STATISTICALLY-CONSTRAINED ROBUST DIFFEOMORPHIC REGISTRATION

Ke Zeng, Aristeidis Sotiras, Christos Davatzikos

Center for Biomedical Image Computing and Analytics
University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

Accurate subject-to-template alignment requires deformation models with high degrees of freedom to account for the high anatomical variability. Without proper regularization, such models tend to match the images aggressively, often producing unrealistic transformations, especially in the presence of noise, or pathologies such as various types of lesions. To improve the robustness of deformable registration, we propose a novel framework, which makes use of statistical deformation models (SDMs) for diffeomorphisms. We present a general approach to constructing such SDMs, and detail how to use them for regularizing a given transformation. To preserve the diffeomorphic property, while making use of linear statistical models, we convert the deformation field into a stationary velocity field through the logarithm operator. To account for learning in a high-dimensional, low-sample size setting, we model the high-dimensional velocity field as a collection of mutually constrained local velocity fields. For each local field, a low-dimensional representation is learned using principal component analysis. To capture possible dependencies across local transformations, canonical correlation analysis is performed on each pair of local velocities in the learned low-dimensional space. Experiments on healthy brain images show that the model can capture the normative variation of subject-to-template deformation fields with sub-millimeter accuracy. The method is validated on simulated brain lesion images and is tested on real brain images with pathologies, producing significantly smoother and more robust results than its non-statistical counterpart.

Index Terms— Statistical deformation models, image registration, statistical learning

1. INTRODUCTION

Spatial alignment of different subjects to a common template is widely used to delineate group differences or spatial patterns of correlation with clinical variables. The success of these studies depends greatly on establishing accurate correspondences between the different anatomies and the template [1]. To account for the high anatomical variability, a large number of degrees of freedom (DOF) are required in a general-purpose registration method to produce accurate subject-to-template mapping. However, such a deformable registration is challenged by the high dimensionality of the search space, as well as by the presence of noise, image artifacts, or pathology, which may result in aggressive and unrealistic transformations, since the pathologies are not part of the normative templates.

To address the above limitations, statistical deformation models (SDMs) have been proposed to learn the normal anatomical variability within the population of interest from a set of exemplary deformations. The resulting SDM is used to constrain the estimated deformation to lie within the subspace of learned deformations, thus endowing registration with increased robustness to noise, artifacts, and pathologies, such as lesions. Importantly, representing the transformations through a compact SDM drastically reduces the number of DOF and may accelerate the registration task. A major challenge in training SDMs is the high dimensionality of the deformation fields compared to the limited sample size. To tackle the dimensionality challenge, a cubic B-spline Free Form Deformation model is typically employed to parametrize the transformations, and principal component analysis (PCA) is used to learn the control point displacements [2, 3]. To further reduce the dimensionality of the problem, Glocker *et al.* [4] proposed to cluster the densely sampled control points into several local parts with similar behavior, while Xue *et al.* [5] used wavelet-based decompositions coupled with PCA in each wavelet band.

However, despite the importance of diffeomorphisms for computational anatomy, previous studies have largely overlooked them when learning an SDM. Only a few studies have considered SDMs that preserve the diffeomorphic property. Qiu et al. [6] used PCA in the space of initial momentum under the large deformation diffeomorphic metric mapping (LDMM). Since the momentum has the same dimensionality as the deformation fields, the method is limited by the inability of PCA to estimate high-dimensional covariance structures given limited training samples [7]. In a similar approach, Zhang et al. used global PCA coupled with a low-frequency Fourier-based representation of the LDDMM velocity field [8] to analyze anatomical variations. However, this approach has not been used for image registration, while its ability to cope with complex local variability is limited by the global nature of the statistical model and the band-limited representation.

We present here an approach to learning an SDM for diffeomorphisms, which can account for the high dimensionality of the problem, while also effectively modeling local variations. To account for the fact that the group of diffeomorphisms is not a linear space [9], we adopt the Log-Euclidean framework [9] and parametrize diffeomorphism by its principal logarithm. The log-domain parametrization allows us to perform vectorial statistics while satisfying the invertibility constraint. To tackle the dimensionality challenge, we treat a dense velocity field as a collection of local velocity fields. Each local velocity field is embedded in a much lower-dimensional space than its global counterpart, allowing PCA to model the variability of each local velocity field using a limited number of observations. To capture possible dependencies across local transformations and thus preserve the global structure of the velocity field, canonical correlation analysis (CCA) is performed to learn the relation between pairs of local velocity fields. Experimental results indicate that the proposed model can capture the normative variation of subject-to-template deformation fields with sub-millimeter accuracy. We also report promising results when validating the method on image data with synthetic and real lesions. To the best of our knowledge, this is the first attempt to capture diffeomorphisms in an SDM, while also demonstrating its potential advantage over non-

statistical methods in registering lesion-bearing images.

2. METHOD

Let us consider a common template $\mathcal{T}$ and a set of training images $\mathcal{M} = \{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_N\}$. Let us also assume that the set of corresponding diffeomorphisms $\{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N\}$ that map $\mathcal{M}$ to $\mathcal{T}$ is known. Each diffeomorphism $\mathcal{D}_i$ is converted into a stationary velocity field (SVF) $\mathcal{V}_i$ by taking its principal logarithm [9]. Working in the logarithm domain has the advantage that the invertibility property can be preserved by linear statistical methods.

To address the dimensionality challenge, each SVF $\mathcal{V}_i$ is subsequently represented by a collection of local velocity fields. Each local velocity field corresponds to a sub-block of $\mathcal{V}_i$, centered at a candidate location. Specifically, if the block size is set to $p_x \times p_y \times p_z$ in the 3D case, then the local velocity filed will be an element in $\mathcal{R}^{p_x \times p_y \times p_z \times 3}$. We use $\mathcal{S}$ to represent the set of all candidate block centers. The set of training blocks centered at $s \in \mathcal{S}$ is denoted by $\mathbf{V}_s := \{\mathbf{v}_{s,1}, \mathbf{v}_{s,2}, \ldots, \mathbf{v}_{s,N}\}$.

2.1. Learning local velocity fields

Principal component analysis is used to capture the variation within local velocity fields. For each candidate location s, we denote by $\bar{\mathbf{v}}_s$ the sample mean, and by $\mathbf{Q}_s$ the sample covariance of the vectorized training set $\mathbf{V}_s \subset \mathcal{R}^{p_s}$. The approximate eigen-decomposition of $\mathbf{Q}^s$ may be written as

$$\mathbf{Q}_s \approx \mathbf{U}_s \mathbf{\Lambda}_s {\mathbf{U}_s}^T,$$

where $\mathbf{U}_s \in \mathcal{R}^{p_s \times k_s}$ ($\mathbf{\Lambda}_s \in \mathcal{R}^{k_s \times k_s}$) is the truncated eigenvector (eigenvalue) matrix so that 95% of the variation in $\mathbf{V}_s$ is preserved in $\{{\mathbf{U}_s}^T \mathbf{v}_{s,1}, \ldots, {\mathbf{U}_s}^T \mathbf{v}_{s,N}\}$.

Given $\mathbf{U}_s$ and $\mathbf{\Lambda}_s$, the local velocity field may be parametrized by a coefficient $\boldsymbol{\beta}_s = [\beta_s^1, \beta_s^2, \ldots, \beta_s^{k_s}] \in \mathcal{R}^{k_s}$:

$$\mathbf{v}_s = \bar{\mathbf{v}}_s + \mathbf{U}_s \boldsymbol{\beta}_s \quad \text{such that } |\beta_s^i| \leq c\sqrt{\mathbf{\Lambda}_s^{i,i}},$$

where c controls the admissible range of each β_s^i. For example, setting $c = 3$ allows β_s^i to be any value less than three standard deviations away from the mean.

2.2. Learning correlation across local velocity fields

The above PCA model is only capable of capturing local variations within each block. To preserve the global structure of the velocity field, we carry out a canonical correlation analysis [10] to link each pair of local fields.

Let us consider two different block centers s_1 and s_2. The objective is to learn, using CCA, potential correlations between velocity field blocks centered at s_1 and s_2. Given two vectors $\mathbf{X} = (x_1, \ldots, x_n)$ and $\mathbf{Y} = (y_1, \ldots, y_m)$, CCA will find the linear combinations of the x_i and y_j that have the maximum correlation with each other. To take advantage of the low-dimensional parametrization learned in Sec. 2.1, we perform CCA on the coefficients $\{\boldsymbol{\beta}_{s_1,1}, \ldots, \boldsymbol{\beta}_{s_1,N}\}$ and $\{\boldsymbol{\beta}_{s_2,1}, \ldots, \boldsymbol{\beta}_{s_2,N}\}$, rather than on the raw samples $\mathbf{V}_{s_1}$ and $\mathbf{V}_{s_2}$. If we keep the first d_{s_1,s_2} pairs of canonical variables, with canonical weights $\mathbf{A}_{s_1,s_2} \in \mathcal{R}^{k_{s_1} \times d_{s_1,s_2}}$, $\mathbf{B}_{s_1,s_2} \in \mathcal{R}^{k_{s_2} \times d_{s_1,s_2}}$ and canonical correlation $\boldsymbol{\rho}_{s_1,s_2} = [\rho_{s_1,s_2}^1, \ldots, \rho_{s_1,s_2}^{d_{s_1,s_2}}] \in \mathcal{R}^{d_{s_1,s_2}}$, then the cross-covariance matrix between $\mathbf{A}_{s_1,s_2}\boldsymbol{\beta}_{s_1}$ and $\mathbf{B}_{s_1,s_2}\boldsymbol{\beta}_{s_2}$ is a diagonal matrix with $\rho_{s_1,s_2}^1, \ldots, \rho_{s_1,s_2}^{d_{s_1,s_2}}$ on the main diagonal.

2.3. The statistical deformation model

We now present the complete statistical model for a dense velocity field $\mathcal{D}$. Since $\mathcal{D}$ is parametrized by its logarithm $\mathcal{V}$, and $\mathcal{V}$ is considered as an ensemble of local blocks $\{\mathbf{v}_s, s \in \mathcal{S}\}$, one only needs to specify the distribution of the local blocks, or more compactly, the distribution of the coefficients $\{\boldsymbol{\beta}_s, s \in \mathcal{S}\}$. Here, we propose the following joint probability density function for the coefficients,

$$p(\boldsymbol{\beta}_s, s \in \mathcal{S}) \propto \prod_{s \in \mathcal{S}} f(\boldsymbol{\beta}_s) \prod_{s_1, s_2 \in \mathcal{S}, s_1 \neq s_2} g(\boldsymbol{\beta}_{s_1}, \boldsymbol{\beta}_{s_2} | \lambda).$$

The unary term $f(\boldsymbol{\beta}_s)$ is a product of indicator functions

$$f(\boldsymbol{\beta}_s) = \prod_{i=1,\ldots,k_s} \mathbf{1}(|\boldsymbol{\beta}_s^i| \leq c\sqrt{\mathbf{\Lambda}_s^{i,i}}),$$

which constrain each element in $\boldsymbol{\beta}_s$ to be within c standard variations from the mean.

The interaction potential $g(\boldsymbol{\beta}_{s_1}, \boldsymbol{\beta}_{s_2} | \lambda)$ is defined as

$$g(\boldsymbol{\beta}_{s_1}, \boldsymbol{\beta}_{s_2} | \lambda) = \exp(-\lambda \boldsymbol{\beta}_{s_1}^T \mathbf{A}_{s_1,s_2}^T \mathbf{\Gamma}_{s_1,s_2} \mathbf{B}_{s_1,s_2} \boldsymbol{\beta}_{s_2}),$$

where $\lambda > 0$ is a hyperparameter that controls the weight of the interaction potential, and $\mathbf{\Gamma}_{s_1,s_2}$ is a $d_{s_1,s_2} \times d_{s_1,s_2}$ diagonal matrix with $\rho_{s_1,s_2}^d / (1 - (\rho_{s_1,s_2}^d)^2)$ on the main diagonal. The interaction term penalizes configurations of $\boldsymbol{\beta}_{s_1}, \boldsymbol{\beta}_{s_2}$ that deviate from the learned linear correlation.

Given a new observation $\mathcal{V} = \log \mathcal{D}$, the most likely underlying coefficient may be recovered as the maximum a posteriori probability (MAP) estimate

$$\tilde{\boldsymbol{\beta}}_s = \arg\max_{\boldsymbol{\beta}_s} p(\mathcal{V} | \boldsymbol{\beta}_s, s \in \mathcal{S}) p(\boldsymbol{\beta}_s, s \in \mathcal{S}),$$

where the conditional likelihood $p(\mathcal{V} | \boldsymbol{\beta}_s, s \in \mathcal{S})$ has the form

$$p(\mathcal{V} | \boldsymbol{\beta}_s, s \in \mathcal{S}) \propto \prod_{s \in \mathcal{S}} \exp(-\|\mathbf{v}_s - (\bar{\mathbf{v}}_s + \mathbf{U}_s \boldsymbol{\beta}_s)\|_2^2).$$

Let $\tilde{\mathcal{V}}$ be the velocity field generated by $\tilde{\boldsymbol{\beta}}_s$ and let $\tilde{\mathcal{D}} = \exp(\tilde{\mathcal{V}})$. We denote by P the mapping from $\mathcal{V}(\mathcal{D})$ to $\tilde{\mathcal{V}}(\tilde{\mathcal{D}})$. The function P may be interpreted as a regularization operator that projects $\mathcal{V}(\mathcal{D})$ onto the admissible subspace learned by the statistical model. Using $P(\mathcal{D})$ instead of $\mathcal{D}$ has the potential to improve the robustness of the registration results with little sacrifice on accuracy.

3. EXPERIMENTAL RESULTS

We used a dataset comprising 380 brain MR images (T1) from healthy adults aged between 45 and 55 to learn the SDM. The procedure proposed in [11] was used to create an unbiased template $\mathcal{T}$ from 20 randomly selected volumes that are part of the normative dataset. $\mathcal{T}$ was used as the reference space across the experiments. We used Log-Demons [12], which is a diffeomorphic registration method, to perform pair-wise registration between the reference image and each subject image. To study the performance of the method on lesion-bearing images, the second dataset of 28 T1 volumes showing stroke lesions was also used. Skull stripping and bias correction were performed for all volumes.

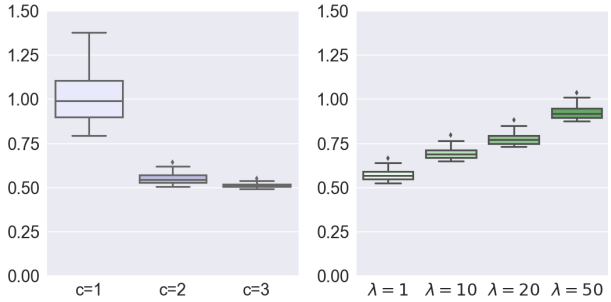

Fig. 1. Box plots showing the RMSD between $\mathcal{D}$ and $\tilde{\mathcal{D}}$. In the left panel λ was set to 0, while c ranged in $1, 2, 3$. In the right panel, c was fixed at $c = 2$, while λ took value in $1, 10, 20, 50$.

3.1. Experiments using the normative dataset

In the first experiment, we demonstrate the capability of the method to capture shape variations within the normative database. The parameter setting is detailed in what follows. The size of the local blocks was set to $16\,\mathrm{mm}$ isotropic, and the distance between block centers was set to $12\,\mathrm{mm}$. Out of the 380 deformation field samples generated by Log-Demons, 200 samples were used for computing the PCA basis, 150 samples were used to learn the canonical correlations, and the other 30 samples were reserved for testing. To make the computation tractable, only canonical correlation pairs with correlation higher than 0.8 were used to define the interaction potential.

With a fixed $\lambda = 0$, we varied the parameter c to examine its influence on the representation ability of the model, which was measured by the root-mean-square difference (RMSD) between $\mathcal{D}$ and $\tilde{\mathcal{D}}$ for the 30 test samples. The results for $c \in 1, 2, 3$ are summarized by the plots in the left panel of Fig. 1. With a widely used value $c = 2$, the method was able to represent left-out samples in the normative database with sub-millimeter accuracy.

We also studied the role of the regularization parameter λ on the representation ability with a fixed $c = 2$. RMSD was again used as the quantitative measure. Results for $\lambda \in 1, 10, 20, 50$ are shown in the right panel of Fig. 1. The RMSD increases with increasing λ, a behavior that was expected as the DOF of the representation reduces with larger λ, resulting in a larger model bias. It is, however, worth noting that even with a fairly large λ ($\lambda = 50$), the discrepancy between $\mathcal{D}$ and $\tilde{\mathcal{D}}$ is still relatively small (median RMSD is below $1\,\mathrm{mm}$). This suggests that the learned correlation-based regularization generalizes well to unseen samples.

3.2. Experiments using simulated lesions

We next evaluated the robustness of the proposed framework in the presence of simulated lesions. 10 annotated T1 images that contain a variety of lesions (necrosis, white matter lesions, tumors, etc.) were used as the base for the simulation. A base image $\mathcal{B}$ was used to plant lesions on a target image $\mathcal{I}$ according to the following procedure. We used a near-affine transformation to coarsely align $\mathcal{B}$ and its manual annotations to $\mathcal{I}$ while preserving the shape of the lesion. Given the aligned lesion mask $\mathcal{M}$, we fused the two images by replacing the intensity values of the voxels in $\mathcal{I}$ that are within $\mathcal{M}$ with the values of the corresponding voxels in $\mathcal{B}$. Gaussian smoothing was performed near the boundary of $\mathcal{M}$ to eliminate image artifacts caused by the fusion process. We denote by $\mathcal{I}'$ the simulated image. 15 random selections from the testing set in Sec. 3.1 were used as the target volume $\mathcal{I}$, resulting in a total number of 150 simulated images

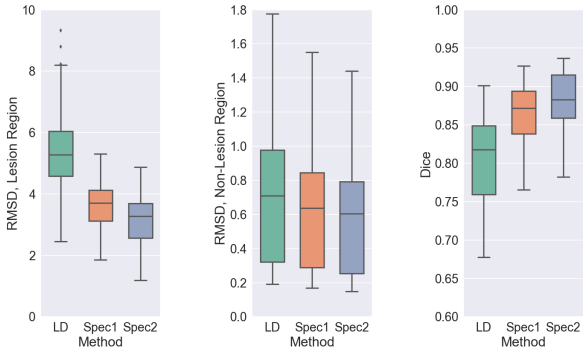

Fig. 2. Box plots summarizing the robustness measures under the simulated setting. From left to right: RMSD between $\mathcal{D}$ (before lesion simulation) and $\mathcal{D}'$ (after lesion simulation) inside the simulated lesion region, RMSD between $\mathcal{D}$ and $\mathcal{D}'$ in the non-lesion region, Dice coefficients between $\mathcal{M} \circ \mathcal{D}$ and $\mathcal{M} \circ \mathcal{D}'$.

(see characteristic examples in the second row, left panel of Fig. 3).

The simulation setting allows us to examine how the registration results change after the lesions are introduced. Assuming Log-Demons returns $\mathcal{D}$ (or $\mathcal{D}'$) when warping $\mathcal{I}$ (or $\mathcal{I}'$) to the template $\mathcal{T}$, we evaluated the difference between $\mathcal{D}$ and $\mathcal{D}'$, along with the difference between $P(\mathcal{D})$ and $P(\mathcal{D}')$. We used the same block size and stride length values as in Sec. 3.1, and set $c = 2$. Two values for the parameter λ were considered, namely $\lambda \in 1, 20$. In what follows, we use the abbreviation LD, Spec1 and Spec2 for Log-Demons, projected Log-Demons with $\lambda = 1$ and $\lambda = 20$, respectively. For brevity we omit the operator P and denote by $\mathcal{D}$ and $\mathcal{D}'$ the deformation fields generated before and after lesion simulation regardless of the method. $\mathcal{D}$ and $\mathcal{D}'$ were compared using the following metrics: the RMSD between $\mathcal{D}$ and $\mathcal{D}'$, and the Dice overlap coefficient between the warped lesion regions $\mathcal{M} \circ \mathcal{D}$ and $\mathcal{M} \circ \mathcal{D}'$.

The box plots in the left two panels of Fig. 2 summarize the distribution of the RMSD between $\mathcal{D}$ and $\mathcal{D}'$, with the RMSD inside the lesion shown on the left, and the RMSD away from the lesion shown in the middle. Spec1 and Spec2 were more robust than LD under both scenarios, with a more notable difference inside the lesion region. The box plot in the right panel shows the Dice overlap coefficients between $\mathcal{M} \circ \mathcal{D}$ and $\mathcal{M} \circ \mathcal{D}'$ obtained by the three methods. Again Spec1 and Spec2 outperformed LD, achieving significantly higher Dice scores. It is also interesting to note that Spec2 performed consistently more robust than Spec1. This is because the employed larger regularization weight reduces model variance. However, this comes at the cost of worse representation errors within normative regions, as the results in Sec. 3.1 suggest. In short, the parameter λ needs to be adjusted on a per-application basis to achieve the desired balance between model fidelity and model robustness.

We conclude the simulation study with a discussion on the representative image slices shown in the left panel of Fig. 3. In all cases, LD produced aggressive deformations to "eliminate" the lesions, resulting in registration results that do not look physically plausible. On the contrary, results produced by Spec1 and Spec2 suggest that such behavior is effectively prevented by the statistical constraints without compromising the registration quality away from the lesions. Spec1 and Spec2 produced much smoother deformations around the lesions, while still being able to accurately align brain structures elsewhere.

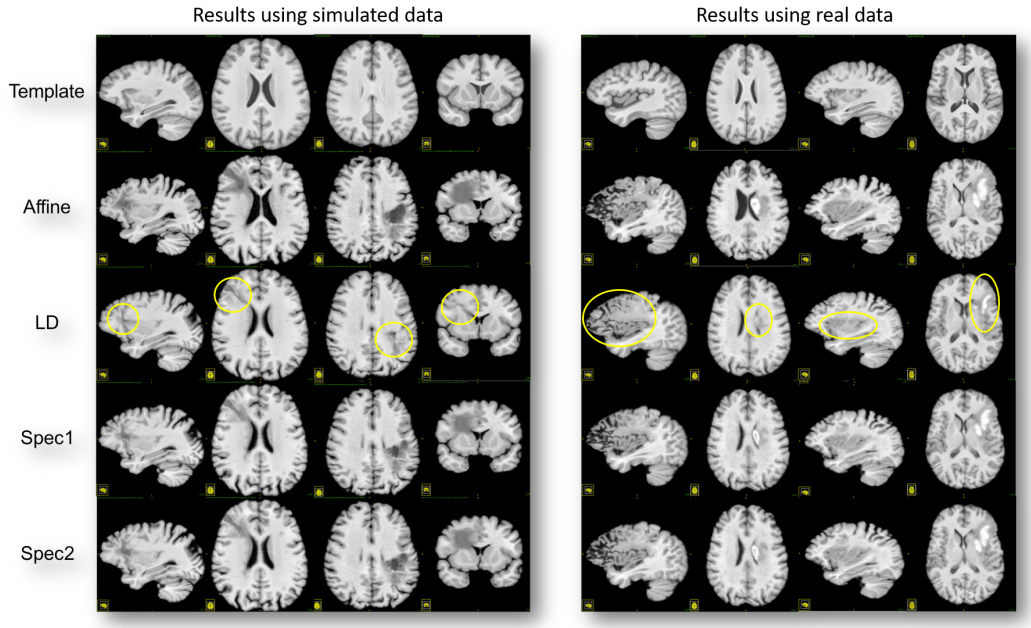

Fig. 3. Representative slices of the registered images. The left panel contains examples from the simulation study, while the right panel contains examples from the clinical study. From top to bottom: template image, affinely aligned subject image, subject image aligned using LD, subject image aligned using Spec1, and subject image aligned using Spec2. The lesion regions that have been aggressively contracted by LD are highlighted by the yellow circle. Note, in particular, the much more natural behavior of the proposed method in these regions.

3.3. Experiments using clinical lesion-bearing data

In this last experiment, we studied the behavior of $\mathcal{D}$ and $\tilde{\mathcal{D}}$ on lesion-bearing images, using real clinical images. In contrast to the simulation setting, we do not have access to the subject's image had he/she been spared by the disease. Therefore, the "before" vs. "after" comparison is no longer feasible. Instead, we quantitatively evaluated the performance of the different approaches to measuring the lesion volume change before and after registration, as well as the transformation smoothness. Visual inspection of the registration results was also performed. The same parameter setting as in Sec. 3.2 was employed.

Fig. 4 summarizes the quantitative comparison between the different methods. The box plot in the left panel shows how the harmonic energy (HE) is distributed for each method. The HEs of the LD deformation fields are significantly higher than those of Spec1 and Spec2. The distribution of the extreme Jacobian determinant (JD) values for each method is shown in the middle panel in Fig. 4. Specifically, absolute log JD values that belong to the top 5 percentile in each volume are summarized in the box plot. The trend is similar to that of the HEs, with LD averaging the highest among the three methods. It is worth noting that LD produces JD values as high as $e^3 \approx 20.09$, while the other two methods do not produce JD values higher than $e^{1.5} \approx 4.48$. We quantified the lesion volume change after the deformable registration by calculating the logarithm of the change ratio $\log(\frac{\textit{Lesion Volume After}}{\textit{Lesion Volume Before}})$. The distribution of the metric for each method is presented in the right panel of Fig. 4. One may again observe LD's problematic behavior since its usage leads to unrealistic lesion volume reduction. This problematic behavior is showcased in the representative results shown in the right panel of Fig. 3. LD maximized image similarity by unrealistically contracting the brain lesion. Contrarily, Spec1 and Spec2 achieved good correspondences in the normative areas, without being influenced by the presence of lesions.

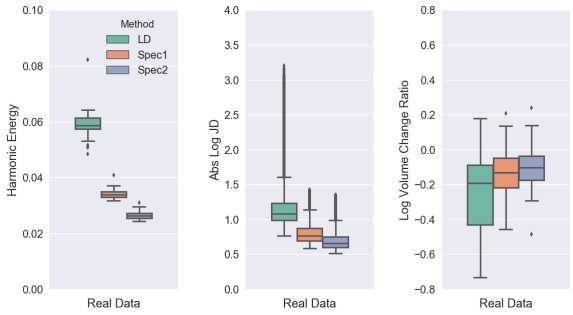

Fig. 4. Boxplots outlining the performance of the three methods on the clinical data. From left to right: harmonic energy of the deformation fields, extreme log Jacobican determinant values and log lesion volume change ratio.

4. CONCLUSION

We introduced a novel registration approach based on a statistical deformation model of diffeomorphisms. The proposed approach leverages the Log-Euclidean framework to preserve diffeomorphism when performing vectorial statistics. Importantly, it employs PCA to effectively learn local statistical models in the high-dimensional, low-sample setting. We additionally used CCA to learn the dependencies between pairs of local velocity fields to ensure that the global structure of the velocity field is preserved. Promising results suggest that the proposed statistical registration method is more robust than its non-statistical counterpart in the presence of abnormalities.

5. REFERENCES

[1] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, July 2013.

[2] Daniel Rueckert, Alejandro F Frangi, and Julia A Schnabel, "Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration," *IEEE transactions on medical imaging*, vol. 22, no. 8, pp. 1014–1025, 2003.

[3] John a. Onofrey, Xenophon Papademetris, and Lawrence H. Staib, "Low-Dimensional Non-Rigid Image Registration Using Statistical Deformation Models From Semi-Supervised Training Data," *IEEE Transactions on Medical Imaging*, vol. 34, no. 7, pp. 1522–1532, 2015.

[4] Ben Glocker, Nikos Komodakis, Nassir Navab, Georgios Tziritas, and Nikos Paragios, "Dense registration with deformation priors," in *Information processing in medical imaging*, 2009, pp. 540–551.

[5] Zhong Xue, Dinggang Shen, and Christos Davatzikos, "Statistical representation of high-dimensional deformation fields with application to statistically constrained 3d warping," *Medical Image Analysis*, vol. 10, no. 5, pp. 740–751, 2006.

[6] A. Qiu, L. Younes, and M. I. Miller, "Principal component based diffeomorphic surface mapping," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 302–311, Feb 2012.

[7] T Tony Cai, Zhao Ren, and Harrison H Zhou, "Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation," *Electronic journal of statistics*, vol. 10, no. 1, pp. 1 – 59, 2016.

[8] Miaomiao Zhang, William M. Wells, and Polina Golland, "Probabilistic modeling of anatomical variability using a low dimensional parameterization of diffeomorphisms," *Medical Image Analysis*, vol. 41, no. Supplement C, pp. 55 – 62, 2017, Special Issue on the 2016 Conference on Medical Image Computing and Computer Assisted Intervention (Analog to MICCAI 2015).

[9] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache, "A log-euclidean framework for statistics on diffeomorphisms," 2006, pp. 924–931.

[10] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, Dec 2004.

[11] Brian B. Avants, Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee, "A reproducible evaluation of {ANTs} similarity metric performance in brain image registration," *NeuroImage*, vol. 54, no. 3, pp. 2033 – 2044, 2011.

[12] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache, "Symmetric log-domain diffeomorphic registration: A demons-based approach," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2008, pp. 754–761.