Preliminary Research Justification

Extending Transformer-Based Inverse Models for FD-DOT

Student: Max Hart
Programme: MSc Artificial Intelligence and Machine Learning
Supervisor: Professor Hamid Dehghani
Date: 2 July 2025

1. Methodological Justifications

1.1 Randomising Source–Detector Geometry

The original work by Dale introduced a transformer-based DOT model capable of real-time imaging under freehand scanning conditions. However, while the scan path was allowed to vary freely, the underlying source–detector (SD) geometry remained fixed during training. This simplification served to reduce computational complexity and preserve architectural consistency, but it implicitly encouraged the model to learn layout-specific mappings between measurements and tissue reconstructions. As a result, the trained model risks overfitting to a single probe configuration, limiting its ability to generalise across different devices, probe layouts, or clinical scenarios.

To address this limitation, this project introduces SD geometry randomisation during training — varying both the separation distances and angular orientations of the SD pairs within each sample. This strategy forces the model to learn a layout-invariant inverse mapping, grounded in the physics of photon transport rather than in memorised geometric patterns. By encountering a wide range of measurement geometries during training, the network becomes robust to variations in scanning hardware, positioning, and scan protocols. Randomising geometry also lays the necessary foundation for spatial-context-aware learning, ensuring the model can properly interpret tissue structure across multiple geometrical configurations — a critical step toward developing a reusable foundation model for generalisable FD-DOT.

1.2 Incorporating Spatial Tissue Context

In conventional DL-DOT models, each source–detector measurement is treated as an isolated signal characterised primarily by its amplitude, phase, and positional metadata. However, this approach overlooks a key factor in optical image reconstruction: the tissue lying between the source and detector significantly influences the light's propagation, attenuation, and scattering behaviour. Two measurements with identical geometry can yield very different signals if one passes through a tumour while the other passes through healthy tissue. Without knowledge of this intervening anatomical context, the model must infer it purely from the FD signal — a difficult and ambiguous task, especially in heterogeneous tissue environments.

To address this, we propose augmenting each measurement token with spatial tissue context — in the form of a local image patch or voxel block extracted from the region between the source and detector. This additional prior provides the model with explicit information about the tissue structure the light has traversed, allowing it to reason more accurately about how the signal relates to the underlying optical properties. It enhances the network's interpretability and reduces reliance on the FD signal alone. Moreover, this anatomical context can help disambiguate measurements that may otherwise appear similar due to noise or geometric alignment. In combination with transformer-based measurement encoding and geometry randomisation, spatial tissue context makes the inverse model more physically grounded, robust to domain shift, and ultimately more clinically applicable.

2. Research Pipeline and Roadmap

Having justified the inclusion of both randomised source–detector geometry and spatial tissue context, we now describe the proposed implementation pathway in detail. The roadmap below outlines what will be implemented, how it will be evaluated, and why each step is critical for developing a reusable foundation model for FD-DOT.

Stage 1: Baseline Reproduction

Objective: Recreate the transformer-based DL-DOT model described in Dale's work, using fixed SD geometry and no anatomical context.

We begin with a simplified training environment using synthetic 2D tissue phantoms (e.g., 256×256 pixels), generated with NIRFAST. Tumour inclusions will be placed randomly within the domain, and a fixed probe layout will be used — for example, SDS separations of 20 mm, 30 mm, and 40 mm. The forward model will simulate frequency-domain (FD) light transport at multiple wavelengths (e.g., 690 nm and 830 nm), generating both amplitude and phase components for each SD pair.

Each token will be formed from the source and detector coordinates $(x_s, y_s), (x_d, y_d)$ and the corresponding FD measurements:

Inputs to the model will consist of a separate token for each SD pair and each wavelength. That is, for every SD pair, we construct two tokens, one for each wavelength (e.g. 690 nm and 830 nm). Each token includes:

$$\text{Token}_{i,\lambda} = [x_s, y_s, x_d, y_d;\ A_\lambda, \phi_\lambda]$$

where (x_s, y_s) and (x_d, y_d) denote the coordinates of the source and detector respectively, and A_λ, ϕ_λ are the amplitude and phase values at wavelength λ.

This formulation ensures that each wavelength-specific signal is treated independently, allowing the transformer to learn how different wavelengths interact with tissue without entangling them prematurely. It also supports modular extension to additional wavelengths in future versions of the model.

These tokens are passed into a transformer encoder that learns global relationships across all SD measurements. The latent representation is then decoded via a CNN into the output volumes for μ_a and μ_s'.

The model will be trained with a composite loss combining voxel-wise RMSE and Dice coefficient:

$$\mathcal{L}_{\text{baseline}} = \lambda_{\text{MSE}} \cdot \text{RMSE}(\hat{\mu}_a, \mu_a) + \lambda_{\text{Dice}} \cdot \text{Dice}(\hat{\mu}_a, \mu_a)$$

This stage acts as a reference implementation and provides a performance baseline for comparison against subsequent stages.

Stage 2: Randomised Source–Detector Geometry

Objective: Improve generalisation by training the model over a broader range of SDS distances and coordinate placements.

Here, we move beyond fixed probe layouts by introducing geometric variation in the source and detector positions. Each training example will sample source and detector locations randomly within the valid tissue boundary. The SDS distance d will be drawn from a uniform distribution, e.g., $d \sim \mathcal{U}(10\,\text{mm}, 50\,\text{mm})$. The orientation will remain fixed — that is, sources and detectors will always be positioned perpendicular to the tissue surface, in accordance with NIRFAST's assumptions. This means the angle θ will not be varied, and the vertical incident assumption will be preserved.

This modification removes the model's dependence on any single probe layout, forcing it to learn spatially invariant mappings from arbitrary SD pairings to tissue property reconstructions. It also reflects more realistic variation in clinical probe placement, without violating simulation constraints.

Stage 2.5: Randomised Phantom Shapes and Probe Placement

Objective: Extend data diversity by simulating more realistic anatomical scenarios through randomly positioned ellipsoidal tissue phantoms embedded within a volumetric cuboid, coupled with robust probe surface placement.

While previous work, including Dale's, primarily simulated tissue within simple cuboidal domains (with or without embedded tumours), such a setup implicitly teaches the inverse model to rely on the geometry of the enclosing volume. In practice, patient anatomy is neither cuboidal nor spatially consistent across individuals or scanning sessions. Thus, generalising to variable tissue morphologies is essential for deploying FD-DOT in realistic clinical environments.

To this end, we propose generating each synthetic sample as follows:

1. Create a large volumetric container — for instance, a $50 \times 50 \times 50$ voxel cuboid that acts as an embedding space but represents air or background.

2. Randomly embed within it a **single large ellipsoidal volume**, which constitutes the healthy tissue. Its centre and radii are sampled anew for each sample, thereby altering its spatial extent and surface curvature. This ensures that each phantom exhibits a unique outer shape.

3. Inside this ellipsoidal tissue, randomly insert 0–5 smaller ellipsoidal tumours with optical contrasts sampled as in Table 5.1 of Dale's thesis.

This randomisation has multiple advantages:

- It decouples the learned inverse mapping from any fixed outer tissue boundary, preventing the model from exploiting consistent cuboid edges.

- By varying the centre and orientation of the ellipsoid, the local curvature and surface normal vectors change, creating diverse photon path distributions and forcing the network to learn more generalised spatial interactions.

- It changes the relative distance and angle to embedded tumours, naturally augmenting the spatial complexity seen by the model.

Probe placement on ellipsoidal surfaces Probes (i.e., source–detector configurations) are then explicitly placed on the **surface of this randomly generated ellipsoid.** This ensures that photons always enter directly into tissue without traversing air — a critical physical constraint. The procedure involves:

- Sampling a position within the bounding cuboid and projecting it onto the nearest point on the ellipsoid's surface, thus guaranteeing valid optical contact.
- Configuring detector offsets (or even fully randomised local placements) around each source point while remaining on the tissue surface.

Randomising the ellipsoid's centre is not merely cosmetic: it alters the local surface geometry where probes attach. This changes the curvature, depth to interior heterogeneities, and expected photon scattering paths. As a result, even if probe positions are sampled similarly in global coordinates, their *local* tissue environment varies significantly, enriching the training set.

Why not mesh the entire cuboid? Only the ellipsoidal tissue volume is meshed and used for FEM simulation. Meshing the entire cuboid would waste computational resources on regions representing air, where no photon propagation occurs. The outer cuboid purely acts as a placement scaffold for randomising spatial relationships.

Expected impact This approach pushes the dataset one step closer to clinical realism. The network can no longer exploit geometric shortcuts tied to fixed phantom shapes and is instead forced to infer tissue optical properties purely from FD measurements and explicit geometry/context cues. In conjunction with geometry-randomised probes and localised context patches (from Stage 3), this design forms a robust pipeline for learning spatially invariant inverse models that generalise to diverse anatomical and scanning conditions.

Stage 3: Integrating Localised Tissue Context

Objective: Improve reconstruction fidelity by explicitly encoding the anatomical structure surrounding each source and detector.

Rather than relying on signal and geometry alone, we introduce local anatomical information to each SD token. Specifically, for each SD pair, we extract two independent tissue patches from the μ_a and/or μ_s' maps — one centred around the source coordinates, and the other around the detector. Each patch (e.g., 32×32 pixels) captures local optical heterogeneity, allowing the model to infer how tissue structure affects light propagation.

Each patch is encoded using a shared CNN encoder:

$$\phi_s = \phi(\text{patch}_{\text{source}}), \quad \phi_d = \phi(\text{patch}_{\text{detector}})$$

These are concatenated to form a context embedding:

$$\phi_{\text{context}} = [\phi_s;\ \phi_d]$$

The final input token becomes:

$$\text{Token}_i = [x_s, y_s, x_d, y_d;\ \mathbf{x}_i;\ \phi_{\text{context}}]$$

This approach allows the model to reason about photon-tissue interactions in a more physically grounded manner, especially in complex or ambiguous regions. It also promotes robustness to noise, spatial distortions, and measurement variance by anchoring the input to known tissue structure.

Stage 4: Transfer Learning and Domain Adaptation (Optional)

Objective: Evaluate the reusability of the trained inverse model across new tissue domains, probe types, or anatomical targets.

Once the full model has been trained with geometry-randomised, context-enriched inputs, we explore its adaptability to more realistic use cases. Fine-tuning will be performed on MRI-derived breast or brain tissue volumes, using a small number of new training samples. Lower layers of the transformer and patch encoder may be frozen, with only the CNN decoder and later attention blocks updated.

This strategy allows the model to serve as a general-purpose inverse solver with minimal overhead for domain-specific deployment. If successful, it could pave the way for robust clinical applications using heterogeneous imaging setups.

Possible extensions include zero-shot evaluation, multi-domain finetuning, or uncertainty-aware learning strategies to further evaluate model robustness.

5. Current Implementation Status and Comparison

This section provides a detailed summary of the current implementation, benchmarking it directly against both Dale's original work and the methodological goals set out in the proposal. It highlights the novel technical aspects achieved so far, explains why these are significant, and outlines the immediate next steps.

5.1 Phantom Geometry and Optical Properties

A completely new data simulation pipeline has been built from scratch using `NIRFASTer-FF`, designed to maximise spatial and optical variability. Unlike Dale's approach, which primarily relied on axis-aligned cuboidal phantoms with static probe arrangements, this implementation generates randomised ellipsoidal healthy tissue volumes embedded within a larger 3D air-filled cuboid. Each ellipsoid's centre, orientation (via random SO(3) rotations), and semi-axes are independently sampled, preventing any consistent outer boundary shape. Multiple smaller ellipsoidal tumours are then embedded with strict geometric checks — ensuring at least 80% of each tumour lies inside the healthy tissue. This substantially reduces artefacts like "floating tumours in air", which were less rigorously controlled in Dale's code.

Optical properties are also enhanced beyond Dale's defaults. Each phantom samples μ_a and μ_s' for the healthy tissue from uniform physiological ranges, then assigns tumour optical contrasts using additional multipliers (1.5–3.5$\times$ for μ_a, 1.5–2.5$\times$ for μ_s'). This ensures every generated phantom presents a unique optical challenge, critical for preventing overfitting in downstream learning.

5.2 Fully Randomised Probe Geometry

Perhaps the most significant departure from Dale's pipeline is in probe configuration. Instead of scanning a rigid multi-detector probe across a flat surface, this implementation extracts explicit surface voxels from the ellipsoidal tissue volume via morphological erosion. Sources are then randomly sampled across this complex surface, and detectors are placed under strict distance constraints (minimum 5 mm) to emulate realistic diffusive sampling. Each source is paired with three detectors, leading to hundreds of spatially unique probe configurations per phantom. This ensures the inverse model must generalise across an effectively unlimited variety of probe layouts, aligning exactly with the proposal's goal of geometry-invariant reconstruction.

5.3 Frequency-Domain Forward Modelling and Data Output

Forward light propagation is modelled using the frequency-domain diffusion equation at 140 MHz, matching typical experimental systems. Measurement noise is explicitly added post-simulation: 2% relative amplitude noise and 2° phase noise, consistent with clinical SNR. Each dataset saves:

- Log-amplitude and phase data reshaped into $(N_{\text{probes}}, 3)$ arrays (3 detectors per source).
- Complete volumetric ground truth maps $(Nx, Ny, Nz, 2)$ for μ_a and μ_s'.
- Full probe coordinate metadata for subsequent patch extraction or geometric encoding.

This structure is intentionally designed so that local patches (Stage 3) can be extracted later during PyTorch dataset loading, without re-running the FEM solver.

5.4 Advantages Over Dale's Implementation

Compared to Dale's thesis and ECBO abstract:

1. The phantom shapes, optical contrasts, and probe geometries are all randomised far more extensively, breaking any implicit spatial priors.
2. Tumour placements are rigorously validated to prevent unphysical intersections with air, unlike simpler volumetric inclusions in the earlier pipeline.
3. Noise modelling is explicitly parameterised and logged for reproducibility.
4. A professional logging system captures mesh quality statistics, phantom compositions, and optode placements, building a robust audit trail for publications.

5.5 Immediate Next Steps

This codebase fully satisfies the spatial generalisation goals articulated in the proposal. The next technical milestones are:

- Extract local volumetric patches around each source and detector position from the saved ground truth. This will enable explicit context embedding for each measurement token.
- Develop the hybrid transformer-CNN architecture to jointly process FD measurements and local tissue context.

- Begin large-scale data generation (e.g., 500 phantoms × 500 probes) to support robust training and statistical analysis.

Overall, this implementation represents a substantial technical advance over Dale's original approach and aligns precisely with the dissertation's ambition of building a generalisable foundation model for FD-DOT.