Article
Magnetotelluric Deep Learning Forward Modeling and Its Application in Inversion

Fei Deng, Jian Hu *, Xuben Wang, Siling Yu, Bohao Zhang, Shuai Li and Xue Li

College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China; dengfei@cdut.edu.cn (F.D.); wxb@cdut.edu.cn (X.W.); yusiling@stu.cdut.edu.cn (S.Y.); zbh981101@163.com (B.Z.); stulishuai@163.com (S.L.); lxsnowberry@163.com (X.L.)

* Correspondence: huda@stu.cdut.edu.cn

Abstract: Magnetotelluric (MT) inversion and forward modeling are closely linked. The optimization and iteration processes of the inverse algorithm require frequent calls to forward modeling. However, traditional numerical simulations for forward modeling are computationally expensive; here, deep learning (DL) networks can simulate forward modeling and significantly improve forward speed. Applying DL for forward modeling in inversion problems requires a high-precision network capable of responding to fine changes in the model to achieve high accuracy in inversion optimization. Most existing MT studies have used a convolutional neural network, but this method is limited by the receptive field and cannot extract global feature information. In contrast, the Mix Transformer has the ability to globally model and extract features. In this study, we used a Mix Transformer to hierarchically extract feature information, adopted a multiscale approach to restore feature information to the decoder, and eliminated the skip connection between the encoder and decoder. We designed a forward modeling network model (MT-MitNet) oriented toward inversion. A sample dataset required for DL forward was established using the forward data generated from the traditional inverse calculation iteration process. The trained network quickly and accurately calculates the forward response. The experimental results indicate a high agreement between the forward results of MT-MitNet and those obtained with traditional methods. When MT-MitNet replaces the forward computation in traditional inversion, the inversion results obtained with it are also highly in agreement with the traditional inversion results. Importantly, under the premise of ensuring high accuracy, the forward speed of MT-MitNet is hundreds of times faster than that of traditional inversion methods in the same process.

Keywords: magnetotelluric inversion; Mix Transformer; deep learning forward

1. Introduction

Magnetotelluric (MT) inversion is a branch of geophysics, and inversion is a method used to determine the corresponding geophysical model (resistivity) from given observations (apparent resistivity and phase). Since the 1950s when the MT method was introduced, MT inversion methods have been playing an important role in various applications, including marine geophysics, environmental geophysics, understanding the electrical structure of the Earth’s crust and upper mantle, geological exploration, and mineral exploration [1,2].

Inversion algorithms can be categorized into indirect inversion and direct inversion based on the relationship between data and models. Indirect inversion requires an initial model and boundary conditions to use iterative optimization methods to find the minimum fitting difference after multiple iterations. Occam inversion [3,4], rapid relaxation inversion [5], nonlinear conjugate gradient inversion [6,7], and Gauss–Newton inversion [8,9] are frequently utilized in the MT inversion of the Earth because of their high accuracy and computational stability, although their efficiency and speed are lower. Direct inversion techniques, on the other hand, perform forward approximations to determine the relationship between field values and models and require no initial model to produce inversion...
results. Nonlinear direct inversion methods such as the BOSTICK inversion [10] algorithm, genetic algorithm, and artificial neural network (ANN) [11] can enhance efficiency; however, their inversion accuracy and stability are restricted, necessitating further development and improvement.

The exponential growth in computing technology has led to the increased use of ANNs and DL research methods in the field of MT. In contrast to the conventional linearized inversion theory approach, the DL-based MT approach involves creating sample datasets, designing and training specific network models, and ultimately achieving fast prediction from data models of the results. The computational efficiency of the model to the result is greatly improved when the network training reaches convergence. For instance, Fang et al. [12] improved the overall quality of inversion on low-frequency seismic data modules using an ANN. Similarly, Moseley et al. [13] predicted the response velocity of acoustic and seismic waves generated in a horizontally layered medium using an ANN. Hansen and Cordua [14] utilized DL to rapidly calculate the forward response for improved efficiency in the inversion of ground-penetrating radar datasets. Li et al. [15] trained an end-to-end ANN to invert the velocity model, fully utilizing the feature information from seismic data. Puzyrev [16] demonstrated the significant potential of deep learning in inversion by predicting resistivity distributions using trained NNs with deeper and more complex layers on small datasets. Puzyrev and Swidinsky [17] discovered that a complex NN could extract richer feature information and predict accurate results but is often expensive to train and does not generalize well. Liu et al. [18] constructed a theoretically driven one-dimensional DL magnetotelluric inversion, but the DL inversion produces low-precision EM responses and cannot be used in two-dimensional geoelectric inversion. Conway et al. [19] accelerated MT inversion by utilizing neural networks for forward computation. However, the anomalies in training geoelectric models were relatively simple, and the average accuracy of the network was only 70%. Liu et al. [20] used a U-shaped neural network architecture for the inversion of resistivity data, whereas Guo et al. [21] applied a supervised descent method along with a priori information to accelerate the inversion process for transient electromagnetic data. However, they used neural networks with less complex layers and structures, and the sample datasets of the networks were artificially generated rather than being obtained based on inversion algorithms. Therefore, it was difficult for the trained network models to have strong generalization ability and higher accuracy.

Although traditional MT forward modeling calculations are highly accurate, they are time-consuming and inefficient. Additionally, the inversion process necessitates multiple calls to forward modeling calculations and multiple iterations to determine the most optimal solution. In contrast to inversion, forward modeling does not offer multiple solutions. Deep learning can approximate the nonlinear mapping relationship of forward modeling, thus enabling quicker prediction from the model to results and enhancing efficiency. This increased efficiency is advantageous for inversion calculations that depend on forward modeling as a vital parameter to solve the problem. As such, the network model should possess a broader receptive field, powerful global modeling capability, and an accurate response to small model changes, allowing for high prediction accuracy in the forward modeling results. Based on experimental results, the main contributions of this article are:

1. The construction of a new forward network dataset to serve MT inversion. To ensure the effectiveness of DL networks, constructing an appropriate dataset for the training samples is crucial for MT forward networks. The inverse forward response model is generated iteratively through an optimization algorithm during the inversion procedure, which differs from real subsurface models or human-made models. We used different geoelectric models for MT forwarding and input the inversion program after obtaining the observed data. By acquiring the forward data generated by the inversion process to build a forward sample dataset that fits the inversion iterative model, an excellent forward network model oriented to inversion was trained.
2. We designed the forward modeling network MT-MitNet for forward computations. MT-MitNet takes Mix Transformer [22] as the backbone of the network, obtains rich feature information of the anomalies in the encoding module, reconstructs anomalous features in the decoding module by combining multiscale ideas, and eliminates the skip connection between the encoder and decoder. After training, MT-MitNet, with global modeling capability, displayed stable and rapid forward modeling calculations, with an average accuracy greater than 95%. MT-MitNet serves the forward calculation in inversion with high accuracy, which improves the overall efficiency of inversion calculation, and the network model has strong generalization ability in the face of multiple anomalies.

2. Methodology
2.1. Dataset Preparation

Preparing sample data is a key step in conducting research on deep learning network models. In the field of computer vision, training networks often use abundant, accurately labeled, and open-source datasets. However, for MT data, it is quite difficult to obtain millions of actual samples. It is worth noting that traditional inversion algorithms, such as the Occam inversion, require a large number of forward computations to reduce the fitting error during the iterative process of inversion computation searching. The data generated by this process (apparent resistivity, phase, and resistivity models) are produced by the inversion algorithm, providing a highly accurate sample for deep-learning-based forward modeling networks. This not only avoids the uncertainty of manually produced forward modeling samples but also facilitates more effective training of the forward modeling network models. The Occam inversion algorithm is based on the idea of smooth inversion to achieve smooth inversion results with strong stability. However, the required multiple forward computations result in time-consuming and inefficient computation. In this study, we utilized the Occam inversion method as an example to conduct experiments on the forward computation process within the deep learning replacement inversion algorithm. The aim was to predict high-precision forward results through deep learning networks and to use them in inversion computation.

In actual engineering environments, the subsurface model is varied within a certain range. The two-dimensional geoelectric model simulated in this study was 4096 m in length and 1024 m in height, with a frequency range between $10^{-2}$ and $10^2$, and the number of frequency points was set to 32. The observation points were spaced 32 m apart, and the number of measurement points was set to 128 so that the size of both the simulated geoelectric model and the original observation data was $32 \times 128$. The anomalies were located in a homogeneous half-space and consisted of positive circles, ellipses, and arbitrary (shaped) closed figures with random distribution locations and resistivity values between $10^{-2}$ and $10^4$ Ωm with a background resistivity of 100 Ωm.

The process of producing the sample database for the MT-inversion-oriented forward network model is shown in Figure 1. The anomalous body dataset comprises 7 anomalous body categories, each divided into 1000 groups based on location in space and resistivity value. Initially, an anomalous body sample is selected from the dataset, which is subjected to MT forward calculation to obtain the apparent resistivity and phase results. These results constitute the observed data necessary for standard Occam inversion calculation. During the iterative process of the Occam inversion algorithm, forward data are generated and saved, and ten data groups are selected randomly and saved in the dataset. The aforementioned process is repeated for each sample set, resulting in a total of 70,000 groups of forward data.
Randomly save ten groups of forward data generated by Occam inversion iterative calculations

2.2. Inversion-Oriented Deep Learning Forward Network Model: MT-MitNet

In practical applications of MT inversion, people are more concerned about the size and electrical characteristics of the anomalous bodies. For the inversion results, they hope to accurately determine the resistance value of the anomaly and the boundary situation of the anomaly delineated from the surrounding normal rock. In the last decade, many researchers have conducted MT forward and inverse studies using recurrent neural networks (RNNs) [23,24] or convolutional neural networks (CNNs) [25–28]. The MT forward results have no obvious edges, and the global information is quite important, but RNNs cannot effectively capture the data-dependent information of ultra-long sequences, and the parallelized computational capability is less efficient and more time-costly. The convolution in CNNs is limited by the perceptual field size and has the problems of translation invariance and focusing only on local information, making it difficult to establish global information at a larger scale. The network model based on the Transformer [29] architecture improves these problems, and the Transformer is able to perform parallel computation and process more information in less time, making up for the lower efficiency of RNNs in
parallel computation. Additionally, the Transformer utilizes a self-attention mechanism, showcasing exceptional learning ability and parallel performance in the processing of sequential data. As a result of this self-attention mechanism being present, the encoding module possesses a larger effective receptive field in comparison with traditional CNN encoders. This allows Transformer-based models to acquire a more comprehensive range of global information while possessing dynamic global modeling capabilities. It can capture information regarding global features and compensate for the limitations of convolutional neural networks in acquiring global information. Based on the excellent performance of the Transformer architecture, Vision Transformer (ViT) [30] and Mix Transformer (MiT) [22] for two-dimensional data have emerged one after another. ViT transforms the input information into vectors, and key features may be ignored, while MiT comprises a hybrid attention module based on ViT architecture, which uses a hierarchical extraction mechanism to obtain feature map information, and it is able to pay attention to both feature extraction and target information integration and has a simple structure and higher efficiency. Considering the various sizes and spatial distributions of anomalies, the magnetoelectric method places great emphasis on achieving accurate overall results. Interactions exist among multiple anomalies within a specific area, and the inversion process entails conducting numerous forward calculations and iterative optimization to obtain the optimal model. The network model should accurately and effectively extract as much feature information about the anomalies as possible. Moreover, using deep learning for forward simulation and applying it in inversion requires a network with high computational precision to accurately respond to small model modifications. MiT has self-attention and hierarchical feature extraction mechanisms, and its receptive field is larger than that of traditional CNNs, allowing it to capture more detailed and complete feature information.

The importance of global information and the need for accuracy led us to choose MiT with global modeling capabilities as the backbone of the network. In addition, MiT has a skip-connection structure that superimposes feature information from the encoding module onto the corresponding decoding module. But the magnetoelectric forward simulation results are progressively smoother, without obvious edges and without requiring redundant superimposition of feature information. The skip-connection structure incorporates the preserved anomalous data and excessive edge information from the encoding module into the forward result [31], leading to inadequate smoothness in the forward results. Therefore, we removed the skip-connection structure between the encoding and decoding modules. Considering the characteristics of forward and inversion MT data, the forward network model MT-MitNet for MT inversion is proposed based on Mix Transformer with low computation and hierarchical extraction of features, as shown in Figure 2.

MT-MitNet consists of an encoding module and a decoding module. The encoding module consists of 4 Mix Transformer Blocks (MiT-Blocks); each MiT-Block consists of efficient self-attention module, mix feed-forward network (Mix-FFN) module, and overlapped patch-merging module. The hierarchical coding module in MiT achieves global modeling through a self-attentive mechanism. The Mix-FFN mainly consists of two lightweight multilayer perceptrons (MLPs), GELU activation, and a 3 × 3 convolution layer. Meanwhile, the Mix-FFN utilizes the translation invariance of convolution operations to incorporate location information into the Mix Transformer. This approach offers increased flexibility compared with the location coding of the original Transformer. The structure of Mix Transformer Block is shown in Figure 3.

Two-dimensional data are not conducive to self-attention calculations. Before entering MiT-Block1, the input data undergo overlapping patch-merging with a convolutional kernel size (K) of 7, stride (S) of 4, and padding (P) of 3 to reduce the data size to 1/4 and obtain a one-dimensional sequence. Then, the data passes through four layers of MiT-Blocks. When entering MiT-Block1, the Overlapping Patch Merging process is skipped, and the output of MiT-Block restores the sequence to two-dimensional. For the subsequent three MiT-Blocks, the parameters of Overlapping Patch Merging (K, S, P) are set to 3, 2, and 1 respectively. After each MiT-Block, the data size reduces to half of the previous size, allowing interaction
between different patches to maintain local continuity and obtain a sequence of feature information. After four MiT-Blocks, we can get $1/4, 1/8, 1/16, 1/32$ features $T_i$, $i \in (1, 2, 3, 4)$, as shown in Equation (1).

$$T_i = \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i.$$  

(1)

where $C_i$ refers to the number of channels of the output, $C_i \in (64, 256, 320, 512)$.

The computational efficiency of the encoding module is mainly constrained by the efficient self-attention layer. The original complexity of multihead attention computation is $O(N^2)$, where $N$ represents the length of the sequence, defined as $N = H \times W$. MiT addresses this issue by introducing a reduction ratio ($R$), reducing the complexity of self-
attention computation to \( O(N^2) \). This allows the network to effectively focus on relevant features, as illustrated in Equation (2). Moreover, unlike ViT, which produces only a single feature, MiT generates multilevel features, akin to CNN. The integration of these multiple layers of features into the decoding module facilitates the training of a more efficient network model.

\[
\begin{align*}
K' &= \text{Reshape}(N, C \cdot R)(K) \\
K &= \text{Linear}(C \cdot R, C)(K') \\
\text{Attention}(Q, K, V) &= \text{Softmax}(\frac{QK}{\sqrt{d_{\text{head}}}})V
\end{align*}
\]  

(2)

where Query (Q), Key (K), and Value (V) have the same dimension \( N \times C \), where Q, K, and V represent input feature sequences. \( R \) refers to the reduction ratio of the efficient self-attention, and the values of \( R \) for the 4 MiT Blocks were set to 64, 16, 4, and 1, respectively. \( K \) represents the decayed \( K \). Linear is a linear layer, and \( d_{\text{head}} \) represents the dimension computed by each attention head, set to 64. The Softmax function is applied to obtain feature weights within the range of 0 to 1, which are then multiplied with the corresponding Value to obtain self-attention.

Magnetoelectric anomaly bodies can manifest at different locations and ranges within a geoelectric model. The encoding module generates multilevel feature information (\( T_i \)) for magnetoelectric forward simulation data. During the upsampling process, the use of a single-size convolutional kernel may result in the loss of some feature information pertaining to multiple anomaly bodies. We employed convolutional kernels of different sizes to handle the hierarchical feature information, allowing each channel to focus more on specific details corresponding to the respective scale. This enhances the decoding process’s ability to recover anomaly positions and ranges. Additionally, smaller convolutional kernels reduce the complexity of parameters and computations without affecting the receptive field. Based on the convolution kernel and hierarchical features at different scales, we developed a multiscale adaptive decoding structure, whose structure is shown in Figure 4.

The decoder consists of four multiscale adaptive modules; each utilizes a multihead design. Each multiscale adaptive module (MSAM) consists of seven convolutional kernels of varying sizes and one convolutional block attention module (CBAM) [32]. The CBAM is a lightweight attention module that can conduct attention operations in both spatial and channel dimensions, thereby accentuating the significance of crucial feature information. Numerous studies have demonstrated consistent performance improvements by integrating CBAM into various models [33,34]. During the decoding process, the first MSAM receives the multilevel feature information (\( T_1 \)) generated by MiT-Block4 and uses it as the input tensor (\( F_1 \)). \( F_1 = H_32 \times W_32 \times 512 \). Using a 1 × 1 convolution, the input tensor is combined with channel information. To extract detailed features, a 3 × 3 convolution is applied. Convolutions of various sizes, including 3 × 7, 7 × 3, 5 × 11, and 11 × 5, are utilized to capture features at different scales along both the horizontal and vertical directions. The output tensors retain the same height (H) and width (W) dimensions as the input tensor but with a reduced channel size of 1/4. By incorporating tensors obtained from various convolutions and establishing residual connections with the input, (\( F_1 \)) integrates multiscale information. In the case of magnetoelectric results within distinct geoelectric models, different aspects are emphasized by utilizing the spatial and channel attention mechanism of CBAM. Subsequently, a 3 × 3 convolution is performed to obtain tensor \( F_1' \). After upsampling, the \( H \) and \( W \) of \( F_1' \) increase by a factor of 2, while the channel count decreases from 512 to 320, resulting in tensor \( F_2 \). \( F_2 = H_{16} \times W_{16} \times 320 \). The above process can be represented by Equation (3). By repeating this process, we obtain \( F_3 = H_{8} \times W_{8} \times 128 \) and \( F_4 = H_{4} \times W_{4} \times 64 \). Finally, \( F_4 \) maintains the same number of channels; after upsampling, the
size becomes $\frac{H}{4} \times \frac{W}{4} \times 64$. Then, through output convolution and reducing the number of channels, the output size of $H \times W \times 1$ is obtained.

\[
\begin{align*}
F'_i &= \text{MSAM}(F_i), i \in (1, 2, 3, 4) \\
F_{j+1} &= \text{Upsample}(F'_j), j \in (1, 2, 3, 4)
\end{align*}
\]  

(3)

where MSAM() denotes the process of the input tensor $F_i$ being processed by MSAM to obtain $F'_i$, and Upsample () denotes the process of upsampling.

Figure 4. Multiscale adaptive module decoding structure.

2.3. Network Training

In this study, following the classification method of Reitermanova training network [35], the 70,000 groups of data in the sample database were randomly divided in a ratio of 8:1:1: 56,000 groups in the training set, 7000 groups in the validation set, and 7000 groups in the test set were obtained.

The input data for network training are continuously downsampled in patch embedding and merging, and after the last encoding module, the size of the obtained data is $1/32$ that of the original size. If the sample data are directly input into the forward network, the semantic information is seriously lost after four downsamplings. To avoid this situation, we interpolated the $32 \times 128$ sample data into the $256 \times 1024$. The encoder continuously acquires the original feature information through four MiT modules and gradually recovers to the output size of the forward result via transpose convolution and upsampling through four multiscale adaptive modules.

The anomalous body model has apparent resistivity values between $10^{-2}$ and $10^4$ $\Omega$m [36]. Sample data with too large a range increase the training time of the network and may make it difficult to converge, while normalized data have better performance in network
training [37]. Therefore, normalizing the sample data before training reduces the stress of training while obtaining a better-performing network model.

In this study, we adopted the idea of maximum–minimum normalization: \( Rho' \) is obtained by taking the logarithm of the resistivity and apparent resistivity data according to Equation (4) with a base of 10, at which time the data range is reduced to between \(-2\) and 4. The maximum and minimum values of each data are counted separately, and the geoelectric model data, apparent resistivity data, and phase data are normalized according to Equations (5)–(7), respectively. When the network training is completed, reflective injection is executed to reduce the data to the original size.

\[
Rho' = \log_{10} Rho
\]  
\[
RHO = \frac{Rho'_1 - Min_1}{Max_1 - Min_1}
\]  
\[
ARHO = \frac{Rho'_2 - Min_2}{Max_2 - Min_2}
\]  
\[
PHS = \frac{Phs - Min_3}{Max_3 - Min_3}
\]

where \( Rho'_1 \) is the resistivity data of the log10-processed geoelectric model, \( Rho'_2 \) is the log10-processed apparent resistivity data, \( Phs \) is the phase data, and \( Max_i \) and \( Min_i \) \((i = 1, 2, 3)\) denote the maximum and minimum values corresponding to each of the resistivity data, apparent resistivity data, and phase data, respectively. \( RHO \) denotes normalized geoelectric model resistivity data, \( ARHO \) denotes normalized apparent resistivity data, and \( PHS \) denotes normalized phase data.

The speed and performance of network convergence are influenced by the optimizer and loss function. For the MT-MitNet forward network model, the Adam [38] adaptive learning rate optimization algorithm is used for training, and the initial learning rate is 0.01. Forward and inversion MT modeling has high requirements for the accuracy of numerical simulations. The SmoothL1Loss loss function performs better for model convergence, anomalous body handling, and gradient explosion prevention, whereas the mean square error loss function creates the problem of gradient explosion at outlier points. Therefore, the SmoothL1Loss loss function was used in this study, which is calculated as follows:

\[
\text{SmoothL1Loss}(X, Y) = \frac{1}{N} \sum_{i=0}^{N} \left\{ \begin{array}{ll}
0.5(X_i - Y_i)^2, & |X_i - Y_i| < 1 \\
|X_i - Y_i| - 0.5, & |X_i - Y_i| \geq 1
\end{array} \right.
\]

where \( X \) is the output of the MT-MitNet forward network, \( Y \) is the result of the conventional algorithm, and \( N \) is the number of samples. \( |X - Y| \) indicates the difference between them. When \( |X - Y| \) is less than 1, the squared error is used; otherwise, the linear error is used.

The experiments were conducted on an NVIDIA GeForce RTX 3090 GPU with 24 G VRAM. The training dataset consisted of 56,000 samples, and the validation and test datasets each contain 7000 samples. For the experiments, we employed the Adam optimizer with an initial learning rate of 0.01. The learning rate was reduced to 1/5 of its original value if the cumulative loss did not decrease for three consecutive epochs. The batch size was set to 2, and the training was performed for 100 epochs. The training was stopped when the validation loss remained within a small range of variation. The MT-MitNet forward network reduced the loss value to \(5.0 \times 10^{-5}\) in the first 10 rounds, stabilizing after 70 epochs. When the training epoch reached 80, the model loss value dropped to \(4 \times 10^{-6}\), the learning rate was \(4.0 \times 10^{-7}\), and the validation loss curve oscillated and stabilized in the range of \(1.1 \times 10^{-5}\) to \(1.4 \times 10^{-5}\), the network model training was finished. The training and validation curves are shown in Figure 5.
3. Experiments

3.1. MT-MitNet Forward Network Model Replaces the Forward Computation Module of Occam Inversion Program

The traditional Occam inversion program is developed in Fortran language, and the inversion starts with the creation or input of the forward response data and the geoelectric model. The inversion algorithm performs several forward runs to find the best iterations for the purpose of reducing the fitting error, but the forward computation is less efficient. Using the trained MT-MitNet network can effectively reduce the forward computation time. However, the Occam inversion program runs in the Fortran development environment, and the forward network model is generated in the Python (PyTorch) environment, both of which cannot be directly accessed and used. In order for the inversion program to successfully use the results predicted by the forward network when calculating the forward response, we wrote the forward network model as a dynamic link library that the inversion program can call. The dynamic link library includes methods for loading and using the MT-MitNet model, data transfer between the Occam program and the MT-MitNet model, and data format conversion. PyTorch [39] provides methods to migrate between different environments; in order to enable the Occam inversion program (Fortran) to call the MT-MitNet model for the forward computation, the steps are as follows: (1) convert MT-MitNet from a PyTorch model to a TorchScript model by utilizing TorchScript; (2) develop a dynamic link library for C++ model calling based on the PyTorch C++ API library (LibTorch), named ForwardScript, which can load the TorchScript model and compute the forward response based on the input model, and outputs it; (3) introduce ForwardScript into the Occam inversion main program (Fortran), replacing the original forward function in the Occam program and calling the ForwardScript deep learning forward dynamic link library instead. The implementation occurs in the form of a dynamic library, so that the Occam inversion program developed by the Fortran program calls the model loading function in the library function when the inversion is executed, calls the forward function in the dynamic link library when the Occam inversion program calculates the forward, and receives the predicted forward result returned by the dynamic library, finally allowing the forward network model to be successfully deployed in the inversion program to complete the replacement of the forward calculation. A simplified flow of the MT-MitNet network model replacing the forward computation module of the Occam inversion program is shown in Figure 6.

The simplified process of MT-MitNet Occam inversion is as follows: the anomaly model and the forward response data are input into the network, and the data are reflected and combined into an array of forward responses at the end of the network computation and passed into the Occam inversion program. The inversion algorithm itself judges the
computation process based on the fit difference and model roughness. If the inversion result does not reach the requirement of the end of the algorithm, the step length is reduced and the search for the optimal Lagrange multiplier continues. The computation finishes when the fit difference of the inversion result meets the fixed value or the model is smooth enough. We evaluated the optimal convergence of the model using the fit difference, which is expressed as Rms.

Figure 6. Simplified MT-MitNet Occam inversion process.

In this study, we deployed the inversion-oriented deep learning forward network model to realize the loading and use of the forward computational model of the replaced Occam inversion program in C++ development environment, and we used the well-trained MT-MitNet to achieve fast and accurate forward computation in order to improve the efficiency of forward computation in inversion.

3.2. Experimental Analysis of the Inversion of MT-MitNet Replacement Forward Calculation

To evaluate the effectiveness of the MT-MitNet forward network in MT inversion applications, we randomly selected a geoelectric model with three anomalies. The model has a background resistivity of 100 Ωm and anomaly resistivity values of 3.16 Ωm, 1 Ωm, and 0.03 Ωm, respectively, from left to right, as shown in Figure 7. Firstly, the MT forward calculation is performed on the geoelectric model to obtain the forward response result. Then, the forward results were combined with the geoelectric model to form the observation data, which were further normalized and input into the Occam inversion program. Both the traditional Occam inversion method and the replacement of Occam’s forward modeling calculation process with MT-MitNet were used separately for computation. The iteration process stopped when the model reached a convergence point (reaches the target fitting difference), as shown in Figure 8.
Figure 7. The three anomaly models and their MT forward response results, from left to right, are anomaly geoelectric model, apparent resistivity results, and phase results.

Figure 8. The results of the 1st iteration, 5th iteration, and 9th iteration of the traditional Occam inversion and MT-MitNet replacement forward computations, from left to right, are resistivity results, apparent resistivity results, and phase results, respectively.

The Occam inversion computed by MT-MitNet replacement forward and the traditional Occam inversion both converged and performed well after nine iterations, with Rms values of 1.805 and 1.88, respectively. Based on the iterative process and results, it could be observed that each round of the inversion iteration showed high correspondence between the model based on the MT-MitNet replacement forward computation and the traditional inversion method. The former also simulated the anomaly model, apparent resistivity, and phase with a high degree of similarity. Nine rounds of iterations of the model fit difference decline curve are shown in Figure 9. The algorithm is computed several times during each iteration to find the minimum fitting error in this round. Figure 9 marks the end position of each iteration with dashed red circles, corresponding to the results of rounds 1, 5, and 9 in Figure 8.

The anomalous body model was computed by stopping at nine iterations. The traditional Occam inversion method computed the forward response in about 2808.6 s, while the MT-MitNet achieved this task in about 4.3 s, indicating a remarkable increase in computational efficiency. After iterative computations, both methods accurately represented the size and location of the anomalous body.
Figure 9. A plot of the Rms from the optimized model searched for at each iteration of the inversion process, as calculated by the traditional Occam inversion (blue) and the MT-MitNet replacement forward calculation (orange). The inversion algorithm invokes the forward computation several times during the inversion process to find the optimal model in an optimization-seeking iteration. The multiple points in the figure represent the optimized models found by the Occam inversion algorithm in each iteration.

In order to compare the accuracy difference between the inversion results of the forward response calculated by the deep learning network and the traditional inversion results, we generated images of the data results using Python 3.7, and according to the characteristics of the images, three different error statistics criteria were adopted in this study to evaluate the inversion results, the apparent resistivity results, and the phase results. The three methods were the average Euclidean distance error statistics, the histogram error statistics, and the average peak signal-to-noise ratio (Peak S/N) error statistics [40]. The computation equations are given in Equations (9)–(11).

\[
\text{Err}_{\text{avg}} = \frac{\sum_{i=1}^{W} \sum_{j=1}^{H} |L_{ij} - K_{ij}|}{W \times H}
\]

where \( W \) and \( H \) represent the width and height of the image, respectively; and \( L_{ij} \) and \( K_{ij} \) denote the results of the traditional Occam inversion and the results of the MT-MitNet computation at point \((i,j)\), respectively.

\[
\text{Err}_{\text{hist}} = \begin{cases} 
\frac{1}{P} \sum_{i=1}^{P} \left(1 - \frac{|P_i - R_i|}{\text{Max}(P, R)}\right), & P_i \neq R_i \\
1, & P_i = R_i
\end{cases}
\]

where \( P \) is the length of the image histogram, and \( P_i \) and \( R_i \) are the values of the forward results saved by the traditional Occam inversion and the results calculated by the MT-MitNet at the point \( i \) on the histogram, respectively.

\[
\text{Err}_{\text{Peak}} = 20 \cdot \log_{10} \left( \frac{\text{MAX}_i}{\sqrt{\sum_{i=1}^{W} \sum_{j=1}^{H} \left|L_{ij} - K_{ij}\right|^2 / W \times H}}\right)
\]

where \( \text{MAX}_i \) is the maximum pixel value of the image; the maximum value after normalization of the data in this study was 1, so \( \text{MAX}_i \) was taken as 1. \( W \) and \( H \) represent the
width and height of the image, respectively; \( L_{ij} \) and \( K_{ij} \) denote the results of traditional Occam inversion results and MT-MitNet computation results at point of \((i, j)\), respectively.

The triple anomaly model undergoes nine iterations in the Occam inversion method using the MT-MitNet replacement forward computation, and its inversion results and forward response are shown in Table 1, along with the inversion results and forward response error of the traditional Occam inversion method.

Table 1. Error of inversion results of three-anomalies model.

<table>
<thead>
<tr>
<th></th>
<th>Average Euclidean Distance (%)</th>
<th>Histogram (%)</th>
<th>Peak S/N (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistivity</td>
<td>4.29</td>
<td>22.60</td>
<td>31.55</td>
</tr>
<tr>
<td>Apparent Resistivity</td>
<td>1.60</td>
<td>19.27</td>
<td>33.36</td>
</tr>
<tr>
<td>Phase</td>
<td>2.58</td>
<td>21.99</td>
<td>32.81</td>
</tr>
</tbody>
</table>

From Table 1, it is evident that the computed results for the anomalous body exhibited an accuracy of more than 95% as indicated by the Euclidean distance, which represents high precision. A Peak S/N result above 30 dB signifies a high degree of similarity between any two compared images. Such a degree of similarity is reflected in the inversion and forward results of the model that are highly in agreement. The evident accuracy indicates that the Occam inversion method with deep learning replacing forward calculation can potentially simulate the traditional Occam forward iterative process. Furthermore, MT-MitNet computed forward results with high accuracy, meeting the requirements of inversion accuracy. The MT-MitNet inversion achieved relatively ideal results when dealing with multiple, complex anomalous bodies.

To test the effectiveness of the MT-MitNet method more comprehensively, 1000 groups of data were randomly selected from the 7000 groups of test sets in the forwarding sample dataset. We evaluated these models under the same experimental conditions as previously described. The model forward responses of these 1000 data groups were combined as observed data, and inversion was performed using the traditional Occam inversion method and the method based on deep learning replacement forward computation, separately. Table 2 gives the average forward elapsed time of the inversion computation for the 1000 data groups for both methods, and Table 3 gives the average error of the forward response computation part of the inversion computation for the 1000 test set data using the MT-MitNet method under the three error evaluation approaches.

Table 2. Forward computation of the average schedule.

<table>
<thead>
<tr>
<th></th>
<th>Average Time Taken for Forward Computation (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Occam</td>
<td>2619.1</td>
</tr>
<tr>
<td>MT-MitNet Occam</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Table 3. Test set data results’ error.

<table>
<thead>
<tr>
<th></th>
<th>Average Euclidean Distance (%)</th>
<th>Histogram (%)</th>
<th>Peak S/N (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistivity</td>
<td>4.25</td>
<td>22.64</td>
<td>31.60</td>
</tr>
<tr>
<td>Apparent Resistivity</td>
<td>1.64</td>
<td>19.32</td>
<td>33.34</td>
</tr>
<tr>
<td>Phase</td>
<td>2.43</td>
<td>21.95</td>
<td>32.89</td>
</tr>
</tbody>
</table>

The results of the test set experiments showed that the inversion-oriented deep-learning-accelerated forward computation of the Occam inversion method is highly similar to the traditional Occam inversion method in terms of inversion results and forward responses. The average MT-MitNet forward time is about 1/653 that of the traditional Occam forward average time, and the forward efficiency is greatly improved. The validation of the test data proved the effectiveness of the inversion-oriented deep learning forwarding method.
To further compare the differences between the MT-MitNet replacement forward computation results and the traditional inversion results, the computation results located at 2000 m and 3088 m of the geoelectric model were extracted for comparison and analysis, as shown in Figures 10 and 11.

The results demonstrate that replacing forward computation with inversion-oriented deep learning yields commendable outcomes. MT-MitNet provides computed values for phase and apparent resistivity that are essentially in agreement with those obtained from traditional methods. Moreover, MT-MitNet accurately identifies the model’s subtle differences and responds appropriately, demonstrating its high accuracy. This finding establishes that this approach renders results with high precision in less time.
4. Conclusions

The proposed deep learning network can precisely simulate the magnetotelluric forward modeling and improve the forward speed. This study applied deep learning to the forward response calculation in MT inversion. By using the forward data generated by traditional inversion calculations, we established a crucial sample dataset for training a deep learning network. We proposed a forward network model named MT-MitNet, based on Mit-Transformer, which can globally model and obtain more abundant feature information and execute faster forward simulations with high accuracy. The experimental results demonstrated that MT-MitNet can accurately simulate the traditional forward method by replacing its traditional forward calculation process in inversion. It has precise forward results that fulfill the requirements of inversion accuracy, with an average resistivity accuracy of 98.4% and an average phase accuracy of 97.42%. MT-MitNet also shows a strong generalization ability in the complex multiple anomalous bodies model.

Author Contributions: Conceptualization, F.D. and J.H.; methodology, F.D. and J.H.; software, F.D. and X.W.; validation, F.D.; formal analysis, S.L., X.L. and B.Z.; investigation, X.W., S.L. and S.Y.; resources, X.W., B.Z., X.L. and S.L.; data curation, S.L., S.Y. and B.Z.; writing—original draft, F.D. and J.H.; writing—review and editing, F.D. and J.H.; visualization, X.W.; supervision, F.D.; project administration, B.Z., S.Y., X.L. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Major Project of China (41930112).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
2. Cagniard, L. Basic theory of the magneto-telluric method of geophysical prospecting. Geophysics 1953, 18, 605–635. [CrossRef]
4. de Groot Hedlin, C.; Constable, S. Occam’s inversion to generate smooth, two-dimensional models from magnetotelluric data. Geophysics 1990, 55, 1613–1624. [CrossRef]


30. Kim, J.; Song, G.; Wang, Y. Improved YOLOv4 marine target detection combined with CBAM. *Symmetry* 2021, 13, 623. [CrossRef]


Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.