Deep Learning Methods for Reconstruction and Analysis of Diffuse Optical Tomography Images of Breast Cancer Lesions

by

Hanene Ben Yedder

M.Sc., University of Quebec in Outaouais, 2014
B.Sc., High School of Telecommunication Sup'Com, 2006

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Fall 2023

Declaration of Committee

Name:	**Hanene Ben Yedder**
Degree:	**Doctor of Philosophy**
Thesis title:	**Deep Learning Methods for Reconstruction and Analysis of Diffuse Optical Tomography Images of Breast Cancer Lesions**
Committee:	**Chair:** Saeid Asgari Adjunct Professor, Computing Science
	Ghassan Hamarneh Supervisor Professor, Computing Science
	Yasutaka Furukawa Committee Member Associate Professor, Computing Science
	Majid Shokoufi Examiner Lecturer, Engineering Science
	Jong Chul Ye External Examiner Professor, Kim Jaechul Graduate School of AI Korea Advanced Institute of Science and Technology

Abstract

The development of an accurate, efficient, portable, and affordable method for identifying breast cancer is critical for both early detection and improved prognosis. Medical imaging modalities play a critical role in cancer screening and treatment monitoring. Diffuse optical tomography (DOT) is a non-invasive imaging modality that can be used in a low-complexity probe design, resulting in an inexpensive portable imaging diagnostic device with low power consumption. In recent years, machine learning techniques have created transformative opportunities for medical image reconstruction and analysis, helping move toward data-driven algorithm designs wherein computational power is augmented with physics priors to push the accuracy and fairness of image driven diagnosis to new limits.

In this thesis, we present multiple deep learning-based medical image reconstruction and analysis approaches for screening breast cancer lesions acquired by DOT. First, an end-to-end image reconstruction model from sensor-domain data is proposed, where physics-based simulation is leveraged to address the lack of available real-world data required for training. Next, we adopt a transfer learning strategy to align and translate the sensor domain distribution between in silico and real-world data and propose a novel loss to promote appearance similarity and penalize artifacts. Following up on this we propose a joint reconstruction and localization solution that simultaneously attends to the most important features while ensuring better lesion localization. Finally, we propose an orthogonal multi-frequency fusion solution for direct prediction of the end task from sensor signal data, increasing diagnosis accuracy at a reduced computational cost.

Extending a portable device with such diagnosis ability promises to improve first-line treatment throughput. These contributions demonstrate the promising role of deep learning in DOT image reconstruction and diagnosis. Combined, our contributions open the path towards personalized medicine for non-invasive portable diagnosis and treatment monitoring of breast cancer in the very near future.

Keywords: Medical imaging; Diffuse optical tomography; Image reconstruction; Deep learning; Object localization and diagnosis; Breast cancer

Dedication

To my respectful parents and beloved family without whose support this thesis was not possible.

Acknowledgements

In profound gratitude, I acknowledge Allah's boundless mercy and blessings, which guide and illuminate my live every day.

First and foremost, I would like to express my heartfelt gratitude to my senior supervisor, Prof. Ghassan Hamarneh, whose guidance, continuous support, and patience throughout my Ph.D. journey have been invaluable. His vast knowledge and extensive experience have been a constant source of motivation for me in both my academic pursuits and daily life. Additionally, I extend my thanks to my examination committee, Prof. Jong Chul Ye, Dr. Majid Shokoufi and Prof. Yasutaka Furukawa, and the defence chair Prof. Saeid Asgari for their time and feedback on this Thesis.

I would like to thank my friends, lab mates, colleagues, and collaborators at the Medical Image Analysis Lab for the wonderful and cherished moments we shared both in the lab and during social setting. Your camaraderie and support have made my time there truly memorable. A special thanks to Ben Cardoen for our brainstorming sessions and his insightful feedback.

I am also deeply grateful to my beloved daughters, husband, family and friends for their encouragement and support throughout my academic journey. Your belief in me has been a constant source of motivation, and I am thankful for having you by my side.

Table of Contents

Declaration of Committee ii

Abstract iii

Dedication iv

Acknowledgements v

Table of Contents vi

List of Tables x

List of Figures xi

List of Acronyms xiv

List of Notations xv

1 Introduction 1

1.1 Background and Motivation 1

1.2 Thesis Contributions 2

1.2.1 Deep Learning for Biomedical Image Reconstruction 3

1.2.2 End-to-End Deep Learning based DOT Image Reconstruction 3

1.2.3 Tackling DL Image Reconstruction Artifacts in Limited-Angle DOT 4

1.2.4 Multitask Deep Learning for Joint Reconstruction and Localization . 5

1.2.5 End-to-End Diagnosis via Deep Orthogonal Multi-Frequency Fusion 6

1.3 Summary of Contributions 7

1.4 Research Contributions Excluded from The Thesis 8

2 Deep Learning for Biomedical Image Reconstruction: A Survey 9

2.1 Introduction 9

2.1.1 Inverse Problem and Challenges 10

2.1.2 Scope of this Survey 13

2.2 Conventional Image Reconstruction Approaches 14

2.2.1 Analytical Methods . . . 15
2.2.2 Iterative Methods . . . 16
2.3 Deep Learning Based Image Reconstruction . . . 19
2.3.1 Deep Learning as Processing Step: Two Step Image Reconstruction Models . . . 21
2.3.2 End-to-End Image Reconstruction: Direct Estimation . . . 23
2.3.3 Raw-to-Task Methods: Task Specific Learning . . . 29
2.4 Medical Training Datasets . . . 30
2.4.1 Real World Datasets . . . 30
2.4.2 Physics Based Simulation . . . 31
2.4.3 Challenge Datasets . . . 31
2.5 Reconstruction Evaluation Metrics . . . 32
2.5.1 Quality . . . 32
2.5.2 Inference Speed . . . 33
2.6 Discussion . . . 34
2.7 Conclusion . . . 38

3 End-to-End Deep Learning Based DOT Image Reconstruction 41
3.1 Introduction . . . 41
3.1.1 Background on Diffuse Optical Tomography Imaging . . . 41
3.1.2 Current DOT Image Reconstruction and Challenges . . . 44
3.1.3 Deep Learning based Methods . . . 45
3.1.4 Contributions . . . 46
3.2 Methodology . . . 47
3.2.1 Physics based Training Data for DOT Reconstruction . . . 47
3.2.2 Phantom Test Dataset . . . 49
3.2.3 Reconstructing Images from DOT Measurements . . . 49
3.2.4 Mathematical Model Formulation . . . 50
3.3 Implementation . . . 51
3.4 Experiments and Results . . . 52
3.4.1 Qualitative Results . . . 52
3.4.2 Quantitative Results . . . 53
3.5 Conclusion . . . 53

4 Tackling DL Image Reconstruction Artifacts in Limited-Angle DOT 55
4.1 Introduction . . . 55
4.2 Methodology . . . 56
4.2.1 Similarity-wise Loss Function . . . 58
4.2.2 Cross Sensors Learning Transfer . . . 58
4.3 Experiments and Results . . . 59

4.3.1 Dataset . . . 59
4.3.2 Implementation . . . 59
4.3.3 Qualitative Results . . . 60
4.3.4 Quantitative Results . . . 61
4.4 Conclusion . . . 61

5 Multitask Deep Learning for Joint Reconstruction and Localization 62
5.1 Introduction . . . 62
5.1.1 DL-based DOT Reconstruction Methods . . . 62
5.1.2 Multitask Learning . . . 63
5.1.3 Contributions . . . 63
5.2 Methodology . . . 64
5.2.1 Reconstructing Images from DOT Measurements . . . 64
5.2.2 Spatial-wise Attention Network . . . 66
5.2.3 Localization-wise Loss Function . . . 66
5.2.4 Domain Learning Module . . . 68
5.3 Datasets . . . 68
5.3.1 In silico Training Data . . . 68
5.3.2 Phantom and Clinical Test Data . . . 69
5.3.3 Implementation . . . 69
5.3.4 Competing Methods . . . 70
5.3.5 Conventional Algorithms Hyper-parameter Selection and Settings . . . 71
5.4 Results . . . 71
5.4.1 Results on Synthetic Data . . . 72
5.4.2 Results on Phantom Data . . . 75
5.4.3 Results on Clinical Patient Data . . . 78
5.4.4 Multitask versus Single-task . . . 79
5.4.5 Network Depth Trade-off: Complexity versus Performance . . . 79
5.4.6 Effect of Varying the Number of Sensors . . . 80
5.5 Discussion . . . 81
5.5.1 Attention-directed Lesion Reconstruction . . . 81
5.5.2 Network Generalization . . . 82
5.5.3 Limitations . . . 82
5.6 Conclusion . . . 83
5.7 Disclaimer . . . 84

6 End-to-End Diagnosis via Deep Orthogonal Multi-frequency Fusion 85
6.1 Introduction . . . 85
6.1.1 Regularization Approaches in DL-reconstruction Methods . . . 86
6.1.2 Multi-frequency DOT . . . 87

6.1.3 Multi-frequency as Data Fusion . . . 88
6.1.4 Towards Direct Medical Image Analysis in DOT . . . 89
6.1.5 Contributions . . . 90
6.2 Methodology . . . 90
6.2.1 Fusion Network . . . 91
6.2.2 Joint Multi-frequency Reconstruction and Diagnosis . . . 93
6.2.3 Direct Prediction: Raw to Task Model . . . 95
6.2.4 Cross Domain Learning . . . 96
6.3 Results . . . 97
6.3.1 Experimental Design . . . 98
6.3.2 Results on Synthetic Data . . . 102
6.3.3 Results on Clinical Data . . . 105
6.3.4 Effect of Lesion Localization on Accuracy . . . 108
6.4 Discussion . . . 108
6.5 Conclusion . . . 110

7 Conclusions, Limitations, and Future Directions 111
7.1 Thesis Summary . . . 111
7.2 Limitations and Future Direction . . . 112

Bibliography 117

List of Tables

Table 2.1 Comparison between different papers that have applied machine learning to the biomedical image reconstruction problem 39

Table 3.1 End-to-End model's quantitative results on phantom data 53

Table 4.1 Quantitative results on 32 phantom experiments. 61

Table 5.1 Competing algorithm are evaluated under different regularization techniques and hyper-parameters on 32 images from the numerical phantoms 72

Table 5.2 Quantitative results on 32 phantom experiments. Methods with a runtime smaller than 20 images per second (< 0.05s) are considered real-time capable. 76

Table 5.3 Quantitative results on real data experiments. 78

Table 5.4 Sensitivity study: The effect of β on performance. 79

Table 5.5 Performance impact of varying number of network layers. 79

Table 5.6 Variation of the model performance over different number of channels. 80

Table 5.7 Trade off between probe complexity and model performance on the synthetic test set. 81

Table 6.1 Regularization approaches in SOTA DL-reconstruction methods . . . 87

Table 6.2 SOTA multi-frequency DOT for image reconstruction and diagnosis . 88

Table 6.3 Summary of variants of our method architectures input and output details' 96

Table 6.4 Optical coefficients distributions on the in silico dataset for wavelengths in 690-850 nm spectrum [147] 97

Table 6.5 Summary of clinical data 99

Table 6.6 Quantitative reconstruction and diagnosis results on in silico test dataset 102

Table 6.7 Quantitative results on clinical dataset using Raw-to-Task++ 107

List of Figures

Figure 1.1 Summary of contributions presented in this thesis 7

Figure 2.1 Data flow in a medical imaging and analysis: From signal to accurate diagnosis 10
Figure 2.2 Signal propagation and feedback loop effects in sender-receiver systems: Modeling and challenges across modalities 11
Figure 2.3 Uniqueness challenges in ill-Conditioned inverse problems: leveraging prior knowledge to resolve conflicting solutions 12
Figure 2.4 The rise of deep learning in biomedical image reconstruction: Examining the last decade's surge in publications 13
Figure 2.5 CT Image reconstruction from sparse view measurements: Exploring 2D projections and conventional filtered back projection method . . 15
Figure 2.6 An iterative workflow for diffuse optical tomography image reconstruction: From probe design to properties visualization. 16
Figure 2.7 Comparison of reconstruction methods for limited-angle CT 18
Figure 2.8 Exploring deep learning-based reconstruction paradigms for raw sensor data: From Two-Step processing to Raw-to-Task models 20
Figure 2.9 Exploring common network architectures for image reconstruction: From MLP to GANs 23
Figure 2.10 Comparing reconstruction methods for Shepp-Logan phantom in 2D CT imaging: Conventional, DL-Based, and Unrolled-Iterative approaches 27
Figure 2.11 Unified biomedical image processing workflow: Joint reconstruction and image analysis for enhanced diagnosis 28

Figure 3.1 A workflow example for DOT image reconstruction: Illumination, detection, and estimation of optical coefficients in tissue 42
Figure 3.2 Spectrum of absorption in the NIR range from 600 nm to 1,000 nm. 43
Figure 3.3 Workflow for in silico training data generation in DOT image reconstruction: Modeling tissue, simulating forward projection, and deep learning training 47

Figure 3.4 Matching spatial distribution of simulated sources and detectors to physical probe layout in DOT 49
Figure 3.5 In silico training pair generation and phantom test pair collection for DOT image reconstruction: Proposed model architecture visualization 50
Figure 3.6 Qualitative comparison of DL-based DOT reconstruction performance vs conventional techniques 52

Figure 4.1 Overall architecture of the proposed model for phantom imaging: Transfer learning (tackling domain shift) and reconstruction network integration 57
Figure 4.2 Qualitative reconstruction performance of our model compared to state of the art techniques 60

Figure 5.1 Architecture of the Deep Residual Attention Model 65
Figure 5.2 In scilico qualitative comparison of reconstruction performance: Multitask-Net model vs. LS, NCG, MLP, and FC-CNN Approaches 73
Figure 5.3 Phantom qualitative comparison of reconstruction performance: Multitask-Net model vs. LS, NCG, MLP, and FC-CNN Approaches 74
Figure 5.4 Ablation study on qualitative reconstruction performance: Comparing Multitask-Net with FCN and RAN architectures 75
Figure 5.5 Reconstruction performance on real patient data 77
Figure 5.6 Impact of varying number of detectors on Multitask-Net qualitative results for lesion detection 81
Figure 5.7 Challenges in multi-lesion reconstruction: detection, occlusion, and localization 83

Figure 6.1 Breast cancer screening workflow paradigms 86
Figure 6.2 Architecture overview of the proposed DOT image reconstruction and diagnosis method 92
Figure 6.3 Qualitative reconstruction performance of absorption coefficients using the FuseNet++ on in silico samples with varying ground truth lesion sizes, locations, and numbers 100
Figure 6.4 Quantitative diagnosis performance of different models when one vs multi-frequency are used 101
Figure 6.5 Diagnosis prediction confusion matrices when one vs multi-frequency inputs are used 103
Figure 6.6 Quantitative diagnosis performance on clinical dataset 104
Figure 6.7 Qualitative reconstruction results in clinical patients with benign and malignant tumors 106
Figure 6.8 Clinical data diagnosis prediction confusion matrices 107

Figure 6.9 Examples of reconstruction failure cases 108
Figure 6.10 Effect of lesion depth and radius on model prediction accuracy . . . 109

List of Acronyms

CNN	Convolutional Neural Network
CT	Computed Tomography
DT	Distance Transform
DOT	Diffuse Optical Tomography
DL	Deep Learning
E2E	End-to-End
FJ	Fuzzy Jaccard
FCN	Fully Convolutional Network
GAN	Generative Adversarial Network
GT	Ground Truth
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MLP	Multi-layer Perceptron
NIR	Near-Infrared
PSNR	Peak Signal to Noise Ratio
ROI	Region Of Interest
SSIM	Structural Similarity Index
TL	Transfer Learning
US	Ultra Sound

List of Notations

y	Input raw data
x	Ground truth image
$\hat{x}$	Reconstructed image
$\mathcal{N}$	Input noise
λ	Wavelength
μ_a, μ_s	Absorption and scattering optical coefficients
l	Diagnosis label
D	Detectors (Sensors)
S	Sources (Emitters)
$\mathcal{F}(\cdot)$	Forward model
$\mathcal{F}^{-1}(\cdot)$	Inverse model
$\mathcal{L}(\cdot)$	Loss function
$\mathcal{R}(\cdot)$	Regularization function
$\alpha, \beta, \gamma, \eta$	Loss weighting hyperparameters
Θ	Learned model parameters

Chapter 1

Introduction

1.1 Background and Motivation

Early screening of breast cancer lesions, the most common form of cancer in women [13], is critical for patient outcome. Although X-ray mammography is considered the "gold standard" for identifying breast cancer lesions, it has potential cumulative health risks due to the dependence on ionizing radiation and low sensitivity (67.8%) in patients with thick breast tissue [225]. Additionally, the timely availability of mammography devices is often problematic, leading to long waiting times. However, both early detection and timely follow-up are critical, both in first detection and in novel personalized medicine-based treatment paradigms. Non-invasive screening via diffuse optical tomography (DOT), a modality based on near-infrared light (NIR), has gained increasing interest over the past few years as an alternative given that NIR light is well tolerated and can penetrate several centimeters into human soft tissue [219], offering several advantages in terms of safety and costs [22]. However, given the ill-posedness of the problem and the absence of an exact analytic inverse transform, standard DOT reconstruction approaches involve approximating the inverse function and remain computationally expensive, making them prohibitive for real-time imaging and delaying their integration into clinical settings [76]. Furthermore, a portable setting and limited power budget of DOT deployments can further complicate the reconstruction task.

This thesis proposal has both clinical and technical motivations. The *clinical* motivation of this proposal resides in the need for accurate and real-time image reconstruction and analysis of portable DOT devices [227], to enable fast and robust screening of diseases such as breast cancer. A subsequent aim is to reduce the burden on the health care system by prioritizing more invasive diagnostic (e.g., CT) for acute cases, thus reducing overall patient wait times. Furthermore, a portable probe DOT device can dramatically impact the quality and quantity of precision diagnostics available to remote locations and under-privileged communities with limited access to other imaging modalities, enabling critical diagnostics for underrepresented communities.

The *technical* motivation is to advance state-of-the-art DOT reconstruction and diagnosis by developing innovative deep learning (DL) methodologies for an important clinical problem while leveraging advances in optical image acquisition techniques. In the process, we specifically need to overcome the following key challenges: reconstruction data in real-world settings, i.e. patient data, has no ground truth and very low availability due to ethical constraints; a probe design optimized for portability, restricting source and sensor counts, results in information loss and reduced measured signal quality; reproducibility across datasets and patients is critical, as is the robustness of the proposed approaches; and finally diagnosis must be balanced in its accuracy, even in the presence of extremely imbalanced data.

While most works focus on conventional reconstruction methods that exploit only priors encoded by human designers and solve each instance separately, in contrast, we investigate DL methodologies as they can exploit implicitly learned feature encodings from the sensor data while reducing the computational cost by performing the optimization or learning stage offline. We adopt an end-to-end solution, i.e. leveraging the image reconstruction task directly from sensor-domain data by mapping sensor measurements to the image domain while approximating the underlying physics of the inverse problem. This methodological choice is motivated by the automatic feature extraction capabilities of DL models while reducing extra computational cost and avoiding the downsides of combining DL and conventional, where one or the other is hybridized as a pre- or post-processing stage to work around suboptimal results using either alone.

The second methodological choice we made was to leverage physics-based simulation and take advantage of cross-domain learning for data distribution adaptation. This *transfer-learning* approach addresses the lack of available real measurements and images required to train the millions of parameters commonly found in state-of-the-art DL models.

The following section outlines the Chapters within the proposed thesis. Each subsection briefly describes the approach, lists specific contributions, and concludes with details about the resulting publication that corresponds to the work.

1.2 Thesis Contributions

In this thesis, we first review the state of the art in deep learning models applied to biomedical images reconstruction and discuss limitations and potential directions (Chapter. 2). Chapter 3 through Chapter 6 present the proposed approaches for limited-angle DOT image reconstruction and diagnosis. In this Chapter, we provide a background on DOT and introduce a preliminary approach to map sensor data to the image domain in Chapter. 3 and tackle image reconstruction artifacts in limited-angle DOT in Chapter. 4. In Chapter 5, we propose a multitask approach for joint image reconstruction and lesion localization. In Chapter 6, we investigate multifrequency data fusion for reconstruction and diagnosis. Fi-

nally, Chapter 7 summarizes the proposed works and contributions and suggests future directions. The following subsections outline the Chapters within this thesis. Each subsection provides a brief summary of the Chapter, emphasizes its significant contributions, and concludes by providing information about the publication associated with the work. The full details of each work are provided in the respective chapters.

1.2.1 Deep Learning for Biomedical Image Reconstruction

Deep learning-based image reconstruction is a new and rapidly growing field with several efforts to leverage transforming signals collected from sensor data into interpretable representations and demonstrate the superior performance of DL models in 2D or 3D reconstruction. We examine 154 research papers for DL based reconstruction from different imaging modalities and present a thorough survey of the studies leveraging the capabilities of the different paradigms.

We review the contributions of existing literature and analyze the works from several dimensions in terms of model design such as architecture, modules, losses, augmentation, data, etc., as well as evaluation aspects such as speed and evaluation metrics. We discuss those dimensions and how future research can solve their limitations. We summarize all examined works into a comprehensive table encoding the analyzed dimensions.

Contributions:

- We present the first thorough survey on deep learning for image reconstruction that classifies and analyzes different reconstruction model paradigms.
- We analyze state of the art models from several aspects, such as architecture, modules, losses, data, evaluation metrics, etc..
- We examine benchmark datasets and commonly used evaluation metrics.
- We discuss different challenges and future directions in this field.

This Chapter was published [31] in the Journal of Artificial Intelligence Review.

> [31] H. Ben Yedder, B. Cardoen, and G. Hamarneh, "Deep Learning for Biomedical Image Reconstruction: A Survey," Artificial Intelligence Review, pp. 1–33, 2020.

1.2.2 End-to-End Deep Learning based DOT Image Reconstruction

DOT image *reconstruction* has been tackled using a wide variety of reconstruction methods; until recently, the research focus has been on model-based algorithms, whose design follows directly from the mathematical problem formulation and are commonly solved iteratively

until convergence [16]. However, model optimization parameters' tuning and computational cost, prohibitive for real-time application, remain the main disadvantages [296]. This work represents a step forward in the use of machine learning in bio-imaging and DOT image reconstruction specifically. In Chapter 3, we evaluate the use of a DL model to reconstruct images directly from their corresponding DOT raw data. Given the scarcity of real-world sensor-image pairs required for training supervised DL models, we leverage physics-based simulation to generate synthetic datasets. We show that our DL based method improves the quality of reconstructed images while enabling real-time image reconstruction.

Contributions:

- We propose the first end-to-end deep learning model for limited angle NIR-based optical imaging.
- We propose the first method that leverages physics-based optical diffusion simulation to generate in silico training datasets.
- We train the reconstruction model using in silico data only and test it on real measurements acquired from a phantom dataset subject to sensor non-idealities and noise.

This work [30] was published in the MICCAI Workshop on Machine Learning for Medical Image Reconstruction.

> [30] H. Ben Yedder, A. BenTaieb, M. Shokoufi, A. Zahiremami, F. Golnaraghi, and G. Hamarneh, "Deep learning based image reconstruction for diffuse optical tomography," in International Workshop on Machine Learning for Medical Image Reconstruction. Springer, 2018, pp. 112–119.

1.2.3 Tackling DL Image Reconstruction Artifacts in Limited-Angle DOT

As the previous Chapter (Chapter. 3) demonstrated the utility of using DL for real-time DOT image reconstruction, this Chapter tackles model artifacts observed in previous results. In a limited-view data acquisition setting, solving the inverse problem is difficult, especially in the presence of noise, due to the highly diffusive nature of light propagation in biological tissues and the sparsity of the recovered information. To this end, we propose a DL network with a novel hybrid loss to promote *appearance similarity* while penalizing artifacts. Furthermore, a transfer learning module renders an in silico trained network applicable to real world data. The novel loss and transfer learning yields consistent improvements in reconstruction quality and reducing lesion localization error.

Contributions:

- We propose the first attempt to align the sensor domain distribution between in silico and real-world data using a transfer learning module while taking advantage of cross-domain learning.
- We propose a novel loss function that dynamically combines the mean squared error loss and the similarity coefficient-based fuzzy Jaccard term.

This work [33] was published in the MICCAI main conference.

[33] H. Ben Yedder, M. Shokoufi, B. Cardoen, F. Golnaraghi, and G. Hamarneh, "Limited- angle diffuse optical tomography image reconstruction using deep learning," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019, pp. 66–74.

1.2.4 Multitask Deep Learning for Joint Reconstruction and Localization

The previous Chapter (Chapter.4) described a method for *reconstructing* breast tissue's optical properties, per pixel. However, *localization*, which identifies those image regions that have characteristics indicative of a breast lesion, may contribute to a more accurate reconstruction of the lesion as a whole. To attend to the most important features, we propose a deep spatial-wise attention network that re-weights features according to their inter-dependencies in feature space and filters noisy ones. Further, a distance transform based loss is introduced to promote more accurate lesion localization in limited-view data acquisition, given its sensitivity to small changes in the size and position of the foreground object. Our results demonstrate an improvement in localization error over the state of the art and show that our method is sufficiently fast to ensure real-time processing that is necessary for deployment in clinical settings.

The ability of our network to successfully find the inverse mapping by learning only from the simulation data without seeing any real or physical phantom data during the training and validation process can be viewed as evidence that it can learn a general inverse function instead of learning a trivial mapping or memorizing examples. The addition of transfer learning and Gaussian noise to simulated data helps the network generalize to unseen data from physical sensors that capture real signals under non-ideal acquisition conditions. Hence, it generalizes well to both phantom and real patient data.

Contributions:

- We present the first work that tests DL-based DOT reconstruction generalization on real patient data.
- We leverage a deep spatial-wise attention network to adaptively re-weight features and attend to the most important ones.

- We introduce a distance transform loss to improve lesion localization.
- We conduct experiments to assess the trade-offs between network characteristics, scanning probe design, and performance.

This Chapter was published [282] in the IEEE Transaction of Medical Imaging.

> [282] H. Ben Yedder, B. Cardoen, M. Shokoufi, F. Golnaraghi, and G. Hamarneh, "Multitask deep learning reconstruction and localization of lesions in limited angle diffuse optical tomography," IEEE Transactions on Medical Imaging, 2021.

1.2.5 End-to-End Diagnosis via Deep Orthogonal Multi-Frequency Fusion

Identifying breast cancer lesions with a portable DOT device can improve early detection. However, critical to this is the more complex problem of discriminating between malignant and benign lesions. In Chapter 6, we investigate *diagnosis* capability of a portable DOT device. In DOT, multiple light frequencies can be leveraged to derive different tissue composition responses and obtain an optimal trade-off between tissue penetration and signal-to-noise ratio. However, these frequency responses share variable overlap in the absorption spectra of the major tissue constituents in the NIR range [245]. Therefore, this could produce suboptimal downstream analysis tasks if intermediate tissue composition estimation includes errors and artifacts. To tackle this problem, we introduce orthogonal fusion for multi-frequency DOT to maximize the independent contribution of each frequency and emphasize their collective strength. Furthermore, we introduce 'direct medical image analysis' (DMIA) for DOT, where we immediately learn the diagnosis rather than relying on intermediate image reconstruction, with the added advantage of drastically reduced computational overhead. We show that our DMIA approach produces more balanced accuracy between benign and malignant lesions. While classical multifrequency DOT reconstruction methods rely on the combination of multiple images, each from a single frequency, our method proposes an end-to-end real-time solution.

Contributions:

- We present the first deep learning framework that investigates the benefit of multi-frequency data on DOT reconstruction and diagnosis quality.
- We investigate the merit of predicting diagnosis from raw data directly without image reconstruction in DOT (direct medical prediction), a novel trend in medical imaging.
- We maximise the independent contribution of each modulation frequency data while emphasizing their collective strength using an orthogonal projection loss.

This Chapter [32] is under review for publication in the journal IEEE Transactions of Medical Imaging.

[32] H. Ben Yedder, B. Cardoen, M. Shokoufi, F. Golnaraghi, and G. Hamarneh, "Orthogonal Multi-frequency Fusion Based Image Reconstruction and Diagnosis in Diffuse Optical Tomography," submitted.

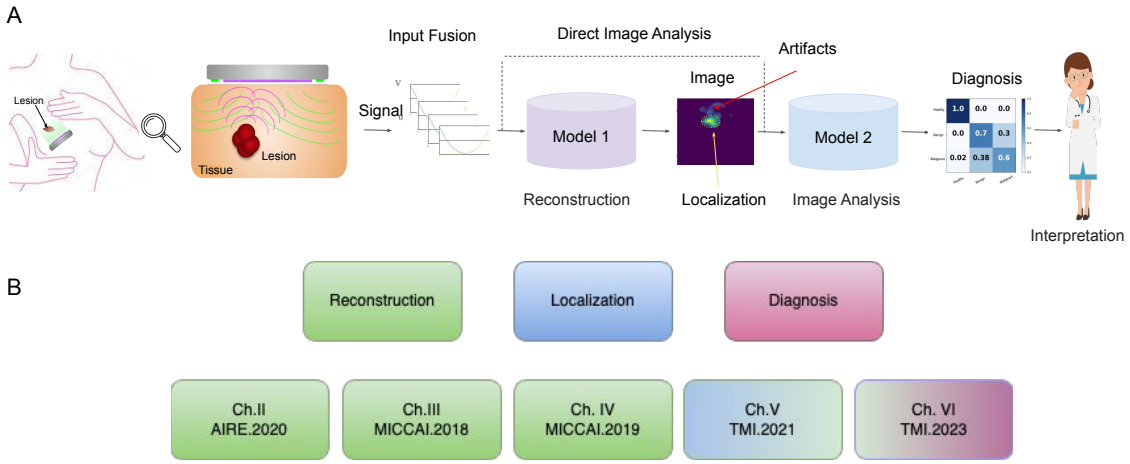

Figure 1.1: Summary of contributions presented in this thesis. (A) Generic representation of a DOT based screening pipeline: A DOT probe acquires signals that are used for image reconstruction and lesion localization followed by diagnosis and interpretation. (B) Summary of the contributions presented in this thesis.

1.3 Summary of Contributions

Our work develops novel computational medical imaging and analysis approaches for screening breast cancer lesions. We tackle the problem of recovering the presence and characteristics of such lesions as well as predicting their malignancy in a critical imaging probe design setting using deep learning based methods. Accurate and real-time image reconstruction and diagnosis with portable DOT devices may enable fast, robust screening of diseases such as breast cancer, where early detection and monitoring are absolutely critical for patient outcome. Technological advancements in DOT imaging and image analysis have the potential to reduce the burden on the health care system by prioritizing more invasive diagnostics for acute cases, thereby reducing overall patient wait times. As such, our work has the potential to directly impact the future line of care screening of breast cancer.

1.4 Research Contributions Excluded from The Thesis

The list below enumerates peer-reviewed publications I was involved with over the course of my PhD studies. While not part of the thesis, these papers highlight the benefits of interdisciplinary science and the cross pollination across domains.

[48] B. Cardoen, **H. Ben Yedder**, S. Lee, and G. Hamarneh. Log-Paradox: Necessary and sufficient conditions for confounding statistically significant pattern reversal under the log-transform, To be submitted to Nature Communications Journal.

[47] B. Cardoen, **H. Ben Yedder**, S. Lee, I. R. Nabi, and G. Hamarneh. DataCurator.jl: Efficient, portable, and reproducible validation, curation, and transformation of large heterogeneous datasets using human-readable recipes compiled into machine verifiable templates, BioInformatics Advances, 2023.

[49] B. Cardoen, **H. Ben Yedder**, A.Sharma, K.C. Chou, I.R. Nabi, and G. Hamarneh. ERGO: efficient recurrent graph optimized emitter density estimation in single molecule localization microscopy. IEEE Transactions on Medical Imaging, 2019, 39(6), pp.1942-1956.

[290] U. Zakia, **H. Ben Yedder**. Dynamic load balancing in SDN-based data center networks. In IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2017, Oct 3, pp.242-247.

[284] **H. Ben Yedder**, U. Zakia, A. Ahmed, and L. Trajković. Modeling prediction in recommender systems using restricted Boltzmann machine. In IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, Oct 5 pp. 2063-2068.

[283] **H. Ben Yedder**, Q. Ding, U. Zakia, Z. Li, S. Haeri, and L. Trajkovic. Comparison of virtualization algorithms and topologies for data center networks. In 26th International Conference on Computer Communication and Networks (ICCCN), 2017 Jul 31, pp. 1-6.

Disclaimer: The author declares that substantial parts of Chapters 2, 3, 4, 5 and 6 of this thesis have been borrowed near-identical from the author's original first-authored publications listed in Section 1.2.

Chapter 2

Deep Learning for Biomedical Image Reconstruction: A Survey

In this Chapter, we provide a comprehensive background and an overview of the state-of-the-art deep-learning based methods in the domain of medical image reconstruction[1]. This foundation will enable us to establish a solid understanding of the current landscape and pave the way for the subsequent exploration of our novel approaches.

2.1 Introduction

Biomedical image reconstruction translates signals acquired by a wide range of sensors into images that can be used for diagnosis and discovery of biological processes in cell and organ tissue. Each biomedical imaging modality leverages signals in different bands of the electromagnetic spectrum, e.g. from gamma rays (Positron emission tomography PET/SPECT)), X-rays (computed tomography (CT)), visible light (microscopy, endoscopy), infrared (thermal images), and radio-frequency (Nuclear magnetic resonance imaging (MRI)), as well as pressure sound waves (in ultrasound (US) imaging) [265]. Reconstruction algorithms transform the collected signals into a 2, 3, or 4-dimensional image.

The accuracy of each reconstruction is critical for discovery and diagnosis. Robustness to noise and generalization cross-modality specifications (e.g., sampling pattern, rate, etc.) and imaging devices parameters' allow a reconstruction algorithm to be used in wider applications. The time required for each reconstruction determines the number of subjects that can be diagnosed as well as the suitability of the technique in operating theatres and emergency situations. The number of measurements needed for a high quality reconstruction impacts the exposure a patient or sample will have to endure. Finally, the hardware requirements define whether a reconstruction algorithm can be used only in a dedicated facility or in portable devices, thus dictating the flexibility of deployment.

[1]The content of this Chapter heavily relies on our published research [31]. Reproduced with permission from Springer Nature.

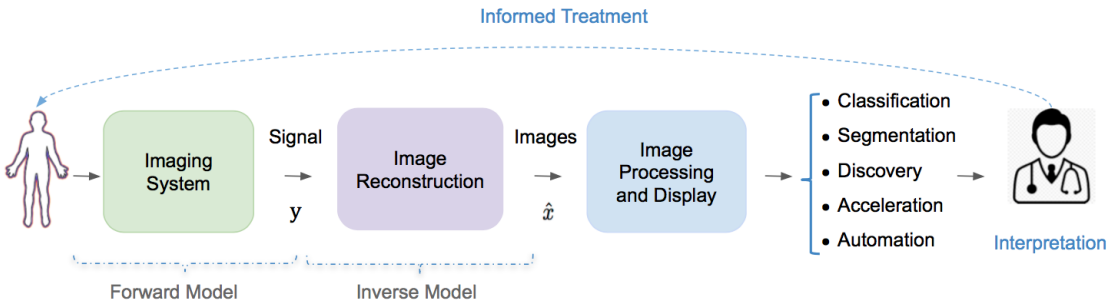

Figure 2.1: Data flow in a medical imaging and image interpretation system. Forward model encodes the physics of the imaging system. The inverse model transforms the collected signals by the acquisition hardware into a meaningful image. The success of diagnosis, evaluation, and treatment rely on accurate reconstruction, image visualization and processing algorithms.

The study of image reconstruction is an active area of research in modern applied mathematics, engineering and computer science. It forms one of the most active interdisciplinary fields of science [85] given that improvement in the quality of reconstructed images offers scientists and clinicians an unprecedented insight into the biological processes underlying disease. Fig. 2.1 provides an illustration of the reconstruction problem and shows a typical data flow in a medical imaging system.

Over the past few years, researchers have begun to apply machine learning techniques to biomedical reconstruction to enable real-time inference and improved image quality in a clinical setting. Here, we first provide an overview of the image reconstruction problem and outline its characteristics and challenges (Section 2.1.1) and then outline the purpose, scope, and the layout of this review (Section 2.1.2).

2.1.1 Inverse Problem and Challenges

From Output to Input

Image reconstruction is the process of forming an interpretable image from the raw data (signals) collected by the imaging device. It is known as an inverse problem where given a set of measurements, the goal is to determine the original structure influencing the signal collected by a receiver given some signal transmission medium properties (Fig. 2.2). Let $y \in \mathbb{C}^M$ represent a set of raw acquired sensor measurements and subject to some unknown noise (or perturbation) vector $\mathcal{N} \in \mathbb{C}^M$ intrinsic to the acquisition process. The objective is to recover the spatial-domain (or spatio-temporal) unknown image $x \in \mathbb{C}^N$ such that:

$$y = \mathcal{A}(\mathcal{F}(x), \mathcal{N}) \tag{2.1}$$

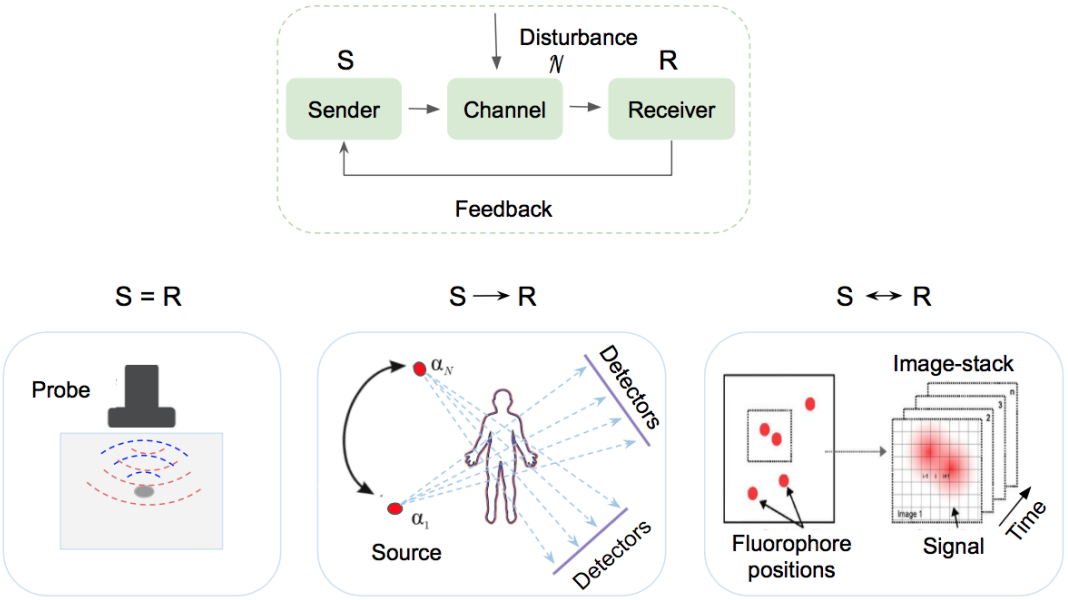

Figure 2.2: Propagation of signals from sender to receiver. While passing through a transmission channel, signals s pick up noise (assuming additive) along the way, until the acquired sensor data $y = s+\mathcal{N}$ reaches the receiver. The properties of the received signal may, via a feedback loop, affect properties of future signal transmission. Sender and receiver modeling differ within modalities. For example the lower figure illustrates in the left ultrasound probe used to send and collect signals (S=R); in the middle: X-ray signal propagates through the subject toward the detector (S $\rightarrow$ R); Single molecule localization microscopy is an instance where S and R are components of a feedback loop (S $\leftrightarrow$ R). The signal, optical measurements of nano-meter precise fluorescent emissions, is reconstructed in a discrete (temporal) X/Y/Z domain. An example of a feedback loop in SMLM is the auto-tuning of the laser power in response to an increase in density (R) that can compromise the spatio-temporal sparsity required for accurate reconstruction [49].

where $\mathcal{F}(\cdot)$ is the forward model operator that models the physics of image-formation, which can include signal propagation, attenuation, scattering, reflection and other transforms, e.g. Radon or Fourier transform. $\mathcal{F}(\cdot)$ can be a linear or a nonlinear operator depending on the imaging modality. $\mathcal{A}$ is an aggregation operation representing the interaction between noise and signal, in the assumption of additive noise $\mathcal{A} = +$.

While imaging systems are usually well approximated using mathematical models that define the forward model, an exact analytic inverse transform $\mathcal{F}^{-1}(\cdot)$ is not always possible. Reconstruction approaches usually resort to iteratively approximate the inverse function and often involve expert-tuned parameters and prior domain knowledge considerations to optimize reconstruction performance.

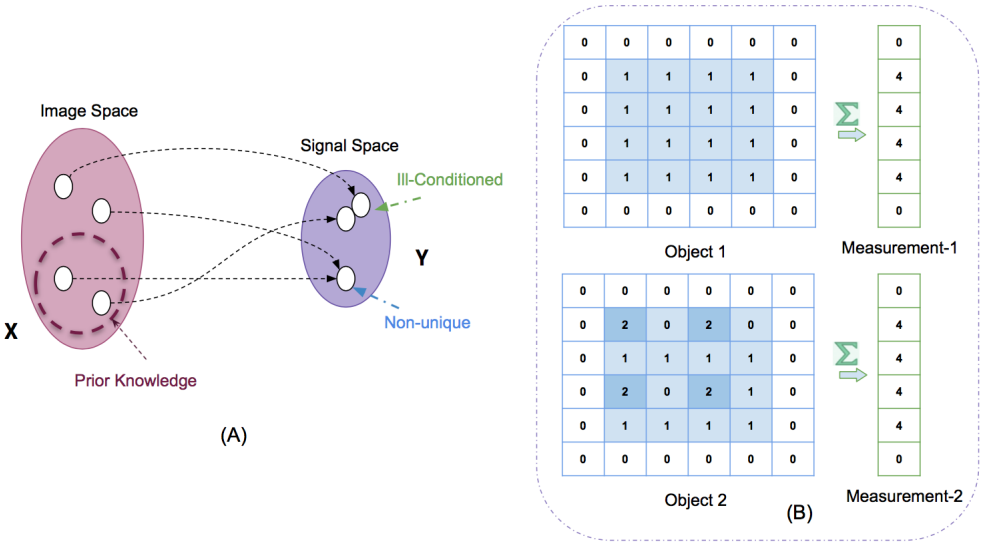

Figure 2.3: Uniqueness and Prior Knowledge in Ill-Conditioned Inverse Problems. (A) A problem is ill-conditioned when two different objects produce very close observed signals. When the observed signals are identical and hence identical reconstructed images, the inverse solution is non-unique. Prior knowledge can be leveraged to rule out certain solutions that conflict with the additional knowledge about the object beyond the measurement vectors. (B) A use case toy example of two objects with the same acquired signals. Prior knowledge about homogeneity of object rules out the second object.

An Ill-Posed Problem

A basic linear and finite-dimensional model of the sampling process leads us to study a discrete linear system of the imaging problem of the form:

$$y = Fx + \mathcal{N}. \tag{2.2}$$

For full sampling, in MRI for instance, we have M = N and the matrix F $\in \mathbb{C}^{M\times N}$ is a well-conditioned discrete Fourier transform matrix. However, when $(M < N)$, there will be many images x that map to the same measurements y, making the inverse an ill-posed problem (Fig. 2.3). Mathematically, the problem is typically under-determined as there would be fewer equations to describe the model than unknowns. Thus, one challenge for the reconstruction algorithm is to select the best solution among a set of potential solutions [170]. One way to reduce the solution space is to leverage domain specific knowledge by encoding priors, i.e. regularization.

Sub-sampling in MRI or sparse-view/limited-angle in CT are examples of how reducing data representation $(M < N)$ can accelerate acquisition or reduce radiation exposure. Additional gains can be found in lowered power, cost, time, and device complexity, albeit at the cost of increasing the degree of ill-posedness and complexity of the reconstruction problem. This brings up the need for sophisticated reconstruction algorithms with high fea-

ture extraction power to make the best use of the collected signal, capture modality-specific imaging features, and leverage prior knowledge. Furthermore, developing high-quality re-

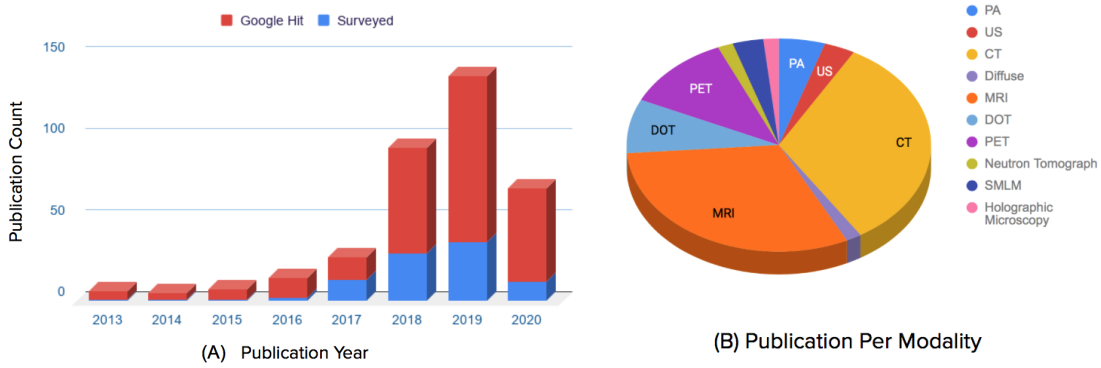

Figure 2.4: (A) The marked increase in publications on biomedical image reconstruction and deep learning in the past 10 years. In red: results obtained by PubMed query that can be found at: `https://bit.ly/recon_hit`. Search query identified 310 contributions which were filtered according to quality, venue, viability and idea novelty, resulting in a total of 95 contributions as representatives covered in this survey(blue). (B) The pie chart represents the frequency of studies per modality covered in this survey.

construction algorithms requires not only a deep understanding of both the physics of the imaging systems and the biomedical structures but also specially designed algorithms that can account for the statistical properties of the measurements and tolerate errors in the measured data.

2.1.2 Scope of this Survey

The field of biomedical image reconstruction has undergone significant advances over the past few decades and can be broadly classified into two categories: conventional methods (analytical and optimization based methods) and data-driven or learning-based methods. Conventional methods (discussed in section 2.2) are the most dominant and have been extensively studied over the last few decades with a focus on how to improve their results [250, 132, 19] and reduce their computational cost [72].

Researchers have recently investigated deep learning (DL) approaches for various biomedical image reconstruction problems (discussed in section 2.3) inspired by the success of deep learning in computer vision problems and medical image analysis [98]. This topic is relatively new and has gained a lot of interest over the last few years, as shown in Fig. 2.4-A and listed in Table 2.1, and forms a very active field of research with numerous special journal issues [256, 258, 205]. MRI and CT received the most attention among studied modalities, as illustrated in Fig. 2.4-B, given their widespread clinical usage, the availability of analytic inverse transform and the availability of public (real) datasets. Several special issues journals devoted to MRI reconstruction have been recently published [146, 156, 215].

Fessler et al. [85] wrote a brief chronological overview on image reconstruction methods highlighting an evolution from basic analytical to data-driven models. [169]) summarized works using CNNs for inverse problems in imaging. Later, Lucas et al. [162] provided an overview of the ways deep neural networks may be employed to solve inverse problems in imaging (e.g., denoising, superresolution, inpainting). As illustrated in Fig. 2.4 since their publications a great deal of work has been done warranting a review. Recently, Mccann et Unser [170] wrote a survey on image reconstruction approaches where they presented a toolbox of operators that can be used to build imaging systems models and showed how a forward model and sparsity-based regularization can be used to solve reconstruction problems. While their review is more focused on the mathematical foundations of conventional methods, they briefly discussed data-driven approaches, their theoretical underpinning, and performance. Similarly, Arridge et al. [14] gave a comprehensive survey of algorithms that combine model and data-driven approaches for solving inverse problems with focus on deep neural based techniques and pave the way towards providing a solid mathematical theory. Zhang et al. [294] provided a conceptual review of some recent DL-based methods for CT with a focus on methods inspired by iterative optimization approaches and their theoretical underpinning from the perspective of representation learning and differential equations.

This survey provides an overview of biomedical image reconstruction methods with a focus on DL-based strategies, discusses their different paradigms (e.g., image domain, sensor domain (raw data) or end to end learning, architecture, loss, etc.) and how such methods can help overcome the weaknesses of conventional non-learning methods. To summarize the research done to date, we provide a comparative, structured summary of works representative of the variety of paradigms on this topic, tabulated according to different criteria, discusses the pros and cons of each paradigm as well as common evaluation metrics and training dataset challenges. The theoretical foundation was not emphasized in this work as it was comprehensively covered in the aforementioned surveys. A summary of the current state of the art and outline of what we believe are strategic directions for future research are discussed.

The remainder of this Chapter is organized as follows: in Section 2.2 we give an overview of conventional methods discussing their advantages and limitations. We then introduce the key machine learning paradigms and how they are being adapted in this field complementing and improving on conventional methods. A review of available data-sets and performance metrics is detailed in Section 2.3. Finally, we conclude by summarizing the current state-of-the-art and outlining strategic future research directions (Section 2.7).

2.2 Conventional Image Reconstruction Approaches

A wide variety of reconstruction algorithms have been proposed during the past few decades, having evolved from analytical methods to iterative or optimization-based methods that

account for the statistical properties of the measurements and noise as well as the hardware of the imaging system [85]. While these methods have resulted in significant improvements in reconstruction accuracy and artifact reduction, are in routine clinical use currently, they still present some weaknesses. A brief overview of these methods' principles is presented in this section outlining their weaknesses.

2.2.1 Analytical Methods

Analytical methods are based on a continuous representation of the reconstruction problem and use simple mathematical models for the imaging system. Classical examples are the inverse of the Radon transform such as filtered back-projection (FBP) for CT and the inverse Fourier transform (IFT) for MRI. These methods are usually computationally inexpensive (in the order of ms) and can generate good image quality in the absence of noise and under the assumption of full sampling/all angles projection [170]. They typically consider only the geometry and sampling properties of the imaging system while ignoring the details of the system physics and measurement noise [85].

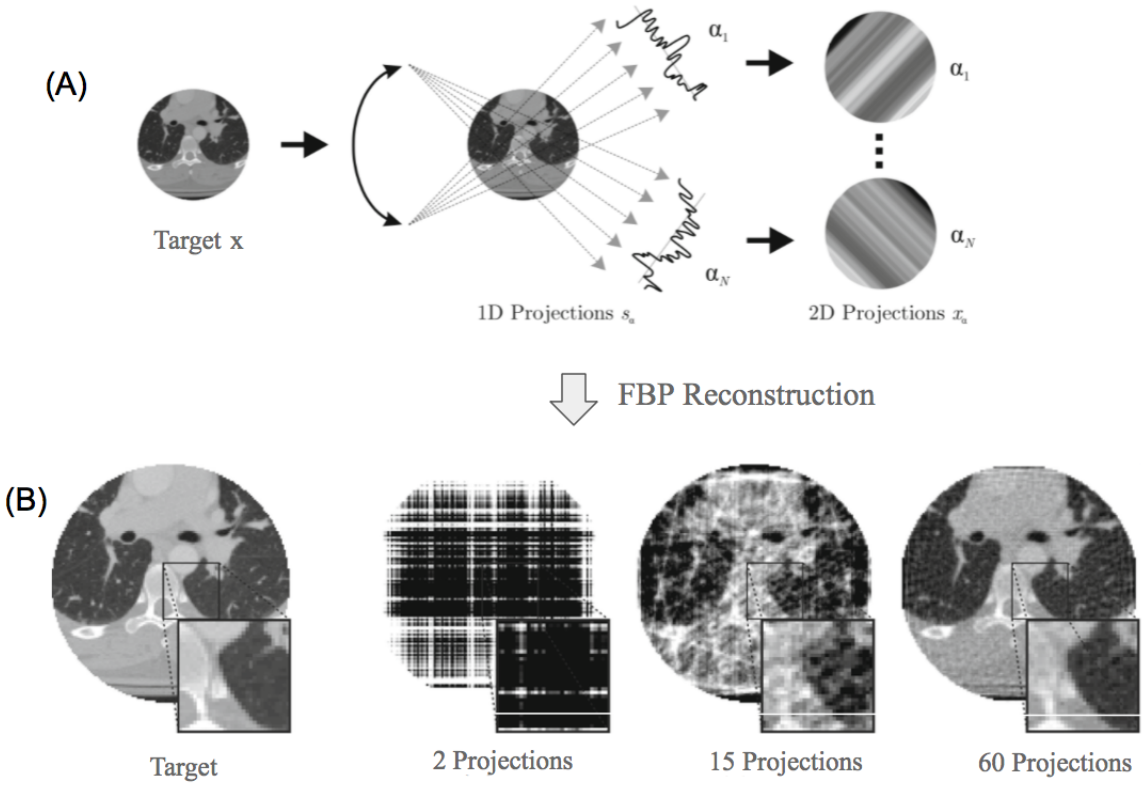

Figure 2.5: CT Image reconstruction from sparse view measurements. (A) Generation of 2D projections from a target 2D CT slice image x for a number of N fixed angles α_i. (B) Image reconstructed using conventional filtered back projection method for different number of a projection angles N. (Modified figures from [246])

When dealing with noisy or incomplete measured data e.g., reducing the measurement sampling rate, analytical methods results deteriorate quickly as the signal becomes weaker. Thus, the quality of the produced image is compromised. Thaler et al. [246] provided examples of a CT image reconstruction using the FBP method for different limited projection angles and showed that analytical methods are unable to recover the loss in the signal (Fig. 2.5), resulting in compromised diagnostic performance.

2.2.2 Iterative Methods

Iterative reconstruction methods, based on more sophisticated models for the imaging system's physics, sensors and noise statistics, have attracted a lot of attention over the past few decades. They combine the statistical properties of data in the sensor domain (raw measurement data), prior information in the image domain, and sometimes parameters of the imaging system into their objective function [170]. Compared to analytical methods iterative reconstruction algorithms offer a more flexible reconstruction framework and better robustness to noise and incomplete data representation problems at the cost of increased computation [205].

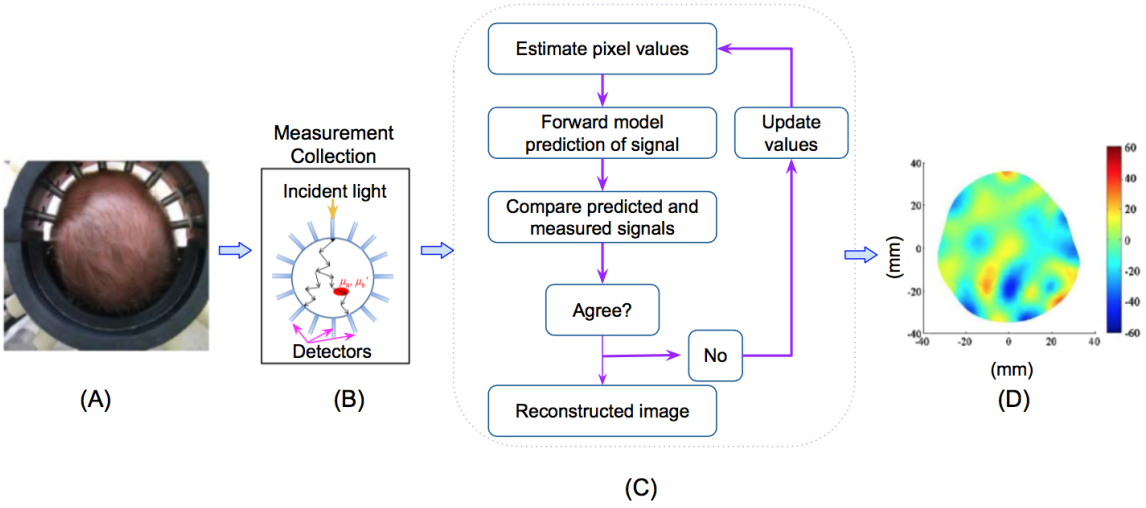

Figure 2.6: Iterative image reconstruction workflow example. (A) Diffuse optical tomography (DOT) fibers brain probe consisting of a set of fibers for illumination and outer bundle fibers for detection [120]. (B) Probe scheme and light propagation modelling in the head, used by the forward model. (C) Iterative approach pipeline. (D) DOT reconstructed image shows the total Hemoglobin (Hb) concentrations in the brain. (Figure licensed under CC-BY 4.0 Creative Commons license.)

Iterative reconstruction methods involve minimizing an objective function that usually consists of a data term and a regularization terms imposing some prior:

$$\hat{x}^* = \underset{\hat{x}}{\operatorname{argmin}} \, \|\mathcal{F}(\hat{x}) - y\| + \lambda \mathcal{R}(\hat{x}), \tag{2.3}$$

where $||\mathcal{F}(\hat{x}) - y\|$ is a data fidelity term that measures the consistency of the approximate solution $\hat{x}$, in the space of acceptable images e.g 2D, 3D and representing the physical quantity of interest, to the measured signal y, which depends on the imaging operator and could include images, Fourier samples, line integrals, etc. $\mathcal{R}(\cdot)$ is a regularization term encoding the prior information about the data, and λ is a hyper-parameter that controls the contribution of the regularization term. The reconstruction error is minimized iteratively until convergence. We note that the solution space after convergence does not need to be singular, for example in methods where a distribution of solutions is sampled (Bayesian neural networks [277, 164]). In such cases, the "$\hat{x}^* \in$" notation replaces the current "$\hat{x}^* =$". However, for consistency with recent art we use the simplified notation assuming a singular solution space at convergence.

The regularization term is often the most important part of the modeling and what researchers have mostly focused on in the literature as it vastly reduces the solution space by accounting for assumptions based on the underlying data (e.g., smoothness, sparsity, spatio-temporal redundancy). The interested reader can refer to [74, 170] for more details on regularization modeling. Fig. 2.6 shows an example of an iterative approach workflow for diffuse optical tomography (DOT) imaging.

Several image priors were formulated as sparse transforms to deal with incomplete data issues. The sparsity idea, representing a high dimensional image x by only a small number of nonzero coefficients, is one dominant paradigm that has been shown to drastically improve the reconstruction quality especially when the number of measurements M or theirs signal to noise ratio (SNR) is low. Given the assumption that an image can be represented with a few nonzero coefficients (instead of its number of pixels), it is possible to recover it from a much smaller number of measurements. A popular choice for a sparsifying transform is total variation (TV) that is widely studied in academic literature. The interested reader is referred to [209] for TV based algorithms modeling details. While TV imposes a strong assumption on the gradient sparsity via the non-smooth absolute value that is more suited to piece-wise constant images, TV tends to cause artifacts such as blurred details and undesirable patchy texture in the final reconstructions (Fig. 2.7 illustrates an example of artifacts present in TV-based reconstruction). Recent work aimed at exploiting richer feature knowledge to overcome TV's weaknesses, for example TV-type variants [298], non-local means (NLM) [299], wavelet approaches [89], and dictionary learning [276]. Non-local means filtering methods, widely used for CT [295], are operational in the image domain and allow

the estimation of the noise component based on multiple patches extracted at different locations in the image [237].

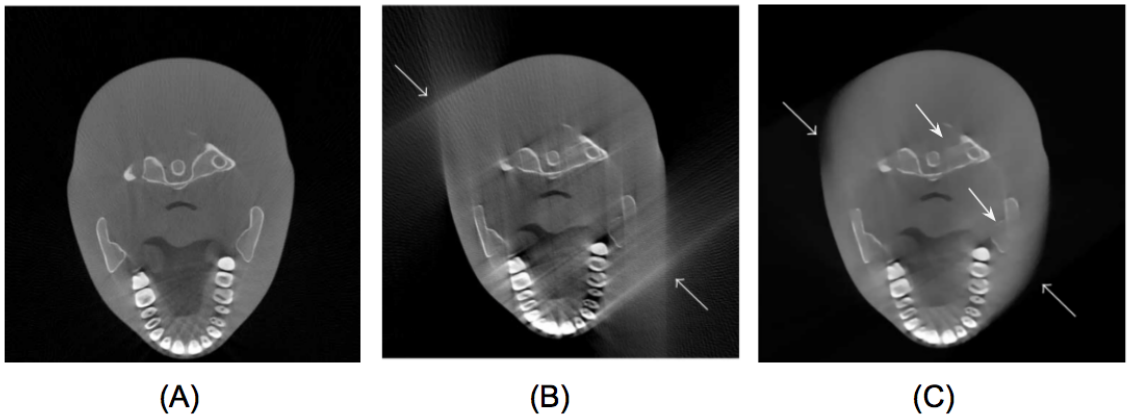

Figure 2.7: Fig. Reconstruction results of a limited-angle CT real head phantom (120 projections) using different conventional reconstruction methods. (A) Reference image reconstructed from 360 projections, (B) Image reconstruction using analytical FBP, (C) Images reconstructed by iterative POCS-TV with initial zero image [108]. Arrows point out to artifacts present in the reconstructed images.(Figure licensed under CC-BY 4.0 Creative Commons license.)

While sparsity measures the first order sparsity, i.e the sparsity of vectors (the number of non-zero elements), some models exploited alternative properties such as the low rank of the data, especially when processing dynamic or time-series data, e.g. dynamic and functional MRI [158, 233] where data can have a correlation over time. Low rank can be interpreted as the measure of the second order (i.e., matrix) sparsity [159]. Low rank is of particular interest to compression given its requirement to a full utilization of the spatial and temporal correlation in images or videos leading to a better exploitation of available data. Recent work combined low rank and sparsity properties for improved reconstruction results [301]. Guo et al.; Otazo et al. [107, 191] decomposed the dynamic image sequence into the sum of a low-rank component, that can capture the background or slowly changing parts of the dynamic object, and sparse component that can capture the dynamics in the foreground such as local motion or contrast changes. The interested reader is referred to [159, 205] for low rank based algorithms modeling survey.

The difficulties of solving the image reconstruction problems motivated the design of highly efficient algorithms for large scale, nonsmooth and nonconvex optimization problems such as the alternating direction method of multipliers (ADMM) [41, 88], primal-dual algorithm [50], iterative shrinkage-thresholding algorithm (ISTA) [68], to name just a few. For instance, using the augmented Lagrangian function [237] of Eq. 2.3, the ADMM algorithm solves the reconstruction problem by breaking it into smaller subproblems that are easier to solve. Despite the problem decomposition, a large number of iterations is still required

for convergence to a satisfactory solution. Furthermore, performance is defined to a large extent by the choice of transformation matrix and the shrinkage function which remain challenging to choose [170].

ISTA is based on a simpler gradient-based algorithm where each iteration leverages hardware accelerated matrix-vector multiplications representing the physical system modeling the image acquisition process and its transpose followed by a shrinkage/soft-threshold step [68]. While the main advantage of ISTA relies on its simplicity, it has also been known to converge quite slowly [294]. Convergence was improved upon by a fast iterative soft-thresholding algorithm (FISTA) [29] based on the idea of Nesterov ([184]).

Overall, although iterative reconstruction methods showed substantial accuracy improvements and artifact reductions over the analytical ones, they still face three major weaknesses: First, iterative reconstruction techniques tend to be vendor-specific since the details of the scanner geometry and correction steps are not always available to users and other vendors. Second, there are substantial computational overhead costs associated with popular iterative reconstruction techniques due to the load of the projection and back-projection operations required at each iteration. The computational cost of these methods is often several orders of magnitude higher than analytical methods, even when using highly-optimized implementations. A trade-off between real-time performance and quality is made in favor of quality in iterative reconstruction methods due to their non-linear complexity of quality in function of the processing time. Finally, the reconstruction quality is highly dependent on the regularization function form and the related hyper-parameters settings as they are problem-specific and require non-trivial manual adjustments. Over-imposing sparsity ($\mathcal{L}1$ penalties) for instance can lead to cartoon-like artifacts [170]. Proposing a robust iterative algorithm is still an active research area [241, 176].

2.3 Deep Learning Based Image Reconstruction

To further advance biomedical image reconstruction, a more recent trend is to exploit deep learning techniques for solving the inverse problem to improve resolution accuracy and speed up reconstruction results. As a deep neural network represents a complex mapping, it can detect and leverage features in the input space and build increasingly abstract representations useful for the end-goal. Therefore, it can better exploit the measured signals by extracting more contextual information for reconstruction purposes. In this section, we summarize works using DL for inverse problems in imaging.

Learning-based image reconstruction is driven by data where a training dataset is used to tune a parametric reconstruction algorithm that approximates the solution to the inverse problem with a significant one-time, offline training cost that is offset by a fast inference time. There is a variety of these algorithms, with some being closely related to conventional methods and others not. While some methods considered machine learning as a reconstruc-

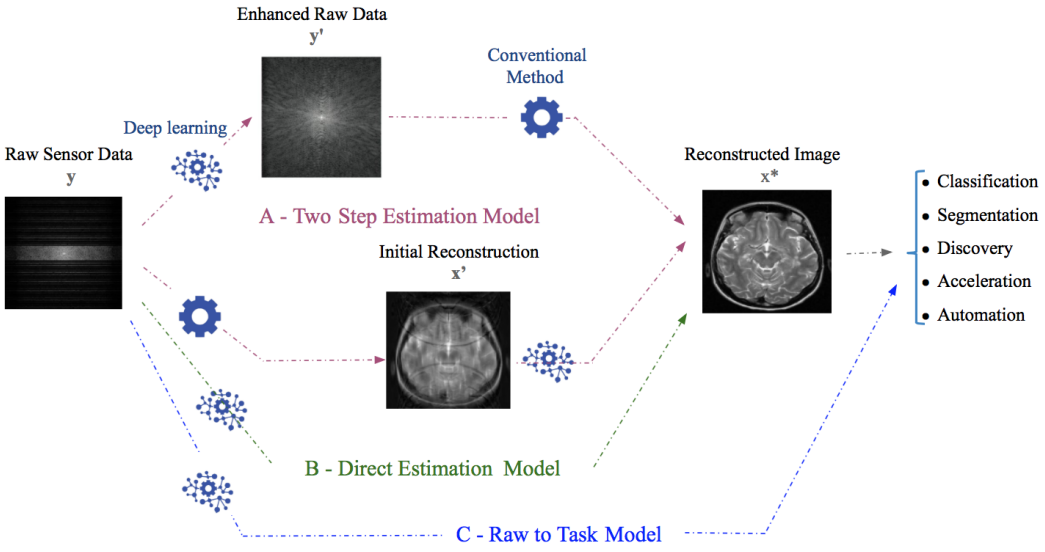

Figure 2.8: Visualization of common deep learning-based reconstruction paradigms from raw sensor data. (A) A two-step processing model is shown where deep learning complements conventional methods (Section 2.3.1). A typical example would be to pre-process the raw sensor data using a conventional approach f_1, enhance the resulting image with a deep learning model f_{θ_2} then perform task processing using f_{θ_3}. The overall function will thus be $f_{\theta_3} \circ f_{\theta_2} \circ f_1$. Vice versa, one can pre-process the raw sensor data using a DL-based approach f_{θ_2} to enhance the collected raw data and recover missing details then apply conventional method f_1 to reconstruct image and finally perform task processing using f_{θ_3}. The overall function will be $f_{\theta_3} \circ f_1 \circ f_{\theta_2}$ in this case. (B) An end-to-end model is shown: The image is directly estimated from the raw sensor data with a deep learning model (2.3.2) followed by downstream processing tasks. The overall function will be $f_{\theta_3} \circ f_{\theta_2}$. (C) Task results can be inferred with or without explicit image reconstruction (Section 2.3.3) using f_{θ_3} function.

tion step by combining a traditional model with deep modeling to recover the missing details in the input signal or enhance the resulting image (Section 2.3.1), some others considered a more elegant solution to reconstruct an image from its equivalent initial measurements directly by learning all the parameters of a deep neural network, in an end-to-end fashion, and therefore approximating the underlying physics of the inverse problem (Section 2.3.2), or even going further and solving for the target task directly (Section 2.3.3). Fig. 2.8 shows a generic example of the workflow of these approaches. Table 2.1 surveys various papers based on these different paradigms and provides a comparison in terms of used data (Table 2.1-Column "Mod.","Samp.","D"), architecture (Table 2.1-Column "TA","Arch."), loss and regularization (Table 2.1-Column "Loss","Reg."), augmentation (Table 2.1-Column "Aug."), etc.

2.3.1 Deep Learning as Processing Step: Two Step Image Reconstruction Models

Complementing a conventional image reconstruction approach with a DL-model enables improved accuracy while reducing the computational cost. The problem can be addressed either in the sensor domain (pre-processing) or the image domain (post-processing) (Fig. 2.8-A, Table 2.1-Column "E2E").

A Pre-Processing Step (Sensor Domain)

The problem is formulated as a regression in the sensor domain from incomplete data representation (e.g., sub-sampling, limited view, low dose) to complete data (full dose or view) using DL methods and has led to enhanced results [129, 157, 59]. The main goal is to estimate, using a DL model, missing parts of the signal that have not been acquired during the acquisition phase in order to input a better signal to an analytical approach for further reconstruction.

Hyun et al. [129] proposed a k-space correction based U-Net to recover the unsampled data followed by an IFT to obtain the final output image. They demonstrated artifact reduction and morphological information preservation when only 30% of the k-space data is acquired. Similarly, Liang et al. [157] proposed a CT angular resolution recovery based on deep residual convolutional neural networks (CNNs) for accurate full view estimation from unmeasured views while reconstructing images using filtered back projection. Reconstruction demonstrated speed-up with fewer streaking artifacts along with the retrieval of additional important image details. Unfortunately, since noise is not only present in the incomplete data acquisition case, but also in the full data as well, minimizing the error between the reference and the predicted values can cause the model to learn to predict the mean of these values. As a result, the reconstructed images can suffer from lack of texture detail.

Huang et al. [125] argue that DL-based methods can fail to generalize to new test instances given the limited training dataset and DL's vulnerability to small perturbations especially in noisy and incomplete data case. By constraining the reconstructed images to be consistent with the measured projection data, while the unmeasured information is complemented by learning based methods, reconstruction quality can be improved. DL predicted images are used as prior images to provide information in missing angular ranges first followed by a conventional reconstruction algorithm to integrate the prior information in the missing angular ranges and constrain the reconstruction images to be consistent to the measured data in the acquired angular range.

Signal regression in the sensor domain reduces signal loss enabling improved downstream results from the coupled analytic method. However, the features extracted by DL methods are limited to the sensor domain only while analytical methods' weaknesses are still present in afterword processing.

A Post-Processing Step (Image Domain)

The regression task is to learn the mapping between the low-quality reconstructed image and its high-quality counterpart. Although existing iterative reconstruction methods improved the reconstructed image quality, they remain computationally expensive and may still result in reconstruction artifacts in the presence of noise or incomplete information, e.g. sparse sampling of data [65, 232]. The main difficulty arises from the non-stationary nature of the noise and serious streaking artifacts due to information loss [58, 5]. Noise and artifacts are challenging to isolate as they may have strong magnitudes and do not obey specific model distributions in the image domain [260]. The automatic learning and detection of complex patterns offered by deep neural networks can outperform handcrafted filters in the absence of complete information [278, 232].

Given an initial reconstruction estimate from a direct inverse operator e.g., FBP [242, 110, 55], IFT [260], or few iterative approach steps [133, 65, 274], deep learning is used to refine the initialized reconstruction and produce the final reconstructed image. For example, Chen et al. [56] used an autoencoder to improve FBP results on a limited angle CT projection. Similarly, Jin et al. [133] enhanced FBP results on a sparse-view CT via subsequent filtering by a U-Net to reduce artifacts. U-Net and other hour-glass shaped architectures rely on a bottleneck layer to encode a low-dimensional representation of the target image.

Generative adversarial networks (GAN) [104] were leveraged to improve the quality of reconstructed images. Wolterink et al. [269] proposed to train an adversarial network to estimate full-dose CT images from low-dose CT ones and showed empirically that an adversarial network improves the model's ability to generate images with reduced aliasing artifacts. Interestingly, they showed that combining squared error loss with adversarial loss can lead to a noise distribution similar to that in the reference full-dose image even when no spatially aligned full-dose and low dose scans are available.

Yang et al. [278] proposed a deep de-aliasing GAN (DAGAN) for compressed sensing MRI reconstruction that resulted in reduced aliasing artifacts while preserving texture and edges in the reconstruction. Remarkably, a combined loss function based on content loss (consisting of a pixel-wise image domain loss, a frequency domain loss and a perceptual VGG loss) and adversarial loss were used. While frequency domain information was incorporated to enforce similarity in both the spatial (image) and the frequency domains, a perceptual VGG coupled to a pixel-wise loss helped preserve texture and edges in the image domain.

Combining DL and conventional methods reduce the computational cost but has its own downsides. For instance, the features extracted by DL methods are highly impacted by the results of the conventional methods, especially in case of limited measurements and the presence of noise where the initially reconstructed image may contain significant and complex artifacts that may be difficult to remove even by DL models. In addition, the information missing from the initial reconstruction is challenging to be reliably recovered by

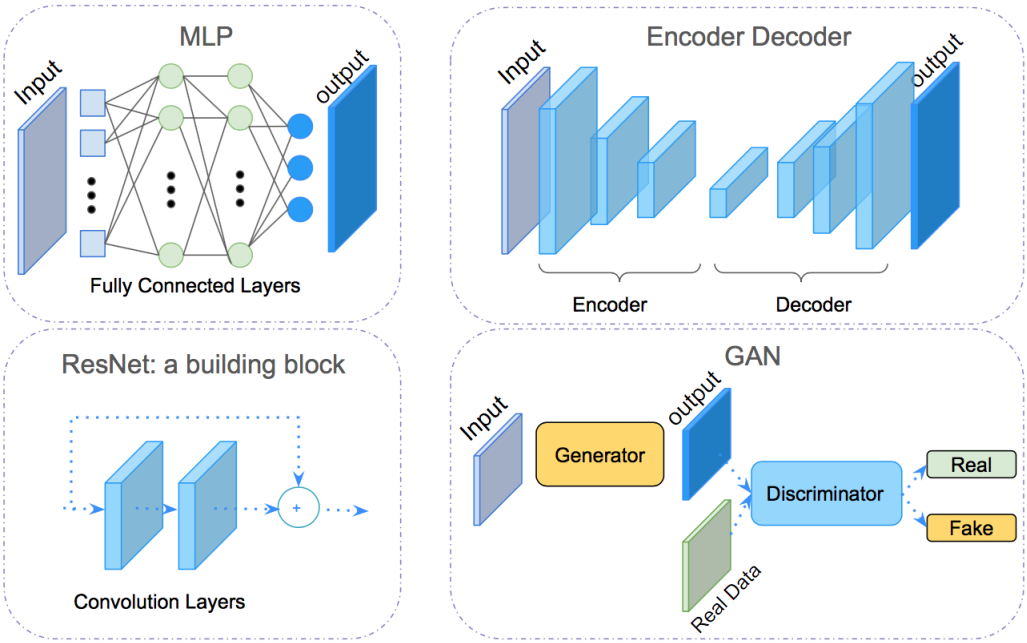

Figure 2.9: Common network architectures used for image reconstruction. From left to right: a multilayer perceptron network based on fully connected layers; an encoder-decoder architecture based convolutional layers; a residual network, e.g. ResNet [116], utilizing skip connections (skip connection is crucial in facilitating stable training when the network is very deep); and a generative adversarial network (GAN). A decoder like architecture includes only the decoder part of the encoder-decoder architecture and may be preceded by a fully connected layer to map measurements to the image space depending on the input data size. A U-Net [210] resembles the encoder-decoder architecture while it uses skip connections between symmetric depths. A GAN [104] includes a generator and descriminor that contest with each other in a min-max game. The generator learns to create more realistic data by incorporating feedback from the discriminator. We refer the interested reader to [7, 143] for a an in depth survey of various types of deep neural network architectures.

post-processing like many inverse problems in the computer vision literature such as image inpainting. Furthermore, the number of iterations required to obtain a reasonable initial image estimate using an iterative method can be hard to define and generally requires a long run-time (in the order of several min) to be considered for real-time scanning. Therefore, the post-processing approach may be more suitable to handle initial reconstructions that are of relatively good quality and not drastically different from the high-quality one.

2.3.2 End-to-End Image Reconstruction: Direct Estimation

An end-to-end solution leverages the image reconstruction task directly from sensor-domain data using a deep neural network by mapping sensor measurements to image domain while approximating the underlying physics of the inverse problem (Fig. 2.8-B). This direct esti-

mation model may represent a better alternative as it benefits from the multiple levels of abstraction and the automatic feature extraction capabilities of deep learning models.

Given pairs of measurement vectors $y \in \mathbb{C}^M$ and their corresponding ground truth images $x \in \mathbb{C}^N$ (that produce y), the goal is to optimize the parameters $\theta \in \mathbb{R}^d$ of a neural network in an end-to-end manner to learn the mapping between the measurement vector y and its reconstructed tomographic image x, which recovers the parameters of underlying imaged tissue. Therefore, we seek the inverse function $\mathcal{A}^{-1}(\cdot)$ that solves:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}\left(\mathcal{A}^{-1}(y,\theta), x\right) + \lambda\mathcal{R}(\mathcal{A}^{-1}(y,\theta)), \tag{2.4}$$

where $\mathcal{L}$ is the loss function of the network that, broadly, penalizes the dissimilarity between the estimated reconstruction and the ground truth. The regularization term $\mathcal{R}$, often introduced to prevent over-fitting, can apply penalties on layer parameters (weights) or layer activity (output) during optimization. $\mathcal{L}1/\mathcal{L}2$ norms are the most common choices. λ is a hyper-parameter that controls the contribution of the regularization term. Best network parameters (θ^*) depend on hyper-parameters, initialization and a network architecture choice.

Recently, several paradigms have emerged for end-to-end DL-based reconstruction the most common of which are generic DL models and DL models that unroll an iterative optimization.

Generic Models

Although some proposed models rely on multilayer perceptron (MLP) feed-forward artificial neural network [197, 39, 83, 259], CNNs remain the most popular generic reconstruction models mainly due to their data filtering and features extraction capabilities. Specifically, encoder-decoder [183, 112], U-Net [253], ResNet [45] and decoder like architecture [286, 272, 307] are the most dominant architectures as they rely on a large number of stacked layers to enrich the level of features extraction. A set of skip connections enables the later layers to reconstruct the feature maps with both the local details and the global texture and facilitates stable training when the network is very deep. Fig. 6.2 illustrates some of the architectures that are widely adopted in medical image reconstruction.

The common building blocks of neural network architectures are convolutional layers, batch normalization layers, and rectified linear units (ReLU). ReLU is usually used to enforce information non-negativity properties, given that the resulting pixels values represent tissue properties e.g., chromophores concentration maps [285], refractive index [237], and more examples in Table 2.1. Batch normalization is used to reduce the internal covariate shift and accelerates convergence. The resulting methods can produce relatively good reconstructions in a short time and can be adapted to other modalities but require a large training dataset and good initialization parameters. Table 2.1-Column "E2E" (check-marked)

summarizes papers by architecture, loss, and regularization for 2D, 3D, 4D, and different modalities.

Zhu et al. [307] proposed a manifold learning framework based decoder neural network to emulate the fast-Fourier transform (FFT) and learn an end-to-end mapping between k-space data and image domains where they showed artifact reduction and reconstruction speed up. However, when trained with an $\mathcal{L}1$ or $\mathcal{L}2$ loss only, a DL-based reconstructed image still exhibits blurring and information loss, especially when used with incomplete data. Similarly, Ben Yedder et al. [30] proposed a decoder like model for DOT image reconstruction. While increased reconstruction speed and lesion localization accuracy are shown, some artifacts are still present in the reconstructed image when training with $\mathcal{L}2$ loss only. This motivated an improved loss function in their follow-up work [33] where they suggested combining $\mathcal{L}2$ with a Jaccard loss component to reduce reconstructing false-positive pixels.

Thaler et al. [246] proposed a Wasserstein GAN (WGAN) based architecture for improving the image quality for 2D CT image slice reconstruction from a limited number of projection images using a combination of $\mathcal{L}1$ and adversarial losses. Similarly, Ouyang et al. [192] used a GAN based architecture with a task-specific perceptual and $\mathcal{L}1$ losses to synthesize PET images of high quality and accurate pathological features. Some attempts were made to reconstruct images observed by the human visual system directly from brain responses using GAN [224, 235]. Shen et al. [224] trained a GAN for functional magnetic resonance imaging (fMRI) reconstruction to directly map between brain activity and perception (the perceived stimulus). An adversarial training strategy which consists of three networks: a generator, a comparator, and a discriminator was adopted. The comparator is a pre-trained neural network on natural images used to calculate a perceptual loss term. The loss function is based on $\mathcal{L}2$ distance in image space, perceptual loss (similarity in feature space) and adversarial loss.

To further enhance results and reduce artifacts due to motion and corruption of k-space signal, Oksuz et al. [189] proposed a recurrent convolutional neural network (RCNN) to reconstruct high quality dynamic cardiac MR images while automatically detecting and correcting motion-related artifacts and exploiting the temporal dependencies within the sequences. Proposed architecture included two sub-networks trained jointly: an artifact detection network that identifies potentially corrupted k-space lines and an RCNN for reconstruction.

To relax the requirement of a large number of training samples, a challenging requirement in a medical setting, simulating data was proposed as an alternative source of training data. However, creating a realistic synthetic dataset is a challenging task in itself as it requires careful modeling of the complex interplay of factors influencing real-world acquisition environment. To bridge the gap between the real and in silico worlds, transfer learning provides a potential remedy as it helps transfer the measurements from the simulation domain

to the real domain by keeping the general attenuation profile while accounting for real-world factors such as scattering, etc.

Ben Yedder et al. [33] proposed a supervised sensor data distribution adaptation based MLP to take advantage of cross-domain learning and reported accuracy enhancement in detecting tissue abnormalities. Zhou et al. [306] proposed unsupervised CT sinograms adaptation, based on CycleGAN and content consistent regularization, to further alleviate the need for real measurement-reconstruction pairs. Interestingly, the proposed method integrated the measurement adaptation network and the reconstruction network in an end-to-end network to jointly optimize the whole network.

Waibel et al. [253] investigated two different variants of DL-based methods for photo-acoustique image reconstruction from limited view data. The first method is a post-processing DL method while the second one is an end-to-end reconstruction model. Interestingly, they showed empirically that an end-to-end model achieved qualitative and quantitative improvements compared to reconstruction with a post-processing DL method. Zhang et al. [297] studied the importance of fully connected layers, commonly used in end-to end model [307, 33], to realize back-projection data from the sensor domain to the image domain and showed that while a back-projection can be learned through neural networks it is constrained by the memory requirements induced by the non-linear number of weights in the fully connected layers. Although several generic DL architectures and loss functions have been explored to further enhance reconstruction results in different ways (resolution, lesion localization, artifact reduction, etc.), a DL-based method inherently remains a black-box that can be hard to interpret. Interpretability is key not only for trust and accountability in a medical setting but also to correct and improve the DL model.

Unrolling Iterative Methods

Unrolling conventional optimization algorithms into a DL model has been suggested by several works [201, 217, 273, 237, 3, 121] in order to combine the benefits of traditional iterative methods and the expressive power of deep models (Table 2.1-Column "E2E"). Rajagopal et al. [203] proposed a theoretical framework demonstrating how to leverage iterative methods to bootstrap network performance while preserving network convergence and interpretability featured by the conventional approaches.

Deep ADMM-Net [237] was the first proposed model reformulating the iterative reconstruction ADMM algorithm into a deep network for accelerating MRI reconstruction, where each stage of the architecture corresponds to an iteration in the ADMM algorithm. In its iterative scheme, the ADMM algorithm requires tuning of a set of parameters that are difficult to determine adaptively for a given data set. By unrolling the ADMM algorithm into a deep model, the tuned parameters are now all learnable from the training data. The ADMM-Net was later further improved to Generic-ADMM-Net [279] where a different variable splitting strategy was adopted in the derivation of the ADMM algorithm

and demonstrated state-of-the-art results with a significant margin over the BM3D-based algorithm [66]. Similarly, the PD-Net [3] adopted neural networks to approximate the proxi-

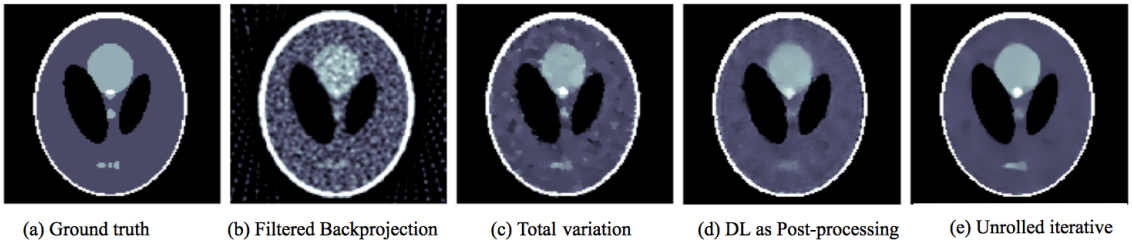

Figure 2.10: Reconstructions of the Shepp–Logan phantom in a two-dimensional CT image reconstruction problem using different reconstruction methods.(a) Reference image. (b)-(c) The standard FBP and TV-regularized reconstruction. (d) FBP reconstruction followed by a U-Net architecture. (e) The learned primal–dual scheme [3]. One can observe the artifacts in the conventional methods removed by the DL-based approaches. Overall, the unrolled-iterative method gave the best results and outperformed even the DL as a post-processing step method. The interested reader is referred to [14] for more details about the training setting. Reproduced with permission from Cambridge University Press.

mal operator by unrolling the primal-dual hybrid gradient algorithm [51] and demonstrated performance boost compared with conventional and two step image reconstruction models. Fig. 2.10 shows results of their proposed method in a simplified setting of a two-dimensional CT image reconstruction problem.

In like manner, Schlemper et al. [217] proposed a cascade convolutional network that embeds the structure of the dictionary learning-based method while allowing end-to-end parameter learning. The proposed model enforces data consistency in the sensor and image domain simultaneously, reducing the aliasing artifacts due to sub-sampling. An extension for dynamic MR reconstructions [218] exploits the inherent redundancy of MR data.

While, the majority of the aforementioned methods used shared parameters over iterations only, Qin et al. [201] proposed to propagate learnt representations across both iteration and time. Bidirectional recurrent connections over optimization iterations are used to share and propagate learned representations across all stages of the reconstruction process and learn the spatio–temporal dependencies. The proposed deep learning based iterative algorithm can benefit from information extracted at all processing stages and mine the temporal information of the sequences to improve reconstruction accuracy. The advantages of leveraging temporal information was also demonstrated in single molecule localization microscopy [49]. An LSTM was able to learn an unbiased emission density prediction in a highly variable frame sequence of spatio-temporally separated fluorescence emissions. In other words, joint learning over the temporal domain of each sequence and across iterations leads to improved de-aliasing.

Semi-supervised and Unsupervised Learning Methods

The majority of deep learning methods for image reconstruction are based on supervised learning where the mapping between signal and reconstructed image is learned. However, the performance of supervised learning is, to a large degree, determined by the size of the available training data which is constrained in a medical setting. In an effort to work around this requirement, Qong et al. [102] applied the deep image prior method [247] to PET image reconstruction. The network does not need prior training matching pairs, as it substitutes the target image with a neural network representation of the target domain but rather requires the patient's own prior information. The maximum-likelihood estimation of the target image is formulated as a constrained optimization problem and is solved using the ADMM algorithm. Qong et al. [103] have extended the approach to the direct parametric PET reconstruction where acquiring high-quality training data is more difficult. However, to obtain high-quality reconstructed images without prior training datasets, registered MR or CT images are necessary. While rigid registration is sufficient for brain regions, accurate registration is more difficult for other regions.

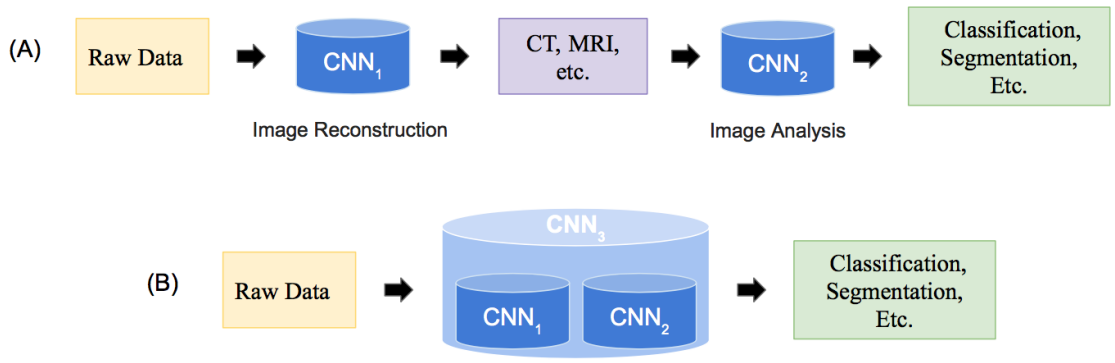

Figure 2.11: Unified Biomedical Image Processing Workflow: Joint Reconstruction and Image Analysis for Enhanced Diagnosis.(A) Biomedical image processing workflow usually involves two steps optimized independently (reconstruction and image analysis) for diagnosis purposes. (B) Jointly solving these tasks using a unified model allows joint parameters tuning and feature sharing.

Instead of using matched high-dose and a low-dose CT, Kang et al. [138] proposed to employ the GAN loss to match the probability distribution where a cycleGAN architecture based cyclic loss and identity loss for multiphase cardiac CT problems is used. Similarly, Li et al. [152] used an unsupervised data fidelity enhancement network that uses an unsupervised network to fine-tune a supervised network to different CT scanning protocol properties. Meng et al. [173] proposed to use only a few supervised sinogram-image pairs to capture critical features (i.e., noise distribution, tissue characteristics) and transfer these features to larger unlabeled low-dose sinograms. A weighted least-squares loss model based TV

regularization term and a KL divergence constraint between the residual noise of supervised and unsupervised learning is used while a FBP is employed to reconstruct CT images from the obtained sinograms.

2.3.3 Raw-to-Task Methods: Task Specific Learning

Typical data flow in biomedical imaging involves solving the image reconstruction task followed by preforming an image processing task (e.g., segmentation, classification) (Fig 2.11-A). Although each of these two tasks (reconstruction and image processing) is solved separately, the useful clinical information extracted for diagnosis by the second task is highly dependent on the reconstruction results. In the raw-to-task paradigm, task results are directly inferred from raw data, where image reconstruction and processing are lumped together and reconstructed image may not be necessarily outputted (Fig. 2.8-C, Fig 2.11-B)

Jointly solving for different tasks using a unified model is frequently considered in the computer vision field, especially for image restoration [216], and has lead to improved results than solving tasks sequentially [196]. The advantages explain the recent attention this approach received in biomedical image reconstruction [238, 124]. For instance, a unified framework allows joint parameters tuning for both tasks and feature sharing where the two problems regularize each other when considered jointly. In addition, when mapping is performed directly from the sensor domain, the joint task can even leverage sensor domain features for further results enhancing while it can be regarded as a task-based image quality metric that is learned from the data. Furthermore, Sbalzarini et al. [216] argues that solving ill-posed problems in sequence can be more efficient than in isolation. Since the output space of a method solving an inverse problem is constrained by forming the input space of the next method, the overall solution space contracts. Computational resources can then be focused on this more limited space improving runtime, solution quality or both.

Image reconstruction is not always necessary for optimal performance. For example, in single-pixel imaging recent work [150] illustrates that an online algorithm can ensure on-par performance in image classification without requiring the full image reconstruction. A typical pipeline in processing compressive sensing requires image reconstruction before classical computer vision (CV) algorithms process the images in question. Recent work [44] has shown that CV algorithms can be embedded into the compressive sensing domain directly, avoiding the reconstruction step altogether. In discrete tomography the full range of all possible projections is not always feasible or available due to time, resource, or physical constraints. A hybrid approach in discrete tomography [193] finds a minimal number of projections required to obtain a close approximation of the final reconstruction image. While an image is still reconstructed after the optimal number of projections is found, determining this optimal set required no reconstructions, a significant advantage compared to an approach where either an iterative or combinatorial selection of reconstructions to determine the optimal set of projections.

2.4 Medical Training Datasets

The performance of learning-based methods is dictated to a large extent by the size and diversity of the training dataset. In a biomedical setting, the need for large, diverse, and generic datasets is non-trivial to satisfy given constraints such as patient privacy, access to acquisition equipment and the problem of divesting medical practitioners to annotate accurately the existing data. In this section, we will discuss how researchers address the trade-offs in this dilemma and survey the various publicly available dataset type used in biomedical image reconstruction literature.

Table 2.1-Columns "Data", "Site" and "Size" summarise details about dataset used by different surveyed papers, which are broadly classified into clinical (real patient), physical phantoms, and simulated data. The sources of used datasets have been marked in the last column of Table 2.1-Columns "Pub". Data" in case of their public availability to other researchers. Since phantom data are not commonly made publicly available, the focus was mainly given to real and simulated data whether they are publicly available or as part of challenges. Used augmentation techniques have been mentioned in Table 2.1-Columns "Aug". Remarkably, augmentation is not always possible in image reconstruction task especially in sensor domain given the non-symmetries of measurements in some case, the nonlinear relationship between raw and image data, and the presence of other phenomena(e.g., scattering). We herein survey the most common source of data and discuss their pros and cons.

2.4.1 Real World Datasets

Some online platforms (e.g., giveascan [163],mridata [180], MGH-USC HCP ([81]), and ukbiobank [38]) made the initiative to share datasets between researchers for image reconstruction task. mridata [180], for example, is an open platform for sharing clinical MRI raw k-space datasets. The dataset is sourced from acquisitions of different manufacturers, allowing researchers to test the sensitivity of their methods to overfitting on a single machine's output while may require the application of transfer-learning techniques to handle different distributions. As of writing, only a subset of organs for well known modalities e.g., MRI and CT are included (Table 2.1-Columns "Site"). Representing the best reconstruction images acquired for a specific modality, the pairs of signal-image form a gold standard for reconstruction algorithms. Releasing such data, while extremely valuable for researchers, is a non-trivial endeavour where legal and privacy concerns have to be taken into account by, for example, de-anonymization of the data to make sure no single patient or ethnographically distinct subset of patients can ever be identified in the dataset. Source of real-word used datasets on surveyed papers has been marked in Table 2.1-Columns "Pub. Data". The dataset sizes remain relatively limited, which can hinder the ability to achieve robust generalization of DL-based methods.

2.4.2 Physics Based Simulation

Physics-based simulation [220, 114, 111] provides an alternative source of training data that allows generating a large and diversified dataset. The accuracy of a physical simulation with respect to real-world acquisitions increases at the cost of an often super-linear increase in computational resources. In addition, creating realistic synthetic datasets is a nontrivial task as it requires careful modeling of the complex interplay of factors influencing real-world acquisition environment. With a complete model of the acquisition far beyond computational resources, a practitioner needs to determine how close to reality the simulation needs to be in order to allow the method under development to work effectively. Transfer learning provides a potential remedy to bridge the gap between real and in silico worlds and alleviates the need for a large clinical dataset [307, 33]. In contrast, the approach of not aiming for complete realism but rather using the simulation as a tool to sharpen the research question can be appropriate. Simulation is a designed rather than a learned model. For both overfitting to available data is undesirable. The assumptions underlying the design of the simulation are more easily verified or shown not to hold if the simulation is not fit to the data, but represents a contrasting view. For example, simulation allows the recreation and isolation of edge cases where a current approach is performing sub-par. As such simulation is a key tool for hypothesis testing and validation of methods during development. For DL-based methods the key advantage simulation offers is the almost unlimited volume of data that can be generated to augment limited real-world data in training. With the size of datasets as one of the keys determining factors for DL-based methods leveraging simulation is essential. Surveyed papers that used simulated data as a training or augmentation data have been marked in Table 2.1-Columns "Pub".

2.4.3 Challenge Datasets

There are only a few challenge (competition) datasets for image reconstruction task e.g., LowDoseCT [243], FastMRI [293], Fresnel [91], and SMLM challenge [119] that includes raw measurements. Low-quality signals can be simulated by artificially subsampling the full-dose acquired raw signal while keeping their corresponding high resolution images pair [102, 274]. While this method offers an alternative source of training data, the downsampling is only one specific artificial approximation to the actual acquisition of low-dose imaging and may not reflect the real physical condition. Hence, performance can be compromised by not accounting for the discrepancy between the artificial training data and real data. Practitioners can leverage techniques such as transfer learning to tackle the discrepancy. Alternatively, researchers collect high-quality images from other medical imaging challenges, e.g., segmentation (MRBrainS challenge [172], MICCAI challenge [34]), and use simulation, using a well known forward model, to generate full and/or incomplete sensor domain pairs.

Here again, only a subset of body scans and diseases for well-studied modalities are publicly available as highlighted in Table 2.1-Columns "Site" and "Pub. Data".

2.5 Reconstruction Evaluation Metrics

2.5.1 Quality

Measuring the performance of the reconstruction approaches is usually performed using three metrics, commonly applied in computer vision, in order to access the quality of the reconstructed images. These metrics are the root mean squared error (RMSE) or normalized mean squared error, structural similarity index (SSIM) or its extended version multiscale structural similarity [263], and peak signal to noise ratio (PSNR).

While RMSE measures the pixel-wise intensity difference between ground truth and reconstructed images by evaluating pixels independently, ignoring the overall image structure, SSIM, a perceptual metric, quantifies visually perceived image quality and captures structural distortion. SSIM combines luminance, contrast, and image structure measurements and ranges between [0,1] where the higher SSIM value the better and SSIM = 1 means that the two compared images are identical.

PSNR (Eq. 2.5) is a well-known metric for image quality assessment which provides similar information as the RMSE but in units of dB while measuring the degree to which image information rises above background noise. Higher values of PSNR indicate a better reconstruction.

$$PSNR = 20 \times \log_{10}\left(\frac{X_{max}}{\sqrt{MSE}}\right), \tag{2.5}$$

where X_{max} is the maximum pixel value of the ground truth image.

Illustrating modality specific reconstruction quality is done by less frequently used metrics such as contrast to noise ratio (CNR) [272] for US. Furthermore, modalities such as US, SMLM or some confocal fluorescence microscopy can produce a raw image where the intensity distribution is exponential or long-tailed. Storing the raw image in a fixed-precision format would lead to unacceptable loss of information due to uneven sampling of the represented values. Instead, storing and more importantly processing the image in logarithmic or dB-scale preserves the encoded acquisition better. The root mean squared logarithmic error (RMSLE) is then a logical extension to use as an error metric [27].

Normalized mutual information (NMI) is a metric used to determine the mutual information shared between two variables, in this context image ground truth and reconstruction [272, 306]. When there is no shared information NMI is 0, whereas if both are identical a score of 1 is obtained. To illustrate NMI's value, consider two images X, Y with random values. When generated from two different random sources, X and Y are independent, yet RMSE(X,Y) can be quite small. When the loss function minimizes RMSE, such cases can induce stalled convergence at a suboptimal solution due to a constant gradient.

NMI, on the other hand, would return zero (or a very small value), as expected. RMSE minimizes average error which may make it less suitable for detailed distribution matching tasks in medical imaging such as image registration where NMI makes for a more effective optimization target [266].

The intersection over union, or Jaccard index, is leveraged to ensure detailed accurate reconstruction [30, 238]. In cases where the object of interest is of variable size and small with respect to the background, an RMSE score is biased by matching the background rather than the target of interest. In a medical context, it is often small deviations (e.g. tumors, lesions, deformations) that are critical to diagnosis. Thus, unlike computer vision problems where little texture changes might not alter the overall observer's satisfaction, in medical reconstruction, further care should be taken to ensure that practitioners are not misled by a plausible but not necessarily correct reconstruction. Care should be taken to always adjust metrics with respect to their expected values under an appropriate random model [90]. The understanding of how a metric responds to its input should be a guideline to its use. As one example, the normalization method in NMI has as of writing no less than 6 [90] alternatives with varying effect on the metric. Table 2.1-Columns "Metrics" surveys the most frequently used metrics on surveyed papers.

2.5.2 Inference Speed

With reconstruction algorithms constituting a key component in time-critical diagnosis or intervention settings, the time complexity is an important metric in selecting methods. Two performance criteria are important in the context of time: throughput measures how many problem instances can be solved over a time period, and latency measures the time needed to process a single problem instance. In a non-urgent medical setting, a diagnosing facility will value throughput more than latency. In an emergency setting where even small delays can be lethal, latency is critical above all. For example, if a reconstruction algorithm is deployed on a single device it is not unexpected for there to be waiting times for processing. As a result latency, if the waiting time is included, will be high and variable, while throughput is constant. In an emergency setting there are limits as to how many devices can be deployed, computing results on scale in a private cloud on the other hand can have high throughput, but higher latency as there will be a need to transfer data offsite for processing. In this regard it is critical for latency sensitive applications to allow deployment on mobile (low-power) devices. To minimize latency (including wait-time), in addition to parallel deployment, the reconstruction algorithm should have a predictable and constant inference time, which is not necessarily true for iterative approaches.

Unfortunately, while some papers reported their training and inference times, (Table 2.1-Columns "Metrics-IS") it is not obvious to compare their time complexity given the variability in datasets, sampling patterns, hardware, and DL frameworks. A dedicated study, out of the scope of this survey, needs to be conducted for a fair comparison. Overall, the

offline training of DL methods bypasses the laborious online optimization procedure of conventional methods, and has the advantage of lower inference time over all but the simplest analytical. In addition, when reported, inference times are usually in the order of milliseconds per image making them real time capable. The unrolled iterative network, used mainly in the MR image reconstruction [237, 279] as the forward and backward projections can be easily performed via FFT, might be computationally expensive for other modality where the forward and backward projections can not be computed easily. Therefore, the total training time can be much longer. In addition when data is acquired and reconstructed in 3D, the GPU memory becomes a limiting factor especially when multiple unrolling modules are used or the network architecture is deep.

2.6 Discussion

Literature shows that DL-based image reconstruction methods have gained popularity over the last few years and demonstrated image quality improvements when compared to conventional image reconstruction techniques especially in the presence of noisy and limited data representation. DL-based methods address the noise sensitivity and incompleteness of analytical methods and the computational inefficiency of iterative methods. In this section we will discuss DL-based methods challenges and illustrate the main trends in DL-based medical image reconstruction: a move towards task-specific learning where image reconstruction is no longer required for the end task, and focus on a number of strategies to overcome the inherent sparsity of manually annotated training data.

Learning: Unlike conventional approaches that work on a single image in isolation and require prior knowledge, DL-based reconstruction methods leverage the availability of large-scale datasets and the expressive power of DL to impose implicit constraints on the reconstruction problem. While DL-based approaches do not require prior knowledge, their performance can improve with it. By not being dependent on prior knowledge, DL-based methods are more decoupled from a specific imaging (sub)modality and thus can be more generalizable. The ability to integrate information from multiple sources without any preprocessing is another advantage of deep neural networks. Several studies have exploited GANs for cross-modal image generation [151, 142] as well as to integrate prior information [149, 36]. Real-time reconstruction is offered by DL-based methods by performing the optimization or learning stage offline, unlike conventional algorithms that require iterative optimization for each new image. The diagnostician can thus shorten diagnosis time increasing the throughput in patients treated. In operating theatres and emergency settings this advantage can be life saving.

Interpretability: While the theoretical understanding and the interpretability of conventional reconstruction methods are well established and strong (e.g., one can prove a method's optimality for a certain problem class), it is weak for the DL-based methods (due

to the black-box nature of DL) despite the effort in explaining the operation of DL-based methods on many imaging processing tasks. However, one may accept the possibility that interpretabilty is secondary to performance as fully understanding DL-based approaches may never become practical.

Complexity: On the one hand, conventional methods can be straightforward to implement, albeit not necessarily to design. On the other, they are often dependent on parameters requiring manual intervention for optimal results. DL-based approaches can be challenging to train with a large if not intractable hyper-parameter space (e.g., learning rate, initialization, network design). In both cases, the hyper-parameters are critical to results and require a large time investment from the developer and the practitioner. In conclusion, there is a clear need for robust self-tuning algorithms, for both DL-based and conventional methods.

Robustness: Conventional methods can provide good reconstruction quality when the measured signal is complete and the noise level is low, their results are consistent across datasets and degrade as the data representation and/or the signal to noise ratio is reduced by showing noise or artifacts (e.g., streaks, aliasing). However, a slight change in the imaging parameters (e.g., noise level, body part, signal distribution, adversarial examples, and noise) can severely reduce the DL-based approaches' performances and might lead to the appearance of structures that are not supported by the measurements [10, 105]. DL based approaches still leave many unsolved technical challenges regarding their convergence and stability that in turn raise questions about their reliability as a clinical tool. A careful fusion between DL-based and conventional approaches can help mitigate these issues and achieve the performance and robustness required for biomedical imaging.

Speed: DL-based methods have the advantage in processing time over all but the most simple analytical methods at inference time. As a result, latency will be low for DL-based methods. However, one must be careful in this analysis. DL-based methods achieve fast inference by training for a long duration, up to weeks, during development. If any changes to the method are needed and retraining is required, even partial, a significant downtime can ensue. Typical DL-based methods are not designed to be adjusted at inference time. Furthermore, when a practitioner discovers that, at diagnosis time, the end result is subpar, an iterative method can be tuned by changing its hyper-parameters. For a DL-based approach, this is non-trivial if not infeasible.

A final if not less important distinction is adaptive convergence. At deployment, a DL-based method has a fixed architecture and weights with a deterministic output. Iterative methods can be run iterations until acceptable performance is achieved. This is a double-edged sword as convergence is not always guaranteed and the practitioner might not know exactly how many more iterations are needed.

Training Dataset: Finally, the lack of large scale biomedical image datasets for training due to privacy, legal, and intellectual property related concerns, limits the application of DL-based methods on health care. Training DL-based models often requires scalable high

performance hardware provided by cloud based offerings. However, deploying on cloud computing and transmitting the training data risks the security, authenticity, and privacy of that data. Training on encrypted data offers a way to ensure privacy during training [97]. More formally a homomorphic encryption algorithm [208] can ensure evaluation (reconstruction) on the encrypted data results are identical after decryption to reconstruction on the non-encrypted data. In practice, this results in an increase in dataset size as compression becomes less effective, a performance penalty is induced by the encryption and decryption routines, and interpretability and debugging the learning algorithm becomes more complex since it operates on human unreadable data.

The concept of federated learning, where improvements of a model (weights) are shared between distributed clients without having to share datasets, has seen initial success in ensuring privacy while enabling improvements in quality [94]. However a recent work by Zhu et al. [308] has shown that if an attacker has access to the network architecture and the shared weights, the training data can be reconstructed with high fidelity from the gradients alone. Data sharing security in a federated setting still presents a concern that requires further investigations.

Simulating a suitable training set also remains a challenge that requires careful tuning and more realistic physical models to improve DL-based algorithm generalization.

Performance: Reconstruction is sensitive to missing raw measurements (false negative, low recall) and erroneous signal (false positive, low precision). Sensitivity in reconstruction is critical to ensure all present signals are reconstructed but high recall is not sufficient as in this setting a reconstructed image full of artifacts and noise can still have high recall. Specificity may be high in reconstructed images, where little to no artifacts or noise is present, but at the cost of omitting information. In the following we discuss how the indicators of both are affected by training regimes and dataset sizes.

Deep learning reconstruction quality varies when using different loss metrics. For instance, when a basic $\mathcal{L}1/L2$ loss in image space is used solely, reconstruction results tend to be worse quality because it generates an average of all possible reconstructions having the same distance in image space, hence the reconstructed images might still present artifacts [224, 30]. Adding feature loss in high dimensional space like perceptual loss and frequency domain loss helps to better constrain the reconstruction to be perceptually more similar to the original image [278, 224, 125]. Furthermore, reconstruction quality varies with dataset size. To analyze the effect of training-dataset size on reconstruction quality, Shen et al. [224] trained their reconstruction model with a variable number of training samples while gradually increasing the dataset size from 120 to 6,000 samples. Interestingly, they reported that while the reconstruction quality improved with the number of training samples, the quality increase trend did not converge which suggests that better reconstruction quality might be achieved if larger datasets are available. Similarly, Sun et al. [240] studied the effect of decreasing the number of measurements and varying the noise level on multiple scattering

based deep learning reconstruction methods and concluded that better feature extraction from higher dimensional space are required for very low sampling cases. In addition, data is usually imbalanced, with the unhealthy (including anomalies) samples being outnumbered by the healthy ones resulting in reduced performance and compromised reconstruction of tumor related structures [267].

While deep networks can take the whole image and produce outputs in a single forward pass, input size can sometimes be a limiting factor as excessive memory cost becomes a challenge for current GPU architectures especially for three-dimensional (3D) images with relatively high resolution. Patch-based training divides the image into possibly overlapping patches and produces outputs for each sub-image in independent forward passes while requiring a reduced amount of memory and time on GPU. For patch-based training, each iteration can take a set of patches (minibatch) for a more stable training where a proper choice of batch size is important for a network convergence and generalization [99]. However, patch-based learning is mostly relevant when training in the same domain, which is the case of deep learning as a pre/post-processing step [59, 271, 274], given the difficulty to map part of the raw signal to a corresponding image patch in end-to-end learning.

Task-specific Learning: Prediction tasks, such as classification or prediction of treatment directly from measurements, that is, omitting the image reconstruction task, is a non-trivial approach that, under the right conditions, can be very worthwhile. Since an explicit encoding of the image is no longer needed, the architecture can be reduced in complexity or redesigned to focus specifically on the task. In addition, focusing on task-specific performance ensures all computational resources are committed to task-specific training rather than expending focus to perhaps unnecessary image reconstruction as an intermediate step. Indeed, as we move towards artificial intelligence based task the field is still undecided if the reconstructed image is required at all for state-of-the-art performance in classification, diagnosis, or treatment prediction as it becomes less important to have a spatio-temporal image representation (i.e. the domain that humans are familiar with). However, one cannot deny the human reliance on visual evidence. Removing the image reconstruction task then can lead to lower interpretability, and perhaps makes it harder for clinicians to trust the outcome. Trust can be regained by increased performance and stability. We observe that from an optimization point of view solving a sequence of two ill-posed problems (reconstruction and task prediction) can be easier compared to solving the entire composite problem at once. When the output of the first ill-posed problem is the input of the second, the search space can be dramatically reduced [216].

Finally, a stumbling block towards adaptation of task-specific training is the lack of standardized datasets. Learning the translation from measurement domain to task domain leads to a field of widely diverse combinations of measurement versus tasks domains, making it less likely to find a standardized challenge dataset offering at least a subset of combinations. The motivation behind task-specific learning can be as simple as a lack of image

training data, forcing the practitioner to remove the image reconstruction task altogether by necessity. An contrasting approach to task-specific learning and image-reconstruction aided learning are multi-output or synthesizing networks. For example, consider MRI where the acquisition device can generate T1 or T2 images, or a combination of both, by changing acquisition parameters. A classification network that expects both as input is then limited in application unless an intermediate network synthesizes the 'missing' images. In other words, rather than omitting the reconstruction, synthesizing networks reconstruct a set of images from a single input [223]. One can then generalize this concept to asking what reconstruction space is ideal for the end task, rather than limiting the method to human interpretable reconstruction (images).

There remain a vast range of challenges and opportunities in the field. So far, most approaches focus on CT and MR image reconstruction while only a handful of approaches exist for the reconstruction of the remaining modalities. Hence, the applicability of deep learning models to these problems is yet to be fully explored. In addition, proposed deep learning architectures are often generic and are not fully optimized for specific applications. For instance, how to optimally exploit spatio-temporal redundancy, or how to exploit multi-spectral data. By addressing these core questions and designing network layers to efficiently learn such data representation, the network architecture can gain a boost in performance and reliability.

2.7 Conclusion

The landscape of medical image reconstruction has witnessed a significant transformation with the recent rise of deep learning techniques. The literature reviewed evidences a notable surge in the adoption of DL-based image reconstruction methods in recent years. This trend has been underscored by their ability to yield remarkable improvements in image quality compared to conventional reconstruction approaches, particularly when confronted with challenges like noisy and limited data representation. As we move forward, it is imperative to address the identified challenges at hand and further leverage the advantages of of DL techniques to continue revolutionizing medical image reconstruction for improved diagnostics and patient care.

Table 2.1: Comparison between different papers that have applied machine learning to the biomedical image reconstruction problem. Mod.: modality; SA.: signal sampling (F: full sampling, SS: sub-sampling, LA: limited angle data, SV: sparse viev, LD: low dose data); MS: multi-spectral input signal; TA: consideration of temporal aspect; D: dimension (PC: point clouds); Site: type of tissue (soft tissue like brain and breast, mix: different part of the body, Abdom: abdominal area, Bio Samp: biological samples, Sim: simulated data); Arch: architecture model (we denote by CNN(s) models that use handcrafted convolutional architectures); E2E: whether the approach is end-to-end or combined (Pos: DL as post-processing, Pre: DL as pre-processing, R2T: Raw-to-task); Loss: used loss function (Comb.loss: combined loss, Perc: perceptual loss, Adver: adversarial loss, CE: cross entropy, Jacc: Jaccard loss, MMD: maximum mean discrepancy); Reg: regularization; Inp: input data (M.Spec Meas: multi-spectral measurements, RF: radio frequencies measurements, PA Meas: photo-acoustic measurements); Out: output data (Chrom Maps: chromophore concentration map, Ref index: refractive index, Img: intensity image, Seg: segmentation mask); Metrics: evaluation metrics (M: MSE or RMSE, P: PSNR, S: SSIM, D: lesion detection metric like Dice and Jaccard, C: contrast to noise ratio and IT: reconstruction inference speed) Aug: used augmentation technique (A: affine, E: elastic, R: rotation transformations); Data: used training data (Sim: simulated data, Clini: clinical data, Phan: phantom data) Pub. Data: public data used.

Reference	Mod.	Year	SA.	MS	TA	D	Site	Arch.	E2E	Loss	Reg.	Inp.	Out.	Aug.	Metrics M|P|S|D|C|IS	Data	Size	Pub. Data
[45]	PA	2018	F	✓		2D	Mix	ResU-Net	✓	MSE		M.Spec Meas	Chrom Maps		✓|✓| -| -| -|✓	Sim	∼2000	
[286]	US	2019	SS			3D	Abdom	Decoder	✓	MSE	L2	RF	Img		-|✓| -| -| -|✓	Clini	500	
[272]	US	2018	F		✓	2D	Mix	CNN	✓	MAE		RF	Elast dist		✓|✓| -| -|✓| -	Sim Phan Clini	∼100	
[110]	CT	2018	SV			2D	Mix	ResU-Net	Pos	Comb. loss		FBP	Img		-|✓|✓|-|-|-	Clini	∼1000	LowDoseCT
[242]	Mult. Scat	2018	F			2D	Bio Samp.	ResU-Net	Pos	MSE		FBP	Ref index		-|✓| -| -| -| -	Sim Phan	1550	Fresnel
[307]	MRI	2018	SS			2D	Brain	Decoder	✓	MSE	L2	k-space	Img	A	✓|✓| -| -| -| -	ImagNet Clini	∼10K	MGH-USCHCP
[56]	CT	2017	LD			2D	Mix	Residual AE	Pos	MSE		FBP	Img	A+R	✓|✓|✓| -| -| -	Sim Clini	∼10K	LowDoseCT
[273]	CT	2018	LA			3D	Mix	Unrolling Iterative	✓	MSE		Sinogram	Img		-|✓|✓| -| -| -	Clini	∼10K	LowDoseCT
[83]	DOT	2019	F			2D	Soft	MLP	✓	MSE		Scattering	Chrom Maps		✓|✓|✓| -| -| -	Sim	∼20K	
[285]	DOT	2019	F			3D	Mix	Decoder	✓	MSE		Scattering	Chrom Maps		-| -| -| -| -| -	Sim Clini	∼1500	
[65]	PET	2019	F			3D	Mix	3DU-Net	Pos	MSE		Sinogram	Img		-| -| -| -|✓| -	Sim Clini	30	
[129]	MRI	2018	SS			2D	Brain	U-Net	Pre	MSE		k-space	Full k-space		✓| -| ✓| -| -| -	Clini	1400	

Reference	Mod.	Year	SA.	MS	TA	D	Site	Arch.	E2E	Loss	Reg.	Inp.	Out.	Aug.	Metrics M|P|S|D|C|IS	Data	Size	Pub. Data
[253]	PA	2018	LV			2D	Vessels	U-Net	✓	MAE		PA Meas	Maps	R	✓| -| -| -| -| -	Sim	3600	
[59]	MRI	2018	SS			2D	Soft	CNN	Pre	MSE		k-space	Img		✓|✓|✓| -| -|✓	Clini	65K	
[157]	CT	2018	SV			2D	Mix	ResNet	Pre	MSE		Sparse Sinogram	Sinog.		✓| -| -| -| -| -	Clini	17K	giveascan
[133]	CT	2017	SV			2D	Mix	U-Net	Pos	MSE		FBP	Img	A	-|✓| -| -| -| -	Sim Clini	1000	
[269]	CT	2017	LD			2D	Thorax cardiac	GAN	Pos	Adver	L2	FBP	Img		✓|✓| -| -| -|✓	Sim Clini	500	
[278]	MRI	2017	SS			2D	Brain	Condi GAN.	Pos	Adver Perc		k-space	Img	A+E	✓|✓| -| -| -|✓	Clini	21K	MICCAI2013
[201]	MRI	2019	SS		✓	3D+t	Cardiac	RCNN	✓	MSE		k-space	Img	A+E +R	✓|✓|✓| -| -|✓	Sim Clini	NA	
[237]	MRI	2016	SS			2D	Brain Chest	Unrolling ADMM	✓	NMSE		k-space	Img		✓|✓| -| -| -|✓	Clini	150	MICCAI2013
[218]	MRI	2018	SS		✓	2D	Cardiac	CNNs	✓	MSE		k-space	Img	A+E	✓|✓| -| -| -|✓	Sim	300	
[183]	SMLM	2018				2D	Sim	Encoder Decoder	✓	MSE	L1	Fluophore emission	SR Img		✓| -| -| -| -|✓	Sim	100K	SMLM
[217]	MRI	2017	SS			2D	Cardiac	CNNs	✓	MSE		k-space	Img	A	✓| -| -| -| -|✓	Sim	300	
[39]	CT	2015	LD			2D	Mix	MLP	Pos	MSE		FBP	img		✓|✓|✓| -| -| -	Clini	100K	
[246]	CT	2018	SV			2D	Mix	wGAN	✓	Adver		LD Sinogram	Img	R+A	✓| -|✓| -| -| -	Clini	NA	
[306]	CT	2019	LA			3D	Mix	GAN	✓	MSE Adver.	Cor coef	LD Sinogram FBP	Img		✓| -|✓| -| -|✓	Clini	5436	LowDoseCT
[30]	DOT	2018	LA			2D	Breast	Decoder	✓	MSE	L2	Scattering	Chrom Maps		✓|✓|✓|✓| -|✓	Phan	4000	
[260]	MRI	2016	SS			2D	Brain	Decoder	Pos	MSE		IFT	Img		-| -| -| -| -| -	Clini	500	
[238]	MRI	2019	SS			2D	Brain	Enc-Dec	R2T	MSE CE		IFT	Img seg	A+E	✓|✓| -|✓| -| -	Clini	192	MRBrainS
[33]	DOT	2019	LA			2D	Breast	Decoder	✓	MSE Jacc.	L2	Scattering	Chrom Maps		✓|✓|✓|✓| -|✓	Phan	20K	
[40]	SMLM	2018	SS			PC	NA	CNN	✓	MMD		Fluophore emission	3D Localiz	A	✓| -| -|✓|✓| -	Sim	100K	SMLM
[112]	PET	2019	F			2D	Sim	Encoder Decoder	✓	MSE		Sinogram	Img	A+R	✓|✓|✓| -| -|✓	Sim	291K	
[189]	MRI	2019	SS		✓	2D+t	Cardiac	3D CNN RCNN	✓	MSE CE		k-space	Img		✓|✓|✓| -| -| -	Sim Clini	300	mridata.org
[203]	CT	2019	SS			3D	Phan	Unrolling Iterative	✓	MSE		Sinogram	Img		✓| -| -| -| -| -	Phan	50K	
[125]	CT	2019	LA			2D	abdomen	U-Net Iterative	✓	MSE		FBP	Artifact Img		✓| -| -| -| -| -	Clini	500	
[259]	DOT	2019	LA			2D	Phan	Stacked AE	✓	MSE		Scattering	Chrom Maps		-|-|-|✓| -|-	Sim	NA	

Chapter 3

End-to-End Deep Learning Based DOT Image Reconstruction

This Chapter focuses on the use of deep learning for image reconstruction in the context of diffuse optical tomography imaging modality. Our primary objective is to address the challenges associated with DOT and introduce a DL-based solution[1]. To achieve this, we begin by providing a comprehensive background on DOT, highlighting its unique characteristics and the difficulties it poses in terms of image reconstruction.

3.1 Introduction

Imaging modalities based on near-infrared light are emerging as tools for biomedical diagnosis, given the non-ionizing nature of infrared light as well as their ability to penetrate a few centimeters into human structures, such as the skull, brain, and breast [67]. The recent progress of optical sensors makes optical-based modalities increasingly attractive. DOT uses near-infrared light to image soft tissues and examination of biological tissue at a macroscopic scale, offering several advantages in terms of safety, costs, portability, and sensitivity to functional changes [12]. This technique has shown great potential in investigating functional brain imaging [53, 302] and breast cancer screening [229, 9].

3.1.1 Background on Diffuse Optical Tomography Imaging

DOT measures the distribution of tissue optical properties as a function of absorption and scattering coefficients. These properties are closely correlated to physiological markers and allow indirect quantitative assessment of tissue malignancy [12, 166]. Indeed, marked variations between healthy and tumor tissue are observed in terms of optical properties and chromophore components (e.g., oxy/deoxy hemoglobin and collagen) [78]. In particular, normal breast tissue and lesions can be separated in terms of optical coefficients at several

[1]This Chapter is based on our published work [30]. Reproduced with permission from Springer Nature.

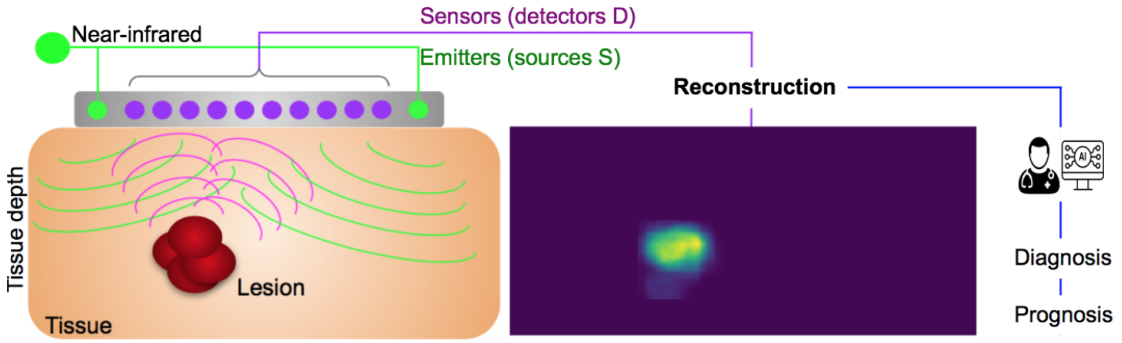

Figure 3.1: DOT image reconstruction workflow example. A DOT-Scan probe consists of two near-infrared sources for tissue illumination and a set of detectors registering the back-scattered photons. Light propagation and scattering in the tissue are schematized. DOT reconstructed image shows the optical coefficients in the tissue. The success of diagnosis and treatment relies on accurate reconstruction and estimation of the optical properties of a medium. © 2021 IEEE.

wavelengths [245, 166]. These properties make DOT a potential promising tool in pre-screening of patients in a clinical setting, saving them from unnecessary exposure to more precise but potentially harmful ionizing modalities such as CT. In such a setting, there is a clear need for both low latency, i.e. method inference speed, preferably real-time, and accurate reconstruction and classification.

The choice of wavelength is crucial in DOT as different wavelengths have varying degrees of penetration depth in biological tissues. Longer wavelengths can penetrate deeper into tissues but may suffer from increased scattering. Shorter wavelengths have shallower penetration but lower scattering. Furthermore, different chromophores, such as hemoglobin, water, and lipids, have varying absorption properties at different wavelengths as shown in Fig. 3.2. By selecting appropriate wavelengths, DOT can target specific chromophores and provide information about tissue composition and oxygenation levels [100]. This wavelength sensitivity is leveraged to analyze optical spectra and reconstruct images of the exposed tissue for diagnostic purposes, given that recovered chromophore concentration changes can convey information about functional brain vascular events, assessing tissue oxygenation, and the characterization and monitoring of breast lesions [76]. In addition, the signal strength and the level of background noise in DOT measurements can also be influenced by the wavelength used.

A DOT scanner is, generally, comprised of an array of emitters and receivers, using low-powered LEDs or lasers, to measure the optical transmission [20] or reflection [229] of light beamed into the tissue at various locations on the tissue surface. DOT projects near-infrared light from one or more light sources and acquires boundary data of the tissue (back-scattered light) using a set of detectors yielding, after reconstruction, cross-sectional

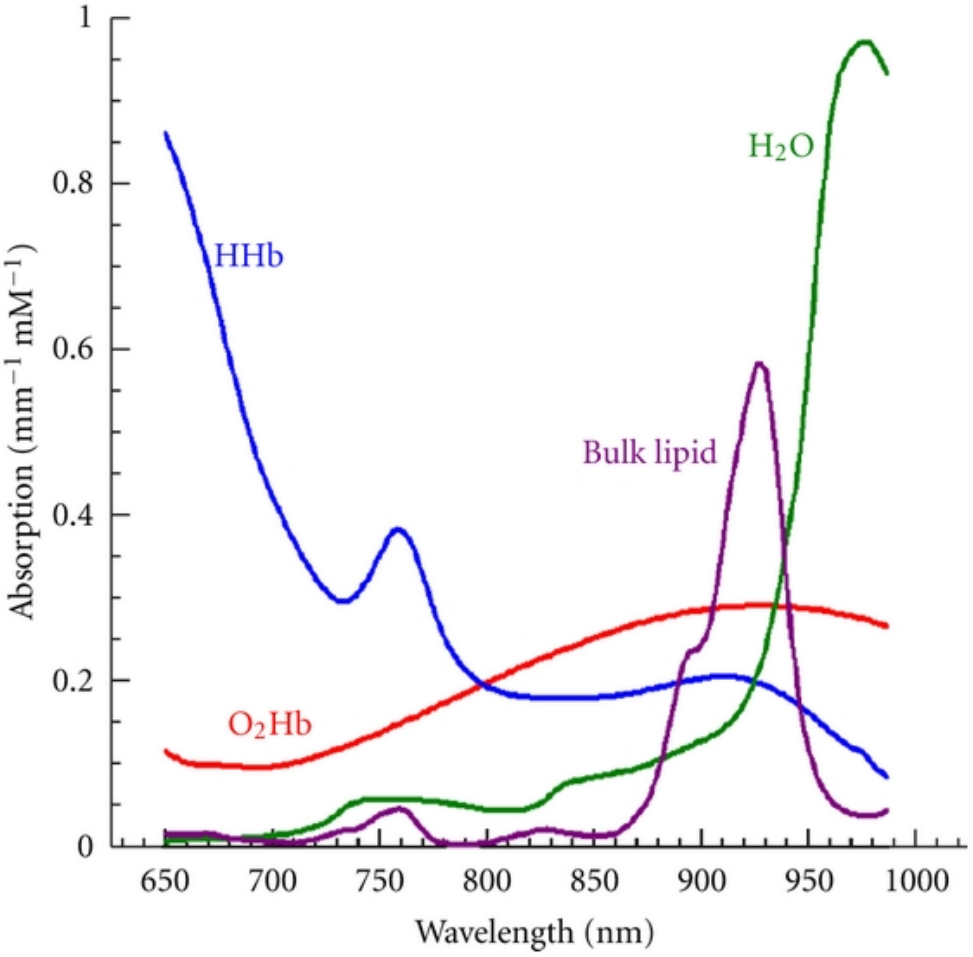

Figure 3.2: The therapeutic optical imaging window, from 600 nm to 900 nm is highly advantageous for in vivo imaging. Hemoglobin and water, which are major absorbers of light in biological tissues, exhibit lower absorption within this window. This enables deeper penetration into biological tissues compared to shorter wavelengths. The reduced scattering in the NIR enables better image quality and signal-to-noise. Reproduced with permission from [118]. Figure licensed under CC-BY 4.0 Creative Commons license.

or volumetric images of tissue [120]. The reconstruction process converts information from the projection (sensor) domain into the image domain to enable interpretation and diagnosis by domain experts or by artificial intelligence (Fig. 3.1). While an optimized probe design enables a reduced hardware complexity and better portability, it increases the complexity of the reconstruction task, especially when the number of sources is limited [282].

A recently developed new functional hand-held diffuse optical breast scanner probe (DOB-Scan) [227], used for real-word data collection in this thesis, has been applied to breast cancer detection as a screening tool and aims to improve the assessment parameters in terms of positive predictive value and accuracy. The probe is currently in clinical tests

for in vivo breast cancer imaging studies. It combines multi-frequency near-infrared light to quantify tissue optical properties in 690 to 850 nm spectra and produces a cross-sectional image of the underneath tissue. The proposed probe uses encapsulated light emitting diodes instead of laser-coupled fiber-optic, which decreases the complexity, size, and cost of the probe while providing accurate and reliable optical properties measurement of the tissue.

3.1.2 Current DOT Image Reconstruction and Challenges

DOT image reconstruction is an inherently ill-posed problem given the nonlinear physics of photon scattering. Although near-infrared light can propagate several centimeters inside the tissue, photons scatter many times and penetrate along random paths through the tissue before reaching the detectors, which makes the reconstruction task difficult. When applied to limited-view data acquisition, allowing reduction of scanner complexity, reconstruction becomes even more challenging (smaller number of measurements compared to the number of unknowns) [15].

The DOT image reconstruction problem has been tackled using a wide variety of reconstruction methods (see [16, 296] for a review). While absolute imaging relies on a single set of measurements to reconstruct spatially distributed optical coefficients, difference imaging aims at recovery of the change in the tissue optical properties based on measurements before and after the change [179]. The reference measurement can be the previous state (e.g., the rest stage in brain DOT) or a reference tissue or phantom. A drawback of difference imaging is that reconstructed images are usually only qualitative [117]. In this manuscript, focus is given to absolute imaging.

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

$$\hat{x}^* = \underset{\hat{x}}{\operatorname{argmin}} \{ \frac{1}{2} \|\mathcal{F}(\hat{x}) - y\|_2^2 + \lambda \mathcal{R}(\hat{x}) \}, \tag{3.1}$$

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

Until recently the focus in DOT reconstruction research has been on model-based algorithms, whose design follows directly from the underlying mathematical problem formulation. A disadvantage of a model based approach is that each instance needs to be optimized independently at reconstruction time, a computationally costly approach that can prohibit real-time application [120]. Reconstruction methods can roughly be divided into

linear approaches, based on inverse scattering theory, and nonlinear iterative approaches, based on model fitting. Linear approaches rely on Born or Rytov approximations, where the linearization is developed from analytical solutions of the diffuse equation for a homogeneous background [120]. This model can be constructed using Green's functions for the diffusion equation. The resultant linear system is usually solved by an iterative method (e.g., conjugate gradients [261]). The Born or Rytov approximation assumes that relative changes induced by anomalies are sufficiently small. When the values of the optical coefficient changes become very large, as in the case of large abnormal lesions, this approximation can lose accuracy in reconstruction [285].

Non-linear methods consider the model in terms of explicit parameters and adjusts these parameters in order to optimize the objective function in (3.1). Methods usually minimize an objective function in an iterative manner until convergence where the distribution of optical properties is searched for by minimizing the difference between the measured and the modelled data. A common approach in iterative reconstruction is to solve the forward model and calculate the Jacobian matrix at each iteration (e.g., Newton-like methods [17]). Alternatively, gradient-based reconstruction techniques use the gradient of the objective function to update the solution offering a reduction in computational complexity, at the cost of slower convergence. Regularization terms have to be added generally as the optimization problems are ill-posed.

Optimization of acquisition time, computational speed along with reconstruction quality remain an active research area in DOT. Despite recent advances, the reconstruction time of conventional DOT reconstruction methods is too high to be suitable for real time application [12]. In this manuscript, we consider throughput of at least 20 images (frames) per second as a minimum real-time equivalent for human experts [289], which corresponds with a refresh rate per image of < 0.05 s. An ideal frame rate of 60 frames per seconds, or 0.016 s per frame, is the minimum sufficient rate to perceive no flickering for normal images [69].

Alternatively, the image reconstruction task can be learned using a deep neural network with a significant off-line training time, up to weeks during prototyping where each model takes a few hours of training, that is offset by a fast inference time as reported in Ben Yedder et al.[31] (chapter 2).

3.1.3 Deep Learning based Methods

Recently, deep learning has emerged as a powerful alternative for biomedical image reconstruction [31, 205, 169] and shown potential in improving resolution accuracy and speeding up reconstruction results especially in the presence of noisy and limited view acquisition.

Convolutional neural networks (CNNs) have previously been applied to medical image reconstruction problems in computed tomography and MRI [56, 217, 169]. Many approaches [133, 39, 185] obtain an initial estimate of the reconstruction using a direct inverse operator or an iterative approach, then use machine learning to refine the estimate and pro-

duce the final reconstructed image, as discussed in Chapter 2. Although this is a straightforward solution, the number of iterations required to obtain a reasonable initial image estimate can be hard to define and in general increases the total reconstruction run-time.

A more elegant solution is to reconstruct an image from its equivalent projection data directly by learning all the parameters of a deep neural network, in an end-to-end fashion and therefore, approximates the underlying physics of the inverse problem. In [307], a unified framework for image reconstruction that allows a mapping between sensor and image domain is proposed. A pre-trained CNN model is used to learn a bidirectional mapping between sensor and image domains where image reconstruction is formulated in a manifold learning framework. The trained model is tested on a variety of MRI acquisition strategies.

While deep learning based image reconstruction has been applied to a variety of medical imaging modalities, they have not yet been used for DOT. In this Chapter, we propose a deep DOT reconstruction method to learn a mapping between raw acquired measurements and reconstructed images. The raw collected data can be considered as image features that approximate nonlinear combinations of image pixel values, which form the desired tissue optical coefficients. Therefore, the raw measured data is a nonlinear function of the desired image pixels values and so performing image reconstruction amounts to learning to invert this nonlinear function. We propose to use deep neural networks to learn, from training data, this nonlinear inverse mapping.

3.1.4 Contributions

We make the following contributions in this Chapter:

(i) We propose the first end-to-end deep learning model for limited angle near-infrared-based optical imaging.

(ii) We leverage a physics-based optical diffusion simulation to generate in silico training datasets and evaluate results on real measurements on physical phantoms collected with the diffuse optical probe (DOB-Scan) [227, 231].

(iii) We train the reconstruction model using in silico data only and test it on real measurements acquired from a phantom dataset subject to sensor non-idealities and noise. We show the utility of our synthetic data generation technique in mimicking real measurements and the generalization ability of our model to unseen phantom datasets.

In section 3.2, we introduce our proposed method for end-to-end reconstruction, the mathematical formulation behind, and detail physics-based optical diffusion training simulation. Model performance results on physical phantoms are presented in Section 3.4. We conclude the Chapter by discussing insights and limitations in Section 3.5. The performance of our proposed system shows that our framework improves reconstruction accuracy when compared against a baseline analytic reconstruction approach.

3.2 Methodology

Our main goal is to reconstruct tomographic images from corresponding sensor-domain sampled data or measurements. To this end we collect training measurements from (a) synthesized tissue geometries with known optical properties using a physics-based simulation of the forward projection operation, and (b) data collected using the probe on physical phantoms. We describe the generation of synthetic training and phantom test datasets as well as the design of the neural network architecture below.

3.2.1 Physics based Training Data for DOT Reconstruction

Our aim here is to create training data pairs in-silico, which include image of optical tissue property and its corresponding detectors measurement. The deep learning model will then be trained to generate the image from the measurement. A set of synthetic 2D images with different lesion shapes, sizes, and locations, representing optical tissue properties, discretized into finite element nodes (2D triangular meshes), and their corresponding forward projection measurements were synthesized (Fig. 3.3). We then assign to these geometries optical transport parameters (absorption and scattering coefficients) similar to real human breast tissue and lesion distribution values [96].

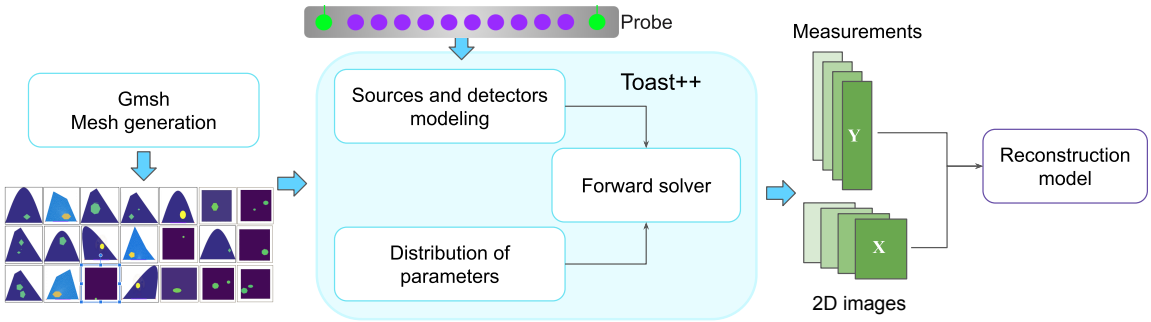

Figure 3.3: Workflow of silico training data generation for DOT image reconstruction. 2D triangular meshes modeling tissue with different lesion sizes and depths were generated using the Gmsh software [93]. The Toast++ software suite was used to simulate the forward projection with a 750 nm wavelength modulation and optical properties approximating breast tissue. Functional handheld probe characteristics were modeled to obtain measurements that mimic real world values. Measurements (y) and ground truth images (x) pairs were saved and fed to the deep learning model for training. © 2021 IEEE.

The time-resolved optical absorption and scattering tomography software (Toast++) [221] was used to simulate the frequency-domain forward model and generate projection measurements for each training mesh. Frequency-domain models or intensity-modulated sources use a frequency ω for modulating the intensity of light. Frequency-domain systems provide information on the medium in the form of amplitude and phase data that permits the recovery of absorption and scattering coefficient parameters. The forward model uses the finite ele-

ment method to solve the diffusion equation that models the light propagation in a diffusive medium. The diffusion equation is given by [221]:

$$-\nabla \cdot D(r)\nabla\varphi(r,\omega) + \left(\mu_a(r) + \frac{i\omega}{c_m(r)}\right)\varphi(r,\omega) = 0, \quad r \in \Omega\backslash\partial\Omega \tag{3.2}$$

with boundary condition [221, 115]:

$$\varphi(\mathbf{m},\omega) + 2\zeta(c_m)D(m)\frac{\partial\varphi(\mathbf{m},\omega)}{\partial\nu} = q(\mathbf{m},\omega), \quad \mathbf{m} \in \partial\Omega \tag{3.3}$$

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

Our functional hand-held probe geometry was modeled accurately in Toast++ to obtain measurements that mimic real values obtained by the DOB probe [227, 229]. The probe is composed of two symmetrical multi-wavelength light-emitting diode (eLED) sources, illuminating the tissue symmetrically and delivering near-infrared light to a body surface at different points, and featuring 128 co-linear equi-spaced detectors (a row) that measure the back-scattered light from the tissue and emitted from the boundary. Each eLED source consists of four different near-infrared (NIR) wavelengths: 690 nm, 750 nm, 800 nm, and 850 nm. These encapsulated light-emitting diodes serve as pencil beam sources, emitting light into the scattering medium. Positioned 0.5 mm apart from each other, the eLEDs function as multi-wavelength pointed illumination sources within the tissue. Correspondingly, the output of the forward model simulation captures a 1D raw intensity diffraction ($1\times$ 256) vector y^s resulting from the scattering of the illuminating light exiting the object. Both LED and all detectors are colinear as depicted in Fig. 3.4.

The absorption coefficients of breast tissue vary between 0.0018 and 0.0025 mm^{-1} for healthy tissue and between 0.09 and 0.16 mm^{-1} for cancerous tissue. Similarly, the scattering coefficients vary between 0.8 and 1 mm^{-1} for healthy tissue and 1 and between 1.6 mm^{-1} for cancerous tissue. Given that relative optical parameters changes at healthy tissue are sufficiently small compared to optical perturbations at abnormalities, background optical parameters were assumed uniform in each sample (the absorption coefficient of background is set to the average of breast tissue similar to [245, 106]) while abnormalities' size, depth, number, and their respective optical parameters were varied. The size of the reconstructed image is 6×6 cm with a resolution of 128×128 pixels.

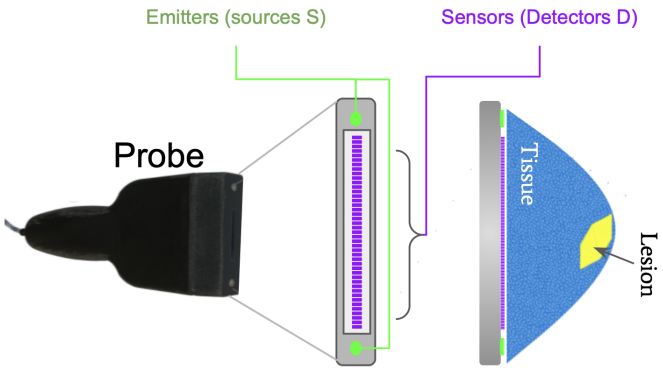

Figure 3.4: The spatial distribution of the simulated sources and detectors matching the layout of the physical probe (left). A sample synthetic mesh is also shown (right).

3.2.2 Phantom Test Dataset

To create physical phantom datasets we rely on a tissue-equivalent solution where an intralipid solution is used to mimic background breast tissue due to its similarity in optical properties to breast fat [228, 87]. Measurements are collected with the DOB-Scan probe. In order to mimic cancerous lesions, a tube with 4 mm cross-sectional diameter was filled with a tumor-like liquid phantom (Indian black ink solution) and was placed at different locations inside the intralipid solution container. The flowchart of synthetic and phantom data acquisition procedures are shown in Fig. 3.5 (Left side).

3.2.3 Reconstructing Images from DOT Measurements

By passing an input measurement through a set of nonlinear transformations one can reconstruct the equivalent image. The proposed architecture consists of a dense layer followed by a set of convolution layers which are designed to efficiently combine features from the first layer with those of deeper layers. The architecture of our proposed model is shown in Fig. 3.5 (right side).

Initial Image Estimate: A fully connected layer, with a ReLu activation, is used as the first layer of the network in order to map the measurement vector to a two-dimensional array that will serve as an initial image estimate. This layer is first pre-trained then included in the deeper architecture including convolutional layers. The goal we seek to achieve using the fully connected layer is to generalize a back projection (BP) operation by learning a weighted combination of the different receptive sensors based on the signal collected from scattered light emitted at different locations in the reconstructed tissue. Empirically we did not observe any improvements in the reconstruction results using more than one fully connected layer. This may be related to the size of the input measurement which is only 256

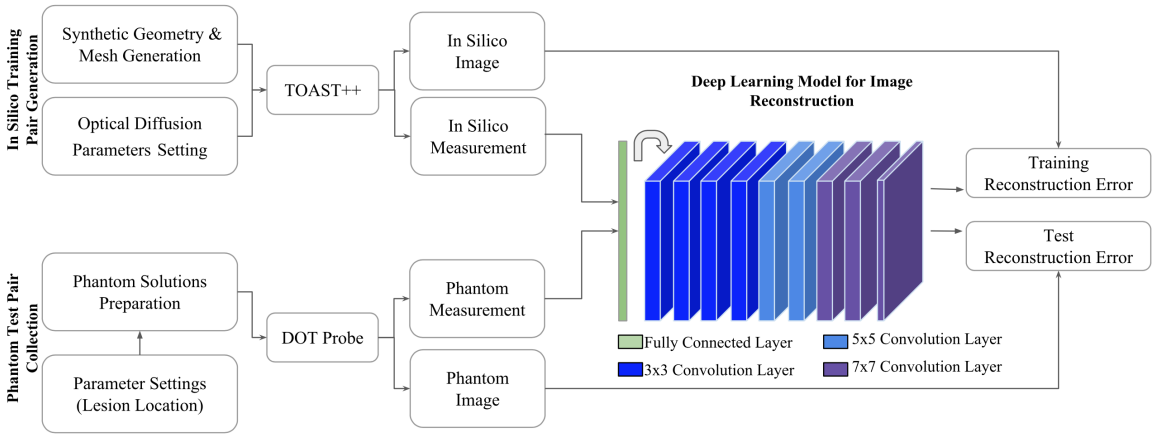

Figure 3.5: In silico training pairs generation using TOAST++ and phantom test pairs collection using DOT-probe are depicted on the left. The overall architecture of the proposed model is shown on the right, where the arrow after the first fully connected layer represents the reshaping procedure before the convolution layers.

dimensional in our dataset. Higher dimensional inputs may benefit from additional layers.

Convolutional Layers: A set of convolutional layers, with 64 channels, are used to refine the first image and produce the final reconstruction image. The non linear ReLU activation and zero-padding are employed at each convolution layer. All feature maps produced by all convolutional layers are set to size 128×128. The size of the convolution filters is increased gradually to cover a larger receptive field at deeper layers and capture local spacial correlations. Details of the architecture are shown in Fig. 3.5.

Integration Layer: The integration layer is a convolutional layer with 7×7 kernel size and a single output channel. It is used to reduce features across the channels from the penultimate layer of the CNN model into a single channel. The output of this layer is the reconstructed image.

3.2.4 Mathematical Model Formulation

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

layer is computed using:

$$o_j^{fc} = g\left(\sum_{i=1}^{N} w_{ij}^{fc} y_i + v_j^{fc}\right), \tag{3.4}$$

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

$$\begin{aligned} o^{cn}[m,n] &= (o^{cn-1} * h)[m,n] \\ &= \sum_i \sum_k h[i,k] o^{cn-1}[m-i, n-k], \end{aligned} \tag{3.5}$$

where the input image is denoted by $o^{cn-1}[m,n]$, kernel by h and m and n are indexes of rows and columns, respectively. We show in the silico reconstruction, Fig. 3.6, that consecutive convolution layers are very important to recover the accurate size and location of the optical anomalies.

3.3 Implementation

We trained the model by minimizing the mean squared error between the reconstructed image and the ground truth synthetic image. We used an L^2 norm penalty on the last convolutional layer output as it facilitates training (i.e. we observed faster convergence using regularization). The model was implemented in Keras and trained for a total of 2,000 epochs on an Nvidia Titan X GPU using batch gradient descent with momentum. The learning rate was set to 0.001 and we used a learning decay of 1e−6, momentum was set to 0.9. All training hyper-parameters were optimized via grid search on a validation set. We sequentially trained the model to first reconstruct an image using the fully connected layer only, then we fine-tuned the entire architecture after including the different convolution layers (Fig. 3.5).

Note that the model was only trained on synthetic data and we kept the phantom data for evaluation only, as depicted in Fig. 3.5. In total, we generated 4,500 synthetic training images and their corresponding simulated DOT measurements and tested our model in 2000 synthetic DOT measurements then in 32 phantom real probe measurements with corresponding ground truth images.

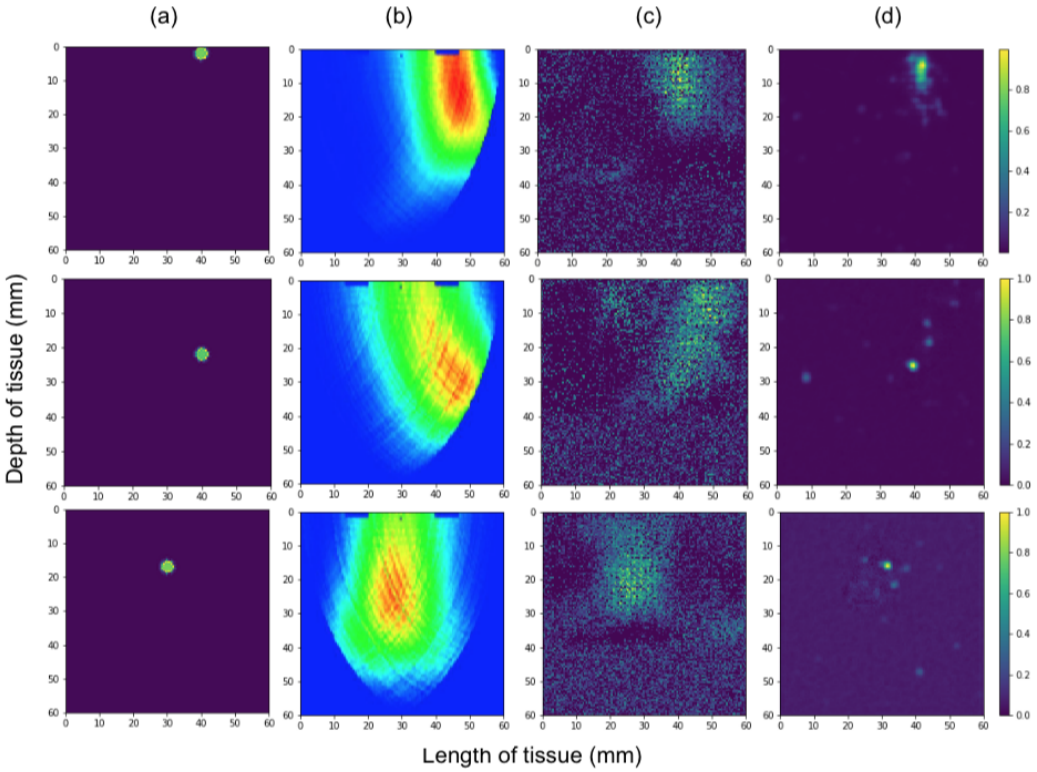

Figure 3.6: Qualitative reconstruction performance of our model compared to conventional techniques. (a)-(d): Ground truth; analytic approach results; generalized BP with one fully connected layer only; and proposed model results.

3.4 Experiments and Results

We compared our results with those obtained by the analytic reconstruction approach described in [227]. Briefly, the analytic method is based on comparing the collected measurement to the measurement of a tissue-equivalent solution with homogeneous value. The resulting difference is then used to perform filtered back-projection and to estimate the spatial location of the lesion.

3.4.1 Qualitative Results

Once trained using the generated synthetic data, our model was tested on the phantom dataset. In Fig. 3.6, we visually compare our proposed reconstruction method to the analytic approach results for phantom cases. Evidently, the images reconstructed by our method are more accurate than those reconstructed by the more conventional analytic approach, when tested on data with a known ground truth. In Fig. 3.6 we show the reconstructed image using only the first fully connected layer which is equivalent to the filtered back-projection operation. Our qualitative results show that reconstructions obtained with one

Table 3.1: Quantitative results scores on 32 phantom test measurements

	Loc. Error (pixel (in mm) , < is better)	MSE (< is better)	PSNR (dB, > is better)	SSIM (> is better)	Jaccard (> is better)	Runtime (ms, < is better)
Analytic approach	77.4 ±32.2 (17 ± 7.1)	0.06 ±0.05	15.08 ±6	0.32 ±0.26	0.5 ±0.19	83.3
Proposed model	33.2 ±23.4 (7.1 ± 5.1)	0.02 ±0.03	20.1 ±4.6	0.46 ±0.28	0.85 ±0.07	7.3

fully connected layer (third column in Fig. 3.6) are on par with reconstructions obtained with the analytic approach (second column in Fig. 3.6).

3.4.2 Quantitative Results

In order to measure the quality of the results, we consider the mean square error as well as the pixel-wise distance between the centre of the lesions in the ground truth image versus the reconstructed image (Loc. Error) where each pixel is 0.22 mm isotropic. The peak signal to noise ratio (PSNR), the SSIM similarity measure, and the Jaccard index (intersection over union) are also calculated. The Jaccard index, used for comparing the similarity and diversity of sample sets, is the ratio of area of overlap between detected and ground truth lesion to the area of their union. This metric is computed after thresholding the reconstructed image to obtain a binary mask where foreground pixels correspond to pixels with highest optical coefficient.

Table 3.1 shows the results for the phantom dataset. This experiment also allows us to evaluate the quality of the synthetic dataset we generated by testing how well a model trained only on synthetic data generalizes to unseen physical phantom images. Results reported in Table 3.1 show that the proposed approach is able to generalize well to the phantom dataset and achieves better performance than the baseline analytic approach in terms of distance (+50%), Jaccard index (+35%), similarity score (+14%) and PSNR (+5db). The high standard deviation in distance metric is mainly due to samples with deep lesion (lesion location $\geq$ 30 mm) since as the lesion depth increases it becomes harder to differentiate the signal from the tumor-free tissue signal. On average, our model achieves an order of magnitude faster reconstruction than the baseline analytic approach.

3.5 Conclusion

This work represents a step forward for both image reconstruction in DOT and the use of machine learning in bio-imaging. We present the first model that leverages physics based forward projection simulators to generate realistic synthetic datasets and we model the inverse problem with a deep learning model where the architecture is tailored to accurately reconstruct images from DOT measurement. We test the method on real acquired projection measurements subject to sensor non-idealities and noise. Results show that our method improves the quality of reconstructed images and shows promising results towards

real-time image reconstruction. Our next phase in this research is to improve further the reconstruction accuracy and lesion localization while reducing artifacts.

Chapter 4

Tackling DL Image Reconstruction Artifacts in Limited-Angle DOT

In the preceding Chapter, the effectiveness of employing deep learning in real-time reconstruction of DOT images was shown. In this Chapter, we focus on addressing model artifacts that were observed in the previous contributions[1]. Dealing with the inverse problem in a scenario of limited-view data acquisition is challenging, particularly when noise is present. This difficulty arises from the highly diffusive characteristics of light propagation in biological tissues and the sparsity of the recovered information.

4.1 Introduction

Most DOT inverse problems consider a circular shape scanner with 16 or more point sources uniformly distributed along the field of view boundary to maximize the number of measurements thereby improving spatial resolution, especially in strongly scattering media. Most recently, a multi-layer perceptron network (MLP) was used to solve DOT image reconstruction problem using high source count [83]. However, increasing the number of sources and detectors adds complexity to the DOT scanner hardware and increases manufacturing cost and computational resources.

One common limitation of existing reconstruction methods is that they perform poorly on data with a very low number of point sources (limited projection data), limited-angle acquisition (e.g, acquisition from one view), or both [227].

Sun et al. [242, 240] address the multiple scattering problem of microwaves in biological samples. They study the effect of decreasing the number of sources, to a limited extent (to a minimum of 20 sources), on deep learning based reconstruction methods for weak and strong scattering scenarios. While their proposed reconstruction model leverages the rich data representation collected from 20 up to 40 point sources, it relies on a computationally

[1]This Chapter is based on our published work [33]. Reproduced with permission from Springer Nature.

expensive analytic reconstruction step to provide a first estimate of the reconstructed image, prohibiting real time inference. Furthermore, their deep model is not optimized in an end-to-end manner.

Limited-angle and limited sources DOT image reconstruction in a strong scattering medium is a challenging task that has been considered in end-to-end fashion for a functional hand-held probe in a clinical trial by Ben Yedder et al. [30]. Yet, their results still suffer from noisy reconstruction and deviation of the reconstruction lesion compared to the ground truth location.

To address the aforementioned limitations, in this Chapter, we propose a deep learning DOT reconstruction method based on a novel loss function and transfer learning to solve the limited-angle and limited sources DOT image reconstruction problem in a strong scattering medium. By adaptively focusing on important features and filtering irrelevant and noisy ones using the Fuzzy Jaccard loss, our network is able to reduce false positive reconstructed pixels and as a result reconstruct more accurate images.

Training machine learning based methods requires a high number of training samples, a challenging requirement in a medical setting, especially with relatively new imaging devices like DOT probes. Synthetic data simulators can provide an alternative source of training data. However, creating a realistic synthetic dataset is a challenging task as it requires careful modelling of the complex interplay of factors influencing real world acquisition environment. A potential remedy is to attempt to bridge the gap between real word acquisition and synthetic data simulation via transfer learning. To the best of our knowledge, this is the first work to employ a Jaccard based loss and transfer learning to the DOT reconstruction problem.

4.2 Methodology

Given pairs of measurement vectors y and their corresponding ground truth image x, our goal is to optimize the parameters θ of a convolutional neural network in an end-to-end manner to learn the mapping between the measurement vector y and its reconstructed tomographic image x, which recovers the optical parameters of underlying imaged tissue. Therefore, we seek the inverse function $\mathcal{F}^{-1}(\cdot)$ that solves:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}\left(\mathcal{F}^{-1}(y,\theta), x\right) + \lambda \mathcal{R}(\mathcal{F}^{-1}(y,\theta)) \tag{4.1}$$

where $\mathcal{L}$ is the loss function of the network that, broadly, penalizes the dissimilarity between the estimated reconstruction and the ground truth. We use an L2 regularization term ($\mathcal{R}$).

The proposed architecture, which extends Yedder et al.'s [30] FCN architecture, Chapter 3, with a transfer learning and novel loss components, is a decoder-like network that consists of a fully connected layer followed by a set of residual layers. The fully connected

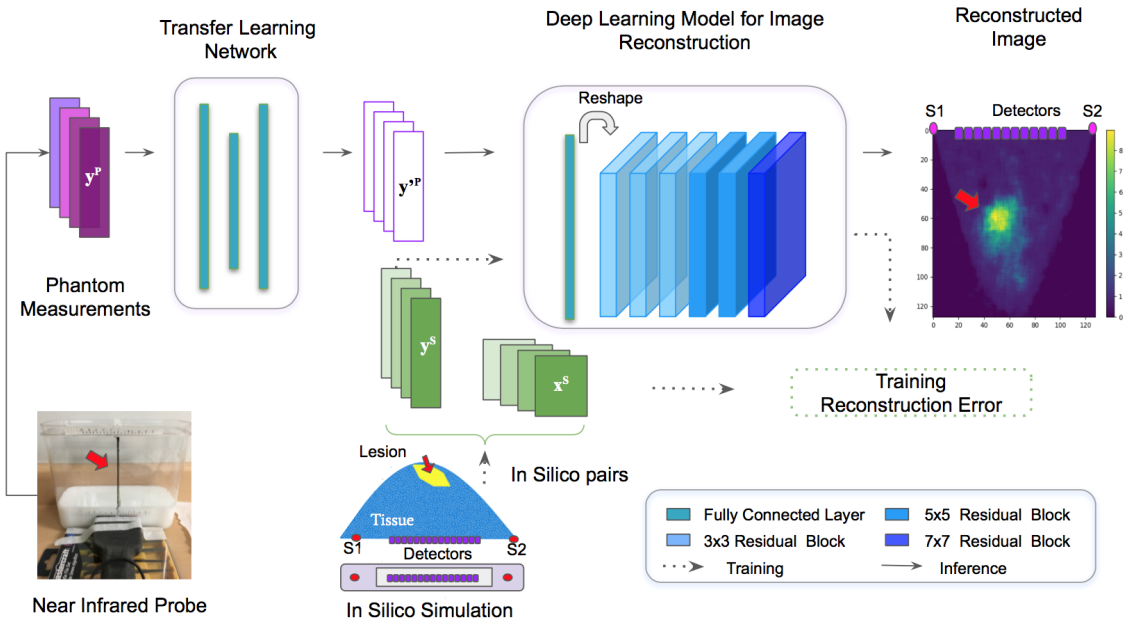

Figure 4.1: Overall architecture of the proposed model: Transfer learning (tackling domain shift) and reconstruction network integration. (lower left) The probe (in black) is positioned to image a phantom (white) with an embedded synthetic lesion (red arrow). The transfer learning network (multilayer perceptron) maps phantom measurements y^p to the domain of in silico training measurements y^s. The mapped measurements are passed through the reconstruction network to produce the reconstructed image $\hat{x}^*$ (rightmost image).

layer maps the measurement vector to a two-dimensional array and provides a coarse image estimate, while the subsequent residual blocks refine the image estimate by passing it through a set of nonlinear transformations to produce the final reconstruction image. The architecture of our proposed model is shown in Fig. 4.1, where each residual block uses convolutions with batch normalization and ReLU.

4.2.1 Similarity-wise Loss Function

To address DOT image reconstruction from a limited information representation (one view with few sources), we propose a novel loss function, $\mathcal{L}$, that dynamically combines two loss terms:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \beta(epoch)\mathcal{L}_{\text{FJ}} \tag{4.2}$$

where $\mathcal{L}_{\text{MSE}}$ is the mean squared error (MSE) loss, which focuses on pixel-wise similarity. $\mathcal{L}_{\text{FJ}}$ is similarity coefficient based fuzzy Jaccard term designed to promote lesion location and appearance similarity while penalizing artifacts. β is a hyper-parameter balancing the two terms and varies with the training epochs to capture the dynamics of this interaction. In particular, $\beta(epoch + \Delta) = \beta_0 + \gamma\beta(epoch)$ with $\gamma > 0$, which allows the network to learn to first reconstruct, via $\mathcal{L}_{\text{MSE}}$, an image estimate that is relatively close to the ground truth image pixel wise distribution and then, via $\mathcal{L}_{\text{FJ}}$, gradually refine that candidate image. In DOT image reconstruction of a breast tissue with zero or more isolated lesions, the majority of the pixels are background, $\mathcal{L}_{\text{FJ}}$ is chosen to address this imbalance. Further, $\mathcal{L}_{\text{FJ}}$ does not require binary values and accounts for the similarity between the foreground as well as the background pixel values. Finally, a log transform of $\mathcal{L}_{\text{FJ}}$ ensures a steep convex gradient,

$$\mathcal{L}_{\text{FJ}} = -\log\left(\frac{\sum_{i=1}^{n}\min(a_i, a_i')}{\sum_{i=1}^{n}\max(a_i, a_i')} + \epsilon\right) \tag{4.3}$$

where the $min(\cdot,\cdot)$ and $max(\cdot,\cdot)$ functions compute a probabilistic intersection and union, respectively while setting $\epsilon = 10^{-5}$ avoids $\log(0)$ domain errors.

4.2.2 Cross Sensors Learning Transfer

In medical imaging settings, transfer learning [80] can be used to bridge the gap between simulated and clinical data by transferring knowledge learned from simulated data to improve the performance of models on clinical data [287]. Domain adaptation deals with the situation where the distribution of data in the source domain is different from the distribution of data in the target domain. The goal is to adapt the model's knowledge from the source domain to perform well in the target domain despite the distribution shift. This is particularly important in medical imaging and relatively new imaging devices such as DOB probes, where obtaining large quantities of annotated clinical data can be challenging and

expensive. As training on real world data is limited by availability of samples, we resort to generating artificial training data via a simulator.

A transfer learning network, implemented as a multilayer perceptron, tackles the domain shift between the real data measurement y^p, as collected from the probe and used during inference, and the in silico data measurement y^s used during training time (Fig. 4.1-upper left). By minimizing a loss $\mathcal{L}_{TL}$, the transfer learning network learns to translate the real world data distribution onto the in silico data distribution while avoiding overfitting on the in silico model. Finally, by retraining or fine-tuning this transfer learning network only, our proposed approach can be generalized to new DOT sensors and or source configurations.

Given the i-th phantom x_i^p we simulate its x_i^s tissue equivalent and derive the corresponding sensor measurements y_i^s, while we collect y_i^p using a physical probe. By minimizing $\mathcal{L}_{TL}$ over N_p phantom experiments, the transfer learning module learns the mapping $\phi(y_i^p) \approx y_i^s$ to ensure it is in the same domain as y^s,

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}_{\mathrm{TL}}(\theta) \quad \text{where} \quad \mathcal{L}_{\mathrm{TL}}(\theta) = \sum_{i=1}^{N_p} ||\phi(y_i^p; \theta) - y_i^s|| \tag{4.4}$$

where a final test reconstructed image is computed as, $\hat{x_i}^* = \mathcal{F}^{-1}(\phi(y_i^p))$.

4.3 Experiments and Results

We compare our proposed approach to the state of the art FCN architecture for limited angle data [30] and the aforementioned MLP approach [83], as well as the analytic reconstruction approach described by Shokoufi et al. [227]. In addition, we evaluate the individual contributions of the terms of our loss function and the transfer learning.

4.3.1 Dataset

To train our network $\mathcal{F}^{-1}(\cdot)$ we use in silico training data pairs (x^s, y^s) as designed in previous Chapter (Chapter 3, Section 3.2.1). It includes images x^s of optical tissue properties, and their corresponding forward projection measurements y^s Simulated lesions have varying sizes, locations, and optical coefficients. Similarly, the phantom test-set based on a tissue-equivalent solution (Chapter 3, Section 3.2.2) was used. All phantom measurements y^p are collected with the DOB-probe.

4.3.2 Implementation

The model was implemented in the Keras framework and trained for a total of 1,000 epochs on an Nvidia Titan X GPU using the Adam optimizer. By optimizing the model's performance on the validation set, we set all hyper-parameters as follows: Batch size to 64; learning rate to 0.001; AMS Grad optimizer set to true; and ($\Delta = 10$, $\beta_0 = 0.2$, $\gamma = 0.002$),

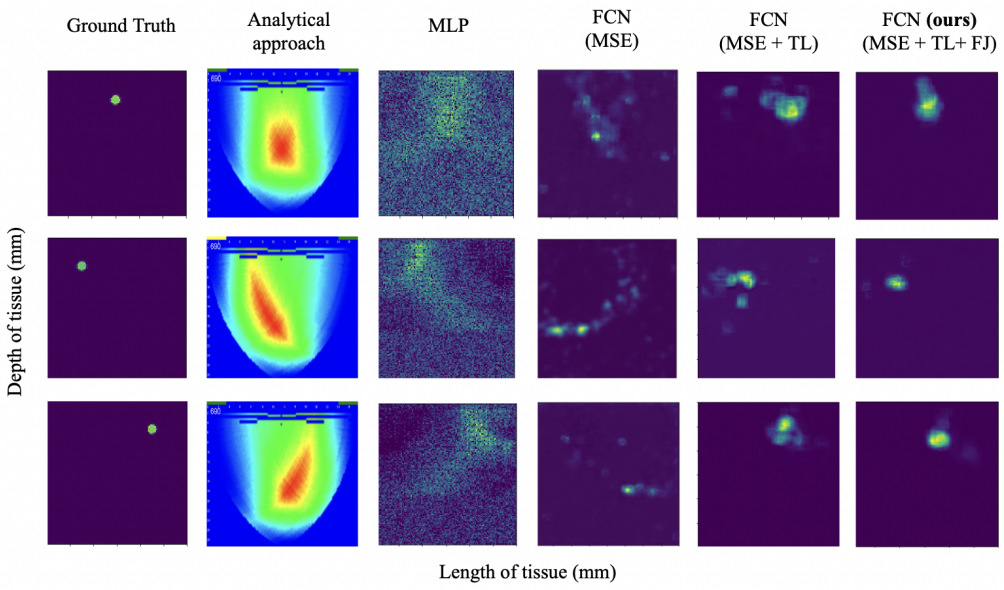

Figure 4.2: Qualitative reconstruction performance of our model compared to state of the art techniques on phantom samples with known lesion ground truth locations. The parabolic shape of the reconstruction produced by the analytical approach is due to the algorithm used.

which describe the update equation of the hyper-parameter β in (5.2); We use a 80/10/10 training/validation/test split of the in silico data.

4.3.3 Qualitative Results

Our model is trained on the in silico data and tested on the phantom dataset. In Fig. 4.2, we visually compare our proposed reconstruction method to the competing methods' results on sample phantom cases. As mentioned earlier, the transfer learning module, tackling domain shift, maps the real world distribution onto the learned in silico distribution. Without such mapping, unsurprisingly, we notice artifacts in the reconstructed image; note the extensive scattering of false positives with different scales and locations (Fig. 4.2 - FCN (MSE)). Adopting transfer learning clearly reduces these artifacts (Fig. 4.2 - FCN (MSE+TL)). Further, observe how incorporating both the new loss term $\mathcal{L}_{FJ}$ and transfer learning module significantly reduces the artifacts and improves lesion localization, which otherwise could compromise diagnosis (Fig. 4.2 - FCN (MSE+TL+FJ)).

While MLP showed good performance in the complete information case, namely a circular shape scanner with 16+ uniformly distributed point sources [83], it underperforms on the limited angle experiments (Fig. 4.2 - MLP). We hypothesize that this difference in performance is due to the convolution operators' ability to extract comprehensive contextual information and synthesize more complex robust features.

Table 4.1: Quantitative results on 32 phantom experiments.

	Loc. Error (pixel (in mm), < is better)	PSNR (dB, > is better)	SSIM (> is better)	Fuzzy Jaccard (> is better)	Runtime (ms, < is better)
Analytic approach [227]	77.4 ± 32.2 (17 ± 7.1)	15.0 ± 6.0	0.32 ± 0.26	0.17 ± 0.15	83.3
MLP [83]	42.0 ± 17.3 (9.3 ± 3.8)	12.5 ± 1.5	0.05 ± 0.03	0.07 ± 0.05	1.2
FCN (MSE) [30]	33.2 ± 23.4 (7.1 ± 5.1)	20.1 ± 4.6	0.46 ± 0.28	0.32 ± 0.06	7.3
FCN (MSE+TL)	16.6 ± 6.60 (3.6 ± 1.4)	20.6 ± 0.4	0.61 ± 0.17	0.45 ± 0.08	11.5
Proposed	**14.8** ± 7.40 (3.2 ± 1.6)	**21.7** ± 0.9	**0.73** ± 0.03	**0.64** ± 0.10	16.9

4.3.4 Quantitative Results

We measure the reconstruction quality via: (i) Lesion localization error, i.e. the distance between the centre of the lesions in the ground truth image versus the reconstructed image; (ii) peak signal to noise ratio (PSNR); (iii) structural similarity index (SSIM); and (iv) the Fuzzy Jaccard [64]. All reconstructed images were first normalized prior to calculating the performance metrics.

Table 4.1 presents the results on the phantom dataset. Using the transfer learning $\phi(\cdot)$ module, we observe ∼10% improvement in Fuzzy Jaccard and ∼16% in SSIM compared to state of the art FCN with MSE only. Adding the new $\mathcal{L}_{FJ}$ loss term boosts the improvement in these two metrics further to ∼34% and ∼33%, respectively. The lesion localization error is also considerably reduced when using transfer learning and $\mathcal{L}_{FJ}$.

4.4 Conclusion

We proposed novel extensions to the deep learning based diffuse optical tomography image reconstruction method presented in previous Chapter. We have shown empirically that our model, trained with the novel hybrid loss function, attains superior quantitative results on multiple evaluation metrics and, qualitatively, improves the reconstructed images, showing fewer artifacts that could compromise clinical diagnosis. The transfer learning module, tackling domain shift, renders an in silico trained network applicable to real world data. More importantly our approach is decoupled from a change in real world measurements and can be generalized to new source configurations. Our next phase in this research is to improve further the lesion localization and validate our approach on real patient data to assess its diagnostic accuracy.

Chapter 5

Multitask Deep Learning for Joint Reconstruction and Localization

In the preceding Chapters, we presented techniques focused on the *reconstruction* of optical properties in breast tissue. Nonetheless, incorporating *localization*, which aims to identify specific regions within the images that exhibit characteristics suggestive of breast lesions, can enhance the accuracy of the reconstruction. To achieve this, we adopt a multitask learning approach, simultaneously addressing DOT-reconstruction and lesion localization in breast cancer tissue[1].

5.1 Introduction

5.1.1 DL-based DOT Reconstruction Methods

Deep neural networks-based techniques have shown promising potential in solving the inverse scattering problem [242, 137, 84, 285]. Preliminary work [242, 137, 249], did not consider end-to-end models but rather two-step approaches where a first image estimate, approximated using conventional methods, is further enhanced using a deep learning model. Recent end-to-end models typically considered a full-view acquisition setup: a circular shape scanner with more than 16 point sources uniformly distributed along the boundary to maximize the number of measurements. For instance, Sun et al.[242] addressed the multiple scattering problem of microwaves in biological samples using a two-step reconstruction method, where an analytical method providing the first image estimate is followed by a convolutional neural network (CNN) for image reconstruction. Feng et al.[84] proposed a multilayer perceptron (MLP) feed-forward artificial neural network for a full-view 2D-DOT image reconstruction. However, performance degrades significantly for limited-angle acquisition where the number of sources is smaller and their placement is co-linear. Recently,

[1]The contents of this Chapter is largely based on our published work © 2021 IEEE Reprinted, with permission, from [282].

based on the Lippman-Schwinger equation and deep convolutional framelets for inverse problems [281], Yoo et al.[285] designed a deep learning approach for 3D inverse scattering problems. The proposed model is based on an autoencoder CNN for a 3D-DOT inverse problem to learn the nonlinear physics of the inverse scattering problem.

Preliminary versions of our work in Chapter 3 and 4 [30, 33] tackled reconstruction of biological tissue parameters from limited-view data acquisition in a strong scattering medium using a decoder-like fully convolutional neural network architectures (FCN and FCN-FJ). In Chapter 4 [33], a weighted combination of mean squared error (MSE) and Fuzzy Jaccard were used as loss function. Models were tested on phantom data only. While advancements have been made, reconstructions are still plagued by artifacts (e.g., false positive reconstructed pixels) and deviations from the ground truth location of the lesion, limiting their practical clinical utility. Furthermore, accurate localization is not necessarily guaranteed by a 'good' reconstruction, especially given the often small ratio of lesion versus background. For accurate diagnosis (e.g., via lesion biopsy) and treatment planning, the location and size of the lesion are paramount.

5.1.2 Multitask Learning

Jointly solving for different tasks using a unified model is frequently considered in the computer vision field and has been shown to improve results in different applications, e.g., taskonomy [291] and skin-multitask [252]. Zamir et al.[291] showed that models optimized to jointly predict several related tasks perform better than models trained on individual tasks separately. Similarly, Kawahara et al.[139] proposed a single model capable of performing multiple tasks on multimodal skin lesion classification while Abhishek et al.[2] reported a performance improvement in skin cancer lesion management prediction when incorporating multitask learning. Sbalzarini [216] asserted that solving ill-posed problems in sequence can be more efficient and easier than in isolation as the overall solution space contracts since jointly solved problems constrained each other's input/output space. Arguably, a unified framework allows feature space sharing and joint parameters tuning for both tasks and where the two problems regularize each other when tackled jointly. In our application as well, we illustrate that a multitask approach leads to a better reconstruction and localization, enabling more accurate diagnosis.

5.1.3 Contributions

We make the following contributions in this Chapter:

(i) To the best of our knowledge, we are the first to leverage reconstruction guided localization multitask learning for DOT reconstruction. We leverage a spatial attention mechanism in combination with a distance transform based loss function. Our network, Multitask-Net, exploits deep layers of spatial-wise attention [122, 211] to attend to important features by filtering irrelevant and noisy ones. Adaptively re-weighting features,

according to their inter-dependencies in feature space, improves the representation ability of our model. Crucially, the distance transform based loss improves lesion reconstruction and localization.

(ii) We leverage a physics-based optical diffusion simulation to generate in silico training datasets and evaluate results on real measurements on physical phantoms and clinical data collected with the diffuse optical spectroscopy probe for breast tissue (DOB-Scan) [227, 231]. We evaluate the generalization ability of our model to unseen data subject to sensor non-idealities and noise.

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

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

In section 5.2, we introduce our proposed method for multitask reconstruction and localization learning followed by domain adaptation learning. Physics-based optical diffusion simulation test dataset, and competing methods are detailed in Section 5.3. Trade-offs of network characteristics and evaluation on various datasets are described in Section 5.4. In silico performance results are presented in Section 5.4.1, results on physical phantoms in Section 5.4.2, and results on real-world data in Section 5.4.3. We conclude the Chapter by discussing insights and limitations on lesion attention-directed reconstruction in DOT in Section 5.5.

5.2 Methodology

5.2.1 Reconstructing Images from DOT Measurements

We extend our previously proposed decoder-like architecture FCN [30] and FCN-FJ [33] consisting of a fully connected layer followed by a set of convolution layers, with a spatial-wise attention mechanism and a new loss component (Fig. 5.1-C). The fully connected layer maps the measurement vector y onto a two-dimensional array by learning a weighted combination of the different receptive sensors based on the signal collected from the back-scattered light collected at different locations in the reconstructed tissue, and therefore provides an initial image estimate. Empirically, we did not witness any improvements in the reconstruction results using more than one fully connected layer. This may be due to the size of the input measurement vector, which is only 256×1 dimensional in our dataset. Higher-dimensional inputs may benefit from additional layers. The output of this layer is reshaped to an 128×128 array and delivered to the subsequent layers for optimization. Then, a set of residual convolutional blocks (6 blocks), with 32 channels and increasing filters size (3, 5, 7), refines the first image via nonlinear transformations and produces the

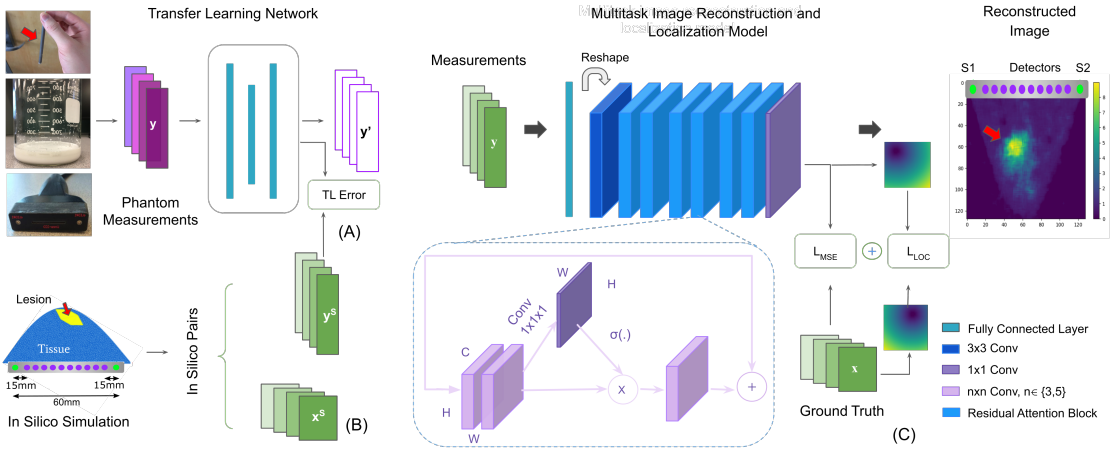

Figure 5.1: The overall architecture of the deep residual attention model. (A) Transfer learning network, tackling domain shift similar to Chapter 4, is used for data distribution adaptation between in silico measurements y^s and real data collected from the probe y. (B) The generation of in-silico training data pairs (ground truth images x^s and their corresponding forward projection measurements y^s) approximating our probe's specification [229]. The red arrow represents the simulated lesion. (C) Architecture of the proposed deep residual attention model. The squeeze and excitation block (inset) squeezes a given feature map U along the channels and excites spatially. The output of the sigmoid layer $\sigma(.)$ represents the relative importance of the spatial information (i, j) of the feature map U. All feature maps produced by all convolutional layers (conv) are set to size 128×128 using (3×3) followed by (5×5) kernels. The similarity-wise loss L_{loc}, leveraged to enforce localization accuracy, computes the mean absolute error between the distance transform of the reconstructed and ground truth images.

final reconstruction image x. Yoo et al.[285] adopted a similar architecture and theoretically justified the reason behind it based on the analysis of deep convolutional framelets [281]. While the mean squared error was used in isolation in our preliminary FCN work [30], a Fuzzy Jaccard loss component $\mathcal{L}_{\mathrm{FJ}}$ was added later FCN-FJ [33] to address the imbalance between the minority pixels corresponding to one or more lesions versus a majority of 'background' pixels corresponding to healthy tissue.

5.2.2 Spatial-wise Attention Network

The framework of our proposed residual attention network (RAN) is shown in Fig. 5.1-C. The first and last convolutional layers are a shallow feature extractor and a reconstruction layer, respectively. The middle layers are residual attention blocks used to extract hierarchical attention-aware features [255, 122]. Since the contributions of features to the reconstruction task vary from different feature maps, we leverage a spatial attention mechanism to prune the irrelevant features and enhance the informative ones by adaptively re-weighting features according to their inter-dependencies in feature space. The spatial attention compresses the feature map U along the channels and excites it spatially to guide the network to pay more attention to the regions of interest; lesion localization in the case of DOT reconstruction. The attention mechanism has been shown effective for natural language processing [280, 251] and computer vision problems tasks [122, 211] as it selectively disregards the noise (less important information) and focuses on what is relevant. Empirically, we did not observe any improvements in the reconstruction results using channel-wise attention [122]. Each residual block uses sequential modules of the form: two convolutions – batch normalization – ReLU and zero-padding followed by a squeeze and excite modules [211].

The spatial squeeze operation is achieved through a convolution generating a projection tensor $q = W \times U$ with weight $W \in \mathbb{R}^{1\times1\times C}$ where each element of the tensor $q \in \mathbb{R}^{H\times W}$ represents the linearly combined representation for all C channels. The projection is then passed through a sigmoid layer σ to rescale activations to [0, 1], which are thereafter used to re-calibrate U spatially:

$$F_{SA}(\mathbf{U}) = \left[\sigma\left(q_{1,1}\right)\mathbf{u}^{1,1}, \cdots, \sigma\left(q_{i,j}\right)\mathbf{u}^{H,W}\right], \tag{5.1}$$

where $\sigma(q_{i,j})$ represents the relative importance of a spatial information (i, j) of a given feature map. As a result, the focus is guided to the lesion location at the cost of loss of attention to healthy tissue, a sacrifice justified given their relative lower importance to the clinician when diagnosis is the end task.

5.2.3 Localization-wise Loss Function

To guide the learning of the RAN network, we propose a novel loss function that combines two loss terms: a mean square error that guides the loss to reconstruct the pixel-wise repre-

sentation of the image and a location based term to strengthen lesion localization accuracy:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \beta \mathcal{L}_{\text{loc}}\,, \tag{5.2}$$

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

Reconstruction Loss

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

Localization Loss

The location-based loss term $\mathcal{L}_{\text{loc}}$ is introduced to emphasize more accurate lesion localization. $\mathcal{L}_{loc}$ is based on the Euclidean distance transform that maps an image to a similar grayscale image, except that the grayscale intensities of points inside the background region are changed to show the distance to the closest foreground boundary from each point. The distance transform is sensitive to small changes in size and position of the foreground object (e.g., offsetting an object by a single pixel can generally change all pixel values of the distance transform image). We leverage this sensitivity to penalize deviation in lesion localization in the reconstruction. $\mathcal{L}_{loc}$ calculates the $\mathcal{L}1$ distance between the distance transform of the ground truth and the reconstructed image pairs (Eq. (5.3), Fig. 5.1-C) and goes to zero when the reconstructed image overlaps with the ground truth. The location-based distance transform loss is given by:

$$\mathcal{L}_{loc}(\theta) = ||DT(\mathcal{F}^{-1}(y, \theta), x) - DT(x)||, \tag{5.3}$$

where

$$\text{DT} = \begin{cases} 0 & \text{if } a_{ij} \in S \\ D(a_{ij}, S) & \text{if } a_{ij} \in R_S\,, \end{cases} \tag{5.4}$$

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

5.2.4 Domain Learning Module

Training a deep learning based model can require a large dataset to converge to optimal performance. However, this is a challenging requirement with relatively new imaging devices such as DOB probes. Physics-based simulation data can provide an alternative source of training samples. Similar to Ben Yedder et al.[33] and Zhou et al. [306], where sensor data distribution adaptation is used to take advantage of cross-domain learning, we leverage transfer learning (Fig. 5.1-A) to tackle domain shift and bridge the gap between real word acquisition and in silico simulated data as proposed in Chapter 4.

5.3 Datasets

To train our tomographic image reconstruction network, we collect training samples from synthesized tissue geometries with known optical properties using a physics-based simulation. We use our recently developed functional hand-held diffuse optical breast scanner (DOB-probe) [227] to collect real data of physical phantoms and from patients for testing our method. No phantom or patient data is used during training and validation.

5.3.1 In silico Training Data

In silico training data pairs, which include images of optical tissue properties and their corresponding detector measurements, as designed in Chapter 3 (Section 3.2.1) were used to train our reconstruction network (Fig. 5.1-B).

A total of 4000 samples of (2D mesh, 256-D vector) data pairs are used where abnormalities' size, depth, number, and their respective optical parameters were varied. Since the absorption coefficients in the lesion are larger than the background, the absorption features due to the lesion emerge in outgoing light at the boundary. Therefore, similar to [28], we focused our in silico dataset on examples that have been shown to be hard to reconstruct and more difficult in treatment.

The in silico data is designed to be consistent with existing prior art [37, 285, 84], where we vary diameters of two-dimensional circular heterogeneous masses representing lesions. We vary the radius from 3-15 mm and tune optical coefficients of background to match that of biological non-lesion tissue. We randomize the number of lesions (0-3), as well as their size, depth, and location to ensure a sufficiently rich dataset matching real world conditions. While simulated data can in principle be infinite in size, we found the current size of our in

silico dataset to be sufficient for our purposes, in that we observed a decreasing return in performance with respect to increasing training time from increases in data set size.

A Gaussian noise layer is used before the dense layer to add independent zero-centered Gaussian noise to each measurement y_i. This noise model simulates the highly varying noise to each individual detector as caused by interference of refracting light as well as sensor noise. We choose $\sigma = 10\%$ of maximum signal value consistent with previous work [285, 84, 242]. For completeness, we note that in addition to this in silico noise, the probe accounts for ambient noise, recorded by capturing a frame without active emitters, which is then subtracted from real data measurement (phantom and clinical) before reconstruction.

5.3.2 Phantom and Clinical Test Data

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

Phantom Data

To analyze the performance of our approach in controlled real experiments setting, breast-mimetic phantoms with known inhomogeneity locations, as described in Chapter 3 (Section 3.2.2), are used (Fig. 5.1-A). The DOB-probe (Fig. 5.1-B) collects phantom measurements y^p. By varying the location of this simulated lesion, we can test the accuracy of the reconstruction method in a number of different scenarios.

Clinical Data

The data is collected from two participants diagnosed with breast cancer, following ethics and institutional review board approval protocol, and subsequently de-identified to protect patient privacy. For each patient, 2D images were scanned from two locations, the first is a sweep over the cancerous lesion location and the second is a sweep over the same region on the opposite healthy breast, as a baseline. Figure 5.5-B illustrates the scan procedure. On average, five measurements were taken on each sweep. The precise location of the cancerous lesion was determined via mammography, ultrasound, showing calcification over the area, and biopsy.

5.3.3 Implementation

The model was implemented in the Keras [60] framework and trained for a total of 500 epochs, on an Nvidia Titan X GPU, with early stopping if the validation loss has not improved in the last 10 epochs. Adam optimizer with default parameters is used [145]. The optimal hyper-parameters such as the number of the filter channels, non-linearity, weights initialization, regularization term, and the batch size were found by trial and error to be optimal for our validation dataset during prototyping.

By optimizing the model's performance on the validation set, we set all hyper-parameters as follows: batch size to 64; learning rate to 10^{-4}; Adam optimizer set to true; and ($\Delta = 10$, $\beta_0 = 0.2$, $\gamma = 0.005$), which describe the update equation of the hyper-parameter β in (5.2); We use a 80/10/10 training/validation/test split of the in silico training data. The model was trained for the first 100 epochs using $\mathcal{L}_{\text{MSE}}$ before introducing $\mathcal{L}_{\text{loc}}$ to avoid noisy reconstruction. A $\mathcal{L}2$ weights' regularization of 10^{-4} was applied to the last residual layer of the network to avoid overfitting to training data. Transfer learning was applied at evaluation time in the phantom and real data while being trained using phantom measurements only. A separate phantom dataset was used to train the transfer learning network.

5.3.4 Competing Methods

We contrast our method results to eight state-of-the-art methods; four of which are conventional methods whereas the others are deep learning based methods. We compare to Least squares-based approach (LS) [70], Levenberg-Marquardt (LM) [221] and nonlinear conjugate gradient (NCG) optimatization based methods [18], and analytical method [229]. The deep learning based methods are: MLP [84], FC-CNN [285], and our preliminary works FCN [30], and FCN-FJ [33]. Inverse solvers are constructed by making use of first derivatives such as NCG-based or second derivatives such as LM-based approaches. Employed algorithms are implemented in the state of the art public software packages of NIRFAST for LS and TOAST++ software for LM and NCG methods.

LS [70] and LM [221] are Jacobian-based approaches. The LM based method employs the modified Levenberg Marquardt (LM) [222], a trust region approach, to minimize the objective function over the search space of model parameters. At each iterative step the Jacobian matrix is recalculated and repeatedly inverted. Optical parameter estimates are then updated. The algorithm converges when the reconstructed optical parameter at the current iteration has not improved over the last two iterations. A first order Tikhonov penalty term is used to regularize reconstruction algorithms. The regularization parameter, found by parameter sensitivity study, is set to 0.01.

NCG is a gradient based method [18]. The objective function describes the difference between the calculated forward measurement and measured data. The inverse solver iteratively computes updates to an initial distribution of the absorption and scattering parameters via minimization of the gradient of the objective function. Total variation (TV) is used to regularise the solution and the object threshold parameter β, found by parameter sensitivity study, is set to 0.1. The finite element discretisation method is used to discretise the computational domain and break down the calculation domain onto the local elements individually. In this Chapter, object geometry is represented by a polygon over which a series of disjoint triangles are generated using the GMesh software suite [93]. We used Isotropic TV regularization [161, 75], as implemented in the Toast++ software suite [221].

The analytical method in [229] is a difference approach that relies on a relative change between optical perturbations at abnormalities and healthy tissue. A homogenous data is collected using a phantom and used as a reference to measure any inhomogeneities in tissue. The difference in absorption between the reference (normal) and abnormal tissues can provide the imaging contrast for tissue diagnostics. Hypothetical curved paths, as explained in [229], are used to recover the absorption coefficient from each illuminated source. The relative change difference is back scattered, using a closed form solution for the diffusion equation, to recover the optical coefficients. Hence, the absorption coefficient at each coordinate (x, y) is calculated as a superposition of back scattered values along the paths of illuminated light from each source. The computed absorption coefficients are then used to recreate two-dimensional cross-sectional images for each wavelength. For the detailed mathematical model description, we refer the reader to [229, 231].

The MLP method [84] is a multilayer perceptron network with two fully connected layers. The FC-CNN [285] uses a fully connected layer followed by a CNN with an encoder-decoder structure with three convolutional layers. The hyperbolic tangent function (tanh) is used as an activation function for the fully connected layer and two first convolutional layers, and rectified linear unit (ReLU) for the last convolutional layer [285]. The FC-CNN implements a similar architecture to our baseline approach FCN [30] where the difference lies in the number of convolution layers, filter, and kernel size. Both architectures were re-implemented as specified in the introducing papers and trained using mean square error between input and output as the loss function. Our preliminary works FCN [19], and FCN-FJ [20] were described in Chapters 3 and 4.

5.3.5 Conventional Algorithms Hyper-parameter Selection and Settings

To validate the conventional algorithms in a quantitative manner and select the best reconstruction parameters, we conducted a study using numerical phantoms and reported results in Table 5.1. Reference mesh coarseness was set to 1.5 mm iin all experiments as we noticed that decreasing it only resulted in longer reconstruction time with negligible to no improvement. The maximum number of iterations was set to 100 epochs for NCG and 30 for LM. Algorithms are optimized over different values of the regularization parameter (λ), TV object threshold parameter (β), and the control parameter (τ) (LM algorithm only). The following optimized parameter values were used for the remainder of the test: $\lambda = 1e - 6$, $\beta = 0.1$, and $\tau = 0.01$.

5.4 Results

In order to measure the quality of the results, we consider: (i) Lesion localization error, i.e. the distance between the centre of the lesions in the ground truth image versus the reconstructed image where each pixel is 0.22 mm isotropic.; (ii) peak signal-to-noise ra-

Table 5.1: Competing algorithm are evaluated under different regularization techniques and hyper-parameters on 32 images from the numerical phantoms

	Reg	Parameters	Loc. Error (pixel, < is better)	PSNR (dB, > is better)	SSIM (> is better)	Fuzzy Jaccard (> is better)
LM	TV	$\beta = 1e-2$, $\lambda = 1e-6$	31.6 ± 14.5	15.8 ± 4.5	0.20 ± 0.02	0.12 ± 0.05
		$\beta = 1e-1$, $\lambda = 1e-6$	31.0 ± 14.3	17.8 ± 5.6	**0.39 ± 0.05**	**0.17 ± 0.03**
		$\beta = 1e-2$, $\lambda = 1e-4$	37.3 ± 23.3	**19.3 ± 11.3**	0.19 ± 0.01	0.09 ± 0.01
	L1	$\tau = 1e-4$, $\lambda = 1e-6$	30.9 ± 17.4	12.6 ± 5.8	0.32 ± 0.12	0.14 ± 0.07
		$\tau = 1e-2$, $\lambda = 1e-6$	**25.1 ± 12.7**	16.6 ± 8.6	0.38 ± 0.09	0.16 ± 0.05
		$\tau = 1e-2$, $\lambda = 1e-4$	40.9 ± 20.3	10.6 ± 9.7	0.20 ± 0.00	0.80 ± 0.01
		$\tau = 1e-2$, $\lambda = 1e-2$	28.2 ± 21.9	10.6 ± 7.3	0.39 ± 0.20	0.18 ± 0.14
NCG	TV	$\beta = 1e-2$, $\lambda = 1e-6$	35.7 ± 19.9	13.5 ± 8.5	0.13 ± 0.02	0.08 ± 0.02
		$\beta = 1e-2$, $\lambda = 1e-4$	37.5 ± 21.4	12.0 ± 8.6	0.14 ± 0.02	0.08 ± 0.02
		$\beta = 1e-1$, $\lambda = 1e-6$	**32.6 ± 16.5**	**16.5 ± 9.2**	**0.33 ± 0.01**	**0.22 ± 0.13**
		$\beta = 1e-0$, $\lambda = 1e-6$	37.4 ± 21.5	12.2 ± 8.7	0.14 ± 0.01	0.10 ± 0.02
		$\beta = 1e-1$, $\lambda = 1e-4$	35.8 ± 19.9	14.2 ± 8.4	0.14 ± 0.01	0.13 ± 0.10
	L1	$\lambda = 1e-3$	40.5 ± 20.8	8.39 ± 8.4	0.15 ± 0.01	0.09 ± 0.03
		$\lambda = 1e-4$	41.8 ± 22.7	12.7 ± 6.0	0.15 ± 0.01	0.09 ± 0.03
		$\lambda = 1e-6$	34.1 ± 19.3	13.6 ± 7.4	0.13 ± 0.01	0.12 ± 0.08

tio (PSNR, $10\log_{10}\frac{\text{MAX}_I}{\text{MSE}}$, where MAX_I is the maximum intensity value of all pixels); (iii) structural similarity index (SSIM) [262]; and (iv) the Fuzzy Jaccard [140].

5.4.1 Results on Synthetic Data

Trained on the in silico data and tested on a separate test set of 80 images, we visually evaluate the reconstruction of our Multitask-Net method and compare results with those of LS [70], NCG [18], MLP [84], and FC-CNN methods [285]. In Fig. 5.2, we show results on some in silico samples with different lesion sizes and depths. Interestingly, while both LS and NCG methods are able to detect lesion presence, LS outperforms in terms of lesion depths and size reconstruction. Images reconstructed by the MLP method provide a quite noisy reconstruction and a rough approximation of the lesion location. Further, filtering via CNN enabled better results by FC-CNN. Overall, our model provides more accurate results as it is able to differentiate between different lesion sizes and recover correctly lesion location and dimension and absorption coefficients despite the limited angle condition. The y-axis scales highlight our model's superiority in reconstructing the optical coefficient values. The difference in performance can be explained by the ability of the convolution operators to extract more comprehensive contextual information and synthesize more complex features.

The last two columns in Fig. 5.2 (e-f) show reconstruction performance in a healthy tissue case as well as a deep lesion case (∼5 cm in depth). Note the model's ability to capture and reconstruct a tumor-free signal. We illustrate the limits of our contribution where deep lesions cannot be distinguished from the background. This is expected given the rapidly reducing strength of reflected signals due to photons scattering many times in random paths before being absorbed by the tissue. A recent study [148] on self-diagnosis

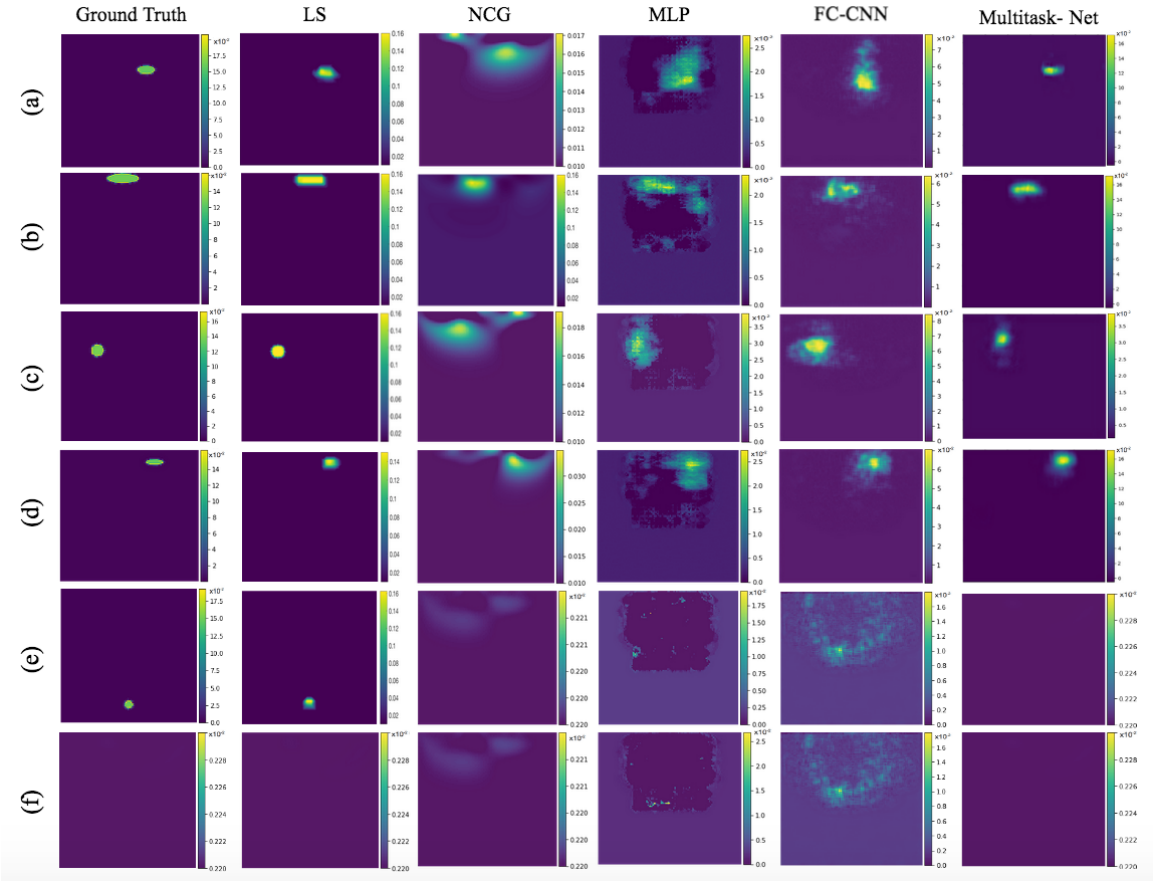

Figure 5.2: Qualitative reconstruction performance of our Multitask-Net model compared to LS, NCG, MLP, and FC-CNN models in silico samples with various ground truth lesions sizes, locations, and absorption coefficients (a-f). Different y-axis scales were needed to highlight the models' performance in reconstructing the optical coefficient values. Evidently, the images reconstructed by our method are more accurate than those reconstructed by the other approaches. The MLP approach presents a rough reconstruction that is further improved by FC-CNN. The NCG method reproduces lesion location, while lesion depth is less accurate. The LS method provides the best overall results.

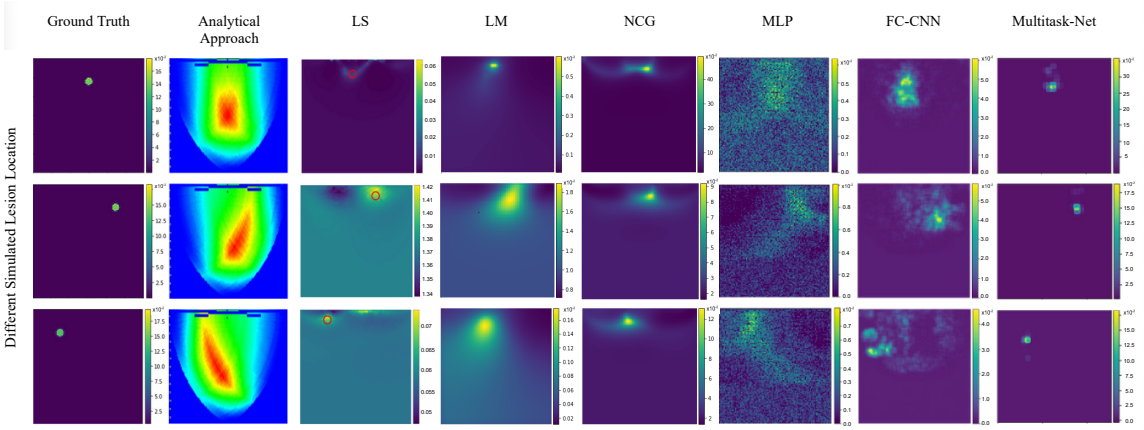

Figure 5.3: Qualitative reconstruction performance of our model compared to state-of-the-art (Analytical, LS, LM, NCG, MLP, and FC-CNN) techniques on phantom samples with known lesion ground truth location. The x-axis and y-axis show lesion location and depth in mm, respectively. Color bar shows the reconstructed absorption coefficient values. Different y-axis scales are unfortunately needed and indicative of the problematic performance of some models in reconstructing the optical coefficient values. The reconstructed image using MLP is equivalent to a back projection operation and gives a rough approximation of lesion localization, while precision of the reconstructed absorption coefficients can be low. Adding convolution layers helps to better filter noise and reduce false-positives while still miss-predicting correct absorption coefficient values required for proper diagnosis. The LM and NCG methods provide good reconstruction with accurate lesion presence detection. Lesion size mis-prediction can be attributed to real data noise and artifacts.

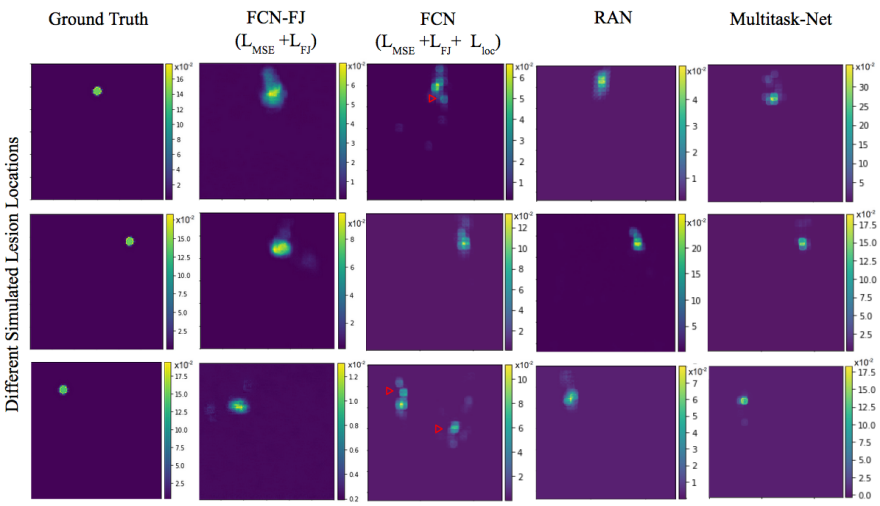

Figure 5.4: Ablation study qualitative reconstruction performance' on phantom samples with known lesion ground truth location. Our proposed method results (last column) are contrasted to FCN with different combinations of loss functions (L_{MSE}, L_{FJ}, and L_{loc}) and RAN architecture with L_{MSE} loss results. By attending to important features, spatial-wise attention model (RAN) filters irrelevant features resulting in improved lesion size and location compared to FCN while reducing false positives (e.g., red triangles). Combining spatial-wise attention and similarity loss provides the best reconstruction results with more accurate reconstruction of absorption coefficients.

confirmed by imaging reported a mean depth of 1.6 cm with a maximum of 6 cm (163 patients, 173 lesions). While this does not preclude deeper lesions in larger population studies, it shows that our performance is aligned with real-world assumptions on lesion depth.

5.4.2 Results on Phantom Data

Trained on the in silico data and tested on the phantom dataset, we visually compare our proposed reconstruction method of absorption coefficients to the competing methods' results on sample phantom cases. We compare to analytical method [229], LS based method [70], LM based method [221], NCG based method [18], MLP model [84], FC-CNN [285], and our preliminary works FCN [30], and FCN-FJ [33].

Qualitative Results on Phantom Data

In Fig. 5.3 and Fig. 5.4, we contrast our results with the state of the art on phantom data. Our DOT reconstruction model is mainly interested in the absorption parameter change as it is related to the hemoglobin concentration changes that mark the presence of anomalies [12].

The analytical approach, while detecting the presence of the lesion, only approximates its location by a rough margin and fails to provide precise reconstructed optical coefficient

Table 5.2: Quantitative results on 32 phantom experiments. Methods with a runtime smaller than 20 images per second (< 0.05s) are considered real-time capable.

		Loss			Loc. Error	PSNR	SSIM	Fuzzy Jaccard	Runtime
		L_{MSE}	L_{FJ}	L_{Loc}	(pixel (in mm), < is better)	(dB, > is better)	(> is better)	(> is better)	(s, < is better)
Conventional	Analytical [229]	na			77.4 ± 32.2 (17 ± 7.1)	15.0 ± 6.0	0.32 ± 0.26	0.17 ± 0.15	83×10^{-3}
	LS [221]	na			31.7 ± 11.7 (6.9 ± 2.6)	13.7 ± 11	0.29 ± 0.07	0.16 ± 0.08	3×10^{2}
	LM [221]	na			25.2 ± 12.7 (5.5 ± 2.8)	16.6 ± 8.6	0.37 ± 0.09	0.16 ± 0.05	5×10^{2}
	NCG [18]	na			32.6 ± 16.5 (7.2 ± 3.6)	16.5 ± 9.2	0.33 ± 0.01	0.22 ± 0.13	2×10^{2}
Deep Learning	MLP [84]	✓	-	-	42.0 ± 17.3 (9.3 ± 3.8)	12.5 ± 1.5	0.05 ± 0.03	0.07 ± 0.05	2×10^{-3}
	FC-CNN [285]	✓	-	-	27.8 ± 18.7 (6.1 ± 4.1)	17.3 ± 7.2	0.38 ± 0.23	0.45 ± 0.11	5×10^{-3}
	FCN [30]	✓	-	-	33.2 ± 23.4 (7.1 ± 5.1)	20.1 ± 4.6	0.46 ± 0.28	0.32 ± 0.06	8×10^{-3}
	FCN-FJ [33]	✓	✓	-	14.8 ± 7.40 (3.2 ± 1.6)	21.7 ± 0.9	0.73 ± 0.03	0.64 ± 0.10	17×10^{-3}
	FCN+L_{Loc}	✓	✓	✓	13.6 ± 7.2 (3.0 ± 1.6)	21.3 ± 2.1	0.74 ± 0.21	0.63 ± 0.05	19×10^{-3}
	RAN	✓	-	-	14.3 ± 7.1 (3.1 ± 1.6)	18.1 ± 6.1	0.76 ± 0.11	**0.70 ± 0.04**	20×10^{-3}
	Multitask-Net	✓	-	✓	**12.1 ± 6.4 (2.6 ± 1.4)**	**23.8 ± 1.7**	**0.77 ± 0.18**	0.69 ± 0.09	23×10^{-3}

values required for proper assessment of lesions (Fig. 5.3- Analytical). LM and NCG based methods (Fig. 5.3- LM, NCG) provide good reconstruction as they are able to detect and localize lesions. Lower performance on lesion size can be in part attributed to noise and artifact on real data information. MLP provides a rough approximation of lesion localization and sizes with very low per pixel values of reconstructed absorption coefficients (Fig. 5.3- MLP). The FC-CNN and FCN-FJ, on the other hand, localize the lesion in a single concentrated area with some error in lesion size and optical coefficients (Fig. 5.3- FC-CNN and Fig. 5.4- FCN-FJ ($\mathcal{L}_{\text{MSE}}$+$\mathcal{L}_{\text{FJ}}$)). Adopting the new loss term $\mathcal{L}_{\text{loc}}$ clearly reduces the lesion localisation error (Fig. 5.4- FCN ($\mathcal{L}_{\text{MSE}}$+$\mathcal{L}_{\text{FJ}}$ +$\mathcal{L}_{\text{loc}}$)). Small artifacts are a disadvantage of this model. The proposed RAN consistently obtains results that minimize both location and size error while miss estimating values of reconstructed absorption coefficients (Fig. 5.4- RAN). Observe how incorporating both the new loss term $\mathcal{L}_{\text{loc}}$ and the RAN in the Multitask-Net significantly reduces the artifacts while improving lesion localization and reconstruction of optical coefficients, which otherwise could compromise diagnosis (Fig. 5.4- Multitask-Net). From these results, it is clear that multitask learning is able to achieve superior performance in both tasks without compromising subtask performance.

Quantitative Results on Phantom Data

Table 5.2 presents the results on the phantom dataset. Yoo et al. architecture presents close results to our preliminary work FCN [30] given the similarity between the used architectures, where the difference lies in terms of model depth and filter sizes. It is worth mentioning that some artifacts are due to the shift in distribution that we overcome using the transfer learning model.

Rows 5 to 8 present an ablation study that shows the contribution of each element of our model. While the distance transform loss allows reduction in lesion localization error, the attention network helps in terms of reconstruction similarity and false positive reduction. We obtain marked improvements in localization error (up to 18%), as well as improvements

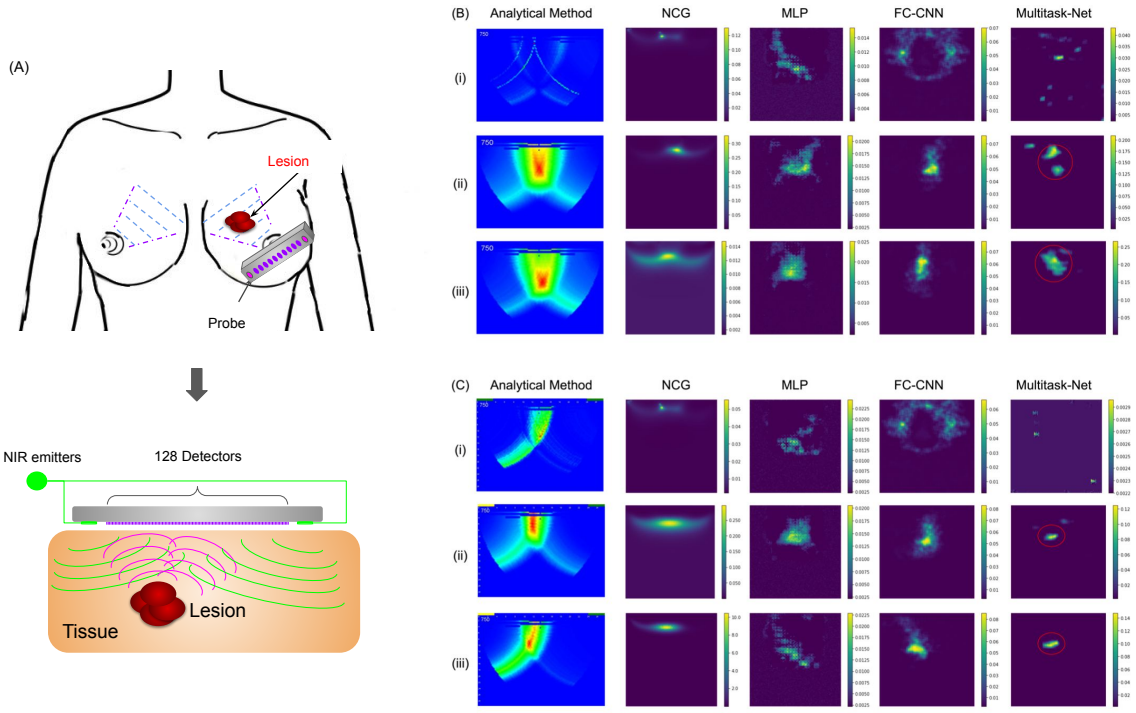

Figure 5.5: Reconstruction performance on real patient data. (A) Clinical data scan procedure. Dashed blue lines delineate the different probe sweeps. (B-C) Qualitative reconstruction performance of our model (last column) compared to analytical, NCG, MLP, and FC-CNN methods on different real patient data. (i) Reconstruction of healthy breast tissue of candidate versus cancerous lesion over different probe positions (ii-iii). Note how the size of the lesion reconstructed by our method more faithfully reflects the relative difference in size. (approximate lesions sizes obtained with ultrasound; patient 1 (B): $22 \times 21 \times 23$ mm, patient 2 (C): $19 \times 16 \times 18$ mm). The minimum of the colorscale is set to the background value, the maximum dynamically to try to report visually consistent results, while accommodating the very high variance between images that would otherwise prevent any comparison.

in SSIM (up to 4%), Fuzzy Jaccard (up to 5%), and PSNR (up to 3%). Interestingly, we found that a RAN performs better than an FCN+$\mathcal{L}_{Loc}$ in terms of Fuzzy Jaccard metric, hence, we can argue that a RAN helps to better focus on lesion area and reduces the false positive pixel.

Computational time for each reconstruction in this study is reported in Table 5.2 (runtime). Once the deep learning model training is completed, the time for the reconstruction is in the order of a few ms, three orders of magnitude faster reconstruction than conventional methods thus enabling real time application. Note that the new loss function has little to no effect on the inference time.

Table 5.3: Quantitative results on real data experiments.

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

5.4.3 Results on Clinical Patient Data

Figure 5.5 presents the reconstructed cross-sectional image of absorption coefficients of breast scans from 2 patients. The probe is placed above the identified lesion in the cancerous breast and at the same location on the opposite healthy breast for each patient (Fig. 5.5-A). We compare with the analytical [229], NCG [18], MLP [84], and FC-CNN [285] methods, reporting visual results in Fig. 5.5- B,C and quantitative results in Table 5.3. As a partial ground truth, both patients underwent ultrasound scans to obtain estimated lesion dimensions, showing irregular masses measuring $22 \times 21 \times 23$ mm for the first patient and $19 \times 16 \times 18$ mm for the second patient, with biopsies confirming that the lesions are cancerous.

In healthy tissue (Fig. 5.5-B:i, C:i), only background signal is reconstructed by almost all methods. In cancerous tissue, all competing approaches and our proposed method detect the lesion. The small artifacts (false positives) in healthy patient data reconstructed by MLP, FC-CNN and Multitask-Net (Fig. 5.5-B:i,C:i) can be attributed in part to the shift in distribution between training and test dataset. While the transfer learning module reduces this gap, it cannot eliminate this gap to the real patient data distribution completely. We observed that such artifacts could be removed by post-processing, e.g., via thresholding, by a qualified clinical operator. At the same time, these failure cases highlight the need for more real-life patient data to better train the transfer learning module. Consistent with the results on phantom data, the analytical, NCG, and MLP methods have a rough approximate outline of the tumor, whereas methods only partly agree with the approximate center of the predicted lesion.

Table 5.3 reports the average of reconstructed area ratio with respect to the ground truth lesion size over different probe sweeps for each patient. Areas of thresholded regions ($> 80\%$ binary threshold), given the pixel dimensions of the image, were compared to the real ground truth area. We observe that, in contrast to the other approaches, our method is able to reconstruct the lesions more faithfully with respect to their relative sizes for patient I while it underestimates lesion size for patient II. In addition, our method obtains a more precise delineation of the lesions.

Table 5.4: Sensitivity study: The effect of β on performance.

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

Table 5.5: Performance impact of varying number of network layers.

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

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

5.4.4 Multitask versus Single-task

We first tested how the combination of different loss functions influences reconstruction and localization performance. We used $\mathcal{L} = \mathcal{L}_{\text{MSE}} + \beta\mathcal{L}_{\text{loc}}$ and explored different values for $\beta \in [0,1]$. Table 5.4 shows the reconstruction performances as well as the localization error on the test set of 80 images for different values of β. Overall, we observed that learning with multiple losses improved both reconstruction and localization performance. In fact, we observed an up to 18% decrease in localization error when using a combination of $\mathcal{L}_{\text{MSE}}$ and $\mathcal{L}_{\text{loc}}$ compared to using $\mathcal{L}_{\text{MSE}}$ in isolation, with increases in SSIM by 6% and Fuzzy Jaccard by 7% using the best weight. Table 5.2 clearly shows that multitask-learning obtains optimal results for all metrics except a marginal drop in Fuzzy Jaccard of 1%.

5.4.5 Network Depth Trade-off: Complexity versus Performance

Increasing the dimensions of the network can allow for improved performance, yet results in larger hardware and training time requirements. With our application domain in hand-held devices, it is critical to evaluate the trade-off between network dimensions and performance. With an exhaustive architecture search beyond the scope of this work, we limit ourselves to

Table 5.6: Variation of the model performance over different number of channels.

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

varying the number of layers and channels. Table 5.6 and 5.5 show experimental performance results that were performed on the in silico dataset.

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

While the deeper model performs better compared to a shallower network (Table 5.5), performance did not improve ($\sim$ 1% variation in all metrics) by adding more residual attention layers compared to the baseline model at the price of more parameters and risking overfitting. A lighter architecture would be more convenient for the deployment in handheld devices with small power requiring a fine balance between accuracy and latency.

Note that the fully connected layer, used to back project the sensor data in the image domain, non-linearly increases the memory requirements of the neural network, thereby limiting the maximum number of additional convolution layers. This layer is deemed compulsory for an inverse scattering problem where direct analytical reconstruction is not well-defined, unlike in magnetic resonance imaging or computed tomography where a well-defined inversion exists using Fourier transform or Radon transform [171].

To observe the dependency on the number of convolution channels, we additionally trained the network with a varying number of channels. We observed marginal variations of the network performance beyond 32 channels: a variation of 1% and 1.5% in SSIM and Fuzzy Jaccard metrics, respectively, beyond 32 channels (Table 5.6). This can be explained by the size and shape of lesions that are generally small, when compared to the background, and smooth objects, therefore they can be easily fitted by a small number of feature maps.

5.4.6 Effect of Varying the Number of Sensors

Given that our probe only has 2 light sources and 128 sensors (detectors), we tested the robustness of our model to the variation of the number of sensors (co-linear equi-spaced detectors) by reducing it gradually from 128 to 16. To train the network for different source-detector configurations, we generated different sets of training data and changed the input sizes of the network accordingly while maintaining the overall network structure. While it is out of the scope of this work to test the results using the DOB-Scan probe (requiring

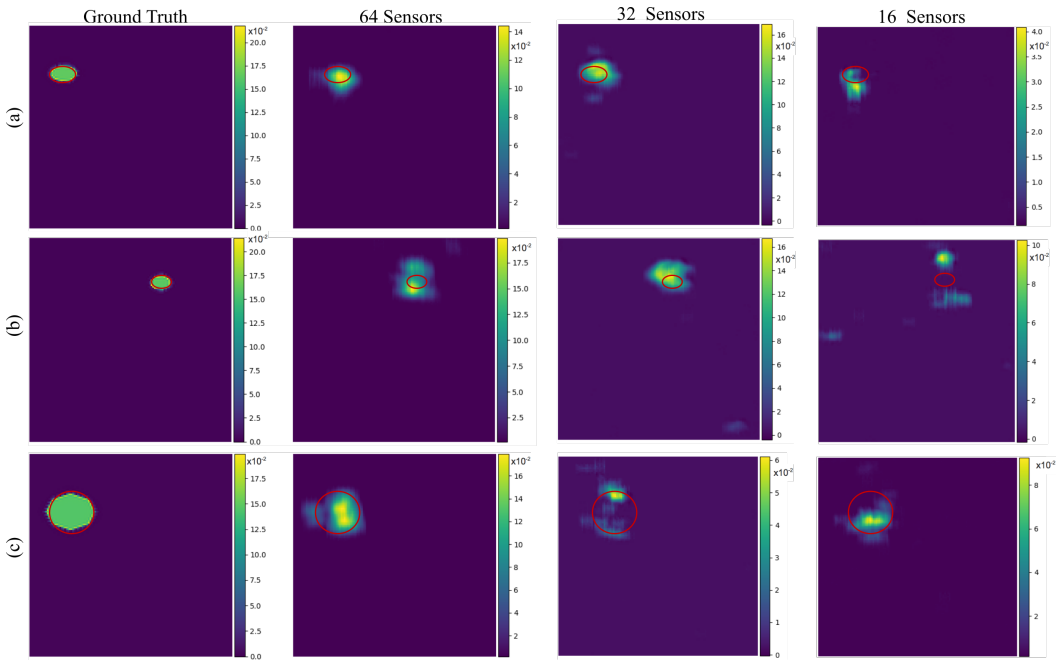

Figure 5.6: Effect of varying number of detectors, while keeping the same number of probe sources', on the Multitask-Net qualitative results' over different lesions sizes, locations, and absorption coefficients (a-c). While the model is still able, in most cases, to detect lesion presence, lesion size and correct localization are impacted by the reduction in the number of sensors and the increase in sparsity of the collected signal.

new hardware design, testing, and manufacturing), we reported experimental performance results for the in silico dataset in Table 5.7 and visual inspection results in Fig. 5.6.

Note the high contrast between the background and the lesion area, despite the increased difficulty of accurately recovering lesion size and location in sparse signal conditions (smaller number of detectors). The results are on par with qualitative results Table 5.7 where we notice an increase in localization error while we observe a slight decrease in SSIM.

5.5 Discussion

5.5.1 Attention-directed Lesion Reconstruction

Table 5.7: Trade off between probe complexity and model performance on the synthetic test set.

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

DOT reconstruction leverages the fact that healthy tissue has optical coefficients that are near uniform, whereas lesions have optical coefficients that diverge markedly and are characterized by high variability. The encoding of a highly variable signal (lesion) requires more information or computational complexity than a near uniform signal. The high variance is explained in part due to high inter-tumor heterogeneity, a leading cause preventing effective treatment or prognosis [28]. Furthermore, for the clinician, it is sufficient that the network correctly recognizes that certain areas are healthy tissue without requiring precise pixel-wise reconstruction. In contrast, the lesion needs not only to be detected, but located and reconstructed precisely in order to enable diagnosis, prognosis, and successful treatment recommendation. Therefore, employing network attention on the lesion is one way of dedicating more computational power where it is required without sacrificing detection of healthy tissue.

5.5.2 Network Generalization

The generalization capability of our network is illustrated by the fact that training occurs on simulation data, yet performance is strong on phantom and real data. Moreover, unlike the analytical approach that requires a reference measurement of healthy tissue to compute the inverse image, our network, similar to Yoo et al.[285], is designed to directly learn from the collected measurement and does not require any reference or prior conditions that might bias the search space. In a clinical setting, such as breast cancer screening, such reference measurement from a homogeneous background may be nontrivial to obtain. Therefore, our model can generalize better to unseen conditions without additional pre- or post-processing techniques. In this work, we obtained good results when trained only on in silico data. We expect improved results as more real data from phantoms and patients is collected. Finally, we used model architecture and implementation details described in Section. 5.2, with good initial results. However, it is likely that other architectures can be found that will yield improved results, e.g., via neural architecture search [79].

5.5.3 Limitations

Although this study provides a proof of concept of the potential advantages of using multitask deep learning for DOT image reconstruction, it suffers from some limitations. First, reconstruction of deep lesions, with a location deeper than 35 mm, remains a challenging task in DOT. When one uses a handheld probe, with a colinear sources-detectors placement, it can be even harder to differentiate the lesions specific signal from tumor-free tissue signal. As a result, the reconstructed image can be biased towards tumor-free tissue. Similarly, while our model is able, in most cases, to detect lesion presence, reconstruction of multiple lesions exhibits some weakness, as shown in Fig. 5.7. Specifically, occluded lesions and lesions situated at the border of the probe can be mis-reconstructed, where only one lesion is reconstructed, lesion location deviates, or tumor-free tissue is reconstructed (false

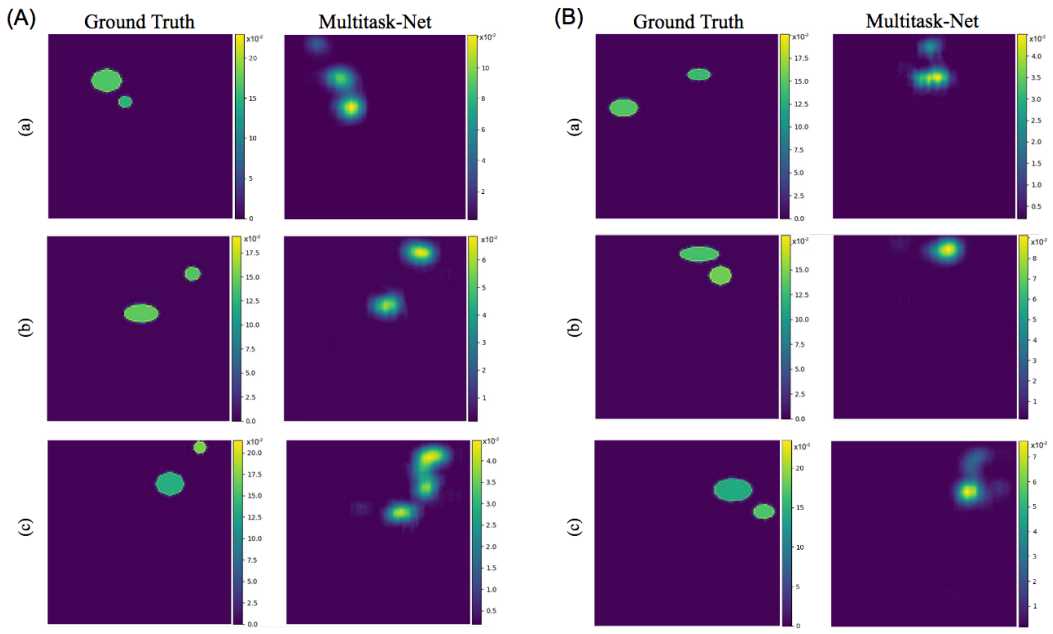

Figure 5.7: Reconstruction of multi-lesions with different sizes and proximity (a-c). (A) While the model is able, in most cases, to detect lesion presence, occluded, edge located and deep lesions can be mis-reconstructed. (B) Some failure reconstruction examples where only one of the lesions is reconstructed.

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

5.6 Conclusion

We present the first work to effectively leverage reconstruction guided localization multitask learning by combining an appearance similarity loss and a reconstruction loss along with a spatial attention mechanism for DOT-reconstruction, with application to reconstructing lesions in breast cancer tissue. We demonstrated improvement in localization error over the state of the art and presented a step forward in DOT image reconstruction and the usage of machine learning in image reconstruction. Leveraging physics-based simulators of light propagation to generate realistic synthetic datasets provided an alternative source of large and diversified training dataset. The critical addition of transfer learning ensures our method is not limited by inherently small real-world datasets. At the same time it is

still capable of performing at high accuracy on such datasets where more complex sources of artifacts and noise are present. Nevertheless, we noted some performance differences in real world data. As more patient data is collected, it will be critical to better understand and address the causes of such differences. Finally, we show that our method is sufficiently fast enough to ensure real time processing, enabling in situ deployment in clinical settings. In the next Chapter, we will focus on multifrequency fusion and the patient diagnosis outcome prediction, which remains the most crucial measure in the medical image domain for evaluating image reconstruction results.

5.7 Disclaimer

Chapter 6

End-to-End Diagnosis via Deep Orthogonal Multi-frequency Fusion

In this Chapter, we explore the diagnostic capability of breast cancer lesions with a portable DOT device. Early detection of breast cancer lesions is crucial, and while a portable DOT device can aid in this process, the challenge lies in accurately distinguishing between malignant and benign lesions. Thus, our investigation focuses on evaluating diagnostic performance using multi-frequencies collected signal[1].

6.1 Introduction

Breast cancer pre-screening is usually carried out using self-breast examinations, which can suffer from high false-positive rates, or clinical breast examinations [63]. Although breast lumps are often benign, such as lipoma, cyst, or hamartoma, lesion malignancies may appear with a non-palpable sign; hence, regular screenings are critical [254]. Identifying breast cancer lesions with a portable DOT device can improve early detection, while avoiding otherwise unnecessarily invasive, ionizing, and more expensive modalities such as CT, as well as enabling pre-screening efficiency. Critical to this capability is not just the identification of lesions but rather the complex problem of discriminating between malignant and benign lesions. Fig. 6.1-A shows a typical breast screening in the medical workflow.

To accurately capture the highly heterogeneous tissue of a cancer lesion embedded in healthy breast tissue with non-invasive DOT, a combination of multiple frequencies can be employed to optimize signal penetration and reduce sensitivity to noise. Through the utilization of distinct frequencies, the DOT system can more precisely measure the tissue's response to light at various depths. Additionally, interactions between light and tissue properties like absorption and scattering exhibit wavelength-dependent variations. The integration of multiple frequencies empowers the DOT system to selectively target different

[1]The content of this Chapter heavily relies on our published research [32]. Reproduced with permission from Creative Commons Attribution (CC BY) license

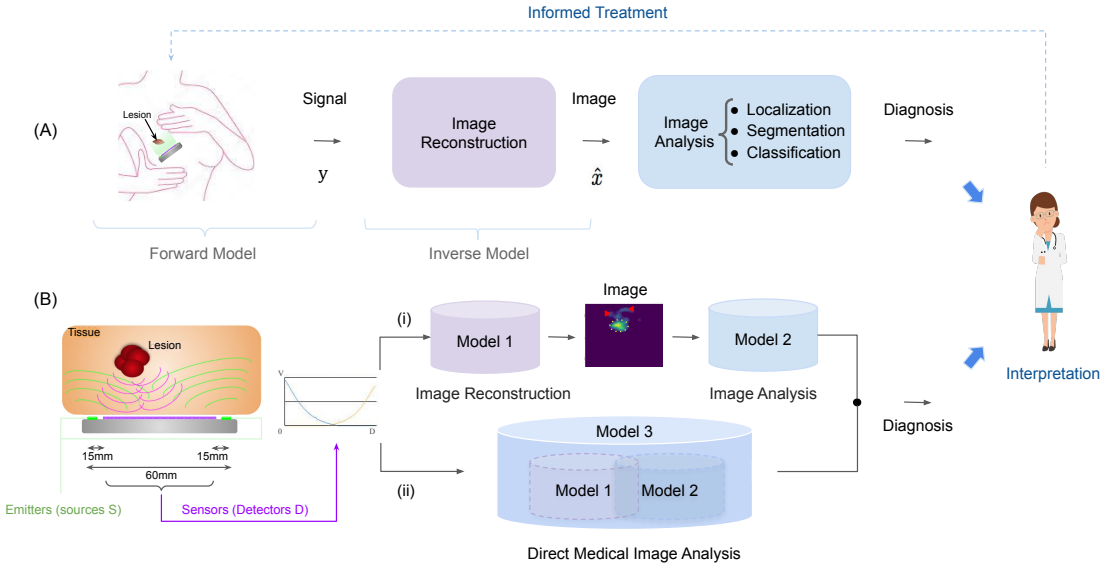

Figure 6.1: Typical breast cancer screening workflow. (A) Images reconstructed by an inverse model, from signals collected by the acquisition hardware, are analyzed for assessment, diagnosis, and treatment prognosis. (B) Screening pipelines can be divided into two paradigms: (i) Accurate reconstruction followed by image based classification. (ii) A direct prediction model omits image reconstruction to focus solely on the ultimate task and can help overcome sub-task errors, e.g., reconstruction induced false positives, marked by red triangles in this scenario, in paradigm (i).

chromophores, resulting in a more accurate reconstruction of the tissue's internal characteristics.

Notably, alterations in chromophore concentrations captured through this process yield information about vital functional events within the brain's vasculature, facilitate the assessment of tissue oxygenation, and facilitate the characterization and ongoing monitoring of breast lesions [76]. Exploiting this wavelength sensitivity, optical spectra can be meticulously analyzed, and images of exposed tissue can be reconstructed, serving diagnostic purposes. However, these responses share variable overlap in the absorption spectra of the major tissue constituents in the red-near infrared range [245] potentially capture common information, and correlate (Fig. 3.2). We investigate multi-frequency DOT fusion for improved reconstruction and more accurate end-to-end identification of malignant versus benign lesions.

6.1.1 Regularization Approaches in DL-reconstruction Methods

Recent studies, e.g., [33, 282, 285, 178], have shown image reconstruction and classification are faster and more accurate when deep learning algorithms are used instead of conventional reconstruction methods. One advantage deep learning based algorithms have over classical reconstruction methods is that they can exploit implicitly learned feature encodings from

Table 6.1: Regularization approaches in state-of-the-art DL-reconstruction methods. D: design approach; FF: feed-forward, I: iterative unrolled based model; M: multi-modal/frequency; S/P: in silico/phantom data; and C: clinical patient data.

	D	M	S/P	C	Approach to mitigate ill-posedness
[213, 30, 186]	FF	×	✓	×	CNN learns the nonlinear end-to-end mapping
[33]	FF	×	✓	×	Promote appearance similarity
[178, 177]	I	×	✓	×	Augment Gauss-Newton algorithm with deep learning
[187]	FF	✓	✓	×	Micro-CT structural prior
[285]	FF	×	✓	×	Model based on Lippmann-Schwinger equation
[82]	FF	×	✓	×	Reflection model as sum of features from different depths
[303]	I	×	✓	×	Data-driven unrolled network promoting appearance similarity
[310, 154, 311]	FF	✓	✓	✓	Multi-modal representation learning (US+DOT)
[282]	FF	×	✓	✓	Deep spatial-wise attention network
Ours	FF	✓	✓	✓	Orthogonal multi-frequency representation learning

the DOT sensor data, whereas classical reconstruction algorithms can exploit only priors encoded by human designers [31]. Recent advances tackle the problem of ill-posedness in a variety of ways. We summarize the closely related approaches in a tabulated overview (Table 6.1), and refer the interested reader to [22, 234] for a more in-depth review.

6.1.2 Multi-frequency DOT

The primary motivation for multi-frequency DOT is to exploit the different but complementary responses of chromophore, tissue components, to multi-frequency excitation, given that chromophores absorb photons at different rates at different wavelengths [100]. A multi-spectral image can be obtained using several LEDs or lasers of multiple wavelengths as illumination. The different LEDs are used consecutively to capture an image per wavelength or combined as one multi-spectral image [226]. The captured multi-frequency data can provide more spatial and contextual information, enabling more robust and accurate identification and discrimination of disease-correlated biological anomalies[166].

While higher frequencies allow for a better separation of optical properties, such as absorption and scattering coefficients, as well as a better detection of small and shallow objects, the limit of the signal-to-noise ratio (SNR) decreases with increased frequency [248, 144], and penetration decreases as frequency increases. Utilizing multi-frequency data for improving DOT image reconstruction and diagnosis has been an active field of research, illustrating that the accuracy of the optical coefficient can be improved using measurements with multiple frequencies [130, 181, 226, 11]. Recent studies, summarized in a tabulated overview (Table 6.2) have shown that using measurements with multiple frequencies can improve the recovery of optical coefficients and provide higher SNR and lower error [52, 226]. Improvement, however, depends on frequencies selection scheme and utilized instrument given the specific noise impact [11]. This finding is supported by Zimek et al.[309], who

Table 6.2: Multi-Frequency DOT for Image Reconstruction and Diagnosis. Di: dimension; Freq: frequency range; D: dataset, S: simulation, P: phantom, C: clinical; and MF:Multi-Frequency.

	Leveraging different frequency schemes	DI	Freq (MHz)	D
[52]	Improve joint optical coefficients recons.	2D	100-250	S
[226]	Enhance fused MF image quality	2D	100-1000	S
[160]	Compensate physiological and noise interference in recons.	3D	361-382	P
[141]	Frequency shifting for reduced recons. ill-posedness	2D	$100+5*i$, $\forall i \in \{0, .., 100\}$	S
[198]	Minimize the effect of phase data and improve contrast	2D	{78, 141, 203}	P
[11]	Evalutate the impact of frequency selection	2D	50-500	S
[245]	Discrimination between malignant and benign tissue	2D	283-472	C

reported that adding dimensions can harm discriminative potential if those dimensions do not improve the signal-to-noise ratio.

Augmenting DOT with ultrasound is finding recent adoption as well, an example of multi-modal fusion [200, 73],[8]. The aforementioned art is based on conventional reconstruction algorithms. To the best of our knowledge, no deep learning-based method has explored the merit of exploiting multiple frequencies in DOT-reconstruction and diagnosis.

6.1.3 Multi-frequency as Data Fusion

Data fusion models mimic higher cognitive abstraction in the human brain by synthesizing information from multiple sources for improved decision-making. While data fusion is non-trivial, the resulting contribution of multiple data sources or multimodal data can significantly improve the performance of deep learning models [300, 109]. The underlying motivation for collecting multi-modal data is to learn the optimal joint representation from rich and complementary features of the same object or scene. In the context of combining multiple information sources to learn more powerful representations, the terms 'early' and 'late' fusion are commonly used [195]. Early fusion refers to concatenating input data from multiple sources in separate channels before presenting it as input to the network, while late fusion involves processing each input data individually and aggregating their output. Mid-fusion restricts cross-data flow to later layers of the network, allowing early layers to specialize in learning and extracting data-specific patterns [182].

Attention mechanisms have been shown to be suitable for the fusion of features that usually suffer from confounding issues such as conflicting or cancelling information, correlation, and noise. Attention provides an approach to learn to select informative subsets of the data, as well as the relationship between data streams, before fusing them into a single comprehensive representation [109, 57]. Transformer based models, based on a multi-head attention architecture, have recently gained increased adoption [42, 182]. However, the high

computational cost and complexity, scaling adversely with input sequence length, remain a significant challenge, especially given the real-time requirement.

Self-supervised learning (SSL) based on a joint embedding architecture, driven by the maximization of the information content of the network branches' embedding, opened the door to the application of joint-embedding SSL to multi-modal signals [25]. The idea is to produce independent embedding variables, removing confounding effects such as partial correlation and avoiding modal collapse between data streams by encouraging architecture diversity between branches, using loss based normalized cross-correlation matrix [292] or explicit variance-preservation term for each embedding [25].

Imposing orthogonal constraints in linear and convolutional neural network layers can act as a form of regularization that can help improve task performance and be beneficial for the network's generalization [24, 123]. Orthogonality in feature space was proposed to encourage intra-class compactness and inter-class separation of the deep features, and has shown improvement in classification tasks [204]. Multi-modal orthogonalization has been used to force uni-modal embeddings to provide independent and complementary information to the fused prediction [43]. Another advantage is that an orthogonal encoding can enforce the learning of a more sparse correlation-free representation. The resulting smaller encoding can reduce architecture dimensions, and serve as an implicit regularization.

6.1.4 Towards Direct Medical Image Analysis in DOT

Traditional computational pipelines in biomedical imaging involve solving tasks sequentially (Fig. 6.1-B.i, e.g., segmentation followed by classification or detection). Although each of these two tasks is usually solved separately, the useful clinical information extracted by the second task is highly dependent on the first task's results. While a 'joint' or multi-stage model where different tasks are lumped together, for example, image reconstruction then classifying diagnosis, can benefit from feature sharing and joint parameters tuning for both tasks, significant computational resources are required to optimize sub-tasks that may not necessarily lead to end-task improvements. In contrast, in the direct medical image analysis [305] (DMIA) paradigm, end task results are directly inferred from raw/original data (e.g., raw sensors or whole image/volume) as illustrates Fig. 6.1-B.ii. Therefore, the model can focus solely on the end task, reclaiming some of the computational resources for improved results while requiring fewer resources. For instance, Wu et al.[270] trained a neural network for joint reconstruction and lung nodule detection from raw acquisitions and showed performance improvement compared to a two-stage approach. Hussain et al.[127] had shown that a segmentation-free kidney volume estimation can help overcome segmentation errors and limitations and reduce the false-positive area estimates. In a similar perspective, Taghanaki et al.[244] investigated a segmentation-free tumor's volume and activity estimation in PET images. Recently, Abhishek et al.[2] illustrated that, in the context of cancerous skin lesions, predicting the management decisions directly can be a simpler problem to ad-

dress than predicting the diagnosis followed by management decisions, as one action can be prescribed to multiple subsets of disease classes.

6.1.5 Contributions

We make the following contributions in this paper:

(i) We investigate the benefit of multi-frequency data on the quality of DOT reconstruction and breast lesion diagnosis. Previously, many works have addressed the multi-frequency reconstruction problem or diagnosis, albeit using conventional methods. Despite the importance of multi-frequency acquisition for chromophore reconstruction, no deep learning framework has investigated multi-frequency fusion nor joint reconstruction and diagnosis to date. Here, we present a novel approach designed to recover the optical properties of breast tissue from multi-frequency data with a deep orthogonal fusion model followed by a diagnosis.

(ii) To the best of our knowledge, this is the first deep learning-based method that investigates the merits of tackling the diagnosis prediction task from raw sensor data directly without image reconstruction in DOT (direct prediction [2]). Results with and without reconstruction are contrasted using a modular pipeline, highlighting the potential of the raw-to-task model for improved accuracy, while reducing computational complexity.

(iii) We extend a fusion network [57] by training models using an orthogonalization loss function [204] to maximize the independent contribution of each frequency data and emphasize their collective strength, with improved predictive performance compared to a single frequency model.

Section 6.2, introduces our proposed model for multi-frequency DOT fusion and defines the two prediction pipelines (raw-to-task and joint reconstruction and diagnosis). Physics-based computational simulation and real patient datasets are detailed in Section 6.3.1. In silico performance results are presented in Section 6.3.2 and results on real-world data in Section 6.3.3. We conclude the paper by discussing insights and limitations on interpretability, speed, and adaptive dynamic treatment in Section 6.4.

6.2 Methodology

Solving the inverse problem in DOT recovers the spatial distribution of a tissue's optical properties $x \in \mathbb{R}^{W \times H}$ based on the measured boundary data $y^i \in \mathbb{R}^{S \times D \times N}$, from S sources (emitters) with D sensors (detectors) at different frequencies $i \in \{1, N\}$. The learned inverse function $\mathcal{F}^{-1}(\cdot)$ maps the raw measurements y to an image estimate $\hat{x}$ while remaining

[2]While our direct prediction contribution omits the 'tomogram' part of DOT, and thus works directly on near-infrared (NIR) sensor data, our fusion contribution applies both to tomogram reconstruction as well as tomogram-free reconstruction. Thus, we continue to use DOT throught the paper instead of NIR.

faithful to the underlying physics constraints. Learning the inverse function $\mathcal{F}^{-1}(\cdot)$ is carried out by solving:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}\left(\mathcal{F}^{-1}(y^i;\theta); x\right) + \lambda\mathcal{R}(\mathcal{F}^{-1}(y^i;\theta))\,, \tag{6.1}$$

where $\mathcal{L}$ and $\mathcal{R}$ are the network loss function and the regularization, θ are the optimized network weights that parameterize $\mathcal{F}^{-1}$. The reconstruction of an image based on the fusion of all raw signals from diverse frequencies is considered as well by using the fusion network described in Section 6.2.1. While reconstructing an accurate 2D/3D image/volume from collected measurements has been the mainstream task in DOT, in a clinical setting, the ultimate purpose is not necessarily obtaining the image itself but rather making an informed clinical diagnosis or management decision, such as lesion detection and classification into predefined classes. To compare the impact of omitting the reconstruction and directly predicting the end task, we implemented two architectures: The first, FuseNet, reflects classical approaches, i.e. a classification module is appended to the output of the reconstruction layer to make a prediction, where the result of the multi-spectral reconstruction is used to supervise the classification task (Section 6.2.2). Whereas the second, Raw-to-Task, uses the same classification module to make a prediction based on the fused raw data directly, i.e. no reconstruction is considered in between. The ultimate goal is to study the ability of deep learning to provide superior prediction based on the raw signal only while reducing model complexity and computational cost (Section 6.2.3).

6.2.1 Fusion Network

Given multi-frequency raw data paired with known diagnosis outcomes, the objective is to learn a robust multi-frequency representation in a supervised learning setting. While many fusion strategies have been proposed in computer vision, natural language processing, and multimodal biomedical data, strategies for fusing data in multi-frequency DOT data remain unexplored in deep learning-based approaches. Inspired by recent methods for multimodal data fusion [57, 43], we adopt a similar attention-based mechanism to control the expressiveness of features from each input frequency before constructing the multi-frequency embedding, while uniquely feeding the raw data directly with no further pre-processing. Let $Y \in \mathbb{R}^{S\times D\times N\times M}$ be a training mini-batch including M tissue samples, each collected using N frequencies such that $Y = [Y_1, Y_2, ..., Y_N]$ where for each frequency i, $Y_i = [y_1^i, ...y_M^i]$ includes data for M samples. When $N > 1$, input measurements from each frequency are combined using the fusion branch (Fusion, Fig. 6.2). To reduce the impact of noisy input features and compress the size of the feature space, each Y_i is first passed through a fully connected layer of length l, with ReLU activation, outputting $Y_i^s \in \mathbb{R}^{l\times 1\times M}$, followed by an attention mechanism that scores the relevance of each feature in Y_i. We define frequencies $\not{i}$ as the set $\{j\}$ such that $j \in \{1, .., N\} \setminus \{i\}$, i.e. for frequencies other than i. A linear transformation W_A of frequencies $Y_{\not{i}}$ that would score the relative importance of each feature in

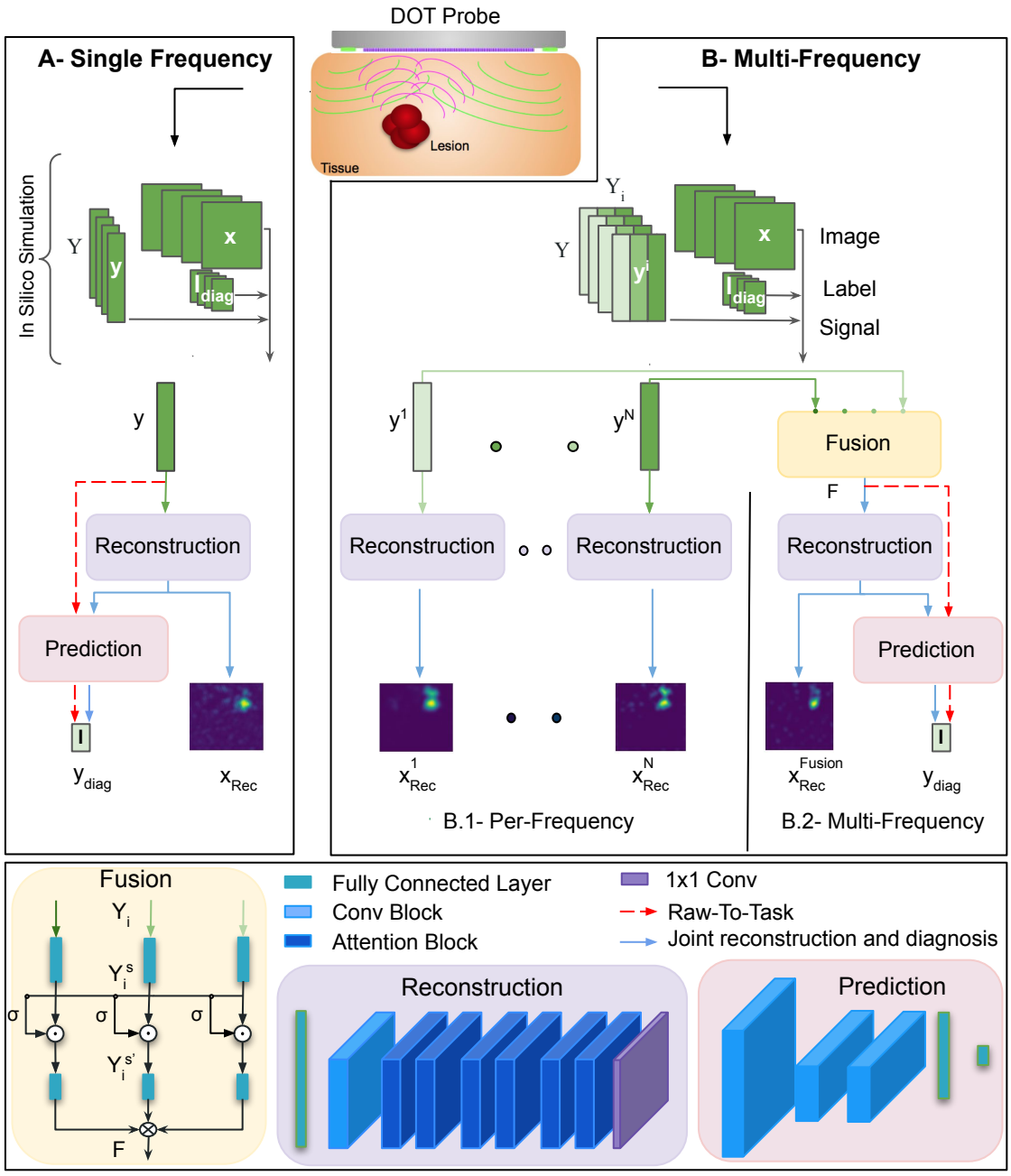

Figure 6.2: Architecture overview of the proposed DOT image reconstruction and diagnosis method. (A) single-frequency and (B) multi-frequency signals (y) along with corresponding ground truth diagnosis labels (l_{diag}) and images (x) are used to train the model. In the single-frequency variant of our method (A), y, is used as input to the image reconstruction, then the resulting image is used for diagnosis prediction. In multi-frequency, note the two variants: (B.1) per-frequency reconstruction and (B.2) multi-spectral reconstruction and diagnosis. For both single and multi-frequency, the red dashed lines depict the raw-to-task flow, where the image reconstruction is skipped and the diagnosis is predicted directly from y. The bottom panel shows the details of the multi-frequency fusion, reconstruction, and prediction modules.

i, is learned. W_A is a learned weight matrix parameters for feature gating. The attention weights vector a_i is then applied to Y_i^s, an element-wise product of scores and features, to form the attention-weighted embedding $Y_i^{s'} \in \mathbb{R}^{l\times 1\times M}$:

$$Y_i^{s'} = a_i * Y_i^s = \sigma\left(W_A * [Y_i]\right) * Y_i^s. \tag{6.2}$$

Finally, attention-weighted embeddings are passed through a fully connected layer of length $l2$, with ReLU activation, then combined through a Kronecker product between all frequency embeddings to capture possible interactions. Each vector is appended by 1 to capture partial interactions between frequencies [57]. The final fused embedding is then defined as:

$$F = \begin{bmatrix} 1 \\ Y_1^{s'} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_2^{s'} \end{bmatrix} \otimes \ldots \otimes \begin{bmatrix} 1 \\ Y_N^{s'} \end{bmatrix}. \tag{6.3}$$

$F \in \mathbb{R}^{l'\times l'\times l'\times M}$, for $N = 3$ and $l' = l2 + 1$, is a N-dimensional hypercube of all frequency interactions.

6.2.2 Joint Multi-frequency Reconstruction and Diagnosis

The task is to recover tissue optical properties and diagnosis outcome given raw signal data. While a single frequency model SF-JRD (Fig. 6.2-A), used as a baseline, relies on a single frequency measurement to reconstruct spatially distributed optical coefficients and predict diagnosis, a multi-frequency model (FuseNet) relies on a joint representation from multiple frequency measurements (Fig. 6.2-B). The multi-frequency model, including a network with multiple branches as shown in (Fig. 6.2-B), inputs N measurements of the same scanned tissue at N frequencies. A multi-spectral image that combines all frequency measurements, using the fusion branch encoding (Fig. 6.2-B.2), is reconstructed and passed to a classification module for diagnosis prediction. Furthermore, a per-frequency image is reconstructed using each modulated frequency signal. As depicted in (Fig. 6.2-B), the FuseNet model outputs are $x_{Rec}^i \ \forall i \in \{1, .., N\}$, x_{Rec}^{Fusion}, and y_{diag} which denote the per-frequency reconstructed image (Fig. 6.2-B.1), the multi-spectral reconstructed image, and the predicted diagnosis label (Fig. 6.2-B.2), respectively.

Using multiple inputs, per frequency network reconstruction branches (Fig. 6.2-B) learn independent representations, where features derived from each input measurement (Y_i) are only useful for the corresponding output x_{Rec}^i. Furthermore, given the differences in initialization, the branches can converge to disconnected modes in weight space, thereby behaving as independently trained neural networks. Empirically, we observe that they converge to distinct optima. For this multi-task reconstruction and prediction model, we extend the multi-task framework [282] and train a model to simultaneously reconstruct a per-frequency image, localize the lesion, and predict the diagnosis.

The reconstruction branch (Fig. 6.2) implements the design detailed in the multi-task framework, proposed in Chapter 5 [282], with a fully connected layer, 128×128, followed by a convolutional layer and 6 residual attention blocks with 32 channels, filters of size of 3 and ReLU activation, to produce the final reconstruction image. While the first and last layers are shallow feature extractors, the attention blocks extract hierarchical attention-aware features with modules of the form: two convolutions followed by squeeze and excite modules. This deep spatial-wise attention network attends to the most important features by reweighting features according to their interdependencies in feature space and filtering noisy ones. In contrast to the difference approach, which uses a reference measurement of healthy tissue to compute the contrasted inverse image, our reconstruction module is designed to learn the mapping directly, from the measured data to the desired output, without the need for any prior knowledge or references that can bias the search space. Furthermore, obtaining such a reference measurement from a homogeneous background in a clinical setting, such as a breast cancer screening, is not trivial; hence, we consider absolute imaging, where the network learns the inverse mapping between sensor measurements and the image domain directly.

The prediction branch (Fig. 6.2) includes 3 convolutional layers with max pooling and two final classification layers. Raw data from different frequencies are passed to the reconstruction branch except for the multi-spectral subnetwork, where raw data from different frequencies are first fused via the fusion branch. The fused features are passed to the reconstruction branch, which outputs a multi-spectral image followed by a classification layer to output the final classification prediction. The multi-task loss (L_{MULTI}) encompasses all three tasks: reconstruction, lesion localization, and diagnosis as a sum of losses for each task is defined as follows:

$$L_{MULTI} = L_{REC} + L_{DIAG}, \tag{6.4}$$

where L_{REC} and L_{DIAG} denote the reconstruction loss and the diagnosis losses, respectively.

Reconstruction Loss

We adopt the reconstruction loss defined in Chapter 5. The mean square error loss L_{MSE} combined with the location loss L_{LOC} guide the image reconstruction and lesion localization of the network as per (6.5). L_{MSE} recovers the pixel-wise representation of the image.

$$\begin{aligned} L_{\text{REC}} &= L_{\text{MSE}} + \beta\ L_{\text{LOC}}, \\ L_{\text{LOC}} &= ||DT(\mathcal{F}^{-1}(y_i, \theta), x) - DT(x)||, \end{aligned} \tag{6.5}$$

where DT denotes the distance transform and computes the Euclidean distance between the image pixel location and the lesion boundaries, θ denotes the parameters of the multi-task model, and $\beta \in [0, 1]$ is a hyper-parameter controlling the contribution of L_{LOC}.

Diagnosis Loss

The diagnosis loss, L_{DIAG}, is a weighted sum of the categorical cross entropy loss L_{CE}, and the orthogonal projection loss L_{OPL}:

$$L_{\text{DIAG}} = L_{CE} + \gamma L_{OPL}, \tag{6.6}$$

$$\begin{aligned} \text{where:} \quad L_{CE} &= L_{CE}\left(x, l_{\text{diag}} \mid \Theta\right) \\ &= -\sum_{j=1}^{n_{\text{diag}}} l_{\text{diag},j} \cdot \log\left(\phi\left(x \mid \Theta\right)_j\right), \\ L_{OPL} &= (1-s) + |d| \\ s &= \sum_{\substack{i,j \in B \\ y_i = y_j}} \langle \mathbf{f}_i, \mathbf{f}_j \rangle, d = \sum_{\substack{i,k \in B \\ y_i \neq y_k}} \langle \mathbf{f}_i, \mathbf{f}_k \rangle, \end{aligned} \tag{6.7}$$

n_{diag}, l_{diag} denote the number of classes in the diagnosis prediction tasks and ground truth label, respectively. $\phi(x|\Theta)_j$ denotes the predicted probability for the j^{th} class by the model parameterized by Θ. $\gamma \in [0, 1]$ is a hyper-parameter balancing the contribution of the L_{OPL}. $|x|$ is the absolute value operator, $< x, y >$ the cosine similarity operator applied on two vectors, and B denotes the mini-batch size.

The orthogonal projection loss L_{OPL}, as defined in [204], is used to maximize separability between classes by enforcing class-wise orthogonality in the intermediate feature space and simultaneously ensuring inter-class orthogonality (d term) and intra-class clustering ((1-s) term) within a mini-batch.

6.2.3 Direct Prediction: Raw to Task Model

The ultimate aim of DOT-based screening is the early identification and classification of breast cancer lesions. Therefore, we investigate if focusing exclusively on the end task, at the cost of omitting the reconstruction of a 2D image, can perform better or worse compared to classification with the intermediate reconstruction. Without the need to reconstruct a 2D image, the architecture and computational complexity reduce significantly, leading to a reduction in power consumption and data computation latency. The classification module is used to make predictions based on the fused raw data, where combined features, extracted from different frequencies using the fusion branch (Section 6.2.1), are passed to a convolutional layer for the prediction task and a final classification layer with the associated loss (Fig. 6.2-dashed lines). The diagnosis loss function, L_{DIAG}, is used to train the model given

Table 6.3: Summary of variants of our method architectures input and output details'. N: number of frequencies; S: sources; D: detectors; H: height; W: width.

	Input	Output	
		Direct prediction	Joint reconstruction and diagnosis
Single-Freq	$Y_1 \in \mathbb{R}^{S \times D}$	$y_{diag} \in \mathbb{R}$	$x_{Rec} \in \mathbb{R}^{H \times W}$
			$y_{diag} \in \mathbb{R}$
		SF-DP	***SF-JRD***
Multi-Freq	$Y \in \mathbb{R}^{S \times D \times N}$	$y_{diag} \in \mathbb{R}$	$x^i_{Rec} \in \mathbb{R}^{H \times W} \quad \forall i \in \{1, .., N\}$
			$x^{Fusion}_{Rec} \in \mathbb{R}^{HxW}$
			$y_{diag} \in \mathbb{R}$
		Raw-to-Task	***FuseNet***

the raw input measurement where:

$$\begin{aligned} L_{\text{CE}} &= L_{\text{CE}}\left(\left(y_i, .., y_N\right), l_{\text{diag}} \mid \Theta\right) \\ &= -\sum_{j=1}^{n_{\text{diag}}} l_{\text{diag},j} \cdot \log\left(\phi\left(y \mid \Theta\right)_j\right), \end{aligned} \tag{6.8}$$

y_i denotes the i^{th} measurement of the raw data and $\phi(y^{(i)}|\Theta)_j$ denotes the predicted probability for the j^{th} class given an input $y^{(i)}$ by the model parameterized by Θ. The orthogonal projection loss L_{OPL} (6.7) is used to maximize separability between classes in the feature space.

FuseNet, Raw-to-Task, SF-JRD and SF-DP models are trained separately while using the same modules: fusion, reconstruction, and prediction modules. Table 6.3 summarises different models input and output details.

6.2.4 Cross Domain Learning

Similar to Chapter 4, we use transfer learning to render an in silico trained network applicable to real world data and reduce the disparities between real-world acquisition y^p and in silico simulated data y^s. A multi-layer perceptron (MLP) network is used to tackle the domain shift by minimizes the transfer learning loss $\mathcal{L}_{TL}$ over N_p sets of real data measurements obtained using a phantom solution and their corresponding tissue-equivalent

Table 6.4: Optical coefficients distributions on the in silico dataset for wavelengths in 690-850 nm spectrum [147]

		Healthy tissue	Benign	Malignant
Absorption $\nu_a(cm^{-1})$	690	0.042 ± 0.013	0.08 ± 0.04	0.110 ± 0.066
	750	0.046 ± 0.024		0.100 ± 0.060
	800	0.052 ± 0.015		0.118 ± 0.096
	850	0.032 ± 0.005		0.124 ± 0.089
Scattering $\nu_s(cm^{-1})$	690	12.9 ± 2.3	19.4 ± 8.4	13.5 ± 4.7
	750	8.70 ± 2.2		11.6 ± 3.9
	800	10.5 ± 1.2		12.2 ± 1.7
	850	8.40 ± 0.4		9.10 ± 1.9

simulated data:

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}_{\mathrm{TL}}(\theta),$$

where

$$\begin{aligned}\mathcal{L}_{\mathrm{TL}}(\theta) = &\sum_{i=1}^{N_p} ||\phi(y_i^p;\theta) - y_i^s|| \\ &+ \alpha \sum_{i=1}^{N_p} \sum_{j=1}^{D-w+1} ||\phi(y^p_{[j-w\ j+w]};\theta) - y^s_{[j-w\ j+w]}||,\end{aligned} \tag{6.9}$$

w is the size of the sliding window, D is the number of detectors, α is a hyper-parameter that is used to control the contribution of the windowed mean absolute error loss. At inference time, the final reconstructed and diagnosis results are computed as:

$$\hat{x}^* = \mathcal{F}^{-1}(\phi(y^p)). \tag{6.10}$$

6.3 Results

We present results on both in-silico and clinical data. Results were obtained by training the model on the in-silico data. A transfer learning network, adapted from Chapter 4 and trained on a phantom dataset, bridges the distributions shift that is unavoidable when switching between in silico and real world data. A Gaussian noise was added to the signal, mimicking real world signal fluctuation, to improve model robustness to sensor noise and mimic the real-world drift of device characteristics on different clinics in between calibrations. This noise model depicts the highly variable noise to each individual detector as caused by sensor noise and interference of refracting light. Consistent with previous work [282, 285, 84], we set $\sigma = 10\%$ of the maximum sensor value. Besides the simulated noise, the probe

accounts for ambient light, the predominant source of noise, as well by capturing a frame without any active emitters and then subtracting it from the actual data measurement, taken during clinical tests, prior feeding it into our model. Performance evaluation captures image reconstruction quality, diagnosis accuracy, and speed. The next section provides details.

6.3.1 Experimental Design

Dataset

Similar to Chapter 3, we simulate light propagation into tissue at different light wavelengths, 690, 750, 800, and 850 nm, illuminating the tissue sequentially, using the physics-based Toast++ software [221]. We collect training samples from synthesized tissues with known optical properties and labels. A set of 2D images with various lesion sizes, shapes, and positions is synthesized. In order to mimic real breast tissue optical parameters', we base the optical properties on realistic optical coefficient values [147, 245] as summarized in Table 6.4. A total of 4000 sample data pairs (256-D $\times$ 4 vectors, 2D image, label) are used to train and test our method. Each sample includes the collected measurement vectors, one for each frequency, the ground truth image, and the diagnosis label. Training dataset size is chosen as a compromise between training time and in-silico performance. We focused on the diversity of simulated scenarios while also being mindful of computational resources. While we note that in silico and phantom data can result in very large datasets, we focused on the diversity of simulated scenarios and adopted a dataset with various lesions number, size, and depth to emulate realistic conditions where the optical properties of the anomaly and surrounding tissues were taken from available in vivo breast tissue experimental data.

Our DOB-probe(Chapter 3, Section 3.2.2) [229, 101] was used to collect real patient data to test our method. Note that the wavelengths share variable overlap in the spectrum (Fig. 3.2) [245], motivating further the need for orthogonal encoding. To train the transfer learning module breast-mimetic phantoms, with known inhomogeneity locations, and DOB-probe were used to collect measurements (Chapter.3).

Following the ethics and institutional review board approval protocol, clinical data were collected from 9 participants diagnosed with breast tumors [230]. In a normal clinical pre-screening exam, a breast is usually divided into four quadrants, and different measurements are collected on each quadrant. Given that the used probe is in clinical trials [229, 226, 282], patients with known cancer localization are considered, and sweeps over the lesion location and the opposite healthy breast are collected. This step was essential to proving that the technology we introduced works well with human tissue. For each patient, height, weight, age, and gender, as well as details of the subjects' breast cancer, briefly summarized in Table 6.5, were recorded. Patients were placed in a supine position, and scans at multiple points over the lesion location and healthy breast were collected. On average, four different measurements (scans) were taken on each breast. Even though no reconstruction ground truth is available for real-world data, it is invaluable to detect robustness and real-world per-

Table 6.5: Summary of clinical data

	Tumor Position	Tumor size (cm)	Tumor Type
Patient 1	Left Breast	$1 \times 0.8 \times 0.7$ $2.5 \times 0.8 \times 0.8$	BI-RADS 7
	Right Breast	$1.1 \times 0.8 \times 0.7$	Benign
Patient 2	Left Breast	$2.2 \times 1.7 \times 1.7$	BI-RADS 4
Patient 3	Left Breast	$1 \times 1 \times 1$	Non-invasive ductal
Patient 4	Left Breast	$2.5 \times 1.7 \times 3.5$	BI-RADS 5
Patient 5	Right Breast	2.4	BI-RADS 4
Patient 6	Right Breast	$2.3 \times 2.2 \times 1.5$	BI-RADS 4
Patient 7	Left Breast	$1.7 \times 1.4 \times 1.2$	BI-RADS 5
Patient 8	Left Breast	$1.6 \times 0.8 \times 0.8$	BI-RADS 5
Patient 9	Right Breast	$2.2 \times 2.1 \times 2.3$	Invasive ductal

formance, with partial ground truth known from other modalities on the same patients. The precise location, size, and type of the tumor lesion were determined via mammography, ultrasound, or biopsy. Note that these details were only used for model performance evaluation and metrics calculation, while raw sensor signal only was inputted to the model. Another advantage of our direct prediction approach is that the absence of pixel-wise ground truth is less problematic compared to reconstruction based classification, as only the diagnosis label is required.

Implementation

Models were implemented in the Keras TensorFlow framework and trained for 100 epochs on an NVIDIA Titan X GPU. By optimizing the model's performance on the validation set, we set all hyper-parameters as follows: batch size to 16, learning rate to 10^{-4}, optimizer set to Adam, and initialization to Xavier. Early stopping was used if the validation loss had not improved within 10 epochs. The in silico data was divided in a 80/10/10% training/validation/test split, and hyper-parameters β (6.5), α and D (6.9), and γ (6.6) were set to 0.2, 0.5, 4 and 0.5, respectively. The fully connected units for the fusion branch were set to 32 and 16 for l and l2, respectively.

Evaluation metrics

To quantify the models' robustness, we look at (i) lesion localization error (Loc. Error); (ii) peak signal-to-noise ratio (PSNR); (iii) structural similarity index (SSIM); and (iv) Fuzzy Jaccard for reconstruction quantification, while the balanced accuracy (BA), F1 score (F1), precision P, recall R, Matthews correlation coefficient (MCC), and confusion matrix are

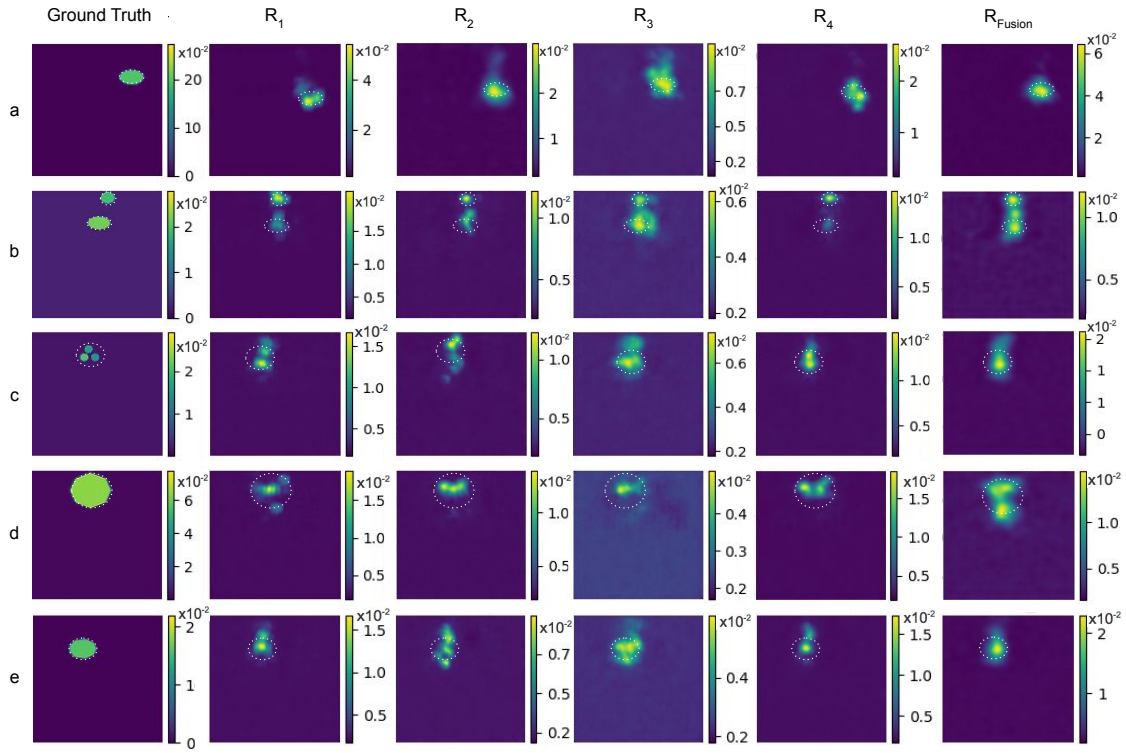

Figure 6.3: Qualitative reconstruction performance of absorption coefficients using the FuseNet++ on in silico samples with varying ground truth lesion sizes, locations, and numbers. Our multi-spectral results (R_{Fusion}) show an overall superiority in terms of generally improved background/foreground contrast and a better differentiation between lesion sizes and lesion localization compared to per-frequency reconstruction results (R_1 to R_4) at wavelengths 690, 750, 800, and 850 nm, respectively.

reported for the classification task quantification.

$$
\begin{aligned}
BA &= \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right), \\
P &= \frac{TP}{(TP+FP)}, \\
R &= \frac{TP}{(TP+FN)}, \\
F1 &= 2\frac{P*R}{P+R}, \\
MCC &= \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}.
\end{aligned}
\tag{6.11}
$$

True positive (TP) is the number of correctly predicted samples as positive, while false positive (FP) is the number of wrongly predicted samples as positive. False negative (FN) is the number of wrongly predicted samples as negative, while true negative (TN) is the number of correctly predicted negative class samples over the number of classes in the

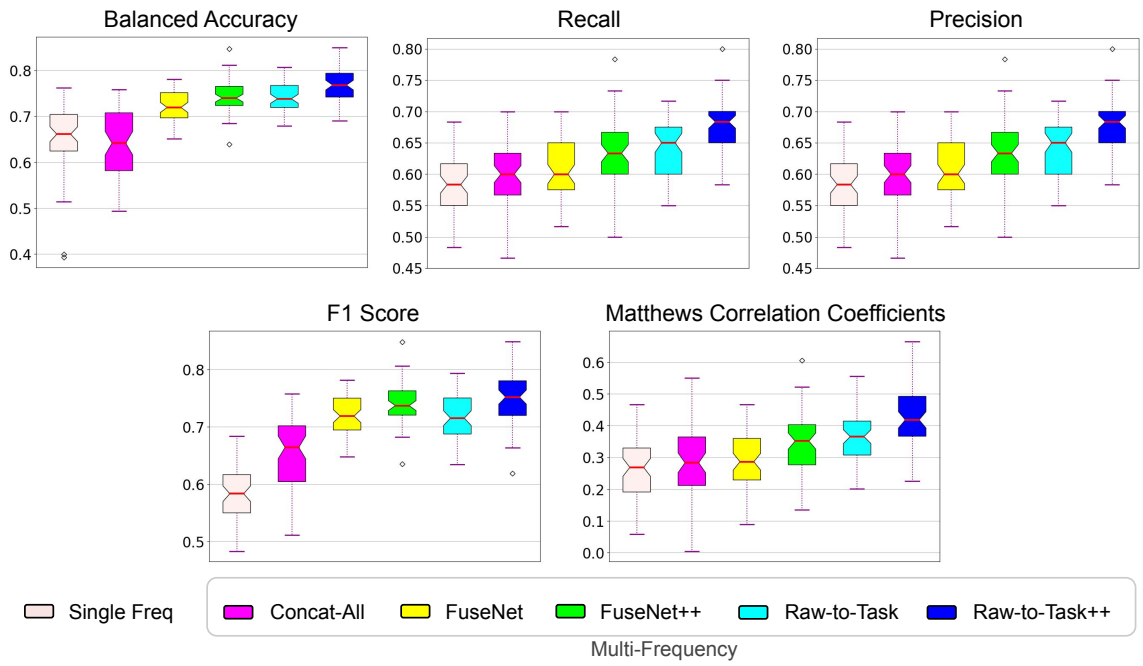

Figure 6.4: Quantitative diagnosis performance of different models when one vs multi-frequency are used. Overall results show improved prediction performances in multi-frequency models. Note the significant improvement when FuseNet is used compared to a simple concatenation (Concat-All). Results using the FuseNet++ enforce the benefit of feature space orthogonality. Raw-to-task++, in which all network capacity is dedicated to the end task, shows an overall performance gain.

Table 6.6: Quantitative results on in silico test dataset.Losses are defined in Section 6.2-B; Loc.error: lesion localization error; PSNR: peak signal-to-noise ratio; SSIM:structural similarity index; BA: balanced accuracy; F1: F1 score. †: value not supported by method, ‡: image reconstruction skipped.

	Loss			Loc. Error	PSNR	SSIM	Fuzzy Jaccard	Runtime	BA	F1
	L_{REC}	L_{CE}	L_{OPL}	(pixel, ↓)	(dB, ↑)	(↑)	(↑)	(ms, ↓)	↑	↑
Single-Freq	✓	✓	†	17.7 ± 21.9	19.1 ± 4.8	0.80 ± 0.05	0.60 ± 0.17	23	0.65	0.65
Concat-All	✓	✓	†	20.4 ± 18.4	19.6 ± 6.2	0.73 ± 0.17	0.61 ± 0.18	28	0.63	0.65
FuseNet	✓	✓	-	17.6 ± 23.3	20.2 ± 4.1	0.88 ± 0.05	0.62 ± 0.19	31	0.72	0.72
FuseNet++	✓	✓	✓	**15.7 ± 12.7**	**21.2 ± 4.4**	**0.89 ± 0.03**	**0.64 ± 0.18**	32	0.74	0.74
Raw-to-Task	†	✓	-			‡		**15**	0.74	0.72
Raw-to-Task++	†	✓	✓			‡		**15**	**0.77**	**0.75**

prediction tasks. Recall quantifies the number of positive class samples properly identified by the model, while precision measures the number of correct positive predictions made by the model. BA, used when quantifying performance on imbalanced data, measures the average accuracy obtained from all classes. MCC measures the quality of multi-class classifications and is informative in cases of skewed class distributions.

For the computational cost at inference, we quantify the forward pass of the model, measured in ms per example.

To evaluate the performance of our models, we contrast the results when using one frequency with many frequencies in the FuseNet and the Raw-to-task model. We present results on in-silico data and clinical data.

6.3.2 Results on Synthetic Data

Trained on the in silico data and tested on a separate test set of 240 images, we compare the reconstruction and prediction performance of our FuseNet and the prediction performance with the Raw-to-Task counterpart.

Joint Reconstruction and Diagnosis

Figure 6.3, illustrates reconstruction results on selected in silico samples with different lesion sizes, numbers, locations, and depths. In order to offer clinicians more details, results based on each frequency separately (R_i) as well as results that use all frequencies are shown, with the latter showing more consistent performance. The joint model successfully exploits the presence of the different frequencies and generally shows an improved background/foreground contrast. For example, the difference in signature for 3 small but proximate lesions is marked in different frequency results (R_1 to R_4) (row c), while a more accurately reconstructed sphere size is provided by the fusion result R_{Fusion} in row (d). Detecting heterogeneity in lesions is critical for correct treatment estimation given that it is a proxy indicator of evolutionary pressure in the lesion, selecting for more resistant cancer sub-populations. Table 6.6 presents the quantitative results of the ablation study, where the contribution of different losses and modular choices of the architecture to model per-

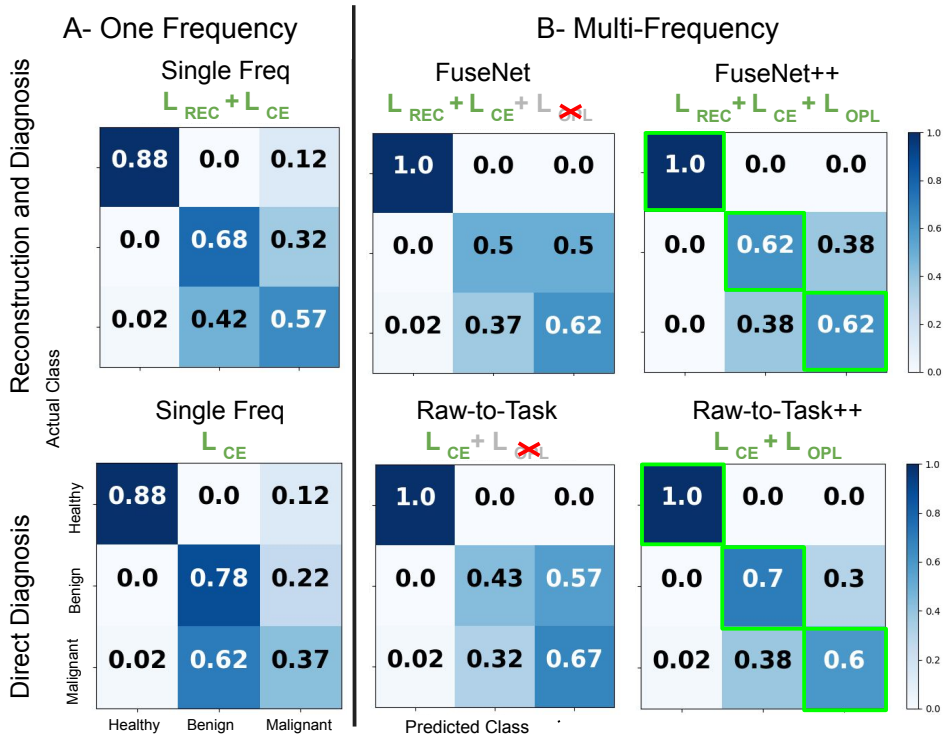

Figure 6.5: Diagnosis prediction confusion matrices when (A) one vs (B) multi-frequency inputs are used. Note the improvement in accuracy of unbiased lesion classification (benign, malign) vs healthy when multiple frequencies are used, as illustrated by the higher values along the diagonal. Results of FuseNet++ highlight the benefit of encouraging orthogonality in enhancing benign vs malignant separability while reducing healthy false negative. Raw-to-task++ further improves separability at the expense of minimal false negative (2%).

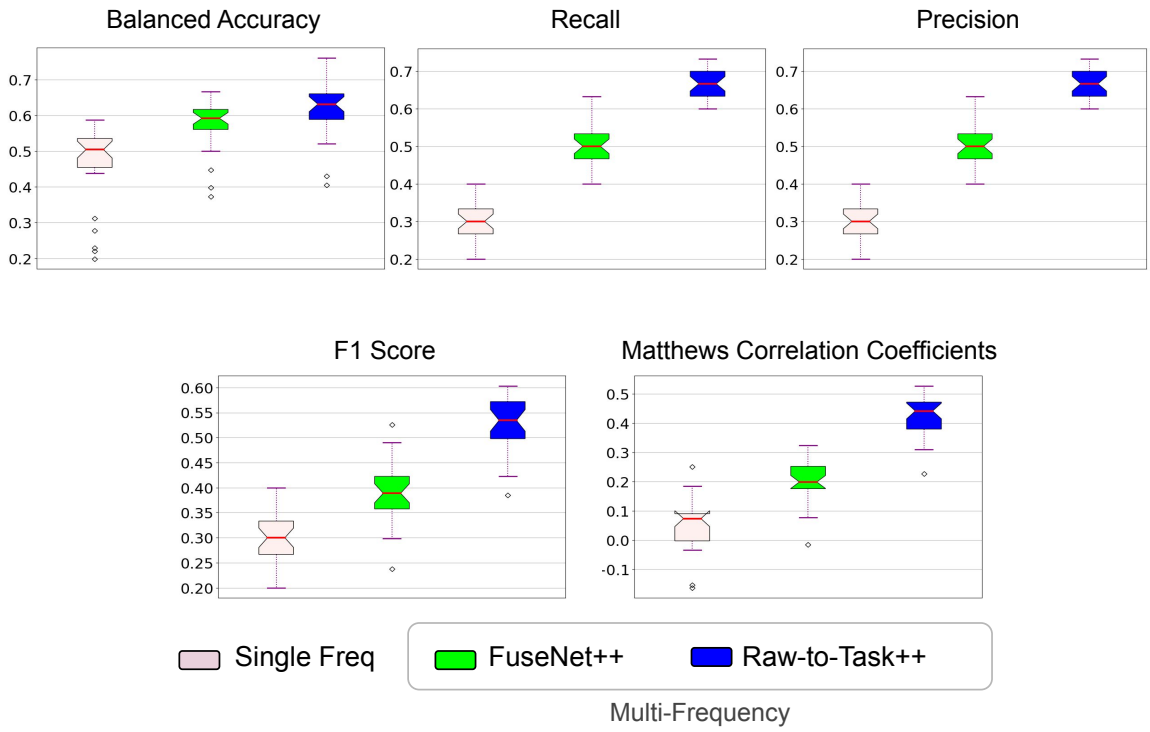

Figure 6.6: Quantitative diagnosis performance when one vs multi-frequency are used on clinical dataset. Overall results show improved prediction performances in multi-frequency models compared to single frequency and indicate that image reconstruction is not necessary for unbiased classification and can even lead to biased results. Note the marked improvement when Raw-to-task++ is used compared to FuseNet++.

formance are quantified. Rows 1 to 4 highlight the benefit of using multi-frequency fusion on the reconstruction task. A naive multiple frequencies concatenation will not necessarily improve results, which agrees with the findings reported by Applegate et al.[11], illustrating the impact of adding noisy dimensions on performances. Nonetheless, we see improved results for FuseNet. When fusion branch and L_{OPL} are used jointly (FuseNet++), the features contribution from each frequency is maximized in contrast to simple features concatenation (Concat-All) at the price of a minimal computational increase (only 9%).

Prediction performance highlighted in Table 6.6 and Fig. 6.4 show an overall improvement when more input frequencies are available, with a boost in performance when FuseNet and FuseNet++ are used. Confusion matrices (Fig. 6.5-A,B) show a clear discrimination between healthy and lesion features when more data, in the form of more frequencies, is available. Further, improved benign and malignant discrimination is observed when feature orthogonality is leveraged (Fig. 6.5-B) as well as a reduction in healthy false negative.

Direct Prediction

In Figure 6.5, similarly to the joint model, the direct prediction model results using a single frequency input (SF-DP) (Fig. 6.5-A) are contrasted with raw-to-task prediction results using multiple frequencies as input (Fig. 6.5-B). A clear discrimination between features is apparent when more data, in the form of multiple frequencies, is available, especially when discriminating between healthy and lesion; the primary application in DOT-based screening deployments. Raw-to-task model significantly reduces computational complexity (Table 6.6-Runtime), enabling lower latency and higher throughput in real medical settings. Next, we tested the contribution of individual loss function terms and architecture component on overall diagnosis performance. Figure 6.4 shows the diagnosis performance on the test set for the best value of γ and highlights the benefits of the feature orthogonality constraint in breast cancer diagnosis, where tumoral and non-tumoral breast lesion differentiation is challenging. Contrasting FuseNet++ and Raw-to-task++ (Fig. 6.4-6.5) illustrates performance gain when all network capacity is dedicated to the end task rather than intermediate ones.

6.3.3 Results on Clinical Data

Figure 6.7 presents the reconstruction performance on breast scans from patients diagnosed with breast tumors. The probe is placed close to the likely location of each identified lesion, and a set of scans are made. The opposite healthy breast, for each patient, is scanned as a contrastive reference. Weak labels were attributed to each set of scans regardless of the probe's closeness to the tumor localization. As a partial ground truth, patients underwent mammography and/or Ultrasound scans to obtain estimated lesion dimensions and biopsies to confirm tumor type. While lesions are accurately reconstructed in most cases, as shown in Fig. 6.7, with clear foreground and background discrimination in R_{Fusion} as well as R_1 to R_4, healthy cases, capturing only background readings, highlight a better robustness of orthogonal fusion, R_{Fusion}, to noise.

Figure 6.6 reports quantitative prediction performance on single vs. multi-frequency data and highlights the overall improved performance when more frequencies are used. Note the biased classification results when image reconstruction supervises the prediction task, FuseNet++, compared to direct prediction from raw data, Raw-to-task++.

The confusion matrix, Fig. 6.8, shows improved discrimination between healthy and lesion features with the raw-to-task model. If we consider that a key feature of the reconstruction based classification is the interpretable angle of such results, we note that the raw-to-task model has the added advantage, in addition to improved performances, that it omits potentially confounding explanations, where reconstruction artifacts can mislead experts. Indeed, in recent work on explainable artificial intelligence, such confounding explainers were identified as a roadblock [131]. Table 6.7 reports raw-to-task model diagnosis

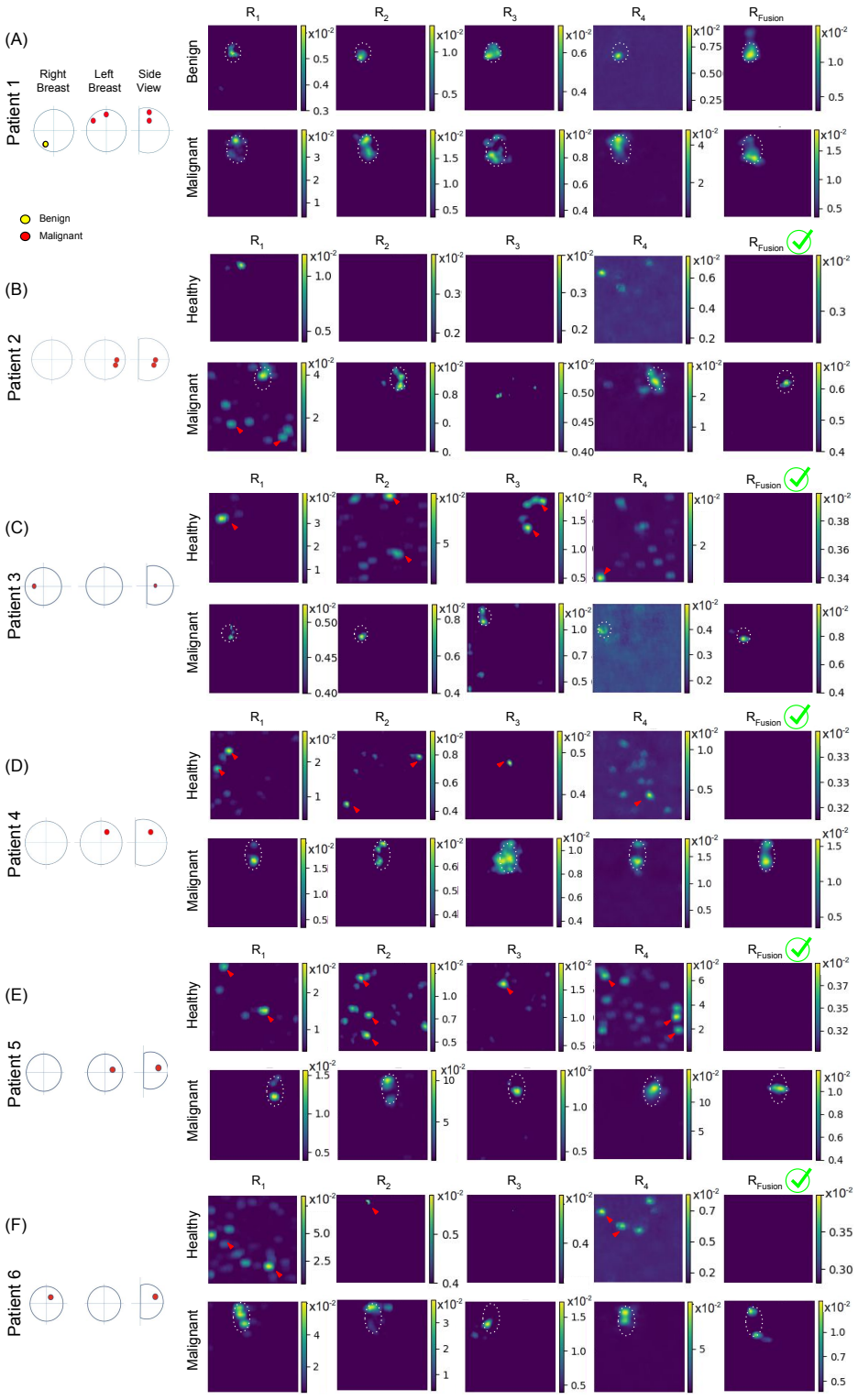

Figure 6.7: Qualitative reconstruction results in clinical patients with benign and malignant tumors. Approximate lesion sizes and locations were obtained with joint modalities (details in Table 6.5). Note, in (A), the ability of FuseNet++ to reconstruct lesions, while, in (B-F), the robustness of orthogonal fusion to noise (R_{Fusion}) compared to (R_1 to R_4) (healthy row) is highlighted.

Table 6.7: Quantitative results on clinical dataset using Raw-to-Task++

	Precision	Recall	F1-score	Number of scans
Healthy	0.71	0.65	0.68	32
Benign	0.11	0.5	0.18	2
Malignant	0.78	0.68	0.73	44
Weighted-Avg	0.73	0.66	0.69	78

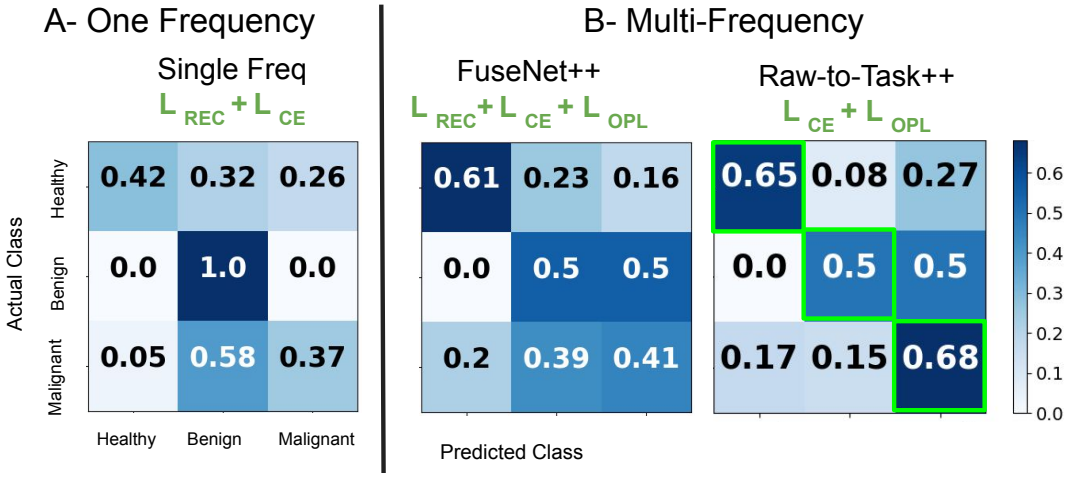

Figure 6.8: Clinical data diagnosis prediction confusion matrices when (A) one vs (B) multi-frequency inputs are used. Note the improvement in accuracy of lesion classification in the Raw-to-Task model, despite the imbalance in the data.

performances on each data class to highlight the clinical data imbalance compared to a balanced training data scheme.

In Figure 6.9, we illustrate some failure cases at the limit of detection capability, with false positives (Fig. 6.9-A,B, marked with red triangles) and false negatives (Fig. 6.9-C). For screening purposes, false negatives are more critical; false positives would eventually be resolved by follow-up diagnosis. Note that the discriminatory power of the detection is limited by tumor depth, shape, and noise level. It may require several scans over breast tissue in order to be captured. The failure cases here are from a single scan measurement only, not aggregated

Although the transfer learning network, trained using phantom data, bridges to some extent the disparity between in silico (training) data and real-world data, its performance on clinical data reveals that it can still be misled by significant real-world variations, such as differences in illumination and noise levels. Additionally, since each tumor is unique, tumor heterogeneity can result in distinct acquisition signatures that may not be present in the training data. These failure cases highlight the need for more clinical data (patient data) to better train the transfer learning module. Current results present a proof of concept, where validation on larger and more diverse datasets is still required.

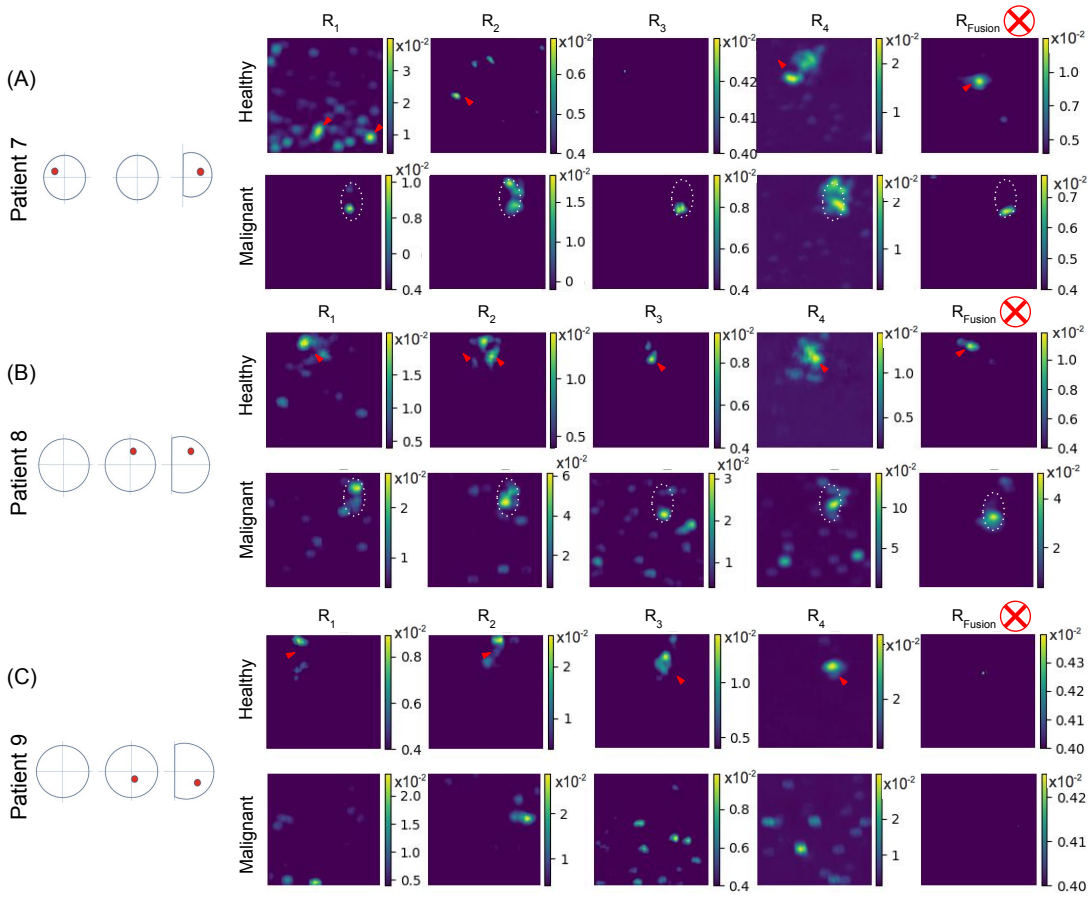

Figure 6.9: Examples of reconstruction failure cases. (A,B) highlight false positive reconstruction cases, marked with red triangles, that remain less critical than false negative cases where a tumor is missed (C). Note the noisy reconstruction in R_1 to R_4, suggesting a quite noisy input signal.

6.3.4 Effect of Lesion Localization on Accuracy

We quantify the effect of lesion location on lesion detection accuracy in Figure 6.10, where we classify whether a lesion is present or not. The penetration depth into breast tissue is approximately half the distance between the source and detectors [236], ∼2.5 cm for our DOT probe. Our results confirm the expected reduction in lesion detection accuracy as the lesions decrease in size or increase in depth.

6.4 Discussion

In order to be an effective tool in clinical settings, a clinician's trust is essential. A combination of good performance, as quantified by accuracy and other metrics, and an interpretable model increases trust. Neither deep learning based reconstruction nor classical iterative algorithms provide a path from pixel to sensor value in a way that a clinician can

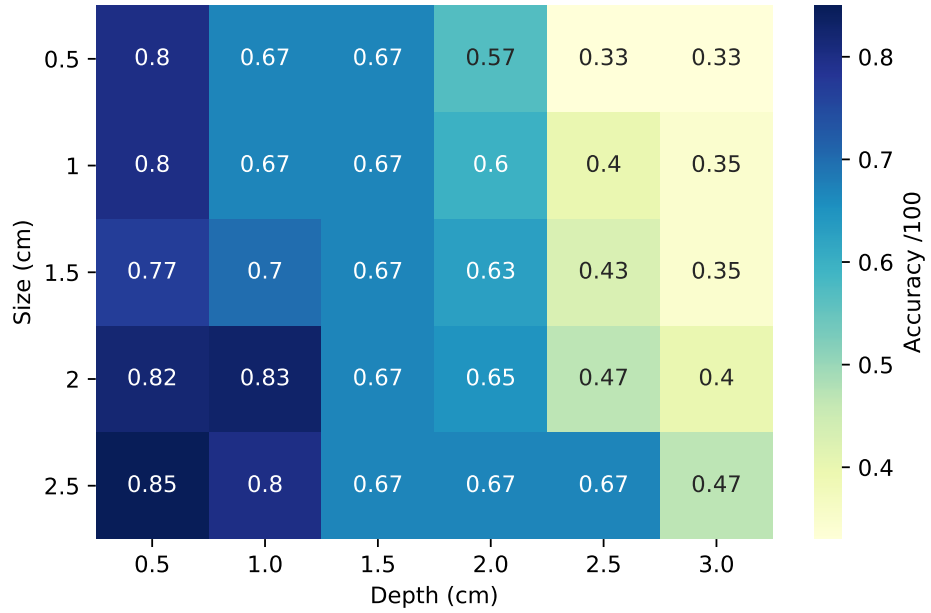

Figure 6.10: Effect of lesion depth and radius on model prediction accuracy. Note how the more superficial (closer to the skin surface) and larger lesions are more accurately detected.

easily understand. While a reconstructed image may seem to increase interpretability, it is typically not created in an interpretable way and is not necessarily causally related to the classification decision. Omitting the reconstructed image, while increasing performance, would not therefore reduce the trust a clinician has in our direct to task contribution.

Cancer treatment regimens, especially for treatment-resistant lesions, are shifting towards adaptive or dynamic treatment models, such as the recent game theory-driven treatment of resistant prostate tumor patients [268]. However, these require accurate, unbiased, and specialized task-specific models. Our raw-to-task approach can be extended to develop models specializing in multiple tasks, not just diagnostics. Examples are prediction of lesion type, progression, localization, and tumor heterogeneity, all the way to successful treatment regimens ahead of time, paving the way for adaptive personalized medicine and disease management [2].

A key focus of this work was to leverage orthogonality in mitigating confounding factors induced by multi-frequency fusion. However, as noted as early as 1936 by Fisher et al.[86], orthogonal representations need not be informative, and thus, in a deep learning setting can also lead to orthogonal or independent encodings that are less or uninformative, as we encountered in our own experiments. The heterogeneity of lesions, especially malignant ones, ensures that no two malignant lesions will likely be the same, thus driving the need for diagnostic capability that focuses on identifying the diverse lesion types, not necessarily the reconstructed image.

6.5 Conclusion

We introduce deep learning based multi-frequency orthogonal fusion for diffuse optical tomography with end-to-end classification of malignancy of breast lesions. Orthogonal fusion of multi-frequency improved both image reconstruction quality and accuracy of tumoral and non-tumoral breast lesions' discrimination. In addition, we show that raw-to-task learning can improve classification without requiring reconstruction in a real time setting.

Chapter 7

Conclusions, Limitations, and Future Directions

7.1 Thesis Summary

Medical imaging is an invaluable resource in medicine as it enables scientists and physicians to peer inside the human body and provides a wealth of information indispensable for understanding, modelling, diagnosis, and treatment of diseases. Despite their impressive advancements and their greater applicability, several improvements in medical imaging can still be envisioned, such as reducing acquisition complexity and processing time to increasing clinical accuracy and throughput, particularly given that total waiting times for medical exams have at least doubled over the last two decades [174]. Furthermore, improving emerging less invasive imaging modalities and the deployment of biomedical imaging in low power handheld devices can potentially reduce the burden on the healthcare system by prioritizing more advanced or invasive diagnostics for acute cases, enable quick monitoring of patients, and facilitate personalized medicine.

Throughout the presented studies in this Thesis, the emphasis was on developing novel and efficient deep learning algorithms specifically tailored for DOT based breast cancer screening, effectively addressing the ill-posed and highly nonlinear nature of the DOT inverse problem in real-time and resource constrained imaging device scenarios. Chapter 2 reviewed the literature on deep learning based methods in the domain of biomedical image reconstruction in general and observed that several studies now report quality improvements when compared to conventional reconstruction techniques, especially in the context of noisy and limited data, and highlight how the field is moving toward producing higher quality images and accurate diagnosis given the most constrained resource budget, scanning time, scanner complexity, and lowest power consumption.

Chapter 3 presents the first published work to evaluate the performance of DL-based method in DOT image reconstruction. While trained on in silico data, validation was confirmed on real-word phantom collected data, subject to sensors non-ideality and noise, and

demonstrated improved performance compared to SOTA conventional approach. To overcome the challenge of limited available data in DOT, data augmentation and transfer learning, tackling domain shift, were applied, with success, in Chapter 3 and Chapter 4. Deep models were trained on large synthetic datasets generated using physics-based simulations and fine-tuned for the specific DOT signal. This approach increased the diversity and size of the training dataset, leading to improved generalization and robustness of the proposed DL models.

Regularization techniques were incorporated in Chapter 4 and Chapter 5 to improve the quality and accuracy of reconstructed images, particularly in limited-angle settings. To improve lesion localization, in Chapter 5, we were the first to utilize distance transform based regularization to promote spatial-wise similarity as well as attention mechanisms to focus on important regions and features, enhancing the accuracy of image reconstruction.

A pathology-based optical signature was proposed in Chapter 6, where a direct clinical prediction was learnt from the raw signal whilst skipping the image reconstruction model. While multiple frequencies can be combined to optimize signal penetration and reduce sensitivity to noise, they share some overlap in terms of the absorption spectra of the major tissue constituents in the red-near infrared range. This potential overlap enables the capture of common information and correlation. To enable better differentiation and quantification of different tissue components and improve both image reconstruction and the accuracy of discriminating tumoral and non-tumoral breast lesions, an orthogonal fusion of multi-frequency data was proposed.

To assess the performance of the algorithms, we employed various image quality assessment metrics such as PSNR and SSIM, as well as evaluated computational complexity. Beyond using standard metrics, we emphasized the critical importance of clinical outcomes by focusing on domain-specific metrics. The evaluation of our methods considered patient diagnosis outcomes, which remain the most valuable measure in the medical imaging domain. We compared the reconstructed images with their in-silico counterparts and evaluated their generalization to real acquired signals in phantom and clinical settings, subject to real-world conditions. The accuracy of tumor and non-tumor breast lesion discrimination was assessed for both the classical and direct prediction pipelines.

7.2 Limitations and Future Direction

The proposed deep learning algorithms have demonstrated notable improvements in both speed and accuracy compared to conventional methods for DOT image reconstruction and classification tasks. One key advantage of deep learning-based algorithms is their ability to leverage implicitly learned feature encodings from the sensor data. However, some open problem in DL remain, namely the requirement for a substantial amount of data to ensure proper generalization, the time-intensive nature of training processes, and the difficulty

of determining optimal hyperparameters for the networks, particularly in multi-objective optimization settings [175]. Here, we summarize selected challenges of high interest, together with promising future directions in this field.

Class imbalance: Medical datasets often exhibit imbalances in the number of samples between different classes and various disease categories, leading to biased training and suboptimal results, especially for the minority class. The imbalanced distribution of samples across different classes can lead to biased model training and a tendency to favor the majority class, neglecting the minority class. Potential approaches to tackle class imbalance include data augmentation techniques to create synthetic samples and exploring techniques like transfer learning. Additionally, using class weights during model training can also help alleviate the impact of class imbalance [23, 49]. Careful evaluation of the model's performance using appropriate metrics [206] can provide a more comprehensive understanding of its effectiveness in handling class imbalance and its overall predictive capability.

Generalization to out-of-distribution data: DL-based DOT models presented in this thesis are trained on in-silico datasets and therefore may not fully capture the wide range of variations and complexities present in real-world scenarios. As a result, when exposed to data that significantly differs from the training distribution, for example if a specific pathology or rare feature is underrepresented or not adequately captured in the training data, or if the data collection exhibits diverse clinical practices, variation in clinical devices or difference in data population, the performance of deep learning-based methods may be degraded. Due to ethical and resource constraints, the real-world patient based dataset size remains limited, and real-world tumor heterogeneity [28] can lead to markedly different acquisition signatures that are unlikely to be present in the simulated training data. Increasing clinical dataset size and testing on carefully curated benchmarks can be a potential remedy to better test DL-based DOT methods and their generalization to out-of-distribution data, although increased data set size does not necessarily ensure improved unbiased performance on minority cases which can have important medical implications. Another trend present in recent art is moving towards self-supervised learning [153] and unsupervised [311] approaches. Additionally, exploring domain generalization techniques like learn domain-invariant features [95], learning from Fourier Transform phase information [275], or improving shortcut learning and feature suppression by measuring feature culpability [26]. Hybrid learning, the concept of reconstructing an image using a secondary modality as prior [187], and continual learning, adapting to changing environments and learning from new data while retaining previous knowledge [71, 199], show promise in this regard as well.

Hallucination: Deep learning models can be prone to artifacts in the reconstructed image. Hallucinations can be especially problematic when dealing with sparse or noisy input data. Although hallucinations, within the context of synthesizing natural images, may be deemed acceptable and even desirable for certain applications such as art, gaming,

and entertainment, particularly in the realm of generative AI, where the emphasis is on artistic, imaginative, or visually pleasing elements rather than strict realism [190, 194], it is crucial to recognize the potential risks in other critical domains. Specifically, in medical applications like tumor diagnosis, where hallucinations can lead to erroneous diagnoses and misleading interpretations [62], and traditional error maps can be insufficient for visualizing and detecting specific hallucinations [35]. The real-world consequences of misdiagnoses can be substantial, impacting patient outcomes and healthcare decisions. While this remains an ongoing challenge, one potential solution is to shift focus away from interpretability by omitting intermediate results, as demonstrated in end-task-specific learning, as proposed in Chapter 6. The potential for reduced trust by the end user is offset by *'trust by results'*: if a diagnosis is more accurate without a reconstructed image, is the image needed at all?

Benchmarks and reproducibility: While deep learning methods in DOT show promise, the lack of open benchmarks and open-source code hinders the evaluation and reproducibility of results. There is a growing need for researchers to embrace open science practices, which encompass encouraging the open sourcing of code [155] and datasets linked to their studies, all while preserving patient privacy [128, 188]. Additionally, creating standardized benchmarks and evaluation metrics can accelerate fair comparisons between different methods and promote advancements in the field [168].

Interpretability and confidence estimation: Deep learning models are often considered black boxes, which limits their adoption in medical imaging applications. Future research can focus on developing methods to enhance the interpretability and explainability of deep learning-based DOT models. This involves visualizing and understanding the learned representations, attributing the model's decision-making process, and providing confidence measures or uncertainty estimates to enhance the trustworthiness of the results. Models based on the iterative unrolling paradigm show promise in this regard [177, 303]. Being able to understand and trust the model's decisions is essential in clinical decision-making and patient care [135, 136]. Uncertainty estimation methods can also be used to generate additional confidence information for end users [54, 214].

Computational complexity: Deep learning frequently require computationally powerful hardware (GPU) to provide real-time results. Network pruning and sparsifying, recently proposed for computer vision tasks [6, 4, 126], presents a promising direction yet to be explored on image reconstruction tasks in order to allow DL-based model processing on CPU and mobile devices. This will be of great interest to emergent mobile scanners (e.g., DOT [228], US [92], CT [212], MRI [113]). Reducing the encoding size in bits of network weights without altering the quality of the prediction has been demonstrated on several reference deep learning image processing networks [239]. A reported decrease in training time of 30-60% is only one benefit. By reducing network weights from 64 or 32 bits to 8 or smaller (i.e. weight quantization) the network requires a smaller memory footprint. As a result, networks that are, at the time of writing, too large to fit on a single GPU can

be reduced to fit on a single GPU. Conversely, networks too large to be deployed on edge devices (handheld scanner, mobile phones, etc.) can become easily deployable in the field without changing their architecture.

Enhanced architecture design: Combining the ideas presented in this thesis with the latest advances in architecture design allows for further improvements and offers another avenue to enhance our contributions to breast cancer prescreening using DOT. For instance, neural architecture search (NAS) enables an automated design of neural network architectures tailored to the specific tasks [288]. The application of NAS ensures that the architecture is finely tuned for modality's unique requirements, especially in the context of resource-constrained and energy-sensitive [165], thereby maximizing its potential in medical imaging. Another possible suggestion is to leverage the diffusion process, which can capture the spatial dependencies and improve the quality of reconstructed images [61, 304]. Moreover, Transformer-based architectures, well-suited for modeling temporal dynamics and handling multi-modal information, hold promising prospects, particularly in light of the reduced data-hungry constraint[46, 77]. Transformers' attention mechanisms enable them to effectively capture complex spatial dependencies and long-range interactions within the data, leading to potential improvements in the overall diagnostic performance.

Multimodal data fusion: Joint multi-modal image reconstruction, such as DOT/CT [21] and PET/MRI [257], has been proposed on several iterative approaches to take advantage of both imaging modalities and led to improving the overall imaging performance, especially avoiding the spatio-temporal artifacts due to scanning with different devices at different times and positions. The idea relies on leveraging the information provided by feature similarity between multiple modalities and complementary information for more accurate and robust end task. While this direction has great interest, it has only just begun to receive consideration [65, 154, 187, 311] and remains a direction for further exploration. Another option is to guide the training of an inferior modality, such as DOT, by a superior modality like CT [167]. Collecting suitably calibrated and registered data on hybrid multi-modality imaging systems remains a key challenge as well. Furthermore, language models present a unique opportunity for expanding multi-modal learning by effectively leveraging existing medical image and text report datasets [202, 264]. This avenue holds great promise in enhancing the fusion of information across modalities and represents a novel direction for exploration.

Adaptive sampling: Attention driven sampling received increasing interest recently especially in a limited data representation context. While adapting the sampling to the reconstruction algorithm showed improved image quality compared to conventional sampling strategies [134], it could be computationally more expensive with uncertain convergence behavior. The development of efficient sampling learning algorithms would be a promising research direction. Recently proposed neural fields-based models open new perspective towards adaptive sampling by moving from classical discretization of the object area/volume

into pixels/voxels at a pre-determined resolution, for modelling diffuse light propagation, to continuous encoding of the optical parameters. This transition opens up new possibilities for adaptive sampling by enhancing the spatial resolution of the reconstruction [207]. This direction presents an exciting opportunity to advance adaptive sampling techniques in the realm of diffuse optical tomography.

Domain specific metrics: The majority of performance evaluation metrics are often oriented towards computer vision tasks, which can render them inadequate for the complex landscape of biomedical imaging. These metrics may lack the ability to capture the true diagnostic accuracy, improved explainability [135], enhanced patient management, efficient triaging, cost-effective screening [167], and more that are integral to the medical domain. We advocate a shift toward medical-oriented evaluations [1, 2, 264] given that reconstructed images are primarily used for a specific diagnosis or treatment purpose. While such measures may be expensive, especially if they require human expert feedback, they will be critical in creating algorithms that can advance the biomedical imaging practice.

It is fascinating to witness how the progress in deep learning has a broad impact across various applications and disciplines, ranging from domain generalization and explainable AI (XAI) to continual learning, neural architecture search (NAS), and more. These advancements hold the potential to greatly enhance research on DOT reconstruction and analysis beyond the scope of this thesis. Moreover, we believe that the contributions made in this thesis will serve as a source of inspiration for further extensions and innovations in other application areas as well. By leveraging these cutting-edge techniques and insights, we anticipate a significant advancement in the field of DOT and its potential to positively impact numerous domains.

Bibliography

[1] Kumar Abhishek and Ghassan Hamarneh. Matthews correlation coefficient loss for deep convolutional networks: Application to skin lesion segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 225–229. IEEE, 2021.

[2] Kumar Abhishek, Jeremy Kawahara, and Ghassan Hamarneh. Predicting the clinical management of skin lesions using deep learning. *Scientific reports*, 11(1):1–14, 2021.

[3] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018.

[4] Alireza Aghasi, Afshin Abdi, Nam Nguyen, and Justin Romberg. Net-trim: Convex pruning of deep neural networks with performance guarantee. In *Advances in Neural Information Processing Systems*, pages 3177–3186, 2017.

[5] Issa Al-Shakhrah and Tariq Al-Obaidi. Common artifacts in computerized tomography: a review. *Applied Radiology*, 32(8):25–32, 2003.

[6] Simon Alford, Ryan Robinett, Lauren Milechin, and Jeremy Kepner. Pruned and structurally sparse neural networks. *arXiv preprint arXiv:1810.00299*, 2018.

[7] Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292, 2019.

[8] Murad Althobaiti, Hamed Vavadi, and Quing Zhu. Diffuse optical tomography reconstruction method using ultrasound images as prior for regularization matrix. *Journal of Biomedical Optics*, 22(2):026002–026002, 2017.

[9] Mirella L Altoe et al. Diffuse optical tomography of the breast: a potential modifiable biomarker of breast cancer risk with neoadjuvant chemotherapy. *Biomedical Optics Express*, 10(8):4305–4315, 2019.

[10] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction-does AI come at a cost? *arXiv preprint arXiv:1902.05300*, 2019.

[11] Matthew B Applegate, Carlos A Gómez, and Darren Roblyer. Frequency selection in frequency domain diffuse optical spectroscopy. In *Optical Tomography and Spectroscopy of Tissue XIV*, volume 11639, page 116390N. International Society for Optics and Photonics, 2021.

[12] MB Applegate, RE Istfan, S Spink, A Tank, and D Roblyer. Recent advances in high speed diffuse optical imaging in biomedicine. *APL Photonics*, 5(4):040802, 2020.

[13] Melina Arnold, Eileen Morgan, Harriet Rumgay, Allini Mafra, Deependra Singh, Mathieu Laversanne, Jerome Vignat, Julie R Gralow, Fatima Cardoso, Sabine Siesling, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66:15–23, 2022.

[14] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.

[15] Simon R Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41, 1999.

[16] Simon R Arridge and John C Schotland. Optical tomography: forward and inverse problems. *Inverse Problems*, 25(12):123010, 2009.

[17] Simon R Arridge and M Schweiger. Photon-measurement density functions. part 2: Finite-element-method calculations. *Applied Optics*, 34(34):8026–8037, 1995.

[18] Simon R Arridge and Martin Schweiger. A gradient-based optimisation scheme for optical tomography. *Optics Express*, 2(6):213–226, 1998.

[19] Sanam Assili. A review of tomographic rconstruction techniques for computed tomography. *arXiv preprint arXiv:1808.09172*, 2018.

[20] Fred S Azar, Kijoon Lee, Ali Khamene, Regine Choe, Alper Corlu, Soren D Konecky, et al. Standardized platform for coregistration of nonconcurrent diffuse optical and magnetic resonance breast images obtained in different geometries. *Journal of Biomedical Optics*, 12(5):051902–051902, 2007.

[21] Reheman Baikejiang, Wei Zhang, and Changqing Li. Diffuse optical tomography for breast cancer imaging guided by computed tomography: A feasibility study. *Journal of X-ray Science and Technology*, 25(3):341–355, 2017.

[22] Ganesh M Balasubramaniam, Ben Wiesel, Netanel Biton, Rajnish Kumar, Judy Kupferman, and Shlomi Arnon. Tutorial on the use of deep learning in diffuse optical tomography. *Electronics*, 11(3):305, 2022.

[23] Debapriya Banik and Debotosh Bhattacharjee. Mitigating data imbalance issues in medical image analysis. In *Data preprocessing, active learning, and cost perceptive approaches for resolving data imbalance*, pages 66–89. IGI Global, 2021.

[24] Nitin Bansal, Xiaohan Chen, and Zhangyang Wang. Can we gain more from orthogonality regularizations in training deep networks? *Advances in Neural Information Processing Systems*, 31, 2018.

[25] Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv preprint arXiv:2105.04906*, 2021.

[26] Nourhan Bayasi, Ghassan Hamarneh, and Rafeef Garbi. Boosternet: Improving domain generalization of deep neural nets using culpability-ranked features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 538–548, 2022.

[27] Bradley Beattie. *Improvements in the robustness and accuracy of bioluminescence tomographic reconstructions of distributed sources within small animals.* PhD thesis, Columbia University, 2018.

[28] Francisco Beca and Kornelia Polyak. Intratumor heterogeneity in breast cancer. In *Novel Biomarkers in the Continuum of Breast Cancer*, pages 169–189. Springer, 2016.

[29] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[30] Hanene Ben Yedder, Aïcha BenTaieb, Majid Shokoufi, Amir Zahiremami, Farid Golnaraghi, and Ghassan Hamarneh. Deep learning based image reconstruction for diffuse optical tomography. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 112–119. Springer, 2018.

[31] Hanene Ben Yedder, Ben Cardoen, and Ghassan Hamarneh. Deep learning for biomedical image reconstruction: A survey. *Artificial Intelligence Review*, pages 1–33, 2020.

[32] Hanene Ben Yedder, Ben Cardoen, Majid Shokoufi, Farid Golnaraghi, and Ghassan Hamarneh. Deep orthogonal multi-frequency fusion for tomogram-free diagnosis in diffuse optical imaging. 2023.

[33] Hanene Ben Yedder, Majid Shokoufi, Ben Cardoen, Farid Golnaraghi, and Ghassan Hamarneh. Limited-angle diffuse optical tomography image reconstruction using deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 66–74. Springer, 2019.

[34] Landman Bennett and Warfield Simon. 2013 cardiac atlas project standard challenge-miccai 2013 grand challenge, 2013. Available at http://masiweb.vuse.vanderbilt.edu/workshop2013/index.php/.

[35] Sayantan Bhadra, Varun A. Kelkar, Frank J. Brooks, and Mark A. Anastasio. On hallucinations in tomographic image reconstruction. *IEEE Transactions on Medical Imaging*, 40(11):3249–3260, 2021.

[36] Sayantan Bhadra, Weimin Zhou, and Mark A Anastasio. Medical image reconstruction with image-adaptive priors learned by use of generative adversarial networks. In *Medical Imaging 2020: Physics of Medical Imaging*, volume 11312, page 113120V. International Society for Optics and Photonics, 2020.

[37] Bo Bi, Bo Han, Weimin Han, Jinping Tang, and Li Li. Image reconstruction for diffuse optical tomography based on radiative transfer equation. *Computational and mathematical methods in medicine*, 2015:286161–23, 2015.

[38] UK Biobank. About uk biobank. *Available at https://www.ukbiobank.ac.uk/about-biobank-uk*, 2014.

[39] David Boublil, Michael Elad, Joseph Shtok, and Michael Zibulevsky. Spatially-adaptive reconstruction in computed tomography using neural networks. *IEEE Transactions on Medical Imaging*, 34(7):1474–1485, 2015.

[40] Nicholas Boyd, Eric Jonas, Hazen P Babcock, and Benjamin Recht. Deeploco: Fast 3D localization microscopy using neural networks. *BioRxiv*, page 267096, 2018.

[41] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[42] Aljaz Bozic, Pablo Palafox, Justus Thies, Angela Dai, and Matthias Nießner. Transformerfusion: Monocular rgb scene reconstruction using transformers. *Advances in Neural Information Processing Systems*, 34, 2021.

[43] Nathaniel Braman, Jacob WH Gordon, Emery T Goossens, Caleb Willis, Martin C Stumpe, and Jagadish Venkataraman. Deep orthogonal fusion: Multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 667–677. Springer, 2021.

[44] Henry Braun, Pavan Turaga, Andreas Spanias, Sameeksha Katoch, Suren Jayasuriya, and Cihan Tepedelenlioglu. Reconstruction-free compressive vision for surveillance applications. *Synthesis Lectures on Signal Processing*, 14(1):1–100, 2019.

[45] Chuangjian Cai, Kexin Deng, Cheng Ma, and Jianwen Luo. End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Optics Letters*, 43(12):2752–2755, 2018.

[46] Yun-Hao Cao, Hao Yu, and Jianxin Wu. Training vision transformers with only 2040 images. In *European Conference on Computer Vision*, pages 220–237. Springer, 2022.

[47] Ben Cardoen, Hanene Ben Yedder, Sieun Lee, Ivan Robert Nabi, and Ghassan Hamarneh. Datacurator. jl: efficient, portable and reproducible validation, curation and transformation of large heterogeneous datasets using human-readable recipes compiled into machine-verifiable templates. *Bioinformatics Advances*, 3(1):vbad068, 2023.

[48] Ben Cardoen, Hanene Ben Yedder, Sieun Lee, Ivan Robert Nabi, and Ghassan Hamarneh. Log-paradox: Necessary and sufficient conditions for confounding statistically significant pattern reversal under the log-transform. *arXiv preprint arXiv:2302.04780*, 2023.

[49] Ben Cardoen, Hanene Ben Yedder, Anmol Sharma, Keng C Chou, Ivan Robert Nabi, and Ghassan Hamarneh. ERGO: Efficient recurrent graph optimized emitter density estimation in single molecule localization microscopy. *IEEE Transactions on Medical Imaging*, 39(6):1942–1956, 2020.

[50] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *JJournal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[51] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[52] Chen Chen, Venkaiah C Kavuri, Xinlong Wang, Ruoyu Li, Hanli Liu, and Junzhou Huang. Multi-frequency diffuse optical tomography for cancer detection. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 67–70. IEEE, 2015.

[53] Chen Chen, Fenghua Tian, Hanli Liu, and Junzhou Huang. Diffuse optical tomography enhanced by clustered sparsity for functional brain imaging. *IEEE Transactions on Medical Imaging*, 33(12):2323–2331, 2014.

[54] Haomin Chen, Catalina Gomez, Chien-Ming Huang, and Mathias Unberath. Explainable medical imaging ai needs human-centered design: guidelines and evidence from a systematic review. *NPJ Digital Medicine*, 5(1):156, 2022.

[55] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017.

[56] Hu Chen, Yi Zhang, Weihua Zhang, Peixi Liao, Ke Li, Jiliu Zhou, and Ge Wang. Low-dose CT via convolutional neural network. *Biomedical Optics Express*, 8(2):679–694, 2017.

[57] Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 2020.

[58] Yang Chen, Zhou Yang, Yining Hu, Guanyu Yang, Yongcheng Zhu, Yinsheng Li, Wufan Chen, Christine Toumoulin, et al. Thoracic low-dose CT image processing using an artifact suppressed large-scale nonlocal means. *Physics in Medicine & Biology*, 57(9):2667, 2012.

[59] Joseph Y Cheng, Feiyu Chen, Marcus T Alley, John M Pauly, and Shreyas S Vasanawala. Highly scalable image reconstruction using deep neural networks with bandpass filtering. *arXiv preprint arXiv:1805.03300*, 2018.

[60] François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[61] Hyungjin Chung, Jeongsol Kim, and Jong Chul Ye. Direct diffusion bridge using data consistency for inverse problems. *arXiv preprint arXiv:2305.19809*, 2023.

[62] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 529–536. Springer, 2018.

[63] Cathy Coleman. Early detection and screening for breast cancer. In *Seminars in oncology nursing*, volume 33, pages 141–155. Elsevier, 2017.

[64] William R Crum, Oscar Camara, and Derek LG Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, 2006.

[65] Jianan Cui, Kuang Gong, Ning Guo, Kyungsang Kim, Huafeng Liu, and Quanzheng Li. CT-guided PET parametric image reconstruction using deep neural network without prior training data. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, page 109480Z. International Society for Optics and Photonics, 2019.

[66] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[67] Alberto Dalla Mora, Davide Contini, Simon Arridge, Fabrizio Martelli, Alberto Tosi, Gianluca Boso, et al. Towards next-generation time-domain diffuse optics for extreme depth penetration and sensitivity. *Biomedical Optics Express*, 6(5):1749–1760, 2015.

[68] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

[69] James Davis, Yi-Hsuan Hsieh, and Hung-Chi Lee. Humans perceive flicker artifacts at 500 hz. *Scientific reports*, 5(1):1–4, 2015.

[70] Hamid Dehghani, Matthew E Eames, Phaneendra K Yalavarthy, Scott C Davis, Subhadra Srinivasan, Colin M Carpenter, Brian W Pogue, and Keith D Paulsen. Near infrared optical tomography using nirfast: Algorithm for numerical model and image reconstruction. *Communications in Numerical Methods in Engineering*, 25(6):711–732, 2009.

[71] Aleksandr Dekhovich, David MJ Tax, Marcel HF Sluiter, and Miguel A Bessa. Continual prune-and-select: class-incremental learning with specialized subnetworks. *Applied Intelligence*, pages 1–16, 2023.

[72] Philippe Despres and Xun Jia. A review of GPU-based medical image reconstruction. *Physica Medica*, 42:76–92, 2017.

[73] Giuseppe Di Sciacca, Giulia Maffeis, Andrea Farina, Alberto Dalla Mora, Antonio Pifferi, Paola Taroni, et al. Evaluation of a pipeline for simulation, reconstruction, and classification in ultrasound-aided diffuse optical tomography of breast tumors. *Journal of Biomedical Optics*, 27(3):036003, 2022.

[74] Bin Dong and Zuowei Shen. Image restoration: a data-driven perspective. In *Proceedings of the International Congress of Industrial and Applied Mathematics (ICIAM)*, pages 65–108. Citeseer, 2015.

[75] A Douiri, M Schweiger, J Riley, and SR Arridge. Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information. *Measurement Science and Technology*, 18(1):87, 2006.

[76] Matthaios Doulgerakis, Adam T Eggebrecht, and Hamid Dehghani. High-density functional diffuse optical tomography based on frequency-domain measurements improves image quality and spatial resolution. *Neurophotonics*, 6(3):035007, 2019.

[77] Siyi Du, Nourhan Bayasi, Ghassan Harmarneh, and Rafeef Garbi. Mdvit: Multi-domain vision transformer for small medical image segmentation datasets. *arXiv preprint arXiv:2307.02100*, 2023.

[78] Turgut Durduran, Regine Choe, Wesley B Baker, and Arjun G Yodh. Diffuse optics for tissue monitoring and tomography. *Reports on Progress in Physics*, 73(7):076701, 2010.

[79] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *Journal of Machine Learning Research*, 20:1–21, 2019.

[80] Zhuang et al. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.

[81] Qiuyun Fan, Thomas Witzel, Aapo Nummenmaa, Koene RA Van Dijk, John D Van Horn, Michelle K Drews, Leah H Somerville, Margaret A Sheridan, Rosario M Santillana, Jenna Snyder, et al. MGH–USC human connectome project datasets with ultra-high b-value diffusion MRI. *Neuroimage*, 124:1108–1114, 2016.

[82] Xiang Fang, Chenyang Gao, Yingxin Li, and Ting Li. Solving heterogenous region for diffuse optical tomography with a convolutional forward calculation model and the inverse neural network. In *Advanced Optical Imaging Technologies III*, volume 11549, pages 50–60. SPIE, 2020.

[83] Jinchao Feng, Qiuwan Sun, Zhe Li, Zhonghua Sun, and Kebin Jia. Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography. *Journal of Biomedical Optics*, 24(5):051407, 2018.

[84] Jinchao Feng, Qiuwan Sun, Zhe Li, Zhonghua Sun, and Kebin Jia. Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography. *Journal of Biomedical Optics*, 24(5):051407, 2018.

[85] Jeffrey A Fessler. Medical image reconstruction: a brief overview of past milestones and future directions. *arXiv preprint arXiv:1707.05927*, 2017.

[86] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

[87] ML Flexman, HK Kim, R Stoll, MA Khalil, CJ Fong, and AH Hielscher. A wireless handheld probe with spectrally constrained evolution strategies for diffuse optical imaging of tissue. *Review of Scientific Instruments*, 83(3):033108, 2012.

[88] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.

[89] Hao Gao, Hengyong Yu, Stanley Osher, and Ge Wang. Multi-energy CT based on a prior rank, intensity and sparsity model (prism). *Inverse Problems*, 27(11):115012, 2011.

[90] Alexander J Gates and Yong-Yeol Ahn. The impact of random models on clustering similarity. *The Journal of Machine Learning Research*, 18(1):3049–3076, 2017.

[91] Jean-Michel Geffrin, Pierre Sabouroux, and Christelle Eyraud. Free space experimental scattering database continuation: experimental set-up and measurement precision. *Inverse Problems*, 21(6):S117, 2005.

[92] Gary Georgeson and Morteza Safai. Portable X-ray backscattering imaging system including a radioactive source, May 23 2017. US Patent 9,658,173.

[93] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331, 2009.

[94] Robin C Geyer, Tassilo Klein, and Moin Nabi. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*, 2017.

[95] Muhammad Ghifary, David Balduzzi, W Bastiaan Kleijn, and Mengjie Zhang. Scatter component analysis: A unified framework for domain adaptation and domain generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7):1414–1430, 2016.

[96] N Ghosh, SK Mohanty, SK Majumder, and PK Gupta. Measurement of optical transport properties of normal and malignant human breast tissue. *Applied Optics*, 40(1):176–184, 2001.

[97] Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin Lauter, Michael Naehrig, and John Wernsing. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *International Conference on Machine Learning*, pages 201–210, 2016.

[98] Evgin Goceri and Numan Goceri. Deep learning in medical image analysis: recent advances and future trends. In *International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing*. IADIS, 2017.

[99] Evgin Goceri and Ali Gooya. On the importance of batch size for deep learning. In *international conference on mathematics. Istanbul, Turkey*, 2018.

[100] Anuradha Godavarty, Suset Rodriguez, Young-Jin Jung, and Stephanie Gonzalez. Optical imaging for breast cancer prescreening. *Breast Cancer: Targets and Therapy*, 7:193, 2015.

[101] Farid Golnaraghi and Majid Shokoufi. Handheld probe and system for imaging human tissue, October 31 2019. US Patent App. 16/479,025.

[102] Kuang Gong, Ciprian Catana, Jinyi Qi, and Quanzheng Li. Pet image reconstruction using deep image prior. *IEEE Transactions on Medical Imaging*, 38(7):1655–1665, 2018.

[103] Kuang Gong, Ciprian Catana, Jinyi Qi, and Quanzheng Li. Direct patlak reconstruction from dynamic pet using unsupervised deep learning. In *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear*

Medicine, volume 11072, page 110720R. International Society for Optics and Photonics, 2019.

[104] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[105] Nina M Gottschling, Vegard Antun, Ben Adcock, and Anders C Hansen. The troublesome kernel: why deep learning for inverse problems is typically unstable. *arXiv preprint arXiv:2001.01258*, 2020.

[106] Dirk Grosenick, Herbert Rinneberg, Rinaldo Cubeddu, and Paola Taroni. Review of optical breast imaging and spectroscopy. *Journal of Biomedical Optics*, 21(9):091311, 2016.

[107] Han Guo, Chenlu Qiu, and Namrata Vaswani. An online algorithm for separating sparse and low-dimensional signal sequences from their sum. *IEEE Transactions on Signal Processing*, 62(16):4284–4297, 2014.

[108] Jingyu Guo, Hongliang Qi, Yuan Xu, Zijia Chen, Shulong Li, and Linghong Zhou. Iterative image reconstruction for limited-angle ct using optimized initial image. *Computational and mathematical methods in medicine*, 2016:1–9, 2016.

[109] Wenzhong Guo, Jianwen Wang, and Shiping Wang. Deep multimodal representation learning: A survey. *IEEE Access*, 7:63373–63394, 2019.

[110] Harshit Gupta, Kyong Hwan Jin, Ha Q Nguyen, Michael T McCann, and Michael Unser. CNN-based projected gradient descent for consistent ct image reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1440–1453, 2018.

[111] Ida Häggström, Bradley J Beattie, and C Ross Schmidtlein. Dynamic PET simulator via tomographic emission projection for kinetic modeling and parametric image studies. *Medical Physics*, 43(6Part1):3104–3116, 2016.

[112] Ida Häggström, C Ross Schmidtlein, Gabriele Campanella, and Thomas J Fuchs. Deeppet: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Medical Image Analysis*, 54:253–262, 2019.

[113] Melanie Hamilton-Basich. Hyperfine receives fda clearance for portable mri technology. *AXIS Imaging News*, 2020.

[114] Robert L Harrison. Monte carlo simulation of emission tomography and other medical imaging techniques. In *AIP conference proceedings*, volume 1204, pages 126–132. AIP, 2010.

[115] Richard C Haskell, Lars O Svaasand, Tsong-Tseh Tsay, Ti-Chen Feng, Matthew S McAdams, and Bruce J Tromberg. Boundary conditions for the diffusion equation in radiative transfer. *JOSA A*, 11(10):2727–2741, 1994.

[116] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[117] Juha Heiskala, Petri Hiltunen, and Ilkka Nissilä. Significance of background optical properties, time-resolved information and optode arrangement in diffuse optical imaging of term neonates. *Physics in Medicine & Biology*, 54(3):535, 2009.

[118] Michel Herranz, Alvaro Ruibal, et al. Optical imaging in breast cancer diagnosis: the next evolution. *Journal of Oncology*, 2012, 2012.

[119] Seamus Holden and Daniel Sage. Imaging: super-resolution fight club. *Nature Photonics*, 10(3):152, 2016.

[120] Yoko Hoshi and Yukio Yamada. Overview of diffuse optical tomography and its clinical applications. *Journal of Biomedical Optics*, 21(9):091312, 2016.

[121] Seyed Amir Hossein Hosseini, Burhaneddin Yaman, Steen Moeller, Mingyi Hong, and Mehmet Akçakaya. Dense recurrent neural networks for inverse problems: History-cognizant unrolling of optimization algorithms. *arXiv preprint arXiv:1912.07197*, 2019.

[122] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[123] Lei Huang, Xianglong Liu, Bo Lang, Adams Wei Yu, Yongliang Wang, and Bo Li. Orthogonal weight normalization: Solution to optimization over multiple dependent stiefel manifolds in deep neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[124] Qiaoying Huang, Dong Yang, Jingru Yi, Leon Axel, and Dimitris Metaxas. FR-Net: Joint reconstruction and segmentation in compressed sensing cardiac MRI. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 352–360. Springer, 2019.

[125] Yixing Huang, Alexander Preuhs, Günter Lauritsch, Michael Manhart, Xiaolin Huang, and Andreas Maier. Data consistent artifact reduction for limited angle tomography with deep learning prior. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 101–112. Springer, 2019.

[126] Zehao Huang and Naiyan Wang. Data-driven sparse structure selection for deep neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 304–320, 2018.

[127] Mohammad Arafat Hussain, Ghassan Hamarneh, Timothy W O'Connell, Mohammed F Mohammed, and Rafeef Abugharbieh. Segmentation-free estimation of kidney volumes in ct with dual regression forests. In *International Workshop on Machine Learning in Medical Imaging*, pages 156–163. Springer, 2016.

[128] P Huston, VL Edge, and E Bernier. Open science/open data: Reaping the benefits of open data in public health. *Canada Communicable Disease Report*, 45(11):252, 2019.

[129] Chang Min Hyun, Hwa Pyung Kim, Sung Min Lee, Sungchul Lee, and Jin Keun Seo. Deep learning for undersampled MRI reconstruction. *Physics in Medicine & Biology*, 63(13):135007, 2018.

[130] Xavier Intes and Britton Chance. Multi-frequency diffuse optical tomography. *Journal of Modern Optics*, 52(15):2139–2159, 2005.

[131] Maia Jacobs, Melanie F Pradier, Thomas H McCoy, Roy H Perlis, Finale Doshi-Velez, and Krzysztof Z Gajos. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational psychiatry*, 11(1):1–9, 2021.

[132] Tanuj Kumar Jhamb, Vinith Rejathalal, and VK Govindan. A review on image reconstruction through MRI k-space data. *International Journal of Image, Graphics and Signal Processing*, 7(7):42, 2015.

[133] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[134] Kyong Hwan Jin, Michael Unser, and Kwang Moo Yi. Self-supervised deep active accelerated MRI. *arXiv preprint arXiv:1901.04547*, 2019.

[135] Weina Jin, Xiaoxiao Li, Mostafa Fatehi, and Ghassan Hamarneh. Guidelines and evaluation of clinical explainable ai in medical image analysis. *Medical Image Analysis*, 84:102684, 2023.

[136] Weina Jin, Xiaoxiao Li, and Ghassan Hamarneh. Evaluating explainable ai on a multi-modal medical imaging task: Can existing algorithms fulfill clinical requirements? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11945–11953, 2022.

[137] Ulugbek S Kamilov et al. Learning approach to optical tomography. *Optica*, 2(6):517–522, 2015.

[138] Eunhee Kang, Hyun Jung Koo, Dong Hyun Yang, Joon Bum Seo, and Jong Chul Ye. Cycle-consistent adversarial denoising network for multiphase coronary ct angiography. *Medical physics*, 46(2):550–562, 2019.

[139] Jeremy Kawahara, Sara Daneshvar, Giuseppe Argenziano, and Ghassan Hamarneh. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE journal of Biomedical and Health Informatics*, 23(2):538–546, 2018.

[140] Jeremy Kawahara and Ghassan Hamarneh. Fully convolutional neural networks to detect clinical dermoscopic features. *IEEE journal of Biomedical and Health Informatics*, 23(2):578–585, 2018.

[141] Huseyin Ozgur Kazanci and O Oral. Frequency shifting model for diffuse optical tomography. *Optical and Quantum Electronics*, 53(11):1–6, 2021.

[142] Vasant Kearney, Benjamin P Ziemer, Alan Perry, Tianqi Wang, Jason W Chan, Lijun Ma, Olivier Morin, Sue S Yom, and Timothy D Solberg. Attention-aware discrimination for mr-to-ct image translation using cycle-consistent generative adversarial networks. *Radiology: Artificial Intelligence*, 2(2):e190027, 2020.

[143] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev https://doi.org/10.1007/s10462-020-09825-6*, 2020.

[144] Hyun Keol Kim, Uwe J Netz, Jürgen Beuthan, and Andreas H Hielscher. Optimal source-modulation frequencies for transport-theory-based optical tomography of small-tissue volumes. *Optics express*, 16(22):18082–18101, 2008.

[145] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[146] Florian Knoll, Kerstin Hammernik, Chi Zhang, Steen Moeller, Thomas Pock, Daniel K Sodickson, and Mehmet Akcakaya. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE Signal Processing Magazine*, 37(1):128–140, 2020.

[147] Alexander B Konovalov, Elina A Genina, and Alexey N Bashkatov. Diffuse optical mammotomography: state-of-the-art and prospects. *Journal of Biomedical Photonics & Engineering*, 2(2):020202–1, 2016.

[148] W.W.M Lam et al. Factors affecting the palpability of breast lesion by self-examination. *Singapore medical journal*, 49(3):228, 2008.

[149] Hengrong Lan, Kang Zhou, Changchun Yang, Jun Cheng, Jiang Liu, Shenghua Gao, and Fei Gao. Ki-gan: Knowledge infusion generative adversarial network for photoacoustic image reconstruction in vivo. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 273–281. Springer, 2019.

[150] Pedro Latorre-Carmona, V Javier Traver, J Salvador Sánchez, and Enrique Tajahuerce. Online reconstruction-free single-pixel image classification. *Image and Vision Computing*, 86:28–37, 2019.

[151] Dan Li, Changde Du, and Huiguang He. Semi-supervised cross-modal image generation with generative adversarial networks. *Pattern Recognition*, 100:107085, 2020.

[152] Danyang Li, Sui Li, Manman Zhu, Qi Gao, Zhaoying Bian, Haiyun Huang, Shanli Zhang, Jing Huang, Dong Zeng, and Jianhua Ma. Unsupervised data fidelity enhancement network for spectral ct reconstruction. In *Medical Imaging 2020: Physics of Medical Imaging*, volume 11312, page 113124D. International Society for Optics and Photonics, 2020.

[153] Linlin Li, Siyuan Shen, Shengyu Gao, Yuehan Wang, Liangtao Gu, Shiying Li, Xingjun Zhu, Jiahua Jiang, Jingyi Yu, and Wuwei Ren. A self-supervised learning approach for high-resolution diffuse optical tomography using neural fields. In *Second Conference on Biomedical Photonics and Cross-Fusion (BPC 2023)*, volume 12753, pages 76–82. SPIE, 2023.

[154] Shuying Li, Menghao Zhang, Minghao Xue, and Quing Zhu. Difference imaging from single measurements in diffuse optical tomography: a deep learning approach. *Journal of Biomedical Optics*, 27(8):086003, 2022.

[155] Zexi Li, Feng Mao, and Chao Wu. Can we share models if sharing data is not an option? *Patterns*, 3(11), 2022.

[156] Dong Liang, Jing Cheng, Ziwen Ke, and Leslie Ying. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal Processing Magazine*, 37(1):141–151, 2020.

[157] Kaichao Liang, Hongkai Yang, Kejun Kang, and Yuxiang Xing. Improve angular resolution for sparse-view CT with residual convolutional neural network. In *Medical Imaging 2018: Physics of Medical Imaging*, volume 10573, page 105731K. International Society for Optics and Photonics, 2018.

[158] Zhi-Pei Liang. Spatiotemporal imagingwith partially separable functions. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 988–991. IEEE, 2007.

[159] Zhouchen Lin. A review on low-rank models in data analysis. *Big Data & Information Analytics*, 1(2&3):139–161, 2016.

[160] Dongyuan Liu, Yao Zhang, Lu Bai, Pengrui Zhang, and Feng Gao. Combining two-layer semi-three-dimensional reconstruction and multi-wavelength image fusion for functional diffuse optical tomography. *IEEE Transactions on Computational Imaging*, 7:1055–1068, 2021.

[161] Wenqi Lu, Jinming Duan, David Orive-Miguel, Lionel Herve, and Iain B Styles. Graph-and finite element-based total variation models for the inverse problem in diffuse optical tomography. *Biomedical Optics Express*, 10(6):2684–2707, 2019.

[162] Alice Lucas, Michael Iliadis, Rafael Molina, and Aggelos K Katsaggelos. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Processing Magazine*, 35(1):20–36, 2018.

[163] Lung Cancer Alliance. Give a scan, Fact Sheet N°282, 2017. `http://www.giveascan.org`, Last accessed on 2020-2-20.

[164] GuanXiong Luo, Na Zhao, Wenhao Jiang, and Peng Cao. Mri reconstruction using deep bayesian inference. *arXiv preprint arXiv:1909.01127*, 2019.

[165] Bo Lyu, Hang Yuan, Longfei Lu, and Yunye Zhang. Resource-constrained neural architecture search on edge devices. *IEEE Transactions on Network Science and Engineering*, 9(1):134–142, 2021.

[166] Shimaa Mahdy, Omnia Hamdy, Mohammed A Hassan, and Mohamed AA Eldosoky. A modified source-detector configuration for the discrimination between normal and diseased human breast based on the continuous-wave diffuse optical imaging approach: a simulation study. *Lasers in Medical Science*, pages 1–10, 2021.

[167] Mayur Mallya and Ghassan Hamarneh. Deep multimodal guidance for medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 298–308. Springer, 2022.

[168] Mark Mazumder, Colby Banbury, Xiaozhe Yao, Bojan Karlaš, William Gaviria Rojas, Sudnya Diamos, Greg Diamos, Lynn He, Alicia Parrish, Hannah Rose Kirk, et al. Dataperf: Benchmarks for data-centric ai development. *arXiv preprint arXiv:2207.10062*, 2022.

[169] Michael T McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017.

[170] Michael T McCann and Michael Unser. Algorithms for biomedical image reconstruction. *arXiv preprint arXiv:1901.03565*, 2019.

[171] Michael T McCann, Michael Unser, et al. Biomedical image reconstruction: From the foundations to deep neural networks. *Foundations and Trends® in Signal Processing*, 13(3):283–359, 2019.

[172] Adriënne M Mendrik, Koen L Vincken, Hugo J Kuijf, Marcel Breeuwer, Willem H Bouvy, Jeroen De Bresser, Amir Alansary, Marleen De Bruijne, Aaron Carass, Ayman El-Baz, et al. MRbrains challenge: online evaluation framework for brain image segmentation in 3T MRI scans. *Computational Intelligence and Neuroscience*, 2015:1, 2015.

[173] Mingqiang Meng, Sui Li, Lisha Yao, Danyang Li, Manman Zhu, Qi Gao, Qi Xie, Qian Zhao, Zhaoying Bian, Jing Huang, et al. Semi-supervised learned sinogram restoration network for low-dose ct image reconstruction. In *Medical Imaging 2020: Physics of Medical Imaging*, volume 11312, page 113120B. International Society for Optics and Photonics, 2020.

[174] Mackenzie Moir and Bacchus Barua. Waiting your turn: Wait times for health care in canada, 2022 report. 2022.

[175] Alejandro Morales-Hernández, Inneke Van Nieuwenhuyse, and Sebastian Rojas Gonzalez. A survey on multi-objective hyperparameter optimization algorithms for machine learning. *Artificial Intelligence Review*, pages 1–51, 2022.

[176] Vahid Moslemi, Vahid Erfanian, and Mansour Ashoor. Estimation of optimized timely system matrix with improved image quality in iterative reconstruction algorithm: A simulation study. *Heliyon*, 6(1):e03279, 2020.

[177] Meghdoot Mozumder, Andreas Hauptmann, Simon R Arridge, and Tanja Tarvainen. Diffuse optical tomography utilizing model-based learning. In *Optics and the Brain*, pages JTu3A–10. Optica Publishing Group, 2022.

[178] Meghdoot Mozumder, Andreas Hauptmann, Ilkka Nissilä, Simon R Arridge, and Tanja Tarvainen. A model-based iterative learning approach for diffuse optical tomography. *IEEE Transactions on Medical Imaging*, 41(5):1289–1299, 2021.

[179] Meghdoot Mozumder, Tanja Tarvainen, Aku Seppänen, Ilkka Nissilä, Simon R Arridge, and Ville Kolehmainen. Nonlinear approach to difference imaging in diffuse optical tomography. *Journal of Biomedical Optics*, 20(10):105001, 2015.

[180] Mridata. mridata.org. `http://mridata.org`, Last accessed on 2019-11-30.

[181] Vicky Mudeng, Wahdiyatun Nisa, and Sena Sukmananda Suprapto. Computational image reconstruction for multi-frequency diffuse optical tomography. *Journal of King Saud University-Computer and Information Sciences*, 2021.

[182] Arsha Nagrani, Shan Yang, Anurag Arnab, Aren Jansen, Cordelia Schmid, and Chen Sun. Attention bottlenecks for multimodal fusion. *Advances in Neural Information Processing Systems*, 34, 2021.

[183] Elias Nehme, Lucien E Weiss, Tomer Michaeli, and Yoav Shechtman. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4):458–464, 2018.

[184] Yurii E Nesterov. A method for solving the convex programming problem with convergence rate o (1/kˆ 2). In *Dokl. akad. nauk Sssr*, volume 269, pages 543–547, 1983.

[185] Thanh Nguyen, Vy Bui, and George Nehmetallah. Computational optical tomography using 3-d deep convolutional neural networks. *Optical Engineering*, 57(4):043111–043111, 2018.

[186] Navid Ibtehaj Nizam, Marien Ochoa, Jason T Smith, and Xavier Intes. Wide-field diffuse optical tomography using deep learning. In *Optical Tomography and Spectroscopy*, pages OW4D–7. Optica Publishing Group, 2022.

[187] Navid Ibtehaj Nizam, Marien Ochoa, Jason T Smith, and Xavier Intes. Deep learning-based fusion of widefield diffuse optical tomography and micro-ct structural priors for accurate 3d reconstructions. *Biomedical Optics Express*, 14(3):1041–1053, 2023.

[188] Natalia Norori, Qiyang Hu, Florence Marcelle Aellen, Francesca Dalia Faraci, and Athina Tzovara. Addressing bias in big data and ai for health care: A call for open science. *Patterns*, 2(10), 2021.

[189] Ilkay Oksuz, James Clough, Bram Ruijsink, Esther Puyol-Antón, Aurelien Bustin, Gastao Cruz, Claudia Prieto, Daniel Rueckert, Andrew P King, and Julia A Schnabel. Detection and correction of cardiac MRI motion artefacts during reconstruction from k-space. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 695–703. Springer, 2019.

[190] Jonas Oppenlaender, Aku Visuri, Ville Paananen, Rhema Linder, and Johanna Silvennoinen. Text-to-image generation: Perceptions and realities. *arXiv preprint arXiv:2303.13530*, 2023.

[191] Ricardo Otazo, Emmanuel Candès, and Daniel K Sodickson. Low-rank plus sparse matrix decomposition for accelerated dynamic mri with separation of background and dynamic components. *Magnetic Resonance in Medicine*, 73(3):1125–1136, 2015.

[192] Jiahong Ouyang, Kevin T Chen, Enhao Gong, John Pauly, and Greg Zaharchuk. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. *Medical Physics*, 46(8):3555–3564, 2019.

[193] Gergely Pap, Gábor Lékó, and Tamás Grósz. A reconstruction-free projection selection procedure for binary tomography using convolutional neural networks. In *International Conference on Image Analysis and Recognition*, pages 228–236. Springer, 2019.

[194] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2337–2346, 2019.

[195] Gaurav Patel and Jose Dolz. Weakly supervised segmentation with cross-modality equivariant constraints. *Medical Image Analysis*, page 102374, 2022.

[196] Grégory Paul, Janick Cardinale, and Ivo F Sbalzarini. Coupling image restoration and segmentation: a generalized linear model/bregman perspective. *International Journal of Computer Vision*, 104(1):69–93, 2013.

[197] Daniel Maria Pelt and Kees Joost Batenburg. Fast tomographic reconstruction from limited data using artificial neural networks. *IEEE Transactions on Image Processing*, 22(12):5238–5251, 2013.

[198] Guy A Perkins, Adam T Eggebrecht, and Hamid Dehghani. Multi-modulated frequency domain high density diffuse optical tomography. *Biomedical Optics Express*, 13(10):5275–5294, 2022.

[199] Matthias Perkonigg, Johannes Hofmanninger, and Georg Langs. Continual active learning for efficient adaptation of machine learning models to changing image acquisition. In *International Conference on Information Processing in Medical Imaging*, pages 649–660. Springer, 2021.

[200] Antonio Pifferi, Alberto Dalla Mora, Laura Di Sieno, Edoardo Ferocino, Alberto Tosi, Enrico Conca, et al. Solus: an innovative multimodal imaging system to improve breast cancer diagnosis through diffuse optics and ultrasounds. In *Optical Tomography and Spectroscopy of Tissue XIV*, volume 11639, page 116390C. International Society for Optics and Photonics, 2021.

[201] Chen Qin, Jo Schlemper, Jose Caballero, Anthony N Price, Joseph V Hajnal, and Daniel Rueckert. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 38(1):280–290, 2018.

[202] Ziyuan Qin, Huahui Yi, Qicheng Lao, and Kang Li. Medical image understanding with pretrained vision language models: A comprehensive study. *arXiv preprint arXiv:2209.15517*, 2022.

[203] Abhejit Rajagopal, Noah Stier, Joyoni Dey, Michael A King, and Shivkumar Chandrasekaran. Towards deep iterative-reconstruction algorithms for computed tomography (CT) applications. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, page 1094856. International Society for Optics and Photonics, 2019.

[204] Kanchana Ranasinghe, Muzammal Naseer, Munawar Hayat, Salman Khan, and Fahad Shahbaz Khan. Orthogonal projection loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12333–12343, 2021.

[205] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *arXiv preprint arXiv:1904.02816*, 2019.

[206] Annika Reinke, Minu D Tizabi, Carole H Sudre, Matthias Eisenmann, Tim Rädsch, Michael Baumgartner, Laura Acion, Michela Antonelli, Tal Arbel, Spyridon Bakas, et al. Common limitations of image processing metrics: A picture story. *arXiv preprint arXiv:2104.05642*, 2021.

[207] Wuwei Ren, Siyuan Shen, Linlin Li, Shengyu Gao, Yuehan Wang, Liangtao Gu, Shiying Li, Xingjun Zhu, Jiahua Jiang, and Jingyi Yu. High-resolution tomographic reconstruction of optical absorbance through scattering media using neural fields. *arXiv preprint arXiv:2304.01682*, 2023.

[208] Ronald L Rivest, Len Adleman, Michael L Dertouzos, et al. On data banks and privacy homomorphisms. *Foundations of Secure Computation*, 4(11):169–180, 1978.

[209] Paul Rodríguez. Total variation regularization algorithms for images corrupted with different noise models: a review. *Journal of Electrical and Computer Engineering*, 2013:10, 2013.

[210] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[211] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 421–429. Springer, 2018.

[212] Alexander Rykkje, Jonathan Frederik Carlsen, and Michael Bachmann Nielsen. Handheld ultrasound devices compared with high-end ultrasound systems: A systematic review. *Diagnostics*, 9(2):61, 2019.

[213] Sohail Sabir et al. Convolutional neural network-based approach to estimate bulk optical properties in diffuse optical tomography. *Applied Optics*, 59(5):1461–1470, 2020.

[214] Patrik Sabol, Peter Sinčák, Pitoyo Hartono, Pavel Kočan, Zuzana Benetinová, Alžbeta Blichárová, L'udmila Verbóová, Erika Štammová, Antónia Sabolová-Fabianová, and Anna Jašková. Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images. *Journal of Biomedical Informatics*, 109:103523, 2020.

[215] Christopher M Sandino, Joseph Y Cheng, Feiyu Chen, Morteza Mardani, John M Pauly, and Shreyas S Vasanawala. Compressed sensing: From research to clinical practice with deep neural networks: Shortening scan times for magnetic resonance imaging. *IEEE Signal Processing Magazine*, 37(1):117–127, 2020.

[216] Ivo F Sbalzarini. Seeing is believing: Quantifying is convincing: Computational image analysis in biology. In *Focus on Bio-Image Informatics*, pages 1–39. Springer, 2016.

[217] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for MR image reconstruction. In *International Conference on Information Processing in Medical Imaging*, pages 647–658. Springer, 2017.

[218] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2018.

[219] Armin Schneider and Hubertus Feussner. *Biomedical engineering in gastrointestinal surgery*. Academic Press, 2017.

[220] Martin Schweiger and Simon R Arridge. The toast++ software suite for forward and inverse modeling in optical tomography. *Journal of Biomedical Optics*, 19(4):040801, 2014.

[221] Martin Schweiger and Simon R Arridge. The toast++ software suite for forward and inverse modeling in optical tomography. *Journal of Biomedical Optics*, 19(4):040801, 2014.

[222] Martin Schweiger, Simon R Arridge, and Ilkka Nissilä. Gauss–newton method for image reconstruction in diffuse optical tomography. *Physics in Medicine & Biology*, 50(10):2365, 2005.

[223] Anmol Sharma and Ghassan Hamarneh. Missing mri pulse sequence synthesis using multi-modal generative adversarial network. *IEEE Transactions on Medical Imaging*, 39(4):1170–1183, 2019.

[224] Guohua Shen, Kshitij Dwivedi, Kei Majima, Tomoyasu Horikawa, and Yukiyasu Kamitani. End-to-end deep image reconstruction from human brain activity. *Frontiers in Computational Neuroscience*, 13, 2019.

[225] Mahesh K Shetty. Screening for breast cancer with mammography: current status and an overview. *Indian journal of surgical oncology*, 1:218–223, 2010.

[226] Nafhatul Zinan Shifa, Md Mahmudur Rahman Sayem, and Md Asiful Islam. Improved image reconstruction using multi frequency data for diffuse optical tomography. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, pages 264–268. IEEE, 2021.

[227] Majid Shokoufi and Farid Golnaraghi. Development of a handheld diffuse optical breast cancer assessment probe. *Journal of Innovative Optical Health Sciences*, 9(02):1650007, 2016.

[228] Majid Shokoufi and Farid Golnaraghi. Development of a handheld diffuse optical breast cancer assessment probe. *Journal of Innovative Optical Health Sciences*, 9(02):1650007, 2016.

[229] Majid Shokoufi and Farid Golnaraghi. Handheld diffuse optical breast scanner probe for cross-sectional imaging of breast tissue. *Journal of Innovative Optical Health Sciences*, 12(02):1950008, 2019.

[230] Majid Shokoufi, Zahra Haeri, Birkanwar S Kharb, and Farid Golnaraghi. Novel handheld diffuse optical spectroscopy probe for breast cancer assessment: Clinical study. *Journal of Biomedical Sciences*, pages 0–0, 2017.

[231] Majid Shokoufi, Zahra Haeri, Zhi Yih Lim, Ramani Ramaseshan, and Farid Golnaraghi. Translation of a portable diffuse optical breast scanner probe for clinical application: a preliminary study. *Biomedical Physics & Engineering Express*, 6(1):015037, 2020.

[232] Ramandeep Singh, Subba R Digumarthy, Victorine V Muse, Avinash R Kambadakone, Michael A Blake, Azadeh Tabari, Yiemeng Hoi, Naruomi Akino, Erin Angel, Rachna Madan, et al. Image quality and lesion detection on deep learning reconstruction and iterative reconstruction of submillisievert chest and abdominal CT. *American Journal of Roentgenology*, 214(3):566–573, 2020.

[233] Vimal Singh, Ahmed H Tewfik, and David B Ress. Under-sampled functional mri using low-rank plus sparse matrix decomposition. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 897–901. IEEE, 2015.

[234] Jason T Smith, Marien Ochoa, Denzel Faulkner, Grant Haskins, and Xavier Intes. Deep learning in macroscopic diffuse optical imaging. *Journal of Biomedical Optics*, 27(2):020901–020901, 2022.

[235] Ghislain St-Yves and Thomas Naselaris. Generative adversarial networks conditioned on brain activity reconstruct seen images. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1054–1061. IEEE, 2018.

[236] C-W Sun. Biophotonics for tissue oxygenation analysis. In *Biophotonics for Medical Applications*, pages 301–320. Elsevier, 2015.

[237] Jian Sun, Huibin Li, Zongben Xu, et al. Deep ADMM-Net for compressive sensing MRI. In *Advances in neural information processing systems*, pages 10–18, 2016.

[238] Liyan Sun, Zhiwen Fan, Xinghao Ding, Yue Huang, and John Paisley. Joint CS-MRI reconstruction and segmentation with a unified deep network. In *International Conference on Information Processing in Medical Imaging*, pages 492–504. Springer, 2019.

[239] Xiao Sun, Jungwook Choi, Chia-Yu Chen, Naigang Wang, Swagath Venkataramani, Vijayalakshmi Viji Srinivasan, Xiaodong Cui, Wei Zhang, and Kailash Gopalakrishnan. Hybrid 8-bit floating point (HFP8) training and inference for deep neural networks. In *Advances in Neural Information Processing Systems*, pages 4901–4910, 2019.

[240] Yu Sun and Ulugbek S Kamilov. Stability of scattering decoder for nonlinear diffractive imaging. *arXiv preprint arXiv:1806.08015*, 2018.

[241] Yu Sun, Brendt Wohlberg, and Ulugbek S Kamilov. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 5(3):395–408, 2019.

[242] Yu Sun, Zhihao Xia, and Ulugbek S Kamilov. Efficient and accurate inversion of multiple scattering with deep learning. *Optics Express*, 26(11):14678–14688, 2018.

[243] Mayo Clinic Datase t. Low dose CT grand challenge. *Available at https://www.aapm.org/GrandChallenge/LowDoseC/*, 2014.

[244] Saeid Asgari Taghanaki, Noirin Duggan, Hillgan Ma, Xinchi Hou, Anna Celler, Francois Benard, et al. Segmentation-free direct tumor volume and metabolic activity estimation from pet scans. *Computerized Medical Imaging and Graphics*, 63:52–66, 2018.

[245] Paola Taroni, Anna Maria Paganoni, Francesca Ieva, Antonio Pifferi, Giovanna Quarto, Francesca Abbate, et al. Non-invasive optical estimate of tissue composition to differentiate malignant from benign breast lesions: A pilot study. *Scientific reports*, 7(1):1–11, 2017.

[246] Franz Thaler, Kerstin Hammernik, Christian Payer, Martin Urschler, and Darko Štern. Sparse-view CT reconstruction using wasserstein GANs. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 75–82. Springer, 2018.

[247] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[248] Mehmet Burcin Unlu, Ozlem Birgul, Roshanak Shafiiha, Gultekin Gulsen, and Orhan Nalcioglu. Diffuse optical tomographic reconstruction using multifrequency data. *Journal of Biomedical Optics*, 11(5):054008, 2006.

[249] Wouter Van den Broek and Christoph T Koch. Method for retrieval of the three-dimensional object potential by inversion of dynamical electron scattering. *Physical review letters*, 109(24):245502, 2012.

[250] Stefaan Vandenberghe, Yves D'Asseler, Rik Van de Walle, T Kauppinen, Michel Koole, Luc Bouwens, Koenraad Van Laere, Ignace Lemahieu, and RA Dierckx. Iterative reconstruction algorithms in nuclear medicine. *Computerized Medical Imaging and Graphics*, 25(2):105–111, 2001.

[251] Ashish Vaswani et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[252] Sulaiman Vesal, Shreyas Malakarjun Patil, Nishant Ravikumar, and Andreas K Maier. A multi-task framework for skin lesion detection and segmentation. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pages 285–293. Springer, 2018.

[253] Dominik Waibel, Janek Gröhl, Fabian Isensee, Thomas Kirchner, Klaus Maier-Hein, and Lena Maier-Hein. Reconstruction of initial pressure from limited view photoacoustic images using deep learning. In *Photons Plus Ultrasound: Imaging and Sensing 2018*, volume 10494, page 104942S. International Society for Optics and Photonics, 2018.

[254] Adrienne G Waks and Eric P Winer. Breast cancer treatment: a review. *Jama*, 321(3):288–300, 2019.

[255] Fei Wang et al. Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2017.

[256] Ge Wang. A perspective on deep imaging. *IEEE access*, 4:8914–8924, 2016.

[257] Ge Wang, Mannudeep Kalra, Venkatesh Murugan, Yan Xi, Lars Gjesteby, Matthew Getzin, Qingsong Yang, Wenxiang Cong, and Michael Vannier. Vision 20/20: Simultaneous CT-MRI—Next chapter of multimodality imaging. *Medical Physics*, 42(10):5879–5889, 2015.

[258] Ge Wang, Jong Chu Ye, Klaus Mueller, and Jeffrey A Fessler. Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging*, 37(6):1289–1296, 2018.

[259] Huiquan Wang, Nian Wu, Yu Cai, Lina Ren, Zhe Zhao, Guang Han, and Jinhai Wang. Optimization of reconstruction accuracy of anomaly position based on stacked auto-encoder neural networks. *IEEE Access*, 7:116578–116584, 2019.

[260] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 514–517. IEEE, 2016.

[261] Yao Wang, Jeng-Hua Chang, Raphael Aronson, Randall Locke Barbour, Harry L Graber, and Jack Lubowsky. Imaging of scattering media by diffusion tomography: an iterative perturbation approach. In *Physiological Monitoring and Early Detection Diagnostic Methods*, volume 1641, pages 58–71. International Society for Optics and Photonics, 1992.

[262] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[263] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[264] Zifeng Wang, Zhenbang Wu, Dinesh Agarwal, and Jimeng Sun. Medclip: Contrastive learning from unpaired medical images and text. *arXiv preprint arXiv:2210.10163*, 2022.

[265] Andrew Webb and George C Kagadis. Introduction to biomedical imaging. *Medical Physics*, 30(8):2267–2267, 2003.

[266] William M Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical image analysis*, 1(1):35–51, 1996.

[267] Bihan Wen, Saiprasad Ravishankar, Luke Pfister, and Yoram Bresler. Transform learning for magnetic resonance image reconstruction: From model-based learning to building neural networks. *arXiv preprint arXiv:1903.11431*, 2019.

[268] Jeffrey West, Yongqian Ma, and Paul K Newton. Capitalizing on competition: An evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment. *Journal of Theoretical Biology*, 455:249–260, 2018.

[269] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*, 36(12):2536–2545, 2017.

[270] Dufan Wu, Kyungsang Kim, Bin Dong, and Quanzheng Li. End-to-end abnormality detection in medical imaging. 2018.

[271] Dufan Wu, Kyungsang Kim, and Quanzheng Li. Computationally efficient deep neural network for computed tomography image reconstruction. *Medical physics*, 46(11):4763–4776, 2019.

[272] Sitong Wu, Zhifan Gao, Zhi Liu, Jianwen Luo, Heye Zhang, and Shuo Li. Direct reconstruction of ultrasound elastography using an end-to-end deep neural network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 374–382. Springer, 2018.

[273] Tobias Würfl, Mathis Hoffmann, Vincent Christlein, Katharina Breininger, Yixin Huang, Mathias Unberath, and Andreas K Maier. Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems. *IEEE Transactions on Medical Imaging*, 37(6):1454–1463, 2018.

[274] Junshen Xu, Enhao Gong, John Pauly, and Greg Zaharchuk. 200x low-dose PET reconstruction using deep learning. *arXiv preprint arXiv:1712.04119*, 2017.

[275] Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14383–14392, 2021.

[276] Qiong Xu, Hengyong Yu, Xuanqin Mou, Lei Zhang, Jiang Hsieh, and Ge Wang. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging*, 31(9):1682–1697, 2012.

[277] Bao Yang, Leslie Ying, and Jing Tang. Artificial neural network enhanced bayesian pet image reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1297–1309, 2018.

[278] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, 2017.

[279] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. ADMM-Net: A deep learning approach for compressive sensing MRI. *CoRR*, abs/1705.06869, 2017.

[280] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.

[281] Jong Chul Ye, Yoseob Han, and Eunju Cha. Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM Journal on Imaging Sciences*, 11(2):991–1048, 2018.

[282] Hanene Ben Yedder, Ben Cardoen, Majid Shokoufi, Farid Golnaraghi, and Ghassan Hamarneh. Multitask deep learning reconstruction and localization of lesions in limited angle diffuse optical tomography. *IEEE Transactions on Medical Imaging*, 41(3):515–530, 2021.

[283] Hanene Ben Yedder, Qingye Ding, Umme Zakia, Zhida Li, Soroush Haeri, and Ljiljana Trajkovic. Comparison of virtualization algorithms and topologies for data center networks. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–6. IEEE, 2017.

[284] Hanene Ben Yedder, Umme Zakia, Aly Ahmed, and Ljiljana Trajković. Modeling prediction in recommender systems using restricted boltzmann machine. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2063–2068. IEEE, 2017.

[285] Jaejun Yoo, Sohail Sabir, Duchang Heo, Kee Hyun Kim, Abdul Wahab, Yoonseok Choi, Seul-I Lee, Eun Young Chae, Hak Hee Kim, Young Min Bae, et al. Deep learning diffuse optical tomography. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 39(4):877 – 887, 2020.

[286] Yeo Hun Yoon, Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye. Efficient b-mode ultrasound image reconstruction from sub-sampled RF data using deep learning. *IEEE Transactions on Medical Imaging*, 38(2):325–336, 2018.

[287] Xiang Yu, Jian Wang, Qing-Qi Hong, Raja Teku, Shui-Hua Wang, and Yu-Dong Zhang. Transfer learning for medical images analyses: A survey. *Neurocomputing*, 489:230–254, 2022.

[288] Zhou Yu, Yuhao Cui, Jun Yu, Meng Wang, Dacheng Tao, and Qi Tian. Deep multimodal neural architecture search. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3743–3752, 2020.

[289] Jie Yuan, Guan Xu, Yao Yu, Yu Zhou, Paul L Carson, Xueding Wang, and Xiaojun Liu. Real-time photoacoustic and ultrasound dual-modality imaging system facilitated with graphics processing unit and code parallel optimization. *Journal of Biomedical Optics*, 18(8):086001, 2013.

[290] Umme Zakia and Hanene Ben Yedder. Dynamic load balancing in sdn-based data center networks. In *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 242–247. IEEE, 2017.

[291] Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3712–3722, 2018.

[292] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021.

[293] Jure Zbontar, Florian Knoll, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, Hersh Chandarana, et al. FastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018.

[294] Haimiao Zhang and Bin Dong. A review on deep learning in medical image reconstruction. *arXiv preprint arXiv:1906.10643*, 2019.

[295] Hao Zhang, Dong Zeng, Hua Zhang, Jing Wang, Zhengrong Liang, and Jianhua Ma. Applications of nonlocal means algorithm in low-dose X-ray CT image processing and reconstruction: A review. *Medical physics*, 44(3):1168–1185, 2017.

[296] Lin Zhang and Guanglei Zhang. Brief review on learning-based methods for optical tomography. *Journal of Innovative Optical Health Sciences*, 12(06):1930011, 2019.

[297] Qiyang Zhang and Dong Liang. Visualization of fully connected layer weights in deep learning ct reconstruction. *arXiv preprint arXiv:2002.06788*, 2020.

[298] Yi Zhang, Yan Wang, Weihua Zhang, Feng Lin, Yifei Pu, and Jiliu Zhou. Statistical iterative reconstruction using adaptive fractional order regularization. *Biomedical Optics Express*, 7(3):1015–1029, 2016.

[299] Yi Zhang, Yan Xi, Qingsong Yang, Wenxiang Cong, Jiliu Zhou, and Ge Wang. Spectral CT reconstruction with image sparsity and spectral mean. *IEEE Transactions on Computational Imaging*, 2(4):510–523, 2016.

[300] Yifei Zhang, Désiré Sidibé, Olivier Morel, and Fabrice Mériaudeau. Deep multimodal fusion for semantic image segmentation: A survey. *Image and Vision Computing*, 105:104042, 2021.

[301] Bo Zhao, Justin P Haldar, Anthony G Christodoulou, and Zhi-Pei Liang. Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints. *IEEE Transactions on Medical Imaging*, 31(9):1809–1820, 2012.

[302] Hubin Zhao, Elisabetta M Frijia, Ernesto Vidal Rosas, Liam Collins-Jones, Greg Smith, Reuben Nixon-Hill, Samuel Powell, Nicholas L Everdell, and Robert J Cooper. Design and validation of a mechanically flexible and ultra-lightweight high-density diffuse optical tomography system for functional neuroimaging of newborns. *Neurophotonics*, 8(1):015011, 2021.

[303] Yongyi Zhao, Ankit Raghuram, Fay Wang, Stephen Hyunkeol Kim, Andreas H Hielscher, Jacob T Robinson, et al. Unrolled-dot: an interpretable deep network for diffuse optical tomography. *Journal of Biomedical Optics*, 28(3):036002, 2023.

[304] Zhizhen Zhao, Jong Chul Ye, and Yoram Bresler. Generative models for inverse imaging problems: From mathematical foundations to physics-driven applications. *IEEE Signal Processing Magazine*, 40(1):148–163, 2023.

[305] Xiantong Zhen and Shuo Li. Towards direct medical image analysis without segmentation. *arXiv preprint arXiv:1510.06375*, 2015.

[306] Bo Zhou, Xunyu Lin, and Brendan Eck. Limited angle tomography reconstruction: Synthetic reconstruction via unsupervised sinogram adaptation. In *International Conference on Information Processing in Medical Imaging*, pages 141–152. Springer, 2019.

[307] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487, 2018.

[308] Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. In *Advances in Neural Information Processing Systems*, pages 14747–14756, 2019.

[309] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(5):363–387, 2012.

[310] Yun Zou, Yifeng Zeng, Shuying Li, and Quing Zhu. Machine learning model with physical constraints for diffuse optical tomography. *Biomedical Optics Express*, 12(9):5720–5735, 2021.

[311] Yun Zou, Yifeng Zeng, Shuying Li, and Quing Zhu. Unsupervised machine learning model for dot reconstruction. In *Optical Tomography and Spectroscopy of Tissue XIV*, volume 11639, pages 23–36. SPIE, 2021.