

MultiMorph: On-demand Atlas Construction

S. Mazdak Abulnaga^{1,2} Andrew Hoopes^{1,2} Neel Dey¹ Malte Hoffmann²
 Bruce Fischl² John Guttag¹ Adrian Dalca^{1,2}

¹MIT Computer Science and Artificial Intelligence Laboratory

²Massachusetts General Hospital, Harvard Medical School

abulnaga@csail.mit.edu

Abstract

We present MultiMorph, a fast and efficient method for constructing anatomical atlases on the fly. Atlases capture the canonical structure of a collection of images and are essential for quantifying anatomical variability across populations. However, current atlas construction methods often require days to weeks of computation, thereby discouraging rapid experimentation. As a result, many scientific studies rely on suboptimal, precomputed atlases from mismatched populations, negatively impacting downstream analyses. MultiMorph addresses these challenges with a feedforward model that rapidly produces high-quality, population-specific atlases in a single forward pass for any 3D brain dataset, without any fine-tuning or optimization. MultiMorph is based on a linear group-interaction layer that aggregates and shares features within the group of input images. Further, by leveraging auxiliary synthetic data, MultiMorph generalizes to new imaging modalities and population groups at test-time. Experimentally, MultiMorph outperforms state-of-the-art optimization-based and learning-based atlas construction methods in both small and large population settings, with a 100-fold reduction in time. This makes MultiMorph an accessible framework for biomedical researchers without machine learning expertise, enabling rapid, high-quality atlas generation for diverse studies.

1. Introduction

We present MultiMorph, a rapid and flexible method for constructing anatomical atlases. An atlas, or deformable template, is a reference image that represents the typical structure within a collection of related images. In biomedical imaging studies, atlases facilitate studying anatomical variability within and across population groups by serving as a common coordinate system for key image analysis tasks such as segmentation [5, 26, 30, 82], shape analysis [1, 25, 52, 62], and longitudinal modeling [36, 66, 67].

Traditional unbiased atlas construction for a population involves solving a computationally intensive iterative opti-

mization problem that often requires several days or weeks of computation. The optimization alternates between aligning (registering) all images to the estimated atlas and updating the atlas in both shape and appearance by averaging the images mapped to the intermediate atlas space [7, 48]. Recent learning-based methods employ a target dataset to explicitly learn an atlas jointly with a registration model [18, 22], yet still require days of training. This necessitates computational infrastructure and machine learning expertise that is unavailable to many biomedical researchers.

Regardless of strategy, an atlas produced from one population of images may not be appropriate for populations that differ from the group used to build the atlas. Re-estimating the atlas is often required for each new experiment. These computational challenges are further compounded by the need to construct atlases for specific image types as many biomedical studies acquire several imaging modalities to highlight different biomedical properties of interest. The repeated, prohibitive computational cost of producing a new atlas leads most scientists to use existing atlases that might not be appropriate for their population group or modality, thereby negatively impacting the analyses in these studies [56].

To meet these challenges, we introduce MultiMorph, a machine learning model that constructs atlases in a single forward pass, requiring only seconds to minutes of computation on a CPU, and no machine learning expertise to use. MultiMorph efficiently generates population- and subgroup-specific atlases, enabling accurate and fine-grained anatomical analyses. We employ a convolutional architecture that processes an arbitrary number of images and computes a set of regularized deformation fields that align the group of images to an atlas space central to that group. The proposed method uses a nonparametric convolutional operation that interacts the intermediate representations of the input images with each other, summarizing and aggregating shared features. Further, by training on diverse imaging modalities alongside supplementary synthetic neuroimaging volumes [28], MultiMorph generalizes to arbitrary imaging modalities at test time. We also intro-

duce a *centrality* layer that ensures that the estimated atlases are unbiased [48]. As a result, MultiMorph rapidly produces high quality atlases for new populations and imaging modalities unseen during training. It further yields more accurate segmentation transfer across population groups than both the most widely used optimization-based approach [7] and recent machine learning approaches [18, 24]. To summarize our contributions:

- We frame atlas construction as a learning-based group registration problem to a central space.
- We present a novel neural network architecture that enables communication between the intermediate representations of a group of images, and show how this can be used to construct accurate group-specific atlases.
- We develop a *centrality* layer that encourages predicted deformations and atlases to be central and unbiased.
- Experimentally, MultiMorph produces atlases that are as good, and often better, than those produced by other methods—and it does it up to 100 times faster.
- We demonstrate the generalizability of the proposed method by constructing atlases for unseen imaging modalities and population groups. These atlases conditioned on age and disease state capture population trends within the data, enabling cross-group analyses.

Our model weights and code are available at <https://github.com/mabulnaga/multimorph>.

2. Related work

Deformable Registration. Deformable registration estimates a dense spatial mapping between image pairs. Traditional methods [4, 6, 47, 60, 69, 71, 75] solve an optimization problem balancing image-similarity and regularization terms to ensure smooth, invertible deformations.

Learning-based methods improve test-time efficiency by training models to directly predict transformations between image pairs, generally enabling faster predictions on new image pairs as compared to traditional methods. Supervised approaches [68, 78, 86, 87] are trained to regress simulated deformations or the outputs of registration solvers, whereas unsupervised methods [9, 15, 19, 21, 29, 32, 37, 38, 40, 54, 59, 61, 65, 80, 84, 89] optimize an unsupervised image-similarity loss and a regularization term in training.

Synthetic Data in Neuroimage Analysis. Recent machine learning-based neuroimage analysis methods have benefited from synthetic training data that extend far beyond real-world variations [10–12, 28, 35, 37, 38, 40, 41, 43, 49, 74]. This domain-randomization strategy trains neural networks on simulated intensity images, synthesized on the fly from a training set of anatomical segmentation maps. As part of the generative model, the images undergo corruption steps simulating common acquisition-related artifacts like distortion [2], low-frequency intensity modulation [77], global

intensity exponentiation [42], resolution reduction, partial voluming [83], among many others. The large variety of data yields shape-biased networks agnostic to the imaging modality. As a result, these models generalize to arbitrary medical images that have the same anatomy as the synthetic training data – largely eliminating the need for retraining to maintain peak performance [23, 39].

Atlas Construction. Deformable atlas construction seeks to find an image that optimally represents a given population, for example, to facilitate atlas-based brain segmentation [5, 26, 30, 82] or to initialize longitudinal morphometric analyses in an unbiased fashion [36, 66, 67].

Iterative atlas construction alternates between registering each image of the population to a current estimate of the atlas and updating the atlas with the average of the moved images until convergence [7, 48, 57, 64, 72]. Another approach computes a spanning tree of pair-wise transforms between subjects to estimate an atlas [44, 73]. Iterative methods on 3D data incur prohibitively long runtimes due to the cost of optimization. Therefore, many studies have used publicly available atlases [27], although these are often not representative of the population being studied.

Recent learning-based atlas construction techniques jointly learn an atlas and a registration network that maps images from the training population to the atlas [16, 18, 20, 22, 24, 31, 76, 79]. These approaches naturally extend to constructing *conditional* atlases, for example conditioning on age [18, 22, 79], or incorporating tasks like segmentation [76]. However, obtaining an atlas for a new population requires machine learning expertise and computational resources for re-training from scratch or fine-tuning a network.

Test-time adaptation for groupwise registration (TAG) [33, 34] maps a group of images to a latent space using a VAE, computes an average of latent vectors, then decodes to estimate an atlas. While this rapidly produces atlases at inference, linearly averaging vectors in a VAE latent space most often does not yield a representation that can be decoded into an unbiased deformable atlas. Further, this model must still be retrained for new imaging modalities or populations. In contrast, MultiMorph directly constructs group-specific atlases from warped images, ensuring fidelity to the data without distortions introduced by latent space aggregation. A single MultiMorph model can generate atlases for a wide variety of imaging modalities and population groups.

Flexible-size Inputs. Recent methods have employed a variety of mechanisms that are flexible to input size, in other applications. For example, in-context learning methods use a flexible-sized input set of input-output example pairs to guide a new image-processing task at inference [14, 17]. Other methods use attention mechanisms across different

inputs to aggregate information among volume slices [85] or tabular data [53]. While cross-attention and variants have been effective for many tasks in vision, they have quadratic memory complexity. At each iteration, our model requires a large set of 3D volumes. Using cross-attention would lead to infeasible memory requirements. In contrast, we propose a flexible feature sharing mechanism with linear complexity to produce central atlases for large groups of 3D images.

3. Methods

3.1. Background

Given two images $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$, deformable registration seeks a nonlinear mapping $\phi : \mathbf{x}_1 \rightarrow \mathbf{x}_2$ that warps one image into the space of the other. The deformation ϕ attempts to align the underlying anatomy captured by the images while maintaining a well-behaved map, and it is traditionally computed by optimization:

$$\arg \min_{\phi} \mathcal{L}_{sim}(\mathbf{x}_2, \mathbf{x}_1 \circ \phi) + \lambda \mathcal{L}_{reg}(\phi), \quad (1)$$

where $\mathcal{L}_{sim}(\mathbf{x}_2, \mathbf{x}_1 \circ \phi)$ measures similarity between image $\mathbf{x}_2$ and the warped image $\mathbf{x}_1 \circ \phi$, $\mathcal{L}_{reg}(\cdot)$ regularizes the map ϕ , and λ is a hyperparameter that balances the two.

Many population-based studies involve groupwise analyses. *Group registration* aligns a collection of m images $\mathcal{X}_m = \{\mathbf{x}_i\}_{i=1}^m$ to an explicit image template $\mathbf{t}$,

$$\arg \min_{\phi_1, \dots, \phi_m} \sum_{i=1}^m \mathcal{L}_{sim}(\mathbf{t}, \mathbf{x}_i \circ \phi_i) + \lambda \mathcal{L}_{reg}(\phi_i). \quad (2)$$

In many scenarios, an explicit template is not available. One can be constructed by iterating a template estimation step, $\hat{\mathbf{t}} = \frac{1}{m} \sum \mathbf{x}_i \circ \phi_i$, and the groupwise optimization (2) until convergence. However, this is computationally expensive and does not scale well to large populations.

Machine learning approaches for pairwise registration use a neural network to predict ϕ as a function of the input images: $f_{\theta}(\mathbf{x}_1, \mathbf{x}_2) = \phi$, where f is a neural network parameterized by θ . Pairwise registration is rapidly computed by a single forward pass of a trained network. Recent methods [18, 22, 24, 76] also estimate a common template $\mathbf{t}$ together with parameters θ in a network, $f_{\theta}(\mathbf{t}, \mathbf{x}) = \phi$.

In this work, we develop a model to directly predict a group-specific set of deformation fields to a central template space. We formulate template construction as a group registration problem given a variable number of inputs.

3.2. Flexible Group Registration

Given a set $\mathcal{X}_m$ of m images from a dataset $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$, we seek to map the images to a space central to $\mathcal{X}_m$. The set $\mathcal{X}_m$ could be an entire population or a subgroup of patients representing an underlying demographic or condition.

Let $\phi_i : \mathbb{R}^d \rightarrow \mathbb{R}^d$ represent the map from $\mathbf{x}_i$ to a central latent image space. We model a function $f_{\theta}(\{\mathbf{x}_1, \dots, \mathbf{x}_m\}) = \{\phi_1, \dots, \phi_m\}$ using a convolutional neural network with learnable parameters θ . The number of parameters of θ is independent of the group size m .

To achieve desirable group registration, we construct f to satisfy the following desiderata:

- **Flexible to group size:** f takes as input a variable number of m images, and computes m maps ϕ to a group-specific central space.
- **Fast:** Computation of $\{\phi_i\}_{i=1}^m$ can be done in a single forward pass of an efficient network.
- **Generalizable:** generalize to unseen datasets $\mathcal{Y}_m$.
- **Unbiased:** the images in $\mathcal{X}_m$ map to a space *central* to that set: $\text{mean}(\{\phi_i\}_{i=1}^m) = 0$.
- **Aligned:** Images $\{\mathbf{x}_i \circ \phi_i\}_{i=1}^m$ mapped to the template space are anatomically aligned.

Satisfying the desiderata leads to a model that can produce flexible templates for user-defined groups on demand. We introduce new methods to achieve these properties below.

3.3. Model

Figure 1 gives an overview of the network architecture of MultiMorph. The network takes a group of a variable number of images and predicts diffeomorphic transformations to a central template space specific to the group. At each network layer, we share features across the inputs using the proposed GroupBlock layer. The network outputs stationary velocity fields, which are then adjusted by a centrality layer to produce an unbiased atlas.

Convolution Layer for Variable Group Size. We propose GroupBlock, a convolutional layer that combines image features across a group. As group registration seeks to align images to a central space, feature communication is helpful to produce a subgroup alignment.

We use a summary statistic to aggregate group features, and communicate the statistic back to each individual group element. Let $c_i^{(l)}$ represent the feature map for input image i at network layer l . The GroupBlock layer aggregates information as follows:

$$\begin{aligned} \bar{c}^{(l)} &= s(\{c_1^{(l)}, \dots, c_m^{(l)}\}) \\ c_i^{(l+1)} &= \text{Conv} \left(\left[c_i^{(l)} \parallel \bar{c}^{(l)} \right]; \theta^{(l)} \right), \end{aligned}$$

where $s(\cdot)$ is the summary statistic across the group dimension, $[\cdot \parallel \cdot]$ is the concatenation operation along the channel dimension, and Conv is a convolutional layer with parameters $\theta^{(l)}$. We use the mean as our summary statistic.

Network. We modify the popular UNet architecture [70], employing a multi-scale structure with residual connections. We replace the standard Conv layers with the proposed GroupBlock feature sharing layer (§3.3). The net-

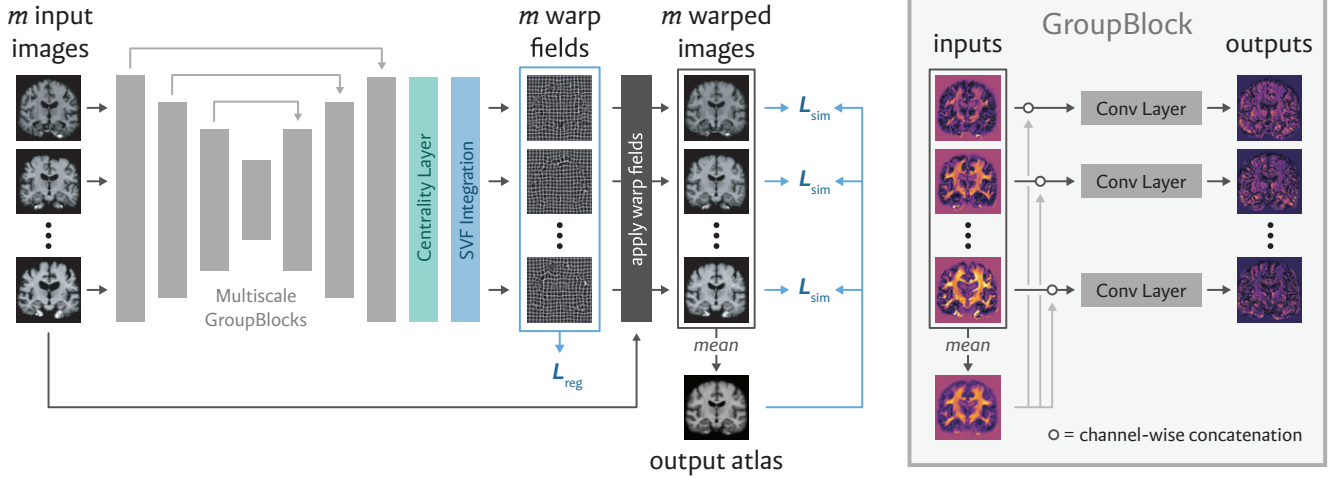


Figure 1. MultiMorph architecture diagram. The model takes in a variable group of m images and constructs an atlas specific to that group. At each layer of the UNet, the proposed GroupBlock mechanism replaces standard convolution kernels. Specifically, it computes the elementwise mean of the intermediate features across the group, and concatenates the resulting features with the individual features. The mechanism enables group interaction by sharing summarized input features across the group. The network outputs m velocity fields mapping images to the group-specific template space. A centrality layer removes any global bias in the average velocity field, before integration and warping the images. The output is a central template representing the shared anatomy of the input group.

work takes as input a group of m images and outputs m d -dimensional stationary velocity fields (SVF).

We use the standard SVF representation of a diffeomorphism [4, 18]. The deformation field is defined through the ordinary differential equation: $\frac{\partial \phi_v^{(t)}}{\partial t} = v \circ \phi_v^{(t)}$, where v is the velocity field. The field is defined for $t \in [0, 1]$ with $\phi_v^{(0)} = Id$, the identity map. The deformation field is obtained by integrating v using the scaling and squaring method [4, 19].

Centrality Layer. Constructing an atlas central to the population group is key to performing unbiased downstream analyses, such as in quantifying anatomical shape differences without bias to any particular structure or subject. A central atlas is one that is “close” to the target population. Many learning approaches use a regularization term to minimize the mean displacement field [18, 22].

We construct a layer that produces a group-specific central template by construction. We subtract the groupwise mean from the output velocity fields: $v_i = v_i^{(L)} - \bar{v}^{(L)}$, where $v_i^{(L)}$ is the final output velocity field for image i , and $\bar{v}^{(L)}$ is the group mean. This centers the velocity fields in the zero-mean Lie subspace.

Template Construction. Given a trained network f_θ , we can construct a template t by aggregating the warped images of the group $\mathcal{X}_m$: $t = g(x_1 \circ \phi_1, \dots, x_m \circ \phi_m)$. We use the mean operation for g .

To apply the map to the group of images, we integrate the SVF to obtain a diffeomorphic displacement field [4, 19]. We then use a spatial transformation function [46] to warp

the images to the central space. The spatial transformer performs a pullback operation with linear interpolation.

3.4. Auxiliary Structural Information

The use of anatomical labelmaps during training of learning-based registration often improves substructure alignment [9]. When segmentation maps are available for some of the images in a set, we use this information to form an atlas segmentation map. Let $\text{seg}[x_i]$ indicate the probabilistic segmentation labelmap of the K structures for image x_i . We construct the labelmap of the template, $\text{seg}[t]$, by taking the set-wise average of the warped probability maps $\text{seg}[t] = \text{mean}_m \{ \text{seg}[x_1] \circ \phi_1, \dots, \text{seg}[x_m] \circ \phi_m \}$.

3.5. Synthetic Training

To aid generalization to unseen modalities, we also train on images synthesized from brain tissue segmentations. For each synthetic training group, we sample K random values uniformly corresponding to K structures. We then use a domain randomization procedure [28] to randomly sample intensity values for each structure, along with a variety of noise patterns and artifacts. This yields groups of synthetic images, where each group exhibits random intensity distributions and tissue contrasts. Supplementary Fig. 11 presents a representative set of example synthetic images.

3.6. Loss

We maximize alignment between the images and anatomical structures of the group and the constructed template, while maintaining a smooth map. For a single image, the

loss is computed as:

$$\mathcal{L}(\phi_i) = \mathcal{L}_{sim}(\mathbf{t}, \mathbf{x}_i \circ \phi_i) + \lambda \mathcal{L}_{reg}(\phi_i) + \gamma \mathcal{L}_{struc}(\text{seg}[\mathbf{t}], \text{seg}[\mathbf{x}_i] \circ \phi_i). \quad (3)$$

The first term $\mathcal{L}_{sim}$ measures pairwise similarity between image $\mathbf{x}_i$ and the template $\mathbf{t}$. We use the normalized cross-correlation objective. The second term regularizes the deformation field to be smooth, $\mathcal{L}_{reg}(\phi_i) = \|\nabla \phi_i\|^2$. When label maps are available during training, we use the third (auxiliary) loss term to align the structures of the training set with the constructed template, using soft-Dice.

Our complete group loss is $\mathcal{L}(\phi_1, \dots, \phi_m) = \frac{1}{m} \sum_{\{i: \mathbf{x}_i \in \mathcal{X}_m\}} \mathcal{L}(\phi_i)$. Since the template $\mathbf{t}$ is constructed by averaging warped images of the group, the loss is dependent on all images of the group.

4. Experiments

We evaluate MultiMorph using 3D brain MRI brain scans, a common setting for atlas construction. We compare MultiMorph against iterative and learning-based approaches in terms of speed, centrality, and accuracy. We also test whether MultiMorph generalizes to new datasets, imaging modalities, and populations that are unseen during training.

4.1. Experimental Setup

Data. We use four public 3D brain MRI datasets. Three datasets—OASIS-1, OASIS-3, and MBB—are used for training, validation, and testing, while IXI serves as an unseen test set. OASIS-1 [58] includes T1-weighted (T1-w) scans of 416 subjects aged 18-96. A hundred OASIS-1 subjects of ages 60 years and older were diagnosed with mild to moderate Alzheimer’s disease (AD), which is correlated with brain atrophy. OASIS-3 [55] contains T1-w and T2-w MRI scans of subjects aged 42-95 years old. We use a subset of 1043 subjects, with 210 diagnosed with mild to severe cognitive dementia. The Mind Brain Body dataset [8] includes T2-w and T2-FLAIR scans of 226 healthy subjects. For each training dataset, we randomly hold out 20% of the subjects for testing, and split the rest into 85% for training and 15% for validation. Each split includes an equal mix of healthy and abnormal subjects of all age ranges. We use the same model for all experiments.

Lastly, to evaluate generalization, we hold out the IXI dataset [45]. We arbitrarily select the Guys Hospital site within IXI and retrieve T1-w, T2-w, and PD-w MRI scans of 319 adult subjects. Importantly, the PD-w MRI modality is not included in any of the training datasets used by our model. These datasets span a large age range and include a mix of disease states and imaging modalities, simulating real-world population studies.

Implementation details. During training, all images within a sampled group have the same acquisition modality. We

apply augmentations, including random exponential scaling, intensity inhomogeneities, and per-voxel Gaussian noise. Additionally, 50% of the sampled training groups contain synthetic images instead of real acquisitions. For preprocessing, using ANTs [81], we affinely align each 3D scan to a common 1-mm isotropic affine reference used in [37, 39]. We extract brain tissue signal using SynthStrip [41] and generate segmentation maps of 30 unique anatomical brain structures using SynthSeg [11].

We train using the Adam optimizer [51] with a learning rate of 10^{-4} . The field regularization hyperparameter is set to $\lambda = 1.0$ and the segmentation-loss weight is $\gamma = 0.5$, both chosen via grid search (see Supplemental Sec. 7). At each training iteration, we randomly sample $m = [2, 12]$ images to form a group and train for 80,000 iterations, using the final saved model. All models are trained on a single RTX8000 GPU. All runtime evaluation results were done on an Intel(R) Xeon(R) Gold 5218 CPU.

Baselines. We evaluate SyGN [7], a widely-used iterative atlas construction method from the ANTs library [81]. Additionally, we compare against AtlasMorph, a learning-based atlas constructor [18] that explicitly *learns* an atlas to best fit the training data. For AtlasMorph, we set the deformation field regularization hyperparameter to $\lambda = 0.1$, as determined via cross-validation. Both MultiMorph and AtlasMorph use the same core registration network.

We also evaluate Aladdin [24], a learning-based method that constructs an average reference atlas during training by learning pairwise registrations. At test time, this atlas serves as the registration target, enabling the generation of new atlases for different population groups. Since Aladdin constructs modality-specific atlases, we train a separate model (with the same capacity as our network) for each modality in our dataset using an optimal regularization loss weight of 10,000, a similarity loss weight of 10, and an image pair loss weight of 0.2, all determined using a grid search. Both AtlasMorph and Aladdin models are trained for 50,000 iterations, followed by 1,500 finetuning iterations per population subgroup to estimate a group atlas at test-time.

Evaluation. We assess the effectiveness of atlas construction techniques in rapidly generating central atlases for new populations. To evaluate registration quality, we compute the Dice score to assess how well the atlas aligns with warped subject scans. We assess field regularity and topology by computing the determinant of the Jacobian of the map, $\det J_\phi(p) = \det(\nabla \phi(p))$ at each voxel p . Locations where $\det J_\phi(p) < 0$ represent folded regions breaking local injectivity. Additionally, we measure atlas centrality by reporting the mean displacement field $\|\bar{\mathbf{u}}\|^2$. Statistical significance is determined using a paired t-test with $p < 0.01$.

Segmentation transfer. As atlases are commonly used for segmentation by warping atlas labels to new target

Table 1. Atlas construction evaluation on 319 brain volumes from IXI. While all baselines were trained or optimized on the full dataset, MultiMorph was not, demonstrating its ability to generalize to entirely new datasets. MultiMorph also generalizes to the PD-w modality not seen during training, demonstrating its capabilities on unseen imaging modalities. * indicates statistical significance ($p < 0.01$).

Modality	Method	Construction time (min.) ($\downarrow$)	Dice ($\uparrow$)	Folds ($\downarrow$)	Centrality $\times 10^{-2}$ ($\downarrow$)
T1-w	ANTs[7]	4345.20	0.863 ± 0.075	524.2 ± 580.04	10.4 ± 30.67
	AtlasMorph[18]	1141.50	0.894 ± 0.015	47.9 ± 29.22	7.8 ± 19.09
	Aladdin[24]	325.20	0.885 ± 0.01	$0.0 \pm 0.0^*$	106.8 ± 97.6
	Ours	10.50	$0.913 \pm 0.006^*$	1.1 ± 1.55	$1.4 \pm 4.32^*$
T2-w	ANTs[7]	4380.60	0.862 ± 0.071	522.6 ± 476.86	18.6 ± 44.286
	AtlasMorph[18]	831.60	0.882 ± 0.018	57.5 ± 31.935	7.8 ± 19.34
	Aladdin[24]	261.00	0.875 ± 0.012	$0.0 \pm 0.125^*$	771.5 ± 744.309
	Ours	10.40	$0.906 \pm 0.007^*$	2.0 ± 2.49	$1.5 \pm 4.683^*$
PD-w	ANTs[7]	4320.20	0.856 ± 0.069	313.1 ± 359.9	12.4 ± 32.805
	AtlasMorph[18]	959.00	0.884 ± 0.018	40.5 ± 26.10	7.4 ± 19.483
	Aladdin[24]	163.80	0.849 ± 0.029	$0.0 \pm 0.0^*$	1175.7 ± 1731.773
	Ours	7.80	$0.900 \pm 0.009^*$	1.601 ± 0.205	$0.9 \pm 3.02^*$

images, we evaluate each method’s segmentation performance. Each atlas is estimated using half the subgroup ($\frac{m}{2}$ images). We randomly sample $\frac{m}{2}$ segmentation label maps to generate the atlas segmentation mask, which is then transferred to the remaining $\frac{m}{2}$ images. Segmentation quality is assessed using the Dice score.

4.2. Results

4.2.1 Generalizing to Unseen Datasets and Modalities

Table 1 presents results for all methods on the IXI dataset, which was entirely held-out for MultiMorph’s training and validation. MultiMorph produces atlases over a $100\times$ faster than ANTs and AtlasMorph, and $30\times$ faster than Aladdin. It consistently achieves the highest Dice score, indicating better anatomical alignment even when constructed on unseen data at test time in a single forward pass. Additionally, MultiMorph yields regular deformation fields with negligible folding and significantly lower bias in the displacement fields, indicating that the constructed atlases are central.

Fig. 2 visualizes sample registration predictions for each modality in IXI and Fig. 3 illustrates example atlases for IXI T1-w and PD-w. Despite never having been trained on this dataset nor having seen the PD-w imaging modality during training, MultiMorph estimates atlases that yield high group alignment in only minutes, demonstrating its potential for

scientific studies requiring specific atlases. We provide additional examples in Supplemental Fig. 10.

4.2.2 Standard Atlas Construction

We now evaluate the ability of MultiMorph to construct population atlases across different age groups and disease states. Specifically, we construct an atlas on the OASIS-3 T1-w test dataset. All baseline models were trained and validated on the test set. Table 3 shows that MultiMorph achieves the highest Dice score while producing atlases 30 – 400 times faster than the baseline methods.

4.2.3 Subgroup Atlas Construction

MultiMorph enables the rapid construction of subgroup atlases for granular population analyses. We evaluate atlases

Table 2. Sub-group atlas construction results. Reported scores are averaged across atlases constructed using subgroups of [5, 10, 20, . . . , 60]. * indicates statistical significance ($p < 0.01$).

Method	Run time (min.) ($\downarrow$)	Dice Transfer ($\uparrow$)	Folds ($\downarrow$)	Centrality $\times 10^{-2}$ ($\downarrow$)
ANTs [7]	436 ± 0.4	0.875 ± 0.009	447 ± 110	8.7 ± 0.1
AtlasMorph [18]	17 ± 1.4	0.893 ± 0.005	50.0 ± 8.7	9.7 ± 0.1
Aladdin [24]	12 ± 0.1	0.877 ± 0.004	$0.0 \pm 0.0^*$	173 ± 3.7
Ours	1.5 ± 0.0	$0.904 \pm 0.002^*$	1.3 ± 0.4	$1.4 \pm 0.04^*$

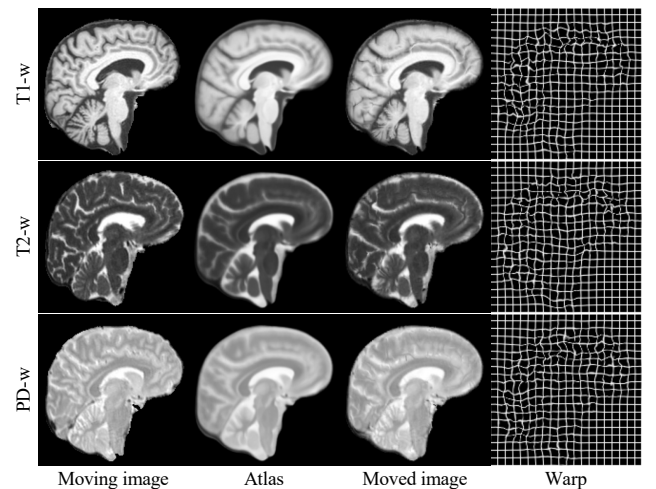


Figure 2. Example images and warps to the constructed atlas on the IXI dataset. Three subjects and three modalities are shown.

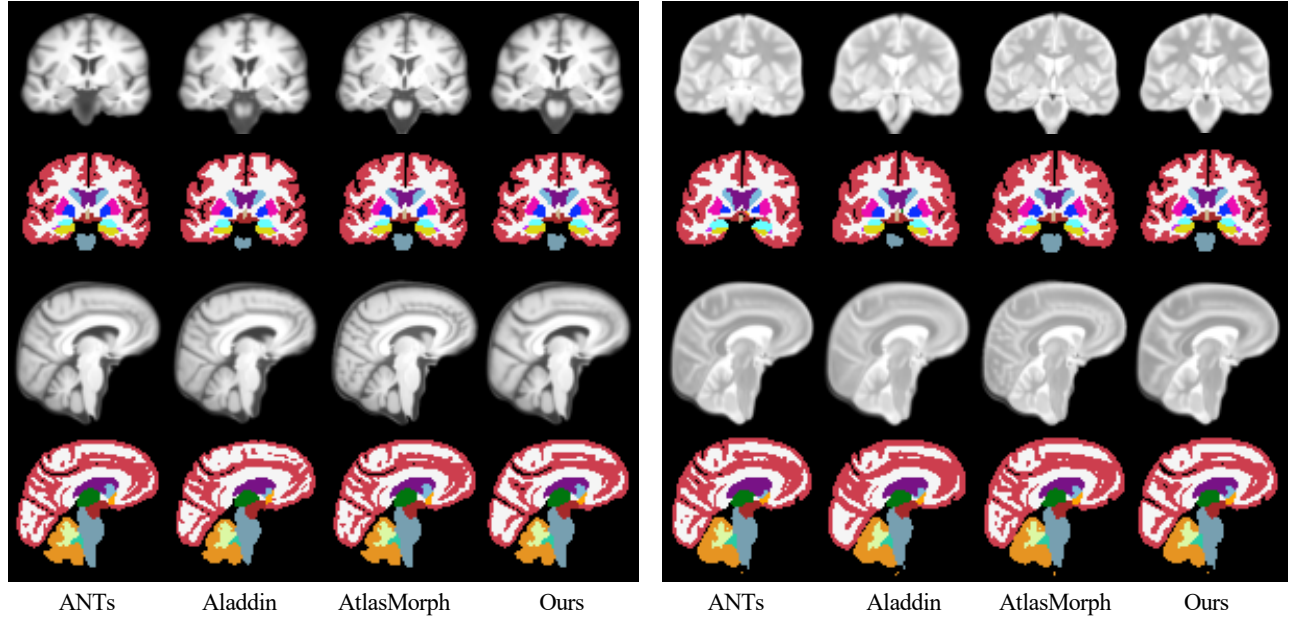


Figure 3. Atlases constructed on the IXI T1-w (left) and IXI PD-w (right) image modality. All baseline methods used the dataset for training or optimization, while our method was never trained on the IXI data. Further, our method was never trained on PD-w images, yet generalizes to this modality.

conditioned on age, age and disease state, as well as random subgroups of the population.

Random Subgroup Analysis. We quantify the effect of subgroup size on atlas quality using the held-out IXI T1-w dataset. Subgroups of $[5, 10, 20, \dots, 60]$ images are randomly sampled, with half used to construct the atlas segmentation and the other half used for evaluation. As in Section 4.2.1, the baselines were trained or optimized on this dataset, whereas MultiMorph was not exposed to any IXI T1-w data during training or validation.

Fig. 4 shows that MultiMorph consistently outperforms baselines, with performance improving as the subgroup

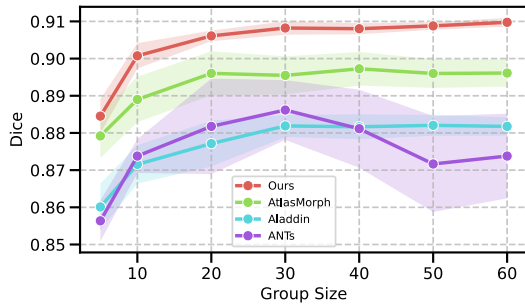


Figure 4. Segmentation transfer performance when varying the number of images used to construct an atlas. Data is taken from the IXI T1-w dataset, which our model did not have access to during training. Our method consistently outperforms the baselines.

size increases. Table 2 reports mean performance across subgroups, with MultiMorph showing better segmentation transfer while maintaining well-behaved deformation fields. Importantly, MultiMorph only requires 1.5 minutes of inference time on a CPU, whereas baselines require fine-tuning or re-optimization, which is both time consuming and requires tens or hundreds of minutes.

Age. We first demonstrate MultiMorph’s ability to create appropriate atlases for user-defined subgroups by grouping healthy OASIS-1 subjects into age bins. We take normal subjects in the validation and test set, and bin them into age ranges $[0 - 19, 20 - 29, \dots, 80 - 89]$. Fig. 5 presents qualitative results, showing anatomical changes consistent with normal aging, such as ventricular enlargement due to brain atrophy [3]. All atlases were generated in under a minute without any fine-tuning.

Diagnosis. Lastly, we examine the effect of dementia on brain aging in the OASIS-3 (T1-w) dataset. We con-

Table 3. Atlas estimation results on 212 subjects from the OASIS-3 T1-w test set. * indicates statistical significance ($p < 0.01$).

Method	Run time (min.) (↓)	Dice Transfer (↑)	Folds (↓)	Centrality $\times 10^{-2}$ (↓)
ANTs [7]	2858	0.886 ± 0.017	765 ± 877	9.5 ± 25.6
AtlasMorph [18]	688	0.881 ± 0.024	50.2 ± 31.9	8.0 ± 0.2
Aladdin [24]	277	0.878 ± 0.016	$0.0 \pm 0.07^*$	175.9 ± 1.8
Ours	5.9	$0.910 \pm 0.014^*$	1.2 ± 2.3	$1.5 \pm 0.05^*$

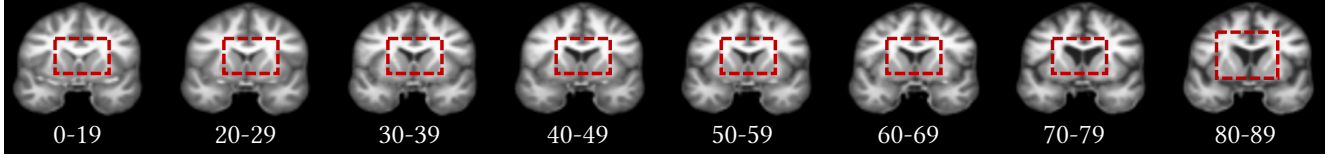


Figure 5. Atlases conditioned on age for healthy subjects in OASIS-1. Ventricle enlargement (red boxes) is observed across time, consistent with neurodegeneration with aging.

struct age-conditioned atlases separately for normal and dementia-diagnosed subjects. Fig. 6 compares brain atrophy across matched age groups. We observe substantial enlargement of the ventricles (outlined in red boxes) and deterioration of the white matter in the dementia group as compared to the controls, consistent with the literature [3, 50, 63, 88].

4.3. Ablation studies

We quantify the effect of several key model components, including the centrality layer (CL), the Group Block (GB) mechanism with varying summary statistics (mean, variance, max), and training without the Dice Loss. Using the OASIS-1 dataset [58], we train our model for 50,000 iterations and assess performance on the test set.

Table 4 summarizes the results. The CL significantly reduced the centrality measure by 1000 $\times$, ensuring unbiased atlas construction, though it came along with a 1 point decrease in Dice. The GB mechanism improved Dice by 1.4 points with negligible degradation of field regularity. We observe no significant performance variation across the various summary statistics tested. Finally, the Dice loss improved performance by over 2 Dice points. Taken holistically, each component strongly contributed to MultiMorph’s performance. We further quantify the impact on subgroup atlas construction in Supplemental Section 6.2 and observe similar trends. Additionally, we assess the impact of training with synthetic data in Supplemental Section 6.1, which improved IXI dataset performance by up to

1.8 Dice points while maintaining field regularity, demonstrating better generalization.

5. Discussion

Limitations and future work. MultiMorph has several avenues for extensions. For example, as it assumes diffeomorphic transformations, MultiMorph cannot accurately construct atlases for neuroimages with topology-changing pathologies. However, this can be addressed by using pathology masks when calculating losses in training [13]. Additionally, MultiMorph is currently only trained for neuroimages, but can be trained on anatomy-agnostic synthetic data [23, 37] to estimate atlases for arbitrary applications. Lastly, our implementation stores all activations in memory at inference, potentially limiting higher group sizes with large 3D volumes in memory-constrained settings.

Conclusion. We presented MultiMorph, a test-time atlas construction framework that works with unseen imaging modalities and any number of input images—without retraining. At its core, MultiMorph leverages a novel convolutional layer for *groups* of images, independent of the number of input samples, enabling efficient and scalable atlas generation. MultiMorph produces unbiased atlases for arbitrary inputs with comparable (and often better) performance, while also being over 100 times faster than previous approaches that require either solving an optimization problem or retraining a model. By making high-quality atlas construction fast, accessible, and adaptable, MultiMorph potentially unlocks new avenues for biomedical research, enabling computational anatomy studies that were previously impractical due to computational constraints.

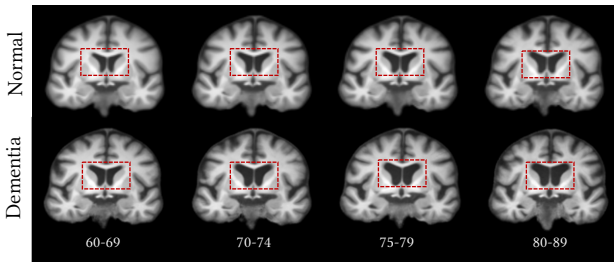


Figure 6. Atlases conditioned on age for normal subjects (top) and subjects with dementia (bottom) from OASIS-3. Visual differences indicate considerable enlargement of ventricles (red boxes) and atrophy of white matter when compared to normal subjects.

Table 4. Model ablations on the Centrality Layer, Group Block mechanism, and Dice loss on the OASIS-1 test set. All proposed components improved atlas construction performance.

Ablation	Dice Transfer ($\uparrow$)	Folds ($\downarrow$)	Centrality $\times 10^{-3}$ ($\downarrow$)
no CL, GB(mean)	0.892 ± 0.018	0.0 ± 0.0	16125 ± 11494
CL, no GB	0.870 ± 0.021	0.1 ± 0.3	9.9 ± 27.4
CL, GB(var)	0.883 ± 0.020	1.5 ± 2.8	12.8 ± 59.27
CL, GB(max)	0.880 ± 0.019	1.5 ± 2.7	12.6 ± 46.69
CL, GB(mean)	0.884 ± 0.020	1.1 ± 1.9	12.0 ± 39.48
CL, GB(mean), Dice	0.919 ± 0.011	5.4 ± 7.5	18.6 ± 61.31

Acknowledgements

We thank Marianne Rakic for her help in the AtlasMorph experiments. We thank Zack Berger for help in proof-reading. Support for this research was provided in part by Quanta Computer Inc. project AIR, the NIH BICCN grants U01 MH117023 and UM1 MH130981, NIH BRAIN CON-NECTS U01 NS132181, UM1 NS132358, NIH NIBIB R01 EB023281, R21 EB018907, R01 EB019956, P41 EB030006, NIH NIA R21 AG082082, R01 AG064027, R01 AG016495, R01 AG070988, the NIH NIMH UM1 MH130981, R01 MH123195, R01 MH121885, RF1 MH123195, NIH NINDS U24 NS135561, R01 NS070963, R01 NS083534, R01 NS105820, R25 NS125599, NIH NICHD R00 HD101553, NIH R01 EB033773, and was made possible by the resources provided by NIH Shared In-strumentation Grants S10 RR023401, S10 RR019307, and S10 RR023043. Additional support was provided by the NIH Blueprint for Neuroscience Research U01 MH093765, part of the multi-institutional Human Connectome Project. Much of the computation resources was performed on hard-ware provided by the Massachusetts Life Sciences Center.

References

- [1] Riza Alp Guler, George Trigeorgis, Epameinondas Antonakos, Patrick Snape, Stefanos Zafeiriou, and Iasonas Kokkinos. Densereg: Fully convolutional dense shape regression in-the-wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6799–6808, 2017. [1](#)
- [2] Jesper LR Andersson, Stefan Skare, and John Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888, 2003. [2](#)
- [3] Liana G Apostolova, Amity E Green, Sona Babakchian, Kristy S Hwang, Yi-Yu Chou, Arthur W Toga, and Paul M Thompson. Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (mci), and alzheimer disease. *Alzheimer Disease & Associated Disorders*, 26(1):17–27, 2012. [7](#), [8](#)
- [4] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007. [2](#), [4](#)
- [5] John Ashburner and Karl J Friston. Unified segmentation. *neuroimage*, 26(3):839–851, 2005. [1](#), [2](#)
- [6] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008. [2](#)
- [7] Brian B Avants, Paul Yushkevich, John Pluta, David Minkoff, Marc Korczykowski, John Detre, and James C Gee. The optimal template effect in hippocampus studies of diseased populations. *Neuroimage*, 49(3):2457–2466, 2010. [1](#), [2](#), [5](#), [6](#), [7](#)
- [8] A Babayan, M Erbey, D Kumral, JD Reinelt, AMF Reiter, J Röbbig, HL Schaare, M Uhlig, A Anwander, PL Bazin, et al. A mind-brain-body dataset of mri, eeg, cognition, emotion, and peripheral physiology in young and old adults. *sci. data* 6, 180308, 2018. [5](#)
- [9] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019. [2](#), [4](#)
- [10] Benjamin Billot, Eleanor Robinson, Adrian V Dalca, and Juan Eugenio Iglesias. Partial volume segmentation of brain mri scans of any resolution and contrast. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII* 23, pages 177–187. Springer, 2020. [2](#)
- [11] Benjamin Billot, Douglas N Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V Dalca, Juan Eugenio Iglesias, et al. Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical image analysis*, 86:102789, 2023. [5](#)
- [12] Benjamin Billot, Colin Magdamo, You Cheng, Steven E Arnold, Sudeshna Das, and Juan Eugenio Iglesias. Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain mri datasets. *Proceedings of the National Academy of Sciences*, 120(9):e2216399120, 2023. [2](#)
- [13] Matthew Brett, Alexander P Leff, Chris Rorden, and John Ashburner. Spatial normalization of brain images with focal lesions using cost function masking. *Neuroimage*, 14(2):486–500, 2001. [8](#)
- [14] Victor Ion Butoi, Jose Javier Gonzalez Ortiz, Tianyu Ma, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Universeg: Universal medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21438–21451, 2023. [2](#)
- [15] Qing Chang, Chenhao Lu, and Mengke Li. Cascading affine and B-spline registration method for large deformation registration of lung X-rays. *Journal of Digital Imaging*, 36(3):1262–1278, 2023. [2](#)
- [16] Zeen Chi, Zhongxiao Cong, Clinton J Wang, Yingcheng Liu, Esra Abaci Turk, P Ellen Grant, S Mazdak Abulnaga, Polina Golland, and Neel Dey. Dynamic neural fields for learning atlases of 4d fetal mri time-series. *arXiv preprint arXiv:2311.02874*, 2023. [2](#)
- [17] Steffen Czolbe and Adrian V Dalca. Neuralizer: General neuroimage analysis without re-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6217–6230, 2023. [2](#)
- [18] Adrian Dalca, Marianne Rakic, John Guttag, and Mert Sabuncu. Learning conditional deformable templates with convolutional networks. *Advances in neural information processing systems*, 32, 2019. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#)
- [19] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis*, 57:226–236, 2019. [2](#), [4](#)
- [20] Maik Dannecker, Vanessa Kyriakopoulou, Lucilio Cordero-Grande, Anthony N Price, Joseph V Hajnal, and Daniel

- Rueckert. Cina: Conditional implicit neural atlas for spatio-temporal representation of fetal brains. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 181–191. Springer, 2024. 2
- [21] Bob D De Vos, Floris F Berendsen, Max A Viergever, Marius Staring, and Ivana Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 204–212. Springer, 2017. 2
- [22] Neel Dey, Mengwei Ren, Adrian V Dalca, and Guido Gerig. Generative adversarial registration for improved conditional deformable templates. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3929–3941, 2021. 1, 2, 3, 4
- [23] Neel Dey, Benjamin Billot, Hallee E Wong, Clinton J Wang, Mengwei Ren, P Ellen Grant, Adrian V Dalca, and Polina Golland. Learning general-purpose biomedical volume representations using randomized synthesis. *arXiv preprint arXiv:2411.02372*, 2024. 2, 8
- [24] Zhipeng Ding and Marc Niethammer. Aladdin: Joint atlas building and diffeomorphic registration learning with pairwise alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20784–20793, 2022. 2, 3, 5, 6, 7
- [25] Pedro F Felzenszwalb. Representation and detection of deformable shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):208–220, 2005. 1
- [26] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002. 1, 2
- [27] Vladimir S Fonov, Alan C Evans, Robert C McKinstry, C Robert Almlie, and DL Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47:S102, 2009. 2
- [28] Karthik Gopinath, Andrew Hoopes, Daniel C Alexander, Steven E Arnold, Yael Balbastre, Adrià Casamitjana, You Cheng, Russ Yue Zhi Chua, Brian L Edlow, Bruce Fischl, et al. Synthetic data in generalizable, learning-based neuroimaging. *Imaging Neuroscience*, 2024. 1, 2, 4
- [29] Karthik Gopinath, Xiaoling Hu, Malte Hoffmann, Oula Puonti, and Juan Eugenio Iglesias. Registration by regression (rbr): a framework for interpretable and flexible atlas registration. *arXiv preprint arXiv:2404.16781*, 2024. 2
- [30] Vicente Grau, AUJ Mewes, M Alcaniz, Ron Kikinis, and Simon K Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE transactions on medical imaging*, 23(4):447–458, 2004. 1, 2
- [31] Christoph Großbröhmer, Ziad Al-Haj Hemidi, Fenja Falta, and Matthias P Heinrich. SINA: Sharp implicit neural atlases by joint optimisation of representation and deformation. In *International Workshop on Biomedical Image Registration*, pages 165–180. Springer, 2024. 2
- [32] Daniel Grzech, Mohammad Farid Azampour, Ben Glocker, Julia Schnabel, Nassir Navab, Bernhard Kainz, and Loïc Le Folgoc. A variational Bayesian method for similarity learning in non-rigid image registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 119–128, 2022. 2
- [33] Ziyi He and Albert C. S. Chung. Instantgroup: Instant template generation for scalable group of brain mri registration, 2024. 2
- [34] Ziyi He, Tony C. W. Mok, and Albert C. S. Chung. Group-wise image registration with atlas of multiple resolutions refined at test phase. In *MICCAI 2023 Workshops*, 2023. 2
- [35] Timothy J Hendrickson, Paul Reiners, Lucille A Moore, Anders J Perrone, Dimitrios Alexopoulos, Erik G Lee, Martin Styner, Omid Kardan, Taylor A Chamberlain, Anurima Mummaneni, et al. Bibsnet: A deep learning baby image brain segmentation network for mri scans. *bioRxiv*, 2023. 2
- [36] Malte Hoffmann, David Salat, Martin Reuter, and Bruce Fischl. Longitudinal FreeSurfer with non-linear subject-specific template improves sensitivity to cortical thinning. In *Proceedings of the International Society for Magnetic Resonance in Medicine*, page 1050. ISMRM, 2020. 1, 2
- [37] Malte Hoffmann, Benjamin Billot, Douglas N Greve, Juan Eugenio Iglesias, Bruce Fischl, and Adrian V Dalca. SynthMorph: learning contrast-invariant registration without acquired images. *IEEE transactions on medical imaging*, 41(3):543–558, 2021. 2, 5, 8
- [38] Malte Hoffmann, Benjamin Billot, Juan E Iglesias, Bruce Fischl, and Adrian V Dalca. Learning MRI contrast-agnostic registration. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 899–903. IEEE, 2021. 2
- [39] Malte Hoffmann, Andrew Hoopes, Bruce Fischl, and Adrian V Dalca. Anatomy-specific acquisition-agnostic affine registration learned from fictitious images. In *Medical Imaging 2023: Image Processing*, page 1246402. SPIE, 2023. 2, 5
- [40] Malte Hoffmann, Andrew Hoopes, Douglas N Greve, Bruce Fischl, and Adrian V Dalca. Anatomy-aware and acquisition-agnostic joint registration with SynthMorph. *Imaging Neuroscience*, 2:1–33, 2024. 2
- [41] Andrew Hoopes, Jocelyn S Mora, Adrian V Dalca, Bruce Fischl, and Malte Hoffmann. SynthStrip: skull-stripping for any brain image. *NeuroImage*, 260:119474, 2022. 2, 5
- [42] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012. 2
- [43] Juan Eugenio Iglesias. A ready-to-use machine learning tool for symmetric multi-modality registration of brain mri. *Scientific Reports*, 13(1):6657, 2023. 2
- [44] Juan Eugenio Iglesias, Marco Lorenzi, Sebastiano Ferraris, Loïc Peter, Marc Modat, Allison Stevens, Bruce Fischl, and Tom Vercauteren. Model-based refinement of nonlinear registrations in 3d histology reconstruction. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada,*

- Spain, September 16-20, 2018, *Proceedings, Part II 11*, pages 147–155. Springer, 2018. 2
- [45] IXI Consortium. IXI dataset. <https://brain-development.org/ixi-dataset/>. 5
- [46] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015. 4
- [47] Rohit Jena, Pratik Chaudhari, and James C Gee. Fireants: Adaptive riemannian optimization for multi-scale diffeomorphic matching. *arXiv preprint arXiv:2404.01249*, 2024. 2
- [48] Sarang Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004. 1, 2
- [49] William Kelley, Nathan Ngo, Adrian V Dalca, Bruce Fischl, Lilla Zöllei, and Malte Hoffmann. Boosting skull-stripping performance for pediatric brain images. *ArXiv*, 2024. 2
- [50] Matthew J Kempton, Tracy SA Underwood, Simon Brunton, Floris Stylios, Anne Schmechtig, Ulrich Ettinger, Marcus S Smith, Simon Lovestone, William R Crum, Sophia Frangou, et al. A comprehensive testing protocol for mri neuroanatomical segmentation techniques: evaluation of a novel lateral ventricle segmentation method. *Neuroimage*, 58(4):1051–1059, 2011. 8
- [51] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [52] Iasonas Kokkinos, Michael M Bronstein, Roee Litman, and Alex M Bronstein. Intrinsic shape context descriptors for deformable shapes. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 159–166. IEEE, 2012. 1
- [53] Jannik Kossen, Neil Band, Clare Lyle, Aidan N Gomez, Thomas Rainforth, and Yarin Gal. Self-attention between datapoints: Going beyond individual input-output pairs in deep learning. *Advances in Neural Information Processing Systems*, 34:28742–28756, 2021. 3
- [54] Julian Krebs, Hervé Delingette, Boris Mailhé, Nicholas Ayache, and Tommaso Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging*, 38(9):2165–2176, 2019. 2
- [55] Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *MedRxiv*, pages 2019–12, 2019. 5
- [56] Jack L Lancaster, Diana Tordesillas-Gutiérrez, Michael Martinez, Felipe Salinas, Alan Evans, Karl Zilles, John C Mazziotta, and Peter T Fox. Bias between mni and talairach coordinates analyzed using the icbm-152 brain template. *Human brain mapping*, 28(11):1194–1205, 2007. 1
- [57] Jun Ma, Michael I Miller, Alain Trouvé, and Laurent Younes. Bayesian template estimation in computational anatomy. *NeuroImage*, 42(1):252–261, 2008. 2
- [58] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507, 2007. 5, 8
- [59] Mingyuan Meng, Lei Bi, Michael Fulham, Dagan Feng, and Jinman Kim. Non-iterative coarse-to-fine transformer networks for joint affine and deformable image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 750–760. Springer, 2023. 2
- [60] Marc Modat, Gerard R Ridgway, Zeike A Taylor, Manja Lehmann, Josephine Barnes, David J Hawkes, Nick C Fox, and Sébastien Ourselin. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–284, 2010. 2
- [61] Tony CW Mok, Zi Li, Yingda Xia, Jiawen Yao, Ling Zhang, Jingren Zhou, and Le Lu. Deformable medical image registration under distribution shifts with neural instance optimization. In *International Workshop on Machine Learning in Medical Imaging*, pages 126–136. Springer, 2023. 2
- [62] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5115–5124, 2017. 1
- [63] Sean M Nestor, Raul Rupsingh, Michael Borrie, Matthew Smith, Vittorio Accomazzi, Jennie L Wells, Jennifer Fogarty, Robert Bartha, and Alzheimer’s Disease Neuroimaging Initiative. Ventricular enlargement as a possible measure of alzheimer’s disease progression validated using the alzheimer’s disease neuroimaging initiative database. *Brain*, 131(9):2443–2454, 2008. 8
- [64] Wolfgang M Pauli, Amanda N Nili, and J Michael Tyszka. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Scientific data*, 5(1):1–13, 2018. 2
- [65] Wei Qiu, Lianjin Xiong, Ning Li, Zhangrong Luo, Yaobin Wang, and Yangsong Zhang. AEAU-Net: an unsupervised end-to-end registration network by combining affine transformation and deformable medical image registration. *Medical & Biological Engineering & Computing*, 61(11):2859–2873, 2023. 2
- [66] Martin Reuter and Bruce Fischl. Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage*, 57(1):19–21, 2011. 1, 2
- [67] Martin Reuter, Nicholas J Schmansky, H Diana Rosas, and Bruce Fischl. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4):1402–1418, 2012. 1, 2
- [68] Marc-Michel Rohé, Manasi Datar, Tobias Heimann, Maxime Sermesant, and Xavier Pennec. Svf-net: learning deformable image registration using shape matching. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pages 266–274. Springer, 2017. 2
- [69] Karl Rohr, H Siegfried Stiehl, Rainer Sprengel, Thorsten M Buzug, Jürgen Weese, and MH Kuhn. Landmark-based elas-

- tic registration using approximating thin-plate splines. *IEEE Transactions on medical imaging*, 20(6):526–534, 2001. [2](#)
- [70] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. [3](#)
- [71] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999. [2](#)
- [72] Stephen J Sawiak, Nigel I Wood, Guy B Williams, A Jennifer Morton, and T Adrian Carpenter. Voxel-based morphometry with templates and validation in a mouse model of huntington’s disease. *Magnetic resonance imaging*, 31(9):1522–1531, 2013. [2](#)
- [73] Dieter Seghers, Emiliano D’Agostino, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Construction of a brain template from mr images using state-of-the-art registration and segmentation techniques. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004: 7th International Conference, Saint-Malo, France, September 26–29, 2004. Proceedings, Part I 7*, pages 696–703. Springer, 2004. [2](#)
- [74] Ziyao Shang, Md Asadullah Turja, Eric Feczko, Audrey Houghton, Amanda Rueter, Lucille A Moore, Kathy Snider, Timothy Hendrickson, Paul Reiners, Sally Stoyell, et al. Learning strategies for contrast-agnostic segmentation via synthseg for infant mri data. In *International Conference on Medical Imaging with Deep Learning*, pages 1075–1084. PMLR, 2022. [2](#)
- [75] Hanna Siebert, Christoph Großbröhmer, Lasse Hansen, and Mattias P Heinrich. Convexadam: Self-configuring dual-optimisation-based 3d multitask medical image registration. *IEEE Transactions on Medical Imaging*, 2024. [2](#)
- [76] Matthew Sinclair, Andreas Schuh, Karl Hahn, Kersten Petersen, Ying Bai, James Batten, Michiel Schaap, and Ben Glocker. Atlas-istn: joint segmentation, registration and atlas construction with image-and-spatial transformer networks. *Medical Image Analysis*, 78:102383, 2022. [2](#), [3](#)
- [77] John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998. [2](#)
- [78] Hessam Sokooti, Bob De Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, and Marius Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part I 20*, pages 232–239. Springer, 2017. [2](#)
- [79] Sophie Starck, Vasiliki Sideri-Lampritsa, Bernhard Kainz, Martin Menten, Tamara Mueller, and Daniel Rueckert. Diff-def: Diffusion-generated deformation fields for conditional atlases. *arXiv preprint arXiv:2403.16776*, 2024. [2](#)
- [80] Haosheng Su and Xuan Yang. Nonuniformly Spaced Control Points Based on Variational Cardiac Image Registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 634–644. Springer, 2023. [2](#)
- [81] Nicholas J Tustison, Philip A Cook, Andrew J Holbrook, Hans J Johnson, John Muschelli, Gabriel A Devenyi, Jeffrey T Duda, Sandhitsu R Das, Nicholas C Cullen, Daniel L Gillen, et al. The antsx ecosystem for quantitative biological and medical imaging. *Scientific reports*, 11(1):9068, 2021. [5](#)
- [82] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Automated model-based tissue classification of mr images of the brain. *IEEE transactions on medical imaging*, 18(10):897–908, 1999. [1](#), [2](#)
- [83] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. A unifying framework for partial volume segmentation of brain mr images. *IEEE transactions on medical imaging*, 22(1):105–119, 2003. [2](#)
- [84] Alan Q Wang, M Yu Evan, Adrian V Dalca, and Mert R Sabuncu. A robust and interpretable deep learning framework for multi-modal registration via keypoints. *Medical Image Analysis*, 90:102962, 2023. [2](#)
- [85] Junshen Xu, Daniel Moyer, P Ellen Grant, Polina Golland, Juan Eugenio Iglesias, and Elfar Adalsteinsson. Svort: iterative transformer for slice-to-volume registration in fetal brain mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–13. Springer, 2022. [3](#)
- [86] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration—a deep learning approach. *NeuroImage*, 158:378–396, 2017. [2](#)
- [87] Sean I Young, Yaël Balbastre, Adrian V Dalca, William M Wells, Juan Eugenio Iglesias, and Bruce Fischl. Superwarp: Supervised learning and warping on u-net for invariant subvoxel-precise registration. In *International Workshop on Biomedical Image Registration*, pages 103–115. Springer, 2022. [2](#)
- [88] Yu Zhang, Norbert Schuff, An-Tao Du, Howard J Rosen, Joel H Kramer, Maria Luisa Gorno-Tempini, Bruce L Miller, and Michael W Weiner. White matter damage in frontotemporal dementia and alzheimer’s disease measured by diffusion mri. *Brain*, 132(9):2579–2592, 2009. [8](#)
- [89] Lei Zhao, Shumao Pang, Yangfan Chen, Xiongfeng Zhu, Ziyue Jiang, Zhihai Su, Hai Lu, Yujia Zhou, and Qianjin Feng. SpineRegNet: Spine Registration Network for volumetric MR and CT image by the joint estimation of an affine-elastic deformation field. *Medical Image Analysis*, 86:102786, 2023. [2](#)