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

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 in oncology and neuroscience. Unlike ionising modalities such as CT or PET, DOT is safe for repeated use 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. Breast cancer is the most common cancer in women, where early detection and treatment monitoring are critical for outcomes. DOT provides functional information such as tumour oxygenation and haemoglobin concentration, not readily accessible with conventional imaging. 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, low-cost alternative to fMRI, valuable at the bedside or in paediatrics [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 shift the bottleneck to computation: reconstruction must be rapid and robust to probe variability, anatomy, and 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. This range maximises haemoglobin contrast while minimising water and lipid absorption, enabling centimetre-scale penetration. The key optical parameters are the absorption coefficient μ_a (mm^{-1}), reflecting chromophore concentration, and the reduced scattering coefficient μ_s' (mm^{-1}), influenced by tissue microstructure. Together, these parameters govern photon fluence and are the quantities to be reconstructed [1, 2].

Measurements use arrays of sources and detectors on the tissue surface. Each source–detector (SD) pair samples a diffuse photon path, with separation (SDS) strongly influencing depth sensitivity. In frequency-domain DOT (FD-DOT), the light is sinusoidally modulated at radiofrequency (e.g. 140 MHz); the detected signal is described by amplitude attenuation and phase shift relative to the input. Using $\log A$ (log-amplitude) together with ϕ (phase) provides complementary sensitivity to μ_a and μ_s'. Short SDS (< 15 mm) probe superficial layers, whereas larger separations (30–40 mm) reach deeper tissue at reduced SNR.

The forward model follows from the frequency-domain diffusion equation and is solved for realistic geometries using the finite element method (FEM) [7]. The operator $\mathcal{F}$ maps spatial fields of μ_a and μ_s' to boundary measurements. Recovering these parameters from sparse, surface-only data is underdetermined and therefore requires regularisation—via smoothness constraints, sparsity-promoting penalties, or data-driven priors learned by neural networks [1].

Terminology. Hereafter, *FD-DOT* denotes frequency-domain measurements of amplitude and phase; *DOT* the modality in general; *DL-DOT* deep learning-based reconstruction.

1.3 Problem Formulation and Notation

This dissertation focuses on frequency-domain diffuse optical tomography (FD-DOT), where 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 use 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:

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

integrating optical and spatial context. For a scan comprising N SD pairs, the full measurement tensor is:

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

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 denotes additive noise, 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 FD-DOT sensitivity to amplitude and phase fluctuations while remaining tractable.

The task is to estimate voxelwise distributions of μ_a and μ_s' on 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 more than deeper ones. This imbalance renders the problem intrinsically unstable in the absence of strong priors. The central challenge is designing models and training strategies that embed spatial priors, handle probe geometry variability, and generalise across diverse phantoms while maintaining fidelity in both μ_a and μ_s'.

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 overlapping sensitivity profiles. Deep tissue contributes weak signals, amplifying 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 variable. 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 when layouts change [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 simplification reflecting typical variability. Even small perturbations can destabilise reconstructions unless considered during training, underscoring the need for noise-aware 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 for reliable DL-DOT deployment.
- **Latency:** Real-time use requires reconstructions in under 0.1 s per volume. Iterative solvers are too slow, 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.

CHAPTER 2

Literature Review

2.1 Conventional Foundations in FD-DOT

This section establishes the minimum physics and algorithmic background required for the learning-based approach pursued later. It traces the standard modelling pipeline from radiative transport to the diffusion approximation, outlines frequency-domain forward modelling with the finite element method (FEM), and summarises classical inverse formulations and regularisation strategies that motivate data-driven alternatives.

2.1.1 From radiative transport to diffusion

At its most fundamental level, photon propagation in tissue is governed by the radiative transport equation (RTE), which evolves radiance over space and angle. The RTE is accurate but high-dimensional and computationally demanding, rendering direct inversion impractical for typical DOT settings [1]. In highly scattering media—typical of the NIR window (650–900 nm)—the angular distribution rapidly approaches isotropy; under this regime, the RTE admits the diffusion approximation, which describes photon fluence via a parabolic partial differential equation [1, 2, 9]. The diffusion model provides a tractable surrogate that is sufficiently accurate away from boundaries or low-scattering regions; its known limitations (e.g., near refractive index mismatches and superficial layers) are well documented [10]. For the frequency-domain case considered here, sinusoidal modulation leads to a complex-valued diffusion equation whose solutions yield amplitude attenuation and phase shift at the tissue boundary—precisely the observables used in FD-DOT.

2.1.2 Frequency-domain forward modelling (FEM in practice)

Realistic head, breast, and organ geometries require numerical solvers. FEM has become the de facto approach in DOT due to its ability to accommodate complex boundaries and heterogeneous optical properties [7]. The tissue domain is discretised into elements; material parameters are piecewise defined, and the weak form of the (complex) diffusion equation is assembled into large sparse linear systems. The resulting forward operator $\mathcal{F}$ maps spatial fields of absorption μ_a (mm^{-1}) and reduced scattering μ_s' (mm^{-1}) to boundary measurements of $\log A$ (log-amplitude) and ϕ (phase). Alternative schemes—finite differences, boundary elements, Monte Carlo accelerations—exist, but FEM remains dominant given its mesh flexibility, ease of incorporating boundary conditions, and direct compatibility with anatomical priors (e.g., from MRI/CT) and frequency-domain extensions [7]. The computational burden is substantial: a single forward solve entails factorising or iteratively solving very large systems, and sensitivity (Jacobian) computations compound costs when performed repeatedly in inversion. These factors underpin the latency observed in conventional FD-DOT pipelines and motivate learned surrogates or end-to-end predictors.

2.1.3 Classical inverse formulations and regularisation

DOT reconstruction seeks spatial maps of (μ_a, μ_s') from sparse, surface-only data. A standard variational formulation is

$$\hat{\mu} \;=\; \arg\min_{\mu} \; \left\| \mathbf{y} - \mathcal{F}(\mu) \right\|_2^2 \;+\; \lambda\, R(\mu),$$

where $\mathbf{y}$ denotes measured $\log A$ and ϕ, $R(\mu)$ encodes prior structure, and $\lambda > 0$ balances data fidelity and regularity. The prevalent solvers linearise the forward map about the current estimate and iterate with Gauss–Newton or Levenberg–Marquardt updates, using FEM-derived Jacobians to compute sensitivities [1]. While effective, these methods incur minutes-per-volume runtimes in 3D and can be brittle in low signal-to-noise regimes.

Regularisation is therefore central. Tikhonov (quadratic) priors suppress noise but blur edges; sparsity and total variation (TV) constraints better preserve localised inclusions at the expense of non-smooth optimisation and parameter tuning [2]. Bayesian formulations (e.g., MAP estimation and approximation-error modelling) offer principled uncertainty handling and hierarchical priors but further increase computational cost [11]. When available, anatomical priors from MRI/CT constrain the solution space and sharpen localisation, but they tie DOT to external imaging and do not remove the need for repeated Jacobian evaluations [1]. Across these variants, three limitations recur: (i) heavy reliance on hand-crafted priors and manual hyperparameter tuning, (ii) sensitivity to model–data mismatch (e.g., boundary conditions, optical heterogeneity), and (iii) prohibitive latency stemming from large-scale PDE solves and sensitivity computations.

Takeaway. The diffusion approximation and FEM provide a physically grounded, widely validated forward model for FD-DOT, and classical inverse solvers remain a valuable reference. However, their computational cost and dependence on hand-crafted priors motivate learning-based approaches that can embed data-driven regularity and achieve near real-time inference once trained. The next section therefore reviews learning-based reconstruction for FD-DOT and related modalities, with emphasis on architectures and representations that address geometry variability, noise, and domain shift.

2.2 Learning-Based Reconstruction for FD-DOT

Deep learning (DL) offers two capabilities that directly address the ill-posedness and latency of FD-DOT: the capacity to learn strong, data-driven priors from examples, and the ability to amortise computation so that inference is near real time once training is complete. Inverse problems in other modalities have shown that learned priors can stabilise reconstructions under noise and sparsity, either by unrolling iterative solvers into trainable networks or by learning direct sensor-to-image mappings; this section focuses on how these ideas translate to FD-DOT.

2.2.1 Why learning helps FD-DOT

Across CT, MRI, and photoacoustics, DL has improved reconstructions via two patterns: *(i) unrolling* classical optimisation into networks whose layers mirror iterations (imparting interpretability and data-consistency), and *(ii) direct inversion* from raw or minimally processed measurements to images. Reviews by Monga *et al.* and Arridge *et al.* synthesise the mathematical foundations of learned inverse problems, including the role of stability, data consistency, and the benefits/risks of data-driven priors [12, 13]. Unrolled methods such as the Learned Primal–Dual algorithm of Adler and Öktem exemplify how physics and learning can be combined to handle non-linear forward operators while remaining task-specific and efficient at test time [14]. These precedents justify a learning-based approach for FD-DOT, where the forward model is expensive and the measurements are sparse and noisy.

2.2.2 DOT-specific DL: what has worked, and what has not

Early DOT studies adopted convolutional encoder–decoders (e.g., U-Nets) to map boundary measurements to absorption (and sometimes scattering) images on simulated or phantom datasets, reporting sharper inclusions and large speed-ups over Tikhonov-regularised FEM [15]. More recent work has scaled to 3D and demonstrated feasibility in vivo: Deng *et al.* presented a three-module network ("FDU-net") trained on simulations and validated on human breast data, showing improved anomaly localisation while maintaining sub-second

inference [16]. End-to-end frameworks for diffuse optical imaging have also appeared (e.g., Periodic-net), indicating the generality of learned reconstructions for optical boundary data [17]. Despite these advances, limitations persist: most published models assume fixed probe geometries, modest datasets, and restricted anatomical variability, leading to degradation under geometry shift and to fragility when moving from simulation to experiment. These gaps motivate architectures that encode geometry explicitly and training regimes that expose models to broader phantom and probe distributions.

2.2.3 CNN encoders/decoders for volumetric DOT

CNNs remain attractive for DOT because they learn spatially local, translation-equivariant priors over volumetric images. In 3D FD-DOT, encoder–decoder topologies (including U-Nets and autoencoders) provide strong inductive bias for recovering compact inclusions and smooth backgrounds. Practical trade-offs arise between resolution and latency (3D convolutions are costly) and between single-parameter networks and multi-head designs that jointly estimate (μ_a, μ_s'). A useful pattern—adopted in this dissertation—is to learn a *spatial latent* with a 3D CNN (via autoencoding ground-truth volumes), then condition or regress that latent from measurements. This separates the tasks of (i) capturing anatomical/structural priors and (ii) fusing measurement information, and it aligns with evidence that learned spatial priors stabilise DOT reconstructions while reducing sensitivity to noise in boundary data.

2.2.4 Attention and transformers for measurement sequences

FD-DOT measurements naturally form sets or sequences of source–detector (SD) pairs, each described by $(\log A, \phi)$ and explicit coordinates of source and detector positions. Self-attention is well suited to aggregate such tokens: it models long-range dependencies across SD pairs, is permutation-flexible, and can integrate explicit spatial embeddings to encode geometry. Positional or geometry-aware embeddings (e.g., from SD coordinates and separations) allow the network to remain path-agnostic while still being *geometry-aware*, addressing the common failure mode under layout changes. Attention pooling further provides global context, which is important when deeper structures only influence long-separation measurements weakly. A constraint for transformers is data hunger; this can be mitigated by strong spatial priors (CNN latent spaces) and aggressive geometry randomisation during training.

2.2.5 Hybrid two-stage designs for FD-DOT

Hybrid CNN–Transformer frameworks combine the strengths of spatial priors and sequence aggregation. Dale *et al.* recently demonstrated high-speed, multi-parameter FD-DOT

using a hybrid design that processes measurement sequences with spatial awareness and reconstructs 3D absorption and reduced scattering at sub-second rates [6]. In this two-stage paradigm (also adopted here), Stage 1 learns a volumetric latent representation from ground-truth (μ_a, μ_s'); Stage 2 consumes SD tokens built from $(\log A, \phi)$ and explicit coordinates, using attention to integrate global information and regress the latent (and hence the volume). Variants include (i) multi-query attention to stabilise global aggregation, (ii) explicit coordinate encodings (SDS, 3D positions, relative vectors), and (iii) selective fine-tuning of the CNN decoder to align the latent space with measurement-conditioned features. The principal benefits are robustness (priors learned from anatomy-like variability) and latency (amortised inference), while typical failure modes—data shift and geometry shift—are addressed by dataset design (phantom/probe diversity, SO(3) rotations) and by geometry-aware tokenisation. This dissertation builds on that direction with systematic geometry randomisation (fixed 256-token sequences), enriched spatial embeddings, and latent alignment strategies aimed at improving generalisation without sacrificing fidelity in both parameters.

2.3 Robustness, Baselines, and the Research Gap

This section consolidates the failure modes that impede deployment of learning-based FD-DOT, reviews a recent path-agnostic baseline, and states the problem this dissertation addresses together with its contributions.

2.3.1 Deployment failure modes

Geometry shift. Handheld and wearable probes rarely follow fixed layouts; source–detector (SD) separations vary with operator handling and patient anatomy. Models trained on a single geometry often degrade when SD patterns change because the mapping from $(\log A, \phi)$ to volume depends on the sampling paths through tissue. Classical FEM-based solvers accommodate arbitrary layouts by recomputing Jacobians [1, 7], but many DL-DOT studies assume fixed arrays or limited variability [15, 16]. This motivates *explicit* geometry handling in learned models and *systematic* geometry randomisation during training.

Noise robustness. FD-DOT measurements are sensitive to coupling, electronic noise, and instrumental drift; small phase errors or amplitude fluctuations can destabilise reconstructions if the learned mapping implicitly overfits to noise patterns. Classical formulations manage this with regularisation and uncertainty modelling [1, 11], whereas learned methods must incorporate noise-aware training and invariances to remain reliable across sessions and systems.

Simulation-to-real gap. Synthetic phantoms enable scale but under-represent anatomical heterogeneity, motion artefacts, and hardware imperfections. As a result, models trained

on narrow synthetic distributions can fail to transfer to experimental or in vivo data [2, 10]. Narrowing this gap requires richer phantom distributions, physics-respecting augmentation, and architectures that remain stable under moderate forward-model mismatch.

2.3.2 Baseline: path-agnostic DL-DOT

Dale *et al.* proposed a hybrid CNN–Transformer framework that processes SD measurements as tokens augmented with explicit spatial information (e.g., source and detector coordinates), enabling the network to integrate arbitrary scanning pathways while reconstructing 3D absorption and reduced scattering at high speed [6]. The approach combines a volumetric prior (learned by a CNN) with a transformer encoder that aggregates measurements via self-attention, achieving multi-parameter reconstructions with sub-second inference and reporting robustness to path variation on held-out layouts [6, 8, 18]. This baseline establishes two key principles for practical DL-DOT: (i) *geometry-aware tokenisation* to mitigate layout dependence, and (ii) *amortised inference* to meet latency constraints. Remaining challenges are shaped by the training distribution: generalisation can still be limited by the diversity of phantoms, measurement noise models, and the extent of geometry randomisation seen during training.

2.3.3 Problem statement and contributions

Building on the baseline principles above, this dissertation targets *geometry- and anatomy-generalised* FD-DOT reconstruction under realistic noise. The central problem is to learn a mapping from measurement sequences to volumetric optical properties that maintains fidelity in both μ_a and μ_s' across widely varying probe layouts and tissue shapes. The specific contributions are:

1. **Diverse phantom and probe generation.** A high-throughput pipeline produces ellipsoidal tissue volumes embedded in air, inducing realistic tissue–air boundaries for surface-aware SD placement. Phantoms undergo random SO(3) rotations to remove spatial bias, and tumour inclusions vary in size and shape to broaden anatomical priors.
2. **Systematic geometry randomisation.** Each phantom yields a large set of SD measurements that are dynamically subsampled into fixed 256-token sequences, enforcing invariance to probe placement and providing strong augmentation against geometry shift.
3. **Hybrid CNN–Transformer with spatially aware embeddings.** Stage 1 learns a 3D spatial latent from ground-truth volumes; Stage 2 aggregates SD tokens built from $(\log A, \phi)$ and explicit coordinates. Multi-query attention stabilises global aggregation, and selective decoder fine-tuning (latent alignment) aligns the spatial prior with measurement-conditioned features, following the hybrid design ethos but developed

independently of specific architectural details [6, 8, 18].

Hypothesis: Combining phantom/probe diversity, systematic geometry randomisation, and a hybrid geometry-aware architecture will improve robustness and generalisation across probe layouts and anatomical variability, while preserving fidelity in absorption and scattering reconstructions. By embedding stronger spatial priors and training under controlled variability, the model is expected to deliver reconstructions that are both accurate and efficient, achieving performance competitive with the best reported DL-DOT systems [6].

2.3.4 From gap to methodology

Chapter 3 introduces the physics-based data pipeline and geometry randomisation protocol, including phantom construction, optical property assignment, probe placement, and noise modelling. Chapter 4 then presents the hybrid CNN–Transformer architecture, followed in Chapter 5 by the training strategy that operationalises it. Chapter 6 reports the experimental results and analysis, with Chapter 7 offering critical discussion and Chapter 8 concluding with reflections and future directions. Together, these chapters translate the research gap into a concrete methodology and evaluation pathway.

CHAPTER 3

Physics-Based Synthetic Data Pipeline

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 $\{\mathbf{x}_{\text{src}}, \mathbf{x}_{\text{det}}\}$

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)

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)

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*, volume 11944 of *Proc. SPIE*, page 119440D. 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, Ville Kolehmainen, Jari P. Kaipio, and Simon R. Arridge. Corrections to linear methods for diffuse optical tomography using approximation error modelling. *Biomedical Optics Express*, 1(1):209–222, 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] Simon R. Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.

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

[15] 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.

[16] Bin Deng, Hanxue Gu, Hongmin Zhu, Ken Chang, Katharina V. Hoebel, Jay B. Patel, Jayashree Kalpathy-Cramer, and Stefan A. Carp. Fdu-net: Deep learning-based three-dimensional diffuse optical image reconstruction. *IEEE Transactions on Medical Imaging*, 42(8):2439–2450, 2023.

[17] Nazish Murad, Min-Chun Pan, and Ya-Fen Hsu. Periodic-net: an end-to-end data driven framework for diffuse optical imaging of breast cancer from noisy boundary data. *Journal of Biomedical Optics*, 28(2):026001, 2023.

[18] Robin Dale. *Deep Learning for Flexible and Real-Time Diffuse Optical Tomography*. PhD thesis, University of Birmingham, 2025.

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