

DDR-Net: Dividing and Downsampling Mixed Network for Diffeomorphic Image Registration

Ankita Joshi¹ and Yi Hong²

¹ Department of Computer Science, University of Georgia
ankita.joshi25@uga.edu

² Department of Computer Science and Engineering, Shanghai Jiao Tong University
yi.hong@sjtu.edu.cn

Abstract. Deep diffeomorphic registration faces significant challenges for high-dimensional images, especially in terms of memory limits. Existing approaches either downsample original images, or approximate underlying transformations, or reduce model size. The information loss during the approximation or insufficient model capacity is a hindrance to the registration accuracy for high-dimensional images, e.g., 3D medical volumes. In this paper, we propose a Dividing and Downsampling mixed Registration network (DDR-Net), a general architecture that preserves most of the image information at multiple scales. DDR-Net leverages the global context via downsampling the input and utilizes the local details from divided chunks of the input images. This design reduces the network input size and its memory cost; meanwhile, by fusing global and local information, DDR-Net obtains both coarse-level and fine-level alignments in the final deformation fields. We evaluate DDR-Net on three public datasets, i.e., OASIS, IBSR18, and 3DIRCADB-01, and the experimental results demonstrate our approach outperforms existing approaches. Codes are available –here–.

Keywords: Diffeomorphic image registration · Dividing and downsampling · Multi-scale registration.

1 Introduction

Deformable image registration establishes pixel- or voxel-level dense correspondences for 2D or 3D image pairs, which form a deformation that transforms images into a common space for comparison and analysis. Such a deformation desires a good property of diffeomorphism, a smooth transformation with a smooth inverse, to ensure the preservation of topology when warping images. Classical image registration models, e.g., LDDMM [6], Stationary Velocity Fields (SVF) [2], successfully estimate diffeomorphic deformations for building dense correspondences between image pairs. However, these models face challenges for practical applications, i.e., providing both fast and accurate solutions. Therefore, researchers have been working on improving the efficiency of diffeomorphic image registration methods [3,17].

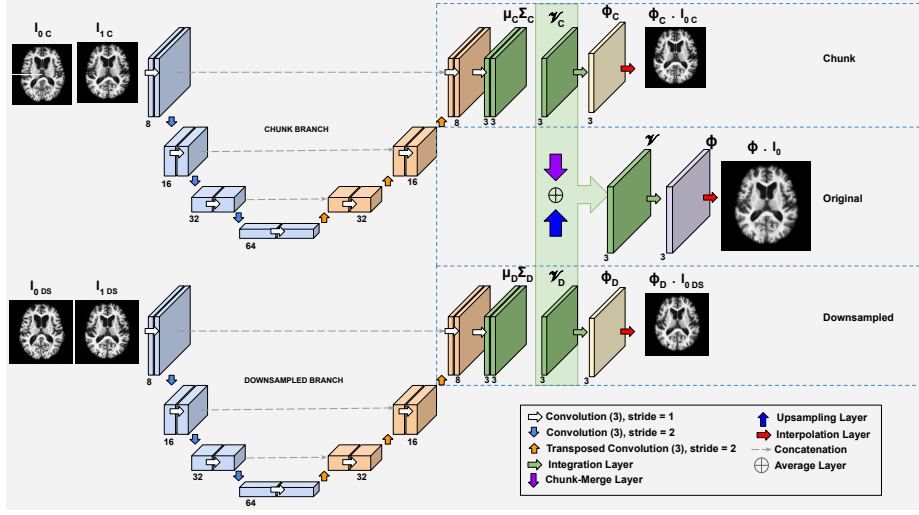


Fig. 1. Architecture of our proposed DDR-Net. Given an image pair I_0 and I_1 , the network estimates deformations at local scale, i.e., chunk branch, the global scale, i.e. the downsampled branch, and the original scale. The deformation ϕ at each level is driven by the corresponding velocity field v , which is sampled from the U-Net outputs, the mean μ and the variance Σ . The deformation at the original resolution is obtained by merging velocity fields generated by the global and local branches and then used to warp the original source images to the corresponding target images by using an interpolation layer. The number of the filters used are under the blocks.

Recently, deep learning based approaches open an alternative to address the above challenges, which motivates our work in this paper. Existing deep registration models focus on tackling the efficiency challenge using supervised [16] or unsupervised techniques [7,12]. Supervised approaches [16] maintain the diffeomorphic property, which is inherited from the classical diffeomorphic model LDDMM, but it requires extra effort to obtain the ground-truth deformations. Meanwhile, its registration accuracy is limited by that of the obtained deformations. The unsupervised approaches [7,12] have shown promising diffeomorphic and efficient registration results by introducing an integration layer into the network design, based on the scaling and squaring method [9]. Due to the flexibility in selecting the network architecture and the loss function, the unsupervised framework has the potential to further improve the registration accuracy. However, because the integration step is computationally expensive and the network faces the memory challenge for high-dimensional images, the unsupervised approaches often work on downsampled images or deformations, which does not fully leverage available information and limits the registration accuracy.

In this paper, we aim to improve the accuracy of unsupervised image registration, using a multi-scale design to integrate image deformations at global, local, and original scales. The difference in our work from current unsupervised

registration models [7,12] is that our model works on downsampled images and chopped chunks, which reduce the computational and memory cost compared to working on the original image size directly. Meanwhile, the integration of these two-scale deformations back to the original scale improves the registration accuracy because of the information fusion at different levels. Previous work [8,11,16] either use multi-scale approaches of downsampling the entire image to improve accuracy of the result or use only patches of the images to reduce the memory cost, but both these techniques do not adequately leverage the data. Therefore, to gain an accuracy boost but not run into memory issues, we propose the Dividing and Downsampling mixed Registration Network (DDR-Net).

Our contributions in this paper are summarized as follows:

- We propose a novel architecture DDR-Net, which can effectively use both global and local features yielding high quality registration performance. Multi-scale information benefits the task of deep image registration.
- We demonstrate an effective way to obtain a trade-off between fully leveraging the available data under limited computing resources and gaining an improved accuracy of diffeomorphic image registration at the same time.
- We conduct extensive experiments on both 2D and 3D datasets with different image types, including brain MRIs and liver CT scans. The experimental results demonstrate that our framework has better registration performance compared to deep-learning-based method VoxelMorph [7] and the classical registration method ANTs SyN [5], in terms of image matching, deformation smoothness, and multi-structure segmentation.

2 Dividing and Downsampling mixed Registration Network (DDR-Net)

Architecture Overview. As shown in Figure 1, our proposed DDR-Net includes three main components: a global branch that handles the registration for downsampled images, a local branch that handles the registration for the cropped local chunks of original images, and an original branch that merges estimated velocity fields from the global and local branches to register original images. Each branch outputs a deformation field ϕ to register image pairs at its corresponding level, and they share some network designs as discussed below.

Backbone Registration. At each scale, we have a diffeomorphic image registration problem. Given an image pair, a source image I_0 and a target image I_1 , each of size $n_x \times n_y \times n_z$, the goal of diffeomorphic image registration is to estimate a smooth deformation field $\phi : \mathbb{R}^{n_x \times n_y \times n_z} \rightarrow \mathbb{R}^{n_x \times n_y \times n_z}$ with a smooth ϕ^{-1} , such that the image deformed from the source, i.e. $\phi \cdot I_0$, is similar to the target image I_1 . Such a diffeomorphic deformation field is driven by a smooth velocity field $v_t, t \in [0, 1]$, via the following differential equation:

$$\frac{d}{dt}\phi = v_t \circ \phi_t, \quad \phi_0 = id. \quad (1)$$

Here, id is an identity deformation. This formulation estimates an optimal velocity field v that drives a deformation field ϕ to match an image pair. So, the registration network has three sub-tasks, i.e., estimating the velocity field, solving Eq. (1) for deformations, and deforming an image with interpolation.

Velocity Field Estimation. The global and local branches follow the same UNet [14] architecture as shown in Fig. 1. The UNet takes in image pairs and outputs the mean μ and the variance Σ for sampling a corresponding stationary velocity field v . Here, the stationary velocity field assumption simplifies the solution of Eq. (1). Given a collection of image pairs $\{(I_0, I_1)\}$, where $I_0, I_1 \in \mathbb{R}^{n_x \times n_y \times n_z}$, the downsampling branch takes the low-resolution image pairs $\{(I_{0D}, I_{1D})\}$, downsampled by half, i.e., $I_{0D}, I_{1D} \in \mathbb{R}^{\frac{n_x}{2} \times \frac{n_y}{2} \times \frac{n_z}{2}}$. The chunk branch receives each input as k divided patches with the same resolution as I_{0D} and I_{1D} , i.e., $k \times \{(I_{0C}, I_{1C})\}$ and $I_{0C}, I_{1C} \in \mathbb{R}^{\frac{n_x}{2} \times \frac{n_y}{2} \times \frac{n_z}{2}}$. These cropped chunks have small overlaps on their boundaries to mitigate the discontinuity of the generated velocity fields at the chunk boundaries, when merging back to form the original high resolution velocity fields. The detailed network architecture in terms of the number of convolution layers and the kernel sizes are shown in Fig. 1.

Deformation Integration. Under that assumption of the stationary velocity field, Equation (1) is simplified with a constant velocity field, and its solution is $\phi = e^v$. Similar to VoxelMorph, we adopt the scaling and squaring algorithm [9] to approximate this solution, which is implemented as a differentiable layer in the network. The downsampling and chunk branches integrate deformations separately, which are driven by their respective velocity fields. In the original branch, we use an averaged global and local velocity field to integrate the deformation. In particular, we upsample the global velocity field to the original resolution using an upsampling layer and merge the k chunks to the original volume using a chunk-merge layer. As a result, we obtain a velocity field at the original resolution which is integrated to get the deformation at the original level.

Image Interpolation. We use an interpolation layer to deform the source image at each branch. For each voxel p in the target image, we compute its location $\phi(p)$ in the source image and compute its intensity value using linear interpolation. This differentiable operation allows the backpropagation of the network errors.

Loss Functions. All the velocity fields generated in the network have the constraint to ensure the smoothness of the deformations. Each branch also outputs a deformed source image to match the corresponding target image at each level, i.e., the downsampled, chunk, and original scales. We use the Mean Square Error (MSE) to measure the matching results and a K-L divergence loss to encourage the smoothness of the velocity field as done in [7,11].

3 Experiments

We evaluate our method on three public datasets, which involves two types of 3D medical images, i.e., brain MRI scans and liver CT scans.

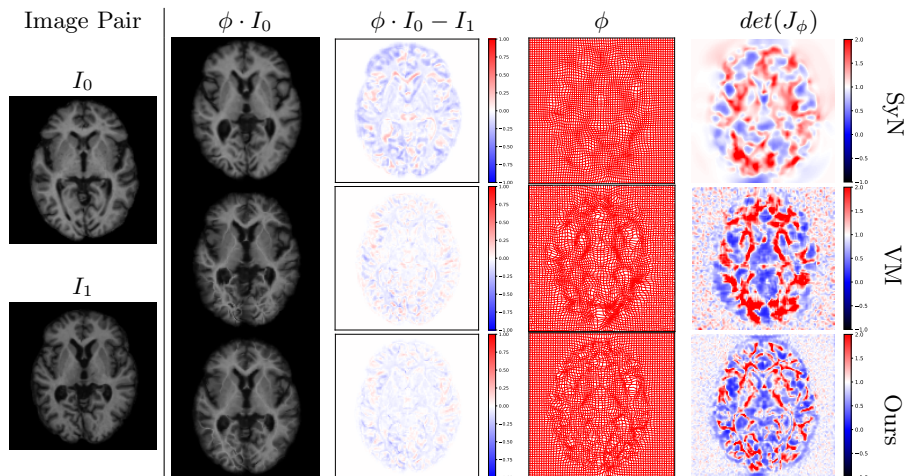


Fig. 2. Registration comparison among SyN, VM (VoxelMorph), and our DDR-Net. Left to right: the image pair median slice from OASIS dataset, the source image I_0 and the target image I_1 ; the warped image; the image difference between warped image and the target image; the deformation ϕ ; the determinant of the deformation Jacobian.

OASIS Dataset [13]. The T1-weighted brain scans from the OASIS dataset underwent pre-processing steps including down-sampling, skull-stripping, intensity normalization to the range $[0, 1]$, bias field correction, and co-registration with affine transformations. We have resulting images of resampled resolution $128 \times 128 \times 128$ and the voxel size of $1.25 \times 1.25 \times 1.25 mm^3$. We collect 360 volume pairs for the 3D experiments.

IBSR18 Dataset [15]. The IBSR18 dataset consists of T1-weighted scans for 18 subjects with dimensions of $256 \times 128 \times 256$. We resampled these images to $128 \times 128 \times 128$. Scans underwent preprocessing steps including skull-stripping, bias field correction, and intensity normalization. Each scan also comes with 84 manually labelled anatomical structures. We collect 306 3D image pairs. Segmentation maps including 28 anatomical structures (see Table 2) were also obtained in this dataset, which followed the same preprocessing steps of resampling.

3DIRCADB-01 Dataset [1]. The 3DIRCADB-01 dataset contains 20 CT scans with masks of the segmented structures available for bone, liver, and skin. Scans underwent preprocessing steps including histogram equalization and intensity normalization to the range $[0, 1]$. We collect and pair the central slices on the axial plan and obtained 380 2D image pairs with a resolution of 512×512 .

Experimental Settings. For all the experiments, we use 70% of the collected data for training, 10% for validation, and 20% for test, after a random shuffle. We use the Adam [10] optimizer with a learning rate of e^{-4} . All the experiments were trained using NVIDIA GeForce TITAN X GPUs.

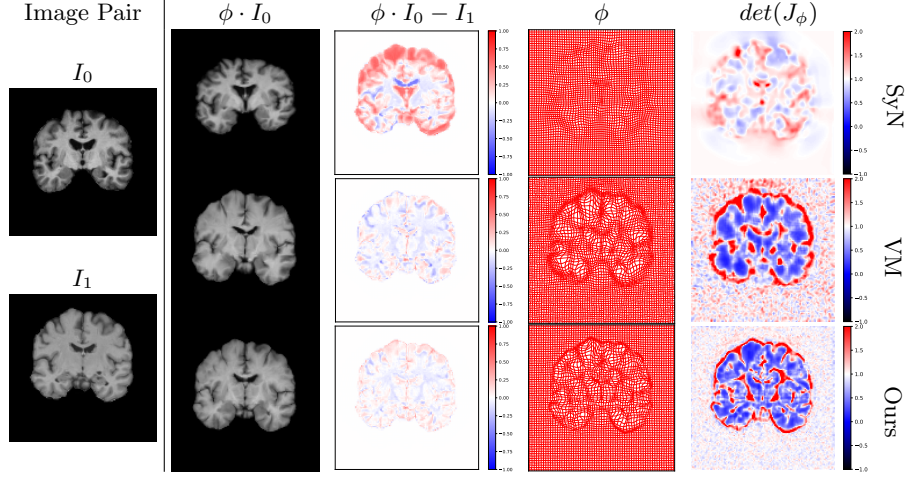


Fig. 3. Registration comparison among SyN, VM (VoxelMorph), and our DDR-Net for the IBSR dataset. Left to right: the same with Fig. 2.

Dataset	Method	RMSE (e^{-3})	Foldings	
			$ \sum det(J_\phi) < 0 $ (e^{-5})	$\frac{1}{N} \sum \delta(det(J_\phi) < 0)$ (Ratio: % ₀₀)
OASIS	SyN [5]	4.79 ± 0.001	495.37 ± 0.03	0.1 ± 0.768
	VM [7]	1.25 ± 0.001	1.46 ± 0.005	0.051 ± 0.014
	DDR-Net	1.10 ± 0.001	1.09 ± 0.000	0.003 ± 0.019
IBSR18	SyN [5]	10.93 ± 0.26	0.00 ± 0.000	0.00 ± 0.00
	VM [7]	2.06 ± 0.001	20.17 ± 0.005	0.05 ± 0.014
	DDR-Net	1.67 ± 0.001	17.266 ± 0.000	0.04 ± 0.117
3DIRCADB-01	SyN [5]	68.04 ± 0.024	476.98 ± 0.001	35.32 ± 5.227
	VM [7]	16.84 ± 0.005	1221.21 ± 0.01	0.72 ± 0.274
	DDR-Net	14.11 ± 0.004	723.89 ± 0.003	0.65 ± 0.247

Table 1. Comparison of the registration performance on all the three datasets.

We measure the image matching after registration using the intensity root mean square error (RMSE), and the smoothness of a deformation by counting the number of its foldings and the absolute sum of negative determinant of its Jacobian [3]. Also, we perform the segmentation task using image registration and report the Dice score in terms of volume overlap.

We compare our approach with ANTsPy package using Symmetric Normalization (SyN) [5]. We use cross-correlation and other default settings, which are optimal for our task with 201 iterations for registration. Another baseline algorithm is VoxelMorph [7], and we also use its default settings for comparison. Initial affine registration is not performed for any of the experiments.

Experimental Results. Figures 2, 3, 4 show qualitative results for OASIS, IBSR18, and 3DIRCADB-01 datasets, respectively. Overall, our approach produces better matching results as compared to the baseline methods, demon-

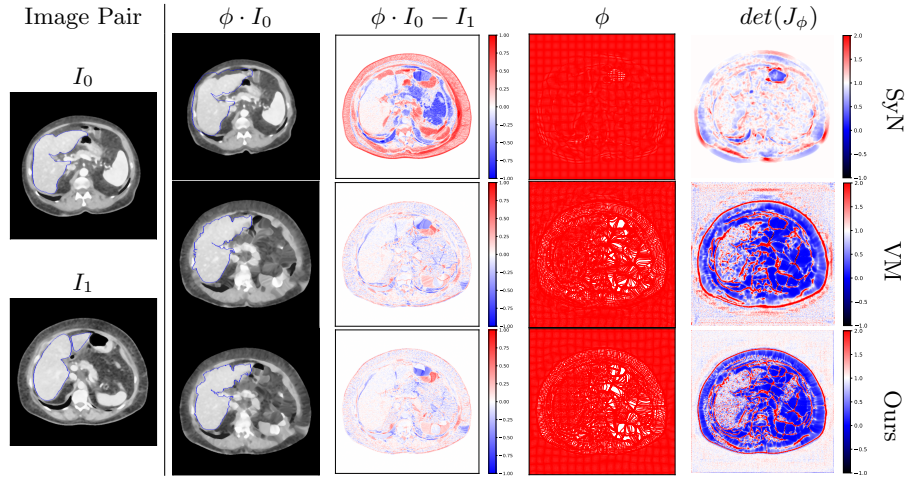


Fig. 4. Registration comparison among SyN, VM (VoxelMorph), and our DDR-Net for the 3DIRCADB-01 dataset. Left to right: the same with Fig. 2.

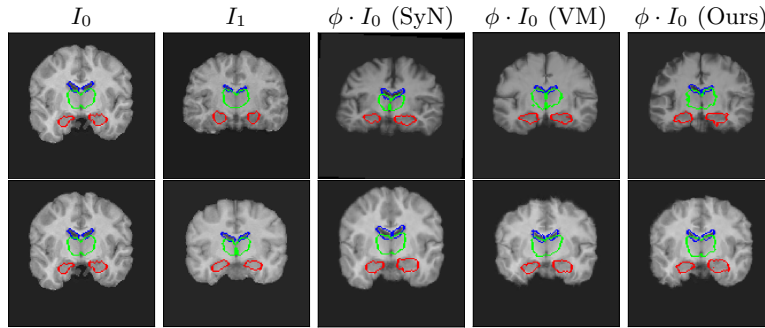


Fig. 5. IBSR18 sampled source, target, and warped images for SyN, VoxelMorph (VM), and our DDR-Net. Blue: ventricles, green: thalami, red: hippocampi.

strated by the image difference plots. VoxelMorph tends to produce unwanted artifacts in its deformed images, which are not seen in our results. The deformation plot and the visualization of the determinant of the Jacobian show that our results have less deformations happening in the background, as expected. Also, our method achieves significantly smaller deformations throughout the image space and provides a better matching result to the target at the same time.

We observe that the VoxelMorph produces more irregular deformation fields compared to the other methods. Table 1 show the quantitative analysis of the registration performance on the OASIS, IBSR18, and 3DIRCADB-01 datasets. We achieve consistently lower number of foldings compared to the baseline methods while maintaining the consistently higher registration accuracy. Table 2 indicates the Dice score of anatomical structures for the IBSR18 and 3DIRACADB-01

Dataset	Region	SyN [4]	VM [7]	DDR-Net
IBSR18	3rd ventricle	0.28 ± 0.14	0.42 ± 0.16	0.44 ± 0.17
	4th ventricle	0.20 ± 0.13	0.39 ± 0.16	0.36 ± 0.16
	amygdala	0.34 ± 0.13	0.27 ± 0.23	0.31 ± 0.20
	brainstem	0.62 ± 0.12	0.69 ± 0.13	0.66 ± 0.16
	caudate	0.48 ± 0.09	0.40 ± 0.21	0.40 ± 0.20
	cerebellum cortex	0.55 ± 0.12	0.59 ± 0.20	0.63 ± 0.13
	cerebellum white matter	0.42 ± 0.14	0.34 ± 0.20	0.44 ± 0.18
	cerebral cortex	0.43 ± 0.09	0.55 ± 0.18	0.58 ± 0.13
	cerebral white matter	0.53 ± 0.06	0.58 ± 0.15	0.59 ± 0.11
	csf	0.22 ± 0.17	0.31 ± 0.16	0.31 ± 0.17
	hippocampus	0.32 ± 0.11	0.17 ± 0.14	0.34 ± 0.21
	lateral ventricle	0.38 ± 0.12	0.53 ± 0.18	0.45 ± 0.17
	pallidum	0.24 ± 0.16	0.23 ± 0.23	0.29 ± 0.24
	putamen	0.29 ± 0.21	0.27 ± 0.25	0.33 ± 0.27
	thalamus	0.68 ± 0.08	0.61 ± 0.18	0.60 ± 0.19
	ventraldc	0.49 ± 0.11	0.49 ± 0.21	0.54 ± 0.18
	Avg. Dice	0.28 ± 0.14	0.43 ± 0.1	0.47 ± 0.1
3DIRCADB-01	bone	0.223 ± 0.106	0.356 ± 0.149	0.369 ± 0.153
	skin	0.889 ± 0.065	0.982 ± 0.012	0.991 ± 0.009
	liver	0.690 ± 0.099	0.731 ± 0.122	0.741 ± 0.131

Table 2. Segmentation comparison using SyN, VoxelMorph (VM) and DDR-Net.

datasets. Our method performs better on most of the anatomical structures for both the IBSR18 and 3DIRCADB-01 datasets. Figure 5 shows the segmentation maps for a few anatomical structures from the IBSR18 dataset.

Regarding the inference time, ANTs SyN, which is tested on CPU since it does not have a GPU implementation, takes on an average 20 minutes, whereas VoxelMorph tested on GPU takes 121 milliseconds, while DDR-Net tested on GPU takes 425 milliseconds to register an image pair from the OASIS dataset. A limited ablation study conducted by us revealed that a single scale UNet architecture working on the full size velocity field integration for 2D input size of 128×128 for a single input image pair takes 26.46 MB of memory, whereas an architecture like DDR-Net for the same input and UNet architecture will take 21.16 MB of memory. The memory requirement was calculated considering the forward pass and parameters in the neural network. For 3D cases, our DDR-Net can take in more images in a batch, compared to a single scale UNet.

4 Conclusion and Discussion

In this paper, we have proposed a diffeomorphic image registration model which estimates smoother deformations and provides a better image matching result. It leverages both the global context and local fine structures of the data effectively to enhance the registration result and handle large deformations as a result of such an architecture. Generally, our approach produced more regular deformation fields, which are significantly smoother than the baseline methods. The dice

scores indicate that our approach is comparable to the existing methods even surpassing them in most cases. Moreover, our architecture shows that simple yet efficient changes in architectures can lead to better deep learning strategies. We believe that an optimal balance of GPU memory and accuracy is essential for registering image pairs in 3D in order to maximize the utilization of 3D information. Our work will push new avenues of research in considering memory efficient registration models in deep learning, which fully leverage the data.

References

1. 3dircadb-01. <https://www.ircad.fr/research/3dircadb/>
2. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 924–931. Springer (2006)
3. Ashburner, J.: A fast diffeomorphic image registration algorithm. *Neuroimage* **38**(1), 95–113 (2007)
4. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* **12**(1), 26–41 (2008)
5. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage* **54**(3), 2033–2044 (2011)
6. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision* **61**(2), 139–157 (2005)
7. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 729–738. Springer (2018)
8. Hering, A., van Ginneken, B., Heldmann, S.: mlvirnet: Multilevel variational image registration network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 257–265. Springer (2019)
9. Higham, N.J.: The scaling and squaring method for the matrix exponential revisited. *SIAM Journal on Matrix Analysis and Applications* **26**(4), 1179–1193 (2005)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
11. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging* **38**(9), 2165–2176 (2019)
12. Krebs, J., Mansi, T., Mailhé, B., Ayache, N., Delingette, H.: Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 101–109. Springer (2018)
13. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience* **19**(9), 1498–1507 (2007)

14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
15. Valverde, S., Oliver, A., Cabezas, M., Roura, E., Lladó, X.: Comparison of 10 brain tissue segmentation methods using revisited ibsr annotations. *Journal of Magnetic Resonance Imaging* **41**(1), 93–101 (2015)
16. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration—a deep learning approach. *NeuroImage* **158**, 378–396 (2017)
17. Zhang, M., Fletcher, P.T.: Finite-dimensional lie algebras for fast diffeomorphic image registration. In: International conference on information processing in medical imaging. pp. 249–260. Springer (2015)