

# A model-based iterative learning approach for diffuse optical tomography

Meghdoot Mozumder, Andreas Hauptmann, *Member, IEEE*, Ilkka Nissilä, Simon R. Arridge, and Tanja Tarvainen

**Abstract**—Diffuse optical tomography (DOT) utilises near-infrared light for imaging spatially distributed optical parameters, typically the absorption and scattering coefficients. The image reconstruction problem of DOT is an ill-posed inverse problem, due to the non-linear light propagation in tissues and limited boundary measurements. The ill-posedness means that the image reconstruction is sensitive to measurement and modelling errors. The Bayesian approach for the inverse problem of DOT offers the possibility of incorporating prior information about the unknowns, rendering the problem less ill-posed. It also allows marginalisation of modelling errors utilising the so-called Bayesian approximation error method. A more recent trend in image reconstruction techniques is the use of deep learning, which has shown promising results in various applications from image processing to tomographic reconstructions. In this work, we study the non-linear DOT inverse problem of estimating the (absolute) absorption and scattering coefficients utilising a ‘model-based’ learning approach, essentially intertwining learned components with the model equations of DOT. The proposed approach was validated with 2D simulations and 3D experimental data. We demonstrated improved absorption and scattering estimates for targets with a mix of smooth and sharp image features, implying that the proposed approach could learn image features that are difficult to model using standard Gaussian priors. Furthermore, it was shown that the approach can be utilised in compensating for modelling errors due to coarse discretisation enabling computationally efficient solutions. Overall, the approach provided improved computation times compared to a standard Gauss-Newton iteration.

**Index Terms**—Deep learning, convolutional neural net-

Paper submitted for review on 20.04.2021. This work was supported in part by The Finnish Cultural Foundation (project 00200746), Academy of Finland (projects (314411, 312342, 336799, 336796, 320166, 334817, 338408), and the CMIC-EPSC platform grant (EP/M020533/1). (*Corresponding author: Meghdoot Mozumder*).

Meghdoot Mozumder is with the Department of Applied Physics, University of Eastern Finland, P.O. Box 1627, 70211 Kuopio, Finland (e-mail: meghdoot.mozumder@uef.fi).

Andreas Hauptmann is with the Research Unit of Mathematical Sciences; University of Oulu, Oulu, Finland and also with the Department of Computer Science, University College London, London WC1E 6BT, United Kingdom.

Ilkka Nissilä is with the Department of Neuroscience and Biomedical Engineering, Aalto University, Aalto University School of Science, P.O. Box 12200, FI-00076 Aalto, Espoo, Finland.

Simon R. Arridge is with the Department of Computer Science, University College London, London WC1E 6BT, United Kingdom.

Tanja Tarvainen is with the Department of Applied Physics, University of Eastern Finland, P.O. Box 1627, 70211 Kuopio, Finland and also with the Department of Computer Science, University College London, London WC1E 6BT, United Kingdom.

works, diffuse optical tomography, absolute imaging.

## I. INTRODUCTION

Diffuse optical tomography (DOT) utilises boundary measurements of near-infrared light to estimate spatially distributed optical absorption and scattering parameters in biological tissues [1]–[3]. The distribution of these optical parameters is useful in obtaining information on tissue function and structure with applications, for example, in imaging of breast cancer [4], [5], prostate imaging [6], [7], neonatal brain imaging [8], functional imaging of the adult brain [9], [10], and pre-clinical small animal imaging [11].

The image reconstruction problem of DOT, is an ill-posed inverse problem. The ill-posedness means that even small errors in measurements or modelling can cause large errors in the image reconstruction. An established strategy to handle the ill-posedness of DOT image reconstruction has been to use regularisation techniques. These techniques utilised assumptions such as smoothness of the solution [12], sparsity [13], [14], or its derivative (total-variation) [15] to obtain stable inversion. In a similar manner, Bayesian estimation utilises prior probability distributions of the unknowns, based on previously available knowledge, to compute the posterior probability distribution as a solution to the inverse problem [16]–[18]. In this regard, the Bayesian approximation error (BAE) approach has become a standard computational technique in ill-posed inverse problems such as DOT [17], [19]. The BAE approach computes statistics of modelling errors, such as reduced model or uncertainties, to compensate these during the solution of the inverse problem, for example [19]–[22]. For more information on image reconstruction problem of DOT and various methodologies, see e.g. [3], [9], [13], [18], [23] and the references therein.

Recently, deep learning methods have shifted the focus of tomographic imaging from classical, purely model-based techniques to data-driven approaches. A significant influence of these techniques has been the availability of large measurement databases, advances in computing capabilities and the potential to learn image features from the data itself. These have led to major improvements for many linear inverse problems, where high quality reference reconstructions can be readily obtained. Most notably, these include X-ray computed tomography [24]–[26], magnetic resonance imaging [27]–[29] and photoacoustic tomography [30]–[32].

For non-linear inverse problems and in particular in DOT, deep learning methods are scarce. This can be explained in parts by the difficulty to obtain high quality training data, costly model evaluations, and limitations of direct reconstruction approaches. Therefore, the initial applications of deep learning methods for DOT have been based on directly learning a non-linear mapping from the boundary measurement data to the spatially distributed optical coefficients [33]–[35] where the reconstruction operator was learned either by classical neural networks [33] or using convolutional neural networks (CNNs) [34], [35]. The approach was utilised in estimating absolute absorption coefficients [33], difference scattering coefficients [34], and spatially-constant values of both the coefficients [35]. The methods were validated with simulations [33], [34] and with two homogeneous experimental phantoms [35]. Improved estimates in terms of estimation accuracy and computation time were reported.

In contrast to directly learning a reconstruction operator, one can combine deep learning with model-based approaches. This can offer the possibilities to overcome some inherent limitations of the learning-based approaches, such as biases due to training samples and need for large training datasets [30]. Furthermore, model-based approaches are improved by providing complementary prior information and information on model uncertainties by training neural-networks to learn image features from a database of target images. A straightforward possibility to include the model equations into the reconstruction is to utilise CNNs as a post-processing tool after an initial reconstruction is obtained, for instance by removing streaking artefacts from filtered back-projection in computerised tomography [24], [25] or sharpening D-bar reconstructions in electrical impedance tomography [36], [37]. Alternatively, an architecture to invert the Lippmann-Schwinger equation for difference imaging of absorption coefficients in DOT was developed in [38].

In this work, we follow the approach of iterative model-based techniques [26], [28], [30] where learned components, given for instance by a CNN, are intertwined with the model equation. In particular, we extend the iterative model-based approach in [30] for solving an inverse problem with a non-linear forward operator. This enables us to tackle the absolute imaging problem of reconstructing both absorption and scattering coefficients in DOT. The image reconstruction problem is solved with a Gauss-Newton algorithm augmented with deep learning. To our knowledge, this is the first study of simultaneous reconstruction of absolute absorption and scattering parameters in DOT utilising deep learning. The use of the iterative model-based learning was chosen because of 1) non-linearity of the inverse problem, 2) to emphasise generalisability of the inversion method, and 3) to learn non-trivial features of images that are difficult to account for using conventional regularisation methods or Gaussian priors. In this approach, model correction is performed implicitly by the network while computing the iterative updates. As such, model-based estimates are enhanced by learned CNN components to obtain more accurate estimates. Further, learning the step length selection parameter, required for updating the estimates in each iteration, provides an accelerated iteration.

The rest of the paper is organised as follows. An introduction to DOT, Bayesian approach to inverse problems, and the proposed model-based learning approach is presented in Sec. II. Implementation of the proposed approach is presented in Sec. III. The numerical simulations and experiments are described in Secs. IV and V. These are followed by discussion in Sec. VI and conclusions in Sec. VII.

## II. METHODS

### A. Diffuse optical tomography

In a typical DOT measurement setup, near-infrared light is introduced into an object from its boundary. Let  $\Omega \subset \mathbb{R}^d$ , ( $d = 2$  or  $3$ ) denote the domain with boundary  $\partial\Omega$  where  $d$  is the (spatial) dimension of the domain. In a diffusive medium, like soft biological tissue, the commonly used light transport model for DOT is the diffusion approximation to the radiative transfer equation [39]. Here, we consider the frequency-domain version of the diffusion approximation [1]

$$\begin{aligned} & \left( -\nabla \cdot \frac{1}{d(\mu_a(r) + \mu'_s(r))} \nabla + \mu_a(r) + \frac{j\omega}{c} \right) \Phi(r) = 0, r \in \Omega, \\ & \Phi(r) + \frac{1}{2\zeta} \frac{1}{d(\mu_a(r) + \mu'_s(r))} \alpha \frac{\partial \Phi(r)}{\partial \hat{n}} = \begin{cases} \frac{q}{\zeta}, & r \in s \\ 0, & r \in \partial\Omega \setminus s \end{cases}, \end{aligned} \quad (1)$$

where  $\Phi(r)$  is the photon fluence,  $\mu_a(r)$  is the absorption coefficient,  $\mu'_s(r)$  is the (reduced) scattering coefficient,  $j$  is the imaginary unit,  $\omega$  is the angular modulation frequency of the input signal and  $c$  is the speed of light in the medium. The parameter  $q$  is the strength of the light source at location  $s \subset \partial\Omega$ , operating at angular modulation frequency  $\omega$ . Further, the parameter  $\zeta$  is a dimension-dependent constant ( $\zeta = 1/\pi$  when  $\Omega \subset \mathbb{R}^2$ ,  $\zeta = 1/2$  when  $\Omega \subset \mathbb{R}^3$ ) and  $\alpha$  is a parameter governing the internal reflection at the boundary  $\partial\Omega$ , and  $\hat{n}$  is an outward unit vector normal to the boundary. The measurable data on the boundary of the object, exitance  $\Gamma(r)$ , is given by

$$\Gamma(r) = -\frac{1}{d(\mu_a(r) + \mu'_s(r))} \frac{\partial \Phi(r)}{\partial \hat{n}} = \frac{2\zeta}{\alpha} \Phi(r). \quad (3)$$

The numerical approximation of the forward model (1)-(3) is typically based on a finite element (FE) approximation [1]. In the FE-approximation, the domain  $\Omega$  is divided into  $N_e$  non-overlapping elements joined at  $N_n$  vertex nodes. The photon fluence in the finite dimensional basis is given by

$$\Phi^h = \sum_{k=1}^{N_n} \phi_k \psi_k(r) \quad (4)$$

where  $\psi_k$  are the nodal basis functions of the FE-mesh and  $\phi_k$  is photon fluence in the nodes of the FE-mesh. We write the finite dimensional approximations for  $\mu_a(r)$  and  $\mu'_s(r)$  as

$$\mu_a(r) \approx \mu_a^h(r) = \sum_{l=1}^{N_n} \mu_{a,l} \psi_l(r) \quad (5)$$

$$\mu'_s(r) \approx \mu'_s{}^h(r) = \sum_{l=1}^{N_n} \mu'_{s,l} \psi_l(r) \quad (6)$$

where  $\mu_{a,l}, \mu'_{s,l}$  denote the absorption and scattering at the nodes of the FE-discretisation.

Typical data types for frequency-domain DOT are the logarithm of amplitude and phase, which is obtained from the real and imaginary parts of the logarithm of complex exitance  $\Gamma = \mathcal{A} \exp(i\phi)$ , as

$$y = \begin{pmatrix} \text{Re} \log(\Gamma) \\ \text{Im} \log(\Gamma) \end{pmatrix} = \begin{pmatrix} \log(\mathcal{A}) \\ \phi \end{pmatrix}, \quad (7)$$

where  $y \in \mathbb{R}^{N_m}$  is the data vector,  $\mathcal{A}$  is the amplitude and  $\phi$  is the phase delay of the measured signal. The FE-approximation of (1)–(3) and (7) is denoted by  $A_h$ , and the observation model is written as

$$y = A_h(\mu_a, \mu'_s) + e \quad (8)$$

where  $e \in \mathbb{R}^{N_m}$  models the random noise in measurements,  $\mu_a = [\mu_{a,1}, \dots, \mu_{a,N_n}] \in \mathbb{R}^{N_n}$  and  $\mu'_s = [\mu'_{s,1}, \dots, \mu'_{s,N_n}] \in \mathbb{R}^{N_n}$  are discretised absorption and scattering coefficients. The sub index  $h$  in the mapping  $A_h$  is a mesh parameter controlling the level of discretization. The operator  $A_h(\mu_a, \mu'_s)$  converges to the continuous forward operator as  $h \rightarrow 0$  and  $N_n \rightarrow \infty$ .

## B. Bayesian estimation

In the Bayesian approach to inverse problems, all the parameters are considered as random variables and the uncertainties of their values are encoded into probability density models [17]. Let us consider the observation model (8). The solution of the inverse problem is the posterior probability density which is obtained through Bayes' theorem, and can be written as

$$\pi(\mu_a, \mu'_s | y) \propto \pi(y | \mu_a, \mu'_s) \pi(\mu_a, \mu'_s) \quad (9)$$

where  $\pi(y | \mu_a, \mu'_s)$  is the likelihood density and  $\pi(\mu_a, \mu'_s)$  is the prior density.

Since we aim at computationally efficient solutions, we compute point estimate(s) from the posterior density, the most typical choice being the *maximum a posteriori* (MAP) estimate. Assuming that the unknowns  $\mu_a$  and  $\mu'_s$  and noise  $e$  are mutually independent and Gaussian distributed, i.e.

$$\mu_a \sim \mathcal{N}(\eta_{\mu_a}, \Gamma_{\mu_a}), \quad \mu'_s \sim \mathcal{N}(\eta_{\mu'_s}, \Gamma_{\mu'_s}), \quad e \sim \mathcal{N}(\eta_e, \Gamma_e),$$

where  $\eta_{\mu_a}, \eta_{\mu'_s}$  and  $\eta_e$  are the means, and  $\Gamma_{\mu_a}, \Gamma_{\mu'_s}$  and  $\Gamma_e$  are the covariance matrices, the MAP estimate is obtained as

$$(\hat{\mu}_a, \hat{\mu}'_s) = \arg \min_{\mu_a, \mu'_s} \left\{ \|L_e(y - A_h(\mu_a, \mu'_s) - \eta_e)\|^2 + \|L_{\mu_a}(\mu_a - \eta_{\mu_a})\|^2 + \|L_{\mu'_s}(\mu'_s - \eta_{\mu'_s})\|^2 \right\} \quad (10)$$

where  $L_{\mu_a}^T L_{\mu_a} = \Gamma_{\mu_a}^{-1}$ ,  $L_{\mu'_s}^T L_{\mu'_s} = \Gamma_{\mu'_s}^{-1}$  and  $L_e^T L_e = \Gamma_e^{-1}$ . The minimisation problem can be solved using an iterative method, such as Gauss-Newton method with iteration of the form

$$(\hat{\mu}_a, \hat{\mu}'_s)_{i+1} = (\hat{\mu}_a, \hat{\mu}'_s)_i + s_i (\delta \hat{\mu}_a, \delta \hat{\mu}'_s)_i \quad (11)$$

where  $(\delta \hat{\mu}_a, \delta \hat{\mu}'_s)_i$  is given by

$$\begin{pmatrix} \delta \hat{\mu}_a \\ \delta \hat{\mu}'_s \end{pmatrix}_i = \left( J_i^T \Gamma_e^{-1} J_i + \begin{pmatrix} \Gamma_{\mu_a}^{-1} & 0 \\ 0 & \Gamma_{\mu'_s}^{-1} \end{pmatrix} \right)^{-1} \begin{pmatrix} J_i^T \Gamma_e^{-1} (y - A_h(\hat{\mu}_a, \hat{\mu}'_s)_i) - \eta_e \\ + \Gamma_{\mu_a}^{-1} (\hat{\mu}_a, i - \eta_{\mu_a}) + \Gamma_{\mu'_s}^{-1} (\hat{\mu}'_s, i - \eta_{\mu'_s}) \end{pmatrix} \quad (12)$$

and  $s_i$  is the step length parameter. Here Jacobian  $J_i$  is the discrete representation of the Fréchet derivative of the non-linear mapping  $A_h(\hat{\mu}_a, \hat{\mu}'_s)_i$  at the current iterate.

## C. Prior

In this work, we use the following two Gaussian forms as our prior distributions for absorption and scattering.

1) *Gaussian Ornstein-Uhlenbeck prior*: To model smooth parameter distributions, we chose the prior model for the unknown parameters (10) as the Ornstein-Uhlenbeck process, which belongs to the Matérn class of covariance functions [40]. Ornstein-Uhlenbeck prior is a Gaussian distribution with the covariance matrix  $\Gamma$  defined as

$$\Gamma_{\mu, mk} = \sigma_\mu^2 \exp\left(-\frac{\|r_m - r_k\|}{\ell}\right) \quad (13)$$

where  $\mu$  denotes the unknown parameters (absorption and scattering),  $\sigma_\mu$  is the standard deviation,  $r_m$  and  $r_k$  are the locations of the FE-discretisation nodes  $m$  and  $k$ , and  $\ell$  is the characteristic length scale which controls the spatial range of correlation. The prior supports correlation between neighborhood discretization points, promoting distributions that can be locally close to homogeneous.

In this work, the means of the prior were  $\eta_{\mu_a} = 0.01 \text{ mm}^{-1}$  and  $\eta_{\mu'_s} = 1 \text{ mm}^{-1}$ , and the standard deviations were  $\sigma_{\mu_a} = 0.0033 \text{ mm}^{-1}$  and  $\sigma_{\mu'_s} = 0.33 \text{ mm}^{-1}$ . The correlation length was  $\ell = 8 \text{ mm}$ .

2) *Gaussian sample-based prior*: For targets which consisted of both non-smooth and smooth features, Gaussian sample-based priors were constructed using a set of sample images  $\{\mu_a, \mu'_s\}^j, j = 1, \dots, N_{\text{samp}}$ . The prior means and covariances were computed as

$$\eta_\mu = \frac{1}{N_{\text{samp}}} \sum_{i=1}^{N_{\text{samp}}} \mu^j \quad (14)$$

$$\Gamma_\mu = \frac{1}{N_{\text{samp}} - 1} \sum_{i=1}^{N_{\text{samp}}} (\mu^j - \eta_\mu)(\mu^j - \eta_\mu)^T. \quad (15)$$

## D. Bayesian approximation error method

In practical applications, the use of a sufficiently dense discretization may be infeasible due to computational resource and time limitations. In such a case, the observation model (8) with a forward operator with a fine discretisation is replaced by an approximate model

$$y = A_H(\bar{\mu}_a, \bar{\mu}'_s) + e \quad (16)$$

where the discretization parameter  $H > h$  and  $(\bar{\mu}_a, \bar{\mu}'_s)$  are the corresponding discretised optical coefficients with discretisation  $H$ . In the Bayesian approximation error approach, instead of just using the approximate model (16) we re-write the observation model (8) in the following way

$$y = A_H(\bar{\mu}_a, \bar{\mu}'_s) + \underbrace{\{A_h(\mu_a, \mu'_s) - A_H(\bar{\mu}_a, \bar{\mu}'_s)\}}_{\varepsilon} + e. \quad (17)$$

Here,  $\varepsilon$  is the discretisation error describing the discrepancy between the accurate forward model and the approximate

model. The Bayesian approximation error method carries out an approximate marginalisation of the posterior over the error  $\varepsilon$ . Following [17], [19], the MAP estimate with the Bayesian approximation error model is obtained as

$$(\hat{\mu}_a, \hat{\mu}'_s) = \arg \min_{\mu_a, \mu'_s} \left\{ \|L_{\varepsilon+e}(y - A_H(\bar{\mu}_a, \bar{\mu}'_s) - \eta_\varepsilon - \eta_e)\|^2 + \|L_{\mu_a}(\bar{\mu}_a - \eta_{\mu_a})\|^2 + \|L_{\mu'_s}(\bar{\mu}'_s - \eta_{\mu'_s})\|^2 \right\} \quad (18)$$

where  $\eta_\varepsilon$  and  $\Gamma_\varepsilon$  are the mean and covariance of the approximation error. Further,  $L_{\varepsilon+e}^\top L_{\varepsilon+e} = (\Gamma_\varepsilon + \Gamma_e)^{-1}$ . In the following sections, we refer to the solution of (18) as the MAP estimate with the Bayesian approximation error (BAE) approach.

### E. Model-based learning using deep Gauss-Newton

In this work, instead of the regular Gauss-Newton update, Eq. (11), we propose to learn an update function for each iteration

$$(\hat{\mu}_a, \hat{\mu}'_s)_{i+1} = G_{\theta_i}((\hat{\mu}_a, \hat{\mu}'_s)_i, (\delta\hat{\mu}_a, \delta\hat{\mu}'_s)_i). \quad (19)$$

The functions  $G_{\theta_i}$  correspond to CNNs with different, learned parameters  $\theta_i$  but with the same architecture. This implies that the update of the optical parameters including the step length selection is now learned from the data during training. The network structure is kept simple and shown in Fig. 1. Due to the representation of each update by a CNN applied to the optical parameters  $(\hat{\mu}_a, \hat{\mu}'_s)_i$  the Gauss-Newton search directions  $(\delta\hat{\mu}_a, \delta\hat{\mu}'_s)_i$ , we refer to the algorithm as a deep Gauss-Newton (DGN). In contrast to the previously proposed model-based learning technique for DOT difference imaging in Ref. [38], we train the DGN layer by layer (layer corresponding here to one iterate), due to the non-linear nature of the absolute imaging problem. Hence, we learn the parameters  $\theta_i$  for each iteration separately.

## III. IMPLEMENTATION

The Toast++ software [41] was utilised in the FE-solution of the diffusion equation using MATLAB (R2017b, Mathworks, Natick, MA). A Python library, Tensorflow (version 1.14.0) [42] was utilised in implementation and training of the DGN algorithm. The simulations, were carried out in a Fujitsu Celcius W550 desktop workstation, with Intel@Xeon(R) W-2125 CPU @ 4.00GHz×8. The training of the DGN were carried out on a NVIDIA Tesla V100 GPU.

### A. Implementation of the deep Gauss-Newton

The architecture chosen for the CNNs performing the update in Eq. (19) is illustrated in Fig. 1. In each iteration, optical coefficients and the corresponding search directions are given as an input to a pipeline. They are expanded to 20 and then 40 channels by a convolutional layer with kernel size of 5 pixels, and dimension  $d = 2$  or 3, including bias and equipped with a ‘leaky’ rectified linear unit (LReLU) as non-linearity, that was defined as

$$\text{LReLU}(\mu) = \max(\mu, 0.1\mu).$$

As opposed to commonly used ReLU, the use of LReLU allow for negative values of input parameters ( $\delta\hat{\mu}_a, \delta\hat{\mu}'_s$ ). The number of channels represent the number of images that are output after each convolution and non-linear operations, representing the number of image variations to be learned. The expansive part of the network serves as a feature extractor (encoder) and the contracting part feature fusion (decoder). The kernel size is the pixel size on which the convolution filters are applied, and relates to the size of spatial variations to be learned. The outputs of the pipelines are added together, and first reduced to 20 channels, equipped with a LReLU, and then to 1 channel without a non-linearity, but multiplication to a scalar value (representing the step length parameter  $s_i$ ). The outcome is added to the current iterate and projected to the positive numbers by a LReLU. Before applying the network, the optical parameters  $(\mu_a, \mu'_s)$  were scaled ( $\mu_a \mapsto 10^2 \times \mu_a, \mu'_s \mapsto \mu'_s$ ) to present values in the same numerical range.

Since the main contribution of this work is not the specific neural network architecture, we used a simple architecture following Ref. [30]. As shown, the network structure is kept rather small with the idea that each  $G_{\theta_i}$  primarily learns how to combine the current iterate parameters and the search directions, in contrast to a large post-processing network.

### B. Training the deep Gauss-Newton

The network was trained by simulations utilising absorption and scattering distributions, i.e. ‘ground-truth images’, drawn from a prior  $\{\mu_{a,\text{true}}, \mu'_{s,\text{true}}\}^j, j = 1, \dots, N_{\text{samp}}$  and corresponding simulated measurement data. The data was simulated using a FE-approximation of the diffusion approximation (1)-(3) and it was corrupted with additive noise. Since CNNs operate on uniform pixel domains, the images drawn in the mesh basis were interpolated to the image basis for the application of the CNN and back for simulation of (1)-(3). Alternatively, one can use graph structures to formulate the problem on the FE mesh directly [43].

The parameters  $\theta_i$  for  $i = 1, \dots, i_{\text{max}}$ , were trained sequentially, where  $i_{\text{max}}$  was the maximum number of iterations. The parameters were trained on each iteration  $i$  by minimising the ‘L<sub>2</sub>-loss’ function

$$\min_{\theta_i} \sum_j \|\mu_{a,i+1}^j - \mu_{a,\text{true}}^j\| + \|\mu_{s,i+1}^j - \mu_{s,\text{true}}^j\|, \quad (20)$$

where  $\mu_{i+1}^j$  at an iterate was given by Eq. (19). For the iteration, the initial guess,  $\mu_1^j$  was chosen as the mean of prior. Thereafter,  $\theta_1$  was trained to minimise the difference between  $\mu_{\text{true}}^j$  and  $\mu_2^j$  (computed using Eq. (19)) for all indices  $j$ . The procedure was repeated to train all  $\theta_i$ ’s. In this work,  $i_{\text{max}}$  was chosen as 5, which has been a typical number of iterations required for convergence of Gauss-Newton iterations in DOT absolute imaging according to our experience. Training of the  $\theta_i$ ’s was carried out by minimising (20) with the TensorFlow’s implementation of the Adam optimiser [44] using batches of size 2, maximum of 10 epochs and step size of  $5 \cdot 10^{-4}$  (learning rate). The estimate  $\mu_{i+1}^j$  at an iterate approaches the accurate solution when trained to a sufficiently low loss. The

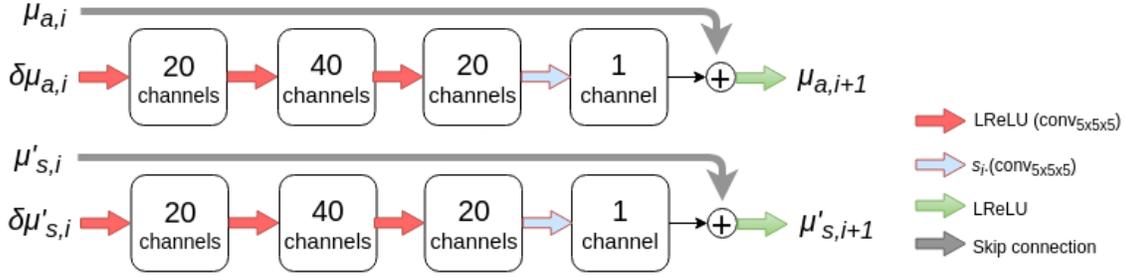


Fig. 1. Diagram of one CNN denoted as  $G_{\theta_i}$ , representing one iteration of the deep Gauss-Newton. The red arrows denote a convolutional layer with  $5 \times 5$  kernel for a 2D image (or  $5 \times 5 \times 5$  kernel for a 3D image), bias and followed by a LReLU. The resulting channels in each layer are indicated in the squares. The blue arrow denotes a convolutional layer followed by a scalar multiplication. The residual update (by the skip connection) is then projected to the positive numbers by the last LReLU (shown as green arrow).

training was terminated when the change of average  $L_2$ -loss (20) was less than 0.1% between two consecutive epochs.

The training procedure is aimed to improve the conventional update of optical parameters (12) based on image features learned from the training samples. These image features and the step length selection are encoded in the learned parameters  $\theta_i$ . The training procedure is summarised in Algorithm 1.

#### Algorithm 1 Training the deep Gauss-Newton

- 1: Draw set  $\{\mu_{a,\text{true}}, \mu'_{s,\text{true}}\}^j, j = 1, \dots, N_{\text{samp}}$ , from prior.
- 2: Generate noisy measurement data using Eq. (8).
- 3: Set  $(\hat{\mu}_a, \hat{\mu}'_s)_1$  as mean of prior (10), for  $N_{\text{samp}}$  cases.
- 4: **function** ITERATE
- 5:    $i \leftarrow 1$
- 6:   **while**  $i < i_{\text{max}}$  **do**
- 7:     Compute  $(\delta\hat{\mu}_a, \delta\hat{\mu}'_s)_i^j$ , using Eq. (12) for  $N_{\text{samp}}$  cases.
- 8:     **function** TRAIN( $(\hat{\mu}_a, \hat{\mu}'_s)_i^j, (\delta\hat{\mu}_a, \delta\hat{\mu}'_s)_i^j, (\mu_{a,\text{true}}, \mu'_{s,\text{true}})^j$ )
- 9:       Train  $\theta_i$ 's by minimizing Eq. (20)
- 10:     **end function** Return  $\theta_i$
- 11:      $(\hat{\mu}_a, \hat{\mu}'_s)_{i+1} \leftarrow G_{\theta_i}((\hat{\mu}_a, \hat{\mu}'_s)_i, (\delta\hat{\mu}_a, \delta\hat{\mu}'_s)_i)$ .
- 12:      $i \leftarrow i + 1$
- 13:   **end while**
- 14: **end function**

### C. Evaluating the deep Gauss-Newton

After training the parameter sets  $\theta_i$ , the learned iterative reconstruction scheme was evaluated by applying the network  $G_{\theta_i}$  at each iteration. This procedure was equivalent to Algorithm 1, starting by setting  $(\hat{\mu}_a, \hat{\mu}'_s)_1$ , calling function 'ITERATE', and skipping the function 'TRAIN'.

## IV. SIMULATIONS

### A. Data generation

In the numerical studies, the domain  $\Omega \subset \mathbb{R}^2$  was a circle with a radius of 35 mm. The measurement setup consisted of 16 sources and 16 detectors modelled as 2 mm wide surface patches located at equi-spaced angular intervals on the boundary. The target optical parameters were either drawn from the Gaussian Ornstein-Uhlenbeck prior, as shown in Fig.

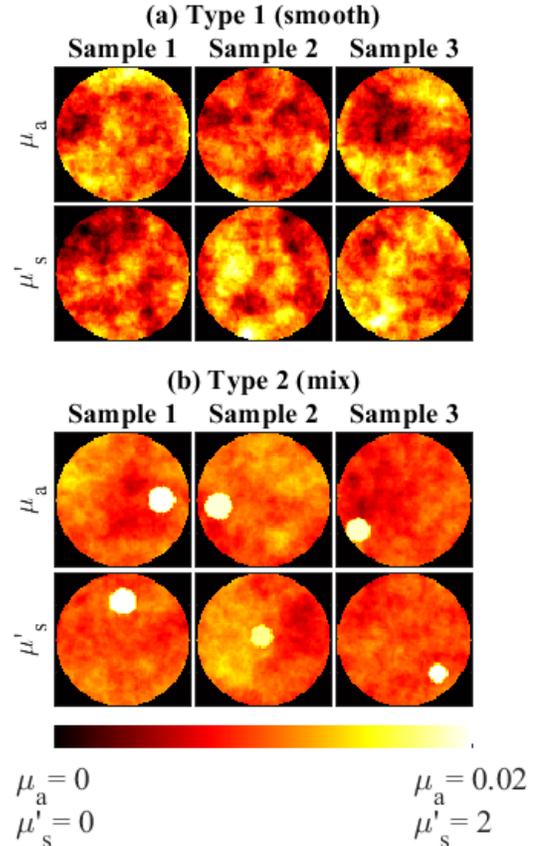


Fig. 2. (a) Three samples of 'smooth' absorption  $\mu_a$  and scattering  $\mu'_s$  images drawn from Ornstein-Uhlenbeck prior and (b) three samples drawn from 'mix' targets.

2 (a), or drawn as a mix of smoothly varying background (drawn from the Ornstein-Uhlenbeck prior) and sharp circular inclusions with varying contrast and radii, as shown in Fig. 2 (b). The optical parameter values were chosen to mimic those in biological tissues [45].

The measurement data were simulated using the FE-approximation of the diffusion approximation (1)-(3), using a 'forward mesh' shown in Fig. 3 (a). Random measurement noise  $e$  drawn from a zero-mean Gaussian distribution

$$\pi(e) = \mathcal{N}(0, \Gamma_e), \quad \Gamma_e = \text{diag}(\sigma_{e,1}^2, \dots, \sigma_{e,N_m}^2), \quad (21)$$

where the standard deviations  $\sigma_{e,k}$  were specified as 1% of the

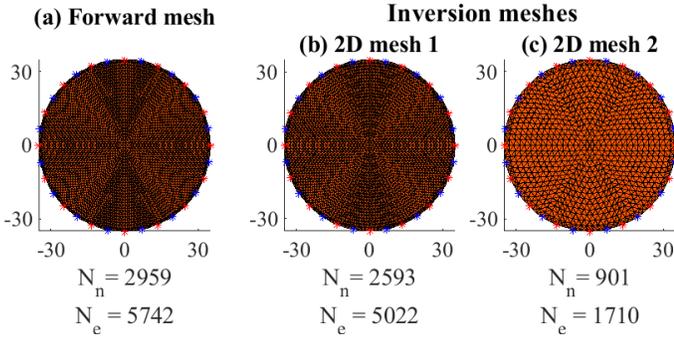


Fig. 3. Meshes used in 2D simulations. Locations of the sources and detectors are shown as red and blue stars on the mesh boundaries. The number of FE-nodes  $N_n$  and elements  $N_e$  for the meshes are also displayed. (a) Forward mesh was used for simulating measurement data. (b)-(c) Inversion meshes 2D mesh 1 and 2D mesh 2 were used for estimating the optical coefficients from the simulated data.

TABLE I

COMPUTATION TIMES FOR TRAINING 1000 ‘SMOOTH’ SAMPLES IN HOURS (H) OR SECONDS (S), USING DIFFERENT INVERSION MESHES. AVERAGE EVALUATION TIME PER SAMPLE IS REPORTED.

Inverse mesh	Training		Evaluation		
	DGN	BAE	GN	BAE	DGN
2D mesh 1	17h		9s		7s
2D mesh 2	19h	127s	4s	4s	3s
3D mesh 1			615s		
3D mesh 2	20h	3319s	112s	132s	96s

simulated noise-free data, was added to the simulated data.

## B. Estimation

Estimation of the optical parameters was first carried out in a ‘2D mesh 1’, shown in Fig. 3 (b). ‘Forward mesh’ and ‘2D mesh 1’ had a similar level of discretisation but they were chosen differently to avoid making an inverse crime [17].

The deep Gauss-Newton (DGN) were trained with a set of 1000 absorption and scattering images as described in Section III-B. The training images were either smooth distributions or mix distributions or both of these. The training times are given in Table I.

Absorption and scattering images were estimated using the model-based DGN as described in Sec. III. For comparison, reconstructions using the conventional Gauss-Newton (GN) Eq. (11) were computed. The conventional GN method used the Orstein-Uhlenbeck prior when estimating smooth targets and the Gaussian sample-based prior when estimating mix targets. The DGN utilised the Gaussian Orstein-Uhlenbeck prior for all cases. The mean  $\eta_e = 0$  and covariance  $\Gamma_e$  of the measurement noise were assumed known.

1) *Case 1: DGN trained with smooth images:* In this case, the proposed DGN was trained using the smooth images. Example reconstructions from one smooth target and one mix target are shown in Fig. 4 (a)-(c). Further, statistics of relative errors of absorption and scattering estimates from 1000 simulated targets and the computation times are shown in Fig. 4 (d)-(e) and in Table I.

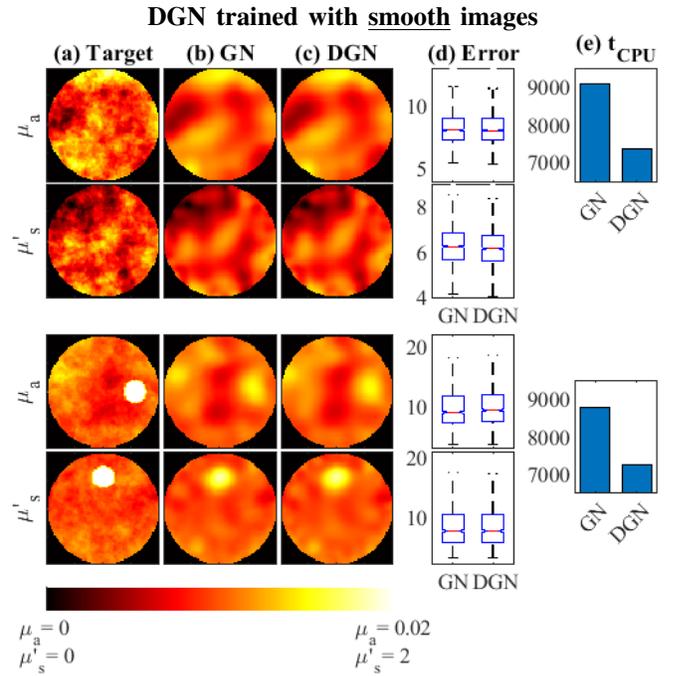


Fig. 4. Reconstructions of (top) smooth and (bottom) mix targets, with conventional GN and DGN trained using smooth images. (a) Target absorption  $\mu_a$  and scattering  $\mu_s$ , (b) estimates using GN and (c) estimates using DGN. Statistics of estimation errors for 1000 evaluation cases are shown as ‘boxplots’ in (d), and the evaluation times in (e).

As shown, DGN provide estimates with similar quality and accuracy as the conventional GN for both smooth and mix targets. The computation times of the estimates with DGN were lower than with the GN. Although the DGN required additional evaluation of CNNs, it still provided a computational advantage since the step-length parameter ( $s_i$ ) was learned and didn’t need to be computed at each iteration.

2) *Case 2: DGN trained with mix targets:* In this case, the DGN was trained with mix images. Example reconstructions from one smooth target and one a mix target are shown in Fig. 5 (a)-(c). Further, statistics of relative errors of absorption and scattering estimates from 1000 simulated targets and the computation times are shown in Fig. 5 (d)-(e) and in Table I.

As it can be seen, the absorption and scattering inclusions of the mix targets are visualised more clearly when estimated using the proposed DGN method. Furthermore, the estimates from the mix targets obtained using DGN have lower errors when compared to the estimates obtained conventional GN. On the other hand, the ‘out-of-distribution’ smooth targets now show higher errors, due to artificial sharpening of the estimated images with the DGN. Again, the computation times of the estimates with the DGN were lower.

3) *Case 3: DGN trained with both smooth and mix targets:* In this case, the DGN was trained with both smooth and mix images, i.e. 50 % of the training samples were smooth images and 50 % were mix images.

Example reconstructions from one smooth target and one mix target are shown in Fig. 6 (a)-(c). Further, statistics of the relative errors of absorption and scattering estimates from 1000 simulated targets and the computation times are shown in Fig. 6 (d)-(e) and in Table I.

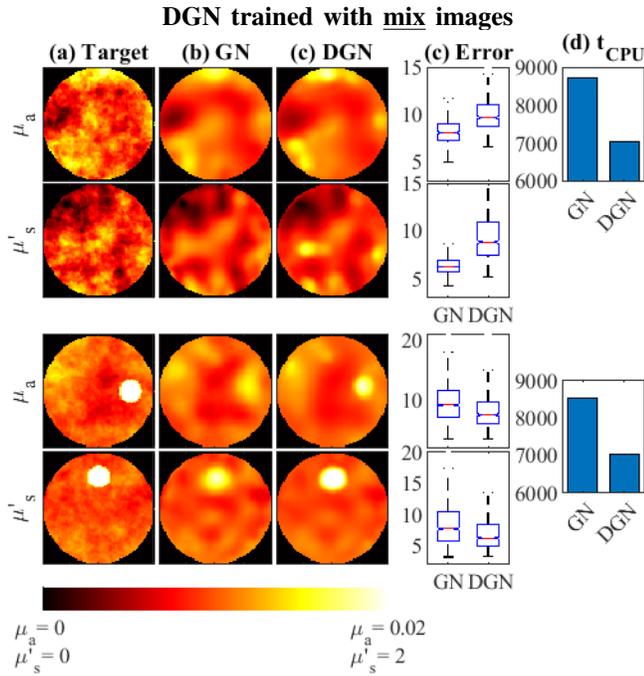


Fig. 5. Reconstructions of (top) mix and (bottom) smooth targets, with conventional GN and DGN trained using mix images. (a) Target absorption  $\mu_a$  and scattering  $\mu'_s$ , (b) estimates using GN and (c) estimates using DGN. Statistics of estimation errors for 1000 evaluation cases are shown in (c), and the evaluation times in (d).

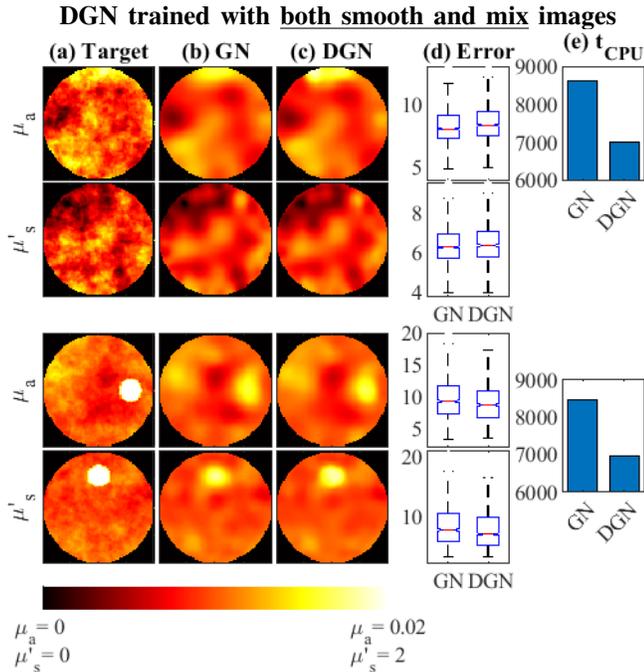


Fig. 6. Reconstructions of (top) smooth and (bottom) mix targets, with conventional GN and DGN trained using both smooth and mix images. (a) Target absorption  $\mu_a$  and scattering  $\mu'_s$ , (b) estimates using GN and (c) estimates using DGN. Statistics of estimation errors for 1000 evaluation cases are shown in (c), and the evaluation times in (d).

As it can be seen, the reconstructions from smooth targets are similar in quality when comparing DGN and GN solutions. Also the relative errors are similar in magnitude. For the mix targets, the reconstructions obtained using DGN show slightly

better contrast than those obtained with GN. However, the difference between these is not as clear as when compared to results obtained using mix training data. Further, the relative errors are similar in magnitude. Computation times of the estimates with the DGN were lower.

### C. Estimation in presence of modelling errors

Then, the DGN method in the presence of modelling errors due to a coarse discretisation was studied. Use of coarse discretisation is beneficial for memory consumption and computation times. In these simulations, a mesh, ‘2D mesh 2’ shown in Fig. 3 (c), was used.

Absorption and scattering images were estimated using the DGN. For comparison, estimates using conventional GN and Gauss-Newton augmented with Bayesian approximation error modelling (BAE) described in Sec. II-D were computed.

For training the DGN, the network parameters were trained using 1000 smooth distributions or mix distributions. For training the BAE method, the statistics of the discretisation errors were calculated with 1000 samples of smooth or mix distributions. These absorption and scattering parameters ( $\mu_a, \mu'_s$ ) were projected from ‘2D mesh 1’ to ‘2D mesh 2’ to obtain  $(\bar{\mu}_a, \bar{\mu}'_s)$ . Thereafter, the accurate forward solutions  $A_h(\mu_a, \mu'_s)$  and approximate forward solutions  $A_H(\bar{\mu}_a, \bar{\mu}'_s)$  were computed. Finally, the mean ( $\eta_\varepsilon$ ) and covariance ( $\Gamma_\varepsilon$ ) of the approximation error  $\varepsilon = A_h(\mu_a, \mu'_s) - A_H(\bar{\mu}_a, \bar{\mu}'_s)$  were computed. The training times of the BAE and DGN are given in Table I. As seen, the DGN training times were higher than training of the conventional BAE.

#### 1) Case 1: DGN and BAE trained with smooth images:

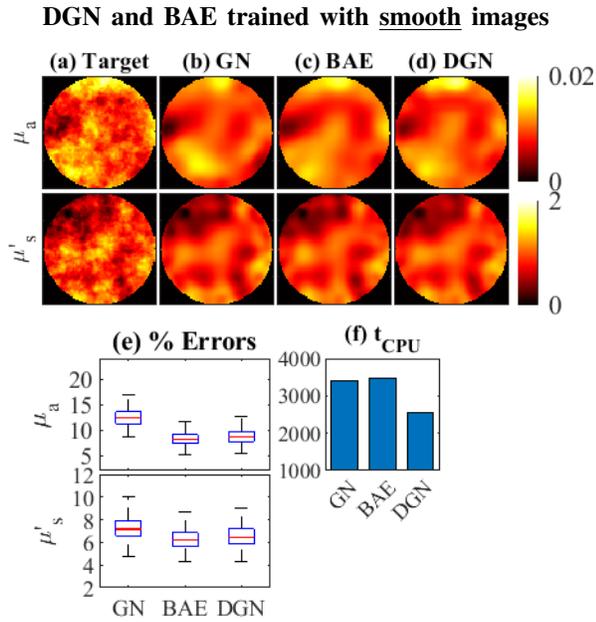
Example reconstructions from a smooth target in the presence of discretisation errors is shown in 7 (a)-(d). The estimates were computed using the proposed DGN and compared against conventional GN and GN augmented with BAE. Further, statistics of the relative errors of the estimates and computation times from 1000 simulated targets, are shown in Fig. 7 (e)-(f) and Table I.

It can be seen that the proposed DGN and the BAE effectively marginalise modelling errors, providing comparable images that have lower errors when compared to the GN. The relative errors obtained with the BAE were comparable to using the DGN. As seen, the evaluation time of the DGN was lower than using the BAE, although the training time of the DGN was higher than the BAE.

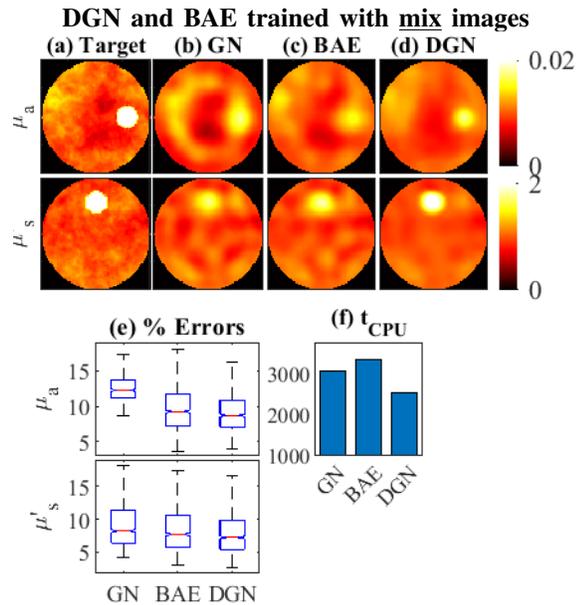
#### 2) Case 2: DGN and BAE trained with mix images:

Example reconstructions from a smooth target obtained using DGN, GN and BAE in the presence of discretisation errors are shown in Fig. 8 (a)-(d). Statistics of the relative errors of the estimates and computation times from 1000 simulated targets are shown in Fig. 8 (e)-(f) and Table I.

As it can be seen, the DGN and the BAE provide images with better quality than the conventional GN that suffers from artefacts due to the modelling errors. Furthermore, the relative errors of DGN and BAE are lower than of GN, with DGN providing the lowest relative errors for both absorption and scattering estimates. The results demonstrate that the DGN can compensate for modelling errors slightly better than the BAE



**Fig. 7.** Reconstructions of smooth targets in the presence of modelling errors. (a) Target absorption  $\mu_a$  and scattering  $\mu'_s$ , (b) estimates using GN, (c) BAE, and (d) DGN. Artefacts due to modelling errors are seen in the GN estimates. These are compensated in the BAE and DGN estimates. Statistics of the estimation errors for 1000 evaluation cases are shown in (e), and the evaluation times in (f).

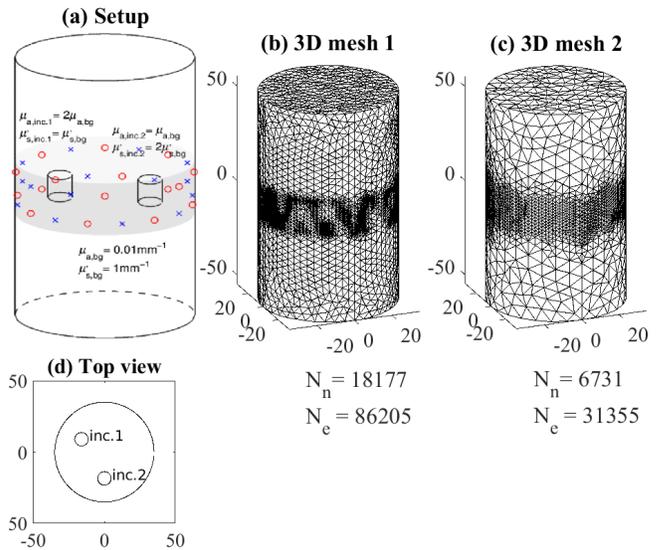


**Fig. 8.** Reconstructions of mix targets in the presence of modelling errors. (a) Target absorption  $\mu_a$  and scattering  $\mu'_s$ , (b) estimates using GN, (c) BAE, and (d) DGN. Artefacts due to modelling errors are seen in the GN estimates. These are compensated in the BAE and DGN estimates. Statistics of the estimation errors for 1000 evaluation cases are shown in (e), and the evaluation times in (f).

in the case of non-smooth targets. The DGN also had lower computation time compared to BAE. However, the training time was higher.

## V. EXPERIMENTS

The phantom experiment was carried out with the frequency-domain DOT instrument at the Aalto University,



**Fig. 9.** (a) DOT experimental setup with position of sources (red circles) and detectors (blue crosses) and two cylindrical inclusions. (b) '3D mesh 1' is a densely discretised 3D mesh used in calculating 'reference' estimate with the GN method. (c) '3D mesh 2' was used in calculating estimates with the DGN, GN and GN augmented BAE. The number of FE-nodes ( $N_n$ ) and elements ( $N_e$ ) in the meshes are also displayed. (d) Top view of the phantom showing location of the inclusions.

Finland [46]. A cylindrical phantom with a radius of 35 mm and height of 110 mm illustrated in Fig. 9 (a) and (d) was studied. The background optical parameters were approximately  $\mu_a = 0.01 \text{ mm}^{-1}$  and  $\mu'_s = 1 \text{ mm}^{-1}$  at wavelength 800 nm, and two cylindrical inclusions which both had the diameter and height of 9.5 mm, were located such that the central planes of the inclusions coincided with the central xy-plane of the cylinder domain. The optical properties of the inclusion 1 were approximately  $\mu_{a,\text{inc.1}} = 0.02 \text{ mm}^{-1}$ ,  $\mu'_{s,\text{inc.1}} = 1 \text{ mm}^{-1}$  (i.e., purely absorption contrast) and the optical properties of the inclusion 2 were  $\mu_{a,\text{inc.2}} = 0.01 \text{ mm}^{-1}$ ,  $\mu'_{s,\text{inc.2}} = 2 \text{ mm}^{-1}$  (i.e., purely scatter contrast), respectively.

The phantom provides absorption and scattering contrast of 2:1, similar to optical parameter variations in tumors [45]. The source and detector configuration in the experiment consisted of 16 sources and 15 detectors arranged in an interleaved order on two rings located 6 mm above and below the central xy-plane of the cylinder domain. The locations of sources and detectors are shown with red circles and blue crosses respectively in Fig. 9 (a). The measurements were carried out at 785 nm with an optical power of 8 mW and angular modulation frequency  $\omega = 100 \text{ MHz}$ . The nearest measurement data from each source position were removed from measured amplitude and phase data. Logarithm of amplitude and phase shift were stored as data, leading to real-valued measurement vectors  $y \in \mathbb{R}^{360}$ .

1) *Reference estimate:* The absorption and scattering parameters were estimated in a densely discretised mesh to provide a reference for the other approaches. A '3D mesh 1' shown in Fig. 9 (b) was used.

To calibrate the source strength and phase coupling of the forward model to the experimental setup, a global calibration was carried out on the measured data as the initial step in

the reconstruction process. Following the initial estimation procedure in [47], the logarithm of source strength and phase coupling were modelled by additive constants. The initialisation step consisted of a four-parameter fit of global background parameters  $\mu_{a1*}$  and  $\mu'_{s1*}$  as well as a global additive shift  $\eta$  of  $\log(\text{amplitude})$  data and a global additive shift  $\phi$  of phase data,

$$(\mu_{a*}, \mu'_{s*}, \eta, \phi) = \arg \min_{\mu_{a*}, \mu'_{s*}, \eta, \phi} \|L_e((y + \Delta y) - A(\mu_{a*}, \mu'_{s*}))\|^2, \quad (22)$$

where  $\Delta y = (\eta, \phi)^T$ . The initialization problem was solved using GN method which resulted in parameter estimates:  $\eta = 3.5$ ,  $\phi = 0.025$ ,  $\mu_{a1*} = 0.01\text{mm}^{-1}$ ,  $\mu'_{s1*} = 0.8\text{mm}^{-1}$ .

Once the initialisation was completed, the measurement data was transformed for the standard GN estimation (11) by the recovered global offsets as  $y \mapsto y + \Delta y$ , and the initial parameter values and the prior means were set to the estimated values  $\mu_{a1*}$  and  $\mu'_{s1*}$ .

Thereafter, the reference estimates were calculated with the standard GN method. In the solution, the measurement noise was assumed to be 1% of the measured absolute values of the log amplitude and phase. Further, the Ornstein-Uhlenbeck prior was used with the same parameter values that were used in the simulations used.

2) *Estimates in presence of model errors:* A coarse ‘3D mesh 2’, shown in Fig. 9 (c) was used in evaluating the proposed DGN approach in the presence of discretisation errors. For comparison, the BAE approach was utilised, and the related minimisation problem was solved using GN. Before reconstructions, the above described global calibration was repeated for the data using the ‘3D mesh 2’.

To train the DGN, 1000 samples of 3D distributed absorption and scattering distributions were drawn from the Ornstein-Uhlenbeck prior and simulated measurement data was calculated using ‘3D mesh 1’. To mimic the experimental situation of unknown source strengths and phase coupling, the corresponding coefficients  $(\eta, \phi)$  were drawn from uniform distributions

$$\eta \sim \mathcal{U}(1, 4), \quad \phi \sim \mathcal{U}(0.01, 0.04)$$

and added to the simulated measurement data. Thereafter, for the training procedure, the ‘3D mesh 2’ was used to compute the four-parameter fit using Eq. (22), transform the data ( $y \mapsto y + \Delta y$ ), and subsequently to train the DGN with the procedure described in Section III-B.

3) *Results:* Estimated absorption and scattering distributions calculated using the DGN, GN and GN augmented with BAE in a coarse discretisation and a GN in a fine discretisation are shown in Fig. 10. Estimates on the central xy-plane of the cylindrical domain are visualised.

As it can be seen in Fig. 10 (a), reference estimates computed in ‘3D mesh 2’ show locations of absorption and scattering inclusions with some boundary artefacts. Absolute and difference imaging reconstructions using the phantom were earlier presented in Refs. [47], [48], and they show similar quality reconstructions. The boundary artefacts are larger in both absorption and scattering estimates obtained with the GN in ‘3D mesh 2’ in Fig. 10 (b). Furthermore,

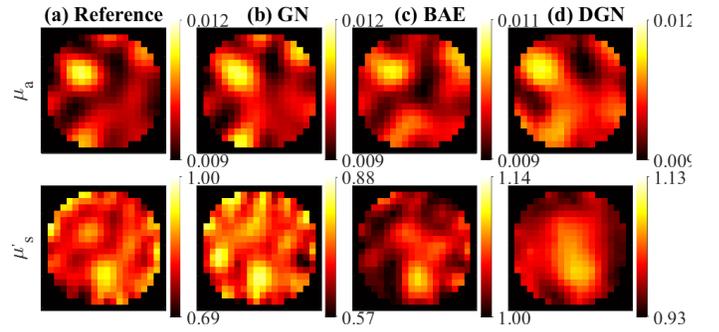


Fig. 10. Absorption  $\mu_a$  and scattering  $\mu'_s$  distributions reconstructed from experimental data. (a) Reference estimate computed with dense mesh ‘3D mesh 1’, (b) GN estimate computed with mesh ‘3D mesh 2’ and (c) DGN estimate compute with ‘3D mesh 2’.

these artefacts are reduced utilising both the GN augmented with BAE and DGN, as seen in Fig. 10 (c)-(d).

Note that the training of the DGN utilised smooth 3D images drawn from Ornstein-Uhlenbeck prior, and not sharp inclusions as were present in the experimental phantom. As such, the evaluation in this case was carried out using a ‘out-of-distribution’ target. The experiment demonstrates that DGN trained on generalised smooth images and coarse meshes can produce images with comparable quality to more dense meshes, using lower computational resources than BAE method.

The training and reconstruction time for the DGN, GN and the BAE method are presented in Table I. It can be seen that the reconstruction time using the experimental data was lower using the proposed DGN, although the training time (using simulated data) was higher.

## VI. DISCUSSION

This work proposed a ‘model-based’ deep-learning approach to absolute imaging of DOT. We demonstrated that model-based learning can provide computational advantages to the standard Bayesian inversion methods, when convolutional neural networks trained with similar images are utilised in the iterative parameter estimation procedure. This was demonstrated utilising smooth and mix targets in Figs. 4, 5 and 6.

In addition to the improvement in computation time, we demonstrated that the method can lead to improved estimates for targets which contain both sharp and smooth features, in Fig. 5. As such, the proposed method can learn non-Gaussian image features more efficiently than standard Gaussian priors. We note that using hyper-priors in Bayesian estimation also allow estimating such sharp image discontinuities [49]. However, utilising hyper-priors in DOT might lead to additional computational burden, for further estimating hyper-parameters related to the priors.

In Figs. 7, 8 we demonstrated the efficacy of the proposed method in compensation of discretisation errors. As shown in 7, the proposed DGN can marginalise discretisation errors with similar accuracy to BAE for smooth targets, providing an advantage in computational time. For mix targets, the proposed DGN outperforms the BAE in estimation accuracy and computational times. This is possibly because CNNs can

learn and compensate non-Gaussian modelling errors more efficiently than BAE, which assumes modelling errors as Gaussian. The authors refer to [50], [51] for more discussions on modelling error corrections using CNNs.

The proposed approach was applied to experimental data and the estimates are presented in Fig. 10. As shown, both BAE and DGN trained on smooth images can effectively compensate modelling errors, and provide accuracy comparable to reference estimates.

The purpose of the study was to present the applicability of the approach to absolute imaging problem of DOT with one relevant architecture. Therefore, we have not presented an in-depth assessment of different deep learning architectures in this work. The proposed DGN approach could be further improved by carrying out a systematic optimisation of the network architectures and involved parameters. Further, the different estimation scenarios, such as the discretisation level and geometry affect on the optimal training procedure as well as modelling errors and experimental system related uncertainties. These topics are part of planned future studies. The authors refer to [27], [32] for other architectures that could also be relevant for DOT.

A classical challenge in DOT is the cross-talk between the optical parameters in the reconstructions. Artifacts due to cross-talk are difficult to spot in smooth or mix images, as were used in the article. Figure 11 shows reconstructions with sharp inclusions solved with GN and DGN. The DGN was trained with sharp inclusions, with random inclusion locations, sizes and contrast. As seen, the cross-talk artefacts that are visible in GN reconstruction are not seen in DGN. We also compared the approach with a sparse recovery scheme that is known to reduce spatial variations in a reconstructed image, whilst preserving sharp discontinuities [13]. We used a sparsity-promoting total-variation (TV) prior [52] implemented with Toast++ software [41] and the conventional GN algorithm. The reconstructed images are shown in Figure 11(d). As seen, use of TV prior results in lower background variations compared to using Gaussian prior in Figure 11(b). However, the cross-talk artifact of the scattering inclusion in the absorption image is still visible.

The study draws connections between statistical Bayesian and learning-based approaches, showing that insights from Bayesian inversion can be used in the design of learned image reconstruction. This opens the possibility for future studies to analyse these connections more carefully.

## VII. CONCLUSIONS

We presented a novel approach for estimating absolute optical parameters in DOT, utilising a model-based iterative deep-learning approach. The results were validated with 2D simulations and a DOT experiment. The results show that the proposed approach leads to improved computational times compared to conventional Gauss-Newton method. Also, the proposed approach can learn non-Gaussian image features and provide improved estimates for targets presenting sharp inclusions. No considerable loss of image quality was reported in situations where imaging targets did not match

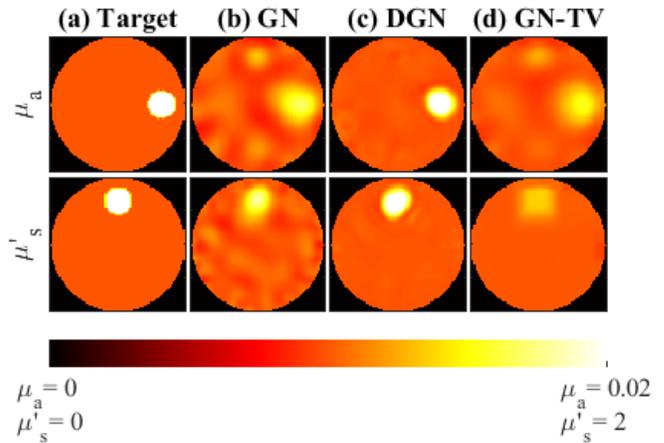


Fig. 11. (a) Target absorption (top) and scattering (bottom) distributions. The reconstructed distributions using (b) Gauss-Newton (GN) algorithm, (c) deep Gauss-Newton (DGN) algorithm and (d) Gauss-Newton algorithm using a total-variation prior (GN-TV).

the training data. Furthermore, the proposed approach was shown to effectively compensate for modelling errors. It was shown to provide improved computational time and estimation accuracy for non-Gaussian targets, compared to the Bayesian approximation error method.

## ACKNOWLEDGEMENT

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

## REFERENCES

- [1] S. R. Arridge, “Optical tomography in medical imaging,” *Inverse Problems*, vol. 15, no. 2, p. R41, 1999.
- [2] A. Gibson, J. Hebden, and S. R. Arridge, “Recent advances in diffuse optical imaging,” *Phys. Med. Biol.*, vol. 50, no. 4, p. R1, 2005.
- [3] T. Durduran, R. Choe, W. B. Baker, and A. G. Yodh, “Diffuse optics for tissue monitoring and tomography,” *Rep. Prog. Phys.*, vol. 73, p. 076701, 2015.
- [4] D. Grosnick, H. Rinneberg, R. Cubeddu, and P. Taroni, “Review of optical breast imaging and spectroscopy,” *J. Biomed. Opt.*, vol. 21, no. 10, p. 091311, 2016.
- [5] J. M. Cochran, D. R. Busch, L. Lin, D. L. Minkoff, M. Schweiger, S. Arridge, and A. G. Yodh, “Hybrid time-domain and continuous-wave diffuse optical tomography instrument with concurrent, clinical magnetic resonance imaging for breast cancer imaging,” *J. Biomed. Opt.*, vol. 24, no. 5, p. 051409, 2019.
- [6] J. He, C. L. Li, B. C. Wilson, C. J. Fisher, S. Ghai, and R. A. Weersink, “A clinical prototype transrectal diffuse optical tomography (trdot) system for in vivo monitoring of photothermal therapy (ptt) of focal prostate cancer,” *IEEE Trans. Biomed. Eng.*, vol. 67, no. 7, pp. 2119–2129, 2019.
- [7] Z. Jiang, D. Piao, G. Holyoak, J. Ritchey, K. Bartels, G. Slobodov, C. Bunting, and J. Krasinski, “Trans-rectal ultrasound-coupled spectral optical tomography of total hemoglobin concentration enhances assessment of the laterality and progression of a transmissible venereal tumor in canine prostate,” *Urology*, vol. 77, no. 1, pp. 237–242, 2011.
- [8] J. C. Hebden, A. Gibson, R. M. Yusof, N. Everdell, E. M. Hillman, D. T. Delpy, S. R. Arridge, T. Austin, J. H. Meek, and J. S. Wyatt, “Three-dimensional optical tomography of the premature infant brain,” *Phys. Med. Biol.*, vol. 47, no. 23, p. 4155, 2002.
- [9] Y. Hoshi and Y. Yamada, “Overview of diffuse optical tomography and its clinical applications,” *J. Biomed. Opt.*, vol. 21, no. 9, p. 091312, 2016.
- [10] M. D. Wheelock, J. P. Culver, and A. T. Eggebrecht, “High-density diffuse optical tomography for imaging human brain function,” *Review of Scientific Instruments*, vol. 90, no. 5, p. 051101, 2019.

- [11] J. Chen, "Optical tomography in small animals with time-resolved monte carlo methods," *Dissertation Abstracts International*, vol. 74, no. 04, 2012.
- [12] S. R. Arridge, M. Schweiger, M. Hiraoka, and D. T. Delpy, "Performance of an iterative reconstruction algorithm for near-infrared absorption and scatter imaging," in *Photon Migration and Imaging in Random Media and Tissues*, vol. 1888. International Society for Optics and Photonics, 1993, pp. 360–371.
- [13] C. Shaw and P. Yalavarthy, "Performance evaluation of typical approximation algorithms for nonconvex  $\ell_p$ -minimization in diffuse optical tomography," *J Opt Soc Am A*, vol. 31, no. 4, pp. 852–862, 2014.
- [14] T. Bhowmik, H. Liu, Z. Ye, and S. Orintara, "Dimensionality reduction based optimization algorithm for sparse 3-d image reconstruction in diffuse optical tomography," *Scientific reports*, vol. 6, no. 1, pp. 1–13, 2016.
- [15] K. D. Paulsen and H. Jiang, "Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization," *Appl. Opt.*, vol. 35, no. 19, pp. 3447–3458, 1996.
- [16] J. C. Ye, K. J. Webb, C. A. Bouman, and R. P. Millane, "Optical diffusion tomography by iterative-coordinate-descent optimization in a bayesian framework," *JOSA A*, vol. 16, no. 10, pp. 2400–2412, 1999.
- [17] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*. Springer, Newyork, 2005.
- [18] S. Diamond, T. Huppert, V. Kolehmainen, M. Franceschini, J. Kaipio, S. Arridge, and D. Boas, "Dynamic physiological modeling for functional diffuse optical tomography," *NeuroImage*, vol. 30, pp. 88–101, 2006.
- [19] S. R. Arridge, J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen, "Approximation errors and model reduction with an application in optical diffusion tomography," *Inv. Probl.*, vol. 22, no. 1, pp. 175–195, Feb. 2006.
- [20] J. Heino, E. Somersalo, and J. Kaipio, "Compensation for geometric mismodelling by anisotropies in optical tomography," *Opt. Express*, vol. 13, no. 1, pp. 296–308, Jan 2005.
- [21] T. Tarvainen, V. Kolehmainen, A. Pulkkinen, M. Vauhkonen, M. Schweiger, S. R. Arridge, and J. P. Kaipio, "An approximation error approach for compensating for modelling errors between the radiative transfer equation and the diffusion approximation in diffuse optical tomography," *Inv. Probl.*, vol. 26, no. 1, p. 015005, jan 2010.
- [22] M. Mozumder, T. Tarvainen, J. P. Kaipio, S. R. Arridge, and V. Kolehmainen, "Compensation of modeling errors due to unknown domain boundary in diffuse optical tomography," *JOSA A*, vol. 31, no. 8, pp. 1847–1855, 2014.
- [23] S. Arridge and J. Schotland, "Optical tomography: forward and inverse problems," *Inverse Problems*, vol. 25, p. 123010 (59pp), 2009.
- [24] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Imag. Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.
- [25] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," *Medical physics*, vol. 44, no. 10, pp. e360–e375, 2017.
- [26] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, no. 12, p. 124007, 2017.
- [27] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, 2017.
- [28] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [29] A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu, and J. A. Steeden, "Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning—proof of concept in congenital heart disease," *Magnetic resonance in medicine*, vol. 81, no. 2, pp. 1143–1156, 2019.
- [30] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge, "Model-based learning for accelerated, limited-view 3-d photoacoustic tomography," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1382–1393, 2018.
- [31] N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning photoacoustic tomography with sparse data," *Nature Machine Intelligence*, vol. 1, no. 10, pp. 453–460, 2019.
- [32] A. Hauptmann and B. T. Cox, "Deep learning in photoacoustic tomography: Current approaches and future directions," *J. Biomed. Opt.*, vol. 25, no. 11, p. 112903, 2020.
- [33] J. Feng, Q. Sun, Z. Li, Z. Sun, and K. Jia, "Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography," *J. Biomed. Opt.*, vol. 24, no. 5, p. 051407, 2018.
- [34] Y. Fan and L. Ying, "Solving optical tomography with deep learning," *arXiv preprint arXiv:1910.04756*, 2019.
- [35] S. Sabir, S. Cho, Y. Kim, R. Pua, D. Heo, K. H. Kim, Y. Choi, and S. Cho, "Convolutional neural network-based approach to estimate bulk optical properties in diffuse optical tomography," *Appl. Opt.*, vol. 59, no. 5, pp. 1461–1470, 2020.
- [36] S. J. Hamilton and A. Hauptmann, "Deep d-bar: Real-time electrical impedance tomography imaging with deep neural networks," *IEEE transactions on medical imaging*, vol. 37, no. 10, pp. 2367–2377, 2018.
- [37] S. J. Hamilton, A. Hänninen, A. Hauptmann, and V. Kolehmainen, "Beltrami-net: domain-independent deep D-bar learning for absolute imaging with electrical impedance tomography (a-EIT)," *Physiological measurement*, vol. 40, no. 7, p. 074002, 2019.
- [38] J. Yoo, S. Sabir, D. Heo, K. H. Kim, A. Wahab, Y. Choi, S.-I. Lee, E. Y. Chae, H. H. Kim, Y. M. Bae *et al.*, "Deep learning diffuse optical tomography," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 877–887, 2019.
- [39] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. Academic, Newyork, 1978.
- [40] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. the MIT Press, Cambridge, MA, 2006.
- [41] M. Schweiger and S. R. Arridge, "The Toast++ software suite for forward and inverse modeling in optical tomography," *J. Biomed. Opt.*, vol. 19, no. 4, p. 040801, 2014.
- [42] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.
- [43] W. Herzberg, D. Rowe, A. Hauptmann, and S. Hamilton, "Graph convolutional networks for model-based learning in nonlinear inverse problems," *IEEE Transactions on Computational Imaging*, DOI: 10.1109/TCI.2021.3132190, 2021.
- [44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [45] S. L. Jacques, "Optical properties of biological tissues: a review," *Phys. Med. Biol.*, vol. 58, no. 11, p. R37, 2013.
- [46] I. Nissilä, T. Nojonen, K. Kotilahti, T. Katila, L. Lipiainen, T. Tarvainen, M. Schweiger, and S. Arridge, "Instrumentation and calibration methods for the multichannel measurement of phase and amplitude in optical tomography," *Rev. Sci. Instrum.*, vol. 76, no. 4, p. 044302, 2005.
- [47] V. Kolehmainen, M. Schweiger, I. Nissilä, T. Tarvainen, S. R. Arridge, and J. P. Kaipio, "Approximation errors and model reduction in three-dimensional diffuse optical tomography," *JOSA A*, vol. 26, no. 10, pp. 2257–2268, 2009.
- [48] M. Mozumder, T. Tarvainen, A. Seppänen, I. Nissilä, S. R. Arridge, and V. Kolehmainen, "Nonlinear approach to difference imaging in diffuse optical tomography," *J. Biomed. Opt.*, vol. 20, no. 10, p. 105001, 2015.
- [49] D. Calvetti and E. Somersalo, "Hypermodels in the Bayesian imaging framework," *Inverse Problems*, vol. 24, no. 3, p. 034013, 2008.
- [50] S. Lunz, A. Hauptmann, T. Tarvainen, C.-B. Schonlieb, and S. Arridge, "On learned operator correction in inverse problems," *SIAM Journal on Imaging Sciences*, vol. 14, no. 1, pp. 92–127, 2021.
- [51] D. Smyl, T. N. Tallman, J. A. Black, A. Hauptmann, and D. Liu, "Learning and correcting non-gaussian model errors," *Journal of Computational Physics*, vol. 432, p. 110152, 2021.
- [52] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D Nonlinear Phenomena*, vol. 60, pp. 259–268, Nov. 1992.