NODEO: A Neural Ordinary Differential Equation Based Optimization Framework for Deformable Image Registration

Yifan Wu* Tom Z. Jiahao* Jiancong Wang Paul A. Yushkevich M. Ani Hsieh James C. Gee
University of Pennsylvania, Philadelphia, PA, USA
{yfwu, zjh}@seas.upenn.edu {jiancong.wang, pauly2}@pennmedicine.upenn.edu
mya@seas.upenn.edu gee@upenn.edu

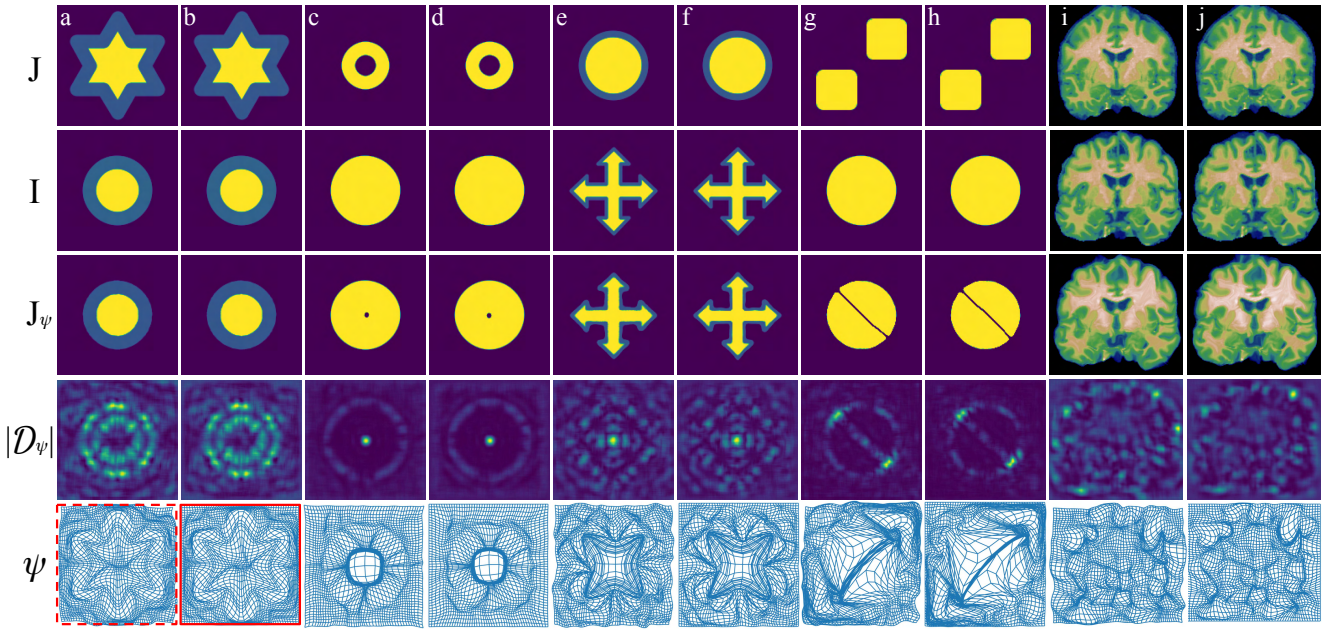

Figure 1. **2D image registration using our framework.** The rows show (J) the moving images, (I) fixed images, (J_ψ) warped moving images, ($|\mathcal{D}_\psi|$) Jacobian determinants of the transformation ψ, and visualization of ψ respectively. The columns (b)(d)(f)(h)(j) incorporate a fixed boundary constraint, as indicated by the box with solid red line, while (a)(c)(e)(g)(i) do not (box with dotted red line).

Abstract

Deformable image registration (DIR), aiming to find spatial correspondence between images, is one of the most critical problems in the domain of medical image analysis. In this paper, we present a novel, generic, and accurate diffeomorphic image registration framework that utilizes neural ordinary differential equations (NODEs). We model each voxel as a moving particle and consider the set of all voxels in a 3D image as a high-dimensional dynamical system whose trajectory determines the targeted deformation field. Our method leverages deep neural networks for their expressive power in modeling dynamical systems, and simultaneously optimizes for a dynamical system between the image pairs and the corresponding transformation. Our formulation allows various constraints to be imposed along the transformation to maintain desired regularities. Our experiment results show that our method outperforms the benchmarks under various metrics. Additionally, we demonstrate the feasibility to expand our framework to register multiple image sets using a unified form of transformation,which could possibly serve a wider range of applications.

1. Introduction

Deformable image registration (DIR) is a process for establishing spatial correspondences between images. The term "deformable" points to the nonlinear and dense nature of the required transformation. DIR has a broad range of applications including normalization of population studies, quantifying changes in longitudinal imaging, accounting for mo-

*Equal contribution.

tions of organ, as a building block of other image analysis algorithms, etc. While there is usually no ground-truth for the optimal transformation, image registration is usually formulated according to and evaluated on three criteria: the accuracy of matching the source and target images in terms of a pre-defined similarity metric, the regularity of the transformation to ensure that it is well-behaved, such as topology preservation, and the speed of the algorithm [34, 37].

Traditional approaches solve DIR as a pair-wise optimization problem. These methods enforce transformation regularity through hard model assumptions. For example, Large Deformation Diffeomorphic Metric Mapping (LDDMM) [7, 23], one of the most influential approaches, solves for diffeomorphisms by formulating the registration problem using a flow Partial Differential Equation (PDE), where the intergral of time-varying velocity fields produces the final deformation. These methods, however, can be challenging to apply if both speed and accuracy are desired, and they potentially limit the performance if different model assumptions are needed [34].

With rapid advancement in machine learning and abundance of medical imaging data, there is increased interest in developing deep learning based methods to solve the DIR problem [6, 11, 12, 25, 45]. These methods significantly reduce the runtime of registration through using neural networks to learn a good sharing representation of a training dataset. Registration on a new image pair then becomes a rapid inference process. However, the generalizablity to unseen data is a long-standing challenge of recent data driven learning based methods, precluding their straightforward application in practice.

Recent development in scientific machine learning has shown promising results in modeling differential equations using neural networks [17, 47], which can describe any system that evolves with "time". Given the demonstrated utility of flow-based approaches as in LDDMM [7] and the known advantages in expressive power of deep neural networks, we ask: *can we integrate the merits of both?* We attempt to analyze this possibility from the following perspectives.

(i) The image registration problem can be viewed as a system-identification problem. Specifically, our task is to find a differential equation whose solution gives the transformation between the images. By imposing few assumptions on the dynamical system, we are able to explore various classes of systems and their solutions to achieve the desired registration while ensuring certain solution properties on demand, such as topology preservation.

(ii) Traditional model-based approaches are naturally constrained by the models they assume and implement. The hard constraints imposed by a given model may potentially be relaxed for better model capacity and regularity. In particular, a learnable flow that allows for penalties on its trajectory may serve as an alternative framework.

(iii) Adopting deep learning in DIR may not require additional data. The expressive power of deep learning stems from its compositional rule of functions. Much of the existing DL-based registration work builds on the foundations of feature learning, which learns the map or flow from supervised labels or a sufficient number of image pairs. Another easily overlooked utility of neural networks is they can serve as a 'parametric' backbone for general optimization problems. For the DIR problem, we can use a network to parameterize the system we aim to identify in (i-ii).

Motivated by the analysis above, we propose *NODEO*, a neural ordinary differential equation based optimization framework that formulates the velocity field optimization as a neural network optimization. Specifically, we treat the set of all pixel/voxel locations in an image as one single evolving system, and parametrize the system's evolving dynamics with a deep neural network. The registration task therefore becomes finding a system whose trajectory's end point is the deformation field that minimizes the dissimilarity between images.

The benefits of our approach are twofold. *First*, starting from the generalized flow field approach and then specializing to image registration, our framework provides enhanced flexibility. Our framework makes embedding proper dynamical constraints (such as spatial smoothness) straightforward, allowing flexibility in terms of defining task-specific "goodness" for our transformations, and producing solutions with the desired properties. We can also easily incorporate boundary conditions on demand. By imposing loss penalties on the intermediate states of the trajectory, our model permits great flexibility in the type and number of assumptions one can impose on the solution. Thus, our framework is straightforward to extend from image pairs to multiple-image sets by adding intermediate supervision. *Second*, we explicitly model the image grid as one high-dimensional system, so that convolutional layers can be naturally leveraged to allow particles to spatially interact with each other within the system dynamics. The proposed solution brings the full expressive power of neural nets, as demonstrated in Figure 1. Note that our approach falls under the family of pair-wise optimization approaches, *i.e.* network parameters are optimized without any additional data.

Overall, the contributions of this paper are: (1) generalization of the flow field approach to registration through high-dimensional dynamical system modeling, (2) unifying the optimization problem of discovering differential equations and their solutions (velocity fields) in one network *without training data*, (3) enhanced flexibility and effectiveness of adding desired regularizations and constraints on solution transformations, and (4) demonstration that the proposed framework can be serve as an alternative approach to optimization-based registration tool that have *proven utility*

across numerous application domains, with state-of-the-art performance over a variety of evaluation metrics.

2. Related Work

2.1. Pair-wise Optimization-based Methods

There are a number of prominent DIR techniques that have been evolved into widely used tools. NiftyReg [24] is a representative work of parametric approaches [29, 30] for describing continuous interpolations with a finite set of parameters utilizing basis functions, b-splines, etc. Demons and its variants [36, 39, 40] are optical flow based non-parametric techniques. SyN [5] is a representative work of greedy techniques and is well-known for producing symmetric solutions. The solution transformations family is a choice one makes based on the application or features one seeks to develop. When it's desired that the tranformation can be large deformation and be topology preserving, which is a property useful in many applications, then flow formulation is the natural family of transformation to use. Large Deformation Diffeomorphic Metric Mapping (LDDMM) is one of the most influential method in this family, which generate the trajectory from the source to target images rather than simply the transformation. LDDMM formulates the registration problem as solving a velocity field that "flows" the source image to the target one. Early LDDMM works directly solve the Euler-Lagrange equation by variational approach. Shooting method is adopted later using Euler-Poincare characteristic for the velocity field to guarantee geodesic path, reducing the optimization space from spatial-temporal velocity fields to initial momentum [7, 23]. The elegant formulation of the LDDMM motivates a number of follow-up efforts to improve the framework, including developments of optimization method using adjoint [41], discretize-then-optimize paradigm to reduce runtime [19, 28], spatial-temporal variant regularizations to relax the constraints [32] and so on. LDDMM is tailored to the registration problem and makes strong assumptions about the dynamics of the flow field, whereas our method is more flexible because we begin with a generic flow field approach and then narrow down to image registration.

2.2. Data-driven Learning-based Methods

With the power of deep learning, there is growing interest in creating learning-based solutions to the DIR problem [6, 10, 11, 45]. By learning a common representation for a collection of images then performing registration in the inference stage, learning-based methods can significantly reduce the runtime [12, 13, 26, 31]. Some efforts attempt to include regularity, such as diffeomorphism, into networks by developing symmetric and reversevable structures or penalty [16, 21, 25]. To better solve large deformations, Xu *et al*. proposed to use neural ODE on image registration to refine the estimated transformation, by modeling the dynamics of the parameters of registration models (e.g., b-splines) [43]. However, they do not use neural ODE to directly describe the transformation between images as a flow like ours. Despite numerous efforts to improve generalizability, such as data augmentation or low-shot learning [9, 33], generalizability remains a long-standing difficulty of current data-driven learning-based approaches. In contrast to the most learning-based methods for registration, which build on feature learning, we employ network to represent a differential equation.

3. Method

3.1. Deformable Image Registration Formulation

The deformable image registration problem can be formulated as follows: consider an unparametrized 3D image as a discrete solid, where the location of the i^{th} voxel/point is given by $\mathrm{x}_i \in \Omega \subseteq \mathbb{R}^3, i \neq j \iff \mathrm{x}_i \neq \mathrm{x}_j$, where Ω is the image domain. This voxel location is known in shape analysis as a landmark coordinate. The location of all voxels or the voxel cloud can be denoted by the ordered set $q = \{x_i\}_{i=1}^N$, where $N = D \times H \times W$ is the total number of voxels in the image, and D, H, and W are the image depth, height and width respectively. As a shorthand, we write Π as the domain of voxel clouds and $q \in \Pi$. We denote the fixed image by I and the moving image by J, which are functions $I, J : \Omega \to \mathbb{R}^d$, mapping each voxel coordinate to the voxel value/intensity. In this work, we only consider scalar-valued, e.g., MRI images, and therefore $d = 1$.

Traditionally, the goal of DIR is to find some *good* transformation $\phi : \Omega \to \Omega$ such that the transformed moving image $J(\phi(\mathrm{x})), \forall \mathrm{x} \in q$ is similar to the fixed image I [7]. Here, ϕ is the spatial transformation that maps the domain of voxels Ω onto itself. An identity map is defined as ϕ_0 such that $\phi_0(\mathrm{x}) = \mathrm{x}$. In many applications [15] of DIR, a *good* property of transformation ϕ is both sufficiently smooth and diffeomorphic in $\mathbb{R}^3$. The latter condition requires that the topology of the moving image is preserved under transformation. In other words, the transformation ϕ should not create folds in Ω. We can overload the function ϕ by defining its application on a voxel cloud as a point-wise transformation on the voxels, i.e. $\phi(q_f) = (\phi(\mathrm{x}_1), \phi(\mathrm{x}_2), ..., \phi(\mathrm{x}_N))^T$, where $q_f = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N)^T$ is the spatially flattened voxel cloud. Similarly, this point-wise transformation can be defined for image J when applied on a transformed voxel cloud.

Following the above definitions, DIR is generally formulated as the minimization of the combined image similarity metric and regularization of the transformation given by

$$\mathcal{J}(\phi; I, J) = \mathcal{S}(J(\phi(q_0)), I) + \mathcal{R}(\phi), \tag{1}$$

where q_0 is the initial voxel cloud when none of the voxels

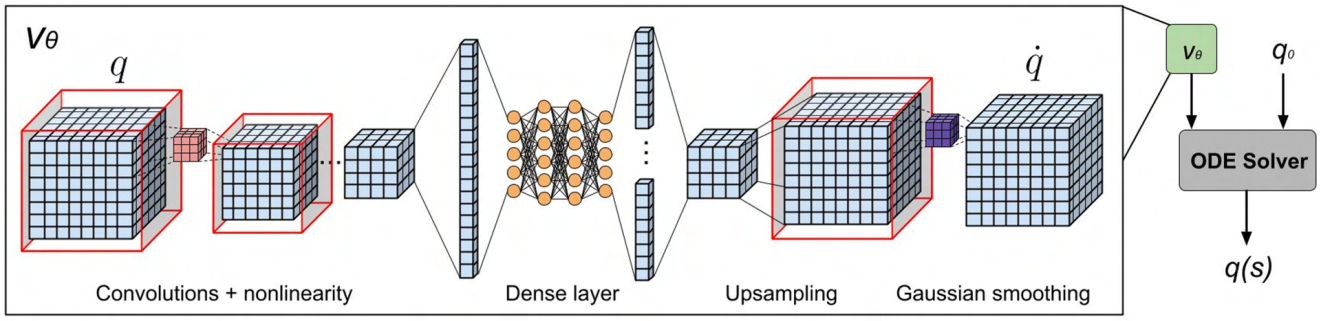

Figure 2. **Framework overview.** Our frame models the vector field $\mathbf{v}_\theta$ as a neural network. A voxel cloud **without intensity** first gets down-sampled using convolutional layers. It then passes through dense layers where time is injected. It then gets upsampled and restored back to the shape of the voxel cloud and smoothened with a gaussian kernel. Note the images are only used for similarity measurement.

in the cloud has undergone transformation. The term $\mathcal{S}(\cdot,\cdot)$ is the similarity metric that measures the difference between the deformed moving image J and the fixed image I. The term $\mathcal{R}(\cdot)$ is the regularization on the transformation.

However, in this work, we reformulate the problem by searching for a transformation $\psi : \Pi \rightarrow \Pi$ which maps the domain of the voxel cloud Π onto itself, while requiring diffeomorphism in Π. This is a marked contrast with most works on pointcloud deformation without intensity [18, 27, 44] that model mapping $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. In this work, we explicitly allow interactions among voxels by searching for a transformation in the high-dimensional cloud space. Our problem formulation has a similar form as Eqn. (1), with the exception that the transformation over the voxel cloud is no longer point-wise, but instead given by

$$\mathcal{J}(\psi; I, J) = \mathcal{S}(J(\psi(q_0)), I) + \mathcal{R}(\psi), \tag{2}$$

where $\psi(q_f) = (\psi(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N))^T$. We can alternatively write it as $\psi(q_f) = (\psi_1(q_f), \psi_2(q_f), ..., \psi_N(q_f))^T$, where ψ_i is called the i^{th} component of the ψ.

3.2. Neural Ordinary Differential Equations

Taking inspiration from the resemblance between residual networks and dynamical systems, Chen et al. [8] first introduced neural ordinary differential equations (NODEs) to approximate infinite depth neural networks. It aims to learn the function f parameterized by θ by defining a loss function of the following form

$$\frac{dz}{dt} = f_\theta(z(t), t), \tag{3}$$

$$\mathcal{L}(\mathrm{z}(t_1)) = \mathcal{L}\left(\mathrm{z}_0 + \int_{t_0}^{t_1} f_\theta(\mathrm{z}(t), t)dt\right). \tag{4}$$

From a system perspective, NODEs are continuous-time models that represent vector fields as neural networks. It has since been adapted as a universal framework for modeling high-dimensional spatiotemporally chaotic systems utilizing convolutional layers [17], demonstrating its ability to capture highly complex behaviors in space and time. Hence, we find it a suitable candidate for our registration task.

Since NODEs often require the numerical solver to take many steps to realize the flows, they are memory-inefficient if all gradients along the the integration steps needs to be stored using traditional backpropagation. Hence many recent works [8, 46] on NODEs have therefore focused on reducing the memory requirements for gradient propagation. Notably, the adjoint sensitivity method (ASM) has enabled constant memory gradient propagation for optimizing NODEs, and we adopt ASM in our work as well. For a brief description of ASM, one can refer to the supplementary. Proofs for its gradient convergence can be found in [8, 17].

ASM enables our framework to interpolate between $t = 0$ and $t = s$ for an arbitrary number of steps with constant memory cost. This is particularly helpful when a temporally smooth diffeomorphic flow is required, as the numerical solver can increase its number of steps to improve the smoothness of $q(t)$ with respect to t. Also, our model can also be extended to multiple or sequential images by imposing constraints on the intermediate states of the trajectory.

3.3. DIR in Dynamical System View

Our work borrows intuition from dynamical systems and treats the trajectory of the entire voxel cloud as the solution to a first-order non-autonomous ordinary differential equation given by

$$\begin{aligned} \frac{dq}{dt} &= \mathcal{K}\mathbf{v}_\theta(q(t), t), \\ s.t.\ q(0) &= q_0, \end{aligned} \tag{5}$$

where $\mathbf{v}_\theta(\cdot)$, as parametrized by θ, is the vector field describing the dynamics of the voxel cloud, q_0 is the initial condition at $t = 0$. We employ Gaussian kernels (for $\Omega \subseteq \mathbb{R}^3$, we use 3D Gaussian kernels), denoted by $\mathcal{K}$, as a

filtering operator to enforce spatial smoothness in Ω. Intuitively $\mathcal{K}$ is to ensure that the velocities of voxels are similar to those of their neighbors. Increasing the kernel size amounts to smoothing over larger voxel space, and therefore will improve the smoothness of the resulting flow; increasing the variance of the kernel amounts to encouraging more individual movements and therefore reduces the smoothness of the resulting flow.

The term *non-autonomous*, or equivalently *time-variant*, *non-stationary* means that the time derivative of q explicitly depends on t [35]. In other words, the velocity field attached to the Eulerian frame changes with time.

The trajectory of q is generated by integrating the ODE in Eqn. (5) with the initial condition q_0. Assuming that the voxel cloud evolves from $t = 0$ to $t = s$, the resulting voxel cloud at $t = s$ denotes the transformation $\psi(q_0)$ given by

$$\psi(q_0) = q(s) = q_0 + \int_0^s \mathcal{K}\mathbf{v}_\theta(q(t), t)dt. \tag{6}$$

Eqn. (6) is referred to as a diffeomorphic flow map in dynamical systems. While the uniqueness and existence theorem [35] only implies diffeomorphism in the high-dimensional space Π, we will show that diffeomorphism can be achieved in Ω by incorporating soft constraints in the optimization task. In practice, the flow map is computed using a numerical integration scheme such as the Euler's method. Note that while s is chosen to be 1 in most existing works, it can be parametrized by the total number of steps taken by the solver and the corresponding step sizes. The task of finding the transformation ψ therefore becomes finding the best set of parameters describing $\mathbf{v}$. The optimization problem therefore becomes:

$$\theta = \arg\min_{\theta\in\Theta} \mathcal{L}_{sim}\left(I, J(q_0 + \int_0^s \mathcal{K}\mathbf{v}_\theta\left(q(t), t\right) dt)\right) + \mathcal{R}(\psi, \mathbf{v}_\theta) + \mathcal{B}(\psi), \tag{7}$$

where Θ is the space of all possible parameters. The different components in the loss function include the similarity metric $\mathcal{L}_{sim}$, the regularizers $\mathcal{R}$, and the boundary conditions $\mathcal{B}$. The individual tasks can employ a different measure that suits the problem for each regularization term here. The similarity loss is $\mathcal{L}_{sim}(I, J) = 1 - NCC(I, J)$, where NCC is the normalized cross correlation given by

$$NCC(I, J) = \frac{1}{N}\sum_{\mathrm{x}\in q(s)} \frac{\sum_{\mathrm{x}_i\in W}(I(\mathrm{x}_i)-\bar{I}(\mathrm{x}))(J(\mathrm{x}_i)-\bar{J}(\mathrm{x}))}{\sqrt{\sum_{\mathrm{x}_i\in W}(I(\mathrm{x}_i)-\bar{I}(\mathrm{x}))^2 \sum_{\mathrm{x}_i\in W}(J(\mathrm{x}_i)-\bar{J}(\mathrm{x}))^2}}, \tag{8}$$

where $\bar{I}(\mathrm{x})$ and $\bar{J}(\mathrm{x})$ are the local mean of a size w^3 window W with x being at its center position, and x_i is an element within this window. In this work we set w as 21.

The regularization term consists of three terms:

$$\mathcal{R}(\psi, \mathbf{v}_\theta) = \lambda_1\mathcal{L}_{Jdet} + \lambda_2\mathcal{L}_{mag} + \lambda_3\mathcal{L}_{smt}. \tag{9}$$

The first term, $\mathcal{L}_{Jdet}$, penalizes negative Jacobian determinants in the transformed voxel cloud and is given by

$$\mathcal{L}_{Jdet} = \frac{1}{N}\sum_{\mathrm{x}\in q(s)} \|\sigma(-(|\mathcal{D}_\psi(\mathrm{x})| + \epsilon))\|_2^2, \tag{10}$$

where $\mathcal{D}_\psi(\mathrm{x})$ is the Jacobian matrix at x under the transformation ψ. Here $\sigma(\cdot) = max(0, \cdot)$ is the ReLU activation function, which is used to select only negative Jacobian determinants. If there are no folds in the transformation, its jacobian determinant $\mathcal{D}_\psi(\mathrm{x})$ should be positive. Lastly, we add a small number ϵ to the Jacobian determinants as an overcorrection. Instead of using $L1$ regularization as in [11, 25], we use $L2$ norm here. Regularization with $L1$ introduces sparsity, reducing the number of folds, while the $L2$ norm can minimize the overall magnitude of folds, thereby avoiding outliers. To adapt to a specific task, one can combine the two. In our framework, $\mathcal{L}_{Jdet}$ is a critical component since it ensures that the flow is diffeomorphic in Ω. In this work, the Jacobian matrix is implemented using the finite difference approximation.

The second term, $\mathcal{L}_{mag}$, regularizes the magnitude of the velocity field along the voxel cloud trajectory and is given by

$$\mathcal{L}_{mag} = \frac{1}{N}\int_0^s \|\mathcal{K}\mathbf{v}_\theta(q(t), t)\|_2^2 dt. \tag{11}$$

This amounts to penalising the *energy* of the flow. In practice, the integral is replaced by a summation of the squared norm along the steps taken by the numerical integration scheme. Lastly, the third term, $\mathcal{L}_{smt}$, is used to regularize the spatial gradients of the transformed voxel cloud and is given by

$$\mathcal{L}_{smt} = \frac{1}{N}\sum_{\mathrm{x}\in q(s)} (\|\nabla_\psi(\mathrm{x})\|_2^2), \tag{12}$$

where $\nabla_\psi(\mathrm{x})$ denotes the spatial gradient around x under the transformation ψ. This encourages spatial smoothness of the transformed voxel cloud. Similar to $\mathcal{L}_{Jdet}$, $\mathcal{L}_{smt}$ is implemented as a discrete approximation. Note that, even though $\mathbf{v}_\theta$ already includes Gaussian filtering, which in turn translates to the spatial smoothness of ψ, the inclusion of $\mathcal{L}_{smt}$ can reduce the degradation in smoothness as a result of numerical integration. The last term in Eqn.(7), $\mathcal{B}(\psi)$, specifies the boundary condition for the transformation ψ. While our MRI registration tasks do not specify any boundary condition ($\mathcal{B}(\psi) = 0$), we will demonstrate its effect through illustrative experiments on 2D images.

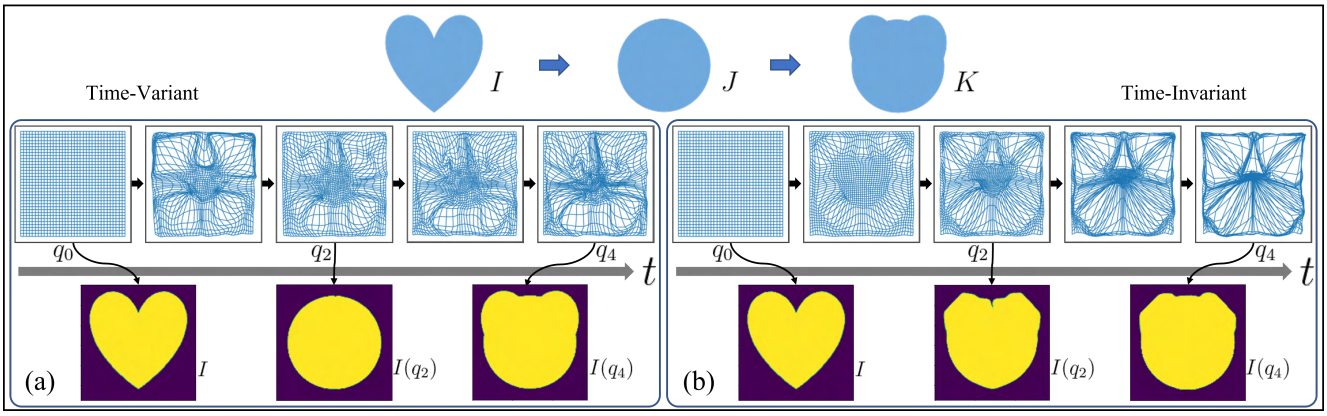

Figure 3. **Discovering transformation on multiple images.** The target images I, J, K are shown in the top row, and the model is tasked to identify a path of transformation from I to K via J. The pictured registration results are from using (a) a time-varying system with explicit time embedding, and (b) a time-invariant system where the transformation does not depend on time. The transformations $\psi(q_t)$ are plotted on top of the warped images during the registration.

4. Method Analysis

4.1. Illustrative Examples in 2D pair images

To demonstrate the properties and capabilities of our framework, we used it to perform registration on a variety of 2D images ($\Omega \subseteq \mathbb{R}^2$) as shown in Figure 1. The 2D brain images have size 144×160, and are slices taken from real brain MRI. All other images have size 144×144, and were hand-drawn. We used Mean Squared Error (MSE) for similarity measurement in Eqn (7), and did not include $\mathcal{L}_{mag}$ and $\mathcal{L}_{smt}$. Based on Figure 1, we note that the resulting transformation ψ is

topology preserving: While the registration warps the moving image as much as possible to look like the fixed image, it preserves the topology of the moving image in 2D. Columns (c) and (d) show that the warped "donut" closely resemble the circle, but the hole in its middle remains. Columns (g) and (h) also illustrate the two disconnected components will not become connected.

diffeomorphic in 2D: Diffeomorphism is a stronger condition than the topology preservation since it requires the transformation to be both continuous and differentiable. There are no visible folds in the visualization of ψ. There are few violations if we inspect the Jacobian determinants closely, and this can be further reduced by increasing the weight on $\mathcal{L}_{Jdet}$. Note that there actually exists no diffeomorphism between I and J, for the examples in columns (e)(f), because the sharp corners make I's manifold non-smooth. Even so, our framework produces a slightly rounded "cross", as the result of a smooth approximation.

enforcing boundary conditions: Columns (a) (c) (e) (g) (i) show registration without boundary conditions, while the remaining columns fix the grids on all four sides. It can be observed that with this boundary condition, the four sides of the resulting ψ remain unchanged.

To analyze the effects of regularization terms, we conducted ablation studies. Figure 4 shows the effect of Gaussian smoothing and $\mathcal{L}_{Jdet}$ regularizer on 2D images. We discovered by applying hard constraints $\mathcal{K}$ to the model allows it to attain the requisite spatial smoothness and continuity. Even though we are modeling a high-dimensional system to take advantage of its expressive capacity, we can achieve diffeomorphic transformation in low dimension, $\mathbb{R}^2$ in this case, by using soft regularizer $\mathcal{L}_{Jdet}$.

4.2. Illustrative Examples in 2D for multiple images

Our framework can also be extended to perform registration on a sequence of images, where the intermediate images act as constraints along the transformations. Figure 3 shows the task of finding the transformation from a heart to a bear, and the image must take the shape of a circle during the intermediate step. Here, we compare a time-variant and time-invariant model. To make the system time-variant, we embed temporal information using positional encoding similar to that of a transformer model [38]. To ensure diffeomorphism during the entire transformation, we applied Eqn. (10) on each of the intermediate steps. It can be seen from Figure 3 that the time-variant system produces a better registration result, demonstrating that incorporating time is important for more constrained tasks such as this. In practice, these intermediate images can be known temporal dynamics between image pairs (*e.g.* infant development [42], disease progression [4], cardiac or respiratory motions). In other words, being able to incorporate intermediate image constraints will allow for embedding domain knowledge into the registration process, and give more accurate transformations.

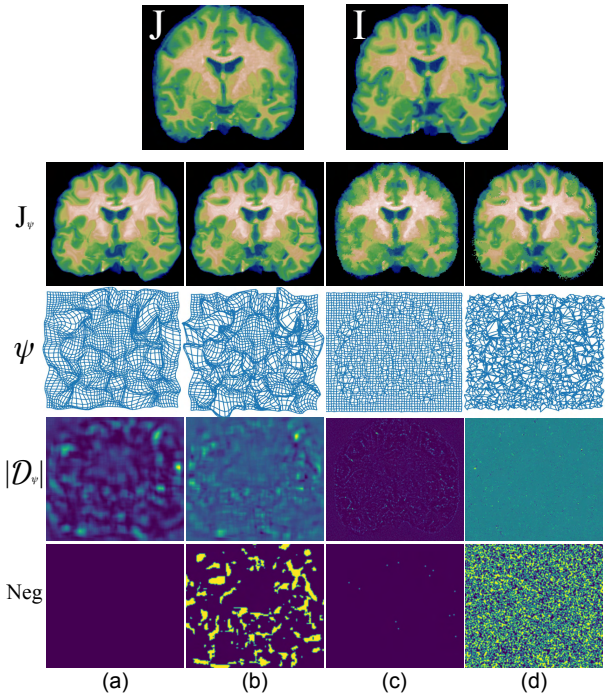

Figure 4. Demonstration of the effect of Gaussian smoothing $\mathcal{K}$ and the $\mathcal{L}_{Jdet}$ regularizer. The rows show (J_ψ) warped moving images, (ψ) grid visualization of the deformation field, ($|\mathcal{D}_\psi|$) Jacobian determinants of the transformation ψ, and (Neg) the regions with negative Jacobian determinants (yellow). The columns shows registration with (a) both $\mathcal{K}$ and $\mathcal{L}_{Jdet}$, (b) only $\mathcal{K}$, (c) only $\mathcal{L}_{Jdet}$, and (d) with neither.

5. Comparison with Benchmarks

5.1. Datasets and Pre-processing

We evaluate our framework on the same registration task as the state-of-the-art [25], which employs atlas-based registration. In this work, the atlases/templates are used as fixed images. We randomly select a small number of images as atlases and use the remainder as moving images to be registered to fixed images. We conduct experiments on two datasets: OASIS and CANDI.

OASIS: The OASIS [22] dataset consists of a collection of cross-sectional T1 MRI scans from 416 subjects. These subjects are aged 18-96 where 100 of them have been clinically diagnosed with mild to moderate Alzheimer's disease. We use standard processing tool FreeSurfer [14] to resample all scans to $1mm \times 1mm \times 1mm$, followed by motion correction and skull stripping. We use the whole brain auto-segmentation provided in FreeSurfer for evaluation. Then we align all images to MNI 152 space [14] by affine transformation. The final images have a size of $160 \times 192 \times 144$ by center cropping. We set five images with IDs $1, 10, 20, 30, 40$ as the atlases, and the remaining images with IDs < 50, as moving images. The total number of moving images is 40, resulting in 200 pairs to be registered.

CANDI: The Child and Adolescent NeuroDevelopment Initiative (CANDI) [20] dataset contains T1 MRI scans of 4 groups, which are Healthy Controls, Schizophrenia Spectrum, Bipolar Disorder, and Bipolar Disorder without Psychosis. We use manually-labeled whole brain segmentation provided in the dataset for skull stripping. Then we transform images along with their respective segmentations to the MNI 152 space and center crop the images to the size $160 \times 192 \times 144$ similar to the OASIS set. We set the first subject of each group as the atlases (fixed images), and the following 5 subjects as moving images. The total number of moving images is 20, resulting in 80 pairs for registration.

5.2. Evaluation Metric

The goal of DIR is to find the spatial correspondence such that the *similarity* of two images is maximized. In diffeomorphic registration, voxels are not allowed to self intersect, which can be guaranteed when the determinants of the Jacobian of the deformation field $\mathcal{D}_\psi(\mathrm{x})$ are non-negative. We follow the convention [10, 11, 16, 25] and measure similarity and diffeomorphism using the following two criteria.

Dice Similarity Coefficient : The dice score measures the ratio between the overlap and union of two spatial regions. Here we calculate dice using the whole-brain segmentation maps. In particular, we evaluate dice between the fixed segmentation and the warped moving segmentation based on the deformation field given by the registration of the two images. We use auto-segmentation maps and compute the average dice across 28 anatomical structures as in VoxelMorph [6]. For the CANDI dataset, as it provides manual labels for 32 structures, we report the average dice for both 28 and 32 structures. One can refer to supplementary material for details of these anatomical structures.

Jacobian Determinant: In our experiment, we report the negative Jacobian ratio denoted as $r^{\mathcal{D}}$, which represents the number of voxels with negative Jacobian determinants versus the total number of voxels for each image. In the meanwhile, we report the sum of negative values of the Jacobian determinant, denoted as $s^{\mathcal{D}}$, which represents the total area/volume of folding in pixel/voxel unit.

5.3. Results

We compare our proposed method with a state-of-the-art learning-based method SYMNet [25], and pairwise optimization-based methods with well-developed software packages including SyN [5], Log-Demons [40] and NiftyReg [24]. SYMNet is considered the leading learning based framework for image registration since it outperforms other existing learning based methods such as the series of VoxelMorph works [6, 10]. Using a deep learning framework to learn the symmetric deformation fields, SYMNet

Table 1. **Comparison with benchmarks.** The top part shows results on our OASIS data setting, dice is averaged over 28 structures. The bottom part shows results on our CANDI data setting, we report both mean dice on 28 and 32 structures. Numbers here are represented as mean or mean $\pm$ std. Note the result on OASIS of SYMNet is from the original paper [25].

OASIS dataset	Avg. Dice (28) ↑	$\mathcal{D}_\psi(\mathrm{x}) \leq 0\ (r^{\mathcal{D}})\downarrow$	$\mathcal{D}_\psi(\mathrm{x}) \leq 0\ (s^{\mathcal{D}})\downarrow$
SYMNet [25]	0.743 ± 0.113	0.026%	-
SyN [1]	0.729 ± 0.109	0.026%	0.005
NiftyReg [2]	0.775 ± 0.087	0.102%	1395.988
Log-Demons [3]	0.764 ± 0.098	0.121%	84.904
NODEO (ours $\lambda_1 = 2.5$)	0.778 ± 0.026	0.030%	34.183
NODEO (ours $\lambda_1 = 2$)	$\mathbf{0.779 \pm 0.026}$	0.030%	61.105
CANDI dataset	Avg. Dice (28) ↑	$\mathcal{D}_\psi(\mathrm{x}) \leq 0\ (r^{\mathcal{D}})\downarrow$	$\mathcal{D}_\psi(\mathrm{x}) \leq 0\ (s^{\mathcal{D}})\downarrow$
SYMNet [25]	0.778 ± 0.091	1.4×10^{-4}%	1.043
SyN [1]	0.739 ± 0.102	0.018%	0.012
NiftyReg [2]	0.775 ± 0.088	0.101%	1395.987
Log-Demons [3]	0.786 ± 0.094	0.071	49.274
NODEO (ours $\lambda_1 = 2.5$)	0.801 ± 0.011	7.5×10^{-8}%	1.574
NODEO (ours $\lambda_1 = 2$)	$\mathbf{0.802 \pm 0.011}$	1.8×10^{-7}%	4.341
CANDI dataset	Avg. Dice (32) ↑	$\mathcal{D}_\psi(\mathrm{x}) \leq 0\ (r^{\mathcal{D}})\downarrow$	$\mathcal{D}_\psi(\mathrm{x}) \leq 0\ (s^{\mathcal{D}})\downarrow$
SYMNet [25]	0.736 ± 0.015	1.4×10^{-4}%	1.043
SyN [1]	0.713 ± 0.177	0.018%	0.012
NiftyReg [2]	0.748 ± 0.160	0.101%	1395.987
Log-Demons [3]	0.744 ± 0.160	0.071	49.274
NODEO (ours $\lambda_1 = 2.5$)	$\mathbf{0.760 \pm 0.011}$	7.5×10^{-8}%	1.574
NODEO (ours $\lambda_1 = 2$)	$\mathbf{0.760 \pm 0.011}$	1.8×10^{-7}%	4.341

achieves reversible registration and yields the best performance in terms of registration accuracy (dice) and quality of the deformation fields measured by $r^{\mathcal{D}}$ and $s^{\mathcal{D}}$. To enable a fair comparison with SYMNet, we employ the same dataset, *e.g.*, OASIS, and data processing practices and use the pre-trained model provided in the official SYMNet repository.

The quantitative results are shown in Table 1. For both the OASIS and CANDI datasets, our method demonstrates a consistent and significant improvement in both mean dice scores over the brain structures as shown in Table 1. Our $r^{\mathcal{D}}$ (all $\leq 0.1\%$) and low values of $s^{\mathcal{D}}$ verify that our method effectively produces diffeomorphic transformations. We visualize the qualitative result of the registration for one example pair image (OASIS-001 and OASIS-002) in Figure 5. We can observe that the ventricular area of the warped brain image obtained with our method is much clearer and does not show any phantom artifacts (cloudy regions where it is black in the fixed image), which demonstrates that our method gives qualitatively better results. See the supplementary material for the full results of benchmarks.

It is important to note that the OASIS dataset generally presents larger deformations between images compared with the CANDI dataset because of the larger age range of the subjects. Therefore we can see lower dice scores for the 28 structures in the OASIS set compared to the CANDI set and more folding in deformation fields. Also, note that 4 out of the 32 structures on CANDI are very small, which are inherently difficult for alignment in registration due to their lower spatial smoothness. This explains the lower dice on the 32 structures compared with that of the 28 structures.

Lastly, we analyze the runtime and model complexity of

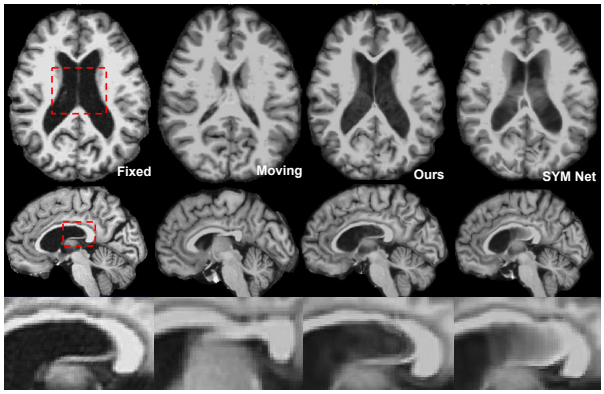

Figure 5. Images showing an example of a registration image pair. Fixed image is OASIS ID001 and Moving image is OASIS ID002. The 3rd column is ours warped image after registration and the 4th column is result of SYMNet [25].

our method and other benchmarks. Experiments shown in Table 1 perform down-sampling on the image before passing it into the network in Figure 2, reducing the runtime by $1/3$ without losing performance. The runtime for SYMNet, SyN, NiftyReg, Log-Demons and ours for one pair image registration are approximately 2s, 25 mins, 70s, 160s, and 80s. While the SYMNet (at inference) takes about 2 seconds to complete registration for one image pair, the pairwise optimization-based methods show difficulty to achieve good performance and fast runtime simultaneously. In comparison, our method takes approximately 80 seconds to register a pair of image (1 step taken), and performs well in terms of similarity and regularity. The number of parameters in our model with the architecture illustrated in Figure 2 is $3/4$ of the number of voxels in the image, demonstrating that the expressive power is well presented without adding model complexity. For the experiments in Table 1, our method uses 3863 MB of memory on a 2080Ti GPU.

6. Conclusion

In this work, we proposed a generic framework for deformable image registration, and investigated the possibility of integrating the merits of both neural netowrks and flow formulations. The resulting models are flexible to incorporate desired transformation regularities and various constraints. We compared our method with benchmarks on several datasets and achieved state-of-the-art results under a variety of metrics. Future works include exploring different ways of time injection into the neural network and applying our methodology to sequential medical data.

7. Acknowledgment

This work was supported by NIH grants R01-NS096720, R01-HL133889, U24-MH114827, RF1-MH124605, RF1-AG069474, NSF IIS 1910308 and Office of Naval Research (ONR) Award No. 14-19-1-2253.

References

[1] https://github.com/ANTsX/ANTsPy. 8

[2] https://github.com/KCL-BMEIS/niftyreg. 8

[3] https://github.com/pyushkevich/greedy. 8

[4] Daniel H Adler, Laura EM Wisse, Ranjit Ittyerah, John B Pluta, Song-Lin Ding, Long Xie, Jiancong Wang, Salmon Kadivar, John L Robinson, Theresa Schuck, et al. Characterizing the human hippocampus in aging and alzheimer's disease using a computational atlas derived from ex vivo mri and histology. *Proceedings of the National Academy of Sciences*, 115(16):4252–4257, 2018. 6

[5] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008. 3, 7

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, A. V. Dalca, and J. Guttag. An unsupervised learning model for deformable medical image registration. In *CVPR*, pages 9252–9260, 2018. 2, 3, 7

[7] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005. 2, 3

[8] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 2018. 4

[9] Hengji Cui, Dong Wei, Kai Ma, Shi Gu, and Yefeng Zheng. A unified framework for generalized low-shot medical image segmentation with scarce data. *IEEE Transactions on Medical Imaging*, 2020. 3

[10] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer, 2018. 3, 7

[11] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis*, 57:226–236, 2019. 2, 3, 5, 7

[12] Adrian V Dalca, Marianne Rakic, John Guttag, and Mert R Sabuncu. Learning conditional deformable templates with convolutional networks. *IEEE TMI: Transactions on Medical Imaging*, 2019. 2, 3

[13] Neel Dey, Mengwei Ren, Adrian V Dalca, and Guido Gerig. Generative adversarial registration for improved conditional deformable templates. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3929–3941, 2021. 3

[14] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012. 7

[15] James C Gee, Martin Reivich, and Ruzena Bajcsy. Elastically deforming a three-dimensional atlas to match anatomical brain images. 1993. 3

[16] Hastings Greer, Roland Kwitt, Francois-Xavier Vialard, and Marc Niethammer. Icon: Learning regular maps through inverse consistency. In *ICCV*, 2021. 3, 7

[17] Tom Z. Jiahao, M. Ani Hsieh, and Eric Forgoston. Knowledge-based learning of nonlinear dynamics and chaos. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2021. 2, 4

[18] Chiyu Jiang, Jingwei Huang, Andrea Tagliasacchi, Leonidas Guibas, et al. Shapeflow: Learnable deformations among 3d shapes. *Advances in Neural Information Processing Systems*, 2020. 4

[19] Ankita Joshi and Yi Hong. Diffeomorphic image registration using lipschitz continuous residual networks. In *Medical Imaging with Deep Learning*, 2022. 3

[20] David N Kennedy, Christian Haselgrove, Steven M Hodge, Pallavi S Rane, Nikos Makris, and Jean A Frazier. Candishare: a resource for pediatric neuroimaging data, 2012. 7

[21] Boah Kim, Dong Hwan Kim, Seong Ho Park, Jieun Kim, June-Goo Lee, and Jong Chul Ye. Cyclemorph: cycle consistent unsupervised deformable image registration. *Medical Image Analysis*, 71:102036, 2021. 3

[22] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507, 2007. 7

[23] Michael I Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2):209–228, 2006. 2, 3

[24] Marc Modat, Gerard R Ridgway, Zeike A Taylor, Manja Lehmann, Josephine Barnes, David J Hawkes, Nick C Fox, and Sébastien Ourselin. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–284, 2010. 3, 7

[25] Tony CW Mok and Albert Chung. Fast symmetric diffeomorphic image registration with convolutional neural networks. In *CVPR*, pages 4644–4653, 2020. 2, 3, 5, 7, 8

[26] Tony CW Mok and Albert Chung. Conditional deformable image registration with convolutional neural network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 35–45. Springer, 2021. 3

[27] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4d reconstruction by learning particle dynamics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5379–5389, 2019. 4

[28] Thomas Polzin, Marc Niethammer, François-Xavier Vialard, and Jan Modersitzki. A discretize–optimize approach for lddmm registration. In *Riemannian Geometric Statistics in Medical Image Analysis*, pages 479–532. Elsevier, 2020. 3

[29] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999. 3

[30] Dinggang Shen and Christos Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration.

IEEE transactions on medical imaging, 21(11):1421–1439, 2002. 3

[31] Zhengyang Shen, Xu Han, Zhenlin Xu, and Marc Niethammer. Networks for joint affine and non-parametric image registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4224–4233, 2019. 3

[32] Zhengyang Shen, Francois-Xavier Vialard, and Marc Niethammer. Region-specific diffeomorphic metric mapping. In *Advances in Neural Information Processing Systems*, 2019. 3

[33] Zhengyang Shen, Zhenlin Xu, Sahin Olut, and Marc Niethammer. Anatomical data augmentation via fluid-based image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 318–328. Springer, 2020. 3

[34] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013. 2

[35] Steven H Strogatz. *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering*. CRC press, 2018. 5

[36] J-P Thirion. Non-rigid matching using demons. In *CVPR*, pages 245–251. IEEE, 1996. 3

[37] Nicholas J Tustison, Brian B Avants, and James C Gee. Learning image-based spatial transformations via convolutional neural networks: A review. *Magnetic resonance imaging*, 64:142–153, 2019. 2

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017. 6

[39] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. In *International conference on medical image computing and computer-assisted intervention*, pages 754–761. Springer, 2008. 3

[40] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009. 3, 7

[41] François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Colin J Cotter. Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241, 2012. 3

[42] Fan Wang, Chunfeng Lian, Zhengwang Wu, Han Zhang, Tengfei Li, Yu Meng, Li Wang, Weili Lin, Dinggang Shen, and Gang Li. Developmental topography of cortical thickness during infancy. *Proceedings of the National Academy of Sciences*, 116(32):15855–15860, 2019. 6

[43] Junshen Xu, Eric Z Chen, Xiao Chen, Terrence Chen, and Shanhui Sun. Multi-scale neural odes for 3d medical image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 213–223. Springer, 2021. 3

[44] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4541–4550, 2019. 4

[45] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396, 2017. 2, 3

[46] Juntang Zhuang, Nicha Dvornek, Xiaoxiao Li, Sekhar Tatikonda, Xenophon Papademetris, and James Duncan. Adaptive checkpoint adjoint method for gradient estimation in neural ode. In *International Conference on Machine Learning*, pages 11639–11649. PMLR, 2020. 4

[47] Juntang Zhuang, Nicha Dvornek, Sekhar Tatikonda, Xenophon Papademetris, Pamela Ventola, and James S Duncan. Multiple-shooting adjoint method for whole-brain dynamic causal modeling. In *International Conference on Information Processing in Medical Imaging*, pages 58–70. Springer, 2021. 2