

# Anatomical Data Augmentation via Fluid-based Image Registration

Zhengyang Shen, Zhenlin Xu, Sahin Olut, and Marc Niethammer

Department of Computer Science, UNC Chapel Hill

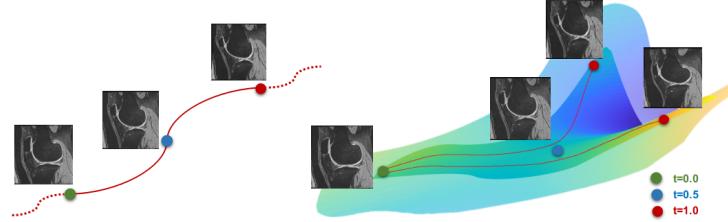
**Abstract.** We introduce a fluid-based image augmentation method for medical image analysis. In contrast to existing methods, our framework generates anatomically meaningful images via interpolation from the geodesic subspace underlying given samples. Our approach consists of three steps: 1) given a source image and a set of target images, we construct a geodesic subspace using the Large Deformation Diffeomorphic Metric Mapping (LDDMM) model; 2) we sample transformations from the resulting geodesic subspace; 3) we obtain deformed images and segmentations via interpolation. Experiments on brain (LPBA) and knee (OAI) data illustrate the performance of our approach on two tasks: 1) data augmentation during training and testing for image segmentation; 2) one-shot learning for single atlas image segmentation. We demonstrate that our approach generates anatomically meaningful data and improves performance on these tasks over competing approaches. Code is available at <https://github.com/uncbiag/easyreg>.

## 1 Introduction

Training data-hungry deep neural networks is challenging for medical image analysis where manual annotations are more difficult and expensive to obtain than for natural images. Thus it is critical to study how to use scarce annotated data efficiently, e.g., via data-efficient models [30,11], training strategies [20] and semi-supervised learning strategies utilizing widely available unlabeled data through self-training [3,16], regularization [4], and multi-task learning [7,31,36].

An alternative approach is data augmentation. Typical methods for medical image augmentation include random cropping [12], geometric transformations [18,15,24] (e.g., rotations, translations, and free-form deformations), and photometric (i.e., color) transformations [14,21]. Data-driven data augmentation has also been proposed, to learn generative models for synthesizing images with new appearance [28,9], to estimate class/template-dependent distributions of deformations [10,19,34] or both [35,6]. Compared with these methods, our approach focuses on a geometric view and constructs a continuous geodesic subspace as an estimate of the space of anatomical variability.

Compared with the high dimensionality of medical images, anatomical variability is often assumed to lie in a much lower dimensional space [1]. Though how to directly specify this space is not obvious, we can rely on reasonable assumptions informed by the data itself. We assume there is a diffeomorphic



**Fig. 1.** Illustration of our fluid-based data augmentation using a 1D (left) and 2D (right) geodesic subspace. We assume a registration from a source to a target image in unit time. In 1D, we can sample along the geodesic path ( $t \in [0, 1]$ ) between the source ( $t = 0$ ) and the target images ( $t = 1$ ). We can also extrapolate  $t \notin [0, 1]$ . In the 2D case, a source and two target images define a two-dimensional geodesic subspace.

transformation between two images, that image pairs can be connected via a geodesic path, and that appearance variation is implicitly captured by the appearance differences of a given image population. For longitudinal image data, we can approximate images at intermediate time points by interpolation or predict via extrapolation. As long as no major appearance changes exist, diffeomorphic transformations can provide realistic intermediate images<sup>1</sup>. Based on these considerations, we propose a data augmentation method based on fluid registration which produces *anatomically plausible* deformations and retains appearance differences of a given image population. Specifically, we choose the Large Deformation Diffeomorphic Metric Mapping (LDDMM) model as our fluid registration approach. LDDMM comes equipped with a metric and results in a geodesic path between a source and a target image which is parameterized by the LDDMM initial momentum vector field. Given two initial momenta in the tangent space of the same source image, we can define a geodesic plane, illustrated in Fig. 1; similarly, we can construct higher dimensional subspaces based on convex combinations of sets of momenta [22]. Our method includes the following steps: 1) we compute a set of initial momenta for a source image and a set of target images; 2) we generate an initial momentum via a convex combination of initial momenta; 3) we sample a transformation on the geodesic path determined by the momentum; and 4) we warp the image and its segmentation according to this transformation.

Data augmentation is often designed for the training phase. However, we show the proposed approach can be extended to the testing phase, e.g., a testing image is registered to a set of training images (with segmentations) and the deep learning (DL) segmentation model is evaluated in this warped space (where it was trained, hence ensuring consistency of the DL input); the predicted segmentations are then mapped back to their original spaces. In such a setting, using

<sup>1</sup> In some cases, for example for lung images, sliding effects need to be considered, violating the diffeomorphic assumption.

LDDMM can guarantee the existence of the inverse map whereas traditional elastic approaches cannot.

**Contributions:** 1) We propose a general fluid-based approach for anatomically consistent medical image augmentation for both training and *testing*. 2) We build on LDDMM and can therefore assure well-behaved diffeomorphic transformations when interpolating *and* extrapolating samples with large deformations. 3) Our method easily integrates into different tasks, such as segmentation and one-shot learning for which we show general performance improvements.

## 2 LDDMM Method

LDDMM [5] is a fluid-based image registration model, estimating a spatio-temporal velocity field  $v(t, x)$  from which the spatial transformation  $\varphi$  can be computed via integration of  $\partial_t \varphi(t, x) = v(t, \varphi(t, x))$ ,  $\varphi(0, x) = x$ . For appropriately regularized velocity fields [8], diffeomorphic transformations can be guaranteed. The optimization problem underlying LDDMM can be written as

$$v^* = \operatorname{argmin}_v \frac{1}{2} \int_0^1 \|v(t)\|_L^2 dt + \operatorname{Sim}(I(1), I_1) \quad \text{s.t.} \quad \partial_t I + \langle \nabla I, v \rangle = 0, \quad I(0) = I_0, \quad (1)$$

where  $\nabla$  denotes the gradient,  $\langle \cdot, \cdot \rangle$  the inner product, and  $\operatorname{Sim}(A, B)$  is a similarity measure between images. We note that  $I(1, x) = I_0 \circ \varphi^{-1}(1, x)$ , where  $\varphi^{-1}$  denotes the inverse of  $\varphi$  in the target image space. The evolution of this map follows  $\partial_t \varphi^{-1} + D\varphi^{-1}v = 0$ , where  $D$  is the Jacobian. Typically, one seeks a velocity field which deforms the source to the target image in unit time. To assure smooth transformations, LDDMM penalizes non-smooth velocity fields via the norm  $\|v\|_L^2 = \langle Lv, Lv \rangle$ , where  $L$  is a differential operator.

At optimality the following equations hold [33] and the entire evolution can be parameterized via the initial vector-valued momentum,  $m = L^\dagger Lv$ :

$$m(0)^* = \operatorname{argmin}_{m(0)} \frac{1}{2} \langle m(0), v(0) \rangle + \operatorname{Sim}(I_0 \circ \varphi^{-1}(1), I_1), \quad (2)$$

$$\text{s.t.} \quad \varphi_t^{-1} + D\varphi^{-1}v = 0, \quad \varphi(0, x) = x, \quad (3)$$

$$\partial_t m + \operatorname{div}(v)m + Dv^T(m) + Dm(v) = 0, \quad m(0) = m_0, \quad v = K \star m, \quad (4)$$

where we assume  $(L^\dagger L)^{-1}m$  is specified via convolution  $K \star m$ . Eq. 4 is the Euler-Poincaré equation for diffeomorphisms (EPDiff) [33], defining the evolution of the spatio-temporal velocity field based on the initial momentum  $m_0$ .

The geodesic which connects the image pair  $(I_0, I_0 \circ \varphi^{-1}(1))$  and approximates the path between  $(I_0, I_1)$  is specified by  $m_0$ . We can sample along the geodesic path, assuring diffeomorphic transformations. As LDDMM assures diffeomorphic transformations, we can also obtain the inverse transformation map,  $\varphi$  (defined in source image space, whereas  $\varphi^{-1}$  is defined in target image space) by solving

$$\varphi(1, x) = x + \int_0^1 v(t, \varphi(t, x)) dt, \quad \varphi(0, x) = x. \quad (5)$$

Computing the inverse for an arbitrary displacement field on the other hand requires the numerical minimization of  $\|\varphi^{-1} \circ \varphi - id\|^2$ . Existence of the inverse map cannot be guaranteed for such an arbitrary displacement field.

### 3 Geodesic Subspaces

We define a geodesic subspace constructed from a source image and a set of target images. Given a dataset of size  $N$ ,  $I_c \in \mathbb{R}^D$  denotes an individual image  $c \in \{1 \dots N\}$ , where  $D$  is the number of voxels. For each source image  $I_c$ , we further denote a target set of  $K$  images as  $\mathbf{I}_K^c$ . We define  $M_K^c := \{m_0^{cj} | \mathcal{M}(I_c, I_j), \mathcal{M} : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}^{D \times d}, I_j \in \mathbf{I}_K^c\}$  as a set of  $K$  different initial momenta, where  $\mathcal{M}$  maps from an image pair to the corresponding initial momentum via Eqs. 2-4;  $d$  is the spatial dimension. We define convex combinations of  $M_K^c$  as

$$C(M_K^c) := \left\{ \tilde{m}_0^c \middle| \tilde{m}_0^c = \sum_{j=1}^K \lambda_j m_0^{cj}, m_0^{cj} \in M_K^c, \lambda_j \geq 0 \forall j, \sum_{j=1}^K \lambda_j = 1 \right\}. \quad (6)$$

Restricting ourselves to convex combinations, instead of using the entire space defined by arbitrary linear combinations of the momenta  $M_K^c$  allows us to retain more control over the resulting momenta magnitudes. For our augmentation strategy we simply sample an initial momentum  $\tilde{m}_0^c$  from  $C(M_K^c)$ , which, according to the EPDiff Eq. 4, determines a geodesic path starting from  $I_c$ . If we set  $K = 2$ , for example, the sampled momentum parameterizes a path from a source image toward two target images, where the  $\lambda_i$  weigh how much the two different images drive the overall deformation. As LDDMM registers a source to a target image in unit time, we obtain interpolations by additionally sampling  $t$  from  $[0, 1]$ , resulting in the intermediate deformation  $\varphi_{\tilde{m}_0^c}^{-1}(t)$  from the geodesic path starting at  $I_c$  and determined by  $\tilde{m}_0^c$ . We can also extrapolate by sampling  $t$  from  $\mathbb{R} \setminus [0, 1]$ . We then synthesize images via interpolation:  $I_c \circ \varphi_{\tilde{m}_0^c}^{-1}(t)$ .

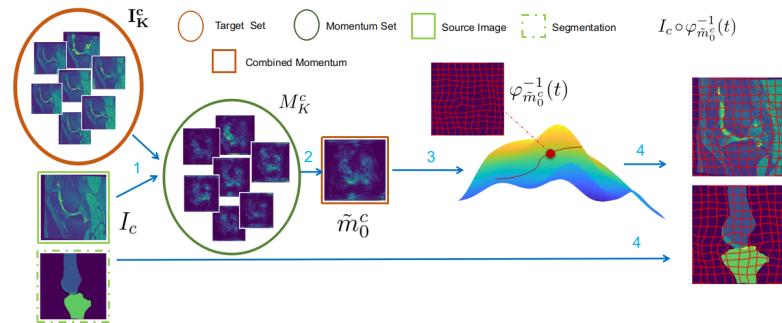
## 4 Segmentation

In this section, we first introduce an augmentation strategy for general image segmentation (Sec. 4.1) and then a variant for one-shot segmentation (Sec. 4.2).

### 4.1 Data augmentation for general image segmentation

We use a two-phase data augmentation approach consisting of (1) pre-augmentation of the training data and (2) post-augmentation of the testing data. During the training phase, for each training image,  $I_c$ , we generate a set of new images by sampling from its geodesic subspace,  $C(M_K^c)$ . This results in a set of deformed images which are anatomically meaningful and retain the appearance of  $I_c$ . We apply the same sampled spatial transformations to the segmentation of the training image, resulting in a new set of warped images and segmentations. We train a segmentation network based on this augmented dataset.

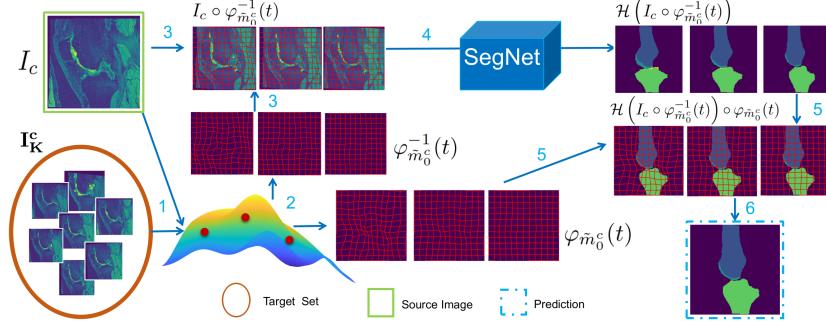
During the testing phase, for each testing image, we also create a set of new images using the same strategy described above. Specifically, we pair a testing image with a set of *training* images to create the geodesic subspace for sampling. This will result in samples that come from a similar subspace that has been used for augmentation during training. A final segmentation is then obtained by warping the predicted segmentations back to the original space of the image to be segmented and applying a label-fusion strategy. Consequently, we expect that the segmentation network performance will be improved as it (1) is allowed to see multiple views of the same image and (2) the set of views is consistent with the set of views that the segmentation network was trained with.



**Fig. 2.** Illustration of the training phase data augmentation. Given a source image  $I_c$  and a set of target images  $\mathbf{I}_K^c$ , a set of momenta  $M_K^c$  is first computed. Then a momentum  $\tilde{m}_0^c \in C(M_K^c)$  is sampled from the convex combination of these momenta  $C(M_K^c)$  defining a geodesic path starting from the source image. Lastly, a transformation  $\varphi_{\tilde{m}_0^c}^{-1}(t)$  is sampled on the geodesic and used to warp the source image and its segmentation.

Fig. 2 illustrates the training phase data augmentation. We first compute  $M_K^c$  by picking an image  $I_c$ ,  $c \in \{1 \dots N\}$  from a training dataset of size  $N$  and a target set  $\mathbf{I}_K^c$  of cardinality  $K$ , also sampled from the training set. We then sample  $\tilde{m}_0^c \in C(M_K^c)$  defining a geodesic path from which we sample a deformation  $\varphi_{\tilde{m}_0^c}^{-1}(t)$  at time point  $t$ . We apply the same strategy multiple times and obtain a new deformation set for each  $I_c$ ,  $c \in \{1 \dots N\}$ . The new image set  $\{I_c \circ \varphi_{\tilde{m}_0^c}^{-1}(t)\}$  consisting of the chosen set of random transformations of  $I_c$  and the corresponding segmentations can then be obtained by interpolation.

Fig. 3 illustrates the testing phase data augmentation. For a test image  $I_c$  and its target set  $\mathbf{I}_K^c$  sampled from the *training set*, we obtain a set of transformations  $\{\varphi_{\tilde{m}_0^c}^{-1}(t)\}$ . By virtue of the LDDMM model these transformations are invertible. For each  $\varphi_{\tilde{m}_0^c}^{-1}(t)$  we can therefore efficiently obtain the corresponding inverse map  $\varphi_{\tilde{m}_0^c}(t)$ . We denote our trained segmentation network by  $\mathcal{H} : \mathbb{R}^D \rightarrow \mathbb{R}^{D \times L}$  which takes an image as its input and predicts its segmentation labels. Here,  $L$  is the number of segmentation labels. Each prediction  $\mathcal{H}(I_c \circ \varphi_{\tilde{m}_0^c}^{-1}(t))$  is warped



**Fig. 3.** Illustration of the testing phase data augmentation. Given a source image  $I_c$  and a set of target images  $\mathbf{I}_K^c$ , a geodesic subspace is determined first. A set of transformations  $\varphi_{\tilde{m}_0}^{-1}(t)$  is then sampled from this space and, at the same time, the corresponding inverse transformations  $\varphi_{\tilde{m}_0^c}(t)$  are obtained. A segmentation network  $\mathcal{H}$  is applied to each warped image and the resulting segmentations  $\mathcal{H}(I_c \circ \varphi_{\tilde{m}_0^c}^{-1}(t))$  are warped back to the source image space. A label fusion strategy is applied to obtain the final segmentation.

back to the space of  $I_c$  via  $\mathcal{H}(I_c \circ \varphi_{\tilde{m}_0^c}^{-1}(t)) \circ \varphi_{\tilde{m}_0^c}(t)$ . The final segmentation is obtained by merging all warped predictions via a label fusion strategy.

**Dataset** The LONI Probabilistic Brain Atlas [25] (LPBA40) dataset contains volumes of 40 healthy patients with 56 manually annotated anatomical structures. We affinely register all images to a mean atlas [13], resample to isotropic spacing of 1 mm, crop them to  $196 \times 164 \times 196$  and intensity normalize them to  $[0, 1]$  via histogram equalization. We randomly take 25 patients for training, 10 patients for testing, and 5 patients for validation.

The Osteoarthritis Initiative [29] (OAI) provides manually labeled knee images with segmentations of femur and tibia as well as femoral and tibial cartilage [2]. We first affinely register all images to a mean atlas [13], resample them to isotropic spacing of 1 mm, and crop them to  $160 \times 200 \times 200$ . We randomly take 60 patients for training, 25 patients for validation, and 52 patients for testing.

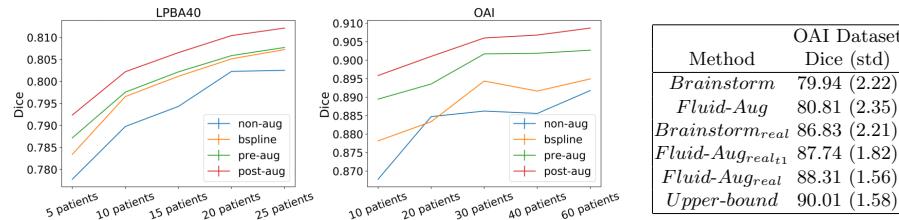
To evaluate the effect of data augmentation on training datasets with different sizes, we further sample 5, 10, 15, 20, 25 patients as the training set on LPBA40 and 10, 20, 30, 40, 60 patients as the training set for OAI.

**Metric** We use the average Dice score over segmentation classes for all tasks in Sec. 4.1 and Sec. 4.2.

**Baselines** *Non-augmentation* is our lower bound method. We use a class-balanced random cropping schedule during training [32]. We use this cropping schedule for all segmentation methods that we implement. We use a U-Net [23] segmentation network. *Random B-Spline Transform* is a transformation locally parameterized by randomizing the location of B-spline control points. Denote  $(\cdot, \cdot)$  as the number of control points distributed over a uniform mesh and the standard deviation of the normal distribution, units are in  $mm$ . The three set-

tings we use are  $(10^3, 3)$ ,  $(10^3, 4)$ ,  $(20^3, 2)$ . During data augmentation, we randomly select one of the settings to generate a new example.

**Settings** During the training augmentation phase (*pre-aug*), we randomly pick a source image and  $K$  targets, uniformly sample  $\lambda_i$  in Eq. 6 and then uniformly sample  $t$ . For LPBA40, we set  $K = 2$  and  $t \in [-1, 2]$ ; for the OAI data, we set  $K = 1$  and  $t \in [-1, 2]$ . For all training sets with different sizes, for both the B-Spline and the fluid-based augmentation methods and for both datasets, we augment the training data by 1,500 cases. During the testing augmentation phase (*post-aug*), for both datasets, we set  $K = 2$  and  $t \in [-1, 2]$  and draw 20 synthesized samples for each test image. The models trained via the augmented training set are used to predict the segmentations. To obtain the final segmentation, we sum the softmax outputs of all the segmentations warped to the original space and assign the label with the largest sum. We test using the models achieving the best performance on the validation set. We use the optimization approach in [17] and the network of [26,27] to compute the mappings  $\mathcal{M}$  on LPBA40 and OAI, respectively.



**Fig. 4.** Segmentation performance for segmentation tasks. The left two plots show Dice scores for the different methods with different training set sizes on the LPBA40 and OAI datasets for general segmentation. Performance increases with training set size. Fluid-based augmentation (pre-aug and post-aug) shows the best performance. The right table compares the performance for one-shot segmentation in Sec. 4.2. Fluid-based augmentation methods perform better than their *Brainstorm* counterparts.

**Results** Fig. 4 shows the segmentation performance on the LPBA40 and the OAI datasets. For training phase augmentation, fluid-based augmentation improves accuracy over non-augmentation and B-Spline augmentation by a large margin on the OAI dataset and results in comparable performance on the LPBA40 dataset. This difference might be due to the larger anatomical differences in the LPBA40 dataset compared to the OAI dataset; such large differences might not be well captured by inter- and extrapolation along a few geodesics. Hence, the OAI dataset may benefit more from the anatomically plausible geodesic space. When test phase augmentation is used in addition to training augmentation, performance is further improved. This shows that the ensemble strategy used by post-aug, where the segmentation network makes a consensus decision based on different views of the image to be segmented, is effective. In practice, we observe that high-quality inverse transformations (that map the segmentations

back to the test image space) are important to achieve good performance. These inverse transformations can efficiently be computed via Eq. 5 for our fluid-based approach.

## 4.2 Data augmentation for one-shot segmentation

We explore one-shot learning. Specifically, we consider single atlas medical image segmentation, where only the atlas image has a segmentation, while all other images are unlabeled. We first review *Brainstorm* [35], a competing data augmentation framework for one-shot segmentation. We then discuss our modifications.

In *Brainstorm*, the appearance of a sampled unlabeled image is first transferred to atlas-space and subsequently spatially transformed by registering to another sampled unlabeled image. Specifically, a registration network  $\mathcal{H}^r$  is trained to predict the displacement field between the atlas  $A$  and the unlabeled images. For a given image  $I_c$ , the predicted transformation to  $A$  is  $\varphi_c(x) = \mathcal{H}^r(I_c, A) + x$ . A set of approximated inverse transformations  $\{\varphi_c^{-1}, c \in 1 \dots N\}$  from the atlas to the image set can also be predicted by the network  $\mathcal{H}^r$ . These inverse transformations capture the anatomical diversity of the unlabeled set and are used to deform the images. Further, an appearance network  $\mathcal{H}^a$  is trained to capture the appearance of the unlabeled set. The network is designed to output the residue  $r$  between the warped image  $I_c \circ \varphi$  and the atlas,  $r_c = \mathcal{H}^a(A, I_c \circ \varphi_c)$ . Finally, we obtain a new set of segmented images by applying the transformations to the atlas with the new appearance:  $\{(A + r_i) \circ \varphi_j^{-1}, i, j \in 1 \dots N\}$ .

We modify the *Brainstorm* idea as follows: 1) Instead of sampling the transformation  $\varphi_c^{-1}$ , we sample  $\varphi_{\tilde{m}_0^c}^{-1}(t)$  based on our fluid-based approach; 2) We remove the appearance network and instead simply use  $\{I_c \circ \varphi_c, c \in 1 \dots N\}$  to model appearance. I.e., we retain the appearance of individual images, but deform them by going through atlas space. This results by construction in a realistic appearance distribution. Our synthesized images are  $\{(I_i \circ \varphi_i) \circ \varphi_{\tilde{m}_0^j}^{-1}(t), \tilde{m}_0^j \in C(M_K^j), t \in \mathbb{R}, i, j \in 1 \dots N\}$ . We refer to this approach as *Fluid-Aug<sub>real</sub>*.

**Dataset** We use the OAI dataset with 100 manually annotated images and a segmented mean atlas [13]. We only use the atlas segmentation for our one-shot segmentation experiments.

**Baseline** *Upper-bound* is a model trained from 100 images and their manual segmentations. We use the same U-net as for the general segmentation task in Sec. 4.1. *Brainstorm* is our baseline. We train a registration network and an appearance network separately, using the same network structures as in [35]. We sample a new training set of size 1,500 via random compositions of the appearance and the deformation. We also compare with a variant replacing the appearance network, where the synthesized set can be written as  $\{(I_i \circ \varphi_i) \circ \varphi_j^{-1}, i, j \in 1 \dots N\}$ . We refer to this approach as *Brainstorm<sub>real</sub>*.

**Settings** We set  $K = 2$  and  $t \in [-1, 2]$  and draw a new training set with 1,500 pairs the same way as in Sec. 4.1. We also compare with a variant where we set  $t = 1$  (instead of randomly sampling it), which we denote *Fluid-Aug<sub>real,t=1</sub>*. Further, we compare with a variant using the appearance network, where the

synthesized set is  $\{(A + r_i) \circ \varphi_{\tilde{m}_0^j}^{-1}(t), m_0^j \in C(M_K^j), t \in \mathbb{R}, i, j \in 1 \dots N\}$ . We refer to this approach as *Fluid-Aug*.

Fig. 4 shows better performance for fluid-based augmentation than for *Brain-storm* when using either real or learnt appearance. Furthermore, directly using the appearance of the unlabeled images shows better performance than using the appearance network. Lastly, randomizing over the location on the geodesic (*Fluid-Aug<sub>real</sub>*) shows small improvements over fixing  $t = 1$  (*Fluid-Aug<sub>real<sub>t=1</sub></sub>*).

## 5 Conclusion

We introduced a fluid-based method for medical image data augmentation. Our approach makes use of a geodesic subspace capturing anatomical variability. We explored its use for general segmentation and one-shot segmentation, achieving improvements over competing methods. Future work will focus on efficiency improvements. Specifically, computing the geodesic subspaces is costly if they are not approximated by a registration network. We will therefore explore introducing multiple atlases to reduce the number of possible registration pairs.

**Acknowledgements:** Research reported in this publication was supported by the National Institutes of Health (NIH) and the National Science Foundation (NSF) under award numbers NSF EECS1711776 and NIH 1R01AR072013. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the NSF.

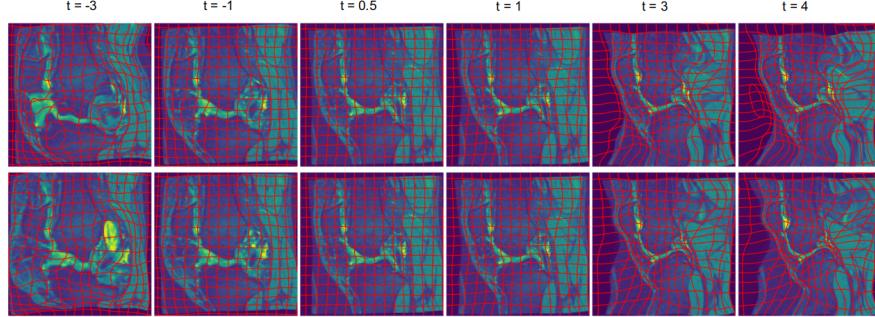
## References

1. Aljabar, P., Wolz, R., Rueckert, D.: Manifold learning for medical image registration, segmentation, and classification. In: Machine learning in computer-aided diagnosis: Medical imaging intelligence and analysis, pp. 351–372. IGI Global (2012)
2. Ambellan, F., Tack, A., Ehlke, M., Zachow, S.: Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative. *Medical image analysis* **52**, 109–118 (2019)
3. Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D.: Semi-supervised learning for network-based cardiac MR image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 253–260. Springer (2017)
4. Baur, C., Albarqouni, S., Navab, N.: Semi-supervised deep learning for fully convolutional networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 311–319. Springer (2017)
5. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* **61**(2), 139–157 (2005)
6. Chaitanya, K., Karani, N., Baumgartner, C.F., Becker, A., Donati, O., Konukoglu, E.: Semi-supervised and task-driven data augmentation. In: International Conference on Information Processing in Medical Imaging. pp. 29–41. Springer (2019)

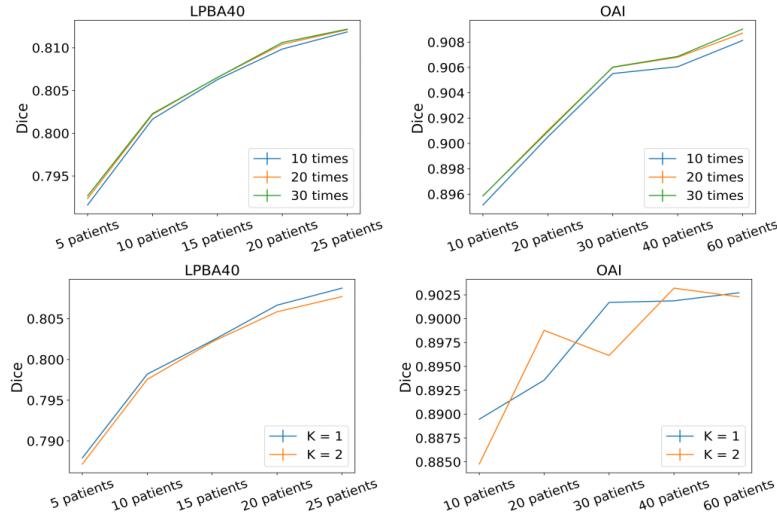
7. Chen, S., Bortsova, G., Juárez, A.G.U., van Tulder, G., de Bruijne, M.: Multi-task attention-based semi-supervised learning for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 457–465. Springer (2019)
8. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics* pp. 587–600 (1998)
9. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321**, 321–331 (2018)
10. Hauberg, S., Freifeld, O., Larsen, A.B.L., Fisher, J., Hansen, L.: Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. In: Artificial Intelligence and Statistics. pp. 342–350 (2016)
11. Heinrich, M.P., Oktay, O., Bouteldja, N.: Obelisk-one kernel to solve nearly everything: Unified 3d binary convolutions for image analysis (2018)
12. Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. In: AMIA Annual Symposium Proceedings. vol. 2017, p. 979. American Medical Informatics Association (2017)
13. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* **23**, S151–S160 (2004)
14. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(2), 236–250 (2005)
15. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on. pp. 565–571. IEEE (2016)
16. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 370–378. Springer (2018)
17. Niethammer, M., Kwitt, R., Vialard, F.X.: Metric learning for image registration. *CVPR* (2019)
18. Oliveira, A., Pereira, S., Silva, C.A.: Augmenting data when training a CNN for retinal vessel segmentation: How to warp? In: 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG). pp. 1–4. IEEE (2017)
19. Park, S., Thorpe, M.: Representing and learning high dimensional data with the optimal transport map from a probabilistic viewpoint. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7864–7872 (2018)
20. Paschali, M., Simson, W., Roy, A.G., Naeem, M.F., Göbl, R., Wachinger, C., Navab, N.: Data augmentation with manifold exploring geometric transformations for increased performance and robustness. *arXiv preprint arXiv:1901.04420* (2019)
21. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on medical imaging* **35**(5), 1240–1251 (2016)
22. Qiu, A., Younes, L., Miller, M.I.: Principal component based diffeomorphic surface mapping. *IEEE transactions on medical imaging* **31**(2), 302–311 (2011)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
24. Roth, H.R., Lee, C.T., Shin, H.C., Seff, A., Kim, L., Yao, J., Lu, L., Summers, R.M.: Anatomy-specific classification of medical images using deep convolutional

- nets. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). pp. 101–104. IEEE (2015)
25. Shattuck, D.W., Mirza, M., Adisetyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* **39**(3), 1064–1080 (2008)
  26. Shen, Z., Han, X., Xu, Z., Niethammer, M.: Networks for joint affine and non-parametric image registration. *CVPR* (2019)
  27. Shen, Z., Vialard, F.X., Niethammer, M.: Region-specific diffeomorphic metric mapping. In: *Advances in Neural Information Processing Systems*. pp. 1096–1106 (2019)
  28. Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: *International workshop on simulation and synthesis in medical imaging*. pp. 1–11. Springer (2018)
  29. The Osteoarthritis Initiative: Osteoarthritis initiative (OAI) dataset. <https://nda.nih.gov/oai/>
  30. Vakalopoulou, M., Chassagnon, G., Bus, N., Marini, R., Zacharaki, E.I., Revel, M.P., Paragios, N.: AtlasNet: multi-atlas non-linear deep networks for medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 658–666. Springer (2018)
  31. Xu, Z., Niethammer, M.: DeepAtlas: Joint semi-supervised learning of image registration and segmentation. *arXiv preprint arXiv:1904.08465* (2019)
  32. Xu, Z., Shen, Z., Niethammer, M.: Contextual additive networks to efficiently boost 3D image segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 92–100. Springer (2018)
  33. Younes, L., Arrate, F., Miller, M.I.: Evolutions equations in computational anatomy. *NeuroImage* **45**(1), S40–S50 (2009)
  34. Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: *International conference on information processing in medical imaging*. pp. 37–48. Springer (2013)
  35. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 8543–8553 (2019)
  36. Zhou, Y., He, X., Huang, L., Liu, L., Zhu, F., Cui, S., Shao, L.: Collaborative learning of semi-supervised segmentation and classification for medical images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2079–2088 (2019)

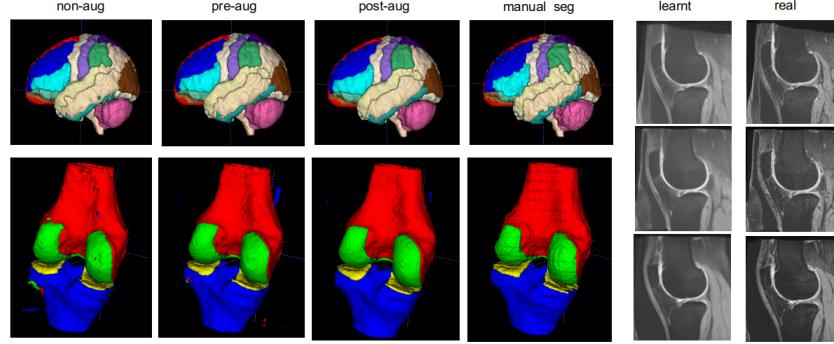
## 6 Supplementary Material



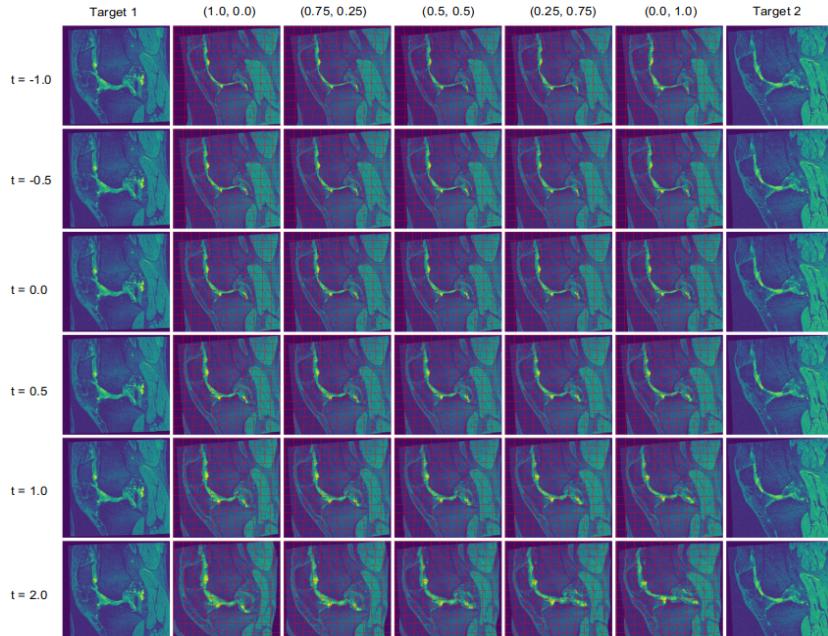
**Fig. 5.** Comparison between inter- and extrapolation of the displacement field (top row) and geodesic inter- and extrapolation (bottom row). We show the center slices from the sagittal view of the 3D MRI knee images. For both methods, we assume they have the same transformation  $\varphi^{-1}(1)$  at  $t = 1$ . Then we compute the displacement-field based inter- and extrapolation as  $\varphi_{affine}^{-1}(t, x) = (\varphi^{-1}(1, x) - x)t + x$ , whereas  $\varphi_{LDDMM}^{-1}$  is obtained via geodesic shooting (based on solving the EPDiff Eq. 4). For large deformations, i.e.,  $t = -3$  and  $t = 4$ , affine extrapolation results in foldings while extrapolation via the LDDMM geodesic results in diffeomorphic transformations.



**Fig. 6.** Ablation study on testing augmentation size (first row) and the choice of  $K$  (second row). For the first row, we evaluate segmentation accuracies for a different numbers of augmentation samples in the test phase.  $N$  times denotes that a testing image is deformed by  $N$  different random transformations drawn from the geodesic subspace. Segmentation accuracies start to saturate for  $N \geq 20$ . For the second row, we evaluate the performance for different choices of  $K$  (i.e., the dimensionality of the geodesic subspace). A larger  $K$  does not necessarily result in better segmentation accuracies.



**Fig. 7.** From the first to the fourth column, we visualize the segmentation results in Sec. 4 on LPBA40 (first row) and OAI (second row); from left to right: results without augmentation (non-aug), training phase augmentation (pre-aug), testing phase augmentation (post-aug), and the manual segmentations. We observe segmentation refinement after *pre-aug* and *post-aug*. For the last two columns, we compare the learnt appearance  $A + r_i$  (fifth column) and the real appearance  $I_i \circ \varphi_i$  (sixth column) in Sec. 4.2, where each row is a patient. The learnt appearance is smoother and hence shows less image texture. In our case, the segmentation network trained using the learnt appearance does not match the noisy testing data well.



**Fig. 8.** Visualization of synthesized images obtained from a source image and its  $K = 2$  geodesic subspace. From left to right: the first and the last column show the two target images. The remaining columns show synthesized images  $I_c \circ \varphi_{\tilde{m}_0^c}^{-1}(t)$ . The column index  $(\cdot, \cdot)$  denotes the mixture weights  $(\lambda_1, \lambda_2)$  in Eq. 6, determining how much target 1 and target 2 influence the overall deformation. From top to bottom: each row refers to the time  $t$  sampled along the geodesic path. For the row  $t = 0$ , the transformation is the identity, i.e.,  $I_c = I_c \circ \varphi_{\tilde{m}_0^c}^{-1}(0)$ . For the row  $t = 1$ , the warped image has similar anatomical structure as target 1 when  $(\lambda_1, \lambda_2) = (1.0, 0.0)$  while is similar to target 2 when  $(\lambda_1, \lambda_2) = (0.0, 1.0)$ . Columns  $(1.0, 0.0)$  and  $(0.0, 1.0)$  show samples on two geodesic paths ( $K = 1$ ) toward target 1 and target 2, respectively.