SOLVING INVERSE PROBLEMS WITH DEEP LEARNING

LEXING YING

ABSTRACT

We discuss some recent work on applying deep learning to inverse problems. On the algorithmic side, we propose two new neural network modules, BCR-Net and Switch-Net, motivated by pseudodifferential and Fourier integral operators that commonly appear in the study of inverse problems. On the application side, we propose neural networks for inverse maps in five applications: electric impedance tomography, optical tomography, inverse acoustic scattering, seismic imaging, and traveltime tomography. In each application, the architecture is motivated by perturbation theory and filtered backprojection, and is implemented using the new modules along with convolution layers. When translation and rotation equivariances are available, appropriate reparameterizations in the data and model domains result in convolutional architectures that are both general and effective. These applications demonstrate that our approach provides a seamless way for combining the mathematical structure of the inverse problems with the power of deep neural networks.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 65N21; Secondary 65R32, 74J25, 86A22

KEYWORDS

Inverse problems, neural networks, deep learning

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 7, pp. 5154–5175
DOI 10.4171/ICM2022/49

1. INTRODUCTION

In the past decade, deep learning (DL) has become the dominant approach in computer vision, image processing, speech recognition, and many other applications in machine learning and data science [26, 34, 40, 42, 43, 48, 58, 60]. From a technical point of view, this success is a synergy of several key developments: (1) deep neural networks (NNs) as a flexible framework for representing high-dimensional functions and maps, (2) simple algorithms such as backpropagation (BP) and stochastic gradient descent (SGD) for tuning the model parameters, (3) efficient general software packages such as Tensorflow [1] and Pytorch [52], and (4) unprecedented computing power provided by GPUs and TPUs. Despite the successes, however, there remain a number of outstanding challenges: (1) NN architectural design is still an art and lacks basic mathematical principles in many cases; (2) NN training often requires an enormous amount of data, which is infeasible in many applications; and (3) a general mathematical theory of deep learning is still lacking.

Many computational problems in physical sciences face the same challenges as those in data science: high-dimensionality, complex or unspecified models, and high computational costs. Some well-known examples include many-body quantum systems, deterministic and stochastic control, molecular dynamics, uncertainty quantification, inverse problems, etc. There is a clear opportunity to leverage the recent developments of DL in the study of these problems. Indeed, the past few years have witnessed a rise of activities in this direction [2, 5, 7, 8, 12, 16, 18, 19, 30, 31, 35–37, 41, 44, 46, 47, 54, 55, 57, 61].

Among these topics, this paper focuses on inverse problems, i.e., recovering unknown interior parameters from boundary measurements. It is a field of enormous importance, with applications in physics, chemistry, medicine, earth sciences, etc. From a computational perspective, many inverse problems are quite challenging for several well-understood reasons: (1) the inverse map from boundary measurements to interior parameters is high-dimensional and nonlinear; (2) asymptotic methods based on perturbation theory often have low accuracy, while fully optimization-based iterative algorithms are often time-consuming; (3) most solution methods are not designed to adapt to data priors, when they are available.

Contributions. We argue that applying deep learning to the study of inverse problems is a fruitful mathematical research direction. On the one hand, NNs offer a flexible tool for representing the high-dimensional inverse maps. They also learn from the data distribution prior effectively via training. On the other hand, the rich mathematical and physical theories behind inverse problems provide guiding principles for designing compact, yet effective NN architectures. As a result, we avoid the need for enormous amounts of data, which are often not available for inverse problems.

The main contributions of this line of study are two-fold. On the *algorithmic* side, we first identify the mathematical operators commonly used in inverse problems, with two such examples being pseudodifferential operators (PDOs) and Fourier integral operators (FIOs) [59]. By leveraging analytical results from partial differential equation (PDE) theory and numerical linear algebra (NLA), we propose novel NN modules for these key types of operators.

On the *application* side, we apply this approach to five different inverse problems: electric impedance tomography, optical tomography, inverse acoustic scattering, seismic imaging, and traveltime tomography. For each application, using the linearized theory and perturbative expansion as a starting point, we approximate the inverse map with a composition sequence of operators. The NN for the inverse map is then assembled using the corresponding modules, along with existing primitives such as convolution neural networks (CNNs). Finally, the weights of the whole network are trained end-to-end with the available training data.

Organization. The rest of the paper is organized as follows. Section 2 describes new NN modules motivated by PDOs and FIOs. Section 3 details the five inverse problems. Finally, Section 4 concludes with a discussion of future directions.

2. NEW, MATHEMATICALLY-MOTIVATED NN MODULES

If one takes a close look at the successful NN architectures in the literature, it is not hard to see that behind each there is a powerful mathematical structure, tabulated as follows.

NN architecture	Mathematical structure
Fully-connected layer	Dense operator
Convolution layer	Translation-invariant local operator
Recurrent neural network (RNN)	Markov chain
ResNet	ODE/time-stepping/semigroup

For the inverse problem theory, two types of commonly occurring operators are pseudodifferential operators (PDOs) and Fourier integral operators (FIOs). In this section, we propose novel NN modules for efficient and accurate representations of these two types of operators.

2.1. Pseudodifferential operators

A pseudodifferential operator (PDO) K is of the form

$$(Kf)(x) \equiv \int k(x,y) f(y) \mathrm{d}y = \int a(x,\xi) e^{2\pi i x\cdot\xi} \hat{f}(\xi) \mathrm{d}\xi, \tag{2.1}$$

where the *symbol* function $a(x,\xi)$ of the PDO is smooth away from the origin of the frequency variable ξ [59]. PDOs are powerful generalizations of standard differential operators. When applied to a function f, the support of the singularities in the output Kf is contained in the singularity support of the input. Some well-known examples of PDOs include the Green's functions of elliptic operators, fractional Laplacians, etc. When a PDO is translation-equivariant, it becomes a convolution and thus can be represented with a convolution layer, though this representation is often not effective for highly nonlocal PDOs. More importantly, non-translation-equivariant PDOs cannot be represented using convolution layers.

One of the key properties of PDOs is that, when discretized with local basis functions, the off-diagonal blocks of the matrix form of a PDO are numerically low-rank [28,29]. This property gives rise to highly effective data-sparse approximations to PDOs, and the one adopted here is based on wavelet analysis. Motivated by this approximation, we propose a novel NN module associated with PDOs by

- representing the data-sparse approximation of PDOs as a (linear) NN,
- enriching its representation power by including intermediate layers and nonlinearities such as the ReLU activation function.

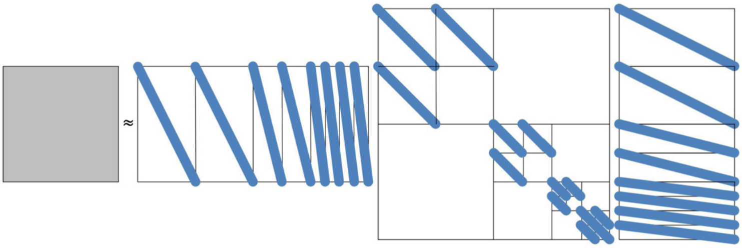


FIGURE 1
The nonstandard wavelet form of a PDO. The first and third matrices on the right-hand side are the inverse and forward transforms for the redundant wavelet/scaling function frame, which can be implemented with fast wavelet transforms in linear complexity. The large middle matrix represents the PDO under this redundant frame, which has a well-defined sparsity pattern with only $O(n)$ nonzero entries.

Wavelet analysis. The data-sparse approximation is based on the nonstandard wavelet form proposed in [9]. Given an $n \times n$ matrix form of a PDO K, the nonstandard form represents the operator in the redundant wavelet/scaling function frame and keeps only $O(n)$ significant coefficients in a well-defined sparsity pattern. Figure 1 illustrates the sparsity pattern, shown in blue.

When applying this PDO to an input vector, the matrix–vector multiplication (MatVec) at each wavelet scale can be written as a three-layer NN with two channels in the middle (see Figure 2 (left) for an illustration). Putting together the networks across all scales gives rise to a linear NN shown in Figure 2 (right).

In order to represent nonlinear operators similar to PDOs, we propose generalizing the architecture in Figure 2 by inserting multiple intermediate layers and including nonlinear activations such as the ReLU function. This results in a new NN module called a *BCR-Net* [17] as shown in Figure 3.

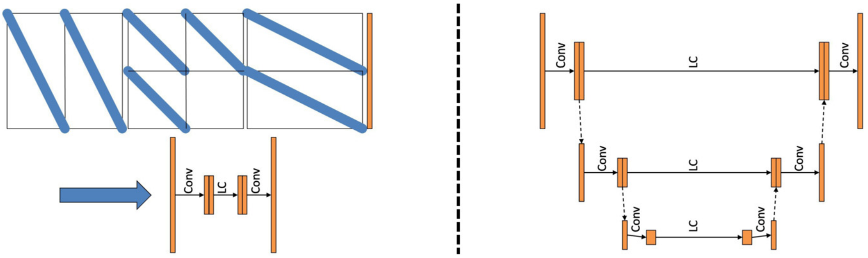


FIGURE 2
A matrix–vector multiplication (MatVec) with an input vector. (Left) The computation at each scale of the wavelet-based data-sparse approximation is a three-layer NN with two channels in the middle. (Right) The NN obtained by merging across all scales. Conv and LC stand for convolution and locally-connected layer, respectively.

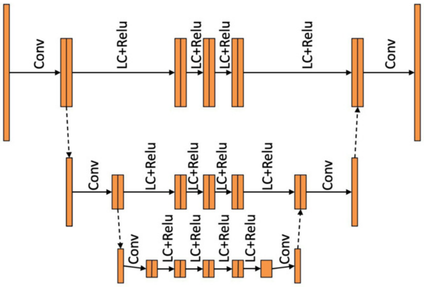


FIGURE 3
The BCR-Net module based on the nonstandard redundant wavelet form for PDOs.

2.2. Fourier integral operators

A Fourier integral operator (FIO) K is of the form

$$(Kf)(x) \equiv \int k(x,y)f(y)\mathrm{d}y = \int a(x,\xi)e^{2\pi i\Phi(x,\xi)}\hat{f}(\xi)\mathrm{d}\xi, \tag{2.2}$$

where the *amplitude* $a(x,\xi)$ of the FIO is smooth away form the origin of ξ and the *phase* $\Phi(x,\xi)$ is homogeneous of degree one in ξ. Viewed as a map from the frequency to the spatial domain (i.e., $\hat{f}$ to Kf), FIOs are generalizations of the Fourier transforms with more general phase and amplitude functions. When applied to a function f, the support of the singularities in the output Kf depends on the input singularities in a well-defined way governed by the Hamiltonian flow of the phase function Φ [59]. Most examples of the FIOs appear in high-frequency wave propagations and scattering theory, and it is for this reason that they are often key to solving wave-based inverse problems.

One key property of FIOs is that, when they are discretized with local basis functions, the resulting matrix representation satisfies the so-called *complementary low-rank property* [45]. More precisely, when the $n \times n$ matrix is partitioned into $\sqrt{n} \times \sqrt{n}$ blocks each of size $\sqrt{n} \times \sqrt{n}$, each block is numerically low-rank. This property allows for an efficient data-sparse approximation, the *butterfly factorization*, to be detailed below. Motivated by the butterfly factorization, we propose a new NN module associated with FIOs by

- representing the butterfly factorization of FIOs as a (linear) NN, and
- enriching its representation power by including intermediate layers and nonlinear activations, such as the ReLU function.

Butterfly factorization. Given an $n \times n$ matrix representation of an FIO K, the simplest form of butterfly factorization partitions the whole matrix into $\sqrt{n} \times \sqrt{n}$ blocks and then computes a low-rank approximation of each block. Figure 4 demonstrates that the low-rank approximations for all blocks can be summarized compactly as the product of three sparse matrices. Notice that the second matrix of the factorization serves essentially as a permutation.

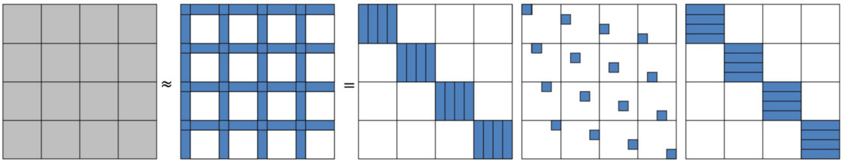


FIGURE 4

Butterfly factorization of an FIO. The middle plot shows the numerical low-rank properties of each $\sqrt{n} \times \sqrt{n}$ block. On the right, the first and third matrices collect the low-rank bases, while the second matrix essentially performs a permutation.

When applying the FIO to an input vector, the MatVec (as shown in Figure 5) can be represented as a three-layer linear NN. Here the first and third matrices become a convolution or locally-connected layer with $\sqrt{n}$ channels, while the second matrix can be implemented with a transpose.

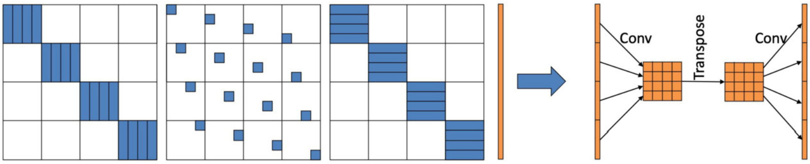


FIGURE 5

A matrix–vector multiplication (MatVec) with an input vector. The computation is represented by a three-layer NN with a transpose operation in the middle. C/LC stands for a convolution or locally-connected layer.

In order to represent nonlinear operators similar to FIOs, we generalize the architecture in Figure 5 by inserting multiple intermediate layers and including nonlinear activations (e.g., ReLU). The resulting new NN module, shown in Figure 6, is called a *Switch-Net* [38].

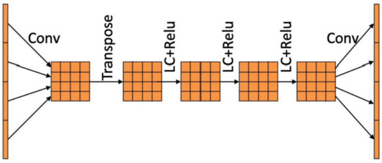


FIGURE 6
The Switch-Net module based on the butterfly factorization for FIOs.

3. INVERSE PROBLEMS

This section describes how to apply deep learning to five inverse problems: electrical impedance tomography, optical tomography, inverse acoustic scattering, seismic imaging, and traveltime tomography. For each problem, we proceed as follows:

- describe the basic setup,
- represent the linearized inverse map as a sequence of operators by following the perturbation theory and filtered backpropagation,
- design the NN architecture by following this sequence and using the new modules in Section 2 as well as CNNs.

Throughout this process, we keep in mind several guiding principles:

- the NN design should adapt to the data collection geometry,
- pre- and post-processing often significantly simplify the NN design, and
- preserving equivariances is the key to efficiency and accuracy.

3.1. Electrical impedance tomography

Consider a rectangular domain Ω (see Figure 7 (left)) with top boundary denoted by Γ. To simplify the presentation, we assume a periodic boundary condition in the horizontal direction. One form of the governing equation for electrical impedance tomography (EIT) is the elliptic equation

$$(Lu)(p) \equiv \big(-\Delta - \eta(p)\big)u(p) = 0, \quad p \in \Omega, \tag{3.1}$$

where we often denote $p = (x, z)$, with x and z being the horizontal and vertical components, respectively. Here $\eta(p)$ is the unknown internal parameter field. In one common form of an EIT experiment, for each boundary point $s \in \Gamma$, we enforce the delta boundary condition $\delta_s(\cdot)$ and then record the normal derivative $\frac{\partial u^s(r)}{\partial n(r)}$ at every point $r \in \Gamma$, where $u^s(\cdot)$ denotes the solution of (3.1) induced by the boundary condition $\delta_s(\cdot)$. The set $d(r, s)$ of boundary measurements is

$$d(r, s) = \frac{\partial u^s(r)}{\partial n(r)} - \frac{\partial u_0^s(r)}{\partial n(r)}, \tag{3.2}$$

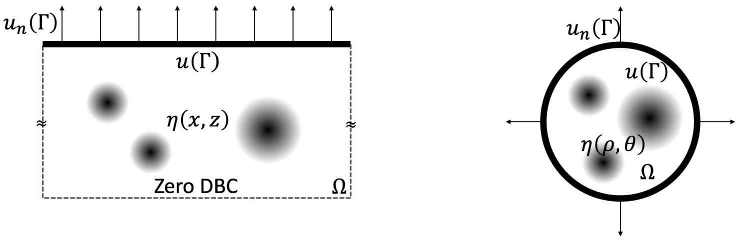


FIGURE 7

Electrical impedance tomography: (left) an experimental setup for a rectangular geometry; (right) an experimental setup in a circular geometry.

where $u_0^s(\cdot)$ stands for the background solution when $\eta(p) \equiv 0$. In technical terms, $d(r,s)$ is the kernel of the *Dirichlet-to-Neumann map* of (3.1). The inverse problem is to recover $\eta(p) \equiv \eta(x,z)$ from $d(r,s)$.

In order to obtain an approximation to the inverse map $d(r,s) \to \eta(x,z)$, we first study how $d(r,s)$ depends on $\eta(x,z)$ in the perturbative regime. Let L_0 be the operator with $\eta(x,z) \equiv 0$ and $G_0 = L_0^{-1}$ be its Green's function. A perturbative analysis in [23] shows that, when η is small, the data $d(r,s)$ can be well approximated with

$$d(r,s) \approx \iint_{(x,z)} \frac{\partial G_0}{\partial n}\big(r,(x,z)\big)\frac{\partial G_0}{\partial n}\big(s,(x,z)\big)\eta(x,z)\mathrm{d}x\mathrm{d}z. \tag{3.3}$$

Due to the translation-equivariance of the background operator L_0, this equation can be simplified when the data d is written in a warped coordinate system (m,h) with $(r,s) \equiv (m+h, m-h)$, namely

$$d(m,h) \approx \iint_{(x,z)} \frac{\partial G_0}{\partial n}\big((m-x)+h,z\big)\frac{\partial G_0}{\partial n}\big((m-x)-h,z\big)\eta(x,z)\mathrm{d}x\mathrm{d}z. \tag{3.4}$$

The key observation is that this is a 1D convolution in m and x with h and z treated as parameters (or as channels in the NN terminology). Furthermore, due to the elliptic nature of the EIT problem, the forward map between η and d is numerically low-rank in h and z. As a result, the number of channels required for this convolution operator is bounded logarithmically in the number of degrees of freedom and the desired accuracy.

The discussion so far shows that, in the small η regime, we can approximate the forward map $K : \eta(x,z) \to d(m,h)$ with a 1D CNN with a small number of channels. The filtered backprojection algorithm suggests that $\eta \approx (K^*K + \varepsilon I)^{-1}K^*d$. This motivates representing the product $(K^*K + \varepsilon I)^{-1}K^*$ as an NN. Regarding K^*, the analysis above for K shows that the adjoint operator K^* can also be approximated with a 1D CNN or BCR-Net with a small number of channels. The operator $(K^*K + \varepsilon I)^{-1}$ is a PDO in the (x,z) domain with global support, which can be approximated with a 2D BCR-Net or even a 2D CNN. Putting them together results in the following NN architecture [23] for the inverse map of the EIT problem:

$$d(m,h) \Rightarrow \text{1D CNN/BCR-Net} \Rightarrow \text{2D CNN/BCR-Net} \Rightarrow \eta(x,z). \tag{3.5}$$

Such an architecture can also be applied directly to the circular geometry (see Figure 7 (right)) if the unknown field η is parameterized in polar coordinates.

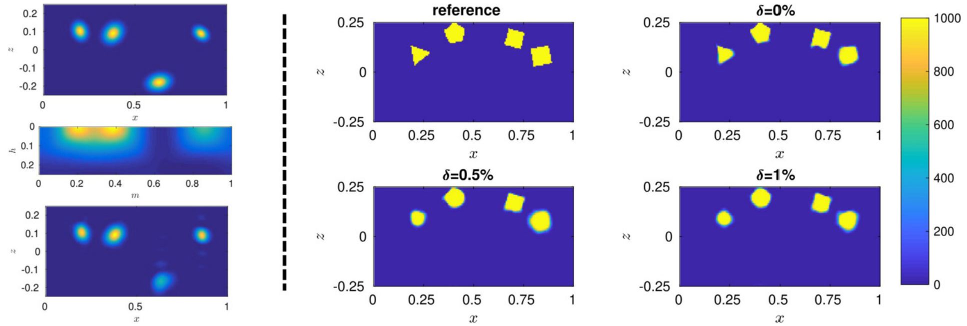


FIGURE 8

Electrical impedance tomography: (left, from top to bottom) the ground truth $\eta(x,z)$, the boundary measurement $d(m,h)$, and the NN reconstruction; (right) the NN reconstructions at different noise levels.

Figure 8 presents a numerical example. The NN has about 70K weights and is trained with about 10K (d,η) training pairs. The left part shows, from top to bottom, the ground truth internal parameters $\eta(x,z)$, the boundary measurements $d(m,h)$ in the warped coordinate system (m,h), and the NN reconstruction obtained by applying the trained NN to $d(m,h)$. The images show that the NN reconstruction is close to the ground truth, though the accuracy gradually deteriorates as the depth z grows due to the nature of the EIT problem. The right part shows how the NN, while trained with noiseless training data, performs under different noise levels in the testing boundary measurement $d(m,h)$. The images demonstrate that the trained NN is robust to measurement noise.

3.2. Optical tomography

Consider a circular domain Ω in 2D (see Figure 9) with boundary $\Gamma = \mathbb{S}^1$. The governing equation for optical tomography (OT) is the radiative transfer equation (RTE)

$$(L\Phi)(p,v) \equiv v \cdot \nabla\Phi(p,v) + \mu_t(x)\Phi(p,v) = \mu(p)\int_{\mathbb{S}^1} \sigma(v \cdot v')\Phi(p,v')\mathrm{d}v',$$
$$(p,v) \in \Omega \times \mathbb{S}^1, \tag{3.6}$$

where σ is a fixed scattering phase with $\int_{\mathbb{S}^1} \sigma(v \cdot v')\mathrm{d}v = 1$. The transport coefficient $\mu_t(p) = \mu_a + \mu(p)$ measures the total absorption, including the *known* physical absorption constant μ_a and the *unknown* scattering strength quantified by the term $\mu(p)$. In a typical OT experiment, for each boundary point $s \in \mathbb{S}^1$, one specifies at s either an isotropic scattering source or a delta source in the normal direction, and records the outgoing flux (denoted by $f^s(\cdot)$) at each point $r \in \mathbb{S}^1$. The set of boundary measurements is given by

$$d(r,s) = f^s(r) - f_0^s(r), \tag{3.7}$$

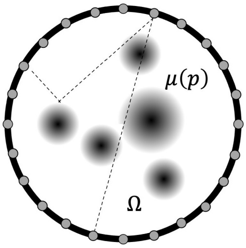


FIGURE 9
Optical tomography. An experimental setup in a circular geometry.

where $f_0^s(\cdot)$ is the outgoing flux when $\mu(x) \equiv 0$. The inverse problem is then to recover $\mu(p)$ from $d(r, s)$.

In order to obtain an approximation for the inverse map $d(r, s) \rightarrow \mu(p)$, we study how $d(r, s)$ depends on $\mu(p)$ in the perturbative regime. By using an equivalent integral formulation [21], one can explicitly derive the perturbative relationship between $\mu(p)$ and $d(r, s)$. However, for the purposes of NN design, a simple observation based on the rotation-equivariance of the experimental setup is sufficient. By introducing a warped coordinate system (s, h) with $(r, s) \equiv (h + s, s)$, the data $d(h, s)$ in the new system can be written as

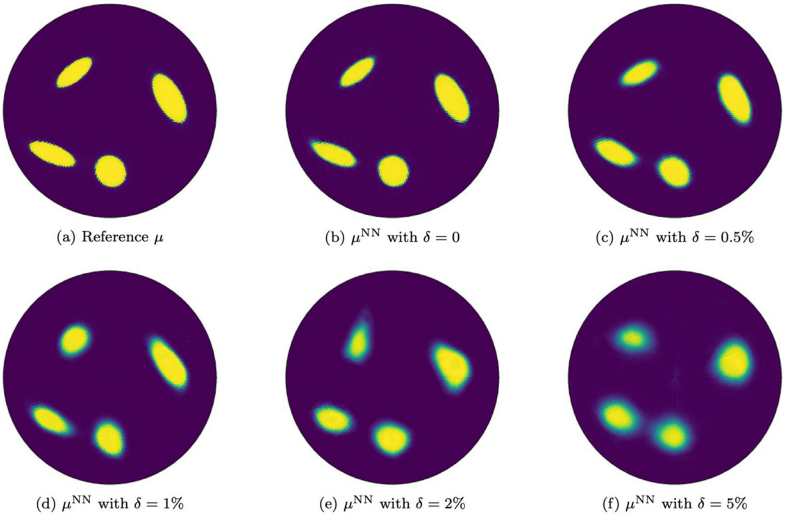
$$d(s, h) \approx \iint_{(\rho,\theta)} k(h, \rho, s - \theta)\mu(\rho, \theta)\mathrm{d}\rho\mathrm{d}\theta, \tag{3.8}$$

which is a 1D convolution in s and θ, with h and ρ treated as parameters (i.e., channels in the NN terminology). Since the RTE (3.6) preserves singularities, especially when the absorption μ_a is weak, this map between η and d is singular in the h and ρ variables. As a result, the number of channels required for the 1D convolution operator can scale with the resolution in ρ and h.

The above discussion shows that, in the small μ regime, we can approximate the forward map $K : \mu(\rho, \theta) \rightarrow d(s, h)$ with a 1D CNN with multiple channels. The filtered backprojection algorithm suggests that $\mu \approx (K^*K + \varepsilon I)^{-1}K^*d$. This again motivates the approach of representing the product $(K^*K + \varepsilon I)^{-1}K^*$ as an NN. As the adjoint to K, the operator K^* can also be approximated with a 1D CNN or BCR-Net with multiple channels. The operator $(K^*K + \varepsilon I)^{-1}$ is a PDO in the (ρ, θ) domain with global support, which can be approximated efficiently with a 2D BCR-Net or CNN. Summarizing these discussions results in the following NN architecture [21] for the OT problem:

$$d(s, h) \Rightarrow \text{1D CNN/BCR-Net} \Rightarrow \text{2D CNN/BCR-Net} \Rightarrow \mu(\rho, \theta). \tag{3.9}$$

Figure 10 presents one numerical example. The resulting NN, with about 50K weights, is trained with a dataset of 8K (d, μ) training pairs. The images show the reference (ground truth) parameter μ, along with the NN reconstructions μ^{NN} at different noise levels. The results suggest that the learned NN representation of the inverse map is quite robust to noise, even though the optical tomography problem is (weakly) ill-posed.


(a) Reference μ (b) μ^{NN} with $\delta=0$ (c) μ^{NN} with $\delta=0.5\%$

(d) μ^{NN} with $\delta=1\%$ (e) μ^{NN} with $\delta=2\%$ (f) μ^{NN} with $\delta=5\%$

FIGURE 10

Optical tomography. Reference solution μ along with the NN reconstructions under different noise levels in the boundary measurement $d(s,h)$.

3.3. Inverse acoustic scattering

Let us consider the acoustic scattering problem in 2D in the frequency domain. The governing equation is the Helmholtz equation

$$(Lu)(p)=\left(-\Delta-\frac{\omega^2}{c(p)^2}\right)u(p)=0, \tag{3.10}$$

where ω is a fixed angular frequency and $c(p)$ is the unknown velocity field. Assume that there exists a known constant background velocity c_0 such that $c(p)-c_0$ is compactly supported in a domain Ω (see Figure 11 (left)). In a typical experimental setup, for each incoming direction $s\in\mathbb{S}^1$, the plane wave $e^{i\omega s\cdot p}$ generates an outgoing scattered field $u^s(p)$ such that $u^s(p)+e^{i\omega s\cdot p}$ is a solution of (3.10). At each unit direction $r\in\mathbb{S}^1$, the far field pattern defined as

$$\hat{u}^s(r)\equiv\lim_{\rho\to\infty}u^s(\rho\cdot r)\sqrt{\rho}e^{-i\omega\rho} \tag{3.11}$$

is recorded. The set of boundary measurements is then $d(r,s)=\hat{u}^s(r)$. Instead of trying to recover $c(p)$ directly, it is often convenient to treat a rescaled index-of-refraction field $\eta(p)\equiv\frac{\omega^2}{c(p)^2}-\frac{\omega^2}{c_0^2}$ as the unknown. The inverse problem is then to recover $\eta(p)$ (equivalently to $c(p)$) from $d(r,s)$.

In order to obtain an approximation for the inverse map $d(r,s)\to\eta(x)$, as usual we consider first how $d(r,s)$ depends on $\eta(x)$ in the perturbative regime. A perturbative analysis for planar incoming waves and far field patterns in [20,38] shows that, when η is small, the data $d(r,s)$ can be approximated up to a smooth amplitude as

$$d(r,s)\approx(K\eta)(r,s)\equiv\int_{p\in\Omega}e^{i\omega(s-r)\cdot p}\eta(p)\mathrm{d}p. \tag{3.12}$$

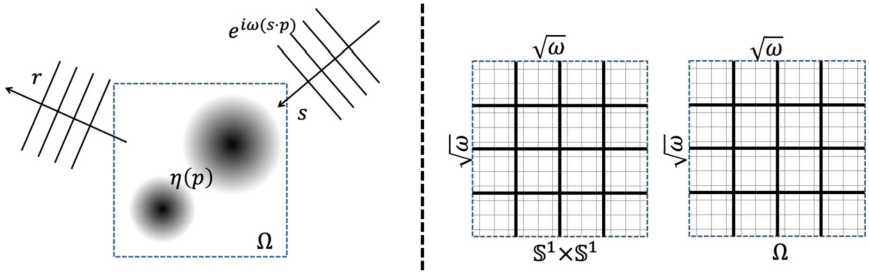


FIGURE 11

Inverse acoustic scattering: (left) an experimental setup in 2D; (right) the complementary low-rank structure of the forward map from $\eta(p)$ to $d(r,s)$.

A rank estimate of the operator kernel [38] shows that K is an FIO from Ω to $\mathbb{S}^1 \times \mathbb{S}^1$ (see Figure 11 (right) for an illustration). As a result, the approximate forward operator from η to d can be represented with a 2D Switch-Net.

The filtered backprojection states that $\eta \approx (K^*K + \varepsilon I)^{-1}K^*d$, thus motivating the approach of representing the product $(K^*K + \varepsilon I)^{-1}K^*$ as an NN. As the adjoint of an FIO is also an FIO [59], the operator K^* can also be approximated with a Switch-Net. The operator $(K^*K + \varepsilon I)^{-1}$ is a PDO in the p variable and can therefore be implemented with a 2D BCR-Net. Concatenating these two modules results in the NN architecture in [38] for the inverse acoustic scattering problem:

$$d(r,s) \Rightarrow \text{2D Switch-Net} \Rightarrow \text{2D BCR-Net} \Rightarrow \eta(p). \tag{3.13}$$

When Ω is a disk, it is natural to parameterize the unknown field η in polar coordinates (ρ,θ). The boundary measurement d is also written in a new coordinate system (h,m) with midpoint $m = \frac{r+s}{2}$ and shift $h = \frac{r-s}{2}$. Under these two new coordinate systems, the rotation-equivariance of the circular geometry implies that the map from $\eta(\rho,\theta)$ to $d(m,h)$ is a 1D convolution in θ and m, with h and ρ treated as the channel dimensions,

$$d(m,h) \approx \int_{\geq 0}\int_0^{2\pi} k(h,\rho,m-\theta)\eta(\rho,\theta)\mathrm{d}\rho\mathrm{d}\theta. \tag{3.14}$$

Following the discussion that leads from (3.8) to (3.9), we can also adopt the following NN architecture [20] for the circular geometry:

$$d(h,m) \Rightarrow \text{1D CNN/BCR-Net} \Rightarrow \text{2D CNN/BCR-Net} \Rightarrow \eta(\rho,\theta). \tag{3.15}$$

Figure 12 gives a numerical example for inverse acoustic scattering. The resulting NN has about 400K weights and is trained with a dataset of 16K (d,η) training pairs. The images show, for two different cases, the reference (ground truth) parameter η, along with the NN reconstructions η^{NN} at different noise levels up to 100%. The results suggest that the learned NN inverse map is highly robust to noise, thanks to the well-posedness of this problem.

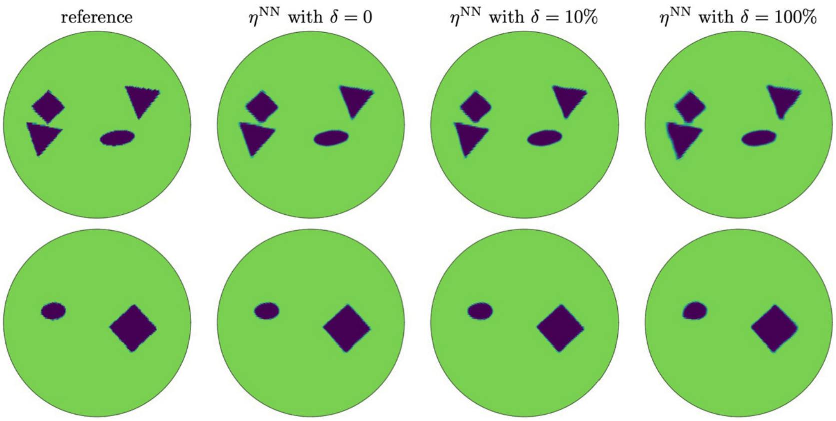


FIGURE 12
Inverse scattering. Each row corresponds to a different test case. For each case, we plot the reference solution, along with the NN reconstructions up to a 100% noise level.

3.4. Seismic imaging

We consider the seismic imaging setting under a simple 2D acoustic model in the frequency domain. The governing equation is again the Helmholtz equation

$$(Lu)(p) = \left(-\Delta - \frac{\omega^2}{c(p)^2}\right)u(p) = f(p), \quad p \in \Omega, \tag{3.16}$$

where ω is a fixed frequency and $c(p)$ is sound speed. We assume that the background velocity $c_0(p)$ is given and the difference between $c(p)$ and $c_0(p)$ is supported in Ω (see Figure 13 (left)). In a typical experimental setup, for each point s on the top surface, one specifies a delta source $f(p) = \delta_s(p)$ and records the wave solution $u^s(\cdot)$ of (3.16) at all points r also on the top surface. The set of boundary measurements is

$$d(r, s) = u^s(r) - u_0^s(r), \tag{3.17}$$

where $u_0^s(\cdot)$ is the solution of some background velocity $c_0(p)$. By again introducing the scaled index-of-refraction field $\eta(p) = \frac{\omega^2}{c(p)^2} - \frac{\omega^2}{c_0(p)^2}$, we obtain the inverse problem of recovering $\eta(p)$ from $d(r, s)$.

As usual, in order to obtain an approximation for the inverse map $d(r, s) \to \eta(p)$, we first study how $d(r, s)$ depends on $\eta(p)$ in the perturbative regime. A perturbative analysis [38] of planar incoming waves and far field patterns shows that, when η is small, the boundary measurement $d(r, s)$ can be well approximated with

$$d(r, s) \approx (K\eta)(r, s) \equiv \int \big(G_0(r, p)G_0(p, s)\big)\eta(p)\mathrm{d}p, \tag{3.18}$$

where $G_0(p)$ is the Green's function of the background operator $L_0 = -\Delta - \omega^2/c_0^2(p)$. A rank estimate of the kernel $G_0(r, p)G_0(p, s)$ in [38] proves that K is an FIO defined between the domain Ω and the product (r, s) space (see Figure 13 (right) for an illustration). As a result, the forward operator from η to d can be approximated with a 2D Switch-Net.

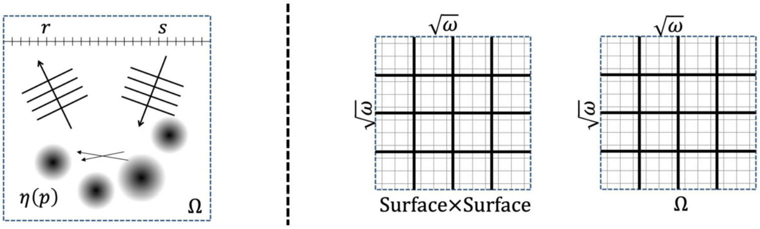


FIGURE 13

Seismic imaging: (left) a simple experimental setup in 2D; (right) the complementary low-rank structure of the forward map from $\eta(p)$ to $d(r,s)$.

The filtered backprojection algorithm again suggests that $\eta \approx (K^*K + \varepsilon I)^{-1}K^*d$, which motivates representing the product $(K^*K + \varepsilon I)^{-1}K^*$ as an NN. As the adjoint of an FIO K, K^* can be approximated with a Switch-Net. Under generic conditions, the operator $(K^*K + \varepsilon I)^{-1}$ is a PDO in the p variable, which can be approximated with a 2D BCR-Net. Putting everything together results in the following NN architecture [38] for the seismic imaging problem:

$$d(r,s) \Rightarrow \text{2D Switch-Net} \Rightarrow \text{2D BCR-Net} \Rightarrow \eta(x,z), \tag{3.19}$$

where x and z are the horizontal and depth coordinates of p, respectively.

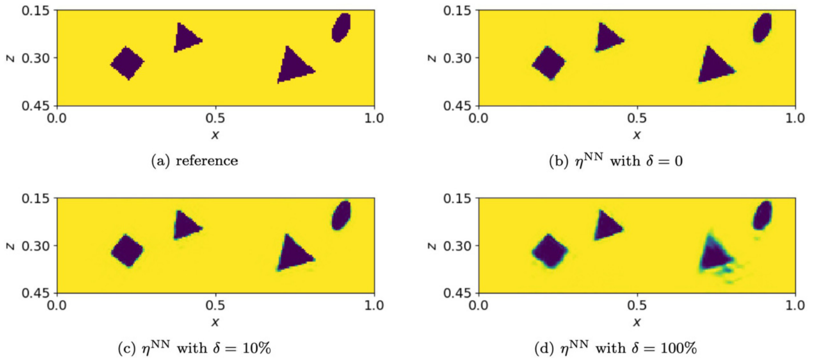
Often in seismic imaging, the background velocity $c_0(p)$ only depends on the depth z (and is independent of the horizontal coordinate x). In this case, we can exploit the translation-equivariance in the horizontal direction and reparameterize the boundary measurement d under the coordinate system (m,h) with $m = \frac{r+s}{2}$ and offset $h = \frac{r-s}{2}$. Under this new coordinate system, the forward map from $\eta(x,z)$ to $d(m,h)$ is a 1D convolution with the offset h and depth z treated as channels,

$$d(m,h) \approx \int_0^Z \int k(h,z,m-x)\eta(x,z)\mathrm{d}x\mathrm{d}z. \tag{3.20}$$

Following the discussion that leads from (3.4) to (3.5), we obtain the following NN architecture [20] for $c_0(p)$ that depends only on depth:

$$d(m,h) \Rightarrow \text{1D CNN/BCR-Net} \Rightarrow \text{2D CNN/BCR-Net} \Rightarrow \eta(x,z). \tag{3.21}$$

Figure 14 shows a numerical example for the seismic inversion problem. The NN has about 1M weights in total and is trained with a dataset of 16K (d,η) pairs. The images show the reference (ground truth) parameter η, along with the NN reconstructions η^{NN} at noise levels up to 100%. The results demonstrate that the learned NN inverse map is quite robust to noise. Notice that the reconstruction quality deteriorates with depth naturally since the boundary measurements are all collected at the top surface.



(a) reference

(b) η^{NN} with $\delta = 0$

(c) η^{NN} with $\delta = 10\%$

(d) η^{NN} with $\delta = 100\%$

FIGURE 14

Seismic imaging. The reference solution along with the NN reconstructions with different levels of noise added to the boundary measurements.

3.5. Traveltime tomography

Here we consider a circular domain Ω in 2D with the boundary denoted by Γ (see Figure 15). The governing equation for traveltime tomography (TT) is the eikonal equation

$$|\nabla u(p)| = \frac{1}{c(p)}, \quad p \in \Omega, \tag{3.22}$$

where $c(p)$ is the unknown velocity field. Assuming that $c(p)$ has a background velocity c_0 (taken to be 1 without loss of generality), we introduce the slowness deviation $\eta(p) \equiv \frac{1}{c(p)} - 1$ and write (3.22) as $|\nabla u(p)| = 1 + \eta(p)$. In a typical setup, we specify the zero boundary condition at each boundary point s, solve for the viscosity solution $u^s(x)$ of (3.22), and record $u^s(r)$ at each boundary point r. The set of boundary measurements is then given by

$$d(r,s) = u^s(r) - u^s_0(r), \tag{3.23}$$

where $u^s_0(r) = \|r - s\|$ is the solution for $\eta(x) \equiv 0$. The inverse problem is to recover $\eta(p) \equiv \frac{1}{c(p)} - 1$ from $d(r,s)$.

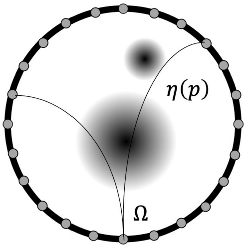


FIGURE 15

Traveltime tomography. Experimental setup in a circular geometry.

To obtain an approximation for the inverse map $d(r, s) \to \eta(p)$, we study how $d(r, s)$ depends on $\eta(p)$ in the perturbative regime. A simple consideration based on the rotation-equivalence of the experimental setup suggests viewing the parameter η in polar coordinates (ρ, θ) and the boundary measurements in the warped coordinates (s, h) with $(r, s) \equiv (h + s, s)$. The data $d(h, s)$ can then be written as [22]

$$d(s, h) \approx (K\eta)(h, s) = \int_{\geq 0} \int_0^{2\pi} k(h, \rho, s - \theta)\eta(\rho, \theta)\mathrm{d}\rho\mathrm{d}\theta, \tag{3.24}$$

which is a 1D convolution in s and θ, with h and ρ treated as parameters (or channels in the NN terminology). Since the viscosity solution of the eikonal equation often has singularities, the number of channels required for the 1D convolution operator can be quite significant.

The discussion above shows that, in the small η regime, we can approximate the forward map $K : \eta(\rho, \theta) \to d(s, h)$ with a 1D CNN or BCR-Net with multiple channels. The filtered backprojection algorithm $\eta \approx (K^* K + \varepsilon I)^{-1} K^* d$ suggests representing the product $(K^* K + \varepsilon I)^{-1} K^*$ as a linear NN and then generalizing to the nonlinear regime. By invoking the same argument used for K, the adjoint operator K^* can also be approximated with a 1D CNN or BCR-Net with a small number of channels. The operator $(K^* K + \varepsilon I)^{-1}$ is a PDO in the (ρ, θ) domain with global support, which can be approximated with a 2D BCR-Net or with a 2D CNN. Summarizing the discussion results in the following NN architecture [22] for traveltime tomography:

$$d(s, h) \Rightarrow \text{1D CNN/BCR-Net} \Rightarrow \text{2D BCR-Net/CNN} \Rightarrow \eta(\rho, \theta). \tag{3.25}$$

Figure 16 gives a numerical example for the traveltime tomography. The NN for the inverse map, with about 640K weights, is trained with a set of 16K (d, η) pairs. The three rows correspond to test examples with negative inclusion $c(p) < 1$, positive inclusion $c(p) > 1$, and mixed inclusion, respectively. For each test example, we plot the reference solution along with the NN reconstructions with different levels of noise added to the boundary measurements. The results show that, even for this ill-posed problem, the NN inverse map is accurate and robust with respect to noise.

4. CONCLUDING REMARKS

In this paper, we discussed our recent work on applying deep learning to inverse problems. On the algorithmic side, we proposed two new NN modules, BCR-Net and Switch-Net. They are motivated by the pseudodifferential and Fourier integral operators, which play key roles in the study of inverse problems. On the application side, we propose NNs that approximate the inverse maps in five settings of interest: electrical impedance tomography, optical tomography, inverse acoustic scattering, seismic imaging, and traveltime tomography. In each application, the architecture is motivated by the perturbation theory and filtered backprojection and is implemented using the new modules along with standard convolution layers. In several cases, we have heavily relied on the special geometry of the domain Ω and the data collection process. When combined with appropriate reparameterizations, this

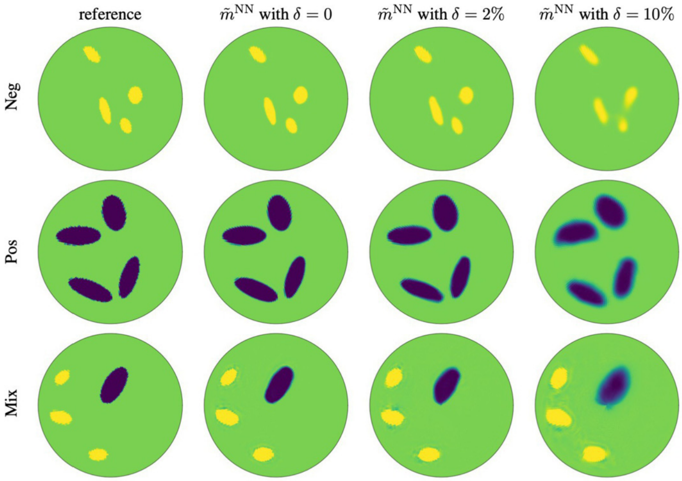


FIGURE 16

Traveltime tomography. The three rows correspond to negative, positive, and mixed inclusions. For each case, the reference solution is shown along with the NN reconstructions with different levels of noise added to the boundary measurements.

often results in NN architectures that are both general and effective. Our approach provides a seamless way that combines the mathematical structure of the inverse problems, the power of deep NNs, and the information in the data prior. Below we list some directions for future research:

- We have considered only the case of complete measurement data. A question of both practical and theoretical importance is how to extend to the case of partial measurement data.
- For wave-based inverse problems, we have focused on a single frequency or a single energy. In many applications, one often has access to boundary measurements at multiple frequencies or energies, or even time-dependent measurements.
- The first part of the proposed NNs is closely related to the migration step in traditional imaging pipelines (such as in seismic imaging). An interesting study would be to compare the intermediate result after the first part of our NN with the migration results to get a more precise understanding of the proposed NNs.

The study of inverse problem using deep learning has grown into a relatively large subject [6, 39, 49, 56, 62]. This paper has solely focused on one approach that is deeply rooted in microlocal analysis (see [3, 10] for related work). There are a few other highly active research directions that we have not discussed here, but which may be of interest to the reader:

- We have not discussed the work on (linear) underdetermined inverse problem in imaging [32,51]. This field is closely connected with sparse recovery problems, such as compressive sensing, matrix completion, phase retrieval, etc. [11].
- In the unrolling or unfolding approach [2,13,15,25,33,50,53,63] for solving inverse problems, one writes the iterative solution algorithm as a ResNet and then trains the network parameters to minimize the reconstruction error. In many cases, this approach leads to a very high quality reconstruction.
- There is also active work studying stability issues when applying deep learning to inverse problems [4,14,24,27], which is particularly important for applications with ill-posed inverse problems.

ACKNOWLEDGMENTS

The author would like to thank Bin Dong, Zach Izzo, and Ozan Öktem for their help in preparing this manuscript.

FUNDING

This work was partially supported by the National Science Foundation under awards DMS-1818449 and DMS-2011699 and by the U. S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program.

REFERENCES

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pp. 265–283, 2016.

[2] J. Adler and O. Öktem, Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Probl.* **33** (2017), no. 12, 124007.

[3] H. Andrade-Loarca, G. Kutyniok, O. Öktem, and P. Petersen, Deep microlocal reconstruction for limited-angle tomography. 2021, arXiv:2108.05732.

[4] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc. Natl. Acad. Sci.* **117** (2020), no. 48, 30088–30095.

[5] M. Araya-Polo, J. Jennings, A. Adler, and T. Dahlke, Deep-learning tomography. *Lead. Edge* **37** (2018), no. 1, 58–66.

[6] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, Solving inverse problems using data-driven models. *Acta Numer.* **28** (2019), 1–174.

[7] L. Bar and N. Sochen, Unsupervised deep learning algorithm for PDE-based forward and inverse problems. 2019, arXiv:1904.05417.

[8] J. Berg and K. Nyström, A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing* **317** (2018), 28–41.

[9] G. Beylkin, R. Coifman, and V. Rokhlin, Fast wavelet transforms and numerical algorithms I. *Comm. Pure Appl. Math.* **44** (1991), no. 2, 141–183.

[10] T. A. Bubba, M. Galinier, M. Lassas, M. Prato, L. Ratti, and S. Siltanen, Deep neural networks for inverse problems with pseudodifferential operators: An application to limited-angle tomography. *SIAM J. Imaging Sci.* **14** (2021), no. 2, 470–505.

[11] E. J. Candès, Mathematics of sparsity (and a few other things). In *Proceedings of the International Congress of Mathematicians, Seoul, South Korea* 123, Citeseer, 2014.

[12] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks. *Science* **355** (2017), no. 6325, 602–606.

[13] J. Cheng, H. Wang, Y. Zhu, Q. Liu, Q. Zhang, T. Su, J. Chen, Y. Ge, Z. Hu, X. Liu, et al., Model-based deep medical imaging: the roadmap of generalizing iterative reconstruction model using deep learning. 2019, arXiv:1906.08143.

[14] M. J. Colbrook, V. Antun, and A. C. Hansen, Can stable and accurate neural networks be computed? – On the barriers of deep learning and Smale's 18th problem. 2021, arXiv:2101.08286.

[15] M. V. de Hoop, M. Lassas, and C. A. Wong, Deep learning architectures for nonlinear operator functions and nonlinear inverse problems. 2019, arXiv:1912.11090.

[16] W. E and B. Yu, The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat.* **6** (2018), no. 1, 1–12.

[17] Y. Fan, C. O. Bohorquez, and L. Ying, BCR-Net: a neural network based on the nonstandard wavelet form. *J. Comput. Phys.* **384** (2019), 1–15.

[18] Y. Fan, J. Feliu-Fabà, L. Lin, L. Ying, and L. Zepeda-Núñez, A multiscale neural network based on hierarchical nested bases. *Res. Math. Sci.* **6** (2019), no. 2, 21.

[19] Y. Fan, L. Lin, L. Ying, and L. Zepeda-Núnez, A multiscale neural network based on hierarchical matrices. *Multiscale Model. Simul.* **17** (2019), no. 4, 1189–1213.

[20] Y. Fan and L. Ying, Solving inverse wave scattering with deep learning. 2019, arXiv:1911.13202.

[21] Y. Fan and L. Ying, Solving optical tomography with deep learning. 2019, arXiv:1910.04756.

[22] Y. Fan and L. Ying, Solving traveltime tomography with deep learning. 2019, arXiv:1911.11636.

[23] Y. Fan and L. Ying, Solving electrical impedance tomography with deep learning. *J. Comput. Phys.* **404** (2020), 109119.

[24] M. Genzel, J. Macdonald, and M. März, Solving inverse problems with deep neural networks–robustness included? 2020, arXiv:2011.04268.

[25] D. Gilton, G. Ongie, and R. Willett, Neumann networks for linear inverse problems in imaging. *IEEE Trans. Comput. Imaging* **6** (2019), 328–343.

[26] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning* 1, MIT Press, Cambridge, 2016.

[27] N. M. Gottschling, V. Antun, B. Adcock, and A. C. Hansen, The troublesome kernel: why deep learning for inverse problems is typically unstable. 2020, arXiv:2001.01258.

[28] L. Greengard and V. Rokhlin, A fast algorithm for particle simulations. *J. Comput. Phys.* **73** (1987), no. 2, 325–348.

[29] W. Hackbusch, L. Grasedyck, and S. Börm, An introduction to hierarchical matrices. *Math. Bohem.* **127** (2002), no. 229–241.

[30] J. Han and A. Jentzen W. E, Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci.* **115** (2018), no. 34, 8505–8510.

[31] J. Han, L. Zhang, R. Car, and W. E, Deep potential: a general representation of a many-body potential energy surface. *Commun. Comput. Phys.* **23** (2018), no. 3, 629–639.

[32] P. Hand and V. Voroninski, Global guarantees for enforcing deep generative priors by empirical risk. In *Conference on learning theory*, pp. 970–978, PMLR, 2018.

[33] A. Hauptmann, J. Adler, S. Arridge, and O. Öktem, Multi-scale learned iterative reconstruction. *IEEE Trans. Comput. Imaging* **6** (2020), 843–856.

[34] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **29** (2012), no. 6, 82–97.

[35] H. Kabir, Y. Wang, M. Yu, and Q.-J. Zhang, Neural network inverse modeling and applications to microwave filter design. *IEEE Trans. Microwave Theory Tech.* **56** (2008), no. 4, 867–879.

[36] Y. Khoo, J. Lu, and L. Ying, Solving for high-dimensional committor functions using artificial neural networks. *Res. Math. Sci.* **6** (2019), no. 1, 1.

[37] Y. Khoo, J. Lu, and L. Ying, Solving parametric PDE problems with artificial neural networks. *European J. Appl. Math.* **32** (2021), no. 3, 421–435.

[38] Y. Khoo and L. Ying, SwitchNet: a neural network model for forward and inverse scattering problems. *SIAM J. Sci. Comput.* **41** (2019), no. 5, A3182–A3201.

[39] Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft, Machine learning in seismology: Turning data into insights. *Seismol. Res. Lett.* **90** (2019), no. 1, 3–14.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012), 1097–1105.

[41] G. Kutyniok, P. Petersen, M. Raslan, and R. Schneider, A theoretical analysis of deep neural networks and parametric PDEs. *Constr. Approx.* (2021), 1–53.

[42] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning. *Nature* **521** (2015), no. 436.

[43] M. K. K. Leung, H. Y. Xiong, L. J. Lee, and B. J. Frey, Deep learning of the tissue-regulated splicing code. *Bioinformatics* **30** (2014), no. 12, i121–i129.

[44] Y. Li, J. Lu, and A. Mao, Variational training of neural network approximations of solution maps for physical models. *J. Comput. Phys.* **409** (2020), 109338.

[45] Y. Li, H. Yang, E. R. Martin, K. L. Ho, and L. Ying, Butterfly factorization. *Multiscale Model. Simul.* **13** (2015), no. 2, 714–732.

[46] Z. Long, Y. Lu, X. Ma, and B. Dong PDE-net, Learning PDEs from data. In *International Conference on Machine Learning*, pp. 3208–3216, PMLR, 2018.

[47] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Process. Mag.* **35** (2018), no. 1, 20–36.

[48] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **55** (2015), no. 2, 263–274.

[49] M. T. McCann and M. Unser, Biomedical image reconstruction: from the foundations to deep neural networks. 2019, arXiv:1901.03565.

[50] V. Monga, Y. Li, and Y. C. Eldar, Algorithm unrolling: interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.* **38** (2021), no. 2, 18–44.

[51] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, Deep learning techniques for inverse problems in imaging. *IEEE J. Sel. Areas Inf. Theory* **1** (2020), no. 1, 39–56.

[52] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., PyTorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32** (2019), 8026–8037.

[53] P. Putzky and M. Welling, Invert to learn to invert. *Adv. Neural Inf. Process. Syst.* **32** (2019), 446–456.

[54] M. Raissi and G. E. Karniadakis, Hidden physics models: machine learning of nonlinear partial differential equations. *J. Comput. Phys.* **357** (2018), 125–141.

[55] M. Raissi, P. Perdikaris, and G. E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378** (2019), 686–707.

[56] S. Ravishankar, J. C. Ye, and J. A. Fessler, Image reconstruction: from sparsity to data-adaptive methods and machine learning. *Proc. IEEE* **108** (2019), no. 1, 86–109.

[57] K. Rudd and S. Ferrari, A constrained integration (CINT) approach to solving partial differential equations using artificial neural networks. *Neurocomputing* **155** (2015), 277–285.

[58] J. Schmidhuber, Deep learning in neural networks: an overview. *Neural Netw.* **61** (2015), 85–117.

[59] E. M. Stein and T. S. Murphy, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals* 3, Princeton University Press, 1993.

[60] I. Sutskever, O. Vinyals, and Q. V. Le, Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process. Syst.* **27** (2014), 3104–3112.

[61] C. Tan, S. Lv, F. Dong, and M. Takei, Image reconstruction based on convolutional neural network for electrical resistance tomography. *IEEE Sens. J.* **19** (2018), no. 1, 196–204.

[62] P. R. Wiecha, A. Arbouet, C. Girard, and O. L. Muskens, Deep learning in nanophotonics: inverse design and beyond. *Photon. Res.* **9** (2021), no. 5, B182–B200.

[63] H.-M. Zhang and B. Dong, A review on deep learning in medical image reconstruction. *J. Oper. Res. Soc. China* (2020), 1–30.

LEXING YING

Department of Mathematics, Stanford University, Stanford, CA 94305-2125, USA,
lexing@stanford.edu