

Remote Sensing Building Damage Assessment Based on Machine Learning

Jiawei Tang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 905462150@qq.com

Shujuan Huang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 349242386@qq.com

Shengquan Yang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: xaitys@126.com

Bozhi Xiao

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 3138379859@qq.com

Abstract—After the occurrence of various types of disasters, including natural disasters and man-made damage, aid workers need accurate and timely data, such as the damage status of buildings, in order to take effective measures for rescue. So as to solve this problem, this paper researches and designs a building damage classification system based on machine learning. The damage assessment system consists of two network models (building extraction network and damage classification network). This article analyzes and designs the structure of each network model, and discusses the principles related to computer vision in machine learning. Buildings in satellite images are segmented through Siamese Convolutional Neural Network, the BottleNeck Module and Feature Pyramid Network are used in the damage classification assessment network to detect damage to buildings in sub-temporal remote sensing images. Subsequently, the model was trained and tested on different disaster events on the xBD dataset. The results show that the building damage detection system based on Siamese-CNN achieves good detection accuracy, and the system has the advantages of simple operation, good timeliness and low resource consumption, and can well meet the needs of disaster assessment.

Keywords—Machine Learning; Remote Sensing; Disaster Assessment; Buildings

I. INTRODUCTION

According to data from the United States Geological Survey, more than one million natural disasters occur around the world every year, with an average of several disasters occurring almost every minute [1]. In the last decade alone, the global death toll from natural disasters has exceeded one million. With the advancement of global urbanization, more densely distributed building areas will inevitably lead to further aggravation of the impact of building damage and collapse after disasters. When current technology is insufficient to reliably predict disasters, how to reduce losses through effective emergency response measures becomes an important issue.

Therefore, an important task in the disaster emergency response process is damage assessment of the disaster-stricken area, and the damage information of surface buildings has become important reference information in the disaster relief emergency response process. Obtaining detailed information of disaster areas through manual survey is slow, dangerous, and timely, which is not conducive to obtaining information.

However, by analyzing satellite image data, it is more convenient to obtain the damage situation in a specific area and make decisions without arriving at the scene. Therefore, using satellite images can obtain disaster area information more efficiently.

At present, the mainstream building damage assessment methods are divided into three types: traditional assessment methods, multispectral sensor-based assessment methods, and deep learning-based assessment methods [2]. Traditional assessment methods consume a lot of resources and time, which is not practical in disaster assessment and will not be discussed here. The following focuses on the assessment method based on multispectral sensors and the assessment method based on deep learning for building damage assessment.

The evaluation method based on multispectral sensors generally uses multispectral satellites to capture the spectral information of different materials in multiple bands such as reflection and absorption of visible light and infrared light for detection [3]. At the same time, multispectral satellites can enhance sensitivity to specific ground objects or phenomena. Therefore, the evaluation method based on multispectral sensors is more accurate on some materials. However, because the images transmitted by spectral satellites are very strict, the images need to be preprocessed, denoised, corrected, etc. And these process will take a long time. Therefore, simpler and faster methods for damage assessment of disaster buildings are needed in disaster assessment.

The evaluation method based on machine learning [4,5] uses complex deep neural networks to enable computers to handle corresponding tasks like humans. This field is called computer vision [6,7] in the field of machine learning. Computer vision research mainly includes the following four types of tasks: the first type of task is image classification [8]; the second type of task is image target detection [9]; the third type of task is image semantic segmentation [10-11]; and the fourth type of task is image instance segmentation [12]. In the building damage assessment method based on deep learning, a computer is used to


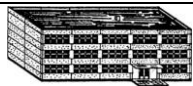








semantically segment the buildings in the image, and then the buildings in the disaster area are classified. Through this method, disaster detection and assessment can be carried out in the disaster-stricken area on the trained deep neural network. Therefore, assessment methods based on machine learning can complete disaster assessment tasks faster and more efficiently.

II. RELATED WORK RESEARCH

A. Building Damage Standards

According to the classification definition of buildings with different degrees of damage by the European Disaster Committee, this definition is also the current reference standard for international research on building damage assessment, as shown in Table 1. According to Table 1, it can be found that since no structural changes have occurred in the damaged structures of Level 1 ‘Undamaged’ and Level 2 ‘Minor Damaged’, ‘Undamaged’ and ‘Minor Damaged’ are usually divided into the same category in building damage assessment algorithms [13].

TABLE I. EUROPEAN DISASTER COMMITTEE TABLE FOR BUILDING DAMAGE ASSESSMENT

Masonry Construction	Fortified Buildings	Damage Level
		Undamaged
		Minor Damaged
		Medium Damaged
		Major Damage
		Destroyed

B. Current status of building damage assessment algorithms

The emergence of machine learning has brought an important turn to the development of the field of building damage assessment based on remote sensing images. Applying computer vision-related technologies in deep learning to building damage assessment can increase the accuracy and efficiency of building damage assessment. In 2016, Researchers integrated neural networks into building disaster detection for the first time, laying the foundation for subsequent building damage assessment through neural network models. In 2018, Duarte et al. [14] constructed the damage evaluation of buildings in remote sensing images as a binary pixel classification problem for the first time. In 2019, Xu et al. [15] compared the performance of different network by evaluate damaged buildings after earthquakes, and trained and tested on different disaster data sets to verify the feasibility of disaster assessment through neural networks. And the same year, Gupta et al. [16] established a large-scale data set for large areas and multiple disasters, providing multi-temporal satellite images before and after disasters. It contains about 700,000 building annotations over 5,000 square kilometers and is mainly used for post-disaster building damage assessment. In 2020, Weber et al. [17] established a semantic segmentation model based on the ResNet50 backbone and trained it through the xBD dataset to achieve graded assessment of building damage. The proposal of the above technologies has opened up a path and space for the research and application of deep learning in the field of building damage assessment, making deep learning faster and more efficient in damage assessment tasks in the field of building damage assessment.

C. Dataset

xBD Dataset [18]: Segmentation, extraction and damage classification of buildings under remote sensing images is a very challenging task due to the complexity of unstructured scenes. Therefore, researchers and research institutions have disclosed many data sets in order to promote the field of disaster assessment. These data sets collect image data using multiple remote sensing

satellites and conduct precise annotations on the images to form accurate label values. This article plans to use the xBD data set commonly used in the field of disaster assessment, as shown in Figure 1.

The xBD dataset is a dataset for building damage assessment released by the Massachusetts Institute of Technology. It is one of the current public datasets of remote sensing images in the world. The data set contains a total of tens of thousands of images, all of which are 1024*1024 high-resolution satellite remote sensing images, marked with 19 different events, including earthquakes, floods, wildfires, volcanic eruptions, etc. These images include pre-disaster and post-disaster images, allowing for adequate research on building damage assessment.



Figure 1. Schematic diagram of xBD dataset

III. TECHNOLOGIES AND NETWORK MODELS USED IN THIS ARTICLE

From an overall perspective, the building damage assessment problem is affected by many aspects such as data sources and evaluation standards. Currently, there is no consistent algorithm framework. This chapter analyzes existing algorithm problems, proposes a set of building damage classification algorithm processes based on Siamese-CNN neural network, and analyzes in detail some of the key technologies involved.

A. Overall design of system

From the perspective of application scenarios, a characteristic of building damage assessment is that it needs to process a large amount of data in a short period of time, which requires high performance of the algorithm. Therefore, the best way is to fully train the model using existing remote sensing image data before and after the

disaster so that it can perform better when processing new data. From the perspective of image classification, high-resolution remote sensing images contain many types of ground objects and have complex background interference, making them more difficult to handle than general natural image classification problems. On the other hand, remote sensing images usually cover a larger area, especially in areas with lower urbanization rates, where the proportion of built-up areas is smaller. Therefore, from the perspective of improving the accuracy of damage assessment and the efficiency of algorithm processing, this paper divides the damage assessment process into two processes: building segmentation extraction and building damage assessment. The figure 2 shows the overall program process of building damage assessment in this article.

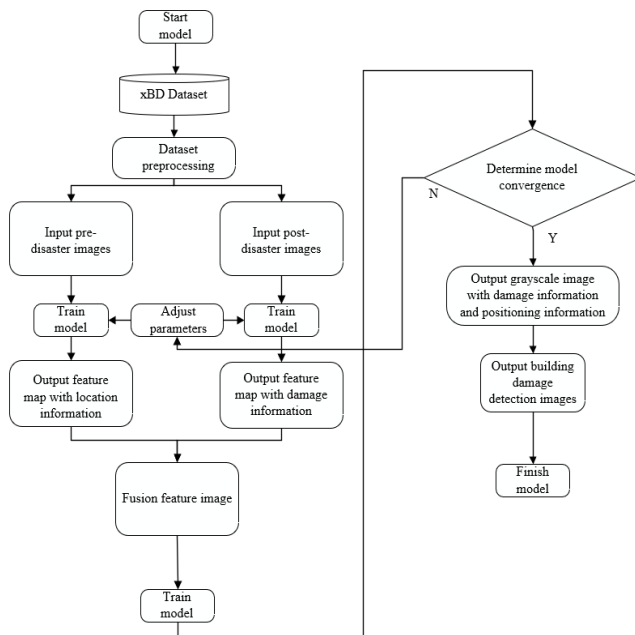


Figure 2. Network model flow chart based on machine learning

B. Key technical analysis

1) *Fully Connected Neural Network basics (FCN):* Fully Connected Neural Networks [19] are the basis of all neural networks. They are a sort of neural network designed from the nervous system of animals and are therefore also known as Artificial Neural Networks. The underlying principle of the fully connected neural network is a mathematical algorithm model for distributed parallel processing. Through the network

complexity of FCN, the interconnection of nodes can be adjusted to achieve the purpose of information processing. Therefore, the fully connected neural network has a good effect in completing the tasks of regression model and classification model. The FCN network structure is shown in Figure 3.

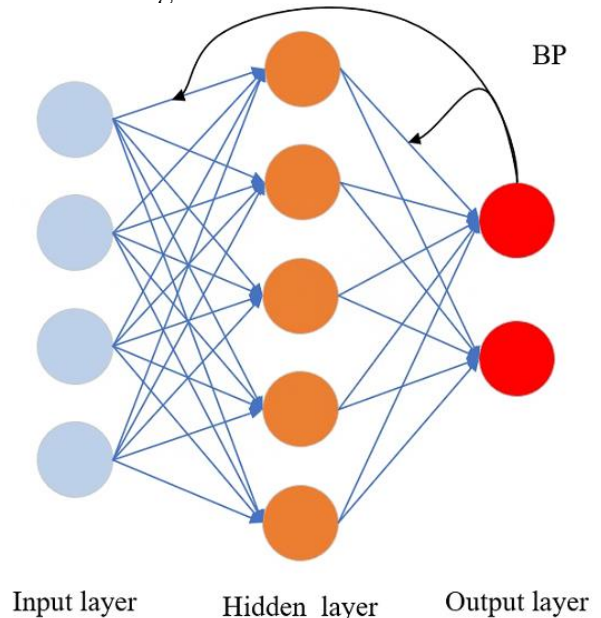


Figure 3. The network structure of the FCN

2) *Convolutional Neural Network basics (CNN):* CNN [20,21] are derived from neuron-based message passing systems in biology. Generally, there are several layers from input to output, each layer contains several neurons, and the neurons in adjacent layers are interconnected. Specific to the field of computational vision, each layer here is a three-dimensional array, and the neuron actually refers to a corresponding item in the three-dimensional array.

The input of the CNN network is the original image. The most important feature of the CNN network is the introduction of alternating convolution layers and pooling layers, which generally obtain the response of a point on the output plane through a small sub-region on the input feature plane. The so-called interconnection of adjacent neurons refers to reorganizing the output of all neurons in a certain layer into a three-dimensional array as the input of the next layer. Generally speaking, as the level of connection increases, the feature plane obtained by the neuron output response decreases, the number of feature

planes increases, and the corresponding receptive field of each neuron response also increases. In general, a convolutional neural network forms a function that obtains an output image from an input image. The basic CNN network structure is shown in Figure 4.

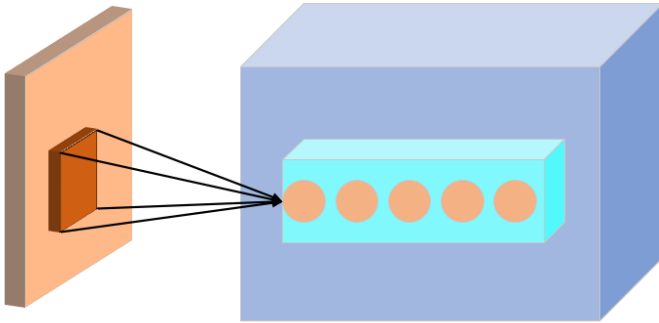


Figure 4. CNN receptive field

a) Convolution layer: The basic operations of convolutional layers [22] are derived from the digital filtering process in traditional image processing. The difference is that the convolution template here is obtained through training and learning, and the template size is generally smaller. All convolution templates in the same convolution layer have the same scale. Only in this way can the outputs of different convolution kernels form feature planes with consistent dimensions. During the forward pass, each convolution kernel is convolved with all feature planes of the input. Specifically, the convolution operation is the dot product of the elements in the filter template and the elements of the receptive field area in the input image. Since the receptive field corresponding to each point on the output feature plane is only a local area in the input plane, the filter template of the convolutional layer learns the local characteristics of the image. The superposition of multiple convolutional layers gradually aggregates these local features into global features that can characterize the entire image.

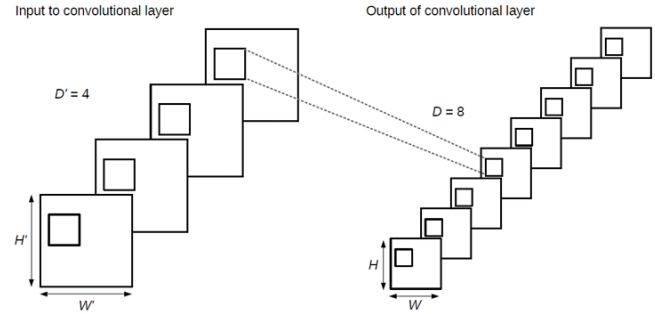


Figure 5. Convolutional layer input and output diagram

b) Activation function: The function of the activation function [23] is to perform a certain nonlinear transformation on the output of each neuron. In a convolutional neural network, the input to the excitation function is generally the output of the convolutional layer. The simplest excitation function is a binary function that outputs 0 or 1 depending on the range of input values. The most commonly used excitation function in early neural networks is the Sigmoid function, which has the following form:

$$f(x) = (1 + \exp(-x))^{-1} \quad (1)$$

This function acts on each element in the three-dimensional vector output by the convolutional layer and does not require a parameter learning process. The most commonly used function in modern neural networks is the *ReLU* function. Specifically, the *ReLU* function outputs 0 for negative values and remains unchanged for positive values. Its form is as follows:

$$f(x) = (0, 1)_{\max} \quad (2)$$

c) Pooling layer: The function of the pooling layer [24] is to perform nonlinear downsampling of the input feature plane. The most commonly used method is Max pooling. Pooling layers generally alternate with convolutional layers (after the activation function), usually reducing the spatial resolution of the input feature plane.

The Max pooling layer is shown in the Figure 6. For selecting a rectangular sub-region in the input feature plane, Max pooling takes the maximum

value of the response in the sub-region as a response on the output feature plane. Simply put, the pooling process is to sample at corresponding intervals and start moving from the sampling center point. In the general convolutional neural network structure, the input feature plane is sampled without overlap. The precise relationship between input and output in a convolutional neural network is as shown in the formula:

$$y_{ijd} = \max\left(\{x_{sxi+s\times j+q,d} | 0 \leq p \leq h-1, 0 \leq q \leq w-1\}\right) \quad (3)$$

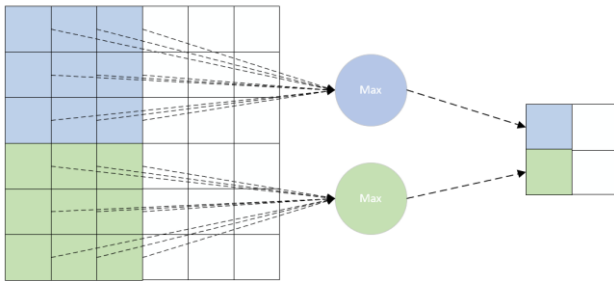


Figure 6. Feature Pyramid Network structure diagram

3) *Siamese Convolutional Neural Network basic (SCN)*: Siamese Convolutional Neural Network [25,26] is a machine learning model used for metric learning. The basic structure of the network contains two identical sub-networks, as shown in Figure 7, which share the same weight parameters. These two sub-networks respectively process two samples to be compared, usually images or other representations of data. Through the design of shared weights, the Siamese network can learn the feature representation of input samples, and the goal of metric learning is to measure the similarity between samples through these feature representations. In the network structure, the output features of two sub-networks are fused together by measuring the similarity or difference between the learning samples in the learning layer. The training process of the network increases the design of a loss function, which encourages similar samples to be closer and dissimilar samples to be further away. Through the backpropagation algorithm, the parameters of the network are adjusted so that it can learn appropriate feature representations and similarity measures. Twin CNN networks are widely used in fields such as face verification, signature verification, target tracking, etc. Its advantage is

that it can learn the similarities between samples and has strong generalization performance.

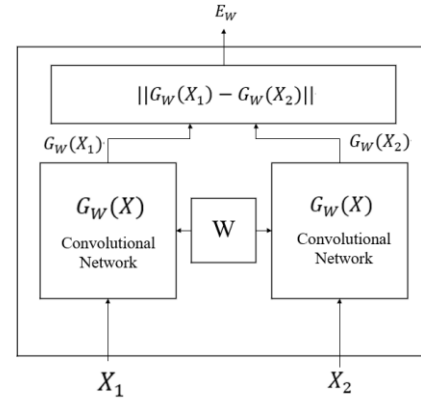


Figure 7. Basic structure diagram of Siamese Convolutional Neural Network

C. Building Damage Classification Network Based On Siamese-CNN

This article intends to design a building damage assessment algorithm based on Siamese-CNN. Siamese-CNN is an algorithm in deep learning. It is a new neural network developed based on CNN. Through Siamese-CNN neural network, features in the image can be extracted and classified based on the features to achieve the purpose of classifying building damage. In the following, the Siamese-CNN network model designed in this article will be focused on.

The Siamese-CNN network model has an encoder-decoder structure. In the encoder module, the images before the disaster and the images after the disaster are input into the convolutional Siamese convolutional neural network, and the images are converted into a series of features. Then entering the decoder module, this article integrates the feature pyramid network and BottleNeck module into the decoder module. This can preserve the features in the input image more abundantly, and can enhance the expression ability of the features, making Siamese-CNN perform better in building damage assessment tasks. Next, we will introduce the feature pyramid network and BottleNeck module used in this article in detail.

Feature Pyramid Networks [27] obtains corresponding feature images by using the multi-scale pyramid structure of the deep convolutional neural network itself. Its core idea is to use multi-

scale feature maps to improve the model's ability to identify different targets. The FPN network structure used in this article is shown in Figure 8. The left side is the bottom-up path, which mainly builds a forward pyramid and extracts feature information at different scales through the special structure of the pyramid. Specifically, after FPN receives multiple feature maps output by the encoder, it immediately performs horizontal convolution. The purpose is to match the number of channels of the bottom feature map to the same number of channels as the top feature image layer by layer for subsequent processing. Then the feature images are upsampled in a top-down path to ensure that the dimensions and sizes of all upper-layer feature images are consistent with the bottom-layer feature images; finally, fusion is performed to obtain a fused multi-scale feature map.

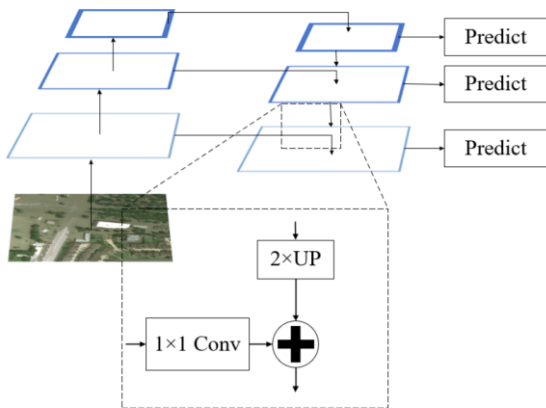


Figure 8. Feature Pyramid Network structure diagram

The BottleNeck module [28] is a widely used structure in deep learning and is often used in ResNet networks. The BottleNeck module can reduce the amount of parameters and calculations when processing larger input data or more complex networks, and can effectively improve network performance. The structure of the BottleNeck module used in this article is shown in Figure 9. It mainly contains three important parts: The first part is dimensionality reduction convolution, which reduces the number of channels and reduces the amount of calculation by using dimensionality reduction convolution with a convolution kernel of 1; The second part is the middle-layer convolution, The middle-layer

convolution with a convolution kernel of 3 is used to perform non-linear changes on the input feature map, which is beneficial to learning more complex features; The last layer is dimensionality-raising convolution, which restores the number of channels of the feature image to the original level by using specific convolutions. In general, as a part of the Siamese-CNN network, the BottleNeck module not only improves the learning efficiency of the network, but also improves the parameter efficiency of the network, so that the Siamese-CNN network has a good effect in dealing with the task of building damage assessment.

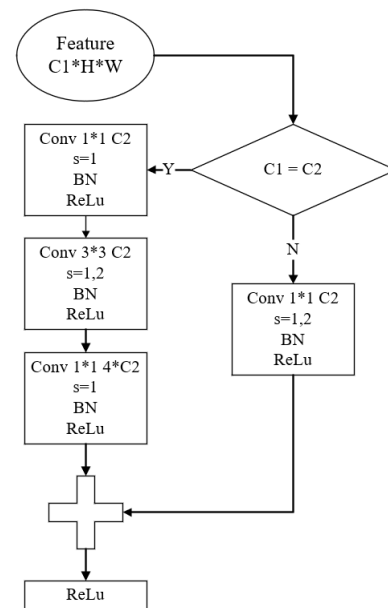


Figure 9. Basic structure of the BottleNeck module

The specific network structure of Siamese-CNN is shown in Figure 10. In the entire network structure, two images, Pre-Damage-Image and Post-Damage-Image, are first input. The Encoder module receives two input images and obtains two sets of feature maps with different resolutions through a combination of multiple rounds of convolution, batch normalization, and activation functions. The first group corresponds to the feature map obtained by Pre-Damage-Image, and the second group corresponds to the feature image obtained by Post-Damage-Image. Taking these feature images as inputs and entering them into Feature Pyramid Networks respectively, FPN will first downsample these feature maps to obtain strong semantic features, then upsample and then

conduct lateral connections to ensure that semantic information can be transferred to feature maps of different scales. After FPN, two feature maps 'Location-FPN-Out' and 'Damage-FPN-Out' are output. Subsequently, the BottleNeck module performed dimension changes, nonlinear transformations and other operations on the 'Location-FPN-Out' and 'Damage-FPN-Out'

feature maps output by FPN to further improve the expressive ability of the feature maps and obtain 'Location-P' and 'Damage-P'. Finally, the feature maps 'Location-P' and 'Damage-P' are resized and fused to obtain the final feature, and then the information of the feature is converted into a mask image as the output of the model.

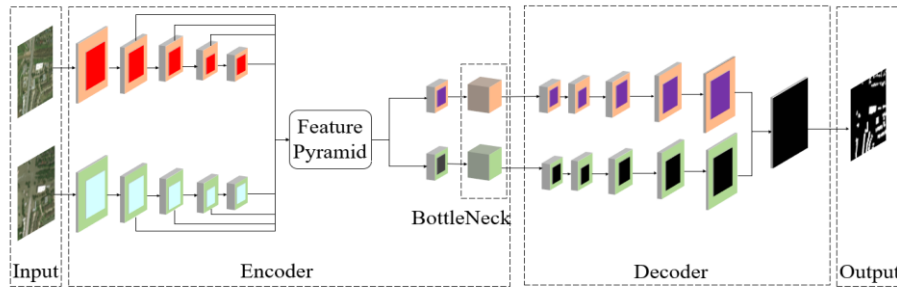


Figure 10. Structure diagram of Siamese-CNN network model

IV. EXPERIMENT AND ANALYSIS

A. Experiment preparation

Based on the disaster level classification table described in the previous section, because of the limitations of remote sensing images, it is difficult to completely distinguish between "Undamaged" and "Minor damaged", so "Minor damaged" is classified as the "Undamaged" level. The building damage classification level table used in this article obtained so far is shown in Table 2.

TABLE II. BASED ON THE BUILDING DAMAGE LEVEL TABLE DEFINED IN THIS ARTICLE

Class	Description
0	Undamaged
1	Minor damage
2	Major damage
3	Destroyed

Since the Encoder stage in Siamese-CNN has a good effect in classifying building damage levels, this article trained a building damage classification network based on Siamese-CNN on the xBD dataset. First, the original data set needs to be preprocessed

accordingly, and the pre-damage-images and post-damage-images must be processed separately to obtain the masked binary image. And adjust the

data set according to the actual situation, and adjust the number of dataset and verification set. Since the area of buildings in satellite images is relatively small, it makes building recognition difficult. Therefore, the data set needs to be preprocessed by random flipping and random segmentation to improve the positioning and classification accuracy of small-area buildings. The image processing is shown in Figure 11. The required training environment configuration is shown in Table 3.

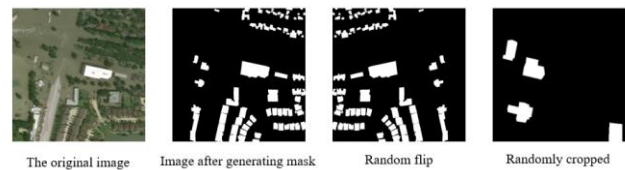


Figure 11. Data processing renderings

TABLE III. TRAINING ENVIRONMENT CONFIGURATION TABLE

Configuration information	Detail
Hardware Configuration	Nvidia RTX 3080 12G
Language	Python 3.8
Main Frame	Pytorch 2.1.0 Cuda11.8
Image information	1024×1024 20248 photos
Optimization Function	Adam
Loss Function	cross entropy loss
Epoch	30
Training time	12h

B. Evaluation metrics

When the model makes predictions, the prediction results for the sample exist in the following situations:

- Predicting positive samples is called True Positive (TP).
- Predicting positive samples is called False Negative (FN).
- Predicting negative samples is called False Positive (FP).
- Predicting negative samples is called True Negative (TN).

From this, the confusion matrix of the model on the validation set can be obtained. The specific form is shown in Table 4.

TABLE IV. CONFUSION MATRIX FORMAL TABLE

Prediction category \ True category	True category	
	Positive sample	Negative sample
Positive sample	TP	FP
Negative sample	FN	TN

- Accuracy: It is the ratio of correctly predicted samples to the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

- Precision: It is the proportion of samples correctly predicted as positive to the total number of samples predicted as positive.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

- Recall: It is the proportion of correctly predicted positive samples to the actual number of positive samples.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

- F₁ Value: In general, precision rate and recall rate affect each other. When the recall rate is high, the precision rate will be very low. In order to ensure that both are high, the F₁ Value is used to measure it. The F₁ Value is essentially the harmonic mean of precision and recall.

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

C. Experimental results

According to the experimental preparation part above, the Siamese-CNN model is trained on the xBD data set. After the training is completed, the training weights are first verified on the xBD validation set, and then the trained network is tested on the xBD test set, F₁ Value is used to evaluate the final performance of the network. The performance results on the validation set are shown in Figure 13, and the results of the evaluation on the xBD test set are shown in Figure 12. (The Table 5 is a supplement and explanation to Figures 12 and 13). It can be observed from the table species data that the model has good results in segmenting buildings and classifying undamaged buildings and collapsed buildings; its performance in classifying lightly damaged buildings is poor. This may be related to the small proportion of lightly damaged buildings in the xBD dataset.

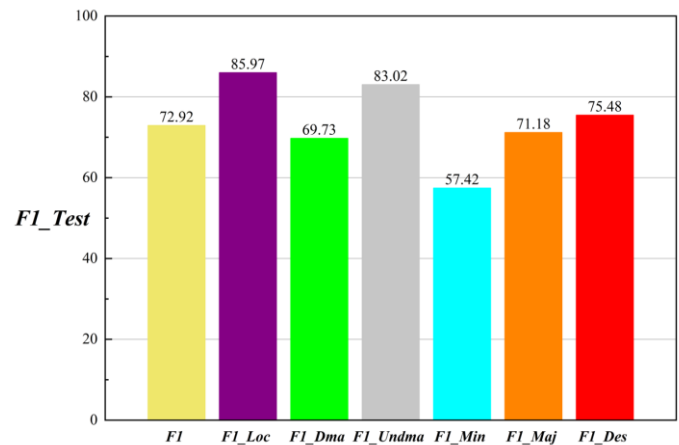


Figure 12. The F1 Value evaluation results on test set

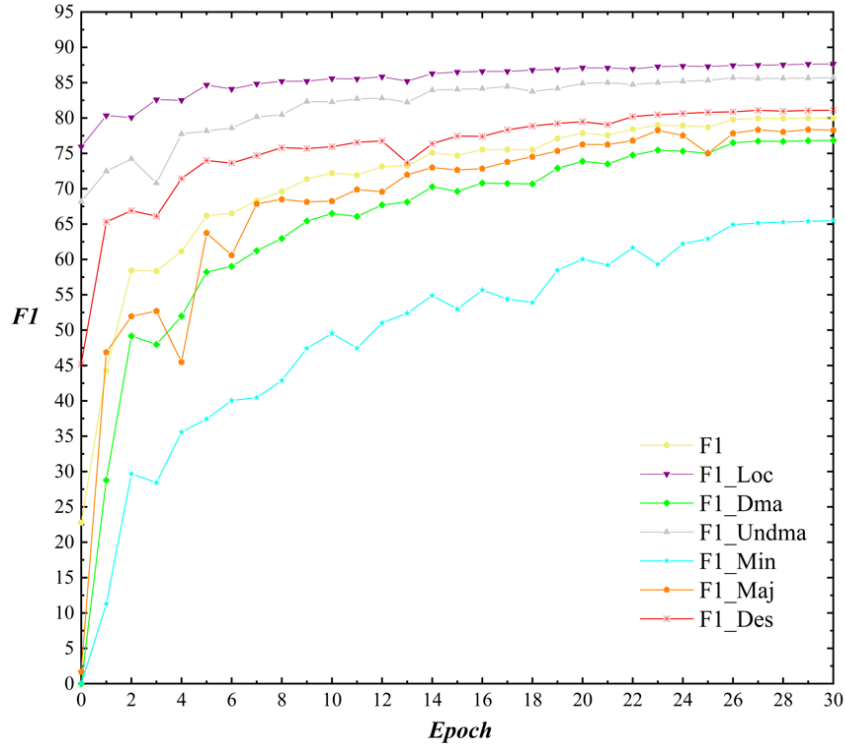


Figure 13. Training results on validation dataset

TABLE V. TRAINING RESULTS ON VALIDATION DATASET

Name	Explanation	Color
F1	The overall F1 value of the building damage assessment on the xBD validation set	Yellow
F1_Loc	F1 values for segmentation of building localization on the xBD validation set	Purple
F1_Dam	F1 value for building damage classification on the xBD validation set	Green
F1_Undam	F1 value for classification of undamaged buildings on the xBD validation set	Grey
F1_Min	F1 value for classification of minor damage buildings on the xBD validation set	Blue
F1_Maj	F1 value for classification of major damage buildings on the xBD validation set	Orange
F1_Des	F1 value for classification of destroyed buildings on the xBD validation set	Red

In order to display the model performance more intuitively, this article randomly selected a set of images from the test set for visual testing. The extracted images were satellite images of hurricane disasters. Predict it through the trained Siamese-CNN model, and visualize the results predicted by the model. The visualization results are shown in the figure 14. (Red in the picture represents destroyed buildings, orange represents major damaged buildings, blue represents minor damaged buildings, and gray represents undamaged buildings) It can be seen that Siamese-

CNN has achieved good results in building damage classification tasks.

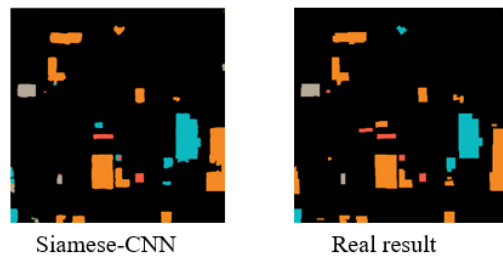


Figure 14. Visual results of testing using Siamese-CNN network model

It can be seen from the experimental results that the Siamese-CNN machine learning model proposed in this article has achieved good results in the building damage assessment task. In actual use of the system, when disaster satellite images are input, the system will first segment the buildings in the entire image and then classify the buildings according to the damage level, truly independently assessing building damage and responding to disasters.

V. CONCLUSIONS

The main purpose of this article is to realize the hierarchical assessment and classification of building damage levels and apply the classification results to disaster loss assessment. In satellite images, the distance between urban buildings is generally small, so it is difficult to accurately distinguish the buildings from the background and correctly position the buildings. The Siamese neural network can train two highly similar images at the same time to improve the accuracy of positioning and damage classification by sharing weights. The Siamese-CNN model used in this article can be applied to building damage detection or building change detection. It uses the timeliness of remote sensing to dynamically obtain building change information and assist other fields.

The follow-up research of this project is as follows:

- 1) Adjust the complexity of the model, study other building damage classification networks, and reduce the complexity of the model as much as possible while ensuring the accuracy of the network and reduce the model's demand for hardware.

- 2) The model can be adjusted and an attention module added to improve the model's ability to extract and capture feature information and improve the accuracy of the model.

REFERENCES

- [1] Shen G. Hwang N S. Spatial-Temporal snapshots of global natural disaster impacts Revealed from EM-DAT for 1900-2015 [J]. *Geomatics, Natural Hazards Risk*, 2019, 10(1):912-934.
- [2] Li S. Song W. Fang L. et al. Deep Learning for Hyperspectral Image Classification: An Overview [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(9):6690-6709.
- [3] Mahdi H. Reza S. Tevmoor S S. et al. Earthquake Damage Region Detection by Multitemporal Coherence Map Analysis of Radar and Multispectral Imagery [J]. *Remote Sensing*, 2021, 13(6):1195-1195.
- [4] Shi D. Ping W. Khushnood A. A survey on deep learning and its applications [J]. *Computer Science Review*, 2021, 40.
- [5] Janiesch C. Zschech P. Heinrich K. Machine learning and deep learning [J]. *Electronic Markets*, 2021, 31(3):1-11.
- [6] Khan I A. Al-Habsi S. Machine Learning in Computer Vision [J]. *Procedia Computer Science*, 2020, 167(C):1444-1451.
- [7] Mahony O N. Campbell S. Carvalho A. et al. Deep Learning vs. Traditional Computer Vision. [J]. *CoRR*, 2019, abs/1910.13796.
- [8] Wang P. Fan F. Wang P. Comparative Analysis of Image Classification Algorithms Based on Traditional Machine Learning and Deep Learning [J]. *Pattern Recognition Letters*, 2020.
- [9] Deng J. Jun D. Xiaojing X. et al. A review of research on object detection based on deep learning [J]. *Journal of Physics: Conference Series*, 2020, 1684(1):012028-.
- [10] Shervin M. Y Y B. Fatih P. et al. Image Segmentation Using Deep Learning: A Survey. [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, PP.
- [11] Ghosh S. Das N. Das I. et al. Understanding Deep Learning Techniques for Image Segmentation [J]. *ACM Computing Surveys (CSUR)*, 2019, 52(4):1-35.
- [12] Wenchao G. Shuang B. Lingxing K. A review on 2D instance segmentation based on deep neural networks [J]. *Image and Vision Computing*, 2022, (prepublish):104401-.
- [13] S. S M. Biswaieet P. Challenges and limitations of earthquake-induced building damage mapping techniques using remote sensing images-A systematic review [J]. *Geocarto International*, 2022, 37(21):6186-6212.
- [14] Duarte D. Nex F. Kerle N. et al. Satellite Image Classification of Building Damages Using Airborne and Satellite Image Samples in a Deep Learning Approach [J]. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2018, 4(2).
- [15] Xu J Z. Lu W. Li Z. et al. Building damage detection in satellite imagery using convolutional neural networks [J]. *arXiv preprint arXiv:1910.06444*, 2019.
- [16] Gupta R. Goodman B. Patel N. et al. Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery [C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, 10-17.
- [17] Weber F. Kané H. Building Disaster Damage Assessment in Satellite Imagery with Multi-temporal Fusion [J]. *arXiv preprint arXiv:2004.05525*, 2020.
- [18] Gupta R. Hosfelt R. Saieev S. et al. xBD: A Dataset for Assessing Building Damage from Satellite Imagery [J]. *arXiv preprint arXiv:1911.09296*, 2019.
- [19] Schuegraf P. Bittner K. Automatic Building Footprint Extraction from Multi-Resolution Remote Sensing Images Using a Hybrid FCN [J]. *ISPRS International Journal of Geo-Information*, 2019, 8(4):191.
- [20] Teia K. Jens L. Felix S. et al. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing

- [1]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 17324-49.
- [21] Jia S. Shaohua G. Yundiang Z. et al. A survey of remote sensing image classification based on CNNs [J]. Big Earth Data, 2019, 3(3):232-254.
- [22] Y. C. X. D. M. L. et al. Dynamic convolution: Attention over convolution kernels [J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 11027-11036.
- [23] Andrea A. Francesco D. Francesco I. et al. A survey on modern trainable activation functions [J]. Neural Networks, 2021, 138(prepublish):14-32.
- [24] Afia Z. Muhammad A. Nazri N M. et al. A Comparison of Pooling Methods for Convolutional Neural Networks [J]. Applied Sciences, 2022, 12(17):8643-8643.
- [25] D. C. Siamese Neural Networks: An Overview [J]. Methods in Molecular Biology, 2021, 219073-94.
- [26] Liu X. 0003 Z. Y. Zhao J. et al. Siamese Convolutional Neural Networks for Remote Sensing Scene Classification [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(8):1200-1204.
- [27] Chen C. Gong W. Chen Y. et al. Object Detection in Remote Sensing Images Based on a Scene-Contextual Feature Pyramid Network [J]. Remote Sensing, 2019, 11(3):339.
- [28] Koonce B. Koonce B. ResNet 50 [J]. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, 2021: 63-72.