Medical Image Analysis 97 (2024) 103249

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Medical image registration via neural fields

Shanlin Sun [a], Kun Han [a], Chenyu You [b], Hao Tang [a], Deying Kong [a], Junayed Naushad [a], Xiangyi Yan [a], Haoyu Ma [a], Pooya Khosravi [a], James S. Duncan [b], Xiaohui Xie [a,*]

[a] University of California, Irvine, Irvine, CA 92697, USA
[b] Yale University, New Haven, CT 06520, USA

ARTICLE INFO

Keywords:
Optimization
Neural fields
Deformable image registration
Neural ODEs
Hybrid Coordinate samplers

ABSTRACT

Image registration is an essential step in many medical image analysis tasks. Traditional methods for image registration are primarily optimization-driven, finding the optimal deformations that maximize the similarity between two images. Recent learning-based methods, trained to directly predict transformations between two images, run much faster, but suffer from performance deficiencies due to domain shift. Here we present a new neural network based image registration framework, called NIR (Neural Image Registration), which is based on optimization but utilizes deep neural networks to model deformations between image pairs. NIR represents the transformation between two images with a continuous function implemented via neural fields, receiving a 3D coordinate as input and outputting the corresponding deformation vector. NIR provides two ways of generating deformation field: directly output a displacement vector field for general deformable registration, or output a velocity vector field and integrate the velocity field to derive the deformation field for diffeomorphic image registration. The optimal registration is discovered by updating the parameters of the neural field via stochastic mini-batch gradient descent. We describe several design choices that facilitate model optimization, including coordinate encoding, sinusoidal activation, coordinate sampling, and intensity sampling. NIR is evaluated on two 3D MR brain scan datasets, demonstrating highly competitive performance in terms of both registration accuracy and regularity. Compared to traditional optimization-based methods, our approach achieves better results in shorter computation times. In addition, our methods exhibit performance on a cross-dataset registration task, compared to the pre-trained learning-based methods.

1. Introduction

3D image registration has a pivotal role in many medical applications (Incoronato et al., 2017; Risholm et al., 2011), such as merging images from different modalities, motion correction, tracking disease progression, and atlas-based image segmentation. Image registration can be categorized into two groups: rigid and non-rigid. Non-rigid registration (also known as deformable registration), considering non-affine coordinate transformations between two images, is more widely used. Diffeomorphic image registration, imposing additional transformation constraints, such as smoothness, invertibility and topology preservation, is often preferred in certain applications. In this paper, we present a new image registration framework that supports both general deformable and specific diffeomorphic image registrations.

Traditional image registration methods (Bajcsy and Kovačič, 1989; Shen and Davatzikos, 2002; Modat et al., 2010; Beg et al., 2005; Avants et al., 2008) approach the problem as an optimization task: finding the optimal coordinate transformations that maximize the similarity between the transformed source image and the target image. These methods usually require hard modeling assumptions on the types of permissible deformations to ensure registration regularity. For instance, NiftyReg (Modat et al., 2010) models deformation fields using B-splines with a set of control points. Flow-based methods model the transformations via a series of time-dependent velocity fields (Beg et al., 2005; Zhang et al., 2017) or stationary velocity fields (Avants et al., 2008), and impose strong assumptions on the space of permissible velocity vector fields. The strong modeling assumptions produce well-behaved transformations, but sometimes also lead to detrimental registration outcomes. To address this, a more flexible framework for modeling permissible transformations is required to improve optimization-based registration. Additionally, these methods can be time-consuming.

Recent advances in deep learning have inspired the development of learning-based image registration methods (Balakrishnan et al., 2019;

* Corresponding author.
E-mail addresses: shanlins@uci.edu (S. Sun), kunh7@uci.edu (K. Han), chenyu.you@yale.edu (C. You), htang6@uci.edu (H. Tang), deyingk@uci.edu (D. Kong), jnaushad@uci.edu (J. Naushad), xiangyy4@uci.edu (X. Yan), haoyum3@uci.edu (H. Ma), pooyak@hs.uci.edu (P. Khosravi), james.duncan@yale.edu (J.S. Duncan), xhx@uci.edu (X. Xie).

https://doi.org/10.1016/j.media.2024.103249
Received 31 August 2022; Received in revised form 24 May 2024; Accepted 21 June 2024
Available online 27 June 2024

Dalca et al., 2018; Mok and Chung, 2020a,b, 2022). The learning-based registration methods are trained to directly output transformations between two images. Although training may take time, predictions are usually generated through a feed-forward model and therefore are very fast. However, in terms of registration accuracy, learning-based methods often lag behind the optimization-based ones under unsupervised settings, even with very complex and large-scale network structures utilized in recent works (Chen et al., 2021a; Mok and Chung, 2022; Shi et al., 2022). Part of the reason is due to the discrepancy between the prediction performances on training data vs. test data. Benefiting from high representational capacity of deep neural networks, learning-based methods can generate high quality transformations between training image pairs, but often generalize poorly on unseen image pairs. Inadequacies in size and diversity of medical datasets accentuate the generalizability issue. To address this issue, recent works (Hering et al., 2021; Siebert et al., 2021; Häger et al., 2020; Zhu et al., 2021) propose a two-step approach. This approach involves using learning models to derive an initial registration, which is then refined using traditional optimization methods.

It is natural to question whether optimization-based registration can also take advantage of the expressive power of deep neural networks. To address this, we propose a framework called **NIR** (Neural Image Registration), which utilizes neural fields to solve medical image registration. Neural fields are a type of neural network, also known as coordinate-based neural multilayer perceptrons (MLPs) or implicit neural representation (INR), that map a point in space and time to a continuous quantity. In a previous study, we demonstrated the effectiveness of neural fields in modeling diffeomorphic transformations for anatomic shape analysis (Sun et al., 2022). This inspired us to use neural fields to model deformable and diffeomorphic registrations between images. NIR provides two ways of modeling image deformations, either directly modeling the displacement vector field or modeling the velocity vector field. In both cases, the neural field within NIR takes as input a 3D coordinate of the source image and outputs a 3D vector (either displacement or velocity) at the location. In the second case, the velocity vector field is further integrated through a Neural Ordinary Differential Equation (ODE) Solver (Chen et al., 2018) to produce the final deformation field, thereby ensuring that the resulting deformation is diffeomorphic.

Modeling deformation fields as coordinate-based MLPs, supplemented with additional features such as Fourier position encoding (Tancik et al., 2020) and periodic activation functions (Sitzmann et al., 2020) in NIR, offers several advantages. First, the neural deformation model is simple and flexible, and yet still has great expressive power. It can use a relatively small number of coefficients to encode signals with an exponentially large frequency support (Yüce et al., 2022). Deformations with high frequencies can be captured by scaling up the number of hidden layers and neurons. Second, neural network can act as "deep prior" (Ulyanov et al., 2018; Gandelsman et al., 2019; Quan et al., 2020; Ren et al., 2020; Williams et al., 2019) in the optimization process. In our neural fields, the weights are shared across the entire space, promoting self-repetition and local similarity in the generated deformations. The "deep prior" of a neural network sometimes is more effective than the explicit smoothness regularization. Fig. 1 demonstrates the advantages of utilizing the "deep prior" of neural networks in deformation representation via a contour registration toy example. Third, different from other neural nets defined on discrete grid coordinates like convolutional neural networks (CNNs), coordinate-based MLPs are defined on the continuous coordinate space. Neural fields can be optimized to model fine deformations with sampled data points and does not require dense input. Consequently, optimizing neural fields is memory-efficient.

The main contributions of our work are summarized as follows:

- We introduce NIR, a novel optimization-based deformable image registration framework that models the displacement field or velocity field via lightweight coordinate-based MLPs with Fourier position encoding and sinusoidal activation functions.
- We further propose a hybrid coordinate sampling scheme, by which two stacked neural fields are separately optimized with two different coordinate samplers to achieve high registration accuracy and regularity.
- NIR is evaluated on two brain MRI datasets and shows competitive registration results in multiple metrics, including intensity similarity between target and transformed moving images, regularity of the transformation.
- Moreover, our framework operates significantly faster than traditional optimization-based methods, while still demonstrating better performance than the pre-trained learning-based methods on a cross-dataset registration experiment.

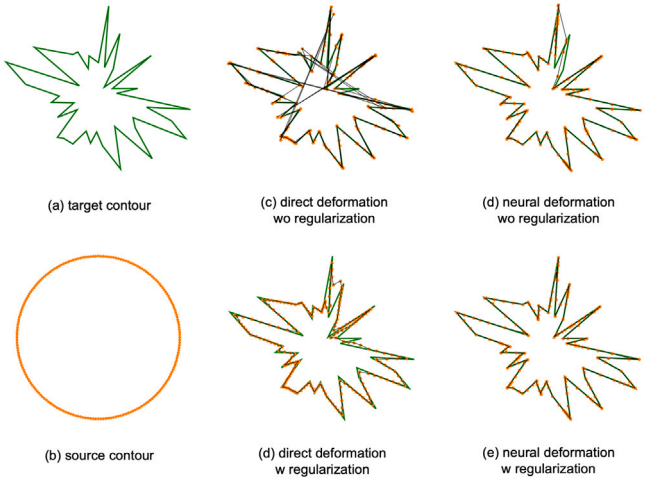


Fig. 1. Contour Registration Toy Example. In this example, the objective of optimization is to deform the circular source contour to align with the target polygonal contour as closely as possible. The results obtained by directly optimizing deformation vectors, with and without regularization, are displayed in (c) and (d), respectively. Conversely, the results shown in (d) and (e) are obtained by representing the (continuous) deformations field through a neural field with and without regularization. The optimization loss is based on the chamfer distance along with normal cosine similarity, and edge length and normal consistency of sampled points can be optionally added as regularization terms. Without a contour topology constraint, optimizing deformation vector fields directly on the source contour can result in a very chaotic deformed contour. However, using neural fields can produce locally smoothed deformation without requiring any regularization. If an explicit contour topology constraint is applied, directly optimized deformation results can be good, but they may also be over-smooth or self-intersected in some regions. In contrast, neural deformation fields can generate nearly perfect contour registration results with high accuracy and regularity if the topology constraint is imposed. This contour registration example demonstrates that neural fields possess a self-prior that enables them to effectively represent deformation maps that exhibit both local smoothness and large deformations, which is the case of brain image registration.

2. Related works

2.1. Optimization-based registration methods

Several studies solve the task of image registration as an optimization problem in the space of displacement vector fields. They optimize the deformable model iteratively with the constraint from a smoothness regularizer which is typically a Gaussian smooth filtering. These include elastic-type models (Bajcsy and Kovačič, 1989), free-form deformation with B-splines (Modat et al., 2010), statistic parametric mapping (Ashburner and Friston, 2000), local affine models (Hellier et al., 2001) and Demons (Thirion, 1998). Diffeomorphic image registration with the attributes of topology preserving and transformation invertibility also achieve remarkable progress in various anatomical studies. Some of the popular methods include Large Diffeomorphic Distance Metric Mapping (LDDMM) (Beg et al., 2005), DARTEL (Ashburner, 2007) and standard symmetric normalization (SyN) (Avants et al., 2008). In this field, the

deformation is modeled by integrating its velocity over time according to the Lagrange transport equation (Christensen et al., 1996; Dupuis et al., 1998) to achieve a global one-to-one smooth and continuous mapping.

2.2. Learning-based registration methods

VoxelMorph (Balakrishnan et al., 2018) utilizes the UNet-like (Ronneberger et al., 2015) structure to directly regress the deformation fields by minimizing the dissimilarity between input and target images. Voxelmorph-diff (Dalca et al., 2019) introduces the diffeomorphic registration and proposes a probabilistic framework. SYM_Net (Mok and Chung, 2020a) provides a symmetric registration method which estimates the forward and backward deformation simultaneously within the space of the diffeomorphic maps. LapIRN (Mok and Chung, 2020b) avoids the local minima of registration in a coarse-to-fine fashion. A recursive cascaded network (Zhao et al., 2019) was proposed to iteratively apply the registration network to the warped moving image and fixed image. DTN (Zhang et al., 2021) deploys a transformer over the CNN backbone to capture the semantic contextual relevance and enhance the extracted feature from backbone. MS-ODENet (Xu et al., 2021) chooses to learn a registration optimizer via a multi-scale neural ODE model and proposes the cross-model similarity metric to alleviate the appearance difference in different contrast levels. Transmorph (Chen et al., 2022) presents a novel image registration by utilizing swin transformer block in the registration framework to identify more precise spatial correspondence. XMorpher (Shi et al., 2022) leverages multi-level semantic correspondence to extract features gradually and enhances the cross attention transformer to facilitate automatic correspondence detection and efficient feature fusion.

2.3. Neural fields for visual computing

2.3.1. Deformation representation

Neural Fields can be used to represent continuous transformation with flexibility. As target geometry and appearance are often modeled with neural fields, it is natural to use neural field to represent the transformation. Niemeyer et al. (2019) performs 4D reconstruction via learned temporal and spatially continuous vector field. Neural Mesh Flow (Gupta, 2020) focuses on generating manifold mesh from images or point clouds via conditional continuous diffeomorphic flow. PointFlow (Yang et al., 2019) incorporates continuous normalizing flows with a principle probabilistic framework to reconstruct 3d point clouds. DiT (Zheng et al., 2021) builds up the dense correspondence across shapes in one category by decomposing DeepSDF (Park et al., 2019) into a deformation network and a single shape representation network.

2.3.2. Medical imaging application

Neural fields have been applied in some medical image analysis tasks, such as 3D image reconstruction or representation. Sun et al. (2021) tries to augment the quantities measured in the sensor domain and reconstructs images with less measurement noise. Shen et al. (2021) predicts the density value at a 3D spatial coordinate, and is supervised by mapping its value back to the sensor domain. Wu et al. (2021) views the 2D slice as the samples from 3D continuous function and reconstructs 3D images from the observed tissue anatomy. NDF (Sun et al., 2022) follows the paradigm of DiT and proposes to model the topology preserving transformation between each organ shape instance and the learned shape template via neural diffeomorphic flow. NeSVoR (Xu et al., 2023) reconstructs a 3D isotropic high-resolution volume from a set of motion-corrupted low-resolution slices with neural fields regressing bias field, volume intensity, and noise variance. Two recent independent works, IDIR (Wolterink et al., 2021) and NODEO (Wu et al., 2022) also proposed optimization-based pair-wise image registration methods utilizing coordinate-based neural networks.

IDIR extends SIREN (Sitzmann et al., 2020) model to optimize the displacement fields of image pairs, thus it is very similar to our simplest displacement-based registration method named NIR-D (refer to Table 1). The primary distinction between IDIR and our NIR-D lies in their coordinate sampling strategy. Specifically, IDIR utilizes random point sampling, but we believe that this approach is not suitable for image registration because it will result in inaccurate image similarity measurements and additional footprint in computing regularization terms, which we will explain in Section 3.4.3 In contrast, we have examined three alternative coordinate sampling methods in our study: downsized sampling, mini-patch sampling, and hybrid sampling. These samplers are specifically designed to accurately measure image similarity and streamline regularization term computation.

NODEO, like our diffeomorphic registration variants, utilizes Neural ODE (Chen et al., 2018) to integrate velocity fields and obtain deformation fields. However, NODEO uses a completely different network architecture. Their neural velocity field is based on a Unet-like 3D CNN model with fully connected bottleneck layers, whereas ours is a simple MLP with coordinate encoding and sinusoidal activation functions. We believe that an MLP-like network is a more appropriate approach to represent complex continuous velocity fields. Firstly, CNN-based networks are primarily designed to capture local features in the input signal, which is a downsampled regular coordinate grid in their case, where neighboring coordinates can be perfectly induced from the center coordinate. Therefore, a CNN-based network cannot provide more than a simple MLP, but with many more parameters to be optimized. Secondly, several theoretical works (Sitzmann et al., 2020; Tancik et al., 2020; Jacot et al., 2018; Yüce et al., 2022) have suggested that using a random Fourier position mapping and a periodic activation function, which we have implemented in our proposed method, can be more effective for modeling neural fields. More importantly, it is crucial to note that continuous velocity fields, which differ from discretized stationary velocity fields (SVFs), cannot be accurately represented by CNN-based models since the diffeomorphic transformation is governed by an ODE that is dependent only on the position and time of the velocity, whereas the output of a CNN is influenced by both its position and its neighbors. As a result, CNN-based networks are not suitable as dynamic functions of Neural ODE for image registration scenarios. We attempted to replicate their results multiple times, but the optimization loss did not converge.

3. Method

3.1. Background

3.1.1. Pairwise image registration

Let $\boldsymbol{T} \in \mathbb{R}^{D\times H\times W}$ and $\boldsymbol{M} \in \mathbb{R}^{D\times H\times W}$ denote the target and moving volumetric images, respectively. Let $\phi : \Omega \subset \mathbb{R}^3 \rightarrow \Omega$ be the deformation field between $\boldsymbol{T}$ and $\boldsymbol{M}$. The unsupervised image registration is commonly formulated as an optimization problem:

$$\hat{\phi} = \arg\min_{\phi} \mathcal{L}(\boldsymbol{T},\ \boldsymbol{M},\ \phi), \tag{1}$$

where the cost function

$$\mathcal{L}(\boldsymbol{T},\ \boldsymbol{M},\ \phi) = \mathcal{L}_{sim}(\boldsymbol{T},\ \boldsymbol{M}\circ\phi) + \lambda_{reg}\cdot\mathcal{L}_{reg}(\phi), \tag{2}$$

includes two terms: (a) $\mathcal{L}_{sim}$, measuring image similarity between the target and warped moving volumes, and (b) $\mathcal{L}_{reg}$, a regularization term on the deformation field. $\boldsymbol{M}\circ\phi$ denotes $\boldsymbol{M}$ warped by the deformation field ϕ. λ_{reg} is a hyperparameter controlling the relative weight of the regularization term .

Registration field ϕ is represented either directly via a displacement field $\mathbf{u}$ with $\phi = \mathbf{Id}+\mathbf{u}$, where $\mathbf{Id}$ is the identity map (Bajcsy and Kovačič, 1989; Balakrishnan et al., 2019), or indirectly via a velocity vector field $\mathbf{v}$, the integration of which leads to ϕ. The second approach is preferred if we require the registration field to be diffeomorphic, i.e., invertible and topology preserving (Beg et al., 2005; Mok and Chung, 2020a).

3.1.2. Neural fields

Both displacement fields and vector fields are modeled by a coordinate-based neural net, referred to as neural field $\mathcal{F}_\theta : \mathbb{R}^3 \to \mathbb{R}^3$, which provides a continuous mapping from 3D coordinate $\boldsymbol{p}$ to the displacement or velocity vector at that position. θ denotes the parameters of the neural net. Neural fields provide a flexible framework for modeling registration field, powerful enough to model highly complex deformations, while maintaining analytic differentiability and allowing us to leverage powerful optimization tools in existing deep learning toolboxes (Frankle and Carbin, 2018).

The neural fields used in this work all consist of a coordinate encoding layer γ, followed by a multilayer perceptron (MLP) whose weights, bias and activation function at the ℓth layer are denoted as $\boldsymbol{W}^{(\ell)}$, $\boldsymbol{b}^{(\ell)}$ and $\rho^{(\ell)}$, respectively. The activities of neurons at each layer are computed sequentially as follows,

$$\begin{aligned} \boldsymbol{z}^{(0)} &= \gamma(\boldsymbol{p}) \\ \boldsymbol{z}^{(\ell)} &= \rho^{(\ell)}\left(\boldsymbol{W}^{(\ell)}\boldsymbol{z}^{(\ell-1)} + \boldsymbol{b}^{(\ell)}\right), \ \ell = 1, \dots, L-1 \\ \mathcal{F}_\theta(\boldsymbol{p}) &= \boldsymbol{W}^{(L)}\boldsymbol{z}^{(L-1)} + \boldsymbol{b}^{(L)}, \end{aligned} \tag{3}$$

where $\boldsymbol{p}$ is the input coordinate and $\mathcal{F}_\theta(\boldsymbol{p})$ denotes the output displacement vector or velocity vector at $\boldsymbol{p}$.

3.2. Overview of NIR

NIR uses neural fields to represent the transformation between two medical images, thus the image registration problem can be solved by optimizing a neural displacement field or neural velocity field. The optimization is solved via stochastic mini-batch gradient descent by finding a stochastic approximation of the objective function Eq. (2) through sampling. This approach contrasts with batch gradient descent, which requires a complete calculation of Eq. (2) and is more memory-demanding.

NIR consists of three main components – Coordinate Sampler (***CS***), Neural Field (***NF***), and Intensity Sampler (***IS***) (Fig. 2). *CS* samples coordinates from the 3D grid points of T, randomly at each step of the optimization. The sampled points are sent to *NF*, which maps position $\boldsymbol{p} \in \mathbb{R}^3$ in the coordinate space of T to position $\boldsymbol{p}' \in \mathbb{R}^3$ in the coordinate space of M. *IS* returns image intensities at query locations of source and target images. Let I_p^T denote the intensity of $\boldsymbol{p}$ on T and $I_{p'}^M$ denote the intensity of $\boldsymbol{p}'$ on M. The sampled image intensities are then used to calculate the similarity loss $\mathcal{L}_{sim}$ (e.g., local normalized cross-correlation loss) between I_p^T and $I_{p'}^M$, as well as the smooth term $\mathcal{L}_{Jdet}$.

The inference mode of NIR is much simpler: the pre-trained neural field takes the whole grid coordinates as input and outputs the deformations at all input coordinates. The warped volume W is then obtained by sampling intensities from the moving volume M given the deformed coordinates.

In Section 3.3, we describe the network design of NF. In Section 3.4, we go over several optimization components, including *CS*, *IS*, and the objective functions. In Section 3.5, we present hybrid coordinate sampling scheme that strikes a balance between registration accuracy and regularity and maintain the optimization efficiency.

3.3. Network design

As illustrated in Fig. 2, NF takes as input a 3D coordinate $\boldsymbol{p} \in \mathbb{R}^3$ in T and outputs the corresponding coordinate $\boldsymbol{p}' \in \mathbb{R}^3$ in M. The transformation from $\boldsymbol{p}$ to $\boldsymbol{p}'$ can be parameterized in two options: (1) use a neural field to directly predict the displacement vector (Fig. 3(a)), or (2) use a neural field to predict the velocity vector, the integral of which leads to the deformation vector (Fig. 3(b)). Both neural displacement field and neural velocity field can be formulated as the Eq. (3) and next we will look into the design of each component in our neural fields.

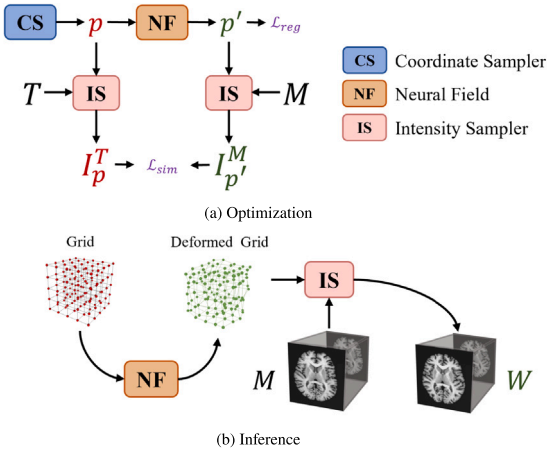


Fig. 2. Overview of NIR, which is a optimization-based pairwise medical image registration framework via neural fields. Plot (b) only presents the transformation of moving volumes via NIR, but the structures associated with the moving volumes can also be transformed in the same way.

3.3.1. Coordinate encoding

Coordinate encoding module maps three-dimensional input coordinates to a higher-dimensional embedding (Mildenhall et al., 2020; Tancik et al., 2020). The mapping can be realized by a family of functionals $e_i : \mathbb{R}^3 \to \mathbb{R}^2$, written as:

$$\gamma(\boldsymbol{p}) = \left[e_1(\boldsymbol{p}), e_2(\boldsymbol{p}), \dots, e_n(\boldsymbol{p})\right] \tag{4}$$

We follow the suggestion from Tancik et al. (2020), encoding coordinates via Fourier mapping, such that

$$e_i(\boldsymbol{p}) = \left[\cos\left(2\pi\boldsymbol{\omega}_i^\top \boldsymbol{p}\right), \sin\left(2\pi\boldsymbol{\omega}_i^\top \boldsymbol{p}\right)\right]^\top, \tag{5}$$

where $\boldsymbol{\omega}_i \in \mathbb{R}^3$ is randomly sampled from a Gaussian distribution with standard deviation σ. The higher the σ, the more likely the model will bias towards the high-frequency signal.

3.3.2. Sinusoidal representation networks (SIRENs)

On top of coordinate encoding layer, the main body of our neural field is a SIREN network (Sitzmann et al., 2020), in which all neurons are activated with sinusoidal functions, i.e., $\rho^{(\ell)} = sin$. Notably, the first layer of SIREN networks can be written as $\boldsymbol{z}^{(1)} = \sin\left(\omega_0\left(\boldsymbol{W}^{(0)}\boldsymbol{z}^{(0)} + \boldsymbol{b}^{(0)}\right)\right)$. Thus, similar to Fourier coordinate mapping, SIRENs can also regulate the spectral bias of the network by adjusting the network hyperparameter ω_0.

Yüce et al. (2022) reveals that the expressive power of coordinate-based MLP with sinusoidal encodings is equivalent to that of a structured signal dictionary, which is restricted to functions that can be expressed as a linear combination of certain harmonics of the coordinate encoding $\gamma(\boldsymbol{p})$. SIREN can be seen as the nested sinusoids and the few coefficients of this network are enough to represent signals with an exponentially large frequency support.

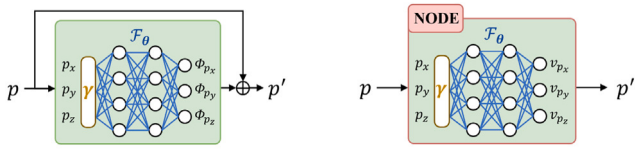
3.3.3. Neural displacement field

Neural displacement field $\mathcal{F}_\theta$ takes as input a 3D location p in T and outputs a displacement vector $\phi_p = [\phi_{p_x}, \phi_{p_y}, \phi_{p_z}]^T = \mathcal{F}_\theta(\boldsymbol{p})$. As a result, the deformed position $\boldsymbol{p}'$ in M is $\boldsymbol{p} + \phi_p$.

3.3.4. Neural velocity field

Under this option, our proposed framework can perform diffeomorphic image registration. Let $\boldsymbol{\Phi}(\boldsymbol{p}, t) : \Omega \subset \mathbb{R}^3 \times [0,1] \mapsto \Omega \subset \mathbb{R}^3$ define a continuous, invertible trajectory from the initial position $\boldsymbol{p} = \boldsymbol{\Phi}(\boldsymbol{p}, 0)$ to the final position $\boldsymbol{p}' = \boldsymbol{\Phi}(\boldsymbol{p}, 1)$, satisfying such ordinary differential equation (ODE) and the initial condition:

$$\frac{\partial \boldsymbol{\Phi}(\boldsymbol{p}, t)}{\partial t} = \boldsymbol{v}(\boldsymbol{\Phi}(\boldsymbol{p}, t), t) \quad \text{s.t.} \quad \boldsymbol{\Phi}(\boldsymbol{p}, 0) = \boldsymbol{p}, \tag{6}$$



(a) Neural Displacement Field (b) Neural Velocity Field

Fig. 3. Neural Fields for Coordinate Deformations – Blue modules indicate the parameters to be optimized. (a) illustrates the neural deformation field that directly transforms the coordinate p in the target volume to the coordinate p' in the moving volume. (b) illustrate the neural velocity field which predicts the stationary velocity vector along the deformation trajectory from p to p'.

where $\boldsymbol{v}(\boldsymbol{p}, t) : \Omega \times [0,1] \mapsto \Omega$ indicates the velocity vector of coordinate $\boldsymbol{p}$ at time t. If $\boldsymbol{v}$ is Lipschitz continuous, a solution to Eq. (6) exists and is unique in the interval $[0,1]$, which ensures that any two deformation trajectories do not cross each other (Dupont et al., 2019). In such discrete cases as Dalca et al. (2018, 2019), the initial value problem (IVP) in Eq. (6) is typically solved with scaling and squaring method (Arsigny et al., 2006).

In this work, we assume that $\boldsymbol{v}$ is continuous and stationary and can be modeled via a neural field, written as $\mathcal{F}_\theta(\boldsymbol{p}) = [\boldsymbol{v}_{p_x}, \boldsymbol{v}_{p_y}, \boldsymbol{v}_{p_z}]^T$. Eq. (6) can be solved with a Differentiable ODE Solver (NODE) (Chen et al., 2018) whose dynamic function is set to be $\mathcal{F}_\theta$. Considering the trade-offs between speed and accuracy, we choose the Fourth-order Runge–Kutta method (rk4) with step size of 0.25 as the ODE solver for our diffeomorohic registration experiments. In the forward pass, the deformed position $\boldsymbol{p}'$ of position $\boldsymbol{p}$ can be estimated by integrating $\mathcal{F}_\theta(\boldsymbol{p})$ from $t = 0$ to $t = 1$ via NODE, formulated as

$$\boldsymbol{p}' = \boldsymbol{\Phi}(\boldsymbol{p}, 1) = \boldsymbol{\Phi}(\boldsymbol{p}, 0) + \int_0^1 \mathcal{F}_\theta(\boldsymbol{\Phi}(\boldsymbol{p}, t))\mathrm{d}t \tag{7}$$

For backpropagation, NODE adopts the adjoint sensitivity method (Pontryagin, 1987), which retrieves the gradient by solving the adjoint ODE backwards in time and allows solving with O(1) memory usage no matter how many steps the ODE solver takes.

3.4. Optimization

In this section, we will introduce the intensity sampler, objective functions as well as coordinate sampler used in our NIR.

3.4.1. Intensity sampler

To utilize gradient-based optimization method, a differentiable intensity sampler is required to estimates the intensities of sub-voxel positions given source images. Same as Jaderberg et al. (2015), Balakrishnan et al. (2019), Mok and Chung (2020a,b), we apply linear interpolation (other interpolation methods can also be applied) as intensity sampler, referred as $\mathcal{IS}^{\text{linear}}$. Given a coordinate $\boldsymbol{c}$ and scans S, the intensity value at $\boldsymbol{c}$, referred to as I_c^S, is obtained based on the intensities of the eight surrounding voxels.

3.4.2. Objective functions

Local normalized cross-correlation is adopted to measure the intensity similarity. Let $\bar{I}_c^S$ denote the intensity mean of local region centering at position c on volume S. In our experiments, $\bar{I}_c^S = \frac{\sum_{c_i} I_{c_i}^S}{w^3}$, where $\boldsymbol{c}_i$ iterates over the local region in the size of w^3. Then local normalized cross-correlation can be defined as below:

$$\mathrm{LNCC}(T, M, \boldsymbol{p}, \boldsymbol{p}') = \frac{\left[\sum_{p_i, p_i'} (I_{p_i}^T - \bar{I}_p^T)(I_{p_i'}^M - \bar{I}_{p'}^M)\right]^2}{\left[\sum_{p_i} (I_{p_i}^T - \bar{I}_p^T)^2\right]\left[\sum_{p_i'} (I_{p_i'}^M - \bar{I}_{p'}^M)^2\right]}, \tag{8}$$

where $\boldsymbol{p}$ denotes the sampled position in the coordinate of target volume T, and $\boldsymbol{p}'$ denotes deformed position in the coordinate of moving volume M.

As for the regularization term, we follow Mok and Chung (2020a) to impose the Jacobian determinant penalty on the predicted deformation field. The Jacobian matrix of the deformation field ϕ at a position $\boldsymbol{p}$ is notated as $J_\phi(\boldsymbol{p})$, where N is the toal number of sampled locations per optimization iteration.

If $|J_\phi(\boldsymbol{p})|$ is positive, it is suggested that the deformation field preserves the local orientation near $\boldsymbol{p}$. Conversely, if $|J_\phi(\boldsymbol{p})|$ is negative, the deformation field reverses the local orientation around $\boldsymbol{p}$. Thus, the local orientation consistency constraint can be defined as

$$\mathrm{LOCC}(\boldsymbol{p}) = \max(0, -|J_\phi(\boldsymbol{p})|), \tag{9}$$

which only penalizes the regions with negative Jacobian determinants. In our experiment, $J_\phi(\boldsymbol{c})$ is approximated as the differences between neighboring deformation vectors.

The intensity similarity $\mathcal{L}_{sim}$ and the regularization term $\mathcal{L}_{reg}$ is the mean value of negative local normalized cross-correlation and local orientation consistency constraint across sampled positions, written as

$$\mathcal{L}_{sim} = \frac{1}{N}\sum_j -\mathrm{LNCC}(T, M, \boldsymbol{p}_j, \boldsymbol{p}_j') \tag{10}$$

$$\mathcal{L}_{reg} = \frac{1}{N}\sum_j \mathrm{LOCC}(\boldsymbol{p}_j) \tag{11}$$

Here, N is the toal number of sampled locations per optimization iteration and $\boldsymbol{p}_j$ denotes the jth location sampled in one batch.

3.4.3. Coordinate sampler

To optimize the parameters of our neural fields, we apply mini-batch stochastic gradient descent method. In other words, we sample a subset of coordinates of the whole image grid to update the model parameters per iteration in optimization. Next, we will discuss three different coordinate samplers: random sampler, downsize sampler and mini-patch sampler.

Random Sampler (Fig. 4(a)) is most commonly used in coordinate-based neural networks (Sitzmann et al., 2020; Niemeyer et al., 2019; Chen et al., 2021b) because the coordinates sampled via a random sampler are distributed across the whole grid and the unbiased sampled coordinates allow for the more stable optimization. But random coordinate sampler is inapplicable in our case. To compute $LNCC$, we need to search closest coordinates among all sampled coordinates to estimate the local intensity mean and correlation, whose consequence is that the optimization speed can be significantly impeded by the searching time. Moreover, randomly sampling coordinate will bring about larger memory consumption for calculating $LOCC$. As we mentioned in Section 3.4.2, we approximate the Jacobian matrix of the deformation field by discretizing the image coordinate space, asking for the coordinates to be sampled in a spatial regularity. If the sampled coordinates are distributed randomly, the Jacobian matrix requires extra memory for the second-order derivatives of deformation field with respect to model parameters during optimization. After all, considering the time and memory deficiency, random coordinate sampler is an impractical choice for our NIR.

Downsize Sampler samples coordinates with specific step size in each dimension as shown in Fig. 4(b). Coordinates sampled by downsize sampler can well cover the entire image coordinate space but the approximation of Jacobian matrix might be of more flaws due to downsizing. The consequence is, the neural fields optimized via downsize coordinate sampler achieve great alignment accuracy but relatively bad local orientation consistency in deformations.

Mini-Patch Sampler randomly selects multiple high-resolution small coordinate blocks as shown in Fig. 4(c). Compared to downsize coordinate sampler, it can provide more accurate Jacobian matrix approximation but the drawback lies in the computation of local normalized cross-correlation. Specifically, the extensive padding operations along

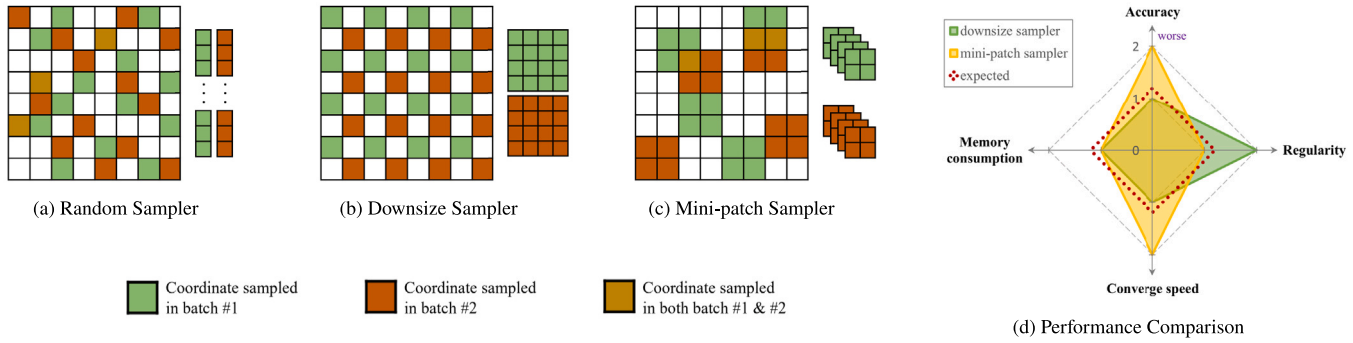


Fig. 4. Coordinate Samplers and Performance Comparisons. (a), (b) and (c) illustrate sampling 16 coordinates per batch from total 64 2D coordinates with three kinds of coordinate samplers. (d) **ranks** (not quantifies) the registration performance of NIR models optimized with two practical coordinate samplers (downsize sampler and mini-patch sampler) in four aspects. The higher ranking in each dimension indicates better performance in that aspect. As is shown in (d), consuming almost the same GPU memory during optimization, compared to NIR optimized with the mini-patch sampler, NIR optimized with the downsize sampler can take less time to converge to a more accurate registration results with more violations in topology preserving. The expected solution, as indicated by the red-dot line, should be of great performance in both registration accuracy and regularity with no or modestly extra computations. For the numerical results supporting the ranking in plot (d), please refer to Table 4.

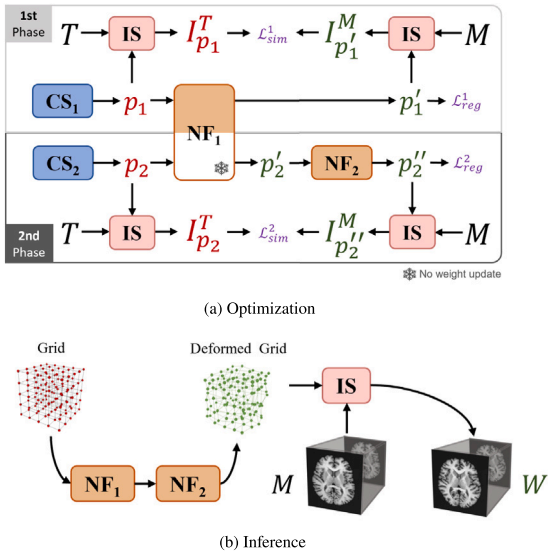


Fig. 5. Overview of NIR with Hybrid Coordinate Sampling Scheme. The optimization is composed of two phases, in which two neural fields (NF_1 and NF_2) are optimized separately. During inference, NIR with hybrid coordinate sampler requires grid coordinates to pass through two neural fields in sequence to get the deformed coordinates.

the patch borders result in the inaccurate local normalized cross-correlation. Thus, the neural fields optimized via mini-patch coordinate sampler are good at registration regularity but bad at alignment accuracy.

Fig. 4(d) demonstrates the rank of the registration performance of two candidate coordinate samplers with the spatial regularity in four criteria — accuracy, regularity, memory consumption, and converge speed. It is apparent in Fig. 4(d) that no coordinate sampling strategy can outperform the others in all criteria. Downsize coordinate sampler is good at criteria in all aspects but the registration regularity, which happens to be the strength of mini-patch coordinate sampler. The expected solution should have high registration accuracy, minor distortions in the deformation field, rapid converge rate as well as little memory consumption, as indicated by the red-dot region in Fig. 4(d).

3.5. NIR with hybrid coordinate sampler

3.5.1. Overview

We intend to enhance the complementarity of the downsize and mini-patch coordinate samplers without the substantial increase in memory and time consumption during optimization. To this end, we propose a hybrid coordinate sampler which performs two different coordinate sampling strategies in two phases of optimization. As shown in Fig. 5, NIR with a hybrid coordinate sampler consists of two concatenated neural fields optimized separately. The first neural field (NF_1) takes in charge of the rough alignment between the moving and target scans and the residual transformation is completed by another neural field (NF_2). In inference, NF_1 and NF_2 deform the whole grid in cascade, which means the output of NF_1 is taken as the input of NF_2 and then NF_2 outputs the final deformed grid. As for optimization, the parameters of NF_1 and NF_2 are updated with the downsize sampler CS_1 and mini-patch sampler CS_2 separately in two phases as depicted in Fig. 5(a).

3.5.2. Optimization

Hybrid coordinate sampler is motivated by three experimental observations indicated by Table 4. (1) the downsize sampler can generate more accurate registration in the price of more distortions in the deformation field; (2) the mini-patch sampler tends to provide over-smooth deformation fields and results in much slower convergence speed; and (3) in the early stage of optimization, the regularity of the deformation field from NIR optimized with the downsize sampler is well-preserved while achieving decent registration accuracy. This could be attributed to the spectral bias of neural fields, which tends to reconstruct lower-frequency signals at the beginning of optimization. The hybrid coordinate sampler leverages these observations by initializing the registration field with the downsize sampler and fine-tuning the results with the mini-patch sampler, achieving a balance between registration accuracy and regularity.

To be specific, the hybrid sampling method is conducted in two phases. In the first phase, NF_1 is optimized with the downsize coordinate sampler CS_1 for a short time. After the first-phase optimization, NF_1 is able to generate the smooth and relatively accurate transformation. Then the goal of the second-phase optimization is to let NF_2 complete the transformation left unfinished by NF_1. We prefer a neural registration field that can align the more detailed structures and does not mess up the underlying topology in the second phase of optimization. For this reason, the input coordinate p_2 for the second-phase optimization are sampled by the mini-patch sampler CS_2 and the initial deformed positions p_2'' for NF_2 are predicted by NF_1. Notably,

taking optimization stability and memory efficiency into account, only NF_2 is optimized in the second-phrase optimization while weights of NF_1 are frozen.

Similar to NIR, the optimization objective functions of NIR with a hybrid sampler is given by:

$$\mathcal{L} = w_1 \cdot (\mathcal{L}^1_{sim} + \lambda_{reg} \cdot \mathcal{L}^1_{reg}) + w_2 \cdot (\mathcal{L}^2_{sim} + \lambda_{reg} \cdot \mathcal{L}^2_{reg}), \tag{12}$$

where $w_1 = 1, w_2 = 0$ in the first phase of optimization and $w_1 = 0, w_2 = 1$ in the second phase. $\mathcal{L}^1_{sim}$, $\mathcal{L}^2_{sim}$, $\mathcal{L}^1_{reg}$ and $\mathcal{L}^2_{reg}$ follows the same definition as introduced in Section 3.4.2.

4. Experiments

4.1. Dataset

All our experiments are conducted on two public 3D brain MR datasets — Mindboggle101 and OASIS.

Mindboggle101 (Klein and Tourville, 2012) is a dataset consisting of 101 T1-weighted MRI scans of healthy individuals from 5 different data sources. Among these scans, 45 are designated as testing data, with 5 serving as moving scans and 40 as target scans. The training dataset comprises 47 scans, while the validation data consists of the remaining scans. For evaluation purposes, we selected 31 cortical regions as outlined in Xu and Niethammer (2019).

OASIS dataset (Marcus et al., 2007) consists of 416 T1-weighted MR images of subjects aged 18 to 96, including individuals with early-stage Alzheimer's Disease (AD). Hoopes et al. (2021) annotated 35 anatomical structures in this dataset, and we use 27 of them for performance evaluation in our experiments. Same as Wu et al. (2022), we select 45 scans, with 40 as target scans and 5 as moving scans for testing. We use 250 of the remaining scans for training the learning-based methods in our comparisons, and select 30 scans for validation.

All MRI images utilized in our experiments undergo the same preprocessing procedures, which include skull stripping, resampling to $1\ \text{mm} \times 1\ \text{mm} \times 1\ \text{mm}$ spacings, affine alignment to the MNI template of T1-weighted MRI imaging (Fonov et al., 2009, 2011), and cropping to a size of $160 \times 192 \times 144$. As all images are already aligned to the MNI template, our experiments focus on the non-linear deformation between pairs of images.

4.2. Implementation details

Network Architecture: The neural displacement field in Section 3.3.3 is composed of four fully connected layers with a hidden feature size of 256. However, due to the computational inefficiency of Neural ODE, we adopt shallower SIREN models for the neural velocity field in Section 3.3.4. This field consists of three fully connected layers with a hidden feature size of 256. The activation scale ω_0 of the first layer of the SIREN model for both neural displacement and velocity field is set to 30 for all experiments. Additionally, regardless of the chosen neural field and coordinate samplers, the dimension of the coordinate embedding and the standard deviation σ of the Gaussian distribution are set to 128 and 3, respectively.

Coordinate Sampler: The down-sampling step size for the downsize coordinate sampler is set to 3 in all dimensions, while the mini-patch coordinate samplers randomly select 5 patches per optimization iteration, with each patch having a size of $32 \times 32 \times 32$.

Objective Function: In computing the local normalized cross-correlation ($LNCC$) for the coordinates sampled from the downsize sampler, a local region size of 9 is set, while a local region size of 27 is used for mini-patches sampling. Throughout the optimization process, the regularization weight λ_{reg} of 1000 is used for our displacement-based deformable registration methods, while our diffeomorphic registration methods use a regularization weight of 100.

Optimization: The network parameters are updated using the Adam optimizer (Kingma and Ba, 2014) with a learning rate of $1e^{-4}$. For NIR, the maximum number of optimization iterations is set to be 900. For NIRs with a hybrid coordinate sampler (NIR-H and NIR-H-Diff), the first phase is optimized for 200 iterations, and the second phase is further optimized for 900 iterations.

Platform: All optimization-based methods are executed on a system equipped with NVIDIA GTX 2080Ti GPUs and an Intel i7-7700K CPU. All learning-based methods are running on a system equipped with NVIDIA A6000 GPUs and an Intel i7-7700K CPU. Across different platforms, we maintained consistency in the versions of software packages, managed via Anaconda (Anon, 2020).

Table 1
Names of options under NIR framework.

Name	Coordinate sampler	Neural field type
NIR-D(-Diff)	Downsize	Displacement (Velocity)
NIR-P(-Diff)	Mini-patch	Displacement (Velocity)
NIR-H(-Diff)	Hybrid	Displacement (Velocity)

4.3. Methods in comparisons

In Section 3.1.2, we have introduced two types of neural fields for the displacement-based deformable registration and diffeomorphic registration respectively, both of which can be integrated into the framework of NIR (Section 3.3) and NIR with a hybrid coordinate sampler (Section 3.5). We provide several options of running NIR and their names are listed in Table 1. The baseline models encompass both learning-based approaches, including VoxelMorph (Balakrishnan et al., 2019), VoxelMorph-diff (Dalca et al., 2019), SYM_Net (Mok and Chung, 2020a), TransMorph (Chen et al., 2022), XMorpher (Shi et al., 2022) and SynthMorph (Hoffmann et al., 2021), as well as optimization-based methods like SyN (Avants et al., 2008), NiftyReg (Modat et al., 2010), IDIR (Wolterink et al., 2021) and NODEO (Wu et al., 2022), and Grid, which can be regarded as a discretized counterpart to our NIR-D. Further elaboration on the training and optimization procedures for these baseline models in our experiments is available in the appendix.

4.4. Experimental setup

In all of our medical image registration experiments, the objective is to transform a moving volume to match a target volume. If the moving volume has associated structure labels, we can map these structures onto the target volume using the transformation obtained from the registration task. Our evaluation metrics (Section 4.5) for the registration results include the similarity between the target and warped volumes/structures, as well as the local orientation consistency of the deformation fields.

In this paper, we conduct three groups of unsupervised registration experiments.

Experiment (1): We adhere to the data splitting protocol as described in Section 4.1. In particular, we leveraged the pre-trained models[1] from the IXI dataset,[2] which comprises 576 brain MR scans. The pre-trained models are available for all learning-based methods, except XMorpher which lacks a pre-trained model. Therefore, we train XMorpher from scratch. For other methods that do provide pre-trained models, we fine-tune them using our target dataset. We proceed to compare the performance of our proposed NIR framework against both learning-based and traditional optimization-based methods on Mindboggle dataset.

[1] https://github.com/junyuchen245/TransMorph_Transformer_for_Medical_Image_Registration/tree/main/IXI/

[2] https://brain-development.org/ixi-dataset/

Table 2
Registration performance comparison on Mindboggle101 dataset and OASIS dataset.

Category	Experiment	(1) Mindboggle101			GPU
	method/metrics	$DSC_s^{(1mm)}$ (↑)	DSC_v (↑)	$\mathbf{J}_{\leq 0}$ (↓)	memory (MB)
Learning-based (pre-trained)	VoxelMorph	0.7075 (0.021)*	0.4753 (0.019)*	2.41e−02 (2.17e−03)*	4985
	VoxelMorph-Diff	0.7057 (0.024)*	0.4692 (0.023)*	1.59e−05 (1.37e−05)*	2923
	SYM_Net	0.7662 (0.020)*	0.5582 (0.019)*	1.28e−04 (2.61e−05)*	3031
	TransMorph	0.7663 (0.020)*	0.5434 (0.022)*	2.20e−02 (2.02e−03)*	8055
	TransMorph-Diff	0.6255 (0.021)*	0.4585 (0.021)*	2.14e−05 (1.69e−05)*	4389
	SynthMorph	0.7358 (0.023)*	0.5149 (0.024)*	3.46e−03 (5.31e−04)*	4985
Learning-based (finetuned)	VoxelMorph	0.7473 (0.022)*	0.5464 (0.024)*	2.26e−02 (2.83e−03)*	10 129
	VoxelMorph-Diff	0.7568 (0.021)*	0.5417 (0.022)*	4.03e−06 (2.16e−05)*	5121
	SYM_Net	0.7725 (0.021)*	0.5763 (0.021)*	5.78e−05 (1.45e−05)*	10 565
	TransMorph	0.7868 (0.020)*	**0.6002 (0.020)**	8.11e−04 (8.64e−04)*	17 203
	TransMorph-Diff	0.6819 (0.023)*	0.5363 (0.022)*	2.95e−06 (5.72e−06)*	9681
	X-Morpher	0.7634 (0.022)*	0.5584 (0.021)*	5.83e−4 (5.36e−4)*	24 459
Optimization-based	NiftyReg	0.7874 (0.024)*	0.5635 (0.026)*	1.01e−03 (9.21e−04)*	–
	SyN	0.7822 (0.019)*	0.5514 (0.020)*	4.40e−06 (4.91e−06)*	–
	Grid	0.7063 (0.022)*	0.5145 (0.023)*	9.61e−04 (2.73e−04)*	4981
	IDIR	0.6986 (0.032)*	0.4819 (0.035)*	**0 (0)**	5183
	NODEO	–	–	–	3863
Ours	NIR-H	0.7809 (0.020)*	0.5561 (0.021)*	1.31e−04 (4.75e−05)*	3341
	NIR-H-Diff	**0.7904 (0.020)**	0.5826 (0.021)	**1.11e−06 (7.84e−07)**	3177
Category	Experiment	(2) OASIS			Running
	method/metrics	$DSC_s^{(1mm)}$ (↑)	DSC_v (↑)	$\mathbf{J}_{\leq 0}$ (↓)	time (s)
Learning-based (pre-trained)	VoxelMorph	0.8412 (0.051)*	0.7873 (0.039)*	1.57e−03 (3.38e−03)*	–
	VoxelMorph-Diff	0.8257 (0.065)*	0.7820 (0.038)*	3.71e−06 (1.63e−06)*	–
	SYM_Net	0.9031 (0.037)	0.8302 (0.026)*	4.10e−05 (1.29e−05)*	–
	TransMorph	0.8865 (0.038)*	0.8218 (0.027)*	2.66e−02 (5.73e−03)*	–
	TransMorph-Diff	0.7141 (0.055)*	0.7667 (0.034)*	4.33e−06 (9.43e−06)*	–
	SynthMorph	0.8782 (0.033)*	0.8191 (0.023)*	1.83e−04 (6.47e−05)*	–
Learning-based (finetuned)	VoxelMorph	0.8962 (0.035)*	0.8274 (0.024)*	5.63e−03 (2.07e−03)*	< 2.0
	VoxelMorph-Diff	0.8913 (0.039)*	0.8197 (0.026)*	6.93e−06 (1.18e−05)*	< 2.5
	SYM_Net	0.9050 (0.028)	0.8397 (0.020)	7.51e−05 (1.84e−05)*	< 3.0
	TransMorph	**0.9164 (0.025)**	**0.8506 (0.020)**	1.41e−02 (3.31e−03)*	< 2.5
	TransMorph-Diff	0.8118 (0.039)*	0.8270 (0.019)*	7.84e−06 (9.47e−06)*	< 3.0
	XMorpher	0.8636 (0.026)*	0.8173 (0.021)*	3.52e−03(2.42e−04)*	< 4.0
Optimization-based	NiftyReg	0.8905 (0.048)*	0.8234 (0.035)*	1.28e−03 (8.99e−04)*	≈ 2521
	SyN	0.9058 (0.026)	0.8371 (0.020)	6.38e−06 (8.49e−06)	≈ 1273
	Grid	0.7897 (0.092)*	0.7613 (0.045)*	8.04e−04 (1.65e−04)*	≈ 3969
	IDIR	0.8042 (0.036)*	0.7748 (0.034)*	**0 (0)**	≈ 3969
	NODEO	–	< 0.783 (-)	3.0e−04 (-)	≈ 80
Ours	NIR-H	0.8984 (0.032)*	0.8274 (0.023)*	1.62e−04 (7.18e−05)*	≈ 90
	NIR-H-Diff	0.9071 (0.034)	0.8382 (0.025)	4.55e−06 (1.05e−05)	≈ 640

The GPU memory consumption for the learning-based methods are "training consumption — inference consumption", but for our proposed methods, are just maximum memory consumption during optimization. The GPU memory consumption in optimizing the hybrid NIR models varies in two phases because the numbers of sampled coordinates in two phases are different. Specifically, the number of coordinates sampled by the downsize sampler and the mini-patch sampler in two phases are 165 888 and 163 840, respectively. Thus, the maximum GPU memory consumption for hybrid NIR models comes from the first phase of optimization. In the case of pre-trained learning-based methods, the GPU memory consumption reflects the cost of inference, whereas in the case of fine-tuned learning-based methods, the GPU memory consumption reflects the cost of training.

Experiment (2): Similar as Experiment (1), we compare our NIR framework with other methods on OASIS dataset under the data split setting from Section 4.1.

Experiment (3): Experiment 3 is designed to assess and compare the generalization capability of various methods for handling image registration between pairs of scans from different datasets. Specifically, 60 pairs of test data are created by matching three randomly selected moving scans of healthy brains from the Mindboggle101 test set with 20 target volumes randomly chosen from the OASIS dataset, which comprises patients with moderate Alzheimer's disease and a Clinical Dementia Rating (CDR) exceeding 0.5. The pre-trained learning-based methods are fine-tuned on the Mindboggle101 training set and incorporate moderate data augmentations, such as "RandGaussianNoise", "RandScaleIntensity", "RandAdjustContrast" and "RandAxisFlip" techniques, implemented by the MONAI (Cardoso et al., 2022) package, to enhance their generalization performance.

The first two experiments aim to evaluate the effectiveness of our proposed methods in brain MRI registration. The third experiment is designed to test the robustness of our optimization-based methods.

To determine the optimal values of certain hyperparameters, such as the learning rate and regularization weight, for our proposed NIR framework, we randomly select 15 image pairs from nine Mindboggle101 validation data to create our validation set.

4.5. Evaluation metrics

All methods in our comparison aim to register the moving volumes to the target volumes. If the moving and target volumes belong to the same dataset, we evaluate the registration performance using the Dice coefficient (DSC) and the ratio of coordinates with a non-positive Jacobian determinant ($\mathbf{J}_{\leq 0}$). However, if the moving and target volumes have different annotations, we use the Structural Similarity Index ($SSIM$) and $\mathbf{J}_{\leq 0}$ for evaluation.

Two types of DSC – volumetric DSC_v and surface DSC_s, are used to evaluate the overlap between two regions. Given the target mask $\mathcal{M}_T$ and warped mask $\mathcal{M}_W$, the volumetric DSC_v can be computed as $DSC_v = \frac{2|\mathcal{M}_T \cap \mathcal{M}_W|}{|\mathcal{M}_T| + |\mathcal{M}_W|}$. According to Reinke et al. (2021), volumetric

measurements may yield similar evaluation scores even if the regions have vastly different shapes, especially in complex structures such as cortical regions. Therefore, boundary-based measures are preferred in such cases. Specifically, we use the surface Dice coefficient (DSC_s) introduced by Nikolov et al. (2018) to assess the alignment accuracy. Unlike volumetric DSC_v, the surface DSC_s evaluates the overlap of two surfaces within a specific tolerance τ., formulated as

$$DSC_s^{(\tau)} = \frac{\left|S_T \cap \mathcal{B}_W^{(\tau)}\right| + \left|S_W \cap \mathcal{B}_T^{(\tau)}\right|}{|S_T| + |S_W|}, \tag{13}$$

where S_i refers to the surfaces of mask $\mathcal{M}_i$, and $\mathcal{B}_i^\tau$ denotes the border regions for the surface S_i within a tolerance τ, which is 1 mm in our experiments. For more details about the surface DSC, please look into Nikolov et al. (2018). Both volumetric and surface DSC range from 0 to 1 and higher score represents better registration accuracy. The final reported scores are the average DSC of all structures over all pairs.

$\mathbf{J}_{\leq 0}$ is a metric that assesses the regularity of deformation fields, measured as the ratio of coordinates with non-positive Jacobian determinant. The Jacobian matrix represents the derivatives of the deformations and reflects the local properties of the deformation field. Only the local regions with a positive Jacobian determinant are transformed in a way that preserves topology and invertibility. Thus, a higher $\mathbf{J}_{\leq 0}$ indicates poorer registration regularity. The calculation of the Jacobian matrix of deformations is described in Section 3.4.2.

The Structural Similarity Index ($SSIM$) (Wang et al., 2004) is a metric that calculates the similarity between two images based on three components: luminance, contrast, and structure. The $SSIM$ value ranges between 0 and 1, where higher values indicate greater similarity between the image pairs. For further information regarding the calculation details, please refer to the original paper (Wang et al., 2004).

In order to evaluate the performance of our proposed NIR-H-Diff method, we will report the mean and standard deviation of all the evaluation metrics used in the experiments in the following tables. In addition, we will perform a paired samples t-test for experiments (1), (2), and (3) between the proposed NIR-H-Diff method and all other comparative methods, and calculate two-sided p-values. A p-value of less than 0.001 will indicate that the scores of NIR-H-Diff and the comparator methods are significantly different and we will mark the scores with an asterisk (*) when our NIR-H-Diff performs significantly better than the corresponding methods on that metric.

4.6. Quantitative comparisons with baselines

In this section, we compare the registration performance of our proposed methods with that of the baseline methods. The reported results of the NIR variants in Tables 2 and 3 were obtained after 900 iterations of optimization.

Table 2 provides a comparison of the performance among all methods in experiments (1) and (2) in terms of registration accuracy, registration regularity, and maximum GPU memory consumption during optimization and inference. The box plots in Fig. 6 demonstrate the Dice's Coefficient of six representative methods for different groups of anatomical structures in experiments (1) and (2). Our proposed methods demonstrate highly competitive performance regarding registration accuracy and regularity while consuming significantly less GPU memory than training or fine-tuning learning-based methods. In experiment (1), NIR-H-Diff outperforms all pre-trained learning-based models across all three evaluation metrics and performs comparably to the finetuned TransMorph in terms of DSC_v. Additionally, our method exhibits superior registration regularity when compared to other competing methods. It can be seen from Fig. 6(a) our advantages in alignment accuracy applied to almost all annotated structure groups. In experiment (2), our methods continue to outperform pre-trained learning-based models in almost every metric, except for SYM-Net on $DSC_s^{(1mm)}$. Even though the finetuned TransMorph and SYM_Net performs better than us in terms of registration accuracy, NIR-H-Diff is significantly superior to other finetuned learning-based models in all evaluation metrics. Based on the results presented in Table 2, we observed that the performance advantage of our proposed methods over the learning-based methods decreased in experiment (2). This outcome may be attributed to the availability of more data for training, all of which were collected from the same institution and followed similar scanning protocols in the OASIS dataset, thereby limiting the exposure of the generalization issue of the learning-based methods in this experiment.

Table 3
Registration performance comparison on experiment (3).

Method	$SSIM$ (↑)	$\mathbf{J}_{\leq 0}$ (↓)
VoxelMorph	0.7459 (0.0072)*	9.68e−02 (3.55e−03)*
VoxelMorph-Diff	0.7471 (0.0061)*	7.09e−06 (1.43e−05)*
SYM_Net	0.7799 (0.0051)*	7.31e−05 (1.82e−05)*
TransMorph	0.7907 (0.0074)*	6.23e−02 (5.68e−03)*
TransMorph-Diff	0.7469 (0.0065)*	7.65e−06 (1.51e−05)*
XMorpher	0.7738 (0.0073)*	1.45e−03(4.23e−03)*
SynthMorph	0.8051 (0.0052)*	7.68e−03 (4.91e−03)*
NiftyReg	0.8389 (0.0057)*	1.36e−03 (8.23e−04)*
SyN	0.8478 (0.0043)*	5.80e−06 (7.16e−06)*
NIR-H	0.8408 (0.0052)*	2.30e−04 (8.91e−05)*
NIR-H-Diff	**0.8530 (0.0034)**	**2.28e−06 (3.71e−06)**

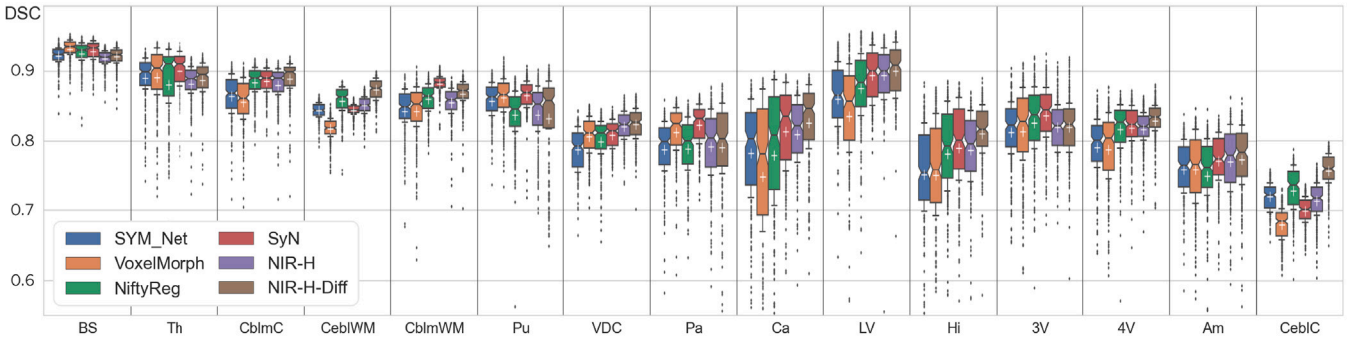
3 target volumes from Mindboggle101 dataset and 20 moving volumes from OASIS dataset are randomly selected to conduct the cross-dataset image registration experiments. The learning-based methods are trained with the training set of Mindboggle101 dataset.

Compared with the optimization-based registration methods (NiftyReg and SyN), our NIR-H-Diff and NIR-H can both provide high-accuracy registration performance but only NIR-H-Diff achieve the top performance in terms of registration regularity. NIR-H is not comparable with diffeomorphic registration method, i.e., SyN, in the metric of $\mathbf{J}_{\leq 0}$. but it only generates about 1/10 folds in deformation fields compared with NiftyReg. Another important criterion to assess the optimization-based methods is the performance relationship with optimization duration, which will be discussed in Section 4.9. Fig. 6(a) and Fig. 6(b) present a closer inspection of registration accuracy of methods in comparison, from which we can tell that our proposed method can achieve the better performance than the other optimization methods in 9 out of 12 structure groups in experiment (1) and 9 out of 15 structure groups in experiment (2).

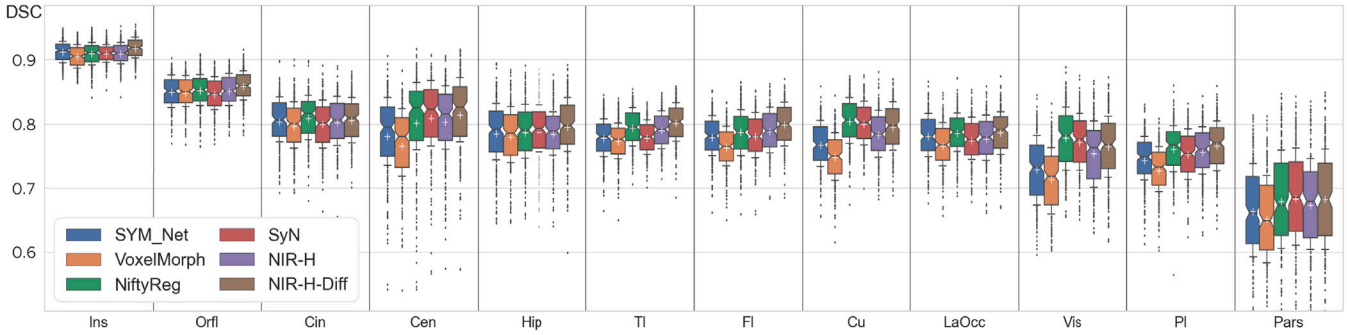
Compared with Grid, our NIR-H is significantly better in terms of registration accuracy and regularity, consuming less GPU memory. Because Grid and NIR-H share a similar optimization process but mainly differ in the ways to describe the deformation fields, the significant advantage of our proposed method may suggest the effectiveness of neural fields in modeling the deformation fields.

Compared with other neural field based methods, specifically IDIR and NODEO, our proposed techniques exhibit notable advantages in terms of registration accuracy. IDIR, which operates as a displacement-based registration method, demonstrates commendable performance concerning $\mathbf{J}_{\leq 0}$ by enforcing deformation smoothness through automatic differentiation. Nevertheless, this achievement is accompanied by significantly higher GPU memory consumption and notably inferior registration accuracy when contrasted with our NIR-H and NIR-H-Diff methods. Regarding NODEO, while their official implementation remains unreleased, making it challenging to reproduce their results, we have managed to perform a preliminary performance comparison based on the reported outcomes, as discussed in Appendix A.2. This comparison has underscored a substantial gap between our methods and NODEO in terms of accuracy and regularity.

Table 3 compares the performance of NIR-H, NIR-H-Diff with six learning-based and two optimization-based methods in experiment (3). As the moving volumes are healthy scans from the Mindboggle101



(a) **Experiment (1)**. We group all 31 structures into 12 groups: Cin (caudal anterior cingulate, rostral anterior cingulate, isthmus cingulate, posterior cingulate), Fl (caudal middle frontal, rostral middle frontal, superior frontal), Pl (inferior parietal, superior parietal, supramarginal), Tl (inferior temporal, middle temporal, superior temporal, transverse temporal), Orfl (lateral orbitofrontal, medial orbitofrontal), LaOcc (lateral occipital), Cen (postcentral, precentral, paracentral), Cu (cuneus, precuneus), Pars (pars opercularis, pars orbitalis, pars triangularis), Hip (entorhinal, parahippocampal), Vis (lingual, fusiform, pericalcarine).



(b) **Experiment (2).** The abbreviations above indicate: brain stem (BS), thalamus (Th), cerebellum cortex (CblmC), cerebral white matter (CeblWM), cerebellum white matter (CblmWM), putamen (Pu), Ventral-DC (VDC), Pallidum (Pa), Caudate (Ca), Lateral Ventricle (LV), Hippocampus (Hi), 3rd Ventricle (3V), 4th Ventricle (4V), Amygdala (Am), and Cerebral Cortex (CeblC).

Fig. 6. Boxplots of Dice's Coefficients for Various Anatomical Structures. Left and right hemispheres are combined together, e.g. we average two Dice's Coefficients of the left and right pair of anatomical structures into one. The white '+'s in the above boxes indicate the average Dice's Coefficients.

dataset and 20 target scans with Alzheimer's disease come from the OASIS dataset, the results in experiment (3) might suggest the robustness of an algorithm against modest domain shift. It is apparent from Table 3 that, compared with optimization-based methods, the learning-based methods learned from the one dataset cannot as well generalize to pair of images coming from different datasets with different health status. Our NIR-H-Diff method can significantly outperform all learning-based, including SynthMorph, as well as optimization-based methods, in the metric of both $SSIM$ and $\mathbf{J}_{\leq 0}$.

4.7. Qualitative comparisons with baselines

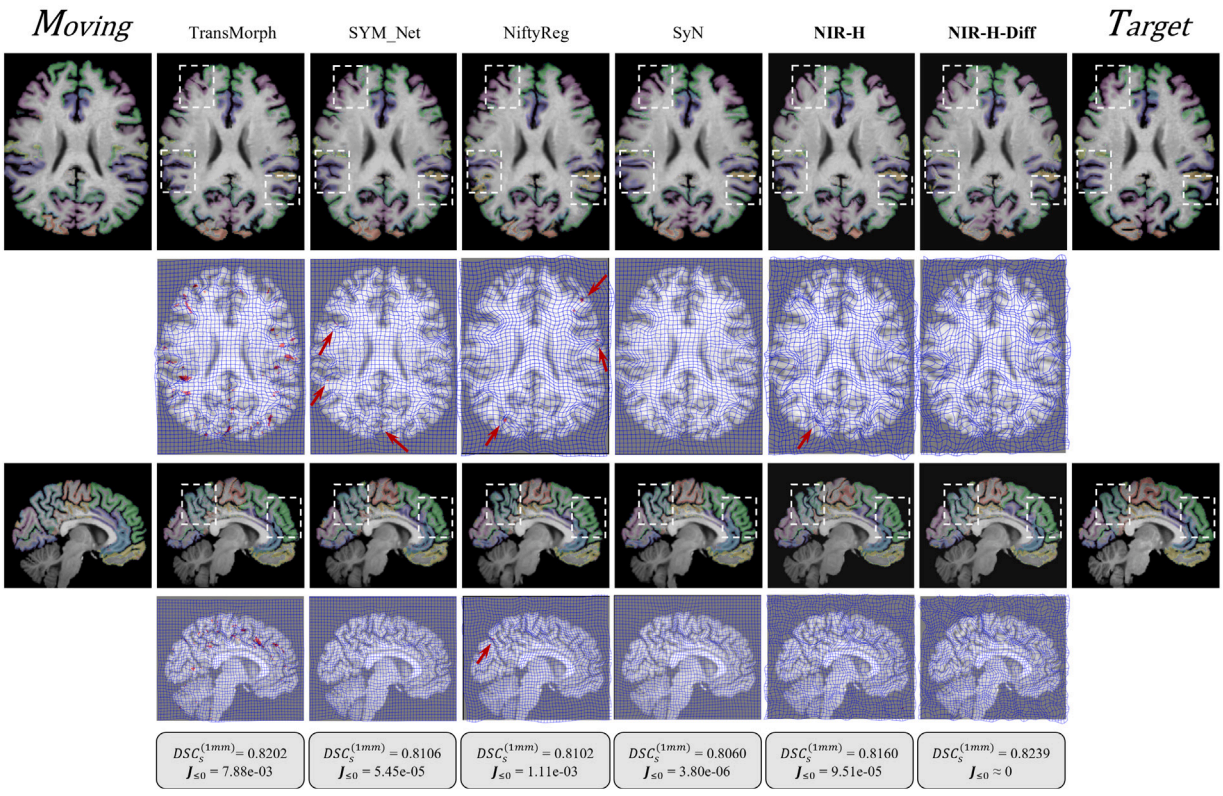
Fig. 7 presents the qualitative comparisons between our proposed and selected benchmarks. Fig. 7 shows that all registration methods are able to align the subcortical regions well, while differences between methods are mainly observed in the cortex regions (indicated by white dotted boxes). NIR-H-Diff and TransMorph exhibit substantially better performance in these regions, but TransMorph achieves great accuracy at the expense of much more foldings in the deformation map. NIR-H-Diff and other methods such as SYM_Net and SyN support diffeomorphic registration, but are not able to generate large deformations when needed. In contrast, our proposed NIR-H-Diff can generate sharp deformation maps without violating diffeomorphism.

4.8. Influence of coordinate samplers

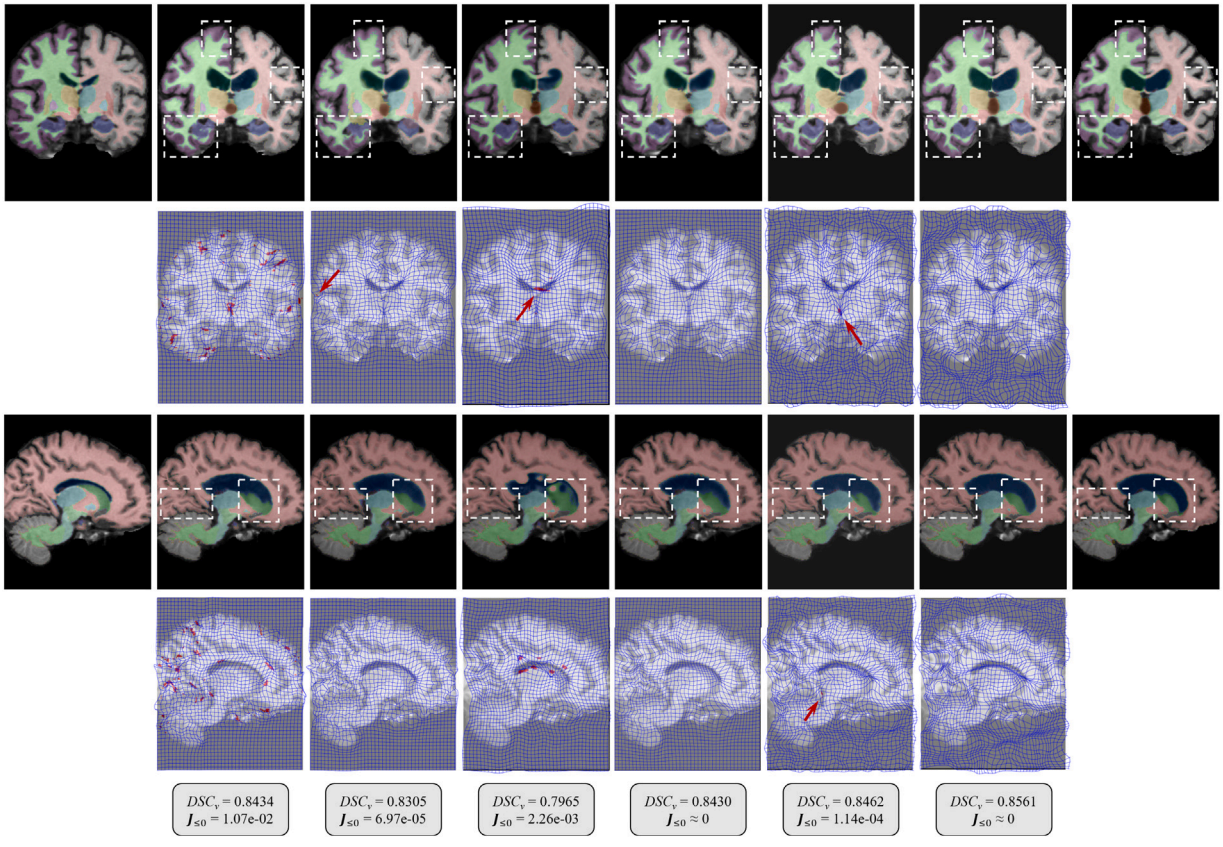
In this section, we investigate the effects of coordinate samplers on the registration results. These experiments are performed on the testing set, and the evaluation is based on $DSC_s^{(1mm)}$ and $\mathbf{J}_{\leq 0}$.

In Fig. 8, we illustrate the impact of different coordinate samplers on the registration performance of a single pair of test data from OASIS datset. The downsize sampler can generate more accurate registration in the price of more distortions in the deformation field while the mini-patch sampler tends to provide the over-smooth deformation fields and results in much slower convergence speed. We also noticed that, in the very early stage of optimization, the regularity of deformation field from NIR optimized with the downsize sampler is well-preserved and at the same time, registration accuracy is quite decent. But as optimization time grows, the fraction of positions with a negative Jacobian determinant increases a lot. To further validate our analysis and design decisions, we present a quantitative comparison. Table 4 shows the diffeomorphic registration performance differences resulting from the choice of coordinate samplers. The table reveals a discernible pattern where the registration accuracy of both NIR-D-Diff and NIR-P-Diff increases over time, but the registration regularity deteriorates. Additionally, the results suggest that NIR-D-Diff can achieve higher registration accuracy more rapidly than NIR-P-Diff, but NIR-P-Diff is capable of sustaining a very low $\mathbf{J}_{\leq 0}$ during the entire optimization process, while NIR-D-Diff is not.

NIR-H-Diff exhibits superior registration regularity compared to NIR-D-Diff and NIR-P-Diff, while maintaining comparable registration accuracy to NIR-D-Diff, without requiring significantly more memory for optimization. It should be noted that the iteration number for NIR-H-Diff in Table 4 refers to the second phase of optimization, during which NIR-H-Diff requires 200 additional iterations compared to the other two methods. Nonetheless, our design achieves our goal of developing an efficient method that can quickly converge to high-quality registration results with good topology preservation.



(a) Experiment (1)



(b) Experiment (2)

Fig. 7. Qualitative Registration Performance Comparison of Different Methods. The models in qualitative comparison are TransMorph, SyM_Net, NiftyReg, NIR-H and NIR-H-Diff. In the above plots, we present two volume pairs from experiment (1) and experiment (2) in two views. The warped volumes generated by different methods are overlapped with the warped structures which are indicated by colors. The key differences in registration quality of different methods are highlighted by the white dotted boxes. The deformation fields are illustrated by the downsized deformed grid in blue and the regions with negative jacobian determinant are colored in red. The last row in Fig. (a) and (b) are the quantitative performance of different registration methods on that image pair. If the $\mathbf{J}_{\leq 0}$ is less than $1e-06$, we take it as ≈ 0.

Table 4
Registration performance differences resulting from coordinate samplers.

Metrics	$DSC_s^{(1mm)}$ (↑)				$\mathbf{J}_{\leq 0}$ (↓)				GPU Memory
method/iteration	100	300	600	900	100	300	600	900	(MB)
NIR-D-Diff	0.7750 (0.022)	**0.7906 (0.021)**	**0.7917** (0.021)	**0.7929** (0.021)	1.75e−06 (1.48e−06)	5.75e−05 (6.36e−06)	1.41e−04 (2.96e−05)	2.04e−04 (7.59e−05)	3177
NIR-P-Diff	0.6875 (0.025)	0.7396 (0.022)	0.7669 (0.019)	0.7792 (0.019)	**0 (0)**	**4.69e−08 (1.64e−08)**	1.13e−06 (7.03e−07)	3.14e−06 (1.19e−06)	**3149**
NIR-H-Diff	**0.7865 (0.022)**	0.7893 (0.021)	0.7897 **(0.020)**	0.7908 **(0.020)**	2.26e−06 (1.42e−06)	8.59e−07 (4.21e−07)	**9.07e−07 5.07e−07**	**1.12e−06 (7.34e−05)**	3177

The table below shows the comparisons of diffeomorphic NIR frameworks optimized via the downsize sampler (NIR-D-Diff), mini-patch sampler (NIR-P-Diff), and hybrid diffeomorphic NIR (NIR-H-Diff). NIR-D-Diff and NIR-P-Diff solely employ NF1 while NIR-H-Diff uses both NF1 and NF2. The comparisons are based on the registration accuracy ($DSC_s^{(1mm)}$), registration regularity ($\mathbf{J}_{\leq 0}$) and converge speed (Iteration). The iteration number of NIR-H-Diff is that of the second phase of optimization. This table supports the qualitative comparisons of different coordinate samplers as shown in Fig. 4(d).

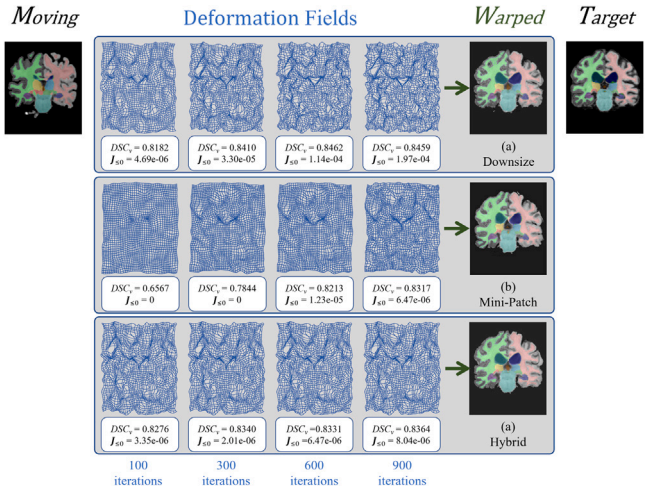


Fig. 8. Comparison of different coordinate samplers (a), (b) and (c) are registration results of NIR optimized with the downsize sampler, mini-patch sampler and hybrid NIR. The above image pair are 'OASIS_OAS1_0001_MR1' (T) and 'OASIS_OAS1_0002_MR1' (M) from the OASIS dataset and we present the registration results over the optimization iterations, generated by the differomorphic NIR. DSC and $J_{\leq 0}$ are the evaluation metrics for registration accuracy and regularity separately.

4.9. Optimization duration v.s. Registration performance

Fig. 9 presents the relationship between registration performance and optimization duration of six optimization-based registration methods in experiment (2), four of which are our proposed methods and the other two are SyN and NiftyReg. It is worth noting that NIR-Diff, NIR-H-Diff and SyN generate diffeomorphic registration fields, while NIR-D, NIR-H and NiftyReg provide deformable registration fields.

As for our proposed methods, we evaluate their registration performance at {100, 300, 600, 900} optimization iterations. To finish 100-iteration optimization, the displacement-based NIR methods take about 9s and the diffeomorphic NIR methods take about 64 s. It needs to be clarified that the optimization iteration of NIR-H and NIR-H-Diff counts from the start of the second-phase optimization.

For SyN, we assessed the registration performance by setting the maximum optimization iteration to {8, 4, 2}, {20, 10, 5}, {60, 30, 15}, and {100, 50, 25} at each level. The average optimization time for each corresponding iteration is approximately 117 s, 261 s, 829 s, and 1273s. Regarding NiftyReg, we evaluated its registration performance with the maximum optimization iteration set to {120, 60, 30}, {400, 200, 100}, {800, 400, 200}, and {1200, 600, 300} at each level. The average optimization time for each corresponding iteration is approximately 309 s, 901 s, 1638s, and 2521s.

Among all methods in comparison, NIR-D has the fastest converge speed and the highest DSC_v (0.8435). However, it also has the highest $\mathbf{J}_{\leq 0}$ (1.08e−03), indicating that it may struggle with preserving topology during the registration process. NIR-H trades off some registration accuracy for improved registration regularity compared to NIR-D. However, as shown in Fig. 9, extending the optimization duration for NIR-H has the potential to improve both DSC_v and $\mathbf{J}_{\leq 0}$.

NiftyReg exhibits relatively poor performance in terms of converge speed, registration accuracy, and registration regularity. This is because the official implementation of NiftyReg has disabled GPU acceleration, making the optimization process using $LNCC$ similarity very time-consuming. As a result, NiftyReg is even slower than methods that support diffeomorphic transformations.

NIR-D-Diff can achieve a decent registration accuracy ($DSC_v \geq 0.83$) within a short time period of approximately 200 s. However, the registration regularity deteriorates as the number of optimization iterations increases. NIR-H-Diff, on the other hand, aims to strike a better balance between registration accuracy and regularity. It achieves similar DSC_v compared to NIR-D-Diff but with significantly greater regularity of deformation fields. Specifically, $\mathbf{J}_{\leq 0}$ remains below 5e-06 during optimization. In experiments (2), SyN demonstrates very strong performance, especially in terms of registration regularity. However, as shown in Fig. 9, our approaches have two main advantages over SyN. Firstly, NIR-D-Diff and NIR-H-Diff can achieve higher DSC_v scores than SyN when optimized for a similar duration. Secondly, as optimization iterations increase in the finer scale, SyN exhibits significantly worse registration regularity and ends with a higher $\mathbf{J}_{\leq 0}$ compared to our NIR-H-Diff.

5. Limitations and future directions

One major limitation of NIR is its running time. Although significantly faster than traditional optimization-based methods, it is still slower than learning-based methods. There are a few potential approaches to address this limitation. Firstly, an adaptive coordinate sampler can be designed that samples coordinates sparsely in regions with easy alignment and densely in regions with large alignment errors. Secondly, NIR can be used in combination with a learning-based method in a two-step approach, where the learning-based method generates an initial registration, followed by fine-tuning through NIR. Third, neural fields can also be integrated into a learning-based framework (Sun et al., 2022; Park et al., 2019; Zheng et al., 2021), where the coordinate-based MLPs and an embedding layer are learned from the training data. During inference, the parameters of coordinate-based MLPs are fixed and merely a latent code associated with the test data is optimized.

In addition, how to introduce surface registration into our image registration framework is a topic worth exploring. NIR establishes correspondence between image pairs to match voxel intensities. It is agnostic to anatomic structures within the images and thus does not always lead to semantically meaningful registrations. One future direction in this regard is to optimize both intensity and shape similarities between the two images. Since shape registration can also be realized

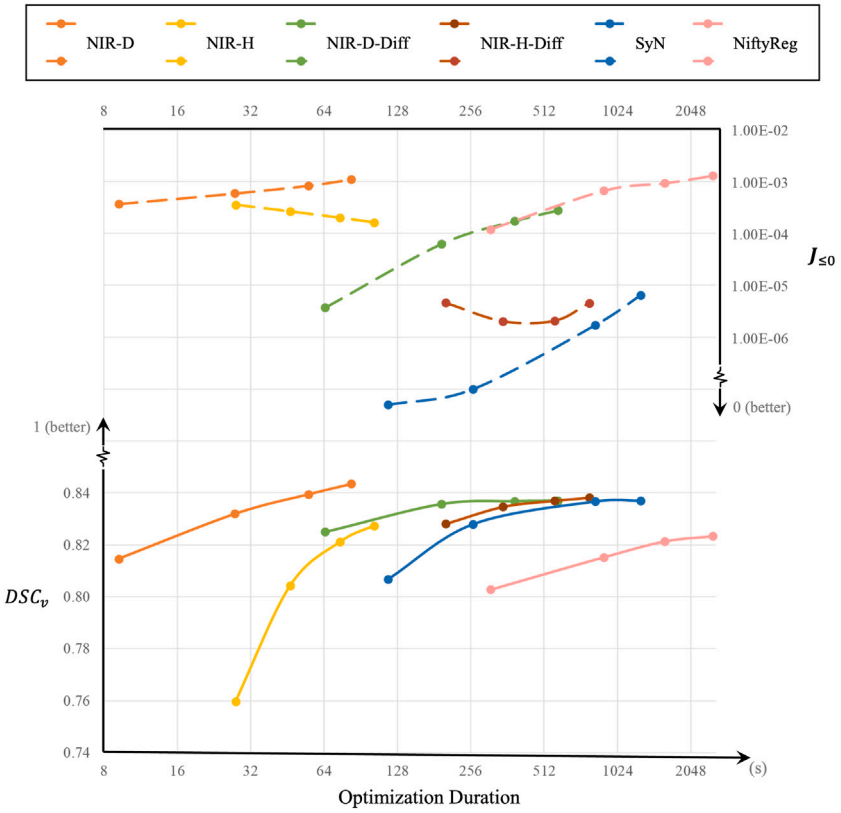


Fig. 9. Optimization Duration v.s. Registration Performance in Experiment (2). The solid and dotted curves respectively illustrate the change of registration accuracy and regularity over optimization duration. In the bottom half of this plot, the higher a solid curve goes, the better registration accuracy it indicates. While in the top half this plot, the lower a dotted curve goes, the better registration regularity it reflects. Thus, visually speaking, a method is preferred if its solid and dotted curves get close over time.

via neural fields as we showed previously (Sun et al., 2022), neural fields provide a promising approach to unify both intensity-based and shape-based registrations within the same framework.

6. Conclusions

We introduce a new optimization-based framework, named NIR, for deformable image registration. NIR employs coordinate-based MLPs with Fourier position encoding and sinusoidal activation functions to model deformation vector fields. The method utilizes the full power of existing deep learning toolboxes to solve the optimization efficiently and demonstrates higher generalizability compared to previous learning-based methods.

We present several options for running NIR, depending on the type of registration (displacement-based or diffeomorphic) and speed requirements. (a) **NIR-D**: the fastest displacement-based deformable registration method with good registration accuracy; (b) **NIR-H**: a rapid displacement-based deformable registration method with a better registration regularity compared to NIR-D; (c) **NIR-D-Diff**: a diffeomorphic registration method with a good registration accuracy and regularity; and (d) **NIR-H-Diff**: a slightly slower diffeomorphic registration method with the best overall performance.

We assess the performance of our methods on two brain MRI datasets against multiple benchmarks and show that they achieve highly competitive results in terms of registration accuracy and regularity. Our approaches outperform traditional optimization-based methods within shorter computation times. Furthermore, our methods do not require significant GPU resources for training and exhibit superior performance on the cross-dataset registration task.

CRediT authorship contribution statement

Shanlin Sun: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Kun Han:** Writing – review & editing, Visualization, Validation, Software, Investigation, Formal analysis. **Chenyu You:** Writing – review & editing, Validation, Investigation, Formal analysis, Conceptualization. **Hao Tang:** Writing – review & editing, Visualization, Validation, Investigation. **Deying Kong:** Writing – review & editing, Methodology, Conceptualization. **Junayed Naushad:** Writing – review & editing, Validation, Formal analysis. **Xiangyi Yan:** Writing – review & editing, Visualization, Validation. **Haoyu Ma:** Visualization, Validation. **Pooya Khosravi:** Writing – review & editing. **James S. Duncan:** Project administration. **Xiaohui Xie:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This project is partially supported by grants from the National Science Foundation (IIS-1715017, DMS-1763272) and UCI IPH Pilot Award.

Appendix. Experiments

A.1. Pre-training setup

Some learning-based registration models were pre-trained on T1-weighted brain MRI scans from the IXI dataset for atlas-to-patient registration. Similar to the preprocessing steps used for the OASIS and Mindboggle datasets, these images were processed using FreeSurfer for skull stripping and affine alignment to MNI305 atlas. The processed images have dimensions of $160 \times 192 \times 144$ and their intensity values are normalized between 0 and 1. The dataset includes a total of 576 MRI volumes, with 403 used for training, 58 for validation, and 115 for testing. The atlas MRI volume is obtained from CycleMorph (Kim et al., 2021). This preprocessed dataset can be downloaded from the TransMorph open-source repository, where the pre-trained model weights of VoxelMorph, VoxelMorph-Diff, TransMorph, and TransMorph-Diff models are also available. Additionally, we pre-trained another baseline model, SYM_Net, on the IXI dataset, following the same data split as the other models.

A.2. Methods in comparisons

VoxelMorph. is a widely recognized learning-based technique that allows for deformable 3D medical image registration between pairs of images. This method employs a large training dataset to learn the desired deformation, rather than optimizing the objective function for each image pair, which can be time-consuming. In our experiments, we adopted the original setup and trained VoxelMorph using $LNCC$ as the similarity metric. We utilized the larger variant with one extra convolutional layer at the output resolution and more channels for later layers, which showed better performance. We also applied a regularization term with a weight of 1 to ensure the smoothness of the predicted displacement field.

VoxelMorph-diff. introduces diffeomorphic registration, which ensures that the transformation between images or surfaces is smooth and invertible. The authors propose a probabilistic framework for diffeomorphic registration, which treats registration as a probabilistic inference problem. In our experiments, we used the optimal settings, with a value of 0.01 for σ and a value of 10 for λ.

SYM_Net. is a learning-based technique that offers diffeomorphic deformable image registration. SYM_Net is a symmetric method that maximizes image similarity within the space of diffeomorphic deformation and estimates the forward and backward transformation simultaneously. In our experiments, we trained SYM_Net using $LNCC$ and the parameter values recommended in the original paper. Specifically, we set the weights for penalizing the negative jacobian determinant, enforcing the smoothness of the velocity field, and constraining the bias for the bidirectional velocity field to 1000, 3, and 0.1, respectively.

Transmorph. presents a novel image registration approach by introducing a transformer structure, which enables more precise identification of the spatial correspondence between fixed and moving images. The authors offer both a deformable registration model (Transmorph) and a diffeomorphic registration model (Transmorph-diff). In our experiments, we trained both models with a hyperparameter value of 1 for the smoothness of the deformation field, as recommended by the authors.

XMorpher. leverages multi-level semantic correspondence to extract features gradually, enabling effective registration. The method also enhances the Cross Attention Transformer (CAT) blocks by establishing an attention mechanism between images to facilitate automatic correspondence detection and efficient feature fusion. In our experiments, we trained XMorpher using the same weights for the image similarity and regularization term as recommended in the original paper.

SynthMorph. introduces a strategy for learning image registration without acquired imaging data, producing powerful networks agnostic to contrast introduced by MRI. We include this methods into experiment (3) and directly used their pre-trained "brains" variant,[3] which is trained using images synthesized from brain label maps, including 26 the largest brain labels of 40 distinct-subject segmentations with brain and non-brain labels from T1w MPRAGE scans of the Buckner40 dataset and a subset of the fMRIDC structural data (Fischl et al., 2002; Van Horn et al., 2001). The "brain" variant of SynthMorph is typically better than the "shape" variant, which is trained using images synthesized from random shapes only.

SyN. is a popular diffeomorphic registration method and we applied the implementation by DIPY (Garyfallidis et al., 2014) with careful parameter tuning on the validation set. Because our test image paris in SyN are in the same modality (T1-weighted), we used NCC as the metric, where sampling radius and the standard deviation of the Gaussian smoothing kernel to be 4 and 2. Since SyN is a iterative-based approach, we set the maximum iteration to {100, 50, 25} for each level to balance the tradeoff between registration accuracy and running time. We did not used their official implementation in the Advanced Normalization Tools (ANTs) (Avants et al., 2009) because it will take over one hour to register one pair of images with CC as the metric with only {40, 20} optimization iterations at two scales, which is inhibitive to practical application.

NiftyReg. is the fast free-form deformation algorithm for non-rigid registration. In our experiments, cubic B-spline interpolation is used to deform moving volumes to optimize $LNCC$ image similarity with the squared Jacobian determinant log as a penalty term. The standard deviation of the Gaussian kernel and the weight of the penalty term are set to be 40 and 0.01 separately, as suggested by Shen et al. (2019). In addition, three scales are used in optimization with the maximum optimization iterations as {1200, 600, 300} for each scale.

IDIR. is also a neural field based optimization method, which randomly samples the same number of coordinates in one iteration. We adjusted the regularization strength of IDIR from its default value 10 to 0.01 for higher registration accuracy.

NODEO. is another neural field based optimization method, which represents the dynamic functions of NODE with 3D Unet. Since they have not released their official implementation, We attempted to replicate their results multiple times, but the optimization loss never converged. Our methods utilized the identical data partitioning scheme as reported in their study for the OASIS dataset. While NODEO reported dice scores across 28 structures, we computed them across 27 structures since they additionally annotated CSF for the OASIS dataset. NODEO reported a mean dice score of 0.779 on the OASIS dataset, and from Fig. 7 of their paper, we can infer that the dice score for CSF was greater than 0.65. Thus, the maximum value of mean dice score across 27 structures for NODEO shall be less than (0.779 * 28 - 0.65) / 27 = 0.783. Also, Negative Jacobian ratio, GPU memory and time consumption can also be found in their paper.

Grid. method treated the displacement vectors of all grid coordinates as independent variables, which are optimized by the same similarity measurement ($LNCC$), regularization term ($LOCC$), and gradient descent optimizer (Adam) as NIR-H uses. 'Grid' applies the multi-scale optimization strategy. At coarser scales, we downsized the original moving and target images and optimized a low-resolution deformation field. At finer scales, we upsampled the moving and target images and initialized the higher-resolution deformation field with the results from the lower-resolution optimization. In practice, we optimized the displacement field at three scales, ranging from a lowest resolution of $40 \times 48 \times 52$ to a highest resolution of $160 \times 192 \times 144$. The regularization weights and iteration numbers were {100, 1000, 2000} and {100, 400, 600}, respectively.

[3] https://github.com/voxelmorph/voxelmorph/tree/dev/data

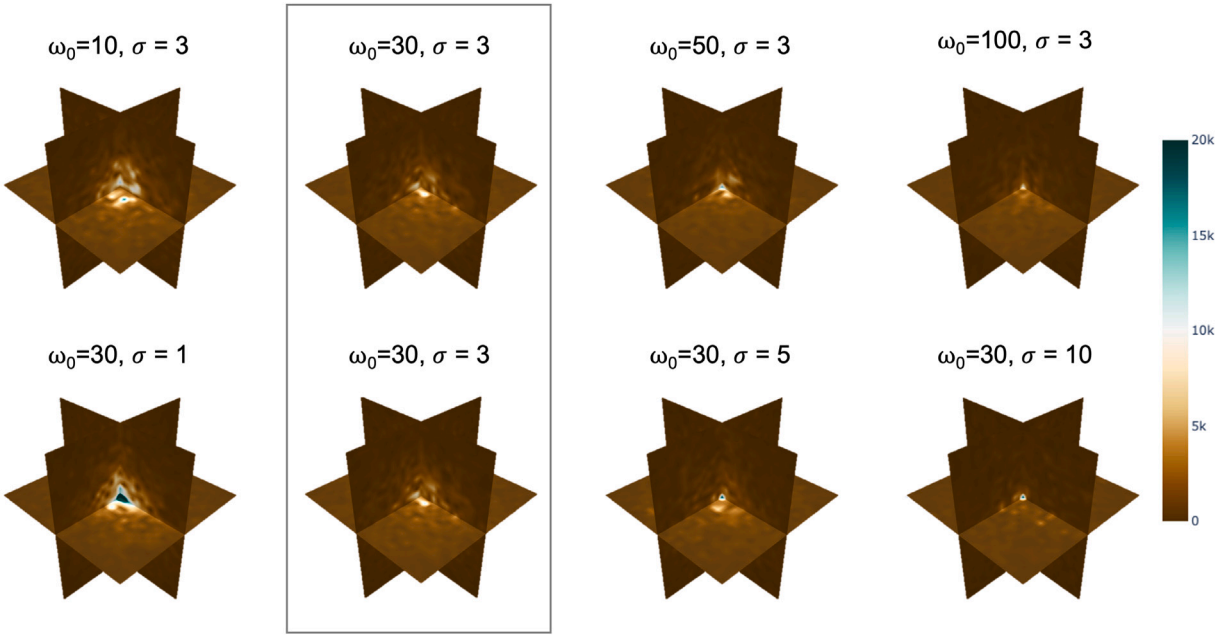


Fig. A.10. Shifted fourier transform of registration fields with different frequency priors ω_0 and σ. The above visualization provides evidence that ω_0 and σ can modulate the frequency of neural deformation fields. In all our experiments, we applied $\omega_0 = 30$ and $\sigma = 3$.

Table A.5
Influence of frequency priors ω_0 and σ.

ω_0	σ	$DSC_s^{(1mm)}$ (↑)	$\mathbf{J}_{\leq 0}$ (↓)
30	1	0.8024 (0.080)	6.64e−05 (1.16e−05)
30	5	0.7824 (0.087)	2.63e−04 (3.60e−05)
30	10	0.7526 (0.089)	1.48e−03 (3.08e−04)
10	3	0.7897 (0.082)	**3.01e−05 (8.96e−06)**
50	3	0.7941 (0.082)	1.80e−04 (2.14e−05)
100	3	0.7789 (0.087)	2.96e−04 (4.59e−05)
30	**3**	**0.8033 (0.081)**	1.27e−04 (1.45e−05)

Increasing values of ω_0 and σ result in more folding in the warping map. However, when ω_0 and σ are either too large or too small, the registration accuracy (as measured by $DSC_s^{(1mm)}$) drops.

Table A.6
Influence of regularity weight λ_{jdet}.

Method	λ_{jdet}	$DSC_s^{(1mm)}$ (↑)	$\mathbf{J}_{\leq 0}$ (↓)
NIR-D-Diff	100	**0.7929 (0.021)**	2.04e−04 (7.59e−05)
	1000	0.7902 (0.021)	3.96e−05 (8.43e−06)
	10 000	0.7856 (0.022)	8.05e−06 (3.26e−06)
NIR-P-Diff	10	0.7823 (0.020)	2.63e−05 (2.44e−05)
	100	0.7792 (0.021)	3.14e−06 (2.03e−05)
NIR-H-Diff	100	0.7908 (0.020)	**1.12e−06 (7.34e−07)**

The effect of λ_{jdet} in the balance of $DSC_s^{(1mm)}$ and $\mathbf{J}_{\leq 0}$ can be observed on NIR-D-Diff and NIR-P-Diff. However, by merely adjusting the scale of λ_{jdet}, both methods cannot outperform NIR-H-Diff in terms of registration accuracy and regularity.

A.3. Influence of frequency priors ω_0 and σ

In our proposed methods, the inclusion of frequency information is crucial to effectively represent complex deformation with a high level of flexibility. We present two approaches for manipulating the frequency of the registration fields in our proposed methods. The first method involves using sine activation functions with a scale factor ω_0 as described in Section 3.3.2. The second method involves using positional embedding through Fourier feature mapping with a factor σ as described in Section 3.3.1. The choice of ω_0 and σ is determined through a hyperparameter sweep on the validation set, and the results of NIR-D-Diff using different frequency priors are presented in Table A.5. To ensure label-agnostic hyperparameter selection, we chose not to apply label-based registration accuracy metrics such as volumetric dice score and surface dice score in our experiments. As indicated in Table A.5, increasing ω_0 and σ leads to more folding in the registration fields. However, excessively large or small ω_0 and σ values result in decreased registration accuracy.

Fig. A.10 illustrates the shifted Fourier Transform of registration fields with different parameters, providing evidence that ω_0 and σ can modulate the frequency of neural representation. The first row of Fig. A.10 shows that registration fields generated by neural fields with smaller ω_0 have more low-frequency components, represented by the central region in the Fourier domain. Similarly, in the second row of Fig. A.10, the registration fields generated by neural fields with smaller σ contain more low-frequency components. Thus, we can expect that a neural field will generate smoother deformations when it has smaller ω_0 and σ. In our experiments, we chose $\omega_0 = 30$ and $\sigma = 3$ as they provide good registration accuracy with only modest irregularity of deformations.

A.4. Influence of regularization weight λ_{reg}

We perform an investigation on the impact of adjusting the regularization weight on the performance of NIR with a hybrid coordinate sampler. The comparison is conducted among the three diffeomorphic variants of NIR, namely NIR-D-Diff, NIR-P-Diff, and NIR-H-Diff, on the validation set from the Mindboggle101 dataset. The results of this investigation are presented in Table A.6. Our experiments aimed to improve the registration regularity of NIR-D-Diff and registration accuracy of NIR-P-Diff by adjusting the weight of the regularization term λ_{reg} in

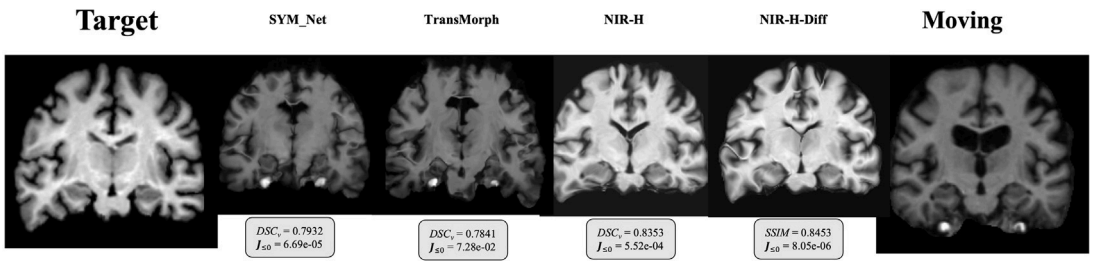


Fig. A.11. Visual Comparison for Experiment (3). The target image comes from a patient with moderate Alzheimer's disease from OASIS dataset and the moving image comes from a healthy person. The colormap of different images covers the complete value range of the given data. In this case, high-intensity artery structures result in overall dark image brightness.

Table A.7
Influence of MRI contrast shift.

	M-M	M-O^h
VoxelMorph	0.8260 (0.0065)	0.8217 (0.0073)
SYM_Net	0.8451 (0.0058)	0.8392 (0.0055)
TransMorph	0.8587 (0.0061)	0.8474 (0.0065)
NIR-H	0.8475 (0.0054)	0.8432 (0.0054)
NIR-H-Diff	**0.8613 (0.0049)***	**0.8546 (0.0048)***

Here are the SSIM scores of registration results on two folds of testing image pairs. M-M represents testing image pairs where moving and target images both come from Mindboggle dataset. M-O^h represents testing image pairs where moving and target images are sourced from Mindboggle and healthy OASIS images, separately.

Table A.8
Influence of population shift.

	O^h-O^{ad}	O^h-O^h
VoxelMorph	0.8325 (0.022)	0.8274 (0.024)
SYM_Net	0.8453 (0.020)	0.8372 (0.022)
TransMorph	**0.8521 (0.020)**	0.8463 (0.021)
NIR-H	0.8392 (0.021)	0.8392 (0.021)
NIR-H-Diff	0.8496 (0.020)	**0.8496 (0.020)***

Here are the volumetric dice scores of registration results of models trained with two folds of image pairs. O^h-O^{ad} represents training image pairs where moving and target images both come from healthy OASIS data. O^h-O^h represents training image pairs where moving and target images are sourced from healthy and Alzheimer's disease OASIS images, separately.

optimization. Our results, presented in Table A.6, show that increasing the scale of λ_{reg} by 100 times for NIR-D-Diff leads to a significant decrease in DSC when comparable registration regularity to NIR-H-Diff is achieved. Decreasing λ_{reg} for NIR-P-Diff does not improve registration accuracy, but rather harms registration regularity. Therefore, we conclude that using a hybrid coordinate sampler, as in NIR-H-Diff, is a more effective approach for achieving a balance between registration accuracy and regularity, compared to simply adjusting the weight of the regularization term.

A.5. Visual comparison for experiment (3)

Fig. A.11 displays qualitative comparisons between our innovative models and two established learning-based methods, SYM_Net and TransMorph. Compared to the registration results from experiment (1), featured in Fig. 7, all methods exhibit a decline in performance in experiment (3), with SYM_Net and TransMorph being particularly affected. A pronounced misalignment of high-intensity arterial structures is noticeable between the target slice and the registration outputs of SYM_Net and TransMorph. This misalignment may be attributed to the hippocampal atrophy observed in the OASIS dataset, an early hallmark of Alzheimer's disease.

A.6. Domain shift analysis

In experiment (3), we investigate the performance gap between the baseline learning-based methods and our proposed methods on cross-dataset registration task. This section presents a comprehensive assessment, focusing on contrast shift, population shift, and task shift analyses.

Contrast shift analysis. To discern the impact of contrast shifts in isolation from population differences, we train learning-based models on the MindBoggle image pairs, same as experiment (1), and evaluate them on two folds of 100 healthy MRI scan pairs from the OASIS and MindBoggle datasets. The results, presented in Table A.7, indicate minimal performance degradation in SSIM scores when aligning healthy MRI scan pairs across different datasets. This suggests that the baseline learning-based methods demonstrate robust generalization capabilities in the presence of contrast shifts alone, likely benefiting from intensity augmentation during training.

Population shift analysis. We separately trained models on 9000 healthy image pairs and 9000 healthy-to-AD image pairs from the OASIS dataset and evaluated them on 100 healthy-to-AD OASIS image pairs. The training data are augmented with random elastic transformation. The findings, detailed in Table A.8, reveal significant performance drops in volumetric Dice scores, highlighting the profound impact of population shifts on registration performance. The limited effectiveness of elastic transformation augmentations in this context suggests the inherent challenge in replicating realistic AD-related morphological changes, such as hippocampal atrophy or ventricular enlargement, through random deformations. Our proposed methods demonstrate significantly better registration accuracy on the testing healthy-to-AD OASIS image pairs.

Task shift analysis. As previously discussed in Table 2, we observed that models pre-trained on IXI image pairs exhibit limited generalizability to the Mindboggle dataset, despite both involving healthy MRI scans. This underscores the influence of task-specific domain shifts, with the initial pre-training focused on atlas-to-patient registration and subsequent fine-tuning on patient-to-patient registration.

References

Anon, 2020. Anaconda Software Distribution. Anaconda Inc, URL https://docs.anaconda.com/.

Arsigny, V., Commowick, O., Pennec, X., Ayache, N., 2006. A log-Euclidean framework for statistics on diffeomorphisms. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part I 9. Springer, pp. 924–931.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. Neuroimage 11 (6), 805–821.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41.

Avants, B.B., Tustison, N., Song, G., et al., 2009. Advanced normalization tools (ANTS). Insight J. 2 (365), 1–35.

Bajcsy, R., Kovačič, S., 1989. Multiresolution elastic matching. Comput. Vis. Graph. Image Process. 46 (1), 1–21.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2018. An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9252–9260.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. VoxelMorph: A learning framework for deformable medical image registration. IEEE Trans. Med. Imaging 38 (8), 1788–1800.

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. 61 (2), 139–157.

Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al., 2022. MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701.

Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y., 2022. Transmorph: Transformer for unsupervised medical image registration. Med. Image Anal. 82, 102615.

Chen, J., He, Y., Frey, E.C., Li, Y., Du, Y., 2021a. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. arXiv preprint arXiv:2104.06468.

Chen, Y., Liu, S., Wang, X., 2021b. Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8628–8638.

Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D., 2018. Neural ordinary differential equations. Adv. Neural Inf. Process. Syst..

Christensen, G.E., Rabbitt, R.D., Miller, M.I., 1996. Deformable templates using large deformation kinematics. IEEE Trans. Image Process. 5 (10), 1435–1447.

Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2018. Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 729–738.

Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2019. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Med. Image Anal. 57, 226–236.

Dupont, E., Doucet, A., Teh, Y.W., 2019. Augmented neural odes. arXiv preprint arXiv:1904.01681.

Dupuis, P., Grenander, U., Miller, M.I., 1998. Variational problems on flows of diffeomorphisms for image matching. Q. Appl. Math. 587–600.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al., 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., Group, B.D.C., et al., 2011. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 54 (1), 313–327.

Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage (47), S102.

Frankle, J., Carbin, M., 2018. The lottery ticket hypothesis: Finding sparse, trainable neural networks. arXiv preprint arXiv:1803.03635.

Gandelsman, Y., Shocher, A., Irani, M., 2019. Double-DIP: Unsupervised image decomposition via coupled deep-image-priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11026–11035.

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I., Contributors, D., 2014. Dipy, a library for the analysis of diffusion MRI data. Front. Neuroinform. 8, 8.

Gupta, K., 2020. Neural Mesh Flow: 3d Manifold Mesh Generation Via Diffeomorphic Flows. University of California, San Diego.

Häger, S., Heldmann, S., Hering, A., Kuckertz, S., Lange, A., 2020. Variable fraunhofer MEVIS RegLib comprehensively applied to Learn2Reg challenge. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 74–79.

Hellier, P., Barillot, C., Mémin, E., Pérez, P., 2001. Hierarchical estimation of a dense deformation field for 3-D robust registration. IEEE Trans. Med. Imaging 20 (5), 388–402.

Hering, A., Hansen, L., Mok, T.C., Chung, A., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., et al., 2021. Learn2Reg: Comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. arXiv preprint arXiv:2112.04489.

Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V., 2021. SynthMorph: Learning contrast-invariant registration without acquired images. IEEE Trans. Med. Imaging 41 (3), 543–558.

Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V., 2021. Hypermorph: Amortized hyperparameter learning for image registration. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 3–17.

Incoronato, M., Aiello, M., Infante, T., Cavaliere, C., Grimaldi, A.M., Mirabelli, P., Monti, S., Salvatore, M., 2017. Radiogenomic analysis of oncological data: A technical survey. Int. J. Mol. Sci. 18 (4), 805.

Jacot, A., Gabriel, F., Hongler, C., 2018. Neural tangent kernel: Convergence and generalization in neural networks. In: Advances in Neural Information Processing Systems, vol. 31.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. In: Advances in Neural Information Processing Systems, vol. 28.

Kim, B., Kim, D.H., Park, S.H., Kim, J., Lee, J.-G., Ye, J.C., 2021. CycleMorph: Cycle consistent unsupervised deformable image registration. Med. Image Anal. 71, 102036.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Klein, A., Tourville, J., 2012. 101 Labeled brain images and a consistent human cortical labeling protocol. Front. Neurosci. 6, 171.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19 (9), 1498–1507.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2020. Nerf: Representing scenes as neural radiance fields for view synthesis. In: European Conference on Computer Vision. Springer, pp. 405–421.

Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. Comput. Methods Programs Biomed. 98 (3), 278–284.

Mok, T.C., Chung, A., 2020a. Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4644–4653.

Mok, T.C., Chung, A., 2020b. Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 211–221.

Mok, T.C., Chung, A., 2022. Affine medical image registration with coarse-to-fine vision transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20835–20844.

Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A., 2019. Occupancy flow: 4d reconstruction by learning particle dynamics. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5379–5389.

Nikolov, S., Blackwell, S., Zverovitch, A., Mendes, R., Livne, M., De Fauw, J., Patel, Y., Meyer, C., Askham, H., Romera-Paredes, B., et al., 2018. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. arXiv preprint arXiv:1809.04430.

Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S., 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 165–174.

Pontryagin, L.S., 1987. Mathematical Theory of Optimal Processes. CRC Press.

Quan, Y., Chen, M., Pang, T., Ji, H., 2020. Self2self with dropout: Learning self-supervised denoising from single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1890–1898.

Reinke, A., Eisenmann, M., Tizabi, M.D., Sudre, C.H., Rädsch, T., Antonelli, M., Arbel, T., Bakas, S., Cardoso, M.J., Cheplygina, V., et al., 2021. Common limitations of image processing metrics: A picture story. arXiv preprint arXiv:2104.05642.

Ren, D., Zhang, K., Wang, Q., Hu, Q., Zuo, W., 2020. Neural blind deconvolution using deep priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3341–3350.

Risholm, P., Golby, A.J., Wells, W., 2011. Multimodal image registration for preoperative planning and image-guided neurosurgical procedures. Neurosurg. Clin. 22 (2), 197–206.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Shen, D., Davatzikos, C., 2002. HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging 21 (11), 1421–1439.

Shen, Z., Han, X., Xu, Z., Niethammer, M., 2019. Networks for joint affine and non-parametric image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4224–4233.

Shen, L., Pauly, J., Xing, L., 2021. NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. arXiv preprint arXiv:2108.10991.

Shi, J., He, Y., Kong, Y., Coatrieux, J.-L., Shu, H., Yang, G., Li, S., 2022. Xmorpher: Full transformer for deformable medical image registration via cross attention. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI. Springer, pp. 217–226.

Siebert, H., Hansen, L., Heinrich, M.P., 2021. Fast 3D registration with accurate optimisation and little learning for Learn2Reg 2021. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 174–179.

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G., 2020. Implicit neural representations with periodic activation functions. Adv. Neural Inf. Process. Syst. 33, 7462–7473.

Sun, S., Han, K., Kong, D., Tang, H., Yan, X., Xie, X., 2022. Topology-preserving shape reconstruction and registration via neural diffeomorphic flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20845–20855.

Sun, Y., Liu, J., Xie, M., Wohlberg, B., Kamilov, U.S., 2021. Coil: Coordinate-based internal learning for imaging inverse problems. arXiv preprint arXiv:2102.05181.

Tancik, M., Srinivasan, P.P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J.T., Ng, R., 2020. Fourier features let networks learn high frequency functions in low dimensional domains. In: Neural Information Processing Systems.

Thirion, J.-P., 1998. Image matching as a diffusion process: An analogy with Maxwell's demons. Med. Image Anal. 2 (3), 243–260.

Ulyanov, D., Vedaldi, A., Lempitsky, V., 2018. Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9446–9454.

Van Horn, J.D., Grethe, J.S., Kostelec, P., Woodward, J.B., Aslam, J.A., Rus, D., Rockmore, D., Gazzaniga, M.S., 2001. The functional magnetic resonance imaging data center (fMRIDC): The challenges and rewards of large–scale databasing of neuroimaging studies. Philos. Trans. R. Soc. London. Ser. B: Biol. Sci. 356 (1412), 1323–1339.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Williams, F., Schneider, T., Silva, C., Zorin, D., Bruna, J., Panozzo, D., 2019. Deep geometric prior for surface reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10130–10139.

Wolterink, J.M., Zwienenberg, J.C., Brune, C., 2021. Implicit neural representations for deformable image registration. In: Medical Imaging with Deep Learning.

Wu, Y., Jiahao, T.Z., Wang, J., Yushkevich, P.A., Hsieh, M.A., Gee, J.C., 2022. NODEO: A neural ordinary differential equation based optimization framework for deformable image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20804–20813.

Wu, Q., Li, Y., Xu, L., Feng, R., Wei, H., Yang, Q., Yu, B., Liu, X., Yu, J., Zhang, Y., 2021. Irem: High-resolution magnetic resonance image reconstruction via implicit neural representation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 65–74.

Xu, J., Chen, E.Z., Chen, X., Chen, T., Sun, S., 2021. Multi-scale neural odes for 3d medical image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 213–223.

Xu, J., Moyer, D., Gagoski, B., Iglesias, J.E., Grant, P.E., Golland, P., Adalsteinsson, E., 2023. NeSVoR: Implicit neural representation for slice-to-volume reconstruction in MRI. IEEE Trans. Med. Imaging.

Xu, Z., Niethammer, M., 2019. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 420–429.

Yang, G., Huang, X., Hao, Z., Liu, M.-Y., Belongie, S., Hariharan, B., 2019. Pointflow: 3d point cloud generation with continuous normalizing flows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4541–4550.

Yüce, G., Ortiz-Jiménez, G., Besbinar, B., Frossard, P., 2022. A structured dictionary perspective on implicit neural representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19228–19238.

Zhang, M., Liao, R., Dalca, A.V., Turk, E.A., Luo, J., Grant, P.E., Golland, P., 2017. Frequency diffeomorphisms for efficient image registration. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 559–570.

Zhang, Y., Pei, Y., Zha, H., 2021. Learning dual transformer network for diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 129–138.

Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al., 2019. Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10600–10610.

Zheng, Z., Yu, T., Dai, Q., Liu, Y., 2021. Deep implicit templates for 3d shape representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1429–1439.

Zhu, W., Huang, Y., Xu, D., Qian, Z., Fan, W., Xie, X., 2021. Test-time training for deformable multi-scale image registration. In: 2021 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 13618–13625.