Fast Predictive Simple Geodesic Regression

Zhipeng Ding[1], Greg Fleishman[3,4], Xiao Yang[1], Paul Thompson[3], Roland Kwitt[4], Marc Niethammer[1,2], The Alzheimer's Disease Neuroimaging Initiative

[1]Department of Computer Science, UNC Chapel Hill, USA
[2]Biomedical Research Imaging Center, UNC Chapel Hill, USA
[3]Imaging Genetics Center, USC, USA
[4]Department of Radiology, University of Pennsylvania, USA
[5]Department of Computer Science, University of Salzburg, Austria

Abstract. Analyzing large-scale imaging studies with thousands of images is computationally expensive. To assess localized morphological differences, deformable image registration is a key tool. However, as registrations are costly to compute, large-scale studies frequently require large compute clusters. This paper explores a fast predictive approximation to image registration. In particular, it uses these fast registrations to approximate a simplified geodesic regression model to capture longitudinal brain changes. The resulting approach is orders of magnitude faster than the optimization-based regression approach and hence facilitates large-scale analysis on a single graphics processing unit. We show results on 2D and 3D brain magnetic resonance images from OASIS and ADNI.

1 Introduction

Imaging studies on brain development, diseases, and aging will continue to dramatically increase in size: ADNI [10] and the Rotterdam study [9] contain thousands of subjects and the UK Biobank [1] targets an order of 100,000 images once completed. Analyzing large-scale studies can quickly become computationally prohibitive; compute clusters are commonly used to parallelize analyses such as deformable image registrations of 3D brain images. Looking ahead, very large-scale studies will require larger compute clusters or efficient algorithms to reduce computational costs. Furthermore, an increasing number of longitudinal studies require efficient algorithms for the analysis of longitudinal image data.

Geodesic regression (GR) [5,12,14] is an attractive approach to capture trends in longitudinal imaging data. GR generalizes linear regression to Riemannian manifolds. Applied to longitudinal image data it compactly expresses spatial image transformations over time. Unfortunately, GR requires the solution of a computationally costly optimization problem. Hence, a simplified, approximate, GR approach (SGR) has been proposed [6] that allows for the computation of the regression geodesic via pairwise image registrations. But even SGR would require months of computation time on a single graphics processing unit (GPU) to process thousands of 3D image registrations for large-scale imaging studies.

However, efficient approaches for deformable image registration have recently been proposed. In particular, for the large displacement diffeomorphic metric

mapping (LDDMM) model, which is the basis of GR approaches for images, registrations can be dramatically sped up by either working with finite-dimensional Lie algebras [18] or by fast predictive image registration (FPIR) [15,16]. FPIR learns a patch-based deep regression model to predict the initial momentum of LDDMM, which characterizes the spatial transformation. By replacing numerical optimization of standard LDDMM registration by a *single* prediction step followed by optional correction steps [16], FPIR can improve speed by orders of magnitude. Besides FPIR, other predictive image registration (i.e., optical flow) approaches exist [2,11,13]. However, FPIR is better suited for brain image registration than these optical flow approaches for the following reasons: (1) FPIR predicts the initial momentum of LDDMM and therefore inherits its theoretical properties. Consequentially, FPIR results in diffeomorphic transformations and geodesics even though predictions are patch-wise. This can not be guaranteed by optical flow prediction methods. (2) Patch-wise prediction allows training of the prediction models based on a very small number of images, utilizing a large number of patches. (3) By using patches instead of full images, predictions for large, high-resolution images are possible despite the memory constraints of current GPUs. (4) None of the existing prediction methods address longitudinal image data. Additionally, as both FPIR and SGR are based on LDDMM, they naturally integrate and hence result in our proposed *fast predictive simple geodesic regression* (FPSGR) approach.

Contributions. (1) *Predictive Geodesic Regression*: We introduce the first fast predictive geodesic regression approach for images. (2) *Large-scale dataset capability*: Our predictive approach facilitates large-scale image registration / regression within a day on a single GPU instead of months for optimization-based methods. (3) *Accuracy*: We show that FPSGR approximates the optimization-based simple GR result well. (4) *Validation*: We showcase the performance of our FPSGR approach by analyzing $> 6,000$ images of the ADNI-1/ADNI-2 datasets.

Organization. Sec. 2 describes FPSGR. Sec. 3 discusses the experimental setup and the training of the prediction models. Sec. 4 presents experimental results for 2D and 3D magnetic resonance images (MRI) of the brain. The paper concludes with a summary and an outlook on future work.

2 Fast Predictive Simple Geodesic Regression (FPSGR)

Our fast predictive simple geodesic regression involves two main components: (1) the fast predictive image registration (FPIR) approach and (2) integration of FPIR into the simple geodesic regression (SGR) formulation. Both FPIR and SGR are based on the shooting formulation of LDDMM [14]. Fig. 1 illustrates our overall approach. The individual components are described in the following.

LDDMM. Shooting-based LDDMM and geodesic regression minimize

$$E(I_0, m_0) = \frac{1}{2}\langle m_0, Km_0\rangle + \frac{1}{\sigma^2}\sum_i d^2(I(t_i), Y^i), \tag{1}$$

$$s.t. \quad m_t + \mathrm{ad}^*_v m = 0, I_t + \nabla I^T v = 0, m - Lv = 0,$$

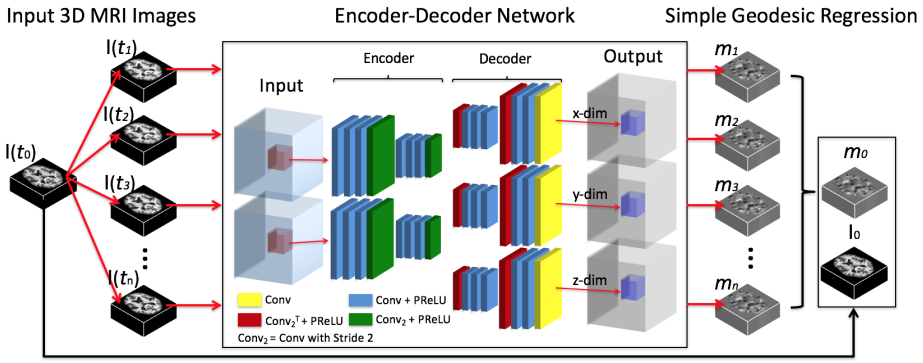

Fig. 1: Principle of Fast Predictive Simple Geodesic Regression (FPSGR). In the Encoder-Decoder network, the inputs are patches from source images and target images at the same position; the outputs are the predicted initial momenta of the corresponding patches. In simple geodesic regression, all the pairwise initial momenta are averaged according to Eq. 2 to produce the initial momentum of the regression geodesic.

where I_0 is the initial image (known for image-to-image registration and to be determined for geodesic regression), m_0 is the initial momentum, K is a smoothing operator that connects velocity v and momentum m as $v = Km$ and $m = Lv$ with $K = L^{-1}$, $\sigma > 0$ is a weight, Y^i is the measured image at time t_i (there will only be one such image for image-to-image registration at $t = 1$), and $d^2(I_1, I_2)$ denotes the image similarity measure between I_1 and I_2 (for example L_2 or geodesic distance); ad^* is the dual of the negative Jacobi-Lie bracket of vector fields: $\mathrm{ad}_v w = -[v, w] = Dvw - Dwv$ and D denotes the Jacobian. The deformation of the source image $I_0 \circ \Phi^{-1}$ can be computed by solving $\Phi_t^{-1} + D\Phi^{-1}v = 0$, $\Phi_0^{-1} = \mathrm{id}$, where id denotes the identity map.

FPIR. Fast predictive image registration [15,16] aims at predicting the initial momentum, m_0, between a source and a target image patch-by-patch. Specifically, we use a deep encoder-decoder network to predict the patch-wise momentum. As shown in Fig. 1, in 3D the inputs are two layers of $15 \times 15 \times 15$ image patches (15×15 in 2D), where the two layers are from the source and target images respectively. Two patches are taken at the same position by two parallel encoders, which learn features independently. The output is the predicted initial momentum in the x, y and z directions (obtained by numerical optimization for the training samples). The network can be divided into Encoder and Decoder. An encoder consists of 2 blocks of 3 $3 \times 3 \times 3$ convolutional layers with PReLU activations, followed by another $2 \times 2 \times 2$ convolution+PReLU with a stride of two, serving as "pooling" operation. The number of features in the first convolutional layer is 64 and increases to 128 in the second. In the Decoder, three parallel decoders share the same input generated from the encoder. Each decoder is the inverse of the encoder except for using 3D transposed convolution layers with a stride of two to perform "unpooling", and no non-linearity layer at the end. To speed up computations, we use patch pruning (i.e., patches outside the brain are not predicted as the momentum is zero there) and a large pixel stride (e.g. 14 for $15 \times 15 \times 15$ patches) for the sliding window of the predicted patches.

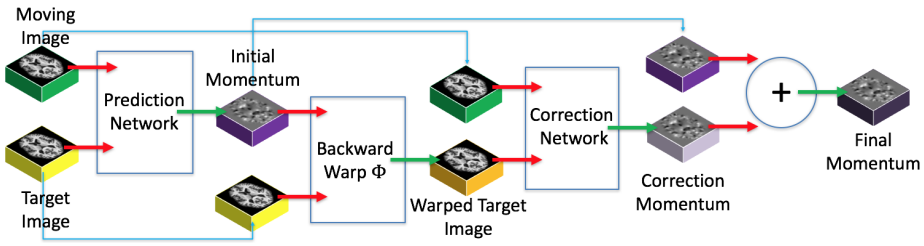

Fig. 2: Prediction + correction network architecture. (1) Predict initial momentum m_p and corresponding backward deformation, Φ. (2) Predict correction of initial momentum, m_c, based on the difference between moving image and warped-back target image. The final momentum is $m = m_p + m_c$. The correction network is trained with the moving images and warped-back target images of the training dataset as inputs.

Correction Network. We also use an additional correction step to improve prediction accuracy, by training a correction network as described in [16]. Fig. 2 illustrates this approach graphically. The correction network has the same structure as the prediction network. Only the inputs and outputs change.

SGR. Determining the initial image, I_0, and the initial momentum, m_0, of Eq. 1 is computationally costly. In simple geodesic regression, the initial image is fixed to the first image of a subject's longitudinal image set. Furthermore, the similarity measure $d(\cdot,\cdot)$ is chosen as the geodesic distance between images and *approximated* so that the geodesic regression problem can by computed by computing pair-wise image registrations with respect to the first image. The approximated optimal m_0 of energy functional (1) for a fixed I_0 is then

$$m_0 \approx \frac{\sum(t_i - t_0)^2 m_0^{Y_i}}{\sum(t_i - t_0)^2} = \frac{\sum(t_i - t_0)\overline{m}_0^{Y_i}}{\sum(t_i - t_0)^2} \tag{2}$$

where $\overline{m}_0^{Y_i}$ is obtained by registering I_0 to Y^i in unit time followed by a rescaling of the momentum to account for the original time duration: $m_0^{Y_i} = \frac{1}{t_i - t_0}\overline{m}_0^{Y_i}$. In our FPSGR approach we compute the momenta $\overline{m}_0^{Y_i}$ via FPIR.

3 Experimental Setup and Training of Prediction Models

We perform experiments on 2D axial MRI brain slices from the OASIS dataset and on 3D images from the ADNI dataset. **2D:** We verify our hypothesis that training FPIR on longitudinal data for longitudinal registrations is preferred over training using cross-sectional data; and that training FPIR on cross-sectional data for cross-sectional registrations is preferred over training using longitudinal data. Comparisons are with respect to registration results obtained by numerical optimization. **3D:** As the ADNI dataset is longitudinal and based on our findings for the 2D data, we train our models with longitudinal registrations only.

Training and Testing of the Prediction Models. We use a set of 150 patients' MRIs of the OASIS dataset for training the 2D model and testing the performance of FPIR. The resampled 2D images of size 128×128 are extracted

2D Longitudinal Test Case Deformation Error [pixel]							
Data Percentile	0.3%	5%	25%	50%	75%	95%	99.7%
Longitudinal Training	**0.0027**	**0.0112**	**0.0267**	**0.0425**	**0.0630**	**0.1222**	**0.2221**
Cross-sectional Training	0.0050	0.0201	0.0475	0.0744	0.1093	0.1862	0.2253
2D Cross-sectional Test Case Deformation Error [pixel]							
Data Percentile	0.3%	5%	25%	50%	75%	95%	99.7%
Longitudinal Training	0.0256	0.1068	0.2669	0.4552	0.7433	1.4966	1.9007
Cross-sectional Training	**0.0120**	**0.0508**	**0.1248**	**0.2047**	**0.3196**	**0.5781**	**0.6973**

Table 1: Deformation error of longitudinal and cross-sectional models tested on longitudinal and cross-sectional data. 2-norm deformation errors in pixels w.r.t. the ground truth deformation obtained by numerical optimization for LDDMM.

from the same axial slice of the 3D OASIS images after affine registration. The ADNI data consists of 6471 3D MR brain images of size $220 \times 220 \times 220$ (3479 ADNI-1, 833 subjects; 2992 ADNI-2, 823 subjects). We use all of the ADNI data. **2D:** We used the first 100 patients for training and the last 50 for testing. For longitudinal training, we registered the first and the second time-point of a patient. For cross-sectional training, we registered a patient's first time point to another patient's second time point. To test longitudinal and cross-sectional registrations, we perform the same type of registrations on the 50 test sets. We compare the deformation error against the LDDMM solution obtained by numerical optimization. Tab. 1 shows the results which confirm our hypothesis that training the prediction model with longitudinal registration cases is preferred for longitudinal registration over training with cross-sectional data. The deformation error is very small for longitudinal training / testing and the predictive method has comparable performance to costly optimization-based LDDMM. These results indicate that it is beneficial to train a prediction model with deformations it is expected to encounter, i.e., relatively small deformations for longitudinal registrations and larger deformations for cross-sectional registration. Hence, for the ADNI data, we train the 3D models using longitudinal registrations only. **3D:** We randomly picked 1/6 patients from each diagnostic category to form a group of 139 patients for training in ADNI-1 and 150 in ADNI-2. The images of the first time-point were registered to all the later time-points within each patient. Matching the distribution of diagnostic categories, we then randomly picked a subset of 165 of these registrations as the training set in ADNI-1 and 140 in ADNI-2. We trained 4 prediction models and their corresponding correction models totaling 8 prediction models, i.e. ADNI-1 Pred-1, ADNI-1 Pred+Corr-1, ADNI-1 Pred-2, ADNI-1 Pred+Corr-2 and analogously for ADNI-2. The training sets within ADNI-1/2 respectively were not overlapping, allowing us to compute predictions (not using the training data) for the *complete* ADNI-1/2 datasets.

Parameter selection. We use the regularization kernel $K = L^{-1} = (-a\nabla^2 - b\nabla(\nabla\cdot) + c)^{-2}$ with $[a, b, c]$ set to $[0.05, 0.05, 0.005]$ for 2D and $[1, 0, 0.1]$ for 3D images; σ is 0.1 for 2D and 3D. We use *Adam* to optimize the network with 10 epochs and learning rates of 0.0005 and 0.0001 in 2D and 3D respectively.

Efficiency. Once trained, the prediction models allow fast computations of registrations. We use a TITAN X(Pascal) GPU and *PyTorch*. In 2D, FPIR took 4 seconds on average to predict 50 pairs of 128×128 MRIs (for the 4 longitudinal/cross-sectional experiments), while a GPU implementation of optimization-based LDDMM took 150 minutes. For the 3D ADNI-1 data ($220 \times$

220×220 MRIs), FPSGR took about one day to predict 2646 pairwise registrations (25s/prediction) and to compute the regression result. Optimization-based LDDMM would require ≈ 40 days of runtime. Runtime for FPIR on ADNI-2 is identical to ADNI-1 as the images have the same dimension.

4 Discussion of experimental Results for 3D ADNI data

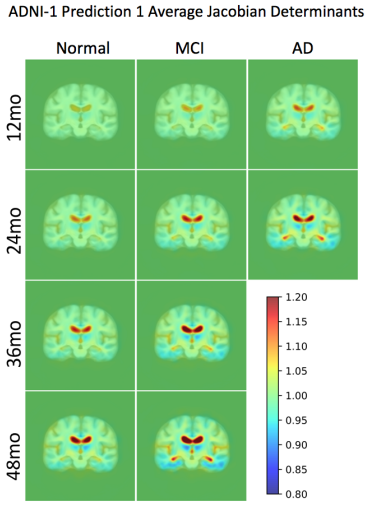

Fig. 3: Average Jacobian determinant over time and diagnostic category for one of the ADNI-1 groups.

We studied 10 experimental groups: all subjects from ADNI-1 using traditional LDDMM, two groups of ADNI-1 with different training data using FPSGR without correction network, and two groups of ADNI-1 using FPSGR with correction network. An analogous set of groups was studied for ADNI-2. We calculated the Jacobian determinant (JD) for every computed deformation. The JDs were then warped to a common coordinate system for the entire ADNI data set using existing deformations from [4,3]. Each such spatially normalized JD was then averaged within a region of interest (ROI). Specifically, we quantify atrophy as $(1 - \frac{1}{|\omega|}\int_\omega |D\phi(x)|dx) \times 100$, where $|\cdot|$ is the determinant, and ω is an area in the temporal lobes which was determined in prior studies [4,3] to be significantly associated with accelerated atrophy in Alzheimer's disease (AD). The resulting scalar value is an estimate of the relative volume change experienced by that region between the baseline and followup image acquisitions. Hence, its sign is positive when the region has lost volume over time and is negative if the region has gained volume over time. See Fig. 3.

ADNI-1		Slope	Intercept
Normal	LDDMM-1	[0.65, **0.73**, 0.81]	[-0.23, **-0.07**, 0.09]
	Pred-1	[0.39, **0.45**, 0.52]	[-0.20, **-0.08**, 0.05]
	Pred+Corr-1	[0.64, **0.71**, 0.77]	[-0.14, **-0.01**, 0.13]
	LDDMM-2	[0.63, **0.72**, 0.81]	[-0.21, **-0.03**, 0.15]
	Pred-2	[0.49, **0.56**, 0.63]	[-0.17, **-0.03**, 0.11]
	Pred+Corr-2	[0.56, **0.64**, 0.71]	[-0.14, **0.00**, 0.14]
MCI	LDDMM-1	[1.55, **1.68**, 1.82]	[-0.28, **-0.04**, 0.19]
	Pred-1	[1.03, **1.13**, 1.23]	[-0.24, **-0.07**, 0.10]
	Pred+Corr-1	[1.39, **1.50**, 1.61]	[-0.14, **0.05**, 0.25]
	LDDMM-2	[1.41, **1.54**, 1.66]	[-0.25, **-0.03**, 0.18]
	Pred-2	[1.15, **1.25**, 1.36]	[-0.20, **-0.02**, 0.16]
	Pred+Corr-2	[1.22, **1.33**, 1.43]	[-0.16, **0.02**, 0.20]
AD	LDDMM-1	[2.07, **2.42**, 2.77]	[-0.28, **0.14**, 0.57]
	Pred-1	[1.31, **1.57**, 1.84]	[-0.22, **0.11**, 0.43]
	Pred+Corr-1	[1.83, **2.13**, 2.42]	[-0.14, **0.22**, 0.58]
	LDDMM-2	[2.03, **2.39**, 2.75]	[-0.29, **0.14**, 0.57]
	Pred-2	[1.66, **1.94**, 2.23]	[-0.21, **0.13**, 0.47]
	Pred+Corr-2	[1.75, **2.03**, 2.32]	[-0.18, **0.16**, 0.51]

Table 2: Slope and intercept values for simple linear regression of volume change over time [Lower end of 95% C.I., **point estimate**, Higher end of 95% C.I.]

We limited our experiments to the applications in [8,7], wherein nonlinear registration/regression is used to quantify atrophy within regions known to be associated to varying degrees with AD (2), Mild Cognitive Impairment (MCI) (1) (including Late Mild Cognitive Impairment), and normal aging (0) in an elderly population.

Moreover, we are interested in the following two critical validations: (1) Are atrophy measurements derived from FPSGR biased to over- or underestimate change? (2) Are those atrophy measurements consistent with those derived from deformations given by the optimization procedure (LDDMM) which produced the training data set? If

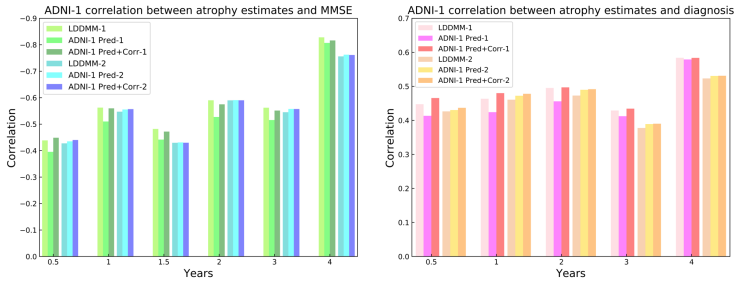

Fig. 4: FPSGR-derived correlations with clinical variables.

experiments show (1) and (2) resolve favorably, then the substantially improved computational efficiency of FPSGR may justify its use for some applications.

Bias. Estimates of atrophy are susceptible to bias [17]. We follow [8] by fitting a straight line (linear regression) through all such atrophy measurements over time in each diagnostic category. The intercept term is an estimate of the atrophy one would measure when registering two scans acquired on the same day; hence it should be near zero and its 95% confidence interval should contain zero. Tab. 2 shows the slopes, intercepts, and 95% confidence intervals for all five groups of ADNI-1. Results (not shown) are similar for ADNI-2. All of them have intercepts that are near zero relative to the range of changes observed and all prediction intercept confidence intervals contain zero. Further, all slopes are positive indicating average volume loss over time, consistent with expectations for an aging and neuro-degenerative population. The slopes capture increasing atrophy with disease severity. We conclude that neither LDDMM optimization nor FPSGR produced deformations with significant bias to over- or underestimate volume change. Note that our LDDMM optimization results and the prediction results show the same trends, and are directly comparable as the results are based on the same test images (same for atrophy measurement).

Atrophy. Atrophy estimates have also been shown to correlate with clinical variables. In accordance with validation (2) from above, we computed the correlation between our atrophy estimates and diagnosis, and also between our atrophy estimates and mini-mental state exam (MMSE) scores (Fig. 4). The magnitudes of correlations we observed for all eight prediction models (four with and four without correction networks) were in the range of -0.4 $\sim$ -0.8 for MMSE and 0.2 $\sim$ 0.6 for diagnosis. Previous studies have reported correlation between atrophy estimates and clinical variables as high as -0.7 for MMSE and 0.5 for diagnosis for 100 subjects [4,3]. All prediction results are comparable with baseline LDDMM results. The correction network generally improves prediction performance.

Jacobian. The average JD images qualitatively agree with prior results [8,7]: severity of volume change increases with severity of diagnosis and time. Change is most substantial in the temporal lobes near the hippocampus (see Fig. 3).

5 Conclusion and Future Work

We proposed a fast approach for geodesic regression (FPSGR) to study longitudinal image data. FPSGR is orders of magnitude faster than existing approaches and facilitates the analysis of large-scale imaging studies. Experiments on the

ADNI-1/2 datasets demonstrate that FPSGR captures expected atrophy trends, exhibits negligible bias and shows high correlations with clinical variables.
Support. This work was supported by NSF grants EECS-1148870/1711776.

References

1. Biobank website: www.ukbiobank.ac.uk 1
2. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: ICCV. pp. 2758–2766 (2015) 2
3. Fleishman, G., Thompson, P.M.: Adaptive gradient descent optimization of initial momenta for geodesic shooting in diffeomorphisms. In: ISBI (2017) 6, 7
4. Fleishman, G., Thompson, P.M.: The impact of matching functional on atrophy measurement from geodesic shooting in diffeomorphisms. In: ISBI (2017) 6, 7
5. Fletcher, P.T.: Geodesic regression and the theory of least squares on Riemannian manifolds. IJCV 105(2), 171–185 (2013) 1
6. Hong, Y., Shi, Y., Styner, M., Sanchez, M., Niethammer, M.: Simple geodesic regression for image time-series. In: WBIR. pp. 11–20 (2012) 1
7. Hua, X., Ching, C.R.K., Mezher, A., Gutman, B., Hibar, D.P., Bhatt, P., Leow, A.D., Jr., C.R.J., Bernstein, M., Weiner, M.W., Thompson, P.M.: MRI-based brain atrophy rates in ADNI phase 2: acceleration and enrichment considerations for clinical trials. Neurobiology of Aging 37, 26–37 (2016) 6, 7
8. Hua, X., Hibar, D.P., Ching, C.R.K., Boyle, C.P., Rajagopalan, P., Gutman, B., Leow, A.D., Toga, A.W., Jr., C.R.J., Harvey, D.J., Weiner, M.W., Thompson, P.M.: Unbiased tensor-based morphometry:improved robustness & sample size estimates for Alzheimer's disease clinical trials. NeuroImage 66, 648–661 (2013) 6, 7
9. Ikram, M.A., van der Lugt, A., Niessen, W.J., Koudstaal, P.J., Krestin, G.P., Hofman, A., Bos, D., Vernooij, M.W.: The Rotterdam scan study: design update 2016 and main findings. European journal of epidemiology 30(12), 1299–1315 (2015) 1
10. Jack, C.R., Barnes, J., Bernstein, M.A., Borowski, B.J., Brewer, J., Clegg, S., Dale, A.M., Carmichael, O., Ching, C., DeCarli, C., et al.: Magnetic resonance imaging in ADNI 2. Alzheimer's & Dementia 11(7), 740–756 (2015) 1
11. Liu, Z., Yeh, R., Tang, X., Liu, Y., Agarwala, A.: Video frame synthesis using deep voxel flow. arXiv preprint arXiv:1702.02463 (2017) 2
12. Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. In: MICCAI. pp. 655–662 (2011) 1
13. Schuster, T., Wolf, L., Gadot, D.: Optical flow requires multiple strategies (but only one network). arXiv preprint arXiv:1611.05607 (2016) 2
14. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In: ISBI. pp. 1219–1222 (2013) 1, 2
15. Yang, X., Kwitt, R., Niethammer, M.: Fast predictive image registration. In: MICCAI DLMIA Workshop. pp. 48–57. Springer (2016) 2, 3
16. Yang, X., Kwitt, R., Niethammer, M.: Quicksilver: Fast predictive image registration-a deep learning approach. NeuroImage (2017), in press 2, 3, 4
17. Yushkevich, P.A., Avants, B.B., Das, S.R., Pluta, J., Altinay, M., Craige, C.: Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3T MRI data. NeuroImage 50(2), 434 – 445 (2010) 7
18. Zhang, M., Fletcher, P.T.: Finite-dimensional Lie algebras for fast diffeomorphic image registration. In: IPMI. pp. 249–260 (2015) 2