Medical Image Analysis (2023)

Contents lists available at ScienceDirect

Medical Image Analysis

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

R2Net: Efficient and Flexible Diffeomorphic Image Registration Using Lipschitz Continuous Residual Networks

Ankita Joshi[a], Yi Hong[b,*]

[a]School of Computing, University of Georgia, Athens, 30602, USA
[b]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

ARTICLE INFO

Article history:
Received ***
Received in final form ***
Accepted ***
Available online ***

Communicated by ***

Keywords: unsupervised diffeomorphic image registration, deep residual networks, Lipschitz continuity, stationary and non-stationary velocity fields, multi-scale registration

ABSTRACT

Classical diffeomorphic image registration methods, while being accurate, face the challenges of high computational costs. Deep learning based approaches provide a fast alternative to address these issues; however, most existing deep solutions either lose the good property of diffeomorphism or have limited flexibility to capture large deformations, under the assumption that deformations are driven by stationary velocity fields (SVFs). Also, the adopted squaring and scaling technique for integrating SVFs is time- and memory-consuming, hindering deep methods from handling large image volumes.

In this paper, we present an unsupervised diffeomorphic image registration framework, which uses deep residual networks (ResNets) as numerical approximations of the underlying continuous diffeomorphic setting governed by ordinary differential equations, which is parameterized by either SVFs or time-varying (non-stationary) velocity fields. This flexible parameterization in our Residual Registration Network (R2Net) not only provides the model's ability to capture large deformation but also reduces the time and memory cost when integrating velocity fields for deformation generation. Also, we introduce a Lipschitz continuity constraint into the ResNet block to help achieve diffeomorphic deformations. To enhance the ability of our model for handling images with large volume sizes, we employ a hierarchical extension with a multi-phase learning strategy to solve the image registration task in a coarse-to-fine fashion. We demonstrate our models on four 3D image registration tasks with a wide range of anatomies, including brain MRIs, cine cardiac MRIs, and lung CT scans. Compared to classical methods SyN and diffeomorphic VoxelMorph, our models achieve comparable or better registration accuracy with much smoother deformations. Our source code is available online at `https://github.com/ankitajoshi15/R2Net`.

1. Introduction

Image registration seeks optimal global or local correspondences between images, which is a fundamental task in medical image analysis and has been an active research field for years (Sotiras et al., 2013; Haskins et al., 2020). Deformable image registration is the process of aligning a pair of images by establishing dense local correspondences between them. Such nonlinear transformations bring images to a common coordinate system for information fusion or for further analysis. Traditional deformable image registration methods, such as Large

*Corresponding author: Tel.: +86-132-6258-0581;
e-mail: yi.hong@sjtu.edu.cn (Yi Hong)

Deformation Diffeomorphic Metric Mapping (LDDMM) Beg et al. (2005), Stationary Velocity Fields (SVF) Arsigny et al. (2006), aim to estimate smooth deformation fields based on optimizing a cost function that balances an image matching term with a smoothness regularity on deformation fields. Specifically, deformable image registration prefers a smooth deformation that has a smooth inverse, i.e., a diffeomorphic deformation. Such diffeomorphic image registration provides satisfactory results in registering medical images because of its capability of preserving topologies in medical scans while achieving good matching accuracy. However, these methods suffer efficiency issues in practice due to their high computational cost in solving complex optimization on high dimensional image pairs.

In recent years, deep learning based methods, like VoxelMorph (Dalca et al., 2018), SyMNet (Mok and Chung, 2020a), have been proposed to address the computational challenge faced by traditional methods. These methods fully leverage the advantages of learning from large amounts of data, resulting in a function that maps the embedding of the input image pairs to the deformation fields for alignment. Due to no need of optimization at the inference stage and the availability of GPU-based implementations, such learning-based approaches highly accelerate the inference stage of image registration. Unfortunately, most deep deformable image registration methods sacrifice the diffeomorphic property to some extent because achieving it in the deep learning framework is challenging. One typical solution is based on a scaling and squaring (SS) strategy, as firstly used in VoxelMorph Dalca et al. (2018) and in Krebs et al. (2018), and later in Hoopes et al. (2021); Mok and Chung (2020b). Such SS-based approaches assume that the deformation field is driven by stationary velocity fields (SVFs), which limit the flexibility of registration models to handle large deformations (Vercauteren et al., 2009). Also, using SS for integrating velocity fields is computationally expensive in terms of time and memory costs, making them difficult to handle high-dimensional and high-resolution images.

To relax the assumption of SVFs and provide efficient solutions for diffeomorphic image registration, we consider the relation between deep residual networks (ResNets) with the Eulerian discretization scheme of ordinary differential equations (ODEs) and propose a Residual Registration Network (R2Net) framework. Our R2Net provides an unsupervised diffeomorphic image registration solution that leverages ODE-based parameterization of diffeomorphisms, using both stationary and non-stationary velocity fields to drive deformation fields. As a result, we have two variants, i.e., SVF-R2Net and NSVF-R2Net, to provide flexibility in capturing large deformations. Furthermore, we propose their multi-scale variants, i.e., SVF-MS-R2Net and NSVF-MS-R2Net, to fully leverage the multiple scales at different resolutions of input image pairs and utilize the available computing resources, so that we do not have to compromise on the size of the input data, or on the resolution of generated velocity fields, or on the deep model sizes, while performing registration of large and high-resolution datasets.

1.1. Background

For a given pair of images, namely, the source or moving image denoted by $I_S(x), x \in \mathbb{R}^d$, and the target or reference image denoted by $I_T(y), y \in \mathbb{R}^d, d = 2, 3$, the goal of image registration is to estimate an optimal transformation $\phi \in \mathcal{T}$, within a set of possible transformations $\mathcal{T}$, which aligns the source image I_S to the target image I_T with the lowest energy cost. That is, image registration can be formulated as an optimization problem that aims to minimize the following energy function:

$$\arg\min_{\phi \in \mathcal{T}} \mathcal{M}(I_T, \phi \cdot I_S) + \mathcal{R}(\phi), \tag{1}$$

where $\mathcal{M}$ quantifies the level of alignment between the deformed source and target images, and penalizes the dissimilarity between them. Also, the transformation ϕ needs some desirable properties like smoothness, which are enforced by the regularizer $\mathcal{R}$ that penalizes the non-smoothness of the transformation.

In particular, deformations may be restricted to a space of mappings with certain desirable properties, such that the deformation has to be diffeomorphic. Diffeomorphism is defined as a one-to-one and smooth (i.e., infinitely differentiable) deformation, having a smooth inverse as well Ashburner (2007). To achieve this, the deformation ϕ is typically driven by smooth velocity vector fields Beg et al. (2005). A diffeomorphic deformation field can be treated as an integral of a smooth time-varying velocity field $v_t, t \in [0, 1]$ via the following ODE:

$$\frac{d}{dt}\phi_t = v_t \circ \phi_t, \quad \phi_0 = id. \tag{2}$$

Here, ϕ_t indicates the deformation at the time point t. ϕ_0 is the identity map and ϕ_1 is the deformation that transforms the source image I_S to the target image I_T. Given the velocity fields $\{v_t\}$ and ϕ_0, the solution ϕ_1 is the numerical integration of Eq. (2), given as

$$\phi_1 = \phi_0 + \int_0^1 v_t(\phi_t)dt. \tag{3}$$

Diffeomorphic deformations can be also parameterized using SVFs as proposed in Arsigny et al. (2006), where the velocity is constant over time ($v_t = V, \forall t$) and is governed by the ODE:

$$\frac{d}{dt}\phi_t = V(\phi_t). \tag{4}$$

For this simplified SVF version, the solution of $\phi(t)$ is represented as the exponential of the velocity V, given as

$$\phi(t) = \exp(tV). \tag{5}$$

1.2. Related Work

Traditional Optimization-Based Methods. Traditional deformable image registration methods aim to find the transformation trajectory between the source and target image pairs, which is an ill-posed two-boundary-value problem since the solution is not-unique and many trajectories could achieve the same goal. To address this issue, various regularizations are used according to some physical assumptions on how the image is allowed to deform, which determines how the estimated deformations will be regularized. Multiple methods (Ashburner, 2007; Glocker et al., 2008; Yeo et al., 2010; Zhang et al., 2017; Avants et al., 2008) have been proposed, which constrains the deformations

to be symmetric, diffeomorphic, volume preserving, etc. To estimate diffeomorphic deformations, classical registration algorithms are proposed to successfully align image pairs with smooth deformations, such as Large Deformation Diffeomorphic Metric Mapping (LDDMM) (Beg et al., 2005; Ceritoglu et al., 2009; Joshi and Miller, 2000), Stationary Velocity Field (SVF) (Arsigny et al., 2006; Hernandez et al., 2007; Ashburner, 2007; Higham, 2005), DARTEL Ashburner (2007), diffeomorphic demons Vercauteren et al. (2009), and symmetric normalization (SyN) Avants et al. (2008).

Among diffeomorphic image registration frameworks, LDDMM and SVF are both fluid-based image registration approaches; the deformations in the LDDMM framework are driven by time-dependent velocity fields, while SVF deformations are driven by a constant velocity field. While SVF has advantages in lower computational memory cost and faster computation, the setting of stationary velocity fields limits its ability to handle large deformations Arsigny et al. (2006). Also, despite its satisfactory results, in the SVF framework the $\exp(\cdot)$ is not surjective, i.e., not all images can be reached by an $\exp(V)$ from the source image, which makes the registration under large deformations uncertain. In other words, SVFs are less flexible, since one can only invert for a subset of deformation maps living on the manifold of diffeomorphisms. It has been shown that SVFs are only adequate for registration problems involving two topologically similar images Mang and Ruthotto (2017). Instead, non-stationary velocity fields are beneficial for applications involving large and highly non-linear transformations, even for the registration of longitudinal data with large motions over time Mang and Ruthotto (2017).

In summary, most traditional methods formulate the image registration problem on Riemannian manifolds and employ an iterative optimization procedure to estimate the deformations for every image pair. Their formulations are elegant and achieve satisfactory registration results in most cases. However, the complex mathematical formulations and high-dimensional optimization make such algorithms computationally expensive. In practice, having a fast, memory-efficient, and even parallelizable image registration approach is critical in clinical applications; so that we can tackle image alignments in a short amount of time or handle high-resolution volumes on a large scale.

Learning-Based Methods. Deep learning based approaches have been shown to provide fast and efficient solutions in the inference stage with highly parallelizable implementations executed on GPUs. Existing deep learning based models tackle the challenging problem of image registration using either supervised (Rohé et al., 2017; Cao et al., 2017; Yang et al., 2017; Fan et al., 2019) or unsupervised techniques (Li and Fan, 2017; De Vos et al., 2019; Vos et al., 2017; Balakrishnan et al., 2019; Dalca et al., 2018) Both methods learn the mapping from image pairs to their deformation fields, while their main difference lies in whether the true deformations are provided for training or not. In particular, supervised approaches maintain the diffeomorphic property, which is inherited from the original image registration models, e.g., LDDMM. However, they require extra effort of obtaining the ground-truth deformations. Meanwhile, the registration accuracy of these supervised methods is limited by that of the provided deformations.

On the other hand, unsupervised approaches reformulate the traditional image registration model in the deep learning framework. Most unsupervised methods make use of spatial transformer networks (STNs) Jaderberg et al. (2015) to allow for differentiable warping and interpolation, which can be integrated using deep neural networks. Unsupervised methods have demonstrated promising image registration accuracy in a variety of image registration tasks Hoopes et al. (2021); Hering et al. (2021); Wu et al. (2022). A simple way to maintain the diffeomorphic property is introducing an integration layer into the network, which solves Eq. 5 in the SVF framework using the scaling and squaring methodology (Higham, 2005; Arsigny et al., 2006). Such SVF-based formulation of solving image registration has been widely used (Krebs et al., 2018; Dalca et al., 2018; Mok and Chung, 2020b,a; Joshi and Hong, 2022).

However, similar to the traditional SVF-based registration methods, although SVF-based registration algorithms produce diffeomorphic deformations they face the same drawback as discussed in the SVF-based traditional algorithms, which are not able to handle large deformations. Along with this issue, another drawback faced by SVF-based deep learning methods is the computational cost of the scaling and squaring step. To reduce the memory cost, the integration step is often carried out using a half-scale of the original size of velocity fields, which leads to missing details in the deformation fields. These issues result in sub-optimal solutions for diffeomorphic image registration using SVF-based deep learning methods.

Currently, there are limited deep learning-based works tackling the diffeomorphic image registration problem based on the parameterization of time-varying velocity fields (Wu et al., 2022; Xu et al., 2021). These two methods solve the deformation equation based on neural ordinary differential equations (NODEs) Chen et al. (2018) and model the problem as learning the optimizer in the image registration formulation. These methods are not as fast as the current unsupervised methods, since they have to perform registration (or learn the correct optimizer settings) for every new image pair in the inference stage. While residual networks (ResNets) based model has been proposed for modeling deformations Ben Amor et al. (2021), they parameterize time-dependent affine velocity fields in order to perform affine diffeomorphic registration on shapes. However, their architecture needs further exploration to fit deformable image registration tasks. ResNets are also used in Yang et al. (2021) to learn deep multi-scale residual representations to boost registration accuracy.

Overall, there is a lack of deep registration models that not only provide diffeomorphic solutions but also use the parameterization of time-varying velocity fields.

Multi-Scale Extensions. Despite unsupervised deep learning-based registration being a popular choice for aligning images, they still face significant challenges in achieving accurate and efficient solutions for the task of diffeomorphic image registration. Firstly, current registration methods are based on a variety of variational auto-encoders (VAEs) or UNets Ronneberger et al. (2015), which suffer from producing over-smooth upsampling velocity fields or learning only low-level statistics rather

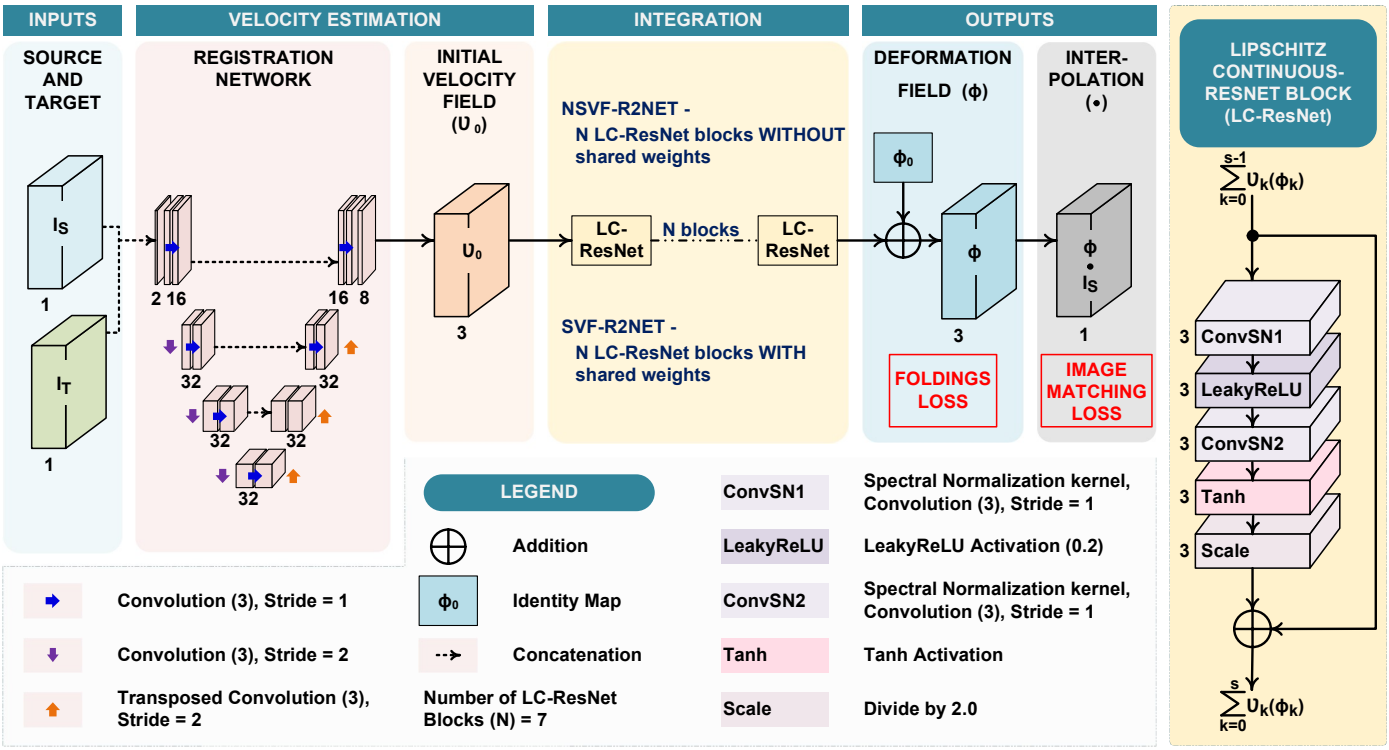

Fig. 1. Overview of our proposed network. The Lipschitz Continuous ResNet block (LC-ResNet) without shared weights corresponds to the time-varying velocity field, while the one with shared weights corresponds to the stationary velocity field. Both models use N=7 LC-ResNet blocks; s in the right LC-ResNet block indicates the s-th block, $s = 1, 2, \cdots, N$.

than high-level semantics (Razavi et al., 2019; Nalisnick et al., 2018). Secondly, as mentioned before, the integration layer is computationally expensive, resulting in a compromise in models or data inputs, such as integrating the velocity fields at half-scale, downsampling the input data size, or reducing the size of the deep learning models. Lastly, registration networks handle high-dimensional medical image volumes, which have a large number of parameters to optimize and increase the difficulty in obtaining an accurate and efficient solution.

A potential solution to address the above issues is to incorporate a hierarchical approach based on a multi-resolution strategy, i.e., using a coarse-to-fine optimization scheme. Existing methods follow a multi-level optimization strategy (Hering et al., 2019; Mok and Chung, 2020b) or apply a multi-scale upsampling design after feature extraction Krebs et al. (2019); Mok and Chung (2020b). However, these approaches still have difficulty in handling large volumes under multiple scales, due to the constraints of limited resources. It is a challenging task to handle high-resolution image volumes while using a multi-scale strategy to leverage limited memory and improve registration accuracy at the same time.

1.3. Contribution

In this paper, we present a Residual Registration Network (R2Net), a diffeomorphic image registration framework, which has the flexibility of generating smooth deformations driven by either stationary or time-varying velocity fields. Also, its multi-scale extension (MS-R2Net) provides the possibility of handling registration between high-dimensional and high-resolution image volumes efficiently. This paper extends a preliminary version of two works presented at the Medical Imaging with Deep Learning (MIDL) 2022 Joshi and Hong (2021) and the IEEE International Symposium on Biomedical Imaging (ISBI) 2022 Joshi and Hong (2022).

In this journal version, we provide theoretical extensions, new results, analysis, and discussion. Theoretically, we study new ways of generating initial velocity fields of our R2Net model, evaluate the correctness of using ResNets to solve ODEs, and have a natural extension to a multi-scale architecture for large image volumes. Experimentally, we show extensive analysis of different architectures and evaluate our models on four datasets with 3D image volumes, including MRI cardiac scans, MRI brain scans, and lung CT images. Our contributions can be summarized as follows:

- We propose a deep diffeomorphic image registration framework, which has flexible parameterizations of deformation fields driven by either stationary velocity fields, SVF-R2Net, or time-varying (non-stationary) velocity fields, NSVF-R2Net. The non-stationary version improves the flexibility of capturing large deformations.
- We demonstrate two efficient multi-scale extensions, MS-NSVF-R2Net and MS-SVF-R2Net, which adopt a dual-phased learning strategy to tackle high-resolution image volumes while fully utilizing computational resources.
- We provide a thorough study of different medical image modalities and anatomies, as well as the time and mem-

ory cost of our registration models. Experiments show our models' efficiency and effectiveness in maintaining diffeomorphic properties and registration accuracy, without sacrificing the original resolution of input images, the integration scale of deformations, or the model size of the backbone deep neural networks.

The remainder of the paper is organized as follows. Section 2 introduces our method and discusses the network architectures. Section 3 describes our conducted experiments, with details regarding the used datasets, settings, and also discussions on experimental results. We conclude our analysis in Section 4.

2. Method

Figure 1 presents an overview of our proposed R2Net. Given a pair of images, we adopt a U-Net to estimate the initial velocity fields, which are integrated via a well-designed residual network to obtain the corresponding deformation fields. Our goal is to use these deformation fields to deform the source image and match the target image as closely as possible; meanwhile, we prefer diffeomorphic deformation fields that preserve the topology of object structures in image scans. Since the residual network (ResNet) plays an essential role in achieving the diffeomorphic deformation, we start with the discussion of our customized ResNet blocks, followed by a detailed description of the basic R2Net and its multi-scale extension.

2.1. ResNet Blocks with Lipschitz Continuity

ODE Solver in Registration Based on ResNet Blocks. In diffeomorphic image registration, solving the ODE (see Eq. 1) that governs the evolution of deformation fields is a core component, which needs to be implemented in the framework of deep neural networks. To address this issue, we employ residual deep networks (ResNets) as numerical schemes of differential equations to integrate stationary or non-stationary velocity fields, since ResNets have connections to the Euler's method used for solving ODEs in scientific computing (Haber and Ruthotto, 2017; Weinan, 2017; Ruthotto and Haber, 2020). For instance, an interpretation of ResNets as incremental flows of diffeomorphisms was recently published in Rousseau et al. (2020) in the context of supervised learning with the application to image classification. Also, there are multiple works that provide insights on ResNets from an aspect of ODEs or partial differential equations (PDEs) Lu et al. (2018); Ruthotto and Haber (2020). These works relate the incremental mapping (residual blocks) defined by ResNets as numerical schemes of differential equations used in diffeomorphic registration models, especially to LDDMM Rousseau et al. (2020).

According to these theoretical insights, we use ResNets to solve the equation of integrating flows given by Eq. 3 and its stationary form as given by Eq. 4. In a typical ResNet-based architecture, given an initial value, like the identity deformation, each learnable residual block takes current states one step forward, which incrementally maps the learned embedded features onto a new space, with an update that takes the form:

$$x_{l+1} = x_l + \mathcal{F}(x_l, \theta_l), \tag{6}$$

where x_l is the input to the l^{th} residual unit and θ_l are the trainable network parameters associated with the l^{th} residual unit. That is, a deep residual network can be also viewed as a discretization of a dynamical system governed by a first-order ODE, where the network layers are viewed as time steps and the network parameters, $\{\theta_l\}$, are viewed as the control to optimize Liu and Theodorou (2019). Such network-based discretization of velocity fields $\{v_t\}$ in Eq. 3 and Eq. 4 can be treated as a simple combination of basis functions using a ResNet mapping in Eq. 6, which can be rewritten as:

$$v_{t+1}(\phi_{t+1}) = v_t(\phi_t) + \mathcal{F}(v_t(\phi_t), \theta_t). \tag{7}$$

Similarly, each mapping block t is viewed as a time-step, and θ_t represents the network parameters.

In this way, the entire ResNet architecture implements the composition of a series of incremental deformation mappings, which is a discretized version of Eq. 3 by replacing the integral with a summation. This interpretation makes $\mathcal{F}(\cdot, \theta_t)$ to be a parameterization of a deformation flow field, and a series of identical ResNet blocks $\mathcal{F}(.; \theta_t)$ integrate the time-dependent velocity fields. Moreover, the interpretation of $\mathcal{F}(\cdot, \theta_t)$ with shared weights is viewed as the numerical implementation of the exponential of velocity fields in Eq. 4, which is the analytical solution of deformation fields driven by stationary velocity fields Rousseau et al. (2020). Overall, the residual blocks $\{\mathcal{F}(\cdot, \theta_t)\}$ resemble the forward Euler method by composing a series of incremental deformation mappings with a given initial value Weinan (2017), which helps us solve the ODEs (Eq. 3 and Eq. 4) required in image registration.

Lipschitz Continuous ResNet (LC-ResNet) Blocks. As pointed out in Younes (2010) (Theorem 8.7), a smoother v yields a smoother deformation ϕ. In the LDDMM framework, an admissible Hilbert space of velocity fields with adequate smoothness conditions is defined as a reproducing kernel Hilbert space (RKHS). A recent work connects neural networks and RKHS Bietti and Mairal (2019) (Proof for Proposition 14), which shows that convolutional neural networks with homogeneous activation functions (e.g., tanh) fall under RKHS. Also, according to the Cauchy-Lipschitz theorem, under adequate smoothness assumptions on the velocity fields v, a Lipschitz continuous integration over time is a well-defined mapping on the space of time-dependent diffeomorphisms. Recently, a number of works advocate the importance of Lipschitz continuity in assuring the generalizability of deep learning models to the perturbation of outputs Yoshida and Miyato (2017); Gouk et al. (2021).

Based on the above theories, we employ the method proposed in Miyato et al. (2018) to enforce Lipschitz continuity in our Residual blocks, i.e., using the spectral normalization for each convolutional layer of the residual blocks. This operation normalizes the spectral norm of the weight matrix $\mathcal{W}$, i.e., $\hat{\mathcal{W}}_{SN} := \frac{\mathcal{W}}{\delta(\mathcal{W})}$, where the function $\delta(\mathcal{W})$ computes the spectral norm of $\mathcal{W}$. After the spectral normalization, the network weights satisfy the Lipschitz constraint $\delta(\hat{\mathcal{W}}_{SN}) = 1$. Having this condition implies that the Hilbert space of admissible velocity fields is an RKHS. In particular, the authors adopt a fast approximation by using the power iteration method (Miy-

ato et al. (2018), Algorithm 1), which replaces the weights $\mathcal{W}$ with $\delta(\mathcal{W})$, the largest singular value of $\mathcal{W}$. This strategy is used to avoid the added computational complexity of computing the singular valued decomposition to compute the eigenvalues. We follow the Keras implementation[1] for computing the spectral norm of the weight matrix.

The architecture of our Lipchitz continuous residual network (LC-ResNet) block is shown on the right of Fig. 1. Each block consists of a convolutional layer, which uses spectral normalization (SN) as a regularizer on the weights. This SN-enforced convolution ensures that the velocity field v is a Lipschitz continuous mapping. After the convolution layer, a point-wise Leaky ReLu activation function is applied to introduce the nonlinearity into the network, which is followed by a second convolutional layer with spectral normalization. We control the magnitude of the velocity field by using a *tanh* layer followed by a scaling layer which further scales the per voxel value by a factor of 2. There are N such ResNet blocks, which indicate N time steps of integration, and each block generates the velocity field at the t-th time step. As a result, we use LC-ResNet blocks to map a smooth initial velocity field to the desired diffeomorphic deformations.

2.2. Basic Model: R2Net

Building upon the LC-ResNet blocks, we design two versions of our Residual Registration Network (R2Net): Stationary Velocity Field based Residual Registration Network (SVF-R2Net) and Non-Stationary Velocity Field based Residual Registration Network (NSVF-R2Net). These two networks share the same architecture as shown in Fig. 1, and their main difference is whether the weights are shared within the residual blocks. Our basic R2Net takes an image pair as input, maps them to an initial velocity field, integrates velocity fields to generate a diffeomorphic deformation, and deforms the source image using generated deformation to match the target image.

Initial Velocity Field Estimation. Assume the source and target images I_S and I_T have the same size of $H \times W \times D$, which are defined over an $n = 3$ dimensional spatial domain $\Omega \subset \mathcal{R}^3$. Different from traditional optimization-based image registration algorithms, the deep registration network uses a convolutional neural network (CNN) to learn the mapping from the source and target images, which is represented by an initial velocity field. The adopted CNN is based on a U-Net Ronneberger et al. (2015), which takes the concatenated source and target images as inputs and predicts a dense initial velocity field v_0. The encoder part of the U-Net uses $3D$ convolutions with a kernel size of $3 \times 3 \times 3$ and a stride of 1, followed by the Leaky ReLU activation functions and $3D$ convolutions with a kernel size of $3 \times 3 \times 3$ and a stride of 2 for downsampling. The number of filters used in each layer is shown in Figure 1. The decoder mirrors the encoder with the convolution replaced by the transposed convolution for upsampling, along with a skip connection to the corresponding encoder block at the same resolution level.

Integration Using LC-ResNet Blocks. Since deformations are driven by initial velocity fields under the government of Eq. 2, this component of R2Net is to learn the mapping from the generated initial velocity field to a diffeomorphic deformation field, i.e., the integration of deformations driven by velocity fields. As discussed in Sec. 2.1, we utilize the ResNet as numerical schemes of differential equations and relate the incremental mappings defined by LC-ResNet blocks to diffeomorphic registration models. The LC-ResNet blocks without sharing weights, which model the case of time-varying velocity fields, constitute the integration component of our NSVF-R2Net. Similarly, we utilize LC-ResNet blocks with shared weights to model the stationary velocity fields, which constitute the integration component of the SVF-R2Net.

Image Interpolation. To measure the goodness of image matching result, we need to generate the deformed source image using the integrated deformation field ϕ and compare it with the target image. In order to warp the original source image volume, we use a linear interpolation layer, which basically computes for every voxel, a corresponding source image voxel location. The values at neighboring voxels are interpolated in order to obtain the intensity value for the location in the target image. This is a differentiable operation and therefore allows for the backpropagation of errors.

2.3. Loss functions

The objective function of traditional image registration includes two terms, i.e., image matching, measuring the similarity between the deformed source and target images, and the smoothness regularizer, measuring the smoothness of the velocity fields. Deep registration networks have similar loss functions; therefore, our networks also have a similarity loss $\mathcal{L}_{sim}$ and a regularizer loss $\mathcal{L}_{reg}$.

Similarity Loss. When measuring the distance between the deformed source image $\phi \cdot I_S$ and the target image I_T, the mean squared error (MSE) is a commonly-used loss function, which is given as

$$\begin{aligned}\mathcal{L}_{sim}(I_T, \phi \cdot I_S) &= \|I_T - \phi \cdot I_S\|_2^2 \\ &= \frac{1}{|\Omega|} \sum_{x \in \Omega} (I_T(x) - \phi(x) \cdot I_S)^2. \end{aligned} \tag{8}$$

Here, Ω is the image spatial domain and $|\Omega|$ indicates the number of pixels or voxels in an image. The MSE loss is suitable for image pairs that have similar intensity distributions and local contrast, like our brain and cardiac image scans.

However, for the task of lung CT image registration, the intensity at the corresponding points of the source and target images vary, because of the altered density of the lung tissue during breathing. For such cases, the MSE is not a desired similarity loss, instead, we use the normalized gradient fields (NGF) loss following the approach proposed in Rühaak et al. (2017); Hering et al. (2019):

$$\mathcal{L}_{sim}(I_T, \phi \cdot I_S) = \int_{\omega} 1 - \frac{\langle \nabla(\phi \cdot I_S), \nabla I_T \rangle_\epsilon^2}{\|\nabla(\phi \cdot I_S)\|_\epsilon^2 \|\nabla I_T\|_\epsilon^2}, \tag{9}$$

where ω is the image domain limited to the region within the lung mask, with $\langle f, g \rangle_\epsilon^2 = \sum_{j=1}^{3} (f_j g_j + \epsilon^2)$, $\|f\|_\epsilon = \sqrt{\langle f, f \rangle_\epsilon}$. Here, $\epsilon > 0$ is the edge parameter that is used to suppress small

[1] https://github.com/IShengFang/SpectralNormalizationKeras

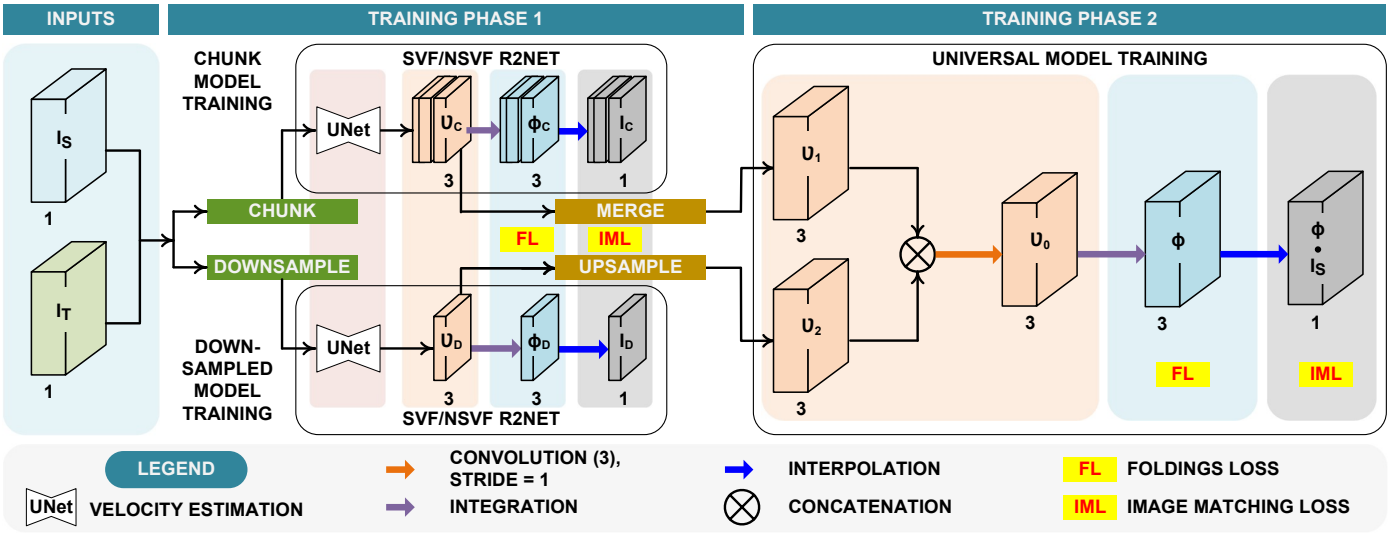

Fig. 2. Multi-scale R2Net architecture with the two-phase registration process. In Phase 1, the chunk and downsampled branches are trained. In phase 2, the universal model is trained. For both Phase 1 and 2, the baseline architecture used is either SVF-R2Net or NSVF-R2Net as shown in Fig. 1.

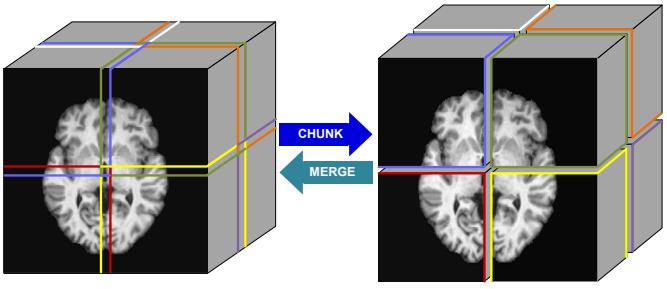

Fig. 3. The chunking process for generating the inputs of the chunk branch in the MS-R2Net architecture. We have an overlap of 2 pixels during the chunking process. During the merging process, we average the values in the overlap region to obtain the original-sized volume. In this figure we can see the original volume being chunked into eight subvolumes.

image noise. We set it to be 1 as used in Hering et al. (2021). This similarity measure helps estimate an accurate alignment, since it will avoid the noise within the lung CT scans.

Regularizer Loss. To encourage the smoothness of the deformation field, we regularize its determinant of the Jacobian, which shows how much each image pixel or voxel was stretched or compressed in the deformation. Typically, a determinant larger than 1.0 indicates an expansion at the pixel/voxel location, and one between 0 and 1 signifies compression, whereas having negative values in the Jacobian determinant means the foldings happening at those voxel positions in the deformations. We aim to keep the Jacobian determinant of the deformation field to be positive to avoid having foldings. We integrate this loss into our overall objective function:

$$\mathcal{L}_{reg}(\phi) = \frac{1}{|\Omega|}\sum_{x\in\Omega} 0.5(\left|\mathcal{J}\left(\phi(x)\right)\right| - \mathcal{J}\left(\phi(x)\right)), \qquad (10)$$

where $\mathcal{J}(\phi(x))$ is the determinant of the Jacobian of the deformation field at its location x.

2.4. *Multi-Scale Variant: MS-R2Net*

As discussed before, most SVF-based deep learning solutions use the scaling and squaring step for integration and solve the registration problem by reducing the number of unknowns either through a coarse parameterization (reducing input/model size) or by using a coarse grid (reducing the velocity field size). Such approximations and simplifications can result in sub-optimal registration quality Himthani et al. (2022); Mang et al. (2019); Brunn et al. (2021). It has also been recently shown that image registration at higher image resolution is more accurate Himthani et al. (2022). Building a multi-scale design has been demonstrated to be an efficient way to improve registration accuracy by avoiding bad local minima Mok and Chung (2020b), but such approaches are unable to process large volumes at multiple scales under limited resource constraints. Accurate registration method for high-dimensional multi-resolution $3D$ images has immense potential for time-sensitive medical studies. Existing methods that follow a multi-level optimization strategy Mok and Chung (2020b); Hering et al. (2019) or apply a multi-scale upsampling design for feature extraction are still unable to process large volumes at multiple scales with available computing resources. Here, we introduce our multi-scale architectures SVF-MSR2Net and NSVF-MSR2Net, as shown in Fig. 2. In both networks, we employ two phases for learning, to further improve the run-time and better utilize the available resources. As a result, we obtain a trade-off between fully leveraging the available data under limited computing resources while at the same time capturing multi-scale features to achieve better registration accuracy.

Phase One: Local and Global Learning. The goal of phase one is to learn initial velocity fields at a reduced size, for instance, half of the original size at each dimension. Chunking and downsampling images are two typical strategies to achieve the size reduction, so that, we can learn the initial velocity fields globally and locally. Therefore, phase one consists of two branches, i.e., a local branch that handles the registration

for the chunked subvolumes of the original input image pairs, and a global branch that handles the registration of the downsampled volumes of the original input volumes.

Chunk branch. The input to the chunk branch is the subvolumes of the original image volumes. We use a chunking process as shown in Fig. 3 to divide/chunk the original 3D source and target volumes into a bunch of subvolumes. For instance, if the original 3D source and target images, I_S and I_T, are of size $H \times W \times D$, then the chunk branch receives each input as k^3 subvolumes with a resolution of $H/k \times W/k \times D/k$. In our experiments, we set $k = 2$. The subvolume from the source image is then concatenated with the corresponding one at the target image to form the input pair of the chunk branch, which follows the basic R2Net. This chunk branch is trained independently until it converges.

Downsampled branch. The image pairs input to the downsampled branch are the downsampled volumes of the original image pairs. Similarly, the downsampled volumes are of sizes $H/d \times W/d \times D/d$ each, where $d = 2$ in our experiments. This branch also takes the basic R2Net architecture and is trained separately until convergence.

In this way, the local branch (chunk sub-volume learning) and the global branch (downsampled volumes learning) are trained separately. Both branches work on image pairs with reduced size, but the chunk branch maintains the original image spacing (i.e., at the original resolution), while the downsampled branch works on a lower resolution. Thanks to the reduced image size, this training process converges faster with reduced memory utilization. After training, for each image pair, we obtain k^3 initial velocity fields $\{v_C\}$ from the local branch, and one initial velocity field v_D from the global branch, which will be used in the next phase of learning.

Phase Two: Integration at Original Scale. In this integration branch, we utilize the local details at the original resolution from the chunk branch and the global context from the downsampled branch. In the second phase, the integration branch takes the combination of the outputs from the two branches estimated in the first phase, resulting in the final output at the original scale of the input image volumes.

Firstly, the multiple velocity fields $\{v_C\}$ from the chunk branch are merged back to the original size, using the merging process as shown in Fig. 3. We keep some overlaps between chunks to reduce the discontinuity at the boundary of each chunk. When merging back, we average the velocity fields at these overlapping regions. As a result, we have the velocity fields v_1 that are of the same resolution as the original input volume, with a size of $H \times W \times D \times p$, $p = 2$ for 2D images and $p = 3$ for 3D images. Similarly, the velocity fields from the downsampled branch v_D are upsampled back to the original resolution, resulting in v_2 with the same size of $H \times W \times D \times p$.

Thus, v_1 and v_2 along with the original source and target images I_S and I_T become the inputs of the integration branch. We concatenate the two velocity fields and pass them to a convolution layer, as shown in Fig. 2. The merged velocity fields are then integrated using LC-ResNet blocks to generate the deformation fields at the original resolution, which warp the original source image to match the original target image. Since phase two has no need of a U-Net-like network to estimate the initial velocity, which saves a lot of GPU memory and makes it possible to integrate deformations at the original high resolutions. As a result, we have multi-scale extensions for both SVF-R2Net and NSVF-R2Net, which are abbreviated as MS-SVF-R2Net and MS-NSVF-R2Net.

3. Experiments and Results

We evaluate our methods on the tasks of brain MRI inter-subject registration, cardiac MRI intra-subject registration, and thoracic CT intra-subject registration, using the following four public datasets.

3.1. Datasets

ACDC. The Automated Cardiac Diagnosis Challenge (ACDC) dataset at STACOM 2017 Bernard et al. (2018) has 150 patients, each having their respective end-diastole (ED) and end-systole (ES) phases. This dataset also contains segmentation for three structures, namely, the left ventricular cavity, myocardium, and the right ventricle. We perform the task of registering the ED frames to the ES frames. To evaluate our methods, we divide 150 subjects of this dataset into sets of 108 for training, 12 for validation, and 30 for testing. Each scan is resampled with a spacing of 1.25×1.25×10 mm, the image intensity is normalized to [0, 1] and then images are cropped to 176×176×16.

EMPIRE10. The EMPIRE10 challenge Murphy et al. (2011) provides a lung dataset that contains 30 pairs of intra-patient thoracic CT scans. Each pair belongs to a single subject. We divide these 30 pairs, subject-wise, into two sets, 20 for training and the rest 10 for testing. Due to the limited training samples, we extend the training set to 200 pairs by random flips of the image scans. Lung regions are cropped from the data with the help of the provided lung masks, the image intensity is normalized to [0, 1], and the volumes are resized to 192×192×192 with a spacing of 1×1×1 mm.

OASIS 3D. We use the OASIS Brain MRI dataset (Marcus et al., 2007; Hoopes et al., 2021), which contains T1-weighted MRI scans for 414 subjects preprocessed with skull-stripping, bias correction, registered and resampled into the freesurfer's Talairach space. After pre-processing, each 3D volume has dimensions of $160 \times 192 \times 224$ with a spacing of 1.25×1.25×1.25 mm. We divide the 414 subjects into sets of 264, 50, and 100 as our training, validation, and test sets, respectively. We then randomly pair images in each set and choose 350 pairs for training, 50 for validation, and 100 for testing.

IBSR18. We use the IBSR18 dataset Valverde et al. (2015), which is pre-processed with skull-stripping, bias correction, registered, and positionally normalized into the Talairach orientation. The MR brain data sets and their manual segmentations are provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available online[2]. It consists of T1-weighted scans for 18 subjects with dimensions

[2] http://www.cma.mgh.harvard.edu/ibsr/

of $256 \times 128 \times 256$, we crop and then pad each scan to the dimension of $224 \times 224 \times 224$ with a voxel spacing of $1 \times 1 \times 1$ mm. Half of this size is the largest one that multi-scale VoxelMorph can handle in our GPU. We divide the subjects into sets of 11, 5, and 3 as our training, test, and validation sets. We then randomly pair images in each set to produce 110 pairs for training, 6 pairs for validation, and 20 pairs for testing.

3.2. Evaluation Metrics and Settings

Similarity. To evaluate the registration performance in terms of image matching, we use two metrics, the root mean squared error (RMSE) and the Dice score. The RMSE measures the intensity difference between the deformed source image and the target image, while the Dice score measures the structure difference using segmentation masks, i.e., computing the overlapping between the deformed segmentation mask of the source image and that of the target image.

Smoothness. To check the diffeomorphic property of our registration model and evaluate the smoothness of its estimated deformations, we compute the Jacobian determinants of the deformation fields and measure the number of voxels with negative Jacobian determinants. We report the absolute number of such voxels, noted as γ^{abs}.

Computational Cost. To evaluate the effectiveness of our models, we measure the resource utilization and report the training time, test time, and memory cost for all our experiments. All the reported values for the time and memory costs are computed by averaging the outputs of 10 runs.

Baselines. We compare our models with a classical registration method, i.e., Symmetric Normalization (SyN) in the ANTsPy package Avants et al. (2008) which is a top-performing classical image registration algorithm Klein et al. (2009). For SyN settings, we use cross-correlation (CC) and Gaussian smoothing with sigma values at each level of (9, 0.02, 0.02), with three scales and 201 iterations, which are optimal for our tasks. Another baseline algorithm is VoxelMorph Dalca et al. (2018), the one with the scaling and squaring approach as a differential layer in the registration network. We use the default settings of VoxelMorph as provided by their implementation[3], except for the heart dataset, where we follow the settings as given in Krebs et al. (2019) and set $\sigma = 0.05$ and $\lambda = 50000$.

Other Settings. We implement our models in Keras with Tensorflow as backend Abadi (2016). For all the experiments we use the Adam Kingma and Ba (2014) optimizer with a learning rate of $1 \times e^{-4}$. All experiments are carried out on $3D$ datasets, which are divided subject-wisely for evaluation. All the deep learning based experiments are trained using a single NVIDIA GeForce TITAN X GPU.

3.3. Experimental Results

Studying Ways of Velocity Estimation. In Fig. 1, we directly use a U-Net output as the estimation of initial velocity fields. Before reaching this final design, we explore other choices like a variational Bayes approach proposed in Dalca et al. (2018); Krebs et al. (2018, 2019), which uses a probabilistic U-Net. We compare the following six different ways to estimate the initial velocity fields, resulting in the non-probabilistic U-Net used in our R2Net. This experiment is performed on NSVF-R2Net and tested on the OASIS 3D dataset under the same settings. The weights of all terms in the loss function are set to 1. All networks are trained until convergence, with a maximum 200 number of epochs.

- *NSVF-R2Net-Prob*. This architecture follows a variational Bayes approach and utilizes a probabilistic U-Net to generate the initial velocity field through a sampling layer as used in Dalca et al. (2018). This architecture is the same as introduced in our previous work Joshi and Hong (2021). Similarly, we use a KL divergence loss on regularizing the mean and variance of the sampled velocity fields to follow a normal distribution. This KL loss helps obtain smooth initial velocity fields.

- *NSVF-R2Net-noKL*. In this architecture, we remove the velocity sampling layers from the NSVF-R2Net-Prob model. Instead, we directly estimate the velocity fields without using the KL loss. That is, this network has no regularizer on velocity fields, and their smoothness is indirectly enforced by our regularizer loss on the Jacobian determinant of the deformation fields. Besides, similar to NSVF-R2Net-Prob, we use the upsampling layers in the U-Net architecture.

- *NSVF-R2Net-Gauss*. Based upon the above NSVF-R2Net-noKL network, we introduce a Gaussian smoothing layer after the output of the U-Net, in order to enforce smoothness on the generated initial velocity fields.

- *NSVF-R2Net-Multi*. Along with the Gaussian smoothing layer, we have an additional diffusion regularizer on the spatial gradients of the initial velocity fields, given as $\sum_{p \in \Omega}(\|\nabla v(p)\|_2^2)$. That is, we apply multiple regularizers on the initial velocity fields to enforce their smoothness.

- *NSVF-R2Net-nonProb*. In this architecture, we regress back to the NSVF-R2Net-noKL version, but replace the upsampling layers in the U-Net decoder with Conv3D Transpose layers. We do not use any regularizer on the velocity fields, but we assume the convolution layers have the ability to smooth out them from coarse to fine by following the decoder.

- *NSVF-R2Net-nonProb-Gauss*. We further explore the necessity of adding additional regularizers on velocity fields. Therefore, we add a Gaussian smoothing layer into NSVF-R2Net-nonProb to smooth the initial velocity fields.

The experimental results are reported in Table 1. The comparison between NSVF-R2Net-Prob and NSVF-R2Net-noKL shows that removing the KL loss would reduce the smoothness of the velocity fields while increasing the accuracy of image matching, as expected. A Gaussian smoothing layer would help increase both matching accuracy and deformation smoothness; however, the additional diffusion regularizer makes things

[3] https://github.com/voxelmorph/voxelmorph

Table 1. Top to bottom: ablation study for studying the estimation of velocity fields; results of all the compared algorithms across three popular datasets namely ACDC, EMPIRE10, OASIS; and lastly the multiscale variants of the algorithms are evaluated on the IBSR18 dataset. For all methods, we measure the Mean Square Error (MSE), Dice score across the available anatomical segmentations provided in the datasets, the number of the negative determinants of the jacobian or foldings, and finally the time and memory costs for all methods. For the Dice score on both OASIS and IBSR datasets, statistical tests demonstrate our NSVF-R2Net-nonProb is not significantly different from the top algorithms on these two datasets.

Experiment	Algorithm	MSE ↓	Dice ↑	Foldings γ^{abs} ↓	Time (ms) ↓	Memory (GB) ↓
Study on Velocity Estimation (e^{-3})	NSVF-R2Net-Prob	$2.68_{\pm 0.00}$	0.61	$3.55_{\pm 11.6}$	838	8.5
	NSVF-R2Net-noKL	$2.33_{\pm 0.00}$	0.64	$16.67_{\pm 119.89}$	**623**	8.5
	NSVF-R2Net-Gauss	$\mathbf{2.32}_{\pm 0.00}$	**0.65**	$14.81_{\pm 117.48}$	645	8.5
	NSVF-R2Net-Multi	$2.55_{\pm 0.00}$	0.62	$34.35_{\pm 28.42}$	817	8.5
	NSVF-R2Net-nonProb	$2.66_{\pm 0.00}$	0.63	$\mathbf{2.72}_{\pm 4.05}$	624	8.5
	NSVF-R2Net-nonProb-Gauss	$3.05_{\pm 0.00}$	0.60	$15.28_{\pm 131.43}$	635	8.5
ACDC (e^{-3})	SyN	$\mathbf{2.76}_{\pm 1}$	0.69	0.00	7 min	-
	VM-Diff	$3.45_{\pm 1}$	0.68	0.00	40	**0.883**
	SVF-R2Net-Prob	$3.58_{\pm 1}$	0.61	0.00	43	0.915
	NSVF-R2Net-Prob	$3.49_{\pm 1}$	0.61	0.00	43	0.915
	SVF-R2Net-nonProb	$3.50_{\pm 1}$	**0.71**	0.00	**36**	0.915
	NSVF-R2Net-nonProb	$3.43_{\pm 1}$	**0.71**	0.00	37	0.915
EMPIRE10	SyN	$\mathbf{0.03}_{\pm 0.01}$	0.91	**0.00**	15 min	-
	VM-Diff	$0.04_{\pm 0.03}$	0.91	$141.9_{\pm 263.8}$	**378**	**8.56**
	SVF-R2Net-Prob	$0.08_{\pm 0.03}$	0.78	**0.00**	876	8.59
	NSVF-R2Net-Prob	$\mathbf{0.03}_{\pm 0.01}$	**0.93**	**0.00**	872	8.59
	SVF-R2Net-nonProb	$\mathbf{0.03}_{\pm 0.01}$	0.90	**0.00**	645	8.59
	NSVF-R2Net-nonProb	$\mathbf{0.03}_{\pm 0.01}$	**0.93**	**0.00**	643	8.59
OASIS(e^{-3})	SyN	$\mathbf{1.08}_{\pm 0.00}$	0.67	$47.70_{\pm 145.14}$	10 min	-
	VM-Diff	$1.10_{\pm 0.00}$	**0.72**	$51.43_{\pm 83.76}$	**450**	8.5
	SVF-R2Net-Prob	$1.54_{\pm 0.0}$	0.69	$28.41_{\pm 17.56}$	840	8.5
	NSVF-R2Net-Prob	$1.34_{\pm 0.0}$	0.70	$\mathbf{27.23}_{\pm 47.44}$	838	8.5
	SVF-R2Net-nonProb	$1.49_{\pm 0.00}$	0.70	$38.84_{\pm 29.36}$	619	8.5
	NSVF-R2Net-nonProb	$1.33_{\pm 0.00}$	0.71	$33.53_{\pm 22.04}$	624	8.5
IBSR18(e^{-3})	SyN	$1.97_{\pm 0.00}$	**0.61**	**0.00**	17 min	-
	MS-VM-Diff	$\mathbf{1.32}_{\pm 0.00}$	0.60	$460.0_{\pm 701.05}$	**647**	11.8
	MS-SVF-R2Net-nonProb	$1.58_{\pm 0.00}$	0.58	$81.85_{\pm 80.98}$	1000	11.8
	MS-NSVF-R2Net-nonProb	$1.46_{\pm 0.00}$	**0.61**	$115.5_{\pm 100.45}$	1000	11.8

worse. This is probably because the deformation field has a very strong constraint of regularity in this case (Gaussian smoothing plus diffusion regularizer plus the already enforced Lipschitz continuity in the residual blocks of our architecture). As a result, it becomes difficult for the model to change the deformation fields far away from the identity map, causing slower convergence and reduced image-matching accuracy. We observed that NSVF-R2Net-nonProb provides the best balancing results on the smoothness of the deformation fields, the accuracy of image matching, and the inference time. The attempt on adding a Gaussian smoothing layer into NSVF-R2Net-nonProb fails, which has decreased registration accuracy and reduced smoothness of deformation fields. This is because the excess regularity constraints on the velocity fields lead to slower convergence and worse image matching.

Overall, the results in Table 1 demonstrate that a variational Bayes approach for estimating velocity fields is not necessary. Due to its good balance between the image matching and deformation smoothness of the NSVF-R2Net-nonProb, we use it to estimate the initial velocity fields in our following experiments.

Evaluating Our Models on Three Datasets. We first evaluate our method on the ACDC dataset Bernard et al. (2018). Table 1 shows that both of our architectures, SVF-R2Net-nonProb and NSVF-R2Net-NonProb, improve registration accuracy in terms of the Dice scores as compared to the baseline architectures. We also observe from Fig. 4 that in the central heart region, where the maximum deformation occurs, R2Net shows better matching results, while VM-Diff suffers proper matching on the boundaries of the heart region. All methods perform well in producing diffeomorphic deformations with zero foldings.

The second experiment is performed on the EMPIRE10 dataset. Lung registration is a particularly challenging task with both large and small deformations Hering et al. (2021) and it is easy to get stuck in local minima Heinrich et al. (2013). It can be seen from Table 1, that both our architectures SVF-R2Net-nonProb and NSVF-R2Net-nonProb improve the registration performance of the baseline architectures in terms of Dice scores. It should also be noted that for respiratory motion, there are multiple intensity changes within the lungs because of the lung tissue alterations caused by breathing Hering et al. (2021), this will lead image matching to be sub-optimal even for the baseline algorithms. However, from the image intensity difference maps in Fig. 4, we can observe that while R2Net architectures show a slight mismatch on the boundaries of the

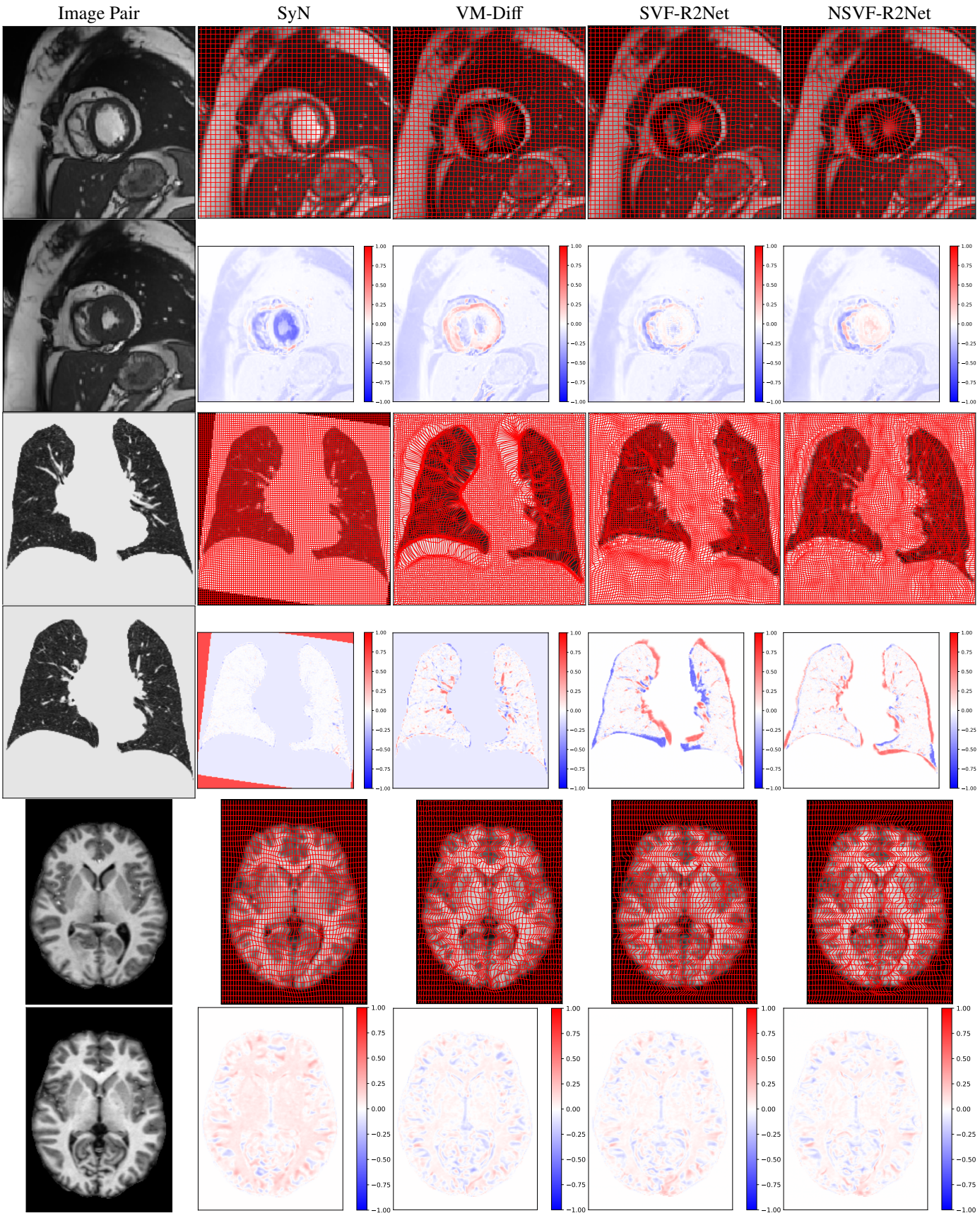

Fig. 4. Registration comparison on cases showing large deformations among SyN, VM-Diff (VoxelMorph), and our models on three datasets, from top to bottom, the ACDC heart MRI dataset, the Empire10 lung CT dataset, and the OASIS brain MRI dataset. The first column shows the input image source (top) and target (bottom) image pair. Next, we show the deformed source image overlayed by the deformation map for all algorithms on the top row and the image intensity difference maps between the deformed source and target image on the bottom row.

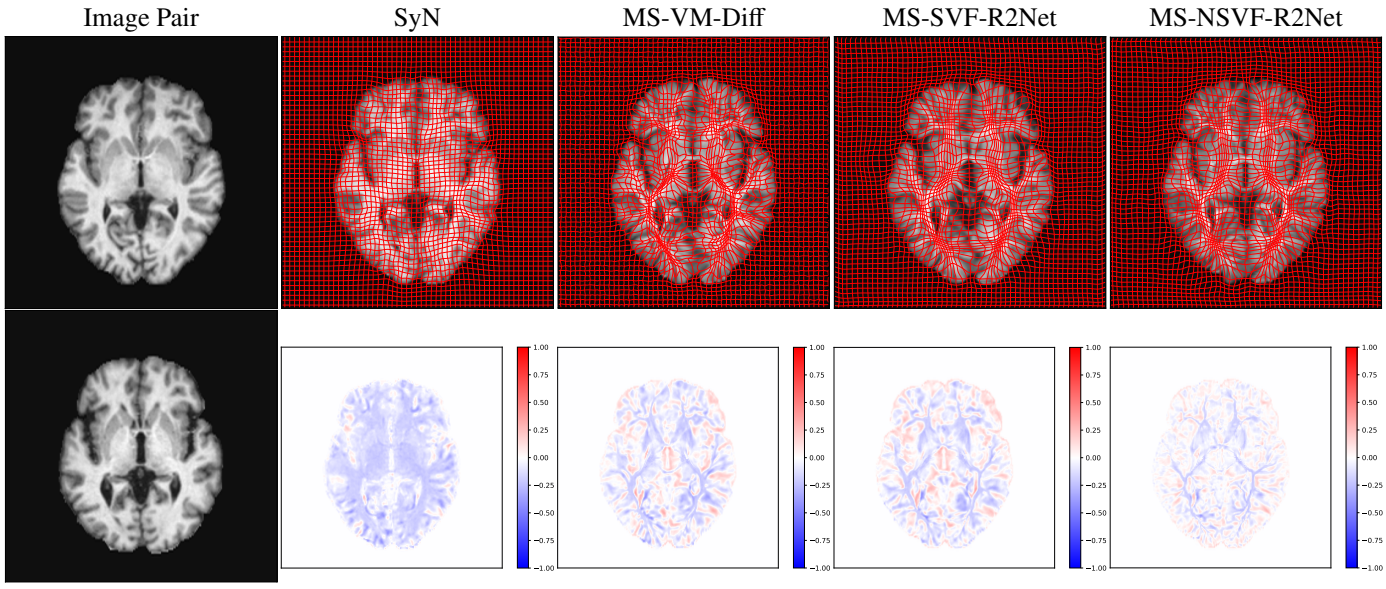

Fig. 5. Multi-scale registration comparison among SyN, multiscale version MS-VM-Diff (VoxelMorph), and our multi-scale models on the IBSR18 dataset. The first column shows the input image source (top) and target (bottom) image pair. Next, we show the deformed source image overlayed by the deformation map for all algorithms on the top row and the image intensity difference maps between the deformed source and target image on the bottom row.

lungs, both SyN and VM-Diff produce a lot of background noise, which makes the image intensity maps slightly red or blue in the background, not in white as our R2Net architectures. Also, all R2Net variants and SyN have no foldings in the deformation fields; however, VM-Diff produces non-diffeomorphic deformations while generating good image-matching results.

Another experiment is performed on the OASIS3D dataset. As shown in Table 1, our architectures SVF-R2Net-nonProb and NSVF-R2Net-nonProb slightly improve the registration accuracy from the probabilistic ones, i.e., SVF-R2Net-Prob and NSVF-R2Net-Prob, in terms of both MSE and Dice scores; however, they also show an increase in the number of foldings. Compared with VM-Diff, NSVF-R2Net-nonProb show comparable Dice scores while with greatly reduced foldings. Qualitative results have been shown in Fig. 4, where the image intensity difference maps show that our models have better matching with more white regions. Smoother deformation fields are also seen in the visualizations of the generated red grids.

Evaluating MS-R2Net Variants on IBSR18 Dataset. To evaluate our multi-scale model, we choose the IBSR18 dataset, which has the largest image volume size among our datasets. We construct two multi-scale versions of our basic R2Net, i.e., MS-SVF-R2Net and MS-NSVF-R2Net. For comparison, we choose SyN and a multi-scale extension VoxelMorph Dalca et al. (2018); Krebs et al. (2018). In particular, we extend VoxelMorph with a U-Net that outputs velocity fields at three different scales, i.e., the original scale and another two with a downsampling factor of 2 and 4, respectively. The integration of the velocity fields, the deformation of the source image, and the matching with the target image are performed on each scale. Also, three KL divergence loss functions are applied at each scale to ensure the smoothness of three velocity fields. We refer to this multi-scale version of VoxelMorph as MS-VM-Diff in our experimental results.

As reported in Table 1, our MS-NSVF-R2Net achieves higher Dice scores compared to other methods and is comparable to SyN. SyN achieves no foldings on this dataset, i.e., it provides completely diffeomorphic results. While both our variants give fewer foldings than MS-VM-Diff, MS-NSVF-R2Net does better than MS-SVF-R2Net in producing better matching results while slightly less smooth deformations. Overall, SyN provides the best results, while needing a much longer time for inference; when compared to MS-VM-Diff, our models produce both a higher Dice score and smoother deformations, with slightly increased inference time in less than 0.4 seconds.

In Fig. 5, we can also see the qualitative results on an image pair sampled from the IBSR18 dataset. It can be seen that the intensity difference map is whiter for our method MS-NSVF-R2Net-nonProb as compared to others. We can also see smooth deformations generated from our MS-R2Net. Besides, Figure 6 shows the deformed images and their corresponding deformations for all image chunks and the downsampled image.

To assess the statistical equivalence of the top-performing algorithms (SyN or VM-Diff) with our NSVF-R2Net-nonProb algorithm for the OASIS and the IBSR18 datasets in the Dice score, we perform the paired two one-sided t-tests (i.e., paired TOST) Wellek (2002). In both cases, we observe that the difference was not statistically significant, and this test, therefore, confirms that our model NSVF-R2Net-nonProb can be considered statistically equivalent to the top-performing algorithm.

Study of Computational Cost. We thoroughly study the computational cost of the proposed architectures. In Table 1, our models show comparable inference time and memory cost, compared to VoxelMorph, which integrates velocity fields at half scale. Specifically, in order to have a fair comparison, we compare the SVF-R2Net and NSVF-R2Net architectures with VoxelMorph, which performs full-scale integration of velocity fields using the scaling and squaring method.

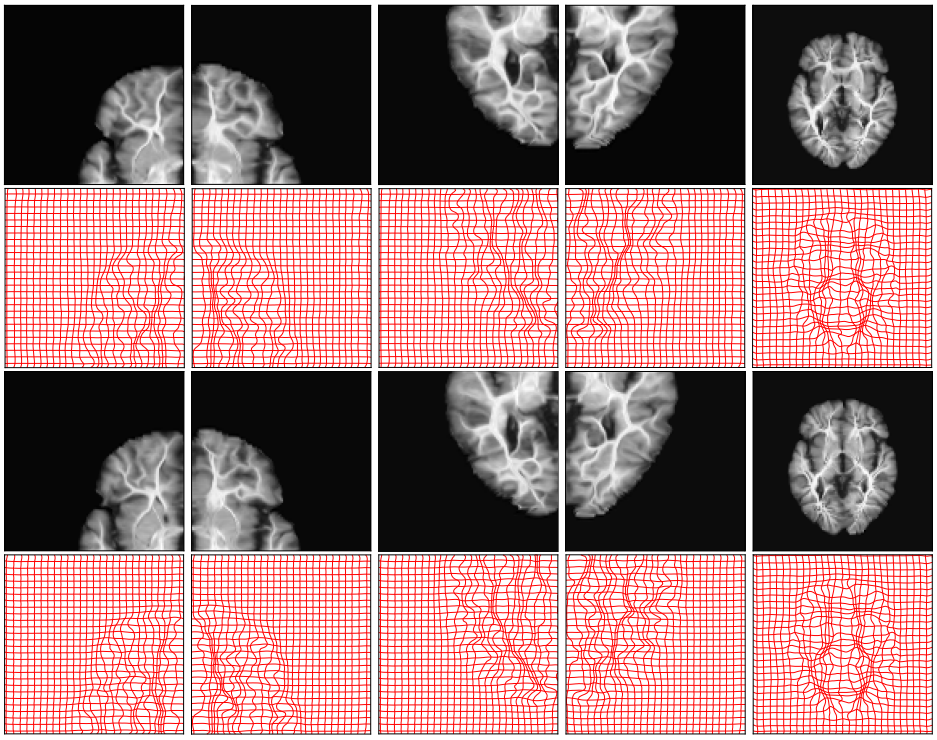

Fig. 6. Deformed images and corresponding deformations ϕ for the chunk branch (the left four columns) and the downsampled branch (the rightmost column) for an image pair sampled from the IBSR18 dataset, the same pair as shown in Fig. 5, which are generated by MS-SVF-R2Net (top two rows) and MS-NSVF-R2Net (bottom two rows).

Table 2. Training Time (sec/iteration) and GPU Memory (GB) consumption for VM-Diff (diffeomorphic VoxelMorph, integrating at the original scale) and our models. All experiments are carried out on one image pair, and the reported values are averaged over 10 runs on the OASIS 3D brain MRI dataset.

Data Size	VM-Diff		SVF-R2Net-Prob		NSVF-R2Net-Prob		SVF-R2Net-nonProb		NSVF-R2Net-nonProb	
n^3	Time	Memory	Time	Memory	Time	Memory	Time	Memory	Time	Memory
64	0.113	1.4	0.170	0.9	0.174	0.9	0.153	0.7	0.156	0.7
96	0.298	2.4	0.432	1.4	0.433	1.4	0.387	1.4	0.390	1.4
112	0.419	4.4	0.651	2.4	0.660	2.4	0.567	2.4	0.577	2.4
128	0.600	8.5	0.915	4.5	0.908	4.5	0.788	2.4	0.799	2.4
144	0.907	8.5	1	4.5	1	4.5	1	4.5	1	4.5
192	–	–	3	8.5	3	8.5	2	8.5	2	8.5
224	–	–	4	11.8	4	11.8	3	11.8	3.5	11.8

Table 2 shows the detailed study of the computational cost of VM-Diff, SVF-R2Net-Prob, NSVF-R2Net-Prob, SVF-R2Net-nonProb, and NSVF-R2Net-nonProb architectures. The public implementation of VoxelMorph was modified, in that, the integration of the velocity fields was done on the original input sizes instead of their default half size. It can be seen that our methods consistently need less memory compared to VM-Diff. Also, VM-Diff cannot work with image sizes 192^3 and higher. However, all four architectures of R2Net can handle these sizes. It is also to be noted that both SVF-R2Net-Prob and NSVF-R2Net-Prob took a little more time in seconds as compared to VM-Diff. This is due to the more number of convolutional layers and the KL-divergence loss in the loss functions. However, SVF-R2Net-nonProb and NSVF-R2Net-nonProb further reduce the time cost, taking just < 0.2 seconds more than VM-Diff for the same input image size.

From this experiment, we can see that all R2Net variants can handle higher resolutions of images and also scale up better, compared to the baseline VM-Diff architecture. This also shows the drawbacks of using scaling and squaring methodology for integrating velocity fields. Considering that medical image volume resolutions are always on the rise, this gives motivation for alternative approaches to be undertaken for the integration of velocity fields such as the one introduced in our work.

Convergence Test for R2Nets. In this section, we test whether our R2Net actually learns how to integrate the velocity fields and generate deformations based on our designed LC-ResNet blocks. That is, we want to evaluate if our LC-ResNet blocks perform as a numerical integration scheme, like the scaling and squaring algorithm to correctly integrate given velocity fields.

We use the following algorithm to determine whether an SVF-R2Net has learned a meaningful model, given an unseen test image pair, I_S and I_T, and a trained model F_θ. We choose SVF-R2Net since it is also parameterized by stationary velocity fields, similar to the scaling and squaring algorithm.

1. In the first step, evaluate a trained SVF-R2Net-nonProb,

and obtain an integrated deformation field, for a given pair of input images:

$$\{v_0, \hat{\phi}_{predicted}\} = F_\theta(I_S, I_T). \tag{11}$$

We obtain an estimated initial velocity field v_0 and the final diffeomorphic deformation driven by this velocity field, which is generated by integrating $\int_0^1 v_t(\phi_t)dt$ using the LC-ResNet blocks.

2. Next, using the same estimated initial velocity field v_0, we pass it through the scaling and squaring layers to obtain the ground-truth integration, which is our expected diffeomorphic deformation $\phi_{expected}$:

$$\phi_{expected} = SS(v_0), \tag{12}$$

where SS denotes the scaling and squaring function.

3. In the last step, we compute the L_2 error between the R2Net predicted deformation and the true/expected one given by the scaling and squaring algorithm:

$$Error(h) = \|\hat{\phi}_{predicted} - \phi_{expected}\|_2. \tag{13}$$

We use the OASIS 3D dataset to conduct our convergence test since our biggest test sample of 100 images is available only for this dataset. As expected, we obtain a score of 3.16e-5 as the mean error and a standard deviation of 0.1e-5 on this. The small standard deviation suggests that we almost get a constant error on most of the predicted deformation fields as compared to the expected ones. Therefore, our LC-ResNet blocks indeed provide an integration of the velocity fields, $\int_0^1 v_t(\phi_t)dt$, as given in Eq. 3 and the corollary shown in Eq. 7.

4. Discussion and Conclusion

In this paper, we have proposed an unsupervised deep diffeomorphic image registration framework, which has flexible parameterizations of deformations fields. Our architectures, SVF-R2Net-nonProb and NSVF-R2Net-nonProb, are based on non-probabilistic UNets for estimating the initial velocity fields, but one for stationary velocity fields and the other for non-stationary (time-varying) velocity fields. In both architectures, we employ Lipschitz-continuous ResNets as numerical schemes of differential equations. We have demonstrated the effectiveness of our approach on varied anatomies and modalities of images for both inter and intra-subject registration tasks. We have outperformed or shown comparable results to both classical and learning-based registration methods in terms of image matching. We have also shown better deformation smoothness and regularity than deep learning based algorithms, by showing a consistently lower number of foldings, across all datasets, which is necessary for the task of diffeomorphic image registration. We also perform a thorough study of the time and computational costs of all R2Net variants. Our architectures have been shown to perform image registration on evaluated datasets in under a second while integrating velocity fields on the original input size and taking the same amount of GPU memory.

We also extend our SVF-R2Net and NSVF-R2Net models into multi-scale variants, namely, MS-SVF-R2Net and MS-NSVF-R2Net. These architectures demonstrate the benefits of fully utilizing the available resources without hampering the resolution of the input images. Compared to traditional and learning based methods, our approaches can offer diffeomorphic guarantees and model large deformations at the same time. We have shown that we are able to fit a dataset of 224×224×224 entirely on a single TiTAN X GPU and the integration has been performed on the original size, while other learning-based methods like VoxelMorph can only handle integration on the half-scale velocity fields.

Lastly, we also provide a convergence test in the form of an algorithm to confirm that SVF-R2Net actually learns to integrate the velocity fields and generate diffeomorphic deformations using our customized LC-ResNet blocks. Theoretically, this can also be shown for the NSVF-R2Net architecture. However, we do not further investigate this aspect in the current work, due to the lack of deep learning based registration methods where the deformations are parameterized using time-varying velocity fields.

Currently, we use a fixed number, i.e., seven, of integration blocks to integrate the velocity fields in our architectures. This setting works on all the datasets that we use in the evaluation, and it also enables us a fair and direct comparison with VoxelMorph, which has seven steps for the scaling and squaring algorithm Arsigny et al. (2006) in the integration layer. However, in the future, we could learn this parameter during the training process, so that we can have an adaptive number of time steps.

Our R2Net framework opens up possibilities for various extensions and applications. For example, in this work we only consider uni-modal registration tasks; however, we can explore the performance of our architectures with modifications in loss functions to accommodate multi-modal image registration. We could also include the label information of the various anatomical regions using their segmentation masks within the training phase as done by Hering et al. (2021); Hoffmann et al. (2021) to improve the overall label-matching accuracy. In the future, due to the similarity with the LDDMM architecture, this model could also be applicable for other tasks such as metamorphic image registration to model the deformation in the presence of appearance changes, for example, to study brain development or disease progression like tumor development.

Declaration of Competing Interest

There is no financial/personal interest or belief that could affect the objectivity of the submitted research results. No conflict of interest exists.

CRediT authorship contribution statement

Ankita Joshi: Conceptualization, Methodology, Implementation, Validation, Investigation, Writing - original draft, Writing - review and editing, Visualization. **Yi Hong:** Conceptualization, Methodology, Resources, Supervision, Writing - review and editing, Funding acquisition.

Acknowledgments

This work was supported by NSFC 62203303, NSF 1755970, and Shanghai Municipal Science and Technology Major Project 2021SHZDZX0102.

References

Abadi, M., 2016. Tensorflow: learning functions at scale, in: Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming, pp. 1–1.

Arsigny, V., Commowick, O., Pennec, X., Ayache, N., 2006. A log-euclidean framework for statistics on diffeomorphisms, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 924–931.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. neuroimage38, 95–113.

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. Medical image analysis 12, 26–41.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging 38, 1788–1800.

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision 61, 139–157.

Ben Amor, B., Arguillère, S., Shao, L., 2021. Resnet-lddmm: Advancing the lddmm framework using deep residual networks. arXiv e-prints , arXiv–2102.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al., 2018. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE transactions on medical imaging 37, 2514–2525.

Bietti, A., Mairal, J., 2019. Group invariance, stability to deformations, and complexity of deep convolutional representations. The Journal of Machine Learning Research 20, 876–924.

Brunn, M., Himthani, N., Biros, G., Mehl, M., Mang, A., 2021. Fast gpu 3d diffeomorphic image registration. Journal of Parallel and Distributed Computing 149, 149–162.

Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D., 2017. Deformable image registration based on similarity-steered cnn regression, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 300–308.

Ceritoglu, C., Oishi, K., Li, X., Chou, M.C., Younes, L., Albert, M., Lyketsos, C., van Zijl, P.C., Miller, M.I., Mori, S., 2009. Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. Neuroimage 47, 618–627.

Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D.K., 2018. Neural ordinary differential equations. Advances in neural information processing systems 31.

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.

De Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis 52, 128–143.

Fan, J., Cao, X., Yap, P.T., Shen, D., 2019. Birnet: Brain image registration using dual-supervised fully convolutional networks. Medical image analysis 54, 193–206.

Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N., 2008. Dense image registration through mrfs and efficient linear programming. Medical image analysis 12, 731–741.

Gouk, H., Frank, E., Pfahringer, B., Cree, M.J., 2021. Regularisation of neural networks by enforcing lipschitz continuity. Machine Learning 110, 393–416.

Haber, E., Ruthotto, L., 2017. Stable architectures for deep neural networks. Inverse problems 34, 014004.

Haskins, G., Kruger, U., Yan, P., 2020. Deep learning in medical image registration: a survey. Machine Vision and Applications 31, 1–18.

Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A., 2013. Mrf-based deformable registration and ventilation estimation of lung ct. IEEE transactions on medical imaging 32, 1239–1248.

Hering, A., Ginneken, B.v., Heldmann, S., 2019. mlvirnet: Multilevel variational image registration network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 257–265.

Hering, A., Häger, S., Moltz, J., Lessmann, N., Heldmann, S., van Ginneken, B., 2021. Cnn-based lung ct registration with multiple anatomical constraints. Medical Image Analysis 72, 102139.

Hernandez, M., Bossa, M.N., Olmos, S., 2007. Registration of anatomical images using geodesic paths of diffeomorphisms parameterized with stationary vector fields, in: 2007 IEEE 11th International Conference on Computer Vision, IEEE. pp. 1–8.

Higham, N.J., 2005. The scaling and squaring method for the matrix exponential revisited. SIAM Journal on Matrix Analysis and Applications 26, 1179–1193.

Himthani, N., Brunn, M., Kim, J.Y., Schulte, M., Mang, A., Biros, G., 2022. Claire—parallelized diffeomorphic image registration for large-scale biomedical imaging applications. Journal of Imaging 8, 251.

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 transactions on medical imaging 41, 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.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. Advances in neural information processing systems 28.

Joshi, A., Hong, Y., 2021. Diffeomorphic image registration using lipschitz continuous residual networks, in: Medical Imaging with Deep Learning.

Joshi, A., Hong, Y., 2022. Efficient diffeomorphic image registration using multi-scale dual-phased learning, in: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 1–5.

Joshi, S.C., Miller, M.I., 2000. Landmark matching via large deformation diffeomorphisms. IEEE transactions on image processing 9, 1357–1370.

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

Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. Neuroimage 46, 786–802.

Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T., 2019. Learning a probabilistic model for diffeomorphic registration. IEEE transactions on medical imaging 38, 2165–2176.

Krebs, J., Mansi, T., Mailhé, B., Ayache, N., Delingette, H., 2018. Unsupervised probabilistic deformation modeling for robust diffeomorphic registration, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, pp. 101–109.

Li, H., Fan, Y., 2017. Non-rigid image registration using fully convolutional networks with deep self-supervision. arXiv preprint arXiv:1709.00799 .

Liu, G.H., Theodorou, E.A., 2019. Deep learning theory review: An optimal control and dynamical systems perspective. arXiv preprint arXiv:1908.10920 .

Lu, Y., Zhong, A., Li, Q., Dong, B., 2018. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations, in: International Conference on Machine Learning, PMLR. pp. 3276–3285.

Mang, A., Gholami, A., Davatzikos, C., Biros, G., 2019. Claire: A distributed-memory solver for constrained large deformation diffeomorphic image registration. SIAM Journal on Scientific Computing 41, C548–C584.

Mang, A., Ruthotto, L., 2017. A lagrangian gauss–newton–krylov solver for mass-and intensity-preserving diffeomorphic image registration. SIAM Journal on Scientific Computing 39, B860–B885.

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. Journal of cognitive neuroscience 19, 1498–1507.

Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y., 2018. Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957 .

Mok, T.C., Chung, A., 2020a. Fast symmetric diffeomorphic image registra-

tion 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.C., 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.

Murphy, K., Van Ginneken, B., Reinhardt, J.M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G.E., Garcia, V., et al., 2011. Evaluation of registration methods on thoracic ct: the empire10 challenge. IEEE transactions on medical imaging 30, 1901–1920.

Nalisnick, E., Matsukawa, A., Teh, Y.W., Gorur, D., Lakshminarayanan, B., 2018. Do deep generative models know what they don't know? arXiv preprint arXiv:1810.09136 .

Razavi, A., Van den Oord, A., Vinyals, O., 2019. Generating diverse high-fidelity images with vq-vae-2. Advances in neural information processing systems 32.

Rohé, M.M., Datar, M., Heimann, T., Sermesant, M., Pennec, X., 2017. Svf-net: Learning deformable image registration using shape matching, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 266–274.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Rousseau, F., Drumetz, L., Fablet, R., 2020. Residual networks as flows of diffeomorphisms. Journal of Mathematical Imaging and Vision 62, 365–375.

Rühaak, J., Polzin, T., Heldmann, S., Simpson, I.J., Handels, H., Modersitzki, J., Heinrich, M.P., 2017. Estimation of large motion in lung ct by integrating regularized keypoint correspondences into dense deformable registration. IEEE transactions on medical imaging 36, 1746–1757.

Ruthotto, L., Haber, E., 2020. Deep neural networks motivated by partial differential equations. Journal of Mathematical Imaging and Vision 62, 352–364.

Sotiras, A., Davatzikos, C., Paragios, N., 2013. Deformable medical image registration: A survey. IEEE transactions on medical imaging 32, 1153–1190.

Valverde, S., Oliver, A., Cabezas, M., Roura, E., Lladó, X., 2015. Comparison of 10 brain tissue segmentation methods using revisited ibsr annotations. Journal of Magnetic Resonance Imaging 41, 93–101.

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45, S61–S72.

Vos, B.D.d., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I., 2017. End-to-end unsupervised deformable image registration with a convolutional neural network, in: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, pp. 204–212.

Weinan, E., 2017. A proposal on machine learning via dynamical systems. Communications in Mathematics and Statistics 1, 1–11.

Wellek, S., 2002. Testing statistical hypotheses of equivalence. Chapman and Hall/CRC.

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.

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.

Yang, T., Tang, Q., Li, L., Bai, X., 2021. Non-rigid medical image registration using multi-scale residual deep fully convolutional networks. Journal of Instrumentation 16, P03005.

Yang, X., Kwitt, R., Styner, M., Niethammer, M., 2017. Quicksilver: Fast predictive image registration–a deep learning approach. NeuroImage 158, 378–396.

Yeo, B.T., Sabuncu, M.R., Vercauteren, T., Holt, D.J., Amunts, K., Zilles, K., Golland, P., Fischl, B., 2010. Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex. IEEE transactions on medical imaging 29, 1424–1441.

Yoshida, Y., Miyato, T., 2017. Spectral norm regularization for improving the generalizability of deep learning. arXiv preprint arXiv:1705.10941 .

Younes, L., 2010. Shapes and diffeomorphisms. volume 171. Springer.

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.