The Impact of Network Depth for the 2D Magnetotelluric Inversion Based on Deep Learning Inversion Models

Xingong TANG[1]，Junhu YU[2] and Zhitao XIONG[3]
1Key Laboratory of Exploration Technologies for Oil and Gas Resources of MOE, Yangtze University, Wuhan, Hubei, China, 430100
[1]Xingong TANG, Key Laboratory of Exploration Technologies for Oil and Gas Resources of MOE, Yangtze University, Wuhan, Hubei,China, tangxg@yangtzeu.edu.cn
[2]Junhu YU, Key Laboratory of Exploration Technologies for Oil and Gas Resources of MOE, Yangtze University, Wuhan, Hubei,China, 1121235150@qq.com
[3]Zhitao XIONG, Key Laboratory of Exploration Technologies for Oil and Gas Resources of MOE, Yangtze University, Wuhan, Hubei,China, 1228409354@qq.com

SUMMARY

This paper investigates the impact of different network depth on the performance of 2D magnetotelluric deep learning inversion models based on residual neural networks. 100,000 electric anomaly body models with different shapes and different scales generated by Gaussian random fields were design and computed. Through a self-developed batch parallel finite-difference forward modeling program based on 2D staggered grid, we conduct forward calculations and use the TE response of the models as training samples for the deep learning inversion model. Three different depths network structures, ResNet-18, ResNet-50, and ResNet-152, are trained on the samples to observe the influence of different network depths on the model performance, respectively. The results of the model inversion show that increasing the network depth within a certain range can improve the inversion performance of the deep learning model. However, beyond a certain threshold, by increasing the depth of the network, it will not lead to improvement of model performance, but increase the computational time.

Keywords: magnetotelluric, two dimensional inversion, neural network inversion, Residual Neural Networks

INTRODUCTION

The principle of magnetotelluric depth sounding inversion is to utilize the surface magnetotelluric field signals obtained from the Magnetotelluric (MT) method to infer the electrical models that conform to the actual underground structures. However, traditional magnetotelluric inversion methods based on iterative algorithms relying on objective function optimization theory always depend on the selection of initial inversion models. The inversion results are greatly influenced by human factors, prone to local minima, and time-consuming, making it difficult to meet the demand for efficient processing and interpretation of measured data. In recent years, the widespread application of deep learning technology has provided new approaches for solving inverse problems.

Puzyrev (2019) proposed a deep fully convolutional inversion network and applied it to the inversion of measured data, achieving good inversion results. Liu (2022) proposed a one-dimensional inversion algorithm PhyDNN based on physics-driven neural networks, and its reliability in processing measured data was verified through comparison with OCCAM inversion. Ling (2023) proposed an 8-layer residual neural network (ResNet1D-8) for AMT data inversion, and its good convergence performance was verified through comparison with results from traditional inversion methods. The MT inversion methods based on deep learning mentioned above can output inversion results in real-time and are more efficient in computation compared to traditional regularization inversion methods, indicating broad application prospects of deep learning technology in the field of magnetotelluric inversion.

Some studies have shown that appropriately increasing network depth can improve the accuracy and generalization ability of the model. However, increasing network depth is not always beneficial. When the network depth exceeds a certain threshold, further increasing depth can lead to a decrease in model performance, but decrease the computational efficiency. This is because excessively deep networks are prone to gradient vanishing or exploding problems, making it difficult for the model to train and converge. Therefore, this paper designs three residual neural networks of different depths to train samples and study the influence of different network depths on model performance.

Methodology

Similar to the objective function in traditional regularization inversion, deep learning-based magnetotelluric inversion also requires the design of a suitable loss function and the minimization of this loss function through model training to obtain the optimal inversion model. However, it has the advantages of not relying on an initial model and not requiring the calculation of the gradient of the objective function. The training of deep learning inversion models mainly involves the following steps:
(1) Construction of the model training dataset. We generate a large number of geological models with varying degrees of smoothness using Gaussian random fields to establish a standard batch data model library for deep learning inversion training. Corresponding response values are obtained through batch forward calculations, and the resistivity models are matched with the response data one by one to construct the sample dataset for training the network model.
(2) Construction and training of the network model. We design three residual neural networks, ResNet-18, ResNet-50, and ResNet-152, to train magnetotelluric inversion models. The three different depth network structures are shown in Table 1. Before training, the data is standardized to improve the network training process and obtain an ideal network model.
(3) Inversion prediction of resistivity. The resistivity data corresponding to the TE polarization mode response are standardized, and the standardized data are input into the trained network model for real-time inversion prediction to obtain the predicted distribution of underground resistivity structure.

The rationality of deep learning training samples is crucial for improving the generalization ability of the inversion model. Training samples consisting solely of models with regular boundaries have relatively simple features. Therefore, in the design of forward initial models, we introduce Gaussian random fields to construct complex and smooth boundary models. The generation method of complex smooth models using Gaussian random fields is illustrated in Figure 1. In the model depicted in Figure 1, the central area contains three blue circles representing low-resistivity points and two red circles representing high-resistivity points. These circles represent the coordinates of the centers of five random anomaly bodies in the central area of the model, each with its own resistivity value range. The smoothness of the model is controlled by setting the range of anomaly bodies and the scale parameter of the Gaussian random field.

The rationality of the deep learning sample set is crucial for enhancing the model's inversion generalization capability. Training models solely composed of samples from rule-based boundaries exhibit slightly insufficient generalization performance. In designing the initial model for forward modeling, Gaussian random fields are introduced to construct complex smooth boundary models. After determining the model boundaries and dividing the model grid, establish a Gaussian random field within the central region of the model. The method of generating complex smooth models using Gaussian random fields is illustrated in Figure 1. In the model shown in Figure 1, the central region comprises three blue-circled low-resistivity points and two red-circled high-resistivity points, representing the coordinates of the five random anomaly body center points within the central region of the model. Each point has its own assigned resistivity range. The smoothness of the model is controlled by setting the anomaly body range and the scale parameter of the Gaussian random field.

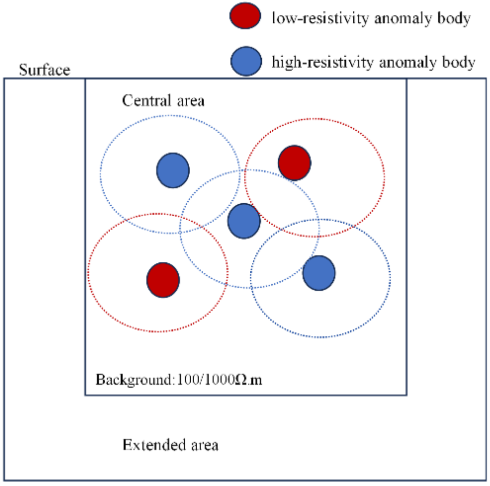

Figure 1. Schematic of smooth model generation by Gaussian random field.

The smaller the scale parameter value, the clearer the boundaries between the model anomaly bodies in the output. Conversely, larger scale parameter values result in more thorough merging of boundaries between the model anomaly bodies, creating smoother boundaries. The generated model schematic is illustrated in Figure 2.

For the sake of simplifying the research, here, only the apparent resistivity values under the TE polarization mode are selected as training sample data, with the original model corresponding to each sample being used as the training label. Effective data preprocessing methods can reduce the difficulty of network training, assisting the network in more accurately and efficiently uncovering and learning the nonlinear mapping relationship between input and output data. Since the background resistivity of the model and the resistivity of the anomaly bodies span multiple orders of magnitude, resulting in significant differences in resistivity value ranges, this study takes the logarithm base 10 of both the sample data and the sample labels. This compression brings all data onto the same scale for network input, effectively reducing differences between strong and weak signals, while also minimizing the impact of noise on model training.

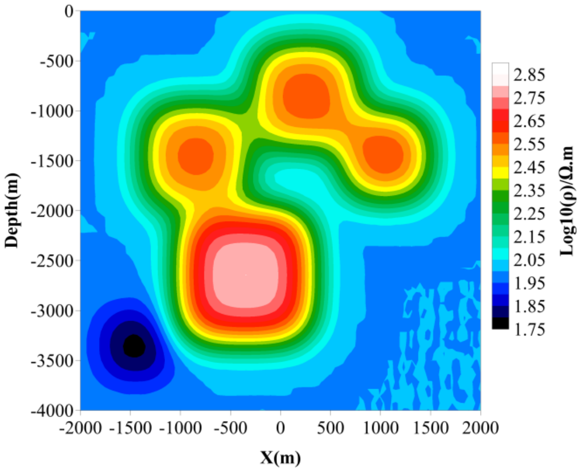

Figure 2. Schematic of smooth boundary model generation.

The complete hyperparameter settings for the neural network model training in this paper are shown in Table 2. The activation function used is ReLU, which is the most commonly used in Convolutional Neural Networks (CNNs). The training optimizer employed is Adam, known for its robustness, fast convergence speed, and suitability for large-scale hyperparameters. The initial learning rate for training is set to 0.01, and as training progresses, the learning rate decays by a factor of 0.8 every 10 iterations. This approach allows the model to gradually reduce the learning rate during training, bringing it closer to the optimal solution, thus improving model performance and convergence speed. Additionally, the "early stopping" training strategy is adopted in this paper, where training stops if the validation set error does not decrease continuously for 50 consecutive times.

This paper constructs a two-dimensional magnetotelluric inversion neural network model based on the TensorFlow 2.4 framework and performs network training and prediction in CPU mode. The hardware configuration used is an Intel(R) Core(TM) i5-12400 CPU @2.50GHz with 16GB RAM. In the network training process, the apparent resistivity under TE polarization from magnetotelluric forward modeling is used as the input data for the network channels, while the model resistivity serves as the expected output data. The sample dataset is divided into training and validation sets in an 8:2 ratio. The three network training processes all converge rapidly, with the error stabilizing after approximately 150 iterations, and no overfitting phenomena are observed. This indicates that the proposed inversion model and network training strategy are reasonable. After the model training is completed, the trained network model is saved for use in the subsequent inversion steps.

Table 1 Different deep residual neural network architectures

Layer name	Output size	18-layer	50-layer	152-layer
Conv1	112×112	7×7，64，stride 2		
Conv2_x	56×56	3×3，max pool，stride 2		
		$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times2$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$
Conv3_x	28×28	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times8$
Conv4_x	14×14	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times2$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times36$
Conv5_x	7×7	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times2$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$

Table 2 Hyper parameter description

Hyper parameter	description
Epoch	200
batchsize	16
activation function	ReLU
learning rate	0.001
optimizer	Adam
gradient clipping	1

Inversion test

Here, we constructed 2000 test sample data sets in the same manner as before to evaluate the performance of different depth deep learning inversion models. To assess the impact of network depth on the inversion training, we randomly selected one geoelectric model for testing purposes.

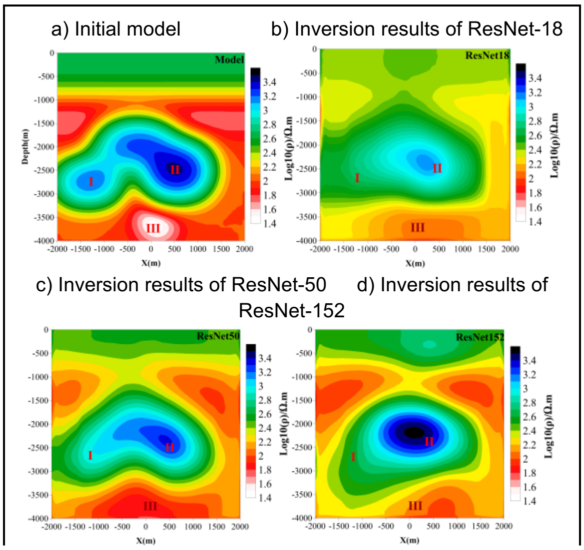
a) Initial model
b) Inversion results of ResNet-18
c) Inversion results of ResNet-50
d) Inversion results of ResNet-152

Figure 3. a) Test initial model for inversion, b) Test results of the ResNet-18 inversion model, c) Test results of the ResNet-50 inversion model, and d) Test results of the ResNet-152 inversion model.

Figure 3 shows the smooth anomaly body model and the inversion results. Figure 3(a) displays the forward model, while Figures 3(b), 3(c), and 3(d) present the inversion results for ResNet-18, ResNet-50, and ResNet-152, respectively. The background resistivity of the model is 100 Ω•m, and the model contains three randomly generated smooth boundary anomaly bodies of different sizes (the blue parts represent high-resistivity anomaly bodies, while the red parts represent low-resistivity). As seen from the inversion results, all three models can effectively recover the positions and sizes of different anomaly bodies for the synthetic data model, demonstrating good inversion performance. Additionally, it can be observed that the inversion results of ResNet-50 and ResNet-152 show better boundary constraints compared with the results of ResNet-18. However, there is no significant improvement in the inversion results for ResNet-152 compared to that of ResNet-50.

CONCLUSIONS

This paper presents 2D MT inversion results with three residual neural network, ResNet-18, ResNet-50, and ResNet-152 to study the influence of network depth on deep learning inversion training. A novel training sample generation method is devised to make the model more suitable for representing complex geoelectric structures. After logarithmic standardization processing of the data, differences between strong and weak signals are effectively reduced while retaining the characteristics of high and low resistivity anomalies in the samples. Finally, the network is tested for inversion using unfamiliar data. The results indicate that shallow networks struggle with deep learning inversion training on large datasets. Appropriately increasing network depth can effectively enhance the training performance of the inversion network. However, excessively deep networks offer limited improvement in the performance of the inversion model and significantly increase inversion training time. This study holds important guiding significance for sample preparation for magnetotelluric data inversion, network parameter design, and network training.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (42274087, 42174083) projects.

REFERENCES

Puzyrev V. 2019. Deep learning electromagnetic inversion with convolutional neural networks. Geophysical Journal International, 218(2): 817-832.

Liu, W., et al. 2022. Physics-driven deep learning inversion with application to magnetotelluric. Remote Sensing, 14(13): 3218.

Ling W, Pan K, Ren Z, et al. 2023. One-dimensional magnetotelluric parallel inversion using a ResNet1D-8 residual neural network. Computers & Geosciences, 180: 105454.