End-to-End NIR-DOT Reconstruction with Hybrid CNN-Transformer

MSc Dissertation

Max Andrew Hart

School of Computer Science

College of Engineering and Physical Sciences

University of Birmingham

2024-25

Abstract

This dissertation presents a novel hybrid CNN-Transformer architecture for enhanced Near-Infrared Diffuse Optical Tomography (NIR-DOT) reconstruction. NIR-DOT is a non-invasive imaging technique with significant potential for brain imaging and cancer detection, but faces challenges in reconstruction accuracy due to the ill-posed nature of the inverse problem.

We introduce a two-stage approach combining the spatial feature learning capabilities of CNNs with the contextual processing power of Transformers, further enhanced by the integration of local tissue context information. Using a comprehensive dataset of 5,000 synthetic phantoms, we demonstrate that our approach significantly improves reconstruction quality over traditional methods and standard deep learning approaches.

Our key innovation—tissue context integration—provides physiologically relevant priors that guide the reconstruction process. Experimental results show improvements in RMSE, SSIM, and PSNR metrics, with particularly notable enhancements in tumor region reconstruction. The proposed method represents a significant advance in NIR-DOT reconstruction technology with potential clinical applications in brain imaging and breast cancer detection.

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 uses near-infrared (NIR) light to probe tissue optical properties, particularly absorption (μ_a) and reduced scattering coefficients (μ_s'). By measuring how NIR photons migrate through tissue, DOT can provide valuable structural and functional information about biological tissue, including haemodynamic changes and tumour localisation [**?**, **?**]. In recent years, Near-Infrared Diffuse Optical Tomography (NIR-DOT) has gained prominence in biomedical imaging due to its non-ionising nature, low cost, and ability to provide unique functional contrast compared to conventional modalities such as MRI or CT.

However, the potential of NIR-DOT has not yet been fully realised in clinical applications. One of the central challenges is the reconstruction of three-dimensional tissue optical property maps from boundary NIR measurements. This inverse problem is severely ill-posed: many different internal tissue configurations can give rise to similar surface measurements. The result is that reconstructions are often noisy, blurred, and highly sensitive to measurement error or modelling assumptions. Overcoming these challenges requires methodological innovation in both data-driven learning and physics-aware modelling.

Recent advances in deep learning, particularly convolutional neural networks (CNNs) and transformers, have revolutionised inverse imaging tasks across modalities. Inspired by these developments, our research leverages a hybrid CNN–transformer architecture for NIR-DOT reconstruction, building upon and extending the work of Robin Dale and colleagues [**?**]. Dale's work demonstrated the feasibility of combining convolutional feature extraction with transformer-based long-range dependency modelling to improve optical imaging reconstruction. Our research pushes this line further by developing a two-stage, student–teacher training paradigm that integrates synthetic phantom generation, robust standardisation, and physics-aware evaluation to achieve state-of-the-art reconstructions in raw optical units.

1.2 Near-Infrared Diffuse Optical Tomography

NIR-DOT operates on the principle of photon migration in highly scattering media. When NIR light in the range of 650–950 nm is introduced into tissue, photons undergo multiple scattering events and absorption before being detected at surface detectors. By modelling this transport, typically with the diffusion approximation to the radiative transfer equation, it is possible to infer the internal distribution of absorption and scattering coefficients.

The forward problem in NIR-DOT consists of solving the diffusion equation given known tissue optical properties and boundary conditions, producing boundary fluxes or detector measurements. The inverse problem, however, is the reconstruction of tissue properties from these measurements. This inversion is complicated by the exponential attenuation of light, the limited number of source–detector pairs, and the inherent ambiguity of mapping low-dimensional measurements back to high-dimensional volumes. Consequently, the problem is both underdetermined and ill-conditioned.

Traditional reconstruction approaches rely on iterative optimisation with regularisation, often requiring prior assumptions about smoothness or sparsity. These methods are computationally expensive and can yield unstable solutions. Data-driven methods, by contrast, can learn mappings directly from simulated or empirical data, enabling faster and more robust reconstructions. In particular, CNNs are adept at learning spatial hierarchies from volumetric data, while transformers excel at capturing long-range dependencies across measurements. Combining these paradigms offers a powerful new direction for NIR-DOT.

1.3 Challenges in NIR-DOT Reconstruction

The central challenge in NIR-DOT is the ill-posed nature of the inverse problem. Specifically:

- **Non-uniqueness:** Different tissue structures can yield nearly identical boundary measurements.
- **Instability:** Small perturbations in measurements can lead to disproportionately large errors in reconstructions.
- **High dimensionality:** Reconstructing 64^3 voxel volumes from relatively few measurements (typically 1000 per phantom) represents an extreme underdetermined problem.
- **Noise sensitivity:** Instrumental noise, modelling errors, and physiological variability further degrade reconstruction accuracy.

From a computational standpoint, another challenge lies in balancing fidelity to physical principles with the flexibility of data-driven learning. Purely physics-based models are slow and fragile, while purely data-driven models risk producing reconstructions that violate known physics. Hybrid approaches must therefore enforce physical consistency while still benefiting from the representational power of deep networks.

1.4 Research Objectives and Contributions

The goal of this dissertation is to design, implement, and evaluate a hybrid CNN–transformer pipeline for NIR-DOT reconstruction that addresses the above challenges through a principled two-stage training framework. The specific objectives are:

1. **Data Simulation:** Develop a high-fidelity synthetic phantom generation and forward-modelling pipeline to create realistic training datasets with controllable optical properties.
2. **Stage 1 Training:** Pre-train a 3D CNN autoencoder on ground truth volumes to learn robust spatial representations of tissue optical properties.
3. **Stage 2 Training:** Introduce a transformer encoder to process NIR measurements, integrating spatial embeddings and sequence modelling to produce latent codes compatible with the Stage 1 decoder.
4. **Student–Teacher Learning:** Employ a teacher–student paradigm where the frozen CNN encoder provides latent targets to guide the transformer-based student network.
5. **Physics-Aware Evaluation:** Evaluate reconstructions in raw physical units (mm^{-1}) with metrics such as Dice coefficient, RMSE, and contrast ratio, ensuring interpretability and clinical relevance.
6. **End-to-End System Integration:** Build a modular, extensible codebase with clean separation of data generation, training, and evaluation, enabling reproducibility and future extensions.

The main contributions of this work are as follows:

- A novel hybrid CNN–transformer architecture for NIR-DOT reconstruction.
- A two-stage training pipeline combining autoencoder pre-training with transformer-based measurement encoding.
- Integration of synthetic phantom simulation with physics-consistent data preprocessing and standardisation.
- A teacher–student learning paradigm for latent-space alignment across stages.
- A comprehensive evaluation framework in raw optical units, ensuring clinically meaningful assessment of reconstructions.

1.5 Dissertation Structure

The remainder of this dissertation is organised as follows:

- **Chapter 2: Literature Review** surveys prior work in NIR-DOT reconstruction, deep learning for inverse problems, and hybrid architectures combining CNNs and transformers.

- **Chapter 3: Methods** details the data simulation pipeline, model architectures, training strategies, and evaluation metrics used in this research.
- **Chapter 4: Experiments** presents experimental design, training protocols, and validation strategies.
- **Chapter 5: Results** reports quantitative and qualitative results for both Stage 1 and Stage 2 training, including ablation studies and comparison with baseline methods.
- **Chapter 6: Discussion** interprets the findings, analyses limitations, and explores implications for clinical translation.
- **Chapter 7: Conclusion** summarises the contributions, highlights potential future research directions, and reflects on the broader impact of this work.

CHAPTER 2

Literature Review

CHAPTER 3

Synthetic Phantom Data Generation

CHAPTER 4

Hybrid CNN-Transformer Architecture

4.1 Architectural Overview

4.2 CNN Autoencoder Design

4.2.1 Encoder Architecture

4.2.2 Decoder Architecture

4.2.3 Residual Connections and Feature Extraction

4.3 NIR Measurement Processing

4.3.1 Spatial Awareness in Measurement Processing

4.3.2 Tissue Patch Extraction and Encoding

4.4 Transformer Encoder Design

4.4.1 Self-Attention Mechanism

4.4.2 Token-Type Embeddings

4.4.3 Feature Enhancement

4.5 Integration Strategy

4.5.1 Information Flow Between Components

4.5.2 Tissue Context Integration

CHAPTER 5

Two-Stage Training Methodology

5.1 Training Strategy Overview

5.2 Stage 1: CNN Autoencoder Pre-Training

5.2.1 Identity Mapping Objective

5.2.2 AdamW with OneCycleLR Optimization

5.2.3 Hyperparameter Selection

5.3 Stage 2: Transformer Enhancement Training

5.3.1 Frozen Decoder Approach

5.3.2 Linear Warmup with Cosine Decay

5.3.3 Differential Weight Decay

5.4 Data Augmentation Strategy

5.5 Implementation Details

5.5.1 Hardware Optimization

5.5.2 Experiment Tracking

CHAPTER 6

Experimental Results and Analysis

6.1 Experimental Setup

6.1.1 Dataset Preparation

6.1.2 Evaluation Metrics

6.1.3 Baseline Methods

6.2 Stage 1 Results: CNN Autoencoder Performance

6.2.1 Reconstruction Quality Metrics

6.2.2 Feature Learning Analysis

6.3 Stage 2 Results: Transformer Enhancement

6.3.1 Baseline Mode Performance

6.3.2 Enhanced Mode with Tissue Context

6.3.3 Comparative Analysis

6.4 Ablation Studies

6.4.1 Impact of Tissue Patch Size

6.4.2 Effect of Transformer Layers and Heads

6.4.3 Sensitivity to Tissue Context Quality

6.5 Visualization and Interpretation

6.5.1 Attention Map Analysis

6.5.2 Reconstruction Quality Visualization

6.5.3 Error Distribution Analysis

CHAPTER 7

Discussion

7.1 Key Findings and Insights

7.2 Clinical Implications

7.3 Limitations of Current Approach

7.4 Computational Efficiency Considerations

CHAPTER 8

Conclusion and Future Work

8.1 Summary of Contributions

8.2 Future Research Directions

8.3 Potential Clinical Applications

Bibliography

APPENDIX A

Implementation Details

APPENDIX B

Additional Experimental Results

APPENDIX C

Mathematical Derivations