Aladdin: Joint Atlas Building and Diffeomorphic Registration Learning with Pairwise Alignment Supplementary Material

This supplementary material provides proofs and additional details for our approach. Specifically, Appx. A derives the Gâteaux variations of our atlas building models leading to the closed-form solution of Eqs. (2.3-2.4); Appx. B discusses the optimization-based atlas building and registration literature in detail; Appx. C presents the proof of the SVF-based Euler-Lagrange equations; Appx. D includes details about experimental settings and hyperparameter tuning.

A. Gâteaux variation of *backward* and *forward* atlas building models

As discussed in Sec. 2, we can deduct the closed-form solutions for the atlas in both *backward* and *forward* models via optimization theory. For simplicity, we define $\mathcal{L}_{sim}$ as the squared L_2 norm, i.e., $\mathcal{L}_{sim}(I,J) = \|I-J\|_2^2 = \langle I-J, I-J\rangle = \int_\Omega (I(x)-J(x))^2\,dx$, where Ω is the domain, $x\in\Omega$ is the position, and $\langle\cdot,\cdot\rangle$ is the usual L_2-product for square integrable vector-fields on Ω. Denote the energy functional of Eq. (2.1) as E_1 and Eq. (2.2) as E_2, i.e.,

$$E_1(\mathcal{I}) = \sum_{i=1}^N \int_\Omega (\mathcal{I}(x) - I_i\circ\Phi_{\alpha_i}^{-1}(x))^2\,dx + \lambda\mathcal{L}_{reg}(\Phi_{\alpha_i}^{-1})\,, \tag{A.1}$$

$$E_2(\mathcal{I}) = \sum_{i=1}^N \int_\Omega (\mathcal{I}\circ\Phi_{\beta_i}^{-1}(x) - I_i(x))^2\,dx + \lambda\mathcal{L}_{reg}(\Phi_{\beta_i}^{-1})\,. \tag{A.2}$$

By Gâteaux variation w.r.t. $\mathcal{I}$, we have

$$\delta E_1(\mathcal{I};d\mathcal{I}) = \frac{\partial}{\partial\epsilon}E_1(\mathcal{I}+\epsilon d\mathcal{I})|_{\epsilon=0} \tag{A.3}$$

$$= \frac{\partial}{\partial\epsilon}\sum_{i=1}^N\Big(\int_\Omega(\mathcal{I}(x)+\epsilon d\mathcal{I}(x) - I_i\circ\Phi_{\alpha_i}^{-1}(x))^2\,dx + \mathcal{L}_{reg}(\Phi_{\alpha_i}^{-1})\Big)|_{\epsilon=0} \tag{A.4}$$

$$= \sum_{i=1}^N\int_\Omega 2\Big(\mathcal{I}(x) - I_i\circ\Phi_{\alpha_i}^{-1}(x)\Big)d\mathcal{I}(x)\,dx \tag{A.5}$$

$$= 2\langle\sum_{i=1}^N \mathcal{I} - I_i\circ\Phi_{\alpha_i}^{-1}, d\mathcal{I}\rangle \stackrel{!}{=} 0\,,\ \forall d\mathcal{I}\,, \tag{A.6}$$

$$\delta E_2(\mathcal{I};d\mathcal{I}) = \frac{\partial}{\partial\epsilon}E_2(\mathcal{I}+\epsilon d\mathcal{I})|_{\epsilon=0} \tag{A.7}$$

$$= \frac{\partial}{\partial\epsilon}\sum_{i=1}^N\Big(\int_\Omega((\mathcal{I}+\epsilon d\mathcal{I})\circ\Phi_{\beta_i}^{-1}(x) - I_i(x))^2\,dx + \mathcal{L}_{reg}(\Phi_{\beta_i}^{-1})\Big)|_{\epsilon=0} \tag{A.8}$$

$$= \sum_{i=1}^N\int_\Omega 2\Big((\mathcal{I}\circ\Phi_{\beta_i}^{-1}(x) - I_i(x))\Big)d\mathcal{I}\circ\Phi_{\beta_i}^{-1}(x)\,dx \tag{A.9}$$

$$= 2\sum_{i=1}^N\int_\Omega \Big(\mathcal{I}\circ\Phi_{\beta_i}^{-1}(x) - I_i\circ\Phi_{\beta_i}\circ\Phi_{\beta_i}^{-1}(x)\Big)d\mathcal{I}\circ\Phi_{\beta_i}^{-1}(x)\,dx \tag{A.10}$$

$$\stackrel{(a)}{=} 2\sum_{i=1}^N\int_{\Omega'} \Big(\mathcal{I}(y) - I_i\circ\Phi_{\beta_i}(y)\Big)d\mathcal{I}(y)|D\Phi_{\beta_i}(y)|\,dy \tag{A.11}$$

$$\stackrel{(b)}{=} 2\sum_{i=1}^N\int_{\Omega} \Big(\mathcal{I}(y) - I_i\circ\Phi_{\beta_i}(y)\Big)|D\Phi_{\beta_i}(y)|d\mathcal{I}(y)\,dy \tag{A.12}$$

$$= 2\langle\sum_{i=1}^N(\mathcal{I} - I_i\circ\Phi_{\beta_i})|D\Phi_{\beta_i}|, d\mathcal{I}\rangle \stackrel{!}{=} 0\,,\ \forall d\mathcal{I}\,, \tag{A.13}$$

where (a) corresponds to a change of variables: i.e., setting $\Phi_{\beta_i}^{-1}(x) = y$ or equivalently $\Phi_{\beta_i}(y) = x$, which results in the Jacobian change of variables $|D\Phi_{\beta_i}(y)|dy = dx$. For (b), the transformation Φ_{β}^{-1}: $\Omega \to \Omega$, where $\Omega \subseteq \mathbb{R}^d$ ($d = 2$ for 2D or $d = 3$ for 3D). Hence, changing variables does not change the domain, i.e. $\Omega^{'} = \Omega$.

B. Related work: optimization-based atlas building and registration approaches

This section briefly introduces optimization-based atlas building and registration approaches.

Optimization-based *backward* atlas building and registration models: Bhatia *et al*. [8] build an atlas by forcing the sum of deformations from all images to be zero and showed good performance in a small deformation setting. To address image datasets with large deformations, Lorenzen *et al*. [29] and Avants *et al*. [4] use an unbiased atlas construction scheme in the space of diffeomorphisms via the large deformation diffeomorphic metric mapping (LDDMM) model [6]. As an extension, Lorenzen *et al*. [30] extended their work to multi-modal image set registration and multi-class atlas formation by minimizing the Kullback-Leibler divergence between the estimated posteriors in a Bayesian framework. Bhatia *et al*. [7] introduced an iterative Expectation-Maximization (EM) framework to simultaneously improve both the alignment of images to their average image, as well as the segmentation of structures in the average space. Van *et al*. [44] use mesh-based representations to generalize a probabilistic atlas building approach to a joint registration and atlas estimation Bayesian inference model, which automatically determines the optimal amount of spatial blurring, the best deformation field flexibility, and the most compact mesh representation. Fletcher *et al*. [20] deveoped a robust brain atlas estimation technique based on the geometric median in the LDDMM framework. To avoid building a fuzzy atlas, Wu *et al*. [49, 47] proposed to average the aligned images according to anatomical shape and distances of local patches, instead of direcly using an intensity average. Wang *et al*. [46] decompose a large-scale groupwise registration problem into a series of small-scale problems, which are easier to solve and thus help registration robustness. Jia *et al*. [26] proposed a hierarchical groupwise registration framework termed ABSORB, which bundles similar images thus reducing registration errors and generating smooth registration paths. Debroux *et al*. [13] proposed a variational model for joint segmentation, registration and atlas generation.

Optimization-based *forward* atlas building and registration models: The *forward* atlas building model is motivated by the notion of deformable templates introduced in [22]. Given a template image I, an entire family of new images with similar anatomical structures is modeled as the orbit of a group of diffeomorphisms $\mathcal{G}$, i.e., $Orb(I) = \{I \circ \phi^{-1} : \phi \in \mathcal{G}\}$. Ma *et al*. [31, 32] formulated atlas building in a Bayesian framework and use Mode Approximation of the EM algorithm to build the atlas. Durrleman *et al*. [17] apply a forward model to build atlases for curves and surfaces. Zhang *et al*. [53] proposed a generative model to jointly estimate the registration regularity, noise variance, and the atlas. Singh *et al*. [43] use a vector initial momentum parameterization of diffeomorphisms for atlas construction. To accommodate complex structural differences across a heterogeneous group of images, Zhang *et al*. [52] proposed a generative Gaussian mixture model for diffeomorphic multi-atlas building.

C. SVF based Euler-Lagrange equation

This section discusses the derivation of the Euler-Lagrange equations of the regularized stationary velocity field *forward* atlas building model (i.e. Eq. (3.4)).

Lemma 1. *The variation of transformation $\Phi_{s,t}^v$ when v is perturbed along h is given by:*

$$\begin{aligned}\partial_h \Phi_{s,t}^v &= \lim_{\epsilon \to 0} \frac{\Phi_{s,t}^{v+\epsilon h} - \Phi_{s,t}^v}{\epsilon} \\ &= D\Phi_{s,t}^v \int_s^t (D\phi_{s,u}^v)^{-1} h \circ \Phi_{s,u}^v \, du \,.\end{aligned} \tag{C.1}$$

Proof. See [6]. □

Theorem 1. *Given a continuous differentiable idealized atlas image $\mathcal{I}$ and a population of noisy observed anatomies $\{I_i\}$ ($i = 1, ..., N$), the $\{v_i^*\}$ minimizing the following energy functional*

$$E_0(\{v_i\}) = \sum_{i=1}^{N} \left[\lambda Reg(v_i) + \|\mathcal{I} \circ \Phi_{1,0}^{v_i} - I_i\|_2^2\right] \tag{C.2}$$

satisfy the Euler-Lagrange equation

$$\lambda \nabla_{v_i^*} \mathit{Reg}(v_i^*) - 2\int_0^1 \Big(|D\Phi_{t,1}^{v_i^*}|(\mathcal{I} \circ \Phi_{t,0}^{v_i^*} - I_i \circ \Phi_{t,1}^{v_i^*})\nabla(\mathcal{I} \circ \Phi_{t,0}^{v_i^*}) \Big)\, dt = 0, \forall i\,. \tag{C.3}$$

Proof. The Euler-Lagrange equation associated to energy functional $E_0(\{v_i\})$ is obtained by setting the Fréchet derivative $\nabla_{v_i} E_0(\{v_i\})$ to zero for each i. Let the velocity v_i be perturbed by an ϵ amount along direction h_i. The Fréchet derivative $\nabla_{v_i} E_0(\{v_i\})$ can be computed from the Gâteaux variation $\partial_{h_i} E_0(\{v_i\})$ by

$$\begin{aligned} \partial_{h_i} E_0(\{v_i\}) &= \lim_{\epsilon \to 0} \frac{E_0(\{v_i + \epsilon h_i\}) - E_0(\{v_i\})}{\epsilon} \\ &= \langle \nabla_{v_i} E_0, h_i \rangle\,. \end{aligned} \tag{C.4}$$

Denote $E_{reg}(v_i) = \mathrm{Reg}(v_i)$ and $E_{sim}(v_i) = \|\mathcal{I} \circ \Phi_{1,0}^{v_i} - I_i\|_2^2$, then we have $\partial_{h_i} E_0(\{v_i\}) = \lambda \partial_{h_i} E_{reg}(v_i) + \partial_{h_i} E_{sim}(v_i)$.

The variation of $E_{reg}(v_i)$ is

$$\partial_{h_i} E_{reg}(v_i) = \langle \nabla_{v_i} \mathrm{Reg}(v_i), h_i \rangle\,.$$

Similar to [6], the variation of $E_{sim}(v_i)$ is

$$\begin{aligned} \partial_{h_i} E_{sim}(v_i) &= 2\Big\langle \mathcal{I} \circ \Phi_{1,0}^{v_i} - I_i, D\mathcal{I} \circ \Phi_{1,0}^{v_i} \partial_{h_i} \Phi_{1,0}^{v_i} \Big\rangle \\ &\overset{(a)}{=} 2\Big\langle \mathcal{I} \circ \Phi_{1,0}^{v_i} - I_i, D\mathcal{I} \circ \Phi_{1,0}^{v_i} \times \Big(-D\Phi_{1,0}^{v_i} \int_0^1 (D\Phi_{1,t}^{v_i})^{-1} h_i \circ \Phi_{1,0}^{v_i}\, dt \Big) \Big\rangle \\ &\overset{(b)}{=} -2\int_0^1 \big\langle \mathcal{I} \circ \Phi_{1,0}^{v_i} - I_i, D(\mathcal{I} \circ \Phi_{1,0}^{v_i}) \times (D\Phi_{1,t}^{v_i})^{-1} h_i \circ \Phi_{1,t}^{v_i} \big\rangle\, dt \\ &= -2\int_0^1 \big\langle (\mathcal{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \circ \Phi_{1,t}^{v_i}, D(\mathcal{I} \circ \Phi_{t,0}^{v_i} \circ \Phi_{1,t}^{v_i}) \times (D\Phi_{1,t}^{v_i})^{-1} h_i \circ \Phi_{1,t}^{v_i} \big\rangle\, dt \\ &\overset{(c)}{=} -2\int_0^1 \big\langle |D\Phi_{t,1}^{v_i}|(\mathcal{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}), D(\mathcal{I} \circ \Phi_{t,0}^{v_i}) h_i \big\rangle\, dt \\ &\overset{(d)}{=} -2\int_0^1 \big\langle |D\Phi_{t,1}^{v_i}|(\mathcal{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla(\mathcal{I} \circ \Phi_{t,0}^{v_i}), h_i \big\rangle\, dt \\ &= \Big\langle -2\int_0^1 \Big(|D\Phi_{t,1}^{v_i}|(\mathcal{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla(\mathcal{I} \circ \Phi_{t,0}^{v_i}) \Big)\, dt, h_i \Big\rangle \end{aligned}$$

where (a) refers to substituting $\partial_{h_i} \Phi_{1,0}^{v_i}$ based on Lemma 1; (b) refers to rewriting $D(\mathcal{I} \circ \Phi_{1,0}^{v_i}) = D\mathcal{I} \circ \Phi_{1,0}^{v_i} D\Phi_{1,0}^{v_i}$; (c) is the chain rule $(D(\mathcal{I} \circ \Phi_{t,0}^{v_i} \circ \Phi_{1,t}^{v_i}) = D_{\Phi_{1,t}^{v_i}}(\mathcal{I} \circ \Phi_{t,0}^{v_i}) D\Phi_{1,t}^{v_i})$ and the Jacobian change of variables. Denote $\Phi_{1,t}^{v_i}(x) = y$, then $\Phi_{t,1}^{v_i}(y) = x$ and the Jacobian change of variables is $|D\Phi_{t,1}^{v_i}(y)| dy = dx$; (d) follows from writing the transpose and changing the notation $\nabla(\mathcal{I} \circ \Phi_{t,0}^{v_i}) = D(\mathcal{I} \circ \Phi_{t,0}^{v_i})^T$.

Collecting terms, the Fréchet derivative $\nabla_{v_i} E_0(\{v_i\})$ is

$$\nabla_{v_i} E_0(\{v_i\}) = \lambda \nabla_{v_i} \mathrm{Reg}(v_i) - 2\int_0^1 \Big(|D\Phi_{t,1}^{v_i}|(\mathcal{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla(\mathcal{I} \circ \Phi_{t,0}^{v_i}) \Big)\, dt\,. \tag{C.5}$$

This equation yields the Euler-Lagrange Eq. (C.3). □

Theorem 2. *Given a continuous differentiable idealized atlas image $\mathcal{I}$ and a population of noisy observed anatomies $\{I_i\}$ $(i = 1, ..., N)$, the $\{v_i^*\}$ minimizing the following energy functional*

$$\begin{aligned} E_1(\{v_i\}) = \sum_{(i,j) \in \Gamma} \Big[&\lambda \mathit{Reg}(v_i) + \|\mathcal{I} \circ \Phi_{1,0}^{v_i} - I_i\|_2^2 + \lambda \mathit{Reg}(v_j) + \|\mathcal{I} \circ \Phi_{1,0}^{v_j} - I_j\|_2^2 \\ &+ \gamma_1 \|I_i \circ \Phi_{0,1}^{v_i} - I_j \circ \Phi_{0,1}^{v_j}\|_2^2 \\ &+ \gamma_2 \Big(\|I_i \circ \Phi_{0,1}^{v_i} \circ \Phi_{1,0}^{v_j} - I_j\|_2^2 + \|I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{1,0}^{v_i} - I_i\|_2^2 \Big) \Big] \end{aligned} \tag{C.6}$$

satisfy the Euler-Lagrange equation

$$
\begin{aligned}
(N-1)\lambda\nabla_{v_i^*}\textit{Reg}(v_i^*) &- 2(N-1)\int_0^1 \Big(|D\Phi_{t,1}^{v_i^*}|(\mathcal{I}\circ\Phi_{t,0}^{v_i^*} - I_i\circ\Phi_{t,1}^{v_i^*})\nabla(\mathcal{I}\circ\Phi_{t,0}^{v_i^*})\Big)\, dt \\
&- 2N\gamma_1\int_0^1 \big(|D\Phi_{t,0}^{v_i^*}|(\frac{\sum_{j=1}^N I_i\circ\Phi_{0,1}^{v_j^*}}{N}\circ\Phi_{t,0}^{v_i^*} - I_i\circ\Phi_{t,1}^{v_i^*})\nabla(I_i\circ\Phi_{t,1}^{v_i^*})\big)\, dt \\
&- 2\gamma_2\int_0^1 \Big((\sum_{j=1}^N |D\Phi_{0,1}^{v_j^*}|)|D\Phi_{t,0}^{v_i^*}|(\frac{\sum_{j=1}^N |D\Phi_{0,1}^{v_j^*}| I_j\circ\Phi_{0,1}^{v_j^*}}{\sum_{j=1}^N |D\Phi_{0,1}^{v_j^*}|}\circ\Phi_{t,0}^{v_i^*} - I_i\circ\Phi_{t,1}^{v_i^*})\nabla(I_i\circ\Phi_{t,1}^{v_i^*})\Big)\, dt \\
&- 2\gamma_2\int_0^1 \sum_{j=1}^N |D\Phi_{t,1}^{v_i^*}|(I_j\circ\Phi_{0,1}^{v_j^*}\circ\Phi_{t,0}^{v_i^*} - I_i\circ\Phi_{t,1}^{v_i^*})\nabla(I_j\circ\Phi_{0,1}^{v_j^*}\circ\Phi_{t,0}^{v_i^*})\, dt = 0\,, \forall i\,.
\end{aligned}
\tag{C.7}
$$

Proof. Similar to the proof of Theorem 1, we need to calculate the Fréchet derivative $\nabla_{v_i}E_1(\{v_i\})$. Denote $E_{pair}^{atlas}(v_i,v_j) = \sum_{(i,j)\in\Gamma}\|I_i\circ\Phi_{0,1}^{v_i} - I_j\circ\Phi_{0,1}^{v_j}\|_2^2$, $E_{pair}^{image}(v_i,v_j) = \sum_{(i,j)\in\Gamma}\|I_i\circ\Phi_{0,1}^{v_i}\circ\Phi_{1,0}^{v_j} - I_j\|_2^2$ and $E_{pair}^{image}(v_j,v_i) = \sum_{(i,j)\in\Gamma}\|I_j\circ\Phi_{0,1}^{v_j}\circ\Phi_{1,0}^{v_i} - I_i\|_2^2$, then we have $\partial_{h_i}E_1(\{v_i\}) = (N-1)\lambda\partial_{h_i}E_{reg}(v_i) + (N-1)\partial_{h_i}E_{sim}(v_i) + \gamma_1\partial_{h_i}E_{pair}^{atlas}(v_i,v_j) + \gamma_2\partial_{h_i}E_{pair}^{image}(v_i,v_j) + \gamma_2\partial_{h_i}E_{pair}^{image}(v_j,v_i)$.

For conciseness, we ignore the calculation of $\partial_{h_i}E_{reg}(v_i)$ and $\partial_{h_i}E_{sim}(v_i)$, as they have been obtained in the proof of Theorem 1.

The variation of $E_{pair}^{atlas}(v_i,v_j)$ is

$$
\begin{aligned}
\partial_{h_i}E_{pair}^{atlas}(v_i,v_j) &= \sum_{(i,j)\in\Gamma} 2\big\langle I_i\circ\Phi_{0,1}^{v_i} - I_j\circ\Phi_{0,1}^{v_j}, DI_i\circ\Phi_{0,1}^{v_i}\partial_{h_i}\Phi_{0,1}^{v_i}\big\rangle \\
&= 2(N-1)\big\langle I_i\circ\Phi_{0,1}^{v_i} - \frac{\sum_{j=1,j\neq i}^N I_j\circ\Phi_{0,1}^{v_j}}{N-1}, DI_i\circ\Phi_{0,1}^{v_i}\partial_{h_i}\Phi_{0,1}^{v_i}\big\rangle \\
&= 2N\big\langle I_i\circ\Phi_{0,1}^{v_i} - \frac{\sum_{j=1}^N I_j\circ\Phi_{0,1}^{v_j}}{N}, DI_i\circ\Phi_{0,1}^{v_i}\partial_{h_i}\Phi_{0,1}^{v_i}\big\rangle \\
&\overset{(a)}{=} 2N\big\langle I_i\circ\Phi_{0,1}^{v_i} - \hat{I}, DI_i\circ\Phi_{0,1}^{v_i}\partial_{h_i}\Phi_{0,1}^{v_i}\big\rangle \\
&\overset{(b)}{=} 2N\Big\langle I_i\circ\Phi_{0,1}^{v_i} - \hat{I}, DI_i\circ\Phi_{0,1}^{v_i}\times\Big(D\Phi_{0,1}^{v_i}\int_0^1 (D\Phi_{0,t}^{v_i})^{-1}h_i\circ\Phi_{0,t}^{v_i}\, dt\Big)\Big\rangle \\
&\overset{(c)}{=} 2N\int_0^1 \big\langle I_i\circ\Phi_{0,1}^{v_i} - \hat{I}, D(I_i\circ\Phi_{0,1}^{v_i})\times(D\Phi_{0,t}^{v_i})^{-1}h_i\circ\Phi_{0,t}^{v_i}\big\rangle\, dt \\
&= 2N\int_0^1 \big\langle (I_i\circ\Phi_{t,1}^{v_i} - \hat{I}\circ\Phi_{t,0}^{v_i})\circ\Phi_{0,t}^{v_i}, D(I_i\circ\Phi_{t,1}^{v_i}\circ\Phi_{0,t}^{v_i})\times(D\Phi_{0,t}^{v_i})^{-1}h_i\circ\Phi_{0,t}^{v_i}\big\rangle\, dt \\
&\overset{(d)}{=} 2N\int_0^1 \big\langle |D\Phi_{t,0}^{v_i}|(I_i\circ\Phi_{t,1}^{v_i} - \hat{I}\circ\Phi_{t,0}^{v_i}), D(I_i\circ\Phi_{t,1}^{v_i})h_i\big\rangle\, dt \\
&\overset{(e)}{=} 2N\int_0^1 \big\langle |D\Phi_{t,0}^{v_i}|(I_i\circ\Phi_{t,1}^{v_i} - \hat{I}\circ\Phi_{t,0}^{v_i})\nabla(I_i\circ\Phi_{t,1}^{v_i}), h_i\big\rangle\, dt \\
&= \Big\langle -2N\int_0^1 \big(|D\Phi_{t,0}^{v_i}|(\hat{I}\circ\Phi_{t,0}^{v_i} - I_i\circ\Phi_{t,1}^{v_i})\nabla(I_i\circ\Phi_{t,1}^{v_i})\big)\, dt, h_i\Big\rangle
\end{aligned}
$$

where (a) is using $\hat{I}$ to represent $\frac{\sum_{j=1}^N I_j\circ\Phi_{0,1}^{v_j}}{N}$; (b) corresponds to substituting $\partial_{h_i}\Phi_{0,1}^{v_i}$ based on Lemma 1; (c) amounts to rewriting $D(I_i\circ\Phi_{0,1}^{v_i}) = DI_i\circ\Phi_{0,1}^{v_i}D\Phi_{0,1}^{v_i}$; (d) is the chain rule $(D(I_i\circ\Phi_{t,1}^{v_i}\circ\Phi_{0,t}^{v_i}) = D_{\Phi_{0,t}^{v_i}}(I_i\circ\Phi_{t,1}^{v_i})D\Phi_{0,t}^{v_i})$ the Jacobian change of variables. Denote $\Phi_{0,t}^{v_i}(x) = y$, then $\Phi_{t,0}^{v_i}(y) = x$ and the Jacobian change of variables is $|D\Phi_{t,0}^{v_i}(y)|dy = dx$; (e) results from writing the transpose and changing the notation $\nabla(I_i\circ\Phi_{t,1}^{v_i}) = D(I_i\circ\Phi_{t,1}^{v_i})^T$.

Before calculating the variation of $E^{image}_{pair}(v_i, v_j)$, we first reformat the equation

$$\begin{aligned}E^{image}_{pair}(v_i, v_j) &= \|I_i \circ \Phi^{v_i}_{0,1} \circ \Phi^{v_j}_{1,0} - I_j\|^2_2 \\ &= \big\langle I_i \circ \Phi^{v_i}_{0,1} \circ \Phi^{v_j}_{1,0} - I_j, I_i \circ \Phi^{v_i}_{0,1} \circ \Phi^{v_j}_{1,0} - I_j\big\rangle \\ &= \big\langle (I_i \circ \Phi^{v_i}_{0,1} - I_j \circ \Phi^{v_j}_{0,1}) \circ \Phi^{v_j}_{1,0}, (I_i \circ \Phi^{v_i}_{0,1} - I_j \circ \Phi^{v_j}_{0,1}) \circ \Phi^{v_j}_{1,0}\big\rangle \\ &\overset{(a)}{=} \big\langle |D\Phi^{v_j}_{0,1}|(I_i \circ \Phi^{v_i}_{0,1} - I_j \circ \Phi^{v_j}_{0,1}), (I_i \circ \Phi^{v_i}_{0,1} - I_j \circ \Phi^{v_j}_{0,1})\big\rangle\end{aligned}$$

where (a) is the Jacobian change of variables. Denote $\Phi^{v_j}_{1,0}(x) = y$, then $\Phi^{v_j}_{0,1}(y) = x$ and the Jacobian change of variables is $|D\Phi^{v_j}_{0,1}(y)|dy = dx$. Hence the variation of $E^{image}_{pair}(v_i, v_j)$ is

$$\begin{aligned}\partial_{h_i} E^{image}_{pair}(v_i, v_j) &= 2 \sum_{(i,j)\in\Gamma} \big\langle |D\Phi^{v_j}_{0,1}|(I_i \circ \Phi^{v_i}_{0,1} - I_j \circ \Phi^{v_j}_{0,1}), DI_i \circ \Phi^{v_i}_{0,1} \partial_{h_i} \Phi^{v_i}_{0,1}\big\rangle \\ &= 2\Big\langle \Big(\sum_{j=1, j\neq i}^{N} |D\Phi^{v_j}_{0,1}|\Big)\Big(I_i \circ \Phi^{v_i}_{0,1} - \frac{\sum_{j=1,j\neq i}^{N} |D\Phi^{v_j}_{0,1}| I_j \circ \Phi^{v_j}_{0,1}}{\sum_{j=1,j\neq i}^{N} |D\Phi^{v_j}_{0,1}|}\Big), DI_i \circ \Phi^{v_i}_{0,1} \partial_{h_i} \Phi^{v_i}_{0,1}\Big\rangle \\ &= 2\Big\langle \Big(\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|\Big)\Big(I_i \circ \Phi^{v_i}_{0,1} - \frac{\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}| I_j \circ \Phi^{v_j}_{0,1}}{\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|}\Big), DI_i \circ \Phi^{v_i}_{0,1} \partial_{h_i} \Phi^{v_i}_{0,1}\Big\rangle \\ &\overset{(a)}{=} 2\Big\langle \Big(\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|\Big)(I_i \circ \Phi^{v_i}_{0,1} - \tilde{I}), DI_i \circ \Phi^{v_i}_{0,1} \partial_{h_i} \Phi^{v_i}_{0,1}\Big\rangle \\ &\overset{(b)}{=} 2\Big\langle \Big(\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|\Big)(I_i \circ \Phi^{v_i}_{0,1} - \tilde{I}), DI_i \circ \Phi^{v_i}_{0,1} \times \Big(D\Phi^{v_i}_{0,1} \int_0^1 (D\Phi^{v_i}_{0,t})^{-1} h_i \circ \Phi^{v_i}_{0,t}\, dt\Big)\Big\rangle \\ &\overset{(c)}{=} 2 \int_0^1 \big\langle \Big(\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|\Big)(I_i \circ \Phi^{v_i}_{0,1} - \tilde{I}), D(I_i \circ \Phi^{v_i}_{0,1}) \times (D\Phi^{v_i}_{0,t})^{-1} h_i \circ \Phi^{v_i}_{0,t}\big\rangle\, dt \\ &= 2 \int_0^1 \big\langle \Big(\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|\Big)(I_i \circ \Phi^{v_i}_{t,1} - \tilde{I} \circ \Phi^{v_i}_{t,0}) \circ \Phi^{v_i}_{0,t}, D(I_i \circ \Phi^{v_i}_{t,1} \circ \Phi^{v_i}_{0,t}) \times (D\Phi^{v_i}_{0,t})^{-1} h_i \circ \Phi^{v_i}_{0,t}\big\rangle\, dt \\ &\overset{(d)}{=} 2 \int_0^1 \big\langle \Big(\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|\Big)|D\Phi^{v_i}_{t,0}|(I_i \circ \Phi^{v_i}_{t,1} - \tilde{I} \circ \Phi^{v_i}_{t,0}), D(I_i \circ \Phi^{v_i}_{t,1}) h_i\big\rangle\, dt \\ &\overset{(e)}{=} 2 \int_0^1 \big\langle \Big(\sum_{j}^{N} |D\Phi^{v_j}_{0,1}|\Big)|D\Phi^{v_i}_{t,0}|(I_i \circ \Phi^{v_i}_{t,1} - \tilde{I} \circ \Phi^{v_i}_{t,0}) \nabla(I_i \circ \Phi^{v_i}_{t,1}), h_i\big\rangle\, dt \\ &= \Big\langle -2 \int_0^1 \Big(\Big(\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|\Big)|D\Phi^{v_i}_{t,0}|(\tilde{I} \circ \Phi^{v_i}_{t,0} - I_i \circ \Phi^{v_i}_{t,1}) \nabla(I_i \circ \Phi^{v_i}_{t,1})\Big)\, dt, h_i\Big\rangle\end{aligned}$$

where (a) is using $\tilde{I}$ to represent $\frac{\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}| I_j \circ \Phi^{v_j}_{0,1})}{\sum_{j=1}^{N} |D\Phi^{v_j}_{0,1}|}$; (b) refers to substituting $\partial_{h_i} \Phi^{v_i}_{0,1}$ based on Lemma 1; (c) amounts to rewriting $D(I_i \circ \Phi^{v_i}_{0,1}) = DI_i \circ \Phi^{v_i}_{0,1} D\Phi^{v_i}_{0,1}$; (d) is the chain rule ($D(I_i \circ \Phi^{v_i}_{t,1} \circ \Phi^{v_i}_{0,t}) = D_{\Phi^{v_i}_{0,t}}(I_i \circ \Phi^{v_i}_{t,1}) D\Phi^{v_i}_{0,t}$) and the Jacobian change of variables. Denote $\Phi^{v_i}_{0,t}(x) = y$, then $\Phi^{v_i}_{t,0}(y) = x$ and the Jacobian change of variables is $|D\Phi^{v_i}_{t,0}(y)|dy = dx$; (e) refers to writing the transpose and changing the notation $\nabla(I_i \circ \Phi^{v_i}_{t,1}) = D(I_i \circ \Phi^{v_i}_{t,1})^T$.

The variation of $E_{pair}^{image}(v_j, v_i)$ is

$$
\begin{aligned}
\partial_{h_i} E_{pair}^{image}(v_j, v_i) &= \sum_{(i,j)\in\Gamma} 2\Big\langle I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{1,0}^{v_i} - I_i, D(I_j \circ \Phi_{0,1}^{v_j}) \circ \Phi_{1,0}^{v_i} \partial_{h_i} \Phi_{1,0}^{v_i} \Big\rangle \\
&\overset{(a)}{=} \sum_{j=1, j\neq i}^{N} 2\Big\langle I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{1,0}^{v_i} - I_i, D(I_j \circ \Phi_{0,1}^{v_j}) \circ \Phi_{1,0}^{v_i} \times \Big(- D\Phi_{1,0}^{v_i} \int_0^1 (D\Phi_{1,t}^{v_i})^{-1} h_i \circ \Phi_{1,t}^{v_i}\, dt \Big) \Big\rangle \\
&= \sum_{j=1}^{N} 2\Big\langle I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{1,0}^{v_i} - I_i, D(I_j \circ \Phi_{0,1}^{v_j}) \circ \Phi_{1,0}^{v_i} \times \Big(- D\Phi_{1,0}^{v_i} \int_0^1 (D\Phi_{1,t}^{v_i})^{-1} h_i \circ \Phi_{1,t}^{v_i}\, dt \Big) \Big\rangle \\
&= -2 \sum_{j=1}^{N} \int_0^1 \big\langle I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{1,0}^{v_i} - I_i, D(I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{1,0}^{v_i}) \times (D\Phi_{1,t}^{v_i})^{-1} h_i \circ \Phi_{1,t}^{v_i} \big\rangle\, dt \\
&\overset{(b)}{=} -2 \sum_{j=1}^{N} \int_0^1 \big\langle (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \circ \Phi_{1,t}^{v_i}, D(I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i} \circ \Phi_{1,t}^{v_i}) \times (D\Phi_{1,t}^{v_i})^{-1} h_i \circ \Phi_{1,t}^{v_i} \big\rangle \\
&\overset{(c)}{=} -2 \sum_{j=1}^{N} \int_0^1 \big\langle |D\Phi_{t,1}^{v_i}| (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}), D(I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i}) h_i \big\rangle \\
&\overset{(d)}{=} -2 \sum_{j=1}^{N} \int_0^1 \big\langle |D\Phi_{t,1}^{v_i}| (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i}), h_i \big\rangle \\
&= \Big\langle -2 \int_0^1 \sum_{j=1}^{N} |D\Phi_{t,1}^{v_i}| (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i})\, dt, h_i \Big\rangle
\end{aligned}
$$

where (a) results from substituting $\partial_{h_i} \Phi_{1,0}^{v_i}$ based on Lemma 1; (b) amounts to rewriting $D(I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{1,0}^{v_i}) = D(I_j \circ \Phi_{0,1}^{v_j}) \circ \Phi_{1,0}^{v_i} D\Phi_{1,0}^{v_i}$; (c) is the chain rule $(D(I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i} \circ \Phi_{1,t}^{v_i}) = D_{\Phi_{1,t}^{v_i}}(I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i}) D\Phi_{1,t}^{v_i})$ and the Jacobian change of variables. Denote $\Phi_{1,t}^{v_i}(x) = y$, then $\Phi_{t,1}^{v_i}(y) = x$ and the Jacobian change of variables is $|D\Phi_{t,1}^{v_i}(y)| dy = dx$; (d) is the result of writing the transpose and changing the notation $\nabla (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i}) = D(I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i})^T$.

Collecting terms, the Fréchet derivative $\nabla_{v_i} E_1(\{v_i\})$ is

$$
\begin{aligned}
\nabla_{v_i} E_1(\{v_i\}) = (N-1)\lambda \nabla_{v_i} \mathrm{Reg}(v_i) &- 2(N-1) \int_0^1 \Big(|D\Phi_{t,1}^{v_i}| (\mathcal{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla (\mathcal{I} \circ \Phi_{t,0}^{v_i}) \Big)\, dt \\
&- 2N\gamma_1 \int_0^1 \big(|D\Phi_{t,0}^{v_i}| (\hat{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla (I_i \circ \Phi_{t,1}^{v_i}) \big)\, dt \\
&- 2\gamma_2 \int_0^1 \Big(\big(\sum_{j=1}^{N} |D\Phi_{0,1}^{v_j}| \big) |D\Phi_{t,0}^{v_i}| (\tilde{I} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla (I_i \circ \Phi_{t,1}^{v_i}) \Big)\, dt \\
&- 2\gamma_2 \int_0^1 \sum_{j=1}^{N} |D\Phi_{t,1}^{v_i}| (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i} - I_i \circ \Phi_{t,1}^{v_i}) \nabla (I_j \circ \Phi_{0,1}^{v_j} \circ \Phi_{t,0}^{v_i})\, dt
\end{aligned} \tag{C.8}
$$

This equation yields the Euler-Lagrange Eq. (C.7). □

D. Experimental Details and Additional Results

This section provides details about our experimental settings and hyperparameter tuning. To ensure fair comparisons between methods with different hyperparameter settings, we use the same random seed throughout all experiments. When training or optimizing models, we obtain the best hyperparameters for each model by tuning on the same hold-out validation dataset with grid search over a subset of hyperparameter combinations. Then the estimated best hyperparameters are used to report the performances in the test dataset. The presented results are from a single instance of each model, not averaged repeated results with different initializations and hyperparameters.

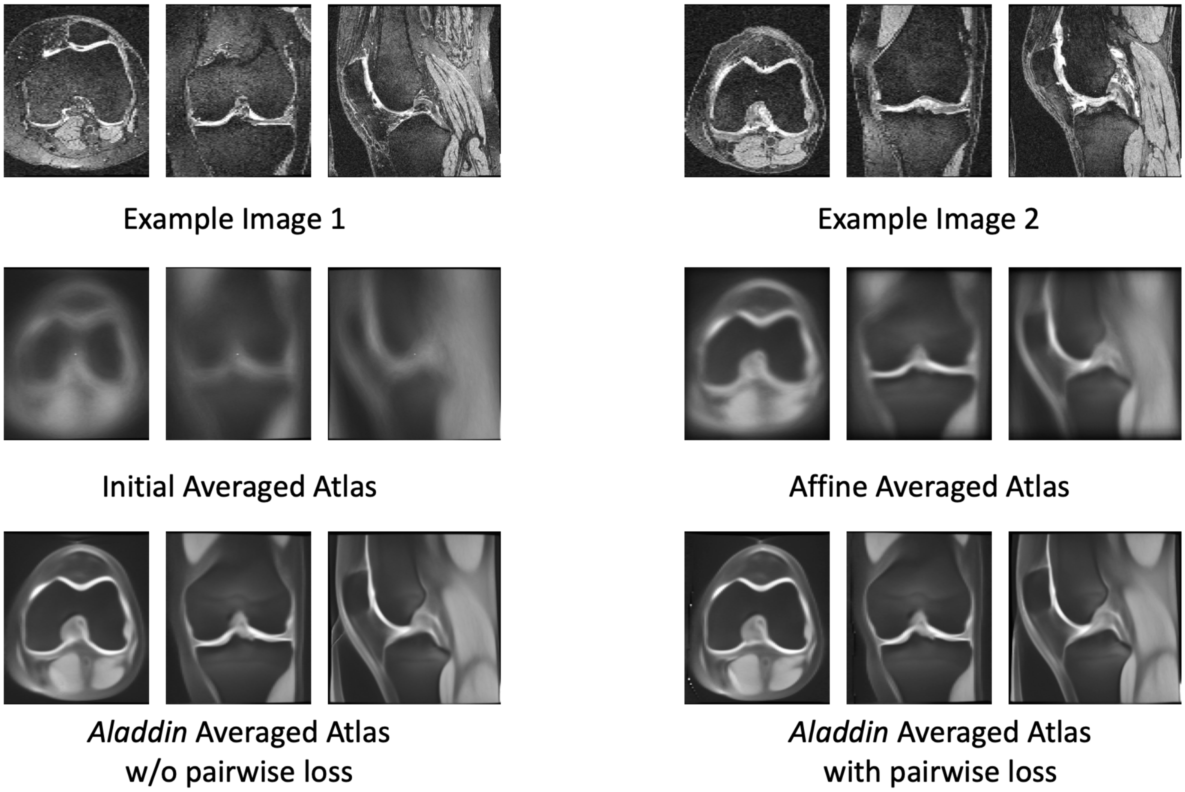

Figure 5: Examples of atlases and images in axial, coronal, and sagittal planes. **Top:** two example images. **Middle:** initial averaged atlas without and with affine pre-alignment. **Bottom:** *Aladdin* results without and with pairwise alignment loss. *Aladdin* obtains a relatively sharp atlas, maintaining the main structures, from the image population.

D.1. Fig. 4 experimental details

In this experiment, we fix the atlas and use the following parameters. The coefficient for the similarity loss is 10.0, $\lambda = 20000.0$, $\gamma_1 = 2.0$, and $\gamma_2 = 5.0$. We use `ADAM` as the optimizer to learn network parameters with a multi-step learning rate over 500 epochs. The initial learning rate is 10^{-4}, which is multiplied by 0.1 at the 300-th epoch and the 420-th epoch. Batch size is 2. We use MSE as the similarity loss.

D.2. Tab. 2 experimental details

Affine Registration: Affine registrations are performed by iteratively registering all images to the average of all images and updating the averaged image based on the resulting warped images. We use `reg_aladin` of Nifty-Reg [40, 36, 34, 38] for affine registration with MSE as similarity loss. These optimization-based affine transformations also serve as the affine pre-alignments for other experiments that require affine pre-registrations.
Joshi *et al*. [27]: We use the *Fast Symmetric Forces Demons Algorithm* [45] (via SimpleITK) as the registration algorithm. The Gaussian smoothing standard deviation for the displacement field is set to 1.2 and the total iteration number is set to 1,000. All the other hyperparameters are default. We use MSE as the similarity loss.
ABSORB [26]: We use *Diffeomorphic Demons* [45] (via SimpleITK) as the registration algorithm. The Gaussian smoothing standard deviation for the displacement field is set to 2.0 and the total iteration number is set to 1,000. We set the ABSORB hyperparameters as follows. We set the neighborhood size to 3, the Gaussian smoothing standard deviation for the displacement field is set to 2.0, the maximum number of levels to 3, the registration to mean to 1, histogram matching to true, and affine registration to false. All the other hyperparameters are default. We use MSE as the similarity loss.
Voxelmorph [5]: We set the regularization coefficient to 2,000 for experiments with affine pre-alignment and to 400 without affine pre-alignments. We use `ADAM` as the optimizer with learning rate 10^{-4}. Batch size is 4. We use MSE as the similarity loss.
He *et al*. [23]: We implement this method based on the descriptions in [23]. We use a 1-step 1-iteration framework to ensure

a fair comparison[9]. We add the atlas according to Eq. (2.3) in the implementation because we found that without an atlas the registration performance is significantly worse than with an atlas. The coefficient for $\mathcal{L}_{grad}$ is 10, the coefficient for $\mathcal{L}_{pres}$ is 1.0, the coefficient for $\mathcal{L}_{sim}$ is 0.4, and the coefficient for $\mathcal{L}_{cycle}$ is 0.1. We use `ADAM` as the optimizer with learning rate 10^{-4} over 500 epochs. Batch size is 8. We use NCC as the similarity loss.

Dalca *et al.* [12]: For experiments with affine pre-alignment we set $\gamma = 0.01$, $\lambda_d = 1.0$, $\lambda_a = 100.0$, and $\sigma^2 = 0.5$. For experiments without affine pre-alignment we set $\gamma = 0.01$, $\lambda_d = 0.2$, $\lambda_a = 20.0$, and $\sigma^2 = 0.5$. For experiments involving pairwise alignment losses, $\gamma_1 = 0.2$ and $\gamma_2 = 0.5$ have the best performance on the validation dataset. We use `ADAM` as the optimizer with learning rate 10^{-4} over 500 epochs. Batch size is 2. We use MSE as the similarity loss.

Aladdin: When specifying the atlas as in Eq. (2.4), the following hyperparameters achieve the best performance in the validation dataset: the coefficient for the similarity loss is 10.0, $\lambda = 1000.0$, $\gamma_1 = 1.0$ and $\gamma_2 = 5.0$. We use `ADAM` as the optimizer to learn network parameters with learning rate 10^{-4} over 500 epochs. Batch size is 2. We use MSE and NCC as the similarity losses in different experiments. When learning the atlas, we use another `SGD` optimizer with learning rate 10^4. When using NCC as the similarity loss, the coefficient of the similarity loss changes to 0.3 and $\gamma_2 = 0.15$.

D.3. Inverse deformation map calculation

For our approach and Dalca *et al.* [12], the inverse transformations are directly available. For all other baseline models, where inverse deformation maps are not available as part of the implementations, we obtain them numerically by solving $\arg\min_\phi \|\Phi\circ\phi - \mathrm{Id}\|_2^2 + \|\phi\circ\Phi - \mathrm{Id}\|_2^2$, where Φ is the known deformation map and ϕ is the sought-for inverse transformation map. We optimize using `ADAM` [28] with learning rate 10^{-4} until convergence. The loss is defined as $\mathcal{L} = \|\Phi\circ\phi - \mathrm{Id}\|_2^2 + \|\phi\circ\Phi - \mathrm{Id}\|_2^2$ and ϕ is the to-be-optimized parameter.

D.4. Example of an atlas built using *Aladdin*

Fig. 5 shows that images differ a lot in the image population. Hence the initial averaged atlas without affine pre-alignment is very fuzzy. Note that the averaging equation in the middle row is Eq. (2.3) and the averaging equation in the bottom row is Eq. (2.4) due to their *backward* and *forward* nature. After the affine alignments, the averaged atlas is clearer and more anatomical structures can be observed. *Aladdin* results in an even clearer structure, which means training images align well in atlas space. Besides, *Aladdin* with pairwise alignment loss results in slightly better alignment for the training images as indicated by the slightly clearer anatomical structures in the bottom row.

D.5. Evaluation measures for atlas building and registration in atlas space

In this work, we define evaluation measures in atlas space (i.e., d_k^{atlas}) as

$$d_k^{atlas} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{Dice}\big(S_i^k \circ \Phi_i\,, \mathbb{V}(\{S_i^k \circ \Phi_i\}_{i=1\ldots M})\big)\,, \tag{D.1}$$

in Sec. 4.1. In this definition, we first use a plurality voting scheme to obtain the consensus segmentation among all warped segmentations and then compare it with each warped segmentation. Note that in some other work [23, 24] evaluation measures in atlas space are defined without a consensus segmentation, i.e.,

$$\tilde{d}_k^{atlas} = \frac{2}{M(M-1)}\sum_{(i,j)\in\Gamma} \mathrm{Dice}\big(S_i^k \circ \Phi_i\,, S_j^k \circ \Phi_j\big)\,, \tag{D.2}$$

where $\Gamma = \{(i,j)|i = 1,2,...,M, j = 1,2,...,M, i < j\}$ is the set of all pairwise index combinations where the first index is smaller than the second index. Both evaluation measures are reasonable but evaluating Eq. (D.1) has $O(N)$ complexity while evaluating Eq. (D.2) has $O(N^2)$ complexity. Therefore, we choose Eq. (D.1) as the evaluation measure in atlas space to save computational time especially for images with multiple structures. The same reasoning also applies to our atlas-as-a-bridge evaluation measure d_k^{bridge} in Sec. 4.1.

D.6. Regularization

Aladdin is the only atlas building approach that uses the bending energy for regularization. The benefit of using the bending energy comes from the fact that it is invariant to affine transformations. Therefore no separate affine pre-registration

[9] If we would use a multi-step and multi-iteration framework to generate better results, to ensure fair comparisons, all the other baseline methods would also need to use a multi-step and multi-iteration approach.

is required. Assume we have an affine transformation of the form $\Phi^{aff} = Ax + x + b$, parameterized s.t. for $A = 0, b = 0$ we obtain the identity transform. Therefore, we have $\frac{\partial \Phi^{aff}}{\partial x} = A + \mathrm{Id}$ and $\frac{\partial^2 \Phi^{aff}}{\partial x^2} = 0$. Hence, the bending energy regularization zeros out any affine contributions. Consequently, in contrast to approaches with zero or first order derivative terms in their regularizer, *Aladdin* can *simultaneously* capture affine and nonparametric deformations.