Research on earthquake magnitude prediction method based on Resnet transfer learning

1st Haomiao Wang
School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture
Beijing 102616, China
E-mail: 1252209619@qq.com

2nd Huaixiu Wang
School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture
Beijing 102616, China
E-mail: wanghxiu@126.com

2024 36th Chinese Control and Decision Conference (CCDC) | 979-8-3503-8778-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CCDC62350.2024.10587759

***Abstract*—To achieve more accurate estimation of microseismic magnitude, a classification model based on deep residual network model transfer learning is constructed to achieve microseismic magnitude prediction. Firstly, the microseismic waveform images were preprocessed as input, and then all weights of the randomly initialized model were selected. The selected values obtained from Resnet18 as the pre trained model were loaded as initialization parameters, and all layers were trained from scratch. Finally, the trained classification model will be applied for earthquake magnitude classification research, ultimately achieving the prediction of microseismic magnitudes. The results show that the prediction accuracy of the model with an error within the range of plus or minus 0.2 can reach 86.33%, and the prediction accuracy of the model with an error within the range of plus or minus 0.3 can reach 94%, demonstrating good classification performance. This model is more stable and reliable, and can effectively solve the problem of missing data labels, verifying the feasibility of using Resnet transfer learning for microseismic magnitude prediction, providing a new technical means for solving magnitude prediction.**

***Keywords*—Earthquake magnitude prediction, Transfer learning, Resnet18 , Residual network**

I. INTRODUCTION

In recent years, with the continuous accumulation and improvement of earthquake data, as well as the emergence of deep learning algorithms, the application of artificial intelligence technology in the field of seismology has entered a new stage. Magnitude, as a measure of earthquake intensity, is one of the important parameters that seismological research needs to consider comprehensively. Therefore, predicting earthquake magnitudes is a highly challenging and crucial issue for seismology [1]. The difficulty of earthquake magnitude prediction lies in the insufficient number of stations used [2], small epicenter distance, missing training samples, and the ability to only utilize underdeveloped P-wave information. In addition, the lack of some earthquake label data and the uncertainty of magnitude evaluation standards are obstacles to the progress of magnitude prediction. These issues may lead to incomplete seismic data and low accuracy in magnitude prediction at present.

Machine learning technology has been proven to be a powerful and effective tool for seismic data analysis and computation. Machine learning can process large amounts of data, including seismic phase picking [3], seismic identification [4], determination of seismic source mechanism [5], seismic magnitude estimation, etc., and is widely used in research in the field of earthquake. Machine learning based methods have achieved great success in achieving advanced performance, primarily due to the availability of large-scale and precisely labeled training datasets. Neural networks have been proven to be powerful tools for seismic signal processing and characterization [6].

The key to the application of machine learning in the field of seismology is data processing, model selection and establishment, and result analysis. With the continuous development of seismology and machine learning, researchers have made the following progress. In China, Liu Tao[7] uses Convolutional Neural Networks (CNN) to train and test the model with seismic acceleration information as input. This method uses M5.5 as the boundary for predicting large and small earthquakes, with an accuracy of 92.3%. However, there are varying degrees of overfitting during the training process, and the acceleration time history records of earthquakes only contain certain earthquake magnitude information, There is an issue of missing raw label data. Similarly, CNN was used to improve the model input. The epicenter distance and depth were used as the corresponding seismic event source information for training as a new dataset. A multi fully connected convolutional neural network model was constructed [8], with a fault tolerance of $\pm$ 0.5. When the initial wave truncation period was 3 seconds, the overall accuracy of the model's magnitude prediction was 89.92%. When the truncation period was 9 seconds, the overall accuracy reached 96.08%. Although this method has a high accuracy, as the window scale continues to increase, there may be a situation where the magnitude prediction results are generally low. Lin Binhua [9] constructed a CNN magnitude prediction model with a 3s waveform input to solve the overfitting phenomenon caused by a large amount of data. He improved the establishment of the magnitude prediction model using 3s waveform data, transforming the prediction problem into a classification problem. The model was tested and analyzed using a new earthquake example in 2019, and the accuracy of the magnitude error range for a single station within $\pm$ 0.3 was 85.6%, The method failed to analyze noise and waveform data within 3 seconds, resulting in overestimation of magnitude prediction results. Chen Wanghao [10] used a method that combines Long Short Term Memory (LSTM) neural networks with CNN. This method also transforms the problem of earthquake magnitude prediction into a classification problem, and combines the excellent image feature extraction ability of convolutional neural networks with the learning and improvement ability of Long Short Term Memory neural networks in time series to study. Strong earthquake data is used as supplementary information for earthquake magnitude prediction, Its method can still achieve an accuracy of 87.12% even with a small amount of data. Abroad, Mousavi [11] designed a regressor consisting of

This work is a national key R & D program support project, project number: 2023YFC3008904

Authorized licensed use limited to: Chaitanya Bharathi Institute of Tech - HYDERABAD. Downloaded on August 23,2024 at 09:01:32 UTC from IEEE Xplore. Restrictions apply.

CNN and Recurrent Neural Network (RNN) to estimate earthquake magnitude using waveform front-end and amplitude information. This regressor is not sensitive to data normalization, so it can utilize waveform amplitude information during the training process. This model can predict local amplitude and magnitude, with an average error close to zero and a standard deviation of approximately 0.2. Apriani [12] used a deep neural network (DNN) for analysis, using a 4-second raw seismic waveform as input. This experiment was compared with a random forest (RF), demonstrating that DNN has better optimization and generalization abilities in large-scale data processing, with a magnitude prediction accuracy of 76%.

Based on the existing problems mentioned above, a transfer learning model is proposed to transfer knowledge to target classification tasks with a small amount of data, in order to solve problems in different related fields. This article pre trains the Resnet18 model on the Imagenet dataset, performs transfer learning on the microseismic waveform dataset, extracts model features, and then puts the extracted feature vectors into the classifier for training. This method uses seismic waveforms as inputs for prediction, transforming the prediction problem into a classification problem, and further optimizing the Resnet18 network parameters to obtain new classification results, improving the accuracy of the prediction model.

II. Data Preprocessing

This study used the STanford Earthquake Dataset (STEAD) data for experimentation. STEAD[13] is a globally labeled seismic map dataset. The STEAD dataset consists of two types: (1) local seismic waveforms (recorded at local distances within 350 kilometers of an earthquake), (2) Seismic noise waveform without seismic signal. These data together include approximately 1.2 million time series or seismic signal records exceeding 19000 hours.

This study selected local seismic waveforms, which contain approximately 1.05 million three component seismic maps (each with a duration of 1 minute), and each seismic waveform contains 35 attributes (labels). Mainly including: source time function, epicenter position, depth, magnitude, magnitude type, arrival time of P and S waves, etc. This experiment used four attributes: magnitude, P-wave arrival time, S-wave arrival time, and waveform stop time.

A. *Sample Selection*

In the selection of samples, this study selected 9000 microseismic signal waveforms with earthquake magnitudes below 3.0. The sample data was taken from relevant waveforms in continuous time series archived by the Integrated Research Institute for Seismology (IRIS). The waveform information includes P-wave and S-wave, and starts 5 to 10 seconds before the arrival of P-wave and ends at least 5 seconds after the arrival of S-wave. The arrival time of P-wave, S-wave, and waveform stop time were marked in the waveform image for subsequent magnitude classification of the image. The ratio of training set to test set is 8:2.

B. *The Magnitude of the Tag*

The allowable error for earthquake magnitude is usually within ± 0.3 [14], so earthquake magnitude prediction can be regarded as a classification problem, and thus earthquake magnitude classification can be carried out. The magnitude classification of this earthquake is based on the magnitude difference of 0.1, which is divided into 30 magnitude levels from 0.1 to 3.0, and each magnitude is guaranteed to have labeled data. The specific classification situation is shown in Table 1. In the actual prediction process, assuming that the model is identified as label 5, as shown in Table 1, the earthquake magnitude is (0.4, 0.5]. In this experiment, it is specified that the final predicted value should be taken as the middle value of the magnitude interval, so it is 0.45. There may be a system error of ± 0.05 in the magnitude error, which is within an acceptable range.

Table.1 Magnitude classification and labels

lable	Magnitude	lable	Magnitude	lable	Magnitude
1	0.0～0.1	11	1.0～1.1	21	2.0～2.1
2	0.1～0.2	12	1.1～1.2	22	2.1～2.2
3	0.2～0.3	13	1.2～1.3	23	2.2～2.3
4	0.3～0.4	14	1.3～1.4	24	2.3～2.4
5	0.4～0.5	15	1.4～1.5	25	2.4～2.5
6	0.5～0.6	16	1.5～1.6	26	2.5～2.6
7	0.6～0.7	17	1.6～1.7	27	2.6～2.7
8	0.7～0.8	18	1.7～1.8	28	2.7～2.8
9	0.8～0.9	19	1.8～1.9	29	2.8～2.9
10	0.9～1.0	20	1.9～2.0	30	2.9～3.0

C. *Data Preprocessing*

Effective data augmentation can effectively prevent model overfitting issues for the images in the training set. Randomly scale and crop the image to 256 first × 256 square image, crop 224 from the center of the square image × The small square of 224 is transformed into a Tensor tensor, and the resulting image is enlarged or reduced according to a predefined range to avoid the problem of losing details. Then, image enhancement is performed on the image, such as flipping, rotating, and adding noise. Applying noise involves changing some pixels of the image while maintaining its overall practical significance. Finally, normalization is performed on the input data, which can effectively solve the problem of model overfitting and quickly improve the convergence of the model. Therefore, the Normalize function is used for batch normalization of data in model training.

For the images in the test set, image enhancement is not performed during the same processing as the training set. The image enhancement performed on the training set ensures that the model has sufficient robustness and is not affected by the predicted image rotation and folding, thus preventing overfitting of the model.

When performing forward prediction of the image, the logits prediction scores for 30 magnitude categories are obtained. Logits is the logarithm of the ratio of events that occur to those that should not occur. In the last layer of this model, it is a linear classification layer, and its output is a Logits that can be positive, negative, or large or small. Assuming the probability of an event occurring is p, then the logits of the event are:

Authorized licensed use limited to: Chaitanya Bharathi Institute of Tech - HYDERABAD. Downloaded on August 23,2024 at 09:01:32 UTC from IEEE Xplore. Restrictions apply.

$$\mathrm{logits}(p) = \log \frac{p}{1-p} \tag{1}$$

After obtaining the logits, perform the softmax operation, and the softmax layer will normalize the input to obtain the probability distribution:

$$p(i) = \frac{e^{a_i}}{\sum_{j=0}^{C} e^{a_j}} \tag{2}$$

Among them is the logits in tensorflow. By using the softmax function, convert the logits of 30 categories into confidence levels for 30 categories, with each confidence level becoming a number between 0 and 1, and the sum of the confidence levels for the 30 categories is 1.

III. NETWORK MODEL CONSTRUCTION

A. *Resnet Model Construction*

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

The research process based on Resnet transfer learning is shown in the figure, including data preparation, model construction, model visualization, and evaluation. The microseismic image dataset is divided into training, validation, and testing.

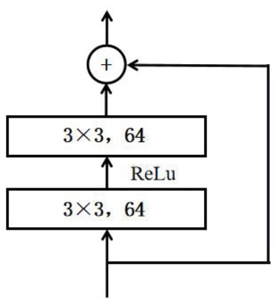


Fig.2 Residual module structure diagram

As the depth of neural networks deepens, problems such as gradient vanishing and degradation arise, which lead to difficulties in model training convergence and low accuracy [15]. In order to establish an effective connection between input and output, this study used residual modules for training. The introduction of residual networks not only solves the problem of vanishing or exploding gradients caused by deepening layers, but also maintains its ability to express features while expanding depth. The module specific residual block structure is shown in Figure 2.Where the specific residual block structure is shown in Figure 3.

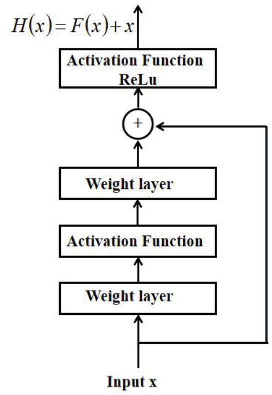


Fig.3 Residual network structure diagram

In order to maintain accuracy in the later layers without causing a decrease, the input x is directly used as the initial result, i.e. F (x)=0, through the shortcut connection of H (x)=F (x)+x. Without increasing network parameters and computational complexity, its structure is quickly connected, enhancing feature transmission. Therefore, in this experiment, the Resnet18 model was chosen based on the characteristics of the data type. This article has improved the fully connected layer network structure, and the improvement results are shown in Figure 4. The 1-layer structure in the left figure has been changed to a 3-layer structure, with each layer structure undergoing a dimensionality reduction. The output results of different layers are selected as the final feature vectors to be extracted according to different needs.

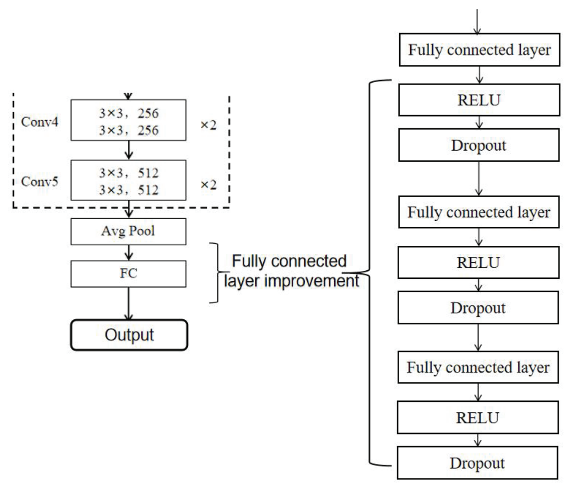


Figure 4 Structural diagram of the fully connected layer

B. *The Resnet Model Based on Transfer Learning*

Due to the large number of parameters and limited training data, insufficient training can lead to very low classification accuracy achieved by deep neural networks. The combination of pre training and fine-tuning is a very effective image classification transfer learning method. As shown by the accuracy of all networks fine tuned on the original dataset, fine tuned networks can easily achieve good accuracy. Transfer learning requires similarity in the feature spaces of the source and target domains. The specific method is to first pre train the model with a large amount of data from the source domain, then transfer the obtained weight parameters, and finally retrain the fully connected layer with a small amount of target data. The transfer learning process of the model is shown in Figure 6.

Authorized licensed use limited to: Chaitanya Bharathi Institute of Tech - HYDERABAD. Downloaded on August 23,2024 at 09:01:32 UTC from IEEE Xplore. Restrictions apply.

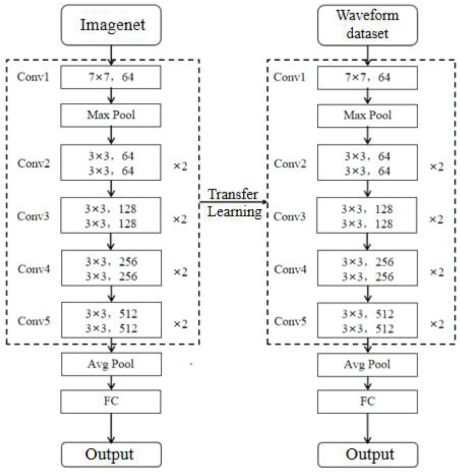


Fig. 6 Structural diagram of the transfer-learning model

The reason for using pre-trained models is as follows: Firstly, training a large model on a large dataset requires more computational power. Secondly, the time spent on training the network is too long, lasting several weeks. Training a new network with pre-trained weights can accelerate the learning process. When conducting transfer learning, there are three strategies to compare and use. The adoption of different transfer learning strategies depends on how different the distribution of the predicted dataset is from that of the large network (Imagenet).

In practical applications, there are three main strategies for transferring Resnet pre-trained models: first, using the Resnet pre-trained model as a feature extractor, freezing the fully connected layer during fine-tuning training, only fine-tuning the last layer of the training model, and using the extracted features to train the classifier to achieve model transfer.

The second is to fine tune the training of all layer strategies. Use the pre trained Resnet18 model as the initialization weight, and after modifying the classification layer, fine tune the weights on all layers of the model. The third is to only fine tune the parameters of the last fully connected layer (fully connected classification layer) of the training model, freeze the other layers, and modify the fully connected layer after loading the pre training model, so that the output of the fully connected layer corresponds to the current number of categories in the dataset. The microseismic waveform image dataset is used as a new input to fine tune the weights of the Resnet model, thereby achieving model transfer. However, Resnet's pre training is conducted on Imagenet, mainly on waveform images, which have significant differences from the original images, and most training targets have relatively different features. This method is suitable for datasets with significant differences in distribution compared to Imagenet. Therefore, this article selects the third weight fine-tuning strategy for transfer learning.

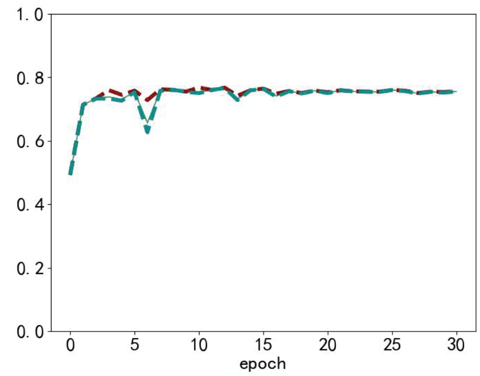
IV. Experimental Application Research

A. Analysis of model results

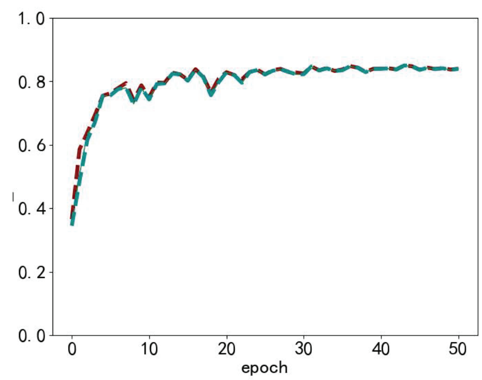
In this study, transfer learning and convolutional neural network were used. Since the magnitude prediction problem is transformed into an image classification problem, the test set accuracy (training accuracy) and F1 score (F1-score) are therefore used to evaluate the training effect of the ResNet transfer learning model. The respective definitions are given below:

$$\begin{cases} Accuracy = \dfrac{TP+TN}{P+N} \\ Precision = \dfrac{TP}{TP+FP} \\ Recall = \dfrac{TP}{TP+FN} = \dfrac{TP}{P} \\ f1 = 2 \times \dfrac{Recall \times Precision}{Recall + Precision} \end{cases} \tag{3}$$

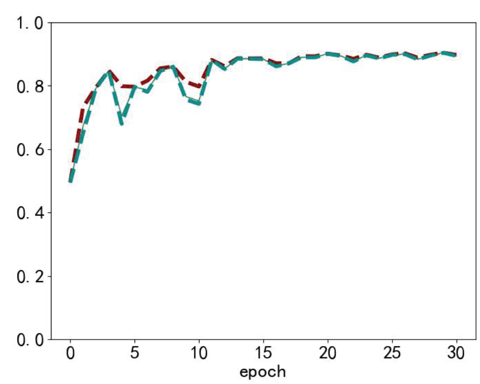
Where TP is the number of true cases, FP is the number of false positive cases, FN is the number of false negative cases, P is the total number of positive cases, and N is the total number of negative cases.



(a) Strategy I model classification evaluation



(b) Strategy II model classification evaluation



(c) Strategy III model classification evaluation

Fig.7 Different model training result graphs

Authorized licensed use limited to: Chaitanya Bharathi Institute of Tech - HYDERABAD. Downloaded on August 23,2024 at 09:01:32 UTC from IEEE Xplore. Restrictions apply.

The test results of the model are shown in Figure 7, and all three images show the results of using Resnet18 for transfer learning. Among them, strategy three has the best effect.

In the experiment, ResNet18 was used to conduct the comparison experiment. The specific magnitude classification accuracy of the six methods is shown in Table 2. In them, the ResNet transfer learning model configuration reached the expected results, and the accuracy of training set and test set were continuously improved with the increase of training time, and finally converged to a high level. Among them, the first configuration accuracy of ResNet18 transfer learning model is 75.44%, the second configuration accuracy of ResNet18 transfer learning model is 80.47%, and the first configuration accuracy of ResNet18 transfer learning model is 85.10%.

The specific data for other results are shown in Table 2, with CNN method and Resnet50 as comparative experiments. In conclusion, the model using Resnet18 as a pre-trained model for transfer learning has a higher prediction accuracy, about 3%, which is more suitable for subsequent microshock magnitude prediction.

Table 2 Accuracy Table of Earthquake Magnitude Classification

Method	Accuracy%	Precision%	Recall%
CNN	79.04%	75.71%	72.13%
TL-M1(resnet18)	75.44%	71.27%	72.52%
TL-M2(resnet18)	80.47%	79.55%	79.18%
TL-M3(resnet18)	85.10%	83.48%	82.60%
TL-M1(resnet50)	72.91%	70.10%	70.97%
TL-M2(resnet50)	76.63%	75.86%	75.37%
TL-M3(resnet50)	82.65%	81.37%	84.00%

B. *Predicted results analysis*

In this study, in order to more intuitively show the error distribution of the estimated magnitudes of different models, the magnitude prediction results were made, and the deviation of the results was recorded as shown in Figure 5. In these, 300 data were selected for prediction. For clear contrast results, do the auxiliary lines y=x + 0.3 and y=x-0.3 in Fig.8. As can be seen in the figure, the traditional CNN model has the lowest accuracy in the more refined magnitude classification, with the most deviations from ± 0.3 and the largest discretization. Among Resnet50, as a comparison experiment, its model accuracy, prediction accuracy and classification performance evaluation indexes are inferior to Resnet18 model.

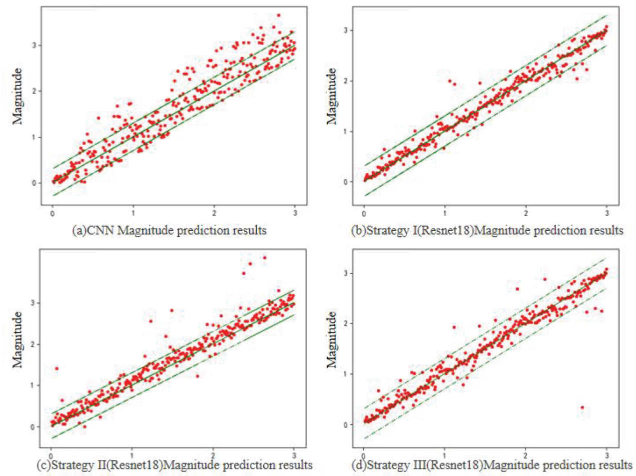


Fig.8 Magnitude prediction result

The training set with transfer learning strategy 3 has the least deviation from ± 0.3 and has less discretized, most concentrated on the diagonal of y=x, and the highest model accuracy. The accuracy of using Resnet18 was generally higher than Resnet50. Therefore, the experiment finally chooses to randomly initialize all the weights of the model, train all layers from scratch, select Resnet18 as the pre-trained model for prediction, the error in ± 0.3 is 282, the accuracy is 94%, the error in ± 0.2 is 241, the accuracy is 80.3%, and the experimental variance is 0.16237. Specific data from the six experiments are shown in Table 3.

Table 3 Error Table for Magnitude Classification

	sample number	Error <0.3 sample number	Error <0.2 sample number	variance
TL-M1 (resnet18)	300	263	186	0.29938
TL-M2 (resnet18)	300	269	210	0.23717
TL-M3 (resnet18)	300	282	241	0.16237
TL-M1 (resnet50)	300	207	155	0.31531
TL-M2 (resnet50)	300	231	165	0.24238
TL-M3 (resnet50)	300	262	193	0.22595
CNN	300	237	178	0.32993

V. CONCLUSION

As one of the three elements of earthquake, earthquake magnitude is also one of the most basic parameters in the field of seismology. The magnitude of the earthquake plays a direct measure of the post-earthquake disaster assessment. This paper uses residual network and transfer learning, uses waveform data as data source, transforms the prediction problem into classification problem, avoids the problem of missing label data and gradient disappearance in the training process, and draws the following conclusions:

For magnitude estimation, most studies use data calculation, while relatively few studies use waveform image data. In this paper, the model using pre-trained mature deep network model weights as initial weights simplify the training process and extracts deep features from them. At the same time, avoid the traditional feature

Authorized licensed use limited to: Chaitanya Bharathi Institute of Tech - HYDERABAD. Downloaded on August 23,2024 at 09:01:32 UTC from IEEE Xplore. Restrictions apply.

extraction methods on the specific attribute data dependence, reduce the complexity of the process.

REFERENCES

[1] Ma Chinzhong. Enlightenment of several important earthquake prediction and prediction results [J]. Journal of seismology, 2014,36 (3): 14.

[2] Li Shan has it. Approaching the earthquake early warning [J]. Disaster Prevention Expo, 2018 (02): 14-23

[3] Simons F J , Dando B , Allen R M . Automatic detection and rapid determination of earthquake magnitude by wavelet multiscale analysis of the primary arrival[J]. Earth & Planetary Science Letters, 2006, 250(1-2):214-223.

[4] Colosimo, G.; Crespi, M.; Mazzoni, A. Real-time GPS seismology with a stand-alone receiver: A preliminary feasibility demon-stration. J. Geophys. Res. 2011,116,B11302.

[5] M A,K S W,Daryono. Earthquake Magnitude Estimation Based on Machine Learning: Application to Earthquake Early Warning System[J]. Journal of Physics: Conference Series,2021,1951(1)

[6] Li Z , Meier M A , Hauksson E , et al. Machine Learning Seismic Wave Discrimination: Application to Earthquake Early Warning[J]. Geophysical Research Letters, 2018.

[7] Liu Tao, Dai Zhijun, Chen Su, etc. Deep learning-based seismic magnitude classification [J]. Journal of seismology, 2022,44 (04): 656-664.

[8] Wei Dongliang, Wang Yanwei, Wang Zifa, etc. Rapid seismic magnitude estimation method based on convolutional neural network [J]. Journal of seismology, 2022,44 (02): 316-326.

[9] Lin Binhua, Venus, Kang Lanchi, Wei Yongxiang, Li Jun, Zhang Yanming, Chen Huifang, Zhou Shiwen. Research on earthquake magnitude determination based on convolutional neural network [J]. Journal of Geophysics, 2021,64 (10): 3600-3611

[10] Chen Wanghao, Yin Ling, Li Feibin, Yang Zihao, Ma Bangjie. Research on earthquake magnitude prediction based on CNN-LSTM [J]. Science and Technology and Innovation, 2020 (11): 53-54. DOI: 10.15913/j.cnki. kjycx. 2020.11.018

[11] Mousavi S M , Beroza G C . A Machine Learning Approach for Earthquake Magnitude Estimation[J]. Geophysical Research Letters, 2019, 47(1).

[12] M A,K S W,Daryono. Earthquake Magnitude Estimation Based on Machine Learning: Application to Earthquake Early Warning System[J]. Journal of Physics: Conference Series,2021,1951(1).

[13] JIA D， WEI D， SOCHE R R， et al． ImageNet: A Large-scale Hierarchical Image Database ［C］ // IEEE Conference on Computer Vision and Pattern Recognition． Miami: IEEE， 2009: 248－255．

[14] Zhao Yang, Liang Yingchun, Xu Jun, Li Dazhou. Improved ResNet18 network model for flower recognition [J]. Computer Technology and Development, 2022,32 (07): 167-172.

[15] Nan Fangfang, Zhang Feng, Gong Gubin, etc. Analysis of automatic earthquake magnitude deviation in Xinjiang [J]. Seismic geomagnetic observation and research, 2019,40 (04): 45-52.

Authorized licensed use limited to: Chaitanya Bharathi Institute of Tech - HYDERABAD. Downloaded on August 23,2024 at 09:01:32 UTC from IEEE Xplore. Restrictions apply.