JGR Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000734

Special Collection:
Advanced machine learning in solid earth geoscience

Key Points:

- Superior performance of Vision Transformer in image-based Electrical Resistivity Tomography (ERT) inversion in comparison to existing models
- Neural network-based colormap calibration of ERT imaging for accurate resistivity mapping and consistent interpretations
- Accuracy, efficiency, resolution, and adaptiveness advantages over traditional numerical inversion methods

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
K. C. Carroll,
kccarr@nmsu.edu

Citation:
Yin, H., Carroll, K. C., Yuan, Y., Jamil, A., Rucker, D. F., Dai, Z., & Soltanian, M. R. (2025). Hybrid vision transformer with convolutional blocks approach for subsurface electrical resistivity tomography inversion. *Journal of Geophysical Research: Machine Learning and Computation*, *2*, e2025JH000734. https://doi.org/10.1029/2025JH000734

Received 11 APR 2025
Accepted 18 OCT 2025

Author Contributions:
Conceptualization: Huichao Yin, Kenneth C. Carroll, Ahsan Jamil, Zhenxue Dai, Mohamad Reza Soltanian

Hybrid Vision Transformer With Convolutional Blocks Approach for Subsurface Electrical Resistivity Tomography Inversion

Huichao Yin[1], Kenneth C. Carroll[1], Yusen Yuan[1], Ahsan Jamil[1], Dale F. Rucker[2], Zhenxue Dai[3,4], and Mohamad Reza Soltanian[5]

[1]Plant & Environmental Sciences Department, New Mexico State University, Las Cruces, NM, USA, [2]HydroGEOPHYSICS, Inc., Tucson, AZ, USA, [3]College of Construction Engineering, Jilin University, Changchun, China, [4]School of Environmental and Municipal Engineering, Qingdao University of Technology, Qingdao, China, [5]Departments of Geosciences and Environmental Engineering, University of Cincinnati, Cincinnati, OH, USA

Abstract Electrical Resistivity Tomography (ERT) is widely used for subsurface imaging. Despite recent advancements in machine learning based ERT inversion using conventional deep learning models such as CNN and UNet, accurately delineating resistivity profiles of natural and artificial subsurface anomalies remains challenging. These challenges stem from multi-scale complex geological heterogeneity, spatial resolution limitations, and the ill-posed nature of ERT inversion. This study proposes a new hybrid image-based ERT inversion method leveraging a Vision Transformer (ViT) model incorporating convolutional blocks to map observed pseudo sections of subsurface apparent resistivity to true resistivity distributions. This novel approach is compared to others including CNN AutoEncoder, UNet, and Latent Diffusion. A synthetic data set generated using forward modeling with varying conductive/resistive heterogeneity zone, facilitates model training using a curriculum strategy to improve generalization across complexity levels. To address the absence of true resistivity data for field observations, a residual CNN model is employed to calibrate resistivity colormap ranges of ERT imaging to ensure consistency in resistivity values represented in both input and output images of the ViT model, maximizing inversion accuracy and result interpretation in applications. The proposed method demonstrated superior performance over frequently used comparative models, as the adopted ViT model excels in inversion accuracy, computational efficiency, adaptability to heterogeneity, and robustness to measurement noise. Inversion results showed significant advantages over traditional Gauss-Newton numerical inversion, which achieved high fidelity in rapidly resolving subsurface anomalies with sharp boundaries and accurately mapped resistivity distributions with high resolution, presenting a promising data-driven approach for large-scale ERT imaging applications.

Plain Language Summary Electrical Resistivity Tomography (ERT) is a widely used geophysical technique for imaging subsurface structures by measuring variations in electrical resistivity. Despite recent progress in inversion techniques using traditional numerical methods and newer, machine-learning models, it remains challenging to accurately characterize complex geological features and artificial anomalies due to limited resolution and the inherently uncertain and spatially variable nature of the subsurface and the model calibration process. This study presents a novel image-based inversion framework that utilizes a Vision Transformer model augmented with convolutional blocks to reconstruct the true subsurface resistivity material-property distributions from ERT measurements. The model is trained using a synthetic data set generated through forward modeling, which enables comparison with the known true material property distribution. A curriculum learning strategy is employed to improve model generalization across different levels of data complexity. To facilitate method applications where true resistivity values are unavailable, a separate residual neural network model is developed to calibrate the color scales of ERT images, ensuring consistency between model input and output interpretations. When benchmarked against the conventional numerical inversion approach, the proposed method shows significant improvements in delineating sharp resistivity contrasts with high resolutions and resolving detailed subsurface features in real-time.

1. Introduction

Electrical Resistivity Tomography (ERT) is a widely utilized geophysical technique for imaging subsurface structures (Binley et al., 2015; Dimech et al., 2022; Ducut et al., 2022; Robinson et al., 2003). In hydrogeological

Data curation: Huichao Yin, Yusen Yuan, Ahsan Jamil, Dale F. Rucker
Funding acquisition: Kenneth C. Carroll
Investigation: Huichao Yin, Ahsan Jamil
Methodology: Huichao Yin, Kenneth C. Carroll, Ahsan Jamil, Zhenxue Dai
Project administration: Mohamad Reza Soltanian
Resources: Kenneth C. Carroll, Dale F. Rucker
Software: Huichao Yin, Kenneth C. Carroll
Validation: Kenneth C. Carroll, Yusen Yuan, Dale F. Rucker
Visualization: Huichao Yin
Writing – original draft: Huichao Yin
Writing – review & editing: Kenneth C. Carroll, Yusen Yuan, Zhenxue Dai, Mohamad Reza Soltanian

studies, ERT is particularly valuable for mapping groundwater resources, identifying aquifers, and localizing groundwater bearing formations (Fan et al., 2023; Tao et al., 2022). Its non-invasive nature (Sakar et al., 2024) and ability to obtain high-fidelity measurements of resistive and conductive subsurface targets, anomalous and heterogeneous features (Bai et al., 2024; Ciampi et al., 2024), has made it broadly applied across various fields including environmental assessments (Bai et al., 2024; Gao et al., 2024; Perrone et al., 2014), water resource management (Liu et al., 2024), and underground mining (Rupesh et al., 2024; Su et al., 2024).

Over the years, numerical simulation methods (Rucker et al., 2017) such as Finite Difference Method (Gao et al., 2020; Thirard et al., 2022), Finite Element Method (Aleardi et al., 2021; Zhao et al., 2024), Integral Equation Method (Dong & Yue, 2025), and Gauss-Newton (GN) Method (Boyle et al., 2018; Rucker et al., 2017) are predominantly relied on for accurately delineating resistivity profiles of subsurface natural target zones such as faults (Ducut et al., 2022; Hung et al., 2024), karst features (Liu et al., 2021), or igneous intrusions (Kasing et al., 2023) in addition to artificial anomalous bodies (Rucker et al., 2012; Pang et al., 2022; B. Wang et al., 2024) such as storage tanks, contamination spills (Gao et al., 2024), and tunnels (Nie et al., 2024). Despite their reliability across diverse geophysical conditions, numerical simulation methods face several challenges. High resistivity can contrasts distorting inversion results, while limited spatial resolution often hinders the accurate detection of fine-scale heterogeneities (Liu et al., 2023), and their inherently ill-posed nature (Jamil et al., 2024; Mudunuru et al., 2022), which allows multiple plausible models, potentially introducing artifacts. Furthermore, numerical simulation methods for ERT inversion often require iterative solutions to large systems of equations (Papadopoulos et al., 2022), making them computationally expensive, especially for high-resolution models (Bai et al., 2024; Shi et al., 2024). The iterative forward modeling and inversion processes can result in longer computation times and substantial memory requirements.

To improve ERT measurement, interpretation, and robustness while also reducing computational complexity, field applications often abstract the geometric and resistive properties of the target, aiming to capture only the essential subsurface properties (Dosoky, 2023; Jamil et al., 2024). Complex geological features can be represented by simplified layered homogeneous structures in two-dimensional (2D) cross-sectional resistivity profiles. Irregularly shaped targets and subsurface anomalies such as caves, water bodies, water tanks, and buried structures can be approximated as ellipsoidal, cylindrical, or rectangular prisms to ease interpretation. However, numerically inverting observational ERT data under these circumstances still suffers from the loss of fine-scale details that could smooth out sharp contrasts, potentially obscuring localized anomalies and introducing non-uniqueness, where multiple resistivity distributions fit the observed data equally well. Additionally, discretized resistivity representations can introduce artificial boundary effects and interpolation errors, which would result in geometric distortions where the assumed model does not align with the true subsurface structure (Rucker et al., 2012). To address such limitations, more advanced approaches have been proposed that integrate complementary data sets. Joint inversion of ERT with ground-penetrating radar (Linde et al., 2006), time-lapse joint inversion with dynamic constraints (Rittgers et al., 2016), and integrated hydrogeophysical–hydrogeological characterization (Abbas et al., 2022) provide improved subsurface resolution and help mitigate the non-uniqueness problem.

Building on these developments, numerous inversion methods have been developed over the years utilizing various data formats including 2D mesh-based approaches (Blanchy et al., 2020; Dai & Samper, 2004; De Peppo et al., 2024) and image-based deep learning techniques (Lan et al., 2023). The 2D inversion represents the subsurface as a structured or unstructured mesh and solves for resistivity distribution along vertical cross-sections (Damavandi et al., 2022) using numerical simulation or machine learning models (Oladeji et al., 2024) such as multilayer perceptron (MLP) (Jamil et al., 2024; Manrique et al., 2023) and K-means (Dick et al., 2024; Doyoro et al., 2025), enabling fault detection and mineral exploration but struggling with strong lateral heterogeneities (Rucker et al., 2017). Recent advances in image-based deep learning approaches utilize convolutional neural networks (CNNs) (Guo et al., 2023), deep neural networks (Liu et al., 2022), ResNet (Yan et al., 2023), UNet (Kong et al., 2023; Liu et al., 2023), physics-informed neural networks (PINNs) (Haruzi & Moreno, 2023), artificial neural network (Chhun et al., 2022), and generative models including generative adversarial network (GAN) (Amiri et al., 2024) and diffusion models (Zhan et al., 2025) to reconstruct resistivity distributions efficiently from observatory ERT data. These offer real-time inversion and noise resilience but require extensive labeled data sets for training.

In comparison to traditional 2D mesh grid inversion, image-based deep learning approaches offer greater flexibility and adaptability across different ERT measurement configurations and spatial scales. While conventional methods require explicit forward modeling tailored to specific electrode arrays such as dipole-dipole, Wenner, and Schlumberger (Dosoky, 2023), deep learning models can be trained across multiple configurations by learning spatial resistivity patterns directly from image data generated (Dong & Zhang, 2024). Additionally, 2D mesh-based inversion relies on fixed discretization, where finer grids increase computational complexity (Gao et al., 2020), whereas deep learning methods enable scale-invariant resistivity reconstructions without predefined grid constraints. This adaptability allows seamless application across surveys with varying electrode spacing, penetration depths, and geological settings, making deep learning approaches ideal for large-scale ERT imaging. Deep learning models, once trained, can provide near-instantaneous predictions by leveraging pre-learned patterns, significantly reducing computational costs by directly mapping measurements to subsurface resistivity distributions.

In recent studies, CNN-based neural networks such as ResNet (Huang et al., 2024) and UNet (Guo et al., 2023) are amongst the most frequently used models for image-based ERT inversions due to their efficiency and spatial feature extraction capabilities. However, the CNN-based models exhibit limitations in capturing long-range dependencies, handling scale variability, and managing sparse or noisy data in ERT inversion (Mauricio et al., 2023). Relationships between distant regions in the input data, such as spatial correlations between resistivity anomalies that are far apart in the apparent resistivity pseudo-section are challenging for CNNs due to their reliance on local convolutional filters, which restricts their ability to model global resistivity variations, requiring deeper architectures or larger kernels to approximate long-range correlations (Azad et al., 2021; Zu et al., 2024). Additionally, CNNs operate on fixed-resolution inputs, making them less adaptable to varying spatial scales and more sensitive to missing or noisy data, as their spatial convolution mechanisms struggle to interpolate incomplete resistivity measurements (Yang et al., 2024).

While the more advanced Vision Transformer and diffusion models have seen frequent successes in image-to-image prediction tasks such as image reconstruction and super-resolution (Roy et al., 2023; Saharia et al., 2023; Zhang et al., 2024), their applications in image-based ERT inversion are still nonexistent. They have the potential to offer significant advantages over traditional models like CNN. ViTs leverage self-attention mechanisms to model global resistivity patterns more effectively (Xiao et al., 2023). Their patch-based processing also allows flexible input sizes, enhancing adaptability across different spatial scales and survey configurations (Dosovitskiy et al., 2020; Jiang et al., 2025; Raghu et al., 2021). Diffusion models further improve inversion robustness through iterative denoising, making them resilient to measurement noise and incomplete data (Taufik et al., 2024; Xia et al., 2024). Additionally, both ViTs and diffusion models have generalization capabilities across various electrode configurations without requiring extensive retraining. Both of them offer promising performance in reconstructing highly complex resistivity structures with superior resolution, whereas ViT excels in computation efficiency (Trappolini et al., 2024; F. Wang et al., 2024), making it suitable for large-scale and high-fidelity ERT applications.

In this study, Vision Transformer (ViT) is employed incorporating convolutional blocks (Yang et al., 2024; Zu et al., 2024) to conduct accurate mapping between apparent resistivity observed in ERT data and true subsurface resistivity distributions in 2D sectional resistivity profiles that consist of multiple anomalous bodies of contrasting resistivity values to their surrounding environments with sharp boundaries. Inversion results from CNN-AutoEncoder, UNet, Latent Diffusion Network (LDM) and Res2DINV (Dosoky, 2023; Jamil et al., 2024; Rupesh et al., 2024) using the industry benchmark GN model (Rucker et al., 2017) are compared for validation. The proposed method aims to conduct effective and reliable real-time ERT inversion in complex domains where heterogeneity is mainly brought by or abstracted as the subsurface existence of one or more anomalous target zones of rectangular or similar shapes, which are frequently found in both natural geological formations and artificial subsurface structures. By improving inversion accuracy, efficiency, preserving sharp boundaries, and utilizing neural network models for resistivity colormap calibrations for field applications, the proposed approach seeks to address the limitations of traditional numerical simulation and existing machine learning methods in resolving fine-scale subsurface characterization in such domains.

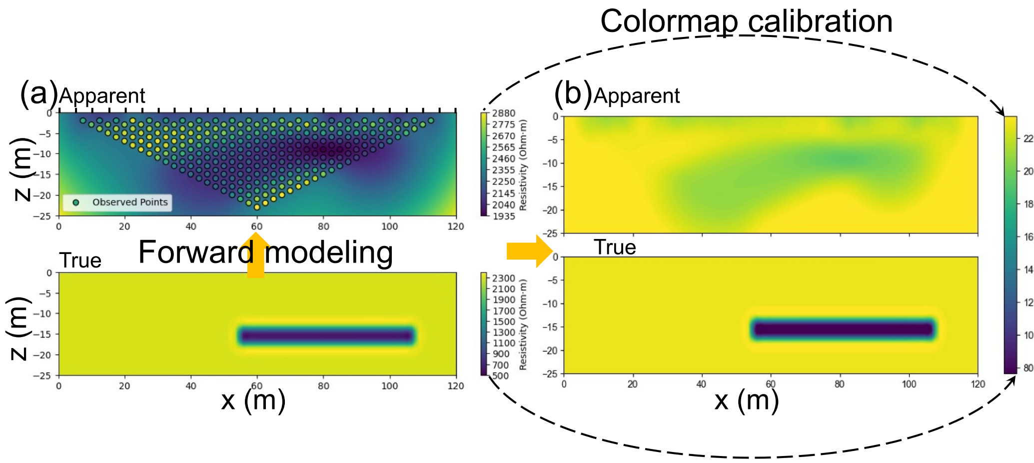

Figure 1. Conceptual diagram of image data generation using (a) forward modeling and (b) its colormap calibration. The generated pseudo section image and the true resistivity profile image are respectively on top and bottom of (a), the x and z axis represent horizontal distance and depth below subsurface, respectively, and darker areas in these images represent the anomalous body with contrasting resistivity values to the background area. The images are then displayed in the shared colormap shown in (b) where its range is designated as the maximal and minimum resistivity values of both the pseudo section and true distribution. This enables effective model training on interpretating the resistivity value differences between the true and apparent resistivity profile images, it also ensures the consistency in resistivity values represented amongst input and output image of the inversion model, so that the inversion result image can be interpreted directly using the calibrated colormap.

2. Data Set Description

The data set for model training in this study was generated using ResIPy (Blanchy et al., 2020), an ER modeling tool employing the finite element method with Neumann boundary conditions to simulate apparent resistivity data. This forward modeling approach created a controlled data set where the true resistivity distributions were known, enabling direct evaluation of inversion performance (Dosoky, 2023; Jamil et al., 2024). The computational domain measured 120 × 25 m with 1–5 embedded anomalies with randomly varied location, size, and resistivity contrast against the surrounding matrix (Figure 1). This assumption was made originally as an abstraction of rectangular or similar-shaped anomalies, providing a simplified yet systematic representation of subsurface targets for controlled training. To ensure control over the model training process, the simulation does not contain scenarios where multiple targets are overlapped within a single domain.

As shown in Figure 1a, a surface array of 25 electrodes spaced at 5 m intervals was used, and the commonly used Dipole-Dipole electrode configurations were simulated according to the true resistivity profile (Bai et al., 2024; Pang et al., 2022; Rucker et al., 2017). Resistivity values are calculated for each converted coordinate in the pseudo section. All contours are then created by using the radial basis function interpolation with multiquadric kernel to generate images of 744 × 154 in width and height in pixels, which is the same ratio of the actual domain (Cui et al., 2021). The interpolation kernel is defined as:

$$\phi(r) = \sqrt{r^2 + c^2} \tag{1}$$

where r is the Euclidean distance between points and c is a shape parameter. This algorithm is adopted in this study for ERT mesh-grid interpolation, as it provides a smooth and globally consistent resistivity profile, effectively handling irregularly spaced data while preserving both local variations and overall structural continuity (Cui et al., 2021). The stability and smooth gradients make it ideal for reconstructing complex resistivity distributions essential for image-based inversion models.

The colormap shared between the true and apparent resistivity profile images are calibrated using the range of the minimum and maximal resistivity value in both profiles (Figure 1b), which is defined as:

$$\rho_{\min} = \mathrm{MIN}\left(\rho_{T_{\min}}, \rho_{O_{\min}}\right) \tag{2}$$

$$\rho_{\max} = \mathrm{MAX}\left(\rho_{T_{\max}}, \rho_{O_{\max}}\right) \tag{3}$$

where $\rho_{T\min}$, $\rho_{T\min}$ and $\rho_{T\max}$, $\rho_{T\max}$ are respectively the minimal and maximal resistivity values for true and observation resistivity profile, $\rho_{\min}$ and $\rho_{\max}$ are respectively minimal and maximal resistivity values that the calibrated resistivity colormap can represent using its darkest or lightest color, both true and apparent resistivity profiles are then plotted with this colormap. This is optimal for augmenting their image feature difference while preserving features for model interpretation. A total of 100,000 synthetic resistivity profile samples were created (10,000 for each target number that ranged from 1 to 5), ensuring an even distribution of conductive and resistive targets with a minimum target-to-matrix resistivity contrast of 2:1 to maintain sufficient heterogeneity. To enhance training stability across different ERT setups, Z-score normalization was applied to all input features, standardizing the data set with a mean of zero and unit variance. This setup, while abstracted, can be employed as representatives of subsurface targets such as buried structures, waste deposits, or water bearing geological formations within a small scale.

The employed forward modeling enables the generation of large-scale and diverse data sets, ensuring deep learning models are trained on a wide range of resistivity distributions and survey conditions for data driven ERT inversion, enhancing generalization to real-world scenarios. Additionally, synthetic data sets can be tailored to specific geological settings and ERT measurement configurations, improving specialization for targeted applications. The generated data set also facilitates curriculum training, allowing models to be trained in a controllable manner on data complexities by gradually increasing the target number for optimizing model performance while minimizing data acquisition costs.

Furthermore, the synthetic data set also allows numerical inversion with the commonly used GN model to validate the method's effectiveness. The GN numerical inversion model provides a mathematically rigorous benchmark based on iterative least-squares optimization, ensuring a reliable reference for evaluating deep learning-based inversion. It generates smooth and physically consistent resistivity profiles, facilitating accurate performance assessment (Boyle et al., 2018). Unlike deep learning models, GN inversion offers interpretable results with sensitivity and uncertainty analysis, allowing for detailed error evaluation. In addition, both methods rely on gradient-based optimization, making GN a logical baseline for comparing how well deep learning models approximate physics-based inversion.

3. Methodology

The proposed method follows the workflow presented in Figure 2, which shows that the ViT model incorporating spatial convolutional blocks is first trained on the synthetic data set prepared using forward modeling and ResIPy, which provides a flexible finite-element simulation framework to generate controlled resistivity scenarios with known ground truth, thereby enabling robust mapping between the apparent and true resistivity profiles. The model was trained starting from simple scenarios where only 1 target body is included, and the number of targets gradually increased to 5 for model training following a curriculum training strategy, eventually allowing the trained model to autonomously determine the complexity of the prediction task in any applications. A selection of deep learning models including residual CNN AutoEncoder, UNet, and LDM was adopted for identical tasks for performance comparison. Since true data in field applications are absent, the image colormap cannot be obtained by directly taking the maximal and minimum value in the true and apparent resistivity profiles, making accurate resistivity colormap predictions essential for calibrating ViT model's input and output image. A CNN model incorporating residual blocks is therefore employed to predict the colormap ranges of apparent resistivity measurements by training on the synthetic data set, taking optimal ranges calibrated on true and apparent resistivity profiles as ground truth (Figure 1b) for its output. With the predicted colormap range, the field observatory resistivity profile can be plotted accordingly, and the predicted results obtained from the ViT model were eventually able to be interpreted as true resistivity distributions.

Finally, the results are examined against those from GN numerical simulation method, using Res2DINV, a widely adopted tool for practical resistivity inversion, to benchmark our approach against a standard reference, and to validate accuracy and reliability for real-world ERT inversion applications.

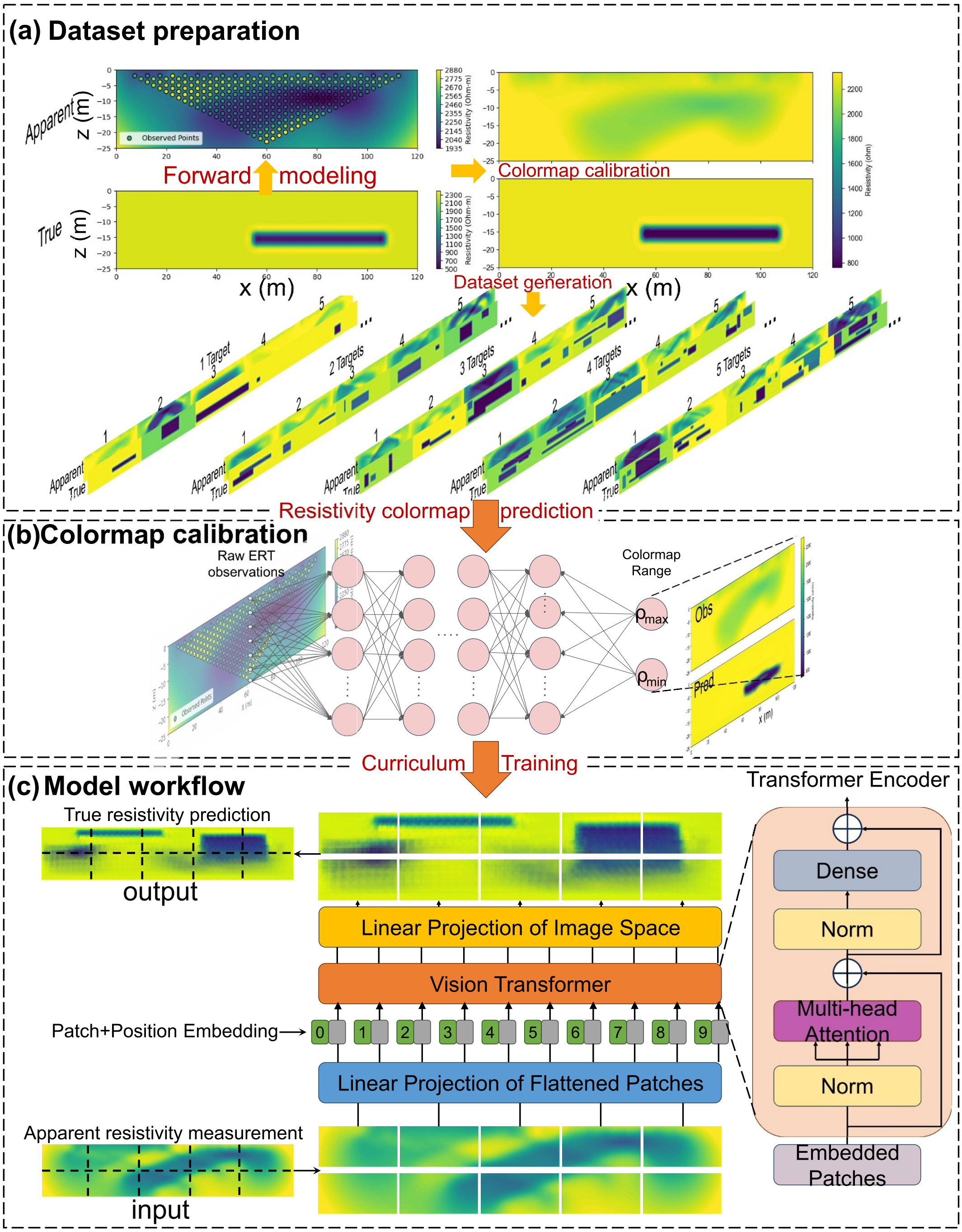

Figure 2. Schematic diagram of proposed method. Section (a) illustrates how image data set containing apparent and true resistivity image pairs are generated across 1 to 5 anomalous targets using forward modeling. The data set is used for curriculum training of the ViT model shown in (c). The inversion result image can be mapped to resistivity values with the colormap obtained from the true and apparent resistivity profiles in the training set, whereas in field applications. It takes another CNN-based model depicted in (b) to predict such resistivity colormap ranges to be shared by input and output image so that the prediction can be performed accurately, and the result can be interpreted directly as in the training set.

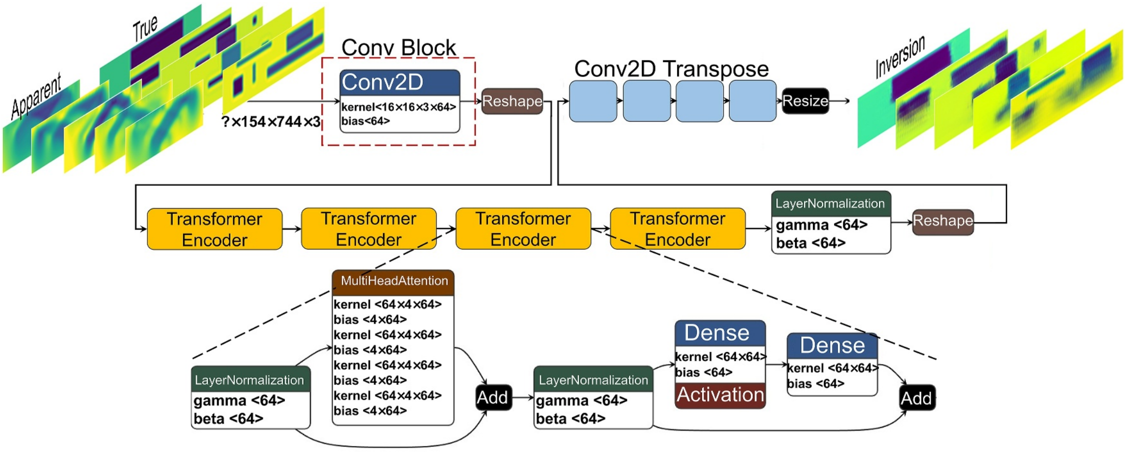

Figure 3. ViT model implemented architecture. Apparent resistivity profiles images are generated using forward modeling as the input of the model with corresponding true resistivity profile images serving as the ground truth. Model output is the predicted true resistivity profile image of a given ERT observation.

3.1. Image-Based Inversion Models and Training Strategies

3.1.1. Vision Transformer

The Vision Transformer (ViT) is a deep learning architecture that applies self-attention mechanisms to image processing tasks, replacing traditional convolutional layers with patch-based attention to model long-range dependencies to enhance spatial consistency and feature generalization (Harsuko et al., 2025). The encoder splits the input image I of size $H \times W \times C$ (height, width, channels) (154, 744, 3 in this study) into fixed-size patches of $P \times P$ (16 × 16 in this study), embeds them into a lower-dimensional latent space using a convolution operation:

$$Z = Wp * I \tag{4}$$

where Wp is a learnable weight matrix applied via a convolutional layer (Conv2D) with a kernel size of $P \times P$ and stride P. To retain spatial information, a learnable 2D positional encoding is applied. The positional encoding matrix is generated as:

$$E_{\text{pos}} = \text{Dense}\left(\left[p_x, p_y\right]\right) \tag{5}$$

where p_x, p_y are the normalized grid coordinates of the patches. The output learnable positional encoding $E\text{pos}$ is added to the patch embeddings as: $Z' = Z + E\text{pos}$.

A stack of transformer blocks applies multi-head self-attention defined as:

$$\text{Attention}\,(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where $Q = Z'W_Q$, $K = Z'W_K$, and $V = Z'W_V$ are the query, key, and value matrices, d_k is the dimension of keys. Each attention output is passed through a position-wise feed-forward network (FFN) allowing the model to learn spatially global resistivity variations:

$$\text{FFN}(x) = \sigma(W_1 x + b_1)\, W_2 + b_2 \tag{7}$$

where W_1,W_2 and b_1,b_2 are learnable weight matrices and bias terms. σ is a ReLU activation function. After transformer processing, the resistivity profile is reconstructed by getting the patches passing through transposed convolution layers (Conv2D Transpose), progressively upsampling feature maps to restore spatial resolution (Figure 3).

For ERT inversion, ViT is particularly suitable due to its ability to capture complex resistivity distributions across large spatial domains without the spatial bias of convolutional filters (Jiang et al., 2025). However, incorporating convolutional blocks within ViT (Figure 3) enhances its suitability by preserving local resistivity structures while leveraging the global attention mechanism to improve resistivity mapping (Zu et al., 2024). This hybrid approach combines CNN's ability to extract fine-scale spatial features with ViT's long-range dependency modeling, making it highly effective for image-based ERT inversion across multiple measurement configurations and varying subsurface scales.

3.1.2. Comparative Models

3.1.2.1. Residual CNN AutoEncoder

The model is designed using an AutoEncoder architecture (Ma et al., 2017) for image-based ERT inversion incorporating multi-scale feature extraction, residual learning, and adaptive upsampling for high-fidelity resistivity reconstruction. The encoder consists of stacked convolution (Conv2D), batch normalization (BatchNormalization), and max pooling (MaxPooling) layers, progressively reducing spatial dimensions while extracting hierarchical resistivity features. A residual block in the bottleneck mitigates vanishing gradients and enhances feature retention (He et al., 2016). The decoder employs convolution transpose (Conv2DTranspose) layers for spatial recovery, with edge padding ensuring size consistency (Figure S1 in Supporting Information S1).

3.1.2.2. UNet

The U-Net-based model adopted here leverages a symmetric encoder-decoder structure with skip connections to preserve fine-scale resistivity features for ERT image inversions (Hiskiawan et al., 2023; Kong et al., 2023). The contracting path extracts multi-scale spatial features through successive Conv2D and MaxPooling layers, progressively reducing resolution while increasing feature depth. The bottleneck layer further refines high-level representations before upsampling. The expanding path uses Conv2DTranspose layers with skip connections, where feature maps from the encoder are concatenated with upsampled representations, enhancing structural preservation. Padding operations are also used across layers to ensure spatial consistency (Figure S2 in Supporting Information S1).

3.1.2.3. Latent Diffusion Model

The LDM is employed integrating spatial attention and latent space representations to enhance resistivity profile reconstruction (Rombach et al., 2022; Zhan et al., 2025). The encoder progressively extracts hierarchical features using Conv2D layers with spatial attention, which enhances feature weighting based on global and local resistivity variations. The encoded latent representation captures high-level subsurface patterns while reducing spatial complexity. The decoder reconstructs resistivity distributions by upsampling through Conv2DTranspose layers, with spatial attention improving feature refinement. A compact latent space facilitates efficient inversion while preserving resistivity structure (Figure S3 in Supporting Information S1). This architecture enhances global feature learning, noise resilience, and scale adaptability, making it well-suited for the studied task.

3.1.3. Curriculum Training

Curriculum training is a deep learning strategy where a model is trained on progressively more complex tasks, starting with simpler examples and gradually introducing more challenging ones. This approach improves model generalization, convergence stability, and feature learning efficiency by enabling the network to first learn fundamental patterns before handling intricate variations (Saad et al., 2024; Shorten & Khoshgoftaar, 2019).

In this study, curriculum training is adopted for data sets containing 1 to 5 anomalous conductive, or resistive, target bodies (areas with substantially contrasting resistivities to their backgrounds), as it allows the model to first learn basic resistivity contrasts with single anomalies before tackling more complex heterogeneous domains. This structured learning process helps to automate model determination on target amount in complex subsurface areas within the studied scale.

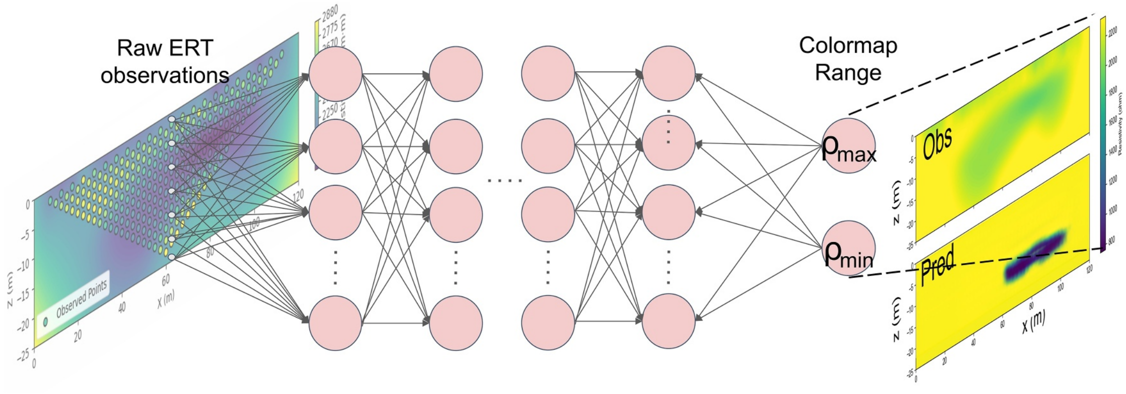

Figure 4. Resistivity colormap calibration for field observations using CNN. $\rho_{\max}$ and $\rho_{\min}$ of the model output is the predicted range of colormap which are the maximal and minimal resistivity values that the lightest and darkest color in the colormap respectively represent. The observatory resistivity image labeled "Obs" is plotted accordingly to be subsequently used as input of the ViT model, and the predicted image labeled "Pred" is also interpreted using this colormap for resistivity distribution. These are performed to be consistent with how the ViT model is trained on apparent and true resistivity images that share a colormap in its training data sets.

3.2. Resistivity Colormap Calibration Model

Using consistent resistivity colormaps for both apparent and true resistivity profiles is crucial for image-based ERT inversion models because it ensures that color intensities and gradients accurately reflect the same range of resistivity values across both images. For a pair of apparent and true resistivity profile images in the synthetic data set used, the colormap range is designated to be within the minimal and maximal resistivity value in both images in order to maximize image difference while preserving adequate image color variation details. When the colormap mapping is uniform, the model can directly learn the relationship between observed distortions and true subsurface structures by analyzing pixel-wise differences and spatial patterns consistently across all data samples, which enhances the model's ability to detect anomalies, recognize resistivity contrasts, and minimize inversion errors, resulting in more accurate resistivity reconstructions. In addition, such calibration ensures the ViT model generates image predictions in consistent colormaps, allowing model outputs to be directly interpreted as definite resistivity value distributions. However, in field observations, colormap range of the apparent resistivity image cannot be directly designated due to the absence of ground truth. Therefore, a calibration model is proposed by training a CNN model to learn optimal colormap ranges from the synthetic data sets to predict the colormap range from a set of apparent resistivity measurements on a given pseudo section, as shown in Figure 4.

The proposed residual CNN model shown in Figure 4 incorporates residual connections and multi-scale feature extraction to efficiently predict colormap ranges ($\rho_{\min}$ and $\rho_{\max}$ in Figure 4). With the colormap, the resistivity value ρ of any pixel on the images can be defined as:

$$\rho = \rho_{\min} + \frac{(\rho_{\max} - \rho_{\min})(P - 0)}{255} \tag{8}$$

where P is the pixel value when the image is plotted in grayscale, where 0 and 255 respectively stand for lightest (white) and darkest (black) color.

The training data set of the model is prepared from the synthetic data set by interpolating observed apparent resistivity values onto a consistent grid using the nearest-neighbor method, which transforms the scattered resistivity measurements into grid arrays. The grid array is then taken as the model input with the associated minimal and maximal resistivity values of the observed and true resistivity as the ground truth of the model output, improving computational efficiency while preserving critical resistivity distribution patterns. The model architecture includes three convolutional blocks with BatchNormalization and MaxPooling (Figure S4 in Supporting Information S1), progressively extracting hierarchical features. A residual connection between intermediate layers enhances gradient flow, preventing vanishing gradient issues during training (Zu et al., 2024). This calibration approach can eventually improve the ViT model's ability to learn resistivity contrasts by maintaining a uniform colormap mapping between observed and true resistivity profiles. Moreover, the predicted colormap allows the inversion image of field observations to be accurately interpreted to resistivity values.

3.3. Gauss-Newton Inversion Algorithm

The GN numerical inversion algorithm is a widely used iterative optimization method in geophysical inversion for estimating subsurface resistivity distributions from observed ERT data (Boyle et al., 2018) in commonly used inversion platforms like RES2DINV (Jamil et al., 2024). It addresses the nonlinear inverse problem by linearizing it through a Taylor series expansion, iteratively minimizing the objective function:

$$\Phi(m) = \frac{1}{2}\|d - f(m)\|^2 \tag{9}$$

where m is the model parameter vector, d represents the observed data, and $f(m)$ is the forward model prediction. The GN update rule for the model parameters is given by:

$$m_{k+1} = m_k - \left(J^{\mathrm{T}}J\right)^{-1}J^{\mathrm{T}}\left(f(m_k) - d\right) \tag{10}$$

where J is the Jacobian matrix of the forward model with respect to the model parameters. This update mechanism minimizes the misfit between observed and predicted data, refining the resistivity model at each iteration (Boyle et al., 2018).

The GN algorithm is widely adopted in the industry due to its robustness, reliability, and interpretability, offering quantifiable uncertainty estimates and sensitivity analysis. These strengths make it a benchmark method for evaluating the performance of image-based deep learning inversion models in field applications where true resistivity profiles are unavailable, providing a physically consistent reference for assessing the accuracy and stability of data-driven resistivity reconstructions (Rucker et al., 2017).

3.4. Inversion Model Training Configuration

The ViT model was trained on the synthetic image data set representing ERT resistivity profiles, with the data set split into a 7:3 training-to-testing ratio to ensure robust model evaluation. All images used for model training are normalized to ensure consistent data scaling, improve model convergence, and enhance the stability and accuracy of the model by maintaining uniform pixel value distributions. The model was compiled using the Adam optimizer, with mean squared error (MSE) as the loss function to measure pixel-wise accuracy and Structural Similarity Index (SSIM) as the evaluation metric to enhance the model's focus on preserving structural details in resistivity reconstructions (Yin et al., 2024b). MSE is defined as:

$$\mathrm{MSE} = \frac{1}{M \cdot N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_P(i,j) - I_G(i,j)\right)^2 \tag{11}$$

where $I_P(i, j)$ and $I_G(i, j)$ are the value of the pixel (i, j) in the predicted and true resistivity profile image, respectively. M and N are respectively the height and width of the image in pixels. SSIM is defined as:

$$\mathrm{SSIM}(x,y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left({\sigma_x}^2 + {\sigma_y}^2 + c_2\right)} \tag{12}$$

where x and y respectively represent the true and predicted resistivity profile image, μ_x and μ_y, σ_x and σ_y respectively represent the luminance mean value and contrast standard deviation of the images, c_1 and c_2 are stabilizing constants and σ_{xy} is the covariance of images. The SSIM thus has a range of 0–1, of which 0 stands for complete structural dissimilarity and 1 for absolute structural similarity between images.

The combination of high SSIM and low MSE values provides a comprehensive indication of high accuracy in image-based ERT inversion (Yin et al., 2024b), where MSE quantifies pixel-wise reconstruction errors, ensuring numerical fidelity, while SSIM evaluates structural preservation, capturing resistivity contrasts and anomaly boundaries, enabling a more robust evaluation of inversion performance across varying subsurface complexities.

During training, the model utilized a batch size of 16 and a 10% validation split within the training set to monitor performance on unseen data. The early stopping callback with a patience of 10 epochs was applied, allowing the model to restore the best weights when validation loss did not improve, preventing overfitting. The model was trained for a maximum of 50 epochs, while the early stopping mechanism could reduce the actual training duration by halting once convergence was achieved.

With this training setup, the model is trained on the 5 training data sets (respectively containing images of 1–5 targets) one after another in a stepwise manner to gradually increase model comprehension of data complexities. The training was implemented with Python 3.9 in a Windows system on an intel i9-13900K @ 3.00GHz CPU with 128GB RAM and a Nvidia RTX 4090 GPU to accelerate computation.

3.5. Robustness and Field Applicability Tests

To complement the architecture comparisons and synthetic training workflow, we designed two additional evaluations aimed at assessing robustness and real-world applicability of the proposed ViT-based inversion approach.

The trained model was evaluated under controlled synthetic perturbations and Gaussian noise to simulate realistic survey uncertainties. Perturbations were introduced by randomly scaling each grid cell resistivity value by a factor of $(1 \pm p\%)$, where p took values of 1%, 2%, and 5%. For example, at $p = 5\%$, a resistivity value of 100 Ω·m could be scaled randomly between 95 and 105 Ω·m. Gaussian noise was superimposed on the pseudo-sections by adding normally distributed random values with zero mean and standard deviations corresponding to 5% and 10% of the overall data amplitude. These controlled variations mimic systematic parameter fluctuations and measurement noise that commonly occur in ERT surveys, enabling assessment of the model's stability under realistic disturbance levels.

In addition to synthetic robustness tests, we incorporated a field demonstration using an electrical resistivity observation from the U.S. DOE Hanford site (Jamil et al., 2024). The Hanford site presents a heterogeneous vadose zone with alternating sandy, silty, and low-permeability layers, compounded by the presence of historical contamination. These conditions make it a particularly challenging test environment for resistivity inversion. Including a sample test in this field provides preliminary assessment of the proposed method under realistic geological and environmental complexities not captured in synthetic data sets.

4. Results and Discussion

4.1. Image-Based Inversion

The convergence of the ViT's loss functions are shown in Figure 5. This figure shows decline in metric value improvements over iterations with the increase of target numbers, as MSE convergence values gradually increased through (a) to (e), only in (a) and (b) did it achieve below or near below 0.01 values. As for SSIM, the convergence value also drops down to below 0.8 starting from (c). Model convergence on 1-target data set clearly out-performing all others. Early stopping begins to take effect starting from 3-target data set to prevent overfitting, as training and validation loss values started show divergence. However, the converging values are still satisfactory across all data sets, since MSE values are still below 0.02 and SSIM values above 0.7, which can be considered high image resemblance in terms of both pixel value and structural similarity in general. With the curriculum training strategy, the finally obtained ViT model has prediction results on the testing data set of images with 1–5 anomalous targets. as shown in Figure 6.

Figure 6 shows low MSE (below 1.5×10^{-3}) across all data sets with various target numbers, whereas SSIM has significantly decreased from above 0.9 to below 0.7 in scenarios with 4 and 5 target subsets due to the tendency of losing prediction fidelity on smaller targets, which exhibit less distinctive image patterns. In general, accurate delineation of major target zone geometrics and resistivity distributions in both target and background areas are achieved, demonstrating the ViT model was able to autonomously determine the amount of targets within the observed domain, producing satisfactory image inversion results across resistivity profiles with various levels of complexity. Since resistivity colormaps are shared between apparent and true resistivity profile images (Figure 1), the model is trained to predict images following consistent colormaps, allowing inversion resistivity values to be directly interpreted using existing $\rho_{\min}$ and $\rho_{\max}$ values in the colormaps (Equation 8). Further performance

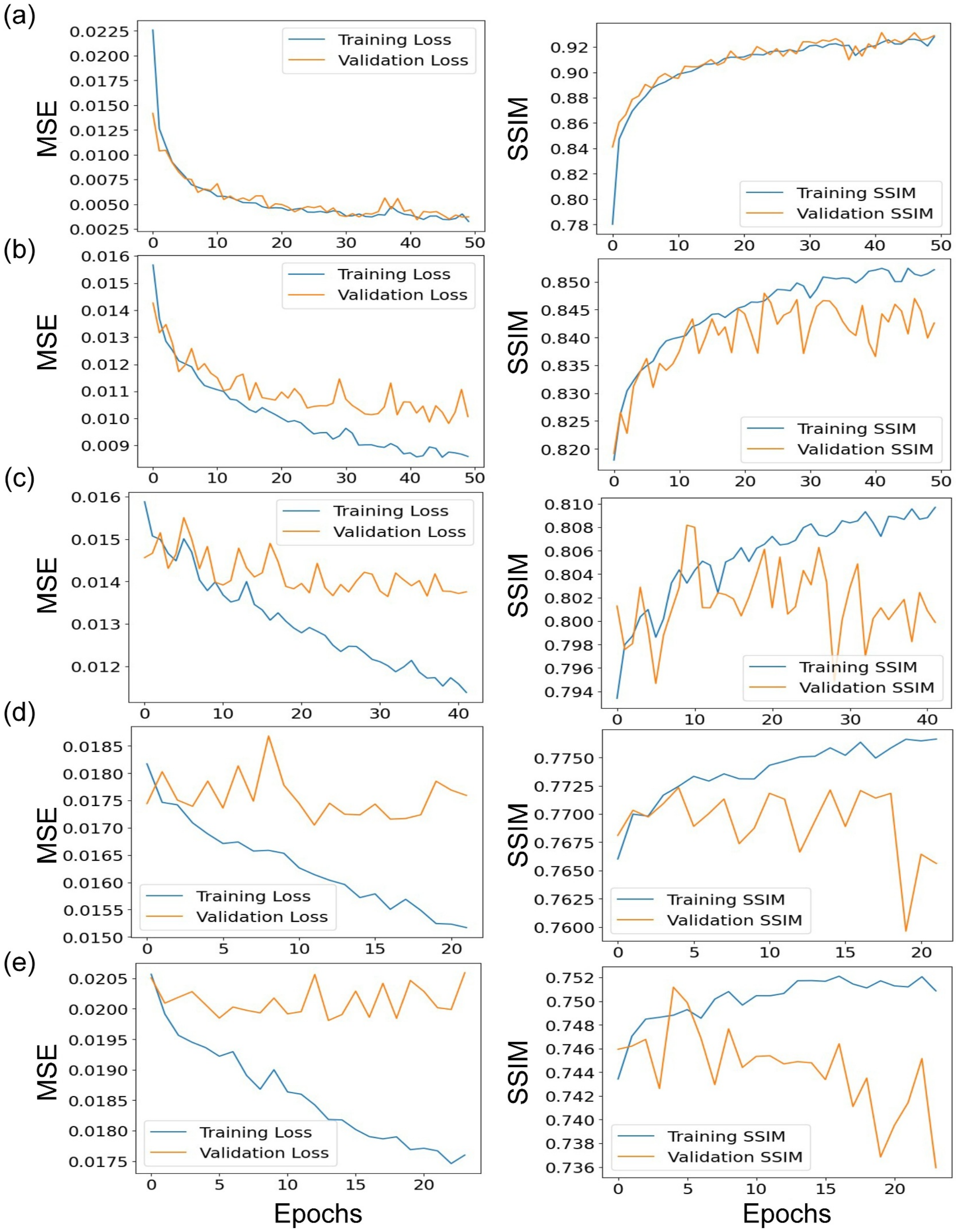

Figure 5. ViT training MSE loss and SSIM convergence across data sets containing (a–e) 1 to 5 targets within a domain. Satisfactory convergences of the metrics are presented in (a) and (b), whereas trends of overfitting are demonstrated in (c), (d), and (e).

evaluation on SSIM and MSE metrics across sub-datasets with 1–5 targets are shown in Table S1 in Supporting Information S1 and Figure 7.

Table S1 in Supporting Information S1 and Figure 7 show the SSIM values declining with increasing target complexity, indicating reduced structural similarity between predicted and true resistivity profiles. In addition, there is a gradual increase in MSE, suggesting growing errors on a pixel level. The confidence intervals' variation trends are inconsistent in Figures 7a and 7b, demonstrating the model's stability decreased in terms of structural

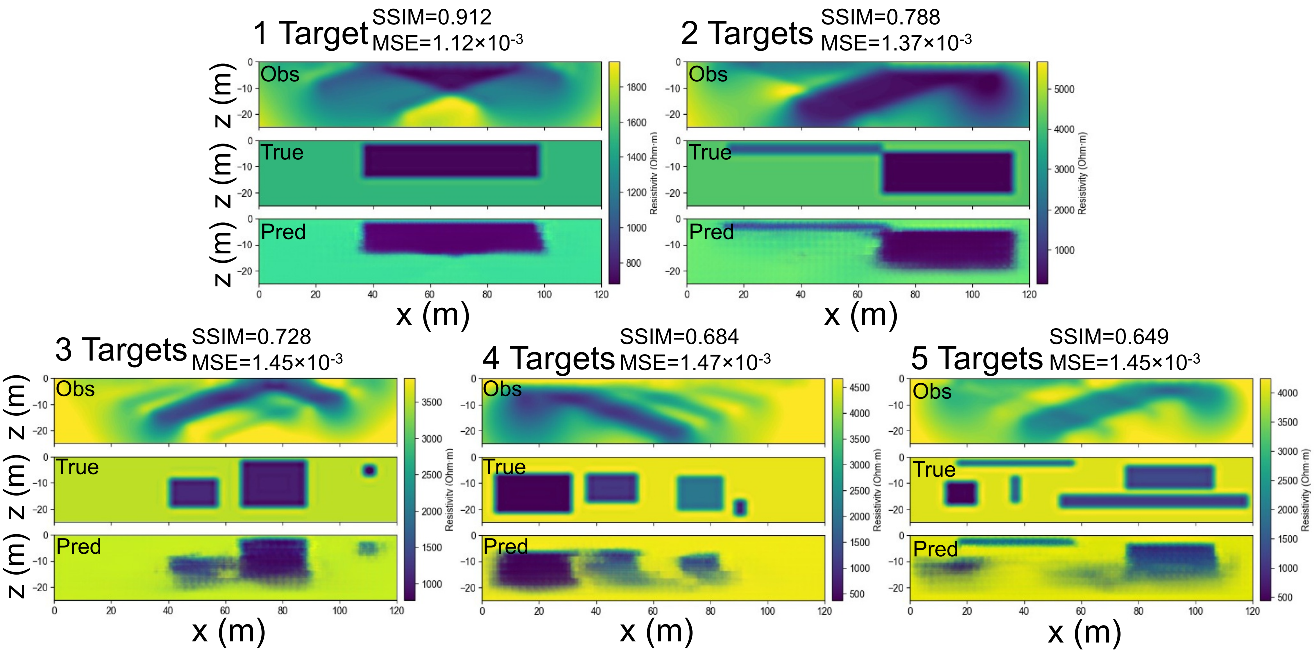

Figure 6. ViT model inversion image samples from domains containing 1 to 5 targets. "Obs" and "True" respectively stand for apparent and true resistivity profile images in the data set. "Pred" stands for the image-based inversion results. The SSIM and MSE metrics are measured on each subset containing 3,000 samples with corresponding target numbers of the testing set for average values, all MSE values shown are for normalized images. The locations and boundaries of anomalous conductive and resistive bodies are accurately presented in obtained results by visual inspection. Both target and background resistivity values are also well reflected as their color tones are visually close to those in the true resistivity profile images.

similarity with more target numbers, but increased in terms of pixel-wise measurements. This also implies that the model performance may be more stable for predicting overall resistivity values for the domain in complex subsurface scenarios, but less stable for acquiring geometric information of targets.

In summary, the curriculum training strategy ensured the ViT model to effectively learn resistivity mapping for all studied target numbers within domain while maintaining generalization capability to resistivity patterns of multiple complexity levels, facilitating accurate image-based ERT inversion for applications across various scenarios.

4.2. Model Comparison and Robustness Evaluation

The selected residual CNN AutoEncoder, UNet, and LDM models are also applied for identical tasks by training on 1-target data set first using identical training setups as the ViT model, except LDM takes no validation set from the training set, since preliminary experiments showed that model performance with a validation set was poorer than without, likely due to the inherent variability of the synthetic data and validation-induced interruptions in

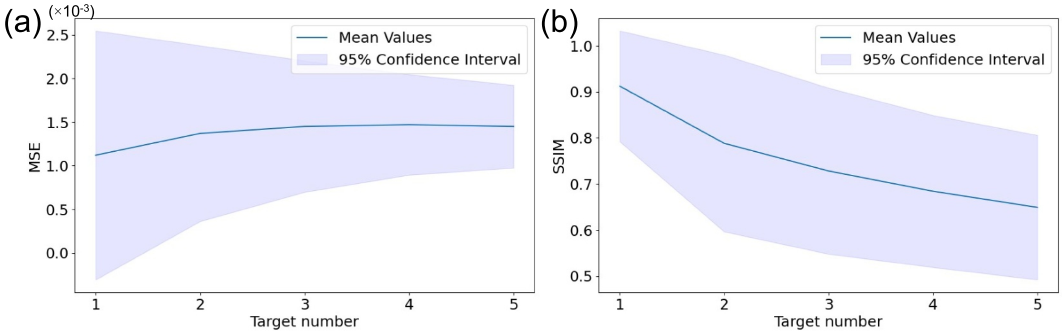

Figure 7. (a) SSIM and (b) MSE variation of ViT model across data sets containing 1 to 5 anomalous targets.

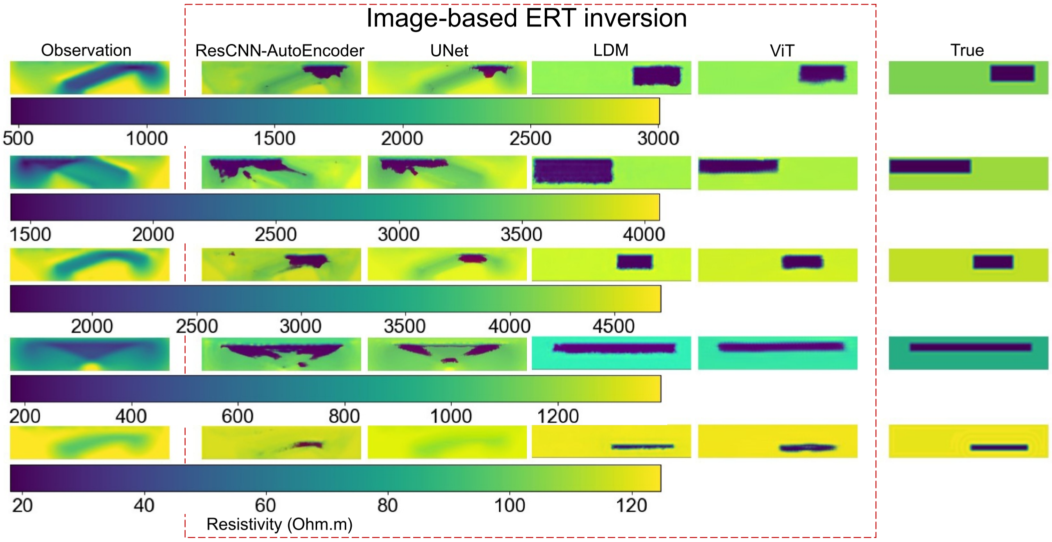

Figure 8. Image-based ERT inversion results comparison of various models for the 1-target data set. The observation images in the first column are the input images of apparent resistivity pseudo sections, while the images of true resistivity profiles in the last column serve as the ground truth of the results. Color and geometric distortions to the ground truth are substantially more obvious in output images of the ResCNN-AutoEncoder and UNet, indicating less accurate resistivity inversion results in comparison with LDM and ViT.

latent space learning. By training without validation, the model maintains stable learning dynamics and robust evaluation on the testing set. The results obtained are shown on Figure 8. The CNN AutoEncoder and UNet models demonstrate reasonable reconstructions but exhibit blurred boundaries and artifacts. The LDM shows sharper anomaly detection, maintaining structural consistency, but struggles with edge definition. The ViT model provides the best performance of accurately capturing anomaly shapes, boundaries, and resistivity values in terms of imaging colors with high contrast and minimal artifacts, closely aligning with true profiles. The detailed performance metrics are shown in Table 1.

Consistent with the visual comparisons in Figure 8, the ViT model achieves the highest SSIM of 0.912 and lowest MSE of 1.12×10^{-3} with relatively low standard deviations, with the shortest training time of 706 s and relatively low standard deviation for SSIM and MSE across the 3,000 samples in the testing set, confirming its superior ability to maintain structural integrity and minimize reconstruction errors with the highest computation efficiency and competent model stability, which is vital for obtaining target geometrics and resistivity distributions in complex domains. Only LDM has comparable performance at the cost of a significantly higher training time of 6,608 s due to its more complicated training mechanism.

As the ViT model excels in the inversion task benefiting from its global feature extraction capabilities, the LDM also offers balanced performance between detail preservation and noise suppression for low-complexity scenarios in which only one target is present. However, in more complex scenarios, LDM's advantages substantially weakened as shown in Figure 9 and Table S2 in Supporting Information S1.

Table 1
Inversion Performance Comparison of Various Models for the 1-Target Data Set

	ViT		CNN-AutoEncoder		UNet		LDM	
Model	SSIM	MSE (10^{-3})	SSIM	MSE (10^{-3})	SSIM	MSE (10^{-3})	SSIM	MSE (10^{-3})
Mean	0.912	1.12	0.859	22.3	0.857	28.4	0.919	1.13
Max	0.984	2.83	0.985	103.2	0.987	93.1	0.990	2.46
Min	0.346	0.294	0.616	3.25	0.605	2.74	0.333	0.176
Std	0.0613	0.728	0.0605	11.5	0.0722	14.7	0.0624	0.520
Training Time (sec)	706		1,354		1,005		6,608	

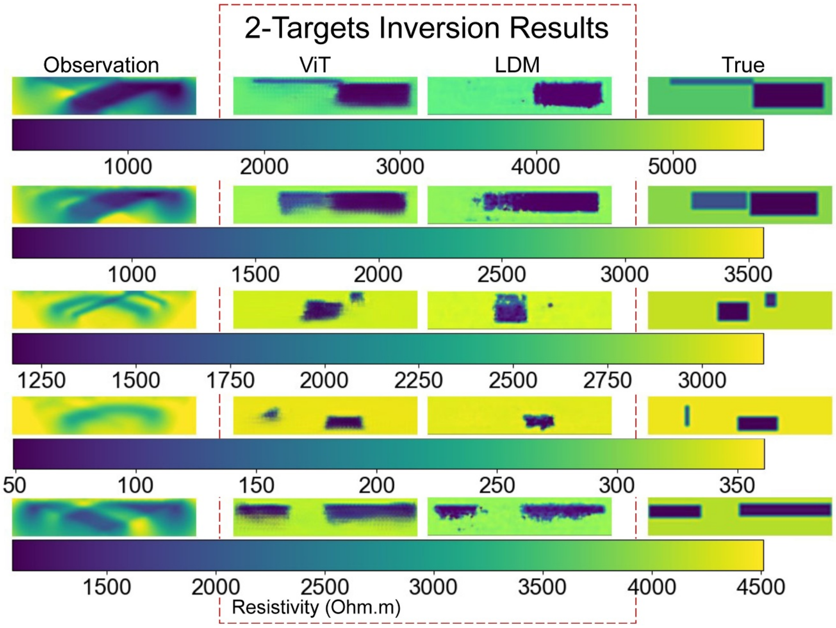

Figure 9. Inversion results of ViT and LDM for the 2-target data set.

Figure 9 shows the ViT model effectively captures both large- and small-scale anomaly targets and maintains relatively clear boundaries, demonstrating robustness in reconstructing more complex resistivity profiles with 2 targets, resulting in an average SSIM value of 0.788. In contrast, the LDM results exhibit inconsistencies and instability, losing or dispositioning smaller targets like in the first, third, and fourth image. In addition, resistivity values are not as accurately reflected as in ViT results in extremely high target to background contrast like the first and fifth image. These lead to a lower average SSIM value of 0.718 and higher MSE standard deviation of 0.590 (Table S2 in Supporting Information S1), indicating LDM's performance substantially degrades with increased complexity, resulting in blurred edges in targets like the fifth image and overall reduced accuracy. This also implies that the LDM model would not be suitable for inversion tasks of complexity levels with more than 1 target due to its instability arises from the reliance on iterative denoising processes, which are sensitive to fine-scale details and may introduce noise during reconstruction. On the other hand, the ViT model takes advantage of self-attention mechanisms and enhances its ability to maintain accuracy across diverse complexity levels.

Beyond direct comparisons across architectures, we further assessed the robustness of the fully trained ViT model under controlled perturbations and Gaussian noise, as detailed in Section 3.5. The evaluation was conducted on the testing set of the 5-target data set, with results summarized in Figure 10.

Despite these increasing disturbance levels, the ViT model consistently preserved coherent resistivity structures and anomaly boundaries, with only gradual degradation in SSIM and MSE metrics. This robustness evaluation demonstrates that, in addition to achieving superior performance against benchmark models, the ViT model maintains stability under realistic observational disturbances.

4.3. Resistivity Colormap Calibration

To take advantage of the promising performance of the ViT model, the model input in the form of observational ERT imaging data needs to be calibrated with the image colormap of true resistivity profile as explained in previous sections. This refers to aligning resistivity colormap ranges based on the maximal and minimal value in both observation and true resistivity profiles, so that the model is able to obtain accurate inversion results by applying the uniform image-to-image mapping learned on the hypothetical data sets. In addition, the model output images can be converted back to resistivity maps with Equation 8 using consistent colormaps shared with the model input images. Since true resistivity profiles for field observations of ERT data are absent for performing such calibration, the proposed CNN-based calibration model is used to predict the range of resistivity colormaps from a given observed resistivity profile with the associated patterns learned on the synthetic data sets.

10.1029/2025JH000734

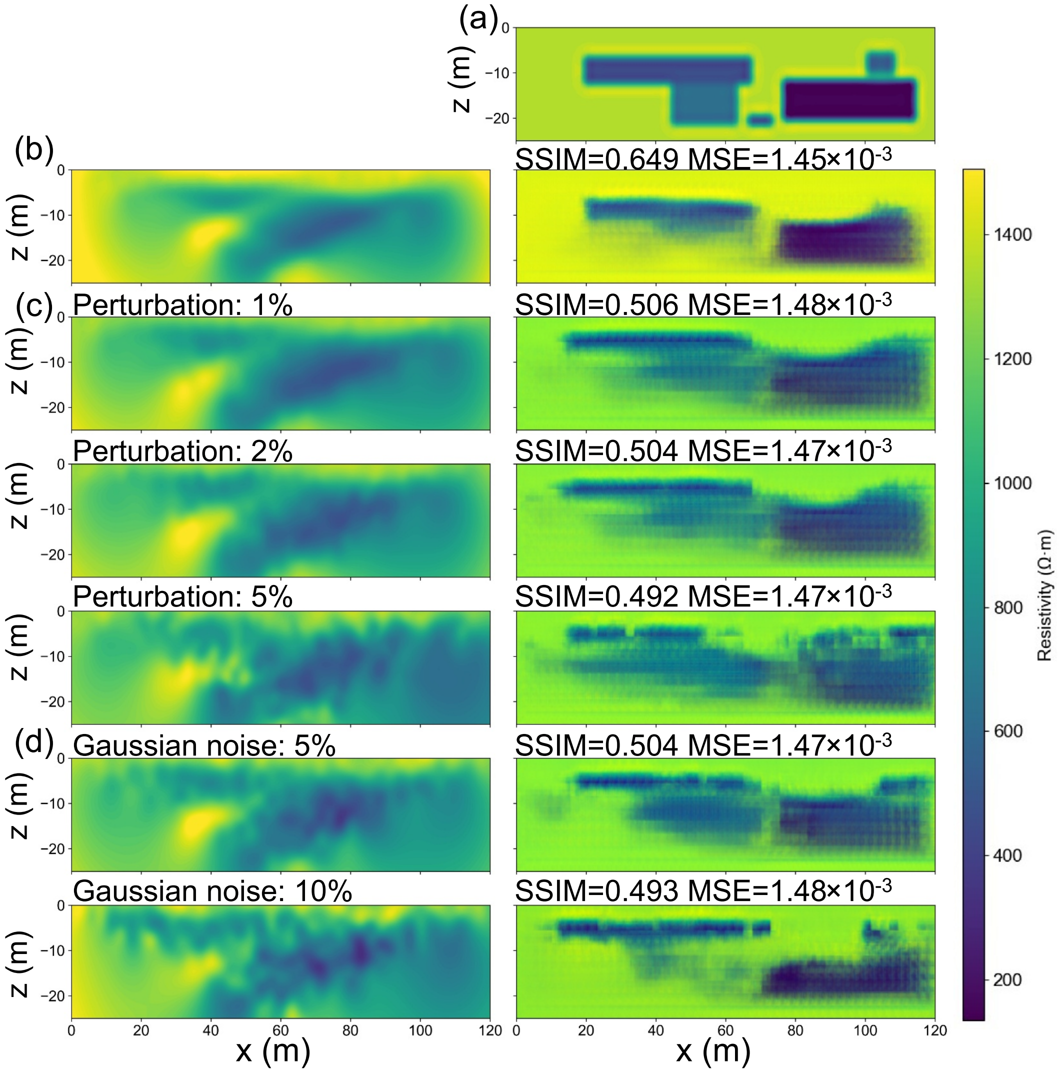

Figure 10. A sample of (a) true ERT profile, its (b) pseudo section (left), pseudo section with synthetic (c) perturbation and (d) Gaussian noise, and corresponding ViT model output (right). Mean SSIM and MSE value for all samples of the testing set is labeled correspondingly.

The grid array data prepared from the original apparent resistivity profiles is also split using a 7:3 ratio for training and testing set for the calibration model. In addition, the training data is further split into training and validation subsets with an 8:2 ratio, ensuring model evaluation stability. The training setup for the adopted residual CNN model employed Adam optimizer, MSE loss, and mean absolute error (MAE) as the evaluation metric. The training process is configured for 100 epochs with a batch size of 32, incorporating early stopping with a patience of 20 epochs to prevent overfitting. The curriculum training strategy previously employed is also applied across data sets containing 1 to 5 targets. The training and validation loss function decreased quickly within 10 iterations for the 1-target data set to below 200×10^3, as shown in Figure S4 in Supporting Information S1. Improvements on convergence value can still be found with over each training on increased target number, eventually reaching below 50×10^3 for 5-targets data set, despite earlier intervention of early stopping to prevent model instability exhibited as the fluctuations of the validation loss. The obtained calibration results with the fully trained model are shown in Figure 11 and Table S3 in Supporting Information S1, prediction values of both upper and lower limits of the resistivity colormap range are closely distributed along the identity line, and apparent resistivity images generated with the predicted colormap range demonstrates high resemblance to the original images calibrated with the true resistivity profile using Equations 2 and 3, which indicates promising performance of the calibration

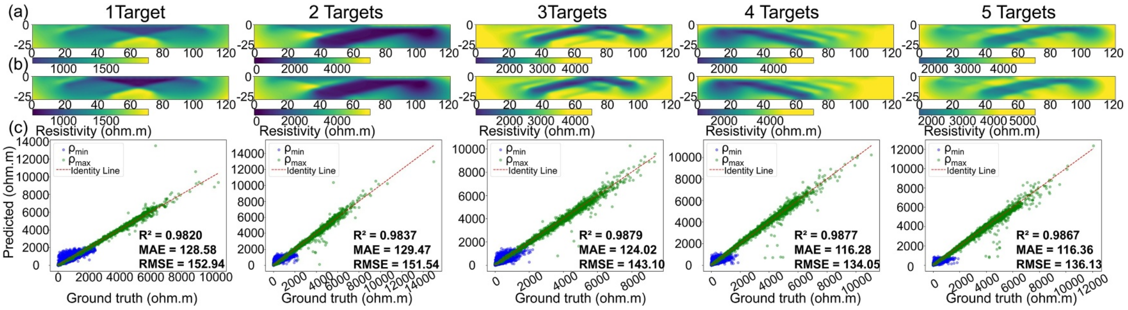

Figure 11. Prediction results of the upper and lower limit of resistivity colormap across 1- to 5-target data set, respectively represented by green and blue scatters in (c), in which both MAE and RMSE are generally within 150 Ωm, and R^2 values are above 0.98, demonstrating very low level of error across data sets in all levels of complexities. Corresponding sample images demonstrate original apparent resistivity image features in (b) are nearly unaffected by imaging with the predicted range of colormap (a), where little color distortions can be found.

model. This also implies that the model's high reliability during the image generation process using observation resistivity profile for ViT model's inputs when true resistivity values are unavailable.

Furthermore, the performance of the ViT inversion model using the calibrated images as inputs is examined by comparing with the original results shown in Figure 6 where true resistivity values are used for colormap calibration. Samples of the comparison are shown in Figure 12, which presents the ViT inversion image results between input images of calibrated colormap and original input images have minor visual differences. Inversion resistivity values are also highly consistent with the original results (Figure 12c) regardless of the data complexity level as the 5-targets data set has the lowest MAE and RMSE and highest R^2. The colormap calibration has no evident impairing effect on ViT performance in terms of image evaluation metrics including SSIM and MSE (Table S4 in Supporting Information S1), which are comparable to those of original results (Table S1 in Supporting Information S1). This further confirms the proposed resistivity colormap calibration method's capability of maintaining ViT performance when true resistivity values are absent for the calibration.

4.4. Inversion Results Comparison With Gauss-Newton Method

With accurate calibration on resistivity colormaps, inversion images from observational resistivity profiles are mapped to continuous resistivity distributions across the entire domain. The inversion image is converted to grayscale images before being mapped to actual resistivity distributions using the calibrated resistivity colormap obtained and Equation 8, utilizing linear conversion between pixel values and resistivity values. This section

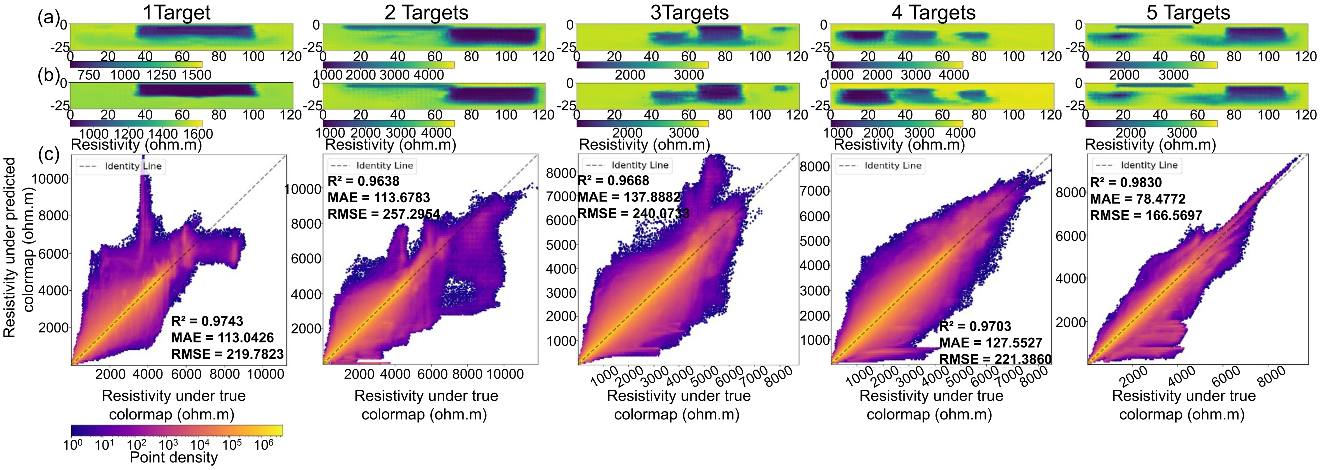

Figure 12. (c) Scatter density plot of inversion resistivity values represented in each pixel of ViT output images across the testing sets of various target numbers for apparent resistivity input image with calibrated colormap against original colormap. Inversion result samples with (a) calibrated colormap and (b) original inversion result samples are also compared. R^2 values in all complexity levels are all above 0.96, MAE and RMSE are respectively below 140 and 260 Ωm. Both visual differences of the image samples and resistivity colormap range value differences are minor.

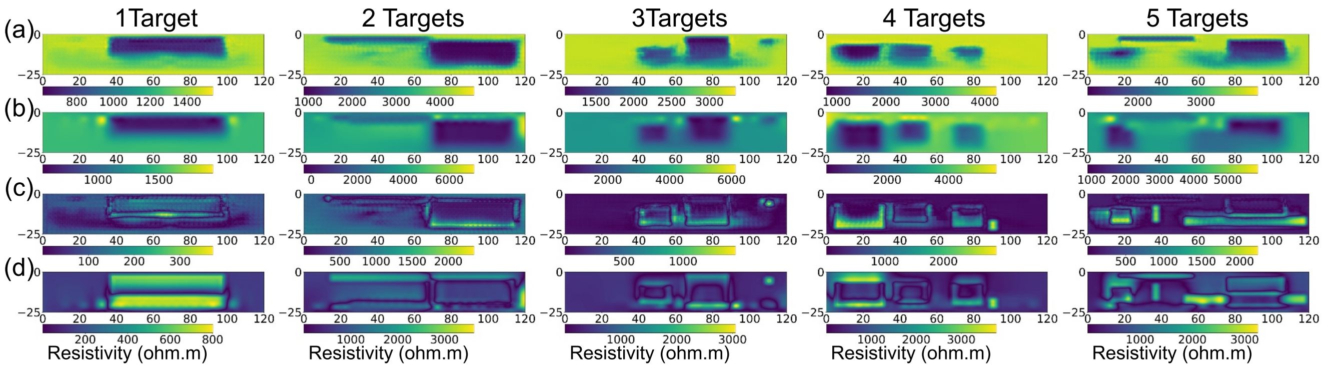

Figure 13. Samples of the inversion results obtained from (a) the proposed method (ViT) being compared with those obtained from (b) Gauss-Newton (GN) method across 1- to 5-target data set, and corresponding error maps between the ground truth and results of the (c) ViT and (d) GN method.

extends the evaluation beyond deep learning benchmarks to include the Gauss–Newton method, thereby situating the proposed model's performance relative to the established inversion standard.

The results of ViT inversion in the format of resistivity distribution maps shown in Figure 12a are subsequently compared with results of the GN numerical simulation method. The ViT model takes the inputs of the images generated from apparent resistivity profiles with calibrated colormap as previously discussed, and the GN method takes original measurements of corresponding apparent resistivity profiles. These are performed in coherent with field applications where true resistivity profiles are absent. Since the proposed method demonstrates high model stability with low standard deviation of performance metrics (Table 1, Table S1 in Supporting Information S1), five representative samples across various data complexity levels are taken for analysis as shown in Figure 13. The anomalous targets delineated in ViT and GN method are visually resemblant in their geometrics, exhibiting consistent shapes, sizes, and positions, whereas target boundaries in ViT results are apparently clearer. Additionally, almost all error maps for ViT results present lower errors with restrained high-error area and reduced overall average error values, proving superior performance of the proposed method in attaining resistivity distributions that are closer to ground truth for both target and background areas. Furthermore, the trained ViT carried out inversions within 20 ms for each sample, substantially outperforms GN's iterative approach, which typically took at least 5 s.

The resistivity values of each pixel in the predicted image are further plotted against corresponding true resistivity values and compared with the GN method, as shown in Figure 14. The measurement on GN method is interpolated using the same radial basis function for true resistivity profile interpolation to match the resolution of the proposed method's image result. Deviation of predicted values from the true values are exhibited in all results on such pixel level measurements, causing none of the scatters to be well-fitted to the identity line. Still, scatters of ViT results are more closely distributed with lower MAE and RMSE values, especially for the 1-target data set results, which also get the lowest MAE and RMSE. However, there is not a general variation trend of these metrics with the increase of data complexity levels.

Figures 13 and 14 proves the proposed method's promising capabilities in both localizing anomalous targets with accurate geometric properties and sharp boundaries. In comparison with the bench marking GN numerical simulation method, the proposed method provides a competent solution for obtaining true resistivity distributions of high fidelity with significantly higher efficiency. Furthermore, the increase of complexity level did not have a major impact on the method's performance, proving the effectiveness of curriculum training in establishing the model's autonomous adaptability to various levels of heterogeneity, which also implies the method's potential and applicability in future applications, especially when an adequate amount of field data can be obtained for model fine-tuning (similar to the curriculum training employed) and transfer learning (Yin et al., 2024a) to suit specific task scenario and requirements.

To further address real-world applicability, a case demonstration using a measurement from the U.S. DOE Hanford site was included (Section 3.5), with results shown in Figure 15. Following the proposed workflow, the pseudo-section of the collected field data was first used as input for image prediction, followed by colormap calibration to match the resistivity range. The obtained image was then reversed to a resistivity distribution for

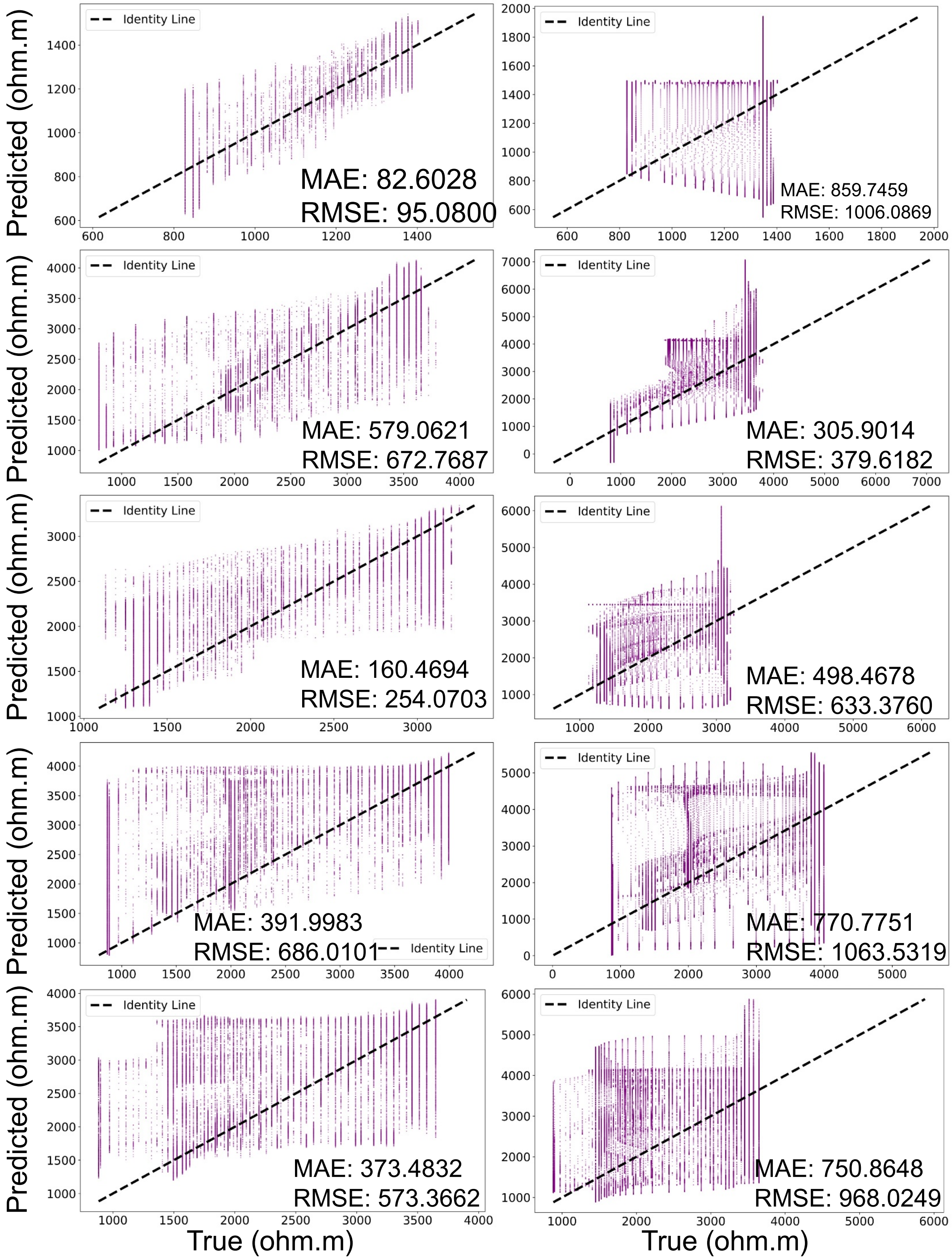

Figure 14. Predicted against true resistivity values plotted across all pixels in each sample image result across 1- to 5-target data set. Left and right columns represent results obtained from the proposed method and Gauss-Newton method, respectively.

comparison with the traditional GN inversion. Although notable discrepancies appear in regions of resistivity contrasts between the GN inversion and our ViT-based model, this demonstration provides important insights into how the proposed method can be extended to real data. Specifically, our approach is capable of extracting coherent structural features with sharper boundaries that align with the GN inversion, suggesting its potential to highlight anomalies that may be poorly resolved in traditional site characterization workflows.

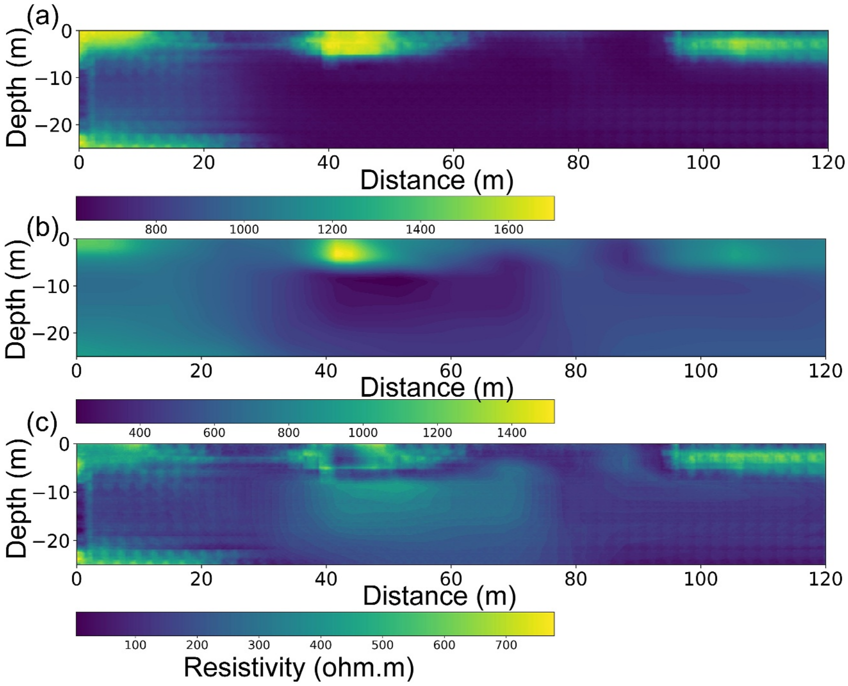

Figure 15. Field measurement inversion results of (a) the proposed method, (b) traditional GN method, and (c) the differentiation between the two methods. Resistive targets at about 40–50 m, and 100–120 m are delineated by the proposed method with sharp contrast to the background, despite some discrepancies with the GN inversion in surrounding areas.

The results illustrate that the proposed method remains effective under these challenging field conditions, supporting its utility for practical ERT applications. It should be clarified that only one representative field sample is used and the current model was trained solely on synthetic data. The performance of the proposed approach could be further improved in future studies by fine-tuning or retraining the network with a sufficient amount of field data to better accommodate diverse site-specific features.

4.5. Discussion

The results are organized to first identify the most effective deep learning architecture and then to compare it against the GN standard. This highlights both the internal benchmarking and the geophysical relevance of the proposed workflow. Despite the method's promising performance demonstrated on forward modeling data sets across varying complexities, several limitations and considerations remain:

Model sensitivity and accuracy: While the proposed ViT model demonstrates strong predictive capability compared to Gauss–Newton inversions, its sensitivity decreases for small-scale targets or scenarios with more than three anomalies. This is evident in reduced fidelity at fine scales (Figure 12) and wider confidence intervals in SSIM and MSE metrics under high-complexity data sets. The model remains effective at capturing large-scale resistivity contrasts and overall structure, but finer details are more prone to smoothing or distortion.

Simplified anomaly geometry: The synthetic data set was generated using rectangular or similarly shaped anomalies. This abstraction was made deliberately to control the training process and provide a systematic benchmark against the Gauss–Newton method. The rectangular target geometry can be considered generally representative of some subsurface features including underground mine shafts, pipes, tunnels, storage tanks, karst features (e.g., caves), or even some sedimentary features such as buried paleo channel deposits. However, natural subsurface structures like fractures, stratified layers, and heterogeneous resistivity distributions, are often more irregular. This simplification may constrain generalization, and an ablation study with more diverse geometric scenarios would be valuable. We emphasize that the ViT-based workflow is flexible: retraining or fine-tuning with data sets including irregular shapes, stratified layering, or fracture-like targets could readily extend applicability.

Noise robustness: We have explicitly evaluated robustness to synthetic perturbations (1%–5%) and Gaussian noise (5%–10%). Results show that large-scale resistivity structures remain identifiable under moderate noise, though fine-scale features are progressively smoothed. This demonstrates stability under observational disturbances, while also highlighting the need for uncertainty quantification methods (e.g., Bayesian deep learning, ensembles) in future work.

Applicability to field-measurement data: As a proof-of-concept, we applied the method to a measurement from the U.S. DOE Hanford site. This demonstrates that the workflow can be extended beyond idealized synthetic conditions, though further systematic validation with diverse field data sets is needed.

To address these limitations in future studies and applications of the method, the flexibility of forward modeling can be utilized for generating tailored training data sets with specific target geometries to meet unique subsurface conditions and reduce potential discrepancies in geological heterogeneity not represented in the synthetic data. By adjusting the geometric complexity of synthetic targets and measurement configurations during data set preparation, the model training process can be customized to address particular geological contexts, such as cylindrical for boreholes, ellipsoidal for geological lenses, or irregular shapes for natural voids and fractures. This adaptability ensures that the proposed method not only provides general solutions for broad ERT applications but also supports specialized model training for site-specific inversion tasks, enhancing the method's practical utility in complex subsurface environments. Future work will therefore focus on systematically expanding synthetic data sets to include irregular geometries, stratified structures, and fracture-like anomalies, while also incorporating uncertainty-aware inversion strategies such as Bayesian deep learning or ensembles. Such developments will further enhance the model's generalization, robustness, and reliability for deployment with real-world field data.

5. Conclusions

The proposed ViT model incorporating convolutional blocks presents a novel approach to image-based ERT inversion, demonstrating robust performance in mapping apparent resistivity profiles to true subsurface resistivity distributions. The study's findings highlight several key conclusions:

- The ViT model has achieved enhanced inversion accuracy, effectively capturing both geometric features and resistivity contrasts of subsurface anomalies, outperforming existing frequently used CNN-based models, and more advanced LDM, particularly in complex geological scenarios.
- The adoption of a curriculum training strategy, starting from simple single-target scenarios and progressing to multi-target complexities, facilitates the model's generalization and adaptability to various subsurface conditions, enabling automated complexity recognition for its applications.
- The proposed residual CNN model for resistivity colormap calibration provides a practical solution for field applications where true resistivity data is unavailable, allowing consistent resistivity mapping and image interpretation by predicting colormap ranges directly from observed ERT data.
- The method exhibited accuracy, efficiency, and reliability advantages when validated against the bench marking GN numerical simulation inversion method, proving its competency and potential in handling ERT data containing various levels of heterogeneity.

This study demonstrates that the proposed method offers a powerful image-based alternative to traditional inversion techniques, contributing to advancements in ERT data interpretation and supporting effective decision-making in subsurface exploration projects. The method's ability to generalize across diverse geological settings, combined with efficient computational performance, makes it well-suited for large-scale and high-fidelity ERT imaging, offering significant advantages for subsurface target characterization in hydrogeological, environmental, and mining sectors.

Future studies can be conducted in several aspects for potential improvements and addressing method limitations, including enhancing model adaptability to small-scale targets and customizing target geometric complexities in synthetic and field training data to handle diverse geological conditions. In addition, the results obtained on ERT data can be assimilated with other geophysical measurements, such as seismic or ground-penetrating radar data, to further enhance inversion accuracy.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The data that support the findings of this study are generated using forward modeling codes (Yin et al., 2025a). The model implementation, trained weights, and sample inversion results are available (Yin et al., 2025b).

Acknowledgments

This work was supported by the NSF (Award Number (FAIN): 2142686) and U.S. Department of Energy Environmental Management Minority Serving Institution Partnership Program (EM-MSIPP) managed by the Savannah River National Laboratory under BSRA contract TOA 800002114. We appreciate assistance of Fayaz Shaik and the entire Carroll Lab.

References

Abbas, M., Deparis, J., Isch, A., Mallet, C., Jodry, C., Azaroual, M., & Baltassat, J. M. (2022). Hydrogeophysical characterization and determination of petrophysical parameters by integrating geophysical and hydrogeological data at the limestone vadose zone of the Beauce aquifer. *Journal of Hydrology*, *615*, 128725. https://doi.org/10.1016/j.jhydrol.2022.128725

Aleardi, M., Vinciguerra, A., & Hojat, A. (2021). Ensemble-based electrical resistivity tomography with data and model space compression. *Pure and Applied Geophysics*, *178*(5), 1781–1803. https://doi.org/10.1007/s00024-021-02730-1

Amiri, H., Vogel, H., & Plümper, O. (2024). New 2D to 3D reconstruction of heterogeneous porous media via deep generative adversarial networks (GANs). *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000178. https://doi.org/10.1029/2024jh000178

Azad, R., Fayjie, A. R., Kauffmann, C., Ben Ayed, I., Pedersoli, M., & Dolz, J. (2021). On the texture bias for few-shot CNN segmentation. In *Proceedings of the IEEE winter conference on applications of computer vision (WACV)*. Online.

Bai, P., Zhao, D., Li, C., Qiao, S., & Liu, Z. (2024). Arbitrary dipole-dipole observation systems and high-precision resistivity imaging algorithms for complex survey areas. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–12. https://doi.org/10.1109/tgrs.2024.3499978

Binley, A., Hubbard, S. S., Huisman, J. A., Revil, A., Robinson, D. A., Singha, K., & Slater, L. D. (2015). The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. *Water Resources Research*, *51*(6), 3837–3866. https://doi.org/10.1002/2015WR017016

Blanchy, G., Saneiyan, S., Boyd, J., McLachlan, P., & Binley, A. (2020). ResIPy, an intuitive open source software for complex geoelectrical inversion/modeling. *Computers & Geosciences*, *137*, 104423. https://doi.org/10.1016/j.cageo.2020.104423

Boyle, A., Wilkinson, P. B., Chambers, J. E., Meldrum, P. I., Uhlemann, S., & Adler, A. (2018). Jointly reconstructing ground motion and resistivity for ERT-based slope stability monitoring. *Geophysical Journal International*, *212*(2), 1167–1182. https://doi.org/10.1093/gji/ggx453

Chhun, K. T., Woo, S. I., & Yune, C.-Y. (2022). Inversion of 2D cross-hole electrical resistivity tomography data using artificial neural network. *Science Progress*, *105*(1), 00368504221075465. https://doi.org/10.1177/00368504221075465

Ciampi, P., Giannini, L. M., Cassiani, G., Esposito, C., & Papini, M. P. (2024). Geo-constrained clustering of resistivity data revealing the heterogeneous lithological architectures and the distinctive geoelectrical signature of shallow deposits. *Engineering Geology*, *337*, 107589. https://doi.org/10.1016/j.enggeo.2024.107589

Cui, M., Luo, C., Li, G., & Pan, M. (2021). The parameterized level set method for structural topology optimization with shape sensitivity constraint factor. *Engineering with Computers*, *37*(2), 855–872. https://doi.org/10.1007/s00366-019-00860-8

Dai, Z., & Samper, J. (2004). Inverse problem of multicomponent reactive chemical transport in porous media: Formulation and applications. *Water Resources Research*, *40*(7). https://doi.org/10.1029/2004WR003248

Damavandi, K., Abedi, M., Norouzi, G. H., & Mojarab, M. (2022). Geoelectrical modelling of a landslide surface through an unstructured mesh. *Bulletin of the Geophysical Observatory*, *63*(2), 337–356. https://doi.org/10.4430/bgo00384

De Peppo, G. P., Cercato, M., & De Donno, G. (2024). Cross-gradient joint inversion and clustering of ERT and SRT data on structured meshes incorporating topography. *Geophysical Journal International*, *239*(2), 1155–1169. https://doi.org/10.1093/gji/ggae326

Dick, M. D., Bery, A. A., Okonna, N. N., Ekanem, K. R., Bashir, Y., & Akingboye, A. S. (2024). A novel machine learning approach for interpolating seismic velocity and electrical resistivity models for early-stage soil-rock assessment. *Earth Science Informatics*, *17*(3), 2629–2648. https://doi.org/10.1007/s12145-024-01303-9

Dimech, A., Attard, G., Farrugia, A., Schembri, J., Farrugia, P. S., & Bianco, L. (2022). A review on applications of time-lapse electrical resistivity tomography (TL-ERT). *Surveys in Geophysics*, *43*, 1475–1518. https://doi.org/10.1007/s10712-022-09731-2

Dong, F., & Yue, S. (2025). A nonlinear sensitivity matrix for improving the imaging quality in electrical resistance tomography. *Measurement*, *243*, 116407. https://doi.org/10.1016/j.measurement.2024.116407

Dong, Y. L., & Zhang, Z. J. (2024). Deep forest modeling: An interpretable deep learning method for mineral prospectivity mapping. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(4), e2024JH000311. https://doi.org/10.1029/2024jh000311

Dosoky, W. (2023). Assessment of three mixed arrays dataset for subsurface cavities detection using resistivity tomography as inferred from numerical modelling. *SN Applied Sciences*, *5*(11), 303. https://doi.org/10.1007/s42452-023-05539-w

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. https://doi.org/10.48550/arXiv.2010.11929

Doyoro, Y. G., Gelena, S. K., & Lin, C.-P. (2025). Improving subsurface structural interpretation in complex geological settings through geophysical imaging and machine learning. *Engineering Geology*, *344*, 107839. https://doi.org/10.1016/j.enggeo.2024.107839

Ducut, J. D., Alipio, M., Go, P. J., Concepcion, R., II, Vicerra, R. R., Bandala, A., & Dadios, E. (2022). A review of electrical resistivity tomography applications in underground imaging and object detection. *Displays*, *73*, 102208. https://doi.org/10.1016/j.displa.2022.102208

Fan, J., Xu, W., Li, X., & Wang, X. (2023). Fine detection of large-scale complex geological structures based on geophysical prospecting techniques. *Environmental Earth Sciences*, *82*(3), 83. https://doi.org/10.1007/s12665-023-10759-9

Gao, J., Smirnov, M., Smirnova, M., & Egbert, G. (2020). 3-D DC resistivity forward modeling using the multi-resolution grid. *Pure and Applied Geophysics*, *177*(6), 2803–2819. https://doi.org/10.1007/s00024-019-02365-3

Gao, W., Xiang, C., Wu, C., Li, X., Zhang, W., Tang, L., et al. (2024). Unveiling heavy metal(loid) contamination and migration at an abandoned smelting site: Integrated geophysical and hydrological analyse. *Chemical Engineering Journal*, *500*, 156853. https://doi.org/10.1016/j.cej.2024.156853

Guo, Q., Liu, B., Wang, Y., & He, D. (2023). A deep learning inversion method for 3-D electrical resistivity tomography based on neighborhood feature extraction. *IEEE Sensors Journal*, *23*(16), 18550–18558. https://doi.org/10.1109/jsen.2023.3293205

Harsuko, R., Cheng, S., & Alkhalifah, T. (2025). Propagating the prior from shallow to deep with a pre-trained velocity-model generative transformer network. *Journal of Geophysical Research: Machine Learning and Computation*, *2*(1), e2024JH000408. https://doi.org/10.1029/2024jh000408

Haruzi, P., & Moreno, Z. (2023). Modeling water flow and solute transport in unsaturated soils using physics-informed neural networks trained with geoelectrical data. *Water Resources Research*, *59*(6), e2023WR034538. https://doi.org/10.1029/2023wr034538

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEe conference on computer vision and pattern recognition (CVPR), Seattle, WA*.

Hiskiawan, P., Chen, C.-C., & Ye, Z.-K. (2023). Processing of electrical resistivity tomography data using convolutional neural network in ERT-NET architectures. *Arabian Journal of Geosciences*, *16*(10), 581. https://doi.org/10.1007/s12517-023-11690-w

Huang, E., Zhang, M., Zhang, Y., Chen, B., Peng, L., & Li, Y. (2024). Online calibration of ERT based on impedance analysis in oil-water two-phase flow measurement. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1–12. https://doi.org/10.1109/tim.2024.3400348

Hung, Y.-C., Wang, H., Wu, P.-L., Liu, H.-C., & Lin, C.-P. (2024). 3D effect and countermeasure of 2D geoelectrical imaging of a subsurface linear structure. *Engineering Geology*, *338*, 107603. https://doi.org/10.1016/j.enggeo.2024.107603

Jamil, A., Rucker, D. F., Lu, D., Brooks, S. C., Tartakovsky, A. M., Cao, H., & Carroll, K. C. (2024). Comparison of machine learning and electrical resistivity arrays to inverse modeling for locating and characterizing subsurface targets. *Journal of Applied Geophysics*, *229*, 105493. https://doi.org/10.1016/j.jappgeo.2024.105493

Jiang, Y., Ma, J., Ning, J., Li, J., Wu, H., & Bao, T. (2025). One-fit-all transformer for multimodal geophysical inversion: Method and application. *Journal of Geophysical Research: Machine Learning and Computation*, *2*(1), e2024JH000432. https://doi.org/10.1029/2024jh000432

Kasing, M., Jirasek, J., Matysek, D., Melnyk, A., Jirman, P., Skupien, P., & Pospisil, L. (2023). Geophysical architecture of tectonized volcanic bodies near flysch thrust fronts (outer Western Carpathians). *Journal of Applied Geophysics*, *213*, 105021. https://doi.org/10.1016/j.jappgeo.2023.105021

Kong, S., Oh, J., Yoon, D., Ryu, D.-W., & Kwon, H.-S. (2023). Integrating deep learning and deterministic inversion for enhancing fault detection in electrical resistivity surveys. *Applied Sciences-Basel*, *13*(10), 6250. https://doi.org/10.3390/app13106250

Lan, X., Zou, C., Peng, C., & Wu, C. (2023). Uncertainty quantification in intelligent-based electrical resistivity tomography image reconstruction with Monte Carlo dropout strategy. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–16. https://doi.org/10.1109/tgrs.2023.3262835

Linde, N., Binley, A., Tryggvason, A., Pedersen, L. B., & Revil, A. (2006). Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data. *Water Resources Research*, *42*(12), W12404. https://doi.org/10.1029/2006WR005131

Liu, B., Wang, C., Liu, Z., Xu, Z., Nie, L., Pang, Y., et al. (2021). Cascade surface and borehole geophysical investigation for water leakage: A case study of the Dehou reservoir, China. *Engineering Geology*, *294*, 106364. https://doi.org/10.1016/j.enggeo.2021.106364

Liu, P., Liu, S., Chen, C., Hong, T., Xiao, Y., & He, S. (2024). Assessment of emergency water sources using electrical resistivity tomography: A case study in the Longmen Shan Fault Zone. *Water*, *16*(20), 2967. https://doi.org/10.3390/w16202967

Liu, X., Lu, Q., & Liu, S. (2023). Resolution enhancement of electrical resistivity tomography based on deep learning. *IEEE Geoscience and Remote Sensing Letters*, *20*, 1–5. https://doi.org/10.1109/lgrs.2023.3320698

Liu, Y., Zou, C., Chen, Q., Zhao, J., & Wu, C. (2022). Optimization of critical parameters of deep learning for electrical resistivity tomography to identifying hydrate. *Energies*, *15*(13), 4765. https://doi.org/10.3390/en15134765

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, *17*(4), 818. https://doi.org/10.3390/s17040818

Manrique, I. I., Caterina, D., Nguyen, F., & Hermans, T. (2023). Quantitative interpretation of geoelectric inverted data with a robust probabilistic approach. *Geophysics*, *88*(3), B151–B166. https://doi.org/10.1190/geo2022-0133.1

Mauricio, J., Domingues, I., & Bernardino, J. (2023). Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Applied Sciences-Basel*, *13*(9), 5521. https://doi.org/10.3390/app13095521

Mudunuru, M. K., Cromwell, E. L. D., Wang, H., & Chen, X. (2022). Deep learning to estimate permeability using geophysical data. *Advances in Water Resources*, *167*, 104272. https://doi.org/10.1016/j.advwatres.2022.104272

Nie, L., Song, Z., Li, Z., & Zhang, S. (2024). Ground-tunnel geological prospecting and treatment methods of small-diameter TBM crossing shallow buried water-rich tunnel: A case study. *Tunnelling and Underground Space Technology*, *153*, 106024. https://doi.org/10.1016/j.tust.2024.106024

Oladeji, E., Parsekian, A., & Grana, D. (2024). Machine learning-based hydrofacies classification: Effects of noise and regularization. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000197. https://doi.org/10.1029/2024jh000197

Pang, C., Gao, W., Wu, P., & Wang, L. (2022). Grouting effect detection within the floor of a coal seam using 3D electric resistivity tomography (ERT) with arbitrary electrode positions. *Applied Sciences-Basel*, *12*(11), 5625. https://doi.org/10.3390/app12115625

Papadopoulos, N., Oikonomou, D., Simyrdanis, K., & Heng, L. M. (2022). Practical considerations for shallow submerged archaeological prospection with 3-D electrical resistivity tomography. *Archaeological Prospection*, *29*(1), 103–123. https://doi.org/10.1002/arp.1841

Perrone, A., Lapenna, V., & Piscitelli, S. (2014). Electrical resistivity tomography technique for landslide investigation: A review. *Earth-Science Reviews*, *135*, 65–82. https://doi.org/10.1016/j.earscirev.2014.04.002

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (2021). Do vision transformers see like convolutional neural networks? In *Proceedings of the 35th annual conference on neural information processing systems (NeurIPS)*. Online.

Rittgers, J. B., Revil, A., Mooney, M. A., Karaoulis, M., Wodajo, L., & Hickey, C. (2016). Time-lapse joint inversion of geophysical data with automatic joint constraints and dynamic attributes. *Geophysical Journal International*, *207*(3), 1401–1419. https://doi.org/10.1093/gji/ggw346

Robinson, D., Jones, S., Wraith, J., Or, D., & Friedman, S. (2003). A review of advances in dielectric and electrical conductivity measurement in soils using time domain reflectometry. *Vadose Zone Journal*, *2*(4), 444–475. https://doi.org/10.2136/vzj2003.4440

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), New Orleans, LA*.

Roy, S. K., Deria, A., Shah, C., Haut, J. M., Du, Q., & Plaza, A. (2023). Spectral-spatial morphological attention transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–15. https://doi.org/10.1109/tgrs.2023.3242346

Rucker, D. F., Crook, N., Glaser, D., & Loke, M. H. (2012). Pilot-scale field validation of the long electrode electrical resistivity tomography method. *Geophysical Prospecting*, *60*(6), 1150–1166. https://doi.org/10.1111/j.1365-2478.2011.01039.x

Rucker, C., Guenther, T., & Wagner, F. M. (2017). pyGIMLi: An open-source library for modelling and inversion in geophysics. *Computers & Geosciences*, *109*, 106–123. https://doi.org/10.1016/j.cageo.2017.07.011

Rupesh, R., Tiwari, P., & Sharma, S. P. (2024). Estimation of geotechnical parameters for coal exploration from quasi-3D electrical resistivity measurements. *Minerals*, *14*(1), 102. https://doi.org/10.3390/min14010102

Saad, O. M., Harsuko, R., & Alkhalifah, T. (2024). SiameseFWI: A deep learning network for enhanced full waveform inversion. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000227. https://doi.org/10.1029/2024jh000227

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., & Norouzi, M. (2023). Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(4), 4713–4726. https://doi.org/10.1109/tpami.2022.3204461

Sakar, C., Schwartz, N., & Moreno, Z. (2024). Physics-informed neural networks trained with time-lapse geo-electrical tomograms to estimate water saturation, permeability and petrophysical relations at heterogeneous soils. *Water Resources Research*, *60*(8), e2024WR037672. https://doi.org/10.1029/2024wr037672

Shi, Y., Zhang, J., You, X., Ma, Z., & Li, J. (2024). Transient electromagnetic inversion to image the shallow subsurface based on convolutional bidirectional long short-term memory neural networks. *Geophysical Journal International*, *239*(1), 173–191. https://doi.org/10.1093/gji/ggae253

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Su, Z., Revil, A., Ghorbani, A., Zhang, X., Zhao, X., & Richard, J. (2024). Combining electrical resistivity, induced polarization, and self-potential for a better detection of ore bodies. *Minerals*, *14*(1), 12. https://doi.org/10.3390/min14010012

Tao, M., Chen, X., Cheng, Q., & Binley, A. (2022). Evaluating the joint use of GPR and ERT on mapping shallow subsurface features of karst critical zone in southwest China. *Vadose Zone Journal*, *21*(1), e20172. https://doi.org/10.1002/vzj2.20172

Taufik, M. H., Wang, F., & Alkhalifah, T. (2024). Learned regularizations for multi-parameter elastic full waveform inversion using diffusion models. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(1), e2024JH000125. https://doi.org/10.1029/2024jh000125

Thirard, G., Thiery, Y., Gourdier, S., Grandjean, G., Maquaire, O., Francois, B., et al. (2022). Hydromechanical assessment of a complex landslide through geophysics and numerical modeling: Toward an upgrade for the Villerville landslide (Normandy, France). *Engineering Geology*, *297*, 106516. https://doi.org/10.1016/j.enggeo.2022.106516

Trappolini, D., Laurenti, L., Poggiali, G., Tinti, E., Galasso, F., Michelini, A., & Marone, C. (2024). Cold diffusion model for seismic denoising. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(2), e2024JH000179. https://doi.org/10.1029/2024jh000179

Wang, B., Wang, W., Jin, F., Tan, H., Liu, N., & Huang, D. (2024). Exploration on electrical resistance tomography in characterizing the slurry spatial distribution in cemented granular materials. *Frontiers of Structural and Civil Engineering*, *18*(3), 365–379. https://doi.org/10.1007/s11709-024-1049-3

Wang, F., Huang, X., & Alkhalifah, T. (2024). Controllable seismic velocity synthesis using generative diffusion models. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000153. https://doi.org/10.1029/2024jh000153

Xia, M., Zhou, Y., Yi, R., Liu, Y.-J., & Wang, W. (2024). A diffusion model translator for efficient image-to-image translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(12), 10272–10283. https://doi.org/10.1109/tpami.2024.3435448

Xiao, H., Li, L., Liu, Q., Zhu, X., & Zhang, Q. (2023). Transformers in medical image segmentation: A review. *Biomedical Signal Processing and Control*, *84*, 104791. https://doi.org/10.1016/j.bspc.2023.104791

Yan, C., Zhang, G., Chen, Y., Huang, S., Zhao, Y., & Wang, J. (2023). One-dimensional structure reparameterized convolutional neural network for two-phase image reconstruction based on ERT. *Measurement Science and Technology*, *34*(10), 105402. https://doi.org/10.1088/1361-6501/ace2df

Yang, Y., Saad, O. M., & Alkhalifah, T. (2024). Deep learning-based 3D microseismic event direct location using simultaneous surface and borehole data. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(4), e2024JH000365. https://doi.org/10.1029/2024jh000365

Yin, H., Carroll, K. C., Yuan, Y., Ahsan, J., Rucker, D. F., Dai, Z., & Mohamad, R. S. (2025a). Synthetic ERT forward modeling dataset for deep learning inversion [Dataset]. *Figshare*. https://doi.org/10.6084/m9.figshare.28746350.v1

Yin, H., Carroll, K. C., Yuan, Y., Ahsan, J., Rucker, D. F., Dai, Z., & Mohamad, R. S. (2025b). Model implementation, trained weights, and inversion outputs for ViT-based ERT inversion [Software and Data]. *Figshare*. https://doi.org/10.6084/m9.figshare.28746353.v1

Yin, H., Zhang, G., Wu, Q., Cui, F., Yan, B., Yin, S., et al. (2024a). Transfer learning with transformer-based models for mine water inrush prediction: A multivariate analysis using sparse and imbalanced monitoring data. *Mine Water and the Environment*, *43*(4), 707–726. https://doi.org/10.1007/s10230-024-01011-2

Yin, H., Zhang, G., Wu, Q., Cui, F., Yan, B., Yin, S., et al. (2024b). Unraveling overlying rock fracturing evolvement for mining water inflow channel prediction: A spatiotemporal analysis using ConvLSTM image reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–17. https://doi.org/10.1109/tgrs.2024.3452937

Zhan, C., Dai, Z., Jiao, J. J., Soltanian, M. R., Yin, H., & Carroll, K. C. (2025). Toward artificial general intelligence in hydrogeological modeling with an integrated latent diffusion framework. *Geophysical Research Letters*, *52*(3), e2024GL114298. https://doi.org/10.1029/2024gl114298

Zhang, H., Yang, Y.-J., & Zeng, W. (2024). Towards semantically continuous unpaired image-to-image translation via margin adaptive contrastive learning and wavelet transform. *Expert Systems with Applications*, *252*, 124132. https://doi.org/10.1016/j.eswa.2024.124132

Zhao, M., Jia, M., Yang, L., Suo, K., Lu, Y., Wei, Q., & Zhang, Y. (2024). Correction methods and applications of ERT in complex terrain. *MethodsX*, *13*, 103012. https://doi.org/10.1016/j.mex.2024.103012

Zu, S., Zhao, P., Ke, C., & Junxing, C. (2024). ResACEUnet: An improved transformer unet model for 3D seismic fault detection. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000232. https://doi.org/10.1029/2024jh000232