Generative Adversarial Registration for Improved Conditional Deformable Templates

Neel Dey
New York University
neel.dey@nyu.edu

Mengwei Ren
New York University
mengwei.ren@nyu.edu

Adrian V. Dalca
MIT, MGH
adalca@mit.edu

Guido Gerig
New York University
gerig@nyu.edu

Abstract

Deformable templates are essential to large-scale medical image registration, segmentation, and population analysis. Current conventional and deep network-based methods for template construction use only regularized registration objectives and often yield templates with blurry and/or anatomically implausible appearance, confounding downstream biomedical interpretation. We reformulate deformable registration and conditional template estimation as an adversarial game wherein we encourage realism in the moved templates with a generative adversarial registration framework conditioned on flexible image covariates. The resulting templates exhibit significant gain in specificity to attributes such as age and disease, better fit underlying group-wise spatiotemporal trends, and achieve improved sharpness and centrality. These improvements enable more accurate population modeling with diverse covariates for standardized downstream analyses and easier anatomical delineation for structures of interest.

1. Introduction

Deformable image registration enables the quantification of geometric dissimilarity via the pairwise warping of a source image to a target. In the context of population studies, pairwise registration of a subject onto a *deformable template* is a central step in standardized analyses, where an ideal template is an unbiased barycentric representation of the (sub-)population of interest [7, 10, 47]. Templates play a key role in diverse large-scale biomedical imaging tasks such as alignment to a common coordinate system [33, 86], brain extraction [38, 43], segmentation [17, 43], and image and shape regression models [31, 67], among others.

While templates can be obtained from a reference database, they are preferably constructed for specific populations by optimizing for an image which minimizes the average deformation to each individual subject. As the template strongly affects subsequent morphometric analysis [82, 89], template construction has received significant attention. Further, as a single template cannot capture the wide structural variability within a population (via age and cohort, for example), we consider *conditional* template estimation with continuous and/or categorical attributes. Conditional templates constructed on image sets with diverse covariates enable sub-population modeling accounting for information learned from the overall population and obviate the need for arbitrary thresholding of demographic information to perform independent analyses [20, 23, 98].

Implicit models for template estimation [2, 7, 10, 47, 58, 96] alternate between registration of each scan to the current template estimate and updating the template based on averages of the warped subject scans. Due to the averaging of aligned image intensities, the resulting templates may blur significantly in regions with high-frequency deformations even alongside shape corrections [10]. Recently, *explicit* template estimation via unsupervised deep networks was proposed in [23] where each stochastic update of a registration network yields a (potentially conditional) template without averaging aligned images and transformations.

However, both implicit and explicit models typically only minimize an image dissimilarity term between the moved template and fixed image (and/or vice-versa) subject to application-specific regularization ensuring a diffeomorphic (smooth, differentiable, and invertible) transformation. As inter-brain variability includes complex topological changes not captured by purely diffeomorphic models, estimated templates are often unrealistic and do not resemble the data that they represent. Sub-optimal appearance impacts downstream applications due to ambiguous and/or implausible anatomical boundaries. For example, in order to register one or more expert-annotated templates to target images for atlas-based segmentation [43, 54, 59], the template(s) must have clearly distinguishable anatomical boundaries to enable expert delineation. Unfortunately, structural anatomical boundaries are often obfuscated by current template estimation approaches.

We present a learning framework to estimate sharp (optionally conditional) templates with realistic anatomy via generative adversarial learning. Our core insight is that in addition to possessing high registration accuracy, the *distribution* of moved template images should be indistinguishable from the real image distribution. We develop a generator comprising of template generation and registration sub-networks and a discriminator which assesses the realism and condition-specificity of the synthesized and warped templates. As adversarial objectives encourage high-frequency detail, the templates gain naturalistic boundaries without the need for ad hoc post-processing. To develop stable and accurate 3D GANs for large medical volumes with highly limited sample and batch sizes, we develop extensive optimization and architectural schemes, augmentation strategies, and conditioning mechanisms.

Our contributions include: (1) a generative adversarial approach to deformable template generation and registration which for the first time uses a realism-based registration regularizer; (2) construction of conditional templates across diverse challenging datasets including neuroimages of pre-term and term-born newborns, adults with and without Huntington's disease, and real-world face images; (3) improvements on current template construction methodologies in terms of centrality and interpretability alongside significantly increased condition-specificity. Code is available at `https://github.com/neel-dey/Atlas-GAN`.

2. Related work

Generative adversarial networks [36] have lead to remarkable progress in high-fidelity image generation [15, 49, 50, 71]. Consequently, GANs for image translation [40, 44, 71, 95, 102] and inverse problems [25, 53, 81] have shown that in addition to reconstruction objectives, adversarial regularizers dramatically increase the visual fidelity of the reconstructions by compensating for high-frequency details typically lost by using reconstruction objectives alone. We apply analogous reasoning in our use of conditional adversarial regularization of registration objectives. For conditional generator networks, modulating every feature map with learned conditional scales and shifts has lead to significantly improved image synthesis [15, 72, 77] over methods where the attribute vector is concatenated to the input.

Deformable image registration is the spatial deformation of a source image to a target. Optical flow registration commonly deployed in computer vision [16, 70] admits deformations which may create anatomically implausible transformations when applied to biomedical images. Instead, a series of more suitable registration algorithms have been developed [21, 28, 46, 78, 87], further leading to several topology-preserving diffeomorphic extensions [8, 12, 19, 91, 93, 99, 84]. More recently, deep networks trained under either supervised [18, 85, 97] or unsupervised [11, 24, 27, 52, 55, 68] registration objectives have emerged, simultaneously offering both greater modeling flexibility and accelerated inference performance.

Generative adversarial registration leveraging simulation has been used in works such as [39] where large-scale finite element simulations of plausible deformations serve as the real domain for a GAN loss alongside supervised registration. Simulated pairs of aligned and mis-aligned image patches have also been used to adversarially optimize a registration network [30, 34]. Our approach is distinct in that we focus on templates and not just registration, we develop adversarial registration techniques accounting for covariates, we process complete 3D volumes and do not use simulation, focusing only on moved template realism.

Template estimation enables standardized analyses of image sets by acting as barycentric representations of a given population. Unconditional template construction has a rich history in medical image analysis [2, 7, 10, 47, 58, 96]. Due to the blurring induced by image and shape averaging of aligned images, popular registration frameworks perform template post-processing and sharpening [10] which may inadvertently create implausible structure [3] and may still fail to resolve structures in highly variable populations. Further, given covariates of interest, ad hoc approaches may ignore shared information by dividing the dataset into sub-populations of interest and constructing templates for each independently. More principled approaches explicitly account for age and potentially other covariates by building spatiotemporal templates and have been extensively validated on pediatric [32, 35, 37, 51, 80, 83] and adult [14, 26, 42, 79] neuroimages.

In this work, we extensively build upon VoxelMorph Templates [23] (referred to as `VXM` in this work). Driven by a generative model, unconditional `VXM` considers a grid of free parameters as a template, which is used together with a training image as input to a registration network [24]. The network estimates a diffeomorphic displacement field between each image and the template. Both the registration network and the template parameters are trained end-to-end under a regularized image matching cost. For conditional `VXM`, a convolutional decoder upsamples an attribute vector to generate a conditional template, which is then similarly end-to-end processed by the registration network. Subsequent sections detail our methodologies and improvements.

3. Methodology

Figure 1 gives an overview of our approach. The generator network (a) & (b) synthesizes a conditional template and deforms it to a fixed image to be assessed by a discriminator (c). The framework is trained end-to-end under a regularized registration and adversarial cost to encourage both registration accuracy and template realism.

Template Generation Sub-network. We develop an architecture whose backbone is agnostic to conditional or unconditional training. For unconditional training, we use a

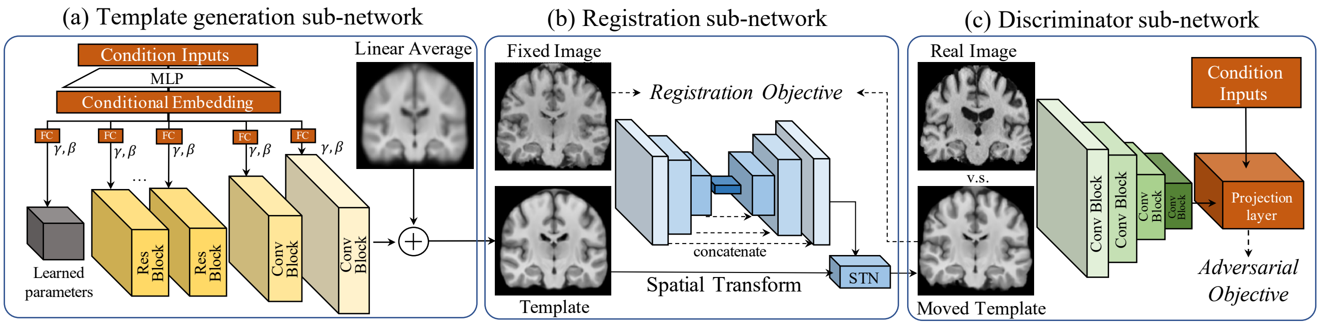


Figure 1. **Overview of the proposed template construction framework**. A template generation network (a) processes an array of learned parameters with a convolutional decoder whose feature-wise affine parameters are learned from input conditions to generate a conditional template. A registration network (b) warps the generated template to a randomly sampled fixed image. A discriminator (c) is trained to distinguish between moved synthesized templates and randomly-sampled images such that realism and condition-specificity is encouraged.

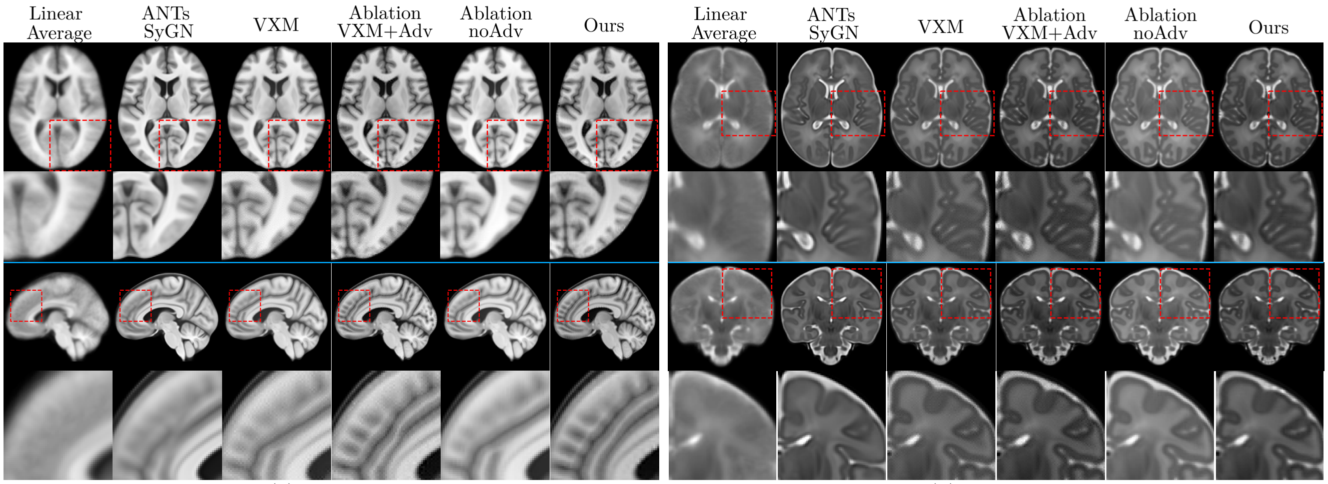
randomly-initialized parameter array (similar to [49, 50]) at half the spatial resolution of the desired template which is processed by convolutional decoder. The decoder output is added to the linear average of training images to generate the unconditional template, such that the network primarily learns to generate high-frequency detail. Checkerboard patterns generated by unconditional VXM are ameliorated in this design by imposing spatial priors through convolutions. However, the central advantage of this backbone architecture is that it enables more parameter-efficient and powerful mechanisms for conditional training, as described below.

For conditional training, given condition vector z and a feature map h_c^i from the ith layer and cth channel, we feature-wise linearly modulate (FiLM [75]) all features h_c^i in the backbone such that $FiLM(h_c^i) = \gamma_c^i(z) * h_c^i + \beta_c^i(z)$, where $\gamma(z)$ and $\beta(z)$ are scale and shift parameters learned from z. As shown in Figure 1(a), we use a four-layer MLP to generate a shared conditional embedding from z which is then linearly projected with weight decay individually to every layer in the template network to generate feature-wise transformations. The primary benefit of this design is that with conditioning at every layer (as opposed to conditional VXM where the only source of conditioning is at its input), the template network has a higher capacity to fit datasets with high variability and synthesize more appropriate templates. A secondary benefit built upon the original VXM architecture is parameter-efficiency. The original VXM design uses a projection from z to a high-dimensional vector at its input. In its neuroimaging experiments, $z \in \mathbb{R}^3$ (i.e., 3 attributes) is projected to a $\mathbb{R}^{\sim 7M}$ vector using a weight matrix with $\sim 21M$ parameters. As the number of conditions increase (with one-hot encoded categorical attributes, e.g.), the rapidly increasing number of parameters in this weight matrix makes learning intractable. Conversely, our framework is relatively insensitive to the dimensionality of the condition vector z, which is processed by a shallow MLP (with 64 units) to generate channel-wise scalars.

Registration Sub-network. We use an established U-Net registration architecture [24], which takes fixed and template images and outputs a time stationary velocity field (SVF) v [4, 5, 65]. When the SVF is integrated over time $t \in [0, 1]$, it yields a diffeomorphic displacement field $\varphi_v^{(t)}$ such that $\frac{\partial \varphi_v^{(1)}}{\partial t} = v(\varphi_v^{(t)})$, where $\varphi_v^{(0)}$ and $\varphi_v^{(1)}$ represent the identity and final displacement fields, respectively. We then use $\varphi_v^{(1)}$ with a spatial transformer [45] to deform the template to the fixed image space.

Discriminator Sub-network. We use a five-layer fully convolutional discriminator network (PatchGAN [44, 102]) to assess realism on local patches of input images. For example, given an input neuroimage volume of $160 \times 192 \times 160$, the discriminator has a receptive field of $63 \times 63 \times 63$. We find that discriminator regularization is essential for stable and artefact-free training, as outlined further below.

For conditional templates, the discriminator is trained to distinguish between real and synthesized images given their categorical and/or continuous covariates. For discriminator conditioning, we build on the projection method [64] commonly used in modern GANs [15, 48] defined as $f(x, y; \theta) = y^T V \phi(x; \theta_\Phi) + \psi(\phi(x; \theta_\Phi); \theta_\Psi)$, where x is the network input, y is the condition, $f(x, y; \theta)$ is the pre-activation discriminator output, $\theta = \{V, \theta_\Phi, \theta_\Psi\}$ are learned parameters such that V is the embedding matrix of y, $\phi(x, \theta_\Phi)$ is the network output given x prior to conditioning, and $\psi(., \theta_\Psi)$ is a scalar function of $\phi(x, \theta_\Phi)$. However, this formulation extends only to either categorical or continuous attributes and does not apply to both types of conditioning, rendering it inadmissible to neuroimaging settings where we are simultaneously interested in attributes such as age (continuous) and disease (often categorical). Fortunately, under mild assumptions of conditional independence of the continuous and categorical attributes given the input, we find that similar analysis to [64] factorizes cleanly into: $f(x, y; \theta) = y_{cat}^T V_{cat} \phi(x; \theta_\Phi) + y_{con}^T V_{con} \phi(x; \theta_\Phi) +$



(a) Predict-HD (b) dHCP

Figure 2. **Unconditional templates** learned from (a) Predict-HD and (b) dHCP. Our synthesized templates yield more neuroanatomically-representative structure (e.g., improved cortical folding for Predict-HD) and appearance (e.g., darker cortical grey matter for dHCP).

$\psi(\phi(x;\theta_\Phi);\theta_\Psi)$, where the *cat* and *con* subscripts indicate categorical and continuous attributes, respectively.

Loss Functions. We define our objective function in the unconditional setting, with straightforward extensions to the conditional scenario. The generator uses a three-part objective including an image matching term, penalties encouraging deformation smoothness and centrality, and an adversarial term encouraging realism in the moved images. For matching, we use a squared localized normalized cross-correlation (LNCC) objective, following standard medical image analysis rationale of requiring intensity standardization in local windows [9]. For deformation regularization, we follow [23] and employ $Reg(\varphi) = \lambda_1\|\bar{u}\|_2^2 + \lambda_2\sum_{p\in\Omega}\|\nabla u(p)\|_2^2 + \lambda_3\sum_{p\in\Omega}\|u(p)\|_2^2$ over voxels p, where u indicates spatial displacement such that $\varphi = Id + u$ and $\bar{u} = \frac{1}{n}\sum_{p\in\Omega} u(p)$ is a moving average of estimated spatial displacements. Briefly, the first term leads to small deformations across the entirety of the dataset, whereas the second and third encourage smooth and small individual deformations, respectively. The adversarial terms used to train generator and discriminator networks correspond to the least-squares GAN [61] objective, chosen for its relative stability. The overall generator loss can be summarized as $L = L_{LNCC} + \lambda_{reg}Reg(\varphi) + \lambda_{GAN}L_{GAN}$, where we use $\lambda_{GAN} = 0.1$ and $\lambda_{reg} = [\lambda_1, \lambda_2, \lambda_3] = [1, 1, 0.01]$ as in [23]. For the discriminator, we employ several forms of regularization, as detailed below.

GAN Stabilization. Generative adversarial training stabilizes and improves with lower image resolutions, higher batch sizes, bigger networks, and higher sample sizes [15]. However, the opposite arises in neuroimaging as images are larger volumes, GPU memory limits training configurations to low batch sizes and small networks, and sample sizes in medical imaging studies are often only a few hundred scans, and thus necessitate careful stabilization. We enforce a 1-Lipschitz constraint on both networks with spectral normalization [63] on every layer, which has been shown to stabilize training and improve gradient feedback to the generator [22]. We further use the R_1 gradient penalty [62] on the discriminator which strongly stabilizes GAN training, defined as $R_1 = \frac{\gamma}{2}\mathbb{E}_{x\sim P_{real}}[\|\nabla D(x)\|_2^2$ where γ is the penalty weight, P_{real} is the real distribution, and D is the discriminator. As discriminator overfitting on limited data is a key cause of GAN instability, we further use differentiable augmentations [48, 90, 100, 101] on *both* real and synthesized images when training the discriminator. We sample random translations for all datasets and further sample from the dihedral D_4 and (a subset of) D_{4h} groups for 2D images and 3D volumes, respectively. Interestingly, brightness/contrast/saturation discriminator augmentations lead to training collapse for neuroimaging datasets, but were found to improve training on a 2D RGB face dataset.

4. Experiments

4.1. Datasets

dHCP. The developing human connectome project (dHCP) provides a dataset of newborns imaged near birth with gestational ages ranging from 29-45 weeks and thus displaying rapid week-to-week structural changes [41]. Spatiotemporal template estimation on dHCP is challenging as premature birth presents decreased cortical folding alongside increased incidences of hemorrhages, hyper-intensities, and lesions [66]. For age-conditioned template construction, we use all 558 T2w MR images from dHCP release 2 preprocessed and segmented by methodologies described in [60]. Images are affine-aligned to a constructed affine template and split at the subject-level (accounting for

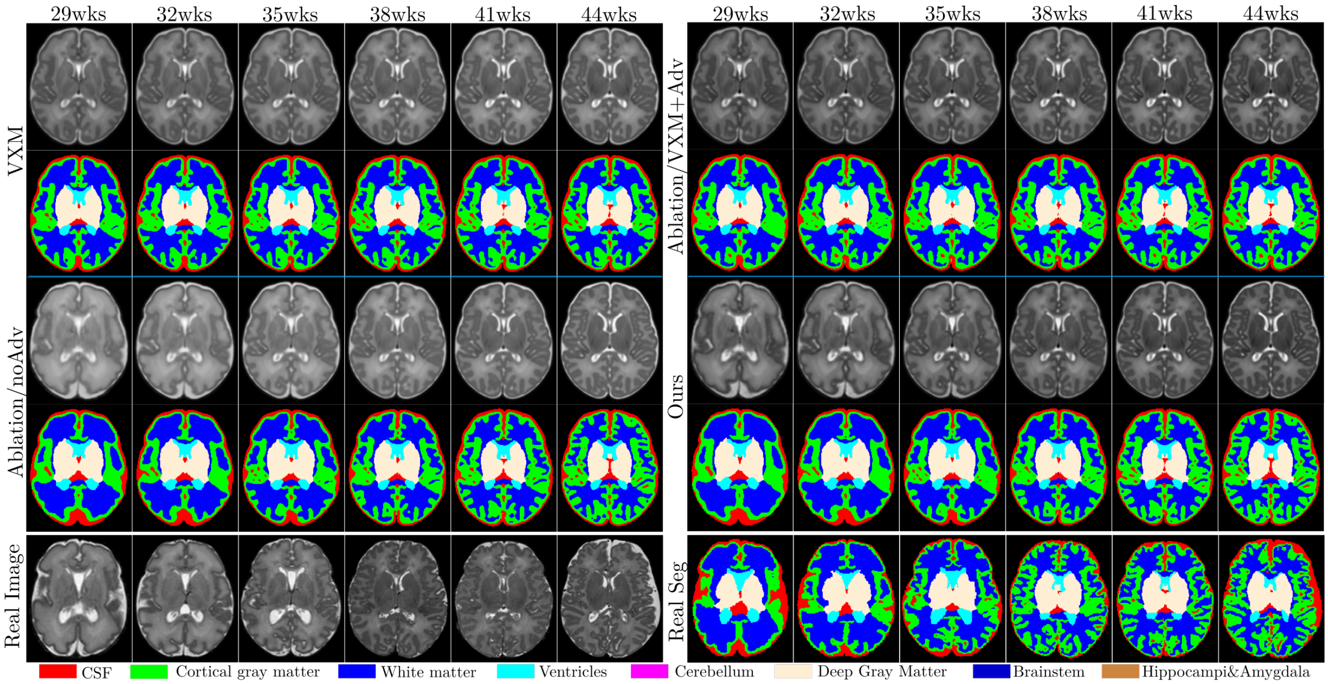


Figure 3. **Age-conditional dHCP templates** alongside template segmentations obtained by [59]. Representative real images and segmentations are visualized in the bottom row.

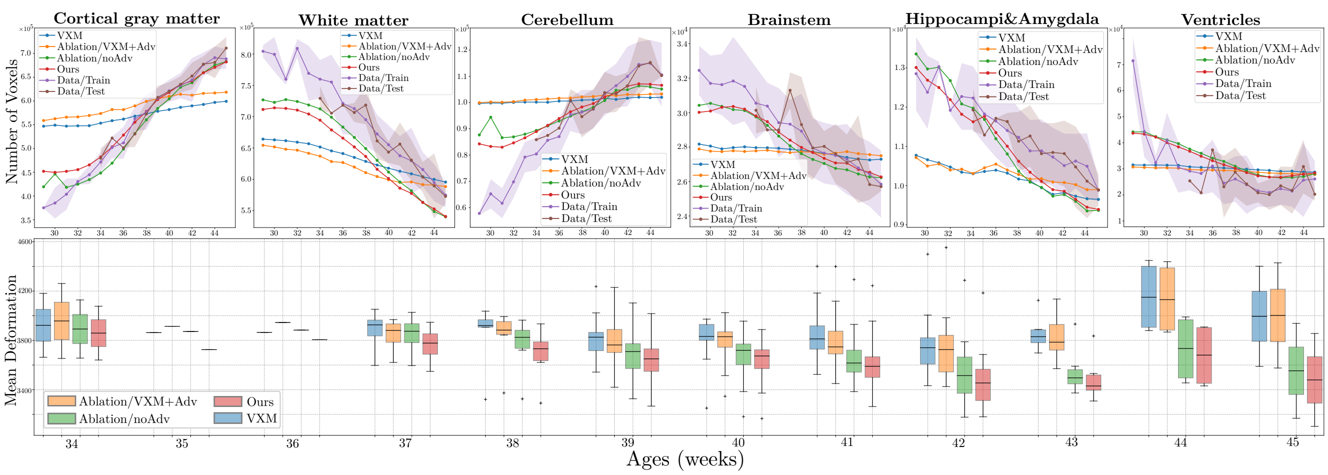


Figure 4. **Top:** volumetric trends of dHCP template segmentations for all methods overlaid upon the volumetric trends for the underlying train (purple) and test (brown) sets. **Bottom:** Mean deformation norms to held-out test data (lower is better) for all conditional methods.

twins and repeat scans), resulting in 458, 15, and 85 scans for training, validation, and testing, respectively.

Predict-HD. We use a longitudinal multi-center and multi-scanner database of healthy controls and subjects with Huntington's disease (HD) [13, 73]. HD is a (typically) adult-onset progressive neurodegenerative disease impairing motor control and cognition [94] which substantially alters brain morphology. We build templates conditioned on age and the presence of the HD genetic mutation. We use 1117 T1w MR images from 388 individuals affine-aligned to MNI [33]. Image preprocessing is described in [74] and image segmentation was performed semi-automatically with labels corresponding to the Neuromorphometrics template [1]. We use 897, 30, 190 images for training, validation, and testing, split at the subject level.

FFHQ-Aging. Face images have been used as experimental vehicles to analyze various qualitative aspects of template construction [10]. We use FFHQ-Aging [69], a database of 70,000 real-world face images providing labels corresponding to (binned) age, gender, and the presence of glasses. FFHQ [49] captures significantly higher variation in terms of age, head pose, and accessories (e.g., hats

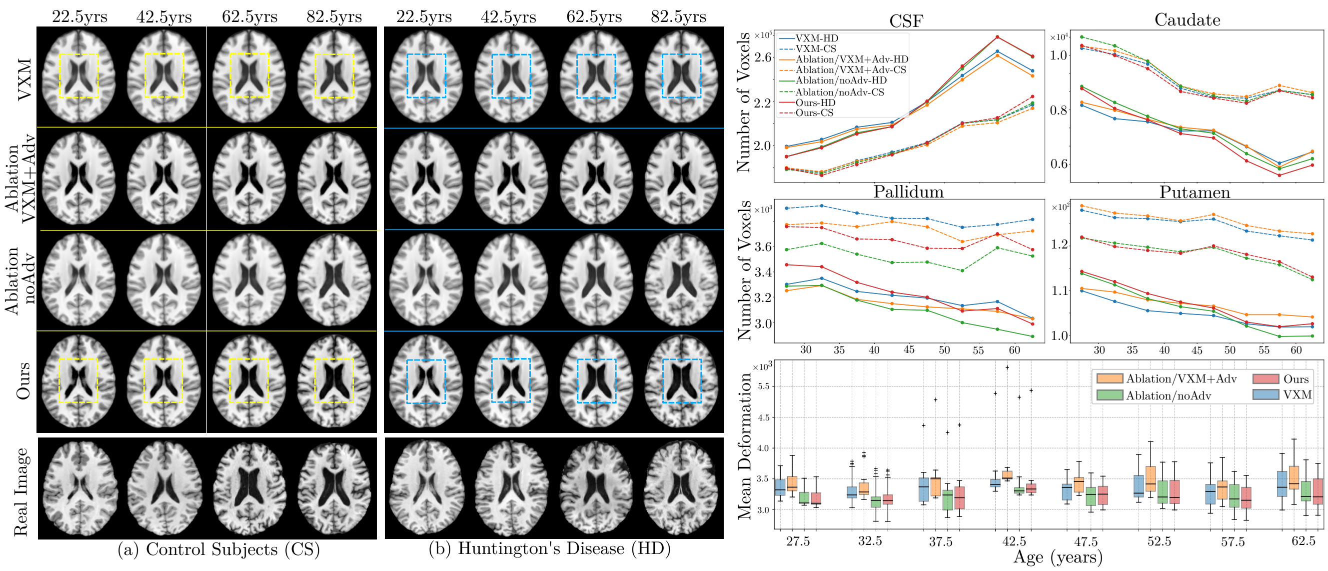


Figure 5. **Left:** Age and cohort-conditional templates for Predict-HD with representative real images visualized in the bottom row. **Top-right:** Inter-cohort volumetric trends from segmented templates for structures of interest for Huntington's disease. **Bottom-right:** Mean deformation norms to held-out test data (lower is better) for all conditional methods.

and costumes) as compared to datasets such as CelebA [57] and is thus a significant challenge. We resize the training images to 128×128, and use age, gender, and the presence of glasses as input conditions. FFHQ-Aging is a challenging dataset, as topological changes (e.g., mouths open or closed) render diffeomorphisms to be a severely limited class of transformations for such images.

4.2. Baselines and Evaluation Strategies

Baselines & Ablations. We first compare with the widely-used unconditional template construction algorithm `SyGN` [10] implemented in the `ANTs` library [92]. We then perform comparisons with a deep network for conditional and unconditional template estimation (`VXM` [23]) trained under its original objective. To isolate our core differences from `VXM`, we use the same registration network for all settings. We use ablated variants to investigate whether adding a discriminator network to the original framework (`Ablation/VXM+Adv`) or whether training our architecture under only a regularized registration cost without a discriminator (`Ablation/noAdv`) yield similar improvements to our combined framework with both architectural changes and discriminator networks (`Ours`). As `Ablation/noAdv` is an ablation of `Ours`, it retains spectral normalization in the template generation branch, which may unnecessarily hamper its performance when trained without an adversarial cost. We do not compare with conventionally-optimized *spatiotemporal* template construction methods as, to our knowledge, there are none that generically apply across diverse image sets (i.e., neonatal T2 MRI, adult T1 MRI, and RGB faces), account for arbitrary covariates, and typically require significant computational resources and domain knowledge.

Evaluation. Constructed templates are difficult to evaluate, as competing properties are often desired. For example, weak deformation regularization enables exact matching of templates to target images at the cost of generating anatomically-impossible deformation magnitudes and irregularities, whereas strong regularization provides smaller and more *central* deformations but produces poor alignment [88]. We posit that preferable templates simultaneously present increased sharpness, accurate alignment, and small and smooth deformations to the target population.

We follow standard methods for quantifying template/MRI sharpness [29, 47, 56, 76] with the entropy focus criterion [6]. To assess registration quality and centrality, we follow the evaluation protocols defined in [23] on held-out test data, including: (1) average Dice coefficients for segmentation labels deformed from the template to the target, corresponding to registration accuracy; (2) mean determinant of the Jacobian matrix $J_\varphi(p)$ of the deformation field φ over voxels p with $|J_\varphi(p)| \leq 0$ indicating local folding of the deformation field and $|J_\varphi(p)| \sim 1$ corresponding to smooth local deformation; (3) average deformation norms to the target images, with lower values indicating improved template centrality given equivalent registration accuracy. Finally, we estimate the norm of the moving average of deformations accumulated over training iterations, with lower values corresponding to increased centrality.

The considered datasets present significant gaps between training and test set age distributions and hence quantitative registration evaluations are performed on a subset of

Table 1. Quantitative evaluations on neuroimaging data for all methods including average Dice scores to test data, norms of accumulated moving average deformations over the course of training, entropy focus criteria (EFC), average Jacobian determinant to test data, and average deformation norms to test data. All deep network methods result in comparable Dice and Jacobian determinants, with our method (`Ours`) demonstrating improvements over baseline `VXM` in essential template qualities such as sharpness and deformation centrality.

Setting		Method	Avg Dice (↑)	‖Mov. Def‖ (↓)	EFC (↓)	Avg $\lvert J_\varphi \rvert$	Avg ‖Def‖ (↓)
Unconditional	Predict-HD	ANTs SyGN	0.75 ± 0.01	-	0.882	1.0000 ± 0.0000	4601 ± 542
		VXM	0.76 ± 0.02	238.83	0.872	0.9987 ± 0.0016	4029 ± 296
		Ablation/VXM+Adv	0.77 ± 0.02	209.42	0.868	0.9986 ± 0.0014	3919 ± 289
		Ablation/noAdv	0.77 ± 0.02	271.00	0.864	0.9979 ± 0.0015	4096 ± 315
		Ours	0.77 ± 0.02	224.68	0.863	0.9987 ± 0.0013	3872 ± 291
Unconditional	dHCP	ANTs SyGN	0.86 ± 0.01	-	0.872	1.0000 ± 0.0000	2687 ± 345
		VXM	0.88 ± 0.01	251.48	0.894	0.9992 ± 0.0013	3883 ± 239
		Ablation/VXM+Adv	0.88 ± 0.01	261.36	0.872	0.9993 ± 0.0014	4029 ± 234
		Ablation/noAdv	0.88 ± 0.01	285.55	0.894	0.9980 ± 0.0013	3987 ± 257
		Ours	0.87 ± 0.01	250.27	0.871	0.9987 ± 0.0012	3815 ± 232
Conditional	Predict-HD	VXM	0.75 ± 0.02	221.68	0.871 ± 0.004	0.9987 ± 0.0013	3867 ± 252
		Ablation/VXM+Adv	0.75 ± 0.02	208.61	0.863 ± 0.004	0.9987 ± 0.0014	3975 ± 286
		Ablation/noAdv	0.75 ± 0.02	199.39	0.874 ± 0.005	0.9987 ± 0.0013	3755 ± 253
		Ours	0.75 ± 0.02	175.01	0.863 ± 0.007	0.9991 ± 0.0013	3756 ± 258
Conditional	dHCP	VXM	0.87 ± 0.01	240.79	0.901 ± 0.006	1.0001 ± 0.0013	3844 ± 230
		Ablation/VXM+Adv	0.87 ± 0.01	215.90	0.887 ± 0.009	0.9991 ± 0.0012	3823 ± 239
		Ablation/noAdv	0.86 ± 0.03	177.81	0.903 ± 0.007	0.9990 ± 0.0015	3666 ± 237
		Ours	0.87 ± 0.01	167.62	0.885 ± 0.009	0.9994 ± 0.0013	3609 ± 226

the overall age range. For Dice evaluation, Predict-HD template segmentation followed [23] and dHCP templates were segmented with [59]. See the appendices for further details.

4.3. Implementation Details

We use a batch size of 1 for 3D neuroimages and a batch size of 32 for 2D planar images. As all datasets considered have highly imbalanced age distributions, we sample minority time-points at higher frequencies when training conditionally. We do not make dataset-specific hyperparameter choices beyond R_1 regularization for GAN stability, where we use $\gamma = 5 \times 10^{-4}, 10^{-3}$, and 5×10^{-3} for Predict-HD, dHCP, and FFHQ-Aging, respectively, corresponding to their respective stabilization requirements. For fair comparison, the hyperparameters and architectures pertaining specifically to registration were matched to the suggested settings for `VXM` for all deep networks. All architectures and remaining design choices are detailed in the appendices.

4.4. Results and Analysis

Figure 2 shows unconditional templates for dHCP and Predict-HD, highlighting that the adversarial approach yields more anatomically-accurate templates. For example, in the sagittal view of the Predict-HD templates (bottom row), we observe anatomical structures more clearly within the red insets when using adversarial regularizers, whereas prior methods are unable to do so. `SyGN` is unable to resolve structures which display high-frequency deformations due to its reliance on averaging aligned intensities and shapes.

For unconditional templates, we observe subtle differences between the proposed method and its reduction `Ablation/VXM+Adv`, with either approach having distinct advantages depending on their intended use. `Ablation/VXM+Adv` trains faster as it does not use 3D convolutions in its template generation branch. Conversely, `Ours` removes subtle checkerboard patterns generated by previous methods. Interestingly, unconditional `Ablation/noAdv` produces stronger moving average deformation magnitudes than other deep network approaches (Table 1), suggesting that the adversarial objective is necessary for this setting to obtain optimal results. However, subsequent analysis of *conditional* template estimation reveals significant differences between these approaches.

Figure 3 provides sample age-conditional templates alongside anatomical segmentations for dHCP. Methods that use input-concatenation architectures for template generation (`VXM` and `Ablation/VXM+Adv`) underfit the rapid week-to-week developmental changes in this dataset, whereas conditionally-modulated architectures (`Ablation/noAdv` and `Ours`) can generate templates which closely follow the underlying trends in the data. **Figure 4** further shows that the conditionally-modulated models better represent volumetric changes in the held-out test set (top row) while showing improved centrality (bottom row). In comparison to `Ablation/noAdv`, the complete framework (`Ours`) better fits spatiotemporal image

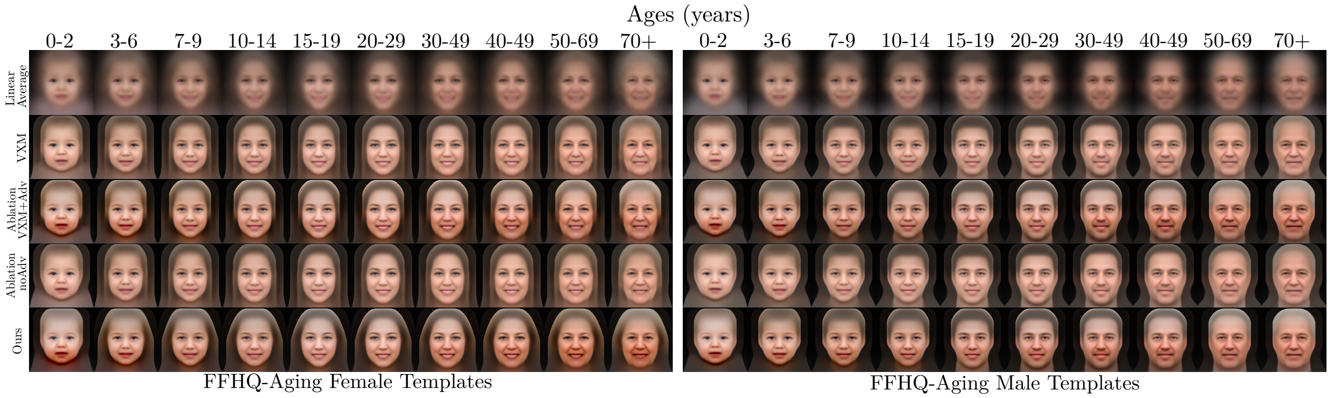


Figure 6. **Age and cohort-conditional FFHQ templates** showing qualitatively improved perceptual fidelity with our framework.

contrast, is sharper, and shows mildly improved centrality (**Table** 1). We underscore that the training and test image volumes correspond to the *affine pre-aligned* data and thus reflect relative volumes in the affine template space.

We observe similar trends in **Figure 5** (left) for Predict-HD. While age and cohort-conditional templates generated by `VXM` and `Ablation/VXM+Adv` do show subtle geometric variation, stronger spatiotemporal changes and dataset-similarity are observed with our complete framework (compare changes in ventricles within the dashed boxes). Analogous improvements in test-set deformation magnitude for Predict-HD are shown in the boxplots (bottom-right). Volume trends of segmented templates for regions pertinent to Huntington's disease are shown in the top-right and follow expected trends such as larger ventricular volumes and smaller basal ganglia volumes in the group with the Huntington's mutation as compared to the controls.

Figure 6 illustrates qualitative FFHQ-Aging templates. Row 1 shows the conditional linear averages of the training set. As above, methods trained without adversarial losses (rows 2 and 4) correctly learn spatiotemporal changes while methods trained with them (rows 3 and 5) present stronger appearance variation and yield improved template perception, e.g. by removing border artefacts in the male cohort. `Ours` further removes border artefacts from the female cohort and increases shape variation and perceptual fidelity (bottom-left row). Further analysis of FFHQ-Aging template construction is presented in the appendices.

Table 1 summarizes quantitative results. All methods achieve comparable Dice coefficients and produce smooth deformations ($|J_\varphi| \sim 1$). However, the proposed techniques generally yield improved sharpness (lower entropy focus criteria) and improved centrality (lower deformation norms) while showing equivalent registration performance, indicating that the constructed templates are more barycentric representations of the data. We stress that Dice coefficients between unconditional and conditional settings cannot be directly compared as the template segmentations were obtained via different approaches. Finally, EFC interpretation requires care. While the numerical differences are subtle, EFC range is more restricted. For example, Gaussian filtering ($\sigma = 1$) of the Predict-HD *unconditional* template from our method only increases EFC from 0.86 to 0.88, indicating that smaller changes are meaningful. Conditional EFC indicate that using a discriminator significantly improves sharpness across scan age ($p < 10^{-5}$ between `VXM` and `Ours` for both datasets), with trends best interpreted via the temporal EFC plots given in the appendices.

5. Conclusions

Explicit methods for the construction of conditional deformable templates using stochastic gradient descent and deep networks are powerful tools for the efficient and flexible modeling of image populations with arbitrary covariates. In this work, we take a generative synthesis approach towards explicit template estimation by constructing a framework which enables training on datasets challenging for generative adversarial networks. The resulting templates are sharp and easy to delineate for domain-experts, are more representative of the underlying demographics, and closely follow typical development in both neonatal MRI of developing pre-term neonates and adult MRI sampled across typical lifespans with and without neurodegeneration. Finally, while this work is motivated from the perspective of neuroimaging, it applies to generic imaging modalities.

Acknowledgements

This work was supported by NIH NS050568, R01EB021391, 1R01AG064027, 1R01HD088125-01A1, 1R01DA038215-01A1, U01NS082086, 1R01AG064027, R01HD055741, 1R01MH118362, 1R34DA050287, R01MH122447, R01ES032294, and NSF MRI1229185. Predict-HD segmentations were provided by the Univ. of Iowa. dHCP is funded by the EU's 7th Framework Programme (FP/2007-13) / ERC Grant Agreement 319456.

References

[1] Neuromorphometrics, Inc. http://www.neuromorphometrics.com/.

[2] Stéphanie Allassonnière, Yali Amit, and Alain Trouvé. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):3–29, 2007.

[3] ANTsX. Fix sharpening artifacts in template construction · issue 956. https://github.com/ANTsX/ANTs/issues/956.

[4] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-euclidean framework for statistics on diffeomorphisms. In Rasmus Larsen, Mads Nielsen, and Jon Sporring, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, pages 924–931, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[5] John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95 – 113, 2007.

[6] D. Atkinson, D. L. G. Hill, P. N. R. Stoyle, P. E. Summers, and S. F. Keevil. Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Transactions on Medical Imaging*, 16(6):903–910, 1997.

[7] Brian Avants and James C Gee. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage*, 23:S139–S150, 2004.

[8] 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.

[9] Brian B. Avants, Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033 – 2044, 2011.

[10] 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.

[11] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

[12] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.

[13] Kevin M Biglan, Ying Zhang, Jeffery Long, Michael Geschwind, Gail Kang, Annie Killoran, Wenjing Lu, Elizabeth McCusker, James A Mills, Lynn A Raymond, et al. Refining the diagnosis of huntington disease: the predict-hd study. *Frontiers in aging neuroscience*, 5:12, 2013.

[14] Alexandre Bône, Olivier Colliot, and Stanley Durrleman. Learning the spatiotemporal variability in longitudinal shape data sets. *International Journal of Computer Vision*, 128(12):2873–2896, July 2020.

[15] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[16] Thomas Brox and Jitendra Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(3):500–513, 2011.

[17] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine*, 104(3):e158–e177, 2011.

[18] Xiaohuan Cao, Jianhua Yang, Jun Zhang, Dong Nie, Minjeong Kim, Qian Wang, and Dinggang Shen. Deformable image registration based on similarity-steered cnn regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 300–308. Springer, 2017.

[19] Yan Cao, Michael I Miller, Raimond L Winslow, and Laurent Younes. Large deformation diffeomorphic metric mapping of vector fields. *IEEE transactions on medical imaging*, 24(9):1216–1230, 2005.

[20] Jieyu Cheng, Adrian V. Dalca, and Lilla Zöllei. Unbiased atlas construction for neonatal cortical surfaces via unsupervised learning. In Yipeng Hu, Roxane Licandro, J. Alison Noble, Jana Hutter, Stephen Aylward, Andrew Melbourne, Esra Abaci Turk, and Jordina Torrents Barrena, editors, *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis*, pages 334–342, Cham, 2020. Springer International Publishing.

[21] G. E. Christensen, S. C. Joshi, and M. I. Miller. Volumetric transformation of brain anatomy. *IEEE Transactions on Medical Imaging*, 16(6):864–877, 1997.

[22] Casey Chu, Kentaro Minami, and Kenji Fukumizu. Smoothness and stability in gans. In *International Conference on Learning Representations*, 2020.

[23] Adrian Dalca, Marianne Rakic, John Guttag, and Mert Sabuncu. Learning conditional deformable templates with convolutional networks. In *Advances in neural information processing systems*, pages 806–818, 2019.

[24] 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.

[25] Giannis Daras, Joseph Dean, Ajil Jalal, and Alexandros G. Dimakis. Intermediate layer optimization for inverse problems using deep generative models, 2021.

[26] Brad C Davis, P Thomas Fletcher, Elizabeth Bullitt, and Sarang Joshi. Population shape regression from random design data. *International journal of computer vision*, 90(2):255–266, 2010.

[27] 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, pages 204–212. Springer, 2017.

[28] Dinggang Shen and C. Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.

[29] Oscar Esteban, Daniel Birman, Marie Schaer, Oluwasanmi O. Koyejo, Russell A. Poldrack, and Krzysztof J. Gorgolewski. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, 12(9):e0184661, sep 2017.

[30] Jingfan Fan, Xiaohuan Cao, Qian Wang, Pew-Thian Yap, and Dinggang Shen. Adversarial learning for mono- or multi-modal registration. *Medical image analysis*, 58:101545, 2019.

[31] James Fishbaugh, Marcel Prastawa, Guido Gerig, and Stanley Durrleman. Geodesic shape regression in the framework of currents. In James C. Gee, Sarang Joshi, Kilian M. Pohl, William M. Wells, and Lilla Zöllei, editors, *Information Processing in Medical Imaging*, pages 718–729, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[32] James Fishbaugh, Martin Styner, Karen Grewen, John Gilmore, and Guido Gerig. Spatiotemporal modeling for image time series with appearance change: Application to early brain development. In Dajiang Zhu, Jingwen Yan, Heng Huang, Li Shen, Paul M. Thompson, Carl-Fredrik Westin, Xavier Pennec, Sarang Joshi, Mads Nielsen, Tom Fletcher, Stanley Durrleman, and Stefan Sommer, editors, *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*, pages 174–185, Cham, 2019. Springer International Publishing.

[33] Vladimir Fonov, Alan C Evans, Kelly Botteron, C Robert Almli, Robert C McKinstry, D Louis Collins, Brain Development Cooperative Group, et al. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1):313–327, 2011.

[34] Yabo Fu, Yang Lei, Tonghe Wang, Kristin Higgins, Jeffrey Bradley, Walter Curran, Tian Liu, and Xiaofeng Yang. Lungregnet: An unsupervised deformable image registration method for 4d-ct lung. *Medical Physics*, 47, 02 2020.

[35] Ali Gholipour, Caitlin K. Rollins, Clemente Velasco-Annis, Abdelhakim Ouaalam, Alireza Akhondi-Asl, Onur Afacan, Cynthia M. Ortinau, Sean Clancy, Catherine Limperopoulos, Edward Yang, Judy A. Estroff, and Simon K. Warfield. A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific Reports*, 7(1), Mar. 2017.

[36] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[37] Piotr A. Habas, Kio Kim, James M. Corbett-Detig, Francois Rousseau, Orit A. Glenn, A. James Barkovich, and Colin Studholme. A spatiotemporal atlas of mr intensity, tissue probability and shape of the fetal brain with application to segmentation. *NeuroImage*, 53(2):460–470, 2010.

[38] Xu Han, Roland Kwitt, Stephen Aylward, Spyridon Bakas, Bjoern Menze, Alexander Asturias, Paul Vespa, John Van Horn, and Marc Niethammer. Brain extraction from normal and pathological images: a joint pca/image-reconstruction approach. *NeuroImage*, 176:431–445, 2018.

[39] Yipeng Hu, Eli Gibson, Nooshin Ghavami, Ester Bonmati, Caroline M Moore, Mark Emberton, Tom Vercauteren, J Alison Noble, and Dean C Barratt. Adversarial deformation regularization for training image registration neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 774–782. Springer, 2018.

[40] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018.

[41] Emer J Hughes, Tobias Winchman, Francesco Padormo, Rui Teixeira, Julia Wurie, Maryanne Sharma, Matthew Fox, Jana Hutter, Lucilio Cordero-Grande, Anthony N Price, et al. A dedicated neonatal brain imaging system. *Magnetic resonance in medicine*, 78(2):794–804, 2017.

[42] W. Huizinga, D.H.J. Poot, M.W. Vernooij, G.V. Roshchupkin, E.E. Bron, M.A. Ikram, D. Rueckert, W.J. Niessen, and S. Klein. A spatio-temporal reference model of the aging brain. *NeuroImage*, 169:11–22, 2018.

[43] Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219, 2015.

[44] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[45] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. *arXiv preprint arXiv:1506.02025*, 2015.

[46] Hans J. Johnson and Gary E. Christensen. Landmark and intensity-based, consistent thin-plate spline image registration. In Michael F. Insana and Richard M. Leahy, editors, *Information Processing in Medical Imaging*, pages 329–343, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[47] Sarang Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.

[48] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data, 2020.

[49] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[50] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.

[51] Heejong Kim, Martin Styner, Joseph Piven, and Guido Gerig. A framework to construct a longitudinal dw-mri infant atlas based on mixed effects modeling of dodf coefficients. In *Computational Diffusion MRI*, pages 149–159. Springer, 2020.

[52] 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.

[53] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017.

[54] Hyeon Woo Lee, Mert R Sabuncu, and Adrian V Dalca. Few labeled atlases are necessary for deep-learning-based segmentation. *arXiv preprint arXiv:1908.04466*, 2019.

[55] Hongming Li and Yong Fan. Non-rigid image registration using fully convolutional networks with deep self-supervision, 2017.

[56] Xiaoxiao Liu, Marc Niethammer, Roland Kwitt, Nikhil Singh, Matt McCormick, and Stephen Aylward. Low-rank atlas image analyses in the presence of pathologies. *IEEE transactions on medical imaging*, 34(12):2583–2591, 2015.

[57] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Large-scale celebfaces attributes (celeba) dataset.

[58] Peter Lorenzen, Marcel Prastawa, Brad Davis, Guido Gerig, Elizabeth Bullitt, and Sarang Joshi. Multi-modal image set registration and atlas formation. *Medical Image Analysis*, 10(3):440–451, 2006. Special Issue on The Second International Workshop on Biomedical Image Registration (WBIR'03).

[59] Antonios Makropoulos, Ioannis S Gousias, Christian Ledig, Paul Aljabar, Ahmed Serag, Joseph V Hajnal, A David Edwards, Serena J Counsell, and Daniel Rueckert. Automatic whole brain mri segmentation of the developing neonatal brain. *IEEE transactions on medical imaging*, 33(9):1818–1831, 2014.

[60] Antonios Makropoulos, Emma C Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.

[61] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[62] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for GANs do actually converge? volume 80 of *Proceedings of Machine Learning Research*, pages 3481–3490, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

[63] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018.

[64] Takeru Miyato and Masanori Koyama. cGANs with projection discriminator. In *International Conference on Learning Representations*, 2018.

[65] M. Modat, P. Daga, M. J. Cardoso, S. Ourselin, G. R. Ridgway, and J. Ashburner. Parametric non-rigid registration using a stationary velocity field. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 145–150, 2012.

[66] Zoltán Molnár and Mary Rutherford. Brain maturation after preterm birth. *Science translational medicine*, 5(168):168ps2–168ps2, 2013.

[67] Marc Niethammer, Yang Huang, and François-Xavier Vialard. Geodesic regression for image time-series. In *International conference on medical image computing and computer-assisted intervention*, pages 655–662. Springer, 2011.

[68] Marc Niethammer, Roland Kwitt, and Francois-Xavier Vialard. Metric learning for image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8463–8472, 2019.

[69] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis, 2020.

[70] Nils Papenberg, Andrés Bruhn, Thomas Brox, Stephan Didas, and Joachim Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, Jan. 2006.

[71] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[72] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019.

[73] JS Paulsen, Douglas R Langbehn, Julie C Stout, Elizabeth Aylward, Christopher A Ross, Martha Nance, Mark Guttman, Shannon Johnson, M MacDonald, Leigh J Beglinger, et al. Detection of huntington's disease decades before diagnosis: the predict-hd study. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(8):874–880, 2008.

[74] Jane S Paulsen, Jeffrey D Long, Christopher A Ross, Deborah L Harrington, Cheryl J Erwin, Janet K Williams, Holly James Westervelt, Hans J Johnson, Elizabeth H Aylward, Ying Zhang, et al. Prediction of manifest huntington's disease with clinical and imaging measures: a prospective observational study. *The Lancet Neurology*, 13(12):1193–1201, 2014.

[75] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron C Courville. Film: Visual reasoning with a general conditioning layer. In *AAAI*, 2018.

[76] Deepthi Rajashekar, Matthias Wilms, M. Ethan MacDonald, Jan Ehrhardt, Pauline Mouches, Richard Frayne,

Michael D. Hill, and Nils D. Forkert. High-resolution t2-FLAIR and non-contrast CT brain atlas of the elderly. *Scientific Data*, 7(1), Feb. 2020.

[77] M. Ren, N. Dey, J. Fishbaugh, and G. Gerig. Segmentation-renormalized deep feature modulation for unpaired image harmonization. *IEEE Transactions on Medical Imaging*, pages 1–1, 2021.

[78] 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.

[79] Jean-Baptiste Schiratti, Stéphanie Allassonniere, Olivier Colliot, and Stanley Durrleman. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Neural Information Processing Systems*, number 28, 2015.

[80] Andreas Schuh, Antonios Makropoulos, Emma C Robinson, Lucilio Cordero-Grande, Emer Hughes, Jana Hutter, Anthony N Price, Maria Murgasova, Rui Pedro AG Teixeira, Nora Tusor, et al. Unbiased construction of a temporally consistent morphological atlas of neonatal brain development. *bioRxiv*, page 251512, 2018.

[81] Maximilian Seitzer, Guang Yang, Jo Schlemper, Ozan Oktay, Tobias Würfl, Vincent Christlein, Tom Wong, Raad Mohiaddin, David Firmin, Jennifer Keegan, Daniel Rueckert, and Andreas Maier. Adversarial and perceptual refinement for compressed sensing mri reconstruction. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 232–240, Cham, 2018. Springer International Publishing.

[82] Matthew L Senjem, Jeffrey L Gunter, Maria M Shiung, Ronald C Petersen, and Clifford R Jack Jr. Comparison of different methodological implementations of voxel-based morphometry in neurodegenerative disease. *Neuroimage*, 26(2):600–608, 2005.

[83] Ahmed Serag, Paul Aljabar, Gareth Ball, Serena J. Counsell, James P. Boardman, Mary A. Rutherford, A. David Edwards, Joseph V. Hajnal, and Daniel Rueckert. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *NeuroImage*, 59(3):2255–2265, 2012.

[84] Zhengyang Shen, François-Xavier Vialard, and Marc Niethammer. Region-specific diffeomorphic metric mapping. In *Advances in Neural Information Processing Systems*, pages 1098–1108, 2019.

[85] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–239. Springer, 2017.

[86] J Talairach. 3-dimensional proportional system; an approach to cerebral imaging. co-planar stereotaxic atlas of the human brain. *Thieme*, pages 1–122, 1988.

[87] J-P Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical image analysis*, 2(3):243–260, 1998.

[88] B.T. Thomas Yeo, Mert R. Sabuncu, Rahul Desikan, Bruce Fischl, and Polina Golland. Effects of registration regularization and atlas sharpness on segmentation accuracy. *Medical Image Analysis*, 12(5):603–615, 2008. Special issue on the 10th international conference on medical imaging and computer assisted intervention - MICCAI 2007.

[89] Paul M Thompson, Roger P Woods, Michael S Mega, and Arthur W Toga. Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. *Human brain mapping*, 9(2):81–92, 2000.

[90] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. On data augmentation for gan training, 2020.

[91] Nicholas J. Tustison and Brian B. Avants. Diffeomorphic directly manipulated free-form deformation image registration via vector field flows. In Benoît M. Dawant, Gary E. Christensen, J. Michael Fitzpatrick, and Daniel Rueckert, editors, *Biomedical Image Registration*, pages 31–39, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[92] Nicholas J Tustison, Philip A Cook, Andrew J Holbrook, Hans J Johnson, John Muschelli, Gabriel A Devanyi, Jeffrey T Duda, Sandhitsu R Das, Nicholas C Cullen, Daniel L Gillen, et al. Antsx: A dynamic ecosystem for quantitative biological and medical imaging. *medRxiv*, 2020.

[93] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1, Supplement 1):S61–S72, 2009. Mathematics in Brain Imaging.

[94] Francis O Walker. Huntington's disease. *The Lancet*, 369(9557):218–228, 2007.

[95] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[96] Guorong Wu, Hongjun Jia, Qian Wang, and Dinggang Shen. Sharpmean: groupwise registration guided by sharp mean image and tree-based registration. *NeuroImage*, 56(4):1968–1981, 2011.

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

[98] Evan M. Yu, Adrian V. Dalca, and Mert R. Sabuncu. Learning conditional deformable shape templates for brain anatomy. In Mingxia Liu, Pingkun Yan, Chunfeng Lian, and Xiaohuan Cao, editors, *Machine Learning in Medical Imaging*, pages 353–362, Cham, 2020. Springer International Publishing.

[99] Miaomiao Zhang and P Thomas Fletcher. Fast diffeomorphic image registration via fourier-approximated lie algebras. *International Journal of Computer Vision*, 127(1):61–73, 2019.

[100] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient gan training, 2020.

[101] Zhengli Zhao, Zizhao Zhang, Ting Chen, Sameer Singh, and Han Zhang. Image augmentations for gan training, 2020.

[102] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.