, W.; Dong, H.; Liu, Y. Unsupervised domain adaptation with self-attention for post-disaster building damage detection. *Neurocomputing* **2020**, *415*, 27–39. [\[CrossRef\]](#)
13. Gupta, R.; Hosfelt, R.; Sajeev, S.; Patel, N.; Goodman, B.; Doshi, J.; Heim, E.; Choset, H.; Gaston, M. xbd: A dataset for assessing building damage from satellite imagery. *arXiv* **2019**, arXiv:1911.09296.
14. Yang, N.; Tang, H. GeoBoost: An Incremental Deep Learning Approach toward Global Mapping of Buildings from VHR Remote Sensing Images. *Remote Sens.* **2020**, *12*, 1794. [\[CrossRef\]](#)
15. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
16. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 1180–1189.
17. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
18. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8697–8710.
19. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
20. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
21. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [\[CrossRef\]](#)
22. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018*; Springer: Cham, Switzerland, 2018; pp. 270–279.
23. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
24. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
25. Agustsson, E.; Tschannen, M.; Mentzer, F.; Timofte, R.; Gool, L.V. Generative adversarial networks for extreme learned image compression. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 221–231.
26. Engin, D.; Genç, A.; Kemal Ekenel, H. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 825–833.
27. Dudhane, A.; Murala, S. Cdnet: Single image de-hazing using unpaired adversarial training. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; pp. 1147–1155.
28. Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.Y.; Isola, P.; Saenko, K.; Efros, A.; Darrell, T. Cycada: Cycle-consistent adversarial domain adaptation. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1989–1998.
29. Van Etten, A.; Lindenbaum, D.; Bacastow, T.M. Spacenet: A remote sensing dataset and challenge series. *arXiv* **2018**, arXiv:1807.01232.