Anisotropic parameter modeling based on deep residual network

Introduction

With the improvement of underground imaging precision, anisotropic media imaging has gradually become a conventional method. Anisotropic parameter modeling provides the final model for subsequent migration imaging and has an important influence on the quality of migration imaging. Therefore, it is necessary to study the anisotropic parameter modeling technology with high precision. As a method of constructing velocity model, tomography has become a common processing tool. However, there are also difficulties in conventional reflection tomography: (1) Travel time difference is difficult to obtain, especially for pre-stack data with low SNR; (2) During ray tracing, a large number of interpolation and test firing are required to fit the coordinates of the shot-receiver point, and there are complex reflection problems (Woodward 2008).

The development of deep neural network brings a series of breakthroughs for data processing. Among them, the most widely used is the deep residual neural network proposed by He et al. (2016), which adopts the residual connection method to effectively optimize the model parameters in back propagation. In recent years, with the steady development of artificial intelligence algorithms and the continuous improvement of computer capabilities, deep learning has been increasingly widely applied in the field of geophysical exploration, and has achieved good results in fracture prediction, first arrival automatic picking, reservoir parameter prediction and other aspects. This article apply the deep residual network to anisotropic parameter modeling work, based on the finite difference algorithm create double anisotropic parameters corresponding to the wave field snapshot data set as the benchmark data sets, and built a deep residual network, and improve the problem that BP neural network parameters is difficult to adjust, has obtained the ideal effect.

Method

In traditional convolutional networks, with the increase of network depth, the multi-layer back propagation of error signals will lead to gradient dispersion and explosion, which will reduce the recognition rate of the network. The residual network uses the residual learning module to improve the structure of the traditional deep convolutional network and avoid the performance degradation that is easy to occur during model training. The specific structure is shown in Figure 1. /2 represents the doubling of the number of channels. Figure 1 (a) is applicable to the case that the input channel and output channel of the residual block are consistent, and Figure 1 (b) is applicable to the case that the input channel and output channel of the residual block are inconsistent, so convolution of 1×1 is required for the input of the residual block to reduce the number of channels. Where, x is the input, y is the expected output of the network, Conv1 and Conv2 are the first and second convolutional layers respectively, BN stands for batch standardization, ReLU is the activation function, and \( F(x) \) is the residual function obtained after a series of processing.

While the Resnet model gradually deepens the CNN network structure, the weight to be determined iteratively in the model and the complexity to be calculated in the training also increase accordingly, which leads to excessive resource occupation during the training, high requirements on the training equipment and low training efficiency. In this paper, in order to improve the convergence rate and shorten the training time of the model, a standardized BN layer is added after the convolutional layer of the Resnet model, that is, the input data of each layer of the neural network is preprocessed into a distribution with a mean value of 0 and a standard deviation of 1. In addition, after the BN layer, the data falls into the region that is more sensitive to the input change after the nonlinear activation function, thus reducing the overfitting of the model and avoiding the problem of gradient explosion in the model.

In this paper, based on the ResNet18 structure of the residual network, a Resnet model suitable for anisotropic parameter modeling with wave field snapshot images is constructed, as shown in Figure 2.
The anisotropic parameter modeling based on residual network is mainly composed of three parts: sample set construction, model training and model application.

1) Sample set construction
The experiment involved building a data set of wave field images that allowed the model to learn about the potential relationship between the wave field images and the anisotropic parameters $\varepsilon$ and $\delta$. In this paper, anisotropic acoustic wave equation (Zhou 2006) is used to simulate. The equation is shown in formula (1).

$$
\frac{1}{\nu_p^2} \frac{\partial^2 p}{\partial t^2} - (1 + 2\delta) \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) p - \frac{\partial^2 p}{\partial z^2} = (1 + 2\delta) \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) q
$$

$$
\frac{1}{\nu_p^2} \frac{\partial^2 q}{\partial t^2} - 2(\varepsilon - \delta) \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) q = 2(\varepsilon - \delta) \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) p
$$

Where, $\nu_p$ is the acoustic velocity, $p$ is the wave field, $q$ is the auxiliary function to compensate the transverse upper wave front loss in VTI medium, $\varepsilon$ and $\delta$ is the anisotropic parameter. The simulation process is as follows: given two anisotropic parameters, a homogeneous medium is adopted, the longitudinal wave velocity is 2137m/s, the grid dimension is 201×201, and the grid spacing is 10m. The source with a 15 Hz dominant frequency Ricker wavelet is located in the center of the model, and the receivers are evenly distributed on the surface. The time step is 1ms, and the maximum recording time is 200ms. Set 0.001 for the interval of two anisotropic parameters, and 322,002 wave field images and anisotropic parameter pairs were generated for the measurements at $-0.4 \leq \varepsilon \leq 0.4$ and $-0.4 \leq \delta \leq \varepsilon$. Due to the limitations of equipment and the study of model performance by exploring a small number of samples, this paper uses the first 99221 data pairs generated to generate a small wave field image data set. In terms of data division, 1000 samples were selected as the verification set and 1000 samples as the test set by random sampling, and the number of training sets was taken as the independent variable to conduct experimental research. Among them, training set, verification set and test set are mutually exclusive.

2) Model training and model application
For the Resnet model, the fixed-structure Resnet18 network is used, and the last layer of 1000 nodes is changed into the output of 2 nodes, as shown in Figure 1 (b). The residual block represented by the dotted line box is the structure (a) in Figure 1, and the solid line box represents the structure (b) in Figure 1. The initial learning rate was set to 0.001, the exponential decay learning rate adjustment...
strategy with a base of 0.5 was adopted, the Adam optimization method was adopted, and the batch size was 4. This model contains approximately $1.0 \times 10^6$ parameters.

**Figure 2 Deep residual network model**

In order to obtain a parameter model that can match the snapshot image of the wave field with its corresponding anisotropic parameters, the snapshot image data of the wave field is taken as input, and the anisotropic parameters are taken as the learning target, namely "label". The BP neural network constructed is trained as follows: From a set of training samples randomly divided into four samples, and put these samples in deep neural network model to forward propagate, the obtained predicted anisotropic parameters and the true anisotropic parameters of the four samples were calculated, and the losses were propagated backward and the network parameters were adjusted layer by layer, and do this until all the training samples are taken. One such training was for an epoch, and it lasted for 80 generations. At the end of each generation, the obtained parameter model was used to predict and verify the anisotropic parameters of the data set, and the variance between the predicted value and the true value of all samples was calculated. In this paper, the generation 1 with the smallest variance of generation 80 is selected as the final model, and the effect of the selected model is tested on the test set.

**Examples**

Change the number of training samples, take training sets with less than 5000 samples as a small number of samples, and training sets with more than 10000 samples as a large number of samples. In the case of a small number of samples, 500 samples were added at a time, and in the case of a large number of samples, 20,000 samples were added at a time. Several models were trained respectively. During the training, the loss convergence of L1 on the training set and the verification set of the model obtained in each generation is recorded. In order to better observe the situation of the model on the small training sample data set and the large training sample data set, the overall loss curve was divided and plotted. Figure 3 shows the L1 loss of residual network. The loss function in the training process converges well due to the use of Adam optimization algorithm. In the verification set, the loss fluctuates sharply, because the deep network model learns knowledge on the training set. The verification set sample is different from the training set sample, and the overfitting on the training set causes some errors on the verification set. Even so, the overall loss on the validation set continues to decline rapidly.

Using models obtained from training sets with different sample numbers, the test set samples were predicted, and the difference between the predicted value and the true value of 1000 pairs of anisotropic parameters was recorded. In order to better understand the confidence of the prediction of the overall deep network model, the error bar diagram as shown in Figure 4 was drawn with 1000 pairs of data, and the two predicted parameters were plotted separately. The overall trend of mean absolute error decreases, but the standard deviation fluctuates to a certain extent, which is related to the different distribution of training set and test set samples, in which there is some overfitting.

**Conclusions**
The anisotropic parameter modeling based on BP neural network can accurately calculate the anisotropic parameters, but to obtain an appropriate network requires complex structural adjustment, which requires a great deal of time and energy. However, the anisotropic parameter modeling based on the deep residual network does not require complex parameter adjustment process, so it has strong universality and the calculation results are more accurate. For practical geological tasks, it is more recommended to use a Resnet model, which is easier to set up, with far fewer network parameters than BP neural network, and faster training speed than BP neural network, which can converge faster in the case of fewer training samples.

**Figure 3** Convergence curve of L1 loss of Resnet model on training set and verification set. Sample number of training set: (a) 500-90000; (b) 10000-90000.

**Figure 4** Mean absolute error and standard deviation of Resnet model on test set. Sample number of training set: (a) 500-90000; (b) 10000-90000.

**References**

