

# A Tailored Semiphysics-Driven Artificial Neural Network for Electromagnetic Full-Wave Inversion

Yanjin Chen, Miao Zhong, Zhen Guan, and Feng Han<sup>ID</sup>, *Senior Member, IEEE*

**Abstract**—This article presents a tailored artificial neural network (ANN) that is designed for electromagnetic (EM) full-wave inversion (FWI). The ANN includes two subnets. The first subnet is a physics-driven fully connected neural network in which the known sensitivity matrix of EM FWI is assigned to the network model parameters. The second subnet is the traditional data-driven convolutional neural network (CNN) U-Net. The role of the first subnet is to convert the scattered EM fields recorded at the receiver array into the preliminary dielectric images of the scatterers with the aid of the fast calculation of the graphics processing unit. Then, the second subnet is used to further refine the scatterer dielectric parameters. In the off-line training, the network parameters in the first subnet are frozen, and only the parameters in the second subnet are optimized. In the online prediction, the whole ANN is used to directly invert the dielectric parameters of the scatterers. The proposed semiphysics-driven ANN is compared with the purely data-driven ANN for the inversion accuracy, training cost, generalization ability, antinoise ability, and so on.

**Index Terms**—Artificial neural network (ANN), convolutional neural network (CNN), electromagnetic (EM) full-wave inversion (FWI).

## I. INTRODUCTION

LECTROMAGNETIC (EM) full-wave inversion (FWI) is to reconstruct the unknown scatterer model parameters, such as locations and dielectric properties by strictly solving the scattering data equation. It has wide applications in medical imaging and diagnosis [1], nondestructive inspection [2], security screening [3], geophysical remote sensing [4], and so on.

Due to the intrinsic nonlinearity of the EM inverse scattering problem [5], iteration is often adopted to find the optimized solution of the scatterer model parameters. In addition, because it is not easy to obtain *a priori* information of the unknown scatterers, the whole inversion domain is usually discretized into a number of grids, and the pixel-based inversion is employed to solve for the model parameters in all the grids simultaneously. One of the commonly used pixel-based FWI methods is to solve the integral equations iteratively [6]. In each iteration of the inversion procedure, the forward

Manuscript received 26 June 2021; revised 26 September 2021; accepted 9 November 2021. Date of publication 29 December 2021; date of current version 8 September 2022. (*Corresponding author: Feng Han.*)

The authors are with the Institute of Electromagnetics and Acoustics and the Fujian Provincial Key Laboratory of Electromagnetic Wave Science and Detection Technology, Xiamen University, Xiamen 361005, China (e-mail: feng.han@xmu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TAP.2021.3137501>.

Digital Object Identifier 10.1109/TAP.2021.3137501

solver is invoked, and the simulated scattered EM fields at the receiver array are updated. Then, the model parameters in all the pixels of the inversion domain are adjusted according to the mismatch between the simulated scattered EM fields and the measured ones [7]. This process continues until the mismatch becomes lower than a prescribed threshold or the maximum iteration number is achieved. Previous researchers have developed several iterative methods in this framework. For example, in the Born iteration method (BIM) [8], the state and data equations are solved alternately. The model parameters of the scatterers are directly updated in each iteration. BIM has the variants of variational BIM (VBIM) and distorted BIM (DBIM) [8], [9]. They are different from BIM since the perturbations of the model parameters are updated in each iteration instead of the model parameters themselves. Compared with BIM, DBIM and VBIM can reach higher inversion accuracy due to the tiny adjustments of the model parameters in each iteration. Contrast source inversion (CSI) is another type of FWI method [10]. Different from the Born-type methods, it has no forward computation, and the contrast and the contrast source are updated alternately until their total mismatch is minimized [11]. The subspace optimization method (SOM) [12] is similar to CSI. However, the induced current is decomposed into the deterministic part and the ambiguous part by singular value decomposition (SVD). The iteration is implemented in the subspace. It converges faster than CSI and has a more robust antinoise ability [13]. These pixel-based iterative FWI methods work efficiently for the reconstruction of scatterers with irregular shapes and arbitrary parameter distribution. However, the iteration also leads to high computational costs. This becomes more obvious when the scatterer has a large electrical size, and the inversion domain is discretized into a large number of grids.

Neural network-based FWI is a hot topic in recent years since it can, to a large extent, effectively circumvent the iteration and lower the computational cost. Chen *et al.* [14] and Massa *et al.* [15] classified it into four categories, the direct learning approach, the physics-assisted learning approach, the learning-assisted objective-function approach, and other approaches. Direct learning is the most straightforward method since it directly converts the scattered fields into the scatterer model parameters. All the physical parameters and mechanisms of EM inverse scattering are implicitly included in the neural network, which is realized by the off-line training. Representative research works have been presented in [16]–[18]. The drawback of direct learning is that the network must spend

an unnecessary cost to train and learn the underlying wave physics of EM scattering. This intuitively increases both the network complexity and training cost. Therefore, researchers proposed the physics-assisted learning approach in which the scattered fields are first converted into preliminary dielectric images of the scatterers by using traditional approximate inverse solvers, e.g., Born approximation (BA) or backpropagation (BP), and thus, both the input and output of the neural network are the model parameter spatial distribution. The network only deals with pure images with gray values and has no relationship with physics. Typical inversion results have been shown in [16] and [19]–[24]. The learning-assisted objective-function approach is another type of hybrid method, but it is opposite to the physics-assisted learning approach. It uses the neural network to obtain the preliminary dielectric images of the scatterers first and then implements the traditional iterative methods to refine the model parameters [25]–[27]. Strictly speaking, among all these neural network-based FWI methods, only inversion by the direct learning approach is instantaneous. Although the approximate inverse solver in the physics-assisted learning approach is also fast, it still needs some time for the central processing unit (CPU) to accomplish the pixel-based inversion since the multiplication of matrices and vectors is usually necessary to acquire the solutions. However, the training cost of the direct learning approach is intuitively higher than that of the physics-assisted learning approach since the network is completely data-driven, and it must learn the physics of EM inverse scattering.

In this article, we propose a tailored artificial neural network (ANN) in which some known EM scattering physical variables are explicitly embedded. That is to say, the ANN is partially physics-driven instead of purely data-driven. Specifically speaking, we divide the whole ANN into two subnets. The first subnet is a fully connected neural network (FCNN) with the activation function removed. The approximate sensitivity matrix of the EM FWI is fixed inside it. It performs like the BA solver in the physics-assisted learning approach and converts the scattered field data into preliminary scatterer model parameters. In order to reduce the number of network parameters in the first subnet, we apply the principal component analysis (PCA) to reduce the dimension of the sensitivity matrix. The second subnet is the standard convolutional neural network (CNN) U-Net that is used to further enhance the accuracy of the inverted scatterer model parameters. It is worth mentioning that the proposed ANN-based inversion is different from the direct learning approach because the known sensitivity matrix is explicitly embedded inside the ANN, and there is no need to learn it in the training. It is also different from the physics-assisted learning approach since the preliminary scatterer model parameters are obtained by a neural network. The conversion can be instantaneous since the graphics processing unit (GPU) instead of CPU is used to accomplish the multiplication of the matrix and the vector. This design guarantees both the real-time inversion merit of the direct learning approach and the low training cost of the physics-assisted learning approach. Compared with the design of cascaded CNNs given in [24], the first subnet in our semiphysics-driven ANN is the direct BA solver and needs no

training. Its weight is determined by the PCA of the sensitivity matrix and remains unchanged in the training process. By contrast, the first subnet in [24] is trained by BP labels. In addition, the second subnet in our semiphysics-driven ANN is only trained by the final permittivity images. However, in [24], the second part includes several cascaded CNNs and is trained with several multiresolution labels by gradually incorporating the high-frequency components. In addition, compared with the two-step deep learning scheme presented in [17], the first subnet in our semiphysics-driven ANN is purely physics-driven. It needs no training and behaves the same as the BA solver. By contrast, the first subnet in [24] must be trained to optimize its network parameters, e.g., weights. The weights of the first subnet in our semiphysics-driven ANN are only solved once by PCA and embedded inside the network in the whole training and prediction process. The PCA operation is offline and will not raise the complexity of the whole procedure.

The rest of this article is organized as follows. In Section II, we briefly introduce the architecture design of the semiphysics-driven ANN and compare it with the purely data-driven ANN. The way to embed the sensitivity matrix of the BA inverse solver into the semiphysics-driven ANN is also discussed in detail. In Section III, a series of numerical experiments are implemented to justify the superiority of the proposed semiphysics-driven ANN with respect to the purely data-driven ANN. In addition, its drawbacks are also discussed. In Section IV, the proposed ANN is used to invert laboratory-measured data. In Section V, conclusions, discussions, and future work are presented.

## II. THEORY

In this section, we briefly introduce the architectures of the semiphysics-driven and purely data-driven ANNs, the reduction of the sensitivity matrix dimension, and the network training configuration.

### A. ANN Architectures

As shown in Fig. 1, the semiphysics- and purely data-driven ANNs have the same framework. Both of them include two subnets. The first subnet includes two fully connected layers. It has no bias value and activation function. The second subnet is a traditional CNN U-Net. The only difference between the two ANNs lies in the weight values of the first subnet. In the semiphysics-driven ANN, the sensitivity matrix elements of BA are assigned to the weights. They are known and fixed in the training process. The first subnet is used to imitate the calculation process of BA and obtains the preliminary scatterer model parameters from the scattered field data. However, because the dimension of the sensitivity matrix  $\mathbf{S}$  is too large, directly embedding it into the first subnet will increase the capacity of the network. We use PCA to reduce its dimension. The compressed sensitivity matrix is denoted by  $\mathbf{S}_{\text{PCA}}$ , as shown in Fig. 1(a), where the superscript  $-1$  stands for the matrix inverse.  $\mathbf{U}_{\text{proj}}$  is the projection matrix, and its specific derivation will be discussed in Section II-B.

By contrast, in the purely data-driven ANN, the first subnet is to obtain the hidden feature map through two trainable

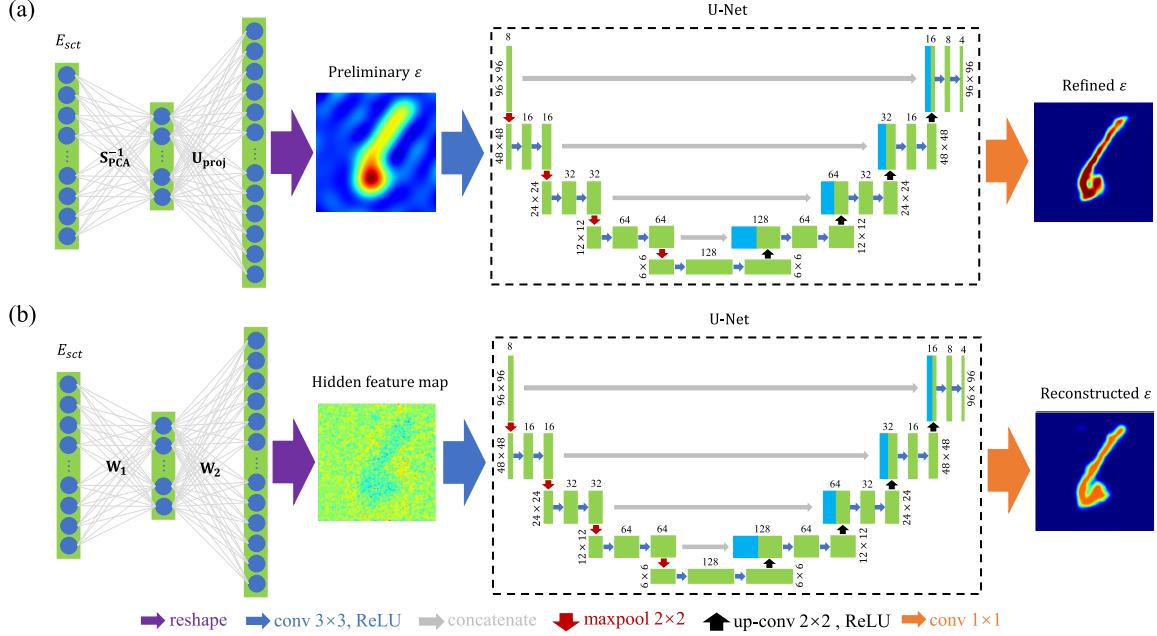


Fig. 1. Architectures of the ANNs. (a) Semiphysics-driven architecture. (b) Purely data-driven architecture.

fully connected layer weights  $\mathbf{W}_1$  and  $\mathbf{W}_2$ , as shown in Fig. 1(b). One should note that  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are optimized in the training process and have no relationship with any EM scattering physics. Their dimensions are the same as those of  $\mathbf{S}_{PCA}^{-1}$  and  $\mathbf{U}_{proj}$ , respectively. The second subnet U-Nets in the two architectures are exactly identical. They include the encoder and the decoder. The encoder consists of four same modules for feature extraction. Each module includes two  $3 \times 3$  convolution layers using the ReLu activation function and a  $2 \times 2$  maximum pooling layer. Following each encoder module, the number of feature map channels is doubled, and the size is reduced by half. The decoder consists of four same modules to restore the original image resolution. Each module includes a  $2 \times 2$  upsampling convolution layer using the ReLu activation function and two  $3 \times 3$  convolution layers using the ReLu activation function. The last layer of the network is a  $1 \times 1$  convolution layer without activation function. The jump connection between the encoder layer and the decoder layer can concatenate the high-level features and low-level features extracted from the network together to remedy the information loss in the downsampling. The details of the U-Net design can be found in [28]. The input of the whole ANN is the measured scattered field vector. Through the two-layer FCNN, it is transformed into the BA image or hidden feature image, and then, the final inversion result is obtained through U-Net.

### B. Reducing Sensitivity Matrix Dimension

The schematic of the 2-D FWI model configuration is shown in Fig. 2. It is assumed that there are  $N_t$  transmitters and  $N_r$  receivers uniformly placed in a circle surrounding the inversion domain  $D$ . The scatterer with the relative permittivity  $\epsilon_r$  is embedded inside the domain  $D$ . The background medium is free space. In the forward scattering computation,  $\epsilon_r$  of the scatterer is known, and the total field  $E_y^{tot}$  in the inversion

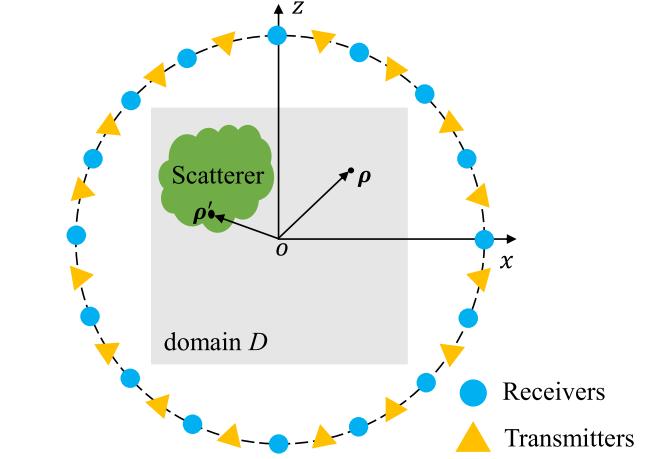


Fig. 2. 2-D FWI model configuration. There are  $N_t$  transmitters and  $N_r$  receivers placed around the inversion domain  $D$ .

domain  $D$  is solved for by the state equation, which is expressed as

$$E_y^{inc}(\rho) = E_y^{tot}(\rho) - j\omega\epsilon_0 \int_D g(\rho, \rho')[\epsilon_r(\rho') - 1] \times E_y^{tot}(\rho') d\rho' \quad (1)$$

where  $E_y^{inc}$  is the incident field in the inversion domain  $D$  when the scatterer is absent. In the inverse scattering computation, the model parameter  $\epsilon_r$  is solved from the scattered field  $E_y^{sct}$  recorded at the receiver array by the data equation, which is expressed as

$$E_y^{sct}(\rho) = j\omega\epsilon_0 \int_D g(\rho, \rho')[\epsilon_r(\rho') - 1] E_y^{tot}(\rho') d\rho' \quad (2)$$

where  $g$  is the 2-D scalar Green's function. It is expressed as [29]

$$g(\rho, \rho') = -\frac{\omega\mu_0}{4} H_0^{(2)}(k_0|\rho - \rho'|) \quad (3)$$

where  $H_0^{(2)}$  is the zeroth-order Hankel function of the second kind and  $k_0 = \omega\sqrt{\epsilon_0\mu_0}$  is the wavenumber in the free space.

It can be seen from the state equation (1) that the total field  $E_y^{tot}$  depends on the model parameter  $\epsilon_r$ . Therefore, the data equation (2) is nonlinear. In the traditional iterative inversion, the total field can be updated by the forward scattering computation in each iteration [30]. However, in the ANN inversion, there is no iteration and the sensitivity matrix is fixed. Thus, we use the BA, i.e., use the incident field to replace the total field, and the data equation becomes

$$E_y^{sct}(\rho) \approx j\omega\epsilon_0 \int_D g(\rho, \rho')[\epsilon_r(\rho') - 1] E_y^{inc}(\rho') d\rho'. \quad (4)$$

We then divide the inversion domain  $D$  into  $N_x \times N_z$  discrete grids and arrange the scattered fields for all the transmitters and receivers in a vector. Equation (4) can be discretized into

$$\mathbf{b}' = \mathbf{S}'\mathbf{x} \quad (5)$$

where

$$\mathbf{b}' = [E_y^{sct}(\rho_{ir}, \rho_{it})] \in \mathbb{C}^{(N_t N_r, 1)} \quad (6a)$$

$$\begin{aligned} \mathbf{S}' &= j\omega\Delta S [g(\rho_{ir}, \rho'_{ix,iz}) E_y^{inc}(\rho'_{ix,iz}, \rho_{it})] \\ &\in \mathbb{C}^{(N_t N_r, N_x N_z)} \end{aligned} \quad (6b)$$

$$\mathbf{x} = [\epsilon_r(\rho'_{ix,iz}) - 1] \in \mathbb{R}^{(N_x N_z, 1)}. \quad (6c)$$

One should note that, in (6),  $it \in [1, N_t]$ ,  $ir \in [1, N_r]$ ,  $ix \in [1, N_x]$ , and  $iz \in [1, N_z]$  are the indexes of transmitters, receivers, discrete grids in the  $\hat{x}$  direction, and discrete grids in the  $\hat{z}$  direction, respectively. The sensitivity matrix  $\mathbf{S}'$  is the same for all training samples since the incident field is the same.  $\Delta S$  is the area of a discrete grid. Equation (5) is a complex-number equation, and a real-number equation can be obtained by splitting it

$$\mathbf{b} = \mathbf{S}\mathbf{x} \quad (7)$$

where

$$\mathbf{b} = \begin{bmatrix} \text{Re}(\mathbf{b}') \\ \text{Im}(\mathbf{b}') \end{bmatrix} \in \mathbb{R}^{(2N_t N_r, 1)} \quad (8a)$$

$$\mathbf{S} = \begin{bmatrix} \text{Re}(\mathbf{S}') \\ \text{Im}(\mathbf{S}') \end{bmatrix} \in \mathbb{R}^{(2N_t N_r, N_x N_z)}. \quad (8b)$$

The sensitivity matrix  $\mathbf{S}$  has a large dimension. Directly embedding it into the first subnet of the semiphysics-driven ANN shown in Fig. 1(a) will lead to a large number of network model parameters. We use PCA to transform  $\mathbf{S}$  into  $\mathbf{S}_{PCA}$ , remove the redundant information, and reduce the dimension of  $\mathbf{S}$ . First, the covariance matrix  $\text{cov}(\mathbf{S})$  of  $\mathbf{S}$  is obtained. Then, the spectral decomposition (SD) of  $\text{cov}(\mathbf{S})$  is performed. The eigenvalue matrix  $\Sigma$  and the eigenvector matrix  $\mathbf{U}$  are rearranged in the order of eigenvalues from large to small. Finally, the first  $K$  columns of the eigenvalue matrix  $\mathbf{U}$  can be taken as the projection matrix  $\mathbf{U}_{proj} \in \mathbb{R}^{(N_x N_z, K)}$  if  $\mathbf{S}_{PCA}$  has the dimensions of  $2N_t N_r \times K$ . This procedure can be expressed by the following equations:

$$\text{cov}(\mathbf{S}) = (\mathbf{S} - \text{mean}(\mathbf{S}))^T (\mathbf{S} - \text{mean}(\mathbf{S})) \quad (9a)$$

$$\mathbf{U}\Sigma\mathbf{U}^T = \text{cov}(\mathbf{S}) \quad (9b)$$

$$\mathbf{U}_{proj} = \mathbf{U}[:, 1 : K] \quad (9c)$$

where the value of  $K$  is usually determined by how much variance information is retained. In this work, the value of  $K$  corresponds to retaining 99% variance information, which means that

$$\frac{\sum_{i=1}^K \sigma_{i,i}}{\sum_{i=1}^{N_x N_z} \sigma_{i,i}} \geq 0.99 \quad (10)$$

where  $\sigma_{i,i}$  is the  $i$ th diagonal element of the matrix  $\Sigma \in \mathbb{R}^{(N_x N_z, N_x N_z)}$ . Note the diagonal elements of  $\Sigma$  are arranged in the descending sequence. Finally, the compressed sensitivity matrix  $\mathbf{S}_{PCA}$  and the Born approximate solution  $\mathbf{x}$  can be acquired by the following equations:

$$\mathbf{S}_{PCA} = \mathbf{S}\mathbf{U}_{proj} \quad (11a)$$

$$\mathbf{x}_{PCA} = \mathbf{S}_{PCA}^{-1} \mathbf{b} \quad (11b)$$

$$\mathbf{x} = \mathbf{U}_{proj} \mathbf{x}_{PCA}. \quad (11c)$$

### C. ANN Training Consideration

In this article, the MNIST [31], which is a database of handwriting digits widely used in the field of machine learning, is employed to train both the semiphysics- and purely data-driven ANNs. The scattered field data are simulated by the stabilized biconjugate-gradient fast Fourier transform (BCGS-FFT) forward solver. The details can be found in [32] and will not be repeated here. In the training process, the parameters of the first subnet in the semiphysics-driven ANN are frozen, and only the U-Net following the first subnet is optimized. That is to say,  $\mathbf{S}_{PCA}^{-1}$  and  $\mathbf{U}_{proj}$  shown in Fig. 1(a) are kept fixed in the training. The output of the first subnet is  $\mathbf{x} = \mathbf{U}_{proj} \mathbf{S}_{PCA}^{-1} \mathbf{b}$ . By contrast, in the purely data-driven ANN, both  $\mathbf{W}_1$  and  $\mathbf{W}_2$  in the first subnet and the following U-Net are optimized simultaneously in the training. The output of its first subnet is  $\mathbf{x} = \mathbf{W}_2 \mathbf{W}_1 \mathbf{b}$ .

The loss function of two ANNs is the mean square error (mse), which is expressed as [33]

$$\text{mse} = \frac{1}{2NN_x N_z} \sum_{i=1}^{NN_x N_z} \|y_i - \hat{y}_i\|^2 \quad (12)$$

where  $N$  is the total number of training samples and  $y_i$  is the inverted permittivity value in a certain discrete grid, while  $\hat{y}_i$  is the corresponding label. The optimizer adopted in this work is Adam [34]. The learning rate is 0.001. The number of training epochs is 500. The batch size is 32. These parameters are the same for all the following numerical examples.

## III. NUMERICAL RESULTS

In this section, we use three numerical cases to validate the advantages and disadvantages of the proposed

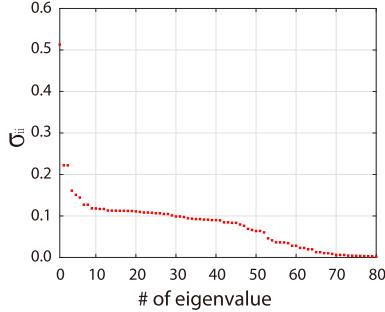


Fig. 3. First 80 eigenvalues of the covariance matrix  $\text{cov}(\mathbf{S})$ .

semiphysics-driven ANN by comparing it with the purely data-driven ANN. As shown in Fig. 2, there are totally 36 transmitters and 36 receivers uniformly placed on a circle with a radius of 3 m. The operating frequency is 300 MHz. Note that the layouts of the transmitter and receiver arrays roughly follow the Nyquist law, i.e., every transmitter or receiver is placed in a half-wavelength interval. The inversion domain with its center located at the origin has the dimensions of  $1.92 \text{ m} \times 1.92 \text{ m}$  and is divided into  $96 \times 96$  pixels. Thus, the dimension of the sensitivity matrix  $\mathbf{S}$  in (8b) is  $96 \times 96 \times 36 \times 36 \times 2 = 23\,887\,872$ . We then do the SD of the covariance matrix of the sensitivity matrix  $\mathbf{S}$ , and the first 80 eigenvalues are displayed in Fig. 3. If we use PCA to reduce the dimension and choose  $K = 68$  to retain 99% variance information, the first subnet in the semiphysics-driven ANN becomes a two-layer FCNN in which  $\mathbf{S}_{\text{PCA}}^{-1}$  and  $\mathbf{U}_{\text{proj}}$  are embedded. The total parameters are reduced to  $96 \times 96 \times 68 + 36 \times 36 \times 2 \times 68 = 802\,944$ , which is only 3.36% of the original amount. In addition, it should be mentioned that the U-Net connected to the first subnet includes 241 657 network parameters, which only take 23% of the total ANN parameters' amount. This also implies that the number of the trainable network parameters of the semiphysics-driven ANN is 23% of that of the purely data-driven ANN. Another interesting comparison is the computation time of the first subnet of the semiphysics-driven ANN with and without PCA applied. Numerical results show that it takes 0.28 s when PCA is applied but 0.38 s without PCA used to predict 3000 samples.

In the first numerical case, both ANNs are trained by the same MNIST datasets with a different number of samples. Both the inversion accuracy for scatterers with different contrasts and the training costs for the different number of samples are compared. In the second case, 400 handwritten letters with different contrasts are used to further test the generalization ability of two trained ANNs. In the third case, the antinoise ability of two ANNs is compared.

In addition, in order to quantitatively evaluate the inversion performance of the proposed ANN, we use the data misfit and model misfit defined in [35, eqs. (16) and (17)]. The data misfit indicates how well the measured scattered fields match the predicted scattered fields. The model misfit indicates how well the reconstructed model parameters match the true model parameters. All the numerical experiments are performed on a workstation with 20-core Xeon E2650 v3 2.3 GHz CPU

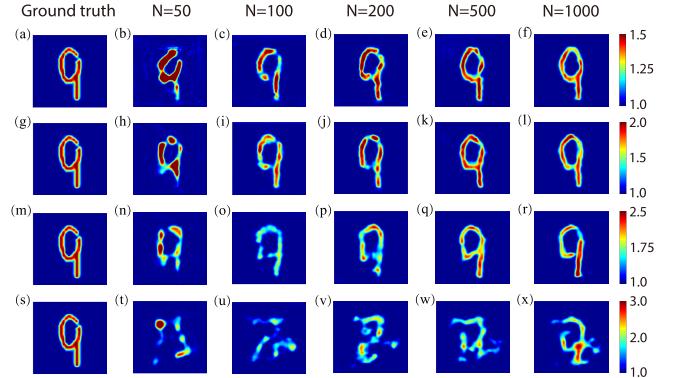


Fig. 4. Ground truths of the digit “9” and the inversion results by the semiphysics-driven ANN for different contrasts when it is trained by datasets with different sizes. The true relative permittivity values of four digits from the first row to the fourth row are 1.5, 2.0, 2.5, and 3.0.

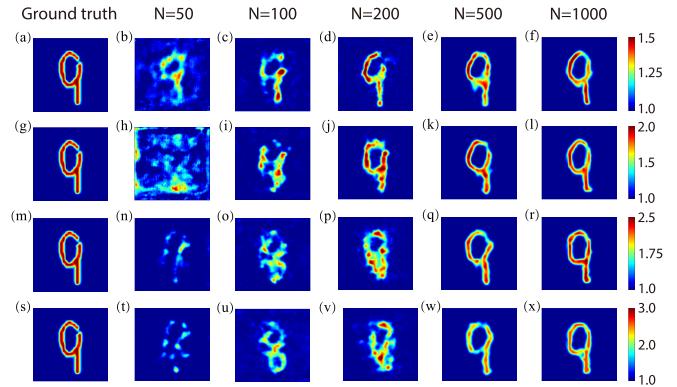


Fig. 5. Ground truths of the digit “9” and the inversion results by the purely data-driven ANN for different contrasts when it is trained by datasets with different sizes. The true relative permittivity values of four digits from the first row to the fourth row are 1.5, 2.0, 2.5, and 3.0.

and 512 GB RAM. ANNs are trained on an NVIDIA Geforce RTX3090 GPU with 24 GB memory.

#### A. Case 1: Handwritten Numerals

The purpose of this case is to compare the inversion accuracy and training cost of two types of ANNs. We use the same handwritten numeral dataset to train both the semiphysics- and purely data-driven ANNs. The relative permittivity of the scatterer varies between 1.2 and 3.0 in the dataset. The sizes of the datasets are chosen as  $N = 50$ ,  $N = 100$ ,  $N = 200$ ,  $N = 500$ , and  $N = 1000$ , respectively.

The digit “9” is selected to test the inversion accuracy of the semiphysics-driven ANN by comparing its results with those from the purely data-driven ANN. Fig. 4 shows the reconstructed scatterer profiles by the semiphysics-driven ANN. From the first row to the fourth row, the results correspond to the scatterer  $\epsilon_r = 1.5, 2.0, 2.5$ , and 3.0, respectively. The first column shows the ground truths. From the second column to the sixth column, the results correspond to the training dataset with the size of  $N = 50, 100, 200, 500$ , and 1000, respectively. In a similar way, Fig. 5 displays the inversion results by

TABLE I  
MODEL MISFITS (%) OF THE RECONSTRUCTED SCATTERER PROFILES FOR DIFFERENT CONTRASTS WHEN TRAINING DATASETS HAVE DIFFERENT SIZES

	Semi-physics-driven ANN					Purely data-driven ANN				
	$N = 50$	$N = 100$	$N = 200$	$N = 500$	$N = 1000$	$N = 50$	$N = 100$	$N = 200$	$N = 500$	$N = 1000$
$\epsilon_r = 1.5$	12.37	5.63	5.43	3.77	3.27	8.42	7.71	7.73	5.04	4.22
$\epsilon_r = 2.0$	13.05	10.24	9.09	6.69	6.30	24.73	14.15	11.18	8.58	6.44
$\epsilon_r = 2.5$	21.38	20.90	18.49	15.27	14.33	26.16	21.02	17.53	13.26	9.77
$\epsilon_r = 3.0$	36.16	35.19	34.72	28.67	27.59	33.99	28.49	25.98	19.60	14.95

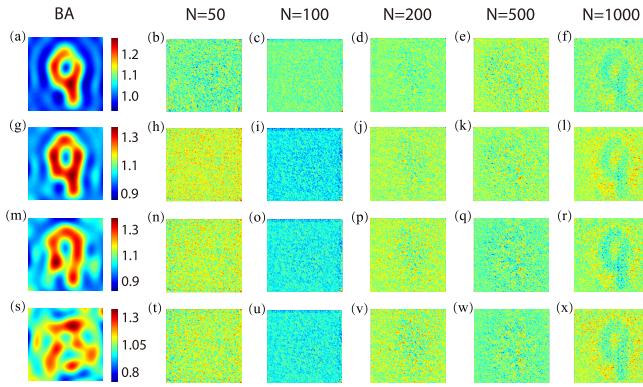


Fig. 6. BA results of the digit “9” for different contrasts by the first subnet in the semiphysics-driven ANN and the hidden feature maps extracted by the first subnet in the purely data-driven ANN when it is trained by datasets with different sizes. The first column shows the BA inversion results. Columns 2–6 show the extracted hidden feature maps by the first subnet of the purely data-driven ANN for the training datasets with different sizes.

the purely data-driven ANN. Fig. 6 shows the reconstructed profiles by the first subnet. The first column shows the BA results that are generated by the first subnet of the semiphysics-driven ANN. Columns 2–6 show the extracted hidden feature maps by the first subnet of the purely data-driven ANN for the training datasets with different sizes.

We then study how the contrast of the scatterer and the training dataset size affect the inversion accuracy and also compare the effect of the semiphysics-driven ANN and that of purely data-driven ANN.

1) *Scatterer Contrast Effect:* It can be seen from Fig. 4 that, with the increase in the scatterer contrast, the inversion results by the semiphysics-driven ANN become worse and worse. This is caused by the BA solver in the first subnet, as shown in the first column of Fig. 6. When the contrast of the scatterer becomes large, the BA results deviate away from the true scatterer profiles severely. In this situation, even the trained U-Net cannot effectively recover the profiles. An interesting observation is that the results from the purely data-driven ANN also show similar variation trends although it has nothing to do with EM scattering physics, as shown in Fig. 5. This is because the nonlinearity of the EM scattering becomes more obvious when the contrast of the scatterer becomes larger. As a result, it is more difficult to optimize the network parameters in the training. Of course, this is only manifested in the final inversion results. The intermediate results of the ANN, e.g., the extracted hidden feature maps shown in Fig. 6, do not display obvious discrepancies for different contrasts. In the purely

data-driven ANN, the intermediate results have no definite physical meaning. Therefore, it is hard to say how they depend on the contrast of the scatterer.

2) *Training Size Effect:* By comparing the results shown in different columns of Figs. 4–6, we can see how the training dataset size affects the inversion performance. A larger training dataset, no matter for the semiphysics- or purely data-driven ANN, usually makes the inversion results better. This is because the network parameters are more fully optimized if more training samples are used in the training. As shown in Fig. 4(n)–(r), the reconstructed profile is poor for the scatterer with  $\epsilon_r = 2.5$  when the training dataset size is 50. However, the inverted model parameter of the scatterer approaches the true value when the size increases to 1000. That is to say, the trained U-Net is capable of remedying the loss of the BA solution. Similar phenomena also occur in the inversion by the purely data-driven ANN. As shown in Fig. 5(t)–(x), even for the scatterer with high contrast, the reconstructed profile becomes better when  $N$  increases. The ANN has the ability to learn the nonlinear relationship between the scattered field data and the scatterer model parameters if there are enough training samples. This is also illustrated by the last column shown in Fig. 6. We can see that the first subnet in the purely data-driven ANN can capture the preliminary image maps when  $N = 1000$ .

3) *Comparison of Two ANNs:* As discussed above, both the scatterer contrast and the training dataset size have significant impacts on the inversion accuracy. However, they show obviously different characteristics for the semiphysics- and purely data-driven ANNs.

For a low contrast scatterer, when the training dataset size is small, the semiphysics-driven ANN outperforms the purely data-driven ANN. This is shown by the subfigures in the top left corners of Figs. 4 and 5. It is also quantitatively proved by the model misfits listed in Table I. Nevertheless, this superiority gradually disappears as the training dataset size increases. For example, when the scatterer has the relative permittivity  $\epsilon_r = 2.0$ , the model misfit of the reconstructed profile by the semiphysics-driven ANN is 10.24% if  $N = 100$ , which is obviously less than the model misfit 14.15% of the reconstructed profile by the purely data-driven ANN. However, when the training dataset size increases to  $N = 1000$ , they are almost the same, as listed in Table I. The reason for the aforementioned comparisons is given as follows. For a low contrast scatterer, the BA solver in the semiphysics-driven ANN is able to provide good preliminary profiles for the following U-Net, as shown in Fig. 6(a) and (g). As a

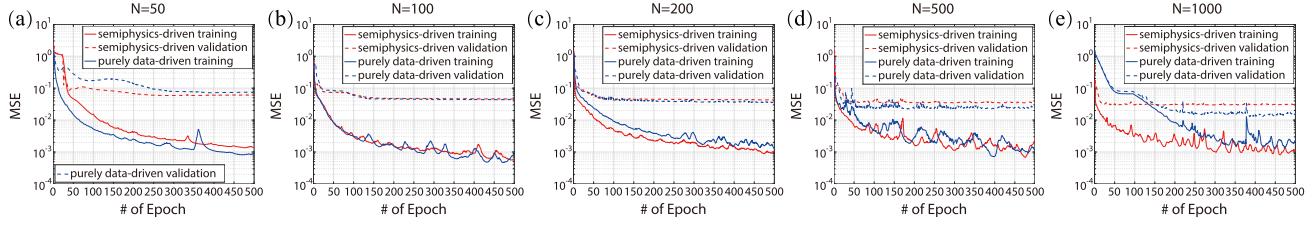


Fig. 7. Comparisons of convergence curves for the training and validation of two ANNs when the datasets have different sizes. (a)  $N = 50$ . (b)  $N = 100$ . (c)  $N = 200$ . (d)  $N = 500$ . (e)  $N = 1000$ .

consequence, the U-Net can produce inverted scatterer parameters with low errors even when the training dataset size is small. This is obviously shown by Fig. 4(c) and (i). By contrast, as shown in Fig. 5(c) and (i), the purely data-driven ANN fails to produce reliable reconstructed scatterer profiles in the online prediction if the training dataset size is limited. This is because the purely data-driven ANN lacks auxiliary physics to extract the preliminary scatterer profiles from the scattered field data and, thus, heavily depends on the training data. As a result, the ANN is not adequately trained, and the final inverted scatterer parameters will show large errors. Fortunately, this can be effectively improved if we increase the training dataset size  $N$  and let the ANN be fully trained, as shown by the subfigures in the top right corners of Figs. 4 and 5. This point is also verified by the hidden feature maps extracted by the first subnet in the purely data-driven ANN shown in the last column of Fig. 6. We can see that just the first subnet will generate good preliminary profiles if  $N$  is increased to 1000.

For a high contrast scatterer, when the training dataset size is small, both the semiphysics-driven ANN and the purely data-driven ANN fail to reconstruct the scatterer model parameters, as shown by the subfigures in the bottom left corners of Figs. 4 and 5. This is also proved by the large model misfit values listed in Table I. Such a failure is caused by the intrinsic nonlinearity of the EM inverse scattering problem. Either the BA solver or the trained ANN cannot effectively overcome the strong nonlinearity if  $N$  is too small. With the increase in the training dataset size, the purely data-driven ANN gradually outperforms the semiphysics-driven ANN, which is shown by subfigures in the bottom right corners of Figs. 4 and 5. That is to say, the whole purely data-driven ANN is effectively trained to fit the nonlinear relationship between the scattered field data and scatterer model parameters when  $N$  is large enough. However, in the semiphysics-driven ANN, the network model parameters in the first subnet, the BA solver, are fixed in the training. It cannot effectively overcome the nonlinearity even if the training dataset size is increased. In this situation, the final inversion results by the following U-Net are still not obviously improved from the poor inversion results by the BA solver even though the U-Net is trained by 1000 samples. This is vividly shown by the comparison between Figs. 4(x) and 5(x) and the corresponding model misfits listed in Table I.

Fig. 7 shows the loss function convergence curves for the training and validation of two types of ANNs when the datasets have different sizes. Because scatterers with different contrast values are present in both the training and validation datasets, the mse values in different training epochs have no obvious difference for the semiphysics-driven ANN and the

purely data-driven ANN. When the training dataset has the size of  $N = 50$ , the validation mse of the semiphysics-driven ANN converges below that of the purely data-driven ANN, as shown in Fig. 7(a). This indicates that the performance of the semiphysics-driven ANN is slightly better than that of the purely data-driven ANN when the training dataset size is small. When the training dataset size increases to  $N = 1000$ , the validation mse of the purely data-driven ANN converges below that of the semiphysics-driven ANN, as shown in Fig. 7(e). This shows that, in this situation, the performance of the purely data-driven ANN is better than that of the semiphysics-driven ANN. Fig. 7(a)–(e) shows that, with the increase in the training dataset size, the performance of the purely data-driven ANN gradually catches up with and exceeds that of the semiphysics-driven ANN.

Another important issue that we want to emphasize here is the training time cost for different dataset sizes and the comparison between two types of ANNs. Because the memory of GPU is limited, we only feed the ANN 32 training samples each time. Therefore, training 1000 samples are significantly more time-consuming than training 50 samples. However, numerical experiments show that the training time in each epoch is almost the same for both types of ANNs if the dataset size is the same although the purely data-driven ANN contains much more network parameters to be optimized than those of the semiphysics-driven ANN. Therefore, in order to achieve the same inversion accuracy, the purely data-driven ANN has a higher training cost than the semiphysics-driven ANN. This is especially obvious when the scatterer contrast is low.

### B. Case 2: Handwritten Letters

In this case, in order to compare the generalization ability of the two ANNs, we use 100 handwritten letters “A” with  $\epsilon_r = 1.5$ , 100 handwritten letters “B” with  $\epsilon_r = 2.0$ , 100 handwritten letters “E” with  $\epsilon_r = 2.5$ , and 100 handwritten letters “K” with  $\epsilon_r = 3.0$  as the testing datasets. Fig. 8 shows the scatter plots of model misfits and data misfits of 100 reconstructed profiles. The data misfits are computed by the forward BCGS-FFT solver. For the letter “A” with  $\epsilon_r = 1.5$ , as shown in the first row of Fig. 8, most blue triangles fall into the lower left regions of the red dots, which means that both the model misfits and data misfits of the inverted profiles by the semiphysics-driven ANN are less than those by the purely data-driven ANN. That is to say, the semiphysics-driven ANN has a stronger generalization ability than the purely data-driven ability when scatterer  $\epsilon_r = 1.5$ . However, we can also see that, with the

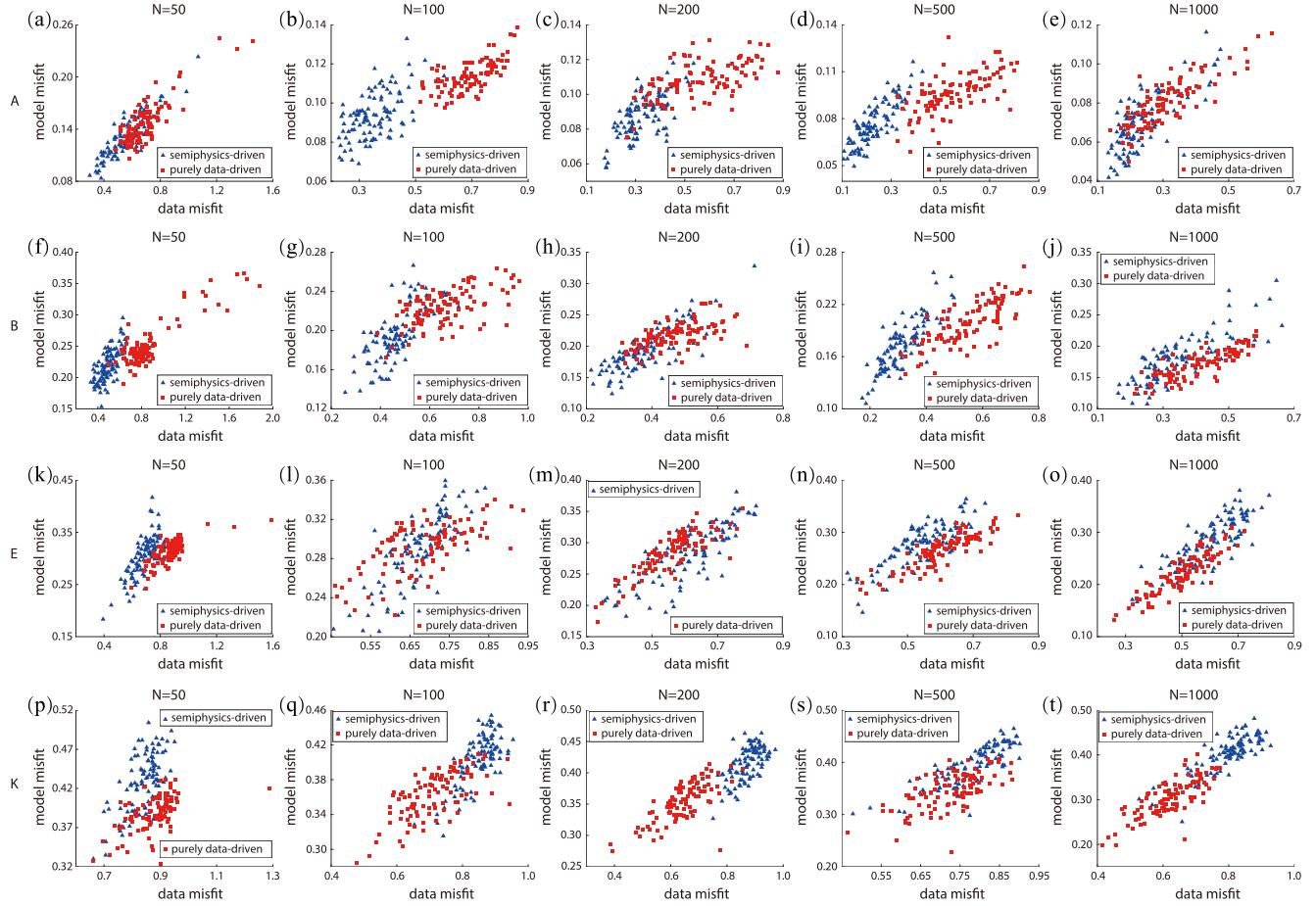


Fig. 8. Scatter plots of data misfits and model misfits of 100 reconstructed scatterer profiles by two ANNs. The first row is for the letters “A” with  $\varepsilon_r = 1.5$ . The second row is for the letters “B” with  $\varepsilon_r = 2.0$ . The third row is for the letters “E,” with  $\varepsilon_r = 2.5$ . The fourth row is for the letters “K” with  $\varepsilon_r = 3.0$ . (a), (f), (k), and (p)  $N = 50$ . (b), (g), (l), and (q)  $N = 100$ . (c), (h), (m), and (r)  $N = 200$ . (d), (i), (n), and (s)  $N = 500$ . (e), (j), (o), and (t)  $N = 1000$ .

increase in the training dataset size, the blue triangles and the red dots gradually mix together. This means the generalization ability superiority of the semiphysics-driven ANN gradually losses. When we increase the scatterer permittivity to  $\varepsilon_r = 2.0$  and 2.5, as shown in the second and third rows for the letters “B” and “E,” this superiority is only maintained for the training dataset size  $N = 50$ . When  $N$  is larger than 50, as shown in Fig. 8(g)–(j) and (l)–(o), it is hard to say the semiphysics-driven ANN outperforms the purely data-driven ANN since the blue triangles and red dots mix together. For the letter “E” with  $\varepsilon_r = 2.5$ , the purely data-driven ANN actually slightly outperforms the semiphysics-driven ANN when  $N = 1000$ , as shown in Fig. 8(o). For the letter “K” with  $\varepsilon_r = 3.0$ , the purely data-driven ANN definitely outperforms the semiphysics-driven ANN, which is displayed in the last row of Fig. 8.

In order to more directly compare the generalization ability of two ANNs, we then choose four different specific samples from four testing sets, respectively. Fig. 9 shows the selected inversion results by the semiphysics-driven ANN, while Fig. 10 shows the selected inversion results by the purely data-driven ANN. Fig. 11 shows the outputs of the first subnet. The first column displays the results of BA, and the second to sixth columns display the hidden feature maps extracted by the first subnet of the purely data-driven ANN. By comparing

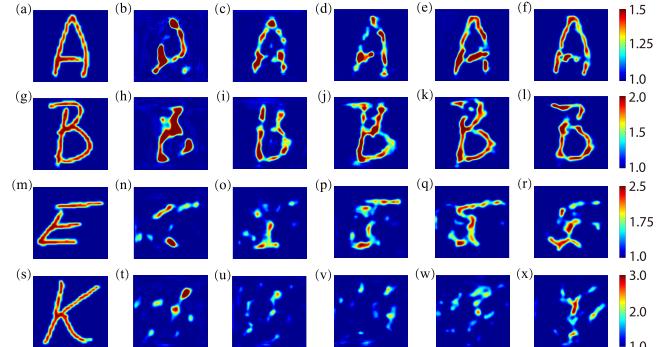


Fig. 9. Ground truths of the letter “A” with  $\varepsilon_r = 1.5$ , “B” with  $\varepsilon_r = 2.0$ , “E” with  $\varepsilon_r = 2.5$ , “K” with  $\varepsilon_r = 3.0$ , and the corresponding inversion result by the semiphysics-driven ANN when it is trained by datasets with different sizes. (a), (g), (m), and (s) Ground truth. (b), (h), (n), and (t)  $N = 50$ . (c), (i), (o), and (u)  $N = 100$ . (d), (j), (p), and (v)  $N = 200$ . (e), (k), (q), and (w)  $N = 500$ . (f), (l), (r), and (x)  $N = 1000$ .

the first two rows in Figs. 9 and 10, we can see that the semiphysics-driven ANN has a stronger generalization ability than the purely data-driven ANN when the scatterer  $\varepsilon_r \leq 2.0$ , which is especially obvious if the training dataset size is small. This is because the BA solver only depends on EM scattering physics and has no relationship with training data. Its generalization ability is theoretically infinitely large. Therefore, the superiority of the whole semiphysics-driven ANN is directly

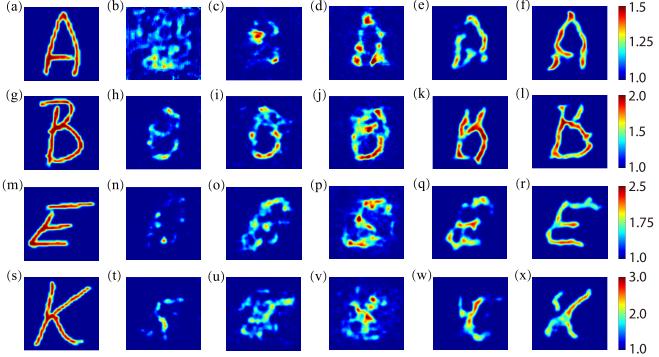


Fig. 10. Ground truths of the letter “A” with  $\epsilon_r = 1.5$ , “B” with  $\epsilon_r = 2.0$ , “E” with  $\epsilon_r = 2.5$ , “K” with  $\epsilon_r = 3.0$ , and the corresponding inversion result by the purely data-driven ANN when it is trained by datasets with different sizes. (a), (g), (m), and (s) Ground truth. (b), (h), (n), and (t)  $N = 50$ . (c), (i), (o), and (u)  $N = 100$ . (d), (j), (p), and (v)  $N = 200$ . (e), (k), (q), and (w)  $N = 500$ . (f), (l), (r), and (x)  $N = 1000$ .

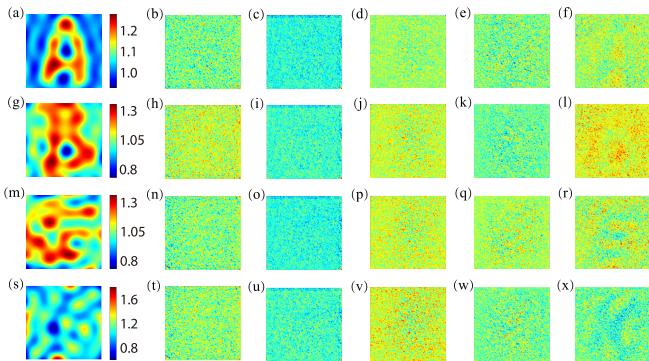


Fig. 11. BA results of four letters by the first subnet in the semiphysics-driven ANN and the hidden feature maps extracted by the first subnet in the purely data-driven ANN when it is trained by datasets with different sizes. The first column shows the BA inversion results of four letters. Columns 2–6 show the hidden feature maps. Rows 1–4 are for the letter “A” with  $\epsilon_r = 1.5$ , “B” with  $\epsilon_r = 2.0$ , “E” with  $\epsilon_r = 2.5$ , and “K” with  $\epsilon_r = 3.0$ , respectively. (a), (g), (m), and (s) BA. (b), (h), (n), and (t)  $N = 50$ . (c), (i), (o), and (u)  $N = 100$ . (d), (j), (p), and (v)  $N = 200$ . (e), (k), (q), and (w)  $N = 500$ . (f), (l), (r), and (x)  $N = 1000$ .

manifested when the scatterer contrast is low, and the training dataset size is small. However, when the dataset size becomes large, its advantage is weakened. The hidden feature maps extracted by the first subnet in the purely data-driven ANN also show good preliminary profiles, as shown in Fig. 11(f) and (l). When the scatterer  $\epsilon_r$  is larger than 2.0, the semiphysics-driven ANN has a poor generalization ability, as shown in the third and fourth rows of Fig. 9. This is because the BA results of scatterers with high contrasts are poor, which are shown in Fig. 11(m) and (s). In this situation, the U-Net connecting to the BA solver cannot distinguish the severely distorted testing letters that are far from the digits in the training datasets. This drawback is effectively overcome by the purely data-driven ANN, supposing that it is trained by enough samples, as shown in Fig. 10(r) and (x). The first subnet of the purely data-driven ANN that takes around 76% amount of the parameters of the whole ANN is fully optimized in the training process when the dataset size is large. Therefore, it has stronger adaptability to scatterers with high contrasts than the BA solver.

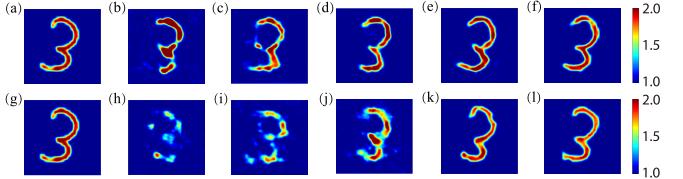


Fig. 12. Ground truths of the digit “3” with the relative permittivity of 2.0 and the inversion results by two ANNs for different training dataset sizes when the scattered field data are contaminated by 10 dB noise. The results in the first row are reconstructed by the semiphysics-driven ANN. The results in the second row are reconstructed by the purely data-driven ANN. (a) and (g) Ground truth. (b) and (h)  $N = 50$ . (c) and (i)  $N = 100$ . (d) and (j)  $N = 200$ . (e) and (k)  $N = 500$ . (f) and (l)  $N = 1000$ .

### C. Case 3: Comparison of the Antinoise Ability of Two ANNs

In this case, we compare the antinoise ability of two ANNs by using them to invert the simulated scattered field data contaminated by 10 dB noise. Here, the noise level is defined according to the signal-to-noise ratio (SNR) of power. Fig. 12 shows the reconstruction results of the digit “3” with the relative permittivity 2.0. The first row and the second row represent the inversion results of the semiphysics-driven ANN and the purely data-driven ANN, respectively. It can be seen that both ANNs have the strong antinoise ability, and their inversion performance is hardly affected even by 10 dB noise. The neural network can naturally suppress the noise effect. Therefore, in the presence of noise, the inversion performance comparison between two ANNs is basically consistent with the conclusion for the previous noise-free cases.

## IV. INVERSION OF LABORATORY-MEASURED DATA

In this section, we use the experimental data measured at the Institute Fresnel [36] to further test two ANNs. The transmitter arrays and the receiver arrays are, respectively, placed on a circle with the radius of 0.72 and 0.76 m surrounding the “twodielTM” profile with the relative permittivity of 3.0. The operating frequency is 3 GHz. Totally, 36 transmitters are used to radiate EM waves, and 49 receivers are used to record data. The measured data are calibrated by multiplying them with a single complex-valued coefficient, which is derived from the ratio of the measured incident field and the simulated one at the receiver located at the opposite position of the source [16], [37]. The inversion domain is set to be 0.192 m  $\times$  0.192 m and is discretized into 96  $\times$  96 pixels. We use the same transmitter and receiver configuration of the experiment to generate simulated training data. Multiple scatterers are allowed to present in the inversion domain. Meanwhile, considering the difference between the measured scattered fields and the simulated scattered fields, we add random noise to the simulated scattered field data in the training dataset.

Fig. 13 shows the reconstructed profiles of two cylinders by two ANNs when they are trained by datasets with different sizes. It can be seen that both ANNs can invert the experimental data. However, because of the simple shapes and small electrical sizes of the scatterers, two ANNs can obtain good inversion results even when the training dataset only has 50 samples. In addition, it is noted that the semiphysics-driven ANN outperforms the purely data-driven ANN when the training dataset size is small, e.g., when  $N = 100$ . Meanwhile,

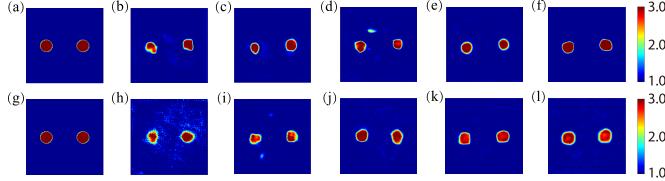


Fig. 13. Ground truths of the “twodielTM” profile and the inversion results by two ANNs when they are trained by datasets with different sizes. The results in the first row are reconstructed by the semiphysics-driven ANN. The results in the second row are reconstructed by the purely data-driven ANN. (a) and (g) Ground truth. (b) and (h)  $N = 50$ . (c) and (i)  $N = 100$ . (d) and (j)  $N = 200$ . (e) and (k)  $N = 500$ . (f) and (l)  $N = 1000$ .

with the increase in the training dataset size, the inversion accuracy is also improving. These phenomena are the same as those observed in the numerical results.

## V. CONCLUSION, DISCUSSION, AND FUTURE WORK

In this article, a semiphysics-driven ANN with the approximate sensitivity matrix embedded inside is proposed, and its performance is compared with a purely data-driven ANN, which has the exact same architecture. The difference between two ANNs lies in their first fully connected subnets. In the semiphysics-driven ANN, the first subnet is the BA solver, and its parameters are fixed in the training. In other words, its solution depends on EM scattering physics and has no relationship with training data. By contrast, in the purely data-driven ANN, the first subnet is trained along with the whole ANN. Both handwritten numeral and letter datasets are used to test two ANNs. Their performance comparisons are summarized as follows.

- For scatterers with low contrasts, both the inversion accuracy and the generalization ability of the semiphysics-driven ANN are better than those of the purely data-driven ANN if the training dataset size is small. The discrepancy is gradually weakened as the dataset size increases. This is because the BA solver is physics-driven, and it can provide good preliminary scatterer profiles for the following U-Net. Therefore, the U-Net can produce reliable inversion results even if it is trained by a limited number of samples. However, because the purely data-driven ANN completely depends on the training data and can only learn the nonlinear mapping between the scattered field data and the scatterer model parameters in the training, only when it is trained by enough samples, its network parameters can be sufficiently optimized.
- For scatterers with high contrasts, both the semiphysics-driven ANN and the purely data-driven ANN fail to obtain reliable scatterer profiles if the training dataset size is small. The nonlinear relationship between the scattered field data and the scatterer model parameters cannot be overcome by the BA solver. It is also difficult for the purely data-driven ANN to learn it if the training samples are limited. However, the purely data-driven ANN outperforms the semiphysics-driven ANN for both the inversion accuracy and the generalization ability if the training dataset size is large enough. The key reason is that

the BA solver distorts the inverted profile too much if the scatterer contrast is too high. As a result, the following U-Net cannot distinguish the distorted profile. This drawback is especially obvious in the tests for generalization ability since the handwritten letters are far from the handwritten numerals. By contrast, because the first subnet in the purely data-driven ANN is effectively trained when there are enough samples, it successfully learns the nonlinear mapping between the scattered field data and the scatterer model parameters.

- The training time in each epoch is roughly the same for both ANNs although their trainable network parameters are obviously different. However, due to the limited size of GPU memory, only a limited number of samples can be fed to the GPU. A dataset with a larger size will cost more time in the training. In this sense, the semiphysics-driven ANN outperforms the purely data-driven ANN for the training cost when the scatterer contrast is low.

From the above discussion, we can see that the bottleneck of the semiphysics-driven ANN is the BA solver. Because its inversion has large errors when the scatterer contrast is high, the following U-Net cannot benefit from the preliminary model parameters provided by it. Therefore, future work will be focused on the improvement of the physics-driven module. A physics-driven mathematical equation solver that needs limited network model parameters to depict but has strong adaptability to all kinds of EM scattering scenarios is desired.

## REFERENCES

- C. Kaye, I. Jeffrey, and J. LoVetri, “Improvement of multi-frequency microwave breast imaging through frequency cycling and tissue-dependent mapping,” *IEEE Trans. Antennas Propag.*, vol. 67, no. 11, pp. 7087–7096, Nov. 2019.
- A. Sophian, G. Y. Tian, D. Taylor, and J. Rudlin, “Electromagnetic and eddy current NDT in weld inspection: A review,” *Insight Non-Destructive Test. Cond. Monit.*, vol. 43, no. 5, pp. 302–306, 2001.
- F. García-Rial, D. Montesano, I. Gómez, C. Callejero, F. Bazus, and J. Grajal, “Combining commercially available active and passive sensors into a millimeter-wave imager for concealed weapon detection,” *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 3, pp. 1167–1183, Mar. 2019.
- G. D. Egbert and A. Kelbert, “Computational recipes for electromagnetic inverse problems,” *Geophys. J. Int.*, vol. 189, no. 1, pp. 251–267, Apr. 2012.
- X. Chen, *Computational Methods for Electromagnetic Inverse Scattering*. Hoboken, NJ, USA: Wiley, 2018.
- Z. Q. Zhang and Q. H. Liu, “Three-dimensional nonlinear image reconstruction for microwave biomedical imaging,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, pp. 544–548, Mar. 2004.
- Y. M. Wang and W. C. Chew, “An iterative solution of the two-dimensional electromagnetic inverse scattering problem,” *Int. J. Imag. Syst. Technol.*, vol. 1, no. 1, pp. 100–108, Jun. 1989.
- F. Li, Q. H. Liu, and L.-P. Song, “Three-dimensional reconstruction of objects buried in layered media using Born and distorted Born iterative methods,” *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 2, pp. 107–111, Apr. 2004.
- N. Zaiping, Y. Feng, Z. Yanwen, and Z. Yerong, “Variational Born iteration method and its applications to hybrid inversion,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 1709–1715, Jul. 2000.
- P. M. van den Berg and R. E. Kleinman, “A contrast source inversion method,” *Inverse Problems*, vol. 13, no. 6, pp. 1607–1620, 1997.
- A. Abubakar and P. M. van den Berg, “Three-dimensional inverse scattering applied to cross-well induction sensors,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 1669–1681, Jul. 2000.
- X. Chen, “Subspace-based optimization method for solving inverse-scattering problems,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 42–49, Jan. 2010.

- [13] Y. Liu *et al.*, "A frequency-hopping subspace-based optimization method for reconstruction of 2-D large uniaxial anisotropic scatterers with TE illumination," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6091–6099, Oct. 2016.
- [14] X. Chen, Z. Wei, M. Li, and P. Rocca, "A review of deep learning approaches for inverse scattering problems," *Prog. Electromagn. Res.*, vol. 167, pp. 67–81, 2020.
- [15] A. Massa, D. Marcantonio, X. Chen, M. Li, and M. Salucci, "DNNs as applied to electromagnetics, antennas, and propagation—A review," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2225–2229, Nov. 2019.
- [16] Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 1849–1860, Apr. 2019.
- [17] H. M. Yao, W. E. I. Sha, and L. Jiang, "Two-step enhanced deep learning approach for electromagnetic inverse scattering problems," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2254–2258, Nov. 2019.
- [18] L.-Y. Xiao, J. Li, F. Han, W. Shao, and Q. H. Liu, "Dual-module NMM-IEM machine learning for fast electromagnetic inversion of inhomogeneous scatterers with high contrasts and large electrical dimensions," *IEEE Trans. Antennas Propag.*, vol. 68, no. 8, pp. 6245–6255, Aug. 2020.
- [19] L. Li, L. G. Wang, F. L. Teixeira, C. Liu, A. Nehorai, and T. J. Cui, "DeepNIS: Deep neural network for nonlinear electromagnetic inverse scattering," *IEEE Trans. Antennas Propag.*, vol. 67, no. 3, pp. 1819–1828, Mar. 2019.
- [20] J. Xiao, J. Li, Y. Chen, F. Han, and Q. H. Liu, "Fast electromagnetic inversion of inhomogeneous scatterers embedded in layered media by Born approximation and 3-D U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1677–1681, Oct. 2020.
- [21] L. Zhang, K. Xu, R. Song, X. Ye, G. Wang, and X. Chen, "Learning-based quantitative microwave imaging with a hybrid input scheme," *IEEE Sensors J.*, vol. 20, no. 24, pp. 15007–15013, Dec. 2020.
- [22] X. Ye, Y. Bai, R. Song, K. Xu, and J. An, "An inhomogeneous background imaging method based on generative adversarial network," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 11, pp. 4684–4693, Nov. 2020.
- [23] Y. Huang, R. Song, K. Xu, X. Ye, C. Li, and X. Chen, "Deep learning-based inverse scattering with structural similarity loss functions," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4900–4907, Feb. 2021.
- [24] K. Xu, C. Zhang, X. Ye, and R. Song, "Fast full-wave electromagnetic inverse scattering based on scalable cascaded convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, early access, Jul. 21, 2021, doi: [10.1109/TGRS.2021.3093100](https://doi.org/10.1109/TGRS.2021.3093100).
- [25] Y. Sanghvi, Y. Kalepu, and U. Khankhoje, "Embedding deep learning in inverse scattering problems," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 46–56, 2020.
- [26] G. Chen, P. Shah, J. Stang, and M. Moghaddam, "Learning-assisted multimodality dielectric imaging," *IEEE Trans. Antennas Propag.*, vol. 68, no. 3, pp. 2356–2369, Mar. 2020.
- [27] Y. Chen, L.-Y. Xiao, J. Zhuo, F. Han, and Q. H. Liu, "Quantitative electromagnetic inversion of irregular scatterers based on a threefold hybrid method," *IEEE Trans. Antennas Propag.*, vol. 69, no. 12, pp. 8664–8674, Dec. 2021.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.*, 2015, pp. 234–241.
- [29] C. A. Balanis, *Advanced Engineering Electromagnetics*. New York, NY, USA: Wiley, 2012, ch. 11.
- [30] Y. Chen, P. Wen, F. Han, N. Liu, H. Liu, and Q. H. Liu, "Three-dimensional reconstruction of objects embedded in spherically layered media using variational Born iterative method," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1037–1041, Jul. 2017.
- [31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [32] T. Lan, N. Liu, Y. Liu, F. Han, and Q. H. Liu, "2-D electromagnetic scattering and inverse scattering from magnetodielectric objects based on integral equation method," *IEEE Trans. Antennas Propag.*, vol. 67, no. 2, pp. 1346–1351, Feb. 2019.
- [33] S.-W. Fu, T.-W. Wang, Y. Tsao, X. Lu, and H. Kawai, "End-to-end waveform utterance enhancement for direct evaluation metrics optimization by fully convolutional neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 9, pp. 1570–1584, Sep. 2018.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [35] T. Lan, N. Liu, F. Han, and Q. H. Liu, "Joint petrophysical and structural inversion of electromagnetic and seismic data based on vol. integral, equation method," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2075–2086, Apr. 2019.
- [36] K. Belkebir and M. Saillard, "Testing inversion algorithms against experimental data," *Inverse Problems*, vol. 17, no. 6, p. 1565, 2001.
- [37] J.-M. Geffrin, P. Sabouroux, and C. Eyraud, "Free space experimental scattering database continuation: Experimental set-up and measurement precision," *Inverse Problems*, vol. 21, no. 6, pp. S117–S130, Dec. 2005.



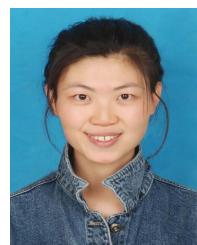
**Yanjin Chen** received the B.S. degree in mechanical and electronic engineering from Hainan University, Hainan, China, in 2017, and the M.S. degree in the electromagnetic field and wireless technology from Xiamen University, Xiamen, China, in 2021.

He is currently a Research Assistant with the Fujian Provincial Key Laboratory of Electromagnetic Wave Science and Detection Technology, Xiamen University. His research interest is applying machine learning techniques to electromagnetic inverse scattering problems.



**Miao Zhong** received the B.S. degree in communication engineering from Nanchang Hangkong University, Nanchang, China, in 2020. She is currently pursuing the master's degree with Xiamen University, Xiamen, China.

Her research interests include electromagnetic inverse scattering and full-wave inversion.



**Zhen Guan** received the B.S. degree in mathematics and applied mathematics and the M.S. degree in pure mathematics from the Tianjin University of Technology and Education, Tianjin, China, in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree in computational science with Xiamen University, Xiamen, China.

Her research interests include the fast algorithms for computational electromagnetics and electromagnetic scattering and inverse scattering problems.



**Feng Han** (Senior Member, IEEE) received the B.S. degree in electronic science from Beijing Normal University, Beijing, China, in 2003, the M.S. degree in geophysics from Peking University, Beijing, in 2006, and the Ph.D. degree in electrical engineering from Duke University, Durham, NC, USA, in 2011.

Since July 2015, he has been with Xiamen University where he is currently an Associate Professor with the Institute of Electromagnetics and Acoustics. He has published over 50 articles in refereed journals. His research interests include electromagnetic scattering and inverse scattering in complex media, fast full-wave electromagnetic inversion based on machine learning, the configuration of the antenna array for electromagnetic inverse problems, and geophysical electromagnetic exploration and inversion.