Unrolled convolutional neural network for full-wave inverse scattering

Yarui Zhang, Marc Lambert, Aurélia Fraysse, Dominique Lesselier

Unrolled Convolutional Neural Network for Full-Wave Inverse Scattering

Yarui Zhang, Marc Lambert, Aurélia Fraysse, and Dominique Lesselier, *Senior Member IEEE*

***Abstract*—An unrolled deep learning scheme for solving full-wave nonlinear inverse scattering problems (ISPs) is proposed. Inspired by the so-called *unrolled* method, an iterative neural network structure combining the contrast source inversion (CSI) method and residual network (ResNet) is designed. By embedding the CSI iterations into the deep learning model, the domain knowledge is well incorporated into the learning process. Thorough numerical tests are carried out to evaluate the performance, stability, robustness, and reliability of the proposed approach. Comparisons with the widely used U-net structure and CSI exhibit the advantage of the proposed approach.**

***Index Terms*—contrast source inversion method, deep learning, inverse scattering, unrolled method**

I. Introduction

ELECTROMAGNETIC inverse scattering problems (ISPs) [1] are concerned with reconstructing the spatial distribution or physical parameters of an unknown scatterer from measured scattered fields. ISPs are faced with in various fields, like medical imaging [2], remote sensing [3], and non-destructive testing [4].

ISPs are generally tackled by optimization methods that minimize the discrepancy between obtained results and expected ones. Deterministic optimization methods such as the Contrast Source Inversion method (CSI) [5], Distorted Born Iterative Method (DBIM) [6], and Subspace-based Optimization Method (SOM) [7] and variants, are widely used to deal with nonlinear ISPs. Yet, such local optimization methods may stall in local minima unless reasonable initialization. Prior information or regularizations such as positivity and sparsity [8] are usually enforced onto the optimization to alleviate ill-posedness.

Deep learning, mostly convolutional neural networks (CNN), has achieved great success for tasks of computer vision, e.g., image segmentation [9], object recognition [10]. Deep learning is also successful for inverse imaging problems [11], [12], [13], like image denoising [14], restoration [15], super-resolution [16], with worthy application in MRI-based medical imaging [17] as well. Recently, deep-learning-based methodologies have been applied to ISPs, as reviewed, e.g., in [18], [19] and [20]. Unlike traditional iterative methods that get the solutions from measurements and the domain knowledge, they benefit from the large data set to learn the unknown solution while the domain knowledge is less involved.

The advantages of applying deep learning tools to solve ISP are twofold. On the one hand, it allows real-time reconstruction whereas traditional iterative methods require computational time. On the other hand, by benefiting from the information contained in large-scale data, deep learning tools can sometimes solve highly nonlinear ISPs for which traditional methods fail. However, collecting large-scale datasets may not be straightforward in real-world applications.

Deep learning approaches for solving ISPs can be roughly categorized into three types as follows.

The contrast distribution can be directly derived from the measured data, the input of the neural network being the scattered fields, and the output being the contrast. In [21], CNN is used to diagnose a damaged micro-structure and the architecture takes the collected data as input and yields the contrast distribution. Nevertheless, directly learning the contrast from the scattered fields requires time to learn the known physical process. As a result, the combination of domain knowledge and neural networks has been much exploited.

Approaches of the second type consist in incorporating physical knowledge into deep learning. In [22] and [23], several physical-based learning approaches are proposed such as the Back-Propagation Scheme (BPS) and Induced Current Learning Method (ICLM), which pre-initialize the input of the neural network based on the domain knowledge. In [24], two physical-guided loss functions are proposed, which enhance noise robustness and improve reconstruction accuracy. These methods can reduce the nonlinearity of the relationship between input and output of the neural network and simplify learning.

The Deep-NIS approach advocated in [25] also neatly fits the category of physics-inspired deep learning tools, once emphasized that there exist many contributions on ISPs that would fit the category, yet lie beyond the scope of the present contribution. In some parallel also, one should cite [26] as involving in original fashion a two-step solution, extraction of so-called feature fragments from the data via an encoder network, mapping those to the contrasts via a decoder network. Also, among many other contributions, the reviewing of which goes beyond the scope of the present contribution, a procedure that embeds a forward modeling neural network in an inversion one is devised in [27] and shows success for a full-wave ISP.

A neural network can be combined also with traditional op-

Yarui Zhang is with Université Paris-Saclay, CentraleSupélec, CNRS, Laboratoire de Génie Electrique et Electronique de Paris, 91192, Gif-sur-Yvette, France.(e-mail: yarui.zhang@geeps.centralesupelec.fr).

Marc Lambert is with Université Paris-Saclay, CentraleSupélec, CNRS, Laboratoire de Génie Electrique et Electronique de Paris, 91192, Gif-sur-Yvette, France.(e-mail: marc.lambert@geeps.centralesupelec.fr).

Aurélia Fraysse is with Laboratoire des signaux et systèmes, Université Paris-Saclay, CNRS, CentraleSupélec, Gif-sur-Yvette, France.(e-mail: aurelia.fraysse@l2s.centralesupelec.fr).

Dominique Lesselier is with Laboratoire des signaux et systèmes, Université Paris-Saclay, CNRS, CentraleSupélec, Gif-sur-Yvette, France.(e-mail: dominique.lesselier@l2s.centralesupelec.fr).

timization methods. This third type still employs the traditional iterative approach to solve the ISPs but uses deep learning to complete the optimization procedure. In [28], the learning process is followed by a refinement stage, where SOM is to refine the result obtained from the previous stage.

To achieve sound retrievals, incorporating domain knowledge into the deep model is of importance. In addition to initializing the network with approximate results based on physical information as often made, the *unrolled* method [29] has been proposed for inverse problems with a fully known forward model to flexibly combine iterative optimization methods with deep learning, the physics guiding the solution as exemplified in the recent contribution [30].

It replaces the components of iterative optimizers with learnable parameters or networks while integrating the knowledge of the direct model. Unlike the aforementioned deep learning schemes, which mainly adopt a single network to map the measurements or the approximate reconstructed results to the ground truth, the unrolled method uses a recurrent block architecture where the number of repeated blocks corresponds with the number of iterations.

Its general framework has been much discussed to solve inverse problems. Generally, deep models are designed based on the proximal gradient descent algorithm, where a trainable CNN can replace the proximal operator. In addition to the proximal operator, other free parameters like the step-size of descent can be learned in training. In [31] and [32], the residual network is used to model the proximal map. Furthermore, following connections between an iterative solver and the CNN structure, some neural networks are designed to imitate the iterative methods in [12], [33] and [34].

In the present contribution, inspired by the unrolled deep learning method as proposed in the field of computer vision and signal processing, the interest of which is, in particular, emphasized in the thorough review [35] a novel unrolled model which combines updating steps of CSI and a classical CNN structure is developed. The entire network has a recurrent block architecture that contains a series of sub-blocks. Each includes a set of CSI update steps and a CNN block, which corresponds to an update of contrast and contrast source, as it will be seen. Then, several strategies are designed to improve the model. Due to the recurrent network structure, a combined loss function is set up, the weight parameters being determined automatically during training. Modifications are also made compared to a general unrolled method to adapt the network to nonlinear inverse problems.

The contribution goes as follows. The forward problem is sketched in Section II. The proposed methods are developed in Section III. Numerical illustrations are in Section IV. Conclusion and perspectives are in Section V.

II. Problem formulation

The scenario is time-harmonic two-dimensional electromagnetic scattering in the transverse magnetic (TM) polarization (time convention $e^{-\mathrm{i}\omega t}$, ω angular frequency). An inhomogeneous linear isotropic scatterer is embedded inside a homogeneous linear isotropic medium $\mathcal{D}$ with permittivity of air ϵ_0 and permeability μ_0. $\epsilon_r(\mathbf{r})$ and $\sigma(\mathbf{r})$ denote relative permittivity and conductivity of the scatterer with $\mathbf{r} \in \mathcal{D}$ as a given observation point. TM waves generated by N_s ideal line sources located outside the scatterer at positions $\mathbf{r}_s$ illuminate it. For each illumination, the scattered fields are collected by N_r ideal receivers at positions $\mathbf{r}_r$ along a circle of observation $\mathcal{S}$, not intersecting the scatterer, and not necessarily jointing with the sources.

The scattered electric field $E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s)$ collected by a receiver at $\mathbf{r}_r$ and associated with a source at $\mathbf{r}_s$ satisfies the integral equation

$$E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s) = \int_D G(\mathbf{r}_r, \mathbf{r}') J(\mathbf{r}', \mathbf{r}_s) d\mathbf{r}', \ \mathbf{r} \in \mathcal{D} \tag{1}$$

letting

$$J(\mathbf{r}, \mathbf{r_s}) = \chi(\mathbf{r}) E^{\text{tot}}(\mathbf{r}, \mathbf{r}_s) \tag{2}$$

$G(\mathbf{r}, \mathbf{r}')$ is the Green's function which represents the electromagnetic response to a line source radiating in free space. It is given by $G(\mathbf{r}, \mathbf{r}') = \frac{-i\omega\mu_0}{4} \mathrm{H}_0^{(1)}(k_B \|\mathbf{r} - \mathbf{r}'\|)$, $\mathrm{H}_0^{(1)}$ zero-order Hankel function of the first kind.

The contrast function is defined as usual as $\chi(\mathbf{r}) = k^2(\mathbf{r}) - k_B^2$ where $k^2(\mathbf{r}) = \omega^2 \epsilon_0 \epsilon_r(\mathbf{r}) \mu_0 + i\omega\mu_0\sigma(\mathbf{r})$, $k_B^2 = \omega^2\epsilon_0\mu_0$. $J(\mathbf{r}, \mathbf{r_s})$ and $E(\mathbf{r}, \mathbf{r}_s)$ are the equivalent current and the total electric field induced within the object by the incident wave.

The total electric field is obtained as

$$E^{\text{tot}}(\mathbf{r}, \mathbf{r_s}) = E^{\text{inc}}(\mathbf{r}, \mathbf{r}_s) + \int_{\mathcal{D}} G(\mathbf{r}, \mathbf{r}') J(\mathbf{r}', \mathbf{r_s}) d\mathbf{r}', \ \mathbf{r} \in \mathcal{D} \tag{3}$$

where $E^{\text{inc}}(\mathbf{r}, \mathbf{r}_s)$ is the incident field.

Using a pulse-basis point-matching method of moments (MoM), the domain $\mathcal{D}$ is discretized into N^2 small square pixels wherein electric field and contrast are considered constant. The discrete version of the previous equations reads

$$E^{\text{diff}}_{i_r, i_s} = \sum_{i=1}^{N} G_{S;i_r,i} J_{i,i_s} \tag{4}$$

$$E^{\text{tot}}_{i,i_s} = E^{\text{inc}}_{i,i_s} + G_{D;i,i} J_{i,i_s} \tag{5}$$

with $i_s = 1, \ldots, N_s$, $i_r = 1, \ldots, N_r$, $i = 1, \ldots, N^2$, and $J_{i,i_s} = \text{diag}(\chi_i) E^{\text{tot}}_{i,i_s}$ the contrast source. Subscripts D and S indicate the location of $\mathbf{r}$, and otherwise, the operators are identical:

$$G_{D,\,S} J_{i,i_s} = \int_{\mathcal{D}} G(\mathbf{r}, \mathbf{r}') J(\mathbf{r}', \mathbf{r}_s) d\mathbf{r}' \qquad \mathbf{r} \in \mathcal{D} \text{ or } \mathbf{r} \in \mathcal{S} \tag{6}$$

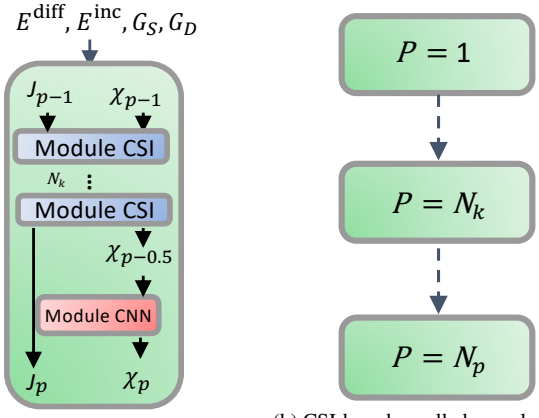
The forward problem is the calculation of $E^{\text{diff}}_{i_r,i_s}$ from knowledge of χ_i and the (non-linear and ill-posed) inverse scattering one is to retrieve χ_i from $E^{\text{diff}}_{i_r,i_s}$.

III. CSI-based unrolling deep learning schema

As underlined, it is key to integrate the physical knowledge into the deep learning approach to get physically meaningful results while reducing the computational burden of learning already known physical laws.

The unrolled method, which imitates iterative optimization approaches, generally holds a network structure with repeated

This article has been accepted for publication in IEEE Transactions on Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAP.2022.3216999



(a) Single sub-block structure

(b) CSI-based unrolled neural network structure

Fig. 1. CSI-based unrolled neural network structure. $p \in \{1 \dots N_p\}$, N_p number of sub-blocks. $k \in \{1 \dots N_k\}$, N_k number of CSI iterations in each.

sub-blocks. Each sub-block contains the iterations of traditional optimization methods and a neural network.

Herein, one proposes a CSI-based unrolled deep learning method (The structure is sketched in Fig. 1), where each sub-block of the network contains two modules: *Module CSI* and *Module CNN*. In the following, the CSI method is first reviewed as the main brick of *Module CSI*. Then, some basic elements of CNN and the network structure adopted are introduced. Details of the approach follow. Last, the U-net [36] structure is briefly presented for comparison.

A. Contrast source inversion method: Module CSI

CSI is much used to tackle inverse scattering. From (4) and (5), state and data equations stand as

$$J_{i,i_s} = \text{diag}\left(\chi_i\right)\left[E^{\text{inc}}_{i,i_s} + G_{D,i,i} J_{i,i_s}\right] \tag{7}$$

$$E^{\text{diff}}_{i,i_s} = \sum_{i=1}^{N} G_{S;i_r,i} J_{i,i_s} \tag{8}$$

The cost function is a linear combination of normalized mismatches in state and data equations:

$$F(J_{i,i_s}, \chi_i) = \frac{\sum_{i_s}\sum_{i_r} \left\|\xi_{i_r,i_s} - \sum_i G_{S;i_r,i} J_{i,i_s}\right\|^2}{\sum_{i_s}\sum_{i_r} \left\|\xi_{i_r,i_s}\right\|^2} + \frac{\sum_{i_s}\sum_{i} \left\|\text{diag}\left(\chi_i\right) E^{\text{inc}}_{i,i_s} - J_{i,i_s} + \text{diag}(\chi_i) G_{D;i,i} J_{i,i_s}\right\|^2}{\sum_{i_s}\sum_{i} \left\|\text{diag}\left(\chi_i\right) E^{\text{inc}}_{i,i_s}\right\|^2} \tag{9}$$

Contrast sources J_{i,i_s} and contrast χ_i are alternatively reconstructed with a conjugate gradient method. At each iteration, the contrast sources are updated by minimizing the entire cost function, then χ_i is determined by minimizing the error in the state equation. Note that the minimization at each iteration is linear. Update steps of CSI at a single iteration k are shown in Algorithm 1. Here, $\rho_{i_s,k}$ is the data error, $r_{i_s,k}$ the state error, $g^J_{i_s,k}$ the gradient of the cost function (9) w.r.t $J_{i_s,k}$ evaluated at $J_{i_s,k-1}$, whereas β_{k-1}, $\nu_{i_s,k}$ the Polak-Ribière conjugate gradient directions, and α^J_k the update size.

Algorithm 1 CSI at iteration k

Input: $J_{i_s,k-1}$, χ_{k-1}, ξ_{i_s}, G_S, G_D, E^{inc}

$\rho_{i_s,k} = \xi_{i_s} - G_S J_{i_s,k-1}$

$r_{i_s,k} = \beta_{k-1} E^{\text{tot}}_{i_s,k-1} - J_{i_s,k-1}$

$$g^J_{i_s,k} = -\frac{G^*_S \rho_{i_s,k}}{\sum_{i_s}\sum_{i_r}\|\xi_{i_r,i_s}\|^2} - \frac{r_{i_s,k} - G^*_D(\chi_{k-1} r_{i_s,k})}{\sum_{i_s}\sum_i \|\chi_i E^{\text{inc}}_{i,i_s}\|^2}$$

$$\nu_{i_s,k} = g^J_{i_s,k} + \frac{\Re\left\{\sum_{i_s}\left\langle g^J_{i_s,k}, g^J_{i_s,k} - g^J_{i_s,k-1}\right\rangle\right\}}{\sum_{i_s}\left\langle g^J_{i_s,k-1}, g^J_{i_s,k-1}\right\rangle}\nu_{i_s,k-1}$$

$$\alpha^J_k = \frac{-\Re\left\{\sum_{i_s}\left\langle g^J_{i_s,k}, \nu_{i_s,k}\right\rangle\right\}}{\frac{\sum_{i_s}\|G_S\nu_{i_s,k}\|^2}{\sum_{i_s}\sum_{i_r}\|\xi_{i_r,i_s}\|^2} + \frac{\sum_{i_s}\|\nu_{i_s,k} - \chi_{k-1} G_D \nu_{i_s,k}\|^2}{\sum_{i_s}\sum_i \|\chi_i E^{\text{inc}}_{i,i_s}\|^2}}$$

$J_{i_s,k} = J_{i_s,k-1} + \alpha^J_k \nu_{i_s,k}$

$E^{\text{tot}}_{i_s,k} = E^{\text{inc}}_{i_s} + G_D J_{i_s,k}$

$$\chi_k = \frac{\sum_{i_s=1}^{N_s} J_{i_s,k} E^{\text{tot}^*}_{i_s,k}}{\sum_{i_s=1}^{N_s}\|E_{i_s,k}\|^2}$$

Output: $J_{i_s,k}$, χ_k

The first sub-block, *Module CSI* –based on the CSI method– aims at providing an immediate initialization of J and χ for the *Module CNN*. Then, in the following sub-blocks, *Module CSI* further refines the results that come from the precedent *Module CNN* (Module CNN, yielding, at the specific step considered, the contrast distribution, is discussed in detail in the following subsections.) Note that the step size α^J is not a learnable parameter and is determined by minimizing both data and state error functions.

B. Convolutional neural networks: Module CNN

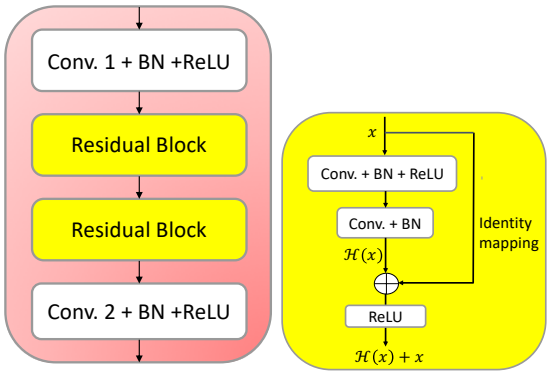
A convolutional neural network (CNN) typically consists of four types of layers: convolutional, rectified linear unit (ReLU), pooling, and fully-connected layer. Traditional artificial neural networks are usually built only with fully connected layers, which brings drawbacks like loss of spatial structure information, computational burden, and over-fitting. CNN can alleviate these problems because the convolutional layer as the core layer in building a CNN produces few parameters.

In the proposed unrolled schema, the network structure in *Module CNN* is stacked by residual blocks. Residual learning has been proposed in [37] to deal with network degradation, which means that adding more layers might degrade learning performance. The entire network structure of *Module CNN* is illustrated in Fig. 2. It thus consists of two residual blocks and two convolutional blocks.

Adding a shortcut between the input x and the output, the network has to learn the residual $\mathcal{H}(x)$ between ground truth and input. It should be easier to optimize than the original mapping, and this shortcut ensures that higher layers perform at least as well as lower ones.

The input of *Module CNN* is the contrast χ obtained by *Module CSI*, which is a tensor of shape

[Number of samples $\times$ Number of channels $\times$ $H \times W$], H and W being image height and width. The number of input/output channels is 2 for a complex-valued image by splitting real and imaginary parts into two different channels; for a real-valued image, the second channel is empty.



(a) Module CNN (b) Residual sub-block

Fig. 2. Network structure of Module CNN and its residual sub-block.

For all convolutional layers, the spatial size of the filters used to obtain the feature maps is 3×3. Inside the residual block, the number of filters for each layer is 64. The number of filters of layers *Conv.*1 and *Conv.*2 in *Module CNN* are 64 and 2, respectively. To keep the spatial size of feature maps unchanged after each convolution, the stride of the convolution operation is set to 1, and a 1-sized zero-padding is added. Each convolutional layer is followed by batch normalization (BN), which stabilizes the neural network, accelerates training convergence, reduces sensitivity to weight initialization, and alleviates overfitting. Finally, a ReLU layer is added as the nonlinear activation layer used to overcome the vanishing gradient problem and add the nonlinearity to the model.

C. CSI-based unrolling deep learning framework: CSI-Net

As depicted in Fig. 1, the network contains a series of sub-blocks. For the first sub-block, the input is the pre-initialized contrast χ and contrast source J, obtained by standard back-propagation.

A closed-form contrast source is obtained as:

$$J_0 = \frac{\|G_S^* E^{\text{diff}}\|^2}{\|G_S G_S^* E^{\text{diff}}\|^2} G_S^* E^{\text{diff}} \tag{10}$$

and the corresponding contrast χ_0 derived from J_0:

$$\chi_0 = \frac{\sum_{i_s=1}^{N_s} J_{i_s,0} E_{i_s,0}^{\text{tot}*}}{\sum_{i_s=1}^{N_s} \|E_{i_s,0}^{\text{tot}}\|^2} \tag{11}$$

Additionally, the inputs include as well auxiliary data $E^{\text{diff}}, E^{\text{inc}}, G_S, G_D$ to perform the CSI iterations. Denoting N_p the total number of sub-blocks and N_k the total number of CSI iterations performed inside each sub-block. In the p-th sub-block, there are two main steps.

Firstly, the input passes through a sequence of *Module CSI*. In the commonly used unrolling deep learning framework, the gradient descent update step is performed only once inside each sub-block, where the step size is considered as a free parameter learned during training. In contrast, the step size α^J here is a constant determined at each iteration, thus not a learnable parameter. Moreover, instead of running a single CSI iteration in each sub-block, the CSI iteration is repeated several times.

This modification comes from the following considerations: Data in *Module CSI* are not normalized, so the initialization of step size is essential and should respect the correct scale. So it is more convenient to determine the step size by following traditional optimization. As it cannot be learned, the single iteration step is replaced by multiple iterations to speed up convergence and amplify improvements brought by *Module CSI*. At the end of this step, a preliminary reconstruction $\chi_{p-0.5}$ of the contrast χ is obtained.

Secondly, the previous result $\chi_{p-0.5}$ is applied as input of *Module CNN*. Note that it should be normalized before being involved in the network. The output of this module is χ_p, which is the final result of the p-th sub-block, and the corresponding contrast source J_p is updated by solving the forward problem.

Since the proposed model has a cascade architecture, the choice of the loss function is important. The loss scale corresponding to each sub-block's output could significantly differ. To equally train each sub-block, a combined loss function built with the normalized mean square error (MSE) is chosen:

$$L = \sum_{i=1}^{N_p} \frac{\lambda_i L_i}{\sum_{i=1}^{N_p} \lambda_i}, \; L_i = \sum_{j=1}^{N_e} (\tilde{x}_j^i - x_j)^2, \; \lambda_i = \frac{L_0}{L_i} \tag{12}$$

where N_e denotes the number of elements in $\tilde{x}^i$, and λ_i is the weighting coefficients of i-th sub-block.

D. U-net deep learning framework for comparison

The U-net structure, originally designed for medical image segmentation, is widely used for ISPs (as in [22], [24] among many). Details of U-net can be found in [36]. Fig. 3 shows the network structure in the present work.

IV. Numerical illustration

Numerical experiments have been carried out to evaluate the performance of the method. Details of the scattering system are as follows: the operating frequency is 400 MHz, the region of interest $\mathcal{D}$ is of $2\,\text{m} \times 2\,\text{m}$ size and is discretized into N square cells. For the forward problem, $N = 32 \times 32$. $N_r = 32$ receivers and $N_s = 16$ transmitters are evenly distributed on a circle of radius $r = 3\,\text{m}$ centered at (0, 0) m. The embedding medium is of $\epsilon_b = 1$.

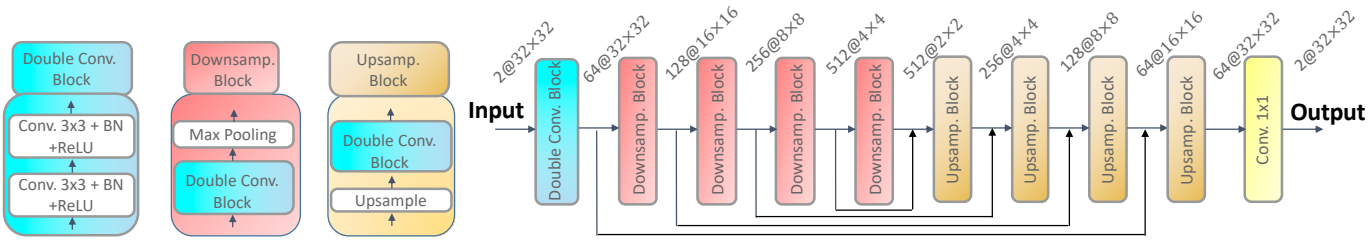
To evaluate the reconstruction quality from different points of view, three criteria are adopted:

1) Relative error of contrast

$$Err_\chi = \frac{\|\tilde{\chi} - \chi\|^2}{\|\chi\|^2} \tag{13}$$

2) Relative error of permittivity over the pixelized domain

$$Err_\varepsilon = \sqrt{\frac{1}{N_x \times N_y} \left[\sum_{i_x}^{N_x} \sum_{i_y}^{N_y} \left| \frac{\tilde{\varepsilon}_{i_x,i_y} - \varepsilon_{i_x,i_y}}{\varepsilon_{i_x,i_y}} \right| \right]} \tag{14}$$



(a)

Fig. 3. The U-net structure. The dimension of the feature maps passing through each layer is noted as [Number of channels@$H \times W$].

3) Structural Similarity Index

$$\mathrm{SSIM}(\tilde{\chi}, \chi) = \frac{(2\mu_{\tilde{\chi}}\mu_{\chi} + C_1) + (2\sigma_{\tilde{\chi}\chi} + C_2)}{(\mu_{\tilde{\chi}}{}^2 + \mu_{\chi}{}^2 + C_1)(\sigma_{\tilde{\chi}}{}^2 + \sigma_{\chi}{}^2 + C_2)} \tag{15}$$

where μ_x and σ_x are the mean and the standard deviation of x, respectively. σ_{xy} is the covariance between x and y. $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$ are to stabilize the division with weak denominator where L the dynamic range of the pixel values. $k_1 = 0.01$ and $k_2 = 0.03$ by default.

Denote the proposed unrolled network as CSI-Net. To check on the advantage brought by *Module CSI*, one removes all *Module CSI* blocks from CSI-Net, and the model becomes a cascaded residual neural network, denoted as CC-ResNet. For comparison, results of CC-ResNet, the widely used U-net structure, and the iterative CSI method with positivity constraint are also displayed. For the latter, the number of iterations has been arbitrarily chosen as $N_k^{\mathrm{CSI}} = 150$ and the algorithm stopped when N_k^{CSI} is reached.

A. Implementation details

Implementation details of the approach, including division of dataset and hyperparameters settings, are as follows: several datasets are used to train the proposed model, each one comprising 2500 samples. Among the 2500 data pairs, 2000 are employed for training, 300 for validation, and 200 for testing. The Adam optimizer [38] is chosen to optimize the loss function. 40 epochs are set in the training process. The size of the mini-batch is 20. The learning rate is set to 0.001, decayed by 0.1 each 10 epochs. N_k and N_p are chosen by compromising between learning performance and computational complexity, here $N_k = 10$ and $N_p = 4$.

Computations are run on a server with NVIDIA RTX A3000 GPU (6 GB). The networks are implemented with Pytorch [39]. It takes 0.63 h for CSI-Net training and 0.03 h for CC-ResNet and U-net. It takes less than 0.4 s to get the reconstruction using CSI-Net, and about 4 s using CSI, U-Net being almost real-time.

B. Test with circular dataset

A synthetically generated circular dataset [22] is used. Each sample contains disks of random quantities: number (between 1 and 3), relative permittivity (between 1.3 and 2), radius (between 0.15 m and 0.4 m), and location. The scatterers may overlap. The collected fields used for training are noise-free. Unless specified, the scatterers are lossless.

The trained models are tested with 200 samples. Fig. 4 shows six representative results obtained by three models and CSI. Results by CSI-Net are smoother, and edges better preserved, especially when multiple scatterers interplay.

In Samples #1, #3 and #4, where the scatterers are very close to each other, only rough estimates are obtained by CC-ResNet and U-net. The shape and relative permittivity obtained are not exact, and the boundaries not clear.

In Samples #2 and #5, with CC-ResNet and U-net, the reconstruction of the weaker scatterers is influenced by the relatively stronger ones, so only the latter are well retrieved. Sample #6 is challenging since an obstacle is inside one another. CC-ResNet and U-net can retrieve the two large obstacles while the small nested one is missed. With CSI, the reconstructed scatterers exhibit clear boundaries and shapes. Yet, compared to CSI-Net, artifacts show up inside the scatterers obtained by CSI.

Table I provides the corresponding reconstructed errors and SSIM. For all test samples #1−#6, CSI-Net generates the lowest reconstruction errors and highest SSIM. In addition, Err_χ, Err_ε and SSIM of all 200 samples obtained by three trained models are displayed in Fig. 5 to provide a global view of the reconstruction performance. CSI-Net ensures the lowest average reconstruction errors but also reduces extreme cases, i.e., decreases the variance of reconstruction errors.

TABLE I
RECONSTRUCTION QUALITY OF THREE MODELS AND CSI. SNR = 20 dB.

	Model	Sample #1	#2	#3	#4	#5	#6
Err_χ	CC-ResNet	0.172	0.121	0.137	0.113	0.101	0.113
	CSI-Net	**0.027**	**0.028**	**0.021**	**0.027**	**0.017**	**0.014**
	U-net	0.166	0.085	0.108	0.100	0.125	0.093
	CSI	0.057	0.067	0.058	0.063	0.055	0.064
Err_ε	CC-ResNet	0.148	0.073	0.098	0.077	0.081	0.073
	CSI-Net	**0.048**	**0.045**	**0.037**	**0.050**	**0.034**	**0.029**
	U-net	0.149	0.074	0.090	0.072	0.096	0.071
	CSI	0.078	0.069	0.063	0.069	0.067	0.065
SSIM	CC-ResNet	0.340	0.475	0.445	0.425	0.488	0.499
	CSI-Net	**0.710**	**0.726**	**0.693**	**0.682**	**0.704**	**0.720**
	U-net	0.203	0.582	0.6514	0.569	0.531	0.669
	CSI	0.478	0.463	0.430	0.443	0.481	0.394

To check on robustness, further simulations are carried out with Gaussian noises of different SNR (20, 10 and 5 dB) added to the scattered fields. In Fig. 6, the mean and variance of results obtained with the real-valued circular dataset are displayed. For the mean error, the growth rate of CSI-Net at

This article has been accepted for publication in IEEE Transactions on Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAP.2022.3216999

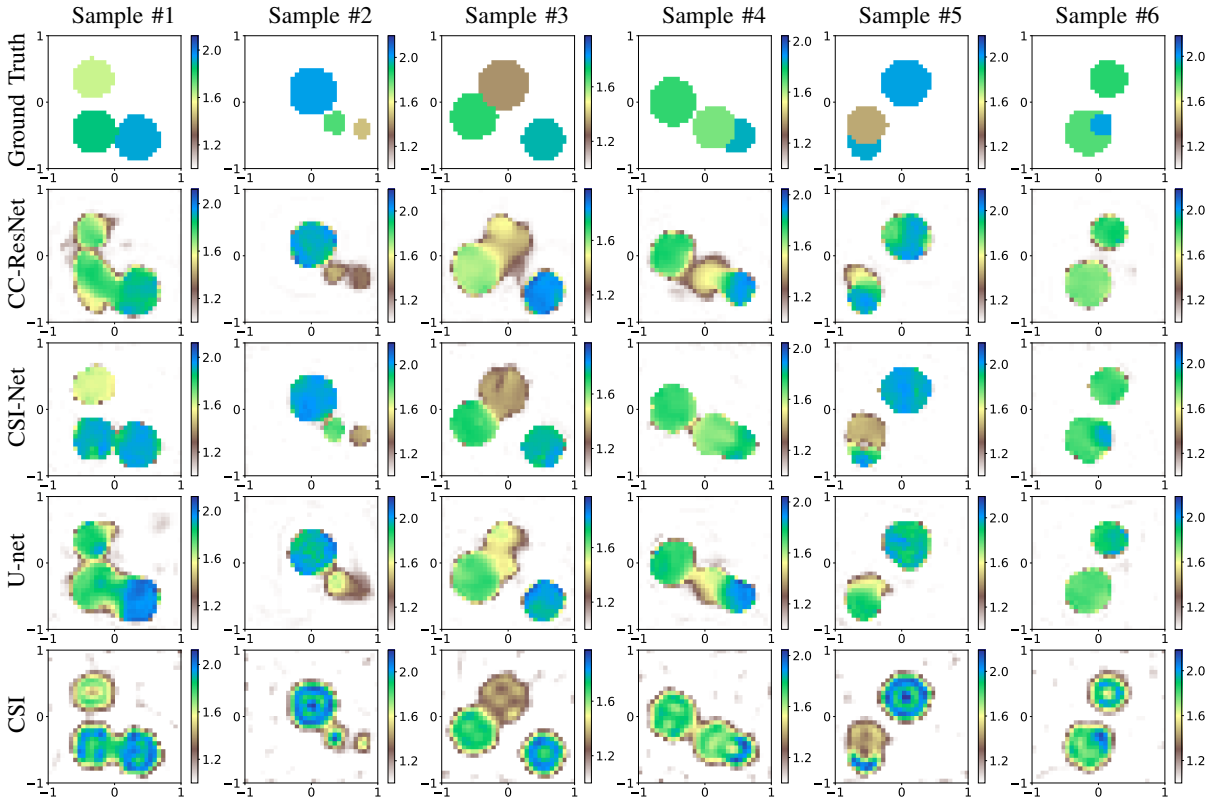
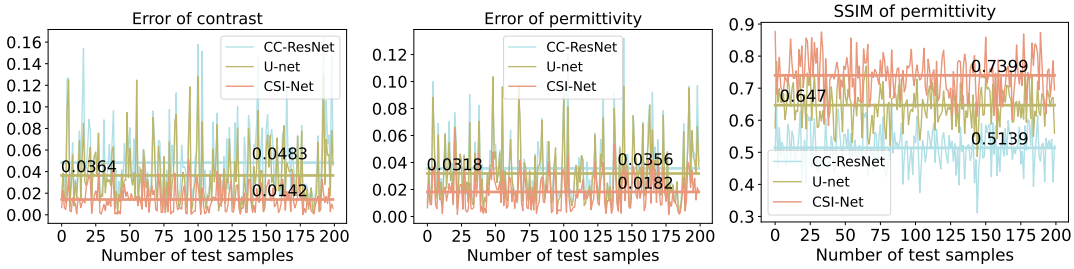


Fig. 4. Relative permittivity of reconstructed profiles from the test dataset. The range of relative permittivity is $[1.3, 2]$. Gaussian noise of SNR = 20 dB is added. The first row is the ground truth of the profiles.



(a) (b) (c)

Fig. 5. Err_χ, Err_ε and SSIM obtained by applying the trained model on 200 test samples. Gaussian noise of SNR = 20 dB is added to collected fields.

20 and 10 dB is larger than for the two other approaches, yet CSI-Net always exhibits the smallest mean error, i.e., it is robust to noise contamination. As for the variance change, improvement by the proposed method lies not only in better overall reconstruction quality but also in a strong reduction of poor results, i.e., confidence level increases.

In general, a larger training dataset leads to better neural network training. However, in the real world, collecting datasets may not be straightforward. So, a neural network should perform well with a small amount of data. Here, the models are trained with the real-valued circular dataset, and the numbers of training pairs are 250, 500, 1000, 1500, and 2000, respectively. The mean value and variance of Err_χ are displayed in Fig. 7. As the training dataset's size increases, the performance of all models improves. But, for a small dataset, U-net performance strongly degrades while CC-ResNet and CSI-Net remain stable. Moreover, CSI-Net consistently outperforms the other two methods.

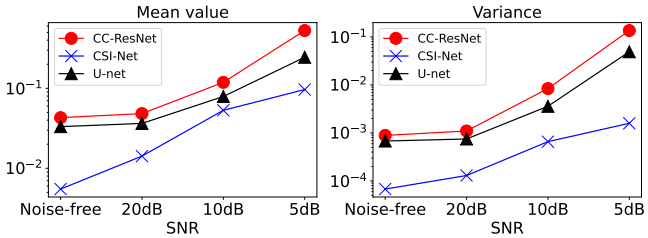


Fig. 6. Noise stability: mean and variance of Err_χ obtained by CC-ResNet, CSI-Net, and U-net on 200 test samples belonging to the circular database as a function of noise levels.

Evaluations of reliability can be performed by showing that the training dataset is not particularly tuned to a proposed

This article has been accepted for publication in IEEE Transactions on Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAP.2022.3216999

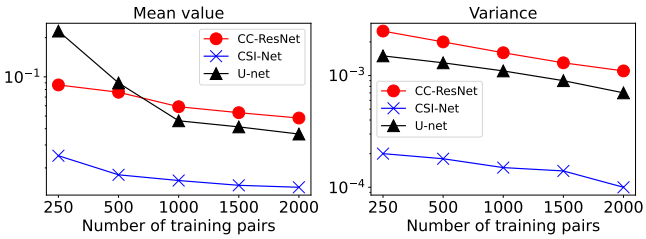


Fig. 7. Mean and variance of Err_χ obtained by CC-ResNet, CSI-Net and U-net on 200 test samples as a function of the number of training pairs.

network. To this end, a real-valued circular dataset including 5000 samples is used. Five datasets of 2000 data pairs are randomly sampled from the latter as training datasets for the models. Fig. 8 illustrates that the performance of the proposed approach remains stable when the dataset changes.

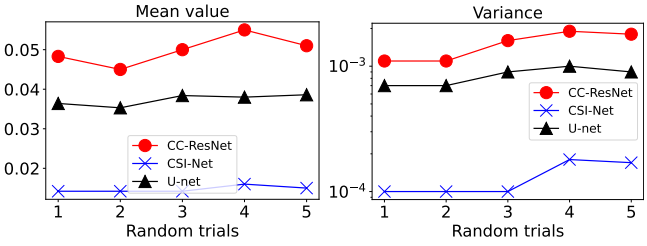


Fig. 8. Mean and variance of Err_χ obtained by CC-ResNet, CSI-Net and U-net on 200 test samples using models trained with five randomly chosen datasets of 2000 data pairs.

Generalization capability is always an issue, so challenging examples must be considered. The models trained with the circular dataset are now tested on six test samples that are not within the range of the training dataset. The results are in Fig. 9. CC-ResNet is not included since all of its results are worse than CSI-Net. Notice that positivity constraint is enforced in CSI, significantly improving its performance. The number of iterations is set to 200. In particular, samples #7 and #8 are the well-known "Austria" profile with relative permittivity of 1.5 and 2, respectively.

CSI-Net achieves the best reconstruction for most of the samples from #7 to #12. For sample #7, U-net only gives a rough reconstruction, where the center of the ring is obscured, and the shape is not continuous. For sample #8, where the level of nonlinearity increases, it fails. As for CSI, the retrieved profiles have a smaller size than the true ones. Also, artifacts show up around the edges. For the other samples, U-net fails to provide meaningful results and CSI can achieve good results, though with more artifacts than CSI-Net results. The Err_χ of the results are in Table II. They show that CSI-Net has a better generalization capability than U-net and can improve the reconstruction quality of several cases compared to CSI.

C. Test with MNIST dataset

The proposed model is trained with a different dataset where the scatterers are modeled from the MNIST database [40], which contains handwritten digits samples. Digits are transformed to scatterers of random relative permittivity between 2 and 2.5. To increase diversity, a circular/rectangle

TABLE II
ERR_χ OF TEST SAMPLES #7-#12. SNR = 20 dB

Model	Sample #7	#8	#9	#10	#11	#12
CSI-Net	**0.038**	0.095	**0.023**	**0.024**	**0.029**	0.052
U-net	0.236	0.477	0.157	0.182	0.235	0.101
CSI	0.098	**0.093**	0.036	**0.024**	0.031	**0.034**

of random relative permittivity between 2 and 2.5 is added to each sample. The diameters/side lengths are between 0.3 m and 0.8 m. The scattering system settings, dataset division, and training hyperparameters are unchanged.

The trained models are tested on 200 samples within the same range of the training dataset. As shown in Fig. 10, samples #13-#16 are four representative samples among the test dataset, and #17-#18 are "Austria" profiles with relative permittivities of 2.5 and 3, respectively. Due to strong nonlinearity, CSI fails to reconstruct all samples so its results are not displayed. U-net can roughly reconstruct the profiles, however, most details are missing. In comparison, CSI-Net achieves better reconstructions for all test samples.

The Err_χ of these test samples obtained by different methods are presented in Table III. Note that its last two columns give the mean value and the variance of Err_χ across all 200 test samples. It is readily seen that CSI-Net outperforms other methods for the cases where the nonlinearity increases.

TABLE III
ERR_χ OF TEST SAMPLES #13-#18, AND 200 TEST SAMPLES. SNR = 20 dB

Model	Sample #13	#14	#15	#16	#17	#18	Mean	Variance
CSI-Net	**0.073**	**0.064**	**0.096**	**0.059**	**0.105**	**0.218**	**0.047**	**0.0004**
U-net	0.249	0.197	0.219	0.162	0.146	0.244	0.097	0.0014

D. Test with lossy scatterers

All previous results are on lossless scatterers. The proposed model is now trained with a complex-valued circular dataset to represent lossy scatterers. The training dataset shares the same key parameters as the real-valued one except for the relative permittivity value. Real parts of relative permittivities are between 1.3 and 2.4; imaginary parts vary from 0 to 1.2. The trained model is tested on the "Austria" profile, and the results are in Fig. 11. Both CSI-Net and CSI exhibit satisfactory results. On the contrary, U-net fails to reconstruct the ring for both real and imaginary parts. Corresponding Err_χ are in Table IV. The last two rows give the mean error and variance over 200 test samples.

V. CONCLUSION

An unrolled deep learning model embedding the iterative CSI optimization method into the CNN structure has been proposed to tackle a full-wave 2-D inverse scattering problem. Indeed, employing deep learning tools to solve such a problem requires learning the laws of scattering. Also, those tools may perform well when traditional optimization methods appear to fail, yet lack of generalization capability and potentially low

This article has been accepted for publication in IEEE Transactions on Antennas and Propagation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAP.2022.3216999

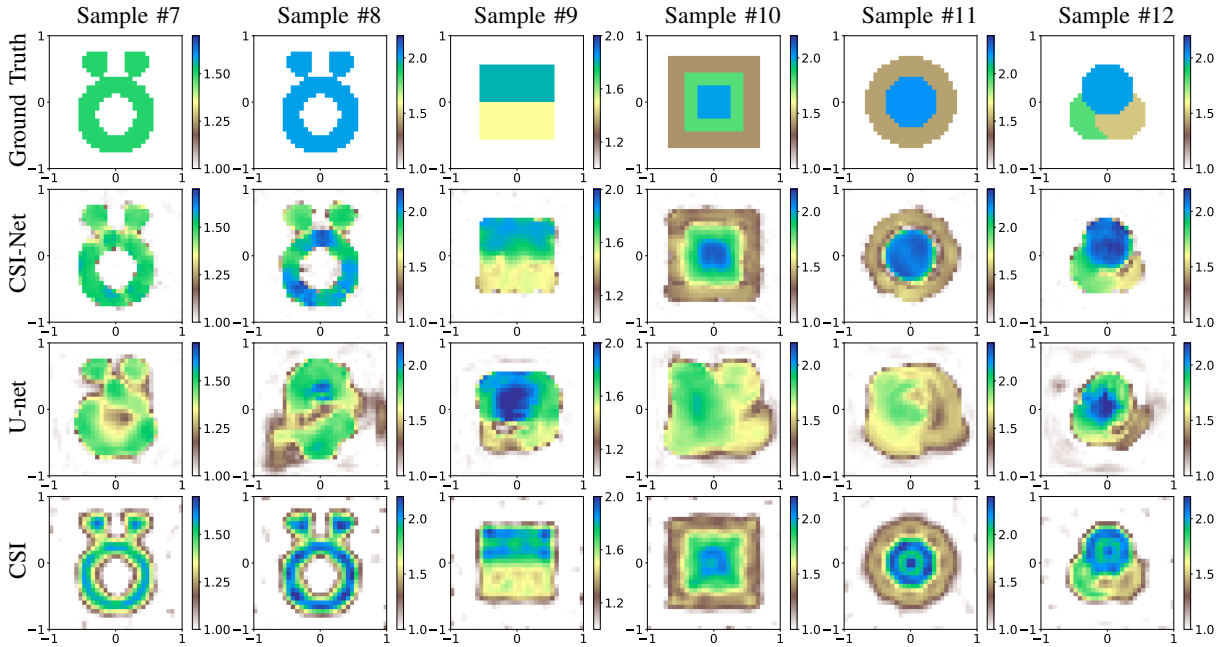


Fig. 9. Reconstructed relative permittivity obtained by CSI-Net, U-net, and CSI. The models are trained with the circular dataset. The test dataset lies outside the range of training data. 20 dB Gaussian noise affects the collected scattered fields.

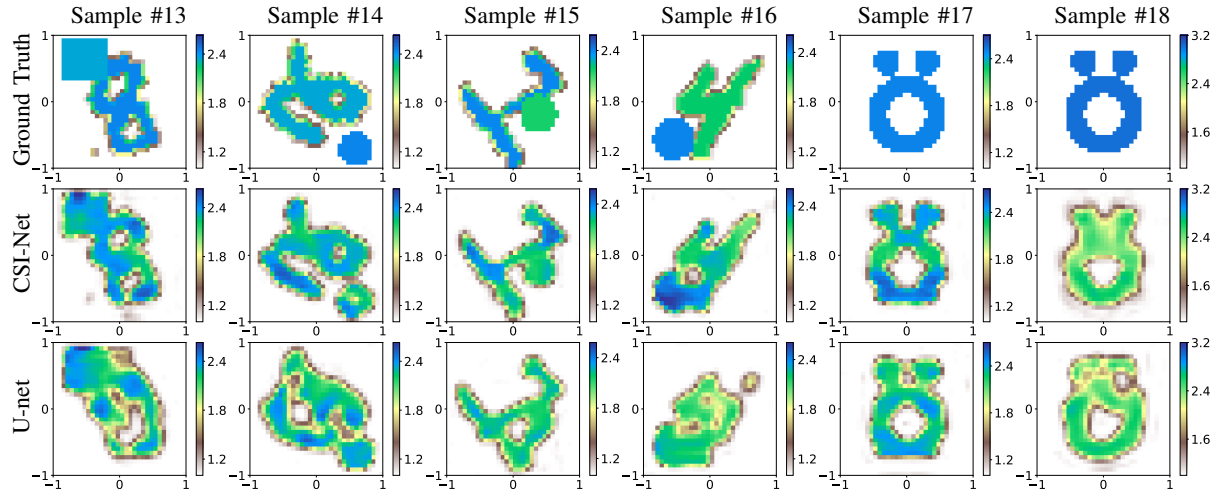


Fig. 10. Reconstructed relative permittivity obtained by CSI-Net, U-net (CSI has been omitted since failing to yield proper reconstruction). The models are trained with the MNIST dataset. The test dataset is outside the range of training data. 20 dB Gaussian noise affects the collected scattered fields.

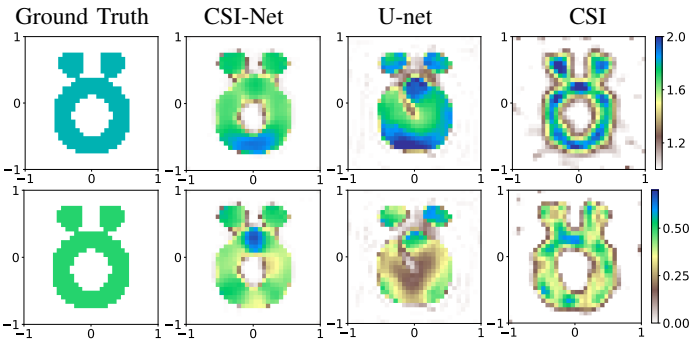


Fig. 11. Reconstructed relative permittivity obtained by models trained with a complex-valued circular dataset. $\Re\{\varepsilon\}$ (top) of ground truth is 1.8 and $\Im\{\varepsilon\}$ (bottom) is 0.5. 20 dB Gaussian noise affects the collected scattered fields.

confidence level tend to impair their usage. So, combining domain knowledge with learning approaches by integrating the CSI updating steps within the network structure appears a fruitful way forward.

Numerical illustrations indeed confirm that the proposed CSI-Net has a better generalization ability than the widely used U-net structure. The confidence level of results is also improved with the help of embedded CSI steps. Furthermore, good results are obtained even when CSI fails.

The fact that the cost function takes in account the outputs of all subblocks with proper weights, and so doing at least yields the same degree of convergence of each said subblock, should be emphasized. Notice however that one is not aware of mathematical proof that would fit the proposed framework, so work into that direction should be helpful.

To summarize, CSI-Net combines the advantages of both

TABLE IV
MEAN AND VARIANCE OF ERR_χ OVER 200 TEST SAMPLES. SNR = 20 dB.

Model	CSI-Net	U-net	CSI
"Austria"	**0.131**	0.269	0.157
Mean	**0.021**	0.040	\
Variance	**0.0004**	0.0011	\

sides. Overall, the proposed approach provides a general scheme for fruitfully mixing iterative optimization and learning. That is, CSI can be replaced by another iterative method while extra constraints should further improve the confidence level of the network structure.

REFERENCES

[1] X. Chen, *Computational Methods for Electromagnetic Inverse Scattering*. Wiley-IEEE Press, Singapore, 2018.

[2] N. K. Nikolova, "Microwave imaging for breast cancer," *IEEE Microw. Mag.*, vol. 12, no. 7, pp. 78–94, 2011.

[3] C. Varotsos and V. Krapivin, *Microwave Remote Sensing Tools in Environmental Science*. Springer Cham, Switzerland, 2020.

[4] N. Ida and N. Meyendorf, *Handbook of Advanced Nondestructive Evaluation*. Springer Cham, Switzerland, 2019.

[5] P. M. Van den Berg and A. Abubakar, "Contrast source inversion method: State of art," *Prog. Electromagn. Res.*, vol. 34, pp. 189–218, 2001.

[6] W. Chew and Y. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," *IEEE Trans. Med. Imaging*, vol. 9, no. 2, pp. 218–225, 1990.

[7] X. Chen, "Subspace-based optimization method for solving inverse-scattering problems," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 1, pp. 42–49, 2010.

[8] Y. Zhang, M. Lambert *et al.*, "Group sparsity penalized contrast source solution method for 2-D non-linear inverse scattering," *IEEE Open J. Antennas Propag.*, vol. 3, pp. 48–58, 2022.

[9] G. Litjens, T. Kooi *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[10] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 3367–3375.

[11] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Mag.*, vol. 34, no. 6, pp. 85–95, 2017.

[12] K. H. Jin, M. T. McCann *et al.*, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[13] A. Lucas, M. Iliadis *et al.*, "Using deep neural networks for inverse problems in imaging: Beyond analytical methods," *IEEE Signal Processing Mag.*, vol. 35, no. 1, pp. 20–36, 2018.

[14] S. Zhang and E. Salari, "Image denoising using a neural network based non-linear filter in wavelet domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, 2005, pp. ii/989–ii/992 Vol. 2.

[15] K. Zhang, W. Zuo *et al.*, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 2808–2817.

[16] Z. Cui, H. Chang *et al.*, "Deep network cascade for image super-resolution," in *Proc. Int. Conf. Comput. Vis.*, D. Fleet, T. Pajdla *et al.*, Eds. Cham: Springer International Publishing, 2014, pp. 49–64.

[17] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Modl: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 394–405, 2019.

[18] A. Massa, D. Marcantonio *et al.*, "Dnns as applied to electromagnetics, antennas, and propagation—a review," *IEEE Antennas Wireless Propagat. Lett.*, vol. 18, no. 11, pp. 2225–2229, 2019.

[19] X. Chen, Z. Wei *et al.*, "A review of deep learning approaches for inverse scattering problems (invited review)," *Prog. Electromagn. Res.*, vol. 167, pp. 67–81, 01 2020.

[20] M. Salucci, M. Arrebola *et al.*, "Artificial intelligence: New frontiers in real time inverse scattering and electromagnetic imaging," *IEEE Trans. Antennas Propagat.*, pp. 1–1, 2022.

[21] P. Ran, Y. Qin *et al.*, "Subwavelength microstructure probing by binary-specialized methods: Contrast source and convolutional neural networks," *IEEE Trans. Antennas Propagat.*, vol. 69, no. 2, pp. 1030–1039, 2021.

[22] Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Trans. Geosci. Remote Sensing*, vol. 57, no. 4, pp. 1849–1860, 2019.

[23] ——, "Physics-inspired convolutional neural network for solving full-wave inverse scattering problems," *IEEE Trans. Antennas Propagat.*, vol. 67, no. 9, pp. 6138–6148, 2019.

[24] Z. Liu, M. Roy *et al.*, "Physics-guided loss functions improve deep learning performance in inverse scattering," *IEEE Trans. Comput. Imaging*, vol. 8, pp. 236–245, 2022.

[25] L. Li, L. G. Wang, and F. L. Teixeira, "Performance analysis and dynamic evolution of deep convolutional neural network for electromagnetic inverse scattering," *IEEE Antennas Wireless Propagat. Lett.*, vol. 18, no. 11, pp. 2259–2263, 2019.

[26] H. M. Yao, L. Jiang, and W. E. I. Sha, "Enhanced deep learning approach based on the deep convolutional encoder–decoder architecture for electromagnetic inverse scattering problems," *IEEE Antennas Wireless Propagat. Lett.*, vol. 19, no. 7, pp. 1211–1215, 2020.

[27] R. Guo, Z. Lin *et al.*, "Physics embedded deep neural network for solving full-wave inverse scattering problems," *IEEE Trans. Antennas Propagat.*, pp. 1–1, 2021.

[28] Y. Sanghvi, Y. Kalepu, and U. K. Khankhoje, "Embedding deep learning in inverse scattering problems," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 46–56, 2020.

[29] G. Ongie, A. Jalal *et al.*, "Deep learning techniques for inverse problems in imaging," *IEEE J. Sel. Areas Commun.*, vol. 1, no. 1, pp. 39–56, 2020.

[30] J. Liu, H. Zhou *et al.*, "Physical model-inspired deep unrolling network for solving nonlinear inverse scattering problems," *IEEE Trans. Antennas Propagat.*, vol. 70, no. 2, pp. 1236–1249, 2022.

[31] M. Mardani, Q. Sun *et al.*, "Neural proximal gradient descent for compressive imaging," in *Adv Neural Inf Process Syst.*, ser. NIPS'18, S. Bengio, H. Wallach *et al.*, Eds. Red Hook, NY, USA: Curran Associates Inc., 2018, p. 9596–9606.

[32] S. Diamond, V. Sitzmann *et al.*, "Unrolled optimization with deep priors," *ArXiv*, vol. abs/1705.08041, 2017.

[33] J. Zhang and B. Ghanem, "ISTA-net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 1828–1837.

[34] L. Li, L. G. Wang *et al.*, "Deepnis: Deep neural network for nonlinear electromagnetic inverse scattering," *IEEE Trans. Antennas Propagat.*, vol. 67, no. 3, pp. 1819–1825, 2019.

[35] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IIEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, N. Navab, J. Hornegger *et al.*, Eds. Springer, Cham, 2015, pp. 234–241.

[37] K. He, X. Zhang *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 770–778.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learning Represent.*, 2015.

[39] A. Paszke, S. Gross *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Adv. Neural Inf. Process. Syst.*, H. M. Wallach, H. Larochelle *et al.*, Eds., vol. 32, 2019, pp. 8024–8035.

[40] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Processing Mag.*, vol. 29, no. 6, pp. 141–142, 2012.

Yarui Zhang received the B.E. degree from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2018 and the M.S. degree in Control, signal and image processing from Université Paris-Saclay, Gif-sur-yvette, France, in 2019. She is currently pursuing the Ph.D. degree in signal and image processing with the Université Paris-Saclay in Group of Electrical Engineering – Paris. Her research interests include inversion and imaging, signal and image processing and inverse scattering problems.

Marc Lambert received the Doctorat en Sciences and Habilitation à Diriger des Recherches degrees from the Université Paris Sud, Orsay, France, in 1994 and 2001, respectively. Since 1995, he has been with Chargé de Recherche, Centre National de la Recherche Scientifique (CNRS), Paris, France, and has carried out his research work with the Laboratoire des Signaux et Systèmes, Joint Laboratory of Supélec, CNRS and Université Paris Sud, Gif-sur-Yvette, France, until 2014. In 2015, he joined the Group of Electrical Engineering Paris. His research focuses on solutions of direct and inverse scattering problems in both electromagnetics and acoustics, and their applications to the characterization of complex objects buried in complex environments from limited datasets, most of his attention being on the numerous theoretical and computational issues which this characterization entails.

Aurélia Fraysse was born in Créteil, France, in 1978. She graduated from University Paris 12 in 2001. She received her Ph. D. degree in mathematics at University Paris 12 in 2005.
In 2006, she was a research assistant at Ecole Nationale Supérieure des Télécommunications (Telecom Paris Tech). She is presently associate-professor at IUT Cachan and researcher with the Laboratoire des Signaux et Systèmes, Université Paris-Saclay, CNRS, CentraleSupélec.

Dominique Lesselier (Senior Member, IEEE) was born in Lons-le-Saunier, France, in August 1953. He received the Engineering degree from Ecole Supérieure d'Electricité (Supélec), Paris, France, in 1975, and the Doctorat d'Etat es Sciences Physiques degree from Université Pierre et Marie Curie, Paris, in 1982. He is with the Centre National de la Recherche Scientifique (CNRS) since October 1981, and he is now Director of Research CNRS Emeritus. He was a Visiting Scholar with the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA, USA, from 1982 to 1983. He belongs to the Laboratoire des Signaux et Systèmes (L2S), Université Paris-Saclay, CNRS, CentraleSupélec, Gif-sur-Yvette, France. As the Director (2006–2009) of the Groupement de Recherche CNRS "GDR Ondes", he managed a network of scientists involved in the science of waves. Dr. Lesselier is a fellow of the Institute of Physics and a member of the Electromagnetics Academy and the International Union of Radio Science, Commission B. He was on the International Advisory Panel of Inverse Problems from 2005 to 2016 after serving on its Editorial Board from 1997 to 2004, and from 2003 to 2019, he has been Associate Editor of Radio Science. Since 1998, he is on the Standing Committee of the Electromagnetic Non-Destructive Evaluation Workshop Series and the International Steering Committee of the International Symposia on Applied Electromagnetics and Mechanics. He was a recipient of the R. W. P. King Award in 1982 from the IEEE Antennas and Propagation Society. His research interests include the development of solution imaging methods of inverse problems under many guises, from mathematics to simulations to applications, and vice versa.