

# DeepFLASH: An Efficient Network for Learning-based Medical Image Registration

Jian Wang  
University of Virginia  
jw4hv@virginia.edu

Miaomiao Zhang  
University of Virginia  
mz8rr@virginia.edu

## Abstract

*This paper presents DeepFLASH, a novel network with efficient training and inference for learning-based medical image registration. In contrast to existing approaches that learn spatial transformations from training data in the high dimensional imaging space, we develop a new registration network entirely in a low dimensional bandlimited space. This dramatically reduces the computational cost and memory footprint of an expensive training and inference. To achieve this goal, we first introduce complex-valued operations and representations of neural architectures that provide key components for learning-based registration models. We then construct an explicit loss function of transformation fields fully characterized in a bandlimited space with much fewer parameterizations. Experimental results show that our method is significantly faster than the state-of-the-art deep learning based image registration methods, while producing equally accurate alignment. We demonstrate our algorithm in two different applications of image registration: 2D synthetic data and 3D real brain magnetic resonance (MR) images. Our code is available at <https://github.com/jw4hv/deepflash>.*

## 1. Introduction

Image registration has been widely used in medical image analysis, for instance, atlas-based image segmentation [2, 16], anatomical shape analysis [20, 33], and motion correction in dynamic imaging [18]. The problem of deformable image registration is typically formulated as an optimization, seeking for a nonlinear and dense (voxelwise) spatial transformation between images. In many applications, it is desirable that such transformations be diffeomorphisms, i.e., differentiable, bijective mappings with differentiable inverses. In this paper, we focus on diffeomorphic image registration highlighting with a set of critical features: (i) it captures large deformations that often occur in brain shape variations, lung motions, or fetal movements;

(ii) the topology of objects in the image remain intact; and (iii) no non-differentiable artifacts, such as creases or sharp corners, are introduced. However, achieving an optimal solution of diffeomorphic image registration, especially for large-scale or high-resolution images (e.g., a 3D brain MRI scan with the size of  $256^3$ ), is computationally intensive and time-consuming.

Attempts at speeding up diffeomorphic image registration have been made in recent works by improving numerical approximation schemes. For example, Ashburner and Friston [3] employ a Gauss-Newton method to accelerate the convergence of large deformation diffeomorphic metric mapping (LDDMM) algorithm [8]. Zhang et al. propose a low dimensional approximation of diffeomorphic transformations, resulting in fast computation of the gradients for iterative optimization [32, 27]. While these methods have led to substantial reductions in running time, such gradient-based optimization still takes minutes to finish. Instead of minimizing a complex registration energy function [8, 31], an alternative approach has leveraged deep learning techniques to improve registration speed by building prediction models of transformation parameters. Such algorithms typically adopt convolutional neural networks (CNNs) to learn a mapping between pairwise images and associated spatial transformations in training dataset [28, 21, 7, 10, 9]. Registration of new testing images is then achieved rapidly by evaluating the learned mapping on given volumes. While the aforementioned deep learning approaches are able to fast predict the deformation parameters in testing, the training process is extremely slow and memory intensive due to the high dimensionality of deformation parameters in imaging space. In addition, enforcing the smoothness constraints of transformations when large deformation occurs is challenging in neural networks.

To address this issue, we propose a novel learning-based registration framework DeepFLASH in a low dimensional bandlimited space, where the diffeomorphic transformations are fully characterized with much fewer dimensions. Our work is inspired by a recent registration algorithm FLASH (Fourier-approximated Lie Algebras for Shoot-

ing) [31] with the novelty of (i) developing a learning-based predictive model that further speeds up the current registration algorithms; (ii) defining a set of complex-valued operations (e.g., complex convolution, complex-valued rectifier, etc.) and complex-valued loss function of transformations entirely in a bandlimited space; and (iii) proving that our model can be easily implemented by a dual-network in the space of real-valued functions with a careful design of the network architecture. To the best of our knowledge, we are the first to introduce the low dimensional Fourier representations of diffeomorphic transformations to learning-based registration algorithms. In contrast to traditional methods that learn spatial transformations in a high dimensional imaging space, our method dramatically reduces the computational complexity of the training process where iterative computation of gradient terms are required. This greatly alleviates the problem of time-consuming and expensive training for deep learning based registration networks. Another major benefit of DeepFLASH is that the smoothness of diffeomorphic transformations is naturally preserved in the bandlimited space with low frequency components. Note that while we implement DeepFLASH in the context of convolutional neural network (CNN), it can be easily adapted to a variety of other neural networks, such as fully connected network (FCN), or recurrent neural network (RNN). We demonstrate the effectiveness of our model in both 2D synthetic and 3D real brain MRI data.

## 2. Background: Deformable Image Registration

In this section, we briefly review the fundamentals of deformable image registration in the setting of LDDMM algorithm with geodesic shooting [26, 29]. While this paper focuses on LDDMM, the theoretical tool developed in our model is widely applicable to various registration frameworks, for example, stationary velocity fields.

Consider a source image  $S$  and a target image  $T$  as square-integrable functions defined on a torus domain  $\Omega = \mathbb{R}^d / \mathbb{Z}^d$  ( $S(x), T(x) : \Omega \rightarrow \mathbb{R}$ ). The problem of diffeomorphic image registration is to find the shortest path (also known as geodesic) of diffeomorphic transformations  $\phi_t \in \text{Diff}(\Omega) : \Omega \rightarrow \Omega, t \in [0, 1]$ , such that the deformed image  $S \circ \psi_1$  at time point  $t = 1$  is similar to  $T$ . An explicit energy function of LDDMM with geodesic shooting is formulated as an image matching term plus a regularization term that guarantees the smoothness of the transformation fields

$$E(v_0) = \frac{\gamma}{2} \text{Dist}(S \circ \psi_1, T) + \frac{1}{2} (Lv_0, v_0), \text{ s.t.},$$

$$\frac{d\psi_t}{dt} = -\mathcal{D}\psi_t \cdot v_t, \quad (1)$$

where  $\gamma$  is a positive weight parameter,  $\mathcal{D}$  denotes a Ja-

cobian matrix, and  $\text{Dist}(\cdot, \cdot)$  is a distance function that measures the similarity between images. Commonly used distance metrics include sum-of-squared difference (L2-norm) of image intensities [8], normalized cross correlation (NCC) [4], and mutual information (MI) [17]. The deformation  $\phi$  is defined as an integral flow of the time-varying Eulerian velocity field  $v_t$  that lies in the tangent space of diffeomorphisms  $V = T\text{Diff}(\Omega)$ . Here  $L : V \rightarrow V^*$  is a symmetric, positive-definite differential operator that maps a tangent vector  $v \in V$  into the dual space  $m \in V^*$ , with its inverse  $K : V^* \rightarrow V$ . The  $(\cdot, \cdot)$  denotes a pairing of a momentum vector  $m \in V^*$  with a tangent vector  $v \in V$ .

The geodesic at the minimum of (1) is uniquely determined by integrating the geodesic constraint, a.k.a. Euler-Poincaré differential equation (EPDiff) [1, 19], which is computationally expensive in high dimensional image spaces. A recent model FLASH demonstrated that the entire optimization of LDDMM with geodesic shooting can be efficiently carried in a low dimensional bandlimited space with dramatic speed improvement [31]. This is based on the fact that the velocity fields do not develop high frequencies in the Fourier domain (as shown in Fig. 1). We briefly review the basic concepts below.

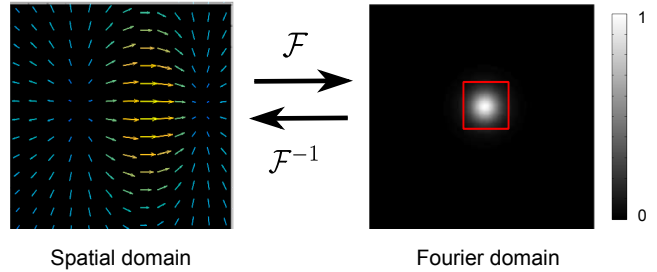


Figure 1. An example of velocity field in spatial domain vs. Fourier domain.

Let  $\widetilde{\text{Diff}}(\Omega)$  and  $\widetilde{V}$  denote the space of Fourier representations of diffeomorphisms and velocity fields respectively. Given time-dependent velocity field  $\tilde{v}_t \in \widetilde{V}$ , the diffeomorphism  $\tilde{\psi}_t \in \widetilde{\text{Diff}}(\Omega)$  in the finite-dimensional Fourier domain can be computed as

$$\tilde{\psi}_t = \tilde{e} + \tilde{u}_t, \quad \frac{d\tilde{u}_t}{dt} = -\tilde{v}_t - \tilde{\mathcal{D}}\tilde{u}_t * \tilde{v}_t, \quad (2)$$

where  $\tilde{e}$  is the frequency of an identity element,  $\tilde{\mathcal{D}}\tilde{u}_t$  is a tensor product  $\tilde{\mathcal{D}} \otimes \tilde{u}_t$ , representing the Fourier frequencies of a Jacobian matrix  $\tilde{\mathcal{D}}$  with central difference approximation, and  $*$  is a circular convolution with zero padding to avoid aliasing<sup>1</sup>.

<sup>1</sup>To prevent the domain from growing infinity, we truncate the output of the convolution in each dimension to a suitable finite set.

The Fourier representation of the geodesic constraint (EPDiff) is defined as

$$\frac{\partial \tilde{v}_t}{\partial t} = -\tilde{K} \left[ (\tilde{D}\tilde{v}_t)^T \star \tilde{m}_t + \tilde{\nabla} \cdot (\tilde{m}_t \otimes \tilde{v}_t) \right], \quad (3)$$

where  $\star$  is the truncated matrix-vector field auto-correlation. The operator  $\tilde{\nabla} \cdot$  is the discrete divergence of a vector field. Here  $\tilde{K}$  is a smoothing operator with its inverse  $\tilde{L}$ , which is the Fourier transform of a commonly used Laplacian operator  $(-\alpha\Delta + I)^c$ , with a positive weight parameter  $\alpha$  and a smoothness parameter  $c$ . The Fourier coefficients of  $\tilde{L}$  is, i.e.,  $\tilde{L}(\xi_1, \dots, \xi_d) = \left( -2\alpha \sum_{j=1}^d (\cos(2\pi\xi_j) - 1) + 1 \right)^c$ , where  $(\xi_1, \dots, \xi_d)$  is a  $d$ -dimensional frequency vector.

### 3. Our Method: DeepFLASH

We introduce a learning-based registration network DeepFLASH in a low dimensional bandlimited space  $\tilde{V}$ , with newly defined operators and functions in a complex vector space  $\mathbb{C}^n$ . Since the spatial transformations can be uniquely determined by an initial velocity  $\tilde{v}_0$  (as introduced in Eq. (3)), we naturally integrate this new parameterization in the architecture of DeepFLASH. To simplify the notation, we drop the time index of  $\tilde{v}_0$  in remaining sections.

Analogous to [28], we use optimal registration results, denoted as  $\tilde{v}^{opt}$ , estimated by numerical optimization of the LDDMM algorithm as part of the training data. Our goal is then to predict an initial velocity  $\tilde{v}^{pre}$  from image patches of the moving and target images. Before introducing DeepFLASH, we first define a set of complex-valued operations and functions that provide key components of the neural architecture.

Consider a  $Q$ -dimensional complex-valued vector of input signal  $\tilde{X}$  and a real-valued kernel  $H$ , we have a complex-valued convolution as

$$H * \tilde{X} = H * \mathcal{R}(\tilde{X}) + iH * \mathcal{I}(\tilde{X}), \quad (4)$$

where  $\mathcal{R}(\cdot)$  denotes the real part of a complex-valued vector, and  $\mathcal{I}(\cdot)$  denotes an imaginary part.

Following a recent work on complex networks [25], we define a complex-valued activation function based on Rectified Linear Unit (ReLU). We apply real-valued ReLU separately on both of the real and imaginary part of a neuron  $\tilde{Y}$  in the output layer, i.e.,

$$\text{CReLU}(\tilde{Y}) = \text{ReLU}(\mathcal{R}(\tilde{Y})) + i\text{ReLU}(\mathcal{I}(\tilde{Y})). \quad (5)$$

#### 3.1. Loss function

Let a labeled training dataset including pairwise images and their associated optimal initial velocity fields be  $\{S_n, T_n, \tilde{v}_n^{opt}\}_{n=1}^N$ , where  $N$  is the number of pairwise images. We model a prediction function  $f(S_n, T_n; W)$  by using a convolutional neural network (CNN), with  $W$  being a

real-valued weight matrix for convolutional layers. We then define a loss function  $\ell$  as

$$\ell(W) = \sum_{n=0}^N \|\tilde{v}_n^{opt} - f(S_n, T_n; W)\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \quad (6)$$

where  $\lambda$  is a positive parameter balancing between function  $f$  and a regularity term  $\text{Reg}(\cdot)$  on the weight matrix  $W$ . While we use  $L_2$  norm as regularity in this paper, it is generic to other commonly used operators such as  $L_1$ , or  $L_0$  norm.

The optimization problem of Eq. (6) is typically solved by using gradient-based methods. The weight matrix  $W$  is updated by moving in the direction of the loss functions steepest descent, found by its gradient  $\nabla_W \ell$ .

#### 3.2. Network Architecture: Decoupled Complex-valued Network

While we are ready to design a complex-valued registration network based on CNN, the implementation of such network is not straightforward. Trabelsi et al. developed deep complex networks to specifically handle complex-valued inputs at the cost of computational efficiency [25]. In this section, we present an efficient way to decouple the proposed complex-valued registration network into a combination of regular real-valued networks. More specifically, we construct a dual-net that separates the real and imaginary part of complex-valued network in an equivalent manner.

Given the fact that the computation of real and imaginary parts in the convolution (Eq. (4)) and the activation function (Eq. (5)) are separable, we next show that the loss defined in Eq. (6) can be equivalently constructed by the real and imaginary part individually. To simplify the notation, we define the predicted initial velocity  $\tilde{v}_n^{pre} \triangleq f(S_n, T_n; W)$  and rewrite Eq. (6) as

$$\ell(W) = \sum_{n=0}^N \|\tilde{v}_n^{opt} - \tilde{v}_n^{pre}\|_{L_2}^2 + \lambda \cdot \text{Reg}(W). \quad (7)$$

Let  $\tilde{v}_n^{opt} = \beta_n + i\mu_n$  and  $\tilde{v}_n^{pre} = \delta_n + i\eta_n$  in Eq. (7), we then obtain

$$\begin{aligned} \ell(W) &= \sum_{n=0}^N \|(\beta_n + i\mu_n) - (\delta_n + i\eta_n)\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \\ &= \sum_{n=0}^N \|(\beta_n - \delta_n) + i(\mu_n - \eta_n)\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \\ &= \sum_{n=0}^N \|\beta_n - \delta_n\|_{L_2}^2 + \|\mu_n - \eta_n\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \\ &= \sum_{n=0}^N \|\mathcal{R}(\tilde{v}_n^{opt}) - \mathcal{R}(\tilde{v}_n^{pre})\|_{L_2}^2 \\ &\quad + \|\mathcal{I}(\tilde{v}_n^{opt}) - \mathcal{I}(\tilde{v}_n^{pre})\|_{L_2}^2 + \lambda \cdot \text{Reg}(W). \end{aligned} \quad (8)$$

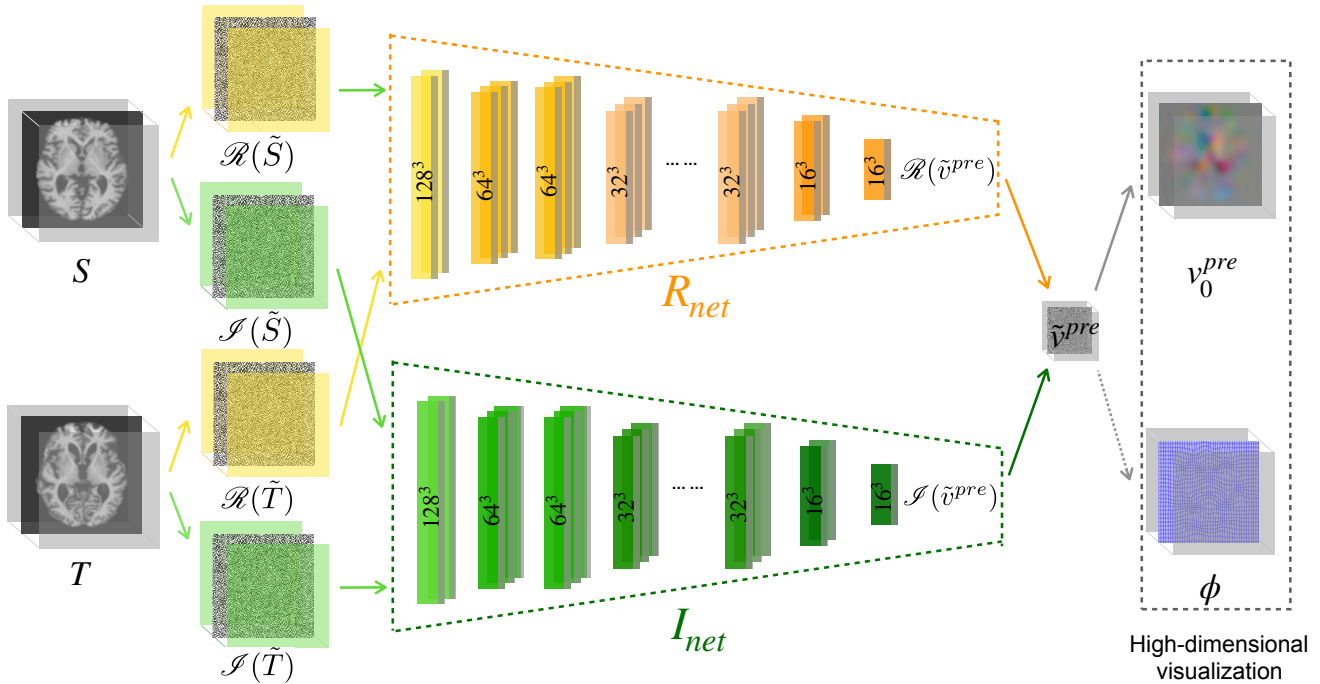


Figure 2. Architecture of DeepFLASH with dual net:  $S$  and  $T$  denote the source and target image from high dimensional spatial domain.  $\mathcal{R}(\tilde{S})$  and  $\mathcal{I}(\tilde{S})$  are the real and imaginary frequency spectrum convert from  $S$ .  $\mathcal{R}(\tilde{T})$  and  $\mathcal{I}(\tilde{T})$  are the real and imaginary frequency convert from  $T$ .  $\tilde{v}^{pre}$  is the low dimensional prediction optimized from our model.  $v_0^{pre}$  and  $\phi$  are the velocity and transformation field we recovered from our low dimensional prediction.

Fig. 2 visualizes the architecture of our proposed model DeepFLASH. The input source and target images are defined in the Fourier space with the real part of frequencies as  $\mathcal{R}(\tilde{S})$  and  $\mathcal{R}(\tilde{T})$  vs. the imaginary parts as  $\mathcal{I}(\tilde{S})$  and  $\mathcal{I}(\tilde{T})$ . We train real and imaginary parts in two individual neural networks  $R_{net}$  and  $I_{net}$ . The optimization stays in a low dimensional bandlimited space without converting the coefficients ( $\mathcal{R}(\tilde{v}^{pre})$ ,  $\mathcal{I}(\tilde{v}^{pre})$ ) to high dimensional imaging domain back and forth. It is worthy to mention that our decoupled network is not constrained by any specific architecture. The CNN network in Fig. 2 can be easily replaced by a variety of the state-of-the-art models, e.g., U-net [22], or fully connected neural network.

### 3.3. Computational complexity

It has been previously shown that the time complexity of convolutional layers [14] is  $O(\sum_{p=1}^P b_{p-1} \cdot h_p^2 \cdot b_p \cdot Z_p^2)$ , where  $p$  is the index of a convolutional layer and  $P$  denotes the number of convolutional layers. The  $b_{p-1}$ ,  $b_p$ ,  $h_p$  are defined as the number of input channels, output channels and the kernel size of the  $p$ -th layer respectively. Here  $Z_p$  is the output dimension of the  $p$ -th layer, i.e.,  $Z_p = 128^3$  when the last layer predicts transformation fields for 3D images with the dimension of  $128^3$ .

Current learning approaches for image registration have

been performed in the original high dimensional image space [7, 28]. In contrast, our proposed model DeepFLASH significantly reduces the dimension of  $Z_p$  into a low dimensional bandlimited space  $z_p$  (where  $z_p \ll Z_p$ ). This makes the training of traditional registration networks much more efficient in terms of both time and memory consumption.

## 4. Experimental Evaluation

We demonstrate the effectiveness of our model by training and testing on both 2D synthetic data and 3D real brain MRI scans.

### 4.1. Data

**2D synthetic data.** We first simulate 3000 “bull-eye” synthetic data (as shown in Fig 3) by manipulating the width  $a$  and height  $b$  of an ellipse equation, formulated as  $\frac{(x-50)^2}{a^2} + \frac{(y-50)^2}{b^2} = 1$ . We draw the parameter  $a, b$  randomly from a Gaussian distribution  $\mathcal{N}(4, 2^2)$  for inner ellipse, and  $a, b \sim \mathcal{N}(13, 4^2)$  for outer ellipse.

**3D brain MRI.** We include 3200 public T1-weighted 3D brain MRI scans from Alzheimers Disease Neuroimaging Initiative (ADNI) [15], Open Access Series of Imaging Studies (OASIS) [13], Autism Brain Imaging Data Ex-

change (ABIDE) [11], and LONI Probabilistic Brain Atlas Individual Subject Data (LPBA40) [23] with 1000 subjects. Due to the difficulty of preserving the diffeomorphic property across individual subjects particularly with large age variations, we carefully evaluate images from subjects aged from 60 to 90. All MRIs were all pre-processed as  $128 \times 128 \times 128$ ,  $1.25\text{mm}^3$  isotropic voxels, and underwent skull-stripped, intensity normalized, bias field corrected and pre-aligned with affine transformation.

To further validate our model accuracy through segmentation labels, we use manually delineated anatomical structures in LPBA40 dataset. We randomly select 100 pairs of MR images with segmentation labels from LPBA40. We then carefully down-sampled all images and labels from the dimension of  $181 \times 217 \times 181$  to  $128 \times 128 \times 128$ .

## 4.2. Experiments

To validate the registration performance of DeepFLASH on both 2D and 3D data, we run a recent state-of-the-art image registration algorithm FLASH [30, 32] on 1000 2D synthetic data and 2000 pairs of 3D MR images to optimal solutions, which are used as ground truth. We then randomly choose 500 pairs of 2D data and 1000 pairs of 3D MRIs from the rest of the data as testing cases. We set registration parameter  $\alpha = 3$ ,  $c = 6$  for the operator  $\hat{\ell}$ , the number of time steps for Euler integration in geodesic shooting as 10. We adopt the band-limited dimension of the initial velocity field  $\tilde{v}^{opt}$  as 16, which has been shown to produce comparable registration accuracy [30]. We set the batch size as 64 with learning rate  $\eta = 1e-4$  for network training, then run 2000 epochs for 2D synthetic training and 5000 epochs for 3D brain training.

Next, we compare our 2D prediction with registration performed in full-spatial domain. For 3D testing, we compare our method with three baseline algorithms, including FLASH [31] (a fast optimization-based image registration method), Voxelmorph [6] (an unsupervised registration in image space) and Quicksilver [28] (a supervised method that predicts transformations in a momentum space). We also compare the registration time of DeepFLASH with traditional optimization-based methods, such as vector momenta LDDMM (VM-LDDMM) [24], and symmetric diffeomorphic image registration with cross-correlation (SyN) from ANTs [5]. For fair comparison, we train all baseline algorithms on the same dataset and report their best performance from published source codes (e.g., <https://github.com/rkwitt/quicksilver>, <https://github.com/balakg/voxelmorph>).

To better validate the transformations generated by DeepFLASH, we perform registration-based segmentation and examine the resulting segmentation accuracy over eight brain structures, including Putamen (Put), Cerebellum (Cer), Caudate (Caud), Hippocampus (Hipp), Insular cor-

tex (Cor), Cuneus (Cune), brain stem (Stem) and superior frontal gyrus (Gyrus). We evaluate a volume-overlapping similarity measurement, also known as Sørensen–Dice coefficient, between the propagated segmentation and the manual segmentation [12].

We demonstrate the efficiency of our model by comparing quantitative time and GPU memory consumption across all methods. All optimal solutions for training data are generated on an i7, 9700K CPU with 32 GB internal memory. The training and prediction procedure of all learning-based methods are performed on Nvidia GTX 1080Ti GPUs.

## 4.3. Results

Fig.3 visualizes the deformed images, transformation fields, and the determinant of Jacobian (DetJac) of transformations estimated by DeepFLASH and a registration method that performed in full-spatial image domain. Note that the value of DetJac indicates how volume changes, for instance, there is no volume change when  $\text{DetJac} = 1$ , while volume shrinks when  $\text{DetJac} < 1$  and expands when  $\text{DetJac} > 1$ . The value of DetJac smaller than zero indicates an artifact or singularity in the transformation field, i.e., a failure to preserve the diffeomorphic property when the effect of folding occurs. Both methods show similar patterns of volume changes over transformation fields. Our predicted results are fairly close to the estimates from registration algorithms in full spatial domain.

Fig.4 visualizes the deformed images and the determinant of Jacobian with pairwise registration on 3D brain MRIs for all methods. It demonstrates that our method DeepFLASH is able to produce comparable registration results with little to no loss of the accuracy. In addition, our method gracefully guarantees the smoothness of transformation fields without artifacts and singularities.

The left panel of Fig.5 displays an example of the comparison between the manually labeled segmentation and propagated segmentation deformed by transformation field estimated from our algorithm. It clearly shows that our generated segmentations align fairly well with the manual delineations. The right panel of Fig.5 compares quantitative results of the average dice for all methods, with the observations that our method slightly outperforms the baseline algorithms.

Fig. 6 reports the statistics of dice scores (mean and variance) of all methods over eight brain structures from 100 registration pairs. Our method DeepFLASH produces comparable dice scores with significantly fast training of the neural networks.

Table.1 reports the time consumption of optimization-based registration methods, as well as the training and testing time of learning-based registration algorithms. Our method can predict the transformation fields between images approximately 100 times faster than optimization-

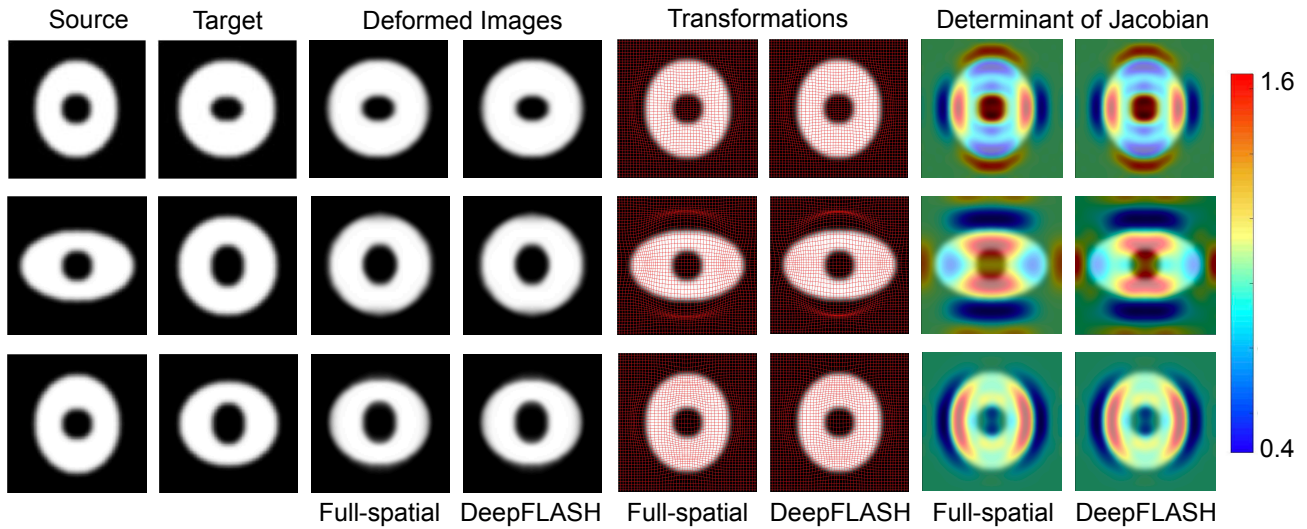


Figure 3. Example of 2D registration results. Left to right: 2D synthetic source, target image, deformed image computed in the full-spatial domain, transformation grids overlaid with source image, and determinant of Jacobian (DetJac) of transformations.

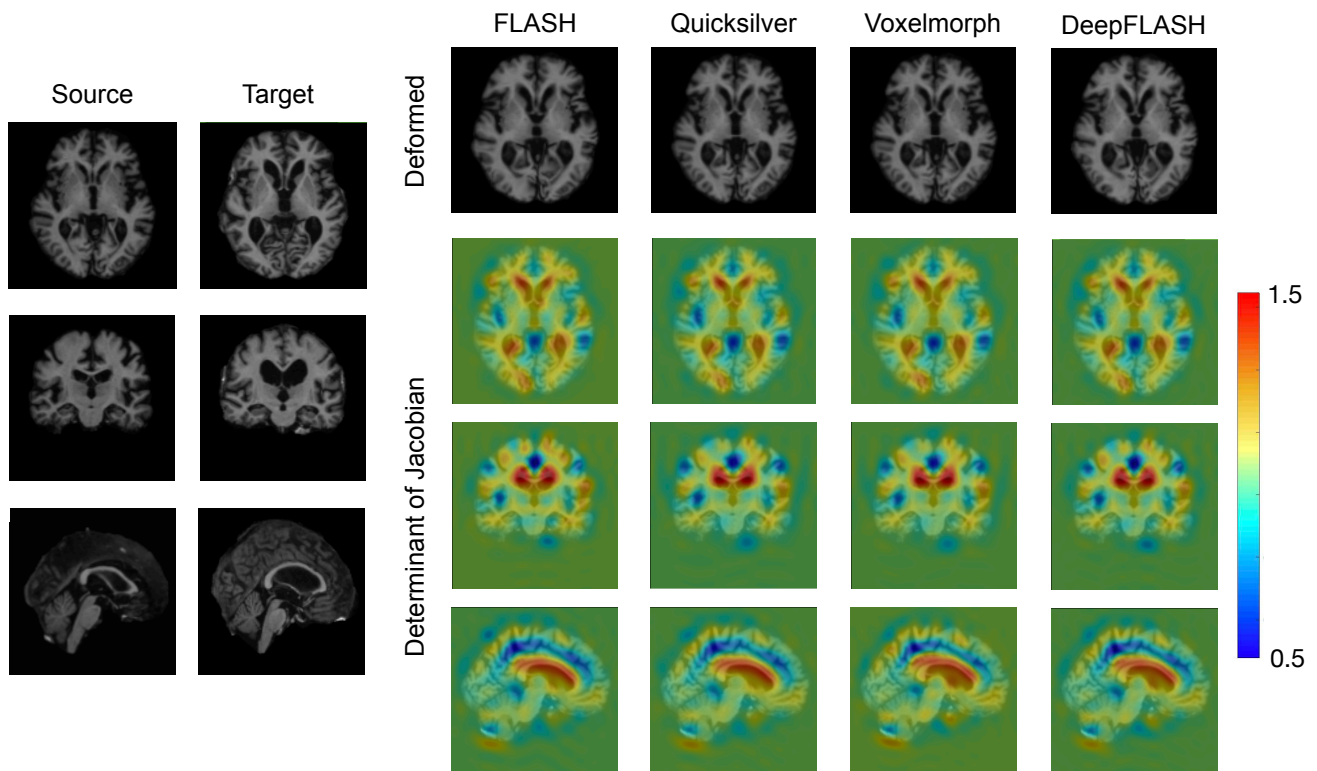


Figure 4. Example of 3D image registration on OASIS dataset. Left panel: axial, coronal, and sagittal view of source and target images. Right panel: deformed images and determinant of Jacobian of the transformations by FLASH, Quicksilver, Voxelmorph, and DeepFLASH.

based registration methods. In addition, DeepFLASH outperforms learning-based registration approaches in testing, while with significantly reduced computational time in training.

Fig.7 provides the average training and testing GPU memory usage across all methods. It has been shown that our proposed method dramatically lowers the GPU footprint compared to other learning-based methods in both training

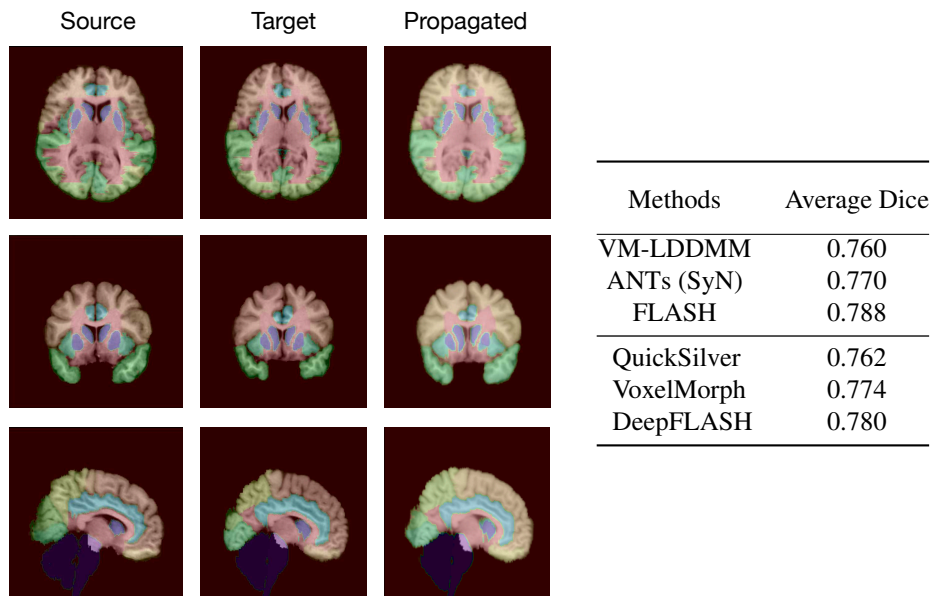


Figure 5. Left: source and target segmentations with manually annotated eight anatomical structures (on LPBA40 dataset), propagated segmentation label deformed by our method. Right: quantitative result of average dice for all methods.

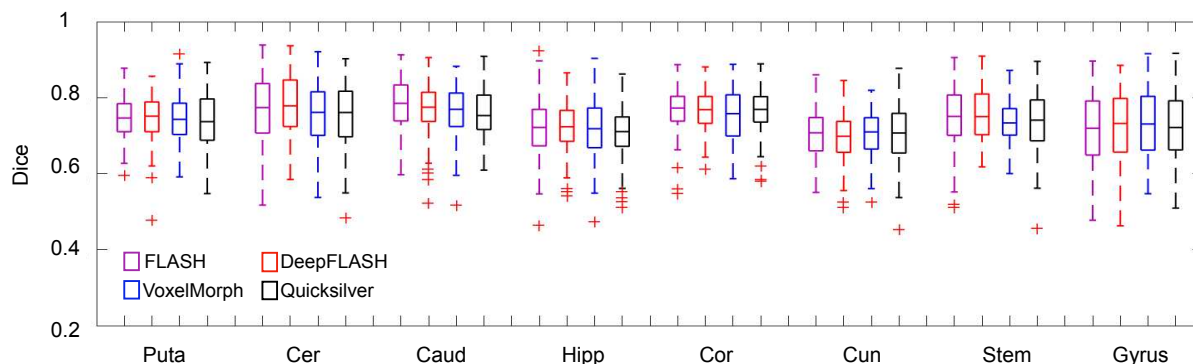


Figure 6. Dice score evaluation by propagating the deformation field to the segmentation labels for four methods on eight brain structures (Putamen(Puta), Cerebellum (Cer), Caudate (Caud), Hippocampus (Hipp), Insular cortex (Cor), Cuneus (Cune), brain stem (Stem), and superior frontal gyrus (Gyrus)).

and testing.

## 5. Conclusion

In this paper, we present a novel diffeomorphic image registration network with efficient training process and inference. In contrast to traditional learning-based registration methods that are defined in the high dimensional imaging space, our model is fully developed in a low dimensional bandlimited space with the diffeomorphic property of transformation fields well preserved. Based on the fact that current networks are majorly designed in real-valued spaces, we are the first to develop a decoupled-network to

solve the complex-valued optimization with the support of rigorous math foundations. Our model substantially lowers the computational complexity of the neural network and significantly reduces the time consumption on both training and testing, while preserving a comparable registration accuracy. The theoretical tools developed in our work is flexible/generic to a wide variety of the state-of-the-art networks, e.g., FCN, or RNN. To the best of our knowledge, we are the first to characterize the diffeomorphic deformations in Fourier space via network learning. This work also paves a way for further speeding up unsupervised learning for registration models.

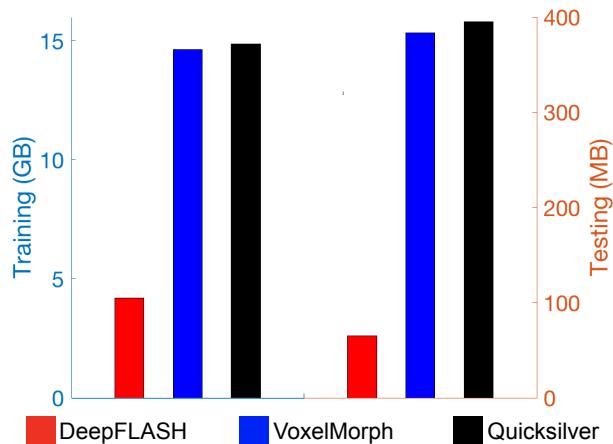


Figure 7. Comparison of average GPU memory usage on our method and learning-based baselines for both training and testing processes.

Table 1. Top: quantitative results of time consumption on both CPU and GPU for optimization-based registration methods (there is no GPU-version of ANTs). Bottom: computational time of training and prediction for our method DeepFLASH and current deep learning registration approaches.

| Methods     | Training Time (GPU hours) | Registration time (sec) |              |
|-------------|---------------------------|-------------------------|--------------|
|             |                           | CPU                     | GPU          |
| VM-LDDMM    | -                         | 1210                    | 262          |
| ANTs(SyN)   | -                         | 6840                    | -            |
| FLASH       | -                         | 286                     | 53.4         |
| Quicksilver | 31.4                      | 122                     | 0.760        |
| Voxelmorph  | 29.7                      | 52                      | 0.571        |
| DeepFLASH   | <b>14.1</b>               | <b>41</b>               | <b>0.273</b> |

## References

- [1] Vladimir Arnold. Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l’hydrodynamique des fluides parfaits. In *Annales de l’institut Fourier*, volume 16, pages 319–361, 1966.
- [2] John Ashburner and Karl. J. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005.
- [3] John Ashburner and Karl J Friston. Diffeomorphic registration using geodesic shooting and gauss–newton optimisation. *NeuroImage*, 55(3):954–967, 2011.
- [4] 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.
- [5] Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.
- [6] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9252–9260, 2018.
- [7] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 2019.
- [8] Mirza 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.
- [9] Tian Cao, Nikhil Singh, Vladimir Jovic, and Marc Niethammer. Semi-coupled dictionary learning for deformation prediction. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 691–694. IEEE, 2015.
- [10] Chen-Rui Chou, Brandon Frederick, Gig Mageras, Sha Chang, and Stephen Pizer. 2d/3d image registration using regression learning. *Computer Vision and Image Understanding*, 117(9):1095–1106, 2013.
- [11] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659, 2014.
- [12] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [13] Anthony F Fotenos, AZ Snyder, LE Gitron, JC Morris, and RL Buckner. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and ad. *Neurology*, 64(6):1032–1039, 2005.
- [14] Kaiming He and Jian Sun. Convolutional neural networks at constrained time cost. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5353–5360, 2015.
- [15] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.
- [16] Sarang Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23, Supplement1:151–160, 2004.
- [17] Michael Leventon, William M Wells III, and Eric Grimson. Multiple view 2d-3d mutual information registration. In *Image Understanding Workshop*, volume 20, page 21. Citeseer, 1997.
- [18] Ruizhi Liao, Esra A Turk, Miaomiao Zhang, Jie Luo, P Ellen Grant, Elfar Adalsteinsson, and Polina Golland. Temporal registration in in-utero volumetric mri time series.



- In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 54–62. Springer, 2016.
- [19] Michael I Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, 24(2):209–228, 2006.
- [20] Marc Niethammer, Yang Huang, and François X Vialard. Geodesic regression for image time-series. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 655–662. Springer, 2011.
- [21] Marc-Michel Rohé, Manasi Datar, Tobias Heimann, Maxime Sermesant, and Xavier Pennec. Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 266–274. Springer, 2017.
- [22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [23] David W Shattuck, Mubeena Mirza, Vitria Adisetiyo, Cornelius Hojatkashani, Georges Salamon, Katherine L Narr, Russell A Poldrack, Robert M Bilder, and Arthur W Toga. Construction of a 3d probabilistic atlas of human cortical structures. *Neuroimage*, 39(3):1064–1080, 2008.
- [24] Nikhil Singh, Jacob Hinkle, Sarang Joshi, and P Thomas Fletcher. A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1219–1222. IEEE, 2013.
- [25] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, João Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep complex networks. *arXiv preprint arXiv:1705.09792*, 2017.
- [26] François X 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.
- [27] Jian Wang, Wei Xing, Robert M Kirby, and Miaomiao Zhang. Data-driven model order reduction for diffeomorphic image registration. In *International Conference on Information Processing in Medical Imaging*, pages 694–705. Springer, 2019.
- [28] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration—a deep learning approach. *NeuroImage*, 158:378–396, 2017.
- [29] Laurent Younes, Felipe Arrate, and Michael I Miller. Evolutions equations in computational anatomy. *NeuroImage*, 45(1):S40–S50, 2009.
- [30] Miaomiao Zhang and P Thomas Fletcher. Finite-dimensional lie algebras for fast diffeomorphic image registration. In *International Conference on Information Processing in Medical Imaging*, pages 249–260. Springer, 2015.
- [31] Miaomiao Zhang and P Thomas Fletcher. Fast diffeomorphic image registration via fourier-approximated lie algebras. *International Journal of Computer Vision*, 127(1):61–73, 2019.
- [32] Miaomiao Zhang, Ruizhi Liao, Adrian V Dalca, Esra A Turk, Jie Luo, P Ellen Grant, and Polina Golland. Frequency diffeomorphisms for efficient image registration. In *International Conference on Information Processing in Medical Imaging*, pages 559–570. Springer, 2017.
- [33] Miaomiao Zhang, William M Wells, and Polina Golland. Low-dimensional statistics of anatomical variability via compact representation of image deformations. In *International conference on medical image computing and computer-assisted intervention*, pages 166–173. Springer, 2016.