CNN Based Fault Recognition with Multi-Scale Fusion Attention Mechanism

Bo Wang\(^1\) and Jinwen Ma\(^2\)

\(^1\)Department of Information and Computation Science, School of Mathematical Sciences, Peking University, Beijing, China. E-mail: 1701210114@pku.edu.cn

\(^2\)Department of Information and Computation Science, School of Mathematical Sciences, Peking University, Beijing, China. E-mail: jwma@math.pku.edu.cn

Abstract. In geological exploration, seismic data interpretation is very important and its key task is fault detection or recognition. In fact, the accurate recognition of fault shape and distribution plays an important role in oil and gas exploration. With the increasing demand of fault interpretation accuracy in oilfield, traditional manual fault interpretation methods are susceptible to subjective factors of interpreters. With the rapid development of deep learning technique, convolutional neural networks (CNNs) has successfully applied to the complicated tasks of image classification. This paper utilizes proposes a CNN based fault recognition framework with a multi-scale fusion attention mechanism. It is demonstrated by the experimental results that our proposed approach is superior to the traditional methods on both the fault interpretation accuracy and efficiency of 3D seismic images, greatly reducing the risk of oil and gas exploration and development. With a GeForce GTX 1080 GPU, the training processing takes about 4 hours and predicts the faults in a 33×33×33 seismic volume using only tens of milliseconds.

1. Introduction

In seismic data interpretation, fault recognition is a very challenging task. Geological faults are important and valuable because they are usually related to the formation of underground traps where oil may accumulate. In the early studies of fault recognition or identification, some conventional methods were used to strengthen the fault characteristics of seismic data, such as the median filtering and the improved numerical fitting analysis method like Randen [1]. In the middle of 1990, Farmer and Bahorich M [2] established the coherent analysis method for 3D seismic data, which is useful for fault analysis. In recent years, different kinds of fault recognition methods have been proposed such as the likelihood function method [3], the optimal surface voting method [4], the end-to-end U-Net method [5] and so on.

Since the VGG (Visual Geometry Group) [6] network, as a recently developed architecture of CNN, took the first place in the positioning task and the second place in the classification task in ILSVRC-2014, it has been widely and successfully applied in various tasks of image classification and object recognition. Its outstanding contribution lies in the use of a small convolution kernel 3×3, which effectively improves the effect of the architecture by increasing the network depth. In fact, the VGG convolution neural network owns good generalization capability for many datasets.

In the field of neural networks, attention mechanism [7] has been developed as a resource allocation scheme which allocates the computing resources to more important tasks for solving the information overload problem with the limited computing power. By introducing an attention mechanism, we can focus on the pieces of input information that are more critical to the current task among all the pieces of input information, reducing the attention to the other pieces of information, so that the information...
overload problem can be solved with the efficiency and accuracy of task processing being improved. As the network deepens, some information will be lost in each layer, and more information will be lost in the last layer. The multi-scale fusion attention mechanism with adding and contacting operations [8] can increase the amount of information for the image features of the task.

Based on the seismic data of eight attributes of the earthquake, we propose a framework of the improved VGG convolutional neural network with the multi-scale fusion attention mechanism for fault recognition. Moreover, a balanced cross-entropy loss function is used to optimize the parameters of the network to overcome the problem of class imbalance. Finally, we achieve the state of the art results.

2. Fault recognition of improved VGG based on multi-scale fusion attention mechanism

With the rapid development of artificial intelligence and deep learning, the emergence and continuous improvement of convolutional neural network have made it one of the increasingly popular algorithms for supervised learning. At present, in many specific research areas such as pattern recognition, convolutional neural network do not need to manually select image related features, and make complex analysis and processing. We can directly input the original image to the network where the successive convolution and pooling operations help to perform the feature extraction and recognition after the supervised training with the artificial labeled data has been accomplished.

2.1. Framework design principle

The convolution operation is performed with the input of the previous layer by using different convolution filter to obtain feature map. The activation function realizes the non-linear transformation of the signal in the convolutional network, which improves the non-linear fitting ability of the entire network model. The pooling layer reduces the size of the feature vector, amount of parameters and effectively prevents overfitting. The fully connected layer is also a classifier layer. After softmax processing, the category labels with the highest probability corresponding to the image are output.

2.2. Improved VGG convolutional neural network

We utilize an improved VGG convolutional neural network as follows. The seismic data are organized into three-dimensional data (3D) form and the convolution operations are also 3D form. We select the number of layers according to the fault recognition task. As a result, we get the following best architecture of the improved VGG convolutional neural network.

![Figure 1. The improved VGG convolutional neural network for 3D fault recognition.](image)

The loss function of the network is designed by

\[
L = -\frac{1}{N} \sum_{i=1}^{N} [\beta y_i \log(p_i) + (1 - \beta) (1 - y_i) \log(1 - p_i)].
\]

(1)

Where \( \beta = \frac{\sum_{i=1}^{N}(1-y_i)}{N} \) is the number of negative classes and \( (1 - \beta) \) is the number of positive classes.
2.3. Improved VGG convolutional neural network with multi-scale fusion attention mechanism

In order to improve the efficiency of the neural network, the attention mechanism tries to select certain important pieces of input information for processing. For example, in a machine reading comprehension task, a long article is given, while the content of the article is then asked. However, the raised questions may be only related to one or two sentences in the paragraph and the rest are irrelevant, so you only need to pick out the relevant fragments for the neural network to process, instead of entering all the content of the article into the neural network. In this way, we can improve the VGG model by calculating the attention values or weights of the pieces of input information into two steps: Firstly, we calculate the attention distribution on all the pieces of input information; Secondly, we calculate the weighted average of each piece of input information according to the attention distribution.

Firstly, we define an attention variable $z \in [1, N]$ to represent the index position of the selected piece of input information. That is, $z = i$ is to indicate that the $i$-th piece of input information is selected, we then calculate that, given a query $q$ and the input information $X = [x_1, \ldots, x_N]$, the attention value or selecting probability $\alpha_i$ of the $i$-th piece of input information is given by

$$\alpha_i = p(i|X, q) = softmax(s(x_i, q)) = \frac{\exp(s(x_i, q))}{\sum_{j=1}^{N} \exp(s(x_j, q))},$$

(2)

The attention distribution consists of these $\alpha_i$, i.e., $\alpha = [\alpha_1, \ldots, \alpha_N]$. $s(x_i, q)$ is the attention scoring function that is expressed by

$$s(x_i, q) = x_i^T q,$$

(3)

Secondly, $\alpha_i$ represents the degree of correlation between the $i$-th piece of input information and the query $q$. The “soft” information selection mechanism is used to give the query result, which is to process all the pieces of input information using a weighted average method to obtain the attention value:

$$Attention(X, q) = \sum_{i=1}^{N} \alpha_i x_i.$$

(4)

The following figure 2 shows the processing of calculating the attention value:

![Figure 2. The flow diagram of the attention mechanism.](image)

We concatenate the outputs of the first three stacks of the VGG convolutional neural network. By deconvoluting the output of the last stack (a type of convolution, equivalent to transposed convolution [9]), we then make dot multiplication separately with each of the first three stacks, and then implement the convolution operations to keep the same output size as the last stack, finally add these output as intermediate results. Our architecture with multi-scale fusion attention mechanism is shown in figure 3.
3. Experimental results

We use 200,000 tagged real data in 360(crossline)×1036(inline)×3500(vertical) real volume dataset for training, validation and testing. Among them, the positive class has 140,000 data, while the negative class has 60,000 data, the training and testing ratios are set to 0.8 and 0.2, respectively. The parameters are configured as: training data is 160,000, testing data is 40,000, training batch size is 32, testing batch size is 64, epoch is 100, learning rate is 0.001, decay rate is 0.99 exponential decay, dropout rate is 0.8, convolution kernel is 3×3×3, maxpooling is 2×2×2 and using Adam optimization method. Due to that our positive data and negative data are not balanced, we use the loss function adjustment, sample adjustment and sample classification threshold shift.

Table 1. Comparison of fault identification performance between different methods.

| Method                          | Accuracy | Precision | Recall | F1-score |
|---------------------------------|----------|-----------|--------|----------|
| Planarity                       | 0.8760   | 0.7365    | 0.7652 | 0.7506   |
| Fault likelihood                | 0.9267   | 0.8679    | 0.8564 | 0.8621   |
| CNN + logistic regression       | 0.9656   | 0.6989    | 0.8021 | 0.7470   |
| CNN + support vector machine    | 0.9689   | 0.6754    | 0.8132 | 0.7380   |
| U-Net                           | 0.9641   | 0.9432    | 0.9518 | 0.9475   |
| Without Attention + VGG         | 0.9790   | 0.9673    | 0.9717 | 0.9695   |
| Multi-scale attention + VGG     | 0.9852   | 0.9735    | 0.9789 | 0.9762   |

As can be seen from the table 1, our proposed method is better than the previous method. The best indicators: Accuracy: 0.9852, Precision: 0.9735, Recall: 0.9789, F1-score: 0.9762. Speed: training time 4 hours, total test time 10 ~ 15 minutes. The total parameters are 120 million.

Data size: (X, Y, Z) = (69, 271, 451), using ((17-51), (68-202), (113-337)) range, ploting the data of size (35, 135, 225). Note: X: Crossline, Y: Inline, Z: Vertical.

Figure 3. The VGG convolutional neural network with multi-scale fusion attention mechanism.
4. Conclusion
We have proposed a framework of improved VGG convolutional neural network with multi-scale fusion attention mechanism for seismic fault recognition, which can achieve the state-of-the-art experimental results on real seismic data. By using the improved VGG network, the effective features from the seismic attributes are extracted. Moreover, by using the attention mechanism, the local features can be better learned, and the predicted fault curve becomes smoother and more continuous. As a result, our proposed approach is much better than the existing methods and quite effective and efficient for fault recognition.

Acknowledgments
This work was supported by the BGP Inc., China National Petroleum Corporation. We gratefully acknowledge the support of NVIDIA Corporation with the GeForce GTX 1080 GPU for this research.

References
[1] Randen T, Pedersen S I, Signer C, et al 1999 Image processing tools for geologic unconformity extraction In Proc. of IEEE Nordic Signal Processing Symp (Norway: Asker) pp 9-11
[2] Bahorich M S and Farmer S L 1995 3D seismic discontinuity for faults and stratigraphic features The Leading Edge 14 pp 1053-58
[3] Wu X and Hale D 2016 3D seismic image processing for faults Geophysics 81 pp IM1-11
[4] Wu X and Fomel S 2018 Automatic fault interpretation with optimal surface voting Geophysics 83 pp O67-82
[5] Wu X, Liang L, Shi Y et al 2019 FaultSeg3D: using synthetic datasets to train an end-to-end convolutional neural network for 3D seismic fault segmentation Geophysics pp 1-36
[6] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale Image Recognition Computer Science (Canada: Montreal)
[7] Vaswani A, Shazeer N, Parmar N, et al 2017 Attention Is All You Need Computer Science (USA:CA)
[8] Huang G, Liu Z, Van Der Maaten L, et al 2017 Densely connected convolutional networks Computer Vision and Pattern Recognition (USA: Hawaii) pp 4700-4708
[9] Matthew D Z, Graham W T and Rob F 2011 Adaptive Deconvolutional Networks for Mid and High Level Feature Learning IEEE (Spain: Barcelona)