Published in final edited form as:

IEEE Trans Med Imaging. 2023 August ; 42(8): 2439–2450. doi:10.1109/TMI.2023.3252576.

FDU-Net: Deep Learning-Based Three-Dimensional Diffuse Optical Image Reconstruction

Bin Deng,
Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Charlestown, MA 02129 USA and Harvard Medical School (HMS), Boston, MA 02115 USA.

Hanxue Gu,
Martinos Center at MGH and are currently with the Department of Electrical and Computer Engineering at Duke University.

Hongmin Zhu,
Martinos Center at MGH and are currently with the People eXperience and Technology, Amazon.com Services LLC.

Ken Chang,
Harvard-MIT Division of Health Sciences and Technology (HST) at the Massachusetts Institute of Technology (MIT) and is currently with the Department of Medicine at Memorial Sloan Kettering Cancer Center.

Katharina V. Hoebel,
HST at MIT.

Jay B. Patel,
HST at MIT.

Jayashree Kalpathy-Cramer,
Martinos Center at MGH and HMS and is currently with the Department of Ophthalmology at the University of Colorado Anschutz Medical Campus.

Stefan A. Carp
Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Charlestown, MA 02129 USA and Harvard Medical School (HMS), Boston, MA 02115 USA.

Abstract

Near-infrared diffuse optical tomography (DOT) is a promising functional modality for breast cancer imaging; however, the clinical translation of DOT is hampered by technical limitations. Specifically, conventional finite element method (FEM)-based optical image reconstruction approaches are time-consuming and ineffective in recovering full lesion contrast. To address this, we developed a deep learning-based reconstruction model (FDU-Net) comprised of a Fully connected subnet, followed by a convolutional encoder-Decoder subnet, and a U-Net for fast, end-to-end 3D DOT image reconstruction. The FDU-Net was trained on digital phantoms that include

---

Corresponding author: B. Deng (bin.deng@mgh.harvard.edu).

B. Deng and H. Gu contributed equally to this work.

randomly located singular spherical inclusions of various sizes and contrasts. Reconstruction performance was evaluated in 400 simulated cases with realistic noise profiles for the FDU-Net and conventional FEM approaches. Our results show that the overall quality of images reconstructed by FDU-Net is significantly improved compared to FEM-based methods and a previously proposed deep-learning network. Importantly, once trained, FDU-Net demonstrates substantially better capability to recover true inclusion contrast and location without using any inclusion information during reconstruction. The model was also generalizable to multi-focal and irregularly shaped inclusions unseen during training. Finally, FDU-Net, trained on simulated data, could successfully reconstruct a breast tumor from a real patient measurement. Overall, our deep learning-based approach demonstrates marked superiority over the conventional DOT image reconstruction methods while also offering over four orders of magnitude acceleration in computational time. Once adapted to the clinical breast imaging workflow, FDU-Net has the potential to provide real-time accurate lesion characterization by DOT to assist the clinical diagnosis and management of breast cancer.

Keywords

Convolutional neural network; Deep Learning; Diffuse optical tomography; Inverse problem; Image reconstruction; Breast cancer

I. INTRODUCTION

Nnear-Infrared (NIR) diffuse optical tomography (DOT) is an emerging functional imaging technology for cost-effective, safe, and noninvasive assessment of breast cancer [1], [2]. Using non-ionizing NIR light, DOT can capture abnormal deep tissue vasculature and metabolism through quantitative *in vivo* tomographic images of total hemoglobin concentration and average tissue hemoglobin oxygenation. Over the years, independent clinical studies have repeatedly demonstrated the potential of DOT in differentiating between malignant and begin breast lesions [3]–[8], as well as in the early prediction of the outcome of neoadjuvant therapy [9]–[11], especially when used in a multi-modal approach in combination with clinical breast imaging modalities.

Despite its sensitivity to cancer pathophysiology and success in pilot studies, DOT faces a few limitations attributed to the intrinsic challenges in solving the ill-posed DOT inverse problem [12]. To ensure a stable solution, regularization is typically required at the cost of the loss of spatial resolution and contrast recovery, which results in the current utility of DOT being predominantly limited to imaging large lesions over 1–1.5 cm in size. Yet, in the diagnostic setting in the U.S., over 60% of all biopsy-confirmed breast cancers detected with screening mammography are smaller than 1.5 cm [13]. Efforts in utilizing more sophisticated reconstruction methods with compressed sensing and sparse regularization have been attempted [14]–[16] but only achieved incremental improvement in image quality. Fundamentally, these algorithms are still built upon the current paradigm of DOT image reconstruction that uses iterative nonlinear optimization together with a finite-element numerical model of light propagation. These conventional methods are not only time-consuming but also require trained experts to optimize performance, making DOT less appealing for clinical translation.

In recent years, the use of various deep neural networks (DNN) architectures has become an active research area for image reconstruction in low-dose computed tomography (CT), positron emission tomography (PET), and compressed sensing magnetic resonance imaging (MRI) [17], [18]. Despite different underlying physics across imaging modalities, consistent and significant improvements have been achieved over existing methods in the quality of recovered images. Strictly data-driven DNN models have shown remarkable generalization with superior noise and sampling defect immunity [19]. Once trained, it takes as little as a few milliseconds to obtain a 2D image slice or a 3D volume using a single feed-forward inference, significantly increasing the reconstruction speed. The ability of deep learning (DL) methods to instantaneously recover high-quality images from noisy, under-sampled acquisitions presents a novel opportunity to address the limitations in DOT image reconstruction.

Indeed, several recent works in the field of nonlinear diffuse tomographic imaging [20]–[25], including fluorescence tomography [26]–[28] and photoacoustic tomography [29], have shown early promise of DL-based approaches for end-to-end direct sensor-to-image domain mapping for optical image reconstruction. In the domain of DOT, Feng *et al.* [20] trained a three-layer neural network and demonstrated significantly improved quantification accuracy over the conventional Tikhonov regularization-based reconstruction method in simulated two-dimensional (2D) circular phantoms. The same group further proposed a 2D Z-Net [24] that also used MRI images as model input together with multispectral optical signals to enhance image quality and demonstrated its ability to differentiate a large malignant lesion from a benign case. Yedder *et al.* [21] proposed a DNN model comprising a fully-connected layer, a set of convolutional layers, and a final integration layer. Though trained on 2D synthetic data, the model could recover optical images from real measurements on liquid phantoms with better quality than the analytical method. Yoo *et al.* [25] trained a three-dimensional (3D) DNN model to invert the nonlinear Lippman-Schwinger equation with simulated optical data and applied the DL model to *in vivo* animal experiments. The model was successful in recovering the 3D location of anomalies more accurately than the iterative FEM-based method using the Levenberg-Marquardt algorithm. Zou *et al.* [22] proposed another 3D machine learning model with physical constraints (ML-PC) implemented in the loss function and trained it with mixed simulated and real measurement data. Compared to the Born conjugate gradient descent method, the ML-PC model showed reduced mean percentage error of reconstructed maximum absorption coefficient within the inclusion and improved depth distribution in phantom cases. The ML-PC approach was further tested in 10 patient cases and demonstrated potential for improved differentiation between malignant and benign lesions.

While promising, prior works using DL-based approaches in DOT image reconstruction have focused primarily on the overall image quality and validation in a small number of sample cases, generally with large sizes and high contrasts. This paper presents a novel 3D deep convolutional neural network, the FDU-Net, to directly reconstruct optical images of absorption coefficients from simulated data with commensurate noise profiles as real DOT measurements. We further compare the performance of FDU-Net with another DL-based network similar to prior work [25] and with conventional FEM-based approaches in 400 simulated test cases (100 each in four size groups of 8-, 10-, 12-, and 16-mm in diameter) with inclusion contrasts aligned with those observed in our prior study on DOT for the diagnosis of breast cancer [4]. Afforded by the existence of ground truth, these testing cases allow us to systematically evaluate the capability of the DL-based model in resolving inclusion contrast, location, and size, especially in conventionally challenging small-size and/or low-contrast cases. This first-of-its-kind systematic performance evaluation aims to determine if the apparent superiority of DL-based methods depends on lesion characteristics such as size and contrast. This work will help us better understand the advantages and limitations of deep learning models to guide the design of future model architecture for the robust reconstruction of DOT images.

II. METHODS

A. Forward Simulation of DOT Measurements

One challenge in adopting deep learning methods for DOT image reconstruction is the lack of ground truth knowledge of the heterogeneous distribution of tissue constituents that give rise to the contrast seen in DOT images. A common approach is to use simulated optical measurements on digital phantoms. In this work, a 3D breast shape was positioned inside a 216 × 144 × 60 mm^3 bounding volume, as shown in Fig. 1. To simulate breast lesions within the breast boundary, a single spherical inclusion with diameters ranging from 8 to 16 mm (mean ± std: 12.00 ± 2.32 mm, uniform distribution) was inserted at random x-y locations at the central depth ($z = 30\ \mathrm{mm}$). Based on this geometry, dual-resolution tetrahedral meshes composed of a fine mesh (mean nodal distance of 0.36 mm) within the inclusion region and a coarse mesh (mean nodal distance of 5.92 mm) outside were generated for each geometry using a MATLAB-based meshing tool *iso2mesh* [30].

As shown in Fig. 1, 48 continuous-wave (CW) sources and 32 detectors, spatially distributed in the same pattern as in our clinical multimodal DOT/mammography breast imaging system [31], were used to generate a set of 1,536 simulated CW optical measurements at 690 nm using a diffusion equation-based forward model [12] for each digital phantom. To this end, absorption coefficients (μ_a) representative of typical tissue optical properties [32]–[34] were randomly assigned to the inclusion (0.17 ± 0.05 cm^{-1}) and homogenous background (0.06 ± 0.01 cm^{-1}), respectively, in each phantom case, resulting in a range of inclusion-to-background absorption coefficient ratios from 1.12 to 5.12 (v), while reduced scattering coefficients (μ_s') were fixed at 9.25 cm^{-1} across phantoms. The choice of this fixed μ_s' value was guided by the bulk μ_s' measured in our prior large-population study [4], which is consistent with widely accepted literature values reported by other groups [35], [36], where the mean and standard deviation of μ_s' at 690 nm is 9.8 ± 2.0 cm^{-1}. The diffusion equation was then numerically solved on the dual-resolution mesh using a MATLAB-based finite-element solver *Redbird-m* [37], [38].

B. Noise Model

To ensure the simulated data resemble the signal amplitude and noise profiles of real measurements, DOT data of a homogenous silicone rubber slab phantom (6.5-cm thick, $\mu_a = 0.024\ \text{cm}^{-1}$ and $\mu_s' = 7.275\ \text{cm}^{-1}$ at 690 nm, made by mixing Smooth-On Silc Pig White & Black pigments into Smooth-On Ecoflex 00–50 silicone) collected by the multimodal optical imaging system in [31] were used to characterize the dynamic range and noise levels. A forward simulation based on the same geometry and optical properties is performed the same way as described above to serve as a reference. Using the set of measured and simulated fluence at each source and detector pair, we first determined the global scaling factor α that brings the simulated data in the range of the experimental data as:

$$arg \min_{\alpha} \{\| \alpha \cdot \Phi_s(s, d) - \Phi_m(s, d) \|_2\} \tag{1}$$

where s is the index of source optodes; d is the index of detector optodes; $\Phi_m(s, d)$ is the measured fluence data averaged over four measurements of the silicone rubber phantom; $\Phi_s(s, d)$ is the simulated data using the forward model described above; and $\| \cdot \|_2$ denotes the $L2$ norm.

Then, two types of pseudo-random noise, namely the signal-independent electronic noise and the signal-dependent shot noise [39], were added to the scaled forward model output using:

$$\Phi_s'(s, d) = \alpha \cdot \Phi_s(s, d) + n_{elec} + n_{shot} \tag{2}$$

$$n_{elec} = \sigma_{elec} \times |U_1| \tag{3}$$

$$n_{shot} = \sigma_{shot} \times U_2 \times \sqrt{|\alpha \cdot \Phi_s(s, d)|} \tag{4}$$

where $\Phi_s'(s, d)$ is the scaled simulated data with noise added; n_{elec} and n_{shot} are electronic and shot noise, respectively; U_1 and U_2 are two independent random variables with the standard normal distribution; σ_{elec} and σ_{shot} are factors that control the level of added electric and shot noise, respectively, which are determined by the following minimization process:

$$arg \min_{\sigma_{elec}, \sigma_{shot}} \{ \| \Phi_s'(s, d) - \Phi_m(s, d) \|_2\} \tag{5}$$

As shown in Fig. 2, the scaled simulated data with added noise (red stars) can accurately recapitulate the characteristics of signals collected from real measurements (blue squares), including the noise floor that is determined by the detection limit of the optical imaging hardware. The same global scaling factor α and noise factors σ_{elec} and σ_{shot} are then applied to all *Redbird-m* forward outputs.

C. Preprocessing and Ground Truth

Redbird-m forward output data with added noise were rescaled using the minimum and maximum values across simulated data from all phantoms in the training dataset, *i.e.,* the population min-max normalization. All 1,536 data points (48 sources × 32 detectors) of each simulation were included as the input to train the deep learning models, including those that are typically excluded in conventional DOT image recon methods due to low signal-to-noise ratio (data to the right of source-detector separation limit shown in cyan dotted line in Fig. 2). To decrease the number of parameters that need to be trained, thus reducing the computation cost, we downsampled the original image volume of 216 × 144 × 60 mm^3 to form the ground truth 3D volume of 72 × 48 × 12 voxels with a resolution of 3mm × 3mm × 5mm. This image downsampling rate was determined by a few considerations. First, the downsampled voxel should be small enough to preserve inclusion information for the smallest 8-mm diameter inclusions. Second, we chose a slightly higher factor of 5 in Z than the 3 × 3 in-plane downsampling rate as DOT is known to have limited Z resolution. Moreover, in the parallel-plate geometry, the XY localization accuracy is of higher clinical importance. The downsampling represents a balanced trade-off between the potential image resolution limited by the physics of NIR light propagation in tissue and training efficiency. The population min-max normalization was also applied to the ground truth optical images of μ_a at 690 nm to boost the rate of convergence. The minimum and maximum values used to normalize the inputs and outputs of the neural network were logged and used later to preprocess the testing dataset and restore model outputs to meaningful values of absorption coefficients, respectively.

D. Deep Learning Neural Networks and Training

1) Network Structure—As shown in Fig. 3, we implemented our DOT image reconstruction framework with a deep neural network, named FDU-Net, consisting of **F**ully connected layers and a 3D convolutional encoder-**D**ecoder followed by a **U-Net** to further enhance inclusion features and reduce noise. First, the front part of the network has two fully connected layers, *i.e.*, the input layer and a hidden layer, activated by the rectified linear unit (ReLU) function and then a reshape operation to transfer the sensor-domain data (1,536 linear array) to image-domain (72 × 48 × 12 voxels). These fully connected layers function as an approximation of the inversion operator that resolves the spatial distribution of absorbers from a finite number of measurement pairs. To enhance the robustness of the fully connected network and prevent overfitting, an empirically determined dropout rate of 0.8 was used [40]. Then, the reshaped volume was passed onto a 3D convolutional autoencoder network to extract inclusion features, *e.g.,* location and contrast, presented as small, localized perturbation. Finally, the intermediate model output was fed into a 3D U-Net [41], [42] with three resolution steps. Each layer in the analysis path consists of two 3×3×3 padded convolutions with ReLU activation followed by a 2×2×2 max pooling operation with strides of two in each dimension for downsampling. Starting with 32 feature channels, the number of feature channels was doubled at each downsampling step. In the synthesis path, each layer consists of an up-convolution of 2×2×2 by strides of two followed by one 3×3×3 convolution with ReLU activation. Skip connections from layers of equal resolution in the analysis path provide the high-resolution features to the synthesis path. The final layer, a 1×1×1 convolution, is used to map the 32-channel features to the unified final 72 × 48 × 12 image volume.

2) Prior-Weighted Loss Function—Since the embedded single spherical inclusion constitutes only a tiny fraction (0.14 – 1.15‰) of the entire 3D image volume and the μ_a contrast between inclusion and background (inclusion-to-background ratio $R = 2.73 \pm 0.87$, range 1.12 – 5.12) is relatively small, attempts to use the typical $L2$ loss function were not effective in recovering such small perturbations. Inspired by the concept of prior-guided reconstruction methods [43]–[46] that have been proven advantageous in recovering small lesion contrast in conventional DOT image reconstruction, we introduced a loss function that penalized more heavily on inaccuracies within the inclusion region-of-interest (ROI). The loss function took advantage of the prior knowledge of the location and size of the inclusion to impose different weights on the $L2$ loss within and outside the inclusion ROI:

$$L_{\text{roi}} = \omega_{roi} * \| y_{roi}^{output} - y_{roi}^{truth} \|_2^2 + (1 - \omega_{roi}) * \| y_{bg}^{output} - y_{bg}^{truth} \|_2^2 \tag{6}$$

where L_{roi} is the prior-weighed loss function; y^{truth} and y^{output} are the ground truth optical image and model output with subscription "roi" and "bg" denotes the values within the inclusion ROI and background image volume, respectively; ω_{roi} and $(1 - \omega_{roi})$ are the weights imposed to the $L2$ loss within and outside the inclusion ROI. By setting the value of ω_{roi} between 0.5 and 1, we control the strength in reinforcing accuracy within the inclusion. This prior-weighted loss function is somewhat analogous to the class weighting strategy used in the work of nnU-Net by Isensee *et al.* for biomedical image segmentation [47].

The ω_{roi} setting was determined empirically. The optimal values used in the fully trained model were $\omega_{roi} = 0.985$ and 0.9, respectively, for the intermediate and final training steps, which are close to the volume fraction of the background. During training, using the prior-weighted loss function as defined in Equation (6) was observed to accelerate the learning process and improve the accuracy of inclusion location recovery using the input data that revealed no explicit information on system configuration. It is important to note that the prior-weighted loss is only relevant during training; once trained, no prior knowledge is used to recover the optical images in the testing dataset using the FDU-Net model.

3) Two-Step Training—Overall, a total of 3,445 cases with randomized lesion sizes, locations, and optical properties were generated and randomly split into training (95%) and validation datasets (5%). We used the TensorFlow [48] framework to train the neural network on a NVIDIA Titan RTX GPU using the Adam gradient descent optimization algorithm [49]. A two-step training strategy was used to train the three-part FDU-Net. First, we trained the fully connected and the convolutional autoencoder network, *i.e.,* the first two parts, together for 300 epochs with starting learning rate $\alpha = 0.001$ and exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$, values recommended by the original Adam paper [49]. Note that though empirically determined, we have validated via training loss plots and recovered image quality that 300 epochs are sufficient to ensure model convergence with these hyperparameters during training. Once trained, the weights of the first two parts were fixed and training of the third part, *i.e.,* the 3D U-Net, proceeded. The learning rate was reset to 0.001, and the network was trained until the validation loss did not reduce further in the past 10 epochs. On a Linux CentOS workstation equipped with two 16-core 2.10 GHz Xeon Gold 6130 CPUs and a Titan RTX GPU with 24 GB memory, it took about 5 hours to complete the two-step training of the 68,625,314 parameters of the FDU-Net.

E. Performance Evaluation and Comparison

A separate set of 400 cases was generated, 100 each for 8-, 10-, 12- and 16-mm diameter inclusions, respectively, with randomized locations and optical properties (inclusion-to-background ratios $R = 2.71 \pm 0.80$, range 1.13 – 4.85). They were used for testing and comparing model performance among various methods. In addition, testing cases with irregularly shaped inclusions and with two separate inclusions of randomized diameters and contrasts were created to evaluate model performance in recovering previously unseen shapes.

Several reference DOT image reconstruction methods were used to put the performance of the FDU-Net in the context of other typically used methods. First, we used the widely adopted finite-element (FEM) based Tikhonov-regulated iterative Gauss-Newtown reconstruction method implemented by an in-house inverse problem solver *Redbird* [38], as the reference to compare performance with the proposed FDU-Net model. Two FEM image reconstruction approaches were used – one used optical data only without prior knowledge of inclusion location to resemble the use cases of standalone DOT systems (FEM-noprior), and the other leveraged the size and location of the inclusion to further regularize the solution [43], [44], resembling the use cases in multimodal implementations of DOT where co-registered anatomical breast images from a complementary medical imaging modality can be used as structural priors (FEM-prior). Regularization parameters used in FEM-based reconstructions were optimized by the standard “L-curve” approach [50] with no further fine-tuning on each individual case.

In addition, to evaluate the performance of FDU-Net against other DL-based networks previously proposed for the same DOT image reconstruction task, we constructed a network we call FD-Net as a DL-based reference model. Structurally, the FD-Net is equivalent to the proposed FDU-Net without the U-Net component, and comprises only the **F**ully connected layers followed by a convolutional encoder-**D**ecoder, similar to the architecture used in Yoo’s work [25]. The FD-Net was also trained using the same hyperparameters and MSE loss as in Yoo’s work. It is worthwhile to point out that despite the similarities in architecture between the reference FD-Net and Yoo’s model, differences exist. For example, the FD-Net has different numbers of input and output channels due to the differences in numbers of DOT data pairs and reconstructed image volume. Moreover, the lack of public benchmark datasets for the purpose of DOT image reconstruction algorithm development makes the FD-Net we trained inevitably different from Yoo’s model. As such, we can only cautiously extend conclusions based on the observations in performance differences between the FDU-Net and FD-Net to a comparison of the FDU-Net vs. Yoo’s model.

To quantitatively compare the overall quality of the recovered images, automated objective image quality metrics, such as root mean squared error (RMSE), structural similarity index

(SSIM), and peak signal-to-noise ratio (PSNR) were calculated by MATLAB functions *"rmse"*, *"ssim"*, and *"psnr"*, respectively, using the 3D ground-truth optical images as reference. Further, to evaluate the performance in inclusion recovery, the primary focus of this work, we used the MATLAB function *"graythresh"* to perform automatic image thresholding based on the Otsu's algorithm [51] to segment the central slice of the recovered optical images into two classes – background and inclusion. Based on the Otsu segmentation, three metrics were quantified on the recovered optical images at the central depth to evaluate the accuracy of inclusion recovery objectively:

1. Inclusion-to-background ratio (R): Defined as the ratio between the mean optical property within the inclusion region and that within the background region. The closer to the ground-truth R, the higher accuracy in contrast recovery.

2. Reconstructed-to-true diameter ratio: Defined as the ratio between the equivalent reconstructed diameter $\left(d_e = \sqrt{A/\pi} \times 2\right)$ calculated on the area of segmented inclusion region (A) and the ground-truth inclusion diameter. The closer to 1, the more accurate in size recovery.

3. Inclusion center offset: Defined as the Euclidean distance on the XY-plane between the maximum pixel on the reconstructed image of the segmented inclusion region and the centroid of the spherical inclusion of ground truth. The closer to 0, the higher the localization accuracy.

III. RESULTS

Fig. 4 shows images of absorption coefficients at 690 nm for three sample cases recovered by the proposed FDU-Net, the FD-Net similar to prior work [25], and conventional FEM-based methods, both with (FEM-prior) and without (FEM-noprior) prior guidance. In all cases, the images recovered by FDU-Net (2nd column) result in lower RMSE, higher SSIM, and higher PSNR than other methods. The background recovery is also more homogenous that closely resembles the ground truth image. More importantly, inclusion contrasts are better preserved, as evidenced in notably higher inclusion-to-background ratios R. The average contrast recovery ratio among these three sample cases is 92.0%, compared to 36.5%, 39.5% and 50.1%, respectively, using FD-Net (3rd column), FEM-noprior (4th column) and FEM-prior (5th column) methods. FDU-Net recovered inclusions are also in the close vicinity of the ground-truth locations. Note that such localization precision is achieved without explicitly informing the reconstruction process about the true inclusion location as used in the FEM-prior method. The reference FD-Net model performs reasonably well in identifying inclusion location; however, it is very limited in recovering contrast. We have also benchmarked the time cost for each reconstruction method. By leveraging GPU computing, the image reconstruction using the trained DL models, *i.e.,* FDU-Net and FD-Net, took about 20 ms to complete through a single inference. In comparison, FEM-noprior and FEM-prior took a total of 12 min and 3 hours, respectively, for a complete reconstruction consisting of 10 iterations of optimization. This means that the DL-based approaches can be four orders of magnitude faster than the conventional FEM-based methods.

Findings in the sample cases are well represented in the group analyses of all 400 testing cases. As shown in Fig. 5a, in all size groups, the median RMSEs of images recovered by FDU-Net (red boxes) are significantly less than other methods. Especially, RMSEs of FDU-Net are less than half of conventional FEM-based methods, suggesting much improved overall quantification accuracy. Furthermore, the SSIM and PSNR of the FDU-Net approach is the highest, while that of the FEM-noprior method is the lowest among all reconstruction methods in all size groups, as shown in Figs. 5b and 5c. Compared to FEM methods, the variation in RMSE and SSIM across images, indicated by the tightness of the box, is also smaller using the FDU-Net method, suggesting consistency in the quality of reconstructed images.

The scatter plots in Fig. 6 further focus on the size-dependent quantification accuracy in the recovery of inclusion contrast R. In all four methods, as the size of the inclusion increases, a higher percentage of contrast is recovered in more cases. When the inclusion size reaches 16-mm diameter (Fig. 6d), the vast majority of recovered contrast values by FDU-Net (red asterisks) track along the ideal recovery reference line. To allow quantitative comparisons, linear fits of the data points in each size group and reconstruction methods were performed. The resulting slopes, intercepts, and coefficients of determination of the fit $\left(\mathrm{R}^2\right)$ are shown in Table 1. Using the FD-Net and FEM-noprior method, the recovered inclusion-to-background ratios R stay around 0.99 and 1.12, respectively, regardless of inclusion size and ground-truth R (evidenced by near 0 slopes). As shown in Figs. 4h and 4i, although this contrast level can still be visually discerned, interpretation of an optical image of this quality can be challenging due to the commensurate background. Using the FEM-prior method, the R^2 values in 12- and 16-mm size groups are improved significantly but remain low in smaller 8- and 10-mm size groups $\left(\mathrm{R}^2 < 0.10\right)$. In contrast, the R^2 values for FDU-Net start from a moderately high value of 0.38 even in the smallest 8-mm diameter size group and increase steadily to 0.78 in the 16-mm size group with a slope of 0.80.

Moreover, to compare the physical properties of recovered inclusions, Fig. 7 shows the box plots of size recovery ratios and inclusion center offsets of the four size groups using DL- and FEM-based methods. Though notably better localization accuracy is seen in the 16-mm group, the FEM-noprior method (black boxes) in general yields overestimation in size (median >3.0) and large offsets from the ground-truth centroids (median >40 mm) in all other size groups, indicating erroneous inclusion recovery. Using the FEM-prior method (blue boxes), the median recovery size ratios and offsets approach ideal values of 1 and 0, respectively, in the 16-mm group. Moderate improvement in the 10- and 12-mm groups over the FEM-noprior is observed. Yet, the FEM-prior method still struggles, as evidenced in the wide spread of the 25th to 75th percentile boxes, especially in size recovery. Interestingly, as shown in Fig. 7b, the reference FD-Net (green boxes) demonstrates consistent localization performance with medians of ~20 mm off the ground-truth locations in all size groups, while showing a comparable performance in reconstructed-to-truth diameter ratio to FEM-based methods. In contrast, the FDU-Net model (red boxes) shows clear superiority in all size groups. For 10-mm diameter and above, especially, the medians of both metrics are close to the ideal values. Even in the smallest 8-mm size group, the size recovery is significantly improved despite the challenge in localization.

Finally, to demonstrate the generalizability of the FDU-Net model, reconstructed images of an irregularly shaped inclusion case and two two-inclusion cases, one of the same sizes but different Rs and the other of different sizes and Rs, are shown in Fig. 8. Although trained exclusively on single spherical inclusion cases, the FDU-Net model could still recover the contrast, shape, and location with precision for an irregularly shaped inclusion as shown in the top row of Fig. 8. Among other methods, the shape recovery is only successful using FEM-prior, where the inclusion shape and location information were used explicitly to guide the image reconstruction. For the two-inclusion cases, the FDU-Net is the only method to correctly identify and recover the most contrast in both cases. Meanwhile, the reference FD-Net recovers only one inclusion, either the one with the higher R or larger size. For the same-size different-R two-inclusion case (middle row), the FDU-Net also preserves the relative contrast between the two inclusions; however, when inclusions of two different sizes are present, the FDU-Net picks up the larger size inclusion better than the smaller ones, as all other methods. The capability of the FDU-Net in recovering previously unexposed shapes further highlights the potential of the FDU-Net model to accurately reconstruct the optical property distributions corresponding to realistically shaped and potentially multi-focal lesions.

Beyond the simulated inclusion cases shown thus far, using the FDU-Net model trained entirely on simulated data, we have further recovered optical images from a real DOT measurement on a 42-year-old woman with an 18×21 mm invasive ductal carcinoma. The measurement protocol, detailed in prior work [31], has been approved by the Institutional Review Board at MGH, and a written consent was obtained from the patient. As shown in Fig. 9, the trained FDU-Net model is able to recover an absorption increase due to the presence of the malignant tumor at the correct location with higher contrast than the conventional FEM-noprior method. This is an encouraging preliminary result obtained without any further training of the FDU-Net using experimental data. Also, note the difference in the optode density between Figs. 9a and 9b – the FDU-Net only used 43.2% of DOT data (only 1,536 CW data pairs at 690 nm) as inputs compared to the FEM method (all 3,552 CW and frequency-domain data pairs at both 690 and 830 nm) during image reconstruction. Again, this result was achieved with a significantly accelerated (20 ms vs. 1.5 hours in this case) and fully automated process. However, there is some mismatch of tumor boundaries between the recovered optical contrast and x-ray mammogram-based tumor marking. This is likely due to multiple factors including the limited resolution of DOT compared to x-ray mammography, the somewhat subjective nature of the lesion boundary marked by a breast radiologist on a mammogram acquired separately from the DOT imaging data, and, importantly, the indirect structural-to-functional contrast relationship of breast tumors [52], [53]. While large-population validation is needed, this case study supports the feasibility of deploying the FDU-Net trained on simulated data for DOT image reconstruction of real patient measurements.

IV. DISCUSSION

In this work, we constructed a 3D deep convolutional neural network and trained it with simulated DOT data on cases with a simple geometry for direct sensor-to-image domain transformation, *i.e.*, the image reconstruction task. We hypothesized that the network could learn the correspondence between optical contrast and boundary measurements across space, and complex internal contrast distributions can then be reconstructed additively from the superposition of several elementary building blocks. Empirically, this hypothesis is supported by our results showing high performance on test cases with multiple or irregularly shaped inclusions, as well as on a real patient case.

The FDU-Net architecture is informed by the structures proposed in AUTOMAP and Yoo *et al.* [19], [25], typically composed of fully connected layers followed by a convolutional encoder-decoder sub-network. Such a general framework has been proven effective in solving various inverse problems by using fully connected layers first to approximate the projection from the sensor domain to the image domain and then utilizing the sparse autoencoder to extract high-level features from the image domain. Although most prior works have tested the feasibility of deep learning for the image reconstruction task in 2D, given that DOT measures the 3D distribution of tissue constituents, it is necessary to utilize 3D networks to recover tomographic images for applications in breast imaging. Specifically, in the scenarios of supplementing digital breast tomosynthesis (DBT) and clinical MRI with functional DOT for breast cancer screening and diagnosis, 3D localization is required to characterize suspicious lesions.

The FDU-Net introduced in this work has made several advances on the previous general framework. First, a 3D U-Net was included in the back end of the model to enhance inclusion features further. By utilizing skip connections that transfer the high-resolution information from the low-level layers of the analysis path to the high-level layers of the synthesis path, U-Nets can leverage both local and global information to enhance the image quality. The resulting three-part FDU-Net leads to not only precise localization but also to a significant improvement in quantification accuracy, the lack of which was a limitation of the previously proposed model [25] similar to the reference FD-Net shown in this paper. To help us better understand the role of each subnet, we have constructed several variations of the FDU-Net and performed ablation tests (results not shown). We found that the convolutional encoder-decoder is effective in suppressing artifacts away from the ground truth location to achieve localization accuracy, which explains the consistent localization performance of the FD-Net shown in Fig. 7b, while the U-Net is essential in enhancing contrast recovery. Second, a unique prior-weighted loss function was introduced during training to effectively characterize small perturbations at the lesion location in the optical reconstructions within a noisy background. This idea is inspired by the structural prior guided approaches [43]–[46] widely used in the conventional FEM methods. The difference is that the prior information is *only used during training* for loss calculation, and no prior knowledge of inclusion size and location is needed during the reconstruction using the trained FDU-Net. As shown in Fig. 8, imposing such prior-weighed loss does not result in the model predicting only single spherical inclusion shapes but rather offering sufficient generalizability needed for the trained FDU-Net model to succeed in previously unexposed cases. Finally, the input layer of the FDU-Net indiscriminately took in all CW data shown in Fig. 2 (red stars), including those that fell within and below the noise floor. Data pruning to get rid of noisy and invalid measurement pairs, a cumbersome preprocessing step typically needed in the FEM-based conventional methods, is now intrinsically part of the FDU-Net image reconstruction learned directly through the strictly data-driven approach. This eliminates the need to change the model structure and train separate models for different applications.

The improved overall image quality reconstructed by the trained FDU-Net model is clearly demonstrated in the three sample cases shown in Fig. 4 and group analysis results on RMSE, SSIM, and PSNR metrics of all 400 test cases shown in Fig. 5. The FDU-Net has demonstrated significantly better performance than other methods across all size groups with all metrics. In line with prior work, a deep learning-based approach outperforms conventional methods in solving the DOT inverse problem, yielding images with a higher level of fidelity to ground truth and less noise. Notably, such improvement is achieved with excellent efficiency. The 3D image reconstruction is done by performing a single feed-forward pass through the trained FDU-Net model using all source-detector data pairs without the additional needs of preprocessing or iterative optimization as in the FEM-based methods. This means a full image reconstruction is completed in less than a second with no subjective intervention, greatly lowering the translation barrier of DOT for real-time breast imaging.

Another main objective of this work is to systematically study the performance of deep learning models in inclusion recovery, especially in conventionally challenging cases. It is known that FEM-based methods are limited in reconstructing small and low-contrast lesions. This is again evidenced in Fig. 4, where in comparison to a typical case with a high-contrast 16-mm diameter large inclusion (Case 2, middle row), cases with either a comparable R but half the diameter (Case 1, top row) or the same diameter but half the R (Case 3, bottom row) failed to be recovered using FEM without the guidance of structural lesion priors (4th column). As expected, with priors, the conventional method demonstrated much-improved performance (5th column), especially in localization precision. However, the recovered inclusion contrasts are significantly lower than the learning-based FDU-Net model (2nd column) in all three cases. In images reconstructed by the FDU-Net, the Rs were fully recovered in the two large inclusion cases. Even in the 8-mm inclusion case, where contrast recovery is expected to be impacted by the 3× in-plane downsampling, near half of R was recovered. In contrast, the reference FD-Net, with no additional U-Net and using standard MSE loss, could identify the inclusion location with a degree of success but severely underestimates the contrast. Although the FD-Net is not exactly the same as the model used in prior work by Yoo et al. [25], this observation is consistent.

From Fig. 6, it is clear that success in inclusion recovery is dependent on size and contrast, although, on a case-by-case basis, the FDU-Net method generally yields the highest recovered inclusion contrasts. Using the FDU-Net, in size groups of 10-mm and above, a linear relationship can be observed with at least 50% of ground-truth R recovered in most cases, and the linear fit fast approaches the ideal line when the inclusion size reaches 16-mm in diameter. In comparison, the FEM-noprior method, as well as the FD-Net model, largely failed in size groups smaller than 16-mm. With the help of prior knowledge of ground-truth inclusion size and location, the FEM-prior method starts to be able to recover ~40% of R in size groups of 12-mm and above, especially in high contrast cases with $R \geq 3$. In the smallest size group of 8-mm diameter inclusion, FDU-Net can still recover a substantial percentage of contrast while other methods mostly failed.

However, as the inclusion size decreases under 10-mm in diameter, it becomes challenging even for the FDU-Net to accurately locate the inclusion. As shown in Fig. 7b, in the 8-mm size group, on average, all four reconstruction methods have recovered centroids over 30 mm off from the ground truth. That is almost four times the inclusion diameter, indicating failure in accurate localization. However, in 10-mm and above size groups, the FDU-Net quickly regains localization precision with median offsets staying around 5 mm, which is commensurate to the radius of the inclusions. In comparison, FEM-prior and FEM-noprior methods achieve similar localization accuracy starting at only 12- and 16-mm diameter size groups, respectively. The reference FD-Net, however, maintains similar levels of localization accuracy of ~20 mm off the ground truth across size groups despite a slow decreasing trend. Moreover, since the edges of inclusions recovered by the reference FD-Net and the FEM-based methods are usually blurry and lack a sharp contrast relative to the background, the Otsu segmentation (selected to ensure objectivity in image interpretation) quite frequently enlarged the inclusion region, resulting in overestimated recovered inclusion size as shown in Fig. 7a. Meanwhile, the average recovered size ratios using the FDU-Net are around 1 in size groups 10-mm and above, close to ground-truth inclusion sizes.

The challenges faced by the FDU-Net in recovering the small 8-mm inclusions can be attributed to a few factors. First, the image volume was downsampled to a 3 mm × 3 mm × 5 mm voxel resolution. This leads to a loss of edge contrast of the spherical inclusion and only a very limited number of voxels representing the full ground-truth R. As a result, the recovered lesion characteristics, i.e., R, size, and XY offset, by the FDU-Net in the 8-mm size group are inferior to other size groups. The most direct impact of the downsampling is probably the recovered size. Moreover, the level of noise added to the simulated DOT data could be too large for the FDU-Net to tease out the subtle perturbation to the signal introduced by the inclusion. The realistic noise profile is equivalent to adding a shot noise at an average level of 74.8% of the amplitude of the clean optical signal and electronic noise that results in a ~50 dB dynamic range (20log10) with the percentage of overall added noise ranging from ~0.1% for the highest signal to 4764.1% for the lowest signal, almost an order of magnitude more noise than those used in previous studies from others and our group [43]. This probably has not only led to larger errors in localizing small 8-mm inclusions but also caused larger, low-contrast lesion cases to be susceptible to artifacts as shown in Fig. 4i.

To better understand the limitations of the FDU-Net, we have defined some empirical criteria to classify recovered images into three categories – success, limited success with artifact, and failure. A case is deemed a success if the center of mass of the highest contrast region reconstructed is within 10 mm of the ground-truth centroid and secondary areas of contrast, if any, are lower than 50% of the highest reconstructed contrast. A case is deemed a limited success with artifact if secondary contrasts are larger than or equal to 50% of the highest reconstructed contrast. All other cases are deemed as a failure. We found that the main driver of artifacts or failed reconstructions was indeed low contrast, more so for smaller inclusions. Over 86% of limited success with artifact cases have ground-truth R smaller than 3. About 97% of failed cases have ground-truth R smaller than 3 and over half of the failed cases are of the 8-mm size group. These observations emphasize that DL models, like the FDU-Net, though effective in improving the optical image quality over the conventional FEM-based methods, cannot bypass the fundamental limitations in physics. When the intrinsic signal-to-noise ratio of the DOT data is low, the performance of model prediction will be suboptimal. *Nevertheless,* the FDU-Net is still effective in improving the detectability of an inclusion. The minimum inclusion-to-background ratios R required for 90% cases to be successfully reconstructed are 3.23, 2.88, 2.15, and 1.39 for 8-, 10-, 12- and 16-mm size group, respectively, using the FDU-Net. In comparison, the conventional FEM-prior method never reaches a 90% success rate for the 8- and 10-mm size groups and requires minimum Rs of 3.56 and 2.41, respectively, for the 12- and 16-mm size groups, which is 65–75% higher than the FDU-Net. We expect further performance improvement, particularly in small and low-contrast cases, by using the multimodal DOT approach, where anatomical images such as DBT and MRI can be used jointly with optical signals as inputs in deep learning models to improve quantification accuracy further, as demonstrated by Feng *et al.* [24].

Despite the demonstrated superior performance of the FDU-Net over previously proposed DL network structures and the conventional FEM-based methods in standalone DOT image reconstruction, our work has limitations. Since the FEM-based methods tend to result in larger artifacts towards the surface in the parallel-plate setting, we have only focused on inclusion recovery at the central depth when comparing performance across methods. The FDU-Net needs to be further trained on cases of various inclusion depths to ensure 3D localization. Further, due to the downsampling, performance evaluation in inclusion sizes smaller than 8-mm is limited. Future work is needed to decrease the voxel size at the model output before using FDU-Net for small lesion cases. Moreover, as mentioned above, spatially co-registered anatomical images will likely be needed as model input to further ensure robust performance in cases with small anomalies. Finally, while we have demonstrated the feasibility of using deep learning models trained on synthetic data directly on real measurements in a patient case, as has been repeatedly shown to be feasible in prior works [19], [22], [24], further validation in large-scale human-subject imaging using trained FDU-Net model for a variety of breast sizes, including fine-tuning of model parameters by training on real measurement data, is needed to demonstrate its effectiveness in improving the sensitivity and specificity of breast cancer diagnosis with DOT.

V. CONCLUSION

In this study, a deep learning-based FDU-Net was used to achieve direct sensor- to image-domain transformation for 3D DOT image reconstruction. The model was built upon a 3-part architecture and trained using a unique prior-weighed loss function to enhance inclusion recovery further. The performance of FDU-Net was systematically studied and compared with a DL network analogous to previously proposed ones and conventional FEM-based methods in 400 simulated data with added realistic noise. Using all source-detector data pairs unbiasedly, the FDU-Net can recover images of significantly improved quantification accuracy in a fraction of a second without human intervention. In inclusions of 10-mm diameter and above, the FDU-Net is capable of recovering over 50% of ground-truth inclusion-to-background ratio with less than 5 mm in-plane localization error and near-ideal size, outperforming other methods in every metric, including the FEM-prior method that leverages ground-truth inclusion size and location to guide the image reconstruction. Even in small 8-mm inclusion cases, where conventional methods have largely failed, the FDU-Net is able to demonstrate reasonable performance. With the excellent model generalizability, the trained FDU-Net bears ample potential to be further developed and used as an effective new approach for DOT image reconstruction in breast cancer imaging.

ACKNOWLEDGMENT

The authors would like to thank Dr. Mansi Saksena at the MGH Breast Imaging Division for providing the radiologic marking of the breast cancer patient case and Ms. Jayne Cormier for acquiring the patient optical imaging data.

This work is supported by the U.S. National Institute of Biomedical Imaging and Bioengineering (K01EB027726), National Cancer Institute (R01CA187595), and NVIDIA Academic GPU Grant. J. Kalpathy-Cramer receives research support from General Electric, Bayer, and Genentech.

REFERENCES

[1]. Tromberg BJ et al. , “Assessing the future of diffuse optical imaging technologies for breast cancer management,” Med. Phys, vol. 35, no. 6, pp. 2443–2451, 2008. [PubMed: 18649477]

[2]. Zhu Q and Poplack S, “A review of optical breast imaging: Multi-modality systems for breast cancer diagnosis.,” Eur. J. Radiol, vol. 129, p. 109067, 2020. [PubMed: 32497943]

[3]. Choe R et al. , “Differentiation of benign and malignant breast tumors by in-vivo three-dimensional parallel-plate diffuse optical tomography,” J. Biomed. Opt, vol. 14, no. 2, p. 024020, 2009. [PubMed: 19405750]

[4]. Fang Q et al. , “Combined optical and X-ray tomosynthesis breast imaging.,” Radiology, vol. 258, no. 1, pp. 89–97, 2011. [PubMed: 21062924]

[5]. Mastanduno MA et al. , “MR-Guided Near-Infrared Spectral Tomography Increases Diagnostic Performance of Breast MRI.,” Clin. Cancer Res, vol. 21, no. 17, pp. 3906–12, 2015. [PubMed: 26019171]

[6]. Chae EY et al. , “Development of digital breast tomosynthesis and diffuse optical tomography fusion imaging for breast cancer detection,” Sci. Rep, vol. 10, no. 1, p. 13127, 2020. [PubMed: 32753578]

[7]. Feng J et al. , “Addition of T2-guided optical tomography improves noncontrast breast magnetic resonance imaging diagnosis.,” Breast Cancer Res, vol. 19, no. 1, p. 117, 2017. [PubMed: 29065920]

[8]. Zhu Q et al. , “Assessment of Functional Differences in Malignant and Benign Breast Lesions and Improvement of Diagnostic Accuracy by Using US-guided Diffuse Optical Tomography in Conjunction with Conventional US.,” Radiology, vol. 280, no. 2, pp. 387–97, 2016. [PubMed: 26937708]

[9]. Choe R et al. , “Diffuse optical tomography of breast cancer during neoadjuvant chemotherapy: a case study with comparison to MRI.,” Med. Phys, vol. 32, no. 4, pp. 1128–39, 2005. [PubMed: 15895597]

[10]. Sajjadi AY et al. , “Normalization of compression-induced hemodynamics in patients responding to neoadjuvant chemotherapy monitored by dynamic tomographic optical breast imaging (DTOBI),” Biomed. Opt. Express, vol. 8, no. 2, pp. 555–569, 2017. [PubMed: 28270967]

[11]. Tromberg BJ et al. , “Predicting responses to neoadjuvant chemotherapy in breast cancer: ACRIN 6691 trial of diffuse optical spectroscopic imaging (DOSI).,” Cancer Res, vol. 76, no. 20, pp. 5933–5944, 2016. [PubMed: 27527559]

[12]. Arridge SR and Schotland JC, “Optical tomography: Forward and inverse problems,” Inverse Probl, vol. 25, no. 12, 2009.

[13]. Lehman CD et al. , “National performance benchmarks for modern screening digital mammography: Update from the Breast Cancer Surveillance Consortium,” Radiology, vol. 283, no. 1, pp. 49–58, 2017. [PubMed: 27918707]

[14]. Süzen M, Giannoula A, and Durduran T, “Compressed sensing in diffuse optical tomography.,” Opt. Express, vol. 18, no. 23, pp. 23676–90, 2010. [PubMed: 21164712]

[15]. Prakash J et al. , “Model-resolution-based basis pursuit deconvolution improves diffuse optical tomographic imaging.,” IEEE Trans. Med. Imaging, vol. 33, no. 4, pp. 891–901, 2014. [PubMed: 24710158]

[16]. Kavuri VC et al. , “Sparsity enhanced spatial resolution and depth localization in diffuse optical tomography.,” Biomed. Opt. Express, vol. 3, no. 5, pp. 943–57, 2012. [PubMed: 22567587]

[17]. McCann MT, Jin KH, and Unser M, “Convolutional neural networks for inverse problems in imaging: A review,” IEEE Signal Process. Mag, vol. 34, no. 6, pp. 85–95, 2017.

[18]. Wang G et al. , “Image Reconstruction is a New Frontier of Machine Learning,” IEEE Trans. Med. Imaging, vol. 37, no. 6, pp. 1289–1296, 2018. [PubMed: 29870359]

[19]. Zhu B et al. , “Image reconstruction by domain-transform manifold learning.,” Nature, vol. 555, no. 7697, pp. 487–492, 2018. [PubMed: 29565357]

[20]. Feng J et al. , “Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography,” J. Biomed. Opt, vol. 24, no. 05, p. 1, 2018.

[21]. Ben Yedder H et al., “Deep learning based image reconstruction for diffuse optical tomography,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11074 LNCS, pp. 112–119, 2018.

[22]. Zou Y et al. , “Machine learning model with physical constraints for diffuse optical tomography,” Biomed. Opt. Express, vol. 12, no. 9, pp. 5720–5735, 2021. [PubMed: 34692211]

[23]. Zhang M et al. , “Deep learning-based method to accurately estimate breast tissue optical properties in the presence of the chest wall.,” J. Biomed. Opt, vol. 26, no. 10, p. 106004, 2021. [PubMed: 34672146]

[24]. Feng J et al. , “Deep-learning based image reconstruction for MRI-guided near-infrared spectral tomography,” Optica, vol. 9, no. 3, p. 264, 2022. [PubMed: 35340570]

[25]. Yoo J et al. , “Deep Learning Diffuse Optical Tomography,” IEEE Trans. Med. Imaging, vol. 39, no. 4, pp. 877–887, 2020. [PubMed: 31442973]

[26]. Smith JT et al. , “Fast fit-free analysis of fluorescence lifetime imaging via deep learning.,” Proc. Natl. Acad. Sci. U. S. A, vol. 116, no. 48, pp. 24019–24030, 2019. [PubMed: 31719196]

[27]. Huang C et al. , “Fast and robust reconstruction method for fluorescence molecular tomography based on deep neural network,” in Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XVII, 2019, vol. 10881, pp. 257–262.

[28]. Guo L et al. , “3D deep encoder–decoder network for fluorescence molecular tomography,” Opt. Lett, vol. 44, no. 8, p. 1892, 2019. [PubMed: 30985768]

[29]. Cai C et al. , “End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging,” Opt. Lett, vol. 43, no. 12, p. 2752, 2018. [PubMed: 29905680]

[30]. Fang Q and Boas DA, “Tetrahedral mesh generation from volumetric binary and grayscale images,” in 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Jun. 2009, pp. 1142–1145.

[31]. Zimmermann BB et al. , “Multimodal breast cancer imaging using coregistered dynamic diffuse optical tomography and digital breast tomosynthesis,” J. Biomed. Opt, vol. 22, no. 4, p. 046008, 2017. [PubMed: 28447102]

[32]. Leff DR et al. , “Diffuse optical imaging of the healthy and diseased breast: a systematic review.,” Breast Cancer Res. Treat, vol. 108, no. 1, pp. 9–22, 2008. [PubMed: 17468951]

[33]. Durduran T et al. , “Diffuse Optics for Tissue Monitoring and Tomography.,” Rep. Prog. Phys, vol. 73, no. 7, p. 076701, 2010. [PubMed: 26120204]

[34]. Grosenick D et al. , “Review of optical breast imaging and spectroscopy.,” J. Biomed. Opt, vol. 21, no. 9, p. 091311, 2016. [PubMed: 27403837]

[35]. Tromberg BJ et al. , “Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy.,” Neoplasia, vol. 2, no. 1–2, pp. 26–40, 2000. [PubMed: 10933066]

[36]. Durduran T et al. , “Bulk optical properties of healthy female breast tissue.,” Phys. Med. Biol, vol. 47, no. 16, pp. 2847–61, 2002. [PubMed: 12222850]

[37]. Fang Q et al. , “A multi-modality image reconstruction platform for diffuse optical tomography,” in Biomedical Optics, 2008, p. BMD24.

[38]. Fang Q et al. , “Combined optical imaging and mammography of the healthy breast: optical contrast derived from breast structure and compression.,” IEEE Trans. Med. Imaging, vol. 28, no. 1, pp. 30–42, 2009. [PubMed: 19116186]

[39]. Boas DA et al. , “Imaging the body with diffuse optical tomography,” IEEE Signal Process. Mag, vol. 18, no. 6, pp. 57–75, 2001.

[40]. Srivastava N et al. , “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” J. Mach. Learn. Res, vol. 15, no. 1, pp. 1929–1958, 2014.

[41]. Ronneberger O, Fischer P, and Brox T, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015, 2015, pp. 234–241.

[42]. Çiçek Ö et al. , “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” in Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2016, 2016, pp. 424–432.

[43]. Deng B et al. , “Characterization of structural-prior guided optical tomography using realistic breast models derived from dual-energy x-ray mammography.,” Biomed. Opt. Express, vol. 6, no. 7, pp. 2366–79, 2015. [PubMed: 26203367]

[44]. Fang Q et al. , “Compositional-prior-guided image reconstruction algorithm for multi-modality imaging.,” Biomed. Opt. Express, vol. 1, no. 1, pp. 223–235, 2010. [PubMed: 21258460]

[45]. Althobaiti M, Vavadi H, and Zhu Q, “Diffuse optical tomography reconstruction method using ultrasound images as prior for regularization matrix,” J. Biomed. Opt, vol. 22, no. 2, p. 026002, 2017. [PubMed: 28152129]

[46]. Zhang L et al. , “Direct regularization from co-registered anatomical images for MRI-guided near-infrared spectral tomographic image reconstruction.,” Biomed. Opt. Express, vol. 6, no. 9, pp. 3618–30, 2015. [PubMed: 26417528]

[47]. Isensee F et al. , “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.,” Nat. Methods, vol. 18, no. 2, pp. 203–211, 2021. [PubMed: 33288961]

[48]. Abadi M et al., “TensorFlow: A System for Large-Scale Machine Learning,” in Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, 2016, pp. 265–283.

[49]. Kingma DP and Ba J, “Adam: A Method for Stochastic Optimization,” 2014.

[50]. Hansen PC, “The L-curve and its use in the numerical treatment of inverse problems,” 2000.

[51]. Otsu N, “A Threshold Selection Method from Gray-Level Histograms,” IEEE Trans. Syst. Man. Cybern, vol. 9, no. 1, pp. 62–66, 1979.

[52]. Weatherall PT et al. , “MRI vs. histologic measurement of breast cancer following chemotherapy: comparison with x-ray mammography and palpation.,” J. Magn. Reson. Imaging, vol. 13, no. 6, pp. 868–75, 2001. [PubMed: 11382946]

[53]. Li A et al. , “Assessing the spatial extent of breast tumor intrinsic optical contrast using ultrasound and diffuse optical spectroscopy.,” J. Biomed. Opt, vol. 13, no. 3, p. 030504, 2008. [PubMed: 18601521]

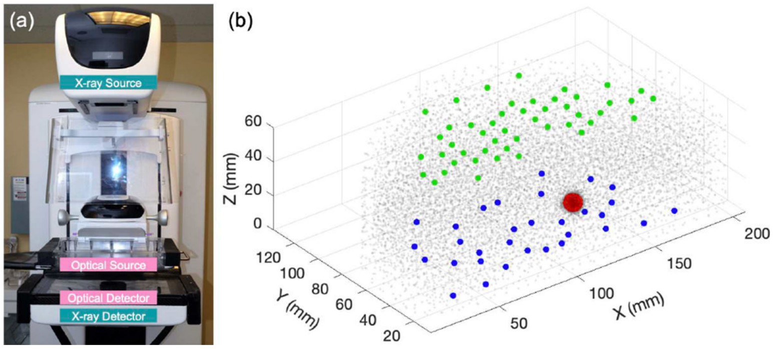

Fig. 1.

(a) Photo of the DOT system integrated with a clinical digital breast tomosynthesis (DBT) device in the transmission geometry, *i.e.,* the DOT source and detector probes arranged on the opposite sides of the compressed breast. (b) A sample phantom geometry with an 8-mm diameter inclusion embedded within a breast-shaped boundary represented by cloud points of the dual-resolution mesh nodes, magenta dots in the fine inclusion region and pale gray dots in the rest of the geometry, respectively. 48 CW sources (green dots) and 32 detectors (blue dots) are plotted as overlays on the top and bottom of the phantom, i*.e.,* superior and inferior in the patient position orientation, respectively.

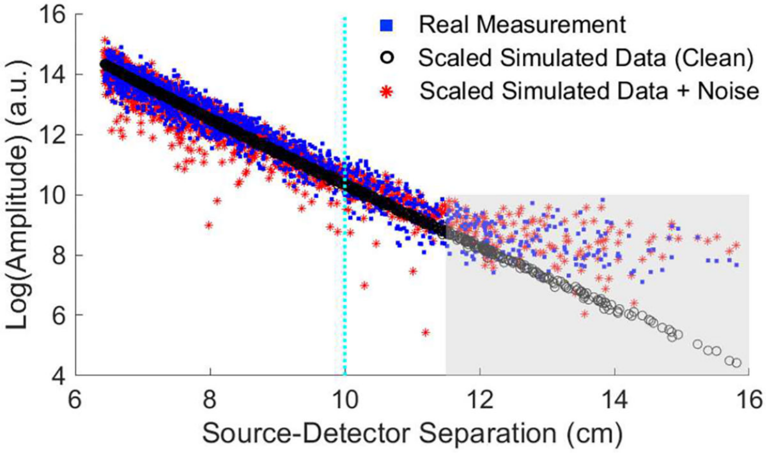

Fig. 2.

Plot of the signal amplitude of all 48×32 source-detector combinations versus the corresponding distances. Black circles: simulated data with only global scaling factor applied; Red stars: scaled simulated data with noise added; Blue squares: averaged real measurements of a homogeneous silicone rubber slab phantom; Gray shaded area: region reaching the noise floor; Cyan dotted line: the source-detector separation limit set for FEM-based methods.

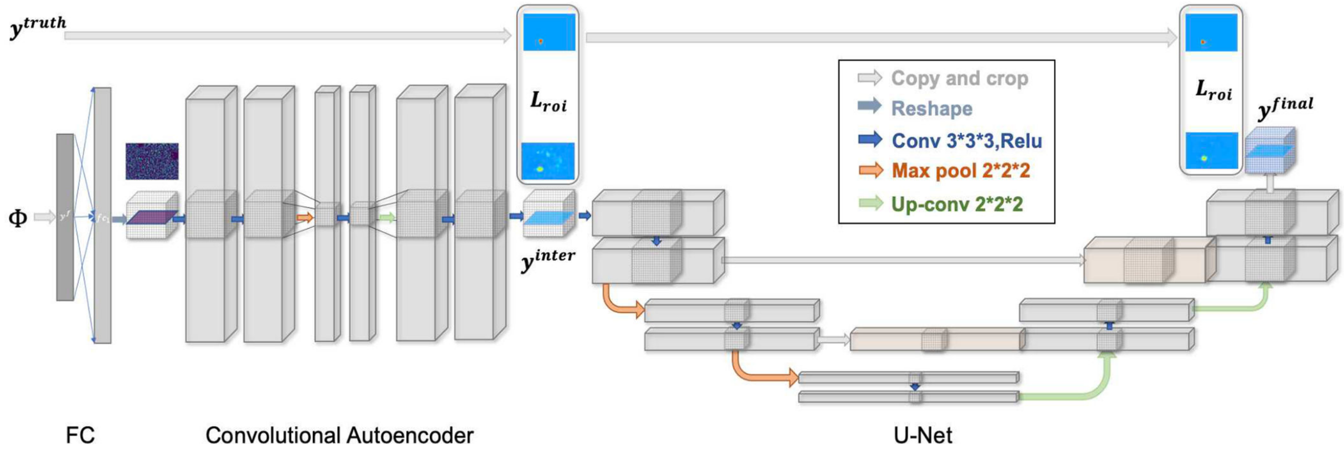

Fig. 3.

Schematic of the proposed deep convolutional neural network, *i.e.,* the FDU-Net. FC: fully connected layers. L_{roi}: prior-weighted loss function; y^{truth}: ground truth 3D optical image volume of μ_a; y^{inter}: intermediate model outputs upon completing the first training step; y^{final}: eventual optical image recovered after the second training step.

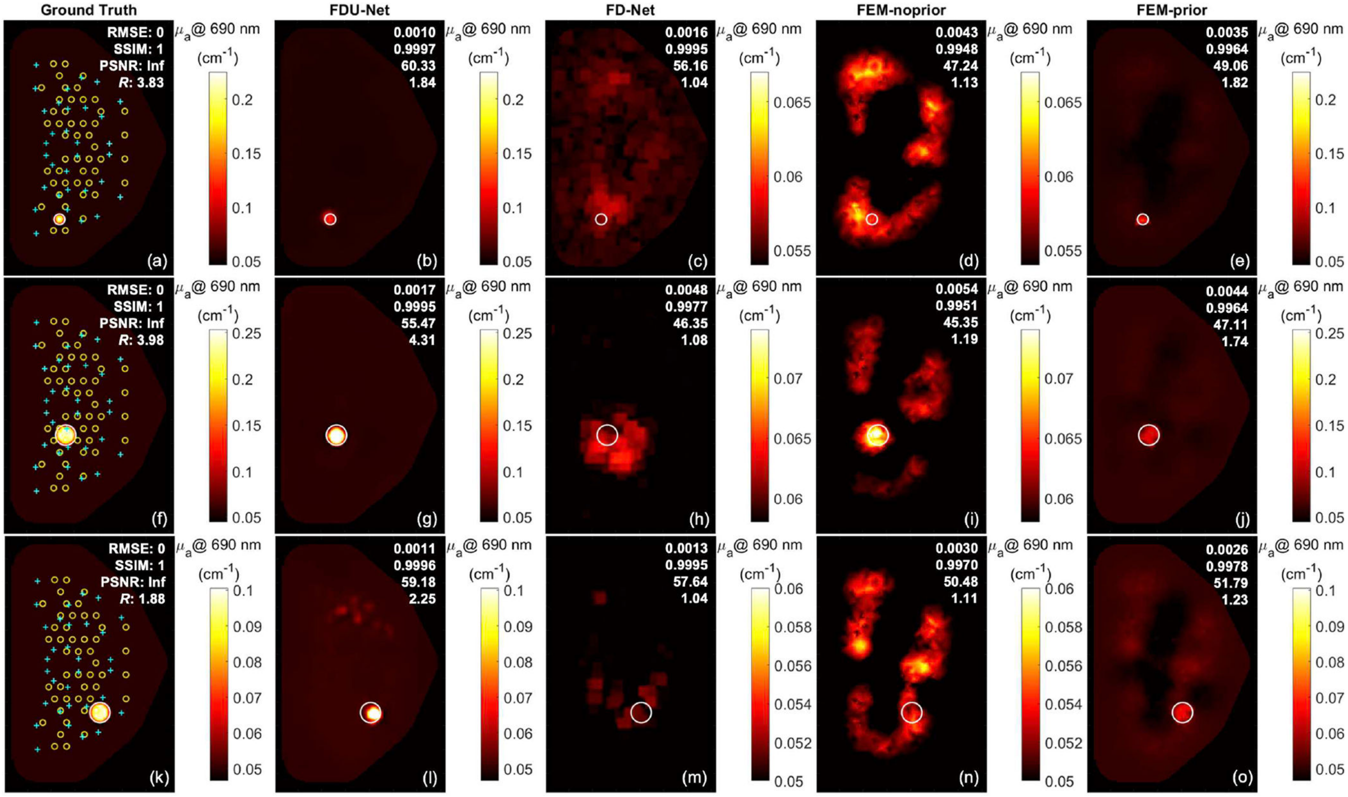

Fig. 4.

Ground-truth (1st column) and recovered absorption coefficients (μ_a) at 690 nm using DL-based (2nd column, proposed FDU-Net, and 3rd column, FD-Net that has similar architecture as previously published in ref [25]) and conventional FEM-based without (4th column, FEM-noprior) and with (5th column, FEM-prior) structural prior guided reconstruction methods for three single spherical inclusion sample cases. Note that only the FEM-prior method used ground-truth inclusion size and location during image reconstruction. Case 1 (top row) has an 8-mm diameter inclusion of ground-truth $R = 3.83$; Case 2 (middle row) has a 16-mm diameter inclusion of $R = 3.98$; and Case 3 (bottom row) has a 16-mm inclusion of $R = 1.88$. RMSE, SSIM, PSNR, and R annotated in the top right corner of each reconstructed image from top to bottom. Yellow circle – source optode; Cyan cross – detector optode.

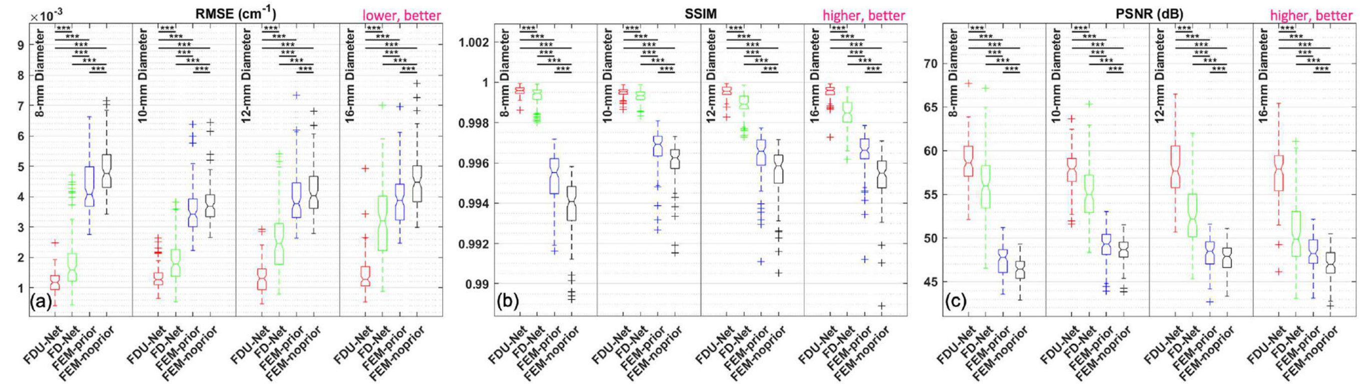

Fig. 5.

Box plots of (a) root mean squared error (RMSE), (b) structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) of 100 test cases in each size group with four reconstruction methods. The bottom and top edges of each box indicate the 25th and 75th percentiles, respectively. Notch line – median. Significance bars of two-sided paired *t*-test between two methods: * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

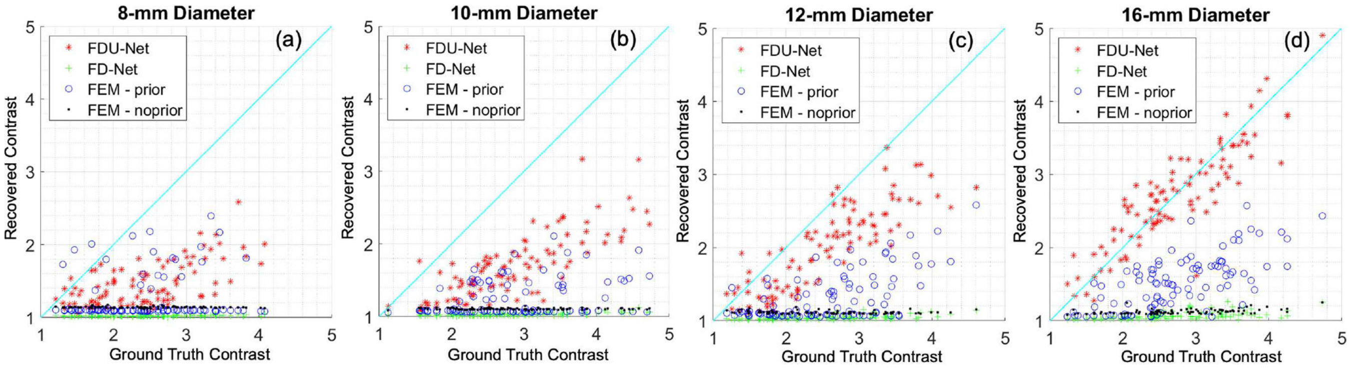

Fig. 6.

Recovered inclusion-to-background ratio R using two DL- and two FEM-based reconstruction methods for 4 different inclusion sizes, 100 test cases in each size group. Cyan solid line – reference for perfect recovery.

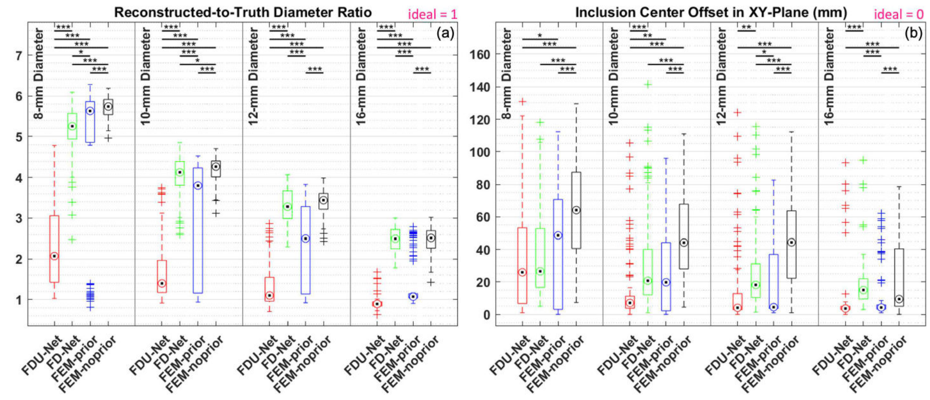

Fig. 7.

Box plot comparisons of inclusion size and localization accuracies in terms of (a) ratios of reconstructed inclusion diameter to ground-truth diameter, and (b) offsets of inclusion center from ground-truth centroid in XY-plane. The bottom and top edges of each box indicate the 25th and 75th percentiles, respectively. Target symbols – median. Significance bars of two-sided paired *t*-test between two methods: * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

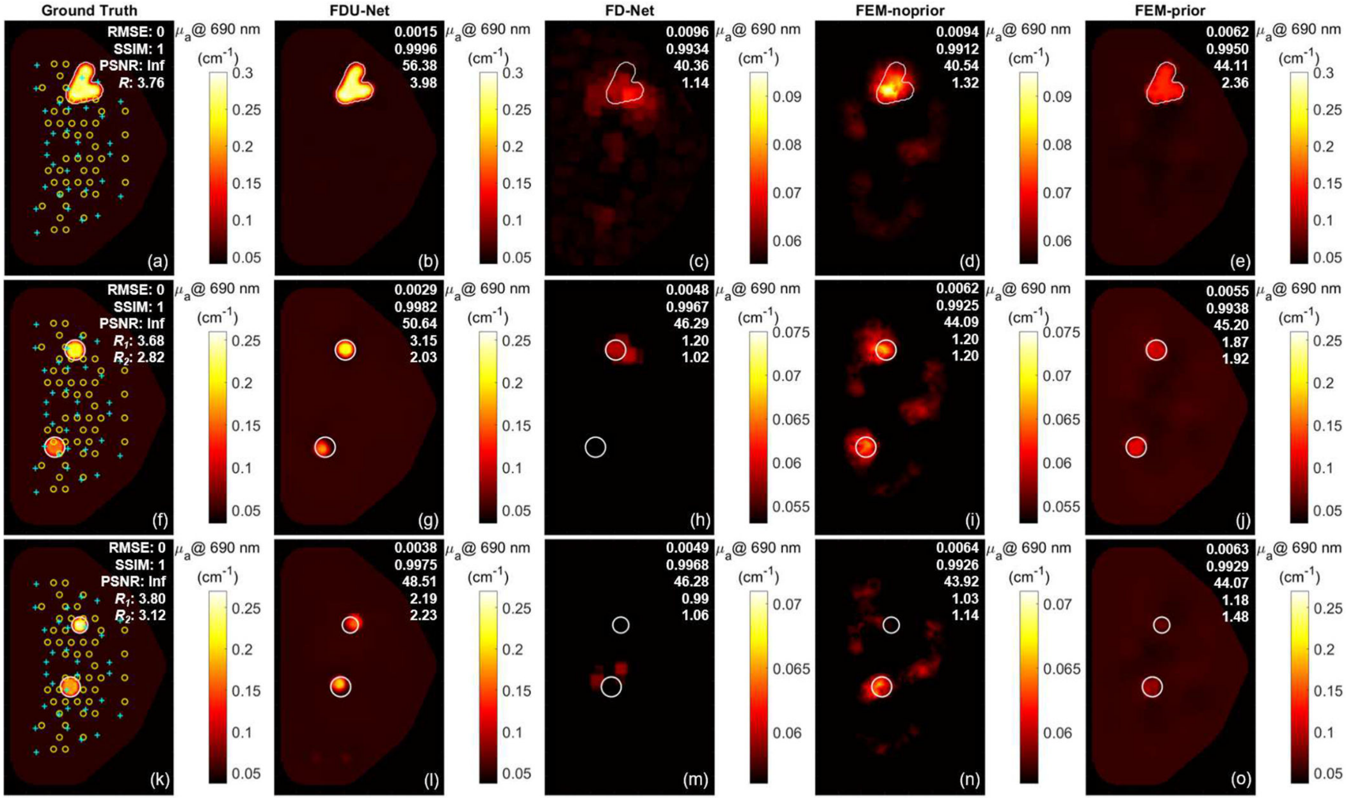

Fig. 8.

Ground-truth (1st column) and recovered absorption coefficients (μ_a) at 690 nm using DL-based (2nd column, proposed FDU-Net, and 3rd column, FD-Net) models trained on only singular spherical inclusion cases and conventional FEM-based without (4th column, FEM-noprior) and with (5th column, FEM-prior) structural prior guided reconstruction methods for three previously unseen cases. Top row – a case with an irregularly shaped inclusion; Middle row – a case with two 16-mm diameter spherical inclusions 83.7 mm apart; Bottom row – a case with one 15.7-mm and one 12.4-mm diameter spherical inclusions 52.6 mm apart. RMSE, SSIM, PSNR, and R annotated in the top right corner of each reconstructed image from top to bottom. R_1 and R_2 – Inclusion-to-background ratios R for the 1st inclusion on top and a 2nd inclusion on the bottom, respectively, in the two-inclusion cases; Yellow circle – source optode; Cyan cross – detector optode.

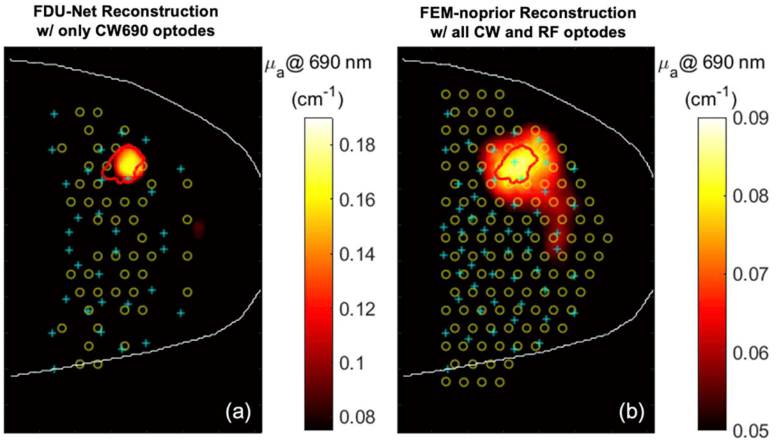

Fig. 9.

Recovered optical images of a patient with a malignant tumor using (left) FDU-Net model trained entirely on simulated data and (right) FEM-based conventional method. White line – breast contour; Red line – tumor marking spatially transformed from ROI drawn by a breast radiologist on separately acquired clinical x-ray mammogram; Yellow circle – source optode; Cyan cross – detector optode. Note that more DOT measurement data were used in the FEM-based reconstruction.

Table 1

Parameters of Linear Regression ($y = ax + b$) between Ground-Truth and Recovered R using Various Reconstruction Methods

Inclusion Diameter	**FDU-Net**			**FD-Net**			**FEM-prior**			**FEM-noprior**		
	a	b	R^2	a	b	R^2	a	b	R^2	a	b	R^2
8 mm	0.26	0.82	0.38	0.005	1.01	0.21	0.06	1.10	0.02	−0.002	1.15	0.02
10 mm	0.42	0.50	0.63	0.017	0.98	0.47	0.09	0.99	0.09	−0.001	1.11	0.01
12 mm	0.63	0.40	0.73	0.021	0.99	0.30	0.26	0.67	0.36	−0.004	1.12	0.03
16 mm	0.80	0.53	0.78	0.039	0.97	0.30	0.34	0.61	0.47	0.021	1.07	0.25