



Improved Objective Functions for Implicit Neural Representations in Deformable Image Registration

Johannes B. Gebauer¹ 

J.GEBAUER@UKE.DE

Maximilian Nielsen¹ 

M.NIELSEN@UKE.DE

Frederic Madesta¹ 

F.MADESTA@UKE.DE

René Werner¹ 

R.WERNER@UKE.DE

Thilo Sentker¹ 

T.SENTKER@UKE.DE

¹ *Institute for Applied Medical Informatics, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany*

Editors: Under Review for MIDL 2026

Abstract

We propose an enhanced multi-scale Implicit Neural Representation (INR) framework for dense deformable image registration, designed to maximize alignment accuracy and deformation regularity. By modeling the transformation as a coordinate-based neural field, we optimize directly on image pairs using a coarse-to-fine dual-branch architecture that splits motion into global and local components. The objective function is driven by mask-guided Normalized Cross-Correlation and curvature regularization to ensure smooth, anatomically plausible motion. Evaluation on the DIR-Lab 4DCT thorax dataset demonstrates state-of-the-art performance with a mean Target Registration Error (TRE) below 1.0 mm. On the more challenging DIR-Lab COPDgene thorax dataset, the model achieves robust alignment with a mean TRE of 1.23 mm, yielding performance comparable to leading classical optimization frameworks. A comprehensive ablation study confirms that the dual-branch design and multi-scale optimization strategy are necessary to achieve these results, enabling precise registration with modest computational overhead. These findings demonstrate that carefully structured INR frameworks can achieve sub-millimeter precision on standard benchmarks while maintaining robustness in large-deformation scenarios. Source code will be made available upon acceptance at <https://github.com/IPMI-ICNS-UKE/INR-DIR>.

Keywords: Implicit Neural Representations, Deformable Image Registration, Multi-Scale Optimization, Thoracic CT

1. Introduction

Accurate deformable image registration (DIR) is essential for numerous clinical applications, ranging from image-guided interventions and quantitative longitudinal analysis to multi-modal image fusion (Sotiras et al., 2013). The challenge lies in estimating a dense, non-rigid transformation field that achieves precise spatial alignment between a moving and a fixed image.

Traditionally, DIR methods, whether based on classical optimization or deep learning (DL) with convolutional neural networks (CNN) (Polzin et al., 2013; Vishnevskiy et al., 2016; Balakrishnan et al., 2019; Hering et al., 2021; Hansen and Heinrich, 2021), rely on representing the deformation field via discrete voxel grids. This discretization leads to inherent limitations on the transformation. First, the spatial resolution of the deformation is

coupled to the grid size, limiting the ability to model fine-grained, sub-voxel motion without excessive computational costs. Second, while smooth transformations are anatomically essential, grid-based methods typically enforce regularity through complex interpolation schemes (e.g., B-splines) or discrete regularization penalties, which can complicate the optimization landscape (Rueckert et al., 2002).

To overcome these structural constraints, we adopt the paradigm of implicit neural representations (INR). INR model signals as continuous functions parameterized by a multi-layer perceptron (MLP), offering a resolution-independent mechanism to encode spatial transformations (Sitzmann et al., 2020). In our framework, the deformation field is modeled as a continuous function of spatial coordinates, where the network predicts the displacement vector. To capture high-frequency components of the deformation that standard MLP often fail to resolve (spectral bias), we utilize a network with periodic activation functions (SIREN), enabling the model to learn fine anatomical details efficiently.

The transition from discrete voxel grids to coordinate-based neural fields, as demonstrated by the INR framework (IDIR) of (Wolterink et al., 2022), offers advantages over conventional optimization schemes. First, the estimated deformation is intrinsically continuous, capable of representing motion at arbitrary spatial resolutions. Second, the inherent differentiability of the MLP allows for the analytic computation of spatial derivatives via automatic differentiation. This simplifies the implementation of complex smoothness priors, such as curvature regularization, directly within the energy functional.

However, while the standard SIREN approach of (Wolterink et al., 2022) is effective at capturing fine details, it is prone to local minima when facing large-magnitude deformations. To address this limitation, we advance the INR-based framework by integrating principles from classical registration theory, specifically multi-resolution strategies and the spectral decoupling of deformations, into a unified multi-scale, dual-branch architecture. Unlike the standard single-network design, our framework explicitly decouples global anatomical motion from fine-grained local deformations. By optimizing these components in a coarse-to-fine manner, our approach ensures robust alignment for large anatomical changes, such as those seen in COPD patients, while retaining the sub-voxel precision characteristic of SIREN-based INR.

2. Method

2.1. Problem Formulation

Let $I_f, I_m : \Omega \rightarrow \mathbb{R}$ denote the fixed and moving images, respectively, defined on a spatial domain $\Omega \subset \mathbb{R}^3$. To restrict the optimization to the relevant anatomical structures, we utilize binary lung masks $M : \Omega \rightarrow \{0, 1\}$, automatically generated using TotalSegmentator (Wasserthal et al., 2023).

The goal of deformable image registration is to estimate a dense, non-linear transformation field $\phi : \Omega \rightarrow \Omega$ that spatially aligns I_m to I_f , such that the warped moving image $I_m \circ \phi$ is anatomically consistent with I_f .

Since estimating ϕ solely from image intensity data is an ill-posed problem, we formulate the registration as an energy minimization task combining a masked data fidelity term and

a regularization prior:

$$\hat{\phi} = \arg \min_{\phi} \mathcal{L}_{\text{sim}}(I_f, I_m \circ \phi; M) + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}}(\phi), \quad (1)$$

where \mathcal{L}_{sim} quantifies the dissimilarity between the fixed and warped images within the region defined by the lung mask M , \mathcal{L}_{reg} promotes smoothness and topological regularity in the deformation field, and $\lambda_{\text{reg}} > 0$ is a hyperparameter controlling the trade-off between alignment accuracy and deformation plausibility.

2.2. Implicit Neural Representation of the Deformation Field

Unlike traditional approaches that parameterize the deformation field on a discrete voxel grid, we model ϕ as a continuous function parameterized by a neural network, referred to as an INR. We define the deformation in a residual form:

$$\phi_{\theta}(\mathbf{x}) = \mathbf{x} + u_{\theta}(\mathbf{x}), \quad (2)$$

where $\mathbf{x} \in \Omega$ represents the spatial coordinates and $u_{\theta} : \Omega \rightarrow \mathbb{R}^3$ is the displacement field predicted by a MLP with trainable parameters θ . This residual formulation provides an inductive bias towards the identity transformation, ensuring stable initialization and faster convergence.

Standard MLP with ReLU activation functions suffer from spectral bias (Rahaman et al., 2019), creating difficulties in learning high-frequency functions. To overcome this and capture fine-grained anatomical details, we adopt the SIREN framework (Sitzmann et al., 2020), which has been successfully applied to medical image registration (Wolterink et al., 2022; Sun et al., 2024). A SIREN network utilizes sinusoidal activation functions and can be expressed as a composition of layers:

$$u_{\theta}(\mathbf{x}) = \mathbf{W}_n(\psi_{n-1} \circ \psi_{n-2} \circ \cdots \circ \psi_0)(\mathbf{x}) + \mathbf{b}_n, \quad (3)$$

where ψ_i denotes the i -th layer of the network. Each layer applies an affine transformation defined by a weight matrix \mathbf{W}_i and a bias vector \mathbf{b}_i , followed by a component-wise sine non-linearity:

$$\psi_i(\mathbf{y}) = \sin(\mathbf{W}_i \mathbf{y} + \mathbf{b}_i). \quad (4)$$

A distinct property of INR is the ability to compute spatial derivatives analytically. The Jacobian of the transformation, $J_{\phi}(\mathbf{x}) = \mathbf{I} + \nabla_{\mathbf{x}} u_{\theta}(\mathbf{x})$, is obtained exactly via automatic differentiation. A key advantage of sinusoidal activations is that they are infinitely differentiable (C^{∞}), enabling the computation of higher-order derivatives (e.g., the Hessian) if required. Furthermore, the sampling strategy for training points is not restricted to a regular grid, allowing for flexible, off-grid optimization.

As described in (Sitzmann et al., 2020), careful initialization is needed for SIREN networks. We follow the proposed scheme and draw the initial weights from a uniform distribution $\mathbf{W}_i \sim \mathcal{U}(-\sqrt{6/n}, \sqrt{6/n})$, where n is the number of inputs to the layer. Additionally, we scale the weights of the first layer by a factor ω_0 to control the spatial frequency spectrum of the initial output.

2.3. Optimization and Regularization

Our framework adopts an instance-specific optimization strategy. Instead of training on a large dataset to learn a global registration function, we optimize the neural field parameters θ directly for a given image pair.

During each iteration, the moving image is resampled at the transformed coordinates $\phi_\theta(\mathbf{x})$ using differentiable linear interpolation. We denote the warped image as $I_m \circ \phi_\theta$. The registration is formulated as the minimization of the following objective with respect to θ :

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}_{\text{sim}}(I_f, I_m \circ \phi_\theta) + \mathcal{L}_{\text{reg}}(u_\theta). \quad (5)$$

For the similarity term \mathcal{L}_{sim} , we employ the Normalized Cross-Correlation (NCC) loss, which is robust to linear intensity variations between scans.

To ensure physically plausible deformations, we apply regularization during optimization. Specifically, to ensure smoothness, we utilize a curvature regularizer based on the Laplacian of the displacement field, as proposed in (Fischer and Modersitzki, 2003):

$$\mathcal{L}_{\text{curv}} = \int_{\Omega} \sum_{i=1}^3 \left(\Delta u_{\theta}^{(i)}(\mathbf{x}) \right)^2 d\mathbf{x}, \quad (6)$$

where $u_{\theta}^{(i)}$ denotes the i -th component of the displacement vector.

We observed that penalizing the Laplacian term alone, rather than the full bending energy as implemented in (Wolterink et al., 2022), yields a computational speed-up by reducing the number of required second-order derivative calculations, without sacrificing registration performance.

Furthermore, to prevent folding in the deformation field, we enforce a regularization penalty on negative Jacobian determinants:

$$\mathcal{L}_{\text{jac}} = \int_{\Omega} \text{ReLU}(-\det(\nabla u_{\theta}(\mathbf{x}))) d\mathbf{x} \quad (7)$$

where the ReLU function penalizes only regions with negative Jacobian determinants (indicating local folding), while preserving differentiability.

Finally, to enforce anatomical consistency, we incorporate a mask-based semantic loss. Specifically, we compute the binary cross-entropy between the fixed and warped moving masks for all five lung lobes (left upper/lower; right upper/middle/lower). This ensures global alignment at the lobar level using segmentations obtained via TotalSegmentator (Wasserthal et al., 2023). The resulting composite regularization objective combines smoothness and semantic constraints:

$$\mathcal{L}_{\text{reg}} = \lambda_{\text{curv}} \mathcal{L}_{\text{curv}} + \lambda_{\text{jac}} \mathcal{L}_{\text{jac}} + \lambda_{\text{mask}} \mathcal{L}_{\text{mask}} \quad (8)$$

2.4. Multi-Scale Dual-Branch Optimization

To effectively recover large deformations, we implement a coarse-to-fine optimization strategy inspired by classical multi-resolution registration techniques. We mitigate the risk of getting trapped in local minima by initially optimizing the network on smoothed image pairs

obtained via Gaussian filtering. The smoothed intensity G_σ at voxel coordinates (i, j, k) is defined using a kernel size of $2\sigma + 1$:

$$G_\sigma[i, j, k] = \frac{1}{(2\pi)^{3/2}\sigma^3} \exp\left(-\frac{i^2 + j^2 + k^2}{2\sigma^2}\right), \quad (9)$$

where σ controls the smoothing scale. This filtration suppresses high-frequency details, forcing the network to focus on global structural alignment in the early stages. We adopt a multi-stage approach, reducing σ in discrete steps ($\sigma \in \{4, 2\}$) as the optimization progresses. In the final stage ($\sigma = 0$), the original non-filtered image is utilized.

To further disentangle global and local motion, we decompose the displacement field into two parallel MLP branches, a coarse branch and a fine branch, as illustrated in Fig. 1. The residual deformation (Eq. 2) is modified to:

$$\phi_\theta(\mathbf{x}) = \mathbf{x} + c_{\text{coarse}} \cdot u_{\theta, \text{coarse}}(\mathbf{x}) + c_{\text{fine}} \cdot u_{\theta, \text{fine}}(\mathbf{x}), \quad (10)$$

where $u_{\theta, \text{coarse}}$ and $u_{\theta, \text{fine}}$ represent the outputs of the respective branches, and $c_{\text{coarse}}, c_{\text{fine}} \in \mathbb{R}$ are additional learnable scaling parameters.

We use this architecture to introduce frequency-specific biases. For the coarse branch, we set the SIREN frequency initialization parameter $\omega_0 = 10$ (lower than the standard $\omega_0 = 30$), explicitly biasing it towards learning low-frequency spatial variations. The training procedure is sequential:

1. **Phase 1 (Coarse):** Only the coarse branch is optimized using heavily smoothed images ($\sigma = 4$). The fine branch is inactive ($c_{\text{fine}} = 0$).
2. **Phase 2 (Refinement):** As training progresses, we reduce σ to 2 and finally to 0. Simultaneously, we activate the fine branch. To ensure a stable transition, c_{fine} is initialized at a lower magnitude relative to the coarse scale (specifically $c_{\text{fine}} = c_{\text{coarse}}/10$), allowing the network to progressively add local details to the established global deformation.

2.5. Implementation Details

We optimize the network by iteratively sampling random coordinates \mathbf{x} strictly from within the fixed image lung mask M_f . At each iteration, we sample a batch of $N = 30,000$ points. We train for a cumulative total of 3,000 steps distributed across the three optimization stages ($\sigma \in \{4, 2, 0\}$). We employ the AdamW optimizer with an initial learning rate of 10^{-3} , which is gradually reduced to 10^{-5} following a cosine annealing schedule.

For the network topology, we utilize the dual-branch architecture with layer dimensions and frequency initializations (ω_0) as defined in Sec. 2.4. The scaling factors are initialized as $c_{\text{coarse}} = 0.5$ and, at the point where the fine branch is activated, $c_{\text{fine}} = 0.1 \cdot c_{\text{coarse}}$. The loss weighting hyperparameters are set to $\lambda_{\text{curv}} = 0.01$, $\lambda_{\text{jac}} = 0.01$, and $\lambda_{\text{mask}} = 0.1$. All experiments were conducted on a single NVIDIA 3090 GPU using PyTorch.

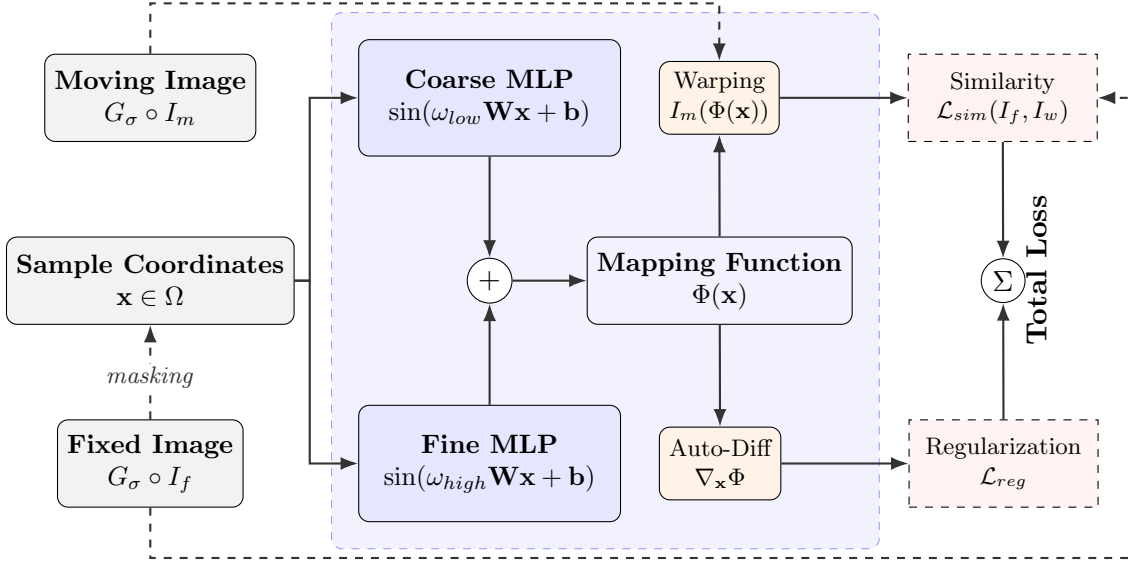


Figure 1: Overview of the proposed multi-scale dual-branch INR framework.

2.6. Datasets

We evaluate the proposed INR-based registration framework using two standard benchmarks for thoracic image registration: the DIR-Lab 4D Computed Tomography (4DCT) dataset (Castillo et al., 2009) and the DIR-Lab COPDgene dataset (Castillo et al., 2013). Both datasets include manually annotated anatomical landmarks, enabling the quantitative assessment of Target Registration Error (TRE).

2.6.1. DIR-LAB 4DCT

The DIR-Lab 4DCT dataset comprises 10 subject-specific 4DCT scans. Each case contains 10 volumetric CT images representing distinct phases of the respiratory cycle (phases 0 through 9), spanning from maximum inhalation to maximum exhalation. To evaluate registration accuracy, we utilize the provided set of 300 expert-identified landmarks for each case, which are defined on the extreme phases (end-inspiration and end-expiration).

The image dimensions and voxel spacings vary across subjects:

- **Cases 1–5:** Spatial resolution of $256 \times 256 \times [99\text{--}112]$ voxels, with anisotropic voxel spacing ranging from $[0.97\text{--}1.16] \times [0.97\text{--}1.16] \times 2.5 \text{ mm}^3$.
- **Cases 6–10:** Higher spatial resolution of $512 \times 512 \times [106\text{--}136]$ voxels, with a fixed voxel spacing of $0.97 \times 0.97 \times 2.5 \text{ mm}^3$.

2.6.2. DIR-LAB COPDGENE

To assess performance on pathology-induced large deformations, we utilize the DIR-Lab COPDgene dataset. This set consists of 10 image pairs derived from patients with Chronic Obstructive Pulmonary Disease (COPD), providing distinct volumetric breath-hold CT images for the inhalation (iBH) and exhalation (eBH) phases.

The spatial parameters for this dataset are grouped as follows:

- **Case 2:** Spatial resolution of $256 \times 256 \times 112$ voxels, with a physical spacing of $0.645 \times 0.645 \times 2.5 \text{ mm}^3$.
- **Cases 1 & 3–10:** Spatial resolution of $512 \times 512 \times [112\text{--}135]$ voxels, with in-plane spacing ranging from 0.586 to 0.742 mm and a fixed slice thickness of 2.5 mm.

2.7. Evaluation Metrics

To quantitatively assess registration performance, we utilize three standard metrics focusing on landmark accuracy, anatomical overlap, and deformation regularity.

The primary metric for registration accuracy is the Target Registration Error (TRE). We follow the standard evaluation protocol for the DIR-Lab benchmarks. For each corresponding landmark pair $(\mathbf{p}_m, \mathbf{p}_f)$ in the moving and fixed domains, the fixed landmark is first transformed by the estimated deformation field to obtain $\mathbf{p}'_f = \phi(\mathbf{p}_f)$. This transformed coordinate is then snapped (rounded) to the nearest integer voxel index. The final TRE is computed as the Euclidean distance between this snapped coordinate and the moving landmark \mathbf{p}_m in physical world coordinates (millimeters). We report the mean TRE and standard deviation over all 300 expert landmarks provided for each case.

To evaluate the spatial overlap of anatomical structures, we compute the Dice Similarity Coefficient (DSC). Since our registration framework focuses specifically on the pulmonary region, we assess alignment accuracy using the segmentations of the five anatomical lung lobes described in Sec. 2.3.

3. Results

We trained our dual-branch INR on all 20 cases described in Sec. 2.6. The resulting TRE values are compared against competing deep learning-based methods in Table 1. For IDIR (Wolterink et al., 2022), we report values obtained using the official implementation with default hyperparameters, as case-wise results for COPDgene were not available in the original publication. On average, our proposed model outperforms the listed methods on both datasets.

While performance is comparable to IDIR on cases with small initial displacements (i.e., 4DCT Cases 1 and 2), our model demonstrates superior robustness in scenarios with large anatomical deformations, particularly within the COPDgene cohort.

Further inspection of the individual network branches confirms that our training scheme effectively separates global and local motion. The coarse branch captures the majority of the deformation, with mean displacement magnitudes $6\times$ and $14\times$ larger than those of the fine branch for the 4DCT and COPDgene datasets, respectively. This decomposition is visually exemplified in Fig. 2.

3.1. Ablation Study

To validate the contributions of the proposed components, we conduct an ablation study comparing the following configurations:

Table 1: Quantitative comparison on DIR-Lab 4DCT and DIR-Lab COPDgene data (Cases 1–10), evaluated by TRE (mean and std) in mm for each individual case, along with the mean performance over all cases. We compare our method against other learning-based approaches: GraphRegNet (deep graphs), VIRNet (CNN), IDIR (INR). Bold indicates the best result.

Method	DIR-Lab 4DCT Case ID										Overall
	1	2	3	4	5	6	7	8	9	10	
no reg.	4.01 (2.91)	4.65 (4.09)	6.73 (4.21)	9.42 (4.81)	7.10 (5.15)	11.10 (6.98)	11.59 (7.87)	15.16 (9.11)	7.82 (3.99)	7.63 (6.54)	8.52 (5.57)
GraphRegNet ¹	0.86 (N/A)	0.90 (N/A)	1.06 (N/A)	1.45 (N/A)	1.60 (N/A)	1.59 (N/A)	1.74 (N/A)	1.46 (N/A)	1.58 (N/A)	1.71 (N/A)	1.39 (N/A)
VIRNet ²	0.99 (0.47)	0.98 (0.46)	1.11 (0.61)	1.37 (1.03)	1.32 (1.36)	1.15 (1.12)	1.05 (0.81)	1.22 (1.44)	1.11 (0.66)	1.05 (0.72)	1.14 (0.76)
IDIR ³	0.76 (0.94)	0.76 (0.94)	0.94 (1.02)	1.32 (1.27)	1.23 (1.47)	1.09 (1.03)	1.12 (1.00)	1.21 (1.29)	1.22 (0.95)	1.01 (1.05)	1.07 (1.10)
Ours	0.76 (0.91)	0.76 (0.91)	0.88 (1.03)	1.27 (1.22)	1.16 (1.46)	0.95 (0.99)	0.93 (0.97)	1.10 (1.25)	1.00 (0.93)	0.94 (0.95)	0.98 (1.07)

Method	DIR-Lab COPDgene Case ID										Overall
	1	2	3	4	5	6	7	8	9	10	
no reg.	25.90 (11.57)	21.77 (6.46)	12.29 (6.39)	30.90 (13.49)	30.90 (14.05)	28.32 (9.20)	21.66 (7.66)	25.57 (13.61)	14.84 (10.01)	22.48 (10.64)	23.46 (10.31)
GraphRegNet ¹	1.38 (N/A)	2.09 (N/A)	1.22 (N/A)	1.58 (N/A)	1.37 (N/A)	1.10 (N/A)	1.19 (N/A)	1.19 (N/A)	0.99 (N/A)	1.38 (N/A)	1.34 (N/A)
VIRNet ²	—	—	—	—	—	—	—	—	—	—	—
IDIR ³	1.55 (1.67)	2.66 (3.57)	1.35 (1.07)	1.46 (1.18)	1.43 (1.54)	66.46 (31.87)	1.39 (1.33)	1.72 (2.00)	1.28 (1.50)	1.62 (1.28)	8.09 (4.70)
Ours	1.11 (1.26)	2.07 (3.19)	1.11 (0.98)	1.10 (0.94)	1.09 (1.16)	1.40 (2.07)	1.04 (1.19)	1.27 (1.68)	0.95 (1.20)	1.21 (1.14)	1.23 (1.48)

¹(Hansen and Heinrich, 2021) ²(Hering et al., 2021) ³(Wolterink et al., 2022)

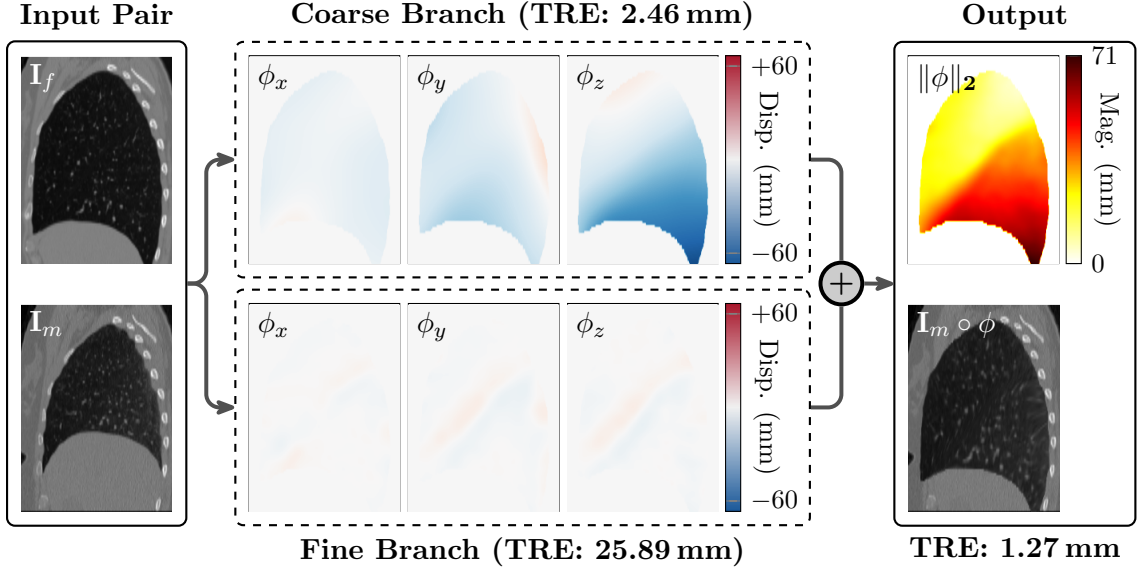


Figure 2: Visual decomposition of the registration field (Case 8, DIR-Lab COPDgene). Left: Fixed (I_f) and moving (I_m) input images. Center: The learned displacement fields are decomposed into x (lateral), y (anterior-posterior), and z (inferior-superior) components (ϕ_x, ϕ_y, ϕ_z) for both the coarse and fine branches. Right: The magnitude field $\|\phi\|_2$ and the resulting warped image $I_m \circ \phi$.

1. **Single-Branch Baseline:** A standard SIREN network (3 hidden layers, 256 units, $\omega_0 = 30$) trained directly on high-resolution images, representing the vanilla INR approach.

2. **Single-Branch + Multi-scale:** The standard SIREN trained using our coarse-to-fine Gaussian smoothing schedule, testing the impact of scheduled learning on a standard architecture.
3. **Dual-Branch (No Multi-scale):** Our proposed dual-branch architecture trained directly on high-resolution images without the smoothing schedule, testing the capability of the architecture alone to decouple motion.
4. **Dual-Branch + Multi-scale (Full Model):** The complete proposed framework, combining the dual-branch architecture with the coarse-to-fine Gaussian smoothing schedule.

To ensure a fair comparison, all ablation models were trained using the identical regularization functions and optimization scheme detailed in Sec. 2.5. Consequently, our single-branch baseline serves as a controlled variant of the IDIR architecture, differing only in the loss functions and training protocol. The quantitative results of this analysis are summarized in Table 2.

On the 4DCT dataset, performance differences between model variants were negligible. However, for the larger deformations in the COPDgene data, both the dual-branch architecture and the blurring schedule were relevant for minimizing TRE. DSC scores remained consistent across all ablation models, indicating robust alignment of anatomical boundaries. The exclusion of the blurring schedule compromised stability, resulting in convergence failure for the dual-branch model in two cases (COPDgene 4 and 6).

Table 2: Ablation Study on DIR-Lab 4DCT and COPDgene datasets. We evaluate the impact of the dual-branch architecture and multi-scale optimization on registration accuracy (TRE) and the mean DSC over the five lung lobes. Bold indicates the best results.

Configuration		DIR-Lab 4DCT		DIR-Lab COPD	
Architecture	Multi-scale	TRE (mm) ↓	DSC ↑	TRE (mm) ↓	DSC ↑
SINGLE-BRANCH	–	1.01 ± 0.18	0.964	1.41 ± 0.34	0.953
SINGLE-BRANCH	✓	1.01 ± 0.19	0.964	1.44 ± 0.36	0.952
DUAL-BRANCH	–	0.97 ± 0.16	0.965	4.55 ± 5.40	0.946
DUAL-BRANCH	✓	0.98 ± 0.16	0.965	1.23 ± 0.30	0.955

4. Discussion

In this work, we proposed an INR framework for DIR, designed to overcome the limitations of fixed-grid discretizations and single-scale neural fields. By introducing a multi-scale, dual-branch architecture, we aimed to combine the conflicting requirements of capturing large-magnitude anatomical shifts while resolving fine-grained local details. Our evaluation on the DIR-Lab 4DCT and COPDgene benchmarks confirms the efficacy of this approach, yielding

state-of-the-art precision on standard respiratory motion (4DCT) and robust performance on pathology-induced (COPDgene) deformations.

The core hypothesis of our study was that standard coordinate-based MLP face a performance barrier when modeling complex, multi-scale deformations, as e.g., present in the DIR-Lab COPD data. Our ablation study reveals that while the single-branch baseline is robust (TRE: 1.41 mm), it hits a performance ceiling and fails to benefit from explicit multi-scale optimization strategies (TRE: 1.44 mm). This suggests an inherent difficulty in simultaneously resolving global and local frequencies within a single stream. Conversely, our proposed dual-branch architecture is unstable in isolation (TRE: 4.55 mm), likely due to the ill-posed nature of signal decomposition without spectral guidance. However, when combined with the coarse-to-fine schedule, the dual-branch framework effectively breaks the performance barrier of the single-stream baseline, utilizing the spectral decoupling to reduce the mean TRE to 1.23 mm.

On the DIR-Lab 4DCT dataset, our method achieved a mean TRE of < 1.0 mm. To our knowledge, this sub-millimeter accuracy represents a new state-of-the-art for INR-based methods and outperforms established DL baselines such as GraphRegNet (Hansen and Heinrich, 2021) and VIRNet (Hering et al., 2021). This validates the hypothesis that instance-specific optimization, free from the constraints of grid resolution and domain shift, can achieve high precision when carefully regularized.

For the COPDgene dataset, we observed a mean TRE of 1.23 mm. While this does not strictly outperform the best discrete optimization methods, e.g., isoPTV with an overall TRE of 0.96 mm (Vishnevskiy et al., 2016), it demonstrates robustness for a gradient-based method. Unlike standard DL approaches that often fail to generalize to the extreme lung motion seen in COPD without extensive pathology-specific training data, our instance-specific approach adapts to the geometry of each pair. Furthermore, our results show a clear improvement over the single-stream INR framework of (Wolterink et al., 2022), confirming that architectural modifications are necessary to scale INRs to large-deformation regimes.

Despite promising results, our approach has limitations inherent to instance-specific INRs. First, the computational cost exceeds that of single-shot learning methods. While we used a conservative 3,000-iteration schedule for experimental consistency, standard cases (e.g., DIR-Lab 4DCT) typically converge to sub-millimeter TREs in approximately 60 seconds. This precludes real-time use but remains viable for offline tasks like radiotherapy planning. Second, although robust to large deformations, modeling physiological sliding motion (e.g., at the pleural boundary) remains challenging due to the continuous nature of vector fields. Future work could investigate spatially adaptive regularization and specialized sampling schemes to better resolve these interfaces.

Overall, we have presented a robust, coordinate-based registration framework for thoracic 4DCT. By effectively combining the continuous representation power of SIREN in a dual-branch architecture with a classical coarse-to-fine optimization strategy, our method offers a powerful tool for medical image analysis where precise alignment is highly relevant.

References

- Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.
- Richard Castillo, Edward Castillo, Rudy Guerra, Valen E Johnson, Travis McPhail, Amit K Garg, and Thomas Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine & Biology*, 54(7):1849, 2009.
- Richard Castillo, Edward Castillo, David Fuentes, Moiz Ahmad, Abbie M Wood, Michelle S Ludwig, and Thomas Guerrero. A reference dataset for deformable image registration spatial accuracy evaluation using the copdgene study archive. *Physics in Medicine & Biology*, 58(9):2861, 2013.
- Bernd Fischer and Jan Modersitzki. Curvature based image registration. *Journal of Mathematical Imaging and Vision*, 18(1):81–85, 2003.
- Lasse Hansen and Mattias P Heinrich. Graphregnet: Deep graph regularisation networks on sparse keypoints for dense registration of 3d lung cts. *IEEE Transactions on Medical Imaging*, 40(9):2246–2257, 2021.
- Alessa Hering, Stephanie Häger, Jan Moltz, Nikolas Lessmann, Stefan Heldmann, and Bram Van Ginneken. Cnn-based lung ct registration with multiple anatomical constraints. *Medical Image Analysis*, 72:102139, 2021.
- Thomas Polzin, Jan Rühaak, René Werner, Jan Strehlow, Stefan Heldmann, Heinz Handels, and Jan Modersitzki. Combining automatic landmark detection and variational methods for lung ct registration. In *Fifth international workshop on pulmonary image analysis*, pages 85–96, 2013.
- Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5301–5310. PMLR, 09–15 Jun 2019.
- Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 2002.
- Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.
- Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.

- Shanlin Sun, Kun Han, Chenyu You, Hao Tang, Deying Kong, Junayed Naushad, Xiangyi Yan, Haoyu Ma, Pooya Khosravi, James S. Duncan, and Xiaohui Xie. Medical image registration via neural fields. *Medical Image Analysis*, 97:103249, 2024. ISSN 1361-8415. doi: <https://doi.org/10.1016/j.media.2024.103249>.
- Valery Vishnevskiy, Tobias Gass, Gabor Szekely, Christine Tanner, and Orcun Goksel. Isotropic total variation regularization of displacements in parametric image registration. *IEEE transactions on medical imaging*, 36(2):385–395, 2016.
- Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023.
- Jelmer M Wolterink, Jesse C Zwienenberg, and Christoph Brune. Implicit neural representations for deformable image registration. In *International Conference on medical imaging with deep learning*, pages 1349–1359. PMLR, 2022.