sensors

MDPI

Article

Classification of Building Damage Using a Novel Convolutional Neural Network Based on Post-Disaster Aerial Images

Zhonghua Hong [1], Hongzheng Zhong [1], Haiyan Pan [1,*], Jun Liu [1,2,3], Ruyan Zhou [1], Yun Zhang [1], Yanling Han [1], Jing Wang [1], Shuhu Yang [1] and Changyue Zhong [3]

[1] College of Information Technology, Shanghai Ocean University, Shanghai 201306, China
[2] National Earthquake Response Support Service, Beijing 100049, China
[3] College of Civil Engineering and Architecture, Guizhou Minzu University, Guiyang 550025, China
* Correspondence: hy-pan@shou.edu.cn; Tel.: +86-021-61900612

Abstract: The accurate and timely identification of the degree of building damage is critical for disaster emergency response and loss assessment. Although many methods have been proposed, most of them divide damaged buildings into two categories—intact and damaged—which is insufficient to meet practical needs. To address this issue, we present a novel convolutional neural network—namely, the earthquake building damage classification net (EBDC-Net)—for assessment of building damage based on post-disaster aerial images. The proposed network comprises two components: a feature extraction encoder module, and a damage classification module. The feature extraction encoder module is employed to extract semantic information on building damage and enhance the ability to distinguish between different damage levels, while the classification module improves accuracy by combining global and contextual features. The performance of EBDC-Net was evaluated using a public dataset, and a large-scale damage assessment was performed using a dataset of post-earthquake unmanned aerial vehicle (UAV) images. The results of the experiments indicate that this approach can accurately classify buildings with different damage levels. The overall classification accuracy was 94.44%, 85.53%, and 77.49% when the damage to the buildings was divided into two, three, and four categories, respectively.

Keywords: building damage; deep learning; earthquake building damage classification net (EBDC-Net); aerial images

Citation: Hong, Z.; Zhong, H.; Pan, H.; Liu, J.; Zhou, R.; Zhang, Y.; Han, Y.; Wang, J.; Yang, S.; Zhong, C. Classification of Building Damage Using a Novel Convolutional Neural Network Based on Post-Disaster Aerial Images. *Sensors* **2022**, *22*, 5920. https://doi.org/10.3390/s22155920

Academic Editors: Lei Deng and Xianglei Liu

Received: 28 June 2022
Accepted: 4 August 2022
Published: 8 August 2022

1. Introduction

As some of the most catastrophic events in nature, earthquakes can cause significant structural damage to buildings [1]. The timely and accurate classification of the degree of building damage is of great importance to the government's emergency response and rescue operations. Remote sensing images can be used to obtain abundant spatiotemporal information in the affected area so that buildings can be evaluated on a large scale, at low cost, and quickly [2].

Convolutional neural networks (CNNs) have powerful feature learning and inference capabilities, as well as strong performance in image processing tasks [3]. Therefore, CNNs are widely used in the damage assessment of buildings. According to the number of images used, building damage assessment methods are classified as dual-temporal and single-temporal methods [4].

The dual-temporal methods extract features from the pre- and post-disaster images, and then determine the localization and the degree of damaged buildings [4]. Wu et al. constructed a Siamese neural network with different backbones to automatically detect damaged buildings, and used the attention gate to filter useless features [5]. Xiao et al. proposed a dynamic cross-fusion network that enables the localization and classification tasks of buildings to share feature information from different levels of the CNN network,

enhancing information exchange across tasks [6]. Adriano et al. developed a damage assessment network by combining multimodal and multi-temporal data to increase the utility of the model under different data sources [7]. Since the dual-temporal methods employ both pre- and post-disaster images, they usually have a high classification accuracy [8]. However, the practicality of these methods is greatly limited due to the accessibility of dual-temporal images [9,10].

Single-temporal methods only use post-disaster images for building damage assessment tasks. Therefore, they are subject to relatively few constraints. Duarte et al. proposed a CNN with multiresolution feature fusion to increase the performance of the model in multiresolution images [11]. Ji et al. explored the use of pre-trained CNN and fine-tuned CNN strategies for the damage classification of buildings after earthquakes [12]. Nex et al. assessed the migration performance of the CNN using images of different locations and spatial resolutions [13]. Ishraq et al. replaced the fully connected layer in the CNN with a global average pooling layer to assess building damage caused by hurricanes [14]. However, the majority of the existing research divides buildings into two categories—intact and damaged—which cannot meet the needs of rescue and post-disaster damage refinement assessment. More detailed classification information about the degree of damage to buildings is needed [15]. Ci et al. combined a CNN with ordinal regression to classify building damage as intact, slightly damaged, severely damaged, or collapsed [16]. Ma et al. used geographic information system (GIS) data to provide evident boundary characteristics of buildings. A CNN model combined with GIS data was proposed to classify building damage into slight damage, moderate damage, and severe damage [17]. However, these studies ignore the impact of the extraction of building damage features on classification accuracy. In post-disaster images, the shape and texture of the building change significantly [18]. Distinguishing between slight damage and severe damage is a challenging task because they share similar characteristics. For instance, the damage characteristics of buildings are mainly manifested in roofs, except that the damaged area is different. Therefore, it is necessary to aggregate similar features to enhance the discrimination of different degrees of building damage [2]. In addition, texture and spatial information around the buildings can provide necessary auxiliary information for evaluation. Exploring the relationship between global features and context features in images helps the model to classify the damage levels of buildings more accurately.

To address the abovementioned issues, this study proposes a novel CNN—namely, the earthquake building damage classification net (EBDC-Net)—for assessment of building damage using post-disaster aerial images. The proposed network is made up of a feature extraction encoder module and a damage classification module. The feature extraction encoder module is used to extract the semantic information and enhance the feature representation capability of different damage levels of buildings from the images, while the damage classification module is used to fuse the global and contextual features to improve the accuracy of damage classification.

The rest of this paper is organized as follows: Section 2 introduces the data sources and the proposed method. The experimental results are presented in Section 3. Section 4 discusses the role of historical earthquake data in new earthquakes. Finally, some conclusions are drawn in Section 5.

2. Materials and Methods

2.1. Data Sources

The datasets used in this study contain post-earthquake images from three different locations. The first comprises post-earthquake aerial images of the 7.1 magnitude earthquake that occurred on 14 April 2010 in Yushu County, Qinghai Province, China [16]. The second comprises post-earthquake aerial images of the 6.5 magnitude earthquake that occurred on 3 August 2014 in Ludian County, Yunnan Province, China [16]. The third comprises post-earthquake UAV images of the 6.4 magnitude earthquake that occurred on 21 May 2021 in Yangbi County, Yunnan Province, China. The Yushu and Ludian datasets are public

datasets, and can be downloaded from https://github.com/city292/build_assessment (accessed on 1 April 2022) [16]. For the Yangbi dataset, the region of interest (ROI) was first cropped from the post-disaster UAV images of the Yangbi earthquake. Then, the ROI was cropped into patches of different sizes according to the resolution of the images and the structural features of the local buildings. Finally, all patches were uniformly resized to 88 × 88 pixels. As shown in Table 1, the datasets classify building damage into four categories: intact, slightly damaged, severely damaged, and collapsed. As shown in Table 2, the number of images with four damage levels in the dataset was counted.

Table 1. Building examples and image information of the four damage levels in the datasets.

Dataset	Intact	Slightly Damaged	Severely Damaged	Collapsed	Image Size	Resolution
Yushu Dataset [16]					88 × 88 Pixel	0.1 m
Ludian Dataset [16]					88 × 88 Pixel	0.2 m
Yangbi Dataset				-	88 × 88 Pixel	0.03–0.2 m

Table 2. Statistics on the number of buildings sampled in the datasets with the four levels of damage.

Damage Level	Description	Ludian Dataset	Yushu Dataset	Yangbi Dataset
L0	Intact	1630	778	928
L1	Slightly damaged	3074	918	202
L2	Severely damaged	1685	665	111
L3	Collapsed	1984	1140	-
Total		8337	3510	1241

To verify the performance of EBDC-Net on the different classification criteria, the buildings with different damage levels were divided into three groups, as shown in Table 3. Group 1 simply divided all of the buildings into non-collapsed and collapsed, without distinguishing the degrees of damage to the buildings. Group 2 contained three categories, namely, intact, severely damaged, and collapsed. Slightly damaged buildings were considered to be intact. For Group 3, a more detailed classification criterion was devised, and all of the buildings were divided into four categories: intact, slightly damaged, severely damaged, and collapsed.

2.2. Methods

As shown in Figure 1, a building damage classification framework—namely, EBDC-Net—is proposed in this study. EBDC-Net is composed of a feature extraction encoder module and a building damage classification module. First, a building damage feature extraction encoder module was constructed to extract the semantic information of different damage levels. In feature extraction, the spatial attention mechanism (SAM) is used to gather similar features in the image to enhance the feature representation ability of the

network. Second, two parallel modules—global feature extraction (GFE) and contextual feature extraction (CFE)—were included in the building damage classification module to fully exploit the global features and contextual features in the images for building damage classification.

Table 3. Distribution of building damage levels for the three classification criteria.

Group	Description	Damage Level
Group 1	Non-collapsed	L0, L1, L2
	Collapsed	L3
Group 2	Intact	L0, L1
	Severely damaged	L2
	Collapse	L3
Group 3	Intact	L0
	Slightly damaged	L1
	Severely damaged	L2
	Collapse	L3

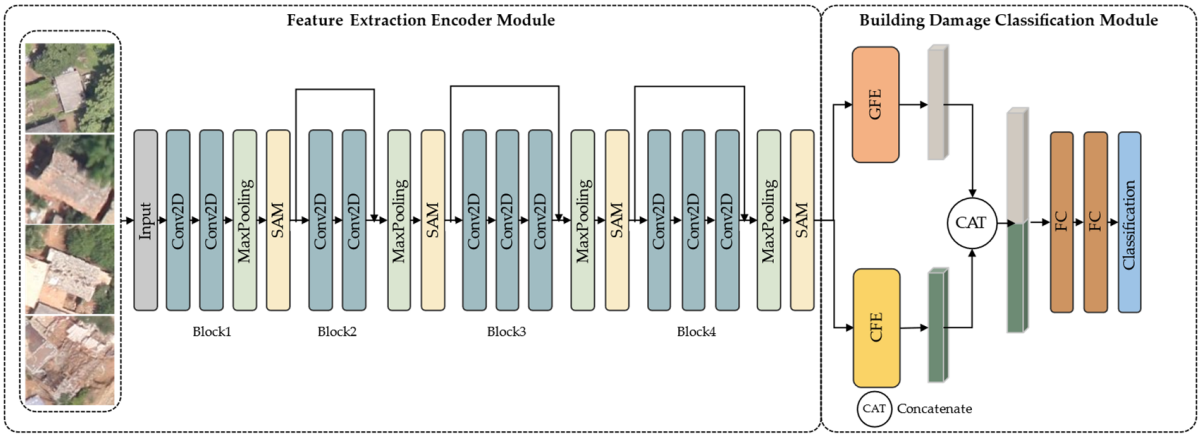


Figure 1. The framework of the proposed EBDC-Net.

2.2.1. Feature Extraction Encoder Module

Extracting useful features from the post-disaster aerial images helps to determine the degree of damage to buildings [19]. As shown in Figure 1, the post-disaster aerial images were used as input data for EBDC-Net. The encoder module of EBDC-Net consists of four convolutional blocks stacked together. In blocks 1 and 2, each convolutional block contains two 2D convolutions (kernel = 3). In blocks 3 and 4, each convolutional block contains three 2D convolutions (kernel = 3). Both convolution operations were followed by a maximum-pooling downsampling operation (kernel = 2). In addition, due to the small size of the input images in this study, there was significant feature loss as the network deepened. The structure of the residual connections can reduce the difficulty of optimization, and enables the training of deeper networks [20]. Therefore, residual connections were added to the last three convolution blocks to alleviate the gradient disappearance and gradient explosion problems during feature extraction.

Although convolutional blocks can extract the semantic information of different levels, in the post-disaster aerial images, the damage features of slightly and severely damaged buildings were scattered in different areas of the images, and the proportion of damage features in the images was small, resulting in a large intraclass variance between different damage categories in the same image. To enhance the feature representation ability of the encoder module and obtain better classification accuracy, a spatial attention mechanism (SAM) was introduced at the end of each convolutional block [21]. The SAM adaptively explores similarities between features at different locations in the image, integrates similar

features at any scale, increases intraclass consistency between different damage classes, and suppresses unwanted information and noise.

As shown in Figure 2, the SAM received the feature maps $F_A \in R^{C\times H\times W}$ extracted from the convolutional block, where C, H, and W represent the channel, height, and width of F_A, respectively. First, F_A was used as an input, and two new feature maps—F_B and F_C—were obtained through two convolutional layers (kernel = 1). The output channels of these two convolutional layers were $C/8$. Second, F_B was reshaped as $R^{N\times(\frac{C}{8})}$, and F_C was reshaped as $R^{(\frac{C}{8})\times N}$, where $N = H \times W$. Subsequently, F_B and F_C underwent matrix multiplication to generate the feature map $F_{S'} \in R^{N\times N}$. Finally, F_S was fed into the softmax layer, and the attention weight map $F_S \in R^{N\times N}$ was generated.

$$F_{Sji} = \frac{\exp(F_{Bi}\cdot F_{Cj})}{\sum_{i=1}^{N}\exp(F_{Bi}\cdot F_{Cj})} \tag{1}$$

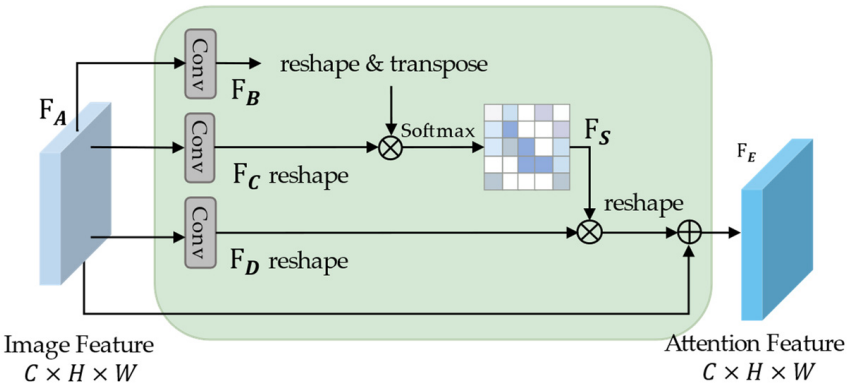


Figure 2. Spatial attention mechanism.

F_{Sji} was used to measure the influence between the features of any two positions in the space. The closer the representation of the features of two positions, the stronger the correlation between them, and the larger the value of F_{Sji}. After obtaining the attention weight map, F_A was fed into a new convolution layer to generate a new feature map F_D. F_D and F_A had the same shape. A matrix was multiplied between the sum space of the attention weight feature map F_S and the feature map F_D, and the result was reshaped as $R^{C\times H\times W}$. Finally, this result was multiplied by a trainable scale factor α and summed with F_A to obtain the final feature map F_E, with α initialized to 0.

$$F_{E_j} = SAM(F_A) = \alpha \sum_{i=1}^{N} F_{Sji}F_{D_i} + F_{A_j} \tag{2}$$

According to Equation (2), the value of each position of F_{E_j} was obtained through the weighted fusion of the original features, with the values in F_S as weights. Therefore, SAM selectively aggregated similar semantic features to improve intraclass compactness and semantic consistency between different damage classes, enabling the network to better distinguish between buildings of different damage classes.

2.2.2. Building Damage Classification Module

When using post-disaster aerial images to classify building damage levels, focusing only on the characteristics of the building itself is not enough to accurately distinguish its damage level. Scenes around buildings can provide necessary auxiliary information for damage assessment. If the global information and contextual dependencies in the images are taken into account, it may improve the final damage classification results. As shown in Figure 1, two parallel modules were designed in the building damage classification module to capture the global information and contextual feature dependencies in the

images, respectively. The feature F extracted by the feature extraction encoder module was used as the input to the building damage classification network. Specifically, F was first fed into the GFE module, where the global feature vector F_G of the image was extracted using a global-level pooling layer. Then, F was fed into the CFE module. In the CFE module, the long short-term memory (LSTM) layer [22] was used to extract the contextual feature dependencies F_C in the image.

As a deep regression neural network, LSTM can handle long-term relationships of memory sequence information [22]. Many remote sensing image classification studies use LSTM to extract spatial and spectral features from images [23–25]. In this study, LSTM was used to explore the contextual dependencies between different regional feature sequences. The generation of feature sequences is the key to learning contextual feature dependencies. As shown in Figure 3, $F \in R^{C \times H \times W}$ was transformed into a feature sequence $V = [x_1, x_2, \cdots, x_K] \in R^C$, where $K = H \times W$. Each C-dimensional feature vector x_k was fed into the LSTM sequentially as a feature sequence.

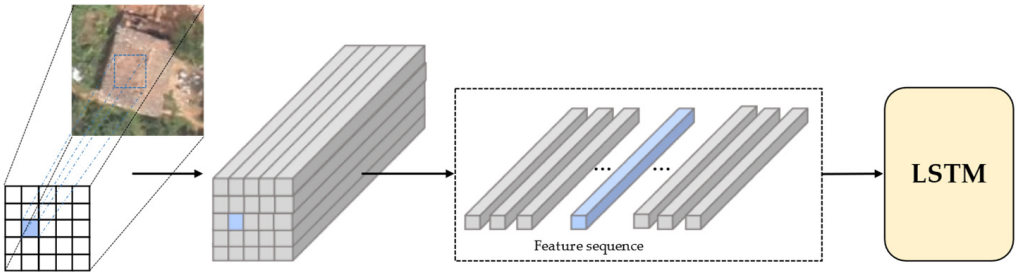


Figure 3. Feature sequence generation.

As shown in Figure 4, LSTM has three inputs: the input value x_k of the current feature sequence, the output value h_{k-1} of the previous feature sequence, and its cell state C_{k-1}. LSTM has two outputs: the output value h_k of the current feature sequence, and its cell state C_k. The forgetting gate f_k combines h_{k-1} and x_k to determine how much of the cell state C_{k-1} of the previous feature sequence is retained in the current feature sequence. The input gate i_k combines h_{k-1} and x_k to determine how much of the input $\widetilde{C}_k$ of the current feature sequence is preserved in the new cell state C_k. The output gate o_k combines the cell state C_k of the current feature sequence to control the current output value h_k. Based on these components, the storage cells and their outputs can be computed as follows:

$$f_k = \sigma\left(W_f[h_{k-1}, x_k] + b_f\right) \tag{3}$$

$$i_k = \sigma(W_i[h_{k-1}, x_k] + b_i) \tag{4}$$

$$\widetilde{C}_k = \tanh(W_C[h_{k-1}, x_k] + b_C) \tag{5}$$

$$C_k = f_k \circ C_{k-1} + i_k \circ \widetilde{C}_k \tag{6}$$

$$o_k = \sigma(W_o[h_{k-1}, x_k] + b_o) \tag{7}$$

$$h_k = o_k \circ tanh(C_k) \tag{8}$$

where σ represents the sigmoid function, $'\circ'$ is the Hadamard product, and W_f, W_i, W_C, W_0, b_f, b_i, b_C, and b_o are learnable weights. The output state of the last feature sequence was used as the contextual feature F_C to describe the contextual feature dependencies in the image. In the contextual feature extraction module, two LSTM layers were stacked, and the output dimensions were set to 256.

After obtaining the global feature F_G and the contextual feature F_C, the two features were fed into the concatenate layer for effective connection to obtain the fusion feature F_{fusion}. Then, F_{fusion} was fed into two fully connected layers of 256 dimensions. It is worth

noting that the dropout strategy was used to avoid overfitting. Finally, the features were processed using the softmax function to output the damage level of the building.

$$F_{fusion} = [F_G, F_C] \tag{9}$$

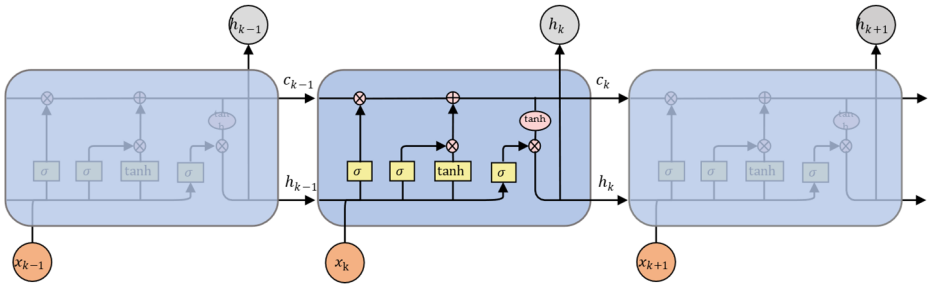


Figure 4. LSTM processing cell.

3. Results

3.1. Implementation Details of the Experiment

All of the experiments and tests in this study were conducted on the same platform, configured with 32 GB RAM, an i7 9800X @3.8 GHz CPU, and a GeForce RTX 2080 Ti GPU. The ratio of the training set, validation set, and test set in the Ludian dataset was 8:1:1, respectively. Since the numbers of images in the Yushu and Yangbi datasets are smaller, the ratio of the training set, validation set, and test set was 6:2:2, respectively.

To obtain the optimal hyperparameters, different batch sizes and learning rates were tested individually, where one of the hyperparameters was fixed.

As shown in Tables 4 and 5, the highest accuracy of the model was achieved when the learning rate was 0.0001 and the batch size was 32. Meanwhile, as shown in Figure 5, the model converged when it was trained for 100 epochs.

Table 4. The effects of different batch sizes on model classification accuracy.

Test	OA (%)	Kappa	MSE
LR-0001-BS-8	68.39	0.56	0.44
LR-0001-BS-16	75.00	0.65	0.28
LR-0001-BS-32	77.49	0.69	0.26
LR-0001-BS-64	75.96	0.67	0.28

Table 5. The effects of different learning rates on model classification accuracy.

Test	OA (%)	Kappa	MSE
LR-001-BS-32	29.02	0.06	0.93
LR-0001-BS-32	77.49	0.69	0.26
LR-00001-BS-32	68.58	0.57	0.40

The model was trained using the SGD optimizer; the weight decay was 0.001, and the cross-entropy loss function was used. The pre-training weight on ImageNet was used to initialize the feature extraction encoder network. Horizontal and vertical flips were used for data enhancement during training. To ensure fairness, the parameters of all of the comparison methods were the same as those of the proposed method. In this study, the overall accuracy (OA), kappa coefficient, and mean square error (MSE) were used as indicators to evaluate the classification accuracy of the model.

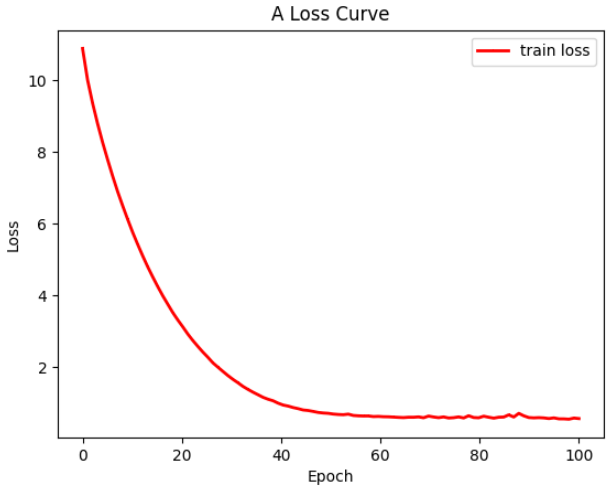


Figure 5. Loss value curve.

3.2. Results of the Comparison of Different Baseline Models

We first compared the performance of different baseline models in the Ludian and Yushu datasets. The classification results for the Yangbi dataset are not presented because there were no collapsed buildings, meaning that it was not possible to divide the buildings into three groups according to the aforementioned grouping criteria. Therefore, the Yangbi dataset was used to discuss the performance of the fine-tuned model. The baseline was constructed by removing the residual connections, SAM module, and CFE module from EBDC-Net. The performance of seven baseline models was compared in Group 1, Group 2, and Group 3, including DenseNet [26], ResNet50 [20], InceptionV3 [27], Xception [28], MobileNet [29], VGG16 [30], and baseline. Tables 6 and 7 show the quantitative comparison of the building damage classification accuracy of different baseline models in the Ludian and Yushu datasets, respectively. As can be seen from Table 6, all of the seven baselines exhibited similar performance for Group 1, with an overall accuracy higher than 90%, due to its relatively simple classification criterion. However, with the increase in the number of building damage categories, the differences in performance between the different models became greater—especially for Group 3, where the classification accuracy dropped dramatically. Among all of the baseline models, the baseline used in this study performed the best on OA, kappa, and MSE in the three groups of the Ludian dataset.

Table 6. The quantitative comparison of different baseline models for classification of building damage in the Ludian dataset.

Model Name	Group 1			Group 2			Group 3		
	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE
DenseNet	91.09	0.76	0.09	79.21	0.65	0.27	69.63	0.58	0.43
ResNet50	92.52	0.79	0.07	81.32	0.68	0.20	71.26	0.61	0.37
InceptionV3	92.62	0.79	0.07	81.41	0.68	0.23	72.99	0.63	0.35
Xception	92.43	0.79	0.08	79.12	0.64	0.26	65.80	0.53	0.43
MobileNet	92.81	0.80	0.07	80.46	0.66	0.24	72.22	0.62	0.33
VGG16	93.29	0.81	0.06	81.61	0.68	0.22	73.37	0.63	0.30
Baseline	**93.39**	**0.82**	**0.06**	**83.52**	**0.72**	**0.18**	**74.23**	**0.64**	**0.30**

The bold font indicates the best accuracy of each indicator.

A similar conclusion can be drawn in the Yushu dataset. As shown in Table 7, for Group 1, all of the baseline models exhibited excellent performance, with an overall accuracy higher than 90%. VGG16 showed the best OA in Group 1, which was 0.42% higher than that of the adopted baseline model. For Groups 2 and 3, the best OA was obtained using the adopted baseline model, which was 0.85% and 1.28% higher than that of VGG-16.

Table 7. The quantitative comparison of different baseline models for classification of building damage in the Yushu dataset.

Model Name	Group 1			Group 2			Group 3		
	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE
DenseNet	92.44	0.82	0.07	76.32	0.60	0.38	63.34	0.50	0.66
ResNet50	93.58	0.85	0.06	76.03	0.61	0.32	63.20	0.50	0.60
InceptionV3	92.58	0.83	0.07	75.46	0.58	0.35	63.62	0.51	0.58
Xception	92.86	0.84	0.07	76.46	0.61	0.29	61.91	0.48	0.59
MobileNet	92.87	0.84	0.07	75.19	0.58	0.39	64.05	0.51	0.59
VGG16	**93.72**	**0.86**	**0.06**	76.18	0.63	0.27	63.91	0.51	0.52
Baseline	93.30	0.85	0.07	**77.03**	**0.64**	**0.26**	**65.19**	**0.53**	**0.47**

The bold font indicates the best accuracy of each indicator.

3.3. Results of Ablation Experiments

In this paper, ablation experiments were performed to demonstrate the contribution of different modules in EBDC-Net to the classification of building damage, where R represents the residual connections, S represents the SAM module, and C represents the CFE module. Tables 8 and 9 show the comparison of the ablation experiments in the Ludian and Yushu datasets, respectively. Compared with the baseline, when the residual connections, SAM module, and CFE module were all added to the model, it showed the highest overall accuracy for the three groups of the two datasets. Compared with the baseline, the OA of EBDC-Net improved by 1.05% and 1.42% for Group 1, 2.01% and 1.99% for Group 2, and 3.26% and 2.43% for Group 3, in the Ludian and Yushu datasets, respectively.

Table 8. The effects of different modules in EBDC-Net on the accuracy of building damage classification in the Ludian dataset.

Model Name	Group 1			Group 2			Group 3		
	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE
Baseline	93.39	0.82	0.07	83.52	0.72	0.18	74.23	0.64	0.30
Baseline +R	93.58	0.82	0.06	83.81	0.72	0.18	75.19	0.66	0.27
Baseline +R+P	93.67	0.82	0.06	84.10	0.72	0.17	76.14	0.67	0.27
Baseline +R+P+C	**94.44**	**0.83**	**0.06**	**85.53**	**0.75**	**0.17**	**77.49**	**0.69**	**0.26**

The bold font indicates the best accuracy of each indicator.

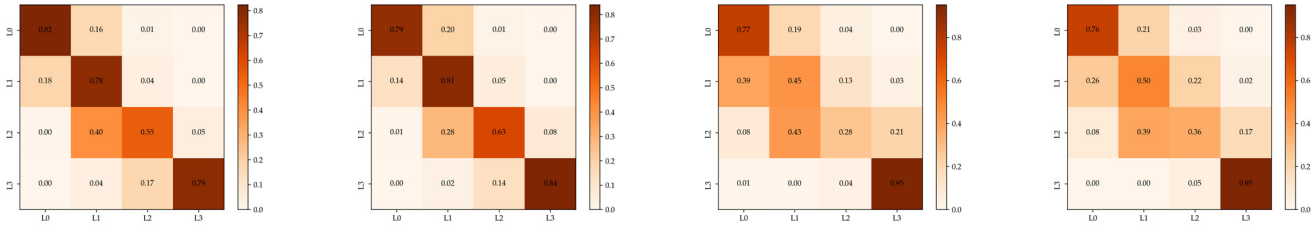
Table 9. The effects of different modules in EBDC-Net on the accuracy of building damage classification in the Yushu dataset.

Model Name	Group 1			Group 2			Group 3		
	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE
Baseline	93.30	0.85	0.07	77.03	0.64	0.26	65.19	0.53	0.47
Baseline +R	93.58	0.86	0.06	78.60	0.64	0.27	65.48	0.53	0.47
Baseline +R+P	93.86	0.86	0.06	78.74	0.65	0.24	66.48	0.66	0.45
Baseline +R+P+C	**94.72**	**0.88**	**0.05**	**79.02**	**0.65**	**0.26**	**67.62**	**0.56**	**0.42**

The bold font indicates the best accuracy of each indicator.

These results indicate that EBDC-Net showed more significant advantages in building damage classification tasks where the categories were more finely divided. This is because the residual connections mitigated the loss of small features as the network deepened. Second, SAM enhanced the representation of damage features in the images, and improved the network's ability to distinguish between intermediate damage classes. Finally, combining global and contextual features of the images improved the classification accuracy. Figure 6 is the confusion matrix between the baseline and EBDC-Net in the Ludian and Yushu datasets. It can be concluded that EBDC-Net is better able to distinguish between buildings

with different levels of damage. Thus, EBDC-Net helps in the fine-grained assessment of building damage.



(**a**) (**b**) (**c**) (**d**)

Figure 6. Confusion matrix in Group 3: (**a**) baseline in the Ludian dataset; (**b**) EBDC-Net in the Ludian dataset; (**c**) baseline in the Yushu dataset; (**d**) EBDC-Net in the Yushu dataset.

3.4. Results of Comparison with Different Building Damage Classification Methods

To verify the effectiveness of EBDC-Net in the classification of building damage, we compared EBDC-Net with four different building damage classification methods. Res-CNN is a model constructed using the CBR module and residual connection [11]. Dense-CNN is a CNN model constructed with dense blocks [13]. The full connection layer in VGG-GAP is replaced by the global average pooling layer [14]. VGG-OR combines the CNN with ordinal regression [16]. As shown in Tables 10 and 11, the EBDC-Net framework proposed in this study showed the best performance in all three groups. The OA was 94.44% and 94.72% in Group 1, 85.33% and 79.02% in Group 2, and 77.49% and 67.62% in Group 3, respectively. Compared to the other four methods, EBDC-Net had a more significant advantage over Group 3 than Groups 1 and 2, with an overall accuracy of 13.5% and 9.42% higher than Re-CNN, 8.34% and 8.13% higher than Dense-CNN, 4.22% and 3.9% higher than VGG-GAP, and 1.92% and 2.57% higher than VGG-OR, respectively.

Table 10. Comparison of different methods for the task of classifying building damage in the Ludian dataset.

Model Name	Group 1			Group 2			Group 3		
	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE
Res-CNN [11]	89.27	0.68	0.10	75.86	0.54	0.37	63.99	0.50	0.49
Dense-CNN [13]	89.08	0.71	0.11	77.68	0.61	0.31	69.15	0.57	0.38
VGG-GAP [14]	93.30	0.81	0.06	83.14	0.71	0.19	73.27	0.64	0.32
VGG-OR [16]	93.29	0.81	0.06	84.58	0.74	0.17	75.57	0.66	0.26
EBDC-Net	**94.44**	**0.83**	**0.06**	**85.53**	**0.75**	**0.17**	**77.49**	**0.69**	**0.26**

The bold font indicates the best accuracy of each indicator.

Table 11. Comparison of different methods for the task of classifying building damage in the Yushu dataset.

Model Name	Group 1			Group 2			Group 3		
	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE	OA (%)	Kappa	MSE
Res-CNN [11]	91.44	0.91	0.09	75.32	0.58	0.37	58.20	0.43	0.64
Dense-CNN [13]	92.15	0.83	0.08	76.03	0.60	0.32	59.49	0.44	0.74
VGG-GAP [14]	94.00	0.87	0.06	78.89	0.67	0.23	63.77	0.51	0.56
VGG-OR [16]	93.30	0.85	0.07	77.75	0.65	0.26	65.05	0.53	0.43
EBDC-Net	**94.72**	**0.88**	**0.05**	**79.02**	**0.65**	**0.26**	**67.62**	**0.56**	**0.42**

The bold font indicates the best accuracy of each indicator.

As shown in Table 12, there was a small amount of debris around the intact buildings in the first and second images, while the buildings in the seventh and eighth images were buried by large debris, and none of their roofs showed significant damage. EBDC-Net enhanced the model's ability to distinguish between texture information and spatial

structure around the buildings by combining global and contextual features. The third and sixth images correspond to slightly damaged and severely damaged buildings, respectively. In both damage classes, the main body of the building was intact, and the damage to the building was scattered across the roof. SAM can aggregate similar features in images, enhancing the network's feature representation, and helping to distinguish buildings in intermediate damage categories.

Table 12. Comparison of different methods in the refined assessment of building damage (L0: intact; L1: slightly damaged; L2: severely damaged; L3: collapsed).

Classification Results	**Images**							
Ground Truth	L0	L0	L1	L1	L2	L2	L3	L3
Dense-CNN	L1	L1	L2	L2	L2	L3	L3	L3
VGG-GAP	L0	L2	L2	L2	L3	L2	L3	L3
VGG-OR	L0	L1	L2	L2	L2	L2	L2	L2
EBDC-Net	L0	L0	L1	L1	L2	L2	L3	L3

4. Discussion

In the building damage classification task, the model learned the damage characteristics of buildings from historical earthquake data. After the earthquake, fine-tuning the model with new data helped to quickly and accurately assess the building damage levels. In this study, three experiments were designed to explore the role of historical data in the post-earthquake building damage assessment task. In Test 1, EBDC-Net was trained using the Ludian dataset and predicted in the Yushu dataset. In Test 2, EBDC-Net was trained and predicted in the Yushu dataset. In Test 3, the Yushu dataset was used to fine-tune EBDC-Net, which was trained in the Ludian dataset.

As shown in Table 13, for Test 1, the OA of the model was 88.30% in Group 1, 69.19% in Group 2, and 56.63% in Group 3. Compared with the results in the Ludian dataset, the classification accuracy of the model in Groups 2 and 3 decreased sharply. As shown in Figure 7, the structure, shape, and style of buildings in the two datasets were very different. The features learned by the model from the Ludian dataset were not enough to represent the features of damaged buildings in the Yushu dataset, leading to low classification accuracy.

Table 13. Impact of historical data on the refined assessment of building damage.

Test Name	**Group 1**			**Group 2**			**Group 3**		
	OA (%)	**Kappa**	**MSE**	**OA (%)**	**Kappa**	**MSE**	**OA (%)**	**Kappa**	**MSE**
Test1	88.30	0.75	0.12	69.19	0.52	0.39	56.63	0.41	0.76
Test 2	94.72	0.88	0.05	79.02	0.65	0.26	67.62	0.56	0.42
Test 3	**95.86**	**0.91**	**0.04**	**80.02**	**0.68**	**0.22**	**68.33**	**0.57**	**0.43**

The bold font indicates the best accuracy of each indicator.

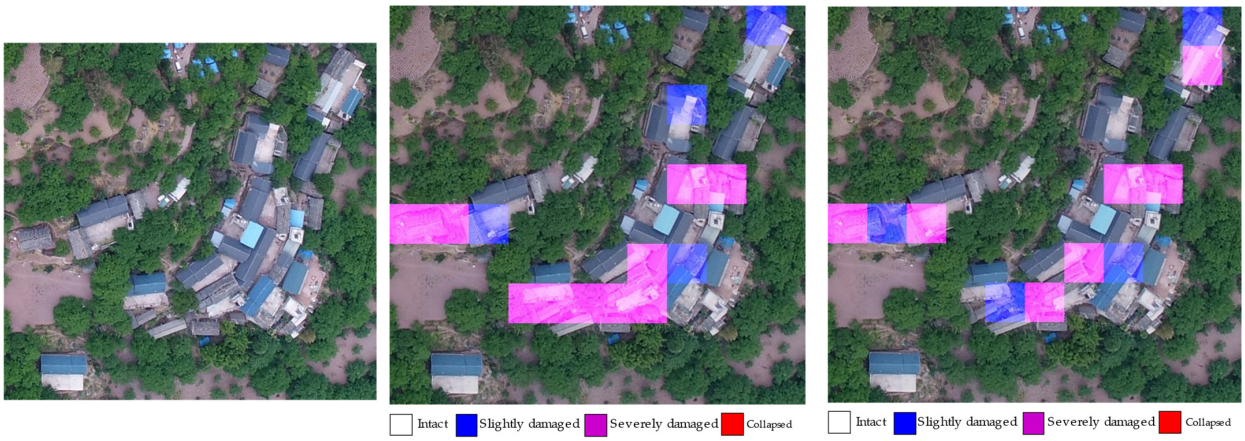
(**a**) (**b**)

Figure 7. Examples of buildings in different areas: (**a**) Ludian dataset; (**b**) Yushu dataset.

Similarly, in Test 2, the OA of the model in Group 2 and Group 3 was 79.02% and 67.62%, respectively, which was lower than the corresponding accuracy in the Ludian dataset. The reason for this is that there were much smaller scales of Yushu dataset images than Ludian dataset images. It was also shown that in the refined assessment of building damage, the number of samples can have a significant impact on the accuracy of the assessment.

However, the accuracy of the model was improved dramatically when the network trained using Ludian images was fine-tuned using a small additional amount of Yushu images. As shown in Table 11, the OA was 95.86%, 80.82%, and 68.33% for the three groups, respectively, which was 7.56% and 1.14% higher for Group 1, 10.83% and 1% for Group 2, and 11.7% and 0.71% for Group 3, compared to Tests 1 and 2, respectively. This is because the historical earthquake data can provide the basic features of the damaged buildings. By adding a small number of images from the testing area, more detailed and local features can be learned, bringing about the improvement of classification accuracy. This indicates that fine-tuning is an effective strategy for the classification of building damage

Through the visual qualitative analysis of the prediction results of some areas, we can intuitively understand the model through the assessment of building damage. In this study, we divided the images into patches, rather than segmentations of individual buildings. Therefore, a patch may contain several buildings. When a building was cropped into two or more patches, the damaged features of the building were retained in the corresponding patches. The EBDC-Net model trained using the Ludian dataset was fine-tuned using the Yangbi dataset. Figure 8 shows the visualization results of the model evaluation and the visual interpretation results (ground truth). The model evaluation results were generally consistent with the visual interpretation results, with an overall accuracy of 75%, a kappa of 0.66, and an MSE of 0.26. In addition, the time needed for UAV image evaluation was tested. The results show that the average processing time for an image of 5474 × 3648 pixels was 19 s.



(**a**) (**b**) (**c**)

Figure 8. Example of assessment results of building damage in the Yangbi dataset: (**a**) original image; (**b**) EBDC-Net assessment results; (**c**) visual interpretation results.

5. Conclusions

In this work, we propose a novel network called EBDC-Net to solve the finer classification problem of damaged buildings after earthquakes. The proposed method was tested using two datasets and compared with four state-of-the-art methods. In addition, the roles of the residual connection, spatial attention mechanism, and contextual feature extraction module were also explored. The experimental results demonstrated the following: (1) in the Ludian and Yushu datasets, the accuracy of the proposed method was at least 1.92% and 2.57% higher compared to state-of-the-art building damage classification methods; (2) with the introduction of the above three strategies, the classification accuracy was improved by 3.26% and 2.43% in the Ludian and Yushu datasets, respectively, compared to the baseline

model; and (3) using the historical earthquake data and the fine-tuned model is a good strategy to quickly classify the buildings damaged in the new earthquake.

The main contributions of this paper can be summarized as follows:

(1) We propose a novel deep-learning-based model to solve the fine-grained classification problem of damaged buildings, which is critical to earthquake rescue and post-disaster damage assessment.
(2) The spatial attention mechanism and the contextual feature extraction module are embedded in EBDC-Net, which can improve the model's ability to classify buildings with different levels of damage.

In the future, we will try to explore the classification of building damage under complex conditions through the use of multimodal and multi-temporal remote sensing images.

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

Funding: This work was funded by the National Key R&D Program of China, grant number 2018YFB0505400, 2017YFC1500906; the National Natural Science Foundation of China, grant number 41871325, 42061073; and the Natural Science and Technology Foundation of Guizhou Province under Grant [2020]1Z056.

Data Availability Statement: The Ludian and Yushu datasets are freely available online, and can be found at https://github.com/city292/build_assessment (accessed on 1 April 2022).

Acknowledgments: We thank Ci et al. for providing a free aerial image dataset to the entire scientific community so that our work could be successfully completed.

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

References

1. Taşkin, G.; Erten, E.; Alataş, E.O. A Review on Multi-temporal Earthquake Damage Assessment Using Satellite Images. In *Change Detection and Image Time Series Analysis 2: Supervised Methods*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2021; pp. 155–221. [CrossRef]
2. Huang, H.; Sun, G.; Zhang, X.; Hao, Y.; Zhang, A.; Ren, J.; Ma, H. Combined multiscale segmentation convolutional neural network for rapid damage mapping from postearthquake very high-resolution images. *J. Appl. Remote Sens.* **2019**, *13*, 022007. [CrossRef]
3. Liu, X.; Deng, Z.; Yang, Y. Recent progress in semantic image segmentation. *Artif. Intell. Rev.* **2019**, *52*, 1089–1106. [CrossRef]
4. Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A.; Zhang, L. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. *Remote Sens. Environ.* **2021**, *265*, 112636. [CrossRef]
5. Wu, C.; Zhang, F.; Xia, J.; Xu, Y.; Li, G.; Xie, J.; Du, Z.; Liu, R. Building damage detection using U-Net with attention mechanism from pre-and post-disaster remote sensing datasets. *Remote Sens.* **2021**, *13*, 905. [CrossRef]
6. Xiao, H.; Peng, Y.; Tan, H.; Li, P. Dynamic Cross Fusion Network for Building-Based Damage Assessment. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6. [CrossRef]
7. Adriano, B.; Yokoya, N.; Xia, J.; Miura, H.; Liu, W.; Matsuoka, M.; Koshimura, S. Learning from multimodal and multitemporal earth observation data for building damage mapping. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 132–143. [CrossRef]
8. 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]
9. Song, D.; Tan, X.; Wang, B.; Zhang, L.; Shan, X.; Cui, J. Integration of super-pixel segmentation and deep-learning methods for evaluating earthquake-damaged buildings using single-phase remote sensing imagery. *Int. J. Remote Sens.* **2020**, *41*, 1040–1066. [CrossRef]
10. Yang, W.; Zhang, X.; Luo, P. Transferability of convolutional neural network models for identifying damaged buildings due to earthquake. *Remote Sens.* **2021**, *13*, 504. [CrossRef]
11. Duarte, D.; Nex, F.; Kerle, N.; Vosselman, G. Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. *Remote Sens.* **2018**, *10*, 1636. [CrossRef]

12. Ji, M.; Liu, L.; Zhang, R.F.; Buchroithner, M. Discrimination of earthquake-induced building destruction from space using a pretrained CNN model. *Appl. Sci.* **2020**, *10*, 602. [CrossRef]
13. 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]
14. Ishraq, A.; Lima, A.A.; Kabir, M.M.; Rahman, M.S.; Mridha, M. Assessment of Building Damage on Post-Hurricane Satellite Imagery using improved CNN. In Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 23–25 March 2022; pp. 665–669. [CrossRef]
15. Cao, C.; Liu, D.; Singh, R.P.; Zheng, S.; Tian, R.; Tian, H. Integrated detection and analysis of earthquake disaster information using airborne data. *Geomat. Nat. Hazards Risk* **2016**, *7*, 1099–1128. [CrossRef]
16. 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]
17. Ma, H.; Liu, Y.; Ren, Y.; Wang, D.; Yu, L.; Yu, J. Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images. *Remote Sens.* **2020**, *12*, 260. [CrossRef]
18. Matin, S.S.; Pradhan, B. Challenges and limitations of earthquake-induced building damage mapping techniques using remote sensing images-A systematic review. *Geocarto Int.* **2021**, 1–27. [CrossRef]
19. Guo, H.; Shi, Q.; Du, B.; Zhang, L.; Wang, D.; Ding, H. Scene-driven multitask parallel attention network for building extraction in high-resolution remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4287–4306. [CrossRef]
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
21. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154. [CrossRef]
22. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
23. Zhou, F.; Hang, R.; Liu, Q.; Yuan, X. Hyperspectral image classification using spectral-spatial LSTMs. *Neurocomputing* **2019**, *328*, 39–47. [CrossRef]
24. Yin, J.; Qi, C.; Chen, Q.; Qu, J. Spatial-spectral network for hyperspectral image classification: A 3-D CNN and Bi-LSTM framework. *Remote Sens.* **2021**, *13*, 2353. [CrossRef]
25. Liu, Q.; Zhou, F.; Hang, R.; Yuan, X. Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1330. [CrossRef]
26. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. [CrossRef]
27. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]
28. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]
29. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
30. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]