Fault Structure Interpretation Based on Deep Learning—Risk Assessment of Agricultural Geological Disaster

Pei Yao1,2,a, Jofrisse Cremilda Samuel1,b*

1 College of Geo-exploration Science and Technology, Jilin University, Changchun, China
2 Geoscience Big Data Analysis and Application Technology Innovation Center of Ministry of Natural Resources, Jilin University, Changchun, China

First author’s e-mail: peiyao@jlu.edu.cn
*Corresponding author’s e-mail: cremildajofrisse@hotmail.com

Abstract. China is a large agricultural country, shouldering the heavy responsibility of agricultural production, facing great pressure and challenges under the influence of geological disasters. Frequent agricultural geological disasters have a serious impact on people's agricultural production and life. The complex geological structure of agricultural production areas, especially the wide distribution and large number of regional fault structures, especially the lack of popularization of geological disaster knowledge among rural residents, is also an important cause of geological disasters. Fault interpretation is the core of complex structure interpretation. The efficiency and degree of interpretation directly affect the progress of agroecological hazard risk assessment.

1. Introduction

In recent years, artificial intelligence has been applied in many fields and achieved good results. Aiming at the risk assessment of agricultural geological disasters, this work attempts to use gravity and magnetic data for fault recognition. This paper presents a method of fault prediction using convolution neural network. The training set is constructed by gravity and magnetic data and artificial interpretation of fault tags; Combined with the characteristics of the fault in the actual work area, the corresponding convolution neural network is constructed; through the experimental analysis to determine the relevant parameters of the network, and network model training, using the trained network model to predict the fault of the actual work area gravity and magnetic data, the accuracy of fault prediction results is significantly higher than the traditional method.

2. Construction of convolution neural network model

The deep learning algorithm used in this paper is AlexNet[1], which is designed by Hinton, the winner of ImageNet competition in 2012, and his student Alex Krizhevsky. Compared with the traditional machine learning algorithm [2][3][4], it is more efficient, and can extract more image features, making the training results better [5]. The model structure of AlexNet algorithm is shown in Figure 1.

Firstly, we preprocess the 1:50000 aeromagnetic data of a certain area in Hebei Province by using data enhancement technology and normalization after depolarization. The square pixel block of 31 pixels high and 31 pixels wide centered on each sampling point is taken as 31 × 31 × 3; The convolution neural network model is constructed, and the parameters of the model are compared through training.
Finally, the Adam optimization function is used, the initial learning rate is set to 0.01, and the activation function is Relu function. The convolution neural network is used in this paper. After a series of feature extraction, the final output is calculated by sofamax to get the category probability. The sum of the probabilities is 1, and the larger value is used to indicate the category of the sampling point, so as to identify the fault. Corresponding to a single data for fault identification, the organic fusion of three different geophysical data will greatly improve the training accuracy. Through the test of the trained model, the results verify the feasibility of using convolution neural network for fault recognition.

![Figure 1 Structure Diagram of AlexNet Model](image)

In the process of model training, each sampling point in the data is taken as the center, and a small image block is formed with the pixels around it. The interpretation result of the center pixel, that is, whether it is a fault, is used as the label of the image block. As shown in Figure 2, after normalizing the 1:250000 aeromagnetic data, the image is taken as the input and whether there are fault points is taken as the output. 19437 image blocks representing fault and 19437 image blocks representing non fault are randomly extracted. Fault recognition can take a two-dimensional or high-dimensional array of each point and some related points around as an image, and take the fault result as the label corresponding to the center point of the image.
3. Model training and analysis

The test environment used in this paper is inter (R) core™ I9-10900 CPU, 64g memory, NVIDIA geforce GTX3080, 10GB video memory, Matlab programming language, using CPU for data reading, preprocessing, GPU for model training acceleration. Our data set size is about 40000, and each sample size is 31 × 31. It is divided into training set and verification set according to the ratio of 7:3. 10 epochs are trained iteratively, and each interval of 60 iterations is verified on the verification set. The prepared data are input into the convolutional neural network according to the requirements of network format. 70% of the data are used as training data, and the remaining 30% are used as test data. After iterative training, the precision curves of training set precision and test set precision with the number of iterations are shown in Figure 3 (a), and the loss curves of training set and test set precision with the number of iterations are shown in Figure 3(b).

Figure 3 Curve of Training Precision and Loss Function
In the training data of artificial interpretation of faults, the number of samples labeled as non-faults is much larger than that labeled faults. If the training data of non-fault data is much more than fault data are input into the network for model training, it is assumed that the network output results are all non-fault. When calculating the accuracy rate, the accuracy ratio of the final network training is always higher because the proportion of non-fault data is relatively large, that is to say, the classification of prediction bias samples is more. This greatly reduces the generalization ability of the model. Therefore, before training the network, the sample equalization is needed, and the same number of non-fault points are randomly selected according to the number of fault points, as shown in Figure 4.

![Figure 4 Number of Samples](image1)

![Figure 5 Confusion Matrix of Training Data](image2)

In the field of deep learning, obfuscation matrix is a kind of visual effect to show the performance of algorithm through characteristic matrix, so obfuscation matrix is also called possibility table or error matrix. Input the data from the training set into the trained convolutional neural network model for fault prediction, and then draw the confusion matrix between the predicted result and the actual fault label. The confusion matrix is shown in Figure 5. The data on the diagonal line from the upper left corner to the lower right corner in the figure are classified correctly, and other positions are classified incorrectly. The sum of the data in the confusion matrix represents the total sample. Therefore, according to the definition of the secondary index standard accuracy of the confusion matrix, the accuracy of the verification training set is about 90.98%, which shows that the network has a good effect in fault prediction.

4. Application of the model in Chongli ore concentration area

In practical work, using deep learning method to train the model, we usually hope that it not only performs well for the known data (training set), but also for the unknown data, that is, it has good generalization ability. The characteristics of fault structure information on gravity and magnetic maps are as follows: (1) the isolines of magnetic anomalies are linear and densely distributed, with monotonicity, inflection point and extremum asymptote (2) The background fields on both sides of the boundary with different characteristics are obviously different, and the strike and shape are significantly different (3) The distribution zone of linear anomaly or beaded anomaly (4) Abnormal common dislocation line or regular twisting line (5) The boundary of different geological units. Therefore, the application effect of our previously trained model in the new area, whether the gravity and magnetic data of different scales are still valid, whether the aeromagnetic data of the new area can be used to speculate the fault structure, that is, whether the method of deep learning for fault structure has generalization, is also an important content of our work. In order to test the effectiveness of the proposed method, we take the 1:50000 aeromagnetic data in Chongli area of Hebei Province as an example to verify the effectiveness and generalization ability of the method. According to the collected 1:50000 geological and mineral maps of Chongli sheet, Changyukou sheet and Zhaochuan sheet, faults are
relatively developed in Chongli ore concentration area of Hebei Province. The representative fault in the study area is Chongli Chicheng deep fault in the north of Chongli sheet, which belongs to the eastern part of Shangyi Chicheng deep fault. In the north of Chongli sheet, the East and west sides of Chongli Laohugou line extend into the adjacent sheet respectively.

The 1:50000 aeromagnetic map of Chongli ore concentration area in Hebei Province is analyzed by deep learning. As shown in Figure 6, the East-West Chongli Chicheng deep fault in the north of Chongli sheet is well recognized. The NW and NE trending brittle faults exposed in Changyukou and Zhaochuan sheets are also well recognized. Most of the known field verified faults have been well recognized in this artificial intelligence recognition. Specifically, 80 faults have been identified in the field. In this work, 66 faults have been identified in the field, and the recognition rate is 82.5%. Generally speaking, the recognition effect of NW, NE and EW faults is better, while the recognition effect of Sn faults is slightly worse. The reason is that the previous training model lacks Sn faults. The recognition ability of the model for faults in different directions can be effectively enhanced by increasing the training samples of Sn faults. The main advantage of deep learning algorithm over traditional algorithms is to continuously improve the robustness and generalization ability of the model through the supplement of samples.

Figure 6 Effect of Deep Learning Method on Fault Identification in Chongli Ore Concentration Area, Hebei Province

5. Conclusion
The work of this paper is a beneficial attempt to extract fracture from gravity and magnetic data by using deep learning method, and the effectiveness of the method is preliminarily verified by establishing the model. And because the gravity and magnetic field information corresponding to the fault boundary is mainly the linear dense distribution zone of magnetic anomaly isoline, or the boundary of different characteristic anomaly field, the model has a certain generalization after it is established, and can identify the fault information of different scales. With the increase of the number of learning models, the recognition accuracy of the proposed method will increase accordingly, which is also the advantage of deep learning with big data.
Acknowledgments
This research was funded by the Geological Survey Project of China Geological Survey (DD20160050) and the basic scientific research project of Jilin University (Knowledge Graph research based on geological and mineral survey). We would like to thank reviewers for carefully reading this paper and their very useful comments.

References
[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
[2] Ye X W, Jin T, Yun C B. A review on deep learning-based structural health monitoring of civil infrastructures. Smart Structures and Systems. 2019(5): 567-585.
[3] Sirajudeen A, Ezhilarasi M. Multi-scale LBP and SVM Classification to Identify Diabetic Retinopathy in Lesions. CURRENT MEDICAL IMAGING REVIEWS. 2015(11):49-55.
[4] Ding F J, Jia X D, Hong T J. Flow Stress Prediction Model of 6061 Aluminum Alloy Sheet Based on GA-BP and PSO-BP Neural Networks. RARE METAL MATERIALS AND ENGINEERING. 49(6): 1840-1853.
[5] Shi X Y, Qiu G, Yin C. An Improved Bearing Fault Diagnosis Scheme Based on Hierarchical Fuzzy Entropy and Alexnet Network. IEEE ACCESS. 2021(9): 61710-61720.