(Received 17 April 2025, Accepted 29 July 2025, Scheduled 6 August 2025)

A Physics-Assisted Learning Method Based on the Improved U-Net for Reconstructing 2-D Dielectric Objects

Zhangyue Zhao and Chun Xia Yang*

Department of Information and Communication Engineering, Shanghai Normal University, Shanghai, China

ABSTRACT: In the past few years, deep learning has emerged as a transformative force in tackling challenges within the realm of electromagnetic inverse scattering, driving remarkable advances and reshaping conventional approaches. Among them, the physics-assisted learning method that combines traditional inverse scattering algorithms with deep neural networks has demonstrated excellent real-time inversion capability and lower computational complexity. For two-dimensional inverse scattering problems, an approximate solution of the target is first obtained using a linear approximation algorithm, followed by mapping learning from low to high precision with a neural network. To enhance both precision and generalizability, this study integrates a transformer module into a CBAM U-Net framework, giving rise to a refined architecture aptly named TransAtten U-Net. By retaining certain positional information while enhancing the correlations between features, the overall feature extraction effect is improved. Through simulation experiments, the paper compares the performance of the proposed TransAtten U-Net two-step method, TransAtten U-Net direct method, and CBAM U-Net two-step method. Experimental results demonstrate that the proposed TransAtten U-Net two-step method not only achieves higher accuracy than the other two approaches, but also exhibits a stronger generalization capability across diverse scenarios, along with enabling real-time imaging.

1. INTRODUCTION

Electromagnetic inverse scattering conundrum stands as a pivotal challenge in the realm of electromagnetic field inversion methodologies, with applications spanning radar detection, medical imaging, non-destructive evaluation, and numerous other domains. The fundamental objective of this problem is to infer the position, geometry, and physical properties of a target object from the known scattered data. In electromagnetic inverse scattering, contrast typically refers to the difference in relative permittivity between the target object and surrounding medium. By measuring the scattered data, the contrast distribution can be estimated, which allows for the deduction of the target object's shape, physical properties, and location.

The inverse scattering problem lies in the challenge of recovering the signal from limited data [1], with issues such as "false solutions" and solution multiplicity often arise, especially when dealing with large-scale sources or the presence of noise. In practical applications, the scattered data is usually incomplete and contaminated by noise. Furthermore, the problem is ill-posed, meaning that different target objects may produce similar scattered data, leading to uncertainty in the inversion results. Due to the profoundly nonlinear and intrinsically ill-posed nature of electromagnetic inverse scattering, the problem is typically recast as an optimization challenge, thus rendering it more manageable [2]. By constructing a cost function, the original nonlinear problem is converted into an optimization problem, which is then solved through multiple iterations. While optimization algorithms offer high accuracy, they tend to be computationally inefficient. Therefore, this paper focuses on exploring deep learning methods that improve computational accuracy while ensuring real-time imaging.

Li et al. provided a comprehensive review of machine learning applications in electromagnetics, highlighting its potential in biomedical imaging where the reconstruction process integrates data, measurement physics, and prior knowledge [3]. Direct inversion methods based on deep learning solve the contrast directly from scattered data [4]. However, these methods [5] can only handle simple problems and lack sufficient accuracy [6]. Physics-assisted learning methods, on the other hand, combine domain knowledge with mathematical models [7–9], embedding this knowledge into the input or internal architecture of neural networks, simplifying the inversion process [10] and allowing effective contrast estimation. Guo et al. proposed a physics-embedded deep neural network to solve full-wave inverse scattering problems, achieving high-accuracy and super-resolution reconstructions by integrating physical modeling into the deep learning framework [11]. Many studies [12] focus on approximating the inversion using domain knowledge estimating the contrast first and then using this estimate as input to a neural network to obtain a more accurate contrast. This strategy streamlines the learning process by redirecting the input space—from raw measurement data to the more structured contrast domain—thus easing the model's interpretive burden. Both the input and output are contrast distribution maps, making it possible to directly apply existing image processing neural network techniques [13, 14], such as U-Net [15].

Through the intricate interlacing of skip connections between its input and output layers, U-Net architecture demonstrates remarkable proficiency in image reconstruction tasks, especially when there exists a strong resemblance between the input and

* Corresponding author: Chun Xia Yang (chunxiay@shnu.edu.cn).

output images [16]. U-Net is a general image-to-image architecture, where its encoder-decoder structure with skip connections plays a critical role in fusing features across different scales. Rather than focusing solely on recovering high-frequency components, U-Net's goal is to capture and refine detailed information by effectively integrating multi-scale features through skip connections. In the context of inverse scattering, U-Net has been applied in various studies to tackle electromagnetic inverse scattering problems. As an illustration, in [17, 18], a U-Net framework is employed to ingest low-resolution approximations and produce high-resolution contrast mappings.

CBAM U-Net, which integrates the Convolutional Block Attention Module (CBAM) into the traditional U-Net architecture, is an advanced image segmentation method [19]. CBAM U-Net has shown significant advantages in improving segmentation accuracy, enhancing robustness, and increasing training efficiency, particularly for tasks with complex backgrounds or unclear target boundaries. It has a wide range of applications and can address the shortcomings of traditional U-Net models in dealing with noisy or blurred targets. The CBAM U-Net model has been successfully applied to many practical engineering problems, especially in medical image segmentation, although the research on microwave imaging is still in its early stages. In this study, an advanced CBAM U-Net framework is used to cope with the complexity brought by the electromagnetic inverse scattering problem.

This paper introduces Transformer module [20] into CBAM U-Net model structure and applies it to electromagnetic inverse scattering problem. First, the feature map is flattened into a sequence, and then Transformer module uses a multi-head self-attention mechanism [21] to effectively capture the mutual relationship of contrast at each position and extract global features. This approach helps to more finely represent the interdependence between local features and overall information, thereby improving the accuracy of contrast.

This paper is organized as follows. This section lays the foundation by explaining the background, research motivation, and the importance of using deep learning methods to solve the electromagnetic inverse scattering problem. In addition, it also outlines the research objectives and key contributions of this study. Section 2 presents the imaging model and related equations for 2D electromagnetic inverse scattering problem. Section 3 describes the working principle of CBAM U-Net and its custom improvements, along with the structure of TransAtten U-Net model. In Section 4, a series of simulation experiments aimed at tackling inverse scattering problems using the MNIST dataset [22] are illustrated, accompanied by trials for generalization assessment. Section 5 encapsulates the key research outcomes and offers a perspective on potential avenues for future exploration.

2. ELECTROMAGNETIC INVERSE SCATTERING PROBLEM

This work delves into two-dimensional electromagnetic inverse scattering conundrum, with the objective of extracting details regarding the location, form, and material characteristics of the scatterer from the scattered data. Fig. 1 illustrates the physical model underlying the inverse scattering problem. The scatterer is located in region D, referred to as the Domain of Interest (DOI), and is illuminated by incident waves generated by transmitters. In the forward problem, the properties of the scatterer are known, and the objective is to compute the scattered field. On the other hand, in the case of the inverse scattering problem [23], the scattered field is captured at the receivers for each illumination, with the primary objective being to deduce the scatterer's location, shape, and physical characteristics based on the measured scattered fields.

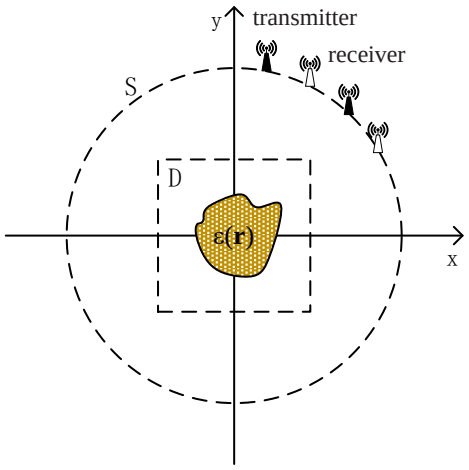

FIGURE 1. Physical model of the electromagnetic inverse scattering problem. The scatterer is located in the domain D (Domain of Interest, DOI), and is characterized by a relative permittivity ϵ_r. Transmitters and receivers are positioned along the measurement surface S outside the DOI, where the scattered fields are recorded.

Assume that there is a background medium with uniform properties in the DOI in free space, with a relative permittivity ϵ_0 and permeability μ_0. Inside the DOI, there exists an unknown non-magnetic dielectric scatterer with a relative permittivity ϵ_r and permeability μ. Several transmitters and receivers are located on the surface S, which is outside the DOI. Each transmitter emits transverse magnetic (TM) waves along the z-axis, while all the receivers simultaneously record the scattered data. When electromagnetic wave propagates and encounters the scatterer, a scattered field is generated. The total field recorded by the receivers is the combined result of both the scattered field and incident field. The forward problem is described by a source-type integral equation [2]:

$$\boldsymbol{E}^{\boldsymbol{t}}(\mathbf{r}) = \boldsymbol{E}^{\boldsymbol{i}}(\mathbf{r}) + k_0^2 \int_D g(\mathbf{r},\mathbf{r}')O(\mathbf{r}')\boldsymbol{E}^{\boldsymbol{t}}(\mathbf{r}')d\mathbf{r}', \quad \mathbf{r} \in D \tag{1}$$

$$\boldsymbol{E}^{\boldsymbol{s}}(\mathbf{r}) = k_0^2 \int_D g(\mathbf{r},\mathbf{r}')O(\mathbf{r}')\boldsymbol{E}^{\boldsymbol{t}}(\mathbf{r}')d\mathbf{r}', \quad \mathbf{r} \in S \tag{2}$$

where $\boldsymbol{E}^{\boldsymbol{t}}(\mathbf{r})$, $\boldsymbol{E}^{\boldsymbol{i}}(\mathbf{r})$, and $\boldsymbol{E}^{\boldsymbol{s}}(\mathbf{r})$ represent the total field, incident field, and scattered field, respectively. $g(\mathbf{r},\mathbf{r}')$ is the two-dimensional Green's function in free space, and k_0 represents the wave number in the background medium. $O(\mathbf{r}) = \epsilon_r(\mathbf{r}) - 1$ is the contrast in relative permittivity.

3. TRANSATTEN U-NET TWO-STEP METHOD

This section introduces the proposed TransAtten U-Net two-step method, which combines Diffraction Tomography (DT) imaging algorithm with an improved U-Net model. The aim is to optimize the contrast distribution in electromagnetic inverse scattering problem. By incorporating the Transformer module, TransAtten U-Net enhances the global feature modeling capability on top of the traditional U-Net, improving both inversion accuracy and computational efficiency. The proposed method consists of the following two steps.

3.1. Step 1: Contrast Estimation via DT

To fully leverage the known physical laws, this paper first employs DT algorithm to invert the contrast distribution. DT algorithm, a method rooted in linear approximation and derived from the Fourier diffraction projection theorem, proves effective for scatterers exhibiting low contrast. Leveraging the Born approximation, this algorithm finds its application primarily in low-frequency domains. It linearly approximates the variation in the scattered field amplitude with respect to the material properties of the target scatterer.

Building on the problem model outlined in Section 2, it is important to note that the scattered field is significantly weaker than the incident field. As a result, within the region $\mathbf{r} \in D$ inside the scatterer, the total field $\boldsymbol{E^t}(\mathbf{r})$ can be effectively approximated by the incident field $\boldsymbol{E^i}(\mathbf{r})$. Thus, Eq. (2) can be approximated as:

$$\boldsymbol{E^s}(\mathbf{r}) \approx k_0^2 \int_D g(\mathbf{r}, \mathbf{r'}) O(\mathbf{r'}) \boldsymbol{E^i}(\mathbf{r'})\, d\mathbf{r'},\ \mathbf{r} \in S \tag{3}$$

Here, the position vectors of the transmitter and receiver in the two-dimensional space are denoted as $\boldsymbol{\rho_T}$ and $\boldsymbol{\rho_R}$, respectively. Under the aforementioned approximation conditions, the scattering field computation equation can be derived as follows [2]:

$$\boldsymbol{E^s}(\boldsymbol{\rho_R}) \approx \frac{i}{8\pi k_0 \sqrt{\rho_R \rho_T}} e^{ik_0(\rho_T+\rho_R)} \tilde{O}(\mathbf{k_R} - \mathbf{k_T}) \tag{4}$$

In Eq. (4), $\tilde{O}$ is the Fourier transform of O, and $\mathbf{k_R} = k_0 \hat{\boldsymbol{\rho_R}}$, $\mathbf{k_T} = -k_0 \hat{\boldsymbol{\rho_T}}$. According to Eq. (4), the contrast of the target scatterer can be related to the scattered field through Fourier transform. After measuring the scattered field, the corresponding parameter transformation can be performed to obtain the data in the Fourier space corresponding to the contrast in the imaging region. The result can then be quickly obtained through the inverse Fourier transform. This method transforms the original complex integral equation into a form suitable for Fourier transform, significantly improving computational efficiency and making it advantageous for real-time imaging.

Using the DT imaging algorithm described above, an approximate solution for the contrast in the imaging region can be obtained. However, since this is a linear approximation method, the inversion accuracy of the contrast may not be satisfactory. Therefore, further accuracy improvements are achieved by applying U-Net on top of this algorithm.

3.2. Step 2: Optimizing Initial Contrast through TransAtten U-Net

In this investigation, a refined version of U-Net is employed to regenerate the contrast, thereby yielding a contrast distribution of superior precision. Transformer module was introduced into CBAM U-Net, resulting in the formation of TransAtten U-Net. Compared with CBAM U-Net, TransAtten U-Net enhances the global feature modeling capability and performs better in complex image tasks.

U-Net consists of encoder, decoder, and skip connections, and was originally designed for image segmentation. CBAM U-Net is an evolution of the traditional U-Net, which introduces convolutional block attention module (CBAM) to enhance the feature extraction process. CBAM uses channel and spatial attention techniques to enable the network to focus on the most relevant details in the image. In the context of electromagnetic inverse scatter imaging, CBAM U-Net transforms low-precision contrast distribution into high-precision distribution.

The architecture of TransAtten U-Net is shown in Fig. 2, which illustrates the dimensionality transformation of feature maps at each layer. The network contains four core components: encoder, bridge, decoder, and skip connections. The encoder is responsible for the initial extraction of features from the input image. The bridge acts as the link between the encoder and decoder, introducing a Transformer module to capture global self-attention within the image. The decoder progressively upsamples the features extracted by the encoder, merging local features from the encoder with global features from the decoder through skip connections, thereby reconstructing the image's high-resolution details. Finally, the image's contrast and finer details are enhanced through the optimization of the loss function. A detailed breakdown of the TransAtten U-Net's structure and the specific improvements made in this paper follows.

3.2.1. Encoder

The encoder comprises a series of convolutional layers, CBAM, and pooling layers, which collaboratively serve to progressively distill features from the image, all the while diminishing its spatial resolution. As the data traverses through the convolutional layers, the feature map's channel count swells, allowing the network to seize progressively more intricate features. CBAM focuses on the most critical information in the image and applies weighting. Pooling operations gradually reduce the spatial resolution.

The detailed computational sequence for the encoder is described as follows: 1) The input is the DT result, and the image size is $1 \times N \times N$. 2) Subsequent to two 3×3 convolution operations, the feature map undergoes an expansion in channel count, rising to 32, while its dimensions are adjusted to $32 \times N \times N$. After each convolution, Batch Normalization (BN) is performed, followed by a rectified linear unit (ReLU) activation function. 3) The feature map is passed through the CBAM module to compute the channel and spatial attention outputs, which are then weighted, yielding the weighted feature map. This feature map is also passed to the decoder section via skip connections. 4) The feature map is subjected to a

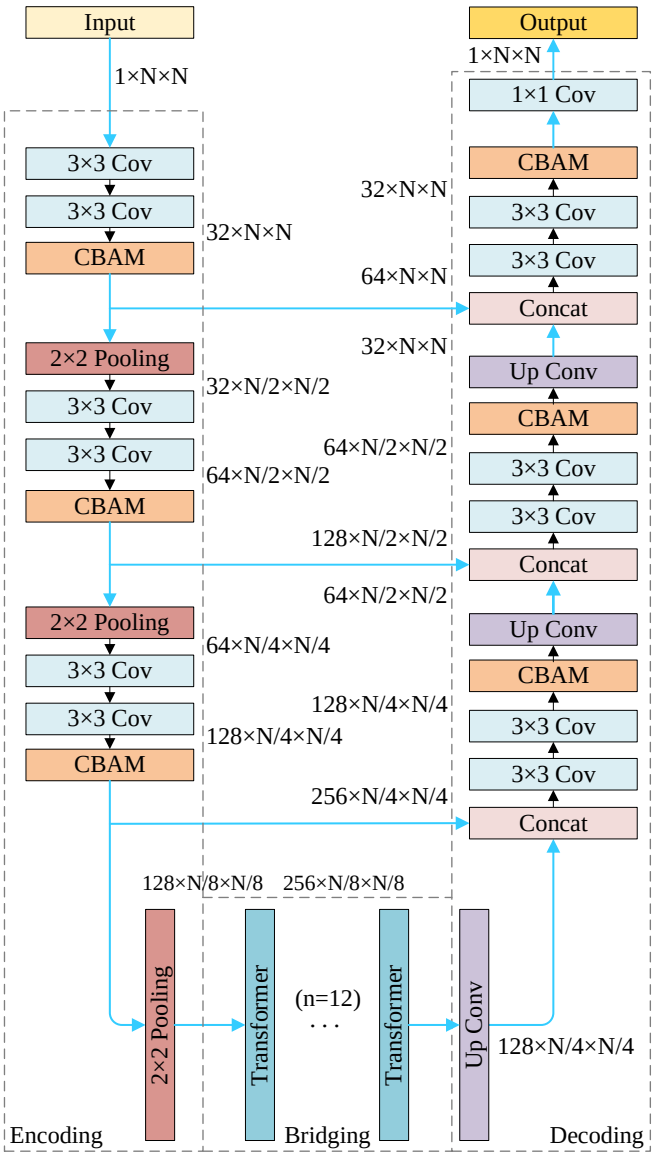

FIGURE 2. TransAtten U-Net Network Structure. The diagram illustrates the architecture of the TransAtten U-Net, which includes the input, encoder, bridge, and decoder components. The network progressively transforms feature maps at each layer, with each feature map's dimensions noted in the format: channels × height × width, and labeled on the figure. Here, N denotes the height (and width) of the input image. Key operations include: 3×3 Convolution (Cov), CBAM (Convolutional Block Attention Module), 2×2 Pooling, Transformer (with 12 layers), Concat (Concatenation), Up Conv (Up-sampling Convolution), and 1×1 Convolution.

2×2 max-pooling operation, which compresses its dimensions to $32 \times N/2 \times N/2$. 5) Steps 2, 3, and 4 are repeated, with the feature map size changing to $64 \times N/4 \times N/4$. 6) The process is repeated again, with the feature map size changing to $128 \times N/8 \times N/8$. Finally, it connects to the bridge section.

3.2.2. Bridge Section

In the bridge part, we replace the convolutional layers in the classic U-Net with a 12-layer Transformer module, as shown in Fig. 3. The main purpose of this part is to provide rich global information to the decoder. The Transformer module uses a multi-head self-attention mechanism to grasp the long-range correlation of different regions on the feature map, effectively overcoming the limitation that traditional convolution operations can only process local information. Each layer of Transformer has a multi-head self-attention mechanism, which enables the model to focus on different areas of the image at the same time, enhancing the diversity and richness of feature representation, thereby having a more detailed understanding of the image.

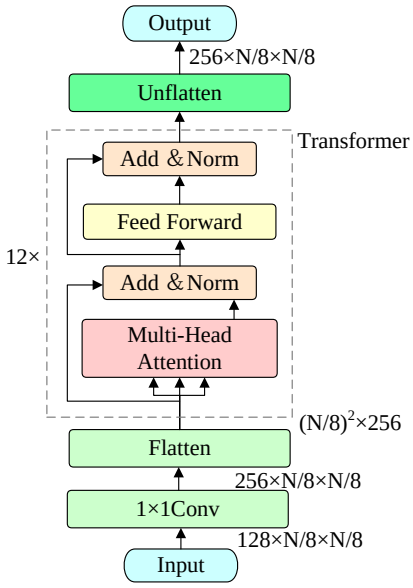

FIGURE 3. Structure of the Bridge Section. The diagram illustrates the structure of the bridge section, including the input, output, and key operations. The dimensions of feature maps at each stage are indicated in the format: channels × height × width, and labeled on the figure. Here, N denotes the height (and width) of the input image to the U-Net. After the Flatten operation, the feature map is reshaped into a sequence of size sequence length × feature dimension = $(N/8)^2 \times 256$. Key operations include: 1×1 Convolution (1×1 Conv), Multi-Head Attention, Transformer (12 layers), Feed Forward, Add & Norm (Add and Normalize), Flatten, and Unflatten.

The multi-head self-attention mechanism works by evaluating the mutual relationship between each element in the input sequence and every other element, and then merging them through weighted summation to produce a new representation. This process includes steps such as querying, typing, mapping, calculating attention scores, and performing weighted summation. In multi-head self-attention, multiple heads work in parallel, and each head focuses on different subspace information, further enhancing the expressiveness of the model. After residual connection and normalization, the final output can be passed to subsequent layers. The advantage of multi-head self-attention is that it can handle long-distance dependencies, which makes it very valuable in applications in the image and text fields.

The detailed calculation sequence of the bridge section is as follows: 1) 1×1 convolution is used for channel expansion, and the feature map size becomes $256 \times N/8 \times N/8$. 2) The

width and height dimensions of the feature map are collapsed into a single one-dimensional sequence, resulting in a sequence of size $(N/8)^2 \times 256$. 3) The expanded feature sequence is processed by the self-attention mechanism, which can capture global contextual dependencies and enhance the depth of feature representation. 4) The output features are weighted and combined according to the information at different positions in the input image, and residual connections and normalization operations are applied to ensure stable gradient propagation. 5) The feature sequence is further processed by a feedforward neural network, and residual connections and normalization are applied to adjust the output for training stability. 6) The processed feature sequence is reshaped to the original feature map dimension ($256 \times N/8 \times N/8$) and finally passed to the decoder part.

3.2.3. Decoder

The decoder gradually rebuilds the image's spatial resolution through upsampling operations, while simultaneously utilizing skip connections to integrate corresponding features from the encoder. In the concluding phase, a convolutional layer is utilized to convert the extracted features into the resultant contrast distribution map.

The decoder is described in detail according to the computation sequence: 1) Through a deconvolution operation, the feature map size is changed to $128 \times N/4 \times N/4$. 2) The feature map procured in step 1 is fused with its counterpart from the encoder along the channel axis, producing a feature map of dimensions $256 \times N/4 \times N/4$. 3) After two 3×3 convolution operations, the feature map size is reduced to $128 \times N/4 \times N/4$. 4) The feature map is modulated through the CBAM (Convolutional Block Attention Module), resulting in a reweighted feature map. 5) Steps 1, 2, 3, and 4 are repeated, and the feature map size becomes $64 \times N/2 \times N/2$. 6) Steps 1, 2, 3, and 4 are repeated again, and the feature map size becomes $32 \times N \times N$. 7) A 1×1 convolutional operation is employed to transform the feature map into the ultimate contrast distribution map, sized $1 \times N \times N$, with the output being produced as the final result.

3.2.4. Advantages of TransAtten U-Net

By means of the aforementioned procedure, TransAtten U-Net adeptly merges the Convolutional Block Attention Module (CBAM) with the Transformer module, thereby refining the contrast distribution map for electromagnetic inverse scattering imaging tasks. The model significantly enhances the accuracy and details of the contrast by performing multi-level feature extraction, global self-attention modeling, and gradually restoring the image resolution. Notably, the introduction of the Transformer module strengthens the network's ability to model global dependencies, which is challenging to achieve with traditional convolutional neural networks.

4. NUMERICAL EXPERIMENTS

To evaluate the model, the DT algorithm and neural network model are trained and tested using simulated generated scattering data. TransAtten U-Net direct method is derived from the proposed method by removing the Physics-Assisted component and directly using the scattering data as input. CBAM U-Net two-step method is also based on the proposed method, with TransAtten U-Net replaced by CBAM U-Net. This section compares the performance and efficiency of the proposed TransAtten U-Net two-step method, TransAtten U-Net direct method, and CBAM U-Net two-step method, emphasizing real-time performance and the higher accuracy exhibited by the proposed method.

4.1. Training the Network with the MNIST Dataset

In this section, electromagnetic simulation experiments are presented, derived from the problem model outlined in Section 2. To simulate the electromagnetic wave propagation process, the parameters for the transmitting electromagnetic wave are first set, and the positions of the transmitter and receiver are determined. The transmitter sequentially emits electromagnetic waves, while the receiver collects the reflected electromagnetic waves from the scatterer. The received echo data contains the scattering characteristics of the target object, which will be used for subsequent imaging. The scattering data obtained from this simulation process provides real and effective training data for DT algorithm and neural network models.

The experimental parameters for this section are set as follows. The incident wave frequency is 300 MHz, corresponding to a wavelength (λ) of 1 meter. A square region of interest (DOI) with a side length of 2λ is positioned in free space, where the background contrast is set to 0. The centroid of the DOI is positioned at the origin of the coordinate framework. The DOI is uniformly subdivided into a 28×28 grid, with the assumption that the contrast within each grid cell remains uniform. To assess the method's performance, the standard MNIST dataset is utilized. Sixteen transmitters and an equal number of receivers are arranged symmetrically along the circumference of a circle, with a radius of 10λ, its center coinciding with the origin of the coordinate system.

Throughout the training process, 5000 samples are randomly selected from the MNIST training set to simulate scatterers in the shape of digits, with the contrast randomly set between 0.2 and 1.5. The DT results are utilized as input to the TransAtten U-Net, functioning as a preliminary approximation of the contrast. The goal of the network is to predict the contrast distribution and learn the fine details of reconstructing the contrast image by comparing it with the real contrast distribution. The model parameters are optimized through backpropagation to minimize the error. During the testing phase, the DT algorithm first performs a preliminary contrast inversion on the new input image to obtain a low-accuracy contrast distribution. The low-resolution DT result is subsequently passed into the trained TransAtten U-Net model, where it undergoes refined contrast reconstruction. The reconstructed results are analyzed both qualitatively and quantitatively, and compared with other methods to evaluate the reconstruction performance.

During the entire training phase, mean square error (MSE) is used as the loss function to refine the network to ensure that the generated contrast distribution map is close to the real scene. In order to improve training efficiency and accelerate conver-

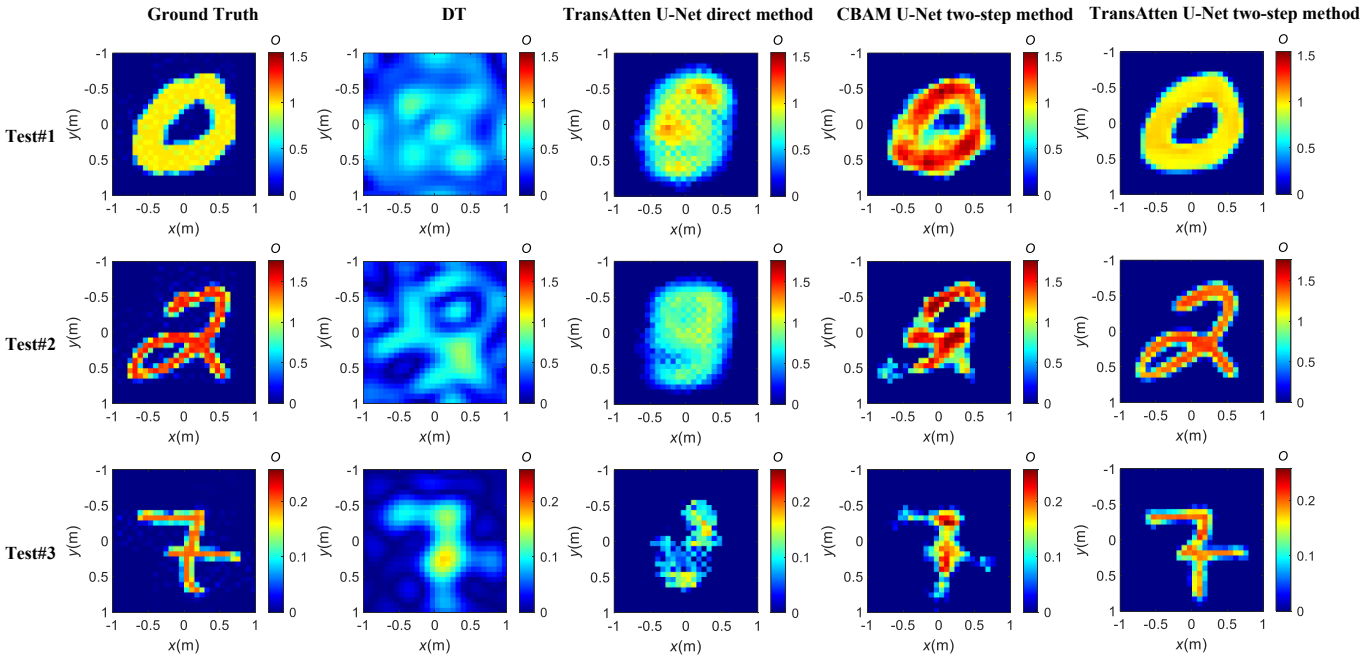

FIGURE 4. Reconstructed results for MNIST test samples within the contrast range of the training set. O denotes the contrast.

gence, a small batch gradient descent strategy is adopted to reduce the amount of calculation and avoid local optimality. The combination of momentum optimization and learning rate decay strategy further smoothes the optimization process and ensures convergence to the global optimum. Through these methods, TransAtten U-Net not only ensures stable training, but also can generate high-precision contrast distribution maps, enhancing the adaptability and generalization ability of the model in complex environments.

4.2. Testing the Network with the MNIST Dataset

A random selection of 400 samples from the MNIST test set is used for testing, with contrast values ranging from 0.2 to 1.5, consistent with the contrast range of the training set. Fig. 4 unveils the authentic and reconstructed contrast distributions for the test samples, juxtaposing the outcomes from the DT algorithm, TransAtten U-Net direct method, CBAM U-Net two-step method, and the proposed method. The initial row of Fig. 4 presents samples with a contrast value of 1; the second row displays samples with a contrast of 1.5; and the third row illustrates samples with a contrast of 0.2. The reconstruction images for each sample are displayed with reference to the color scale of the target image. Compared with the other methods, the proposed method provides more accurate reconstructed images both for high and low contrast situations. Especially in the high-contrast scenario, the inversion result from the DT algorithm shows significant deviation from the true distribution, while TransAtten U-Net significantly reduces this difference.

Mean Squared Error (MSE) measures the mean of the squared deviations in pixel intensities between two images, where a lesser value signals a closer resemblance. In contrast, Structural Similarity Index Measure (SSIM) takes into account the human visual system, assessing the similarity in brightness, contrast, and structure; a larger SSIM value signifies a greater degree of similarity. To objectively evaluate the quality of the reconstructed results, normalized MSE and SSIM are employed as performance metrics. Table 1 presents the MSE and SSIM values for the four methods across 400 test samples. The MSE for DT algorithm is 0.0811, for TransAtten U-Net direct method is 0.0644, for CBAM U-Net two-step method is 0.0205, and for the proposed method is 0.0078, which is significantly lower than the other methods. The SSIM for DT algorithm is 0.1747, for TransAtten U-Net direct method is 0.2910, for CBAM U-Net two-step method is 0.7728, and for the proposed method is 0.8734, clearly higher than the other methods. Both of these evaluation metrics serve to demonstrate that the proposed method markedly enhances the fidelity of the reconstructed image. Fig. 5 shows the MSE statistics for the TransAtten U-Net direct method, CBAM U-Net two-step method, and proposed method. In all the above experiments, the average reconstruction time per instance for TransAtten U-Net was 0.0223 seconds. The experiments were conducted on a personal laptop equipped with an NVIDIA GeForce MX450 GPU (2 GB GDDR6 memory, 896 CUDA cores).

TABLE 1. MSE and SSIM for four methods.

Method	MSE	SSIM
DT	0.0811	0.1747
TransAtten U-Net direct method	0.0644	0.2910
CBAM U-Net two-step method	0.0205	0.7728
TransAtten U-Net two-step method	0.0078	0.8734

On the MNIST test set, the proposed method demonstrates more accurate reconstruction results under both high and low contrast conditions. It achieves the lowest MSE and highest SSIM, significantly outperforming the DT algorithm, TransAt-

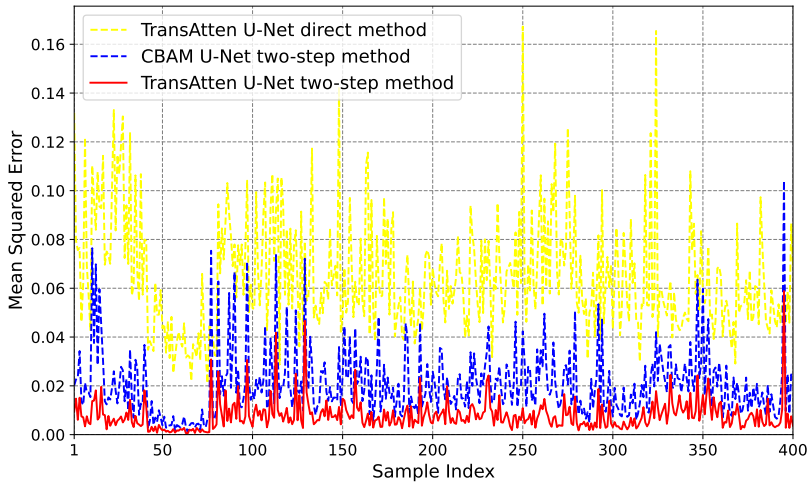

FIGURE 5. MSE statistics for the TransAtten U-Net direct method, the CBAM U-Net two-step method, and the proposed method.

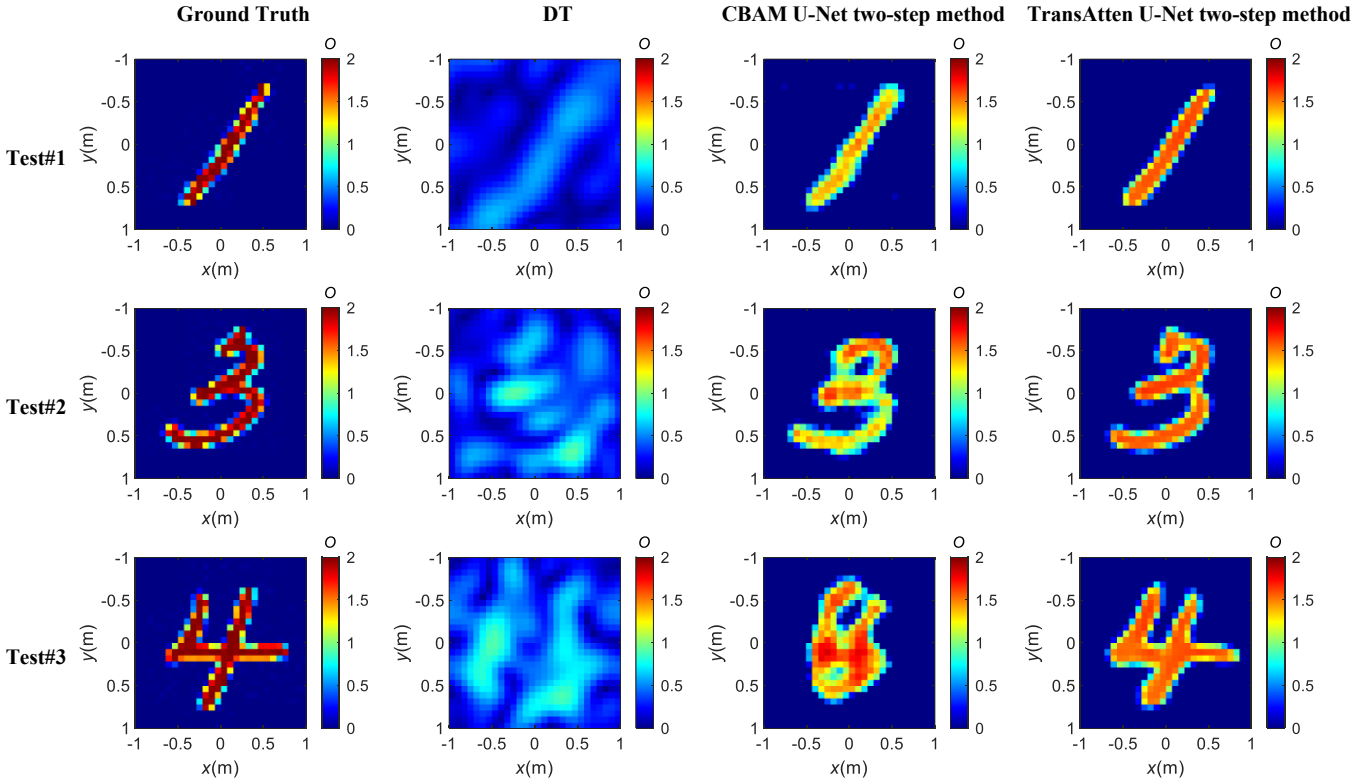

FIGURE 6. Reconstructed results for MNIST test samples with contrast of 2. O denotes the contrast.

ten U-Net direct method, and CBAM U-Net two-step method, while maintaining real-time imaging capability. This result clearly demonstrates that the introduction of physics-assisted mechanisms and the improvements made to the CBAM U-Net each makes independent and crucial contributions to the performance enhancement of the proposed method.

4.3. Generalization Tests

In the following generalization tests, we excluded the method with significantly inferior performance from the previous section. Since TransAtten U-Net direct method did not yield satisfactory results in initial testing (see Subsection 4.2), we focused on comparing the performance of the proposed method, CBAM U-Net two-step method, and DT algorithm.

A random selection of test samples from the MNIST test set is used for generalization testing, with contrast values of 2, which exceed the contrast range of the training set. Fig. 6 shows the true and reconstructed contrast distributions for the test samples, comparing the results of the DT algorithm, CBAM U-Net two-step method, and proposed method. Each row of Fig. 6 rep-

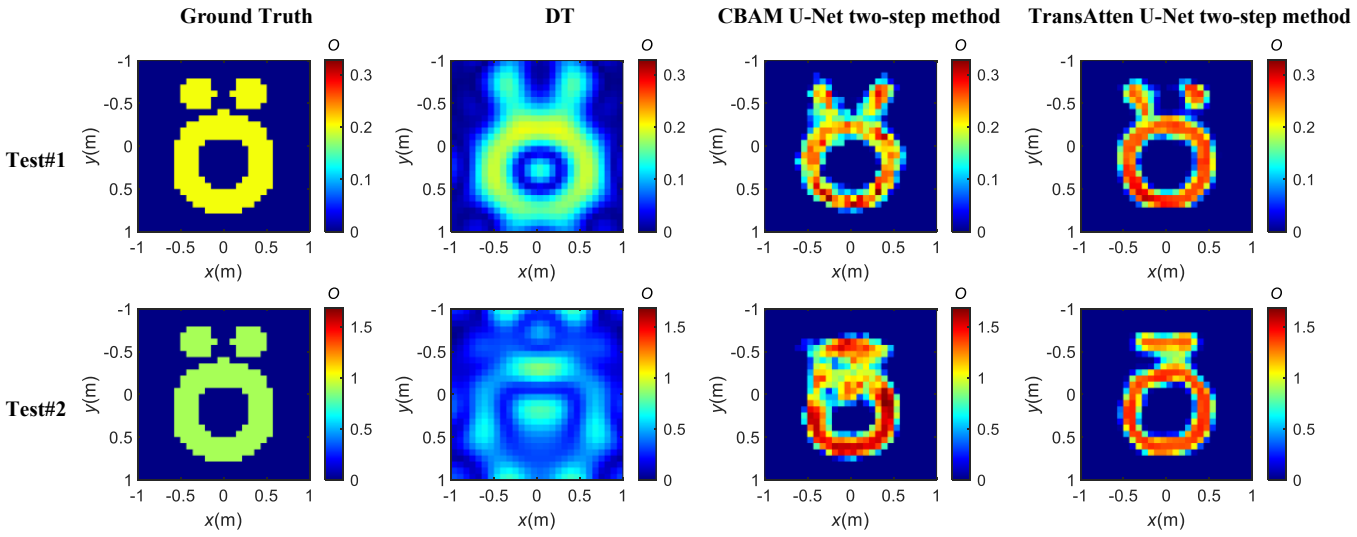

FIGURE 7. Reconstructed results for the Austria Profile. O denotes the contrast.

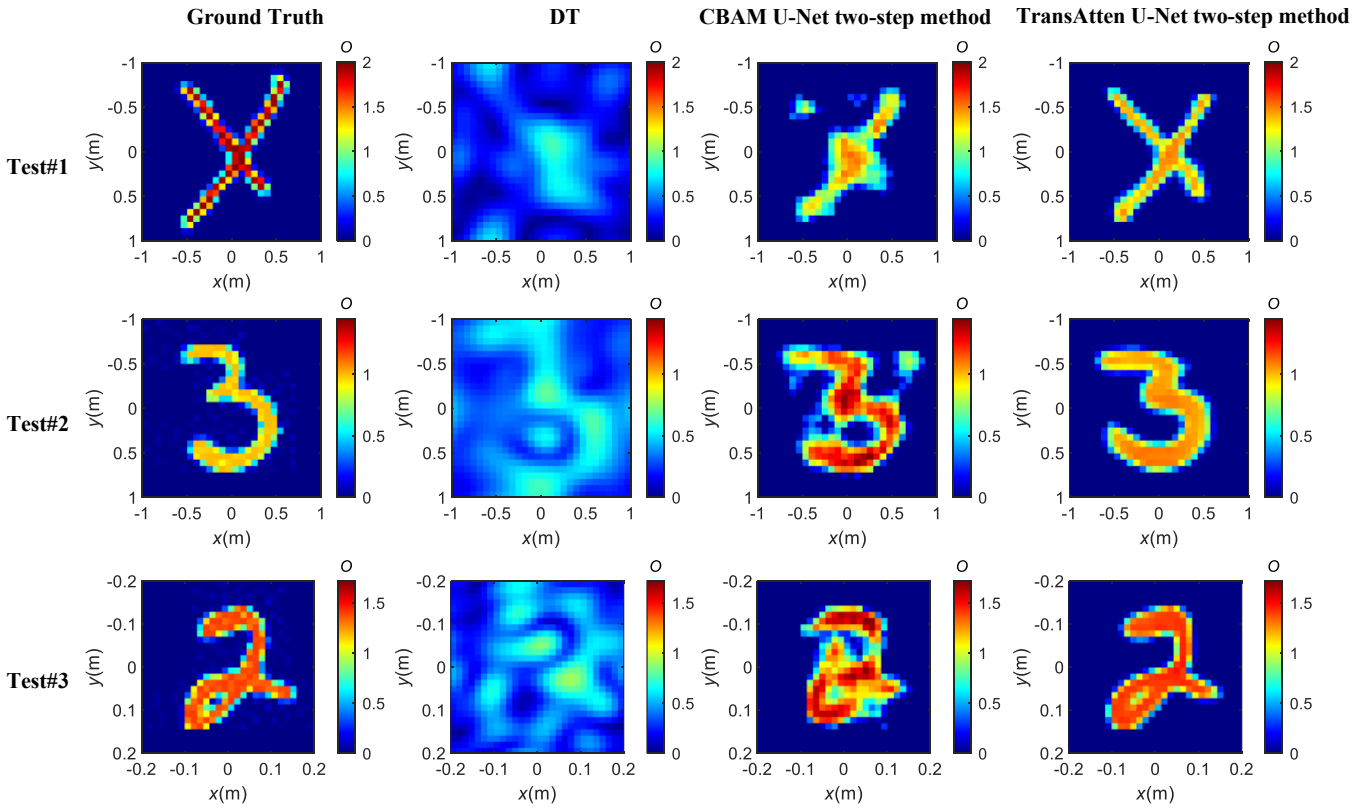

FIGURE 8. Generalization test results for handwritten letters, noisy MNIST, and varying transmitter/receiver configurations. O denotes the contrast.

resents a group of test samples, with the reconstructed images displayed according to the color scale of the target image. Compared with the other methods, the proposed method demonstrates more accurate reconstruction results, even for contrast that exceeds the training set.

"Austria Profile" is selected for the generalization test. The Austria Profile consists of two discs and an annular shape, widely accepted in various research fields, including inverse scattering. Fig. 7 shows the true and reconstructed contrast distributions for the Austria Profile, comparing the results of the DT algorithm, CBAM U-Net two-step method, and proposed method. Each row of Fig. 7 represents a group of test samples, with the reconstructed images displayed according to the color scale of the target image. Compared with other methods, the proposed method achieves more accurate reconstruction images, even for shapes that exceed the training set.

Random test samples of handwritten letters are selected from the EMNIST [24] test set for generalization testing, with con-

trast values of 2, which exceed the contrast range of the training set. The first row in Fig. 8 presents the ground truth and reconstructed contrast distributions of the handwritten letters, comparing the results of the DT algorithm, CBAM U-Net two-step method, and proposed method. Compared with other methods, the proposed method achieves more accurate reconstructed images, even for a shape and contrast that exceed the training set.

Test samples are randomly selected from the MNIST test set, with 20% Gaussian white noise added for the generalization test. The second row in Fig. 8 presents the ground truth and reconstructed contrast distributions for the test samples, comparing the outcomes of the DT algorithm, CBAM U-Net two-step method, and proposed method. Compared with the other methods, the proposed method achieves more accurate reconstructed images, even in the presence of significant noise.

Test samples are randomly selected from the MNIST test set, where the number of transmitters and receivers and the transmitting wave frequency are varied for the generalization test. The number of transmitters and receivers is set to 24, and the transmitting wave frequency is adjusted to 1.5 GHz, corresponding to a wavelength (λ) of 0.2 m. The side length of the region of interest (DOI) and the distribution radius of transmitters and receivers are scaled proportionally to the wavelength change. The third row in Fig. 8 presents the ground truth and reconstructed contrast distributions for the test samples, comparing the results of the DT algorithm, CBAM U-Net two-step method, and proposed method. Compared with other methods, the proposed method achieves more accurate reconstructed images, even for test samples with different incident wave configurations and frequencies.

The results of the various generalization tests notably show that the proposed method outperforms both the DT algorithm and CBAM U-Net two-step method. Compared with the other methods, it demonstrates stronger adaptability and stability.

5. CONCLUSION

This paper introduces a two-step method that combines TransAtten U-Net with physics-assisted learning to tackle the challenges of electromagnetic inverse scattering problems. In the first step, DT algorithm is used to obtain an initial contrast estimate of the target, which serves as a low-resolution input for further optimization by TransAtten U-Net. Building upon CBAM U-Net, the introduction of Transformer module leverages multi-head self-attention mechanisms to enhance the network's ability to model global information, significantly improving the fine reconstruction of contrast images. Experimental results demonstrate that the proposed TransAtten U-Net two-step method achieves higher computational accuracy than the traditional DT algorithm, TransAtten U-Net direct method, and CBAM U-Net two-step method, while maintaining real-time imaging capability. Moreover, the proposed method shows stronger generalization ability than the CBAM U-Net two-step method. Subsequent inquiries could delve into the incorporation of supplementary physical principles, with the intention of augmenting the model's versatility and resilience in more convoluted electromagnetic inverse scattering contexts.

REFERENCES

[1] Battaglia, G. M., A. F. Morabito, R. Palmeri, and T. Isernia, "Effective non-iterative phase retrieval of 2-D bandlimited signals with applications to antenna characterization and diagnostics," *IEEE Transactions on Antennas and Propagation*, Vol. 71, No. 8, 6444–6453, 2023.

[2] Chen, X., *Computational Methods for Electromagnetic Inverse Scattering*, John Wiley & Sons, 2018.

[3] Li, M., R. Guo, K. Zhang, Z. Lin, F. Yang, S. Xu, X. Chen, A. Massa, and A. Abubakar, "Machine learning in electromagnetics with applications to biomedical imaging: A review," *IEEE Antennas and Propagation Magazine*, Vol. 63, No. 3, 39–51, 2021.

[4] Chen, X., Z. Wei, L. Maokun, and P. Rocca, "A review of deep learning approaches for inverse scattering problems (invited review)," *Electromagnetic Waves*, Vol. 167, 67–81, 2020.

[5] Bevacqua, M. T., S. D. Meo, L. Crocco, T. Isernia, and M. Pasian, "Millimeter-waves breast cancer imaging via inverse scattering techniques," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, Vol. 5, No. 3, 246–253, 2021.

[6] Benny, R., T. A. Anjit, and P. Mythili, "Deep learning based non-iterative solution to the inverse problem in microwave imaging," *Progress In Electromagnetics Research M*, Vol. 109, 231–240, 2022.

[7] Zhou, H., X. Huang, and Y. Wang, "Nonlinear inverse scattering imaging method based on iterative multi-scale network," *Chinese Journal of Radio Science*, Vol. 37, No. 6, 1019–1024, 2022.

[8] Guo, R., T. Huang, M. Li, H. Zhang, and Y. C. Eldar, "Physics-embedded machine learning for electromagnetic data imaging: Examining three types of data-driven imaging methods," *IEEE Signal Processing Magazine*, Vol. 40, No. 2, 18–31, 2023.

[9] Zhou, Y., Y. Zhong, Z. Wei, T. Yin, and X. Chen, "An improved deep learning scheme for solving 2-D and 3-D inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, Vol. 69, No. 5, 2853–2863, 2021.

[10] Xu, K., Z. Qian, Y. Zhong, J. Su, H. Gao, and W. Li, "Learning-assisted inversion for solving nonlinear inverse scattering problem," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 71, No. 6, 2384–2395, 2023.

[11] Guo, R., Z. Lin, T. Shan, X. Song, M. Li, F. Yang, S. Xu, and A. Abubakar, "Physics embedded deep neural network for solving full-wave inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, Vol. 70, No. 8, 6148–6159, 2022.

[12] Yao, H. M., W. E. I. Sha, and L. Jiang, "Two-step enhanced deep learning approach for electromagnetic inverse scattering problems," *IEEE Antennas and Wireless Propagation Letters*, Vol. 18, No. 11, 2254–2258, 2019.

[13] Gao, Y., H. Liu, X. Wang, and K. Zhang, "On an artificial neural network for inverse scattering problems," *Journal of Computational Physics*, Vol. 448, 110771, 2022.

[14] Salucci, M., M. Arrebola, T. Shan, and M. Li, "Artificial intelligence: New frontiers in real-time inverse scattering and electromagnetic imaging," *IEEE Transactions on Antennas and Propagation*, Vol. 70, No. 8, 6349–6364, 2022.

[15] Ronneberger, O., P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015*, 234–241, 2015.

[16] Xia, Y. and S. He, "A lightweight deep learning model for full-wave nonlinear inverse scattering problems," *Progress In Electromagnetics Research M*, Vol. 128, 83–88, 2024.

[17] Sun, Y., Z. Xia, and U. S. Kamilov, "Efficient and accurate inversion of multiple scattering with deep learning," *Optics Express*, Vol. 26, No. 11, 14 678–14 688, 2018.

[18] Wei, Z. and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 4, 1849–1860, 2019.

[19] Woo, S., J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19, 2018.

[20] Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, Vol. 30, No. 1, 2, 2017.

[21] Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16 × 16 words: Transformers for image recognition at scale," *ArXiv Preprint ArXiv:2010.11929*, 2020.

[22] LeCun, Y., C. Cortes, C. Burges, *et al.*, "Mnist handwritten digit database," 2010.

[23] Cakoni, F., D. Colton, and H. Haddar, *Inverse Scattering Theory and Transmission Eigenvalues*, SIAM, 2022.

[24] Cohen, G., S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: Extending MNIST to handwritten letters," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2921–2926, Anchorage, AK, USA, 2017.