Improved Probabilistic Diffeomorphic Registration with CNNs

Adrian V. Dalca* CSAIL, MIT MGH, HMS
Guha Balakrishnan CSAIL, MIT
John Guttag CSAIL, MIT
Mert R. Sabuncu ECE and BME, Cornell

Abstract

Classical image registration methods have seen significant technical development, but often require significant runtime. In contrast, learning-based registration methods achieve impressive test accuracy and runtime, but often omit classical development. In recent months, preliminary methods have joined these paradigms, providing fast runtime while maintaining classical guarantees. In this work, we extend these methods by improving theoretical approximations, introducing a bidirectional loss to enable accurate inverse deformations, and analyzing algorithm behaviour in the presence of limited training data or given training segmentation maps. We demonstrate improved behaviour and accuracy in a range of experiments.

1 Introduction

Deformable registration computes a dense correspondence between two images, and is fundamental to many medical image analysis tasks. Classical methods such as elastic-type models [3], B splines [27], dense vector fields [29], or discrete methods [9, 14] yield deformations by solving an optimization problem. Diffeomorphic transforms, which ensure properties like topology preservations, have seen extensive methodological development [1, 2, 6, 31]. However, since these methods solve a complex optimization problem, they often require substantial runtime for a given image pair.

Recent learning-based methods propose to train neural networks that map image pairs to deformations. Supervised approaches require ground truth registration fields, often derived via more conventional registration tools [26, 28, 30]. Unsupervised or self-supervised methods [4, 5, 10, 20] use a spatial transformer [17] to enable end-to-end training without explicit supervision. However, these methods tend to lack the theoretical guarantees and properties provided by classical methods.

Methods introduced in the past six months use classical concepts to derive neural-network based solutions that maintain desired properties such as diffeomorphic guarantees and uncertainty estimates [8, 19]. However, these preliminary methods have only been demonstrated in limited settings. In this work, we extend and analyze several aspects of one such method, VoxelMorph [8].

- VoxelMorph mathematical development approximates a spatially smooth deformation prior by introducing independence at every voxel. We introduce a new approximation that encourages deformations to be more consistent with a smoothness prior.
- The VoxelMorph loss function encourages an accurate forward deformation. We introduce a bidirectional loss function that explicitly encourages accurate forward *and* backwards warps, resulting in improved performance in subject-to-subject registration.
- VoxelMorph was demonstrated in the context of a large training dataset. Here, we perform an analysis of test performance based on different numbers of training subjects.

*adalca@mit.edu

Medical Imaging meets NeurIPS 2018, Montréal, Canada.

- VoxelMorph uses only MRI images during training. Here, we show that adding segmentation maps during training significantly improves test-time accuracy.

2 Methods

We first briefly review diffeomorphic VoxelMorph [8], then describe our proposed extensions.

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be the images to be registered, and let $\boldsymbol{z}$ be a latent variable that parametrizes a transformation function $\boldsymbol{\phi_z}$. In VoxelMorph, $\boldsymbol{z}$ is a stationary velocity field, which is integrated to a diffeomorphic deformation $\boldsymbol{\phi_z}$. The multivariate normal prior $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{\Lambda}_z^{-1})$ encourages spatial smoothness by letting $\boldsymbol{\Lambda}_z = \lambda \boldsymbol{L}$,where λ denotes a parameter controlling the scale of the velocity field $\boldsymbol{z}$, and $\boldsymbol{L}$ is the Laplacian of a neighborhood graph defined on a voxel grid. The model further assumes that $\boldsymbol{x}$ is a noisy observation of warped image $\boldsymbol{y}$: $p(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{y} \circ \boldsymbol{\phi_z}, \sigma^2 \mathbb{I})$, where σ^2 reflects the variance of additive image noise.

The goal is to find the posterior probability of registration $p(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$, giving the most likely registration field $\boldsymbol{\phi_z}$ as well as an estimate of uncertainty for each voxel. However, since $p(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$ is intractable, the authors use a variational approximation modeled as a multivariate normal: $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_{z|x,y}, \boldsymbol{\Sigma}_{z|x,y})$, where the functions $\boldsymbol{\mu}_{z|x,y}$ and $\boldsymbol{\Sigma}_{z|x,y}$ are estimated using a convolutional neural network. By minimizing the distance $\min_\psi \mathrm{KL}\left[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})||p(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})\right]$, the authors arrive at the network loss function: $\mathcal{L}(\boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2\sigma^2 K} \sum_k ||\boldsymbol{x} - \boldsymbol{y} \circ \boldsymbol{\phi}_{z_k}||^2 + \mathrm{KL}\left[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})||p(\boldsymbol{z})\right]$, where $\boldsymbol{z}_k$ is a sample from $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$. The first term encourages image matching, and the second encourages the velocity field to be smooth as specified by the prior. Registration for a new image pair $\{\boldsymbol{x}, \boldsymbol{y}\}$ is computed by evaluating the neural network, sampling from $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$, and computing $\boldsymbol{\phi}_{z_k}$.

2.1 Proposed Approximate Posterior

In VoxelMorph, the covariance $\boldsymbol{\Sigma}_{z|x,y}$ is diagonal, leading to independent noise in $\boldsymbol{z}$ at each voxel. This leads to smoothness only being encouraged through $\boldsymbol{\mu}_{z|x,y}$, and stationary velocity fields $\boldsymbol{z}_k$ sampled from $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$ can be noisy. We introduce a new approximation to encourage smoothness:

$$\widetilde{\boldsymbol{\Sigma}} z|x, y = \boldsymbol{C}_{\sigma_c} \boldsymbol{D} \boldsymbol{D}^T \boldsymbol{C}_{\sigma_c}^T, \tag{1}$$

where $\boldsymbol{D}$ is a diagonal matrix outputted by the neural network and $\boldsymbol{C}_{\sigma_c}$ is a fixed smoothing convolution matrix such that $\boldsymbol{C}_{\sigma_c} \boldsymbol{w}$ is equivalent to 3D convolution of $\boldsymbol{w}$ by a gaussian filter with variance σ_c^2. We choose σ_c such that the smoothing operation matches the scale of the prior determined by λ: $\frac{1}{\sqrt{2\pi\sigma^{3/2}}} = (\lambda * 6)^{-1}$. Sampling from $\widetilde{q}_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_{z|x,y}, \widetilde{\boldsymbol{\Sigma}}_{z|x,y})$ is then achieved using the *reparametrization trick* [18]: $\boldsymbol{z}_k = \boldsymbol{\mu}_{z|x,y} + \boldsymbol{C}_{\sigma_c} \boldsymbol{D} \boldsymbol{r}$, where $\boldsymbol{r}$ is a sample from the standard normal. Intuitively, compared to the original approximation, this sampling procedures smoothes the term added to the mean.

2.2 Bidirectional cost function

The VoxelMorph loss encourages accurate deformation $\boldsymbol{\phi}$ through the image matching term, but does not constrain the inverse deformation $\boldsymbol{\phi}^{-1}$. We propose to encourage accuracy for both deformations:

$$\mathcal{L}(\boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{4\sigma^2 K} \sum_k \left(||\boldsymbol{x} - \boldsymbol{y} \circ \boldsymbol{\phi}_{z_k}||^2 + ||\boldsymbol{y} - \boldsymbol{x} \circ \boldsymbol{\phi}_{z_k}^{-1}||^2 \right) + \mathrm{KL}\left[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})||p(\boldsymbol{z})\right]. \tag{2}$$

Because of the stationary velocity field representation, computing the inverse deformation field $\boldsymbol{\phi}_{\boldsymbol{z}}^{-1}$ can be achieved by taking the negative of the velocity field: $\boldsymbol{\phi}_{\boldsymbol{z}}^{-1} = \boldsymbol{\phi}_{-\boldsymbol{z}}$ [1, 24], enabling the computation of both fields inside one network.

2.3 Training with Limited Data

Learning-based methods like VoxelMorph replace classical pair-specific optimization over deformation fields with a global optimization of network parameters. The generalizability of the deformations depends on the amount and variability of training data. In the original work, VoxelMorph was trained on thousands of scans. In this work, we train the models on various numbers of scans and

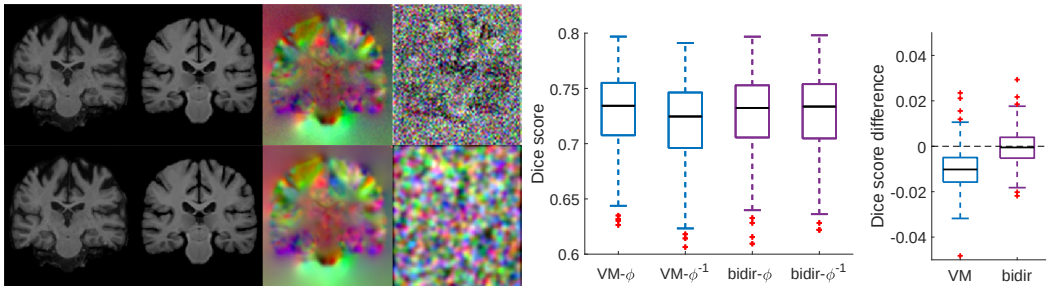

Figure 1: Left: Example image pair, sample velocity field, and standard deviation across several samples for original VoxelMorph (top) and our smoother distribution (bottom). Right: VoxelMorph (blue) results in less accurate *inverse* warp ϕ^{-1} compared to deformations ϕ. The proposed bi-directional loss (purple) leads to both deformations performing equally well, and on par with the VoxelMorph forward deformations.

evaluate the accuracy of the resulting models at test time. Furthermore, the resulting deformation can be interpreted as simply an approximation or initialization to the optimal deformation, and we experiment with improving it using instance-specific optimization.

2.4 Training with additional information: Segmentation Maps

VoxelMorph was trained using MRI images. Here, we leverage segmentation maps during training to improve test-time accuracy. Anatomical label maps are sometimes available during training. These are produced by human experts or automated pipelines. Letting $\mathbf{s}_x^l, \mathbf{s}_y^l \circ \boldsymbol{\phi}$ be the set of voxels of structure l for $\mathbf{x}$ and $\mathbf{y} \circ \boldsymbol{\phi}$, respectively, we quantify the volume overlap for structure l using the Dice score [12]. We expand the VoxelMorph loss by adding a segmentation term $\alpha \mathcal{L}_{seg}$:

$$\mathcal{L}_{seg}(\mathbf{s}_x, \mathbf{s}_y \circ \boldsymbol{\phi}) = -\sum_l \text{Dice}(\mathbf{s}_x^l, \mathbf{s}_y^l \circ \boldsymbol{\phi}) = -\sum_l 2 \cdot \frac{|\mathbf{s}_x^l \cap (\mathbf{s}_y^l \circ \boldsymbol{\phi})|}{|\mathbf{s}_x^l| + |\mathbf{s}_y^l \circ \boldsymbol{\phi}|}. \tag{3}$$

Segmentation maps are not not used to register a test image pair.

3 Experiments

For most experiments below, we use the data used in the original VoxelMorph paper, and focus on 3D atlas-based registration. Specifically, we register each scan to an atlas computed using external data [13]. We use a large-scale, multi-site dataset of 7829 T1-weighted brain MRI scans from eight publicly available datasets: ADNI [25], OASIS [21], ABIDE [11], ADHD200 [23], MCIC [15], PPMI [22], HABS [7], and Harvard GSP [16]. We performed standard pre-processing, including

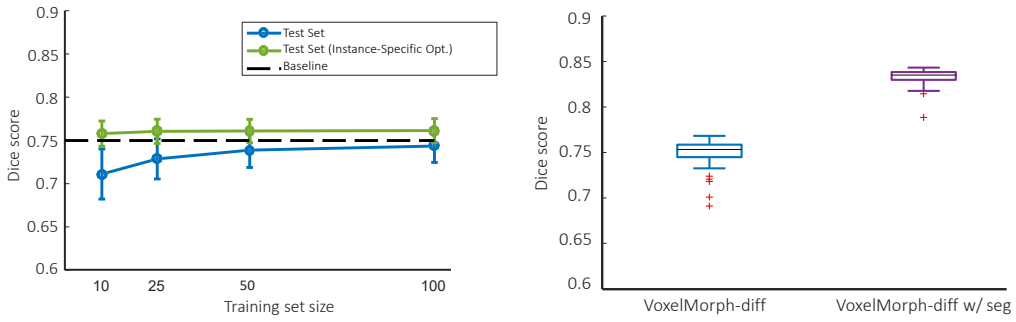

Figure 2: Left: effect of training set size on accuracy (blue line), and instance-specific deformation optimization (green line) after these are initialized to the VoxelMorph results (blue line). Right: significant improvement in test performance when segmentation is used during training.

resampling to 1mm isotropic voxels, affine spatial normalization, brain extraction, and segmentation of 29 anatomical labels using FreeSurfer [13], and crop the final images to $160 \times 192 \times 224$. We split the dataset into 7329, 250, and 250 volumes for train, validation, and test sets respectively.

To evaluate registration quality, we register each subject to an atlas, propagate the segmentation map using the resulting warp, measure volume overlap using the Dice metric for each label, and report the average Dice metric across labels.

Proposed Approximate Posterior We train a network using the proposed smoothing approximate posterior (1), as well as the original VoxelMorph approximation. We find that the two methods result in similar performance for the test set: (0.752 ± 0.021). However, we also evaluate the diffeomorphic property of the final deformations $\boldsymbol{\phi}_{\boldsymbol{z}}$ by counting the number of voxels p such that $|J_\phi(p)| \leq 0$, where J is the Jacobian matrix, and find that VoxelMorph samples average 2.6 non-diffeomorphic voxels, while our proposed method yields only 1.0. Figure 1 Left shows that the new posterior yields significantly smoother stationary velocity fields, consistent with the model prior.

Bidirectional cost function We train a registration network using the proposed bi-directional cost function, as well as the original VoxelMorph network, for both atlas-based registration, as well as subject-to-subject registration. For atlas-based registration, both methods perform similarly. However, Figure 1 Right shows that for subject to subject registration, the VoxelMorph loss leads to the inverse deformations performing lower (0.716 ± 0.042) than the forward deformations (0.726 ± 0.043), while the new bidirectional cost function leads to both deformations performing equally well $(0.725 \pm 0.042$ and 0.725 ± 0.041 Dice scores), on par with the VoxelMorph forward deformations.

Training Set Size and Instance-Specific Optimization We train VoxelMorph on subsets of different sizes from our training dataset. We evaluate Dice score on the test set as well as the test set when each deformation is further individually optimized for each test image pair using gradient descent on each stationary velocity field $\boldsymbol{z}$ (instance-specific optimization). Figure 2 Left shows that a small training set size of 10 scans results in slightly lower test Dice scores compared to larger training set sizes. There is no significant difference when training with 100 scans compared to the full dataset of over 7000 scans. Further optimizing the deformation on each test image pair results in better test Dice scores regardless of training set size.

Training with Segmentation Maps We train a VoxelMorph network using the proposed segmentation loss along with the original VoxelMorph scan. Figure 2 Right demonstrates that using segmentations at training time yields a significant registration improvement at test time.

Conclusion We introduce several extensions and analyses for VoxelMorph. We demonstrate that the proposed approximating posterior results in smoother velocity fields, and the bi-directional cost function improves accuracy of the *inverse* deformation. We show that VoxelMorph can be used even with limited training data, and using segmentations at training can dramatically improve performance.

References

[1] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007.

[2] B.B. Avants et al. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.

[3] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46:1–21, 1989.

[4] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. An unsupervised learning model for deformable medical image registration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9252–9260, 2018.

[5] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: A learning framework for deformable medical image registration. In *arXiv preprint arXiv:1809.05231*, 2018.

[6] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision*, 61:139–157, 2005.

[7] Alexander Dagley, Molly LaPoint, Willem Huijbers, Trey Hedden, Donald G McLaren, Jasmeer P Chatwal, Kathryn V Papp, Rebecca E Amariglio, Deborah Blacker, Dorene M Rentz, et al. Harvard aging brain study: dataset and accessibility. *NeuroImage*, 2015.

[8] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. *arXiv preprint arXiv:1805.04605*, 2018.

[9] Adrian V Dalca, Andreea Bobu, Natalia S Rost, and Polina Golland. Patch-based discrete registration of clinical brain images. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 60–67. Springer, 2016.

[10] Bob D de Vos, Floris F Berendsen, Max A Viergever, Marius Staring, and Ivana Išgum. 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*, pages 204–212. 2017.

[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):659–667, 2014.

[12] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

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

[14] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab, and Nikos Paragios. Dense image registration through mrfs and efficient linear programming. *Medical image analysis*, 12(6):731–741, 2008.

[15] Randy L Gollub, Jody M Shoemaker, Margaret D King, Tonya White, Stefan Ehrlich, Scott R Sponheim, Vincent P Clark, Jessica A Turner, Bryon A Mueller, Vince Magnotta, et al. The mcic collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics*, 11(3):367–388, 2013.

[16] Avram J Holmes, Marisa O Hollinshead, Timothy M O'Keefe, Victor I Petrov, Gabriele R Fariello, Lawrence L Wald, Bruce Fischl, Bruce R Rosen, Ross W Mair, Joshua L Roffman, et al. Brain genomics superstruct project initial data release with structural, functional, and behavioral measures. *Scientific data*, 2, 2015.

[17] Max Jaderberg, Karen Simonyan, and Andrew Zisserman. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.

[18] Diederik P Kingma and Jimmy Ba. ADAM: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[19] Julian Krebs, Tommaso Mansi, Boris Mailhé, Nicholas Ayache, and Hervé Delingette. Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 101–109. Springer, 2018.

[20] Hongming Li and Yong Fan. Non-rigid image registration using fully convolutional networks with deep self-supervision. *preprint arXiv:1709.00799*, 2017.

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

[22] Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.

[23] Michael P Milham, Damien Fair, Maarten Mennes, Stewart HMD Mostofsky, et al. The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, 6:62, 2012.

[24] M. Modat, I.J.A. Simpson, M.J. Cardoso, D.M. Cash, N. Toussaint, N.C. Fox, and S. Ourselin. Simulating neurodegeneration through longitudinal population analysis of structural and diffusion weighted mri data. *Medical Image Computing and Computer-Assisted Intervention*, LNCS 8675:57–64, 2014.

[25] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford R Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni). *Alzheimer's & Dementia*, 1(1):55–66, 2005.

[26] 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 (MICCAI)*, pages 266–274. Springer, 2017.

[27] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformation: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[28] Hessam Sokooti, Bob de Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, and Marius Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 232–239. Springer, 2017.

[29] J.P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

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

[31] 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.