Deep Learning Schemes for Full-Wave Nonlinear Inverse Scattering Problems

Zhun Wei and Xudong Chen

Abstract—The paper is devoted to solving a full-wave inverse scattering problem, which is aimed at retrieving permittivities of dielectric scatterers from the knowledge of measured scattering data. Inverse scattering problems are highly nonlinear due to multiple scattering, and iterative algorithms with regularizations are often used to solve such problems. However, they are associated with heavy computational cost, and consequently they are often time consuming. This paper proposes the convolution neural network (CNN) technique to solve full-wave inverse scattering problems. We introduce and compare three training schemes based on U-Net CNN including direct inversion, back-propagation, and dominant current schemes (DCS). Several representative tests are carried out, including both synthetic and experimental data, to evaluate the performances of the proposed methods. It is demonstrated that the proposed DCS outperforms the other two schemes in terms of accuracy and is able to solve typical inverse scattering problems fastly within one second. The proposed deep learning inversion scheme is promising in providing quantitative images in real time.

Index Terms—Inverse scattering, convolution neural network, dielectric scatterers.

I. INTRODUCTION

INVERSE scattering problems (ISP) are concerned with determining either physical or geometrical properties of scatterers from the measured scattered fields. The inverse scattering technique has wide applications in various fields, such as estimating physical parameters from the observations of external or internal radiant energy in remote sensing [1], detecting human organs and biological systems in biomedical imaging and diagnosis [2]. One of the most important advantages of inversion approaches is that it avoids destructive evaluation. In order to detect the inhomogeneities in a medium, one only needs to collect scattered fields outside the medium non-destructively. Due to the intrinsic ill-posedness and nonlinearity of ISP, iterative optimization methods with regularizations are usually used to solve ISP, such as Born iterative method [3], distorted Born iterative method [4], contrast source-type inversion (CSI) method [5]–[8], and subspace optimization method (SOM) [9]–[11]. By minimizing the objective function that quantifies the mismatch between calculated and measured data iteratively, properties of the unknown scatterers are reconstructed. The main drawback of these iterative methods is that they are time consuming and thus not suitable for real-time reconstruction. Under some conditions, the ISP can also be
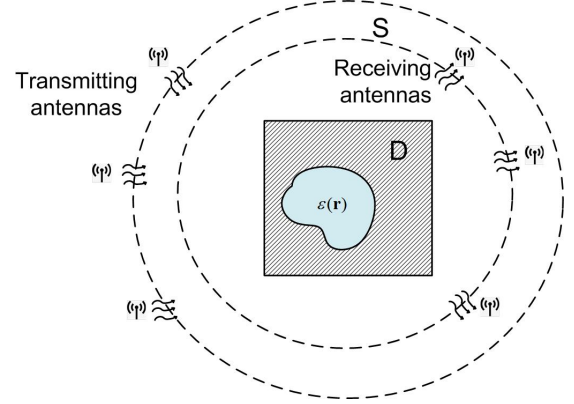
Fig. 1. A typical schematic of inverse scattering problems.

approximately solved by noniterative methods. For example, the inverse problem can be decomposed into several linear equations and each linear equation can be solved without iteration in the extended Born approximation method [12] and back-propagation method [13]. Noniterative methods are able to provide reconstruction results in a quite short time period, but the results are not accurate, especially for strong scatterers.

Methodologies based on artificial neural network have also been proposed and applied to extract rather general information about the geometric and electromagnetic properties of scatterers [14], [15]. Nevertheless, most of these methods use a few parameters to represent scatterers, such as their positions, sizes, shapes, and piece-wise constant permittivities. However, the scope of such a parameterization approach is limited, since the number of scatterers can be arbitrary and the scatterers can be spatially inhomogeneous. A more versatile approach to represent scatterers is to use pixel basis, i.e., the value of permittivity of each pixel is an independent parameter. Recently, deep learning has attracted intensive attention for providing state-of-the-art performance for image classification [16], [17] and segmentation [18], [19]. Neural networks with regression features have also provided impressive results on inverse problems, such as signal denoising [20], deconvolution [21] and interpolation [22], [23]. In [19], a U-net deep convolution neural network has been designed for segmentation problems, and it has been further applied to solve an ill-posed linear equation $b = Kx$ in [24], where K, x, and b are the measurement operator, unknowns and measurement data, respectively.

In this paper, we have proposed three inversion schemes based on U-net CNN to solve nonlinear inverse scattering problems, where scatterers are represented by pixel bases.

First, a direct inversion scheme (DIS) is proposed, where complex scattered fields are directly used to retrieve profiles of relative permittivities. Second, by combining the CNN with the non-iterative algorithm, a back-propagation scheme (BPS) is proposed. Third, considering the drawbacks in the BPS, a dominant current scheme (DCS) is further devised. We test and compare the three schemes, and show promising results from both synthetic data (circular-cylinder and MNIST datasets [25]) and experimental data. It is found from the tests that DCS shows significant advantages over the other two schemes, and is able to reconstruct images with typical sizes within 1 second.

This paper is organized as follows. In section II, formulation of the problem is introduced. In section III, we proposed three CNN-based schemes to solve nonlinear inverse scattering problems, where scatterers are represented by pixel bases. In sections IV and V, tests with numerical and experimental data are respectively conducted. Finally, we summarize our work in section VI.

It is worth introducing the notations used throughout the paper. We use $\overline{\overline{X}}$ and $\overline{X}$ to denote the matrix and vector of the discretized operator or parameter X, respectively. We use the operation diag($\cdot$) to denote diagonalizing a vector to a diagonal matrix. Furthermore, the superscripts $*$, T, and H denote the complex conjugate, transpose, and conjugate transpose of a matrix or vector, respectively. Finally, we use $||\cdot||$ and $||\cdot||_F$ to denote Euclidian length of a vector and Frobenius norm of a matrix, respectively.

II. FORMULATION OF THE PROBLEM

The configuration under study is presented in Fig. 1. For convenience of presentation, we consider a two-dimensional (2-D) transverse-magnetic (TM) case, where the longitude direction is along $\hat{z}$. In free-space background, nonmagnetic scatterers are located in a domain of interest (DOI) $D \subset R^2$, and illuminated by N_i line sources located at $\mathbf{r}_p^i$ with $p = 1, 2, ..., N_i$. For each incidence, the scattered field is measured by N_r antennas located at $\mathbf{r}_q^s$ with $q = 1, 2, ..., N_r$.

The forward problem is described by two equations. The first one is the electric field integral equation (EFIE), which is also known as the Lippmann-Schwinger equation,

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

where $E^t(\mathbf{r})$ and $E^i(\mathbf{r})$ denote total and incident electric field, respectively. $k_0 = \omega\sqrt{\mu_0 \varepsilon_0}$ is the wavenumber of homogeneous medium background, and $g(\mathbf{r}, \mathbf{r}')$ is the 2-D free space Green's function. The contrast current density $I(\mathbf{r})$ is defined as $I(\mathbf{r}) = \xi(\mathbf{r}) E^t(\mathbf{r})$ with the contrast $\xi(\mathbf{r}) = \varepsilon_r(\mathbf{r}) - 1$. The second equation of the forward problem is described as,

$$E^s(\mathbf{r}) = k_0^2 \int_D g(\mathbf{r}, \mathbf{r}') I(\mathbf{r}') d\mathbf{r}', \text{ for } \mathbf{r} \in S, \quad (2)$$

where $E^s(\mathbf{r})$ is the scattered field on the measurement surface S. The first equation in forward problem describes the wave-scatterer interaction in domain D and is usually called as state equation. The second equation describes the scattered field as a re-radiation of the induced contrast current and is called as data equation.

For inverse problems, the goal is to reconstruct the relative permittivities $\varepsilon_r(\mathbf{r})$ ($r \in D$) from the measured scattered field $E_p^s(\mathbf{r})$ of N_i incidences. If we denote Ψ as the operator of solving the corresponding forward problem, the nonlinear equation [26]:

$$E_p^s = \Psi(\varepsilon_r) \quad (3)$$

does not have an exact solution of ε_r in the presence of noise. Instead of directly solving (3), an objective function for optimization problem is usually constructed as

$$\text{Min} : f(\varepsilon_r) = \sum_{p=1}^{N_i} ||\Psi(\varepsilon_r) - E_p^s||^2 + \alpha T(\varepsilon_r), \quad (4)$$

where $T(\varepsilon_r)$ is the regularization used to balance in data fitting and stability of the solution. α is a constant regularization coefficient. It is well-known that (4) is nonlinear and nonconvex, and difficult to solve due to the presence of local minima [26]–[28]. There are many iterative optimization algorithms proposed to solve the nonlinear optimization problems, such as [3]–[5], [11]. In this paper, we focus on inverse scattering problems which exclude scenarios where high frequency parts of the objects are dominant, such as corrugated ones.

III. CNN SCHEMES FOR NONLINEAR INVERSE PROBLEMS

Before introducing CNN schemes to solve ISP, it is convenient to write both the state and data equations in discretized forms. Specifically, we use the method of moment (MOM) with the pulse basis function and the delta testing function to discretize the domain D into $M \times M$ subunits [29], and the centers of subunits are located at $\mathbf{r}_n$ with $n = 1, 2, ..., M^2$. We obtain the following discretized forms of (1) and (2), respectively:

$$\overline{E}^t = \overline{E}^i + \overline{\overline{G}}_D \cdot \overline{I} \quad (5)$$

and

$$\overline{E}^s = \overline{\overline{G}}_S \cdot \overline{I}, \quad (6)$$

where the vectors $\overline{E}^t$, $\overline{E}^i$, $\overline{E}^s$, and $\overline{I}$ are M^2 dimensions with the nth element being $\overline{E}^t(n) = E^t(\mathbf{r}_n)$, $\overline{E}^i(n) = E^i(\mathbf{r}_n)$, $\overline{E}^s(n) = E^s(\mathbf{r}_n)$, and $\overline{I}(n) = I(\mathbf{r}_n)$, respectively. The matrix $\overline{\overline{G}}_D$ and $\overline{\overline{G}}_S$ are $M^2 \times M^2$ and $N_r \times M^2$ dimensions with $\overline{\overline{G}}_D(n, n') = k_0^2 A_{n'} g(\mathbf{r}_n, \mathbf{r}_{n'})$ and $\overline{\overline{G}}_S(q, n') = k_0^2 A_{n'} g(\mathbf{r}_q, \mathbf{r}_{n'})$, respectively. Here, $A_{n'}$ is the area of the n'th subunits, and we have $n = 1, 2, ..., M^2$, $n' = 1, 2, ..., M^2$, and $q = 1, 2, ..., N_r$.

Similarly, the contrast $\xi(\mathbf{r})$ of domain D is discretized into an M^2-dimensional vector $\overline{\xi}$ with $\overline{\xi}(n) = \xi(\mathbf{r}_n)$. For the sake of convenience, we also use another discretized form of the contrast in the paper, which is a matrix $\overline{\overline{\xi}}$ with $M \times M$ dimensions directly reshaped from $\overline{\xi}$. Further, from the definition of contrast current density $I(\mathbf{r})$, the discretized forms of $I(\mathbf{r})$ can be computed as $\overline{I} = \text{diag}(\overline{\xi}) \cdot \overline{E}^t$.

A. Direct Inversion Scheme (DIS)

To solve ISP with CNN, the first thought would be a regression of unknowns directly from measured data, which is referred to as the direct inversion scheme (DIS). In DIS for inverse scattering problems, the inputs of the CNN are the complex values of scattered field and the outputs are the contrasts $\overline{\overline{\xi}}$ of domain D. This scheme is similar to that used in some previous literatures such as [14], but the proposed direct inversion scheme (DIS) differs in two aspects. The first one is that the outputs of DIS are the unknowns at each pixel of D, while the outputs of [14] are only four or three unknown variables that are used to parameterize scatterers, such as radii or positions of cylinders. The second is that CNN is used in the proposed DIS, which is a deep-learning network involving many hidden layers, and thus it is different from the single hidden-layer neural network as done in [14].

B. Back-Propagation Scheme (BPS)

Although it is possible to train a CNN to regress directly from the measured scattered field $E_p^s(\mathbf{r})$ to relative permittivity ε_r, the learning process can be simplified by performing noniterative inversion algorithms first before training. In this section, a back-propagation scheme (BPS) based on CNN for inverse scattering problem is proposed.

The BPS assumes that induced current in BPS I^b is proportional to the back-propagated field, which is an effective approach in reconstructing weak-scattering scatterers [13], [26]:

$$\overline{I}^b = \chi \cdot \overline{\overline{G}}_S^H \cdot \overline{E}^s. \tag{7}$$

A cost function F^b based on (6) can be defined as:

$$F^b(\chi) = ||\overline{E}^s - \overline{\overline{G}}_S \cdot (\chi \cdot \overline{\overline{G}}_S^H \cdot \overline{E}^s)||^2. \tag{8}$$

The minimum of $F(\chi)$ requires the derivative of $F(\chi)$ with respect to χ to be zero, and an analytical solution of χ can be obtained:

$$\chi = \frac{(\overline{E}^s)^T \cdot (\overline{\overline{G}}_S \cdot (\overline{\overline{G}}_S^H \cdot \overline{E}^s))^*}{||\overline{\overline{G}}_S \cdot (\overline{\overline{G}}_S^H \cdot \overline{E}^s)||}. \tag{9}$$

With χ, the induced current $\overline{I}^b$ can be derived from (7), and the updated total electrical field of BPS $\overline{E}^{t,b}$ can consequently be obtained from (5) with

$$\overline{E}^{t,b} = \overline{E}^i + \overline{\overline{G}}_D \cdot \overline{I}^b. \tag{10}$$

For each incidence p, the contrast $\overline{\xi}_p^b$ and $\overline{I}_p^b$ satisfy the following relationship based on the definition of induced current:

$$\overline{I}_p^b = \text{diag}(\overline{\xi}_p^b) \cdot \overline{E}_p^{t,b}. \tag{11}$$

Combining all incidences of (11) leads to a least squares problem, and an analytical solution can be obtained for the nth element of the contrast $\overline{\xi}^b(n)$ with

$$\overline{\xi}^b(n) = \frac{\sum_{p=1}^{N_i} \overline{I}_p^b(n) \cdot [\overline{E}_p^{t,b}(n)]^*}{\sum_{p=1}^{N_i} ||\overline{E}_p^{t,b}(n)||^2}. \tag{12}$$

In the learning process of BPS, the contrasts $\overline{\overline{\xi}}^b$ (reshaped from $\overline{\xi}^b$) are the inputs of the CNN, and the outputs are the true contrasts of the domain D. We mention in passing two issues to avoid any possible confusion. The first is that the introduced BPS is valid for arbitrary incidence fields and for both near and far filed measurements, which is different from the filtered back-propogation (FBP) in diffraction tomography (DT). The second is that the term "back-propgation" in BPS should be distinguished from the error back-propgation algorithm with chain rule calculation used to update weights in CNN.

C. Dominant Current Scheme (DCS)

The motivation of dominant current scheme (DCS) lies in the fact that, for nonlinear problem, it is difficult to directly solve a large number of unknowns from limited measurements. If one can obtain a solution for the dominant part of unknowns, and use the solution as an initial value, the problem would be easier to solve.

In the first step of DCS, we turn our attention from directly computing contrast to obtaining a dominant component of induced current. The dominant current should have two features. The first one is that it contains most of the important features of the unknown objects, and the second is that it should be robust in noisy environment.

If a singular value decomposition is conducted on $\overline{\overline{G}}_S$, one can obtain $\overline{\overline{G}}_S = \sum_n \overline{u}_n \sigma_n \overline{\nu}_n^H$ with $\sigma_1 \geq \sigma_2 ... \geq \sigma_{M^2} \geq 0$, where $\overline{\mu}_n$, $\overline{\nu}_n$, and σ_n are the left, right singular vectors, and singular values of $\overline{\overline{G}}_S$. By considering orthogonality of the singular vectors and the relationship

$$\overline{\overline{G}}_S \cdot \overline{\nu}_n = \sigma_n \overline{u}_n, \tag{13}$$

an induced current $\overline{I}^+$ from the first L dominant singular values can be uniquely calculated from the data equation (6) with

$$\overline{I}^+ = \sum_{j=1}^{L} \frac{\overline{\mu}_j^H \cdot \overline{E}^s}{\sigma_j} \overline{\nu}_j. \tag{14}$$

Since the first L singular values are larger than the remaining ones, the induced currents calculated from (14) are more stable when the measured scattered field $\overline{E}^s$ is contaminated by noise. Nevertheless, $\overline{I}^+$ has missed some information in the current spanned by the other $M^2 - L$ singular values I^-, where the actual current should be $\overline{I} = \overline{I}^+ + \overline{I}^-$. To compensate the missing information from $\overline{I}^-$, we modify the induced current $\overline{I}^+$ as the dominant current $\overline{I}^d$ in DCS by

$$\overline{I}^d = \overline{I}^+ + \overline{I}^l, \tag{15}$$

where the low-frequency induced current $\overline{I}^l$ is represented by:

$$\overline{I}^l = \overline{\overline{F}}^l \cdot \overline{\alpha}^l \tag{16}$$

with $\overline{\overline{F}}^l$ and $\overline{\alpha}^l$ be low-frequency matrix composed by the first m low-frequency Fourier bases and their corresponding coefficients, respectively. It is noted that, in this paper, we directly use 1D Fourier bases and the currents have been reshaped to vectors. Both L and m are constant empirical
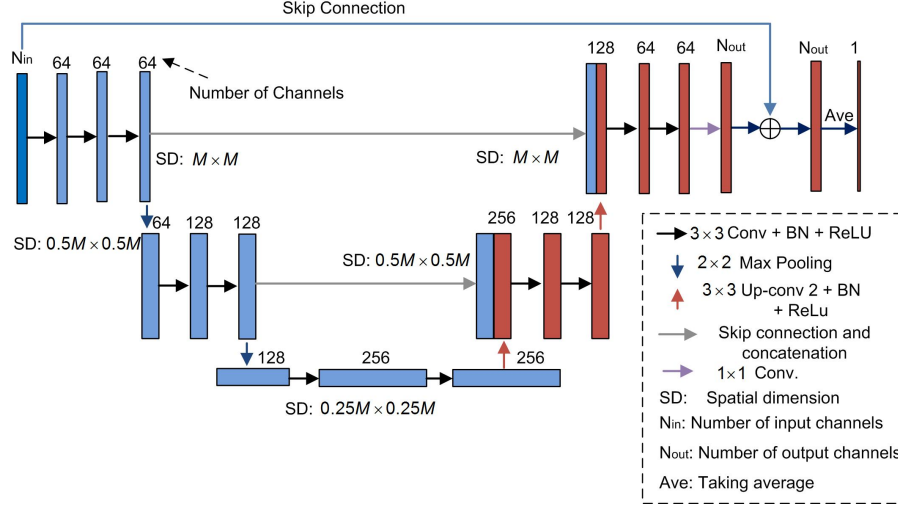
Fig. 2. The U-net architecture for the proposed three CNN schemes: DIS, BPS, and DCS, where BN denotes batch normalization. The details of U-net can be found in [19], [24].

numbers, and the method to choose them will be introduced in Section IV.

The second step of DCS is to obtain an approximate value of $\overline{\alpha}^l$. If we multiply both side of (5) by $\mathrm{diag}(\overline{\xi})$, a state equation based on induced current is obtained

$$\overline{I} = \mathrm{diag}(\overline{\xi}) \cdot [\overline{E}^i + \overline{\overline{G}}_D \cdot \overline{I}]. \tag{17}$$

By replacing the induced current $\overline{I}$ in both state equation, (17), and data equation, (6), with the dominant induced current $\overline{I}^d$, a cost function with respect to $\overline{\alpha}^l$ is obtained by considering both normalized state equation and data equation:

$$F(\overline{\alpha}^l) = [\frac{||\overline{\overline{G}}_f \cdot \overline{\alpha}^l - \overline{G}_e||^2}{||\overline{E}^s||^2} + \frac{||\overline{\overline{A}} \cdot \overline{\alpha}^l - \overline{B}||^2}{||\overline{I}^+||^2}], \tag{18}$$

where $\overline{\overline{G}}_f = \overline{\overline{G}}_S \cdot \overline{\overline{F}}^l$, $\overline{G}_e = \overline{\overline{G}}_S \cdot \overline{I}^+ - \overline{E}^s$, $\overline{\overline{A}} = \overline{\overline{F}}^l - \mathrm{diag}(\overline{\xi}_b) \cdot (\overline{\overline{G}}_D \cdot \overline{\overline{F}}^l)$, and $\overline{B} = \mathrm{diag}(\overline{\xi}_b) \cdot (\overline{E}^i + \overline{\overline{G}}_D \cdot \overline{I}^+) - \overline{I}^+$. Here, we start by assuming $\overline{\xi} = \overline{\xi}^b$ and consequently the cost function depends solely on $\overline{\alpha}^l$.

To avoid matrix inversion, an approximate analytical solution of $\overline{\alpha}^l$ is given by minimizing the cost function (18) along the gradient direction $\overline{\rho}$ with

$$\overline{\alpha}^l = \beta\overline{\rho}, \tag{19}$$

The gradient direction is calculated as

$$F'(\overline{\alpha}^l = 0) = \overline{\rho} = (\frac{\overline{\overline{G}}_f^H \cdot \overline{G}_e}{||\overline{E}^s||^2} + \frac{-\overline{\overline{A}}^H \cdot \overline{B}}{||\overline{I}^+||^2}), \tag{20}$$

and the objective function is quadratic with respect to β and its minimizer is given by $\beta = Num/Den$. Here, the numerator and denominator are respectively

$$Num = \frac{-(\overline{\overline{G}}_f \cdot \overline{\rho})^H \cdot \overline{G}_e}{||\overline{E}^s||^2} + \frac{(\overline{\overline{A}} \cdot \overline{\rho})^H \cdot \overline{B}}{||\overline{I}^+||^2}, \tag{21}$$

and

$$Den = \frac{||\overline{\overline{G}}_f \cdot \overline{\rho}||^2}{||\overline{E}^s||^2} + \frac{||\overline{\overline{A}} \cdot \overline{\rho}||^2}{||\overline{I}^+||^2}. \tag{22}$$

With $\overline{\alpha}^l$, the dominant induced current $\overline{I}^d$ is easily derived following (15). Then, the final step is to calculate contrast $\overline{\xi}_p^d$ of DCS based on $\overline{I}^d$. Specifically, according to (5), the total electrical field of DCS $\overline{E}_p^{t,d}$ for the pth incidence can be updated as:

$$\overline{E}_p^{t,d} = \overline{E}_p^i + \overline{\overline{G}}_D \cdot \overline{I}_p^d. \tag{23}$$

Then, based on the definition

$$\overline{I}_p^d = \mathrm{diag}(\overline{\xi}_p^d) \cdot \overline{E}_p^{t,d}, \tag{24}$$

the nth element of the contrast $\overline{\xi}_p^d$ for the pth incidence is obtained with

$$\overline{\xi}_p^d(n) = \frac{\overline{I}_p^d(n) \cdot [\overline{E}_p^{t,d}(n)]^*}{||\overline{E}_p^{t,d}(n)||^2}. \tag{25}$$

In the learning process of DCS, the contrasts of each incidence $\overline{\overline{\xi}}_p^d$ (reshaped from $\overline{\xi}_p^d$) are put into different input-channels of CNN, and each corresponding output-channel is the true contrasts $\overline{\overline{\xi}}$ of the domain D. Consequently, there are N_i pairs of input- and output-channels of DCS, whereas there is only one pair of input- and output-channel for DIS and BPS.
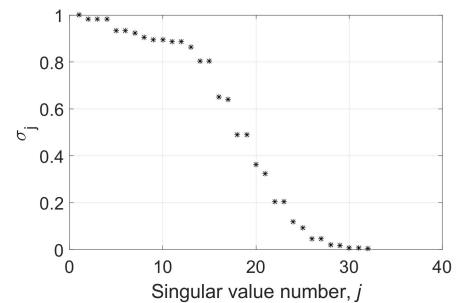

Fig. 3. The singular values of G_S normalized to its maximum value.
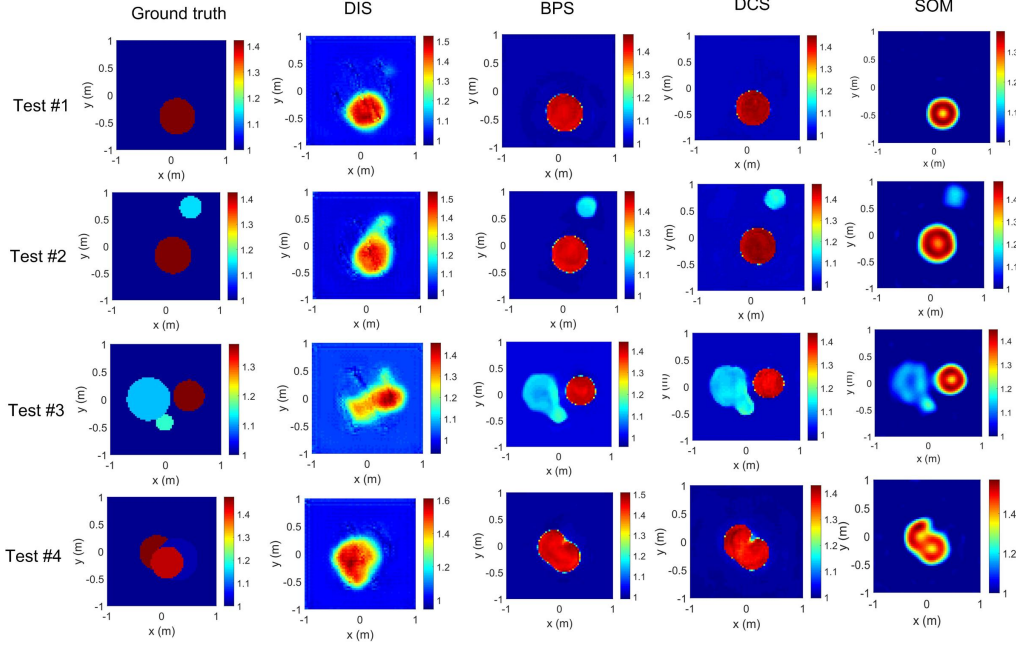
Fig. 4. Example One: Reconstructed relative permittivity profiles from scattered fields with 5% noise for DIS, BPS, DCS, and iterative method (SOM), where the relative permittivity is between 1 and 1.5. The first column shows the ground truth images for 4 representative tests.

It is important to note that, in (15), we construct dominant current $\overline{I}^d$ from $\overline{I}^l$ instead of $\overline{I}^-$ for the following reasons. Firstly, without the presence of I^-, only the first L singular values and vectors are needed in (14) and consequently, only a thin-type SVD is sufficient to obtain $\overline{I}^+$, where the computational cost are largely reduced. Secondly, most information of natural images is concentrated in low spatial frequency bands. By considering the low frequency band only, the number of unknowns is largely reduced. Furthermore, with the use of Fourier bases, fast Fourier transform (FFT) can be used directly in the subsequent calculation with low computational cost. Lastly, studies have shown that deep architecture properties of CNNs, namely their strong learning capability and high representational capacity, are well suited to image restoration from degraded inputs [30]. Thus, it is a good choice that we exclude the high frequency part of induced currents that is easily contaminated by noise from the inputs of CNNs, and let CNNs to restore this part by training.

D. U-net Convolutional Neural Network

In this paper, we implement the proposed three CNN schemes using the U-net architecture, which is originally designed for segmentation [19]. The U-net CNN has also been applied to solve the linear equation $b = Kx$ in [24]. In this paper, U-net is used to solve full-wave nonlinear inverse scattering problems.

As presented in Fig. 2, the U-net architecture consists of a contracting path (left side) and an expansive path (right side). The contracting path consists of repeated application of 3×3 convolutions, batch normalization, and rectified linear unit (ReLU), which is followed by a 2×2 max pooling operation. The expansive path is similar to contracting path, but the max pooling in contracting path is replaced by a 3×3

up convolution in expansive path. In expansive path, there are also two concatenations with the correspondingly cropped feature maps from the contracting path. For DCS, the number of output channels N_{out} in Fig. 2 is equal to the number of incidences N_i, and an average operation is consequently conducted on the outputs of all incidences to obtain the final results.

The U-net CNN architecture is well suited for the the ISP for the following reasons:

- A skip connection is inserted from the input of the neural network to its output layer in the U-net architecture. This approach is particular well suited to BPS and DCS proposed in the manuscript since the inputs and outputs share similar important features, and it also mitigates the vanishing gradient problem during training [24], [31].
- Due to downsampling in contracting path of U-net, the effective receptive field of the network increases significantly, which is of paramount importance for ISP since a large field of view over the input image can significantly improve the prediction at each pixel of the output image [32].
- Batch Normalization (BN) is used to alleviate the internal covariate shift, i.e., the change in the distribution of network activations due to the change in network parameters during training [33]. Specifically, through a normalization step that fixes the means and variances of layer inputs, BN offers a much faster training and reduces the dependence of gradients on the scale of the parameters or their initial values [33]. Consequently, the robustness of the architecture to initializations is increased.

Further, we have modified the structure of U-net, and applied a multiple-channel scheme (MCS) in DCS. The MCS adopted in DCS can be comprehended as a kind of "da-

TABLE I
RELATIVE ERRORS R_e FOR THE TESTS. NA REPRESENTS THE SITUATION
WHERE THE SCHEME FAILS TO OBTAIN ANY MEANINGFUL RESULTS.

CNN schemes	DIS	BPS	DCS	SOM
Example 1	5.8%	1.7%	1.8%	3%
Example 2	14%	6.5%	5%	6.4%
Example 3	NA	11.4%	9.7%	6%
"Austria" Tests	NA	10.4%	8.5%	7.6%
Lossy scatterers	NA	6.7%	5.5%	9.47%

ta augmentations". In machine learning, data augmentation (DA) [34] is a process to reduce overfitting and increase robustness by artificially enlarging the dataset using label-reserving transformations, such as mirroring image in the horizontal and vertical directions during training. MCS shares the properties of enlarging data set and increasing robustness with DA. Specifically, The MCS adopted in DCS enlarges the data set by utilizing the information from all incidences and increases the robustness by taking average of all outputs in different channels. Nevertheless, MCS differs from DA in two aspects. Firstly, the outputs of DA change accordingly with the transformation of inputs, whereas the outputs in each channel of MCS are kept the same in spite of the change of the inputs in each channel (here in DCS, an incidence corresponds to one pair of input- and output-channel). Secondly, data from DA is used as a separate training set and applied independently in the training process. In MCS, data in all channels are applied simultaneously in a single process.

E. Computational complexity

To compare the performance of different methods, one of the crucial criteria is the the computational complexity. Since the DOI is descretized into $M \times M$ pixels, the computational cost C_1 for iterative algorithms to solve (4) can be described as [26]

$$C_1 = O(N_o N_i N_f M^2 \log M^2), \tag{26}$$

where $M^2 \log M^2$ is the computational cost of matrix-vector multiplication with FFT in each iteration of the forward solver, and N_f is the number of iterations for solving the forward problem. N_o denotes the number of iterations during optimizations. In iterative methods, the key to solve inverse scattering problems with low computational complexities is to reduce the value of N_o since the value of other parameters in (26) can barely be reduced.

For DIS, the computational cost includes the basic operations in CNN, such as convolutions, additions, calculations of ReLU function, and max poolings. Among them, the complexity is dominated by convolutions. Specifically, if there are Q_i input feature maps, Q_o output feature maps, and the feature map size and convolution kernel size are $M \times M$ and $K_f \times K_f$ ($K_f = 3$ in this paper), respectively, the computational workload in the convolution layer is in the order of $O(M^2 K_f^2 Q_i Q_o)$ [35].

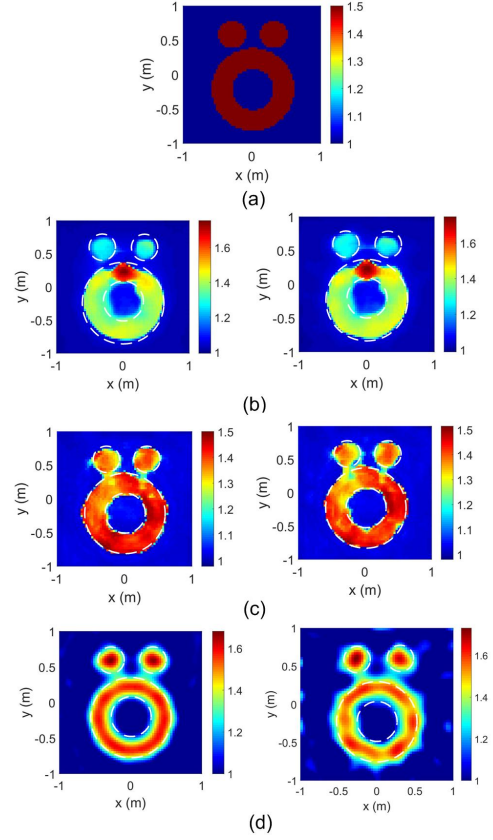To compute the input of BPS, the computational cost is dominant by computing $G_D(I)$, where the computational cost

Fig. 5. "Austria" tests: image reconstruction of "Austria" profile. (a) Ground truth profile of Austria. Reconstructed relative permittivity profiles for (b) BPS, (c) DCS, and (d) iterative method (SOM) with 5% (left) and 20% (right) Gaussian noise presented. It is noted that the dashed lines in both (b), (c), and (d) denote the true boundaries of cylinders in (a).

is $O(N_f M^2 \log M^2)$ if FFT is applied in the matrix-vector multiplication.

For DCS, there is an additional thin SVD on $\overline{\overline{G}}_S$ involved, and the computational cost of a thin SVD is $O(N_r^2 M^2)$ [36]. It is important to note that although we have $N_i - 1$ more input and output feature maps in the first and last convolution layers for the proposed DCS, no additional feature map is added in other convolution layers. It is known from Fig. 2, that there are 14 convolution layers in the CNN architecture of this paper, and the number of feature maps in other convolution layers except input and output layer are larger than N_i. Consequently, the overhead computation due to multiple incidences is not significant in DCS.

In CNN based inversion methods, an advantage is that each operation in the CNN is ideal for GPU-based parallelization since all operations are simple and local. Moreover, in the test process, only one forward process is needed to obtain the reconstructed results without iterations.

IV. NUMERICAL RESULTS

This section presents some results using synthetic data (circular-cylinder and MNIST datasets [25]) to evaluate the performance of the proposed CNN schemes in reconstructing relative permittivities from scattered field. In all tests, reconstructions are conducted at a single frequency, i.e., frequency
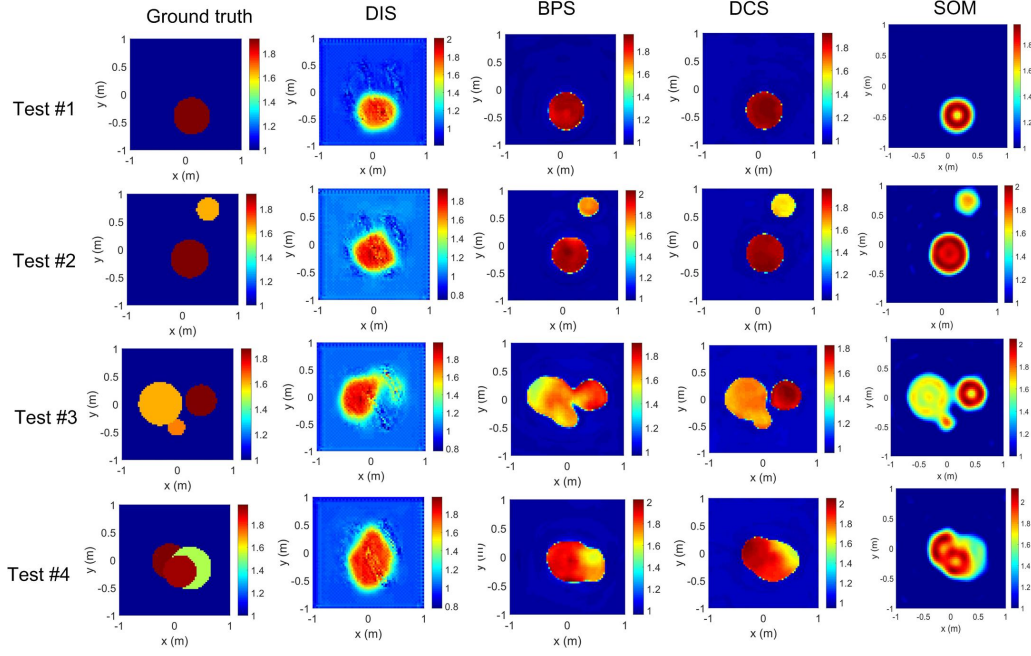
Fig. 6. Example Two: Reconstructed relative permittivity profiles from scattered fields with 5% noise for DIS, BPS, DCS, and iterative method (SOM), where the relative permittivity is between 1.5 and 2. The first column shows the ground truth images for 4 representative tests.

hoping technique is not involved. It is noted that, in this paper, we use the term "cylinder" to denote "circular-cylinder" in the following text for brevity.

A. Numerical setup

In the numerical tests, we consider a DOI D with the size of 2×2 m^2 and discretize the domain into 64×64 pixels. There are 16 line sources and 32 line receivers equally placed on a circle of radius 3 m centered at $(0, 0)$ m. For the scatterers in D, a cylinder dataset is synthetically generated comprising 500 images of cylinders of random relative permittivity, radius, number and location. It is important to stress that overlapping between cylinders is allowed. Due to the limited size of D, the radii of random generated cylinders are between 0.15 m and 0.4 m, and the number is between 1 and 3. Among the 500 profiles, 475 of them are used to train CNNs with the proposed schemes, and 25 images are used to test the trained CNNs. In numerical tests, we limit the relative permittivities in a range of 0.5 to better compare the results in the same level of nonlinearity. It is important to note that, the proposed CNN schemes also work for a large range of permittivities, and an example is given in the experimental validation part.

The scattered fields from N_i incidences are generated numerically using method of moment (MOM) and recorded into a matrix $\overline{\overline{E}}^s$ with the size of $N_r \times N_i$. Then, additive white Gaussian noise $\overline{\overline{n}}$ is added to $\overline{\overline{E}}^s$. The resultant noisy matrix $\overline{\overline{E}}^s + \overline{\overline{n}}$ is treated as the measured scattered field that is used to reconstruct relative permittivities, and the noise level is quantified as $(||\overline{\overline{n}}||_F / ||\overline{\overline{E}}^s||_F)$. The operating frequency is 400 MHz, and unless otherwise stated, *a priori* information is that the scatterers are lossless and have nonnegative contrast [6]. It is noted that parameters, such as number of measurements,

locations of line sources and receivers, and frequencies can slightly affect the performance. In this paper, we follow the settings of these parameters in the previous literature [9] for a better comparison.

In order to quantitatively evaluate the performance of different schemes, a relative error R_e is also defined:

$$R_e = \frac{1}{M_t} \sum_{j=1}^{M_t} ||\overline{\overline{\varepsilon}}_r^t - \overline{\overline{\varepsilon}}_r^r||_F / ||\overline{\overline{\varepsilon}}_r^t||_F \qquad (27)$$

where $\overline{\overline{\varepsilon}}_r^t$ and $\overline{\overline{\varepsilon}}_r^r$ are true and reconstructed relative permittivity profiles, respectively, and M_t is the number of tests conducted.

B. Implementation Details

For the empirical parameters L and m, usually a successive range of integer L and m, instead of a single one, works for a reconstruction problem. Specifically, for the value of L, it is determined by $\overline{\overline{G}}_s$ which is affected by measurement setup, such as electrical dimension of DOI and frequencies. Usually, a successive range of integer L ($8 \leq L \leq 25$ in our setup), instead of a single one, works for a reconstruction problem. The best performance is usually achieved based on the criterion that the Lth singular value σ_L is around 50% of the first singular value σ_1. As presented in Fig. 3, since σ_{18} is around 50% of σ_1, the best performance will be obtained when L is around 18. In this paper, L and m values are empirically chosen as 15 and 600, respectively. We choose $L = 15$ to better compare the results in the previous literature [9], and $m = 600$ works for all the tests we have done.

In all the CNN schemes, MatConvNet toolbox [37] is used to implement the proposed CNN schemes. For the training process, we use personal computer with CPU (3.4 GHz Intel Core i7 Processor and 16 GB RAM). It is noted that, since
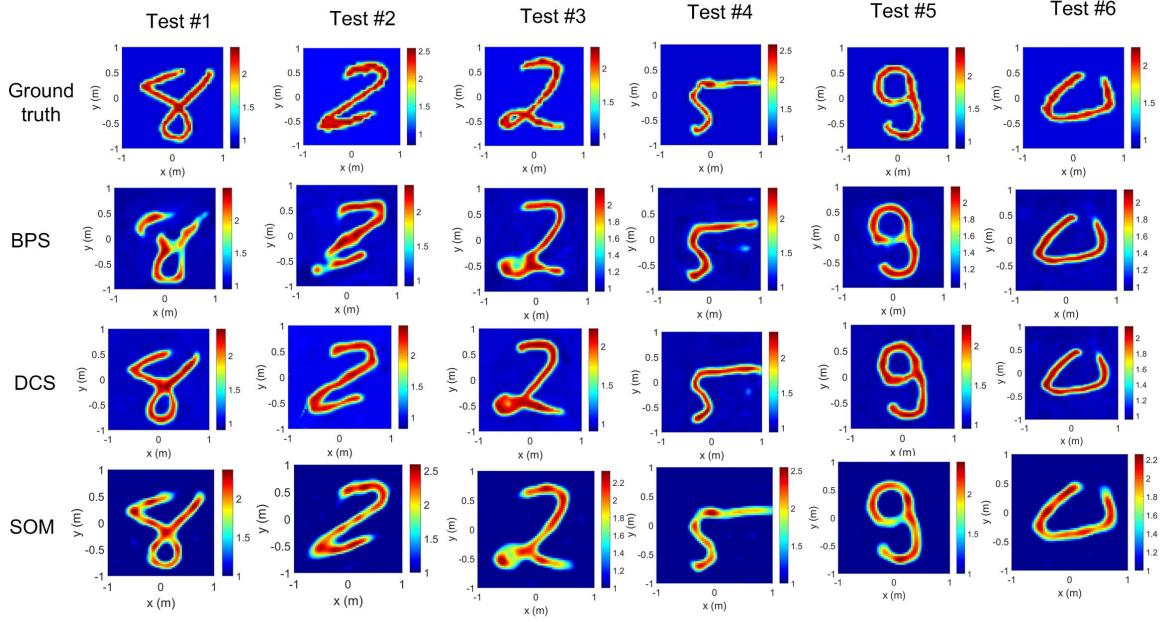
Fig. 7. Example Three (tests with MNIST database): Reconstructed relative permittivity profiles from scattered fields with 5% noise for BPS, DCS, and iterative method (SOM), where the relative permittivity is between 2 and 2.5. The first row shows the ground truth images for 6 representative tests.
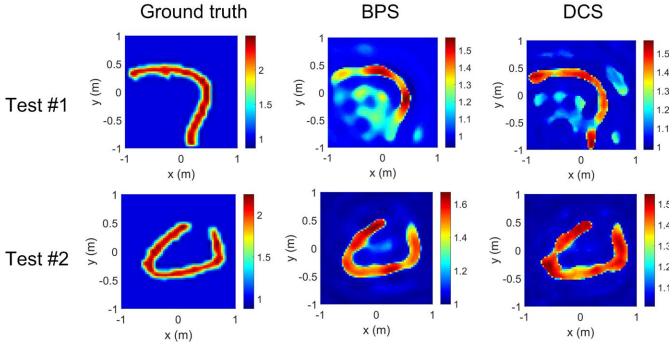

Fig. 8. Use the network trained with circular-cylinders in Example 1 to test the MNIST database in Example 3.

each operation of CNN is simple and local, both the training and test processes are ideal for GPU-based parallelization. Consequently, the time spent on both training and test can be significantly reduced by GPU calculation.

The cost function used for training in this paper is Euclidean loss with $\frac{1}{N_o} \sum_{i=1}^{N_o} (x_i - t_i)^2$, where N_o is the total number of outputs. x_i and t_i are the outputs of the last layer and targets (labels) of the CNN schemes, respectively. The hyperparameters for training are as follows: learning rate decreasing logarithmically from 10^{-6} to 10^{-8}; batchsize equals 1; momentum equals 0.99. Further, maximum 200 epoches are set for training unless there are no apparent changes in the cost function.

C. Tests with circular-cylinders

In the first example, the relative permittivity is set between 1 and 1.5 with 5% Gaussian noise presented in scattered field. The CNN network is firstly trained with different schemes, and then 25 tests with different profiles are conducted on the

trained network. In Fig. 4, reconstructed relative permittivity profiles of 4 representative tests are presented, and each test is finished within 1 second using the personal computer. Fig. 4 suggests that DIS can only reconstruct very simple profiles such as a single cylinder and in addition the reconstructed profiles do not present sharp boundaries. For some difficult profiles, the reconstructed relative permittivity profiles only show the position of profiles but with wrong shapes. In comparison, both BPS and DCS can obtain satisfying results for all tests. To evaluate the performance quantitatively, the relative errors are further computed for the 25 tests and shown in the row "Example 1" of Table I. It shows that, for weak-scattering scatterers (relative permittivity between 1 and 1.5), the relative errors R_e are in the same level for BPS and DCS, which are much lower than that of DIS.

To better compare the performance of BPS and DCS, we intentionally add a special test ("Austria profile") for the trained CNN presented in Fig. 5(a), where the parameters of the profile is actually out of the training range. Specifically, the "Austria" profile consists of two disks and one ring. Both of the disks have a radius of 0.2 m and centered at (0.3, 0.6) and (-0.3, 0.6) m. The ring has an inner radius of 0.3 m and an exterior radius of 0.6 m. The background is air and the relative permittivities of scatterers are 1.5. The "Austria" is special for two reasons. Firstly, the "Austria" is a challenging profile that is well-known in the community of inverse scattering, thus it acts as a representative challenging test. Secondly, as shown in Fig. 5(a), the "Austria" presents two features that are out of the range of training parameters. One is that there are 4 cylinders presented in the profile and the other is that the largest radius of cylinders is 0.6 m. Consequently, it is challenging for the trained CNNs.

Figs. 5(b) and (c) present the reconstructed relative permittivity profiles for BPS and DCS, respectively, where both
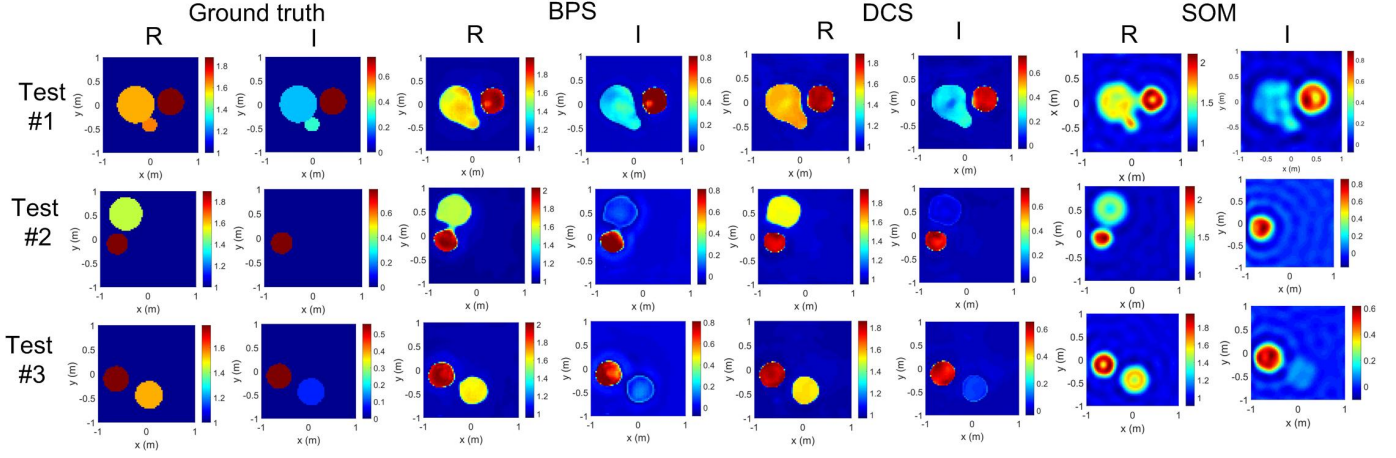
Fig. 9. Lossy scatterers tests: Reconstructed relative complex permittivity profiles from scattered fields with 5% noise for BPS, DCS, and iterative method (SOM), where the real part and imaginary part of relative permittivity are in the range of 1.5-2 and 0-1, respectively. The first two columns show the ground truth images for 3 representative tests. It is noted that R and I in the figure denote real and imaginary parts, respectively.

5% (left) and 20% (right) Gaussian noises are considered in the scattered fileds. The case with 20% noise is considered since it is common that, in practice, the noise level in the test process may be much larger than that in the training process. Specifically, for the tests with 20% noise case, we directly use the trained network with only 5% noise presented in the training process. It can be seen from Fig 5(b) that, although the positions are correctly displayed for BPS, the radii of the reconstructed ring and disks are apparently smaller than those of true profiles (dash line in Fig 5(b)). Further, there is an obvious artifact on the upper part of the ring. In comparison, DCS can obtain satisfying results for the "Austria" profile without any apparent discrepancy or shrunken cylinder. The relative errors for the "Austria" tests are also presented in Table I, and it shows that DCS quantitatively achieves better results than BPS does.

In the second example, the scatterers with the relative permittivities between 1.5 and 2 are considered, which means that the effects of multiple scattering become stronger than those of the first example and consequently the nonlinearity of the ISP is increased [26]. In Fig. 6, we present the reconstructed relative permittivity profiles of four representative tests for the three proposed schemes. It can be observed that while both BPS and DCS consistently outperform DIS, the DCS presents much better reconstruction results than BPS does, such as in the Tests #3 and #4 of Fig. 6. In the row "Example 2" of Table I, the relative errors R_e for the second example is also presented, and it is found that R_e for BPS increases significantly and becomes larger than that of DCS.

D. Tests with MNIST Database

In the third example, the unknown objects are modeled by MNIST database [25], which is a database of handwriting digits widely used in the field of machine learning. Instead of recognizing and classifying the digit in the database, we are devoted to quantitatively reconstructing the profile where the scatterers are represented by the digits. To this end, the scatterers are set to be dielectric with a random relative

permittivities between 2 and 2.5 and some representative examples are presented in the first row of Fig. 7. The MNIST training data contains 70,000 images of handwritten digits, and each digit image has a size of 28×28 pixels. In our example, we employ 475 images for training purpose and 25 images for testing purpose. In scattering problems, all images are assigned a dimension of 2×2 m^2 and then discretized into to 64×64 pixels.

Similar to the first two examples, reconstructed profiles from some representative examples are presented in Fig. 7 and the relative errors of all tests are included in the row "Example 3" of Table I. Since DIS scheme fails for this example, the results are excluded from Fig. 7. It is found from the MNIST example that both BPS and DCS are able to reconstruct the profile with satisfying results, but some discrepancies are shown for BPS.

To further test the proposed methods, we also consider an extreme case: we use the network trained with circular-cylinders in Example 1 to test the MNIST database in Example 3, where the two databases differ significantly with regard to both the profiles and ranges of relative permittivities. In Fig. 8, we present two representative tests with both BPS and DCS. It is seen that despite the results degrade compared with the results obtained in Example 3, Fig. 8 still shows that the reconstructed profiles are qualitatively satisfying, i.e, the shape, size, position of scatterers generally agree with the ground truth. The reason is that the proposed deep learning schemes train and learn the relationship between ground-truth permittivities and the inputs (12) and (25) in the natural pixel bases regardless of the shapes of scatterers. The input signals (12) and (25) to the learning system are derived from wave physics equations that are discretized into nature pixel bases. On the other hand, it is also noted that the degradation in this example can hardly refrain us from applying the proposed CNN schemes since before solving an inverse problem, we usually have a basic knowledge about the range to be solved.

It can be concluded from the first three examples that DCS outperforms BPS for challenging profiles, such as "Austria", and scatterers with larger permittivities. As mentioned
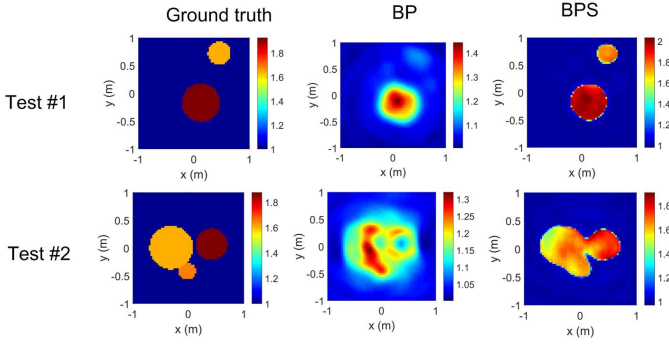
Fig. 10. Two tests of reconstructed results by BPS and BP from Example 2.

Fig. 11. True profile of "FoamDielExt" [38]: The large cylinder (SAITEC SBF 300) has a diameter of 80 mm with the relative permittivity $\varepsilon_r = 1.45 \pm 0.15$; The small cylinder (berylon) has a diameter of 31 mm with the relative permittivity $\varepsilon_r = 3 \pm 0.3$. Here, "$\pm$" denotes the range of uncertainty for the experimental value.

previously, instead of exploring an approximate solution of scatterer's profile that contains a large number of unknowns, DCS obtains an approximate solution for the dominant part of unknowns, which decreases the difficulties of the nonlinear problem. Moreover, since only a single solution is input into the CNN for BPS, the input can hardly capture all features of the profiles. In comparison, in DCS, results from different incidences are filled into different channels of the input images, which is helpful to capture more features of the profiles. Furthermore, by taking an average operation of the outputs corresponding to all incidences, the solutions are also more stable compared with a single output in BPS.

E. Tests with lossy scatterers

All of the above examples are based on the dielectric profiles of which are lossless. In order to further validate the versatility of the proposed methods, we try to reconstruct lossy scatterers with the proposed CNN schemes. The setups in the training process are the same with those of Example 2 except that complex permittivities are considered in this example. Specifically, the imaginary and real parts of $\overline{\overline{\xi}}^b$ and $\overline{\overline{\xi}}_p^d$ obtained from Eqs. (12) and (25) are put into different channels of inputs of BPS and DCS, respectively. In the first two columns of Fig. 9, the true profiles of three representative tests are depicted. The real and imaginary parts of relative permittivities are in the range of 1.5-2 and 0-1, respectively, which are referred to the previous literature [10]. The reconstructed results by both BPS and DCS are also presented in Fig. 9, and it is seen that both CNN schemes achieve satisfying results for lossy scatterers. In the last row of Table 1, we further quantitatively compare the relative errors from 25 tests, and it suggests that DCS achieves better performance than that of BPS.

F. Comparisons with iterative algorithm and noniterative method

In this paper, we also compare the proposed CNN schemes with widely used iterative algorithms. Specifically, SOM [9] is used to reconstruct all the profiles in previous examples and the results are compared with those of CNN schemes. For fair comparisons, all the parameters pertinent to implementation of SOM are set to the suggested values, and the number of
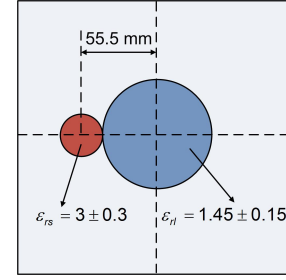
maximum iterations has been set to 150 unless there is no apparent change in the cost function.

As presented in Figs. 3-6, it is found that BPS and DCS obtain much sharper boundaries than those of SOM, and the reason is that the large number of training data offers strong constraints for CNN schemes. To quantitatively compare the performance, we have also calculated the relative errors of SOM for all the tests, as presented in Table 1. It can be seen that DCS outperforms SOM quantitatively for the first two and the last examples. However, in "Austria" tests and Example 3, SOM achieves better results than those of CNN schemes. For "Austria" tests, the reason is that the profile parameters, such as the number of circles and radii of "Austria", are out of training ranges of CNN schemes, which suggests a limitation that the performance of the trained CNN will degrade if the test example is out of training dataset. For Example 3, the relative permittivity is large and the ISP is highly nonlinear, thus some important features have been lost in the inputs of CNN schemes, which suggests that SOM outperforms the proposed CNN schemes in solving highly nonlinear ISPs.

Further, we have compared average time taken for reconstructing one test by all the methods. By using personal computer with CPU (3.4 GHz Intel Core i7 Processor and 16 GB RAM), it takes about 7 hours for training CNNs, and less than 1 second for testing with the proposed CNN schemes. Each CNN scheme differs slightly with respect to both training and testing time. In comparison, for SOM with FFT acceleration, it takes around 196 seconds to finish the reconstruction, which suggests that the tests by CNN schemes are hundreds of times faster than those by the iterative method.

Besides the iterative algorithm, to verify the comparative enhancement provided by CNN schemes over usual noniterative methods, we also compare CNN schemes with back propoagation (BP), where the results of BP are actually the inputs of BPS. We have tested the performance of BP in Examples 1, 2 and 3 and compared it with BPS. Specifically, the returned resulting relative errors for BP are 6.97%, 17.39%, and 28.05% for Examples 1, 2, and 3, respectively. In comparison, as presented in Table 1 of the manuscript, the relative errors for BPS are 1.7%, 6.5%, and 11.4% for Examples 1, 2, and 3, respectively. In Fig. 10, we further present two tests of reconstructed results by BP and BPS in Example 2, and it can be seen that BPS outperforms BP significantly.
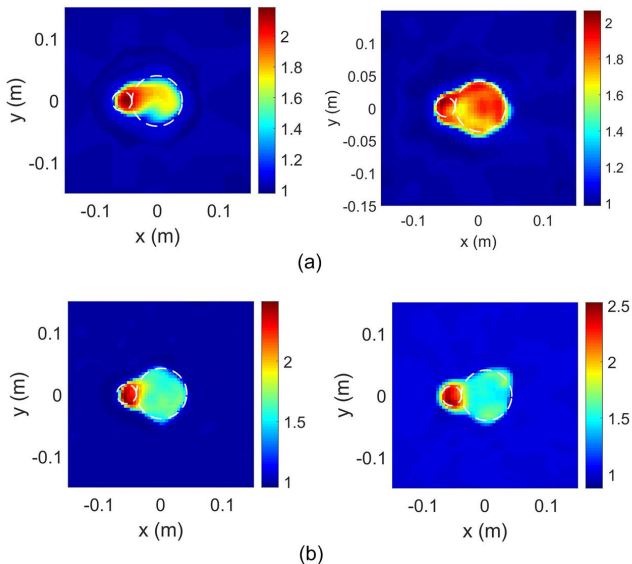
Fig. 12. Experimental data tests: image reconstruction of "FoamDielExt". Reconstructed relative permittivity profiles for (a) BPS (b) and DCS at 3 GHz (Left) and 4 GHz (Right). It is noted that the dashed lines in both (a) and (b) denote the true boundaries of cylinders in Fig. 11.

V. TESTS WITH EXPERIMENTAL DATA

To further validate the proposed BPS and DCS, tests have also been conducted with experimental data measured at Institue Fresnel [38]. As presented in Fig. 11, a "FoamDielExt" profile with TM case is considered in this section. The transmitting and receiving antennas are both wide band ridged horn antennas, and the source-object center and object-receiver distances are $d_e = 1.67$ m. There are 8 transmitters and 241 receivers in this data set. The "FoamDielExt" consists of two cylinders, where the large cylinder (SAITEC SBF 300) has a diameter of 80 mm with the relative permittivity $\varepsilon_r = 1.45 \pm 0.15$, and the small cylinder (berylon) has a diameter of 31 mm with the relative permittivity $\varepsilon_r = 3 \pm 0.3$. To calibrate the experimental data, one single complex coefficient is multiplied with the data, where the coefficient is simply derived from the ratio of the measured incident field and the simulated one at the receiver located at the opposite of the source [38].

We know *a priori* that the scatterers are distributed within a 30 cm × 30 cm DOI. In both BPS and DCS, the training data are chosen to be similar to those used in Examples 1 and 2, with the difference that we have increased the range of relative permittivity (1.5-3.3). Since we have no knowledge of the noise level, we train neural network using noiseless synthetic data and test the network using measured experimental data. This approach has the advantage that it does not require a large number of experimental data and save a lot of time in solving practical problems.

In Fig. 12, we present the reconstructed relative permittivity profiles from BPS and DCS at both 3 GHz and 4 GHz, where both of them are obtained within 1 second. For both cases, the values of m keep the same as in the previous section ($m = 300$). Since the lower the frequency is, the faster the singular values of G_S drops, the value of L for 4 GHz is

larger than that for 3 GHz. Specifically, according to the criterion introduced in Section III, i.e., the Lth singular value σ_L is around 50% of the first singular value σ_1, $L = 15$ and $L = 20$ are chosen for 3 GHz and 4 GHz, respectively. It is found that, although BPS has successfully detected the positions of scatterers in DOI, the shape and value differ from the true profile of "FoamDielExt" for both cases. In comparison, DCS is able to obtain satisfying results that are much better than those of BPS. Specifically, the relative errors are 14.3% and 12.2% for BPS and DCS, respectively, which quantitatively validate our observations in Fig. 12. In addition, it is known that the nonlinearity of ISP increases when frequency increases, and consequently it is more challenging to solve ISP. It can be observed from the right figure of Fig. 12(b) that, some distortions are observed for DCS at 4 GHz, which also suggests a limitation of DCS. Nevertheless, the results from experimental data further verify our conclusion that, compared with BPS, DCS has a better performance when dealing with high nonlinear inverse scattering problems.

VI. CONCLUSION

In this paper, we proposed three inversion schemes including DIS, BPS and DCS based on a U-net CNN to solve full-wave inverse scattering problems, where scatterers are represented by pixel bases. It is found that the computation time for all the CNN schemes in reconstructing a typical 64×64 image is below one second, which is much faster than that reported for iterative reconstruction. In numerical simulations, we show that DIS can only reconstruct some simple profiles from the measured scattered fields and the reconstructed image does not exhibit sharp boundary. In comparison, both BPS and DCS are able to reconstruct satisfying results with sharp boundaries given the measured scattered fields. Even for the tests well out of the range of the training databases, the proposed BPS and DCS are able to obtain satisfying results despite the degradation on performance since the training is conducted on natural pixel bases.

By comparing with iterative method (SOM), it is also found that the proposed BPS and DCS obtain much sharper boundaries than those of SOM, and the reason is that the larger number of training data offers strong constraints for CNN schemes. Although DCS quantitatively achieves much better results than SOM in some examples, there are two cases where SOM outperforms the proposed DCS. The first case is that the test examples are out of the range of training process, such as "Austria" test. The second case is that the ISP is highly nonlinear, such as Example 3 in the numerical tests. Further, it can be seen that the training of CNN requires a large set of data consisting of experimental data for various true scatterers, which is a challenging and time-consuming process in practice. For this drawback, an alternative approach is presented in this paper, where we use synthetic data from simulations to train the network, and then reconstruct relative permittivities from calibrated experimental data.

REFERENCES

[1] H. Kagiwada, R. Kalaba, S. Timko, and S. Ueno, "Associate memories for system identification: Inverse problems in remote sensing," *Mathematical and Computer Modelling*, vol. 14, pp. 200–202, 1990.

[2] A. Quarteroni, L. Formaggia, and A. Veneziani, *Complex systems in biomedicine*. New York;Milan;: Springer, 2006.

[3] Y. M. Wang and W. C. Chew, "An iterative solution of the two-dimensional electromagnetic inverse scattering problem," *International Journal of Imaging Systems and Technology*, vol. 1, no. 1, pp. 100–108, 1989.

[4] W. C. Chew and Y. M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted born iterative method," *IEEE Transactions on Medical Imaging*, vol. 9, no. 2, pp. 218–225, 1990.

[5] T. M. Habashy, M. L. Oristaglio, and A. T. de Hoop, "Simultaneous non-linear reconstruction of two-dimensional permittivity and conductivity," *Radio Science*, vol. 29, no. 4, pp. 1101–1118, 1994.

[6] M. v. d. B. Peter and E. K. Ralph, "A contrast source inversion method," *Inverse Problems*, vol. 13, no. 6, p. 1607, 1997.

[7] P. M. v. d. Berg, A. L. v. Broekhoven, and A. Abubakar, "Extended contrast source inversion," *Inverse Problems*, vol. 15, no. 5, p. 1325, 1999.

[8] S. Lin-Ping, Y. Chun, and L. Qing Huo, "Through-wall imaging (twi) by radar: 2-d tomographic results and analyses," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2793–2798, 2005.

[9] X. Chen, "Subspace-based optimization method for solving inverse-scattering problems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 42–49, 2010.

[10] Y. Zhong and X. Chen, "An FFT twofold subspace-based optimization method for solving electromagnetic inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 3, pp. 914–927, 2011.

[11] K. Xu, Y. Zhong, and G. Wang, "A hybrid regularization technique for solving highly nonlinear inverse scattering problems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 1, pp. 11–21, 2018.

[12] T. M. Habashy, R. W. Groom, and B. R. Spies, "Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering," *Journal of Geophysical Research: Solid Earth*, vol. 98, no. B2, pp. 1759–1775, 1993.

[13] K. Belkebir, P. C. Chaumet, and A. Sentenac, "Superresolution in total internal reflection tomography," *Journal of the Optical Society of America A*, vol. 22, no. 9, pp. 1889–1897, 2005.

[14] S. Caorsi and P. Gamba, "Electromagnetic detection of dielectric cylinders by a neural network approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 820–827, 1999.

[15] I. T. Rekanos, "Neural-network-based inverse-scattering technique for online microwave medical imaging," *IEEE Transactions on Magnetics*, vol. 38, no. 2, pp. 1061–1064, 2002.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015: 18th International Conference*, 2015, pp. 234–241.

[20] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2012, pp. 341–349.

[21] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014, pp. 1790–1798.

[22] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[23] G. Riegler, M. Rther, and H. Bischof, "Atgv-net: Accurate depth super-resolution," in *Computer Vision ECCV 2016: 14th European Conference*, 2016, pp. 268–284.

[24] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[26] X. Chen, *Computational Methods for Electromagnetic Inverse Scattering*. Wiley, 2018.

[27] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer-Verlag, Berlin, Germany, 2nd edn., 1998.

[28] A. Kirsch, *An introduction to the mathematical theory of inverse problem*. Springer, New York, 1996.

[29] A. F. Peterson, S. L. Ray, and R. Mittra, *Computational methods for electromagnetics*. Wiley-IEEE Press New York, 1998, vol. 2.

[30] R. Wang and D. Tao, "Non-local auto-encoder with collaborative stabilization for image restoration," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2117–2129, 2016.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[32] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. European Conf. Computer Vision*, 2016, pp. 694–711.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, p. 448456.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[35] J. Cong and B. Xiao, "Minimizing computation in convolutional neural networks," in *24th International Conference on Artificial Neural Networks, Proceedings*, 2014, pp. 281–290.

[36] G. W. Stewart, *Matrix algorithms*. Philadelphia: Society for Industrial and Applied Mathematics, 1998.

[37] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proc. ACM Int. Conf. Multimedia*, 2015, pp. 689–692.

[38] G. Jean-Michel, S. Pierre, and E. Christelle, "Free space experimental scattering database continuation: experimental set-up and measurement precision," *Inverse Problems*, vol. 21, no. 6, p. S117, 2005.