Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration

Adrian V. Dalca[1,2,3], Guha Balakrishnan[1], John Guttag[1], and Mert R. Sabuncu[3]

[1] Computer Science and Artificial Intelligence Lab, MIT
[2] Martinos Center for Biomedical Imaging, Massachusetts General Hospital, HMS
[3] School of Electrical and Computer Engineering, Cornell University

Abstract. Traditional deformable registration techniques achieve impressive results and offer a rigorous theoretical treatment, but are computationally intensive since they solve an optimization problem for each image pair. Recently, learning-based methods have facilitated fast registration by learning spatial deformation functions. However, these approaches use restricted deformation models, require supervised labels, or do not guarantee a diffeomorphic (topology-preserving) registration. Furthermore, learning-based registration tools have not been derived from a probabilistic framework that can offer uncertainty estimates. In this paper, we present a probabilistic generative model and derive an unsupervised learning-based inference algorithm that makes use of recent developments in convolutional neural networks (CNNs). We demonstrate our method on a 3D brain registration task, and provide an empirical analysis of the algorithm. Our approach results in state of the art accuracy and very fast runtimes, while providing diffeomorphic guarantees and uncertainty estimates. Our implementation is available online at `https://github.com/voxelmorph/voxelmorph`.

Keywords: medical image registration, diffeomorphic registration, probabilistic modeling, convolutional neural networks, variational inference, uncertainty estimation

1 Introduction

Deformable registration computes a dense correspondence between two images, and is fundamental to many medical image analysis tasks. Traditional methods solve an optimization over the space of deformations, such as elastic-type models [5,29], B splines [28], dense vector fields [31], or discrete methods [10,14]. Constraining the allowable transformations to diffeomorphisms ensures certain desirable properties, such as preservation of topology. Diffeomorphic transforms have seen extensive methodological development, yielding state-of-the-art tools, such as LDDMM [7,33], DARTEL [3], and SyN [4]. Unfortunately, these methods often demand substantial time and computational resources to run for a given image pair.

Recent methods have proposed to train neural networks that map a pair of input images to an output deformation. These approaches usually require ground truth registration fields, often derived via more conventional registration tools,

which can introduce a bias and necessitate significant preprocessing [26,30,32]. Some preliminary papers [11,21] explore unsupervised strategies that build on the spatial transformer network [17], but are only demonstrated with constrained deformation models such as affine or small displacement fields. Furthermore, they have only been validated on limited volumes, such as 3D patches or 2D slices. A recent paper avoids these pitfalls, but still does not provide topology-preserving guarantees or probabilistic uncertainty estimates, which yield meaningful information for downstream image analysis [6].

In this paper we present a formulation for registration as conducting variational inference on a probabilistic generative model. This framework naturally results in a learning algorithm that uses a convolutional neural network with an intuitive cost function. We introduce novel *diffeomorphic integration* layers combined with a transform layer to enable unsupervised end-to-end learning for diffeomorphic registration. We present extensive experiments, demonstrating that our algorithm achieves state of the art registration accuracy while providing diffeomorphic deformations, fast runtime and estimates of registration uncertainty.

1.1 Diffeomorphic Registration

Although the method presented in this paper applies to a multitude of deformable representations, we choose to work with diffeomorphisms, and in particular with a stationary velocity field representation [3]. Diffeomorphic deformations are differentiable and invertible, and thus preserve topology. Let $\phi : R^3 \to R^3$ represent the deformation that maps the coordinates from one image to coordinates in another image. In our implementation, the deformation field is defined through the following ordinary differential equation (ODE):

$$\frac{\partial \phi^{(t)}}{\partial t} = v(\phi^{(t)}) \tag{1}$$

where $\phi^{(0)} = Id$ is the identity transformation and t is time. We integrate the stationary velocity field v over $t = [0, 1]$ to obtain the final registration field $\phi^{(1)}$.

We compute the integration numerically using *scaling and squaring* [2], which we briefly review here. The integration of a stationary ODE represents a one-parameter subgroup of diffeomorphisms. In group theory, v is a member of the Lie algebra and is exponentiated to produce $\phi^{(1)}$, which is a member of the Lie group: $\phi^{(1)} = \exp(v)$. From the properties of one-parameter subgroups, for any scalars t and t', $\exp((t + t')v) = \exp(tv) \circ \exp(t'v)$, where $\circ$ is a composition map associated with the Lie group. Starting from $\phi^{(1/2^T)} = p + v(p)$ where p is a map of spatial locations, we use the recurrence $\phi^{(1/2^{t-1})} = \phi^{(1/2^t)} \circ \phi^{(1/2^t)}$ to obtain $\phi^{(1)} = \phi^{(1/2)} \circ \phi^{(1/2)}$. T is chosen so that $v \approx 0$.

2 Methods

Let x and y be 3D images, such as MRI volumes, and let z be a latent variable that parametrizes a transformation function $\phi_z : R^3 \to R^3$. We use a generative model

to describe the formation of $\boldsymbol{x}$ by warping $\boldsymbol{y}$ into $\boldsymbol{y} \circ \boldsymbol{\phi_z}$. We propose a variational inference method that uses a neural network of convolutions, diffeomorphic integration, and spatial transform layers. We learn the network parameters in an unsupervised fashion, i.e., without access to ground truth registrations. We describe how the network yields fast diffeomorphic registration of a new image pair $\boldsymbol{x}$ and $\boldsymbol{y}$, while providing uncertainty estimates.

2.1 Generative Model

We model the prior probability of $\boldsymbol{z}$ as:

$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{\Sigma}_z), \tag{2}$$

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Our work applies to a wide range of representations $\boldsymbol{z}$. For example, $\boldsymbol{z}$ could be a low-dimensional embedding of a dense displacement field, or even the displacement field itself. In this paper, we let $\boldsymbol{z}$ be a stationary velocity field that specifies a diffeomorphism through the ODE (1). We let $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ be the Laplacian of a neighborhood graph defined on a voxel grid, where $\boldsymbol{D}$ is the graph degree matrix, and $\boldsymbol{A}$ is a voxel neighbourhood adjacency matrix. We encourage spatial smoothness of $\boldsymbol{z}$ by letting $\boldsymbol{\Sigma}_z^{-1} = \boldsymbol{\Lambda}_z = \lambda \boldsymbol{L}$, where $\boldsymbol{\Lambda}_z$ is a precision matrix and λ denotes a parameter controlling the scale of the velocity field $\boldsymbol{z}$.

We let $\boldsymbol{x}$ be a noisy observation of warped image $\boldsymbol{y}$:

$$p(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{y} \circ \boldsymbol{\phi_z}, \sigma^2 \mathbb{I}), \tag{3}$$

where σ^2 reflects the variance of additive image noise.

We aim to estimate the posterior registration probability $p(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$. Using this, we can obtain the most likely registration field $\boldsymbol{\phi_z}$ for a new image pair $(\boldsymbol{x}, \boldsymbol{y})$ via MAP estimation, along with an estimate of uncertainty for this registration.

2.2 Learning

With our assumptions, computing the posterior probability $p(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$ is intractable. We use a variational approach, and introduce an approximate posterior probability $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$ parametrized by $\boldsymbol{\psi}$. We minimize the KL divergence

$$\min_{\psi} \mathrm{KL}\left[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y}) \| p(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})\right]$$

$$= \min_{\psi} \mathbf{E}_q\left[\log q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y}) - \log p(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})\right]$$

$$= \min_{\psi} \mathbf{E}_q\left[\log q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y}) - \log p(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{y})\right] + \log p(\boldsymbol{x}; \boldsymbol{y})$$

$$= \min_{\psi} \mathrm{KL}\left[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y}) \| p(\boldsymbol{z})\right] - \mathbf{E}_q\left[\log p(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{y})\right], \tag{4}$$

which is the negative of the *variational lower bound* of the model evidence [19]. We model the approximate posterior $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$ as a multivariate normal:

$$q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_{z|x,y}, \boldsymbol{\Sigma}_{z|x,y}), \tag{5}$$

where $\boldsymbol{\Sigma}_{z|x,y}$ is diagonal.
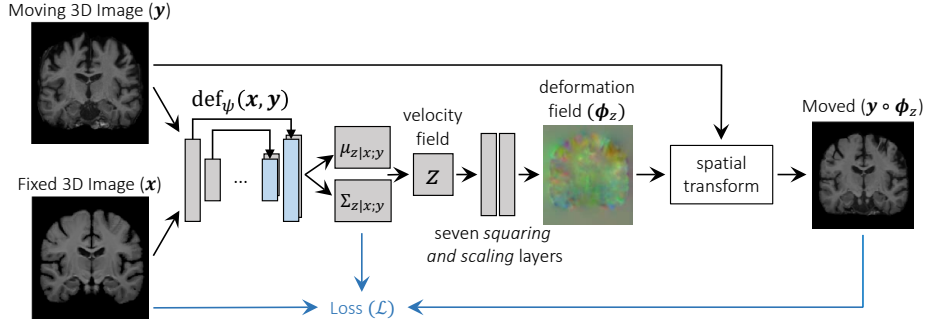
Fig. 1. Overview of end-to-end unsupervised architecture. The first part of the network, $\mathrm{def}_\psi(\boldsymbol{x}, \boldsymbol{y})$ takes the input images and outputs the approximate posterior probability parameters representing the velocity field mean, $\boldsymbol{\mu}_{z|x;y}$, and variance, $\boldsymbol{\Sigma}_{z|x;y}$. A velocity field $\boldsymbol{z}$ is sampled and transformed to a diffeomorphic deformation field $\boldsymbol{\phi}_z$ using novel differentiable *squaring and scaling* layers. Finally, a spatial transform warps $\boldsymbol{y}$ to obtain $\boldsymbol{y} \circ \boldsymbol{\phi}_z$.

We estimate $\boldsymbol{\mu}_{z|x,y}$, and $\boldsymbol{\Sigma}_{z|x,y}$ using a convolutional neural network $\mathrm{def}_\psi(\boldsymbol{x}, \boldsymbol{y})$ parameterized by $\boldsymbol{\psi}$, as described below. We can therefore learn the parameters $\boldsymbol{\psi}$ by optimizing the variational lower bound (4) using stochastic gradient methods. Specifically, for each image pair $\{\boldsymbol{x}, \boldsymbol{y}\}$ and samples $\boldsymbol{z}_k \sim q_\psi(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$, we can compute $\boldsymbol{y} \circ \boldsymbol{\phi}_{z_k}$, with the resulting loss:

$$
\mathcal{L}(\boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{y})
$$
$$
= -\mathbf{E}_q\left[\log p(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{y})\right] + \mathrm{KL}\left[q_\psi(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})||p(\boldsymbol{z})\right] \tag{6}
$$
$$
= \frac{1}{2\sigma^2 K} \sum_k ||\boldsymbol{x} - \boldsymbol{y} \circ \boldsymbol{\phi}_{z_k}||^2 + \frac{1}{2}\left[\mathrm{tr}(\lambda \boldsymbol{D}\boldsymbol{\Sigma}_{z|x;y} - \log|\boldsymbol{\Sigma}_{z|x;y}|) + \boldsymbol{\mu}_{z|x,y}^T \boldsymbol{\Lambda}_z \boldsymbol{\mu}_{z|x,y}\right] + \mathrm{const},
$$

where K is the number of samples used. In our experiments, we use $K = 1$. The first term encourages the warped image $\boldsymbol{y} \circ \boldsymbol{\phi}_{z_k}$ to be similar to $\boldsymbol{x}$. The second term encourages the posterior to be close to the prior $p(\boldsymbol{z})$. Although the variational covariance $\boldsymbol{\Sigma}_{z|x,y}$ is diagonal, the last term spatially smoothes the mean: $\boldsymbol{\mu}_{z|x,y}^T \boldsymbol{\Lambda}_z \boldsymbol{\mu}_{z|x,y} = \frac{\lambda}{2} \sum \sum_{j \in N(I)} (\boldsymbol{\mu}[i] - \boldsymbol{\mu}[j])^2$, where $N(i)$ are the neighbors of voxel i. We treat σ^2 and λ as fixed hyper-parameters.

2.3 Neural Network Framework

We design the network $\mathrm{def}_\psi(\boldsymbol{x}, \boldsymbol{y})$ that takes as input $\boldsymbol{x}$ and $\boldsymbol{y}$ and outputs $\boldsymbol{\mu}_{z|x,y}$ and $\boldsymbol{\Sigma}_{z|x,y}$, based on a 3D UNet-style architecture [27]. The network includes a convolutional layer with 16 filters, four downsampling layers with 32 convolutional filters and a stride of two, and finally three upsampling convolutional layers with 32 filters. All convolutional layers use LeakyReLu activations and a 3x3 kernel.

To enable unsupervised learning of parameters $\boldsymbol{\psi}$ using (6), we must form $\boldsymbol{y} \circ \boldsymbol{\phi}_z$ to compute the data term. We first implement a layer that samples a new $\boldsymbol{z}_k \sim \mathcal{N}(\boldsymbol{\mu}_{z|x,y}, \boldsymbol{\Sigma}_{z|x,y})$ using the "re-parameterization trick" [19].

We propose novel *scaling and squaring* network layers to compute $\boldsymbol{\phi}_{z_k} = \exp(\boldsymbol{z}_k)$. Specifically, these involve compositions within the neural network architecture using a differentiable spatial transformation operation. Given two 3D vector fields $\boldsymbol{a}$ and $\boldsymbol{b}$, for each voxel p such a layer computes $(\boldsymbol{a} \circ \boldsymbol{b})(p) = \boldsymbol{a}(\boldsymbol{b}(p))$, a non-integer voxel location $\boldsymbol{b}(p)$ in $\boldsymbol{a}$, using linear interpolation. Starting with $\boldsymbol{v}^{(1/2^T)} = \boldsymbol{z}_k$, we compute $\boldsymbol{v}^{(1/2^{t+1})} = \boldsymbol{v}^{(1/2^t)} \circ \boldsymbol{v}^{(1/2^t)}$ recursively using these layers, leading to $\boldsymbol{v}^{(1)} \triangleq \boldsymbol{\phi}_{z_k} = \exp(\boldsymbol{z}_k)$. In our experiments, we use $T = 7$.

Finally, we use a spatial transform layer to warp volume $\boldsymbol{y}$ according to the computed diffeomorphic field $\boldsymbol{\phi}_{z_k}$. This network results in three outputs, $\boldsymbol{\mu}_{z|x,y}, \boldsymbol{\Sigma}_{z|x,y}$ and $\boldsymbol{y} \circ \boldsymbol{\phi}_{z_k}$, which are used in the model loss (6).

In summary, the neural network takes as input $\boldsymbol{x}$ and $\boldsymbol{y}$, computes $\boldsymbol{\mu}_{z|x,y}$ and $\boldsymbol{\Sigma}_{z|x,y}$, samples a new $\boldsymbol{z}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, computes a diffeomorphic $\boldsymbol{\phi}_{z_k}$ and applies it to $\boldsymbol{y}$. Since all the steps are designed to be differentiable, we learn the network parameters using stochastic gradient descent based methods on the loss (6). The framework is summarized in Figure 1.

Our implementation uses Keras [8] with a Tensorflow backend [1] and the ADAM optimizer [18]. We implement our method as part of the VoxelMorph package, with both implementations available online at `https://github.com/voxelmorph/voxelmorph`.

2.4 Registration and Uncertainty

Given learned parameters, we approximate registration of a new scan pair $(\boldsymbol{x}, \boldsymbol{y})$ using $\boldsymbol{\phi}_{\hat{z}_k}$. We first obtain $\hat{\boldsymbol{z}}_k$ using

$$\hat{\boldsymbol{z}}_k = \arg \max_{\boldsymbol{z}_k} p(\boldsymbol{z}_k | \boldsymbol{x}; \boldsymbol{y}) = \boldsymbol{\mu}_{z|x;y}, \tag{7}$$

by evaluating the neural network $\text{def}_{\psi}(\boldsymbol{x}, \boldsymbol{y})$ on the two input images. We then compute $\boldsymbol{\phi}_{\hat{z}_k}$ using the *scaling and squaring* network layers. We also obtain $\boldsymbol{\Sigma}_{z|x,y}$, enabling an estimation of the uncertainty of the velocity field $\boldsymbol{z}$ at each voxel j:

$$H(\boldsymbol{z}[j]) \approx \mathbf{E}\left[-\log q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})\right] = \frac{1}{2} \log 2\pi \boldsymbol{\Sigma}_{z|x;y}[j, j]. \tag{8}$$

We also estimate uncertainty in the deformation field $\boldsymbol{\phi}_z$ empirically. We sample several representations $\boldsymbol{z}_{k'} \sim q_{\psi}(\boldsymbol{z}|\boldsymbol{x}; \boldsymbol{y})$, propagate them through the diffeomorphic layers to compute $\boldsymbol{\phi}_{z_k'}$, and compute the empirical diagonal covariance $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\phi}_z}[j, j]$ across samples. The uncertainty is then $H(\boldsymbol{\phi}[j]) \approx \frac{1}{2} \log 2\pi \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\phi}_z}[j, j]$.

3 Experiments

We focus on 3D atlas-based registration, a common task in population analysis. Specifically, we register each scan to an atlas computed using external data [13].

Data and Preprocessing. We use a large-scale, multi-site dataset of 7829 T1-weighted brain MRI scans from eight publicly available datasets: ADNI [25], OASIS [22], ABIDE [12], ADHD200 [24], MCIC [15], PPMI [23], HABS [9], and

Method	Avg. Dice	GPU sec	CPU sec	$|J_\Phi| \leq 0$	Uncertainty
Affine only	0.567 (0.157)	0	0	0	No
ANTs (SyN)	0.750 (0.135)	-	9059 (2023)	6505 (3024)	No
VoxelMorph	0.750 (0.137)	0.554 (0.017)	144 (1)	18096 (4230)	No
Ours	**0.753** (0.137)	**0.451** (0.011)	**51** (0.2)	**0.7** (2.0)	**Yes**

Table 1. Summary of results: mean Dice scores over all anatomical structures and subjects (higher is better), mean runtime; and mean number of locations with a non-positive Jacobian of each registration field (lower is better). All methods have comparable Dice scores, while our method and the original VoxelMorph are orders of magnitude faster than ANTs. Only our method achieves both high accuracy and fast runtime while also having nearly zero non-negative Jacobian locations and providing uncertainty prediction.

Harvard GSP [16]. Acquisition details, subject age ranges and health conditions are different for each dataset. We performed standard pre-processing steps on all scans, including resampling to 1mm isotropic voxels, affine spatial normalization and brain extraction for each scan using FreeSurfer [13]. We crop the final images to $160 \times 192 \times 224$. Segmentation maps including 29 anatomical structures, obtained using FreeSurfer for each scan, are used in evaluating registration results. We split the dataset into 7329, 250, and 250 volumes for train, validation, and test sets respectively, although we underscore that the training is unsupervised.

Evaluation Metric. To evaluate a registration algorithm, we register each subject to an atlas, propagate the segmentation map using the resulting warp, and measure volume overlap using the Dice metric. We also evaluate the diffeomorphic property, a focus of our model. Specifically, the Jacobian matrix $J_\phi(p) = \nabla\phi(p) \in \mathcal{R}^{3\times3}$ captures the local properties of ϕ around voxel p. The local deformation is diffeomorphic, both invertible and orientation-preserving, only at locations for which $|J_\phi(p)| > 0$ [3]. We count all other voxels, where $|J_\phi(p)| \leq 0$.
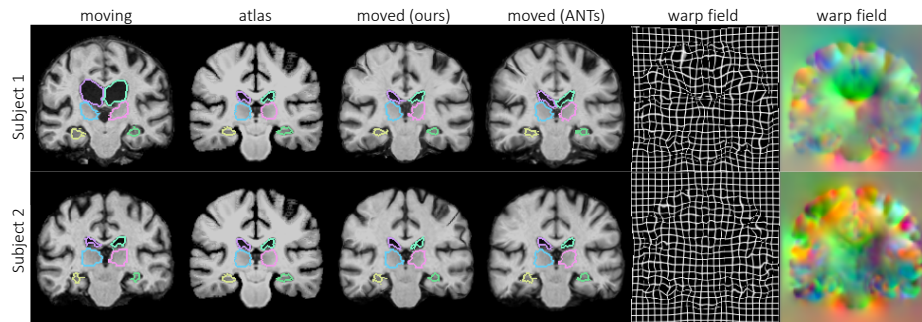

Fig. 2. Example MR slices of input moving image, atlas, and resulting warped image for our method and ANTs, with overlaid boundaries of ventricles, thalami and hippocampi. Our resulting registration field is shown as a warped grid and RGB image, with each channel representing dimension. Due to space constraints, we omit VoxelMorph examples, which are *visually* similar to our results and ANTs.

Baseline Methods. We compare our approach with the popular ANTs software package using Symmetric Normalization (SyN) [4], a top-performing algorithm [20]. We found that the default ANTs settings were sub-optimal for our task, so we performed a wide parameter and similarity metric search across a multitude of datasets. We identified top performing parameter values on the Dice metric and used cross-correlation as the ANTs similarity measure. We also test our recent CNN-based method, VoxelMorph, which aims to produce fast registration but does not yield diffeomorphic results or uncertainty estimates [6]. We sweep the regularization parameter using our validation set, and use the optimal parameters in our results.

Results on Test set: Figure 2 shows representative results. Figure 3 illustrates Dice results on a range of anatomical structures, and Table 1 gives a summary of the results. Not only does our algorithm achieve state of the art Dice results and the fastest runtimes, but it also produces diffeomorphic registration fields (having nearly no non-negative Jacobian voxels per scan) and yields uncertainty estimates.

Specifically, all methods achieve comparable Dice results on each structure and overall. Our method and VoxelMorph require a fraction of the ANTs runtime to register two images: less than a second on a GPU, and less than a minute on a CPU (for our method). To the best of our knowledge, ANTs does not have a GPU implementation. Algorithm runtimes were computed for an NVIDIA TitanX GPU and a Intel Xeon (E5-2680) CPU, and exclude preprocessing common to all methods. Importantly, while our method achieves positive Jacobians at nearly all voxels, the flow fields resulting from the baseline methods contain a few thousand locations of non-positive Jacobians. This can be alleviated with increased spatial regularization, but this in turn leads to a drop in performance on the Dice metric.

Uncertainty. Figure 4 shows representative uncertainty maps, unique to our model. The velocity field is more certain near anatomical structure edges, and
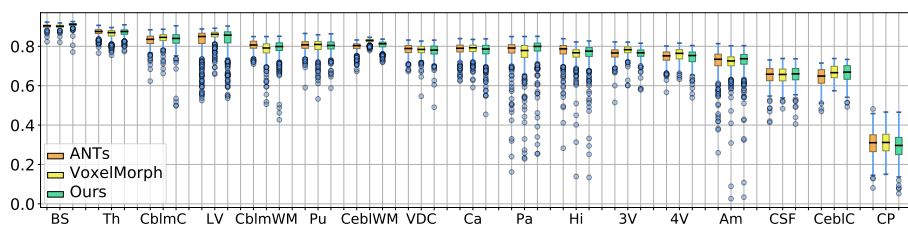

Fig. 3. Boxplots indicating Dice scores for anatomical structures for ANTs, VoxelMorph, and our algorithm. Left and right hemisphere structures are merged for visualization, and ordered by average ANTs Dice score. We include the brain stem (BS), thalamus (Th), cerebellum cortex (CblmC), lateral ventricle (LV), cerebellum white matter (CblmWM), putamen (Pu), cerebral white matter (CeblWM), ventral DC (VDC), caudate (Ca), pallidum (Pa), hippocampus (Hi), 3rd ventricle (3V), 4th ventricle (4V), amygdala (Am), CSF (CSF), cerebral cortex (CeblC), and choroid plexus (CP).
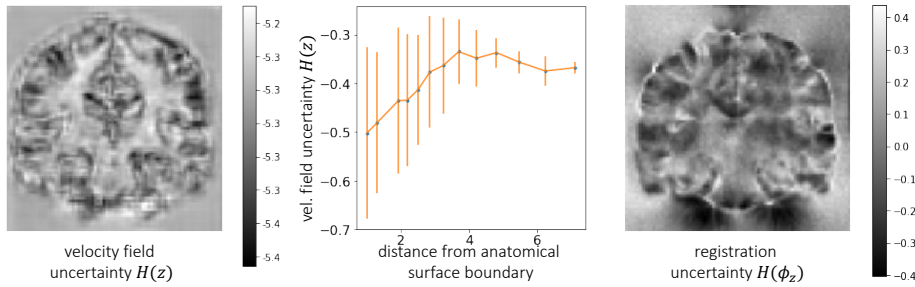
Fig. 4. Example velocity field uncertainty $H(\boldsymbol{z})$ (left) indicates low uncertainty near structure boundaries, as seen in the line graph (middle). This correlation is less obvious in the final registration field uncertainty $H(\boldsymbol{\phi}_z)$ (right).

less confident in homogenous scan regions, such as the white matter or ventricle interior.

Parameter Analysis. We perform a grid search for the two fixed hyperparameters λ and σ^2. We train a model for each parameter pair and evaluate Dice on the validation set. We search 30 values within two orders of magnitude around meaningful initial values for both parameters: $\sigma^2 \sim (0.07)^2$, the variance of the intensity difference between an *affinely* aligned image and the atlas, and $\lambda = 10000$, equivalent to a diagonal standard deviation of 1 voxel for $\boldsymbol{\phi}_z$. We found $\sigma^2 \sim (0.035)^2$ and $\lambda \in (20000, 100000)$ to perform well, and set $\lambda = 70,000$.

4 Conclusion

We propose a probabilistic model for diffeomorphic image registration and derive a learning algorithm that makes use of a convolutional neural network and an intuitive resulting loss function. To achieve unsupervised, end-to-end learning for diffeomorphic registrations, we introduce novel *scaling and squaring* differentiable layers. Our derivation is generalizable. For example, $\boldsymbol{z}$ can be a low dimensional embedding representation of a deformation field, or the displacement field itself. Our algorithm can infer the registration of new image pairs in under a second. Compared to traditional methods, our method is significantly faster, and compared to recent learning based methods, our method offers diffeomorphic guarantees, and provides natural uncertainty estimates for resulting registrations.

5 Acknowledgments

References

1. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow:

Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

2. Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 924–931. Springer, 2006.

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

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

6. Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. An unsupervised learning model for deformable medical image registration. *arXiv preprint arXiv:1802.02604*, 2018.

7. 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.

8. François Chollet. Keras. `https://github.com/fchollet/keras`, 2015.

9. 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.

10. Adrian V Dalca, Andreea Bobu, Natalia S Rost, and Polina Golland. Patch-based discrete registration of clinical brain images. In *MICCAI-PATCHMI Patch-based Techniques in Medical Imaging*. Springer, 2016.

11. 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 *DLMIA*, pages 204–212. Springer, 2017.

12. A. Di Martino 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.

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

14. B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. 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 OKeefe, 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. D.P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv:1312.6114*, 2013.

20. Arno Klein, Jesper Andersson, Babak A Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E Christensen, D Louis Collins, James Gee, Pierre Hellier, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802, 2009.

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

22. 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.

23. 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.

24. 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.

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 Alzheimers disease: the Alzheimers 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 *MICCAI*, pages 266–274. Springer, 2017.

27. O. Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.

28. 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.

29. D. Shen and C. Davatzikos. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imag.*, 21(11):1421–1439, 2002.

30. 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 *MICCAI*, pages 232–239, Cham, 2017. Springer.

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

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

33. 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.