Towards Generalisable Inverse Modelling in Frequency-Domain Diffuse Optical Tomography Using a Hybrid CNN–Transformer

MSc Dissertation

Max Hart

School of Computer Science

College of Engineering and Physical Sciences

University of Birmingham

2024-25

Abstract

Acknowledgements

Abbreviations

ACB	Apple Banana Carrot

Contents

Abstract ii

Acknowledgements iii

Abbreviations iv

List of Figures ix

List of Tables x

1 Introduction 1

- 1.1 Background and Motivation 1
- 1.2 Frequency-Domain Diffuse Optical Tomography 2
- 1.3 Problem Formulation and Notation 2
- 1.4 Challenges in FD-DOT Reconstruction 3
- 1.5 Research Objectives and Contributions 4
- 1.6 Dissertation Structure 5

2 Literature Review 6

- 2.1 Optical Physics and Modelling Foundations 6
  - 2.1.1 From Radiative Transport to Diffusion Approximation 6
  - 2.1.2 Numerical Forward Modelling in FD-DOT 7
- 2.2 Conventional DOT Reconstruction Methods 7
  - 2.2.1 Iterative Inversion Frameworks 7
  - 2.2.2 Regularisation Strategies 8
- 2.3 Deep Learning in Medical Inverse Problems 8
  - 2.3.1 Overview Across Modalities 8
  - 2.3.2 DOT-Specific Applications 9
- 2.4 Neural Architectures for DOT 9
  - 2.4.1 CNN Approaches 9
  - 2.4.2 Transformers and Attention 9

2.4.3 Hybrid and Two-Stage Frameworks 9
2.5 Challenges of Generalisation in DL-DOT 9
2.5.1 Geometry Shift 9
2.5.2 Noise and Measurement Variability 10
2.5.3 Simulation-to-Real Gap 10
2.6 Baseline and Research Gap 10
2.6.1 Dale's Hybrid FD-DOT 10
2.6.2 Open Gaps and Research Opportunity 10

3 Physics-Based Synthetic Data Pipeline **11**
3.1 Forward Modelling in the Frequency Domain 11
3.2 Geometric Phantom Construction 11
3.2.1 Ellipsoidal Tissue and Inclusion Design 11
3.2.2 SO(3) Rotations and Spatial Bias Mitigation 11
3.3 Optical Property Assignment 11
3.4 Surface Extraction and Probe Placement 11
3.4.1 Binary Morphological Surface Extraction 11
3.4.2 Surface-Constrained Source–Detector Placement 11
3.5 Probe Geometry Randomisation and Tokenisation (Contribution) 11
3.6 Noise Model 11
3.7 Dataset Composition and Preprocessing 11
3.7.1 Standardisation and Leakage Prevention 11
3.7.2 HDF5 Design and DataLoader Throughput 11

4 Proposed Hybrid CNN–Transformer Model **12**
4.1 Architectural Overview 13
4.2 Stage 1: CNN Autoencoder (Teacher Prior) 13
4.2.1 Encoder: Residual Blocks and Downsampling 13
4.2.2 Latent Space Design (d_z=256) 13
4.2.3 Decoder: Progressive Upsampling and Reconstruction 13
4.3 Stage 2: Spatially-Aware Measurement Embedding (Contribution) 13
4.3.1 Signal Branch for $\{\log A, \varphi\}$ 13
4.3.2 Position Branch for $\{\mathbf{x}_{\text{src}}, \mathbf{x}_{\text{det}}\}$ 13
4.3.3 Fusion and Token Formation 13
4.4 Transformer Encoder 13
4.4.1 Attention, Depth, and Positional Handling 13
4.4.2 Regularisation and Capacity Control 13
4.5 Aggregation via Multi-Query Attention Pooling (Contribution) 13
4.6 Latent Alignment and Decoder Reuse 13
4.7 Inference Pathway and Computational Cost 13

5 Training Strategies and Optimisation 14
5.1 Stage 1: CNN Pre-Training . . . 14
5.1.1 Loss Function and Schedules . . . 14
5.1.2 Stability Tricks (Mixed Precision, Gradient Clipping) . . . 14
5.2 Stage 2: Transformer Training with Latent Alignment . . . 14
5.2.1 Latent RMSE-Only Objective (Contribution) . . . 14
5.2.2 Decoder Unfreezing Protocol (Contribution) . . . 14
5.2.3 Measurement Subsampling and Augmentation . . . 14
5.2.4 Optimisers, LR Schedules, and Weight Decay . . . 14
5.3 Implementation Details and Reproducibility . . . 14

6 Experimental Results and Analysis 15
6.1 Experimental Setup . . . 15
6.1.1 Datasets and Splits . . . 15
6.1.2 Metrics (Latent RMSE; Voxel RMSE/SSIM for μ_a, μ_s') . . . 15
6.1.3 Hardware and Runtime Reporting . . . 15
6.2 Stage 1 Results: Autoencoder Reconstruction Quality . . . 15
6.3 Stage 2 Results: Transformer Enhancement . . . 15
6.4 Ablations (Contribution-Focused) . . . 15
6.4.1 Mean Pooling vs Multi-Query Attention . . . 15
6.4.2 Fixed Geometry vs Randomised Geometry + L=256 . . . 15
6.4.3 With vs Without Decoder Unfreezing . . . 15
6.4.4 Embedding Variants: With/Without Position Branch . . . 15
6.4.5 Sequence Length Sensitivity: $L \in \{128, 256, 512\}$ . . . 15
6.5 Generalisation to Held-Out Probe Layouts (Contribution) . . . 15
6.6 Comparative Analysis with Prior Work (Dale Baseline) . . . 15
6.7 Qualitative Visualisation and Error Analysis . . . 15

7 Discussion 16
7.1 Key Findings . . . 16
7.2 Robustness to Geometry and Noise Shift . . . 16
7.3 Clinical Implications of a Geometry-Robust DOT Model . . . 16
7.4 Limitations and Threats to Validity . . . 16
7.5 Computational Efficiency and Practical Deployment . . . 16
7.6 Positioning Against Prior Work (Dale) and Field Impact . . . 16

8 Conclusion and Future Work 17
8.1 Summary of Contributions . . . 17
8.2 Future Directions . . . 17
8.3 Final Remarks . . . 17

A Implementation and Hyperparameters 20

B Extended Dataset Examples and Probe Layouts 21

C Additional Quantitative Results **22**

D Mathematical Derivations **23**

E Reproducibility Checklist and Ethics Statement **24**

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Background and Motivation

Diffuse optical tomography (DOT) is a non-invasive imaging modality that reconstructs the optical properties of tissue from near-infrared (NIR) light measurements. By exploiting the wavelength-dependent absorption and scattering of NIR photons in biological tissue, DOT enables three-dimensional imaging of physiological parameters such as blood volume and oxygenation. These parameters are biomarkers of vascularisation and haemodynamics, making DOT attractive for oncology and neuroscience. Unlike ionising modalities such as computed tomography (CT) or positron emission tomography (PET), DOT is inherently safe for repeated use, cost-effective, and portable. These features make it well suited to longitudinal monitoring, point-of-care screening, and intraoperative guidance [1, 2].

These advantages translate directly into clinical impact. Among the most widely studied applications is breast cancer imaging. Breast cancer remains the most common cancer in women worldwide, and both early detection and accurate monitoring of treatment response are critical determinants of outcome. DOT can provide functional information such as tumour oxygenation and haemoglobin concentration, which are not readily accessible using conventional imaging techniques. In particular, DOT has been investigated for monitoring patient response to neoadjuvant chemotherapy, where frequent, non-invasive, and low-cost imaging is required but impractical with MRI or mammography [3]. In neuroscience, DOT-based functional imaging has been applied to study brain activation and cerebral oxygenation, offering a portable and low-cost alternative to fMRI that is especially valuable for bedside or paediatric use [4].

Recent advances in instrumentation have further expanded the potential of DOT. Handheld and wearable systems are now capable of real-time scanning and adapting to varied patient geometries [5]. These developments, however, shift the bottleneck towards computation: reconstruction must be both rapid and robust to probe variability, anatomical diversity, and measurement noise. Conventional algorithms based on iterative inversion of the diffusion equation with repeated Jacobian updates remain prohibitively slow, often requiring minutes per reconstruction even on high-performance hardware. Deep learning-based DOT (DL-DOT) has therefore emerged as a promising paradigm, offering sub-second inference while maintaining or improving upon the fidelity of physics-based solvers [6].

1.2 Frequency-Domain Diffuse Optical Tomography

DOT relies on the propagation of near-infrared (NIR) light, typically within the 650–900 nm optical window, through biological tissue. This range maximises haemoglobin contrast while minimising water and lipid absorption, enabling photons to penetrate several centimetres. The key optical parameters are the absorption coefficient, μ_a (mm^{-1}), which reflects the concentration of chromophores, and the reduced scattering coefficient, μ_s' (mm^{-1}), which is influenced by tissue microstructure and morphology. Together, these parameters govern the photon fluence distribution and form the fundamental quantities to be reconstructed [1, 2].

Measurements are typically acquired using arrays of light sources and detectors placed on the tissue surface. Each source–detector (SD) pair samples a diffuse photon path through tissue, with the corresponding source–detector separation (SDS) strongly influencing depth sensitivity. In frequency-domain DOT (FD-DOT), the light is sinusoidally modulated at a radiofrequency (e.g. 140 MHz), and the detected signal is characterised by both amplitude attenuation and phase shift relative to the input. The combination of $\log A$ (log-amplitude) and ϕ (phase) provides complementary sensitivity to μ_a and μ_s'. Short SDS values (< 15 mm) primarily probe superficial layers, whereas larger separations (30–40 mm) improve sensitivity to deeper tissue, albeit with reduced signal-to-noise ratio.

The forward model of light transport is derived from the frequency-domain diffusion equation, which, for realistic geometries, is solved numerically using the finite element method (FEM) [7]. The forward operator $\mathcal{F}$ maps spatial distributions of μ_a and μ_s' to measurable boundary data. The corresponding inverse problem is to recover these parameters from sparse, surface-only measurements. Because this problem is underdetermined, additional constraints must be imposed to stabilise reconstructions. These constraints — collectively known as regularisation — may take the form of enforcing smoothness, assuming sparsity of inclusions, or applying data-driven priors learned by neural networks [1].

Terminology note: From this point forward, this dissertation will use *FD-DOT* to denote the specific imaging problem studied here, i.e. frequency-domain measurements of amplitude and phase in the NIR window. The term *DOT* will be used only when referring to the modality in general, while *DL-DOT* will be used specifically to denote deep learning-based reconstruction methods.

1.3 Problem Formulation and Notation

This dissertation focuses on frequency-domain diffuse optical tomography (FD-DOT), in which the photon field is sinusoidally modulated at frequency f (Hz). For each source–detector (SD) pair, the measurement is expressed as a complex value:

$$M = Ae^{i\phi},$$

where A is the detected amplitude and ϕ is the phase shift relative to the source. Because amplitudes span several orders of magnitude, reconstructions are typically performed using the logarithm of amplitude, $\log A$, together with ϕ. These two quantities form the core measurement features for each SD pair.

Each SD pair is further represented by the three-dimensional coordinates of both the source and detector positions, (x_s, y_s, z_s) and (x_d, y_d, z_d). The complete feature vector for a single pair is therefore:

$$m_i = \{\log A_i, \phi_i, x_{s,i}, y_{s,i}, z_{s,i}, x_{d,i}, y_{d,i}, z_{d,i}\},$$

which integrates optical information with spatial context. For a scan comprising N SD pairs, the full measurement tensor is:

$$\mathbf{y} = \{m_1, m_2, \dots, m_N\} \in \mathbb{R}^{N\times 8}.$$

The forward model of FD-DOT is governed by the frequency-domain diffusion equation, which for realistic geometries is solved numerically using the finite element method (FEM). This defines the mapping:

$$\mathbf{y} = \mathcal{F}(\mu_a, \mu_s') + \epsilon,$$

where $\mathcal{F}$ denotes the FEM-based forward operator from optical properties to boundary measurements, and ϵ represents additive measurement noise. In this work, ϵ is modelled as Gaussian perturbations with 0.5% relative variance on $\log A$ and $\pm 0.5^\circ$ absolute variance on ϕ. This simplified model captures the dominant sensitivity of FD-DOT systems to amplitude and phase fluctuations while remaining computationally tractable.

The reconstruction task is to estimate voxelwise distributions of μ_a and μ_s' across a three-dimensional grid. This study adopts a $64 \times 64 \times 64$ discretisation at 1 mm resolution, yielding approximately 2.6×10^5 voxels per parameter, or about 5.2×10^5 unknowns in total. Denoting the reconstructions by $\hat{\mu}_a$ and $\hat{\mu}_s'$, the inverse mapping can be expressed as:

$$\mathcal{G} : \mathbf{y} \mapsto \{\hat{\mu}_a, \hat{\mu}_s'\},$$

where $\mathcal{G}$ is implemented by a learned neural network.

This inverse problem is severely underdetermined. Even with $N = 1000$ SD pairs, the measurement tensor $\mathbf{y} \in \mathbb{R}^{1000\times 8}$ contains far fewer entries than the hundreds of thousands of voxel values to be recovered. Moreover, measurement sensitivity is non-uniform, with superficial voxels contributing disproportionately more than deeper ones. This imbalance renders the problem intrinsically unstable in the absence of strong priors. The central challenge addressed in this dissertation is therefore the design of models and training strategies that embed spatial priors, incorporate probe geometry variability, and generalise across anatomically diverse tissue phantoms while maintaining fidelity in both μ_a and μ_s' reconstructions.

1.4 Challenges in FD-DOT Reconstruction

The challenges in FD-DOT arise from both the physics of light transport and the requirements of clinical deployment:

- **Ill-posedness:** Sparse, surface-only measurements must be mapped to dense three-dimensional volumes. Each detector records photons that have undergone multiple scattering events, producing broad and overlapping sensitivity profiles. Deep tissue regions contribute disproportionately weak signals, which amplifies inversion instability. Without strong priors, reconstructions overfit superficial structures while failing to capture deeper inclusions.
- **Geometry shift:** In handheld or wearable systems, probe geometry is rarely fixed. Source–detector separations vary with operator handling, patient anatomy, and motion. FEM-based solvers can accommodate arbitrary layouts by recomputing Jacobians, but most deep learning models are trained on fixed geometries and degrade significantly when the layout changes [8]. Overcoming this limitation is essential for achieving path-agnostic and clinically practical FD-DOT.
- **Noise robustness:** FD-DOT measurements are influenced by electronic noise, coupling variability, and instrumental fluctuations. In this work, noise is modelled as Gaussian perturbations (0.5% on $\log A$ and $\pm 0.5^\circ$ on ϕ), a deliberate simplification that reflects typical system variability. Even small perturbations, however, can destabilise reconstructions if not explicitly considered during training, highlighting the need for noise-aware learning pipelines.
- **Sim-to-real gap:** Large-scale synthetic datasets enable supervised training but cannot fully replicate the heterogeneity of patient anatomy, motion artefacts, or hardware imperfections. This mismatch introduces a domain gap between simulated and clinical data, which must be narrowed to enable reliable deployment of DL-DOT systems.
- **Latency:** Real-time use requires reconstructions in under 0.1 s per volume. Iterative solvers are far too slow, often demanding hundreds of iterations per scan. Learned inverse solvers are therefore essential to achieve clinically viable runtimes while maintaining image quality.

1.5 Research Objectives and Contributions

The overarching aim of this dissertation is to advance diffuse optical tomography (DOT) towards models that generalise across diverse probe geometries and anatomies, addressing a central limitation of existing deep learning-based DOT (DL-DOT) approaches. Building on the hybrid CNN–Transformer paradigm introduced by Dale [6, 8], this work investigates new strategies in data generation, architectural refinement, and evaluation to improve robustness and clinical viability. The principal objectives and contributions are as follows:

1. **Phantom and probe diversity for generalisation:** A high-throughput phantom generation pipeline was developed that extends well beyond Dale's slab-like tissue models. Ellipsoidal tissue volumes were embedded inside cubic air domains, creating tissue–air boundaries from which local surface patches defined source–detector placement. Tumour inclusions varied in size and shape, while probe positions were randomly distributed across accessible surfaces. To avoid spatial bias, phantoms were randomly rotated in three dimensions using the full SO(3) rotation group.
2. **Systematic geometry randomisation:** Each phantom produced 1000 source–detector

(SD) measurements, dynamically subsampled into fixed 256-token sequences. This enforces invariance to probe placement, provides strong data augmentation, and addresses degradation under geometry shift—one of the key barriers to clinical deployment of DL-DOT.

3. **Hybrid CNN–Transformer framework:** A two-stage hybrid network was implemented in line with Dale's design philosophy, but developed independently without access to specific architectural details. Stage 1 trains a 3D CNN autoencoder on ground-truth absorption and scattering volumes to establish a spatial latent representation. Stage 2 employs a transformer encoder with spatially aware embeddings that fuse optical measurements ($\log A$, ϕ) with explicit source–detector coordinates, enabling robust volumetric reconstruction.
4. **Architectural refinements:** Beyond the baseline design, improvements include an enhanced spatial embedding scheme, multi-query attention for global aggregation, and selective fine-tuning of the CNN decoder. Additional refinements are described in Chapter 4. Together, these modifications increase the expressivity and stability of the learned solver.
5. **Evaluation of robustness:** The framework was assessed on held-out phantoms from the synthetic test set, providing unseen anatomical shapes and probe configurations. Comparative reference to Dale's baseline work contextualises the results. The aim was not to surpass prior performance but to investigate whether the proposed strategies improved generalisation under phantom and probe variability.

In summary, this dissertation explores how phantom diversity, geometry randomisation, and architectural refinements can be combined within a hybrid CNN–Transformer framework to investigate pathways towards more generalisable DL-DOT. The focus is on testing robustness to geometry variation, phantom diversity, and measurement noise, with the broader goal of informing future work on clinically viable reconstruction methods.

1.6 Dissertation Structure

The remainder of this dissertation is organised as follows. Chapter 2 reviews the theoretical and methodological background, covering NIR light transport physics, conventional DOT reconstruction algorithms, and recent developments in DL-DOT, with emphasis on CNN and transformer architectures. Chapter 3 details the synthetic data generation pipeline, including phantom construction, optical property assignment, probe placement strategies, and noise modelling. Chapter 4 describes the proposed hybrid CNN–Transformer framework, outlining the architectural design, spatial embedding strategies, and latent alignment mechanism. Chapter 5 explains the training methodology, optimisation schedules, and implementation considerations. Chapter 6 presents the experimental results, including ablation studies and comparative evaluations against baseline methods. Chapter 7 provides a critical discussion of findings, highlighting robustness, limitations, and clinical implications. Finally, Chapter 8 concludes with a summary of contributions and outlines directions for future research.

This structure is intended to guide the reader from fundamental principles, through data generation and model development, to experimental validation and critical discussion. The next chapter therefore sets the stage by introducing the theoretical background of light transport and conventional DOT reconstruction methods.

CHAPTER 2

Literature Review

2.1 Optical Physics and Modelling Foundations

The reconstruction of optical properties in diffuse optical tomography (DOT) depends critically on accurate forward models of photon transport. While Chapter 1 outlined the principles of frequency-domain DOT, this section situates these concepts within the broader theoretical and computational literature. Two themes dominate: the simplification of photon transport from the radiative transport equation to the diffusion approximation, and the numerical strategies used to solve the resulting forward problem in realistic geometries.

2.1.1 From Radiative Transport to Diffusion Approximation

At its most fundamental level, light in tissue is described by the radiative transport equation (RTE), which tracks photons as they scatter and are absorbed across both space and angle. The RTE is exact but extremely challenging to solve in practice, as it requires keeping track of too many variables simultaneously [1].

Fortunately, biological tissue is typically highly scattering: photons undergo many scattering events for each absorption. In this regime, the angular distribution of light quickly becomes nearly isotropic, meaning it is no longer necessary to model the exact photon directions. By exploiting this simplification, the RTE reduces to the diffusion equation, which captures the bulk flow of photons while ignoring fine angular detail. This transition was formalised in seminal work by Arridge *et al.* (1999) [1], and further reviewed by Gibson *et al.* (2005) [2] and Boas *et al.* (2001) [9].

The diffusion approximation has since become the standard model in DOT because it balances realism with tractability. It works particularly well in the near-infrared window (650–900 nm), where scattering dominates over absorption. Its main limitations arise near tissue boundaries or in low-scattering regions, where the assumption of isotropy breaks down [10]. Even so, for most DOT applications the diffusion model provides sufficient accuracy to form the foundation of both classical and modern reconstruction algorithms.

2.1.2 Numerical Forward Modelling in FD-DOT

In practice, solving the diffusion equation in realistic geometries requires numerical discretisation. The finite element method (FEM) is the most widely adopted approach in DOT because of its flexibility with complex tissue boundaries and heterogeneous optical properties [7]. FEM discretises the tissue volume into small elements, enabling accurate representation of domains such as breast or brain geometries. The forward operator $\mathcal{F}$ is then constructed to map spatially varying absorption and scattering coefficients to predicted boundary measurements of log-amplitude and phase.

Alternative schemes, such as finite difference (FDM) or boundary element methods, have also been explored, though they are less common in modern DOT. FEM remains dominant because it integrates naturally with anatomical priors from MRI/CT and supports frequency-domain extensions, where sinusoidal modulation is incorporated into the diffusion equation. The computational cost, however, is significant: each forward solve involves very large systems of equations, and repeated sensitivity calculations are required for iterative inverse solvers. These bottlenecks are central to why conventional FD-DOT reconstructions are slow (often minutes per scan) and why deep learning has become a compelling alternative [10, 7].

In summary, the literature on optical modelling establishes a clear progression: the diffusion approximation provides a tractable surrogate to the full radiative transport equation, while FEM-based solvers have become the de facto standard for handling anatomical complexity. These foundations naturally lead to the next challenge: how to invert these forward models to reconstruct images, which is the focus of the following section.

2.2 Conventional DOT Reconstruction Methods

While the diffusion equation provides the forward model for photon transport, the inverse problem of recovering absorption and scattering maps from boundary data is considerably more challenging. This section reviews the classical literature on DOT reconstruction, focusing on iterative inversion frameworks and the role of regularisation in stabilising solutions.

2.2.1 Iterative Inversion Frameworks

Conventional DOT reconstruction is typically framed as an optimisation task: given a forward operator $\mathcal{F}$, the goal is to recover $\{\mu_a, \mu_s'\}$ such that the predicted boundary data match the measured values. A standard formulation is:

$$\hat{\mu} = \arg\min_{\mu} \ \|\mathbf{y} - \mathcal{F}(\mu)\|^2 + \lambda R(\mu),$$

where $\mu = \{\mu_a, \mu_s'\}$ are the optical parameters, $\mathbf{y}$ are the boundary measurements, $R(\mu)$ is a regularisation term, and λ balances data fidelity against prior assumptions.

The most common approach to solving this optimisation is Jacobian-based iterative inversion, in which the forward model is linearised around a current guess and updated step by step. Gauss–Newton and Levenberg–Marquardt schemes are widely used examples [1]. These methods can produce reasonable reconstructions, but they are slow: each iteration requires solving a large

FEM system and recomputing sensitivities, often leading to runtimes of several minutes per scan.

To improve stability, priors are introduced. Tikhonov regularisation is the classical choice, penalising large parameter deviations to suppress noise. More advanced approaches, such as sparsity or total variation (TV) penalties, encourage sharper boundaries and are particularly useful for detecting localised tumours [2]. The trade-off, however, is that such priors must be tuned carefully and can make solutions more sensitive to noise.

2.2.2 Regularisation Strategies

Because DOT is highly underdetermined, regularisation plays a central role. Smoothness priors are the simplest option, enforcing gradual variation in optical properties. This works well for largely homogeneous tissue but risks blurring small inclusions.

Probabilistic approaches extend this by framing DOT as a Bayesian inference problem, where optical parameters are treated as random variables with prior distributions. Maximum a posteriori (MAP) formulations, as reviewed by Tarvainen *et al.* (2010) [11], provide both stability and a means to quantify uncertainty, though at the cost of added computation.

A further strategy is to incorporate structural information from other imaging modalities such as MRI or CT, using anatomical priors to guide reconstructions [1]. This reduces ambiguity and improves localisation, but also limits flexibility since DOT becomes dependent on external data.

In summary, classical DOT solvers have evolved from basic iterative updates to more sophisticated regularised and anatomically guided methods. While they remain valuable, their reliance on slow Jacobian computations, hand-tuned priors, and limited robustness to noise has motivated a shift towards deep learning-based approaches, where priors can be learned directly from data. This transition is the focus of the next section.

2.3 Deep Learning in Medical Inverse Problems

The rapid success of deep learning (DL) in vision and language tasks has naturally extended to medical imaging, where inverse problems such as tomographic reconstruction are central. This section situates DOT within that broader trend by first reviewing DL applications across established modalities before narrowing to DOT-specific efforts.

2.3.1 Overview Across Modalities

Inverse problems in medical imaging—whether reconstructing attenuation maps in CT, proton densities in MRI, or absorption distributions in photoacoustics—share a common challenge: recovering high-dimensional fields from sparse or indirect measurements. Deep learning has been applied to these problems in two main ways: *post-processing* of conventionally reconstructed images to remove artefacts, and *direct inversion* where networks learn mappings from raw measurement data to images without iterative solvers [12].

In CT, convolutional networks have been trained to denoise and de-streak low-dose reconstructions, reducing radiation exposure while maintaining diagnostic quality. In MRI, unrolled neural networks embed physics-based constraints directly into the architecture, accelerating acquisition and reconstruction. Photoacoustic tomography has similarly leveraged DL to suppress limited-view artefacts and boost spatial resolution. Across these domains, the key advantage of DL is its ability to learn data-driven priors that stabilise ill-posed inversions while enabling near real-time inference once trained.

2.3.2 DOT-Specific Applications

Compared with CT or MRI, DL research in DOT is still at an early stage, but several studies have demonstrated its potential. Early work by Feng *et al.* (2020) trained a U-Net to reconstruct absorption maps from simulated DOT data, reporting sharper images and fewer artefacts than Tikhonov-regularised FEM solvers [13]. Subsequent studies have explored refinements such as residual CNNs and adversarial objectives, though most remain constrained to fixed probe geometries and relatively small synthetic datasets.

These initial applications highlight both promise and limitation. CNNs can achieve reconstructions orders of magnitude faster than iterative solvers, but their ability to generalise across different geometries and anatomies is weak. This limitation motivates more flexible architectures—such as hybrid CNN–Transformer designs—that can explicitly incorporate spatial context and better handle variability in patient anatomy and probe placement.

In summary, while DL has already reshaped other imaging modalities, its role in DOT is still emerging. The literature demonstrates feasibility but also reveals bottlenecks in robustness and generalisation. These observations naturally set the stage for the next section, which focuses on CNN-based approaches in DOT.

2.4 Neural Architectures for DOT

2.4.1 CNN Approaches

Detail 2D/3D CNN encoder–decoders, U-Nets, and autoencoders applied to DOT, including performance trade-offs.

2.4.2 Transformers and Attention

Review the adoption of attention in medical imaging, benefits of long-range modelling, and emerging applications in inverse problems.

2.4.3 Hybrid and Two-Stage Frameworks

Discuss hybrid CNN–Transformer designs, teacher–student training, and latent alignment strategies from the literature.

2.5 Challenges of Generalisation in DL-DOT

2.5.1 Geometry Shift

Summarise literature on probe variability, geometry-aware networks, and failures under shift.

2.5.2 Noise and Measurement Variability

Review noise modelling in synthetic vs. experimental DOT data, and robustness strategies in ML.

2.5.3 Simulation-to-Real Gap

Survey domain adaptation, transfer learning, and physics-informed approaches proposed to narrow the gap.

2.6 Baseline and Research Gap

2.6.1 Dale's Hybrid FD-DOT

Review Dale's recent baseline work, outlining his contributions (latent alignment, pooling, geometry assumptions).

2.6.2 Open Gaps and Research Opportunity

Position this dissertation: prior work struggles with geometry-agnostic generalisation and diverse phantom representations. Identify the niche where this project contributes.

CHAPTER 3

Physics-Based Synthetic Data Pipeline

3.1 Forward Modelling in the Frequency Domain

3.2 Geometric Phantom Construction

3.2.1 Ellipsoidal Tissue and Inclusion Design

3.2.2 SO(3) Rotations and Spatial Bias Mitigation

3.3 Optical Property Assignment

3.4 Surface Extraction and Probe Placement

3.4.1 Binary Morphological Surface Extraction

3.4.2 Surface-Constrained Source–Detector Placement

3.5 Probe Geometry Randomisation and Tokenisation (Contribution)

3.6 Noise Model

3.7 Dataset Composition and Preprocessing

3.7.1 Standardisation and Leakage Prevention

3.7.2 HDF5 Design and DataLoader Throughput

CHAPTER 4

Proposed Hybrid CNN–Transformer Model

4.1 Architectural Overview

4.2 Stage 1: CNN Autoencoder (Teacher Prior)

4.2.1 Encoder: Residual Blocks and Downsampling

4.2.2 Latent Space Design (d_z=256)

4.2.3 Decoder: Progressive Upsampling and Reconstruction

4.3 Stage 2: Spatially-Aware Measurement Embedding (Contribution)

4.3.1 Signal Branch for $\{\log A, \varphi\}$

4.3.2 Position Branch for $\{\mathbf{x}_{\mathrm{src}}, \mathbf{x}_{\mathrm{det}}\}$

4.3.3 Fusion and Token Formation

4.4 Transformer Encoder

4.4.1 Attention, Depth, and Positional Handling

4.4.2 Regularisation and Capacity Control

4.5 Aggregation via Multi-Query Attention Pooling (Contribution)

4.6 Latent Alignment and Decoder Reuse

4.7 Inference Pathway and Computational Cost

CHAPTER 5

Training Strategies and Optimisation

5.1 Stage 1: CNN Pre-Training

5.1.1 Loss Function and Schedules

5.1.2 Stability Tricks (Mixed Precision, Gradient Clipping)

5.2 Stage 2: Transformer Training with Latent Alignment

5.2.1 Latent RMSE-Only Objective (Contribution)

5.2.2 Decoder Unfreezing Protocol (Contribution)

5.2.3 Measurement Subsampling and Augmentation

5.2.4 Optimisers, LR Schedules, and Weight Decay

5.3 Implementation Details and Reproducibility

CHAPTER 6

Experimental Results and Analysis

6.1 Experimental Setup

6.1.1 Datasets and Splits

6.1.2 Metrics (Latent RMSE; Voxel RMSE/SSIM for μ_a, μ_s')

6.1.3 Hardware and Runtime Reporting

6.2 Stage 1 Results: Autoencoder Reconstruction Quality

6.3 Stage 2 Results: Transformer Enhancement

6.4 Ablations (Contribution-Focused)

6.4.1 Mean Pooling vs Multi-Query Attention

6.4.2 Fixed Geometry vs Randomised Geometry + L=256

6.4.3 With vs Without Decoder Unfreezing

6.4.4 Embedding Variants: With/Without Position Branch

6.4.5 Sequence Length Sensitivity: $L \in \{128, 256, 512\}$

6.5 Generalisation to Held-Out Probe Layouts (Contribution)

6.6 Comparative Analysis with Prior Work (Dale Baseline)

6.7 Qualitative Visualisation and Error Analysis

CHAPTER 7

Discussion

7.1 Key Findings

7.2 Robustness to Geometry and Noise Shift

7.3 Clinical Implications of a Geometry-Robust DOT Model

7.4 Limitations and Threats to Validity

7.5 Computational Efficiency and Practical Deployment

7.6 Positioning Against Prior Work (Dale) and Field Impact

CHAPTER 8

Conclusion and Future Work

8.1 Summary of Contributions

8.2 Future Directions

8.3 Final Remarks

Bibliography

[1] Simon R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41–R93, 1999.

[2] Adam P. Gibson, Jeremy C. Hebden, and Simon R. Arridge. Recent advances in diffuse optical imaging. *Physics in Medicine and Biology*, 50(4):R1–R43, 2005.

[3] Bruce J Tromberg, Zheng Zhang, Anaïs Leproux, Thomas D O'Sullivan, Albert E Cerussi, Philip M Carpenter, Rita S Mehta, Darren Roblyer, Wei Yang, Keith D Paulsen, et al. Predicting responses to neoadjuvant chemotherapy in breast cancer: Acrin 6691 trial of diffuse optical spectroscopic imaging. *Cancer Research*, 76(20):5933–5944, 2016.

[4] Adam T. Eggebrecht, Benjamin R. White, Silvina L. Ferradal, Cheng Chen, You Zhan, Abraham Z. Snyder, Hamid Dehghani, and Joseph P. Culver. Mapping distributed brain function and networks with diffuse optical tomography. *Nature Photonics*, 8(6):448–454, 2014.

[5] Roy A Stillwell and Thomas D O'Sullivan. A real-time fully handheld frequency domain near infrared spectroscopy imaging system. In *Multiscale Imaging and Spectroscopy III*, page PC119440D. SPIE, 2022.

[6] Robin Dale, Biao Zheng, Felipe Orihuela-Espina, Nicholas Ross, Thomas D O'Sullivan, Scott Howard, and Hamid Dehghani. Deep learning-enabled high-speed, multi-parameter diffuse optical tomography. *Journal of Biomedical Optics*, 29(7):076004, 2024.

[7] Hamid Dehghani, Matthew E Eames, Phaneendra K Yalavarthy, Sean 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.

[8] Robin Dale, Nicholas Ross, Scott Howard, Thomas D O'Sullivan, and Hamid Dehghani. Transformer-encoder for real-time dot scanning. In *European Conference on Biomedical Optics (ECBO)*, 2025.

[9] David A. Boas, Daniel H. Brooks, Eric L. Miller, Charles A. DiMarzio, Misha Kilmer, Robert J. Gaudette, and Quan Zhang. Imaging the body with diffuse optical tomography. *IEEE Signal Processing Magazine*, 18(6):57–75, 2001.

[10] Simon R. Arridge and John C. Schotland. Optical tomography in medical imaging: theory, models, and applications. *Inverse Problems*, 25(12):123010, 2009.

[11] Tanja Tarvainen, Marko Vauhkonen, Simon R. Arridge, and Jari P. Kaipio. Bayesian image reconstruction in diffuse optical tomography. *IEEE Transactions on Medical Imaging*, 29(6):1028–1041, 2010.

[12] Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.

[13] Xu Feng, Wei Chen, Long Wei, and Fei Gao. Deep learning-based image reconstruction for diffuse optical tomography. *Biomedical Optics Express*, 11(11):6366–6381, 2020.

APPENDIX A

Implementation and Hyperparameters

APPENDIX B

Extended Dataset Examples and Probe Layouts

APPENDIX C

Additional Quantitative Results

APPENDIX D

Mathematical Derivations

APPENDIX E

Reproducibility Checklist and Ethics Statement