Hybrid CNN-Transformer Architecture with Tissue Context Integration for Enhanced NIR-DOT Reconstruction

Master of Science Dissertation

Max 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

Abstract **ii**

Acknowledgements **iii**

Abbreviations **iv**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**
1.1 Background and Motivation 1
1.2 Near-Infrared Diffuse Optical Tomography 1
1.3 Challenges in NIR-DOT Reconstruction 1
1.4 Research Objectives and Contributions 1
1.5 Dissertation Structure 1

2 Literature Review **2**
2.1 Physics of NIR Light Propagation 2
2.2 Classical NIR-DOT Reconstruction Methods 2
2.3 Deep Learning for Medical Image Reconstruction 2
2.4 CNN Approaches in NIR-DOT 2
2.5 Transformers in Medical Imaging 2
2.6 Hybrid Architectures and Multi-Stage Training 2
2.7 Contextual Integration in Medical Imaging 2
2.8 Research Gap and Opportunity 2

3 Synthetic Phantom Data Generation **3**
3.1 Physics-Based Forward Modeling 3
3.2 Geometric Phantom Construction 3
3.3 Optical Property Assignment 3
3.4 Surface Extraction and Probe Placement 3
3.5 Frequency-Domain Measurement Simulation 3

3.6 Dataset Composition and Statistics 3
3.7 Validation of Simulated Data 3

4 Hybrid CNN-Transformer Architecture 4
4.1 Architectural Overview 4
4.2 CNN Autoencoder Design 4
4.2.1 Encoder Architecture 4
4.2.2 Decoder Architecture 4
4.2.3 Residual Connections and Feature Extraction 4
4.3 NIR Measurement Processing 4
4.3.1 Spatial Awareness in Measurement Processing 4
4.3.2 Tissue Patch Extraction and Encoding 4
4.4 Transformer Encoder Design 4
4.4.1 Self-Attention Mechanism 4
4.4.2 Token-Type Embeddings 4
4.4.3 Feature Enhancement 4
4.5 Integration Strategy 4
4.5.1 Information Flow Between Components 4
4.5.2 Tissue Context Integration 4

5 Two-Stage Training Methodology 5
5.1 Training Strategy Overview 5
5.2 Stage 1: CNN Autoencoder Pre-Training 5
5.2.1 Identity Mapping Objective 5
5.2.2 AdamW with OneCycleLR Optimization 5
5.2.3 Hyperparameter Selection 5
5.3 Stage 2: Transformer Enhancement Training 5
5.3.1 Frozen Decoder Approach 5
5.3.2 Linear Warmup with Cosine Decay 5
5.3.3 Differential Weight Decay 5
5.4 Data Augmentation Strategy 5
5.5 Implementation Details 5
5.5.1 Hardware Optimization 5
5.5.2 Experiment Tracking 5

6 Experimental Results and Analysis 6
6.1 Experimental Setup 7
6.1.1 Dataset Preparation 7
6.1.2 Evaluation Metrics 7
6.1.3 Baseline Methods 7
6.2 Stage 1 Results: CNN Autoencoder Performance 7
6.2.1 Reconstruction Quality Metrics 7
6.2.2 Feature Learning Analysis 7
6.3 Stage 2 Results: Transformer Enhancement 7
6.3.1 Baseline Mode Performance 7
6.3.2 Enhanced Mode with Tissue Context 7
6.3.3 Comparative Analysis 7
6.4 Ablation Studies 7
6.4.1 Impact of Tissue Patch Size 7
6.4.2 Effect of Transformer Layers and Heads 7

6.4.3 Sensitivity to Tissue Context Quality 7
6.5 Visualization and Interpretation 7
6.5.1 Attention Map Analysis 7
6.5.2 Reconstruction Quality Visualization 7
6.5.3 Error Distribution Analysis 7

7 Discussion **8**
7.1 Key Findings and Insights 8
7.2 Clinical Implications 8
7.3 Limitations of Current Approach 8
7.4 Computational Efficiency Considerations 8

8 Conclusion and Future Work **9**
8.1 Summary of Contributions 9
8.2 Future Research Directions 9
8.3 Potential Clinical Applications 9

A Implementation Details **11**

B Additional Experimental Results **12**

C Mathematical Derivations **13**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

1.2 Near-Infrared Diffuse Optical Tomography

1.3 Challenges in NIR-DOT Reconstruction

1.4 Research Objectives and Contributions

1.5 Dissertation Structure

CHAPTER 2

LITERATURE REVIEW

2.1 Physics of NIR Light Propagation

2.2 Classical NIR-DOT Reconstruction Methods

2.3 Deep Learning for Medical Image Reconstruction

2.4 CNN Approaches in NIR-DOT

2.5 Transformers in Medical Imaging

2.6 Hybrid Architectures and Multi-Stage Training

2.7 Contextual Integration in Medical Imaging

2.8 Research Gap and Opportunity

CHAPTER 3

SYNTHETIC PHANTOM DATA GENERATION

3.1 Physics-Based Forward Modeling

3.2 Geometric Phantom Construction

3.3 Optical Property Assignment

3.4 Surface Extraction and Probe Placement

3.5 Frequency-Domain Measurement Simulation

3.6 Dataset Composition and Statistics

3.7 Validation of Simulated Data

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