

## 2D magnetotelluric inversion based on ResNet

LiAn Xie<sup>a</sup>, Bo Han<sup>a</sup>, Xiangyun Hu<sup>b,\*</sup>, Ningbo Bai<sup>c</sup>

<sup>a</sup> Institute of Geological Survey, China University of Geosciences, Wuhan, 430074, China

<sup>b</sup> Hubei Subsurface Multi-Scale Imaging Key Laboratory, School of Geophysics and Geomatics, China University of Geosciences, Wuhan, 430074, China

<sup>c</sup> Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan, 430074, China



### ARTICLE INFO

**Keywords:**  
Magnetotellurics  
2D inversion  
Residual network  
Deep learning

### ABSTRACT

In this study, a deep learning algorithm was applied to two-dimensional magnetotelluric (MT) data inversion. Compared with the traditional linear iterative inversion methods, the MT inversion method based on convolutional neural networks (CNN) does not rely on the selection of the initial model parameters and does not fall into the local optima. Although the CNN inversion models can provide a clear electrical interface division, their inversion results may remain prone to abrupt electrical interfaces as opposed to the actual underground electrical situation. To solve this issue, a neural network with a residual network architecture (ResNet-50) was constructed in this study. With the apparent resistivity and phase pseudo-section data as the inputs and with the resistivity parameters of the geoelectric model as the training labels, the modified ResNet-50 model was trained end-to-end for producing samples according to the corresponding production strategy of the study area. Through experiments, the training of the ResNet-50 with the dice loss function effectively solved the issue of over-segmentation of the electrical interface by the cross-entropy function, avoided its abrupt inversion, and overcame the computational inefficiency of the traditional iterative methods. The proposed algorithm was validated against MT data measured from a geothermal field prospect in Huanggang, Hubei Province, which showed that the deep learning method has opened up broad prospects in the field of MT data inversion.

### 1. Introduction

Magnetotellurics (MTs) is a geophysical exploration technique that infers underground variations in electrical conductivity and resistivity from the Earth's natural variations in electric and magnetic fields. Widely used in resource exploration and geological research, MTs have the advantages of being robust to high and low resistivity and reaching greater resource exploration depths. Due to the strong nonlinearity and uncertainty of MT data, its accurate and precise inversion remains a difficult problem. Currently, the common methods of MT data inversion are based on linearized iterative algorithms using the gradient information of the objective function, such as Newton, quasi-Newton (QN), Gauss Newton (GN), Occam inversion (Siripunvaraporn et al., 2002), and nonlinear conjugate gradient (NLCG) methods (Newman et al., 2002). Compared to gradient based first-order methods, Occam method is a second-order method. The Occam method proposed by Constable is a classical Gauss Newton method. Each iteration of this method will update the regularization factor, which is famous for its inversion stability and is widely used in magnetotelluric one-dimensional, two-dimensional and three-dimensional data inversion (Constable et al. (1987);

Groot-Hedlin and Constable, 1990; Siripunvaraporn and Sarakorn, 2000). Smith (1991) proposed a fast relaxation inversion algorithm for magnetotelluric data, greatly improving the inversion speed. These deterministic methods have achieved satisfactory results in practice (Liu et al., 2021) but are sensitive to the initial model parameters and easily fall into the local optimal solution.

With the progress of computer performance and forward algorithms, more nonlinear inversion methods have been proposed. simulated annealing inversion (Shi and Wang, 1998; Sharma, 2012), ant colony algorithm inversion (Liu et al., 2015), genetic algorithm inversion (Schwarzbach et al., 2005), particle swarm optimisation algorithm inversion (Shaw and Srivastava 2007) and Bayesian inversion (Guo et al., 2011; Di et al., 2020), these nonlinear methods have been implemented in electromagnetic fields and have achieved good results. Furthermore, There are some previous applications of ANN to 2D MT inversion (Montahaei and Oskooi, 2014; El-Qady and Ushijima 2001), which demonstrates the feasibility of artificial intelligence in the field of MT inversion.

Recently, the deep learning (DL) algorithm has provided new ideas for processing MT data. As a branch of machine learning, the DL

\* Corresponding author. Institute of Geological Survey, China University of Geosciences, Wuhan, 430074, China.

E-mail address: [xyhu@cug.edu.cn](mailto:xyhu@cug.edu.cn) (X. Hu).

algorithm is based on the development of artificial neural networks (ANN), which is a data-driven, statistical, and approximate method. Following training, generalized inductive learning, and continuous updating of network node weights, the DL algorithm can make the network response approximate to any nonlinear complex behaviors. Based on DL, mapping the relationships between subsurface physical properties and the generated physical field data can help in solving the geophysical inversion issue without requiring cyclic iterative computation of the model responses, thus rendering its computational efficiency favorable. Several studies have explored various DL-based methods of geophysical prospecting. Liu et al. (2021) proposed a DL model for resistivity data inversion by incorporating a smoothing constraint and depth weighting into loss function to reduce false anomalies and improve the inversion accuracy. Zhang et al. (2021) used a convolutional neural network (CNN) architecture called U-Net for gravity data inversion and reconstructed a residual density model. Moghadas (2020) implemented CNN-based one-dimensional (1D) inversion for electromagnetic induction (EMI) data and instantaneously estimated the number of subsurface conductivity layers. Li et al. (2020) used the DL technique for the rapid imaging of time-domain airborne electromagnetic data. Many studies have also applied the DL algorithm to seismic full-waveform inversion (Zhang and Alkhalifah, 2019) and other geophysical fields, such as gravity (He et al., 2021; Yang et al., 2022) and magnetism. Previous studies have showed that DL-based methods offer high computational efficiency, accurate and precise predictions, and broad prospects for a wide range of applications.

At present, the application of the DL algorithm to geophysical data inversion still remains in its infancy. To the best of our knowledge, the application of DL to MT data inversion still possesses shortcomings in terms of validation and an abrupt electrical interface. By adjusting the ResNet-50 structure and its loss function, this study transformed 2D MT data inversion from being a prediction issue via a continuous regression model to that via a logistic classification model for improving computational cost and efficiency. After MT forward and inverse problems were introduced, the DL algorithm and its network structure, numerical experiments, and validation were presented.

## 2. MT theory

### 2.1. MT forward problem

The production of dataset was based on the finite element method for 2D MT forward modeling. Maxwell's equations in 2D MT can be divided into the polarization modes of transverse electric (TE) and transverse magnetic (TM). From Maxwell's equations (Siripunvaraporn et al., 2002), it can be deduced that, for the TE mode,  $E_x$  should satisfy the following partial differential equation:

$$\frac{\partial}{\partial y} \left( \frac{1}{i\omega\mu} \frac{\partial E_x}{\partial y} \right) + \frac{\partial}{\partial z} \left( \frac{1}{i\omega\mu} \frac{\partial E_x}{\partial z} \right) + \sigma E_x = 0, \quad (1)$$

For the TM mode, the equation satisfied by the x-component of the magnetic field ( $H_x$ ) is as follows:

$$\frac{\partial}{\partial y} \left( \frac{1}{\sigma} \frac{\partial H_x}{\partial y} \right) + \frac{\partial}{\partial z} \left( \frac{1}{\sigma} \frac{\partial H_x}{\partial z} \right) + i\omega\mu H_x = 0, \quad (2)$$

where  $E_x$  is the x-component of the electric field;  $\sigma$  is the electrical conductivity;  $\omega$  is the angular frequency; and  $\mu$  is the permeability.

To solve Eqs. (1) and (2), the boundary conditions in the TE mode must be set as follows:

$$\begin{aligned} E_x|_{AD} &= 1, \frac{\partial E_x}{\partial y} \Big|_{AB,CD} = 0 \\ \frac{\partial E_x}{\partial y} \Big|_{BC} &+ \sqrt{-i\omega\mu\sigma} E_x \Big|_{BC} = 0 \end{aligned} \quad (3)$$

Moreover, the boundary conditions in the TM mode should be as follows:

$$\begin{aligned} H_x|_{AD} &= 1, \frac{\partial H_x}{\partial y} \Big|_{AB,CD} = 0 \\ \frac{\partial E_x}{\partial y} \Big|_{BC} &+ \sqrt{-i\omega\mu\sigma} E_x \Big|_{BC} = 0 \end{aligned} \quad (4)$$

AB and CD are the left and right boundaries, respectively, at infinity. AD and BC are the upper boundary on the ground and the lower boundary at infinity, respectively. The left and right lower boundaries are approximately infinite, and the upper boundary is z.

For the TE mode, Assuming the upper boundary  $z = -h$  is infinite, and  $E_x$  is a constant value:

$$\frac{\partial E_x}{\partial z} \Big|_{z=-h} = i\omega\mu \quad (5)$$

For the TM mode,  $H_y$  at the upper boundary z is a constant value:

$$H_y \Big|_{z=0} = 1 \quad (6)$$

Therefore, the MT boundary problem can be summarized as follows:

$$\begin{cases} \nabla \cdot (\tau \nabla u) + \lambda u = 0 & \in \Omega \\ u = 1 & \in AB \\ \frac{\partial u}{\partial n} = 0 & \in AD, BC \\ \frac{\partial u}{\partial n} + ku = 0 & \in CD \end{cases} \quad (7)$$

where u refers to the electric field component  $E_x$  in Eq. (1) or the magnetic field component  $H_x$  in Eq. (2).

Further, the boundary is substituted by a functional construction, and the problem of transformation becomes equivalent to the following variational problem:

$$\begin{cases} F(u) = \int_{\Omega} \left[ \frac{1}{2} \tau (\nabla u)^2 - \frac{1}{2} \lambda u^2 \right] d\Omega + \int_{CD} \frac{1}{2} \tau k u^2 d\Gamma \\ u|_{AB} = 1 \\ \delta F(u) = 0 \end{cases} \quad (8)$$

The apparent resistivity and phase expressions of the surface observation points can be obtained through substitution and derivation under the boundary conditions. In the TE mode, the following are involved:

$$Z_{TE} = \frac{E_x}{\left( \frac{1}{i\omega\mu} \frac{\partial E_x}{\partial z} \right)} \quad (9)$$

$$\begin{aligned} \rho_a^{TE} &= \frac{1}{\omega\mu} |Z_{TE}|^2 \\ \varphi^{TE} &= \arctan \frac{\text{Im}[Z_{TE}]}{\text{Re}[Z_{TE}]} \end{aligned} \quad (10)$$

The TM mode involved the following:

$$Z_{TM} = \frac{\frac{1}{\sigma} \frac{\partial H_x}{\partial z}}{H_x} \quad (11)$$

$$\begin{aligned} \rho_a^{TE} &= \frac{1}{\omega\mu} |Z_{TM}|^2 \\ \varphi^{TE} &= \arctan \frac{\text{Im}[Z_{TM}]}{\text{Re}[Z_{TM}]} \end{aligned} \quad (12)$$

where  $Z_{TE}$  and  $Z_{TM}$  are the wave impedances of the TE and TM modes,

respectively;  $\rho_a^{\text{TE}}$  is the apparent resistivity of the TE mode;  $\rho_a^{\text{TM}}$  is the apparent resistivity of the TM mode; and  $\varphi^{\text{TE}}$  and  $\varphi^{\text{TM}}$  are the phases of the TE and TM modes, respectively.

Given the above formulas, the apparent resistivity and phase data under the corresponding modes were estimated from the calculations of the electric or magnetic field components. In this study, the triangular finite element method was used for 2D MT forward calculations. Based on the rectangular element grid, the entire grid was divided into the target and extended regions. The rectangular grid was divided into triangles, and the triangular biquadratic interpolation was performed to obtain the shape function. The stiffness matrix was formed by substituting the differential equation of the 2D finite element variational problem of MT, and the electric or magnetic field values of each grid node were obtained from the boundary conditions. The component values were obtained by the difference, whereas the wave impedance, apparent resistivity, and phase values were obtained by substituting Eqs. (7)–(10). The target areas were the storage areas of geological body and data acquisition, divided into the uniform grids. According to certain strategies, resistivity anomaly was filled in the storage area of geological body to generate multiple models. Apparent resistivity and phase pseudo-section data obtained via forward modeling were used as double-channel features, and the corresponding geoelectric model outputs were used as labels to obtain the training data.

## 2.2. MT inversion problem

The existence of resistivity anomalies in the underground space is the basic condition of the MT detection method. The observed responses caused by different resistivity anomalies have different localization characteristics, and there also exists a spatial correspondence between the observed responses and the underground resistivity anomalies. The core idea of applying the DL algorithm to 2D MT data inversion was to build training datasets through a large amount of forward data of different resistivity models, learn and conclude from experience about these feature relationships, constantly update the weights of the network during the training, and map the nonlinear relationships between the inputs of apparent resistivity and phase data and the output of resistivity anomalous bodies. This mapping can be expressed as follows:

$$m = F(\rho_a, \varphi) \quad (13)$$

where  $m$  is the resistivity model;  $F$  is the mapping function;  $\rho_a$  is the apparent resistivity data; and  $\varphi$  is the phase data.

## 3. DL methodology

### 3.1. Convolutional layer

A deep CNN is composed of convolution layers, pooling layers, and fully connected layers of different scales and involves operations and processing, such as nonlinear activation functions and batch regularization (BatchNorm, BN), among which the convolutional layer plays a

role in feature extraction.

**Fig. 1** is a schematic diagram of the convolution layer in CNN, which can be expressed as follows:

$$y^l = f(w^l * y^{l-1} + b^l) \quad (14)$$

where  $w^l$  and  $b^l$  represent the convolution kernel and the bias value in the layer  $l$ ;  $*$  represents the convolution operation;  $y^{l-1}$  and  $y^l$  refer to the input and output of the layer  $l$ ; and the output of each convolution layer is the nonlinear weight of the output of its previous layer. Each convolution layer usually contains more than one convolution kernel, through which the characteristic information of multiple channels is outputted.  $w^{l,k}(k=1, \dots, K)$  is used to represent the  $k$ 'th convolution kernel in the  $l$ -th convolution layer. Then, in the  $l$ -th layer, for the input value  $y^{l-1}$  with the number of channels  $C$ , convolution is carried out through the  $k$ 'th convolution kernel, and the output value at the position  $(i, j)$  can be expressed as follows:

$$y_{k,i,j}^l = f \left( \sum_{c=1}^C \sum_{m=1}^M \sum_{n=1}^N w_{c,m,n}^{l,k} y_{c,i-\left\lceil \frac{M}{2} \right\rceil + m, j - \left\lceil \frac{N}{2} \right\rceil + n}^{l-1} + b^{l,k} \right) \quad (15)$$

where  $C$ ,  $M$ , and  $N$  are the number of channels, height, and width of the convolution kernel  $w^{l,k}$ , respectively. The number of channels in the convolution kernel should be the same as the input value  $y^{l-1}$ .  $f$  is the nonlinear activation function used to introduce nonlinear characteristics into the CNN.

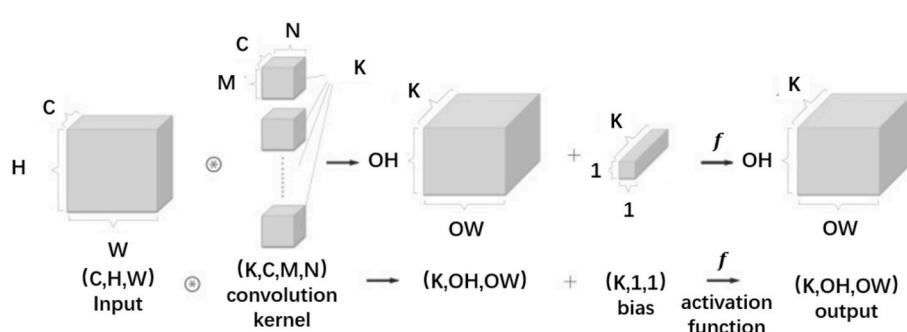
The convolution and combination operations only have linear computational capabilities. When the training data are complex, and the required mapping is nonlinear, the participation of the activation function is required to increase the nonlinear expression ability of the ANN to learn and map any complex nonlinear functions. The common activation functions, such as  $\text{sigmod}(x)$ ,  $\tan h(x)$ , and  $\text{ReLU}(x)$ , are defined in Eqs. (14)–(16) below:

$$\text{sigmod}(x) = 1 - \frac{1}{1 + e^{-x}} \quad (16)$$

$$\tan h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (17)$$

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (18)$$

Different activation functions exhibit different advantages and disadvantages. For example,  $\text{sigmod}(x)$  is smooth and easy to differentiate but causes gradient disappearance.  $\tan h(x)$  can slow down the gradient disappearance but leads to high computational cost.  $\text{ReLU}(x)$  shows low computational cost and can quickly converge during training but may cause neuronal inactivation. Overall, the more the number of hidden layers of CNN are, the more the network parameters and nonlinear activation functions are, and the stronger the nonlinear expression



**Fig. 1.** Schematic diagram of the convolution layer.

ability of CNN is.

### 3.2. Pooling layer

A CNN typically arranges the pooling of connection layers among a series of convolution layers, the essence of which is down-sampling. As the dimension of data becomes increasingly high after multiple convolutions, and the graph of features does not change significantly, a large number of parameters are generated, which results in not only the difficulty of training the network, but also the issue of overfitting. Therefore, data are typically compressed by pooling connections among the convolution layers in order to reduce the data dimensionality and the number of parameters. In particular, the pooling operation serves to aggregate statistics between a pixel and its surrounding data, reduce the size of the feature map, and take the average or maximum value of its adjacent areas so as to further reduce the number of parameters. This can be used to fuse the feature maps of the previous layer and prevent overfitting as the parameters of adjacent regions have a strong correlation. There are two common pooling operations: maximum pooling and average pooling.

**Fig. 2** shows the maximum-pooling operation. Average pooling works in a similar manner, except that the maximum value is replaced by the average value.

The pooling operation accomplishes the following three main functions: (1) reduction in the sensitivity of the convolution layer to the spatial position; (2) reduction in the sensitivity of the spatial subsampling process; and (3) reductions in the size of the input array and the computing resources for the subsequent neural network layer.

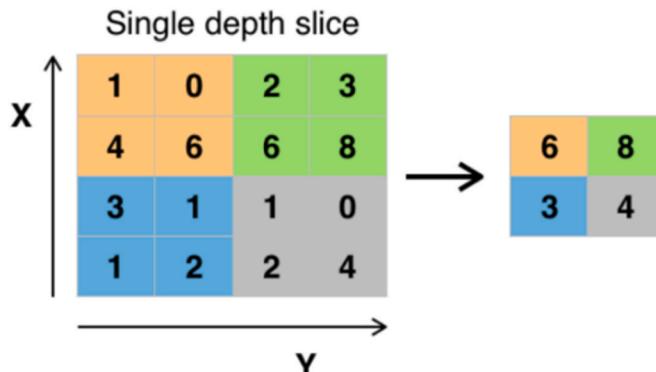
### 3.3. Loss function

The loss function is an operation used to measure the difference between the predicted and observed (label) values. In the training stage of the model, once the training data of each batch were sent to the model, the predicted value was outputted through forward propagation, and the difference between the predicted and real values was calculated through the loss function to obtain the loss value. Then, the model updated each parameter in the network through backpropagation in order to reduce the loss between the observed and predicted values. Thus, the model predictions were close to the observed values, while the best-fit network mapped the relationship based on the training data.

This study used the dice loss function (Dice\_Loss) and the cross-entropy loss function (CE\_Loss), which could not cope with a strong imbalance of positive and negative samples.

$$L = - \sum_{i=1}^N y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}) \quad (19)$$

where  $L$  is the cross-entropy loss function for inversion;  $N$  is the total



**Fig. 2.** Schematic diagram of maximum pooling operation.

number of grids;  $i$  is the grid number;  $y^{(i)}$  for the  $i$ 'th grid label value of resistivity; and  $\hat{y}^{(i)}$  is the  $i$  th grid resistivity classification prediction.

$$L_{dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (20)$$

where  $L_{dice}$  is the dice loss function; and  $X$  and  $Y$  are the label and prediction samples, respectively. The cross-entropy loss function equitably treats each point in the grid for error calculation when the positive and negative samples are strongly imbalanced. Thus, the positive samples with a small proportion are submerged by more negative samples, resulting in an abrupt electrical interface for the inversion results. The calculation error of the Dice\_Loss function is related to not only the predicted value of the current grid points, but also the value of other neighboring points in the relevant area based on the intersection operation of the predicted samples and label samples. This in turn increases the influence of positive samples with a small proportion on the updating of network weights, thus making the network fit in the right direction.

## 4. ResNet framework

### 4.1. Residual module

ResNet was proposed by He et al. (2015), aiming to solve the “degradation phenomenon” of deep neural networks; that is, a significant decrease in the model accuracy without warning after the accuracy saturation with the deepening of the network layer. The increasing depth and number of network layers increase the number of activation functions introduced, conducive to the improvement of the nonlinear transformation ability of the network to better perform the extraction and transformation of data features. However, when it reaches a certain level, the network almost completely loses its ability to convert linearly. This was where He et al. (2015) proposed the ResNet structure with the residual module of “skip-connection”.

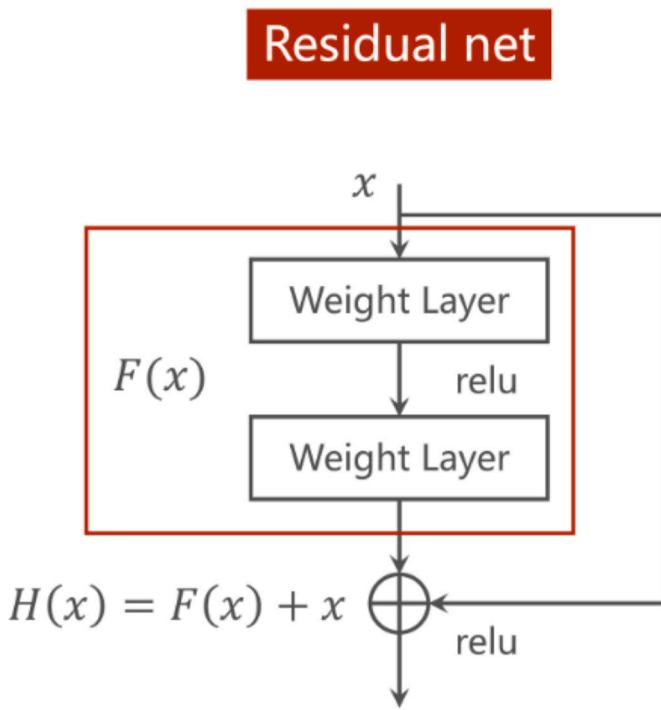
### 4.2. ResNet-50 structure

The core concept of the ResNet-50 lies in the skip-connection module. As shown in **Fig. 3**, on the basis of retaining the output  $F(x)$  of the nonlinear layer, a short connection is directly introduced from the input to the output  $H(x)$ , thus making the overall mapping of this module appear as follows:

$$H(x) = F(x) + x \quad (21)$$

**Fig. 4** shows the structure of the ResNet-50, which consists of five stage layers and one fully connected layer to form a deep CNN.

The first-stage layer, conv1, is the convolution layer of the interconnection network input, composed of 64 convolution kernels with a stride of two. The size of each convolution kernel is  $7 \times 7$  pixels. The remaining 48 layers form 16 residual module structures, referred to as Bottleneck modules in the ResNet-50. The four stages, conv2\_x to conv5\_x, are stacked by different bottleneck layers whose structure in each stage is different. The network structure of conv2\_x in the second stage is shown as an example on the right side of **Fig. 4**. The input in this stage first passes through a maximized pooling layer with a stride length of two and a size of  $3 \times 3$  and then passes through three stacked bottlenecks, as shown in **Fig. 4**. Each Bottleneck structure, shown in the red box in **Fig. 4**, is composed of the three convolution layers: conv1, conv2, and conv3. In conv2, the convolution kernel size was  $3 \times 3$ , the number of channels was 64, the step size was one, and the padding value was one. The raw input data of each bottleneck module are added to the output values of its last convolution layer, conv3, through a bottleneck module. The skip-connection channel passes through a convolution layer whose function is to ensure that the number of channels is consistent when the input and output are added through the jump



**Fig. 3.** Schematic diagram of the “Jump connection” module.

connection. The output of each convolution layer is followed by *ReLU* activation function and batch normalization (BN) so that the average value of the feature map is 0 with the variance of 1.

In the entire ResNet-50 structure, the input information is decircled to the output through multiple jump connections, which protect the integrity of the input to a large extent, reduce the loss of the convolutional layer to the MT response, and increase the weight of the contribution of original information to the network mapping in the training process. In the design of this study, the fully connected layer of the ResNet-50 used for a single classification task was abandoned. Instead, several convolution layers and up-sampling were used to restore the size of the feature layer to the required size corresponding to the grid size of the geoelectric model. Finally, a softmax classifier was used to classify the obtained features according to the preset resistivity range of each grid in order to obtain the 2D MT inversion results from the ResNet-50.

## 5. Experiments

### 5.1. Datasets and training

In this study, based on the Pytorch framework, a full CNN model using the ResNet-50 as the main trunk was constructed. A large number of 2D MT forward data were used to train the network to learn the mapping relationship between the apparent resistivity data and the resistivity anomaly model. The strategy for generating the training dataset was to control different characteristic values of abnormal bodies in the geoelectric model at a fixed grid of 120 km in length  $\times$  50 km in depth according to a certain step size, combine them to generate multiple resistivity values, and trap combinations to form different geoelectric models. These characteristic values included the length, thickness, position in the grid, resistivity value, and number of resistivity transition layers. Due to the large number of the variables, the number of samples obtained according to the permutation combination was too large, resulting in unbearable computational cost. Therefore, a certain production strategy was developing for datasets, such as controlling the step size of the variable variation to reduce the total number of samples and adding random values to variables to increase sample diversity. Simultaneously, the representativeness of the samples was improved to the extent possible. Accordingly, 5% Gaussian noise was added to some sample features to prevent the network from underfitting and improve the generalizability performance of the network.

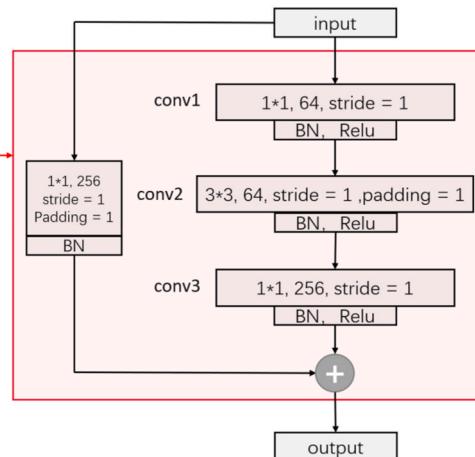
Training was completed in 50 rounds, with a batch size of 16 and with a sample size of 13,824. The code was based on the Pytorch framework, while Adam was used as the training optimizer. The initial learning rate was set to 0.001, while the activation function used was ReLU. The regularization parameter of dropout was set to 0.1. The training curve is shown in Fig. 5, where the total loss value fell to 0.0535, while the loss value of the verification set decreased to 0.1168.

### 5.2. Verification of theoretical model

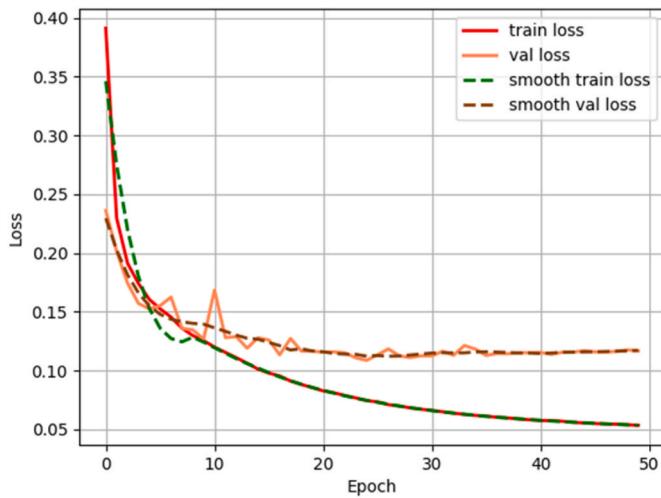
In this study, a high-low resistivity model combined with the different noise levels and the model selected from the verification dataset were used to theoretically verify the training results. The network performed well for the inversion results of the simple high-low resistivity combined anomalous body model, as shown in Fig. 6. The resistivity value, size, and location distribution of the anomalous body were well restored.

The ResNet-50 inversion results for the model response with the Gaussian noise are shown in Fig. 7. The original geoelectric model could be well restored in the noise response tests of 5% and 10%, as shown in Fig. 6, which exceeded the noise level in the training set. When the

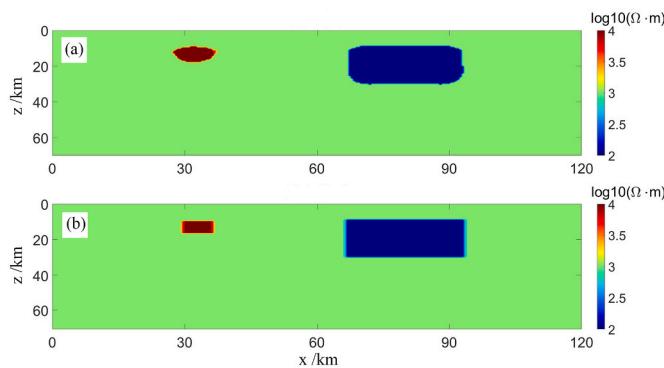
layer name	output size	ResNet-50
conv1	112×112	7×7, 64, stride 2
conv2_x	56×56	3×3 max pool, stride 2 [1 × 1, 64] [3 × 3, 64] [1 × 1, 256] × 3
conv3_x	28×28	[1 × 1, 128] [3 × 3, 128] × 4 [1 × 1, 512]
conv4_x	14×14	[1 × 1, 256] [3 × 3, 256] × 6 [1 × 1, 1024]
conv5_x	7×7	[1 × 1, 512] [3 × 3, 512] × 3 [1 × 1, 2048]
	1×1	average pool, 1000-d fc, softmax



**Fig. 4.** Schematic diagram of the ResNet-50 network (left) and its Bottleneck modules (right).



**Fig. 5.** Training curves.



**Fig. 6.** Inversion results of the simple high-low resistivity combined anomalous body model according to (a) the ResNet-50 with TE-mode and (b) the geoelectric model.

Gaussian noise level reached 20%, the reduction of the anomalous body in the geoelectric model exhibited a large deviation in size, whereas the location and resistivity of the anomalous body still remained restored.

The hold-out method was used to divide the entire data into the training and validation datasets according to a certain proportion. The

creation and best-fit of the model were based on the training dataset via supervised learning, whereas the validation dataset was not used to train the model but used to test the predictive performance of the trained model. Several samples were selected from the validation dataset for testing, as illustrated in Fig. 8. The ResNet-50 inversion results performed well for both simple and complex samples.

In order to further validate the algorithm, a theoretical model inversion test was conducted by modeling in addition. As shown in Fig. 9, the response obtained from the inversion results of the test model through forward modeling is basically consistent with the response of the original geoelectric model. This test model is independent of the training and validation sets, and the results verify the universality and basic accuracy of the network.

One of the most known issues in MT is the screening effect of a deeper conductor above a shallower and thick conductive layer.

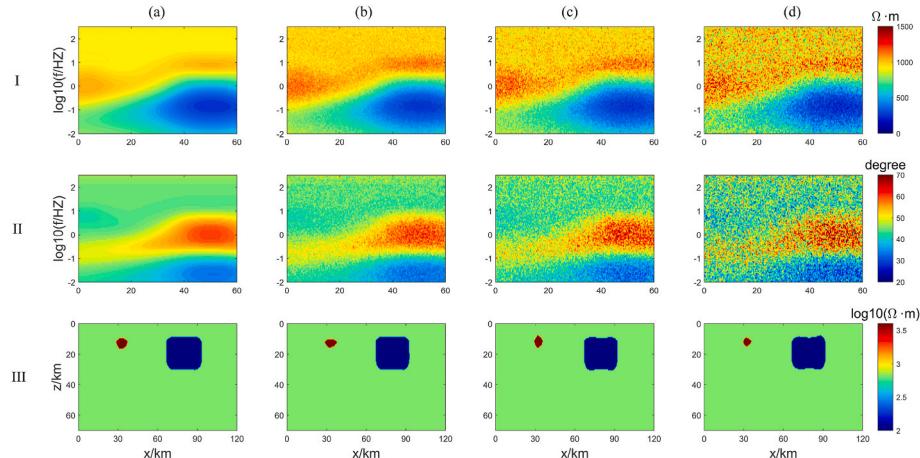
The network performs well in the inversion of the screening effect model, as shown in Fig. 10. The resistivity values, sizes, and position distributions of the low resistivity body and the target anomalous body below have been well restored.

### 5.3. Inversion of measured data

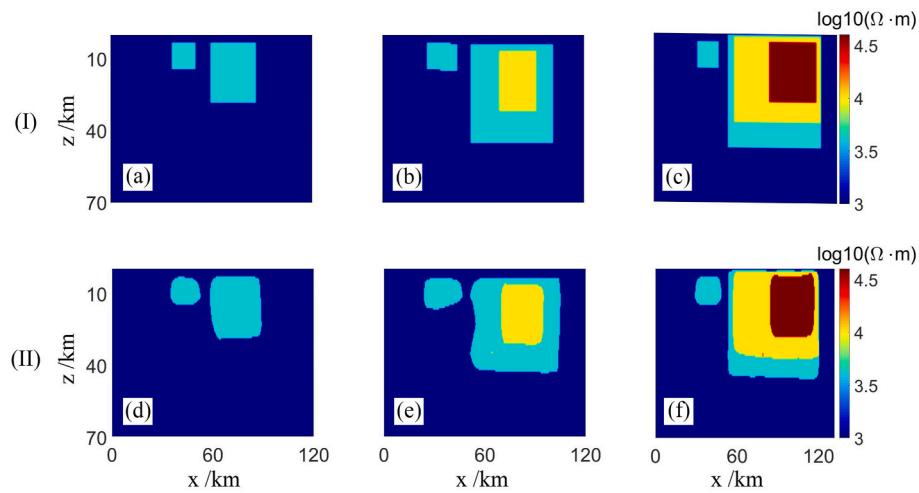
As shown in Fig. 11, measured MT data were obtained from a pre-feasibility exploration project of a deep high-temperature geothermal resource in Huanggang, Hubei Province. The design section length was 57 km, the point distance was 1 km, and the acquisition equipment was the V8 multifunctional electric method of the Phoenix Company in Canada.

Fig. 12 shows a comparison of the inversion results predicted from the different ResNet-50 models to the measured data of a geothermal field prospect in Huanggang. The first (a) and second (b) inversion results were the results of the ResNet-50 trained by using the functions of CE\_Loss and Dice\_Loss, respectively. The third inversion results (c) were derived from the Nonlinear conjugate gradient method. The CE\_Loss function exhibited a strong learning ability for the dominant classes in the sample. Due to the small proportions of the geoelectric classes of the medium and low resistivity in the entire sample size, the CE\_Loss function failed to accurately classify them with a high enough loss value.

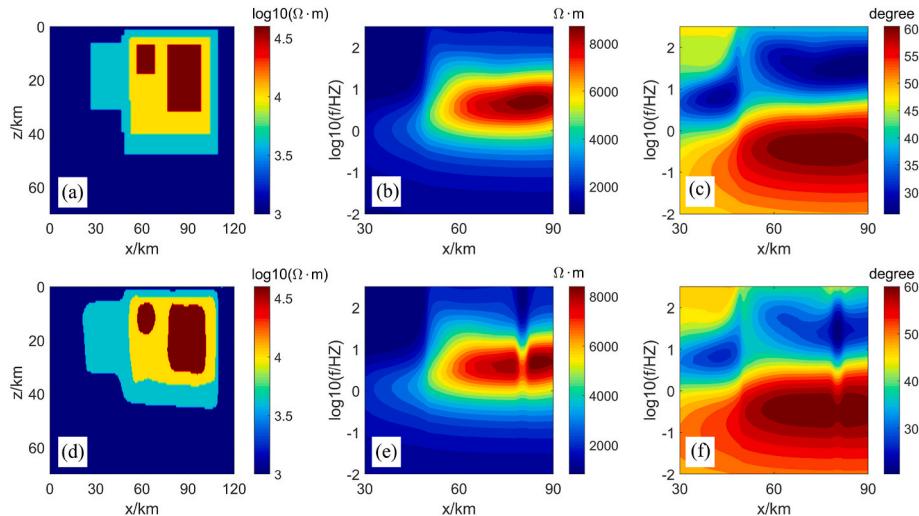
As a result, the final inversion ability of the network remained insufficient for the medium-low resistivity and medium-high resistivity samples, which results in the abnormal body on the left being mistakenly identified as high resistivity, and the positions of the two high resistivity abnormal bodies on the right being restored incorrectly. The ResNet-50 with CE\_Loss only learned the relationships for the high and low



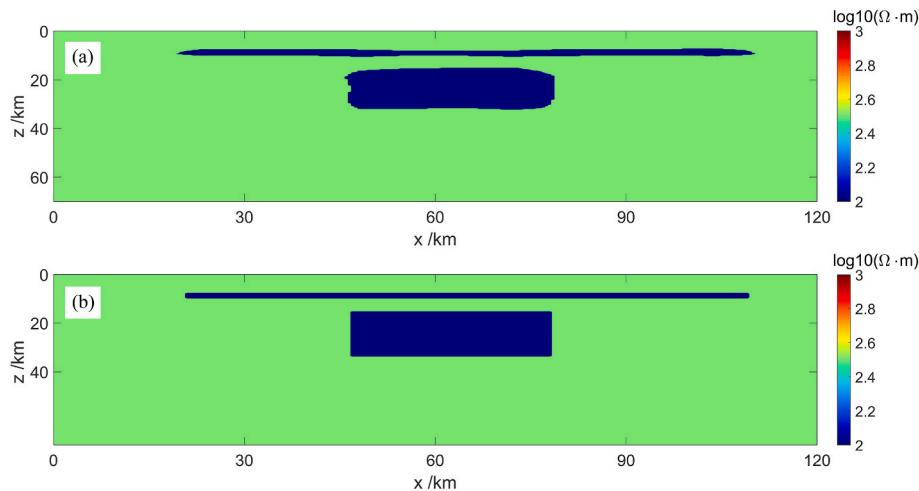
**Fig. 7.** Inversion results of the MT response in the TE mode with the different levels of Gaussian noise. Rows (I) and (II) separately show the apparent resistivity and phase data with the different levels of Gaussian noise. Row (III) shows the ResNet-50 inversion results under the different levels of Gaussian noise. Columns (a) to (d) represent the different levels of Gaussian noise added to the MT response, corresponding to 0%, 5%, 10%, and 20%, respectively.



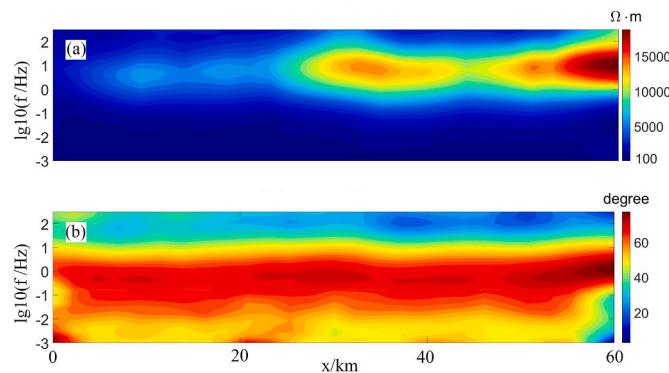
**Fig. 8.** Inversion results of the MT response in validation dataset. Rows (I) shows the ResNet-50 inversion results. Rows (II) shows the geoelectric model.



**Fig. 9.** Inversion results of test model. (a), (b) and (c) are respectively the geoelectric model, forward apparent resistivity response and forward phase response of the test model, and (d), (e) and (f) are respectively the inversion results based on the test model forward response, forward apparent resistivity response based on the inversion results of ResNet-50, and forward phase response based on the inversion results of ResNet-50.



**Fig. 10.** Inversion results of screening effect model according to (a) the ResNet-50 with TE-mode and (b) the geoelectric model.



**Fig. 11.** Measured MT data of a geothermal field prospect in Huanggang: (a) apparent resistivity pseudo-section data of the TE mode and (b) phase pseudo-section data of the TE mode.

resistivity samples, thus also resulting in the appearance of an abrupt electrical interface. Dice\_Loss function to some extent solved the shortcomings of the former. Given the overall correlation between the predicted and observed values based on the forward propagation results, the penalty for the prediction error of the medium-low resistivity and high resistivity samples with a small proportion could be improved. In addition, the sudden electrical interfaces in the inversion results caused by the limited learning ability of the ResNet-50 using CE\_Loss in the presence of the imbalanced sample size (the low proportion of the samples) could be effectively reduced. Compared with the inversion results of the conjugate gradient algorithm, the position of the basic abnormal trap was the same; however, the “focusing effect” was better.

## 6. Conclusions

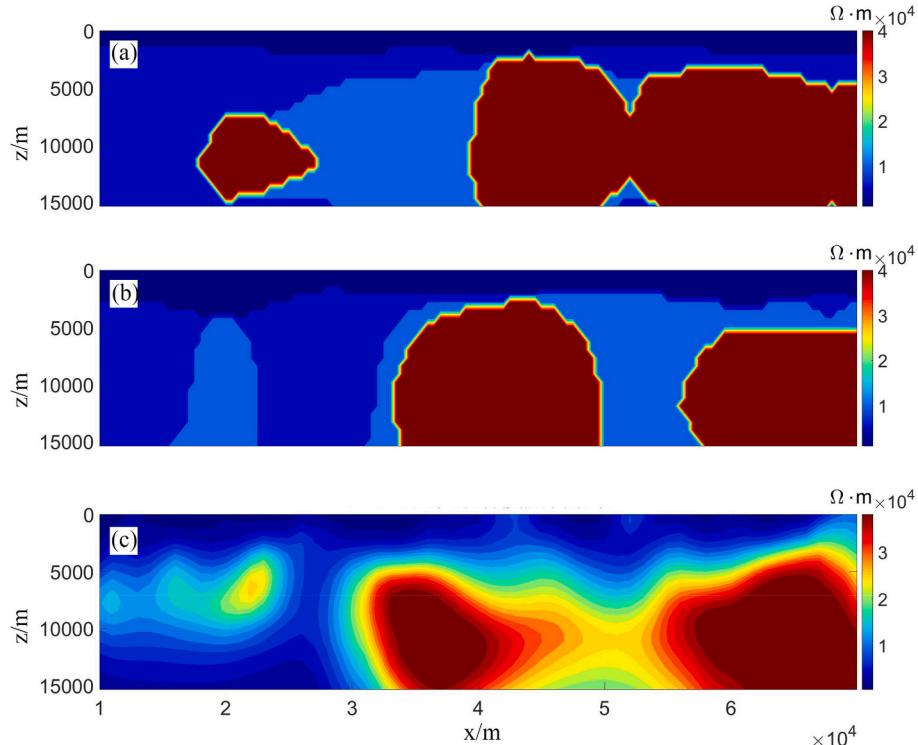
In this study, ResNet-50 CNN was redesigned, adapted, and debuted to construct a CNN structure suitable for 2D MT data inversion. The 2D

MT forward response was calculated via the triangular finite element method to produce the training dataset. The trained network basically captured the nonlinear relationship between the MT forward response and the geoelectric model, yielding better inversion results with “focusing effect”. In the verification of the theoretical model, the noise response of the geoelectric model was observed, and the denoise level exceeded the expected value of the noise level in the training dataset, which proved the robustness of the improved algorithm. The addition of Dice Loss as the loss function to ResNet-50 significantly reduced the issue of an abrupt electrical interface in the measured MT data inversion. Compared with the traditional deterministic inversion methods, the DL-based inversion algorithm does not require multiple iterations for model predictions when used in applications, thus significantly improving the computational cost and efficiency of MT data inversion. Moreover, it does not rely on the selection of the initial model parameters and does not fall into the local optima, which makes it perform well in screening effect testing.

At the current level of geophysical exploration, observation data inevitably contains noise and errors, which in general leads to the multiplicity of geophysical inversion solutions. There are often many models used to fit observation data, and the complexity can be arbitrarily high. However, due to issues such as ill posed nature, they often cannot achieve complete fitting results. If we change our thinking and seek to develop the fitting model in the most simplified direction to reflect the most basic feature information of the underground real model, the computational cost may be smaller and the goal may be easier to achieve. The methods in our work may fit this idea. By combining enough simple models for training, a “simple enough” network model can be fitted, which can restore the simplest model in the inversion of measured data and highlight the most fundamental features of the real geoelectric cross-section.

## Authorship statement

LiAn Xie wrote the draft and code, built the network and dataset,



**Fig. 12.** Comparisons of the inversion results to the measured MT data of a geothermal field prospect in Huanggang: (a) ResNet-50 inversion results using CE\_Loss (rms:11.474); (b) ResNet-50 inversion results using Dice\_Loss (rms:4.825); (c) inversion results using the Nonlinear conjugate gradient algorithm (rms:2.741).

completed the training, and finished the experiment. Ningbo Bai derived part of the theory and revised the expression of the formula and draft, and checked the code. Bo Han supervised the method and provided the methodological support. Xiangyun Hu provided data and improved conceptualization.

### Declaration of competing interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

### Acknowledgements

This research was co-funded by the National Natural Science Foundation of China (No. 42220104002, 42174095, U1812402, and 41630317), the Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education (No. GLAB2022ZR10) and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan).

### References

- Constable, S.C., Parker, R.L., Constable, C.G., 1990. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data[J]. *Geophysics* 52 (3), 289–300.
- Di, Q.Y., Xue, G.Q., Yin, C.C., Li, X., 2020. New methods of controlled-source electromagnetic detection in China[J]. *Sci. China Earth Sci.* 63 (9), 1268–1277.
- El-Qady, G., Ushijima, K., 2001. Inversion of DC resistivity data using neural networks [J]. *Geophys. Prospect.* 49 (4), 417–430.
- Guo, R.W., Dosso, S.E., Liu, J.X., Dettmer, J., Tong, X.Z., 2011. Non-linearity in Bayesian 1-D magnetotelluric inversion[J]. *Geophys. J. Int.* 185 (2), 663–675.
- He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (9), 1904–1916.
- He, S.Y., Cai, H.Z., Liu, S., Xie, J.T., Hu, X.Y., 2021. Recovering 3D basement relief using gravity data through convolutional neural networks[J]. *J. Geophys. Res. Solid Earth* 126 (10).
- Li, J.F., Liu, Y.H., Yin, C.C., Ren, X.Y., Su, Y., 2020. Fast imaging of time-domain airborne EM data using deep learning technology[J]. *Geophysics* 85 (5), E163–E170.
- Liu, W., Xi, Z.Z., Wang, H., Zhang, R.Q., 2021. Two-dimensional deep learning inversion of magnetotelluric sounding data[J]. *J. Geophys. Eng.* 18 (5), 627–641.
- Moghadas, D., 2020. One-dimensional deep learning inversion of electromagnetic induction data using convolutional neural network[J]. *Geophys. J. Int.* 222 (1), 247–259.
- Montahaei, M., Oskooi, B., 2014. Magnetotelluric inversion for azimuthally anisotropic resistivities employing artificial neural networks[J]. *Acta Geophys.* 62 (1), 12–43.
- Newman, G.A., Hoversten, G.M., Alumbaugh, D.L., 2002. Three-dimensional magnetotelluric modeling and inversion: application to sub-salt imaging[M]. *Methods Geochem. Geophys.* 35, 127–152. Elsevier.
- Schwarzbach, C., Borner, R.U., Spitzer, K., 2005. Two-dimensional inversion of direct current resistivity data using a parallel, multi-objective genetic algorithm[J]. *Geophys. J. Int.* 162 (3), 685–695.
- Sharma, S.P., 2012. VFSARES—a very fast simulated annealing FORTRAN program for interpretation of 1-D DC resistivity sounding data from various electrode arrays. *Comput. Geosci.* 42, 177–188.
- Shaw, R., Srivastava, S., 2007. Particle swarm optimization: a new tool to invert geophysical data[J]. *Geophysics* 72 (2), F75–F83.
- Shi, X.M., Wang, J.Y., 1998. One dimensional magnetotelluric sounding inversion using simulated annealing. *Earth Sci. J. China Univ. Geosci.* 23 (5), 542–546 (in Chinese).
- Siripunvaraporn, W., Egbert, G., Lenbury, Y., 2002. Numerical accuracy of magnetotelluric modeling: a comparison of finite difference approximations[J]. *Earth Planets Space* 54 (6), 721–725.
- Smith, J.T., Booker, J.R., 1991. Rapid inversion of two-and three-dimensional magnetotelluric data[J]. *J. Geophys. Res. Solid Earth* 96 (B3), 3905–3922.
- Yang, Q.G., et al., 2022. 3-D gravity inversion based on deep convolution neural networks[J]. *Geosci. Rem. Sens. Lett. IEEE* 19.
- Zhang, Z.D., Alkhalifah, T., 2019. Regularized elastic full-waveform inversion using deep learning[J]. *Geophysics* 84 (5), R741–R751.
- Zhang, Z.H., et al., 2021. Joint gravity and gravity gradient inversion based on deep learning[J]. *Chin. J. Geophys.* 64 (4), 1435–1452.