

Multitask Deep Learning Reconstruction and Localization of Lesions in Limited Angle Diffuse Optical Tomography

Hanene Ben Yedder, Ben Cardoen, *Student member, IEEE*, Majid Shokoufi, Farid Golnaraghi, and Ghassan Hamarneh, *Senior Member, IEEE*

Abstract—Diffuse optical tomography (DOT) leverages near-infrared light propagation through tissue to assess its optical properties and identify abnormalities. DOT image reconstruction is an ill-posed problem due to the highly scattered photons in the medium and the smaller number of measurements compared to the number of unknowns. Limited-angle DOT reduces probe complexity at the cost of increased reconstruction complexity. Reconstructions are thus commonly marred by artifacts and, as a result, it is difficult to obtain an accurate reconstruction of target objects, e.g., malignant lesions. Reconstruction does not always ensure good localization of small lesions. Furthermore, conventional optimization-based reconstruction methods are computationally expensive, rendering them too slow for real-time imaging applications. Our goal is to develop a fast and accurate image reconstruction method using deep learning, where multitask learning ensures accurate lesion localization in addition to improved reconstruction. We apply spatial-wise attention and a distance transform based loss function in a novel multitask learning formulation to improve localization and reconstruction compared to single-task optimized methods. Given the scarcity of real-world sensor-image pairs required for training supervised deep learning models, we leverage physics-based simulation to generate synthetic datasets and use a transfer learning module to align the sensor domain distribution between in silico and real-world data, while taking advantage of cross-domain learning. Applying our method, we find that we can reconstruct and localize lesions faithfully while allowing real-time reconstruction. We also demonstrate that the present algorithm can reconstruct multiple cancer lesions. The results demonstrate that multitask learning provides sharper and more accurate reconstruction.

Index Terms—Diffuse optical tomography, image reconstruction, deep learning, spatial-wise attention, distance transform, lesion localization, multitask Learning, transfer learning, handheld probe.

I. INTRODUCTION

NEAR-infrared diffuse optical tomography is an optical imaging technique used for the examination of biological tissue at a macroscopic scale and has witnessed increased interest as a non-invasive and low-cost breast cancer screening alternative to ionizing X-ray mammography, the de facto screening technique [1], [2]. DOT projects near-infrared light from one or more light sources and acquires boundary data of the tissue (back-scattered light) using a set of detectors yielding, after reconstruction, cross-sectional or volumetric images of tissue [3]. The reconstruction process converts information from the projection (sensor) domain into the image domain to enable interpretation and diagnosis by domain experts or by artificial intelligence (Fig. 1).

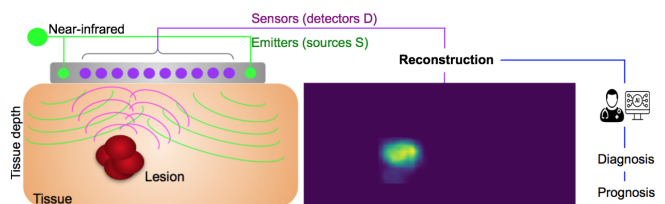


Fig. 1. DOT image reconstruction workflow example. A DOT-Scan probe consists of two near-infrared sources for tissue illumination and a set of detectors registering the back-scattered photons. Light propagation and scattering in the tissue are schematized. DOT reconstructed image shows the optical coefficients in the tissue. The success of diagnosis and treatment relies on accurate reconstruction and estimation of the optical properties of a medium.

DOT allows quantitative imaging of optical properties as it measures and visualizes the distribution of tissue absorption and scattering properties. These optical parameters are correlated to physiological markers such as blood oxygenation and tissue metabolism, allowing quantitative assessment of tissue malignancy [4].

Accepted on September 21 2021.

Correspondence to be addressed to {hbenyedd, hamarneh}@sfu.ca

This study was conducted through the Fraser Health-ethical approval reference number FHREB 2014-065 – PI Dr. Farid Golnaraghi– and supported in part by grants from Natural Sciences and Engineering Research Council of Canada (NSERC). We thank Nvidia Corporation for the donation of Titan X GPUs used in this research. This research was enabled in part by support provided by Westgrid (www.westgrid.ca) and Compute Canada (www.computeCanada.ca). The authors are grateful for the reviewers’ feedback which greatly improved the manuscript. The authors also thank Dr. Jeremy Kawahara and Ashish Sinha, from the Medical Image Analysis Lab, for their suggestions during paper revision.

Hanene Ben Yedder, Ben Cardoen, and Ghassan Hamarneh are with the Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, BC Canada V5A 1S6. e-mail: {hbenyedd, bcardoen, hamarneh}@sfu.ca

Majid Shokoufi and Farid Golnaraghi are with School of Mechatronic Systems Engineering, Simon Fraser University, BC Canada V5A 1S6. e-mail: {mshokouf, mfgolnar}@sfu.ca

A. Challenges in DOT Image Reconstruction

Given that the optical properties of normal and diseased tissues diverge, it is possible to detect breast cancer lesions using these optical properties reconstructed from collected boundary measurements of light propagation within a tissue [5]. The key challenge lies in properly determining optical properties given the nonlinear physics of photon scattering and the ill-posedness of the problem. Furthermore, although near-infrared light can propagate several centimeters inside the tissue, photons scatter many times and penetrate along random paths through the tissue before reaching the detectors, which makes the reconstruction task difficult.

When applied to limited-view data acquisition, allowing reduction of scanner complexity, reconstruction becomes even more challenging (smaller number of measurements compared to the number of unknowns) [6]. A handheld probe [1] can reduce system complexity, such as lower hardware and software requirements. However, the colinear placement of sources and detectors, in some probe designs (e.g., probe shown in Figure 1), can result in a limited view of the target and artifacts.

B. Current DOT Image Reconstruction Approaches

The DOT image reconstruction is an ill-posed problem that has been tackled using a wide variety of reconstruction methods (see [7], [8] for a review).

1) *Conventional Methods*: While absolute imaging relies on a single set of measurements to reconstruct spatially distributed optical coefficients, difference imaging aims at recovery of the change in the tissue optical properties based on measurements before and after the change [9]. The reference measurement can be the previous state (e.g., the rest stage in brain DOT) or a reference tissue or phantom. A drawback of difference imaging is that reconstructed images are usually only qualitative [10]. In this manuscript, focus is given to absolute imaging.

The image reconstruction problem is commonly formulated as recovering the optical properties distribution \hat{x}^* (2D image or 3D volume) that minimizes the reconstruction error between the boundary measurement sample y and the forward projection $\mathcal{F}(\cdot)$ from a possible reconstructed image \hat{x} :

$$\hat{x}^* = \underset{\hat{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathcal{F}(\hat{x}) - y\|_2^2 + \lambda \mathcal{R}(\hat{x}) \right\}, \quad (1)$$

where the first term aims to enforce data fidelity, $\mathcal{R}(\cdot)$ is a regularization term used to constrain the inverse problem to yield plausible solutions, and λ is a hyper-parameter that controls the contribution of \mathcal{R} . Common regularization choices include, for instance, standard Tikhonov regularization $R(x) = \lambda \|x\|_2^2$ and total variation (TV) regularization $R(x) = \lambda \|\nabla x\|_1$.

Until recently the focus in DOT reconstruction research has been on model-based algorithms, whose design follows directly from the underlying mathematical problem formulation. A disadvantage of a model based approach is that each instance needs to be optimized independently at reconstruction time, a computationally costly approach that can prohibit real-time application [3]. Reconstruction methods can roughly be

divided into linear approaches, based on inverse scattering theory, and nonlinear iterative approaches, based on model fitting. Linear approaches rely on Born or Rytov approximations, where the linearization is developed from analytical solutions of the diffuse equation for a homogeneous background [3]. This model can be constructed using Green's functions for the diffusion equation. The resultant linear system is usually solved by an iterative method (e.g., conjugate gradients [11]). The Born or Rytov approximation assumes that relative changes induced by anomalies are sufficiently small. When the values of the optical coefficient changes become very large, as in the case of large abnormal lesions, this approximation can lose accuracy in reconstruction [12].

Non-linear methods consider the model in terms of explicit parameters and adjusts these parameters in order to optimize the objective function in (1). Methods usually minimize an objective function in an iterative manner until convergence where the distribution of optical properties is searched for by minimizing the difference between the measured and the modelled data. A common approach in iterative reconstruction is to solve the forward model and calculate the Jacobian matrix at each iteration (e.g., Newton-like methods [13]). Alternatively, gradient-based reconstruction techniques use the gradient of the objective function to update the solution offering a reduction in computational complexity, at the cost of slower convergence. Regularization terms have to be added generally as the optimization problems are ill-posed.

Optimization of acquisition time, computational speed along with reconstruction quality remain an active research area in DOT. Despite recent advances, the reconstruction time of conventional DOT reconstruction methods is too high to be suitable for real time application [4]. In this manuscript, we consider throughput of at least 20 images (frames) per second as a minimum real-time equivalent for human experts [14], which corresponds with a runtime per image of < 0.05 s. An ideal frame rate of 60 (or 0.016 s) is the minimum sufficient rate to perceive no flickering for normal images [15].

Alternatively, the image reconstruction task can be learned using a deep neural network with a significant off-line training time, up to weeks during prototyping where each model takes a few hours of training, that is offset by a fast inference time as reported in Ben Yedder et al. [16].

2) *Deep Learning Methods*: Recently, deep learning has emerged as a powerful alternative for biomedical image reconstruction [16]–[18] and shown potential in improving resolution accuracy and speeding up reconstruction results especially in the presence of noisy and limited view acquisition. The capability of deep neural networks to learn complex functions automatically with fast inference makes their application to this problem ideal.

Deep neural networks-based techniques have shown promising potential in solving the inverse scattering problem [12], [19]–[21]. Preliminary work [19], [20], [22], did not consider end-to-end models but rather two-step approaches where a first image estimate, approximated using conventional methods, is further enhanced using a deep learning model. Recent end-to-end models typically considered a full-view acquisition setup: a circular shape scanner with more than 16 point

sources uniformly distributed along the boundary to maximize the number of measurements. For instance, Sun et al. [19] addressed the multiple scattering problem of microwaves in biological samples using a two-step reconstruction method, where an analytical method providing the first image estimate is followed by a convolutional neural network (CNN) for image reconstruction. Feng et al. [21] proposed a multilayer perceptron (MLP) feed-forward artificial neural network for a full-view 2D-DOT image reconstruction. However, performance degrades significantly for limited-angle acquisition where the number of sources is smaller and their placement is co-linear. Recently, based on the Lippman-Schwinger equation and deep convolutional framelets for inverse problems [23], Yoo et al. [12] designed a deep learning approach for 3D inverse scattering problems. The proposed model is based on an autoencoder CNN for a 3D-DOT inverse problem to learn the nonlinear physics of the inverse scattering problem.

Preliminary versions of our work [24], [25] tackled reconstruction of biological tissue parameters from limited-view data acquisition in a strong scattering medium using a decoder-like fully convolutional neural network architectures (FCN and FCN-FJ). In [25], a weighted combination of mean squared error (MSE) and Fuzzy Jaccard were used as loss function. Models were tested on phantom data only. While advancements have been made, reconstructions are still plagued by artifacts (e.g., false positive reconstructed pixels) and deviations from the ground truth location of the lesion, limiting their practical clinical utility. Furthermore, accurate localization is not necessarily guaranteed by a ‘good’ reconstruction, especially given the often small ratio of lesion versus background. For accurate diagnosis (e.g., via lesion biopsy) and treatment planning, the location and size of the lesion are paramount.

C. Multitask learning

Jointly solving for different tasks using a unified model is frequently considered in the computer vision field and has been shown to improve results in different applications, e.g., taskonomy [26] and skin-multitask [27]. Zamir et al. [26] showed that models optimized to jointly predict several related tasks perform better than models trained on individual tasks separately. Similarly, Kawahara et al. [28] proposed a single model capable of performing multiple tasks on multimodal skin lesion classification while Abhishek et al. [29] reported a performance improvement in skin cancer lesion management prediction when incorporating multitask learning. Sbalzarini [30] asserted that solving ill-posed problems in sequence can be more efficient and easier than in isolation as the overall solution space contracts since jointly solved problems constrained each other’s input/output space. Arguably, a unified framework allows feature space sharing and joint parameters tuning for both tasks and where the two problems regularize each other when tackled jointly. In our application as well, we illustrate that a multitask approach leads to a better reconstruction and localization, enabling more accurate diagnosis.

D. Contributions

We make the following contributions in this paper:

(i) To the best of our knowledge, we are the first to propose multitask learning for DOT reconstruction. We leverage a spatial attention mechanism in combination with a distance transform based loss function. Our network, Multitask-Net, exploits deep layers of spatial-wise attention [31], [32] to attend to important features by filtering irrelevant and noisy ones. Adaptively re-weighting features, according to their inter-dependencies in feature space, improves the representation ability of our model. Crucially, the distance transform based loss improves lesion reconstruction and localization.

(ii) We leverage a physics-based optical diffusion simulation to generate *in silico* training datasets and evaluate results on real measurements on physical phantoms and clinical data collected with the diffuse optical spectroscopy probe for breast tissue (DOB-Scan) [1], [33].

(iii) To the best of our knowledge, this is the first work that tests deep learning based DOT image reconstruction on clinical data of real patients.

(iv) To bridge the gap between real word acquisition and *in silico* data simulation, transfer learning is used. We evaluate the generalization ability of our model to unseen data subject to sensor non-idealities and noise.

(v) Extensive experiments on network characteristics trade-offs’ and scanning probe characteristics are investigated. The performance evaluation of our proposed model shows that our framework improves reconstruction and localization accuracy when compared against competing methods.

In section II, we introduce our proposed method for multitask reconstruction and localization learning followed by domain adaptation learning. Physics-based optical diffusion simulation test dataset, and competing methods are detailed in Section III. Trade-offs of network characteristics and evaluation on various datasets are described in Section IV. *In silico* performance results are presented in Section IV-A, results on physical phantoms in Section IV-B, and results on real-world data in Section IV-C. We conclude the paper by discussing insights and limitations on lesion attention-directed reconstruction in DOT in Section V.

II. METHODOLOGY

We aim to reconstruct an image directly from the sensor domain using a deep neural network. Given a set of raw DOT acquisitions $y \in \mathbb{R}^{S \times D}$, from S sources (emitters) with D sensors (detectors), and their corresponding tomographic image $x \in \mathbb{R}^{W \times H}$, which represents the tissue’s optical coefficients (Fig. 1), we seek to learn the inverse function $\mathcal{F}^{-1}(\cdot)$ that maps between sensor measurements and image domain while remaining true to the constraints of the underlying physics of the inverse problem. Therefore, we seek to learn the inverse function $\mathcal{F}^{-1}(\cdot)$ in an end-to-end fashion that solves:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(\mathcal{F}^{-1}(y; \theta); x) + \lambda \mathcal{R}(\mathcal{F}^{-1}(y; \theta)), \quad (2)$$

where \mathcal{L} is the network loss function that penalizes the dissimilarity between the reconstructed image and the ground truth. \mathcal{R} is a regularization term that enforces prior, θ are the optimized network parameters or weights.

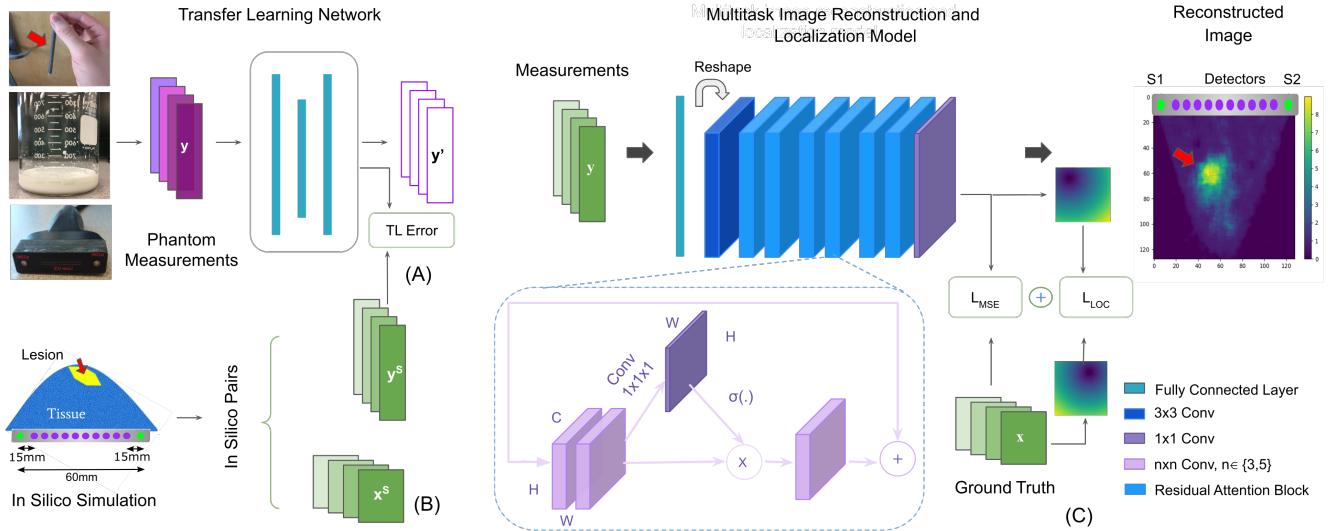


Fig. 2. The overall architecture of the proposed model. (A) Transfer learning network used for data distribution adaptation between in silico measurements y^s and real data collected from the probe y . (B) The generation of in-silico training data pairs (ground truth images x^s and their corresponding forward projection measurements y^s) approximating our probe’s specification [34]. The red arrow represents the simulated lesion. (C) Architecture of the proposed deep residual attention model. The squeeze and excitation block (inset) squeezes a given feature map U along the channels and excites spatially. The output of the sigmoid layer $\sigma(\cdot)$ represents the relative importance of the spatial information (i, j) of the feature map U . All feature maps produced by all convolutional layers (conv) are set to size 128×128 using (3×3) followed by (5×5) kernels. The similarity-wise loss L_{Loc} , leveraged to enforce localization accuracy, computes the mean absolute error between the distance transform of the reconstructed and ground truth images.

A. Reconstructing Images from DOT Measurements

We extend our previously proposed decoder-like architecture FCN [24] and FCN-FJ [25] consisting of a fully connected layer followed by a set of convolution layers, with a spatial-wise attention mechanism and a new loss component (Fig. 2-C). The fully connected layer maps the measurement vector y onto a two-dimensional array by learning a weighted combination of the different receptive sensors based on the signal collected from the back-scattered light collected at different locations in the reconstructed tissue, and therefore provides an initial image estimate. Empirically, we did not witness any improvements in the reconstruction results using more than one fully connected layer. This may be due to the size of the input measurement vector, which is only 256×1 dimensional in our dataset. Higher-dimensional inputs may benefit from additional layers. The output of this layer is reshaped to an 128×128 array and delivered to the subsequent layers for optimization. Then, a set of residual convolutional blocks (6 blocks), with 32 channels and increasing filters size (3, 5, 7), refines the first image via nonlinear transformations and produces the final reconstruction image x . We refer the reader to the Appendix-A for the mathematical formulation.

While the mean squared error was used in isolation in our preliminary FCN work [24], a Fuzzy Jaccard loss component \mathcal{L}_{FJ} was added later FCN-FJ [25] to address the imbalance between the minority pixels corresponding to one or more lesions versus a majority of ‘background’ pixels corresponding to healthy tissue.

B. Spatial-wise Attention Network

The framework of our proposed residual attention network (RAN) is shown in Fig. 2-C. The first and last convolu-

tional layers are a shallow feature extractor and a reconstruction layer, respectively. The middle layers are residual attention blocks used to extract hierarchical attention-aware features [31], [35]. Since the contributions of features to the reconstruction task vary from different feature maps, we leverage a spatial attention mechanism to prune the irrelevant features and enhance the informative ones by adaptively re-weighting features according to their inter-dependencies in feature space. The spatial attention compresses the feature map U along the channels and excites it spatially to guide the network to pay more attention to the regions of interest; lesion localization in the case of DOT reconstruction. The attention mechanism has been shown effective for natural language processing [36], [37] and computer vision problems tasks [31], [32] as it selectively disregards the noise (less important information) and focuses on what is relevant. Empirically, we did not observe any improvements in the reconstruction results using channel-wise attention [31]. Each residual block uses sequential modules of the form: two convolutions – batch normalization – ReLU and zero-padding followed by a squeeze and excite modules [32].

The spatial squeeze operation is achieved through a convolution generating a projection tensor $q = W \times U$ with weight $W \in \mathbb{R}^{1 \times 1 \times C}$ where each element of the tensor $q \in \mathbb{R}^{H \times W}$ represents the linearly combined representation for all C channels. The projection is then passed through a sigmoid layer σ to rescale activations to $[0, 1]$, which are thereafter used to re-calibrate U spatially:

$$F_{SA}(U) = [\sigma(q_{1,1}) \mathbf{u}^{1,1}, \dots, \sigma(q_{i,j}) \mathbf{u}^{H,W}] \quad (3)$$

Where $\sigma(q_{i,j})$ represents the relative importance of a spatial information (i, j) of a given feature map. As a result, the focus

is guided to the lesion location at the cost of loss of attention to healthy tissue, a sacrifice justified given their relative lower importance to the clinician when diagnosis is the end task.

C. Similarity-wise Loss Function

To guide the learning of the RAN network, we propose a novel loss function that combines two loss terms: a mean square error that guides the loss to reconstruct the pixel-wise representation of the image and a location based term to strengthen lesion localization accuracy:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \beta \mathcal{L}_{\text{loc}}, \quad (4)$$

where \mathcal{L}_{MSE} is the mean square error loss function commonly used in image reconstruction tasks, \mathcal{L}_{loc} is the location-based loss term, and β is a hyper-parameter balancing the contribution of the \mathcal{L}_{loc} term. We first allow the network to learn to reconstruct, via \mathcal{L}_{MSE} , an image estimate that is relatively close to the ground truth image pixel-wise distribution and then, via \mathcal{L}_{loc} , refine that candidate image using $\beta(\text{epoch} + \Delta) = \beta_0 + \gamma\beta(\text{epoch})$ with $\gamma > 0$.

1) *Reconstruction Loss*: \mathcal{L}_{MSE} measures the pixel-wise intensity difference between ground truth and reconstructed images and allows the network to learn an image estimate that is relatively close to the ground truth image pixel distribution. However, \mathcal{L}_{MSE} evaluates pixels in isolation, ignoring the overall image structure. Furthermore, \mathcal{L}_{MSE} can be biased by matching the background rather than the target of interest, a concern exacerbated in DOT image reconstruction of breast tissue with zero or more small isolated lesions, where the majority of the pixels are background (healthy tissue).

2) *Localization Loss*: The location-based loss term \mathcal{L}_{loc} is introduced to emphasize more accurate lesion localization. \mathcal{L}_{loc} is based on the Euclidean distance transform that maps an image to a similar grayscale image, except that the grayscale intensities of points inside the background region are changed to show the distance to the closest foreground boundary from each point. The distance transform is sensitive to small changes in size and position of the foreground object (e.g., offsetting an object by a single pixel can generally change all pixel values of the distance transform image). We leverage this sensitivity to penalize deviation in lesion localization in the reconstruction. \mathcal{L}_{loc} calculates the \mathcal{L}_1 distance between the distance transform of the ground truth and the reconstructed image pairs (Eq. (5), Fig. 2-C) and goes to zero when the reconstructed image overlaps with the ground truth. The location-based distance transform loss is given by:

$$\mathcal{L}_{\text{loc}}(\theta) = \|DT(\mathcal{F}^{-1}(y, \theta), x) - DT(x)\| \quad (5)$$

where

$$DT = \begin{cases} 0 & \text{if } a_{ij} \in S \\ D(a_{ij}, S) & \text{if } a_{ij} \in R_S, \end{cases} \quad (6)$$

and $D(a_{ij}, S)$ refers to the Euclidean distance between the image pixel location a_{ij} and lesion boundaries of a shape S . Lesion boundaries are computed during the training by simply thresholding the training images based on the mean and standard deviation ($\mu + \sigma$) of each image.

D. Transfer Learning Network

Training a deep learning based model can require a large dataset to converge to optimal performance. However, this is a challenging requirement with relatively new imaging devices such as DOB probes. Physics-based simulation data can provide an alternative source of training samples. Similar to Ben Yedder *et al.* [25] and Zhou *et al.* [38], where sensor data distribution adaptation is used to take advantage of cross-domain learning, we leverage transfer learning to bridge the gap between real word acquisition and *in silico* simulated data. An MLP (Fig. 2-A) tackles the domain shift between the real data measurement y^p collected from the probe and the *in silico* training measurement y^s . By minimizing the transfer learning loss \mathcal{L}_{TL} over N_p real data measurements collected from the probe using a phantom solution and their tissue equivalent simulated data, the transfer learning module learns the mapping $\phi(y_i^p) \approx y_i^s$ to ensure it is in the same domain as y^s :

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}_{\text{TL}}(\theta) \quad \text{where} \quad \mathcal{L}_{\text{TL}}(\theta) = \sum_{i=1}^{N_p} \|\phi(y_i^p; \theta) - y_i^s\|. \quad (7)$$

The final reconstructed image, at inference time, is computed as:

$$\hat{x}^* = \mathcal{F}^{-1}(\phi(y^p)). \quad (8)$$

III. DATASETS

To train our tomographic image reconstruction network, we collect training samples from synthesized tissue geometries with known optical properties using a physics-based simulation. We use our recently developed functional hand-held diffuse optical breast scanner (DOB-probe) [1] to collect real data of physical phantoms and from patients for testing our method. No phantom or patient data is used during training and validation.

A. *In silico* Training Data

In silico training data pairs, which include images of optical tissue properties and their corresponding detector measurements, were used to train our reconstruction network (Fig. 2-B). A set of synthetic 2D images with different lesion shapes, sizes, and locations, representing optical tissue properties, discretized into finite element nodes (2D triangular meshes), and their corresponding forward projection measurements were synthesized (Fig. 3).

The time-resolved optical absorption and scattering tomography software (Toast++) [40] was used to simulate the frequency-domain forward model and generate projection measurements for each training mesh. Frequency-domain models or intensity-modulated sources use a frequency ω for modulating the intensity of light. Frequency-domain systems provide information on the medium in the form of amplitude and phase data that permits the recovery of absorption and scattering coefficient parameters. The forward model uses the finite element method to solve the diffusion equation that

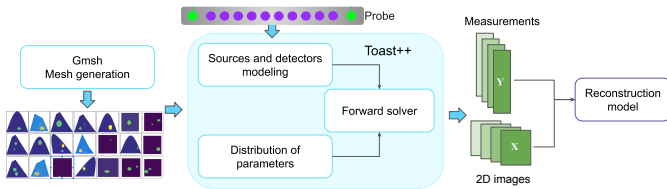


Fig. 3. Workflow of silico training data generation for DOT image reconstruction. 2D triangular meshes modeling tissue with different lesion sizes and depths were generated using the Gmsh software [39]. The Toast++ software suite was used to simulate the forward projection with a 750 nm wavelength modulation and optical properties approximating breast tissue. Functional handheld probe characteristics were modeled to obtain measurements that mimic real world values. Measurements (y) and ground truth images (x) pairs were saved and fed to the deep learning model for training.

models the light propagation in a diffusive medium. The diffusion equation is given by [40]:

$$-\nabla \cdot D(r) \nabla \varphi(r, \omega) + \left(\mu_a(r) + \frac{i\omega}{c_m(r)} \right) \varphi(r, \omega) = 0, r \in \Omega \setminus \partial\Omega \quad (9)$$

with boundary condition [40], [41]:

$$\varphi(\mathbf{m}, \omega) + 2\zeta(c_m)D(m) \frac{\partial \varphi(\mathbf{m}, \omega)}{\partial \nu} = q(\mathbf{m}, \omega), \quad \mathbf{m} \in \partial\Omega \quad (10)$$

where $\mu_a(r)$ and $D(r)$ denote the absorption and the diffusion coefficients at position r , respectively; $c_m(r)$ represents the speed of light in the medium; $q(r, \omega)$ represents the light source term; ζ is a term incorporating the refractive index mismatch at the tissue-air boundary; q is a source distribution on boundary $\partial\Omega$ of domain Ω ; $\partial\nu$ is the outward boundary normal; and $\varphi(r, \omega)$ is the generated photon density distribution with a modulation frequency ω . Synthesized meshes define the medium and distribution of parameters (absorption, scattering, refractive index).

Our functional hand-held probe geometry was modeled accurately in Toast++, as illustrated in Fig. 2-B, to obtain measurements that mimic real values obtained by the DOB probe [1], [34]. The probe is composed of 2 LED light sources, illuminating the tissue symmetrically and delivering near-infrared light to a body surface at different points, and featuring 128 co-linear equi-spaced detectors that measure the back-scattered light from the tissue and emitted from the boundary. Correspondingly, the output of the forward model simulation captures a 1D raw intensity diffraction (1×256) vector y^s resulting from the scattering of the illuminating light exiting the object.

A total of 4000 samples of (2D mesh, 256-D vector) data pairs are used. Realistic human breast tissue optical parameters' distribution values [42] were considered to mimic both healthy and cancerous tissue optical coefficients. The absorption coefficients of breast tissue vary between 0.0018 and 0.0025 mm^{-1} for healthy tissue and between 0.09 and 0.16 mm^{-1} for cancerous tissue. Similarly, the scattering coefficients vary between 0.8 and 1 mm^{-1} for healthy tissue and 1 and between 1.6 mm^{-1} for cancerous tissue. Given that relative optical parameters changes at healthy tissue are sufficiently small compared to optical perturbations at abnormalities, background optical parameters were assumed uniform

in each sample (the absorption coefficient of background is set to the average of breast tissue similar to [43], [44]) while abnormalities' size, depth, number, and their respective optical parameters were varied. Furthermore, since the absorption coefficients in the lesion are larger than the background, the absorption features due to the lesion emerge in outgoing light at the boundary. Therefore, similar to [45], we focused our in silico dataset on examples that have been shown to be hard to reconstruct and more difficult in treatment. The size of the reconstructed image is 6×6 cm with a resolution of 128×128 pixels.

The in silico data is designed to be consistent with existing prior art [12], [21], [46], where we vary diameters of two-dimensional circular heterogeneous masses representing lesions. We vary the radius from 3-15 mm and tune optical coefficients of background to match that of biological non-lesion tissue. We randomize the number of lesions (0-3), as well as their size, depth, and location to ensure a sufficiently rich dataset matching real world conditions. While simulated data can in principle be infinite in size, we found the current size of our in silico dataset to be sufficient for our purposes, in that we observed a decreasing return in performance with respect to increasing training time from increases in data set size.

A Gaussian noise layer is used before the dense layer to add independent zero-centered Gaussian noise to each measurement y_i . This noise model simulates the highly varying noise to each individual detector as caused by interference of refracting light as well as sensor noise. We choose $\sigma = 10\%$ of maximum signal value consistent with previous work [12], [19], [21]. For completeness, we note that in addition to this in silico noise, the probe accounts for ambient noise, recorded by capturing a frame without active emitters, which is then subtracted from real data measurement (phantom and clinical) before reconstruction.

B. Phantom and Clinical Test Data

Two test datasets were used in our experiments: An in silico phantom dataset to validate our method, and a clinical dataset to ensure generalizability to a clinical setting.

1) *Phantom Data*: To analyze the performance of our approach in controlled real experiments setting, breast-mimetic phantoms with known inhomogeneity locations are used (Fig. 2-A). The phantom data set is composed of a solution of intralipid 20% emulsion (Fresenius Kabi Inc.) that mimics background breast tissue with its optical properties [1], [47]. The cancerous lesion is simulated by a 4 mm diameter tube filled with a black ink liquid (Higgins India Ink) that matches the optical properties of tumor lesions [47]. The DOB-probe (Fig. 2-B) collects phantom measurements y^p . By varying the location of this simulated lesion, we can test the accuracy of the reconstruction method in a number of different scenarios.

2) *Clinical Data*: The data is collected from two participants diagnosed with breast cancer, following ethics and institutional review board approval protocol, and subsequently de-identified to protect patient privacy. For each patient, 2D images were scanned from two locations, the first is a sweep over the

cancerous lesion location and the second is a sweep over the same region on the opposite healthy breast, as a baseline. Figure 7-B illustrates the scan procedure. On average, five measurements were taken on each sweep. The precise location of the cancerous lesion was determined via mammography, ultrasound, showing calcification over the area, and biopsy.

C. Implementation

The model was implemented in the Keras [48] framework and trained for a total of 500 epochs, on an Nvidia Titan X GPU, with early stopping if the validation loss has not improved in the last 10 epochs. Adam optimizer with default parameters is used [49]. The optimal hyper-parameters such as the number of the filter channels, non-linearity, weights initialization, regularization term, and the batch size were found by trial and error to be optimal for our validation dataset during prototyping.

By optimizing the model’s performance on the validation set, we set all hyper-parameters as follows: batch size to 64; learning rate to 10^{-4} ; Adam optimizer set to true; and ($\Delta = 10$, $\beta_0 = 0.2$, $\gamma = 0.005$), which describe the update equation of the hyper-parameter β in (4); We use a 80/10/10 training/validation/test split of the in silico training data. The model was trained for the first 100 epochs using \mathcal{L}_{MSE} before introducing \mathcal{L}_{loc} to avoid noisy reconstruction. A \mathcal{L}_2 weights’ regularization of 10^{-4} was applied to the last residual layer of the network to avoid overfitting to training data. Transfer learning was applied at evaluation time in the phantom and real data while being trained using phantom measurements only. A separate phantom dataset was used to train the transfer learning network.

D. Competing Methods

We contrast our method results to eight state-of-the-art methods; four of which are conventional methods whereas the others are deep learning based methods. We compare to Least squares-based approach (LS) [50], Levenberg-Marquardt (LM) [40] and nonlinear conjugate gradient (NCG) optimization based methods [51], and analytical method [34]. The deep learning based methods are: MLP [21], FC-CNN [12], and our preliminary works FCN [24], and FCN-FJ [25]. Inverse solvers are constructed by making use of first derivatives such as NCG-based or second derivatives such as LM-based approaches. Employed algorithms are implemented in the state of the art public software packages of NIRFAST for LS and TOAST++ software for LM and NCG methods.

LS [50] and LM [40] are Jacobian-based approaches. The LM based method employs the modified Levenberg Marquardt (LM) [52], a trust region approach, to minimize the objective function over the search space of model parameters. At each iterative step the Jacobian matrix is recalculated and repeatedly inverted. Optical parameter estimates are then updated. The algorithm converges when the reconstructed optical parameter at the current iteration has not improved over the last two iterations. A first order Tikhonov penalty term is used to regularize reconstruction algorithms. The regularization parameter, found by parameter sensitivity study, is set to 0.01. We refer

the reader to the Appendix-B,C for mathematical formulation and parameters search details.

NCG is a gradient based method [51]. The objective function describes the difference between the calculated forward measurement and measured data. The inverse solver iteratively computes updates to an initial distribution of the absorption and scattering parameters via minimization of the gradient of the objective function. Total variation (TV) is used to regularise the solution and the object threshold parameter β , found by parameter sensitivity study, is set to 0.1. The finite element discretisation method is used to discretise the computational domain and break down the calculation domain onto the local elements individually. In this paper, object geometry is represented by a polygon over which a series of disjoint triangles are generated using the GMesh software suite [39]. We used Isotropic TV regularization [53], [54], as implemented in the Toast++ software suite [40]. For detailed configuration settings, please see appendix-B,C.

The analytical method in [34] is a difference approach that relies on a relative change between optical perturbations at abnormalities and healthy tissue. A homogenous data is collected using a phantom and used as a reference to measure any inhomogeneities in tissue. The difference in absorption between the reference (normal) and abnormal tissues can provide the imaging contrast for tissue diagnostics. Hypothetical curved paths, as explained in [34], are used to recover the absorption coefficient from each illuminated source. The relative change difference is back scattered, using a closed form solution for the diffusion equation, to recover the optical coefficients. Hence, the absorption coefficient at each coordinate (x, y) is calculated as a superposition of back scattered values along the paths of illuminated light from each source. The computed absorption coefficients are then used to recreate two-dimensional cross-sectional images for each wavelength. For the detailed mathematical model description, we refer the reader to [33], [34].

The MLP method [21] is a multilayer perceptron network with two fully connected layers. The FC-CNN [12] uses a fully connected layer followed by a CNN with an encoder-decoder structure with three convolutional layers. The hyperbolic tangent function (tanh) is used as an activation function for the fully connected layer and two first convolutional layers, and rectified linear unit (ReLU) for the last convolutional layer [12]. The FC-CNN implements a similar architecture to our baseline approach FCN [24] where the difference lies in the number of convolution layers, filter, and kernel size. Both architectures were re-implemented as specified in the introducing papers and trained using mean square error between input and output as the loss function. Our preliminary works FCN [19], and FCN-FJ [20] were briefly described in Section 2-A

IV. RESULTS

In order to measure the quality of the results, we consider: (i) Lesion localization error, i.e., the distance between the centre of the lesions in the ground truth image versus the reconstructed image; (ii) peak signal-to-noise ratio (PSNR),

$10 \log_{10} \frac{\text{MAX}_I}{\text{MSE}}$, where MAX_I is the maximum intensity value of all pixels; (iii) structural similarity index (SSIM) [55]; and (iv) the Fuzzy Jaccard [56].

A. Results on Synthetic Data

Trained on the in silico data and tested on a separate test set of 80 images, we visually evaluate the reconstruction of our Multitask-Net method and compare results with those of LS [50], NCG [51], MLP [21], and FC-CNN methods [12]. In Fig. 4, we show results on some in silico samples with different lesion sizes and depths. Interestingly, while both LS and NCG methods are able to detect lesion presence, LS outperforms in terms of lesion depths and size reconstruction. Images reconstructed by the MLP method provide a quite noisy reconstruction and a rough approximation of the lesion location. Further, filtering via CNN enabled better results by FC-CNN. Overall, our model provides more accurate results as it is able to differentiate between different lesion sizes and recover correctly lesion location and dimension and absorption coefficients despite the limited angle condition. The y-axis scales highlight our model superiority in reconstructing the optical coefficient values. The difference in performance can be explained by the ability of the convolution operators to extract more comprehensive contextual information and synthesize more complex features.

The last two columns in Fig. 4 (e-f) show reconstruction performance in a healthy tissue case as well as a deep lesion case (~ 5 cm in depth). Note the model's ability to capture and reconstruct a tumor-free signal. We illustrate the limits of our contribution where deep lesions cannot be distinguished from the background. This is expected given the rapidly reducing strength of reflected signals due to photons scattering many times in random paths before being absorbed by the tissue. A recent study [57] on self-diagnosis confirmed by imaging reported a mean depth of 1.6 cm with a maximum of 6 cm (163 patients, 173 lesions). While this does not preclude deeper lesions in larger population studies, it shows that our performance is aligned with real-world assumptions on lesion depth.

B. Results on Phantom Data

Trained on the in silico data and tested on the phantom dataset, we visually compare our proposed reconstruction method of absorption coefficients to the competing methods' results on sample phantom cases. We compare to analytical method [34], LS based method [50], LM based method [40], NCG based method [51], MLP model [21], FC-CNN [12], and our preliminary works FCN [24], and FCN-FJ [25].

1) *Qualitative Results on Phantom data:* In Fig. 5 and Fig. 6, we contrast our results with the state of the art on phantom data. Our DOT reconstruction model is mainly interested in the absorption parameter changes as it is related to the hemoglobin concentration changes that mark the presence of anomalies [4].

The analytical approach, while detecting the presence of the lesion, only approximates its location by a rough margin and fails to provide precise reconstructed optical coefficient values required for proper assessment of lesion (Fig. 5- Analytical).

LM and NCG based methods (Fig. 5- LM, NCG) provide good reconstruction as they are able to detect and localize lesions. Lower performance on lesion size can be in part attributed to noise and artifact on real data information. MLP provides a rough approximation of lesion localization and sizes with very low per pixel values of reconstructed absorption coefficients (Fig. 5- MLP). The FC-CNN and FCN-FJ, on the other hand, localize the lesion in a single concentrated area with some error in lesion size and optical coefficients (Fig. 5- FC-CNN and Fig. 6- FCN-FJ ($\mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{FJ}}$)). Adopting the new loss term \mathcal{L}_{loc} clearly reduces the lesion localisation error (Fig. 6- FCN ($\mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{FJ}} + \mathcal{L}_{\text{loc}}$)). Small artifacts are a disadvantage of this model. The proposed RAN consistently obtains results that minimize both location and size error while miss estimating values of reconstructed absorption coefficients (Fig. 6- RAN).

Observe how incorporating both the new loss term \mathcal{L}_{loc} and the RAN in the Multitask-Net significantly reduces the artifacts while improving lesion localization and reconstruction of optical coefficients, which otherwise could compromise diagnosis (Fig. 6- Multitask-Net). From these results, it is clear multitask learning is able to achieve superior performance in both tasks without compromising subtask performance.

2) *Quantitative Results on Phantom data:* Table I presents the results on the phantom dataset. Yoo et al. architecture presents close results to our preliminary work FCN [24] given the similarity between the used architectures, where the difference lies in terms of model depth and filter sizes. It is worth mentioning that some artifacts are due to the shift in distribution that we overcome using the transfer learning model.

Rows 5 to 8 present an ablation study that shows the contribution of each element of our model. While the distance transform loss allows reduction in lesion localization error, the attention network helps in terms of reconstruction similarity and false positive reduction. We obtain marked improvements in localization error (up to 18%), as well as improvements in SSIM (up to 4%), Fuzzy Jaccard (up to 5%), and PSNR (up to 3%). Interestingly, we found that a RAN performs better than an FCN+ \mathcal{L}_{Loc} in terms of Fuzzy Jaccard metric, hence we can argue that a RAN helps to better focus on lesion area and reduces the false positive pixel.

Computational time for each reconstruction in this study is reported in Table I (runtime). Once the deep learning model training is completed, the time for the reconstruction is in the order of a few ms, three orders of magnitude faster reconstruction than conventional methods thus enabling real time application. Note that the new loss function has little to no effect on the inference time.

C. Results on Clinical Patient Data

Figure 7 presents the reconstructed cross-sectional image of absorption coefficients of breast scans from 2 patients. The probe is placed above the identified lesion in the cancerous breast and at the same location on the opposite healthy breast for each patient (Fig. 7- A). We compare with the analytical [34], NCG [51], MLP [21], and FC-CNN [12] methods, reporting visual results in Fig. 7- B,C and quantitative results

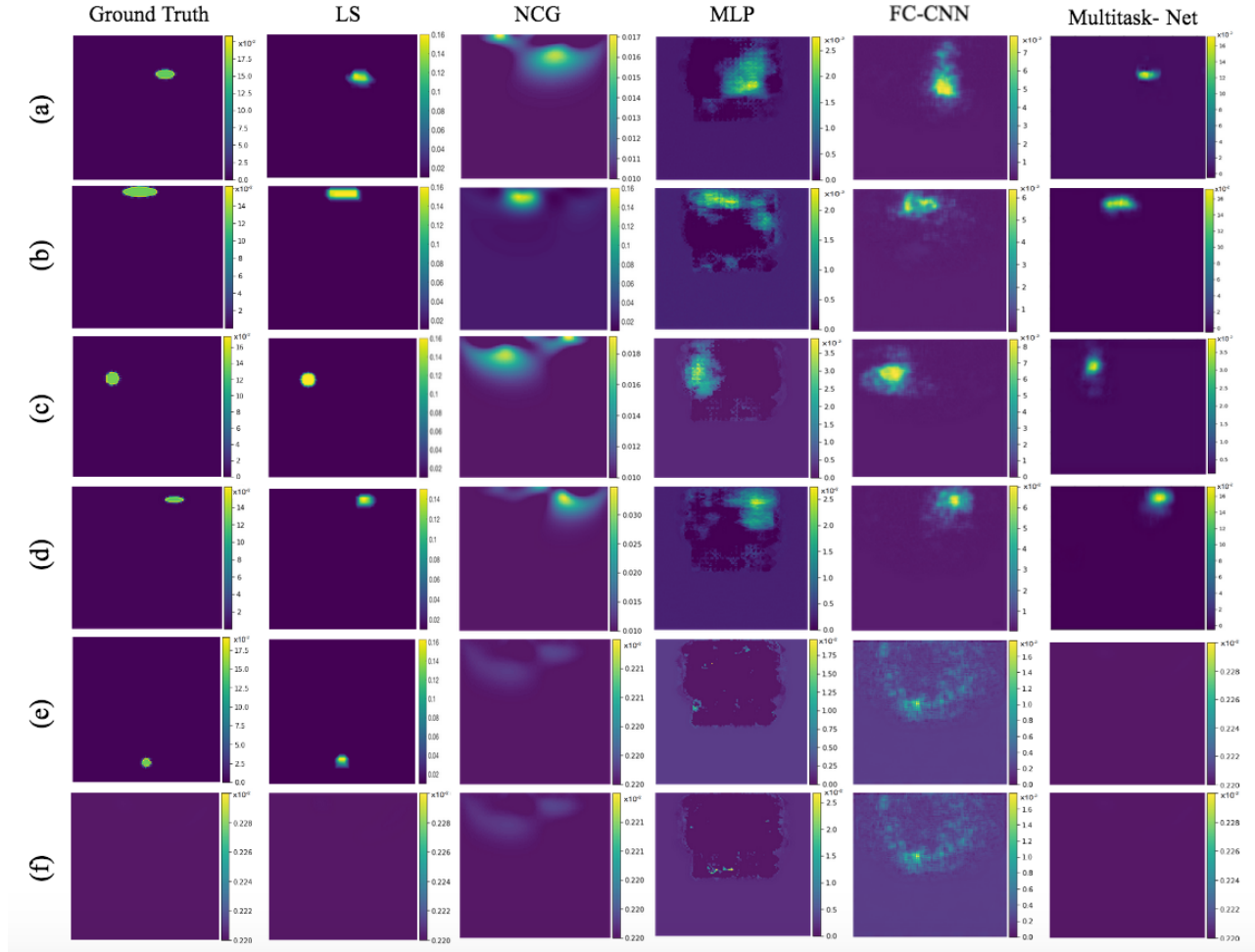


Fig. 4. Qualitative reconstruction performance of our model compared to LS, NCG, MLP, and FC-CNN models in silico samples with various ground truth lesions sizes, locations, and absorption coefficients (a-f). Different y-axis scales were needed to highlight the models' performance in reconstructing the optical coefficient values. Evidently, the images reconstructed by our method are more accurate than those reconstructed by the other approaches. The MLP approach presents a rough reconstruction that is further improved by FC-CNN. The NCG method reproduces lesion location, while lesion depth is less accurate. The LS method provides the best overall results.

TABLE I

QUANTITATIVE RESULTS ON 32 PHANTOM EXPERIMENTS. METHODS WITH A RUNTIME SMALLER THAN 20 IMAGES PER SECOND ($< 0.05s$) ARE CONSIDERED REAL-TIME CAPABLE.

		Loss			Loc. Error	PSNR	SSIM	Fuzzy Jaccard	Runtime
		L_{MSE}	L_{FJ}	L_{Loc}	(pixel, $<$ is better)	(dB, $>$ is better)	($>$ is better)	($>$ is better)	(s, $<$ is better)
Conventional	Analytical [34]	na			77.4 ± 32.2	15.0 ± 6.0	0.32 ± 0.26	0.17 ± 0.15	83×10^{-3}
	LS [40]	na			31.7 ± 11.7	13.7 ± 11	0.29 ± 0.07	0.16 ± 0.08	3×10^2
	LM [40]	na			25.2 ± 12.7	16.6 ± 8.6	0.37 ± 0.09	0.16 ± 0.05	5×10^2
	NCG [51]	na			32.6 ± 16.5	16.5 ± 9.2	0.33 ± 0.01	0.22 ± 0.13	2×10^2
	MLP [21]	✓	-	-	42.0 ± 17.3	12.5 ± 1.5	0.05 ± 0.03	0.07 ± 0.05	2×10^{-3}
	FC-CNN [12]	✓	-	-	27.8 ± 18.7	17.3 ± 7.2	0.38 ± 0.23	0.45 ± 0.11	5×10^{-3}
Deep Learning	FCN [24]	✓	-	-	33.2 ± 23.4	20.1 ± 4.6	0.46 ± 0.28	0.32 ± 0.06	8×10^{-3}
	FCN-FJ [25]	✓	✓	-	14.8 ± 7.40	21.7 ± 0.9	0.73 ± 0.03	0.64 ± 0.10	17×10^{-3}
	FCN+ L_{Loc}	✓	✓	✓	13.6 ± 7.20	21.3 ± 2.1	0.74 ± 0.21	0.63 ± 0.05	19×10^{-3}
	RAN	✓	-	-	14.3 ± 7.10	18.1 ± 6.1	0.76 ± 0.11	0.70 ± 0.04	20×10^{-3}
	Multitask-Net	✓	-	✓	12.1 ± 6.41	23.8 ± 1.7	0.77 ± 0.18	0.69 ± 0.09	23×10^{-3}

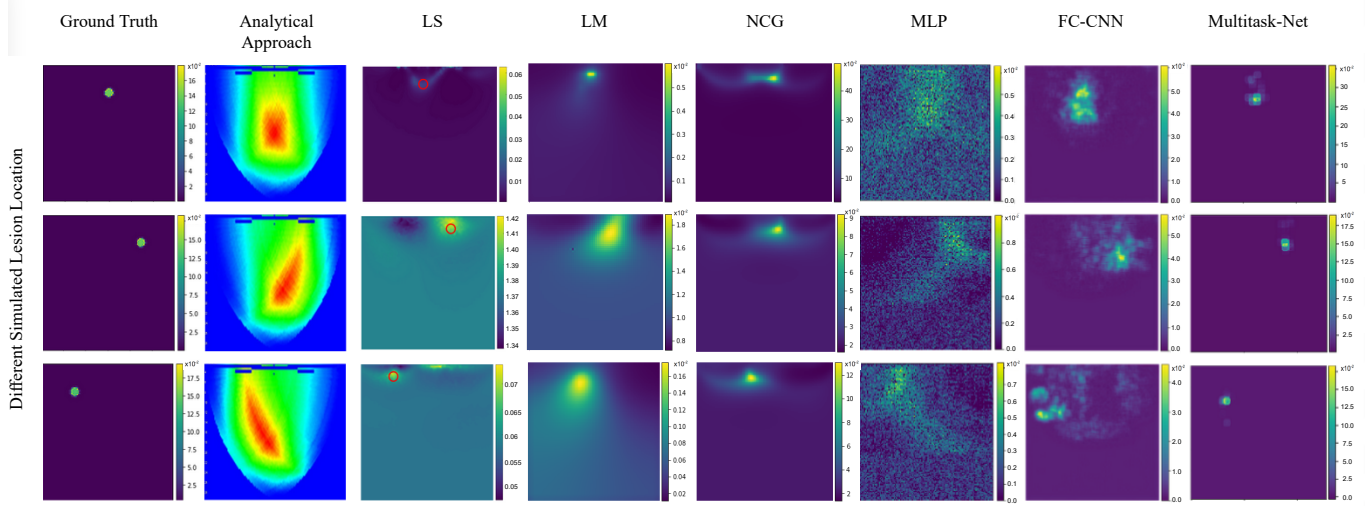


Fig. 5. Qualitative reconstruction performance of our model compared to state-of-the-art (Analytical, LS, LM, NCG, MLP, and FC-CNN) techniques on phantom samples with known lesion ground truth location. The x-axis and y-axis show lesion location and depth in mm, respectively. Color bar shows the reconstructed absorption coefficient values. Different y-axis scales are unfortunately needed and indicative of the problematic performance of some models in reconstructing the optical coefficient values. The reconstructed image using MLP is equivalent to a back projection operation and gives a rough approximation of lesion localization, while precision of the reconstructed absorption coefficients can be low. Adding convolution layers helps to better filter noise and reduce false-positives while still miss-predicting correct absorption coefficient values required for proper diagnosis. The LM and NCG methods provide good reconstruction with accurate lesion presence detection. Lesion size mis-prediction can be attributed to real data noise and artifacts.

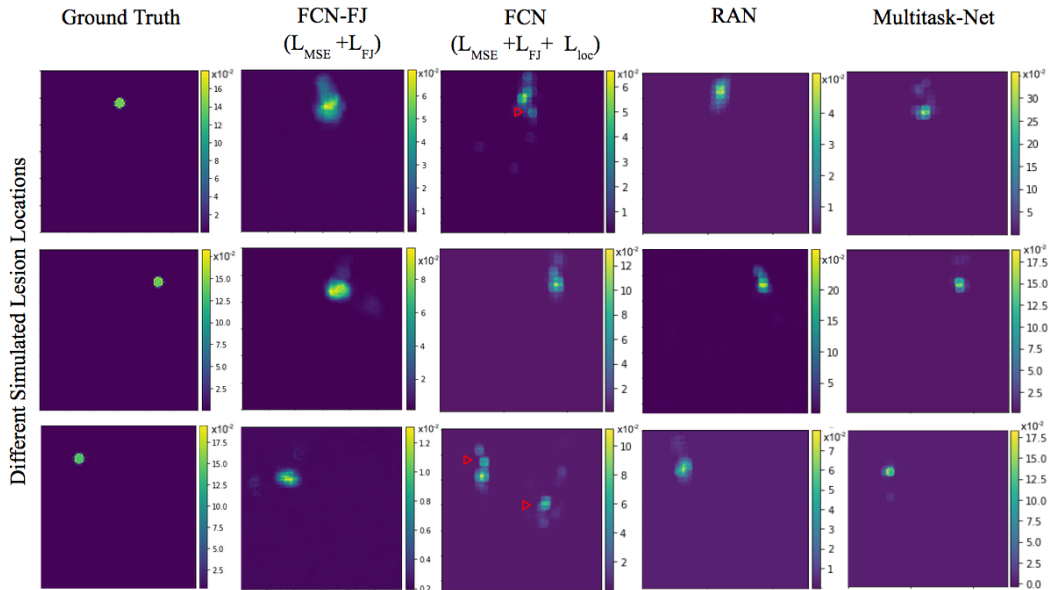


Fig. 6. Ablation study qualitative reconstruction performance on phantom samples with known lesion ground truth location. Our proposed method results (last column) are contrasted to FCN with different combinations of loss functions (L_{MSE} , L_{FJ} , and L_{loc}) and RAN architecture with L_{MSE} loss results. By attending to important features, spatial-wise attention model (RAN) filters irrelevant features resulting in improved lesion size and location compared to FCN while reducing false positives (e.g., red triangles). Combining spatial-wise attention and similarity loss provides the best reconstruction results with more accurate reconstruction of absorption coefficients.

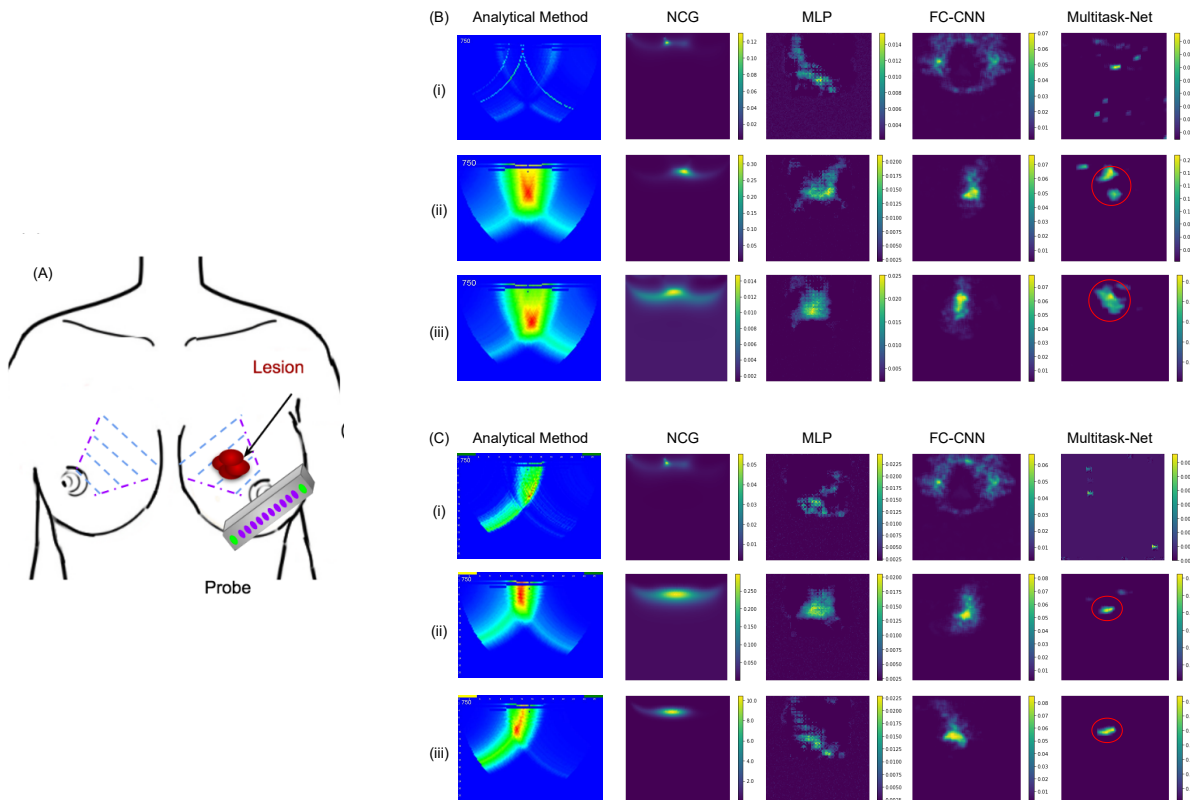


Fig. 7. (A) Clinical data scan procedure. Dashed blue lines delineate the different probe sweeps. (B-C) Qualitative reconstruction performance of our model (last column) compared to analytical, NCG, MLP, and FC-CNN methods on different real patient data. (i) Reconstruction of healthy breast tissue of candidate versus cancerous lesion over different probe positions (ii-iii). Note how the size of the lesion reconstructed by our method more faithfully reflects the relative difference in size. (approximate lesions sizes obtained with ultrasound; patient 1 (B): $22 \times 21 \times 23$ mm, patient 2 (C): $19 \times 16 \times 18$ mm). The minimum of the colorscale is set to the background value, the maximum dynamically to try to report visually consistent results, while accommodating the very high variance between images that would otherwise prevent any comparison.

in Table II. As a partial ground truth, both patients underwent ultrasound scans to obtain estimated lesion dimensions, showing irregular masses measuring $22 \times 21 \times 23$ mm for the first patient and $19 \times 16 \times 18$ mm for the second patient, with biopsies confirming that the lesions are cancerous.

In healthy tissue (Fig. 7-B:i, C:i), only background signal is reconstructed by almost all methods. In cancerous tissue, all competing approaches and our proposed method detect the lesion. The small artifacts (false positives) in healthy patient data reconstructed by MLP, FC-CNN and Multitask-Net (Fig. 7-B:i,C:i) can be attributed in part to the shift in distribution between training and test dataset. While the transfer learning module reduces this gap, it cannot eliminate this gap to the real patient data distribution completely. We observed that such artifacts could be removed by post-processing, e.g. via thresholding, by a qualified clinical operator. At the same time these failure cases highlight the need for more real-life patient data to better train the transfer learning module.

Consistent with the results on phantom data, the analytical, NCG, and MLP methods have a rough approximate outline of the tumor, whereas methods only partly agree with the approximate center of the predicted lesion.

Table II reports the average of reconstructed area ratio with respect to the ground truth lesion size over different probe sweeps for each patient. Areas of thresholded regions ($> 80\%$

TABLE II
QUANTITATIVE RESULTS ON REAL DATA EXPERIMENTS.

	Reconstructed area ratio ($>$ is better)	
	Patient I	Patient II
Analytical	0.18 ± 0.31	0.61 ± 0.03
NCG	0.41 ± 0.37	0.40 ± 0.25
MLP	0.54 ± 0.16	0.34 ± 0.19
FC-CNN	0.44 ± 0.10	0.45 ± 0.08
Multitask-Net	0.70 ± 0.30	0.30 ± 0.34

binary threshold), given the pixel dimensions of the image, were compared to the real ground truth area. We observe that, in contrast to the other approaches, our method is able to reconstruct the lesions more faithfully with respect to their relative sizes for patient I while it underestimates lesion size for patient II. In addition, our method obtains a more precise delineation of the lesions.

While the transfer learning network bridges the gap to some extent between phantom (training) data and real world data, the results on real world data illustrate that high variability in, e.g., illumination and noise levels can still mislead the network. Given that no two tumors are alike, tumor heterogeneity leads to markedly different acquisition signatures that are not necessarily present in the training data.

TABLE III
SENSITIVITY STUDY: THE EFFECT OF β ON PERFORMANCE.

β	Loc. Error (pixel, < is better)	PSNR (dB, > is better)	SSIM (> is better)	Fuzzy Jaccard (> is better)
0	17.8	27.88	0.90	0.71
0.2	13.8	27.84	0.92	0.72
0.4	12.5	27.73	0.96	0.77
0.6	13.1	27.82	0.92	0.72
0.8	13.5	27.79	0.94	0.74
1	14.1	27.69	0.95	0.77

TABLE IV
PERFORMANCE IMPACT OF VARYING NUMBER OF NETWORK LAYERS.

	Loc. Error (pixel, < is better)	PSNR (dB, > is better)	SSIM (> is better)	Fuzzy Jaccard (> is better)
Shallow	17.74	24.38	0.87	0.58
Baseline	14.01	25.87	0.94	0.68
Deep	14.51	25.61	0.93	0.69

D. Multitask versus Single-task

We first tested how the combination of different loss functions influences the reconstruction and localization performance. We used $\mathcal{L} = \mathcal{L}_{\text{MSE}} + \beta\mathcal{L}_{\text{loc}}$ and explored different values for $\beta \in [0,1]$. Table III shows the reconstruction performances as well as the localization error on the test set of 80 images for different values of β . Overall, we observed that learning with multiple losses improved both reconstruction and localization performance. In fact, we observed up to 18% decrease in localization error when using a combination of \mathcal{L}_{MSE} and \mathcal{L}_{loc} compared to using \mathcal{L}_{MSE} in isolation, with increases in SSIM by 6% and Fuzzy Jaccard by 7% using the best weight. Table I clearly shows that multitask-learning obtains optimal results for all metrics except a marginal drop in Fuzzy Jaccard of 1%.

E. Network Depth Trade-off: Complexity versus Performance

Increasing the dimensions of the network can allow for improved performance, yet results in larger hardware and training time requirements. With our application domain in hand-held devices, it critical to evaluate the trade-off between network dimensions and performance. With an exhaustive architecture search beyond the scope of this work, we limit ourselves to varying the number of layers and channels. Table V and IV show experimental performance results that were performed on the in silico dataset.

First, we test the variation of model performance over different depths of the network. We use our proposed model as a baseline and constructed a shallower and a deeper model by adding and subtracting two residual blocks with 32 channels from the baseline, respectively.

While the deeper model performs better compared to a shallower network (Table IV), performance did not improve ($\sim 1\%$ variation in all metrics) by adding more residual attention layers compared to the baseline model at the price of more parameters and risking overfitting. A lighter architecture would be more convenient for the deployment in handheld devices with small power requiring a fine balance between accuracy and latency.

TABLE V
VARIATION OF THE MODEL PERFORMANCE OVER DIFFERENT NUMBER OF CHANNELS.

	Loc. Error (pixel, < is better)	PSNR (dB, > is better)	SSIM (> is better)	Fuzzy Jaccard (> is better)
8 channels	14.92	25.62	0.88	0.63
16 channels	13.65	25.75	0.92	0.66
32 channels	14.02	25.87	0.94	0.69
64 channels	13.95	25.86	0.93	0.67

TABLE VI
TRADE OFF BETWEEN PROBE COMPLEXITY AND MODEL PERFORMANCE ON THE SYNTHETIC TEST SET.

	Loc. Error (pixel, < is better)	PSNR (dB, > is better)	SSIM (> is better)	Fuzzy Jaccard (> is better)
128 sensors	14.03	25.87	0.94	0.69
64 sensors	16.98	21.56	0.92	0.48
32 sensors	17.78	21.53	0.88	0.46
16 sensors	18.66	21.45	0.88	0.45

Note that the fully connected layer, used to back project the sensor data in the image domain, non-linearly increases the memory requirements of the neural network, thereby limiting the maximum number of additional convolution layers. This layer is deemed compulsory for an inverse scattering problem where direct analytical reconstruction is not well-defined, unlike in magnetic resonance imaging or computed tomography where a well-defined inversion exists using Fourier transform or Radon transform [58].

To observe the dependency on the number of convolution channels, we additionally trained the network with a varying number of channels. We observed marginal variations of the network performance beyond 32 channels: a variation of 1% and 1.5% in SSIM and Fuzzy Jaccard metrics, respectively, beyond 32 channels (Table V). This can be explained by the size and shape of lesions that are generally small, when compared to the background, and smooth objects, therefore they can be easily fitted by a small number of feature maps.

F. Effect of Varying the Number of Sensors

Given that our probe only has 2 light sources and 128 sensors (detectors), we tested the robustness of our model to the variation of the number of sensors (co-linear equispaced detectors) by reducing it gradually from 128 to 16. To train the network for different source-detector configurations, we generated different sets of training data and changed the input sizes of the network accordingly while maintained the overall network structure. While it is out of the scope of this work to test the results using the DOB-Scan probe (requiring new hardware design, testing, and manufacturing), we reported experimental performance results for the in silico dataset in Table VI and visual inspection results in Fig. 8.

Note the high contrast between the background and the lesion area, despite the increased difficulty to accurately recover lesion size and location in sparse signal conditions (smaller number of detectors). The results are on par with qualitative results Table VI where we notice an increase in localization error while we observe a slight decrease in SSIM.

V. DISCUSSION

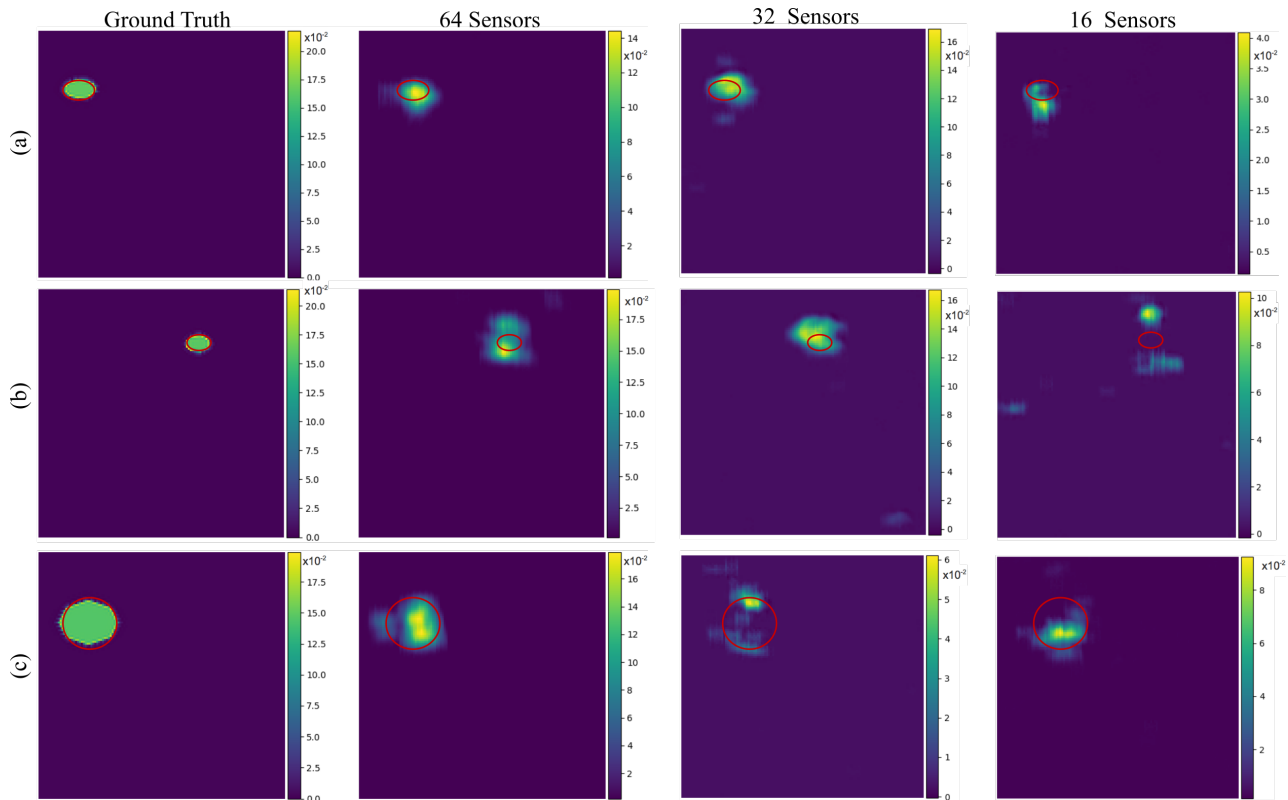


Fig. 8. Effect of varying number of detectors, while keeping the same number of probe sources, on the Multitask-Net qualitative results for different lesions sizes, locations, and absorption coefficients (a-c). While the model is still able, in most cases, to detect lesion presence, lesion size and correct localization are impacted by the reduction in the number of sensors and the increase in sparsity of the collected signal.

A. Attention-directed Lesion Reconstruction

DOT reconstruction leverages the fact that healthy tissue has optical coefficients that are near uniform, whereas lesions have optical coefficients that diverge markedly and are characterized by high variability. The encoding of a highly variable signal (lesion) requires more information or computational complexity than a near uniform signal. The high variance is explained in part due to high inter-tumor heterogeneity, a leading cause preventing effective treatment or prognosis [45]. Furthermore, for the clinician it is sufficient that the network recognizes correctly that certain areas are healthy tissue without requiring precise pixel-wise reconstruction. In contrast, the lesion needs not only to be detected, but located and reconstructed precisely in order to enable diagnosis, prognosis, and successful treatment recommendation. Therefore, employing network attention on the lesion is one way of dedicating more computational power where it is required, without sacrificing detection of healthy tissue.

B. Network Generalization

The generalization capability of our network is illustrated by the fact that training occurs on simulation data, yet performance is strong on phantom and real data. Moreover, unlike the analytical approach that requires a reference measurement of healthy tissue to compute the inverse image, our network, similar to Yoo *et al.* [12], is designed to directly learn from the collected measurement and does not require any

reference or prior conditions that might bias the search space. In a clinical setting, such as breast cancer screening, such reference measurement from a homogeneous background may be nontrivial to obtain. Therefore, our model can generalize better to unseen conditions without additional pre- or post-processing techniques. In this work, we obtained good results when trained only on *in silico* data. We expect improved results as more real data of phantoms and patients are collected. Finally, we used model architecture and implementation details described in III-B.3, with good initial results. However, it is likely that other architectures can be found that will yield improved results, e.g., via neural architecture search [59].

C. Limitations

Although this study provides a proof of concept of the potential advantages of using multitask deep learning to DOT image reconstruction, it suffers from some limitations. First, reconstruction of deep lesions, with a location deeper than 35 mm, remains a challenging task in DOT. When one uses a handheld probe, with a colinear sources-detectors placement, it can be even harder to differentiate the lesions specific signal from tumor-free tissue signal. As a result, the reconstructed image can be biased towards tumor-free tissue. Similarly, while our model is able, in most cases, to detect lesion presence, reconstruction of multiple lesions exhibit some weakness, as shown in Fig. 9. Specifically, occluded lesions and lesions situated at the border of the probe can be misreconstructed, where only one lesion is reconstructed, lesion

location deviates, or tumor-free tissue is reconstructed (false negative). This problem is mainly due to the weaknesses of the collected signal given the limited angle condition. While multiple sweeps or variations in the scanning angle can help in the latter case, it is unlikely to help in the case of deep lesion(s). Second, since the real dataset only contains 2 patients with few sweeps, it is not an exhaustive list of all cases, and therefore the proposed method requires validation on larger and more diverse datasets. However, this is not a technical limitation of the model since more tests can be done as more real data become available.

VI. CONCLUSION

We present the first work to effectively leverage multitask learning by combining an appearance similarity loss and a reconstruction loss along with a spatial attention mechanism to DOT-reconstruction, with application to reconstructing lesions in breast cancer tissue. We demonstrated improvement in localization error over the state of the art and presented a step forward in DOT image reconstruction and the usage of machine learning in image reconstruction. Leveraging physics-based simulators of light propagation to generate realistic synthetic datasets provided an alternative source of large and diversified training dataset. The critical addition of transfer learning ensures our method is not limited by inherently small real-world datasets. At the same time it is still capable of performing at high accuracy on such datasets where more complex sources of artifacts and noise are present. Nevertheless, we noted some performance differences in real world data. As more patient data is collected, it will be critical to better understand and address the causes of such differences. Finally, we show that our method is sufficiently fast enough to ensure real time processing, enabling in situ deployment in clinical settings.

REFERENCES

- [1] M. Shokoufi and F. Golnaraghi, "Development of a handheld diffuse optical breast cancer assessment probe," *Journal of Innovative Optical Health Sciences*, vol. 9, no. 02, p. 1650007, 2016.
- [2] M. L. Altoe *et al.*, "Diffuse optical tomography of the breast: a potential modifiable biomarker of breast cancer risk with neoadjuvant chemotherapy," *Biomedical Optics Express*, vol. 10, no. 8, pp. 4305–4315, 2019.
- [3] Y. Hoshi and Y. Yamada, "Overview of diffuse optical tomography and its clinical applications," *Journal of Biomedical Optics*, vol. 21, no. 9, p. 091312, 2016.
- [4] M. Applegate, R. Istfan, S. Spink, A. Tank, and D. Roblyer, "Recent advances in high speed diffuse optical imaging in biomedicine," *APL Photonics*, vol. 5, no. 4, p. 040802, 2020.
- [5] H. Dehghani, S. Srinivasan, B. W. Pogue, and A. Gibson, "Numerical modelling and image reconstruction in diffuse optical tomography," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1900, pp. 3073–3093, 2009.
- [6] S. R. Arridge, "Optical tomography in medical imaging," *Inverse problems*, vol. 15, no. 2, p. R41, 1999.
- [7] S. R. Arridge and J. C. Schotland, "Optical tomography: forward and inverse problems," *Inverse problems*, vol. 25, no. 12, p. 123010, 2009.
- [8] L. Zhang and G. Zhang, "Brief review on learning-based methods for optical tomography," *Journal of Innovative Optical Health Sciences*, vol. 12, no. 06, p. 1930011, 2019.
- [9] M. Mozumder, T. Tarvainen, A. Seppänen, I. Nissilä, S. R. Arridge, and V. Kolehmainen, "Nonlinear approach to difference imaging in diffuse optical tomography," *Journal of biomedical optics*, vol. 20, no. 10, p. 105001, 2015.
- [10] J. Heiskala, P. Hiltunen, and I. Nissilä, "Significance of background optical properties, time-resolved information and optode arrangement in diffuse optical imaging of term neonates," *Physics in Medicine & Biology*, vol. 54, no. 3, p. 535, 2009.
- [11] Y. Wang, J.-H. Chang, R. Aronson, R. L. Barbour, H. L. Graber, and J. Lubowsky, "Imaging of scattering media by diffusion tomography: an iterative perturbation approach," in *Physiological Monitoring and Early Detection Diagnostic Methods*, vol. 1641. International Society for Optics and Photonics, 1992, pp. 58–71.
- [12] J. Yoo *et al.*, "Deep learning diffuse optical tomography," *IEEE Transactions on Medical Imaging*, 2019.
- [13] S. R. Arridge and M. Schweiger, "Photon-measurement density functions. part 2: Finite-element-method calculations," *Applied Optics*, vol. 34, no. 34, pp. 8026–8037, 1995.
- [14] J. Yuan, G. Xu, Y. Yu, Y. Zhou, P. L. Carson, X. Wang, and X. Liu, "Real-time photoacoustic and ultrasound dual-modality imaging system facilitated with graphics processing unit and code parallel optimization," *Journal of biomedical optics*, vol. 18, no. 8, p. 086001, 2013.
- [15] J. Davis, Y.-H. Hsieh, and H.-C. Lee, "Humans perceive flicker artifacts at 500 hz," *Scientific reports*, vol. 5, no. 1, pp. 1–4, 2015.
- [16] H. Ben Yedder, B. Cardoen, and G. Hamarneh, "Deep learning for biomedical image reconstruction: A survey," *Artificial Intelligence Review*, pp. 1–33, 2020.
- [17] S. Ravishankar, J. C. Ye, and J. A. Fessler, "Image reconstruction: From sparsity to data-adaptive methods and machine learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 86–109, 2019.
- [18] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, 2017.
- [19] Y. Sun, Z. Xia, and U. S. Kamilov, "Efficient and accurate inversion of multiple scattering with deep learning," *Optics express*, vol. 26, no. 11, pp. 14 678–14 688, 2018.
- [20] U. S. Kamilov *et al.*, "Learning approach to optical tomography," *Optica*, vol. 2, no. 6, pp. 517–522, 2015.
- [21] J. Feng, Q. Sun, Z. Li, Z. Sun, and K. Jia, "Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography," *Journal of biomedical optics*, vol. 24, no. 5, p. 051407, 2018.
- [22] W. Van den Broek and C. T. Koch, "Method for retrieval of the three-dimensional object potential by inversion of dynamical electron scattering," *Physical review letters*, vol. 109, no. 24, p. 245502, 2012.
- [23] J. C. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 991–1048, 2018.
- [24] H. Ben Yedder, A. BenTaieb, M. Shokoufi, A. Zahiremami, F. Golnaraghi, and G. Hamarneh, "Deep learning based image reconstruction for diffuse optical tomography," in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2018, pp. 112–119.
- [25] H. Ben Yedder, M. Shokoufi, B. Cardoen, F. Golnaraghi, and G. Hamarneh, "Limited-angle diffuse optical tomography image reconstruction using deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 66–74.
- [26] A. R. Zamir, A. Sax, W. Shen, L. J. Guibas, J. Malik, and S. Savarese, "Taskonomy: Disentangling task transfer learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3712–3722.
- [27] S. Vesal, S. M. Patil, N. Ravikumar, and A. K. Maier, "A multi-task framework for skin lesion detection and segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 285–293.
- [28] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-point checklist and skin lesion classification using multitask multimodal neural nets," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 538–546, 2018.
- [29] K. Abhishek, J. Kawahara, and G. Hamarneh, "Predicting the clinical management of skin lesions using deep learning," *Scientific reports*, vol. 11, no. 1, pp. 1–14, 2021.
- [30] I. F. Sbalzarini, "Seeing is believing: Quantifying is convincing: Computational image analysis in biology," in *Focus on Bio-Image Informatics*. Springer, 2016, pp. 1–39.
- [31] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [32] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in

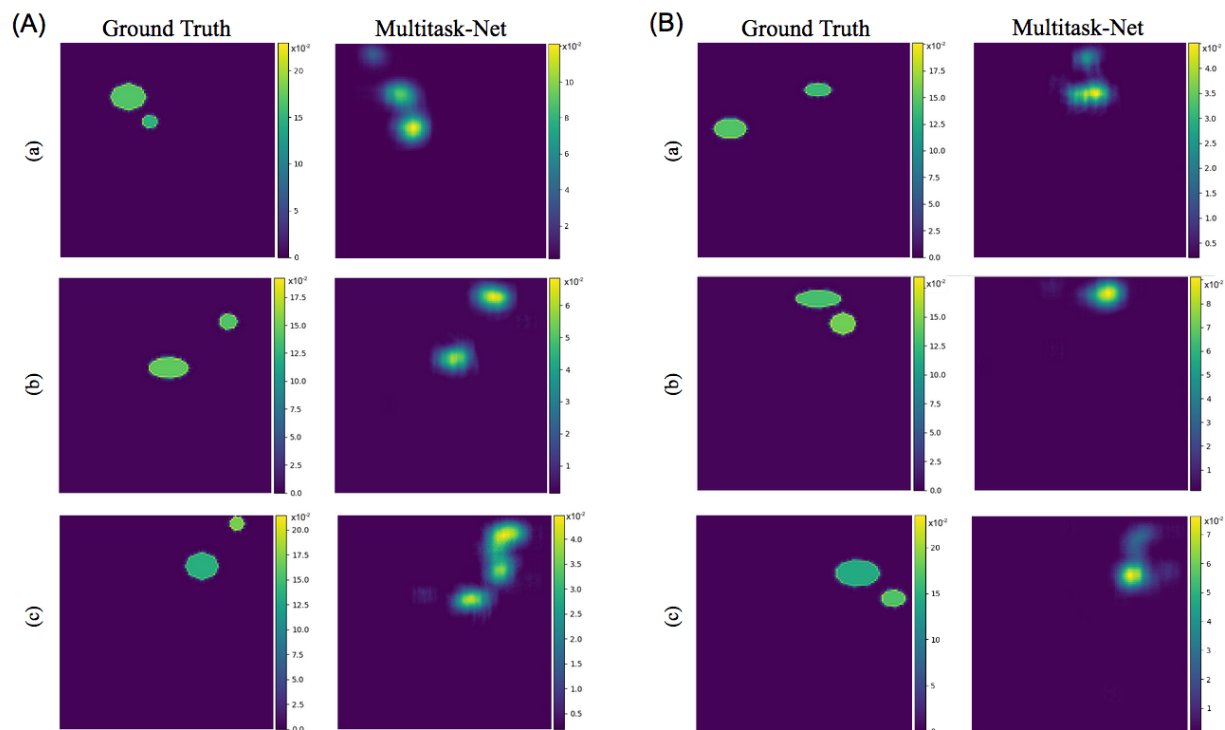


Fig. 9. Reconstruction of multi-lesions with different sizes and proximity (a-c). (A) While the model is able, in most cases, to detect lesion presence, occluded, edge located and deep lesions can be mis-reconstructed. (B) Some failure reconstruction examples where only one of the lesions is reconstructed.

International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 421–429.

- [33] M. Shokoufi, Z. Haeri, Z. Y. Lim, R. Ramaseshan, and F. Golnaraghi, “Translation of a portable diffuse optical breast scanner probe for clinical application: a preliminary study,” *Biomedical Physics & Engineering Express*, vol. 6, no. 1, p. 015037, 2020.
- [34] M. Shokoufi and F. Golnaraghi, “Handheld diffuse optical breast scanner probe for cross-sectional imaging of breast tissue,” *Journal of Innovative Optical Health Sciences*, vol. 12, no. 02, p. 1950008, 2019.
- [35] F. Wang *et al.*, “Residual attention network for image classification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3156–3164.
- [36] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, “Hierarchical attention networks for document classification,” in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480–1489.
- [37] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [38] B. Zhou, X. Lin, and B. Eck, “Limited angle tomography reconstruction: Synthetic reconstruction via unsupervised sinogram adaptation,” in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 141–152.
- [39] C. Geuzaine and J.-F. Remacle, “Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities,” *International journal for numerical methods in engineering*, vol. 79, no. 11, pp. 1309–1331, 2009.
- [40] M. Schweiger and S. R. Arridge, “The toast++ software suite for forward and inverse modeling in optical tomography,” *Journal of biomedical optics*, vol. 19, no. 4, p. 040801, 2014.
- [41] R. C. Haskell, L. O. Svaasand, T.-T. Tsay, T.-C. Feng, M. S. McAdams, and B. J. Tromberg, “Boundary conditions for the diffusion equation in radiative transfer,” *JOSA A*, vol. 11, no. 10, pp. 2727–2741, 1994.
- [42] N. Ghosh, S. Mohanty, S. Majumder, and P. Gupta, “Measurement of optical transport properties of normal and malignant human breast tissue,” *Applied Optics*, vol. 40, no. 1, pp. 176–184, 2001.
- [43] P. Taroni, A. M. Paganoni, F. Ieva, A. Pifferi, G. Quarto, F. Abbate, E. Cassano, and R. Cubeddu, “Non-invasive optical estimate of tissue composition to differentiate malignant from benign breast lesions: A pilot study,” *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.
- [44] D. Grosenick, H. Rinneberg, R. Cubeddu, and P. Taroni, “Review of optical breast imaging and spectroscopy,” *Journal of biomedical optics*, vol. 21, no. 9, p. 091311, 2016.
- [45] F. Beca and K. Polyak, “Intratumor heterogeneity in breast cancer,” in *Novel Biomarkers in the Continuum of Breast Cancer*. Springer, 2016, pp. 169–189.
- [46] B. Bi, B. Han, W. Han, J. Tang, and L. Li, “Image reconstruction for diffuse optical tomography based on radiative transfer equation,” *Computational and mathematical methods in medicine*, 2015, 2015.
- [47] M. Flexman, H. Kim, R. Stoll, M. Khalil, C. Fong, and A. Hielscher, “A wireless handheld probe with spectrally constrained evolution strategies for diffuse optical imaging of tissue,” *Review of Scientific Instruments*, vol. 83, no. 3, p. 033108, 2012.
- [48] F. Chollet *et al.*, “Keras,” <https://github.com/fchollet/keras>, 2015.
- [49] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [50] H. Dehghani, M. E. Eames, P. K. Yalavarthy, S. C. Davis, S. Srinivasan, C. M. Carpenter, B. W. Pogue, and K. D. Paulsen, “Near infrared optical tomography using nifast: Algorithm for numerical model and image reconstruction,” *Communications in numerical methods in engineering*, vol. 25, no. 6, pp. 711–732, 2009.
- [51] S. R. Arridge and M. Schweiger, “A gradient-based optimisation scheme for optical tomography,” *Optics Express*, vol. 2, no. 6, pp. 213–226, 1998.
- [52] M. Schweiger, S. R. Arridge, and I. Nissilä, “Gauss–newton method for image reconstruction in diffuse optical tomography,” *Physics in Medicine & Biology*, vol. 50, no. 10, p. 2365, 2005.
- [53] W. Lu, J. Duan, D. Orive-Miguel, L. Herve, and I. B. Styles, “Graph- and finite element-based total variation models for the inverse problem in diffuse optical tomography,” *Biomedical optics express*, vol. 10, no. 6, pp. 2684–2707, 2019.
- [54] A. Douiri, M. Schweiger, J. Riley, and S. Arridge, “Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information,” *Measurement Science and Technology*, vol. 18, no. 1, p. 87, 2006.
- [55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image

- quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [56] J. Kawahara and G. Hamarneh, “Fully convolutional neural networks to detect clinical dermoscopic features,” *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 578–585, 2018.
- [57] W. Lam *et al.*, “Factors affecting the palpability of breast lesion by self-examination,” *Singapore medical journal*, vol. 49, no. 3, p. 228, 2008.
- [58] M. T. McCann, M. Unser *et al.*, “Biomedical image reconstruction: From the foundations to deep neural networks,” *Foundations and Trends® in Signal Processing*, vol. 13, no. 3, pp. 283–359, 2019.
- [59] T. Elsken, J. H. Metzen, and F. Hutter, “Neural architecture search: A survey,” *Journal of Machine Learning Research*, vol. 20, pp. 1–21, 2019.

APPENDIX A

ROLE OF THE FULLY CONNECTED AND CONVOLUTIONAL LAYERS

The fully connected layer emulates an approximate inverse of the forward operator that can be further refined by subsequent convolution layers. During the training process, the error between the predicted output and the ground truth distribution of absorption coefficients is backward propagated from the output layer to the fully connected layer to adjust the weights and biases. By repeatedly applying this procedure for each sample in the training set, the learning process can converge. Given randomly initialized weights w_{ij}^l and bias variables b_{ij}^l at each layer (l), the output o_j of the j 'th neuron at the fully connected (fc) layer is computed using:

$$o_j^{fc} = g \left(\sum_{i=1}^N w_{ij}^{fc} y_i + v_j^{fc} \right), \quad (11)$$

where $g(\cdot)$ is the neuron activation function, and y is the measurement vector input to the network. Hence, the fully connected layer learns a weighted combination of the different receptive sensors, based on the signal collected (y), and its reshaped output provides a first image estimate that is subsequently enhanced by subsequent layers. Filters (kernels) pass over the image estimate and transform it based on the weights in each filter. The final feature map values at each convolutional (cn) layer are then computed according to the following formula:

$$\begin{aligned} o^{cn}[m, n] &= (o^{cn-1} * h)[m, n] \\ &= \sum_i \sum_k h[i, k] o^{cn-1}[m-i, n-k], \end{aligned} \quad (12)$$

where the input image is denoted by $o^{cn-1}[m, n]$, kernel by h and m and n are indexes of rows and columns, respectively. We have shown in the silico reconstruction, Fig. 4, shows that consecutive convolution layers are very important to recover the accurate size and location of the optical anomalies. Yoo *et al.* [12] adopted a similar architecture and theoretically justified the reason behind it based on the analysis of deep convolutional framelets [23].

APPENDIX B

IMAGE RECONSTRUCTION REGULARISATION DETAILS

We consider the regularized output least-squares approach to the inverse problem in equation (1) where x represents the searched optical parameter. The first order Tikhonov penalty term takes the form [52]:

$$\mathcal{R}(x) = \|\mathbb{L}\Delta x\|_2^2, \quad (13)$$

where $\mathbb{L}^T \mathbb{L} \in \mathbb{R}^{N \times N}$ denotes the Laplacian, $\Delta x_k = x_k - x_0$, and x_0 is an initial estimate. We define $\mathbb{L}^T \mathbb{L}$ as:

$$(\mathbb{L}^T \mathbb{L})_{ij} = \begin{cases} n(i) & \text{if } i = j \\ -1 & \text{if } j \text{ is neighbour of } i \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where $n(i)$ is the number of neighbours of basis component i . The four-connected pixel neighbourhood in 2D are used. The isotropic total variation penalty term of the form [53]:

$$\mathcal{R}(\mu) = \sum_{i=1}^M \sqrt{|(D_x \mu)_i|^2 + |(D_y \mu)_i|^2} \quad (15)$$

where D_x is a matrix of size $M \times N$ representing the discrete partial derivative of μ , the searched optical parameter, along the x direction. D_y is the derivative matrix along the y direction. M is the number of triangles in the mesh and N is number of nodes.

TABLE VII
COMPETING ALGORITHM ARE EVALUATED UNDER DIFFERENT REGULARIZATION TECHNIQUES AND HYPER-PARAMETERS ON 32 IMAGES FROM THE NUMERICAL PHANTOMS

	Reg	Parameters	Loc. Error (pixel, < is better)	PSNR (dB, > is better)	SSIM (> is better)	Fuzzy Jaccard (> is better)
LM	TV	$\beta = 1e-2, \lambda = 1e-6$	31.6 ± 14.5	15.8 ± 4.5	0.20 ± 0.02	0.12 ± 0.05
		$\beta = 1e-1, \lambda = 1e-6$	31.0 ± 14.3	17.8 ± 5.6	0.39 ± 0.05	0.17 ± 0.03
		$\beta = 1e-2, \lambda = 1e-4$	37.3 ± 23.3	19.3 ± 11.3	0.19 ± 0.01	0.09 ± 0.01
	LI	$\tau = 1e-4, \lambda = 1e-6$	30.9 ± 17.4	12.6 ± 5.8	0.32 ± 0.12	0.14 ± 0.07
		$\tau = 1e-2, \lambda = 1e-6$	25.1 ± 12.7	16.6 ± 8.6	0.38 ± 0.09	0.16 ± 0.05
		$\tau = 1e-2, \lambda = 1e-4$	40.9 ± 20.3	10.6 ± 9.7	0.20 ± 0.00	0.80 ± 0.01
		$\tau = 1e-2, \lambda = 1e-2$	28.2 ± 21.9	10.6 ± 7.3	0.39 ± 0.20	0.18 ± 0.14
NCG	TV	$\beta = 1e-2, \lambda = 1e-6$	35.7 ± 19.9	13.5 ± 8.5	0.13 ± 0.02	0.08 ± 0.02
		$\beta = 1e-2, \lambda = 1e-4$	37.5 ± 21.4	12.0 ± 8.6	0.14 ± 0.02	0.08 ± 0.02
		$\beta = 1e-1, \lambda = 1e-6$	32.6 ± 16.5	16.5 ± 9.2	0.33 ± 0.01	0.22 ± 0.13
		$\beta = 1e-0, \lambda = 1e-6$	37.4 ± 21.5	12.2 ± 8.7	0.14 ± 0.01	0.10 ± 0.02
		$\beta = 1e-1, \lambda = 1e-4$	35.8 ± 19.9	14.2 ± 8.4	0.14 ± 0.01	0.13 ± 0.10
	LI	$\lambda = 1e-3$	40.5 ± 20.8	8.39 ± 8.4	0.15 ± 0.01	0.09 ± 0.03
		$\lambda = 1e-4$	41.8 ± 22.7	12.7 ± 6.0	0.15 ± 0.01	0.09 ± 0.03
		$\lambda = 1e-6$	34.1 ± 19.3	13.6 ± 7.4	0.13 ± 0.01	0.12 ± 0.08

APPENDIX C

HYPER-PARAMETER SELECTION AND SETTINGS

To validate the conventional algorithms in a quantitative manner and select the best reconstruction parameters, we conducted a study using numerical phantoms and reported results in Table VII. Reference mesh coarseness was set to 1.5 mm in all experiments as we noticed that decreasing it only resulted in longer reconstruction time with negligible to no improvement. The maximum number of iterations was set to 100 epochs for NCG and 30 for LM. Algorithms are optimized over different values of the regularization parameter (λ), TV object threshold parameter (β), and the control parameter (τ) (LM algorithm only).

The following optimized parameter values were used for the remainder of the test: $\lambda = 1e-6$, $\beta = 0.1$, and $\tau = 0.01$.

To demonstrate the effect of the number of sources/detectors on conventional reconstruction results, we used a modified in silico design, including a third source in the middle of the probe and increased number (44 more) of co-linear detectors equidistantly positioned between sources (Fig. 10-A). As reported in Fig. 10-B, performance improves with denser probe configuration. Note the more localized lesion reconstruction

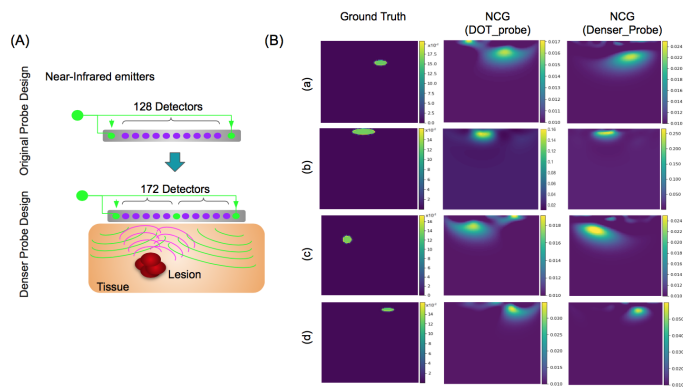


Fig. 10. Effect of the number of sources/detectors on NCG reconstruction results. (A) A slightly modified *in silico* probe design with an additional source in the middle of the probe and few more detectors equidistantly positioned. (B) Reconstruction results using original and *in silico* probe design based on best selected hyper-parameters.

with the modified probe design, using the best parameters selected with the original probe design. Results are expected to improve in real and phantom cases as well. However, we cannot at this time manufacture a modified probe to verify this expected result on real data. In addition, such a probe design would not align with our foreseen deployment in clinical settings, where we aim to minimize hardware complexity.