2D magnetotelluric imaging method based on visionary self-attention mechanism and data science

Yaohua Luo[1,2], Jiachen Li[1], Xuben Wang[2*†], Junjie Zong[1*†] and Haoyu Tang[1]

*Correspondence:
xbwang@cdut.edu.cn;
2023050913@stu.cdut.edu.cn
[1]College of Computer Science and
Cyber Security (Pilot Software
College), Chengdu University of
Technology, No. 1, East Third Road,
Erxianqiao, Chengdu, Chenghua
District 610059, China
[2]Key Laboratory of Earth
Exploration and Information
Techniques (Chengdu University of
Technology), Ministry of Education,
No. 1, East Third Road, Erxianqiao,
Chengdu, Chenghua District
610059, China
[†]Equal contributors

Abstract

2D magnetotelluric (MT) imaging detects underground structures by measuring electromagnetic fields. This study tackles two issues in the field: traditional methods' limitations due to insufficient forward modeling data, and the challenge of multiple solutions in complex scenarios. We introduce an enhanced 2D MT imaging approach with a novel self-attention mechanism, involving: 1. Generating diverse geophysical models and responses to increase data variety and volume. 2. Creating a Swin–Unet-based 2D MT Imaging network with self-attention for better modeling and relation capture, incorporating a MT sample generator using real data to lessen large-scale supervised training dependence, and refining the loss function for optimal validation. This method also includes eliminating MT background response to boost training efficiency and reduce training time. 3. Applying a transverse electric/transverse magnetic method for comprehensive 2D MT data response. Tests show that our method greatly improves 2D MT imaging's accuracy and efficiency, with excellent generalization.

Keywords: 2D MT Imaging; Deep Learning; Self-Attention Mechanism; Swin–Unet

1 Introduction

In the magnetotelluric (MT) method, natural electromagnetic fields are used to investigate the structure of the Earth's electrical conductivity. The electromagnetic (EM) fields at the surface of the Earth behave almost like plane waves, with most of their energy being reflected and a small amount propagating vertically downward into the Earth. The amplitude, phase, and directional relationships between electric (E) and magnetic (H or B) fields on the surface depend on the distribution of electrical conductivity in the subsurface. By use of computer models, field measurement programs can be designed to study regions of interest within the Earth from depths of a few tens of meters to the upper mantle [1]. Practitioners encounter several intricate challenge related to forward modeling for two-dimensional electromagnetic imaging on the Earth's surface [2–4]. Firstly, the subsurface's complexity, characterized by multi-faceted, heterogeneous, and intricate geological formations such as varied rock strata, mineral deposits, and aquifers, presents significant

modeling difficulties. Accurately simulating the electromagnetic response of these structures, particularly at geological interfaces with irregular boundaries and substantial depth variations, is a formidable task.

Secondly, model parameterization and precision are pivotal. Electromagnetic forward modeling necessitates a parameterized depiction of the subsurface, encompassing properties like conductivity. The exactitude of these parameters is vital for faithfully simulating subsurface features and yielding dependable imaging results. Nonetheless, achieving precise parameterization is challenging due to the inherent complexity of the subsurface.

Thirdly, the issue of resolution and depth imaging is salient. The attenuation and scattering of electromagnetic signals during propagation hinder the high-resolution imaging of deep-seated structures. This phenomenon diminishes the resolution of deeper formations, complicating the acquisition of accurate subsurface information.

To address these challenges, various numerical simulation methods are employed in MT forward modeling, including the finite element method, the integral equation method, and the finite difference method. These techniques are well-suited for handling the large computational domains required for simulating electromagnetic behavior in complex subsurface structures. However, the process of solving the large sparse matrices that arise during these simulations is computationally intensive and often requires advanced iterative solvers, such as Krylov subspace methods. Despite their effectiveness, these solvers can sometimes fail to converge, especially when applied to highly complex geological models.

To enhance convergence and efficiency, multigrid methods have been introduced as an auxiliary tool for iterative solvers. Combining these advanced techniques is often necessary to achieve successful forward modeling in MT studies (e.g., using a block Gauss–Seidel (GS) smoothing algorithm to improve the convergence of geometric multigrid (GMG), which assists Krylov subspace iterative solvers). Furthermore, a significant number of theoretical models are needed to support these complex simulations and ensure accurate data interpretation. To counter these challenges, our approach involves the generation of diverse geophysical theoretical models and electromagnetic response samples, thereby augmenting data variety and volume. Electromagnetic inversion entails deriving electrical models from resistivity and phase measurements obtained at the Earth's surface. Magnetotelluric sounding inversion methods can be broadly categorized into linear and nonlinear inversions. Linear inversion methods, including the Gauss–Newton method, conjugate gradient method, and Marquardt method, converge rapidly but tend to depend heavily on the initial model and are prone to becoming trapped in local optima. Nonlinear inversion methods, on the other hand, not only overcome the limitations of linear inversion, but also effectively avoid becoming trapped in local optima, making them a focus of scholarly research. Common nonlinear inversion methods include Particle Swarm Optimization (PSO), Monte Carlo simulation, genetic algorithms, and neural network algorithms. Due to its straightforward principles, independence from initial models, global optimization capabilities, and fast convergence, PSO is applied across a wide variety of fields such as electrical engineering, geophysics, and machine learning. The objective is to create models that not only align with observed data but also approximate the actual subsurface conditions accurately [5–7]. Electromagnetic inversion is fundamentally an optimization problem, where the goal is to minimize an objective function Φ [8], comprising the data function (Φ_d) and the constraint function (Φ_m), both modulated by a regulariza-

tion parameter λ (the Lagrange multiplier) [9].

$$\Phi = \Phi_d + \lambda \Phi_m. \tag{1}$$

Conventional electromagnetic inversion methods often demand substantial computational resources and extended durations for completion [10, 11]. Furthermore, the acquisition of large-scale supervised samples is challenging. To surmount these impediments, pretraining strategies are proposed. By initially training models on an extensive, unlabeled dataset of electromagnetic responses, they can internalize richer and more universal features, enhancing their performance and generalization on actual samples. Subsequent fine-tuning on a smaller, labeled dataset enables swifter convergence and enhanced performance compared to traditional training methodologies. Additionally, networks leveraging self-attention mechanisms, exemplified by the Swin–Unet network based on the Swin transformer, are deemed more efficacious for pretraining tasks than conventional convolutional neural network (CNN) architectures [12]. Therefore, this study proposes the Swin–Unet network as the ideal architecture for the two-dimensional electromagnetic imaging of Earth.

2 Related work

1) Dataset Generation for Theoretical Geophysical Electromagnetic Models: We innovated a method for producing datasets comprising theoretical geophysical electromagnetic models and two-dimensional electromagnetic imaging data. Utilizing the SimPEG framework [13], we systematically generated geophysical theory models, computed two-dimensional electromagnetic resistivity response maps in a parallelized manner, and archived them as training samples.

2) Application of Visual Network Models with Self-Attention Mechanisms: Our exploration into the application of self-attention-based visual network models in two-dimensional electromagnetic imaging is noteworthy. Employing Swin–Unet as the foundational network architecture, we processed real measurement data using the Neo4j graph database. We designed pretraining tasks, enhanced loss functions by informing them using geophysical priors, and trained the network with forward-modeled samples. A comprehensive assessment of the model's performance was conducted to validate its efficacy.

3) Two-Dimensional Electromagnetic Imaging Network for transverse electric/transverse magnetic (TE/TM) Joint Modes: We introduced a specialized two-dimensional electromagnetic imaging network adept at integrating information from both TE and TM modes. This integration significantly enhances the accuracy of imaging deep-seated anomalies.

3 Methodology

In this segment, we elaborate on the Transformer's encoder and decoder, the Swin–Unet network architecture, the refined loss function, the method used to expedite model training via elimination of background electromagnetic responses, and the merits of the TE/TM joint mode.

3.1 Transformer's encoder and decoder

The transformer network, a deep neural network model, adopts an encoder–decoder framework [14]. The encoder transforms the input sequence into a set of representations,

Figure 1 Transformer encoder–decoder structure [15]
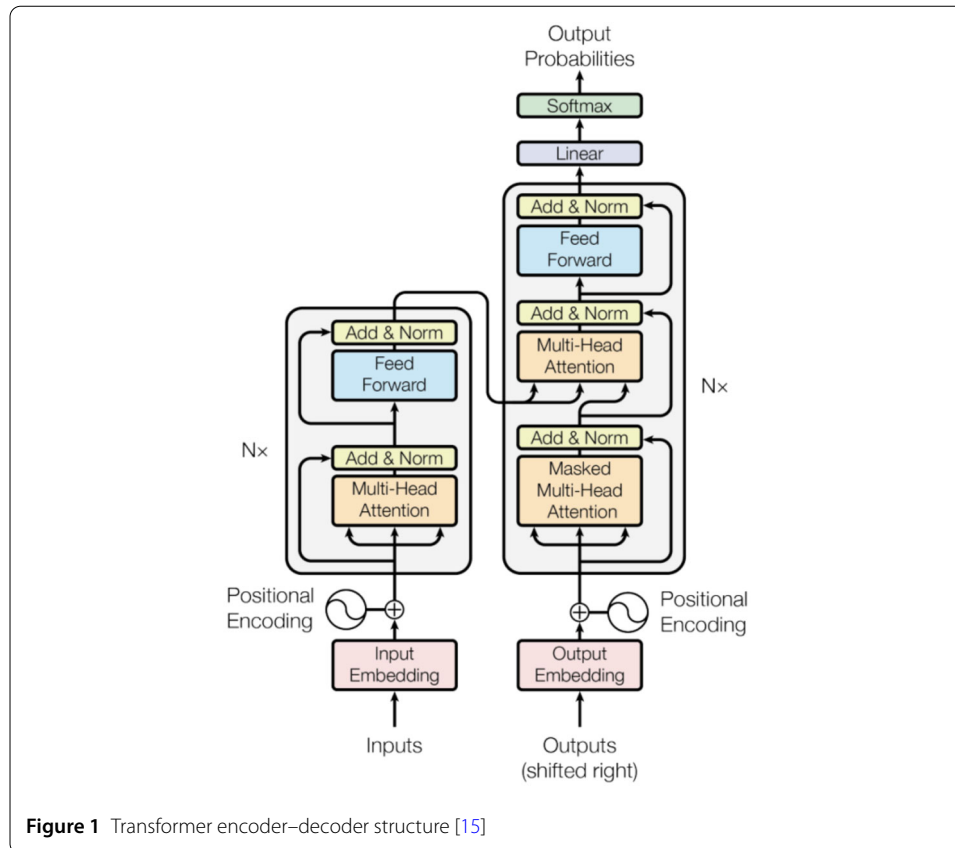
while the decoder leverages these representations to generate the output sequence. In the Transformer network, the encoder consists of multiple identical layers, each comprising a self-attention mechanism layer and a feedforward neural network layer. The self-attention layer evaluates the significance of each position in the input sequence, and the feedforward layer conducts nonlinear transformations on these positional representations. Similarly, the decoder is composed of identical layers, each containing a self-attention mechanism layer and an encoder–decoder attention layer. This latter layer enables the decoder to concentrate on relevant positions in the input sequence during output generation. The encoder–decoder structure of the Transformer is depicted in Fig. 1.

In the described architecture, each layer of the encoder is composed of three sequential modules: a self-attention module, a multi-head attention module, and a fully connected layer module. Additionally, three residual connections are integrated, each bypassing one of these modules. These connections merge with the original output through layer normalization to produce the final output. The decoder, on the other hand, comprises multiple identical layers, each containing a self-attention sub-layer, an encoder–decoder attention sub-layer, and a fully connected feedforward neural network. The input for the decoder includes the target sequence's embedding vector and a weighted sum of the position vectors from the encoder's output, where weights are computed by the encoder–decoder attention sub-layer. The decoder iteratively generates the output sequence, employing the softmax function at each position to convert the output into a probability distribution [16], with the highest probability determining the output for that position.

Figure 2 Swin–Unet network structure [18]
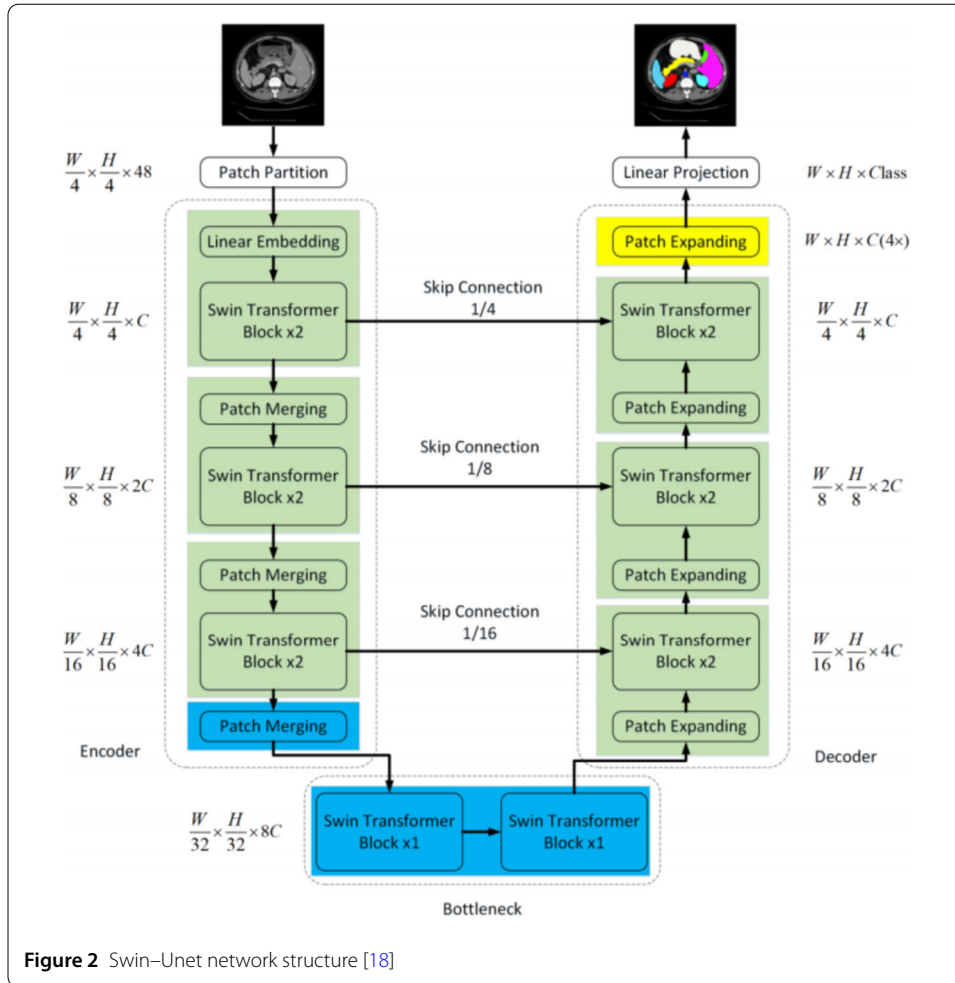
3.2 Network architecture

Swin–Unet, a U-shaped image segmentation network, is built upon the Swin transformer architecture [17] and incorporates the U-shaped structure and skip connections from Unet [18], with Swin transformer blocks forming the encoder's backbone. Additionally, a symmetric decoder block extension, termed the patch-expanding layer, is integrated for image reconstruction. The network's architecture is illustrated in Fig. 2.

The encoder in Swin–Unet mirrors the Swin transformer's setup [18]. The initial input image of dimensions $W \times H \times C$ is transformed into a vector of dimensions $\frac{W}{4} \times \frac{H}{4} \times C$ through patch partitioning and embedding. This vector undergoes feature learning in two successive Swin Transformer blocks. The features' dimensions and resolution remain unaltered within these blocks. Patch-merging layers amalgamate smaller patches into larger ones, concurrently augmenting feature dimensions and achieving multiscale feature fusion. The encoder's structure is characterized by shift-based window self-attention computation, facilitating the learning of feature relationships from local to global scopes.

3.3 Loss function

In the realm of two-dimensional electromagnetic imaging, where model outputs are categorized similarly to image segmentation networks, the choice of loss function is pivotal [19]. The cross-entropy loss function, prevalently used in image segmentation [20], aims to

minimize the discrepancy between predicted and actual outputs, thus effectively addressing misclassifications and multi-class challenges. However, its use is limited by its inability to differentiate the significance of pixels, which is crucial in electromagnetic imaging tasks, particularly when detecting smaller anomalies. The cross-entropy function's computational bias towards the background can lead to small anomalies being overlooked.

Focal Loss is a variant of cross-entropy loss designed to address class imbalance in image segmentation tasks. In scenarios where there is a significant imbalance between foreground and background pixels, standard cross-entropy loss tends to focus more on the majority class (background), leading to suboptimal performance for the minority class (foreground or objects of interest). Focal Loss mitigates this by down-weighting the loss for well-classified examples and focusing more on hard-to-classify examples.

The mathematical expression for Focal Loss is:

$$\text{Focal Loss} = -\alpha \left(1 - p_t\right)^\gamma \log\left(p_t\right),$$

where:

p_t is the model's estimated probability for the correct class.

α is a balancing factor to adjust the importance of positive and negative examples.

γ is the focusing parameter that reduces the loss contribution of well-classified examples.

In Swin–Unet, Focal Loss is especially useful for pixel-wise classification tasks, as it enhances performance in detecting smaller or less frequent classes, ensuring that the model learns features from underrepresented pixels in segmentation tasks.

Dice Loss is a metric-based loss function commonly used for segmentation tasks to measure the overlap between the predicted segmentation and the ground truth. It is particularly effective in handling imbalanced datasets, where the number of background pixels far exceeds the number of foreground pixels.

The Dice coefficient (D) is a measure of similarity between two sets, and the Dice Loss is defined as:

$$\text{Dice Loss} = 1 - \frac{2 \times \left|Y \cap \hat{Y}\right|}{|Y| + \left|\hat{Y}\right|},$$

where:

Y represents the true labels (ground truth mask).

$\hat{Y}$ represents the predicted labels (predicted mask).

$\left|Y \cap \hat{Y}\right|$ denotes the intersection of the true and predicted labels (i.e., the number of correctly classified pixels).

$|Y| + \left|\hat{Y}\right|$ represent the number of pixels in the true labels and predicted labels, respectively.

This formula measures the overlap between the predicted and true segmentations, ensuring that the model accurately captures the details of both foreground and background in segmentation tasks. A Dice Loss close to 0 indicates that the model's predictions closely align with the true segmentation.

In the context of Swin–Unet, using this Dice Loss formula aims to maximize the overlap between the true segmentation and the predicted segmentation, thereby improving

the accuracy of image segmentation, especially in scenarios with an imbalance between foreground and background.

In image segmentation tasks, merging Dice Loss with Focal Loss creates a comprehensive loss function. This synergistic approach balances the impact of different classes during training and elevates the model's performance in complex scenarios.

The Swin–Unet model in this study employs a combination of these loss functions, tailored to the specific challenges of two-dimensional electromagnetic imaging, to ensure nuanced and accurate anomaly detection.

This revised explanation offers a clearer, more technical insight into the network structure and loss functions used, and shows how they align with the sophisticated nature of research in electromagnetic imaging.

The loss functions used in the Swin–Unet model in the original paper are as follows:

$$Loss_{total} = \alpha Loss_{Focal} + (1 - \alpha) \, Loce_{Dice},$$

where α is the predefined weight of the loss function, defaulting to 0.4.

In two-dimensional electromagnetic imaging, anomalies often occupy a smaller spatial area compared to the background in the samples. This disparity leads to a scenario where $Loss_{Focal}$ tends to converge more rapidly than $Loce_{Dice}$. During the middle and later stages of training, the relatively stable $Loss_{Focal}$ disproportionately influences the total loss function $Loss_{total}$, potentially undermining the effectiveness of the combined loss function. To address this imbalance, we propose a loss function with dynamically adjustable weight coefficients, thus enhancing the training process:

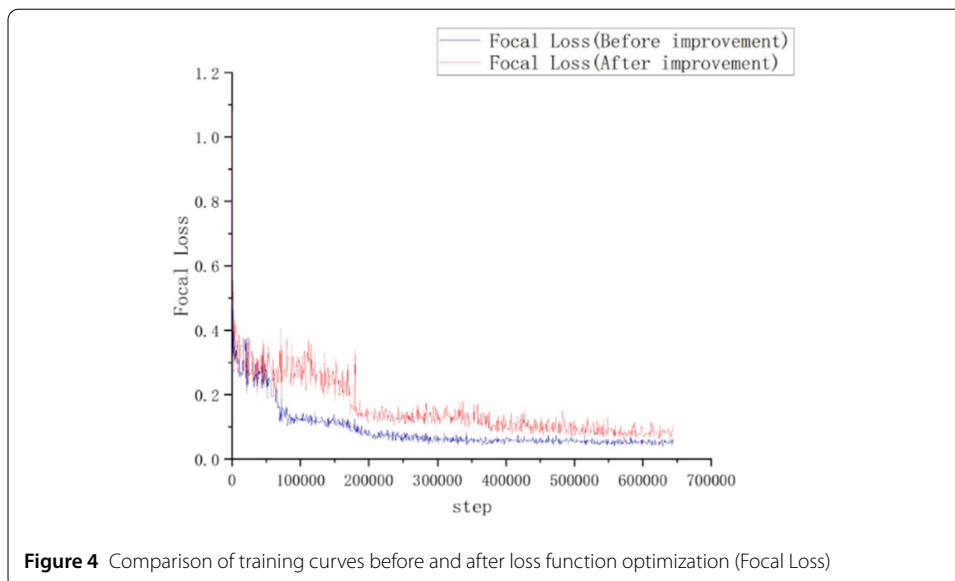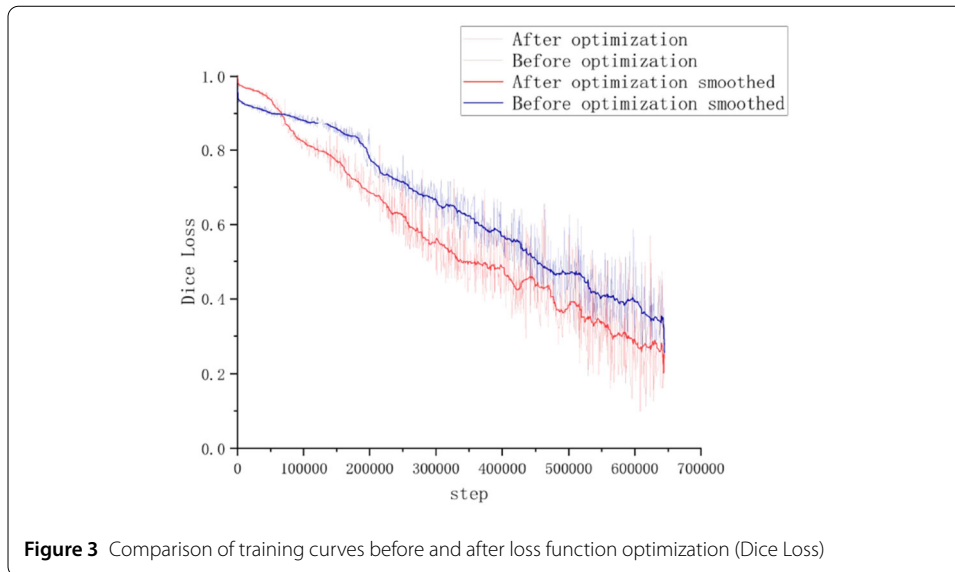$$Loss'_{total} = \alpha' Loss_{Focal} + \left(1 - \alpha'\right) Loce_{Dice}. \tag{2}$$

Here, α' is defined as:

$$\alpha' = \frac{1}{1 + e^{-\left(\text{Loss Focal}^{-\beta}\right)}} \cdot \alpha_{\text{variable}} + \alpha_{\text{static}}. \tag{3}$$

In this formulation, α_{variable} and α_{static} are adjustable hyperparameters of the training process. The parameter β represents a threshold for $Loss_{Focal}$, signaling its convergence when it reaches or drops below this threshold. When the value of $Loss_{Focal}$ exceeds β, the value of α' leans towards α; conversely, as $Loss_{Focal}$ falls below β, α' tends towards 0.

As illustrated in Figs. 3 and 4, the optimized loss function enables Dice Loss to converge more rapidly. Although the training curve for Focal Loss shows greater fluctuation compared to the pre-optimization phase, it still remains within a lower numerical range. Thus, the refinement of the loss function markedly enhances the efficiency and effectiveness of model training.

The model training utilizes the following hyperparameter configuration to optimize performance and meet these challenges: the batch size is set to 48, optimized for 12 classification categories. To prevent overfitting, a dropout rate of 0.1 is used. The model processes patches of size 2×2, combined with an embedding dimension of 96, and employs Swin Transformer layers with depths of [2, 6] and identical settings for decoder depths, enhancing the model's representational capability.

Figure 3 Comparison of training curves before and after loss function optimization (Dice Loss)

Figure 4 Comparison of training curves before and after loss function optimization (Focal Loss)

The adamw optimizer is used, with a learning rate set to 0.00015, a warm-up strategy applied in the first 10 epochs, and a total of 3000 training iterations. The attention mechanism is configured with a varied number of heads at different levels, [3, 6, 12, 21], and a window size of 7×7. The ratio of the dimensionality of the hidden layers in the MLP layer to that in the attention mechanism is set to 4. The careful selection of these hyperparameters aims to improve the model's ability to detect small-scale anomalies in images while maintaining awareness of broader features The specific model depths and hyperparameters are shown in Table 1.

3.4 Model training acceleration method

This paper introduces an innovative approach to mitigate the influence of background electromagnetic responses in forward modeling for two-dimensional electromagnetic imaging. In many instances, a significant portion of the electromagnetic responses gen-

Table 1 Training hyperparameter list

Hyperparameter name	Setting value	Description
batch_size	48	Batch Size
num_classes	12	Number of Classes
drop_rate	0.1	Dropout Probability
patch_size	2	Patch Size
embed_dim	96	Embedding Dimension
depths	[2, 6]	Depth of Each Swin Transformer Layer
decoder_depths	[2, 6]	Depth of Each Decoder
optimizer.name	adamw	Optimizer
learning_rate	0.0005	Learning Rate
warmup_epochs	10	Warmup Epochs
epochs	3000	Training Epochs
num_heads	[3, 6, 12, 21]	Number of Attention Heads
window_size	7	Window Size
MLP_ratio	4	Ratio of MLP Layer Hidden Dimension to Attention Mechanism Dimension

erated corresponds to background noise, which can hinder the model's learning efficacy [22, 23]. Our method involves treating the vectors corresponding to zero-value blocks in input sample regions as null vectors, effectively disregarding any relative positional offsets. This approach sets all self-attention computations associated with these regions to zero, thereby masking the irrelevant background information. Consequently, this strategy allows the model to concentrate more intently on detecting and processing anomaly-related responses during its training phase. Additionally, incorporating samples that partially represent anomaly responses during training can bolster the model's capacity to discern and understand the relationships present in the data, thereby enhancing overall accuracy. The implementation process encompasses the following steps:
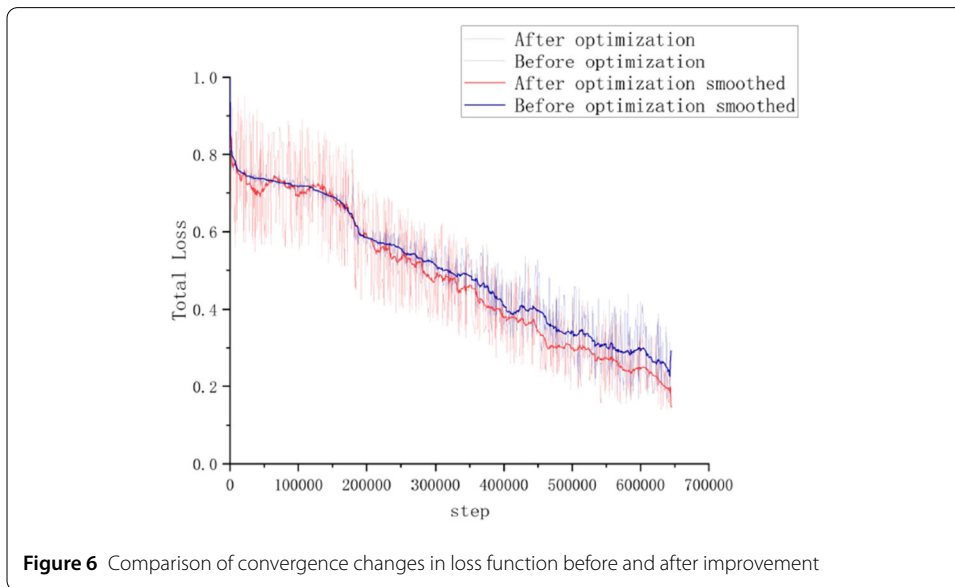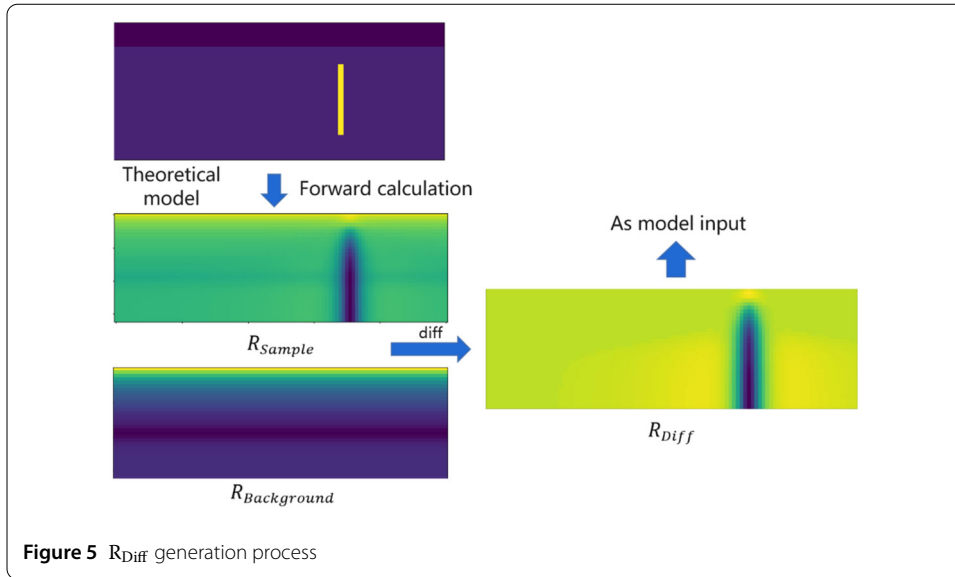
1. Integrate a masking layer atop the relative positional bias matrix in the original network model. This layer sets the relative positional bias for zero vectors to zero.

2. Employ a sample generation program to create background electromagnetic responses for two-dimensional electromagnetic imaging in scenarios devoid of anomalies, denoted as $R_{Background}$.

3. Compute the difference between the input electromagnetic responses R_{Sample} of the training samples and $R_{Background}$ as follows:

$$R_{Diff} = \text{smaller\_zeroing}(R_{Sample} - R_{Background})$$

Here, R_{Diff} encapsulates the anomaly response information within R_{Sample}. The function 'smaller_zeroing' nullifies elements in the matrix whose absolute values are below a set threshold. This step is crucial, as additive noise is inherent in the generation of both background electromagnetic responses and anomaly responses. The exclusion of this noise is imperative to prevent its interference with self-attention calculations. Figure 5 illustrates the process of generating R_{Diff}.

4. Adapt the model's data loader for training to accommodate R_{Diff} data. Incorporate a switching function in the data loader, enabling it to alternate between loading input samples from R_{Sample} and R_{Diff}.

5. Introduce a training hyperparameter λ, signifying the likelihood of the data loader selecting R_{Diff} for input data loading. As the training progresses, λ gradually diminishes to zero. Post-training, the model requires only the original electromagnetic response maps

Figure 5 R_{Diff} generation process

Figure 6 Comparison of convergence changes in loss function before and after improvement

for imaging purposes, obviating the need for input masking responses. Figure 3 depict the loss function's convergence curves during training, both with and without the background electromagnetic response elimination method. The graph demonstrates that this method significantly augments the model's training efficiency.

Next, we will introduce how patch embedding is performed in the network. In the Swin Transformer, to handle large-scale data, the input data needs to be divided into many small patches, each of which is embedded into a vector.

As shown in Fig. 7, the input two-dimensional raw data first passes through a two-dimensional convolution layer, where both the stride and kernel_size of the convolution layer are set to the same size as the hyperparameter patch_size, and the bias is set to 0. The number of output channels of the convolution layer is set to the dimension of the embedding vector. In the figure, embed_dim is exemplified with a value of 1.

Figure 7 The computation process of patch embedding

Figure 8 Convolution computation process
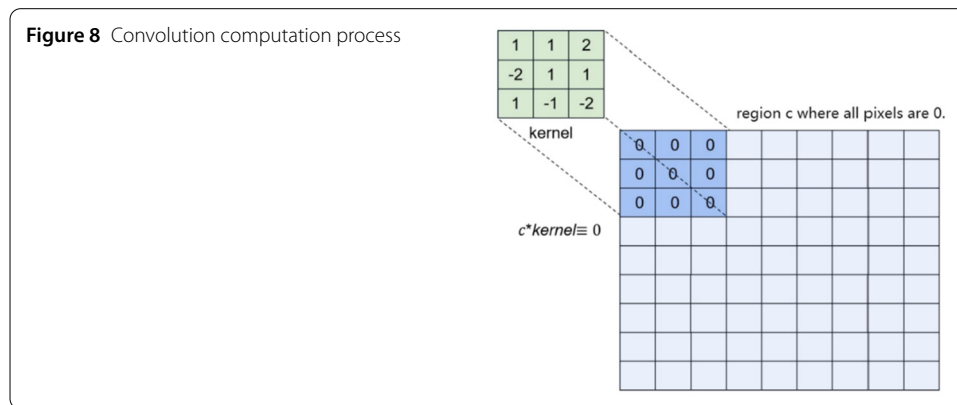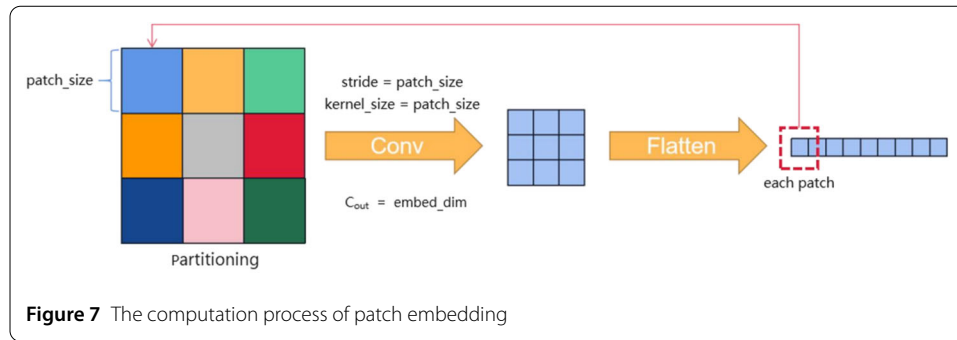
After the convolution, the data is flattened and undergoes a dimensional transformation through a linear embedding operation, ultimately resulting in the vector representation corresponding to each patch.

The computation process of the convolution operation is shown in Fig. 8. As can be seen, when the convolution bias is set to 0, the result of any convolution kernel scanning an area where all values are zero is also 0. The result of the convolution is then organized through a linear embedding operation into a set of vectors corresponding to the patch regions. For patches where all values are zero, the corresponding vector representation is a zero vector.

3.5 Transformer relative position encoding for TE/TM joint mode

In the context of electromagnetic imaging, the TE/TM joint mode amalgamates data from both the TE and TM modes. This combined approach leverages the strengths of each mode to counterbalance their respective limitations, offering a more holistic insight into subsurface structures [24], particularly helpful in discerning electrical variations in vertical and horizontal orientations. The application of this integrative mode facilitates high-resolution electromagnetic imaging across a broader depth spectrum, thereby rendering the interpretation of underground structures more accurately.

To effectively train models utilizing TE and TM joint modes, it is essential to concurrently input dual resistivity response maps corresponding to these modes [21]. This cross-modal input necessitates significant alterations in the network architecture, thus increasing computational and storage demands and prolonging training durations. In response, we approach TE and TM data as a unified modality, directly concatenating the two response maps as network inputs. However, this straightforward concatenation introduces a challenge: the boundary regions of the two maps, which are not initially contiguous,

might be misinterpreted as adjacent by the model post-concatenation [25]. To mitigate this issue, we introduce a Transformer relative position encoding scheme specifically tailored for the TE/TM joint mode.

The Swin transformer employs relative position encoding to adeptly capture the relative positional information within sequence data [26]. This method, in contrast to traditional absolute position encoding, is more suitable for managing sequences of varying lengths, particularly for extensive sequences. In the Swin Transformer, relative position encoding is implemented by inserting relative position embeddings before each self-attention layer. These embeddings comprise two components: one from absolute position encoding and another from relative position encoding. The latter adopts a sinusoidal position encoding pattern, assigning each position a fixed-length vector and encoding positions using sine and cosine functions. This strategy effectively retains the relative distance information among adjacent sequence positions, thus enhancing the capture of local patterns.

In the Swin transformer, the equation for each windowed self-attention computation is formulated as follows:
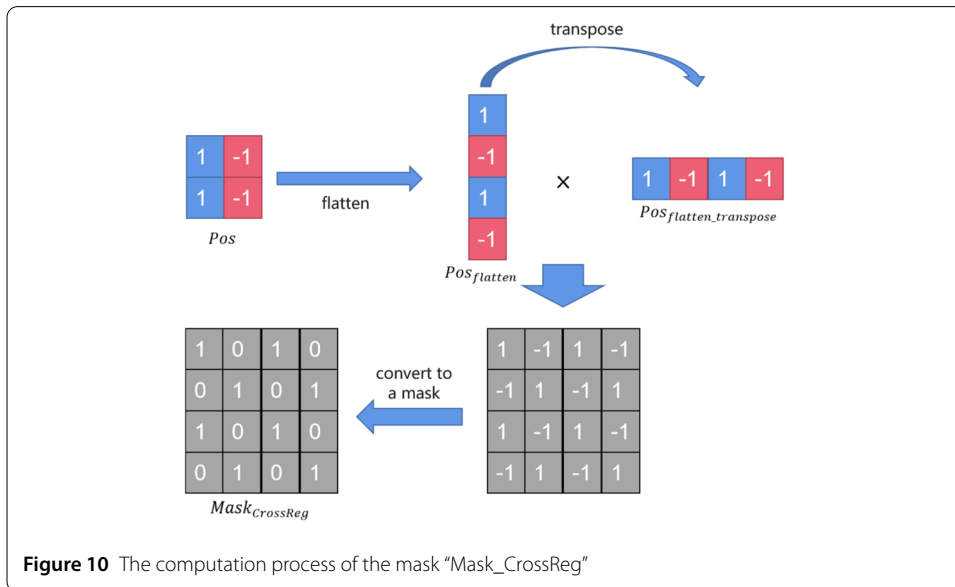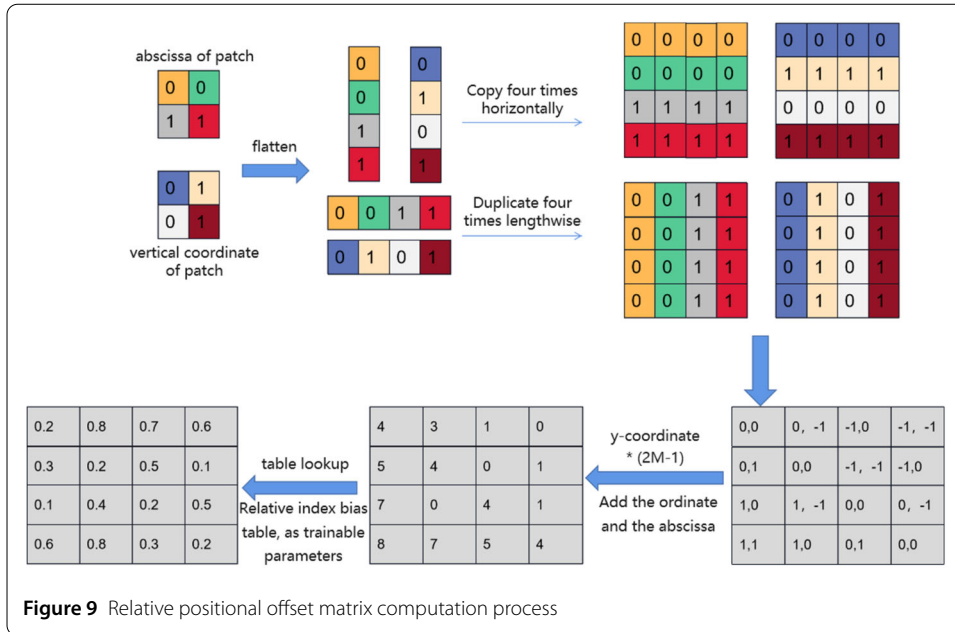
$$Attention\,(Q, K, V) = SoftMax\left(QK^{T}/\sqrt{d_k} + B\right)V. \tag{4}$$

Here, Q, K, V, and d_k are analogous to their counterparts in the standard self-attention computation. Q, K, and V denote matrices composed of query, key, and value vectors, respectively, while d_k represents the dimensions of the key vectors. B symbolizes the relative position of the encoding bias matrix. Notably, the dimensions of Q and K are $w^2 \times N$, where w is the window size and N is the embedding vector dimension. Consequently, the dimensions of QK^T and B are $w^2 \times w^2 \times N$. Given the uniformity of the position encoding offset index matrix across the embedding vector dimensions, it suffices to compute a single two-dimensional offset index matrix, then replicate this matrix along the embedding vector dimension. The offset index matrix, of size $w^2 \times w^2$, represents each patch within the window, with each column denoting the relative position index for each patch, inclusive of itself. Upon generating the index matrix, each index fetches the relative position bias from the relative position bias table (a trainable tensor). Figure 9 elucidates the computation process of the relative position bias matrix for a window size of 2.

To rectify the positional information discrepancies when concurrently inputting TE/TM response data into the network, we propose an advanced strategy for transformer relative position encoding, specifically tailored for TE/TM joint mode:

1) Patch Embedding with Type Markers: During the patch-embedding process for each small block region, we append a TE/TM type marker at the end of the corresponding vector. The TE mode is denoted by a marker value of −1 and the TM mode by 1. Consequently, this alteration modifies the output data format of the patch embedding from the original structure (batch, window_nums, window_size × window_size, N) to (batch, window_nums, window_size × window_size, N + 1).

2) Data Transformation Operations: The network architecture is adapted to segregate the type markers prior to executing transformation operations and then to reintegrate them afterward. The patch-merging process scales up low-resolution feature maps to align with high-resolution maps and concatenates them in channel order, thereby altering the feature map size. To address this, we implement separate patch-merging operations for the original data and the type markers and then merge them, ensuring the integrity of the type of information.

Figure 9 Relative positional offset matrix computation process

Figure 10 The computation process of the mask "Mask_CrossReg"

With these steps, TE/TM type marker information for each patch becomes accessible during windowed self-attention computation. The marker information tensor is denoted as "Pos". We then generate a cross-region mask "Mask_CrossReg" using the marker information tensor:

$$\text{Mask}_{\text{CrossReg}} = \text{Pos}_{\text{flatten}} \times \text{Pos}_{\text{flattentrans}}$$

As illustrated in Fig. 10, the regions where $\text{Mask}_{\text{CrossReg}} \left[\text{Mask}_{\text{CrossReg}} > 0 \right]$ indicate that the corresponding rows and columns belong to small block regions of the same type. Conversely, regions where $\text{Mask}_{\text{CrossReg}} \left[\text{Mask}_{\text{CrossReg}} = 0 \right]$ signify that rows and columns correspond to small block regions of different types.

Figure 11 Subsurface spatial grid distribution
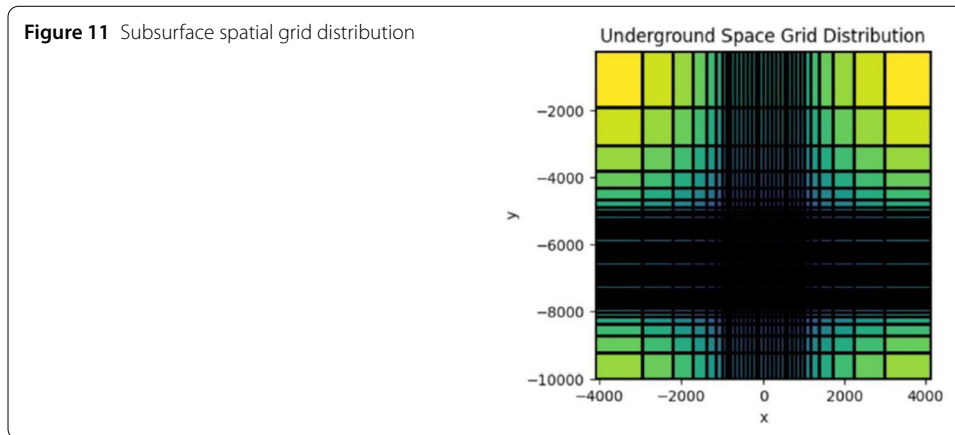
Subsequently, "Mask_CrossReg" is applied to the relative position bias matrix. In the resulting matrix, after masking operation, elements equating to 0 are replaced by a cross-region bias δ, a trainable parameter, typically initialized as 1. This modified relative position encoding maintains the same relative position bias for small blocks of the same type that was present prior to modification. For small blocks of different types, their relative position bias is substituted by the cross-region bias δ.

4 Experimental results and analysis

This section delineates the acquisition of the pre-training dataset, quantitative evaluation, and imaging analysis under the TE mode, including a comparative assessment of the TE mode versus the TE/TM joint mode.

4.1 Dataset

To generate a diverse range of geological models, we developed a script that automates the process. The geophysical simulation framework SimPEG was employed for forward modeling calculations, augmented by the Intel Math Kernel Library [27] to parallelize these computations, thereby acquiring 2D electromagnetic apparent resistivity response maps of the theoretical models. The implementation entailed the following steps:

1. Spatial Grid Construction: An underground spatial grid was established, extending from −4 km to 4 km in the X-direction and from 0 km to 10 km in the Y-direction, as depicted in Fig. 11.

2. Automatic Generation of Theoretical Models: A program was created for the automated generation of theoretical models. This program randomly determines the number, size, and distribution of anomalies within these models. The surrounding rock's resistivity was fixed at 100 $\Omega \cdot m$. For the anomalies, resistivity values were randomly selected from either a low resistivity range (1–100 $\Omega \cdot m$) or a high resistivity range (200–10,000 $\Omega \cdot m$). The aim of this randomization of positions and arrangements of anomalies is to enhance the diversity of the training samples. Figure 12 illustrates the varied theoretical model samples generated. To simplify the model computations, resistivity values were classified into 12 categories based on their magnitude, as summarized in Table 2.

3. Electromagnetic Forward Modeling: The SimPEG framework was utilized to develop the electromagnetic forward modeling program. Measurement points were placed along the x-axis at 100-meter intervals, totaling 81 points. The program also set 55 frequency points within a specified range, as detailed in Table 3.
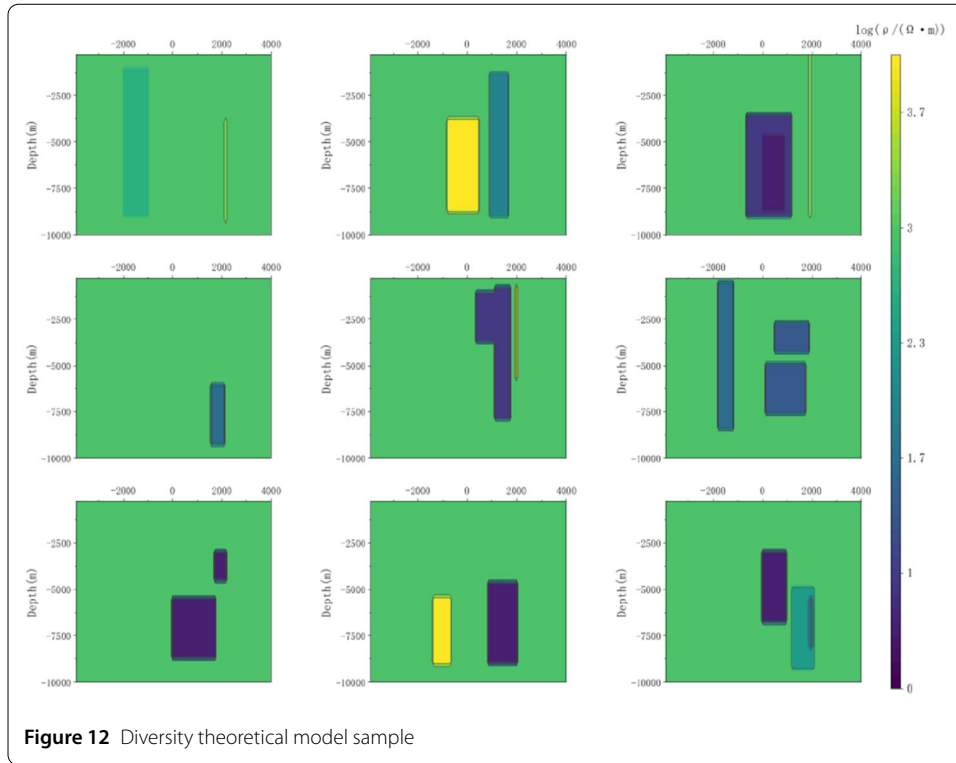
Figure 12 Diversity theoretical model sample

Table 2 Classification of theoretical model anomaly body resistivity

Classification code	resistivity range ($\Omega \cdot$ m)
1	(1, 5]
2	(5, 10]
3	(10, 25]
4	(25, 50]
5	(50, 200]
6	(200, 500]
7	(500, 1000]
8	(1000, 1500]
9	(1500, 2000]
10	(2000, 3000]
11	(3000, 5000]
12	(5000, 10000)

Table 3 Frequency range

Frequency/Hz									
0.034	0.044	0.057	0.686	0.858	0.108	0.137	0.177	0.228	0.274
0.343	0.434	0.549	0.709	0.915	1.098	1.373	1.739	2.197	2.838
3.662	4.394	5.493	6.958	8.789	11.35	14.64	17.57	21.97	27.83
35.15	45.41	58.59	70.31	87.89	111.3	140.6	181.6	234.3	281.25
351.5	445.3	562.5	726.5	937	1218	1593	2062	2625	3281
4125	5156	6562	8343	10,406					

The SimPEG framework facilitated the electromagnetic forward modeling calculations, generating apparent resistivity response maps for both TM and TE modes. These response maps served as input samples, with the data from theoretical models providing labeled samples. Each pair of input and labeled samples was assigned a unique ID to maintain a

Table 4 Performance benchmark testing of 2D electromagnetic forward modeling accelerated with MKL parallelization

Number of enabled CPU cores	Time for 100 calculations (seconds)	Average time per single calculation (seconds)
1 (Without MKL)	incomplete	631
8	4614.5	46.14
16	2861.3	28.61

one-to-one correspondence. The use of the Intel Math Kernel Library significantly expedited the computations [27], reducing the computation time by more than ten times compared to non-parallel processing. The efficiency benchmarks are presented in Table 4. The sample generation script demonstrated high efficiency, producing 60,000 training samples. Using 16-core parallel acceleration on a workstation equipped with an Intel i9-13900k CPU, the generation process spanned approximately 20 days.
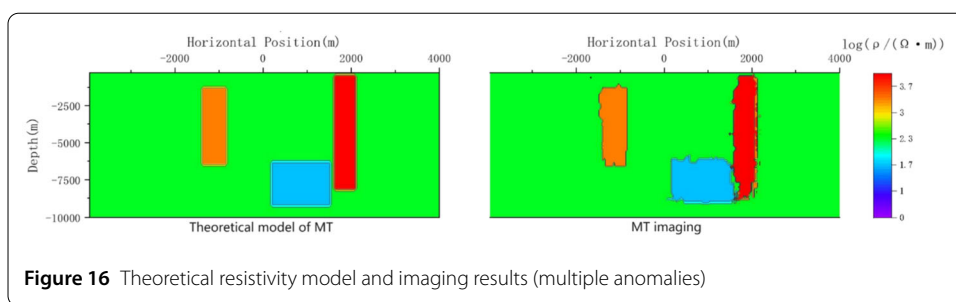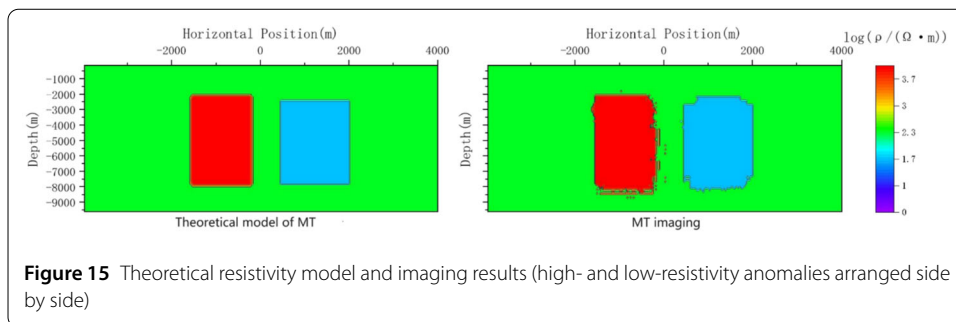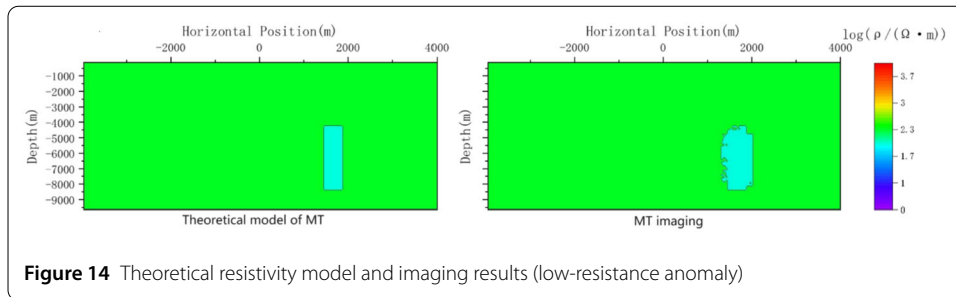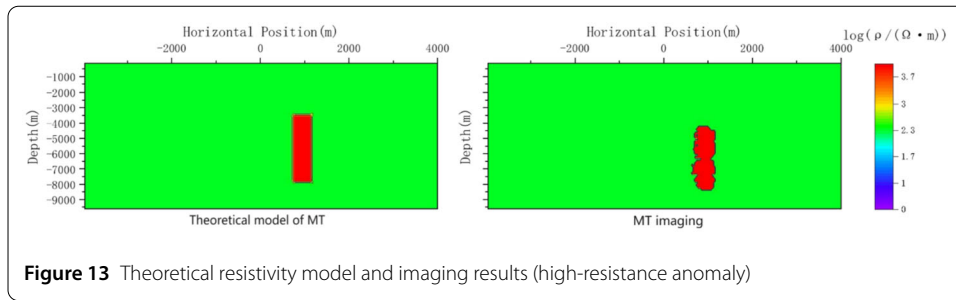
4.2 Real dataset

For pretraining, a substantial dataset of unsupervised real electromagnetic responses from the Earth is necessary. This research utilized data compiled by the Incorporated Research Institutions for Seismology (IRIS) spanning from 2000 to the present, encompassing over 5200 geoelectric measurement points globally. The data can be accessed at IRIS Electromagnetic Transfer Functions (http://ds.iris.edu/spud/emtf). However, the individual data from these measurement points cannot be directly used for pretraining a 2D imaging network. It is essential to connect these points along specific orientations to create measurement lines. The frequency point responses at each measurement point on these lines were arranged based on their spatial positions, forming resistivity response images for each measurement line, which were then used as pretraining samples.

Additionally, the Neo4j graph database was employed to manage the measurement point data. Utilizing Neo4j's graph storage structure, which facilitates adjacency properties, spatial and other relational data were attached to the nodes, allowing for efficient O(1) time complexity when executing spatial-relationship-related queries, thus achieving near-real-time query speeds.
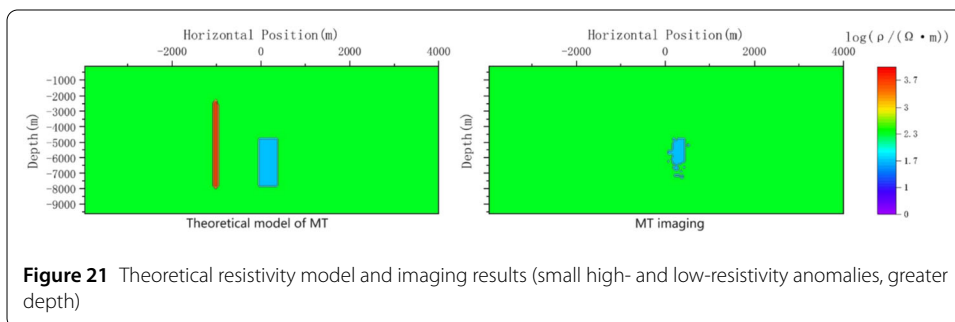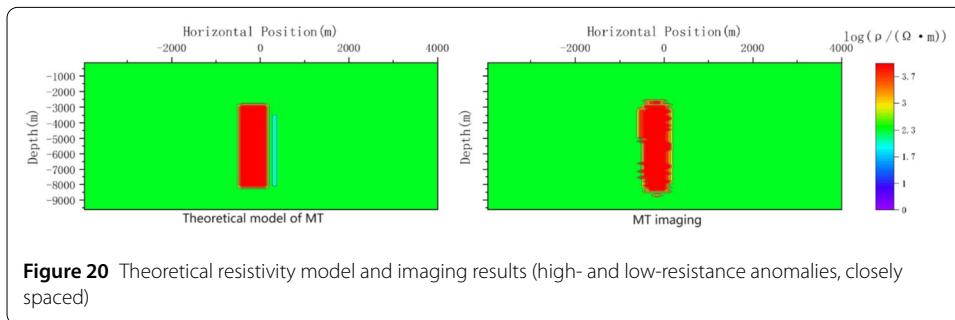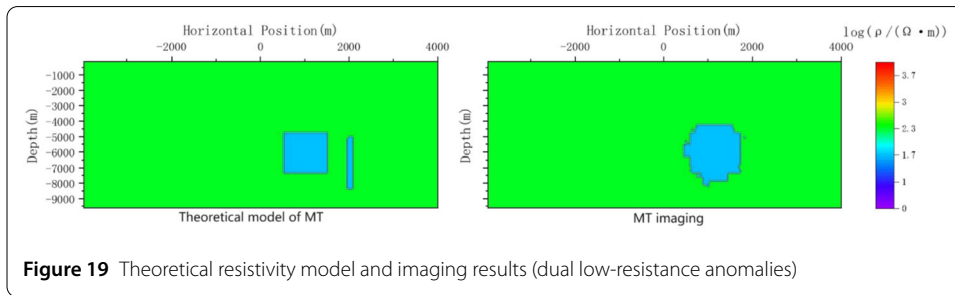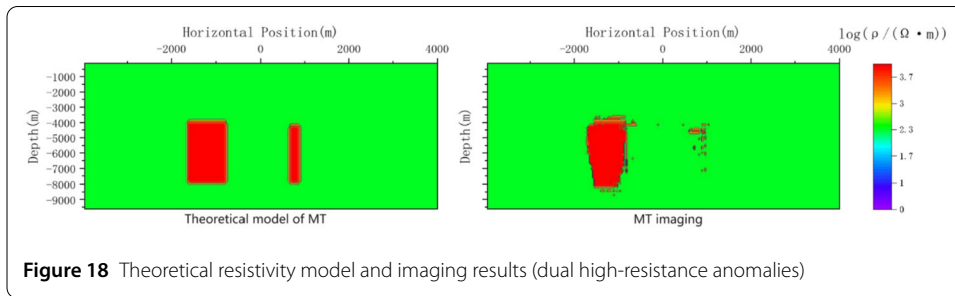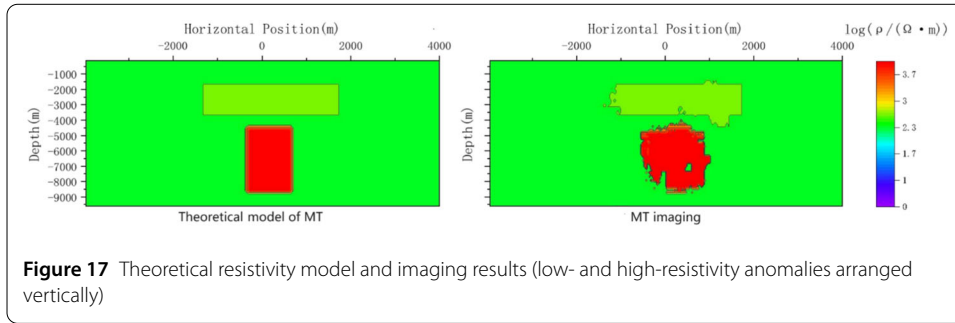
4.3 TE mode imaging analysis

To validate the efficacy of our geoelectric electromagnetic imaging method utilizing the Swin–Unet network, we conducted numerical simulation experiments. These experiments covered nine distinct scenarios involving various shapes and locations of anomalous bodies. The experimental findings reveal that this method can complete an imaging process in approximately 30 milliseconds, significantly outpacing traditional inversion algorithms in terms of computational efficiency. Figures 13 to 21 display the imaging outcomes for these nine scenarios. The first column in each figure illustrates the theoretical resistivity model, while the second column presents the imaging results obtained from deep learning predictions.

The deep learning model demonstrated high accuracy in scenarios with larger, shallow-depth anomalous bodies or simpler geological structures, closely mirroring the theoretical models. However, challenges arose with smaller, deeper anomalous bodies and in situations with multiple closely spaced anomalies. The primary challenges were as follows:

Figure 13 Theoretical resistivity model and imaging results (high-resistance anomaly)

Figure 14 Theoretical resistivity model and imaging results (low-resistance anomaly)

Figure 15 Theoretical resistivity model and imaging results (high- and low-resistivity anomalies arranged side by side)

Figure 16 Theoretical resistivity model and imaging results (multiple anomalies)

1. Depth Sounding Sensitivity: Geoelectric methods are more receptive to low-resistivity anomalies but offer lower resolution for high-resistivity ones. The vertical resolution decreases with depth, leading to unreliable deep learning model imaging for deeper anomalies. When anomalies are proximate, their response data intermix, complicating identification.

2. Mode Limitations: The experiments relied solely on TE mode response data, which focus on shallow subsurface structures. The absence of TM mode data, which are more sensitive to deeper structures, limited the model's accuracy for deeper anomalies.
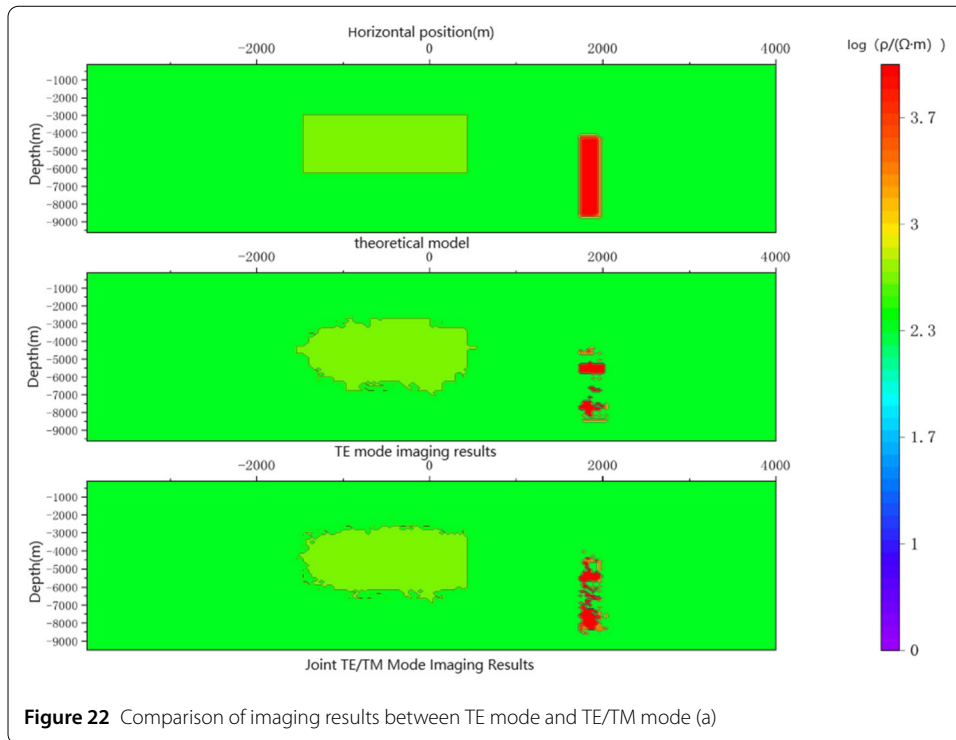
Figure 17 Theoretical resistivity model and imaging results (low- and high-resistivity anomalies arranged vertically)

Figure 18 Theoretical resistivity model and imaging results (dual high-resistance anomalies)

Figure 19 Theoretical resistivity model and imaging results (dual low-resistance anomalies)

Figure 20 Theoretical resistivity model and imaging results (high- and low-resistance anomalies, closely spaced)

Figure 21 Theoretical resistivity model and imaging results (small high- and low-resistivity anomalies, greater depth)

Figure 22 Comparison of imaging results between TE mode and TE/TM mode (a)
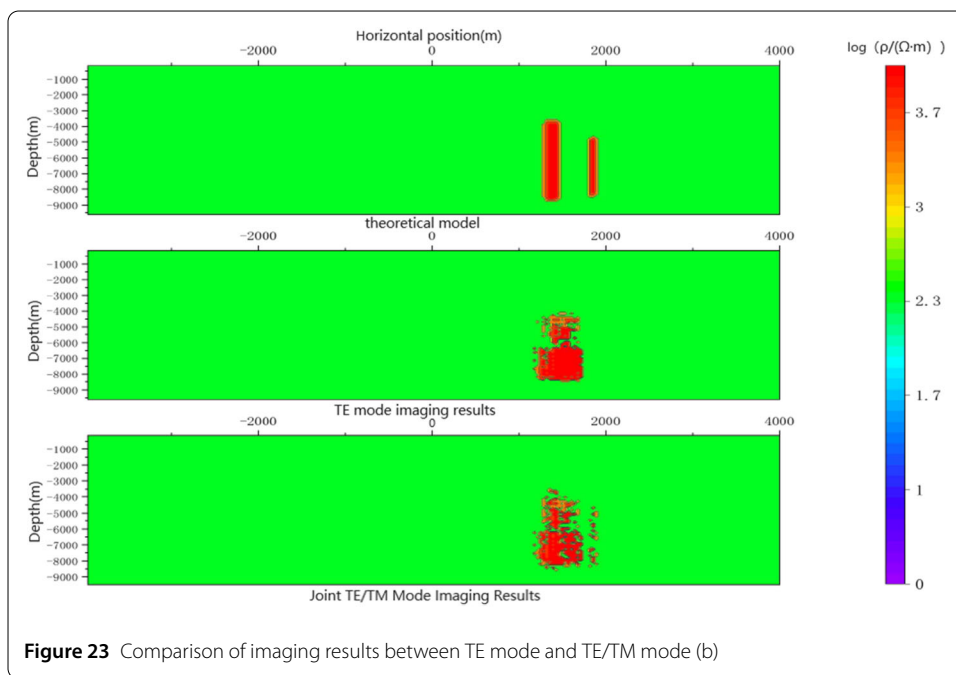
Figure 23 Comparison of imaging results between TE mode and TE/TM mode (b)

4.4 TE/TM Joint Imaging Experimental Results

Imaging results from the TE/TM joint mode model are showcased in Figs. 22, 23, and 24. In these figures, the first column represents the theoretical resistivity models, the second column displays results from the TE mode-trained deep learning network, and the third column presents predictions from the TE/TM joint mode-trained network.
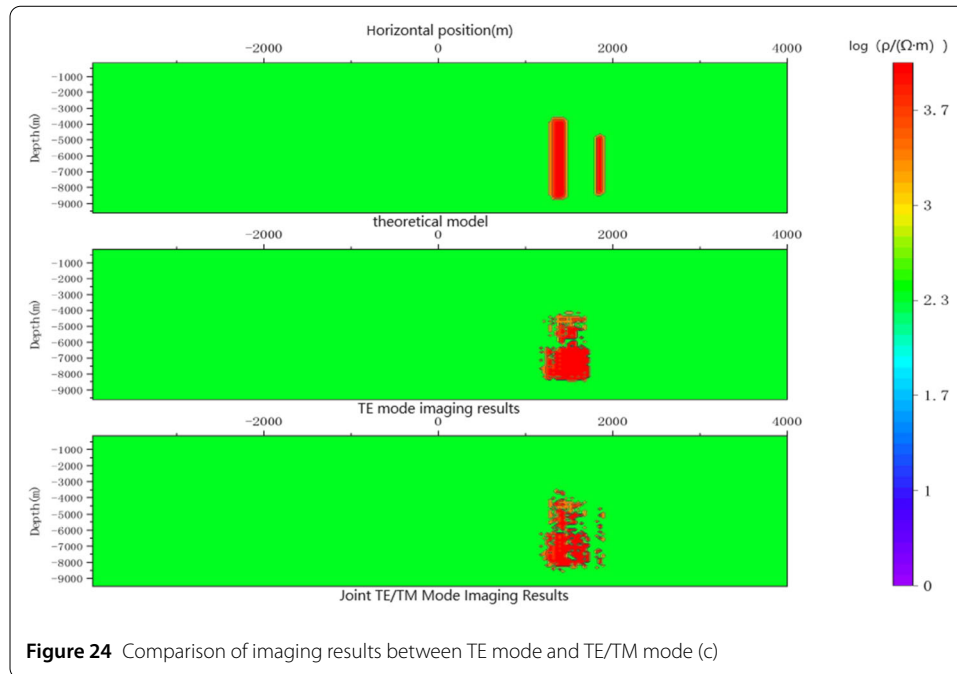
Figure 24 Comparison of imaging results between TE mode and TE/TM mode (c)

Table 5 Comparison of performance of 2D geoelectromagnetic imaging methods based on deep learning (a)

	IOU (small dataset)	DSC (small dataset)	Boundary IOU (small dataset)	IOU (small dataset)	DSC (small dataset)	Boundary IOU (small dataset)
PSPNet	0.924	0.837	0.207	0.956	0.854	0.290
Unet	0.925	0.838	0.209	0.954	0.853	0.287
Swin–Unet	0.919	0.832	0.200	0.971	0.865	0.324

For deep-seated and small anomalies, the neural network trained with TE/TM joint mode data showed marked improvements in imaging accuracy over the model trained using the TE mode only.

4.5 Quantitative evaluation

For quantitative assessment, traditional image segmentation network architectures like the Unet network [28] and the PSPNet network [29] were selected for comparison. We adapted the data loaders of these networks to the specific requirements of geoelectric 2D imaging, setting with the in_channel network hyperparameter as 1 and the n_classes hyperparameter as 12. Both the Unet and PSPNet networks were trained on our constructed geoelectric 2D imaging dataset. To evaluate network performance across different data scales, training was conducted on both a small-scale dataset (6000 pairs) and a large-scale dataset (60,000 pairs), with quantitative evaluations carried out using validation sets. The evaluation results for the TE mode are presented in Table 5.

Given that anomalies in the training dataset generally occupy a smaller area compared to the background (surrounding rock), the Intersection over Union (IOU) metric tends to yield high values. Taking various metrics into account, the results demonstrated that on the smaller dataset, the CNN-based geoelectric 2D imaging method achieved a notably higher accuracy than the Swin–Unet network-based approach. However, the Swin–Unet-based method outperformed the CNN-based method when trained on the larger dataset.

Table 6 Comparison of performance of 2D geoelectromagnetic imaging methods based on deep learning (b).

Method	IOU	DSC	Boundary IOU
Swin–Unet (TE)	0.971	0.865	0.324
Swin–Unet (TE/TM)	0.964	0.897	0.329

The quantitative evaluation results for the TE/TM joint mode are shown in Table 6. It is observed that networks trained with the TE/TM joint mode registered slightly lower IOU scores compared to those trained solely with the TE mode. Yet, they exhibited significantly better Dice similarity coefficient (DSC) and boundary IOU scores than the TE mode-only networks. The IOU and DSC are metrics evaluating the overlap between predicted and actual segmentation results, although they differ in computation. While the IOU is the ratio of the intersection to the union of true positives, false positives, and false negatives, the DSC measures the ratio of twice the intersection to the sum of true positives and the sum of false positives and false negatives. Hence, a high IOU can occur with substantial overlap in segmentation results, but DSC might be lower if there are numerous false positives or negatives. Networks trained with the TE/TM joint mode showed a stronger capability in reducing errors in geoelectric imaging and superior performance in delineating anomaly boundaries.

5 Discussion

Our study extensively explored 2D geoelectric electromagnetic imaging technology, with a focus on deep learning-based approaches. While significant advancements were made, there remain areas for improvement:

1. Data Requirements: The current dataset is somewhat limited for effective pretraining, necessitating collaboration with more organizations to achieve a more expansive collection of geoelectric sounding.

2. Model and Data Realism: Theoretical models differ from actual geological conditions. Optimizing the shapes of anomalies in these models and expanding the number of training samples are thus crucial.

3. Network Scale: Limited hardware capabilities restrict data categorization, impacting the precise identification of anomalies. Expanding network scale through enhanced hardware or supercomputing platforms could improve accuracy.

Future research directions:

1. Data augmentation: Investigate data augmentation techniques to improve the model's robustness against different noise levels and uncertainties.

2. Model simplification: Explore more streamlined model architectures to reduce computational costs and improve inference speed while maintaining imaging quality.

3. Cross-domain applications: Research the application of this method in other geophysical imaging techniques to assess its generality and adaptability.

4. Uncertainty analysis: Conduct uncertainty analysis on model outputs to evaluate the impact of various factors on imaging results, thereby improving the

In summary, while our exploration of 2D geoelectric electromagnetic imaging has made substantial strides, addressing these limitations and pursuing these future research directions will be essential to enhance the robustness, accuracy, and applicability of deep learning approaches in geophysical exploration. This will ultimately contribute to a deeper understanding of subsurface structures and more effective resource management.

6 Conclusions

We have developed an advanced method for two-dimensional magnetotelluric imaging based on a visual self-attention mechanism. Our research aims to address the complex and time-consuming issues present in traditional two-dimensional magnetotelluric inversion imaging methods. These traditional methods, such as finite element, finite difference, and finite volume methods, face large-scale nonlinear inverse problems and the issue of multiple possible solutions, meaning different subsurface conductivity structures could fit the same measured data, making it challenging to find a unique solution.

To overcome these challenges, we have embraced the latest advancements in deep learning technology. We utilized neural network models to process electromagnetic response data directly, generating corresponding subsurface conductivity images. This approach significantly improved computational efficiency and reduced the issue of multiple solutions. However, the accuracy of the model predictions heavily depends on the distribution of the sample data, and discrepancies between theoretical models and actual geological settings might affect performance.

To address these issues, we took the following key steps:

1. We developed a geophysical theoretical model generator to produce various geoelectric theoretical models in batches.

2. We used a forward program based on the finite volume method to generate magnetotelluric responses, with sample generation efficiency being enhanced via parallel acceleration optimization.

3. We collected real magnetotelluric data from open source data sources and constructed a magnetotelluric pre-training sample generator based on a graph database to enhance the model's generalization to real data and reduce its dependence on extensive supervised training data.

4. We built a two-dimensional magnetotelluric imaging network based on the Swin–Unet model. The self-attention mechanism in this network offers superior modeling capabilities and better captures the relationships present in the data, thereby improving the model's accuracy and generalization.

5. We introduced a training acceleration method based on prior geophysical knowledge and improved the existing loss function to effectively enhance the training efficiency of the model.

Additionally, to address the issue of incomplete information on deep anomalous structures in the model when only TE mode response data are available, we applied a TE/TM joint mode magnetotelluric 2D imaging method. The experimental results from this method demonstrate that our approach significantly enhances the accuracy and efficiency of two-dimensional magnetotelluric imaging, along with improved generalization capabilities.

Author contributions
Conceptualization, LY and LJ; methodology, WX and ZJ; software, WX and TH; validation, LY, LJ and ZJ; formal analysis, TH; investigation, ZJ; resources, LJ; data curation, WX; writing—original draft preparation, LY; writing—review and editing, LY; visualization, LJ; supervision, WX; project administration, TH; funding acquisition, ZJ. All authors have read and agreed to the published version of the manuscript.

Data availability

Declarations

Competing interests

The authors declare no conflicts of interest.

References

1. Vozoff, K.: The magnetotelluric method. In: Nabighian, M.N. (ed.) Electromagnetic Methods in Applied Geophysics: Applications/Part B, vol. 2, pp. 641–711. Society of Exploration Geophysicists, Tulsa (1991). https://doi.org/10.1190/1.9781560802686.ch8. ISBN (print): 978-1-56080-022-4, ISBN (online): 978-1-56080-268-6
2. Ren, Z.Y., Chen, C.J., Tang, J.T., et al.: A new integral equation approach for 3D magnetotelluric modeling. Chin. J. Geophys. **60**(11), 4506–4515 (2017). https://doi.org/10.6038/cjg20171134
3. Xue, G.Q., Li, X., Di, Q.Y.: Research progress in TEM forward modeling and inversion calculation. Prog. Geophys. **23**(4), 1165–1172 (2008)
4. Yan, H.T., Zhang, L., Zhang, J.F., Zhang, C., Liu, W., Chu, J.Q.: Application of global weak-form mesh-free methods in two-dimensional magnetotelluric forward. Prog. Geophys. **34**(2), 658–667 (2019). https://doi.org/10.6038/pg2019BB0482
5. Gan, L., Wu, Q.J., Huang, Q.H., Zhang, H.Q., Tang, R.J.: Structure constrained joint inversion of magnetotelluric data and receiver function. Chin. J. Geophys. **65**(11), 4460–4470 (2022). https://doi.org/10.6038/cjg2022Q0319
6. Tan, H.D.: Three-dimensional rapid relaxation inversion for the magnetotelluric method. Chin. J. Geophys. **46**(6), 850–854 (2003)
7. Yan, L.J., Hu, W.B.: Non-linear inversion with the quadratic function approaching method for magnetotelluric data. Chin. J. Geophys. **47**(5), 935–940 (2004)
8. Liu, W., Wang, H., Xi, Z., Zhang, R., Huang, X.: Physics-driven deep learning inversion with application to magnetotelluric. Remote Sens. **14**(13), 3218 (2022). https://doi.org/10.3390/rs14133218
9. Kang, M., Hu, X.Y., Kang, J., et al.: Compared of magnetotelluric 2D inversion methods. Prog. Geophys. (in Chinese) **32**(2), 0476 (2017). https://doi.org/10.6038/pg20170205.
10. Xu, Y.X., Wang, J.Y.: A multiresolution inversion of one-dimensional magnetotelluric data. Chin. J. Geophys. **41**(05), 704–711 (1998)
11. Luo, H.M., Wang, J.Y., Zhu, P.M., Shi, X.M., He, G.M., Chen, A.P., Wei, M.: Quantum genetic algorithm and its application in magnetotelluric data inversion. Chin. J. Geophys. **52**(1), 260–267 (2009)
12. Yang, C.Q., Zhou, Y.T., He, H., et al.: Global context and attention-based deep convolutional neural network for seismic data denoising, Geophys. Prospect. Petroleum **60**, 751–762 (2021)
13. Cockett, R., Kang, S., Heagy, L.J., et al.: SIMPEG: an open source framework for simulation and gradient-based parameter estimation in geophysical applications. Comput. Geosci. **85**, 142–154 (2015)
14. Liu, W., Lu, X.: Research progress of transformer based on computer vision. Comput. Eng. Appl. **58**(6), 1–16 (2022)
15. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. Advances in Neural Information Processing Systems, **30** (2017)
16. Peng, A., Cao, D.P.: Research and application of logging lithology identification based on deep learning. Prog. Geophys. **33**(3), 1029–1034 (2018). https://doi.org/10.6038/pg2018BB0319
17. Xu, G.X., Feng, C., Ma, F.: Review of medical image segmentation based on UNet. J. Front. Comput. Sci. Technol. **17**(8), 1776–1792 (2023)
18. Cao, H., Wang, Y., Chen, J., et al.: Swin–unet: Unet-like pure transformer for medical image segmentation. In: Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III, pp. 205–218. Springer (2023).
19. Yin, X.H., Wang, Y.C., Li, D.Y.: Survey of medical image segmentation technology based on U-Net structure improvement. Ruan Jian Xue Bao/J. Softw. **32**(2), 519–550 (2021). (in Chinese). Available online: http://www.jos.org.cn/1000-9825/6104.htm
20. Zhao, X., Li, H., Su, A., Zhang, H., Liu, J., Gu, G.: Adhesive leukocyte segmentation algorithm based on weighted loss function. J. Jilin Univ. Sci. Ed. **59**(1), 85–91 (2021)
21. Han, Z.Q., Yuan, H.L., He, S.M., Feng, B.: Magnetic source transient electromagnetic measured data processing based on the simulating magnetotelluric 2D inversion technique. Prog. Geophys. **31**(2), 517–524 (2016). https://doi.org/10.6038/pg20160203
22. Xue, G.Q., Li, X., Di, Q.Y.: The progress of TEM in theory and application. Prog. Geophys. **22**(4), 1195–1200 (2007)
23. Wang, S.M., Wang, J.Y.: Application of higher-order statistics in magnetotelluric data processing. Chin. J. Geophys. **47**(5), 928–934 (2004)
24. Xiong, B., Luo, T.Y., Cai, H.Z., Liu, Y.L., Wu, Y.Q., Guo, S.N.: Two-dimensional magnetotelluric inversion of topography. Geophys. Geochem. Explor. (3), 587–593 (2016). https://doi.org/10.11720/wtyht.2016.3.22
25. Xu, L.B., Wei, W.B., Jin, S., et al.: Study of deep electrical structure along a profile from northern ordos block to yinshan orogenic belt. Chin. J. Geophys. **60**(2), 575–584 (2017). https://doi.org/10.6038/cjg20170212
26. Wang, W.L., Wang, T.J., Chen, J.C., You, W.B.: Medical image segmentation method combining multi-scale and multi-head attention. J. Zhejiang Univ. Sci. A **56**(9), 1796–1805 (2022)
27. Kong, Y., Sun, B.Q., Zhang, X.Z., et al.: Application of intel MKL in GNSS data processing with bernese GNSS software. J. Geod. Geodyn. **40**(7), 736–740 (2020)
28. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241. Springer, Berlin (2015)

29. Zhao, H., Shi, J., Qi, X., et al.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.