TITLE PAGE

Citation Format:

Aapo Peräkorpi, Matteo Bregonzioa, Gianluca Valentini, Giovanna Tramontin, and Andrea Farina, "Direct application of deep learning for diffuse optical tomography", Proc. of SPIE Vol. 13314, Optical Tomography and Spectroscopy of Tissue XVI, edited by Sergio Fantini, Paola Taroni, 1331405 (February 2025)

Copyright notice:

DOI abstract link:

https://doi: 10.1117/12.3042940

Direct application of deep learning for diffuse optical tomography

Aapo Peräkorpi[a, b], Matteo Bregonzio[a], Gianluca Valentini[b, c], Giovanna Tramontin[c, b], and Andrea Farina[c]

[a]Datrix SpA, Foro Buonaparte 71, 20121 Milan, Italy
[b]Politecnico di Milano, Dipartimento di Fisica, Piazza L. da Vinci 32, 20133 Milan, Italy
[c]Consiglio Nazionale delle Ricerche, IFN, Piazza L. da Vinci 32, 20133 Milan, Italy

ABSTRACT

Diffuse optical tomography (DOT) is a non-invasive imaging technique that uses near-infrared light to visualise tissue properties by reconstructing optical characteristics in scattering media. Time-domain DOT (TD-DOT) enhances this process by measuring photon time-of-flight, providing depth-resolved data but at a high computational cost. This study investigates the potential of machine learning (ML) for accelerating TD-DOT image reconstruction using simulated data. A neural network with an autoencoder architecture and a 3D U-Net refinement stage was trained on simulated datasets. The model achieved 96.4% precision in detecting the inclusion positions and reconstructed 3D optical properties in 0.25 seconds, with root mean squared errors of 20.1% for voxelwise absorption and 17.2% for scattering coefficients. These results demonstrate ML's feasibility for fast, accurate TD-DOT reconstruction, highlighting future prospects for real-world validation and improved data compression techniques.

Keywords: Diffuse optical tomography, Hadamard, Structured illumination, Machine learning, Monte Carlo, Image reconstruction

1. INTRODUCTION

Diffuse optical tomography (DOT) is a non-invasive imaging technique that uses near-infrared light (630-950 nm) to characterize biological tissues by reconstructing the internal optical properties of scattering media. By analysing how light diffuses through tissue, DOT enables the visualisation of tissue composition, oxygenation, and blood flow, making it valuable in a wide range of medical applications, including functional brain imaging and breast cancer detection.[1,2] Time-domain DOT (TD-DOT) specifically measures the temporal distribution of photons as they propagate through the medium, providing depth-resolved information. By using time-of-flight data and spatial resolution the contrast can be improved, offering a more detailed picture of the tissue's optical characteristics. Compared to systems using continuous wave or intensity-modulated light, time-domain data contains more information, although its modelling and image reconstruction are computationally and hardware wise demanding.[3,4]

One way to address the computational challenges of TD-DOT is through single-pixel imaging (SPI), a computational imaging method that uses a single detector to reconstruct high-resolution images . Instead of capturing a 2D image directly, the sample is illuminated with a sequence of spatial light patterns, and the corresponding reflected or transmitted light intensity is measured. This approach allows image reconstruction using compressive sensing techniques, reducing the amount of data required across various spectral ranges.[5] Hadamard patterns are particularly advantageous in SPI due to their orthogonality and binary nature, enabling efficient encoding and decoding of spatial information. Applying these patterns in TD-DOT enhances measurement sensitivity and reduces noise by maximizing light collection efficiency.[6]

Further author information: (Send correspondence to Aapo Peräkorpi)
Aapo Peräkorpi: E-mail: aapo.perakorpi@datrixgroup.com

Optical Tomography and Spectroscopy of Tissue XVI, edited by Sergio Fantini, Paola Taroni, Proc. of SPIE Vol. 13314, 1331405 · © 2025 SPIE
1605-7422 · doi: 10.1117/12.3042940

Structured illumination, especially in transmittance geometries, complements SPI by spatially modulating light projected onto the tissue surface. In TD-DOT, Fourier or Hadamard patterns are preferred for structured illumination due to their mathematical properties allowing for data compression, facilitating robust image reconstruction through computational inversion techniques.[6,7]

While the data compression in TD-DOT makes the image reconstruction computationally faster it is still orders of magnitudes slower than the advances in machine learning (ML) could allow. In particular neural networks and deep learning algorithms can be trained to perform fast, accurate image reconstructions by learning complex mappings between measured data and tissue properties, bypassing the need for time-consuming iterative inverse problem algorithms. Most work in the ML field related to DOT has focused on improving iterative reconstruction approaches by optimising individual steps in the process.[8,9] However, recent studies have explored the use of ML for direct reconstruction.[10,11]

This study investigates the use of Machine Learning for reconstructing time-domain diffuse optical tomography (TD-DOT) with simulated data. Monte Carlo (MC) simulations were used to model the radiative transport equation (RTE), simulating photon propagation through diffuse media. MC simulations are employed due to their capacity to model photon propagation in highly scattering media by tracking individual photon trajectories, accounting for both scattering and absorption events. Unlike approximate methods, MC simulations offer precise solutions across diverse geometries and optical property configurations. However, they suffer from high computational costs and inherent noise due to statistical variability. Recent advancements in GPU computing have dramatically accelerated MC simulations, enabling their application in both forward and inverse problem-solving frameworks.[12,13]

2. METHODS

2.1 Data generation

The open-source Monte Carlo Extreme (MCX) software,[14] specifically its mesh-based implementation integrated with MATLAB (MMC),[15] was utilized to conduct simulations. The MMC code facilitates comprehensive simulations of photon transport for various sample geometries and illumination-detection configurations including structured illumination. It supports flexible mesh generation, allowing specification of absorption, scattering coefficients, anisotropy factors, and refractive indices for different sample regions.

Key simulation outputs include photon entry/exit coordinates and photon partial path lengths within each medium. These metrics provide an in-depth understanding of light transport dynamics in turbid media. The simulated detected local intensity was calculated as shown in Equation 1.

$$I(t) = \sum_p w_p(t) = \sum_p \exp\left(-\sum_m l_{p,m} \cdot \mu_{a,m}\right), \tag{1}$$

where $I(t)$ is the local intensity at time t, $w_p(t)$ is the weight of the p-th detected photon at time t, $l_{p,m}$ is the partial path length of a photon through the region m and $\mu_{a,m}$ is the absorption coefficient of the m-th region. Finally $t_{f,p}$ is the total time-of-flight of a detected photon.

Finally the simulated light intensity was convolved with realistic impulse response function taken from an experimental setup. This method of calculating the $I(t)$ allows the bining of the light intensity not only in the spatial surface of the detector but in time bins as well.[16]

To train a machine learning model, a diverse dataset was generated using the following workflow:

1. Mesh Generation: A variety of 3D geometries (spheres, cubes, pyramids, cylinders) with different sizes, orientations, and optical properties were created using iso2mesh library.[17] The dimensions of the background slab were 65×65×15 mm^3, where the 32×32 mm^2 region in the center represented the area of illumination and detection. The slab's thickness, corresponding to the 15 mm distance between the source and detector planes, ensured photons propagated through a controlled medium. The inclusion sizes ranged from approximately 3 mm to 10 mm in diameter, with orientations and positions randomized but constrained to lie between the source and detector planes. The background slab had optical properties spanning reduced scattering coefficients of 0.5-1.5

mm^{-1} and absorption coefficients of 0.005-0.015 mm^{-1}. The optical properties of the inclusions were randomized to be between 1.2 to 2.0 times that of the background absorption and the scattering. This process yielded 6000 optically unique geometries.

2. Photon Simulation: Each geometry was simulated by running MMC on an NVIDIA GeForce RTX 3050 GPU, with 10^8 simulated photons per run. The photons were weighted based on their flight paths and binned into source, detection, and time bins for analysis.

For structured illumination, the 32×32 mm^2 source illumination area was divided into 1024 bins (32×32 grid), each corresponding to Hadamard basis pattern. These patterns were created by constructing Hadamard matrices arranged in dyadic (Paley) order,[18] then rearranged using a zigzag algorithm as described by.[19] This ordering was chosen in order to avoid possible x vs. y bias arising from the fact that typical orderings have striped patterns at the start and padding operation in the 1D convolutions in the neural network could make these more pronounced. The detection area, also 32×32 mm^2, employed a corresponding 32×32 Hadamard grid. Time bins were selected by analysing the photon detection histogram (Figure 1) and were manually set to: 0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1.0, 1.5, 2.0, and 5.0 ns.

3. Data Preprocessing: Hadamard patterns were constructed by aggregating photon counts from corresponding source bins. These patterns served as inputs for the neural network. The reconstruction volume covered the center 32×32 mm^2 region where the patterns were projected. This (32x32x15 mm^3) region was discretized into a 64×64×30 voxel grid, encoding optical properties and inclusion locations as shown in Figure 2. To reduce noise from limited photon counts, a Gaussian blur was applied to the simulated intensity measurements.

4. Data Storage: Preprocessed data was saved as .mat files, converted into .zarr format, and transferred to a GPU server for deep learning training.

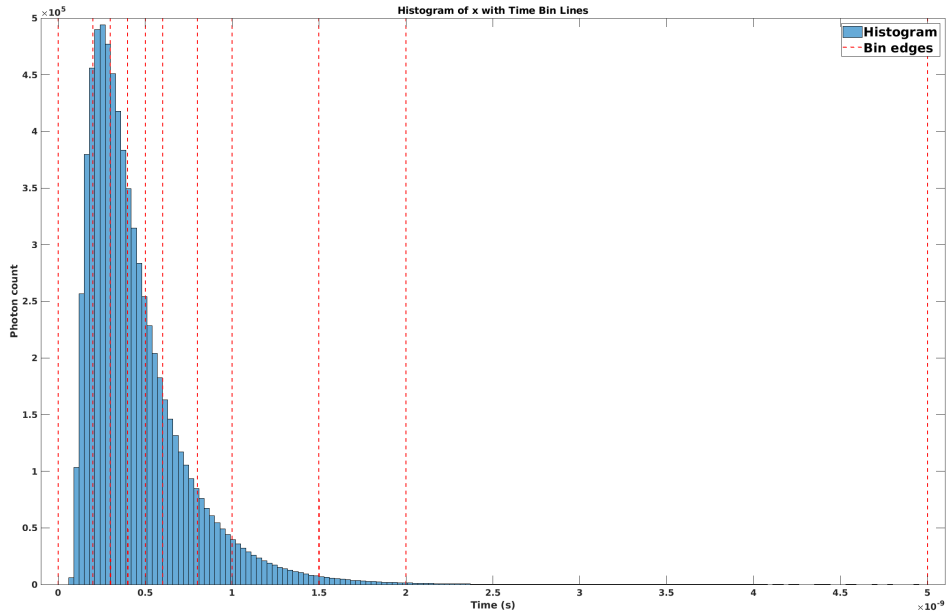

Figure 1. Histogram of the photon time-of-flights from source, through the simulated phantom, to the detector. Red lines represent selected edges of time bins.

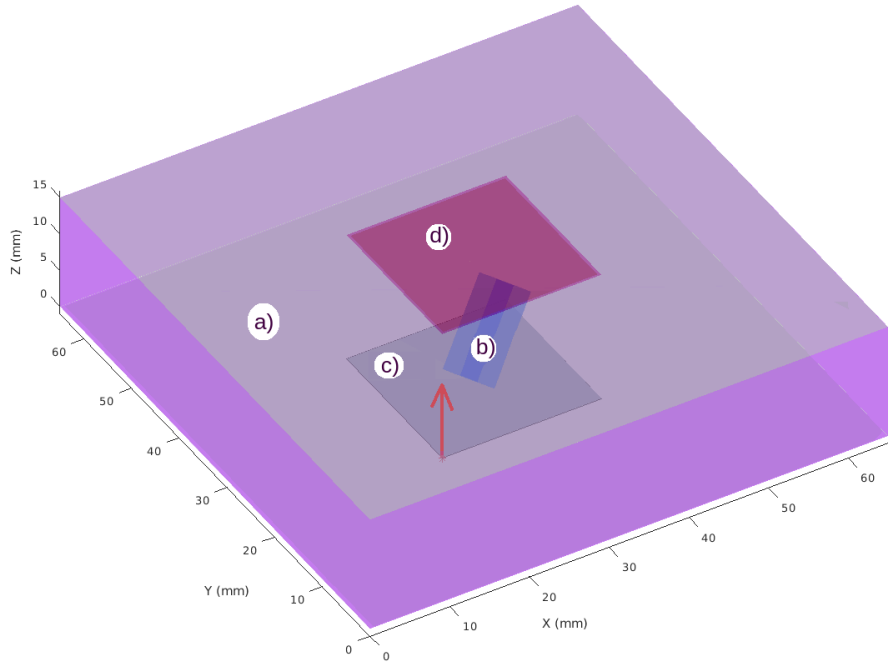

Figure 2. Illustration of the simulated measurement geometry: a) Phantom background composed of an optically homogeneous material; b) a small inclusion embedded within the background exhibiting differing optical properties; c) structured illumination source projected onto a specified plane (direction indicated by the red arrow); and d) detector positioned above the phantom to capture transmitted diffuse light.

2.2 Machine learning

The neural network aimed to reconstruct a 64×64×30 voxel structure with three channels: absorption coefficient, scattering coefficient, and a binary inclusion mask. Inputs consisted of 10,240 images (size 32×32) representing light intensity distributions, derived from 10 time steps and 1,024 Hadamard basis illumination patterns. The light intensities were normalized to fall within the range [0, 1] to ensure numerical stability during training.

The model was built using PyTorch and featured approximately 200 million parameters. The architecture was inspired by,[10] but omitted the fully connected layer to avoid the computational overhead associated with the large dimensionality of the input data. Instead, the architecture included a more complex encoder section, as shown in Figure 3.

The encoder was designed to process the varied inputs efficiently and utilized 2D convolutions for processing 32×32 spatial images, 1D convolutions for encoding the Hadamard patterns, and a gated recurrent unit (GRU) for temporal analysis across the 10 time steps. GRU was chosen over LSTM due to its computational efficiency, which was critical given the overall size of the model.

The decoder employed 3D convolutions to reconstruct the voxelized structure from the features extracted by the encoder. Activation functions included leaky ReLU with a negative slope of 0.01, which provided non-linear modelling capability while mitigating vanishing gradient issues. Batch normalization and max-pooling were used to improve training stability and reduce overfitting, while a dropout rate of 0.35 added further regularization.

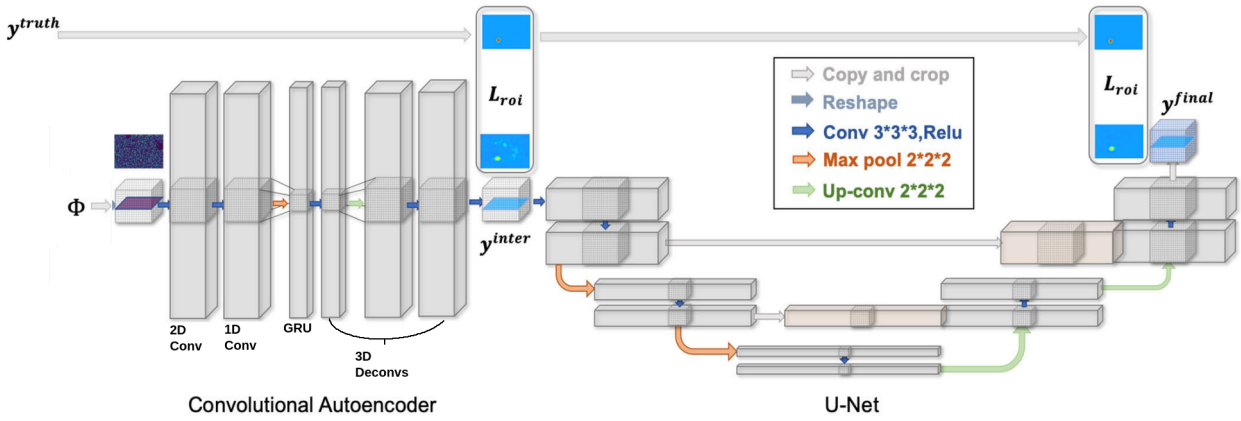

Figure 3. The schematic depicts the neural network architecture that integrates an autoencoder with a U-net. The prior-weighted loss function, $L : roi$, guides the optimisation process. Key elements include y^{truth}, the ground truth 3D optical image volume representing μ_a, μ_s, and binary mask for the inclusion, y^{inter} the output of the autoencoder model after its training and y^{final} the final optical image reconstructed following U-net training.
ϕ denotes the 4D dataset of light intensity measurements, which serves as the model's input. Within the encoder section of the autoencoder, the x-y spatial dimensions are initially processed through 2D convolutional layers. Subsequently, the illumination pattern dimension is analysed using 1D convolutional layers, and the temporal dimension is processed through a gated recurrent unit (GRU). The combined features are then passed into a 3D deconvolutional section, reconstructing the 3D structure. This 3D representation is further refined using a 3D U-net, giving the final optical image reconstruction. The image is edited version adapted from paper [10] .

To enhance the predictions, the autoencoder's output was passed through a second-stage network, a more traditional 3D Unet. This U-net focused on correcting errors and artifacts from the autoencoder. The first model excelled at predicting the binary inclusion mask but struggled to provide clear separation in the absorption and scattering coefficients between the background and the inclusion. The U-net was able to refine these predictions by combining the sharp binary mask with the continuous optical property values predicted by the autoencoder.

Training was performed on an NVIDIA L40S GPU using the AdamW optimiser with a learning rate of 0.0001 and weight decay of 0.005 for the autoencoder. For the U-net, a higher learning rate of 0.01 and weight decay of 0.001 were used. A learning rate scheduler, was applied to dynamically adjust the learning rate based on validation performance, with a patience of 20 epochs and a reduction factor of 0.1.

The loss function was a combination of weighted binary cross-entropy (BCE), weighted mean squared error (MSE), and total variation loss. BCE and MSE losses were scaled by a factor of 100 to address observed numerical instability. The weights for BCE and MSE were dynamically calculated for each sample based on the background-to-foreground voxel ratio in the binary masks. For example, the dynamic weighting ensured that contributions from the foreground inclusion typically significantly smaller in volume than the background were sufficiently emphasized during training. Total variation loss remained static and was used to enforce smoothness in the predicted optical properties.

Gradient issues were a challenge for the larger autoencoder model, necessitating a lower learning rate to stabilize training. The model was trained for 100 epochs with a 1:1.8:7.2 test-validation-train split ratio. This setup provided sufficient data for robust learning while maintaining adequate evaluation datasets for testing and validation.

No additional data augmentation techniques were employed during training, as the dataset generated through the simulation process was deemed sufficiently extensive. However, the dataset's preparation steps, including Gaussian blurring and normalization of all input channels to the range [0, 1], contributed to improved model performance.

Overall, the architecture and training strategy achieved strong reconstruction performance, effectively leveraging the structured illumination data to predict absorption, scattering, and inclusion masks within a voxelized domain.

3. RESULTS

The precision of the model stands around 90% e.g. True positives / (False positives + False negatives) is around 0.9 across the test dataset, as illustrated in Figure 4. The binary mask predicting the inclusion location is more accurate and has less noise as well as artifacting than the other two channels (absorption and scattering coefficients) which predict continuous real values instead of binary ones. The model is significantly faster compared to traditional methods for inverse problems, which can take several minutes.[4] In contrast, it reconstructs the 3D distribution of variables in just 0.25 seconds.

While the contrast is good between foreground and background, voxel values are less accurately predicted. The relative error in absorption and scattering prediction across the test dataset stands around 15%, the peak signal to noise ratio was around 45 for absorption and 70 for scattering. Structural similarity index was on average 3 for absorption and 1.3 for scattering. These results are presented in Figures 5 and 6.

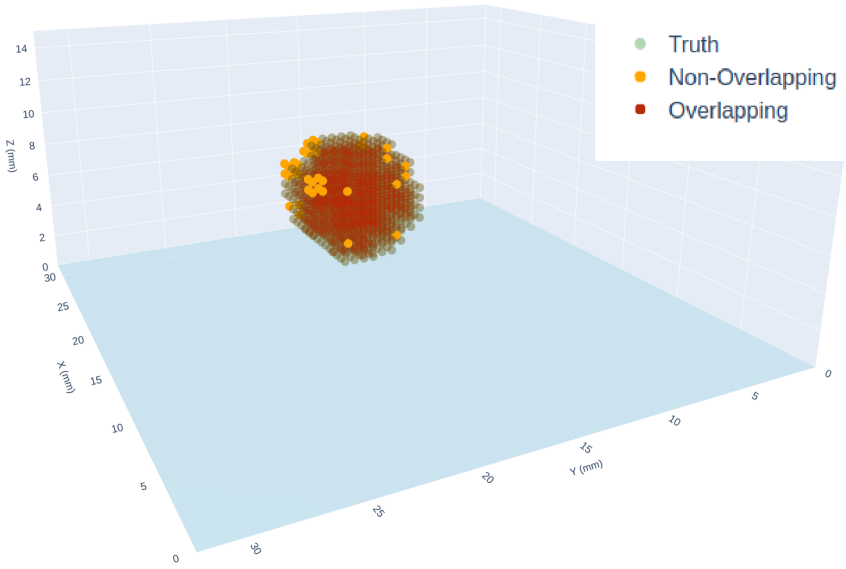

Figure 4. 3D representation of predicted binary mask. The true location of inclusion is indicated by green markers. Regions where the prediction of model overlaps with true location are shown in red, while areas without overlap are displayed in orange.

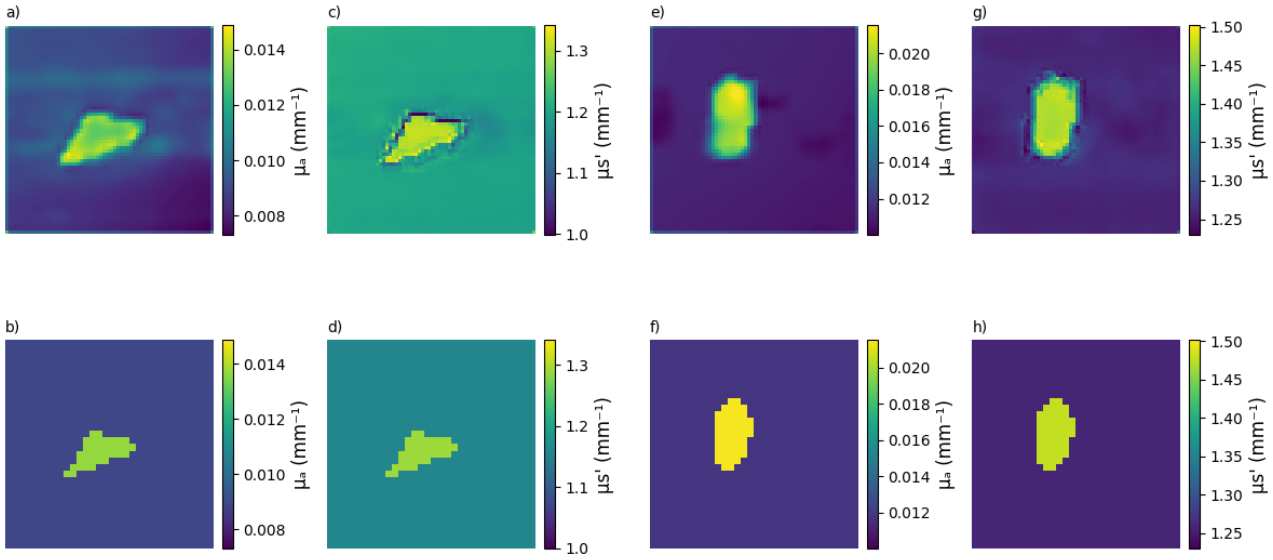

Figure 5. 2D cutoff images of optical properties with two inclusion objects pyramid (columns 1 and 2) and cylinder (columns 3 and 4). Top row: model prediction, bottom row: ground truth. (a, b) Absorption for the background and pyramid inclusion, (c, d) scattering for the background and pyramid inclusion, (e, f) Absorption for the background and cylinder inclusion and (g, h) scattering for the background and cylinder inclusion.

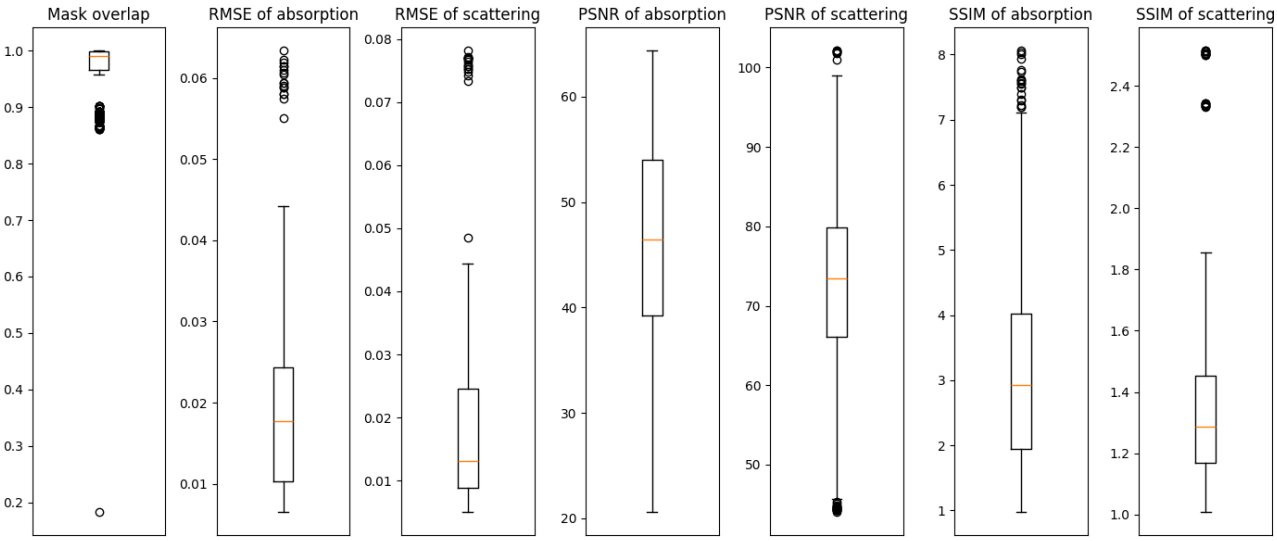

Figure 6. Box plots of scattering and absorption metrics (left to right): Mask overlap (higher = better), root mean squared error (RMSE) of absorption (lower = better), RMSE of scattering, peak signal-to-noise ratio (PSNR) of absorption (higher = better), PSNR of scattering, structural similarity index (SSIM) of absorption (higher = better) and SSIM of scattering.

4. CONCLUSION

The objective of this study was to evaluate the feasibility and advantages of neural network-based image reconstruction compared to traditional methods, particularly regarding computational speed and accuracy. Similar results to those reported in previous studies[10,11] were achieved, although the input here was more complex, incorporating time-domain and structured illumination dimensions. From the presented results it can be seen that it is possible to rapidly reconstruct time-domain diffuse optical tomography (TD-DOT) images accurately. This will allow for rapid development and examination of data compression techniques in TD-DOT.

Some caveats should be highlighted. The study was carried out *in-silico* and future studies should be conducted to validate the results shown here with real physical phantoms and *in-vivo*. Potential improvements to the reconstruction method could be made by reducing the number of Hadamard basis patterns. This should also be investigated in future studies as it would significantly reduce the computational requirements and the measurement time.

This work demonstrates an accurate and rapid reconstruction technique for TD-DOT with structured illumination *in-silico* using machine learning. With this, it appears to be possible to avoid computationally costly solutions to the inverse problem with ML.

Acknowledgements

Funded by the European Union (GA 101072354). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15, 1999.

[2] A. P. Gibson, J. C. Hebden, and S. R. Arridge. Recent advances in diffuse optical imaging. *Physics in Medicine & Biology*, 50(4), 2005.

[3] M. Schweiger and S. R. Arridge. Application of temporal filters to time resolved data in optical tomography. *Physics in Medicine & Biology*, 44(7):1699–1717, 1999.

[4] M. Mozumder and T. Tarvainen. Time-domain diffuse optical tomography utilizing truncated fourier series approximation. *Journal of the Optical Society of America A*, 37:182–191, 2020.

[5] M. F. Duarte, M. A. Davenport, D. Takhar, and et al. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.

[6] A. Farina, M. Betcke, C. D'Andrea, and et al. Multiple-view diffuse optical tomography system based on time-domain compressive measurements. *Optics Letters*, 42(14):2822, 2017.

[7] M. Mireles, E. Xu, R. Ragunathan, and Q. Fang. Medium-adaptive compressive diffuse optical tomography. *Biomedical Optics Express*, 15:5128–5142, 2024.

[8] J. Yoo, S. Sabir, D. Heo, K. Kim, and et al. Deep learning diffuse optical tomography. *IEEE Transactions on Medical Imaging*, 39(4):877–887, 2020.

[9] M. Mozumder, A. Hauptmann, and T. Tarvainen. A model-based iterative learning approach for diffuse optical tomography. *IEEE Transactions on Medical Imaging*, 41(5):1289–1299, 2022.

[10] B. Deng, H. Gu, H. Zhu, K. Chang, and et al. Fdu-net: Deep learning-based three-dimensional diffuse optical image reconstruction. *IEEE Transactions on Medical Imaging*, 42(8):2439–2450, 2023.

[11] R. Dale, B. Zheng, F. Orihuela, and et al. Deep learning-enabled high-speed, multi-parameter diffuse optical tomography. *Journal of Biomedical Optics*, 29(7):076004, 2024.

[12] L. Yu, F. Nina-Paravecino, D. Kaeli, and Q. Fang. Scalable and massively parallel monte carlo photon transport simulations for heterogeneous computing platforms. *Journal of Biomedical Optics*, 23(1):010504, 2018.

[13] M. Ardakani, L. Yu, D. Kaeli, and Q. Fang. Framework for denoising monte carlo photon transport simulations using deep learning. *Journal of Biomedical Optics*, 27(8):083019, 2022.

[14] Q. Fang and S. Yan. Graphics processing unit-accelerated mesh-based monte carlo photon transport simulations. *Journal of Biomedical Optics*, 24(11):115002, 2019.

[15] MathWorks. Matlab, 2023.

[16] S. Yan, R. Yao, X. Intes, and Q. Fang. Accelerating monte carlo modeling of structured-light based diffuse optical imaging via photon sharing. *Optics Letters*, 45(10):2842–2845, 2020.

[17] Qianqian Fang and David Boas. Tetrahedral mesh generation from volumetric binary and gray-scale images. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pages 1142–1145, 2009.

[18] R. Paley. On orthogonal matrices. *Journal of Mathematics and Physics*, 1933.

[19] S. Huang, B. Chen, B. Chen, and C. Chan. Enhancing compressive single-pixel imaging with zig-zag-ordered walsh-hadamard light modulation. *IEEE Photonics Technology Letters*, 36(12):803–806, 2024.