Araştırma Makalesi *BAUN Fen Bil. Enst. Dergisi, 28(2), 921-936, (2026)*

DOI:10.25092/baunfbed. 1949121 *J. BAUN Inst. Sci. Technol., 28(2), 921-936, (2026)*

Deep learning algorithms for through-the-wall imaging

Xiaofan JIA[1], Abdulkadir C. YUCEL[1,*]

[1]*Nanyang Technological University, College of Engineering, The School of Electrical and Electronic Engineering, West Jurong, Singapore*

Geliş Tarihi (Received Date): 12.05.2026
Kabul Tarihi (Accepted Date):22.06.2026

Abstract

Fast through-the-wall imaging (TWI) by inverse scattering is highly desirable for many security and civilian applications. TWI by inverse scattering requires the solution of an ill-posed and nonlinear set of equations whose solution is attained iteratively (e.g., via a distorted Born iterative method). The iterative solution requires long execution times and often does not converge, limiting the applicability of inverse scattering to real-world TWI scenarios. To address these issues, this study investigates deep learning-augmented inverse-scattering schemes that combine physics-based modeling with the learning capability of neural networks. Using images after a few distorted Born iterations as input, three neural networks, namely a traditional convolutional neural network (CNN), U-net, and mask region-based CNN (Mask R-CNN) are leveraged to obtain final high-quality images of the scatterers behind the walls and their performance is compared. Among these three techniques, the traditional CNN regresses scatterer positions and restores dielectric profiles, while U-net segments scatterers and Mask R-CNN detects scatterers. Numerical results show that U-net achieves the highest accuracy, while Mask R-CNN offers competitive accuracy and removes boundary defects, making it highly effective for TWI applications.

Keywords: *Convolutional neural network (CNN), deep learning, inverse scattering, Mask R-CNN, through-the-wall imaging, U-net.*

Duvar arkası görüntüleme için derin öğrenme algoritmaları

Öz

Ters saçılmaya dayalı hızlı duvar arkası görüntüleme (DAG), birçok güvenlik ve sivil uygulama için önemli bir ihtiyaçtır. Ancak ters saçılma tabanlı DAG, çözümü yinelemeli

[*]Abdulkadir C. Yucel, acyucel@ntu.edu.sg, https://orcid.org/0000-0001-9920-4043
Xiaofan Jia, xfjia@ntu.edu.sg, https://orcid.org/0000-0001-8534-3755

olarak elde edilen kötü koşullu ve doğrusal olmayan denklem takımlarının çözülmesini gerektirir; örneğin bu amaçla Distorted Born iteratif yöntemi kullanılabilir. Bu yinelemeli çözüm süreçleri uzun hesaplama süreleri gerektirmekte ve çoğu zaman yakınsamama problemiyle karşılaşmakta, dolayısıyla ters saçılmanın gerçek dünya DAG senaryolarındaki uygulanabilirliğini sınırlamaktadır. Bu sorunları ele almak amacıyla bu çalışmada, fiziğe dayalı modellemeyi sinir ağlarının etkin öğrenme kabiliyetiyle birleştiren derin öğrenme destekli ters saçılma algoritmaları incelenmektedir. Birkaç Distorted Born iterasyonu sonrasında elde edilen görüntüler girdi olarak kullanılarak, duvar arkasındaki saçıcıların nihai yüksek kaliteli görüntülerini elde etmek için üç farklı sinir ağı mimarisi—geleneksel evrişimsel sinir ağı (CNN), U-net ve maske bölge tabanlı CNN (Mask R-CNN)—kullanılmış ve performansları karşılaştırılmıştır. Bu üç teknik arasında geleneksel CNN saçıcı konumlarını kestirmekte ve dielektrik profilleri geri kazandırmakta, U-net saçıcıları bölütlemekte, Mask R-CNN ise saçıcıları tespit etmektedir. Sayısal sonuçlar, U-net'in en yüksek doğruluğu sağladığını; Mask R-CNN'in ise rekabetçi doğruluk sunarken sınır kusurlarını giderdiğini ve bu yönüyle DAG uygulamaları için oldukça etkili bir yöntem olduğunu göstermektedir.

Anahtar kelimeler: *Evrişimsel sinir ağı (CNN), derin öğrenme, ters saçılma, maske bölge tabanlı CNN (Mask R-CNN), duvar arkası görüntüleme (DAG), U-net*

1. Introduction

Both civilian and security applications significantly benefit from through-the-wall imaging (TWI), a non-destructive technique for imaging obscured objects and scatterers behind walls [1, 2]. This technique processes electromagnetic (EM) signals to detect and image objects within a domain of interest. Often, TWI is desired to be performed by solving an inverse scattering problem (ISP), yielding high-resolution reconstructions compared to radar-based approaches [1]. This paper improves these reconstructions by hybridizing a standard ISP solution approach with deep learning algorithms.

ISP requires the solution of a system of equations that is ill-posed and nonlinear. To do this, objective-function (model-based) approaches are traditionally used [3]. These approaches can be either iterative or noniterative [4]. Noniterative methods, such as back-propagation, Born approximation [5], and the dominant current scheme, compute results fast but are limited to reconstructing low-contrast scatterers [6]. On the other hand, iterative methods, including Born iterative method [7], distorted Born iterative method (DBIM) [8], and contrast source inversion [9], are capable of reconstructing high-contrast scatterers with finer resolution [6]. Despite their clear advantages, their practical applications are limited due to high computational costs [4].

To break the trade-off between computational cost and output accuracy, deep learning algorithms have been employed to solve ISP. In these approaches, the computational burden is shifted to the offline training stage, which is only done once. Thereafter, fast and accurate reconstructions can be attained online [3]. Supervised learning was first applied to ISP using purely data-driven methods, in which deep learning algorithms serve as direct regression models to infer target properties from measured scattered-field data [4, 10]. Later, physical principles have been leveraged to improve the efficiency of deep learning algorithms [11, 12]. Deep learning algorithms have since been deployed to learn the descent directions for fast convergence of iterative approaches [13], provide good initial estimates for ISP models [14, 15], and generate final images [4, 6, 16-20]. However, neural networks are used as 'black boxes' in these approaches and often lack interpretability. The latest trend tackles this issue by integrating physical principles into neural networks through two approaches: physically constrained optimization, which applies additional physical constraints to the loss function during training [21, 22], and the combination of physical methods with neural networks to form physics-informed neural networks [23-26]. Supervised learning requires ground-truth dielectric profiles and contrast values as labels, which can be difficult to obtain in complex scenarios [27]. Additionally, the neural networks' generalizability can be significantly hindered if the testing data deviates from the training data distribution [28]. These limitations have led to growing interest in unsupervised learning, including self-supervised learning for ISP [27-30]. However, all these works are solutions for general ISPs aiming at the reconstruction of the shape, position, and dielectric profile of the scatterers in the investigation domain.

In this study, three deep learning-augmented schemes are specifically designed for TWI applications, where neural networks detect scatterers from blurred DBIM outputs. In the first scheme, a traditional CNN [31] regresses the scatterer coordinates and the permittivity profile. The second scheme combines the DBIM with a U-net [32] for image segmentation, refining the permittivity profile obtained from DBIM and distinguishing scatterers from the background. The third employs Mask Region-based CNN (Mask R-CNN) [33], an object detection algorithm, to accurately extract scatterer shapes and positions from DBIM output. A performance comparison of the three schemes demonstrates that U-net achieves the highest accuracy, whereas Mask R-CNN excels at eliminating defects in retrieved scatterers through shape constraints. Both U-net and Mask R-CNN-based schemes surpass traditional CNN-augmented algorithm in accuracy, image quality, and generalizability. The rest of this paper is organized as follows. Section II briefly explains DBIM and introduces three deep learning schemes. In Section III, the performance and generalizability of the three neural networks are compared via numerical examples. Finally, the conclusions are drawn in Section IV.

2. Deep learning-augmented schemes

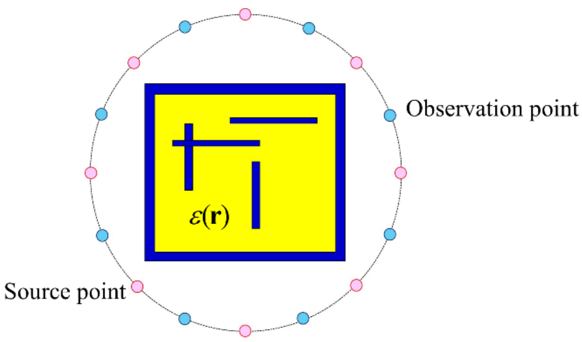

Figure 1. TWI Scenario.

The 2D scenario under consideration for TWI is shown in Figure 1. In this scenario, N_t line sources and N_r observation points are placed along a circle around the walls for illuminating the investigation domain and measuring the scattered fields, respectively. The problem of TWI via inverse scattering is to reconstruct the permittivity profile of the investigation domain $\varepsilon(\mathbf{r})$ from the fields obtained at observation points. To leverage the underlying physics of ISP and reduce the learning burden on neural networks, inverse scattering schemes incorporate an objective-function approach using deep learning algorithms. Rough estimates of the permittivity profiles are first obtained by the DBIM from raw measurement data and then used as input to a traditional CNN, U-Net, or Mask R-CNN to improve reconstruction accuracy.

2.1. Distorted Born iterative method

The implementation details of the distorted Born iterative method (DBIM) [8] are given in this section. First, the numerical methods pertaining to determining the E-fields generated by relevant sources are described. Then, the distorted Born method is presented as an iterative method that matches numerically determined E-fields to measured E-fields to determine the permittivity of a scatterer.

2.1.1. Solution of the forward problem

It is assumed that there are N_t transmitting line sources driven with current $J_{trans}^{(l)}(\boldsymbol{\rho})=\delta(\boldsymbol{\rho}-\boldsymbol{\rho}_{t_l})$, where $\boldsymbol{\rho}_{t_l}$ is a source location and $l=1,...,N_t$, and N_r receivers measure the E-fields at locations $\boldsymbol{\rho}_{r_m}$, where $m=1,\ldots,N_r$. The transmitters and receivers reside outside a scatterer whose permittivity $\varepsilon_r(\boldsymbol{\rho})$ is unknown and its support is denoted by Ω, where $\boldsymbol{\rho}=(x,y)$ is the position in Cartesian coordinates. The scatterer is discretized by a regular grid consisting of N rectangular elements of dimensions $\Delta_x\times\Delta_y$. Furthermore, all field quantities and the permittivity are assumed constant within each grid element. By applying the volume equivalence principle to the rectangular grid elements with centers $\mathbf{p}_i$, where $i=1,\ldots,N$, the E-field generated by the l^{th} transmitter can be determined as

$$
\begin{aligned}
\mathbf{x}_k^{(l)} &= \mathbf{Z}_k^{(-1)}\mathbf{x}_{\text{inc}}^{(l)}, \quad l=1,...,N_t \\
\mathbf{Z}_k &= \left(\mathbf{I}-\mathbf{Z_0}Diag\left\{\boldsymbol{\varepsilon}_k-\mathbf{1}\right\}\right) \\
\mathbf{x}_0^{(l)} &= \left[\frac{H_0^{(2)}\left(k_0\left|\mathbf{p}_1-\boldsymbol{\rho}_l'\right|\right)}{4j} \;\cdots\; \frac{H_0^{(2)}\left(k_0\left|\mathbf{p}_N-\boldsymbol{\rho}_l'\right|\right)}{4j}\right]^T \\
\left(\mathbf{Z_0}\right)_{i,j} &= k_0^2\int_\Omega \frac{H_0^{(2)}\left(k_0\left|\mathbf{p}_i-\boldsymbol{\rho}'\right|\right)}{4j}P_j\left(\boldsymbol{\rho}'\right)d\boldsymbol{\rho}' \\
E_k^{(l)}\left(\boldsymbol{\rho}\right) &= \sum_{i=1}^{N}P_i\left(\boldsymbol{\rho}\right)x_{i,k}^{(l)} \\
\boldsymbol{\varepsilon}_k\left(\boldsymbol{\rho}\right) &= \sum_{i=1}^{N}P_i\left(\boldsymbol{\rho}\right)\varepsilon_{i,k} \\
\mathbf{x}_k^{(l)} &= \left[x_{1,k}^{(l)}\ldots x_{N,k}^{(l)}\right]^T, \quad \boldsymbol{\varepsilon}_k=\left[\varepsilon_{1,k}\ldots\varepsilon_{N,k}\right]^T.
\end{aligned}
\tag{1}
$$

Here $P_j(\boldsymbol{\rho})$ is a pulse basis having support on the j^{th} rectangular element, $H_0^{(2)}(\cdot)$ is the zeroth-order Hankel function of the second kind, $Diag\{\mathbf{y}\}$ is an operator that generates a diagonal matrix with diagonal entries chosen from $\mathbf{y}$, $\mathbf{1}$ is a vector consisting of entries of all ones, and k_0 is the free-space wavenumber of the source wave. Furthermore, the received E-field at each of the elements can be determined as

$$\mathbf{b}_k^{(l)} = \mathbf{E}_{\mathbf{inc}}^{(l)} + \mathbf{Z}_{\mathbf{scat_0}} Diag\left\{\boldsymbol{\varepsilon}_k - \mathbf{1}\right\} \mathbf{x}_k^{(l)}, \quad l = 1, ..., N_t$$
$$\mathbf{E}_{\mathbf{inc}}^{(l)} = \left[\frac{H_0^{(2)}\left(k_0 \left|\boldsymbol{\rho}_{r_1} - \boldsymbol{\rho}'_{t_l}\right|\right)}{4j} \cdots \frac{H_0^{(2)}\left(k_0 \left|\boldsymbol{\rho}_{r_{N_r}} - \boldsymbol{\rho}'_{t_l}\right|\right)}{4j} \right]^T \tag{2}$$
$$\left(\mathbf{Z}_{\mathbf{scat_0}}\right)_{i,j} = k_0^2 \int_{\Omega} \frac{H_0^{(2)}\left(k_0 \left|\boldsymbol{\rho}_{\mathbf{r}_i} - \boldsymbol{\rho}'\right|\right)}{4j} P_j\left(\boldsymbol{\rho}'\right) d\boldsymbol{\rho}'$$

2.1.2. Distorted Born iterative method

Assume that $\mathbf{b}_{true}^{(l)}$ is the true measured E-field by the scatterer, $\mathbf{x}_{true}$ is the true E-field on rectangular element centers, and $\boldsymbol{\varepsilon}_{true}$ is the permittivity of the scatterer rectangular elements. The volume equivalence principle dictates that

$$\mathbf{b}_{true}^{(l)} - \mathbf{b}_k^{(l)} = \mathbf{Z}_{\mathbf{scat},k} Diag\left\{\mathbf{x}_{true}\right\} \boldsymbol{\delta\varepsilon}_k, \quad l = 1, ..., N_t$$
$$\boldsymbol{\delta\varepsilon}_{k+1} = (\boldsymbol{\varepsilon}_{true} - \boldsymbol{\varepsilon}_k) \tag{3}$$
$$\mathbf{Z}_{\mathbf{scat},k} = \mathbf{Z}_{\mathbf{scat_0}} (\mathbf{I} + \frac{ik_0}{\eta}\left(\boldsymbol{\varepsilon}_k - \mathbf{1}\right) \mathbf{Z}_k^{-1} \mathbf{Z_0})$$

where η is the impedance of free space. (3) enables the determination of the true permittivity profile. However, it requires measurement of the unknown true E-field in the domain. Distorted Born approximates $\mathbf{x}_{true}$ by $\mathbf{x}_{\varepsilon_k}$ in (3) to update an estimate of the true permittivity $\boldsymbol{\varepsilon}_k$ to $\boldsymbol{\varepsilon}_{k+1} = \boldsymbol{\varepsilon}_k + \boldsymbol{\delta\varepsilon}_k$. The Born approximation requires that updates of $\boldsymbol{\varepsilon}_k$ are small enough to ensure that $\mathbf{x}_{k+1} \approx \mathbf{x}_k$. Furthermore, since (3) is known to be ill-conditioned, regularization must be employed to ensure that the solution is bounded. Here we employ L^2 regularization by minimizing $\left\|\boldsymbol{\delta\varepsilon}_{\mathbf{k+1}}\right\|_{\mathbf{2}}^{\mathbf{2}}$. The final Distorted Born algorithm is provided in Algorithm 1.

Algorithm 1: Distorted Born iterative method

Inputs:	$\mathbf{b}_{true} = \left[\mathbf{b}_{true}^{(1)\,T} \ldots \mathbf{b}_{true}^{(N_t)\,T}\right]^T$, $\left(\mathbf{x}_0^{(1)\,T} \ldots \mathbf{x}_0^{(N_t)\,T}\right)$, $\left(\mathbf{E}_{\mathbf{inc}}^{(1)T} \ldots \mathbf{E}_{\mathbf{inc}}^{(N_t)T}\right)$, threshold($thresh$), number of iterations ($N_{\mathbf{it}}$)
1	**Initiation**: $\boldsymbol{\varepsilon}_0 = [\varepsilon_0 \ldots \varepsilon_0]^T$, where $l = 1, \ldots, N_t$, and the relative residual error $(RRE) = 1$
2	**for** $k = 0$ to $N_{\mathbf{it}}$ **do**
3	update E-scattered $\mathbf{b}_k = \begin{bmatrix} \mathbf{E}_{\mathbf{inc}}^{(1)} + \mathbf{Z}_{\mathbf{scat_0}} Diag\left\{\boldsymbol{\varepsilon}_k - \mathbf{1}\right\} \mathbf{x}_k^{(1)} \\ \mathbf{E}_{\mathbf{inc}}^{(N_t)} + \mathbf{Z}_{\mathbf{scat_0}} Diag\left\{\boldsymbol{\varepsilon}_k - \mathbf{1}\right\} \mathbf{x}_k^{(N_t)} \end{bmatrix}$

evaluate scattering matrix* $\mathbf{K} \leftarrow \begin{bmatrix} \mathbf{Z}_{\mathbf{scat}_k} Diag\left\{\mathbf{x}_k^{(1)}\right\} \\ \vdots \\ \mathbf{Z}_{\mathbf{scat}_k} Diag\left\{\mathbf{x}_k^{(N_t)}\right\} \end{bmatrix}$

determine regularization parameter $\gamma = \max\left(\frac{RRE^2}{2}, 10^{-4}\right)$

determine permittivity update $\boldsymbol{\delta\varepsilon}_\mathbf{k} = \left(\mathbf{K}^\mathbf{H}\mathbf{K} + \gamma\mathbf{I}\right)\mathbf{K}^\mathbf{H}(\mathbf{b}_{true} - \mathbf{b}_k)$

update permittivity $\boldsymbol{\varepsilon}_{k+1} = \boldsymbol{\varepsilon}_k + \boldsymbol{\delta\varepsilon}_k$

update E-total estimate in domain $\mathbf{x}_{k+1}^{(j)} = \mathbf{Z}_{k+1}^{-1}\mathbf{x}_0^{(l)}$, where
$l = \{1, \dots, N_t\}$

compute $RRE = \|\mathbf{b}_k\| / \|\mathbf{b}_\mathbf{0}\|$

If $RRE < thresh$, **then** break

* For details of computing $\mathbf{Z}_{\mathbf{scat}_0}, \mathbf{Z}_0$, $\mathbf{Z}_{\mathbf{scat}_k}$, and $\mathbf{Z}_{k+1}$ refer to the main text.

2.2. Neural networks

The implementation details of the traditional CNN [31], U-net [32], and Mask R-CNN [33] are given in this section. These neural network algorithms are used to reconstruct the interior walls of a small residential unit layout, as depicted in Figure 1, without loss of generality, while the outer walls are already known and omitted during the reconstruction.

2.2.1. Traditional CNN

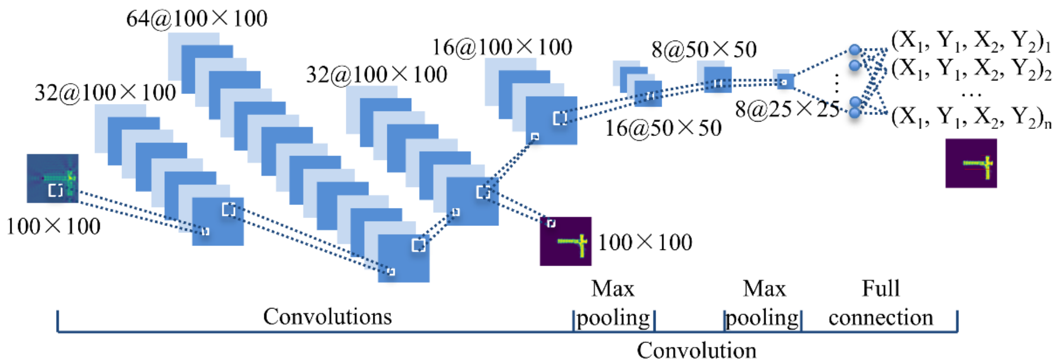

Figure 2. Traditional CNN architecture for the proposed scheme, where 3×3 convolutions (stride = 1) and 2×2 max pooling are applied.

A traditional CNN with the architecture shown in Figure 2 is used to determine the coordinates of the bottom-left and top-right corners of each interior wall. Since multi-task learning enhances both prediction accuracy and learning efficiency of a deep learning algorithm, the network is designed not only to predict coordinates (upper branch) but also to reconstruct permittivity profile (lower branch) simultaneously. In the CNN architecture [Figure 2], the first three convolutional layers extract features from input data generated with DBIM, which is followed by another convolutional layer for permittivity profile, which consists of repeated application of 3×3 convolutions and 2×2 max-pooling operators, along with two fully connected layers to further capture and downsample features. Rectified linear unit (ReLU) serves as the activation function for nonlinear mapping, except for the last profile reconstruction layer, where the sigmoid function is applied. Batch normalization (BN) and dropout are performed to prevent overfitting. As the number of output neurons corresponds to the coordinates of scatterers, the number of objects must be given as prior knowledge to the traditional CNN.

For the multi-task learning algorithm, a combined loss function is defined as:

$$L = L_{\varepsilon} + \alpha L_{r}\,. \tag{4}$$

Here L_{ε} and L_{r} denote the loss of permittivity profile reconstruction (cross-entropy loss) and the regression loss of wall coordinates (mean squared error loss), respectively, and α is the weighting coefficient.

2.2.2. U-net

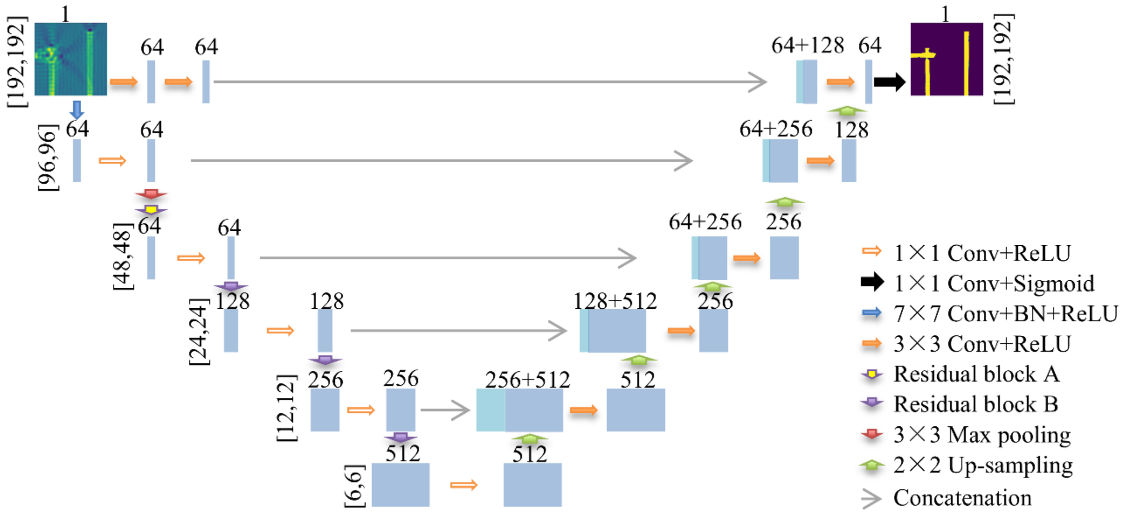

Figure 3. U-net architecture for the proposed scheme, where Conv and BN represent convolution and batch normalization, respectively.

Although it is possible to regress corner coordinates directly by the traditional CNN model to determine interior wall locations, the number of scatterers must be provided *a priori* and the model does not generalize well for non-rectangular objects. To remove these restrictions, a U-net [32] is used. The U-net [Figure 3] was first proposed for biomedical image segmentation [32] and has been successfully applied to solve ISPs, as alluded to in the introduction.

As depicted schematically in Figure 3, the U-net comprises three parts: the contracting path (left), the expanding path (right), and the skip connections. During the contraction, repeated convolutional blocks and residual blocks [Figure 4] reduce spatial resolution and capture the context of inputs. The expanding path consists of up-sampling and convolution within convolutional blocks, merging local details with global information. This feature fusion is significant for TWI segmentation due to the common features shared between inputs and outputs.

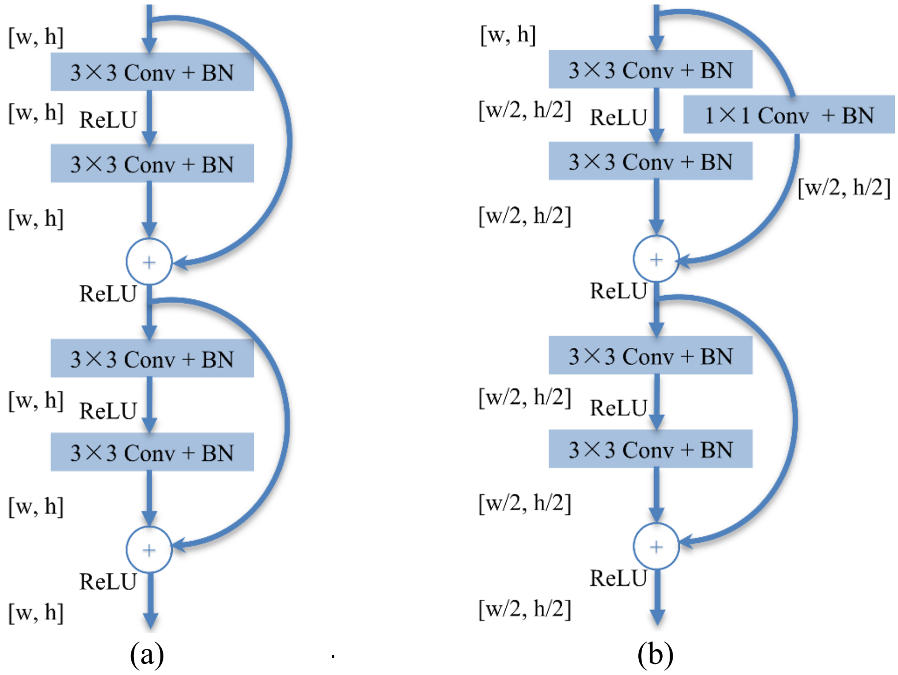

Figure 4. (a) Residual blocks A and (b) residual blocks B applied in U-net. [w, h] denotes the dimensions of input images.

2.2.3. Mask R-CNN

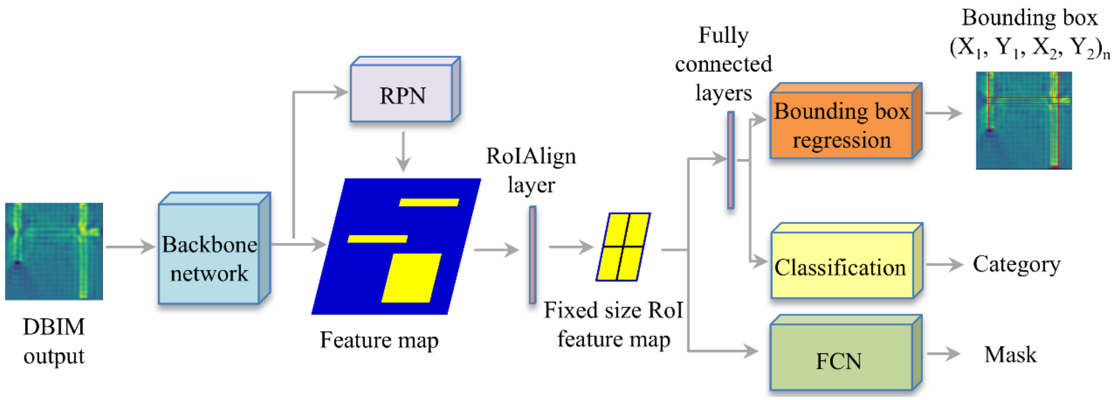

Figure 5. Schematic illustration of Mask R-CNN for the proposed algorithm.

To extract the shapes and positions of objects within the investigation domain, the deep learning algorithms are designed to solve the classical object detection task, in which object locations are annotated with several bounding boxes regardless of their exact shapes. Mask R-CNN [33] performs object detection and image segmentation in parallel. As shown in Figure 5, Mask R-CNN consists of three steps: i) *Feature extraction*: A backbone network with a sophisticated architecture extracts features from the input image. Meanwhile, a feature pyramid network that merges features from different levels is introduced to further enhance the backbone network. ii) *Proposal and alignment of regions of interest*: In the region proposal network (RPN), anchors at different scales and aspect ratios slide on the output feature maps from the backbone network to generate regions of interest (RoIs). Then, the RoI dimensions are adjusted by the RoIAlign layer via bilinear interpolation to a fixed size to meet the input requirements of the subsequent prediction network. iii) *Object detection*: The prediction network consists of three parallel branches: bounding box regression layers, fully connected layers for classification, and the fully convolutional network (FCN) for mask generation. In TWI, output bounding boxes are treated as the wall boundaries.

3. Numerical results

This section presents numerical results comparing the performance of three schemes for TWI. The investigation domain with dimensions of $20\times 20\ \mathrm{m}^2$ is discretized using a grid of 112×112 pixels. The domain consists of horizontally and vertically oriented homogeneous walls that are randomly positioned and enclosed by the exterior walls, as shown in Figure 1. Each wall has a permittivity of $1.5\varepsilon_0$ and a thickness of 6 pixels, where ε_0 is the permittivity of the free space. Non-wall regions are assumed to be air with permittivity ε_0. The proposed scheme and traditional schemes are used to reconstruct the size, position, orientation, and permittivity profile of the interior walls.

A total of $N_i = 72$ line sources and $N_r = 72$ observation points are assumed to be equally spaced on a $15.2\ \mathrm{m}$-diameter circle centered on the domain. The line sources are assumed to operate at a frequency of $80\ \mathrm{MHz}$. The ISP is performed using DBIM with a regularization factor

$$\gamma = \max(\sigma\chi / 20, \sigma\times 10^{-4}). \tag{5}$$

Here σ is the maximum singular value of the discretized system for forward scattering and χ is the error in scattered fields at the observation points.

Images from the 15$^{\text{th}}$ iteration of the DBIM are the input to the neural networks. Before inputting these images, the pixels associated with the exterior walls are removed. This results in 100×100 pixel images. The training, validation, and test data are generated from domains containing one to three interior walls.

3.1. Model training

All deep learning algorithms are implemented in Python using PyTorch [34]. The computations are performed on a server with an NVIDIA Tesla P100 GPU. For neural networks, inputs are the images obtained from the DBIM, and the groundtruths are permittivity profiles of the domain. Three different datasets of 1000 images are generated for one, two, and three-wall scenarios. 75% of images in the dataset are used for training, 10% for validation, and 15% for testing.

3.1.1. Traditional CNN

With images from DBIM as inputs, the desired outputs of the traditional CNN are permittivity profiles and wall coordinates. Three traditional CNNs are separately trained and tested on the three datasets using Adam optimization [35] with a learning rate of 0.06, batch size of 64, α of 10, and dropout rate of 0.5 in the fully connected layer. The maximum training epoch is set to 500 unless no apparent decrease in the loss function is observed.

3.1.2. U-net

Three different datasets are merged into a single dataset and used to train and test a single U-net with binary segmentation outputs. Input images are resized to 192×192 pixels to satisfy the requirements of the U-net input layer. The network is trained with the Adam optimizer to minimize a combined loss of binary cross-entropy loss with dice loss [36]. For residual blocks, we fine-tune the parameters of a ResNet-18 pre-trained on ImageNet [37]. The learning rate is initially set to 10^{-4} and subsequently divided by 10 after every 250 epochs. The batch size is set to 32, and the U-net is trained for 500 epochs. The best model is saved for testing, and the outputs are resized to their original dimension of 100×100.

3.1.3. Mask R-CNN

101-layer ResNeXt [38] with a feature pyramid network is chosen as the backbone network, providing the best trade-off between speed and accuracy [39]. Taking wall locations as targets, we fine-tune a model pre-trained on ImageNet [37] by wall images to obtain a single Mask R-CNN. The Mask R-CNN's loss function is a weighted sum of losses from RPN and the multi-task predictive network [33]. The training hyperparameters are maximum iterations of 5000, batch size of 2, and the detection threshold in the testing of 0.96.

3.2. Testing results

For the groundtruth and predicted location of a particular wall, Intersection over Union (IoU) refers to the ratio between the areas of their overlap and their union, i.e.,

$$\mathrm{IoU} = \frac{\mathrm{Target} \cap \mathrm{Prediction}}{\mathrm{Target} \cup \mathrm{Prediction}}. \tag{6}$$

Additionally, pixel accuracy reports the percent of correctly reconstructed or classified pixels in the image. Table 1 shows quantitative results.

Table 1. Quantitative evaluation of three schemes

		Average IoU	Average pixel accuracy
DBIM + Traditional CNN	One-wall	0.587669	0.990190
	Two-wall	0.269190	0.939339
	Three-wall	0.205940	0.839929
DBIM + U-net		**0.987029**	**0.998860**
DBIM + Mask R-CNN		0.733824	0.981657

3.2.1. Traditional CNN

Profile reconstructions and coordinate regressions shown in Figure 6 suggest that the traditional CNN can predict wall coordinates accurately in simple one-wall cases. However, for more difficult scenarios, the model outputs rough wall locations with distorted widths or lengths. This is because the number of outputs increases with the number of walls, which consequently challenges the learning ability of the relatively shallow traditional CNN. The comparison shown in Table 1 demonstrates that the performance of traditional CNN degrades as the number of walls increases.

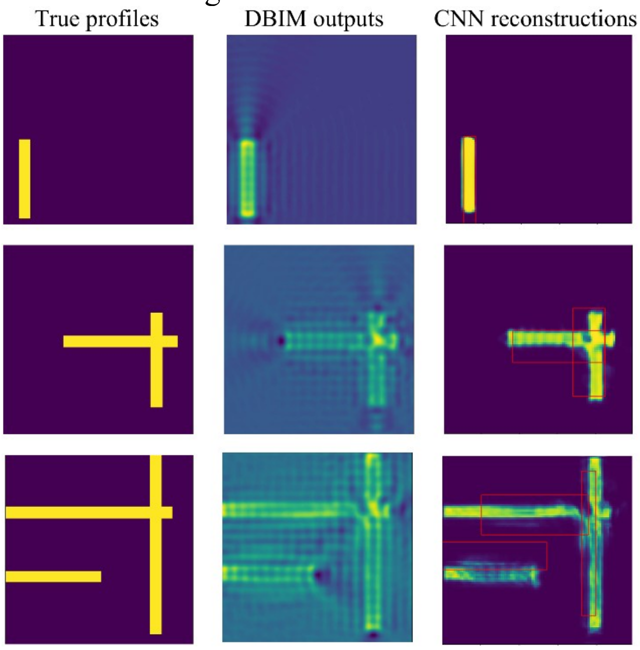

Figure 6. Profile reconstruction and coordinate regression of traditional CNN for one-wall, two-wall, and three-wall scenarios.

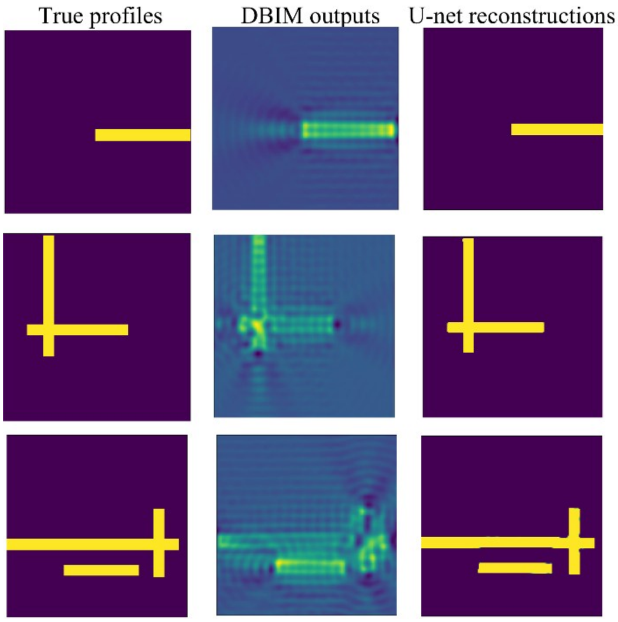

Figure 7. U-net segmentation outputs for one-wall, two-wall, and three-wall scenarios.

3.2.2. U-net

The segmentation outputs from the U-net are presented in Figure 7. The model gives highly accurate wall profiles in one-wall and two-wall scenarios. Although U-net reconstructions fill in missing profile details in distorted Born outputs, defects around the boundary may still be observed in the three-wall cases. The average IoU and pixel accuracy of U-net on the test set are close to 1, confirming that U-net significantly outperforms traditional CNN in solving highly nonlinear scenarios with an increasing number of walls.

3.2.3. Mask R-CNN

Mask R-CNN outputs highly accurate wall positions in one-wall and two-wall cases without boundary defects. Figure 8 only shows the challenging three-wall detections, where walls intersect, and the spacing between walls is narrow. The Mask R-CNN can localize walls accurately even in highly blurry DBIM images. The slightly lower average pixel accuracy of Mask R-CNN [Table 1] results from the quantization error during post-processing, where predicted bounding box coordinates are rounded to integers to generate binary outputs.

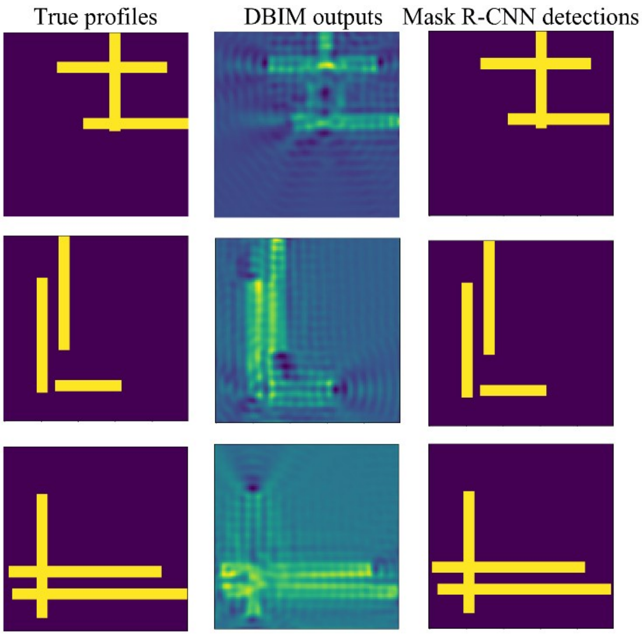

Figure 8. Mask R-CNN detections for three-wall scenarios.

3.2.4. Comparison of schemes

For further comparison, the three proposed schemes are trained on a fixed dataset and tested on the same scenario; the outputs are shown in Figure 9. The performance of the traditional CNN degrades as the number of walls increases. Without shape constraints, defects along the boundary and within the wall are obvious in U-net segmentations. On the other hand, Mask R-CNN localizes the wall more accurately with clear boundaries even in challenging three-wall scenarios. The Mask R-CNN-augmented algorithm benefits from the similarity in shape between walls and detection bounding boxes.

3.3. Test of generalization ability

The traditional CNN scheme has limited generalizability because a separate model is required for each assumed number of walls, whereas a well-pretrained U-net or Mask R-CNN can predict an arbitrary number of walls. Figure 10 depicts the three-wall scenario test results of Mask R-CNN and U-net trained on one-wall and two-wall data. Distinct and accurate wall boundaries confirm the better generalization ability of Mask R-CNN than U-net.

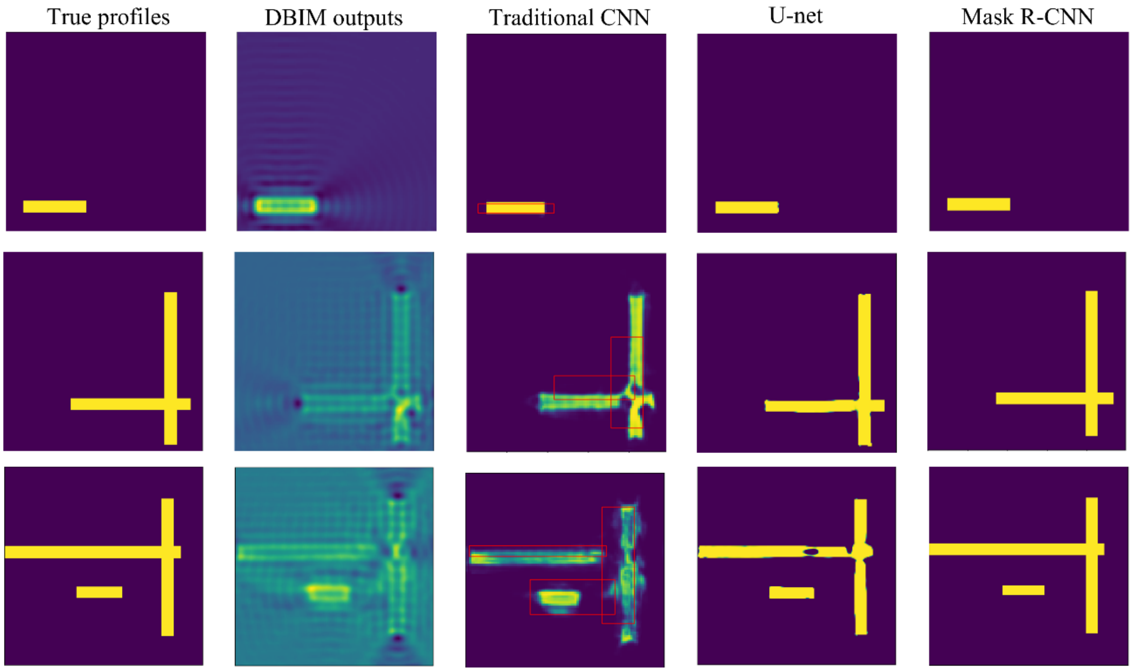

Figure 9. Comparison of traditional CNN, U-net, and Mask R-CNN schemes.

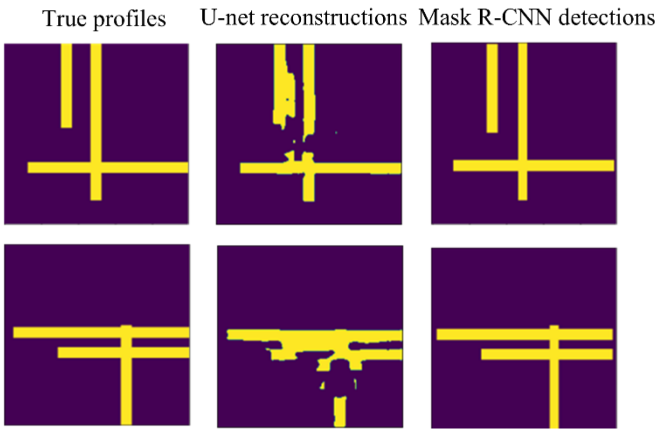

Figure 10. Test on three-wall images with Mask R-CNN or U-net trained on one-wall and two-wall scenarios.

4. Conclusion

In this paper, three deep learning-augmented inverse scattering schemes were compared for TWI applications. By utilizing the images from initial DBIM iterations as inputs, a traditional CNN regresses scatterer coordinates and the permittivity profile, while the U-net refines the input and separates scatterers from the background, and Mask R-CNN detects scatterers from the low-resolution input without producing boundary defects. It was shown that U-net and Mask R-CNN-based schemes outperform the traditional CNN-augmented algorithm in terms of accuracy, image quality, and generalization. U-net outperforms the other two quantitatively, while Mask R-CNN is capable of removing defects in retrieved scatterers using shape constraints.

Acknowledgements

This work was conducted during the PhD studies of Xiaofan Jia under the supervision of Prof. Abdulkadir C. Yucel (the main/primary author). The work was supported by Nanyang Technological University via a Start-Up Grant (Award No. 001096-00001) awarded to Prof Yucel. The authors are thankful to Dr. Luis J. Gomez for fruitful discussions during the implementation of the distorted Born iterative method.

References

[1] Chu, Y., Xu, K., Zhong, Y., Ye, X., Zhou, T., Chen, X., and Wang, G., Fast microwave through wall imaging method with inhomogeneous background based on Levenberg–Marquardt algorithm. **IEEE Transactions on Microwave Theory and Techniques**, 67, 1138–1147, (2019).

[2] Chen, R., Wei, Z., and Chen, X., Three dimensional through-wall imaging: Inverse scattering problems with an inhomogeneous background medium. **Proceedings, 2015 IEEE 4th Asia-Pacific Conference on Antennas and Propagation (APCAP)**, 505–506, (2015).

[3] Chen, X., Wei, Z., Li, M., and Rocca, P., A review of deep learning approaches for inverse scattering problems. **Progress In Electromagnetics Research**, 167, 67–81, (2020).

[4] Wei, Z., and Chen, X., Deep-learning schemes for full-wave nonlinear inverse scattering problems. **IEEE Transactions on Geoscience and Remote Sensing**, 57, 1849–1860, (2019).

[5] Slaney, M., Kak, A., and Larsen, L., Limitations of imaging with first-order diffraction tomography. **IEEE Transactions on Microwave Theory and Techniques**, 32, 860–874, (1984).

[6] Xiao, J., Li, J., Chen, Y., Han, F., and Liu, Q. H., Fast electromagnetic inversion of inhomogeneous scatterers embedded in layered media by Born approximation and 3-D U-Net. **IEEE Geoscience and Remote Sensing Letters**, (2019).

[7] Wang, Y., and Chew, W. C., An iterative solution of the two-dimensional electromagnetic inverse scattering problem. **International Journal of Imaging Systems and Technology**, 1, 100–108, (1989).

[8] Chew, W. C., and Wang, Y., Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method. **IEEE Transactions on Medical Imaging**, 9, 218–225, (1990).

[9] Van Den Berg, P. M., and Kleinman, R. E., A contrast source inversion method. **Inverse Problems**, 13, 1607, (1997).

[10] Yao, Y. M., Sha, W. E. I., and Jiang, L., Two-step enhanced deep learning approach for electromagnetic inverse scattering problems. **IEEE Antennas and Wireless Propagation Letters**, 18, 2254–2258, (2019).

[11] Wei, Z., and Chen, X., Physics-inspired convolutional neural network for solving full-wave inverse scattering problems. **IEEE Transactions on Antennas and Propagation**, 67, 6138–6148, (2019).

[12] Sun, J., Niu, Z., Innanen, K. A., Li, J., and Trad, D. O., A theory-guided deep-learning formulation and optimization of seismic waveform inversion. **Geophysics**, 85, R87–R99, (2020).

[13] Guo, R., Song, X., Li, M., Yang, F., Xu, S., and Abubakar, A., Supervised descent learning technique for 2-D microwave imaging. **IEEE Transactions on Antennas and Propagation**, 67, 3550–3554, (2019).

[14] Chen, G., Shah, P., Stang, J., and Moghaddam, M., Learning-assisted multi-modality dielectric imaging. **IEEE Transactions on Antennas and Propagation**, 68, 2356–2369, (2019).

[15] Sanghvi, Y., Kalepu, Y., and Khankhoje, U. K., Embedding deep learning in inverse scattering problems. **IEEE Transactions on Computational Imaging**, 6, 46–56, (2020).

[16] Sun, Y., Xia, Z., and Kamilov, U. S., Efficient and accurate inversion of multiple scattering with deep learning. **Optics Express**, 26, 14678–14688, (2018).

[17] Li, L., Wang, L. G., Teixeira, F. L., Liu, C., Nehorai, A., and Cui, T. J., DeepNIS: Deep neural network for nonlinear electromagnetic inverse scattering. **IEEE Transactions on Antennas and Propagation**, 67, 1819–1825, (2019).

[18] Khoshdel, V., Ashraf, A., and LoVetri, J., Enhancement of multimodal microwave-ultrasound breast imaging using a deep-learning technique. **Sensors**, 19, 4050, (2019).

[19] Ye, X., Bai, Y., Song, R., Xu, K., and An, J., An inhomogeneous background imaging method based on generative adversarial network. **IEEE Transactions on Microwave Theory and Techniques**, (2020).

[20] Xu, K., Qian, Z., Zhong, Y., Su, J., Gao, H., and Li, W., Learning-assisted inversion for solving nonlinear inverse scattering problem. **IEEE Transactions on Microwave Theory and Techniques**, 71, 2384–2395, (2022).

[21] Hu, H.-J., Xiao, L.-Y., Li, J., and Liu, Q. H., A hybrid forward-inverse neural network with the transceiver-configuration-independent technique for the wideband electromagnetic inverse scattering problem. **IEEE Transactions on Geoscience and Remote Sensing**, 62, 1–11, (2024).

[22] Liu, Z., Roy, M., Prasad, D. K., and Agarwal, K., Physics-guided loss functions improve deep learning performance in inverse scattering. **IEEE Transactions on Computational Imaging**, 8, 236–245, (2022).

[23] Zhou, H., Cheng, Y., Zheng, H., Liu, Q., and Wang, Y., Deep unfolding contrast source inversion for strong scatterers via generative adversarial mechanism. **IEEE Transactions on Microwave Theory and Techniques**, 70, 4966–4979, (2022).

[24] Yin, T., Wang, C.-F., Xu, K., Zhou, Y., Zhong, Y., and Chen, X., Electric flux density learning method for solving three-dimensional electromagnetic scattering problems. **IEEE Transactions on Antennas and Propagation**, (2022).

[25] Guo, R., Lin, Z., Shan, T., Song, X., Li, M., Yang, F., Xu, S., and Abubakar, A., Physics embedded deep neural network for solving full-wave inverse scattering problems. **IEEE Transactions on Antennas and Propagation**, 70, 6148–6159, (2021).

[26] Yao, H. M., Guo, R., Li, M., Jiang, L., and Ng, M. K. P., Enhanced supervised descent learning technique for electromagnetic inverse scattering problems by the deep convolutional neural networks. **IEEE Transactions on Antennas and Propagation**, 70, 6195–6206, (2022).

[27] Zhou, H., Xu, C., Cheng, Y., Liu, Q., Hu, S., and Wang, Y., Self-supervised learning of physics-guided deep unfolding parallel networks for solving nonlinear inverse scattering problems. **IEEE Transactions on Microwave Theory and Techniques**, 72, 5204–5217, (2024).

[28] Huang, Q., Li, C., Ye, X., Xu, K., and Song, R., Meta-learning-assisted untrained neural network for electromagnetic inverse scattering problems. **IEEE Transactions on Antennas and Propagation**, (2025).

[29] Song, R., Li, M., Xu, K., Li, C., and Chen, X., Electromagnetic inverse scattering with an untrained SOM-net. **IEEE Transactions on Microwave Theory and Techniques**, 70, 4980–4990, (2022).

[30] Hu, Y.-D., Wang, X.-H., Zhou, H., and Wang, L., A priori knowledge-based physics-informed neural networks for electromagnetic inverse scattering. **IEEE Transactions on Geoscience and Remote Sensing**, 62, 1–9, (2024).

[31] LeCun, Y., Bengio, Y., and Hinton, G., Deep learning. **Nature**, 521, 436–444, (2015).

[32] Ronneberger, O., Fischer, P., and Brox, T., U-Net: Convolutional networks for biomedical image segmentation. **Proceedings, International Conference on Medical Image Computing and Computer-Assisted Intervention**, 234–241, Springer International Publishing, Cham, (2015).

[33] He, K., Gkioxari, G., Dollár, P., and Girshick, R., Mask R-CNN. **Proceedings, IEEE International Conference on Computer Vision**, 2961–2969, (2017).

[34] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A., Automatic differentiation in PyTorch. **Proceedings, 31st Conference on Neural Information Processing Systems**, (2017).

[35] Kingma, D. P., and Ba, J., Adam: A method for stochastic optimization. **arXiv Preprint arXiv:1412.6980**, (2014).

[36] Jadon, S., A survey of loss functions for semantic segmentation. **arXiv Preprint arXiv:2006.14822**, (2020).

[37] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., ImageNet: A large-scale hierarchical image database. **Proceedings, 2009 IEEE Conference on Computer Vision and Pattern Recognition**, 248–255, (2009).

[38] Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K., Aggregated residual transformations for deep neural networks. **Proceedings, IEEE Conference on Computer Vision and Pattern Recognition**, 1492–1500, (2017).

[39] Wu, Y., Kirillov, A., Massa, F., Lo, W., and Girshick, R., Detectron2. **GitHub**, https://github.com/facebookresearch/detectron2, (2019).