

Convolutional and Long Short-term Memory Neural Network for Earthquake Detection

Xuefan Xu^{1,2}, Yingxue Wang¹⁺, Lian Zou² and Yifeng Liu¹

¹ China Academy of Electronics and Information Technology, Beijing, 100041, China

² School of Electronic Information, Wuhan University, Wuhan, 430072, China

Abstract. Earthquake detection is a foundation and critical work link in the study of seismology. Traditional detection algorithms have the disadvantages of low detection sensitivity, low computational efficiency, and poor general applicability. In this paper, we develop a novel earthquake detection method based on convolutional neural network (CNN) and long short-term memory network (LSTM). It extracts the high-level features in the seismic signal and learns the time-frequency characteristics of the main phase by recording the three component data on a single station. We train the network using 4932 waveform windows (2236 positive windows and 2696 negative windows) recorded in Wenchuan, China, with a window size of 30s. Using the trained model to test a continuous waveform for one day, compared to the long-short window energy ratio method (STA/LTA), all manually selected seismic events were successfully detected. To explore the impact of different window sizes and LSTM layers on the detection results, we use a larger dataset (derived from Oklahoma, USA) for network training. The test results show that our method not only has a good generalization ability for cataloging events, but also detects micro-seismic events that are not included in the catalog. The detection accuracy of cataloging events reaches 100%. Our results indicate that this method has fast, efficient and scalable superior performance in earthquake detection.

Keywords: earthquake detection, CNN, LSTM.

1. Introduction

Large and small earthquakes occur every year around the world. With the continuous increase of seismic stations, the recorded continuous seismic waveform data has also increased significantly for each earthquake. It is increasingly unrealistic to process these seismic data entirely by hand. How to process massive seismic observation data quickly and efficiently is a challenging task.

At present, the mainstream methods for automatic seismic detection and identification at domestic and foreign are long-and-short window ratio method (STA/LTA) [1], template matching detection algorithm [2]-[5] and fingerprint similarity threshold method [6]. Traditional methods have different focuses on the qualitative indicators of seismic detection. Although STA/LTA can find unknown sources, has high universal applicability, can detect earthquakes in real time, and has high computational efficiency, but lacks sensitivity to low SNR events, so its detection sensitivity is low. Although the template matching method solves the problem of low detection sensitivity, the method needs to determine the template waveform in advance, so its general applicability is low. For the fingerprint similarity threshold method, although the efficiency of the similarity search can be improved, the output is limited to repeated events.

The deep learning method is a subset of artificial neural networks (ANNs) that has emerged and revolutionized many sciences and engineering domains in recent years [7]-[9]. In particular, recent advancements in deep learning (DL) frameworks based on deep neural networks drastically improved

⁺ Corresponding author. Tel.: + 8613269736099;
E-mail address: wangyingxue@cslslab.net.

accuracy in image recognition, natural language processing and other applications [10]. Limin et al. [11] used deep neural network to identify singular points on a fingerprint. Alexiei et al. [12] made use of Convolutional Neural Networks (CNNs) to forecast the next period price direction with respect to the current price. Recently, Wisal et al. [13] shed light on the utilization of a deep convolutional neural network (DCNN) for facial emotion recognition from videos. The advantage of neural networks is that they have more hidden layers. By multi-level abstraction, it has a stronger description ability than simply increasing the number of neurons. It can abstract the required low /Medium/Advanced features from the original data, which eliminates the need for manual engineering of important pre-processing and features [14]. Perol et al. [15] used the CNN method to identify and classify induced earthquakes in the central United States. While achieving higher recognition capabilities than template matching methods, computational time-consuming and storage usage was significantly lower than the FAST method. The convolutional neural network is a kind of deep neural network. Using convolution kernel as mediation, the inherent local structure existing in the data is identified by convolution/downsampling operation, and the number of weight parameters is reduced, which can effectively avoid over-fitting. Titos et al. [16] introduced recurrent neural networks (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU) to detect and classify continues sequences of volcano-seismic events at the Deception Island Volcano, Antarctica. Yue W et al. [17] developed a region-based densely connected network method to detect events from seismic signals. With the continuous development of machine learning and deep learning technology in recent years, the number of researchers using this technology for seismological research has increased year by year. This has led to the development of many excellent seismic detection algorithms [18]-[22].

In this study, we consider seismic detection as a two-class supervised learning problem. Using the technique of deep learning, a seismic detection method combining a convolutional neural network and a long short-term memory network is proposed. The time-frequency characteristics in the seismic signal are learned by recording three component data on a single station. Among them, the Convolutional Neural Network (CNN) extracts the advanced features of seismic waveforms through convolution/downsampling operations. After the convolution feature extraction, the feature quantities are redistributed into the sequence and then fed into the LSTM network. The LSTM layer will learn and model the timing patterns of the data. At the end of the network is the fully connected layer, which performs advanced reasoning and maps the learned sequence model from the final step to the desired output class. We used more than 30 G of raw seismic waveform data, divided the training and test sets, and then demonstrated the performance of the network. The model is highly scalable, and it generalizes the seismic signals that have never appeared during training. Compared with traditional algorithms, the detection accuracy is higher, and the operation speed is faster.

2. Methodology

2.1. Long-short-term memory networks

Long-short-term memory networks are a special kind of recurrent neural networks. They were first proposed by Hochreiter & Schmidhuber in 1997. The original intention was to solve the problem of long-term dependence on neural networks. The key to LSTM is the state of the cell, which transfers information from the previous unit to the next unit with minimal linear interaction. Also, the LSTM controls the discarding or adding information through the "gate" to achieve the function of forgetting or remembering. An LSTM has three of these gates, to protect and control the cell state, namely the forgotten gate, the input gate, and the output gate. The structure of multiple LSTM layers is as shown in Fig. 1 below.

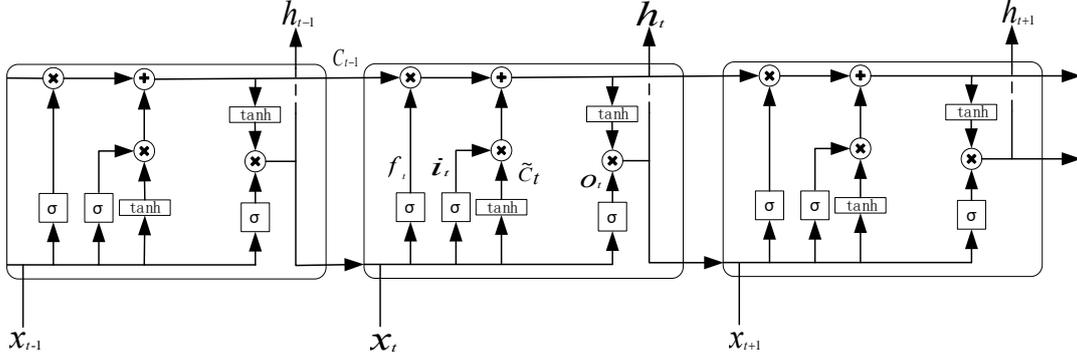


Fig. 1: Multiple LSTM layers

In the figure, the definition of cell memory and each gate is as follows

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

Where f_t is forget gate, i_t is input gate, O_t is the output gate, and \tilde{C}_t is candidate value for replacing the memory. The matrices W_f, W_i, W_o, W_C and b_f, b_i, b_o, b_C are associated weights and bias terms, and σ and \tanh are activation functions:

$$\sigma = \frac{1}{1 + e^{-z}} \quad (5)$$

$$\tanh = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (6)$$

The value of memory cell at each time step will be set using the candidate value at current step (\tilde{C}_t) and previous value (C_{t-1}) based on input and forget gates:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (7)$$

Where $*$ is element-wise multiplication. The final state at time t, h_t , is obtained based on output gate and the value of memory cell:

$$h_t = O_t * \tanh(C_t) \quad (8)$$

LSTM has a variety of applications in the technology field. LSTM-based systems can learn translation languages, control robotics, image analysis, document summarization, speech recognition image recognition, handwriting recognition, control of chat bots, predictive disease, click-through rate and stock, synthetic music, and more. In this paper, we apply LSTM to seismic waveform sequence learning, and the results have achieved good classification detection results.

2.2. The network architecture

The network structure we proposed is shown in Fig. 2. In this network structure, we use an 8-layer convolutional neural network, and the number of layers in the LSTM layer was not specified in consideration of the needs of the experiment.

The initial inputs of the network structure are three-component seismograms of known earthquakes. Due to a large amount of data, we create the dataset and convert the waveform data of the original fixed window and the corresponding label into the .tfrecords file format corresponding to tensorflow. The records data file is a binary file that stores seismic waveform data and tags in a unified manner. It can make better use of memory and quickly copy, move, read, and store in tensorflow. The convolutional neural network (CNN) is used as feature extractor because it can extract spatial structure of input channels through convolution operation. At each convolution layer, all three channels are convolved and summed using a linear 1D filter. We use a convolution kernel size of 3, which extracts the high-frequency characteristics of the input signal.

At the same time, unlike standard convolution, we set the stride size to 2 instead of 1, and the feature amount will be reduced by half after each convolution/downsampling, which will speed up the computational efficiency. With progressive downsampling, deep input signals have an exponentially increasing receptive field, which allows them to extract low-frequency features. During the convolution process, we set the fill mode to ‘SAME’ zero padding to ensure that the feature size after convolution remains the same.

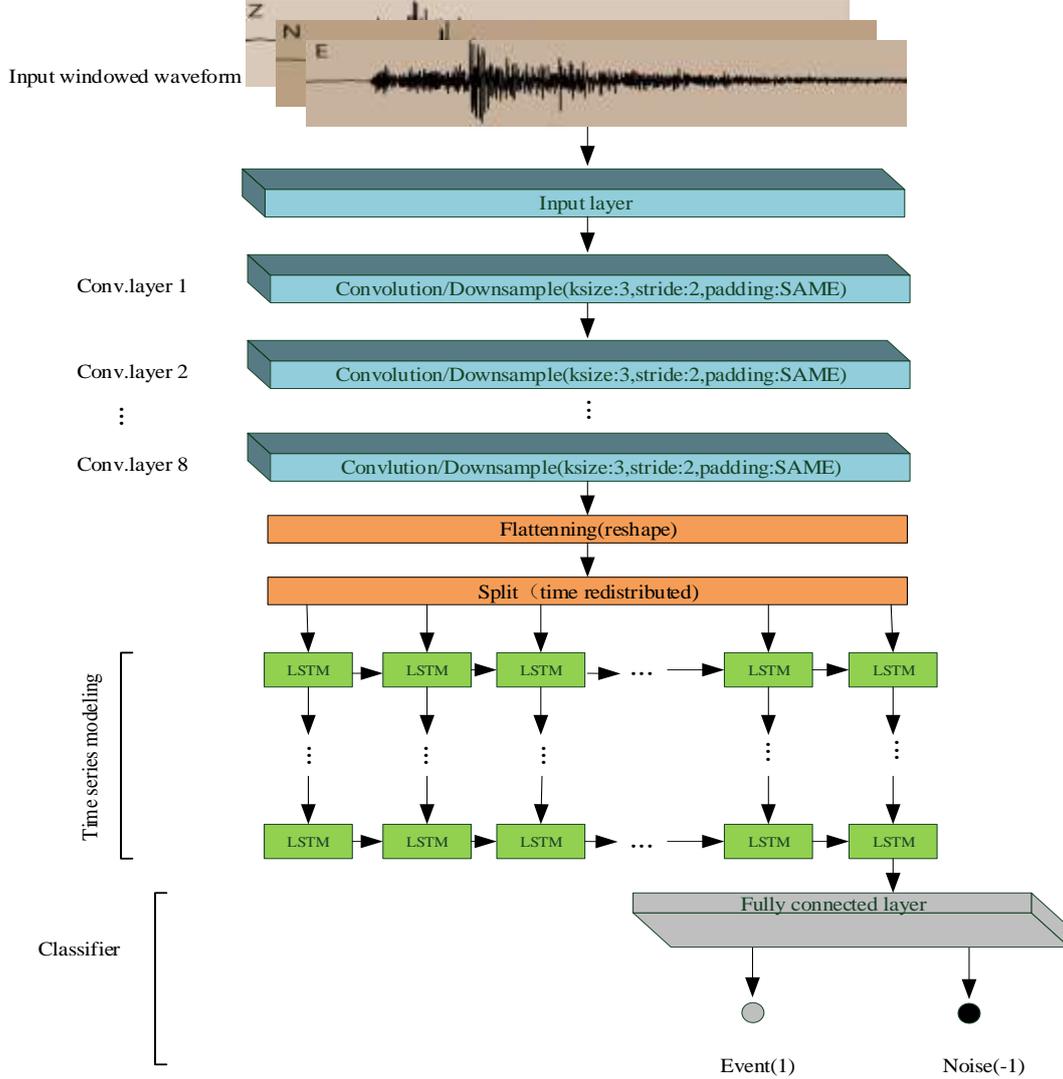


Fig. 2: The architecture of our proposed deep neural networks.

After CNN feature extraction, we will extract the extracted features tensor in one-dimensional and sequence segmentation. It is then fed into the LSTM module, which will learn and model the timing patterns of the data. Since each of our sequences is just trying to predict a classification probability, we only use the last element of the LSTM top-level output sequence. Finally, enter them into the fully connected layer, which performs advanced reasoning and maps the learned sequence model from the final step to the desired output class. Here, the classification activation function we use is the softmax function. Softmax function is a generalization of the logistic function that maps a length-p vector of real values to length-K vector of values, as follows:

$$P(y = j|x) = \frac{e^{x^T W_j}}{\sum_{k=1}^K e^{x^T W_k}} \quad (9)$$

Here, this can be seen as the composition of K linear functions $x \mapsto x^T W_1, \dots, x \mapsto x^T W_K$, where $x^T W$ denotes the inner product of x and W .

Overall, the model is an end-to-end learning framework that can characterize seismic data with high precision. It is scalable to meet many applications for seismic data processing automation.

2.3. Experimental results

A large number of original seismic flow data and corresponding seismic event catalogs we used in the study were published by Perol et al. (2017) on Github. We used this raw seismic stream data to create training and test data sets for two different window size events and noises based on the corresponding catalog. Among them, the 10s window size training set contains 2709 events, 1263847 noise; the test set contains 209 events, 231168 noise. The 30s window size training set contains 2709 events, 424409 noise, and the test set contains 209 events and 161,280 noises. To get a better model performance improvement, we preprocess the data set by removing the mean and normalization. We trained the model on one GeForce GTX 980 Ti-PCI-E GPU. We optimize the model parameters by minimizing an L2-regularized cross-entropy loss function. The optimization algorithm chosen is the ADAM algorithm. When training and testing the models for different window sizes, we performed two, five, and eight layers on the LSTM layer.

To more intuitively reflect the process of training and test evaluation, we use the matplotlib library to visualize it. As shown in Fig. 3 below

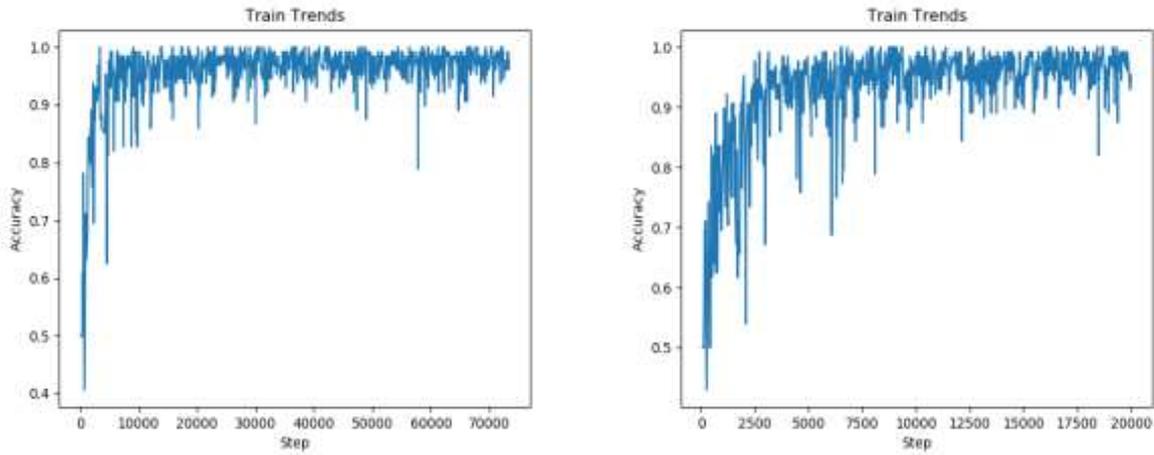


Fig. 3: Accuracy rate trend chart for different window size training processes. The left side indicates the size of the 10s window, the right side indicates the size of the 30s window.

Table 1 shows the results of the selected 10s-window size for earthquake detection. From Table 1, we can see that our model has achieved 100% accuracy for the detection of events. This shows that our model has good detection and recognition capabilities. Also, for the detection of noise, the effect is not obtained as the same as the event. This is because we created the dataset based on the manually selected earthquake catalog. For some earthquakes with small magnitudes, it is inevitable that there will be some leakage. By comparing the results of the different LSTM layers, we conclude that the effect of using the 5-layer LSTM layer is optimal.

Table 1: The results of the selected 10s-window size for earthquake detection

Model	Average Loss	Average accuracy (events)	Average accuracy (noise)
10s_window/2_layers_CNN+LSTM	0.0500	100%	69%
10s_window/5_layers_CNN+LSTM	0.0424	100%	97.6%
10s_window/8_layers_CNN+LSTM	0.0333	100%	72.2%

Table 2 shows the results of the selected 30s-window size for earthquake detection. It can be seen that, as the 10s window size, when the number of LSTM layers is 5, the detection result is the best. At the same time, for our datasets, from the experimental results of different window sizes, the detection of the window size of 10s is better than the window size of the 30s. We found that the selection of window size is a very important step in earthquake detection preprocessing.

Table 2: The results of the selected 30s-window size for earthquake detection

Model	Average Loss	Average accuracy (events)	Average accuracy (noise)
30s_window/2_layers_CNN+LSTM	0.0786	100%	69%
30s_window/5_layers_CNN+LSTM	0.0851	100%	88.8%
30s_window/8_layers_CNN+LSTM	0.0772	99.5%	73%

Similarly, we selected the corresponding best model for the test set of two different window sizes to make the test verification process trend graph. The results are shown in Fig. 4 below.

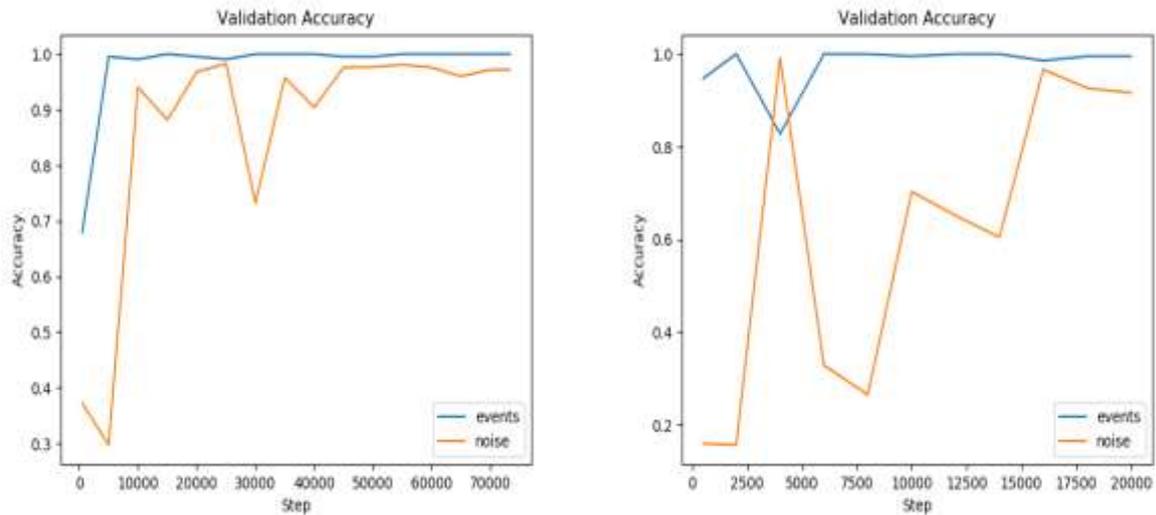


Fig. 4: Accuracy rate trend chart for different window size testing processes. Blue and red lines represent events and noise, respectively. The left side indicates the size of the 10s window, the right side indicates the size of the 30s window.

In addition to the performance tuning experiments of the model, we also used a small data set to perform experiments comparing the performance of the traditional STA/LTA method with the proposed model. Our training set contains 2,236 events and 2,696 noise in 30s window size samples. The test set was a continuous waveform of July 25, 2008, on which we manually picked 44 cataloging events. Through experimental testing, our model was able to quickly and accurately detect all 44 cataloging events, and also detected some uncataloged microseisms (confirmed by AR+AIC method). Compared to our model, only the 41 events in the catalog were detected using the STA/LTA method. The experimental test results are shown in Table 3 below:

Table 3: Comparison of two detection results

Method	Number of detection
STA/LTA	41
CNN-LSTM	44

Fig. 5, Fig. 6, and Fig. 7 below are the spectrograms of the three components we selected for the detected event and noise.

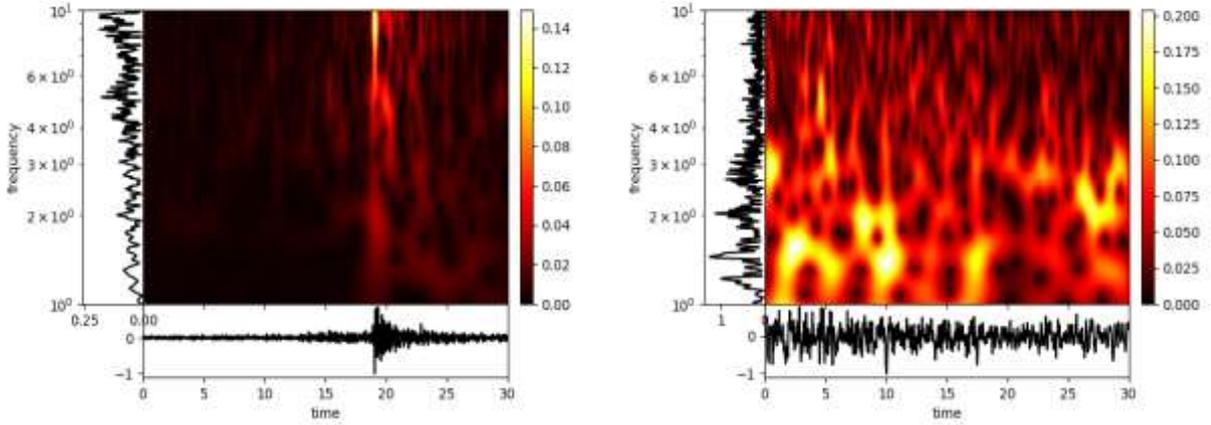


Fig. 5: Spectrogram corresponding to the east and west components (E)

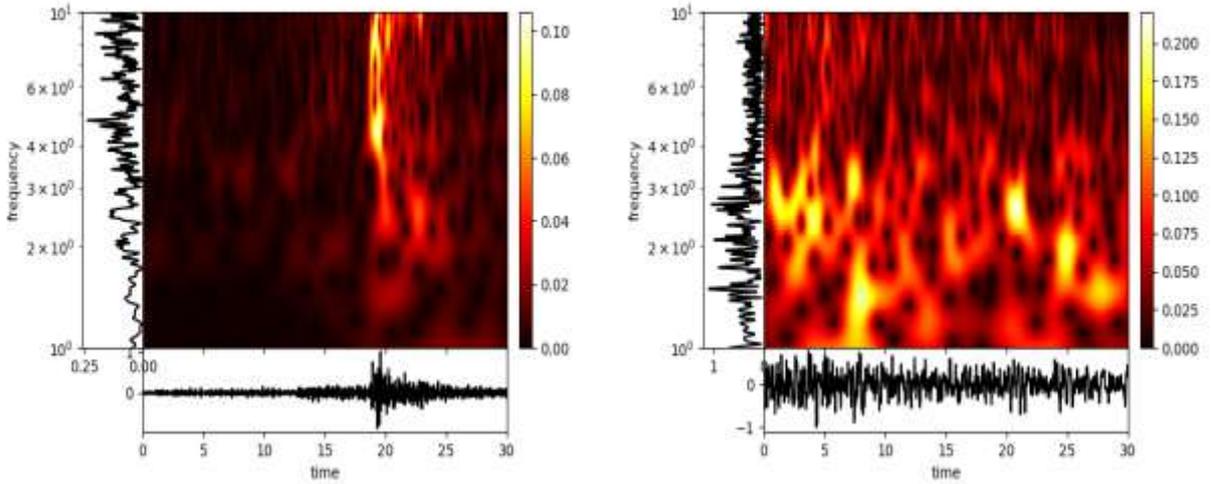


Fig. 6: Spectrogram corresponding to the north and south components (N)

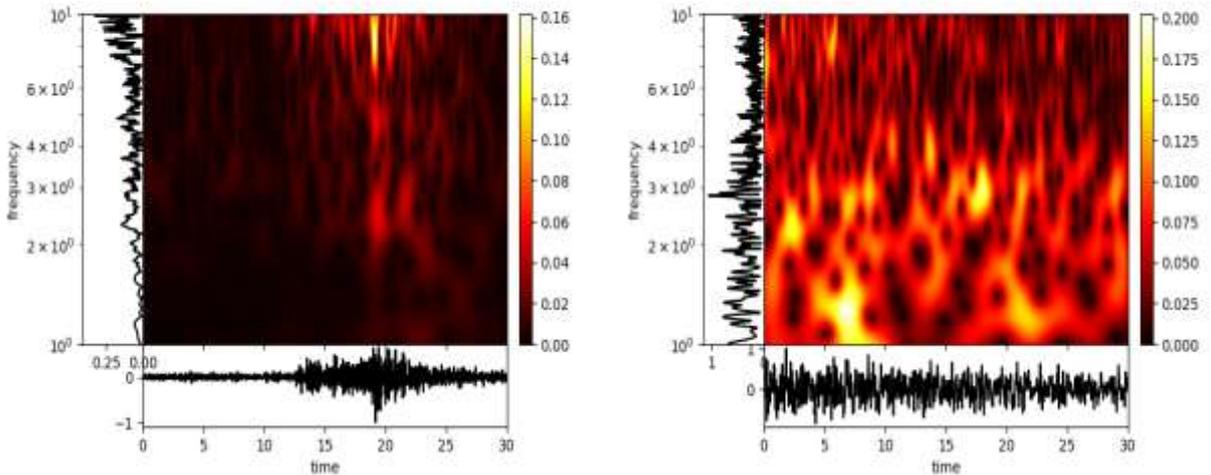


Fig. 7: Spectrogram corresponding to the vertical component (Z)

This paper focuses on how to combine CNN and LSTM for seismic detection and explores the generalization capabilities of models in different LSTM layers. The framework enables fast and accurate detection of seismic events. The learning model also has a good detection effect on microseisms that are not

included in the catalog. Compared with the traditional STA/LTA method, our method has higher detection sensitivity. As long as the model is trained, our method can be quickly applied to seismic stream files for detection and classification. The architecture is very flexible and can be easily extended.

3. Acknowledgments

We would like to thank Perol, author of the paper "Convolutional neural network for earthquake detection and location", for providing a large amount of original seismic flow data and corresponding catalogs on the Github in Oklahoma, USA. Similarly, we would also like to thank the official of the Tianchi aftershock capture competition, providing data on the Wenchuan aftershock in China. This Project was granted financial support from China Postdoctoral Science Foundation (Grant No.2018M630188), the National Key Research and Development Program (2017YFC0821603), and the China Electronic Technology Group Corporation Joint Fund (614B0801010a).

4. References

- [1] Allen R . Automatic earthquake recognition and timing from single trace[J]. *Bull.seism.soc.am*, 1978, 68(5):1521-1532.
- [2] Gibbons, S. J., and Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1), 149–166.
- [3] Shelly D R, Beroza G C, Satoshi I. Non-volcanic tremor and low-frequency earthquake swarms[J]. *Nature*, 2007,446(7133):305-307.
- [4] Ross, Z. E., Rollins, C., Cochran, E. S., Hauksson, E., Avouac, J.-P., and Ben-Zion, Y. (2017). Aftershocks driven by afterslip and fluid pressure sweeping through a fault-fracture mesh.*Geophysical Research Letters*, 44(16), 8260–8267.
- [5] Li, Z., Peng, Z., Hollis, D., Zhu, L., and McClellan, J. (2018). High-resolution seismic event detection using local similarity for Large-N arrays. *Scientific Reports*, 8(1), 1646.
- [6] Yoon C E, O Reilly O, Bergen K J, et al. Earthquake detection through computationally efficient similarity search[J]. *Science Advances*, 2015,1(11):e1501057.
- [7] Donahue J, Hendricks L A, Guadarrama S, et al. Long-term Recurrent Convolutional Networks for Visual Recognition and Description[M]. Elsevier, 2015.
- [8] Ord óñez F J, Roggen D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition[J]. *Sensors*, 2016,16(1):115.
- [9] Bartz C, Herold T, Yang H, et al. Language Identification Using Deep Convolutional Recurrent Neural Networks[J]. 2017:880-889.
- [10] Valeriy Gavrishchaka, Zhenyi Yang, Rebecca Miao, and Olga Senyukova, "Advantages of Hybrid Deep Learning Frameworks in Applications with Limited Data," *International Journal of Machine Learning and Computing* vol. 8, no. 6, pp. 549-558, 2018.
- [11] Limin Liu, "Fingerprint Analysis and Singular Point Definition by Deep Neural Network," *International Journal of Machine Learning and Computing* vol. 8, no. 6, pp. 625-629, 2018.
- [12] Alexiei Dingli and Karl Sant Fournier, "Financial Time Series Forecasting – A Deep Learning Approach," *International Journal of Machine Learning and Computing* vol. 7, no. 5, pp. 118-122, 2017.
- [13] Wisal Hashim Abdulsalam, Rafah Shihab Alhamdani, and Mohammed Najm Abdullah, "Facial Emotion Recognition from Videos Using Deep Convolutional Neural Networks," *International Journal of Machine Learning and Computing* vol. 9, no. 1, pp. 14-19, 2019.
- [14] Mousavi S M , Zhu W , Sheng Y , et al. CRED: A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection[J]. 2018.
- [15] Perol T, Gharbi M, Denolle M. Convolutional neural network for earthquake detection and location[J]. *Science Advances*, 2018,4(2).

- [16] Titos M, Bueno A, Garcia L, et al. Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018,PP(99):1-13.
- [17] Yue W, Lin Y, Zheng Z, et al. DeepDetect: A Cascaded Region-Based Densely Connected Network for Seismic Event Detection[J]. IEEE Transactions on Geoscience & Remote Sensing,PP(99):1-14.
- [18] Kong Q, Allen R M. Seismic waveform classification using deep learning: Agu Fall Meeting, 2017[C].
- [19] Chen Y. Fast waveform detection for microseismic imaging using unsupervised machine learning[J]. Geophysical Journal International, 2018,215(2):1185-1199.
- [20] Zhu W, Beroza G C. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method[J]. 2018.
- [21] Li W, Nakshatra F, Narvekar N, et al. Seismic Data Classification Using Machine Learning[M]. 2018.
- [22] Zhou Z, Zhang Z, Wu Y, et al. Earthquake Detection in 1-D Time Series Data with Feature Selection and Dictionary Learning[J]. 2018.