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

Hanene Ben Yedder, Ben Cardoen, Majid Shokoufi, Farid Golnaraghi, and Ghassan Hamarneh

Abstract—Diffuse optical tomography (DOT) leverages near-infrared light propagation through tissue to assess its optical properties and identify abnormalities. DOT image reconstruction from limited-angle data acquisition is severely ill-posed due to the highly scattered photons in the medium and the relatively small number of collected projections. Reconstructions are thus commonly marred by artifacts and, as a result, it is difficult to obtain 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. Both quantitative and qualitative results on phantom and real data indicate the superiority of our multitask method in the reconstruction and localization of lesions in tissue compared to state-of-the-art methods. 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, transfer learning, portable hand-held, clinical trial.

I. INTRODUCTION

Correspondence to be addressed to {hbenyedder, 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) and MITACS. We thank Nvidia Corporation for the donation of Titan X GPUs used in this research and Compute Canada for HPC resources.

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

NEAR-infrared diffuse optical tomography is a relatively new 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-probe consists of a set of near-infrared sources for tissue illumination and LED 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].

A. Challenges in DOT Tissue 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. While no exact inverse approach exists a priori, conventional reconstruction algorithms are sensitive to reconstruction parameters (e.g., boundary conditions, regularization) and are more suited when strong assumptions are justified, which is usually not feasible in practice.

The problem can be relaxed in the case of a low scattering medium using the linear first Born approximation [3], however, resolving this problem in a highly scattering medium remains hard given the severely ill-posed nonlinear inverse [6]. 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) [7]. A linear probe [1], for instance, reduces system complexity (lower hardware and software requirements), however, the directional nature of the transducer results in a limited view of the target and artifacts.

B. Current Approaches to Address DOT Inverse Problem

Conventional methods for reconstructing tissue properties from optical propagation rely on computationally intensive iterative optimization [3], [5], [8], which are too slow for real-time imaging applications. Furthermore, their reconstruction quality is highly dependent on the regularization function and related hyper-parameters settings that often require expert parameters tuning. Recently, deep learning has emerged as a powerful alternative for biomedical image reconstruction [9]–[11] 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]–[15]. While preliminary works did not consider end-to-end model inversion of distribution [12], [13], [16], 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. [12] 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. [14] proposed a multilayer perceptron (MLP) feed-forward artificial neural network for a full-view 2D-DOT image reconstruction. Similarly, Yoo et al. [15] proposed an autoencoder CNN for a 3D-DOT inverse problem. However, performance degrades significantly for limited-angle acquisition where the number of sources is smaller and their placement is co-linear.

Preliminary versions of our work [17], [18] tackled reconstruction of biological tissue parameters from limited-view data acquisition in a strong scattering medium using a decoder-like fully convolutional neural network architecture (FCN)

with a weighted combination of mean squared error (MSE) and Fuzzy Jaccard as loss function and tested only on phantom data. 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. Incorporating multitask learning has been shown to improve results in other applications, e.g., taskonomy [19] and skin-multitask [20]. In our application as well, we illustrate that a multitask approach leads to a better reconstruction and localization, enabling more accurate diagnosis.

C. Contributions

We make the following contributions in this paper:

- 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 [21], [22] 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.
- 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) [1], [23].
- To the best of our knowledge, this is the first work that tests image deep learning based DOT reconstruction on clinical data of real patients.
- 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.
- 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 baseline analytical and state-of-the-art deep learning based reconstruction approaches.

In section II, we introduce our proposed method for multitask reconstruction and localization learning followed by domain adaptation learning. Physics-based optical diffusion simulation and test dataset 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-C, results on physical phantoms in Section IV-D, and results on real-world data in Section IV-E. 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:

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

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.

A. Reconstructing Images from DOT Measurements

We extend our previously proposed decoder-like architecture [17], [18] 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, mimicking the filtered back-projection layer, maps the measurement vector y into 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 . While the mean squared error was used in isolation in our preliminary work [17], a Fuzzy Jaccard loss component $\mathcal{L}_{FJ}$ was added later [18] 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 convolutional 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 [21], [25]. 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 [26], [27] and computer vision problems tasks [21], [22] 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 [22]. Each residual block uses sequential modules of the form: two convolutions – batch normalization – ReLU and zero-padding followed by a squeeze [21] and excite module.

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 to rescale activations to $[0, 1]$, which are thereafter used to re-calibrate U spatially. 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:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \beta \mathcal{L}_{\text{loc}}, \quad (2)$$

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

1) *Reconstruction Loss*: $\mathcal{L}_{\text{MSE}}$ 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, $\mathcal{L}_{\text{MSE}}$ evaluates pixels in isolation, ignoring the overall image structure. Furthermore, $\mathcal{L}_{\text{MSE}}$ 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 $\mathcal{L}_{\text{loc}}$ is introduced to emphasize more accurate lesion localization. $\mathcal{L}_{\text{loc}}$ is based on the Euclidean distance transform that maps an image to a similar gray-level image, except that the gray-level 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

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^p . (B) In-silico training data pairs generation simulating our probe's specification [24]. 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 $\sigma(\cdot)$ corresponds to the relative importance of a spatial information (i, j) of the feature map U . All feature maps produced by all convolutional layers are set to size 128×128 using (3×3) followed by (5×5) kernels. The similarity-wise loss $\mathcal{L}_{loc}$, leveraged to enforce localization accuracy, computes the mean absolute error between the distance transform of the reconstructed image and the ground truth one.

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. (3), 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:

$$\mathcal{L}_{loc}(\theta) = \|DT(\mathcal{F}^{-1}(y; \theta); x) - DT(x)\| \quad (3)$$

where

$$DT = \begin{cases} 0 & \text{if } a_{ij} \in S \\ D(a_{ij}, S) & \text{if } a_{ij} \in R_S, \end{cases} \quad (4)$$

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 Yedder et al. [18] and Zhou et al. [28], 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 :

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}_{TL}(\theta) \quad \text{where} \quad \mathcal{L}_{TL}(\theta) = \sum_{i=1}^{N_p} \|\phi(y_i^p; \theta) - y_i^s\|. \quad (5)$$

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

$$\hat{x}^* = \mathcal{F}^{-1}(\phi(y^p)). \quad (6)$$

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).

Fig. 3. Workflow of DOT in silico training data generation for DOT image reconstruction. 2D triangular meshes modeling tissue with different lesion sizes and depths were generated using a Gmsh software [29] and fed to the time-resolved optical absorption and scattering tomography software (Toast++) for forward projection simulation using 750 nm wavelength modulation and breast like optical properties.

The Toast++ software suite [30] 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 models the light propagation in a diffusive medium. The diffusion equation is given by:

$$-\nabla \cdot D(r) \nabla \varphi(r, \omega) + \left(\mu_a(r) + \frac{i\omega}{c_m(r)} \right) \varphi(r, \omega) = q_0(r, \omega) \quad (7)$$

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; $q_0(r, \omega)$ represents the light source term; and $\varphi(r, \omega)$ is the generated photon rate 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 modelled accurately in Toast++, as illustrated in Fig. 2-B, to obtain measurements that mimic real values obtained by the DOB probe [1], [24]. The probe is comprised 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, 128-D vector) data pairs are used. Realistic human breast tissue optical parameters distribution values [31] were considered to mimic both healthy and cancerous tissue optical coefficients. The absorption coefficients of breast tissue range around 0.002 mm^{-1} and 0.16 mm^{-1} for healthy and cancerous tissue, respectively. The scattering coefficients are in the range of 0.6 mm^{-1} to 1.6 mm^{-1} . The size of the reconstructed image is $6 \times 6 \text{ cm}$ with a resolution of 128×128 pixels. A Gaussian noise ($\sigma = 10\%$ of maximum sensor value) was added to the measurements to mimic real-world noise.

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], [32]. 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 [32]. 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, showing calcification over the area, and biopsy.

C. Implementation

The model was implemented in the Keras [33] framework and trained for a total of 500 epochs on an Nvidia Titan X GPU using the Adam optimizer. 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 (2); 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 $\mathcal{L}_{\text{MSE}}$ before introducing $\mathcal{L}_{\text{loc}}$ 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.

IV. RESULTS

A. Multitask versus Single-task

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) [34]; and (iv) the Fuzzy Jaccard [35].

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 $\beta \in [0, 1]$. Figure 4-A shows the reconstruction

Fig. 4. Sensitivity study. A: The effect of β on performance (learning rate fixed at $1\text{-e}4$). B: Variation of the model performance over different number of channels. A marginal variation of 1% and 1.5% in SSIM and Fuzzy Jaccard metrics, respectively, beyond 32 channels. C: Trade off between probe complexity and model performance on the synthetic test set ($n=80$ images). D: Performance impact of varying number of network layers. Shallow and deep models are constructed by subtracting and adding two residual blocks with 32 channels from our proposed model used as a baseline. A lighter architecture would be more convenient for the deployment in handheld devices with small power requiring a fine balance between accuracy and latency.

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 $\mathcal{L}_{\text{MSE}}$ and $\mathcal{L}_{\text{loc}}$ compared to using $\mathcal{L}_{\text{MSE}}$ in isolation, with an increase 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%.

B. 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 is 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. Figures 4-B,D 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 (Fig 4-D), 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. 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 [36].

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 (Fig. 4-B). 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.

C. 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

Fig. 5. Qualitative reconstruction performance of our model compared to the state-of-the-art MLP technique on 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 MLP approach.

our Multitask-Net method and compare results with those of the the MLP method [14]. In Fig. 5, we show results on some in silico samples with different lesion sizes and depths. Interestingly, while the images reconstructed by MLP method provide only a very noisy reconstruction and a rough approximate of the lesion location, 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. 5 (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 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 [37] 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.

D. 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 an analytical approach described in Shokoufi et al. [38], an MLP model [14], and the fully convolutional network (FCN) [18].

1) *Qualitative Results on Phantom data:* In Fig. 6, we contrast our results with the state of the art on phantom data. 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. 6- Analytical). 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]. Similar results are obtained for MLP with very low per pixel values of reconstructed absorption coefficients (Fig. 6- MLP). The FCN, on the other hand, localizes the lesion in a single concentrated area with some error in lesion size and optical coefficients (Fig. 6- FCN). Adopting the new loss term $\mathcal{L}_{loc}$ clearly reduces the lesion localisation error (Fig. 6- FCN ($\mathcal{L}_{MSE} + \mathcal{L}_{FJ} + \mathcal{L}_{loc}$)). 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 optical coefficients reconstruction, 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. 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. Furthermore, as reported in our previous work [17], [18], our model allows real-time reconstruction. Note that the new loss function has little to no effect on the inference time.

E. Results on Real 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). Given the small dataset size, we only report visual results and compare with the analytical approach [1]. As 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 both methods. In cancerous tissue, both the analytical approach and our proposed method detect the lesion. Consistent with the results on phantom data, the analytical approach has a very rough approximate outline of the tumor, whereas our method agrees with the approximate center of the predicted lesion, but is more localized.

We observe that, in contrast to the analytical approach, our method is able to reconstruct the lesions more faithfully with

Fig. 6. Qualitative reconstruction performance of our model compared to state-of-the-art 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 coefficients 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 the filtered back-projection operation and gives a rough approximation of lesion localization, while precision of the reconstructed absorption coefficients are very low. Adding convolution layers helps to better filter noise and reduce false-positive while still miss-predicting correct absorption coefficients values required for proper diagnostic. 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 absorption coefficients reconstruction.

TABLE I
QUANTITATIVE RESULTS ON 32 PHANTOM EXPERIMENTS.

	Loss			Loc. Error (pixel, <, is better)	PSNR (db >, is better)	SSIM (>, is better)	Fuzzy Jaccard (>, is better)
	L_{MSE}	L_{FJ}	L_{Loc}				
Analytical [1]		na	-	77.4 ± 32.2	15.0 ± 6.0	0.32 ± 0.26	0.17 ± 0.15
MLP [14]	✓	-	-	42.0 ± 17.3	12.5 ± 1.5	0.05 ± 0.03	0.07 ± 0.05
DL [17]	✓	-	-	33.2 ± 23.4	20.1 ± 4.6	0.46 ± 0.28	0.32 ± 0.06
FCN [18]	✓	✓	-	14.8 ± 7.40	21.7 ± 0.9	0.73 ± 0.03	0.64 ± 0.10
FCN+ L_{Loc}	✓	✓	✓	13.6 ± 7.20	21.3 ± 2.1	0.74 ± 0.21	0.63 ± 0.05
RAN	✓	-	-	14.3 ± 7.10	18.1 ± 6.1	0.76 ± 0.11	0.70 ± 0.04
Multitask-Net	✓	-	✓	12.1 ± 6.41	23.8 ± 1.7	0.77 ± 0.18	0.69 ± 0.09

Fig. 7. (A) Clinical data scan procedure. Dashed blue lines delineate the different probe sweeps. (B-C) Qualitative reconstruction performance of our model compared to state-of-the-art analytical technique on different real patient data. (i) Reconstruction of healthy breast of candidate versus cancerous lesion over different probe position (ii-iii). Note how the size of the lesion reconstructed by our method more faithfully reflects the relative difference in size. (approximate lesions sizes obtain with ultrasound; patient 1 (B): $22 \times 21 \times 23$ mm, patient 2 (C): $19 \times 16 \times 18$ mm)

respect to their relative sizes. In addition, our method obtains a more precise delineation of the lesions.

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 equi-spaced 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-probe (requiring new hardware design, testing, and manufacturing), we reported experimental performance results for the in silico dataset in Fig. 4-C and visual inspection results in Fig. 8.

Note the high contrast between the background and the lesion area despite the increased difficulty to properly recover lesion size and location by increasing the sparsity of the signal (lower number of detectors). The results are on par with

Fig. 8. Effect of varying number of detectors on the qualitative result. 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.

qualitative results (Fig. 4-C) where we notice an increase in localization error while we observe a slight decrease in SSIM.

V. DISCUSSION

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 that 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 [39]. 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 ability of our network to successfully find the inverse mapping by learning only from the simulation data without seeing any real or physical phantom data during the training/validation process can be viewed as evidence that it can learn a general inverse function instead of learning a trivial mapping or memorizing examples. The addition of transfer learning and Gaussian noise to simulated data helps further the network to generalize to unseen data from physical sensors that capture real signals under non-ideal acquisition conditions. Moreover, unlike the analytical approach that requires a reference measurement of healthy tissue to compute the inverse image, our network 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 on the 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 [40].

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 (lesions location ≥ 35 mm) remains a challenging task in DOT, especially in case of a linear probe as it is harder to differentiate the signal from the tumor-free tissue signal. As a result, the reconstructed image can be biased towards tumor-free tissue. Similarly, occluded lesions and lesions situated at the border of the probe can be miss-reconstructed where only one lesion is reconstructed, lesion location deviates, or tumor-free tissue is reconstructed. 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 warrants validation on larger and more diverse datasets. However, this is not a technical limitation of the model as more tests can be done as more real data become available.

VI. CONCLUSION

We are the first 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 forward projection simulators 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 the inherently small real-world datasets, without compromising accuracy where more complex sources of artifacts and noise are present. We confirmed that on real-world data we obtained results consistent with the analytical method, yet much faster to compute, paving the way for real-time application in a clinical setting.

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] O. K. Lee and J. C. Ye, "Accurate inversion of absorption and scattering in diffuse optical tomography without iterative green's function update," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 661–664.
- [7] S. R. Arridge, "Optical tomography in medical imaging," *Inverse problems*, vol. 15, no. 2, p. R41, 1999.
- [8] V. E. Pera *et al.*, "Spatial second-derivative image processing: an application to optical mammography to enhance the detection of breast tumors," *Journal of Biomedical Optics*, vol. 8, no. 3, pp. 517–525, 2003.
- [9] H. Ben Yedder, B. Cardoen, and G. Hamarneh, "Deep learning for biomedical image reconstruction: A survey," *Artificial Intelligence Review*, pp. 1–33, 2020.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] U. S. Kamilov *et al.*, "Learning approach to optical tomography," *Optica*, vol. 2, no. 6, pp. 517–522, 2015.
- [14] 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.
- [15] J. Yoo *et al.*, "Deep learning diffuse optical tomography," *IEEE Transactions on Medical Imaging*, 2019.
- [16] 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.
- [17] H. B. 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.
- [18] H. B. 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 421–429.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] F. Chollet *et al.*, "Keras," <https://github.com/fchollet/keras>, 2015.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] W. Lam *et al.*, "Factors affecting the palpability of breast lesion by self-examination," *Singapore medical journal*, vol. 49, no. 3, p. 228, 2008.
- [38] M. Shokoufi, "Multi-modality breast cancer assessment tools using diffuse optical and electrical impedance spectroscopy," Ph.D. dissertation, Simon Fraser University, 2016.
- [39] F. Beca and K. Polyak, "Intratumor heterogeneity in breast cancer," in *Novel Biomarkers in the Continuum of Breast Cancer*. Springer, 2016, pp. 169–189.
- [40] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *Journal of Machine Learning Research*, vol. 20, pp. 1–21, 2019.