THE EXTRACTION OF POST-EARTHQUAKE BUILDING DAMAGE INFORMATIOM BASED ON CONVOLUTIONAL NEURAL NETWORK

Meng Chen [1], Xiaoqing Wang [1, *], Aixia Dou [1], Xiaoyong Wu [1]

[1] Institute of Earthquake Forecasting, China Earthquake Administration, Beijing 100036, China - (hpu_cm, wangxiaoq517)@163.com

Commission III, ICWG III/IVa

KEY WORDS: Earthquake, Seismic Damage Information, Extraction, Deep Learning, Convolutional Neural Network

ABSTRACT:

The seismic damage information of buildings extracted from remote sensing (RS) imagery is meaningful for supporting relief and effective reduction of losses caused by earthquake. Both traditional pixel-based and object-oriented methods have some shortcoming in extracting information of object. Pixel-based method can’t make fully use of contextual information of objects. Object-oriented method faces problem that segmentation of image is not ideal, and the choice of feature space is difficult. In this paper, a new stratage is proposed which combines Convolution Neural Network (CNN) with imagery segmentation to extract building damage information from remote sensing imagery. the key idea of this method includes two steps. First to use CNN to predicate the probability of each pixel and then integrate the probability within each segmentation spot. The method is tested through extracting the collapsed building and uncollapsed building from the aerial image which is acquired in Longtoushan Town after Ms 6.5 Ludian County, Yunnan Province earthquake. The results show that the proposed method indicates its effectiveness in extracting damage information of buildings after earthquake.

1. INTRODUCTION

China, surrounded by the world's two major seismic zones (the Eurasian seismic belt, the circum-pacific seismic belt), suffers many serious earthquake disasters. With the development of remote sensing technology, the acquisition capacity and the quality of remote sensing images have been greatly improved, which provide favourable conditions for extracting seismic damage information of buildings from remote sensing images. There are many previous works have been done to analyse the remote sensing imagery. According to basic unit of classification, the method of extracting earthquake damage information can be summed up as pixel-based method and object-based method. The pixel-based classification method can't fully use the spectral, shape, texture and contextual information of the image, which makes the accuracy of the extraction and classification of the buildings relatively low. Baatz and Schäpe (1999) firstly proposed the object-oriented method to deal with high resolution remote sensing image. The key technology is multi-scale segmentation based on the minimum principle of heterogeneity (producing object), and the classification system is based on fuzzy logic and fuzzy mathematics (information extraction) (Metzler, et al. 2007). But the object-oriented method is usually difficult to choose the feature used for the image classification.

Recently, deep learning has become state-of-the-art solution for visual recognition (Nogueira, et al. 2017). Given its success, deep learning has been intensively used in several distinct tasks of different domains (Ian Goodfellow, et al. 2016; Bengio and Yoshua, 2009). In remote sensing field, compared with the traditional classification method based on pixel or object, the artificial neural network classification algorithm has strong ability of self-learning and fault tolerance (Haykin, 2008). There are several CNN architectures have been proposed in analysis of remote sensing imagery. Saito, et al. (2016) introduced a CNN-based framework, which is used to extract building and road. Basaeed, et al. (2016) proposed a region segmentation technique for remote sensing images using a boosted committee of CNNs coupled with inter-band and intra-band fusion. The proposed method is a fusion framework consisting of a set of thirty boosted networks that derive individual probability maps on the location of region boundaries from the different multi-spectral bands and combines them into one using an averaging inter-band fusion scheme. Because a set of thirty boosted networks are used, it will increase the cost of time. Längkvist, et al. (2016) shown how a CNN can be applied to multispectral ortho imagery and a digital surface model (DSM) of a small city for a full, fast and accurate per-pixel classification. But in earthquake disaster region, DSM data usually can’t be acquired in time. Saito and Aoki (2015) used CNN to learn mapping from raw pixel values in aerial imagery to three object labels (buildings, roads, and others). It applies a patch-based approach, and the boundary of the label object is very sharp.

In order to overcome the shortcomings of pixel-based and object-oriented methods, a new stratagem that object-oriented and CNN are combined is proposed to extract the information of damaged buildings from remote sensing image.

This article include the method, technology route and basic architecture of CNN. Then experiment and results are described. The Last is discussion and conclusion.

2. METHODS

2.1 Basic Idea of the Methods

The basic idea to extract building damage from RS image by using the method of CNN combined with imagery segmentation is shown in Figure 1. The main workflow is divided into two steps. The first step is divided into two parts. One is the segmentation of remote sensing imagery; The other is to use well-

* Corresponding author

trained CNN to predicate pixel's probability belongs to a certain probability and then generate probability patches. The second step is to combine segmentation spots and probability map to integrate the category of every segmentation spots.

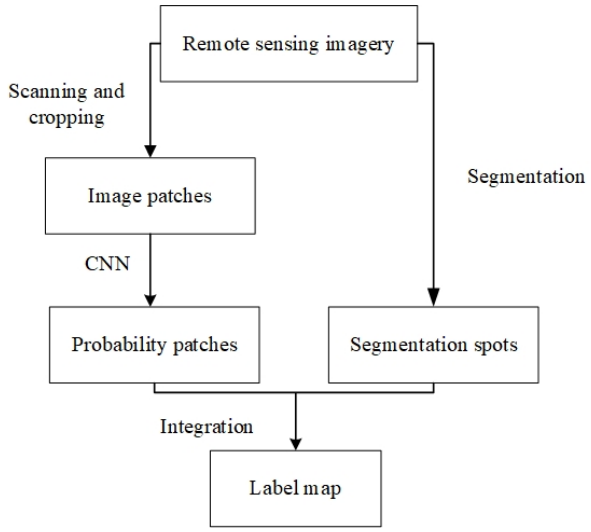

Figure 1. Workflow map (modified from Yuanming Shu, 2014).

Multi-scale segmentation method is used to segment the remote sensing imagery into segmentation spots, which will be unit for integrate the probability. During the segmentation process three key parameters (scale parameter, image layer weights, composition of homogeneity criterion) will be set.

CNN is one of the import methods for the research. Let a fixed size window scan through the entire remote sensing imagery with a stride s. At each location, it will produce an image patch *N* with the same size of the window, which is the input of CNN. The operation of CNN produces a probability patch $\hat{\mathrm{m}}$ with a fixed size ($s_{\hat{m}} \times s_{\hat{m}}$). The $\hat{\mathrm{m}}$ represents the probability of pixels within the boundary which has the size of $s_{\hat{m}} \times s_{\hat{m}}$ at the centre of image patch *N*. Then assemble all probability patches and form the probability map of the entire remote sensing imagery. Some pixels in different probability patches may present the probability of the same pixel in image *N*. In this situation, the max value of pixel in these probability patches is chosen as the probability value of pixel in image *N*. At last, the value of every pixel within a segmentation spot is integrated to obtain the average probability of every segmentation spot.

The goal of this article is extract building damage information, including collapsed building, un-collapsed building and background. Its focus operation is to use trained wall CNN to predict a multi-channel label image $\hat{m}$ from an input image patch N. Let K_c be the number of categories that we want to extract, while it is easy to understand image $\hat{m}$ has K_c channels. The main work is to use input image patch N and the corresponding ground truth map $\tilde{m}$ which labels the category of every pixel to train CNN. Then we can use the well-trained CNN model to predicate the category of the raw pixels.

2.2 Basic Architecture of CNN

The basic architecture of CNN usually consists of alternatively piled convolutional layer, fully connect layer and predicate layer. The convolutional layer usually includes the operation of convolution, non-liner transformation and pooling.

The convolutional layer usually connects input imagery or feature maps. The operation of convolution is explained as following. Assume that an imagery or feature map having the size of $s_n \times s_n$ with K_N-channels and K two-dimensional filter-kernels having the size of $w_f \times h_f$ are taken as inputs of the convolutional layer. The output maps will be in the size of $(s_n - \mathrm{w}_f + 1) \times (s_n - \mathrm{h}_f + 1)$ with *K*-channels. Each channel of the output image is called a filter site (Alshehhi, et al. 2017). In the case that the convolution process is not slide 1 pixel, the stride r which effects the size of filter site is required (Nogueira, et al. 2016). If $r > 1$, the size of an output map from convolution process is decreased to $((s_n - \mathrm{w}_f)/r + 1) \times ((s_n - \mathrm{h}_f)/\mathrm{r} + 1)$ The convolution process is defined as following:

$$x_k(i,j) = \sum_{k'=1}^{K_N} \left\{ \sum_{p=0}^{w_f-1} \sum_{q=0}^{h_f-1} x_{k'}(i \cdot r + p,\ j \cdot r + q) \cdot h_k(p,q) \right\} + b_k \quad (1)$$

where $x_{k'}(i,j)$ = pixel value at (i,j) in k'-th channel of an input image or of a feature map

$x_k(i,j)$ = pixel value at (i,j) of k-th filter site

$h_k(p,q)$ = weight value at(p,q) on k-th filter

b_k = a bias parameter of k-th filter that is shared among all locations (p,q)

The second operation of convolutional layer is non-liner transformation which is also called activation function. There are many activation functions that can be used, such as sigmoids, hyperbolic tangents, and rectified linear units (ReLUs) (Krizhevsky, et al. 2012). But rectified function is currently the mostly used because ReLUs are known to offer some practical advantages in the convergence of the training procedure. In this paper, we use ReLU function (Nair and Hinton 2010), as follows:

$$x_k'(i,j) = max\big(0, x_k(i,j)\big) \quad (2)$$

After the process of activation function, the next step is pooling, which takes the filter sites operated by non-liner transformation performs subsampling to them by considering maximum or average value in $w_p \times h_p$ pooling window. this pooling window is set to slip at a stride *t*. In this article max-pooling is used. Let us assume that $x_k'(i,j)$ is an output of the previous activation operator and by applying max-pooling, the output $x_k''(i,j)$ is expressed as following:

$$x_k''(i,j) = \max_{0 \le i' \le w_p - 1, 0 \le j' \le h_p - 1} \big(x_k'(i \cdot t + i', j \cdot t + j')\big) \quad (3)$$

when the suitable number of convolutional layers are stacked, the next several layers are usually set to be fully connected layers to comprehensively use the entire features of the image patch. But the fully connected layer will hugely increase the number of parameters needed training and increase the cost of computation. So a dropout stratagem (Hinton, et al. 2012) is adopt. In some CNN architectures, they usually don't have only one fully connected layer, and in this article the CNN architecture has two fully connected layers.

Fully connect layer is usually followed by A classifier layer. The operation of the classifier is described as following. Let assume $\boldsymbol{x} = [x_1,\ x_2, \dots \dots, x_L]^T$ denotes the output of fully connected layer and softmax function is applied to each $\boldsymbol{x}$ to convert into probability vector which is reshaped into the form $\hat{m}_\tau = [\hat{m}_{\tau,1},\ \hat{m}_{\tau,2}, \dots \hat{m}_{\tau,k_c} \dots, \hat{m}_{\tau,K_c}]^T$, $(\tau = 1,2,\cdots, s_{\hat{m}} \times s_{\hat{m}})$. The expression of the function as follow:

$$\hat{m}_{\tau,k_c} = \frac{\exp(x \cdot w_{\tau,k_c})}{\sum_{j=1}^{K_c} \exp(x \cdot w_{\tau,j})} \quad (4)$$

where w_{τ,k_c} = the τ-th weight vector which connect to the k_c-th probability value of τ-th pixel output unit

In this article, we adopt the CNN architecture which is shown in figure 2. The CNN is trained by minimizing the negative log likelihood using mini-batch stochastic gradient descent with momentum (Bengio and Yoshua, 2012)

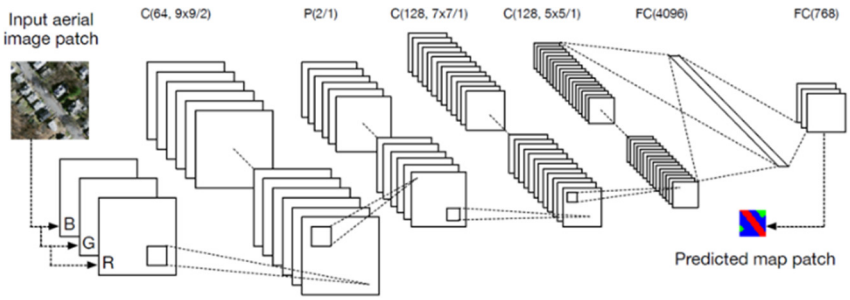

Figure 2. the CNN architecture (Saito and Aoki 2015)

After the training of the network, it operates the test imagery, then we can acquire the probability map. The map with K_c channels, expresses the probability of each pixel belong to a certain category. If the probability map is used to produce the label map according to a threshold, it will also cause a lot of salt-and-pepper noise similar as the pixel-based method. This phenomenon can be avoided by combining it with the segmentation of imagery. In the probability map, we integrate the value of the pixel within the boundary of a certain segmentation spot, the function is expressed as following:

$$p_{k_c} = \frac{1}{Q} \sum_{\tau=1}^{Q} \hat{m}_{\tau,k_c} \quad (5)$$

where $\hat{m}_{\tau,k_c}$ is the probability of the τ-th pixel in k_c-th channel within a certain segmentation spot

Q is the total number of pixels within a certain segmentation spot

p_{k_c} is the probability of the segmentation spot belongs to the k_c-th category

3. EXPERIMENT AND RESULTS

3.1 Dataset

A destructive earthquake with Ms6.5 occurred at 16:30 on August 3, 2014 in Ludian County (27.1° N, 103.3°E), Yunnan Province. The postearthquake aerial image was acquired with area of about 12 km^2 and resolution of 0.3m in Longtoushan Town of Ludian. Some collapsed and uncollapsed buildings are found in the imagery. This imagery is used for experiment.

The imagery need to be preprocessed to train CNN. Firstly, the ground truth map $\hat{m}$ is obtained (Figure 3) by using object-based classification method and then corrected artificially for each object spot to suitable category. Secondly, we scan and crop the imagery and ground truth map to patches which has size of 64 × 64 pixels. These patches are divided into three sets: training (9000 patches), validation (2000 patches) and testing (500 patches). Every data and ground truth map are rotated randomly, to increase the number of training data.

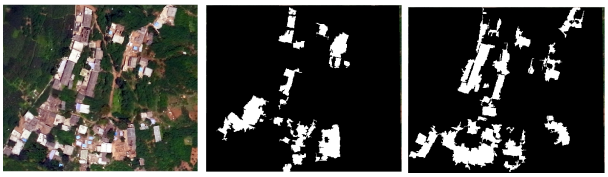

Figure 3. The original map (left) and the truth ground map of collapsed building (middle) and un-collapsed building(right)

3.2 Experiment

The imagery is segmented with a set of parameters as scale parameter=50, image Layer weights=1, composition of homogeneity criterion=0.5. As a result the boundary of each segmentation spot is shown in figure 4.

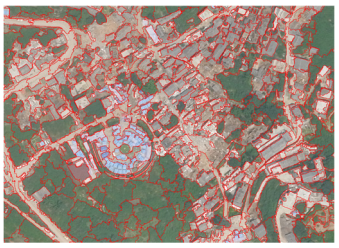

Figure 4. The Multi-scale segmentation of remote sensing imagery

Before the training of CNN, we firstly set the hyper parameters of the net wok. The fine-tuning was done to reduce the training iteration. In this article, hyper parameters as set as following (Alshehhi, et al. 2017): mini-batch size of 128 with the momentum of 0.9. The training was regularized by weight decay set to 0.0005, and dropout regularization for all fully connected layers with dropout ratio set to 0.5. We initialized the weights in each layer with a random number drawn from a zero-mean Gaussian distribution with standard deviation 0.01. The learning rate is started with 0.0005 with initial bias set to constant 0.1.

Then well-trained model is chosen to predicate the probability of every pixel belonging to a certain category, such as the probability map of unclooapsed building which is shown as figure 5. The Combination of the segmentation spots and the probability map (as is shown in figure 6.) is used to integrate the probability of every segmentation spot belonging to a certain category by function (5).

Figure 5. The probability map of uncollapsed building

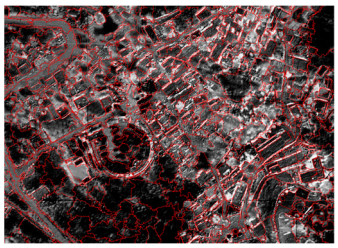

Figure 6. The probability overlay the boundary of segmentation spots

3.3 Result

Through the above processing, the seismic damage information of the building is extracted from the aerial image in Longmenshan Town (Figure 7). The result is shown in figure 8. The result is compared with the visual interpreted one to verify the accuracy of this method. The test result is shown in table 1. The overall accuracy is 0.93, Kappa is 0.86, which show a good classification result.

Figure 7. Aerial photograph in LongTouMountain-town of Ludian acquired after Ludian earthquake

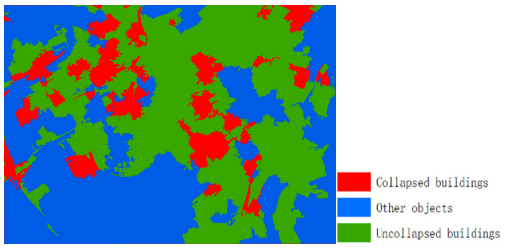

Figure 8. the classification result of building damage

Items	Uncollapsed buildings	Collapsed buildings
Visual interpretation	107	37
Extraction by proposed method	114	30
Accuracy	93 %	81%

Table 1. The accuracy test table

4. CONCLUSION

Through the experiment, it is found that the method of combining CNN and segmentation can have a good result in extracting collapsed and uncollapsed buildings. But there still exists some defect comparing the extracted building damage map with the ground-truth map. Firstly, some bare soil or un-collapsed buildings are classified as collapsed buildings and vice versa. Secondly, small number collapsed buildings are not extracted. The reasons can be summarized as following: firstly, the number of training data is too small, what's more the ground truth map may have error; secondly, the collapsed buildings and the bare land are easy to be confused, because both of them have similar spectral signature; thirdly, the number of training of different categories have some difference. In concrete terms, there are more number of training patches are belong to the background (other objects), this phenomenon may result to error. In the future work, we will try to focus on taking the artificially defined features such as topological relations into account, to obtain better effects of seismic damage extraction.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers. The aerial image is provided by Beijing Anxiang Power Technology Co Ltd. The research was supported by national key research and development program of the Ministry of China (2017YFB0504104).

REFERENCES

Rasha Alshehhi, Prashanth Reddy Marpu, Wei Lee Woon, Mauro Dalla Mura, 2017. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, pp. 139-149.
http://dx.doi.org/10.1016/j.isprsjprs.2017.05.002

Baatz, Martin, and A Schäpe. (1999). Object-oriented and multi-scale image analysis in semantic networks. In: 2nd International Symposium: Operationalization of Remote Sensing, 1, pp. 16-20
http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Object-Oriented+and+Multi-Scale+Image+Analysis+in+Semantic+Networks#0

Basaeed, Essa, Harish Bhaskar, and Mohammed Al-Mualla 2016 Supervised remote sensing image segmentation using boosted convolutional neural networks. *Knowledge-Based Systems*, 99, pp.19-27.
http://dx.doi.org/10.1016/j.knosys.2016.01.028

Bengio, Yoshua 2012 *Neural Networks: Tricks of the Trade.* Springer-Verlag Berlin Heidelberg, pp.437-478.
https://doi.org/10.1007/978-3-642-35289-8_26

Bengio, Yoshua, 2009. Learning Deep Architecture for AI. *Foundations and Trends in Machine Learning*, 2(1), pp. 1–127.
https://doi.org/10.1561/2200000006

Haykin, Simon S 2008 *Neural Networks and Learning Machine*. China Machine Press, pp. 32 - 41

Hinton, Geoffrey E, Srivastava, Nitish, Krizhevsky, Alex, Sutskever, Ilya, Salakhutdinov, Ruslan R. 2012 Improving neural networks by preventing co-adaptation of feature detectors. *Computer Science*, 3 (4), pp. 212-223.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, 2016. *Deep Learning*, MIT Press,
http://www.deeplearningbook.org.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. (2012): ImageNet classification with deep convolutional neural networks In: International Conference on Neural Information Processing Systems, pp. 1097-1105.

Längkvist, Martin, et al., 2016. Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks. *Remote Sensing*, 8(4), pp.329.

Metzler, V., T. Aach, and C. Thies, 2007. Object-oriented Image Analysis by Evaluating the Casual Object Hierarchy of a Partitioned Reconstructive Scale-space. Proceedings of ISMM2002, pp. 657-658.

Nair, Vinod, and Geoffrey E. Hinton. (2010): Rectified linear units improve restricted boltzmann machines. In: International Conference on International Conference on Machine Learning. pp. 807-814.
doi: 10.1.1.165.6419

Nogueira, Keiller, Otávio A. B. Penatti, and Jefersson A. dos Santos, 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, 61, pp. 539-556.

Saito, Shunta, and Yoshimitsu Aoki, 2015. Building and road detection from large aerial imagery. In: *Image Processing: Machine Vision Applications*, San Francisco, Vol. VIII, pp. 1814–1821.
doi: 10.1117/12.2083273

Saito, Shunta, Takayoshi Yamashita, and Yoshimitsu Aoki, 2016. Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks. *Electronic Imaging* 60(1), pp. 10402-1/10402-9.

Yuanming Shu, 2014. Deep Convolutional Neural Networks for Object Extraction from High Spatial Resolution Remotely Sensed Imagery. Waterloo, Ontario, Canada, pp. 26.