

2018

# Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography

Lähivaara, Timo

Acoustical Society of America (ASA)

---

Tieteelliset aikakauslehtiartikkelit

In copyright 1.0

<http://dx.doi.org/10.1121/1.5024341>

---

<https://erepo.uef.fi/handle/123456789/6616>

*Downloaded from University of Eastern Finland's eRepository*

# Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography

Timo Lähivaara, Leo Kärkkäinen, Janne M. J. Huttunen, and Jan S. Hesthaven

Citation: [The Journal of the Acoustical Society of America](#) **143**, 1148 (2018); doi: 10.1121/1.5024341

View online: <https://doi.org/10.1121/1.5024341>

View Table of Contents: <http://asa.scitation.org/toc/jas/143/2>

Published by the [Acoustical Society of America](#)

---

## Articles you may be interested in

[Passive, broadband suppression of radiation of low-frequency sound](#)

The Journal of the Acoustical Society of America **143**, EL67 (2018); 10.1121/1.5022192

[Multiple scattering and scattering cross sections](#)

The Journal of the Acoustical Society of America **143**, 995 (2018); 10.1121/1.5024361

[Towards real-time assessment of anisotropic plate properties using elastic guided waves](#)

The Journal of the Acoustical Society of America **143**, 1138 (2018); 10.1121/1.5024353

[Effects of flow gradients on directional radiation of human voice](#)

The Journal of the Acoustical Society of America **143**, 1173 (2018); 10.1121/1.5025063

[Modeling analysis of ultrasonic attenuation and angular scattering measurements of suspended particles](#)

The Journal of the Acoustical Society of America **143**, 1049 (2018); 10.1121/1.5024233

[Time-domain sound field reproduction using the group Lasso](#)

The Journal of the Acoustical Society of America **143**, EL55 (2018); 10.1121/1.5022280

---

# Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography

Timo Lähivaara,<sup>1,a)</sup> Leo Kärkkäinen,<sup>2,b)</sup> Janne M. J. Huttunen,<sup>2,b)</sup> and Jan S. Hesthaven<sup>3</sup>

<sup>1</sup>*Department of Applied Physics, University of Eastern Finland, Kuopio, Finland*

<sup>2</sup>*Nokia Technologies, Espoo, Finland*

<sup>3</sup>*Computational Mathematics and Simulation Science, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*

(Received 27 September 2017; revised 4 January 2018; accepted 29 January 2018; published online 22 February 2018)

The feasibility of data based machine learning applied to ultrasound tomography is studied to estimate water-saturated porous material parameters. In this work, the data to train the neural networks is simulated by solving wave propagation in coupled poroviscoelastic-viscoelastic-acoustic media. As the forward model, a high-order discontinuous Galerkin method is considered, while deep convolutional neural networks are used to solve the parameter estimation problem. In the numerical experiment, the material porosity and tortuosity is estimated, while the remaining parameters which are of less interest are successfully marginalized in the neural networks-based inversion. Computational examples confirm the feasibility and accuracy of this approach.

© 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5024341>

[JFL]

Pages: 1148–1158

## I. INTRODUCTION

Measuring the porous properties of a medium is a demanding task. Different parameters are often measured by different application-specific methods, e.g., the porosity of rock is measured by weighing water-saturated samples and comparing their weight with dried samples. The flow resistivity of the damping materials can be computed from the pressure drop caused by the gas flowing through the material. In medical ultrasound studies, the porosity of the bone is determined indirectly by measuring the ultrasound attenuation as the wave passes through the bone.

For the porous material characterization, information carried by waves provides a potential way to estimate the corresponding material parameters. Ultrasound tomography (UST) is one technique that can be used for material characterization purposes. In this technique, an array of sensors is placed around the target. Typically, one of the sensors is acting as a source while others are receiving the data. By changing the sensor that acts as a source, a comprehensive set of wave data can be recorded which can be used to infer the material properties. For further details on the UST, we refer to Ref. 1, and references therein.

The theory of wave propagation in porous media was rigorously formulated in the 1950s and 1960s by Biot.<sup>2–5</sup> The model was first used to study the porous properties of bedrock in oil exploration. Since then, the model has been applied and further developed in a number of different fields.<sup>6–9</sup> The challenge of Biot's model is its computational complexity. The model produces several different types of waveforms, i.e., the fast and slow pressure waves and the shear wave, the computational simulation of which is a

demanding task even for modern supercomputers. Computational challenges further increase when attempting to solve inverse problems, as the forward model has to be evaluated several times.

In this work, we consider a process for the parameter estimation, comprising two sub-tasks: (1) forward model: the simulation of the wave fields for given parameter values and (2) inverse problem: estimation of the parameters from the measurements of wave fields. The inverse problem is solved using deep convolutional neural networks that provide a framework to solve the inverse problems: we can train a neural network as a model from wave fields to the parameters. Over the last decades, neural networks have been applied in various research areas such as image recognition,<sup>10,11</sup> cancer diagnosis,<sup>12,13</sup> forest inventory,<sup>14,15</sup> and groundwater resources.<sup>16,17</sup> Deep convolutional neural networks (CNNs) are a special type of deep neural networks<sup>11,18,19</sup> that employ convolutions instead of matrix multiplications (e.g., to reduce the number of unknowns). In this study, we employ deep convolutional neural networks to two-dimensional ultrasound data.

The structure of the rest of the paper is as follows. First, in Sec. II, we formulate the poroviscoelastic and viscoelastic models and describe the discontinuous Galerkin method. Then, in Sec. III, we describe the neural networks technique for the prediction of material parameters. Numerical experiments are presented Sec. IV. Finally, conclusions are given in Sec. V.

## A. Model justification

The purpose of this paper is to study the feasibility of using a data based machine learning approach in UST to characterize porous material parameters. In the synthetic model setup, we have a cylindrical water tank including an elastic shell layer. Ultrasound sensors are placed inside the water tank. In the model, we place a cylindrically shaped

<sup>a)</sup>Electronic mail: timo.lahivaara@uef.fi

<sup>b)</sup>Present address: Nokia Bell Labs, Espoo, Finland.

porous material sample in water and estimate its porosity and tortuosity from the corresponding ultrasound measurements by a convolutional neural network. Figure 1 shows a schematic drawing of a UST setup studied in this work.

The proposed model has a wide range of potential applications. Cylindrically shaped core samples can be taken of bedrock or manmade materials, such as ceramics or concrete, to investigate their properties. In addition, from the medical point of view, core samples can be taken, for example, of cartilage or bones for diagnosing purposes. Depending on the application, the model geometry and sample size together with sensor setup needs to be scaled.

## II. WAVE PROPAGATION MODEL

In this paper, we consider wave propagation in isotropic coupled poroviscoelastic-viscoelastic-acoustic media. In Secs. II A and II B, we follow Refs. 20–23 and formulate the poroviscoelastic and viscoelastic wave models. The discontinuous Galerkin method is discussed in Sec. II C.

### A. Biot's poroelastic wave equation

We use the theory of wave propagation in poroelastic media developed by Biot in Refs. 2–5. We express Biot equations in terms of solid displacement  $\mathbf{u}_s$  and the relative displacement of fluid  $\mathbf{w} = \phi(\mathbf{u}_f - \mathbf{u}_s)$ , where  $\phi$  is the porosity and  $\mathbf{u}_f$  is the fluid displacement. In the following,  $\rho_s$  and  $\rho_f$  denote the solid and fluid densities, respectively. We have

$$\rho_a \frac{\partial^2 \mathbf{u}_s}{\partial t^2} + \rho_f \frac{\partial^2 \mathbf{w}}{\partial t^2} = \nabla \cdot \mathbf{T}, \quad (1)$$

■ shell ■ fluid ■ porous sample  
sensors: × source ○ receivers

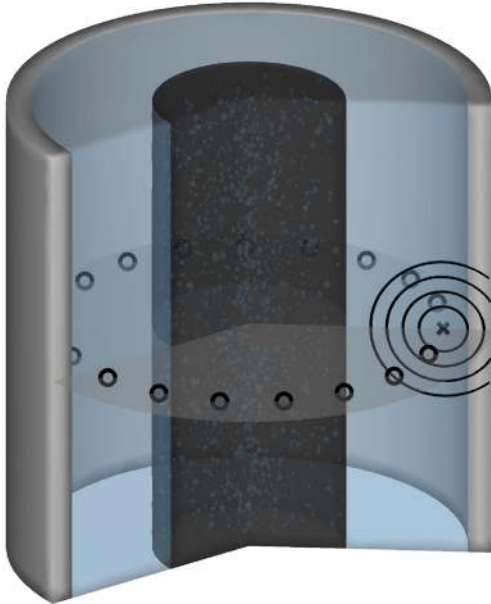


FIG. 1. (Color online) Schematic idea of the ultrasound tomography used to estimate porous material parameters. Acoustic wave generated by the source interact with the porous material which can be further seen on the data recovered by the receivers.

$$\rho_f \frac{\partial^2 \mathbf{u}_s}{\partial t^2} + \frac{\rho_f \tau}{\phi} \frac{\partial^2 \mathbf{w}}{\partial t^2} + \frac{\eta}{k} \frac{\partial \mathbf{w}}{\partial t} = \nabla \cdot \mathbf{T}_f, \quad (2)$$

where  $\rho_a = (1 - \phi)\rho_s + \phi\rho_f$  is the average density,  $\mathbf{T}$  is the total stress tensor, and  $\mathbf{T}_f$  is the fluid stress tensor. In Eq. (2),  $\tau$  is the tortuosity,  $\eta$  is the fluid viscosity, and  $k$  is the permeability.

The third term in Eq. (2) is a valid model at low frequencies, when the flow regime is laminar (Poiseuille flow). At high frequencies, inertial forces may dominate the flow regime. In this case, the attenuation model may be described in terms of viscous relaxation mechanics as discussed, for example, in Refs. 21 and 23. In this paper, we use the model derived in Ref. 23. The level of attenuation is controlled by quality factor  $Q_0$ .

One can express the stress tensors as

$$\mathbf{T} = 2\mu_{fr}\mathbf{E} + \left( \kappa_{fr} + \alpha^2 M - \frac{2}{3}\mu_{fr} \right) \text{tr}(\mathbf{E})\mathbf{I} - \alpha M \zeta \mathbf{I}, \quad (3)$$

$$\mathbf{T}_f = M(\alpha \text{tr}(\mathbf{E}) - \zeta)\mathbf{I}, \quad (4)$$

where  $\mu_{fr}$  is the frame shear modulus,  $\mathbf{E} = \frac{1}{2}(\nabla \mathbf{u}_s + (\nabla \mathbf{u}_s)^T)$  denote the solid strain tensor,  $\kappa_{fr}$  is the frame bulk modulus,  $\text{tr}(\cdot)$  is the trace,  $\mathbf{I}$  is the identity matrix, and  $\zeta = -\nabla \cdot \mathbf{w}$  is the variation of the fluid content. The effective stress constant  $\alpha$  and modulus  $M$ , given in Eqs. (3) and (4), can be written as  $\alpha = 1 - \kappa_{fr}/\kappa_s$ , where  $\kappa_s$  is the solid bulk modulus, and  $M = \kappa_s/(\alpha - \phi(1 - \kappa_s/\kappa_{fr}))$ , where  $\kappa$  is the fluid bulk modulus.

### B. Viscoelastic wave equation

The following discussion on the elastic wave equation with viscoelastic effects follows Carcione's book,<sup>20</sup> in which a detailed discussion can be found. Expressed as a second order system, the elastic wave equation can be written in the following form:

$$\rho_e \frac{\partial^2 \mathbf{u}_e}{\partial t^2} = \nabla \cdot \mathbf{S} + \mathbf{s}, \quad (5)$$

where  $\rho_e$  is the density,  $\mathbf{u}_e$  the elastic displacement,  $\mathbf{S}$  is a stress tensor, and  $\mathbf{s}$  is a volume source. In the two-dimensional viscoelastic (isotropic) case considered here, components of the solid stress tensor  $\mathbf{S}$  may be written as<sup>20</sup>

$$\begin{aligned} \sigma_{11} = & (\lambda_e + 2\mu_e)\epsilon_{11} + \lambda_e\epsilon_{22} + (\lambda_e + \mu_e) \sum_{\ell=1}^{L_e} \nu_1^{(\ell)} \\ & + 2\mu_e \sum_{\ell=1}^{L_e} \nu_{11}^{(\ell)}, \end{aligned} \quad (6)$$

$$\begin{aligned} \sigma_{22} = & (\lambda_e + 2\mu_e)\epsilon_{22} + \lambda_e\epsilon_{11} + (\lambda_e + \mu_e) \sum_{\ell=1}^{L_e} \nu_1^{(\ell)} \\ & - 2\mu_e \sum_{\ell=1}^{L_e} \nu_{11}^{(\ell)}, \end{aligned} \quad (7)$$

$$\sigma_{12} = 2\mu_e\epsilon_{12} + 2\mu_e \sum_{\ell=1}^{L_e} \nu_{12}^{(\ell)}, \quad (8)$$

where  $\mu_e$  and  $\lambda_e$  are the unrelaxed Lamé coefficients,  $\epsilon_{11}$ ,  $\epsilon_{22}$ , and  $\epsilon_{12}$  are the strain components, and  $L_e$  is the number of relaxation terms. The memory variables  $\nu_1^{(\ell)}$ ,  $\nu_{11}^{(\ell)}$ , and  $\nu_{12}^{(\ell)}$  satisfy

$$\frac{\partial \nu_1^{(\ell)}}{\partial t} = -\frac{\nu_1^{(\ell)}}{\tau_{\sigma\ell}^{(1)}} + \phi_{1\ell}(0)(\epsilon_{11} + \epsilon_{22}), \quad (9)$$

$$\frac{\partial \nu_{11}^{(\ell)}}{\partial t} = -\frac{\nu_{11}^{(\ell)}}{\tau_{\sigma\ell}^{(2)}} + \frac{\phi_{2\ell}(0)(\epsilon_{11} - \epsilon_{22})}{2}, \quad \ell = 1, \dots, L_e, \quad (10)$$

$$\frac{\partial \nu_{12}^{(\ell)}}{\partial t} = -\frac{\nu_{12}^{(\ell)}}{\tau_{\sigma\ell}^{(2)}} + \phi_{2\ell}(0)\epsilon_{12}, \quad (11)$$

where

$$\phi_{k\ell}(t) = \frac{1}{\tau_{\sigma\ell}^{(k)}} \left( 1 - \frac{\tau_{\epsilon\ell}^{(k)}}{\tau_{\sigma\ell}^{(k)}} \right) \left( \sum_{\ell=1}^n \frac{\tau_{\epsilon\ell}^{(k)}}{\tau_{\sigma\ell}^{(k)}} \right)^{-1} \times \exp(-t/\tau_{\sigma\ell}^{(k)}), \quad k = 1, 2. \quad (12)$$

In Eq. (12),  $\tau_{\epsilon\ell}^{(k)}$  and  $\tau_{\sigma\ell}^{(k)}$  are relaxation times corresponding to dilatational ( $k = 1$ ) and shear ( $k = 2$ ) attenuation mechanisms.

The acoustic wave equation can be obtained from the system by setting the Lamé coefficient  $\mu_e$  to zero.

### C. Discontinuous Galerkin method

Wave propagation in coupled poroviscoelastic-viscoelastic-acoustic media can be solved using the discontinuous Galerkin (DG) method (see, e.g., Refs. 22–26), which is a well-known numerical approach to numerically solve differential equations. The DG method has properties that makes it well-suited for wave simulations, e.g., the method can be effectively parallelized and it can handle complex geometries and, due to its discontinuous nature, large discontinuities in the material parameters. These are all properties that are essential features for the method to be used in complex wave problems. Our formulation follows Ref. 27, where a detailed account of the DG method can be found.

### III. ESTIMATING MATERIAL PARAMETERS BY NEURAL NETWORKS

The aim of this paper is to estimate porous material parameters by applying artificial neural networks trained to

simulated data. Compared to traditional inverse methods, neural networks have an advantage that they allow computationally efficient inferences. In other words, after the network has been learned, inferences can be carried out using the network without evaluating the forward model. Furthermore, the neural networks provide a straightforward approach to marginalize uninteresting parameters in the inference.

First we will give a brief summary to deep neural networks. For a wider representation of the topic, the reader is directed to the review articles by LeCun *et al.*<sup>11</sup> and Bengio,<sup>19</sup> or to the book by Buduma.<sup>28</sup>

We consider the following supervised learning task. We have a set of input data  $\{X_\ell\}$  with a labels  $\{Y_\ell\}$  ( $\ell = 1, \dots, N$ ). In our study, the input data comprises of the measured ultrasound fields considered as “images” such that rows correspond to temporal data and columns to receiver and the outputs are corresponding material parameters. The “image” should be interpreted as two dimensional data, not a traditional picture; there is no color mapping used in our algorithm. The aim is to find a function  $\Theta$  between the inputs and outputs,

$$Y = \Theta(X).$$

The task is to find a suitable form for the function and learn it from the given data. Contrary to traditional machine learning in which the features are pre-specified, deep learning has an advantage that features are learned from the data.

Neural networks are a widely used class of functions in machine learning. Figure 2 (left) shows an example of a neural network with one hidden layer. Each circular node represents an artificial neuron and an arrow represents a connection between the output of one neuron to the input of the other. This network has four inputs  $x_1, \dots, x_4$  (input layer), five hidden units with output activations  $a_1, \dots, a_5$  (hidden layer), and two outputs  $y_1$  and  $y_2$  (output layer). Neurons in the hidden layer are computational units that, for a given input  $(x_1, \dots, x_4)$ , calculate the activations as

$$a_j = \sigma \left( \sum_{i=1}^4 w_{ji}^{(1)} x_i + b_j^{(1)} \right), \quad j = 1, \dots, 5,$$

where  $w_{ji}^{(1)}$  is a weight assigned to the connection between the  $i$ th input,  $b_j^{(1)}$  is a bias term and  $j$ th activation, and  $\sigma$  is a non-linear function (nonlinearity). Similarly, the neurons at output layer are units for which the output is calculated as

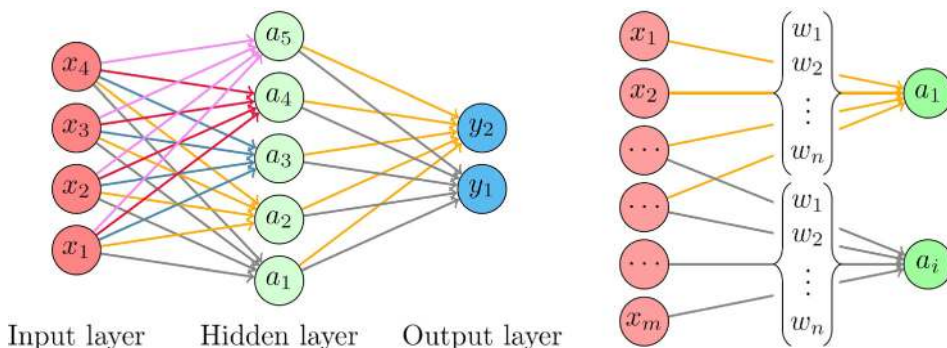


FIG. 2. (Color online) Left: Fully connected network with one hidden layer. Each arrow represents multiplication of the input or activation by a weight. Right: A convolutional layer.  $x_1, \dots, x_m$  are the inputs to the layers and  $a_1, \dots, a_i$  are the outputs of the layer. Note that the same set of the weights  $w_1, \dots, w_n$  are used for all outputs.

$$y_j = \sigma \left( \sum_{i=1}^5 w_{ji}^{(2)} a_i + b_j^{(2)} \right), \quad j = 1, \dots, 2,$$

where  $w_{ji}^{(2)}$  and  $b_j^{(2)}$  are again weights and biases.

At present, the most typically used nonlinearity is the rectified linear unit (ReLU),  $\sigma(x) = \max(0, x)$ .<sup>29</sup> Over the past decades, smoother nonlinearities such as the sigmoid function  $\sigma(x) = (1 + \exp(-x))^{-1}$  or tanh have been used as nonlinearity, but the ReLU typically allows better performance in the training of deep neural architectures on large and complex datasets.<sup>11</sup> Some layers can also have different linearities than others, but here we use the same  $\sigma$  for notational convenience.

It is easy to see that the neural network can be presented in matrix form as

$$A = \sigma(w^{(1)} \cdot X + b^{(1)}), \quad Y = \sigma(w^{(2)} \cdot A + b^{(2)}),$$

where  $X = (x_1, \dots, x_4)^T$ ,  $Y = (y_1, y_2)^T$ , and  $A = (a_1, \dots, a_5)^T$ .  $w^{(1)}$  and  $w^{(2)}$  are matrices/tensors comprising of the weights  $w_{ij}^{(1)}$  and  $w_{ij}^{(2)}$ , respectively,  $b^{(1)}$  and  $b^{(2)}$  are the vectors comprising of the bias variables  $b_i^{(1)}$  and  $b_i^{(2)}$ , respectively, and the function  $\sigma$  operates to the vectors element-wise. The network can be also written as a nested function  $\Theta_\theta(X) = \sigma(w^{(2)}(\sigma(w^{(1)} \cdot X + b^{(1)})) + b^{(2)})$ , where  $\theta$  represents the unknowns of the network [ $\theta = (w^{(1)}, b^{(1)}, w^{(2)}, b^{(2)})$ ].

Deeper networks can be constructed similarly. For example, a network with two hidden layers can be formed as  $A^{(1)} = \sigma(w^{(1)} \cdot X + b^{(1)})$ ,  $A^{(2)} = \sigma(w^{(2)} \cdot A^{(1)} + b^{(2)})$ , and  $Y = \sigma(w^{(3)} \cdot A^{(2)} + b^{(3)})$ , or  $\Theta_\theta(X) = \sigma(w^{(3)} \cdot \sigma(w^{(2)} \cdot (\sigma(w^{(1)} \cdot X + b^{(1)})) + b^{(2)}) + b^{(3)})$ , where  $\theta = (w^{(1)}, b^{(1)}, w^{(2)}, b^{(2)}, w^{(3)}, b^{(3)})$ .

The neural networks described above are called fully connected: there are connections between all nodes of the network. A disadvantage of such fully connected networks is that total number of free parameters can be enormous especially when the dimensions of input data is high (e.g., high resolution pictures) and/or the network has several hidden layers. CNNs were introduced<sup>18,30</sup> to overcome this problem. Figure 2 (right) shows an example of a 1D-convolutional layer. In a convolutional layer, the matrix multiplication is replaced with a convolution: for example, in 1D, the activations of the layer are calculated as

$$A = \sigma(w^{(1)} * X + B^{(1)}),$$

where the asterisk denotes the discrete 1D-convolution and  $w^{(1)}$  is a set of filter weights (a feature to be learned). In 2D, the input  $X$  can be considered as an image and the convolution  $w^{(1)} * X$  is two dimensional with a convolution matrix or mask  $w^{(1)}$  (a filtering kernel).

In a practical solution, however, only one convolution mask (per layer) is typically not enough for good performance. Therefore, each layer includes multiple masks that are applied simultaneously and the output of the convolutional layer is a bank (channels) of images. This three dimensional object forms the input of the next layer such that each

input channel has its own bank of filtering kernels and the convolution is effectively three dimensional.

Usually, convolutional neural networks also employ pooling<sup>10,31</sup> for down-sampling (to reduce dimension of activations). See, for example, LeCun *et al.*<sup>11</sup> or Buduma<sup>28</sup> for details.

The training of the neural networks is based on a set of training data  $\{X_\ell, Y_\ell\}$ . The purpose is to find a set of weights and biases that minimize the discrepancy between the outputs  $\{Y_\ell\}$  and the corresponding predicted values given by the neural networks  $\{\Theta_\theta(X_\ell)\}$ . Typically, in regression problems, this is achieved by minimizing the quadratic loss function  $f(\theta)$  over the simulation dataset

$$f(\theta) = f(\theta; \{X_\ell\}, \{Y_\ell\}) = \frac{1}{N_{nn}} \sum_{\ell=1}^{N_{nn}} (\Theta_\theta(X_\ell) - Y_\ell)^2 \quad (13)$$

to obtain the network parameters, weights, and biases, of the network. The optimization problem could be solved using gradient descent methods. The gradient can be calculated using back propagation which essentially is a procedure of applying the chain rule to the lost function in an iterative backward fashion.<sup>11,28,32</sup> The computation of the predictions  $\Theta_\theta(X_\ell)$  and its gradients is a computationally expensive task in the case of large training dataset. Therefore, during the iteration, the cost and gradients are calculated for mini-batches that are randomly chosen from the training set. This procedure is called stochastic gradient descent (SGD).<sup>18,28</sup>

The validation of the network is commonly carried out using a test set, which is either split from the original data or collected separately. This test set is used to carry final evaluation of the performance. This is done due to the tendency of deep networks for overfitting, which comes out as good performance in the training set but very poor performance in the test set (i.e., the network “remembers” the training samples instead of learning to generalize). Third set, the validation data set, is also traditionally used to evaluate performance during the training process.

#### IV. NUMERICAL EXPERIMENT

In this section, we present the results obtained from testing the data driven approach to estimate porous material parameters in ultrasound tomography. In this paper, initial conditions are always assumed to be zero and we apply the free condition as a boundary condition.

In the following results, time integration is carried out using an explicit low-storage Runge-Kutta scheme.<sup>33</sup> For each simulation, the length of the time step  $\Delta t$  is computed from

$$\Delta t = \left( \frac{h_{\min}^\ell}{2c_{\max}^\ell (N^\ell)^2} \right)_{\min}, \quad \ell = 1, \dots, K, \quad (14)$$

where  $c_{\max}^\ell$  is the maximum wave speed,  $N^\ell$  is the order of the polynomial basis,  $h_{\min}^\ell$  is the smallest distance between

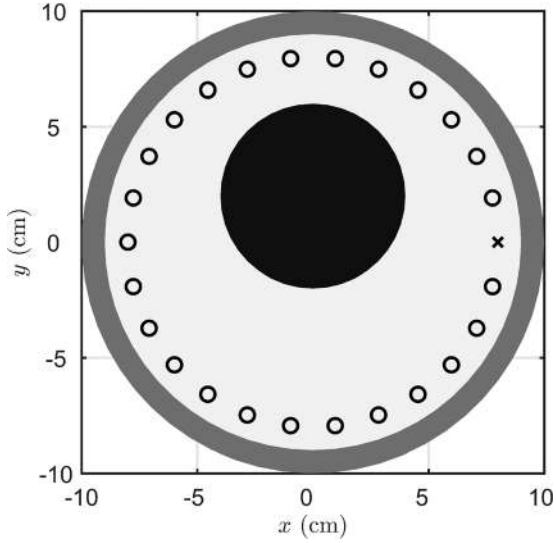


FIG. 3. Figure shows the problem geometry. In the graph, the circles denote the receivers and the cross denotes the source.

two vertices in the element  $\ell$ , and  $K$  is the number of elements.

### A. Model setup

Let us first introduce the model problem. Figure 3 shows the studied two dimensional problem geometry. The propagation medium contains three subdomains: a cylindrical shaped poroelastic inclusion (black), a fluid (light gray), and a solid shell (dark gray). The computational domain is a circle with radius of 10 cm with a 1 cm thick shell layer. A circular shaped poroelastic inclusion with a radius of 4 cm is vertically shifted to 2 cm from the center of the circular domain. Shifting the target from the center avoids symmetrically positioned sensors (with respect to  $x$  axis) to receive equal signal and therefore have more information from the target.

A total of 26 uniformly distributed ultrasound sensors are located at a distance of 8 cm from the center of the circle. The source is introduced on the strain components  $\varepsilon_{11}$  and  $\varepsilon_{22}$  by the first derivative of a Gaussian function with frequency  $f_0 = 40$  kHz, a time delay  $t_0 = 1.2/f_0$ . The sensor that is used as a source does not collect data in our simulations. Receivers collect solid velocity components  $u_s$  and  $v_s$ . In the following, the simulation time is 0.4 ms. Note that recorded

TABLE I. Table lists the minimum and maximum values used for uniform distributions for each unknown physical parameter.

Variable name	Symbol	Minimum	Maximum
Solid density	$\rho_s$ (kg/m <sup>3</sup> )	1000.0	5000.0
Solid bulk modulus	$\kappa_s$ (GPa)	15.0	70.0
Frame bulk modulus	$\kappa_{fr}$ (GPa)	5.0	20.0
Frame shear modulus	$\mu_{fr}$ (GPa)	3.0	14.0
Tortuosity	$\tau$	1.0	4.0
Porosity	$\phi$	0.01	0.99
Permeability	$k$ (m <sup>2</sup> )	$1.0 \times 10^{-10}$	$1.0 \times 10^{-7}$
Quality factor	$Q_0$	20.0	150.0

TABLE II. Table lists the relaxation times (two mechanisms) used with the shell layer. Relaxation times are computed using the nonlinear optimization method discussed in detail in Ref. 34.

	$\tau_{e\ell}^{(1)}$	$\tau_{\sigma\ell}^{(1)}$	$\tau_{e\ell}^{(2)}$	$\tau_{\sigma\ell}^{(2)}$
$\ell = 1$	$1.239 \times 10^{-4}$	$1.176 \times 10^{-4}$	$1.249 \times 10^{-4}$	$1.165 \times 10^{-4}$
$\ell = 2$	$5.319 \times 10^{-6}$	$5.042 \times 10^{-6}$	$5.370 \times 10^{-6}$	$4.999 \times 10^{-6}$

data is downsampled to a sampling frequency of 800 kHz on each receiver.

The inclusion is fully saturated with water. The fluid parameters are: the density  $\rho_f = 1020$  kg/m<sup>3</sup>, the fluid bulk modulus  $\kappa_f = 2.295$  GPa, and the viscosity  $\eta = 1.0 \times 10^{-3}$  Pa s. All other material parameters of the inclusion are assumed to be unknown. In this paper, we assume a relatively wide range of possible parameter combinations, see Table I. Furthermore, the unknown parameters are assumed to be uncorrelated. The physical parameter space gives  $\sim 1.5$  kHz as an upper bound for Biot's characteristic frequency

$$f_c = \frac{\eta\phi}{2\pi\tau\rho_f k} \quad (15)$$

and hence we operate in Biot's high-frequency regime ( $f_c < f_0$ ) in all possible parameter combinations.

For the fluid subdomain (water), we set the density  $\rho_e = 1020$  kg/m<sup>3</sup>, the first Lamé parameter  $\lambda_e = 2.295$  GPa, and the second Lamé parameter  $\mu_e = 0$ . The elastic shell layer has the following parameters:  $\rho_e = 2000$  kg/m<sup>3</sup>,  $\lambda_e = 12.940$  GPa, and  $\mu_e = 5.78$  GPa. The relaxation times for the viscoelastic attenuation in the shell layer are given in Table II.

The derived wave speeds for each subdomain are given in Table III. For the poroelastic inclusion, both the minimum and maximum wave speeds are reported. It should be noted that the reported values for the wave speeds in the inclusion correspond to the values generated by sampling the material parameters and hence they may not correspond to the global maximum and minimum values. A detailed approach for calculating wave speeds is given in Ref. 23.

It should be noted that, depending on the application, only some of the material parameters are unknowns of interest. In this work, we focus on estimating the porosity and tortuosity of the inclusion while the solid density and bulk modulus, frame bulk and shear modulus, permeability, and quality factor are marginalized in the neural networks-based inversion algorithm, discussed in detail below.

### B. Training, validation, and test data

For the convolutional neural networks algorithm used in this work, the recorded wave data is as expressed as images

TABLE III. Derived wave speeds for each subdomain. For the inclusion, both the minimum/maximum values are given. For the inclusion, values are based on sampling the material parameters.

Subdomain	$c_p^I$ (m/s)	$c_p^{II}$ (m/s)	$c_s$ (m/s)
Inclusion	1900/20189	196/1983	833/11578
Fluid	1500	—	—
Shell	3500	—	1700

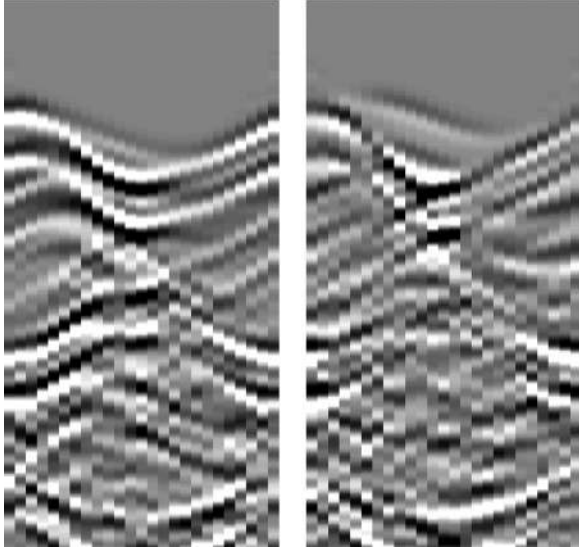


FIG. 4. Example images used for training the neural networks algorithm. Note that neural networks algorithm uses the original pixel values of the image  $X$  [see Eq. (16)]. Time is shown on the vertical-axis and receiver index on the horizontal-axis. Images correspond to samples where the shear wave speeds  $c_s$  is minimum (left) or maximum (right) (see Table III).

$X \in \mathbb{R}^d$ ,  $d = N_t \times N_r$ , where  $N_t$  denotes the number of time steps and  $N_r$  the number of receivers. The input dimension is  $d = 320 \times 25$ , as the data  $X$  can be seen as a 2D-image, comprising 25 pixels, corresponding to the receiver positions, times 320 pixels corresponding to the time evolution of the signal.

As the data on each column of the image, we use the following:

$$X(:, \ell) = \frac{x_r^\ell (u_s^\ell - u_{s,noi}^\ell) + y_r^\ell (v_s^\ell - v_{s,noi}^\ell)}{\sqrt{(x_r^\ell)^2 + (y_r^\ell)^2}}, \quad \ell = 1, \dots, N_r, \quad (16)$$

where  $(x_r^\ell, y_r^\ell)$  are the coordinates of the  $\ell$  th receiver and  $(u_s^\ell, v_s^\ell)$  are the horizontal and vertical velocity components on the  $\ell$  th receiver. In Eq. (16) subscript *noi* denotes the data that is simulated without the porous inclusion. Two example images are shown in Fig. 4.

We have generated a training data set comprising 15 000 samples using computational grids that have  $\sim 3$

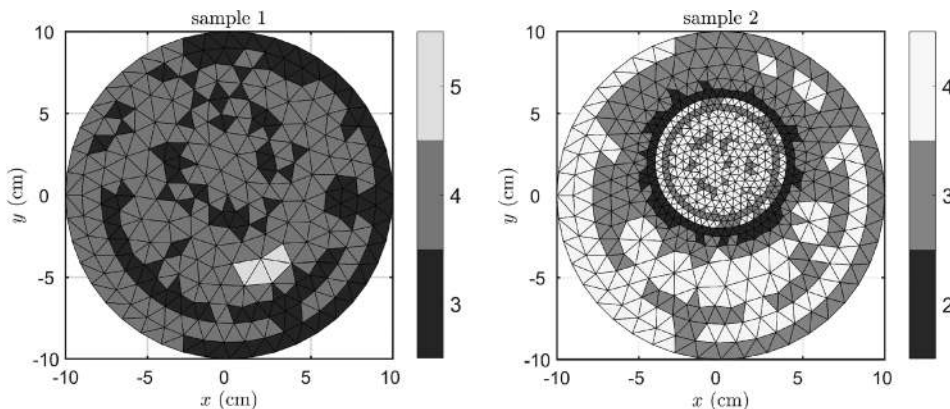


FIG. 5. Two example grids used in the computations. The colorbar shows the order of the basis functions.

TABLE IV. The convolutional neural network architecture used in this work. The convolutional layers uses periodic convolution (padding) in the receiver position-direction and no padding in the temporal direction. The total number of unknowns in the network is 3.38 M.

Layer k	Type and non-linearity	Input size	Output size
	Input	$320 \times 25$	$320 \times 25 \times 1$
1	Convolution layer ( $5 \times 3 \times 1$ filter, 20 filters) + Layer normalization + ReLU + Max-pooling ( $2 \times 2$ )	$320 \times 25 \times 1$	$316 \times 25 \times 20$
2	Convolution layer ( $5 \times 3 \times 20$ filter, 25 filters) + Layer normalization + ReLU + Max-pooling ( $2 \times 2$ )	$158 \times 13 \times 20$	$154 \times 13 \times 25$
3	Vectorization Fully connected layer + Layer normalization + ReLU	$77 \times 7 \times 25$ 13475	$77 \times 7 \times 25$ 13475 250
4	Fully connected layer	250	1
	Output		1

elements per wavelength. The physical parameters for each sample are drawn from the uniform distribution (bounds given in Table I). The order of the basis functions is selected separately for each element of the grid. The order  $N_\ell$  of the basis function in element  $\ell$  is defined by

$$N_\ell = \left\lceil \frac{2\pi a h_{\max}^\ell}{\lambda_\ell^w} + b \right\rceil, \quad (17)$$

where  $\lambda_\ell^w = c_{\min}^\ell / f_0$  is the wavelength,  $c_{\min}^\ell$  is the minimum wave speed, and  $\lceil \cdot \rceil$  is the ceiling function. The parameters  $a$  and  $b$  control the local accuracy on each element. Following Refs. 35 and 36, we set  $(a, b) = (1.0294, 0.7857)$ .

Figure 5 shows two examples of computational grids and the corresponding basis order selection on each triangle. Example grids consist of 466 elements and 256 vertices ( $h_{\min} = 0.81$  cm and  $h_{\max} = 1.95$  cm) (sample 1) and 1156 elements and 601 vertices ( $h_{\min} = 0.32$  cm and  $h_{\max} = 1.75$  cm) (sample 2).

Figure 6 shows two snapshots of the scattered solid velocity field ( $\sqrt{(u_s - u_{s,noi})^2 + (v_s - v_{s,noi})^2}$ ) for two time instants. At the first time instant, the transmitted fast pressure wave and also the first reflected wave front is clearly visible. At the second time, all wave components have reflected back from the left surface of the phantom. Furthermore, more



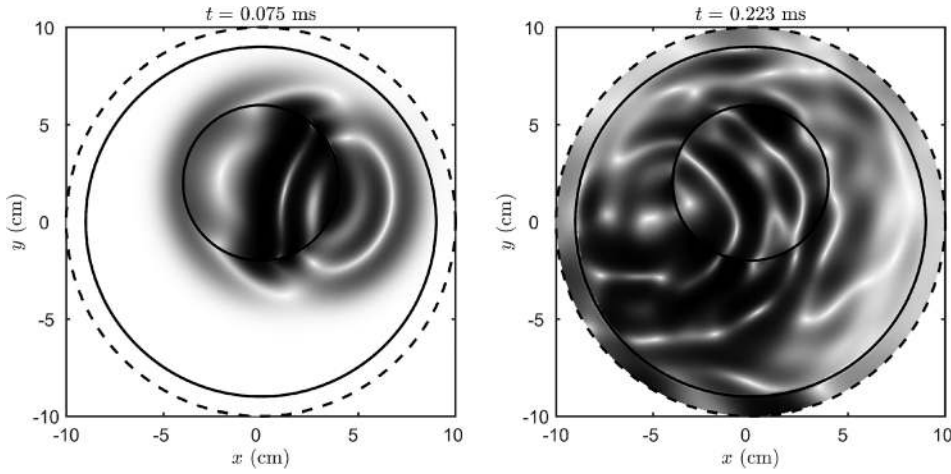


FIG. 6. Wave interaction with the poroelastic material at two times. The title shows the time. In the graphs, solid black lines show the inclusion/water and water/shell interfaces while the dotted black line shows the exterior boundary.

complicated wave scattering patterns can be seen inside the porous inclusion. These wave fields demonstrate how the received signals are obtained as combinations of multiple wave fronts.

To include observation noise in the training, each image in the simulation set is copied five times and each of the copies is corrupted with Gaussian noise of the form

$$X_{\ell}^{\text{noised}} = X_{\ell} + A\epsilon^A + B|X_{\ell}|\epsilon^B, \quad (18)$$

where  $\epsilon^A$  and  $\epsilon^B$  are independent zero-mean identically distributed Gaussian random variables. The second term represents additive white noise and the last term represents noise relative to the signal strength. To represent a wide range of different noise levels, for each sample image, the coefficients  $A$  and  $B$  are randomly chosen such that the standard deviations of the white noise component is between 0.03%–5% (varying logarithmically), and the standard deviations of the relative component is between 0%–5%. The total number of samples in the training set is  $N_{mn} = 5 \times 15\,000 = 75\,000$ .

Furthermore, two additional data sets were generated: a validation data set and a test set, that both comprise 3000 samples. In machine learning, the validation data set is traditionally used to evaluate performance during the training process and the test set is used for the final evaluation of the network.

These data sets are generated similarly to the training set, except computational grids were required to have  $\sim 4$  elements per wavelength to avoid inverse crime.<sup>37</sup> Furthermore, for the test set, the non-uniform basis order parameters are  $(a, b) = (1.2768, 1.4384)$ <sup>35</sup> [Eq. (17)] and the noise is added in a more systematic manner (instead of choosing  $A$  and  $B$  randomly) to study the performance with different noise levels (see Sec. IV D).

### C. Convolutional neural networks architecture

The neural network architecture is shown in Table IV and Fig. 7. The CNN architecture is similar to ones used for image classification, for example, ALEXNET,<sup>10</sup> with some exceptions. In our problem, the input data can be considered as one channel images instead of color images with three channels (i.e., we do not apply any color mapping). Our network also lacks the SOFTMAX layer, which is used as an outmost layer to provide the classification. Instead, the outmost two layers are simply fully connected layers. In addition, our network also has a smaller number of convolutional and fully connected layers with smaller dimensions in filter banks, which leads to significantly smaller number of unknowns. We wish, however, to note that the purpose was not to find the most optimal network architecture. For example, there can be other architectures that

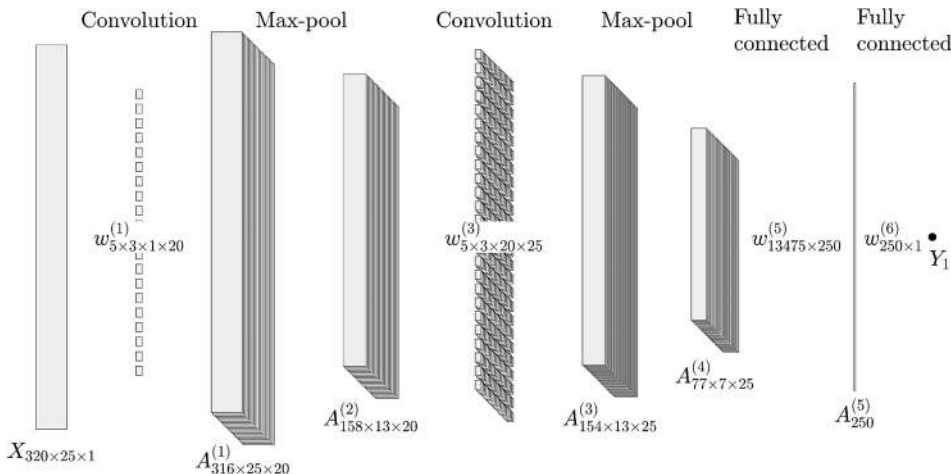


FIG. 7. A flow diagram of the network used in the study. Flow is from left to right (connecting arrows are not shown for graphical clarity). The sizes of the boxes scale with the memory size of the elements. The size/structure of the neurons and weights are also displayed in the subscript of  $A^{(i)}$  and  $w^{(i)}$  for each layer  $i$ . The matrices of the weights of the full connected layers  $w^5$  and  $w^6$  are not drawn for graphical clarity.

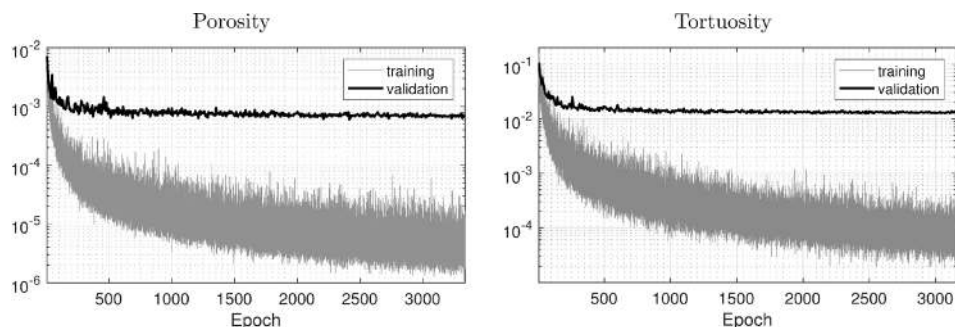


FIG. 8. Training and validation loss for porosity (left) and tortuosity (right) as function of number of epochs (full training cycles in stochastic optimization).

can provide similar performance with an even smaller number of unknowns. Furthermore, similar performance can also be achieved with fully connected networks with  $\sim 3$  layers, but at the expense of significantly larger number of unknowns.

The loss function is chosen to be quadratic [Eq. (13)]. The implementation is carried out using `TENSORFLOW`, which is a `PYTHON` toolbox for machine learning. The optimization was carried out with the `ADAM` optimizer.<sup>38</sup> The batch size for the stochastic optimization is chosen to be 50 samples.

#### D. Results

Figure 8 shows the loss of the training and validation data. The loss is shown for the two unknown parameters of interest, e.g., porosity and tortuosity. In both cases, we observe that the network has practically reached its generalization capability at least after 2000 full training cycles. The accuracy of the network is affected by the marginalization over all other parameters. In principle, the effect of the other parameters could compensate the changes that using, for example, a different porosity would cause, leaving the waveform of the measurement intact.

We have applied the trained network to predict porosity and tortuosity from images of the test that are corrupted with

the white noise component with low noise level (Fig. 9), moderate noise level (Fig. 10), and high noise level (Fig. 11). The figures also include error histograms. Table V shows statistics of the prediction error for different noise levels. Figure 12 shows the maximum absolute error and the root-mean-square error [the square root of Eq. (13)] as a function of the noise level in white noise component. The predictions are slightly positively biased with smaller noise levels, but positive bias is diminished with higher noise levels. Such a behavior might be due to the discretization error in forward models (the simulation of training and testing data were carried out using different levels of discretizations) which may dominate with lower observation noise levels but becomes negligible with higher noise levels. On the contrary, with high levels of noise, the predictions are negatively biased especially for larger values of porosity and tortuosity.

We have also studied the effect of the relative noise [the third term in Eq. (18)] to the results. Figure 13 shows the maximum absolute error and the root-mean-square error as a function of the noise level in the relative component. As we can see, the predictions are almost unaffected by the relative error even with significantly high noise levels ( $\sim 5\% - 10\%$ ).

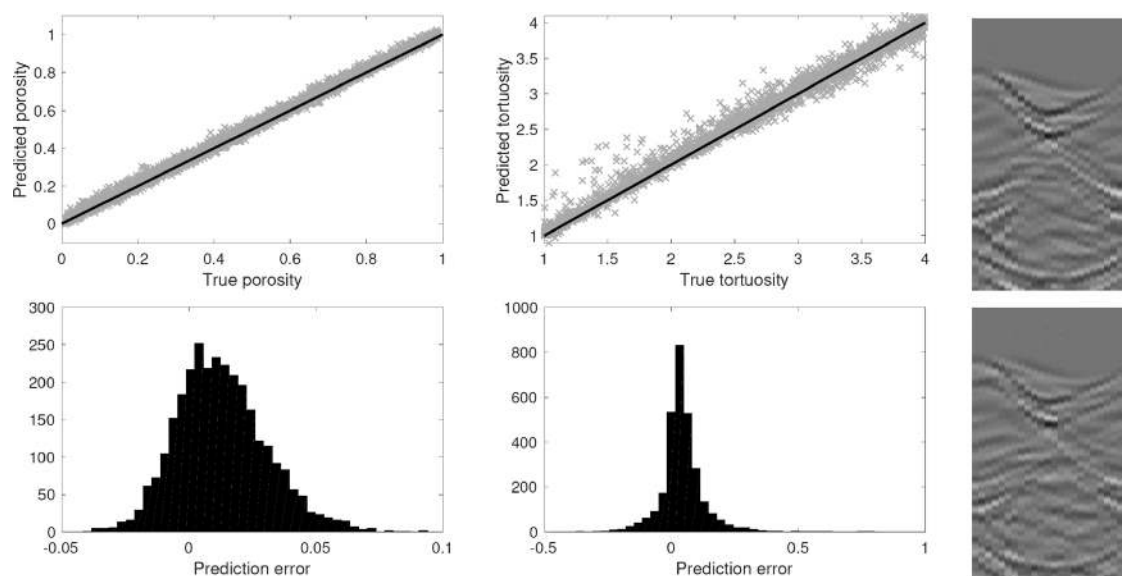


FIG. 9. Predicted porosities (left) and tortuosities (middle) for the test data with white noise of 0.8% (relative to the maximum absolute value of the signal in the training set). Bottom row (left and middle) shows histograms of the prediction error (difference between the predicted and true values). The right column shows two examples of the sample images in this noise level.

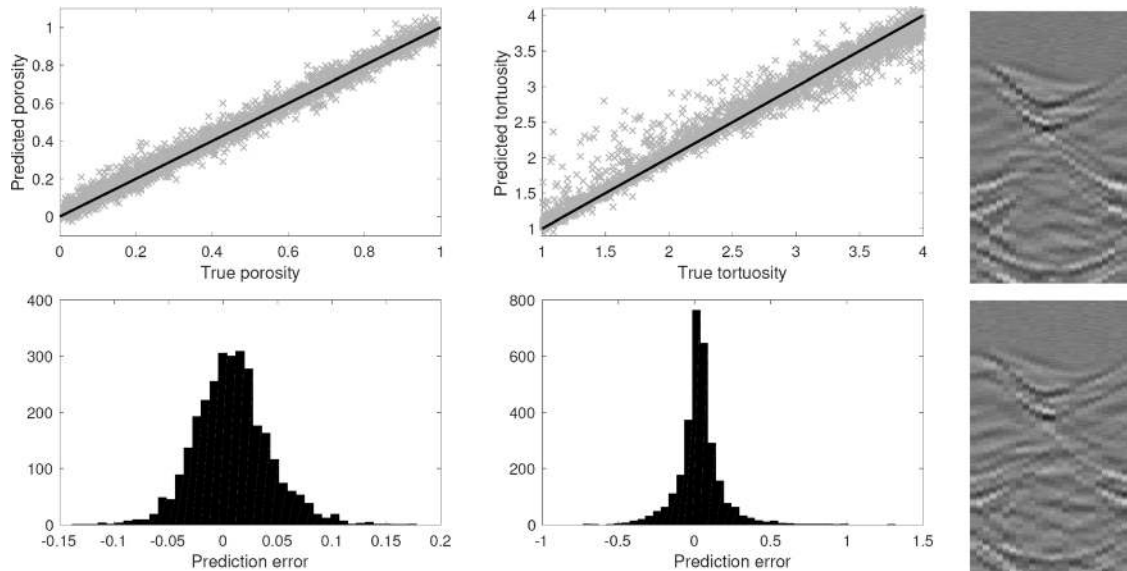


FIG. 10. Predicted porosities (left) and tortuosities (middle) for the test data with white noise of 3.2% (relative to the maximum absolute value of the signal in the training set). Bottom row (left and middle) shows histograms of the prediction error (difference between the predicted and true values). The right column shows two examples of the sample images in this noise level.

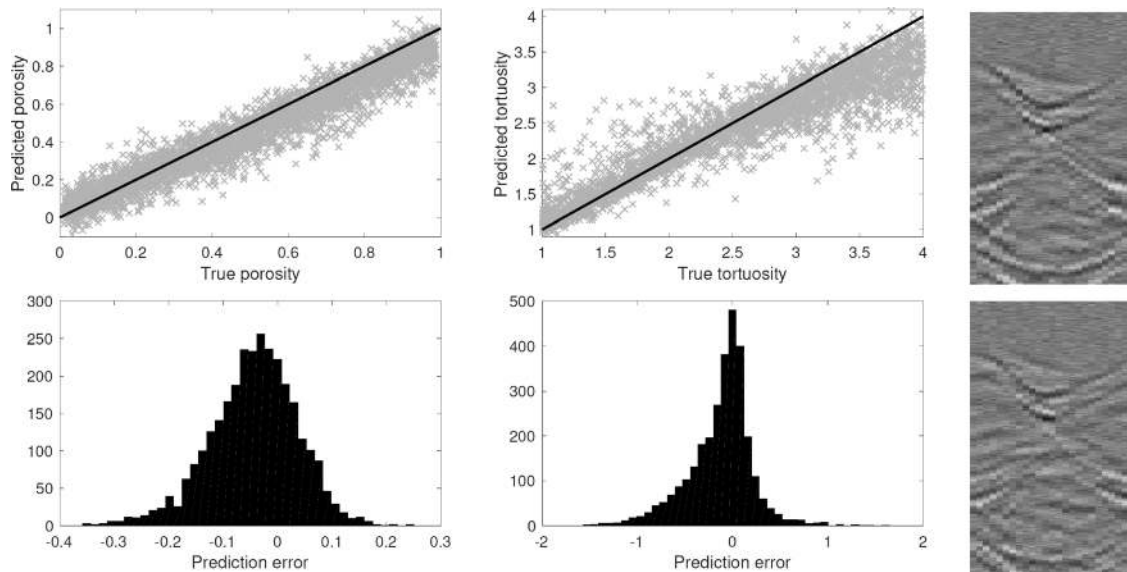


FIG. 11. Predicted porosities (left) and tortuosities (middle) for the test data with white noise of 8.1% (relative to the maximum absolute value of the signal in the training set). Bottom row (left and middle) shows histograms of the prediction error (difference between the predicted and true values). The right column shows two examples of the sample images in this noise level.

TABLE V. Bias, standard deviation, kurtosis, and percentiles (25% and 75%) for the errors of the predicted porosities and tortuosities with different noise levels.

Noise level	Porosity					Tortuosity				
	Bias	S.D.	Kurtosis	25%	75%	Bias	S.D.	Kurtosis	25%	75%
0%	0.012	0.016	3.2	0.0011	0.023	0.041	0.087	14	0.0092	0.07
0.16%	0.012	0.016	3.2	0.0012	0.023	0.042	0.087	14	0.0097	0.07
0.49%	0.013	0.017	3.3	0.0013	0.023	0.044	0.091	15	0.0087	0.073
0.81%	0.013	0.018	3.5	0.00078	0.024	0.046	0.096	15	0.0076	0.075
1.6%	0.013	0.023	4.1	-0.002	0.026	0.047	0.11	14	0.0016	0.08
3.2%	0.0087	0.035	4.4	-0.014	0.028	0.035	0.16	11	-0.026	0.085
4.9%	-0.0029	0.049	4.2	-0.034	0.027	-0.0012	0.22	8.1	-0.085	0.082
8.1%	-0.043	0.081	3.6	-0.093	0.0099	-0.12	0.34	5.1	-0.29	0.062

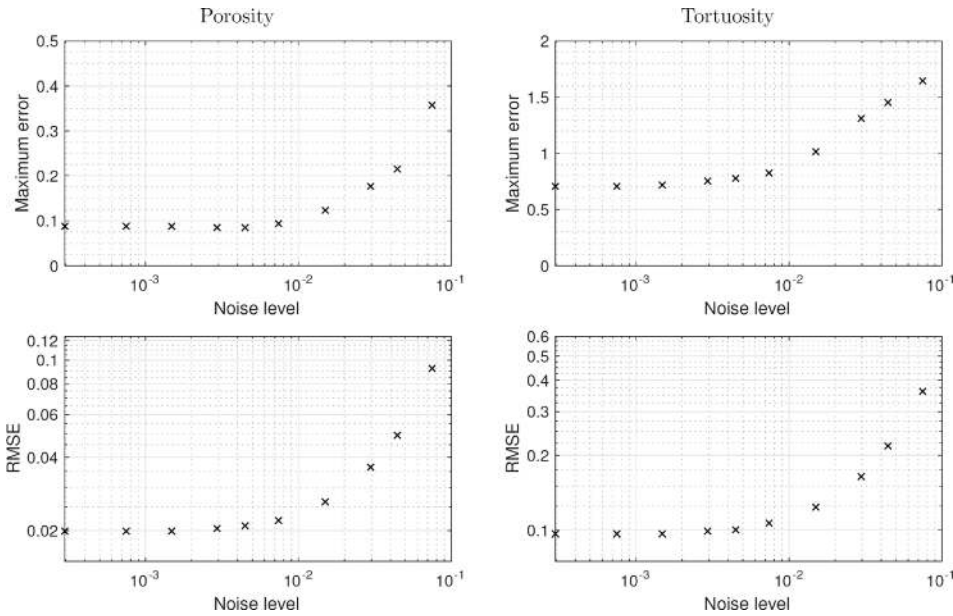


FIG. 12. The maximum absolute errors and the root-mean-square errors (RMSE) as a function of the noise level. The  $x$  axis is the noise level relative to the maximum of the (noise-free) signal in the training set.

## V. CONCLUSIONS

In this paper, we proposed the use of CNNs for estimation of porous material parameters from synthetic ultrasound tomography data. In the studied model, ultrasound data were generated in a water tank into which the poroelastic material sample was placed. A total of 26 ultrasound sensors were positioned in the water. One of the sensors generated the source pulse while others were used in the receiving mode. The recorded velocity data were represented as images which were further used as an input to the CNNs.

In the experiment, the parameter space for the porous inclusion was assumed to be large. For example, the porosity of the inclusion was allowed to span the interval from 1% to 99%. The selected parameter space models a different type of materials. We estimated the porosity and tortuosity of the porous material sample while all other material parameters were considered as nuisance parameters (see Table I). Based

on the results, it seems that these parameters can be estimated with acceptable accuracy with a wide variety of noise levels, while the nuisance parameters are successfully marginalized. The error histograms for both porosity and tortuosity show excellent accuracy in terms of root-mean-square error and bias.

We have marginalized our inference of porosity and tortuosity over six other material parameters (listed in Table I), which makes it possible to detect the primary porous material parameters from the waveforms without the knowledge of the values of the nuisance parameters, even if they can have a significant impact on the waveforms themselves. The success in the marginalization significantly increases the potential of the neural networks for material characterization.

Future studies should include more comprehensive investigation of model uncertainties, including geometrical inaccuracies (positioning and size of the material sample),

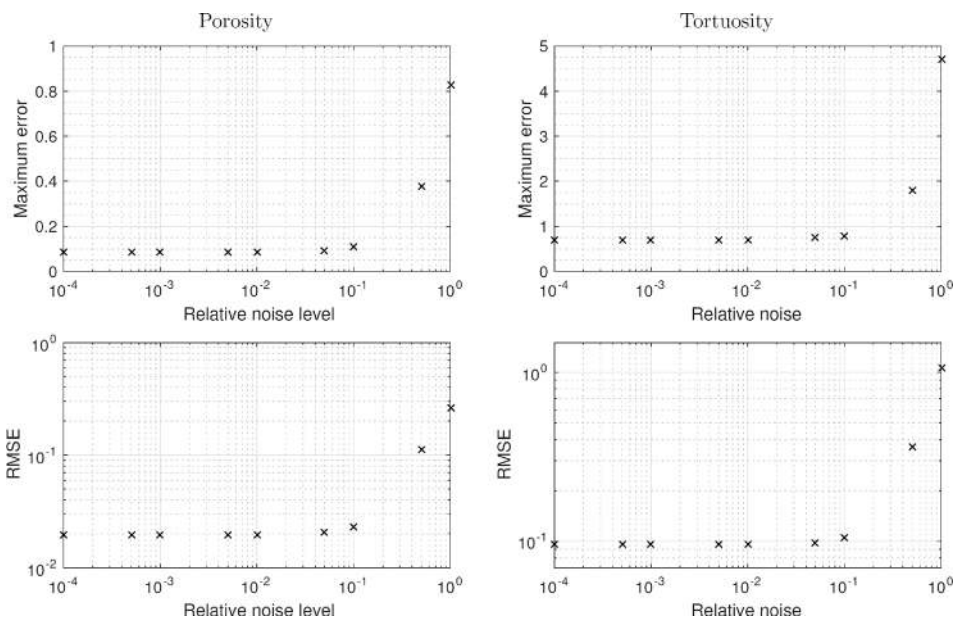


FIG. 13. The maximum absolute errors and the RMSE as a function of the relative noise  $[B$  in Eq. (18)].

fluid parameter changes, viscoelastic parameters of the shell layer, material inhomogeneities, and the sensor setup. In addition, the extension to the three spatial dimensions together with actual measurements are essential steps to guarantee the effectiveness of the proposed method.

## ACKNOWLEDGMENTS

This work has been supported by the strategic funding of the University of Eastern Finland and by the Academy of Finland (Project No. 250215, Finnish Centre of Excellence in Inverse Problems Research). This article is based upon work from COST Action DENORMS CA-15125, supported by COST (European Cooperation in Science and Technology).

- <sup>1</sup>N. Duric, P. J. Littrup, C. Li, O. Roy, and S. Schmidt, "Ultrasound tomography: A decade-long journey from the laboratory to clinic," in *Ultrasound Imaging and Therapy*, edited by A. Karellas and B. R. Thomadsen (CRC Press, Boca Raton, FL, 2015).
- <sup>2</sup>M. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid. I. Low frequency range," *J. Acoust. Soc. Am.* **28**(2), 168–178 (1956).
- <sup>3</sup>M. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid. II. Higher frequency range," *J. Acoust. Soc. Am.* **28**(2), 179–191 (1956).
- <sup>4</sup>M. Biot, "Mechanics of deformation and acoustic propagation in porous media," *J. Appl. Phys.* **33**(4), 1482–1498 (1962).
- <sup>5</sup>M. Biot, "Generalized theory of acoustic propagation in porous dissipative media," *J. Acoust. Soc. Am.* **34**(5), 1254–1264 (1962).
- <sup>6</sup>N. Sebaa, Z. Fellah, M. Fellah, E. Ogam, A. Wirgin, F. Mitri, C. Depollier, and W. Lauriks, "Ultrasonic characterization of human cancellous bone using the Biot theory: Inverse problem," *J. Acoust. Soc. Am.* **120**(4), 1816–1824 (2006).
- <sup>7</sup>M. Yvonne Ou, "On reconstruction of dynamic permeability and tortuosity from data at distinct frequencies," *Inv. Probl.* **30**(9), 095002 (2014).
- <sup>8</sup>T. Lähivaara, N. Dudley Ward, T. Huttunen, Z. Rawlinson, and J. Kaipio, "Estimation of aquifer dimensions from passive seismic signals in the presence of material and source uncertainties," *Geophys. J. Int.* **200**, 1662–1675 (2015).
- <sup>9</sup>J. Parra Martinez, O. Dazel, P. Göransson, and J. Cuenca, "Acoustic analysis of anisotropic poroelastic multilayered systems," *J. Appl. Phys.* **119**(8), 084907 (2016).
- <sup>10</sup>A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Red Hook, NY, 2012), pp. 1097–1105.
- <sup>11</sup>Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
- <sup>12</sup>P. S. Maclin, J. Dempsey, J. Brooks, and J. Rand, "Using neural networks to diagnose cancer," *J. Med. Syst.* **15**(1), 11–19 (1991).
- <sup>13</sup>L. A. Menendez, F. J. de Cos Juez, F. S. Lasheras, and J. A. Riesgo, "Artificial neural networks applied to cancer detection in a breast screening programme," *Math. Comput. Model.* **52**(7–8), 983–991 (2010).
- <sup>14</sup>P. Muukkonen and J. Heiskanen, "Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data," *Remote Sens. Environ.* **99**(4), 434–447 (2005).
- <sup>15</sup>H. Niska, J.-P. Skön, P. Packalen, T. Tokola, M. Maltamo, and M. Kolehmainen, "Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs," *IEEE Trans. Geosci. Remote Sens.* **48**(3), 1076–1085 (2009).
- <sup>16</sup>I. N. Daliakopoulos, P. Coulibaly, and I. K. Tsanis, "Groundwater level forecasting using artificial neural networks," *J. Hydrol.* **309**(1–4), 229–240 (2005).
- <sup>17</sup>S. Mohanty, K. Jha, A. Kumar, and K. P. Sudheer, "Artificial neural network modeling for groundwater level forecasting in a River Island of Eastern India," *Water Resour. Manag.* **24**(9), 1845–1865 (2010).
- <sup>18</sup>Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).
- <sup>19</sup>Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.* **2**(1), 1–127 (2009).
- <sup>20</sup>J. Carcione, *Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic and Porous Media* (Elsevier, Amsterdam, 2001).
- <sup>21</sup>C. Morency and J. Tromp, "Spectral-element simulations of wave propagation in porous media," *Geophys. J. Int.* **175**(1), 301–345 (2008).
- <sup>22</sup>L. Wilcox, G. Stadler, C. Burstedde, and O. Ghattas, "A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media," *J. Comput. Phys.* **229**, 9373–9396 (2010).
- <sup>23</sup>N. Dudley Ward, T. Lähivaara, and S. Eveson, "A discontinuous Galerkin method for poroelastic wave propagation: Two-dimensional case," *J. Comput. Phys.* **350**, 690–727 (2017).
- <sup>24</sup>M. Käser and M. Dumbser, "An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—I. The two-dimensional isotropic case with external source terms," *Geophys. J. Int.* **166**(23), 855–877 (2006).
- <sup>25</sup>J. de la Puente, M. Dumbser, M. Käser, and H. Igel, "Discontinuous Galerkin methods for wave propagation in poroelastic media," *Geophysics* **73**(5), T77–T97 (2008).
- <sup>26</sup>G. Gabard and O. Dazel, "A discontinuous Galerkin method with plane waves for sound-absorbing materials," *Int. J. Numer. Meth. Eng.* **104**(12), 1115–1138 (2015).
- <sup>27</sup>J. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications* (Springer, Berlin, 2007).
- <sup>28</sup>N. Buduma, *Fundamentals of Deep Learning, Designing Next-Generation Machine Intelligence Algorithms* (O'Reilly Media, Sebastopol, CA, 2017).
- <sup>29</sup>R. Hahnloser, M. A. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature* **405**, 947–951 (2000).
- <sup>30</sup>K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.* **36**, 193–202 (1980).
- <sup>31</sup>D. Scherer, A. Müller, and S. Behnke, *Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition* (Springer, Berlin, 2010), pp. 92–101.
- <sup>32</sup>S. Linnainmaa, "The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors" (in Finnish), Master's thesis, University of Helsinki, (1970), <http://www.idsia.ch/~juergen/linnainmaa1970thesis.pdf> (Last viewed February 9, 2018).
- <sup>33</sup>M. Carpenter and C. Kennedy, "Fourth-order 2N-storage Runge-Kutta schemes," Technical Report No. NASA-TM-109112 (1994).
- <sup>34</sup>E. Blanc, D. Komatitsch, E. Chaljub, B. Lombard, and Z. Xie, "Highly accurate stability-preserving optimization of the Zener viscoelastic model, with application to wave propagation in the presence of strong attenuation," *Geophys. J. Int.* **205**(1), 427–439 (2016).
- <sup>35</sup>T. Lähivaara and T. Huttunen, "A non-uniform basis order for the discontinuous Galerkin method of the acoustic and elastic wave equations," *Appl. Numer. Math.* **61**, 473–486 (2011).
- <sup>36</sup>T. Lähivaara and T. Huttunen, "A non-uniform basis order for the discontinuous Galerkin method of the 3D dissipative wave equation with perfectly matched layer," *J. Comput. Phys.* **229**, 5144–5160 (2010).
- <sup>37</sup>J. Kaipio and E. Somersalo, "Statistical inverse problems: Discretization, model reduction and inverse crimes," *J. Comput. Appl. Math.* **198**(2), 493–504 (2007).
- <sup>38</sup>D. Kingma and J. Ba, "Adam: A method for stochastic optimization," ArXiv e-print (2014).