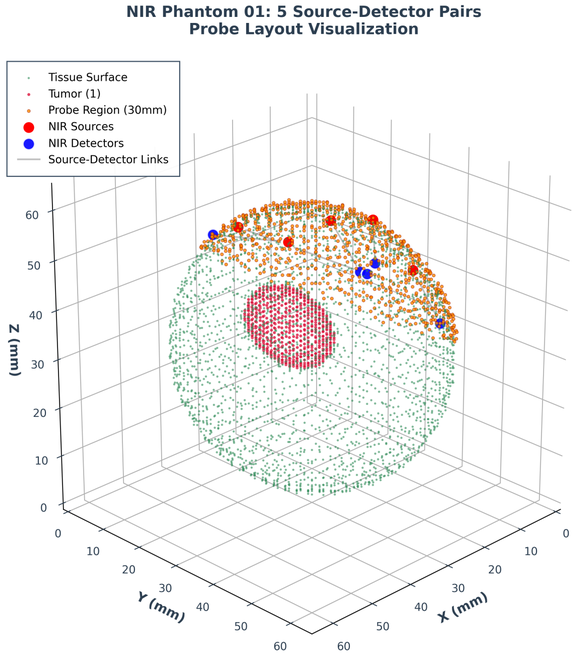
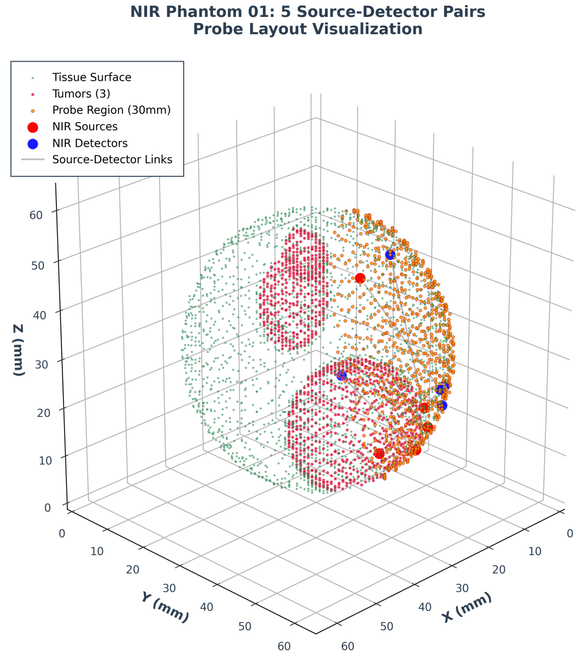
Towards Generalisable Inverse Modelling in NIR-DOT

A Two-Stage CNN–Transformer with Latent Alignment



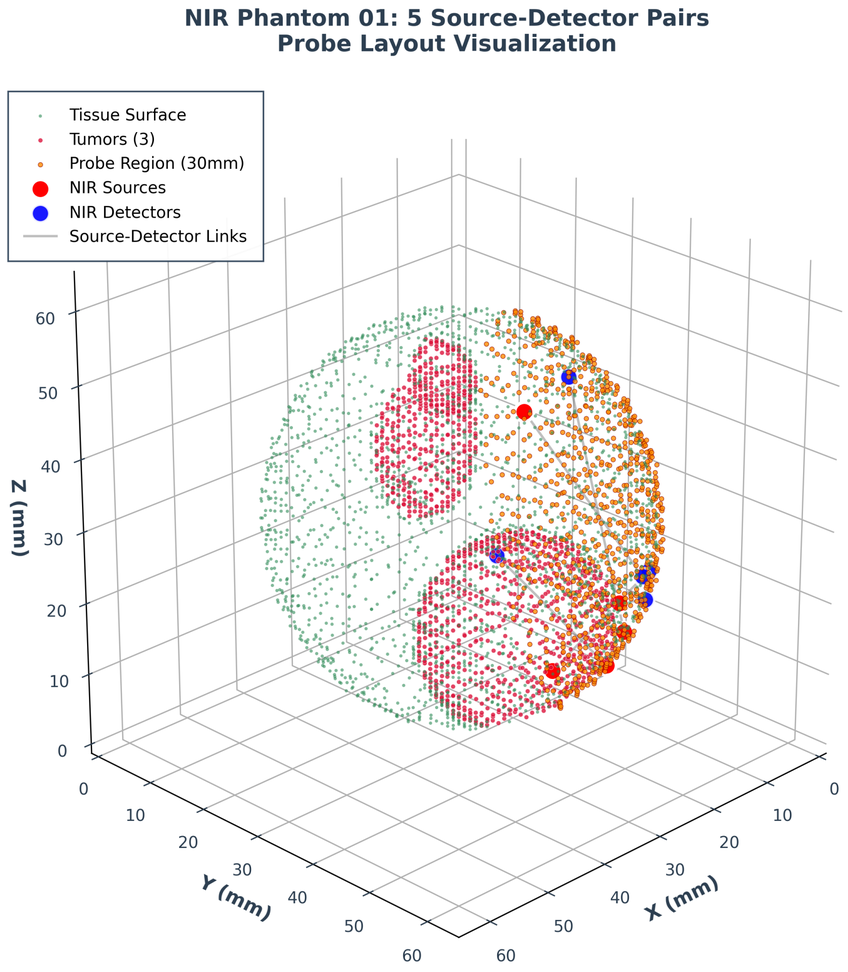
Max Hart



MSc AI/ML Dissertation Presentation

Introduction

- **What is NIR-DOT?**
  - A non-invasive imaging method that shines near-infrared (NIR) light into tissue and measures how it scatters and absorbs.
- **What are probes?**
  - Arrays of sources (light-emitters) and detectors (sensors) placed on the skin surface. They capture measurement signals (log-amplitude & phase).
- **What are phantoms?**
  - Synthetic 3D models of tissue (with tumours or inclusions) used to generate realistic training data. These allow controlled experiments.
- **What data do we use?**
  - Thousands of simulated measurements from phantoms, solved with a physics-based forward model. Each dataset pairs surface probe signals with the ground truth internal tissue maps (absorption & scattering).
- **The goal:**
  - From surface probe data -> reconstruct 3D images of tissue (optical properties), enabling low-cost portable diagnostic imaging.

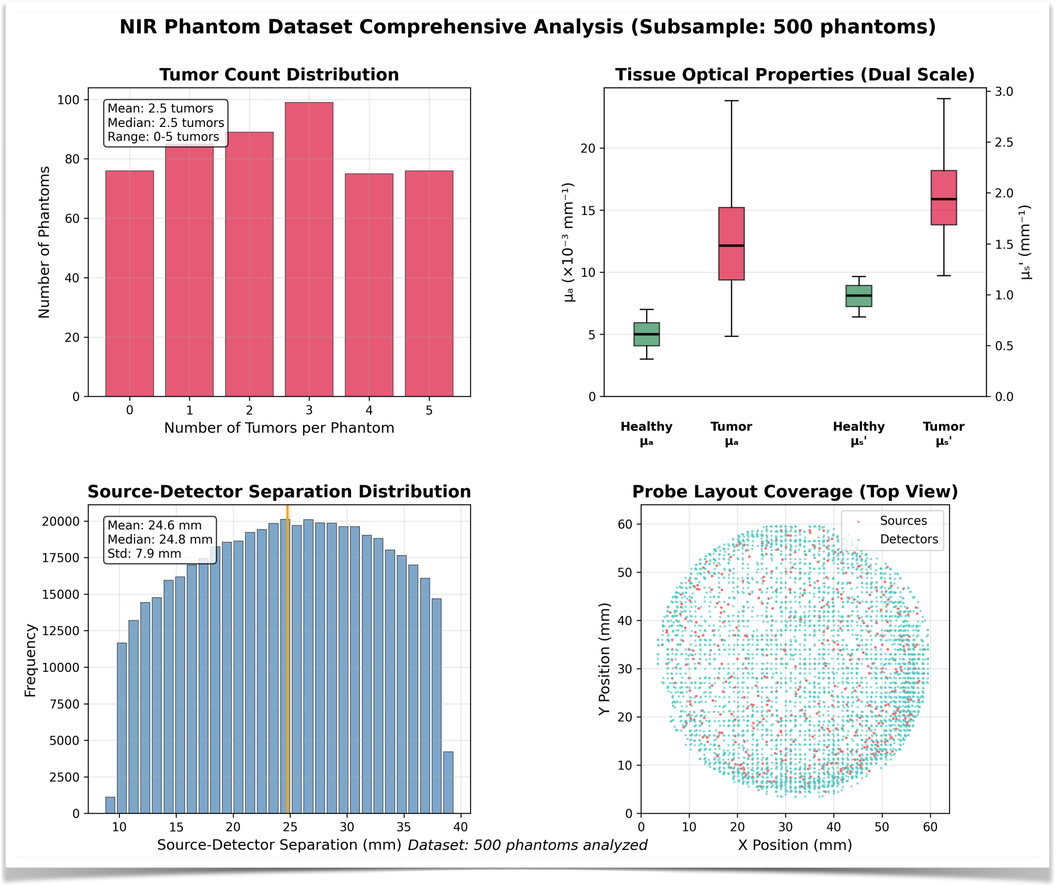


Motivation and Goals

- **Make DOT truly real-time and flexible.** Iterative recon is too slow in clinic; DL-DOT can achieve sub-second latency, but most models are locked to fixed scan paths.
- **Generalise beyond prescribed scanning.** Prior DL-DOT memorises geometry-specific shortcuts; we explicitly encode measurement physics so one model handles arbitrary probe layouts and path lengths.
- **Quantitative recovery of μ_a and μ'_s.** Leverage FD amplitude + phase to separate absorption and reduced scattering in 3D—pushing from relative changes to true quantitative imaging.
- **Train for robustness, not memorisation.** Heavy randomisation of probe positions, rotations, densities, and tissue properties forces the model to learn forward-physics patterns—not dataset quirks.
- **Clinical payoff.** A single, path-agnostic model cuts re-training, adapts to real handheld usage, and moves DL-DOT closer to routine bedside imaging.

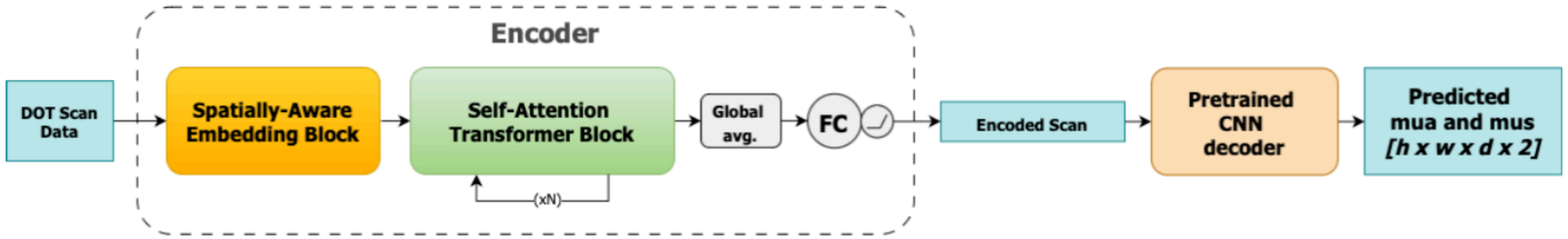
Phantom Generation & Simulation

- **Physics-accurate forward model.** NIRFASTer-FF (140 MHz) with tetrahedral FEM (~1 mm^3 mesh elements) generates FD amplitude + phase for 3D μ_a and μ'_s.
- **Clinically realistic probe geometry.** Surface-aware placement with 50 strategic sources (Poisson-disk) × 20 detectors each → 1000 measurements/phantom; SDS 10–40 mm.
- **Controlled variability.** Randomised tissue/tumour ellipsoids (pose/size), tumour counts, and optical properties encourage learning physics — not memorised layouts.
- **High-throughput generation.** 10,000 phantoms synthesised with full ground-truth (μ_a, μ'_s), yielding ~4.8GB total (HDF5-compressed).
- **Rich measurement tensors.** Each phantom stores 1000 measurements with 8 features per record: [log-amplitude, phase, src_x, src_y, src_z, det_x, det_y, det_z]



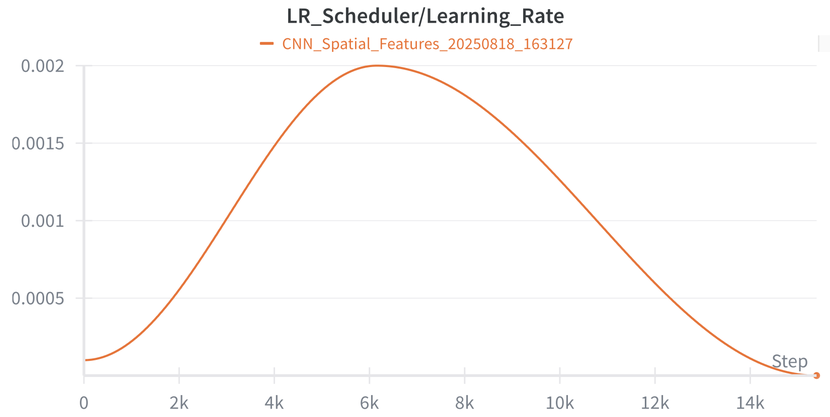
Model Architecture

- **Two-Stage Hybrid Pipeline**
  - Stage 1: 3D CNN autoencoder (16 -> 256 channels, ~7M params) learns spatial features -> 256-D latent bottleneck.
  - Stage 2: NIR measurements (1000 x 8 ->256 x 8) -> spatially aware embedding -> transformer -> global pooling -> frozen CNN decoder
- **Spatially-Aware Embedding**
  - Splits signals [log_amp, phase] and positions [src_xyz, det_xyz], projected via MLPs -> fused into 256-D tokens.
  - Encodes source-detector geometry explicitly, ensuring path-length physics is preserved.
- **Transformer Backbone**
  - 8-layer, 8-head Transformer (~4M params) processes 256 spatially-aware tokens.
  - Multi-query attention pooling mechanism uses 4 learnable queries to extract complementary global representations, fused into a 256-D encoded scan for the CNN.
- **Teacher-Student Latent Alignment**
  - Transformer latent (student, 256-D) is trained to match CNN encoder latent (teacher, 256-D) via RMSE loss, with frozen decoder for end-to-end validation.



End-to-End Training Strategy

- **Stage 1 - CNN Pre-training with OneCycleLR**
  - Pre-train a 3D CNN autoencoder using OneCycleLR “”super-convergence” to rapidly converge on robust volumetric spatial features.
- **Stage 2 - Transformer Fine-Tuning with Cosine Decay**
  - Attach frozen CNN decoder; train transformer encoder with linear warmup -> cosine decay.
  - Differential regularisation stabilises attention layers while leveraging CNN features.
  - Custom AdamW parameter groups: tuned weight decay across embeddings, attention, and normalisation layers for balanced optimisation.
- **Mixed Precision & Optimised Data Pipeline**
  - AMP with bf16, PyTorch compilation, channels-last memory, and prefetch workers maximise training throughput.
  - Trained on NVIDIA A100 (40GB HBM, 200GB RAM).
- **Measurement Subsampling for Stability & Augmentation**
  - 1000 simulated measurements per phantom -> subsampled to 256 per scan.
  - Provides fixed sequence length and ~4x augmentation without data leakage.



CNN Learning Schedular

Transformer Learning Schedular

Evaluation Framework

- **Multi-Channel Evaluation with Raw Unit Conversion**
  - RMSE, Dice, and Contrast Ratio computed separately for absorption (μ_a) and scattering (μ'_s).
  - Inputs (optical properties, source-detector positions, measurements) normalised; metrics reported post-standardisation
- **Latent RMSE Training Objective + End-to-End Validation**
  - Stage 2 optimised on latent-space RMSE (256D teacher-student alignment).
  - Progress monitored with full reconstructions through frozen CNN decoder each epoch.
- **Deterministic Validation Protocol**
  - Fixed validation seeds ensure consistent phantom sets across experiments.
  - Eliminates metric variance from random sampling -> clean comparison of architectures/hyperparameters.
  - All reconstructions shown are from a 10% holdout test set (80/10/10 split used).
- **Physics-Aware constraints & Clinical Ranges**
  - Enforced contrast limits ($\mu_a \leq 0.0245$, $\mu'_s \leq 2.95$ mm^{-1}) to maintain physiological realism.
  - 3-view slice grids assess reconstructions quality and artefact suppression.
- **Comprehensive Logging & Early Stopping**
  - W&B captures learning curves, attention maps, and 24-image reconstructions.
  - Best checkpoint selected by validation RMSE with 25-epoch patience.

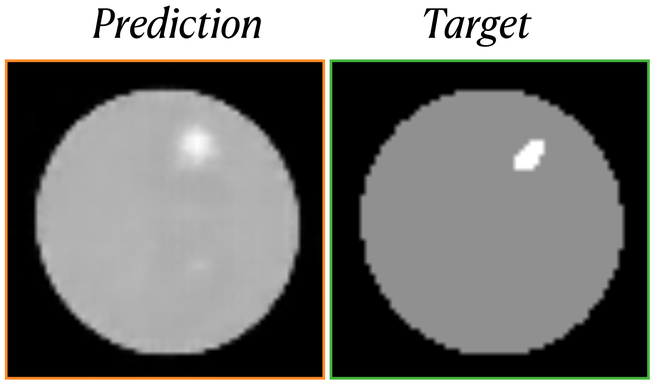
$$RMSE(\hat{y}, y, \mu) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i^{(\mu)} - y_i^{(\mu)})^2}$$

$$SDC(\hat{M}, M) = \frac{2|\hat{M} \bigcap M|}{|\hat{M}| + |M|}$$

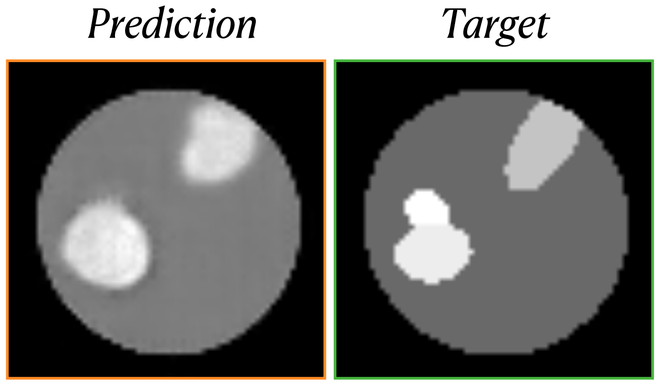
$$CR(\hat{y}, y, \mu) = \frac{\left\langle \hat{y}_M^{(\mu)} \right\rangle}{\left\langle \hat{y}_{\neg M}^{(\mu)} \right\rangle} \Big/ \frac{\left\langle y_M^{(\mu)} \right\rangle}{\left\langle y_{\neg M}^{(\mu)} \right\rangle}$$

Results Stage 1 — (CNN Autoencoder)

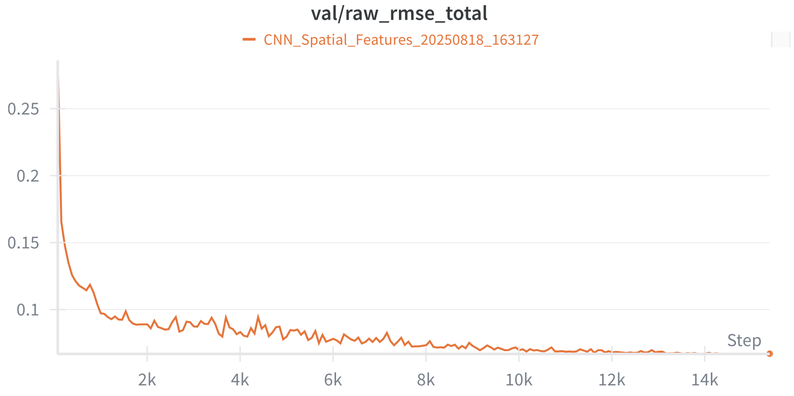
- Strong Convergence with OnceCycleLR
  - Best **Raw RMSE: 0.0669** (μ_a: 0.0007, μ'_s:0.0947)
  - **Dice = 0.727, Contrast = 0.883**
- Stable Training: No overfitting, smooth loss curves.
- Clear Reconstructions:
  - Sharp inclusions in μ_a channel.
  - Blurred but consistent structures in μ'_s.

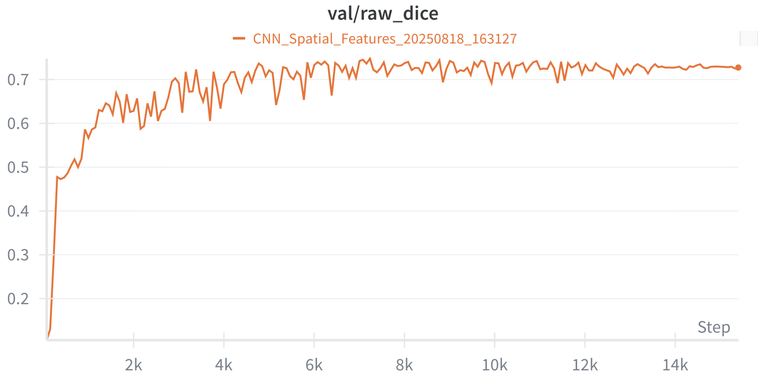


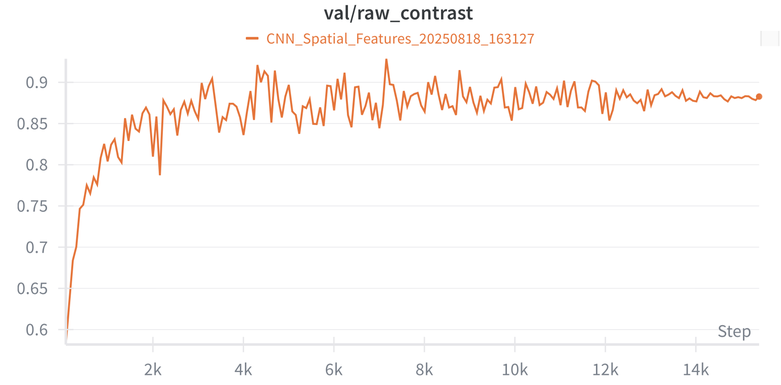
μ_a



μ'_s

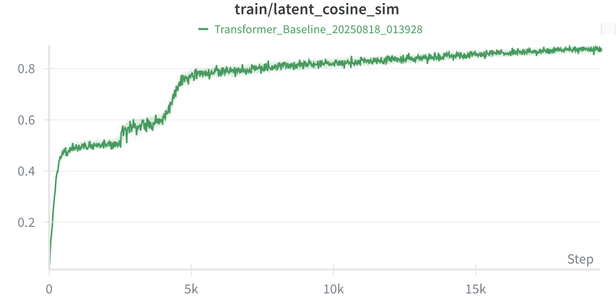


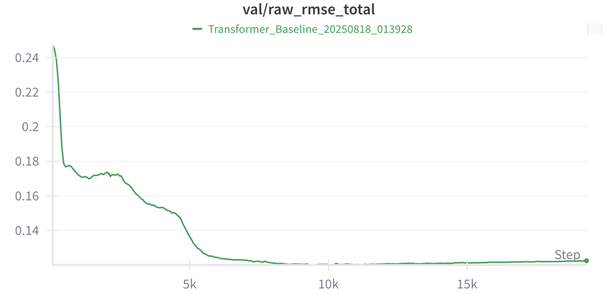


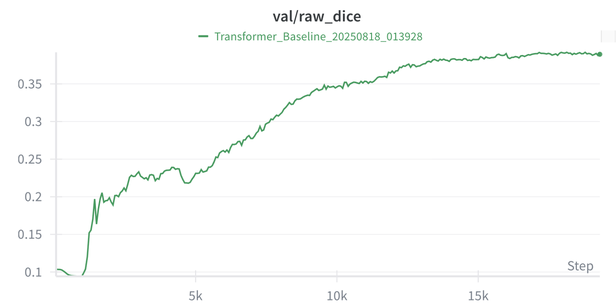


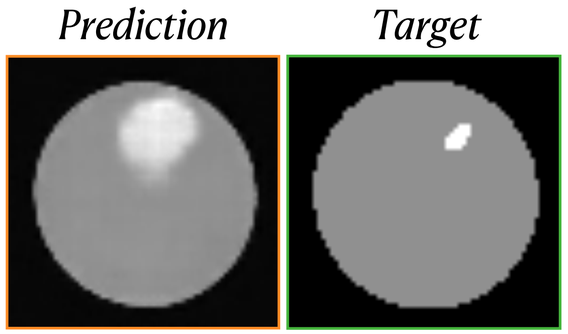
Results Stage 2 — Transformer + Frozen CNN Decoder

- **Teacher-Student Alignment**
  - Latent RMSE ↓ steadily; cosine similarity ↑ (>0.8), confirming strong alignment between transformer output and CNN latent space.
- **Reconstruction Quality**
  - Predictions capture global tissue structure with clearer feature enhancement than Stage 1; tumour boundaries are partially recovered, highlighting progress towards sharper localisation.
- **Validation Metrics**
  - RMSE (μ_a, μ_s) ↓ significantly after ~5k steps and stabilises.
  - Dice scores improve (~0.35 total), demonstrating consistent spatial overlap with ground truth.
  - Contrast recovery strengthens over training, with reconstructions trending towards more accurate boundary delineation.
- **Attention Dynamics**
  - Entropy ↓, indicating the model is focusing on informative measurements.
  - Feature enhancement metric ↓, showing latent features becoming more compact and structured.
- **Overall**
  - Stage 2 establishes a robust mapping from measurements -> latent features, achieving smoother, coherent reconstructions with improved global fidelity compared to Stage 1.

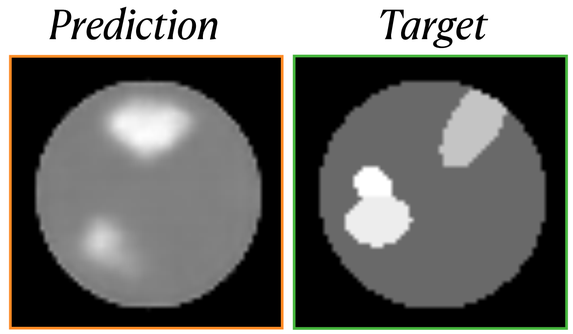








μ_a



μ'_s

Conclusion & Impact

- **End-to-End NIR-DOT Reconstruction Pipeline Demonstrated**
  - Hybrid CNN-Transformer with latent alignment achieves reliable μ_a and μ'_s recovery from simulated measurements.
  - First DL-DOT model to generalise across tissue shapes and probe geometries without retraining.
- **Scalable Two-Stage Training Framework**
  - Latent-only Stage 2 enables efficient transformer optimisation while preserving CNN decoder quality.
- **Ready for Enhanced Spatial Modelling**
  - Tissue patch integration and multi-modal fusion architecture prepared for improved reconstruction accuracy
- **Clear Path to Clinical Translation**
  - Framework supports larger dataset scaling, extended training, and validation with real NIR-DOT hardware and patient data.

Questions & Discussion