

Learn an Index Operator by CNN for Solving Diffusive Optical Tomography: A Deep Direct Sampling Method

Guo, Ruchi; Jiang, Jiahua; Li, Yi

DOI:

[10.1007/s10915-023-02115-7](https://doi.org/10.1007/s10915-023-02115-7)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Guo, R, Jiang, J & Li, Y 2023, 'Learn an Index Operator by CNN for Solving Diffusive Optical Tomography: A Deep Direct Sampling Method', *Journal of Scientific Computing*, vol. 95, no. 1, 31.
<https://doi.org/10.1007/s10915-023-02115-7>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.



Learn an Index Operator by CNN for Solving Diffusive Optical Tomography: A Deep Direct Sampling Method

Ruchi Guo¹ · Jiahua Jiang^{2,3} · Yi Li⁴

Received: 29 August 2021 / Revised: 14 November 2022 / Accepted: 7 December 2022 /
Published online: 8 March 2023
© The Author(s) 2023

Abstract

In this work, we investigate the diffusive optical tomography (DOT) problem in the case that limited boundary measurements are available. Motivated by the direct sampling method (DSM) proposed in Chow et al. (SIAM J Sci Comput 37(4):A1658–A1684, 2015), we develop a deep direct sampling method (DDSM) to recover the inhomogeneous inclusions buried in a homogeneous background. In this method, we design a convolutional neural network to approximate the index functional that mimics the underlying mathematical structure. The benefits of the proposed DDSM include fast and easy implementation, capability of incorporating multiple measurements to attain high-quality reconstruction, and advanced robustness against the noise. Numerical experiments show that the reconstruction accuracy is improved without degrading the efficiency, demonstrating its potential for solving the real-world DOT problems.

Keywords Deep learning · Inverse problems · Direct sampling methods · Diffusive optical tomography · Reconstruction algorithm

1 Introduction

Diffusive optical tomography (DOT) is a promising imaging technique with many clinical applications, such as screening for breast cancer and the development of cerebral images [18, 19, 22]. As its mechanism, the input flux of near-infrared (NIR) photons is used to

✉ Jiahua Jiang
j.jiang.3@bham.ac.uk

Ruchi Guo
ruchig@uci.edu

Yi Li
liyi3@shanghaitech.edu.cn

¹ Department of Mathematics, University of California, Irvine, CA, USA

² School of Mathematics, University of Birmingham, Birmingham, UK

³ Institute of Mathematical Science, ShanghaiTech University, Pudong, China

⁴ School of Information Science and Technology, ShanghaiTech University, Pudong, China

illuminate the body and the output flux is measured on the surface of the body. In this process, chromophores in the NIR window such as oxygenated and deoxygenated hemoglobin, water, and lipid, are abundant in the body tissue, and a weighted sum of their contributions gives different absorption coefficients [13]. Then the measured pairs of the fluxes provide some information for detecting and reconstructing the optical properties inside the body by creating images of the distribution of absorption coefficients inside the body. Let the absorption medium occupy an open bounded connected domain $\Omega \subseteq \mathbb{R}^2$ with a piecewise C^2 boundary and D be a subdomain of Ω representing the inhomogeneous inclusions. The absorption coefficient of the media in Ω is described by a non-negative function $\mu \in L^\infty(\Omega)$, while μ_0 is the absorption coefficient of the homogeneous background medium and the support of $\mu - \mu_0$ occupies the subdomain D . We consider the DOT model that the potential $u \in H^1(\Omega)$ representing the photon density field satisfies the following equation:

$$-\Delta u_\omega + \mu u_\omega = 0 \quad \text{in } \Omega, \quad (1.1)$$

$$\frac{\partial u_\omega}{\partial \mathbf{n}} = g_\omega \quad \text{on } \partial\Omega, \quad \omega = 1, 2, \dots, N, \quad (1.2)$$

where $g_\omega \in H^{-1/2}(\partial\Omega)$ denotes the surface flux along $\partial\Omega$, and $f_\omega = u_\omega|_{\partial\Omega}$ is the measurement of the surface potential on the boundary. In this paper, we consider the Neumann boundary value problem for simplicity, and the proposed algorithm can be also applied to the Robin boundary value problem. The overarching goal of the DOT problem is to recover the geometry of the inhomogeneous inclusions D from the N pairs of potential-flux data $(g_\omega, f_\omega)_{\omega=1}^N$, referred as Cauchy data pairs. It is noted that the considered model can be viewed as a special case of the more general one [7] which has more complicated jump conditions.

The inverse procedure in the DOT problem can be described by the well-known Neumann-to-Dirichlet (NtD) mapping

$$\Lambda_\mu : H^{-1/2}(\partial\Omega) \rightarrow H^{1/2}(\partial\Omega), \quad g \mapsto u|_{\partial\Omega}, \quad (1.3)$$

where u is the trace of the solution of (1.1) according to the boundary condition $g \in H^{-1/2}(\partial\Omega)$. It is known that the NtD mapping preserves all the information needed to recover μ [7]; see also the discussion in Sect. 3. However, approximating the whole NtD mapping requires a large number of Cauchy data pairs, which hinders the application in many practical situations. To circumvent the application limitation, in this work, we focus on developing the reconstruction algorithm that only requires a reasonably small number of Cauchy data pairs.

Over the past decades, much effort has been rewarded with many promising developments of solving the DOT problem. One type of the widely used algorithms are iterative methods. The augmented Lagrange method [1] and the shape optimization methods [42, 43] reconstruct the inhomogeneous inclusions by minimizing a functional measuring the observed and simulated data. To alleviate the onerous computational costs caused by the necessity of using a fine mesh, the multigrid methods [33, 38] have been proposed to resolve this issue. Another standard approach formulates an integral function involving the Green's function corresponding to background absorption coefficient, and use the Born type iteration [37, 39] to solve this integral equation until convergence. Nonetheless, the major issue of these approaches is that a large number of the forward PDEs need to be solved in the iterative process, which is time-consuming and computationally infeasible in many practical situations such as three-dimensional problems. Thus, for ultrasound-modulated DOT, the works in [2–4] provide the hybrid method that is capable of reconstructing the parameters in known inclusions with only a few iterations.

Another popular group of methods for solving the DOT problem are non-iterative methods. The joint sparsity methods [31, 32] have been investigated to solve the aforementioned integration equation by a non-iterative approach in which an ill-conditioned matrix needs to be inverted and some suitable preconditioners or regularization are demanded. The well-known factorization methods described in [7, 12, 28] compute the spectral information of the boundary data mapping and determine the inclusion shape by checking the convergence property of a series involving the spectral data. Based on a nonlinear Fourier transform, the D-bar methods in [35, 36] reconstruct the absorption coefficient by solving a boundary integral equation. Nevertheless, the severe ill-posedness of the original inverse problem will manifest itself in the computation of the boundary integral equations, which brings strenuousness in computation. This is improved in [29, 30] by linearizing the reconstruction problem and reducing it to a well-posed boundary-value problem for a coupled system of elliptic equations such that it can be efficiently solved. Besides these classical approaches, vast development of deep neural networks-based algorithms has been undergone for solving the DOT problem, for instance, using CNN to learn the nonlinear photon scattering physics and reconstruct the 3D optical anomalies [41] and FDU-Net that is able to fast recover the irregular-shaped inclusions with accurate localization for breast DOT [20], and some detailed surveys can be found in the recent monologue [6, 40].

Recently, direct sampling methods (DSM) arise as very appealing non-iterative strategies to solve many geometric inverse problems.

The critical component of DSM is to construct a certain index function indicating the shape and location of the inclusions through a duality product operating on the boundary data and some probing functions. The DSM has been proven to be highly robust, effective and efficient for many inverse problems, such as electrical impedance tomography (EIT) [16], inverse scattering [25, 26], moving potential reconstruction [17] and the inversion of Radon transform for computed tomography [14]. For the DOT problem, the original DSM [15] designs an elegant index function format for the case of a single measurement pair. However, only a single measurement results in insufficient accuracy of the reconstruction in more complicated cases, which stymies its application in practical scenarios of medical imaging. Indeed, the closed form of explicit index function for multiple Cauchy data pairs or complex-shaped domain may become very intricate and deriving it with conventional mathematical approaches will be very challenging.

On the one hand, deep learning recently has become an alternative approach to canonical mathematical derivation. On the other hand, it is worth mentioning that the index function of DSM can be regarded as a mapping from a specially generated data manifold to the inclusion distribution. The data manifold consists of data functions built up by solving the forward problem with background coefficient, and this underlying mathematical structure serves as our crucial guideline to design the neural networks (NN) such that the complicated nonlinear mapping can be learned by data. In our previous work for EIT [23], we have found and shown that the corresponding data-driven index function can successfully capture the essential structure of the true index function, smooth out the noise and outperform the conventional DSM.

To address this barrier from the original DSM, in this work we develop a novel deep direct sampling method (DDSM) for solving the DOT problem based a convolutional network (CNN). The main benefits of the proposed DDSM contain:

- It is easy to implement for 2D and particularly 3D cases which challenge many conventional approaches.

- The index function in the DDSM is able to incorporate multiple Cauchy data pairs which enhances the reconstruction quality and robustness with respect to noise.
- The DDSM inherits the high efficiency of DSM [15], which is benefitted from its offline-online decomposition structure. For given measurements, the reconstruction is based on the fast evaluation of the CNN-based index function in the online stage, which has almost the same speed as the original DSM.
- Similar to DSM, constructing data functions by solving elliptic equations smooths out the noise on the boundary such that the resulting NN is highly robust against the noise.
- It is capable of incorporating very limited boundary data points to achieve satisfactory reconstruction.
- It can yield adequate reconstruction even if the absorbing coefficients of materials to be recovered are significantly different from those used for training, which is of much practical interest as only very rough guess is needed.

The rest of the paper is organized as follows. In Sect. 2, we briefly review the original DSM. The development of our proposed DDSM and its theoretical justification are introduced in Sect. 3. The numerical experiments to validate the advantages of the DDSM are provided in Sect. 4. Finally, Sect. 5 summarizes the research findings.

2 Direct Sampling Methods

In this section, we briefly review the conventional DSM proposed by Chow et al. in [15] for DOT that will serve as the framework guiding us to design suitable neural networks. The main idea is to approximate an index function satisfying

$$\mathcal{I}(x) = \begin{cases} 1 & \text{if } x \in D, \\ 0 & \text{if } x \in \Omega \setminus \overline{D}, \end{cases} \tag{2.1}$$

which depends on the given Cauchy data. Since the key structure of the index function in [15] assumes only a *single* pair of Cauchy data, we follow this assumption in this section.

We begin with introducing a family of probing functions $\eta_x(\xi)$ with $x \in \Omega$ on $\partial\Omega$ which are the fundamental ingredients for both the theory of uniqueness of DOT [7] and the DSM [15]. Consider the following diffusion equation with homogeneous background medium absorption coefficient μ_0 :

$$-\Delta w_x + \mu_0 w_x = \delta_x \quad \text{in } \Omega; \quad w_x = 0 \quad \text{on } \partial\Omega, \tag{2.2}$$

where $\delta_x(\xi)$ is the delta function associated with $x \in \Omega$. For a fixed point $x \in \Omega$, the probing function η_x is defined as the surface flux of w_x over $\partial\Omega$:

$$\eta_x(\xi) := \frac{\partial w_x(\xi)}{\partial \mathbf{n}}, \quad \xi \in \partial\Omega. \tag{2.3}$$

An essential component of DSM is the dual product,

$$\langle \phi, \psi \rangle_{\partial\Omega, s} := \int_{\partial\Omega} (-\Delta_{\partial\Omega})^s \phi \psi \, ds \quad \text{on } H^{2s}(\partial\Omega) \times L^2(\partial\Omega), \tag{2.4}$$

where $(-\Delta_{\partial\Omega})^s$ is a certain fractional order of the surface Laplacian operator defined on $\partial\Omega$. The index function can be defined as

$$\mathcal{I}(x) = c_0(x) \frac{\langle \eta_x, f - \Lambda_{\mu_0} g \rangle_{\partial\Omega, s}}{|\eta_x|_Y} \tag{2.5}$$

where $c_0(x)$ is a constant for normalization specified later, and $|\cdot|_Y$ is an algebraic function of seminorms in $H^{2s}(\partial\Omega)$ (itself may not be a norm or seminorm), and a typical choice in [15] is $|\cdot|_Y = |\cdot|_{H^1(\partial\Omega)}^{1/2} |\cdot|_{L^2(\partial\Omega)}^{3/4}$. In the DSM for various geometric inverse problems [15, 16, 26], this duality product structure plays an important role in index functions since it can effectively remove the errors contained in the data $f - \Lambda_{\mu_0}g$ by the high smoothness of η_x near $\partial\Omega$.

Note that the format of the index function in (2.4) may not be computationally effective since probing functions η_x change with respect to x . Therefore, an alternative characterization of the index function is derived in [15] which is the foundation of our neural network. Let φ denote the solution of the following equation also with only the background absorption coefficient

$$-\Delta\varphi + \mu_0\varphi = 0 \quad \text{in } \Omega; \quad \varphi = -(-\Delta\partial\Omega)^s(f - \Lambda_{\mu_0}g) \quad \text{on } \partial\Omega. \tag{2.6}$$

Then the index function in (2.5) can be effectively computed through φ . To show the relationship between the duality product and φ , we provide a brief derivation for $s = 0$. Based on the equation (2.2), the definition of duality product in (2.3) and Green’s formula, we have

$$\begin{aligned} \langle \eta_x, f - \Lambda_{\mu_0}g \rangle_{\partial\Omega, 0} &= \int_{\partial\Omega} \eta_x(f - \Lambda_{\mu_0}g)ds = - \int_{\partial\Omega} \frac{\partial w_x}{\partial \mathbf{n}} \varphi ds \\ &= \int_{\partial\Omega} w_x \frac{\partial \varphi}{\partial \mathbf{n}} - \frac{\partial w_x}{\partial \mathbf{n}} \varphi ds = \int_{\Omega} w_x \Delta\varphi - \varphi \Delta w_x dx \tag{2.7} \\ &= \int_{\Omega} w_x(\mu_0\varphi) - \varphi \Delta w_x dx = \varphi(x). \end{aligned}$$

The same result can be obtained for $s = 1$ in [15]. With (2.5) and (2.7), the index function can be equivalently rewritten as

$$\mathcal{I}(x) = c_0(x) \frac{\varphi(x)}{|\eta_x|_Y}, \tag{2.8}$$

where the typical choice for c_0 is $(\|\varphi\|_{L^\infty(\Omega)} + \varphi(x))^{-1}$ [15].

It is highlighted that the index function depends on the Cauchy data only through the function φ while the value of $|\eta_x|_Y$ is only based on the geometry of Ω . This mathematical structure guides us to set φ as the input of the networks designed to approximate the index function, which will be detailed in the next section. Given the importance of φ , we shall call it the Cauchy difference function in our following discussion. Note that φ is readily solvable and only needs to be done once from (2.6) for a single pair of Cauchy data (f, g) . Since there is only one PDE to solve in the reconstruction procedure, the cost is much lower than the optimization-based methods. Our proposed DDSM inherits this advantage, namely only N PDEs are required to be solved for N Cauchy data pairs.

Besides φ , the index function in (2.8) is determined by the probing functions η_x computed from (2.2) and (2.3). However, the evaluation of η_x requires solving (2.2) for every x that needs to be located, which maybe inefficient. To address this issue, the explicit formulas are provided in [15] for some specific domains; for instance, the probing function corresponding to a unit circle is

$$\eta_x(\xi) = \frac{1}{2\pi} \sum_{n \in \mathbb{Z}} \frac{J_n(i\sqrt{\mu_0}r_x)}{J_n(i\sqrt{\mu_0})} e^{in(\theta_x - \theta_\xi)}, \quad \xi \in \mathbb{S}^1, \tag{2.9}$$

where J_n are the Bessel functions, and (r_x, θ_x) is the polar coordinate for a point x . However, it has not escaped our notice that such explicit formulas may not be available for all type

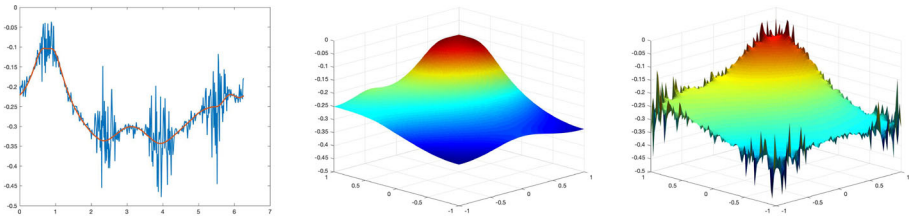


Fig. 1 The boundary data with independent and identically Gaussian noise (left), the generated $\varphi(x)$ without noise (middle), $\varphi(x)$ with noise (right)

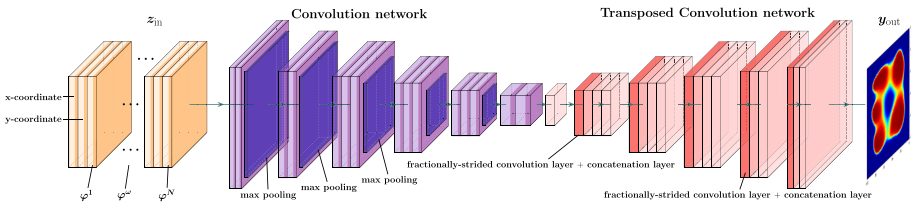


Fig. 2 The structure of CNN for 2D DOT problem. The input is a 3D matrix in $\mathbb{R}^{n_x \times n_y \times (N+2)}$ and the output is a matrix in $\mathbb{R}^{n_1 \times n_2}$

of geometries. Different from the analytical typed index function (2.5) in the conventional DSM, the index function in the proposed DDSM is represented by a neural network and learned from data, which is not in a closed form but capable of handling more complicated and practical problems.

3 Deep Direct Sampling Methods

Despite the successful application of the original DSM, we believe some aspects can be further improved with the recently developed DNN techniques, for which conventional mathematical derivation may face some difficulties.

First, the index function format above may only involve a *single* Cauchy data pair, which may limit its accuracy. Moreover, our numerical results show that including one pair of Cauchy data is robust for the spatially-invariant noise used in [15]. But when the noise become highly spatially variant for boundary data points, for example independent and identically Gaussian distribution illustrated in the left plot of Fig. 1, the reconstruction may not be robust as shown by the numerical results in Fig. 5, which brings out the importance and necessity of including multiple measurements. However, it has been unclear so far how to theoretically develop an explicit index function with a closed form that systematically incorporate multiple Cauchy data pairs through canonical mathematical derivation, though some basic strategies can be applied such as average, maximum or product of each individual index function. Second, even for the case of a single measurement, the format of the index function $\mathcal{I}(x)$ may not be the optimal approximation to the true index function, for example the empirical choice of the tuning parameter s and norm $|\cdot|_p$. We believe this is where the DNN models can exploit their advantages to replace some theoretical derivation by data driven approaches such that the more optimal index function can be obtained.

Table 1 Relationship between DSM for DOT problems and image segmentation problems

	DSM for DOT	Image segmentation
x	mesh point in Ω	pixel
$\phi^\omega(x)$	Cauchy difference functions	image features (channels)
\mathcal{I}	index of inclusion distribution	dense prediction

Therefore, to enhance the performance of DSMs based the aforementioned aspects, in this work, we propose a *Deep Direct Sampling Method (DDSM)* by mimicking the underlying mathematical structure suggested by the DSM. For simplicity, we mainly discuss the two-dimensional case and the method can be readily and naturally extended to the three-dimensional case. We have implemented the method for both the 2D and 3D DOT problems where the numerical examples and reconstruction results will be provided in Sect. 4.

3.1 Neural Network Structure

We note that the index function (2.8) suggests the existence of a non-linear mapping or operator from φ to the location of x , namely whether it is inside or outside the inclusions. We shall see in Sect. 3.2 that this operator theoretically exists if the NtD mapping is given, i.e., all the Cauchy data pairs are available, but it is not easy to derive an explicit index function to fully incorporate the information with limited Cauchy data pairs.

Motivated by the classical DSM, we define *Cauchy difference functions* $\{\varphi^\omega\}_{\omega=1}^N$, where $\varphi^\omega (1 \leq \omega \leq N)$ is the solution of (2.6) with the boundary value formed by the ω -th pair of Cauchy data $f_\omega - \Lambda_{\sigma_0} g_\omega$, namely

$$-\Delta \varphi^\omega + \mu_0 \varphi^\omega = 0 \quad \text{in } \Omega; \quad \varphi^\omega = -(-\Delta_{\partial\Omega})^s (f_\omega - \Delta_{\partial\Omega} g_\omega) \quad \text{on } \partial\Omega. \tag{3.1}$$

Then, the input of the DNN is designed as $(x, \varphi^1, \dots, \varphi^N)$. Using these data functions as the input is one of the main difference between our proposed DDSM and the learning-based approaches reviewed in the literature, which brings quite some advantages. First, it extends the boundary data to the domain interior that mimic images and features, for instance the input can be treated as a $(N + 2)$ -channel image, and the nonlinear operator to be trained can be viewed as semantic image segmentation process [11, 24] partitioning a digital image into multiple segments (set of pixels) based on two characteristics: inside or outside the inclusions. The relationship is listed in Table 1 for readers from various background. This analogy provides us with some well-established tools such as the architecture of the CNN [24]. Then, the new index function is assumed to take the form

$$\mathcal{I} = \mathcal{F}_{\text{CNN}}(x, \varphi^1, \dots, \varphi^N) : [H^1(\Omega)]^{2N+2} \rightarrow L^2(\Omega), \tag{3.2}$$

where \mathcal{F}_{CNN} is a function expressed by CNN with parameters to be trained. This structure agrees with the theory discussed in Sect. 3.2. Note that \mathcal{I} becomes an operator from one Sobolev space to another, so in the following discussion we shall call it an index operator, which is one of the main differences from the original DSM. In addition, a remarkable feature of generating φ as the inputs is that the noise can be smoothed out at the boundary through solving the elliptic type PDEs, namely they are highly smoothed inside the domain. For example, in Fig. 1, φ on the right has rough behavior only near the boundary but very smooth inside the domain while φ without noise is shown in the middle.

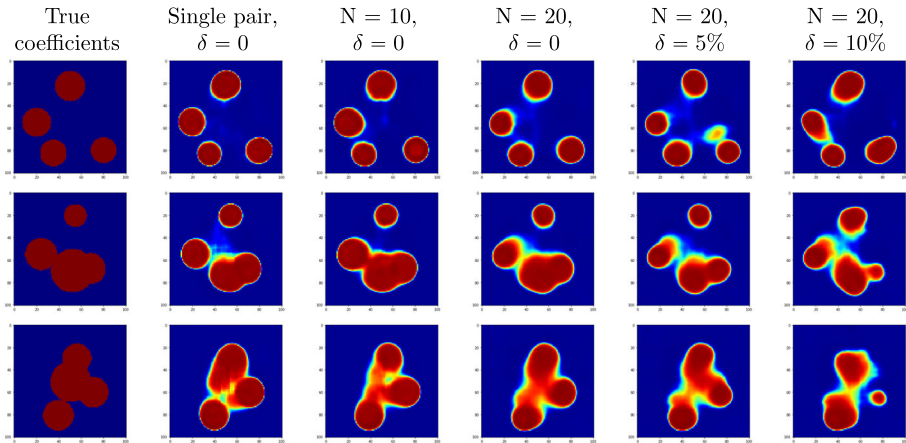


Fig. 3 Reconstruction for 3 cases in **Scenario 1** (4 circles) with different Cauchy data number and noise level: Case 1(top), Case 2(middle) and Case 3(bottom)

Now we proceed to describe the detailed structure of the proposed CNN. For simplicity, we first assume that Ω has rectangular shape and leave the general situation for later discussion about the implementation details. Then we suppose that Ω is discretized by a $n_1 \times n_2$ Cartesian grid where n_1 and n_2 are corresponding to the direction of x_1 and x_2 respectively. Based on the previous explanation, the input of the CNN is a 3D matrix (a 4D tensor for 3D problems). For example, for an inclusion sample with N Cauchy difference functions $\{\varphi^\omega\}_{\omega=1}^N$ generated from $\{g_\omega, f_\omega\}_{\omega=1}^N$, the corresponding input denoted by $z_{in} \in \mathbb{R}^{n_x \times n_y \times (N+2)}$ is a stack of $N+2$ matrices in $\mathbb{R}^{n_1 \times n_2}$, where the first two slices are formed by spatial coordinates x_1 and x_2 respectively, and the rest N slices are the numerical solutions $\varphi^1(x), \dots, \varphi^N(x)$ computed at the Cartesian grid points. The pictorial elucidation of z_{in} is shown in Fig. 2.

The overall configuration of the proposed CNN consists of two parts – convolution and transposed convolution networks as illustrated in Fig. 2. The convolution network corresponds to feature extractor that transforms the input to multidimensional feature representation, where the i -th block can be represented as

$$\phi_i^c(z_{conv}) = \mathcal{M}(\kappa(\beta(W_{conv} * z_i + b_{conv}))), \tag{3.3}$$

in which z_i is the input image, \mathcal{M} is the max-pooling layer that mainly help in highlighting the most prominent features of the input image by selecting the maximum element from each non-overlapping subregions, W_{conv} is the convolution filter for the 2D convolutional layer, κ indicates the activation function, β is batch normalization layer [24], $*$ denotes the convolution operation, and b_{conv} refers the bias. The transposed convolution network is more like a backwards-strided convolution that produces the output by dilating the extracted feature with the learned filter, where the i -th block can be expressed as

$$\phi_i^t(z_{trans}) = \mathcal{C}(\kappa(\beta(\mathcal{T}(z_i, W_{trans}, b_{trans})))), \tag{3.4}$$

where \mathcal{T} refers a fractionally-strided convolution layer, W_{trans} and b_{trans} are the corresponding transposed convolutional filter and the bias, \mathcal{C} is a concatenation layer, and other notations are the same as (3.3). Θ is defined as all the unknown parameters to be learned in training, which includes the convolution and transposed convolutional filters, as well as the biases. In

this work, we consider two activation functions: the ReLu and sigmoid

$$\kappa(z) = \max\{0, z\} \quad \text{and} \quad \kappa(z) = \frac{1}{1 + e^{-z}}. \tag{3.5}$$

It is known that choosing ReLu can significantly enhance the sharpness of reconstruction. But our experience suggests that only using ReLu may not yield satisfactory reconstructions sometimes for the situation that inclusions are far away from the training data set. Nevertheless, the numerical results show that using the sigmoid activation in the last layer may improve the reconstruction for those situations.

Then the entire CNN model can be represented as

$$\mathbf{y}_{\text{out}} = \phi_{N_t}^t \circ \dots \circ \phi_1^t \circ \phi_{N_c}^c \circ \dots \circ \phi_1^c(z_{\text{in}}), \tag{3.6}$$

where the output \mathbf{y}_{out} is a $n_1 \times n_2$ matrix which is supposed to approximate an inclusion distribution, i.e., the entire index function values on the domain Ω . Here the complicated DNN function can approximate the index functional or operator, namely

$$\phi_{N_t}^t \circ \dots \circ \phi_1^t \circ \phi_{N_c}^c \circ \dots \circ \phi_1^c \approx \mathcal{I},$$

where, again, the input is not the data at individual points but on the entire domain. As mentioned before, different from the conventional DSM, there is no closed form for the DNN-based index functional, but those parameters to be trained may better approximate the true index operator for the current available data.

Let S be the total number of training samples. To measure the accuracy of the CNN model (3.6), we employ the mean squared error (MSE) as the loss function

$$\mathcal{L}_{\text{loss}}(\Theta) = \frac{1}{S} \sum_{\ell=1}^S (\mathbf{y}_{\text{out}}(z_{\text{in}}^\ell) - \mathcal{I}^\ell)^\top (\mathbf{y}_{\text{out}}(z_{\text{in}}^\ell) - \mathcal{I}^\ell), \tag{3.7}$$

where \mathcal{I}^ℓ is the true distribution (index values) corresponding to the ℓ -th inclusion sample, z_{in}^ℓ denotes the input image that is related to Cauchy difference functions $\{\varphi^{(\ell, \omega)}\}_{\omega=1}^N$ for the ℓ -th inclusion sample. To reduce the infeasibility of gradient descent algorithm for large datasets, stochastic gradient descent (SGD) is implemented to find the minimization of the loss function (3.7), for which we omit the details here.

3.2 Unique Determination

Note that the proposed DNN model implicitly assumes that inclusion distribution can be uniquely determined by those data functions $\{\varphi^\omega\}_{\omega=1}^N$. In this subsection, we show that this is indeed true for $N \rightarrow \infty$ and $\mu_0 > 0, \mu \geq \mu_0$ in D , i.e., the boundary data are all known in the Sobolev space $H^{-1/2}(\partial\Omega) \times H^{1/2}(\partial\Omega)$. Let us recall some known results at first. Define the difference between the operators Λ_μ and Λ_{μ_0} as

$$\tilde{\Lambda}_\mu = \Lambda_\mu - \Lambda_{\mu_0} : H^{-1/2}(\partial\Omega) \rightarrow H^{1/2}(\partial\Omega). \tag{3.8}$$

According to [7], based on the assumption that $\mu_0 > 0$ and $\mu \geq \mu_0$ in D , it is known that $\tilde{\Lambda}_\mu$ is a self-adjoint and positive operator, and thus it admits the eigenpairs $(\lambda_\omega, v_\omega), \omega = 1, 2, \dots$, with $\lambda_\omega > 0$, such that

$$\tilde{\Lambda}_\mu v_\omega = \lambda_\omega v_\omega, \quad \omega = 1, 2, \dots, \tag{3.9}$$

where v_ω form an orthogonal basis of the space $H^{-1/2}(\partial\Omega)$. Let $\mathcal{R}(\tilde{\Lambda}_\mu^{1/2})$ be the range of the operator $\tilde{\Lambda}_\mu^{1/2}$. By the factorization theory for DOT [7], it is known that the information of $\mathcal{R}(\tilde{\Lambda}_\mu^{1/2})$ together with the probing functions can be used to determine the inclusion distribution. Here we shall employ this theory to show that the data functions can also uniquely determine the inclusion distribution.

Theorem 3.1 *Given a collection of orthonormal basis functions $\{g_\omega\}_{\omega=1}^\infty$ in $H^{-1/2}(\partial\Omega)$, let $\{g_\omega, \Lambda_\mu g_\omega\}_{\omega=1}^\infty$ be the corresponding Cauchy data pairs and let $\{\varphi^\omega\}_{\omega=1}^\infty$ be the generated data functions. Then the inclusion can be uniquely determined by the data functions $\{\varphi^\omega\}_{\omega=1}^\infty$.*

Proof According to [7], we know that $x \in D$ if and only if the corresponding probing function $\eta_x(\xi) \in \mathcal{R}(\tilde{\Lambda}_\mu^{1/2})$. By the Picard’s criterion, it is further equivalent to the convergence of the following sequence

$$S(x; \{v_\omega\}_{\omega=1}^\infty) = \sum_{\omega=1}^\infty \frac{(\eta_x, v_\omega)_{\partial\Omega}^2}{|\lambda_\omega|} < \infty. \tag{3.10}$$

Let us first focus on $\{g_\omega\}_{\omega=1}^\infty$ being exactly the eigenfunctions $\{v_\omega\}_{\omega=1}^\infty$ in (3.9). Then, using $v_\omega = \lambda_\omega^{-1} \tilde{\Lambda}_\mu v_\omega$ and the identity in (2.7), i.e., using the integration by parts, we have

$$\begin{aligned} (\eta_x, v_\omega)_{\partial\Omega} &= \lambda_\omega^{-1} (\eta_x, \tilde{\Lambda}_\mu v_\omega)_{\partial\Omega} \\ &= \lambda_\omega^{-1} (\eta_x, (\Lambda_\mu - \Lambda_{\mu_0}) v_\omega)_{\partial\Omega} = \lambda_\omega^{-1} \varphi^\omega(x) \end{aligned} \tag{3.11}$$

where φ^ω are the data functions corresponding to $g_\omega = v_\omega$ solved from (2.6) with $s = 0$. Then, combining (3.10) and (3.11), we have

$$S(x; \{v_\omega\}_{\omega=1}^\infty) = \sum_{\omega=1}^\infty \frac{|\varphi^\omega(x)|^2}{|\lambda_\omega|^3}. \tag{3.12}$$

Since $|\lambda_\omega| = \|\varphi^\omega\|_{L^2(\partial\Omega)} / \|v_\omega\|_{L^2(\partial\Omega)}$, the inclusion distribution is determined by the convergence property of the sequence in (3.12) which is further determined by φ^ω . Next, for general orthonormal basis $\{g_\omega\}_{\omega=1}^\infty$ as the basis, following the argument in [5, Theorem 3.8], we can express $\{v_\omega\}_{\omega=1}^\infty$ by the expansion of $\{g_\omega\}_{\omega=1}^\infty$ which is then used to express $\{\varphi^\omega\}_{\omega=1}^\infty$ corresponding to $\{g_\omega\}_{\omega=1}^\infty$. Plugging it in (3.12) we have the new series to determine the inclusion shape that only depends on $\{\varphi^\omega\}_{\omega=1}^\infty$. \square

Remark 3.2 According to the theory above, we can conclude that the whole uniqueness does not depend on the value of μ_0 and μ , which only requires they are distinguished from each other. Indeed, this theoretical property in a certain sense can manifest itself in the proposed DDSM since the nice reconstruction can be obtained for the different values of μ even if they are significant from those used for training.

Remark 3.3 In our work for EIT [23], a similar data function is generate, but the input of the fully neural network is the gradient of the Cauchy difference function. This difference is theoretically supported by the format of φ in the generated series (3.12).

Suppose that the data functions $\{g_\omega\}$ are exactly the eigenfunctions $\{v^\omega\}$, $\omega = 1, 2, \dots$, and let φ^ω be the corresponding harmonic extensions. Then, the argument of Theorem 3.1 enables us to explicitly construct a function approximating the true indicator function: given any $\epsilon > 0$, there exists a sufficiently large integer N and a function \mathcal{I}_N depending on φ^ω , $\omega = 1, \dots, N$, such that

$$\|\mathcal{I} - \mathcal{I}_N\|_{L^\infty(\Omega)} \leq \epsilon. \tag{3.13}$$

To see this, according to the series in (3.12), we define $\mathcal{S}_N(x) := \sum_{\omega=1}^N \frac{|\varphi^\omega(x)|^2}{|\lambda_\omega|^3}$. Theorem 3.1 suggests that there is a constant ρ such that $\rho > \mathcal{S}_N(x)$, $\forall x \in D$, and given each $\epsilon > 0$ there is an integer N such that $\mathcal{S}_N(x) > 4\rho\epsilon^{-2}/\pi^2$, $\forall x \notin D$. Then, we can define \mathcal{I}_N as

$$\mathcal{I}_N(x) = 1 - \frac{2}{\pi} \arctan\left(\frac{\pi\epsilon}{2\rho}\mathcal{S}_N(x)\right) \quad (3.14)$$

The desired inequality in (3.13) can be simply verified by the basic inequality $z > \arctan(z) \geq \frac{\pi}{2} - z^{-1}$, $\forall z > 0$. However, it is worthwhile to mention that such function \mathcal{I}_N is generally not computable in practice as it still requires the full spectrum information of $\tilde{\Lambda}_\mu$. If there are only limited data pairs available, the approximation formula remains unknown in theory. Here, instead we borrow the mathematical structure described above to construct CNN for approximation.

3.3 Implementation Details

In order to perform the convolution operator, the discretization of Ω needs to be chosen as Cartesian grids which are natural for rectangular domain. Note that the data functions φ^ω are solved from the equations merely with the background absorption coefficient, so there is no need to require the mesh to align with the interface, and a simple Cartesian mesh may yield satisfactory numerical solutions. As for a general shaped domain, we just need to immerse it into a rectangle such that the Cartesian grid can be generated on the whole rectangular fictitious domain. However, if the data functions φ^ω are computed on a general triangular mesh, they need to be reevaluated at the generated Cartesian grid points since the values at the triangle mesh points are not able to directly support the CNN computation. To alleviate the computational burden, we can prepare these data functions before training rather than solve them during training. It should be noted that the DDSM indeed requires more memories compared with other deep learning approaches that directly input the original boundary point data, since the newly generated data functions are at least 2D.

4 Numerical Experiments

In this section, we present numerical experiments to demonstrate that our newly developed DDSMs are effective and robust for the reconstruction of inhomogeneous inclusions in the DOT. For such medical imaging problems, in general only limited data can be obtained from real clinical environments, but instead, simulation or the so-called synthetic data are cheap, easily accessible and are not subject to the objective factors. As suggested by many works in the literature [8, 21, 23, 27, 44], these features make synthetic data suitable for training DNN, and the resulting network can be further enhanced by realistic data from clinics. So we shall sample the inclusion distribution by randomly creating some basic geometric objects in Ω which are then used to generate synthetic data set for training and testing.

4.1 2D Problem Setting and Data Generation

Let the boundary (interface) of the i -th ($i = 1, 2, \dots, M$) basic geometric object be represented by a level-set function $\Gamma_i(x_1, x_2)$, then define the boundary of the inclusion as the zero level

set of

$$\Gamma(x_1, x_2) = \min_{i=1,2,\dots,M} \{\Gamma_i(x_1, x_2)\}. \quad (4.1)$$

Then the support of the inclusions is the subset $\{(x_1, x_2) : \Gamma(x_1, x_2) < 0\}$. More specifically, we consider the follow two scenarios with different basic geometric objects for training data generation:

Scenario 1: Γ_i are 5 random circles with the radius sampled from $\mathcal{U}(0.2, 0.4)$ and the center sampled from $\mathcal{U}(-0.7, 0.7)$.

Scenario 2: Γ_i are 4 random ellipses with the longer axis, the eccentricity and the center points sampled from $\mathcal{U}(0.2, 0.6)$, $\mathcal{U}(0, 0.9)$ and $\mathcal{U}(-0.7, 0.7)$, respectively.

Here $\mathcal{U}(a, b)$ denotes the uniform distribution over $[a, b]$. Sampling circles or ellipses are widely used in many deep learning methods for generating synthetic data in DOT [21, 27]. However, a major difference is that the format (4.1) allows those basic inclusions to touch and overlap with each other such that the overall inclusion distribution could be much more geometrically complicated, which makes the reconstruction more challenging. Besides, we set $\mu_0 = 0$ and $\mu_1 = 50$ for our experiments.

For the boundary conditions, following our previous work [23], we still use Fourier functions as the applied surface flux boundary data g_ω since they naturally form the orthogonal bases on the boundary. In particular, we pick the first N modes:

$$\begin{aligned} g_\omega(\theta) &= \cos(\omega\theta) \quad \omega = 1, 2, \dots, N/2 \quad \text{and} \quad g_\omega(\theta) \\ &= \sin((\omega - N/2)\theta), \quad \omega = N/2 + 1, \dots, N, \end{aligned} \quad (4.2)$$

where $\theta \in [0, 2\pi)$ is the angle of $(x_1, x_2) \in \partial\Omega$, and we choose $g_1(\theta) = \cos(\theta)$ for the case of a single measurement, i.e., $N = 1$, and $N = 10, 20$ for the case of multiple measurements (Fig. 6).

For the diffusion–reaction equation with a discontinuous absorption coefficient, the solutions generally have smooth first-order derivatives [10] such that typical second-order schemes may achieve optimal convergence for sufficiently fine meshes. But if the jump of the coefficient is large, there might be a thin layer around the inclusion interface [10] requiring extremely fine meshes to resolve [9]. Here, for simplicity, we focus on the case of moderate jump and apply a uniform 100×100 Cartesian mesh to solve the forward model (1.1). Nevertheless, as we only need the simulated solutions on the boundary, which is away from the coefficient discontinuity, the numerical solution may not be effected much by the discontinuity. For each inclusion sample with every boundary condition in (4.2), we can obtain the numerical solutions u_ω and $f_\omega = u_\omega|_{\partial\Omega}$, and the data pairs $(f_\omega, g_\omega)_{\omega=1}^N$ are used in (3.1) to generate data functions as the input of the proposed DNN. Even so, numerical errors can not be avoided in this procedure, which are then considered as noise in the data. In fact, such errors are much smaller than the artificial noise described below.

On the one hand, it is generally difficult to obtain the accurate knowledge of the physiological noise presented in human data. On the other hand, as discussed in Sect. 3 and shown in the left plot of Fig. 1, using boundary data to generate the data functions φ^ω can smooth out the noise in the sense that the interior of the data functions can be highly smooth. Both the original DSM [15, 16] and our recent work using DDSM for solving EIT [23] indicate that this mechanism can significantly enhance the robustness with respect to the noise. Therefore, instead of adding noise in the training set, we will add very large noise in the test data. This is intentionally to test the robustness of the proposed DNN with respect to noise that is not contained in the training set. In particular, we apply the following point-wise noise on the

Table 2 Comparison of NNs with $f - f_0$ as the input and with ϕ as the input ($\mu_0 = 0, \mu = 10$)

Methods	1 data pair			10 data pair		
	MSE	CrossEntropy	Dice	MSE	CrossEntropy	Dice
f	0.0538	0.1776	0.2446	0.0440	0.1516	0.1880
$\varphi(x)$	0.0301	0.1058	0.1605	0.0310	0.1090	0.1595
improvement	43.95%	40.44%	34.38%	29.56%	28.11%	15.19%

synthetic measured data

$$f_{\omega}^{\delta}(x) = (1 + \delta G(x))f_{\omega}, \tag{4.3}$$

where δ is the signal-to-noise ratio chosen as 0, 5% and 10%, $G(x)$ are Gaussian random variables with standard norm distribution which are assume to be independently identical with respect to the points x . Different from the set-up in [15] that uses a spatially-invariant noise, such noise will cause very rough data on the boundary as shown in Fig. 1, challenging the robustness of the reconstruction algorithms more seriously.

4.2 2D Reconstructions

Now we present reconstruction results for each scenario in 2D, and explore the performance of the proposed algorithm for some more challenging situations.

Effect of ϕ . To begin with, we first present a group of results to demonstrate that using φ as the input to the DNNs can indeed significantly improve the reconstruction accuracy. For this purpose, we train a CNN with merely $f - f_0$ as the input and compare the performance with the proposed DDSM by examining their average accuracy on a test set in terms of different metrics. The numerical results are reported in Table 2 where we can see the DDSM outperforms the approach with $f - f_0$ as the input a lot. In addition, the employment of ϕ can improve the performance in many aspects. When applying the NNs trained with $\mu_0 = 0, \mu = 10$ to the data generated with $\mu_0 = 0, \mu = 100$, we still observe significant improvement. As for the reconstruction for out-of-distribution, i.e., the shape to be reconstructed is very different from the training data set, the improvement is also huge, see the numerical example of Fig. 9.

Some basic results: The reconstruction results for three basic cases are provided in Figs. 3 and 4 for each scenario. The figures clearly show the accurate reconstruction of the proposed algorithm. In particular, for the third case of Fig. 4, the true inclusion has a concave portion near the domain center away from the boundary data which is generally difficult to be captured. But the proposed algorithm can recover it quite satisfactorily.

We observe that the proposed algorithm using single or multiple measurements almost gives comparably good reconstruction. However, our further numerical results suggest that using multiple measurements can significantly enhance the robustness of the reconstruction with respect to noise. To demonstrate this, for the first case in Scenario 1 (the top one in Fig. 3), we present the reconstruction of the single pair and 10 pairs with noise in Fig. 5. It is observed that even 5% noise can totally destroy the reconstruction for the single measurement case while the performance of 10 pairs of measurements is much better but still worse than 20 pairs. It is also worth mentioning that, even for the single measurement, the reconstruction is still highly robust with respect to the spatially-invariant noise used in [15], i.e., G is independent of x in (4.3), of which the results are omitted here (Fig. 6).

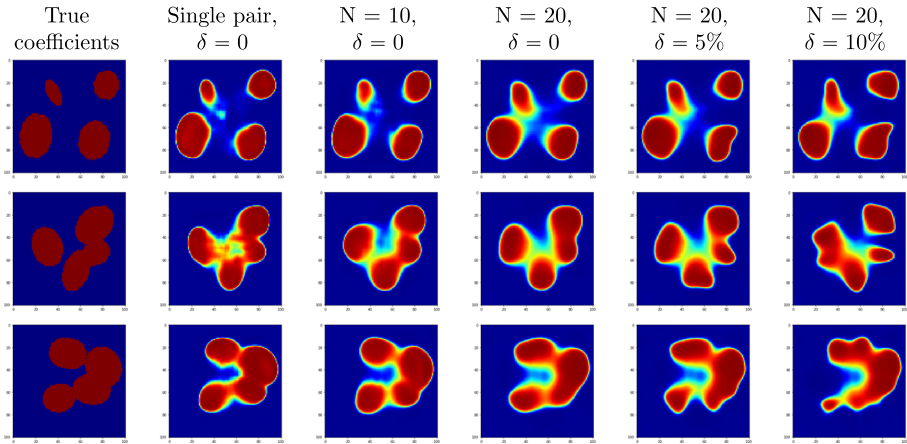


Fig. 4 Reconstruction for 3 cases in **Scenario 2** (4 ellipses) with different Cauchy data number and noise level: Case 1 (top), Case 2 (middle) and Case 3 (bottom)

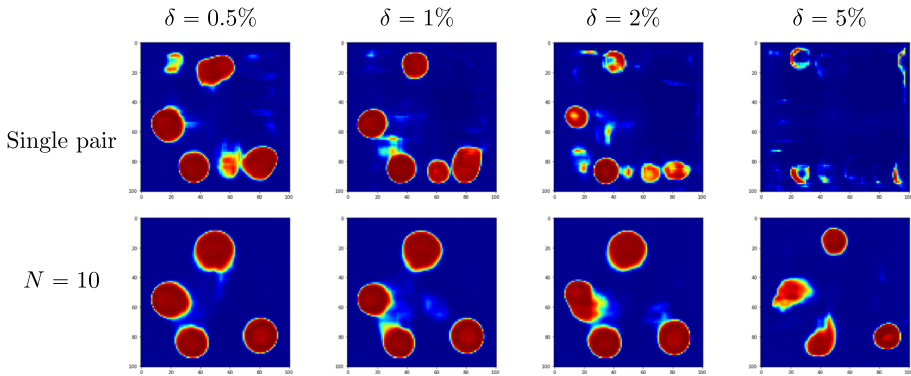


Fig. 5 Comparison of reconstruction with respect to noise for the single measurement and 10 pairs of measurements

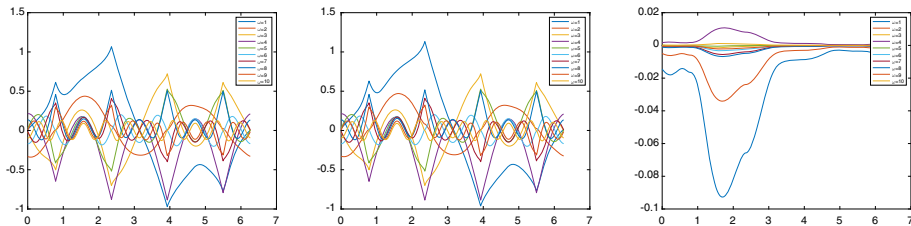


Fig. 6 Plots of $g_{1,\omega}|_{\partial\Omega}$, $g_{2,\omega}|_{\partial\Omega}$ and $g_{1,\omega}|_{\partial\Omega} - g_{2,\omega}|_{\partial\Omega}$ versus the polar angle θ (of points on the boundary) where $g_{1,\omega}$ and $g_{2,\omega}$ correspond to inclusion distribution with and without the center inclusion respectively

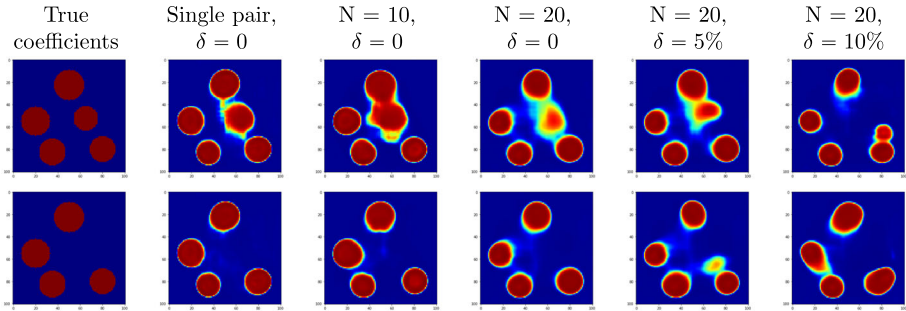


Fig. 7 Comparison of reconstruction for two close inclusion distribution

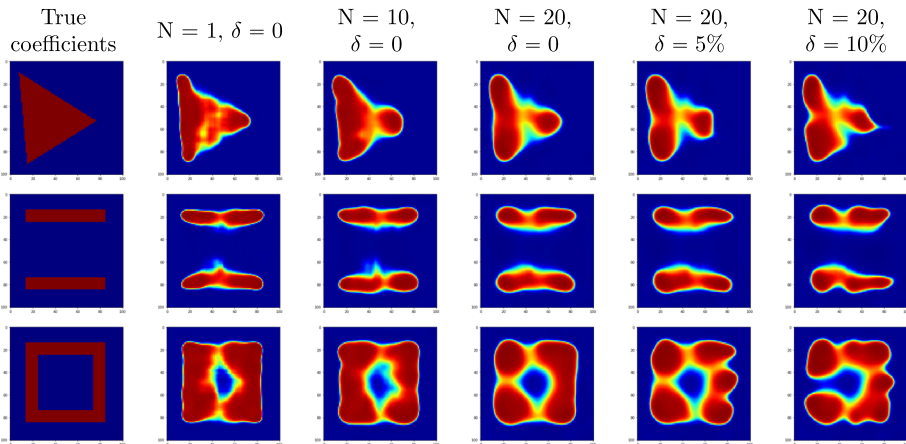


Fig. 8 CNN-DDSM reconstruction for 3 special inclusion shapes: one triangle (top), two long rectangular bars (middle) and a rectangular ring (bottom)

Sensitivity to Data: The DOT is well-known to be highly ill-posed which means the inclusions are very insensitive to the boundary data. To examine this phenomenon and test the sensitivity of the algorithm with respect to the data, we consider two different inclusion distributions in Fig. 7, where the only difference is the center inclusion. We plot the corresponding f_ω by fixing the applied flux g_ω , $\omega = 1, 2, \dots, 10$, which are indeed quite close to each other. Namely, the center inclusion is extremely insensitive to noise, which makes its reconstruction very difficult. The results in Fig. 7 show that the center inclusion can still be clearly captured. We highlight that for 20 pairs of measurements the center inclusion is captured even with 5% noise which is even larger than the relative difference of their boundary data. These results demonstrate that the proposed algorithm can dig out the small difference buried in boundary data for various inclusion distribution but still keep the robustness with respect to the noise.

Out-of-distribution: We note that the basic geometric shape such as circles or ellipses may be more appropriate to be chosen to imitate the tumors for the clinical application of the DDSM. But even so, it may not be expected that the shape to be reconstructed is always covered by or close to the training data set, which triggers us to further investigate the capability of the DDSM for the inclusion distribution that has the geometry out of the scope of the training data, that is, they can not be generated by those basic geometry objects. Here

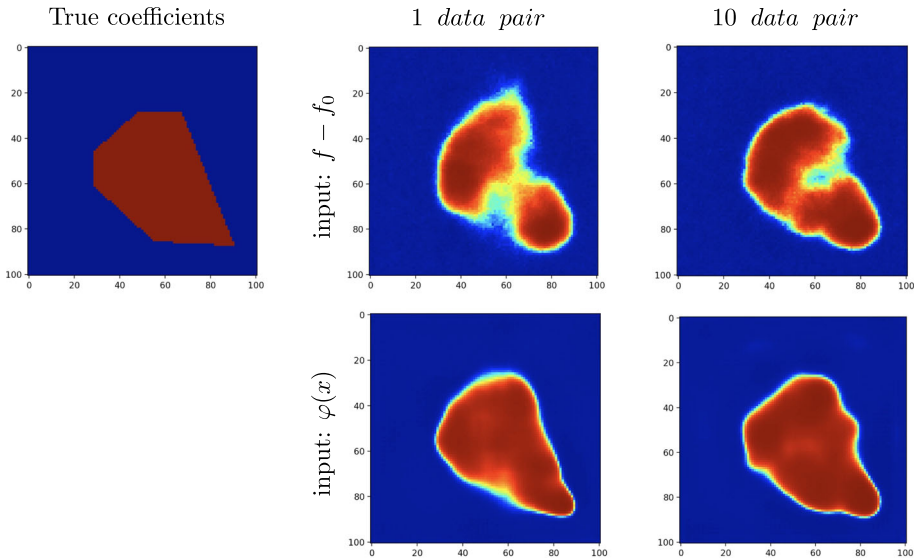


Fig. 9 DDSM reconstruction for a hexagon (out-of-distribution) with different inputs: input $f - f_0$ without harmonic extensions (top), input $\varphi(x)$ (bottom). MSE loss: 0.0471 for the input f and 0.0169 for the input φ (1 data pair); 0.0262 for the input $f - f_0$ and 0.0125 for the input φ (10 data pair). The improvement by using φ is 64.12% (1 data pair) and 52.29% (10 pairs)

we present the reconstruction for three different shapes: a triangle, two bars and an annulus in Fig. 8. In particular, the annulus has a hollow center which is difficult to be captured by the lights sourced at the boundary. But we can still observe the satisfactory reconstruction of their basic geometric properties for all these inclusions. Similar to the previous results, the reconstruction by 20 parts is still robust with respect to the large noise. Note that one can add these shapes to the training data set if some more accurate reconstruction is needed.

In addition, we also apply the trained DNN to a hexagonal inclusion, see the reconstruction in Fig. 9. Here, we specifically compare two different DNNs where one has φ as the input which is the key for the proposed DDSM, and the other one trivially has the $f - f_0$ as the input.

Limited data points: Note that all the previous results are generated by assuming that the data are available at every boundary grid point for using finite difference methods to generate data functions φ^ω , which may not be practical. Indeed, according to DOT experiments, see [34] for example, even if a camera may receive the light at every point, the light sources can be placed on only a few points on the boundary. So we shall explore this issue by limiting the data points where the Neumann data g_ω are available. In particular, we consider the case that there are $4L$, $L = 8, 16$, data points on the boundary with L points equally distributed on each side of the domain. The data g_ω are assumed to be obtained at these points which are then linearly interpolated to generate functions on the boundary. These functions are then used to generate the corresponding Dirichlet data functions f_ω available at all the boundary mesh points. Then the interpolated Neumann data and the simulated Dirichlet data are used to generate data functions φ^ω as the input to the same DNN used above, and the reconstruction results are presented in Fig. 10 for the first case in Fig. 4. As we can see, the three relatively larger ellipses are reconstructed quite satisfactory even with 10% noise. However, the result for the small ellipse is not as good as the others. We guess it may be due to its small size

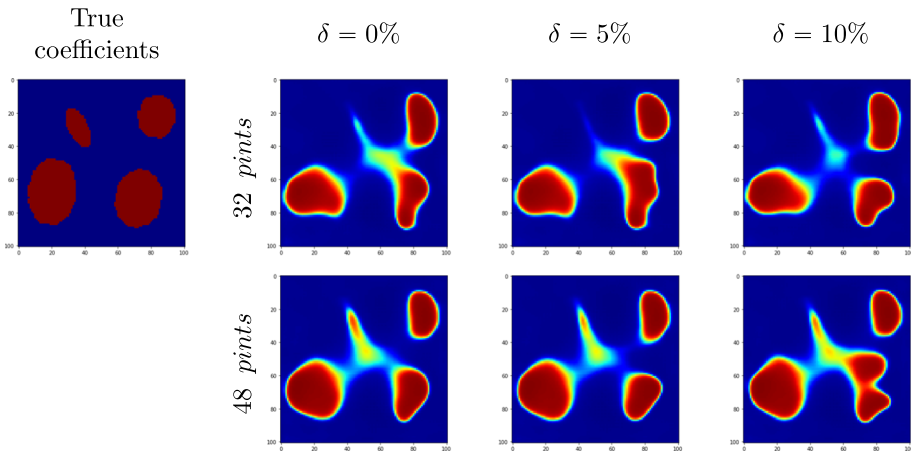


Fig. 10 DDSM reconstruction for random ellipses with limited data points and $N = 20$

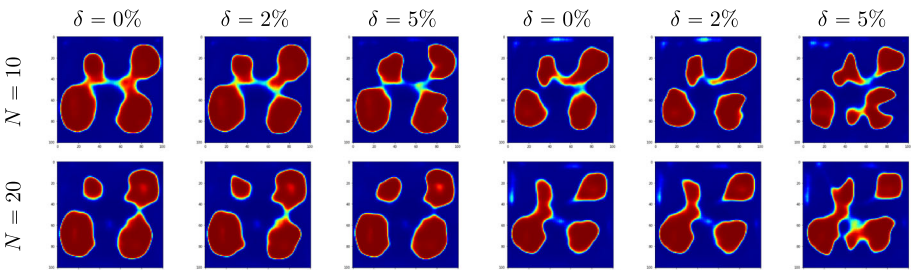


Fig. 11 CNN-DDSM reconstruction for different μ : $(\mu_0, \mu) = (0, 200)$ (left three columns) and $(1, 100)$ (right three columns)

that receives too little light for passing its geometric information to the boundary. Despite its inaccurate shape, the algorithm still tells us that an inclusion exists around the upper-left corner.

Reconstruction for different μ : Now we demonstrate that the proposed algorithm can be used to obtain reasonable reconstruction even though the true absorption coefficient values are much different from those used to generate training data. Here we also use the first case in Fig. 4 of the scenario 2 as an example. But we generate the boundary data with two different groups of absorption coefficient values $(\mu_0, \mu_1) = (0, 200)$ and $(1, 100)$ where the background and inclusion coefficient values may both vary. The same DNN is used predict the inclusion geometry of which the results are presented in Fig. 11. Although the reconstruction is not as good as the one shown in Fig. 4, we can see that all the four ellipses are clearly captured and the reconstruction is still stable with respect to noise. We highlight that this feature is very important for the practical clinical situations as the material property of patients' tumors may vary and may not be known accurately. In spite of this, the proposed algorithm still has the potential to detect the appearance and shapes of the tumors to certain extend.

In addition, we also apply the trained DNNs to a non-piecewise-constant function μ which has a Gaussian distribution here, see Fig. 12 for the reconstruction. In this case, the inclusion boundary can be understood as a diffusion interface. Note that the training data set used here

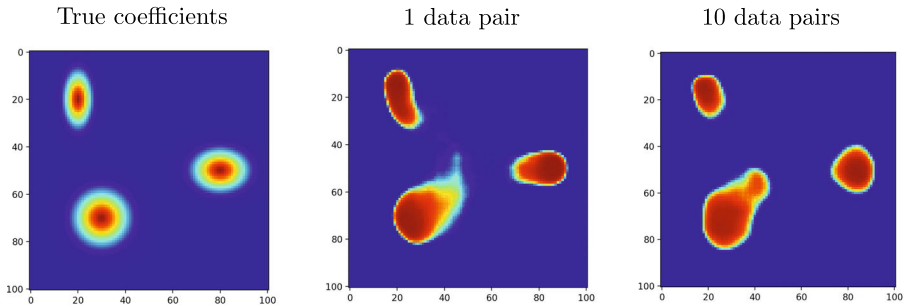


Fig. 12 DDSM reconstruction for μ with a random Gaussian distribution

is still generated by μ as piecewise constants. Indeed, we can still observe quite reasonable reconstructions of the location.

Comparison of DSM and DDSM: we also present a group of numerical examples to compare the classical DSM and the proposed DDSM. The classical DSM is implemented through the formula (2.8), where $\varphi(x)$ is obtained by solving (2.6) with $s = 1$. The numerical results are reported in Fig. 13. Specifically, we have implemented the formula with the boundary data generated from various frequencies. It can be certainly observed that the proposed DDSM can generate sharper reconstruction than the classical DSM with a single data pair. Of course, with multiple data pairs, the reconstruction can be improved a lot. Note that the classical DSM may not handle multiple data pairs systematically. In addition, we also present the plots of the normalized φ which do not show any reasonable information. Thus, the process by CNN is critical to dig the hidden information to produce accurate reconstructions.

4.3 3D Reconstruction

In this subsection, we apply the DDSM to 3D DOT problems. We consider a cubic domain $\Omega = (-1, 1) \times (-1, 1) \times (-1, 1)$, and two ellipsoids with the level-set functions $\Gamma_i(x_1, x_2, x_3)$, $i = 1, 2$, which have the axis length, rotation angles and center points sampled from $\mathcal{U}(0.4, 0.6)$, $\mathcal{U}(0, 2\pi)$ and $[\mathcal{U}(-0.4, 0.4)]^3$, respectively. Similar to (4.1), we let the inclusions be generated by the following function involving the random variables described above

$$\Gamma(x_1, x_2, x_3) = \min_{i=1,2} \{\Gamma_i(x_1, x_2, x_3)\}. \tag{4.4}$$

In the 3D case, we employ the first 9 spheric harmonic functions below and map them to the surface of the domain

$$\begin{aligned} \Re Y_l^m(\theta, \phi) &= \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos(\theta)) \sin(m\phi) \\ \Im Y_l^m(\theta, \phi) &= \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos(\theta)) \cos(m\phi) \end{aligned} \tag{4.5}$$

to generate the flux boundary data, where $P_l^m(z)$ are the Legendre polynomials with $l = 0, 1, 2, 0 \leq m \leq l$, of which some examples are plotted in Fig. 14. The other setting-ups are similar to the 2D case.

The reconstruction results of two typical cases are presented in Fig. 15, where the two

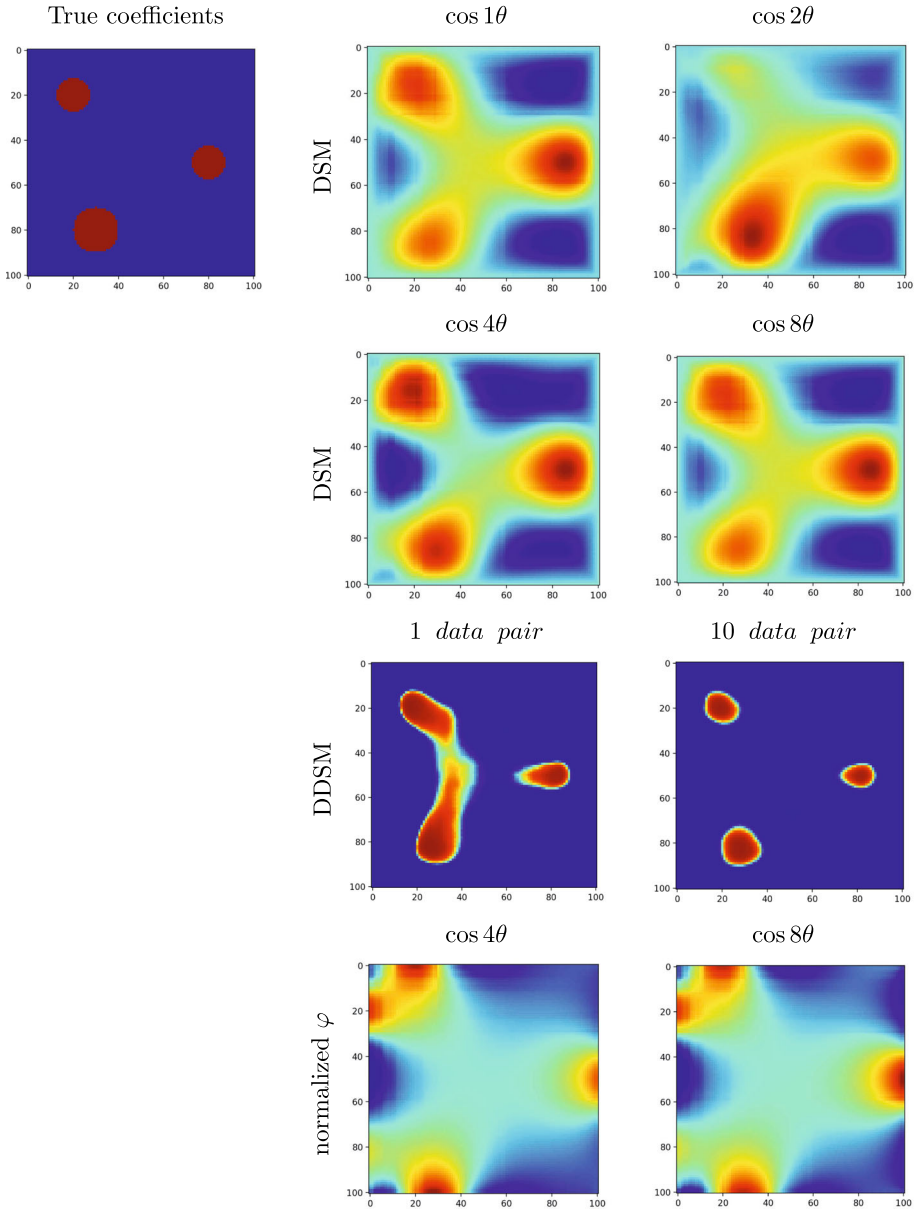


Fig. 13 Comparison of DSM reconstruction and DDSM reconstruction for 3 circles. The MSE of the results of DSM with frequency $\cos \theta$, $\cos 2\theta$, $\cos 4\theta$, $\cos 8\theta$ are 0.0512, 0.0532, 0.0502 and 0.0487. The MSE of the results of DDSM with 1 and 10 data pairs are 0.0429 and 0.0161, respectively. The last row is the normalized φ , i.e., $\varphi/(\|\varphi\|_{L^\infty(\Omega)} + |\varphi(\mathbf{x})|)$

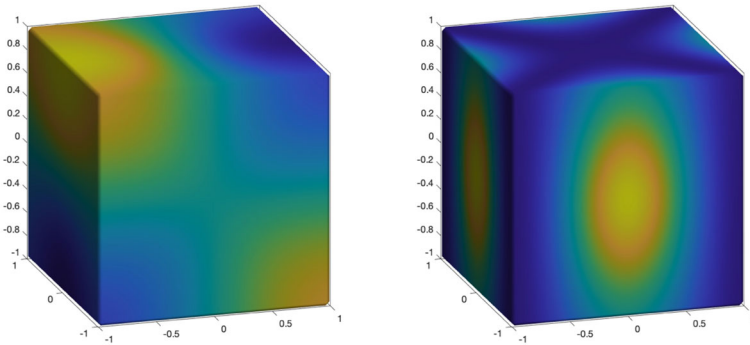


Fig. 14 Harmonic functions on the surface of Ω

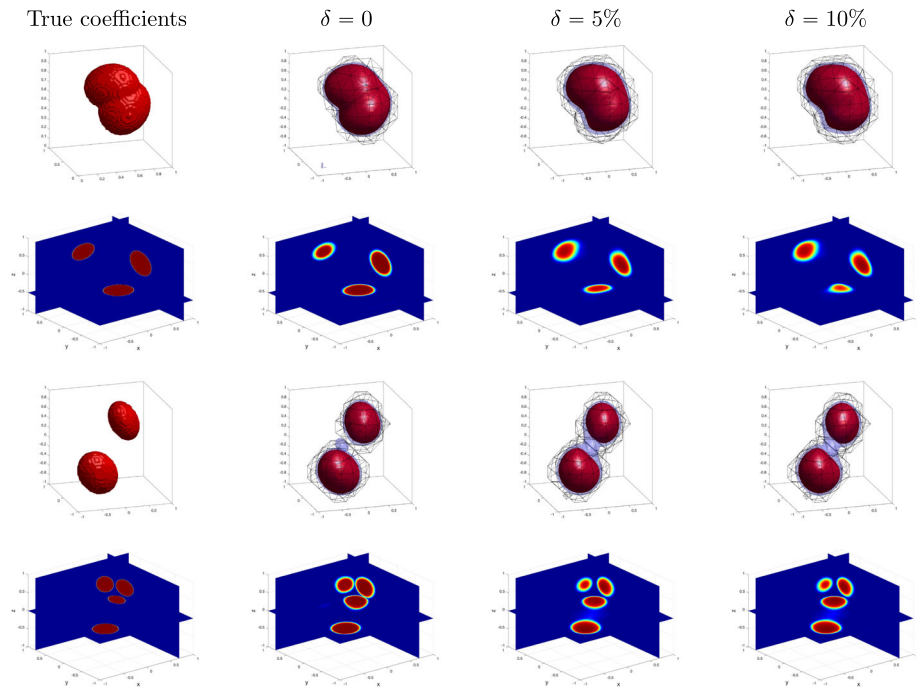


Fig. 15 Reconstruction for 2 cases with different Cauchy data number and noise level: Case 1(top) and Case 2(bottom)

ellipsoids are glued together in the first case and separated in the second one. To show the reconstructed inclusion distribution, we employ the 3D density plots: the red, blue and mesh surface are corresponding to the isosurface with the values 0.75, 0.06 and 0.025, respectively, which is the first row of each case. In addition, we also plot some cross sections which are given in the second row of each case. As shown by all these plots, the reconstruction results are quite accurate and also robust with respect to large noise. Note that for 3D DOT problems, solving 3D forward problems for iterative methods can be highly expensive. Given the efficiency, accuracy and robustness of the proposed DDSM, we believe it can be very attractive for further real-world applications.

Fig. 16 Data points on two faces of the cubic domain

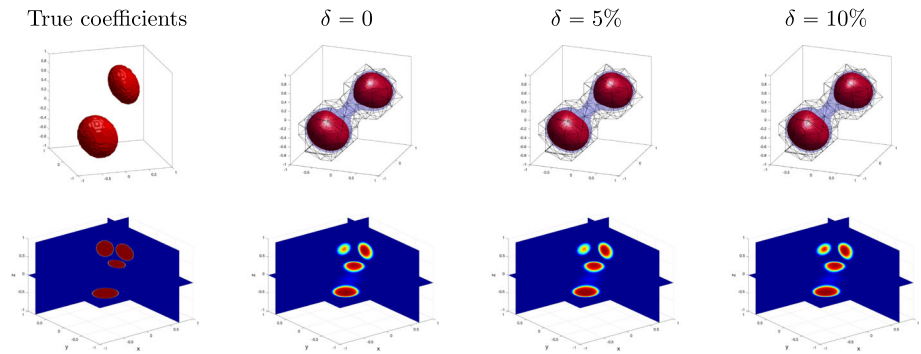
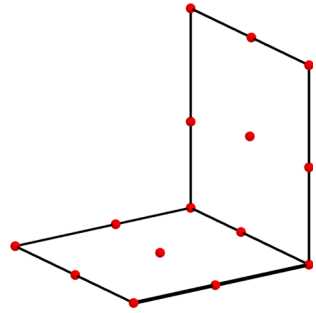


Fig. 17 Reconstruction with limited data points on the boundary

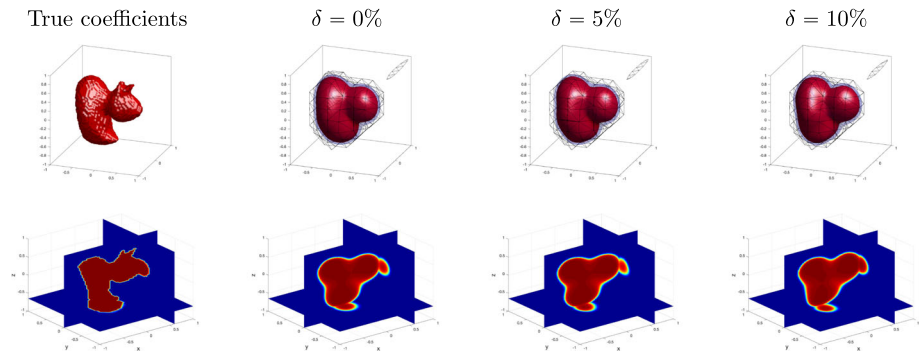


Fig. 18 Reconstruction for a duck-shaped inclusion with different noise level

For the second case, we also test the performance of the DDSM for that the boundary data are only available at a few points not every grid point. Specifically, as only the low-frequency data are used for the 3D examples, we assume there are only 9 points evenly distributed on each face of the domain, see Fig. 16 for an illustration, then there are only 26 points distributed on the boundary. Similar to the 2D case, linear interpolation is used to generate data functions on each face. The reconstruction results are presented in Fig. 17, where we can see the shape can be recovered quite well even though the data is extremely limited.

Moreover, we apply the DNN to an inclusion which has a duck shape that is not in the scope of the training data set, i.e., it is certainly not a union of two ellipsoids (Fig. 18). In

this case, we can still observe that the reconstruction can capture the basic duck shape which is still robust with respect to the large noise. But the geometry of the beak is lost as the training data set is only sampled from two ellipsoids. It can be expected that more detailed geometry can not be recovered accurately. However, one can still add more inclusion samples with various geometric randomness to the training data set in order to attain better accuracy (Fig. 18).

5 Conclusions

Inspired by the DSM [15], this paper proposes a novel deep direct sampling method for the DOT problem, where the neural network architecture is designed to approximate the index functional. Hybridizing the DSM and CNN has several advantages. First, the index functional approximated by the CNN is capable of systematically incorporating multiple Cauchy data pairs, which can improve the reconstruction quality and the algorithmic robustness against the noise. Second, once the neural network is well trained, the data-driven index functional can be executed efficiently, which inherits the main benefit of the conventional DSM. Third, the DDSM can successfully handle challenging cases such as limited data and 3D reconstruction problems, which shows great potential in its practical application. Various numerical experiments justify these findings. Hence, we believe the proposed DDSM provides an efficient technique and a new promising direction for solving the DOT problem.

Acknowledgements The second author and the third author were funded by the startup funding from ShanghaiTech University (No. 2020F0203-000-16), Shanghai Science and Technology Innovation Program (No. 21YF1429100) and National Natural Science Foundation of China (No. 12101406). The first author was funded by NSF DMS-2012465.

Funding The authors have not disclosed any funding.

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Competing interests The authors have not disclosed any competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abdoulaev, G.S., Ren, K., Hielscher, A.H.: Optical tomography as a PDE-constrained optimization problem. *Inverse Prob.* **21**(5), 1507–1530 (2005)
2. Ammari, H., Bossy, E., Garnier, J., Nguyen, L. H., Seppecher, L.: A reconstruction algorithm for ultrasound-modulated diffuse optical tomography. *Proc. Am. Math. Soc.*, **142**, 3221–3236 (2014)
3. Ammari, H., Garnier, J., Nguyen, L.H., Seppecher, L.: Reconstruction of a piecewise smooth absorption coefficient by an acousto-optic process. *Commun. Partial Differ. Equ.* **38**(10), 1737–1762 (2013)

4. Ammari, H., Nguyen, L.H., Seppecher, L.: Reconstruction and stability in acousto-optic imaging for absorption maps with bounded variation. *J. Funct. Anal.* **267**(11), 4361–4398 (2014)
5. Anagnostopoulos, K., Charalambopoulos, A., Kleefeld, A.: The factorization method for the acoustic transmission problem. *Inverse Problem* (2013). <https://doi.org/10.1088/0266-5611/29/11/115015>
6. Applegate, M., Istfan, R., Spink, S., Tank, A., Roblyer, D.: Recent advances in high speed diffuse optical imaging in biomedicine. *APL Photon.* **5**(4), 040802 (2020)
7. Bal, G.: Reconstructions in impedance and optical tomography with singular interfaces. *Inverse Probl.* **21**(1), 113–131 (2004)
8. Ben Yedder, H., BenTaieb, A., Shokoufi, M., Zahiremami, A., Golnaraghi, F., Hamarneh, G.: Deep learning based image reconstruction for diffuse optical tomography. In: Knoll, F., Maier, A., Rueckert, D. (eds.) *Machine Learning for Medical Image Reconstruction*, pp. 112–119. Springer, Cham (2018)
9. Brayanov, I.A.: Numerical solution of a two-dimensional singularly perturbed reaction–diffusion problem with discontinuous coefficients. *Appl. Math. Comput.* **182**(1), 631–643 (2006)
10. Brayanov, I.A., Vulkov, L.G.: Finite volume difference methods for convection-dominated problems with interface. In: Lirkov, I., Margenov, S., Wasniewski, J., Yalamov, P. (eds.) *Large-Scale Scientific Computing*, pp. 429–437. Springer, Berlin (2004)
11. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2017)
12. Cheney, M.: The linear sampling method and the MUSIC algorithm. *Inverse Probl.* **17**(4), 591–595 (2001)
13. Choe, R.: *Diffuse Optical Tomography and Spectroscopy of Breast Cancer and Fetal Brain*. PhD thesis, University of Pennsylvania (2005)
14. Chow, Y.T., Han, F., Zou, J.: A direct sampling method for the inversion of the radon transform. *SIAM J. Imag. Sci.* **14**(3), 1004–1038 (2021)
15. Chow, Y.T., Ito, K., Liu, K., Zou, J.: Direct sampling method for diffusive optical tomography. *SIAM J. Sci. Comput.* **37**(4), A1658–A1684 (2015)
16. Chow, Y.T., Ito, K., Zou, J.: A direct sampling method for electrical impedance tomography. *Inverse Probl.* **30**(9), 095003 (2014)
17. Chow, Y.T., Ito, K., Zou, J.: A time-dependent direct sampling method for recovering moving potentials in a heat equation. *SIAM J. Sci. Comput.* **40**(4), A2720–A2748 (2018)
18. Culver, J.P., Choe, R., Holboke, M.J., Zubkov, L., Durduran, T., Slem, A., Ntziachristos, V., Chance, B., Yodh, A.G.: Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: evaluation of a hybrid frequency domain/continuous wave clinical system for breast imaging. *Med. Phys.* **30**(2), 235–247 (2003)
19. Dehghani, H., Pogue, B.W., Jiang, S., Brooksby, B., Paulsen, K.D.: Three-dimensional optical tomography: resolution in small-object imaging. *Appl. Opt.* **42**(16), 3117–3128 (2003)
20. Deng, B., Gu, H., Carp, S.A.: Deep learning enabled high-speed image reconstruction for breast diffuse optical tomography. In: *Optical Tomography and Spectroscopy of Tissue XIV*, vol. 11639, p. 116390B. International Society for Optics and Photonics (2021)
21. Fan, Y., Ying, L.: Solving optical tomography with deep learning. [arXiv:1910.04756](https://arxiv.org/abs/1910.04756) (2019)
22. Gu, X., Zhang, Q., Bartlett, M., Schutz, L., Fajardo, L.L., Jiang, H.: Differentiation of cysts from solid tumors in the breast with diffuse optical tomography I. *Acad. Radiol.* **11**(1), 53–60 (2004)
23. Guo, R., Jiang, J.: Construct deep neural networks based on direct sampling methods for solving electrical impedance tomography. *SIAM J. Sci. Comput.* **43**(3), B678–B711 (2021)
24. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*, pp. 448–456. PMLR (2015)
25. Ito, K., Jin, B., Zou, J.: A direct sampling method to an inverse medium scattering problem. *Inverse Probl.* **28**(2), 025003 (2012)
26. Ito, K., Jin, B., Zou, J.: A direct sampling method for inverse electromagnetic medium scattering. *Inverse Probl.* **29**(9), 095018 (2013)
27. Yoo, J., Sabir, S., Heo, D., Kim, K.H., Wahab, A., Choi, Y., Lee, S.-I., Chae, E.Y., Kim, H.H., Baeb, Y.M., et al.: Deep learning diffuse optical tomography. *IEEE Trans Med Imaging* **39**(4), 877–887 (2020)
28. Kirsch, A.: Characterization of the shape of a scattering obstacle using the spectral data of the far field operator. *Inverse Probl.* **14**(6), 1489–1512 (1998)
29. Klibanov, M., Lucas, T., Frank, R.: New imaging algorithm in diffusion tomography. In: *Optical Tomography and Spectroscopy of Tissue: Theory, Instrumentation, Model, and Human Studies II*, Bellingham, WA, Society of Photo-Optical Instrumentation Engineers (1997)
30. Klibanov, M.V., Lucas, T.R., Frank, R.M.: A fast and accurate imaging algorithm in optical/diffusion tomography. *Inverse Probl.* **13**(5), 1341–1361 (1997)

31. Lee, O., Kim, J.M., Bresler, Y., Ye, J.C.: Compressive diffuse optical tomography: noniterative exact reconstruction using joint sparsity. *IEEE Trans. Med. Imaging* **30**(5), 1129–1142 (2011)
32. Lee, O., Ye, J.C.: Joint sparsity-driven non-iterative simultaneous reconstruction of absorption and scattering in diffuse optical tomography. *Opt. Express* **21**(22), 26589–604 (2013)
33. Oh, S., Milstein, A.B., Bouman, C.A., Webb, K.J.: Multigrid inversion algorithms with applications to optical diffusion tomography. In: Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA (2002)
34. Ren, W., Jiang, J., Mata, A.D.C., Kalyanov, A., Ripoll, J., Lindner, S., Charbon, E., Zhang, C., Rudin, M., Wolf, M.: Multimodal imaging combining time-domain near-infrared optical tomography and continuous-wave fluorescence molecular tomography. *Opt Express* **28**(7), 9860–9874 (2020)
35. Tamminen, J., de Hoop, M. V., Lassas, M., Siltanen, S.: D-bar method and exceptional points at positive energy: a computational study. In: Proceedings of the Project Review, Geo-Mathematical Imaging Group, 1 (2014)
36. Tamminen, J., Tarvainen, T., Siltanen, S.: The d-bar method for diffuse optical tomography: a computational study. *Exp. Math.* **26**(2), 225–240 (2017)
37. Yao, Y., Pei, Y., Wang, Y., Barbour, R.L.: Born type iterative method for imaging of heterogeneous scattering media and its application to simulated breast tissue. In: Proc. SPIE 2979, Optical Tomography and Spectroscopy of Tissue: Theory, Instrumentation, Model, and Human Studies II, 18 August (1997)
38. Ye, J.C., Bouman, C.A., Webb, K.J., Millane, R.P.: Nonlinear multigrid algorithms for Bayesian optical diffusion tomography. *IEEE Trans. Image Process.* **10**(6), 909–922 (2001)
39. Ye, J.C., Webb, K.J., Millane, R.P., Downar, T.J.: Modified distorted born iterative method with an approximate Fréchet derivative for optical diffusion tomography. *J. Opt. Soc. Am. A* **16**(7), 1814–1826 (1999)
40. Yedder, H.B., BenTaieb, A., Shokoufi, M., Zahiremami, A., Golnaraghi, F., Hamarneh, G.: Deep learning based image reconstruction for diffuse optical tomography. In: International Workshop on Machine Learning for Medical Image Reconstruction, pp. 112–119. Springer, Berlin (2018)
41. Yoo, J., Sabir, S., Heo, D., Kim, K.H., Wahab, A., Choi, Y., Lee, S.-I., Chae, E.Y., Kim, H.H., Bae, Y.M., et al.: Deep learning diffuse optical tomography. *IEEE Trans. Med. Imaging* **39**(4), 877–887 (2019)
42. Zacharopoulos, A., Schweiger, M., Kolehmainen, V., Arridge, S.: 3d shape based reconstruction of experimental data in diffuse optical tomography. *Opt. Express* **17**(21), 18940–18956 (2009)
43. Zacharopoulos, A.D., Arridge, S.R., Dorn, O., Kolehmainen, V., Sikora, J.: Three-dimensional reconstruction of shape and piecewise constant region values for optical tomography using spherical harmonic parametrization and a boundary element method. *Inverse Prob.* **22**(5), 1509–1532 (2006)
44. Zhang, L., Zhang, G.: Brief review on learning-based methods for optical tomography. *J. Innov. Opt. Health Sci.* **12**(06), 1930011 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.